

# **Classificação e Análise do Uso do Solo Com Base em Pontos de Interesse**

Relatório de projeto apresentado para a obtenção do grau de  
Mestre em Informática e Sistemas  
Especialização em Tecnologias da Informação e do Conhecimento

**Autor**

**Renato Ribeiro de Andrade**

**Orientadores**

**Prof. Dra. Ana Cristina da Costa Oliveira Alves**

Professora do Departamento de Engenharia Informática e de Sistemas  
Instituto Superior de Engenharia de Coimbra

**Prof. Dr. Carlos Lisboa Bento**

Professor do Departamento de Engenharia Informática  
Universidade de Coimbra



E que o mínimo que a gente faça seja, a cada momento,  
o melhor que afinal se conseguiu fazer.

Lya Luft, 2004



Este trabalho é dedicado à minha amada irmã,  
sem a qual eu, de maneira alguma, teria chegado onde cheguei;  
por ter assumido o papel de mãe quando a nossa,  
por motivos alheios à nossa vontade, já não pôde estar presente;  
pelo trabalho ao qual desde muito cedo se dedicou para me sustentar;  
por ter aberto mão de incontáveis oportunidades na vida, em favor da minha criação;  
pelas noites de sono que perdeu a cuidar de mim;  
e pelo inexplicável amor e imensurável carinho que me dedica,  
apesar da distância geográfica que agora nos separa.



# Agradecimentos

A realização deste projeto não teria sido possível sem a ajuda e o suporte de várias pessoas, dentre as quais algumas se destacam.

Primeiramente, gostaria de agradecer a Deus por me permitir chegar até aqui e por se fazer presente a cada instante da minha vida, por me ajudar a enfrentar as dificuldades do dia a dia; pela esperança que me oferece constantemente e por me mostrar o caminho a seguir, mesmo quando as incertezas parecem se sobrepor à autoconfiança.

Gostaria ainda de expressar a minha gratidão a todos os meus professores, por todo o conhecimento compartilhado e sobretudo aos meus orientadores, Professora Doutora Ana Alves e Professor Doutor Carlos Bento, pelo esforço, dedicação, empenho, comprometimento e pelas constantes preocupações com a qualidade e relevância científica do meu trabalho.

Aos meus irmãos e ao meu pai, por todas as vezes que abriram mão das suas próprias necessidades em favor das minhas, pelo seu papel na minha educação, por acreditarem no meu potencial e por me apoiarem nos momentos mais difíceis da minha caminhada, mesmo agora que estou geograficamente tão distante.

A um grande amigo e companheiro de muitas jornadas, com quem, direta ou indiretamente, tenho dividido os principais momentos da vida; pela convivência diária, pela ajuda e suporte que me dedicou, e por compreender as dificuldades que enfrentei durante realização das várias etapas deste curso de mestrado, sobretudo no desenvolvimento deste projeto.

À minha falecida tia, que embora tenha tido um papel fundamental na minha infância e no meu crescimento, partiu ao fim do segundo semestre do curso, sem que eu tivesse sequer a oportunidade de me despedir; por me incentivar a abraçar cada oportunidade que surgiu no meu caminho e pelo orgulho de quem sou, que sempre fez questão de demonstrar.

A todos aqueles que de alguma maneira contribuíram com o meu crescimento pessoal e intelectual e que estiveram presentes nos momentos importantes da minha vida, por toda e qualquer ajuda ou suporte que me tenham dado, principalmente em relação à minha jornada acadêmica.

Finalmente, mas não menos importante, gostaria ainda de destacar a minha gratidão aos meus amigos, que estiveram sempre prontos a me ajudar nas mais diversas situações; e aos meus colegas de curso, pelos trabalhos em grupo, pelas aulas de português, e pelos momentos de alegria e descontração que juntos pudemos desfrutar.



# Resumo

A gestão dos espaços urbanos tem se tornado um desafio cada vez maior para a administração pública. A evolução das cidades ocasiona um aumento significativo na sua complexidade, em termos de distribuição geográfica. Com o crescimento da população, muitas regiões enfrentam problemas no que diz respeito ao planeamento urbano. À medida que as mudanças ocorrem na dinâmica das grandes metrópoles, garantir a qualidade de vida das pessoas, a sustentabilidade e o equilíbrio económico dos diversos setores, revela-se uma crescente preocupação. Por outro lado, com as recentes evoluções da tecnologia, surgiram também oportunidades para colmatar as necessidades enfrentadas neste contexto. A disseminação dos sensores remotos levou a um aumento exponencial no volume de dados georreferenciados gerados diariamente. Aliado ao avanço da capacidade computacional e à modernização dos Sistemas de Informação Geográfica (SIG), a disponibilidade dos dados georreferenciados impulsionou a adoção de um vasto conjunto de metodologias, utilizadas para as mais variadas finalidades nos estudos dos espaços urbanos.

Este trabalho foi desenvolvido com o objetivo de explorar o potencial dos dados colaborativos, nomeadamente os Pontos de Interesse (POIs), para a caracterização das regiões geográficas. O foco principal do estudo foi a região metropolitana de Lisboa, por ser uma das zonas mais importantes do país. Foi desenvolvida uma aplicação que realiza a coleta automática dos POIs na Internet. Os dados coletados foram tratados e organizados para permitir a sua utilização. Com a ajuda de outros dois conjuntos de dados de validação, foi implementado um método para a classificação dos espaços, onde foram testadas diferentes técnicas. A área estudada foi dividida em unidades que representam áreas de tamanho padrão e cada unidade foi classificada quanto ao seu tipo de uso ou cobertura. Como forma de aprofundar o estudo, foi proposto um método automático que permite a análise das correlações existentes entre os tipos de uso e cobertura dos espaços e as mais diversas categorias de empresas, serviços e POIs localizados na região de estudo. Foi ainda proposto um segundo método, também automático, que permite a análise das correlações entre as diferentes categorias de POIs.

O estudo realizado é um contributo para a administração pública, considerando que constitui uma alternativa aos métodos atuais de caracterização do uso e da ocupação dos espaços, que envolvem um elevado investimento financeiro. Além disso as metodologias propostas para a análise das correlações podem ser empregadas também no setor privado, como forma de produzir conhecimento para ajudar a suportar a tomada de decisões no que diz respeito à escolha das localizações mais adequadas para cada tipo de empresa ou serviço.

**Palavras-Chave:** espaços urbanos, dados georreferenciados, *data mining*, *machine learning*, correlação espacial cruzada.



# Abstract

The management of urban spaces has become an increasing challenge for public administration. The evolution of cities has a significant impact on their complexity, in terms of geographic distribution. With population growth, many regions face problems related to urban planning. As changes occur in the dynamics of big cities, ensuring people's quality of life, sustainability and the economic balance of the various economic sectors is becoming a growing concern. On the other hand, with recent developments in technology, new opportunities have also emerged to address the needs faced in this context. The popularisation of remote sensors has led to an exponential increase in the volume of georeferenced data generated daily. Combined with the advancement of computational power and the modernisation of Geographic Information Systems (GIS), the availability of georeferenced data has driven the adoption of a wide range of methodologies, used for the most varied purposes in the study of urban spaces.

This work was developed with the objective of exploring the potential of collaborative data, namely Points of Interest (POIs), to characterise geographic spaces. The main focus of the study was the metropolitan region of Lisbon, as one of the most important areas in the country. An application was created to collect POIs automatically on the Internet. The collected data was processed and organised to make it ready for use. Adopting two other well-established datasets as ground truth data, a method for land use classification was proposed and implemented, in which different techniques were tested. The studied region was divided into units that represent areas of standard size and each unit was classified according to its type of use or coverage. As a way to deepen the study, another automatic method was proposed and applied through the analysis of the correlations between land use/land cover and the most diverse categories of companies, services and POIs located in the study region. A second method, also automatic, was proposed, which allows the analysis of the correlations between the different categories of POIs.

The study carried out is a contribution to public administration, considering that it constitutes an alternative to the current methods of characterising land use and land cover, which involve a high monetary investment. In addition, the proposed methodologies for analysis of correlations can also be used in the private sector, as a way of producing knowledge to support the decision-making process, with regard to choice of the most suitable locations for each type of company or service.

**Keywords:** land use and land cover analysis, georeferenced data, data mining, machine learning, spatial cross correlation.



# Índice

Lista de Figuras .....	15
Lista de Tabelas .....	17
Acrónimos .....	21
1 Introdução .....	1
1.1 Enquadramento .....	1
1.2 Motivação .....	1
1.3 Objetivos .....	2
1.4 Metodologia do trabalho .....	3
1.5 Principais contributos .....	3
1.5.1 Publicações .....	4
1.6 Estrutura do documento .....	5
2 Fundamentação teórica .....	7
2.1 Caracterização do uso dos espaços .....	7
2.1.1 Dados .....	9
2.1.2 Métodos .....	10
2.1.3 Combinações de dados e métodos .....	17
2.2 Mapas de densidade/concentração ( <i>heatmaps</i> ) .....	20
2.3 Correlação espacial .....	21
2.3.1 Autocorrelação espacial .....	21
2.3.2 Correlação espacial cruzada .....	22
2.3.3 Coeficiente de correlação de Pearson .....	24
2.3.4 Coeficiente de correlação ponto-bisserial .....	25
3 Ferramentas e tecnologias .....	27
3.1 Docker .....	27
3.2 PostgreSQL .....	28

3.3	PostGIS.....	29
3.4	QGIS.....	30
3.5	RapidMiner.....	31
3.6	Linguagens de programação.....	32
3.6.1	PHP .....	32
3.6.2	Java .....	33
3.6.3	Python .....	33
3.6.4	R.....	34
4	Dados .....	35
4.1	Pontos de interesse .....	35
4.1.1	Facebook Places.....	35
4.1.2	O dataset.....	36
4.2	Dados de uso e ocupação dos espaços ( <i>Land Use/Land Cover</i> ) .....	51
4.2.1	CORINE.....	52
4.2.2	Carta de Uso e Ocupação do Solo - COS .....	54
4.3	Tabelas iniciais da base de dados.....	57
5	Extração de Conhecimento .....	59
5.1	Classificação do uso dos espaços .....	59
5.1.1	O software desenvolvido para a geração dos <i>datasets</i> .....	60
5.1.2	Experiências baseadas no COS.....	60
5.1.3	Experiências baseadas no CORINE.....	72
5.2	Análise das correlações .....	90
5.2.1	Mapas de densidade .....	90
5.2.2	Experiências e resultados.....	93
6	Conclusões .....	115
6.1	Revisão geral do tema .....	115
6.2	Revisão dos objetivos e do trabalho realizado .....	115
6.3	Contributos .....	117
6.3.1	Contributos teóricos.....	117
6.3.2	Contributos práticos.....	117

6.4	Limitações .....	118
6.5	Trabalho futuro.....	119
7	Referências.....	121
	Apêndice A. <i>Scripts</i> SQL.....	129
	Apêndice B. Formatos dos ficheiros utilizados .....	145
	Apêndice C. Modelo da base de dados .....	147
	Apêndice D. Parametrizações .....	149
	Apêndice E. Matrizes de confusão.....	151
	Apêndice F. Conjuntos de atributos.....	159
	Apêndice G. Categorias dos Pontos de Interesse.....	185
	Apêndice H. Correlações .....	187



# Lista de Figuras

Figura 1: Topologia de uma rede neuronal simples.....	15
Figura 2: Duas possíveis configurações para a grelha de coordenadas. ....	39
Figura 3: Fluxograma do algoritmo de criação da grelha. ....	40
Figura 4: Disposição adotada para a definição dos pontos da grelha. ....	41
Figura 5: Grelha inicial de coordenadas. ....	42
Figura 6: Grelha final de coordenadas. ....	42
Figura 7: Distribuição dos POIs das subcategorias da categoria “Non Business Places”. ....	48
Figura 8: Pontos de interesse na categoria “Business” e em suas subcategorias.....	49
Figura 9: Total de registos na categoria “Business” e nas suas subcategorias. ....	50
Figura 10: Distribuição dos POIs das categorias “Food & Beverage” e “Local Service”.....	51
Figura 11: Amostra dos polígonos do COS. ....	61
Figura 12: Exemplo da proximidade geográfica entre os espaços comerciais e industriais.....	71
Figura 13: Exemplo da dimensão dos polígonos do CORINE. ....	73
Figura 14: <i>Grid</i> construída para as experiências baseadas no CORINE. ....	73
Figura 15: Distribuição geográfica dos POIs das subcategorias da categoria “Business”. ....	86
Figura 16: Exemplo da ocorrência dos POIs das categorias selecionadas manualmente. ....	87
Figura 17: Exemplo da proximidade geográfica entre algumas classes.....	88
Figura 18: Comparação entre <i>heatmaps</i> com raios diferentes. ....	107
Figura 19: POIs de algumas categorias em relação a determinadas classes. ....	108
Figura 20: Distribuição dos POIs de algumas categorias em relação à classe “121”. ....	109
Figura 21: Lista ordenada das categorias obtidas no site do Facebook. ....	145
Figura 22: Mapa de categorias. ....	146
Figura 23: Principais tabelas da base de dados. ....	147
Figura 24: Matriz de confusão do NB com <i>cross-validation</i> , no cenário 1, do COS. ....	151
Figura 25: Matriz de confusão do RF com <i>cross-validation n</i> , no cenário 1, do COS.....	151
Figura 26: Matriz de confusão do SVM com <i>cross-validation</i> , no cenário 1, do COS.....	151
Figura 27: Matriz de confusão do NN com <i>cross-validation</i> , no cenário 1, do COS.....	151
Figura 28: Matriz de confusão do NB, com <i>split validation</i> , no cenário 1, do COS. ....	152
Figura 29: Matriz de confusão do NN com <i>split validation</i> , no cenário 1, do COS.....	152
Figura 30: Matriz de confusão do NB com <i>cross-validation</i> , no cenário 2, do COS. ....	152
Figura 31: Matriz de confusão do RF com <i>cross-validation</i> , no cenário 2, do COS.....	152
Figura 32: Matriz de confusão do SVM com <i>cross-validation</i> no cenário 2, do COS. ....	152
Figura 33: Matriz de confusão do NN com <i>cross-validation</i> , no cenário 2, do COS.....	153

Figura 34: Matriz de confusão do NB com <i>split validation</i> , no cenário 2, do COS. ....	153
Figura 35: Matriz de confusão do NN com <i>split validation</i> , no cenário 2, do COS. ....	153
Figura 36: Matriz de confusão do NB com <i>cross-validation</i> , no cenário 3, do COS. ....	153
Figura 37: Matriz de confusão do RF com <i>cross-validation</i> , no cenário 3, do COS. ....	153
Figura 38: Matriz de confusão do SVM com <i>cross-validation</i> , no cenário 3, do COS. ....	154
Figura 39: Matriz de confusão do NN com <i>cross-validation</i> , no cenário 3, do COS. ....	154
Figura 40: Matriz de confusão do NB com <i>split validation</i> , no cenário 3, do COS. ....	154
Figura 41: Matriz de confusão do NN com <i>split validation</i> , no cenário 3, do COS. ....	154
Figura 42: Matriz de confusão do NB com <i>cross-validation</i> , no cenário 4, do COS. ....	154
Figura 43: Matriz de confusão do RF com <i>cross-validation</i> no cenário 4, do COS. ....	155
Figura 44: Matriz de confusão do SVM com <i>cross-validation</i> , no cenário 4, do COS. ....	155
Figura 45: Matriz de confusão do NN com <i>cross-validation</i> , no cenário 4, do COS. ....	155
Figura 46: Matriz de confusão do NB com <i>split validation</i> , no cenário 4, do COS. ....	155
Figura 47: Matriz de confusão do NN com <i>split validation</i> , no cenário 4, do COS. ....	155
Figura 48: Matriz de confusão do NB com <i>cross-validation</i> , no cenário 5, do COS. ....	155
Figura 49: Matriz de confusão do RF com <i>cross-validation</i> , no cenário 5, do COS. ....	156
Figura 50: Matriz de confusão do SVM com <i>cross-validation</i> no cenário 5, do COS. ....	156
Figura 51: Matriz de confusão do NN com <i>cross-validation</i> , no cenário 5, do COS. ....	156
Figura 52: Matriz de confusão do NB com <i>split validation</i> , no cenário 5, do COS. ....	156
Figura 53: Matriz de confusão do NN com <i>split validation</i> , no cenário 5, do COS. ....	156
Figura 54: Matriz de confusão obtida no cenário 1, do CORINE. ....	156
Figura 55: Matriz de confusão obtida no cenário 2, do CORINE. ....	157
Figura 56: Matriz de confusão obtida no cenário 2, do CORINE. ....	157
Figura 57: matriz de confusão obtida no cenário 4, do CORINE. ....	157
Figura 58: Matriz de confusão obtida no teste 1, do cenário 5, do CORINE. ....	157
Figura 59: Matriz de confusão obtida no teste 2, do cenário 5, do CORINE. ....	157
Figura 60: Matriz de confusão obtida no teste 3, do cenário 5, do CORINE. ....	157
Figura 61: Matriz de confusão obtida no teste 4, do cenário 5, do CORINE. ....	158
Figura 62: Matriz de confusão obtida no teste 5, do cenário 5, do CORINE. ....	158
Figura 63: Matriz de confusão obtida no teste 6, do cenário 5, do CORINE. ....	158
Figura 64: Matriz de confusão obtida no teste 7, do cenário 5, do CORINE. ....	158

# Lista de Tabelas

Tabela 1: Dados e métodos utilizados nos estudos analisados. ....	18
Tabela 2: Descrição dos atributos do <i>dataset</i> utilizado no estudo. ....	37
Tabela 3: Categorias de níveis 1 e 2 dos POIs. ....	46
Tabela 4: Classes do CORINE <i>land cover</i> . ....	52
Tabela 5: Classes do <i>dataset</i> COS. ....	55
Tabela 6: Cenários de classificação realizados com o <i>dataset</i> COS. ....	61
Tabela 7: Resultados obtidos no cenário 1, dos testes com o COS. ....	62
Tabela 8: <i>Precision</i> e <i>recall</i> obtidos com <i>cross-validation</i> no cenário 1, com o COS. ....	63
Tabela 9: <i>Precision</i> e <i>recall</i> obtidos com <i>split validation</i> no cenário 1, como COS. ....	63
Tabela 10: Resultados obtidos no cenário 2, dos testes com o COS. ....	64
Tabela 11: <i>Precision</i> e <i>recall</i> obtidos com <i>cross-validation</i> no cenário 2, do COS. ....	64
Tabela 12: <i>Precision</i> e <i>recall</i> obtidos com <i>split validation</i> no cenário 2, do COS. ....	65
Tabela 13: Resultados obtidos no cenário 3, dos testes com o COS. ....	65
Tabela 14: <i>Precision</i> e <i>recall</i> obtidos com <i>cross-validation</i> no cenário 3, do COS. ....	66
Tabela 15: <i>Precision</i> e <i>recall</i> obtidos com <i>split validation</i> no cenário 3, do COS. ....	66
Tabela 16: Resultados obtidos no cenário 4, dos testes com o COS. ....	67
Tabela 17: <i>Precision</i> e <i>recall</i> obtidos com <i>cross-validation</i> no cenário 4, com o COS. ....	67
Tabela 18: <i>Precision</i> e <i>recall</i> obtidos com <i>split validation</i> no cenário 4, do COS. ....	67
Tabela 19: Resultados obtidos no cenário 5, dos testes com o COS. ....	69
Tabela 20: <i>Precision</i> e <i>recall</i> obtidos com <i>cross-validation</i> no cenário 5, do COS. ....	69
Tabela 21: <i>Precision</i> e <i>recall</i> obtidos com <i>split validation</i> no cenário 5, do COS. ....	69
Tabela 22: Reagrupamento das classes do CORINE. ....	74
Tabela 23: Cenários de classificação realizados com o <i>dataset</i> CORINE. ....	75
Tabela 24: resultados do cenário 1, dos testes com o CORINE. ....	77
Tabela 25: Categorias dos POIs localizados nas áreas da classe “Artificial Surfaces”. ....	77
Tabela 26: Resultados do cenário 2, dos testes com o CORINE. ....	79
Tabela 27: Categorias selecionadas no cenário 3. ....	79
Tabela 28: Resultados do cenário 3, dos testes com o CORINE. ....	80
Tabela 29: Resultados do cenário 4, dos testes com o CORINE. ....	81
Tabela 30: Resultados do teste 1, do cenário 5, com base no CORINE. ....	82
Tabela 31: Resultados do teste 2, do cenário 5, com base no CORINE. ....	82
Tabela 32: Resultados do teste 3, do cenário 5, com base no CORINE. ....	83

Tabela 33: Resultados do teste 4, do cenário 5, com base no CORINE.....	83
Tabela 34: Resultados do teste 5, do cenário 5, com base no CORINE.....	84
Tabela 35: Resultados do teste 6, do cenário 5, com base no CORINE.....	84
Tabela 36: Resultados do teste 7, do cenário 5, com base no CORINE.....	85
Tabela 37: Maiores correlações com a classe 1, do CORINE (raio: 1250m).....	95
Tabela 38: Maiores correlações com a classe 1, do CORINE (raio: 2500m).....	96
Tabela 39: Maiores correlações com a classe 1, do CORINE (raio: 5000m).....	96
Tabela 40: Maiores correlações com a classe 1.1, do CORINE (raio: 1250m).....	97
Tabela 41: Maiores correlações com a classe 1.1, do CORINE (raio: 2500m).....	97
Tabela 42: Maiores correlações com a classe 1.1, do CORINE (raio: 5000m).....	98
Tabela 43: Maiores correlações com a classe 111, do CORINE (raio: 1250m).....	98
Tabela 44: Maiores correlações com a classe 111, do CORINE (raio: 2500m).....	99
Tabela 45: Maiores correlações com a classe 111, do CORINE (raio: 5000m).....	99
Tabela 46: Maiores correlações com a classe 112, do CORINE (raio: 1250m).....	100
Tabela 47: Maiores correlações com a classe 112, do CORINE (raio: 2500m).....	100
Tabela 48: Maiores correlações com a classe 112, do CORINE (raio: 5000m).....	101
Tabela 49: Maiores correlações com a classe 121, do CORINE (raio: 1250m).....	101
Tabela 50: Maiores correlações com a classe 121, do CORINE (raio: 2500m).....	102
Tabela 51: Maiores correlações com a classe 121, do CORINE (raio: 5000m).....	102
Tabela 52: Correlações entre as classes de LULC e as subcategorias da categoria Business. ....	103
Tabela 53: Maiores correlações entre a classe 1, do CORINE, e as categorias.....	104
Tabela 54: Maiores correlações entre a classe 1.1, do CORINE, e as categorias.....	104
Tabela 55: Maiores correlações entre a classe 111, do CORINE, e as categorias.....	105
Tabela 56: Maiores correlações entre a classe 112, do CORINE, e as categorias.....	105
Tabela 57: Maiores correlações entre a classe 121, do CORINE, e as categorias.....	106
Tabela 58: Correlações entre as subcategorias (nível 2) da categoria Business (raio: 1250m)..	112
Tabela 59: Correlações entre as subcategorias (nível 2) da categoria Business (raio: 2500m)..	112
Tabela 60: Correlações entre as subcategorias (nível 2) da categoria Business (raio: 5000m)..	113
Tabela 61: Valores dos parâmetros utilizados nos classificadores. ....	149
Tabela 62: Conjunto de atributos 1.....	159
Tabela 63: Conjunto de atributos 2.....	161
Tabela 64: Conjunto de atributos 3.....	163
Tabela 65: Conjunto de atributos 4.....	166
Tabela 66: Conjunto de atributos 5.....	170
Tabela 67: Conjunto de atributos 6.....	172
Tabela 68: Conjunto de atributos 7.....	174
Tabela 69: Conjunto de atributos 8.....	176
Tabela 70: Conjunto de atributos 9.....	178
Tabela 71: Conjunto de atributos 10.....	181

Tabela 72: Conjunto de atributos 11.....	182
Tabela 73: Categorias utilizadas para a geração dos primeiros <i>heatmaps</i> .....	185
Tabela 74: Correlações entre as categorias e a classe 1, do CORINE.....	187
Tabela 75: Correlações entre as categorias a classe 1.1, do CORINE.....	189
Tabela 76: Correlações entre as categorias e a classe 111, do CORINE.....	191
Tabela 77: Correlações entre as categorias e a classe 112, do CORINE.....	193
Tabela 78: Correlações entre as categorias e a classe 121, do CORINE.....	195
Tabela 79: Correlações entre as classes e as subcategories (nível 3) da categoria “Business”..	197
Tabela 80: Correlações entre as categorias (raio dos <i>heatmaps</i> : 1250m).....	203
Tabela 81: Correlações entre as categorias (raio dos <i>heatmaps</i> : 2500m).....	204
Tabela 82: Correlações entre as categorias (raio dos <i>heatmaps</i> : 5000m).....	204



# Acrónimos

POIs – Points Of Interest

SIG – Sistema de Informação Geográfica

SQL – Structured Query Language

LULC – Land Use and Land Cover

LSBN – Location-Based Social Networks

OSM – OpenStreetMap

OOC – Object-Oriented Classification

LDA – Latent Dirichlet Allocation

HSC – Hierarchical Semantic Cognition

RF – Random Forest

SVM – Support Vector Machine

DCNN - Deep Convolutional Neural Network

SOM – Self-Organizing Maps

LS – Laplacian Score

NB – Naïve Bayes

NN – Neural Network

GPS – Global Positioning System

K-NN – K-Nearest Neighbor

ELM – Extreme Learning Machine

JDL – Joint Deep Learning

LS – Laplacian Score

MLE – Maximum Likelihood Estimation

LDA – Latent Dirichlet Allocation

DMR – Dirichlet Multinomial Regression

HSC – Hierarchical Semantic Cognition

GLCM – Gray-Level Co-occurrence Matrix

SIFT – Scale Invariant Feature Transform  
pLSA – Probabilistic Latent Semantic Analysis  
IDW – Inverse Distance Weight  
KDE – Kernel Density Estimation  
IHSC – Inverse Hierarchical Semantic Cognition  
ESTARFM – Space-Time Fusion Algorithm  
DCCA – Detrended Cross-Correlation Analysis  
RMT – Random Matrix Theory  
DTW – Dynamic Time Warping  
TI – Tecnologia de Informação  
API – Application Programming Interface  
PHP – Hypertext Preprocessor  
SDK – Software Development Kit  
SGBD – Sistema Gerenciador de Base de Dados  
JSON – JavaScript Object Notation  
WORA – Write Once, Run Anywhere  
JVM – Java Virtual Machine  
CSV – Comma-Separated Values  
GNU – General Public License  
UNIX – Uniplex Information and Computing System  
URL – Uniform Resource Locator  
ETL – Extract, Transform, Load  
COS – Carta de Uso e Ocupação do Solo  
UMC – Unidade Mínima Cartográfica  
SNIG – Sistema Nacional de Informação Geográfica  
CRS - Coordinate Reference System

# 1 Introdução

Neste capítulo é apresentado o trabalho realizado, por meio do seu enquadramento, abordando também a motivação para a sua realização, os principais objetivos, a metodologia seguida, os principais contributos do estudo e, por fim, a estrutura deste documento.

## 1.1 Enquadramento

Com o passar dos anos os espaços urbanos se tornam inevitavelmente mais complexos, exigindo um esforço cada vez maior no que diz respeito à sua gestão [1], [2]. Problemas como a falta de espaço e o crescimento desordenado das cidades são uma realidade cada vez mais comum em muitos países. À medida que a população das grandes metrópoles aumenta, aumentam também as preocupações com assuntos relacionados a diversos problemas, como por exemplo a sustentabilidade, o trânsito, a poluição, o transporte público, a qualidade de vida das pessoas, entre outros [3], [4]. Além disso, com o crescimento das necessidades de expansão das cidades, e considerando as demandas específicas de cada segmento de indústria, empresa ou serviço, planejar questões ligadas às infraestruturas e ao tipo de uso que será dado aos espaços, revela ser um problema complexo, que exige grande esforço por parte dos profissionais envolvidos. Neste contexto, o estudo dos espaços urbanos tem se tornado cada vez mais importante, tendo em vista o seu contributo indispensável para os mais diversos setores da economia e principalmente para a administração pública, no que diz respeito ao planeamento e a manutenção das cidades modernas.

## 1.2 Motivação

No âmbito dos estudos dos espaços urbanos, nota-se uma grande variedade de trabalhos relacionados à geração e manutenção de informações relacionadas não só à cobertura das regiões, como também aos vários tipos de uso atribuídos às diversas áreas geográficas. Na prática, tais informações normalmente dão origem aos mapas de uso e ocupação do solo, que por sua vez são utilizados nos mais variados setores da administração pública, principalmente no que diz respeito à gestão dos espaços e das infraestruturas [5]–[7]. São, portanto, informações essenciais para o planeamento urbano, que exercem uma influência direta nas restrições relacionadas à ocupação do território. Estas restrições impedem por exemplo, que uma indústria se instale numa região residencial e desta forma, possuem uma estrita relação com a localização das diversas empresas, serviços e outros pontos de interesse distribuídos nos centros urbanos.

Com o passar dos anos, os mapas de uso e ocupação do território se tornaram uma ferramenta essencial em vários domínios. Entender de maneira eficiente a forma como cada espaço é utilizado é uma questão central na dinâmica das cidades modernas [8]. Com a evolução da tecnologia, surgiram também muitas iniciativas para auxiliar a criação e a atualização deste tipo de informação. Novos estudos sugerem uma grande diversidade de métodos que podem ser utilizados para esta finalidade. Contudo, os mapas de uso e ocupação do solo ainda dependem, em sua maioria, de conjuntos de dados que implicam um elevado investimento financeiro para geração ou aquisição [8], [9]. Além disso, em diversos casos, a criação destes conjuntos de dados é também um processo bastante dispendioso em termos de tempo. Neste sentido, um dos grandes desafios que se coloca à comunidade científica, é o uso de tipos alternativos de dados, que sejam preferencialmente gratuitos ou de baixo custo, para a caracterização dos espaços urbanos.

Outro grande desafio enfrentado no contexto dos espaços urbanos modernos, diz respeito à localização geográfica das diversas categorias de empresas, serviços e outros pontos de interesse. Determinar a zona mais adequada para a instalação de uma empresa é uma tarefa com elevado grau de complexidade e que está diretamente relacionada à existência da infraestrutura adequada e, conseqüentemente, ao planeamento urbano [10]–[13]. Com base nas diversas possibilidades trazidas pelos avanços tecnológicos dos dias atuais, surgem ideias cada vez mais inovadoras, que apresentam um grande potencial para auxiliar na solução da questão. Neste sentido, uma etapa fundamental que pode contribuir nas fases iniciais da resposta ao problema, diz respeito à análise das correlações existentes entre cada setor e o tipo de uso praticado em cada região do espaço urbano. Desta forma, a proposta de metodologias que possam ser utilizadas para evidenciar as ligações entre a classificação dos espaços e os tipos de empresas, serviços ou pontos de interesse que neles se localizam, é outra demanda com grande relevância nesta área do conhecimento.

## 1.3 Objetivos

Os objetivos centrais deste trabalho são os seguintes:

- Realizar uma investigação para destacar os principais métodos e os tipos de dados mais frequentemente utilizados ao longo dos últimos anos para a classificação do uso e da cobertura dos espaços, nos diversos estudos científicos, tendo em conta os seus objetivos;
- Analisar o potencial dos dados gratuitos e colaborativos, nomeadamente os pontos de interesse, para caracterização dos espaços urbanos e seminaturais;
- Propor uma metodologia para a análise das correlações entre variados grupos de pontos geográficos e a caracterização das regiões, aplicada ao seguinte caso prático: estudo das ligações existentes entre os diversos tipos de uso e ocupação dos espaços e as várias categorias de empresas, serviços e outros pontos de interesse localizados na região metropolitana de Lisboa;

- Propor uma metodologia para análise das correlações existentes entre concentrações distintas de pontos geográficos, aplicada ao seguinte caso prático: estudo das ligações existentes entre as diferentes categorias de empresas, serviços e outros pontos de interesse instalados na região metropolitana de Lisboa.

## 1.4 Metodologia do trabalho

Na etapa preliminar do projeto foi feita uma definição de alto nível das tarefas a serem executadas. Foi realizado também nesta altura, o planeamento da arquitetura a ser aplicada na configuração dos servidores que foram disponibilizados. Ao longo do desenvolvimento do estudo, foram realizadas reuniões semanais com os orientadores, para acompanhamento do trabalho e para definição a um nível mais detalhado, das atividades a serem desempenhadas nas etapas seguintes. Nestas reuniões, quando necessário, eram também definidas novas tarefas e realizada a exclusão daquelas que já não eram necessárias. Sempre que preciso, as tarefas eram reajustadas para melhor se adequar à realidade do projeto. A título informal, foram também realizadas incontáveis discussões sobre o desenvolvimento do estudo, clarificando questões como as ferramentas e os tipos e fontes de dados a serem adotados. Os problemas encontrados foram resolvidos de forma pontual, à medida que surgiam. Para comunicação formal foi utilizado o correio eletrónico e para questões informais recorreu-se ao uso de um software de mensagens instantâneas. Os documentos e os conjuntos de dados gerados, tal como os softwares desenvolvidos, foram armazenados em repositórios remotos, com acesso através da Internet.

## 1.5 Principais contributos

O desenvolvimento deste trabalho envolveu diferentes etapas, algumas das quais dizem respeito à vertente de engenharia, que se traduz em questões mais técnicas, nomeadamente no conhecimento reunido e documentado sobre o uso de diferentes tecnologias; e ao desenvolvimento de ferramentas para colmatar as necessidades inerentes a cada fase do trabalho, desde a coleta dos dados até ao resultado final das análises. Neste sentido, os principais contributos do trabalho são:

- Desenvolvimento de uma aplicação para coleta de pontos de interesse na Internet, que constituem o principal conjunto de dados utilizado para a realização do projeto;
- Desenvolvimento de duas aplicações para extração de conjuntos de dados específicos, a partir de uma base de dados espacial, os quais foram utilizados como dados de entrada para os métodos utilizados;
- Desenvolvimento de uma aplicação baseada nas bibliotecas de uso livre e de código-fonte aberto, de um SIG, para a geração de mapas de densidade e exportação dos mesmos em formato matricial;

- Criação de um método automático para a análise de correlações com base em dados com valores binários ou contínuos, extraídos a partir do processamento de imagens em formato vetorial;
- Criação de diversos *scripts* SQL, para suprir as várias necessidades que surgiram ao longo do desenvolvimento do trabalho, os quais podem ser úteis para a replicação ou continuidade deste estudo.

Finalmente, e ainda mais importante, o estudo traz diversos contributos na vertente científica, os quais envolveram uma investigação mais aprofundada, dando origem a alguns trabalhos que foram publicados. Nas principais etapas do projeto, alguns pontos de destaque são:

- Investigação dos tipos de dados e métodos (e respetivos objetivos) adotados no decorrer dos últimos anos para a caracterização dos espaços urbanos;
- Investigação do potencial dos modelos treinados a partir de métricas extraídas de POIs, para classificação do uso e da ocupação dos espaços geográficos;
- Estudo, com recurso à uma metodologia própria, das ligações existentes entre os diferentes tipos de uso e ocupação dos espaços geográficos e as várias categorias de empresas, serviços e outros pontos de interesse localizados na área estudada;
- Estudo, com recurso a uma metodologia própria, das ligações existentes entre as diversas categorias de empresas, serviços e pontos de interesse instalados na área estudada.

### 1.5.1 Publicações

Ao longo do desenvolvimento deste estudo, foram realizadas duas publicações científicas, estando uma terceira já submetida a uma revista, à espera de uma decisão por parte dos revisores. Os trabalhos já publicados são os seguintes:

- R. Andrade, A. Alves and C. Bento, "Exploring different combinations of data and methods for urban land use analysis: a survey", in Workshop on Ambient Intelligence for promoting Sustainable Behaviors (BRAINS). Part of the 15th European Conference on Ambient Intelligence, 2019.
- R. Andrade, A. Alves and C. Bento "An overview of different data types and methods for urban land use analysis", in Poster Session at the 15th European Conference on Ambient Intelligence, 2019.

## 1.6 Estrutura do documento

O restante deste documento está organizado da seguinte forma. No capítulo 2 são apresentados os fundamentos teóricos relacionados aos objetivos do projeto. São também abordados os tipos de dados e os métodos, tal como os seus objetivos, utilizados para a caracterização dos espaços urbanos nos diversos estudos publicados no decorrer dos últimos anos. São ainda apresentados conceitos teóricos ligados aos restantes temas envolvidos no estudo.

O capítulo três apresenta as diversas ferramentas e tecnologias adotadas ao longo do desenvolvimento deste trabalho. É realizada uma descrição da finalidade para a qual foi utilizada cada uma delas, tal como as suas principais características e os motivos que influenciaram a sua escolha. Em cada tópico é feita um pequeno enquadramento, de maneira a fornecer uma melhor interpretação do contexto onde geralmente é utilizada aquela tecnologia.

No capítulo 4 são discutidos os dados utilizados no estudo. Em relação aos pontos de interesse, é apresentada a forma como foram coletados, o tratamento aplicado e uma análise exploratória, com o objetivo de fornecer uma interpretação geral das principais categorias existentes, do volume e da distribuição geográfica dos dados, etc. É também feita uma apresentação dos outros dois conjuntos de dados utilizados, que representam o uso e a cobertura dos espaços no território nacional, num dos casos, e na região da Europa, no outro.

Uma das partes mais importantes do conteúdo é explorada no capítulo 5, onde são apresentados o trabalho realizado e os resultados obtidos em duas etapas específicas do estudo. A primeira delas está relacionada à investigação do potencial dos dados dos pontos de interesse para a caracterização dos espaços. A segunda etapa trata da metodologia proposta e dos resultados obtidos no que diz respeito à análise das correlações entre as diferentes classes de uso e ocupação das regiões e as várias categorias das empresas e pontos de interesse. É também explorado o trabalho realizado no sentido de analisar as ligações existentes entre a localização geográfica das empresas e dos pontos de interesse instalados na região estudada.

Por último, são apresentadas as conclusões gerais obtidas com base no trabalho realizado. Inicialmente é feita uma análise geral do tema e dos objetivos. Em seguida são abordados os principais contributos, teóricos e práticos. São também listadas as eventuais limitações do estudo. Por fim, são apresentadas as sugestões para a continuidade deste trabalho.



## 2 Fundamentação teórica

Neste capítulo é apresentada a fundamentação teórica para os conceitos envolvidos no estudo. Primeiramente são abordadas as questões relacionadas à classificação do uso e da ocupação dos espaços. Em seguida, o conteúdo é direcionado para as metodologias utilizadas no contexto das análises das correlações com base em dados georreferenciados.

### 2.1 Caracterização do uso dos espaços

Nos últimos anos, o estudo dos espaços urbanos tem se tornado cada vez mais importante devido a questões ligadas a diversos domínios. Embora muitos desses domínios estejam relacionados com o crescimento sustentável das cidades, observa-se uma grande variedade de tópicos no que diz respeito aos trabalhos nesta área científica. Entre os assuntos que dependem diretamente ou indiretamente do estudo dos espaços urbanos, destacam-se por exemplo a preservação ambiental, a redução da poluição, o planeamento da expansão urbana, a criação e a manutenção das infraestruturas, o trânsito e os transportes, a qualidade de vida, a segurança pública, o controlo de doenças contagiosas, entre muitos outros. Contudo, a medida que as cidades evoluem e se expandem, a manutenção de informações relacionadas às áreas urbanizadas se torna um desafio cada vez maior para a comunidade científica e para a administração pública.

Entender a maneira como as pessoas utilizam e interagem com os espaços urbanos e como essas áreas normalmente mudam ao longo dos anos, se tornou uma questão fundamental nos tempos atuais [8], [14]. Desta forma, devido à crescente disponibilidade de informações georreferenciadas e imagens de sensores remotos, vários métodos, que podem ser utilizados com base em um ou mais tipos de dados, se tornaram populares para a extração de conhecimento no contexto do uso e da ocupação dos espaços geográficos. Neste sentido, é notável o esforço dos investigadores para melhoria dos resultados, por meio da combinação de várias técnicas e diferentes tipos de dados, traduzindo-se frequentemente em metodologias relativamente eficazes. O assunto central do estudo do uso e da ocupação do solo é frequentemente referido pelo seu termo em inglês, *Land Use and Land Cover* – LULC, e envolve 3 conceitos fundamentais:

- A cobertura dos espaços (*land cover*): uma determinada região pode ser ocupada por florestas, áreas inundadas, superfícies impermeáveis, rios, etc. [15],[16].
- A utilização dos espaços (*land use*): caracteriza-se pelo propósito da existência do espaço, e.g. lazer, recreação, etc. [15],[17].

- Regiões funcionais (*functional zones*): áreas onde são executadas as atividades humanas, e.g. áreas comerciais, industriais, residenciais, etc. [4].

Nos estudos de LULC, muitos investigadores focam-se na cobertura dos espaços em detrimento das análises em larga escala das regiões funcionais, o que faz com que os mapas neste grau de detalhe se tornem menos comuns. Além disso, muitos dos trabalhos nesta área baseiam-se na interpretação visual de imagens, que normalmente passam pela extração de padrões que eventualmente possam existir por exemplo no formato, na cor e na textura dos elementos. Neste sentido, os elementos da imagem geralmente são conhecidos como “objetos”. Tais objetos são basicamente componentes visuais ou “entidades” com alguma relevância na imagem, como um prédio, uma casa, uma ponte, entre outros [18].

Contudo, Além das regiões funcionais serem mais abrangentes, em termos espaciais, que os simples objetos, elas também são semanticamente diferentes destes. Por exemplo, enquanto uma área residencial é uma região funcional, uma construção está relacionada à cobertura do espaço. Considerando que estes dois tipos de unidades estão em níveis semânticos diferentes, os métodos tradicionais baseados na interpretação visual de objetos, por si próprios, não são capazes de caracterizar as regiões funcionais [5]. Mesmo assim existem muitas abordagens baseadas no uso de imagens de sensores remotos, que permitem a classificação das áreas urbanas com base numa análise morfológica, com recurso à extração de características espectrais e texturais que representam as informações das regiões [4]. Tais abordagens têm evoluído significativamente nos últimos tempos, tendo em vista que por meio delas é possível revelar padrões, considerando a presença, a forma, o tamanho e a eventual distribuição espacial das construções e dos espaços seminaturais [7], [19], [20].

No contexto dos estudos dos espaços urbanos, as propriedades extraídas das imagens, com base nos padrões de cores, texturas e formas são normalmente conhecidas como *low-level semantic features*, enquanto que as demais, geralmente extraídas de dados relacionados ao comportamento e as atividades humanas, como por exemplo as informações socioeconômicas, são chamadas *high level semantic features* [1], [5]. Conforme referido, embora existam diversos estudos relacionados aos espaços urbanos que se baseiam na interpretação de imagens, e mesmo considerando a recente evolução que as abordagens para esta finalidade sofreram nos últimos tempos, ainda assim a classificação do uso dos espaços com base apenas nas *low-level features* dificilmente alcança resultados relevantes, tendo em vista que diferentes objetos podem ter as mesmas propriedades físicas e objetos idênticos podem ter atributos diferentes. Por este motivo, os especialistas geralmente recomendam a combinação de vários tipos de dados e de diferentes métodos, de maneira a alcançar melhores resultados. Por exemplo, na análise de um conjunto de imagens, os objetos podem ser classificados com base nas suas características visuais, e adicionando-se informações relacionadas às atividades humanas, torna-se possível a identificação do tipo de uso de cada zona [1].

Como forma de aumentar a relevância dos resultados obtidos, muitos autores sugerem o uso de diferentes combinações de métodos e dados [1], [21]. Além disso, devido ao rápido desenvolvimento tecnológico vivenciado nos últimos tempos, novos métodos têm sido propostos, baseados nas novas funcionalidades promovidas pelo avanços nos sistemas SIG, nas inovações no que diz respeito ao tratamento de grandes volumes de informação georreferenciada e nas imagens de sensores remotos [2]. Relativamente aos dados, como forma de melhorar os resultados, é comum a recomendação de combinações de diversos tipos, como por exemplo pontos de interesse, dados de atividades em redes sociais georreferenciadas (*Location-Based Social Networks – LSBN*), imagens de sensores remotos, entre outros, de maneira a permitir a construção de modelos mais robustos [1], [4], [8], [14]

### 2.1.1 Dados

Uma importante questão relativamente à classificação do uso das regiões, diz respeito à melhoria contínua dos resultados obtidos nos estudos. Uma das formas de alcançar este objetivo é o uso de atributos extraídos a partir de diferentes fontes de dados. Nesta secção são abordados os principais tipos de dados utilizados para caracterização das regiões geográficas, com base no conjunto de estudos analisados, que foram publicados nos últimos 5 anos.

#### 2.1.1.1 Imagens de sensores remotos

Devido ao recente desenvolvimento dos diversos tipos de sensores remotos, está disponível atualmente uma grande quantidade de dados gerados por meio dos mesmos. Entre os dados mais comuns estão as imagens de sensores remotos [22], que podem ser encontradas por exemplo em *datasets* públicos disponibilizados por universidades [8], agências governamentais [6], etc. As imagens utilizadas neste contexto têm a sua origem principalmente em sensores instalados em satélites e aeronaves e são particularmente úteis para a extração de informações sobre a cobertura dos espaços. Essas informações podem ser combinadas com outros tipos de dados para a caracterização funcional das regiões urbanas [5].

Tipicamente, por meio do uso de imagens de sensores remotos, a extração de informações sobre a cobertura das regiões é baseada nas propriedades físicas dos objetos, considerando as diversas características como a sua distribuição ao longo do espaço, além de padrões constituídos pela sua cor, textura e forma [1], [22]. Em 2017, foi proposto uma metodologia para a classificação do uso dos espaços urbanos, combinando modelos de classificação de cenas e o Google Tensorflow [23]. No ano seguinte, 2018, foram utilizadas imagens de sensores remotos em conjunto com pontos de interesse e dados de redes rodoviárias para mapear zonas funcionais urbanas [24].

#### 2.1.1.2 Dados colaborativos

O uso dos diversos tipos de aplicações colaborativas têm gerado atualmente uma vasta quantidade de dados, em diversos domínios. Algumas dessas aplicações, geralmente utilizadas em dispositivos

móveis, produzem dados frequentemente adotados em estudos relacionados ao uso dos espaços. Um exemplo de projeto que gera uma grande quantidade de dados colaborativos é o OpenStreetMap – OSM [25]. Os tipos mais frequentes de dados colaborativos neste contexto são:

- Pontos de interesse ([2], [14], [20], [26], [27]);
- Mensagens de texto ([3], [4], [28], [29]);
- Registos de atividades em LBSNs ([4], [14], [30]);
- Mapas colaborativos ([1], [8], [31]).

### 2.1.1.3 Dados de trajetos de táxi

Dados das viagens de táxi podem facilmente fornecer informações sobre pontos de origem e destino, além das distâncias e do tempo de cada viagem. Contudo, tendo em conta que os locais de desembarque dos passageiros nem sempre representam o seu destino final [32], o uso desse género de informação torna-se um desafio, a certo nível. Além disso, normalmente esse tipo de *dataset* não fornece informações sobre o propósito das viagens. Desta forma, a combinação dos dados de trajetos de táxi com outros, torna-se muitas vezes necessária. Um exemplo da adoção desse tipo de dados pode ser visto em [33], onde os autores utilizaram os dados de viagens de táxi em conjunto com pontos de interesse e informações de trânsito para a descoberta de regiões funcionais na área estudada.

### 2.1.2 Métodos

Com o intuito de melhorar os resultados obtidos nos estudos dos espaços urbanos, outra peculiaridade que se observa nos trabalhos mais recentes é a combinação de diferentes métodos. A utilização desses diferentes métodos foi também impulsionada pela disponibilidade dos dados que se nota atualmente, como mencionado no tópico anterior. No conjunto de estudos analisado, nota-se uma diversidade de métodos. Alguns deles são, entretanto, utilizados para preparação dos dados ou apenas em parte do trabalho. No que diz respeito à classificação dos espaços, os métodos observados com maior frequência são os seguintes:

- *Object-Oriented Classification* – OOC ([2], [19]);
- *Latent Dirichlet Allocation* – LDA ([4], [14]);
- *K-means* ([1], [26]);
- *Hierarchical Semantic Cognition* – HSC ([5], [21]);
- *Random Forest* – RF ([8], [29]);
- *Support Vector Machine* – SVM ([6], [34]);
- *Deep Convolutional Neural Network* – DCNN ([22], [35]).

Neste trabalho, para a realização das experiências apresentadas na secção 5.1, foram testados vários métodos diferentes para classificação do uso e da ocupação dos espaços. Sendo assim, os métodos utilizados são descritos a seguir com maior detalhe.

### 2.1.2.1 Naïve Bayes - NB

O Naïve Bayes está entre alguns dos algoritmos preditivos mais comuns nos dias atuais. É usado para as mais diversas tarefas de classificação e tem as suas raízes nos conceitos da probabilidade e estatística. O NB potencializa as relações probabilísticas entre os atributos e aquilo que se quer prever. Em geral, o algoritmo também assume fortemente a independência entre os atributos e, apesar desta independência nem sempre ser verdadeira, a robustez do classificador acaba por ultrapassar as limitações causadas. Conforme indicado pelo nome, o algoritmo foi criado com base no teorema de Bayes, resultado do trabalho de Thomas Bayes, publicado em 1763. O teorema é considerado um dos mais importantes conceitos na área da estatística, tendo em conta que expressa matematicamente como um grau de confiança subjetiva muda para considerar novas evidências [36]. A compreensão do teorema de Bayes é essencial para o entendimento do algoritmo. São abordados nos parágrafos seguintes, os detalhes sobre o algoritmo, de acordo com [36].

Considere que  $X$  seja um conjunto de atributos e  $Y$  a classe, ou seja, aquilo se quer prever. Desta forma, sendo  $X = \{X_1, X_2, X_3, \dots, X_n\}$  e  $X_i$  um atributo individual, a probabilidade do resultado  $P(Y)$  é denominada “probabilidade a priori”, e pode ser calculada com base no conjunto de dados de treino. Em resumo, a probabilidade a priori demonstra a probabilidade de um determinado resultado em um determinado conjunto de dados. Sendo assim,  $P(Y|X)$  é uma “probabilidade condicional”, isto é, a probabilidade daquele resultado dada a evidência, ou seja, quando o valor de  $X$  é conhecido.  $P(Y|X)$  é também conhecida como “probabilidade a posteriori”. O teorema de Bayes pode ser representado pela Equação 1.

$$P(Y|X) = \frac{P(Y) \times P(X|Y)}{P(X)}$$

Equação 1: Fórmula do teorema de Bayes.

Na equação mostrada acima,  $P(X|Y)$  é uma outra probabilidade, chamada “probabilidade condicional da classe”. Em suma, é a probabilidade de existirem as condições, dado um resultado, e tal como  $P(Y)$ , pode ser calculada com base nos dados de treino. Relativamente à  $P(X)$ , esta é basicamente a probabilidade da evidência e considerando que é a mesma para todas as classes, não precisa ser calculada, podendo ser assumida como uma constante. Considerando um conjunto de dados com  $n$  elementos, o cálculo poderia ser representado como se vê a seguir, na Equação 2.

$$P(Y|X) = \frac{P(Y) \times \prod_{i=1}^n P(X_i|Y)}{P(X)}$$

Equação 2: Fórmula para o cálculo da probabilidade condicional.

Durante a sua execução, o NB primeiramente calcula a probabilidade a priori, isto é, a probabilidade de ocorrência de cada classe. Em seguida é feito o cálculo da probabilidade condicional, que representa a probabilidade de cada valor de atributo para cada classe. Conhecidos esses valores, o algoritmo aplica o cálculo apresentado na equação acima, para saber qual classe apresenta maior probabilidade, sendo esta a escolhida como resultado.

O algoritmo NB é bastante simples, porém robusto. É capaz de lidar com valores nulos nos exemplos de teste, um problema que algoritmos como as árvores de decisão dificilmente são capazes de ultrapassar em casos onde o atributo em falta foi utilizado na estrutura do modelo, nos nós mais próximos à raiz. Entretanto, apesar de simples e robusto, o NB apresenta limitações para lidar com conjuntos de dados de treino incompletos, atributos contínuos, e com casos onde o conjunto de dados apresenta muitos atributos dependentes.

### 2.1.2.2 *Random Forest* - RF

Como indicado pelo nome, o algoritmo *Random Forest* é basicamente um conjunto de árvores de decisão [37], onde cada uma delas trabalha com uma parte do conjunto de dados. Mais especificamente, o termo *random* refere-se às amostras aleatórias, enquanto a palavra *forest* refere-se ao seu significado literal: florestas [38]. O algoritmo foi primeiramente impulsionado por um estudo relativamente recente [39] e utiliza um conceito de ensemble, conhecido como *bagging* (um acrónimo do método *Bootstrap Aggregating*), proposto em um trabalho preliminar do mesmo autor [40].

Na técnica *bagging*, em cada iteração uma amostra aleatória dos dados de treino é considerada para a construção do modelo. Embora essa abordagem seja também uma maneira indireta de se adicionar aleatoriedade à construção do modelo, a literatura aponta algumas desvantagens em utilizar uma técnica de *bagging* diretamente com árvores de decisão. Entre essas desvantagens, destaca-se o facto de que as opções de divisão nos níveis mais altos da árvore são prováveis, em termos estatísticos, de permanecerem pouco variáveis em relação às amostras. Desta forma, as árvores criadas apresentam maior similaridade, o que limita as possibilidades de redução dos erros. Para solução do problema, uma das alternativas é diversificar os modelos das árvores de decisão, utilizando modelos aleatórios, com menor correlação com as amostras. Por seguir esta abordagem, o RF geralmente alcança melhores resultados em comparação com a utilização direta do *bagging*. Para reduzir o erro de generalização, o algoritmo faz randomização tanto na seleção de exemplos de treino, quanto na seleção de atributos [36], [41].

Para controlar a aleatoriedade do critério de seleção, um parâmetro aleatório inteiro  $q \leq d$  é utilizado, onde  $d$  representa o número total de atributos do conjunto de dados. É então criada uma árvore de decisão. Para cada nó, ao invés de considerar todos os atributos, é criado aleatoriamente um subconjunto  $S$ , do tamanho  $q$ . Quanto maior for o valor de  $q$ , maior será a correlação das árvores geradas e conseqüentemente mais similares a uma árvore gerada sem

nenhuma aleatoriedade. Uma forma de contornar o problema é utilizar valores pequenos para o parâmetro, em relação ao número total de atributos, o que também permite que as árvores cresçam de maneira mais eficaz, tendo em vista que é preciso considerar um número menor de atributos em cada nó. Em contrapartida, um subconjunto  $S$  constituído por mais atributos pode gerar maior *accuracy*, o que também é importante. Portanto, o ideal é encontrar um ponto de equilíbrio, que pode ser determinado por  $q = \log_2(d) + 1$  [41].

Em casos onde o número de atributos do conjunto de dados é muito reduzido, torna-se difícil a atribuição de um valor para  $q$  que seja menor que  $d$ . Nesse tipo de situação é utilizado um número  $L \leq d$  de atributos que serão combinados. Assim, para cada nó, são selecionados  $L$  atributos de forma aleatória. Esses atributos são então combinados de forma linear com coeficientes gerados com base em um intervalo  $[-1,1]$ . Desta forma, o subconjunto  $S$  de atributos é criado com recurso a um número  $q$  de combinações [41]. Tal como no caso anterior, a divisão sempre acontece com base no conjunto de atributos  $S$ , o que produz divisões aleatórias e multivariadas.

### 2.1.2.3 Support Vector Machine - SVM

Algoritmos como o SVM são relativamente novos, tal como muitos métodos lineares. Apesar do nome, frequentemente traduzido como “máquinas de vetores de suporte”, parecer bastante confuso para um algoritmo de aprendizagem computacional, trata-se de um método bastante robusto e tem sido aplicado com sucesso em vários domínios. O SVM é um algoritmo com bases multidisciplinares, tendo sido desenvolvido a partir de teorias originárias de três áreas principais: informática, estatística e matemática. Uma das primeiras introduções formais aos conceitos envolvidos foi apresentada em um estudo publicado em 1995 [42]. A seguir é apresentada uma descrição detalhada do algoritmo, a qual se baseia essencialmente em [36].

O SVM funciona com base num princípio que visa estabelecer fronteiras à volta de pontos que pertencem a uma mesma classe. Tais fronteiras são frequentemente chamadas de hiperplano. De uma forma simplificada, considerando apenas dois atributos, o hiperplano pode ser uma reta ou uma curva. Já em três dimensões, pode ser um plano ou uma superfície irregular complexa. Consequentemente, quando se considera um número muito grande de dimensões, o hiperplano torna-se difícil de visualizar. Sendo assim, o termo “hiperplano” é uma forma genérica de descrever as fronteiras em mais de três dimensões. Uma vez estabelecidas tais fronteiras, a classificação é uma questão de verificar onde é que os novos pontos se enquadram melhor. O que pode ser visto como uma vantagem do SVM, é que o algoritmo precisa somente de um conjunto de “pontos-chave” que servirão para identificar e definir as fronteiras, os quais são chamados vetores de suporte. Esses pontos são chamados de vetores porque são basicamente agrupamentos dos vários atributos e dos seus respetivos valores, que representam cada exemplo do conjunto de dados utilizado.

No SVM, o número de fronteiras é definido conforme o conjunto de dados utilizado. Desta forma, o número mais adequado é aquele que consegue definir os limites mais eficientes para todas as classes, de maneira a minimizar, o máximo possível, o número de classificações incorretas. De uma forma geral, uma fronteira que consegue maximizar a distância geométrica média entre as classes que separa é ainda mais eficaz. Sendo assim, o SVM realiza uma otimização com o objetivo de maximizar essa distância  $n$ -dimensional, que é denominada “margem”. Entretanto, raramente é possível encontrar conjuntos de dados separáveis linearmente. E nos casos onde existe esta separação, podem haver pontos que acabam por ficar exatamente na linha das fronteiras. Nesse tipo de situação, o algoritmo aplica uma penalização para cada um dos pontos que se encontra dentro da margem, tendo em vista que o hiperplano ideal seria aquele com o menor número possível de pontos dentro dessa área. Desta forma o classificador pode escolher o hiperplano com a menor agregação de penalidade.

De acordo com Kotu & Deshpande [36], na grande maioria das situações, entretanto, os dados não podem ser separados por uma linha reta. Nos casos onde o hiperplano ideal seria uma elipse ou um círculo, por exemplo, um truque simples pode ser aplicado para transformar variáveis. Considerando duas variáveis  $x$  e  $y$ , uma nova variável  $z$  pode ser criada da seguinte forma:  $z = \sqrt{(x^2 + y^2)}$ . Nesse exemplo a variável  $z$  nada mais é do que a equação para um círculo. Sendo assim, considerando agora as dimensões  $x$  e  $z$ , tem-se um problema possível de se solucionar linearmente. As “funções de núcleo” (*kernel functions*) são, portanto, uma forma de transformar espaços não lineares em lineares. A maioria das implementações do SVM disponibilizam várias funções de núcleo, desde funções simples até funções sigmóides.

As três etapas fundamentais no algoritmo são: (1) encontrar as fronteiras para cada classe; (2) definir o melhor hiperplano,  $H$ , ou seja, o que maximiza a distância entre cada classe e as fronteiras, isto é, a margem; (3) classificar: determinar em qual lado do hiperplano os exemplos se encaixam.

Para definir as fronteiras é usada a seguinte estratégia: o algoritmo faz a ligação entre todos os pontos de uma mesma classe, evidenciando um contorno à volta deles, que é então definido como fronteira (também chamada de *convex hull*). Na etapa seguinte do algoritmo, qualquer hiperplano, como por exemplo, no caso de dois atributos  $x_1$  e  $x_2$ , pode ser representado por  $H = b + w \times x = 0$ , onde  $b$  é um termo de interceção chamado *bias*,  $x$  é  $(x_1, x_2)$  e  $w$  é o peso  $(w_1, w_2)$ . A margem do hiperplano pode ser calculada por:  $\frac{2}{\sqrt{(w_0 \times w_0)}}$ .

Para a maximização da margem é necessário aplicar um processo bem estabelecido na teoria matemática da otimização, conhecido como programação quadrática [43]. Para o cálculo do  $w_0$ , que pode ser definido com base em apenas alguns poucos vetores de suporte, usa-se  $w_0 = \sum |y_i x_i|$ , onde  $y_i$  são as classes,  $x_i$  os vetores e os  $i$ 's são coeficientes (diferentes de zero para tais

vetores). Uma vez definido o hiperplano, para a classificação basta apenas substituir o exemplo  $x$  a ser classificado, na equação, revelando assim a classe a qual pertence.

A descrição matemática completa para os pormenores do algoritmo pode ser encontrada em algumas publicações, onde se destaca [42] e [44]. É importante também destacar, que no trabalho apresentado no tópico 5.1, foi utilizada uma implementação específica do SVM [45], que utiliza a técnica “um-contra-um” [46] para realizar classificações polinomiais.

#### 2.1.2.4 Neural Network - NN

As redes neurais artificiais são uma forma de simulação das redes neurais biológicas. O seu funcionamento é similar ao do sistema nervoso humano, que é composto por tipos específicos de células, chamadas de “neurónios”. Esse tipo de célula recebe estímulos, faz “cálculos” e transmite os resultados para os outros neurónios. Os pontos de ligação entre os neurónios são chamados “sinapses”. Nos seres vivos, o processo de aprendizagem se dá com base na mudança da intensidade das ligações sinápticas, que normalmente acontece em resposta a estímulos externos. [41]. As redes neurais artificiais têm um histórico de construção, isto é, receberam contributos ao longo dos anos até ao ponto em que estão atualmente.

Apesar da teoria original das redes neurais envolver muitos termos da biologia, o seu conceito central tem uma base matemática relativamente simples. Num exemplo simples de rede neuronal, sendo  $Y = 1 + 2X_1 + 3X_2 + 4X_3$  um modelo linear onde  $Y$  é o resultado e  $X_1, X_2$  e  $X_3$  são atributos de entrada, 1 é a interceção e 2, 3 e 4 são os coeficientes. Desta forma,  $X_1, X_2$  e  $X_3$  são valores de entrada e passarão, cada um, por um “nó”. Posteriormente serão multiplicados pelos seus respectivos pesos. Há ainda um nó que possui somente o valor da interceção e, portanto, não recebe nenhuma entrada [36]. Os valores de todas as ligações serão enviados e sumarizados em um último nó, para permitir a previsão do resultado  $Y$ , como representado na Figura 1.

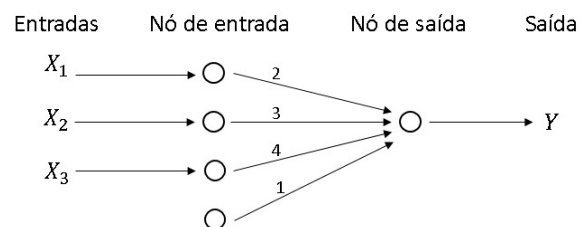


Figura 1: Topologia de uma rede neuronal simples. Adaptado de [36].

É importante clarificar alguns termos utilizados no âmbito da teoria das redes neurais, onde destacam-se as “unidades”, que são os nós de uma NN; a “camada de saída” (ou “nós de saída”), que é a última camada de uma rede neuronal; a “função de agregação”, que é executada pela camada de saída, a qual também pode ter uma “função de transferência”. Junto com a função de agregação e a função de transferência, a camada de saída também executa a “função de ativação”. Outro conceito importante é o do “perceptrão”, que é uma rede neuronal simples. Para

resolver problemas complexos e não lineares, as redes neurais normalmente apresentam “camadas ocultas” entre as camadas de entrada e de saída. Uma camada oculta por sua vez, contém vários nós ligados às outras camadas e realiza uma função de ativação.

A parte mais importante do treino de um modelo baseado em NN, é encontrar os pesos das ligações, dada uma topologia e a função de ativação. Cada exemplo do conjunto de dados de treino é utilizado pelo modelo para estimar o erro entre o resultado previsto e o real. Esse valor estimado é então utilizado para ajustar os pesos de maneira a reduzir o erro do próximo exemplo de treino. Esse processo é repetido até que o valor do erro seja reduzido a um intervalo “aceitável” [36]. Neste ajusteste as redes neurais artificiais utilizam o conceito de *back propagation*, que será explicado mais adiante.

O tipo de rede neuronal mais simples, chamado perceptron, possui basicamente uma única camada de entrada e uma de saída, como no exemplo mostrado anteriormente. Trata-se de um modelo de rede neuronal também denominado “camada única”, considerando que os nós de entrada fazem somente a transmissão dos valores recebidos, sem nenhum cálculo computacional, ficando as funções matemáticas a cargo do nó de saída [41]. Como referido, para solucionar problemas que apresentam um grau de complexidade mais elevado, normalmente são utilizadas as redes neurais “multicamadas”. Os parágrafos seguintes abordam o funcionamento desse tipo de topologia, assumindo-se, a título de exemplo, a existência de dois valores possíveis para a classe,  $\{-1, +1\}$ .

Para além das camadas de entrada e de saída, as redes neurais multicamadas possuem também camadas ocultas, que por sua vez podem incluir outras camadas, onde os nós estão ligados entre si, ilustrando o conceito de *multilayer feed-forward network*. Além disso os nós de uma camada estão completamente ligados aos nós da camada seguinte. A topologia desse tipo de rede é determinada automaticamente, tendo por base o número de camadas e o número de nós em cada uma delas, sendo esses parâmetros informados pelo utilizador. É também possível utilizar outros tipos de funções, ao contrário das NN de camada única, que permitem somente o uso de funções lineares das entradas. Por exemplo, sendo  $\overline{W} = (w_1 \dots w_d)$  os pesos para as ligações de  $d$  entradas diferentes para um neurónio de saída (considerando um conjunto de dados com  $d$  atributos) e  $b$  o *bias* associado à função de ativação, o resultado  $z_i$  para o elemento  $\overline{X}_i$  da posição  $i$  de uma sequência  $(x_i^1 \dots x_i^d)$ , é dado pelo cálculo mostrado a seguir, na Equação 3 [41].

$$z_i = \sum_{j=1}^d w_j \frac{1}{1 + e^{-x_i^j}} + b$$

Equação 3: Cálculo do valor esperado, executado em uma camada oculta de uma rede neuronal.

Neste caso, se  $z_i$  é uma função calculada nas camadas ocultas, deixa, portanto, de ser o resultado final da predição e a saída é então propagada para a próxima camada.

Considerando que os valores da saída esperados para os neurónios de uma camada oculta não são conhecidos, a correção dos pesos quando um exemplo de treino é classificado de maneira incorreta torna-se um problema. É onde se enquadra o *back propagation*. O algoritmo possui a fase de avanço (*forward phase*), onde as instâncias de treino são introduzidas nos nós de entrada, resultando em uma espécie de “cascata” de cálculos nas camadas da rede neuronal, com base nos pesos atuais. O resultado predito é então comparado com o valor real para saber se a classificação foi correta. Numa segunda etapa, a fase inversa (*backward phase*), são aprendidos os erros no sentido contrário ao fluxo da rede neuronal. Com base no erro das camadas finais é feita uma estimativa do erro nas camadas iniciais. Nas camadas ocultas, a estimativa do erro é calculada em função dos erros estimados e dos pesos das ligações dos nós à sua frente [41]. Com base nessa estimativa é então atualizado o peso das ligações de cada nó.

### 2.1.3 Combinações de dados e métodos

O uso dos métodos tradicionais baseados na interpretação de imagens de sensores remotos para classificação do uso e da ocupação dos espaços não é uma questão simples, tendo em conta que as propriedades físicas das regiões têm se tornado cada vez mais sofisticadas devido ao aumento da complexidade dos sistemas urbanos. Apesar de apresentar algumas limitações, o uso das imagens ainda é bastante comum nos dias atuais, graças ao aumento da sua disponibilidade, causado pela popularização dos sensores remotos, frequentes em drones, aviões, satélites, etc. Em combinação com outros tipos de dados, as imagens de sensores remotos são uma forte componente nas análises dos espaços geográficos. Como mostrado na Tabela 1, muitos estudos, como por exemplo [14] e [27] sugerem a sua combinação com dados de atividades em LSBNs, trajetórias de táxis, POIs, entre outros.

Apesar de vários estudos se focarem em objetivos similares, nota-se com alguma frequência a adoção de métodos diferentes. Por exemplo, para a classificação de uso dos espaços e para o mapeamento de regiões funcionais, [21] e [22] utilizaram, ambos, imagens de sensores remotos e dados colaborativos. Entretanto, enquanto o primeiro aplicou diferentes métodos para segmentar e classificar as imagens, o segundo adotou uma abordagem baseada em redes neurais convolucionais. De maneira similar, vários trabalhos sugerem abordagens que incluem o uso de dados colaborativos em conjunto com métodos como o LDA [14], *Place2Vec* [29], modelos bayesianos [27] e vários outros.

Conforme se observa na Tabela 1, o tipo de dados mais comum nos estudos analisados são os pontos de interesse. Trata-se de uma categoria adotada por muitos autores, principalmente pela ligação direta que possui com o comportamento humano, possibilitando, até certo nível, revelar a maneira como as pessoas utilizam os espaços. Além disso, os pontos de interesse estão frequentemente relacionados com as atividades registadas nas LBSN, sendo este o principal motivo pelo qual os dois tipos de dados normalmente são utilizados em conjunto [1], [8]. Verifica-

se ainda que vários outros tipos de dados colaborativos são gerados pelas pessoas em suas atividades diárias. Desta forma, em muitos países os dados colaborativos estão disponíveis em larga escala, o que incentiva a sua utilização em casos onde outros *datasets* não estão disponíveis, como por exemplo os dados do planejamento urbano ou dados GPS (Global Positioning System), tal como as trajetórias de táxi.

