

Instituto Politécnico de Coimbra  
Instituto Superior de Contabilidade  
e Administração de Coimbra

Tiago André Gomes Alves

Análise e previsão de audiências no mercado televisivo  
e de comunicação em Portugal

Análise e previsão de audiências no mercado televisivo e de comunicação em Portugal

Tiago André Gomes Alves

ISCAC | 2021

Coimbra, julho de 2021





Instituto Politécnico de Coimbra  
Instituto Superior de Contabilidade  
e Administração de Coimbra

Tiago André Gomes Alves

## Análise e previsão de audiências no mercado televisivo e de comunicação em Portugal

Dissertação submetida ao Instituto Superior de Contabilidade e Administração de Coimbra para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Análise de Dados e Sistemas de Apoio à Decisão realizada sob a orientação do Professor Doutor Fernando Paulo dos Santos Rodrigues Belfo.

Coimbra, julho de 2021

## **TERMO DE RESPONSABILIDADE**

Declaro ser o autor desta dissertação, que constitui um trabalho original e inédito, que nunca foi submetido a outra Instituição de ensino superior para obtenção de um grau académico ou outra habilitação. Atesto ainda que todas as citações estão devidamente identificadas e que tenho consciência de que o plágio constitui uma grave falta de ética, que poderá resultar na anulação da presente dissertação.

## **DEDICATÓRIA**

Aos meus avós Amílcar, António, Celestina, Laurinda, Maria e à Ti Cidália.

## **AGRADECIMENTOS**

Este trabalho reflete um percurso de vários anos, que só foi possível concluir com a ajuda e incentivo de diversas pessoas, desde familiares, amigos e professores.

Em primeiro lugar, gostava de deixar expresso um agradecimento especial à minha família, pelo papel fulcral que desempenham: aos meus pais Florbela e Jacinto e irmãos Bernardo e Matilde, sem esquecer todos os avós, tios e primos (já está, Cátia!). Estou eternamente grato por todo o apoio, paciência e confiança que depositaram em mim durante este percurso. Sem eles não seria mais difícil, mas sim impossível.

Quero agradecer também ao meu orientador nesta etapa, o Professor Doutor Fernando Belfo, por toda a disponibilidade, motivação, compreensão e suporte que sempre apresentou. Foi um percurso longo, mas o Professor esteve sempre presente, ajudando a guiar este projeto com ricos contributos, num trabalho incansável. Pretendo também referir a Professora Doutora Isabel Pedrosa, uma das docentes mais atenciosas e dedicadas das quais tive o privilégio de ser aluno e que está diretamente ligada ao meu gosto pelo estudo dos dados e das audiências.

Queria agradecer ainda a todas as minhas colegas de trabalho: Catarina, Ema, Patrícia, Paula e Vera que, para além de profissionais de exceção, são pessoas amigas, atenciosas, dedicadas e formam a melhor equipa que podia pedir. Obrigado por acreditarem em mim desde o início e por toda a ajuda inexcelável.

O meu sincero agradecimento também a todos os meus amigos, que são uma presença constante na minha vida e com os quais qualquer celebração tem muito mais significado.

Resta-me agradecer a todos professores e colegas que foram ajudando no meu trajeto, tornando-me hoje mais capaz do que ontem e menos do que amanhã.

A todos os que tornaram isto possível, o meu muito obrigado!

## **RESUMO**

O futebol é um desporto que foi transformado num negócio à escala global devido ao seu mediatismo e universalidade. Apesar das origens humildes, movimentam milhões de euros em todo o mundo e tem impacto em toda a sociedade. Uma das fontes de receitas que alimenta o negócio do futebol é a publicidade, pois constitui um espaço que diversas marcas aproveitam para se promover e fazer crescer os seus negócios rapidamente. Associado à publicidade estão as audiências televisivas dos jogos, pois são um barómetro do número de pessoas alcançadas em cada encontro e permitem aos diversos agentes envolvidos no desporto aferir os públicos a que conseguem chegar e em que quantidade.

Sabendo da importância dos dados audiométricos no negócio do futebol, este projeto pretende encontrar fatores que justifiquem e prevejam as audiências televisivas, no contexto dos campeonatos portugueses. Foram considerados todos os jogos da Primeira e Segunda Liga de Portugal, nas épocas de 2018/2019 e de 2019/2020, esta última marcada por uma interrupção causada pela Covid-19 e, posteriormente, com jogos disputados à porta fechada, uma situação transversal a todo o globo e sem precedentes.

A metodologia utilizada foi a CRISP-DM e tem seis fases: entendimento do negócio, entendimento dos dados, preparação dos dados, modelação, avaliação do modelo e implementação do modelo. O conjunto de dados é constituído por 798 transmissões de jogos de futebol, caracterizados através de 55 variáveis, assentes em atributos do jogo, do desempenho das equipas, do contexto geográfico, económico e demográfico em que as equipas estão inseridas e ainda relativos às audiências que essas transmissões alcançaram.

O modelo final proposto é o mais simples dos quatro modelos analisados e inclui o poder de compra dos municípios associados às equipas, os valores de mercado dos plantéis, o preço dos canais televisivos e o número de jogos em simultâneo como as variáveis independentes que mostraram assumir influência significativa nas audiências televisivas.

Este estudo propõe um modelo a utilizar futuramente em novos trabalhos pela Academia e um conhecimento valioso para as partes interessadas no negócio do futebol, tais como as televisões, clubes e patrocinadores, poderem otimizar as audiências televisivas dos adeptos no futuro. Algumas limitações deste projeto foram apresentadas e alguns trabalhos para o futuro foram também propostos.

Palavras-chave: audiências, televisão, futebol, campeonato português, CRISP-DM, regressão, mineração de dados

## **ABSTRACT**

Football is a sport that turned into a global scale business due to its media coverage and universal character. Although its humble roots, it moves millions of euros everywhere and impacts several spheres of society. One of the sources of revenue in the football's business is publicity since it constitutes a platform that many brands use to self-promote and make their business quickly grow. Associated with advertisement, there are television audiences of sporting events, being the barometer for the number of people reached in every match and allowing the involved agents in sport to assess the type and quantity of audience that may be achieved.

Knowing the importance of audiometric data in the football business, this research pretends to find factors that justify and predict TV audiences in the context of Portuguese championships. Were considered every game from Portugal's first and second division in the 2018/2019 and 2019/2020 seasons, this last season affected by an interruption caused by Covid-19 and, posteriorly, with games disputed behind closed doors, a reality happening across the world and without precedents.

The methodology used in this project was CRISP-DM and contemplate six steps: business understanding, data understanding, data preparation, modelling, evaluation, and deployment. The data set analyzed was constituted by 789 football games, characterized through 55 variables, from game information, team's performance, the team's geographic, economic and demographic context and data related to audiences that the game's transmission reached.

The final proposed model is the simplest of the four analyzed models and includes the buying power of inhabitants from the club location, squad's market value, the price of TV subscription and the number of matches happening simultaneously as the independent variables that showed significant influence on TV audiences.

This project proposes a model that can be used in the Academy and gives valuable knowledge to the interested parts in the football business, such as television broadcasters, clubs and sponsors, to optimize fans' reach and TV audiences in the future. Some limitations of this project were presented, and some future research was proposed.

Keywords: tv audiences, television, football, portuguese football, CRISP-DM, regression, data mining

## **ÍNDICE GERAL**

1	INTRODUÇÃO .....	1
2	REVISÃO DA LITERATURA .....	3
2.1	Mercado Televisivo e de Comunicação .....	3
2.1.1	Evolução histórica global da televisão.....	3
2.1.2	Evolução história da televisão em Portugal .....	4
2.1.3	Diferentes abordagens dos órgãos de comunicação .....	5
2.1.4	Futuro do mercado da televisão .....	6
2.2	Possíveis influências nas audiências televisivas .....	9
2.2.1	Duração .....	10
2.2.2	Dia da semana .....	11
2.2.3	Hora da emissão .....	13
2.2.4	Temporada .....	15
2.2.5	Número do episódio .....	16
2.2.6	Periodicidade.....	16
2.2.7	Classe social dos telespetadores .....	17
2.2.8	Partidas em simultâneo .....	18
2.2.9	Outras influências nas emissões desportivas .....	19
2.3	Previsão aplicada no âmbito de audiências .....	20
2.3.1	Aprendizagem automática na previsão de audiências de televisão .....	20
2.3.2	Previsão de audiências de televisão através das redes sociais .....	23
2.4	Negócio das transmissões televisivas de jogos de futebol .....	25
2.4.1	A Premier League .....	26
2.4.2	O futebol português .....	29
3	METODOLOGIA .....	34

4	ENTENDIMENTO DOS DADOS .....	38
4.1	Dados de televisão.....	38
4.2	Unidade de análise .....	38
4.3	Recolha dos dados.....	39
4.3.1	Procedimento de recolha de dados.....	39
4.3.2	Dados de jogos e desempenho .....	40
4.3.3	Dados geográficos, económicos e demográficos .....	41
4.3.4	Dados de audiências.....	42
4.4	Descrição dos dados .....	43
4.5	Exploração dos dados.....	47
4.5.1	Primeira ou Segunda Liga.....	47
4.5.2	Hora da transmissão .....	48
4.5.3	Dia da semana .....	51
4.5.4	Distância geográfica entre equipas .....	52
4.5.5	Classificação das equipas.....	55
4.5.6	População do município.....	57
4.5.7	Poder de compra dos munícipes .....	59
4.5.8	Participações na competição .....	62
4.5.9	Antiguidade na competição .....	63
4.5.10	Classificação da época anterior.....	65
4.5.11	Valor de mercado dos plantéis.....	66
4.5.12	Golos do jogo.....	67
4.5.13	Capacidade dos estádios .....	68
4.5.14	Canais premium .....	69
4.5.15	Jogos em simultâneo .....	71
5	PREPARAÇÃO DOS DADOS .....	73

5.1	Seleção dos dados .....	73
5.2	Limpeza dos dados .....	73
5.3	Construção dos dados.....	73
5.4	Integração dos dados .....	74
5.5	Formatação dos dados .....	76
6	MODELAÇÃO.....	77
6.1	Seleção da técnica de modelação .....	77
6.2	Geração do desenho de testes.....	78
6.3	Construção do modelo.....	78
6.4	Revisão do modelo.....	81
7	DISCUSSÃO DOS RESULTADOS .....	84
7.1	Avaliação dos resultados.....	84
7.2	Implementação prática dos modelos .....	84
7.3	Fatores não significativos.....	85
7.4	Fatores significativos ausentes do modelo final.....	86
7.5	Poder de compra dos munícipes.....	88
7.6	Valor de mercado dos plantéis .....	89
7.7	Canais premium .....	89
7.8	Jogos em simultâneo .....	90
8	CONCLUSÃO.....	92
	REFERÊNCIAS BIBLIOGRÁFICAS .....	95
	APÊNDICES .....	101
	APÊNDICE 1: Amostra da base de dados.....	102
	ANEXOS .....	104
	ANEXO I: Classificação final da Primeira Liga em 2018-2019 .....	105

ANEXO II: Valor de mercado das equipas da Primeira Liga, no início da época 2018-2019 .....	106
ANEXO III: Distância geográfica entre o estádio do Futebol Clube do Porto e Grupo Desportivo de Chaves .....	107

## **ÍNDICE DE FIGURAS**

Figura 1- Evolução do número de horas de conteúdo original produzido pela Netflix (Statista) Fonte: Stoll (2020)	7
Figura 2- Gastos estimados com conteúdo original (Variety) Fonte: (Spangler, 2020)	7
Figura 3 - Consumo médio semanal de vídeo por país (em horas) (Deloitte) Fonte: (Boehm et al., 2018)	8
Figura 4 - Programas mais vistos de 28 de junho de 2020 Fonte: (CAEM/GFK, 2021)	10
Figura 5 - Audiência média dos principais canais de televisão portugueses no 1º semestre de 2019 por dia da semana Fonte: (CAEM/GFK, 2021)	13
Figura 6 - Consumo televisivo ao longo do dia (Abril de 2021) Fonte: (CAEM/GFK, 2021)	14
Figura 7 - Jogadores do Arsenal depois do 1º jogo televisionado (1937)	27
Figura 8 - Custo dos contratos televisivos da Premier League de 2010 a 2022 Fonte: Gazapo (2018)	29
Figura 9 - Principais marcos do mercado das transmissões televisivas do futebol português	30
Figura 10 - Fases da metodologia CRISP-DM	35
Figura 11 – Diversas proveniências para a construção do conjunto de dados	39
Figura 12 - Audímetro (1 e 3) e comando (2) Fonte: Imagem por Reprodução/GfK em Espalha-Factos (2020)	42
Figura 13 - Relação entre o rat# médio, o número de jogos e as horas do jogo	50
Figura 14 - Número de jogos e rating médio por dia da semana	52
Figura 15 - Distribuição geográfica das equipas no mapa de Portugal	53
Figura 16 - Distribuição do rat# por distância geográfica entre equipas	54
Figura 17 - Dispersão de jogos com o Rat# face à soma da distância pontual das duas equipas para o líder	55

Figura 18 - Relação entre o rating e a diferença de população residente no concelho das equipas _____	58
Figura 19 - Relação entre o rating e a soma de população residente no concelho das equipas _____	59
Figura 20 - Rat# por soma de participações na competição dos 2 clubes de um jogo _	63
Figura 21 - Rat# por soma de antiguidade na competição _____	64
Figura 22 - Rat# por soma da classificação no último ano _____	65
Figura 23 - Rat# por número de jogos em paralelo _____	71
Figura 24 - Processo de integração de dados _____	75

## **ÍNDICE DE TABELAS**

Tabela 1 – Rating médio, número e percentagem de jogos por dia da semana _____	51
Tabela 2 - Classes de SomaDist _____	56
Tabela 3 - Relação entre o número de golos e as audiências médias _____	67
Tabela 4 - Audiências médias por preço de subscrição de canal _____	70
Tabela 5 - Distribuição das variáveis pelos 3 modelos iniciais de treino _____	79
Tabela 6 - Variáveis significativas dos 3 modelos iniciais _____	80
Tabela 7 - Valores-p do modelo com as 13 variáveis inicialmente significativas _____	80
Tabela 8 – Resumo dos principais indicadores dos quatro modelos _____	82

## **Lista de abreviaturas, acrónimos e siglas**

BBC .....	British Broadcasting Corporation
BMO .....	Bank of Montreal
BSkyB.....	British Sky Broadcasting-
CAEM.....	Comissão de Análise de Estudos de Meios
CRISP-DM.....	Cross Industry Standard Process for Data Mining
ERC.....	Entidade Reguladora para a Comunicação Social
EUA .....	Estados Unidos da América
F1 .....	Fórmula 1
FCP .....	Futebol Clube do Porto
FPF.....	Federação Portuguesa de Futebol
GBM .....	Gradient Boosting Machine
GFK .....	Growth from Knowledge
INE.....	Instituto Nacional de Estatística
ITV.....	Independent Television
NBA.....	National Basketball Association
NFL.....	National Football League
OCS.....	Órgão de Comunicação Social
RMSE.....	Root Mean Square Error
RTP .....	Rádio e Televisão de Portugal
SCP .....	Sporting Clube de Portugal
SEN .....	Sistema Estatístico Nacional
SIC .....	Sociedade Independente de Comunicação
SLB .....	Sport Lisboa e Benfica
TVI.....	Televisão Independente
UEFA.....	União das Federações Europeias de Futebol
VOD.....	Vídeo On Demand

## **1 INTRODUÇÃO**

Nos dias que correm todos os conteúdos produzidos têm diversos objetivos e públicos definidos. Um meio de comunicação como uma televisão, um jornal, uma revista, um blog tem identificado um ou mais públicos-alvo e por norma pretende, dentro desse(s) público(s), atingir o maior número de pessoas.

Nesse sentido, as audiências assumem uma enorme importância, pois permitem medir não só quantas pessoas consumiram determinado conteúdo, mas também em que horas, durante quanto tempo, ou em que dispositivo o fizeram. Outra das valências das audiências é a caracterização das pessoas que consumiram o objeto em estudo, revelando a sua classe social, localização, género, faixa etária, entre outros aspetos. Através das audiências é possível comparar diversos meios de comunicação ou tipologias de conteúdos e encontrar faixas de maior consumo. Este importante conhecimento poderá ajudar a definir melhor toda a estratégia de posicionamento de um meio no mercado.

Uma das principais fontes de financiamento dos meios de comunicação é a publicidade que conseguem vender, o que ainda reforça mais a importância das audiências de cada um desses meios. À partida, quanto maior for o número de pessoas alcançadas por um artigo de um site, ou de uma série televisiva, maior será o rendimento obtido em publicidade com o respetivo conteúdo. Na perspetiva de uma empresa, que queira publicitar o seu serviço ou produto, é não só importante que a sua mensagem chegue a muita gente, mas também que chegue às pessoas corretas. O conceito de marketing direcionado explica isso mesmo e permite que uma mensagem publicitária tenha mais impacto no seu público-alvo, traduzindo-se assim em maiores receitas para a empresa. Com base no rico conhecimento adquirido das audiências, nomeadamente através dos atributos específicos que as caracterizam podem-se desenvolver cada vez mais e melhores campanhas de marketing de base de dados. O marketing de base de dados é um processo de marketing orientado à informação tornado possível pela tecnologia de base de dados que permite desenvolver, testar, implementar, medir e modificar programas e estratégias de marketing personalizados (Belfo, 2019).

Com todas estas possibilidades, valências e importância das audiências, o objetivo principal deste projeto é obter modelos que expliquem e consigam prever as audiências das transmissões de jogos de futebol. Os componentes da estrutura do futebol em Portugal

estão interrelacionados entre si, podendo-se assumir, por exemplo, que tanto o futebol amador como o profissional possuem fortes relações de dependência entre si.

As audiências televisivas possuem uma importância transversal a todo o desporto, pois é uma das fontes que lhe garante maior financiamento. Um dos indicadores mais imediatos da importância da televisão para os clubes de futebol é aquilo que aconteceu desde a chegada da pandemia da Covid-19 e o impacto que a mesma gerou na indústria do desporto. Numa fase inicial, o futebol, tal como muitas atividades económicas, parou. No entanto, a partir do momento em que se conseguiram agilizar formas do “espetáculo” continuar, esta indústria foi das que conseguiu manter a sua atividade central em funcionamento, apesar de todos os constrangimentos e dificuldades impostas pela pandemia. Isso ficou também patente na realização do recente Campeonato Europeu de Futebol de 2020, comumente referido como UEFA Euro 2020 ou simplesmente Euro 2020 e que ocorreu no verão de 2021, mantendo-se, contudo, o nome UEFA Euro 2020. Os jogos desta competição decorreram em 11 diferentes países, uma particularidade inédita, e todos eles com a presença de público. A final, realizada no estádio de Wembley, em Londres, decorreu mesmo com o impressionante número, em ambiente de pandemia, de 67.173 espectadores, num estádio com capacidade total para 90.000 espectadores.

Esta necessidade urgente do regresso do futebol e posterior pressão para que não voltasse a sofrer nova paragem resulta de necessidade dos clubes e federações continuarem a obter financiamento para sobreviverem, financiamento esse que encontra nos direitos televisivos uma fonte indispensável, que lhe permite cobrir as perdas resultantes de outras origens que deixaram de contar, ou foram gravemente afetadas, como a bilhética ou venda de *merchandising*.

Em resumo, este projeto pretende entender melhor o que influencia a audiência televisiva de um jogo de futebol, num desporto que é tão dependente desse simples indicador, permitindo assim a que os todos os agentes da indústria tenham mais informação útil para tratar e maximizar os resultados do “seu” desporto.

## **2 REVISÃO DA LITERATURA**

Esta etapa do projeto pretende fazer uma revisão dos grandes temas que suportam e serão abordados ao longo do projeto.

Como as audiências em estudo dizem respeito a jogos transmitidos pela televisão, numa fase inicial será dado um contexto relacionado com o mercado televisivo, tanto a nível global, como em Portugal, desde o seu aparecimento, passando pela massificação e chegando até aos dias que correm, onde surgem novas formas de consumo como o *streaming*.

Outro dos grandes temas revistos são as audiências, já que estas são uma das grandes âncoras da publicidade televisiva e a principal métrica em estudo neste trabalho. Em relação às audiências serão apresentados alguns dos possíveis fatores que as influenciam, tanto no contexto desportivo como da restante programação televisiva e também alguns trabalhos de previsão aplicados neste âmbito.

A revisão de literatura aborda ainda o negócio das transmissões televisivas, num resumo dos principais acontecimentos aplicado em dois mercados: o inglês, que serviu de exemplo para todo o futebol europeu e o mercado português, que vai ser trabalhado neste projeto.

### **2.1 Mercado Televisivo e de Comunicação**

#### **2.1.1 Evolução histórica global da televisão**

O início da idade da televisão, na década de 1950, deu origem à comunicação visual e estimulou a emergência de teorias interdisciplinares dos *media* (Briggs & Burke, 2009). O surgimento da televisão veio revolucionar todo o paradigma da comunicação e da vida em sociedade. É essencialmente através dela que a maioria das pessoas se informa, entretém, forma opiniões, aprende e conhece lugares que nem sequer sabiam existir. São diversas as utilidades que uma televisão tem nos dias que correm, e apesar de recorrentemente a sua morte ser anunciada, a grande maioria dos lares continua a ter uma, seja ela grande, pequena, velha ou de última geração.

A televisão como eletrodoméstico de entretenimento continua a estar presente desde o seu período de massificação na grande maioria das casas, sendo as suas valências cada vez mais diversas, o que lhe tem proporcionado um longo período de vida. Uma dessas novas valências surge com o aparecimento da Internet, que veio modificar o consumo

televisivo, tornando-o mais diversificado e possibilitando o aparecimento de novos formatos. Estas evoluções provocam constantes alterações no comportamento do telespetador, que atualmente pode estar a consumir conteúdos televisivos sem estar perante uma televisão, podendo para isso utilizar equipamentos como o *smartphone*, ou o computador.

São cada vez mais os formatos televisivos que existem e com a panóplia de equipamentos no mercado, os telespetadores têm tendência para modificar o seu comportamento consoante o que estão a ver. Há os programas ditos familiares que, como o nome indica, são feitos a pensar num consumo em grupo e seio familiar. Por outro lado, existem outros programas que tanto são consumidos em família ou com os amigos, como também individualmente, não tendo uma linha tão definida. Estas diversas maneiras de visualizar um programa influenciam as audiências de determinado conteúdo em cada plataforma. A título de exemplo um canal pode disponibilizar os mesmos conteúdos na emissão normal de televisão e no seu site, e a audiência contextualizada ser bastante discrepante. É com base nessa ideia que têm surgido tipologias como as webséries, que são séries concebidas para serem vistas exclusivamente através da Internet e por norma têm mais consumo em dispositivos como computadores e *smartphones*.

### **2.1.2 Evolução história da televisão em Portugal**

Segundo Lopes & Pereira (2007), a televisão surge na Europa, como “a voz do Estado”, não tendo Portugal sido exceção a esta “regra”. Promovida por Marcelo Caetano nasce, em 1957, a Rádio e Televisão Portuguesa (RTP), que desde essa altura “foi sendo gerida pelo Estado” (Fernandes, 2013). Esse controlo não era absoluto numa fase inicial, mas “a partir de 1968, ano em que nasceu o segundo canal, o Estado assumiu a maioria do capital, ficando na posse de sessenta por cento das ações” (Torres, 2011).

O paradigma em Portugal, no que a televisões diz respeito, alterou-se radicalmente em 1992, mais concretamente a 6 de outubro, quando “pelas mãos de Francisco Pinto Balsemão nasce a Sociedade Independente de Comunicação (SIC)” (Fernandes, 2013). A SIC foi o primeiro canal privado a aparecer em Portugal, mas não foi o único por muito tempo, pois “a 20 de fevereiro de 1993 nasce a TVI (Televisão Independente), um projeto ligado à igreja e dirigido por Roberto Carneiro” (Fernandes, 2013). A entrada em cena destas duas estações fez com que o mercado quisesse saber mais sobre as audiências

televisivas e estas passassem a ser um termo de comparação e competição entre os diversos canais.

Desde essa altura até aos dias de hoje têm surgido novos métodos de ver televisão, com forte incidência para as plataformas móveis. O crescimento do consumo de conteúdos em dispositivos móveis, como *smartphones* ou *tablets*, acentuou-se no mercado dos consumidores mais jovens, que não só veem os conteúdos televisivos em diversos dispositivos distintos, como os consomem de uma maneira diferente, quer seja através de gravações, ou *on demand* pela Internet. O conteúdo produzido pelas televisões tem assim de se adaptar aos novos consumidores e, conseqüentemente, às novas formas de consumo. Nesse sentido as principais televisões em Portugal têm apostado nos últimos anos em plataformas para responder a essa procura como a RTP Play, OPTO SIC ou o TVI Player.

### **2.1.3 Diferentes abordagens dos órgãos de comunicação**

Aquando da criação de um meio de comunicação são definidas algumas diretrizes, mais ou menos evidentes, que definem a abordagem desse mesmo meio no contexto social e competitivo em que se insere. Com a evolução e a passagem do tempo as diretrizes acima referidas sofrem mutações de diversas escalas, existindo sempre pontos de concordância entre os diversos meios presentes no mercado.

Desde a sua fase inicial até à revolução de abril de 1974, a televisão em Portugal cumpria a função de “megafone do regime salazarista-marcelista” (Torres, 2011). O papel que a RTP desempenhava nessa altura sofreu alterações, tendo a estação pública sido nacionalizada após ocupação militar. A abordagem programática da RTP sofreu alterações com a nacionalização. Se antes da revolução dominavam os programas de ficção e informação, depois dela surgiram programas de entretenimento, o que veio mudar o paradigma da sociedade de consumo. A mudança que se seguiu, no contexto dos media em Portugal, surge no final dos anos 80. Existia uma “necessidade de se diversificar a oferta devido aos diversos gostos do público” (Galamba, 2014) que, após uma revisão constitucional em 1989, permite o aparecimento dos canais privados. Em Portugal, os canais privados surgiram no início dos anos 90, seguiu-se a televisão por cabo e mais recentemente a internet.

Uma das grandes divisões entre os objetivos dos órgãos de comunicação surge da condição se o mesmo é público ou privado. Uma televisão pública existe para servir a população e fornecer um serviço que não pode ser garantida pelas entidades privadas. Isto

traduz-se, segundo Cádima (2010), na aposta em “conteúdos que sejam claramente distintos das ofertas comerciais”, seguindo alguns princípios como “a universalidade, a diversidade e pluralismo de ideias e conteúdos, a informação substantiva e contextualizada, a promoção da cidadania e de uma democracia participativa e o desenvolvimento da literacia mediática” (Galamba, 2014). Já o objetivo principal das televisões privadas, ou comerciais, passa por maximizar o seu lucro, tendo por isso outro tipo de preocupações aquando da construção das suas grelhas televisivas, procurando ter a maior audiência possível, de forma a gerar mais receitas provenientes de publicidade.

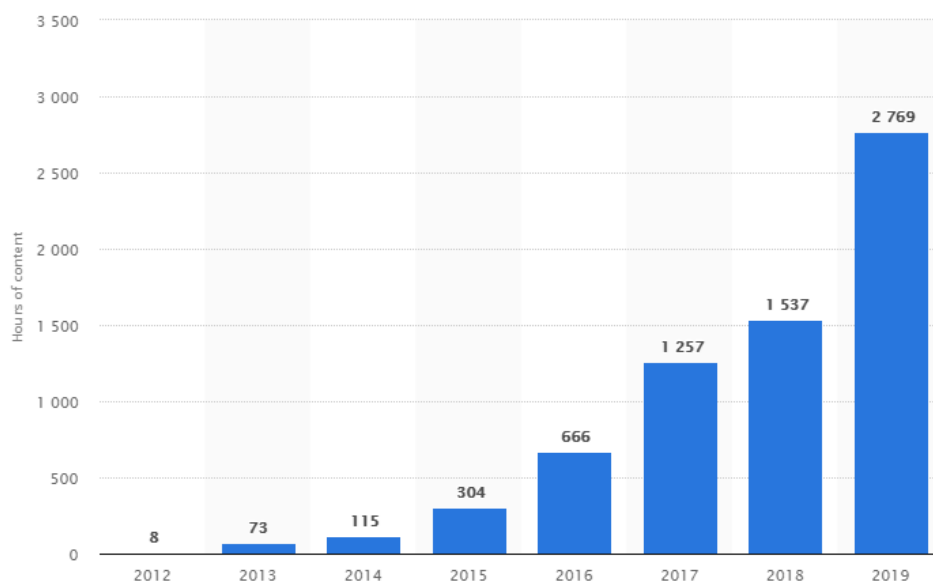
#### **2.1.4 Futuro do mercado da televisão**

O mercado da televisão e vídeo está a atravessar mais uma revolução e é difícil prever a extensão e real impacto da mesma. Às tradicionais emissoras televisivas, sejam elas regionais, nacionais ou internacionais, têm-se juntado nos últimos anos algumas plataformas de *streaming*, como a Netflix, a HBO e a Amazon Prime que mudaram o paradigma do consumo, bem como o valor de mercado dos conteúdos produzidos. Ao consumo tradicional, em que os telespetadores se sentavam à frente da televisão e “absorviam” a programação imposta pelas estações televisivas, juntou-se o consumo *on demand*, onde cada pessoa vê o conteúdo que quer, num dispositivo à sua escolha, no horário que considerar adequado e ao seu próprio ritmo. Este tipo de consumo já está consolidado no mercado de TV e vídeo e coexiste com o tradicional.

Nos dias que correm, os serviços de *streaming* já não funcionam apenas como uma plataforma para reproduzir vídeos e séries, também estão a investir em produzir e licenciar o seu próprio conteúdo (Boehm et al., 2018). À data, a Netflix é a empresa que domina no setor do Vídeo On Demand (VOD), tendo no final de 2019 cerca de 167 milhões assinantes pagantes (Netflix Inc., 2019) e apesar de ser essencialmente um depósito de filmes, séries e documentários pré-existentes, a empresa sediada na Califórnia, Estados Unidos da América (EUA), tem apostado no desenvolvimento de conteúdo original. Foi em 2013 que a Netflix começou a lançar o seu conteúdo original próprio com as séries "House of Cards," "Hemlock Grove," "Arrested Development" e "Orange Is the New Black" (Netflix, 2020) e desde então tem crescido consecutivamente, como demonstra a Figura 1, registando 371 originais no ano de 2019, número que, de acordo com Bridge (2019) na *Variety*, supera toda a produção do setor televisivo nos EUA em 2005. Para além dos valores significativos na quantidade dos originais lançados no

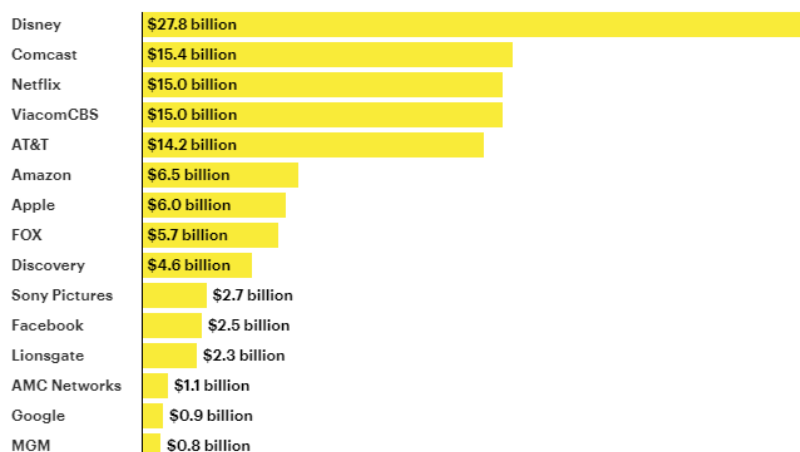
## *Análise e previsão de audiências no mercado televisivo e de comunicação em Portugal*

mercado, a qualidade também se tem evidenciado. Na última edição dos Globos de Ouro, a Netflix teve 34 nomeações, mais do que qualquer outra empresa de entretenimento (Lang, 2019). Em 2019, os gastos estimados da Netflix na produção de conteúdo próprio fixaram-se nos 15,3 mil milhões de dólares, estando previsto um aumento de 2 mil milhões para o ano seguinte.



*Figura 1- Evolução do número de horas de conteúdo original produzido pela Netflix (Statista)*  
*Fonte: Stoll (2020)*

A análise feita pelo BMO Capital Markets estima ainda que a Netflix venha a despende cerca de 26 mil milhões de dólares em conteúdo, no ano de 2026 (Morris, 2020). Este investimento surge pelo facto de o catálogo de originais da Netflix ser consideravelmente menor do que o da sua concorrência (Disney, NBC, HBO), o que não impede essa mesma concorrência de investir quantias avultadas, como mostra a Figura 2.



*Figura 2- Gastos estimados com conteúdo original (Variety)*  
*Fonte: (Spangler, 2020)*

O sucesso adjacente ao aparecimento das plataformas de *streaming* trouxe para esse mercado outros concorrentes, desde canais de televisão tradicionais e são exemplo disso serviços como a RTP Play (RTP), o BBC iPlayer (BBC), ou mesmo os próprios criadores de conteúdo, como a Disney e a Fox.

Uma das grandes funcionalidades potencializadas pelos serviços de *streaming* foram os serviços *on demand*, que “mudaram a procura dos espetadores no consumo de tv e vídeo” (Boehm et al., 2018). Um inquérito da Deloitte mostra que, na Alemanha, um mercado de televisão mais conservador, 44% dos consumidores assiste a conteúdos baseados em plataformas de vídeo *on demand* pelo menos uma vez por semana. Este sucesso do vídeo *on demand* faz com que as pessoas esperem cada vez mais por “conteúdo relevante acessível a qualquer altura, em qualquer lugar e no formato que melhor se ajustar às suas necessidades” (Boehm et al., 2018).

A revolução nas formas de consumo não é algo exclusivo a uma região específica, nem a países com um poder de compra equivalente. A Figura 3 equipara países de realidades diferentes, como o Japão, os EUA, a França e o Brasil e mostra, entre os mercados televisivos representados, como está o seu equilíbrio entre consumo de vídeo em direto nos tradicionais canais de televisão e nas plataformas de *streaming*.

