

2025

**Diogo Miguel Garcia
Gonçalves**

**O LOOK TO BOOK COMO FERRAMENTA DE
GESTÃO: UMA ABORDAGEM BASEADA EM
DADOS NO SETOR TURÍSTICO**

2025

**Diogo Miguel Garcia
Gonçalves**

**O LOOK TO BOOK COMO FERRAMENTA DE
GESTÃO: UMA ABORDAGEM BASEADA EM
DADOS NO SETOR TURÍSTICO**

Dissertação apresentada à Faculdade de Ciências Sociais e Tecnologia da Universidade Europeia, para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Gestão realizada sob a orientação científica da Professora Doutora Mariana Sofia Barreira Cavique Santos, Professora Auxiliar da Universidade Europeia

Agradecimentos

Gostaria de iniciar estes agradecimentos com uma palavra especial à minha orientadora, Professora Doutora Mariana Cavique. Muito obrigado por todo o seu acompanhamento e apoio incondicional, desde a licenciatura até à conclusão deste mestrado. Este estudo só existe graças à sua orientação, pelo que esta dissertação também lhe pertence.

Gostaria também de agradecer à equipa de *Business Optimization* da empresa Viagens Abreu, pela disponibilidade e pela autorização para a utilização dos dados que serviram de base a este estudo.

Aos meus pais, deixo o meu maior e mais profundo agradecimento. Sem vocês nada disto teria sido possível. Muito obrigado por todo o apoio e suporte incondicionais ao longo da vida, por fazerem de tudo para que nada me faltasse e por me darem a oportunidade de cumprir mais um objetivo. Sem o vosso apoio e sacrifício, nunca teria chegado até aqui, isto é tudo graças a vocês. Esta conquista é vossa tanto quanto é minha!

À minha namorada, que foi o meu pilar e maior suporte nesta jornada, expresso o mais sincero agradecimento. Nos momentos bons e maus, apoiou-me mais que ninguém, motivou-me e encorajou-me a nunca desistir. Muito obrigado, meu amor, por nunca me deixares desistir, por estares sempre disposta a ajudar-me e por me aconselhares em cada detalhe, e acima de tudo, por estares sempre a meu lado!

Aos meus pais e à minha namorada, muito obrigado por acreditarem em mim e por estarem sempre a meu lado. Foram a minha motivação nos dias em que pensei que não conseguiria continuar. Por isso mesmo, esta dissertação não é apenas minha, é nossa!

À minha família e amigos, gostaria imenso de vos agradecer e enaltecer pelo vosso apoio e encorajamento constantes, pelas gargalhadas e memórias e por ajudarem a tornar esta jornada académica mais leve. Um enorme obrigado a todos vocês, aos que ainda irão festejar comigo este feito e também aos que já não podem, mas continuam a ser uma enorme fonte de inspiração e motivação.

Por fim, gostaria de agradecer aos meus colegas de faculdade, o meu grupo de trabalho, por tornarem esta jornada académica mais divertida e memorável. Um especial agradecimento ao amigo que a faculdade me deu, sempre presente nesta caminhada e sempre disponível para ajudar e apoiar.

Resumo

No atual contexto da economia digital, a medição de métricas de desempenho tornou-se crucial para apoiar a gestão e a tomada de decisão. No setor do turismo, esta necessidade é ainda mais evidente, dado o peso das plataformas digitais de pesquisa e reserva no percurso do consumidor e a multiplicidade de fatores que moldam a conversão. Entre as métricas disponíveis, o rácio Look to Book (L2B) mede a eficácia de conversão, ao expressar a proporção de pesquisas que resultam em reservas. Isoladamente, esta métrica traduz a relação entre procura e ação. No entanto, em articulação com variáveis como preço, tipologia de quarto ou regime alimentar, permite compreender padrões de conversão com maior profundidade.

A presente dissertação tem como objetivo central identificar os fatores que influenciam significativamente o rácio L2B. Para tal, foi utilizada uma base de dados real do setor turístico, correspondente ao mês de dezembro de 2024, à qual foram aplicados modelos de regressão linear múltipla, com o objetivo de testar a associação entre o rácio e um conjunto de variáveis explicativas.

Com base nos resultados obtidos, observa-se que o preço médio tem efeito negativo na conversão, enquanto determinadas tipologias de quarto e regimes alimentares mostram impactos positivos. Verificou-se ainda um resultado inesperado nas tarifas não reembolsáveis, cuja influência positiva contrasta com o esperado. No seu conjunto, os fatores apresentam magnitudes próximas, o que reforça o carácter multifatorial do L2B.

Em termos de contributos, a nível académico, a investigação acrescenta evidência empírica à literatura sobre métricas de conversão, ao oferecer uma análise quantitativa que enriquece a compreensão do L2B e dos fatores que o influenciam. Do ponto de vista da gestão, o estudo demonstra a utilidade e o potencial do L2B enquanto ferramenta de apoio à decisão, que disponibiliza orientações práticas que podem sustentar estratégias de marketing, operações e *revenue management*. Para além disso, destaca-se a sua relevância para o *Business Intelligence*, ao oferecer um indicador capaz de ser integrado em *dashboards* e sistemas de monitorização, o que potencia uma gestão mais analítica e orientada por dados.

Palavras-Chave: *Look to Book*, Gestão, Turismo, *Business Intelligence*, Regressão Linear Múltipla

Abstract

In the current context of the digital economy, the measurement of performance metrics has become crucial to support management and decision-making. This is especially relevant in the tourism sector, where digital platforms are central to the search and booking process and where multiple factors shape conversion outcomes. The Look to Book (L2B) ratio captures the efficiency with which searches are converted into bookings. While it reflects the link between demand and action on its own, when analyzed together with variables such as price, room type, or meal plan, it provides valuable insights into conversion dynamics.

The main objective of this dissertation is to identify the factors that significantly influence the L2B ratio. For this purpose, a real dataset from the tourism sector, covering December 2024, was analyzed using multiple linear regression models to assess the association between the ratio and a set of explanatory variables.

The results indicate that average price exerts a negative effect on conversion, whereas certain room types and meal plans have positive impacts. An unexpected outcome emerged with non-refundable rates, whose positive influence contrasts with conventional expectations in the literature. Overall, the factors display comparable magnitudes, underscoring the multifactorial nature of the L2B.

This study makes two main contributions. From an academic perspective, it provides empirical evidence that enriches the literature on conversion metrics through a quantitative analysis of the L2B and its determinants. From a managerial standpoint, it highlights the utility and potential of the L2B as a decision-support tool, offering practical guidance for strategies in marketing, operations, and revenue management. Moreover, its relevance for Business Intelligence is emphasized, as the ratio can be integrated into dashboards and monitoring systems, fostering more analytical and data-driven management practices.

Keywords: Look to Book, Management, Tourism, Business Intelligence, Multiple Linear Regression

Índice

I – Introdução.....	1
II - Revisão da Literatura.....	2
2.1 – Conversão do Consumidor e o Conceito de Look to Book (L2B).....	2
2.1.1 – Fatores Influentes na Taxa de Conversão.....	6
2.2 – Modelação Preditiva e <i>Business Intelligence</i>	8
2.2.1 – Métricas de Conversão e KPI nas Plataformas Digitais.....	10
III - Metodologia.....	12
3.1 – Abordagem Metodológica do Estudo.....	12
3.2 – Recolha e Fontes de Dados.....	13
3.3 – Preparação dos Dados.....	15
3.3.1 – Recodificação das Categorias de Quarto.....	15
3.3.2 – Agregação do <i>Dataset</i> de Reservas.....	17
3.3.3 – Construção do <i>Dataset</i> de Análise: Junção dos <i>Datasets</i>	19
3.3.4 – Definição das Variáveis.....	20
3.4 – Modelo Estatístico.....	21
IV – Análise e Exploração Estatística de Dados.....	23
4.1 – Fase 1: Análise Exploratória Inicial.....	24
4.1.1 Distribuição da Variável Dependente (L2B).....	24
4.1.2 Distribuição das Variáveis Independentes.....	27
4.1.3 Relação entre Variáveis: Correlação e Multicolinearidade.....	29
4.1.4 Considerações finais da Fase 1.....	32
4.2 - Transformações Aplicadas aos Dados.....	32
4.2.1 - <i>Binning</i>	33
4.2.2 - Criação de <i>Dummies</i>	33
4.2.3 - Transformação logarítmica de <i>Avg_Rate</i>	34
4.2.4 - Transformação logit de L2B.....	34
4.2.5 – Resumo das Transformações na Preparação para Regressão.....	35
4.3 - Fase 2: Análise Exploratória após Transformações.....	36
4.3.1 – Distribuição das Variáveis Transformadas.....	36
4.3.2 – <i>Outliers</i> após as Transformações.....	38
4.3.3 – Multicolinearidade após as Transformações.....	40
4.3.4 – Correlações entre Variáveis.....	41

4.3.5 – Impacto Global das Transformações	43
4.4 – Comparação entre L2B Original e L2B_logit.....	44
4.5 – Conclusão da Análise Exploratória de Dados	45
V – Modelação e Validação Estatística	46
5.1 - Análise do Modelo Completo Inicial	47
5.1.1 – Introdução e Enquadramento	47
5.1.2 – Resultados da Estimação.....	47
5.1.3 – Análise Preliminar dos Coeficientes	48
5.1.4 – Conclusão sobre o Modelo Completo.....	49
5.2 – Análise do Modelo Reduzido.....	49
5.2.1 – Introdução e Justificação da Redução	49
5.2.2 – Variáveis Eliminadas e Critérios de Exclusão	50
5.2.3 – Resultados da Estimação do Modelo Reduzido	51
5.2.4 – Testes aos Pressupostos e Interpretação.....	53
5.2.5 – Conclusão sobre o Modelo Reduzido	56
5.3 – Análise do Modelo Robusto.....	57
5.3.1 – Introdução e Justificação Teórica	57
5.3.2 – Resultados da Estimação do Modelo Robusto	58
5.3.3 – Interpretação dos Resultados	60
5.3.4 – Conclusão final sobre o modelo robusto como modelo final	64
VI – Discussão de Resultados.....	65
6.1 – Interpretação dos Coeficientes Significativos	65
6.2 – Comparação de Resultados com a Literatura Existente	67
6.3 – Implicações para a Gestão.....	69
6.4 – Síntese dos Resultados	71
VII – Conclusão.....	72
7.1 – Conclusão do Estudo.....	72
7.2 – Contribuições para a Gestão e para a Academia	74
7.3 – Limitações do Estudo.....	76
7.4 – Recomendações para Estudos Futuros.....	77
VII-Referências Bibliográficas	79

Índice de figuras

Figura 1 - Nuvem dos pares de palavras mais frequentes em “Room”	16
Figura 2 – Gráfico de barras dos pares de palavras mais frequentes em “Room”.....	16
Figura 3 – Representação da transformação dos dados	19
Figura 4 – Distribuição do L2B	25
Figura 5 – Boxplot do L2B	26
Figura 6 – Histogramas das Variáveis Explicativas	27
Figura 7 – Boxplots das Variáveis Explicativas	28
Figura 8 – Heatmap das Correlações entre Variáveis.....	30
Figura 9 – Scatterplots entre L2B e variáveis independentes	31
Figura 10 – Distribuição do L2B Logit	37
Figura 11 – Distribuição de log_Avg_Rate	38
Figura 12 – Boxplot de L2B_logit.....	39
Figura 13 – Boxplot de log_Avg_Rate	40
Figura 14 – Heatmap de L2B original	41
Figura 15 – Heatmap de L2B_logit	42
Figura 16 – Comparação entre L2B Original e L2B_logit	44
Figura 17 – Histograma dos resíduos do modelo reduzido	53
Figura 18 – QQ-Plot dos resíduos do modelo reduzido.....	54
Figura 19 – Gráfico de resíduos vs valores ajustados do modelo reduzido.....	55
Figura 20 – QQ-Plot dos resíduos do modelo robusto	61
Figura 21 – Coeficientes Estimados do Modelo Robusto	63

Índice de tabelas

Tabela 1 – Fases do processo de conversão digital no turismo	3
Tabela 2 – Variáveis presentes no dataset de reservas.....	14
Tabela 3 – Variáveis presentes no dataset de pesquisas	14
Tabela 4 - Variáveis utilizadas na análise	20
Tabela 5 – Comparação entre L2B Original e L2B_logit.....	45
Tabela 6 – Variáveis Eliminadas do Modelo Completo	50

Lista de Abreviaturas

ADR – *Average Daily Rate*

AIC – *Akaike Information Criterion*

AI – *All Inclusive*

BB – *Bed & Breakfast*

B2B – *Business to Business*

B2C – *Business to Consumer*

BIC – *Bayesian Information Criterion*

BI – *Business Intelligence*

CUG – *Closed User Groups*

EDA – *Exploratory Data Analysis*

FB – *Full Board*

HAC – *Heteroskedasticity and Autocorrelation Consistent*

HB – *Half Board*

IC – *Intervalos de Confiança*

KPI – *Key Performance Indicator*

L2B – *Look to Book*

MQO – *Mínimos Quadrados Ordinários*

OLS – *Ordinary Least Squares*

OTA – *Online Travel Agency*

RevPAR – *Revenue per Available Room*

RO – *Room Only*

VIF – *Variance Inflation Factor*

I – Introdução

A digitalização tem transformado profundamente diversos setores, sendo o turismo um dos mais impactados. A crescente integração de ferramentas digitais nos processos de pesquisa, planejamento e reserva alterou de forma decisiva o comportamento dos consumidores, que passaram a valorizar não apenas a conveniência, mas também a personalização e a confiança no ambiente online. Este fenómeno, impulsionado pela disseminação de plataformas digitais, redes sociais e sistemas de reserva em tempo real, trouxe consigo novas oportunidades e desafios para a gestão das organizações. De acordo com Barykin et al. (2021), a digitalização no turismo assume hoje uma relevância estratégica, ao passo que estudos como os de Tussyadiah e Zach (2017), sublinham a forma como os consumidores se tornaram coprodutores da experiência, ao influenciarem e serem influenciados por avaliações, recomendações e conteúdos digitais. Desta forma, o comércio digital no turismo tornou-se, assim, um espaço mediado por dados, algoritmos e interfaces, onde a eficiência da conversão depende de múltiplos fatores, desde o preço até à percepção de valor da oferta (Barykin et al., 2021; Gretzel et al., 2015).

Neste contexto, torna-se essencial compreender métricas que permitam avaliar a eficácia da conversão em ambientes digitais. Entre elas, destaca-se o rácio Look to Book (L2B), que mede o desempenho da conversão de pesquisas em reservas. Isoladamente, o L2B expressa a relação entre procura e ação, refletindo a proporção de reservas face ao volume de pesquisas. No entanto, o seu verdadeiro valor analítico advém da articulação com variáveis adicionais, como o preço, a tipologia de quarto ou o regime alimentar, que ajudam a explicar porque é que determinados padrões de conversão são mais ou menos eficazes.

A questão central que orienta esta dissertação consiste em identificar os fatores que influenciam significativamente o rácio L2B. A partir de um conjunto de dados reais do setor turístico, procura-se compreender de que forma variáveis como o preço médio, as tipologias de quarto, os regimes alimentares ou as políticas tarifárias, se relacionam com a eficiência da conversão digital. O objetivo do estudo é duplo, por um lado, contribuir para o aprofundamento da literatura sobre métricas de conversão em contextos digitais, ao aplicar metodologias quantitativas a um fenómeno ainda pouco estudado. Por outro lado, explorar o potencial do L2B enquanto ferramenta de gestão, evidenciando a sua utilidade prática para áreas como o marketing, as operações, o *revenue management* e o *Business Intelligence*.

Este trabalho procura, assim, articular duas dimensões complementares. Do ponto de vista académico, pretende oferecer evidência empírica que reforce o debate sobre métricas de desempenho no comércio digital no turismo, destacando o L2B como um objeto válido de análise. Do ponto de vista da gestão, a dissertação procura demonstrar que a análise do L2B fornece pistas úteis para a tomada de decisão estratégica, permitindo diagnosticar problemas de conversão, priorizar segmentos de clientes e apoiar a construção de modelos de previsão mais ajustados à realidade dos mercados digitais.

No que respeita à estrutura, a dissertação encontra-se organizada em sete capítulos. O presente capítulo que corresponde à introdução, o segundo capítulo onde se desenvolve a revisão da literatura. O terceiro capítulo descreve a metodologia utilizada. O quarto capítulo apresenta a exploração estatística inicial dos dados, enquanto o quinto dedica-se à modelação e validação estatística. O sexto capítulo apresenta a discussão dos resultados obtidos e as suas implicações práticas. Por fim, o sétimo capítulo sintetiza as principais conclusões, apresenta as contribuições académicas e de gestão, reconhece as limitações do estudo e apresenta as recomendações para estudos futuros.

II - Revisão da Literatura

2.1 – Conversão do Consumidor e o Conceito de Look to Book (L2B)

A conversão do consumidor, um conceito amplamente explorado no âmbito do marketing digital, refere-se ao processo através do qual os potenciais clientes realizam ações específicas que resultam numa compra efetiva (Xie & Lee, 2020; Moe & Fader, 2004).

Do ponto de vista do comportamento do consumidor, a literatura aponta para uma reorganização das etapas clássicas de decisão, como a consciencialização, a consideração, a intenção e a ação, as quais são mais vezes expostas em ambientes digitais. Vários autores têm sublinhado que, estas etapas de decisão são influenciadas por variáveis como a apresentação visual da oferta, a facilidade de navegação, os sistemas de recomendação personalizados, a reputação online e a compatibilidade com dispositivos móveis (Xie & Lee, 2020; Wu et al., 2024).

No setor da hospitalidade e turismo, este processo envolve uma sequência de etapas comportamentais que refletem o percurso do cliente, desde o primeiro contacto com a oferta até à realização da reserva. Segundo Xie e Lee (2020), esse percurso pode ser segmentado em quatro

fases, auxiliares da reorganização das etapas clássicas de decisão, sendo estas: a consciencialização, a consideração, a interação digital e a reserva.

Tabela 1 – Fases do processo de conversão digital no turismo

Fase	Descrição
1. Consciencialização	O consumidor toma conhecimento do produto ou serviço (ex.: hotel) através de motores de busca, redes sociais ou anúncios online.
2. Consideração	O consumidor compara diferentes ofertas com base em fatores como preço, avaliações, localização e comodidades.
3. Interação Digital	O consumidor realiza ações como clicar em anúncios ou visitar a página do hotel, o que demonstra um maior envolvimento com a oferta.
4. Reserva	O consumidor finaliza a transação através da realização da reserva, que leva à conclusão do processo de conversão.

Fonte: Xie e Lee (2020)

Estas fases foram também conceptualizadas por Moe e Fader (2004), que analisaram o comportamento dinâmico de conversão em ambientes de comércio eletrónico e demonstraram que a conversão resulta de um processo dinâmico de decisão, onde múltiplas variáveis contextuais e individuais influenciam o resultado final.

Embora os seus estudos se tenham centrado em plataformas genéricas de *e-commerce*, as dinâmicas descritas são altamente relevantes para o setor turístico, especialmente quando aplicadas a sistemas de reservas online. Neste enquadramento, de plataformas online no turismo, a conversão é o resultado de uma série de decisões e interações, influenciada pela perceção de valor por parte do cliente, pela facilidade de navegação da plataforma e pela clareza da informação.

É neste contexto que surge a necessidade de métricas específicas para avaliar a eficácia do processo de conversão. A métrica *Look to Book* (L2B), representa o rácio entre o número de pesquisas efetuadas por um utilizador (look) e o número de reservas concluídas (book), conforme representado na Equação 1.

$$L2B = \frac{N^{\circ} \text{ de Reservas}}{N^{\circ} \text{ de Pesquisas}} \quad (1)$$

Desta forma, um L2B elevado é desejável, já que traduz uma maior taxa de conversão, de interesse em ação, que reflete aspetos como a atratividade da oferta, a clareza da proposta de valor ou a qualidade da experiência digital proporcionada ao utilizador. Este desempenho positivo pode refletir uma proposta de valor bem alinhada com as expetativas do público-alvo.

Por outro lado, um rácio L2B baixo deve ser entendido como sinal de alerta, na medida em que pode indicar obstáculos entre a procura gerada e a conversão efetiva. Estes obstáculos podem decorrer de fatores, como preços pouco competitivos, falhas na experiência do utilizador, ou uma comunicação ineficaz da oferta disponível. Assim, o L2B torna-se um indicador sensível não só à atratividade da oferta, mas também à eficácia do processo digital de captação e conversão.

Ainda que um L2B elevado represente, em termos operacionais, uma maior eficácia na conversão, também pode refletir situações de baixa procura ou de limitação na variedade de opções apresentadas ao cliente (Xie & Lee, 2020). Sendo assim, a correta interpretação do L2B deve ser complementada com um enquadramento contextual que considere variáveis de oferta, procura e comportamento do consumidor, como volume total de tráfego, o tempo médio de navegação ou a taxa de rejeição. Esta abordagem permite uma compreensão mais contextualizada e potencia decisões mais informadas e eficazes.

Esta métrica tem sido utilizada por empresas do setor para monitorizar o desempenho das suas plataformas, em especial por intermediários como *bedbanks* (entidades grossistas que distribuem inventário de hotéis para outras agências), OTAs (*Online Travel Agencies* (ex.: Booking.com, Expedia)), bem como por plataformas de venda B2B (*Business-to-Business*), e B2C (*Business-to-Consumer*). Embora o termo não se encontre padronizado na literatura académica, a lógica subjacente ao L2B, enquadra-se na análise de métricas de conversão digitais adaptadas à realidade da indústria do turismo.

Importa esclarecer que o rácio L2B, na sua formulação base, corresponde à relação entre o número de reservas efetuadas e o número de pesquisas realizadas. Embora esta métrica não integre diretamente variáveis como datas de entrada e saída, destino, tipologia de quarto ou faixa de preço, o seu valor pode ser significativamente enriquecido quando analisado em articulação com esses

atributos. Esta abordagem cruzada permite compreender melhor os contextos em que o L2B é mais elevado, o que facilita a identificação de padrões de procura e de fatores que favorecem a conversão.

Embora existam outras métricas que avaliem a conversão, o L2B destaca-se pela sua simplicidade, aplicabilidade imediata no setor do turismo e potencial para ser desdobrado em análises segmentadas com grande valor operacional. Em contraste, no comércio eletrónico tradicional, a conversão é normalmente definida como o rácio entre visitantes e transações concluídas, sem necessariamente considerar os filtros e interações detalhadas ao longo do processo (Mariani, 2020).

Embora o termo *Look to Book* ainda não surja na literatura científica, vários estudos discutem conceitos similares sob outras denominações, como *search-to-booking conversion rate* ou *click-to-purchase ratio*. Por exemplo, Cezar e Ögüt (2015) realizaram uma análise ao impacto de fatores como avaliações de clientes, sistemas de recomendação e ordem de posicionamento nas listas de resultados, onde concluíram que estes elementos influenciam diretamente a taxa de conversão em plataformas de reserva online.

Adicionalmente, trabalhos como os de Buhalis e Law (2008) e Xiang et al. (2015), ajudam a compreender o papel crescente das tecnologias de informação no turismo, ao dar destaque à importância dos motores de busca, da apresentação visual da oferta e da qualidade dos sistemas de recomendação. Os autores Buhalis e Law (2008) e Xiang et al. (2015) alertam para o facto de o comportamento do consumidor digital ser fortemente influenciado por estímulos contextuais e pela arquitetura da informação, elementos que podem condicionar o sucesso da conversão e, conseqüentemente, afetar indicadores como o L2B.

Em síntese, o rácio *Look to Book* pode ser considerado um indicador prático do desempenho comercial digital, especialmente útil em plataformas de intermediação turística. A sua análise permite compreender melhor a eficácia das estratégias de captação e conversão de procura, bem como identificar eventuais fragilidades em cada etapa da experiência do consumidor. Contudo, apesar da sua utilização em contextos empresariais específicos, existe uma lacuna evidente na literatura científica no que respeita à análise quantitativa dos fatores que influenciam o L2B, com base em dados reais. Esta constatação justifica a necessidade de mais investigação, através de

abordagens estatísticas que explorem os determinantes desta métrica em ambientes operacionais concretos.

2.1.1 – Fatores Influentes na Taxa de Conversão

A taxa de conversão constitui um dos principais indicadores de desempenho no comércio digital turístico, ao evidenciar a eficácia na transformação de visualizações ou interações online em reservas concretas. Compreender os fatores que influenciam este indicador é fundamental para a otimização das estratégias comerciais das plataformas de distribuição turística.

A literatura científica tem identificado diversos elementos que condicionam a taxa de conversão nas plataformas digitais de turismo, apontando para uma multiplicidade de dimensões, sendo estas económicas, tecnológicas, percetivas e comportamentais, que, em conjunto, moldam as decisões dos consumidores ao longo do processo de reserva.

Entre os fatores económicos, o preço e a perceção de valor por parte do consumidor destacam-se como uns dos elementos mais influentes. Diversos estudos demonstram uma correlação negativa entre preços elevados e intenção de reserva, sobretudo em mercados com elevada transparência informacional (Mohapatra et al., 2023). No entanto, esta relação é modulada por variáveis como a reputação da unidade, a fidelização do cliente ou a qualidade percebida pelo mesmo, o que pode facilitar a tolerância a preços superiores desde que justificados por atributos valorizados (Wang et al., 2021).

As características do alojamento, como o tipo de quarto, o regime de alimentação, a vista ou até mesmo a presença de comodidades adicionais, influenciam significativamente o comportamento de reserva. A investigação de Tussyadiah e Zach (2017), evidencia que estas especificações são cuidadosamente analisadas pelos consumidores durante a fase de decisão, sendo determinantes na escolha final e na perceção de valor da oferta.

Outro fator bastante estudado é a reputação online, criada pelas avaliações e classificações de clientes anteriores. Avaliações positivas sobre o produto geram confiança, reduzem a perceção de risco e aumentam a probabilidade de conversão. Zhao et al. (2015) demonstram que a qualidade, quantidade e atualidade das avaliações têm impacto direto na intenção de reserva, sendo este efeito particularmente relevante em ambientes altamente competitivos.

A posição do alojamento nas páginas de resultados das plataformas online é, por sua vez, um fator que influencia significativamente a decisão de reserva. A análise comportamental revela que os utilizadores tendem a clicar nas primeiras opções exibidas, o que cria a percepção de que a sua visibilidade se deve à qualidade ou popularidade do alojamento (Masiero & Nicolau, 2015). Esta dinâmica tem implicações diretas tanto na estratégia comercial dos fornecedores como nos algoritmos de ordenação das plataformas.

Adicionalmente, a percepção de escassez e urgência, provocada por mensagens como “Último quarto disponível” ou “Reservado recentemente”, aumenta a pressão temporal e reduz a procrastinação. Tussyadiah e Zach (2017) referem que estas táticas contribuem para a aceleração da decisão de reserva, sobretudo em consumidores mais indecisos ou em contextos de alta procura.

