INSTITUTO
UNIVERSITÁRIO
DE LISBOA

Impact of Covid-19 Pandemic on the Dynamics of Accommodation Sharing: A Hosts Perspective

Filipa Camarate Gonçalves

Master in Computer Science and Business Management

Supervisor:
Doctor Fernando Manuel Marques Batista, Associate Professor, Iscte – University Institute of Lisbon

Co-Supervisor:
Doctor Ricardo Daniel Santos Faro Marques Ribeiro, Associate Professor, Iscte – University Institute of Lisbon

September, 2024

TECHNOLOGY
AND ARCHITECTURE

Department of Information Science and Technology

Impact of Covid-19 Pandemic on the Dynamics of Accommodation Sharing: A Hosts Perspective

Filipa Camarate Gonçalves

Master in Computer Science and Business Management

Supervisor:
Doctor Fernando Manuel Marques Batista, Associate Professor,
Iscte – University Institute of Lisbon

Co-Supervisor:
Doctor Ricardo Daniel Santos Faro Marques Ribeiro, Associate Professor, Iscte – University Institute of Lisbon

September, 2024

To my grandfather Manuel and my grandmother Maria

Acknowledgment

I would like to begin by thanking my mother, Fátima, and my boyfriend, Miguel, for all the support and motivation they gave me throughout the elaboration of this master's thesis.

I also wish to thank my supervisors, Professor Fernando Batista and Professor Ricardo Ribeiro, for their availability and all the guidance they provided during this journey.

Lastly, my heartfelt thanks go to my family and friends for their unwavering encouragement, which helped me overcome this challenge.

Resumo

O presente estudo investiga os efeitos da pandemia de COVID-19 na economia de alojamentos partilhados, analisando especificamente como os anfitriões do Airbnb no distrito de Lisboa ajustaram a forma como disponibilizam os seus alojamentos em resposta a mudanças nas condições de mercado. À medida que plataformas como o Airbnb transformaram os modelos tradicionais de hospedagem, compreender estas adaptações é crucial para identificar tendências no comportamento dos consumidores e nas implicações económicas.

Utilizando modelos avançados de classificação de dados, a presente investigação examina as alterações nas características dos alojamentos antes e depois da pandemia. Os resultados revelam transformações significativas nos serviços disponibilizados, destacando um foco crescente em comodidades modernas, sustentabilidade e distinções regionais. Os alojamentos urbanos tendem a enfatizar a praticidade e o conforto, enquanto os alojamentos rurais mantêm características tradicionais que atraem hóspedes que procuram espaço e tranquilidade.

A presente investigação contribui com informações valiosas sobre a evolução do panorama da economia de alojamentos partilhados, ilustrando a resiliência e adaptabilidade dos anfitriões face a perturbações globais. O trabalho futuro é proposto para explorar tendências a longo prazo, melhorar as metodologias de classificação para regiões menos representadas e realizar estudos comparativos em diferentes regiões, aprofundando assim a compreensão neste campo dinâmico.

PALAVRAS CHAVE: *economia de alojamentos partilhados*, *airbnb*, *anfitrião*, *Lisboa*, *impacto da COVID-19*

Abstract

This study investigates the effects of the COVID-19 pandemic on the accommodation sharing economy, specifically analyzing how Airbnb hosts in the Lisbon district adjusted their listings in response to changing market conditions. As platforms like Airbnb have transformed traditional hospitality models, understanding these adaptations is crucial for identifying trends in consumer behavior and economic implications.

Employing advanced data classification models, the research examines shifts in the characteristics of accommodations before and after the pandemic. The findings reveal significant transformations in provided services, highlighting an increased focus on modern amenities, sustainability, and regional distinctions. Urban accommodations tended to emphasize practicality and comfort, while rural listings retained traditional features that appeal to guests seeking spaciousness and tranquility.

This research contributes valuable insights into the evolving landscape of the accommodation sharing economy, illustrating the resilience and adaptability of hosts amid global disruptions. Future work is proposed to explore longitudinal trends, improve classification methodologies for minority regions, and conduct comparative studies across different regions, further enhancing understanding in this dynamic field.