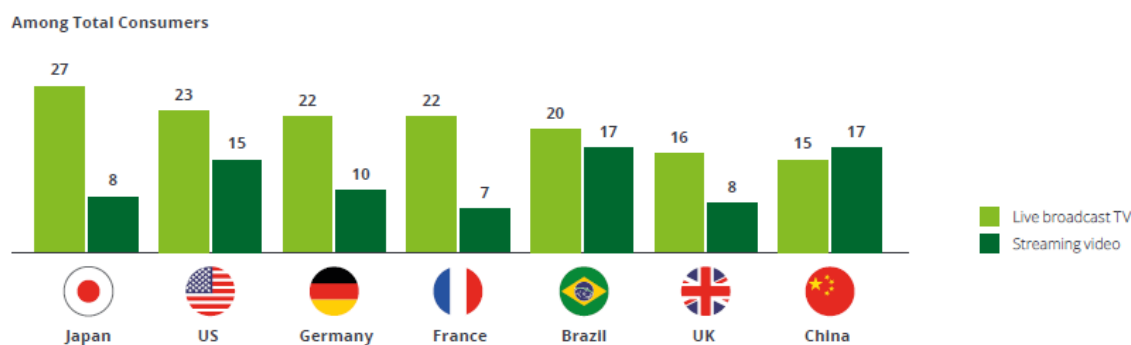


Figura 3 - Consumo médio semanal de vídeo por país (em horas) (Deloitte)  
Fonte: (Boehm et al., 2018)

O Japão é a nação em que a diferença entre as horas médias de consumo de televisão é maior face ao *streaming*. Em contrapartida a China é o único país representado onde já são vistas, semanalmente, mais horas de vídeo em *streaming* do que televisão em direto. Este cenário em constante mudança faz com que o mercado da TV e vídeo seja algo muito difícil de prever, considerando as empresas de diferentes espectros que estão ‘na luta’, mas também fatores de diversa ordem: sociais, tecnológicos, económicos, ambientais e políticos que funcionam como influenciadores do caminho futuro deste mercado.

## **2.2 Possíveis influências nas audiências televisivas**

As audiências podem ser medidas diversas formas, mas as mais comuns são o *rating*, em percentagem e milhares, e o *share*, em percentagem. O *rating* televisivo corresponde à audiência de um determinado programa ou estação num período específico de tempo, podendo ser expresso como uma percentagem da população da audiência, ou em milhares. De acordo com a Comissão de Análise de Estudos de Meios (CAEM, 2021), que é a associação responsável pelas audiências em Portugal um lado, *rating* em milhares, simbolizado por *rat#*, representa o número médio de milhares de indivíduos presentes na audiência do evento/canal. No caso do *rating* em percentagem, simplificado por *rat%*, pode ser calculado de acordo com a Equação 1.

$$rating_e = \frac{\text{número de indivíduos sintonizadas no programa ou estação e}}{\text{número de indivíduos com TV}}$$

*Equação 1 – Fórmula de cálculo do “rating” de um programa ou estação televisiva*

Por outro lado, o *share* televisivo, o qual é a audiência de um determinado programa de televisão ou período de tempo, expresso como uma percentagem da população que assiste aquele programa de TV naquele horário específico face às que assistem a qualquer programa em qualquer estação. O *share* de uma estação pode ser calculado de acordo com a Equação 2.

$$share_e = \frac{\text{número de residências sintonizadas numa estação e}}{\text{número de residências sintonizadas numa qualquer estação}}$$

*Equação 2 – Fórmula de cálculo do “share” de um programa ou estação televisiva*

Por exemplo, se 1.000 casas têm televisão, 500 dessas casas têm as televisões ligadas e 100 casas estão a ver um determinado programa de uma estação, então o *rating* dessa estação é de 10 por cento e o *share* desse programa é de 100 a dividir por 500, ou seja 20 por cento.

As audiências televisivas são a base do mercado publicitário das televisões e podem ser influenciadas por fatores de diversa ordem. A nível de *share*, o resultado obtido por um programa varia sempre de acordo com o resto dos canais, mas aspetos como a sua duração ou o dia e hora em que é transmitido são controláveis pela estação televisiva e têm também o seu impacto na prestação do mesmo. Na Figura 4 estão ordenados os 15 programas mais vistos do dia 28 de junho de 2020, nos canais de televisão portugueses, por *rating* decrescente. Um aspeto que mostra a importância deste tipo de classificações no estatuto

que os canais querem passar para os anunciantes e telespetadores é o das diversas divisões criadas para os mesmos conteúdos, o que lhes permite ter mais programas entre os mais vistos do dia, fortalecendo assim a sua imagem. Por exemplo, ao saberem de antemão que os programas “Isto é Gozar com Quem Trabalha” e “Quem Quer Namorar com o Agricultor?”, na SIC e “Big Brother”, no caso da TVI vão captar uma grande quantidade de público, tendo em conta não só os seus formatos e protagonistas, como também a hora (horário nobre) e dia (domingo), as estações optam por segmentar esses mesmos conteúdos fazendo assim com que os seus programas ocupem mais posições em ranking.

**Ranking de programas mais vistos - 28 de Junho de 2020 (Domingo)**

Canal	Início	Programa	Total dia (Live + VOSDAL)		
			rat%	rat#	sh%
1 SIC	21:58	ISTO E GOZAR COM QUEM TRABALHA - A ENTREVISTA	15,7	1.484	28,7
2 SIC	21:44	ISTO E GOZAR COM QUEM TRABALHA	15,4	1.453	28,7
3 TVI	22:46	BIG BROTHER - A EXPULSAO	13,9	1.314	29,3
4 TVI	21:59	BIG BROTHER - ALTO RISCO	12,8	1.209	23,9
5 SIC	19:58	JORNAL DA NOITE	12,7	1.199	26,1
6 SIC	22:35	QUEM QUER NAMORAR COM O AGRICULTOR? III - DIAS FELIZES	12,6	1.189	24,2
7 SIC	22:15	QUEM QUER NAMORAR COM O AGRICULTOR? III	11,7	1.105	22
8 TVI	23:36	BIG BROTHER - AS NOMEAÇÕES	10,7	1.017	31,1
9 SIC	23:00	QUEM QUER NAMORAR COM O AGRICULTOR? III - O AMOR ACONTECE	10,3	971	21,8
10 SIC	13:00	PRIMEIRO JORNAL	9,9	936	25,7
11 SIC	23:23	QUEM QUER NAMORAR COM O AGRICULTOR? III - CONVERSA EM DIA	9,5	895	21,1
12 TVI	19:57	JORNAL DAS 8	9,1	866	18,8
13 SIC	23:51	QUEM QUER NAMORAR COM O AGRICULTOR? III - INTIMIDADES	8,4	794	22,4
14 SIC	14:11	FAMA SHOW	8,3	784	22
15 SIC	12:10	VIDA SELVAGEM	7,6	718	22

*Figura 4 - Programas mais vistos de 28 de junho de 2020*

*Fonte: (CAEM/GFK, 2021)*

Se estas duas estações, SIC e TVI, não optassem por esta estratégia, por exemplo se o programa “Quem Quer Namorar com o Agricultor” mantivesse o mesmo nome nas grelhas durante toda a emissão, só iria ocupar um dos lugares no top-15, estando assim vulnerável a ocupar uma posição inferior no ranking e com uma presença menos notada, pelo facto se só aparecer uma ocasião em vez das 5 que surgem no exemplo da Figura 4.

Esta é uma das estratégias utilizadas pelas estações para se evidenciarem no contexto de um ranking. Também, para fazerem parte desses mesmos rankings necessitam de sobressaírem na audiência que acumulam, tendo em consideração fatores de diversa ordem, como os que em seguida se irão apresentar.

### **2.2.1 Duração**

Um programa tem uma duração estipulada, que é influenciada por muitos aspetos. Uma das razões que impõe a necessidade de uma duração aliada a um programa é que o

conjunto de todos os conteúdos programáticos de uma estação forma a grelha desse mesmo canal. No entanto, a duração de um conteúdo varia tendo em conta a sua tipologia, mercado/país em que se insere e as audiências históricas ou esperadas do mesmo. Existem certos programas com uma duração imposta mais rígida que, seja pelo seu sucesso ou formato em causa, é complicado alterar, mas para os restantes programas importa perceber se as audiências dos mesmos se alteram consoante o tempo que os mesmos ocupam na grelha. Esta diferença pode exemplificar-se utilizando dois desportos: o futebol e o ciclismo. Um jogo de futebol tradicional é constituído por duas partes de 45 minutos e é transmitido na íntegra, por vezes com algum tempo de antevisão e de rescaldo associado. Já uma etapa de ciclismo é de uma duração mais irregular, dependendo da prova em que se insere, o tipo de etapa ou a distância do percurso. Esta variedade na distância obriga a equipa que gere a programação televisiva da mesma a decidir que percentagem do percurso deve ou não transmitir, de forma a maximizar os seus objetivos. Dependendo do público que assiste a este desporto é importante perceber se o canal beneficia com a transmissão integral de todas as etapas, ou é mais produtivo segmentar a emissão, escolhendo apenas os momentos com mais impacto televisivo.

A duração de um mesmo programa também pode variar consoante o país. Por exemplo, no caso do “Hell’s Kitchen”, a versão norte-americana do programa tem cerca de 40 minutos, ao passo que em Portugal, a adaptação do formato surge com emissões que chegam perto das 2 horas. Esta discrepância tem em conta o tipo de espetador e os hábitos de consumo do mesmo, podendo fazer com que o programa tenha sucesso em dois países diferentes, com uma duração completamente díspar.

### **2.2.2 Dia da semana**

De uma maneira global, o consumo altera consoante o dia da semana ou estação/época do ano. Um paralelismo que se pode encontrar em relação a este aspeto é a diferente adesão que um centro comercial tem durante a semana face ao fim de semana ou no mês de janeiro em comparação com o de dezembro. As razões que levam a que existam grandes disparidades neste tipo de consumo são várias, como por exemplo o tempo livre da maioria das pessoas ou o facto de se viver ou não uma época festiva e consumista como o Natal.

Em Portugal, o “Preço Certo” é um programa com uma longevidade assinalável e sinónimo de sucesso de audiências, principalmente entre os espetadores mais velhos.

Apesar deste sucesso, os *ratings* a que este programa chega apresentam uma certa sazonalidade consoante se trata do horário de verão ou de inverno, obtendo melhores registos no segundo caso. Esta particularidade deve-se ao facto de o mesmo ter um horário fixo das 19h às 20h, período em que no horário de inverno muitos dos telespetadores assíduos do programa já se encontram em casa, mas no horário de verão os mesmos telespetadores terem uma rotina diferente e não estarem disponíveis para assistir televisão àquela hora.

No caso das audiências televisivas essa alteração nos padrões de consumo é notória e apresenta características diferentes de uma estação televisiva para outra. Analisando as audiências médias semestrais, através da Figura 5, para a RTP1, SIC e TVI no ano de 2019 é possível descortinar algumas tendências. Tal como é possível observar na Figura 5, no primeiro semestre desse ano, a RTP1 teve a 4ª feira e o sábado como os seus dias com maior audiência média, sendo o domingo o seu dia menos expressivo. Nos dias com maior tráfego na emissão da RTP1 destacaram-se as transmissões desportivas, de competições como a Liga das Nações e a Taça de Portugal. Já no domingo a procura associada ao canal público é mais diminuta devido à oferta de grandes formatos por parte dos canais da concorrência (“Quem Quer Namorar com o Agricultor?” e “Dança com as Estrelas”). O domingo é o dia em que as estações comerciais (SIC e TVI) apostam mais e conseguem chegar a um número mais elevado de pessoas, tanto no primeiro como no segundo semestre. A oferta programática destes dois canais é bastante semelhante e pelo que indicam os resultados é do agrado do grande público. Outro pormenor interessante da análise feita às audiências médias por dia da semana surge na média da SIC às quintas-feiras. Este dia é o segundo mais relevante para o canal, o que pode ser explicado pelo facto de a estação de Paço de Arcos transmitir nesse dia jogos da Liga Europa.

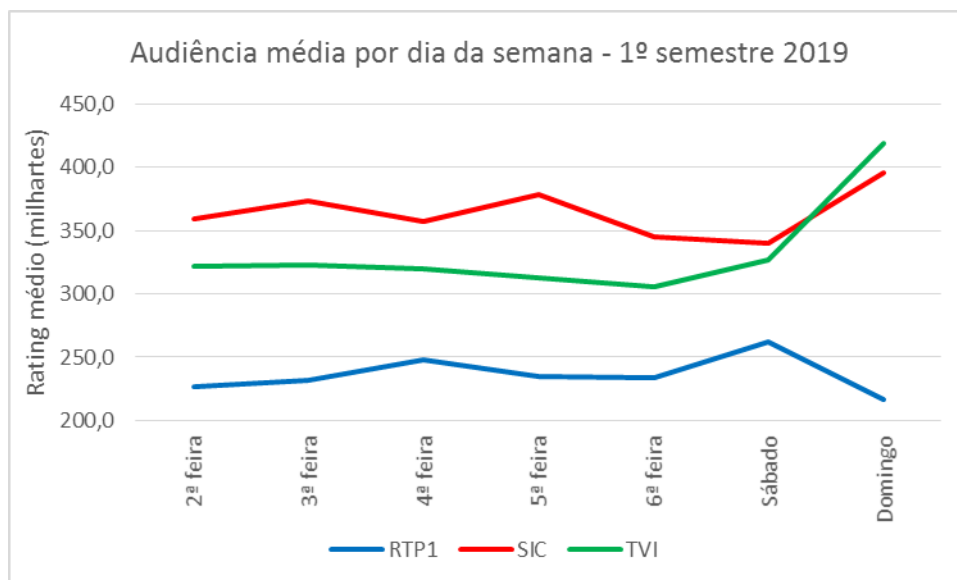


Figura 5 - Audiência média dos principais canais de televisão portugueses no 1º semestre de 2019 por dia da semana  
Fonte: (CAEM/GFK, 2021)

De uma forma global, da análise às audiências dos canais por dia da semana tiram-se duas conclusões. Por um lado, as transmissões desportivas influenciam positivamente as audiências de um canal, mesmo que não sejam de uma frequência regular. Por outro lado, o fim-de-semana é o período em que os portugueses consomem mais televisão, o que se traduz numa aposta reforçada e diferenciada por parte de todos os canais nesses dias.

### 2.2.3 Hora da emissão

As horas da emissão de um programa são tidas em conta para que seja maximizado o número de telespetadores disponíveis e interessados a consumir esse conteúdo. A faixa da manhã é normalmente destinada a exibir programas para públicos infantis ou seniores, ou seja, pessoas fora da idade ativa e que de uma forma global têm mais disponibilidade para ver televisão a essas horas. Faz parte do papel de uma estação de televisão adaptar a sua oferta à grande maioria de pessoas que assiste aos seus programas.

A Figura 6 representa o consumo televisivo médio em Portugal ao longo do dia, no mês de abril de 2021. A interpretação desta informação é fulcral no planeamento de uma grelha programática, pois permite identificar quais os períodos de maior consumo. Uma das aplicações mais imediatas que daqui pode surgir está relacionada com as grandes produções, ou conteúdos com maior preço de aquisição. Quando uma televisão faz um elevado investimento económico necessita de calcular com detalhe em que horário o vai “colocar no ar”, para que o consiga rentabilizar da melhor forma. Em contrapartida os

períodos como a madrugada, ou seja, de menor consumo, ficam destinadas a programas de arquivo, ou repetições, que são conteúdos com um preço associado já esbatido.

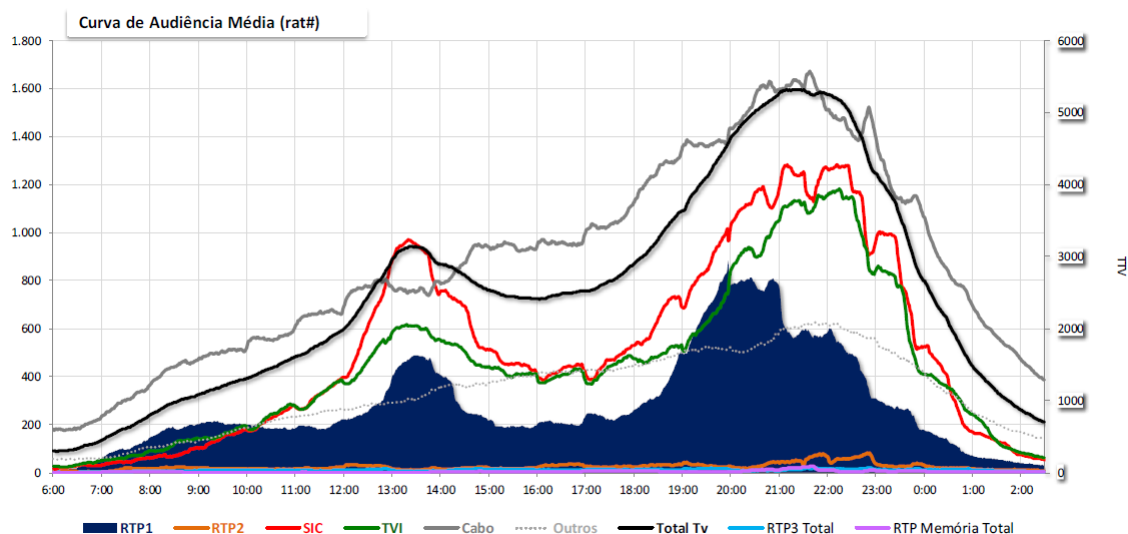


Figura 6 - Consumo televisivo ao longo do dia (Abril de 2021)

Fonte: (CAEM/GFK, 2021)

Da análise ao consumo neste mês é possível identificar um sentido crescente de consumo desde as 6h00 até ao final da hora de almoço tradicional, perto das 14h00. O primeiro pico de consumo do dia verifica-se entre as 13h00 e as 14h00, sendo normalmente explorado pelas televisões generalistas para transmitir noticiários. Depois desse período, associado à pausa para almoço, regista-se a primeira quebra de consumo do dia televisivo que dura até meio da tarde. Com o final do dia de trabalho convencional, o número de pessoas a ver televisão volta a subir e atinge, entre as 18h00 e as 19h00, o mesmo valor do pico da hora de almoço. A curva de consumo nesta altura do dia já segue novamente em ritmo ascendente e atinge o seu auge na faixa 21h00-22h00, período pertencente ao designado *prime-time* televisivo. Seguindo a lógica de programar os conteúdos mais apelativos para que agradem a uma maior quantidade de público é nestas horas de maior consumo que as televisões vão fazer um investimento mais forte.

Apesar de esta curva de consumo ser apenas referente a um dos meses do ano é exemplificativa do que é o consumo televisivo em Portugal de uma forma genérica. Obviamente não é uma estrutura estanque e oscila consoante outros fatores, como as idades dos telespetadores, os dias da semana, ou o horário em vigor, mas oferece uma visão global do consumo e explica alguns comportamentos das televisões.

No desporto, uma pesquisa dedicada a estudar as influências no consumo televisivo de jogos da NFL, Tainsky (2010) confirmou que o facto de um jogo ser disputado no horário nobre tinha influência positiva no número de espetadores alcançados.

#### **2.2.4 Temporada**

Existem tipologias de programas que são segmentados por temporadas, existindo alterações nas audiências entre as diferentes edições. Essas alterações dependem do desempenho que as mesmas geraram na(s) temporada(s) anterior(es) e a expectativa, ou interesse que a atual desperta nos telespetadores.

Um exemplo disso mesmo é o “The Voice Portugal”, que é um formato internacional adaptado pela RTP1 e que conta já com 7 temporadas. A estreia deste *talent show* em Portugal foi no ano de 2011 e a edição mais recente terminou em janeiro de 2020, existindo normalmente uma temporada/edição por ano. Nas primeiras duas temporadas a quota de mercado (share) deste programa rondou os 20%, 19,8% e 20,5% mais concretamente. Esse valor subiu para 24,1% e 25% nas duas edições seguintes, que foram até à data o auge deste programa em termos de audiências televisivas em Portugal. A partir da 5ª edição, o programa registou a primeira quebra de audiências da sua história (21,3% de share), tendência que continuou na 6ª temporada (17,7% share). A temporada mais recente, a 7ª, sofreu uma mudança na equipa de jurados e conseguiu manter a quota de mercado da anterior. Estas variações nas audiências do “The Voice Portugal” são obviamente fruto de outros fatores para além da antiguidade do programa, tais como a concorrência existente, o talento e carisma dos concorrentes de cada edição, a equipa de jurados e apresentadores, a estratégia de redes sociais, entre muitos outros, mas a “idade” de um programa também mostra ter o seu impacto, pois nem todos os conteúdos têm a mesma vida útil.

Realizando um paralelo com as transmissões do futebol, cada temporada televisiva pode corresponder a uma época. Nesse caso, também se podem esperar alterações na audiência global de época para época, fruto de aspetos como a qualidade dos intervenientes, o equilíbrio competitivo do campeonato, ou a promoção de cada um dos jogos. Com o passar dos anos, os adeptos de futebol têm acesso a competições estrangeiras com maior facilidade e criam laços com clubes e jogadores internacionais através das redes sociais, sendo por isso expectável que exista uma fatia de público potencial a “fugir” da Liga

Portuguesa, que é considerada mais periférica face a ligas como a inglesa, espanhola e alemã, por exemplo.

Tal como o público tem acesso a mais programas de entretenimento, quer seja através da internet ou da televisão por cabo, a quantidade de futebol disponível para assistir também tem crescido e, assim sendo, a evolução da temporada ou época, no caso do desporto, pode provocar alterações nas audiências e deve ser um aspeto a ter em conta pela indústria da televisão.

### **2.2.5 Número do episódio**

Existindo variações de temporada para temporada, é legítimo que ao longo da mesma o interesse nos episódios de um programa se altere, oscilando, por exemplo, de acordo com o seu interesse mediático no caso de um noticiário, da fase em que se encontra quando se trata de um *talent show*, se é um episódio especial numa novela ou consoante o convidado num programa de entrevistas.

Analisando o caso prático da 1ª temporada na novela Nazaré (SIC) descobre-se um padrão nos resultados por episódio durante os 10 meses de emissão. Em termos de *rating*, os resultados mais expressivos concentram-se na reta final da novela, mais concretamente nos últimos 4 meses. A média diária de telespetadores alcançados é também mais elevada nos últimos 30 episódios (1.480 de *rating*#) do que nos 30 iniciais (1.282 de *rating*#), factos que atestam a validade elevada que estes formatos conservam junto do público.

Se no subcapítulo anterior a temporada de um programa foi comparada com a época desportiva, no caso dos episódios a alusão é com a fase da competição em causa. De uma prova como um Campeonato da Europa é de esperar que um jogo das meias-finais, ou final tenha mais espetadores do que um da fase de grupos, pois são teoricamente mais decisivos e mediáticos. No entanto não é algo linear e outros fatores, como as seleções que vão a jogo podem modificar a audiência esperada. Nesse tipo de competições em específico, o interesse mediático associado à prova é muito dependente se Portugal está em prova, ou já foi eliminado, aspeto que vai alterar a promoção dos jogos por parte do canal que os emite o que, conseqüentemente, terá reflexo nas audiências.

### **2.2.6 Periodicidade**

A periodicidade da transmissão de um conteúdo também pode impactar o seu desempenho ao nível das audiências. São várias as estratégias que se podem utilizar aquando da

disponibilização dos episódios de uma série ao público. A Netflix generalizou a ação de lançar temporadas inteiras de uma só vez, mas nem em todas as plataformas o sucesso desta medida se comprova. O lançamento de todos os episódios de uma série num só dia faz com que a mesma reúna uma atenção mediática nesse mesmo dia, mas por outro lado torna-a mais suscetível a ser esquecida nas semanas seguintes.

O RTP Lab é uma iniciativa da RTP que pretende financiar projetos para integrarem o seu catálogo de conteúdos digitais (RTP, 2020). Dos sete projetos vencedores exibidos em 2019, existiram algumas diferenças na distribuição dos mesmos, pois alguns foram para o Youtube e para a RTP Play, o *streaming* da RTP, outros foram só para uma destas plataformas. A nível da periodicidade existiram também diferenças. Dois dos sete conteúdos tiveram os seus episódios lançados semanalmente, num dia fixo e os restantes foram disponibilizados na íntegra no dia da estreia. Da análise feita, as 2 séries com periodicidade semanal (On C@ll e Frágil) tiveram os melhores resultados, face às restantes em termos de visualizações no Youtube. Já na RTP Play o cenário é semelhante, com Frágil e On C@ll a ocuparem respetivamente o 1º e 3º lugar entre os conteúdos com maior procura. Este exemplo demonstra que é importante refletir sobre a estratégia de disponibilização dos episódios, pois é algo que pode ter influência no número de espetadores alcançados.

### **2.2.7 Classe social dos telespetadores**

A classe social dos telespetadores é uma das particularidades que tem impacto nas audiências de um programa, em duas vertentes distintas. Apesar de existirem conteúdos considerados consensuais, ou transversais, há outros que estão diretamente ligados a determinados estratos sociais. Numa grelha de programas pode existir oferta que é destinada a classes sociais mais desfavorecidas, mas também para outras com mais tranquilidade económica.

Ao objetivo de uma estação de televisão querer alcançar o máximo de telespetadores possíveis, poderá acrescentar-se o fator da classe social desses telespetadores. Esta preocupação na estratégia das emissoras surge com as eventuais exigências dos anunciantes, os quais pretendem ver as suas ações de *marketing* expostas a um público com maior poder de compra possível, para que possam maximizar os lucros decorrentes das campanhas publicitárias televisivas.

Num artigo que estuda a procura relacionada com as transmissões televisivas da National Football League (NFL) (Tainsky, 2010), o autor propõe dois modelos que incluem uma variável de ordem económica, relativa ao rendimento médio dos habitantes da área metropolitana em que a equipa participante se insere. A variável que representava o rendimento da população foi considerada significativa e com um coeficiente negativo para ambos os modelos. Num estudo anterior a este concluiu-se que o rendimento tinha um impacto positivo no consumo de desporto ao vivo (Baimbridge et al., 1996). Devido a isso, Tainsky (2010) sugere que esse impacto distinto do rendimento económico da população se deve a que “os custos associados à audiência na televisão são limitados quando comparado a assistir a um jogo presencialmente”. Essa justificação da insignificância do custo de assistir a um jogo pela televisão sofre alterações, consoante o conforto financeiros que os telespetadores possuam, mas também com a situação atual do mercado televisivo e dos direitos de transmissão. Em 2021, um adepto de um clube português pode ter que ser assinante de 3 canais *premium* o que significa uma despesa de cerca de 50 euros acrescido ao preço do pacote contratado com a operadora, caso pretenda assistir a todos os jogos do seu clube, visto que a oferta em sinal aberto é bastante reduzida e o mercado está a elevar os seus níveis de fragmentação, com a recente chegada de novos canais.

### **2.2.8 Partidas em simultâneo**

A calendarização de uma competição é efetuada tendo em conta uma grande quantidade de partes interessadas, que vão desde o adepto que se desloca ao estádio, ao que assiste na televisão, passando pelos clubes participantes, patrocinadores e operadores televisivos. Se do ponto de vista dos adeptos a marcação dos jogos é algo mais complexo, porque as preferências variam individualmente, consoante a vida privada e condições de cada um, no caso dos patrocinadores e operadores televisivos a ideia passa por ter a maior exposição possível, ou seja, têm preferência por jogos calendarizados a horas que potenciem as audiências televisivas. A calendarização depende da possibilidade dos jogos serem disputados em simultâneo, fragmentando conseqüentemente a audiência de cada um dos jogos ou aumentando o interesse gerado à volta de uma competição.

No caso do principal campeonato alemão de futebol, a Bundesliga, os horários em que os jogos são disputados são estanques, salvo raras exceções que sucedam. O modelo tradicional nesta competição consiste em 6 jogos disputados ao sábado (cinco às 15:30 e

outro às 18:30, na hora local), 2 disputados ao domingo (um às 15:30 e o outro às 17:30) e um noutro dia da semana. Numa pesquisa de Schreyer & Schmidt (2016), foi analisada a procura televisiva para quase 1500 jogos do futebol alemão e uma das variáveis em estudo definia quantos jogos da Bundesliga estavam a ser disputados simultaneamente. Em relação à calendarização chegou-se à conclusão de que “o sábado e o domingo são dias com mais visualizações, face às sextas-feiras”, mas no que respeita ao número de jogos em simultâneo não foram encontradas evidências de que essa variável afetasse as audiências de um jogo.

A UEFA estipula, no 48º artigo dos seus estatutos, que cada uma das associações pode definir “um período de duas horas e meia no sábado ou domingo, durante o qual qualquer transmissão de futebol pode ser proibida” (UEFA, 2003), aplicadas a jogos internacionais. Esta medida visa proteger as ligas domésticas de cada país, pois evita uma maior competição televisiva para os jogos nacionais, demonstrando assim que um maior número de jogos em paralelo pode afetar o interesse gerado à volta de um jogo. Outra das imposições da UEFA é a proibição da marcação de jogos das principais ligas de nacionais de cada país em horários coincidentes com os dos seus jogos na Liga dos Campeões e Liga Europa, que corrobora a medida acima descrita, mas desta vez em benefício das competições internacionais da UEFA.

### **2.2.9 Outras influências nas emissões desportivas**

Uma transmissão desportiva tem características muito diferenciadoras em relação a todos os outros tipos de programas. Enquanto a maioria dos programas é gravado previamente, as emissões de desporto vivem muito do direto. Na recente paragem globalizada do desporto, provocada pela pandemia de Covid-19, os canais desse segmento ocuparam a sua grelha com inúmeros jogos históricos que tinham em arquivo, de forma a alimentar a procura dos telespetadores de conteúdo desportivo. No entanto, quando as competições começaram a regressar, os jogos históricos desapareceram rapidamente da programação dos canais de desporto. Esta particularidade mostra que, embora um acontecimento desportivo antigo acalente algum interesse dos consumidores, a atualidade continua a reinar.

Um desporto como o ciclismo profissional depende quase exclusivamente de patrocinadores e é por isso que os corredores têm tanta publicidade nos seus equipamentos. Os patrocinadores no mercado pretendem, ao apoiar as equipas, a maior

exposição possível, no local onde a prova decorre, mas essencialmente nas transmissões televisivas, onde é possível alcançar um maior número de espetadores. Foi seguindo esta lógica de raciocínio que Rodríguez, Pérez, Puente, & Rodríguez (2012) tentaram perceber quais os fatores “determinantes para as audiências televisivas no ciclismo profissional”. Da pesquisa inicial feita pela equipa de analistas espanhóis, chegou-se à conclusão de que, de uma forma genérica, a competitividade e as especificidades das etapas, bem como as principais características dos ciclistas, como a nacionalidade, têm impacto nas audiências televisivas do ciclismo. Com esta consideração, estes investigadores estabeleceram um modelo de estimação do *rating*, em percentagem, para as provas de ciclismo baseado na “competitividade, tipo de etapa, tipo de volta, nacionalidade do líder, calendarização, audiência de herança, tipo de canal, programação nos outros canais, ano” (Rodríguez et al., 2012).

### **2.3 Previsão aplicada no âmbito de audiências**

As audiências de um conteúdo media, para além de ajudarem a medir o sucesso atual de um programa, oferecem a possibilidade de se estabelecerem expectativas acerca dos resultados futuros. Com uma previsão de audiências eficaz, as partes interessadas conseguirão negociar acordos publicitários de forma mais informada, existindo a possibilidade de fazer evoluir este mercado. Os modelos preditivos podem ainda ser importantes para se perceber o impacto que a alteração de cada variável provoca no total de telespetadores alcançados, servindo assim como ferramenta para melhorar a competitividade do canal televisivo.

#### **2.3.1 Aprendizagem automática na previsão de audiências de televisão**

A atual capacidade computacional disponível, aliada a técnicas e algoritmos de aprendizagem automática ou aprendizado de máquina (*machine learning*, em inglês), permitem processar enormes quantidades de informação disponível em variados conjuntos de dados de diversas áreas. A aprendizagem automática tem sido cada vez mais utilizado nos mais diversos domínios, seja no setor público, como nas áreas fiscal (Seiça et al., 2019), educação (C. Pimenta et al., 2018) e medicina (Brandão et al., 2021) ou, no sector privado, como no marketing (Cui et al., 2006), na indústria de media e entretenimento (Sereday & Cui, 2017a), na indústria de eventos (Loureiro et al., 2014), no turismo (P. Pimenta et al., 2011) e em muitas outras áreas, contribuindo para a criação

de novos conhecimentos e ajudando as organizações a definir estratégias que lhes permitam aumentar seu desempenho.

Nos Estados Unidos da América, as redes de televisão vendem a maioria do seu espaço publicitário no “*upfront*”, que é “um grupo de eventos que acontece anualmente na primavera (...) uma festa de lançamento para introduzir novos programas e (...) um mercado para os anunciantes comprarem tempo comercial na televisão bem antes do previsto” (Sereday & Cui, 2017b). Os diversos agentes envolvidos neste evento procuram garantir o melhor negócio para a sua parte, os diretores das estações de televisão procuram valorizar os seus programas ao máximo e os anunciantes buscam o espaço ideal para o seu anúncio, de acordo com os seus objetivos individuais. A grande base deste negócio milionário, no caso apresentado pelo Nielsen Journal e que pode ser extrapolado para diversas realidades, são as audiências de TV (*ratings*), ou seja a quantidade de pessoas que se espera que assista a um programa.

A importância das audiências de televisão leva a que as empresas invistam cada vez mais nas previsões de resultados, tornando-os mais rápidos e precisos e dando a essas empresas uma maior capacidade de antecipação das decisões que deverão tomar. Estas projeções têm assumido nos últimos anos uma complexidade maior, fruto da alteração nos hábitos de consumo, que tornam o mercado mais competitivo e diversificado. Os espetadores “estão a consumir cada vez mais media em diferentes dispositivos e através de diferentes canais” (Sereday & Cui, 2017b), o que torna o seu comportamento mais imprevisível e variável. Além disto, o consumo fora das horas de transmissão original (*on demand*) também tem aumentado significativamente. Isto são tudo aspetos que prejudicam a precisão das previsões de audiências e valorizam ainda mais as ferramentas e modelos que o conseguem fazer melhor dentro do mercado.

Foi pedido ao Nielsen Journal que fizesse projeções para várias televisões, em mais de 40 segmentos demográficos, abrangendo diferentes tipos de consumos (direto, diferido) e segmentos temporais (hora, dia). Dos anos em análise, 2011 a 2016, foram utilizados dados do 1º trimestre, com a missão de se prever *ratings* do 4º trimestre desses mesmos anos, até ao 3º trimestre do ano seguinte. Os dados de audiências utilizados, como *ratings*, espetadores únicos, audiência média minuto a minuto (AA%), habitações com televisão, foram complementados com outras variáveis passíveis de influenciar o sucesso de um programa, como os gastos com marketing, o desempenho nas redes sociais e outras características como género e hora de emissão.