Tendo em conta o panorama atual do consumo digital, é importante destacar a influência que o conteúdo visual exerce na taxa de conversão, uma vez que imagens e vídeos assumem um papel determinante na forma como o consumidor avalia e confere credibilidade à oferta apresentada. De acordo com El-Said (2020), concluiu-se que imagens de alta qualidade, coerentes com a descrição do alojamento, reforçam a credibilidade da oferta e aumentam a intenção de reserva, atuando assim como substitutos parciais da experiência tangível.

Finalmente, a influência das redes sociais tem-se afirmado como determinante no comportamento de reserva, ao influenciarem tanto a descoberta de destinos como a confiança nas escolhas dos consumidores. O conteúdo partilhado em plataformas digitais, como Instagram ou TikTok, contribui para a construção da imagem dos destinos e dos alojamentos, sobretudo entre os públicos mais jovens. O estudo de Petković et al. (2024), confirma que a presença ativa e positiva nas redes sociais tem um impacto significativo na intenção de visita e reserva.

Em síntese, a taxa de conversão nas plataformas digitais de turismo resulta da interação de um conjunto de fatores interdependentes, que vão desde dimensões objetivas, como o preço e os atributos do alojamento, até aspetos subjetivos, como a reputação online ou o conteúdo visual. A incorporação destes fatores numa lógica estratégica permite não apenas otimizar o desempenho individual de cada unidade de alojamento, mas também desenvolver políticas de distribuição mais eficazes, alinhadas com o comportamento do consumidor digital contemporâneo.

2.2 – Modelação Preditiva e *Business Intelligence*

A crescente disponibilidade de dados nas organizações tem vindo a reforçar práticas analíticas no apoio à tomada de decisão estratégica. Entre essas práticas, destacam-se a modelação preditiva e o *Business Intelligence* (BI), duas abordagens fundamentais na era da gestão orientada por dados. Em contextos de gestão, estas ferramentas assumem um papel cada vez mais central na monitorização de desempenho, alocação eficiente de recursos e antecipação de tendências de mercado. Como referem Sharda et al. (2020), estas metodologias possibilitam uma gestão mais proativa e fundamentada em evidência, crucial em ambientes cada vez mais competitivos e voláteis.

A modelação preditiva pode ser definida como a aplicação de métodos estatísticos e algoritmos para prever resultados futuros com base em padrões identificados em dados históricos (Wooldridge, 2019). Esta abordagem é versátil e pode ser aplicada a várias áreas da gestão, como o planeamento estratégico, a análise de risco, a previsão de rotatividade de clientes ou a otimização de operações. Na prática, esta técnica pode ser aplicada através de ferramentas como *Python*, R ou IBM SPSS, que permitem estimar procura, prever flutuações de receitas ou antecipar desvios de desempenho. Um exemplo relevante é o estudo de Chong et al. (2017), no qual modelos preditivos foram utilizados para antecipar vendas com base em sentimentos dos consumidores, avaliações online e estratégias promocionais, evidenciando a utilidade desta técnica na gestão comercial e de marketing.

De forma complementar, o *Business Intelligence* envolve um conjunto de processos, arquiteturas e tecnologias que visam transformar dados brutos em informação relevante para apoiar a gestão. O seu foco incide na recolha, integração e visualização interativa de dados, permitindo análises e decisões mais ágeis e baseadas em métricas concretas (Popovič et al., 2012). Esta orientação analítica favorece o desenvolvimento de culturas organizacionais *data-driven*, promovendo uma maior capacidade de resposta às exigências do mercado.

No plano operacional, plataformas como *Power BI*, *Tableau* ou *QlikSense* têm sido amplamente adotadas por gestores de diferentes setores, precisamente por facilitarem a criação de *dashboards* interativos, o acompanhamento de KPIs (*Key Performance Indicators*), a disseminação de insights e a tomada de decisões junto das equipas de gestão. Tal como referem Sharda et al. (2020), estas

ferramentas contribuem decisivamente para a democratização do acesso à informação e para o desenvolvimento de culturas organizacionais orientadas para a análise e a melhoria contínua.

No domínio do turismo, estas abordagens têm sido amplamente adotadas, impulsionadas pela digitalização dos canais de venda e pela crescente disponibilidade de dados transacionais e comportamentais. A modelação preditiva tem-se revelado particularmente útil na previsão de taxas de ocupação, receitas esperadas (RevPAR), níveis de procura e métricas de conversão, como o L2B.

Por outro lado, o BI tem permitido às organizações turísticas monitorizar a *performance* comercial, ajustar políticas de *pricing* e melhorar a experiência do cliente através da personalização da oferta (Fuchs et al., 2014; Del Vecchio et al., 2018). A capacidade de cruzar dados comportamentais, contextuais e transacionais torna estas ferramentas indispensáveis em ambientes altamente dinâmicos, como *bedbanks* ou agências online (OTAs), onde o volume e a variedade de dados são particularmente elevados.

É neste contexto que a métrica L2B adquire uma importância estratégica. Apesar de, na sua forma base, este rácio não incorporar diretamente variáveis como datas de entrada e saída, destino ou tipologia de quarto, o seu valor analítico aumenta significativamente quando combinado com atributos contextuais. Esta análise segmentada permite identificar contextos onde a taxa de conversão é mais elevada, facilitando decisões como o ajuste de preços, a reformulação da oferta ou o reforço de campanhas promocionais. A análise estatística do L2B através de modelos explicativos, como a regressão linear múltipla, permite estimar o impacto relativo de diferentes fatores no desempenho comercial, conforme a literatura econométrica (Gujarati & Porter, 2009; Wooldridge, 2019). Assim, a simplicidade operacional e a aplicabilidade imediata do L2B, tornam-no particularmente relevante em ambientes de intermediação digital, como os *bedbanks* ou as OTAs.

Porém, autores como Tussyadiah e Zach (2017) e Hair et al. (2019) reforçam a importância da análise exploratória de dados (EDA) como etapa prévia essencial para garantir a validade estatística dos modelos e a robustez das conclusões. A articulação entre EDA, modelação preditiva e BI garante uma base sólida para decisões fundamentadas em dados, reduzindo o risco de enviesamento e aumentando o valor prático das análises.

Por fim, destaca-se o potencial do L2B enquanto KPI digital. Quando incorporado em *dashboards* personalizados, pode ser cruzado com outros indicadores operacionais, como o tempo médio de navegação, taxa de rejeição ou duração da sessão, proporcionando uma visão mais integrada do comportamento do utilizador. O acompanhamento contínuo destes KPIs, através de ferramentas como *Power BI* ou *Tableau*, contribui para a criação de uma cultura de decisão orientada por dados (Del Vecchio et al., 2018), e estabelece uma ligação direta entre a análise quantitativa e a gestão estratégica, tanto no setor do turismo como na generalidade das organizações que operam em ambientes digitais.

2.2.1 – Métricas de Conversão e KPI nas Plataformas Digitais

Num cenário marcado pela crescente digitalização da economia, a utilização de indicadores de desempenho tornou-se fundamental para apoiar a tomada de decisão baseada em dados. Em particular, a definição e monitorização de Key Performance Indicators (KPIs) são práticas centrais da gestão contemporânea, funcionando como instrumentos que permitem alinhar objetivos organizacionais, acompanhar resultados operacionais e responder com agilidade às exigências de ambientes dinâmicos e competitivos (Sharda, Delen, & Turban, 2020). Neste sentido, os KPIs digitais surgem como uma extensão natural das métricas tradicionais, adaptando-se à realidade dos modelos de negócio baseados em plataformas tecnológicas, como é o caso do turismo online.

Na gestão, os KPIs podem assumir diferentes funções consoante a área funcional da organização. Na área financeira, métricas como a margem operacional ou o retorno do investimento continuam a ser amplamente utilizadas. Por outro lado, no domínio do marketing digital, indicadores como o *click-through rate*, o custo por aquisição ou o *engagement rate*, ganham destaque. Independentemente da sua natureza, a utilidade dos KPIs reside na sua capacidade de tornar a *performance* organizacional em dados mensuráveis, comparáveis e orientados para a melhoria contínua. A sua integração em sistemas de *Business Intelligence*, como *dashboards* analíticos interativos, permite às organizações uma monitorização mais eficaz e agir de forma proativa perante desvios ou oportunidades (Popović et al., 2012).

No setor do turismo, onde a experiência do consumidor e a competitividade da oferta são fatores críticos, os KPIs têm vindo a assumir um papel de destaque. Estas métricas medem a eficácia com que uma plataforma transforma o interesse dos utilizadores, traduzido em pesquisas, cliques ou visualizações, em ações concretas, como reservas. Entre estas métricas, destaca-se o L2B, um KPI

particularmente relevante em contextos de elevada intermediação digital, como *bedbanks*, OTAs ou plataformas B2B.

Embora o cálculo do L2B seja relativamente simples, a análise combinada entre este rácio e variáveis complementares, permite identificar padrões de conversão, diagnosticar potenciais obstáculos à reserva e ajustar campanhas promocionais ou estratégias de *pricing*. Como salientam Xiang et al. (2015), a análise cruzada entre métricas de conversão e dados comportamentais favorece decisões mais ágeis e bem fundamentadas, com impacto direto na proposta de valor digital apresentada ao consumidor.

Simultaneamente, métricas clássicas do desempenho hoteleiro continuam a desempenhar um papel relevante na gestão operacional, nomeadamente:

- Taxa de Ocupação: percentagem de quartos vendidos face ao total disponível;
- ADR (Average Daily Rate): receita média por quarto vendido;
- RevPAR (Revenue per Available Room): receita média por quarto disponível, resultante da conjugação entre a taxa de ocupação e o ADR.

Embora estas métricas sejam eficazes na avaliação do desempenho operacional, a sua utilidade em ambientes puramente digitais é limitada. Por isso, a reunião entre KPIs digitais de conversão, como o L2B, e indicadores operacionais clássicos permite uma leitura mais integrada do desempenho, ao cruzar indicadores pré-reserva com resultados financeiros reais (Sigala, 2018; Del Vecchio et al., 2018).

Para maximizar o valor dos KPIs, a sua interpretação deve ser feita em função do posicionamento estratégico da organização. Por exemplo, *bedbanks* ou intermediários digitais beneficiam particularmente da análise do L2B, pois variações neste rácio podem indicar problemas de disponibilidade, falhas no conteúdo ou preços pouco competitivos. Já operadores hoteleiros com venda direta poderão continuar a privilegiar o ADR ou o RevPAR, especialmente quando integrados com KPIs digitais que reflitam a jornada do consumidor.

A literatura sugere que a eficácia da monitorização de KPIs depende da capacidade tecnológica da organização e da integração dos indicadores em sistemas analíticos (Chong et al., 2017). Plataformas como *Power BI* ou *Tableau*, são hoje amplamente utilizadas para visualizar e cruzar

dados provenientes de diferentes fontes, visualizar *insights* relevantes de forma intuitiva e acionável e reforçar a cultura organizacional orientada por dados (Sharda et al., 2020). A integração de métricas como o L2B nestas plataformas permite não apenas antecipar desvios, mas também personalizar a oferta, otimizar a experiência digital e alinhar decisões estratégicas com os objetivos organizacionais.

Em síntese, os KPIs e métricas de conversão digital representam instrumentos indispensáveis para uma gestão eficaz, informada e orientada por dados. A sua articulação com sistemas de *Business Intelligence*, bem como a sua análise cruzada com variáveis contextuais, permite às organizações atuar com maior proatividade, rigor e visão estratégica. Esta lógica aplica-se tanto à gestão no setor do turismo como às restantes áreas de negócio, uma vez que as implicações e boas práticas associadas a estes indicadores são transversais a outras áreas da gestão, servindo de ponte natural para a abordagem dos desafios específicos enfrentados pelas empresas no ecossistema online.

III - Metodologia

3.1 – Abordagem Metodológica do Estudo

A presente investigação é de natureza quantitativa e tem como objetivo analisar de que forma diferentes fatores influenciam o rácio *Look to Book* (L2B), com auxílio de dados disponibilizados pela empresa Viagens Abreu. A investigação procura compreender de que forma variáveis como o preço, o tipo de quarto ou o plano de refeições, influenciam o desempenho de conversão medido através do L2B.

A estratégia metodológica adotada parte de conhecimento previamente explorado na literatura sobre conversão digital, comportamento do consumidor e métricas de performance no turismo (Kotler et al., 2021; Petković et al., 2024), com o objetivo de validar empiricamente a relação entre variáveis independentes e a taxa de conversão observada.

Trata-se de um estudo transversal, uma vez que os dados utilizados foram recolhidos num período temporal específico, o mês de dezembro de 2024, o que proporciona uma representação estática das dinâmicas do fenómeno em análise nesse momento. Adicionalmente, este é um estudo aplicado, uma vez que utiliza dados reais provenientes da empresa e procura gerar conhecimento com aplicabilidade prática à realidade empresarial. Os resultados alcançados poderão constituir

um contributo relevante para a gestão, orientando a definição de estratégias de melhoria das empresas.

A metodologia de análise baseia-se em modelação estatística, que será conduzida através de um modelo de regressão linear múltipla (*Ordinary Least Squares* - OLS), que permite estimar a relação entre a variável dependente (L2B) e um conjunto de variáveis explicativas selecionadas. A escolha deste modelo deve-se à sua ampla aplicação em estudos na área da Gestão e do Marketing (Hair et al., 2019; Wooldridge, 2020), bem como à sua capacidade para captar relações lineares entre variáveis contínuas e categóricas.

3.2 – Recolha e Fontes de Dados

Os dados utilizados neste estudo foram obtidos a partir de registos internos da empresa Viagens Abreu, mais especificamente da área de *Business Optimization*, com foco na análise da performance comercial dos clientes que utilizam a plataforma digital da empresa para efetuar pesquisas e reservas de alojamento.

A escolha desta organização como objeto de estudo justifica-se pela disponibilidade de dados internos, que permitem uma análise aprofundada sobre o comportamento do cliente e os fatores influentes no rácio L2B. Ao mesmo tempo, o conhecimento obtido neste estudo pode gerar benefícios diretos para a gestão estratégica e tomada de decisão da empresa.

A recolha de dados foi realizada durante o mês de dezembro de 2024, a seleção deste mês visa garantir uma amostra com volume representativo de pesquisas e reservas, devido à conjugação de reservas de fim de ano, campanhas promocionais e dinâmicas específicas da época natalícia.

Os dados analisados foram obtidos a partir de dois tipos de ficheiros operacionais:

- Conjunto de 64 ficheiros CSV com dados sobre as reservas dos 1000 hotéis com maior número de reservas, incluindo variáveis como preço da reserva, categoria do quarto e tipo de regime alimentar incluído. A partir dele foi formado o *dataset de reservas*, cujas variáveis são apresentadas na Tabela 2.
- Relatório mensal de atividade, que apresenta o número de pesquisas realizadas por cada hotel. A partir deste foi construído o *dataset de pesquisas*, cujas variáveis são apresentadas na Tabela 3.

Tabela 2 – Variáveis presentes no *dataset* de reservas

Variável	Descrição
<i>Code</i>	Representa o código que identifica um hotel específico por fornecedor, ou seja, o mesmo hotel pode ter vários códigos diferentes porque pode ser oferecido por vários fornecedores.
<i>Hotel</i>	Representa o nome do hotel como aparece em sistema.
<i>Master</i>	Representa o código que identifica um único hotel, independentemente do seu fornecedor.
<i>Room</i>	Representa a categoria de quarto, de acordo com o sistema.
<i>Meal Plan</i>	Representa o tipo de regime alimentar incluído na reserva.
<i>Status</i>	Representa o estado em que a reserva se encontra.
<i>Rate</i>	Representa o valor do quarto na reserva.
<i>In</i>	Representa a data de check-in do cliente.
<i>Out</i>	Representa a data de check-out do cliente.

Fonte: Autor (2025)

Tabela 3 – Variáveis presentes no *dataset* de pesquisas

Variável	Descrição
<i>Code</i>	Representa o código que identifica um hotel específico por fornecedor, ou seja, o mesmo hotel pode ter vários códigos diferentes porque pode ser oferecido por vários fornecedores.
<i>Name</i>	Representa o nome do hotel como aparece em sistema.
<i>Master</i>	Representa o código que identifica um único hotel, independentemente do seu fornecedor.
<i>City</i>	Representa a cidade onde localiza o hotel.
<i>Country</i>	Representa a sigla do país onde se localiza o hotel.
<i>Globally Requested</i>	Representa o número de pesquisas por “Code”.

Fonte: Autor (2025)

Para assegurar o cálculo da métrica L2B, procedeu-se à integração dos dois conjuntos de dados, um que contém o total mensal de pesquisas (*dataset* de pesquisas) e outro que contém as reservas

individuais realizadas ao longo do mesmo período (*dataset* de reservas). As reservas foram previamente agregadas por hotel, que permitiu a combinação com os dados de pesquisas. A junção foi realizada com base na variável comum “*Code*”.

A partir do conjunto de dados integrado (*dataset* de análise), foi calculada a métrica *Look to Book* (L2B), definida como o rácio entre o número de pesquisas e o número de reservas efetuadas por cada hotel, ilustrado na Equação 1.

Todo o processamento dos dados foi realizado em Excel e em *Python*, onde se utilizou o ambiente *Jupyter Notebook* e as bibliotecas *Pandas* e *NumPy*, garantindo rastreabilidade, reprodutibilidade e eficiência no tratamento de grandes volumes de informação.

Na secção seguinte, será apresentado em detalhe o processo de preparação dos dados e a forma como os dois *datasets* foram integrados, incluindo as etapas de recodificação das categorias de quartos, agregação e junção.

3.3 – Preparação dos Dados

3.3.1 – Recodificação das Categorias de Quarto

Durante o processo de preparação dos dados, verificou-se que a variável “*Room*”, associada ao tipo de quarto, apresentava uma elevada diversidade de categorias, resultante de formatos distintos, idiomas variados e diferentes níveis de detalhe. Esta situação resultava numa falta de uniformidade que dificultava a análise comparativa entre unidades hoteleiras e impedia a utilização direta da variável nas fases seguintes do estudo.

De modo a ultrapassar esta limitação, procedeu-se à atribuição de novos rótulos às categorias de quarto, com o objetivo de reduzir a granularidade. Esta recodificação foi feita com base na leitura de padrões frequentes na variável “*Room*”, através de nuvens de palavras ou gráficos de barras.

Esta recodificação, resultou na criação de quatro categorias agregadas, suficientemente distintas entre si, mas que preservam a lógica comercial das tipologias analisadas: *Room*, *Deluxe*, *Non-Refundable Room* e Outros.

As categorias foram definidas da seguinte forma:

- *Room*: Inclui quartos identificados como Single Room, Standard Room, Double Room, Triple Room ou variações semelhantes, desde que não contenham elementos associados a luxo ou serviços premium, como *jacuzzi*, piscina, vista mar, vista jardim, ou variações próximas. Esta categoria foi associada à tipologia mais simples e habitual.
- *Deluxe*: Inclui todos os quartos com descrições como Deluxe, King, Queen, Suite, Superior, bem como aqueles que mencionam explicitamente algum elemento de conforto superior ou luxo, tais como *jacuzzi*, *whirlpool*, piscina, vista mar, vista jardim, vista piscina, menções ao regime *All Inclusive*, ou designações próximas. Esta categoria foi associada ao segmento de quartos de valor mais elevado.
- *Non-Refundable Room*: Inclui todos os quartos cuja descrição mencione explicitamente “Non-Refundable” ou variações próximas. Esta tipologia foi interpretada como representativa de ofertas mais económicas, uma vez que estas condições estão habitualmente associadas a preços mais baixos.
- Outros: Categoria residual atribuída a descrições de quartos que não se enquadravam em nenhuma das anteriores, como por exemplo *Apartment*, *Bungalow*, ou outras designações menos frequentes e também aos valores nulos encontrados neste campo.

O mapeamento foi realizado em *Python*, com base em regras textuais aplicadas às descrições originais da variável.

3.3.2 – Agregação do *Dataset* de Reservas

Conforme descrito previamente, os dados de reservas utilizados foram recolhidos a partir de um conjunto de 64 ficheiros CSV extraídos ao longo do mês de dezembro de 2024. Cada ficheiro incluía registos ao nível da reserva individual, com diversas variáveis operacionais.

A primeira etapa do tratamento consistiu em juntar todos esses ficheiros num único *dataset*, recorrendo ao *Python* e utilizando a biblioteca *pandas*, para poder agrupar todos os ficheiros. Após

a consolidação dos dados brutos, foi necessário agregá-los ao nível mensal por hotel, de forma a alinhar a granularidade com o dataset de pesquisas. Para garantir essa compatibilidade entre ambas as fontes, utilizou-se a variável “*Code*” como chave de agregação, por ser o identificador único de cada hotel e a única variável, presente nos dois conjuntos de dados.


Durante o processo de agregação, aplicaram-se diferentes estratégias de tratamento consoante o tipo de variável. No caso da variável “*Rate*” (preço por noite), foi calculada a média mensal por hotel. Para isso, somaram-se todos os valores de “*Rate*” associados a reservas confirmadas de cada hotel (“*Code*”), dividindo-se esse total pelo número de ocorrências válidas, criando-se assim a variável “*Avg_Rate*”.

Para além da média do preço por noite, foi também contabilizado o número total de reservas confirmadas por cada hotel ao longo do mês de dezembro. Esta contagem corresponde ao número de vezes que cada código de hotel (“*Code*”) surgiu nos registos, o que permitiu criar a variável denominada “*N_Reservas*”. Esta variável representa o número de reservas por hotel no período em análise, sendo essencial para análises complementares como, por exemplo, o cálculo do rácio *Look to Book* (L2B).

As variáveis categóricas, como o tipo de quarto (*Room*) e o plano de refeições (*Meal Plan*), foram transformadas ao converter cada categoria em variáveis indicadoras binárias. Posteriormente, contabilizaram-se as frequências absolutas por hotel, permitindo caracterizar a oferta por unidade hoteleira. No caso da variável “*Room*”, após a recodificação em quatro categorias, como *Room*, *Deluxe*, *Non-Refundable Room* e *Outros*, cada uma foi representada como coluna distinta no dataset, sendo as ocorrências mensais agregadas por hotel (*Code*). O mesmo procedimento foi aplicado à variável “*Meal Plan*”, originando colunas correspondentes às categorias originais: *All Inclusive* (AI), *Bed & Breakfast* (BB), *Full Board* (FB), *Half Board* (HB) e *Room Only* (RO).

O resultado final desta fase foi um ficheiro único, onde cada linha representa um hotel distinto, com os principais atributos agregados, ao nível mensal, e preparados para posterior junção com o *dataset* de pesquisas, tal como se pode verificar na Figura 3.

Figura 3 – Representação da transformação dos dados



Code	Room	Meal Plan	Rate
1521782	313	Room	BB 188.8
43321	313	Room	BB 213.6
1902477	313	Room	BB 213.6
1464309	313	Room	BB 356
99027	313	Room	BB 285.6
1235546	313	Room	BB 284.8
2146803	313	Room	BB 285.6
916772	313	Room	BB 213.6
478167	313	Room	BB 285.6
180450	313	Room	BB 476

Code	Avg_Rate	N_Reservas	Deluxe	Non-Refundable Room	Outros	Room	AI	BB	FB	HB	RO
0	313	244.283333	288	0	0	0	288	0	288	0	0
1	346	267.574574	258	0	0	0	258	0	258	0	0
2	2382	419.620553	253	0	205	0	48	0	253	0	0
3	4402	537.750198	253	0	0	0	253	0	253	0	0
4	4890	210.000000	8	0	0	0	8	0	0	0	8
5	5754	267.235314	207	0	0	0	207	0	0	0	207
6	11832	289.304663	208	6	0	0	202	0	29	0	179
7	11833	327.248462	208	0	0	0	208	0	27	0	181
8	11836	452.603542	288	5	0	0	283	0	27	0	261
9	11898	337.169447	253	0	253	0	0	0	0	0	253

Fonte: Autor (2025)

3.3.3 – Construção do *Dataset de Análise: Junção dos Datasets*

Após a transformação e agregação das variáveis provenientes do dataset de reservas, foi necessário integrar esta informação com os dados relativos às pesquisas efetuadas por cada hotel. Estes dados, essenciais para o cálculo do rácio L2B, encontravam-se num ficheiro distinto, o dataset de pesquisas, que apresentava, à escala mensal, o número total de pesquisas por hotel durante o mês de dezembro de 2024.

Para alinhar os conjuntos de dados, procedeu-se à junção dos ficheiros pela variável *Code*, que identifica unicamente cada hotel. Esta integração permitiu combinar o volume de reservas com o número de pesquisas correspondentes, o que garantiu consistência temporal e granularidade. Com a base unificada, calculou-se a variável dependente do estudo, o rácio *Look to Book (L2B)*, indicador obtido através da divisão do número de reservas pelo número de pesquisas por hotel, uma métrica central no setor do turismo.

Por fim, foi realizada uma verificação final da coerência dos dados, para assegurar a ausência de registos com pesquisas iguais a zero, de forma a evitar divisões inválidas, e garantir a integridade de todos os campos necessários à análise. A base final, ficou constituída por variáveis numéricas agregadas, indicadores sobre o tipo de produto oferecido (como o tipo de quarto e o plano de refeições), valores médios por hotel (como o preço por noite) e a variável dependente L2B, pronta para ser utilizada na fase de modelação estatística.

3.3.4 – Definição das Variáveis

Após todo o processo de consolidação, limpeza e preparação dos dados, foi possível definir o conjunto final de variáveis que servem de base à análise estatística. Estas variáveis refletem não só os aspetos quantitativos ligados à procura e ao preço, mas também as características qualitativas dos produtos turísticos oferecidos pelos hotéis, como o tipo de quarto e o plano de refeições.

A Tabela 4 resume todas as variáveis utilizadas na análise e descreve o seu significado, tipo e função no estudo:

Tabela 4 - Variáveis utilizadas na análise

Nome da Variável	Tipo	Descrição	Papel na Análise
<i>Code</i>	Categórica	Identificador único de cada hotel na base de dados.	Identificador / Agrupador
<i>Avg_Rate</i>	Numérica	Preço médio por noite por hotel, calculado com base nas reservas com estado confirmado.	Variável explicativa
L2B	Numérica	Rácio entre o número total de reservas e o número total de pesquisas por hotel.	Variável dependente
<i>Room</i>	Numérica	Número de ocorrências de quartos simples ou normais por hotel.	Variável explicativa
<i>Deluxe</i>	Numérica	Número de ocorrências de quartos com características de luxo ou gama superior.	Variável explicativa
<i>Non-Refundable_Room</i>	Numérica	Número de ocorrências de quartos classificados como não reembolsáveis.	Variável explicativa
Outros	Numérica	Número de ocorrências de categorias de quarto que não se enquadram nas classificações anteriores.	Variável explicativa

<i>All Inclusive</i> (AI)	Numérica	Número de ocorrências de reservas com regime de tudo incluído.	Variável explicativa
<i>Bed and Breakfast</i> (BB)	Numérica	Número de ocorrências de reservas com pequeno-almoço incluído.	Variável explicativa
<i>Full Board</i> (FB)	Numérica	Número de ocorrências de reservas com pensão completa.	Variável explicativa
<i>Half Board</i> (HB)	Numérica	Número de ocorrências de reservas com meia pensão.	Variável explicativa
<i>Room Only</i> (RO)	Numérica	Número de ocorrências de reservas sem qualquer plano de refeições incluído.	Variável explicativa

Fonte: Autor (2025)

Este conjunto de variáveis foi escolhido por representar de forma equilibrada a combinação entre volume de procura, características do produto e opções tarifárias. Constitui, assim, a base sobre a qual será desenvolvida a modelação estatística apresentada na próxima secção.

