

Anfitriões no Airbnb: Uma análise segmentada multinível baseada em avaliações, propriedades e características

Maria João Ferreira Gonçalves

M

2024





ANFITRIÕES NO AIRBNB: UMA ANÁLISE SEGMENTADA MULTINÍVEL BASEADA EM
AVALIAÇÕES, PROPRIEDADES E CARACTERÍSTICAS

Maria João Ferreira Gonçalves

Dissertação

Mestrado em Marketing

Orientado por
Professor Doutor Pedro José Ramos Moreira de Campos

2024

Agradecimentos

Em primeiro lugar, gostaria de expressar os meus sinceros agradecimentos a todos aqueles que, de uma forma ou de outra, contribuíram para a realização desta dissertação. Cada gesto de apoio, cada palavra de incentivo e todas as conversas motivadoras fizeram a diferença e foram essenciais para alcançar esta etapa.

Ao meu orientador, Professor Doutor Pedro Campos, o meu profundo agradecimento pela orientação, paciência e disponibilidade ao longo de todo este processo. O seu conhecimento e constante apoio foram cruciais para a realização deste trabalho. Agradeço-lhe também por nunca deixar de acreditar em mim e me dar a confiança necessária para continuar este projeto. Esta dissertação não teria sido possível sem o seu valioso contributo.

Aos meus pais, que sempre me apoiaram incondicionalmente e me deram conselhos ao longo de todo o meu percurso. Sou eternamente grata por tudo o que fizeram e continuam a fazer por mim.

À minha irmã, que, com os seus dotes de culinária, deu-me a força para nunca desistir. Obrigada pela paciência e pelos momentos de partilha.

Ao Paulo, agradeço o seu apoio inabalável, paciência e palavras de encorajamento fundamentais para que eu conseguisse superar todos os desafios que este processo trouxe. Obrigada por estares sempre ao meu lado, por acreditares em mim e por me incentivares a nunca desistir, mesmo nos momentos mais complicados.

Às minhas companheiras de mestrado, que partilharam comigo esta viagem académica cheia de desafios, altos e baixos, mas também de muitas conquistas. O vosso apoio e companheirismo foi essencial. Obrigada por estarem comigo neste percurso.

A todos, o meu sincero obrigado!

Resumo

No contexto da economia de partilha, o Airbnb destaca-se como uma plataforma que conecta anfitriões e hóspedes, oferecendo experiências de alojamento personalizadas. Este estudo apresenta uma metodologia inovadora ao aplicar a segmentação multinível no contexto dos anfitriões do Airbnb, destacando-se por tratar dados estruturados hierarquicamente e interligados, e tem como objetivo identificar os diferentes perfis de anfitriões do Airbnb.

A metodologia adotada incluiu a recolha de dados a partir do site *Inside Airbnb*, focando-se na cidade do Porto. A análise foi conduzida através de duas abordagens principais: o *clustering* uninível, que analisa cada nível de dados de forma independente, e o *clustering* multinível, que permite uma análise mais complexa, considerando a interdependência entre diversos fatores. Esta metodologia permitiu uma segmentação mais detalhada dos diferentes tipos de anfitriões.

Os resultados evidenciaram a existência de variados perfis de anfitriões. E a análise multinível demonstrou ser a metodologia mais eficaz, oferecendo uma segmentação mais precisa dos mesmos. Em contraste, o *clustering* uninível, apresentou resultados menos detalhados, perdendo nuances importantes nas interações entre os níveis de dados.

Este estudo apresenta uma abordagem inovadora ao aplicar a segmentação multinível no contexto dos anfitriões do Airbnb, oferecendo uma visão mais detalhada sobre as interações entre os diferentes níveis de dados. Contribui para a literatura existente sobre alojamento P2P ao analisar o impacto das características dos anfitriões e das propriedades na satisfação dos hóspedes, e oferece insights práticos para anfitriões e gestores de plataformas sobre como otimizar a gestão das suas propriedades e interações com os clientes.

Palavras-chave: Clustering multinível, Airbnb, anfitriões, alojamento P2P, segmentação de mercado.

Abstract

In the context of the sharing economy, Airbnb stands out as a platform that connects hosts and guests, offering personalized lodging experiences. This study introduces an innovative methodology by applying multilevel segmentation to Airbnb hosts, focusing on hierarchically structured and interconnected data. The primary objective is to identify the different profiles of Airbnb hosts.

The methodology involved data collection from the Inside Airbnb website, focusing on hosts in the city of Porto. The analysis was conducted using two main approaches: single-level clustering, which examines each data level independently (reviews, properties, and hosts), and multilevel clustering, which provides a more complex analysis by considering the interdependence between guest reviews, property characteristics, and host profiles. This methodology allowed for a more detailed segmentation of different types of hosts.

The results highlighted the existence of various host profiles, ranging from professional hosts to individual hosts with a single property. The multilevel analysis proved to be the most effective methodology, offering a more precise segmentation of hosts. In contrast, single-level clustering, though simpler, yielded less detailed results, missing important nuances in the interactions between the data levels.

This study presents an innovative approach by applying multilevel segmentation in the context of Airbnb hosts, providing a more detailed view of the interactions between different data levels (reviews, properties, and hosts). It contributes to the existing literature on peer-to-peer accommodation by analyzing the impact of host and property characteristics on guest satisfaction, offering practical insights for hosts and platform managers on how to optimize property management and interactions with customers.

Keywords: Multilevel clustering. Airbnb, hosts, P2P accommodation, market segmentation.

Índice

Agradecimentos.....	i
Resumo.....	ii
Abstract.....	iii
1. Introdução.....	1
2. Revisão de Literatura.....	6
2.1. Economia de Partilha.....	6
2.1.1. <i>Conceito</i>	6
2.1.2. <i>Economia partilhada vs. Economia de plataforma</i>	7
2.1.3. <i>A economia partilhada e o Turismo</i>	8
2.2. Alojamentos Peer-to-peer (P2P).....	9
2.2.1. <i>Conceito e surgimento</i>	9
2.2.1. <i>Motivação de uso e Experiências Co-criadas (Benefícios e Riscos)</i>	10
2.3. Airbnb.....	10
2.3.1. <i>O modelo de negócio</i>	10
2.3.2. <i>Evolução do Airbnb em Portugal</i>	11
2.4. Os Anfitriões.....	13
2.4.1. <i>Perfil do anfitrião</i>	13
2.4.2. <i>Avaliações</i>	14
2.4.3. <i>Propriedades</i>	14
2.5. Clustering multinível e clustering uninível.....	15
3. Metodologia.....	18
3.1. Problema e questão de investigação.....	18
3.2. Modelo conceptual e hipóteses de investigação.....	19
3.3. Contexto da Investigação.....	22
3.4. Metodologia de Investigação.....	23
3.4.1. <i>Recolha e Organização dos Dados</i>	25
3.4.2. <i>Análise Multi-nível e Análise Uninível</i>	28
4. Análise e Discussão de Resultados.....	34
4.1. Análise descritiva dos clusters.....	34
4.2. Comparação Modelo A e B.....	38
4.2.1. <i>Indicadores de Qualidade dos Clusters</i>	38

4.2.2. <i>Análise da Matriz de Confusão entre os clusters</i>	39
4.2.3. <i>Análise da concordância entre os clusters</i>	40
4.2.4. <i>Discussão de resultados</i>	41
5. Conclusão.....	44
5.1 Considerações finais	44
5.2. Contribuições teóricas e práticas	45
5.3 Limitações e direções de pesquisa futura	47
Referências Bibliográficas	50
Anexos	58
Anexo 1 – Literatura existente acerca dos tipos de anfitriões.....	58
Anexo 2- Número de anúncios na plataforma Airbnb em Lisboa em 2024 por freguesia	59
Anexo 3- Número de anúncios na plataforma Airbnb no Porto em 2024 por freguesia	59
Anexo 4- Descrição da implementação do programa em python.....	59
Anexo 5 – Comparação Tabela de Contingência	61
Anexo 6 – Comparação dos Centroides A vs B.....	61
Anexo 7 - Centroides das Métricas.....	61
Anexo 8 - Discussão geral de resultados.....	62

Índice de Figuras

Figura 1- Evolução do número de anúncios no Airbnb	12
Figura 2- Modelo Conceptual.....	21
Figura 3- Modelo A	22
Figura 4- Modelo B	22
Figura 5 - Estrutura da análise de dados do Modelo A	30
Figura 6- Estrutura da análise de dados ao nível das reviews	31
Figura 7- Estrutura da análise de dados ao nível das propriedades	31
Figura 8- Estrutura da análise de dados ao nível dos anfitriões	32
Figura 9- Estrutura geral da análise dos dados	33
Figura 10 - Dados resultantes dos clusters do Modelo A	35
Figura 11 - Dados resultantes dos clusters do Modelo B.....	37

Índice de Tabelas

Tabela 1- Definições Economia de Partilha, por ordem cronológica	7
Tabela 3- Metodologias de estudos relacionados	24
Tabela 4 - Avaliação da coesão dos clusters	39

1. Introdução

O turismo tem representado um dos pilares fundamentais da economia portuguesa, evidenciado até junho de 2024, marcos impressionantes no setor, com a receção de mais de 14 milhões de hóspedes e de mais de 35 milhões de dormidas (INE, 2024). Estes números destacam-se por Portugal se ter transformado num destino turístico global, e demonstram o crescimento do turismo face a 2023. O ano de 2024 foi importante pois ultrapassou pela primeira vez os dados de 2019, isto porque depois da pandemia os números diminuíram drasticamente e apenas em 2023 se registou um aumento (INE, 2024). Assim, é possível verificar que foi alcançado um novo pico em 2024 do número de hóspedes e dormidas, resultante do aumento das infraestruturas turísticas e da diversificação das experiências oferecidas aos visitantes.

Devido à grande oferta de alojamentos disponíveis, a ascensão da procura por alojamentos alternativos, impulsionada por turistas que procuram experiências mais autênticas e personalizadas, catalisou a emergência de alojamentos Peer-to-Peer (P2P) e, por consequente, plataformas dedicadas a facilitar essas transações (Sthapit & Jiménez-Barreto, 2018). Aplicações de plataformas como o Airbnb revolucionaram a maneira como as pessoas viajam, permitindo que qualquer pessoa com algum espaço disponível possa rentabilizá-lo, transformando residências privadas em potenciais estabelecimentos de hospedagem (Vartiak et al., 2023). A popularidade destas plataformas evidencia uma mudança significativa no comportamento do consumidor e destaca a importância da flexibilidade, conveniência e autenticidade nas escolhas de alojamento dos novos turistas.

No caso da plataforma Airbnb, os anfitriões, proprietários dos estabelecimentos, assumem um papel ainda mais importante em relação às outras plataformas, isto porque o Airbnb opera num modelo de negócio que equilibra as necessidades e a importância de anfitriões e hóspedes (Qiu et al., 2023), estabelecendo uma relação simbiótica entre as duas partes. Os anfitriões, neste contexto, transcendem o típico papel de meros fornecedores de alojamento, passando a atuar como parceiros cruciais na criação de experiências de viagem memoráveis (Fradkin et al., 2021). O sistema de *reviews* mútuas do Airbnb, permite que tanto anfitriões quanto hóspedes se avaliem, e é um pilar fundamental na preservação da confiança na plataforma. (Airbnb.inc, 2023) As avaliações abrangem vários aspetos da experiência, desde a precisão do anúncio até ao cumprimento das regras da casa. Este sistema ajuda os usuários a tomar decisões informadas, e contribui para a melhoria contínua da qualidade do

serviço oferecido pela plataforma, estabelecendo um ciclo virtuoso de feedback e melhorias constantes (Xue et al., 2022).

Neste cenário próspero, o aumento de anfitriões de alojamentos P2P tornou-se uma tendência, impulsionada pela procura de alojamentos que ofereçam uma imersão mais autêntica na cultura local (A. Belarmino & Y. Koh, 2020). Em Portugal, esta tendência é ainda mais notável, visto que é um país onde a hospitalidade é uma característica enraizada e valorizada (Sousa et al., 2021). Os anfitriões locais têm desempenhado um papel crucial em atender à crescente procura, fornecendo alojamentos, com experiências personalizadas que refletem o verdadeiro espírito português (Sthapit, Björk, Coudounaris, et al., 2022).

O aumento significativo do número de anfitriões na plataforma Airbnb gerou a necessidade de compreender e segmentar estes diferentes tipos de anfitriões de forma mais detalhada. A maioria dos estudos concentra-se nas avaliações como o principal critério para avaliar o desempenho e o impacto dos anfitriões (D. Chen et al., 2024). No entanto, esta abordagem limita a análise ao não considerar outras variáveis, como as características das propriedades e as práticas de gestão dos anfitriões (Farmaki & Kaniadakis, 2020). Para colmatar esta lacuna, torna-se essencial adotar uma análise que integre múltiplas características, permitindo a identificação de perfis distintos de anfitriões, baseando-se em dados além das avaliações. Estudos recentes, como os de Abrate et al. (2022) e Tussyadiah (2016), já começaram a explorar a segmentação de anfitriões com base em variáveis adicionais, como a motivação económica, o número de propriedades geridas e a experiência acumulada. Este tipo de variáveis permite captar de forma mais precisa a diversidade de práticas e interações personalizadas que os anfitriões utilizam para enriquecer a experiência dos hóspedes, proporcionando uma visão mais holística e integrada do ecossistema do Airbnb.

A adoção de uma abordagem metodológica mista, conforme sugerido por Qiu et al. (2023), é essencial para compreender a complexidade dos anfitriões no Airbnb. Dado que, os estudos anteriormente referidos, apenas analisam as variáveis relativas ao anfitrião de forma isolada e não reconhecem a existência da ligação natural que existe entre elas, pois um anfitrião é caracterizado não só pelas suas avaliações, mas também pelas suas características e pelas suas propriedades. Assim, existe a necessidade de capturar uma visão mais completa e multidimensional dos anfitriões, integrando tanto as perspetivas dos hóspedes, expressas nas avaliações, quanto outros atributos dos anfitriões que impactam a experiência de hospedagem. Tal abordagem enriquece a análise, destacando não apenas o que é

explicitamente avaliado pelos hóspedes, mas também os fatores subtis e menos óbvios que contribuem para a definição de um hospede.

Face a isso, é imperativo representar esta conexão observada nestas variáveis, tendo em conta que os dados referentes aos anfitriões, avaliações e propriedades no Airbnb estão interligados (Zal et al., 2024). Assim, a análise destas variáveis não deve ser realizada de forma independente, já que atuam conjuntamente, interagindo e influenciando-se entre si (Zamani et al., 2019). Nesse sentido, as metodologias multinível surgem como as mais apropriadas para captar tal complexidade, uma vez que possibilitam o tratamento de dados organizados em diferentes camadas, permitindo compreender de forma mais detalhada estas interações entre as variáveis (Zal et al., 2024).

Neste seguimento, esta investigação tem como principal foco os anfitriões do Airbnb, com o objetivo de identificar e segmentar os diferentes tipos de anfitriões na plataforma, utilizando uma análise inovadora, que vai além das abordagens tradicionais da literatura, que pouco exploraram a aplicação de segmentação multinível. Assim, esta investigação irá explorar e compreender o mercado dos anfitriões, através da realização de uma segmentação uninível e multinível baseada nas avaliações dos alojamentos, nas informações acerca dos mesmos e dos seus alojamentos, de forma a responder às seguintes questões de investigação:

Q1: Qual das metodologias (*clustering* multinível ou *clustering* uninível) é mais eficaz na segmentação de mercado dos anfitriões?

Q2: Quais os diferentes tipos de anfitriões que existem dentro da plataforma Airbnb?

A nível teórico, é de esperar que este estudo enriqueça a literatura sobre economia de partilha e alojamentos peer-to-peer (P2P), fornecendo uma análise detalhada das metodologias de segmentação multinível aplicadas ao contexto de plataformas como o Airbnb. Ao integrar variáveis que vão além das avaliações dos hóspedes, contribui para uma compreensão mais profunda e holística dos diferentes perfis de anfitriões e das suas práticas de gestão, abordando uma lacuna significativa na investigação existente. Esta análise inovadora oferece *insights* sobre como as interações entre os anfitriões, as suas propriedades e as avaliações dos hóspedes moldam os diferentes tipos de segmentos, uma perspetiva ainda pouco explorada na academia. Para além disso, espera-se que esta investigação abra novas vias de pesquisa sobre segmentação de mercado em plataformas P2P e promova o uso de metodologias multinível noutras áreas de estudo onde os dados hierárquicos estão presentes.

Em termos práticos, os resultados deste estudo oferecem *insights* valiosos para

diferentes partes interessadas no ecossistema do Airbnb. Para os anfitriões, os resultados da segmentação multinível podem fornecer orientações sobre como otimizar as suas práticas de gestão e melhorar a experiência dos hóspedes, ao identificar os fatores específicos que influenciam as avaliações positivas e o sucesso das suas propriedades. Para os gestores de plataformas como o Airbnb, os resultados ajudam a refinar as estratégias de apoio e formação oferecidas aos anfitriões. Ao compreender os diferentes perfis e como cada grupo de anfitriões se relaciona com o sucesso das suas propriedades, a plataforma pode adaptar a comunicação e o apoio fornecido a cada tipo, melhorando a experiência de quem usa o serviço e aumentando a competitividade no mercado.

Para responder às questões de investigação e atingir os objetivos definidos, este estudo adotou uma abordagem quantitativa, centrada na análise de *clusters* uninível e multinível. A amostra foi composta por dados recolhidos a partir da plataforma *Inside Airbnb*, que disponibiliza informações públicas sobre anfitriões, propriedades e avaliações de hóspedes. Para garantir uma análise adequada, foram extraídos dados de 10 anfitriões com um número significativo de propriedades na área do Porto, selecionando 10 propriedades por anfitrião e 10 avaliações por propriedade, culminando num total de 100 propriedades e 1000 avaliações. As variáveis analisadas incluíram características dos anfitriões (e.g., número de propriedades, taxa de resposta), características das propriedades (e.g., tipo de propriedade, localização), e as avaliações dos hóspedes em seis dimensões (e.g., limpeza, precisão, relação qualidade/preço). A análise de dados foi conduzida em duas fases principais. Primeiro, foi aplicada uma abordagem uninível, onde cada nível de dados (avaliações, propriedades e anfitriões) foi analisado de forma independente. Em seguida, a abordagem multinível foi utilizada para integrar os dados em diferentes níveis, permitindo uma visão mais abrangente e detalhada das interações entre este tipo de dados. Esta combinação de metodologias permite comparar os resultados de ambas as abordagens e identificar qual oferece uma segmentação mais precisa e útil dos anfitriões.

Posto isto, esta dissertação está estruturada em cinco secções. A primeira, corresponde à introdução, oferece uma visão geral do estudo, justificando a sua relevância teórica e prática, assim como a metodologia adotada. Na secção seguinte, é apresentada a revisão da literatura, abordando os conceitos centrais de economia de partilha, alojamentos peer-to-peer, e plataformas como o Airbnb. A secção 2.1 explora as principais abordagens metodológicas para segmentação, com um foco particular no clustering multinível e uninível. A secção 3 descreve em pormenor a metodologia utilizada para a recolha e análise dos dados,

enquanto a secção 4 é dedicada à apresentação e discussão dos resultados, com especial destaque para a comparação entre as abordagens multinível e uninível. Por fim, na secção 5, são apresentadas as considerações finais, que incluem as principais conclusões do estudo, as suas limitações e sugestões para futuras investigações.