Na tarefa de exploração dos dados foram encontradas algumas características importantes sobre o consumo televisivo: “o prime-time ainda é de longe o período do dia mais popular” e “espetadores mais velhos têm muito mais probabilidade de assistir TV do que as gerações mais jovens, e as mulheres, geralmente, assistem mais TV que os homens.” (Sereday & Cui, 2017b). Outras das ilações retiradas foram a de que o consumo de televisão é maior aos fins-de-semana e que tem descido nos últimos anos, apesar do crescimento do consumo em diferido.

Dos diversos modelos e algoritmos de aprendizagem automática testados pelos investigadores da Nielsen Journal, como regressão linear, redes neurais, árvores de decisão e *Gradient Boosting Machine* (GBM), foi este último o que revelou oferecer uma melhor combinação entre precisão e escalabilidade para o projeto em causa. O GBM é uma técnica aprendizagem automática que produz um modelo de previsão, baseando-se noutros modelos mais fracos e de menor dimensão, geralmente com recurso às árvores de decisão. Ou seja, o GBM combina um conjunto de árvores de decisão para daí extrair conhecimento que de cada uma delas individualmente não seria possível obter. Os investigadores da Nielsen Journal utilizaram a XGBoost, uma biblioteca de software de código aberto que fornece uma estrutura de aumento de gradiente de regularização disponível para várias linguagens de programação e que tem como características a eficiência, flexibilidade, rapidez, robustez e portabilidade. Esta variante “penaliza modelos excessivamente agressivos”, o que representa um combate ao sobreajustamento (*overfitting*, em inglês), que corresponde a uma situação em que os modelos estão demasiado ajustados aos dados históricos utilizados. No contexto da previsão de audiências de televisão é muito importante evitar o sobreajustamento, pois a televisão está em constante mudança e o que determina o sucesso de um conteúdo no dia de hoje, amanhã pode ter uma importância reduzida, ou mesmo nula.

Na construção do modelo desenvolvido por Sereday & Cui (2017b), os dados foram divididos aleatoriamente em dois conjuntos: o de treino e o de testes, feitos através de validação cruzada. Com esta divisão, o modelo vai aprender a fazer as previsões com base no conjunto de treino, aprendizagem essa que será aplicada nos dados de testes. Depois de o modelo ser treinado, as variáveis a usar, bem como os respetivos pesos, serão definidas de acordo os resultados obtidos nos testes de validação cruzada, atenuando assim o sobreajustamento do modelo ao conjunto de treino. Após o estabelecimento das

variáveis finais, o modelo da Nielsen foi comparado com as projeções feitas pelos seus clientes, num novo conjunto de dados e para diferentes grupos demográficos.

Os resultados obtidos pelo modelo dos investigadores da Nielsen Journal foram satisfatórios, mantendo-se dentro das expectativas. O desempenho do modelo foi mais impressionante, face ao do cliente, nos dados mais granulares, ou seja, ao nível das faixas horárias, em quase todas as televisões. Nos dados agregados de cada televisão os resultados foram menos claros e nalguns casos o modelo do cliente revelou-se mais fiável. Este facto pode ter algumas justificações, como ter sido “capaz de capturar um conhecimento interno único que poderia explicar os efeitos de alto nível que podem influenciar todos os programas” (Sereday & Cui, 2017b) e de esses mesmos modelos necessitarem de menos variáveis de previsão. Em análises posteriores, chegaram à conclusão que um modelo misto, composto pelo modelo dos investigadores Sereday & Cui (2017b) e pelo modelo dos clientes, supera em 90% das vezes cada uma das projeções individualmente, o que leva a crer que se consegue extrair um conhecimento exclusivo da “mistura” dos dois modelos.

As conclusões finais levaram Sereday & Cui (2017b) a considerar o seu modelo robusto para a previsão de *ratings* futuros, utilizando os princípios de aprendizagem automática, especialmente para os dados mais granulares. Como os canais possuem conhecimento interno de relevo, chegou-se à conclusão que um modelo híbrido, combinando o da Nielsen e o dos clientes, pode ser a melhor abordagem para o futuro.

### **2.3.2 Previsão de audiências de televisão através das redes sociais**

Desde o seu surgimento e massificação, as redes sociais são um dos principais meios que os telespetadores utilizam para comentar um programa. É visível, principalmente em grandes formatos como os *reality shows*, o impacto que a programação televisiva tem nos conteúdos mais badalados nas redes sociais. Um dos principais indicadores deste impacto são as tendências no Twitter, que reúnem os termos ou *hashtags* mais utilizadas nas últimas horas e oferecem uma perspetiva, que pode ser local (cidade ou país) ou mundial, de qual o “assunto do momento”. Por norma, as notícias ou eventos de maior impacto estão entre as tendências do Twitter, o que valida esta medição como um bom barómetro de popularidade de certo assunto. Por esse motivo, as televisões trabalham para que os seus programas figurem nestas listas, pois é sinal que está muita gente a ver o programa e pode ainda servir de chamariz para atrair novos espetadores.

Em Portugal, alguns dos programas de maior relevo como o “The Voice Portugal”, “O Programa da Cristina” e o “Big Brother” estão regularmente nas tendências, tanto através de *hashtags* (ex: #thevoicport), como através de termos relacionados (ex: o nome de um convidado que esteja n”O Programa da Cristina” ou de um concorrente do “Big Brother”).

No geral, o desporto também atrai um grande número de comentários nas redes sociais. Ao observarem esta tendência as estações de desporto, como a SportTV e a Eleven Sports, criaram *hashtags* dedicadas a certas competições ou desportos, para agregarem os diversos interessados nessas competições ou desportos, conseguindo assim interagir com os fãs e tornando as emissões mais dinâmicas. Esta opção faz com que os espetadores se sintam parte da transmissão a que estão a assistir e traz também novos assinantes aos seus canais, o que gera mais retorno financeiro. Apesar de o futebol ser o desporto com maior número de interessados em Portugal, é comum ver os termos #NBAnaSPORTTV e #F1Eleven nas tendências, o que demonstra que a estratégia utilizada por estes canais privados dá frutos, mesmo em desportos que em Portugal têm uma base de fãs mais reduzida, como é o caso da National Basketball Association (NBA) e da Fórmula 1 (F1).

Para além do Twitter, o Facebook também é uma rede social bastante utilizada para partilhar opiniões sobre conteúdo televisivo, facto que levou muitos canais a criarem páginas nessa plataforma, não só institucionais, mas também dedicadas a determinados programas. As páginas de fãs, criadas pelas televisões, surgiram com o propósito de “interagir com os espetadores e obter respostas deles, de forma a aumentar os *ratings*” (Hsieh et al., 2013). Numa pesquisa publicada em 2013, foi analisado o conteúdo gerado pelas páginas de fãs de um programa e o histórico de resultados de audiências desse mesmo conteúdo, utilizando redes neurais artificiais, para fazer previsões dos *ratings* de TV. Para além do poder comercial que a previsão dos *ratings* oferece a uma estação, a análise do impacto das páginas de fãs num programa pode levar um canal a decidir se investe mais na gestão dessas páginas ou não. A construção do modelo baseou-se em algoritmos de *backpropagation*, numa 1ª fase para “treinar” a rede neural artificial e posteriormente usando essa mesma rede para “fazer previsões de *ratings* para os programas futuros” (Hsieh et al., 2013).

As redes neurais consistem em modelos computacionais que fazem o processamento de dados à semelhança do que acontece com o cérebro humano. Nestas redes existem diversos nós, aqui também designados por neurónios, com ligações entre si e que têm um

peso associado a uma dessas ligações. É através do treino da rede que é gerado conhecimento, pois em cada iteração do modelo, o peso associado a cada ligação pode ser ajustado, o que permite uma diminuição do erro. O estudo desenvolvido por Hsieh, Seng-cho, Cheng, e Wu (2013) focou-se em 4 séries de televisão, com exibição semanal e conjugou os *ratings* de cada episódio, com as interações que as publicações de cada série têm semanalmente no Facebook. Os dados retirados do Facebook dizem respeito ao número de publicações, gostos, comentários e partilhas, tanto da parte dos fãs como dos gestores da página. Já os dados de *rating* usados são as audiências obtidas pelo 1º episódio da série e o episódio anterior. Tendo em conta a metodologia utilizada todos os dados “foram normalizados antes da fase de treino e teste, para que estejam entre 0 e 1” (Hsieh et al., 2013).

Os resultados obtidos por esta investigação foram significativos na maioria das situações, provando que é possível efetuar previsão de audiências com base em dados provenientes das redes sociais, neste caso o Facebook. As métricas provenientes das páginas de fãs foram divididas em dois períodos temporais: um conjunto com dados referentes à última semana antes do episódio ir para o ar e outro com uma abrangência de 3 semanas anteriores. De uma forma global os dados das publicações apenas uma semana antes de o episódio ser lançado parecem produzir previsões mais acertadas, o que se pode justificar com a volatilidade associada às redes sociais. Um dos casos em que as previsões estiveram mais longe do *rating* real sucederam em episódios que coincidiram com eventos especiais noutros canais (estreias ou episódios finais), o que provocou uma divisão da audiência. Outro dos fatores que prejudicou a previsão da audiência de um episódio foi a tipologia de publicações da página da série nas redes sociais, que apelava à interação dos fãs diretamente e fez com que o modelo sobreavaliasse um episódio. Estas as situações poderão vir a ser corrigidas e prevenidas em futuros modelos. Este e outros modelos idênticos mostraram-se viáveis na sua implementação e geradores de conhecimento muito importante para a tomada de decisão das partes interessadas.

## **2.4 Negócio das transmissões televisivas de jogos de futebol**

O negócio associado ao futebol tem por si só características que o diferenciam em larga escala dos outros negócios. Atualmente, os clubes de futebol profissionais funcionam como empresas que têm como potenciais consumidores os seus sócios, adeptos e simpatizantes. No entanto, os clubes funcionando como empresas logram de alguns

benefícios que uma marca dita normal dificilmente atingirá, como o facto de os seus consumidores serem “mais apaixonados em relação ao seu clube de futebol do que à sua marca preferida” e demonstrarem “um grande nível de lealdade ao seu clube”, o que os leva a assumir “comportamentos irracionais nos seus hábitos de consumo”, adquirindo produtos sem ter em atenção fatores como o “preço e a qualidade” (Bühler et al., 2006). A indústria do futebol, ciente destas características dos seus adeptos/clientes, tem então procurado explorar cada vez mais essa paixão (muitas vezes irracional), que milhões de pessoas têm por este desporto em diversos aspetos do negócio, sendo o das transmissões televisivas um dos mais significativos.

#### **2.4.1 A Premier League**

À semelhança da origem do desporto em si, as primeiras iniciativas que visaram transmitir, ao vivo, um jogo de futebol surgiram em Inglaterra. Esse passo inicial é dado pela televisão inglesa BBC a 16 de setembro de 1937, com um jogo entre o Arsenal e a sua equipa de reservas Figura 1(ver Figura 7), calendarizado com o objetivo de ser televisionado (Silveira, 2010). No ano seguinte registou-se ainda a primeira transmissão de um jogo internacional, com o Inglaterra x Escócia e a final da Taça de Inglaterra, entre o Huddersfield Town e o Preston North End. Estas primeiras emissões eram essencialmente experimentais, visto que a massificação da televisão nas casas das pessoas ainda estava longe de ser uma realidade. Para se ter uma ideia, o encontro acima citado da final Taça de Inglaterra de 1938, entre o Huddersfield e o Preston, o primeiro jogo a ser transmitido na íntegra, teve cerca de 94 mil espetadores no estádio, mas apenas 10 mil lares possuíam uma televisão.



*Figura 7 - Jogadores do Arsenal depois do 1º jogo televisionado (1937)*

Estas iniciativas do futebol inglês foram um prenúncio do que se viria a passar na evolução da indústria do desporto pois, ao nível das transmissões televisivas, o futebol mundial tem seguido a sua evolução consoante o que se faz em Inglaterra. Entre os anos 50 e metade da década de 80, a BBC e a ITV dividiram entre si o duopólio das transmissões desportivas definidas como de interesse nacional, numa fase em que existia um receio dos dirigentes desportivos, que temiam uma quebra nas receitas de bilheteira com a eventual retransmissão dos jogos em grandes lugares públicos como salas de cinema (Scelles et al., 2020).

Antes de 1983 nenhuns direitos televisivos foram pagos para transmitir jogos em direto. Nesse ano surge o primeiro contrato de direitos televisivos, com duração até 1985, estabelecido com a BBC e a ITV e fixado nos 2,6 milhões de libras por ano. Estas duas estações assinaram ainda outro acordo de 1986 até 1988, no valor de 3,1 milhões de libras anuais, até que houve uma rutura e as duas estações televisivas deixaram de cooperar. O duopólio permitia à BBC e à ITV a aquisição de direitos a preços muito mais acessíveis do que caso estivessem a competir uma com a outra. As elevadas audiências levaram a que a federação inglesa pedisse mais dinheiro no ano de 1985, proposta recusada por ambas as estações. Essa parceria conheceu o fim no ano de 1988, quando a ITV quebrou o acordo e adquiriu individualmente os direitos televisivos por uma verba de 11 milhões de libras anuais, 3,5 vezes mais que o anterior acordo (Gratton & Solberg, 2007).

O desporto começou a tomar as rédeas das grelhas televisivas na década de 90, não só através do futebol mas também com o ténis, o basquetebol, o rugby e a F1, o que fez

aumentar o preço de aquisição das respetivas transmissões televisivas, devido à disputa entre as diversas estações (Santamaría & Díaz, 2014). Essa década coincide ainda com criação da Premier League, uma competição que surge da rutura dos principais clubes ingleses com a Football League, de forma a concentrarem em si maior volume de receitas, vindas de fontes como os direitos televisivos e os contratos de patrocínio. No que concerne aos direitos televisivos a proposta mais vantajosa para a liga recém-criada vem da British Sky Broadcasting (BSkyB), que adquire os direitos pela quantia de 38,3 milhões de libras por época e procede a algumas alterações estratégicas, como a alteração do horário dos jogos e o aumento do número de partidas com cobertura televisiva: eram transmitidos 18 jogos por época no acordo anterior e passam a ser 60 (Christos et al., 2018). Esta chegada da BSkyB alterou o paradigma dos contratos televisivos no futebol inglês, que superaram em 250% o valor anual que vigorava até então, o que possibilitou um aumento nos prémios monetários das equipas e conseqüente crescimento das mesmas. Com maior orçamento disponível, os clubes conseguiram atrair melhores jogadores, o que levou a um incremento de qualidade do espetáculo desportivo, beneficiando assim as audiências televisivas e assistências nos estádios (Scelles et al., 2020). Como era expectável, aquando da renegociação do contrato em 1997, a *Premier League* conseguiu um acordo ainda mais favorável para os seus clubes, na ordem dos 167,5 milhões de libras anuais, um crescimento de 337% em relação ao acordo anterior e no contrato de 2001 a 2004 atingiu o valor de 367 milhões de libras por temporada (+119% face ao contrato anterior). Esta sequência de contratos televisivos cada vez mais volumosos sofreu um pequeno revés em 2004, num período em que a BSkyB não enfrentou grande concorrência pela aquisição dos direitos, conseguindo assim celebrar o acordo por 333 milhões de libras por época (-9%) e aumentar o número de jogos transmitidos. O número de jogos passou de 66 passaram para 138, diminuindo assim o preço por jogo em 57%. De forma a acabar com este monopólio da BSkyB a Comissão Europeia interveio e insistiu para que pelo menos um dos pacotes de direitos vendidos pela Premier League fosse atribuído a uma estação diferente. Como resultado dessa intervenção, os direitos para o contrato seguinte, relativos aos anos de 2007 a 2010, foram divididos entre a BSkyB (4 pacotes, 92 jogos por época) e a Setanta (2 pacotes e 46 jogos por ano), num acordo que rendeu na totalidade uma média de 567 milhões de libras por época, mantendo assim a tendência típica de valorização dos direitos que havia sofrido um revés recentemente (Gratton & Solberg, 2007).

Como se mostra na Figura 8, desde 2010 os preços praticados para aquisição dos direitos aumentaram consideravelmente nos últimos acordos até ao mais recente, de 2019 a 2022. Em dezembro de 2017, a Premier League licitou os pacotes para as transmissões do período de 2019 a 2022 com a BT e a Sky, como sucedeu nos dois triénios anteriores.

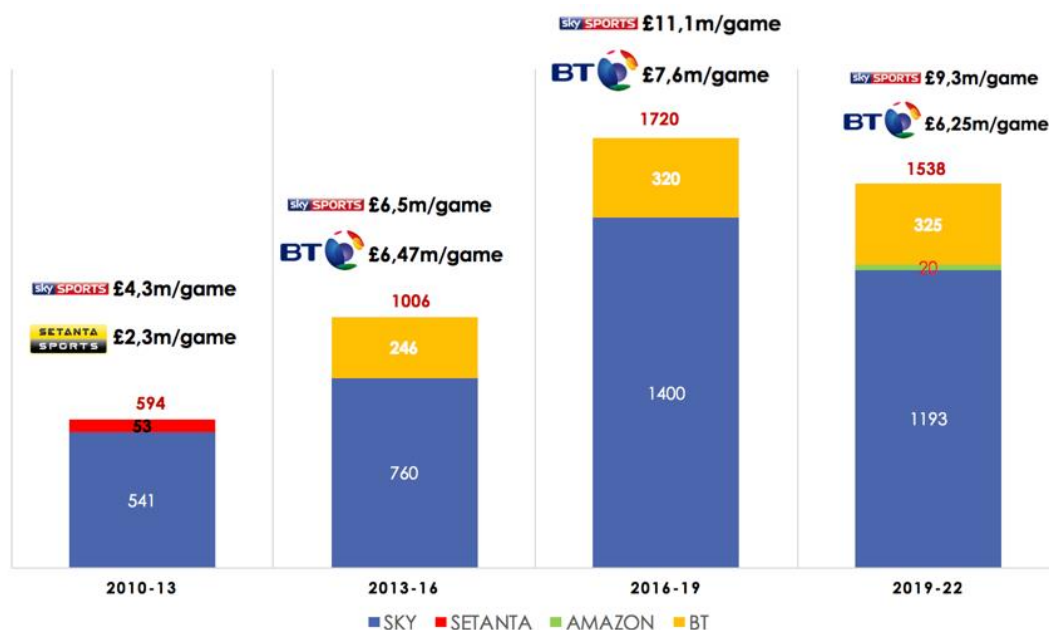


Figura 8 - Custo dos contratos televisivos da Premier League de 2010 a 2022  
Fonte: Gazapo (2018)

Depois dessa licitação, as duas estações acordaram uma parceria para partilha de conteúdo para esse período, em que os canais de cada televisão estavam disponíveis na plataforma de cada uma delas, melhorando assim o pacote que cada uma delas pode oferecer aos seus clientes (Dudley, 2017). Esta quebra nas receitas registada no mais recente acordo, com a entrada de uma gigante tecnológica neste mercado, a Amazon, demonstram que estamos perante um cenário de mudança e que cada vez mais terão de ser consideradas as novas formas de consumo e estratégias de distribuição (Gazapo, 2018).

#### 2.4.2 O futebol português

Em Portugal, o primeiro jogo de futebol a ser emitido na televisão foi um amigável entre o Sporting Clube de Portugal (SCP) e o Áustria Viena, em 1958. O jogo foi disputado no Estádio José de Alvalade, situado próximo dos estúdios da RTP, a única televisão portuguesa na época e a qual foi a estação responsável por essa efeméride (Marques, 2014). Em jogos a contar para o campeonato, a primeira transmissão em direto para a televisão data de 1978, 30 anos após o jogo do Sporting x Áustria Viena, num jogo que opôs o Vitória de Setúbal e o Belenenses (Nogueira, 2018). Nesse jogo o número de

portugueses que conseguiu assistir já foi maior pois, ao contrário de 1958, o número de lares com televisão já era significativo e a “caixinha mágica” já era um aparelho mais massificado.

As transmissões continuaram a partir dessa data. No entanto, a grande alteração relativamente ao mercado das transmissões televisivas do futebol português sucede em 1998 (ver Figura 9), ano em que surge o primeiro canal português dedicado exclusivamente à produção e transmissão de conteúdos desportivos, a SportTV. Nos anos seguintes a SportTV foi conseguindo aumentar a sua base de subscritores, à medida que ia também aumentando o número de horas de conteúdo transmitido. Esta rápida integração no mercado foi facilitada pelo facto do canal transmitir eventos de grande interesse, como os jogos do campeonato nacional (Rodrigues, 2017). Sendo a SportTV um canal *premium* que desde a sua criação agradou aos adeptos de desporto em Portugal era notório que os direitos televisivos podiam ser uma boa fonte de receitas para os clubes, tal como já sucedia noutros países como Inglaterra e França.

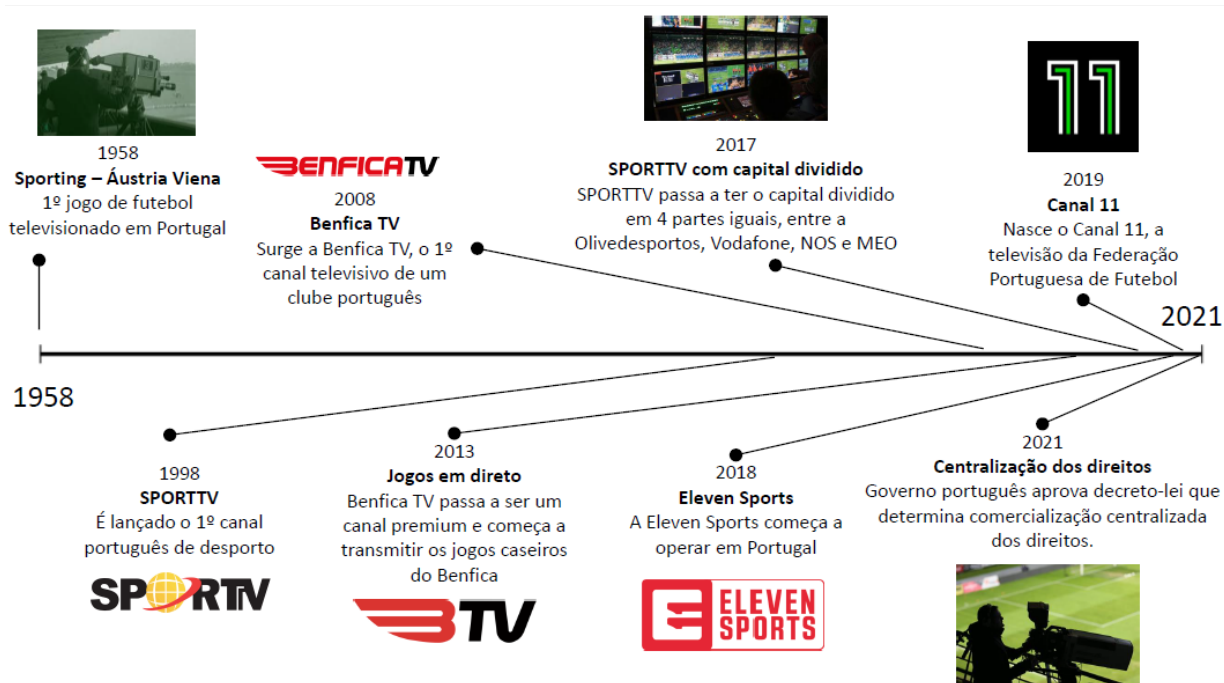


Figura 9 - Principais marcos do mercado das transmissões televisivas do futebol português

Apesar de existirem semelhanças da importância que os direitos televisivos assumem, em Portugal e noutros países, o montante que a eles está associado e a forma como é distribuído é bastante díspar. Estima-se que para a época de 2011/2012 os clubes portugueses tenham gerado cerca de 60 milhões de euros em direitos de transmissão nacionais, um valor bastante mais reduzido ao praticado nas ligas Big-5 (Inglaterra,

Espanha, Itália, Alemanha e França), onde todas as ligas conseguiram acordos acima dos 400 milhões de euros. Mais de metade dos cerca de 60 milhões registados pela liga portuguesa em 2011/2012 estavam concentrados em 3 clubes: Sporting Clube de Portugal (SCP), Futebol Clube do Porto (FCP) e Sport Lisboa e Benfica (SLB), normalmente designados apenas por Sporting, Porto e Benfica. Essa particularidade entronca noutra das diferenças que o futebol jogado em Portugal tem face a outras ligas, que consiste nos clubes negociarem os seus direitos individualmente e não coletivamente, provocando assim maiores desequilíbrios nas receitas auferidas entre os clubes mais e menos competitivos. Rodrigues (2017) chegou a uma correlação de Pearson de 0,79 (fortemente positiva) entre as receitas televisivas geradas por um clube e a sua classificação final da época 2015/2016, o que demonstra bem o impacto que a distribuição de receitas pode ter no desempenho de uma equipa na competição.

Como demonstra o cronograma da Figura 9, o mercado português sofreu um “abanão” em 2013, quando o Benfica decidiu não vender os seus direitos à Controlinveste Media, entidade que cedia os direitos à SportTV, e começou a explorar essa fonte de receitas através do seu canal próprio, a Benfica TV. O clube lisboeta justificou a sua opção por considerar que conseguia maior volume de receitas explorando a transmissão dos seus próprios jogos, num movimento que se revelou certo. Na época de 2013/2014 a estimativa apontava para receitas totais da Primeira Liga na ordem dos 75 milhões de euros, com os “3 grandes” a ficarem com cerca de 72% do bolo total. Em média os restantes 13 clubes receberam 1,6 milhões de euros, um valor 14 vezes inferior ao do Benfica: 22,744 milhões de euros (Almeida, 2015). Numa época mais recente, a época de 2015/2016, a soma dos contratos televisivos dos clubes perfeitou a quantia de 120 milhões de euros, o que representou uma valorização de 100% face ao ano de 2012. Este valor, mais uma vez, é distribuído maioritariamente pelos grandes, que receberam 80 milhões de euros (67%) (Rodrigues, 2017). Os principais rivais do Benfica, o Sporting e o Porto, seguiram o caminho das “águias” e também decidiram avançar no projeto de ter um canal televisivo. No caso do Sporting, estreou o seu canal em 2014 (Fantastic, 2014) e o Porto completou a aquisição do já existente Porto Canal em 2015 (Sapo Desporto, 2015). Apesar de todos os “3 grandes” possuírem um canal televisivo, só o do Benfica tem carácter *premium* e transmite exclusivamente os jogos da equipa principal na Primeira Liga. No caso do Porto e do Sporting as transmissões próprias resumem-se às equipas de

formação e secundárias, bem como de outras modalidades, não tendo por isso influenciado diretamente o mercado dos direitos televisivos na Primeira Liga.

Na época de 2018/2019, as receitas televisivas representavam para os clubes da liga portuguesa 32% da receita média, acima de outras fontes como os contratos de patrocínio e publicidade (24%) e dos prémios da UEFA (20%). O relatório apresentado pela UEFA (2018) retrata ainda que, para esta época, o valor total dos direitos televisivos em Portugal foi de 143 milhões de euros, o 7º mais elevado do velho continente e apenas inferior ao das 5 ligas que superam Portugal no ranking da UEFA, as Big-5, e à liga turca. Em termos de receita média por equipa o valor em Portugal é de 7,9 milhões de euros por equipa, enquanto na liga inglesa, a 1ª do ranking, o valor médio por equipa é de 143,2 milhões de euros, ou seja, cada equipa inglesa da *Premier League* recebe anualmente um valor médio correspondente a direitos televisivos, idêntico a todas as equipas da Primeira Liga portuguesa juntas.

Todas as ligas que apresentam maior volume de receitas que Portugal trabalham num modelo centralizado e em que aspetos relativos à despesa desportiva, como a classificação são tidos em conta na atribuição de receitas, ao contrário do que sucede no nosso país, onde ainda prevalece a negociação individualizada. Como já foi referido, este sistema agrava as desigualdades entre os clubes de pequena e grande dimensão. Em Portugal, o rácio do “dinheiro das televisões” recebido entre o clube que mais ganha e a mediana oscila entre os 10 e 15, sendo a maior discrepância registada no continente europeu, num valor muito acima do rácio médio, que é de 2,3 (UEFA, 2018).

Mesmo existindo alguns aspetos positivos deste mercado em Portugal, como o facto de os clubes terem contratos televisivos assinados até 2028, ou seja, de longa duração, a atual desigualdade na distribuição do bolo total dos direitos é uma ameaça à competitividade e desenvolvimento da liga na próxima década. Países que atualmente estão atrás de Portugal no ranking, como a Bélgica, Holanda, Rússia, Ucrânia e Turquia podem aproveitar esta fragilidade e ultrapassar Portugal no ranking, aumentando a competitividade dos seus clubes, através de fatores como a melhor distribuição de receitas (Gouveia & Pereira, 2021).

De forma a atenuar esta situação, o Governo português já interveio e aprovou um decreto-lei que proíbe as sociedades desportivas de comercializarem os direitos de transmissão a partir da época 2028/2029, “obrigando-as” assim a adotarem um modelo centralizado.

Este modelo centralizado é aquele que vigora nas ligas mais competitivas e com maior saúde financeira e vai ao encontro do recomendado pela Autoridade da Concorrência e trabalhos como o de Almeida (2015) e Rodrigues (2017).

O mercado das transmissões televisivas em Portugal sofreu umas alterações recentemente pois, no início de 2017, a SportTV passou a ser detida em quatro partes de 25% pela Olivedesportos, Vodafone, MEO e NOS, numa operação que, segundo Joaquim Oliveira, o presidente do conselho de administração do canal, "complementa e fortalece" a estratégia de "crescimento e inovação" da estação" (Faria, 2017). Esta alteração trouxe os principais operadores de comunicação para estrutura do principal canal privado de desporto em Portugal e antecedeu a chegada de mais dois novos canais para o segmento do desporto.

A chegada dos dois novos canais televisivos relacionados com desporto em Portugal deu-se nos anos de 2018 e 2019. No primeiro desses dois anos assinalou-se a chegada da Eleven Sports a Portugal, canal que pela primeira vez trouxe competição relevante à SportTV no mercado dos canais de desporto por assinatura (ver Figura 9). A Eleven posicionou-se no mercado com uma mensalidade mais baixa que a concorrência e garantiu os direitos de competições como a Liga dos Campeões, a liga espanhola, alemã e francesa, para além de outros desportos como F1, NFL, que anteriormente eram propriedade da SportTV, enfraquecendo a posição do canal português. A outra movimentação no mercado deveu-se à criação de um canal novo de origem, embora neste caso não se trate de um canal premium, mas sim de acesso livre e disponível do pacote geral das operadoras: o Canal 11. Este canal é propriedade da Federação Portuguesa de Futebol (FPF), e aposta em transmissões de escalões inferiores, como o Campeonato de Portugal, e campeonatos de formação como a Liga Revelação, para além do acompanhamento às seleções nacionais e uma maior cobertura do futebol feminino. Estes dois novos canais vieram "abandar" com o panorama atual das transmissões do futebol em Portugal.

### **3 METODOLOGIA**

Este capítulo terá, numa primeira fase, a apresentação e explicação da metodologia que acompanhará este projeto, a metodologia CRISP-DM. Depois dessa fase serão apresentadas as tarefas que a metodologia define, de modo a suportar sustentadamente a fase dos resultados do capítulo seguinte.

A *Cross Industry Standard Process for Data Mining* (CRISP-DM) é uma metodologia de investigação que pode ser definida como um modelo com processos hierárquicos que estabelece as diferentes fases na implantação de um projeto de *Data Mining* (mineração de dados), consistindo num conjunto de tarefas descritas em quatro níveis de abstração (da generalidade à especificidade): fases, tarefas genéricas, as tarefas específicas e as instâncias do processo (Chapman et al., 2000).

Por se tratar de um estudo na área da mineração de dados, a metodologia utilizada neste estudo suportou-se na metodologia CRISP-DM. Concebida em 1966, o seu modelo de referência é composto por 6 fases principais, respetivamente, o entendimento do negócio, o entendimento dos dados, a preparação dos dados, a modelação, a avaliação do modelo e a implementação do modelo (Chapman et al., 2000). A pertinência e atualidade da metodologia CRISP-DM na área da mineração de dados e do *business intelligence* continuam a ser evidentes em publicações anteriores (Belfo & Andreica, 2018; Huber et al., 2019).

Esta metodologia pode ser utilizada em qualquer projeto de *data mining*, das mais variadas áreas de negócio. Este modelo de referência associa, ao ciclo de vida de um projeto de *data mining*, as 6 fases referidas anteriormente com tarefas específicas e relação entre essas mesmas tarefas, tal como é apresentado na Figura 10.

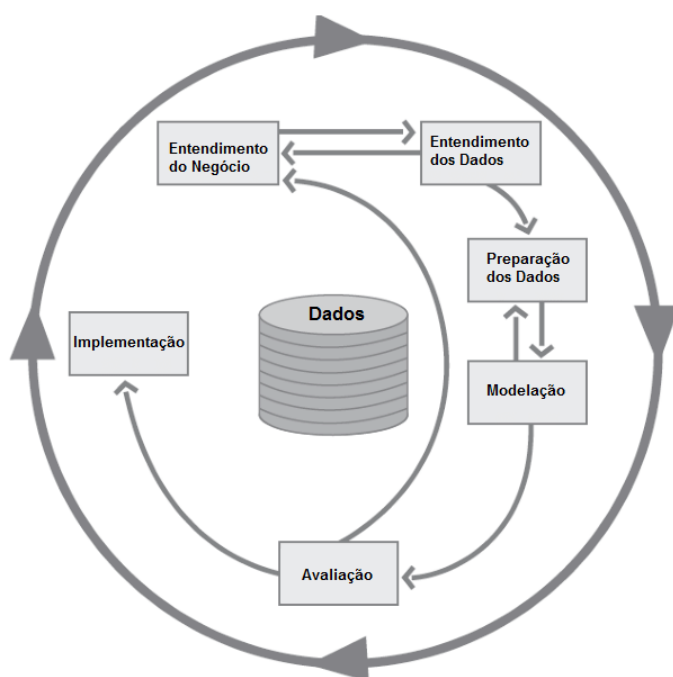


Figura 10 - Fases da metodologia CRISP-DM

Fonte: Adaptado de Chapman et al. (2000)

Neste projeto, as 6 fases existentes que constam na Figura 10 são aplicadas da seguinte forma:

- Entendimento do negócio:

Esta fase corresponde à revisão de literatura efetuada. Em resumo, ao se representarem as preferências e hábitos de consumo do público, as audiências assumem um papel preponderante na tomada de decisão dos meios de comunicação. As previsões feitas na área das audiências oferecem por isso a cada estação uma vantagem na negociação de contratos publicitários, e um conhecimento antecipado do real potencial dos seus conteúdos. Quanto menor for o erro associado aos modelos de previsão, mais capacitada e confiante estará a entidade em causa na abordagem ao mercado, pois consegue antecipar eventuais problemas. A projeção do desempenho de um conteúdo, através das métricas de audiências, pode levar a esse mesmo conteúdo sofrer ajustes, ao nível de orçamento, guião, estrutura programática, horário de emissão, plataforma em que é transmitido ou em casos mais críticos se decida pelo término do projeto. As diversas possibilidades inerentes à projeção de audiências constituem uma mais-valia para os meios de comunicação, permitindo-lhes ter mais controlo e

conhecimento relativamente aos conteúdos que disponibilizam ou querem vir a disponibilizar.