No que diz respeito aos métodos, devido à indisponibilidade de dados para validação dos resultados, muitos investigadores têm adotado o uso de técnicas não supervisionadas, onde métodos de *clustering* como o *spectral clustering* e o *K-Nearest Neighbors* (KNN) são comuns. Entretanto, relativamente ao conjunto de estudos analisado, o método mais comum nesta categoria é o *K-means*, devido a sua simplicidade e efetividade para propósitos como o agrupamento de POIs, regiões funcionais ou mesmo zonas geográficas. Alguns exemplos de estudos que adotaram técnicas de *clustering* são [1] e [47]. No caso do primeiro, os autores concluíram que a sua utilização em conjunto com outros métodos produziu resultados satisfatórios para a classificação da cobertura das regiões estudadas.

Em casos onde havia conjuntos de dados disponíveis para a validação, os autores optaram pela utilização de métodos de aprendizagem supervisionada. Neste contexto, o *Random Forest* (RF) é um dos algoritmos mais comuns, devido à sua efetividade como classificador, considerando um balanço entre performance e resultados, em tarefas como por exemplo a classificação do uso dos espaços e das regiões funcionais. Entretanto existem também outras razões para o uso deste método. O algoritmo foi utilizado por exemplo em [5] e [48]. No caso particular deste segundo estudo, os autores optaram pela utilização do RF por considerarem o método escalável e robusto para lidar com conjuntos de dados contendo um grande número de atributos.

A Tabela 1 apresenta os tipos de dados e os métodos utilizados nos estudos analisados, para caracterização dos espaços geográficos, tal como os seus objetivos. Como se pode observar, verificou-se a ocorrência de vários métodos diferentes. As técnicas foram empregadas para os mais diversos propósitos. Alguns exemplos incluem algoritmos como o *Naïve Bayes* – NB, *Extreme Learning Machine* – ELM, *Word2Vec*, *Skeleton-based decomposition*, *Multiresolution segmentation*, *Place2Vec*, *Joint Deep Learning* – JDL e vários outros. Apesar de haver casos onde métodos diferentes são utilizados para objetivos similares, os *datasets* adotados também são frequentemente diferentes, tornando difícil a comparação dos resultados e das conclusões.

Tabela 1: Dados e métodos utilizados nos estudos analisados.

Estudo	Dados	Método	Objetivo
[3]	Atividades em LBSN	<i>Self-Organizing Maps</i> – SOM	Segmentação das regiões
		<i>Spectral Clustering</i>	Deteção dos tipos de uso das regiões
		<i>Laplacian Score</i> – LS	Seleção de atributos
[48]	Atividades em LBSN POIs	<i>Clustering</i> (vários métodos)	Inferir os tipos de uso das regiões
		<i>Naïve Bayes</i> – NB	
		<i>Support Vector Machine</i> – SVM	
		Random Forest – RF	Classificação do uso dos espaços

		<i>POI Matching algorithm</i>	Mapear POIs entre <i>datasets</i> diferentes
[49]	POIs Dados de Censo Limites das cidades	<i>Bayesian networks; Tree-based Learners; Instance-based Learners; Rule-based Learners</i>	Classificar POIs
		<i>Maximum likelihood estimation – MLE</i>	Desagregação e estimativa de uso dos espaços
		<i>Multi-resolution segmentation</i>	Segmentação de imagens
		<i>K-means</i>	Agrupamento de dados em classes
[47]	Mapas de cobertura do solo Imagens de sensores remotos	<i>Central tendency measures</i>	Obtenção de métricas de cada classe
		<i>K-Nearest Neighbor – K-NN</i>	
		<i>Extreme Learning Machine – ELM</i>	Deteção do tipo de cobertura dos espaços
		<i>Support Vector Machine – SVM</i>	
		<i>Dilatation</i>	Remoção de detalhes desnecessários nos mapas
		<i>Subfields-based parallel thinning algorithm</i>	Extração das ruas, estradas e autoestradas
[33]	POIs Trajetórias de táxi Dados de transportes públicos	<i>Two-pass algorithm</i>	Geração de zonas segmentadas
		<i>Latent Dirichlet Allocation – LDA</i>	Descoberta de tópicos com base em padrões de mobilidade
		<i>Dirichlet Multinomial Regression – DMR</i>	
		<i>Hierarchical Semantic Cognition – HSC</i>	Classificação de regiões funcionais
[5]	Imagens de sensores remotos POIs	<i>Multiresolution Segmentation</i>	Segmentação de imagens
		<i>Random Forest – RF</i>	Caracterização do uso das regiões
		<i>ISO- DATA algorithm</i>	Agrupamento de padrões
		<i>Cellular automata model</i>	Geração de zonas de uso nos espaços
[8]	Rede de estradas do OSM Imagens de sensores remotos POIs Atividades em LBSN	<i>Random Forest – RF</i>	Classificação do uso dos espaços
		<i>Object-based classification</i>	Classificação de imagens pré-processadas
		<i>Gray-Level Co-occurrence Matrix – GLCM</i>	Calcular atributos
		<i>Scale Invariant Feature Transform – SIFT</i>	Extração de <i>features</i> em imagens
		<i>K-means</i>	Classificação de <i>features</i>
[1]	Rede de estradas do OSM Imagens de sensores remotos POIs Atividades em LBSN	<i>Probabilistic Latent Semantic Analysis – pLSA</i>	Identificação de atributos
		<i>Latent Dirichlet Allocation – LDA</i>	
		<i>Support Vector Machine – SVM</i>	Classificação dos tipos de uso dos espaços
		<i>Latent Dirichlet Allocation – LDA</i>	Geração de sumários com base em tópicos
[14]	POIs Atividades em LBSN	<i>K-means</i>	Agrupamento de regiões semanticamente similares
		<i>Delaunay triangulation spatial constraints</i>	
		<i>Ward clustering</i>	Identificação de relações topológicas
		<i>Greedy algorithm</i>	Construção de documentos baseados nas TAZ
[2]	POIs Zonas de Análises de Tráfego – TAZ	<i>Word2Vec</i>	Extração de vetores de POIs
		<i>K-Means</i>	Agrupamento das TAZ
		<i>Random Forest – RF</i>	Classificação do uso dos espaços
		<i>TF-IDF algorithm</i>	Transformação das frequências de palavras em atributos
[23]	Imagens de sensores remotos	<i>Random Forest – RF</i>	Classificação de padrões de uso dos espaços
		<i>Google Inception</i>	Deteção de padrões de uso dos espaços
[4]	POIs Mensagens de texto <i>Building blocks</i>	<i>Latent Dirichlet Allocation – LDA</i>	Cálculo de atributos
		<i>Random Forest – RF</i>	Classificação das regiões funcionais
[24]	Imagens de sensores remotos POIs Redes de estradas	<i>Example-based feature extraction</i>	Produção de mapas de cobertura dos espaços
		<i>Multi-resolution segmentation</i>	Segmentação de imagens
		<i>Object-based classification</i>	Classificação da cobertura dos espaços
[22]	Imagens de sensores remotos Redes de estradas	<i>Skeleton-based decomposition</i>	Decomposição de imagens multiespectrais
		<i>Semi-transfer deep convolutional neural network</i>	Mapeamento do uso dos espaços
		<i>Inverse Distance Weight – IDW</i>	Construção de ligações de diversos tipos
[27]	Atividades em LBSN Imagens de sensores remotos Trajetórias de táxi POIs	<i>Kernel Density Estimation – KDE</i>	Inferir os tipos de usos das construções
		<i>Bayesian model</i>	Calcular probabilidades dos propósitos dos passageiros
		<i>Multiresolution segmentation</i>	Segmentação de imagens
[21]	Imagens de sensores remotos, Redes de estradas	<i>Hierarchical Semantic Cognition – HSC</i>	Classificação de forma hierárquica
		<i>Inverse Hierarchical Semantic Cognition – IHSC</i>	Otimização dos resultados da classificação
[50]	Imagens de sensores remotos	<i>Object-based convolutional neural network</i>	Classificação do uso dos espaços

[9]	POIs	<i>Support Vector Machine - SVM</i>	Classificação do uso e ocupação do solo
		<i>K-means</i>	Agrupamento de POIs e áreas vizinhas
		<i>POI frequency analysis</i>	Anotação das funções de cada região
[29]	POIs Matrizes Origem-Destino – OD	<i>Random Forest – RF</i>	Avaliação e comparação da performance dos modelos
		<i>Place2vec</i>	Extração e classificação das regiões funcionais
[35]	Imagens de sensores remotos	<i>ResNet-50 DCNN</i>	Extração de atributos das imagens
		<i>Space-Time Fusion Algorithm – ESTARFM</i>	Fusão de dados
[6]	Imagens de sensores remotos	<i>Multiresolution segmentation</i>	Segmentação das imagens
		<i>Support vector machine – SVM</i>	Extração dos tipos de uso e ocupação do solo
[51]	Imagens de sensores remotos	<i>Joint Deep Learning – JDL</i>	Classificação do uso e da ocupação do solo

## 2.2 Mapas de densidade/concentração (*heatmaps*)

Os mapas de concentração são uma forma de agrupamento visual de dados. O seu uso como representação da intensidade de fenômenos tem sido adotado em diversos domínios. Embora esse tipo de mapa não tenha uma definição exata na cartografia, é normalmente utilizado para a exploração de grandes quantidades de dados com o objetivo de ajudar a identificar concentrações relevantes. Para fins de visualização, os valores das densidades são classificados em intervalos e esses intervalos são representados por um gradiente de cores [52]–[54].

O método adotado neste estudo para a geração dos mapas de concentração, foi a estimativa de densidade por kernel (*Kernel Density Estimation – KDE*). O uso deste método com dados georreferenciados serve para identificar a densidade de pontos numa zona centralizada em cada coordenada, revelando estatísticas locais com base nos pontos dentro de um raio, que são representados como clusters [52]. O método KDE funciona através da definição de uma superfície simétrica (kernel) para cada coordenada. É então avaliada a distância dessa superfície até um ponto central, com base numa função matemática. Depois é feita a sumarização dos valores das superfícies para esse ponto. Esse processo é repetido para todos os pontos e as sumarizações permitem estimar a densidade das distribuições [55].

O conceito do método surgiu em 1950 e foi inicialmente aplicado a dados univariados, entretanto no contexto dos dados georreferenciados é aplicado a dados multivariados com o objetivo de destacar a distribuição espacial e a intensidade dos pontos. O cálculo da estimativa de densidade baseada em kernel [55] é representado na Equação 4, onde:

$\hat{f}(u, v)$  é a estimativa da densidade na localização  $(u, v)$ ;

$n$  é a quantidade de exemplos;

$h$  é o raio;

$k$  é a função kernel;

$d_i$  é a distância entre os pontos  $(u, v)$  e a localização da observação de ordem  $i$ .

$$\hat{f}(u, v) = \frac{1}{nh^2} \sum_{i=1}^n K\left(\frac{d_i}{h}\right)$$

Equação 4: Cálculo da estimativa de densidade baseada em kernel.

Por meio da definição do kernel sobre cada um dos pontos, é criada uma superfície suave e contínua. Em termos computacionais, para a utilização do KDE, há que se ter em conta algumas parametrizações, nomeadamente o raio, o tamanho dos pixels e a função do kernel [56]. O raio determina a distância de influência à volta de cada ponto. Um raio menor fará com que o *heatmap* tenha detalhes mais finos e menores, enquanto que a utilização de um valor mais alto para este parâmetro dará origem a um mapa mais suave. O tamanho do pixel determina a resolução do *raster* a ser gerado. Em relação à função do kernel, esta, por sua vez, é uma função que controla a taxa na qual a influência de um ponto diminui à medida que a distância do ponto aumenta.

## 2.3 Correlação espacial

Além de revelar a ligação existente entre variáveis ou métricas, os coeficientes de correlação permitem realizar previsões para o valor de uma variável com base em outras. As correlações entre duas variáveis indicam que quando uma delas está presente em certo nível, há uma tendência para que a outra também figure. O facto de haver essa tendência não obriga entretanto uma relação causal, mas releva pura e simplesmente que uma variável ou métrica tende a crescer ou decrescer em função dos valores de outra [57]. As correlações são utilizadas em diversas áreas do conhecimento, com as mais distintas finalidades.

Analisar e modelar associações espaciais é uma questão essencial para os trabalhos de investigação em diferentes domínios das ciências sociais e ambientais. No que diz respeito à análise de dados espaciais, medir associações significa avaliar padrões geográficos de similaridade nas ocorrências em atributos ou conjuntos de dados [58]. Neste contexto, as correlações são divididas em duas categorias distintas. A primeira delas é a autocorrelação espacial. Trata-se de um tipo de correlação envolvendo uma única amostra, sendo, portanto, a relação de uma variável, por exemplo a população das regiões, com ela mesma, geralmente recorrendo ao uso de uma métrica espacial, como a distância euclidiana. Em contrapartida, existe também o conceito de correlação espacial cruzada. Esta, por sua vez, pode ser descrita como uma forma de relacionar mais de uma variável, com recurso à uma medida espacial contínua. Enquanto o primeiro é já um conceito bastante consolidado na literatura, a teoria e as metodologias para o segundo ainda não foram bem construídas, apesar de ser um termo já frequente nos estudos mais recentes [59]. A autocorrelação já se encontra disponível entre os conjuntos de ferramentas analíticas existentes em vários sistemas SIG. Desta forma, faz-se necessário propor métodos para análises das correlações espaciais cruzadas, de maneira a ajudar na solidificação do conceito.

### 2.3.1 Autocorrelação espacial

A autocorrelação espacial pode ser definida como sendo a correlação existente entre os valores de uma determinada variável em um plano bidimensional, que por sua vez é, ou referenciado localmente, ou ligado a uma estrutura espacial sobre a qual se assenta [60]. Em outras palavras, o

objetivo da autocorrelação espacial é medir até que nível o valor de uma variável em uma determinada localização está ligado ao valor desta mesma variável nas áreas vizinhas [61]. Por meio da adoção de métricas estatísticas baseadas no índice de Moran [62] e no coeficiente de Geary [63], ao longo dos últimos anos foi desenvolvida uma fundamentação teórica bastante consolidada para a análise das autocorrelações espaciais. Esse tipo de correlação tem sido aplicado em diversos domínios para a análise de fenômenos naturais e humanos. A autocorrelação espacial pode ser quantificada através da adoção de diferentes métricas. Embora o coeficiente de Geary seja muito utilizado, o índice Moran I é considerado o método estatístico mais comum para obtenção de indicadores com variação geográfica para esse tipo de análise nas amostras [64]. Esses dois índices são inversamente relacionados.

### 2.3.2 Correlação espacial cruzada

Como forma de complementar as análises de autocorrelação espacial, recentemente surgiu o conceito de correlação espacial cruzada. Apesar de ser possível encontrar alguns trabalhos que utilizam esse tipo de correlação, os métodos adotados variam frequentemente. Esse tipo de correlação normalmente lida com problemas de séries temporais e de dimensões espaciais [59]. Para lidar com séries temporais não estacionárias, é possível utilizar o método de correlação cruzada destendenciada (*Detrended Cross-Correlation Analysis* – DCCA [65]). Já para contemplar as flutuações, pode ser utilizada a teoria das matrizes aleatórias (*Random Matrix Theory* – RMT [66]). Considerando que tanto o método DCCA quanto a matriz RTM são teorias bem formatadas, os estudos que envolvem análise de correlação cruzada com base em séries temporais se desenvolveram notavelmente nos últimos anos [59]. Entretanto, substituindo-se as séries temporais pelos parâmetros espaciais, surge o problema da correlação espacial cruzada.

Alguns estudos, como por exemplo [28], utilizam o Chi-quadrado para analisar as correlações entre métricas extraídas dos dados em formato *raster*. No caso específico deste estudo, as correlações foram utilizadas para analisar a ligação entre o volume de mensagens de texto enviadas em uma LBSN e a cobertura das zonas, em termos das construções existentes, com o objetivo de perceber se as atividades na rede social ocorriam somente nas áreas urbanas ou não. Entretanto, embora o teste do Chi-quadrado possa ser utilizado para a análise de correlações com base em dados espaciais, tal como as tabelas de contingência, esses métodos não são a melhor alternativa para testes formais, considerando que variam conforme as unidades de medida [67].

Utilizando uma combinação de pontos de interesse, informações de controle de trânsito em tempo real e dados das redes de estradas da cidade de Pequim, na China, foi realizado um estudo [68] com o objetivo de analisar os impactos do tipo de uso das regiões e os congestionamentos de trânsito. No estudo, os autores utilizaram um método baseado em técnicas de *clustering* para o agrupamento dos dados. Para a análise das correlações foi utilizada uma regressão linear múltipla com base em 4 variáveis independentes. Embora os resultados alcançados tenham sido

satisfatórios, de acordo com os autores, o trabalho realizado baseou-se em ferramentas comerciais, como por exemplo o ArcGIS e o SPSS, o que dificulta a sua reprodução dada a dependência das licenças dos softwares.

Em 2017 foi publicado um estudo [69] onde os autores basearam-se em padrões de mobilidade obtidos a partir de dados de uma LBSN (Twitter) para inferir os tipos de uso das regiões urbanas, na região de Chicago, USA. Foram também utilizados dados de inquéritos relacionados a viagens e mapas parciais de LULC. Para o agrupamento dos dados foi utilizado um método de *clustering*. Com o objetivo de encontrar similaridades entre os tipos de uso das regiões com base no comportamento dos utilizadores da rede social, os autores utilizaram uma medida de distorção dinâmica no tempo (*Dynamic Time Warping – DTW*). Foi conduzida também uma análise de correlações onde foi utilizado o coeficiente de correlação de Pearson, baseado em métricas extraídas dos dados.

Tendo sido observado na província de Qinghai, na China, que as modificações de LULC ocorrem mais frequentemente em zonas específicas do território, foi realizado um estudo [70] com o objetivo de revelar as ligações existentes entre as vilas e templos e as alterações nos tipos de uso das regiões envolventes. Foram utilizados dois conjuntos de dados que representam os templos e vilas e outros dois representando o uso e a ocupação do solo em duas épocas distintas. Os autores recorreram ao uso de diagramas de Voronoi, métodos de *clustering* e da técnica *Cross K-function* para propor indicadores que podem ser usados para analisar e modelar a associação espacial existente entre combinações de diferentes proporções de dados vetoriais, como por exemplo linhas e pontos, pontos e polígonos e linhas e polígonos.

Com recurso à utilização de métodos de *clustering* e técnicas de *deep learning*, por meio de algoritmos como por exemplo o *Word2Vec*, e com base em pontos de interesse e dados de análise de tráfego por zonas, foi realizado um estudo [2] na província de Cantão, na China. O trabalho teve como principal objetivo a caracterização do uso e da ocupação das regiões a uma escala refinada. Os autores analisaram as correlações existentes entre vetores construídos com base nos pontos de interesse, com o objetivo de ajudar a avaliar o seu potencial para estudos deste género. Os vetores obtidos foram agrupados por meio do *K-Means* e foi então medida a similaridade do cosseno para analisar as ligações existentes entre os grupos formados. Foi também realizado o agrupamento das zonas de tráfego, que representam os tipos de uso e ocupação das regiões, o que permitiu medir a sua similaridade com os grupos de vetores formados por pontos de interesse.

Também na China, foi realizado um outro estudo [71] para analisar as ligações existentes entre os tipos de uso das regiões e a concentração de metais pesados em uma bacia hidrográfica no planalto de Qinghai-Tibet. Foram utilizadas amostras de solo coletadas na superfície e no subsolo para verificação das diferentes concentrações dos poluentes. Foram também utilizados dados de sensores remotos. Para analisar as ligações entre os dados extraídos das amostras e os

vários tipos de cobertura do solo nos locais de extração, os autores recorreram ao cálculo do coeficiente de correlação de Pearson, por meio do uso das bibliotecas para análise espacial disponíveis na linguagem R. Foram também analisadas as correlações por meio de uma coleção de ferramentas estatísticas disponível em um software proprietário (ArcGIS), onde as concentrações de metais pesados foram representadas em uma camada e os tipos de cobertura do solo, em outra. Embora o trabalho apresentado na seção 5.2 também tenha recorrido à linguagem R para a análise da correlação existente entre diferentes camadas de dados do tipo *raster*, o método utilizado é diferente. Enquanto o estudo conduzido por Dai et al. [71] baseou-se no uso de um sistema GIS proprietário, a metodologia proposta neste trabalho recorre à utilização de tecnologias gratuitas para um objetivo similar.

### 2.3.3 Coeficiente de correlação de Pearson

O coeficiente de correlação de Pearson, também chamado de “coeficiente de correlação produto-momento” é uma das métricas mais comuns para inferir ligações entre variáveis. É utilizado para dados contínuos, isto é, medidos em uma escala de intervalo ou proporção. Apesar dos créditos pelo desenvolvimento deste coeficiente terem sido atribuídos a Karl Pearson, pela publicação da fórmula em 1895, Francis Ysidro Edgeworth, um cientista britânico, realizou uma publicação 3 anos antes, onde apresentava os cálculos para um coeficiente muito semelhante. Entretanto, devido ao facto deste último ter incluído o seu trabalho em um artigo estatístico com uma estrutura bastante complicada, os seus contributos só foram notados anos depois [57].

O coeficiente de correlação de Pearson permanece bastante popular nos dias atuais principalmente porque apresenta algumas características bastante úteis. Trata-se de uma medida que é fácil de programar em qualquer linguagem que suporte operações básicas aritméticas. Normalmente encontram-se também funcionalidades para a realização do cálculo em softwares como o Microsoft Excel e o SPSS e até mesmo em conjuntos de ferramentas incluídos em sistemas GIS como o ArcGIS. Várias tecnologias de programação, como por exemplo o Python e o R também incluem bibliotecas para se trabalhar com o coeficiente. Além disso é relativamente fácil de interpretar, tendo em conta que é uma medida da associação linear que quantifica a tendência de variação que uma variável apresenta em função de outra [72].

Em termos matemáticos, para duas sequências,  $\{x_i\}$  e  $\{y_i\}$ , ambas de tamanho  $N$ , o coeficiente de correlação produto-momento pode ser calculado por meio da Equação 5 [73], onde,  $\bar{x}$  e  $\bar{y}$  representam as médias das sequências.

$$r = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2 \sum_{i=1}^N (y_i - \bar{y})^2}}$$

Equação 5: Fórmula para o cálculo da correlação de Pearson.

Os valores possíveis de serem obtidos na equação variam entre -1 e 1. Em termos de interpretação, o valor 1 representa uma correlação positiva perfeita, enquanto -1 demonstra uma correlação negativa perfeita. Neste caso o valor 0 representa a ausência de correlações [72].

#### 2.3.4 Coeficiente de correlação ponto-bisserial

O coeficiente de correlação ponto-bisserial é uma variação do coeficiente de correlação de Pearson. Por meio dele é possível analisar as associações existentes entre variáveis dicotômicas (binárias) e contínuas. Basicamente uma variável dicotômica é qualitativa e assume somente dois valores possíveis. Alguns exemplos desse tipo de variável são: “masculino ou feminino”, “licenciado ou não licenciado”, “sucesso ou falha”, etc. Sendo assim, para o cálculo do coeficiente de correlação ponto-bisserial, a variável dicotômica é sempre convertida para um formato numérico, assumindo o valor “0”, para uma classe, e o valor “1” para a outra [74].

Tal como no coeficiente de correlação produto-momento, os valores possíveis de serem obtidos no resultado da correlação ponto-bisserial, variam entre -1 e 1 e são, portanto, interpretados da mesma maneira. A fórmula para realizar o cálculo do coeficiente é mostrada a seguir, na Equação 6 [74], [75], onde, considerando que a variável dicotômica é  $X$ :

$Y$  é a variável contínua;

$\Sigma Y_1$  é a soma dos valores de  $Y$  para todas as observações onde o valor de  $X$  é igual a 1;

$\Sigma Y$  é a soma de todos os valores de  $Y$ ;

$\Sigma Y^2$  é a soma de todos os valores de  $Y$  levantados ao quadrado;

$N_1$  é o número de observações onde o valor de  $X$  é igual a 1;

$N_0$  é o número de observações onde o valor de  $X$  é igual a 0;

$N$  é o número total de observações;

$$r_{pb} = \frac{N \Sigma Y_1 - N_1 \Sigma Y}{\sqrt{N_1 N [N \Sigma Y^2 - (\Sigma Y)^2]}}$$

Equação 6: Fórmula para o cálculo da correlação ponto-bisserial.



## 3 Ferramentas e tecnologias

Neste capítulo são apresentadas as principais ferramentas envolvidas no desenvolvimento do estudo. Também são apresentadas as tecnologias adotadas, como por exemplo as linguagens de programação utilizadas para construir a aplicação para coleta dos dados e o software para geração dos mapas de densidade. De maneira geral, são abordados desde a tecnologia de containerização escolhida para executar a aplicação de coleta dos dados, o motor de base de dados para manipulação de dados georreferenciados, o sistema SIG adotado, o software utilizado para as análises relacionadas à caracterização dos espaços e as linguagens de programação com base nas quais foram desenvolvidas as aplicações envolvidas nas diversas etapas do estudo.

### 3.1 Docker

Devido aos recentes progressos tecnológicos, a relação entre as empresas e o uso dos softwares modernos se tornou cada vez mais evidente. De certa forma, esta relação reforça a necessidade das empresas, mesmo aquelas que não são dos setores da Tecnologia de Informação – TI, terem cada vez mais o domínio das diversas ferramentas informáticas existentes. Em outras palavras, a grande maioria das empresas atuais torna-se indiretamente parte da indústria de software [76], seja pela dependência do desenvolvimento de novos sistemas para uso interno ou ainda pela criação de novas experiências baseadas em software para disponibilizar aos seus clientes.

O ambiente informático para suportar o uso dos softwares modernos, seja por parte das equipas internas das empresas ou por parte do cliente final, deve ser preferencialmente uma etapa do trabalho que não demande grandes esforços de criação e manutenção. Os requisitos desses ambientes podem variar de acordo com as especificidades de cada caso, entretanto de modo geral, o uso de arquiteturas baseadas em virtualização pode fornecer o isolamento necessário em muitos desses ambientes.

Docker é uma ferramenta de containerização, de uso gratuito, que permite a criação de containers para a instalação e execução de aplicações de maneira rápida e eficiente [77]. Os containers Docker disponibilizam ambientes onde é possível adicionar e configurar recursos e dependências para as aplicações. De forma geral, os containers são construídos sobre o núcleo do sistema operativo, tirando proveito dos seus recursos para criar um isolamento que permite a execução dos processos de maneira independente [78]–[80].

Entre as vantagens do uso de containers, destaca-se não só o isolamento necessário para a execução das aplicações, mas também o facto do baixo consumo de recursos do sistema operativo

*host* e do hardware. Os containers são construídos ao mesmo nível das aplicações, porém de maneira isolada, e apesar de partilharem o mesmo núcleo do sistema operativo *host*, eles não podem acedê-lo diretamente [78]. Isso significa que qualquer alteração feita é refletida somente no container e não no sistema operativo em si. A principal diferença entre um container e uma máquina virtual é que uma máquina virtual executa um sistema operativo completo, incluindo o seu próprio núcleo.

É importante ressaltar que o núcleo do sistema operativo não disponibiliza tudo o que é necessário para executar aplicações. Existem muitas APIs e serviços que são oferecidos por bibliotecas que situam-se acima do nível do núcleo [78]. Sendo assim, devido ao facto de o container ser isolado, torna-se necessário a existência de uma cópia desses serviços e APIs. Neste contexto surge o conceito das imagens. Uma imagem é basicamente uma camada sobre a qual o container é criado, e que fornece os serviços que o núcleo não disponibiliza. Desta forma, em cada imagem existem diferentes camadas empilhadas umas sobre as outras e em cada camada existem diferentes ficheiros agrupados.

Para construir as imagens, a ferramenta Docker utiliza um documento de texto chamado Dockerfile. Um ficheiro Dockerfile contém basicamente os comandos que podem ser executados em linha de comandos para a construção da imagem. Durante a construção da imagem, esses comandos são executados de forma sequencial [81]. Portanto, por meio do uso desses ficheiros, é possível automatizar a criação das imagens, permitindo desta forma que se possa reproduzir e disponibilizar rapidamente o ambiente contendo todas as dependências e configurações necessárias para testar ou executar os softwares, seja na fase de desenvolvimento ou mesmo em produção. De acordo com a Microsoft [78], isso reduz o custo e acelerara o desenvolvimento e a execução de aplicações.

Os containers Docker podem ser executados tanto em computadores e servidores baseados em Windows quanto em Linux. Por este e por todos os outros motivos já referidos, a tecnologia Docker foi escolhida para execução da aplicação desenvolvida para coletar os dados utilizados neste estudo. Pelo facto de ter sido criada em PHP, o desenvolvimento e a execução desta aplicação exigiram um ambiente que tivesse um servidor WEB e um interpretador da linguagem instalado e configurado. Sendo assim, a ferramenta facilitou a criação deste ambiente, permitindo que o mesmo seja replicado facilmente no futuro, caso necessário. A aplicação referida será devidamente apresentada no capítulo seguinte.

## 3.2 PostgreSQL

A evolução da tecnologia impulsiona um aumento cada vez mais evidente na quantidade de dados gerados, seja pelos utilizadores finais, pelos próprios dispositivos móveis, transações automáticas nos sistemas informáticos ou por diversas outras fontes. Desta forma, armazenar e gerir os dados,

revela-se uma tarefa desafiadora nos dias atuais, de modo que os Sistemas de Gestão de Bases de Dados (SGBD) tornaram-se cada vez mais populares, existindo atualmente uma grande variedade de softwares para esta mesma finalidade.

O PostgreSQL é um SGBD relacional<sup>1</sup> bastante conhecido na indústria e entre a comunidade científica [82]. Conforme descrito no site oficial [83], a ferramenta é *open source* e está a ser desenvolvida e efetivamente melhorada há mais de 30 anos, o que faz com que seja bastante confiável, robusta e tenha boa performance. Além disso, o PostgreSQL possui uma ampla comunidade de utilizadores a nível mundial. A ferramenta foi desenvolvida na Universidade da Califórnia [84]. Apesar de ter sido inicialmente criado para plataformas UNIX, atualmente o PostgreSQL está também disponível para outros sistemas operativos, como por exemplo Windows e Mac OS.

Existe um grande número de extensões disponíveis para o PostgreSQL, como por exemplo a extensão PostGIS, desenvolvida para permitir a gestão e a manipulação de dados georreferenciados. O SGBD também traz uma vasta gama de funcionalidades que foram criadas para facilitar o desenvolvimento de aplicações, reforçar a proteção da integridade dos dados, fornecer um nível seguro de tolerância a falhas e facilitar a gestão, seja de grandes ou pequenas quantidades de dados [85]. O PostgreSQL é também conhecido pela sua flexibilidade, permitindo criar tipos de dados e funções personalizadas, suportando igualmente a escrita de código em diferentes linguagens de programação [86]. Além disso, de acordo com o site oficial [85], o SGBD provou ser altamente escalável tanto na quantidade de dados que pode gerir, quanto na quantidade de utilizadores que é capaz de suportar.

Neste estudo foi proposto um modelo relacional para armazenar os dados coletados. Sendo assim e por todos os demais motivos apresentados, foi adotado o PostgreSQL como SGBD. Por se tratar de um estudo que se baseia na utilização de dados georreferenciados, também foi utilizada a extensão PostGIS. No caso específico deste projeto, uma outra grande vantagem da adoção do PostgreSQL, foi a capacidade que este possui de manipular dados no formato JSON [87], pois como pode ser visto no capítulo 4, os dados coletados trazem um atributo estruturado neste formato.

### 3.3 PostGIS

Devido à popularização das tecnologias de GPS (*Global Positioning System*), a geração de dados georreferenciados tornou-se uma realidade cada vez mais evidente ao longo dos últimos anos. Com o aumento do volume desses dados, a comunidade científica enfrentou uma mudança de paradigma na análise de dados, tendo em vista a necessidade de lidar com um novo atributo que representa a

---

<sup>1</sup> <https://www.oracle.com/database/what-is-a-relational-database/> [último acesso: 11/03/2020]

posição geográfica dos registos. Neste contexto surgiu uma forte necessidade da existência de bases de dados capazes de armazenar e manipular dados georreferenciados. Para responder a esta necessidade, diferentes abordagens foram adotadas pelas diferentes comunidades e empresas responsáveis pelos SGBDs existentes. Para os utilizadores do PostgreSQL, esta resposta veio principalmente por meio da criação da extensão PostGIS.

De acordo com o site oficial [88], para além das funcionalidades básicas para se trabalhar com dados georreferenciados, a extensão traz diversas outras, que normalmente não são encontradas em SGBDs que nativamente suportam esse tipo de dados. Além disso a ferramenta é de código fonte aberto, está disponível para uso gratuito e conforme descrito pela fundação OSGeo [89], permite o uso de vários outros tipos de dados, como por exemplo dados geométricos, geográficos, *raster*, topografias, etc. Em outras palavras, por meio do uso da extensão PostGIS, pode-se armazenar e manipular tipos de dados georreferenciados como qualquer outro tipo convencional suportado pelo PostgreSQL. Dessa forma, com as novas funcionalidades adicionadas pela extensão, é possível responder questões como “quais coordenadas estão dentro dos limites de uma determinada zona?”, ou “qual o tamanho de um determinado distrito?” [90].

Neste estudo, o trabalho realizado é baseado na extração de conhecimento a partir de dados de Pontos de Interesse. Portanto, a característica mais importante dos dados neste caso é essencialmente a georreferenciação, não sendo possível a obtenção dos mesmos resultados com o uso de outros tipos de dados que não fossem espaciais. Desta forma, adotou-se a extensão PostGIS para permitir que fossem aproveitadas as vantagens da georreferenciação dos dados por meio do uso da tecnologia PostgreSQL, evitando assim a necessidade de utilização de diferentes ferramentas para manipular e gerir conjuntos e ficheiros separados de dados, o que normalmente é feito em casos onde não se opta pelo uso de uma base de dados espacial.

### 3.4 QGIS

Com o aumento da disponibilidade dos dados georreferenciados, ocasionado pela popularização das tecnologias móveis baseadas em geolocalização, a necessidade de sistemas informáticos cada vez mais eficientes para exploração e análise desse tipo de dados, tornou-se uma realidade cada vez mais evidente. Neste cenário surgiram os sistemas SIG. De acordo com o National Geographic [91], os sistemas SIG são softwares desenvolvidos para capturar, armazenar, explorar e analisar dados georreferenciados, permitindo que as pessoas e organizações encontrem padrões e relações espaciais existentes nos dados. São, portanto, indispensáveis para diferentes questões relacionadas à aquisição, processamento, distribuição, uso, manutenção e preservação dos dados espaciais.

Atualmente existem diversos sistemas SIG no mercado. Alguns deles são proprietários, como é o caso do ArcGIS, e outros são de uso livre. Nesta segunda categoria, o Quantum GIS (QGIS) provavelmente é um dos softwares mais conhecidos pela comunidade [92]. O QGIS é de

código fonte aberto e de uso livre sob a licença GNU<sup>2</sup> (*General Public License*). O projeto é de responsabilidade da fundação OSGeo<sup>3</sup>, embora também receba diversos contributos da comunidade, desde novas funcionalidades a *bug fixes* e *bug reports*, documentações, suporte aos utilizadores, etc. [93]. Por ser amplamente utilizado no meio académico e na indústria, o software conta com uma vasta comunidade a nível mundial, o que facilita a solução de problemas. Além disso, está disponível para diferentes plataformas, como por exemplo Windows, Linux e Mac OS.

O QGIS conta atualmente com um grande número de *plug-ins*, que estendem as suas funcionalidades nativas. Além disso, novos *plug-ins* podem ser criados pelo utilizador, caso seja necessário. Outra vantagem deste software é que possibilita o uso dos módulos GRASS<sup>4</sup> por meio da sua interface gráfica [94]. O QGIS também suporta diversos formatos de dados, incluindo *Shapefiles*, *rasters*, *GeoTIFFs* e muitos outros.

Além das características citadas anteriormente, o QGIS também traz uma interface de programação em Python bastante robusta, que permite estender as funcionalidades nativas, automatizar a realização de tarefas ou criar aplicações *stand alone*, que não obrigam o utilizador a ter o modo gráfico do QGIS em execução, para o seu funcionamento. Esta flexibilidade foi fundamental para a etapa do estudo onde se realizou a análise das correlações, considerando que foi desenvolvida uma aplicação para automatizar o processo de criação dos mapas de densidade utilizados. Além disso, o QGIS também foi essencial para a manipulação, visualização e análise dos dados, como se verifica com base no conteúdo apresentado nos próximos capítulos.

### 3.5 RapidMiner

Uma questão essencial em todas as áreas científicas diz respeito às ferramentas disponíveis. Relativamente à análise de dados não é diferente. Existem vários softwares, que oferecem diferentes ferramentas para facilitar o trabalho dos cientistas nessa área. Apesar desta diversidade, alguns deles, entretanto, trazem um conjunto mais amplo de funcionalidades. É o caso do RapidMiner [95]. Apesar de ser um produto comercial, está disponível para os estudantes uma licença específica, que permite o uso do software. O RapidMiner é uma suíte de ferramentas que disponibiliza várias soluções para a análise de dados, incluindo integrações com várias outras tecnologias, como por exemplo os sistemas de gestão de base de dados [96]. Além disso permite realizar o carregamento e a transformação de dados a partir de várias fontes e inclui um conjunto completo de funcionalidades para aprendizagem computacional.

No que diz respeito às questões de usabilidade, o software apresenta uma interface amigável e intuitiva, com uma curva de aprendizagem relativamente baixa. No RapidMiner, as

<sup>2</sup> <https://www.gnu.org/licenses/gpl-3.0.en.html> [último acesso: 11/03/2020]

<sup>3</sup> <https://www.osgeo.org/> [último acesso: 11/03/2020]

<sup>4</sup> <https://grass.osgeo.org/> [último acesso: 11/03/2020]

análises são desenhadas em um esquema de fluxo, onde o utilizador escolhe a operação que deseja implementar e adiciona-a à parte do processo onde pretende que seja executada. As operações estão disponíveis em uma lista e cada uma delas está devidamente documentada. O software oferece ao utilizador um conjunto relativamente amplo de tutoriais onde são mostrados exemplos de implementação para as várias operações. É possível ainda fazer a depuração dos processos, bastando adicionar um ponto de interrupção nas operações onde o utilizador preferir [96].

Relativamente às funcionalidades, estão disponíveis diversos algoritmos de *data mining* e *machine learning*, desde técnicas simples de *clustering* até classificadores mais complexos, como as redes neuronais. O software oferece também operadores para avaliação da performance dos algoritmos, que podem ser utilizados tanto com base em *split validation* quanto com *cross-validation*. É possível ainda visualizar os detalhes do modelo criado, como por exemplo uma árvore de decisão, com seus nós e ramos [96]. Por todos os motivos referidos, a ferramenta foi escolhida para as análises relacionadas às questões da caracterização do uso e da ocupação dos espaços, apresentada no tópico 5.1.

## 3.6 Linguagens de programação

Conforme citado anteriormente, neste estudo optou-se pela automação de muitas das tarefas realizadas. Para permitir tal automação, diferentes softwares foram criados, em diferentes linguagens de programação. Esta seção apresenta as linguagens de programação utilizadas no desenvolvimento desses softwares.

### 3.6.1 PHP

De acordo com o site oficial [97], o PHP (*Hypertext Preprocessor*) é uma linguagem de programação de código fonte aberto, amplamente utilizada no meio académico e na indústria. É especialmente recomendada para desenvolvimento de softwares para a Web. Trata-se de uma linguagem interpretada do lado do servidor, o que a torna bastante segura e independente das plataformas utilizadas pelo cliente. Além disso é flexível, rápida, estável e conta com uma vasta comunidade de utilizadores a nível mundial [98].

Tendo em vista que os dados iniciais deste estudo foram coletados por meio de uma API disponibilizada pelo Facebook, optou-se pela utilização do PHP no desenvolvimento da aplicação que automatizou a recolha dos dados. Esta escolha deve-se não só às diversas vantagens já mencionadas da linguagem, mas também ao facto de o Facebook disponibilizar um SDK<sup>5</sup> (*Software Development Kit*) construído nesta mesma linguagem, o que facilitou o desenvolvimento da aplicação.

---

<sup>5</sup> <https://github.com/facebookarchive/php-graph-sdk/tree/master/docs> [último acesso: 11/03/2020]

### 3.6.2 Java

Conforme descrito pela Oracle, o Java é uma linguagem de programação e também uma plataforma computacional desenvolvida inicialmente pela Sun Microsystems, em 1995. Entre as principais vantagens do Java, a empresa destaca a velocidade, a segurança e a confiabilidade [99]. Além disso a linguagem baseia-se no conceito WORA (*Write Once, Run Anywhere*), o que significa que as aplicações desenvolvidas podem ser executadas em todas as plataformas que suportam o Java, tendo em vista que neste caso a compilação do código é feita para *byte code*<sup>6</sup>, que por sua vez é interpretado pela JVM (*Java Virtual Machine*) [100].

Neste estudo, a geração dos mapas de densidade utilizados para representação dos dados e para a análise das correlações foi feita em duas etapas. Uma delas foi a criação de ficheiros do tipo CSV (*Comma-Separated Values*) contendo os dados dos pontos de interesse e a outra foi a leitura desses ficheiros e a geração dos *heatmaps* em si. Sendo assim, optou-se pela utilização do Java para a criação de duas aplicações de linha de comandos, que se ligam à uma base de dados PostgreSQL, buscam os dados com base em critérios pré-definidos e geram os ficheiros contendo, no caso de uma delas, os pontos de interesse e no caso da outra, as métricas a serem utilizadas para a classificação do uso dos espaços.

### 3.6.3 Python

De acordo com o site oficial [101], o Python é uma linguagem de programação de alto nível, orientada a objetos. Por se tratar de uma linguagem interpretada, o Python pode também ser utilizado para o desenvolvimento de *scripts*. É relativamente simples, tendo uma curva de aprendizagem considerada baixa, o que reduz os custos de desenvolvimento e manutenção e ao mesmo tempo a torna altamente atrativa, principalmente pelo aumento de produtividade que proporciona. Outro aspeto importante é que o Python suporta módulos e *packages*, o que incentiva a modularização e facilita a reutilização de código. Ainda de acordo com o site, o interpretador está disponível gratuitamente para as principais plataformas, tal como as várias bibliotecas *standard*.

Conforme referido anteriormente neste capítulo, o QGIS suporta o uso do Python para o desenvolvimento de extensões ou para a automação de tarefas, permitindo até mesmo a criação de aplicações *standalone*, que podem ser executadas em linha de comandos e não dependem do modo gráfico do software. Por este motivo, aliado a todas as vantagens da linguagem, optou-se pela utilização da mesma no processo de geração automática dos mapas de densidade, utilizados tanto na exploração visual dos dados, quanto na análise de correlações.

---

<sup>6</sup> <https://www.javatpoint.com/java-bytecode> [último acesso: 11/03/2020]

### 3.6.4 R

Conforme descrito no site oficial [102], além de ser uma linguagem de programação, o R é também um ambiente computacional para análises estatísticas. É gratuito para uso, está disponível sob a licença GNU. A linguagem R suporta uma grande variedade de técnicas estatísticas e além disso é altamente extensível. Entre as vantagens do R, destacam-se a alta qualidade dos gráficos gerados. Também podem ser geradas fórmulas e símbolos matemáticos, quando necessário. Em termos de plataformas, o R pode ser compilado e executado em vários ambientes UNIX, Windows e Mac OS. A tecnologia inclui uma suíte de ferramentas para a manipulação e análise de dados, cálculos e visualizações.

Neste estudo, para a análise das correlações com base nos mapas de densidade, optou-se pela utilização da linguagem R. Além das vantagens apresentadas, o R permite a utilização de diferentes bibliotecas. Uma delas, chamada simplesmente de “*raster*”, traz diversas funcionalidades para análise estatística de camadas desse mesmo tipo, incluindo a funções para o cálculo de correlações e o suporte à reprojeções de dados georreferenciados.

## 4 Dados

Neste capítulo são apresentados os dados envolvidos na realização do estudo. Foram utilizadas duas categorias diferentes de dados. A primeira delas é onde se enquadram os pontos de interesse, enquanto que a segunda é constituída por conjuntos de dados georreferenciados que representam o uso e a cobertura dos espaços no país. Desta forma, neste capítulo são apresentadas as principais características dos dados adotados. No caso dos pontos de interesse, tendo em vista que o processo estabelecido desde a coleta até a representação e a extração de conhecimento a partir dos dados envolveu diferentes etapas, também são apresentados os passos seguidos desde a coleta até à análise exploratória dos mesmos.

### 4.1 Pontos de interesse

Nos últimos anos nota-se uma rápida e clara disseminação das tecnologias móveis baseadas em geolocalização. Neste contexto e principalmente devido aos sensores GPS existentes nos dispositivos móveis, como por exemplo os *smartphones*, surgiu uma oportunidade de negócio explorada rapidamente por empresas do setor: as LSBN, como por exemplo o Facebook, o Twitter e o Instagram. O surgimento dessas redes sociais proporcionou um notável aumento na disponibilidade de vários tipos de dados georreferenciados. Um exemplo amplamente conhecido dessa categoria de dados, são os Pontos de Interesse - POIs.

De acordo com o site do projeto OpenStreetMap [103], o termo Ponto de Interesse é usado na cartografia para representar uma localização em particular, em um mapa. Trata-se, portanto, de uma referência a um ponto específico, que pode ser por exemplo um restaurante, um hotel, uma indústria, um centro de saúde, etc. De acordo com Yao et al. [2], os POIs têm sido amplamente utilizados em estudos relacionados aos espaços urbanos. Ainda segundo a visão de Jiang et al. [49], grandes quantidades de pontos de interesse estão disponíveis online. É o caso por exemplo, dos que foram utilizados neste trabalho.

#### 4.1.1 Facebook Places

A democratização do acesso à Internet vivenciada nos últimos anos, aliada à popularização dos dispositivos móveis tal como os *smartphones* e *tablets*, revelou uma crescente necessidade de as empresas estarem cada vez mais presentes *online*. A evolução do comportamento do consumidor na era digital, obrigou o estabelecimento de novas formas de *marketing* e entre essas novas práticas, destaca-se divulgação dos produtos e serviços nas redes sociais. Neste cenário e dado o

advento das redes sociais baseadas em geolocalização, surgiram diferentes opções através das quais as empresas podem disponibilizar informações e realizar campanhas na Internet. Uma dessas opções é a criação de páginas no Facebook, que é uma das redes sociais com maior número de utilizadores atualmente [104].

Durante a criação de uma página de uma empresa, o Facebook permite que seja adicionada a localização geográfica. Neste caso a página fica automaticamente disponível como um “Facebook Place”, permitindo que os utilizadores façam *check-in* e partilhem fotos e comentários sobre o local [105]. De acordo com Doug Gross [106], quando um utilizador faz *check-in* em um determinado local, a ideia é que os seus amigos que estão próximo possam visualizar esta atividade e possivelmente também visitar aquele mesmo local. Desta forma, o Facebook Places aumenta geralmente a visibilidade das empresas e permite que os utilizadores encontrem produtos e serviços mais facilmente, o que torna a ferramenta bastante atrativa não só para quem atende diretamente o consumidor final, mas para qualquer segmento do mercado.

O Facebook Places possui um conjunto de categorias<sup>7</sup>, as quais estão organizadas de maneira hierárquica, o que permite distinguir facilmente os diversos tipos de pontos de interesse existentes. Este sistema de categorias está em Inglês e na data da coleta dos dados para a realização do estudo, vinculadas aos POIs localizados em Portugal, existiam 1225 categorias diferentes, embora o número possibilidades seja bem maior no total a nível mundial. Numa consulta simples ao serviço fornecido pelo Facebook, foram obtidas 1556 categorias, organizadas em 8 grupos principais, que se dividem em vários níveis, cada um com diferentes subníveis.

O Facebook disponibiliza uma API que permite buscar diversos tipos de informação. Alguns dos seus serviços fornecem dados relacionados às contas dos utilizadores, outros sobre as páginas existentes, eventos ou lugares, sempre de acordo com as leis de proteção de dados em vigor. Especificamente no caso deste último serviço, é possível realizar buscas em zonas específicas do território. De acordo com o site oficial [107], a API do Facebook fornece informações sobre mais de 140 milhões de locais distribuídos ao longo dos diversos países. Conforme abordado na secção 4.1.2.1, esta foi a opção utilizada para obter o conjunto de dados que permitiu a realização do trabalho.

#### 4.1.2 O dataset

Para este estudo foram coletados 171 177 pontos de interesse em Portugal continental, por meio de um serviço na API do Facebook. O *dataset* obtido possui 45 atributos, conforme se pode ver a seguir, na Tabela 2, onde os atributos sublinhados são específicos dos pontos de interesse da categoria “restaurants”. Entre os atributos existentes, alguns foram criados simplesmente por questões de *debug* para que se pudesse ter maior controlo sobre a execução das requisições

---

<sup>7</sup> <https://www.facebook.com/pages/category/> [último acesso: 11/03/2020]

efetuadas pela aplicação que foi desenvolvida para a coleta dos dados, nomeadamente os atributos “autoink\_pk”, “datetime”, “url” e “coords”.

Tabela 2: Descrição dos atributos do *dataset* utilizado no estudo.

<b>Atributo</b>	<b>Descrição</b>
autoink_pk	Chave primária da tabela na base de dados local
id	Identificador do POI na base de dados do Facebook
name	Nome do ponto de interesse
about	Informação diversa
description	Descrição
category_list	Lista das categorias a que pertence
checkins	Número de check-ins
is_permanently_closed	Se está permanentemente encerrado
is_verified	Se foi verificado formalmente ou não
link	Link para a página do POI no Facebook
single_line_address	Endereço completo
website	Link do site oficial na Internet
hours	Horário de funcionamento
espec_fld_city	Cidade
espec_fld_city_id	Identificador da cidade na base de dados do Facebook
espec_fld_country	País
espec_fld_country_code	Código do país
espec_fld_located_in	Zona onde o POI está localizado
espec_fld_latitude	Latitude
espec_fld_longitude	Longitude
espec_fld_name	Nome da zona onde se localiza
espec_fld_region	Região onde se situa
espec_fld_region_id	Identificador da região geográfica, na base de dados do Facebook
espec_fld_state	Estado onde se localiza o POI
espec_fld_street	Rua
espec_fld_zip	Código postal
<u>espec_fld_delivery</u>	Se faz entregas ou não
<u>espec_fld_catering</u>	Se serve refeições
<u>espec_fld_groups</u>	Se é um local adequado para grupos
<u>espec_fld_kids</u>	Se é adequado para crianças
<u>espec_fld_outdoor</u>	Se possui lugares na parte externa
<u>espec_fld_pickup</u>	Se o possui serviço de coleta
<u>espec_fld_reserve</u>	Se faz reservas
<u>espec_fld_takeout</u>	Se vende comida preparada para viagem
<u>espec_fld_waiter</u>	Se tem serviço de mesa
<u>espec_fld_walkins</u>	Se faz atendimento sem reserva

<u>espec fld breakfast</u>	Se serve pequeno almoço
<u>espec fld coffee</u>	Se serve cafés
<u>espec fld dinner</u>	Se serve jantar
<u>espec fld drinks</u>	Se serve bebidas
<u>espec fld lunch</u>	Se serve refeições além do pequeno almoço, almoço ou jantar
datetime	Data e hora em que o POI foi descarregado da API
url	URL do pedido realizado à API
coords	Coordenadas geográficas do ponto de interesse
geom_point	Representação da localização geográfica por meio de um atributo do tipo <i>point</i> .

Os pontos de interesse são os principais dados deste trabalho. Embora o *dataset* obtido seja bastante completo, quando se trata da análise dos dados, os atributos importantes variam dependendo do estudo que se pretende. No caso deste estudo em particular, considera-se que a localização e as categorias às quais um determinado ponto de interesse pertence, são as características mais relevantes. Em resumo, as categorias representam o perfil de cada ponto de interesse, ou mesmo o segmento de cada empresa ou serviço, enquanto que a localização geográfica permite perceber qual a correlação existente entre cada categoria, por meio da proximidade dos POIs vinculados a ela, como será abordado no capítulo 5, onde também é apresentado um estudo no qual é realizada uma análise das ligações entre as categorias e os tipos de uso e ocupação dos espaços geográficos.

#### 4.1.2.1 Coleta

Embora existam muitos *datasets* disponíveis na Internet, há sempre uma preocupação eminente relativamente à qualidade dos dados. Essa preocupação torna-se ainda mais clara quando o assunto são POIs, principalmente porque em sua maioria, os *datasets* disponíveis gratuitamente nesta categoria, são normalmente criados a partir de dados colaborativos. Facilmente se encontram casos onde os dados trazem registos como por exemplo "semáforos", "rotundas", "passadeiras", etc. Considerando também que esses dados muitas vezes não têm um sistema organizado de categorias, a realização de estudos baseados nesses *datasets* muitas vezes pode implicar um grande esforço no sentido de organizar e limpar os dados, o que frequentemente inviabiliza a sua realização. Nesse sentido, os dados do Facebook Places mostraram-se mais adequados para este projeto. Entretanto obtê-los a nível do território nacional implica um imenso esforço de coleta, que claramente não justifica a sua execução de forma manual.

##### 4.1.2.1.1 Automação do processo de coleta dos dados

Considerando que o Facebook disponibiliza em uma das suas APIs um serviço para realizar pesquisas, verificou-se que seria tecnicamente viável a construção de um software para automatizar a coleta dos dados. Neste sentido, algo que facilitou o desenvolvimento do software

foi o SDK<sup>8</sup> disponível gratuitamente para desenvolvedores. Este SDK foi criado em PHP, motivo pelo qual, dentre outros, optou-se pela utilização desta linguagem para a aplicação pretendida. O kit de desenvolvimento funciona com base em um registo feito do aplicativo no Facebook Developers, sem o qual não é possível a utilização da API.

A aplicação foi desenvolvida com recurso à orientação a objetos e possui duas funcionalidades principais. Entretanto, para discorrer sobre tais funcionalidades, é necessário abordar em primeiro lugar, o funcionamento do serviço utilizado na API. Existem diferentes possibilidades para as consultas dos dados neste serviço. Uma delas é informar uma posição geográfica como parâmetro na requisição. Esta foi a possibilidade que melhor se adequou às necessidades no caso deste estudo. Neste tipo de consulta, deve-se informar também um raio dentro do qual a busca vai ser realizada. Em resumo, a API retorna todos os pontos de interesse dentro do raio informado, tendo como centro a posição geográfica definida. Desta forma, uma das funcionalidades da aplicação que foi criada no âmbito do projeto, para a coleta dos dados, é a geração de uma “grelha”, contendo pontos distribuídos uniformemente ao longo do território. Com base nesta grelha, a outra função executada pela aplicação é a coleta dos pontos de interesse, realizando as requisições necessárias para cada coordenada.

#### 4.1.2.1.1 Geração da grelha de coordenadas

Um dos desafios ao se utilizar este tipo de buscas é a distribuição dos pontos de maneira que permita a cobertura de todo o território que se pretende. Tendo em conta que nos testes realizados o maior raio dentro do qual se conseguiu obter resultados foi 2000 metros, foi necessário encontrar uma maneira de criar uma grelha que considerasse esta configuração. Como se pode ver na parte “A” da Figura 2, caso as coordenadas sejam posicionadas muito longe umas das outras na grelha, várias partes do território ficam descobertas. Em contrapartida, como mostrado na parte “B” da imagem, uma forma mais adequada seria o posicionamento dos pontos de maneira a criar áreas de interseção entre eles. Neste caso, embora durante a execução das requisições haja repetição de resultados, consegue-se garantir que são obtidos todos os POIs distribuídos ao longo do território.

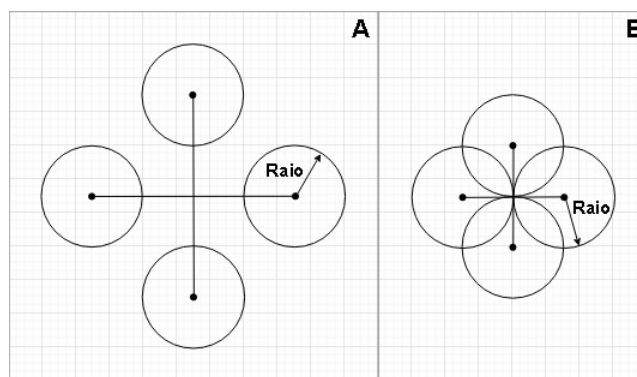


Figura 2: Duas possíveis configurações para a grelha de coordenadas.

<sup>8</sup> <https://developers.facebook.com/docs/php/gettingstarted> [último acesso: 11/03/2020]

Tendo sido encontrada a configuração mais adequada para a criação da grelha, foi utilizado um cálculo matemático para gerar as coordenadas. Foram estabelecidas inicialmente 4 coordenadas à volta da área que seria processada: uma no canto superior esquerdo, uma no canto superior direito, uma no canto inferior esquerdo e outra no canto inferior direito. Essas coordenadas serviram de limite para a execução do algoritmo que determinou os pontos na grelha.

O algoritmo criado pode ser visto no fluxograma representado na Figura 3. Em resumo, funciona da seguinte maneira: em ciclo e partindo da parte inferior esquerda, o algoritmo calcula novos pontos, pela adição de uma distância pré-definida, ou seja, 2000 metros, à **latitude**. Ao atingir o limite superior, é calculada uma nova coordenada que servirá como ponto de partida, na parte inferior esquerda da grelha, por meio da adição de mais 1000 metros à **longitude**. Para a criação desta nova coordenada, de maneira alternada, são adicionados 1000 metros à **latitude** para que o novo ponto “base” fique ligeiramente “acima” dos gerados anteriormente, como destacado em vermelho, na Figura 4. Este processo se repete até que seja atingido o limite superior direito da grelha, quando é então finalizado.

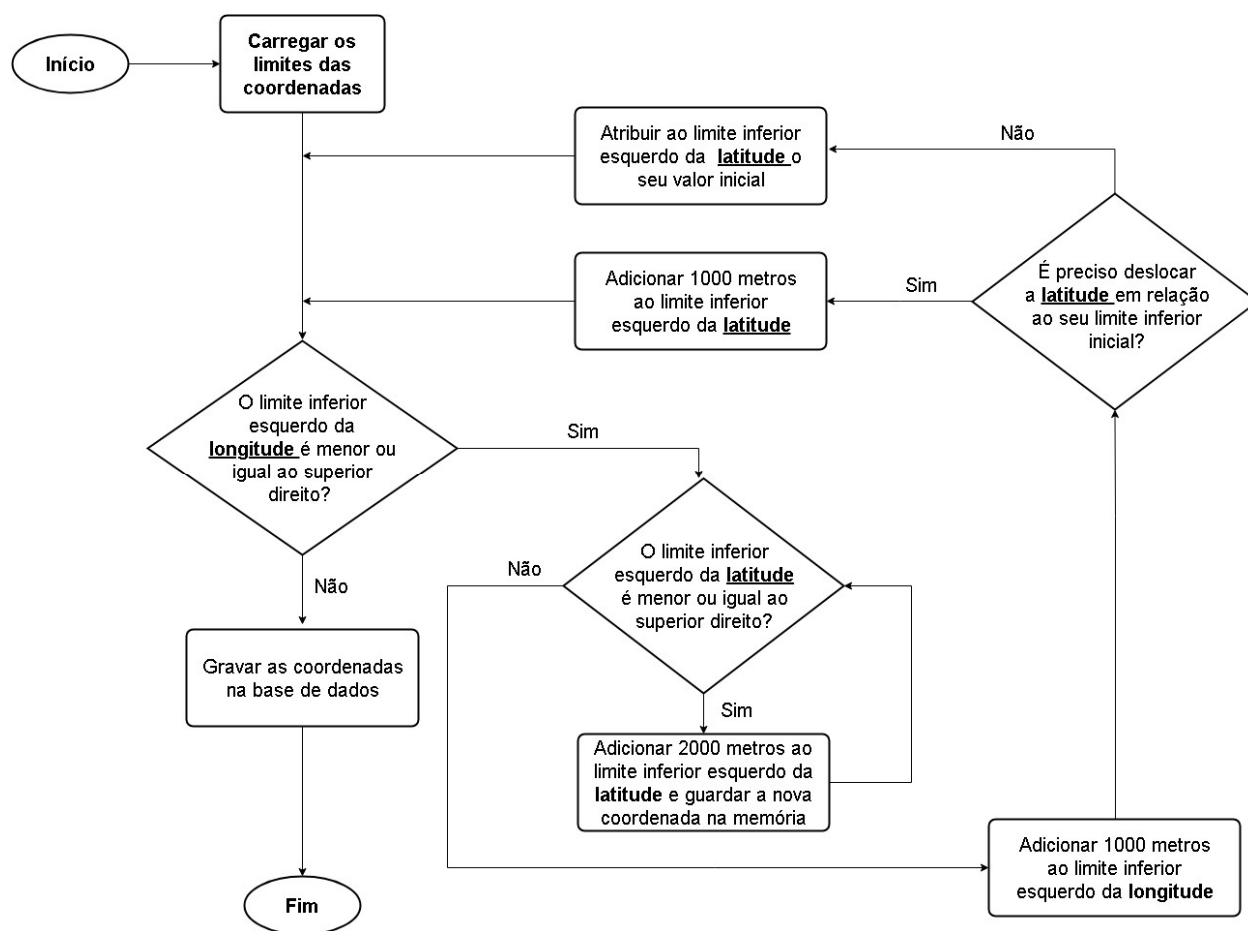


Figura 3: Fluxograma do algoritmo de criação da grelha.

A figura a seguir apresenta o formato adotado para a geração da grelha de coordenadas, onde se pode notar que os pontos em vermelho estão deslocados em relação aos pretos, ou seja, não seguem a mesma linha, de maneira a permitir uma melhor cobertura da área, considerando o tamanho do raio dentro do qual serão coletados os POIs.

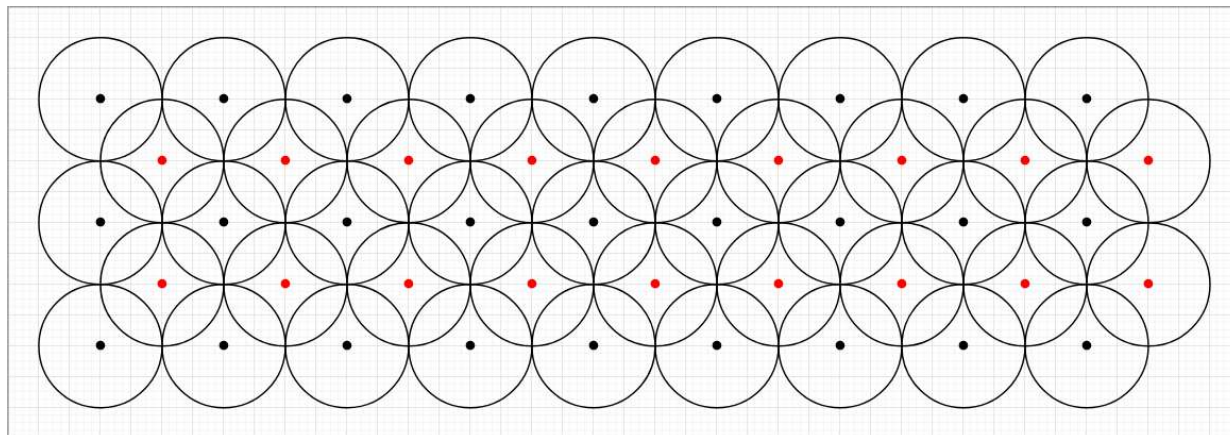


Figura 4: Disposição adotada para a definição dos pontos da grelha.

A grelha inicial criada é basicamente um quadrado contendo diversas coordenadas com espaçamento padrão entre elas. Ao projetar a grelha em um mapa, nota-se que este quadrado está posicionado sobre o território pretendido, conforme mostrado na Figura 5, onde se pode ver a grelha inicial contendo as coordenadas geradas pela aplicação. Nota-se do lado esquerdo uma visão geral, com pouco *zoom*, enquanto que do lado direito observa-se uma ampliação em uma determinada zona do mapa. Como se pode imaginar, a grelha possui uma grande quantidade de pontos que não são necessários. Neste caso, devem ser considerados somente os pontos que se encontram dentro dos limites do território. Para se chegar a este resultado, tendo em vista que a grelha foi armazenada em uma tabela da base de dados, optou-se pela realização de uma consulta, recorrendo aos recursos do PostGIS. Nesta consulta, foi utilizada também uma tabela contendo um polígono que representa os limites do território português. Esta tabela foi obtida através da importação de um ficheiro do tipo *shapefile*, obtido no site da Direção Geral de Território<sup>9</sup>.

<sup>9</sup> [http://www.dgterritorio.pt/cartografia\\_e\\_geodesia/cartografia/carta\\_administrativa\\_oficial\\_de\\_portugal\\_caop/](http://www.dgterritorio.pt/cartografia_e_geodesia/cartografia/carta_administrativa_oficial_de_portugal_caop/)  
[último acesso: 11/03/2020]

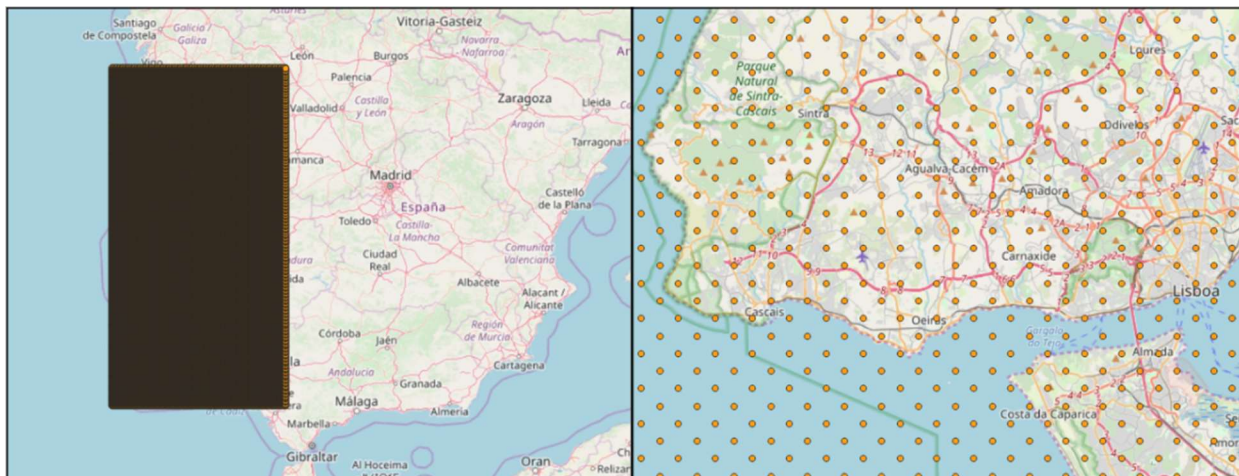


Figura 5.:Grelha inicial de coordenadas.

A seguir, na Figura 6, é mostrada a grelha final contendo as coordenadas a serem utilizadas como referências para a busca dos pontos de interesse. Observa-se do lado esquerdo o panorama completo e do lado direito o foco em uma determinada zona. Após a filtragem dos pontos por meio de uma consulta com recurso às funções espaciais, realizada diretamente à base de dados, restaram na grelha somente as coordenadas localizadas dentro dos limites do país. Na imagem, nota-se ainda que os pontos que restaram após o processo de filtragem permanecem com o espaçamento original. Neste estudo não foram consideradas as ilhas, portanto os dados em questão são referentes a Portugal Continental. O código SQL executado gerou uma nova tabela na base de dados e o *script* utilizado pode ser visualizado no apêndice A.1.

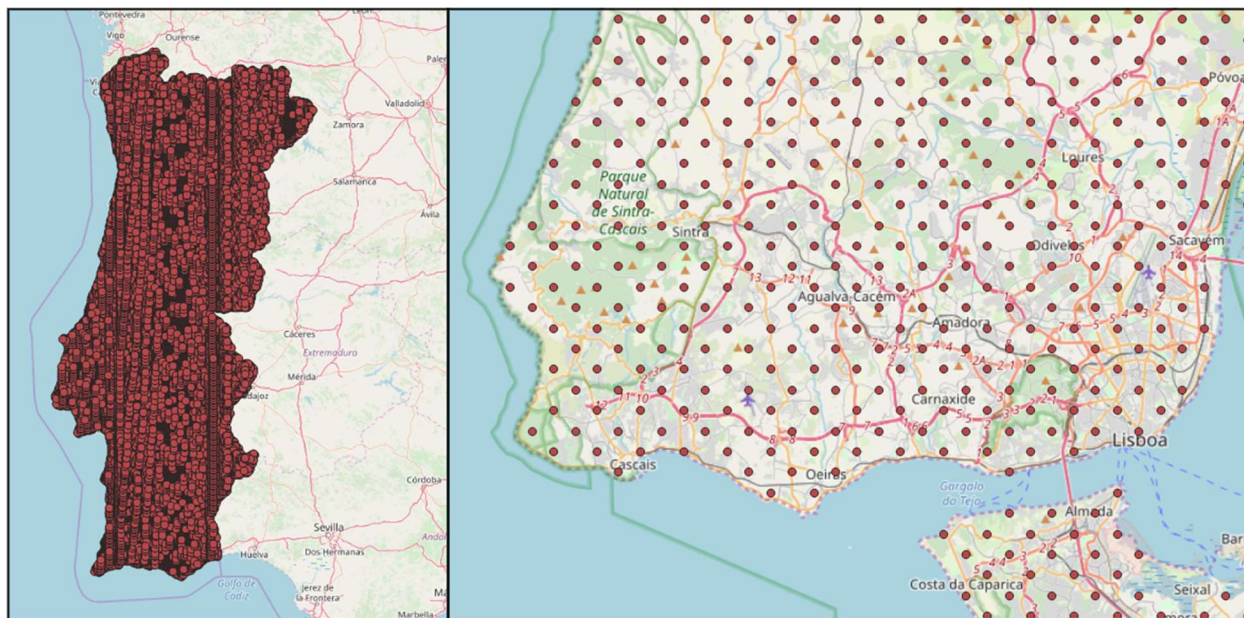


Figura 6: Grelha final de coordenadas.