3.4 – Modelo Estatístico

Para compreender quais os fatores que influenciam o rácio L2B, recorreu-se à aplicação de um modelo de regressão linear múltipla. Esta técnica estatística permite avaliar o efeito conjunto de várias variáveis independentes sobre uma variável dependente contínua, tornando-se particularmente útil em contextos onde se pretende compreender relações complexas entre fenómenos.

Segundo Hair et al. (2019), a regressão linear é uma das técnicas mais utilizadas em estudos de gestão e análise de desempenho, uma vez que permite estimar a força, direção e significância estatística dos efeitos de diversos preditores em simultâneo, o que possibilita a identificação de fatores mais relevantes para a explicação do fenómeno em análise. No presente estudo, o objetivo é avaliar, em que medida, determinados atributos relacionados com os hotéis, as reservas e as ofertas disponíveis influenciam a conversão, medida através do rácio L2B.

A estimativa do modelo foi realizada através do método dos Mínimos Quadrados Ordinários (MQO), mais conhecido como *Ordinary Least Squares* (OLS), que constitui o procedimento

estatístico mais comum em modelos de regressão. Esta abordagem estatística clássica procura ajustar a melhor linha possível aos dados, minimizando a soma dos quadrados dos resíduos, ou seja, diminuindo o total dos erros entre os valores observados e os valores previstos pelo modelo. Trata-se, assim, de um método conhecido pela sua robustez e simplicidade, e amplamente validado na literatura, próprio para estudos com dados contínuos e múltiplos preditores (Wooldridge, 2020).

No entanto, para que os coeficientes estimados por MQO sejam estatisticamente válidos, é necessário que determinados pressupostos fundamentais da regressão linear sejam respeitados. Estes pressupostos, amplamente reconhecidos na literatura econométrica (Gujarati & Porter, 2009; Wooldridge, 2020), são também conhecidos como condições do teorema de *Gauss-Markov*, asseguram que os estimadores obtidos sejam os melhores, lineares e não-viesados (BLUE – *Best Linear Unbiased Estimators*). Estes pressupostos são os seguintes:

- Linearidade do modelo: Parte-se do princípio de que existe uma relação linear entre a variável dependente e cada uma das variáveis independentes. Caso contrário, o modelo poderá estar mal especificado.
- Normalidade dos resíduos: Os resíduos (erros do modelo) devem seguir uma distribuição aproximadamente normal. Esta condição é importante para garantir a validade dos testes de significância estatística aplicados aos coeficientes.
- Homoscedasticidade: A variância dos resíduos deve ser constante para todos os níveis das variáveis independentes. A violação desta condição (heterocedasticidade) compromete os intervalos de confiança e os testes de hipóteses.
- Ausência de multicolinearidade: As variáveis independentes não devem apresentar fortes correlações entre si. A presença de multicolinearidade (existência de relações lineares entre preditores) dificulta a interpretação dos coeficientes e pode comprometer a estabilidade das estimativas.

Na prática, é frequente que dados reais, sobretudo dados operacionais recolhidos em contexto empresarial, não cumpram na íntegra todos os pressupostos teóricos exigidos pela regressão linear. De acordo com Gujarati e Porter (2009), os dados reais raramente seguem o “cenário ideal”, logo o desafio de aplicar modelos estatísticos a dados empíricos reside justamente na identificação e mitigação destas limitações, de forma a garantir a validade das conclusões.

Quanto à significância global do modelo, esta foi avaliada através da estatística F, em vez da tradicional Análise de Variância (ANOVA). Importa referir que a estatística F é conceptualmente e matematicamente equivalente à ANOVA global, uma vez que ambas comparam a variância explicada pelo modelo com a variância residual, ao testar se, em conjunto, as variáveis independentes contribuem significativamente para explicar a variabilidade da variável dependente. Este teste global constitui um critério fundamental para justificar a utilização de qualquer modelo de regressão como estrutura válida para fins interpretativos e preditivos (Field, 2018; Gujarati & Porter, 2009; Wooldridge, 2016).

No contexto da regressão linear múltipla, a estatística F testa a hipótese nula conjunta de que todos os coeficientes associados às variáveis independentes são iguais a zero, isto é:

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_k = 0$$

Importa sublinhar que o valor absoluto da estatística F, por si só, não constitui critério suficiente para avaliar a validade global do modelo. Desta forma, o elemento decisivo é o *p-value* associado, que indica se a variância explicada pelo modelo é significativamente superior à variância não explicada. Somente quando o *p-value* associado à estatística F se encontra abaixo do nível de significância definido ($\alpha = 0,05$) se pode concluir que o modelo, no seu conjunto, é estatisticamente significativo e adequado para fins interpretativos e preditivos.

Ao longo deste estudo, a estatística F foi utilizada em alternativa à ANOVA como critério global de validação dos modelos de regressão, de forma a garantir consistência metodológica e evitar redundância na apresentação dos resultados.

IV – Análise e Exploração Estatística de Dados

A análise exploratória de dados (EDA – *Exploratory Data Analysis*) representa uma etapa fundamental em qualquer investigação quantitativa, que permite uma compreensão mais clara da estrutura e das características dos dados, antes da aplicação de técnicas estatísticas mais complexas. No âmbito da regressão linear múltipla, a EDA assume um papel particularmente relevante, pois possibilita a deteção de padrões, tendências, *outliers* e assimetrias que, se ignorados, podem comprometer a validade dos pressupostos do modelo.

De acordo com autores como Hair et al. (2019) e Field (2018), a qualidade das inferências estatísticas depende fortemente da adaptação aos pressupostos exigidos pelos modelos utilizados.

A EDA constitui, por isso, um instrumento essencial não apenas para o diagnóstico de eventuais distorções ou desequilíbrios, mas também para a tomada de decisões informadas sobre transformações, limpeza ou ajustamentos necessários à robustez da análise.

Neste capítulo, procede-se a uma exploração estatística cuidadosa dos dados, de forma a garantir que a base utilizada na modelação preditiva apresenta as condições ideais para gerar resultados fiáveis e bem fundamentados.

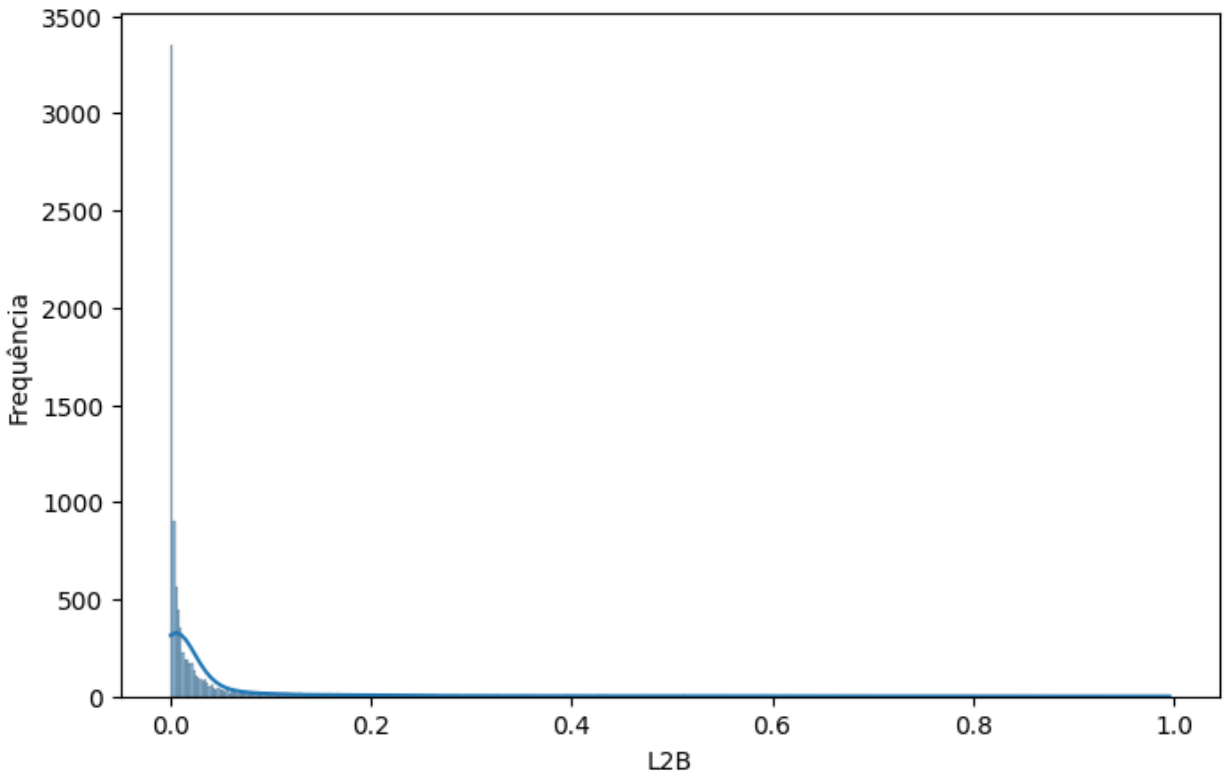
4.1 – Fase 1: Análise Exploratória Inicial

A primeira fase da Análise Exploratória de Dados (EDA) teve como objetivo compreender o comportamento da variável dependente e das variáveis explicativas presentes no estudo, com o propósito de diagnosticar padrões, assimetrias, *outliers* e eventuais violações dos pressupostos exigidos pela regressão linear múltipla. A análise combinou abordagens gráficas (histogramas, *boxplots*, matrizes de correlação e *scatterplots*) e numéricas (estatísticas descritivas), permitindo obter um retrato fiel da estrutura e qualidade dos dados.

4.1.1 Distribuição da Variável Dependente (L2B)

A análise da variável dependente (L2B), revelou uma distribuição significativamente assimétrica à direita, com uma concentração massiva de observações muito próximas de zero, uma cauda longa e valores pontuais a aproximar-se de 1, tal como se pode observar na Figura 4.

Figura 4 – Distribuição do L2B



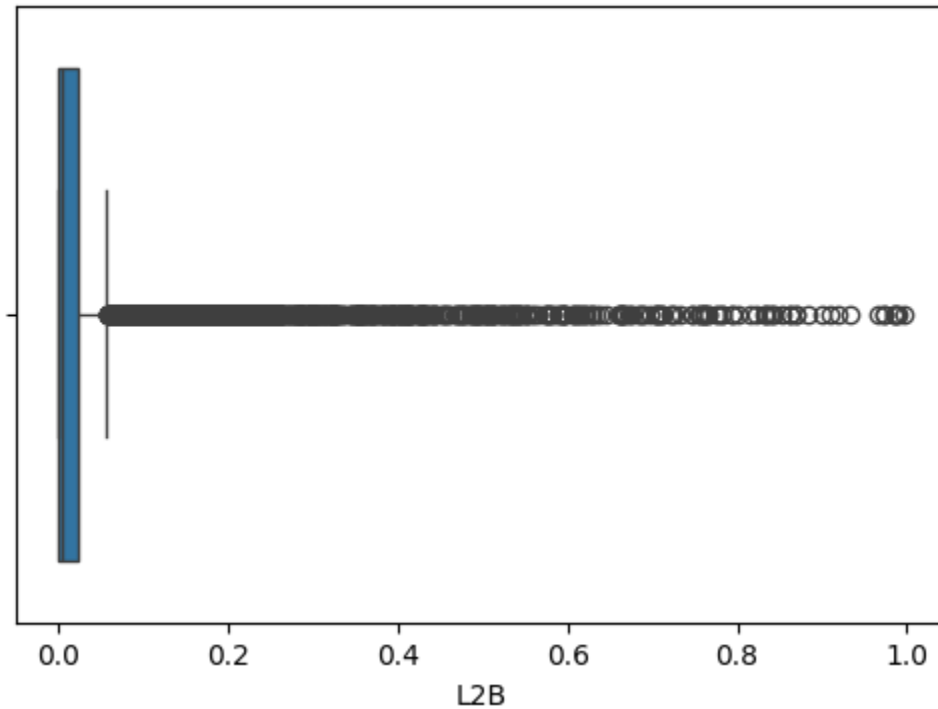
Fonte: Autor (2025)

Este padrão representa uma realidade expectável no setor de vendas online, onde a maioria das pesquisas não resulta em reservas efetivas.

Os valores da análise descritiva confirmam esta leitura, uma vez que a média do L2B é de apenas 0,041, enquanto a mediana, ainda mais baixa (0,005), revela a predominância de taxas de conversão muito reduzidas. O desvio padrão é relativamente elevado, cerca de 0,109, o que demonstra uma dispersão significativa face à média, e o valor máximo de 0,996 confirma a existência de casos pontuais de conversão quase total, que constituem *outliers* estatísticos.

Por sua vez, o *boxplot*, representado na Figura 5, reforça visualmente esta evidência, que mostra múltiplos *outliers* na cauda superior da distribuição.

Figura 5 – Boxplot do L2B



Fonte: Autor (2025)

Adicionalmente, o cálculo dos coeficientes de assimetria (*skewness*) e curtose reforça a necessidade de tratamento estatístico da variável, apresentando os seguintes resultados:

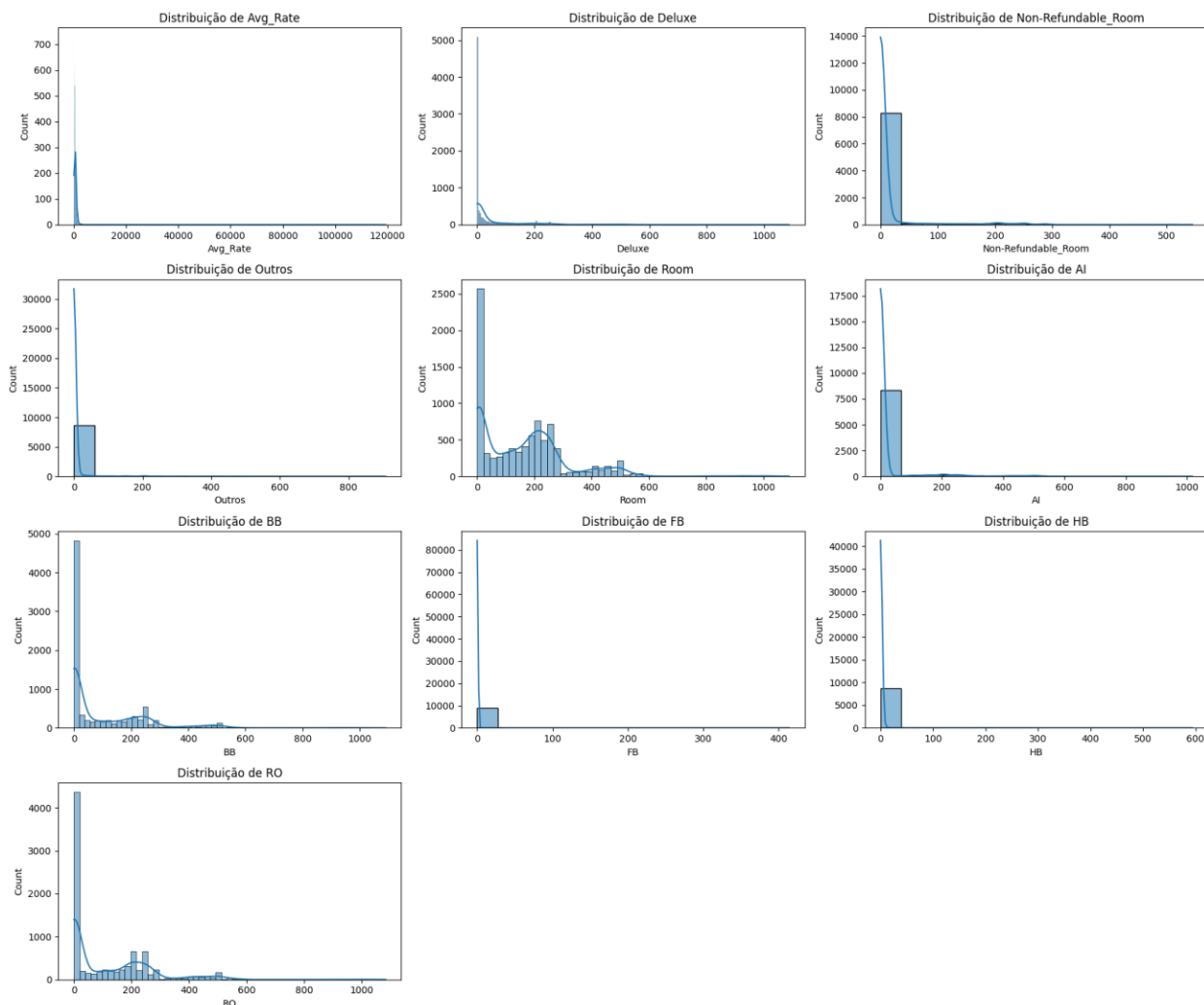
- *Skewness*: 4.533 - Valor substancialmente superior a zero, o que indica uma distribuição bastante enviesada à direita, ou seja, a maioria dos valores encontra-se concentrada em níveis baixos, com uma cauda longa que se estende para valores mais elevados;
- *Curtose*: 24.111 - Valor significativamente superior a 3, o que caracteriza uma distribuição leptocúrtica, ou seja, com caudas mais pesadas do que uma distribuição normal e maior incidência de valores extremos (*outliers*).

Desta forma, estas características comprometem a normalidade e a homocedasticidade necessárias à regressão linear, tornando previsível a necessidade de aplicar, mais adiante, uma transformação à variável dependente L2B.

4.1.2 Distribuição das Variáveis Independentes

A avaliação das variáveis explicativas revelou padrões distintos, mas com uma característica comum, distribuições assimétricas e presença de valores extremos. O histograma, que se pode observar na Figura 6, ilustra este fenómeno para variáveis como *Avg_Rate*, *Room*, *Deluxe*, e os diferentes regimes de refeição (BB, RO, AI, etc.).

Figura 6 – Histogramas das Variáveis Explicativas

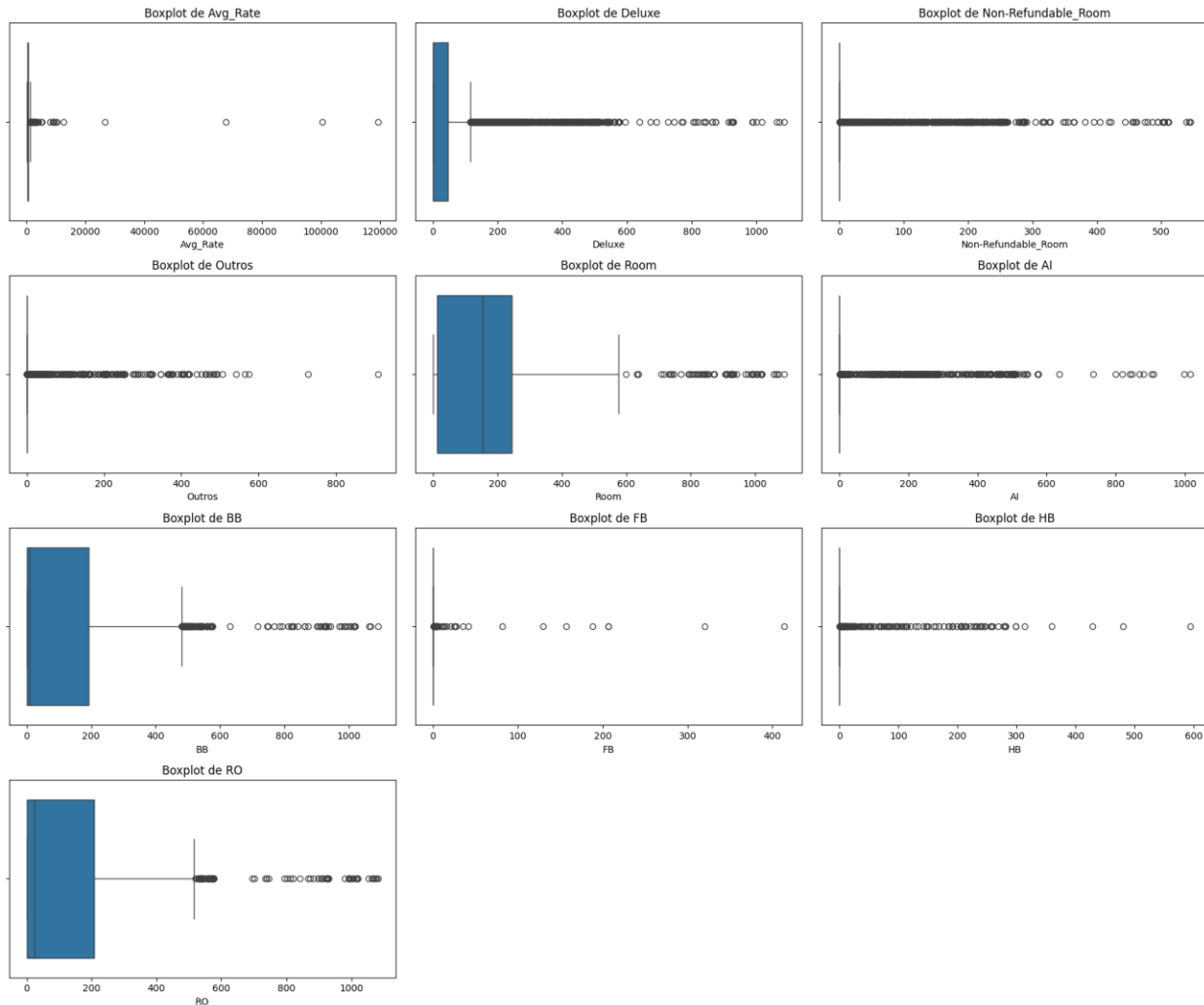


Fonte: Autor (2025)

A variável *Avg_Rate* (preço médio por noite) apresenta uma média de 611,94 euros, mas com um valor máximo extremamente elevado de 119.348,80 euros, o que evidencia *outliers* de grande magnitude. No entanto, a mediana (474,75 euros) é substancialmente inferior à média de 611.94

euros, sugerindo enviesamento positivo. O histograma (Figura 6), mostra uma concentração acentuada de valores baixos, enquanto o *boxplot* confirma a presença de *outliers* significativos, tal como se pode observar na Figura 7.

Figura 7 – Boxplots das Variáveis Explicativas



Fonte: Autor (2025)

Este padrão reforça a necessidade de aplicar uma transformação de suavização à variável *Avg_Rate*, de forma a reduzir a assimetria e minimizar o impacto de valores extremos na modelação.

Por outro lado, variáveis como *Deluxe* e *Room* a assimetria é igualmente evidente. No caso de *Deluxe*, a média é de 55,85 quartos, mas o valor máximo atinge 1.086, e a mediana é de apenas 1

quarto, que demonstra que a maioria dos hotéis não oferece esta tipologia em larga escala, enquanto alguns poucos têm uma disponibilidade muito acima da média. A variável *Room* segue um comportamento semelhante, com uma média de 161,82 quartos, uma mediana de 154 e um valor máximo de 1.090, refletindo diferenças significativas na capacidade entre unidades hoteleiras.

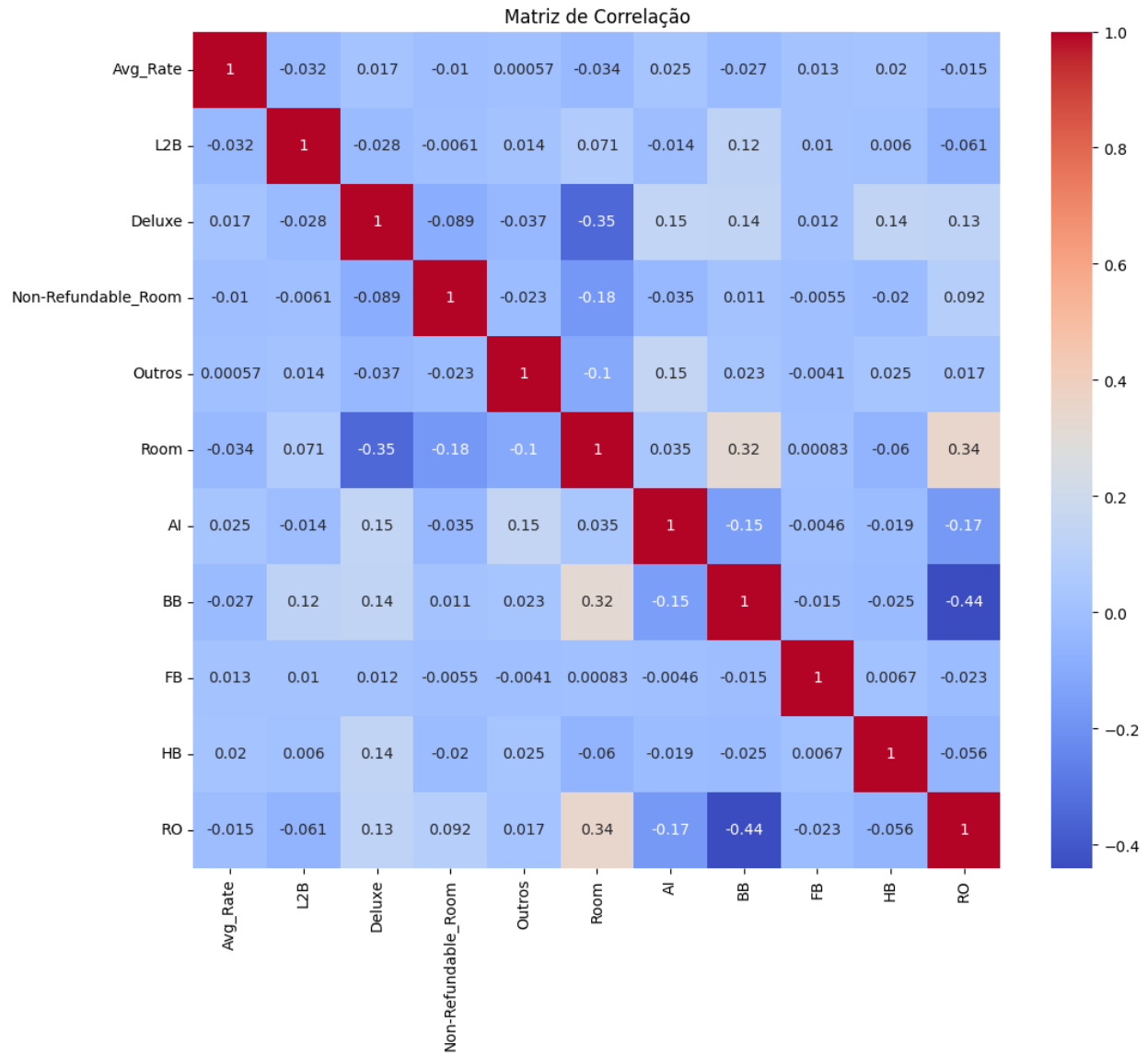
Quanto às variáveis associadas a regimes de refeição (AI, BB, FB, HB, RO), a assimetria é ainda mais evidente. Por exemplo, a variável BB apresenta uma média de 99,13 quartos, mas tem uma mediana de apenas 8 e um máximo de 1.090, enquanto a variável RO tem uma média de 117,42, uma mediana de 24 e um máximo de 1.082. Estes resultados revelam concentrações acentuadas em torno de zero ou valores baixos, acompanhadas de *outliers* significativos, reflexo de que nem todos os hotéis oferecem estes regimes de forma generalizada.