- Entendimento dos dados

A etapa do conhecimento dos dados permite uma familiarização inicial com os dados, de forma a conhecer as suas características, potencialidades, qualidades e utilidade para o projeto. Desta fase podem resultar também algumas hipóteses importantes para o desenvolvimento dos modelos, tal como subconjuntos de dados.

- Preparação dos dados

Esta é a fase que delimita quais dos dados do conjunto inicial transitam para serem integrados no modelo, através de tarefas como a seleção, limpeza e transformação dos mesmos. A forma como esta fase foi desenvolvida é apresentada mais à frente neste relatório.

- Modelação dos dados

Para cada problema de *data mining* existem diversas técnicas possíveis, que podem ser aplicadas, cada uma delas com características diferenciadoras. Esta fase consiste em, do universo de técnicas disponíveis escolher aquelas que se julgam mais adequadas, contando que por vezes essa escolha possa levar novamente à fase anterior, para orientar os dados com a metodologia escolhida. A forma como esta fase foi desenvolvida é também apresentada mais à frente neste relatório.

- Avaliação

Após a construção dos modelos é importante validar se os mesmos vão ao encontro do estipulado previamente. Na sequência da avaliação, consoante o resultado, podem surgir 3 cenários diferentes: avançar com o projeto para a fase de implementação, retroceder algumas fases da metodologia CRISP-DM para corrigir alguns aspetos do processo, ou num caso mais extremo recomeçar todo o processo. Quando é necessário retroceder ou começar um novo processo, a própria avaliação já constitui uma fonte de conhecimento, pois oferece uma nova abertura sobre a problemática, o que permite encarar o problema com novas informações.

Igualmente, apresenta-se mais à frente neste relatório a forma como esta fase foi desenvolvida neste projeto.

- **Implementação**

A derradeira fase é a de implementação, que se resume à aplicação prática, na realidade do negócio, do conhecimento extraído das etapas anteriores. Esta etapa corresponde ao propósito final da criação dos modelos, que é a utilização dos mesmos por parte do cliente. Esta fase não foi concluída na prática, mas, após a apresentação do conhecimento gerado, prevê-se que tal suceda no futuro.

## **4 ENTENDIMENTO DOS DADOS**

A unidade de análise definida para este estudo foi a da “transmissão de jogo de futebol”. Os dados escolhidos para avaliar o desempenho da transmissão televisiva, no que toca às audiências, foram escolhidos considerando essa unidade de análise.

Tal como se desenvolveu no capítulo da revisão de literatura, a audiência de um jogo poderá depender de diversos fatores e condicionantes, os quais foram tidos em conta enquanto atributos a considerar nos modelos desenvolvidos. Esses atributos (variáveis explicativas ou exógenas) pretendem contemplar diversos aspetos relacionados com um jogo, como as questões competitivas, aspetos geográficos, demográficos e económicos, característicos a cada um dos jogos. A audiência correspondente à variável a explicar (ou endógena), pretendendo-se verificar o impacto que cada variável explicativa tem ou não nessa mesma audiência.

### **4.1 Dados de televisão**

Existem diversas métricas para avaliar o desempenho de uma emissão de televisão, porém neste projeto foram escolhidas as métricas mais utilizadas no contexto português:

- **Rating%:** percentagem média da população portuguesa que assistiu a um determinado programa/canal, num intervalo definido;
- **Rating#:** número médio de pessoas, em milhares, que viu o conteúdo em análise;
- **Share%:** o share percentual representa a percentagem da população que visualizou um certo programa/canal, considerando todas as pessoas que estavam a ver televisão nesse período de tempo. A título de exemplo, um evento tem 50% de share, se for visto por metade das pessoas que estavam a ver televisão aquando da sua emissão.

### **4.2 Unidade de análise**

Como referido na secção acima a unidade de análise escolhida foi a transmissão de jogo de futebol. Este tipo de eventos tem, por norma, a duração de 90 minutos que são divididos em 2 partes, com 45 minutos cada, divididas por um intervalo de 15 minutos. Estas durações são sempre variáveis, tendo em conta o tempo de compensação dado pelos árbitros em cada uma das partes e também depende se o jogo tem ou não um prolongamento. Uma transmissão televisiva tradicional respeita a estrutura dos jogos, “aproveitando” para fazer um intervalo publicitário coincidente com o descanso do jogo

e adicionando-lhes pequenos segmentos de previsão, análise e comentário em curtos espaços de tempo no início e final de cada parte.

A unidade escolhida é o jogo por se acreditar que possui o nível certo de especificidades que podem ajudar a explicar o que causa um maior impacto nas audiências.

### 4.3 Recolha dos dados

A etapa da recolha dos dados revelou alguma complexidade fruto da natureza dos mesmos, o que se traduziu num processo bastante demorado. Como não existia um conjunto de dados que contemplasse todos os aspetos em estudo, foi necessária a criação de um conjunto de raiz.

#### 4.3.1 Procedimento de recolha de dados

Como a unidade de análise em estudo é a transmissão televisiva de um jogo de futebol foi necessário recolher informação em diversas fontes, sobre alguns dos aspetos que caracterizam essa mesma transmissão. Como demonstra a Figura 11, a caracterização de cada transmissão foi feita através de 4 tipos de informação: dados do jogo, dados de desempenho das equipas, dados relativos ao contexto geográfico, económico e demográfico em que as equipas estão inseridas e por último, dados relativos aos resultados de audiências que essa mesma transmissão obteve.

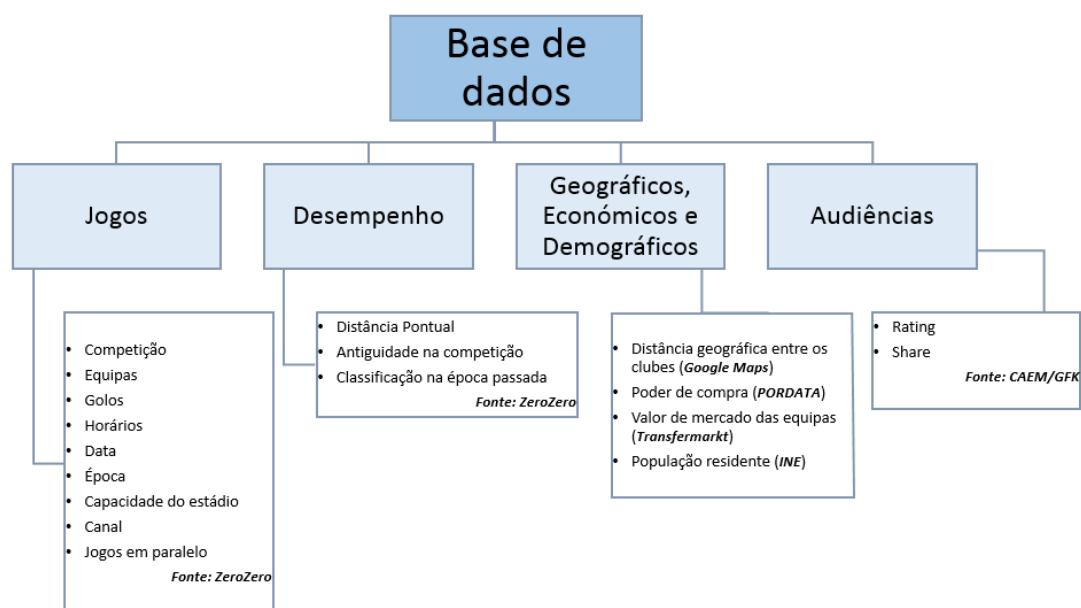


Figura 11 – Diversas proveniências para a construção do conjunto de dados

#### **4.3.2 Dados de jogos e desempenho**

A base de dados começou a ser construída pelos dados que representam o jogo de futebol, como a competição a que pertence, as equipas que o disputam, o número de golos marcado, o horário e data em que se disputou, a época desportiva em que se insere, a capacidade do estádio, o canal de transmissão e o número de jogos em paralelo da mesma competição que foram transmitidos em simultâneo.

Este processo iniciou-se a 27 de julho de 2020, terminando no dia 21 de agosto do mesmo ano, demorando assim cerca de um mês. Estes dados relativos aos jogos foram retirados do Zerozero, um portal português dedicado ao futebol e a outras modalidades desportivas, que contém uma das maiores bases de dados de acesso gratuito no mundo. Todas as competições e jogos considerados constavam no Zerozero e por isso as outras fontes de informação relacionadas com os aspetos do jogo serviram apenas para cruzamento e validação dos dados.

O Zerozero, que é um Órgão de Comunicação Social (OCS) inscrito na Entidade Reguladora para a Comunicação Social (ERC), possui também o registo da tabela classificativa das épocas transatas (ver ANEXO I), atualizado jornada a jornada, o que permitiu recolher a informação dos dados de desempenho das equipas. Estes dados foram apenas considerados para as competições no formato campeonato, tendo em contas as suas especificidades e traduziram-se na distância pontual de cada uma das equipas em jogo ao 1º classificado, no início da jornada em que o jogo decorreu, e o historial desses mesmos clubes na competição, seja através da antiguidade na mesma ou a classificação obtida na época anterior.

O preenchimento de alguns dos dados de cada um dos registos (jogo), fez-se com a utilização de tabelas auxiliares, que agilizaram todo o processo. Por exemplo, como foi criada uma tabela com a listagem das equipas, sempre que se inseria uma delas no registo de um jogo, algumas informações como o valor de mercado, a antiguidade na competição e a classificação no último ano eram preenchidas automaticamente, por se considerarem estanques durante toda a época, evitando assim a sua inserção jogo a jogo. Para os jogos de campeonato, o tempo estimado para preenchimento dos detalhes de cada jornada foi de 10 minutos e 40 segundos, o que tendo em conta a dimensão das jornadas da Liga Portuguesa (9 jogos por jornada), dá uma média de 1 minuto e 11 segundos por jogo.

### **4.3.3 Dados geográficos, económicos e demográficos**

Os dados considerados geográficos, económicos e demográficos, ao contrário do que aconteceu com os dados referentes aos jogos, foram recolhidos de diversas fontes diferentes. A recolha da distância entre as duas equipas foi feita através do Google Maps, considerando a distância, em quilómetros (km), entre os estádios das equipas em jogo, feita através de carro (ver ANEXO III). Nas situações em que a viagem de carro não é viável, por exemplo num jogo entre uma equipa de Portugal continental e uma do arquipélago da Madeira, é considerada a distância em linha reta entre o ponto em que se situam ambos os estádios, um dado que também é facultado pelo Google Maps. Tendo em conta que a grande maioria dos encontros do campeonato português se vão repetindo, tanto na mesma época como nas seguintes, também foi criada uma tabela auxiliar que reúne as distâncias entre todos os clubes do campeonato, facilitando assim o processo de preenchimento da base de dados nesse aspeto. Com o auxílio da tabela auxiliar, o processo de preenchimento das distâncias de cada jogo foi ficando mais eficaz ao longo do processo, pois para a grande maioria dos encontros passou a existir um histórico de distância associada. Tendo em conta que os mapas da Google Maps são alvo de constantes atualizações, a escolha desta plataforma foi feita tendo por base a sua consistência, usabilidade e fiabilidade.

O poder de compra associado ao município em que cada equipa está estabelecida é uma variável que apenas foi considerada nos jogos entre equipas portuguesas e para a sua recolha utilizou-se a PORDATA, uma base de dados organizada e desenvolvida pela Fundação Francisco Manuel dos Santos, que possui dados oficiais e certificados tanto de Portugal, como da Europa. Tal como no poder de compra, os dados referentes à população residente, no caso dos clubes nacionais, foram considerados tendo em conta o município e a sua extração foi feita do portal do INE, um órgão que integra o Sistema Estatístico Nacional (SEN), que produz e difunde as estatísticas oficiais portuguesas, assegurando a supervisão e coordenação técnico-científica do SEN. Para o caso das seleções a população residente considerada foi a do total dos países que as mesmas representam e nos clubes internacionais a das localidades a que pertencem. Apesar de serem dados relativos aos plantéis desportivos, o valor de plantel é uma métrica que o portal Zerozero não contempla, por isso essa informação foi recolhida no Transfermarkt, que é um portal alemão especializado neste tipo de informação e usado globalmente.

#### 4.3.4 Dados de audiências

Os dados de audiências considerados, share e *rating*, são da responsabilidade da CAEM/GfK, as quais asseguram os serviços de audimetria para a televisão em Portugal desde 2012 e foram retirados do sistema de audimetria de televisão (CAEM/GfK, 2021). Em Portugal, como não há a possibilidade de recolher a informação relativa a todas as casas, as audiências televisivas são estabelecidas por um painel de 1.100 lares, que representam cerca de 3.000 pessoas. Para tal, existe um audímetro por cada televisão dos lares pertencentes ao painel, aparelho que utiliza a tecnologia de *audiomatching*, de forma a identificar que canal está a ser transmitido e os habitantes da casa, através de um comando, sinalizam a sua presença na divisão onde a televisão está ligada. Conforme mostra a Figura 12, o comando (elemento 2 da Figura 12) tem vários botões, cada um associado ao perfil de um dos habitantes do lar e a interação com o mesmo é a única vertente em que existe ação humana neste processo.



Figura 12 - Audímetro (1 e 3) e comando (2)  
Fonte: Imagem por Reprodução/GfK em Espalha-Factos (2020)

De seguida, o sistema valida a informação recebida, tanto a indicada pelos painelistas, como a recolhida pelo sistema de *audiomatching* e “é feita a ponderação”. A cada indivíduo que compõe a amostra é atribuído um peso, “para que o conjunto de indivíduos chamados nesse dia representem a população total“, como explica Joelma Garcia, responsável pela área de audiências televisivas da GfK (Ferreira, 2020).

Para a criação da base de dados usada neste projeto, a extração das audiências associada a cada jogo foi feita utilizando a YUMI (Marktest, 2017), um *software* de análise, parametrizando as horas em que os jogos estiveram a ser emitidos.

#### **4.4 Descrição dos dados**

A base de dados criada é baseada na unidade de análise definida nos capítulos anteriores, a transmissão de futebol. Partindo desse pressuposto contém 798 registos, um por cada linha, sendo cada um correspondente a um jogo de futebol da Primeira ou Segunda Liga, referentes às épocas desportivas 2018/2019 e 2019/2020. Os atributos que representam cada um dos registos são 55, sendo alguns originais, outros auxiliares, que são gerados através do cruzamento ou interpretação dos originais, não esquecendo os que foram definidos para auxiliar na interpretação e leitura da base de dados (ver Quadro 1 e Quadro 2 do Apêndice 1).

Apresenta-se em seguida a definição dos atributos:

- **Divisão** – Variável numérica que varia entre 0 e 1 e pretende definir a importância das competições no contexto em que se inserem. O número “1” representa o principal campeonato do país, a Primeira Liga neste caso e o “2” a Segunda Liga.
- **Competição** – Designação da competição a que o jogo pertence.
- **Equi1** – Equipa que joga na condição de visitada.
- **Equi2** – Equipa que joga na condição de visitante.
- **Hora** – Hora de início do jogo
- **Data** – Dia em que o jogo é disputado, no formato dd-mm-aaaa.
- **DiaSemana** – Dia da semana em que o jogo é disputado.
- **Mês** – Mês do ano em que o jogo é disputado.
- **Ano** – Ano em que o jogo é disputado.
- **Época** – Período desportivo no qual o jogo se enquadra, no formato aaaa-aaaa.
- **DistAdv** – Distância, em quilómetros (km), entre o estádio das duas equipas, no caso dos jogos de clubes. Nos encontros de seleções, não disputados em campo neutro, considera-se a distância entre a capital económica do país visitante com o local em que a partida se disputa. Nos jogos disputados em campo neutro, ou seja, num recinto que não pertence a nenhuma das equipas, é calculada a distância entre as duas capitais.

- **DistPontualEqui 1/2** – Distância pontual da equipa 1 face ao 1º classificado do grupo ou campeonato a que o jogo se refere. Não se aplica a jogos a eliminar, casos em que se aplica o valor “0” neste atributo.
- **DistPontualEqui 2** – Distância pontual da equipa 2 ao 1º classificado do grupo ou campeonato a que o jogo se refere. Não se aplica a jogos a eliminar, casos em que se aplica o valor “0” neste atributo.
- **SomaDist** – Soma das distâncias pontuais das 2 equipas em jogo, relativamente ao 1º classificado.
- **DifDist** – Valor absoluto da diferença entre as distâncias pontuais das 2 equipas em relação ao 1º classificado.
- **PodCompEqui 1** - Poder de compra *per capita* do município a que equipa 1 pertence. Os valores são dados em percentagem (%), onde 100% representa a média de poder de compra *per capita* de Portugal. Só são consideráveis para os jogos entre equipas nacionais.
- **PodCompEqui 2** - Poder de compra *per capita* do município a que equipa 2 pertence. Os valores são dados em percentagem (%), onde 100% representa a média de poder de compra *per capita* de Portugal. Só são consideráveis para os jogos entre equipas nacionais.
- **DifPodComp** – Valor absoluto da diferença entre o poder de compra associado às duas equipas.
- **SomaPodComp** – Soma do poder de compra associado às duas equipas.
- **MedPondPodPop** – Média do poder de compra das duas equipas em jogo, ponderada com a população afeta a cada uma delas.
- **PopEqui 1** – População residente do município a que a equipa 1 faz parte, no caso dos clubes nacionais. Nas seleções nacionais é considerada a população total do país e nos clubes estrangeiros, a população da cidade em que estes se inserem.
- **PopEqui 2** – População residente do município a que a equipa 2 faz parte, no caso dos clubes nacionais. Nas seleções nacionais é considerada a população total do país e nos clubes estrangeiros, a população da cidade em que estes se inserem.
- **DifPop** – Valor absoluto da diferença entre a população residente de cada uma das equipas.
- **SomaPop** – Soma da população residente associada a cada uma das equipas.

- **PartiEqui 1** – Número de participações da equipa 1 na competição em causa, apenas considerável para jogos do campeonato.
- **PartiEqui 2** – Número de participações da equipa 2 na competição em causa, apenas considerável para jogos do campeonato.
- **DifParti** – Diferença entre o histórico de participações na prova das duas equipas.
- **SomaParti** – Soma do número de participações das duas equipas no campeonato.
- **AntiCompEqui 1** – Número de presenças consecutivas da equipa 1, desde a última entrada. Apenas são considerados dados para os campeonatos nacionais.
- **AntiCompEqui 2** – Número de presenças consecutivas da equipa 2, desde a última entrada. Apenas são considerados dados para os campeonatos nacionais.
- **DifAntiComp** – Valor absoluto da diferença entre o número de participações consecutivas das duas equipas, na competição considerada.
- **SomaAntiComp** – Soma do número de participações consecutivas que as duas equipas possuem.
- **ClaUAnoEqui 1** – Classificação da equipa 1 da edição anterior da competição, apenas considerando o campeonato nacional. Para as equipas recém-promovidas assume-se a classificação que tiveram na divisão inferior, adicionando o número total de equipas no campeonato em causa. Por exemplo: um clube recém-promovido à Primeira Liga de Portugal, que terminou em 2º lugar da Segunda Liga na época transata, vai ter como valor na variável ClaUAnoEqui o número 20, resultante da soma da sua classificação na época anterior (2) com o número de equipas da Primeira Liga (18).
- **ClaUAnoEqui 2** – Classificação da equipa 2 da edição anterior da competição, apenas considerando o campeonato nacional. Para as equipas recém-promovidas assume-se a classificação que tiveram na divisão inferior, adicionando o número total de equipas no campeonato em causa. Por exemplo: um clube recém-promovido à Primeira Liga de Portugal, que terminou em 2º lugar da Segunda Liga na época transata, vai ter como valor na variável ClaUAnoEqui o número 20, resultante da soma da sua classificação na época anterior (2) com o número de equipas da Primeira Liga (18).
- **DifClaUAno** – Diferença entre a posição no campeonato obtida pelas duas equipas, na época anterior.

- **SomaClaUAño** – Soma das posições classificativas ocupadas pelas duas equipas, na época anterior.
- **VMEqui 1/2** – Valor de mercado, em milhões de euros (€), do plantel da equipa 1/2, na época a que o jogo diz respeito. Para os clubes pertencentes às principais ligas europeias considera-se o valor a 1 de agosto do ano em que a época se inicia e aos restantes o valor global da época. No caso das seleções nacionais, a variável VMEqui toma o valor de mercado que o plantel dessas mesmas seleções apresentava no 1º dos dois anos que uma época engloba.
- **VMEqui 2** – Valor de mercado, em milhões de euros (€), do plantel da equipa 2, na época a que o jogo diz respeito.
- **DifVM** - Diferença do valor de plantel das duas equipas.
- **SomaVM** – Soma do valor de plantel das equipas em jogo.
- **Golos** – Número de golos marcados no jogo.
- **CapEstadio** – Capacidade do estádio em que o jogo é disputado.
- **Canal** – Canal televisivo em que o jogo é transmitido. Quando o jogo é simultaneamente emitido num canal aberto e num canal *premium* considera-se o canal aberto, pela possibilidade que os mesmos têm em chegar a um número maior de telespetadores.
- **PreçoCanal** – Mensalidade, em euros (€), associada à subscrição do canal em que o jogo foi transmitido. Para os jogos disponíveis em sinal aberto, o PreçoCanal é considerado 0 euros.
- **Paralelo** – Número de jogos, da mesma competição, disputados num horário coincidente com a da partida em causa. Nas competições com fase de grupos são considerados apenas os jogos do mesmo grupo e nas fases a eliminar das provas europeias são apenas tidos em conta os jogos de equipas portuguesas transmitidos em simultâneo. A variável Paralelo não é considerada nos jogos de carácter particular.
- **Rat#** - Número médio de pessoas, em milhares, que viu o jogo em análise.
- **Rat%** - Percentagem média da população portuguesa que assistiu ao jogo.
- **Shr%** - Percentagem da população que que televisionou o jogo, considerando todas as pessoas que estavam a ver televisão nesse período de tempo. A título de exemplo, uma partida de futebol tem 50% de share (Shr%), se for visto por metade das pessoas que estavam a ver televisão aquando da sua emissão.

## **4.5 Exploração dos dados**

Na fase de exploração de dados poderão surgir algumas ilações sobre o tipo de respostas que o conjunto de dados (*dataset*) oferecerá. Esta etapa poderá também contribuir para a fase de limpeza dos dados ou transformação dos mesmos.

Neste capítulo, faz-se uma exploração dos dados para cada uma das principais variáveis presentes, em especial, em relação aos resultados de audiências a que estão associadas, com o objetivo de perceber qual o impacto que cada uma delas pode ter nas transmissões televisivas. Algumas das variáveis tiveram de ser adaptadas a valores numéricos, ou enquadradas em classes para obter uma melhor representatividade e interpretação dos dados.

### **4.5.1 Primeira ou Segunda Liga**

A base de dados contém jogos das duas principais ligas de futebol em Portugal, a Primeira e a Segunda Liga. É expectável que os jogos da Primeira Liga sejam muito mais atrativos para o público, pois é onde estão presentes as equipas com maior dimensão social. A Primeira Liga tem melhores jogadores e é a competição que apura o campeão nacional. A Primeira Liga tem todos os seus jogos televisionados. Já na Segunda Liga, o número de jogos transmitidos a cada jornada é variável, sendo que, para a base de dados em estudo, são transmitidos em média 3,2 encontros na televisão por jornada.

A Primeira Liga, por ter mais jogos transmitidos do que a Segunda, é a competição mais representada na base de dados com 612 jogos dos 798 totais, ou seja 76,7% dos casos presentes no *dataset* são referentes ao principal escalão do futebol português. A percentagem alusiva à Segunda Liga, 23,3%, é ligeiramente inferior à esperada, pois na época 2019/2020 esta competição não retomou à atividade após a interrupção provocada pela pandemia de Covid-19, tendo-se realizado apenas 24 das 34 jornadas previstas regularmente. Caso tivesse tudo acontecido conforme esperado e mantendo a média de jogos transmitidos por jornada até então, a Segunda Liga estaria representada com uma percentagem entre os 25% e 26%. Em termos de audiências médias, a Primeira Liga regista um *rating* médio de 68,8 mil telespetadores e a Segunda Liga fixa-se nos 15,2 mil telespetadores. Os encontros que representam as marcas de audiência máximas em ambas as competições envolvem jogos entre Benfica e Porto: no caso da 1ª Liga, com 375,5 mil telespetadores, e 59,2 mil telespetadores no jogo da 2ª Liga, que envolveu as mesmas duas equipas secundárias, vulgo equipas B, de ambos os clubes. O jogo com mais

audiência da 2ª Liga não consegue superar a média dos 612 encontros da 1ª Liga, mas coincide com a marca mínima registada num jogo do principal escalão, um Paços de Ferreira x Moreirense (5/1/2019), que não foi além de 1500 telespetadores.

#### **4.5.2 Hora da transmissão**

A hora de transmissão tem influência na audiência que o mesmo consegue atingir, em qualquer conteúdo televisivo. O consumo varia ao longo do dia e de país para país, tendo em conta as rotinas de cada povo e a altura em que mais gente consegue ver televisão. Essa oscilação no consumo verifica-se também nos jogos de futebol, o que faz com que a marcação dos jogos tenha que ter em conta a transmissão do mesmo na televisão, de forma a captar uma maior audiência, mas também outros fatores como a conveniência para os adeptos que se deslocam aos estádios.

Dos horários presentes na base de dados considera-se apenas o valor das horas, excluindo os minutos da análise, por se considerar que o valor destes não provoca grandes oscilações no consumo.

Da observação à Figura 13 sabe-se que as faixas horárias em que começam mais jogos são as 20h (179 jogos – 22% do total de jogos), 15h (154 jogos – 19%) e 11h (114 jogos – 14%). Destas faixas destaca-se a particularidade de às 11h apenas se terem disputados jogos da Segunda Liga, o que traz outra evidência: 61,3% dos jogos da Segunda Liga (114 de 186) que são televisionados começam entre as 11h00 e as 11h59, um período matinal caracterizado por menor consumo televisivo. Esta opção de calendarização pelas entidades competentes pode-se explicar por ser um período de menor concorrência com jogos de outras ligas e consequentemente mais espaço livre na grelha programática das televisões.

Influenciado pelo facto de serem apenas jogos da 2ª Liga e jogados em períodos de menor consumo televisivo, o *rat#* médio dos jogos das 11h é o mais baixo, com 11,6 mil telespetadores por jogo. Das 3 faixas seguintes (12h, 13h e 14h) apenas a última tem um jogo, com 24,7 de *rat#* não dando para extrair daí grandes conclusões, a não ser que é um período que a Liga Portugal não explora com frequência. Os jogos das 15h e das 16h obtêm um *rating* médio semelhante, 16,7 mil e 15,3 respetivamente e apresentam uma característica desviante da tendência encontrada na análise às audiências por divisão. Apesar de terem um número total de partidas bastante díspar, 154 às 15h e 43 às 16h, as duas faixas são constituídas por cerca de 80% de jogos da 1ª Liga. O pormenor que vai

contra a tendência verificada no tópico 4.5.1 - Primeira ou Segunda Liga, é o facto de o principal campeonato não se conseguir superiorizar, em termos de audiência média, face à 2ª Liga. Tendo em conta os jogos das 15h, a Segunda Liga consegue uma audiência média (17,5 mil) superior aos encontros da Primeira Liga (16,5 mil), já nos encontros das 16h a 1ª Liga sobressai, mas com uma vantagem mínima (15,5 mil versus 14,4 mil). Estes resultados que contrariam a tendência natural de uma Primeira Liga mais pujante, podem ser justificados pelas equipas mais frequentes nestes jogos: apenas 4 dos 35 jogos da 2ª divisão não tem a presença de Benfica B ou Porto B, ou seja apesar de serem jogos de uma competição inferior estão, em larga maioria, envolvidas equipas com uma grande base de adeptos e que disputam o título na 1ª Liga. Outra das eventuais explicações está relacionada com as televisões responsáveis pela transmissão dos jogos. No caso dos encontros da 1ª Liga neste horário, é necessário subscrever a SportTV, um canal com um custo mensal de 24 euros, já na 2ª Liga, para além da SportTV, encontram-se jogos transmitidos na BTV (mensalidade de 10 euros) e no Porto Canal, canal incluído gratuitamente nos pacotes de televisão por cabo. Existindo jogos na 2ª Liga emitidos em canais de acesso mais facilitado, tendo em conta o preço de acesso, a base de espetadores possíveis de captar é maior, o que favorece as audiências desses encontros, que à partida são menos atrativos para o público.

Como é visível na Figura 13, os jogos que se iniciam a partir das 17 horas já se encontram noutra patamar, com médias de rat# por jogo situadas entre os 79,5 mil (19h) e os 93,5 mil telespetadores (21h). Nestes períodos horários os jogos da Segunda Liga com acompanhamento televisivo já são mais raros e representam apenas 7% do total dos casos. Ao contrário do que sucede no período da manhã, a fase da tarde-noite é mais “concorrida” em termos de calendarização de jogos das principais ligas, tanto nacionais como internacionais, não existindo tantas oportunidades para exibir encontros da Segunda Liga. Deste extenso período, entre as 17h e as 21h, é possível fazer uma divisão em duas classes, com semelhanças dentro das mesmas. Uma classe destinada ao período da tarde, que inclui jogos começados às 17h, 18h ou 19h, onde existem em média 75 jogos por hora e com um rat# de 81 mil telespetadores.

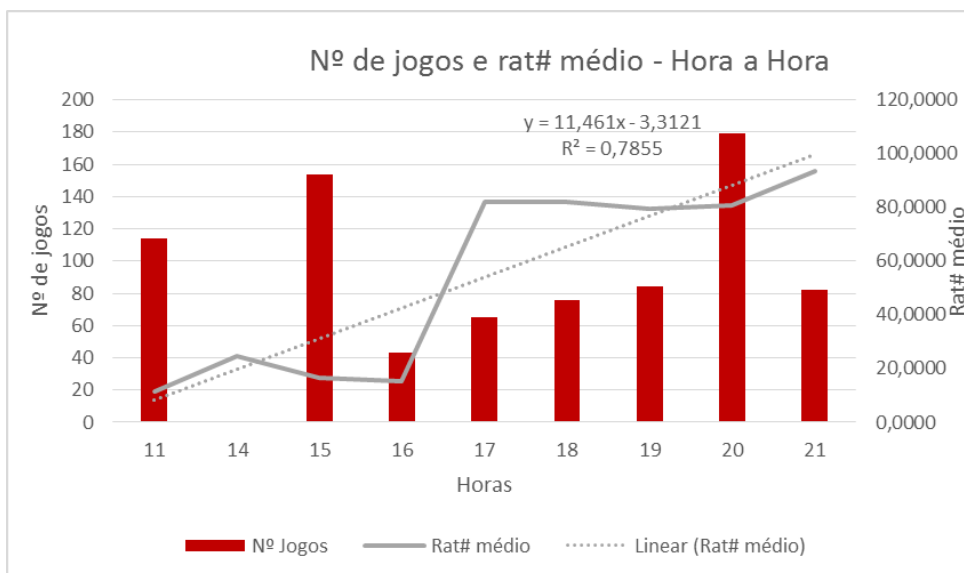


Figura 13 - Relação entre o rat# médio, o número de jogos e as horas do jogo

A outra classe é destinada ao período da noite, com os jogos começados às 20h e 21h, faixas essas em que o consumo televisivo é maior, vulgo *prime-time*, o que faz desses horários os mais cobiçados por todas as partes interessadas, sendo por norma destinados aos jogos mais atrativos. As características desta segunda classe são notórias em dois aspetos que a Figura 13 nos mostra: a faixa das 20h é aquela em que se iniciam mais jogos (22% do total da base de dados) e é na faixa das 21h que o *rating* médio por jogo é mais elevado, chegando aos 93,5 mil telespetadores, uma marca que supera em 37 mil indivíduos a média geral.

Como era de prever encontram-se evidências para que a marcação da hora de um jogo tenha impacto nas audiências do mesmo, não sendo, no entanto, um fator de impacto isolado. A hora a que um encontro se inicia leva em conta diversas partes do espetáculo desportivo, desde os adeptos, às equipas, sem esquecer a vertente financeira, representada pelas televisões, patrocinadores e gabinetes comerciais dos clubes e da Liga. Os efeitos provocados pela hora de início podem ainda ser mais ou menos impactantes aliados a aspetos como o canal em que a transmissão vai decorrer, as equipas envolvidas, ou a fase do ano/temporada que está a decorrer. É por isso fulcral que todas estas partes sejam consideradas e que aquando da definição das horas se perceba o impacto que isso possa trazer.

### 4.5.3 Dia da semana

Os dias da semana em que se disputam os jogos possuem uma abordagem idêntica à das horas, visto que a sua definição deve ter em conta a calendarização da época desportiva, mas também as oscilações de consumo televisivo.

O fim-de-semana é o período em que o consumo televisivo é maior e, para além disso, é também a fase em que se concentram maior número de jogos de futebol. Cerca de 70% do total dos jogos considerados na base de dados decorreram ao fim de semana, conforme demonstra a Tabela 1. Os dias adjacentes ao fim de semana, ou seja, a 2ª e 6ª feira, vêm de seguida como dias mais ocupados pelos jogos da 1ª e 2ª Liga, concentrando entre si 20% dos casos. Nos 3 dias restantes (3ª, 4ª e 5ª feira) é quando se realizam menor número de jogos nos campeonatos nacionais, nas épocas em análise. Estes são os dias da semana destinados aos jogos das competições europeias, sendo essa também uma das razões para a menor frequência de partidas de campeonato a meio da semana.