#### 4.1.2.1.1.2 Coleta dos pontos de interesse

Uma vez tendo sido construída a grelha contendo todas as coordenadas devidamente definidas, a aplicação pôde então ser utilizada para a coleta dos pontos de interesse. O software quando executado, realiza uma busca na base de dados para obter uma coordenada que ainda não tenha sido processada. Essa coordenada é passada como parâmetro para a chamada do serviço. Para cada pedido efetuado, a API retorna no máximo 100 resultados. Caso existam mais de 100 pontos de interesse que cumpram os critérios da busca, o serviço faz a paginação dos dados retornados. Neste cenário, a API retorna também uma URL para a próxima requisição. Sempre que esta URL é retornada, a aplicação utiliza-a para o próximo pedido, até que não haja mais resultados para aqueles parâmetros. Este ciclo se repete até que seja atingido o limite de requisições permitido pela API ou até que seja atingido o número de requisições definido como parâmetro ao iniciar a execução da aplicação.

A aplicação realiza o controlo de todas as coordenadas já processadas e também daquelas que estão em processamento. Isso permite que várias instâncias do software sejam executadas em paralelo, permitindo acelerar o processo de coleta dos dados. Este controlo é feito por meio de atributos na base de dados. Durante a execução do algoritmo, todos os pontos de interesse obtidos são gravados numa tabela específica, que posteriormente é utilizada como base para o tratamento dos dados. Conforme mencionado anteriormente, muitos dos POIs coletados são repetidos, tendo em vista a configuração da grelha.

#### 4.1.2.2 Tratamento

A extração de conhecimento a partir dos dados depende diretamente da qualidade do *dataset* utilizado. O processo seguido para limpar e organizar os dados depende do estado em que estes se encontram. No geral, existem casos onde é necessária a aplicação de um processo completo de ETL (*Extract, Transform, Load*). O ETL inclui a obtenção de dados de diferentes fontes, uma etapa de transformação e uma última de carregamento [108]. Em geral, quando se trata da análise de dados, a etapa da extração não é necessária em casos onde o *dataset* já está formatado, exceto nas situações em que se pretende reunir informações armazenadas em mais de uma base de dados, ou mesmo quando se pretende criar cópias da base de dados original, de maneira que a aplicação do processo de transformação não afete o *dataset* inicial. No caso deste estudo, pode-se considerar que o processo de tratamento consistiu essencialmente numa filtragem para evitar a repetição dos registos e na transformação dos dados. Além disso, foi feita também uma organização das categorias dos POIs.

##### 4.1.2.2.1 Limpeza dos dados

Tendo em vista o funcionamento dos métodos de busca dos dados na API, a grelha de coordenadas gerada na etapa anterior à coleta dos dados incluía diversos pontos que, quando processados pela aplicação de coleta, geraram muitos registos repetidos na base de dados, devido à sobreposição dos raios. Desta forma, após a coleta, a tabela de pontos de interesse inicial foi filtrada, dando

origem a uma segunda tabela onde os pontos de interesse já não estavam duplicados. Para a filtragem, utilizou-se o campo “link”, tendo em vista que este é um atributo único para cada ponto de interesse. O código SQL utilizado para a filtragem dos pontos de interesse pode ser visualizado no apêndice A.2.

#### 4.1.2.2.2 Organização das categorias

Conforme abordado na parte introdutória desta secção, o conjunto de dados no qual se baseou este projeto é bastante completo, incluindo diversos atributos. Um deles, nomeadamente o atributo “category\_list”, originalmente estava no formato JSON (*JavaScript Object Notation*), onde se encontravam as categorias as quais pertence cada ponto de interesse. Como se pode notar com base no conteúdo apresentado mais à frente, no capítulo 5, as categorias dos POIs são de especial importância para a realização deste estudo. Considerando que cada POI pode pertencer a uma ou mais categorias, optou-se por buscar uma forma que pudesse facilitar as consultas realizadas à base de dados envolvendo este campo.

Embora o PostgreSQL disponibilize várias funcionalidades para se trabalhar com campos do tipo JSON, realizar buscas numa tabela com grande volume de dados, aplicando filtros com base em critérios baseados em atributos desse tipo, torna-se bastante ineficiente em termos de desempenho. Neste sentido e tendo em vista que o *dataset* encontra-se numa base de dados relacional, foi criada uma nova tabela para armazenar as diversas categorias existentes. Considerando que um ponto de interesse pode pertencer a mais de uma categoria, foi também criada uma tabela de ligação entre os POIs e as categorias. A consulta utilizada para extrair as categorias do campo original e guardá-las na nova tabela pode ser vista no apêndice A.3.

##### 4.1.2.2.2.1 Organização das categorias de forma hierárquica

Embora ultrapassado o problema da extração das categorias, houve ainda a necessidade de organizá-las hierarquicamente. Inicialmente, assumiu-se que isso poderia ser feito por meio de uma busca via API, tendo em vista que as categorias retornadas trazem o próprio identificador e também o identificador da categoria à qual pertence, caso seja uma subcategoria. A primeira busca realizada foi pelos chamados “placetopics”<sup>10</sup>, que são uma família específica de categorias do Facebook. Os dados obtidos foram importados para a base de dados, através de instruções SQL apresentadas no apêndice A.4. Entretanto, verificou-se, que entre as categorias importadas, 308 eram subcategorias de outras que não existiam no conjunto obtido. Além disso, das 1225 categorias extraídas dos pontos de interesse, 570 não existiam nesta lista. Tendo em vista tais inconsistências, foi necessário buscar uma alternativa para solucionar o problema da hierarquia.

Com base numa pesquisa realizada na documentação do Facebook, foi encontrada uma página onde, de acordo com o site oficial [109], estariam listadas todas as categorias existentes.

---

<sup>10</sup> <https://blog.reachlocal.com/facebook-place-topics-how-to-add-your-place-subcategory>  
[último acesso: 11/03/2020]

Como não havia forma de descarregar essa informação num formato que pudesse ser importado diretamente para a base de dados, realizou-se este trabalho de forma parcialmente manual, com a ajuda do Microsoft Word, no qual foi possível colar os dados copiados do site, criando automaticamente uma lista hierárquica numerada. Formatando corretamente e copiando as informações para o Microsoft Excel, conseguiu-se então exportar um ficheiro CSV que foi posteriormente importado na base de dados. As instruções SQL utilizadas para a importação do ficheiro podem ser vistas no apêndice A.5.

Após a importação das categorias obtidas no site e a criação de uma nova tabela na base de dados para armazená-las, primeiramente tentou-se fazer o mapeamento das mesmas para as categorias extraídas dos pontos de interesse, através do nome. Contudo, verificou-se que 57 das categorias dos pontos de interesse não figuravam na lista obtida no site. Ao chegar-se a esta conclusão, ainda foi feita uma tentativa com a seguinte estratégia: verificar se as categorias que não existiam na lista descarregada no site, existiam na lista obtida via API na tentativa anterior. Por meio desta análise, foi verificado que 47 delas não existiam em nenhuma das 2 listas. Face a um problema deste género, concluiu-se de imediato que nem a lista de categorias do site, nem a lista dos “placetopics” da API serviriam para organizar a hierarquia das categorias originais.

Os problemas encontrados revelaram-se um grande desafio, que veio a ser superado após uma extensiva pesquisa. Com base nesta pesquisa, evidenciou-se a existência de uma nova possibilidade de buscas na API, por meio de um serviço denominado “fb\_page\_categories”. Realizada a busca, os dados obtidos foram guardados num ficheiro, no formato JSON, que foi posteriormente importado para a base de dados, por meio das instruções mostradas no apêndice A.6. Após a importação, verificou-se que a lista obtida é mais completa em relação as 2 anteriores, contendo 1556 categorias, enquanto a que fora obtida no site possui 1512 registos e a outra 710. Ao consultar os dados, verificou-se que mesmo assim, 33 das categorias extraídas dos pontos de interesse não figuravam na lista recém-importada.

Considerando a inexistência de correspondência entre os 33 registos nas duas tabelas de categorias, realizou-se ainda uma vasta pesquisa na Internet à procura de uma solução. Apesar do esforço, a solução não foi encontrada, pelo que decidiu-se realizar um mapeamento manual das categorias extraídas dos pontos de interesse que não existem nos dados importados para esta última tabela. Esta tarefa foi realizada com base numa análise cuidadosa dos pontos de interesse existentes em cada categoria envolvida. Como se pode ver no apêndice B.2, foi elaborada uma folha de cálculo onde foi registada a correspondência de cada um dos itens em questão. Com base no mapa estabelecido, foi criado um ficheiro CSV. Uma vez importado para a base de dados, este ficheiro gerou uma tabela de ligação, que foi posteriormente carregada com as demais correspondências, que já existiam entre as categorias originais dos pontos de interesse e a lista obtida no serviço “fb\_page\_categories”. O *script* utilizado para esta finalidade pode ser visto no apêndice A.7.

Considerando que a lista final de categorias dos POIs é relativamente extensa, a seguir, na Tabela 3, são mostradas apenas aquelas que pertencem aos níveis 1 e 2. A lista completa foi disponibilizada na Internet, numa folha de cálculo disponível em <http://tiny.cc/395voz>.

Tabela 3: Categorias de níveis 1 e 2 dos POIs.

Nível 1	Nível 2	Nível 1	Nível 2
Businesses	Advertising/Marketing	Other	Brand
	Agriculture		Cause
	Arts & Entertainment		Color
	Automotive, Aircraft & Boat		Community
	Beauty, Cosmetic & Personal Care		Course
	Commercial & Industrial		Cuisine
	Education		Diseases
	Finance		Editorial/Opinion
	Food & Beverage		Election
	Hotel & Lodging		Event
	Legal		Exchange Program
	Local Service		Fan Page
	Media/News Company		Harmonized Page
	Medical & Health		Just For Fun
	Non-Governmental Organization (NGO)		Language
	Nonprofit Organization		Mood
	Public & Government Service		Nationality
	Real Estate		Not a Business
	Science, Technology & Engineering		Political Ideology
	Shopping & Retail		Profile
Sports & Recreation	Satire/Parody		
Travel & Transportation	Sports Season		
Community Organization	Armed Forces	Surgeries	
	Charity Organization	Ticket Sales	
	Community Service	Topic	
	Country Club / Clubhouse	TypeAhead	
	Environmental Conservation Organization	University (NCES)	
	Government Organization	University Status	
	Labor Union	Work Position	
	Political Organization	Work Project	
	Political Party	Work Status	
	Private Members Club	Actor	
	Religious Organization	Artist	
	Social Club	Athlete	

	Sorority & Fraternity	Author
	Sports Club	Band
	Youth Organization	Blogger
Feed for Workplace	-	Chef
Interest	Literary Arts	Comedian
	Performance Art	Dancer
	Performing Arts	Designer
	Science	Digital Creator
	Sports	Editor
Media	Visual Arts	Entrepreneur
	Art	Fashion Model
	Books & Magazines	Film Director
	Concert Tour	Fitness Model
	Media Restoration Service	Gamer
	Music	Government Official
	Show	Journalist
	TV & Movies	Motivational Speaker
	Theatrical Play	Musician
	Theatrical Productions	Musician/Band
Non-Business Places	Automated Teller Machine (ATM)	News Personality
	Campus Building	Orchestra
	City Infrastructure	Producer
	Government Building	Scientist
	Landmark & Historical Place	Spiritual Leader
	Locality	Sports Promoter
	Meeting Room	Talent Agent
	Outdoor Recreation	Video Creator
	Public Toilet	Writer
	Religious Place of Worship	
Residence		

#### 4.1.2.3 Exploração

De maneira a permitir uma melhor interpretação dos dados obtidos, foi realizada uma análise exploratória preliminar. Esta análise foi inicialmente feita via consultas à base de dados, permitindo a obtenção das características principais do *dataset*, apresentadas na parte introdutória desta secção. Posteriormente foi feita também uma exploração a nível de visualização, por meio da utilização do QGIS. Esta análise exploratória constitui uma etapa importante do trabalho, tendo em vista que permitiu, entre outras questões, observar a distribuição geográfica dos pontos de interesse em cada categoria. Conforme já referido, foram coletados 171177 pontos de interesse. É

importante voltar a reforçar que cada ponto de interesse pode pertencer a mais de uma categoria, e é relativamente grande o número de registos onde se observa esta peculiaridade.

Em termos exploratórios, uma das primeiras observações que se faz é a seguinte: a começar pelas categorias de nível superior, verifica-se que dentre as 8 existentes a nível mundial, as categorias “Non-Business Places”, “Other” e “Feed for Workplace” não possuem dados em Portugal ligados diretamente a elas, embora as duas primeiras possuam subcategorias às quais existem pontos de interesse vinculados, em território nacional. Considerando as subcategorias incluídas na categoria “Other”, apesar de estas possuírem um total de 10845 pontos de interesse, numa análise cuidadosa desses dados, observa-se um número elevado de registos que à partida poderiam ser considerados irrelevantes para o estudo, como por exemplo os POIs das subcategorias “Election”, “Brend”, “Cause”, “Fan Page”, “Nationality”, “Political Ideology”, “Topic”, “Event”, “Work Position”, entre muitos outros. Assim, tendo em vista que os pontos de interesse nesta categoria representam apenas 6.3% dos registos do *dataset*, aponta-se que, caso seja necessário, a mesma pode ser desconsiderada no estudo principal.

Relativamente à categoria “Non Business Places”, esta possui um total de 25951 pontos de interesse vinculados às suas subcategorias, apesar de nenhum deles estar ligado diretamente a ela. Representando mais de 15% dos exemplos do *dataset*, entre tais subcategorias encontram-se por exemplo, tópicos como “Religious Place of Worship”, “Residence” e “Outdoor Recreation”, que incluem pontos de interesse que, de acordo com a literatura [9], parecem ser relevantes para o âmbito deste estudo. A distribuição dos pontos de interesse destas subcategorias pode ser vista, a seguir, na Figura 7.

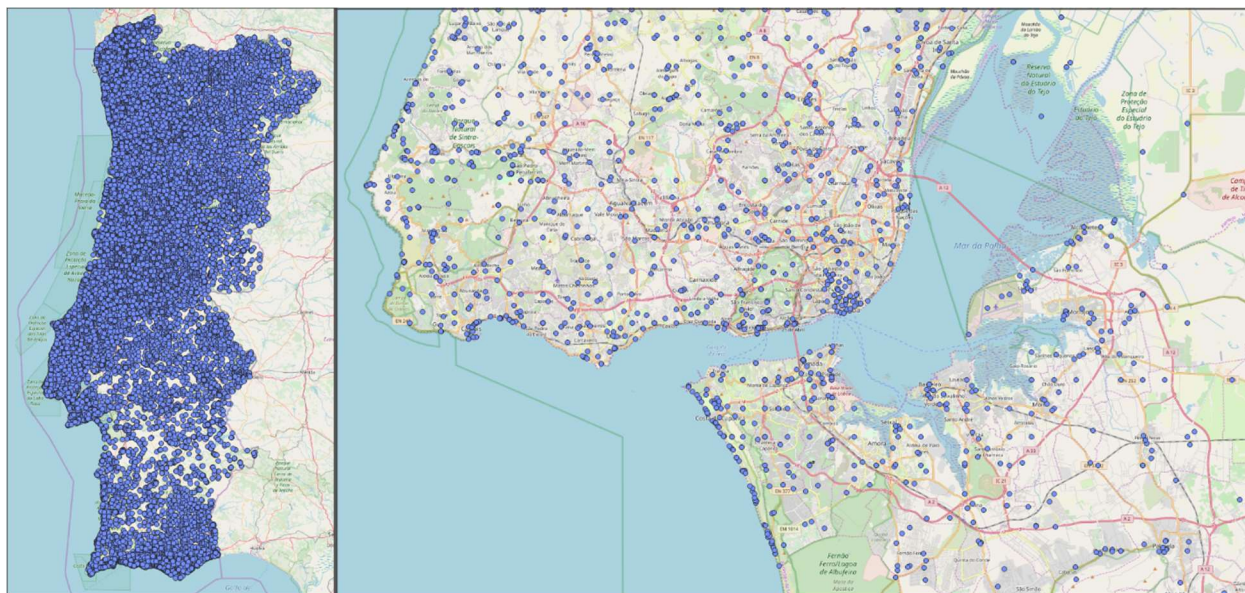


Figura 7: Distribuição dos POIs das subcategorias da categoria “Non Business Places”.

Ainda a nível de topo, encontra-se a categoria “Media”, onde mesmo considerando as suas subcategorias, representa um total de apenas 0.56% dos pontos de interesse coletados, ou seja, 962 registos. Ao analisar tais subcategorias, observa-se tópicos como “Art”, “Show”, “Music”, “Concert Tour”, entre outros, indicando que a mesma possui pouca ou nenhuma relevância para o estudo principal. Sendo assim, aponta-se que, caso necessário, esta categoria, juntamente com as suas subcategorias, pode ser desconsiderada nas análises. A exemplo desta, tem-se ainda a categoria “Public Figure”, que possui 1872 registos e inclui subcategorias como “Band”, “Musician”, “Orchestra”, “Sports Promoter”, entre algumas outras. Uma análise mais pormenorizada dos POIs incluídos nestas categorias, revelou registos que à partida também não parecem relevantes para o projeto. Da mesma forma a nível superior, encontram-se ainda as categorias “Interest” e “Community Organization”, que juntamente com suas subcategorias, somam respetivamente 0,39 e 2,83% dos registos do *dataset*. Tal como no caso anterior, considera-se, com base nos POIs que incluem, que estas parecem não apresentar relevância, tanto para a caracterização dos espaços geográficos, quanto para a análise das correlações.

Diferente dos casos abordados nos parágrafos anteriores, a categoria “Business” inclui a grande maioria dos dados coletados. Ao realizar uma busca pelos pontos de interesse existentes nesta categoria e nas suas subcategorias, é obtido um total de 173435 registos, indicando que muitos dos pontos de interesse neste grupo estão vinculados a mais de uma categoria. Com base nos exemplos analisados, verifica-se que esta categoria parece demonstrar grande relevância para o que se pretende no estudo. As subcategorias “Advertising/Marketing”, “Comercial & Industrial”, “Finance”, “Food & Beverage”, “Hotel & Lodging” e “Local Service”, são algumas das que estão vinculadas ao segundo nível desta categoria, que no total são 22. A seguir, na Figura 8, pode-se observar a distribuição geográfica dos pontos de interesse referidos, onde também é possível notar uma evidente densidade nos registos.



Figura 8: Pontos de interesse na categoria “Business” e em suas subcategorias.

Conforme se pode ver na Figura 9, ao analisar separadamente as subcategorias incluídas na categoria “Business”, observa-se que a maior parte dos pontos de interesse pertencem ao segmento “Food & Beverage”, onde se encontram por exemplo bares e restaurantes, totalizando 40594 registos. Em segundo lugar, encontra-se a categoria “Local Service”, com 34002 registos. Nesta categoria se encontram subcategorias como “Computer Repair Service”, “Moving & Storage Service”, “Party & Entertainment Service”, “Shoe Repair Shop” e “Weight Loss Center”. Juntas, estas duas categorias principais representam um total de pouco mais de 43% dos pontos de interesse do segmento “Business”. Na Figura 10, é mostrada a distribuição geográfica dos pontos de interesse da categoria “Food & Beverage”, na parte superior, e da categoria “Local Service”, na parte inferior. Embora sejam bastante representativas, as duas referidas categorias de topo, apresentam pouca densidade de pontos na região do Alentejo, sendo que as regiões de maior densidade se concentram ao longo da costa, na parte norte e no extremo sul do país.

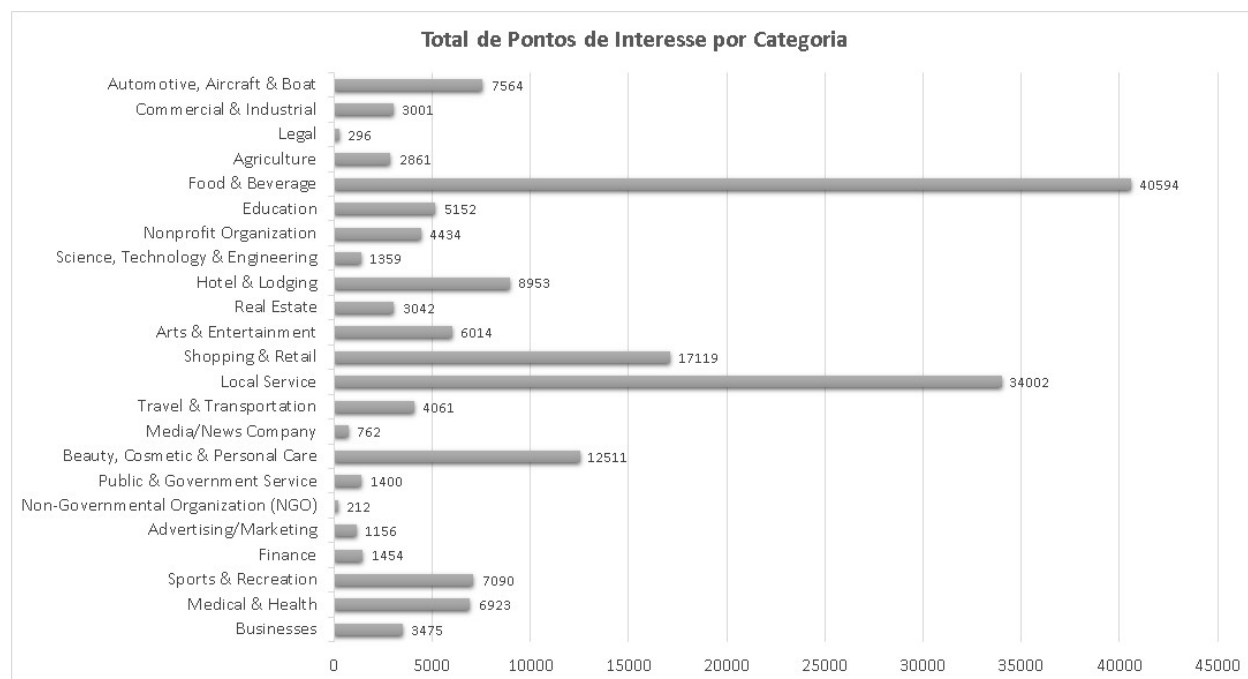


Figura 9: Total de registos na categoria “Business” e nas suas subcategorias.

Apesar de representarem dois segmentos diferentes, as categorias “Food & Beverage” e “Local Service” contêm pontos de interesse bastante próximos uns dos outros, em termos de localização geográfica. Em cenários como este, a aglomeração de pontos de categorias diferentes numa mesma localização geográfica ou em regiões próximas, pode indicar uma forte correlação entre as categorias. Neste caso em específico, em termos numéricos, verifica-se que a categoria “Food & Beverage” contém 6592 registos a mais que a categoria “Local Service”, ou seja, é mais de 19% superior. A correlação entre estas duas categorias será estudada no capítulo 5, onde o são apresentados, no tópico 5.2, os resultados das análises das correlações entre os mapas de densidade gerados a partir dos pontos de interesse existentes em cada segmento.

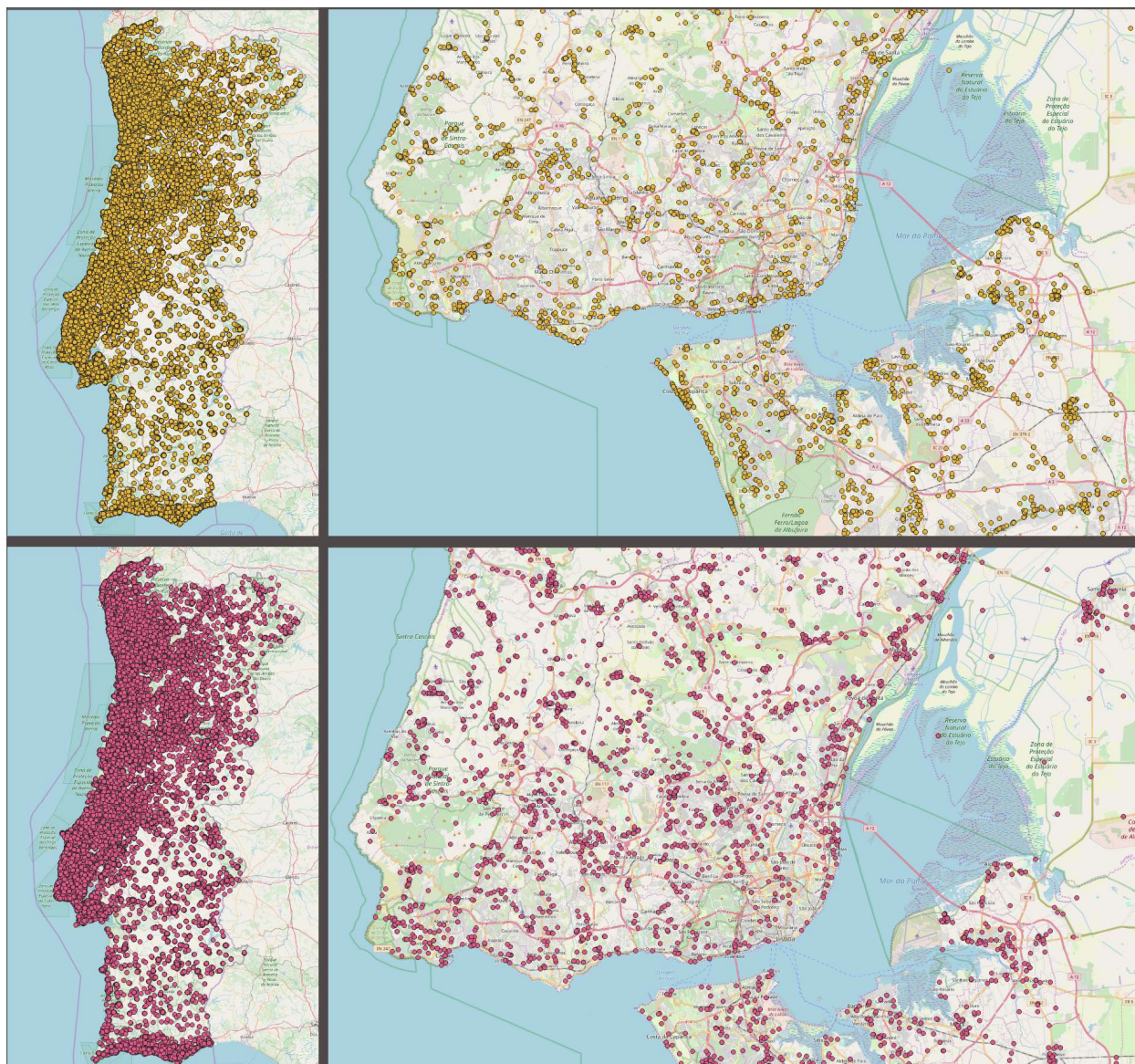


Figura 10: Distribuição dos POIs das categorias “Food & Beverage” e “Local Service”.

## 4.2 Dados de uso e ocupação dos espaços (*Land Use/Land Cover*)

Nos últimos anos nota-se uma crescente popularização do uso dos sensores remotos, o que resulta numa grande quantidade de dados georreferenciados. Alguns dos dados produzidos por esses sensores, são essencialmente imagens. Com a evolução da tecnologia, as imagens de sensores remotos passaram a apresentar cada vez uma maior resolução e com base na sua interpretação, são produzidos diferentes *datasets*, sobretudo mapas de uso e ocupação das regiões.

Atualmente existem diferentes iniciativas que visam a criação de mapas dinâmicos do uso e da ocupação dos espaços. No que diz respeito ao território português, dentre as iniciativas

existentes destacam-se o CORINE *Land Cover* e a Carta de Uso e Ocupação do Solo – COS, ambas sob responsabilidade da Direção Geral de Território - DGT<sup>11</sup>. Nesta secção são abordados os conjuntos de dados produzidos por ambos os projetos.

#### 4.2.1 CORINE

O CORINE *Land Cover* [110] é um conjunto de dados que representa a cobertura e o uso dos espaços na região da Europa, criado por iniciativa da Agência Europeia do Ambiente (EEA) em parceria com os Estados-Membros. Trata-se, portanto, de uma cartografia temática que é produzida principalmente com base em imagens de satélite, sendo que soluções semiautomáticas são aplicadas em alguns países, recorrendo-se ao uso de informações locais, processamento de imagens e interpretação e generalização, com recurso a sistemas SIG. O CORINE está atualmente disponível tanto em formato vetorial quanto matricial. Em termos de resolução, possui uma Unidade Mínima Cartográfica – UMC de 25 hectares. De acordo com o site oficial, a iniciativa de criação do CORINE data de 1985, tendo sido disponibilizada a primeira versão em 1990. Nos anos seguintes, nomeadamente em 2000, 2006, 2012, e 2018, foram lançadas novas atualizações.

Conforme apresentado na Tabela 4, segundo a nomenclatura oficial do CORINE [111] os dados do conjunto estão divididos essencialmente em 3 níveis hierárquicos, traduzindo-se num total de 44 classes ao nível mais pormenorizado. É importante clarificar que em algumas partes do estudo, com o intuito de facilitar o entendimento, estas mesmas classes são referidas pelo seu nome traduzido para o português, conforme reforçado no capítulo 5. A descrição detalhada de cada classe do conjunto também pode ser consultada on-line<sup>12</sup>.

Tabela 4: Classes do CORINE *land cover*.

Nível 1	Nível 2	Nível 3
<b>1. Artificial surfaces</b>	1.1. Urban fabric	111. Continuous urban fabric
		112. Discontinuous urban fabric
	1.2. Industrial, commercial and transport units	121. Industrial or commercial units
		122. Road and rail networks and associated land
		123. Port areas
		124. Airports
	1.3. Mine, dump and construction sites	131. Mineral extraction sites
		132. Dump sites
		133. Construction sites
	1.4. Artificial, non-agricultural vegetated areas	141. Green urban areas
		142. Sport and leisure facilities

<sup>11</sup> [http://www.dgterritorio.pt/a\\_dgt/](http://www.dgterritorio.pt/a_dgt/) [último acesso: 11/03/2020]

<sup>12</sup> [https://land.copernicus.eu/user-corner/technical-library/corine-land-cover-nomenclature-guidelines/docs/pdf/CLC2018\\_Nomenclature\\_illustrated\\_guide\\_20190510.pdf](https://land.copernicus.eu/user-corner/technical-library/corine-land-cover-nomenclature-guidelines/docs/pdf/CLC2018_Nomenclature_illustrated_guide_20190510.pdf) [último acesso: 11/03/2020]

<b>2. Agricultural areas</b>	2.1. Arable land	211. Non-irrigated arable land
		212. Permanently irrigated land
		213. Rice fields
	2.2. Permanent crops	221. Vineyards
		222. Fruit trees and berry plantations
		223. Olive groves
	2.3. Pastures	231. Pastures
		241. Annual crops associated with permanent crops
		242. Complex cultivation patterns
		243. Land principally occupied by agriculture, with significant areas of natural vegetation
	2.4. Heterogeneous agricultural areas	244. Agro-forestry areas
		3.1 Forests
312. Coniferous forest		
313. Mixed forest		
3.2. Scrub and/or herbaceous vegetation associations		321. Natural grasslands
	322. Moors and heathland	
	323. Sclerophyllous vegetation	
	324. Transitional woodland-shrub	
3.3. Open spaces with little or no vegetation	331. Beaches, dunes, sands	
	332. Bare rocks	
	333. Sparsely vegetated areas	
	334. Burnt areas	
	335. Glaciers and perpetual snow	
<b>4. Wetlands</b>	4.1. Inland wetlands	411. Inland marshes
		412. Peat bogs
	4.2. Maritime wetlands	421. Salt marshes
		422. Salines
		423. Intertidal flats
<b>5. Water bodies</b>	5.1. Inland waters	511. Water courses
		512. Water bodies
	5.2. Marine waters	521. Coastal lagoons
		522. Estuaries
		523. Sea and ocean

Algumas versões do CORINE *land cover* estão disponíveis para a região de Portugal continental e das ilhas, sob a responsabilidade da Direção Regional do Ordenamento do Território e Ambiente. Estas versões podem ser obtidas no site do Sistema Nacional de Informação Geográfica – SNIG<sup>13</sup> e em vários outros, como o iGeo<sup>14</sup>, por exemplo. Contudo, a versão 2018

<sup>13</sup> <https://snig.dgterritorio.gov.pt/rndg/srv/por/catalog.search#/search?anysnig=corine> [último acesso: 11/03/2020]

<sup>14</sup> <http://www.igeo.pt/DadosAbertos/Listagem.aspx> [último acesso: 11/03/2020]

para o país ainda não está disponível nos repositórios locais. Mesmo assim, considerando que não existem restrições legais quanto ao uso do CORINE, todas as versões existentes do conjunto de dados, referentes ao território europeu, podem ser obtidas em diferentes formatos por meio do site oficial<sup>15</sup>, mediante registo prévio.

Para este estudo, foram utilizados os dados da versão 2018, do CORINE, em formato vetorial. Não estando disponível esta versão contendo somente os polígonos referentes à cobertura do território português, houve a necessidade de se recorrer à versão europeia, obtida por meio do serviço de descarregamento existente no site oficial. Desta forma, foi descarregado o *dataset* em formato *SQLite*, que foi posteriormente importado para a base de dados e para o QGIS. No caso deste trabalho, um dos pontos negativos da utilização desta versão foi especificamente o facto de representar uma zona muito abrangente, ao invés de apenas o território nacional. Por este motivo, foi necessário realizar algum tratamento dos dados, de maneira a permitir que os mesmos fossem estudados de maneira eficiente, tendo em conta os recursos computacionais disponíveis.

Para filtrar somente os dados referentes ao território português, após a sua importação para a base de dados, foram isolados somente os polígonos que representavam áreas totalmente ou parcialmente localizadas dentro das fronteiras do país. Embora houvessem diversas alternativas à filtragem por meio do PostGIS, como por exemplo a realização de uma interseção de polígonos no QGIS, optou-se por realizar esta operação na base de dados, por ser uma maneira rápida e eficiente de alcançar o objetivo pretendido. Desta forma, por meio da utilização do código SQL apresentado no apêndice A.11, foi criada uma nova tabela para guardar os dados do CORINE, referentes à cobertura do território nacional. Em termos de tratamento dos dados, para além da filtragem, foi realizada também uma reorganização das classes, de maneira a permitir a classificação. Este assunto é abordado mais adiante, no tópico 5.1.3.

#### 4.2.2 Carta de Uso e Ocupação do Solo - COS

De forma semelhante ao CORINE, a Carta de Uso e Ocupação do Solo [112] é uma cartografia temática disponível em formato vetorial, que visa caracterizar o uso e a ocupação do território em Portugal continental, tendo em vista que as regiões autónomas possuem os seus próprios programas de cartografia. O projeto é de responsabilidade da Direção Geral de Território - DGT<sup>16</sup>, sendo financiado pelo Fundo Ambiental, e já foram disponibilizadas diversas versões do *dataset*, nomeadamente nos anos de 1995, 2007, 2010, 2015 e 2018. Segundo o site oficial [113], o conjunto de dados é produzido com base na interpretação visual de imagens aéreas de alta resolução. A Carta de Uso e Ocupação do Solo possui uma UMC de 1 hectare.

O COS é constituído por polígonos que representam as diversas classes de uso e ocupação do solo, sendo que neste contexto, cada polígono representa uma área do território que é igual ou

<sup>15</sup> <https://land.copernicus.eu/pan-european/corine-land-cover> [último acesso: 11/03/2020]

<sup>16</sup> <http://www.dgterritorio.pt/> [último acesso: 11/03/2020]

superior a 1 hectare, onde predomina em mais de 75% da área delimitada uma determinada classe, independente de existirem ou não outras classes na zona representada, desde que não ultrapassem 25% do polígono. As 83 classes do *dataset* estão organizadas em 5 níveis hierárquicos, ou seja, mais 35 classes em relação à sua versão anterior e cuja nomenclatura foi elaborada por um processo de consulta pública na época da produção da primeira versão da edição disponibilizada no ano de 2007. A documentação completa das classes da Carta de Uso e Ocupação do Solo pode ser consultada on-line<sup>17</sup>. A Tabela 5 apresenta algumas delas, nomeadamente aquelas de nível 1, 2 e 3, onde se encontram todas as que foram utilizadas no estudo.

Para o trabalho realizado neste projeto, as classes mais importantes são por exemplo, a “1.1 Tecido edificado”, que inclui “1.1.1 Tecido edificado contínuo” (correspondente à classe “111. Continuous urban fabric”, do CORINE), que por sua vez representa os espaços urbanos mais densos e “1.1.2 Tecido edificado descontínuo” (correspondente à classe “112. Discontinuous urban fabric”, do CORINE), onde se enquadram as áreas artificiais menos densas. Também foi de grande importância a classe “1.2 Indústria, comércio e instalações agrícolas”, onde estão representadas as zonas industriais.

Tabela 5: Classes do *dataset* COS.

Nível 1	Nível 2	Nível 3
<b>1.Territórios artificializados</b>	1.1 Tecido edificado	1.1.1 Tecido edificado contínuo
		1.1.2 Tecido edificado descontínuo
		1.1.3 Espaços vazios em tecido edificado
	1.2 Indústria, comércio e instalações agrícolas	1.2.1 Indústria
		1.2.2 Comércio
		1.2.3 Instalações agrícolas
	1.3 Infraestruturas	1.3.1 Infraestruturas de produção de energia
		1.3.2 Infraestruturas de águas e tratamento de resíduos
	1.4 Transportes	1.4.1 Redes viárias e ferroviárias e espaços associados
		1.4.2 Áreas portuárias
		1.4.3 Aeroportos e aeródromos
	1.5 Áreas de extração de inertes, áreas de deposição de resíduos e estaleiros de construção	1.5.1 Áreas de extração de inertes
		1.5.2 Áreas de deposição de resíduos
		1.5.3 Áreas em construção
	1.6 Equipamentos	1.6.1 Equipamentos desportivos
		1.6.2 Equipamentos de lazer e parques de campismo
		1.6.3 Equipamentos culturais

<sup>17</sup> [http://mapas.dgterritorio.pt/atom-dgt/pdf-cous/COS2018/ET-COS-2018\\_v1.pdf](http://mapas.dgterritorio.pt/atom-dgt/pdf-cous/COS2018/ET-COS-2018_v1.pdf) [último acesso: 11/03/2020]

		1.6.4 Cemitérios
		1.6.5 Outros equipamentos e instalações turísticas
	1.7 Parques e jardins	1.7.1 Parques e jardins
<b>2.Agricultura</b>	2.1 Culturas temporárias	2.1.1 Culturas temporárias de sequeiro e regadio e arrozais
	2.2 Culturas permanentes	2.2.1 Vinhas
		2.2.2 Pomares
		2.2.3 Olivais
	2.3 Áreas agrícolas heterogéneas	2.3.1 Culturas temporárias e/ou pastagens melhoradas associadas a culturas permanentes
		2.3.2 Mosaicos culturais e parcelares complexos
2.3.3 Agricultura com espaços naturais e seminaturais		
2.4 Agricultura protegida e viveiros	2.4.1 Agricultura protegida e viveiros	
<b>3.Pastagens</b>	3.1 Pastagens	3.1.1 Pastagens melhoradas
		3.1.2 Pastagens espontâneas
<b>4.Superfícies agroflorestais (SAF)</b>	4.1 Superfícies agroflorestais (SAF)	4.1.1 Superfícies agroflorestais (SAF)
<b>5.Florestas</b>	5.1 Florestas	5.1.1 Florestas de folhosas
		5.1.2 Florestas de resinosas
<b>6.Matos</b>	6.1 Matos	6.1.1 Matos
<b>7. Espaços descobertos ou com pouca vegetação</b>	7.1 Espaços descobertos ou com pouca vegetação	7.1.1 Praias, dunas e areais
		7.1.2 Rocha nua
		7.1.3 Vegetação esparsa
<b>8. Zonas húmidas</b>	8.1 Zonas húmidas	8.1.1 Zonas húmidas interiores
		8.1.2 Zonas húmidas litorais
<b>9.Massas de água superficiais</b>	9.1 Massas de água interiores	9.1.1 Cursos de água
		9.1.2 Planos de água
	9.2 Aquicultura	9.2.1 Aquicultura
		9.3.1 Salinas
	9.3 Massas de água de transição e costeiras	9.3.2 Lagoas costeiras
		9.3.3 Desembocaduras fluviais
9.3.4 Oceano		

Atualmente não existem restrições legais aplicadas ao uso do COS, embora seja recomendado pela DGT que qualquer publicação baseada no conteúdo do *dataset* inclua um texto de atribuição dos créditos à instituição. O *dataset* está disponível em NUTS 1 e em NUTS 3, o que significa que podem também ser descarregados os dados referentes às várias partes específicas do país. No caso particular deste estudo, optou-se pela utilização da versão com cobertura de todo o território nacional, com referencia ao ano de 2018, que foi obtida por meio do SNIG, embora os

links para descarregamento dos dados também estejam presentes no site da Direção Geral de Território<sup>18</sup>.

### 4.3 Tabelas iniciais da base de dados

Tendo sido coletados os pontos de interesse e obtidos os dados do CORINE e do COS, foi então criada uma base de dados espacial para armazená-los. Esta base de dados inicial contava apenas com um reduzido conjunto de tabelas, que deram posteriormente origem a um vasto conjunto de outras entidades, incluindo algumas funções e visualizações, desenvolvidas para colmatar as várias necessidades que surgiram durante a realização do estudo. Este modelo inicial, isto é, apenas com as tabelas relacionadas aos pontos de interesse e as suas respetivas categorias, contendo também as tabelas onde foram armazenados os dados dos outros dois conjuntos, pode ser consultado no apêndice C.1. É ainda importante reforçar que algumas operações realizadas no QGIS também resultaram em dados que foram importados, em momentos posteriores, para esta mesma base de dados. A descrição resumida de cada tabela do modelo inicial é apresentada a seguir.

- Tabela “coordinates”: guarda as coordenadas criadas para a coleta dos dados;
- Tabela “log”: guarda a informação da execução da aplicação de coleta dos dados;
- Tabela “dgt\_portugal”: guarda os limites geográficos do território;
- Tabela “dgt\_counties\_portugal”: guarda os limites geográficos de cada concelho;
- Tabela “dgt\_parishes”: guarda os limites geográficos de cada freguesia;
- Tabela “clc2018\_eu\_v20”: guarda os dados originais do CORINE referentes a todo o território europeu;
- Tabela “clc\_legend”: guarda os códigos e nomes das classes do CORINE;
- Tabela “clc2018\_pt\_lisbon\_grid\_final\_v2\_full”: guarda a *grid* que foi criada para uso nas análises de uso e ocupação dos espaços geográficos;
- Tabela “points\_of\_interest”: guarda os pontos de interesse coletados;
- Tabela “poi\_categories”: guarda as categorias iniciais, extraídas dos POIs coletados;
- Tabela “poi\_vs\_categories”: tabela de ligação entre os pontos de interesse e as categorias a que pertencem;
- Tabela “fbpages\_categ\_api\_v1”: guarda a lista completa das categorias obtidas na API do Facebook, com a indicação da hierarquia;
- Tabela “poi\_categories\_to\_fbpage\_categ\_api\_v1”: faz a ligação entre as categorias dos pontos de interesse coletados e a lista completa das categorias, obtida na API do Facebook

<sup>18</sup> [http://www.dgterritorio.pt/dados\\_abertos/cos/](http://www.dgterritorio.pt/dados_abertos/cos/) [último acesso: 11/03/2020]



## 5 Extração de Conhecimento

Com o objetivo de extrair informação útil e relevante dos conjuntos de dados disponíveis, foram conduzidos dois estudos diferentes. O primeiro deles está relacionado com a classificação do uso e ocupação dos espaços e envolveu diferentes cenários, baseados em duas metodologias distintas. O segundo visa analisar as correlações existentes entre as várias categorias dos pontos de interesse, bem como a sua ligação com as diferentes classes de uso e cobertura das regiões geográficas. São, portanto, abordados neste capítulo, os dois estudos estabelecidos nos domínios supracitados. São também apresentados e discutidos segundo o ponto de vista do autor, os resultados obtidos em cada uma das etapas do trabalho.

### 5.1 Classificação do uso dos espaços

Neste estudo foram realizadas diferentes experiências para classificação do uso dos espaços, cujo conceito foi abordado na fundamentação teórica, no tópico 2.1. Algumas dessas experiências basearam-se no uso do *dataset* COS, descrito na secção 4.2.2, enquanto outras acentaram-se sobre o uso do CORINE, abordado no tópico 4.2.1. Contudo, as diversas métricas para a realização de todas essas experiências, foram extraídas dos pontos de interesse, envolvendo, em todos os casos, as categorias nas quais os mesmos se agrupam e em alguns cenários, os seus respetivos níveis hierárquicos. Este tópico descreve as principais experiências realizadas para a caracterização das regiões geográficas, envolvendo espaços urbanos, semi-naturais e em alguns casos, áreas naturais.

É importante reforçar que para as experiências realizadas, foram utilizadas duas metodologias distintas. A primeira delas envolveu a utilização dos centróides dos vários polígonos de cada *dataset*. Considerando que cada polígono representa uma classe de utilização do espaço, o centróide passa a representar o polígono e consequentemente a sua respetiva classe. Esta metodologia, descrita na secção 5.1.2, traduziu-se em resultados inferiores aos obtidos na segunda, que foi baseada na criação de uma *grid* formada por células de um tamanho predefinido. Esta grelha é basicamente uma camada posicionada em um sistema GIS, sobre outra, que representa o *dataset* escolhido. Nesta metodologia, cada uma das células da grelha assume o papel da classe relacionada à área que mais predomina em seu perímetro. Por exemplo, uma célula constituída por 20% de espaço urbano contínuo e 80% de áreas industriais e comerciais, é classificada como “industrial e comercial”, tendo em vista que é maioritariamente constituída por áreas desta mesma classe. Esta segunda metodologia é abordada no tópico 5.1.3.

### 5.1.1 O software desenvolvido para a geração dos *datasets*

Ao longo deste estudo foram seguidas várias etapas, focadas essencialmente em objetivos distintos, optando-se, sempre que possível, pelo desenvolvimento de mecanismos automáticos para a realização das tarefas envolvidas. Estas etapas fazem parte da estratégia seguida para diminuir a complexidade do problema à volta do qual se deu o estudo em questão. Desta forma, como se pode notar de acordo com o conteúdo apresentado até aqui, alguns dos pontos importantes deste trabalho passam pela criação de softwares. Para além daqueles já abordados anteriormente, foram também desenvolvidos outros dois, para a extração dos conjuntos de dados a serem utilizados tanto na caracterização dos espaços, quanto na análise das correlações.

Para a geração dos conjuntos de dados sobre os quais se baseou o estudo apresentado nos tópicos 5.1.2 e 5.1.3, o aplicativo de linha de comando criado estabelece uma ligação à base de dados e realiza consultas específicas. Para o desenvolvimento do software, foi utilizada a linguagem Java. Ao executar o aplicativo, é apresentado um menu de opções em modo texto, onde o utilizador escolhe a opção correspondente aos conjuntos de dados que deseja extrair. A partir de então, o software realiza uma ou mais buscas, consoante a opção escolhida. Para cada pesquisa realizada, é criado, no diretório parametrizado, um ficheiro do tipo CSV, onde são inseridos os resultados obtidos. Estão implementadas todas as consultas SQL utilizadas para a geração dos dados empregados no estudo. Entretanto, o aplicativo foi desenvolvido com recurso ao padrão “estratégia”, de maneira a facilitar a implementação de qualquer nova consulta, caso seja necessário, garantindo assim uma maior flexibilidade.

### 5.1.2 Experiências baseadas no COS

Conforme introduzido nas seções anteriores, as experiências iniciais foram realizadas com base nos centróides dos polígonos do *dataset*. Desta forma, através do códigoSQL mostrado no apêndice A.8, foram estabelecidos os centróides para cada polígono. Uma amostra do resultado do processo pode ser vista a seguir, na Figura 11, que mostra os polígonos do COS, com seus respectivos centróides, onde são representados, em azul, os espaços urbanos mais densos e em castanho as áreas predominante industriais e comerciais. Considerando que cada polígono representa o espaço onde há predominância de uma determinada classe de uso do espaço, as métricas para a classificação dos polígonos foram bastante diversificadas, conforme detalhado nos tópicos a seguir.

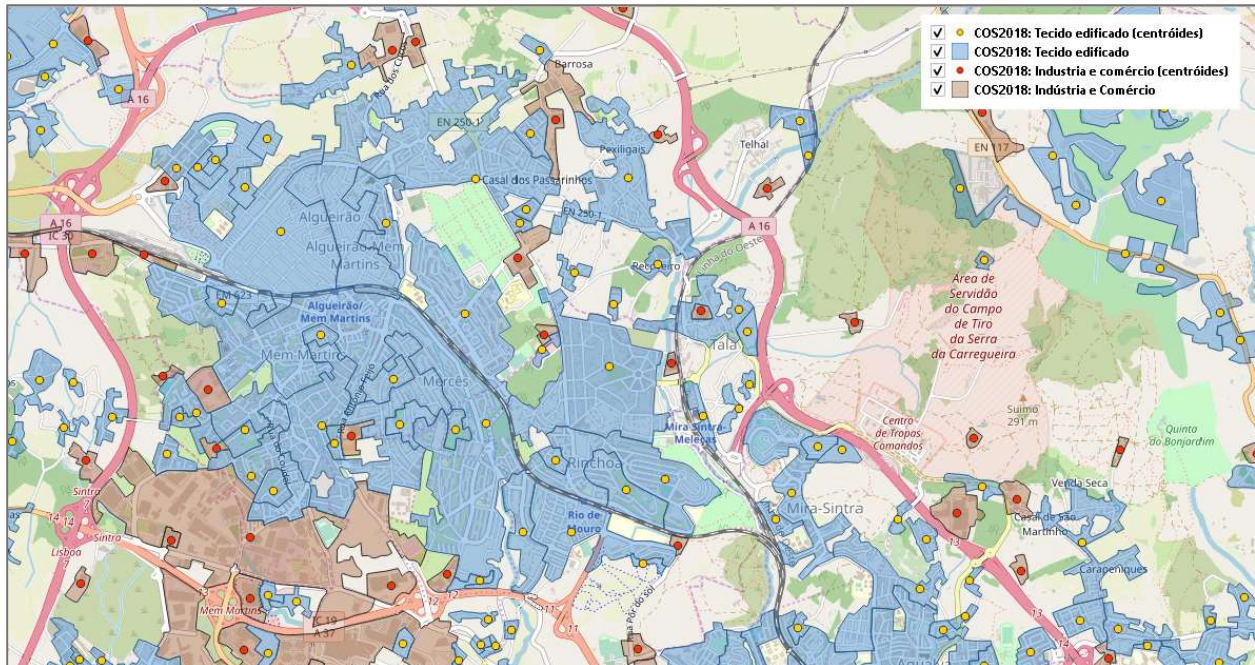


Figura 11: Amostra dos polígonos do COS.

Foram realizados diferentes experiências para classificação do uso dos espaços, com base em diferentes conjuntos de atributos. Nesta seção, serão abordados os testes realizados com base na Carta de Uso e Ocupação do Solo. De forma introdutória, a Tabela 6 mostra uma descrição resumida de cada um dos cenários realizados com base neste conjunto de dados e nos atributos específicos, selecionados em cada caso, os quais são descritos com maior detalhe nas seções seguintes.

Para efeitos de comparação, foram utilizados diferentes classificadores, nomeadamente *Naive Bayes* - NB, *Random Forest* - RF, *Support Vector Machine* - SVM e *Neural Net* – NN. Os parâmetros utilizados nos algoritmos podem ser consultados no apêndice D.1. Quanto à validação, utilizou-se *cross-validation* com 10 *folds* e *stratified sample*. Relativamente às métricas de avaliação, foram utilizadas as seguintes: *accuracy*, *F-measure*, *Kappa*, *weighted mean recall* e *weighted mean precision*. Considerando que a rede neuronal foi o classificador que em geral obteve melhor resultado, foram também realizados testes de comparação entre este classificador e o Naive Bayes, utilizando-se *split validation*, com 70% dos exemplos utilizados para a construção do modelo e os restantes 30% para o treino.

Tabela 6: Cenários de classificação realizados com o *dataset* COS.

Cenário	Descrição da experiência de classificação
Cenário 1	Classes “industrial”, “comercial” e “instalações agrícolas” utilizando 22 atributos extraídos das subcategorias da categoria “Business”, dos pontos de interesse.
Cenário 2	Classes “industrial”, “comercial” e “instalações agrícolas”, recorrendo-se a um conjunto de 44 atributos, extraídos das mesmas categorias do cenário anterior.

<b>Cenário 3</b>	Classes “industrial e comercial” e “instalações agrícolas” utilizando o mesmo conjunto de atributos do cenário 2.
<b>Cenário 4</b>	Classes “tecido edificado” e “indústria e comércio” com utilização dos mesmos atributos do caso anterior.
<b>Cenário 5</b>	Classes “tecido edificado” e “indústria e comércio” e onde foi feita uma seleção automática de atributos extraídos das categorias de nível 2, 3, 4 e 5 dos POIs.

### 5.1.2.1 Cenário 1

Para primeira experiência baseada no *dataset* COS, extraiu-se um conjunto de vários atributos dos pontos de interesse, com base nas categorias às quais pertencem. Alguns dos atributos extraídos representam a distância do centróide do polígono até o ponto de interesse mais próximo, de cada categoria. Outros representam a distância até o mais distante. Considerando a possível relevância das subcategorias a nível 2, da categoria “Business”, optou-se neste primeiro cenário, pela utilização das mesmas. A descrição de cada uma das 44 métricas extraídas pode ser consultada no apêndice F.1. Para auxiliar a geração do conjunto de dados utilizado neste cenário e alguns dos próximos, foram utilizados os comandos SQL apresentados no apêndice A.9.

No primeiro cenário executado com base neste *dataset*, realizou-se a classificação dos espaços urbanos em três classes distintas: indústria, comércio e instalações agrícolas. Foi extraído um conjunto aleatório, entretanto sem repetição, de 1250 registros de cada classe, traduzindo-se num conjunto balanceado, com um total de 3750 exemplos. Ao analisar uma matriz de correlação, observa-se à partida que os atributos que representam a distância entre o centróide do polígono e o ponto de interesse mais distante de cada categoria, possuem baixa correlação com as classes, apresentando valores que variam entre -0.092 e -0.073, sendo portanto irrelevantes para esta experiência. Por outro lado, os 22 atributos que representam distância entre o centróide e o ponto de interesse mais próximo de cada categoria, mostraram correlações entre 0.305 e 0.479, sendo portanto mais indicados para a classificação.

Eliminados os atributos que apresentaram correlações inferiores a 0.3 com as classes, foi realizado o treino e o teste dos classificadores. A seguir, na Tabela 7, podem ser vistos os resultados obtidos em cada caso. As matrizes de confusão geradas com base no teste de cada classificador podem ser consultados no apêndice E.1.

Tabela 7: Resultados obtidos no cenário 1, dos testes com o COS.

Cross-validation					Split validation	
	NB	RF	SVM	NN	NB	NN
<b>Accuracy</b>	57.57	58.00	42.64	65.49	56.98	66.76
<b>F-Measure</b>	55.43	53.08	34.19	65.19	54.87	67.13
<b>Kappa</b>	0.36	0.37	0.14	0.48	0.36	0.50
<b>Weighted mean recall</b>	57.57	58.00	42.64	65.49	56.98	66.76
<b>Weighted mean precision</b>	58.45	58.70	46.90	66.05	57.99	68.33

Como se pode observar na Tabela 8, ao analisar a *precision* e o *recall* de cada classe, nota-se claramente que o RF, tal como o NB, alcançam resultados próximos, enquanto que o SVM obtém resultados inferiores. Em contrapartida, a rede neuronal é o classificador que apresenta melhores resultados. Nota-se que frequentemente os classificadores apresentam resultados menos satisfatórios relativamente às classes “Industrial” e “Comercial”. Ao observar a matriz de confusão gerada para cada um desses classificadores, conclui-se que no geral os mesmos distinguem relativamente bem a classe “Instalações agrícolas” das demais, entretanto têm dificuldade em distinguir as outras duas, sendo provavelmente este o motivo da baixa performance na classificação. Esta observação levou à elaboração de outro cenário, onde as categorias “Indústria” e “Comércio” figuram juntas numa mesma classe, mantendo-se em separado a classe “Instalações agrícolas”. Este cenário será abordado mais adiante.

Tabela 8: *Precision e recall* obtidos com *cross-validation* no cenário 1, com o COS.

Cross-validation								
Classe	Naïve Bayes		Random Forest		SVM		Neural Net	
	Prec.	Rec.	Prec.	Rec.	Prec.	Rec.	Prec.	Rec.
<b>Industrial</b>	42.00%	24.56%	44.75%	12.96%	8.58%	1.84%	54.64%	48.48%
<b>Comercial</b>	52.18%	86.00%	49.22%	91.28%	39.49%	97.68%	63.85%	70.80%
<b>Instalações agrícolas</b>	81.02%	62.16%	81.50%	69.76%	91.03%	28.40%	76.89%	77.20%

Conforme introduzido anteriormente, considerando que em termos gerais a rede neuronal obteve melhores resultados em comparação com os demais classificadores, realizou-se ainda um outro teste comparativo entre esta e o NB, utilizando-se o *split validation*. Os resultados podem ser observados na Tabela 9. Como era suposto, e também com base nesses dados e nas matrizes de confusão apresentadas no apêndice E.1, em geral os resultados para a classe “Industrial” foram inferiores aos demais, reforçando a ideia de que provavelmente os atributos escolhidos não permitem que os classificadores façam bem a distinção entre as áreas industriais e as demais.

Tabela 9: *Precision e recall* obtidos com *split validation* no cenário 1, como COS.

Split validation				
Classe	Naïve Bayes		Neural Net	
	Precision	Recall	Precision	Recall
<b>Industrial</b>	40.37%	23.47%	53.35%	65.87%
<b>Comercial</b>	51.20%	85.07%	69.52%	58.40%
<b>Instalações agrícolas</b>	82.39%	62.40%	82.13%	76.00%

### 5.1.2.2 Cenário 2

Considerando a baixa performance apresentada pelos classificadores no caso anterior, decidiu-se realizar um outro teste, adicionando novos atributos ao *dataset*. Desta vez foi realizada uma contagem dos pontos de interesse de cada uma das 22 subcategorias da categoria “Business”,

existentes dentro de um raio de 10000 metros a partir do centróide de cada polígono. Foram também adicionados outros 22 atributos que representam a proporção de POIs de cada categoria, existentes num raio de 10000 metros a partir do centróide de cada polígono. Desta forma, para o cálculo da proporção de cada categoria, foi considerado o total de pontos de interesse desta, dentro do raio, a dividir pelo número total de pontos de interesse da mesma. Foram também realizados testes considerando 5000 metros e 2500 metros de raio, entretanto como os resultados foram inferiores, optou-se por manter o valor de 10000 metros. Os atributos adicionados ao conjunto de dados, são descritos no apêndice F.2.

Tendo sido adicionados os novos atributos, verificou-se logo à partida recorrendo a uma matriz de correlação, que a relevância entre os aqueles que representam a quantidade de pontos de interesse de cada categoria dentro do raio, é a mesma dos que representam a proporção. Desta forma, foram mantidos somente os da quantidade, tendo em vista que representam um menor esforço computacional em termos de extração. Como resultando foi obtido um *dataset* agora com 44 atributos, ou seja 22 que já existiam e 22 novos. Como é possível observar na Tabela 10, considerando também um total de 1250 exemplos de cada classe, os resultados alcançados com a adição dos novos atributos são poucos superiores aos obtidos no teste anterior.

Tabela 10: Resultados obtidos no cenário 2, dos testes com o COS.

Cross-validation					Split validation	
	NB	RF	SVM	NN	NB	NN
<b>Accuracy</b>	58.93	58.21	58.00	65.49	58.67	64.89
<b>F-Measure</b>	56.73	53.45	57.77	65.23	56.37	65.17
<b>Kappa</b>	0.38	0.37	0.37	0.48	0.38	0.47
<b>Weighted mean recall</b>	58.93	58.21	58.00	65.49	58.67	64.89
<b>Weighted mean precision</b>	57.78	59.87	57.70	65.90	58.08	65.68

A seguir é possível notar, de acordo com a Tabela 11, que em geral a rede neuronal é o classificador com melhor resultado. As matrizes de confusão geradas para cada modelo podem ser visualizadas no apêndice E.2. Nota-se ainda, de acordo com os dados apresentados, que mesmo com os novos atributos, continua não sendo possível distinguir entre as três classes com resultados satisfatórios, reforçando a ideia apresentada no tópico anterior, da necessidade de criação de um novo teste onde as categorias “indústria” e “comércio” figuram juntas numa mesma classe.

Tabela 11: *Precision e recall* obtidos com *cross-validation* no cenário 2, do COS.

Cross-validation								
Classe	Naïve Bayes		Random Forest		SVM		Neural Net	
	Precision	Recall	Precision	Recall	Precision	Recall	Precision	Recall
<b>Industrial</b>	47.88%	27.12%	48.33%	13.92%	50.16%	49.12%	54.46%	48.88%
<b>Comercial</b>	55.32%	78.56%	49.19%	92.40%	56.24%	52.64%	63.95%	70.80%
<b>Instalações agrícolas</b>	70.17%	71.12%	81.96%	68.32%	66.59%	72.24%	77.17%	76.80%

Nota-se, de acordo com o conteúdo mostrado na Tabela 12, que com a utilização do *split validation*, os resultados obtidos são bastante semelhante aos resultados alcançados com a utilização do *cross-validation*, mostrados acima. Além disso, tal como nos casos anteriores, de acordo com a *precision* e o *recall* pode-se concluir que a classe “instalações agrícolas” é mais facilmente distinguida das demais, pelos classificadores.

Tabela 12: *Precision* e *recall* obtidos com *split validation* no cenário 2, do COS.

Split validation				
Classe	Naïve Bayes		Neural Net	
	Precision	Recall	Precision	Recall
<b>Industrial</b>	47.52%	25.60%	52.13%	58.67%
<b>Comercial</b>	53.19%	80.00%	67.54%	62.13%
<b>Instalações agrícolas</b>	73.54%	70.40%	77.37%	73.87%

### 5.1.2.3 Cenário 3

Tendo por base os resultados dos testes anteriores, decidiu-se ainda realizar uma outra experiência com a classificação dos espaços em duas categorias: (1) “industrial e comercial”, formada pelos polígonos que as categorias “indústria” e “comércio”; e (2) “instalações agrícolas”. Os atributos utilizados foram os mesmos que apresentaram maior correlação com as classes no caso anterior, baseados nas subcategorias (nível 2) da categoria “Business”, nomeadamente: (1) a distância do centróide de cada polígono até o ponto de interesse mais próximo de cada subcategoria; e (2) a quantidade de pontos de interesse de cada subcategoria dentro de um raio de 10000 metros a partir do centróide do polígono. Neste cenário também foram utilizados 1250 exemplos de cada classe. As matrizes de confusão geradas por cada classificador podem ser consultadas no apêndice E.3. Como se pode ver a seguir, na Tabela 13, os resultados obtidos foram bastante superiores aos alcançados anteriormente.

Tabela 13: Resultados obtidos no cenário 3, dos testes com o COS.

Cross-validation					Split validation	
	NB	RF	SVM	NN	NB	NN
<b>Accuracy</b>	78.64	82.94	78.70	84.72	78.27	83.40
<b>F-Measure</b>	78.59	82.85	78.69	84.71	78.23	83.32
<b>Kappa</b>	0.57	0.66	0.57	0.69	0.57	0.67
<b>Weighted mean recall</b>	78.64	82.94	78.70	84.72	78.27	83.40
<b>Weighted mean precision</b>	78.91	83.64	78.80	84.99	78.49	84.04

Conforme suposto com base nos resultados anteriores, as classes “indústria e comércio” e “instalações agrícolas” são classificadas com mais eficiência, tendo em vista que são naturalmente mais distintas entre si, diferentemente das classes “indústria” e “comércio”, considerando que também é possível haver espaços comerciais próximos a zonas que são predominantemente

industriais, e vice versa. Nota-se, de acordo com as informações da Tabela 14, que em geral os classificadores alcançaram melhor resultado em comparação com os cenários anteriores, o que era esperado, tendo em conta que foram consideradas somente duas classes.

Tabela 14: *Precision e recall* obtidos com *cross-validation* no cenário 3, do COS.

Cross-validation								
Classe	Naïve Bayes		Random Forest		SVM		Neural Net	
	Prec.	Rec.	Prec.	Rec.	Prec.	Rec.	Prec.	Rec.
<b>Industria e Comércio</b>	76.21%	83.28%	78.88%	89.96%	79.96%	76.60%	83.41%	86.68%
<b>Instalações agrícolas</b>	81.57%	74.00%	88.32%	75.92%	77.54%	80.80%	86.14%	82.76%

Em relação aos testes com o *split validation*, nota-se, de acordo com os dados mostrados na Tabela 15, que apesar dos resultados obtidos serem um pouco inferiores àqueles apresentados com a utilização do *cross-validation*, em geral são também superiores aos dos testes anteriores. É possível verificar ainda, que o NB apresenta uma *precision* bastante inferior à da rede neuronal, na classe “indústria e comércio”. Nota-se também na classe “instalações agrícolas”, que a rede neuronal obteve um *recall* relativamente superior.

Tabela 15: *Precision e recall* obtidos com *split validation* no cenário 3, do COS.

Split validation				
Classe	Naïve Bayes		Neural Net	
	Precision	Recall	Precision	Recall
<b>Industria e Comércio</b>	75.98%	82.67%	88.72%	76.53%
<b>Instalações agrícolas</b>	80.99%	73.87%	79.37%	90.27%

#### 5.1.2.4 Cenário 4

Apesar de se verificar que é possível distinguir com eficiência a classe “indústria e comércio”, da classe “instalações agrícolas”, acredita-se que seja de maior valor em termos de aplicação prática, conseguir classificar também os espaços residenciais. Embora se considere que o cenário mais adequado para este estudo seja conseguir classificar os espaços urbanos em pelo menos 4 categorias, nomeadamente: (1) “agricultura”, (2) “florestas”, (3) “espaços residenciais” e (4) “espaços comerciais ou industriais”, decidiu-se que a melhor opção para atingir este objetivo seria por etapas. Na experiência realizada nesta etapa, optou-se por considerar apenas as classes “tecido edificado”, que representa os espaços residenciais, e “indústria e comércio”, afim de verificar se os classificadores apresentam resultados relevantes no que diz respeito a estas duas classes.

Como se pode observar na Tabela 16, utilizando os mesmos atributos do cenário anterior e por meio de um amostra composta por 4400 exemplos da classe “tecido edificado” e outros 4400 da classe “industrial e comercial”, os resultados obtidos foram bastante semelhantes aos

apresentados nos cenários 1 e 2. As matrizes de confusão geradas com base no teste de cada modelo podem ser consultadas no apêndice E.4. Tendo em conta que neste caso são consideradas somente 2 classes, os resultados indicam que os dados utilizados provavelmente não possuem grande potencial para a classificação dos espaços aquando da utilização da atual metodologia. Nesta altura do estudo, suspeitou-se que o provável motivo tivesse uma relação direta com as categorias escolhidas, o que levou à elaboração de um novo teste, que é apresentado no tópico a seguir.

Tabela 16: Resultados obtidos no cenário 4, dos testes com o COS.

Cross-validation					Split validation	
	NB	RF	SVM	NN	NB	NN
<b>Accuracy</b>	61.48	58.08	62.08	64.30	61.97	65.11
<b>F-Measure</b>	60.24	52.32	62.07	63.91	60.83	65.02
<b>Kappa</b>	0.23	0.16	0.24	0.29	0.24	0.30
<b>Weighted mean recall</b>	61.48	58.08	62.08	64.30	61.97	65.11
<b>Weighted mean precision</b>	63.11	65.53	62.10	65.20	63.54	65.27

Conforme mostrado abaixo, na Tabela 17, verifica-se que em termos de *precision* e *recall*, os resultados mais equilibrados são aqueles apresentados pelo SVM, ao contrário do caso discutido no cenário 1. Apesar disso, como se pode verificar ao considerar também as demais métricas apresentadas anteriormente, neste cenário a rede neuronal continua a ser o classificador com melhores resultados.

Tabela 17: *Precision* e *recall* obtidos com *cross-validation* no cenário 4, com o COS.

Cross-validation								
Classe	Naïve Bayes		Random Forest		SVM		Neural Net	
	Precision	Recall	Precision	Recall	Precision	Recall	Precision	Recall
<b>Ind. e Comércio</b>	58.48%	79.11%	54.77%	92.84%	62.44%	60.61%	61.86%	74.57%
<b>Tecido edificado</b>	67.73%	43.84%	76.51%	23.32%	61.74%	63.55%	67.99%	54.02%

Ainda de acordo com os dados mostrados na Tabela 18, é possível concluir que no geral os resultados obtidos tanto pelo NB quanto pela rede neuronal utilizando-se o *split validation*, não apresentam grande discrepância se comparados com aqueles obtidos aquando da utilização do *cross-validation*. No geral os resultados foram bastante abaixo daquilo que se esperava, pelo que se optou pela geração de novos *datasets* com atributos baseados em outras categorias dos pontos de interesse. Este cenário é discutido no tópico a seguir.

Tabela 18: *Precision* e *recall* obtidos com *split validation* no cenário 4, do COS.

Split validation				
Classe	Naïve Bayes		Neural Net	
	Precision	Recall	Precision	Recall
<b>Industria e Comércio</b>	58.93%	79.02%	63.71%	70.23%
<b>Tecido edificado</b>	68.16%	44.92%	66.84%	60.00%

### 5.1.2.5 Cenário 5

Realizados os testes preliminares discutidos nas secções anteriores, optou-se por executar uma seleção automática de atributos, de maneira a manter somente aqueles que se mostrassem mais relevantes para a classificação. Devido às limitações dos servidores utilizados, no que diz respeito ao processamento de grande volume de dados, a experiência apresentada neste tópico considerou somente a região metropolitana de Lisboa, composta por 18 municípios, nomeadamente os Concelhos de Alcochete, Almada, Amadora, Barreiro, Cascais, Lisboa, Loures, Mafra, Moita, Montijo, Odivelas, Oeiras, Palmela, Seixal, Sesimbra, Setúbal, Sintra, Vila Franca de Xira.