Com o auxílio visual da Figura 7, confirma-se a presença de *outliers*, o padrão comum de forte dispersão e alta assimetria em praticamente todas as variáveis. Esta heterogeneidade sugere que a utilização direta destas variáveis no modelo pode comprometer os resultados, sendo aconselhável aplicar transformações, como normalizações, criação de *bins* ou categorização, para controlar estas distorções.

4.1.3 Relação entre Variáveis: Correlação e Multicolinearidade

A matriz de correlação, apresentada na Figura 8, revela coeficientes baixos entre a variável dependente (L2B) e as variáveis independentes, o que sugere a ausência de relações lineares fortes. As correlações mais significativas concentram-se entre variáveis do mesmo domínio, como *Room* e *Deluxe*, com um coeficiente negativo de -0,35, ou como BB e RO, com um coeficiente negativo de -0,44, padrões estes que indicam alguma redundância de informação entre variáveis associadas à tipologia de quartos.

Figura 8 – Heatmap das Correlações entre Variáveis



Fonte: Autor (2025)

Estas correlações justificam a necessidade de refletir sobre uma possível agregação ou recodificação de variáveis para evitar sobreposição de efeitos explicativos no modelo final.

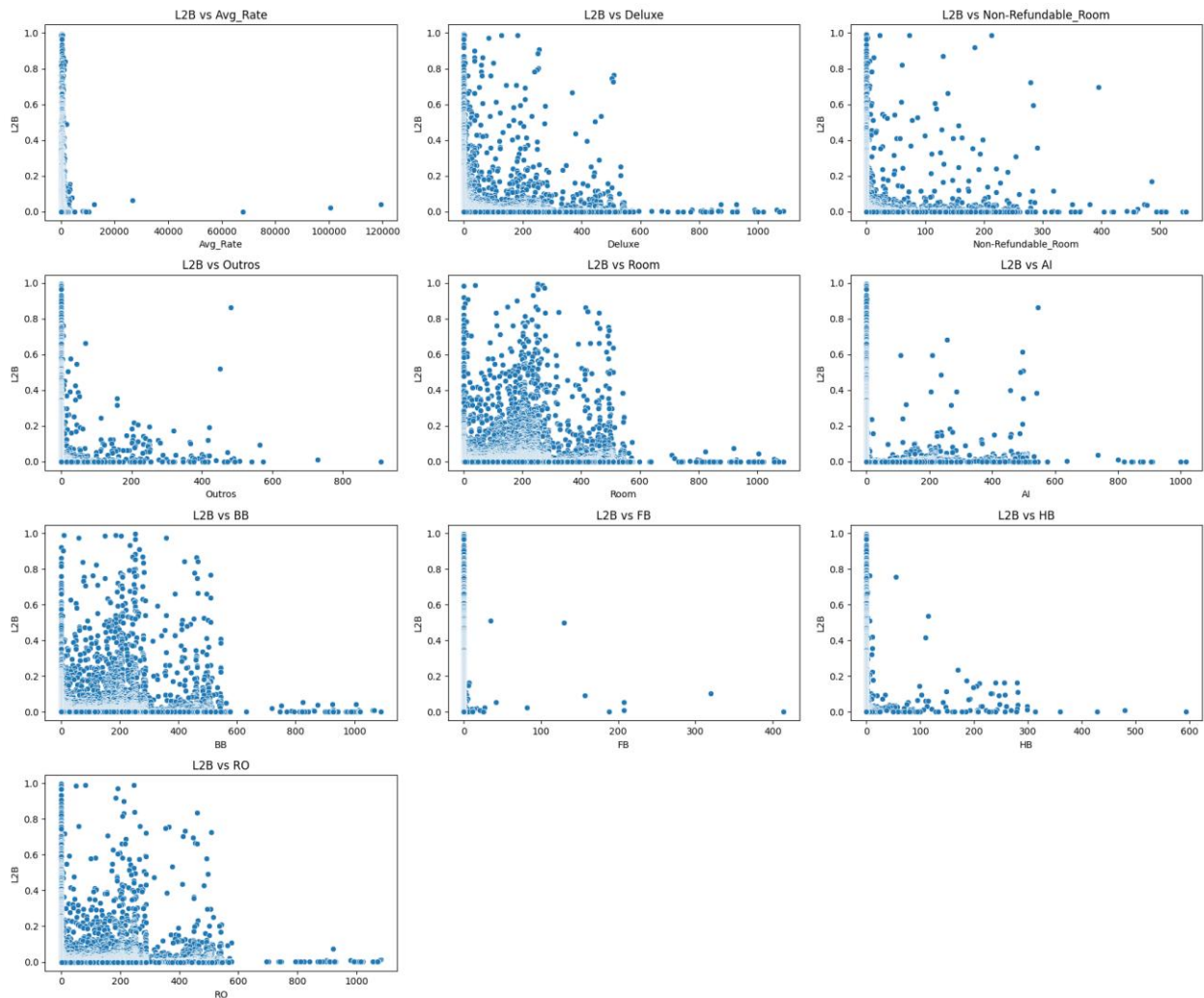
O diagnóstico de multicolinearidade, realizado através do teste *Variance Inflation Factor* (VIF), confirmou esta situação, pois todas as variáveis, exceto *Avg_Rate*, com um VIF aproximadamente igual a 1,00, apresentaram valores de VIF infinitos, que é um indicador de colinearidade quase

perfeita. Este cenário compromete a estabilidade dos coeficientes e reforça a necessidade de reestruturar ou agregar variáveis antes da modelação.

Os diagramas de dispersão ilustram a fraca relação linear entre o L2B e as variáveis explicativas, tal como se pode verificar na Figura 9.

Em grande parte dos casos, observa-se uma elevada concentração de pontos junto ao eixo do L2B onde o valor é igual a 0, o que dificulta a deteção de tendências claras. Este padrão reforça a necessidade de transformação da variável dependente, não só para melhorar a distribuição, mas também para potenciar a sua relação funcional com as variáveis explicativas.

Figura 9 – Scatterplots entre L2B e variáveis independentes



Fonte: Autor (2025)

Por exemplo, mesmo hotéis com centenas de quartos num determinado regime (como BB ou RO) podem apresentar taxas de conversão (L2B) próximas de zero, enquanto outros com oferta muito reduzida obtêm valores de L2B substancialmente superiores. Isto evidencia a complexidade da relação entre oferta e conversão e reforça a utilidade de métodos de transformação que realcem padrões ocultos.

4.1.4 Considerações finais da Fase 1

A primeira fase da Análise Exploratória de Dados permitiu identificar aspetos críticos para a preparação dos dados para a modelação:

- O L2B apresenta assimetria elevada, curtose acentuada e presença de *outliers*, justificando a aplicação de uma transformação;
- O *Avg_Rate* contém outliers extremos que podem distorcer a análise, o que reforça a necessidade de uma transformação logarítmica;
- As variáveis de tipologia de quartos e regimes de refeição apresentam concentrações de zeros e caudas longas, recomendando a aplicação de *binning*;
- A multicolinearidade detetada exige reestruturação das variáveis, de forma a eliminar redundâncias e assegurar a estabilidade da modelação.

Conclui-se, assim, que a base de dados, na sua configuração original, exige um conjunto de transformações que serão detalhadas na fase seguinte da análise exploratória.

4.2 - Transformações Aplicadas aos Dados

Após a análise exploratória inicial, constatou-se que várias variáveis apresentavam assimetria acentuada, concentrações excessivas de valores muito baixos ou nulos, bem como *outliers* de magnitude elevada. Estas características comprometiam a normalidade, a homocedasticidade e a estabilidade dos coeficientes estimados, violando pressupostos fundamentais da regressão linear múltipla. Para resolver estas limitações e preparar o *dataset* para a modelação, foram aplicadas quatro transformações principais: *binning*, criação de variáveis *dummy*, transformação logarítmica do *Avg_Rate* e transformação *logit* do L2B.

4.2.1 - *Binning*

Perante as distribuições assimétricas, caudas longas e elevada concentração de valores nulos observadas em várias variáveis explicativas, procedeu-se ao *binning*, isto é, à criação de escalões, ou intervalos, definidos a partir da análise dos percentis empíricos e da dispersão real de cada variável. A lógica foi selecionar intervalos mais curtos nas zonas onde a densidade de observações é elevada e intervalos mais largos onde surgem poucos valores e mais extremos, assegurando sempre que cada intervalo tenha um número suficiente de observações para suportar inferência estatística e interpretação.

Esta abordagem foi aplicada às variáveis de tipologia de quarto (*Deluxe*, *Non-Refundable_Room*, *Outros_Room*) e de regime de refeição (AI, BB, FB, HB, RO), onde se ajustou os limites aos percentis e à dispersão específica de cada uma.

Por exemplo, num registo onde o hotel teve 18 ocorrências de quartos *Deluxe*, a análise da distribuição revelou Percentil 25% = 0, Percentil 50% = 1, Percentil 75% = 5, Percentil 90% = 20 e máximo = 1086. Com base nestes marcos e de forma a garantir intervalos representativos, segmentou-se a variável nos escalões 0, 1-5, 6-20, 21-100 e >100. Assim, um registo com *Deluxe* = 18 foi colocado no escalão 6-20.

No entanto, em variáveis mais dispersas, adotaram-se intervalos mais alargados, como por exemplo, *Room* foi segmentado em 1-50, 51-150, 151-300, 301-500 e >500, enquanto BB foi segmentado em 1-10, 11-50, 51-200 e >200, para reduzir a influência de *outliers*, captar efeitos não lineares e simplificar a leitura dos efeitos em modelação.

4.2.2 - Criação de *Dummies*

A partir dos escalões definidos na etapa anterior, criaram-se variáveis *dummy*, ou seja, variáveis binárias (0 ou 1) que indicam se um registo pertence ou não a um determinado intervalo, para permitir a inclusão destas categorias num modelo linear sem impor uma ordem artificial entre os escalões.

Esta abordagem permitiu modelar efeitos não lineares e específicos de cada categoria, sem impor uma relação ordinal artificial entre os *bins*. Por exemplo, um hotel que registou 320 ocorrências de quartos *Room*, a variável *Room_binned_301-500* teria um valor igual a 1 e 0 nos restantes intervalos de *Room*, o que preserva a natureza qualitativa das categorias resultantes do *binning*.

Para cada conjunto de *dummies* oriundo da mesma variável, deixou-se uma categoria de fora, que funciona como categoria de referência. Esta prática, amplamente recomendada na literatura, evita a multicolinearidade perfeita pela soma das *dummies* ser igual a 1, problema conhecido como *dummy variable trap*, e garante que cada coeficiente estimado seja interpretado em relação à categoria base, normalmente definida como a mais frequente ou a que representa a situação mais neutra do ponto de vista operacional.

4.2.3 - Transformação logarítmica de *Avg_Rate*

A variável *Avg_Rate* apresentava valores muito assimétricos e *outliers* severos, com um valor máximo superior a 119.000 euros, o que poderia distorcer os coeficientes do modelo.

De forma a estabilizar a variância e reduzir o peso de *outliers*, aplicou-se a transformação logarítmica natural (\log_{1p}), que deu origem à variável *log_Avg_Rate*. Esta transformação comprime a escala, aproxima a distribuição a um formato mais simétrico e melhora a adaptação da variável aos pressupostos da regressão linear múltipla, favorecendo relações mais lineares com a variável dependente.

4.2.4 - Transformação logit de L2B

A variável dependente L2B é um rácio limitado ao intervalo entre 0 e 1 e, na sua forma original, apresentava assimetria acentuada e curtose elevada. Para a tornar mais adequada à modelação linear, aplicou-se a transformação *logit*, amplamente utilizada em variáveis proporcionais, por converter o intervalo restrito (0,1) no domínio real $(-\infty, +\infty)$ (Hosmer et al., 2013; Menard, 2002). Esta conversão reduz a influência de valores extremos, promove maior simetria e assegura uma relação mais próxima da linearidade entre a variável dependente e os preditores. De forma genérica, a transformação *logit* é definida pela Equação 2:

$$\text{logit}(p) = \ln\left(\frac{p}{1-p}\right) \quad (2)$$

No entanto, quando a variável assume valores muito próximos de 0 ou 1, a aplicação direta desta transformação pode conduzir a resultados indefinidos, o que cria instabilidade nos cálculos. De modo a mitigar este problema, a literatura aplicada recomenda, neste contexto, a introdução de um pequeno termo de ajuste ε que impede a ocorrência de valores extremos ou indefinidos (Long, 1997; Wooldridge, 2010). No presente estudo, adotou-se $\varepsilon = 10^{-6}$, em linha com práticas comuns

em regressão logística, o que resultou na forma ajustada da transformação logit, apresentada na Equação 3:

$$L2B_{logit} = \ln\left(\frac{L2B + \varepsilon}{1 - L2B - \varepsilon}\right) \quad (3)$$

Esta versão ajustada assegura que o L2B nunca atinge valores que resultem em resultados indefinidos, o que torna a variável transformada mais estável do ponto de vista estatístico. Em termos práticos, a transformação *logit* ajustada atenua a assimetria, reduz substancialmente a influência de *outliers* e melhora a conformidade com os pressupostos de normalidade dos resíduos e homocedasticidade, sendo, por isso, adequada para a aplicação em regressão linear múltipla.

A aplicação desta técnica originou a variável *L2B_logit*, tendo sido mantida em paralelo a versão original L2B para comparação na fase seguinte.

4.2.5 – Resumo das Transformações na Preparação para Regressão

Em conjunto, estas transformações fortalecem o *dataset* para regressão linear múltipla, na medida em que:

- O *binning* reduz a influência de *outliers*, capta potenciais relações não lineares e melhora a legibilidade substantiva dos efeitos, que melhora a interpretação dos resultados.
- As variáveis *dummies* representam corretamente os intervalos definidos e evitam multicolinearidade ao fixar uma referência, o que permite interpretações diferenciais claras.
- A transformação logarítmica no *Avg_Rate* estabiliza a escala de preços e reduz a dominância dos valores extremos.
- A transformação *logit* no L2B corrige a forte assimetria e torna a variável dependente adequada para modelação linear.

O *dataset* final, assim transformado, apresenta melhores condições para satisfazer os pressupostos da regressão linear múltipla, garantindo maior robustez estatística e validade interpretativa dos resultados. No entanto, de modo a preservar a possibilidade de comparação entre a variável dependente na sua forma original e transformada, criaram-se duas cópias da base de dados resultante das etapas anteriores (*df_transformada*):

- `df_transformada_L2B_original` - mantém a variável L2B na sua forma original.
- `df_transformada_L2B_logit` - inclui a variável *L2B_logit* obtida através da transformação *logit*.

Esta abordagem permite realizar análises paralelas e avaliar, na fase seguinte, qual das versões apresenta melhor desempenho e adaptação aos pressupostos da regressão.

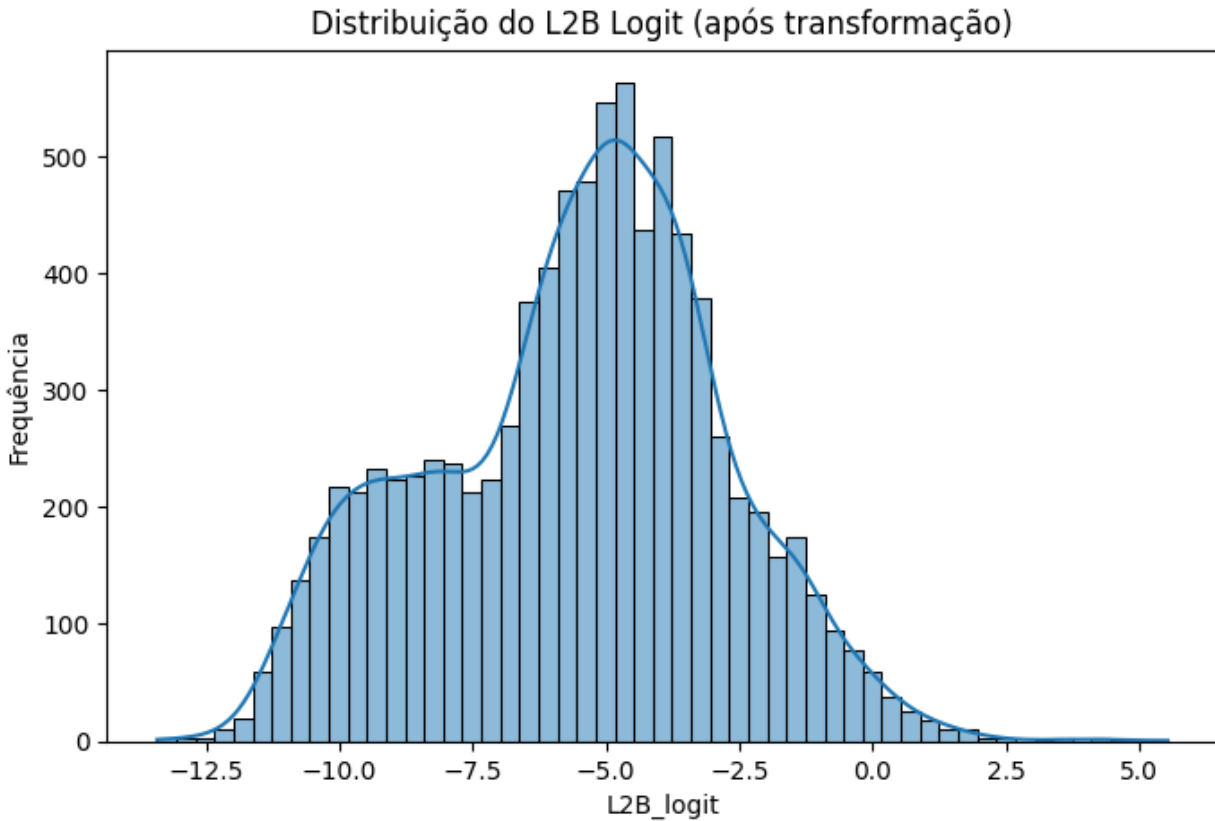
4.3 - Fase 2: Análise Exploratória após Transformações

Após a aplicação das transformações descritas na secção anterior, procedeu-se a uma nova análise exploratória dos dados. O objetivo foi perceber, de forma clara e fundamentada, como estas alterações afetaram a estrutura e a qualidade do *dataset*. Esta etapa funcionou, assim, como um momento de validação, permitindo verificar melhorias na forma das distribuições, redução de *outliers*, controlo da multicolinearidade e reforço da adaptação estatística para a regressão linear múltipla.

4.3.1 – Distribuição das Variáveis Transformadas

A transformação *logit* aplicada ao L2B revelou-se eficaz na correção da sua forte assimetria positiva, observada na fase inicial, tal como evidencia a Figura 4. Na sua forma original, o L2B apresentava valores fortemente concentrados próximos de zero, com *skewness* de 4,533 e *kurtosis* de 24,111, que evidenciava uma distribuição fortemente enviesada e leptocúrtica. Após a transformação *logit*, a distribuição tornou-se mais equilibrada e simétrica, com maior dispersão ao longo da escala. Este resultado confirma que a transformação cumpriu o seu propósito: mapear o intervalo (0, 1) para $(-\infty, +\infty)$, que torna a variável mais compatível com pressupostos estatísticos, como a normalidade, e facilita a modelação linear, tal como comprova a Figura 10.

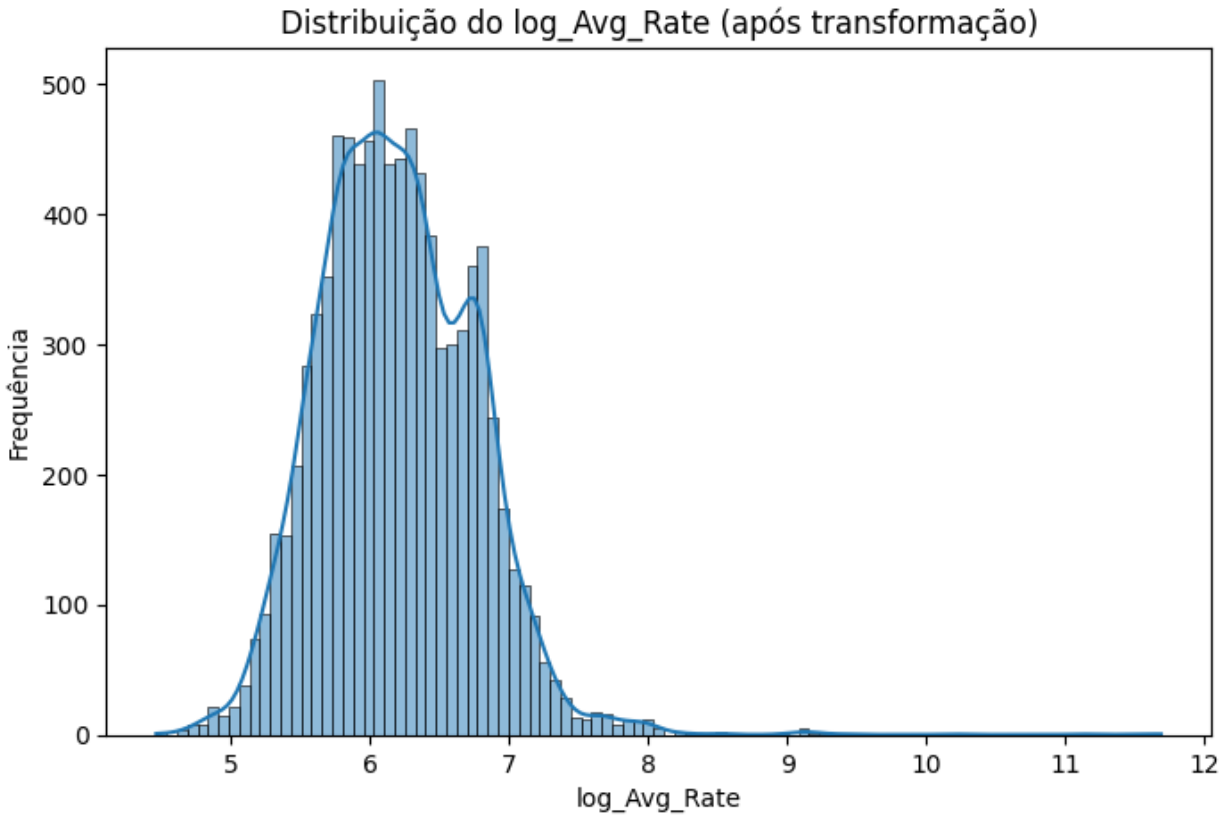
Figura 10 – Distribuição do L2B *Logit*



Fonte: Autor (2025)

No caso da variável *Avg_Rate*, a aplicação da transformação logarítmica, que deu origem à variável, *log_Avg_Rate*, resultou numa distribuição mais compacta e próxima da simetria, tal como se pode observar na Figura 11. Esta modificação atenuou o efeito de valores extremos, cujo máximo ultrapassava 119.000 euros, na fase inicial, e contribuiu para reduzir a influência desproporcional de *outliers* e estabilizar a variância associada a esta variável, o que facilita a interpretação dos coeficientes na fase de modelação.

Figura 11 – Distribuição de \log_Avg_Rate

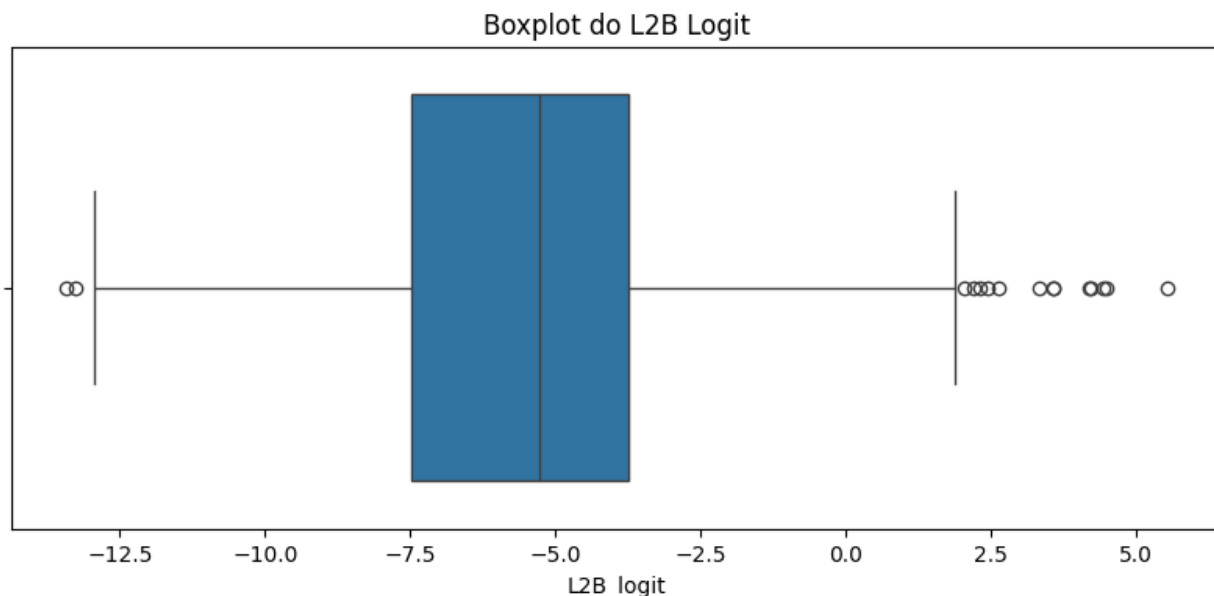


Fonte: Autor (2025)

4.3.2 – *Outliers* após as Transformações

A análise de *outliers* antes e depois das transformações, ilustra de forma clara os benefícios destas alterações. No *boxplot* da variável L2B original, ilustrado na Figura 5, o número de *outliers* identificados foi de 1.272. Após a transformação *logit*, o número de *outliers* reduziu drasticamente para apenas 15 na forma transformada, como se pode comprovar na Figura 12, o que evidencia uma melhoria substancial na estabilidade e robustez estatística da variável dependente.