2. Revisão de Literatura

Este capítulo apresenta uma revisão da literatura que procura ser abrangente, explorando as nuances e complexidades dos anfitriões no Airbnb e todo o seu envolvente. Esta revisão da literatura sintetiza os conhecimentos existentes no tema, e identifica lacunas estabelecendo uma base para futuras investigações.

2.1. Economia de Partilha

2.1.1. Conceito

Embora partilhar não seja um conceito novo (Felson & Spaeth, 1978), com o aumento de várias práticas e serviços relacionados com este conceito, a investigação acerca da economia de partilha expandiu-se e, por isso, o conceito de economia de partilha foi definido de várias maneiras por vários autores.

Na literatura existente acerca da economia de partilha existem autores que relacionam a economia partilhada com a necessidade de colaboração e de se tratar de um produto ou serviço peer-to-peer (Hamari et al., 2016), seguindo o conceito mais tradicional de que, nesta economia, o consumo é colaborativo através da partilha de bens que não são utilizados em toda a sua extensão de forma comercial. No entanto, existem autores que descrevem a economia de partilha como uma economia de acesso (Khalek & Chakraborty, 2023), que é caracterizada pela prestação de um serviço a pedido dos clientes, facilitando o acesso a certos recursos sem haver a necessidade de propriedade por parte dos clientes, trata-se de um serviço *on-demand*, ou seja, os consumidores podem solicitar e utilizar esses recursos somente quando necessário, ao invés de possuí-los permanentemente.

Gerwe e Silva (2018) definem a economia de partilha como um sistema que permite o acesso a bens físicos e humanos pouco utilizados através de plataformas online que facilitam as interações entre utilizadores. Este conceito baseia-se em 4 pilares que são a capacidade subutilizada, como um lugar a mais num carro, o acesso temporário, onde não é necessário ter propriedade de um bem, as transações de pessoa para pessoa, operações que não envolvem negócios ou empresas e, por fim, as plataformas online que facilitam as transações entre utilizadores.

Tabela 1- Definições Economia de Partilha, por ordem cronológica

Autor	Definições
Felson and Spaeth (1978)	“Eventos em que uma ou mais pessoas consomem bens ou serviços económicos no processo de envolvimento em atividades conjuntas com uma ou mais outras pessoas.”
Hamari et al. (2016)	“A atividade baseada na troca de informações entre pares [peer-to-peer] de obter, dar ou partilha do acesso a bens e serviços, coordenada através de serviços online de base comunitária.”
Gerwe e Silva (2018)	“Um sistema socioeconómico que permite aos pares conceder acesso temporário aos seus ativos físicos e humanos subutilizados através de plataformas online.”
Khalek & Chakraborty (2023)	“Sistema de intercambio que designa várias transações realizadas conforme a procura medianas por plataformas digitais, principalmente em colaboração entre indivíduos, que defendem a utilização de recursos ociosos sem necessidade de propriedade. O SE é amplamente identifica de como uma convergência de dois modos distintos de consumo - consumo colaborativo e o consumo baseado no acesso.”

Como é visível, existe alguma divergência acerca dos conceitos da economia de partilha. Existem autores que defendem que a economia partilhada não inclui uma remuneração financeira (Belk, 2007), enquanto outros acreditam que a economia de partilha é diferente do conceito de partilha, sendo esta mais orientada para o lucro do que para a orientação social (Saravade et al., 2021). O desafio ao definir a economia de partilha é no momento que entram as plataformas digitais, pois já não se trata apenas de ações de partilha, mas sim de transações dentro de uma plataforma onde, na maioria das vezes, são realizadas transações financeiras que geram lucros para os donos da plataforma, indo contra o conceito inicial da partilha.

2.1.2. Economia partilhada vs. Economia de plataforma

A economia de partilha, no mundo digital, começou como uma extensão das práticas

de partilha tradicionais, como a partilha de boleias, mas rapidamente se transformou devido à ascensão da internet e das plataformas online (Ke Rong & Yining Luo, 2023).

Acquier et al. (2017) defende que a economia partilhada se baseia em 3 pilares: na economia de acesso, na economia de plataforma e na economia baseada na comunidade. A economia de plataforma é então definida por empresas que utilizam tecnologia digital para facilitar transações entre consumidores e fornecedores de bens ou serviços. Este modelo económico baseia-se em plataformas digitais que conectam diretamente os consumidores e permitem a troca, compra, ou venda de serviços e produtos. Exemplos disso são a Uber e o Airbnb (Schor et al., 2020). Esta economia tem uma relação intrínseca com a economia partilhada. A economia partilhada, ou economia colaborativa, é um sistema onde indivíduos compartilham bens ou serviços, muitas vezes facilitado por uma plataforma digital. Embora a economia de plataforma não se limite apenas à partilha, muitas das suas manifestações mais conhecidas são, de facto, formas de economia partilhada.

Apesar de muitos autores afirmarem que a economia de plataforma faz parte da economia de partilha (Ke Rong & Yining Luo, 2023; Wirtz et al., 2019), é importante perceber a distinção destes dois conceitos. Na economia de plataforma existem transações comerciais entre consumidores e fornecedores, o que vai contra o conceito inicial de partilha, a partilha de acesso a bens ou serviços, sem uma transação monetária direta. Posto isto, surge a questão se o Airbnb e a Uber, por exemplo, são realmente exemplos de economia de partilha, pois apesar utilizarem o conceito de partilha, a sua estrutura sugere que se inclina mais para um modelo de negócios comercial do que para a ideia pura de partilha colaborativa, característica da economia de partilha (Hall et al., 2022).

Portanto, embora existam sobreposições, principalmente devido ao uso de plataformas digitais para facilitar a partilha de recursos, a economia de plataforma e a economia de partilha são distintas. A economia de plataforma abrange uma gama mais ampla de atividades comerciais e não se limita ao compartilhamento no sentido puro da palavra (Acs et al., 2021).

2.1.3. A economia partilhada e o Turismo

A economia de partilha tem exercido um impacto notável no setor do turismo (K. Rong & Y. Luo, 2023). Caracterizada pelo uso de plataformas online que facilitam a partilha de recursos entre indivíduos, esta nova economia transformou a maneira como as pessoas viajam e visitam novos destinos (Valentinas et al., 2021).

Uma das indústrias mais afetadas pela economia de partilha foi a indústria hoteleira, pois plataformas como Airbnb e VRBO apresentaram alternativas mais atraentes e destacaram-se dos hotéis (Wirtz et al., 2019). Além de afetar as taxas de ocupação e os preços, a economia partilhada também tem influenciado as expectativas dos hóspedes. Existe cada vez mais uma procura por experiências personalizadas e autênticas, um dos pontos mais fortes dos alojamentos compartilhados o que desafia os hotéis a inovar em serviços e ofertas (Chang & Sokol, 2022).

Ainda existem muitos turistas que preferem o ambiente e os serviços de um hotel tradicional. No entanto, muitos destes estabelecimentos têm vindo a adaptar-se, incluindo novas tecnologias para também melhorar a experiência do hospede (Srovnalíkova et al., 2020).

O aluguer de carros era também um tipo de indústria bastante comum no turismo, mas com o aparecimento de economia partilhada, plataformas como a Uber e a Bolt surgiram e mudaram para sempre o setor dos transportes privados (Boateng et al., 2019).

Assim, a economia de partilha está a alterar o conceito de turismo em vários aspetos, criando tanto oportunidades de novos negócios, como desafios para os negócios mais tradicionais. Esta oferece aos turistas mais escolhas e experiências personalizadas, o que força outros negócios a inovarem e a revolucionarem-se.

2.2. Alojamentos Peer-to-peer (P2P)

2.2.1. Conceito e surgimento

O surgimento da economia da partilha, também conhecida como economia colaborativa, juntamente com a alteração nas preferências dos consumidores, desempenhou um papel fundamental na emergência dos alojamentos P2P. Este fenómeno reflete uma mudança significativa na forma como as pessoas escolhem viver as suas experiências de viagem e interagem com os serviços de alojamento (Sthapit & Jiménez-Barreto, 2018).

Este conceito revolucionou a indústria da hospedagem, baseando-se no princípio de que as pessoas podem alugar as suas propriedades, como casas ou quartos que não estão a ser usados para turistas, a um preço mais económico, em alternativa aos hotéis tradicionais. Esta ideia, conforme indicado pelo nome 'peer-to-peer', envolve uma transação direta entre indivíduos (Belk, 2014). O elemento que realmente distingue e dá singularidade ao P2P é a possibilidade de estabelecer interações mais extensas e pessoais entre o consumidor (os hóspedes) e o fornecedor (os anfitriões) (Amanda Belarmino & Yoon Koh, 2020).

À medida que o alojamento P2P se tornou uma questão relevante no mundo do

turismo e da hotelaria, a investigação acerca deste fenómeno tem vindo a aumentar. Este interesse académico tem intensificado o debate sobre os impactos do alojamento P2P, tanto a nível económico quanto sociocultural. Os estudos e discussões em torno deste tema têm procurado compreender melhor as implicações que este modelo de hospedagem tem nas dinâmicas locais e globais do turismo (Cheng et al., 2022; Nakamura et al., 2024).

2.2.1. Motivação de uso e Experiências Co-criadas (Benefícios e Riscos)

Os alojamentos P2P, como dito anteriormente, diferenciam-se ao oferecerem uma experiência mais individual e particular a cada um dos hóspedes. Os utilizadores de alojamentos P2P são frequentemente motivados pela procura de experiências autênticas e interações culturais significativas (A. Belarmino & Y. Koh, 2020).

Segundo La et al, (2022), a localização é um dos maiores motivos de escolher este tipo de alojamento, devido a conseguirem ter acesso a recursos mais convenientes que normalmente não iriam ter acesso no alojamento tradicional como os hotéis. Ao usufruírem de uma melhor localização, os hóspedes afirmam usufruírem de uma estadia como se fossem residentes do local (Garg, 2020).

No entanto, fatores como avaliações positivas, comunicação transparente, e informações detalhadas sobre o anfitrião contribuem para a criação de uma sensação de confiança (Park & Tussyadiah, 2020). Conhecer o anfitrião, mesmo que virtualmente, pode acrescentar um elemento pessoal à experiência da estadia. Descrições detalhadas com informações sobre a propriedade, o motivo de se terem tornado anfitriões e as suas interações anteriores com os hóspedes, podem fazer com que estes se sintam mais conectados e seguros (Dongzhi Chen et al., 2024).

Num alojamento peer-to-peer existe uma grande probabilidade que os hóspedes tenham variados encontros com o anfitrião, o que resultam numa troca de interações e impressões (Meng & Cui, 2020). O Airbnb é um dos melhores exemplos de plataformas que capacitam este tipo de experiências, isto porque os hóspedes no Airbnb não são vistos como agentes passivos, estes estão envolvidos ativamente na criação de suas experiências de viagem, interagindo com os anfitriões e o ambiente ao seu redor, originando experiências co-criadas (Sthapit, Björk, & Coudounaris, 2022).

2.3. Airbnb

2.3.1. O modelo de negócio

Ao reinventar as estadias em casas particulares e as experiências turísticas, o Airbnb

trouxe ao setor do turismo uma das inovações mais fascinantes da última década. Esta plataforma alterou para sempre os alojamentos de curta duração em cidades que apresentam uma alta procura, disponibilizando um tipo de alojamento mais acessível, o alojamento P2P (Dolnicar, 2019), provocando reações intensas no setor comercial estabelecido de alojamento turístico (Zach et al., 2019).

Atualmente, o Airbnb é um exemplo paradigmático de uma plataforma de mercado de duas faces, que conecta anfitriões com hóspedes, criando um ecossistema no qual ambos os lados do mercado são atendidos simultaneamente, beneficiando-se mutuamente do serviço (Fradkin et al., 2021). Os serviços oferecidos pelo Airbnb permitem a maximização da utilização de quartos ou habitações subutilizadas (Vartiak et al., 2023). O funcionamento do Airbnb é inteiramente digital, estendendo-se desde a publicação do anúncio do espaço de alojamento (pelo anfitrião) até à partilha de impressões sobre a estadia na forma de uma avaliação (pelo hóspede). Deste modo, e tendo como principal objetivo a preservação da confiança entre os utilizadores da plataforma, o Airbnb possibilita que tanto anfitriões quanto hóspedes realizem avaliações mútuas (Airbnb, Inc, 2023). Este mecanismo incentiva um comportamento responsável de ambas as partes e oferece aos anfitriões uma ferramenta para expressar suas experiências com os hóspedes (Xue et al., 2022).

O modelo de negócio do Airbnb distingue-se significativamente de plataformas de reservas de alojamentos tradicionais como o Booking.com, especialmente no que diz respeito ao papel dos anfitriões e à dinâmica de avaliações recíprocas. Ao contrário de plataformas como o Booking.com, onde o foco principal está na experiência do hóspede (Vartiak et al., 2023), o Airbnb estabelece um equilíbrio entre as necessidades e importância dos anfitriões e dos hóspedes (Qiu et al., 2023). Esta abordagem bidirecional garante que ambas as partes tenham voz ativa na plataforma, promovendo um ambiente de respeito mútuo e cooperação (Fradkin et al., 2021). Os anfitriões no Airbnb não são apenas fornecedores de alojamento, mas também parceiros essenciais que contribuem ativamente para a experiência única de cada hóspede. Por isso, o Airbnb desenvolveu uma comunidade de anfitriões diversificada, onde partilham experiências, dicas e práticas recomendadas (Sthapit, Björk, Coudounaris, et al., 2022). Devido a esta vasta rede de anfitriões com uma ampla variedade de estilos de vida, experiências e diversidade cultural é necessário compreender os diferentes tipos de anfitriões existentes na plataforma (Qiu et al., 2023).

2.3.2. Evolução do Airbnb em Portugal

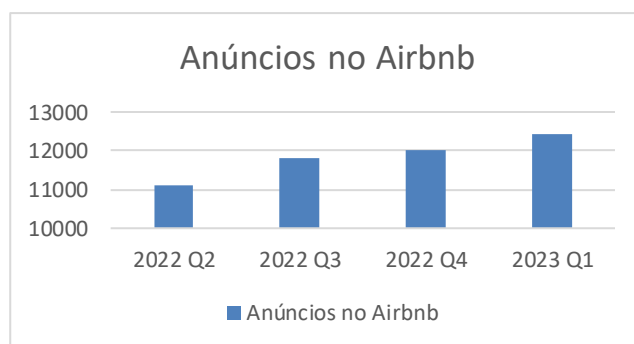
O Airbnb registou um crescimento expressivo desde que iniciou operações em

Portugal. Em 2009, os anfitriões portugueses começaram a receber hóspedes pela plataforma, e desde então o número de anúncios e reservas aumentou significativamente. No final de 2023, o Airbnb contava com 13 602 propriedades na região do Porto registadas na plataforma e 780 922 avaliações nas propriedades. Em junho de 2024, registou um aumento, contando com 14 306 propriedades na região do Porto e 849 963 avaliações nas propriedades. Dado isto, na região do Porto, em menos de um ano, o número de propriedades aumentou 5% e o número de avaliações registou um aumento de 9%. Já em relação a Lisboa, em 2023 o Airbnb registava 22 605 propriedades e 1 329 832 de avaliações nas propriedades e até junho de 2024 registou 23 624 propriedades, representando um aumento de 5% e 1 468 635 avaliações, representando um aumento de 10%. (Inside Airbnb, 2024) Além disto, o Airbnb em 2023, devolveu aos municípios do Porto e Lisboa 14.9 milhões de euros, relativamente à taxa turística cobrada aos turistas, representando um aumento de 10% face a 2022. ([Dinheiro Vivo, 2024](#))

Este crescimento não se limitou às grandes cidades, estendendo-se a áreas rurais e regiões menos conhecidas, como o Alentejo e as ilhas dos Açores e Madeira. Em 2022, dois terços das estadias no Airbnb foram feitas fora das principais áreas urbanas, demonstrando a tendência de dispersão do turismo .

Nos dados da Figura 1 é possível observar que no espaço de um ano verificou-se um aumento de cerca de 10% no número de alojamentos. com estes dados não podemos afirmar se esta tendência já existia antes e durante a pandemia.

Figura 1- Evolução do número de anúncios no Airbnb



Comparando o número de anúncios total das cidades de Lisboa (22 605) e Porto (13 448) em 2024, é possível observar que a capital apresenta um número superior de anúncios e em comparação com 2019, Lisboa (22 242) apresentou um crescimento no número de

anúncios, ao contrário do Porto (16 230). A comparação com o ano de 2019 é necessária pois foi o maior pico de anúncios e reservas de sempre, num ano antes da pandemia, pois este afetou significativamente o número de reservas (Santos & Oliveira Moreira, 2021).

Como observado nos Anexos 2 e 3, as cinco freguesias com mais anúncios em Lisboa são Santa Maria Maior (13.7%), Misericórdia (10.3%), Arroios (8.8%), Cascais e Estoril (6.4%) e Santo António (6%), enquanto no Porto as cinco freguesias com mais anúncios no Porto são Cedofeita, Ildefonso, Sé, Miragaia, Nicolau, Vitória (50.5%), Bonfim (10.1%), Santa Marinha e São Pedro da Afurada (5 %), Lordelo de Ouro e Massarelos (4%) e Paranhos (4%)

2.4. Os Anfitriões

2.4.1. Perfil do anfitrião

O papel significativo desempenhado pelos anfitriões do Airbnb no crescimento do sucesso do alojamento entre pares (P2P) é amplamente reconhecido na literatura. Os anfitriões do Airbnb são considerados parceiros essenciais no serviço de partilha e entre membros proeminentes da comunidade online do Airbnb. Estes asseguram que os serviços de qualidade sejam continuamente prestados sob a marca “Airbnb”, num ambiente de alta procura. A importância dos perfis dos anfitriões no Airbnb reside no impacto significativo que estes perfis têm sobre as decisões dos hóspedes. Segundo Hans et al. (2024) os sinais sociais e emocionais transmitidos pelas imagens de perfil dos mesmos influenciam a intenção de reserva dos hóspedes. Aspetos como a legibilidade, a expressão de sentimentos positivos e a perspetiva centrada no hóspede ajudam também a construir uma perceção de confiança (Zhang et al., 2020).

Além disso, as descrições que refletem a personalidade do anfitrião e detalham a experiência oferecida podem diferenciar os alojamentos, tornando-os mais atraentes para os hóspedes. No entanto, os estudos atuais acerca das descrições do anfitrião focam-se principalmente no estudo do vocabulário e das frases e não em características a um nível mais abrangente, como a identificação de diferentes estratégias de anfitriões (Yan et al., 2023). Estudos como o conduzido por Tussyadiah exploram as estratégias de autoapresentação, assim, os perfis dos anfitriões foram explorados mais detalhadamente para identificar diferenças através das descrições e identificou cinco tipos de anfitriões: o cidadão global, o especialista local, o pessoal, o sucedido e o criativo (Tussyadiah, 2016).

2.4.2. Avaliações

Dentro da literatura emergente sobre os anfitriões do Airbnb, existem numerosos estudos que se concentram nas interações entre anfitriões e hóspedes, visto que estas são fundamentais para o comportamento dos anfitriões. Farmaki e Stergiou (2019) revelam que as interações sociais entre os mesmos podem ser um fator importante na mitigação de sentimentos de solidão para ambas as partes. Moon et al. (2019) indicam ainda que a satisfação e as intenções comportamentais de anfitriões e hóspedes tendem a ser atribuídas às suas interações.