Para possibilitar a execução da experiência realizada neste cenário, foram realizadas algumas operações à base de dados, com recurso ao código disponível no apêndice A.10. Foram extraídos dos pontos de interesse, atributos baseados em todas as categorias de nível 2, 3, 4 e 5. Alguns desses atributos representam, tal como nos cenários anteriores, a distância entre o centróide do polígono e o ponto de interesse mais próximo, de cada categoria. De maneira semelhante, os demais atributos representam a quantidade de pontos de interesse de cada categoria, existentes dentro de um raio de 10000 metros a partir do centróide de cada polígono.

Para a seleção de atributos, utilizou-se um método de *ranking* [114], onde foram testados diferentes alternativas, incluindo pesos atribuídos com base em correlações e nos critérios *information gain*<sup>19</sup>, *gain ratio*<sup>20</sup>, *gini index*<sup>21</sup>, entre outros. Entretanto, o melhor resultado foi obtido com base no uso do SVM. Desta forma, recorrendo ao uso do *Support Vector Machine*<sup>22</sup>, foi feita a atribuição de pesos a cada um dos atributos, obtendo um *dataset* final onde foram selecionados 63 atributos, os quais são apresentados no apêndice F.3.

Ao analisar os datasets extraídos para este cenário, verificou-se logo à partida que a devido aos limites geográficos da região escolhida, foram gerados somente 1360 exemplos da classe “industrial e comercial”. Utilizando-se este valor como parâmetro para o balanceamento do conjunto de dados, obteve-se um total de 2720 exemplos, um número relativamente pequeno, se comparado com o teste realizado no cenário anterior. Como se pode notar com base nas informações apresentadas na Tabela 19, utilizando-se o *cross-validation* com *stratified sample*, obtém-se um resultado inferior ao que foi alcançado com a utilização dos 44 atributos apresentados nos cenários anteriores. De forma similar, os resultados apresentados pelos classificadores aquando da utilização do *split validation*, também ficam aquém daqueles alcançados no teste anterior.

---

<sup>19</sup> <http://tiny.cc/wqoklz> [último acesso: 11/03/2020]

<sup>20</sup> <http://tiny.cc/vsoklz> [último acesso: 11/03/2020]

<sup>21</sup> <http://tiny.cc/6toklz> [último acesso: 11/03/2020]

<sup>22</sup> <http://tiny.cc/uuoklz> [último acesso: 11/03/2020]

Tabela 19: Resultados obtidos no cenário 5, dos testes com o COS.

Cross-validation					Split validation	
	NB	RF	SVM	NN	NB	NN
<b>Accuracy</b>	57.35	55.70	58.64	62.35	56.25	59.44
<b>F-Measure</b>	55.51	49.56	58.64	62.35	53.77	59.26
<b>Kappa</b>	0.15	0.11	0.17	0.25	0.13	0.19
<b>Weighted mean recall</b>	57.35	55.70	58.64	62.35	56.25	59.44
<b>Weighted mean precision</b>	58.83	61.04	58.65	62.76	57.96	59.61

No que diz respeito aos resultados da *precision* e do *recall* de cada classe, torna-se evidente, de acordo com os dados mostrados na Tabela 20 e nas matrizes de confusão apresentadas no apêndice E.5, que com base no conjunto de dados utilizado e nos atributos selecionados, os classificadores não conseguem resultados satisfatórios na distinção entre as classes “indústria e comércio” e “tecido edificado”. Conforme observado nos casos anteriores, a rede neuronal obtém valores pouco superiores aos demais modelos. Mesmo assim, os resultados alcançados ficam muito aquém daquilo que era esperado para este cenário.

Tabela 20: *Precision* e *recall* obtidos com *cross-validation* no cenário 5, do COS.

Cross-validation								
Classe	Naïve Bayes		Random Forest		SVM		Neural Net	
	Precison	Recall	Precison	Recall	Precison	Recall	Precison	Recall
<b>Industria e Comércio</b>	62.44%	36.91%	68.95%	20.74%	58.57%	59.04%	62.23%	62.87%
<b>Tecido edificado</b>	55.22%	77.79%	53.35%	90.66%	58.71%	58.24%	62.48%	61.84%

Ainda de acordo com as informações apresentadas na Tabela 21, nota-se que em geral, com o *split validation*, utilizando-se os parâmetros referidos, os resultados alcançados são ainda inferiores àqueles obtidos com o *cross-validation*, sendo que mesmo assim, a rede neuronal resulta num modelo provavelmente melhor se comparado ao NB.

Tabela 21: *Precision* e *recall* obtidos com *split validation* no cenário 5, do COS.

Split validation				
Classe	Naïve Bayes		Neural Net	
	Precison	Recall	Precison	Recall
<b>Industria e Comércio</b>	61.64%	33.09%	58.32%	66.18%
<b>Tecido edificado</b>	54.27%	79.41%	60.91%	52.70%

### 5.1.2.6 Considerações acerca das experiências realizadas

Nas seções anteriores foram apresentadas 5 experiências realizadas com base no *dataset* COS, edição 2018. No cenário 1 foi realizado um teste preliminar, com a classificação do uso do solo em 3 categorias: "Indústria", "Comércio" e "Instalações agrícolas". Foram realizadas experiências baseadas em diferentes classificadores, recorrendo-se ao uso tanto do *cross-validation*, quanto do *split validation*. As amostras foram extraídas de maneira extratificada, sendo porém que os *datasets* foram posteriormente balanceados em termos de classes. Os resultados obtidos foram relativamente inferiores aos que se esperava, mostrando sobretudo que os modelos criados apresentavam pouca assertividade para distinguir as classes "indústria" e "comércio" entre si e entre a classe "instalações agrícolas".

Considerando os resultados obtidos no cenário 1, algumas hipóteses foram levantadas relativamente ao ocorrido. Tendo em conta que os 22 atributos utilizados representam a menor distância entre o centróide de cada polígono e o ponto de interesse mais próximo, de cada subcategoria, a nível 2, da categoria "Business", suspeitou-se que os mesmos seriam insuficientes para treinar os modelos de forma eficiente. Resolveu-se portanto testar a segunda hipótese, originando o cenário 2, onde são consideradas as mesmas classes, entretanto são adicionados 22 novos atributos ao *dataset*. Desta forma, cada um dos novos atributos gerados representa a quantidade de pontos de interesse, de cada uma das mesmas subcategorias, existentes dentro de um raio a partir do centróide do polígono. Para a escolha do raio, foram testadas as seguintes medidas: 2500, 5000 e 10000 metros. Tendo em conta que os melhores resultados foram obtidos com o raio de 10000 metros, adotou-se esta medida como parâmetro para geração dos atributos adicionados ao conjunto de dados.

Realizado o treino e o teste dos modelos do cenário 2, observou-se de imediato, com base nos resultados obtidos, que os mesmos apresentaram uma performance bastante inferior ao que se esperava. Ao analisar de maneira mais pormenorizada os cenários 1 e 2, foi possível concluir que os modelos apresentam uma notável dificuldade na distinção das classes. Um dos prováveis motivos seria a similaridade entre as classes "indústria" e "comércio", provavelmente pela proximidade geográfica que as regiões de ambas as classes apresenta. Essa conclusão está perfeitamente alinhada com o que se nota na dinâmica das cidades modernas: os espaços industriais frequentemente apresentam muita atividade comercial, sendo que também nota-se no geral que existem indústrias de vários setores instaladas em espaços comerciais. Explorando este ponto de vista, a Figura 12 mostra um exemplo da situação apresentada, onde se pode observar nas zonas em destaque, a proximidade geográfica existente entre os espaços comerciais, mostrados em verde, e industriais, representados pelos polígonos em vermelho.

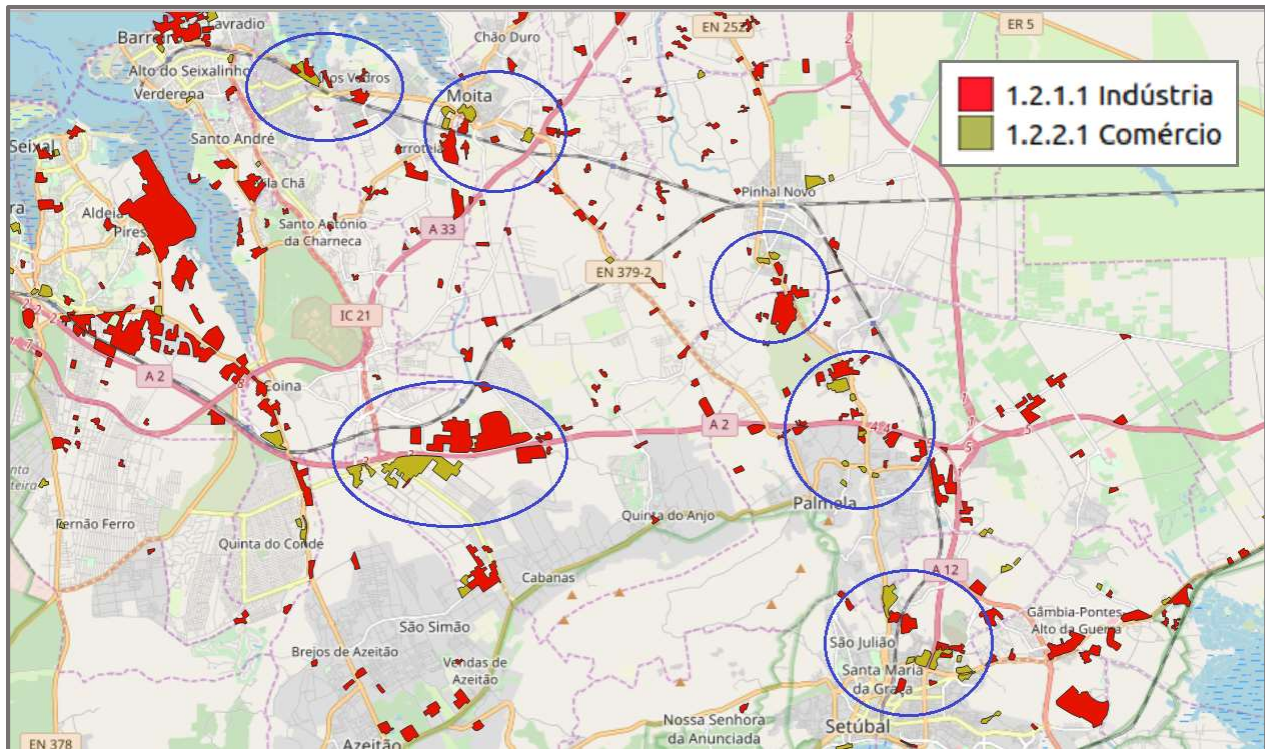


Figura 12: Exemplo da proximidade geográfica entre os espaços comerciais e industriais.

A proximidade que se nota entre as zonas comerciais e industriais pode ser um dos motivos para a baixa performance apresentada nos testes. Considerando que alguns dos atributos representam a distância entre os centróides dos polígonos e o ponto de interesse mais próximo, de cada categoria, logo, se um centróide é muito próximo do outro, as distâncias para ambos até aos POIs também serão muito próximas. O mesmo acontece com a quantidade de pontos de interesse de cada categoria dentro do raio. Tendo por base a proximidade dos centróides, esta quantidade também será muito próxima.

Seguindo as evidências apresentadas nos cenários 1 e 2, optou-se por um novo teste onde os polígonos que representam as zonas industriais e comerciais, figuravam juntos numa mesma classe. Desta forma o cenário 3 considerava as classes “indústria e comércio” e “instalações agrícolas”. Apesar do resultado obtido ser muito superior àqueles alcançados nos casos anteriores, um classificador que apenas consegue distinguir entre estas duas categorias de uso do solo, seria pouco útil em termos de aplicação prática. Desta forma, de maneira a tentar encaminhar os testes na direção de um modelo mais robusto, optou-se por uma nova experiência no cenário 4, onde os classificadores pudessem distinguir os espaços residenciais dos industriais e comerciais. Este cenário apresentou um resultado pouco satisfatório, provavelmente devido à pouca relevância dos atributos escolhidos, no que diz respeito à distinção destas duas classes.

Partindo do pressuposto que os atributos utilizados nos testes anteriores não permitiam a criação de um modelo mais eficiente para alcançar o objetivo do estudo, optou-se pela extração de

atributos baseados em um conjunto mais amplo de categorias, permitindo desta forma um método de seleção automática de atributos. Assim, o cenário 5 tinha o objetivo de distinguir, tal como na experiência anterior, entre as classes “indústria e comércio” e “tecido edificado”. Foram selecionados com base na atribuição de pesos, recorrendo ao uso do SVM, 63 atributos, extraídos das categorias de nível 2, 3, 4 e 5. Entretanto, devido às limitações incontornáveis do poder de processamento dos servidores utilizados, o experimento foi limitado à zona metropolitana de Lisboa. Provavelmente por este motivo, que levou a uma redução muito significativa no número de exemplos, os resultados alcançados foram pouco satisfatórios.

Partindo de uma análise geral de todos os cenários apresentados, chegou-se à conclusão que seria necessária uma metodologia diferente da atual para, tendo em conta os recursos computacionais e os dados disponíveis, proporcionar a possibilidade de alcançar resultados mais satisfatórios nesta etapa do estudo. Com base neste pressuposto, decidiu-se então realizar um teste com os polígonos do CORINE *Land Cover*. Os testes realizados com este conjunto de dados apresentaram resultados muito mais relevantes que aqueles obtidos até então. Estas novas experiências são portanto apresentados na seção seguinte.

### 5.1.3 Experiências baseadas no CORINE

Posteriormente aos testes realizados com o COS, foram realizadas várias experiências com base no CORINE *Land Cover*, edição 2018. Ao analisar o *dataset*, verifica-se com muita clareza, no que diz respeito às classes “Continuous urban fabric”, “Discontinuous urban fabric” e “Industrial or commercial units”, que o mesmo apresenta polígonos relativamente grandes, como se pode observar na Figura 13, onde é mostrado um exemplo da dimensão dos polígonos do *dataset*. Na imagem vê-se um polígono da classe “112. Discontinuous urban fabric” representando uma área de mais de 556 hectares. Desta forma, e tendo em conta as evidências que indicavam que a metodologia utilizada nas experiências anteriores poderia não ser a mais adequada para o objetivo pretendido, optou-se por um novo método que pudesse representar o espaço de maneira mais uniforme.

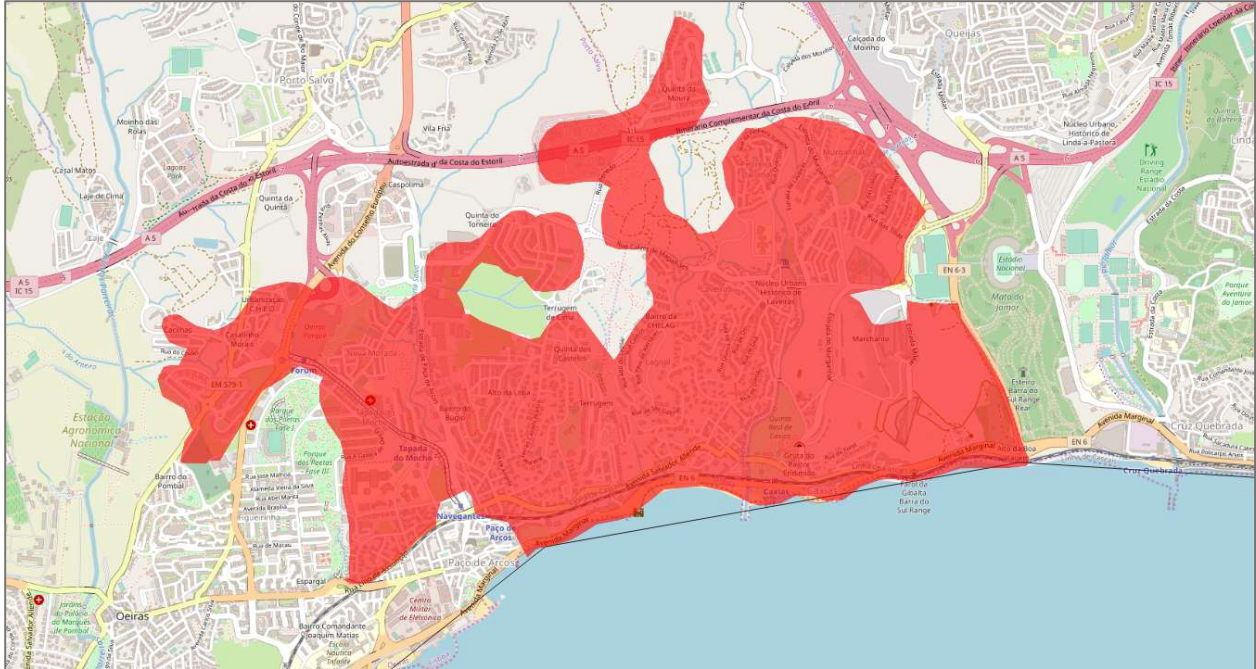


Figura 13: Exemplo da dimensão dos polígonos do CORINE.

Para os testes realizados com este novo conjunto de dados, optou-se portanto por uma metodologia similar a que foi utilizada por Calegari et al. [9], baseada na construção de uma *grid*, cujas células representam cada parte do espaço estudado. Desta forma, focando-se na área metropolitana de Lisboa, foi construída uma grelha, onde cada célula ocupa uma área de 250 x 250 metros, como mostrado a seguir, na Figura 14. Nesta figura, a parte esquerda representa a grelha inicial, enquanto a direita representa o resultado de operação de interseção entre a grelha original e um polígono que representa os limites oficiais da região do estudo, de maneira a manter somente as células da zona pretendida.



Figura 14: *Grid* construída para as experiências baseadas no CORINE.

Após a criação da grelha, realizou-se um reagrupamento das classes existentes no novo conjunto de dados, reduzindo as mesmas ao conjunto apresentado a seguir, na Tabela 22. Após esta etapa, a grelha criada foi sobreposta, no QGIS, sobre os polígonos do CORINE e em seguida foram extraídos os percentuais de cada classe, existente em cada célula da grelha. Desta forma, para cada célula considerou-se a classe que mais predomina na área em que representa. Por exemplo, se em uma célula 10% do território pertence à classe “Continuous urban fabric”, 15% à classe “Port areas” e os restantes 75% à classe “Industrial or commercial units and public facilities”, esta última, como é a que mais predomina, é assumida como sendo a classe a que pertence a célula.

Tabela 22: Reagrupamento das classes do CORINE.

<b>Classe</b>	<b>Subclasse</b>
<b>Class 1.1 Urban fabric</b>	111. Continuous urban fabric
	112. Discontinuous urban fabric
<b>Class 1.2 Industrial, commercial and transport units</b>	121. Industrial or commercial units and public facilities
	122. Road and rail networks and associated land
	123. Port areas
	124. Airports
<b>Class 1.3 Mine, dump and construction sites</b>	131. Mineral extraction sites
	132. Dump sites
	133. Construction sites
<b>Class 1.4 Artificial non-agricultural vegetated areas</b>	141. Green urban areas
	142. Sport and leisure facilities
<b>2. Agricultural areas</b>	2. Agricultural areas
<b>3. Forest and semi-natural areas</b>	3. Forest and semi-natural areas
<b>4. Wetlands</b>	4. Wetlands
<b>5. Water bodies</b>	5. Water bodies

As *layers* resultantes das operações realizadas no QGIS foram importados para a base de dados e em seguida foi aplicada a consulta mostrada no apêndice A.12 para obter a classe predominante em cada célula, de acordo com a regra explicada acima.

Conforme mencionado no capítulo 4, as principais classes do CORINE adotadas para este estudo são, em muitos casos, referidas pelo seu nome traduzido para o português. A classe “1. Artificial surfaces”, frequentemente chamada de “zonas artificiais” é uma das mais utilizadas nas análises. As suas subclasses, tal como “111. Continuous urban fabric”, que representa as áreas urbanas mais densas, “112. Discontinuous urban fabric”, que representa as zonas artificiais menos densas e “121. Industrial, commercial and transport units”, onde se enquadram as áreas industriais, são também algumas das mais importantes para as análises realizadas. Além disso, são também utilizadas as classes “2. Agricultural areas”, onde estão representadas as regiões de agricultura, e “3. Forest and semi-natural areas”, onde se enquadram áreas como parques ambientais e matas.

Tendo sido realizado todo este trabalho de preparação dos dados, foram então extraídos os *datasets* com os atributos a serem utilizados nas experiências de classificação do uso dos espaços. Abaixo, são descritos de maneira abreviada, na Tabela 23, todos os cenários executados para os testes que se basearam no CORINE. É importante reforçar, que nos cenários 1, 2, 3 e 4, foram consideradas as mesmas 5 classes de uso e ocupação do solo, enquanto nos testes 1, 2, 3 e 4 do cenário 5, foram utilizadas 4 classes. Entretanto, nos testes 5, 6 e 7 do cenário 5, foram consideradas 3 classes de *Land Use/Land Cover*.

Tabela 23: Cenários de classificação realizados com o *dataset* CORINE.

<b>Cenário</b>	<b>Descrição da experiência de classificação do uso do solo</b>
<b>Cenário 1</b>	Classes “áreas urbanas contínuas”, “áreas urbanas descontínuas”, “industrial, comercial e instalações públicas”, “agricultura” e “florestas e áreas seminaturais” e 44 atributos.
<b>Cenário 2</b>	Classes “áreas urbanas contínuas”, “áreas urbanas descontínuas”, “industrial, comercial e instalações públicas”, “agricultura” e “florestas e áreas seminaturais” e 200 atributos.
<b>Cenário 3</b>	Classes “áreas urbanas contínuas”, “áreas urbanas descontínuas”, “industrial, comercial e instalações públicas”, “agricultura” e “florestas e áreas seminaturais” e 58 atributos.
<b>Cenário 4</b>	Classes “áreas urbanas contínuas”, “áreas urbanas descontínuas”, “industrial, comercial e instalações públicas”, “agricultura” e “florestas e áreas seminaturais” e 94 atributos selecionados de forma automática.
<b>Cenário 5.1</b>	Classes "Áreas urbanas descontínuas", "Agricultura", "Áreas urbanas contínuas" e "Industrial, comercial e instalações públicas" e 144 atributos
<b>Cenário 5.2</b>	Classes "Florestas e áreas seminaturais", "Áreas urbanas descontínuas", "Áreas urbanas contínuas" e "Industrial, comercial e instalações públicas" e 157 atributos
<b>Cenário 5.3</b>	Classes "Florestas e áreas seminaturais", "Agricultura", "Áreas urbanas contínuas", "Industrial, comercial e instal. públicas" e 142 atributos
<b>Cenário 5.4</b>	Classes "Florestas e áreas seminaturais", "Agricultura", "Áreas urbanas descontínuas" e "Áreas urbanas contínuas" e 221 atributos
<b>Cenário 5.5</b>	Classes "Agricultura", "Áreas urbanas contínuas" e "Industrial, comercial e instalações públicas" e 221 atributos
<b>Cenário 5.6</b>	Classes "Áreas urbanas contínuas e descontínuas", "Agricultura" e "Industrial, comercial e instalações públicas" e 96 atributos
<b>Cenário 5.7</b>	Classes "Agricultura/florestas/áreas seminaturais", "Áreas urbanas contínuas e descontínuas" e "Industrial, comercial e instalações públicas" e 93 atributos

É importante destacar que para os testes realizados nos cenários 1, 2 e 3, os pontos de interesse incluídos nas subcategorias das categorias utilizadas também são considerados. Em outras palavras, se uma categoria “C” é subcategoria de “B”, que por sua vez pertence a “A”, então neste caso todos os pontos de interesse vinculados a “B” e “C” são considerados na listagem dos

pontos de interesse de “A”. Esta situação é discutida nas considerações acerca dos testes realizados com base no CORINE, no tópico 5.1.3.6.

Considerando que em todas as experiências realizadas para a classificação do uso do solo até esta altura, o classificador com melhor resultado foi a rede neuronal, e tendo em conta os recursos computacionais disponíveis, optou-se pela utilização deste classificador nos testes a seguir. Por questões práticas, para o treino e o teste do modelo também se optou pela utilização somente do *cross-validation* com 10 *folds* e *stratified sample*, sendo desta forma, que os resultados obtidos com o *split validation* não são apresentados.

Embora existam diferenças nas categorias com base nas quais foram gerados os vários atributos utilizados em cada cenário, as *features* em si foram obtidas em todos os casos por meio dos critérios adotados anteriormente, ou seja: (1) a distância do centroide da célula até o ponto de interesse mais próximo, de cada categoria e (2), a quantidade de pontos de interesse de cada categoria, existente dentro de um raio de 10000 metros a partir do centroide de cada célula da *grid*. Relativamente ao número de exemplos, nas experiências realizadas nos cenários 1, 2, 3 e 4, foi utilizado um total de 3000 registos de cada classe, com exceção daquela que representa as zonas industriais, para a qual foram considerados 2667 exemplos. Nos restantes cenários, o número de exemplos utilizados é descrito mais adiante.

#### 5.1.3.1 Cenário 1

De maneira a dar continuidade à exploração do potencial de alguns dos atributos utilizados nas experiências realizadas com o COS, optou-se, neste primeiro teste baseado no CORINE, pelas subcategorias da categoria “Business” para a extração das *features*. Para esta experiência, foram consideradas 5 classes, nomeadamente: (1) “Áreas urbanas contínuas”, (2) “Áreas urbanas descontínuas”, (3) “Industrial / Comercial”, (4) “Agricultura” e (5) “Florestas e áreas seminaturais”. Nesta experiência foram considerados os mesmos 44 atributos empregados no cenário 2, dos testes realizados com base nos dados do COS, ou seja, aqueles que são apresentados nos apêndices F.1 e F.2.

Neste cenário, para a classificação das células da grid em 5 categorias diferentes, conforme se pode notar a seguir, na Tabela 24, os resultados alcançados, embora ainda aquém do esperado, foram bastante superiores aos obtidos aquando da utilização de mais de duas classes nos testes com o COS. A matriz de confusão gerada com base no treino do modelo pode ser consultada no apêndice E.6. É possível verificar com base nos resultados, que a classe “industrial / comercial” parece bastante distinta das demais, tal como a classe “áreas urbanas contínuas”. Em contrapartida os resultados mostram que as áreas industriais e as zonas urbanas menos densas são mais difíceis de identificar.

Tabela 24: resultados do cenário 1, dos testes com o CORINE.

Cross-validation					
Medida	Valor		Classe	Precision	Recall
Accuracy	71.34		Florestas e áreas seminaturais	71.45%	71.17%
F-Measure	71.01		Agricultura	67.80%	63.03%
Kappa	0.64		Áreas urbanas desconínuas	66.04%	57.43%
Weighted mean recall	71.51		Áreas urbanas contínuas	78.46%	87.17%
Weighted mean precision	71.46		Industrial / Comercial	71.23%	78.74%

### 5.1.3.2 Cenário 2

Neste segundo cenário das experiências baseadas no CORINE, foram consideradas as mesmas 5 classes do teste anterior. Entretanto, tendo em conta a importância da seleção de atributos para qualquer processo de extração de conhecimento a partir dos dados, foram consideradas novas categorias para a geração das *features* utilizadas neste caso. Assim sendo, a metodologia utilizada para a escolha das categorias dos pontos de interesse passou pela filtragem daqueles geograficamente localizados dentro da área dos polígonos da classe “Artificial surfaces”, do CORINE. Esta classe foi escolhida pelo facto de melhor representar as áreas urbanas. Foram, portanto, isolados os pontos de interesse e a seguir foi obtida a lista de categorias a que pertencem. Esta operação foi realizada com recurso ao código SQL mostrado no apêndice A.13. No total, foram obtidas 100 categorias neste processo, as quais são listadas a seguir, na Tabela 25.

Tabela 25: Categorias dos POIs localizados nas áreas da classe “Artificial Surfaces”.

Producer	Residence
Writer	Event
Music	Labor Union
Concert Tour	Real Estate
Author	Shopping & Retail
Athlete	Landmark & Historical Place
Artist	Meeting Room
Journalist	Social Club
News Personality	Campus Building
Chef	Charity Organization
Comedian	Legal
Dancer	Commercial & Industrial
Designer	Science, Technology & Engineering
Entrepreneur	Fitness Model
Government Official	Private Members Club
Cuisine	Blogger
Media/News Company	Political Organization

Non-Governmental Organization (NGO)	Gamer
Education	Film Director
Nonprofit Organization	Hotel & Lodging
Cause	Editor
Religious Place of Worship	Outdoor Recreation
Work Position	Performance Art
Community	Band
Work Project	Literary Arts
Political Party	Theatrical Play
Science	Sports
Art	Books & Magazines
Orchestra	Finance
Sports Promoter	Talent Agent
Travel & Transportation	Government Building
Just For Fun	Automated Teller Machine (ATM)
Arts & Entertainment	TypeAhead
Beauty, Cosmetic & Personal Care	Visual Arts
Medical & Health	Musician
Public & Government Service	Food & Beverage
Government Organization	Agriculture
Sorority & Fraternity	Actor
Ticket Sales	Brand
Community Service	City Infrastructure
Musician/Band	Motivational Speaker
Automotive, Aircraft & Boat	Media Restoration Service
Youth Organization	TV & Movies
Country Club / Clubhouse	Advertising/Marketing
Sports & Recreation	Performing Arts
Religious Organization	Local Service
Exchange Program	Theatrical Productions
Armed Forces	Fashion Model
Sports Club	Video Creator
Environmental Conservation Organization	Locality

Isoladas as categorias e extraídos os atributos a serem utilizados, foi feita uma seleção manual daqueles mais relevantes, com recurso a uma matriz de correlação. Foram selecionados 188 atributos, com base nos quais foram então realizados o treino e o teste da rede neuronal. Os resultados obtidos podem ser consultados logo a seguir, na Tabela 26. Como se pode notar, a performance do modelo foi pouco superior em relação ao observado cenário anterior, sendo que o classificador mostra as mesmas tendências para as mesmas classes, como já se havia verificado anteriormente.

Tabela 26: Resultados do cenário 2, dos testes com o CORINE.

Cross-validation				
Medida	Valor	Classe	Precision	Recall
Accuracy	72.33	Florestas e áreas seminaturais	72.54%	69.20%
F-Measure	72.14	Agricultura	66.63%	64.63%
Kappa	0.65	Áreas urbanas desconínuas	64.23%	61.40%
Weighted mean recall	72.50	Áreas urbanas contínuas	81.40%	87.23%
Weighted mean precision	72.39	Industrial / Comercial	75.89%	80.01%

### 5.1.3.3 Cenário 3

No seguimento dos testes executados nos cenários anteriores, foi ainda realizada uma nova experiência, baseada em atributos extraídos a partir de outras categorias. Semelhantemente ao que foi feito em [9], neste cenário optou-se por uma seleção manual das categorias, onde foram escolhidas aquelas que mostram maior semelhança às que foram utilizadas pelos autores. Desta forma, no total foram escolhidas 29 categorias, o que resultou num conjunto de 58 atributos. As categorias selecionadas podem ser consultadas a seguir, na Tabela 27. O *script* SQL utilizado para esta operação pode ser consultado no apêndice A.14.

Tabela 27: Categorias selecionadas no cenário 3.

Bank	Parking garage lot
Bar	Pharmacy drugstore
Barber shop	Police station
College university	Pub
Fast food restaurant	Public garden
Furniture	Public government service
Gym physical fitness center	Public school
High school	Public service
Hospital	Public square plaza
Hotel	Public swimming pool
Junior high school	Public utility company
Library	School
Middle school	Supermarket
Mobile home park	Traffic school
Park	

Mesmo considerando que nesta experiência foi utilizada a mesma quantidade de exemplos e o mesmo método para a geração do *sample* utilizado no cenário 2, e tendo ainda em conta que o número de categorias utilizado foi bem menor em relação aquele pelo qual se optou no caso anterior, é possível notar, de acordo com as informações mostradas abaixo, na Tabela 28, que os

resultados obtidos foram bastante semelhantes. Em comparação com o modelo obtido no teste precedente, pode-se observar uma diferença muito pequena em termos de performance, o que não era esperado, tendo em vista que na experiência anterior foi utilizado um total de 100 categorias, o que resultou num *dataset* com 200 atributos.

Tabela 28: Resultados do cenário 3, dos testes com o CORINE.

Cross-validation				
Medida	Valor	Classe	Precision	Recall
Accuracy	71.04	Florestas e áreas seminaturais	70.82%	70.37%
F-Measure		Agricultura	65.97%	63.47%
Kappa	0.64	Áreas urbanas desconínuas	65.18%	53.90%
Weighted mean recall	71.24	Áreas urbanas contínuas	78.37%	88.40%
Weighted mean precision	70.80	Industrial / Comercial	72.74%	80.05%

#### 5.1.3.4 Cenário 4

Ao analisar os resultados obtidos nos cenários anteriores, observa-se que mesmo com todos os progressos alcançados, os modelos obtidos apresentam uma performance ainda inferior àquilo que era esperado como *output* para esta fase do estudo. Desta forma, considerando a grande importância em qualquer estudo deste gênero, daquilo que se conhece por *feature selection*, na experiência realizada neste quarto cenário optou-se pela adoção de um método de seleção automática de atributos. Neste sentido, de maneira a disponibilizar um número mais abrangente de *features*, passou-se, portanto, a considerar todas as categorias de nível 2, 3, 4 e 5, dos pontos de interesse, para a geração das mesmas.

É importante destacar para a geração dos atributos a serem utilizados que neste cenário, não foram considerados os pontos de interesse vinculados às subcategorias das categorias utilizadas. Desta forma, se a categoria “B” é uma subcategoria de “A”, ao calcular por exemplo a distância entre o centroide da célula e o ponto de interesse mais próximo, da categoria “A”, suponha que dos que estão ligados diretamente a esta, a menor distância seja 900 metros. Neste caso, mesmo que exista um ponto de interesse ligado à categoria “B” a 500 metros, o resultado da operação será sempre 900. Este pormenor é bastante importante pois, como se está a trabalhar com as categorias de nível 2, 3, 4 e 5, se fossem consideradas as subcategorias, neste caso para as categorias de nível mais alto muitas das distâncias calculadas seriam iguais as das de nível mais baixo, fazendo com que vários atributos do conjunto de dados tivessem o mesmo valor, apenas variando o nome do atributo.

Para a geração dos atributos com base nas categorias dos níveis escolhidos, considerando que somam um total de 1484, deparou-se novamente com as limitações do poder computacional disponível. Desta forma, mesmo considerando somente os pontos de interesse da área

metropolitana de Lisboa, há que se ter em conta que a realização dos cálculos necessários para a geração das 2968 *features* para cada célula da *grid*, é uma operação bastante dispendiosa em termos de processamento. Por esta razão, foram extraídos vários *datasets* separados, cada um com determinado conjunto de atributos, os quais foram posteriormente objetos de uma operação de junção, no RapidMiner. O *script* SQL utilizado para a geração das várias tabelas e dos índices que facilitaram a extração dos conjuntos de dados utilizados neste cenário, pode ser consultado no apêndice A.10.

Quanto à seleção de atributos, optou-se por um método de *ranking* [114], através do qual é atribuído um determinado peso para cada *feature* com base em um determinado critério. O método foi testado com base em diferentes medidas, nomeadamente: *Information gain*, *Gain ratio* e *Gini index*, adotando-se no fim das contas, o critério que resultou no conjunto de atributos mais relevante para o que se pretendia. Desta forma, com base no *Gain ratio*, foi realizada a seleção automática de 94 *features*, que são apresentadas no apêndice F.4. Os resultados obtidos no treino do modelo gerado são mostrados abaixo, na Tabela 29, onde é possível observar que a performance do classificador ficou equiparada ao que se conseguiu obter nos cenários anteriores, estando ligeiramente abaixo do que se alcançou no cenário 2.

Tabela 29: Resultados do cenário 4, dos testes com o CORINE.

Cross-validation				
Medida	Valor	Classe	Precision	Recall
Accuracy	71.56	Florestas e áreas seminaturais	72.12%	70.03%
F-Measure	71.34	Agricultura	67.27%	62.77%
Kappa	0.64	Áreas urbanas desconínuas	63.59%	60.20%
Weighted mean recall	71.71	Áreas urbanas contínuas	81.32%	86.90%
Weighted mean precision	71.54	Industrial / Comercial	72.12%	78.67%

### 5.1.3.5 Cenário 5

Com base no mesmo conjunto de atributos e nos mesmos critérios de seleção de *features* apresentados no caso anterior, neste cenário foram realizados 7 testes com diferentes classes. As matrizes de confusão geradas com base no treino dos classificadores podem ser consultadas no apêndice E.6. A seguir são apresentados os resultados obtidos em cada caso, os quais são também objeto de uma discussão apresentada no tópico 5.1.3.6.

No primeiro teste realizado foram utilizados 144 atributos., mostrados no apêndice F.5. Neste teste, as 4 classes envolvidas são: (1) “áreas urbanas contínuas”, para a qual foram utilizados 3043 exemplos; (2) “áreas urbanas descontínuas”, para a qual foram considerados 4000 exemplos; (3) “indústria, comércio e instalações públicas”, para a qual foram utilizados 2667 exemplos; e (4) “agricultura”, para a qual foram também utilizados 4000 exemplos. O resultado obtido na

avaliação do classificador pode ser visto a seguir, na Tabela 30. Embora os dados indiquem que o modelo gerado é melhor, se comparado aos anteriores, os espaços urbanos menos densos ainda parecem ser um desafio em termos de classificação.

Tabela 30: Resultados do teste 1, do cenário 5, com base no CORINE.

Cross-validation				
Medida	Valor	Classe	Prec.	Recall
Accuracy	77.41	Áreas urbanas desconínuas	69.94%	71.65%
F-Measure	77.39	Agricultura	83.32%	76.30%
Kappa	0.70	Áreas urbanas contínuas	78.74%	88.24%
Weighted mean recall	77.89	Industrial, comercial e instal. publicas	79.17%	75.37%
Weighted mean precision	78.22			

Tal como no caso anterior, neste segundo teste também se optou pela utilização de 4 classes distintas, nomeadamente: (1) “áreas urbanas contínuas”; (2) “áreas urbanas descontínuas”; (3) “indústria, comércio e instalações públicas”; e (4) “florestas e áreas seminaturais”. Foram considerados 3043 exemplos para a primeira, 4000 para a segunda, 2667 para a terceira e 4000 para a última. Foi também utilizado um conjunto de 157 atributos, que podem ser apreciados no apêndice F.6. Como mostram os resultados apresentados na Tabela 31, o modelo obtido não apresentou grandes vantagens em relação ao anterior.

Tabela 31: Resultados do teste 2, do cenário 5, com base no CORINE.

Cross-validation				
Medida	Valor	Classe	Prec.	Recall
Accuracy	78.53	Florestas e áreas seminaturais	87.27%	82.45%
F-Measure	78.48	Áreas urbanas descontínuas	74.31%	70.28%
Kappa	0.71	Áreas urbanas contínuas	76.28%	87.91%
Weighted mean recall	78.74	Industrial, comercial e instal. publicas	75.09%	74.35%
Weighted mean precision	78.65			

Para além dos testes anteriores, realizou-se ainda uma nova experiência com a utilização de outro conjunto de classes, baseado em 142 *features*., apresentadas no apêndice F.7. Neste caso o modelo foi treinado para distinguir entre as classes “áreas urbanas contínuas”, “indústria, comércio e instalações públicas”, “agricultura” e “florestas e áreas seminaturais”. Foram utilizados respetivamente 3043 e 2667 exemplos para as duas primeiras e 4000 para cada uma das duas últimas. Com base na Tabela 32, nota-se que o teste do classificador apresentou resultados próximos aos anteriores.

Tabela 32: Resultados do teste 3, do cenário 5, com base no CORINE.

Cross-validation				
Medida	Valor	Classe	Prec.	Recall
Accuracy	78.01	Florestas e áreas seminaturais	72.90%	79.07%
F-Measure	77.86	Agricultura	76.93%	66.12%
Kappa	0.70	Áreas urbanas contínuas	87.54%	90.04%
Weighted mean recall	78.94	Industrial, comercial e instal. publicas	76.60%	80.50%
Weighted mean precision	78.57			

No quarto teste realizado neste cenário, onde foram empregados 158 atributos, optou-se por verificar até que ponto uma rede neuronal conseguiria distinguir entre as classes “áreas urbanas contínuas”, “áreas urbanas descontínuas”, “agricultura” e “florestas e áreas seminaturais”. Os atributos utilizados podem ser consultados no apêndice F.8. Neste teste foram utilizados 4000 exemplos para cada classe, com exceção daquela que representa as áreas urbanas contínuas, para a qual foram considerados 3043 exemplos no total. Nota-se com clareza, ao analisar as informações apresentadas na Tabela 33, que mesmo de tendo sido considerado um número bem maior de *features* e o mesmo número de classes dos testes anteriores, os resultados obtidos foram inferiores, com exceção da última classe.

Tabela 33: Resultados do teste 4, do cenário 5, com base no CORINE.

Cross-validation				
Medida	Valor	Classe	Prec.	Recall
Accuracy	73.34	Florestas e áreas seminaturais	72.84%	75.58%
F-Measure	73.15	Agricultura	70.04%	64.65%
Kappa	0.64	Áreas urbanas descontínuas	69.42%	67.75%
Weighted mean recall	74.28	Áreas urbanas contínuas	82.29%	89.16%
Weighted mean precision	73.92			

Com o propósito de avaliar a capacidade de um modelo treinado para reconhecer três das classes de uso do solo que se considera serem de grande importância no caso deste estudo, o quinto teste efetuado teve por base as áreas urbanas mais densas, as de indústria / comércio / instalações públicas e as zonas de agricultura. Relativamente ao número de exemplos de cada classe, foram considerados respectivamente 3043, 2667 e 4000. Neste teste optou-se por um conjunto de 221 atributos, que são listados no apêndice F.9. Os dados da Tabela 34 mostram que o modelo obtido é bastante robusto se comparado aos anteriores, indicando resultados significativamente melhores.

Tabela 34: Resultados do teste 5, do cenário 5, com base no CORINE.

Cross-validation				
Medida	Valor	Classe	Prec.	Recall
Accuracy	90.21	Agricultura	93.68%	91.55%
F-Measure	90.20	Áreas urbanas contínuas	89.75%	93.53%
Kappa	0.85	Industrial, comercial e instal. publicas	85.59%	84.40%
Weighted mean recall	89.83			
Weighted mean precision	89.84			

No seguimento dos casos anteriores, o sexto teste realizado explorou o potencial de um modelo treinado também com base em três classes distintas. A primeira delas é basicamente a união das classes “áreas urbanas contínuas” e “áreas urbanas descontínuas”, enquanto que a segunda, “agricultura” e a terceira, “industrial, comercial e instalações públicas”, mantiveram-se inalteradas tal como no quinto teste. Foram considerados 4000 exemplos para cada uma das duas primeiras e outros 2667 para a última. O treino foi baseado na utilização de 96 atributos, apresentados no apêndice F.10. A Tabela 35 mostra os resultados obtidos no treino do modelo. Como se pode notar, os resultados apresentados, embora inferiores aos do teste anterior, mostram um modelo bastante robusto.

Tabela 35: Resultados do teste 6, do cenário 5, com base no CORINE.

Cross-validation				
Medida	Valor	Classe	Prec.	Recall
Accuracy	82.31	Áreas urbanas contínuas e descontínuas	81.22%	81.17%
F-Measure	82.34	Agricultura	86.65%	83.08%
Kappa	0.73	Industrial, comercial e instal. públicas	77.98%	82.86%
Weighted mean recall	82.37			
Weighted mean precision	82.16			

Por fim e de forma semelhante ao caso anterior, neste último teste foi utilizada a classe “áreas urbanas contínuas e descontínuas”, para a qual foram considerados 4000 exemplos. Foi também criada uma nova classe, formada pela união dos polígonos que representam as categorias “agricultura” e “florestas e áreas seminaturais”. Para esta também se utilizou um total de 4000 exemplos. Mantendo-se inalterada, foi ainda utilizada a classe “industrial, comercial e instalações públicas”, para a qual recorreu-se a um total de 2667 exemplos. Para este teste, foi adotado um conjunto composto por 93 atributos, que são mostrados no apêndice F.11. Como se pode concluir com base nos resultados apresentados na Tabela 36, o modelo criado equipara-se ao que foi apresentado caso anterior.

Tabela 36: Resultados do teste 7, do cenário 5, com base no CORINE.

Cross-validation				
Medida	Valor	Classe	Prec.	Recall
Accuracy	81.88	Agricultura/florestas/áreas seminaturais	84.08%	84.90%
F-Measure	81.89	Áreas urbanas contínuas e descontínuas	79.97%	81.15%
Kappa	0.72	Industrial, comercial e instal. públicas	81.44%	78.58%
Weighted mean recall	81.83			
Weighted mean precision	81.75			

### 5.1.3.6 Considerações acerca das experiências realizadas

Nos tópicos anteriores, foram apresentados vários cenários que representam diferentes experiências, cujo objetivo é a caracterização do uso do solo na área metropolitana de Lisboa. As experiências foram realizadas com base em atributos extraídos a partir de várias combinações de categorias de pontos de interesse. Os testes apresentados nos cenários referidos utilizaram para treino e teste dos classificadores, o *dataset* CORINE, edição 2018. Para a criação dos modelos, recorreu-se ao uso do *cross-validation*, com 10 *folds* e *estratified sample*, optando-se pelo uso de uma rede neuronal.

O primeiro cenário apresentado explora uma experiência onde as *features* foram extraídas a partir das subcategorias da categoria “Business”. Foram escolhidas 5 classes de uso do solo, nomeadamente: (1) “Áreas urbanas contínuas”, (2) “Áreas urbanas descontínuas”, (3) “Industrial / Comercial”, (4) “Agricultura” e (5) “Florestas e áreas seminaturais”. Utilizando-se um total de 44 atributos, obteve-se um resultado superior a todos os testes realizados aquando da utilização do COS, onde se consideravam mais de 2 classes. Observa-se com bastante clareza, com base nos resultados apresentados, que a classe “áreas urbanas contínuas” é a que melhor se distingue das demais. Considerando que esta classe representa as áreas mais densas nos espaços urbanos, nota-se, conforme o exemplo apresentado na Figura 15, que há uma clara concentração de pontos de interesse da categoria “Business” e das suas subcategorias à volta dessas zonas, evidenciando uma relação clara entre esta classe de uso do solo e a referida categoria, na região estudada. Na imagem é mostrado um exemplo da distribuição geográfica dos pontos de interesse das várias subcategorias da categoria “Business” (em verde), em relação aos polígonos da classe “111. Continuous urban area” (em vermelho).

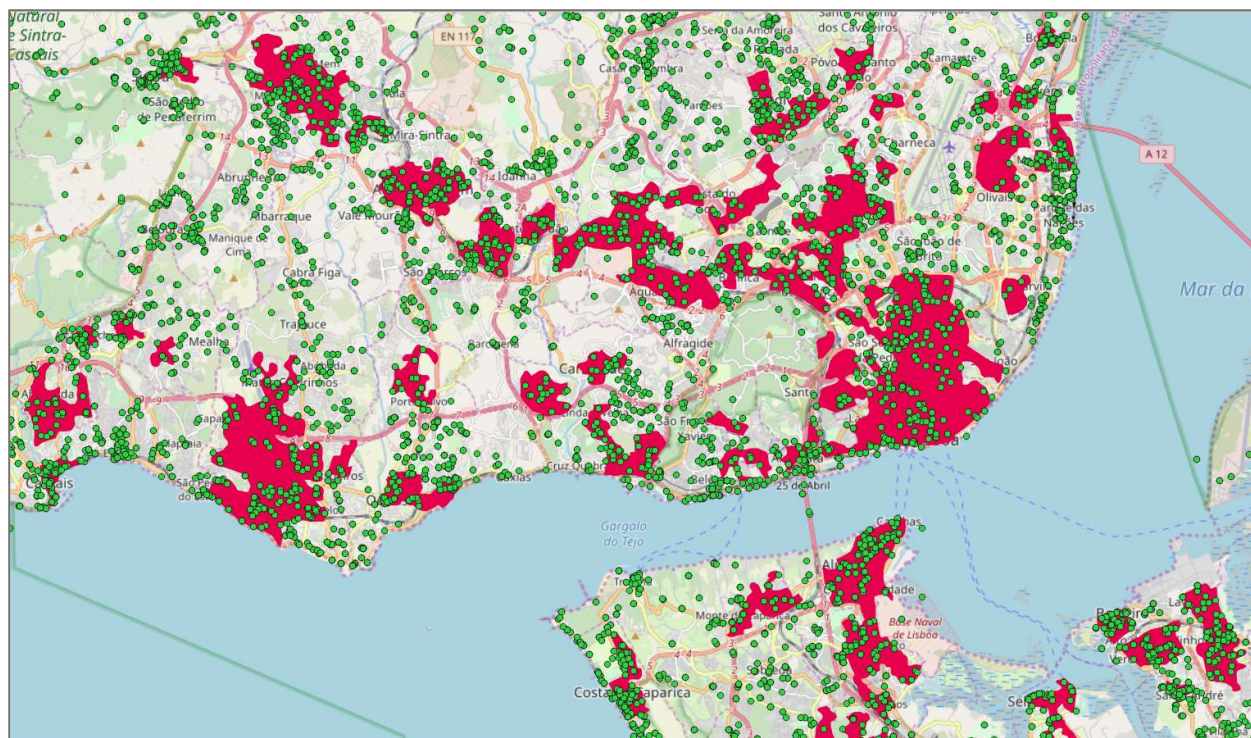


Figura 15: Distribuição geográfica dos POIs das subcategorias da categoria “Business”.

Na tentativa de explorar ainda mais o potencial das diversas categorias dos pontos de interesse, foi realizada uma nova experiência no segundo cenário, onde foram consideradas as mesmas 5 classes de uso do solo do caso anterior. Desta vez, foram escolhidas as categorias que figuravam com maior frequência nos pontos de interesse localizados dentro dos limites geográficos dos polígonos da classe “Urban surfaces”, que representa a maior parte dos espaços urbanos no CORINE. Sendo assim, era esperado que os modelos gerados apresentassem maior performance na identificação tanto das áreas de indústria e comércio, como das zonas mais densas das cidades. Os resultados obtidos contribuem para a sustentação dessa hipótese, tendo em vista que ambas as classes apresentaram maior *precision* e *recall* em comparação com o cenário anterior.

Na sequência dos testes realizados, o terceiro cenário baseou-se numa seleção manual de categorias onde foram escolhidas aquelas que mais se assemelham às que foram utilizadas no estudo conduzido por Calegari et al. [9]. Desta forma, as 29 categorias utilizadas deram origem a um total de 58 atributos que foram adotados na criação do modelo. Considerando as mesmas 5 classes, tal como nos casos anteriores, observa-se, contudo, que os resultados alcançados foram ligeiramente inferiores, entretanto muito próximos daqueles obtidos nas experiências realizadas até então. Partindo do pressuposto que as categorias selecionadas são as que melhor representam os pontos de interesse considerados na escolha por exemplo, da melhor zona para compra de um imóvel residencial (bancos, ginásios, supermercados, bares, escolas, universidades, restaurantes, etc.), assume-se que as mesmas dariam origem a bons atributos para a identificação das classes pretendidas. Esse pressuposto parte do princípio que, como se pode verificar na Figura 16, os

pontos de interesse das categorias escolhidas figuram geralmente próximos às zonas industriais/comerciais e residenciais (áreas urbanas contínuas e descontínuas).

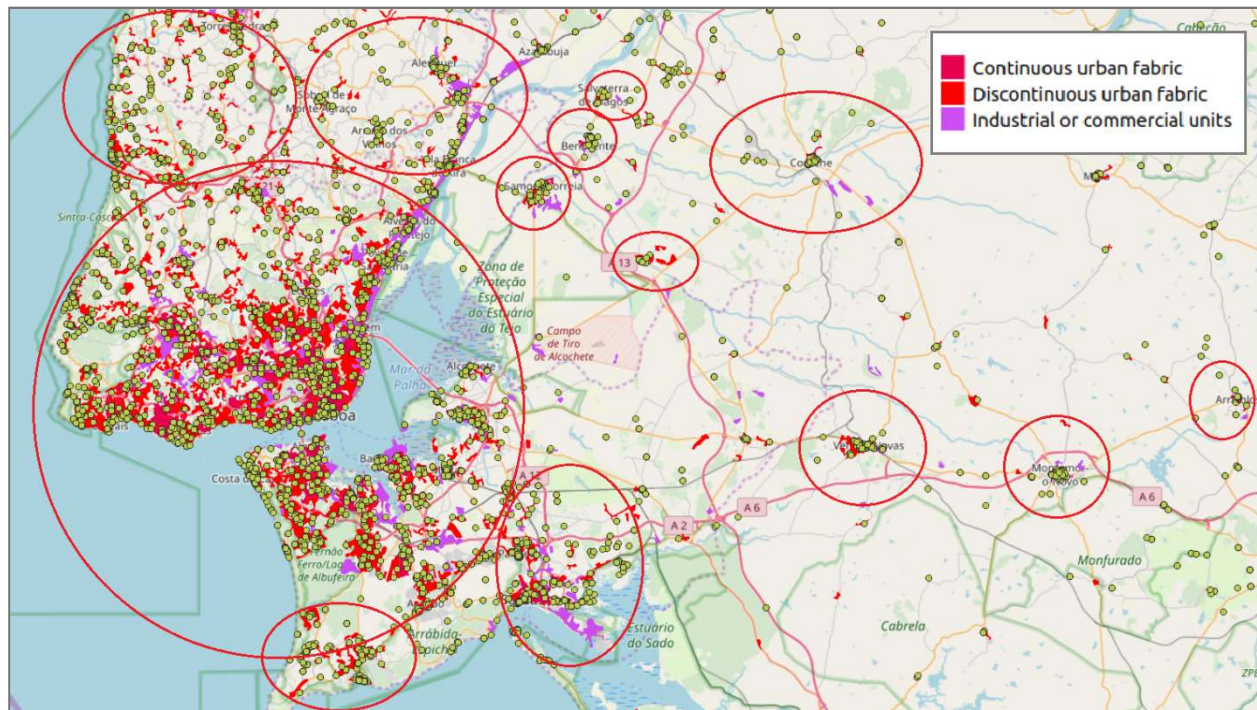


Figura 16: Exemplo da ocorrência dos POIs das categorias selecionadas manualmente.

Embora os resultados obtidos no cenário 3 sejam pouco inferiores aos obtidos anteriormente, os dados mostram que o modelo originado é equivalente aos apresentados até então, tendo em vista que com apenas 58 atributos, conseguiu-se um modelo que se equipara àquele que foi gerado no cenário 2, com recurso a um conjunto de 188 atributos. Mesmo assim considera-se que os resultados obtidos tanto no cenário 1, quanto no 2 e no 3, são inferiores ao que se espera para a utilização num caso prático, pelo que se prosseguiu com o estudo, realizando-se outras duas experiências.

Seguindo-se as direções apontadas pelas experiências anteriores, foi então realizado o cenário 4, onde o teste executado contou com uma seleção automática de atributos, o que supostamente deveria melhorar significativamente os resultados. Entretanto, contrariamente ao que era esperado, os resultados obtidos nesta quarta experiência mostram que o modelo gerado não conseguiu destacar-se em termos de performance quando comparado aos demais, indicando que provavelmente terá sido atingido o potencial máximo dos atributos disponíveis, relativamente às 5 classes escolhidas. Nesta experiência, o método de seleção de atributos teve por base a atribuição de peso com recurso aos vários critérios descritos anteriormente. Entretanto, tendo em conta as limitações dos recursos computacionais disponíveis, não foram testados outros métodos de seleção de atributos, como por exemplo o *forward selection* ou o *backward elimination*. Além disso, acredita-se que seria importante explorar o potencial da geração de atributos, neste caso.

Explorando a suspeita de se ter atingido o potencial máximo das *features* disponíveis, optou-se por analisar a capacidade de classificação dos modelos treinados com menos classes. Desta forma, diferente dos casos anteriores, alguns dos testes realizados no cenário 5 foram baseados em conjuntos de 4 classes, enquanto outros tiveram por base 3 categorias distintas de uso dos espaços. Como se pode observar com base nos resultados do cenário 4, há indicações de que as classes “áreas urbanas descontínuas”, “agricultura” e “florestas e áreas seminaturais” são semelhantes entre si.

Esta hipótese pode ser comprovada com base na Figura 17, onde é possível verificar que os polígonos que representam as zonas urbanas menos densas (classe “112. Discontinuous urban fabric”) estão cercados por áreas agricultáveis e florestas (classes “2. Agricultural areas” e “3. Forest and semi-natural areas”). Além disso, de acordo com a nomenclatura oficial das classes do CORINE [111], a classe "112. Discontinuous urban fabric" representa áreas onde as estruturas urbanas estão associadas a áreas de vegetação e onde há predominância das superfícies permeáveis. Essas áreas normalmente ocorrem nas periferias das zonas urbanas, tal como as áreas mais utilizadas para a agricultura e as florestas.

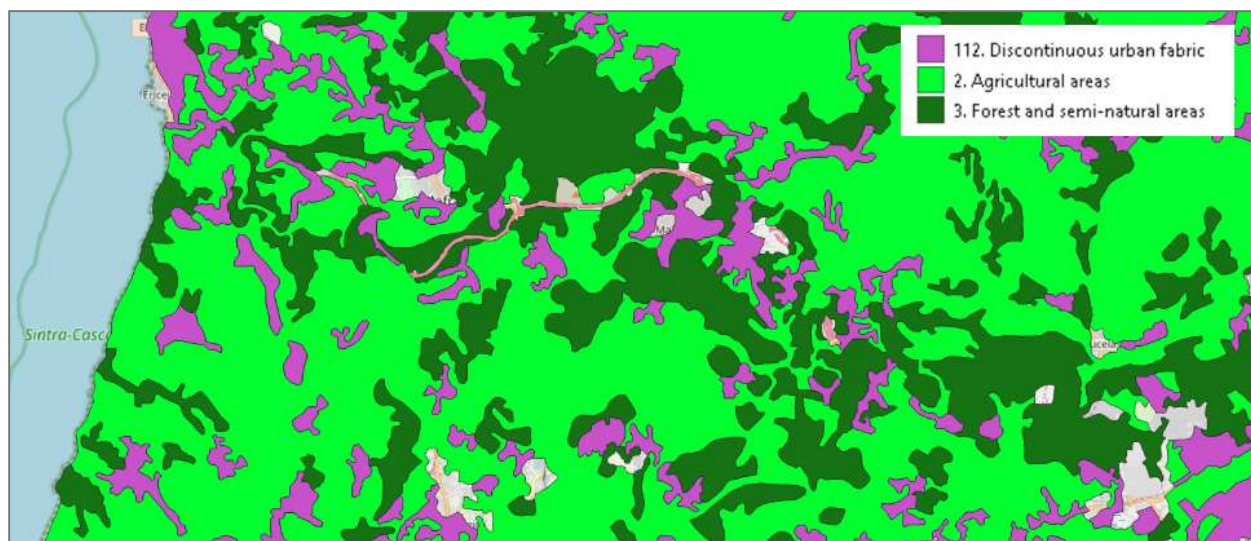


Figura 17: Exemplo da proximidade geográfica entre algumas classes.

Desta forma, partindo do conjunto de classes utilizado nos testes anteriores e considerando a semelhança observada entre algumas delas, optou-se, no primeiro teste do cenário 5, pela eliminação da classe “florestas e áreas seminaturais”, observando-se logo uma melhora nos resultados obtidos no teste do modelo. De forma semelhante, no segundo teste não foi considerada a classe “agricultura”, sendo que o modelo obtido ficou equiparado ao anterior. Nesta mesma direção, o terceiro teste foi realizado sem a classe “áreas urbanas contínuas”, tendo sido alcançado um resultado similar aos dois anteriores. Em contrapartida, de maneira a comprovar que a melhoria nos resultados obtidos nestes três testes não se deve apenas ao facto de os modelos terem sido criados com base em 4 classes, ao invés de 5, o quarto teste considerou as classes “áreas urbanas

contínuas”, “áreas urbanas descontínuas”, “agricultura” e “florestas e áreas seminaturais”. O resultado obtido suporta a ideia de que estas três últimas categorias de uso do solo são pouco distintas entre si, sendo, portanto, mais difíceis de classificar, se comparadas às demais.

Nota-se, desta forma, ao analisar os resultados dos três primeiros testes, que provavelmente devido às similaridades na localização geográfica tanto das áreas de agricultura, quanto das áreas de florestas e das zonas urbanas menos densas, os modelos criados apresentam menor eficácia na distinção dos polígonos que representam essas regiões. Sendo assim, eliminando-se qualquer uma das três classes, o resultado obtido é semelhante, sendo possível notar uma melhoria nos modelos em relação aqueles que utilizaram as 5 classes.

Com base nesta conclusão, foram realizados outros três testes, cada um com o objetivo de caracterizar o uso do solo em 3 diferentes classes. Primeiramente foram consideradas as classes que se mostraram mais distintas umas das outras. “áreas urbanas contínuas”, “agricultura” e “industrial, comercial e instalações públicas”. O resultado obtido claramente suporta a ideia da distinção existente entre essas três categorias. Sendo assim, de forma a tentar obter um modelo mais robusto e adequado para uso em termos práticos, optou-se por juntar as classes que possivelmente apresentariam alguma similaridade. Portanto, para o teste seguinte as classes “áreas urbanas contínuas” e “áreas urbanas descontínuas” foram fundidas numa só. Neste teste, foram ainda consideradas as classes “industrial, comercial e instalações públicas” e “agricultura”, resultando num modelo capaz de distinguir com relativa eficiência entre tais categorias de uso do solo.

Por fim, no último teste realizado nesta experiência, foram consideradas as classes “áreas urbanas contínuas e descontínuas” e “industrial, comercial e instalações públicas”. Foi também criada uma nova classe, na qual figuram juntas as categorias “agricultura” e “florestas e áreas seminaturais”. Essa experiência mostrou que é possível treinar um modelo baseado em rede neuronal, para distinguir com relativa eficácia entre essas três categorias de uso do solo.

De uma forma geral, ao comparar a metodologia adotada nas experiências realizadas com base no CORINE com aquela pela qual se optou nos testes baseados no COS, nota-se claramente as diferenças nos resultados. De uma forma geral, o método utilizado nesta secção provou ser mais eficiente em relação ao anterior. Alguns motivos podem ser apontados para esta maior eficiência, nomeadamente o facto de terem sido utilizadas os polígonos com tamanho uniforme, representando áreas geográficas de dimensões similares. Entretanto, verificou-se que o facto desta metodologia se traduzir num maior número de exemplos para criação dos modelos foi determinante para a melhoria nos resultados.

## 5.2 Análise das correlações

Como complementação ao trabalho apresentado na secção 5.1, foi realizado um estudo ainda mais aprofundado, envolvendo a análise das correlações existentes entre as várias categorias dos pontos de interesse e as diferentes classes de uso e ocupação dos espaços. O estudo procurou ainda, revelar padrões com base na existência de eventuais ligações entre as localizações geográficas dos pontos de interesse de cada categoria em relação às demais. Estas análises permitem responder questões como por exemplo “qual a intensidade da ligação entre cada tipo de uso dos espaços e cada categoria de empresa?”; “que tipos de pontos de interesse ou empresas, geralmente existem nas zonas mais centrais das cidades?”; “é comum a existência de restaurantes nas zonas residenciais menos densas?”; “quais são as categorias de pontos de interesse ou de empresas mais comuns nas áreas industriais?”. Também é possível perceber dúvidas como “existem categorias de pontos de interesse que geralmente são encontradas próximo dos supermercados?” ou “onde há igrejas, há também espaços comerciais?”.

Para a análise das correlações, optou-se pela utilização de mapas de densidade. A seguir é apresentado o método utilizado para a sua geração. São também descritos os softwares desenvolvidos para esta finalidade. Em seguida são abordadas as experiências realizadas, tal como os resultados obtidos e uma análise dos mesmos.

### 5.2.1 Mapas de densidade

A geração dos mapas de densidade passou essencialmente por duas etapas. A primeira delas envolveu a extração dos dados das tabelas do PostgreSQL, num formato que pudesse ser facilmente interpretado para a criação dos *heatmaps*, enquanto que a segunda foi a importação dos ficheiros gerados na etapa anterior e a geração dos *rasters* a partir de onde foi extraída a informação estudada. Desta forma, para tirar partido de uma característica importante das bibliotecas do QGIS, nomeadamente a capacidade de construir *layers* a partir de documentos de texto estruturado, foi criado um software para a geração de ficheiros em formato *CSV*. Os ficheiros criados são posteriormente importados para um segundo software, também desenvolvido no âmbito deste trabalho, que faz efetivamente a geração dos mapas de densidade. Esta aplicação informática é considerada mais importante que a primeira, tendo em vista que sem ela não teria sido possível a aplicação da metodologia utilizada.

Apesar do QGIS disponibilizar na sua interface gráfica uma funcionalidade de ligação direta à base de dados, a implementação da mesma numa aplicação *standalone* desenvolvida a partir das bibliotecas do software não seria uma tarefa trivial. Além disso considera-se que seria importante para possíveis replicações deste estudo, não obrigar a utilização de um sistema de gestão de base de dados nesta etapa. Em outras palavras, adotando-se o uso de ficheiros de texto como *input* para o software que faz a geração dos mapas de densidade, qualquer interessado poderia fazer uso do mesmo, sendo suficiente dispor apenas da latitude e da longitude de cada

ponto de interesse. Desta forma, acredita-se que neste sentido foi proporcionada uma certa flexibilidade metodológica, que se considera importante para um trabalho desta magnitude.

#### 5.2.1.1 Automação do processo de extração dos dados

Conforme introduzido no tópico anterior, o ponto de partida para a geração dos mapas de densidade foi, naturalmente, a seleção e a extração dos conjuntos de dados. Para permitir que os mesmos fossem criados de maneira automática, foi desenvolvido um software que faz a geração de ficheiros do tipo *CSV*, a partir de uma ligação à base de dados. Para tanto, recorreu-se ao uso do Java, tendo em conta que a tecnologia permite que se trabalhe com ficheiros de texto, de uma maneira relativamente fácil. Além disso, as aplicações desenvolvidas com recurso a esta linguagem de programação podem ser executadas em várias plataformas, incluindo Windows e Linux. Relativamente à manipulação dos dados, o Java permite o uso do JDBC (*Java Database Connectivity*)<sup>23</sup>, o que facilita as implementações. Além disso permite também o uso de encriptação SSL (*Secure Sockets Layer*)<sup>24</sup>, caso seja necessário, proporcionando desta forma mais segurança no que diz respeito às ligações. As características da tecnologia são apresentadas no tópico 3.6.2.

O software é, portanto, uma ferramenta de linha de comando, que realiza a pesquisa das categorias à base de dados e, para cada uma delas, vai buscar os pontos de interesse relacionados. Para a listagem das categorias, o software está desenvolvido para permitir o uso de diferentes “estratégias”, as quais podem ser facilmente implementadas de acordo com as eventuais necessidades. Relativamente aos pontos de interesse utilizados para a construção do ficheiro, a aplicação lista aqueles localizados dentro da zona metropolitana de Lisboa, tendo em conta que esta foi a região escolhida para o estudo. Além disso, para cada categoria, são ainda considerados os pontos de interesse vinculados também às suas subcategorias. Em termos de *output*, para cada categoria retornada na busca, é gerado um ficheiro com as coordenadas de todos os POIs que pertencem à mesma, respeitadas as devidas restrições. Esse ficheiro fica então disponível no diretório parametrizado, estando, a partir de então, pronto para ser utilizado na geração dos mapas de densidade.

#### 5.2.1.2 O software desenvolvido para criação dos mapas de densidade

Os mapas de densidade foram o recurso adotado para a representação das informações relacionadas à concentração dos pontos de interesse em cada região geográfica dentro da zona onde se focou este estudo. Considerando o grande número de categorias analisadas, de maneira a automatizar o processo de geração dos mesmos, foi então desenvolvido um software com recurso às bibliotecas do QGIS. Para tanto, recorreu-se ao uso do Python, tendo em conta que a API adotada<sup>25</sup> é

---

<sup>23</sup> <https://docs.oracle.com/javase/8/docs/technotes/guides/jdbc/> [último acesso: 11/03/2020]

<sup>24</sup> <https://www.openssl.org/docs/> [último acesso: 11/03/2020]

<sup>25</sup> <https://qgis.org/api/> [último acesso: 11/03/2020]

disponibilizada nesta linguagem. Trata-se, portanto, de um aplicativo de linha de comando, que gera os mapas de densidade de acordo com os dados de *input* carregados a partir de ficheiros de texto. O software está preparado para ser utilizado tanto no Windows, quanto no Linux, tendo sido testado no Windows Server 2016 e no Ubuntu 18.04.

Ao ser carregado, o aplicativo faz primeiramente a definição de todas as variáveis de ambiente necessárias ao seu funcionamento. Em seguida, realiza uma listagem dos ficheiros que contém dados de entrada a serem utilizados. Tais ficheiros devem obrigatoriamente conter a latitude e a longitude de cada ponto. O software realiza então a leitura dos dados existentes nos ficheiros. Para cada ficheiro lido, é criada em memória, uma *layer* contendo a informação geográfica dos pontos carregados. A partir desta camada, é gerado o mapa de concentração. Desta forma, cada *heatmap* gerado representa um determinado ficheiro de entrada. Sempre que a aplicação gera um mapa de densidade, realiza logo em seguida a gravação do mesmo em formato GeoTIFF<sup>26</sup>, no diretório parametrizado. Para além dos diretórios de entrada e saída de dados, o software permite ainda a definição dos mesmos parâmetros disponíveis quando se está a utilizar a funcionalidade de geração de mapas de concentração, por meio da interface gráfica do QGIS. Embora sejam vários estes parâmetros, os mais importantes são: o tamanho do pixel, o raio e o tipo da função de *kernel* a ser considerado nos cálculos.