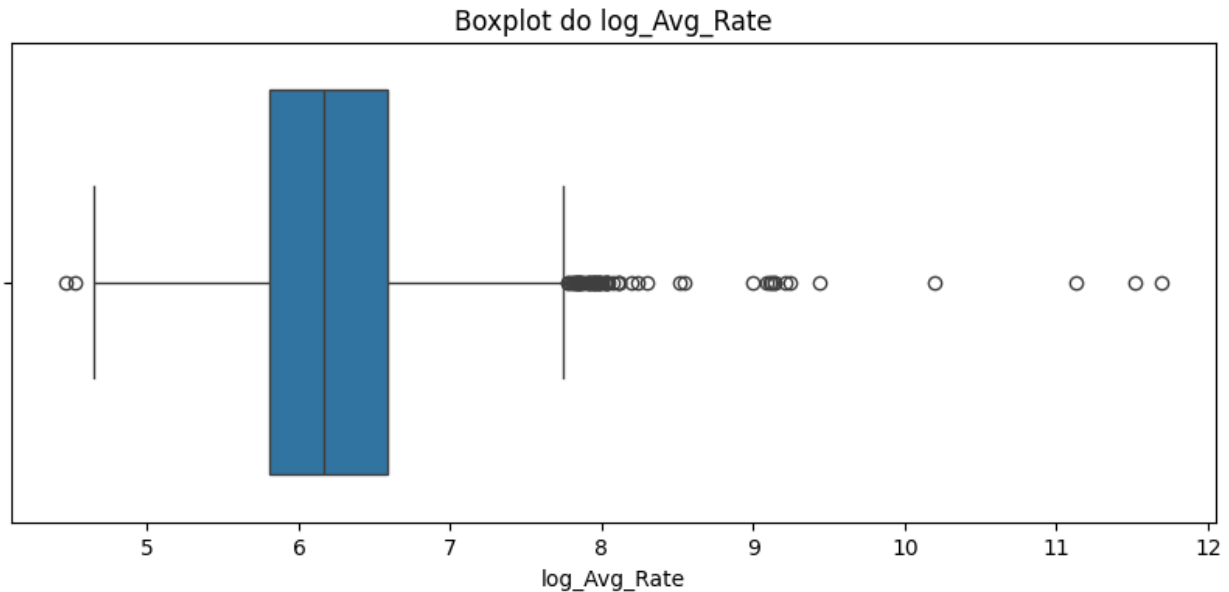
Figura 12 – *Boxplot* de *L2B_logit*



Fonte: Autor (2025)

Da mesma forma, antes da transformação logarítmica à variável *Avg_Rate*, o *boxplot* da variável original, ilustrado na Figura 7, identificava um número total de *outliers* igual a 296. No entanto, após a transformação logarítmica, o número de *outliers* reduziu para apenas 64 na variável transformada *log_Avg_Rate*, como se pode comprovar na Figura 13, o que evidencia uma melhoria substancial na estabilidade e robustez estatística desta variável.

Figura 13 – *Boxplot de log_Avg_Rate*



Fonte: Autor (2025)

Este resultado traduz-se numa diminuição significativa de distorções na estimação dos coeficientes e na melhoria da qualidade do modelo preditivo.

4.3.3 – Multicolinearidade após as Transformações

A avaliação dos fatores de inflação da variância (VIF), após as transformações, revelou uma estrutura estatística estável, com níveis de multicolinearidade considerados baixos ou moderados. No caso da versão com o L2B original, as variáveis com VIF mais elevados foram *Room_binned_151-300*, com um VIF de 4.28, e *BB_binned_>200*, com um VIF de 4.60, entre outras com valores acima de 2. De forma semelhante, na versão com L2B_logit, as variáveis *RO_binned_>200* e *BB_binned_>200* apresentaram os VIF mais elevados, 4.95 e 4.60, respetivamente. No entanto, em ambos os casos, todos os valores de VIF obtidos ficaram abaixo do limiar crítico de 10, o que sugere que não há evidência preocupante de multicolinearidade severa (Kutner et al., 2005).

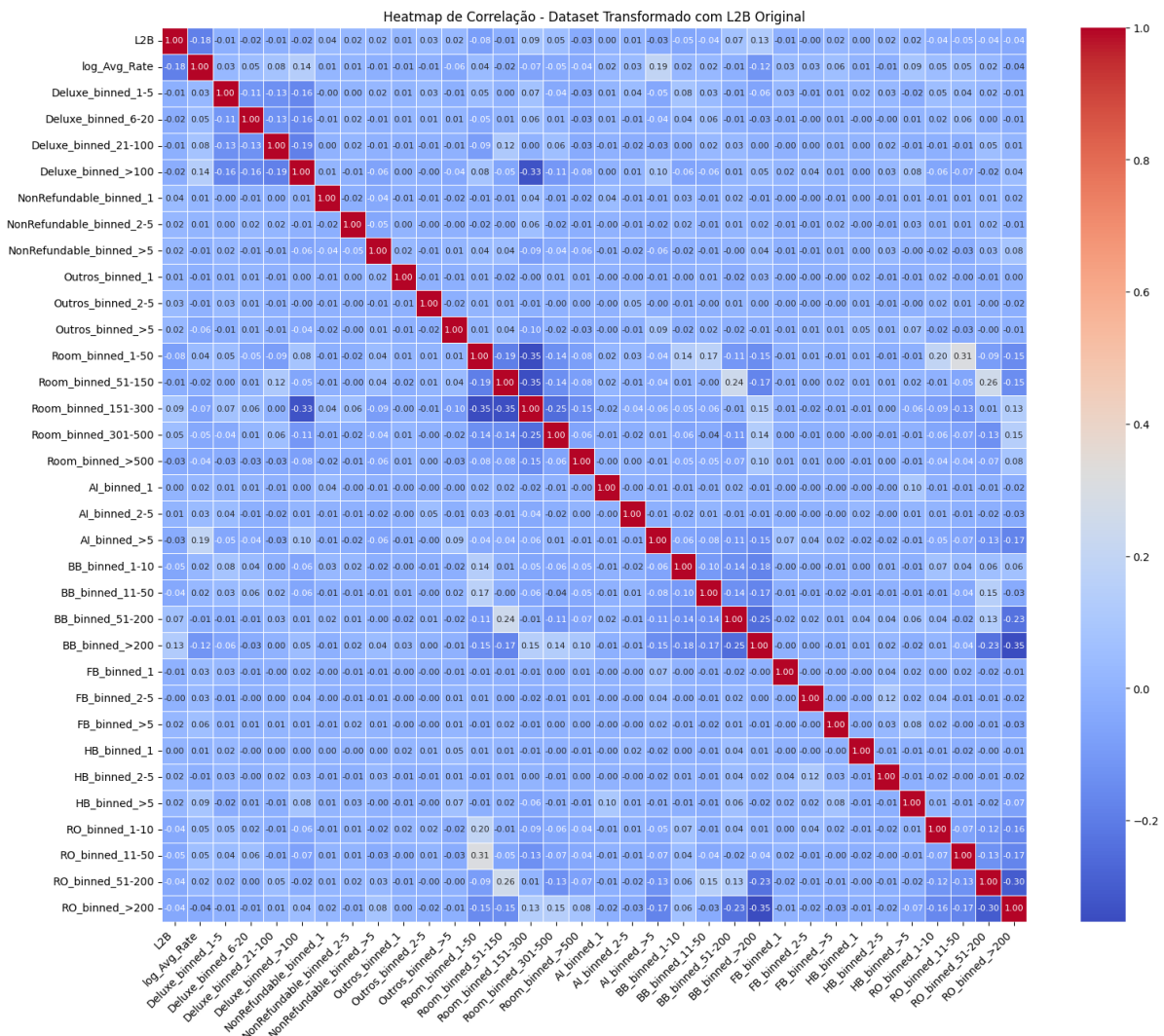
Esta melhoria resulta, em grande parte, da reestruturação das variáveis categóricas em intervalos mais homogéneos, através do *binning*, e da criação cuidada de variáveis *dummy*, respeitando o princípio da exclusão de uma categoria por variável. Estes procedimentos permitiram reduzir redundâncias e garantir uma representação estatisticamente válida das categorias, assegurando

maior precisão na estimação dos coeficientes e maior robustez no modelo a aplicar na fase seguinte.

4.3.4 – Correlações entre Variáveis

A análise das matrizes de correlação reforça a adaptação das transformações aplicadas. Após as transformações aplicadas, foram gerados dois *heatmaps* de correlação, com base na variável dependente original (L2B) e na variável transformada (*L2B_logit*). Em ambos os casos, os padrões de associação revelam coeficientes de correlação reduzidos entre as variáveis independentes, indicando um nível moderado de associação e uma diminuição de relações excessivamente fortes.

Figura 14 – Heatmap de L2B original



Em particular, observou-se que o *L2B_logit* tende a apresentar valores de correlação mais altos com diversos preditores, que sugere uma relação estatística mais forte e, potencialmente, mais linear.

As correlações entre variáveis independentes mantiveram-se, em geral, moderadas, o que complementa os resultados obtidos pela análise de VIF e confirma que a estrutura do dataset não sofre de colinearidade excessiva. Esta combinação de análises permite validar que as transformações não comprometeram a independência entre os preditores e contribuíram para melhorar a relação com a variável dependente.

4.3.5 – Impacto Global das Transformações

Em conjunto, as transformações aplicadas contribuíram para a robustez e adaptação estatística do *dataset*, com impacto direto na preparação para regressão linear múltipla. Os principais benefícios observados foram:

- A transformação *logit* do L2B que promoveu a simetria da distribuição e reduziu drasticamente o número de *outliers*, o que aumenta a adaptação aos pressupostos de normalidade e homocedasticidade.
- A transformação logarítmica do *Avg_Rate* estabilizou a escala dos preços e reduziu o peso dos valores extremos.
- O *binning* de variáveis categóricas, permitiu representar escalões de frequência de forma estruturada e reduzir a influência de outliers.
- A criação de variáveis *dummy* para cada intervalo, permitiu a incorporação de categorias não ordinais sem impor hierarquias artificiais e garantiu, simultaneamente, a ausência de redundância, ao excluir uma categoria por variável.

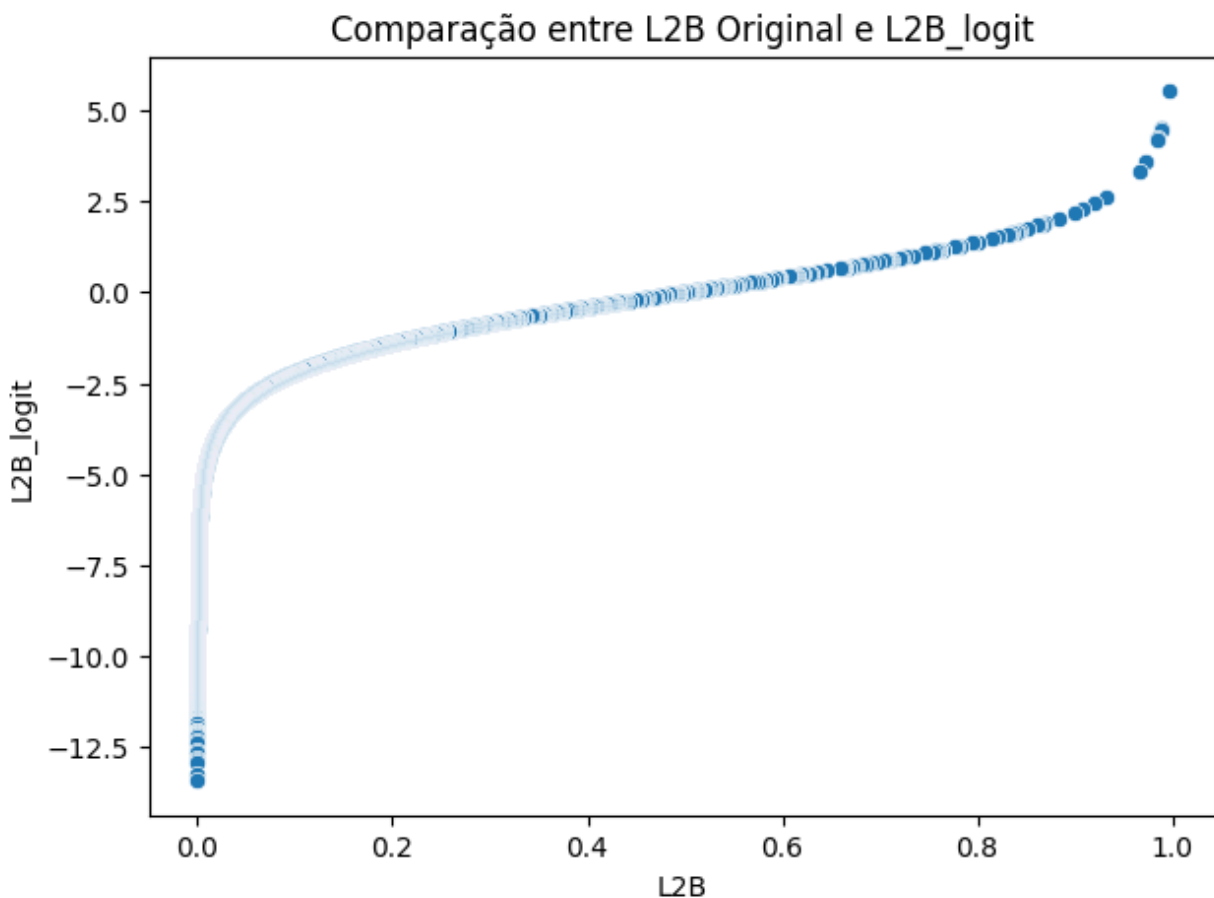
Assim, o *dataset* final, nas versões *df_transformada_L2B_original* e *df_transformada_L2B_logit*, encontra-se significativamente mais robusto e preparado para a fase seguinte, onde será comparada a adaptação das duas versões da variável dependente e selecionada aquela que oferece melhor desempenho para a modelação preditiva.

4.4 – Comparação entre L2B Original e L2B_logit

Após a aplicação da transformação *logit* à variável dependente L2B, procedeu-se à comparação direta entre a versão original e a transformada, com o objetivo de avaliar os ganhos estatísticos e a adaptação para utilização em regressão linear múltipla.

A Figura 16 apresenta um gráfico de dispersão, no qual o eixo das abscissas representa os valores originais de L2B e o eixo das ordenadas representa os valores transformados de *L2B_logit*.

Figura 16 – Comparação entre L2B Original e *L2B_logit*



Fonte: Autor (2025)

Visualmente, o gráfico comparativo demonstra que a transformação *logit* atua de forma significativa sobre a forma e a escala da distribuição, tornando-a mais simétrica e reduzindo de forma expressiva a presença de valores extremos.

Importante salientar que, apesar desta alteração, a relação entre as duas variáveis mantém-se monotónica crescente, o que significa que a ordem relativa dos registos não é alterada, ou seja, sempre que um registo apresenta um valor de L2B superior a outro na escala original, continuará a apresentar um valor superior também na escala *logit*. Deste modo, a transformação preserva a estrutura ordinal dos dados, intervindo apenas na normalização e na adaptação estatística para modelação.

A análise quantitativa desta comparação encontra-se resumida na Tabela 5, onde se apresentam os principais critérios de avaliação. Observa-se que o L2B original apresenta uma assimetria acentuada, onde 1272 *outliers* foram identificados, enquanto o *L2B_logit* reduz esse número para apenas 15 *outliers*. Além disso, a correlação com os preditores aumenta de fraca para moderada, e a adaptação global aos pressupostos da regressão linear múltipla é substancialmente melhorada.

Tabela 5 – Comparação entre L2B Original e *L2B_logit*

Critério	L2B	L2B_logit
Assimetria	Elevada	Reduzida
Outliers	> 1200	~ 15
Correlação com preditores	Fraca	Moderada
Adaptação à regressão	Insuficiente	Adequada

Fonte: Autor (2025)

Com base nestes resultados, a variável *L2B_logit* foi selecionada como variável dependente final para a modelação. A sua utilização assegura maior estabilidade estatística, melhor linearidade em relação aos preditores e uma redução significativa da influência de valores extremos, proporcionando assim maior conformidade com os pressupostos da regressão linear múltipla.

4.5 – Conclusão da Análise Exploratória de Dados

A Análise Exploratória de Dados (EDA) permitiu avaliar, de forma estruturada, a qualidade e adaptação da base de dados para aplicação em regressão linear múltipla.

O processo foi conduzido em duas fases distintas. Numa primeira fase, identificaram-se problemas de assimetria, elevado número de *outliers*, sinais de multicolinearidade e correlações fracas entre

a variável dependente e os preditores. Estes resultados justificaram a aplicação de transformações como *binning*, criação de variáveis *dummy*, transformação logarítmica da *Avg_Rate* e aplicação da função *logit* ao L2B.

A segunda fase confirmou melhorias substanciais, tais como a redução drástica de *outliers*, a maior simetria das distribuições, as correlações mais consistentes e a eliminação de problemas de multicolinearidade. A comparação direta entre o L2B original e o *L2B_logit* evidenciou a superioridade da variável transformada em termos de estabilidade estatística e adaptação aos pressupostos da regressão linear múltipla.

Deste modo, definiu-se como *dataset* final a base de dados *df_transformada_L2B_logit*, que inclui a variável dependente *L2B_logit* e todas as variáveis independentes transformadas e validadas. Esta estrutura garante robustez estatística, validade interpretativa e alinhamento com os objetivos do estudo, ou seja, a investigação pode prosseguir para a construção e validação do modelo estatístico.

V – Modelação e Validação Estatística

O presente capítulo tem como objetivo estimar e validar um modelo estatístico que permita identificar e quantificar os fatores que influenciam significativamente o rácio *Look to Book* (L2B), no contexto empresarial.

O procedimento adotado, segue uma abordagem quantitativa baseada na Regressão Linear Múltipla, uma metodologia amplamente reconhecida pela sua capacidade de analisar relações lineares entre uma variável dependente contínua e múltiplas variáveis independentes, que considera simultaneamente o efeito de diversos fatores explicativos (Hair et al., 2019; Wooldridge, 2016). Ao adotar este tipo de abordagem quantitativa, pretende-se contribuir diretamente para uma gestão estratégica mais informada, ao fornecer bases sólidas para decisões operacionais e comerciais.

O modelo é construído a partir do *dataset* final definido no capítulo anterior (*df_transformada_L2B_logit*), preparado para assegurar robustez estatística, validade interpretativa e alinhamento com os objetivos do estudo. Serão aplicados testes de pressupostos ao modelo para compreender se o mesmo é válido ou não, bem como indicadores de qualidade do ajustamento, de forma a fundamentar a escolha da estrutura mais adequada.

5.1 - Análise do Modelo Completo Inicial

5.1.1 – Introdução e Enquadramento

O primeiro passo da modelação estatística consistiu na estimação do modelo completo, que integrou a totalidade das variáveis explicativas consideradas relevantes na etapa anterior de preparação e validação dos dados. No total, foram incluídas 33 variáveis independentes, abrangendo tanto indicadores quantitativos, como a variável *log_Avg_Rate*, logaritmo do preço médio por noite, como variáveis categóricas transformadas em bins, representativas de diferentes categorias de quarto, como *Deluxe*, *Non-Refundable*, *Outro* e *Room*, e de diferentes regimes alimentares como AI, BB, FB, HB e RO.

A decisão de incluir a totalidade destas variáveis nesta fase inicial teve como propósito avaliar, de forma abrangente, o contributo individual e conjunto de cada fator potencialmente relevante na explicação da variabilidade do *L2B_logit*. Este ponto de partida mais amplo permite, numa fase subsequente, distinguir as variáveis efetivamente significativas daquelas cujo impacto é residual.

5.1.2 – Resultados da Estimação

O modelo foi estimado pelo método dos Mínimos Quadrados Ordinários (OLS). Os principais indicadores globais do modelo foram os seguintes:

- O número de variáveis, um total de 33 variáveis explicativas, que reflete o objetivo inicial de captar toda a informação disponível.
- O coeficiente de determinação, com um valor de R^2 igual a 0,121, o que indica que cerca de 12,1% da variabilidade da variável dependente (*L2B_logit*), é explicada pelo conjunto das variáveis independentes. Embora este valor possa parecer modesto, é consistente com a literatura em estudos aplicados ao comportamento de procura e métricas de conversão no turismo e noutros setores digitais, onde múltiplos fatores não observados, como variáveis qualitativas e comportamentais, influenciam fortemente os resultados (Hair et al., 2019; Wooldridge, 2016).
- Os critérios de informação, como *Akaike Information Criterion* (AIC) e *Bayesian Information Criterion* (BIC), com valores iguais a 42030,4 e 42274,2, respetivamente, onde ambos penalizam a complexidade do modelo, que permite comparações futuras com

versões mais parcimoniosas. O BIC, em particular, penaliza mais fortemente modelos com elevado número de parâmetros.

- A estatística F, com F igual a 36,77, e um *p-value*, onde $p < 0,001$, revelam que o modelo, no seu conjunto, é estatisticamente significativo, o que confirma a validade da sua utilização para fins interpretativos e preditivos.

Estes resultados iniciais mostram que o modelo tem poder explicativo, ainda que limitado, e que a sua complexidade justifica uma análise detalhada às variáveis individuais.

5.1.3 – Análise Preliminar dos Coeficientes

Nesta fase inicial, procedeu-se à análise preliminar dos coeficientes, *p-values* e intervalos de confiança de cada variável explicativa. O objetivo não foi validar definitivamente o modelo, mas identificar padrões de significância estatística e possíveis redundâncias, que constituíssem uma base para decisões subsequentes relativas à simplificação do modelo.

A análise individual permitiu identificar variáveis com efeitos estatisticamente significativos e outras cuja contribuição para o modelo é residual.

- Exemplos de variáveis significativas:
 - *log_Avg_Rate* apresentou um coeficiente negativo (-0,736; $p < 0,001$), sugerindo que, mantendo-se as restantes variáveis constantes, aumentos no preço médio tendem a reduzir o *L2B_logit*, indicando que preços mais elevados se associam a menor probabilidade de conversão.
 - *Deluxe_binned_>100* evidenciou um efeito positivo robusto (+0,959; $p < 0,001$), sugerindo que volumes muito elevados nesta categoria se associam a maior probabilidade de conversão.
 - *NonRefundable_binned_1* mostrou um impacto positivo significativo (+1,040; $p < 0,001$), possivelmente associado a clientes mais decididos a confirmar a reserva.
- Exemplos de variáveis não significativas:
 - *NonRefundable_binned_>5* apresentou $p = 0,483$ e intervalo de confiança que inclui zero, sugerindo ausência de efeito consistente.

- *Outros_binned_1* revelou $p = 0,155$, também com intervalo de confiança abrangendo zero, indicando que a sua contribuição para o modelo é estatisticamente irrelevante.

A tabela integral com os coeficientes estimados, respetivos *p-values* e intervalos de confiança encontra-se disponível no Anexo I, de forma a assegurar rastreabilidade e transparência nas decisões metodológicas.

5.1.4 – Conclusão sobre o Modelo Completo

O modelo completo revelou-se estatisticamente significativo no seu conjunto, o que confirma que as variáveis incluídas contribuem, em parte, para explicar a variabilidade do *L2B_logit*. Contudo, a análise preliminar dos coeficientes evidenciou que certas variáveis apresentam contributos residuais ou não significativos, aspeto que merece uma investigação mais aprofundada.

Desta forma, apesar de o modelo ser útil como ponto de partida para compreender o efeito global das variáveis, a sua complexidade e redundâncias justificam a transição para um modelo reduzido, que permitirá explorar alternativas mais parcimoniosas.

5.2 – Análise do Modelo Reduzido

5.2.1 – Introdução e Justificação da Redução

Após a estimação inicial do modelo completo, verificou-se que, embora o mesmo fosse estatisticamente significativo no seu conjunto, várias variáveis apresentavam contributos residuais ou ausência de significância estatística. A manutenção de todas as 33 variáveis explicativas revelou-se, assim, pouco eficiente do ponto de vista da parcimónia e da utilidade prática para a gestão.

De acordo com a literatura, modelos mais parcimoniosos tendem a ser preferíveis, uma vez que facilitam a interpretação, evitam complexidade excessiva e aumentam a aplicabilidade dos resultados em contextos reais de tomada de decisão (Burnham & Anderson, 2002; Hair et al., 2019). Além disso, a simplificação estrutural permite que os resultados sejam traduzidos em indicadores de interpretação mais intuitiva e acessível, que, por sua vez, se convertem em recomendações práticas e acionáveis para a gestão.

Desta forma, a decisão de reduzir o número de variáveis baseou-se na análise preliminar dos seguintes critérios, coeficientes estimados, *p-values* e intervalos de confiança (IC) a 95%, com o

objetivo de reforçar a parcimónia do modelo. Tal como foi referido na secção 5.1, existiam variáveis cujo intervalo de confiança incluía o valor zero, o que sugere que o efeito estimado da variável pode ser nulo, ou variáveis que apresentavam *p-values* superiores a 0,05, que é um indicador de ausência de efeito estatisticamente consistente.

Tal como referem Field (2018) e Wooldridge (2016), a utilização destes critérios, amplamente recomendados em estudos econométricos e de análise de regressão, constituiu a base para a eliminação subsequente de variáveis, de modo a assegurar que o modelo final retivesse apenas os fatores mais robustos e relevantes.

Deste modo, a redução do modelo não originou apenas uma decisão estatística, mas também metodológica e estratégica, onde se procurou equilibrar a validade científica com a utilidade prática, de forma a criar uma ferramenta analítica simultaneamente rigorosa e aplicável à realidade da gestão no setor turístico.

5.2.2 – Variáveis Eliminadas e Critérios de Exclusão

De acordo com os critérios definidos anteriormente, foram eliminadas oito variáveis do modelo completo, de modo a certificar a parcimónia do modelo. Tal como defendem Hair et al. (2019) e Field (2018), a exclusão destas variáveis permite preservar a validade estatística e concentrar o modelo nos fatores mais relevantes para explicar a variabilidade de *L2B_logit*.

A Tabela 6 ilustra as razões estatísticas da exclusão de cada variável, ao apresentar os seus valores correspondentes.

Tabela 6 – Variáveis Eliminadas do Modelo Completo

Variável	Coefficiente	p - value	IC 95% Inferior	IC 95% Superior	Critério de Exclusão
<i>NonRefundable_binned_>5</i>	- 0,071	0,483	- 0,268	0,127	p > 0,05 e IC inclui 0
<i>Outros_binned_1</i>	0,590	0,155	- 0,223	1,403	p > 0,05 e IC inclui 0

<i>AI_binned_1</i>	- 1,143	0,326	- 3,422	1,137	p > 0,05 e IC inclui 0
<i>AI_binned_2-5</i>	0,666	0,221	- 0,400	1,732	p > 0,05 e IC inclui 0
<i>FB_binned_1</i>	0,319	0,683	- 1,213	1,852	p > 0,05 e IC inclui 0
<i>FB_binned_2-5</i>	1,198	0,112	- 0,278	2,674	p > 0,05 e IC inclui 0
<i>RO_binned_1-10</i>	0,107	0,406	- 0,146	0,360	p > 0,05 e IC inclui 0
<i>RO_binned_>200</i>	0,111	0,408	- 0,152	0,374	p > 0,05 e IC inclui 0

Fonte: Autor (2025)

A eliminação destas variáveis resultou num modelo mais conciso, composto por 25 variáveis explicativas, que preserva os fatores estatisticamente relevantes e reduz o ruído introduzido por variáveis redundantes.

Este processo assegura que a análise subsequente se centra nos determinantes mais sólidos do rácio L2B, em conformidade com as recomendações metodológicas de parcimónia e com a necessidade de produzir resultados que sejam não apenas válidos do ponto de vista estatístico, mas também úteis e aplicáveis na prática da gestão.