Neste contexto, as avaliações dos anfitriões no Airbnb assumem um papel crucial. Estas avaliações não apenas refletem a qualidade da estadia e a experiência do hóspede, mas também influenciam diretamente a reputação e o sucesso dos anfitriões na plataforma. Estudos como o de Ye (2023) destacam que hóspedes com avaliações históricas positivas tendem a ter experiências mais satisfatórias em reservas futuras, sugerindo uma relação bidirecional entre satisfação do hóspede e as avaliações do anfitrião. Assim, tornam-se um elemento-chave na construção de confiança e na promoção de experiências positivas tanto para os hóspedes quanto para os anfitriões.

No Airbnb, as avaliações desempenham um papel fundamental na forma como os hóspedes e anfitriões interagem, afetando a percepção de qualidade e a confiança na plataforma. Para além dos comentários em formato de texto, os hóspedes têm de avaliar vários aspetos da estadia que são considerados essenciais para a análise da experiência. Entre esses aspetos encontram-se: limpeza, que avalia o estado de higiene da propriedade; comunicação com o anfitrião, relacionada com a facilidade e prontidão na troca de informações; check-in, que abrange a simplicidade e conveniência do processo de entrada na propriedade; localização, referindo-se à proximidade de pontos de interesse e à adequação do local; relação preço-qualidade, que mede se o custo está alinhado com a qualidade da estadia; e precisão, avaliando se a descrição da propriedade corresponde à realidade percebida pelo hóspede. Estes critérios garantem uma avaliação mais objetiva da experiência e ajudam outros utilizadores a tomar decisões mais informadas (Airbnb, Inc., 2023; Xue et al., 2022). O sistema de feedback contínuo possibilita ainda um ciclo de melhorias para os anfitriões, permitindo a adaptação às expectativas e necessidades dos hóspedes (Fradkin et al., 2021).

2.4.3. Propriedades

A literatura sobre os *hosts* no Airbnb identifica dois tipos principais de anfitriões: 'profissionais' e 'não profissionais'. Os anfitriões profissionais tendem a gerir múltiplos

anúncios, que podem incluir propriedades próprias e de terceiros. Os anfitriões não profissionais compartilham espaços extras nas suas casas, muitas vezes motivados pela interação social e pelo desejo de obter um maior poder económico. (Abrate et al., 2022).

O estudo de Farmaki & Kaniadakis (2020) apresenta uma análise detalhada dos tipos de anfitriões no Airbnb, identificando quatro categorias principais: profissionais emergentes, anfitriões individuais com motivação econômica, e anfitriões coabitantes, com orientações econômicas ou sociais. Esta categorização reflete uma diversidade na forma como os anfitriões abordam a hospedagem no Airbnb, variando desde aqueles que gerem múltiplas propriedades com uma perspectiva de negócios, até aqueles que compartilham o seu espaço extra por razões sociais ou económicas.

Na sequência da análise da literatura, é apresentada o Anexo 1 que condensa e sintetiza os diversos tipos de anfitriões do Airbnb identificados em pesquisas acadêmicas ao longo dos anos. Como analisado, as investigações atuais tendem a concentrar-se predominantemente nos hóspedes e nas dinâmicas da sua interação com os anfitriões, sendo que a maioria dos estudos se baseia exclusivamente nas avaliações das propriedades disponibilizadas em plataformas como o Airbnb (Chen et al., 2024; Cheng & Jin, 2019; Guttentag et al., 2018; Julião et al., 2023).

A literatura existente limita-se frequentemente a distinguir os anfitriões em categorias binárias (Xie et al., 2021) sem uma análise aprofundada que contemple a diversidade e a complexidade dos diferentes tipos de anfitriões presentes na referida plataforma. Observa-se uma notória insuficiência na literatura no que concerne à exploração de outras dimensões que caracterizam os anfitriões do Airbnb, estendendo-se para além das avaliações online (Qiu et al., 2023).

2.5. Clustering multinível e clustering uninível

"O *clustering* (ou agrupamento) é uma técnica essencial em análise de dados, usada para dividir grandes conjuntos de dados em grupos homogêneos, com o objetivo de descobrir padrões e agrupamentos naturais. Ele é amplamente utilizado em diversas áreas, desde biologia e medicina até marketing e análise de redes sociais, permitindo identificar semelhanças e diferenças em grandes volumes de dados (Hareesha et al., 2023). No contexto de estudos de mercado, particularmente em plataformas de economia de partilha como o Airbnb, o *clustering* pode oferecer insights valiosos sobre padrões de comportamento de clientes e anfitriões (Guttentag, D., 2019). Tradicionalmente, o *clustering* é realizado de forma uninível, ou seja, sem considerar a existência de hierarquias ou níveis nos dados. Esta

abordagem consiste em agrupar diferentes tipos de variáveis, ignorando a complexidade inerente às relações entre diferentes camadas (Oyewole & Thopil, 2023).

Embora o *clustering* uninível apresente a vantagem de simplicidade e facilidade de implementação, ele apresenta também limitações significativas, especialmente quando aplicado a dados que naturalmente organizam-se em estruturas hierárquicas. O *k-means*, por exemplo, exige que o número de clusters seja definido antecipadamente e agrupa os dados com base na distância entre pontos, o que pode levar a resultados insatisfatórios em contextos onde os dados não possuem uma estrutura uniforme (Ghosh & Kumar, 2013). Oyewole & Thopil (2023) afirmam que os métodos tradicionais de clustering (ex., k-means) ou clustering baseado em densidade (ex., DBSCAN), são exemplos clássicos de clustering uninível, onde os dados são agrupados num único nível sem considerar hierarquias intrínsecas ou múltiplas camadas de dados. Ao não ter em conta as inter-relações entre diferentes níveis de dados, esta abordagem pode resultar na perda de nuances importantes e numa compreensão superficial dos padrões observados (Holodinsky et al., 2020).

Em contraste, o *clustering* multinível, proposto nesta tese, adota uma abordagem mais robusta ao reconhecer e incorporar a hierarquia dos dados. Estudos como os apresentados por Lebedev e Sukhoparov (2023) mostram que a aplicação destes modelos pode melhorar significativamente os indicadores de qualidade em análises de dados, especialmente em ambientes que requerem processamento de grandes volumes de dados. Este tipo de abordagem permite uma melhor adaptação às variações

A principal vantagem do *clustering* multinível reside na sua capacidade de capturar e representar relações complexas entre diferentes níveis de dados. Uma vez que, oferece uma visão mais holística e integrada, possibilitando a identificação de dependências e interações que não seriam evidentes numa análise uninível (Ren et al., 2022). Isto é particularmente relevante em contextos onde os dados apresentam uma estrutura em nível natural, como é o caso de plataformas de partilha de alojamento. Além disso, a abordagem multinível facilita a flexibilidade e a personalização, permitindo definir os clusters específicos em cada nível, adaptando a análise às necessidades particulares de negócios ou de pesquisa. No contexto empresarial, isso pode traduzir-se em estratégias de marketing mais direcionadas e eficazes, uma vez que as empresas podem identificar com precisão quais atributos são mais valorizados pelos diferentes segmentos de clientes. (Chatterjee & Pasquier, 2020)

No entanto, o *clustering* multinível não está isento de desafios. A sua implementação requer uma maior capacidade computacional e metodológica, dado que envolve a análise

simultânea de múltiplos níveis de dados (Yang & Tan, 2024). Estas dificuldades podem tornar o processo mais demorado e complexo, exigindo técnicas sofisticadas de modelação e análise de dados. Além disso, os resultados obtidos podem ser mais difíceis de interpretar e comunicar, dada a complexidade adicional introduzida pela consideração de múltiplos níveis.

Em suma, a escolha entre *clustering* uninível e multinível deve ser guiada pela natureza dos dados e pelos objetivos específicos da análise. Enquanto o *clustering* uninível oferece simplicidade e rapidez, podendo ser adequado para análises iniciais ou menos complexas, o *clustering* multinível proporciona uma profundidade de análise necessária para compreender sistemas complexos e hierárquicos. No contexto do Airbnb, onde as interações entre avaliações, propriedades e anfitriões desempenham um papel crucial, o *clustering* multinível pode revelar padrões e insights valiosos que, de outra forma, permaneceriam ocultos. Dessa forma, o uso de *clustering* multinível enriquece a análise académica e fornece uma base mais sólida para a tomada de decisões estratégicas no mercado de hospitalidade e turismo.

3. Metodologia

Esta secção faz uma síntese do problema e da questão de investigação, sendo de seguida apresentado o desenho da investigação, com um enquadramento ao contexto em que a mesma ocorre. É ainda desdobrada a abordagem metodológica para que se atinjam os objetivos propostos e, por fim, indicados os critérios de análise de dados para dar corpo ao estudo empírico.

3.1. Problema e questão de investigação

A crescente atenção académica aos tipos de anfitriões no Airbnb tem sido uma tendência marcante na literatura recente. Muitos estudos (Farmaki & Kaniadakis, 2020; Abrate et al., 2022; Tussyadiah, 2016; Guttentag et al., 2018) têm-se concentrado em identificar e categorizar diferentes perfis de anfitriões, distinguindo entre anfitriões profissionais e não profissionais, e os seus respetivos comportamentos na plataforma. Estes estudos exploram fatores como a motivação económica, o número de propriedades geridas, e o grau de envolvimento pessoal, destacando como diferentes tipos de anfitriões afetam a gestão das propriedades e a experiência dos hóspedes, através destas análises com foco na segmentação dos perfis dos anfitriões é possível obtermos uma visão mais clara das dinâmicas de mercado no Airbnb, permitindo compreender como cada perfil pode contribuir para a diversidade do serviço oferecido

No entanto, apesar desta valiosa contribuição, verifica-se uma lacuna significativa na literatura quanto à exploração de outras facetas que caracterizam os anfitriões do Airbnb. Sendo a interdependência entre anfitriões, hóspedes e a plataforma uma característica central dos alojamentos P2P torna-se necessário adotar uma abordagem metodológica mais abrangente (S. Ye et al., 2023). Isto porque, na economia de partilha, as preferências e comportamentos não ocorrem em isolamento, mas sim num ambiente onde as ações dos anfitriões e hóspedes afetam diretamente a experiência de ambas as partes, o que reforça a ideia de interdependência. (Proserpio et al., 2018) Este conceito é suportado por modelos de reciprocidade e confiança, que indicam que os resultados em mercados P2P são mais interdependentes e menos previsíveis do que em mercados tradicionais, devido à natureza colaborativa e informal da plataforma (Proserpio et al., 2018). Esta abordagem deve ir além da análise isolada de variáveis, reconhecendo as ligações entre os diferentes níveis de dados (Zal et al., 2024). Apenas desta forma será possível capturar uma imagem mais completa e multidimensional, refletindo de maneira mais precisa o impacto conjunto dessas variáveis na

experiência de hospedagem (Qiu et al., 2023).

Como mencionado, a análise das variáveis pertencentes aos alojamentos P2P, como as avaliações dos hóspedes e as características do alojamento ou anfitrião, não deve ser feita isoladamente, uma vez que atuam de forma conjunta, interagindo e influenciando-se mutuamente (Zamani et al., 2019). Assim, as metodologias multinível surgem como as mais adequadas para captar esta complexidade, uma vez que permitem tratar dados organizados em diferentes níveis, capturando as relações entre as avaliações dos hóspedes, as características das propriedades e o comportamento dos anfitriões de maneira mais precisa e abrangente (Zal et al., 2024).

Assim, o objetivo principal desta investigação é explorar e compreender de forma aprofundada qual das metodologias — *clustering* multinível ou uninível — é mais eficaz na segmentação de mercado dos anfitriões da plataforma Airbnb. Através da comparação entre os dois modelos, pretende-se identificar qual o método que consegue capturar de forma mais precisa os padrões distintos e os fatores-chave que definem os diferentes segmentos de mercado entre os anfitriões do Airbnb, fornecendo uma base sólida para a tomada de decisões estratégicas na plataforma.

Consequentemente, esta investigação pretende explorar as seguintes questões de investigação:

Q1: Qual das metodologias (*clustering* multinível ou *clustering* uninível) é mais eficaz na segmentação de mercado dos anfitriões?

Q2: Quais os diferentes tipos de anfitriões que existem dentro da plataforma Airbnb?

3.2. Modelo conceptual e hipóteses de investigação

A crescente importância das plataformas de economia partilhada, como o Airbnb, tem vindo a suscitar um elevado interesse académico na análise das características e motivações dos seus anfitriões. Segundo Hamari et al. (2016), as motivações dos anfitriões variam desde o desejo de gerar uma fonte adicional de rendimento até à procura por interações sociais e culturais. De forma semelhante, Farmaki & Kaniadakis (2020) sublinham que a segmentação dos anfitriões em perfis distintos, como profissionais e não profissionais, tem sido amplamente discutida, destacando diferentes comportamentos e estratégias no mercado consoante o grau de profissionalização.

No entanto, apesar dos avanços, Qiu et al. (2023) identificam uma lacuna importante na literatura existente. Estes autores defendem a necessidade de adotar abordagens analíticas que avaliem os anfitriões para além das variáveis isoladas, como as características das

propriedades ou as avaliações dos hóspedes, e que integrem uma visão holística, incorporando outros atributos que reflitam de forma mais completa as suas dinâmicas. Guttentag et al. (2017) reforçam esta perspectiva ao sublinhar que a escolha de uma estadia pelos hóspedes é motivada por uma interação complexa entre vários fatores, incluindo a qualidade da interação com o anfitrião, a propriedade em si e a experiência vivida.

Gerwe & Silva (2018) apontam para a complexidade inerente às interações no contexto da economia partilhada, sugerindo que a utilização de modelos analíticos mais sofisticados é essencial para capturar a profundidade dessas interações. Neste sentido, Jiang et al. (2024) defendem que a análise das variáveis relacionadas com anfitriões, propriedades e avaliações dos hóspedes deve ser feita de forma interdependente, considerando as interações complexas que influenciam a experiência global dos utilizadores. Assim, torna-se evidente que uma análise fragmentada destas variáveis não oferece uma compreensão completa da dinâmica entre anfitrião e hóspede.

Face a esta revisão, torna-se claro que existe uma lacuna significativa na literatura que justifica a adoção de uma abordagem mais abrangente e interligada na análise dos anfitriões, das suas propriedades e das avaliações dos hóspedes. Posto isto, propõem-se a seguinte hipótese de investigação para o presente estudo, de forma a testar uma abordagem que considere as variáveis como dependentes:

H1: Uma abordagem multinível apresenta uma segmentação mais eficaz dos anfitriões, ao capturar de forma mais detalhada as interações complexas entre diferentes níveis de dados (avaliações, características das propriedades e informações dos anfitriões), em comparação com abordagens uninível.

Estudos como o de Lebedev e Sukhoparov (2023) destacam a capacidade dos modelos multinível para lidar com dados hierárquicos e capturar as interações subtis entre variáveis. Estes modelos são particularmente adequados para contextos complexos, como o Airbnb, onde diferentes níveis de dados — anfitriões, propriedades e avaliações — estão interligados. Ao permitir uma análise mais profunda das interações entre estas variáveis, a análise multi-nível proporciona uma segmentação mais completa e detalhada, pois é particularmente eficaz em contextos onde a complexidade dos dados requer uma abordagem hierárquica. No contexto do Airbnb, onde as características das propriedades e das avaliações estão interligadas, a análise multinível oferece uma segmentação mais precisa e ajustada às nuances dos dados (Ren et al., 2022). No entanto, devido à diferença na complexidade entre as abordagens, espera-se que a correspondência entre os clusters seja reduzida. De acordo

com Hareesha et al. (2023) e Oyewole & Thopil (2023), uma abordagem uninível, por ignorar as hierarquias dos dados, resulta em agrupamentos menos precisos quando comparado a abordagens multinível. Além disto, em dados com estrutura hierárquica, as metodologias multinível tendem a capturar melhor as interações entre variáveis, resultando em clusters mais diferenciados (Chatterjee & Pasquier, 2020). Dado isto foram adicionalmente criadas as seguintes hipóteses:

H1.1: Uma abordagem multinível apresenta uma melhor qualidade de segmentação em comparação com abordagens uninível;

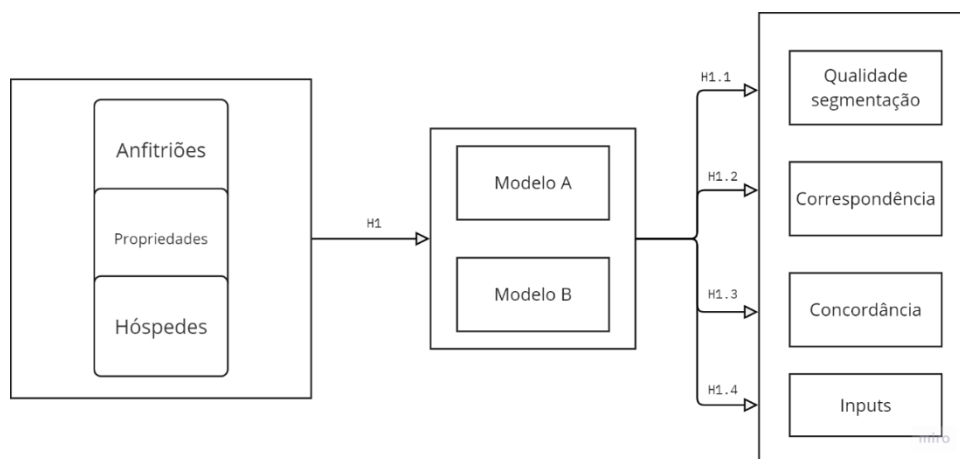
H1.2: Uma abordagem multinível não apresenta correspondência com abordagens uninível;

H1.3: Uma abordagem multinível não apresenta concordância com abordagens uninível;

H1.4: Uma abordagem multinível, devido à sua complexidade, proporcionará insights mais ricos e será mais eficaz na identificação de padrões em comparação com abordagens uninível.

Depois das hipóteses devidamente formuladas e justificadas através da revisão de literatura e respectivas lacunas, estas são representadas em conjunto, de uma forma mais simplificada, no modelo conceptual:

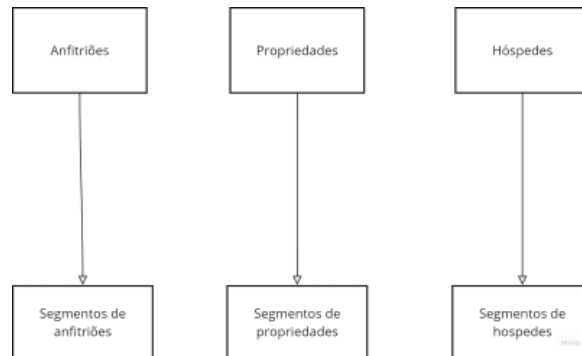
Figura 2- Modelo Conceptual



O Modelo A centra-se em analisar as variáveis de forma agregada ao nível do anfitrião, desconsiderando a hierarquia inerente aos dados. Neste modelo, as avaliações dos hóspedes, características das propriedades e atributos do anfitrião são consolidados num único nível de análise, permitindo uma avaliação simplificada. Por exemplo, ao invés de examinar como cada propriedade individualmente afeta as avaliações, este modelo resume as informações, tratando-as como representativas do anfitrião de forma geral. Embora esta abordagem facilite a quantificação do impacto de cada variável no perfil do anfitrião e seja útil para identificar tendências gerais, ela ignora as interações complexas entre diferentes

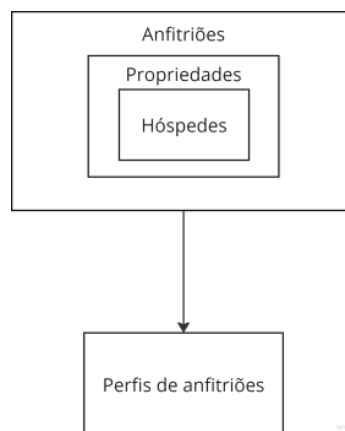
níveis de dados, como a influência mútua entre características específicas das propriedades e as avaliações recebidas.