Dia da Semana	Rat# médio	Nº de jogos	% de jogos
2ª feira	74,0	66	8%
3ª feira	77,7	22	3%
4ª feira	71,4	32	4%
5ª feira	80,1	23	3%
6ª feira	59,5	94	12%
sábado	53,0	263	33%
domingo	49,2	298	37%
		<hr/>	
		798	

Tabela 1 – Rating médio, número e percentagem de jogos por dia da semana

Em termos de audiências médias, os dias da semana com menor número de jogos são aqueles que apresentam melhores resultados. Pela observação à Figura 14 são notórias as melhores médias de audiências para os encontros calendarizados nos dias de semana face ao que sucede no fim-de-semana. De 2ª a 5ª feira, o rat# médio supera a fasquia dos 70 mil telespetadores em todos os dias, ao contrário do que sucede na 6ª feira (rat# médio= 59,5 mil), sábado (rat# médio = 53 mil) e domingo (rat# médio = 49,2 mil).

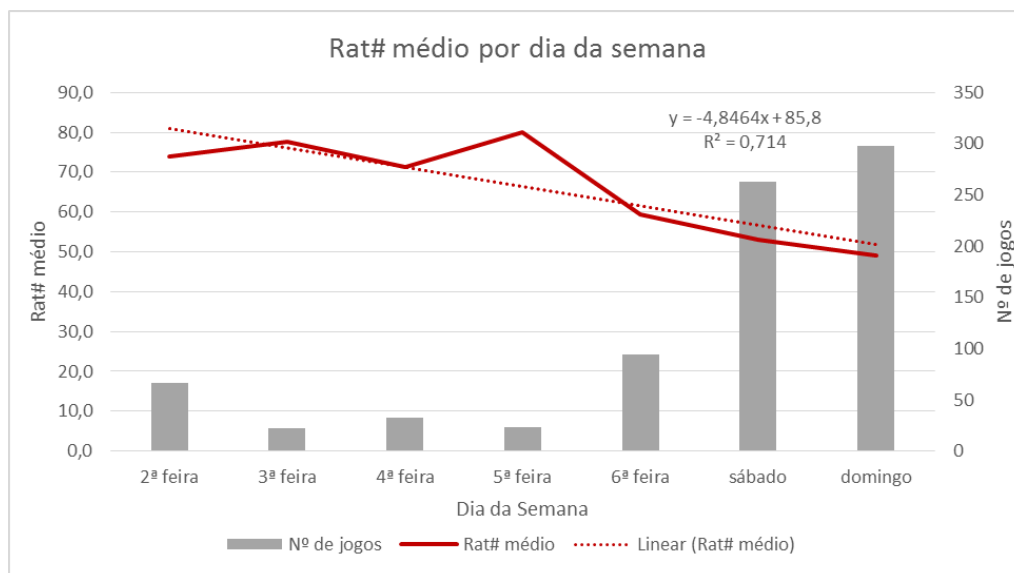


Figura 14 - Número de jogos e rating médio por dia da semana

Desta análise, evidencia-se que nos dias de semana conseguem-se em média maiores audiências, apesar do consumo televisivo generalizado ser mais baixo, mas a diferença face às médias que se registam nos fins-de-semana não é muito expressiva, tendo em conta a diferença abismal que existe em termos de oferta desportiva na televisão em cada um destes períodos. Cabe, portanto, às entidades interessadas e competentes definir uma estratégia de marcação dos jogos, sabendo o que acarreta cada uma das escolhas.

#### 4.5.4 Distância geográfica entre equipas

A Figura 15 apresenta a distribuição geográfica dos clubes incluídos no conjunto de dados, no mapa de Portugal. Esta distribuição foi atribuída consoante a localização dos estádios em que cada um joga na condição de visitado. Diretamente relacionado com essas localizações surge a variável que representa a distância entre um e outro clube, atributo esse que pode influenciar as audiências de um jogo de diferentes formas.

Caso se trate de um encontro em que a variável “distância geográfica entre equipas” tenha um valor muito baixo, ou seja, o estádio das duas equipas é muito próximo um do outro, podemos estar perante uma rivalidade regional, aspeto que proporciona encontros em que não está apenas em causa a vertente desportiva, mas também um valor sentimental provocado pelo bairrismo. Estes jogos, conhecidos na gíria como dérbis, têm a particularidade de atrair espetadores que não acompanham regularmente o desporto, mas que por se tratar de um acontecimento social, prestam uma maior atenção comparativamente com outras situações.

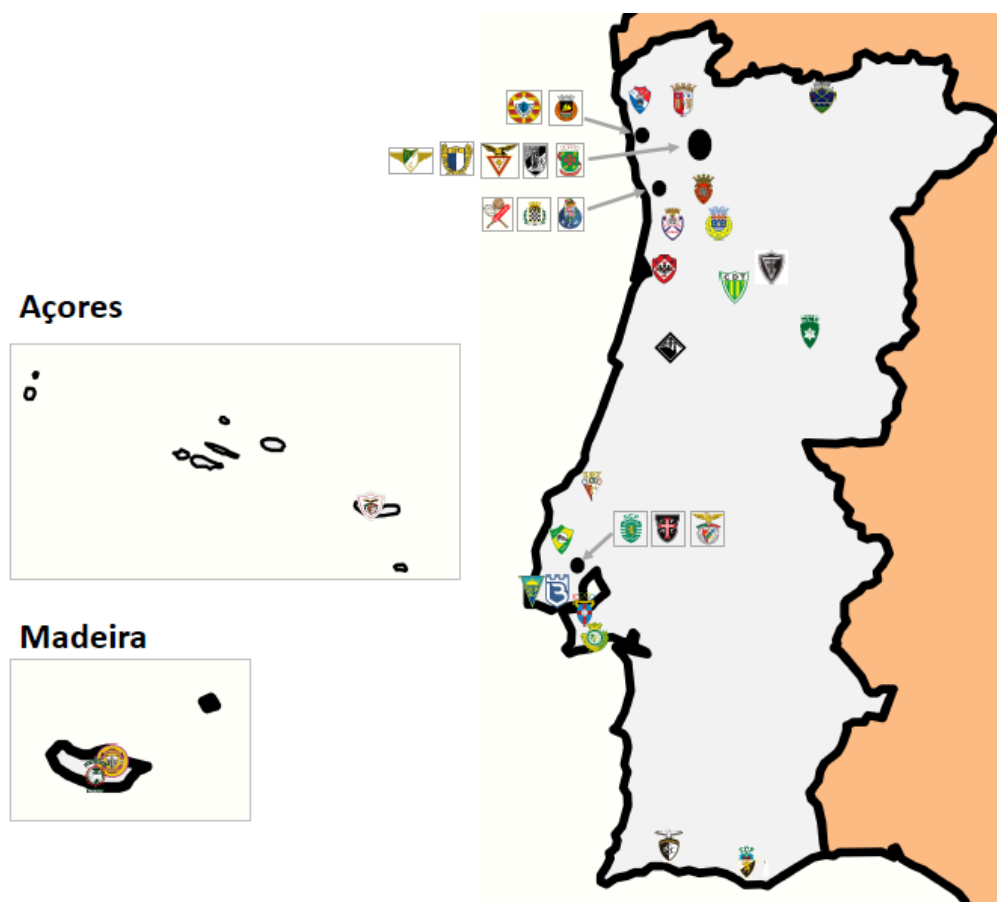


Figura 15 - Distribuição geográfica das equipas no mapa de Portugal

Nos jogos em que a distância é mais elevada o impacto nas audiências televisivas pode-se notar pela dificuldade dos adeptos visitantes se deslocaram presencialmente ao estádio. Imaginando um jogo que coloca frente-a-frente uma equipa algarvia e uma minhota, a distância é um fator que dificulta que os adeptos decidam ir assistir presencialmente ao jogo da sua equipa ao estádio, podendo significar mais de 5 horas de viagem em cada sentido. Este fator, aliado a outros aspetos como o dia da semana ou a hora, podem levar a que a opção passe por assistir ao encontro pela televisão, beneficiando assim as audiências, mas prejudicando a assistência do estádio.

A distância geográfica média entre as equipas, nos 798 jogos que compõem a base de dados, é de 445 km, sendo a distância mínima registada (4 km) nos confrontos entre Benfica e Sporting e o maior percurso (1.611 km) nos jogos do Chaves contra o Santa Clara. Um denominador comum dos jogos com maior distância entre adversários é que incluem clubes insulares, sejam eles dos Açores ou da Madeira, verificando-se a presença do Santa Clara, Marítimo ou Nacional em todos os encontros com uma distância superior a 800 km. O número de casos em que esta condição se verifica, distância maior que 800

km, é de 168 jogos, que registam uma média de 40,7 mil telespetadores. Do lado que reúne os confrontos locais, jogos com uma distância inferior a 50 km, o número de casos é aproximado (159), mas os jogos registam uma média de audiências com mais 14,7 mil telespetadores, registando um rat# de 55,4 mil. Este aspeto pode corroborar com a lógica dos dérbis serem jogos de interesse para uma maior franja de público.

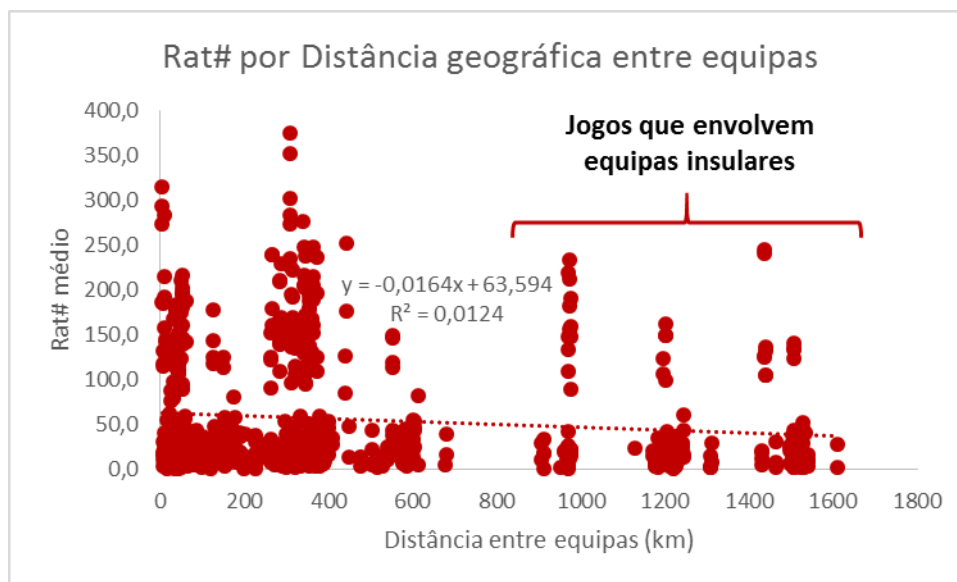


Figura 16 - Distribuição do rat# por distância geográfica entre equipas

Segmentando a análise das distâncias pelas duas ligas obtém-se que a média de distâncias percorridas na 1ª Liga é cerca do dobro da registada na 2ª Liga, com 502 km em média para a 1ª Liga versus 259 km em média para a 2ª Liga. Nos anos de análise, 2 dos 3 clubes insulares jogam as duas épocas na 1ª Liga e o outro clube, o Nacional da Madeira, compete uma temporada na 1ª Liga e outra na 2ª, fator que provoca uma diferença tão significativa entre as duas competições. Se a análise de distâncias médias por jogo for feita excluindo os clubes das ilhas os valores da 1ª e 2ª liga são mais equilibrados: 230 km na 1ª Liga e 222km na 2ª Liga.

A distribuição das audiências consoante a distância entre equipas, representada através da Figura 16, demonstra que a faixa dos 200 aos 399 km (266 jogos – 33% da amostra) é aquela em que o *rating* médio por jogo é mais elevado, ascendendo aos 75,9 mil telespetadores. A outra faixa que tem uma representatividade semelhante na base de dados é a que vai dos 0 aos 199 km e os seus 296 jogos (37% da amostra) alcançam um rat# médio de 51,2 mil, que se situa abaixo do valor geral para a audiência média (56,3 mil). Das restantes faixas analisadas nota para o facto de a faixa dos 400-799 km e dos mais de 1400 km terem um número de jogos aproximados, 68 e 62 respetivamente, e os seus

ratings médios serem também muito semelhantes, variando entre 40,2 mil nos 400-700 km e 40,6 nos jogos entre equipas que distam 1400 km ou mais entre si.

#### 4.5.5 Classificação das equipas

A distância pontual em relação ao primeiro classificado é uma variável que pretende representar não apenas a ordem classificativa geral, mas a classificação relativa de uma equipa. Como se trata de uma variável que mede uma diferença de pontos para o primeiro, é tido em conta não o número de lugares que distam do topo da tabela, mas a distância pontual para esse topo. Ou seja, evidencia-se a possibilidade que esta equipa ainda terá de chegar ao primeiro lugar.

Com a derivação resultante da soma das distâncias das duas equipas num jogo surge a variável SomaDist. Esta variável vai assumir um valor superior nos jogos entre equipas que estão a ter menos rendimento desportivo e menor para os jogos das equipas do topo da tabela, ou seja, aqueles em que provavelmente se espera que recolham mais audiência porque se enfrentam dois candidatos ao título. A SomaDist tem como mínimo o valor 0, que pode surgir em duas ocasiões: em todos os jogos da 1ª jornada do campeonato, ou quando é um jogo entre duas equipas empatadas pontualmente no 1º lugar do campeonato e como máximo 117, que representa um jogo entre duas equipas de fundo da tabela perto do final da época.

A Figura 17 revela a distribuição de audiências consoante a variável SomaDist e demonstra a tendência esperada inicialmente: a média de *rating* por jogo vai diminuindo quanto maior é o valor associado à SomaDist.

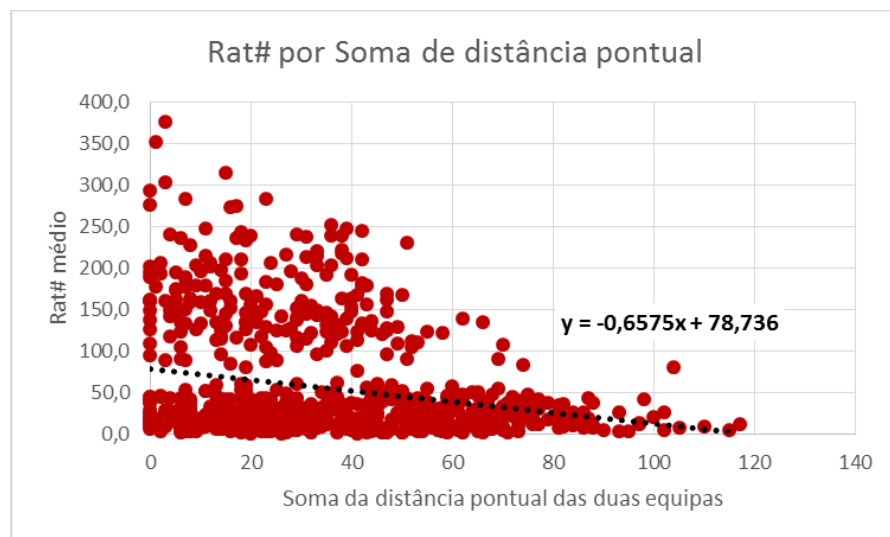


Figura 17 - Dispersão de jogos com o Rat# face à soma da distância pontual das duas equipas para o líder

Na Tabela 2 está dividida a amostra total por classes de 20 pontos, as médias mais elevadas são de 71,4 rat# e 61,9 rat# e dizem respetivamente respeito aos jogos com SomaDist de 0 a 20 e entre 21 e 40. No sentido oposto surgem as médias mais baixas: 28,4 rat#, 19 rat# e 20,5 rat#, referentes aos jogos com maior soma de distância pontual.

Outro dos aspetos que revela a tendência associada a esta variável é dada pela equação de regressão linear do gráfico:  $\text{Rat\#} = -0,6575 \text{ SomaDist} + 78,736$ . Desta equação é possível observar o sinal do coeficiente associado à variável SomaDist, que assume o sinal negativo. Este sinal negativo significa que, de forma global, quanto maior é soma das distâncias pontuais, menores são as audiências esperadas nesse jogo. O valor-p associado a esta equação de regressão linear é inferior a 0,05 e por essa razão a regressão é considerada significativa.

SomaDist	Rat# por jogo	Nº de jogos
0-20	71,4	284
21-40	61,9	225
41-60	46,5	156
61-80	28,4	96
81-100	19	30
101-117	20,5	7

*Tabela 2 - Classes de SomaDist*

Ainda relacionada com a diferença pontual existe a variável DifDist, que se assume como uma maneira de representar o equilíbrio entre as equipas que se defrontam. Nos 798 jogos analisados, esta variável oscila entre o 0 (encontros com equipas empatadas na classificação geral) e 54, tendo como média 9,52. No plano teórico, os jogos com menor diferença na distância pontual das duas equipas serão mais equilibrados pelo facto das equipas estarem classificadas numa posição aproximada e, por outro lado, os jogos com maior diferença representarão os confrontos entre equipas dos extremos opostos da tabela classificativa. Dos dados recolhidos observa-se que os jogos em que a DifDist é menor ou igual a 3, ou seja, jogos em que a posição das equipas é bastante aproximada e existe a possibilidade de uma alteração na classificação, o *rating* médio é de aproximadamente 41 mil telespetadores. No espectro oposto da variável, jogos com a DifDist maior ou igual a 31 pontos, a média de *rating* situa-se nos 154 mil. Estas conclusões levam a que, numa primeira instância, se considere que o público tem uma maior demanda por jogos desequilibrados, com equipas de valias muito díspares. No entanto, essa preferência pode também ser explicada pelo facto de nos jogos com maior DifDist constarem,

normalmente, as equipas consideradas 3 grandes (Benfica, Sporting e Porto) que possuem uma base de adeptos muito maior do que todas as outras equipas do campeonato. Ao contrário do que sucede com a variável SomaDist, neste caso a equação linear do gráfico tem um valor positivo associado à variável, demonstrando assim a tendência de maior audiência quanto perante maior desequilíbrio pontual entre as duas equipas.

#### **4.5.6 População do município**

Outra das variáveis que caracteriza social e demograficamente as equipas é a população residente nos concelhos a que as mesmas pertencem. Em teoria, espera-se que uma equipa inserida num grande centro urbano tenha uma maior massa adepta e conseqüentemente existam mais potenciais interessados em assistir aos jogos da mesma na televisão. Seguindo a mesma linha de raciocínio, é de esperar que uma equipa pertencente a um concelho com menor população atraia menos valores de audiências. Do rol de equipas presentes na base de dados, as que pertencem a concelhos com menor população são: Arouca, Chaves, Sporting Covilhã e Tondela, todas elas com menos de 50 mil habitantes. No extremo oposto, ou seja, das equipas situadas em maiores centros urbanos, surgem o Benfica, Casa Pia (2ª Liga) e Sporting, com mais de 500 mil habitantes e ainda o Porto, Boavista e Estoril (2ª Liga), que apresentam cerca de 215 mil habitantes nos anos analisados.

Se, por um lado, há indícios de que quanto maior for o desequilíbrio pontual entre duas equipas num jogo, maior audiência parece existir, então, por outro lado, também é possível que quanto maior for o desequilíbrio entre as dimensões das populações das duas equipas, maior audiência ao jogo possa existir.

A variável auxiliar DifPop representa a desigualdade, ao nível da população que é afeta a cada um dos clubes em jogo. Através desta variável, pode-se tentar perceber se os desníveis demográficos ajudam a impactar as audiências, partindo do pressuposto que quanto maior for a variável DifPop, maior será a diferença na base de adeptos das equipas. A Figura 18 parece mostrar a divisão em dois grupos de jogos, tendo em atenção a variável DifPop. Um primeiro grupo, com DifPop entre 0 e 200.000 (77% dos jogos) tem um *rating* médio de 35,2 mil telespetadores. O segundo grupo tem uma DifPop entre 290.000 e 500.000 (23% dos jogos) e um *rating* médio de 127,1 mil telespetadores.

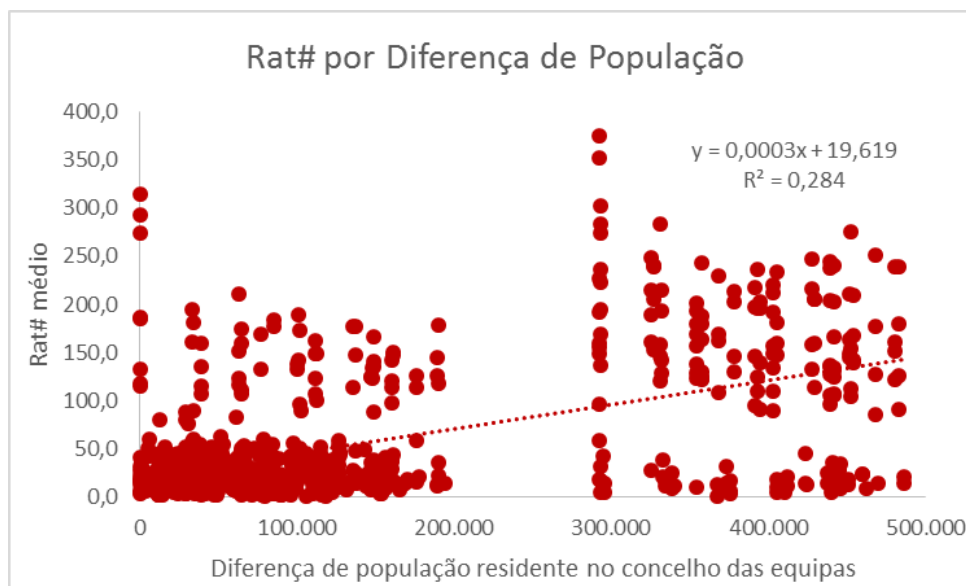


Figura 18 - Relação entre o rating e a diferença de população residente no concelho das equipas

Numa primeira análise, aos dados surge a evidência que a diferença entre população afeta a cada clube influencia fortemente as audiências, pois a média de espetadores alcançados é significativamente superior na classe em que a DifPop é mais elevada. No entanto, dos 183 jogos com DifPop superior a 290 mil, 73% incluem o Benfica ou o Sporting, 2 dos 3 grandes, fator que pode “desequilibrar a balança”, pois é de senso comum que a base de adeptos destes clubes não encontra paralelo em Portugal. Excluindo estes jogos da equação sobram apenas encontros do Benfica B e do Casa Pia com DifPop superior a 290 mil, todos eles a contar para a 2ª Liga, jogos esses que obtêm uma média de 18,4 mil de *rating*, bastante inferior à da classe de DifPop a que pertencem e cerca de metade da classe com DifPop entre 0 e 200 mil.

Será de esperar que quanto maior for a dimensão da população do concelho, das duas equipas de um jogo, mais elevado será naturalmente o *rating* desses jogos. A variável SomaPop tem um valor médio de 336 mil e representa a soma da população das localidades de cada uma das equipas presentes num encontro. Para melhor compreensão dos dados relativos à SomaPop, representados na Figura 19, procedeu-se a uma divisão dos mesmos nas 3 seguintes classes:

- SomaPop entre 0 e 240.000 (292 jogos): com um *rating* médio de 20,2 mil telespetadores;
- SomaPop entre 240.000 e 440.000 (318 jogos): com um *rating* médio de 46,2 mil telespetadores;

- SomaPop superior a 500.000 (188 jogos): com um *rating* médio de 129,4 mil telespetadores.

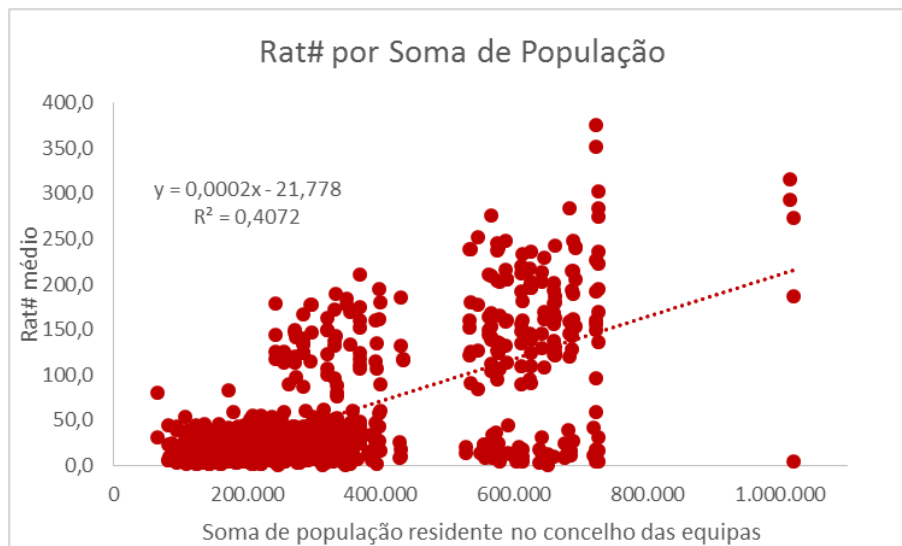


Figura 19 - Relação entre o rating e a soma de população residente no concelho das equipas

Como no caso da DifPop, parece confirmar-se que os valores médios de audiências vão aumentando consoante o número de população das equipas em jogo. No entanto, também aqui se verifica um forte impacto da presença dos clubes grandes. O Benfica e o Sporting dominam as audiências na classe de SomaPop superior a 500.000, impulsionando assim as audiências médias da mesma. Já na classe média, SomaPop entre 240 e 440 mil, o *rating* médio é mais modesto (46,2 mil), mas conta com os jogos do Porto, também considerado um dos “3 grandes”. Retirando os encontros do Porto da análise, a média para essa faixa desce para 24,3 mil, o que mesmo assim é um resultado acima da média geral de todos os jogos sem 3 grandes (21,7 mil telespetadores). Isto pode indicar que as variáveis populacionais têm influência nos resultados das audiências alcançadas, mas em menor escala do que parecem ter quando se suportam também em jogos dos “3 grandes”. Resumidamente a população do concelho das equipas parece ter influência nas audiências dos encontros, porque a ela está associada a dimensão do clube, mas não é ainda possível determinar a sua magnitude pelo facto de os clubes grandes influenciarem os dados em larga escala.

#### 4.5.7 Poder de compra dos municípios

A exploração dos dados relacionados com o poder de compra servirá para entender se este importante aspeto do contexto social dos habitantes de uma região pode inflacionar as audiências dos jogos das suas equipas. A associação da população residente num

concelho, em que uma equipa está sediada, aos seus adeptos é muito limitativa, principalmente no caso dos clubes de maior dimensão que têm grande representação e domínio por todo o país. Nos casos dos outros clubes acredita-se que a sua base de fãs se situe geograficamente no município da equipa e localidades vizinhas.

A variável *DifPodComp* estabelece a diferença entre o poder de compra associado às regiões das duas equipas. Assim sendo, esta variável assume os valores mais baixos nos jogos entre equipas pertencentes à mesma zona, ou de realidades económicas parecidas e os valores mais altos quando se trata de confrontos entre equipas de contextos mais desequilibrados. Nos encontros com *DifPodComp* entre 0% e 1%, apenas os jogos com equipas grandes presentes conseguiram superar os 100 mil telespetadores, sobressaindo os confrontos regionais, também conhecidos como *dérbis*: Sporting x Benfica e Porto x Boavista.

A importância e capacidade fora do comum de atrair espetadores por parte destes encontros já tinha “saltado à vista” no capítulo 4.5.4 - Distância geográfica entre equipas. Excluindo os casos específicos relacionados com os *dérbis*, a tendência associada à variável *DifPodComp* é de que há maiores audiências quanto maior for a diferença no poder de compra. As médias mais baixas de *rating* por jogo verificam-se nas franjas de valores mais baixas. Quando *DifPodComp* está entre 0 e 19% a média de *rating* por jogo é de 30,4 mil espetadores e nos casos em que essa diferença está entre os 20% e os 39% observa-se uma média de 18,8 mil. Em sentido inverso, os valores médios mais altos surgem quando a *DifPodComp* alcança valores mais elevados. Quando *DifPodComp* está entre 80% e 119%, o *rating* médio é de 110 mil telespetadores por jogo e quando *DifPodComp* está entre 120% e 149%, o *rating* médio é de 116,5 mil telespetadores por jogo.

Esta tendência revela que os jogos com equipas pertencentes a contextos sociais mais desequilibrados são mais atrativos para o público, sem considerar os casos particulares acima apresentados presentes nas diferenças de 0% e 1%.

Ao contrário da diferença no poder de compra, a soma do poder de compra das duas equipas agrega a capacidade de adquirir bens e serviços associado às duas equipas de um jogo, em vez de o distanciar. Nos jogos analisados, a variável *SomaPodComp* tem um valor médio de 243% e varia entre 154% e 440%. Do total de 798 jogos, 379 (47%) reúnem um valor de *SomaPodComp* entre 149% e 229% e alcançam uma média de 22

mil telespetadores por jogo, valor que não chega a metade da média global da base de dados (56,3 mil). Nos encontros que envolvem equipas que somam maior poder de compra encontram-se as médias mais elevadas de *rating*. Quando a SomaPodComp se situa entre 295% e 334%, o *rating* médio de 127,1 mil telespetadores por jogo. Nos casos em que a SomaPodComp figura entre 342% e 440%, o *rating* médio é de 133,1 mil telespetadores por jogo.

Através da comparação destas médias de audiências tão díspares sobressai a tendência de maior audiência dos telespetadores por jogos que envolvem equipas de contextos sociais mais favorecidos. Outra das explicações que daqui pode surgir está associada à transmissão dos jogos em canais premium, tipologia de canal associada a 95% dos jogos presentes na base de dados. Como os encontros são na sua grande maioria transmitidos em canais através de assinatura, pode surgir daqui um entrave para os adeptos de contextos mais pobres assistirem aos jogos das equipas da sua terra e assim não contribuírem para as audiências dos mesmos.

Para enquadrar melhor os dados relativos ao poder de compra e à população das equipas estabeleceu-se uma média ponderada entre essas duas variáveis, chamada MedPondPodPop. Esta média vai então oferecer uma representação do poder de compra das localidades afetas às equipas em jogo, consoante a população que essas mesmas localidades têm.

A distribuição dos valores afetos à média ponderada em discussão pode-se dividir em 3 grandes grupos. O primeiro grupo compreende os valores mais baixos entre os 78% e os 120% e representa perto de metade dos jogos analisados. Do conjunto destes 383 jogos chega-se a uma média de *rating* de 21,9 mil, tendo como partidas com *rating* mais elevado dois jogos que envolvem o Sporting de Braga, um contra o Rio Ave e outro contra o V. Guimarães. O conjunto intermédio contempla a MedPondPodPop entre os 121% e 158% (227 jogos, com 28% do total analisado), com médias de audiências de 53,8 rat# e 1,7% share. Explorando os jogos de maior audiência neste conjunto surgem dois a envolverem o Porto, um contra o Braga (210,9 de rat#) e outro com o V. Guimarães (197,4 de rat#), duas equipas minhotas. Observando os jogos da Primeira Liga que captaram menos telespetadores a equipa que mais se destaca é o Belenenses, com um jogo frente ao Boavista (2,6 de rat#) e outro com o Famalicão (4,4 de rat#). Destes dois conjuntos de jogos ressalta à vista a capacidade do Sporting de Braga e Vitória de Guimarães de sobressair entre os jogos mais procurados, algo à partida esperado visto se tratar dos

clubes, para além dos “3 grandes”, com maior número de adeptos e lutarem constantemente pelas posições cimeiras da tabela.

O último lote de jogos é aquele em que a média ponderada assume valores mais altos. Este grupo contempla uma MedPondPodPop de 190% a 220% e abrange 188 jogos (24% do total analisado) e onde as audiências atingem patamares mais elevados, tanto a nível de média como de registo único. A média nesta situação é de 129,4 rat#, ou seja, mais de 75 mil espetadores médios por jogo face ao conjunto intermédio, e os valores máximos alcançados superam os 350 mil telespetadores, em dois jogos envolvendo o Benfica e o Porto. Nos jogos menos apetecíveis para os telespetadores destacam-se partidas da Segunda Liga, que contam com a presença de equipas lisboetas, como o Benfica B e o Casa Pia.

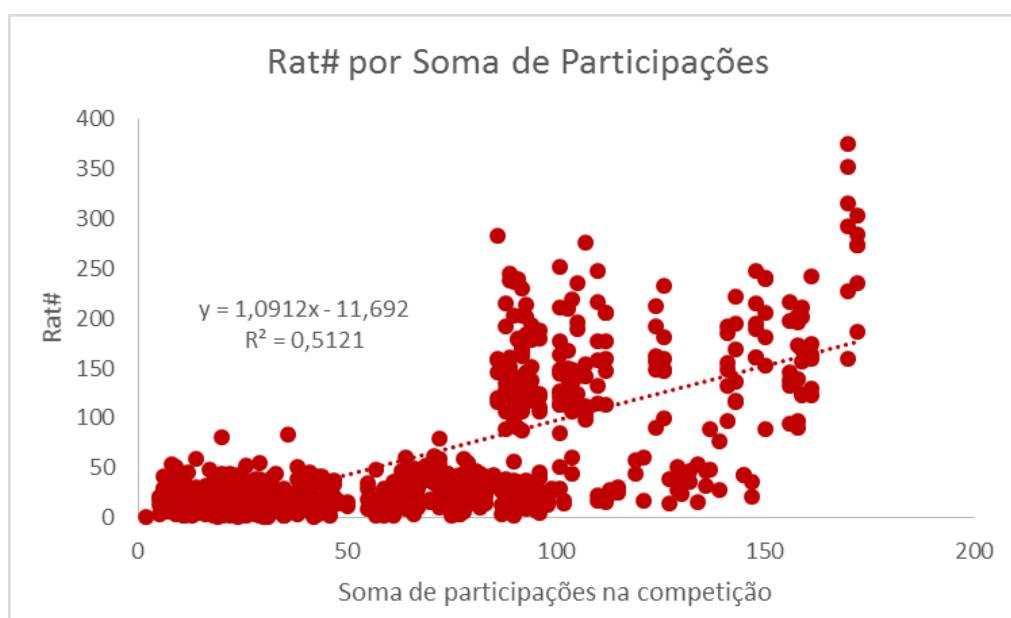
Desta análise resulta que as audiências televisivas podem ser fortemente beneficiadas quando se trata de um jogo com equipas que possuem elevado poder de compra nos concelhos a que pertencem. No entanto, estes resultados carecem de contextualização e de uma análise mais profunda, pois nos registos de audiências mais baixos de cada classe surgem os encontros da Segunda Liga e nesses casos o facto de existir uma MedPondPodPop elevada não parece ter importância suficiente para “ajudar” os jogos a chegar a mais gente.