### 5.2.1.3 Geração dos mapas de densidade

Nesta etapa, os mapas de densidade foram utilizados para a extração da informação estudada. Na sequência das experiências apresentadas na secção 5.1, organizou-se esta fase do projeto com base nos resultados obtidos até então. Conforme apresentado no tópico 5.1.3.5, o último teste realizado no quinto cenário das experiências de classificação de uso dos espaços, baseadas nas classes do CORINE, focou-se na identificação das zonas: (1) comerciais e industriais, incluindo instalações públicas, (2) áreas urbanas contínuas e descontínuas e (3) regiões de agricultura, florestas e áreas seminaturais. Foram gerados diversos atributos, com base nas categorias de nível 2, 3, 4 e 5, dos pontos de interesse. Dentre esses atributos, foram selecionados de maneira automática para a criação do modelo, aqueles que apresentaram maior relevância. Na primeira experiência desta etapa, de maneira a complementar o estudo anterior, a geração dos mapas de densidade teve por base uma lista de 68 categorias, nomeadamente aquelas a partir das quais foram extraídos os atributos utilizados anteriormente. Esta lista de categorias pode ser consultada no apêndice G.1. Para a preparação dos dados nesta etapa do trabalho, recorreu-se ao uso dos comandos SQL apresentados no apêndice A.15. De forma semelhante, numa segunda etapa das análises, foram utilizadas as subcategorias da categoria “Business”. Os motivos desta escolha são apresentados mais adiante.

---

<sup>26</sup> <https://gdal.org/drivers/raster/gtiff.html> [último acesso: 11/03/2020]

Identificadas as categorias dos pontos de interesse que seriam alvo das análises nesta etapa do estudo, e recorrendo-se ao uso do software desenvolvido, foram então extraídos os pontos de interesse, dando origem aos ficheiros *CSV* a serem utilizados. Com base neles, procedeu-se à geração dos mapas de densidade, através da ferramenta criada para esta finalidade. Nesta operação, após a realização de diversas provas de conceito, optou-se pela utilização de pixels que representam 50 metros quadrados, tendo em conta que uma resolução mais fina daria origem a *rasters* muito detalhados, ocupando desta forma demasiado espaço de armazenamento. Implicaria também num alto consumo de memória e num elevado custo de processamento, nas etapas seguintes. Em contrapartida, uma resolução mais baixa poderia não apresentar um grau relevante de informação para a análise das correlações. Relativamente ao tamanho do raio, foram realizados testes com 1250, 2500 e 5000 metros, de maneira a permitir uma análise da forma como os valores das correlações variam consoante a área de influência considerada para cada POI.

### 5.2.2 Experiências e resultados

Relativamente às classes de uso do solo, recorrendo-se aos *scripts* SQL apresentados no apêndice A.16, as mesmas foram primeiramente isoladas em tabelas separadas na base de dados. A seguir foi realizada uma operação no QGIS, para criar camadas em *raster* com base nos polígonos que delimitam as regiões onde predomina cada classe. Essas camadas foram posteriormente exportadas em formato GeoTIFF. Desta forma, cada *raster* gerado representa uma categoria de uso do solo. Foram considerados numa mesma *layer* os polígonos que representam a classe “1. Artificial surfaces” (nível 1). Em outra camada foram representados os polígonos da classe “1.1. Urban fabric” (nível 2). Foram ainda isoladas em *layers* separadas, as classes “111. Continuous urban fabric” e “112. Discontinuous urban fabric” (nível 3). Além disso, foi também gerado um *raster* que representa a classe “121. Industrial or commercial units and public facilities” (nível 3).

Os *rasters* gerados para representar as classes de uso de solo do CORINE, foram utilizados como parte dos dados de entrada para o passo seguinte do estudo. A outra parte dos *inputs* foram os mapas de concentração. Com base em cada categoria dos pontos de interesse, daquelas listadas no apêndice G.1, foi feita a geração do primeiro conjunto de *heatmaps*. De forma similar, foram gerados os mapas de densidade para as subcategorias da categoria “Business”. A partir dos dados de entrada, foram obtidas as informações empregadas nas análises apresentadas nos tópicos seguintes. Portanto, foram gerados 68 *heatmaps* com raio de 1250 metros, outros 68 com raio de 2500 metros e ainda mais 332 com raio de 5000 metros. Desses últimos 332, 68 representam as categorias selecionadas com base no estudo anterior, e 264 as subcategorias da categoria “Business”. De forma automatizada, foram calculadas as correlações entre cada mapa de densidade e cada categoria de uso do solo. De maneira similar e igualmente automática, foram também calculadas as correlações entre os *heatmaps* criados, resultando numa matriz que mostra a correlação de cada categoria com todas as outras.

### 5.2.2.1 Automação do cálculo das correlações

As correlações constituem a informação central no que diz respeito às análises realizadas nesta etapa do trabalho. Em termos gerais, foram utilizados *rasters* que representam 5 classes CORINE, além de 468 mapas de densidade. Portanto, relativamente às correlações entre os tipos de uso e cobertura do solo e a concentração dos pontos de interesse de cada categoria, foram calculadas 2340 correlações. Num processo similar, foram realizados também os cálculos para os *heatmaps* que representam subconjunto contendo 22 categorias, o que resulta numa matriz com um total de 484 correlações. Considerando que cada *raster*, seja dos que representam o uso do solo ou daqueles que representam a concentração de pontos de interesse, é constituído por um total de 4 849 343 pixels e tendo em conta que o cálculo da correlação é feito pixel a pixel, conclui-se facilmente que não é viável a realização manual desta tarefa. Portanto, recorreu-se ao uso da linguagem R para o cálculo das correlações de forma automatizada.

Em R, o cálculo das correlações a partir de *rasters* implica algumas exigências. Obriga por exemplo, que os *inputs* estejam no mesmo sistema de coordenadas<sup>27</sup> e que tenham o mesmo número de pixels, isto é, o número de colunas de ambos os *inputs* deve ser igual, tal como o número de linhas. Considerando que os *rasters* resultantes do QGIS ou da aplicação desenvolvida para geração dos mapas de densidade, embora tenham o mesmo sistema de coordenadas (*Coordinate Reference System* – CRS), possuem geralmente tamanhos diferentes, foi necessário padronizá-los, em termos de *extent*. Para isso, existem algumas alternativas. Uma delas seria implementar uma funcionalidade em Python utilizando o core do QGIS para a tarefa. Entretanto uma prova de conceito realizada revelou que esta reprojeção por meio do QGIS parece alterar o tamanho do pixel em alguns casos, pelo que se procurou outra alternativa. A solução encontrada passou pelo *resample*<sup>28</sup> por meio de algumas das funcionalidades disponíveis no R.

Além do *resample* dos dados de entrada, o código desenvolvido implementa diferentes versões, embora com o mesmo resultado, das funcionalidades que realizam o cálculo de dois tipos de correlação. A primeira delas é a correlação ponto-bisserial, utilizada para obter a correlação entre as classes de uso do solo e os mapas de densidade. Não muito diferente, a segunda é o próprio coeficiente de correlação de Pearson. A implementação funciona de maneira similar para os dois casos. Primeiramente, são carregados os dados de entrada, que são transformados em camadas e empilhados. No caso da correlação ponto-bisserial, cada camada de *land use* representa os dados binários, sendo que onde há predominância da classe, os pixels possuem o valor “1”. Em contrapartida, os mapas de densidade representam os valores contínuos. Para os dados de entrada, são, portanto, extraídos os valores de cada pixel. Em seguida são preenchidos com “0” os espaços correspondentes aos pixels que não possuem valor. São então calculadas as correlações entre cada camada binária e cada camada contínua. O resultado é gravado em um documento de texto em

---

<sup>27</sup> <http://tiny.cc/v747qz> [último acesso: 11/03/2020]

<sup>28</sup> <https://www.rdocumentation.org/packages/raster/versions/3.0-12/topics/resample> [último acesso: 11/03/2020]

formato CSV, no diretório parametrizado para as saídas. Quanto ao cálculo das correlações entre os mapas de densidade, numa das versões implementadas, é gerado um ficheiro para cada *heatmap*, que apresenta a correlação deste com todos os outros. Na outra versão, é gerado apenas um *output*, ou seja, um ficheiro contendo uma matriz que mostra a correlação entre todas as camadas envolvidas no cálculo.

#### 5.2.2.2 O uso e ocupação dos espaços e as concentrações dos pontos de interesse

Conforme introduzido nos tópicos anteriores, tendo sido gerados os ficheiros em formato *raster*, foram então calculadas as correlações. Nesta secção, são apresentados e discutidos os resultados alcançados no estudo onde foram analisadas as correlações existentes entre os pontos de interesse e as classes de uso do solo.

##### 5.2.2.2.1 Apresentação dos resultados

Nos tópicos seguintes são apresentados os resultados obtidos em duas experiências distintas. A diferença entre elas consiste nas categorias dos pontos de interesse que foram utilizadas. A primeira considerou um conjunto de 68 categorias de vários níveis, enquanto a segunda baseou-se nas subcategorias da categoria “Business”.

##### 5.2.2.2.1.1 Experiência preliminar

Partindo do nível mais alto dos tipos de uso e ocupação do solo escolhidos, as primeiras correlações extraídas foram, portanto, relativas à classe “1. Artificial surfaces”. A lista completa dos resultados obtidos para esta classe pode ser consultada no apêndice H.1.1.1. Ao analisar as correlações obtidas para os mapas de concentração gerados com um raio de 1250 metros, nota-se, conforme mostrado na Tabela 37, que as 5 maiores correlações encontradas variam de 0.409 a 0.490.

Tabela 37: Maiores correlações com a classe 1, do CORINE (raio: 1250m).

Classe de uso do solo -> Categoria	Correlação
Artificial surfaces -> Local service	0.490635
Artificial surfaces -> Sports & recreation	0.454211
Artificial surfaces -> Education	0.436287
Artificial surfaces -> Medical health	0.423117
Artificial surfaces -> Beauty cosmetic personal care	0.409037
Artificial surfaces -> Shopping retail	0.405836
Artificial surfaces -> Nonprofit organization	0.385625
Artificial surfaces -> Real estate	0.383834
Artificial surfaces -> Restaurant	0.381753
Artificial surfaces -> Beauty salon	0.376329

Considerando uma área de influência de 2500 metros para cada ponto na geração dos mapas de densidade, nota-se, ao realizar o cálculo das correlações, que os resultados obtidos variam em

diferentes aspetos, em relação àqueles apresentados acima. Conforme mostrado na Tabela 38, não somente são encontradas maiores correlações, como também a presença de categorias que antes não figuravam na lista. É o caso por exemplo, da categoria “Shopping & retail”. Outro detalhe intrigante, diz respeito à categoria “Education”, que agora apresenta maior correlação que a categoria “Sports & recreation”.

Tabela 38: Maiores correlações com a classe 1, do CORINE (raio: 2500m).

<b>Classe de uso do solo -&gt; Categoria</b>	<b>Correlação</b>
Artificial surfaces -> Local service	0.541463
Artificial surfaces -> Education	0.535278
Artificial surfaces -> Sports & recreation	0.532571
Artificial surfaces -> Medical health	0.519108
Artificial surfaces -> Shopping & retail	0.503866
Artificial surfaces -> Real estate	0.496376
Artificial surfaces -> Nonprofit organization	0.482451
Artificial surfaces -> Health beauty	0.479770
Artificial surfaces -> Beauty cosmetic personal care	0.473118
Artificial surfaces -> Business service	0.463478

Relativamente às correlações extraídas com base nos mapas de densidade gerados com um raio de 5000 metros, nota-se ainda mais variações, em comparação com os casos anteriores. Conforme mostrado na Tabela 39, a categoria “Education” figura no topo dos resultados. Observa-se também a presença de novas categorias, como por exemplo “Real estate” e “Health beauty”. Além disso, os dados mostram ainda que dentre as 10 categorias apresentadas, a menor correlação é equivalente a 0.53, um valor superior à grande maioria daqueles apresentados anteriormente.

Tabela 39: Maiores correlações com a classe 1, do CORINE (raio: 5000m).

<b>Classe de uso do solo -&gt; Categoria</b>	<b>Correlação</b>
Artificial surfaces -> Education	0.594404
Artificial surfaces -> Medical health	0.581258
Artificial surfaces -> Health beauty	0.566915
Artificial surfaces -> Sports & recreation	0.559584
Artificial surfaces -> Real estate	0.556196
Artificial surfaces -> Shopping retail	0.554263
Artificial surfaces -> Nonprofit organization	0.546400
Artificial surfaces -> Local service	0.543361
Artificial surfaces -> Business service	0.539077
Artificial surfaces -> Gym physical fitness center	0.530743

Posteriormente às análises apresentadas, foram calculadas as correlações para a classe “Urban fabric”. A lista completa dos resultados obtidos para esta classe pode ser consultada no

apêndice H.1.1.2. Na Tabela 40 são mostradas as correlações mais significativas que foram alcançados por meio da utilização dos *heatmaps* com raio de 1250 metros. Nota-se que as correlações obtidas foram todas inferiores a 0.5. Entretanto as categorias que figuram no topo da lista estavam também presentes entre aquelas já discutidas nos casos anteriores.

Tabela 40: Maiores correlações com a classe 1.1, do CORINE (raio: 1250m).

Classe de uso do solo -> Categoria	Correlação
Urban fabric -> Local service	0.492297
Urban fabric -> Beauty cosmetic personal care	0.433745
Urban fabric -> Education	0.425247
Urban fabric -> Medical health	0.419407
Urban fabric -> Sports & recreation	0.414012
Urban fabric -> Beauty salon	0.403519
Urban fabric -> Shopping retail	0.390312
Urban fabric -> Restaurant	0.390245
Urban fabric -> Real estate	0.384663
Urban fabric -> Nonprofit organization	0.380908

No que diz respeito aos resultados obtidos aquando da utilização dos mapas de concentração com raio de 2500 metros, é possível notar, de acordo com as informações apresentadas na Tabela 41, que ocorrem variações não apenas em relação ao valor das correlações, como também quanto às próprias categorias. Nota-se por exemplo a presença de várias outras categorias com maiores correlações que a categoria “Beauty, cosmetic & personal care”, que antes apresentava a segunda maior correlação dentre todas.

Tabela 41: Maiores correlações com a classe 1.1, do CORINE (raio: 2500m).

Classe de uso do solo -> Categoria	Correlação
Urban fabric -> Local service	0.511213
Urban fabric -> Education	0.496747
Urban fabric -> Medical health	0.481450
Urban fabric -> Sports & recreation	0.472074
Urban fabric -> Real estate	0.471171
Urban fabric -> Shopping retail	0.461363
Urban fabric -> Beauty cosmetic personal care	0.461298
Urban fabric -> Nonprofit organization	0.449480
Urban fabric -> School	0.437897
Urban fabric -> Beauty salon	0.434780

Diferentemente do resultado obtido com a classe “1. Artificial surfaces”, nos cálculos baseados nos mapas de concentração com raio de 5000 metros, nota-se que a classe “1.1 Urban fabric” apresenta correlações superiores a 0.5 apenas em relação às categorias “Education” e

“Medical health”. Conforme mostrado na Tabela 42, as 5 classes que apresentaram maiores correlações são as mesmas do caso anterior, entretanto em ordens diferentes.

Tabela 42: Maiores correlações com a classe 1.1, do CORINE (raio: 5000m).

<b>Classe de uso do solo -&gt; Categoria</b>	<b>Correlação</b>
Urban fabric -> Education	0.516448
Urban fabric -> Medical health	0.504722
Urban fabric -> Real estate	0.498155
Urban fabric -> Local service	0.485274
Urban fabric -> Sports & recreation	0.484821
Urban fabric -> Shopping retail	0.479884
Urban fabric -> Health beauty	0.476286
Urban fabric -> Business service	0.476275
Urban fabric -> Nonprofit organization	0.470295
Urban fabric -> Event planner	0.467368

Em relação à classe “111. Continuous urban fabric”, os resultados obtidos podem ser consultados, na íntegra, no apêndice H.1.1.3. A seguir, na Tabela 43 são listadas as correlações mais significativas com base nos mapas de concentração com raio de 1250 metros. Nota-se a presença de algumas categorias que não figuravam nos resultados dos testes anteriores, como por exemplo “Arts & entertainment” e “Tattoo & piercing shop”.

Tabela 43: Maiores correlações com a classe 111, do CORINE (raio: 1250m).

<b>Classe de uso do solo -&gt; Categoria</b>	<b>Correlação</b>
Continuous urban fabric -> Health beauty	0.336645
Continuous urban fabric -> Beauty cosmetic personal care	0.334934
Continuous urban fabric -> Government organization	0.321434
Continuous urban fabric -> Education	0.314048
Continuous urban fabric -> Nonprofit organization	0.311784
Continuous urban fabric -> Beauty salon	0.303882
Continuous urban fabric -> Arts entertainment	0.300491
Continuous urban fabric -> Restaurant	0.297522
Continuous urban fabric -> Medical health	0.295117
Continuous urban fabric -> Tattoo piercing shop	0.293407

A Tabela 44 contém as informações referentes aos resultados obtidos com base na utilização de mapas de concentração com raio de 2500 metros. Em relação ao caso anterior, é possível notar que embora no geral as correlações obtidas sejam maiores, algumas categorias apresentam uma realidade contrária. É o caso por exemplo da categoria “Beauty, cosmetic & personal care”. Esta tendência já havia sido verificada para a mesma categoria nos resultados das correlações obtidas para a classe “1.1. Urban fabric”.

Tabela 44: Maiores correlações com a classe 111, do CORINE (raio: 2500m).

Classe de uso do solo -> Categoria	Correlação
Continuous urban fabric -> Government organization	0.388695
Continuous urban fabric -> Health beauty	0.377241
Continuous urban fabric -> Nonprofit organization	0.370761
Continuous urban fabric -> Education	0.365230
Continuous urban fabric -> Arts entertainment	0.352019
Continuous urban fabric -> Tattoo piercing shop	0.335994
Continuous urban fabric -> Medical health	0.335982
Continuous urban fabric -> Real estate	0.331566
Continuous urban fabric -> Beauty cosmetic personal care	0.320673
Continuous urban fabric -> Restaurant	0.301355

Considerando um raio de 5000 metros para a geração dos mapas de concentração, as maiores correlações obtidas variam entre 0.29 e 0.38, conforme dados da Tabela 45. Contudo, as 5 categorias que alcançaram o topo da lista são as mesmas do caso anterior. Observa-se ainda que a ordem em que tais categorias figuram, não apresentou nenhuma alteração, contrariamente ao que se vê em vários dos casos já abordados.

Tabela 45: Maiores correlações com a classe 111, do CORINE (raio: 5000m).

Classe de uso do solo -> Categoria	Correlação
Continuous urban fabric -> Government organization	0.387747
Continuous urban fabric -> Health beauty	0.362785
Continuous urban fabric -> Nonprofit organization	0.357865
Continuous urban fabric -> Education	0.351652
Continuous urban fabric -> Arts entertainment	0.349422
Continuous urban fabric -> Medical health	0.322424
Continuous urban fabric -> Real estate	0.320566
Continuous urban fabric -> Tattoo piercing shop	0.309617
Continuous urban fabric -> Gym physical fitness center	0.305373
Continuous urban fabric -> Accessories	0.290454

Ao realizar-se as experiências para a classe “112. Discontinuous urban fabric”, foram obtidas as correlações apresentadas no apêndice H.1.1.4. Para os mapas de concentração criados com raio de 1250 metros, os resultados mais significativos são apresentados a seguir, na Tabela 46. Em comparação com as principais correlações obtidas para a classe “111. Continuous urban fabric” aquando da utilização dos *heatmaps* com a mesma medida de raio, nota-se, que apenas 3 categorias são comuns, nomeadamente: “Beauty, cosmetic & personal care”, “Education” e “Medical health”.

Tabela 46: Maiores correlações com a classe 112, do CORINE (raio: 1250m).

<b>Classe de uso do solo -&gt; Categoria</b>	<b>Correlação</b>
Discontinuous urban fabric -> Local service	0.418331
Discontinuous urban fabric -> Sports & recreation	0.319053
Discontinuous urban fabric -> Medical health	0.308090
Discontinuous urban fabric -> Education	0.303851
Discontinuous urban fabric -> Beauty cosmetic personal care	0.301516
Discontinuous urban fabric -> Shopping retail	0.292252
Discontinuous urban fabric -> Business service	0.288893
Discontinuous urban fabric -> Event planner	0.287297
Discontinuous urban fabric -> Beauty salon	0.284942
Discontinuous urban fabric -> School	0.283388

Relativamente aos resultados obtidos com base nos mapas de concentração gerados com raio de 2500 metros, é possível observar, recorrendo-se às informações da Tabela 47, o mesmo padrão encontrado no caso anterior. Ao se comparar os resultados com aqueles obtidos para a classe “111. Continuous urban fabric” com base nos mesmos *heatmaps*, nota-se que 3, das 10 categorias que figuram entre aquelas com maior correlação, aparecem em ambos os casos. São elas: "Education", "Medical health" e "Real estate".

Tabela 47: Maiores correlações com a classe 112, do CORINE (raio: 2500m).

<b>Classe de uso do solo -&gt; Categoria</b>	<b>Correlação</b>
Discontinuous urban fabric -> Local service	0.437754
Discontinuous urban fabric -> Sports & recreation	0.376771
Discontinuous urban fabric -> Event planner	0.369952
Discontinuous urban fabric -> Education	0.355873
Discontinuous urban fabric -> Medical health	0.355272
Discontinuous urban fabric -> Shopping retail	0.352931
Discontinuous urban fabric -> Business service	0.348352
Discontinuous urban fabric -> School	0.347112
Discontinuous urban fabric -> Real estate	0.346100
Discontinuous urban fabric -> Pet service	0.341191

A análise dos mapas de concentração com raio de 5000 metros reforçou ainda mais os indícios já observados anteriormente. Com base nos dados apresentados na Tabela 48, nota-se mais uma vez, que 80% das categorias são diferentes daquelas que figuram entre as 10 que apresentaram maiores correlações com a classe “111. Continuous urban fabric”. Neste caso, as únicas categorias em comum são “Education” e “Medical health”.

Tabela 48: Maiores correlações com a classe 112, do CORINE (raio: 5000m).

<b>Classe de uso do solo -&gt; Categoria</b>	<b>Correlação</b>
Discontinuous urban fabric -> Local service	0.426830
Discontinuous urban fabric -> Event planner	0.407672
Discontinuous urban fabric -> Sports & recreation	0.403159
Discontinuous urban fabric -> Business service	0.392970
Discontinuous urban fabric -> Medical health	0.389592
Discontinuous urban fabric -> Education	0.386135
Discontinuous urban fabric -> Shopping retail	0.385060
Discontinuous urban fabric -> Real estate	0.383178
Discontinuous urban fabric -> School	0.379443
Discontinuous urban fabric -> Pet service	0.378690

A classe “121. Industrial or comercial units and public facilities” foi a última envolvida nas experiências. Os resultados obtidos são listados integralmente no apêndice H.1.1.5. Primeiramente foram realizados os cálculos com base em mapas de concentração com raio de 1250 metros. Conforme informações apresentadas na Tabela 49, as correlações obtidas para esta classe foram muito menores, comparativamente com as demais. Além disso, entre as categorias com maior correlação, nota-se a presença de 2 novas, nomeadamente “Industrial or commercial” e “Retail company”.

Tabela 49: Maiores correlações com a classe 121, do CORINE (raio: 1250m).

<b>Classe de uso do solo -&gt; Categoria</b>	<b>Correlação</b>
Industrial commercial -> Gym physical fitness center	0.147872
Industrial commercial -> Shopping & retail	0.138123
Industrial commercial -> Industrial or commercial	0.137897
Industrial commercial -> Retail company	0.126261
Industrial commercial -> Sports & recreation	0.123912
Industrial commercial -> Local service	0.122221
Industrial commercial -> Business service	0.115872
Industrial commercial -> Health beauty	0.106894
Industrial commercial -> Education	0.106472
Industrial commercial -> Pet store	0.105676

As informações da Tabela 50 representam os resultados mais relevantes dentre aqueles verificados na experiência baseada nos *heatmaps* com raio de 2500 metros. Nota-se, que embora pouco superiores aos valores apresentados anteriormente, as correlações obtidas ainda foram consideravelmente baixas, num contexto geral. Em relação aos resultados da tabela anterior, verifica-se que das 10 categorias apresentadas, 9 são iguais. A categoria “Medical health” é a única que figura apenas na lista atual.

Tabela 50: Maiores correlações com a classe 121, do CORINE (raio: 2500m).

Classe de uso do solo -> Categoria	Correlação
Industrial commercial -> Gym physical fitness center	0.186125
Industrial commercial -> Shopping & retail	0.180924
Industrial commercial -> Health beauty	0.178808
Industrial commercial -> Local service	0.163768
Industrial commercial -> Pet store	0.161306
Industrial commercial -> Sports recreation	0.156006
Industrial commercial -> Business service	0.155598
Industrial commercial -> Retail company	0.154985
Industrial commercial -> Medical health	0.152495
Industrial commercial -> Commercial industrial	0.152279

No que diz respeito aos mapas de concentração com raio de 5000 metros, nota-se, de acordo com os resultados mostrados na Tabela 51, que as categorias "Beauty cosmetic personal care", "Beauty salon", "Education" e "Tattoo piercing shop" são aquelas que não aparecem nos resultados obtidos para os *heatmaps* com raio de 2500 metros. Além disso, verifica-se a presença da categoria "Education", diferentemente dos casos anteriores.

Tabela 51: Maiores correlações com a classe 121, do CORINE (raio: 5000m).

Classe de uso do solo -> Categoria	Correlação
Industrial commercial -> Health beauty	0.223848
Industrial commercial -> Gym physical fitness center	0.214244
Industrial commercial -> Shopping & retail	0.208365
Industrial commercial -> Medical health	0.192464
Industrial commercial -> Education	0.189012
Industrial commercial -> Beauty cosmetic personal care	0.185460
Industrial commercial -> Tattoo piercing shop	0.183696
Industrial commercial -> Beauty salon	0.181789
Industrial commercial -> Local service	0.181151
Industrial commercial -> Sports recreation	0.176009

#### 5.2.2.2.1.2 Análise das subcategorias da categoria "Business"

Ao analisar os mapas de concentração com raio de 5000 metros, tendo em conta que são os que em geral apresentam resultados mais significativos, verifica-se um padrão relativamente importante. Considerando um conjunto de 50 categorias, ou seja, as 10 que obtiveram maiores correlações com cada classe de uso do solo, nota-se que 45 delas pertencem à categoria "Business". Desta forma, tendo em conta os indícios da relevância desta categoria, optou-se por aprofundar o estudo, analisando mais 22 subcategorias a nível 2 da categoria "Business" e outras 242 a nível 3, sabendo-se que neste último caso eram 386 no total, e que, entretanto, foram consideradas apenas aquelas que tinham pontos de interesse na área estudada.

Conforme introduzido anteriormente, na etapa preliminar desta análise verificou-se que os mapas de concentração com raio de 1250 e 2500 metros, no geral, apresentam resultados menos expressivos. Além disso, considerando também os recursos computacionais disponíveis, optou-se por utilizar somente os *heatmaps* com raio de 5000 metros, nesta fase do estudo.

No que se refere às categorias de nível 2, tratando-se de um número mais reduzido, os resultados obtidos são apresentados, na íntegra, na Tabela 52. Para uma melhor visualização, foram realizados alguns ajustes na intensidade das cores, de maneira a destacar os valores mais significativos. Nota-se por exemplo, que 11, das 22 categorias analisadas, apresentam uma correlação superior a 0,5 com a classe “1. Artificial surfaces”, incluindo as categorias “Automotive aircraft boat” e “Public government service”, que não constavam no conjunto de categorias utilizado na experiência anterior.

Tabela 52: Correlações entre as classes de LULC e as subcategorias da categoria Business.

	<b>Artificial surfaces</b>	<b>Urban fabric</b>	<b>Cont. urban fabric</b>	<b>Disc. urban fabric</b>	<b>Ind./commercial</b>
Advertising marketing	<b>0.45586</b>	<b>0.39835</b>	0.18764	<b>0.34594</b>	0.15564
Agriculture	0.16900	0.16328	0.01871	0.17528	0.04590
Arts entertainment	<b>0.52657</b>	<b>0.45675</b>	<b>0.34942</b>	<b>0.31939</b>	0.14973
Automotive aircraft boat	<b>0.50288</b>	<b>0.41717</b>	0.18010	<b>0.37172</b>	0.23576
Beauty cosmetic personal care	<b>0.50700</b>	<b>0.45150</b>	0.25630	<b>0.36700</b>	0.18546
Commercial industrial	<b>0.30676</b>	0.25873	0.06915	0.25503	0.13384
Education	<b>0.59440</b>	<b>0.51645</b>	<b>0.35165</b>	<b>0.38613</b>	0.18901
Finance	<b>0.44887</b>	<b>0.40691</b>	0.20078	<b>0.34814</b>	0.12526
Food beverage	<b>0.49620</b>	<b>0.45350</b>	0.26067	<b>0.36676</b>	0.13942
Hotel lodging	0.21258	0.21661	0.06770	0.20787	0.00946
Legal	0.28774	0.21187	0.15025	0.15496	0.09990
Local service	<b>0.54336</b>	<b>0.48527</b>	0.21919	<b>0.42683</b>	0.18115
Media news company	<b>0.46487</b>	<b>0.38315</b>	<b>0.32191</b>	0.25136	0.12903
Medical health	<b>0.58126</b>	<b>0.50472</b>	<b>0.32242</b>	<b>0.38959</b>	0.19246
Non-governmental organization	<b>0.35316</b>	0.29205	0.20495	0.21485	0.10099
Nonprofit organization	<b>0.54640</b>	<b>0.47030</b>	<b>0.35786</b>	<b>0.32997</b>	0.15826
Public government service	<b>0.52534</b>	<b>0.44272</b>	<b>0.34724</b>	<b>0.30467</b>	0.15481
Real estate	<b>0.55620</b>	<b>0.49816</b>	<b>0.32057</b>	<b>0.38318</b>	0.17411
Science technology engineering	<b>0.42773</b>	<b>0.35700</b>	0.20533	0.28864	0.18521
Shopping retail	<b>0.55426</b>	<b>0.47988</b>	0.28111	<b>0.38506</b>	0.20836
Sports recreation	<b>0.55958</b>	<b>0.48482</b>	0.25944	<b>0.40316</b>	0.17601
Travel transportation	<b>0.49223</b>	<b>0.41366</b>	0.24661	<b>0.32946</b>	0.14457

Relativamente ao nível 3 das categorias, os resultados são apresentados na íntegra, no apêndice H.1.2. Primeiramente foram analisadas as correlações obtidas para a classe “1. Artificial surfaces”, cujo resultado pode ser visualizado na Tabela 53. Nas informações apresentadas verifica-se a presença de 5 categorias com correlação superior a 0.5. Nota-se ainda que dentre elas, as categorias “Sports & recreation venue” e “School”, não aparecem no estudo preliminar. Entretanto os pontos de interesse ligados às mesmas estão incluídos, no caso da primeira, na categoria “Sports & recreation” e no caso da segunda, na categoria “Education”.

Tabela 53: Maiores correlações entre a classe 1, do CORINE, e as categorias.

<b>Classe de uso do solo -&gt; Categoria</b>	<b>Correlação</b>
Artificial surfaces -> Sports recreation venue	0.552843
Artificial surfaces -> Business service	0.539077
Artificial surfaces -> School	0.521566
Artificial surfaces -> Event planner	0.517203
Artificial surfaces -> Shopping mall	0.503175
Artificial surfaces -> Real estate agent	0.491157
Artificial surfaces -> Restaurant	0.489071
Artificial surfaces -> Medical center	0.483880
Artificial surfaces -> Beauty salon	0.483867
Artificial surfaces -> Apparel clothing	0.477751

No que diz respeito à classe “1.1. Urban fabric”, as principais correlações obtidas são apresentadas a seguir, na Tabela 54. Dentre os resultados, talvez o maior destaque esteja relacionado à categoria “Photography videography”. Entretanto, tal como nos resultados anteriores, trata-se de uma subcategoria da categoria “Local service”. É também o caso da subcategoria “Shopping mail”, que pertence à categoria “Shopping & retail”.

Tabela 54: Maiores correlações entre a classe 1.1, do CORINE, e as categorias.

<b>Classe de uso do solo -&gt; Categoria</b>	<b>Correlação</b>
urban fabric -> Business service	0.476275
urban fabric -> Sports recreation venue	0.469036
urban fabric -> Event planner	0.467368
urban fabric -> School	0.466534
urban fabric -> Restaurant	0.448290
urban fabric -> Real estate agent	0.437189
urban fabric -> Beauty salon	0.431286
urban fabric -> Pet service	0.427631
urban fabric -> Photography videography	0.424154
urban fabric -> Shopping mall	0.414636

Na sequência das análises apresentadas, foram calculadas as correlações para a classe “111. Continuous urban fabric”, cujo resultado pode ser apreciado na Tabela 55. Com base nas informações apresentadas, nota-se a presença de algumas novas categorias, como por exemplo “Cultural center”, “Doctor”, “Dance night club”, “Mobile phoneshop”. Contudo, de forma similar aos casos anteriores, tratam-se de subcategorias de outras já discutidas.

Tabela 55: Maiores correlações entre a classe 111, do CORINE, e as categorias.

<b>Classe de uso do solo -&gt; Categoria</b>	<b>Correlação</b>
Continuous urban fabric -> Specialty school	0.346595
Continuous urban fabric -> Performance event venue	0.345985
Continuous urban fabric -> Cultural center	0.331412
Continuous urban fabric -> Tattoo piercing shop	0.309617
Continuous urban fabric -> Shopping mall	0.306724
Continuous urban fabric -> Real estate service	0.293065
Continuous urban fabric -> Doctor	0.289052
Continuous urban fabric -> Dance night club	0.288269
Continuous urban fabric -> Mobile phone shop	0.283756
Continuous urban fabric -> Real estate agent	0.282763

Relativamente à classe “112. Discontinuous urban fabric”, os resultados apresentados na Tabela 56, mostram também a presença de algumas subcategorias das categorias já mencionadas, sendo este o caso das duas últimas, “Home garden store” e “Motor vehicle company”. Entretanto talvez um dos principais pontos de atenção neste caso, esteja relacionado à comparação com os resultados obtidos em relação à classe anterior. Desta forma, tratando-se de duas classes de uso do solo que apresentam bastante distinção entre si é possível notar uma clara diferença entre as categorias que figuram em cada um dos casos.

Tabela 56: Maiores correlações entre a classe 112, do CORINE, e as categorias.

<b>Classe de uso do solo -&gt; Categoria</b>	<b>Correlação</b>
Discontinuous urban fabric -> Event planner	0.407672
Discontinuous urban fabric -> Business service	0.392970
Discontinuous urban fabric -> School	0.379443
Discontinuous urban fabric -> Pet service	0.378690
Discontinuous urban fabric -> Sports recreation venue	0.378632
Discontinuous urban fabric -> Photography videography	0.359445
Discontinuous urban fabric -> Restaurant	0.357109
Discontinuous urban fabric -> Beauty salon	0.354404
Discontinuous urban fabric -> Home garden store	0.350968
Discontinuous urban fabric -> Motor vehicle company	0.342588

Por fim, as correlações apresentadas na Tabela 57 foram obtidas com base na classe “121. Industrial or commercial units and public facilities”. Neste caso, nota-se no topo da lista a presença de 3 categorias, nomeadamente: “Automotive service”, “Automotive dealership” e “Automotive store”. Verifica-se que são, portanto, subcategorias da categoria “Automotive, Aircraft & Boat”, que não figura entre as 68 utilizadas no estudo preliminar.

Tabela 57: Maiores correlações entre a classe 121, do CORINE, e as categorias.

Classe de uso do solo -> Categoria	Correlação
Industrial commercial -> Automotive service	0.220140
Industrial commercial -> Automotive dealership	0.215573
Industrial commercial -> Automotive store	0.196853
Industrial commercial -> Sports recreation venue	0.192609
Industrial commercial -> Apparel clothing	0.189551
Industrial commercial -> Tattoo piercing shop	0.183696
Industrial commercial -> Beauty salon	0.181789
Industrial commercial -> Shopping mall	0.178505
Industrial commercial -> Party entertainment service	0.178106
Industrial commercial -> Transportation service	0.176565

#### 5.2.2.2.2 Considerações acerca das experiências

No tópico anterior foram realizados dois estudos. Inicialmente foram apresentados os resultados alcançados nas experiências realizadas com o objetivo de obter as correlações existentes entre as várias classes de uso e ocupação do solo e as categorias dos pontos de interesse. Foram escolhidas 5 classes do *dataset* CORINE *Land Cover*: (1) “1. Artificial surfaces”, (2) “1.1. Urban fabric”, (3) “111. Continuous urban fabric”, (4) “112. Discontinuous urban fabric” e (5) “121. Industrial or commercial units and public facilities”. Para cada uma das 68 categorias de pontos de interesse selecionadas, foram gerados de maneira automática, mapas de concentração com raios de 1250, 2500 e 5000 metros. As correlações entre os *heatmaps* e as classes escolhidas foram calculadas programaticamente. O segundo estudo realizado foi algo similar ao primeiro, apresentando diferenças principalmente a nível das categorias selecionadas. Nele foram consideradas 22 subcategorias, a nível 2, da categoria “Business” e outras 242, a nível 3. Os mapas de concentração utilizados foram gerados com raio de 5000 metros.

No que diz respeito ao estudo preliminar, em relação à classe “1. Artificial surfaces”, observa-se que as categorias “Education”, “Medical health” e “Sports & recreation” figuram sempre entre as 5 com maiores correlações. Ao analisar os resultados de maneira mais pormenorizada, verifica-se que para as três categorias, a medida que o raio do *heatmap* aumenta, os valores das correlações também aumentam. Porém nota-se uma realidade diferente relativamente à categoria “Local service”, cuja correlação diminui no caso dos mapas de concentração com raio de 5000 metros. Como mostrado na Figura 18, recorrendo-se a uma análise

exploratória, é possível verificar que para esta categoria, quando é considerada uma maior área de influência para os pontos, formam-se no mapa de concentração, regiões de grande densidade fora das zonas onde há predominância da referida classe, o que poderá ser um dos motivos que levam a correlação a diminuir com o aumento do raio.

Em outras palavras, como se pode verificar na dinâmica das cidades, alguns segmentos de empresas ou serviços, atendem zonas mais amplas do território, enquanto outros servem somente à comunidade local, um bairro ou uma vizinhança. Um exemplo disso são os hospitais, que prestam muitos atendimentos aos utentes que vivem em localidades mais distantes. O mesmo se observa relativamente às escolas e universidades. Em contrapartida, um serviço local geralmente visa atender apenas aos residentes das áreas mais próximas. Desta forma, o que se verifica em relação às categorias “Education”, “Medical health” e “Local service” está perfeitamente alinhado com aquilo que se observa nos espaços urbanos modernos. Enquanto os pontos de interesse das duas primeiras se concentram em zonas mais centrais (como se pode ver na Figura 19), os da última tendem a se localizar, na maioria dos casos, nas extremidades ou mesmo fora das zonas representadas pela classe “1. Artificial surfaces”, provocando o efeito que se nota na Figura 18, onde são mostrados os polígonos desta classe (em verde), sobrepostos aos mapas de concentração que representam os pontos de interesse da categoria “Local service”. Na imagem do lado esquerdo é mostrado um mapa de concentração com raio de 1250 metros, enquanto que no *heatmap* do lado direito, o raio é de 5000 metros. É importante destacar que nos mapas de concentração mostrados, as áreas mais claras representam as zonas de maior densidade.

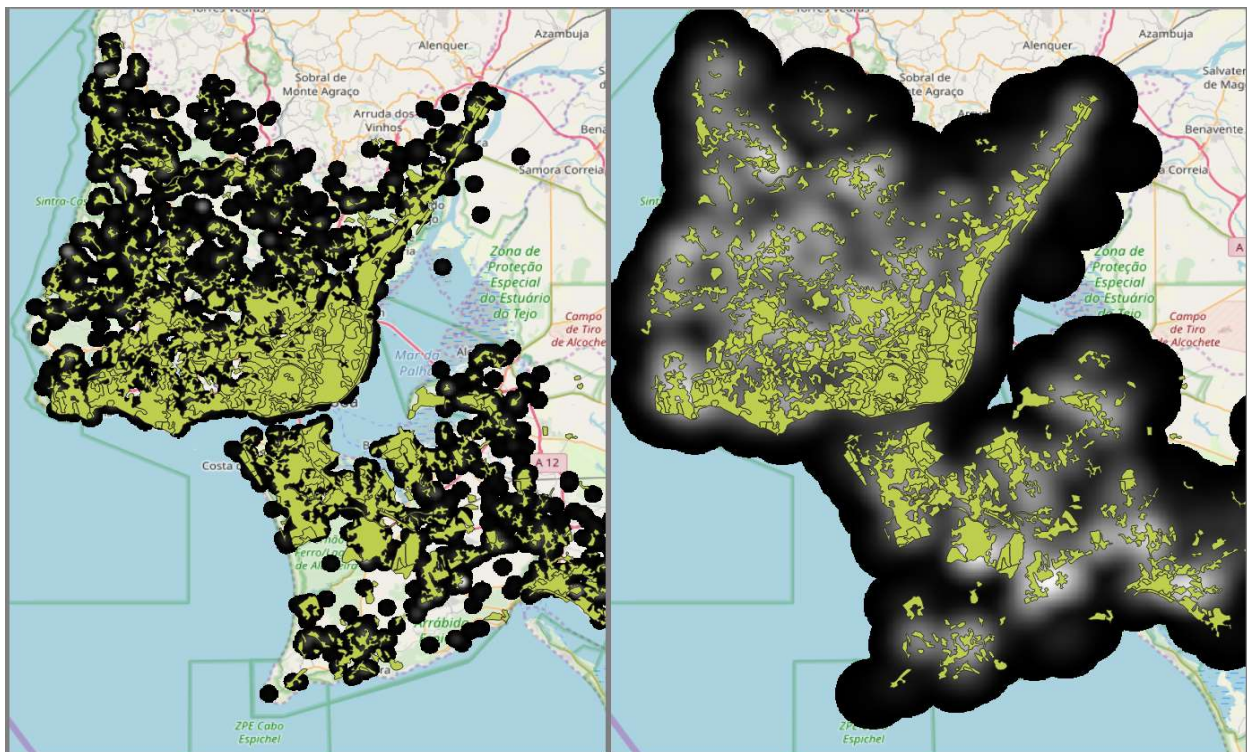


Figura 18: Comparação entre *heatmaps* com raios diferentes.

Relativamente à classe “1.1. Urban fabric”, observa-se a mesma tendência para a categoria “Local service”. Em outras palavras, comparando os resultados obtidos com os *heatmaps* que possuem raio de 2500 metros, aos alcançados aquando da utilização daqueles com raio de 1250 metros, nota-se um aumento do valor da correlação. Entretanto observa-se que a correlação obtida com um raio de 5000 metros é menor que nos outros dois casos. Considerando que a classe “1. Artificial surfaces” é composta em sua maioria pelos polígonos da classe “1.1 Urban fabric”, pode-se assumir que um provável motivo para o ocorrido seja o que foi apresentado anteriormente. O facto de uma das classes incluir todas as zonas representadas pela outra, faz também com se verifique mais um padrão: a presença das categorias “Education”, “Medical health” e “Sports & recreation” entre aquelas com maiores correlações. Como se pode ver na Figura 19, os pontos de interesse das três categorias localizam-se, em sua maioria, nas zonas representadas pelas duas classes, o que explica os valores obtidos para as correlações nos dois casos. Na imagem, os pontos de interesse da categoria “Medical health” são mostrados em amarelo, os da categoria “Education”, em vermelho e os da categoria “Sports & recreation”, em azul. Do lado esquerdo, vê-se os polígonos que representam a classe “1. Artificial areas” e do lado direito, os da classe “1.1. Urban fabric”.

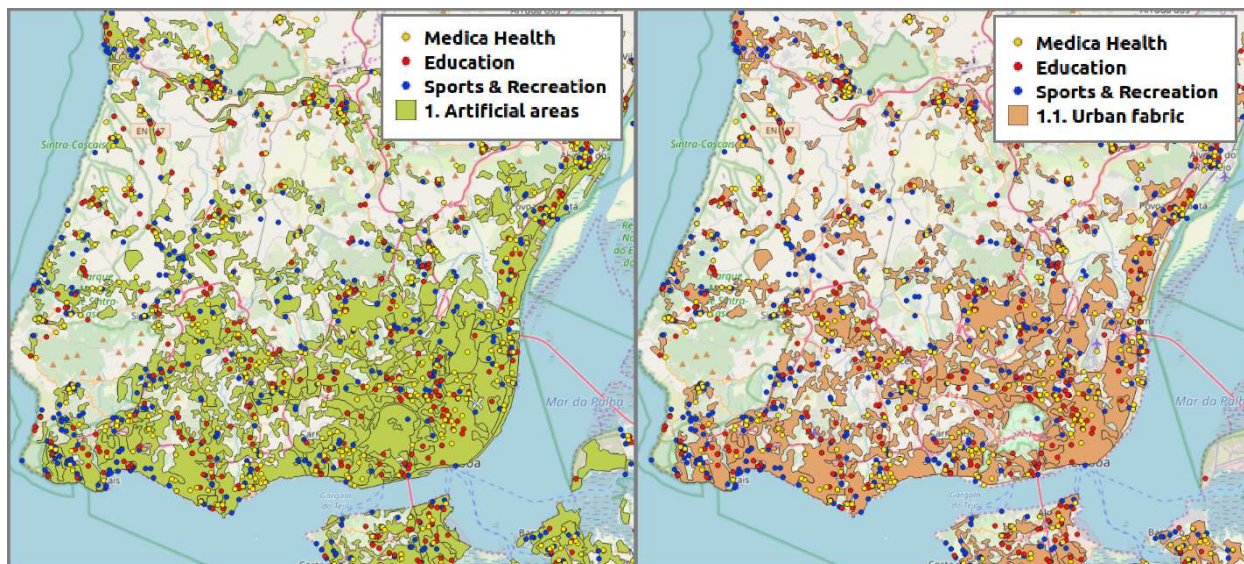


Figura 19: POIs de algumas categorias em relação a determinadas classes.

Considerando um nível mais baixo na hierarquia dos tipos de uso do solo presente nos dados do CORINE, foram realizadas as mesmas experiências para a classe “111. Continuous urban fabric”, que representa as zonas urbanas mais densas. Tal como nos resultados obtidos anteriormente, nota-se, independentemente do raio, a presença da categoria “Education” entre aquelas que apresentam maior correlação. O mesmo acontece com a classe “112. Discontinuous urban fabric” (zonas artificiais menos densas), no que diz respeito às categorias “Medical health” e “Sports & recreation”. Esta observação justifica plenamente o motivo destas três categorias figurarem entre aquelas com maiores correlações no caso da classe “1.1. Urban fabric” e

consequentemente “1. Artificial surfaces”. Além disso, outra observação importante apoia-se nas diferenças dos valores das correlações. Considerando que no nível mais baixo na hierarquia figuram as classes mais específicas, observa-se, provavelmente devido ao tamanho mais reduzido dos polígonos, que as correlações apresentadas diminuem conforme o grau de detalhe do uso e ocupação do solo aumenta.

Em relação às experiências realizadas com a classe “121. Industrial or commercial units and public facilities”, observa-se que no geral os valores das correlações são bastante inferiores àqueles obtidos nos casos anteriores. Nota-se a presença da categoria “Shopping & retail” entre os 10 principais resultados, independentemente do raio considerado. É possível verificar a presença de algumas das categorias já discutidas, como por exemplo a categoria “Education”, “Medical health” e “Sports & recreation”, embora não em todos os casos. Outras categorias verificadas são por exemplo “Gym physical fitness center”, “Commercial & industrial” e “Local service”. Como se pode verificar na Figura 20, os polígonos da referida classe representam áreas muito menores se comparados aos daquelas abordadas anteriormente. Nota-se com clareza que, apesar de existirem áreas onde ocorram grandes concentrações de pontos de interesse das categorias mencionadas, dentro dos limites da classe em questão, há também a presença deles em regiões onde não se verifica a predominância da classe. Esta análise visual justifica, portanto, as baixas correlações obtidas.

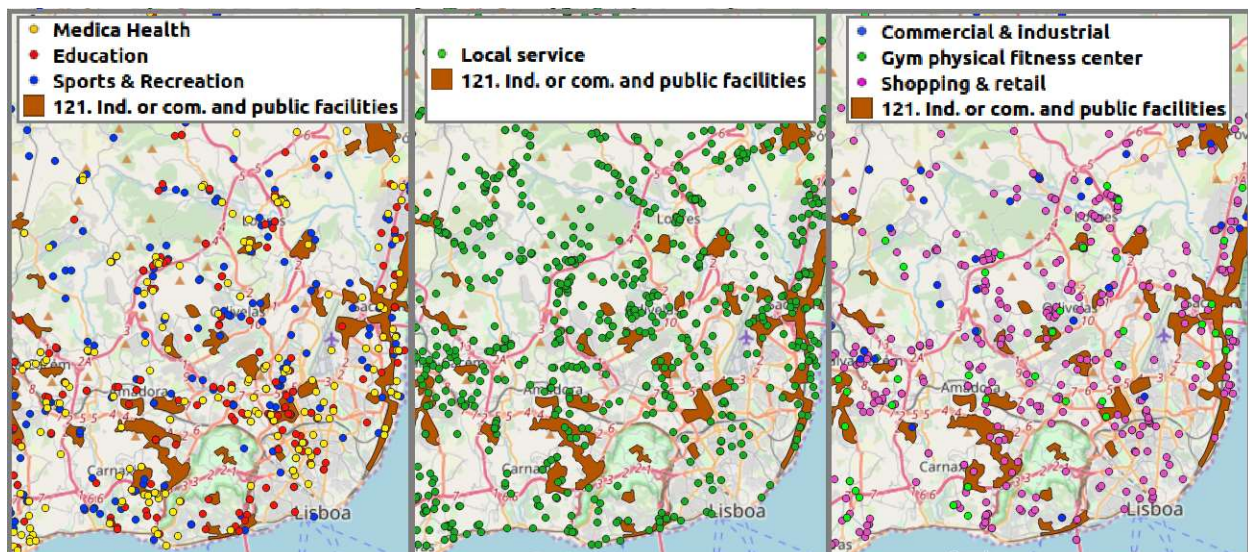


Figura 20: Distribuição dos POIs de algumas categorias em relação à classe “121”.

De forma geral, considerando que as classes envolvidas no estudo representam as zonas urbanas, é possível verificar através de uma análise com um grau de detalhe mais refinado, a forma como varia a correlação das diversas categorias, não apenas em função do raio utilizado nos mapas de concentração, mas também no que diz respeito ao uso ou ao tipo de cobertura de cada região. Nota-se por exemplo nas experiências baseadas na classe de nível 1, que independente da área de

influência do *heatmap*, 3 categorias foram sempre verificadas entre aquelas com maiores correlações. No entanto, ao se analisar as classes de nível 3, observa-se que 2 dessas categorias na verdade apresentam maior correlação com as zonas urbanas menos densas e a outra com as áreas onde se predomina uma cobertura artificial mais densa. Entretanto, quando a classe de uso do solo representa zonas mais pequenas, torna-se mais difícil perceber a relação destas com as diversas categorias. Portanto, pode-se concluir que a análise ao nível mais baixo das classes de uso e cobertura do solo pode apresentar resultados mais precisos, desde que os polígonos não representem áreas muito pequenas do território.

Observa-se também que a área de influência a ser considerada na geração do mapa de concentração é um parâmetro importante. Conforme abordado, existem situações onde *heatmaps* com raios muito grandes podem indicar baixas correlações entre as classes de uso do solo e as categorias, mesmo em situações onde as correlações apresentadas eram maiores aquando da adoção de um raio menor. Entretanto, serviços diferentes possuem áreas de influência diferentes. Desta forma, é recomendado que este parâmetro seja avaliado com cautela para a condução de um estudo deste género. Neste sentido, a realização de testes baseados em mapas de concentração com diferentes raios pode representar uma alternativa mais adequada.

Realizadas as análises com base nas 68 categorias selecionadas, verificou-se que entre os resultados mais relevantes estavam sempre presentes as subcategorias de uma mesma categoria de topo. Ao analisar este conjunto de categorias que apresentava maiores correlações em relação a todas as classes, verificou-se que 90% delas pertenciam à categoria “Business”. Considerando as indicações da possível importância desta categoria, optou-se por aprofundar ainda mais o estudo, considerando numa nova fase, as suas subcategorias de nível 2 e de nível 3.

Nesta segunda etapa do estudo, relativamente ao nível 2, verifica-se a presença das categorias “Automotive aircraft boat” e “Public government service”, entre aquelas com maior grau de correlação com a classe “1. Artificial areas”. O facto destas duas categorias não estarem presentes na lista utilizada anteriormente indica, por um lado, que embora o primeiro conjunto seja bastante completo, mesmo assim dentre as subcategorias da categoria “Business” encontram-se ainda outras que podem ser importantes neste tipo de análise. Por outro lado, considerando que no conjunto de categorias analisadas foram reveladas apenas 2 novas com correlações relevantes, pode-se afirmar também, que até certo ponto, o método de seleção das categorias utilizado anteriormente foi relativamente eficiente, caso contrário este número poderia ser ainda maior.

No que diz respeito à análise das categorias de nível 3, as análises revelaram também duas correlações importantes com a classe “1. Artificial surfaces”. Dentre os resultados mais relevantes, nota-se a ocorrência das categorias “Sports & recreation venue” e “School”. Recorrendo-se às correlações obtidas com as categorias de nível 2, verificam-se entre as mais significativas, a categoria “Sports & recreation”, onde se inclui a primeira, e “Education”, onde está incluída a segunda. Considerando que os mapas de densidade gerados incluem, para cada categoria, as suas

subcategorias, pode-se concluir que em certos casos é possível obter informações mais detalhadas aquando da adoção das categorias mais específicas, desde que estas possuam um número suficientemente representativo de pontos de interesse.

Ao realizar uma comparação entre os resultados obtidos para as classes “111. Continuous urban fabric” e “112. Discontinuous urban fabric”, verifica-se que no caso das categorias de nível 2, muitas daquelas que apresentavam alta correlação com uma das classes, apresentavam também com a outra. Em contrapartida, nota-se uma clara distinção ao se analisar os principais resultados obtidos para as categorias de nível 3. Neste caso, ao analisar os 10 resultados mais relevantes para cada uma das duas classes, nota-se que as categorias apresentadas são completamente diferentes. Considerando a distinção natural entre as classes de uso do solo em questão, seria aceitável supor que nas zonas representadas por cada uma delas, os pontos de interesse existentes fossem de natureza distinta. Portanto, é possível concluir que o estudo a um nível mais baixo das categorias, apresenta maior potencial para revelar mais detalhes no que diz respeito às correlações com as classes mais específicas de uso e ocupação do solo, o que provavelmente não poderia ser alcançado aquando da utilização apenas das categorias dos níveis superiores.

No geral, foram analisados e discutidos os resultados obtidos com base nas 68 categorias selecionadas para o estudo preliminar, além daqueles alcançados com recurso às subcategorias da categoria “Business”. Neste sentido, verifica-se que mesmo no caso do primeiro conjunto, as categorias que figuravam entre os principais resultados estavam, na sua grande maioria, ligadas à categoria “Business”, tornando evidente o potencial deste segundo conjunto para o estudo dos espaços urbanos.

### 5.2.2.3 Correlações entre as categorias dos pontos de interesse

Com base nos resultados da experiência anterior, optou-se pela realização de um estudo ainda mais aprofundado, onde foram analisadas as correlações existentes entre as diferentes subcategorias de nível 2 da categoria “Business”. Portanto, foram gerados mapas de concentração para cada uma delas. Neste processo foram consideradas as distâncias de 1250, 2500 e 5000 metros de raio. Os mapas de densidade gerados foram então comparados entre si, com o objetivo de extrair as correlações existentes entre as concentrações dos pontos de interesse das categorias representadas por eles. Os resultados obtidos são apresentados a seguir.

#### 5.2.2.3.1 Apresentação dos resultados

Realizadas as análises a partir dos mapas de concentração, foram então obtidas as correlações em três matrizes, cada uma delas para um conjunto específico de *heatmaps*, de acordo com os respetivos raios. As matrizes obtidas podem ser consultadas na íntegra, no apêndice H.2. Considerando os mapas de concentração com raio de 1250 metros, correlações cujos valores são iguais ou superior a 0.5, são destacados na Tabela 58. Nota-se que as categorias “Beauty cosmetic personal care”, “Food beverage” e “Local service” são aquelas de maior relevância, tendo em vista

que apresentam correlações relativamente altas com um maior número de categorias. No caso desta última, apresenta, inclusive, correlações superiores a 0.6 com outras 3 categorias.

Tabela 58: Correlações entre as subcategorias (nível 2) da categoria Business (raio: 1250m).

	[3]	[4]	[5]	[7]	[9]	[12]	[14]	[16]	[18]	[20]	[21]
[3] Arts entertainment	1										
[4] Automotive aircraft boat	0.24	1									
[5] Beauty cosmetic personal care	0.44	0.4	1								
[7] Education	0.47	0.33	0.59	1							
[9] Food beverage	<b>0.59</b>	0.33	<b>0.67</b>	0.52	1						
[12] Local service	0.51	0.56	<b>0.68</b>	<b>0.63</b>	<b>0.64</b>	1					
[14] Medical health	0.43	0.36	<b>0.63</b>	0.58	0.54	<b>0.66</b>	1				
[16] Nonprofit organization	0.49	0.24	0.51	0.5	0.55	0.51	0.46	1			
[18] Real estate	0.36	0.31	0.51	0.44	0.45	0.55	0.48	0.37	1		
[20] Shopping retail	0.42	0.44	0.58	0.49	0.56	<b>0.67</b>	0.55	0.41	0.46	1	
[21] Sports recreation	0.48	0.37	0.59	0.53	<b>0.61</b>	<b>0.62</b>	0.52	0.48	0.46	0.49	1

Relativamente à análise do conjunto de *heatmaps* criados com base na adoção de um raio de 2500 metros, os resultados são apresentados na Tabela 59. Embora os valores obtidos sejam maiores em relação ao caso anterior, destacando-se as 10 correlações mais significativas, verifica-se que entre eles encontram-se os da categoria “Beauty cosmetic personal care”. Entretanto, diferente do que se verificava anteriormente, a correlação apresentada com a categoria “Medical health” é menor do que a que se obteve em relação à categoria “Shopping & retail”.

Tabela 59: Correlações entre as subcategorias (nível 2) da categoria Business (raio: 2500m).

	[3]	[4]	[5]	[7]	[9]	[12]	[14]	[16]	[18]	[20]	[21]
[3] Arts entertainment	1										
[4] Automotive aircraft boat	0.38	1									
[5] Beauty cosmetic personal care	0.53	0.58	1								
[7] Education	0.63	0.50	0.70	1							
[9] Food beverage	0.66	0.46	<b>0.74</b>	0.65	1						
[12] Local service	0.60	0.72	<b>0.76</b>	<b>0.75</b>	<b>0.73</b>	1					
[14] Medical health	0.57	0.52	0.71	<b>0.76</b>	0.65	<b>0.75</b>	1				
[16] Nonprofit organization	0.65	0.38	0.6	0.67	0.66	0.63	0.62	1			
[18] Real estate	0.53	0.49	0.64	0.63	0.60	0.68	0.63	0.55	1		
[20] Shopping retail	0.56	0.65	<b>0.73</b>	0.67	0.69	<b>0.79</b>	0.70	0.57	0.63	1	
[21] Sports recreation	0.58	0.54	0.70	0.69	<b>0.73</b>	<b>0.75</b>	0.67	0.61	0.65	0.65	1

Finalmente, considerando os mapas de concentração com raio de 5000 metros, verifica-se um aumento significativo nos valores das correlações. Conforme informações apresentadas na Tabela 60, nota-se claramente resultados próximos a 0.9, como é o caso da correlação entre a

categoria “Local service” e “Shopping & retail”. Ao destacar os resultados mais importantes, é possível observar, em relação à categoria “Education”, que esta agora apresenta uma correlação relativamente alta em relação à categoria “Sports & recreation”, superando, inclusive, a correlação com a categoria “Medical health”, ao contrário do que se verificava anteriormente.

Tabela 60: Correlações entre as subcategorias (nível 2) da categoria Business (raio: 5000m).

	[3]	[4]	[5]	[7]	[9]	[12]	[14]	[16]	[18]	[20]	[21]
[3] Arts entertainment	1										
[4] Automotive aircraft boat	0.52	1									
[5] Beauty cosmetic personal care	0.62	0.75	1								
[7] Education	0.78	0.67	0.77	1							
[9] Food beverage	0.73	0.63	<b>0.83</b>	0.78	1						
[12] Local service	0.67	<b>0.84</b>	<b>0.83</b>	<b>0.83</b>	0.82	1					
[14] Medical health	0.73	0.69	0.78	<b>0.88</b>	0.77	0.82	1				
[16] Nonprofit organization	0.82	0.56	0.7	<b>0.83</b>	0.77	0.74	0.79	1			
[18] Real estate	0.69	0.66	0.73	0.79	0.74	0.77	0.75	0.73	1		
[20] Shopping retail	0.71	0.82	<b>0.85</b>	0.82	0.81	<b>0.88</b>	<b>0.83</b>	0.75	0.75	1	
[21] Sports recreation	0.69	0.71	0.78	<b>0.83</b>	<b>0.85</b>	<b>0.86</b>	0.81	0.76	0.81	0.8	1

### 5.2.2.3.2 Considerações acerca das experiências

Na experiência realizada foram selecionadas 22 categorias, com base nas etapas anteriores do estudo. Para os pontos de interesse que pertencem a cada uma delas, foram gerados mapas de concentração com 3 medidas distintas de raio. Os mapas de concentração gerados foram então submetidos à uma análise com base no cálculo das correlações existentes entre eles.

O resultado da análise demonstra que embora algumas categorias não tenham alcançado o topo dos resultados no que diz respeito à sua correlação com determinadas classes de uso do solo, como se pode verificar na experiência conduzida no tópico 5.2.2.2, ainda assim demonstram alguma ligação com outras categorias frequentemente encontradas em áreas onde há predominância das mesmas classes. É o caso da categoria “Food & beverage”. Embora esta categoria não tenha alcançado correlações superiores a 0.5, em nenhum dos casos do estudo anterior, nota-se que a mesma apresenta uma ligação bastante significativa com a categoria “Local service”, que por sua vez demonstrou correlações relativamente altas com a classe “1. Artificial surfaces”. Na prática, pode-se concluir, portanto, que existem casos onde a predominância dos pontos de interesse de uma determinada categoria pode não estar ligada a um tipo específico de uso ou ocupação do solo, mas ainda assim poderá ter relação com a existência de alguns segmentos específicos de empresas ou serviços naquela região geográfica.

Numa análise mais pormenorizada das correlações obtidas com base nos mapas de concentração com raio de 2500 metros, verifica-se que a categoria “Beauty cosmetic personal

care” passa a apresentar uma correlação mais significativa com a categoria “Shopping & retail”, do que com a categoria “Medical health”, diferente do que se verificava nos resultados obtidos com os *heatmaps* com raio de 1250 metros. O mesmo acontece com a categoria “Education”, nas correlações baseadas nos mapas de concentração com raio de 5000 metros, onde se verifica que a sua correlação com a categoria “Sports & recreation” passa a ser maior do que com a categoria “Medical health”, ao contrário do que foi observado com a utilização dos *heatmaps* com raio de 2500 metros. Desta forma, tal como nas análises das correlações entre as categorias dos pontos de interesse e as classes de uso e ocupação do solo, verifica-se, aquando do aumento da área de influência considerada no mapa de concentração, que nem todas as correlações variam na mesma proporção, o que indica que, ao contrário do que se imagina, *heatmaps* com raios muito grandes podem ocultar algumas correlações, ou revelar outras. Portanto, pode-se concluir que seria importante uma investigação mais aprofundada para se perceber a área de influência mais adequada para cada categoria dos pontos de interesse e para cada tipo de empresa ou serviço.

## 6 Conclusões

Neste capítulo são apresentadas as conclusões do estudo. Primeiramente é feita uma revisão geral do tema. Em seguida são revistos os objetivos e o trabalho realizado. Na sequência são abordados os seus principais contributos e as suas limitações. Por fim, são descritas as sugestões para o trabalho futuro.

### 6.1 Revisão geral do tema

O estudo dos espaços urbanos tem se tornado cada vez mais importante nos dias atuais, dada a sua importância principalmente para a administração pública, no que diz respeito à gestão das cidades. Deles dependem diversas questões, desde as infraestruturas até a qualidade de vida das pessoas. Neste domínio, um dos grandes problemas está relacionado à geração e manutenção das mais variadas informações dos espaços, como por exemplo os mapas de uso e ocupação do território, que normalmente dependem de conjuntos de dados com elevados custos financeiros. Desta forma faz-se necessário explorar o potencial das fontes de dados gratuitos para esta finalidade. Na dinâmica dos espaços urbanos modernos, outra questão fundamental para a economia é a atração de investimento local ou estrangeiro, onde encontrar a localização mais adequada para cada tipo de empresa é um grande desafio. Entretanto, com a evolução da tecnologia, tem se tornado possível a extração do conhecimento com base em informações geradas pelos estudos dos espaços urbanos, para suportar a tomada de decisões por parte dos investidores.

### 6.2 Revisão dos objetivos e do trabalho realizado

O primeiro objetivo apontado neste estudo, foi a investigação dos estudos dos espaços urbanos, realizados ao longo dos últimos anos, com o intuito de apontar os principais tipos de dados utilizados para a caracterização dos espaços urbanos, tal como os métodos e os seus respetivos objetivos. Foram analisados diversos trabalhos publicados em diferentes países. Foi realizada uma seleção daqueles mais relevantes para o que se pretendia. Após uma análise mais pormenorizada do conjunto selecionado, foram identificados alguns tipos de dados disponíveis gratuitamente na Internet, com relativo potencial para uso neste estudo. Dentre eles, um dos que mais se destacou foram os dados dos pontos de interesse.

Outro objetivo importante do trabalho realizado, foi a análise do potencial dos pontos de interesse para a identificação dos tipos de uso e ocupação do território. Foi selecionada a zona metropolitana de Lisboa como região a ser estudada. Para a automação da coleta dos dados na

Internet, foi desenvolvido um software. Os dados coletados foram tratados e organizados para permitir a sua utilização no estudo. Foram realizadas várias experiências, destacando a representatividade dos dados para a caracterização dos espaços em diferentes cenários. Em cada cenário foram utilizados diferentes conjuntos de atributos, gerados com base nos pontos de interesse. Para facilitar a geração dos conjuntos de atributos utilizados, foi desenvolvida uma aplicação informática. Após a realização das experiências, foram discutidos os resultados apresentados em cada cenário, tendo sido realizada uma análise cuidadosa no sentido de fornecer uma melhor interpretação em cada um dos casos. Esta etapa do trabalho revelou que os pontos de interesse possuem um relativo potencial para a caracterização dos espaços, embora alguns tipos de cobertura do solo sejam naturalmente difíceis de identificar com base nesse tipo de dados.

A identificação das ligações existentes entre cada categoria de empresas, serviços e outros pontos de interesse, e cada tipo de uso e ocupação do território, foi outro objetivo importante do trabalho. Para esta análise, recorreu-se ao uso de outros dois conjuntos de dados, além daquele já coletado anteriormente. Foram utilizados mapas de concentração para representar as várias categorias dos pontos de interesse. De maneira a permitir uma análise mais aprofundada, foram consideradas diferentes medidas de raio para a geração dos mapas de concentração. Foram desenvolvidas duas aplicações informáticas para permitir a realização desta etapa do estudo: uma delas para a preparação dos dados a serem utilizados e a outra para a criação dos mapas de concentração. Além disso, foi também desenvolvido um método automático para calcular as correlações existentes entre cada tipo de uso e cobertura do solo e cada categoria. Os resultados revelaram elevados graus de correlação em diferentes casos. Foi realizada uma análise pormenorizada no sentido de fornecer um entendimento mais aprofundado dos resultados obtidos. Nesta etapa, o estudo revelou vários padrões importantes, que podem ser utilizados para suportar as decisões relacionadas às localizações mais adequadas para cada tipo de empresa, serviço ou ponto de interesse. A metodologia adotada provou ser capaz de cumprir com relativa eficácia, o objetivo proposto.

Por fim, o último objetivo estabelecido para o estudo foi perceber a ligação existente entre as várias categorias de empresas, serviços e outros pontos de interesse. Para tal, foi proposto um método automático, também baseado em mapas de concentração. Com base nos resultados obtidos nas etapas anteriores, foram selecionadas diferentes categorias. Os resultados foram analisados e discutidos de maneira cuidadosa, no sentido de permitir a apresentação das conclusões obtidas. O estudo mostrou que a metodologia utilizada é capaz de permitir a extração do conhecimento no sentido proposto. A aplicação prática da metodologia, realizada nesta etapa do trabalho, demonstra o seu potencial para revelar padrões que possam promover o apoio à decisão no que diz respeito à escolha das localizações para as empresas, serviços e outros pontos de interesse.

## 6.3 Contributos

Considerando as diversas etapas envolvidas no estudo, é possível destacar diversos contributos teóricos e práticos, promovidos pelo trabalho realizado, os quais são apresentados a seguir.

### 6.3.1 Contributos teóricos

Um dos pontos de destaque do estudo, é essencialmente o trabalho de investigação realizado. Com base neste trabalho, foi possível contribuir para a literatura com duas publicações científicas. A primeira delas promove um *overview* dos principais tipos de dados e métodos utilizados para caracterização dos tipos de uso e ocupação do solo, no contexto dos espaços urbanos. De maneira muito mais aprofundada, a segunda aborda também o mesmo tema, fornecendo uma descrição mais detalhada tanto no que diz respeito aos dados quanto em relação aos métodos. Além disso fornece uma visão sistemática dos objetivos do uso de cada método, além de promover uma análise comparativa e uma discussão focada na combinação dos vários tipos de dados e dos diferentes métodos observados entre o conjunto de estudos analisado.