Figura 3- Modelo A



O Modelo B, por sua vez, adota uma abordagem multinível, explorando a interdependência entre as variáveis para compreender como estas, em conjunto, influenciam a definição do perfil ideal de anfitrião. A força deste modelo está na sua capacidade de captar as inter-relações complexas entre os diferentes níveis de dados, como as avaliações, as características das propriedades e os atributos do anfitrião, que podem passar despercebidas em análises mais simples. Essa abordagem proporciona uma visão mais detalhada e abrangente. O Modelo B introduz uma inovação metodológica ao empregar uma segmentação multinível, superando a limitação de estudos anteriores

Figura 4- Modelo B



3.3. Contexto da Investigação

Para contextualizar a lacuna identificada na literatura e responder às questões de investigação propostas, esta dissertação concentra-se no setor do turismo, com especial foco

no fenómeno dos alojamentos peer-to-peer (P2P) e na diversidade de perfis de anfitriões na plataforma Airbnb. O objetivo é realizar uma análise aprofundada dos diferentes tipos de anfitriões, investigando as suas características e práticas de gestão, de forma a identificar padrões e perfis distintos. Esta investigação visa contribuir para uma melhor compreensão das dinâmicas que moldam o papel dos anfitriões no ecossistema do Airbnb.

O setor do turismo é uma das áreas de maior relevância económica para Portugal, contribuindo significativamente para o PIB nacional. (INE, 2023) Nos últimos anos, Portugal tem-se consolidado como um dos destinos turísticos mais procurados na Europa, atraindo milhões de visitantes anualmente. (INE, 2023). (Pang et al., 2024) Neste contexto, a figura do anfitrião no Airbnb ganha destaque, não só como fornecedor de alojamento, mas também como um elemento-chave na criação de experiências de viagem personalizadas e memoráveis (Sthapit & Jiménez-Barreto, 2018). No entanto, a crescente popularidade do Airbnb também levanta questões sobre a profissionalização dos anfitriões e as suas motivações (Farmaki & Kaniadakis, 2020).

Segundo os dados mais recentes, Portugal tem assistido a um aumento expressivo no número de anúncios no Airbnb, com um impacto significativo no mercado imobiliário local e na oferta de alojamento turístico (Inside Airbnb, 2024). Este crescimento tem gerado debates sobre os efeitos económicos e sociais deste fenómeno, especialmente no que diz respeito à gentrificação e à subida dos preços dos imóveis em áreas turísticas (Cunha & Lobão, 2022; Lopes et al., 2020)

Em resumo, o contexto desta investigação é moldado pelas transformações do turismo em Portugal e pelo papel emergente dos anfitriões do Airbnb como agentes económicos e culturais.

3.4. Metodologia de Investigação

A revisão da literatura revelou vários estudos que abordaram a identificação de perfis de anfitriões no Airbnb, alinhando-se com o objetivo desta investigação. Farmaki e Kaniadakis (2020), através de entrevistas semiestruturadas com anfitriões na Suécia, adotaram uma abordagem temática para identificar quatro tipos distintos de anfitriões, destacando as dinâmicas centrais nas interações entre anfitriões, hóspedes e a plataforma. Abrate et al. (2022), por outro lado, realizaram uma análise longitudinal dos dados de alojamentos em Milão e Roma, identificando anfitriões profissionais e não-profissionais com base numa análise de regressão de efeitos aleatórios. Este estudo investigou o impacto da variação de preços e do número de propriedades geridas no desempenho dos anfitriões.

Tussyadiah (2016) complementa este corpo de literatura com uma análise de descrições de anfitriões em Nova Iorque, utilizando uma abordagem de pré-processamento de texto e análise de clusters hierárquicos para explorar diferenças entre perfis, com base em dados de 12.785 alojamentos. Estes estudos, embora distintos nas suas metodologias, partilham o foco na identificação de perfis de anfitriões, fornecendo uma base valiosa para a presente investigação. Na Tabela 3 é possível observar de forma esquematizada as diferentes metodologias.

Tabela 2- Metodologias de estudos relacionados

Autores	Metodologias
Tussyadiah (2016)	Análise de descrições de anfitriões do Airbnb em Nova York com pré-processamento do texto, análise de cluster hierárquica, análises de variância (ANOVA), testes t e análises de regressão
Farmaki & Kaniadakis (2020)	Análise de entrevistas semiestruturadas com anfitriões
Chatterjee & Pasquier (2020)	Análise de registos de vendas que combina várias configurações algorítmicas de <i>clustering</i> para gerar uma solução de <i>clustering</i> de consenso em vários níveis
Abrate et al (2022)	Análise longitudinal dos dados de alojamentos do Airbnb em Milão e Roma com regressões de efeitos aleatórios

Assim para responder às questões de investigação, este estudo assume uma natureza exploratória, cuja metodologia assenta numa abordagem quantitativa. Os dados foram recolhidos através do site do Inside Airbnb e são obtidos a partir de informações publicamente disponíveis no site Airbnb. Foram analisados, limpos e agregados e disponibilizados para acesso público. A partir deste site foram retirados dois ficheiros com as variáveis a serem analisadas posteriormente. No caso da segmentação multinível, esta é mais utilizada em áreas como a saúde (Ren et al., 2022) e ciências da computação (Wang et al., 2024) devido à sua capacidade de lidar com complexidades e heterogeneidade dos dados.

No clustering uninível (Modelo A), não se tem em consideração a hierarquia existente nos dados. Este método tradicional ignora que os dados podem estar organizados em diferentes níveis. Em vez disso, foca-se apenas num único nível de agrupamento, que, neste caso, seria o nível dos anfitriões. No contexto dos anfitriões, cada um pode ter várias

propriedades, e cada propriedade pode ter várias avaliações feitas por hóspedes. Se tentarmos agrupar diretamente os anfitriões com base em todas as avaliações das suas propriedades, o resultado poderá ser distorcido, porque algumas propriedades podem ter mais avaliações do que outras, e avaliações repetidas poderiam enviesar os resultados. Para evitar essa distorção, a solução no clustering uninível é agregar os dados ao nível do anfitrião antes de fazer o agrupamento. Isto significa resumir ou condensar as avaliações de todas as propriedades associadas a um anfitrião, usando estatísticas como a média, a mediana ou o desvio padrão. Um anfitrião que tem três propriedades (uma com avaliações excelentes, outra com avaliações medianas, e outra com poucas avaliações) teria as suas avaliações agregadas num só valor por categoria, que representaria a média de todas as avaliações das suas propriedades. Assim, em vez de considerar cada avaliação individualmente, consideramos um único valor agregado para cada anfitrião. Depois de agregados os dados, o clustering pode ser feito com base nesses valores resumidos.

Por outro lado, o clustering multinível é uma abordagem mais sofisticada, que reconhece e utiliza a estrutura hierárquica dos dados. Em vez de colapsar os dados num único nível, esta abordagem explora os diferentes níveis de organização que existem nos dados. Existe um nível inferior, que representa as avaliações feitas por hóspedes; acima desse nível temos um nível intermédio, que é o nível das propriedades. Cada anfitrião pode ter várias propriedades, e as propriedades têm várias avaliações. E, finalmente, há o nível superior, que corresponde ao anfitrião. O clustering multinível permite obter conclusões em cada um destes níveis.

Esta abordagem multinível é particularmente útil quando se quer compreender como os padrões observados em níveis inferiores (por exemplo, as avaliações das propriedades) influenciam os padrões em níveis superiores (por exemplo, o sucesso geral do anfitrião). Tem em conta todas as informações dos níveis anteriores e não apenas as informações agregadas do nível do anfitrião.

3.4.1. Recolha e Organização dos Dados

Os dados foram obtidos a 2 de abril de 2024, onde foi transferido um ficheiro com as avaliações das propriedades na zona do Porto e outro ficheiro com as características dos anfitriões e das propriedades da mesma zona, estes dados são referentes até ao dia 18 de março de 2024. Considerando a extensão da amostra, 13600 entradas de propriedades, optou-se por seleccionar dez anfitriões, com base tanto no número de propriedades geridas quanto na sua natureza. Priorizou-se anfitriões que se apresentavam como pessoas

individuais ou casais com um número significativo de propriedades.

De forma, a analisar cada avaliação através de clusters foi necessário codificar o texto em números, assim, foram definidas 6 características:

- **Limpeza:** Avalia comentários dos hóspedes relacionados à limpeza da propriedade.
- **Comunicação:** Examina a interação entre hóspede e anfitrião, incluindo a resolução de problemas e sugestões.
- **Check-in:** Foca na experiência do hóspede com o processo de check-in e seu nível de conforto com o método adotado.
- **Localização:** Qualifica a localização da propriedade e se essa localização tem implicações na experiência do hospede.
- **Relação preço/qualidade:** Analisa comentários sobre o custo-benefício da propriedade e a disposição dos hóspedes para retornar.
- **Precisão:** Verifica a correspondência entre as expectativas criadas pelo anúncio e a realidade da propriedade.

Estas dimensões foram escolhidas em alinhamento com os critérios de avaliação usados pelo Airbnb, que requer que os hóspedes avaliem cada uma delas ao deixarem uma avaliação. Tal alinhamento assegura maior relevância e comparabilidade dos dados analisados.

Devido ao grande volume de avaliações, a análise das características mencionadas foi realizada por meio de um programa criado para análise com integração de uma API, baseada em algoritmos de machine learning e desenvolvida e escrita em Python.

O algoritmo em causa baseia-se na análise de sentimentos, que tem evoluído significativamente com o desenvolvimento de técnicas de Processamento de Linguagem Natural (NLP) (Carmona et al., 2022), especialmente com o surgimento de modelos de linguagem avançados como o ChatGPT. Tradicionalmente, a análise de sentimentos era realizada através de métodos manuais ou semiautomatizados, que dependiam de regras fixas e de frameworks predefinidos (Manickas & Shea, 1997). Estes métodos, embora oferecessem uma interpretação robusta dos dados, eram limitados em termos de escalabilidade, além de dependerem fortemente da subjetividade dos analistas, o que dificultava a replicação dos resultados e introduzia enviesamentos nas análises (Jeong & Lee, 2024).

No entanto, a introdução de modelos de linguagem como o ChatGPT trouxe uma verdadeira revolução na forma de realizar a análise de sentimentos. Uma vez que consegue analisar

grandes volumes de dados de maneira contextual e precisa, capturando nuances emocionais e variações de polaridade que os métodos tradicionais não eram capazes de identificar. Por exemplo, no estudo sobre análise de feedback em avaliações de hotéis (Jeong & Lee, 2024), o uso de ChatGPT facilitou a extração de sentimentos relacionados a falhas de serviço de forma automatizada e precisa, algo que anteriormente seria difícil e moroso com a codificação manual. A integração de ChatGPT com Python permite ainda a codificação direta dos sentimentos em valores numéricos, facilitando a análise quantitativa posterior.

Este modelo não só melhora a precisão, como é mais eficiente, permitindo que grandes volumes de texto, como avaliações de clientes, sejam processadas de forma escalável. (Jeong & Lee, 2024) Por isso, a utilização de ChatGPT em conjunto com Python para a codificação de sentimentos nas avaliações torna-se essencial nesta dissertação. O processo avaliativo executado pelo modelo, de acordo com o prompt enviado, começa com a identificação das categorias principais nas quais as *reviews* serão analisadas. Estas categorias incluem: Limpeza, Precisão, Check-in, Comunicação, Localização, e Relação Preço x Qualidade. A categoria de Limpeza envolve a avaliação das menções relativas à higiene do local, que se foca na percepção dos utilizadores sobre a limpeza do espaço. A categoria de Precisão avalia a correspondência entre as descrições apresentadas nos anúncios e a realidade percebida pelos utilizadores, incluindo a conformidade com as fotografias disponibilizadas. O Check-in refere-se à facilidade e eficácia do processo de entrada e acesso ao local, sendo um aspeto importante para a experiência do hóspede. A Comunicação é analisada com base na eficácia, prontidão e clareza da interação entre os anfitriões e os hóspedes. A categoria de Localização examina a conveniência do imóvel em termos de proximidade a pontos de interesse, comodidades e opções de transporte. Por fim, a Relação Preço x Qualidade avalia se o custo do local está em consonância com a qualidade e os benefícios percebidos pelos utilizadores.

Após a identificação das categorias, o algoritmo realiza uma leitura atenta do texto para identificar qualquer menção ou descrição que se enquadre nas categorias especificadas. Esta leitura minuciosa é essencial para assegurar que todas as nuances e detalhes relevantes são capturados, proporcionando uma análise abrangente do feedback dos hóspedes. O processo de extração de informações relevantes segue-se à leitura, onde, para cada categoria, o modelo extrai informações específicas que indiquem uma avaliação positiva, negativa ou neutra. Esta extração pode basear-se tanto em descrições explícitas feitas pelos utilizadores como em inferências implícitas. As informações extraídas são então classificadas de acordo

com uma escala numérica que varia de 0 a 5, de acordo com a escala do Airbnb no momento de avaliação. Esta classificação permite uma avaliação quantitativa das categorias mencionadas, onde uma nota de 0 indica que a categoria não foi mencionada no texto. Uma nota de 1 reflete uma avaliação muito negativa, com críticas severas ou problemas graves identificados, enquanto uma nota de 5 denota uma avaliação positiva, com elogios claros e uma expressão de satisfação por parte do hóspede.

Com base nestas classificações, o modelo atribui notas e comentários detalhados para cada categoria.

O resultado deste processo consiste na criação de um resumo estruturado, no qual cada categoria é listada com a nota correspondente e um comentário justificativo da avaliação. Este método sistemático e detalhado assegura uma análise consistente e objetiva dos textos, permitindo uma avaliação clara e fundamentada para cada categoria. A resposta do modelo é posteriormente processada para extrair as notas numéricas, utilizando expressões regulares na função `getNotes()`. Estas notas são garantidas para estar dentro do intervalo de 0 a 5, assegurando a coerência e a fiabilidade da avaliação.

Finalmente, os resultados das avaliações são registados e armazenados para posterior análise. As notas atribuídas são compiladas numa lista de dicionários, em que cada dicionário contém o ID do imóvel ou anfitrião e as respetivas notas para cada categoria. Estes resultados são então gravados num ficheiro CSV na pasta de *output*, através da função `writeCsv()`. O ficheiro CSV é subsequentemente convertido para o formato Excel, facilitando a sua manipulação e garantindo a consistência dos dados.

Foram obtidas e organizadas as informações relativas às avaliações, de seguida, através do ficheiro *listings.csv* foram retirados os dados relativos às propriedades e aos anfitriões. Nas propriedades foram selecionados os dados relativos ao tipo de propriedade, à capacidade e ao número de camas e nos anfitriões foram selecionados os dados relativos ao número de propriedades no Airbnb, se é *superhost* e a taxa de resposta. Através da organização destes dados com o respetivo anfitrião e propriedade previamente selecionados, juntamente com a avaliação, foi possível avançar para o *clustering* multinível e uninível.

3.4.2. Análise Multi-nível e Análise Uninível

A análise de dados nesta investigação foi conduzida utilizando duas abordagens distintas de clustering, denominadas Modelo A e Modelo B, implementadas através do software de programação R.

O Modelo A é baseado numa abordagem mais clássica, onde os dados são

segmentados de forma independente em três níveis distintos.

A primeira etapa foi a preparação dos dados, esta etapa é crucial e envolve a organização e limpeza dos dados para garantir que eles estejam corretos e utilizáveis. Foi usada a métrica Gower, própria para dados mistos (qualitativos e quantitativos). Quando se utiliza o método de Gower, que é uma métrica de dissimilaridade implementada na função `daisy` do pacote `cluster` no R, a normalização dos dados já é tratada automaticamente. O coeficiente de Gower é projetado para lidar com diferentes tipos de variáveis (numéricas, categóricas, binárias, etc.) e ele normaliza as variáveis numéricas para que todas contribuam de maneira equivalente à distância final, independentemente de suas magnitudes originais. Portanto, ao utilizar o coeficiente de Gower através do `daisy`, não é necessário normalizar explicitamente os dados. Após as matrizes de dissimilaridade, foi necessário determinar quantos grupos seriam ideais para segmentar os dados. Para isso, foi utilizado o `NbClust`, do pacote `NbClust` no R que é usado para determinar o número ideal de *clusters* com base em diferentes métodos e índices de validação. Quando se parametriza com o método `Ward.2` (uma variante do método de Ward) e o índice de silhueta, o processo funciona da seguinte forma:

- Método de `Ward.2`: Este método de agrupamento hierárquico minimiza a soma dos quadrados dentro dos clusters. A diferença entre o método de Ward tradicional e o `Ward.2` é apenas na forma como as distâncias são computadas entre clusters (usando uma versão alternativa da distância euclidiana quadrática).
- Índice de Silhueta: Avalia a qualidade dos clusters medindo o quão próximos os pontos dentro de um cluster estão uns dos outros, em comparação com os pontos de outros clusters. O valor da silhueta varia de -1 a 1, onde valores mais próximos de 1 indicam que os clusters são bem separados e compactos.

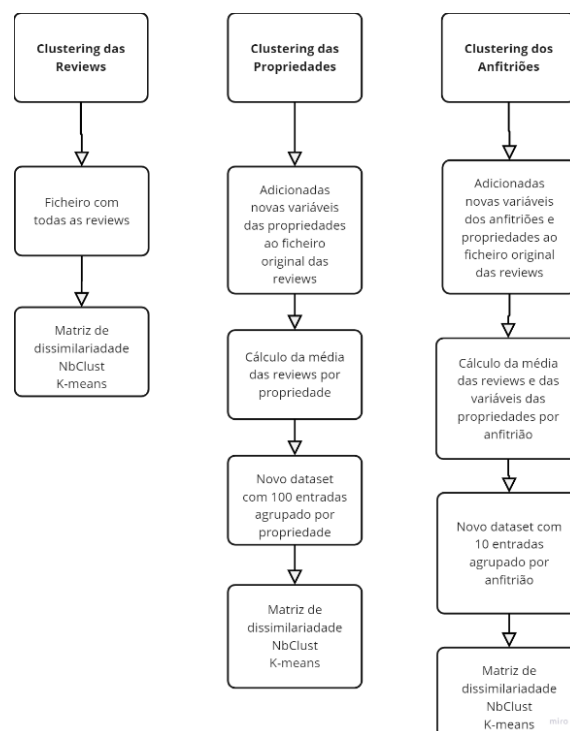
Com o número ideal de clusters identificado, é aplicado o algoritmo `k-means` para realizar o clustering. O `k-means` é uma técnica que distribui pontos de dados em `k` grupos, onde `k` é o número de clusters determinado anteriormente. O algoritmo trabalha de forma iterativa para minimizar a distância entre os pontos de dados e os centros dos clusters, garantindo que os dados dentro de um cluster sejam o mais semelhantes possível. Para este modelo, foram realizadas 3 análises individuais a 3 *datasets* diferentes, sendo que para as avaliações foi utilizando o ficheiro original com todas as 1000 entradas e, a partir daí foi realizada a análise de *clustering*. Para a análise das propriedades foram adicionadas 3 novas

variáveis ao ficheiro original – o tipo de alojamento, o número de pessoas que cada alojamento suportava e o número de camas que existiam no alojamento. De forma que a análise seja feita a nível da propriedade e como cada propriedade tem 10 *reviews* associadas e assim, obtivemos um novo dataset com 100 entradas sendo cada entrada uma propriedade. Por fim, ao ficheiro inicial com as *reviews* e com as variáveis das propriedades foram adicionadas 3 novas variáveis dos anfitriões – se o anfitrião é um *superhost*, quantas propriedades o anfitrião gere e a taxa de resposta do anfitrião aos hóspedes. A partir deste novo ficheiro é feita a análise de clusters resultando em 6 clusters.