#### **4.5.8 Participações na competição**

A quantidade de participações de um clube nas competições está associada a um histórico mais alargado nas mesmas e a um período mais alargado que poderá possibilitar a captação de mais adeptos e simpatizantes e, conseqüentemente, maiores audiências. Foi este racional que motivou a observação das variáveis relacionadas com as participações e a sua eventual implicação nas audiências.

A variável DifParti, representa a diferença do número de participações em diferentes edições passadas da competição dos dois clubes, que competem num determinado jogo. Nota-se novamente a preferência dos consumidores pelos encontros entre equipas de contextos distintos. Existem 171 registos referentes a jogos que colocam frente a frente equipas cuja DifParti é igual ou inferior a 5 edições e a média desses jogos é de 35,6 mil de *rating*, um registo inferior à média global de 56,3 rat#. Realce ainda para o facto de os jogos que envolvem, dérbis e clássicos, Benfica, Sporting e Porto estarem neste lote, facto que impulsiona a média total, sabendo-se que os confrontos estas equipas são sempre dos

mais vistos da época. Por outro lado, as maiores audiências médias surgem nos jogos com DifParti maior que 70 edições (130,2 de rat# em média nos 88 jogos), faixa essa que é maioritariamente composta pelos confrontos dos “3 grandes” com as equipas mais inexperientes do campeonato. O impacto destes 3 clubes também se visualiza na análise da variável SomaParti, pois quando a soma de participações das equipas não supera as 80 a média de *rating* fixa-se nos 19,9 mil e esta condição, SomaParti  $\leq$  80, não inclui qualquer um dos seus jogos pelo facto de, nas épocas analisadas, Benfica, Sporting e Porto terem sempre mais de 80 participações na Primeira Liga.



*Figura 20 - Rat# por soma de participações na competição dos 2 clubes de um jogo*

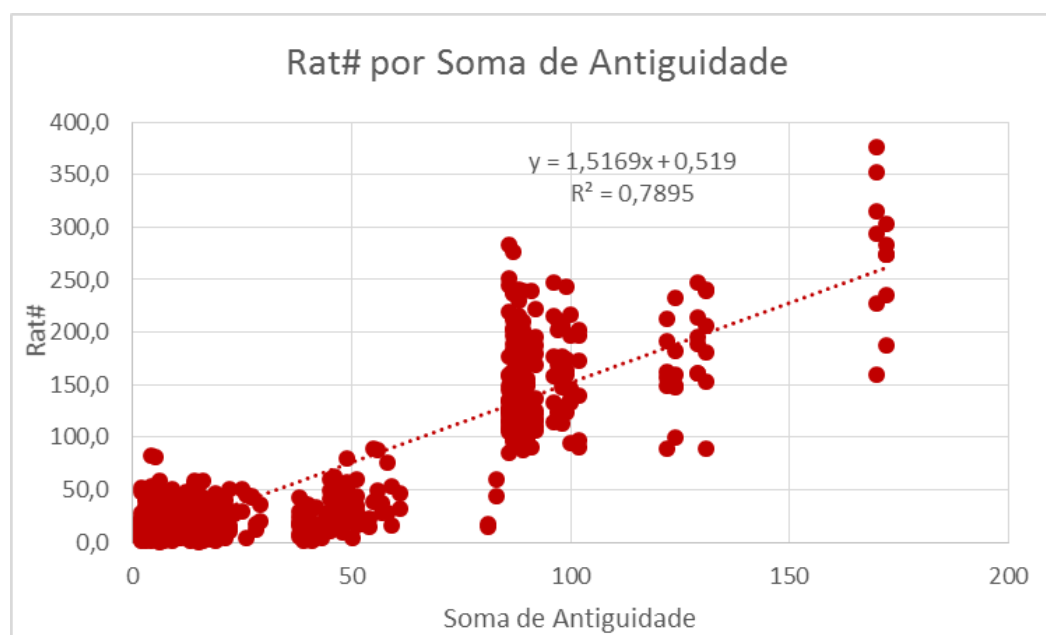
Na Figura 20 observa-se que a barreira dos 100 mil telespetadores só é ultrapassada em jogos em que as duas equipas somam participações em edições passadas (SomaParti) superior a 80. Outros dos aspetos que é possível verificar é que a partir das 140 participações, a maioria dos jogos fica acima dos 150 mil telespetadores, o que demonstra a preferência dos consumidores de futebol aos confrontos que envolvam equipas com maior histórico na competição.

#### **4.5.9 Antiguidade na competição**

A variável AntiComp representa o número de participações consecutivas na competição em causa que as equipas têm. A sua exploração é pertinente para observar de que forma o nível de consolidação que as equipas têm na liga que disputam influencia as audiências dos seus jogos. Na diferença de antiguidade entre as equipas, representada por DifAntiComp, o conjunto de encontros com os valores mais baixos, entre 0 e 15 neste

caso, que são os confrontos entre equipas com uma antiguidade aproximada, a média de *rating* por jogo é de 25,9 mil. Na classe de jogos seguinte, com *DifAntiComp* entre 22 e 48, o *rat#* médio registado é de 57,2, valor aproximado à média global da base de dados: 56,3. A classe com as diferenças mais díspares, *DifAntiComp* entre 70 e 85, reflete os encontros entre equipas com um historial mais distinto e prova disso é o facto de todos os jogos englobarem um dos “3 grandes”, Benfica, Sporting ou Porto. A audiência média destes jogos é de 155,6 mil telespetadores.

A outra variável resultante da antiguidade na competição é a *SomaAntiComp*, que soma as épocas que os dois clubes acumulam consecutivamente e tem um comportamento semelhante à da diferença entre esses mesmos valores.



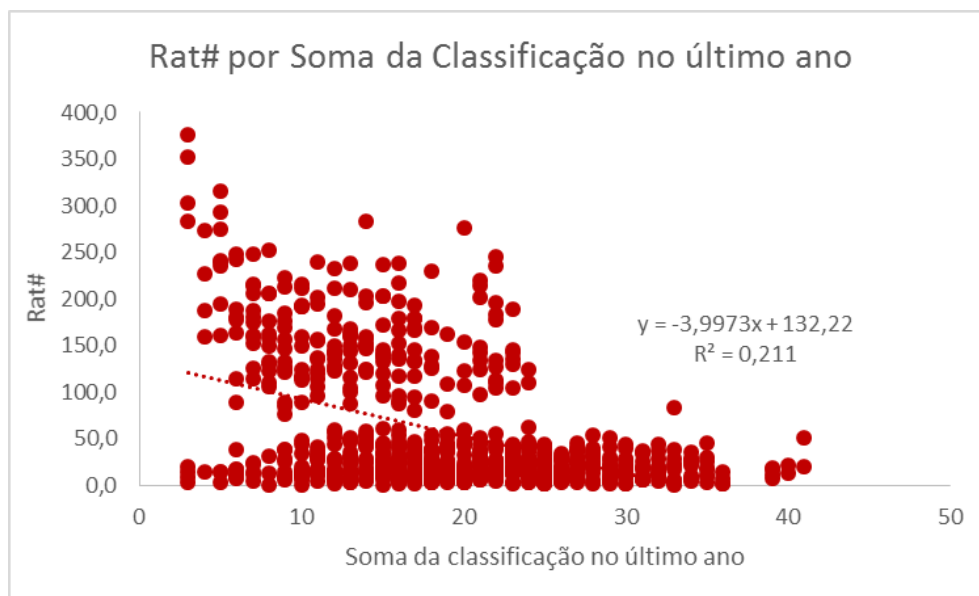
*Figura 21 - Rat# por soma de antiguidade na competição*

Conforme mostra a Figura 21, a distribuição das audiências por *SomaAntiComp* está representada em 3 classes de valores. O espectro que reúne os valores mais pequenos, *SomaAntiComp* menor ou igual a 61, tem todos os jogos em que não está presente uma equipa dos “3 grandes”, encontros esses que chegam a uma média de 21,6 mil de *rating*. Na 2ª classe, *rat#* médio = 155,5, estão todos os jogos dos “3 grandes”, excluindo os encontros entre eles, e os jogos do Sporting de Braga contra o Marítimo, duas das equipas que há mais tempo consecutivo se mantêm na Primeira Liga. O grupo de jogos com maior *SomaAntiComp* diz respeito aos confrontos de Benfica, Sporting e Porto entre si e supera a média de 273 mil telespetadores.

Da análise feita a estas variáveis é notória que a antiguidade que as equipas acumulam na competição lhes confere um estatuto superior face aos clubes mais “novatos”, permitindo assim a esses emblemas mais históricos atingir valores de audiências bastante superiores. Relativamente à diferença, confirma-se novamente a tendência para a preferência dos jogos entre equipas de contextos diferentes, isto é, teoricamente desequilibrados.

#### **4.5.10 Classificação da época anterior**

A variável que define a classificação das equipas no último ano enquadra-se na categoria de desempenho e o seu estudo pretende abordar qual o impacto que esse prestígio/competitividade vindo da época transata traz às audiências. Na abordagem feita à variável auxiliar DifClaUAno, que estabelece a diferença classificativa das equipas em jogo na temporada anterior, não foi encontrada visualmente nenhuma correlação de relevo, detalhe que pode ser sintomático de uma indiferença dos telespetadores acerca desta particularidade.



*Figura 22 - Rat# por soma da classificação no último ano*

Analisando a variável que soma as classificações de ambas as equipas, SomaClaUAno, já se estabelecem algumas hipóteses resultantes da sua correlação com as audiências. Ao somar as posições das duas equipas, esta variável será tanto menor quanto melhor posicionadas estas equipas tiverem ficado no ano anterior, o que tradicionalmente equivale a um jogo com intervenientes mais fortes e com maiores ambições competitivas. Da análise à Figura 22 é notória a preferência dos consumidores pelos jogos com menor

SomaClaUano, pois os únicos encontros que se conseguem aproximar e superar das audiências mais elevadas estão reunidos entre a SomaClaUano de 3 a 24.

No outro extremo de análise verifica-se o inverso. Os jogos com SomaClaUano maior ou igual a 25 (195 casos) têm um *rating* médio de 19 mil. Neste tipo de jogos, entre equipas mais mal classificadas na época anterior, são apenas 4 os jogos, entre 195 casos, que superam a fasquia dos 50 mil telespetadores. Entre esses encontros destaca-se o Portimonense x Gil Vicente (3/6/2020), jogo que marcou o regresso da Primeira Liga pós paragem devido à pandemia de Covid-19 e o Gil Vicente x Famalicão (9/6/2020), encontro que permitiu ao Famalicão alcançar o Sporting na 4ª posição, cimentando assim a sua época sensação de regresso à Primeira Liga.

#### **4.5.11 Valor de mercado dos plantéis**

A valorização económica dos plantéis é um dos indicadores que ajuda no papel de identificar quais as equipas mais bem apetrechadas e conseqüentemente favoritas à vitória. Da relação entre esta variável de cada uma das equipas surgem dois novos indicadores DifVM e SomaVM que dizem, respetivamente, respeito à diferença e à soma do valor de mercado das duas equipas que jogam entre si. Começando pela DifVM, os jogos com maior equilíbrio entre o valor de mercado não se refletiram genericamente em maiores audiências, excetuando os encontros que envolveram Benfica, Sporting e Porto. No conjunto destes jogos, com uma DifVM entre os 0 e 100 milhões de euros estão representados 78% dos casos, com uma média de 27,1 mil de *rating*. À medida que se analisam as outras faixas de valores, os valores de audiências médias vão subindo. Com uma diferença do valor de mercado das duas equipas entre DifVM entre 100 e 200 milhões de euros, existe uma média de 144,3 mil de *rating* e se essa diferença for DifVM superior a 200 milhões de euros, uma média de 162,9 mil de *rating*.

Analisando a variável que representa o total do valor de mercado das duas equipas em jogo, SomaVM, parece ser evidente de que quanto menor for o valor de mercado das equipas menor será a capacidade de o jogo em questão gerar uma audiência significativa. Nos jogos em que essa variável possui um valor inferior a 150 milhões de euros, a audiência média é de 21,7 mil telespetadores, um valor muito baixo quando comparado aos valores obtidos entre os 160 e os 250 milhões de euros, com 147,5 mil telespetadores de média e na faixa superior a 250 milhões de euros, com 184,1 mil telespetadores em média nos 94 jogos. Tendo em conta estes registos parece confirmar-se a preferência

acentuada dos fãs de futebol pelos encontros que reúnem os plantéis mais fortes. A expectativa nesses jogos é de maior espetacularidade, o que os torna mais atrativos como transmissão televisiva, apelando até a públicos que consomem futebol menos regularmente.

#### **4.5.12 Golos do jogo**

Os golos são a estatística base do futebol. São eles a parte principal do espetáculo, que dita o ritmo do jogo e decide quem ganha ou perde. É por eles que um telespetador se senta no sofá para assistir a um jogo, para ver qual equipa consegue marcar um número de golos superior ao do adversário. Associado a esse momento de euforia, que é um golo, é fulcral perceber qual o impacto que estes momentos têm na audiência de um jogo, partindo-se do pressuposto inicial que um adepto quer ver quantos mais golos melhor, pois os mesmos estão associados a espetacularidade e são os golos que traçam a linha entre os vencedores e os perdedores.

Analisando as audiências médias dos jogos, consoante o número de golos marcado em cada um deles, regista-se uma tendência de subida nas audiências quanto maior for o número de golos, conforme demonstra a Tabela 3. Entre os jogos em que não há golos, até àqueles em que se registam 7, o *rating* médio vai crescendo. A única quebra na audiência média encontra-se entre os 7 e 8 golos, mas convém ter em conta que nesses casos são médias feitas com muito poucas ocorrências.

<b>Nº de golos</b>	<b>Jogos</b>	<b>Rat# médio</b>
0	60	36,6
1	177	47,7
2	186	54,7
3	162	55
4	112	68,4
5	58	72,3
6	29	73
7	7	92,9
8	4	57,1
9	2	64,2
10	1	219,9

*Tabela 3 - Relação entre o número de golos e as audiências médias*

As audiências médias mais baixas registam-se nos jogos com 0 golos (36,6 rat#, em 60 casos), 1 golo (47,7 rat#, em 177 casos) e 2 golos (54,7 rat#, em 186 jogos). Como os jogos com um número elevado de golos são mais raros e apresentam poucas ocorrências,

foram agrupados numa nova categoria, de forma a que a audiência desses jogos seja contextualizada de melhor maneira. Essa nova categoria inclui os encontros com 5 golos ou mais e apresenta 101 casos, que atingem uma média de 74,6 mil telespetadores. A audiência destes jogos situa-se acima dos valores médios de rat# no total da base de dados (56,3 mil).

Em certa medida os dados parecem corroborar a hipótese inicial, de que os jogos mais apetecíveis para os espetadores são aqueles em que se marcam mais golos, apesar de quando a quantidade de golos é muito acima do normal não se conseguirem tirar muitas conclusões diretas, devido à escassez de ocorrências.

O impacto dos golos nas audiências é uma reflexão importante para as diversas partes do espetáculo futebolístico, pois ao se comprovar que um jogo por ter maior número de golos atrai mais telespetadores devem ser tomadas medidas para fomentar esse acontecimento, como eventualmente o aumento do tempo útil de jogo, o aumento da dimensão da baliza, as condições para a prática do desporto ou o recrutamento e formação de treinadores e jogadores.

#### **4.5.13 Capacidade dos estádios**

O estádio em que se disputa um jogo pode ter muita influência no desenrolar dos acontecimentos do mesmo. Tendo isso em consideração decidiu-se incluir uma variável no modelo que representasse qualitativamente as condições de um estádio e a maneira escolhida foi a capacidade do mesmo, apesar das limitações que esse valor por si só acarreta. De um estádio com maior capacidade espera-se que possibilite melhores condições para a realização televisiva do jogo, as diversas infraestruturas estejam mais bem cuidadas (como é o caso do estado de conservação da relva), consiga captar mais público presencialmente (o que poderá ser um fator que aumenta a emoção com o decorrer do jogo) e que lá joguem as equipas mais bem apetrechadas, sendo assim expectável audiências mais significativas consoante o número de lugares do estádio seja maior. Os dados da lotação foram analisados em faixas com a amplitude de 10 mil lugares, excetuando os jogos nos estádios que tinham mais de 40 mil lugares. Como só existem 3 estádios em Portugal com mais de 40 mil lugares, que é o caso dos recintos dos “3 grandes”, os jogos lá disputados foram todos agrupados na mesma faixa.

Nas 3 primeiras faixas de valores (0-10 mil, 10-20 mil e 20-30 mil), as audiências médias variam entre 34,8 mil, 40,7 mil e 48 mil telespectadores, respetivamente. A tendência de

crescimento faixa a faixa mantém-se para os jogos disputados em recintos que possuem entre 30 e 40 mil lugares, onde se regista uma média de 50,7 mil. No entanto, os valores alcançados pela classe que reúne os jogos disputados nos estádios do Benfica, Sporting e Porto é consideravelmente superior, superando os 172 mil telespetadores por jogo.

Estes dados demonstram que o grande diferencial de audiências surge em jogos nos maiores estádios, pelo facto de estes mesmos estádios albergarem, por norma, encontros com as equipas tradicionalmente mais fortes e com maior legião de adeptos.

#### **4.5.14 Canais premium**

Sendo o futebol não só o desporto, mas o conteúdo televisivo que mais gera audiências em muitos países, como é o caso de Portugal, a sua rentabilização através de canais premium foi algo que aconteceu com naturalidade. Nas épocas em análise, para acompanhar todos os jogos mais relevantes dos campeonatos portugueses, os telespetadores tinham de subscrever 2 diferentes conjuntos de canais premium: a BTV e SportTV. Associado a isso e para verem jogos em canais como o Porto Canal ou a SportTV + é necessária a ter um contrato com uma operadora de telecomunicação para ter o serviço de televisão por cabo. A SportTV era a que apresentava o pacote mais caro, com um preço mínimo de cerca de 24 euros mensais. A BTV optou por um preço a rondar os 10 euros. No caso da BTV, tratando-se do canal de um clube, o Benfica, apenas transmite jogos dessa mesma equipa, todos eles disputados com o Benfica e o Benfica B (Segunda Liga) na condição de visitado. Como demonstrado em diversas análises anteriores o Benfica é o um dos clubes que alcança maiores audiências televisivas e isso estende-se à televisão do clube, pois é a mesma que é responsável pelos seus jogos caseiros, algo único em Portugal. Estas partidas alcançam sempre um *rating* superior a 150 mil espetadores, algo que já era expectável pela análise feita à variável CapEstadio. A relação entre jogos emitidos pela BTV e encontros disputados no Estádio da Luz (com cerca de 65 mil lugares) é quase perfeita, o que torna a variável CapEstadio e PreçoCanal muito semelhantes na representação de alguns jogos.

A SportTV está responsável pelos restantes jogos e as audiências dos mesmos percorrem quase toda a gama de valores, oscilando entre 0,5 mil (Estoril x Académica – 18/1/2020) e os 351,6 mil telespetadores (Porto x Benfica – 2/3/2019). Existem ainda alguns jogos que são transmitidos no Porto Canal e na SportTV +, canais que não têm o estatuto de premium, pois têm o acesso aberto a todos os clientes com um pacote de TV por cabo

normal. A primeira estação emite as partidas da equipa B do Porto quando a mesma joga em casa, enquanto a SportTV + foi o canal escolhido pela SportTV a partir da época de 2019/2020 para aumentar a cobertura da Segunda Liga e fazê-la chegar a mais pessoas, transmitindo alguns encontros no seu canal de “livre acesso”. Neste conjunto de dados a BTV está representada em 62 jogos e os 5 canais pagos da SportTV são responsáveis pelo acompanhamento de 700 encontros (mais de 11 vezes mais). No entanto, nos 10 jogos com maior audiência (acima de 270 mil telespetadores) existe um equilíbrio com 5 pertencentes a cada um destes dois canais.

Para operacionalização e análise de uma variável que caracterize a visualização por tipo de canal, dividiram-se os canais em 3 faixas consoante o seu preço. A 1ª faixa inclui os canais com um custo de 0 euros, ou seja, aqueles que emitem jogos, mas não são premium. A faixa dos 9,90 euros que é destinada à BTV e por fim a faixa dos 23,99, alusiva aos 5 canais SportTV que são *premium*, conforme ilustra a Tabela 4.

<b>Canais</b>	<b>Preço</b>	<b>Nº de jogos</b>	<b>Rat# médio</b>
<b>Porto Canal e SportTV +</b>	0,00 €	36	24,4
<b>BTV</b>	9,90 €	62	133
<b>SportTV (1,2,3,4 e 5)</b>	23,99 €	700	71,1

*Tabela 4 - Audiências médias por preço de subscrição de canal*

Apesar de terem o acesso mais facilitado, pela não obrigação de um pagamento mensal extra, os jogos emitidos nos canais que não são *premium* não se destacaram dos restantes e atingem uma audiência média de 24,4 mil telespetadores, a mais baixa dos tipos em análise. Convém ter em conta que associado aos modestos registos destes 2 canais (Porto Canal e SportTV +) está o facto de os mesmos apenas transmitirem jogos da Segunda Liga. Em contraponto, a BTV é o canal que regista as melhores audiências médias (133 mil telespetadores) por jogo, sabendo que para isso contribui o facto de transmitir em exclusivo os jogos caseiros da sua equipa principal, mas também da formação B. A média dos 5 canais SportTV situa-se nos 71,1 mil telespetadores, ou seja, acima de média global da base de dados, mas muito aquém dos registos da BTV.

Os campeonatos profissionais de futebol em Portugal têm sido nos últimos anos um exclusivo dos canais *premium* e os consumidores já não estão acostumados a ter jogos em sinal aberto. No entanto, se esta variável for considerada significativa poderá contribuir para uma reflexão acerca do preço dos canais e conseqüente democratização do acesso ao futebol por parte da população.

#### 4.5.15 Jogos em simultâneo

A análise às audiências dos jogos, consoante a quantidade de partidas diferentes a decorrer em simultâneo, é fulcral. Ela poderá ajudar as partes interessadas sobre eventuais sobreposições de transmissões televisivas de jogos na tomada de decisões da sua calendarização. Estando a decorrer mais do que um jogo na mesma faixa horária, com características semelhantes (da mesma competição, por exemplo), é expectável que a audiência de cada um deles seja afetada, devido à fragmentação do público pelos diversos ‘focos de interesse’. As entidades que calendarizam o futebol em Portugal já mostram preocupações nesse sentido e o reflexo disso é o facto de 72% dos jogos analisados terem decorrido sem nenhum outro em simultâneo. Considerando aqueles que tiveram entre 0 e 1 jogos em paralelo da mesma competição essa percentagem sobe para 93%.

Dos dados analisados é possível constatar uma diminuição das audiências médias por jogo, à medida que o número de jogos em simultâneo é maior. Essa situação é visível na Figura 23 e na linha de regressão linear que aí se representa entre estas duas variáveis. Como os interessados em consumir futebol são, para todos os jogos quase sempre os mesmos, acontece uma diminuição das audiências médias e máximas, quando há mais do que uma partida, em determinada hora.

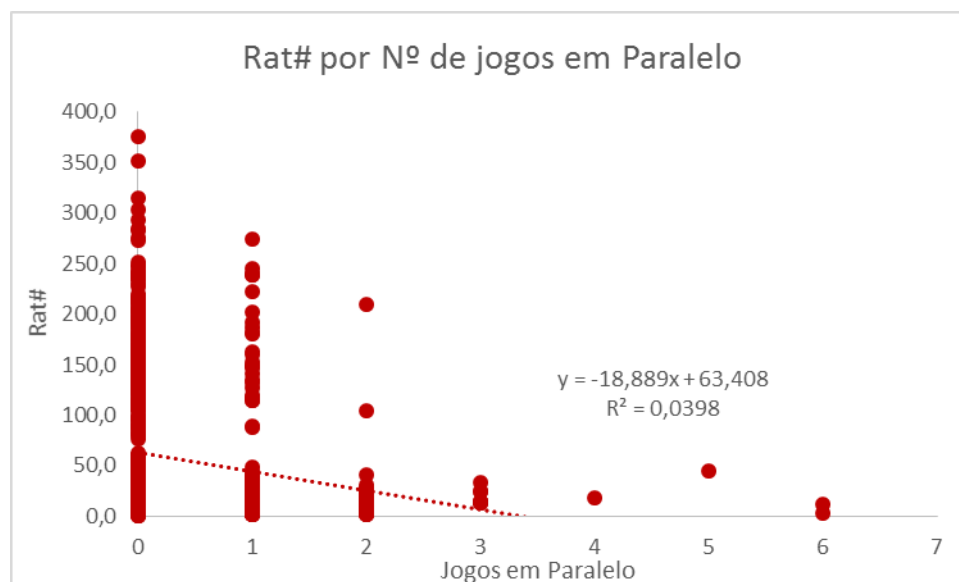


Figura 23 - Rat# por número de jogos em paralelo

Os jogos que decorrem em faixas horárias mais sobrecarregadas, ou seja, com mais três ou mais encontros a serem transmitidos na mesma altura, concentram-se quase exclusivamente entre os 0 e 50 mil de *rating* e são todos referentes à 2ª Liga. O mesmo padrão de comportamento verifica-se nos encontros onde a variável “número de jogos em

paralelo” assume o valor 2. Com 2 encontros, mas a fugir a esta “regra” encontram-se o Benfica x Portimonense (30/10/2019 – 209,8 rat#) e o Paços de Ferreira x Sporting (31/10/2019 – 104,1 rat#). O teto de audiências fixado nos 50 mil telespetadores poderá ser um alerta de qual o limite de público alcançável com determinado número de jogos transmitidos ao mesmo tempo e talvez seja essa a razão para o facto de ser uma opção de marcação tão rara, principalmente na Primeira Liga.

Para os jogos disputados com apenas um outro encontro em simultâneo, o limite máximo de *rating* é de 275 mil telespetadores, não existindo nenhum jogo que conseguisse ultrapassar essa barreira de audiências, neste conjunto de dados. A maior parte da amostra é de jogos sem nenhuma outra partida transmitida em simultâneo, o que limita um pouco as conclusões ao nível das audiências mais residuais.

Conclui-se então que o número de jogos transmitidos em paralelo poderá afetar as audiências televisivas dos mesmos. Este aspeto pode já ser tido em conta pelas partes interessadas responsáveis pela calendarização, pois na grande parte dos casos nota-se a distribuição da oferta programática pela grelha televisiva, fugindo, sempre que possível à sobreposição de jogos.

## **5 PREPARAÇÃO DOS DADOS**

### **5.1 Seleção dos dados**

O conjunto de dados recolhidos tem diversas fontes, proporcionando o aparecimento de muitos atributos, representados nas colunas. Desse conjunto de atributos foram excluídos alguns que apenas serviam como auxílio na navegação da base de dados e/ou identificação do registo/jogo em causa, tais como: a identificação do nome da competição (Competição), o nome dos clubes envolvidos (Equi1 e Equi2), o dia, mês, ano e época futebolística em que o jogo é disputado (Data, Mês, Ano e Época), a designação nominal do canal televisivo (Canal). Os dados de audiências inicialmente recolhidos foram o *rating* (em milhares e em percentagem) e o share televisivo. De forma a tornar o projeto mais objetivo, foi apenas escolhido o *rating* em milhares (Rat#) como representação das audiências, devido ao facto de os outros valores (Rat% e Shr%) serem, maioritariamente, valores de percentagem baixos, portanto, não permitindo uma identificação imediata do número pessoas alcançadas pela transmissão do jogo.

Para além dos atributos que foram excluídos, houve outros que foram criados, conforme explicitado na subsecção 5.3 - Construção dos dados.

Em termos de registos, ou seja, as linhas da base de dados, não foi feita qualquer seleção, ou seja, todos os jogos recolhidos inicialmente permaneceram até ao final do processo.

### **5.2 Limpeza dos dados**

Para esta etapa não foi necessário qualquer tipo de intervenção. Das fontes onde foram recolhidos os dados não foi detetado qualquer problema relacionado com a qualidade dos mesmos, nem teve de ser realizada nenhuma etapa de correção para o posterior tratamento dos mesmos.

### **5.3 Construção dos dados**

Neste projeto foram estabelecidos conceitos com um certo grau de abstração e por isso difíceis de representar através de um atributo. Daí surgiu a necessidade de construir novos atributos que oferecessem um significado aproximado a esses mesmos conceitos.

Tendo em conta a informação disponível é impossível chegar diretamente ao poder de compra dos adeptos de uma equipa e por essa razão associou-se, a cada equipa, o poder de compra dos habitantes do concelho a que a mesma pertence. Por não existir um valor

mais preciso, foi feita esta aproximação, seguindo a lógica de que as pessoas de um certo município tendencialmente apoiarão os principais clubes do mesmo.

Outro dos atributos em que isto se verifica é a base de adeptos de cada equipa onde, devido ao facto de os clubes contabilizarem os sócios de maneiras diferentes e fazerem a contagem dos mesmos em períodos não coincidentes, a população de cada município foi o valor encontrado para representar a base de apoio e de potenciais telespetadores de cada equipa.

Outros novos atributos construídos dizem respeito às diferenças e somas dos atributos referentes a cada equipa. Os valores que uma equipa tem em cada variável necessitam de “confronto” com os da outra equipa, para que possa haver uma adequação para cada jogo, e para isso foram construídos atributos auxiliares de diferença e soma, em relação a esses mesmos valores. Esses dados auxiliares estão presentes nas variáveis da distância pontual, poder de compra, população, participações e antiguidade na competição, classificação da época anterior e valor de mercado do plantel, já referidos anteriormente neste relatório. Trabalhando dados das duas equipas agrupados consegue-se que os mesmos digam respeito a toda a unidade de análise, a transmissão de jogo, e não apenas a uma só equipa.

Por fim, foram feitas transformações de dados para o atributo referente ao dia da semana e do horário de início do jogo. No dia da semana estava associado um valor nominal (segunda-feira, terça-feira, ...) e, para efeitos de modelação, essa variável deu origem a 7 novas variáveis do tipo booleana, uma alusiva a cada um dos dias da semana. A título de exemplo: se um jogo foi disputado a uma quarta-feira, a variável alusiva a esse dia vai ter o valor 1 e nos outros 6 dias da semana terá o valor 0. A alteração relacionada com o horário também teve como objetivo facilitar o processo de modelação e consistiu em transformar o horário inicial do jogo apenas no valor da hora, considerando que o valor dos minutos não acrescentaria valor ao processo. Os jogos iniciados às 18:15, 18:30, 18:45 têm todos como valor de hora de início 18, estendendo-se este processo a todas as faixas horárias.

#### **5.4 Integração dos dados**

A base de dados tem 6 grandes fontes: PORDATA, Transfermarkt, INE, CAEM/GFK, Zerozero e Google Maps, conforme demonstra a Figura 24. A integração destas fontes uma com as outras é sempre feita tendo por base a unidade de análise que é a transmissão de jogo.

Inicialmente foi feito o levantamento do total de jogos para as épocas e competições analisadas, sendo para isso usado o Zerozero. Delas foi possível extrair a informação relativa à competição em causa, às equipas em jogo, ao dia e hora do jogo, à distância pontual das equipas, à localização e capacidade do estádio, ao histórico de participações, à antiguidade e classificação no ano anterior de cada equipa, ao número de golos do jogo e ainda ao canal em que o mesmo é transmitido.



*Figura 24 - Processo de integração de dados*

Posteriormente e tendo por base a informação relativa à morada dos dois clubes é calculada a distância entre as mesmas, com o auxílio do Google Maps. Tendo a morada de cada clube como base, foram recolhidos os dados referentes ao poder de compra e população residente dos municípios a que as mesmas pertencem, utilizando para isso o PORDATA e o INE, respetivamente.

A recolha de dados no Transfermarkt serviu para aferir o valor de mercado dos plantéis que se defrontam em cada encontro. Para cada uma das épocas em estudo foi extraída do Zerozero a lista de equipas participantes em cada competição, servindo essa lista como base para o apuramento de valores extraídos do portal TransferMarkt (ver ANEXO II). Depois de dessa recolha todas as equipas ficaram com um valor de mercado associado.

Os últimos dados a serem recolhidos foram provenientes da CAEM/GFK e são referentes às audiências televisivas de cada jogo, tendo esse levantamento sido feito com base nos dados do Zerozero, referentes à hora e canal da transmissão em causa.

## **5.5 Formatação dos dados**

As tarefas compreendidas na formatação de dados não foram necessárias neste projeto pois os dados foram recolhidos na forma prevista inicialmente. As únicas modificações que ocorreram são as resultantes do processo explicitado no subcapítulo 5.3 - Construção dos dados.

## **6 MODELAÇÃO**

A fase de modelação começou com a escolha da técnica mais adequada a utilizar. Do rol de possibilidades existentes, a escolha recaiu na técnica da regressão, dado que esta técnica “fornece” uma função matemática que descreve a relação entre uma variável a explicar e as várias variáveis explicativas. O objetivo principal deste projeto passa por encontrar atributos que justifiquem as audiências televisivas de um jogo de futebol e perceber qual o impacto que esses mesmos atributos têm sobre essa variável. Utilizando a regressão esperou-se que o objetivo de chegar a um modelo preditivo simplificado fosse alcançado, tendo como base valores históricos de audiências de duas épocas desportivas recentes e os valores característicos de um jogo de futebol a eles associado.

### **6.1 Seleção da técnica de modelação**

Como descrito no ponto anterior, a técnica escolhida para representar o modelo preditivo simplificado foi a regressão linear múltipla. A variável que se pretende descrever é o *rating* televisivo obtido por um jogo de futebol, com base em variáveis relacionadas com o mesmo, como a hora, o valor de mercado das equipas, o número de golos, entre outras.

A análise da regressão focou-se essencialmente em 3 indicadores. O primeiro valor foi o coeficiente de determinação,  $R^2$ , que demonstra o grau de explicação do modelo. O modelo tem mais poder explicativo quanto mais perto este coeficiente estiver do valor 1. Para escolher quais as variáveis têm ou não impacto nos valores do *rating* é também importante olhar para o valor-p de cada uma delas.

O valor-p (probabilidade de significância), é a probabilidade de se obter uma estatística de teste igual ou mais extrema que aquela observada em uma amostra, sob a hipótese nula, ou seja, a hipótese de a referida variável não ser significativa no modelo. Por exemplo, em testes de hipótese, pode-se rejeitar a hipótese nula a 5% caso o valor-p seja menor que 5%. Assim, uma outra interpretação para o valor-p, é que este é o menor nível de significância com que se rejeitaria a hipótese nula. Em termos gerais, um valor-p pequeno significa que a probabilidade de obter um valor da estatística de teste como o observado é muito improvável, levando assim à rejeição da hipótese nula. Os níveis de significância mais usuais são 1%, 5% e 10%. Nos modelos produzidos neste projeto serão apenas consideradas variáveis cujo valor-p seja inferior ao nível de confiança estabelecido de 0,10 (10%).