Finalmente, e ainda mais importante, acredita-se que o trabalho apresentado no capítulo 5 seja também um forte contributo teórico. Foram aplicados diferentes métodos para a análise do potencial dos dados dos pontos de interesse para classificação do uso do solo. Foi também desenvolvida e testada uma metodologia para análise das ligações existentes entre os tipos de uso e ocupação do território e as diversas categorias de empresas e serviços. Foram ainda investigadas as ligações existentes entre as diferentes categorias dos pontos de interesse. Os resultados obtidos em cada uma das 3 etapas abordadas sugere que as metodologias seguidas podem ser altamente relevantes em estudos para caracterização dos espaços urbanos e principalmente, para a extração de mais conhecimento, que somado àquele gerado pelos métodos tradicionalmente adotados pela indústria, pode vir a suportar as decisões tanto na administração pública, quanto no setor privado, no que diz respeito à escolha da localização mais adequada para a instalação de cada tipo de empresa ou serviço.

### 6.3.2 Contributos práticos

No decorrer das atividades realizadas para o cumprimento dos vários objetivos deste estudo, foram adotadas diferentes estratégias para colmatar as necessidades emergentes. Algumas dessas estratégias exigiram o domínio de um determinado conjunto de tecnologias. Um dos pontos mais relevantes da aplicação dessas tecnologias foi sem dúvidas, a criação das várias aplicações informáticas e dos demais métodos para as automações das quais dependeu quase que na sua totalidade, a realização deste estudo. Além disso, certamente há que se destacar o tratamento e a manipulação dos dados, presente em todas as fases do trabalho.

Na fase inicial, para a coleta dos dados, foi desenvolvido um software que esteve em execução durante determinado período de tempo. Finalizada esta tarefa, foi desenvolvida uma nova aplicação informática para a extração de conjuntos específicos de dados. E depois uma terceira. E da mesma forma ainda uma outra, para a geração dos mapas de densidade. Para a análise das correlações também foi desenvolvido um método automático. Todas as ferramentas desenvolvidas, tal como os *scripts* para a criação do modelo da base de dados, estão disponíveis online, no seguinte repositório: <https://gitlab.com/andradest/projeto-final-mis>. Neste mesmo repositório está também o ficheiro para a criação automática e a respetiva parametrização do ambiente computacional necessário para a execução da aplicação de coleta dos dados. Da mesma forma, os *datasets* utilizados estão também disponíveis online, no endereço <http://tiny.cc/i6uwoz>.

## 6.4 Limitações

Ao analisar o conteúdo apresentado no capítulo 5, é possível identificar alguns pontos que poderiam ser explorados de maneira ainda mais aprofundada caso estivesse disponível mais algum poder computacional. Em relação ao estudo do potencial dos pontos de interesse para a criação dos modelos para caracterização do uso e da ocupação do território, destaca-se por exemplo a seleção de atributos. Embora se tenha optado por uma seleção automática em vários cenários, devido ao grande número de atributos gerados, alguns métodos comuns na literatura, como por exemplo o *forward selection* e o *backward elimination*, não puderam ser testados. Além disso, ainda no que diz respeito ao trabalho de *feature engineering*, considera-se que poderia ser analisada a possibilidade de geração de novos atributos, que eventualmente poderiam ser utilizados para melhorar a performance dos modelos.

No que diz respeito ao estudo das ligações existentes entre as diferentes categorias das empresas, serviços e outros pontos de interesse, e das eventuais correlações entre elas e os tipos de uso e ocupação do território, as análises realizadas neste trabalho tiveram por base mapas de densidade com média resolução, o que também se atribui às referidas limitações no poder computacional disponível. Desta forma, com base no trabalho desenvolvido, não é possível determinar se seriam verificadas variações na intensidade das correlações consoante a resolução dos mapas de densidade utilizados.

No geral, mesmo considerando que os conjuntos de dados utilizados abrangiam todas as regiões do país, devido à limitação do poder computacional disponível durante a realização do estudo, foi escolhida a região metropolitana de Lisboa para as análises. Neste sentido, considera-se também que seria importante investigar os resultados obtidos com base em outras zonas do território, como por exemplo a área metropolitana do Porto, permitindo assim uma comparação entre eles, que serviria de complemento ao trabalho aqui apresentado.

## 6.5 Trabalho futuro

As evoluções tecnológicas vivenciadas ao longo dos últimos anos possibilitaram o surgimento de várias oportunidades de investigação no que diz respeito aos espaços urbanos. Neste contexto, tendo por base o estudo apresentado, é possível apontar algumas sugestões para o trabalho futuro. Primeiramente, em relação à utilização dos pontos de interesse para caracterização do uso e da cobertura do solo, considera-se importante a avaliação da qualidade dos dados coletados em diferentes fontes. Um estudo comparativo envolvendo pontos de interesse provenientes de diferentes repositórios poderia ser útil para os futuros estudos nesta linha de investigação. Além disso, e conforme sugerido por diferentes autores, faz sentido a combinação de outras fontes de dados para reforçar o potencial dos modelos criados para a classificação dos espaços.

O enriquecimento de dados seria outra sugestão importante para o trabalho futuro. A fusão de pontos de interesse provenientes de diferentes fontes, por exemplo, possivelmente resultaria numa base de dados mais completa, sobre a qual poderia ser aplicada a metodologia utilizada neste estudo. Uma base de dados mais rica poderia gerar resultados mais precisos, tanto na classificação dos espaços quanto no que diz respeito a análise das correlações. Neste sentido, seria importante investigar técnicas para a detecção do mesmo ponto de interesse em diferentes fontes, de maneira a evitar repetições. Além disso seria também necessário adotar medidas para limpeza dos dados, eliminando potenciais *outliers* que eventualmente acabam por prejudicar os resultados.

Neste estudo foi apresentada uma metodologia para a análise das correlações importantes para as várias categorias dos pontos de interesse, cujo objetivo final é fornecer informações para a tomada de decisão no sentido da escolha das localizações mais adequadas para a instalação de empresas ou serviços. Neste sentido, é evidente que a tomada de decisão é um processo complexo e que envolve diferentes variáveis. Tendo em vista a importância desta questão para a economia local, a investigação de novas formas de extração do conhecimento baseado em dados gratuitos ou de baixo custo, para auxiliar a administração pública e as empresas neste processo de tomada de decisão, seria talvez uma das mais importantes recomendações para a continuidade deste trabalho.



## 7 Referências

- [1] X. Liu *et al.*, “Classifying urban land use by integrating remote sensing and social media data,” *Int. J. Geogr. Inf. Sci.*, vol. 31, no. 8, pp. 1675–1696, 2017.
- [2] Y. Yao *et al.*, “Sensing spatial distribution of urban land use by integrating points-of-interest and Google Word2Vec model,” *Int. J. Geogr. Inf. Sci.*, vol. 31, no. 4, pp. 825–848, 2017.
- [3] V. Frias-Martinez and E. Frias-Martinez, “Spectral clustering for sensing urban land use using Twitter activity,” *Eng. Appl. Artif. Intell.*, vol. 35, pp. 237–245, 2014.
- [4] H. Xing and Y. Meng, “Integrating landscape metrics and socioeconomic features for urban functional region classification,” *Comput. Environ. Urban Syst.*, vol. 72, no. February, pp. 134–145, 2018.
- [5] X. Zhang, S. Du, and Q. Wang, “Hierarchical semantic cognition for urban functional zones with VHR satellite images and POI data,” *ISPRS J. Photogramm. Remote Sens.*, vol. 132, pp. 170–184, 2017.
- [6] Z. Deng, X. Zhu, Q. He, and L. Tang, “Land use/land cover classification using time series Landsat 8 images in a heavily urbanized area,” *Adv. Sp. Res.*, 2019.
- [7] Z. Qi, A. G. O. Yeh, X. Li, and Z. Lin, “A novel algorithm for land use and land cover classification using RADARSAT-2 polarimetric SAR data,” *Remote Sens. Environ.*, vol. 118, pp. 21–39, 2012.
- [8] Y. Zhang, Q. Li, H. Huang, W. Wu, X. Du, and H. Wang, “The combined use of remote sensing and social sensing data in fine-grained urban land use mapping: A case study in Beijing, China,” *Remote Sens.*, vol. 9, no. 9, 2017.
- [9] G. Re Calegari, E. Carlino, D. Peroni, and I. Celino, “Extracting urban land use from linked open geospatial data,” *ISPRS Int. J. Geo-Information*, vol. 4, no. 4, pp. 2109–2130, 2015.
- [10] H. Shu and P. ping Xiong, “Reallocation planning of urban industrial land for structure optimization and emission reduction: A practical analysis of urban agglomeration in China’s Yangtze River Delta,” *Land use policy*, vol. 81, no. November 2018, pp. 604–623, 2019.
- [11] J. H. Dunning, “Location Enterprise : the Multinational Neglected,” *J. Int. Bus. Stud.*, vol. 40, no. 1, pp. 45–66, 1998.
- [12] Z. Kapsalyamova, T. Mezher, N. Al Hosany, and I. T. Tsai, “Are low carbon cities attractive to cleantech firms? Empirical evidence from a survey,” *Sustain. Cities Soc.*, vol. 13, pp. 125–138, 2014.
- [13] I. Grossmann, “Perspectives for Hamburg as a port city in the context of a changing global environment,” *Geoforum*, vol. 39, no. 6, pp. 2062–2072, 2008.

- [14] S. Gao, K. Janowicz, and H. Couclelis, “Extracting urban functional regions from points of interest and human activities on location-based social networks,” *Trans. GIS*, vol. 21, no. 3, pp. 446–467, 2017.
- [15] R. Coffey, “The difference between ‘land use’ and ‘land cover,’” *Michigan State University Extension*, 2013. [Online]. Available: [https://www.canr.msu.edu/news/the\\_difference\\_between\\_land\\_use\\_and\\_land\\_cover](https://www.canr.msu.edu/news/the_difference_between_land_use_and_land_cover). [Accessed: 12-Dec-2019].
- [16] NOAA, “National Ocean Service website,” *What is the difference between land cover and land use?*, 2018. [Online]. Available: <https://oceanservice.noaa.gov/facts/lclu.html>. [Accessed: 13-Mar-2019].
- [17] NRCan, “Land Cover & Land Use,” *Natural Resources Canada*, 2019. [Online]. Available: <https://www.nrcan.gc.ca/earth-sciences/geomatics/satellite-imagery-air-photos/satellite-imagery-products/educational-resources/9373>. [Accessed: 03-May-2019].
- [18] T. Blaschke *et al.*, “Geographic Object-Based Image Analysis - Towards a new paradigm,” *ISPRS J. Photogramm. Remote Sens.*, vol. 87, pp. 180–191, 2014.
- [19] R. Brennan and T. L. Webster, “Object-oriented land cover classification of lidar-derived surfaces,” *Can. J. Remote Sens.*, vol. 32, no. 2, pp. 162–172, 2006.
- [20] B. Gong, J. Im, and G. Mountrakis, “An artificial immune network approach to multi-sensor land use/land cover classification,” *Remote Sens. Environ.*, vol. 115, no. 2, pp. 600–614, 2011.
- [21] X. Zhang, S. Du, and Q. Wang, “Integrating bottom-up classification and top-down feedback for improving urban land-cover and functional-zone mapping,” *Remote Sens. Environ.*, vol. 212, no. December 2017, pp. 231–248, 2018.
- [22] B. Huang, B. Zhao, and Y. Song, “Urban land-use mapping using a deep convolutional neural network with high spatial resolution multispectral remote sensing imagery,” *Remote Sens. Environ.*, vol. 214, no. April, pp. 73–86, 2018.
- [23] Y. Yao, H. Liang, X. Li, J. Zhang, and J. He, “Sensing urban land-use patterns by integrating Google Tensorflow and scene-classification models,” *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. - ISPRS Arch.*, vol. 42, no. 2W7, pp. 981–988, 2017.
- [24] J. Song, T. Lin, X. Li, and A. V. Prishchepov, “Mapping Urban Functional Zones by Integrating Very High Spatial Resolution Remote Sensing Imagery and Points of Interest: A Case Study of Xiamen, China,” *Remote Sens.*, vol. 10, no. 11, p. 1737, 2018.
- [25] OpenStreetMap, “About OpenStreetMap,” 2020. [Online]. Available: <https://www.openstreetmap.org/about>. [Accessed: 05-Dec-2019].
- [26] J. Yuan, Y. Zheng, and X. Xie, “Discovering Regions of Different Functions in a City Using Human Mobility and POIs,” *SIGKDD Int. Conf. Knowl. Discov. Data Min.*, 2012.
- [27] X. Liu *et al.*, “Characterizing mixed-use buildings based on multi-source big data,” *Int. J. Geogr. Inf. Sci.*, vol. 32, no. 4, pp. 738–756, 2018.

- [28] M. Jendryke, T. Balz, S. C. McClure, and M. Liao, “Putting people in the picture: Combining big location-based social media data and remote sensing imagery for enhanced contextual urban information in Shanghai,” *Comput. Environ. Urban Syst.*, vol. 62, pp. 99–112, 2017.
- [29] W. Zhai, X. Bai, Y. Shi, Y. Han, Z.-R. Peng, and C. Gu, “Beyond Word2vec: An approach for urban functional region extraction and identification by combining Place2vec and POIs,” *Comput. Environ. Urban Syst.*, vol. 74, no. August 2018, pp. 1–12, 2019.
- [30] J. C. García-Palomares, M. H. Salas-Olmedo, B. Moya-Gómez, A. Condeço-Melhorado, and J. Gutiérrez, “City dynamics through Twitter: Relationships between land use and spatiotemporal demographics,” *Cities*, vol. 72, no. September 2017, pp. 310–319, 2018.
- [31] J. J. Arsanjania and E. Vaz, “An assessment of a collaborative mapping approach for exploring land use patterns for several European metropolises,” *Int. J. Appl. Earth Obs. Geoinf.*, vol. 35, no. PB, pp. 329–337, 2015.
- [32] L. Gong, X. Liu, L. Wu, and Y. Liu, “Inferring trip purposes and uncovering travel patterns from taxi trajectory data,” *Cartogr. Geogr. Inf. Sci.*, vol. 43, no. 2, pp. 103–114, 2016.
- [33] N. J. Yuan, Y. Zheng, X. Xie, Y. Wang, K. Zheng, and H. Xiong, “Discovering urban functional zones using latent activity trajectories,” *IEEE Trans. Knowl. Data Eng.*, vol. 27, no. 3, pp. 712–725, 2015.
- [34] X. Zhang and S. Du, “A Linear Dirichlet Mixture Model for decomposing scenes: Application to analyzing urban functional zonings,” *Remote Sens. Environ.*, vol. 169, pp. 37–49, 2015.
- [35] E. Flores, M. Zortea, and J. Scharcanski, “Dictionaries of deep features for land-use scene classification of very high spatial resolution images,” *Pattern Recognit.*, vol. 89, pp. 32–44, 2019.
- [36] V. Kotu and B. Deshpande, *Data Science Concepts and Practice*, Second Edi. Morgan Kaufmann, 2019.
- [37] J. R. Quinlan, “Induction of decision trees,” *Mach. Learn.*, vol. 1, no. 1, pp. 81–106, 1986.
- [38] V. K. Ayyadevara, “Pro machine learning algorithms: a hands-on approach to implementing algorithms in Python and R.” Apress, 2018.
- [39] L. Breiman, “Random Forests,” *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
- [40] L. Breiman, “Bagging Predictors,” *Mach. Learn.*, vol. 24, no. 2, pp. 123–140, 1996.
- [41] C. C. Aggarwal, *Data Mining: The Textbook*. Springer Publishing Company, Incorporated, 2015.
- [42] C. Cortes and V. Vapnik, “Support-vector networks,” *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.
- [43] R. Fletcher, *Practical Methods of Optimization*. USA: Wiley-Interscience, 1987.
- [44] A. J. Smola and B. Schölkopf, “A tutorial on support vector regression,” *Stat. Comput.*, vol.

- 14, no. 3, pp. 199–222, 2004.
- [45] C. C. Chang and C. J. Lin, “LIBSVM: A Library for support vector machines,” *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 1–39, 2011.
- [46] S. Knerr, L. Personnaz, and G. Dreyfus, “Single-layer learning revisited: a stepwise procedure for building and training a neural network BT - Neurocomputing,” 1990, pp. 41–50.
- [47] S. S. Durduran, “Automatic classification of high resolution land cover using a new data weighting procedure: The combination of k-means clustering algorithm and central tendency measures (KMC-CTM),” *Appl. Soft Comput. J.*, vol. 35, pp. 136–150, 2015.
- [48] X. Zhan, S. V. Ukkusuri, and F. Zhu, “Inferring Urban Land Use Using Large-Scale Social Media Check-in Data,” *Networks Spat. Econ.*, vol. 14, no. 3–4, pp. 647–667, 2014.
- [49] S. Jiang, A. Alves, F. Rodrigues, J. Ferreira, and F. C. Pereira, “Mining point-of-interest data from social networks for urban land use classification and disaggregation,” *Comput. Environ. Urban Syst.*, vol. 53, pp. 36–46, 2015.
- [50] C. Zhang *et al.*, “An object-based convolutional neural network (OCNN) for urban land use classification,” *Remote Sens. Environ.*, vol. 216, no. June, pp. 57–70, 2018.
- [51] C. Zhang *et al.*, “Joint Deep Learning for land cover and land use classification,” *Remote Sens. Environ.*, vol. 221, no. November 2018, pp. 173–187, 2019.
- [52] S. V. S. K. Ukkusuri and C. Yang, “Transportation analytics in the era of big data.” Springer, Cham, 2019.
- [53] J. Oksanen, C. Bergman, J. Sainio, and J. Westerholm, “Methods for deriving and calibrating privacy-preserving heat maps from mobile sports tracking application data,” *J. Transp. Geogr.*, vol. 48, pp. 135–144, 2015.
- [54] Y. Zhang, L. Liu, and H. Wang, “A new perspective on the temporal pattern of human activities in cities: The case of Shanghai,” *Cities*, vol. 87, no. March 2018, pp. 196–204, 2019.
- [55] B. P. Y. Loo and T. K. Anderson, *Spatial Analysis Methods of Road Traffic Collisions*, 1st ed. USA: CRC Press, Inc., 2015.
- [56] A. Cutts and A. Graser, “Learn QGIS : your step-by-step guide to the fundamental of QGIS 3.4.” Packt Publishing, 2018.
- [57] S. L. Jackson, *Research Methods and Statistics A Critical Thinking Approach*. 2009.
- [58] R. Haining, “Spatial Data Analysis in the Social and Environmental Sciences,” R. Haining, Ed. Cambridge: Cambridge University Press, 1990.
- [59] Y. Chen, “A new methodology of spatial cross-correlation analysis,” *PLoS One*, vol. 10, no. 5, pp. 1–20, 2015.
- [60] D. A. Griffith, “Spatial Autocorrelation,” A. B. T.-I. E. of H. G. (Second E. Kobayashi, Ed. Oxford: Elsevier, 2020, pp. 355–366.

- [61] S. Wang, H. Liu, H. Pu, and H. Yang, “Spatial disparity and hierarchical cluster analysis of final energy consumption in China,” *Energy*, vol. 197, p. 117195, 2020.
- [62] P. A. P. Moran, “Notes on Continuous Stochastic Phenomena,” *Biometrika*, vol. 37, no. 1/2, pp. 17–23, Apr. 1950.
- [63] R. C. Geary, “The Contiguity Ratio and Statistical Mapping,” *Inc. Stat.*, vol. 5, no. 3, pp. 115–146, Apr. 1954.
- [64] L. Anselin, “Local Indicators of Spatial Association—LISA,” *Geogr. Anal.*, vol. 27, no. 2, pp. 93–115, 1995.
- [65] B. Podobnik and H. E. Stanley, “Detrended cross-correlation analysis: A new method for analyzing two nonstationary time series,” *Phys. Rev. Lett.*, vol. 100, no. 8, pp. 1–11, 2008.
- [66] V. Plerou, P. Gopikrishnan, B. Rosenow, L. A. N. Amaral, and H. E. Stanley, “Universal and nonuniversal properties of cross correlations in financial time series,” *Phys. Rev. Lett.*, vol. 83, no. 7, pp. 1471–1474, 1999.
- [67] L. F. Batista, N. N. Imai, L. H. da S. Rotta, and F. S. Y. Watanabe, “Análise de correlação espacial de variáveis relacionadas à colonização de macrófitas submersas,” *An. XVI Simp. Bras. Sensoriamento Remoto - SBSR, Foz do Iguaçu, PR, Bras. 13 a 18 abril 2013, INPE*, no. 1986, pp. 6917–6922, 2013.
- [68] T. Zhang, L. Sun, L. Yao, and J. Rong, “Impact Analysis of Land Use on Traffic Congestion Using Real-Time Traffic and POI,” *J. Adv. Transp.*, vol. 2017, no. 1985, 2017.
- [69] A. Soliman, K. Soltani, J. Yin, A. Padmanabhan, and S. Wang, “Social sensing of urban land use based on analysis of Twitter users’ mobility patterns,” *PLoS One*, vol. 12, no. 7, p. e0181657, 2017.
- [70] L. Guo, S. Du, R. Haining, and L. Zhang, “Global and local indicators of spatial association between points and polygons: A study of land use change,” *Int. J. Appl. Earth Obs. Geoinf.*, vol. 21, no. 1, pp. 384–396, 2012.
- [71] L. Dai *et al.*, “Geostatistical analyses and co-occurrence correlations of heavy metals distribution with various types of land use within a watershed in eastern QingHai-Tibet Plateau, China,” *Sci. Total Environ.*, vol. 653, pp. 849–859, 2019.
- [72] R. K. Pearson, “Exploratory data analysis using R.” 2018.
- [73] P. Dalgaard, “Introductory statistics with R,” in *Statistics and computing*, 2008.
- [74] F. J. Gravetter and L. B. Wallnau, *Statistics for the behavioral sciences*, 10th Editi. Cengage Learning, 2017.
- [75] R. B. Ewen, *Workbook for Introductory Statistics for the Behavioral Sciences*, Second Edi. Academic Press, 1976.
- [76] Docker, “Docker: The Modern Platform for High-Velocity Innovation,” 2019. [Online]. Available: <https://www.docker.com/why-docker>. [Accessed: 27-Nov-2019].
- [77] Opensource.com, “What is Docker?,” 2019. [Online]. Available:

- <https://opensource.com/resources/what-docker>. [Accessed: 27-Nov-2019].
- [78] Microsoft, “Windows and containers,” 2019. [Online]. Available: <https://docs.microsoft.com/en-us/virtualization/windowscontainers/about/>. [Accessed: 27-Nov-2019].
- [79] Microsoft, “What is Docker?,” 2018. [Online]. Available: <https://docs.microsoft.com/en-us/dotnet/architecture/microservices/container-docker-introduction/docker-defined>. [Accessed: 27-Nov-2019].
- [80] Red Hat, “What is DOCKER?,” 2019. [Online]. Available: <https://www.redhat.com/en/topics/containers/what-is-docker>. [Accessed: 27-Nov-2019].
- [81] Docker, “Dockerfile Reference,” 2019. [Online]. Available: <https://docs.docker.com/engine/reference/builder/>. [Accessed: 28-Nov-2019].
- [82] A. F. Sveen, “Efficient storage of heterogeneous geospatial data in spatial databases,” *J. Big Data*, vol. 6, no. 1, 2019.
- [83] PostgreSQL, “New to PostgreSQL?,” 2019. [Online]. Available: <https://www.postgresql.org/>. [Accessed: 28-Nov-2019].
- [84] PostgreSQLTutorial.com, “What is PostgreSQL?,” 2019. [Online]. Available: <http://www.postgresqltutorial.com/what-is-postgresql/>. [Accessed: 28-Nov-2019].
- [85] PostgreSQL, “About PostgreSQL,” 2019. [Online]. Available: <https://www.postgresql.org/about/>. [Accessed: 28-Nov-2019].
- [86] PostgreSQL, “Chapter 39. Procedural Languages,” 2019. [Online]. Available: <https://www.postgresql.org/docs/9.5/xplang.html>. [Accessed: 28-Nov-2019].
- [87] PostgreSQL, “9.15. JSON Functions and Operators,” 2019. [Online]. Available: <https://www.postgresql.org/docs/10/functions-json.html>. [Accessed: 28-Nov-2019].
- [88] PostGIS, “About PostGIS,” 2019. [Online]. Available: <https://postgis.net/>. [Accessed: 02-Dec-2019].
- [89] OSGeo, “PostGIS,” 2019. [Online]. Available: <https://www.osgeo.org/projects/postgis/>. [Accessed: 02-Dec-2019].
- [90] T. Tjukanov, “Why should you care about PostGIS? — A gentle introduction to spatial databases,” 2018. [Online]. Available: <https://medium.com/@tjukanov/why-should-you-care-about-postgis-a-gentle-introduction-to-spatial-databases-9eccd26bc42b>. [Accessed: 02-Dec-2019].
- [91] National Geographic, “GIS (Geographic Information System),” 2019. [Online]. Available: <https://www.nationalgeographic.org/encyclopedia/geographic-information-system-gis/>. [Accessed: 02-Dec-2019].
- [92] S. Khan and K. Mohiuddin, “Empirical Evaluation of Open Source QGIS with Contemporary Proprietary GIS Systems - A Study,” *J. Multi Discip. Eng. Technol.*, vol. 12, no. 1, pp. 31–38, 2018.

- [93] QGIS, “QGIS - The Leading Open Source Desktop GIS,” 2019. [Online]. Available: <https://www.qgis.org/en/site/about/index.html>. [Accessed: 01-Dec-2019].
- [94] C. Dempsey, “Getting Started With QGIS: Open Source GIS,” 2012. [Online]. Available: <https://www.gislounge.com/getting-started-with-qgis-open-source-gis/>. [Accessed: 02-Dec-2019].
- [95] RapidMiner, “RapidMiner | Data Science & Machine Learning Platform,” 2020. [Online]. Available: <https://rapidminer.com/>. [Accessed: 18-Jun-2019].
- [96] CompareCamp, “RapidMiner Review,” 2020. [Online]. Available: <https://comparecamp.com/rapidminer-review-pricing-pros-cons-features/>. [Accessed: 06-May-2020].
- [97] PHP, “What is PHP?,” 2019. [Online]. Available: <https://www.php.net/manual/en/intro-whatis.php>. [Accessed: 03-Dec-2019].
- [98] E. McGrath, “12 ‘Must-Know’ Advantages of PHP,” 2015. [Online]. Available: <https://www.vandelaydesign.com/advantages-of-php/>. [Accessed: 03-Dec-2019].
- [99] Oracle, “What is Java technology and why do I need it?,” 2019. .
- [100] T. Noergaard, “Section II. Embedded Hardware,” in *Embedded Systems Architecture (Second Edition)*, Second Edi., T. Noergaard, Ed. Newnes, 2013, p. 87.
- [101] Python, “What is Python? Executive Summary,” 2019. [Online]. Available: <https://www.python.org/doc/essays/blurb/>. [Accessed: 04-Dec-2019].
- [102] R Foundation, “What is R?,” 2019. [Online]. Available: <https://www.r-project.org/about.html>. [Accessed: 05-Dec-2019].
- [103] OpenStreetMap, “Points of interest,” 2019. [Online]. Available: [https://wiki.openstreetmap.org/wiki/Points\\_of\\_interest](https://wiki.openstreetmap.org/wiki/Points_of_interest). [Accessed: 09-Dec-2019].
- [104] Zephoria Digital Marketing, “The Top 20 Valuable Facebook Statistics – Updated November 2019,” 2019. [Online]. Available: <https://zephoria.com/top-15-valuable-facebook-statistics/>. [Accessed: 10-Dec-2019].
- [105] A. Pilon, “What You Need to Know About Facebook Places for Your Business,” 2019. [Online]. Available: <https://smallbiztrends.com/2019/07/facebook-places.html>. [Accessed: 10-Dec-2019].
- [106] D. Gross, “Facebook Places: Here’s how it works,” 2010. [Online]. Available: <http://edition.cnn.com/2010/TECH/social.media/08/19/facebook.places.explain/index.html>. [Accessed: 11-Dec-2019].
- [107] Facebook, “Places Graph,” 2019. [Online]. Available: <https://developers.facebook.com/products/places/>. [Accessed: 10-Dec-2019].
- [108] Microsoft, “Extração, transformação e carregamento (ETL),” 2019. [Online]. Available: <https://docs.microsoft.com/pt-pt/azure/architecture/data-guide/relational-data/etl>. [Accessed: 16-Dec-2019].

- [109] Facebook, “All categories,” 2019. [Online]. Available: <https://www.facebook.com/pages/category/>. [Accessed: 16-Dec-2019].
- [110] Copernicus, “CORINE Land Cover,” 0, 2020. [Online]. Available: <https://land.copernicus.eu/pan-european/corine-land-cover>. [Accessed: 02-Mar-2020].
- [111] B. Kosztra, G. Büttner, G. Hazeu, and S. Arnold, “Updated CLC illustrated nomenclature guidelines,” 2019. [Online]. Available: [https://land.copernicus.eu/user-corner/technical-library/corine-land-cover-nomenclature-guidelines/docs/pdf/CLC2018\\_Nomenclature\\_illustrated\\_guide\\_20190510.pdf](https://land.copernicus.eu/user-corner/technical-library/corine-land-cover-nomenclature-guidelines/docs/pdf/CLC2018_Nomenclature_illustrated_guide_20190510.pdf). [Accessed: 03-Oct-2020].
- [112] Direção-Geral do Território, “Especificações técnicas da Carta de Uso e Ocupação do Solo (COS) de Portugal Continental para 2018,” 2019. [Online]. Available: [http://mapas.dgterritorio.pt/atom-dgt/pdf-cous/COS2018/ET-COS-2018\\_v1.pdf](http://mapas.dgterritorio.pt/atom-dgt/pdf-cous/COS2018/ET-COS-2018_v1.pdf). [Accessed: 05-Dec-2019].
- [113] Direção Geral de Território, “Carta de Uso e Ocupação do Solo de Portugal Continental (COS),” 2020. [Online]. Available: [http://www.dgterritorio.pt/dados\\_abertos/cos/](http://www.dgterritorio.pt/dados_abertos/cos/). [Accessed: 02-Mar-2020].
- [114] J. Novaković, P. Strbac, and D. Bulatović, “Toward optimal feature selection using ranking methods and classification algorithms,” *Yugosl. J. Oper. Res.*, vol. 21, no. 1, pp. 119–135, 2011.

# Apêndice A. *Scripts* SQL

Neste apêndice são apresentados os *scripts* SQL utilizados para colmatar as necessidades emergentes nas diferentes etapas do estudo.

## A.1. Filtragem das coordenadas

Depois da aplicação desenvolvida para coleta dos dados realizar a criação da grelha inicial de coordenadas, foi necessário realizar a filtragem dos pontos gerados. Esse processo originou uma nova tabela na base de dados, chamada “coordinates”. O código SQL utilizado foi o seguinte:

```
CREATE TABLE poi.coordinates AS (  
  SELECT id,  
         latitude,  
         longitude,  
         geom_point,  
         pristine,  
         in_progress  
  FROM coordinates_initial c,  
       dgt_portugal p  
  WHERE st_intersects(c.geom_point, p.geom)  
);
```

## A.2. Filtragem dos pontos de interesse

O código SQL mostrado a seguir foi utilizado para gerar uma nova tabela de pontos de interesse, com base na antiga, aplicando-se uma filtragem aos dados para evitar a repetição dos mesmos.

```
CREATE TABLE poi.points_of_interest AS (  
  SELECT DISTINCT ON (link)  
         autoink_pk,  
         name,  
         about,  
         description,  
         category_list :: json,  
         checkins,  
         is_permanently_closed,  
         is_verified,  
         link,  
         single_line_address,  
         website,  
         hours,  
);
```

```

    espec_fld_city,
    espec_fld_city_id,
    espec_fld_country,
    espec_fld_country_code,
    espec_fld_located_in,
    espec_fld_latitude,
    espec_fld_longitude,
    espec_fld_name,
    espec_fld_region,
    espec_fld_region_id,
    espec_fld_state,
    espec_fld_street,
    espec_fld_zip,
    espec_fld_delivery,
    espec_fld_catering,
    espec_fld_groups,
    espec_fld_kids,
    espec_fld_outdoor,
    espec_fld_pickup,
    espec_fld_reserve,
    espec_fld_takeout,
    espec_fld_waiter,
    espec_fld_walkins,
    espec_fld_breakfast,
    espec_fld_coffee,
    espec_fld_dinner,
    espec_fld_drinks,
    espec_fld_lunch,
    datetime,
    url,
    coords,
    id
  FROM fbplaces_refined_categ_list
);
ALTER TABLE poi.points_of_interest
  ADD PRIMARY KEY (autoink_pk);

SELECT AddGeometryColumn('points_of_interest', 'geom_point', 4326, 'POINT', 2);
START TRANSACTION;
UPDATE points_of_interest
  SET geom_point = st_setsrid(st_point(espec_fld_longitude, espec_fld_latitude),
4326);
COMMIT ;

```

### A.3. Criação da tabela de categorias

O *script* a seguir foi utilizado para extrair as categorias do campo JSON “categoria\_list” dos pontos de interesse, armazená-las numa nova tabela e estabelecer a ligação entre esta e os POIs.

```

CREATE TABLE poi.poi_categories AS (
  SELECT DISTINCT ON (category_id) category_id, category_name

```

```

FROM (SELECT DISTINCT ON (category_name) category_id, category_name
      FROM (SELECT cjl.category_id :: bigint AS category_id,
                  cjl.category_name :: text
                  FROM points_of_interest p
                  CROSS JOIN LATERAL (
                    SELECT d.elem :: json ->> 'id' AS category_id,
                          d.elem :: json ->> 'name' AS category_name
                    FROM json_array_elements_text(p.category_list) AS d (elem)
                  ) cjl
            ) cat0
      ) cat1
ORDER BY cat1.category_id
);
CREATE TABLE poi.poi_vs_categories AS (
  SELECT DISTINCT ON (p.autoink_pk, cat.category_id) p.autoink_pk AS poi_id,
                 cat.category_id
  FROM points_of_interest p
  CROSS JOIN LATERAL (
    SELECT d.elem :: json ->> 'name' AS category_name
           -- string_agg(d.elem::text, ', ') AS list
    FROM json_array_elements_text(p.category_list) AS d (elem)
  ) cjl
  INNER JOIN poi_categories cat ON cat.category_name = cjl.category_name
  ORDER BY p.autoink_pk ASC, cat.category_id ASC
);
ALTER TABLE poi_vs_categories
  ADD PRIMARY KEY (poi_id, category_id);
ALTER TABLE poi_vs_categories
  ADD FOREIGN KEY (category_id) REFERENCES poi_categories (category_id);
ALTER TABLE poi_vs_categories
  ADD FOREIGN KEY (poi_id) REFERENCES points_of_interest (autoink_pk);

```

## A.4. Importação das categorias obtidas via API

As categorias obtidas via API por meio da procura pelos Place Topics foram importadas para a base de dados através da utilização das seguintes instruções SQL.

```

CREATE UNLOGGED TABLE poi.fbplaces_categories_temp1 (
  doc json
);
COPY poi.fbplaces_categories_temp1 FROM '/tmp/FacebookCategories_JSON_3.json' WITH
(FORMAT TEXT);
CREATE TABLE poi.fbplaces_categories_temp2 (
  id          BIGINT          NOT NULL,
  name        VARCHAR(250) NOT NULL,
  parent_ids  JSON
);
CREATE TYPE categ AS (

```

```

    id          BIGINT,
    name        VARCHAR(250),
    parent_ids  JSON
);

START TRANSACTION;
INSERT INTO fbplaces_categories_temp2 (
    SELECT p.id :: BIGINT, p.name, p.parent_ids
    FROM fbplaces_categories_temp1 v
    CROSS JOIN LATERAL json_populate_recordset(null :: categ, doc) AS p
);

DROP TABLE fbplaces_categories_temp1;
DROP TYPE categ;

CREATE TABLE poi.fbplaces_categ_placetopics_api AS (
    WITH aux(id, name, parent_id) AS (
        SELECT c.id,
               c.name,
               json_array_elements(c.parent_ids) :: TEXT AS parent_id
        FROM fbplaces_categories_temp2 c
    ) SELECT id,
            name,
            replace(parent_id, '"', '') :: BIGINT AS parent_id
    FROM aux
);

ALTER TABLE fbplaces_categ_placetopics_api
    ADD PRIMARY KEY (id);
DROP TABLE fbplaces_categories_temp2;
COMMIT ;

```

## A.5. Importação das categorias do site para a base de dados

Após a criação do ficheiro mostrado no apêndice B.1, o mesmo foi exportado no formato CSV e importado na base de dados. O *script* SQL mostrado a seguir foi usado para criar uma nova tabela contendo os atributos “parent\_id” e “parent\_code”, que se referem à categoria a qual pertence cada subcategoria.

```

START TRANSACTION;

CREATE TABLE poi.fbplaces_categ_site_v2 AS (
    SELECT c1.id,
           c1.code,
           c1.category_name,
           c2.code AS parent_code,
           c2.id AS parent_id
    FROM fbplaces_categ_site_v1 c1
    LEFT JOIN fbplaces_categ_site_v1 c2 ON c2.code = LEFT(c1.code, LENGTH(c1.code) -

```

```

POSITION('.' in REVERSE(c1.code)))
                                AND LENGTH(c1.code) > 1);
ALTER TABLE fbplaces_categ_site_v2
  ADD PRIMARY KEY (id);
ALTER TABLE fbplaces_categ_site_v2
  ADD FOREIGN KEY (parent_id) REFERENCES fbplaces_categ_site_v2 (id);

COMMIT ;

```

## A.6. Importação dos dados do endpoint “fb\_page\_categories”

As categorias obtidas no *endpoint* “fb\_page\_categories” foram salvas num ficheiro em formato JSON, que foi importado para a base de dados por meio da utilização do seguinte código SQL.

```

CREATE UNLOGGED TABLE poi.fb_page_categories_temp1 (
  doc json
);

-- Ler o ficheiro JSON
COPY poi.fb_page_categories_temp1 FROM '/tmp/FacebookPageCategories.json' WITH
(FORMAT TEXT);
SELECT * FROM fb_page_categories_temp1;

-- DROP TYPE pagecateg;
CREATE TYPE pagecateg AS (
  id BIGINT,
  name VARCHAR(250),
  api_enum VARCHAR(50),
  parent_id BIGINT,
  fb_page_categories JSON
);

-- Criar a tabela fbpages_categ_api_v1 para guardar as categorias
ROLLBACK;
START TRANSACTION;
CREATE TABLE poi.fbpages_categ_api_v1 AS (
  WITH RECURSIVE category_tree(id, name, parent_id, fb_page_categories) AS (
    SELECT parent.id,
           parent.name,
           null :: BIGINT AS parent_id,
           parent.fb_page_categories
    FROM fb_page_categories_temp1
    CROSS JOIN LATERAL json_populate_recordset(null :: pagecateg, doc) AS parent
    UNION ALL
    SELECT son.id,
           son.name,
           t.id AS parent_id,
           son.fb_page_categories
    FROM category_tree AS t
    CROSS JOIN LATERAL json_populate_recordset(null :: pagecateg,
t.fb_page_categories) AS son

```

```

        WHERE t.fb_page_categories IS NOT NULL
    ) SELECT t.id,
           t.name,
           t.parent_id
       FROM category_tree t);

-- Criar a tabela fbpages_categ_api_v2 para guardar as categorias, incluindo o nível
de cada uma
CREATE TABLE poi.fbpages_categ_api_v2 AS (
    WITH RECURSIVE category_tree(id, name, parent_id, fb_page_categories) AS (
        SELECT parent.id,
               parent.name,
               null :: BIGINT AS parent_id,
               parent.fb_page_categories,
               1 AS _level
        FROM fb_page_categories_temp1
        CROSS JOIN LATERAL json_populate_recordset(null :: pagecateg, doc) AS parent
        UNION ALL
        SELECT son.id,
               son.name,
               t.id AS parent_id,
               son.fb_page_categories,
               (t._level + 1) AS _level
        FROM category_tree AS t
        CROSS JOIN LATERAL json_populate_recordset(null :: pagecateg,
t.fb_page_categories) AS son
        WHERE t.fb_page_categories IS NOT NULL
    ) SELECT t.id,
           t.name,
           t.parent_id,
           t._level
       FROM category_tree t
    );

ALTER TABLE fbpages_categ_api_v2 ADD PRIMARY KEY (id);
ALTER TABLE fbpages_categ_api_v2 ADD FOREIGN KEY (parent_id) REFERENCES
fbpages_categ_api_v2(id);

COMMIT ;

```

## A.7. Carregamento da tabela de mapeamento das categorias

As categorias extraídas dos pontos de interesse foram mapeadas para aquelas obtidas no *endpoint* “fb\_pages\_categories” em duas etapas: (1) a importação do mapeamento feito de forma manual, apresentado no apêndice B.2 e (2) a inserção da ligação para os registos que já existiam nas duas tabelas. O *script* SQL utilizado para esta segunda etapa é apresentado a seguir.

```

INSERT INTO poi_categories_to_fbpage_categ_api_v1
    SELECT c1.category_id AS original_id, c2.id AS mapped_id

```

```

FROM poi_categories c1
INNER JOIN fbpages_categ_api_v2 c2 ON c2.id = c1.category_id
ORDER BY category_id;

```

## A.8. Criação de uma tabela com os centroides de cada polígono

O comando SQL a seguir, exemplifica a criação de uma nova tabela espacial na base de dados para armazenar os polígonos de determinadas classes do *dataset* COS2018. Neste trecho de código, destaca-se a função “ST\_PointOnSurface”, da extensão PostGIS. Esta função determina um ponto para representar o centro do polígono.

```

CREATE TABLE cos.cos2018_pt_industria_comercio AS
SELECT cospt.gid
      , cospt.cos2018_lg
      , cospt.area
      , cospt.geom
      , ST_Transform(ST_PointOnSurface(cospt.geom), 4326) AS centroid
FROM cos.cos2018_portugal cospt
WHERE cospt.cos2018_lg IN ('1.2.1.1 Indústria', '1.2.2.1 Comércio', '1.2.3.1
Instalações agrícolas');

```

## A.9. Isolamento dos pontos de interesse da categoria “Business”

De maneira a facilitar a execução das consultas utilizadas para gerar os *datasets*, foi criada uma nova tabela para armazenar todos os pontos de interesse da categoria “Business”, tendo em conta que neste caso se um determinado ponto de interesse “A”, encontra-se na categoria “Xpto”, entretanto a categoria “Xpto” é uma subcategoria de “Business”, considera-se que “A” é um ponto de interesse de “Business”. Desta forma, foi criada a função “points\_of\_interest\_by\_category\_v1”, que recebe uma determinada categoria e retorna todos os POIs existentes nesta, tal como nas suas subcategorias, de forma recursiva.

```

CREATE FUNCTION poi.points_of_interest_by_category_v1(category_id NUMERIC)
RETURNS TABLE(
  _1st_categ_id BIGINT
  , _1st_categ_name TEXT
  , _2nd_categ_id BIGINT
  , _2nd_categ_name TEXT
  , subcategory_id BIGINT
  , subcategory_name TEXT
  , geom_point GEOMETRY
  , latitude DOUBLE PRECISION
  , longitude DOUBLE PRECISION
  , poi_autoink_pk INT

```

```

) AS
$body$
SELECT *
FROM (
    WITH level2 AS (
        SELECT lvl1.id AS _1st_categ_id
            , lvl1.name AS _1st_categ_name
            , lvl2.id AS _2nd_categ_id
            , lvl2.name AS _2nd_categ_name
        FROM poi.fbpages_categ_api_v2 lvl1
        INNER JOIN poi.fbpages_categ_api_v2 lvl2 ON lvl2.parent_id = lvl1.id
    OR lvl2.id = lvl1.id
        WHERE lvl1.id = $1
    ), low_level AS (
        WITH RECURSIVE categories AS (
            SELECT lvl2._1st_categ_id
                , lvl2._1st_categ_name
                , lvl2._2nd_categ_id
                , lvl2._2nd_categ_name
                , lvl2._2nd_categ_id AS subcategory_id
                , lvl2._2nd_categ_name AS subcategory_name
            FROM level2 lvl2
            UNION
            SELECT cat._1st_categ_id
                , cat._1st_categ_name
                , cat._2nd_categ_id
                , cat._2nd_categ_name
                , son.id
                , son.name
            FROM categories cat
            INNER JOIN poi.fbpages_categ_api_v2 son ON son.parent_id =
cat.subcategory_id
            WHERE son._level > 2
        ) SELECT *
        FROM categories
    ) SELECT cat._1st_categ_id
        , cat._1st_categ_name
        , cat._2nd_categ_id
        , cat._2nd_categ_name
        , cat.subcategory_id
        , cat.subcategory_name
        , poi.geom_point
        , poi.espec_fld_latitude AS latitude
        , poi.espec_fld_longitude AS longitude
        , poi.autoink_pk AS poi_autoink_pk
    FROM low_level cat
    INNER JOIN poi.poi_categories_to_fbpage_categ_api_v1 map ON
map.maped_id = cat.subcategory_id
    INNER JOIN poi.poi_vs_categories pvc ON
pvc.category_id = map.original_id
    INNER JOIN poi.points_of_interest poi ON
poi.autoink_pk = pvc.poi_id
    ) s0
$body$
LANGUAGE SQL;

```

```
CREATE TABLE poi.pois_business_full AS
SELECT categ.id AS category_id
      , pois.geom_point AS geom
      , pois.latitude
      , pois.longitude
FROM poi.fbpages_categ_api_v2 categ
INNER JOIN LATERAL poi.points_of_interest_by_category_v1(categ.id) pois ON true
WHERE categ.parent_id = 1896171907331974;
```

## A.10. Isolar os POIs de diferentes níveis de categorias

Devido ao facto de as consultas realizadas à base de dados para extração dos pontos de interesse das categorias de nível 2, 3, 4 e 5 envolverem a manipulação de uma grande massa de dados, foram criadas algumas tabelas, por meio da utilização do código SQL mostrado a seguir.

```
CREATE TABLE poi.pois_level2_lisbon AS (
  SELECT cat.id AS category_id
        , poi.geom_point AS geom
        , poi.autoink_pk AS poi_id
        , poi.espec_fld_latitude AS latitude
        , poi.espec_fld_longitude AS longitude
        , (st_transform(geom_point, 3857)) AS geom_3857
  FROM poi.fbpages_categ_api_v2 cat
  INNER JOIN poi.poi_categories_to_fbpage_categ_api_v1 map ON map.maped_id =
cat.id
  INNER JOIN poi.poi_vs_categories pvc ON pvc.category_id = map.original_id
  INNER JOIN poi.points_of_interest poi ON poi.autoink_pk = pvc.poi_id
  INNER JOIN borders.dgt_counties_portugal dgt ON st_within(poi.geom_point,
dgt.wkb_geometry)
  WHERE cat._level = 2
        AND dgt.concelho ~*
' (ALCOCHETE|ALMADA|AMADORA|BARREIRO|CASCAIS|LISBOA|LOURES|MAFRA|MOITA|MONTIJO|ODIVELA
S|OEIRAS|PALMELA|SEIXAL|SESIMBRA|SETÚBAL|SINTRA|VILA FRANCA DE XIRA)'
);
CREATE INDEX pois_level2_lb_categ_id_idx ON poi.pois_level2_lisbon (category_id);
CREATE INDEX pois_level2_lb_geom_idx ON poi.pois_level2_lisbon USING GIST(geom);
CREATE INDEX pois_level2_lb_geom_3857_idx ON poi.pois_level2_lisbon USING
GIST(geom_3857);

CREATE TABLE poi.pois_level3_lisbon AS (
  SELECT cat.id AS category_id
        , poi.geom_point AS geom
        , poi.autoink_pk AS poi_id
        , poi.espec_fld_latitude AS latitude
        , poi.espec_fld_longitude AS longitude
        , (st_transform(geom_point, 3857)) AS geom_3857
  FROM poi.fbpages_categ_api_v2 cat
  INNER JOIN poi.poi_categories_to_fbpage_categ_api_v1 map ON map.maped_id =
cat.id
  INNER JOIN poi.poi_vs_categories pvc ON pvc.category_id = map.original_id
  INNER JOIN poi.points_of_interest poi ON poi.autoink_pk = pvc.poi_id
```

```

INNER JOIN borders.dgt_counties_portugal dgt ON st_within(poi.geom_point,
dgt.wkb_geometry)
WHERE cat._level = 3
AND dgt.concelho ~*
'(ALCOCHETE|ALMADA|AMADORA|BARREIRO|CASCAIS|LISBOA|LOURES|MAFRA|MOITA|MONTIJO|ODIVELA
S|OEIRAS|PALMELA|SEIXAL|SESIMBRA|SETÚBAL|SINTRA|VILA FRANCA DE XIRA)'
);
CREATE INDEX pois_level3_lb_categ_id_idx ON poi.pois_level3_lisbon(category_id);
CREATE INDEX pois_level3_lb_geom_idx ON poi.pois_level3_lisbon USING GIST(geom);
CREATE INDEX pois_level3_lb_geom_3857_idx ON poi.pois_level3_lisbon USING
GIST(geom_3857);

CREATE TABLE poi.pois_level4_lisbon AS (
SELECT cat.id AS category_id
, poi.geom_point AS geom
, poi.autoink_pk AS poi_id
, poi.espec_fld_latitude AS latitude
, poi.espec_fld_longitude AS longitude
, (st_transform(geom_point, 3857)) AS geom_3857
FROM poi.fbpages_categ_api_v2 cat
INNER JOIN poi.poi_categories_to_fbpage_categ_api_v1 map ON map.maped_id =
cat.id
INNER JOIN poi.poi_vs_categories pvc ON pvc.category_id = map.original_id
INNER JOIN poi.points_of_interest poi ON poi.autoink_pk = pvc.poi_id
INNER JOIN borders.dgt_counties_portugal dgt ON st_within(poi.geom_point,
dgt.wkb_geometry)
WHERE cat._level = 4
AND dgt.concelho ~*
'(ALCOCHETE|ALMADA|AMADORA|BARREIRO|CASCAIS|LISBOA|LOURES|MAFRA|MOITA|MONTIJO|ODIVELA
S|OEIRAS|PALMELA|SEIXAL|SESIMBRA|SETÚBAL|SINTRA|VILA FRANCA DE XIRA)'
);
CREATE INDEX pois_level4_lb_categ_id_idx ON poi.pois_level4_lisbon(category_id);
CREATE INDEX pois_level4_lb_geom_idx ON poi.pois_level4_lisbon USING GIST(geom);
CREATE INDEX pois_level4_lb_geom_3857_idx ON poi.pois_level4_lisbon USING
GIST(geom_3857);

CREATE TABLE poi.pois_level5_lisbon AS (
SELECT cat.id AS category_id
, poi.geom_point AS geom
, poi.autoink_pk AS poi_id
, poi.espec_fld_latitude AS latitude
, poi.espec_fld_longitude AS longitude
, (st_transform(geom_point, 3857)) AS geom_3857
FROM poi.fbpages_categ_api_v2 cat
INNER JOIN poi.poi_categories_to_fbpage_categ_api_v1 map ON map.maped_id =
cat.id
INNER JOIN poi.poi_vs_categories pvc ON pvc.category_id = map.original_id
INNER JOIN poi.points_of_interest poi ON poi.autoink_pk = pvc.poi_id
INNER JOIN borders.dgt_counties_portugal dgt ON st_within(poi.geom_point,
dgt.wkb_geometry)
WHERE cat._level = 5
AND dgt.concelho ~*
'(ALCOCHETE|ALMADA|AMADORA|BARREIRO|CASCAIS|LISBOA|LOURES|MAFRA|MOITA|MONTIJO|ODIVELA
S|OEIRAS|PALMELA|SEIXAL|SESIMBRA|SETÚBAL|SINTRA|VILA FRANCA DE XIRA)'
);

```

```
CREATE INDEX pois_level5_lb_categ_id_idx ON poi.pois_level5_lisbon(category_id);
CREATE INDEX pois_level5_lb_geom_idx ON poi.pois_level5_lisbon USING GIST(geom);
CREATE INDEX pois_level5_lb_geom_3857_idx ON poi.pois_level5_lisbon USING
GIST(geom_3857);
```

## A.11. Filtragem dos polígonos do CORINE em território nacional

O script SQL a seguir, mostra a criação de uma nova tabela espacial na base de dados para armazenar os polígonos do CORINE *Land Cover*, que intercetam o território nacional. Para esta operação, recorreu-se ao uso da função “ST\_Intersects”, da extensão PostGIS. Da forma como foi utilizada neste caso, a função irá retornar positivo para todos os polígonos que estão completamente, ou em parte dentro do país. Esta mesma operação também pode ser feita com a utilização de um sistema GIS.

```
CREATE TABLE clc2018_portugal AS (
  SELECT clc.*
  FROM clc2018_eu_v20 clc
  INNER JOIN dgt_portugal pt ON ST_Intersects(
    st_transform(st_setsrid(clc.wkb_geometry, 3035), 4326), pt.geom)
);
```

## A.12. Criação da tabela para guardar os dados das células da grid

Após as operações realizadas no QGIS para a criação da *grid* contendo as células que representam o espaço a ser estudado na área metropolitana de Lisboa, com recurso ao *dataset* CORINE *Land Cover*, os dados foram importados para a base de dados e foi aplicada a consulta SQL mostrada abaixo para criação de uma tabela onde são guardadas as células e as suas respetivas classes de uso do solo.

```
CREATE TABLE clc.clc2018_pt_lisbon_grid_final_v2_full AS
  WITH label AS (
    SELECT grd.cell_id, MAX(class_percentage) AS percentage
    FROM clc.clc2018_pt_lisbon_grid_final_v2 grd
    GROUP BY grd.cell_id
    ORDER BY grd.cell_id
  ) SELECT lbl.cell_id
    , grd1.cell_area
    , grd2.class
    , grd2.legend
    , lbl.percentage
    , grd1.wkb_geometry
    , ST_PointOnSurface(grd1.wkb_geometry) AS centroid
  FROM clc.clc2018_pt_lisbon_grid_intersected grd1
```

```

INNER JOIN label lbl ON lbl.cell_id = grd1.fid
INNER JOIN clc.clc2018_pt_lisbon_grid_final_v2 grd2 ON grd2.cell_id = grd1.fid
AND grd2.class_percentage = lbl.percentage;

CREATE INDEX clc2018_pt_lb_grd_fn_v2_full_wkb_geom_idx ON
clc.clc2018_pt_lisbon_grid_final_v2_full USING GIST(wkb_geometry);
CREATE INDEX clc2018_pt_lb_grd_fn_v2_full_centroid_idx ON
clc.clc2018_pt_lisbon_grid_final_v2_full USING GIST(centroid);

```

### A.13. Filtragem das categorias dos POIs da área delimitada pelos polígonos da classe “Artificial Surfaces” do CORINE

Um dos métodos utilizados para a geração de atributos a serem utilizados nas experiências com o CORINE, baseou-se nas categorias dos pontos de interesse localizados dentro das áreas delimitadas pelos polígonos da classe “Artificial Surfaces”. O script SQL a seguir, mostra como foram obtidas tais categorias e também como foi criada a tabela “pois\_clc\_artificial\_surfaces\_v2”, que contém todos os pontos de interesse vinculados às mesmas.

```

CREATE TABLE clc.clc2018_pt_artificial_surfaces AS (
  SELECT objectid
    , code_18
    , remark
    , area_ha
    , id
    , c18
    , ogc_fid
    , grid_code
    , clc_code
    , label1
    , label2
    , label3
    , clc.centroid
    , ST_Transform(wkb_geometry, 4326) AS wkb_geometry
  FROM clc.clc2018_portugal clc
  INNER JOIN clc.clc_legend leg ON leg.clc_code = clc.code_18
  WHERE leg.label1 ILIKE 'Artificial surfaces'
);

CREATE TABLE poi.pois_clc_artificial_surfaces_v1 AS
  SELECT p.*
    FROM poi.points_of_interest p
    INNER JOIN clc.clc2018_pt_artificial_surfaces a ON ST_Within(p.geom_point,
a.wkb_geometry);

CREATE TABLE poi.categories_clc_artificial_surfaces AS (
  WITH base_categories AS (
    SELECT DISTINCT ON (cat.id) cat.id AS category_id
      , cat.name AS category_name
      , cat_level

```

```

    , cat.parent_id
  FROM poi.pois_clc_artificial_surfaces_v1 p
  INNER JOIN poi.poi_vs_categories pvc ON pvc.poi_id = p.autoink_pk
  INNER JOIN poi.poi_categories pca ON pca.category_id = pvc.category_id
  INNER JOIN poi.poi_categories_to_fbpage_categ_api_v1 map ON map.original_id
= pca.category_id
  INNER JOIN poi.fbpages_categ_api_v2 cat ON cat.id = map.maped_id
  WHERE cat._level > 1
), categories AS (
  SELECT base.category_id
    , base.category_name
    , CASE WHEN base._level = 2 THEN base.category_id
      WHEN base._level = 3 THEN n5.id
      WHEN base._level = 4 THEN n4.id
      WHEN base._level = 5 THEN n3.id
      ELSE n2.id END AS parent_id
    , CASE WHEN base._level = 2 THEN base.category_name
      WHEN base._level = 3 THEN n5.name
      WHEN base._level = 4 THEN n4.name
      WHEN base._level = 5 THEN n3.name
      ELSE n2.name END AS parent_name
    , CASE WHEN base._level = 2 THEN base._level
      WHEN base._level = 3 THEN n5._level
      WHEN base._level = 4 THEN n4._level
      WHEN base._level = 5 THEN n3._level
      ELSE n2._level END AS parent_level
  FROM base_categories base
  LEFT JOIN poi.fbpages_categ_api_v2 n5 ON n5.id = base.parent_id
  LEFT JOIN poi.fbpages_categ_api_v2 n4 ON n4.id = n5.parent_id
  LEFT JOIN poi.fbpages_categ_api_v2 n3 ON n3.id = n4.parent_id
  LEFT JOIN poi.fbpages_categ_api_v2 n2 ON n2.id = n3.parent_id
) SELECT DISTINCT ON (cat.parent_id) cat.parent_id AS category_id
    , cat.parent_name AS category_name
    , cat.parent_level AS category_level
  FROM categories cat
);

CREATE TABLE poi.pois_clc_artificial_surfaces_v2 AS
  SELECT categ.category_id
    , pois.latitude
    , pois.longitude
    , pois.geom_point AS geom
    , st_transform(pois.geom_point, 3857) AS geom_3857
  FROM poi.categories_clc_artificial_surfaces categ
  INNER JOIN LATERAL poi.points_of_interest_by_category_v1(categ.category_id) pois
ON true
  INNER JOIN poi.pois_clc_artificial_surfaces_v1 pasu ON pasu.autoink_pk =
pois.poi_autoink_pk;

```

## A.14. Seleção manual de categorias

De maneira a selecionar as categorias mais semelhantes àquelas utilizadas por Calegari et al. [9], foi realizada uma consulta à base de dados para obter uma lista aproximada daquilo que se pretendia. Em seguida, por meio de uma inspeção visual aos resultados retornados, foram escolhidas as categorias, originando uma nova tabela na base de dados. O quadro a seguir mostra os comandos utilizados para esta operação.

```
SELECT *
FROM poi.poi_categories
WHERE category_name ~*
'(Bank|Bar|BookShop|Cinema|Clothes|DepartmentStore|Doctors|Emergency
|Farm|FastFood|FuelStation|Furniture|Gym|Hospital|Hotel|Library|Marketplace
|Nightclub|Park|Parking|Pharmacy|PlaceOfWorship|Police|Pub|Restaurant|School|Supermar
ket|University)';

CREATE TABLE poi.selected_categories_manual AS
SELECT cat.id AS category_id, cat.name AS category_name
FROM poi.poi_categories pca
INNER JOIN poi.poi_categories_to_fbpage_categ_api_v1 map ON map.original_id =
pca.category_id
INNER JOIN poi.fbpages_categ_api_v2 cat ON cat.id = map.maped_id
WHERE category_id IN (2219, 2601, 2602, 2629, 2637, 110152272401235,
110207655727714, 110290705711626,
115090141929327, 124887510918208, 127806747288007,
133152263416981, 133576170041936,
134381433294944, 139386576124160, 147714868971098,
164243073639257, 169758603141095,
169896676390222, 184405378265823, 192803624072087,
199405806739848, 200027430011994,
200121186665577, 218693881483234, 436846369665656,
923358744476601, 957437364355057,
1153901557995608);
```

## A.15. Preparação dos dados para geração dos heatmaps

Os comandos SQL mostrados no quadro a seguir, foram utilizados para: (1) criar uma tabela na base de dados, com todas as categorias que se mostraram mais relevantes na última experiência apresentada no tópico 0 e (2) isolar os pontos de interesse geograficamente localizados na região metropolitana de Lisboa.

```
CREATE TEMPORARY TABLE lulc_poi_correlated_categories (
category TEXT
);
COPY lulc_poi_correlated_categories FROM '/tmp/lulc_poi_correlated_categories.txt'
WITH (FORMAT TEXT);
```

```

CREATE TABLE poi.lulc_poi_correlated_categories AS
  SELECT cat.id
         , cat.name
         , Lower(regex_replace(cat.name, '^[^a-zA-Z0-9]+', '_', 'g')) AS
sanitized_name
         , cat.parent_id
         , cat._level
         FROM poi.fbpages_categ_api_v2 cat
         WHERE Lower(regex_replace(cat.name, '^[^a-zA-Z0-9]+', '_', 'g')) IN (SELECT
category FROM lulc_poi_correlated_categories)
         ORDER BY name;

CREATE TABLE poi.points_of_interest_lb_metrop_area AS
SELECT *
  FROM poi.points_of_interest p
  INNER JOIN borders.dgt_lisbon_metropolitan_area b ON st_within(p.geom_3857,
b.wkb_geometry);

```

## A.16. Criação das tabelas que representam as classes do CORINE

Para permitir a operação de criação dos *rasters* no QGIS, representando cada categoria de uso do solo que seria posteriormente estudada, foram criadas tabelas distintas na base de dados, contendo os polígonos que se pretendia em cada caso. Para tal, foi utilizado o *script SQL* mostrado a seguir.

```

CREATE TABLE clc.clc2018_pt_artificial_areas AS
  SELECT objectid, code_18, c18, class, legend, wkb_geometry, 1 AS weight
         FROM clc.clc2018_pt_lisbon_metropolitan_area_v2 clc
         WHERE clc.class BETWEEN 100 AND 199;

CREATE TABLE clc.clc2018_pt_urban_fabric AS
  SELECT objectid, code_18, c18, class, legend, wkb_geometry, 1 AS weight
         FROM clc.clc2018_pt_lisbon_metropolitan_area_v2 clc
         WHERE clc.class IN (111, 112);

CREATE TABLE clc.clc2018_pt_cont_urban_fabric AS
  SELECT objectid, code_18, c18, class, legend, wkb_geometry, 1 AS weight
         FROM clc.clc2018_pt_lisbon_metropolitan_area_v2 clc
         WHERE clc.class = 111;

CREATE TABLE clc.clc2018_pt_discont_urban_fabric AS
  SELECT objectid, code_18, c18, class, legend, wkb_geometry, 1 AS weight
         FROM clc.clc2018_pt_lisbon_metropolitan_area_v2 clc
         WHERE clc.class = 112;

CREATE TABLE clc.clc2018_pt_industrial_comercial AS
  SELECT objectid, code_18, c18, class, legend, wkb_geometry, 1 AS weight
         FROM clc.clc2018_pt_lisbon_metropolitan_area_v2 clc
         WHERE clc.class = 121;

```



# Apêndice B. Formatos dos ficheiros utilizados

## B.1. Ficheiro de categorias obtidas no site do Facebook

A Figura 21 mostra o formato dos dados obtidos no site do Facebook após tratamento. Considerando que a lista disponível online, embora organizada hierarquicamente, não possui qualquer numeração nos itens, foi necessária alguma manipulação dos dados para se chegar ao formato apresentado na imagem. O tratamento aplicado é explicado na secção 4.4.

	A	B
1	code	category_name
2	1	Businesses
3	1.1	Advertising/Marketing
4	1.1.1	Advertising Agency
5	1.1.2	Copywriting Service
6	1.1.3	Internet Marketing Service
7	1.1.4	Market Research Consultant
8	1.1.5	Marketing Agency
9	1.1.6	Media Agency
10	1.1.7	Merchandising Service
11	1.1.8	Public Relations Agency
12	1.1.9	Social Media Agency
13	1.1.10	Telemarketing Service
14	1.2	Agriculture

Figura 21: Lista ordenada das categorias obtidas no site do Facebook.

## B.2. Mapa de categorias

Conforme mencionado na secção 4.1.2.2.2.1, foi realizado um mapeamento manual das categorias extraídas dos pontos de interesse que não existem na lista obtida via API, no *endpoint* “fb\_page\_categories”.

id	name	map_to_id	name	parent_id
1113	Fictional Character	1602034176774683	Actor	1602034176774683
1308	Literary Editor	622132518208039	Editor	1602034176774683
2000	Food	273819889375819	Restaurant	1562965077339698
2200	Company	1896171907331974	Businesses	null
2243	Medical Company	145118935550090	Medical & Health	1896171907331974
2246	Agriculture Company	1574325646194878	Agriculture	1896171907331974
2252	Food & Beverage Company	1562965077339698	Food & Beverage	1757592557789532
2500	Local Business	1758418281071392	Local Service	1896171907331974
2600	Organization	369730359717478	Other	
2632	Pet	144982405562750	Pet Service	1758418281071392
2717	Personal Website	2700	Personal Blog	2202
132942110106445	Dental Equipment	143491166059053	Medical Equipment Supplier	145118935550090
152142351517013	Corporate Office	1896171907331974	Businesses	
162813670434541	Market	1153901557995608	Supermarket	150108431712141
170992992946914	Marketing Consultant	123377808095874	Marketing Agency	1757592557789532
176831012360626	Professional Service	1758418281071392	Local Service	1896171907331974
177721448951559	Workplace & Office	139745066094977	Business Center	492376511113483
190648534311833	Recycling & Waste Management	976586885773142	Recycling Center	806240166180298
		121989644542468	Waste Management Company	147714868971098
197097220301977	Historical Place	209889829023118	Landmark & Historical Place	683513901834713
205854706429555	Licensed Financial Representative	161516070564222	Financial Service	1022050661163852
265274547167554	Architectural Engineer	260609830972978	Structural Engineer	297544187300691
360320934014496	Lifestyle Service	1758418281071392	Local Service	1896171907331974
620655754769617	Education Company	2250	Education	1896171907331974
821084571355604	Fashion Company	557045641143373	Design & Fashion	1758418281071392
989113024538740	Farm Equipment Service	165245226927543	Commercial & Industrial Equipment Su	243290832429433
1025733874177680	Home Improvement Service	108427109235243	Home Improvement	1758418281071392
1126714320723207	Landscape Architect	128853933850116	Landscape Company	108427109235243
1422088814483501	Marine	114253171987086	Marine Supply Store	1628617144122365
1566895303614991	Technology Company	297544187300691	Science, Technology & Engineering	1896171907331974
1722308648017694	Blacksmith	816612951820333	Metal & Steel Company	243290832429433
1736039333278498	Brand Agency	123377808095874	Marketing Agency	1757592557789532
1745787725638128	Landscape Designer	1711765319078442	Architectural Designer	108427109235243
1897630797130989	Beauty/Cosmetics Company	139225689474222	Beauty, Cosmetic & Personal Care	1896171907331974

Figura 22: Mapa de categorias.

# Apêndice C. Modelo da base de dados

Este apêndice é dedicado às informações relacionadas à estrutura da base de dados desenvolvida e utilizada no estudo.

## C.1. Modelo inicial

A Figura 23 apresenta o modelo inicial da base de dados, contendo apenas as principais tabelas, onde foram armazenados os pontos de interesse, tal como as suas respectivas categorias, e os dados do CORINE e do COS.

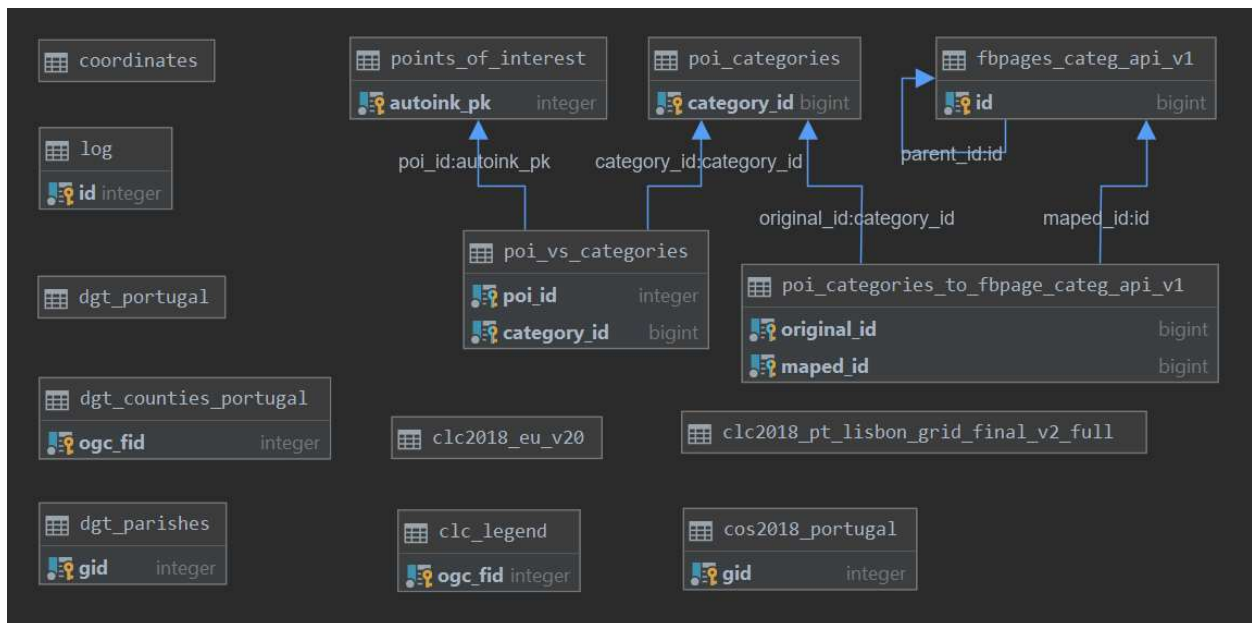


Figura 23: Principais tabelas da base de dados.



# Apêndice D. Parametrizações

Neste apêndice são apresentados os valores utilizados nos parâmetros envolvidos nos algoritmos utilizados no estudo.