Finalmente, é avaliada a qualidade dos clusters formados utilizando novamente o índice de silhueta. Se os clusters tivessem uma boa coesão interna (dados semelhantes dentro de um cluster) e uma boa separação externa (diferenças claras entre clusters), considerávamos que a segmentação estava correta. Os resultados foram então armazenados em arquivos CSV para facilitar a análise posterior e a visualização dos padrões encontrados em cada nível de dados.

De forma geral, a análise dos dados no Modelo A foi realizada da forma como está assinalada na Figura 5.

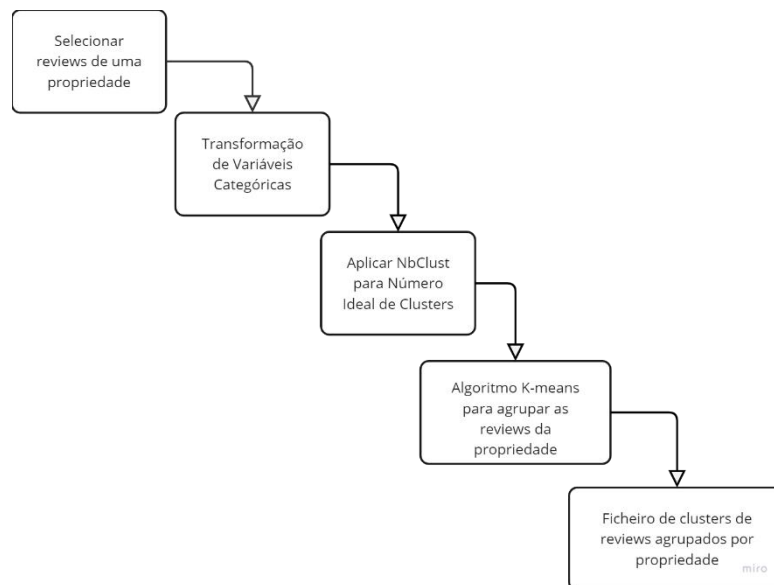
Figura 5 - Estrutura da análise de dados do Modelo A



O Modelo B, adota uma abordagem mais inovadora e complexa, denominada pela autora como *clustering* multinível. A primeira etapa consiste na utilização do ficheiro original

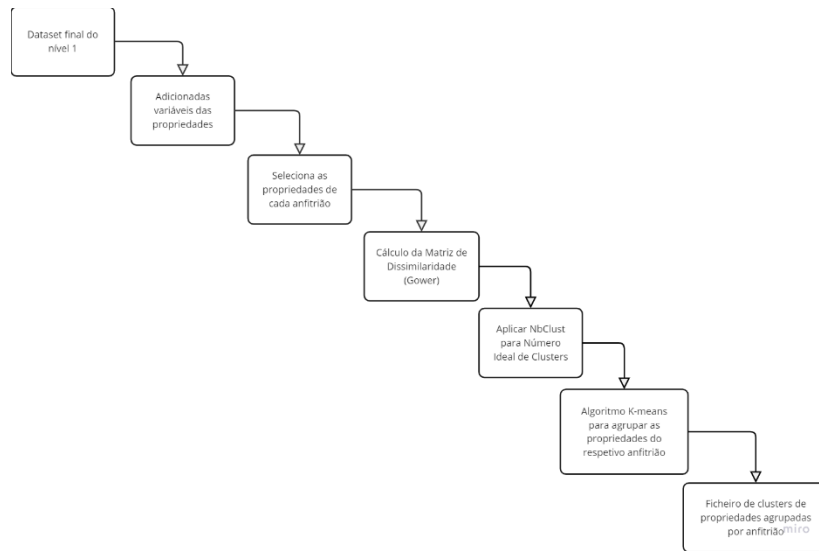
contendo as avaliações (*reviews*) dos hóspedes. O foco inicial é agrupar as avaliações de cada propriedade, considerando variáveis como limpeza, comunicação, check-in, localização, entre outras. Para cada nível deste modelo é aplicado o mesmo processo que o modelo A, através do R é realizada a transformação das variáveis categóricas com o método de Gower e também o algoritmo NbClust é aplicado, com o intuito de determinar o número ótimo de clusters. Em seguida, o algoritmo de k-means é utilizado para agrupar as avaliações, tendo em conta as semelhanças entre elas. O resultado desta etapa são 873 clusters, cada um representando um grupo de avaliações agrupadas por propriedade. Este nível permite uma análise detalhada da forma como as avaliações de cada propriedade são distribuídas, identificando padrões de satisfação e comportamentos dos hóspedes, a estrutura da análise do primeiro nível pode ser observada na Figura 6.

Figura 6- Estrutura da análise de dados ao nível das reviews



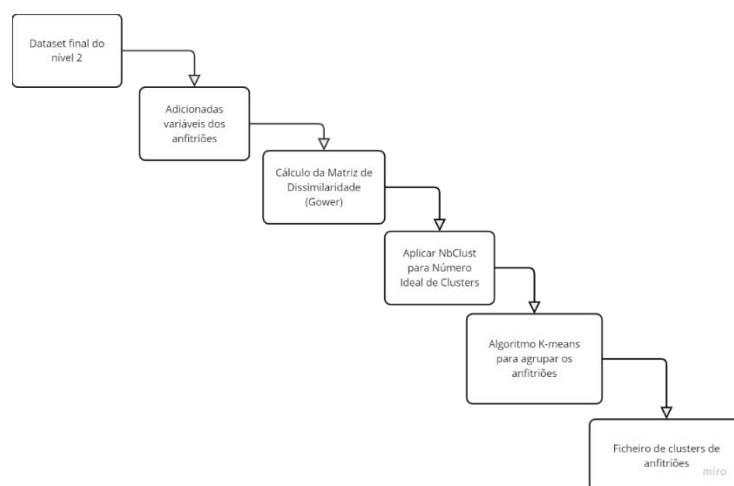
Na segunda etapa, são adicionadas ao conjunto de dados três variáveis adicionais relativas às propriedades: tipo de alojamento, capacidade da propriedade e número de camas, como observado na Figura 7. O objetivo aqui é realizar um novo processo de *clustering*, desta vez agrupando as propriedades por anfitrião. A partir do *dataset* gerado no primeiro nível, onde as propriedades já foram agrupadas pelas suas avaliações. Nesta fase, são obtidos 28 clusters de propriedades organizadas por anfitrião, o que permite identificar padrões na forma como diferentes anfitriões gerem as suas propriedades, com base nas características das mesmas e nas avaliações recebidas.

Figura 7- Estrutura da análise de dados ao nível das propriedades



Por fim, o terceiro nível integra as informações obtidas nas fases anteriores e adiciona mais três variáveis referentes aos anfitriões: se o anfitrião é *superhost*, número de propriedades que o anfitrião gere e taxa de resposta aos hóspedes. Com estas novas variáveis, é realizado o clustering final, agora a nível dos anfitriões. Assim, utiliza-se novamente todo o processo de clustering para agrupar os anfitriões com base nas propriedades que gerem e nas características individuais de cada um. O resultado desta última etapa é a obtenção de 5 clusters de anfitriões, onde cada cluster representa um conjunto de anfitriões com perfis semelhantes, considerando tanto as características das suas propriedades como o seu desempenho enquanto gestores.

Figura 8- Estrutura da análise de dados ao nível dos anfitriões

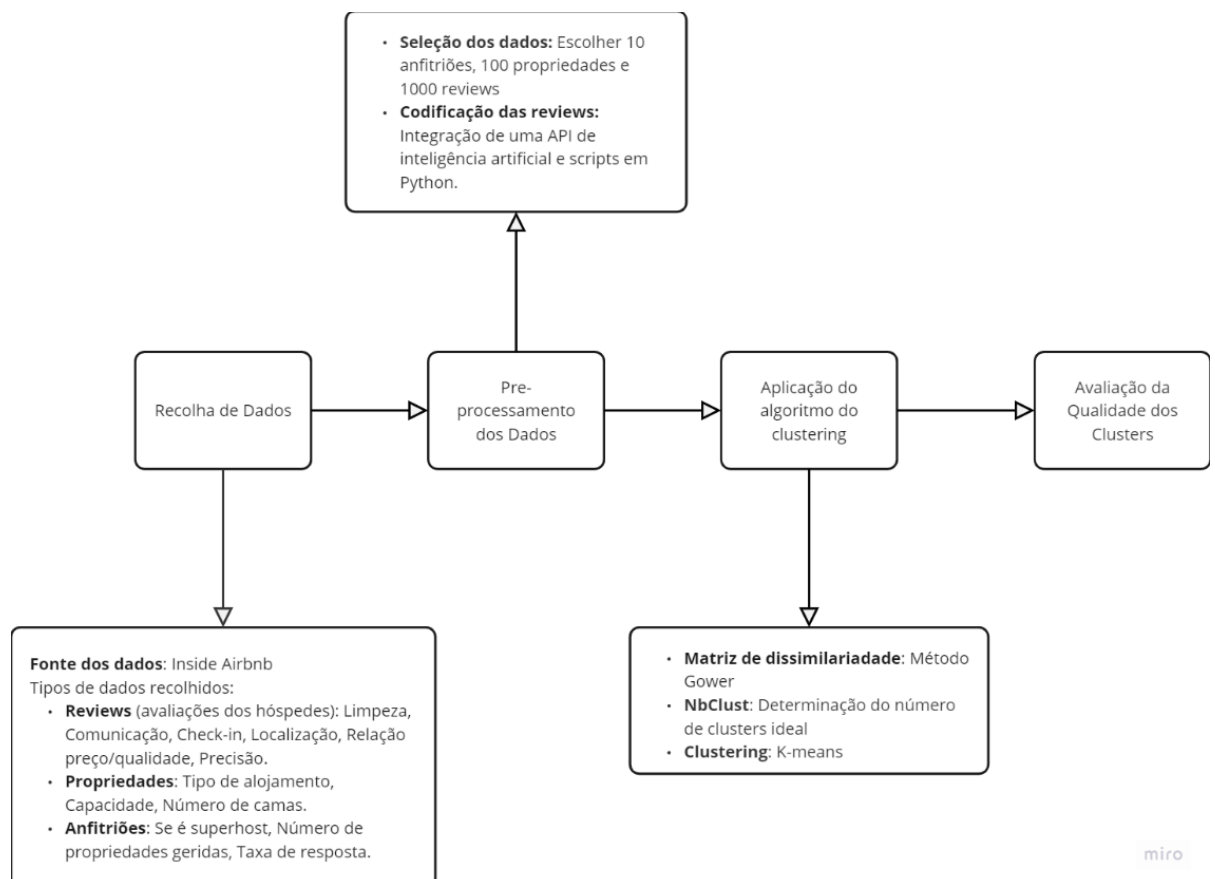


A escolha entre o Modelo A e o Modelo B deve ser guiada pelos objetivos específicos da análise. O Modelo A, com sua simplicidade, é ideal para situações em que se deseja uma

visão clara e independente dos diferentes níveis de dados. Já o Modelo B, com sua complexidade e capacidade de integrar diferentes camadas de informações, é mais adequado para análises que exigem uma compreensão detalhada das interações entre variáveis. Ambas as abordagens, ao serem aplicadas no contexto do Airbnb, contribuem para uma melhor compreensão dos padrões de segmentação, oferecendo uma base sólida para a tomada de decisões estratégicas e para o desenvolvimento de estratégias de marketing mais direcionadas e eficazes.

Assim, o processo deste estudo assenta em 4 pilares principais, a recolha de dados, o pré-processamento dos dados, a análise de clusters com o método de gower, o NbClust e o algoritmo de K-means e, por fim, a avaliação da qualidade dos clusters que será abordado no próximo capítulo. De forma geral, este processo pode ser representado por este esquema:

Figura 9- Estrutura geral da análise dos dados



4. Análise e Discussão de Resultados

4.1. Análise descritiva dos clusters

Nesta secção, discutem-se os resultados obtidos através da análise de clusters aplicada ao Modelo A e ao Modelo B. Através da aplicação do algoritmo de *clustering*, foi possível agrupar os anfitriões em clusters distintos, que refletem diferentes níveis de desempenho e perfis de atuação.

A análise de clusters realizada no Modelo A permitiu identificar seis grupos distintos de anfitriões.

O primeiro cluster caracteriza-se por anfitriões com um número relativamente baixo de propriedades geridas (27), indicando um perfil menos profissional e com propriedades de apenas uma cama disponível. No entanto, apesar de não serem *superhosts* em grande parte dos casos, estes anfitriões conseguem obter pontuações elevadas em comunicação (4.76), o que sugere uma abordagem eficaz na interação com os hóspedes. Além disso, as propriedades neste cluster apresentam uma boa localização (4.52), um aspeto valorizado pelos hóspedes e que contribui positivamente para a experiência geral. A relação preço-qualidade também é bem classificada (4.41).

O segundo cluster inclui anfitriões mais experientes, muitos dos quais gerem um número significativamente maior de propriedades (em média 106), no entanto estas propriedades oferecem, em média, apenas uma cama. Estes anfitriões, com um perfil mais profissional, destacam-se pela elevada pontuação em limpeza (4.73), a localização é igualmente um ponto forte, com uma média de 4.86, sugerindo que estes anfitriões operam em áreas bastante atrativas e procuradas pelos turistas. As avaliações refletem também uma boa relação preço-qualidade (4.51), demonstrando que estes anfitriões conseguem alinhar bem as expectativas dos hóspedes com o valor que oferecem, o que é crucial para fidelizar clientes.

No terceiro cluster, observam-se anfitriões com um número moderado de propriedades (cerca de 93), e que a sua propriedade, em média, dispõe de 5 camas e acomodam 7 pessoas. Neste segmento alguns dos anfitriões são *superhosts* e apresentam uma boa performance na comunicação (4.45) e no check-in (4.36). Contudo, este grupo enfrenta desafios no que diz respeito à precisão (3.97), no entanto, a localização das propriedades neste cluster é altamente valorizada (4.86).

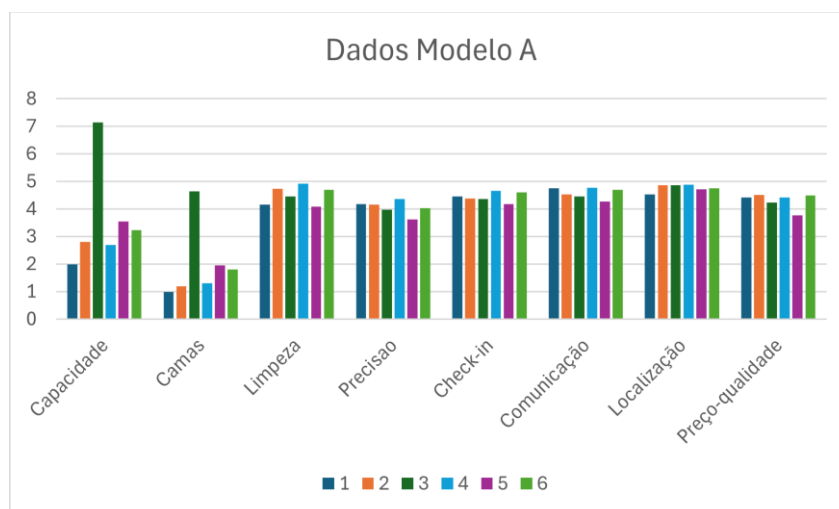
O quarto cluster destaca-se pela excelência em quase todas as dimensões analisadas em relação aos outros clusters. Os anfitriões deste grupo são *superhosts* e possuem um alto número de propriedades, demonstrando perfis profissionais que conseguem obter

pontuações muito altas na limpeza (4.92), check-in (4.66), e comunicação (4.77), o que sugere uma gestão muito eficiente e cuidada das interações com os hóspedes e da manutenção das propriedades. As suas propriedades estão também bem localizadas (4.88), além disso, a relação preço-qualidade (4.42) é elevada.

O quinto cluster inclui anfitriões que apresentam um número moderado de propriedades (58.5) e que acomodam em média 4 pessoas, dispondo de 2 camas. A precisão (3.61) e a relação preço-qualidade (3.77) deste segmento são as categorias com as pontuações mais baixas em relação aos outros clusters, sugerindo que estes anfitriões enfrentam dificuldades em alinhar as expectativas dos hóspedes com a realidade das suas propriedades. A comunicação (4.26) e limpeza (4.08) também estão abaixo da média, o que pode indicar que estes anfitriões não conseguem responder de forma tão eficaz às necessidades dos hóspedes. No entanto, as suas propriedades ainda têm uma localização razoável (4.72).

Por fim, o sexto cluster apresenta anfitriões com um menor número de propriedades (37), mas que são *superhosts*, as suas propriedades oferecem cerca de 2 camas e acomodam cerca de 3 pessoas. Este grupo de anfitriões possui também um desempenho moderado em várias dimensões, com algumas áreas a necessitar de melhorias. Estes anfitriões destacam-se pela elevada pontuação na limpeza (4.69) e *check-in* (4.60), sugerindo um forte compromisso com a qualidade da experiência do hóspede. A localização das propriedades (4.75) é também bastante boa, o que contribui para uma experiência global positiva. A relação preço-qualidade (4.49) é uma das mais altas entre os clusters, o que pode ser um fator importante para a retenção de clientes e a criação de uma base de hóspedes satisfeitos.

Figura 10 - Dados resultantes dos clusters do Modelo A



Assim, de forma geral, é possível compreender que o cluster 5, no que diz respeito às avaliações, apresenta-se abaixo da média, e o cluster 4 apresenta os melhores resultados na maior parte das avaliações. Excluindo o cluster 3, que representa propriedades de maior área, a maior parte das propriedades dos anfitriões dispõe de 1-2 camas e acomoda 2-4 pessoas. Além disso, os anfitriões mais experientes e profissionalizados (clusters 2 e 4) obtêm consistentemente melhores resultados em dimensões como limpeza e localização, os anfitriões com menos propriedades e experiência (clusters 1, 3 e 5) enfrentam desafios em áreas como precisão e relação preço-qualidade.

A análise de clusters no Modelo B adota uma abordagem multinível, o contrário do Modelo A, que foca na análise de cada variável de forma isolada, o Modelo B permite observar como é que estas variáveis combinam para definir diferentes perfis de anfitriões. Neste contexto, foram identificados cinco clusters distintos para o Modelo B, que representam diferentes tipos de anfitriões, desde os que gerem um elevado número de propriedades, até aos *superhosts* com menos propriedades.