5.2.3 – Resultados da Estimação do Modelo Reduzido

A estimação do modelo reduzido foi realizada através do método dos Mínimos Quadrados Ordinários, com um total de 25 variáveis explicativas, após a exclusão das oito variáveis anteriormente identificadas como estatisticamente irrelevantes.

Esta versão reduzida representou uma simplificação face ao modelo completo, resultante da eliminação de oito variáveis com contributos estatisticamente irrelevantes, que visa reforçar a

parcimónia, reduzir redundâncias e preservar apenas os fatores com contributos consistentes para a explicação do *L2B_logit*.

Os principais indicadores globais do modelo reduzido foram os seguintes:

- O coeficiente de determinação, com um valor de R^2 igual a 0,120, confirmou que cerca de 12% da variabilidade do *L2B_logit* é explicada pelo modelo. Mantendo-se assim um valor modesto e compatível com a realidade descrita em estudos empíricos aplicados ao comportamento de procura e métricas de conversão no turismo e noutros setores digitais (Hair et al., 2019; Wooldridge, 2016).
- Os critérios de informação, concretamente o *Akaike Information Criterion* (AIC), com valor igual a 42.020, e o *Bayesian Information Criterion* (BIC), com valor igual a 42.210, apresentaram valores ligeiramente inferiores aos do modelo completo. Este decréscimo sugere uma melhoria na parcimónia e um melhor equilíbrio entre qualidade do ajustamento e simplicidade estrutural.
- A Estatística F, com F igual 48,19, e um *p-value*, onde p é inferior 0,001, confirma que o modelo, no seu conjunto, é estatisticamente significativo. Tal como referido anteriormente, a estatística F é conceptualmente equivalente à ANOVA global, permitindo validar a relevância conjunta das variáveis independentes na explicação do *L2B_logit* (Field, 2018; Gujarati & Porter, 2009).

No que respeita aos coeficientes individuais, vários fatores mantiveram-se estatisticamente significativos e coerentes com os resultados do modelo completo, o que reforça a robustez das conclusões.

- O coeficiente de *log_Avg_Rate*, que apresentou um valor negativo de -0,732 e um *p-value* onde $p < 0,001$), o que confirma que aumentos no preço médio se associam a uma redução do rácio L2B, ou seja, a preços mais elevados corresponde menor probabilidade de conversão.
- A variável *Deluxe_binned_>100* manteve um efeito positivo robusto, com um coeficiente igual a 1,001 e um *p-value* com $p < 0,001$, indicador de que volumes muito elevados nesta categoria de quartos estão associados a maior probabilidade de reserva.

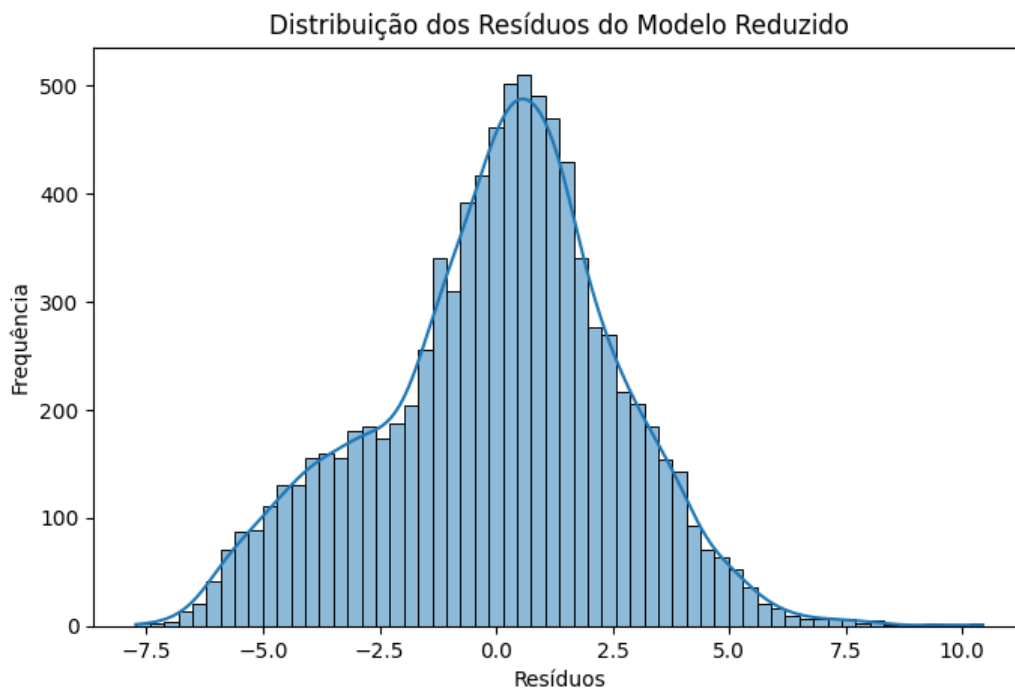
- A categoria *NonRefundable_binned_1* mostrou um impacto positivo significativo, ao apresentar um coeficiente igual a 1,033, em conjunto, com um *p-value* onde $p < 0,001$, que sugere que reservas não reembolsáveis, em pequena escala, tendem a estar associadas a clientes mais decididos e, conseqüentemente, maior probabilidade de conversão.

A preservação destes efeitos estatisticamente significativos demonstra que o processo de redução não comprometeu os principais determinantes identificados, por outro lado, reforçou a clareza e a facilidade de interpretação do modelo. Em síntese, o modelo reduzido constitui uma versão mais parcimoniosa e equilibrada, que conserva a capacidade explicativa necessária e proporciona condições para a obtenção de resultados úteis e aplicáveis em contextos de gestão.

5.2.4 – Testes aos Pressupostos e Interpretação

Após a estimação do modelo reduzido, avaliou-se o cumprimento dos pressupostos da regressão linear múltipla. Esta verificação foi essencial para compreender a robustez do modelo, uma vez que a validade das inferências depende de condições como a normalidade, a homocedasticidade e a independência dos resíduos.

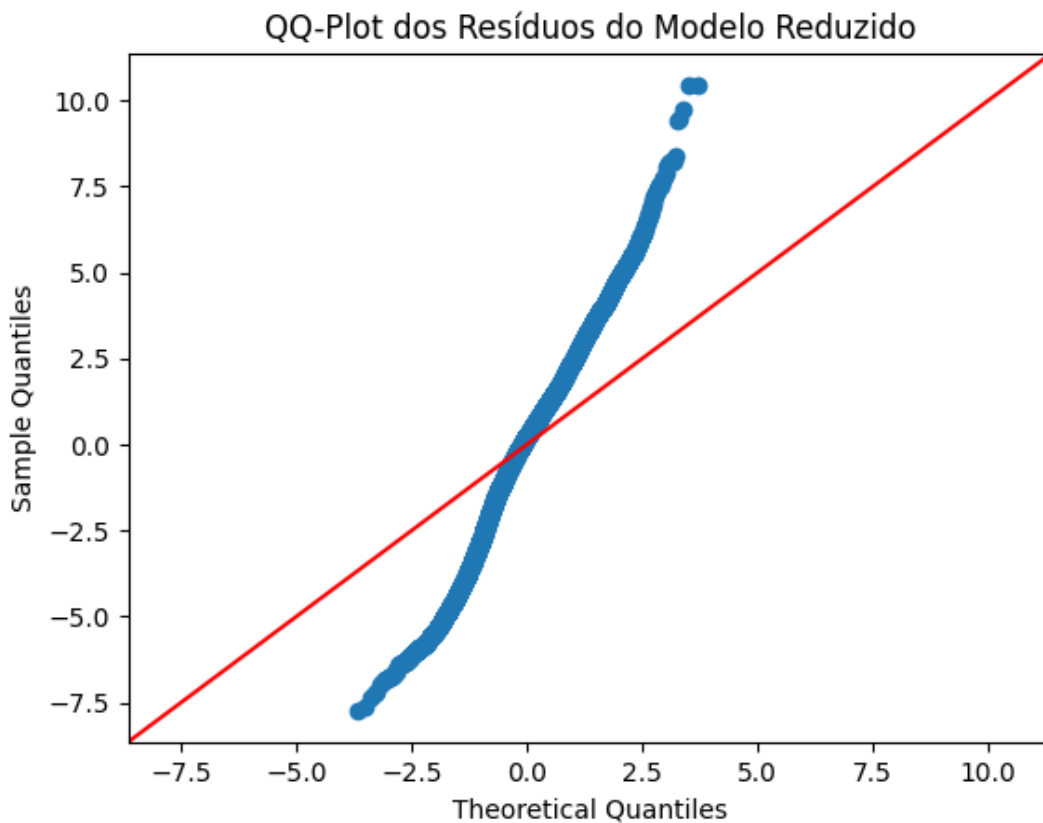
Figura 17 – Histograma dos resíduos do modelo reduzido



Fonte: Autor (2025)

No que diz respeito à normalidade, o teste realizado foi o de *Kolmogorov-Smirnov*, que apresentou uma estatística de 0,0481 e um p-valor inferior a 0,001, o que suportou a rejeição da hipótese nula de distribuição normal, ou seja, que os resíduos não seguem uma distribuição normal, o que revela a violação do pressuposto. Esta conclusão é consistente com os resultados gráficos, como se pode verificar pelo histograma dos resíduos, apresentado na Figura 17, onde se observa uma distribuição assimétrica, e pelo *QQ-Plot*, que se pode verificar na Figura 18, onde os pontos se distanciam da linha de referência, o que confirma a falta de ajustamento à normalidade.

Figura 18 – *QQ-Plot* dos resíduos do modelo reduzido

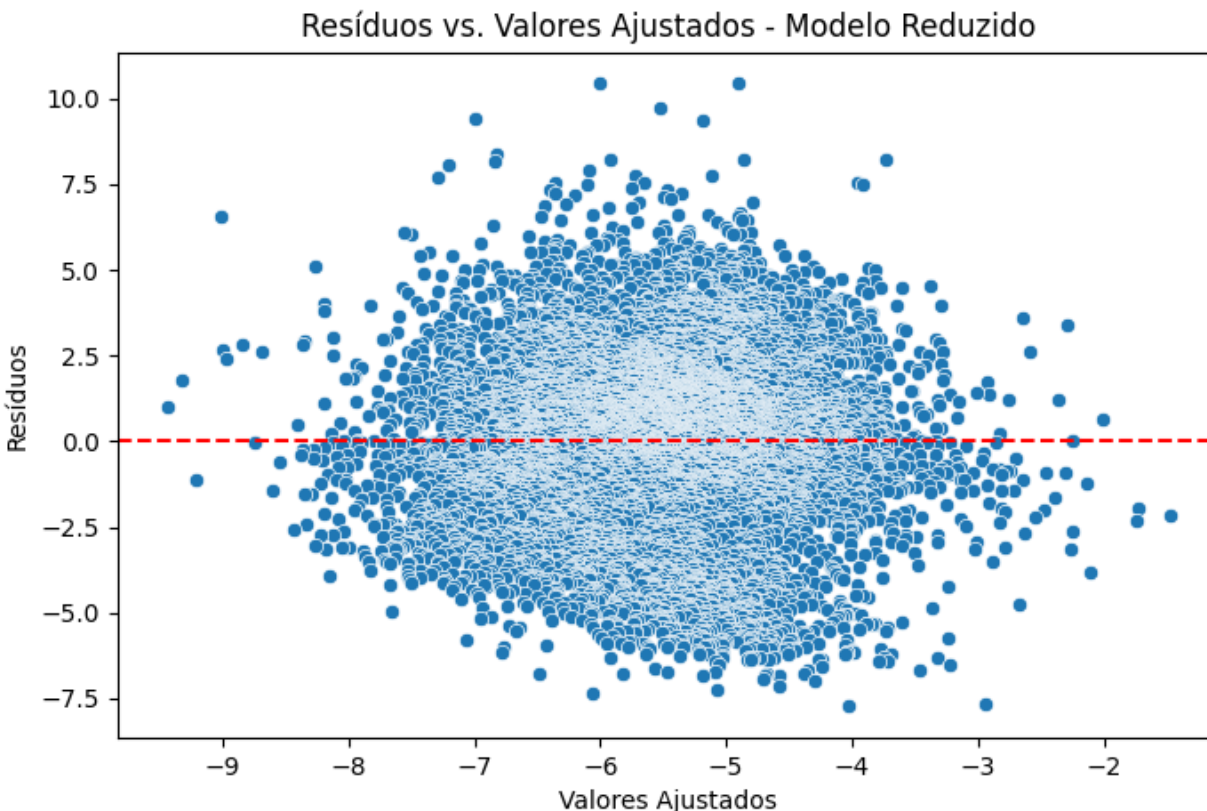


Fonte: Autor (2025)

Quanto à homocedasticidade, ou variância constante dos erros, foi realizada uma análise através do teste de *Breusch-Pagan*. O resultado obtido, onde a estatística teve um valor igual a 695,1050 com *p-value* inferior 0,001, sustentou a rejeição da hipótese nula de variância constante, o que confirma a presença de heterocedasticidade. Esta evidência também se refletiu graficamente no

diagrama de dispersão dos resíduos face aos valores ajustados, apresentado na Figura 19, onde se observam padrões não aleatórios em vez de uma dispersão homogénea.

Figura 19 – Gráfico de resíduos vs valores ajustados do modelo reduzido



Fonte: Autor (2025)

Relativamente à independência dos resíduos, o teste de *Durbin-Watson* apresentou um valor de 0,2325, bastante distante do valor de referência 2. Este resultado traduz uma forte autocorrelação positiva, o que compromete a validade dos erros-padrão quando estimados através do método OLS tradicional.

A análise da multicolinearidade foi realizada na fase da EDA, através do cálculo do *Variance Inflation Factor* (VIF). Os valores encontrados revelaram-se baixos e aceitáveis, todos abaixo do limiar de 5, pelo que este pressuposto não suscitou preocupações adicionais na versão reduzida do modelo.

Apesar das falhas metodológicas identificadas, o modelo preservou fatores estatisticamente robustos e consistentes com o modelo completo. Embora a tabela completa com todos os coeficientes se encontre disponível em anexo, mais concretamente no Anexo II, importa destacar que várias variáveis do modelo completo mantiveram significância estatística elevada no modelo reduzido, que é um indicador positivo quanto à consistência dos resultados. Entre os exemplos mais relevantes, encontram-se:

- A variável *log_Avg_Rate*, com um coeficiente igual a -0,7315 e um *p-value* onde $p < 0,001$, que confirma o efeito negativo do preço médio no rácio L2B.
- A variável *Deluxe_binned_>100*, com um coeficiente de 1,0008 e um *p-value* com $p < 0,001$, que indica maior probabilidade de conversão para hotéis com volumes muito elevados nesta tipologia.
- A variável *NonRefundable_binned_1*, com um coeficiente de 1,0329 e um *p-value* com $p < 0,001$, associado a maior propensão para conversão em tarifas não reembolsáveis.
- A variável *Outros_binned_>5*, com um coeficiente de 0,8435 e um *p-value* onde $p < 0,001$, que reforça o impacto positivo de reservas superiores nesta categoria.

Estes resultados evidenciam que o modelo reduzido conseguiu captar de forma estável os principais determinantes do rácio L2B e ainda preservar a consistência com a realidade observada no mercado. No entanto, as violações dos pressupostos da regressão, como heterocedasticidade, autocorrelação e ausência de normalidade, não permitem que este modelo seja considerado solução final, sendo necessária a aplicação de métodos estatísticos mais robustos que corrijam estas limitações e assegurem validade estatística às conclusões.

5.2.5 – Conclusão sobre o Modelo Reduzido

A estimação do modelo reduzido representou um avanço importante face ao modelo completo, ao proporcionar uma estrutura mais parcimoniosa e uma análise mais centrada nos fatores estatisticamente significativos. A redução para 25 variáveis permitiu eliminar redundâncias e concentrar a análise nos determinantes com maior poder explicativo, alinhando-se assim com as recomendações metodológicas que evidenciam a importância de modelos simplificados para favorecer a sua interpretação e a sua utilidade prática (Burnham & Anderson, 2002; Hair et al., 2019).

Contudo, apesar deste ganho em simplicidade e da preservação de determinantes-chave, como o efeito negativo do preço médio e o impacto positivo de determinadas categorias de quartos e regimes, os testes realizados aos pressupostos evidenciaram limitações críticas. A ausência de normalidade dos resíduos, a presença de heterocedasticidade e a forte autocorrelação positiva comprometem a validade estatística do modelo quando estimado pelo método dos Mínimos Quadrados Ordinários. Tais violações afetam diretamente a fiabilidade dos erros-padrão e, conseqüentemente, a robustez das conclusões deduzidas.

Desta forma, ainda que o modelo reduzido tenha sido benéfico para clarificar os fatores mais consistentes na explicação do rácio L2B, este não reúne condições para ser aceite como modelo final. A transição para uma estimação robusta torna-se, portanto, indispensável, de modo a corrigir estas limitações e garantir que as conclusões criadas não são apenas estatisticamente válidas, mas também relevantes e aplicáveis em contextos de gestão e apoio à decisão.

5.3 – Análise do Modelo Robusto

5.3.1 – Introdução e Justificação Teórica

Após a análise do modelo reduzido, verificou-se que, apesar de ser mais parcimonioso e estatisticamente informativo do que o modelo completo, continuava a apresentar falhas relevantes quanto ao cumprimento dos pressupostos clássicos da regressão linear múltipla, nomeadamente heterocedasticidade, autocorrelação dos resíduos e desvios à normalidade. Estas violações comprometem a validade estatística do modelo quando estimado pelo método tradicional dos Mínimos Quadrados Ordinários.

De modo a ultrapassar as limitações referidas, optou-se pela aplicação de um modelo robusto, com correção de erros-padrão através do método de *Heteroskedasticity and Autocorrelation Consistent* (HAC). Este procedimento, de acordo com Gujarati e Porter (2009) e Wooldridge (2016), ajusta a matriz de variâncias e covariâncias, de forma a neutralizar os efeitos adversos da heterocedasticidade e da autocorrelação, sem alterar os coeficientes estimados do modelo. Assim, preserva-se a estrutura do modelo reduzido, mas garante-se que os erros-padrão, os intervalos de confiança e os testes de significância associados são estatisticamente válidos e fiáveis.

No contexto do modelo robusto, é importante destacar que os testes tradicionais de pressupostos, como *Breusch–Pagan*, *Kolmogorov–Smirnov* ou *Durbin–Watson*, deixam de constituir critério de

exclusão. Embora continuem a sinalizar violações quando aplicados, a correção HAC garante que os resultados permanecem estatisticamente consistentes e válidos, mesmo em presença dessas falhas nos pressupostos, conforme demonstrado por Long e Ervin (2000) e Hayes e Cai (2007).

Do ponto de vista empírico, esta abordagem revela-se particularmente adequada em estudos aplicados a dados reais de natureza complexa, como é o caso do rácio L2B. Em bases de dados desta natureza, com dados reais recolhidos, a diversidade de padrões de procura, as diferenças entre tipologias de produto e a heterogeneidade de comportamento dos clientes, tornam expectável a violação dos pressupostos clássicos. Neste cenário, a utilização de técnicas robustas não é apenas uma escolha metodológica prudente, mas uma necessidade para garantir que os resultados extraídos mantêm validade estatística e utilidade prática.

Desta forma, o modelo robusto consolida-se como a versão final e mais adequada da regressão linear múltipla aplicada ao estudo, ao resolver as limitações do modelo reduzido, onde as violações dos pressupostos comprometiam a confiança nos resultados, e garantir, assim, conclusões estatisticamente fidedignas, consistentes com a realidade operacional do setor e, sobretudo, vantajosas para apoiar a tomada de decisão na gestão.

5.3.2 – Resultados da Estimação do Modelo Robusto

Para realizar a estimação do modelo robusto, recorreu-se à correção HAC, que permitiu obter um modelo estatisticamente consistente, ao preservar a estrutura do modelo reduzido, enquanto corrige as falhas identificadas nos pressupostos clássicos da regressão linear múltipla.

Os indicadores globais permitem caracterizar a sua qualidade e compará-lo diretamente com os modelos anteriores:

- O modelo manteve um total de 25 variáveis independentes, que reflete a redução previamente realizada. Esta parcimónia assegura maior simplicidade interpretativa face ao modelo completo, que incluía 33 variáveis.
- O coeficiente de determinação com um valor de R^2 igual a 0,120, afirma que cerca de 12% da variabilidade do *L2B_logit* é explicada pelo modelo. Este valor mantém-se idêntico ao do modelo reduzido, onde R^2 era igual a 0,120, e praticamente idêntico ao do modelo completo, onde R^2 era igual a 0,121, o que reforça que a exclusão de variáveis redundantes não comprometeu o poder explicativo do modelo.

- Quanto aos critérios de informação, mais propriamente, o *Akaike Information Criterion* (AIC), com valor de 42.020, e o *Bayesian Information Criterion* (BIC), de 42.210, apresentaram valores ligeiramente inferiores aos observados no modelo reduzido, onde AIC teve um valor igual a 42.030 e BIC teve um valor igual 42.210. Esta ligeira descida indica uma melhoria na parcimónia e qualidade do ajustamento, que, por sua vez, se traduz numa melhoria no equilíbrio entre complexidade estrutural e capacidade explicativa (Burnham & Anderson, 2002).
- A Estatística F, onde o valor F foi de 50,29, com *p-value* inferior 0,001, confirma que o modelo é estatisticamente significativo no seu conjunto. Tal como nos modelos anteriores, esta evidência valida a sua utilização para fins interpretativos e preditivos, ao assegurar que as variáveis explicativas, enquanto conjunto, têm impacto significativo na variabilidade do *L2B_logit*.

A análise individual dos coeficientes permite identificar os fatores que exercem impacto estatisticamente significativo sobre o rácio L2B, onde se destacam tanto variáveis já relevantes nos modelos anteriores como novas evidências reforçadas pela robustez do método.

Entre os exemplos mais consistentes e mantidos ao longo das diferentes etapas de estimação, encontram-se:

- O coeficiente de *log_Avg_Rate*, com valor negativo de -0,732 e com *p-value* onde $p < 0,001$, confirma que aumentos no preço médio estão associados a reduções no rácio L2B, o que reforça a evidência de que preços mais elevados diminuem a probabilidade de conversão;
- A variável *Deluxe_binned_>100*, com coeficiente de 1,001 e com *p-value* inferior 0,001, manteve um efeito positivo robusto, que é um indicador de que volumes muito elevados nesta categoria de quartos estão associados a maior propensão de reserva;
- A variável *NonRefundable_binned_1*, cujo coeficiente foi de 1,033 e o *p-value* foi de $p < 0,001$, reforça a relevância das tarifas não reembolsáveis na decisão de reserva, possivelmente associadas a clientes mais decididos a confirmar a reserva;

- A variável *Outros_binned_>5*, com coeficiente de 0,8435 e com *p-value* inferior 0,001, o que indica um impacto positivo significativo quando o número de reservas nesta categoria ultrapassa cinco.

Para além destas variáveis, o modelo robusto evidenciou ainda a relevância de novas categorias, como *Room_binned_151-300*, onde o coeficiente expressivo de 1,954 e o *p-value* menor que 0,001, revela que hotéis com reservas neste intervalo apresentam maior probabilidade de conversão.

De forma consistente com os modelos anteriores, observa-se que os principais determinantes identificados no modelo completo foram preservados, mas agora estimados com maior fiabilidade estatística devido à robustez do método HAC.

No seu conjunto, estes resultados permitem concluir que o modelo robusto não assegura apenas a validade estatística das inferências, ao ajustar erros-padrão, intervalos de confiança e testes de significância de forma robusta, como também preserva os principais determinantes identificados nas fases anteriores. Desta forma, o modelo robusto consolida-se como a versão final e mais adequada da regressão linear múltipla aplicada ao estudo, ao resolver as limitações do modelo reduzido, em que as violações dos pressupostos comprometiam a confiança nos resultados, e gerar conclusões consistentes com a realidade operacional do setor, oferecendo à gestão uma base sólida para a tomada de decisão.

5.3.3 – Interpretação dos Resultados

A estimação do modelo robusto permitiu ultrapassar as limitações detetadas no modelo reduzido, ao garantir que os coeficientes e respetivas inferências estatísticas se mantêm válidos, mesmo perante a violação de pressupostos clássicos. Esta robustez metodológica oferece maior confiança na interpretação dos resultados e permite avançar para uma análise centrada nos determinantes estatisticamente mais consistentes do rácio L2B.

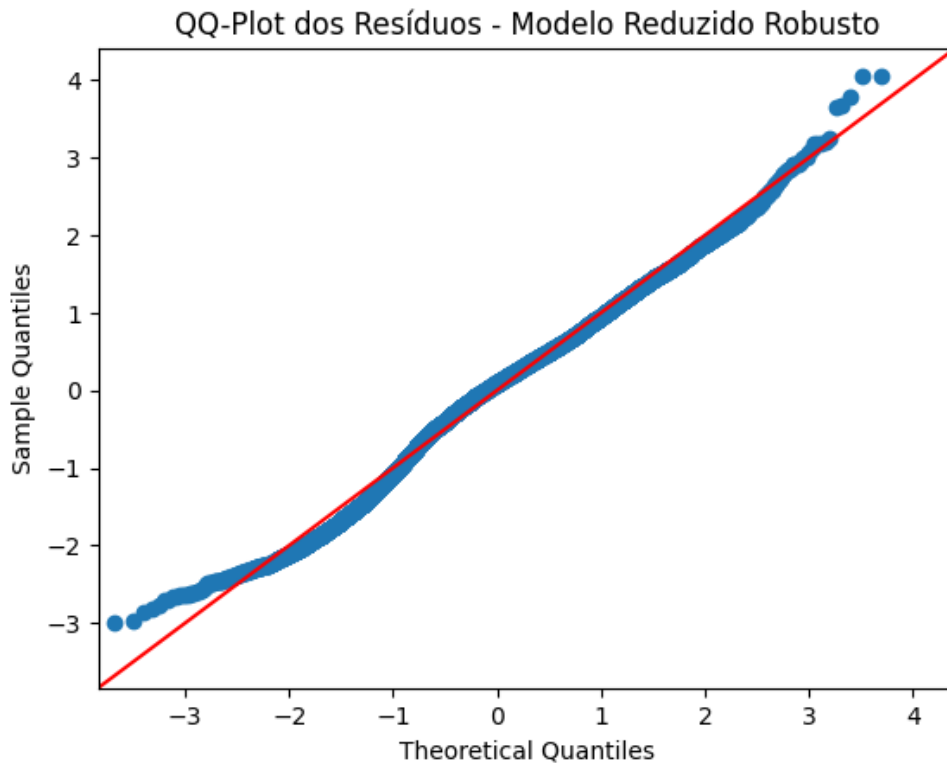
Embora os indicadores globais do modelo já tenham sido apresentados na secção anterior, importa salientar que se mantiveram inalterados face ao modelo reduzido. O coeficiente de determinação ($R^2 = 0,120$), os critérios de informação (AIC = 42.020; BIC = 42.210) e a estatística F (50,29; $p < 0,001$) confirmam que o modelo é estatisticamente significativo no seu conjunto. Assim, a

verdadeira mais-valia do modelo robusto reside na fiabilidade acrescida das estimativas e na interpretação substantiva dos seus coeficientes.