## D.1. Classificadores

A seguir, na Tabela 61, são apresentados os parâmetros utilizados nos vários classificadores que foram adotados para a caracterização dos espaços geográficos.

Tabela 61: Valores dos parâmetros utilizados nos classificadores.

<b>Classificador</b>	<b>Parâmetro</b>	<b>Valor</b>
<b>Naïve Bayes - NB</b>	Laplace correction	True
<b>Random Forest - RF</b>	Number of trees	100
	Criterion	Gain ratio
	Maximal deph	10
	Gess subset ratio	True
	Voting strategy	Confidence
<b>Support Vector Machine - SVM</b>	SVM type	C-SVC
	Kernel type	rbf
	Gamma	0
	C	0
	Cache size	80
	Epsilon	0.001
	Shrinking	True
<b>Neural Network - NN</b>	Training cycles	200
	Learning rate	0.01
	Momentum	0.9
	Shuffle	True
	Normalize	True
	Error epsilon	1.00E-04



# Apêndice E. Matrizes de confusão

## E.1. Classes “Industrial”, “Comercial” e “Instalações agrícolas”, do *dataset* COS e 22 atributos

As figuras a seguir mostram as matrizes de correlação obtidas na experiência de classificação realizada utilizando-se os polígonos do *dataset* COS, com base nos 22 atributos que representam a distância entre o centroide de cada polígono e o ponto de interesse mais próximo de cada subcategoria de nível 2 da categoria “Business”. Este cenário considerou 3 classes de uso do solo: “industrial”, “comercial” e “instalações agrícolas”.

	true Industrial	true Comercial	true Instalacoes Agricolas	class precision
pred. Industrial	307	109	315	42.00%
pred. Comercial	827	1075	158	52.18%
pred. Instalacoes Agricolas	116	66	777	81.02%
class recall	24.56%	86.00%	62.16%	

Figura 24: Matriz de confusão do NB com *cross-validation*, no cenário 1, do COS.

	true Industrial	true Comercial	true Instalacoes Agricolas	class precision
pred. Industrial	162	62	138	44.75%
pred. Comercial	937	1141	240	49.22%
pred. Instalacoes Agricolas	151	47	872	81.50%
class recall	12.96%	91.28%	69.76%	

Figura 25: Matriz de confusão do RF com *cross-validation n*, no cenário 1, do COS.

	true Industrial	true Comercial	true Instalacoes Agricolas	class precision
pred. Industrial	23	16	229	8.58%
pred. Comercial	1205	1221	666	39.49%
pred. Instalacoes Agricolas	22	13	355	91.03%
class recall	1.84%	97.68%	28.40%	

Figura 26: Matriz de confusão do SVM com *cross-validation*, no cenário 1, do COS.

	true Industrial	true Comercial	true Instalacoes Agricolas	class precision
pred. Industrial	606	283	220	54.64%
pred. Comercial	436	885	65	63.85%
pred. Instalacoes Agricolas	208	82	965	76.89%
class recall	48.48%	70.80%	77.20%	

Figura 27: Matriz de confusão do NN com *cross-validation*, no cenário 1, do COS.

	true Industrial	true Comercial	true Instalacoes Agricolas	class precision
pred. Industrial	88	37	93	40.37%
pred. Comercial	256	319	48	51.20%
pred. Instalacoes Agricolas	31	19	234	82.39%
class recall	23.47%	85.07%	62.40%	

Figura 28: Matriz de confusão do NB, com *split validation*, no cenário 1, do COS.

	true Industrial	true Comercial	true Instalacoes Agricolas	class precision
pred. Industrial	247	142	74	53.35%
pred. Comercial	80	219	16	69.52%
pred. Instalacoes Agricolas	48	14	285	82.13%
class recall	65.87%	58.40%	76.00%	

Figura 29: Matriz de confusão do NN com *split validation*, no cenário 1, do COS.

## E.2. Classes “Industrial”, “Comercial” e “Instalações agrícolas”, do dataset COS e 44 atributos

As imagens abaixo mostram as matrizes de correlação obtidas na experiência de classificação realizada utilizando-se os polígonos do *dataset* COS, com base em 44 atributos, dos quais 22 representam a distância entre o centroide de cada polígono e o ponto de interesse mais próximo de cada subcategoria de nível 2 da categoria “Business”, e os restantes 22 são essencialmente a quantidade de pontos de interesse de cada uma dessas subcategorias, localizados dentro de um raio de 10000 metros a partir do centroide de cada polígono. Este cenário, tal como o anterior, considerou 3 classes de uso do solo: “industrial”, “comercial” e “instalações agrícolas”.

	true Industrial	true Comercial	true Instalacoes Agricolas	class precision
pred. Industrial	339	119	250	47.88%
pred. Comercial	682	982	111	55.32%
pred. Instalacoes Agricolas	229	149	889	70.17%
class recall	27.12%	78.56%	71.12%	

Figura 30: Matriz de confusão do NB com *cross-validation*, no cenário 2, do COS.

	true Industrial	true Comercial	true Instalacoes Agricolas	class precision
pred. Industrial	174	51	135	48.33%
pred. Comercial	932	1155	261	49.19%
pred. Instalacoes Agricolas	144	44	854	81.96%
class recall	13.92%	92.40%	68.32%	

Figura 31: Matriz de confusão do RF com *cross-validation*, no cenário 2, do COS.

	true Industrial	true Comercial	true Instalacoes Agricolas	class precision
pred. Industrial	614	393	217	50.16%
pred. Comercial	382	658	130	56.24%
pred. Instalacoes Agricolas	254	199	903	66.59%
class recall	49.12%	52.64%	72.24%	

Figura 32: Matriz de confusão do SVM com *cross-validation* no cenário 2, do COS.

	true Industrial	true Comercial	true Instalacoes Agricolas	class precision
pred. Industrial	611	291	220	54.46%
pred. Comercial	429	885	70	63.95%
pred. Instalacoes Agricolas	210	74	960	77.17%
class recall	48.88%	70.80%	76.80%	

Figura 33: Matriz de confusão do NN com *cross-validation*, no cenário 2, do COS.

	true Industrial	true Comercial	true Instalacoes Agricolas	class precision
pred. Industrial	96	32	74	47.52%
pred. Comercial	227	300	37	53.19%
pred. Instalacoes Agricolas	52	43	264	73.54%
class recall	25.60%	80.00%	70.40%	

Figura 34: Matriz de confusão do NB com *split validation*, no cenário 2, do COS.

	true Industrial	true Comercial	true Instalacoes Agricolas	class precision
pred. Industrial	220	118	84	52.13%
pred. Comercial	98	233	14	67.54%
pred. Instalacoes Agricolas	57	24	277	77.37%
class recall	58.67%	62.13%	73.87%	

Figura 35: Matriz de confusão do NN com *split validation*, no cenário 2, do COS.

### E.3. Classes “Indústria e comércio” e “Instalações agrícolas”, do dataset COS e 44 atributos

As figuras apresentadas a seguir mostram as matrizes de correlação obtidas na experiência de classificação realizada utilizando-se os polígonos do *dataset* COS, com base em 44 atributos, dos quais 22 representam a distância entre o centroide de cada polígono e o ponto de interesse mais próximo de cada subcategoria de nível 2 da categoria “Business”, e os restantes 22 são essencialmente a quantidade de pontos de interesse de cada uma dessas subcategorias, localizados dentro de um raio de 10000 metros a partir do centroide de cada polígono. Este cenário, diferente dos anteriores, considerou 2 classes de uso do solo: “indústria e comércio” e “instalações agrícolas”.

	true Industria e comercio	true Instalacoes Agricolas	class precision
pred. Industria e comercio	2082	650	76.21%
pred. Instalacoes Agricolas	418	1850	81.57%
class recall	83.28%	74.00%	

Figura 36: Matriz de confusão do NB com *cross-validation*, no cenário 3, do COS.

	true Industria e comercio	true Instalacoes Agricolas	class precision
pred. Industria e comercio	2249	602	78.88%
pred. Instalacoes Agricolas	251	1898	88.32%
class recall	89.96%	75.92%	

Figura 37: Matriz de confusão do RF com *cross-validation*, no cenário 3, do COS.

	true Indústria e comercio	true Instalacoes Agricolas	class precision
pred. Indústria e comercio	1915	480	79.96%
pred. Instalacoes Agricolas	585	2020	77.54%
class recall	76.60%	80.80%	

Figura 38: Matriz de confusão do SVM com *cross-validation*, no cenário 3, do COS.

	true Indústria e comercio	true Instalacoes Agricolas	class precision
pred. Indústria e comercio	2167	431	83.41%
pred. Instalacoes Agricolas	333	2069	86.14%
class recall	86.68%	82.76%	

Figura 39: Matriz de confusão do NN com *cross-validation*, no cenário 3, do COS.

	true Indústria e comercio	true Instalacoes Agricolas	class precision
pred. Indústria e comercio	620	196	75.98%
pred. Instalacoes Agricolas	130	554	80.99%
class recall	82.67%	73.87%	

Figura 40: Matriz de confusão do NB com *split validation*, no cenário 3, do COS.

	true Indústria e comercio	true Instalacoes Agricolas	class precision
pred. Indústria e comercio	574	73	88.72%
pred. Instalacoes Agricolas	176	677	79.37%
class recall	76.53%	90.27%	

Figura 41: Matriz de confusão do NN com *split validation*, no cenário 3, do COS.

## E.4. Classes “tecido edificado” e “indústria e comércio”, do dataset COS e 44 atributos

Nesta secção são apresentadas as matrizes de correlação obtidas numa experiência de classificação do uso do solo que considerou as classes “indústria e comércio” e “tecido edificado”. A experiência foi realizada utilizando-se os polígonos do *dataset* COS, com base em 44 atributos, dos quais 22 representam a distância entre o centroide de cada polígono e o ponto de interesse mais próximo de cada subcategoria de nível 2 da categoria “Business”, e os restantes 22 são essencialmente a quantidade de pontos de interesse de cada uma dessas subcategorias, localizados dentro de um raio de 10000 metros a partir do centroide de cada polígono. É importante ressaltar que este cenário difere-se dos demais devido ao uso da classe “tecido edificado”.

	true Industrial ou Comercial	true Tecido Edificado	class precision
pred. Industrial ou Comercial	3481	2471	58.48%
pred. Tecido Edificado	919	1929	67.73%
class recall	79.11%	43.84%	

Figura 42: Matriz de confusão do NB com *cross-validation*, no cenário 4, do COS.

	true Industrial ou Comercial	true Tecido Edificado	class precision
pred. Industrial ou Comercial	4085	3374	54.77%
pred. Tecido Edificado	315	1026	76.51%
class recall	92.84%	23.32%	

Figura 43: Matriz de confusão do RF com *cross-validation* no cenário 4, do COS.

	true Industrial ou Comercial	true Tecido Edificado	class precision
pred. Industrial ou Comercial	2667	1604	62.44%
pred. Tecido Edificado	1733	2796	61.74%
class recall	60.61%	63.55%	

Figura 44: Matriz de confusão do SVM com *cross-validation*, no cenário 4, do COS.

	true Industrial ou Comercial	true Tecido Edificado	class precision
pred. Industrial ou Comercial	3281	2023	61.86%
pred. Tecido Edificado	1119	2377	67.99%
class recall	74.57%	54.02%	

Figura 45: Matriz de confusão do NN com *cross-validation*, no cenário 4, do COS.

	true Industrial ou Comercial	true Tecido Edificado	class precision
pred. Industrial ou Comercial	1043	727	58.93%
pred. Tecido Edificado	277	593	68.16%
class recall	79.02%	44.92%	

Figura 46: Matriz de confusão do NB com *split validation*, no cenário 4, do COS.

	true Industrial ou Comercial	true Tecido Edificado	class precision
pred. Industrial ou Comercial	927	528	63.71%
pred. Tecido Edificado	393	792	66.84%
class recall	70.23%	60.00%	

Figura 47: Matriz de confusão do NN com *split validation*, no cenário 4, do COS.

## E.5. Classes “tecido edificado” e “indústria e comércio”, do dataset COS e 44 atributos

A seguir são apresentadas várias imagens onde se pode observar as matrizes de correlação obtidas numa experiência de classificação do uso do solo que, tal como no teste anterior, considerou as classes “indústria e comércio” e “tecido edificado”. A experiência foi realizada utilizando-se os polígonos do *dataset* COS, com base em 64 atributos, selecionados de forma automática. Os atributos utilizados são descritos na secção 5.1.2.5.

	true Tecido Edificado	true Industrial ou Comercial	class precision
pred. Tecido Edificado	502	302	62.44%
pred. Industrial ou Comercial	858	1058	55.22%
class recall	36.91%	77.79%	

Figura 48: Matriz de confusão do NB com *cross-validation*, no cenário 5, do COS.

	true Tecido Edificado	true Industrial ou Comercial	class precision
pred. Tecido Edificado	282	127	68.95%
pred. Industrial ou Comercial	1078	1233	53.35%
class recall	20.74%	90.66%	

Figura 49: Matriz de confusão do RF com *cross-validation*, no cenário 5, do COS.

	true Tecido Edificado	true Industrial ou Comercial	class precision
pred. Tecido Edificado	803	568	58.57%
pred. Industrial ou Comercial	557	792	58.71%
class recall	59.04%	58.24%	

Figura 50: Matriz de confusão do SVM com *cross-validation* no cenário 5, do COS.

	true Tecido Edificado	true Industrial ou Comercial	class precision
pred. Tecido Edificado	855	519	62.23%
pred. Industrial ou Comercial	505	841	62.48%
class recall	62.87%	61.84%	

Figura 51: Matriz de confusão do NN com *cross-validation*, no cenário 5, do COS.

	true Tecido Edificado	true Industrial ou Comercial	class precision
pred. Tecido Edificado	135	84	61.64%
pred. Industrial ou Comercial	273	324	54.27%
class recall	33.09%	79.41%	

Figura 52: Matriz de confusão do NB com *split validation*, no cenário 5, do COS.

	true Tecido Edificado	true Industrial ou Comercial	class precision
pred. Tecido Edificado	270	193	58.32%
pred. Industrial ou Comercial	138	215	60.91%
class recall	66.18%	52.70%	

Figura 53: Matriz de confusão do NN com *split validation*, no cenário 5, do COS.

## E.6. Matrizes de confusão obtidas nos testes com o CORINE

Nesta secção são apresentadas as matrizes de confusão obtidas através dos testes realizados aos modelos treinados nos diversos cenários de teste realizados com base na utilização do CORINE *Land Cover*. Cada cenário envolveu diferentes conjuntos de atributos, extraídos com base em diferentes categorias dos pontos de interesse, conforme abordado na secção 5.1.3.

	true Florestas e áreas se...	true Agricultura	true Áreas urbanas desco...	true Áreas urbanas contin...	true Industrial / comercial e...	class precision
pred. Florestas e áreas se...	2135	551	209	6	87	71.45%
pred. Agricultura	461	1891	322	17	98	67.80%
pred. Áreas urbanas desco...	213	322	1723	194	157	66.04%
pred. Áreas urbanas conti...	39	49	405	2615	225	78.46%
pred. Industrial / comercial...	152	187	341	168	2100	71.23%
class recall	71.17%	63.03%	57.43%	87.17%	78.74%	

Figura 54: Matriz de confusão obtida no cenário 1, do CORINE.

	true Florestas e áreas se...	true Agricultura	true Áreas urbanas desco...	true Áreas urbanas contin...	true Industrial / comercial e...	class precision
pred. Florestas e áreas se...	2076	486	204	18	78	72.54%
pred. Agricultura	493	1939	370	18	90	66.63%
pred. Áreas urbanas desc...	266	380	1842	207	173	64.23%
pred. Áreas urbanas conti...	37	53	316	2617	192	81.40%
pred. Industrial / comercial...	128	142	268	140	2134	75.89%
class recall	69.20%	64.63%	61.40%	87.23%	80.01%	

Figura 55: Matriz de confusão obtida no cenário 2, do CORINE.

	true Florestas e áreas se...	true Agricultura	true Áreas urbanas desco...	true Áreas urbanas contin...	true Industrial / comercial e...	class precision
pred. Florestas e áreas se...	2111	512	271	11	76	70.82%
pred. Agricultura	459	1904	417	12	94	65.97%
pred. Áreas urbanas desc...	232	322	1617	175	135	65.18%
pred. Áreas urbanas conti...	52	84	369	2652	227	78.37%
pred. Industrial / comercial...	146	178	326	150	2135	72.74%
class recall	70.37%	63.47%	53.90%	88.40%	80.05%	

Figura 56: Matriz de confusão obtida no cenário 2, do CORINE.

	true Florestas e áreas se...	true Agricultura	true Áreas urbanas desco...	true Áreas urbanas contin...	true Industrial / comercial e...	class precision
pred. Florestas e áreas se...	2101	506	226	2	78	72.12%
pred. Agricultura	481	1883	317	12	106	67.27%
pred. Áreas urbanas desc...	264	390	1806	174	206	63.59%
pred. Áreas urbanas conti...	33	46	341	2607	179	81.32%
pred. Industrial / comercial...	121	175	310	205	2098	72.12%
class recall	70.03%	62.77%	60.20%	86.90%	78.67%	

Figura 57: matriz de confusão obtida no cenário 4, do CORINE.

	true Áreas urbanas descontinuas	true Agricultura	true Áreas urbanas contínuas	true Industrial / comercial e inst...	class precision
pred. Áreas urbanas descontinuas...	2866	681	257	294	69.94%
pred. Agricultura	466	3052	15	130	83.32%
pred. Áreas urbanas contínuas	409	83	2685	233	78.74%
pred. Industrial / comercial e ins...	259	184	86	2010	79.17%
class recall	71.65%	76.30%	88.24%	75.37%	

Figura 58: Matriz de confusão obtida no teste 1, do cenário 5, do CORINE.

	true Florestas e áreas semi-nat...	true Áreas urbanas descontinuas	true Áreas urbanas contínuas	true Industrial / comercial e inst...	class precision
pred. Florestas e áreas semi-n...	3298	348	7	126	87.27%
pred. Áreas urbanas descontinuas...	473	2811	225	274	74.31%
pred. Áreas urbanas contínuas	58	490	2675	284	76.28%
pred. Industrial / comercial e ins...	171	351	136	1983	75.09%
class recall	82.45%	70.28%	87.91%	74.35%	

Figura 59: Matriz de confusão obtida no teste 2, do cenário 5, do CORINE.

	true Florestas e áreas semi-nat...	true Agricultura	true Áreas urbanas contínuas	true Industrial / comercial e inst...	class precision
pred. Florestas e áreas semi-n...	3163	987	55	134	72.90%
pred. Agricultura	596	2645	40	157	76.93%
pred. Áreas urbanas contínuas	65	96	2740	229	87.54%
pred. Industrial / comercial e ins...	176	272	208	2147	76.60%
class recall	79.07%	66.12%	90.04%	80.50%	

Figura 60: Matriz de confusão obtida no teste 3, do cenário 5, do CORINE.

## Classificação e Análise do Uso do Solo com Base em Pontos de Interesse

	true Florestas e áreas semi-nat...	true Agricultura	true Áreas urbanas descontinuas	true Áreas urbanas contínuas	class precision
pred. Florestas e áreas semi-n...	3023	762	351	14	72.84%
pred. Agricultura	581	2586	504	21	70.04%
pred. Áreas urbanas descontinuu...	338	561	2710	295	69.42%
pred. Áreas urbanas contínuas	58	91	435	2713	82.29%
class recall	75.58%	64.65%	67.75%	89.16%	

Figura 61: Matriz de confusão obtida no teste 4, do cenário 5, do CORINE.

	true Agricultura	true Áreas urbanas contínuas	true Industrial / comercial e instalações...	class precision
pred. Agricultura	3662	38	209	93.88%
pred. Áreas urbanas contínuas	118	2846	207	89.75%
pred. Industrial / comercial e instalaçõe...	220	159	2251	85.59%
class recall	91.55%	93.53%	84.40%	

Figura 62: Matriz de confusão obtida no teste 5, do cenário 5, do CORINE.

	true Áreas urbanas contínuas e descon...	true Agricultura	true Industrial / comercial e instalações...	class precision
pred. Áreas urbanas contínuas e desco...	3247	411	340	81.22%
pred. Agricultura	395	3323	117	86.65%
pred. Industrial / comercial e instalaçõe...	358	266	2210	77.98%
class recall	81.17%	83.08%	82.86%	

Figura 63: Matriz de confusão obtida no teste 6, do cenário 5, do CORINE.

	true Agricultura, florestas e áreas semi...	true Áreas urbanas contínuas e descon...	true Industrial / comercial e instalações...	class precision
pred. Agricultura, florestas e áreas sem...	3363	449	149	84.90%
pred. Áreas urbanas contínuas e desco...	397	3199	346	81.15%
pred. Industrial / comercial e instalaçõe...	240	352	2172	78.58%
class recall	84.08%	79.97%	81.44%	

Figura 64: Matriz de confusão obtida no teste 7, do cenário 5, do CORINE.

# Apêndice F. Conjuntos de atributos

Neste apêndice são apresentados os vários conjuntos de atributos utilizados no estudo para a caracterização dos espaços geográficos.

## F.1. Atributos utilizados no cenário 1 dos testes baseados no COS

Na Tabela 62, são apresentados os atributos utilizados no cenário 1, dos testes realizados com base nos dados do COS. Este conjunto foi também utilizado como parte do grupo de atributos na qual se baseou a primeira experiência baseada no CORINE.

Tabela 62: Conjunto de atributos 1.

Atributo	Descrição
min_d_advertising_marketing	Distância entre o centróide do polígono e o ponto de interesse mais próximo da categoria "Advertising/Marketing"
min_d_agriculture	Distância entre o centróide do polígono e o ponto de interesse mais próximo da categoria "Agriculture"
min_d_commercial_industrial	Distância entre o centróide do polígono e o ponto de interesse mais próximo da categoria "Commercial & Industrial"
min_d_education	Distância entre o centróide do polígono e o ponto de interesse mais próximo da categoria "Education"
min_d_finance	Distância entre o centróide do polígono e o ponto de interesse mais próximo da categoria "Finance"
min_d_food_beverage	Distância entre o centróide do polígono e o ponto de interesse mais próximo da categoria "Food & Beverage"
min_d_hotel_lodging	Distância entre o centróide do polígono e o ponto de interesse mais próximo da categoria "Hotel & Lodging"
min_d_legal	Distância entre o centróide do polígono e o ponto de interesse mais próximo da categoria "Legal"
min_d_local_service	Distância entre o centróide do polígono e o ponto de interesse mais próximo da categoria "Local Service"
min_d_media_news_company	Distância entre o centróide do polígono e o ponto de interesse mais próximo da categoria "Media/News Company"
min_d_medical_health	Distância entre o centróide do polígono e o ponto de interesse mais próximo da categoria "Medical & Health"
min_d_non_governmental_organization_ngo_	Distância entre o centróide do polígono e o ponto de interesse mais próximo da categoria "Non-Governmental Organization (NGO)"
min_d_nonprofit_organization	Distância entre o centróide do polígono e o ponto de interesse mais próximo da categoria "Nonprofit Organization"
min_d_public_government_service	Distância entre o centróide do polígono e o ponto de interesse mais próximo da categoria "Public & Government Service"
min_d_science_technology_engineering	Distância entre o centróide do polígono e o ponto de interesse mais próximo da categoria "Science, Technology & Engineering"
min_d_beauty_cosmetic_personal_care	Distância entre o centróide do polígono e o ponto de interesse mais próximo da categoria "Beauty, Cosmetic & Personal Care"
min_d_arts_entertainment	Distância entre o centróide do polígono e o ponto de interesse mais próximo da categoria "Arts & Entertainment"

min_d_automotive_aircraft_boat	Distância entre o centróide do polígono e o ponto de interesse mais próximo da categoria "Automotive, Aircraft & Boat"
min_d_real_estate	Distância entre o centróide do polígono e o ponto de interesse mais próximo da categoria "Real Estate"
min_d_shopping_retail	Distância entre o centróide do polígono e o ponto de interesse mais próximo da categoria "Shopping & Retail"
min_d_sports_recreation	Distância entre o centróide do polígono e o ponto de interesse mais próximo da categoria "Sports & Recreation"
min_d_travel_transportation	Distância entre o centróide do polígono e o ponto de interesse mais próximo da categoria "Travel & Transportation"
max_d_advertising_marketing	Distância entre o centroide do polígono e o ponto de interesse mais distante da categoria "Advertising/Marketing"
max_d_agriculture	Distância entre o centroide do polígono e o ponto de interesse mais distante da categoria "Agriculture"
max_d_commercial_industrial	Distância entre o centroide do polígono e o ponto de interesse mais distante da categoria "Commercial & Industrial"
max_d_education	Distância entre o centroide do polígono e o ponto de interesse mais distante da categoria "Education"
max_d_finance	Distância entre o centroide do polígono e o ponto de interesse mais distante da categoria "Finance"
max_d_food_beverage	Distância entre o centroide do polígono e o ponto de interesse mais distante da categoria "Food & Beverage"
max_d_hotel_lodging	Distância entre o centroide do polígono e o ponto de interesse mais distante da categoria "Hotel & Lodging"
max_d_legal	Distância entre o centroide do polígono e o ponto de interesse mais distante da categoria "Legal"
max_d_local_service	Distância entre o centroide do polígono e o ponto de interesse mais distante da categoria "Local Service"
max_d_media_news_company	Distância entre o centroide do polígono e o ponto de interesse mais distante da categoria "Media/News Company"
max_d_medical_health	Distância entre o centroide do polígono e o ponto de interesse mais distante da categoria "Medical & Health"
max_d_non_governmental_organization_ngo_	Distância entre o centroide do polígono e o ponto de interesse mais distante da categoria "Non-Governmental Organization (NGO)"
max_d_nonprofit_organization	Distância entre o centroide do polígono e o ponto de interesse mais distante da categoria "Nonprofit Organization"
max_d_public_government_service	Distância entre o centroide do polígono e o ponto de interesse mais distante da categoria "Public & Government Service"
max_d_science_technology_engineering	Distância entre o centroide do polígono e o ponto de interesse mais distante da categoria "Science, Technology & Engineering"
max_d_beauty_cosmetic_personal_care	Distância entre o centroide do polígono e o ponto de interesse mais distante da categoria "Beauty, Cosmetic & Personal Care"
max_d_arts_entertainment	Distância entre o centroide do polígono e o ponto de interesse mais distante da categoria "Arts & Entertainment"
max_d_automotive_aircraft_boat	Distância entre o centroide do polígono e o ponto de interesse mais distante da categoria "Automotive, Aircraft & Boat"
max_d_real_estate	Distância entre o centroide do polígono e o ponto de interesse mais distante da categoria "Real Estate"
max_d_shopping_retail	Distância entre o centroide do polígono e o ponto de interesse mais distante da categoria "Shopping & Retail"
max_d_sports_recreation	Distância entre o centroide do polígono e o ponto de interesse mais distante da categoria "Sports & Recreation"
max_d_travel_transportation	Distância entre o centroide do polígono e o ponto de interesse mais distante da categoria "Travel & Transportation"

## F.2. Atributos utilizados no cenário 2 dos testes baseados no COS

Na Tabela 63, são listados os atributos que foram adicionados ao conjunto apresentado no tópico anterior, para a realização de uma nova experiência.

Tabela 63: Conjunto de atributos 2.

<b>Atributo</b>	<b>Descrição</b>
amount_advertising_marketing	Quantidade de POIs da categoria "Advertising/Marketing" existentes dentro do raio
amount_agriculture	Quantidade de POIs da categoria "Agriculture" existentes dentro do raio
amount_commercial_industrial	Quantidade de POIs da categoria "Commercial & Industrial" existentes dentro do raio
amount_education	Quantidade de POIs da categoria "Education" existentes dentro do raio
amount_finance	Quantidade de POIs da categoria "Finance" existentes dentro do raio
amount_food_beverage	Quantidade de POIs da categoria "Food & Beverage" existentes dentro do raio
amount_hotel_lodging	Quantidade de POIs da categoria "Hotel & Lodging" existentes dentro do raio
amount_legal	Quantidade de POIs da categoria "Legal" existentes dentro do raio
amount_local_service	Quantidade de POIs da categoria "Local Service" existentes dentro do raio
amount_media_news_company	Quantidade de POIs da categoria "Media/News Company" existentes dentro do raio
amount_medical_health	Quantidade de POIs da categoria "Medical & Health" existentes dentro do raio
amount_non_governmental_organization_ngo_	Quantidade de POIs da categoria "Non-Governmental Organization (NGO)" existentes dentro do raio
amount_nonprofit_organization	Quantidade de POIs da categoria "Nonprofit Organization" existentes dentro do raio
amount_public_government_service	Quantidade de POIs da categoria "Public & Government Service" existentes dentro do raio
amount_science_technology_engineering	Quantidade de POIs da categoria "Science, Technology & Engineering" existentes dentro do raio
amount_beauty_cosmetic_personal_care	Quantidade de POIs da categoria "Beauty, Cosmetic & Personal Care" existentes dentro do raio
amount_arts_entertainment	Quantidade de POIs da categoria "Arts & Entertainment" existentes dentro do raio
amount_automotive_aircraft_boat	Quantidade de POIs da categoria "Automotive, Aircraft & Boat" existentes dentro do raio
amount_real_estate	Quantidade de POIs da categoria "Real Estate" existentes dentro do raio
amount_shopping_retail	Quantidade de POIs da categoria "Shopping & Retail" existentes dentro do raio

amount_sports_recreation	Quantidade de POIs da categoria "Sports & Recreation" existentes dentro do raio
amount_travel_transportation	Quantidade de POIs da categoria "Travel & Transportation" existentes dentro do raio
prop_advertising_marketing	Proporção de POIs da categoria "Advertising/Marketing" existentes dentro do raio
prop_agriculture	Proporção de POIs da categoria "Agriculture" existentes dentro do raio
prop_commercial_industrial	Proporção de POIs da categoria "Commercial & Industrial" existentes dentro do raio
prop_education	Proporção de POIs da categoria "Education" existentes dentro do raio
prop_finance	Proporção de POIs da categoria "Finance" existentes dentro do raio
prop_food_beverage	Proporção de POIs da categoria "Food & Beverage" existentes dentro do raio
prop_hotel_lodging	Proporção de POIs da categoria "Hotel & Lodging" existentes dentro do raio
prop_legal	Proporção de POIs da categoria "Legal" existentes dentro do raio
prop_local_service	Proporção de POIs da categoria "Local Service" existentes dentro do raio
prop_media_news_company	Proporção de POIs da categoria "Media/News Company" existentes dentro do raio
prop_medical_health	Proporção de POIs da categoria "Medical & Health" existentes dentro do raio
prop_non_governmental_organization_ngo_	Proporção de POIs da categoria "Non-Governmental Organization (NGO)" existentes dentro do raio
prop_nonprofit_organization	Proporção de POIs da categoria "Nonprofit Organization" existentes dentro do raio
prop_public_government_service	Proporção de POIs da categoria "Public & Government Service" existentes dentro do raio
prop_science_technology_engineering	Proporção de POIs da categoria "Science, Technology & Engineering" existentes dentro do raio
prop_beauty_cosmetic_personal_care	Proporção de POIs da categoria "Beauty, Cosmetic & Personal Care" existentes dentro do raio
prop_arts_entertainment	Proporção de POIs da categoria "Arts & Entertainment" existentes dentro do raio
prop_automotive_aircraft_boat	Proporção de POIs da categoria "Automotive, Aircraft & Boat" existentes dentro do raio
prop_real_estate	Proporção de POIs da categoria "Real Estate" existentes dentro do raio
prop_shopping_retail	Proporção de POIs da categoria "Shopping & Retail" existentes dentro do raio
prop_sports_recreation	Proporção de POIs da categoria "Sports & Recreation" existentes dentro do raio
prop_travel_transportation	Proporção de POIs da categoria "Travel & Transportation" existentes dentro do raio

### F.3. Atributos utilizados no cenário 5 dos testes baseados no COS

Na Tabela 64, são apresentados os atributos utilizados no cenário 5, dos testes realizados com base nos dados do COS.

Tabela 64: Conjunto de atributos 3.

<b>Atributo</b>	<b>Descrição</b>
amount_accountant	Quantidade de pontos de interesse da categoria "Accountant" existentes dentro do raio definido a partir do centróide
amount_alcohol_addiction_treatment_center	Quantidade de pontos de interesse da categoria "Alcohol addiction treatment center" existentes dentro do raio definido a partir do centróide
amount_astrologist_psychic	Quantidade de pontos de interesse da categoria "Astrologist psychic" existentes dentro do raio definido a partir do centróide
amount_automation_service	Quantidade de pontos de interesse da categoria "Automation service" existentes dentro do raio definido a partir do centróide
amount_beach_resort	Quantidade de pontos de interesse da categoria "Beach resort" existentes dentro do raio definido a partir do centróide
amount_brewery	Quantidade de pontos de interesse da categoria "Brewery" existentes dentro do raio definido a partir do centróide
amount_building_materials	Quantidade de pontos de interesse da categoria "Building materials" existentes dentro do raio definido a partir do centróide
amount_car_wash	Quantidade de pontos de interesse da categoria "Car wash" existentes dentro do raio definido a partir do centróide
amount_carpenter	Quantidade de pontos de interesse da categoria "Carpenter" existentes dentro do raio definido a partir do centróide
amount_comedian	Quantidade de pontos de interesse da categoria "Comedian" existentes dentro do raio definido a partir do centróide
amount_commercial_industrial	Quantidade de pontos de interesse da categoria "Industrial or commercial" existentes dentro do raio definido a partir do centróide
amount_cycling_studio	Quantidade de pontos de interesse da categoria "Cycling studio" existentes dentro do raio definido a partir do centróide
amount_dairy_farm	Quantidade de pontos de interesse da categoria "Dairy farm" existentes dentro do raio definido a partir do centróide
amount_deck_patio_builder	Quantidade de pontos de interesse da categoria "Deck patio builder" existentes dentro do raio definido a partir do centróide
amount_drug_addiction_treatment_center	Quantidade de pontos de interesse da categoria "Drug addiction treatment center" existentes dentro do raio definido a partir do centróide
amount_farm	Quantidade de pontos de interesse da categoria "Farm" existentes dentro do raio definido a partir do centróide
amount_fencing_club	Quantidade de pontos de interesse da categoria "Fencing club" existentes dentro do raio definido a partir do centróide
amount_fish_chips_restaurant	Quantidade de pontos de interesse da categoria "Fish chips restaurant" existentes dentro do raio definido a partir do centróide
amount_flea_market	Quantidade de pontos de interesse da categoria "Flea market" existentes dentro do raio definido a partir do centróide

amount_foodservice_distributor	Quantidade de pontos de interesse da categoria "Foodservice distributor" existentes dentro do raio definido a partir do centróide
amount_gamer	Quantidade de pontos de interesse da categoria "Gamer" existentes dentro do raio definido a partir do centróide
amount_halfway_house	Quantidade de pontos de interesse da categoria "Halfway house" existentes dentro do raio definido a partir do centróide
amount_highway	Quantidade de pontos de interesse da categoria "Highway" existentes dentro do raio definido a partir do centróide
amount_internet_cafe	Quantidade de pontos de interesse da categoria "Internet cafe" existentes dentro do raio definido a partir do centróide
amount_landscape_company	Quantidade de pontos de interesse da categoria "Landscape company" existentes dentro do raio definido a partir do centróide
amount_medical_center	Quantidade de pontos de interesse da categoria "Medical center" existentes dentro do raio definido a partir do centróide
amount_nursing_home	Quantidade de pontos de interesse da categoria "Nursing home" existentes dentro do raio definido a partir do centróide
amount_psychologist	Quantidade de pontos de interesse da categoria "Psychologist" existentes dentro do raio definido a partir do centróide
amount_rent_to_own_store	Quantidade de pontos de interesse da categoria "Rent to own store" existentes dentro do raio definido a partir do centróide
amount_robotics_company	Quantidade de pontos de interesse da categoria "Robotics company" existentes dentro do raio definido a partir do centróide
amount_social_club	Quantidade de pontos de interesse da categoria "Social club" existentes dentro do raio definido a partir do centróide
amount_solar_energy_service	Quantidade de pontos de interesse da categoria "Solar energy service" existentes dentro do raio definido a partir do centróide
amount_speakeasy	Quantidade de pontos de interesse da categoria "Speakeasy" existentes dentro do raio definido a partir do centróide
amount_wheel_rim_repair_service	Quantidade de pontos de interesse da categoria "Wheel rim repair service" existentes dentro do raio definido a partir do centróide
d_accessories	Distância até o ponto de interesse mais próximo, da categoria "Accessories"
d_advertising_agency	Distância até o ponto de interesse mais próximo, da categoria "Advertising agency"
d_arts_entertainment	Distância até o ponto de interesse mais próximo, da categoria "Arts entertainment"
d_automotive_aircraft_boat	Distância até o ponto de interesse mais próximo, da categoria "Automotive aircraft boat"
d_baby_goods_kids_goods	Distância até o ponto de interesse mais próximo, da categoria "Baby goods kids goods"
d_bar_grill	Distância até o ponto de interesse mais próximo, da categoria "Bar grill"
d_cafe	Distância até o ponto de interesse mais próximo, da categoria "Cafe"
d_cargo_freight_company	Distância até o ponto de interesse mais próximo, da categoria "Cargo freight company"
d_city	Distância até o ponto de interesse mais próximo, da categoria "City"

d_community	Distância até o ponto de interesse mais próximo, da categoria "Community"
d_european_restaurant	Distância até o ponto de interesse mais próximo, da categoria "European restaurant"
d_food_beverage	Distância até o ponto de interesse mais próximo, da categoria "Food beverage"
d_home	Distância até o ponto de interesse mais próximo, da categoria "Home"
d_home_improvement	Distância até o ponto de interesse mais próximo, da categoria "Home improvement"
d_hotel	Distância até o ponto de interesse mais próximo, da categoria "Hotel"
d_hotel_lodging	Distância até o ponto de interesse mais próximo, da categoria "Hotel lodging"
d_hotel_resort	Distância até o ponto de interesse mais próximo, da categoria "Hotel resort"
d_industrial_company	Distância até o ponto de interesse mais próximo, da categoria "Industrial company"
d_local_service	Distância até o ponto de interesse mais próximo, da categoria "Local service"
d_media_news_company	Distância até o ponto de interesse mais próximo, da categoria "Media news company"
d_motor_vehicle_company	Distância até o ponto de interesse mais próximo, da categoria "Motor vehicle company"
d_nail_salon	Distância até o ponto de interesse mais próximo, da categoria "Nail salon"
d_nonprofit_organization	Distância até o ponto de interesse mais próximo, da categoria "Nonprofit organization"
d_performance_event_venue	Distância até o ponto de interesse mais próximo, da categoria "Performance event venue"
d_portuguese_restaurant	Distância até o ponto de interesse mais próximo, da categoria "Portuguese restaurant"
d_public_swimming_pool	Distância até o ponto de interesse mais próximo, da categoria "Public swimming pool"
d_soccer_field	Distância até o ponto de interesse mais próximo, da categoria "Soccer field"
d_transportation_service	Distância até o ponto de interesse mais próximo, da categoria "Transportation service"
d_wine_spirits	Distância até o ponto de interesse mais próximo, da categoria "Wine spirits"

## F.4. Atributos do cenário 4 dos testes baseados no CORINE

Na Tabela 65, são apresentados os atributos utilizados no cenário 4, dos testes realizados com base nos dados do CORINE.

Tabela 65: Conjunto de atributos 4.

<b>Atributo</b>	<b>Descrição</b>
amount_commercial_industrial_10000m	Quantidade de pontos de interesse da categoria "industrial or commercial", localizados dentro de um raio de 10000 metros a partir do centróide da célula
amount_education_10000m	Quantidade de pontos de interesse da categoria "education", localizados dentro de um raio de 10000 metros a partir do centróide da célula
amount_local_service_10000m	Quantidade de pontos de interesse da categoria "local service", localizados dentro de um raio de 10000 metros a partir do centróide da célula
amount_nonprofit_organization_10000m	Quantidade de pontos de interesse da categoria "nonprofit organization", localizados dentro de um raio de 10000 metros a partir do centróide da célula
amount_beauty_cosmetic_personal_care_10000m	Quantidade de pontos de interesse da categoria "beauty cosmetic personal care", localizados dentro de um raio de 10000 metros a partir do centróide da célula
amount_arts_entertainment_10000m	Quantidade de pontos de interesse da categoria "arts entertainment", localizados dentro de um raio de 10000 metros a partir do centróide da célula
amount_real_estate_10000m	Quantidade de pontos de interesse da categoria "real estate", localizados dentro de um raio de 10000 metros a partir do centróide da célula
amount_shopping_retail_10000m	Quantidade de pontos de interesse da categoria "shopping & retail", localizados dentro de um raio de 10000 metros a partir do centróide da célula
amount_sports_recreation_10000m	Quantidade de pontos de interesse da categoria "Sports & recreation", localizados dentro de um raio de 10000 metros a partir do centróide da célula
amount_school_10000m	Quantidade de pontos de interesse da categoria "school", localizados dentro de um raio de 10000 metros a partir do centróide da célula
amount_restaurant_10000m	Quantidade de pontos de interesse da categoria "restaurant", localizados dentro de um raio de 10000 metros a partir do centróide da célula
amount_bar_10000m	Quantidade de pontos de interesse da categoria "bar", localizados dentro de um raio de 10000 metros a partir do centróide da célula
amount_event_planner_10000m	Quantidade de pontos de interesse da categoria "event planner", localizados dentro de um raio de 10000 metros a partir do centróide da célula
amount_beauty_salon_10000m	Quantidade de pontos de interesse da categoria "beauty salon", localizados dentro de um raio de 10000 metros a partir do centróide da célula
amount_tattoo_piercing_shop_10000m	Quantidade de pontos de interesse da categoria "tattoo piercing shop", localizados dentro de um raio de 10000 metros a partir do centróide da célula
amount_real_estate_agent_10000m	Quantidade de pontos de interesse da categoria "real estate agent", localizados dentro de um raio de 10000 metros a partir do centróide da célula
amount_stadium_arena_sports_venue_10000m	Quantidade de pontos de interesse da categoria "stadium arena sports venue", localizados dentro de um raio de 10000 metros a partir do centróide da célula
amount_health_beauty_10000m	Quantidade de pontos de interesse da categoria "health beauty", localizados dentro de um raio de 10000 metros a partir do centróide da célula
amount_product_service_10000m	Quantidade de pontos de interesse da categoria "product service", localizados dentro de um raio de 10000 metros a partir do centróide da célula
amount_european_restaurant_10000m	Quantidade de pontos de interesse da categoria "european restaurant", localizados dentro de um raio de 10000 metros a partir do centróide da célula

amount_italian_restaurant_10000m	Quantidade de pontos de interesse da categoria "italian restaurant", localizados dentro de um raio de 10000 metros a partir do centróide da célula
amount_portuguese_restaurant_10000m	Quantidade de pontos de interesse da categoria "portuguese restaurant", localizados dentro de um raio de 10000 metros a partir do centróide da célula
amount_veterinarian_10000m	Quantidade de pontos de interesse da categoria "veterinarian", localizados dentro de um raio de 10000 metros a partir do centróide da célula
amount_photographer_10000m	Quantidade de pontos de interesse da categoria "photographer", localizados dentro de um raio de 10000 metros a partir do centróide da célula
amount_alternative_holistic_health_service_10000m	Quantidade de pontos de interesse da categoria "alternative holistic health service", localizados dentro de um raio de 10000 metros a partir do centróide da célula
amount_hair_salon_10000m	Quantidade de pontos de interesse da categoria "hair salon", localizados dentro de um raio de 10000 metros a partir do centróide da célula
amount_art_museum_10000m	Quantidade de pontos de interesse da categoria "art museum", localizados dentro de um raio de 10000 metros a partir do centróide da célula
amount_car_dealership_10000m	Quantidade de pontos de interesse da categoria "car dealership", localizados dentro de um raio de 10000 metros a partir do centróide da célula
amount_baby_children_s_clothing_store_10000m	Quantidade de pontos de interesse da categoria "baby children s clothing store", localizados dentro de um raio de 10000 metros a partir do centróide da célula
amount_gym_physical_fitness_center_10000m	Quantidade de pontos de interesse da categoria "gym physical fitness center", localizados dentro de um raio de 10000 metros a partir do centróide da célula
d_commercial_industrial	Distância do centroide da célula até o ponto de interesse mais próximo, da categoria "industrial or commercial"
d_local_service	Distância do centroide da célula até o ponto de interesse mais próximo, da categoria "local service"
d_medical_health	Distância do centroide da célula até o ponto de interesse mais próximo, da categoria "medical health"
d_nonprofit_organization	Distância do centroide da célula até o ponto de interesse mais próximo, da categoria "nonprofit organization"
d_beauty_cosmetic_personal_care	Distância do centroide da célula até o ponto de interesse mais próximo, da categoria "beauty cosmetic personal care"
d_sports_recreation	Distância do centroide da célula até o ponto de interesse mais próximo, da categoria "Sports & recreation"
d_automated_teller_machine_atm_	Distância do centroide da célula até o ponto de interesse mais próximo, da categoria "automated teller machine atm "
d_marketing_agency	Distância do centroide da célula até o ponto de interesse mais próximo, da categoria "marketing agency"
d_public_relations_agency	Distância do centroide da célula até o ponto de interesse mais próximo, da categoria "public relations agency"
d_agricultural_cooperative	Distância do centroide da célula até o ponto de interesse mais próximo, da categoria "agricultural cooperative"
d_school	Distância do centroide da célula até o ponto de interesse mais próximo, da categoria "school"
d_restaurant	Distância do centroide da célula até o ponto de interesse mais próximo, da categoria "restaurant"
d_campground	Distância do centroide da célula até o ponto de interesse mais próximo, da categoria "campground"
d_cottage	Distância do centroide da célula até o ponto de interesse mais próximo, da categoria "cottage"

## Classificação e Análise do Uso do Solo com Base em Pontos de Interesse

d_inn	Distância do centroide da célula até o ponto de interesse mais próximo, da categoria "inn"
d_service_apartments	Distância do centroide da célula até o ponto de interesse mais próximo, da categoria "service apartments"
d_hotel	Distância do centroide da célula até o ponto de interesse mais próximo, da categoria "hotel"
d_legal_service	Distância do centroide da célula até o ponto de interesse mais próximo, da categoria "legal service"
d_bicycle_repair_service	Distância do centroide da célula até o ponto de interesse mais próximo, da categoria "bicycle repair service"
d_design_fashion	Distância do centroide da célula até o ponto de interesse mais próximo, da categoria "design fashion"
d_event_space	Distância do centroide da célula até o ponto de interesse mais próximo, da categoria "event space"
d_pet_service	Distância do centroide da célula até o ponto de interesse mais próximo, da categoria "pet service"
d_event_planner	Distância do centroide da célula até o ponto de interesse mais próximo, da categoria "event planner"
d_dentist_dental_office	Distância do centroide da célula até o ponto de interesse mais próximo, da categoria "dentist dental office"
d_defense_company	Distância do centroide da célula até o ponto de interesse mais próximo, da categoria "defense company"
d_tattoo_piercing_shop	Distância do centroide da célula até o ponto de interesse mais próximo, da categoria "tattoo piercing shop"
d_art_gallery	Distância do centroide da célula até o ponto de interesse mais próximo, da categoria "art gallery"
d_party_supply_rental_shop	Distância do centroide da célula até o ponto de interesse mais próximo, da categoria "party supply rental shop"
d_pet_store	Distância do centroide da célula até o ponto de interesse mais próximo, da categoria "pet store"
d_thrift_consignment_store	Distância do centroide da célula até o ponto de interesse mais próximo, da categoria "thrift consignment store"
d_rental_shop	Distância do centroide da célula até o ponto de interesse mais próximo, da categoria "rental shop"
d_hindu_temple	Distância do centroide da célula até o ponto de interesse mais próximo, da categoria "hindu temple"
d_religious_center	Distância do centroide da célula até o ponto de interesse mais próximo, da categoria "religious center"
d_health_beauty	Distância do centroide da célula até o ponto de interesse mais próximo, da categoria "health beauty"
d_product_service	Distância do centroide da célula até o ponto de interesse mais próximo, da categoria "product service"
d_home_decor	Distância do centroide da célula até o ponto de interesse mais próximo, da categoria "home decor"
d_appliance_manufacturer	Distância do centroide da célula até o ponto de interesse mais próximo, da categoria "appliance manufacturer"
d_accountant	Distância do centroide da célula até o ponto de interesse mais próximo, da categoria "accountant"
d_italian_restaurant	Distância do centroide da célula até o ponto de interesse mais próximo, da categoria "italian restaurant"

d_mediterranean_restaurant	Distância do centroide da célula até o ponto de interesse mais próximo, da categoria "mediterranean restaurant"
d_portuguese_restaurant	Distância do centroide da célula até o ponto de interesse mais próximo, da categoria "portuguese restaurant"
d_lounge	Distância do centroide da célula até o ponto de interesse mais próximo, da categoria "lounge"
d_cleaning_service	Distância do centroide da célula até o ponto de interesse mais próximo, da categoria "cleaning service"
d_veterinarian	Distância do centroide da célula até o ponto de interesse mais próximo, da categoria "veterinarian"
d_photographer	Distância do centroide da célula até o ponto de interesse mais próximo, da categoria "photographer"
d_hospitality_service	Distância do centroide da célula até o ponto de interesse mais próximo, da categoria "hospitality service"
d_landscape_company	Distância do centroide da célula até o ponto de interesse mais próximo, da categoria "landscape company"
d_general_dentist	Distância do centroide da célula até o ponto de interesse mais próximo, da categoria "general dentist"
d_orthodontist	Distância do centroide da célula até o ponto de interesse mais próximo, da categoria "orthodontist"
d_alternative_holistic_health_service	Distância do centroide da célula até o ponto de interesse mais próximo, da categoria "alternative holistic health service"
d_psychotherapist	Distância do centroide da célula até o ponto de interesse mais próximo, da categoria "psychotherapist"
d_psychologist	Distância do centroide da célula até o ponto de interesse mais próximo, da categoria "psychologist"
d_massage_service	Distância do centroide da célula até o ponto de interesse mais próximo, da categoria "massage service"
d_motorcycle_manufacturer	Distância do centroide da célula até o ponto de interesse mais próximo, da categoria "motorcycle manufacturer"
d_car_dealership	Distância do centroide da célula até o ponto de interesse mais próximo, da categoria "car dealership"
d_oil_lube_filter_service	Distância do centroide da célula até o ponto de interesse mais próximo, da categoria "oil lube filter service"
d_accessories	Distância do centroide da célula até o ponto de interesse mais próximo, da categoria "accessories"
d_fishing_store	Distância do centroide da célula até o ponto de interesse mais próximo, da categoria "fishing store"
d_golf_course_country_club	Distância do centroide da célula até o ponto de interesse mais próximo, da categoria "golf course country club"
d_horse_riding_school	Distância do centroide da célula até o ponto de interesse mais próximo, da categoria "horse riding school"
d_ocean	Distância do centroide da célula até o ponto de interesse mais próximo, da categoria "ocean"
d_surfing_spot	Distância do centroide da célula até o ponto de interesse mais próximo, da categoria "surfing spot"
d_heating_ventilating_air_conditioning_service	Distância do centroide da célula até o ponto de interesse mais próximo, da categoria "heating ventilating air conditioning service"
d_gym_physical_fitness_center	Distância do centroide da célula até o ponto de interesse mais próximo, da categoria "gym physical fitness center"

## F.5. Atributos do teste 1, do cenário 5, com base no CORINE

Na Tabela 66, são apresentados os atributos utilizados no teste 1, do cenário 5, das experiências realizadas com base nos dados do CORINE.

Tabela 66: Conjunto de atributos 5.

Atributos	
amount_sports_club_10000m	d_travel_company
amount_commercial_industrial_10000m	d_geographical_place
amount_education_10000m	d_recreation_spot
amount_local_service_10000m	d_hindu_temple
amount_medical_health_10000m	d_health_beauty
amount_nonprofit_organization_10000m	d_product_service
amount_beauty_cosmetic_personal_care_10000m	d_home_decor
amount_arts_entertainment_10000m	d_kitchen_cooking
amount_real_estate_10000m	amount_candy_store_10000m
amount_shopping_retail_10000m	amount_european_restaurant_10000m
amount_sports_recreation_10000m	amount_italian_restaurant_10000m
amount_event_10000m	amount_portuguese_restaurant_10000m
d_charity_organization	amount_lounge_10000m
d_concert_tour	amount_veterinarian_10000m
d_video_creator	amount_photographer_10000m
d_commercial_industrial	amount_alternative_holistic_health_service_10000m
d_education	amount_hair_salon_10000m
d_local_service	amount_nail_salon_10000m
d_medical_health	amount_car_dealership_10000m
d_nonprofit_organization	amount_accessories_10000m
d_beauty_cosmetic_personal_care	amount_clothing_store_10000m
d_arts_entertainment	amount_tour_agency_10000m
d_real_estate	d_appliance_manufacturer
d_shopping_retail	d_clothing_company
d_sports_recreation	d_candy_store
d_automated_teller_machine_atm_	d_european_restaurant
d_event	d_hot_dog_joint
amount_school_10000m	d_italian_restaurant
amount_restaurant_10000m	d_pizza_place
amount_bar_10000m	d_portuguese_restaurant
amount_design_fashion_10000m	d_seafood_restaurant
amount_pet_service_10000m	d_lounge
amount_business_service_10000m	d_cleaning_service

amount_event_planner_10000m	d_animal_shelter
amount_beauty_salon_10000m	d_dog_walker
amount_tattoo_piercing_shop_10000m	d_pet_groomer
amount_real_estate_agent_10000m	d_veterinarian
amount_health_beauty_10000m	d_photographer
amount_product_service_10000m	d_home_mover
amount_home_decor_10000m	d_house_painting
d_movie	d_landscape_company
d_public_relations_agency	d_orthodontist
d_manufacturer_supplier	d_alternative_holistic_health_service
d_academic_camp	d_psychotherapist
d_school	d_internet_company
d_financial_service	d_hair_salon
d_investment_management_company	d_nail_salon
d_restaurant	d_massage_service
d_bar	d_history_museum
d_campground	d_motorcycle_manufacturer
d_service_apartments	d_car_dealership
d_legal_service	d_automotive_body_shop
d_design_fashion	d_automotive_restoration_service
d_pet_service	d_accessories
d_business_service	d_beauty_supply_store
d_event_planner	d_fireplace_store
d_home_improvement	d_bicycle_shop
d_movie_television_studio	d_fishing_store
d_dentist_dental_office	d_horseback_riding_center
d_public_utility_company	d_laser_tag_center
d_beauty_salon	d_amateur_sports_team
d_tattoo_piercing_shop	d_school_sports_team
d_art_gallery	d_car_rental
d_real_estate_agent	d_tour_agency
d_real_estate_service	d_beach
d_department_store	d_hiking_trail
d_pet_store	d_surfing_spot
d_thrift_consignment_store	d_health_wellness_website
d_sports_fitness_instruction	amount_gym_physical_fitness_center_10000m
d_sports_league	d_gym_physical_fitness_center
d_boat_ferry_company	d_food_tour_agency
d_rental_shop	d_sightseeing_tour_agency

## F.6. Atributos do teste 2, do cenário 5, com base no CORINE

Na Tabela 67, são apresentados os atributos utilizados no teste 2, do cenário 5, das experiências realizadas com base nos dados do CORINE.

Tabela 67: Conjunto de atributos 6.

Atributos	
amount_local_service_10000m	d_kitchen_cooking
amount_medical_health_10000m	d_website
amount_nonprofit_organization_10000m	amount_portuguese_restaurant_10000m
amount_beauty_cosmetic_personal_care_10000m	amount_veterinarian_10000m
amount_arts_entertainment_10000m	amount_photographer_10000m
amount_shopping_retail_10000m	amount_alternative_holistic_health_service_10000m
amount_sports_recreation_10000m	amount_hair_salon_10000m
d_charity_organization	amount_car_dealership_10000m
d_government_organization	d_appliance_manufacturer
d_religious_organization	d_public_school
d_social_club	d_tax_preparation_service
d_sports_club	d_tea_room
d_concert_tour	d_african_restaurant
d_dancer	d_fast_food_restaurant
d_fitness_model	d_indian_restaurant
d_commercial_industrial	d_italian_restaurant
d_local_service	d_portuguese_restaurant
d_medical_health	d_lounge
d_nonprofit_organization	d_supermarket
d_science_technology_engineering	d_ballroom
d_beauty_cosmetic_personal_care	d_child_care_service
d_automated_teller_machine_atm_	d_cleaning_service
d_cause	d_animal_shelter
amount_restaurant_10000m	d_dog_walker
amount_bar_10000m	d_pet_groomer
amount_business_service_10000m	d_veterinarian
amount_event_planner_10000m	d_photographer
amount_beauty_salon_10000m	d_home_mover
amount_health_beauty_10000m	d_uniform_supplier
amount_product_service_10000m	d_carpet_cleaner
d_newspaper	d_contractor
d_movie	d_house_painting
d_public_relations_agency	d_landscape_company
d_school	d_general_dentist

d_financial_service	d_pregnancy_care_center
d_investment_management_company	d_alternative_holistic_health_service
d_bakery	d_family_therapist
d_caterer	d_psychotherapist
d_farmers_market	d_speech_pathologist
d_restaurant	d_ophthalmologist
d_service_apartments	d_optometrist
d_lawyer_law_firm	d_psychologist
d_legal_service	d_internet_company
d_design_fashion	d_hair_salon
d_locksmith	d_health_spa
d_pet_service	d_massage_service
d_photography_videography	d_motorcycle_manufacturer
d_printing_service	d_car_dealership
d_business_service	d_motorcycle_dealership
d_event_planner	d_automotive_restoration_service
d_movie_television_studio	d_accessories
d_dentist_dental_office	d_beauty_supply_store
d_public_utility_company	d_computer_store
d_tattoo_piercing_shop	d_awning_supplier
d_amusement_theme_park	d_fireplace_store
d_arcade	d_fishing_store
d_art_gallery	d_outdoor_equipment_store
d_circus	d_surf_shop
d_antique_store	d_skateboard_park
d_home_garden_store	d_fitness_trainer
d_party_supply_rental_shop	d_amateur_sports_team
d_pet_store	d_bike_rental
d_retail_company	d_car_rental
d_thrift_consignment_store	d_limo_service
d_bookstore	d_bay
d_sports_recreation_venue	d_playground
d_rental_shop	d_hiking_trail
d_transportation_service	d_surfing_spot
d_highway	d_evangelical_church
d_shopping_district	d_education_website
d_park	d_news_media_website
d_recreation_spot	amount_gym_physical_fitness_center_10000m
d_church	d_peruvian_restaurant
d_hindu_temple	d_bridal_shop
d_religious_center	d_women_s_clothing_store

d_cars	d_gym_physical_fitness_center
d_electronics	d_pilates_studio
d_health_beauty	d_yoga_studio
d_product_service	

## F.7. Atributos do teste 3, do cenário 5, com base no CORINE

Na Tabela 68, são apresentados os atributos utilizados no teste 3, do cenário 5, das experiências realizadas com base nos dados do CORINE.

Tabela 68: Conjunto de atributos 7.

Atributos	
amount_charity_organization_10000m	d_consulate_embassy
amount_government_organization_10000m	d_prison_correctional_facility
amount_sports_club_10000m	d_state
amount_theatrical_productions_10000m	d_computers_brand_
amount_blogger_10000m	d_health_beauty
amount_orchestra_10000m	amount_art_school_10000m
amount_sports_promoter_10000m	amount_medical_school_10000m
amount_media_news_company_10000m	amount_dance_school_10000m
amount_non_governmental_organization_ngo__10000m	amount_community_college_10000m
amount_nonprofit_organization_10000m	amount_steakhouse_10000m
amount_public_government_service_10000m	amount_vegetarian_vegan_restaurant_10000m
amount_travel_transportation_10000m	amount_gay_bar_10000m
amount_campus_building_10000m	amount_convention_center_10000m
amount_city_infrastructure_10000m	amount_business_consultant_10000m
amount_brand_10000m	amount_employment_agency_10000m
amount_ticket_sales_10000m	amount_personal_coach_10000m
d_science	amount_orthodontist_10000m
d_government_organization	amount_surgical_center_10000m
d_orchestra	amount_mental_health_service_10000m
d_legal	amount_dermatologist_10000m
d_campus_building	amount_pediatrician_10000m
d_ticket_sales	amount_plastic_surgeon_10000m
amount_magazine_10000m	amount_auditorium_10000m
amount_newspaper_10000m	amount_comedy_club_10000m
amount_tv_network_10000m	amount_live_music_venue_10000m
amount_college_university_10000m	amount_performance_art_theatre_10000m
amount_bank_10000m	amount_art_museum_10000m
amount_insurance_company_10000m	amount_history_museum_10000m
amount_cheese_shop_10000m	amount_emergency_roadside_service_10000m
amount_bar_10000m	amount_travel_agency_10000m

amount_legal_service_10000m	amount_public_square_plaza_10000m
amount_broadcasting_media_production_company_10000m	amount_education_website_10000m
amount_radio_station_10000m	amount_news_media_website_10000m
amount_publisher_10000m	amount_science_website_10000m
amount_medical_research_center_10000m	amount_local_travel_website_10000m
amount_cultural_center_10000m	d_medical_school
amount_library_10000m	d_dance_school
amount_tattoo_piercing_shop_10000m	d_flight_school
amount_dance_night_club_10000m	d_caribbean_restaurant
amount_performance_event_venue_10000m	d_gay_bar
amount_museum_10000m	d_hookah_lounge
amount_real_estate_service_10000m	d_convention_center
amount_mobile_phone_shop_10000m	d_employment_agency
amount_shopping_mall_10000m	d_endodontist
amount_stadium_arena_sports_venue_10000m	d_orthodontist
amount_airline_company_10000m	d_pediatric_dentist
amount_cruise_line_10000m	d_surgical_center
amount_travel_company_10000m	d_allergist
amount_marina_10000m	d_dermatologist
amount_street_10000m	d_software_company
amount_consulate_embassy_10000m	d_comedy_club
amount_geo_entity_10000m	d_performance_art_theatre
amount_neighborhood_10000m	d_emergency_roadside_service
amount_port_10000m	d_oil_lube_filter_service
amount_state_10000m	d_skateboard_park
amount_monument_10000m	d_skydiving_center
amount_clothing_brand_10000m	d_charter_bus_service
amount_website_10000m	d_church_of_christ
d_tv_network	d_science_website
d_cheese_shop	amount_asian_fusion_restaurant_10000m
d_campground	amount_goan_restaurant_10000m
d_medical_research_center	amount_baby_children_s_clothing_store_10000m
d_cultural_center	amount_maternity_nursing_clothing_store_10000m
d_tattoo_piercing_shop	amount_gym_physical_fitness_center_10000m
d_real_estate_service	d_goan_restaurant
d_beauty_store	d_cuban_restaurant
d_video_game_store	d_peruvian_restaurant
d_stadium_arena_sports_venue	d_reflexologist
d_airline_company	d_contemporary_art_museum
d_cruise_line	d_modern_art_museum
d_boat_ferry_company	d_maternity_nursing_clothing_store

## F.8. Atributos do teste 4, do cenário 5, com base no CORINE

Na Tabela 69, são apresentados os atributos utilizados no teste 4, do cenário 5, das experiências realizadas com base nos dados do CORINE.

Tabela 69: Conjunto de atributos 8.

Atributos	
amount_charity_organization_10000m	d_defense_company
amount_government_organization_10000m	d_tattoo_piercing_shop
amount_theatrical_productions_10000m	d_real_estate_service
amount_blogger_10000m	d_airline_company
amount_orchestra_10000m	d_consulate_embassy
amount_local_service_10000m	d_prison_correctional_facility
amount_media_news_company_10000m	d_geographical_place
amount_non_governmental_organization_ngo__10000m	d_hindu_temple
amount_nonprofit_organization_10000m	d_health_beauty
amount_public_government_service_10000m	amount_medical_school_10000m
amount_beauty_cosmetic_personal_care_10000m	amount_community_college_10000m
amount_arts_entertainment_10000m	amount_portuguese_restaurant_10000m
amount_shopping_retail_10000m	amount_vegetarian_vegan_restaurant_10000m
amount_sports_recreation_10000m	amount_beer_bar_10000m
amount_travel_transportation_10000m	amount_gay_bar_10000m
amount_campus_building_10000m	amount_convention_center_10000m
amount_city_infrastructure_10000m	amount_orthodontist_10000m
amount_brand_10000m	amount_surgical_center_10000m
d_science	amount_alternative_holistic_health_service_10000m
d_charity_organization	amount_mental_health_service_10000m
d_government_organization	amount_dermatologist_10000m
d_religious_organization	amount_pediatrician_10000m
d_orchestra	amount_plastic_surgeon_10000m
d_local_service	amount_auditorium_10000m
d_nonprofit_organization	amount_performance_art_theatre_10000m
d_beauty_cosmetic_personal_care	amount_art_museum_10000m
d_arts_entertainment	amount_history_museum_10000m
d_automated_teller_machine_atm_	amount_clothing_store_10000m
d_ticket_sales	amount_travel_agency_10000m
amount_magazine_10000m	amount_public_square_plaza_10000m
amount_newspaper_10000m	amount_education_website_10000m
amount_tv_network_10000m	amount_news_media_website_10000m
amount_college_university_10000m	amount_science_website_10000m
amount_bank_10000m	amount_local_travel_website_10000m

amount_insurance_company_10000m	d_appliance_manufacturer
amount_cheese_shop_10000m	d_medical_school
amount_restaurant_10000m	d_accountant
amount_bar_10000m	d_caribbean_restaurant
amount_event_planner_10000m	d_mediterranean_restaurant
amount_broadcasting_media_production_company_10000m	d_portuguese_restaurant
amount_radio_station_10000m	d_lounge
amount_publisher_10000m	d_convention_center
amount_medical_research_center_10000m	d_veterinarian
amount_cultural_center_10000m	d_home_mover
amount_library_10000m	d_endodontist
amount_dance_night_club_10000m	d_orthodontist
amount_performance_event_venue_10000m	d_pediatric_dentist
amount_museum_10000m	d_surgical_center
amount_real_estate_service_10000m	d_alternative_holistic_health_service
amount_mobile_phone_shop_10000m	d_psychotherapist
amount_stadium_arena_sports_venue_10000m	d_allergist
amount_cruise_line_10000m	d_dermatologist
amount_travel_company_10000m	d_software_company
amount_marina_10000m	d_comedy_club
amount_street_10000m	d_performance_art_theatre
amount_consulate_embassy_10000m	d_motorcycle_manufacturer
amount_geo_entity_10000m	d_emergency_roadside_service
amount_neighborhood_10000m	d_accessories
amount_port_10000m	d_cabinet_countertop_store
amount_state_10000m	d_fireplace_store
amount_monument_10000m	d_golf_course_country_club
amount_clothing_brand__10000m	d_horse_riding_school
amount_health_beauty_10000m	d_car_rental
amount_website_10000m	d_charter_bus_service
d_tv_network	d_hiking_trail
d_politician	d_church_of_christ
d_marketing_agency	d_science_website
d_agricultural_cooperative	d_local_travel_website
d_bubble_tea_shop	amount_asian_fusion_restaurant_10000m
d_cheese_shop	amount_goan_restaurant_10000m
d_restaurant	amount_maternity_nursing_clothing_store_10000m
d_campground	amount_gym_physical_fitness_center_10000m
d_cottage	d_goan_restaurant
d_service_apartments	d_cuban_restaurant
d_bicycle_repair_service	d_peruvian_restaurant

d_event_space	d_reflexologist
d_event_planner	d_contemporary_art_museum
d_movie_television_studio	d_modern_art_museum
d_medical_research_center	d_maternity_nursing_clothing_store

## F.9. Atributos do teste 5, do cenário 5, com base no CORINE

Na Tabela 70 são apresentados os atributos utilizados no teste 5, do cenário 5, das experiências realizadas com base nos dados do CORINE.

Tabela 70: Conjunto de atributos 9.