O cluster 1 é composto por anfitriões que não possuem e que possuem o estatuto de *superhost* e que gerem em média 58,5 propriedades, tendo estas propriedades em média 2 camas e acomodam cerca de 4 pessoas. Estes anfitriões obtêm pontuações intermédias nas várias categorias, com a limpeza a atingir 3,12 e a precisão 2,73. No entanto, o desempenho ao nível da comunicação e da localização revela-se ligeiramente inferior, com uma média de 2,61 e 4,22, respetivamente. A relação preço-qualidade também é das mais baixas, com uma média de 1,79. Este cluster parece representar anfitriões com um volume considerável de propriedades, mas com avaliações ligeiramente abaixo da média, sobretudo no que diz respeito à relação custo-benefício.

Já o cluster 2 agrupa anfitriões que, em média, gerem 100 propriedades e também podem ser ou não *superhosts*. Estes anfitriões conseguem manter boas classificações na precisão (3,35), comunicação (3,15), e localização (4,47). Apesar da comunicação ser uma das pontuações mais altas, o check-in apresenta um dos resultados mais baixos (3,15). Este grupo de anfitriões acomoda cerca de 3 pessoas e apresenta em média 2 camas nas suas propriedades. Nos outros clusters as propriedades são sempre alugadas por completo, no entanto para este cluster existem propriedades que são alugadas por completo, mas também propriedades que são apenas um único quarto.

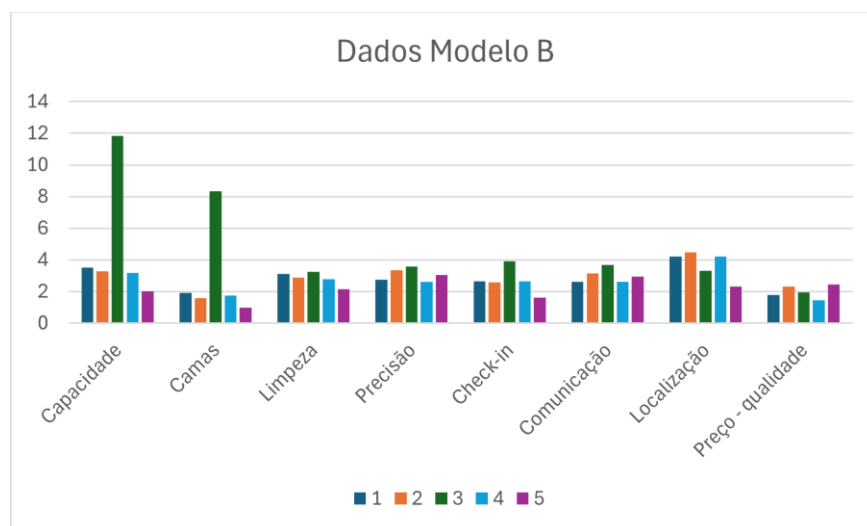
O cluster 3 distingue-se pelo número consideravelmente elevado de camas oferecidas (em média, 11,82 camas por propriedade) e pela elevada capacidade de alojamento. Estes

anfitriões, que são *superhosts*, gerem em média 94 propriedades. A classificação em precisão é relativamente elevada (3,56), assim como o check-in (3,91), a limpeza (3,25) e a comunicação (3,69), sugerindo um bom nível de organização operacional. No entanto, apesar da capacidade de acomodação e das boas avaliações nas outras características, a relação preço-qualidade é moderada (1,95), o que pode indicar que os hóspedes consideram o preço um pouco alto para o que é oferecido.

O cluster 4 compreende anfitriões que gerem cerca 37 propriedades e que também são *superhosts*. Este cluster apresenta capacidade média de alojamento (3,18) e um número de camas (1,76) relativamente parecido aos outros clusters. As avaliações para este grupo são equilibradas, destacando-se a limpeza (3,18) e o check-in (2,66). No entanto, a relação preço-qualidade é bastante inferior (1,45), como também a precisão (2,60) sugerindo que os hóspedes podem não estar totalmente satisfeitos com a experiência recebida nestas propriedades.

Por fim, o Cluster 5 representa anfitriões que, em média, gerem apenas 27 propriedades, sendo este grupo constituído por anfitriões sem o estatuto de *superhost*. Estes anfitriões recebem avaliações mais positivas na relação preço-qualidade (2,46) em relação aos outros segmentos, no entanto a limpeza, o check-in e a localização apresenta os dados mais baixos de todos os clusters.

Figura 11 - Dados resultantes dos clusters do Modelo B



4.2 Comparação Modelo A e B

Neste capítulo, procede-se à comparação crítica das duas metodologias de *clustering* aplicadas à segmentação dos anfitriões da plataforma Airbnb: o Modelo A, baseado numa abordagem uninível, e o Modelo B, que adota uma abordagem multinível, mais robusta e complexa. Esta análise visa avaliar a eficácia e a adequação de cada uma das metodologias, considerando as suas capacidades para captar as nuances comportamentais e operacionais dos anfitriões, em função das características das propriedades e dos anfitriões e das avaliações dos hóspedes.

Diversas métricas foram utilizadas para aferir a qualidade dos clusters gerados, incluindo o método de silhueta, o índice Davies-Bouldin, e uma análise de concordância entre os clusters das duas abordagens, através da construção de uma matriz de contingência. Além disso, a análise foi complementada com medidas avançadas de comparação entre agrupamentos, como o Adjusted Rand Index (ARI) e a Normalized Mutual Information (NMI), entre outras, para garantir uma avaliação exaustiva e académica do desempenho de cada modelo.

4.2.1. Indicadores de Qualidade dos Clusters

A qualidade intrínseca dos agrupamentos gerados por cada uma das abordagens foi aferida, em primeira instância, através das medidas: método de silhueta (Rousseeuw, 1987), que mede a semelhança de um ponto com o seu próprio cluster com outros clusters, e varia de -1 a 1, em que um valor próximo de 1 indica pontos bem agrupados, 0 indica agrupamentos sobrepostos e -1 indica pontos atribuídos ao agrupamento errado, e o índice Davies-Bouldin (Davies & Bouldin, 1979), é utilizado para avaliar a qualidade dos clusters, é definido como o rácio de semelhança médio de cada cluster com o seu cluster mais semelhante, em que valores mais baixos indicam um melhor agrupamento. Estes índices são amplamente reconhecidos pela literatura (Esther & Zielinska, 2023) como medidas sólidas para avaliar a coesão interna dos clusters e a sua separação face a outros grupos. A Tabela 4 sintetiza os resultados obtidos:

Tabela 3 - Avaliação da coesão dos clusters

Abordagem	Método de Silhueta	Índice Davies-Bouldin
Abordagem Multinível	0,176	0,830
Abordagem Uninível – Clientes	0,264	1,940
Abordagem Uninível – Propriedades	0,395	0,931
Abordagem Uninível – Anfitriões	0,388	0,278

A análise dos valores revela disparidades interessantes, a abordagem multinível, embora apresente um método de silhueta inferior (0,176) comparativamente às abordagens uninível, obtém um índice Davies-Bouldin de 0,830, o que sugere uma separação relativamente eficaz entre os clusters formados.

Estes resultados indicam que, embora a abordagem multinível consiga capturar uma maior complexidade estrutural dos dados, as abordagens uninível apresentam uma segmentação mais nítida, com maior homogeneidade intra-cluster e uma melhor separação entre os diferentes grupos de anfitriões.

4.2.2. Análise da Matriz de Confusão entre os clusters

Uma estratégia complementar para a comparação entre as duas abordagens foi a criação de uma matriz de confusão, também conhecida como matriz de contingência. Esta matriz permite examinar a correspondência entre os clusters formados pelo Modelo A (uninível) e o Modelo B (multinível), fornecendo uma visão mais profunda sobre o grau de sobreposição ou divergência entre os agrupamentos resultantes de cada metodologia. (Hastie et al., 2009)

A matriz foi construída através do mapeamento dos anfitriões em ambos os modelos, conforme ilustrado nos Anexos 5 e 6. A análise desta matriz revela que há pouca correspondência direta entre os clusters formados pelas duas abordagens, o que sugere que os dois modelos segmentam os anfitriões de maneiras significativamente diferentes. Os resultados da matriz de confusão indicam uma correspondência limitada entre os clusters do Modelo A e do Modelo B. Isto significa que os dois modelos estão a capturar diferentes aspetos dos anfitriões. Tal pode ser justificado pelo facto de o Modelo B adotar uma

abordagem multinível, o que lhe permite captar interações entre diferentes variáveis e níveis de informação (como as propriedades e os anfitriões), enquanto o Modelo A utiliza uma abordagem mais simples e direta.

4.2.3. Análise da concordância entre os clusters

Para enriquecer a análise, recorreremos a várias métricas avançadas de concordância entre os clusters, que foram calculadas para quantificar a semelhança entre os agrupamentos das duas abordagens. As métricas incluídas nesta análise são o Adjusted Rand Index (ARI), a Normalized Mutual Information (NMI), a Homogeneidade, a Completude e a V-measure. Estas medidas fornecem um panorama detalhado sobre o nível de sobreposição entre os clusters gerados pelos dois modelos, permitindo uma avaliação mais rigorosa da sua coerência e podem ser analisadas ao detalhe no Anexo 7.

Em relação, ao ARI (Hubert & Arabie, 1985), que mede a similaridade entre dois agrupamentos, variando entre -1 e 1, sendo que 1 indica uma concordância perfeita entre os clusters gerados pelos dois modelos e 0 significa que a concordância entre os clusters não é melhor do que o esperado ao acaso. Neste caso, o ARI é de 0.617 sugere que há uma correspondência moderada entre os clusters gerados pelo Modelo A e o Modelo B. Embora não haja uma correspondência perfeita, há uma sobreposição significativa, o que indica que ambos os modelos estão a capturar padrões semelhantes nos dados, mas com algumas variações.

Já em relação ao NMI (Estevez et al., 2009), que é uma medida de dependência entre dois agrupamentos, sendo que 1 significa que as duas partições são idênticas e 0 indica que as partições são completamente independentes, o valor apresentado foi de 0.853 o que indica uma elevada correspondência entre os agrupamentos gerados pelas duas abordagens, isto sugere que os clusters formados no Modelo A e no Modelo B partilham um elevado grau de informação comum, o que é um forte indicativo de que, apesar de utilizarem metodologias diferentes, ambos os modelos identificam padrões semelhantes nos dados.

A homogeneidade mede o grau em que os clusters contêm exemplos que pertencem à mesma classe (ou cluster original). De forma geral, um cluster é considerado homogêneo se todos os seus elementos estiverem corretamente agrupados no mesmo cluster em ambas as abordagens. Um valor de 0.805 indica que os clusters gerados por uma das abordagens (Modelo A ou Modelo B) têm um bom nível de consistência interna. No entanto, não é perfeito, o que sugere que existem algumas inconsistências na forma como os elementos são agrupados entre as duas abordagens.

Dentro destas análises, temos também a completude que avalia se todos os elementos de uma determinada classe estão contidos no mesmo cluster. Um valor de 0.907 indica que os clusters gerados têm uma elevada completude, ou seja, a maioria dos elementos que pertencem ao mesmo cluster no Modelo A também pertencem ao mesmo cluster no Modelo B. Este valor elevado sugere que, em grande parte, os clusters formados pelas duas abordagens são consistentes na forma como agrupam os elementos, com poucos elementos dispersos ou mal classificados entre os diferentes clusters.

Por fim, a V-measure (Rosenberg & Hirschberg, 2007) é a média entre a homogeneidade e a completude, proporcionando um equilíbrio entre estas duas métricas. Um valor de 0.853 indica que há uma boa correspondência entre os agrupamentos em termos de homogeneidade e completude. A V-measure confirma que, globalmente, as duas abordagens de clustering (Modelo A e Modelo B) produzem agrupamentos bastante semelhantes, capturando características semelhantes dos anfitriões e propriedades, com uma divisão consistente e bem estruturada.

4.2.4. Discussão de resultados

Após a validação das hipóteses do modelo estrutural através da análise dos clusters realizou-se a discussão dos respetivos resultados. No anexo 8 encontra-se uma tabela-síntese com a discussão de resultados.

Como analisado anteriormente, uma abordagem multinível é muito mais vantajosa para este tipo de estudo, devido às suas variáveis dependentes (Lebedev & Sukhoparov, 2024; Proserpio et al., 2018), por isso, os resultados deste estudo corroboram a validade das hipóteses propostas, demonstrando a eficácia de uma abordagem mais fragmentada e detalhada, representada pelo Modelo B (multinível), em comparação com abordagens isoladas, como o Modelo A (uninível).

A Hipótese H1.1, que afirmava que o Modelo B apresentaria uma melhor qualidade de segmentação em comparação com o Modelo A, não foi confirmada. Ao contrário do esperado, o Método de Silhueta revelou um valor mais elevado para a abordagem uninível, especialmente para as propriedades (0,395) e os anfitriões (0,388), enquanto a abordagem multinível registou um valor significativamente mais baixo (0,176). Isto sugere que a coesão interna dos clusters foi superior nas abordagens uninível, contrariamente à previsão inicial. Estes resultados desafiam a literatura que frequentemente argumenta a favor da eficácia superior dos modelos multinível na segmentação de dados complexos (Lebedev & Sukhoparov, 2023; Ren et al., 2022), sugerindo que, no contexto deste estudo, as abordagens

uninível foram mais eficazes na criação de clusters coesos. No entanto, estes valores podem refletir a simplicidade e a clareza dos dados quando analisados em níveis isolados, como propriedades ou *hosts*, em vez de capturar interações complexas entre diferentes variáveis. Por outro lado, o Índice Davies-Bouldin, que mede a separação entre clusters, apresentou resultados mais favoráveis para a abordagem uninível, particularmente para os *hosts* (0,278) e as propriedades (0,931), enquanto a abordagem multinível obteve um valor intermediário de 0,830. Isto reforça a ideia de que as abordagens uninível, ao analisar variáveis isoladas, permitiram uma melhor separação entre os grupos formados, oferecendo maior distinção entre clusters.

No que respeita à Hipótese H1.2, que previa uma baixa correspondência entre os clusters gerados pelos dois modelos, os resultados da matriz de confusão confirmam esta previsão. Apenas 45% dos clusters coincidiam entre os modelos, o que sugere que o Modelo A, ao tratar as variáveis de forma isolada, não foi capaz de capturar as interações entre as características dos anfitriões, as propriedades e as avaliações dos hóspedes. Estes dados complexos e hierárquicos exigem abordagens mais sofisticadas, como o Modelo B, para que se possa obter uma segmentação detalhada e precisa (Guttentag et al., 2017). A correspondência limitada entre os clusters confirma que o Modelo A não consegue identificar as nuances que são capturadas pela abordagem mais fragmentada do Modelo B.

Adicionalmente, a Hipótese H1.3, que previu uma concordância moderada entre os dois modelos, também foi validada. As métricas de concordância, nomeadamente o Adjusted Rand Index (ARI) com um valor de 0,53 e a Normalized Mutual Information (NMI) de 0,62, indicam que, embora os modelos apresentem algumas semelhanças nos clusters gerados, o Modelo B oferece uma segmentação mais detalhada e diferenciada. Estes resultados estão de acordo com os estudos de Chatterjee & Pasquier (2020), que destacam a capacidade das abordagens multinível para identificar padrões mais refinados nos dados, resultando numa segmentação mais eficaz.

Por fim, a Hipótese H1.4, que previa que uma abordagem mais fragmentada e multinível proporcionaria insights mais ricos e eficazes, foi confirmada através da análise dos resultados dos clusters. No Modelo A, a segmentação reflete a análise de variáveis de forma isolada, o que resulta em clusters com perfis mais simplificados, como segmentos de anfitriões menos profissionais ou mais profissionais e anfitriões que se destacam na maior parte das categorias com mais experiência e anfitriões que tem dificuldades em alinhar as expectativas dos hóspedes. Já no Modelo B é possível observar que dentro dos anfitriões que

gerem muitas propriedades existem aqueles que apresentam avaliações estáveis e outros que apresentam problemas em certas categorias, sendo que ambos são *superhosts*, e também segmentos de anfitriões em que apresentam poucas propriedades, mas são *superhost* e os segmentos de anfitriões em que apresentam poucas propriedades, mas não são *superhost*. Assim é possível observar em ambos os modelos anfitriões profissionais e não profissionais, mas no modelo B temos uma visão mais clara dos *superhosts*, nos dois modelos conseguimos tirar conclusões em relação ao volume das propriedades, mas no modelo B é mais claro que nem sempre o maior volume de propriedades ou a maior capacidade de alojamento resultam em melhores avaliações gerais. Além disto, no modelo B é possível observar uma maior complexidade na relação do preço com a qualidade e um impacto da capacidade do alojamento na percepção de qualidade. Mas o fator talvez mais importante visível apenas no modelo B é a interação complexa entre as variáveis (como número de propriedades, capacidade de alojamento e avaliações) que interagem entre si. Assim, estes resultados são consistentes com os achados de Oyewole & Thopil (2023), que indicam que abordagens multinível permitem uma análise mais rica e detalhada em contextos onde múltiplos níveis de variáveis estão interligados e, por isso, a H1 é válida.

5. Conclusão

Este último capítulo é dedicado ao resumo dos principais pontos abordados neste estudo com as considerações finais, a descrição dos contributos para a Academia e para a Gestão, as limitações encontradas no decorrer deste estudo e as sugestões para futuras Investigações.

5.1 Considerações finais

O objetivo desta dissertação foi realizar uma análise detalhada dos diferentes perfis de anfitriões na plataforma Airbnb, aplicando uma metodologia de segmentação multinível. Ao contrário das abordagens uninível, que tendem a analisar variáveis de forma isolada, a abordagem multinível permitiu capturar as interações complexas entre diferentes níveis de dados, como as avaliações dos hóspedes, as características das propriedades e dos anfitriões. Esta metodologia revelou-se crucial para uma compreensão mais completa das dinâmicas que influenciam o sucesso dos anfitriões na plataforma (Qiu et al., 2023; Zal et al., 2024).

As principais descobertas deste estudo foram: (1) o Modelo B (multinível) apresentou uma segmentação mais eficaz dos anfitriões do que o Modelo A (uninível), conforme esperado nas hipóteses formuladas (Hareesha et al., 2023); (2) as características dos anfitriões, como o número de propriedades geridas, a localização dos alojamentos e o envolvimento com os hóspedes, influenciam significativamente as avaliações e, conseqüentemente, o sucesso na plataforma (Abrate et al., 2022); (3) a complexidade do modelo multinível permitiu identificar interações subtis entre variáveis que o modelo uninível não captou (Chatterjee & Pasquier, 2020); (4) os perfis de anfitriões mais profissionais, que gerem várias propriedades, tendem a ter avaliações mais consistentes, enquanto anfitriões com uma única propriedade demonstram maior variabilidade nas suas avaliações (Farmaki & Kaniadakis, 2020).

A segmentação multinível, ao contrário da abordagem uninível convencional, possibilitou a detecção de interações mais complexas entre as variáveis, as quais poderiam passar despercebidas em análises mais simples. Esta metodologia permitiu, por exemplo, explorar de forma mais detalhada como as características dos anfitriões e das propriedades influenciam as avaliações dos hóspedes, gerando insights mais profundos sobre os fatores que contribuem para o desempenho dos anfitriões na plataforma. Com essa abordagem, foi possível obter uma compreensão mais apurada das dinâmicas que influenciam o sucesso, algo que as abordagens tradicionais não teriam captado com o mesmo nível de precisão.