KEYWORDS: *accommodation sharing economy*, *airbnb*, *host*, *Lisbon*, *COVID-19 impact*

Contents

List of Figures

List of Tables

CHAPTER 1

Introduction

1.1. Motivation

In the context of accommodation sharing, individuals leverage online platforms to rent out their houses or rooms to others in exchange for financial compensation [1]. The appeal of sharing accommodations lies in enhanced social interaction and the cost-effectiveness of securing quality lodgings, thereby transforming hospitality services, diverging from the traditional hotel industry model [2], [3].

Airbnb is a prominent and groundbreaking intermediary in the accommodation sharing economy that has significantly transformed the global hospitality industry. Airbnb experienced substantial growth, exerting an impact on traditional accommodation providers in the latter part of the 21st century's second decade, through the introduction of a technology-driven distribution platform [4], [5]. This innovative model quickly gained traction, inspiring the adoption of similar approaches by various tourism intermediaries, particularly in urban settings. While there is a surge in scientific research focusing on the accommodation sharing economy, much of the attention is directed towards Airbnb within the realm of tourism academia, leaving other intermediaries and peer-to-peer markets beyond urban areas relatively overlooked [6], [7].

The growing importance of analysing information from shared accommodation economy platforms reflects an intrinsic need to understand and explore the complex dynamics of this sector. As the sharing economy transforms how people seek and offer accommodation, the ability to analyse data from these platforms becomes crucial in identifying trends, consumer behaviour patterns, and economic implications [8].

The ease of access and availability of such data also empower an increasing number of analyses within the shared accommodation economy. Notably, many of these data are readily provided by independent scraping-focused websites, including Inside Airbnb (http://insideairbnb.com) and AirDNA (https://www.airdna.co/), among others, aggregating valuable information from these platforms. Authors such as Koh *et al.* [9] and Mousavi and Zhao [10] stand out for leveraging independent scraping-focused websites, particularly emphasizing Inside Airbnb, in the execution of their research endeavours. Similarly, Ndaguba and Zyl [11] directed their focus towards AirDNA for the same purpose.

Additionally, direct extraction of data from peer-to-peer accommodation platforms using scraping techniques on these platforms is a viable alternative. Authors like Stors and Baltes [12], Chang and Li [13], and Kim, Park, and Yi [14] exemplify researchers who employed scraping techniques to gather essential data for their respective research initiatives. This facilitated and diversified accessibility makes these data invaluable resources for studies and analyses in the context of the shared accommodation economy.

Furthermore, the capability of computer science methods to create value through the available information represents a significant opportunity for both individual studies and comprehensive reviews. Advanced techniques such as data mining, text analysis, and machine learning can unveil valuable insights, emerging patterns, and practical implications for the various stakeholders involved in the shared accommodation economy.

The remarkable growth of the peer-to-peer (P2P) accommodation market in the second decade of the 21st century faced a temporary setback due to the COVID-19 pandemic. The global spread of the COVID-19 virus and its declaration as a pandemic significantly impacted the tourism industry [15], marking a crisis that brought substantial changes to the world, particularly affecting the tourism and hospitality sectors [16], [17]. The negative repercussions of the COVID-19 pandemic extended to the tourism and hospitality industry, impacting platforms like Airbnb [18]. Nearly every sector worldwide felt the effects of the pandemic, with the hotel industry standing out as one of the most severely affected [19].

Considering the challenges posed by the COVID-19 pandemic to the accommodation sharing economy, there arises a critical need for research to explore suitable resources, encompassing data types and computer science methodologies. Understanding the nuanced impact of the pandemic on this evolving sector becomes imperative, prompting a focused examination into the tools and analytical approaches that can illuminate the dynamics of the accommodation sharing economy during these unprecedented times.

1.2. Goals and research questions

The objective of this study is to identify and comprehend potential changes observed in the accommodation sharing economy caused by the COVID-19 pandemic. Specifically, the aim is to understand whether there were any and what changes occurred from the perspective of hosts, by studying how they made their accommodations available on Airbnb before and after the pandemic period. This analysis will focus on the Lisbon district in Portugal, providing a localized understanding of the impacts within this particular region.

Particularly, the study aims to answer the following research questions:

Q1: How has the listing of shared accommodations on Airbnb by hosts in the Lisbon district changed before and after the Covid-19 pandemic? What specific changes have occurred?