Tal como mencionado anteriormente, os testes tradicionais de pressupostos deixam de constituir critério de exclusão, aquando utilizada a correção HAC, embora continuem a sinalizar violações quando aplicados.

No entanto, a melhoria proporcionada pela aplicação da robustez pode ser ilustrada através da análise dos resíduos. No *QQ-Plot* do modelo reduzido, apresentado anteriormente na Figura 18, observava-se um afastamento pronunciado dos pontos face à linha de referência, o que evidenciava sérias falhas na aproximação à normalidade. Por outro lado, no *QQ-Plot* do modelo robusto, representado na Figura 20, os pontos alinham-se de forma bastante mais próxima da linha, tornando-se num indicador de melhoria clara no ajustamento. Embora não se atinja uma normalidade perfeita, esta evolução demonstra como o modelo robusto permite conclusões estatísticas mais fiáveis e consistentes, mesmo em contextos de dados complexos.

Figura 20 – *QQ-Plot* dos resíduos do modelo robusto



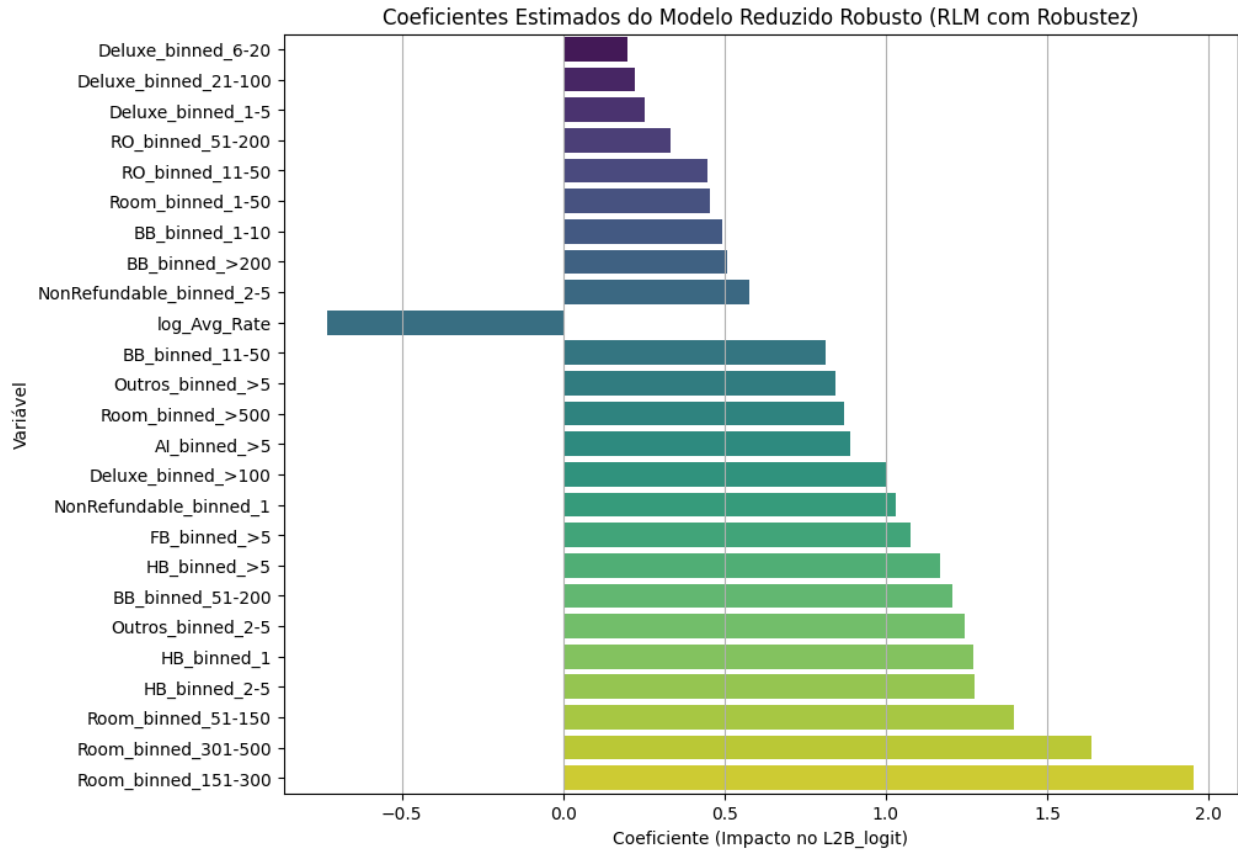
Fonte: Autor (2025)

Quanto à interpretação substantiva, a análise individual dos coeficientes evidencia os fatores que mais influenciam o rácio L2B. Entre os mais relevantes destacam-se:

- A variável *log_Avg_Rate*, que apresentou um coeficiente negativo de -0,732 e um *p-value* inferior a 0,001, o que confirma que aumentos no preço médio se associam a uma diminuição significativa do L2B. Ou seja, preços mais elevados reduzem a probabilidade de conversão, resultado consistente com a literatura em *revenue management* e comportamento do consumidor.
- O coeficiente de *Room_binned_151-300* revelou-se positivo e expressivo, com um valor igual a 1,954 e um *p-value* onde $p < 0,001$, que indica que hotéis com reservas neste intervalo apresentam maior probabilidade de conversão, o que reflete um efeito de escala associado a volumes de procura intermédios.
- A variável *HB_binned_2-5* apresentou um coeficiente positivo de 1,246 e um *p-value* menor que 0,001, sugestão de que o regime de meia pensão, quando associado a um número moderado de reservas, potencia a decisão de compra.
- A variável *BB_binned_51-200*, também com coeficiente positivo de 1,107 e um *p-value*, onde $p < 0,001$, reforça a importância do regime de pequeno-almoço incluído, que continua a constituir uma das ofertas mais valorizadas pelos clientes.
- O coeficiente de *NonRefundable_binned_1*, demonstrou ser um valor positivo igual a 1,033, em conjunto com um *p-value* inferior a 0,001, o que confirma o impacto das tarifas não reembolsáveis, associadas a maior propensão de conversão por parte de clientes decididos e menos sensíveis a cancelamentos.

No entanto, a Figura 21, apresenta graficamente todos os coeficientes estimados e ilustra a direção e magnitude dos efeitos de cada coeficiente, ao evidenciar o contributo diferenciado das variáveis explicativas para o L2B.

Figura 21 – Coeficientes Estimados do Modelo Robusto



Fonte: Autor (2025)

A relevância destas variáveis vai para além da significância estatística, ao traduzir-se em implicações práticas para diferentes áreas da gestão. O preço médio, os regimes alimentares, as tipologias de quarto e as políticas tarifárias constituem dimensões que não só explicam o comportamento do L2B, mas também orientam decisões estratégicas. Na gestão estratégica, os resultados permitem reforçar o posicionamento competitivo através da definição de preços que maximizem a conversão sem comprometer a rentabilidade. Na gestão de operações, orientam a alocação de recursos para tipologias de quartos e regimes com maior impacto na probabilidade de reserva. Na gestão de marketing, permitem segmentar clientes de acordo com a sua sensibilidade ao preço ou a determinados regimes, ao ajustar campanhas promocionais. Por fim, na perspetiva de *Business Intelligence*, estes fatores podem ser integrados em *dashboards* de monitorização contínua, ao transformar resultados estatísticos em instrumentos de gestão operacionais e acessíveis.

Em síntese, o modelo robusto não se limita a corrigir falhas do modelo reduzido, isto porque ao oferecer resultados estatisticamente válidos e consistentes, o modelo garante uma base mais fiável para a tomada de decisão. Enquanto o modelo reduzido já apontava fatores relevantes, a ausência de validade estatística poderia comprometer a confiança dos gestores nas conclusões obtidas. O modelo robusto ultrapassa esta limitação, pois fornece evidência científica sólida e aplicável à prática. Desta forma, consolida-se como uma ferramenta de apoio estratégico, que permite que as decisões sobre preços, regimes e tipologias sejam baseadas em dados mais fiáveis, com maior potencial de orientar o desempenho organizacional da gestão no setor do turismo e em contextos de mercado semelhantes.

5.3.4 – Conclusão final sobre o modelo robusto como modelo final

A estimação do modelo robusto, com correção HAC, consolidou-se como a solução metodológica mais adequada para este estudo. Ao preservar a estrutura parcimoniosa do modelo reduzido e, simultaneamente, corrigir as limitações associadas às violações dos pressupostos da regressão linear múltipla, o modelo robusto assegura validade estatística às estimativas e consistência às inferências obtidas.

Em contraste com o modelo completo, excessivamente complexo, e com o modelo reduzido, que embora parcimonioso não cumpria com os pressupostos fundamentais, o modelo robusto representa um equilíbrio entre simplicidade estrutural e rigor metodológico. Esta combinação garante maior confiança científica e maior utilidade prática, ao produzir resultados que se mantêm fiáveis e replicáveis em contextos de análise e de gestão.

Deste modo, o modelo robusto é apresentado como o modelo final desta investigação, pois responde de forma clara e fundamentada à questão de investigação, “Identificar os fatores que influenciam significativamente o rácio *Look to Book* (L2B)”. A robustez assegurada pela correção HAC, reforça que as conclusões extraídas refletem padrões reais de comportamento, sem distorções provocadas por falhas estatísticas.

Do ponto de vista da gestão, esta decisão traduz-se em benefícios concretos. O modelo final fornece uma base mais sólida para apoiar a tomada de decisão em áreas como a gestão estratégica, a gestão de operações, o marketing e *Business Intelligence*. Ao oferecer resultados estatisticamente consistentes e alinhados com a realidade operacional, garante que as organizações dispõem de

evidência quantitativa fiável para definir políticas de preços, segmentar clientes, ajustar regimes alimentares ou otimizar a composição de portfólios hoteleiros.

Em síntese, o modelo robusto não constitui apenas uma alternativa técnica face às limitações do modelo reduzido, mas uma opção metodológica que conjuga validade científica e relevância prática. A sua adoção assegura que as recomendações originadas por este estudo assentam em fundamentos estatísticos sólidos, que transformam a análise em conhecimento útil e aplicável, capaz de orientar a tomada de decisão e criar valor na gestão do setor turístico e em contextos de mercado semelhantes.

VI – Discussão de Resultados

A discussão dos resultados constitui uma etapa essencial em qualquer investigação académica, visto que é neste momento que os resultados estatísticos adquirem significado, ao serem interpretados à luz da teoria e das implicações para a gestão. A relevância desta fase é amplamente reconhecida na literatura, já que permite transformar análises estatísticas em conhecimento aplicável, capaz de gerar contributos tanto para a ciência como para a ação empresarial (Zikmund et al., 2018; Hair et al., 2019). Neste capítulo, os fatores identificados como determinantes do rácio L2B são analisados de forma crítica, com o objetivo de compreender o seu significado, de os relacionar com a evidência teórica existente e de avaliar o seu impacto em contextos de gestão.

6.1 – Interpretação dos Coeficientes Significativos

A análise dos coeficientes significativos do modelo robusto evidencia, em primeiro lugar, a influência determinante do preço médio na probabilidade de conversão. O efeito negativo associado à variável *log_Avg_Rate*, confirma que os clientes mantêm uma elevada sensibilidade ao preço, característica típica de mercados onde a comparação de ofertas é imediata e onde a elasticidade da procura desempenha um papel central. Este resultado sugere que aumentos no preço médio comprometem diretamente o rácio L2B, mesmo quando a diferenciação do produto está presente. Do ponto de vista da gestão, esta evidência reforça a necessidade de políticas tarifárias dinâmicas e cuidadosamente calibradas, capazes de equilibrar a maximização da receita com a preservação da taxa de conversão.

No que respeita às tipologias de quarto, destacam-se efeitos positivos em intervalos específicos de reservas, com maior expressão em escalões intermédios como o *Room_binned_151-300*. Este

resultado pode ser interpretado como um reflexo de um efeito de escala, em que unidades hoteleiras que atingem volumes moderados de reservas aparentam dispor de maior atratividade, possivelmente pela combinação de reputação consolidada, maior visibilidade em plataformas digitais e condições competitivas ajustadas à procura. Num mercado cada vez mais digitalizado, atributos como a percepção de confiança, a reputação e a facilidade de escolha tornam-se fatores críticos no comportamento do consumidor, assumindo um papel decisivo na conversão.

Por outro lado, a influência positiva de categorias mais exclusivas, como o *Deluxe*, em determinados intervalos, sugere que existe também espaço para uma procura orientada para valor acrescentado, em que a diferenciação pela qualidade do quarto constitui um fator de decisão. Assim, observa-se que a conversão não é apenas uma função do preço, mas também da percepção de qualidade e da adequação da oferta a diferentes segmentos de mercado, refletindo estratégias de posicionamento diferenciadas.

Os regimes alimentares revelam igualmente um contributo expressivo para explicar a probabilidade de conversão. A presença de pequeno-almoço incluído em intervalos específicos, confirma a relevância de benefícios tangíveis que aumentam o valor percebido pelo cliente, o que reduz a necessidade de procurar serviços adicionais. Da mesma forma, a meia pensão surge como um fator de diferenciação valorizado, sobretudo em volumes moderados de reservas, ao conjugar conveniência e previsibilidade de custos. Estes resultados refletem a importância do design do produto turístico, em que a combinação entre alojamento e serviços complementares constitui uma proposta de valor mais atrativa para diferentes perfis de consumidores. A gestão deve, assim, reconhecer que os regimes alimentares funcionam não apenas como um detalhe operacional, mas como uma dimensão estratégica da oferta, com impacto direto no L2B.

Por fim, as políticas tarifárias associadas a opções não reembolsáveis demonstram impacto positivo na taxa de conversão, em particular quando o número de reservas enquadradas neste regime é reduzido. Este resultado pode ser explicado pela existência de um perfil de cliente mais decidido e menos avesso ao risco, que valoriza a garantia de disponibilidade a um preço competitivo, mesmo em detrimento da flexibilidade de cancelamento. Do ponto de vista da empresa, este comportamento traduz-se numa oportunidade de assegurar receitas firmes e reduzir incertezas operacionais. No entanto, é igualmente um alerta para a necessidade de equilibrar este

tipo de tarifas com alternativas mais flexíveis, de modo a não excluir consumidores que preferem preservar opções de alteração ou cancelamento.

6.2 – Comparação de Resultados com a Literatura Existente

O efeito do preço médio sobre o rácio L2B revelou-se negativo, o que confirma que os consumidores demonstram elevada sensibilidade às variações tarifárias. Este resultado está em consonância com a literatura que reconhece o preço como um dos fatores centrais na explicação da conversão, reforçando a relevância do *revenue management*, enquanto prática de gestão que procura maximizar a receita através do ajustamento dinâmico de preços (Cross et al., 2009). A evidência obtida encontra suporte em Mohapatra et al. (2023), que demonstraram a importância da dispersão de preços em plataformas digitais, e reforça a noção de que, em mercados altamente competitivos, mesmo diferenças marginais de preço podem traduzir-se em variações significativas no comportamento de reserva.

Por outro lado, a influência positiva das tarifas não reembolsáveis representa um resultado que contrasta parcialmente com os resultados da literatura, uma vez que a maioria dos estudos distingue a flexibilidade como elemento decisivo no processo de reserva (El-Said, 2020). A evidência obtida neste estudo sugere que, em determinados segmentos, a rigidez da política tarifária pode estar associada a uma maior probabilidade de conversão, possivelmente devido à presença de clientes com decisão mais firme e baixa aversão ao risco. Este resultado pode ser compreendido à luz de investigações sobre comportamento em *e-commerce*, que apontam para a tendência de clientes mais determinados concluírem o processo de compra de forma mais célere e consistente (Moe & Fader, 2004). Desta forma, este estudo acrescenta uma perspetiva complementar à literatura, ao demonstrar que as tarifas não reembolsáveis, habitualmente vistas como potenciais barreiras, podem, em certos contextos, funcionar como estímulo à conversão.

Para além do preço, também os atributos da oferta hoteleira se revelaram determinantes na explicação do L2B. Os resultados associados às tipologias de quarto e aos regimes alimentares, confirmam a importância de características tangíveis na decisão de reserva, o que vai de encontro com estudos que destacam o peso dos atributos específicos da oferta na escolha dos consumidores. Masiero e Nicolau (2015) demonstraram que a seleção de hotéis está fortemente condicionada por características específicas da oferta, incluindo a tipologia de quarto e os serviços adicionais associados. O impacto positivo de tipologias mais exclusivas, como o *Deluxe*, bem como de

categorias associadas a volumes médios de reservas, reforça a evidência de que a percepção de qualidade e a adequação da oferta a diferentes segmentos, exercem influência decisiva na conversão.

De igual modo, os efeitos observados em regimes alimentares, como o pequeno-almoço incluído (BB) ou a meia pensão (HB), encontram suporte em investigações que realçam o valor acrescentado percebido pelos clientes quando os pacotes incluem serviços complementares. Estudos como os de Zhao et al. (2015) e El-Said (2020), destacam que a integração de atributos adicionais aumenta a confiança e reduz a incerteza associada à reserva, o que favorece a decisão de compra. Assim, a evidência empírica obtida neste estudo, confirma a relevância de atributos funcionais como diferenciadores no processo de decisão, ao ampliar a compreensão de como fatores tangíveis reforçam a atratividade da oferta.

Apesar disso, importa reconhecer que fatores intangíveis como a confiança e a reputação, frequentemente salientados pela literatura como determinantes da conversão em contextos digitais (Cezar & Ögüt, 2015; Zhao et al., 2015), não puderam ser avaliados diretamente no presente estudo, uma vez que as variáveis disponíveis não captavam dimensões como rankings de pesquisa, avaliações de clientes ou imagem da marca. Buhalis e Law (2008) já haviam sublinhado o papel transformador da informação online no turismo, antecipando a centralidade da reputação digital no processo de decisão do consumidor, ao passo que Gretzel et al. (2015) desenvolveram o conceito de *smart tourism*, que enfatiza a confiança como elemento estrutural na adoção de soluções digitais.

Este estudo insere-se na linha de investigação sobre a utilização dos dados e a importância do *Big Data* (Mariani, 2020), procurando compreender como a análise de grandes volumes de dados pode contribuir para a compreensão de padrões de comportamento e melhoria da experiência do utilizador.

Em síntese, a comparação com a literatura demonstra que os resultados obtidos confirmam grande parte das evidências produzidas em estudos prévios, em especial no que respeita à sensibilidade ao preço e à valorização de características específicas da oferta hoteleira, mas também introduzem contributos originais que enriquecem o debate académico, nomeadamente o impacto positivo das tarifas não reembolsáveis. Desta forma, este estudo contribui para consolidar e expandir o

conhecimento sobre os determinantes da conversão em contexto turístico, ao oferecer uma perspectiva empírica que complementa de forma crítica os avanços registados na literatura.

6.3 – Implicações para a Gestão

A análise dos resultados obtidos permite retirar um conjunto de implicações práticas para diferentes áreas da gestão, ao transformar os coeficientes estatisticamente significativos do modelo robusto em orientações aplicáveis ao contexto organizacional. Estas implicações refletem não apenas a relevância do preço, das tipologias de quarto e dos regimes alimentares como fatores determinantes do rácio L2B, mas também a forma como tais evidências podem ser operacionalizadas em diferentes domínios da gestão, desde a definição de preços até ao suporte estratégico baseado em *Business Intelligence*.

No domínio do *revenue management*, os resultados confirmam que o preço médio exerce impacto negativo sobre a conversão, evidência que reforça a necessidade de políticas de *pricing* dinâmicas. A sensibilidade do consumidor ao preço sugere que aumentos excessivos comprometem a taxa de reserva, o que impõe aos gestores o desafio de encontrar um ponto de equilíbrio entre rentabilidade e atratividade. A influência positiva das tarifas não reembolsáveis, demonstra ainda que determinados segmentos de clientes valorizam a decisão firme e menos flexível, o que sugere a possibilidade de segmentar a oferta tarifária em função do perfil de risco dos consumidores. Deste modo, o *revenue management* deve ser entendido como uma prática de afinação constante, capaz de ajustar os preços em tempo real e de maximizar a probabilidade de conversão sem comprometer a margem de lucro.

Ao nível do marketing e do produto, a evidência empírica mostra que regimes como o pequeno-almoço incluído (BB) e a meia pensão (HB) aumentam significativamente a propensão de reserva. Estes resultados indicam que serviços complementares funcionam como sinais de valor acrescentado que reduzem a incerteza e aumentam a confiança do consumidor, constituindo-se como elementos de diferenciação. As campanhas de comunicação devem, portanto, enfatizar a conveniência e a experiência proporcionada por estes regimes, reforçando a sua visibilidade em *websites*, motores de busca e plataformas digitais. Do ponto de vista da segmentação, torna-se possível identificar perfis de clientes mais propensos a adquirir pacotes integrados, ajustando a proposta de valor às expectativas de conforto, segurança e conveniência.

Na gestão operacional, os resultados relativos às tipologias de quarto revelam que tanto categorias exclusivas, como o *Deluxe*, quanto intervalos associados a volumes médios de reservas, apresentam impacto positivo na conversão. Este padrão sugere que a atratividade não depende apenas de preços baixos, mas também da adequação da oferta a diferentes segmentos de mercado. Para os gestores, tal implica a necessidade de planejar cuidadosamente o inventário, assegurando a disponibilidade de tipologias com maior impacto na conversão. Além disso, a evidência demonstra a importância de equilibrar oferta premium, dirigida a segmentos de maior poder aquisitivo, com categorias intermédias que garantem escala e consistência de reservas. A gestão operacional deve, assim, alinhar a alocação de recursos com os segmentos mais relevantes para assegurar eficiência e maximizar a taxa de conversão.

No campo da gestão comercial, os resultados fornecem instrumentos para melhorar tanto a relação com fornecedores, como a gestão das próprias equipas comerciais. Em termos de fornecedores, torna-se essencial negociar condições que assegurem a disponibilidade de regimes valorizados, como BB e HB, bem como a inclusão de tarifas não reembolsáveis em contratos estratégicos. Paralelamente, a análise do L2B permite definir objetivos mais precisos para as equipas comerciais, ao avaliar o desempenho por tipo de cliente ou parceiro, como agências de viagens (B2B), consumidores finais (B2C) ou *Closed User Groups* (CUG), e ao incentivar uma abordagem baseada em evidência. A informação produzida pode ainda apoiar programas de formação, ao capacitar as equipas a argumentarem junto de clientes e parceiros, com base em dados que demonstram claramente quais os atributos mais determinantes na conversão.

O papel do *Business Intelligence* assume particular relevância, ao traduzir os resultados estatísticos em instrumentos operacionais para a gestão. A integração do L2B e das variáveis explicativas em *dashboards*, permite acompanhar, em tempo real, o desempenho por segmento, tipologia de quarto, regime alimentar ou tarifa. Estes indicadores podem ser estruturados como KPIs que orientam a monitorização contínua da conversão, possibilitando a deteção rápida de anomalias ou quebras inesperadas de desempenho. A utilização de ferramentas de BI, garante que a informação não se limita a relatórios estáticos, mas que se transforma em conhecimento acionável, acessível e atualizado, capaz de sustentar decisões em diferentes áreas da organização.

Desta forma, o *Business Intelligence* consolida-se como a ponte entre a análise estatística e a prática de gestão, ao permitir que os resultados sejam incorporados em processos de acompanhamento diário e em revisões periódicas de desempenho.

Finalmente, a gestão estratégica e a tomada de decisão de longo prazo beneficiam diretamente da integração destes resultados. A compreensão dos fatores que influenciam o L2B possibilita a definição de prioridades organizacionais, apoiando decisões sobre tipo de produto, alocação de recursos e posicionamento competitivo. A informação proveniente dos sistemas de BI, constitui um suporte essencial para a formulação estratégica, que fornece aos decisores evidência sólida para antecipar tendências, projetar cenários de procura e alinhar objetivos de médio e longo prazo. Para além do contributo tático, o L2B transforma-se assim num instrumento estratégico, capaz de orientar a organização na adaptação a mercados voláteis e altamente competitivos, o que promove uma cultura de decisão assente em dados e não apenas em intuição.

Em síntese, as implicações identificadas demonstram que os resultados do modelo robusto não se esgotam na sua interpretação estatística, mas oferecem contributos concretos para a gestão em múltiplas áreas funcionais. Desde o *pricing* até ao planeamento estratégico, passando pela comunicação de produto, gestão de inventário, negociação comercial e desenvolvimento de *dashboards*, o L2B revela-se uma métrica transversal com elevado potencial de aplicação prática. A consolidação destas implicações reforça a relevância do estudo para a gestão contemporânea, ao transformar evidência empírica em orientações que sustentam decisões mais eficazes, integradas e fundamentadas.

6.4 – Síntese dos Resultados

O presente estudo procurou responder à questão central “Quais os fatores que influenciam o rácio Look to Book (L2B)?”, com base na análise realizada, foi possível identificar um conjunto de variáveis que exercem impacto significativo sobre esta métrica de conversão, que permite compreender de que forma elementos da oferta se refletem no comportamento do consumidor.

Os resultados demonstraram que o L2B é explicado principalmente pelo preço médio, pelas tipologias de quarto, pelos regimes alimentares e pelas políticas tarifárias. O preço médio revelou um efeito negativo, o que confirma que aumentos nas tarifas reduzem a probabilidade de reserva. As tipologias de quarto, evidenciaram um impacto positivo em escalões intermédios e em categorias premium, sugerindo que tanto a perceção de qualidade, como a adequação da oferta a

diferentes segmentos, contribuem para a decisão do consumidor. Também os regimes alimentares, em particular o pequeno-almoço (BB) e a meia pensão (HB), apresentaram influência positiva, funcionando como sinais de conveniência e valor acrescentado. Finalmente, as tarifas não reembolsáveis mostraram-se associadas a maior propensão para a conversão, o que reflete um perfil de clientes mais decididos e menos sensíveis ao risco de cancelamento.

Importa salientar que, enquanto métrica de conversão, o L2B reflete diretamente os fatores considerados neste estudo, ainda que possa igualmente ser influenciado por variáveis não incluídas na base de dados, como localização, sazonalidade ou reputação digital. Esta limitação metodológica reforça a necessidade de futuras investigações explorarem dimensões complementares que permitam captar com maior abrangência a complexidade do processo de decisão.

No entanto, uma das evidências mais relevantes do presente trabalho é que os coeficientes obtidos apresentam magnitudes relativamente próximas, o que indica que o L2B não depende de um único determinante isolado, mas resulta antes de uma combinação de fatores interligados. Este carácter multifatorial, confere maior robustez às conclusões, o que demonstra que a conversão não pode ser explicada apenas pelo preço ou por um atributo específico, mas pela interação de várias dimensões da oferta.

Conclui-se, assim, que o rácio *Look to Book*, é influenciado por um conjunto de fatores relacionados com preço, tipologias de quarto, regimes alimentares e políticas tarifárias, cuja análise integrada permite compreender de forma mais profunda o comportamento do consumidor. Esta dissertação contribui para consolidar o L2B como uma métrica de gestão relevante, ao oferecer uma base empírica que sustenta a tomada de decisão em ambientes turísticos cada vez mais competitivos e orientados por dados.