Adicionalmente, este estudo representa uma das primeiras aplicações da segmentação multinível num conjunto de dados de plataformas peer-to-peer (P2P) como o Airbnb, abrindo caminho para novas formas de análise que capturam as interações dinâmicas entre anfitriões e hóspedes. Até agora, a maior parte das investigações sobre Airbnb utilizava abordagens uninível, que, embora úteis, limitam a capacidade de compreender como os diferentes níveis de dados influenciam mutuamente os resultados. Ao aplicar a metodologia multinível, este estudo mostrou que estas interações interdependentes podem ser não só observadas, mas também quantificadas e exploradas de forma mais precisa e rigorosa, o que marca uma evolução significativa na forma como as dinâmicas entre anfitriões e hóspedes são analisadas (Zal et al., 2024).

Em conclusão, esta dissertação contribui para a literatura sobre plataformas peer-to-peer (P2P), ao apresentar uma abordagem inovadora de segmentação multinível, destacando as interações entre anfitriões, propriedades e avaliações (Tussyadiah, 2016; Guttentag et al., 2018).

5.2. Contribuições teóricas e práticas

Esta dissertação contribui significativamente para o enriquecimento da teoria sobre a segmentação no contexto das plataformas peer-to-peer (P2P), como o Airbnb, ao aplicar e testar uma abordagem inovadora de segmentação multinível. Até agora, a literatura sobre a economia de partilha e plataformas digitais tem-se focado principalmente em abordagens uninível, que analisam variáveis de forma isolada, sem considerar as interações complexas entre os diferentes níveis de dados, como avaliações dos hóspedes, características das propriedades e atributos dos anfitriões. O presente estudo preenche essa lacuna ao aplicar uma metodologia inovadora de segmentação multinível, proporcionando uma visão mais completa e detalhada das dinâmicas que moldam o sucesso dos anfitriões (Abrate et al., 2022; Qiu et al., 2023).

Do ponto de vista teórico, este estudo marca uma inovação ao utilizar a segmentação multinível numa plataforma P2P pela primeira vez, oferecendo uma nova metodologia para compreender melhor as dinâmicas entre anfitriões e hóspedes. A segmentação multinível, ao contrário da tradicional abordagem uninível, permitiu identificar interações subtis entre variáveis que, de outra forma, não seriam capturadas. Por exemplo, ao analisar como as características dos anfitriões e das propriedades interagem com as avaliações dos hóspedes, esta abordagem demonstrou ser capaz de fornecer insights mais ricos e precisos sobre as dinâmicas que moldam o sucesso dos anfitriões na plataforma (Chatterjee & Pasquier, 2020).

Trabalhos prévios (Farmaki & Kaniadakis, 2020; Guttentag, 2019) tinham já salientado a necessidade de abordar estas interações de forma mais complexa, algo que este estudo vem concretizar de forma pioneira. O modelo teórico proposto poderá, assim, servir de base para futuras investigações que explorem a economia de partilha em maior profundidade

Outro contributo fundamental reside na aplicação desta metodologia a diferentes tipos de anfitriões. O estudo revelou que os anfitriões podem ser agrupados em diferentes segmentos com base nas suas práticas de gestão, incluindo anfitriões profissionais, que gerem múltiplas propriedades, e anfitriões individuais, focados em gerir uma única propriedade. Esta distinção, revelada pela abordagem multinível, ajuda a compreender as diferenças de comportamento e desempenho entre os anfitriões e como estas afetam as suas avaliações (Farmaki & Kaniadakis, 2020; Abrate et al., 2022). Estudos anteriores, como o de Tussyadiah (2016), já sugeriam a importância da segmentação, mas este trabalho demonstra como a metodologia multinível pode capturar de forma mais eficaz as complexidades envolvidas. Assim, este estudo amplia a compreensão da segmentação dos anfitriões e das estratégias de gestão adotadas.

Quanto às contribuições práticas, esta dissertação oferece insights valiosos tanto para os anfitriões do Airbnb quanto para os gestores de plataformas P2P. Os anfitriões podem utilizar os resultados para ajustar as suas práticas de gestão, dependendo do seu perfil, de forma a maximizar as suas receitas e melhorar as suas avaliações. Anfitriões que gerem várias propriedades, por exemplo, podem adotar uma abordagem mais padronizada e eficiente, enquanto anfitriões individuais podem focar-se em proporcionar experiências personalizadas para otimizar a satisfação dos hóspedes (Farmaki & Kaniadakis, 2020). Além disso, as plataformas podem utilizar a segmentação multinível para otimizar as suas estratégias de apoio, adaptando-se às necessidades específicas de diferentes perfis de anfitriões, melhorando a retenção e a performance geral da plataforma (Qiu et al., 2023) (Tussyadiah, 2016; Farmaki & Kaniadakis, 2020).

Por fim, a aplicação desta metodologia também tem implicações práticas para a gestão de plataformas digitais. Ao identificar perfis mais detalhados de anfitriões, o Airbnb pode implementar estratégias de personalização de serviços e criar programas de formação específicos para os diferentes tipos de anfitriões, promovendo uma experiência positiva para os hóspedes e, conseqüentemente, uma maior satisfação e lealdade dos utilizadores (Abrate et al., 2022). A segmentação multinível permite ainda uma melhor análise de como diferentes fatores, como o volume de propriedades e a interação com os hóspedes, influenciam

diretamente a reputação e o sucesso dos anfitriões, permitindo decisões mais estratégicas e informadas

Outra implicação prática importante está relacionada com o uso da segmentação multinível como uma ferramenta estratégica para a gestão de plataformas digitais. Esta abordagem permite que as plataformas reconheçam e respondam melhor às necessidades específicas de diferentes segmentos de anfitriões, promovendo a sua retenção e sucesso a longo prazo. Ao adotar uma abordagem personalizada com base nos perfis segmentados, o Airbnb e outras plataformas semelhantes podem aumentar a satisfação dos seus utilizadores e garantir uma maior competitividade no mercado da economia de partilha (Guttentag et al., 2018).

Em suma, esta dissertação não só contribui para o avanço teórico da segmentação em plataformas P2P, como também oferece implicações práticas valiosas para a gestão de anfitriões e plataformas. As descobertas apresentadas ampliam a compreensão das interações complexas entre os diferentes níveis de dados, abrindo novas perspetivas para o estudo da economia de partilha e fornecendo ferramentas práticas para melhorar a gestão e o desempenho dos anfitriões e das plataformas digitais.

5.3 Limitações e direções de pesquisa futura

Como em qualquer investigação, esta dissertação apresenta algumas limitações que devem ser reconhecidas. Em primeiro lugar, o tamanho e a composição da amostra podem limitar a generalização dos resultados. Os dados analisados foram recolhidos de uma amostra muito pequena do que é Airbnb, principalmente as avaliações, o que pode limitar muito os resultados obtidos. Além disso, embora os dados recolhidos da plataforma Inside Airbnb forneçam informações detalhadas sobre os anfitriões, as propriedades e as avaliações, as conclusões tiradas a partir desta amostra não podem ser generalizadas para o universo completo de anfitriões da plataforma, uma vez que as características regionais, culturais e demográficas podem influenciar os padrões de comportamento dos anfitriões e hóspedes (Qiu et al., 2023)

Outra limitação importante está relacionada com a complexidade inerente à metodologia multinível. Embora esta abordagem tenha permitido capturar interações mais detalhadas entre diferentes níveis de dados, como avaliações dos hóspedes e características das propriedades, a implementação desta metodologia exigiu um maior esforço computacional e tempo para processar os dados. Além disso, o uso de metodologias complexas como o clustering multinível pode gerar resultados difíceis de interpretar ou

comunicar, especialmente para audiências não especializadas (Chatterjee & Pasquier, 2020). Esta complexidade pode ter também afetado a clareza dos resultados, dado que algumas variáveis foram analisadas de forma interdependente, o que pode ter dificultado a interpretação exata do impacto de cada variável individual.

Uma limitação adicional é o uso de um programa desenvolvido especificamente para este estudo. Embora o programa em python tenha sido bem-sucedido na análise dos dados do Airbnb, ele não foi testado com outros tipos de dados ou em diferentes contextos. O facto de o programa ser personalizado para esta investigação significa que os resultados obtidos são, até certo ponto, dependentes das suas funcionalidades e capacidades. Sem testes adicionais com outros conjuntos de dados ou estudos, a aplicabilidade e eficácia da ferramenta em cenários diferentes ainda não foram confirmadas, o que limita o seu potencial uso em investigações futuras (Farmaki & Kaniadakis, 2020).

Além disso, a escolha entre as abordagens uninível e multinível reflete-se nas limitações dos modelos. O Modelo A (uninível), mais simples, permite uma análise mais clara e rápida das variáveis de forma isolada, mas perde nuances importantes ao não considerar as interações complexas entre os diferentes níveis de dados (Abrate et al., 2022). Já o Modelo B (multinível), embora mais detalhado, exigiu uma maior capacidade de processamento e apresentou uma maior complexidade na análise e comunicação dos resultados. Este modelo é mais adequado para captar a profundidade das interações entre os diferentes níveis, mas a sua implementação e interpretação podem ser desafiantes para os utilizadores da plataforma ou os investigadores que não têm familiaridade com este tipo de análise (Oyewole & Thopil, 2023)

As direções para pesquisas futuras podem focar-se em expandir o estudo para outras áreas geográficas e culturais, de forma a verificar se os resultados obtidos nesta investigação se mantêm em diferentes contextos. Seria também relevante realizar uma análise comparativa entre cidades com diferentes níveis de maturidade no uso da plataforma, de modo a identificar padrões que possam emergir em mercados mais estabelecidos versus mercados em crescimento. Além disso, futuras investigações poderiam incorporar variáveis adicionais que não foram exploradas neste estudo, como o preço por noite, a descrição em texto do anfitrião, ou as fotos das propriedades.

Por fim, uma outra linha de investigação poderia explorar a aplicação de abordagens metodológicas mistas, combinando dados quantitativos e qualitativos. A integração de métodos qualitativos, como entrevistas com anfitriões, poderia fornecer uma visão mais rica

sobre as motivações e desafios enfrentados pelos anfitriões, complementando os achados quantitativos. Além disso, futuras pesquisas poderiam explorar como diferentes perfis de anfitriões podem adaptar as suas estratégias em resposta às mudanças de mercado, como as novas regulamentações locais ou alterações nas preferências dos consumidores (Zal et al., 2024).

Em suma, esta dissertação abre novas perspectivas para o estudo da segmentação de anfitriões no Airbnb, mas também deixa espaço para investigações adicionais que possam expandir o conhecimento sobre este fenômeno em contextos mais amplos.

Referências Bibliográficas

- Abrate, G., Sainaghi, R., & Mauri, A. G. (2022). Dynamic pricing in Airbnb: Individual versus professional hosts. *Journal of Business Research*, 141, 191-199. <https://doi.org/https://doi.org/10.1016/j.jbusres.2021.12.012>
- Acquier, A., Daudigeos, T., & Pinkse, J. (2017). Promises and paradoxes of the sharing economy: An organizing framework. *Technological Forecasting and Social Change*, 125, 1-10. <https://doi.org/https://doi.org/10.1016/j.techfore.2017.07.006>
- Acs, Z. J., Song, A. K., Szerb, L., Audretsch, D. B., & Komlósi, É. (2021). The evolution of the global digital platform economy: 1971–2021. *Small Business Economics*, 57(4), 1629-1659. <https://doi.org/10.1007/s11187-021-00561-x>
- Belarmino, A., & Koh, Y. (2020). A critical review of research regarding peer-to-peer accommodations [Article]. *International Journal of Hospitality Management*, 84, Article 102315. <https://doi.org/10.1016/j.ijhm.2019.05.011>
- Belarmino, A., & Koh, Y. (2020). A critical review of research regarding peer-to-peer accommodations. *International Journal of Hospitality Management*, 84, 102315. <https://doi.org/https://doi.org/10.1016/j.ijhm.2019.05.011>
- Belk, R. (2007). Why Not Share Rather Than Own? *The ANNALS of the American Academy of Political and Social Science*, 611(1), 126-140. <https://doi.org/10.1177/0002716206298483>
- Belk, R. (2014). You are what you can access: Sharing and collaborative consumption online. *Journal of Business Research*, 67(8), 1595-1600. <https://doi.org/https://doi.org/10.1016/j.jbusres.2013.10.001>
- Boateng, H., Kosiba, J. P. B., & Okoe, A. F. (2019). Determinants of consumers' participation in the sharing economy: A social exchange perspective within an emerging economy context [Article]. *International Journal of Contemporary Hospitality Management*, 31(2), 718-733. <https://doi.org/10.1108/IJCHM-11-2017-0731>
- Carmona, M. A., Aranda, R., Rodríguez González, A., Fajardo-Delgado, D., Sánchez, M., Espinosa, H., Martínez-Miranda, J., Guerrero, R., Bustio-Martínez, L., & Díaz-Pacheco, A. (2022). Natural Language Processing applied to Tourism Research: A Systematic Review and Future Research Directions. *Journal of King Saud University - Computer and Information Sciences*, 34. <https://doi.org/10.1016/j.jksuci.2022.10.010>
- Chang, H. H., & Sokol, D. D. (2022). How incumbents respond to competition from innovative disruptors in the sharing economy—The impact of Airbnb on hotel performance [Article]. *Strategic Management Journal*, 43(3), 425-446. <https://doi.org/10.1002/smj.3201>

- Chatterjee, S., & Pasquier, N. (2020). A Multi-level Consensus Clustering Framework for Customer Choice Modelling in Travel Industry. In (pp. 142-157). Springer International Publishing. https://doi.org/10.1007/978-3-030-60036-5_10
- Chen, D., Zhang, W., Bi, J.-W., Qiu, H., & Lyu, J. (2024). Hosts' online affinities and their impacts on the number of online reviews on peer-to-peer platforms. *Tourism Management*, *100*, 104817. <https://doi.org/https://doi.org/10.1016/j.tourman.2023.104817>
- Chen, D., Zhang, W., Bi, J. W., Qiu, H., & Lyu, J. (2024). Hosts' online affinities and their impacts on the number of online reviews on peer-to-peer platforms [Article]. *Tourism Management*, *100*, Article 104817. <https://doi.org/10.1016/j.tourman.2023.104817>
- Cheng, M., Houge Mackenzie, S., & Degarege, G. A. (2022). Airbnb impacts on host communities in a tourism destination: an exploratory study of stakeholder perspectives in Queenstown, New Zealand. *Journal of Sustainable Tourism*, *30*(5), 1122-1140. <https://doi.org/10.1080/09669582.2020.1802469>
- Cheng, M., & Jin, X. (2019). What do Airbnb users care about? An analysis of online review comments [Article]. *International Journal of Hospitality Management*, *76*, 58-70. <https://doi.org/10.1016/j.ijhm.2018.04.004>
- Cunha, A. M., & Lobão, J. (2022). The effects of tourism on housing prices: applying a difference-in-differences methodology to the Portuguese market [Article]. *International Journal of Housing Markets and Analysis*, *15*(4), 762-779. <https://doi.org/10.1108/IJHMA-04-2021-0047>
- Davies, D., & Bouldin, D. (1979). A Cluster Separation Measure. *Pattern Analysis and Machine Intelligence, IEEE Transactions on, PAMI-1*, 224-227. <https://doi.org/10.1109/TPAMI.1979.4766909>
- Dolnicar, S. (2019). A review of research into paid online peer-to-peer accommodation: Launching the Annals of Tourism Research Curated Collection on peer-to-peer accommodation. *Annals of Tourism Research*, *75*, 248-264. <https://doi.org/https://doi.org/10.1016/j.annals.2019.02.003>
- Estevez, P., Tesmer, M., Perez, C., & Zurada, J. (2009). Normalized Mutual Information Feature Selection. *IEEE transactions on neural networks / a publication of the IEEE Neural Networks Council*, *20*, 189-201. <https://doi.org/10.1109/TNN.2008.2005601>
- Esther, E., & Zielinska, T. (2023). Comparative Analysis of the Clustering Quality in Self-Organizing Maps for Human Posture Classification. *Sensors*, *23*. <https://doi.org/10.3390/s23187925>
- Farmaki, A., & Kaniadakis, A. (2020). Power dynamics in peer-to-peer accommodation: Insights from Airbnb hosts. *International Journal of Hospitality Management*, *89*,

102571. <https://doi.org/https://doi.org/10.1016/j.ijhm.2020.102571>

- Felson, M., & Spaeth, J. L. (1978). Community Structure and Collaborative Consumption: A Routine Activity Approach. *American Behavioral Scientist*, 21(4), 614-624. <https://doi.org/10.1177/000276427802100411>
- Fradkin, A., Grewal, E., & Holtz, D. (2021). Reciprocity and Unveiling in Two-Sided Reputation Systems: Evidence from an Experiment on Airbnb. *Marketing Science*, 40(6), 1013-1029. <https://doi.org/10.1287/mksc.2021.1311>
- Garg, A. (2020). Factors Influencing Generation Y to Choose Airbnb. In (pp. 45-65). <https://doi.org/10.1108/978-1-83982-688-720201004>
- Gerwe, O., & Silva, R. (2018). Clarifying the Sharing Economy: Conceptualization, Typology, Antecedents, and Effects. *Academy of Management Perspectives*, 34. <https://doi.org/10.5465/amp.2017.0010>
- Ghosh, S., & Kumar, S. (2013). Comparative Analysis of K-Means and Fuzzy C-Means Algorithms. *International Journal of Advanced Computer Science and Applications*, 4(4). <https://doi.org/10.14569/ijacsa.2013.040406>
- Guttentag, D. (2019). Progress on Airbnb: a literature review. *Journal of Hospitality and Tourism Technology*, 10. <https://doi.org/10.1108/JHTT-08-2018-0075>
- Hall, C. M., Prayag, G., Safonov, A., Coles, T., Gössling, S., & Naderi Koupaie, S. (2022). Airbnb and the sharing economy [Editorial]. *Current Issues in Tourism*, 25(19), 3057-3067. <https://doi.org/10.1080/13683500.2022.2122418>
- Hamari, J., Sjöklint, M., & Ukkonen, A. (2016). The Sharing Economy: Why People Participate in Collaborative Consumption. *Journal of the Association for Information Science and Technology*, 67, 2047-2059. <https://doi.org/10.1002/asi.23552>
- Han, T.-Y., Bi, J.-W., Wei, Z.-H., & Yao, Y. (2024). Visual cues and consumer's booking intention in P2P accommodation: Exploring the role of social and emotional signals from hosts' profile photos. *Tourism Management*, 102, 104884. <https://doi.org/https://doi.org/10.1016/j.tourman.2024.104884>
- Hareesha, D., Ahamed, S. F., Brahmaiah, V. P., Kalpana, D., Arun, M., & Satyanarayana, P. (2023). A novel multi-level clustering mechanism using heuristic approach for secure data transmission in WSN sector with various trust computation [Article]. *Intelligent Decision Technologies*, 17(4), 1417-1433. <https://doi.org/10.3233/IDT-220314>
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition (Springer Series in Statistics)*.