Outros indicadores muito importantes para a análise dos modelos obtidos pela regressão, são os coeficientes associados a cada uma das variáveis, permitindo assim saber em que sentido e proporção essa variável influencia as audiências.

## **6.2 Geração do desenho de testes**

De forma a analisar a qualidade dos modelos apresentados foi realizado um procedimento para testar a eficácia dos mesmos. Neste projeto a base de dados total foi dividida em dois subconjuntos: um para treinar os modelos e outro para o testar a qualidade dos mesmos.

Em relação à metodologia de divisão foi decidido que 2/3 dos jogos, ou seja 532, seriam destinados ao conjunto de treino e os restantes 266 serviriam para testar os modelos obtidos. O facto de a base de dados possuir 798 jogos, que é um número múltiplo de 3, facilitou o processo de divisão. No que diz respeito aos critérios de atribuição de cada jogo a um ou a outro modelo, começou-se por ordenar a base de dados por ordem cronológica, misturando assim os jogos da Primeira e Segunda Liga, e de seguida associar alternadamente dois jogos ao subconjunto de treino e outro ao subconjunto de testes, até perfazer os 798 encontros.

Depois de encontrados os modelos mais adequados, os mesmos passarão por um teste de fiabilidade que vai permitir uma comparação entre eles, até à escolha do mais indicado. Para a realização desses testes serão considerados dois indicadores: o coeficiente de determinação, conhecido por  $R^2$ , que expressa a quantidade da variância dos dados que é explicada pelo modelo linear e a raiz quadrada da média do quadrado dos erros (conhecida pela sigla RMSE, do inglês *Root Mean Square Error*), com cada erro medido como a diferença entre cada valor real da audiência para cada jogo transmitido e o valor previsto pelo modelo para essa audiência. De um bom modelo espera-se que tenha o  $R^2$  o mais perto possível de 1 e o RMSE o mais baixo possível.

## **6.3 Construção do modelo**

Depois de realizada a divisão da totalidade do conjunto de dados, entre conjunto de treino e de teste, iniciou-se a fase de construção do modelo, baseado na técnica de regressão. Devido ao grande número de variáveis exógenas e às limitações impostas pelo *software* utilizado, Microsoft Excel 2013, em que para calcular uma regressão apenas se podem utilizar 16 variáveis preditivas, separaram-se, numa primeira fase, as 43 variáveis numéricas explicativas do modelo, em 3 grupos, consoante o apresentado na Tabela 5. A

ordem escolhida seguiu a disposição inicial das inserções na base de dados, não tendo sido contemplado qualquer critério de escolha.

<b>Variáveis</b>		
<b>Modelo 1</b>	<b>Modelo 2</b>	<b>Modelo 3</b>
Divisão	DifPodComp	ClaUAnoEqui2
Hora	SomaPodComp	DifClaUAno
SegF	MedPondPodPop	SomaClaUAno
TerF	PopEqui1	VMEqui1
QuaF	PopEqui2	VMEqui2
QuiF	DifPop	DifVM
SexF	SomaPop	SomaVM
Sab	PartiEqui1	Golos
Dom	PartiEqui2	CapEstadio
DistAdv	DifParti	PreçoCanal
DistPontualEqui1	SomaParti	Paralelo
DistPontualEqui2	AntiCompEqui1	
SomaDist	AntiCompEqui2	
DifDist	DifAntiComp	
PodCompEqui1	SomaAntiComp	
PodCompEqui2	ClaUAnoEqui1	

*Tabela 5 - Distribuição das variáveis pelos 3 modelos iniciais de treino*

Nesta fase, foi importante perceber se as diferentes variáveis se revelaram significativas em cada um dos 3 modelos iniciais, tendo para isso considerado o valor-p de todas elas nestes modelos iniciais que começaram o processo. Após a construção dos 3 modelos iniciais acima (Tabela 5) apresentados foram encontradas 13 variáveis com valor-p inferior a 0,05, ou seja, consideradas significativas, presentes na Tabela 6, que serão a base do modelo final. As restantes variáveis foram descartadas

Variável	Valor-p
Divisão	4,42E-22
Hora	2,53E-02
DistAdv	1,98E-07
SomaDist	2,74E-05
DifDist	3,81E-25
PodCompEqui1	5,33E-42
PodCompEqui2	2,05E-35
SomaAntiComp	1,58E-29
SomaClaUAno	1,61E-02
SomaVM	2,96E-48
CapEstadio	8,40E-03
PreçoCanal	1,69E-14
Paralelo	8,54E-08

Tabela 6 - Variáveis significativas dos 3 modelos iniciais

Depois de terem sido encontradas estas 13 variáveis, o subconjunto de treino foi readaptado, para que no novo cálculo da regressão apenas seja considerado o valor destes 13 atributos e o *rating* dos 532 jogos em análise. Apenas com as variáveis significativas resultantes dos três modelos iniciais foi feita uma nova regressão da qual resultam os valores-p presentes na Tabela 7.

Variável	Valor-p	Legenda
Divisão	0,083	Valor-p superior a 0,10
Hora	0,171	valor-p entre 0,05 e 0,10
DistAdv	0,007	valor-p inferior a 0,05
SomaDist	0,493	Valor-p superior a 0,10
DifDist	0,066	valor-p entre 0,05 e 0,10
PodCompEqui1	0,031	valor-p inferior a 0,05
PodCompEqui2	2E-07	valor-p inferior a 0,05
SomaAntiComp	0,757	Valor-p superior a 0,10
SomaClaUAno	0,083	valor-p entre 0,05 e 0,10
SomaVM	2E-53	valor-p inferior a 0,05
CapEstadio	0,536	Valor-p superior a 0,10
PreçoCanal	3E-08	valor-p inferior a 0,05
Paralelo	5E-06	valor-p inferior a 0,05

Tabela 7 - Valores-p do modelo com as 13 variáveis inicialmente significativas

Com base nas propriedades destas variáveis propõem-se 3 modelos de regressão, que conjugam as 9 variáveis que preservam o estatuto de significativas aquando combinadas entre si.

O primeiro modelo proposto assenta numa regressão feita com as variáveis: *DistAdv*, *PodCompEqui1*, *PodCompEqui2*, *SomaVM*, *PreçoCanal* e *Paralelo* que são aquelas com o valor-p mais baixo e inferior a 0,05. De uma forma genérica, consoante as categorias dos atributos, estão representadas duas das três tipologias de dados de entrada considerados na Figura 11 (subsecção 4.3.1 - Procedimento de recolha de dados), que são os dados relativos aos jogos (preço do canal que transmite o jogo e o número de encontros em paralelo) e os elementos geográficos, económicos e demográficos, como a distância geográfica entre equipas, o poder de compra dos habitantes das suas localidades e o valor de mercado dos plantéis. Da regressão realizada com estas variáveis surge o seguinte modelo:

$$\begin{aligned} \mathbf{Rating} = & 15,29351 - 0,00389 \mathit{DistAdv} + 5,39076 \mathit{PodCompEqui1} \\ & + 9,382777 \mathit{PodCompEqui2} + 0,544125 \mathit{SomaVM} \\ & - 1,01174 \mathit{PreçoCanal} - 6,50485 \mathit{Paralelo} \end{aligned}$$

Este modelo apresenta um  $R^2 = 0,8927$  e com a utilização desta regressão nos 266 jogos de teste chega-se a um  $RMSE = 21,7059$ .

#### **6.4 Revisão do modelo**

Nesta fase e como alternativas ao primeiro modelo foram gerados mais dois modelos, com algumas variações à estrutura do mesmo.

O segundo modelo proposto resulta de uma derivação do primeiro, utilizando as 4 variáveis desse mesmo modelo com o valor-p inferior a 0,05, que são *PodCompEqui2*, *SomaVM*, *PreçoCanal* e *Paralelo*. Este modelo surge da oportunidade de apresentar uma versão mais simplificada, isto é, com menor número de variáveis, tentando não perder capacidade de previsão, sendo escolhidas as variáveis consideradas mais significativas, por terem o valor-p mais reduzido. A equação do segundo modelo é representada por:

$$\begin{aligned} \mathbf{Rating} = & 25,91598 + 8,140901 \mathit{PodCompEqui2} + 0,556076 \mathit{SomaVM} \\ & - 1,24724 \mathit{PreçoCanal} - 7,07296 \mathit{Paralelo} \end{aligned}$$

O  $R^2$  deste modelo é de 0,8913 e após calcular os valores de *rating* para os dados de teste o  $RMSE$  assume um valor aproximado ao do primeiro modelo, ou seja 21,8504.

Ao contemplar apenas a variável relativa ao poder de compra referente a um dos concelhos, o segundo modelo pode ficar desequilibrado na sua abordagem prática. Por esse motivo foi gerado o modelo 2' que, às 4 variáveis do segundo modelo adiciona

também o poder de compra da equipa que não estava representado, o PodCompEqui1. A equação do modelo 2' é a seguinte:

$$\begin{aligned} \text{Rating} = & 15,13797 + 5,2 \text{ PodCompEqui1} + 9,46014 \text{ PodCompEqui2} \\ & + 0,54564 \text{ SomaVM} - 1,0797 \text{ PreçoCanal} - 6,9355 \text{ Paralelo} \end{aligned}$$

Este modelo apresenta um  $R^2$  igual a 0,892 e um RMSE de 21,7808, um valor ligeiramente mais baixo do que dos dois modelos anteriores.

Por fim, o terceiro modelo proposto contempla uma regressão feita com todas as 9 variáveis que preservam o estatuto de significativas, apresentando um valor-p inferior a 10%, tendo assim uma proposta teoricamente mais explicativa. A equação que define este modelo é a seguinte:

$$\begin{aligned} \text{Rating} = & 22,99029 - 11,4908 \text{ Divisão} - 0,00642 \text{ DistAdv} + 0,194821 \text{ DifDist} \\ & + 5,8,817987 \text{ PodCompEqui1} + 13,99489 \text{ PodCompEqui2} \\ & + 0,260519 \text{ SomaClaUAno} + 0,5149 \text{ SomaVM} \\ & - 1,29215 \text{ PreçoCanal} - 6,7651 \text{ Paralelo} \end{aligned}$$

Apesar de este modelo contemplar mais 3 variáveis que o primeiro, mais 5 variáveis do que o segundo e mais 4 do que no modelo 2', a explicação dos dados que o mesmo oferece não é muito superior às registadas anteriormente:  $R^2 = 0,8979$ . Também a raiz quadrada da média dos erros apresenta um valor aproximado ao dos outros dois modelos, com um valor de 20,9256.

Para a fase de revisão do modelo são ainda comparados os quatro modelos obtidos, através de indicadores da sua fiabilidade como o  $R^2$ , o  $R^2$  ajustado e o RMSE, calculados tanto para o conjunto de dados de treino como de teste.

Os quatro modelos são bastante semelhantes entre si, tanto a nível de variáveis utilizadas, como nos valores dos indicadores de fiabilidade, como demonstra a Tabela 8, que resume os principais indicadores dos quatro modelos apresentados. Nesta tabela também se podem ver valores de RMSE muito semelhantes, ligeiramente mais baixos para o conjunto de dados de treino, tal como seria de esperar.

Modelo	$R^2$ treino	$R^2$ Ajust. treino	RMSE treino	$R^2$ teste	$R^2$ Ajust. teste	RMSE teste	Variáveis
1	0,8927	0,8917	21,7059	0,8903	0,8882	23,5683	6
2	0,8913	0,8907	21,8504	0,8901	0,8889	23,5696	4
2'	0,8920	0,8912	21,7808	0,8919	0,8903	23,4030	5
3	0,8979	0,8964	20,9256	0,8935	0,8902	23,2011	9

Tabela 8 – Resumo dos principais indicadores dos quatro modelos

Por um lado, o modelo 3 é aquele que apresenta os valores de  $R^2$  e  $R^2$  ajustado mais elevados, bem como o RMSE mais baixo. Por outro lado, o modelo 2 e o modelo 2' são aqueles que têm menor número de variáveis, respetivamente cinco e seis. O princípio da Navalha de Ockham estipula que, havendo várias formas viáveis para representar e justificar um conjunto de factos se deve escolher aquela que for a mais simples (Spade & Panaccio, 2002).

Assim, considerando o princípio da Navalha de Ockham, dever-se-á optar pelo modelo 2 ou pelo modelo 2', porque são os modelos mais simples e com valores de  $R^2$ ,  $R^2$  ajustado e RMSE muito bons e também aproximados aos dos restantes modelos. De entre estes dois, dá-se a preferência à utilização do modelo 2' mais equilibrado, já que considera o “poder de compra” dos adeptos das duas equipas e não apenas de uma.

Em resumo, o modelo 2', com apenas 5 variáveis exógenas e ainda com um poder explicativo muito significativo, é considerado o mais recomendado e eficaz nas tarefas de explicar e prever as audiências televisivas para os jogos de televisão do futebol português.

## **7 DISCUSSÃO DOS RESULTADOS**

### **7.1 Avaliação dos resultados**

Do modelo escolhido observa-se que existem variáveis preditivas de duas das três categorias consideradas na Figura 11, referentes à transmissão de jogo que é a unidade de análise, e aos fatores geográficos, económicos e demográficos.

Neste modelo, três das cinco variáveis preditivas têm um cariz económico, geográfico ou demográfico associado. A equação contempla o poder de compra dos habitantes do concelho a que os dois clubes em confronto pertencem e também o valor de mercado dos seus plantéis. O relacionamento destas variáveis com as audiências televisivas é o esperado, tendo em conta as suas características, pois quanto maior o poder de compra e o valor de mercado dos plantéis, maiores são as audiências prevista. Fica assim comprovada a preferência dos telespetadores pelos jogos das equipas cujo plantel é mais valioso e também a importância de os adeptos terem uma maior disponibilidade financeira, para conseguirem assistir aos jogos pela televisão.

As variáveis significativas de características do jogo no modelo escolhido são o canal em que o jogo é transmitido, representado pelo preço da sua subscrição mensal e o número de encontros que decorrem em paralelo. O modelo valida o esperado inicialmente, ou seja, um encontro tem mais audiências quanto mais barato for o preço do canal em que “passa” e quantos menos jogos decorrerem em simultâneo.

### **7.2 Implementação prática dos modelos**

Os fatores considerados significativos no modelo selecionado serão extremamente importantes para todas as partes interessadas, nomeadamente, a entidade organizadora dos campeonatos, as televisões que atuam no mercado, os clubes participantes, os adeptos, os jogadores, os treinadores e patrocinadores.

A implementação prática dos modelos apresentados pode ser feita por qualquer uma das partes interessadas com relativa facilidade, pois todas as variáveis consideradas neste estudo são de acesso gratuito e estão disponíveis para consulta *online*.

Algumas das partes interessadas, como a televisão, os clubes, a direção das competições ou agências publicitárias possuem ainda informação própria, que podem conjugar com os modelos propostos neste trabalho, criando novos modelos, de forma a atingir resultados mais específicos tendo em conta os seus interesses.

Numa perspetiva de implementação corrente dos modelos para previsão da audiência em cada jogo futuro, as bases de dados podem ser atualizadas com maior frequência e assim possibilitarem resultados mais satisfatórios.

Como este é um mercado que está em constante mudança, existe o risco de os modelos perderem eficácia com fatores como a entrada no mercado de novos canais televisivos, o surgimento de novas formas de consumo, políticas de captação de público ao estádio, nova legislação, alteração na estrutura diretiva dos clubes ou até interesse gerado por outros desportos ou campeonatos concorrentes.

### **7.3 Fatores não significativos**

Do conjunto de variáveis analisadas houve muitas que não foram consideradas significativas e por essa razão não constam no modelo final.

Em relação à calendarização não se encontrou um impacto relevante quanto aos dias da semana em que os jogos são disputados, nem quanto à hora a que se iniciam nas audiências televisivas. Numa primeira modelação as horas ainda se mostraram significativas, mas perderam esse estatuto quando conjugadas com todas as variáveis de valor-p inferior a 5%. Estas características relativas à marcação dos jogos podem não ter grande importância nas audiências televisivas, mas afetam outros assuntos relacionados com o público, como a presença no estádio.

Na variável que soma a distância pontual a que as equipas estão do topo da tabela, também não se encontrou significância, ao contrário da diferença entre esse valor nas duas equipas. Pode isto indicar que, para os telespetadores, o equilíbrio, ou falta dele, é mais relevante dentro do próprio jogo do que no contexto global da competição.

O comportamento das variáveis referentes ao poder de compra populacional revelou-se apenas significativo nos valores individuais de cada equipa. As variáveis da diferença do poder de compra e a soma desse indicador, para as duas equipas, não entraram neste modelo. Também, em relação aos dados demográficos, não foi encontrada significância que justificasse a presença de variáveis alusivas ao número de habitantes por concelho de cada equipa no modelo. O facto de este tipo de variáveis representar e caracterizar de certa forma as equipas pode ter influenciado a sua não presença no modelo, já que essa caracterização das mesmas é também feita por outras variáveis, tornando-as por isso inconsequentes e de pouca utilidade.

Os aspetos que contemplam o historial das equipas, como as participações na competição em causa, ou o número de épocas consecutivas em que cada clube nelas compete, também não foram consideradas no modelo final. A não presença destas variáveis pode dever-se ao diferente espectro de valores que assumem nos casos da 1ª e da 2ª Liga. Para estudos futuros seria pertinente avaliar o impacto destas variáveis, aplicadas apenas a uma só competição.

Nas variáveis que medem o desempenho desportivo, considerado no modelo como sendo a classificação do ano anterior das duas equipas, apenas a soma desse valor de cada uma das equipas consta no modelo 3, não figurando no modelo escolhido 2'.

Relativamente ao valor de mercado dos plantéis em jogo, só a soma dos valores das duas equipas integra o modelo 2', podendo-se explicar pelo facto da diferença de qualidade das equipas já ser apurada por outras variáveis, tirando assim valor ao atributo que estabelece a diferença entre a qualidade teórica dos plantéis.

A variável que pretendia representar a espetacularidade dos encontros, o número de golos, também não faz parte do modelo final. Quando um telespetador decide ver um jogo e lhe dá audiência, não sabe se o encontro vai ter muitos ou poucos golos e talvez esse seja um dos aspetos que “atira” esta variável para fora do modelo. Em trabalhos futuros, uma variável como os golos poderia ser estudada no aspeto de retenção de espetadores durante a maior parte da transmissão, consoante a sua cadência e altura do jogo a que surgem, de modo a validar melhor a sua importância.

Por fim, também a variável da capacidade do estádio se assume como um fator não significativo para os dados analisados. Provavelmente, à semelhança de algumas das variáveis que não integram o modelo, a capacidade do estádio não é escolhida como preditiva por já existirem outros atributos que “forneçam uma justificação” melhor ao dela.

#### **7.4 Fatores significativos ausentes do modelo final**

As variáveis que representam a divisão a que o jogo pertence, a distância geográfica entre os clubes, a classificação das mesmas, tanto na época atual como na época anterior foram consideradas significativas, mas não entram no modelo final escolhido (modelo 2'), pelas razões apresentadas no tópico 7.1 - Avaliação dos resultados.

O comportamento de todas essas variáveis que foram excluídas no modelo final é o esperado. No caso da divisão, os jogos da Primeira Liga geram mais interesse e têm conseqüentemente maiores audiências televisivas, sendo de esperar que quanto menor for o valor desta variável maior seja o número de telespetadores alcançados, pois esta variável tem a si associada um sinal negativo, diminuindo o valor do *rating* previsto à medida que passa de 1 para 2, ou seja, da 1ª para a 2ª Liga.

A variável que estabelece a distância geográfica entre os dois clubes em jogo é uma variável que pode influenciar vários comportamentos dos adeptos. A distância pode servir de catalisador de audiências quando se trata de confrontos entre equipas vizinhas, nos casos em que as distâncias “obrigam” os adeptos a terem de assistir ao jogo pela televisão, pela impossibilidade de irem presencialmente ao estádio. A variável também pode representar o fenómeno de confrontos entre equipas sem rivalidade histórica e que se mostrem desinteressantes para o espetador comum. No contexto dos dados analisados, a variável associada à distância assume um sinal negativo, o que significa que a cada quilómetro de distância entre as equipas, as audiências saem prejudicadas. Conclui-se assim que os telespetadores têm normalmente preferência por jogos entre equipas mais próximas.

No estudo feito, a classificação das equipas é medida através da distância pontual que as mesmas têm ao 1º classificado, para que o peso classificativo não fique apenas associado à posição na tabela, mas considerando o que essa mesma posição reflete. A variável que afere a distância pontual que as duas equipas têm entre elas foi considerada significativa, com sinal positivo. Fica assim genericamente demonstrado que, nos campeonatos em análise, existe uma preferência por partidas mais desequilibradas, quando se defrontam equipas com um desempenho muito díspar.

A classificação da época anterior é mais uma das medidas de desempenho e qualidade das equipas. As classificações do ano anterior representam não só a capacidade histórica das equipas, como também a expectativa para a época em andamento. Dentro das variáveis relacionadas com esta classificação destaca-se a soma da classificação de ambas as equipas, por ser a variável relacionada com a classificação anterior que é considerada significativa, apesar de não constar no modelo final. Os jogos com maior soma de classificação são teoricamente disputados entre as equipas mais fracas e por isso era de esperar que não fossem potenciadores de audiências. No entanto, o sinal e coeficiente desta mesma variável no modelo indica o inverso. A variável *SomaClAUAno* tem sinal

positivo, um comportamento considerado inesperado deste atributo e que se pode dever ao facto de o conjunto de dados, sob a qual incidiu a regressão, conter jogos da Primeira e Segunda Liga misturados, fazendo assim com que existam jogos com valores semelhantes de SomaClaUA<sub>no</sub>, mas referentes a competições de categoria e interesse gerado.

## **7.5 Poder de compra dos municípios**

O poder de compra dos municípios revelou-se uma das variáveis significativas do modelo selecionado (modelo 2').

Nesta investigação definiu-se o poder de compra dos municípios associados à localização de um clube como o indicador da capacidade aquisitiva dos adeptos dos mesmos. Para assistir aos jogos dos campeonatos nacionais de futebol em Portugal é necessário ter uma assinatura de televisão por cabo e, para a maioria dos jogos, ativar a subscrição de canais *premium*. Resumidamente, os jogos não são de acesso livre, sendo por isso necessário algum investimento mensal para que os interessados consumam futebol nacional pela televisão. Da equação do modelo constam as variáveis do poder de compra associados, tanto à equipa caseira, como à que joga na condição de visitante e as duas apresentam um sinal positivo, corroborando a ideia de que a qualidade do poder aquisitivo de determinada franja de adeptos vai influenciar as audiências das equipas que os mesmos apoiam.

Por cada unidade de poder de compra dos municípios da equipa visitada encontra-se o coeficiente 8,817987, valor que para a equipa visitante ascende a 13,99489. Na prática, a cada unidade acrescida no poder de compra da equipa visitada, espera-se que as audiências subam 8.818 telespetadores e 13.995 telespetadores quando se trata do poder de compra dos municípios da equipa forasteira. A diferença de 5,1769, isto é, 5.177 telespetadores no contexto em que é aplicado, entre um e outro coeficiente demonstra que, para as audiências televisivas, é mais impactante o poder de compra dos adeptos visitantes, aspeto que se pode justificar pelo facto dos adeptos caseiros, devido à proximidade geográfica, conseguirem assistir aos jogos presencialmente, ao contrário dos adeptos visitantes que, para se deslocarem ao estádio, necessitam de uma maior logística e investimento.

Embora com escalas diferenciadas a ter em conta, o poder de compra dos municípios é o atributo que tem maior coeficiente na equação do modelo, o que demonstra a importância do mesmo. Este aspeto pode levar a uma reflexão acerca do preço dos canais *premium* e

do investimento que um adepto necessita de fazer para poder ver os jogos dos campeonatos nacionais. Com medidas como a transmissão de jogos em sinal aberto, ou a redução do preço dos canais, o poder de compra não teria provavelmente tanto impacto e, conseqüentemente, o número de adeptos angariados aumentaria.

## **7.6 Valor de mercado dos plantéis**

O valor de mercado dos plantéis revelou-se uma das variáveis significativas do modelo selecionado (modelo 2').

Para representar a valia das equipas enquadrou-se, nos dados analisados, o valor de mercado dos jogadores que constituíam as mesmas. No plano teórico, os fãs de futebol assumem preferência por assistir aos encontros em que estão os melhores jogadores, por esperarem um espetáculo mais atrativo. No modelo escolhido entra a variável SomaVM, que agrupa o valor de mercado dos plantéis das duas equipas que disputam um jogo, validando-se assim a importância deste atributo para a explicação de audiências do conjunto de dados em estudo. Conforme era previsto, SomaVM é uma variável com sinal positivo no seu coeficiente, ou seja, quanto maior for o valor de mercado dos jogadores maiores são as audiências televisivas esperadas para esse jogo. O coeficiente associado a SomaVM é de 0,5149, o que se traduz num aumento de 515 telespetadores previstos a cada 1 milhão de euros de valor de mercado das equipas em campo. De um encontro onde a soma do valor de mercado das equipas seja de 125 milhões de euros espera-se que este fator contribua com um total de 64.363 telespetadores a mais no valor das audiências televisivas do jogo.

Apresentam-se os cálculos para o exemplo apresentado anteriormente:

$$Rat\# = \dots + 0,5149 \text{ SomaVM} + \dots$$

$$Rat\# = \dots + (0,5149 * 125) + \dots$$

$$Rat\# = \dots + 64,3625 + \dots$$

## **7.7 Canais premium**

O canal em que um jogo é transmitido é também uma das variáveis significativas do modelo selecionado (modelo 2').

À semelhança do poder de compra dos munícipes, que também consta no modelo e está discutido no subcapítulo 7.5, o preço de subscrição dos canais em que os jogos são

transmitidos é considerado significativo e impactante nos resultados de audiências. Esta variável é importante para todas as organizações envolvidas no jogo, especialmente para as televisões, pois oferece uma perspetiva do impacto que o preço do seu canal tem no nº de telespetadores que conseguem atingir. Na etapa anterior à exploração dos dados, esperava-se que quanto mais caro fosse o preço de aquisição do canal menores seriam as audiências dos seus jogos, pelo facto de à medida que o preço subisse, o número de pessoas com capacidade de aquisição para o subscrever fosse menor. A regressão efetuada vai ao encontro da hipótese inicial, pois a variável PreçoCanal assume um sinal negativo, com um coeficiente de -1.29215. Este valor significa que, no modelo apresentado, as audiências televisivas esperadas de um jogo sofrem uma quebra de 1.292 espetadores por cada euro a mais no custo de subscrição mensal do canal. Com este importante conhecimento caberá às partes interessadas e com capacidade de influência no preço dos canais premium optar pela estratégia que melhor se adegue aos seus interesses. Importa também lembrar que os canais com o preço igual a 0 euros não são obrigatoriamente de acesso livre, pois para que os telespetadores tenham acesso aos mesmos necessitam de ser assinantes de um pacote de televisão.

## **7.8 Jogos em simultâneo**

O número de jogos emitidos em simultâneo é a última das variáveis significativas do modelo selecionado (modelo 2').

Todos os mercados têm um espectro limitado de interessados e o futebol, apesar de ser consensualmente considerado o desporto-rei, não é exceção. Era importante validar se o facto de existirem outros jogos de futebol a decorrer em simultâneo tinha um prejuízo significativo nas audiências dos encontros, o que se confirmou como verdadeiro. A variável Paralelo tem um coeficiente de -6,7651, o que para o modelo em causa representa uma perda de 6.765 telespetadores num encontro por cada outro jogo da mesma competição que decorra em simultâneo.

O aspeto dos jogos em paralelo já é tido em conta pela Liga Portugal, entidade responsável pela organização da Primeira e Segunda Liga, pois é notória a preocupação em espalhar os jogos ao longo da jornada evitando sobreposições. Nos jogos sem adeptos presentes nos estádios, no período marcado pela Covid-19, a Liga optou também pela marcação regular de encontros em dias de semana diferentes dos tradicionais (6ª feira, sábado, domingo e 2ª feira), de modo a evitar “competir” com outras ligas de futebol

*Análise e previsão de audiências no mercado televisivo e de comunicação em Portugal*

internacional, demonstrando assim não só uma preocupação com a sobreposição de jogos alusivos à própria competição, mas também de campeonatos externos.

## **8 CONCLUSÃO**

Esta dissertação tinha com objetivo base entender e justificar as audiências televisivas de um jogo de futebol dos campeonatos portugueses, consoante as características desse mesmo encontro e das equipas que o disputam. O tema das audiências foi o escolhido por ser um fator com grande peso nas receitas financeiras das diversas partes interessadas no negócio do futebol.

Ao longo de todo o trabalho, uma das conclusões mais imediatas é o peso que os três clubes considerados grandes em Portugal, Benfica, Porto e Sporting, têm. Estes três conjuntos conseguem chegar a valores de audiências inalcançáveis por todos os restantes clubes. Ao contrário do que acontece noutros países, como Inglaterra, a grande maioria dos fãs de futebol simpatiza com um destes três, remetendo o seu clube local para segundo plano, o que proporciona enormes discrepâncias na importância de um jogo com e sem um dos “3 grandes” ao nível de audiências televisivas e, conseqüentemente, retorno gerado.

Relativamente aos fatores que influenciam as audiências de um encontro de futebol, destacaram-se o poder de compra dos munícipes da localidade das equipas, o valor de mercado dos jogadores em campo, o preço do canal em que o jogo é emitido e a quantidade de encontros da mesma competição a decorrer em paralelo. Todos estes atributos se revelaram significativos na tarefa de previsão e conseqüente explicação das audiências. Os atributos em causa são alusivos a diversas vertentes de um jogo e por isso são suscetíveis a contribuir para as diferentes partes interessadas que envolvem o futebol, desde os clubes, televisões, organizadores das ligas, patrocinadores, jogadores, treinadores e adeptos.

Como esta dissertação permitiu definir estas variáveis como sendo significativas e chegar a um coeficiente para as mesmas, a indústria do futebol pode trabalhar certos fatores de forma a otimizar as audiências televisivas. Teoricamente, no caso de um clube, as preocupações que o modelo obtido fornece já são do conhecimento do mesmo, pois confirma-se a ideia inicial de que quanto mais forte for o valor do plantel, mais atrativo vai ser para um telespetador ver essa equipa jogar. As televisões, com base no modelo proposto nesta dissertação, podem chegar a novas abordagens acerca do preço de subscrição do seu canal, pois ficou comprovado que a oscilação do mesmo tem impacto imediato no *rating* dos jogos. Os patrocinadores são uma das partes mais interessadas no

fenómeno das audiências e por isso terão uma visão mais macro de todas as variáveis, percebendo assim que aspetos lhes permitirão chegar com maior acutilância aos seus públicos e negociar contratos com mais informação útil. Também as organizações que gerem as ligas de futebol devem ter uma visão global acerca das diferentes influências nas audiências, pois esses atributos refletir-se-ão num menor ou maior ganho económico para as mesmas, o que impulsionará ou condicionará o nível competitivo das provas que organizam. Um dos aspetos mais imediatos surge do impacto que o número de jogos disputados em simultâneo tem, no número de telespetadores de cada jogo. Esse tema pode ser alvo de reflexão nas tarefas de calendarização, sem esquecer outros aspetos como a distância geográfica entre as equipas, ou a classificação competitiva das mesmas, que apesar de não pertencerem ao modelo final também foram consideradas significativas.

Para a Academia, este projeto fornece uma base de estudo que explora diferentes partes de uma transmissão televisiva de um jogo, podendo assim ser adaptada para outros países, desportos ou competições. Tendo o cenário daquilo que se passa nas audiências das ligas profissionais em Portugal, comprovou-se que o estudo das audiências televisivas tem muitos caminhos para explorar e não depende apenas de fatores intrínsecos à programação televisiva, como a hora de transmissão, as figuras que nela surgem, ou o canal que a emite.

Este projeto apresenta algumas limitações, que podem impactar o resultado dos modelos a que se chegou. Os dados relativos ao valor de mercado dos plantéis de cada clube foram retirados no início de cada época e considerados estanques ao longo das mesmas, sabendo-se que na realidade não é bem assim e oscilam com o decorrer do campeonato e, portanto, não sendo a representação mais fiel ao longo do mesmo, no que concerne à mais-valia de cada equipa. Numa base de trabalho mais regular, o estudo podia ser feito jogo a jogo, tendo assim informação mais aproximada acerca deste atributo. Existem ainda outros dados que são tentativas de representar uma realidade pouco concreta, como é o caso do poder de compra dos adeptos de uma equipa e a quantidade de apoiantes que a mesma tem. Para essas variáveis decidiu-se associar a cada equipa os valores do poder de compra do município a que a mesma pertence e a população residente serviu de reflexo para a base de apoio. Caso existisse acesso ao número de sócios e preço das quotas atualizado de todas as equipas, o trabalho seria mais fiel neste tipo de representação. Também relacionado com os adeptos importa referir a presença das equipas B do Porto e do Benfica que não possuem uma postura semelhante aos restantes plantéis, pois os seus

apoiantes são apenas uma percentagem dos da equipa principal, os seus plantéis são rotativos e tanto podem incluir jogadores da equipa principal, como juniores, sendo assim mais difícil associar à mesma um valor de mercado e até mesmo caracterizar a sua base de fãs. Outra das limitações surgiu com a pandemia de Covid-19, pois alguns dos encontros da segunda época em estudo foram disputados sem adeptos no estádio, facto que não é considerado uma situação normal, afetando assim o tipo de transmissão televisiva. O fator do número de espetadores presenciais não foi tomado em conta, impossibilitando assim a perceção do real impacto que as transmissões pós-Covid sofreram. Nestas épocas, a Primeira Liga tem transmissão televisiva integral de todos os encontros, mas o mesmo não se verifica na Segunda Liga. Este aspeto provoca mais uma possível distorção em relação às conclusões obtidas sobre esse campeonato.

Considerando as descobertas ao longo do projeto e as limitações acima referidas seria pertinente, para trabalhos futuros, abordar as audiências, limitando o grupo de equipas em análise fazendo, por exemplo, estudos só com os jogos das “3 equipas grandes”, ou todos os jogos em que nenhuma delas está presentes, de forma a descortinar o comportamento das variáveis estudadas num ambiente de jogos mais equilibrados.