Atributos	
amount_community_service_10000m	d_product_service
amount_sports_club_10000m	d_home_decor
amount_commercial_industrial_10000m	d_vitamins_supplements
amount_education_10000m	d_sports_event
amount_local_service_10000m	amount_dance_school_10000m
amount_media_news_company_10000m	amount_candy_store_10000m
amount_nonprofit_organization_10000m	amount_bar_grill_10000m
amount_beauty_cosmetic_personal_care_10000m	amount_european_restaurant_10000m
amount_arts_entertainment_10000m	amount_italian_restaurant_10000m
amount_real_estate_10000m	amount_pizza_place_10000m
amount_shopping_retail_10000m	amount_portuguese_restaurant_10000m
amount_sports_recreation_10000m	amount_lounge_10000m
amount_event_10000m	amount_veterinarian_10000m
d_science	amount_photographer_10000m
d_charity_organization	amount_personal_coach_10000m
d_community_service	amount_alternative_holistic_health_service_10000m
d_athlete	amount_hair_salon_10000m
d_entrepreneur	amount_nail_salon_10000m
d_orchestra	amount_live_music_venue_10000m
d_producer	amount_performance_art_theatre_10000m
d_gamer	amount_car_dealership_10000m
d_commercial_industrial	amount_accessories_10000m
d_education	amount_clothing_store_10000m
d_local_service	amount_cosmetics_store_10000m
d_medical_health	amount_beauty_supply_store_10000m
d_nonprofit_organization	amount_furniture_store_10000m
d_beauty_cosmetic_personal_care	amount_tour_agency_10000m
d_arts_entertainment	amount_news_media_website_10000m
d_sports_recreation	d_urban_farm

d_automated_teller_machine_atm_	d_appliance_manufacturer
d_campus_building	d_elementary_school
d_city_infrastructure	d_dance_school
d_cause	d_community_college
d_ticket_sales	d_candy_store
d_event	d_african_restaurant
d_just_for_fun	d_bar_grill
amount_tv_network_10000m	d_caribbean_restaurant
amount_school_10000m	d_diner
amount_restaurant_10000m	d_european_restaurant
amount_bar_10000m	d_hot_dog_joint
amount_design_fashion_10000m	d_italian_restaurant
amount_event_planner_10000m	d_pizza_place
amount_radio_station_10000m	d_portuguese_restaurant
amount_doctor_10000m	d_beer_bar
amount_beauty_salon_10000m	d_lounge
amount_tattoo_piercing_shop_10000m	d_convention_center
amount_dance_night_club_10000m	d_cleaning_service
amount_performance_event_venue_10000m	d_dj
amount_automotive_store_10000m	d_dog_walker
amount_real_estate_agent_10000m	d_veterinarian
amount_real_estate_service_10000m	d_photographer
amount_pet_store_10000m	d_home_mover
amount_shopping_mall_10000m	d_employment_agency
amount_stadium_arena_sports_venue_10000m	d_personal_coach
amount_transportation_service_10000m	d_portable_toilet_rentals
amount_health_beauty_10000m	d_solar_energy_service
amount_product_service_10000m	d_endodontist
amount_home_decor_10000m	d_orthodontist
amount_website_10000m	d_pediatric_dentist
d_magazine	d_hospital
d_tv_network	d_alternative_holistic_health_service
d_tv_show	d_counselor
d_gaming_video_creator	d_physical_therapist
d_public_relations_agency	d_psychotherapist
d_manufacturer_supplier	d_allergist
d_school	d_dermatologist
d_test_preparation_center	d_software_company
d_college_university	d_tanning_salon
d_investing_service	d_water_park
d_food_delivery_service	d_live_music_venue

Classificação e Análise do Uso do Solo com Base em Pontos de Interesse

d_bar	d_performance_art_theatre
d_lawyer_law_firm	d_motorcycle_manufacturer
d_legal_service	d_boat_dealership
d_dry_cleaner	d_car_dealership
d_party_entertainment_service	d_automotive_customization_shop
d_pet_service	d_automotive_restoration_service
d_business_service	d_accessories
d_event_planner	d_cosmetics_store
d_movie_television_studio	d_beauty_supply_store
d_radio_station	d_fireplace_store
d_dentist_dental_office	d_furniture_store
d_medical_research_center	d_kiteboarding_center
d_pharmaceutical_company	d_golf_instructor
d_doctor	d_tennis_stadium
d_cultural_center	d_boat_rental
d_hair_replacement_service	d_charter_bus_service
d_tattoo_piercing_shop	d_tour_agency
d_art_gallery	d_beach
d_bingo_hall	d_hiking_trail
d_automotive_store	d_surfing_spot
d_real_estate_agent	d_church_of_christ
d_real_estate_service	d_evangelical_church
d_e_cigarette_store	d_methodist_church
d_mobile_phone_shop	d_health_wellness_website
d_party_supply_rental_shop	d_home_garden_website
d_pet_store	d_recreation_sports_website
d_shopping_mall	d_society_culture_website
d_stadium_arena_sports_venue	amount_baby_children_s_clothing_store_10000m
d_cruise_line	amount_gym_physical_fitness_center_10000m
d_rental_shop	d_goan_restaurant
d_transportation_service	d_brazilian_restaurant
d_marina	d_cuban_restaurant
d_consulate_embassy	d_chiropractor
d_prison_correctional_facility	d_reflexologist
d_county	d_contemporary_art_museum
d_port	d_modern_art_museum
d_geographical_place	d_baby_children_s_clothing_store
d_hindu_temple	d_maternity_nursing_clothing_store
d_pharmaceuticals	d_gym_physical_fitness_center
d_health_beauty	d_sightseeing_tour_agency
d_jewelry_watches	

## F.10. Atributos do teste 6, do cenário 5 com base no CORINE

Na Tabela 71, são apresentados os atributos utilizados no teste 6, do cenário 5, das experiências realizadas com base nos dados do CORINE.

Tabela 71: Conjunto de atributos 10.

Atributos	
amount_sports_club_10000m	d_public_utility_company
amount_commercial_industrial_10000m	d_tattoo_piercing_shop
amount_education_10000m	d_real_estate_service
amount_local_service_10000m	d_pet_store
amount_nonprofit_organization_10000m	d_rental_shop
amount_beauty_cosmetic_personal_care_10000m	d_geographical_place
amount_arts_entertainment_10000m	d_hindu_temple
amount_real_estate_10000m	d_health_beauty
amount_shopping_retail_10000m	d_product_service
amount_sports_recreation_10000m	d_home_decor
amount_event_10000m	amount_european_restaurant_10000m
d_charity_organization	amount_portuguese_restaurant_10000m
d_commercial_industrial	amount_veterinarian_10000m
d_education	amount_photographer_10000m
d_local_service	amount_alternative_holistic_health_service_10000m
d_medical_health	amount_hair_salon_10000m
d_nonprofit_organization	amount_nail_salon_10000m
d_beauty_cosmetic_personal_care	amount_car_dealership_10000m
d_arts_entertainment	amount_accessories_10000m
d_shopping_retail	amount_tour_agency_10000m
d_sports_recreation	d_appliance_manufacturer
d_automated_teller_machine_atm_	d_caribbean_restaurant
d_event	d_european_restaurant
amount_school_10000m	d_italian_restaurant
amount_restaurant_10000m	d_portuguese_restaurant
amount_bar_10000m	d_lounge
amount_design_fashion_10000m	d_cleaning_service
amount_business_service_10000m	d_dog_walker
amount_event_planner_10000m	d_veterinarian
amount_beauty_salon_10000m	d_photographer
amount_tattoo_piercing_shop_10000m	d_home_mover
amount_real_estate_agent_10000m	d_alternative_holistic_health_service
amount_health_beauty_10000m	d_psychotherapist
amount_product_service_10000m	d_hair_salon

amount_home_decor_10000m	d_massage_service
d_magazine	d_history_museum
d_movie	d_motorcycle_manufacturer
d_public_relations_agency	d_car_dealership
d_manufacturer_supplier	d_accessories
d_school	d_beauty_supply_store
d_restaurant	d_fireplace_store
d_bar	d_tour_agency
d_design_fashion	d_hiking_trail
d_pet_service	d_surfing_spot
d_business_service	amount_gym_physical_fitness_center_10000m
d_event_planner	d_cuban_restaurant
d_movie_television_studio	d_gym_physical_fitness_center
d_dentist_dental_office	d_sightseeing_tour_agency

## F.11. Atributos do teste 7, do cenário 5, com base no CORINE

Na Tabela 72, são apresentados os atributos utilizados no teste 7, do cenário 5, das experiências realizadas com base nos dados do CORINE.

Tabela 72: Conjunto de atributos 11.

Atributos	
amount_commercial_industrial_10000m	d_art_gallery
amount_education_10000m	d_party_supply_rental_shop
amount_local_service_10000m	d_pet_store
amount_nonprofit_organization_10000m	d_retail_company
amount_beauty_cosmetic_personal_care_10000m	d_thrift_consignment_store
amount_arts_entertainment_10000m	d_rental_shop
amount_real_estate_10000m	d_recreation_spot
amount_shopping_retail_10000m	d_hindu_temple
amount_sports_recreation_10000m	d_religious_center
d_charity_organization	d_health_beauty
d_government_organization	d_product_service
d_religious_organization	d_home_decor
d_sports_club	d_kitchen_cooking
d_commercial_industrial	amount_european_restaurant_10000m
d_education	amount_portuguese_restaurant_10000m
d_local_service	amount_veterinarian_10000m
d_medical_health	amount_photographer_10000m
d_nonprofit_organization	amount_alternative_holistic_health_service_10000m
d_beauty_cosmetic_personal_care	amount_hair_salon_10000m

d_shopping_retail	amount_car_dealership_10000m
d_sports_recreation	amount_accessories_10000m
d_automated_teller_machine_atm_	d_appliance_manufacturer
amount_school_10000m	d_portuguese_restaurant
amount_restaurant_10000m	d_cleaning_service
amount_bar_10000m	d_dog_walker
amount_design_fashion_10000m	d_veterinarian
amount_business_service_10000m	d_photographer
amount_event_planner_10000m	d_house_painting
amount_beauty_salon_10000m	d_landscape_company
amount_tattoo_piercing_shop_10000m	d_general_dentist
amount_real_estate_agent_10000m	d_alternative_holistic_health_service
amount_health_beauty_10000m	d_psychotherapist
amount_product_service_10000m	d_psychologist
amount_home_decor_10000m	d_hair_salon
d_public_relations_agency	d_message_service
d_school	d_car_dealership
d_financial_service	d_automotive_restoration_service
d_investment_management_company	d_accessories
d_restaurant	d_fireplace_store
d_service_apartments	d_fishing_store
d_legal_service	d_car_rental
d_design_fashion	d_hiking_trail
d_pet_service	d_surfing_spot
d_business_service	amount_gym_physical_fitness_center_10000m
d_event_planner	d_women_s_clothing_store
d_dentist_dental_office	d_gym_physical_fitness_center
d_tattoo_piercing_shop	



# Apêndice G. Categorias dos Pontos de Interesse

Neste apêndice são apresentadas as categorias nas quais se baseou o processo de geração dos mapas de concentração, utilizados nas análises das correlações.

## G.1. Categorias utilizadas para a geração dos *heatmaps*

Na Tabela 73, são apresentadas as 68 categorias utilizadas para a geração do conjunto inicial de *heatmaps*, as quais foram selecionadas com base nos resultados das etapas anteriores do estudo.

Tabela 73: Categorias utilizadas para a geração dos primeiros *heatmaps*.

<b>Nome da Categoria</b>	<b>Nível</b>	<b>Categoria a que pertence</b>
Accessories	4	Apparel & Clothing
Alternative & Holistic Health Service	4	Medical Service
Appliance Manufacturer	4	Manufacturer/Supplier
Art Gallery	3	Arts & Entertainment
Arts & Entertainment	2	Businesses
Automated Teller Machine (ATM)	2	Non-Business Places
Automotive Restoration Service	4	Automotive Service
Bar	3	Food & Beverage
Beauty, Cosmetic & Personal Care	2	Businesses
Beauty Salon	3	Beauty, Cosmetic & Personal Care
Business Service	3	Local Service
Car Dealership	4	Automotive Dealership
Car Rental	4	Rental Shop
Charity Organization	2	Community Organization
Cleaning Service	4	In-Home Service
Commercial & Industrial	2	Businesses
Dentist & Dental Office	3	Medical & Health
Design & Fashion	3	Local Service
Dog Walker	4	Pet Service
Education	2	Businesses
European Restaurant	4	Restaurant
Event Planner	3	Local Service
Financial Service	3	Finance
Fireplace Store	4	Home & Garden Store
Fishing Store	4	Sporting Goods Store
General Dentist	4	Dentist & Dental Office

Government Organization	2	Community Organization
Gym/Physical Fitness Center	5	Fitness Venue
Hair Salon	4	Beauty Salon
Health/Beauty	3	Brand
Hiking Trail	4	Recreation Spot
Hindu Temple	3	Religious Place of Worship
Home Decor	3	Brand
House Painting	4	Home Improvement
Investment Management Company	3	Finance
Kitchen/Cooking	3	Brand
Landscape Company	4	Home Improvement
Local Service	2	Businesses
Massage Service	4	Spa
Medical & Health	2	Businesses
Nonprofit Organization	2	Businesses
Party Supply & Rental Shop	3	Shopping & Retail
Pet Service	3	Local Service
Pet Store	3	Shopping & Retail
Photographer	4	Photography Videography
Portuguese Restaurant	4	Restaurant
Product/Service	3	Brand
Psychologist	4	Doctor
Psychotherapist	4	Therapist
Public Relations Agency	3	Advertising/Marketing
Real Estate	2	Businesses
Real Estate Agent	3	Real Estate
Recreation Spot	3	Outdoor Recreation
Religious Center	3	Religious Place of Worship
Religious Organization	2	Community Organization
Rental Shop	3	Travel & Transportation
Restaurant	3	Food & Beverage
Retail Company	3	Shopping & Retail
School	3	Education
Service Apartments	3	Hotel & Lodging
Shopping & Retail	2	Businesses
Sports Club	2	Community Organization
Sports & Recreation	2	Businesses
Surfing Spot	4	Recreation Spot
Tattoo & Piercing Shop	3	Beauty, Cosmetic & Personal Care
Thrift & Consignment Store	3	Shopping & Retail
Veterinarian	4	Pet Service
Women's Clothing Store	5	Clothing Store

# Apêndice H. Correlações

Neste apêndice são apresentados os resultados obtidos nas análises das correlações, com base na metodologia proposta no estudo.

## H.1. Correlações com os tipos de uso dos espaços

São listadas nesta secção, as correlações obtidas entre as diversas categorias de empresas, serviços e pontos de interesse, e as classes de uso e ocupação dos espaços.

### H.1.1. Categorias selecionadas com base nas etapas anteriores

Neste tópico são listados os resultados referentes à classe “1. Artificial surfaces”, do conjunto de dados CORINE *land cover*.

#### H.1.1.1. Classe “1. Artificial surfaces”

A seguir, na Tabela 74, são exibidos os resultados das correlações existentes entre as categorias dos pontos de interesse e a classe “1. Artificial surfaces”.

Tabela 74: Correlações entre as categorias e a classe 1, do CORINE.

Categoria	Correlação		
	1250m	2500m	5000m
Accessories	0.239753	0.351576	0.471549
Alternative holistic health service	0.251465	0.333983	0.418646
Appliance manufacturer	0.058328	0.090323	0.139926
Art gallery	0.126139	0.201676	0.308650
Arts entertainment	0.371180	0.459783	0.526573
Automated teller machine atm	-0.004757	-0.009518	-0.019062
Automotive restoration service	0.060023	0.113861	0.194716
Bar	0.270877	0.349536	0.425957
Beauty cosmetic personal care	0.409037	0.473118	0.507004
Beauty salon	0.376329	0.442976	0.483867
Business service	0.365486	0.463478	0.539077
Car dealership	0.215675	0.315805	0.420280
Car rental	0.090745	0.155755	0.247797
Charity organization	0.089404	0.155181	0.260631

Cleaning service	0.129868	0.192806	0.272442
Industrial or commercial	0.225553	0.277556	0.306763
Dentist dental office	0.187040	0.270418	0.383022
Design fashion	0.213462	0.311523	0.405882
Dog walker	0.075617	0.119240	0.181946
Education	0.436287	0.535278	0.594404
European restaurant	0.192560	0.262692	0.329452
Event planner	0.333139	0.448664	0.517203
Financial service	0.180590	0.266704	0.389106
Fireplace store	0.047513	0.058557	0.082112
Fishing store	0.058298	0.079092	0.098789
General dentist	0.127814	0.202473	0.320938
Government organization	0.282857	0.373079	0.434436
Gym physical fitness center	0.347311	0.453277	0.530743
Hair salon	0.340372	0.409374	0.459095
Health beauty	0.362843	0.479770	0.566915
Hiking trail	0.020639	0.017967	0.037750
Hindu temple	0.005807	0.011453	0.016749
Home decor	0.221950	0.316675	0.429224
House painting	0.048360	0.063777	0.112382
Investment management company	0.057132	0.098175	0.162607
Kitchen cooking	0.109755	0.184211	0.263343
Landscape company	0.036618	0.038474	0.041028
Local service	0.490635	0.541463	0.543361
Massage service	0.216972	0.278411	0.331588
Medical health	0.423117	0.519108	0.581258
Nonprofit organization	0.385625	0.482451	0.546400
Party supply rental shop	0.123654	0.211624	0.326903
Pet service	0.314710	0.409131	0.467801
Pet store	0.186665	0.279353	0.360150
Photographer	0.228423	0.354142	0.450433
Portuguese restaurant	0.280306	0.359452	0.427442
Product service	0.278578	0.399081	0.504859
Psychologist	0.150015	0.227307	0.312892
Psychotherapist	0.095791	0.173397	0.274266
Public relations agency	0.063085	0.110692	0.202832
Real estate	0.383834	0.496376	0.556196
Real estate agent	0.324266	0.426421	0.491157
Recreation spot	0.092008	0.092549	0.100242
Religious center	0.109757	0.173937	0.238342
Religious organization	0.230341	0.325257	0.405817

Rental shop	0.120882	0.193283	0.284714
Restaurant	0.381753	0.447932	0.489071
Retail company	0.160378	0.257209	0.389797
School	0.356183	0.456315	0.521566
Service apartments	0.046272	0.079946	0.132065
Shopping & retail	0.405836	0.503866	0.554263
Sports club	0.263823	0.385736	0.488840
Sports & recreation	0.454211	0.532571	0.559584
Surfing spot	0.080715	0.082160	0.085718
Tattoo piercing shop	0.263480	0.368652	0.462021
Thrift consignment store	0.060740	0.091968	0.133480
Veterinarian	0.235496	0.335969	0.433840
Women s clothing store	0.205936	0.288669	0.366015

### H.1.1.2. Classe “1.1. Urban fabric”

A seguir, na Tabela 75, são exibidos os resultados das correlações existentes entre as categorias dos pontos de interesse e a classe “1.1. Urban fabric”.

Tabela 75: Correlações entre as categorias a classe 1.1, do CORINE.

Categoria	Correlação		
	1250m	2500m	5000m
Accessories	0.238135	0.329806	0.405189
Alternative holistic health service	0.252083	0.315571	0.374629
Appliance manufacturer	0.046348	0.070612	0.110803
Art gallery	0.125723	0.194372	0.273900
Arts entertainment	0.333571	0.415395	0.456747
Automated teller machine atm	-0.003928	-0.007858	-0.015737
Automotive restoration service	0.045077	0.076179	0.152944
Bar	0.277883	0.341801	0.394510
Beauty cosmetic personal care	0.433745	0.461298	0.451502
Beauty salon	0.403519	0.434780	0.431286
Business service	0.350055	0.424485	0.476275
Car dealership	0.196898	0.282896	0.351178
Car rental	0.081836	0.139953	0.209844
Charity organization	0.088728	0.147791	0.226964
Cleaning service	0.120046	0.167487	0.230839
Industrial or commercial	0.193824	0.233682	0.258735
Dentist dental office	0.191667	0.254251	0.327035
Design fashion	0.213992	0.296841	0.373752

Dog walker	0.093198	0.144754	0.200111
Education	0.425247	0.496747	0.516448
European restaurant	0.201725	0.259819	0.306311
Event planner	0.328620	0.423188	0.467368
Financial service	0.196241	0.267613	0.369599
Fireplace store	0.052035	0.058684	0.083687
Fishing store	0.051143	0.069673	0.088574
General dentist	0.131394	0.192516	0.274595
Government organization	0.263196	0.329681	0.357340
Gym physical fitness center	0.306061	0.392527	0.449273
Hair salon	0.366983	0.402742	0.412201
Health beauty	0.352597	0.429884	0.476286
Hiking trail	0.027018	0.025575	0.045004
Hindu temple	0.009718	0.019251	0.025434
Home decor	0.231389	0.307250	0.383981
House painting	0.054315	0.064102	0.099522
Investment management company	0.065660	0.096083	0.153233
Kitchen cooking	0.108139	0.173704	0.220485
Landscape company	0.046398	0.051378	0.046143
Local service	0.492297	0.511213	0.485274
Massage service	0.223585	0.259689	0.289119
Medical health	0.419407	0.481450	0.504722
Nonprofit organization	0.380908	0.449480	0.470295
Party supply rental shop	0.123697	0.190873	0.287820
Pet service	0.321758	0.395940	0.427631
Pet store	0.158866	0.224941	0.292009
Photographer	0.231530	0.337179	0.405278
Portuguese restaurant	0.288166	0.352858	0.390330
Product service	0.275952	0.372988	0.437551
Psychologist	0.159599	0.222997	0.290224
Psychotherapist	0.100169	0.179806	0.277154
Public relations agency	0.048037	0.085162	0.163934
Real estate	0.384663	0.471171	0.498155
Real estate agent	0.323735	0.401994	0.437189
Recreation spot	0.103178	0.106024	0.119385
Religious center	0.125975	0.184863	0.244700
Religious organization	0.223978	0.292665	0.355584
Rental shop	0.100880	0.171095	0.242321
Restaurant	0.390245	0.434730	0.448290
Retail company	0.117248	0.195093	0.308577
School	0.359392	0.437897	0.466534

Service apartments	0.051211	0.071283	0.106962
Shopping & retail	0.390312	0.461363	0.479884
Sports club	0.268332	0.359333	0.426382
Sports & recreation	0.414012	0.472074	0.484821
Surfing spot	0.091448	0.094874	0.103977
Tattoo piercing shop	0.268338	0.349713	0.401283
Thrift consignment store	0.046103	0.071510	0.096385
Veterinarian	0.236712	0.321828	0.390732
Women s clothing store	0.203010	0.258290	0.303397

### H.1.1.3. Classe “111. Continuous urban fabric”

A seguir, na Tabela 76, são exibidos os resultados das correlações existentes entre as categorias dos pontos de interesse e a classe “111. Continuous urban fabric”.

Tabela 76: Correlações entre as categorias e a classe 111, do CORINE.

Categoria	Correlação		
	1250m	2500m	5000m
Accessories	0.222984	0.277211	0.290454
Alternative holistic health service	0.230906	0.258612	0.246379
Appliance manufacturer	0.008532	0.016312	0.032271
Art gallery	0.116778	0.175022	0.215204
Arts entertainment	0.300491	0.352019	0.349422
Automated teller machine atm	-0.001885	-0.003772	-0.007554
Automotive restoration service	0.051890	0.085075	0.115104
Bar	0.261223	0.280239	0.272425
Beauty cosmetic personal care	0.334934	0.320673	0.256295
Beauty salon	0.303882	0.295597	0.238152
Business service	0.191143	0.235201	0.260219
Car dealership	0.077615	0.131892	0.183411
Car rental	0.070774	0.119395	0.149042
Charity organization	0.058455	0.108670	0.193636
Cleaning service	0.061292	0.076191	0.099142
Industrial or commercial	0.064586	0.067646	0.069155
Dentist dental office	0.167525	0.220199	0.252298
Design fashion	0.125233	0.170385	0.189971
Dog walker	-0.002951	0.008196	0.018892
Education	0.314048	0.365230	0.351652
European restaurant	0.174380	0.200779	0.190314
Event planner	0.151468	0.195096	0.217030

Financial service	0.092027	0.122424	0.158531
Fireplace store	-0.003772	-0.007554	-0.012516
Fishing store	0.026137	0.044494	0.047536
General dentist	0.130056	0.192512	0.241956
Government organization	0.321434	0.388695	0.387747
Gym physical fitness center	0.248024	0.297802	0.305373
Hair salon	0.292382	0.292820	0.234836
Health beauty	0.336645	0.377241	0.362785
Hiking trail	-0.003772	-0.007076	-0.000985
Hindu temple	-0.003266	-0.006539	-0.009904
Home decor	0.123038	0.178406	0.205740
House painting	0.010486	0.011656	-0.001457
Investment management company	0.007962	0.008284	0.028335
Kitchen cooking	0.047943	0.099397	0.119782
Landscape company	-0.003772	-0.006582	-0.009950
Local service	0.247872	0.251577	0.219195
Massage service	0.135338	0.132467	0.120877
Medical health	0.295117	0.335982	0.322424
Nonprofit organization	0.311784	0.370761	0.357865
Party supply rental shop	0.078473	0.142685	0.219846
Pet service	0.146996	0.191120	0.188709
Pet store	0.100479	0.132142	0.165122
Photographer	0.154180	0.197422	0.203485
Portuguese restaurant	0.192248	0.214431	0.202699
Product service	0.201782	0.251614	0.255131
Psychologist	0.098585	0.130558	0.167701
Psychotherapist	0.022982	0.082315	0.152609
Public relations agency	0.022799	0.038383	0.081198
Real estate	0.283620	0.331566	0.320566
Real estate agent	0.255817	0.295864	0.282763
Recreation spot	0.036956	0.030070	0.021002
Religious center	0.087290	0.102191	0.109038
Religious organization	0.173776	0.199245	0.188156
Rental shop	0.073234	0.117355	0.143392
Restaurant	0.297522	0.301355	0.267123
Retail company	0.080046	0.116799	0.173452
School	0.219200	0.263917	0.264438
Service apartments	-0.003253	0.003226	0.032748
Shopping & retail	0.265025	0.300274	0.281114
Sports club	0.191419	0.223850	0.232212
Sports & recreation	0.265380	0.280053	0.259437

Surfing spot	0.030755	0.022914	0.013816
Tattoo piercing shop	0.293407	0.335994	0.309617
Thrift consignment store	0.036058	0.035796	0.038758
Veterinarian	0.146954	0.208694	0.224789
Women s clothing store	0.170974	0.198045	0.203396

#### H.1.1.4. Classe “112. Discontinuous urban fabric”

A seguir, na Tabela 77, são exibidos os resultados das correlações existentes entre as categorias dos pontos de interesse e a classe “112. Discontinuous urban fabric”.

Tabela 77: Correlações entre as categorias e a classe 112, do CORINE.

Categoria	Correlação		
	1250m	2500m	5000m
Accessories	0.143041	0.216295	0.294573
Alternative holistic health service	0.154376	0.210777	0.285111
Appliance manufacturer	0.047903	0.071075	0.107689
Art gallery	0.076062	0.120772	0.188271
Arts entertainment	0.207190	0.270776	0.319391
Automated teller machine atm	-0.003390	-0.006783	-0.013585
Automotive restoration service	0.021506	0.037851	0.108043
Bar	0.166330	0.228221	0.292777
Beauty cosmetic personal care	0.301516	0.341119	0.366999
Beauty salon	0.284942	0.325331	0.354404
Business service	0.288893	0.348352	0.392970
Car dealership	0.179699	0.246461	0.294621
Car rental	0.052526	0.090772	0.153351
Charity organization	0.067468	0.105874	0.147200
Cleaning service	0.101521	0.147005	0.205988
Industrial or commercial	0.183694	0.227350	0.255029
Dentist dental office	0.122004	0.163007	0.227473
Design fashion	0.171778	0.240201	0.316569
Dog walker	0.107894	0.160227	0.217150
Education	0.303851	0.355873	0.386135
European restaurant	0.129520	0.180526	0.239524
Event planner	0.287297	0.369952	0.407672
Financial service	0.170658	0.234494	0.329928
Fireplace store	0.061463	0.071215	0.102561
Fishing store	0.043237	0.053789	0.073575
General dentist	0.074885	0.108593	0.173670
Government organization	0.114949	0.152004	0.184066

Gym physical fitness center	0.206033	0.275915	0.336220
Hair salon	0.249928	0.290423	0.334565
Health beauty	0.208067	0.272773	0.333965
Hiking trail	0.032957	0.033214	0.051848
Hindu temple	0.012953	0.025699	0.034680
Home decor	0.192864	0.247447	0.319152
House painting	0.055857	0.066336	0.114241
Investment management company	0.070237	0.104717	0.158301
Kitchen cooking	0.095635	0.140737	0.182313
Landscape company	0.055040	0.062331	0.058304
Local service	0.418331	0.437754	0.426830
Massage service	0.176895	0.219686	0.259889
Medical health	0.308090	0.355272	0.389592
Nonprofit organization	0.254631	0.298831	0.329970
Party supply rental shop	0.095795	0.135392	0.201462
Pet service	0.282052	0.341191	0.378690
Pet store	0.123207	0.180278	0.237724
Photographer	0.175106	0.270608	0.344716
Portuguese restaurant	0.217735	0.278686	0.328136
Product service	0.198332	0.280228	0.351773
Psychologist	0.125133	0.178975	0.234206
Psychotherapist	0.100915	0.157519	0.227997
Public relations agency	0.041619	0.074954	0.140076
Real estate	0.275116	0.346100	0.383178
Real estate agent	0.221688	0.287819	0.335461
Recreation spot	0.096303	0.103509	0.123951
Religious center	0.093317	0.151844	0.216087
Religious organization	0.155225	0.218836	0.296911
Rental shop	0.072810	0.127431	0.193610
Restaurant	0.273476	0.321961	0.357109
Retail company	0.087542	0.155096	0.251809
School	0.283388	0.347112	0.379443
Service apartments	0.060225	0.079369	0.103037
Shopping & retail	0.292252	0.352931	0.385060
Sports club	0.195612	0.280645	0.352234
Sports & recreation	0.319053	0.376771	0.403159
Surfing spot	0.086506	0.094921	0.110530
Tattoo piercing shop	0.136934	0.205154	0.279094
Thrift consignment store	0.031785	0.060887	0.087527
Veterinarian	0.185169	0.246630	0.315883
Women s clothing store	0.132945	0.180358	0.228677

### H.1.1.5. Classe “121. Industrial or commercial units and public facilities”

A seguir, na Tabela 78, são exibidos os resultados das correlações existentes entre as categorias dos pontos de interesse e a classe “121. Industrial or commercial and public facilities”.

Tabela 78: Correlações entre as categorias e a classe 121, do CORINE.

Categoria	Correlação		
	1250m	2500m	5000m
Accessories	0.051752	0.089466	0.152994
Alternative holistic health service	0.062058	0.108312	0.144913
Appliance manufacturer	0.047366	0.062354	0.057100
Art gallery	0.016186	0.033778	0.077796
Arts entertainment	0.102776	0.127739	0.149729
Automated teller machine atm	-0.001763	-0.003528	-0.007065
Automotive restoration service	0.044024	0.077306	0.086526
Bar	0.047531	0.077081	0.111152
Beauty cosmetic personal care	0.071525	0.134492	0.185460
Beauty salon	0.059685	0.124177	0.181789
Business service	0.115872	0.155598	0.172233
Car dealership	0.091993	0.126632	0.167468
Car rental	0.039198	0.063924	0.114459
Charity organization	0.007709	0.011925	0.027236
Cleaning service	0.027543	0.063511	0.102368
Industrial or commercial	0.137897	0.152279	0.133841
Dentist dental office	0.039747	0.081272	0.130706
Design fashion	0.050148	0.087318	0.122632
Dog walker	-0.004513	-0.001566	0.019213
Education	0.106472	0.146992	0.189012
European restaurant	0.023550	0.046335	0.081125
Event planner	0.071882	0.116514	0.145995
Financial service	0.026842	0.060480	0.093901
Fireplace store	0.004184	0.019466	0.023669
Fishing store	0.005015	0.012147	0.018556
General dentist	0.035146	0.066430	0.106194
Government organization	0.052579	0.078196	0.107710
Gym physical fitness center	0.147872	0.186125	0.214244
Hair salon	0.051417	0.116550	0.171314
Health beauty	0.106894	0.178808	0.223848
Hiking trail	-0.003528	-0.007065	-0.012870
Hindu temple	-0.003055	-0.006116	-0.004841
Home decor	0.038853	0.081428	0.139783

House painting	0.009754	0.026066	0.060871
Investment management company	0.003072	0.011935	0.015408
Kitchen cooking	0.043466	0.064451	0.108181
Landscape company	-0.003528	-0.006545	-0.001609
Local service	0.122221	0.163768	0.181151
Massage service	0.053784	0.101505	0.142192
Medical health	0.095927	0.152495	0.192464
Nonprofit organization	0.068180	0.114443	0.158256
Party supply rental shop	0.033118	0.071637	0.098511
Pet service	0.071945	0.114094	0.145315
Pet store	0.105676	0.161306	0.171990
Photographer	0.065060	0.118020	0.163409
Portuguese restaurant	0.048544	0.072078	0.110730
Product service	0.068520	0.109926	0.165801
Psychologist	0.020594	0.046498	0.080666
Psychotherapist	-0.004599	-0.004461	0.015211
Public relations agency	0.050195	0.070575	0.090621
Real estate	0.082436	0.132347	0.174106
Real estate agent	0.075461	0.124893	0.168569
Recreation spot	-0.007821	-0.012566	-0.019267
Religious center	0.007656	0.028561	0.045278
Religious organization	0.088017	0.139784	0.157635
Rental shop	0.039639	0.061084	0.102520
Restaurant	0.076196	0.106331	0.130997
Retail company	0.126261	0.154985	0.173245
School	0.072813	0.117801	0.166124
Service apartments	-0.003054	0.000450	0.017230
Shopping & retail	0.138123	0.180924	0.208365
Sports club	0.049449	0.112095	0.171663
Sports & recreation	0.123912	0.156006	0.176009
Surfing spot	-0.009468	-0.013750	-0.019894
Tattoo piercing shop	0.055096	0.109220	0.183696
Thrift consignment store	0.050384	0.059934	0.069478
Veterinarian	0.057758	0.090790	0.137023
Women s clothing store	0.070224	0.124533	0.171322

## H.1.2. Subcategorias a nível 3 da categoria “Business”

Este tópico apresenta as correlações obtidas entre as subcategorias, a nível 3, da categoria “Business”, e as classes de uso e ocupação dos espaços escolhidas. Os resultados podem ser visualizados a seguir, na Tabela 79.

Tabela 79: Correlações entre as classes e as subcategories (nível 3) da categoria “Business”.

	Artificial surfaces	Urban fabric	Cont. urban fabric	Disc. urban fabric	Industrial/commercial
Academic camp	0.08746	0.06704	0.02475	0.06215	0.03811
Advertising agency	0.27469	0.24437	0.11870	0.21016	0.07579
Aerospace company	0.12331	0.06599	0.04960	0.04665	0.09499
Agricultural cooperative	0.14559	0.15053	0.02874	0.15498	0.03605
Agricultural service	0.09406	0.08753	0.00492	0.09691	0.03206
Airline company	0.23337	0.18539	0.18156	0.10678	0.04339
Airline industry service	0.11335	0.08530	0.02099	0.08512	0.07908
Amusement theme park	0.20326	0.17205	0.09952	0.13879	0.08422
Antique store	0.18072	0.17077	0.03324	0.17546	0.02844
Apparel clothing	<b>0.47775</b>	<b>0.40815</b>	0.28044	<b>0.30371</b>	0.18955
Aquarium	0.03125	0.02463	0.00060	0.02772	-0.00706
Arcade	0.20168	0.19545	0.09732	0.16672	0.06610
Archaeological service	0.12439	0.09998	0.08894	0.06275	0.02859
Art gallery	<b>0.30865</b>	0.27390	0.21520	0.18827	0.07780
Arts crafts store	<b>0.30124</b>	0.26916	0.13605	0.22841	0.12338
Astrologist	0.08356	0.07819	0.06354	0.05253	0.02945
Astrologist psychic	0.05280	0.04301	-0.00069	0.04941	0.02251
Auction house	0.08782	0.06272	-0.01015	0.07731	0.03984
Automation service	0.05538	0.03239	0.03264	0.01813	0.06747
Automotive dealership	<b>0.47660</b>	<b>0.38330</b>	0.21597	<b>0.31249</b>	0.21557
Automotive service	<b>0.43239</b>	<b>0.35200</b>	0.15328	<b>0.31290</b>	0.22014
Automotive store	<b>0.41892</b>	<b>0.34835</b>	0.13151	<b>0.32127</b>	0.19685
Bakery	<b>0.32166</b>	<b>0.30844</b>	0.15214	0.26391	0.07243
Bank	0.16050	0.15101	0.19010	0.06269	0.02217
Bar	<b>0.42596</b>	<b>0.39451</b>	0.27242	0.29278	0.11115
Barber shop	<b>0.39640</b>	<b>0.36369</b>	0.22942	0.28241	0.14090
Beauty salon	<b>0.48387</b>	<b>0.43129</b>	0.23815	<b>0.35440</b>	0.18179
Beauty store	<b>0.39432</b>	<b>0.35238</b>	0.25730	0.25348	0.12271

Classificação e Análise do Uso do Solo com Base em Pontos de Interesse

Beauty supplier	0.02919	0.03228	0.04572	0.01047	-0.00267
Bed and breakfast	0.06309	0.08338	-0.01370	0.10289	-0.01531
Betting shop	0.00674	0.01739	-0.01313	0.02737	-0.00923
Bicycle repair service	0.12519	0.11425	0.07386	0.08769	0.03102
Big box retailer	0.09679	0.08095	0.02848	0.07585	0.04271
Bingo hall	0.18042	0.15390	0.18142	0.07098	0.07770
Biotechnology company	0.04638	0.05141	0.00839	0.05376	0.00053
Boat ferry company	0.04287	-0.00077	-0.00468	0.00182	0.09820
Book magazine distributor	0.21039	0.18090	0.12096	0.13653	0.03915
Bookstore	0.29565	0.27751	0.18817	0.20794	0.06033
Boutique store	0.03769	0.03365	0.01205	0.03141	0.02964
Brewery	0.23017	0.20850	0.08296	0.18985	0.07214
Broadcasting media production	<b>0.36004</b>	0.29268	0.24560	0.19218	0.07447
Bubble tea shop	0.11768	0.10857	0.12183	0.05361	0.05555
Bus line	0.11225	0.09283	0.03727	0.08433	0.04085
Business service	<b>0.53908</b>	<b>0.47627</b>	0.26022	<b>0.39297</b>	0.17223
Butcher shop	0.21775	0.21404	0.07734	0.19939	0.05131
Cabin	0.12272	0.04995	0.02689	0.04144	0.05187
Cafe	<b>0.37483</b>	<b>0.34666</b>	0.15053	<b>0.30839</b>	0.11683
Cafeteria	<b>0.32281</b>	0.27405	0.15957	0.22046	0.12522
Campground	0.08335	0.09215	0.01587	0.09587	-0.01354
Casino	0.19470	0.17300	0.11750	0.12951	0.06590
Caterer	<b>0.31872</b>	<b>0.30246</b>	0.13357	0.26779	0.06810
Cheese shop	0.18567	0.14206	0.14297	0.07960	0.04266
Chemical company	0.11264	0.09309	0.00954	0.10059	0.05583
Child protective service	0.04072	0.04001	0.00853	0.04068	0.00056
Cigarette store	0.19369	0.18961	0.13441	0.13872	0.04793
Circus	0.12854	0.08297	0.04499	0.06865	0.10810
Collectibles store	0.24169	0.19055	0.16116	0.12439	0.09297
College university	<b>0.35727</b>	0.26628	0.27209	0.14685	0.12339
Commercial industrial equipment supplier	0.13431	0.10496	0.02261	0.10659	0.09039
Commercial real estate agency	0.16209	0.15467	0.10001	0.11870	0.02362
Community center	0.10430	0.10851	0.11700	0.05632	0.00707
Computer repair service	0.20164	0.18796	0.04201	0.19000	0.04473
Convenience store	0.17833	0.15030	0.10539	0.11062	0.07576
Cottage	0.07795	0.07847	-0.02089	0.10144	0.00644
Cruise line	0.13768	0.12006	0.15861	0.04555	0.03207
Cultural center	0.28329	0.25632	<b>0.33141</b>	0.10137	0.08010
Cultural gifts store	0.14470	0.14693	0.00796	0.16284	0.03902
Dance night club	<b>0.30283</b>	0.26796	0.28827	0.13946	0.06551
Defense company	0.11706	0.10249	0.02101	0.10470	0.03830

Deli	0.12899	0.10746	0.01351	0.11468	0.01674
Dentist dental office	<b>0.38302</b>	<b>0.32704</b>	0.25230	0.22747	0.13071
Department store	0.11843	0.11226	-0.00445	0.13047	0.02413
Design fashion	<b>0.40588</b>	<b>0.37375</b>	0.18997	<b>0.31657</b>	0.12263
Dessert shop	<b>0.40275</b>	<b>0.36224</b>	0.21098	0.29137	0.13105
Discount store	0.11411	0.09357	0.06454	0.06949	0.06597
Distillery	0.00654	0.00593	-0.00591	0.01015	-0.00098
Doctor	<b>0.47085</b>	<b>0.39982</b>	0.28905	0.28926	0.14466
Dry cleaner	0.18823	0.18683	0.08625	0.16326	0.03436
Educational consultant	0.25135	0.23716	0.09027	0.21830	0.06017
Educational research center	0.09522	0.08754	0.07041	0.05924	0.01215
Educational supply store	0.12640	0.11917	0.12909	0.06152	0.02138
Electronics store	0.27641	0.23969	0.13134	0.19755	0.11933
Engineering service	0.28891	0.24355	0.11618	0.21067	0.12330
Environmental service	0.10319	0.10109	0.03824	0.09318	0.01459
Escape game room	-0.00259	0.00364	-0.00755	0.00849	-0.00706
Event planner	<b>0.51720</b>	<b>0.46737</b>	0.21703	<b>0.40767</b>	0.14600
Event space	<b>0.32666</b>	0.27340	0.14353	0.22895	0.13151
Fabric store	0.22860	0.19111	0.17611	0.11643	0.07400
Farm	0.14613	0.13591	0.00920	0.14957	0.03760
Farmers market	0.13594	0.13952	0.03984	0.13605	0.02950
Financial service	<b>0.38911</b>	<b>0.36960</b>	0.15853	<b>0.32993</b>	0.09390
Flea market	0.08436	0.10216	-0.01072	0.12258	-0.00379
Florist	0.24754	0.21509	0.13072	0.16987	0.05936
Food consultant	0.19420	0.14501	0.11512	0.09899	0.07005
Food delivery service	0.28241	0.23972	0.17911	0.17009	0.08473
Food stand	0.09815	0.08024	0.04210	0.06721	0.05247
Food truck	0.03653	0.02109	0.00852	0.01913	0.03259
Food wholesaler	0.25010	0.18907	0.11458	0.14950	0.15443
Foodservice distributor	0.27839	0.25010	0.11322	0.21984	0.11449
Forestry logging	0.04894	0.03413	-0.00699	0.04292	0.02535
Funeral service cemetery	0.15050	0.13942	0.07444	0.11603	0.03953
Gift shop	0.19921	0.18183	0.12396	0.13586	0.03501
Glass mirror shop	0.09331	0.05141	0.01876	0.04779	0.09189
Grocery store	<b>0.34534</b>	<b>0.31203</b>	0.13467	0.27806	0.10357
Gun store	0.08807	0.06865	0.00032	0.07805	0.06620
Hair extensions service	0.10523	0.08652	0.05785	0.06530	0.03629
Hair removal service	<b>0.31655</b>	0.29610	0.14244	0.25543	0.10272
Hair replacement service	0.12662	0.11437	0.09630	0.07491	0.04604
Hobby store	0.12336	0.10916	0.07305	0.08235	0.05430
Home garden store	<b>0.42330</b>	<b>0.38393</b>	0.15034	<b>0.35097</b>	0.14055

Classificação e Análise do Uso do Solo com Base em Pontos de Interesse

Home improvement	<b>0.38793</b>	<b>0.35075</b>	0.11628	<b>0.33276</b>	0.14530
Hostel	0.14482	0.15926	0.08780	0.13095	-0.00716
Hotel	0.26695	0.25690	0.12427	0.22123	0.02134
Hotel services company	0.13741	0.10096	0.05695	0.08228	0.00992
housing assistance service	-0.00282	0.00336	-0.00755	0.00818	-0.00706
Image consultant	0.23386	0.20782	0.09003	0.18501	0.05942
In home service	<b>0.36057</b>	<b>0.30612</b>	0.15051	0.26221	0.11947
Industrial company	0.16535	0.13558	0.00351	0.15246	0.08576
Information technology company	<b>0.36582</b>	<b>0.31797</b>	0.20335	0.24531	0.14080
Inn	0.06741	0.08055	-0.01024	0.09768	-0.01415
Insurance company	<b>0.33695</b>	0.29464	0.12093	0.26614	0.12847
Internet cafe	0.18407	0.15858	0.02368	0.16707	0.06309
Internet marketing service	0.20336	0.18898	0.04576	0.18901	0.07699
Investing service	0.18485	0.14130	0.06038	0.12626	0.04625
Investment management company	0.16261	0.15323	0.02833	0.15830	0.01541
Karaoke	0.20052	0.17142	0.13643	0.11682	0.08466
Laundromat	<b>0.38810</b>	<b>0.34339</b>	0.11637	<b>0.32433</b>	0.13107
Law enforcement agency	0.11938	0.10819	0.05666	0.09067	0.01472
Lawyer law firm	0.21269	0.15990	0.08773	0.13172	0.08148
Legal service	0.17865	0.11835	0.07752	0.09025	0.05602
Library	0.28421	0.21805	0.18288	0.14323	0.07198
Locksmith	0.06912	0.08782	-0.01628	0.10944	-0.01450
Lodge	0.07755	0.09541	0.00875	0.10368	-0.00355
Lottery retailer	-0.00304	-0.00237	-0.00568	0.00057	0.00451
Makeup artist	<b>0.33690</b>	<b>0.30084</b>	0.12933	0.26838	0.12442
Manufacturer supplier	0.22794	0.19919	0.07920	0.18139	0.06854
Market research consultant	0.10042	0.09298	0.04836	0.07812	0.04335
Marketing agency	<b>0.30297</b>	0.23684	0.14954	0.18383	0.15073
Media agency	0.16907	0.15322	0.09207	0.12161	0.06073
Medical center	<b>0.48388</b>	<b>0.41410</b>	0.28027	<b>0.31059</b>	0.15544
Medical equipment manufacturer	0.10488	0.11556	-0.00643	0.13537	-0.00030
Medical equipment supplier	0.06343	0.05852	0.02231	0.05385	0.01350
Medical research center	0.20610	0.17835	0.21431	0.07991	0.02516
Medical service	<b>0.44073</b>	<b>0.38812</b>	0.26451	0.29005	0.15535
Merchandising service	0.02622	0.00163	-0.00430	0.00433	0.04519
Metal steel company	0.13650	0.13059	0.05291	0.11836	0.04128
Mining company	0.06641	0.04101	-0.00917	0.05201	0.05019
Mobile phone shop	0.29989	0.25788	0.28376	0.13058	0.10647
Motel	0.13688	0.09095	0.04055	0.08030	0.11339
Motor vehicle company	<b>0.43084</b>	<b>0.37171</b>	0.14072	<b>0.34259</b>	0.17441
Movie music store	0.11896	0.11357	0.05003	0.10062	0.01653

Movie television studio	0.16096	0.13964	0.06785	0.12008	0.04186
Movie theater	<b>0.32397</b>	0.27472	0.22723	0.18228	0.11454
Moving storage service	0.21004	0.16630	0.03652	0.16848	0.11715
Museum	0.21493	0.16690	0.15500	0.10099	0.04257
Music production studio	0.26911	0.25619	0.16287	0.19821	0.08247
Musical instrument store	0.22218	0.20480	0.12704	0.16027	0.05930
Newsstand	0.08984	0.07489	0.02688	0.06987	0.01368
Optician	0.03480	0.03507	-0.00755	0.04431	0.01182
Outdoor sporting goods company	0.26453	0.21092	0.09384	0.18634	0.13328
Outlet store	0.07590	0.06555	0.04172	0.05068	0.04653
parking garage lot	0.20480	0.14424	0.08058	0.11800	0.08491
Party entertainment service	<b>0.47138</b>	<b>0.40217</b>	0.25754	<b>0.31007</b>	0.17811
Party supply rental shop	<b>0.32690</b>	0.28782	0.21985	0.20146	0.09851
Passport visa service	0.13327	0.14832	0.02092	0.15697	0.01512
Performance event venue	<b>0.46986</b>	<b>0.40046</b>	<b>0.34599</b>	0.25723	0.12804
Personal chef	0.01718	0.01958	0.02427	0.00834	-0.00270
Pet service	<b>0.46780</b>	<b>0.42763</b>	0.18871	<b>0.37869</b>	0.14531
Pet store	<b>0.36015</b>	0.29201	0.16512	0.23772	0.17199
Pharmaceutical company	0.18525	0.15408	0.10244	0.11663	0.08183
Pharmacy drugstore	0.28525	0.25222	0.06824	0.24814	0.12080
Photography videography	<b>0.46728</b>	<b>0.42415</b>	0.21527	<b>0.35944</b>	0.16366
Plastic company	0.05410	0.01726	-0.00738	0.02391	0.10366
Pool billiard hall	0.14849	0.13458	0.10009	0.09576	0.02567
Pop up shop	0.07966	0.03387	-0.00640	0.04227	0.05751
Printing service	0.29012	0.24700	0.11897	0.21300	0.12144
Property management company	0.21396	0.20600	0.07997	0.18872	0.03902
Public relations agency	0.20283	0.16393	0.08120	0.14008	0.09062
Public service	<b>0.32925</b>	0.27881	0.17571	0.21659	0.10945
Public utility company	<b>0.39751</b>	<b>0.33215</b>	0.21899	0.25246	0.15025
Publisher	<b>0.32541</b>	0.23440	0.19582	0.15442	0.13325
Race track	0.13633	0.07221	0.02600	0.06733	0.10575
Radio station	<b>0.37022</b>	<b>0.30310</b>	0.25915	0.19625	0.09708
Railroad company	0.08426	0.08526	0.04057	0.07380	0.00057
Real estate agent	<b>0.49116</b>	<b>0.43719</b>	0.28276	<b>0.33546</b>	0.16857
Real estate appraiser	0.03946	0.03627	0.00149	0.04047	-0.00189
Real estate company	<b>0.32793</b>	0.28852	0.14751	0.24388	0.09722
Real estate developer	<b>0.30466</b>	0.27160	0.16047	0.21715	0.10416
Real estate investment firm	0.25841	0.22694	0.12461	0.18689	0.07916
Real estate service	<b>0.46741</b>	<b>0.40423</b>	0.29307	0.29197	0.17432
Real estate title development	0.29018	0.24809	0.17395	0.18260	0.06585
Rent to own store	-0.00190	-0.00484	-0.00735	-0.00129	0.00620

Classificação e Análise do Uso do Solo com Base em Pontos de Interesse

Rental shop	0.28471	0.24232	0.14339	0.19361	0.10252
Restaurant	<b>0.48907</b>	<b>0.44829</b>	0.26712	<b>0.35711</b>	0.13100
Retail company	<b>0.38980</b>	<b>0.30858</b>	0.17345	0.25181	0.17325
Robotics company	0.05011	0.02950	0.01968	0.02229	0.06594
Rv park	0.01226	0.01194	-0.00304	0.01536	0.00586
Sandwich shop	0.14607	0.12580	0.09883	0.08648	0.02963
School	<b>0.52157</b>	<b>0.46653</b>	0.26444	<b>0.37944</b>	0.16612
Seasonal store	0.06167	0.04403	0.01128	0.04368	0.03670
Service apartments	0.13206	0.10696	0.03275	0.10304	0.01723
Sewing alterations	0.22584	0.20507	0.12631	0.16099	0.07585
Shoe repair shop	0.03820	0.02541	-0.00730	0.03315	0.03933
Shopping mall	<b>0.50318</b>	<b>0.41464</b>	<b>0.30672</b>	0.29597	0.17851
Shopping service	0.22253	0.20952	0.09293	0.18527	0.05764
Skin care service	<b>0.37860</b>	<b>0.33820</b>	0.17855	0.28263	0.13198
Smoothie juice bar	0.19017	0.17892	0.10668	0.14249	0.02242
Social media agency	0.19596	0.20165	0.06932	0.18989	0.01010
Social service	0.23545	0.23362	0.15146	0.17906	0.04152
Solar energy company	0.12696	0.11631	0.04430	0.10704	0.04056
Souvenir shop	0.04649	0.04563	0.01632	0.04260	0.02760
Spa	<b>0.39177</b>	<b>0.34006</b>	0.15175	<b>0.30017</b>	0.14551
Specialty school	<b>0.46966</b>	<b>0.40601</b>	<b>0.34659</b>	0.26320	0.16169
Sporting goods store	<b>0.36803</b>	<b>0.31194</b>	0.16141	0.26256	0.14500
Sports fitness instruction	<b>0.35465</b>	<b>0.31503</b>	0.20411	0.24153	0.10612
Sports league	0.21908	0.22282	0.11196	0.18947	0.02681
Sports recreation venue	<b>0.55284</b>	<b>0.46904</b>	0.27080	<b>0.37863</b>	0.19261
Sports team	<b>0.42867</b>	<b>0.37425</b>	0.17278	<b>0.32703</b>	0.11315
Stadium arena sports venue	<b>0.46191</b>	<b>0.39190</b>	0.26315	0.29514	0.15121
Tattoo piercing shop	<b>0.46202</b>	<b>0.40128</b>	<b>0.30962</b>	0.27909	0.18370
Test preparation center	0.23127	0.20766	0.09526	0.18181	0.07698
Textile company	0.15815	0.12469	0.08772	0.09160	0.07398
Therapist	<b>0.44248</b>	<b>0.38752</b>	0.17465	<b>0.34108</b>	0.14317
Thrift consignment store	0.13348	0.09639	0.03876	0.08753	0.06948
Tobacco store	0.11931	0.10551	0.07408	0.07760	0.03669
Toy store	0.27370	0.25345	0.17853	0.18607	0.07325
Transit hub	0.27819	0.18982	0.12314	0.14544	0.11381
Transportation service	<b>0.38727</b>	<b>0.32354</b>	0.17635	0.26719	0.17656
Travel company	<b>0.40246</b>	<b>0.35868</b>	0.20610	0.29011	0.08234
Travel service	0.19816	0.19198	0.06088	0.18372	0.02215
Trophies engraving shop	0.03603	0.04905	-0.00755	0.06024	-0.00706
Tutor teacher	0.26087	0.25196	0.10336	0.22762	0.05856
Vacation home rental	0.13902	0.15828	0.00859	0.17542	-0.01309

Video game store	0.08280	0.05877	0.02002	0.05544	0.04600
Vintage store	0.17789	0.16068	0.05093	0.15378	0.04150
Waste management company	0.11676	0.06844	0.02894	0.06134	0.07748
Weight loss center	0.15354	0.10706	0.12961	0.04742	0.08483
Wholesale supply store	0.03070	0.02849	-0.01659	0.04201	0.02048
Wine beer spirits store	0.19321	0.17421	0.12808	0.12481	0.03855
Winery vineyard	0.05726	0.06329	0.01278	0.06476	-0.00102
Writing service	0.03319	0.04315	-0.01070	0.05532	-0.01001
Zoo	0.22025	0.15267	0.11984	0.10501	0.09385

## H.2. Correlações entre as categorias

São listadas nesta secção, as correlações obtidas entre as diversas categorias de empresas, serviços e pontos de interesse (umas com as outras).

### H.2.1. Mapas de densidade com raio de 1250 metros

Neste tópico são listados os resultados obtidos aquando da utilização dos mapas de densidade gerados com raio de 1250 metros, que podem ser visualizados a seguir, na Tabela 80.

Tabela 80: Correlações entre as categorias (raio dos *heatmaps*: 1250m).

	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]	[13]	[14]	[15]	[16]	[17]	[18]	[19]	[20]	[21]	[22]
[1] Advertising marketing	1	0.13	0.23	0.28	0.22	0.15	0.22	0.13	0.21	0.12	0.06	0.37	0.16	0.28	0.07	0.21	0.10	0.19	0.09	0.25	0.26	0.20
[2] Agriculture	0.13	1	0.15	0.21	0.17	0.12	0.17	0.09	0.19	0.17	0.04	0.27	0.05	0.17	0.06	0.17	0.09	0.11	0.07	0.19	0.16	0.14
[3] Arts entertainment	0.23	0.15	1	0.24	0.44	0.19	0.47	0.23	0.59	0.30	0.10	0.51	0.31	0.43	0.16	0.49	0.38	0.36	0.14	0.42	0.48	0.38
[4] Automotive aircraft boat	0.28	0.21	0.24	1	0.40	0.33	0.33	0.21	0.33	0.09	0.14	0.56	0.12	0.36	0.06	0.24	0.16	0.31	0.20	0.44	0.37	0.25
[5] Beauty cosmetic personal care	0.22	0.17	0.44	0.40	1	0.26	0.59	0.44	0.67	0.23	0.18	0.68	0.18	0.63	0.14	0.51	0.29	0.51	0.22	0.58	0.59	0.29
[6] Commercial industrial	0.15	0.12	0.19	0.33	0.26	1	0.23	0.19	0.26	0.11	0.06	0.41	0.06	0.26	0.05	0.18	0.14	0.22	0.24	0.32	0.23	0.18
[7] Education	0.22	0.17	0.47	0.33	0.59	0.23	1	0.33	0.52	0.20	0.16	0.63	0.21	0.58	0.16	0.50	0.37	0.44	0.23	0.49	0.53	0.29
[8] Finance	0.13	0.09	0.23	0.21	0.44	0.19	0.33	1	0.35	0.14	0.16	0.41	0.13	0.42	0.08	0.26	0.24	0.35	0.14	0.34	0.30	0.20
[9] Food beverage	0.21	0.19	0.59	0.33	0.67	0.26	0.52	0.35	1	0.48	0.16	0.64	0.22	0.54	0.16	0.55	0.34	0.45	0.17	0.56	0.61	0.40
[10] Hotel lodging	0.12	0.17	0.30	0.09	0.23	0.11	0.20	0.14	0.48	1	0.08	0.32	0.08	0.23	0.05	0.25	0.13	0.20	0.03	0.23	0.37	0.31
[11] Legal	0.06	0.04	0.10	0.14	0.18	0.06	0.16	0.16	0.16	0.08	1	0.20	0.08	0.16	0.06	0.17	0.10	0.18	0.07	0.17	0.16	0.14
[12] Local service	0.37	0.27	0.51	0.56	0.68	0.41	0.63	0.41	0.64	0.32	0.20	1	0.21	0.66	0.14	0.51	0.34	0.55	0.29	0.67	0.62	0.43
[13] Media news company	0.16	0.05	0.31	0.12	0.18	0.06	0.21	0.13	0.22	0.08	0.08	0.21	1	0.21	0.08	0.26	0.23	0.18	0.11	0.16	0.19	0.15
[14] Medical health	0.28	0.17	0.43	0.36	0.63	0.26	0.58	0.42	0.54	0.23	0.16	0.66	0.21	1	0.13	0.46	0.33	0.48	0.22	0.55	0.52	0.34
[15] Non governmental organization	0.07	0.06	0.16	0.06	0.14	0.05	0.16	0.08	0.16	0.05	0.06	0.14	0.08	0.13	1	0.22	0.12	0.12	0.05	0.11	0.14	0.07
[16] Nonprofit organization	0.21	0.17	0.49	0.24	0.51	0.18	0.50	0.26	0.55	0.25	0.17	0.51	0.26	0.46	0.22	1	0.35	0.37	0.17	0.41	0.48	0.34
[17] Public government service	0.10	0.09	0.38	0.16	0.29	0.14	0.37	0.24	0.34	0.13	0.10	0.34	0.23	0.33	0.12	0.35	1	0.26	0.16	0.29	0.29	0.25
[18] Real estate	0.19	0.11	0.36	0.31	0.51	0.22	0.44	0.35	0.45	0.20	0.18	0.55	0.18	0.48	0.12	0.37	0.26	1	0.19	0.46	0.46	0.29
[19] Science technology engineering	0.09	0.07	0.14	0.20	0.22	0.24	0.23	0.14	0.17	0.03	0.07	0.29	0.11	0.22	0.05	0.17	0.16	0.19	1	0.22	0.21	0.14
[20] Shopping retail	0.25	0.19	0.42	0.44	0.58	0.32	0.49	0.34	0.56	0.23	0.17	0.67	0.16	0.55	0.11	0.41	0.29	0.46	0.22	1	0.49	0.33
[21] Sports recreation	0.26	0.16	0.48	0.37	0.59	0.23	0.53	0.30	0.61	0.37	0.16	0.62	0.19	0.52	0.14	0.48	0.29	0.46	0.21	0.49	1	0.38
[22] Travel transportation	0.20	0.14	0.38	0.25	0.29	0.18	0.29	0.20	0.40	0.31	0.14	0.43	0.15	0.34	0.07	0.34	0.25	0.29	0.14	0.33	0.38	1

## H.2.2. Mapas de densidade com raio de 2500 metros

Neste tópico são listados os resultados obtidos aquando da utilização dos mapas de densidade gerados com raio de 2500 metros, que podem ser visualizados a seguir, na Tabela 81.

Tabela 81: Correlações entre as categorias (raio dos *heatmaps*: 2500m).

	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]	[13]	[14]	[15]	[16]	[17]	[18]	[19]	[20]	[21]	[22]
[1] Advertising marketing	1	0.21	0.36	0.46	0.35	0.31	0.37	0.27	0.35	0.18	0.18	0.54	0.28	0.44	0.16	0.35	0.25	0.35	0.28	0.42	0.41	0.34
[2] Agriculture	0.21	1	0.21	0.32	0.26	0.27	0.27	0.17	0.3	0.23	0.05	0.39	0.09	0.27	0.11	0.24	0.15	0.18	0.14	0.28	0.25	0.22
[3] Arts entertainment	0.36	0.21	1	0.38	0.53	0.29	0.63	0.34	0.66	0.35	0.18	0.6	0.46	0.57	0.29	0.65	0.56	0.53	0.26	0.56	0.58	0.54
[4] Automotive aircraft boat	0.46	0.32	0.38	1	0.58	0.49	0.5	0.35	0.46	0.13	0.23	0.72	0.21	0.52	0.15	0.38	0.31	0.49	0.36	0.65	0.54	0.39
[5] Beauty cosmetic personal care	0.35	0.26	0.53	0.58	1	0.41	0.7	0.53	0.74	0.26	0.22	0.76	0.27	0.71	0.22	0.6	0.39	0.64	0.36	0.73	0.7	0.41
[6] Commercial industrial	0.31	0.27	0.29	0.49	0.41	1	0.38	0.27	0.38	0.16	0.09	0.57	0.11	0.36	0.13	0.31	0.24	0.33	0.39	0.47	0.38	0.29
[7] Education	0.37	0.27	0.63	0.5	0.7	0.38	1	0.5	0.65	0.26	0.27	0.75	0.37	0.76	0.3	0.67	0.53	0.63	0.41	0.67	0.69	0.49
[8] Finance	0.27	0.17	0.34	0.35	0.53	0.27	0.5	1	0.48	0.24	0.24	0.53	0.26	0.57	0.18	0.37	0.4	0.45	0.26	0.48	0.45	0.34
[9] Food beverage	0.35	0.3	0.66	0.46	0.74	0.38	0.65	0.48	1	0.55	0.22	0.73	0.32	0.65	0.26	0.66	0.48	0.6	0.29	0.69	0.73	0.56
[10] Hotel lodging	0.18	0.23	0.35	0.13	0.26	0.16	0.26	0.24	0.55	1	0.13	0.37	0.13	0.29	0.11	0.3	0.22	0.28	0.07	0.29	0.46	0.43
[11] Legal	0.18	0.05	0.18	0.23	0.22	0.09	0.27	0.24	0.22	0.13	1	0.28	0.19	0.24	0.12	0.25	0.22	0.27	0.13	0.27	0.24	0.28
[12] Local service	0.54	0.39	0.6	0.72	0.76	0.57	0.75	0.53	0.73	0.37	0.28	1	0.32	0.75	0.25	0.63	0.47	0.68	0.44	0.79	0.75	0.57
[13] Media news company	0.28	0.09	0.46	0.21	0.27	0.11	0.37	0.26	0.32	0.13	0.19	0.32	1	0.38	0.18	0.43	0.41	0.35	0.24	0.29	0.31	0.27
[14] Medical health	0.44	0.27	0.57	0.52	0.71	0.36	0.76	0.57	0.65	0.29	0.24	0.75	0.38	1	0.25	0.62	0.54	0.63	0.37	0.7	0.67	0.52
[15] Non governmental organization	0.16	0.11	0.29	0.15	0.22	0.13	0.3	0.18	0.26	0.11	0.12	0.25	0.18	0.25	1	0.35	0.25	0.26	0.14	0.22	0.25	0.18
[16] Nonprofit organization	0.35	0.24	0.65	0.38	0.6	0.31	0.67	0.37	0.66	0.3	0.25	0.63	0.43	0.62	0.35	1	0.51	0.55	0.31	0.57	0.61	0.51
[17] Public government service	0.25	0.15	0.56	0.31	0.39	0.24	0.53	0.4	0.48	0.22	0.22	0.47	0.41	0.54	0.25	0.51	1	0.45	0.28	0.46	0.45	0.44
[18] Real estate	0.35	0.18	0.53	0.49	0.64	0.33	0.63	0.45	0.6	0.28	0.27	0.68	0.35	0.63	0.26	0.55	0.45	1	0.34	0.63	0.65	0.46
[19] Science technology engineering	0.28	0.14	0.26	0.36	0.36	0.39	0.41	0.26	0.29	0.07	0.13	0.44	0.24	0.37	0.14	0.31	0.28	0.34	1	0.37	0.36	0.25
[20] Shopping retail	0.42	0.28	0.56	0.65	0.73	0.47	0.67	0.48	0.69	0.29	0.27	0.79	0.29	0.7	0.22	0.57	0.46	0.63	0.37	1	0.65	0.48
[21] Sports recreation	0.41	0.25	0.58	0.54	0.7	0.38	0.69	0.45	0.73	0.46	0.24	0.75	0.31	0.67	0.25	0.61	0.45	0.65	0.36	0.65	1	0.53
[22] Travel transportation	0.34	0.22	0.54	0.39	0.41	0.29	0.49	0.34	0.56	0.43	0.28	0.57	0.27	0.52	0.18	0.51	0.44	0.46	0.25	0.48	0.53	1

## H.2.3. Mapas de densidade com raio de 5000 metros

Neste tópico são listados os resultados obtidos aquando da utilização dos mapas de densidade gerados com raio de 2500 metros, que podem ser visualizados a seguir, na Tabela 82.

Tabela 82: Correlações entre as categorias (raio dos *heatmaps*: 5000m).

	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]	[13]	[14]	[15]	[16]	[17]	[18]	[19]	[20]	[21]	[22]
[1] Advertising marketing	1	0.35	0.54	0.65	0.54	0.5	0.6	0.47	0.56	0.3	0.36	0.71	0.49	0.65	0.33	0.54	0.49	0.56	0.54	0.64	0.63	0.54
[2] Agriculture	0.35	1	0.27	0.45	0.4	0.46	0.37	0.33	0.46	0.32	0.12	0.53	0.13	0.4	0.21	0.36	0.24	0.28	0.3	0.41	0.4	0.32
[3] Arts entertainment	0.54	0.27	1	0.52	0.62	0.39	0.78	0.53	0.73	0.4	0.35	0.67	0.67	0.73	0.5	0.82	0.75	0.69	0.45	0.71	0.69	0.7
[4] Automotive aircraft boat	0.65	0.45	0.52	1	0.75	0.63	0.67	0.55	0.63	0.21	0.35	0.84	0.34	0.69	0.34	0.56	0.49	0.66	0.6	0.82	0.71	0.56
[5] Beauty cosmetic personal care	0.54	0.4	0.62	0.75	1	0.58	0.77	0.64	0.83	0.33	0.3	0.83	0.37	0.78	0.38	0.7	0.5	0.73	0.56	0.85	0.78	0.54
[6] Commercial industrial	0.5	0.46	0.39	0.63	0.58	1	0.51	0.39	0.53	0.25	0.17	0.72	0.18	0.49	0.23	0.47	0.35	0.44	0.56	0.63	0.53	0.42
[7] Education	0.6	0.37	0.78	0.67	0.77	0.51	1	0.66	0.78	0.36	0.41	0.83	0.58	0.88	0.55	0.83	0.71	0.79	0.62	0.82	0.83	0.7
[8] Finance	0.47	0.33	0.53	0.55	0.64	0.39	0.66	1	0.63	0.4	0.3	0.67	0.45	0.71	0.34	0.56	0.58	0.58	0.45	0.64	0.64	0.56
[9] Food beverage	0.56	0.46	0.73	0.63	0.83	0.53	0.78	0.63	1	0.6	0.34	0.82	0.42	0.77	0.47	0.77	0.59	0.74	0.5	0.81	0.85	0.69
[10] Hotel lodging	0.3	0.32	0.4	0.21	0.33	0.25	0.36	0.4	0.6	1	0.23	0.45	0.18	0.38	0.23	0.38	0.27	0.36	0.15	0.38	0.53	0.54
[11] Legal	0.36	0.12	0.35	0.35	0.3	0.17	0.41	0.3	0.34	0.23	1	0.4	0.35	0.35	0.26	0.38	0.38	0.35	0.23	0.39	0.38	0.49
[12] Local service	0.71	0.53	0.67	0.84	0.83	0.72	0.83	0.67	0.82	0.45	0.4	1	0.43	0.82	0.44	0.74	0.59	0.77	0.64	0.88	0.86	0.7
[13] Media news company	0.49	0.13	0.67	0.34	0.37	0.18	0.58	0.45	0.42	0.18	0.35	0.43	1	0.58	0.39	0.64	0.67	0.52	0.43	0.47	0.46	0.47
[14] Medical health	0.65	0.4	0.73	0.69	0.78	0.49	0.88	0.71	0.77	0.38	0.35	0.82	0.58	1	0.47	0.79	0.73	0.75	0.59	0.83	0.81	0.71
[15] Non governmental organization	0.33	0.21	0.5	0.34	0.38	0.23	0.55	0.34	0.47	0.23	0.26	0.44	0.39	0.47	1	0.58	0.48	0.45	0.32	0.42	0.47	0.43
[16] Nonprofit organization	0.54	0.36	0.82	0.56	0.7	0.47	0.83	0.56	0.77	0.38	0.38	0.74	0.64	0.79	0.58	1	0.73	0.73	0.53	0.75	0.76	0.71
[17] Public government service	0.49	0.24	0.75	0.49	0.5	0.35	0.71	0.58	0.59	0.27	0.38	0.59	0.67	0.73	0.48	0.73	1	0.61	0.5	0.63	0.6	0.67
[18] Real estate	0.56	0.28	0.69	0.66	0.73	0.44	0.79	0.58	0.74	0.36	0.35	0.77	0.52	0.75	0.45	0.73	0.61	1	0.54	0.75	0.81	0.63
[19] Science technology engineering	0.54	0.3	0.45	0.6	0.56	0.56	0.62	0.45	0.5	0.15	0.23	0.64	0.43	0.59	0.32	0.53	0.5	0.54	1	0.58	0.58	0.48
[20] Shopping retail	0.64	0.41	0.71	0.82	0.85	0.63	0.82	0.64	0.81	0.38	0.39	0.88	0.47	0.83	0.42	0.75	0.63	0.75	0.58	1	0.8	0.65
[21] Sports recreation	0.63	0.4	0.69	0.71	0.78	0.53	0.83	0.64	0.85	0.53	0.38	0.86	0.46	0.81	0.47	0.76	0.6	0.81	0.58	0.8	1	0.72
[22] Travel transportation	0.54	0.32	0.7	0.56	0.54	0.42	0.7	0.56	0.69	0.54	0.49	0.7	0.47	0.71	0.43	0.71	0.67	0.63	0.48	0.65	0.72	1