- Holodinsky, J. K., Austin, P. C., & Williamson, T. S. (2020). An introduction to clustered data and multilevel analyses. *Family Practice*, 37(5), 719-722. <https://doi.org/10.1093/fampra/cmaa017>
- Hubert, L., & Arabie, P. (1985). Comparing partitions. *Journal of Classification*, 2(1), 193-218. <https://doi.org/10.1007/bf01908075>
- Jeong, N., & Lee, J. (2024). An Aspect-Based Review Analysis Using ChatGPT for the Exploration of Hotel Service Failures. *Sustainability*, 16(4), 1640. <https://doi.org/10.3390/su16041640>
- Julião, J., Gaspar, M., Farinha, L., & Trindade, M. A. M. (2023). Sharing economy in the new hospitality: consumer perspective [Article]. *Journal of Hospitality and Tourism Insights*, 6(4), 1447-1463. <https://doi.org/10.1108/JHTI-08-2021-0198>
- Khalek, S. A., & Chakraborty, A. (2023). Access or collaboration? A typology of sharing economy. *Technological Forecasting and Social Change*, 186, 122121. <https://doi.org/https://doi.org/10.1016/j.techfore.2022.122121>
- La, L., Xu, F., Hu, M., & Xiao, C. (2022). Location of Airbnb and hotels: the spatial distribution and relationships. *Tourism Review*, 77(1), 209-224. <https://doi.org/10.1108/TR-10-2020-0476>
- Lebedev, I. S., & Sukhoparov, M. E. (2024). Improving the Quality Indicators of Multilevel Data Sampling Processing Models Based on Unsupervised Clustering. *Emerging Science Journal*, 8(1), 355-371. <https://doi.org/10.28991/esj-2024-08-01-025>
- Lopes, M. C., Liberato, D., Alcn, E., & Liberato, P. (2020). Social tourism development and the population ageing: Case study in portugal and spain. *Smart Innovation, Systems and Technologies*,
- Manickas, P. A., & Shea, L. J. (1997). Hotel Complaint Behavior and Resolution: A Content Analysis. *Journal of Travel Research*, 36, 68 - 73.
- Meng, B., & Cui, M. (2020). The role of co-creation experience in forming tourists' revisit intention to home-based accommodation: Extending the theory of planned behavior. *Tourism Management Perspectives*, 33, 100581. <https://doi.org/https://doi.org/10.1016/j.tmp.2019.100581>
- Moon, H., Miao, L., Hanks, L., & Line, N. D. (2019). Peer-to-peer interactions: Perspectives of Airbnb guests and hosts. *International Journal of Hospitality Management*, 77, 405-414. <https://doi.org/https://doi.org/10.1016/j.ijhm.2018.08.004>
- Nakamura, S., Baskaran, A., & Selvarajan, S. K. (2024). Impact of Airbnb on the hotel industry in Japan [Article]. *Journal of Destination Marketing and Management*, 31, Article 100841. <https://doi.org/10.1016/j.jdmm.2023.100841>

- Oyewole, G. J., & Thopil, G. A. (2023). Data clustering: application and trends [Article]. *Artificial Intelligence Review*, 56(7), 6439-6475. <https://doi.org/10.1007/s10462-022-10325-y>
- Pang, J., Fu, X., & Qi, Y. (2024). The role of peer-to-peer interactions in sharing accommodation: From cocreation experience to peer customer engagement. *International Journal of Hospitality Management*, 120, 103764. <https://doi.org/https://doi.org/10.1016/j.ijhm.2024.103764>
- Park, S., & Tussyadiah, I. P. (2020). How Guests Develop Trust in Hosts: An Investigation of Trust Formation in P2P Accommodation. *Journal of Travel Research*, 59(8), 1402-1412. <https://doi.org/10.1177/0047287519884654>
- Proserpio, D., Xu, W., & Zervas, G. (2018). You get what you give: theory and evidence of reciprocity in the sharing economy. *Quantitative Marketing and Economics*, 16(4), 371-407. <https://doi.org/10.1007/s11129-018-9201-9>
- Qiu, H., Chen, D., Bi, J. W., Lyu, J., & Li, Q. (2022). The construction of the affinity-seeking strategies of Airbnb homestay hosts [Article]. *International Journal of Contemporary Hospitality Management*, 34(3), 861-884. <https://doi.org/10.1108/IJCHM-10-2020-1157>
- Qiu, W. W., Yu, H. T., Tsai, C. H. K., Zhu, D. J., Chen, M. H., & Kim, H. J. (2023). Understanding the value of host-guest intimacy behind online reviews of Airbnb [Article]. *International Journal of Hospitality Management*, 115, Article 103599. <https://doi.org/10.1016/j.ijhm.2023.103599>
- Ren, L., Zhao, D., Zhao, X., Chen, W., Li, L., Wu, T., Liang, G., Cai, Z., & Xu, S. (2022). Multi-level thresholding segmentation for pathological images: Optimal performance design of a new modified differential evolution [Article]. *Computers in Biology and Medicine*, 148, Article 105910. <https://doi.org/10.1016/j.compbiomed.2022.105910>
- Rong, K., & Luo, Y. (2023). Toward born sharing: The sharing economy evolution enabled by the digital ecosystems [Article]. *Technological Forecasting and Social Change*, 196, Article 122776. <https://doi.org/10.1016/j.techfore.2023.122776>
- Rong, K., & Luo, Y. (2023). Toward born sharing: The sharing economy evolution enabled by the digital ecosystems. *Technological Forecasting and Social Change*, 196, 122776. <https://doi.org/https://doi.org/10.1016/j.techfore.2023.122776>
- Rosenberg, A., & Hirschberg, J. (2007). *V-Measure: A Conditional Entropy-Based External Cluster Evaluation Measure*.
- Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20, 53-65. [https://doi.org/https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/https://doi.org/10.1016/0377-0427(87)90125-7)

- Santos, N., & Oliveira Moreira, C. (2021). Uncertainty and expectations in Portugal's tourism activities. Impacts of COVID-19 [Article]. *Research in Globalization*, 3, Article 100071. <https://doi.org/10.1016/j.resglo.2021.100071>
- Saravade, S., Felix, R., & Fırat, A. F. (2021). From Solidity to Liquidity: Macro-Level Consumption Patterns in the Sharing Economy. *Journal of Macromarketing*, 41(2), 284-296. <https://doi.org/10.1177/0276146720958184>
- Schor, J. B., Attwood-Charles, W., Cansoy, M., Ladegaard, I., & Wengronowitz, R. (2020). Dependence and precarity in the platform economy [Article]. *Theory and Society*, 49(5-6), 833-861. <https://doi.org/10.1007/s11186-020-09408-y>
- Sousa, V., Faustino, P., & Rebelo, C. (2021). International perception of Portugal as a tourism destination [Article]. *Journal of Tourism and Development*, 36(2), 575-591. <https://doi.org/10.34624/rtd.v36i2.8797>
- Srovnalíková, P., Semionovaite, E., Baranskaite, E., & Labanauskaite, D. (2020). Evaluation of the Impact of Sharing Economy on Hotel Business [Article]. *Journal of Tourism and Services*, 11(20), 150-169. <https://doi.org/10.29036/jots.v11i20.145>
- Sthapit, E., Björk, P., & Coudounaris, D. N. (2022). Airbnb: what determines a memorable experience? *Consumer Behavior in Tourism and Hospitality*, 17(1), 42-55. <https://doi.org/10.1108/cbth-06-2021-0158>
- Sthapit, E., Björk, P., Coudounaris, D. N., & Stone, M. J. (2022). A new conceptual framework for memorable Airbnb experiences: guests' perspectives. *International Journal of Culture, Tourism and Hospitality Research*, 16(1), 75-86. <https://doi.org/10.1108/ijcthr-01-2021-0002>
- Sthapit, E., & Jiménez-Barreto, J. (2018). Exploring tourists' memorable hospitality experiences: An Airbnb perspective. *Tourism Management Perspectives*, 28, 83-92. <https://doi.org/https://doi.org/10.1016/j.tmp.2018.08.006>
- Tussyadiah, I. (2016). Strategic Self-presentation in the Sharing Economy: Implications for Host Branding. In A. Inversini & R. Schegg (Eds.), (pp. 695-708): Springer, Cham.
- Valentinas, N., Petroké, I., Bačiulienė, V., & Tetiana, V. (2021). The Impact of the Sharing Economy as an Ecosystem on the Tourism Sector. *Journal of Tourism and Services*, 12(22), 66-88. <https://doi.org/10.29036/jots.v12i22.241>
- Vartiak, L., Garbarova, M., & Fabus, J. (2023). Comparative Analysis of Business Models of Selected Accommodation Portals. *Academic Journal of Interdisciplinary Studies*, 12(6), 55. <https://doi.org/10.36941/ajis-2023-0153>
- Wang, H., Zhang, W., & Ma, X. (2024). Contrastive and adversarial regularized multi-

- level representation learning for incomplete multi-view clustering [Article]. *Neural Networks*, 172, Article 106102. <https://doi.org/10.1016/j.neunet.2024.106102>
- Wirtz, J., So, K. K. F., Mody, M. A., Liu, S. Q., & Chun, H. H. (2019). Platforms in the peer-to-peer sharing economy. *Journal of Service Management*, 30(4), 452-483. <https://doi.org/10.1108/josm-11-2018-0369>
- Xie, K., Heo, C. Y., & Mao, Z. E. (2021). Do professional hosts matter? Evidence from multi-listing and full-time hosts in Airbnb. *Journal of Hospitality and Tourism Management*, 47, 413-421. <https://doi.org/10.1016/j.jhtm.2021.04.016>
- Xue, L., Leung, X. Y., & Ma, S. (2022). What makes a good “guest”: Evidence from Airbnb hosts' reviews. *Annals of Tourism Research*, 95, 103426. <https://doi.org/https://doi.org/10.1016/j.annals.2022.103426>
- Yan, Y., Lu, B., & Xu, T. (2023). Exploring the Relationship between Host Self-Description and Consumer Purchase Behavior Using a Self-Presentation Strategy. *Systems*, 11(8), 430. <https://doi.org/10.3390/systems11080430>
- Yang, Z., & Tan, Y. (2024). The methods for improving large-scale multi-view clustering efficiency: a survey. *Artificial Intelligence Review*, 57(6), 153. <https://doi.org/10.1007/s10462-024-10785-4>
- Ye, Q., Liang, S., Wei, Z., & Law, R. (2023). Effects of reputation on guest satisfaction: from the perspective of two-sided reviews on Airbnb. *International Journal of Contemporary Hospitality Management*, 35(10), 3718-3736. <https://doi.org/10.1108/ijchm-09-2022-1170>
- Ye, S., Lei, S. I., Zhao, X., Zhu, L., & Law, R. (2023). Modeling tourists' preference between hotels and peer-to-peer (P2P) sharing accommodation: a pre- and post-COVID-19 comparison. *International Journal of Contemporary Hospitality Management*, 35(4), 1423-1447. <https://doi.org/10.1108/IJCHM-12-2021-1556>
- Zach, F., Nicolau, J., & Sharma, A. (2019). Disruptive innovation, innovation adoption and incumbent market value: The case of Airbnb. *Annals of Tourism Research*, 80. <https://doi.org/10.1016/j.annals.2019.102818>
- Zal, S., Guo, L., Tang, C., & Zhang, J. (2024). Exploring the role of the service provider in sharing economy services. *Journal of Services Marketing*, 38(6), 746-760. <https://doi.org/10.1108/JSM-05-2023-0171>
- Zamani, E. D., Choudrie, J., Katechos, G., & Yin, Y. (2019). Trust in the sharing economy: the AirBnB case. *Industrial Management & Data Systems*, 119(9), 1947-1968. <https://doi.org/10.1108/IMDS-04-2019-0207>
- Zhang, L., Yan, Q., & Zhang, L. (2020). A text analytics framework for understanding the relationships among host self-description, trust perception and purchase

behavior on Airbnb [Article]. *Decision Support Systems*, 133, Article 113288.
<https://doi.org/10.1016/j.dss.2020.113288>

Anexos

Anexo 1 – Literatura existente acerca dos tipos de anfitriões

Tipo de Anfitrião	Descrição
O Cidadão Global (Tussyadiah, 2016)	Anfitriões que valorizam a interação cultural, abertos a conhecer pessoas de diferentes partes do mundo.
O Especialista Local (Tussyadiah, 2016)	Anfitriões com conhecimento local profundo, oferecem uma experiência imersiva na cultura local.
O Pessoal (Tussyadiah, 2016)	Anfitriões focados em proporcionar uma experiência pessoal e autêntica, enfatizando características pessoais.
O Sucedido (Tussyadiah, 2016)	Anfitriões com um certo nível de realização, destacam sua ocupação profissional e educação.
O Criativo (Tussyadiah, 2016)	Anfitriões envolvidos em indústrias criativas, como artes e design, oferecendo uma experiência única e inovadora.
Profissionais Emergentes (Farmaki & Kaniadakis, 2020)	Anfitriões que gerem múltiplas propriedades, tratam Airbnb como um negócio, foco na otimização de receitas.
Anfitriões Individuais com Motivação Económica (Farmaki & Kaniadakis, 2020)	Anfitriões que possuem uma ou duas propriedades, motivados principalmente pela geração de uma fonte de rendimento adicional.
Anfitriões Coabitantes Economicamente Orientados (Farmaki & Kaniadakis, 2020)	Anfitriões que compartilham o seu espaço pessoal, motivados economicamente e pela experiência, com ênfase no aspeto financeiro.
Anfitriões Coabitantes Socialmente Orientados (Farmaki & Kaniadakis, 2020)	Anfitriões que compartilham o seu espaço pessoal, mais focados nas interações sociais e na experiência cultural de hospedar.
Anfitrião amigável e competente (Qu et al, 2022)	Anfitriões com características como ser amigável e acolhedor, além disso, demonstra habilidade e eficiência na gestão do seu espaço e na prestação de serviços.
Anfitriões profissionais (Abrate et al, 2022)	Anfitriões que gerem múltiplos anúncios, que podem incluir propriedades próprias e de terceiros.
Anfitriões não profissionais (Abrate et al, 2022)	Anfitriões que compartilham espaços extras nas suas casas, muitas vezes motivados pela interação social ou económica.

Anexo 2- Número de anúncios na plataforma Airbnb em Lisboa em 2024 por freguesia

Cidade	Total
Santa Maria Maior	3222
Misericórdia	2429
Arroios	2080
Cascais e Estoril	1512
Santo António	1410

Anexo 3- Número de anúncios na plataforma Airbnb no Porto em 2024 por freguesia

Cidade	Total
Cedofeita, Santo Ildefonso, Sé, Miragaia, São Nicolau e Vitória	7202
Bonfim	1445
Santa Marinha e São Pedro da Afurada	711
Paranhos	572
Lordelo de Ouro e Massarelos	560

Anexo 4- Descrição da implementação do programa em python

O programa implementado no script main.py, começa por solicitar ao utilizador os

nomes dos ficheiros que contêm as *reviews* e as informações relativas aos anfitriões e propriedades. De seguida, o programa utiliza a função `cleanLog()` do módulo `utils` para efetuar a limpeza do ficheiro de log (`log.txt`), preparando-o para registar novas entradas. Este procedimento assegura que o ficheiro de log mantém apenas registos relevantes para a execução corrente. Após esta etapa, a função `readFile()` é invocada para proceder à leitura do ficheiro que contém as informações dos anfitriões e propriedades. Esta leitura é realizada recorrendo a bibliotecas especializadas na manipulação de ficheiros Excel, nomeadamente `pandas` e `openpyxl`, garantindo a correta importação dos dados para o ambiente de trabalho.

Num segundo momento, a função `getIds()` é empregue para extrair os identificadores (IDs) a partir do dataframe previamente carregado do ficheiro dos anfitriões e propriedades. A extração dos IDs é realizada sob a forma de uma lista, com a aplicação de filtros que eliminam valores nulos ou inválidos, assegurando a integridade dos dados a serem processados. Subsequentemente, a função `getReviews()` é chamada para abrir e processar o ficheiro de *reviews*. Os dados deste ficheiro são carregados num dataframe, onde os IDs presentes são comparados com a lista de IDs extraída inicialmente, permitindo a identificação de correspondências e garantindo que apenas os IDs válidos e relevantes são considerados.

Para cada ID validado, a função `findReview()` é utilizada para localizar a *review* correspondente no dataframe de *reviews*. Esta função realiza uma pesquisa estruturada, assegurando que cada *review* é corretamente associada ao ID respetivo. As *reviews* localizadas são então preparadas para análise, um processo que inclui a substituição de tags HTML, como `
`, por quebras de linha, garantindo assim a legibilidade e consistência dos textos antes de serem submetidos a análises subsequentes.

Finalmente, a função `analiseReview()` do módulo `bot` é responsável pela interação com a API do OpenAI, utilizando o modelo `gpt-4` para a análise textual das *reviews*. Nesta etapa, um prompt específico é formulado e enviado ao modelo, solicitando uma análise detalhada baseada em seis categorias predefinidas: Limpeza, Precisão, Check-in, Comunicação, Localização e Relação Preço-Qualidade. Esta abordagem permite obter uma avaliação sistemática e abrangente das *reviews*, orientada para critérios de qualidade relevantes no contexto da hospitalidade.

Esta abordagem possibilitou a avaliação de cada comentário numa escala de um a cinco para cada dimensão analítica. Comentários que não mencionassem explicitamente uma dimensão recebiam automaticamente pontuação zero, indicando a ausência de avaliação para

essa característica específica. Expressões como “irei voltar com certeza” ou “não valeu o preço” influenciavam a pontuação da característica relação preço/qualidade, enquanto comentários como “faltou água quente, mas o anfitrião resolveu muito rápido” ou “o anfitrião deu as indicações erradas” impactavam a pontuação na comunicação. A capacidade do sistema de processar estas nuances permitiu uma avaliação precisa das mil avaliações nas dimensões estabelecidas.

Anexo 5 – Comparação Tabela de Contingência

Hosts	Modelo B	Modelo A
1	2	4
2	2	2
3	3	3
4	2	3
5	4	6
6	1	5
7	1	5
8	4	6
9	4	6
10	5	1

Anexo 6 – Comparação dos Centroides A vs B

Cluster_Centroids_A	1	2	3	4	5
1	0	0	0	0	1
2	0	1	0	0	0
3	0	1	1	0	0
4	0	1	0	0	0
5	2	1	0	0	0
6	0	0	0	3	0

Anexo 7 - Centroides das Métricas

Adjusted Rand Index (ARI)	0.6170212765957447
Normalized Mutual Information (NMI)	0.8537076662093104
Homogeneity	0.8056404908414813
Completeness	0.9078744580780113
V-Measure	0.8537076662093104

Anexo 8 - Discussão geral de resultados

Hipóteses	Descrição	Resultado
H1	Uma abordagem multinível apresenta uma segmentação mais eficaz dos anfitriões, ao capturar de forma mais detalhada as interações complexas entre diferentes níveis de dados (avaliações, características das propriedades e informações dos anfitriões), em comparação com abordagens uninível.	Suportada
H1.1	Uma abordagem multinível apresenta uma melhor qualidade de segmentação em comparação com abordagens uninível.	Não suportada
H1.2	Uma abordagem multinível não apresenta correspondência com abordagens uninível.	Suportada
H1.3	Uma abordagem multinível não apresenta concordância com abordagens uninível.	Suportada
H1.4	Uma abordagem multinível, devido à sua complexidade, proporcionará insights mais ricos e será mais eficaz na identificação de padrões em comparação com abordagens uninível.	Suportada

FACULDADE DE ECONOMIA