VII – Conclusão

7.1 – Conclusão do Estudo

O presente estudo teve como objetivo central identificar os fatores que influenciam o rácio *Look to Book* (L2B), uma métrica que mede a eficiência da conversão de pesquisas em reservas, sendo reconhecido como um indicador pertinente para avaliar o desempenho em contextos turísticos. De forma a responder à questão central da investigação, que consistiu em identificar os fatores que

influenciam o L2B, o estudo percorreu um caminho estruturado em três grandes etapas. Numa primeira fase, realizou-se a análise exploratória de dados, que permitiu avaliar a qualidade e a adequação estatística do conjunto de variáveis disponíveis, assegurando assim que o *dataset* final reunia condições para sustentar a modelação. De seguida foi feita a estimação de três modelos de regressão linear múltipla, o modelo completo, o modelo reduzido e o modelo robusto, onde este último foi escolhido como modelo final devido à sua maior validade inferencial. Por fim, a discussão dos resultados permitiu interpretar os coeficientes significativos, relacioná-los com a literatura existente e refletir sobre as suas implicações para a gestão. Este percurso metodológico foi sustentado, de forma transversal, por uma preparação rigorosa dos dados, que contribuiu para a fiabilidade das conclusões.

Os resultados obtidos evidenciam que o L2B é influenciado por diferentes fatores que atuam de forma complementar. O preço médio revelou-se determinante, ao exercer um efeito negativo sobre a taxa de conversão, o que confirma que aumentos das tarifas reduzem a propensão para a reserva. As tipologias de quarto demonstraram impacto positivo em categorias específicas, o que sugere que tanto os segmentos intermédios como as opções *premium*, podem reforçar a atratividade junto dos consumidores. Os regimes alimentares, nomeadamente o pequeno-almoço (BB) e a meia pensão (HB), destacaram-se igualmente pela sua relevância, ao funcionarem como atributos de conveniência e de valor acrescentado. Por último, as tarifas não reembolsáveis apresentaram influência positiva, o que revela que, em determinados contextos, a rigidez contratual pode estar associada a maior probabilidade de conversão. Importa sublinhar que os coeficientes estimados, apresentaram magnitudes relativamente próximas, o que indica que o L2B não resulta de um único determinante isolado, mas antes da interação de múltiplos fatores, o que reforça o carácter multifatorial desta métrica.

De forma global, este estudo confirma a utilidade do L2B enquanto indicador empírico de conversão, ao demonstrar que pode ser explicado e monitorizado com base em atributos concretos da oferta. Do ponto de vista académico, o trabalho contribui para o reforço da literatura sobre métricas de desempenho, em particular no turismo e na gestão, ao fornecer evidência quantitativa sobre os determinantes do L2B. No plano prático da gestão, os resultados consolidam o L2B como uma métrica estratégica, capaz de auxiliar e sustentar, de forma mais assertiva, o processo de tomada de decisões orientadas por dados em áreas como *revenue management*, marketing e

operações. Assim, ao responder à questão de investigação, esta dissertação acrescenta conhecimento sobre os fatores que influenciam o L2B e contribui para a consolidação desta métrica como instrumento analítico e de gestão. Além disso, abre caminho para futuras investigações que explorem dimensões complementares não consideradas no presente estudo.

7.2 – Contribuições para a Gestão e para a Academia

Para além dos resultados empíricos obtidos, importa destacar os contributos que este estudo oferece, tanto ao avanço do conhecimento científico como à prática da gestão. A análise do rácio L2B ultrapassa a dimensão estatística, ao trazer consigo implicações relevantes para a literatura existente e para a aplicação concreta em contextos organizacionais.

Na vertente académica, este estudo contribui para o aprofundamento da literatura sobre métricas de conversão, ao demonstrar que o L2B pode ser explicado por um conjunto de fatores observáveis e quantificáveis, como o preço médio, as tipologias de quarto, os regimes alimentares e as políticas tarifárias. A utilização de técnicas de exploração estatística e de regressão linear múltipla robusta (Hair et al., 2019; Wooldridge, 2020), acrescenta valor metodológico ao evidenciar como técnicas quantitativas podem ser aplicadas a dados do setor turístico e de gestão e utilizadas para analisar fenómenos de comportamento do consumidor. Para além disso, a identificação de um resultado que diverge do padrão habitualmente identificado na literatura, como o impacto positivo das tarifas não reembolsáveis, oferece uma perspetiva inovadora, que enriquece o debate académico ao abrir espaço para novas interpretações e linhas de investigação (Moe & Fader, 2004; El-Said, 2020). Desta forma, este estudo não só confirma tendências teóricas relevantes, como também amplia o conhecimento existente através da introdução de resultados que renovam o debate científico.

Do ponto de vista da gestão, os resultados obtidos traduzem-se em aplicações práticas de relevo. Em primeiro lugar, na área de marketing, a análise do L2B permite identificar discrepâncias entre interesse e conversão, como situações em que existem muitas pesquisas, mas poucas reservas, sobretudo quando segmentado por canal, tipologia de quarto ou regime alimentar. Estas discrepâncias sinalizam se o problema está no preço, na proposta de valor ou na adequação da comunicação, fornecendo pistas objetivas para que as equipas ajustem estratégias de promoção, redefinam mensagens e adaptem campanhas a segmentos específicos. Neste sentido, o L2B pode apoiar decisões sobre a alocação de orçamentos publicitários e sobre o tipo de conteúdos mais eficazes em cada canal, em conformidade com a perspetiva de Kotler e Keller (2021), segundo a

qual a eficácia de uma campanha depende não apenas da atratividade da oferta, mas também da sua pertinência para o público-alvo.

Na área das operações e *revenue management*, o L2B revela-se um instrumento essencial de diagnóstico e monitorização contínua. Um L2B persistentemente baixo, face ao histórico ou a produtos comparáveis, sinaliza desalinhamento entre preço e procura ou propostas pouco competitivas. Na prática, este sinal permite adotar medidas corretivas rápidas, como a reavaliação de níveis tarifários, o ajuste de políticas de flexibilidade e cancelamento, a redistribuição de inventário para tipologias e regimes com melhor desempenho, bem como a renegociação de condições com fornecedores sempre que necessário. Estas práticas alinham-se com os princípios fundamentais do *revenue management*, que enfatizam a importância da gestão dinâmica da oferta e da utilização de indicadores de desempenho, para maximizar a rentabilidade (Cross et al., 2009).

Entre estas dimensões operacionais e a camada mais tecnológica do *Business Intelligence*, destaca-se ainda o papel do L2B no processo de tomada de decisão. Em contextos marcados por elevada incerteza e forte pressão competitiva, a existência de indicadores claros e quantificáveis é essencial para reduzir o risco e aumentar a assertividade das escolhas estratégicas. O L2B permite aos gestores priorizar segmentos de clientes, regimes alimentares ou tipologias de quarto com maior impacto na conversão, tornando mais eficiente a alocação de recursos. Neste sentido, a métrica não é apenas um reflexo do desempenho comercial, mas um instrumento ativo de apoio a decisões críticas, que se alinha com a crescente valorização da gestão orientada por dados (Kotler & Keller, 2021).

Na vertente do *Business Intelligence* e da gestão estratégica, o contributo torna-se ainda mais evidente. A integração do L2B em dashboards e sistemas de monitorização contínua permite transformar resultados estatísticos em ferramentas operacionais de apoio à decisão. A segmentação por canais, como agências de viagens (B2B), consumidores finais (B2C) ou grupos fechados (CUG), combinada com a monitorização de KPIs específicos, possibilita um controlo mais rigoroso da eficácia comercial e a deteção precoce de anomalias, como quebras abruptas de procura ou taxas de conversão invulgarmente baixas (Zikmund et al., 2018). Para além disso, o L2B pode apoiar a formulação de estratégias de médio e longo prazo, ao ser integrado em modelos de previsão baseados em séries temporais ou regressões, que permitem simular cenários futuros e antecipar tendências de procura (Hair et al., 2019). Este estudo demonstra, assim, que o L2B não

deve ser visto apenas como indicador retrospectivo, mas como um instrumento estratégico de antecipação, planejamento e capaz de sustentar uma gestão orientada por dados e mais adaptativa face a mercados voláteis.

Para além das aplicações imediatas, a investigação sistemática dos fatores que influenciam o L2B pode servir de base à construção de modelos de *benchmarking* internos. A identificação de atributos com maior eficiência de conversão, como determinados regimes alimentares ou categorias de quarto, fornece aos gestores a possibilidade de replicar essas práticas eficazes em diferentes mercados ou produtos, o que potencia ganhos de eficácia global. De forma complementar, a incorporação do L2B nos processos de gestão contribui para uma cultura organizacional orientada por dados, na qual a intuição é constantemente validada pela evidência empírica, fortalecendo a capacidade de resposta das empresas em ambientes competitivos (Kotler & Keller, 2021).

Em síntese, este estudo oferece uma dupla contribuição. A nível académico, contribui para a consolidação do L2B como objeto de análise científica, reforça a literatura sobre métricas de conversão e metodologias quantitativas aplicadas à gestão, enquanto introduz evidência inovadora. Quanto à área de gestão, esta investigação demonstra que o L2B pode ser utilizado como um KPI estratégico, capaz de sustentar decisões mais informadas e de contribuir para melhorias de desempenho em áreas como marketing, operações, *revenue management* e *Business Intelligence*. Embora o trabalho não esgote as possibilidades de investigação, os contributos aqui apresentados consolidam o papel do L2B como uma métrica estratégica, com impacto simultâneo na teoria e na prática da gestão.

7.3 – Limitações do Estudo

Tal como em qualquer investigação académica, este estudo apresenta um conjunto de limitações que importa reconhecer, de modo a enquadrar devidamente o alcance das conclusões obtidas. Estas limitações não invalidam a relevância dos resultados, mas estabelecem fronteiras claras para a sua interpretação e abrem espaço a futuros desenvolvimentos.

A primeira limitação está relacionada com o horizonte temporal restrito, uma vez que a análise se baseou apenas em dados referentes a um único mês, dezembro de 2024. Embora este recorte temporal tenha permitido uma avaliação detalhada e controlada, limita a possibilidade de captar sazonalidades ou padrões que se manifestam em períodos mais longos.

Em segundo lugar, os resultados não podem ser generalizados para outros contextos, visto que foram obtidos a partir de um conjunto específico de dados e de um recorte temporal delimitado.

Uma vez que a análise se centrou nos atributos presentes na base de dados disponível, outra limitação está relacionada com a ausência de variáveis contextuais, como a localização ou as datas da estadia, que permitiriam analisar padrões de sazonalidade e de duração das reservas, bem como aspetos relacionados com a navegação na plataforma e fatores ligados à confiança e reputação online, como avaliações ou rankings digitais. Adicionalmente, o impacto positivo das tarifas não reembolsáveis deve ser interpretado com prudência, já que diverge daquilo que a literatura maioritariamente sugere. Este resultado poderá também refletir a ausência de fenómenos captados por variáveis não incluídas na base de dados, como o perfil de cliente ou o canal de reserva.

Em suma, estas limitações não invalidam a solidez do estudo, mas sublinham a necessidade de prudência na extrapolação dos resultados. Além disso, evidenciam que o L2B, apesar de se afirmar como uma métrica útil e robusta, deve ser interpretado em função do contexto e das características dos dados disponíveis.

7.4 – Recomendações para Estudos Futuros

Através dos resultados obtidos nesta investigação, surgem novas oportunidades de estudo, cuja exploração poderá não só aprofundar e expandir o conhecimento académico sobre métricas de conversão e o *Look to Book* (L2B), como também reforçar a utilidade prática desta métrica enquanto ferramenta de apoio à gestão.

Uma primeira linha de investigação passa pelo alargamento do horizonte temporal. A inclusão de dados de vários meses ou anos permitiria captar sazonalidades e variações entre diferentes anos ou épocas, o que reforçaria a robustez dos resultados e a sua utilidade para decisões de planeamento.

Em segundo lugar, a análise poderá ser enriquecida com a inclusão de variáveis contextuais adicionais, como destino, país de origem do cliente ou datas da estadia, e com a integração de fatores digitais associados à confiança e reputação, como avaliações ou rankings online, de forma a tornar a análise ao L2B mais aprofundada e concreta. Neste âmbito, o caso das tarifas não reembolsáveis merece estudos adicionais, de modo a compreender melhor os mecanismos por detrás do resultado que diverge do esperado. Tal análise exigiria informação adicional, como perfis de clientes ou canais de reserva.

Outra via de aprofundamento consiste na utilização de modelos preditivos mais sofisticados, como técnicas de *machine learning* ou modelos de séries temporais avançados, que poderiam reforçar a capacidade de previsão do L2B e captar relações não lineares entre variáveis. Para além do ganho técnico, estas metodologias oferecem às organizações instrumentos de previsão que fornecem apoio aos gestores no processo de tomada de decisão estratégica, ao antecipar cenários de procura e reduzir incertezas em contextos competitivos.

Também se destaca a possibilidade de desenvolver estudos de *benchmarking* e replicação. O L2B poderá ser utilizado como KPI comparativo em diferentes produtos, segmentos ou canais, e a replicação do modelo em empresas ou mercados distintos permitiria avaliar até que ponto os determinantes identificados se mantêm consistentes, reforçando a validade externa das conclusões. Por exemplo, seria pertinente comparar mercados nacionais e internacionais, ou analisar diferenças entre hotéis independentes e cadeias hoteleiras, de modo a perceber se os fatores que influenciam a conversão se mantêm constantes em contextos variados.

Por fim, investigações futuras poderiam analisar combinações de fatores, em vez de avaliar cada variável de forma isolada. Através de metodologias como testes de hipóteses entre grupos, seria possível mapear interações entre variáveis disponíveis nas bases de dados, como por exemplo, destino, preço, datas da estadia, tipologia de quarto e regime alimentar. Do ponto de vista académico, tal abordagem permitiria explorar interações complexas entre determinantes do L2B, contribuindo para modelos explicativos mais robustos. Do ponto de vista da gestão, possibilitaria identificar configurações de atributos mais eficientes, fornecendo recomendações práticas de elevado valor para decisões comerciais e estratégicas, que aumentassem a probabilidade de conversão.

Em síntese, estas recomendações de investigação oferecem oportunidades para superar as limitações do presente estudo e ampliar a compreensão do L2B enquanto métrica de conversão. O seu aprofundamento permitirá não apenas reforçar a literatura existente, mas em particular disponibilizar ferramentas mais sólidas e aplicáveis à prática da gestão.

VII-Referências Bibliográficas

Barykin, S. E., de la Poza, E., Khalid, B., Kapustina, I. V., Kalinina, O. V., & Iqbal, K. M. (2021). Tourism industry: Digital transformation. In B. Khan, M. Kuofie, & S. Suman (Eds.), *Handbook of research on future opportunities for technology management education* (pp. 414–434). IGI Global. <https://doi.org/10.4018/978-1-7998-8327-2.ch025>

Buhalis, D., & Law, R. (2008). Progress in information technology and tourism management: 20 years on and 10 years after the Internet – The state of eTourism research. *Tourism Management*, 29(4), 609–623. <https://doi.org/10.1016/j.tourman.2008.01.005>

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach* (2nd ed.). Springer.

Cezar, A., & Öğüt, H. (2015). Analyzing conversion rates in online hotel booking: The role of customer reviews, recommendations, and rank order in search listings. *International Journal of Contemporary Hospitality Management*, 27(6), 1344–1365. <https://doi.org/10.1108/IJCHM-05-2014-0249>

Chong, A. Y. L., Li, B., Ngai, E. W. T., Ch'ng, E., & Lee, P. K. C. (2017). Predicting online product sales via online reviews, sentiments, and promotion strategies: A big data architecture and neural network approach. *International Journal of Operations & Production Management*, 37(11), 1534–1552. <https://doi.org/10.1108/IJOPM-03-2015-0151>

Cross, R. G., Higbie, J. A., & Cross, D. Q. (2009). Revenue management's renaissance: A rebirth of the art and science of profitable revenue generation. *Cornell Hospitality Quarterly*, 50(1), 56–81. <https://doi.org/10.1177/1938965508328716>

Del Vecchio, P., Mele, G., Ndou, V., & Secundo, G. (2018). Creating value from social big data: Implications for smart tourism destinations. *Information Processing & Management*, 54(5), 847–860. <https://doi.org/10.1016/j.ipm.2017.10.006>

El-Said, O. A. (2020). Impact of online reviews on hotel booking intention: The moderating role of brand image, star category, and price. *Tourism Management Perspectives*, 33, 100604. <https://doi.org/10.1016/j.tmp.2019.100604>

Field, A. (2018). *Discovering statistics using IBM SPSS Statistics* (5th ed.). Sage.

- Fuchs, M., Höpken, W., & Lexhagen, M. (2015). Applying business intelligence for knowledge generation in tourism destinations – A case study from Sweden. In H. Pechlaner & E. Smeral (Eds.), *Tourism and leisure* (pp. 109–129). Springer. https://doi.org/10.1007/978-3-658-06660-4_11
- Gretzel, U., Sigala, M., Xiang, Z., & Koo, C. (2015). Smart tourism: Foundations and developments. *Electronic Markets*, 25(3), 179–188. <https://doi.org/10.1007/s12525-015-0196-8>
- Gujarati, D. N., & Porter, D. C. (2009). *Basic econometrics* (5th ed.). McGraw-Hill.
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate data analysis* (8th ed.). Cengage.
- Hayes, A. F., & Cai, L. (2007). Using heteroskedasticity-consistent standard error estimators in OLS regression: An introduction and software implementation. *Behavior Research Methods*, 39(4), 709–722. <https://doi.org/10.3758/BF03192961>
- Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (3rd ed.). Wiley. <https://doi.org/10.1002/9781118548387>
- Kotler, P., Kartajaya, H., & Setiawan, I. (2021). *Marketing 5.0: Technology for humanity*. John Wiley & Sons.
- Kotler, P., & Keller, K. L. (2021). *Marketing management* (16th ed.). Pearson.
- Kutner, M. H., Nachtsheim, C. J., Neter, J., & Li, W. (2005). *Applied linear statistical models* (5th ed.). McGraw-Hill/Irwin.
- Long, J. S. (1997). *Regression models for categorical and limited dependent variables*. Sage.
- Long, J. S., & Ervin, L. H. (2000). Using heteroscedasticity consistent standard errors in the linear regression model. *The American Statistician*, 54(3), 217–224. <https://doi.org/10.1080/00031305.2000.10474549>
- Mariani, M. (2020). Big data and analytics in tourism and hospitality: A perspective article. *Tourism Review*, 75(1), 299–303. <https://doi.org/10.1108/TR-06-2019-0259>
- Masiero, L., & Nicolau, J. L. (2015). Choice behaviour in online hotel booking. *Tourism Economics*, 22(3), 671–678. <https://doi.org/10.5367/te.2015.0464>

- Menard, S. (2002). *Applied logistic regression analysis* (2nd ed.). Sage. <https://doi.org/10.4135/9781412983433>
- Moe, W. W., & Fader, P. S. (2004). Dynamic conversion behavior at e-commerce sites. *Journal of Consumer Research*, 30(3), 392–402. <https://doi.org/10.1287/mnsc.1040.0153>
- Mohapatra, D., Mohapatra, D. P., & Dubey, R. S. (2023). Price dispersion across online platforms: Evidence from hotel room prices in London (UK). *Applied Economics*, 56(52), 6598–6610. <https://doi.org/10.1080/00036846.2023.2275219>
- Petković, T., Tepavčević, J., & Blešić, I. (2024). The influence of social networks on the intent to visit hotels. *BizInfo (Blace)*, 15(2), 31–37. <https://doi.org/10.5937/bizinfo2402031P>
- Popović, A., Hackney, R., Coelho, P. S., & Jaklič, J. (2012). Towards business intelligence systems success: Effects of maturity and culture on analytical decision making. *Decision Support Systems*, 54(1), 729–739. <https://doi.org/10.1016/j.dss.2012.08.017>
- Sharda, R., Delen, D., & Turban, E. (2020). *Analytics, data science, and artificial intelligence: Systems for decision support* (11th ed.). Pearson.
- Sigala, M. (2018). New technologies in tourism: From multi-disciplinary to anti-disciplinary advances and trajectories. *Tourism Management Perspectives*, 21, 151–155. <https://doi.org/10.1016/j.tmp.2017.12.003>
- Tussyadiah, I. P., & Zach, F. J. (2017). Identifying salient attributes of peer-to-peer accommodation experience. *Journal of Travel & Tourism Marketing*, 34(5), 636–652. <https://doi.org/10.1080/10548408.2016.1209153>
- Wang, Y., So, K. K. F., & Sparks, B. A. (2021). Technology readiness and customer satisfaction with travel technologies: A cross-country study. *Journal of Travel Research*, 60(5), 563–577. <https://doi.org/10.1177/0047287516657891>
- Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data* (2nd ed.). MIT Press.
- Wooldridge, J. M. (2020). *Introductory econometrics: A modern approach* (7th ed.). Cengage.

- Wu, W., Xu, C., Zhao, M., Li, X., & Law, R. (2024). Digital tourism and smart development: State-of-the-art review. *Sustainability*, *16*(23), 10382. <https://doi.org/10.3390/su162310382>
- Xiang, Z., Du, Q., Ma, Y., & Fan, W. (2015). A comparative analysis of major online review platforms: Implications for social media analytics in hospitality and tourism. *Tourism Management*, *58*, 51–65. <https://doi.org/10.1016/j.tourman.2016.10.001>
- Xiang, Z., Wang, D., O’Leary, J. T., & Fesenmaier, D. R. (2015). Adapting to the Internet: Trends in travelers’ use of the web for trip planning. *Journal of Travel Research*, *54*(4), 511–527. <https://doi.org/10.1177/0047287514522883>
- Xie, K. L., & Lee, Y. J. (2020). Hotels at fingertips: Informational cues in consumer conversion from search, click-through, to book. *Journal of Hospitality and Tourism Technology*, *11*(1), 49–67. <https://doi.org/10.1108/JHTT-03-2017-0026>
- Zhao, X. (R.), Wang, L., Guo, X., & Law, R. (2015). The influence of online reviews to online hotel booking intentions. *International Journal of Contemporary Hospitality Management*, *27*(6), 1343–1364. <https://doi.org/10.1108/IJCHM-12-2013-0542>
- Zikmund, W. G., Babin, B. J., Carr, J. C., & Griffin, M. (2018). *Business research methods* (9th ed.). Cengage.

Anexo I

Variável	Coefficiente	P-valor	IC 95% Inferior	IC 95% Superior
const	- 3,238766455	5,6973E-21	- 3,912785405	-2,564747505
log_Avg_Rate	- 0,736441982	6,45421E-44	- 0,839738105	-0,63314586
Deluxe_binned_1-5	0,245895694	0,009147949	0,061007502	0,430783887
Deluxe_binned_6-20	0,195681776	0,04461083	0,00471126	0,386652293
Deluxe_binned_21-100	0,214114471	0,015220613	0,041218997	0,387009945
Deluxe_binned_>100	0,959387822	2,023E-18	0,745078376	1,173697267
NonRefundable_binned_1	1,04020704	2,26382E-05	0,559233265	1,521180816
NonRefundable_binned_2-5	0,567021103	0,002499622	0,199494781	0,934547424
NonRefundable_binned_>5	- 0,070662935	0,483496348	- 0,268339347	0,127013477
Outros_binned_1	0,590105067	0,154727693	- 0,222699131	1,402909266
Outros_binned_2-5	1,223301657	0,000473127	0,537551165	1,909052148
Outros_binned_>5	0,832036682	5,23436E-09	0,55301384	1,111059523
Room_binned_1-50	0,452024748	2,21589E-05	0,243253281	0,660796214
Room_binned_51-150	1,386866089	3,51974E-35	1,168159709	1,60557247
Room_binned_151-300	1,918107601	3,83362E-60	1,689993935	2,146221268
Room_binned_301-500	1,592339601	4,81459E-28	1,309017667	1,875661534
Room_binned_>500	0,810167553	3,10278E-05	0,429137096	1,19119801
AI_binned_1	- 1,142834694	0,325756476	-3,42237408	1,136704691
AI_binned_2-5	0,666041143	0,220756567	- 0,400094463	1,732176748
AI_binned_>5	0,972596735	7,55312E-10	0,66313848	1,282054991
BB_binned_1-10	0,500434839	1,50921E-06	0,296630991	0,704238686
BB_binned_11-50	0,835153009	7,65808E-14	0,616527233	1,053778786
BB_binned_51-200	1,260869972	2,38764E-30	1,045790778	1,475949167
BB_binned_>200	0,60091628	1,3559E-05	0,330334036	0,871498524
FB_binned_1	0,319074259	0,683185404	- 1,213412991	1,851561509
FB_binned_2-5	1,198020093	0,111518994	- 0,277510441	2,673550626
FB_binned_>5	1,114028316	0,044814897	0,025782022	2,20227461
HB_binned_1	1,256069837	0,000383393	0,563008252	1,949131421
HB_binned_2-5	1,230264906	0,000640202	0,524127552	1,93640226
HB_binned_>5	1,209780758	1,01832E-08	0,79614079	1,623420726

RO_binned_1-10	0,107307145	0,405694564	-0,14565542	0,360269711
RO_binned_11-50	0,499572201	0,000130126	0,24374675	0,755397652
RO_binned_51-200	0,407667884	0,000237232	0,190347684	0,624988084
RO_binned_>200	0,110979969	0,40751966	- 0,151660362	0,373620301

Anexo II

Variável	Coefficiente	p-valor	IC 95% Inferior	IC 95% Superior
log_Avg_Rate	-0,7315	<0,001	-0,834	-0,629
Deluxe_binned_1-5	0,2527	0,007	0,068	0,437
Deluxe_binned_6-20	0,1983	0,042	0,007	0,389
Deluxe_binned_21-100	0,2244	0,010	0,053	0,396
Deluxe_binned_>100	1,0008	<0,001	0,815	1,187
NonRefundable_binned_1	1,0329	<0,001	0,553	1,513
NonRefundable_binned_2-5	0,5772	0,002	0,210	0,944
Outros_binned_2-5	1,2456	<0,001	0,561	1,930
Outros_binned_>5	0,8435	<0,001	0,568	1,119
Room_binned_1-50	0,4544	<0,001	0,249	0,660
Room_binned_51-150	1,3975	<0,001	1,229	1,566
Room_binned_151-300	1,9557	<0,001	1,801	2,110
Room_binned_301-500	1,6360	<0,001	1,423	1,849
Room_binned_>500	0,8693	<0,001	0,620	1,119

AI_binned_>5	0,8911	<0,001	0,649	1,133
BB_binned_1-10	0,4945	<0,001	0,375	0,614
BB_binned_11-50	0,8122	<0,001	0,714	0,910
BB_binned_51-200	1,2071	<0,001	1,119	1,295
BB_binned_>200	0,5102	<0,001	0,356	0,664
FB_binned_>5	1,0780	0,130	-0,306	2,462
HB_binned_1	1,2699	<0,001	0,923	1,617
HB_binned_2-5	1,2763	<0,001	0,911	1,641
HB_binned_>5	1,1686	<0,001	0,824	1,514
RO_binned_11-50	0,4482	<0,001	0,259	0,637
RO_binned_51-200	0,3339	<0,001	0,198	0,470