Q2: Are there differences in the key characteristics of accommodations across different areas of the Lisbon district? If so, what are these differences?

In other words, the first research question seeks to understand the changes that have occurred in how Airbnb hosts in the Lisbon district list their properties before and after the Covid-19 pandemic. The second research question, on the other hand, seeks to identify whether there are differences in the most significant characteristics of accommodations across different areas of the Lisbon district, and if so, what those differences are.

1.3. Methodology

To determine whether and how the Covid-19 pandemic has influenced the way hosts list their accommodations on Airbnb across different regions of the Lisbon district, the first task is to develop data classification models that connect the amenities of accommodations with their geographic locations for the years preceding and following the pandemic. The second task is to identify the most representative amenities for each region within Lisbon.

The Cross Industry Standard Process for Data Mining (CRISP-DM) methodology is applied. This approach includes the following stages, as outlined by Azevedo and Santos [20]:

- Business Understanding: Taking the business perspective into account, it focuses on understanding the problem, objectives, and project requirements. In the present study, this phase involves conducting a systematic literature review aimed at understanding what has already been investigated in the context of the accommodation sharing economy and identifying existing gaps, so that this study can contribute new knowledge that has not yet been explored.
- Data Understanding: Interpretation, exploration, and evaluation of the quality of the collected data. In this study, after selecting the data source to be used, a review of all available databases and the data they contain was conducted to identify those that best align with the objectives of the present research and how they can be utilized most effectively.
- Data Preparation: This phase involves the extraction and preparation of the data identified as relevant to the research. Specifically, it was necessary to segment and normalize the original data to ensure that it could be included in the classification models as efficiently as possible, ultimately allowing for better results.
- Modeling: Application of classification algorithms that allowed for the identification of the characteristics of accommodations that most distinguish the various regions of the Lisbon district.
- Evaluation: Assessment of the results obtained through the techniques applied in the modeling phase and selection of the most suitable algorithm for calculating the probabilities of membership of each independent variable to the various classes defined by the dependent variable.
- Deployment: Although the CRISP-DM methodology was used as the foundation, the Deployment stage, which corresponds to the strategic implementation of the obtained model, was not carried out due to its exclusion from the scope of this research.

After completing the Evaluation stage, probability calculations for class membership were conducted, revealing the most representative characteristics of each region in the Lisbon district, which subsequently addressed the second research question.

1.4. Document Structure

This research is organized as follows: Chapter 1 provides the introduction and Chapter 2 reviews the state of the art. Chapter 3 describes the data used in the study, including its collection and preparation for analysis. Chapter 4 presents the results obtained from the research. Finally, Chapter 5 offers the conclusions and suggests directions for future work based on the findings of this investigation.

CHAPTER 2

Literature Review

The present research utilized a Systematic Literature Review (SLR) methodology to identify the predominant themes, research methodologies, and conclusions drawn from existing studies in the shared accommodation economy. This approach is characterized by Okoli and Schabram [21] as "a systematic, explicit, and reproducible method for identifying, evaluating, and synthesizing the existing body of completed and recorded work produced by researchers, scholars, and practitioners". The adoption of SLR enhances literature reviews by introducing transparency and rigor into the process [22]. The protocol employed in this review adheres to a traditional approach pioneered by Kitchenham [23], encompassing the stages of Planning, Conducting, and Reporting, as illustrated in Table 2.1.

2.1. Planning the review

The present review aims to offer a comprehensive analysis of the landscape of studies related to the shared accommodation economy, providing valuable insights for researchers, professionals, and those interested in the subject. Specifically, the review seeks to map and highlight key subjects within the shared accommodation economy, with a focus on identifying prevalent research themes.

Concurrently, the review aims to establish a relationship between identified themes and the methods employed in the examined studies. This entails a critical analysis of the methodological approaches used by researchers to investigate phenomena related to the shared accommodation economy. This connection between themes and methods allows for a deeper understanding of the scientific approach adopted in the reviewed research.

Finally, the SLR aims to link the investigated themes with the conclusions reached by the studies. This involves examining how research in the shared accommodation economy has contributed to advancing knowledge, identifying patterns, gaps, and emerging trends. This correlation between themes and conclusions provides a comprehensive view of the current state of investigation in this specific field.