Este trabalho focou-se nos campeonatos portugueses, mas era interessante replicar o estudo noutros países, competições e plataformas de distribuição. O consumo de futebol é diferente entre Portugal e outras regiões, entre a Liga Portuguesa e outras competições, bem como nos acessos via televisão face ao online. Ao definir um diferente espetro de análise e com o conhecimento obtido deste projeto, pode ser possível acumular mais informação e conhecer melhor e mais especificamente cada mercado.

## REFERÊNCIAS BIBLIOGRÁFICAS

- Almeida, N. P. D. B. de. (2015). *Os Direitos Televisivos no Futebol Português – Análise ao modelo atual e proposta de um modelo alternativo*. Universidade de Lisboa.
- Baimbridge, M., Cameron, S., & Dawson, P. (1996). Satellite television and the demand for football: A whole new Ballgame? *Scottish Journal of Political Economy*, 43(3), 317–333.
- Belfo, F. P. (2019). Método de Desenvolvimento de Campanha de Marketing através do Marketing de Base de Dados. *Proceedings of the 14th Iberian Conference on Information Systems and Technologies (CISTI'2019)*, 1–8. <https://doi.org/10.23919/CISTI.2019.8760892>
- Belfo, F. P., & Andreica, A. B. (2018). A Comprehensive Methodology to Implement Business Intelligence and Analytics Through Knowledge Discovery in Databases. *Mining Intelligence and Knowledge Exploration. MIKE 2018. Lecture Notes in Computer Science*, 11308, 102–111. [https://doi.org/https://doi.org/10.1007/978-3-030-05918-7\\_10](https://doi.org/https://doi.org/10.1007/978-3-030-05918-7_10)
- Boehm, K., Klein, F., Lee, P., Mogg, A., Raab, J., & Esser, R. (2018). The future of the TV and video landscape by 2030. *Deloitte*, 8, 1–32.
- Brandão, L., Belfo, F. P., & Silva, A. (2021). Wavelet-based cancer drug recommender system. *Procedia Computer Science, Communications in Computer and Information Science*, 181, 487–494. <https://doi.org/https://doi.org/10.1016/j.procs.2021.01.194>
- Bridge, G. (2019). *Netflix Released More Originals in 2019 Than the Entire TV Industry Did in 2005*. <https://variety.com/2019/tv/news/netflix-more-2019-originals-than-entire-tv-industry-in-2005-1203441709/>
- Briggs, A., & Burke, P. (2009). *Social History Of The Media* (Polity Press (ed.); 3<sup>a</sup>).
- Bühler, A., Rennhak, C., & Nufer, G. (2006). *The nature of sports marketing* (Issue January 2006).
- Cádima, F. R. (2010). Especificidade e interesse público dos conteúdos do Serviço Público de Televisão no contexto da migração para o digital. *Congreso Euro-Iberoamericano de Alfabetización Mediática y Culturas Digitales Sevilla*. <http://hdl.handle.net/11441/56275>

CAEM/GFK. (2021). *Estudo TAM*. <http://caem.pt/tv>

CAEM. (2021). *CAEM / Glossário*. CAEM. <http://www.caem.pt/glossario>

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Daimlerchrysler, T. R., Shearer, C., & Daimlerchrysler, R. W. (2000). CRISP-DM 1.0, Step-by-step data mining guide. In *SPSS inc* (Vol. 78, pp. 1–78). <https://doi.org/10.1017/CBO9781107415324.004>

Christos, K., Panagiotis, A., Athanasios, L., & Douvis, J. (2018). The English Premier League TV Rights Selling Model – Historical Study. *Scientific Journal of Education, Sports, and Health, XIX*(1), 80–88. <https://doi.org/10.29081/gsjesh.2018.19.1.07>

Cui, G., Wong, M. L., & Lui, H.-K. (2006). Machine learning for direct marketing response models: Bayesian networks with evolutionary programming. *Management Science, 52*(4), 597–612.

Dudley, G. (2017). *Sharing content: Why BT and Sky agreed their new pay-TV partnership*. Sports Pro. <https://www.sportspromedia.com/interviews/sharing-content-why-bt-and-sky-agreed-their-new-pay-tv-partnership>

Fantastic. (2014). *Sporting TV inicia hoje as suas emissões*. Fantastic. <https://www.fantastictv.pt/2014/07/sporting-tv-inicia-hoje-as-suas-emissoes.html?m=1>

Faria, D. (2017). *Capital da Sport TV dividido em quatro*. Correrio Da Manhã. <https://www.cmjornal.pt/tv-media/detalhe/capital-da-sport-tv-dividido-em-quatro>

Fernandes, P. A. dos S. (2013). *A Influência das Audiências nos Alinhamentos Televisivos : A Greve Geral*. [https://eg.uc.pt/bitstream/10316/35910/1/A Influencia das Audiencias nos Alinhamentos Televisivos.pdf](https://eg.uc.pt/bitstream/10316/35910/1/A%20Influencia%20das%20Audiencias%20nos%20Alinhamentos%20Televisivos.pdf)

Ferreira, B. (2020). *Como é feita a medição de audiências em Portugal?* Espalha-Factos. <https://espalhafactos.com/2021/06/12/audiencias-como-e-feita-a-medicao-de-tv-em-portugal/>

Galamba, S. A. M. (2014). *A Televisão em Portugal : Um Estudo sobre Géneros Televisivos nos Canais Generalistas em Perspetiva Comparada*. ISCTE - Instituto Universitário de Lisboa.

Gazapo, C. (2018). *TV Rights in Football - Premier League Analysis*. Sports Business Institute Barcelona. <https://www.sbibarcelona.com/newsdetails/index/403>

- Análise e previsão de audiências no mercado televisivo e de comunicação em Portugal*
- Google Maps. (2021). *Distância geográfica entre o estádio do Futebol Clube do Porto e Grupo Desportivo de Chaves.*
- Gouveia, C., & Pereira, R. (2021). Professional football in Portugal: preparing to resume after the COVID-19 pandemic. *Soccer & Society*, 22(1–2), 103–114. <https://doi.org/10.1080/14660970.2020.1796653>
- Gratton, C., & Solberg, H. A. (2007). *The Economics of Sports Broadcasting*. Routledge.
- Hsieh, W., Seng-cho, T., Cheng, Y., & Wu, C. (2013). Predicting TV Audience Rating with Social Media. *IJCNLP 2013 Workshop on Natural Language Processing for Social Media, October*, 1–5. <https://www.aclweb.org/anthology/W13-4201.pdf>
- Huber, S., Wiemer, H., Schneider, D., & Ihlenfeldt, S. (2019). DMME: Data mining methodology for engineering applications—a holistic extension to the CRISP-DM model. *Procedia Cirp*, 79, 403–408. <https://doi.org/10.1016/j.procir.2019.02.106>
- Lang, B. (2019). *Netflix Dominates Golden Globe Nominations With ‘Marriage Story,’ ‘Unbelievable,’ ‘The Crown.’* <https://variety.com/2019/film/awards/golden-globe-nominations-2019-top-nominees-netflix-marriage-story-1203429203/>
- Lopes, F., & Pereira, S. (2007). Estudos sobre programação televisiva : os programas de informação e os conteúdos para a infância. *Conferência Internacional Informação e Programação de Serviço Público Num Contexto Competitivo*, 1–31. <http://hdl.handle.net/1822/8714>
- Loureiro, A., Lourenço, J., Costa, E., & Belfo, F. (2014). Indução de Árvores de Decisão na Descoberta de Conhecimento: Caso de Empresa de Organização de Eventos. In *VI Congresso Internacional de Casos Docentes em Marketing Público e Não Lucrativo*.
- Marktest. (2017). *YUMI - Your Unique Media Intelligence Solution*. <https://www.marktest.com/wap/private/images/Docs/MediaMonitor/yumi.pdf>
- Marques, S. (2014). *O dia em que Portugal viu o primeiro jogo em direto na televisão. Mais Futebol*. <https://maisfutebol.iol.pt/o-dia-em-que-portugal-viu-o-primeiro-jogo-em-direto-na-tv-139199172>
- Morris, C. (2020). *Netflix will spend over \$17 billion on content in 2020: Analyst*. Fortune. <https://fortune.com/2020/01/16/netflix-spending-content-2020-17-billion/>

- Análise e previsão de audiências no mercado televisivo e de comunicação em Portugal*
- Netflix. (2020). *About Netflix - Homepage*. <https://media.netflix.com/en/about-netflix>
- Netflix Inc. (2019). *Netflix 2019 Quarterly Earnings - Letter to Shareholders*. 1–13. [https://s22.q4cdn.com/959853165/files/doc\\_financials/2019/q4/FINAL-Q4-19-Shareholder-Letter.pdf](https://s22.q4cdn.com/959853165/files/doc_financials/2019/q4/FINAL-Q4-19-Shareholder-Letter.pdf)
- Nogueira, C. (2018). *No Title*. Diário de Notícias. <https://www.dn.pt/desportos/o-1-jogo-do-campeonato-em-direto-na-tv-foi-ha-40-anos-e-vitor-baptista-marcou-10068744.html>
- Pimenta, C., Ribeiro, R., Sá, V., & Belfo, F. P. (2018). Fatores que Influenciam o Sucesso Escolar das Licenciaturas numa Instituição de Ensino Superior Portuguesa. In *Atas da 18ª Conferência da Associação Portuguesa de Sistemas de Informação (CAPSI 2018) Associação Portuguesa de Sistemas de Informação*. Associação Portuguesa de Sistemas de Informação.
- Pimenta, P., Belfo, F., & Trigo, A. (2011). Study the Impact of Booking, com User Scores and Reviews in Hotel Management. In M. M. Cruz-Cunha, J. Varajão, P. Powell, & R. Martinho (Eds.), *Book of abstracts of the CENTERIS 2011, Conference on ENTERprise Information Systems* (Vol. 30, pp. 8–9).
- Reprodução/GfK em Espalha-Factos. (2020). *Audímetro*. Espalha-Factos. <https://espalhafactos.com/2021/06/12/audiencias-como-e-feita-a-medicao-de-tv-em-portugal/>
- Rodrigues, D. (2017). *Os Direitos Televisivos no Futebol Profissional – Uma nova proposta de modelo de centralização para a I Liga*. Universidade de Évora.
- Rodríguez, C., Pérez, L., Puente, V., & Rodríguez, P. (2012). The determinants of television audience for professional cycling: The case of Spain. *Journal of Sports Economics*, 16(1), 26–58. <https://doi.org/10.1177/1527002512471536>
- RTP. (2020). *RTP Lab | Regulamento*. RTP. <https://media.rtp.pt/rtplab/regulamento-2020/>
- Santamaría, J. V. G., & Díaz, G. A. (2014). El mercado español de televisión. *Telos*, 96, 115–124.
- Sapo Desporto. (2015). *FC Porto oficializa compra do Porto Canal*. Sapo Desporto. <https://desporto.sapo.pt/futebol/primeira-liga/artigos/fc-porto-oficializa-compra-do-porto-canal>

- Scelles, N., Dermit-Richard, N., & Haynes, R. (2020). What drives sports TV rights ? A comparative analysis of their evolution in English and French men ' s football first divisions , 1980 – 2020. *Soccer & Society*, 21(5), 491–509. <https://doi.org/10.1080/14660970.2019.1681406>
- Schreyer, D., & Schmidt, S. L. (2016). Game Outcome Uncertainty and Television Audience Demand : New Evidence from German Football. *German Economic Review*, 1956, 1–22. <https://doi.org/10.1111/geer.12120>
- Seiça, A., Trigo, A., & Belfo, F. P. (2019). LexiNB - Uma Abordagem Bietápica de Classificação de Sentimentos em Tweets Relacionados com as Autoridades Fiscais Portuguesas. *Proceedings of the 19.<sup>a</sup> Conferência Da Associação Portuguesa de Sistemas de Informação (CAPSI'2019) Held in Lisboa, Portugal, 11-12 October 2019. Paper 5.*
- Sereday, S., & Cui, J. (2017a). Using machine learning to predict future tv ratings. *Data Science, Nielsen*, 1(3), 3–12.
- Sereday, S., & Cui, J. (2017b). Using Machine Learning to Predict Future TV Ratings. *Journal of Measurement – Nielsen*, 1(February), 1–13. <https://www.nielsen.com/us/en/insights/report/2017/using-machine-learning-to-predict-future-tv-ratings/>
- Silveira, M. T. da. (2010). *Futebol da televisão: molduras audiovisuais. Trabalho de Conclusão de Curso – Habilitação Jornalismo.* Universidade Federal do Rio Grande do Sul, Porto Alegre.
- Spade, P. V., & Panaccio, C. (2002). *William of Ockham.* <https://plato.stanford.edu/entries/ockham/>
- Spangler, T. (2020). *Netflix Projected to Spend More Than \$17 Billion on Content in 2020.* Variety. <https://variety.com/2020/digital/news/netflix-2020-content-spending-17-billion-1203469237>
- Stoll, J. (2020). *Number of hours of first-run original content released by Netflix worldwide from 2012 to 2019.* Statista. <https://www.statista.com/statistics/882490/netflix-original-content-hours/>
- Tainsky, S. (2010). Television Broadcast Demand for National Football League Contests. *Journal of Sports Economics*, 629–640. <https://doi.org/10.1177/1527002509355636>

*Análise e previsão de audiências no mercado televisivo e de comunicação em Portugal*

Torres, E. C. (2011). *A Televisão e o Serviço Público*. Relógio D'Água Editores.

Transfermarkt. (2021). *Valor de mercado das equipas da Primeira Liga, no início da época 2018-2019*.

UEFA. (2003). *Regulations governing the implementation of Article 48 of the UEFA Statutes*.

UEFA. (2018). The European Club Footballing Landscape. In *Club Licensing Benchmarking Report: Financial Year 2018*.  
[https://www.uefa.com/MultimediaFiles/Download/OfficialDocument/uefaorg/ClubLicensing/02/63/79/75/2637975\\_DOWNLOAD.pdf](https://www.uefa.com/MultimediaFiles/Download/OfficialDocument/uefaorg/ClubLicensing/02/63/79/75/2637975_DOWNLOAD.pdf)

Zerozero. (2021). *Classificação final da Liga NOS em 2018-2019*.

## **APÊNDICES**

## APÊNDICE 1: Amostra da base de dados

Quadro 1 - Primeira parte da base de dados

Id	Divisão	NacInt	Competição	Equi1	Equi2	Horári	Data	DiaSemana	Seg	Ter	Qua	Qui	Sex	Sab	Dor	Mês	Ano	Época	DistAdv	DistP	Dist	Son	Difc
1	1	Nacional	Liga Portuguesa Benfica	V. Guimarães		20:30	10-08-2018	sexta-feira	0	0	0	0	1	0	0	Agosto	2018	2018-2019	362	0	0	0	0
2	1	Nacional	Liga Portuguesa V. Setúbal	Aves		16:30	11-08-2018	sábado	0	0	0	0	0	1	0	Agosto	2018	2018-2019	381	0	0	0	0
3	1	Nacional	Liga Portuguesa Tondela	Belenenses SAD		19:00	11-08-2018	sábado	0	0	0	0	0	1	0	Agosto	2018	2018-2019	279	0	0	0	0
4	1	Nacional	Liga Portuguesa Porto	Chaves		21:00	11-08-2018	sábado	0	0	0	0	0	1	0	Agosto	2018	2018-2019	151	0	0	0	0
5	1	Nacional	Liga Portuguesa Marítimo	Santa Clara		16:00	12-08-2018	domingo	0	0	0	0	0	0	1	Agosto	2018	2018-2019	972	0	0	0	0
6	1	Nacional	Liga Portuguesa Feirense	Rio Ave		16:00	12-08-2018	domingo	0	0	0	0	0	0	1	Agosto	2018	2018-2019	63	0	0	0	0
7	1	Nacional	Liga Portuguesa Moreirense	Sporting		18:30	12-08-2018	domingo	0	0	0	0	0	0	1	Agosto	2018	2018-2019	354	0	0	0	0
8	1	Nacional	Liga Portuguesa Braga	Nacional		20:30	12-08-2018	domingo	0	0	0	0	0	0	1	Agosto	2018	2018-2019	1239	0	0	0	0
9	1	Nacional	Liga Portuguesa Portimonense	Boavista		20:15	13-08-2018	segunda-feira	1	0	0	0	0	0	0	Agosto	2018	2018-2019	552	0	0	0	0
10	1	Nacional	Liga Portuguesa Aves	Tondela		20:30	17-08-2018	sexta-feira	0	0	0	0	1	0	0	Agosto	2018	2018-2019	160	3	3	6	0
11	1	Nacional	Liga Portuguesa Chaves	Portimonense		16:30	18-08-2018	sábado	0	0	0	0	0	1	0	Agosto	2018	2018-2019	682	3	3	6	0
12	1	Nacional	Liga Portuguesa Boavista	Benfica		19:00	18-08-2018	sábado	0	0	0	0	0	1	0	Agosto	2018	2018-2019	315	0	0	0	0
13	1	Nacional	Liga Portuguesa Sporting	V. Setúbal		21:00	18-08-2018	sábado	0	0	0	0	0	1	0	Agosto	2018	2018-2019	53	0	0	0	0
14	1	Nacional	Liga Portuguesa Nacional	Moreirense		16:00	19-08-2018	domingo	0	0	0	0	0	0	1	Agosto	2018	2018-2019	1228	3	3	6	0
15	1	Nacional	Liga Portuguesa Rio Ave	Marítimo		16:00	19-08-2018	domingo	0	0	0	0	0	0	1	Agosto	2018	2018-2019	1210	3	0	3	3
16	1	Nacional	Liga Portuguesa Belenenses SAD	Porto		18:30	19-08-2018	domingo	0	0	0	0	0	0	1	Agosto	2018	2018-2019	320	0	0	0	0
17	1	Nacional	Liga Portuguesa Santa Clara	Braga		20:30	19-08-2018	domingo	0	0	0	0	0	0	1	Agosto	2018	2018-2019	1528	3	0	3	3
18	1	Nacional	Liga Portuguesa V. Guimarães	Feirense		20:15	20-08-2018	segunda-feira	1	0	0	0	0	0	0	Agosto	2018	2018-2019	82	3	0	3	3
19	1	Nacional	Liga Portuguesa Marítimo	Chaves		20:30	24-08-2018	sexta-feira	0	0	0	0	1	0	0	Agosto	2018	2018-2019	1312	3	3	6	0
20	1	Nacional	Liga Portuguesa Portimonense	Santa Clara		16:30	25-08-2018	sábado	0	0	0	0	0	1	0	Agosto	2018	2018-2019	1507	6	5	11	1
21	1	Nacional	Liga Portuguesa Benfica	Sporting		19:00	25-08-2018	sábado	0	0	0	0	0	1	0	Agosto	2018	2018-2019	4	0	0	0	0
22	1	Nacional	Liga Portuguesa Porto	V. Guimarães		21:00	25-08-2018	sábado	0	0	0	0	0	1	0	Agosto	2018	2018-2019	51	0	6	6	6
23	1	Nacional	Liga Portuguesa V. Setúbal	Nacional		16:00	26-08-2018	domingo	0	0	0	0	0	0	1	Agosto	2018	2018-2019	972	3	6	9	3
24	1	Nacional	Liga Portuguesa Feirense	Boavista		16:00	26-08-2018	domingo	0	0	0	0	0	0	1	Agosto	2018	2018-2019	33	0	3	3	3
25	1	Nacional	Liga Portuguesa Tondela	Rio Ave		18:30	26-08-2018	domingo	0	0	0	0	0	0	1	Agosto	2018	2018-2019	168	5	3	8	2
26	1	Nacional	Liga Portuguesa Braga	Aves		20:30	26-08-2018	domingo	0	0	0	0	0	0	1	Agosto	2018	2018-2019	35	2	5	7	3
27	1	Nacional	Liga Portuguesa Moreirense	Belenenses SAD		20:15	27-08-2018	segunda-feira	1	0	0	0	0	0	0	Agosto	2018	2018-2019	369	3	3	6	0
28	1	Nacional	Liga Portuguesa Chaves	Braga		19:00	31-08-2018	sexta-feira	0	0	0	0	1	0	0	Agosto	2018	2018-2019	128	4	0	4	4
29	1	Nacional	Liga Portuguesa V. Guimarães	Tondela		21:15	31-08-2018	sexta-feira	0	0	0	0	1	0	0	Agosto	2018	2018-2019	177	4	5	9	1
30	1	Nacional	Liga Portuguesa Belenenses SAD	V. Setúbal		16:30	01-09-2018	sábado	0	0	0	0	0	1	0	Setembro	2018	2018-2019	53	3	4	7	1
31	1	Nacional	Liga Portuguesa Santa Clara	Boavista		19:00	01-09-2018	sábado	0	0	0	0	0	1	0	Setembro	2018	2018-2019	1503	5	3	8	2
32	1	Nacional	Liga Portuguesa Sporting	Feirense		21:00	01-09-2018	sábado	0	0	0	0	0	1	0	Setembro	2018	2018-2019	284	0	0	0	0

## Análise e previsão de audiências no mercado televisivo e de comunicação em Portugal

Quadro 2 - Segunda parte da base de dados

Id	Poc	Poc	Dif	Sor	Med	PopEq	PopEq	DifPop	SomaP	Par	P	D	S	A	A	D	S	Clau	Cl	Di	So	VME	VME	DifV	Som	Go	CapE	Canal	Preço	Par	Rat#	Rat%	Shr%
1	220%	91%	128%	311%	190%	507.220	152.792	354.428	660.012	85	74	11	159	85	12	73	97	2	9	7	11	233,75	38	195,75	271,75	5	65.642	BTV	9,90 €	0	201,5	2%	5%
2	108%	86%	22%	193%	99%	115.758	68.221	47.537	183.979	71	5	66	76	15	2	13	17	14	13	1	27	12,55	9,6	2,95	22,15	2	18.642	SPTV1	23,99 €	0	23,9	0%	1%
3	75%	157%	82%	232%	146%	26.548	176.218	149.670	202.766	4	1	3	5	4	1	3	5	11	12	1	23	15,23	13,18	2,05	28,41	1	5.000	SPTV5	23,99 €	0	18,0	0%	1%
4	158%	79%	79%	237%	146%	215.284	39.345	175.939	254.629	85	16	69	101	85	1	84	86	1	6	5	7	230,88	16,88	214	247,76	5	50.399	SPTV1	23,99 €	0	125,8	1%	4%
5	114%	108%	6%	222%	112%	104.129	67.864	36.265	171.993	39	4	35	43	37	1	36	38	7	20	13	27	21,78	8,8	12,98	30,58	1	10.600	SPTV3	23,99 €	1	6,1	0%	0%
6	85%	96%	11%	181%	89%	138.525	79.579	58.946	218.104	7	25	18	32	3	11	8	14	16	5	11	21	10,25	15,55	5,3	25,8	2	5.401	SPTV5	23,99 €	1	8,8	0%	0%
7	91%	220%	128%	311%	190%	152.792	507.220	354.428	660.012	9	85	76	94	5	85	80	90	15	3	12	18	15,8	158,88	143,08	174,68	4	6.153	SPTV1	23,99 €	0	138,1	1%	4%
8	107%	114%	7%	221%	110%	181.919	104.129	77.790	286.048	63	19	44	82	44	1	43	45	4	19	15	23	74,45	8,2	66,25	82,65	6	30.286	SPTV2	23,99 €	0	26,5	0%	1%
9	104%	158%	54%	261%	147%	55.416	215.284	159.868	270.700	16	56	40	72	2	5	3	7	10	8	2	18	21,95	11,73	10,22	33,68	2	6.000	SPTV1	23,99 €	0	29,0	0%	1%
10	86%	75%	11%	161%	83%	68.221	26.548	41.673	94.769	5	4	1	9	2	4	2	6	13	11	2	24	9,6	15,23	5,63	24,83	4	5.441	SPTV1	23,99 €	0	42,6	0%	1%
11	79%	104%	24%	183%	93%	39.345	55.416	16.071	94.761	16	16	0	32	1	2	1	3	6	10	4	16	16,88	21,95	5,07	38,83	2	9.000	SPTV5	23,99 €	0	16,5	0%	1%
12	158%	220%	62%	377%	201%	215.284	507.220	291.936	722.504	56	85	29	141	5	85	80	90	8	2	6	10	11,73	233,75	222,02	245,48	2	28.263	SPTV1	23,99 €	0	149,3	2%	5%
13	220%	108%	112%	327%	199%	507.220	115.758	391.462	622.978	85	71	14	156	85	15	70	100	3	14	11	17	158,88	12,55	146,33	171,43	3	50.080	SPTV2	23,99 €	0	94,9	1%	3%
14	114%	91%	23%	206%	101%	104.129	152.792	48.663	256.921	19	9	10	28	1	5	4	6	19	15	4	34	8,2	15,8	7,6	24	3	5.586	SPTV5	23,99 €	1	4,1	0%	0%
15	96%	114%	18%	211%	106%	79.579	104.129	24.550	183.708	25	39	14	64	11	37	26	48	5	7	2	12	15,55	21,78	6,23	37,33	4	9.065	SPTV2	23,99 €	1	9,1	0%	0%
16	157%	158%	1%	314%	157%	176.218	215.284	39.066	391.502	1	85	84	86	1	85	84	86	12	1	11	13	13,18	230,88	217,7	244,06	5	37.593	SPTV1	23,99 €	0	160,1	2%	6%
17	108%	107%	1%	215%	107%	67.864	181.919	114.055	249.783	4	63	59	67	1	44	43	45	20	4	16	24	8,8	74,45	65,65	83,25	6	12.500	SPTV2	23,99 €	0	10,8	0%	0%
18	91%	85%	7%	176%	88%	152.792	138.525	14.267	291.317	74	7	67	81	12	3	9	15	9	16	7	25	38	10,25	27,75	48,25	1	29.865	SPTV1	23,99 €	0	25,7	0%	1%
19	114%	79%	35%	194%	105%	104.129	39.345	64.784	143.474	39	16	23	55	37	1	36	38	7	6	1	13	21,78	16,88	4,9	38,66	3	10.600	SPTV1	23,99 €	0	29,6	0%	1%
20	104%	108%	4%	211%	106%	55.416	67.864	12.448	123.280	16	4	12	20	2	1	1	3	10	20	10	30	21,95	8,8	13,15	30,75	4	6.000	SPTV1	23,99 €	0	17,7	0%	1%
21	220%	220%	0%	439%	220%	507.220	507.220	0	1.014.440	85	85	0	170	85	85	0	170	2	3	1	5	233,75	158,88	74,87	392,63	2	65.642	BTV	9,90 €	0	293,2	3%	9%
22	158%	91%	66%	249%	130%	215.284	152.792	62.492	368.076	85	74	11	159	85	12	73	97	1	9	8	10	230,88	38	192,88	268,88	5	50.399	SPTV1	23,99 €	0	123,8	1%	3%
23	108%	114%	7%	222%	111%	115.758	104.129	11.629	219.887	71	19	52	90	15	1	14	16	14	19	5	33	12,55	8,2	4,35	20,75	3	18.642	SPTV5	23,99 €	1	4,7	0%	0%
24	85%	158%	73%	243%	129%	138.525	215.284	76.759	353.809	7	56	49	63	3	5	2	8	16	8	8	24	10,25	11,73	1,48	21,98	2	5.401	SPTV1	23,99 €	1	29,3	0%	1%
25	75%	96%	21%	171%	91%	26.548	79.579	53.031	106.127	4	25	21	29	4	11	7	15	11	5	6	16	15,23	15,55	0,32	30,78	2	5.000	SPTV2	23,99 €	0	20,0	0%	1%
26	107%	86%	21%	193%	101%	181.919	68.221	113.698	250.140	63	5	58	68	44	2	42	46	4	13	9	17	74,45	9,6	64,85	84,05	4	30.286	SPTV1	23,99 €	0	42,1	0%	1%
27	91%	157%	65%	248%	126%	152.792	176.218	23.426	329.010	9	1	8	10	5	1	4	6	15	12	3	27	15,8	13,18	2,62	28,98	2	6.153	SPTV1	23,99 €	0	31,7	0%	1%
28	79%	107%	28%	186%	102%	39.345	181.919	142.574	221.264	16	63	47	79	1	44	43	45	6	4	2	10	16,88	74,45	57,57	91,33	1	9.000	SPTV1	23,99 €	0	14,2	0%	1%
29	91%	75%	16%	166%	89%	152.792	26.548	126.244	179.340	74	4	70	78	12	4	8	16	9	11	2	20	38	15,23	22,77	53,23	1	29.865	SPTV2	23,99 €	0	14,3	0%	0%
30	157%	108%	49%	264%	137%	176.218	115.758	60.460	291.976	1	71	70	72	1	15	14	16	12	14	2	26	13,18	12,55	0,63	25,73	0	37.593	SPTV1	23,99 €	0	11,0	0%	1%
31	108%	158%	50%	266%	146%	67.864	215.284	147.420	283.148	4	56	52	60	1	5	4	6	20	8	12	28	8,8	11,73	2,93	20,53	6	12.500	SPTV5	23,99 €	0	20,2	0%	1%
32	220%	85%	135%	304%	191%	507.220	138.525	368.695	645.745	85	7	78	92	85	3	82	88	3	16	13	19	158,88	10,25	148,63	169,13	1	50.080	SPTV1	23,99 €	0	109,0	1%	3%

## **ANEXOS**

## ANEXO I: Classificação final da Primeira Liga em 2018-2019

### CLASSIFICAÇÃO

		P	J	V	E	D	GM	GS	DG	
1	Benfica	87	34	28	3	3	103	31	+72	+
2	FC Porto	85	34	27	4	3	74	20	+54	+
3	Sporting	74	34	23	5	6	72	33	+39	+
4	SC Braga	67	34	21	4	9	56	37	+19	+
5	V. Guimarães	52	34	15	7	12	46	34	+12	+
6	Moreirense	52	34	16	4	14	39	44	-5	+
7	Rio Ave	45	34	12	9	13	50	52	-2	+
8	Boavista	44	34	13	5	16	34	40	-6	+
9	Belenenses SAD	43	34	10	13	11	42	51	-9	+
10	Santa Clara	42	34	11	9	14	43	45	-2	+
11	Marítimo	39	34	12	3	19	26	44	-18	+
12	Portimonense	39	34	11	6	17	44	59	-15	+
13	V. Setúbal	36	34	8	12	14	28	39	-11	+
14	CD Aves	36	34	10	6	18	35	49	-14	+
15	CD Tondela	35	34	9	8	17	40	54	-14	+
16	GD Chaves	32	34	8	8	18	34	57	-23	+
17	Nacional	28	34	7	7	20	33	73	-40	+
18	Feirense	20	34	3	11	20	27	64	-37	+

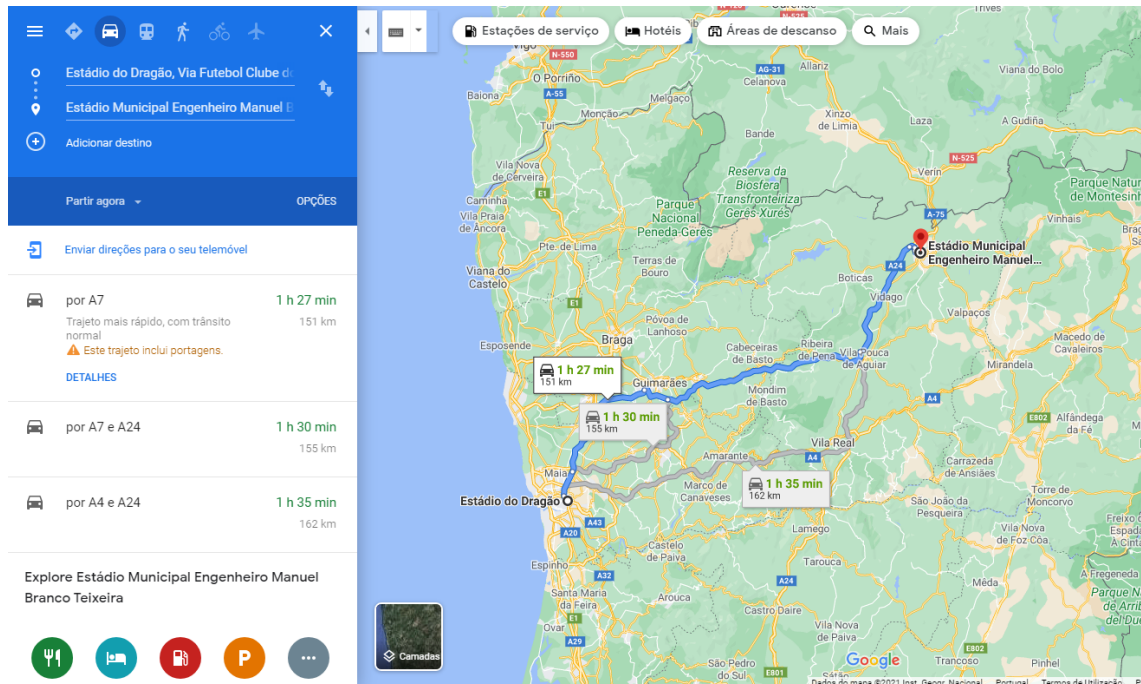
Fonte: Zerozero (2021)

## ANEXO II: Valor de mercado das equipas da Primeira Liga, no início da época 2018-2019

Compacto		Mais	
#	Clube	Liga	Valor01/08/2018 †
1	 Sport Lisboa e Benfica	Liga Bwin	233,75 M €
2	 Futebol Clube do Porto	Liga Bwin	230,88 M €
3	 Sporting Clube de Portugal	Liga Bwin	158,88 M €
4	 Sporting Clube de Braga	Liga Bwin	74,45 M €
5	 Vitória Sport Clube	Liga Bwin	38,00 M €
6	 FC Famalicão	Liga Portugal 2	7,10 M €
7	 Boavista Futebol Clube	Liga Bwin	11,73 M €
8	 FC Paços de Ferreira	Liga Portugal 2	10,15 M €
9	 Portimonense Futebol SAD	Liga Bwin	21,95 M €
10	 Clube Desportivo Santa Clara	Liga Bwin	8,80 M €
11	 Moreirense Futebol Clube	Liga Bwin	15,80 M €
12	 Clube Desportivo Tondela	Liga Bwin	15,23 M €
13	 Gil Vicente Futebol Clube		1,93 M €
14	 Club Sport Marítimo	Liga Bwin	21,78 M €
15	 Belenenses SAD	Liga Bwin	13,18 M €
16	 Futebol Clube de Vizela		375 mil €
17	 GD Estoril Praia	Liga Portugal 2	12,10 M €
18	 Futebol Clube Arouca	Liga Portugal 2	8,53 M €

Fonte: Transfermarkt (2021)

### ANEXO III: Distância geográfica entre o estádio do Futebol Clube do Porto e Grupo Desportivo de Chaves



Fonte: Google Maps (2021)