To acquire the information needed to achieve the previously established objectives, the following research expressions were formulated:

- First research expression: (accommodation OR Airbnb) AND (amenities OR host OR description) AND (mining OR analysis OR learning)
- Second research expression: (accommodation OR Airbnb) AND (reviews OR feedback OR rating) AND (mining OR analysis OR learning)

Table 2.1. Stages of the SLR methodology

Planning the review	Conducting the review	Reporting the review
Motivation	**Search results**	• Presentation and analysis of the results
• The growing importance of analysing information provided by shared accommodation economy platforms • Availability and accessibility of information in the shared accommodation economy • The ability of computer science methods to create value through the datasets made available	• Following the application of filters, a total of 68 articles were successfully obtained	
Objectives	**Data extraction and analysis**	
• Identification of predominant themes, research methods and conclusions drawn from the existing studies in the shared accommodation economy	• Article publication type analysis • Article publication year distribution • Article quality analysis	
Protocol		
• Search databases • Define research expressions • Define inclusion, exclusion, and quality criteria		

The research expressions have been carefully crafted to provide a comprehensive and balanced approach, focusing on both the host and guest perspectives in the dynamics of the shared accommodation economy.

The first research expression is specifically designed to deepen the understanding of practices and strategies adopted by hosts. The terms "amenities", "host", and "description" direct attention to the accommodation's features, the host's role and behaviour, and the provided descriptions, enabling a more specific analysis from the owner's perspective.

On the other hand, the second research expression is more geared towards the guest experience. By exploring "reviews", "feedback", and "rating", this expression aims to comprehend guests' evaluations and perceptions, offering an in-depth view of the stay's quality and interactions with hosts.

Incorporating the terms "mining", "analysis", and "learning" into both research expressions is pivotal for directing the investigation through advanced methodologies like data mining, text analysis, and machine learning. This technical aspect is indispensable for extracting valuable insights from the available data, establishing a robust foundation to comprehend practices and trends in the shared accommodation economy.

Having completed this initial phase, three databases deemed most appropriate for the scope of this research were selected, namely:

- ACM Digital Library (https://dl.acm.org/)
- Scopus (https://www.scopus.com/)
- Web of Science (https://www.webofscience.com/)

In the initial stage, the search expressions were entered into the three databases without the application of filters, unveiling a significant diversity among the studies. This diversity emphasized the importance of implementing filters to streamline and focus the research.

Therefore, various filters were implemented to ensure the quality and relevance of the studies included in this investigation. Initially, the first filter ensured that the search expressions were exclusively applied to the title, abstract, or keywords of each study. Subsequently, the second filter restricted the search to studies written solely in English. The third filter involves excluding studies with publication dates outside the considered appropriate timeframe, which spans from 2018 to 2023. The fourth filter excludes studies outside the Computer Science field from the research. The fifth filter ensured the removal of duplicate articles across the three databases, preventing redundant analysis of the same article. The sixth filter removes articles of low quality from the research. To accomplish this, it was determined that an article deemed of good quality falls within the Q1 or Q2 categories according to Scimago (https://www.scimagojr.com/), and a conference paper of good quality belongs to categories A or B in the ERA ranking, as indicated by the Conference Ranks website (http://www.conferenceranks.com/). Finally, the seventh filter involved manual review of articles, facilitating a more precise and in-depth analysis of each, ensuring that all are relevant and deserving of inclusion in this investigation.

2.2. Conducting the review

Following the application of all filters, the first search expression yielded forty-three articles, while the second search expression produced twenty-five articles. The specific number of articles resulting from each filtering phase for the first and second search expressions is provided in Tables 2.2 and 2.3, respectively. Detailed information about the articles can be found in the tables A.1 and Table A.2 in Appendix A.

After appropriately filtering the studies, it is possible to analyze the main characteristics of the final sample of articles. The source, year of publication, and quality of the

TABLE 2.2. Filtering steps for the first research expression

DB	No filter	First filter	Second filter	Third filter	Fourth filter	Fifth filter	Sixth filter	Seventh filter
Scopus	28876	1261	1206	686	127	126	70	39
IEEE Xplore	8284	32	32	16	16	1	0	0
Web of Science	1191	960	915	597	52	13	8	4
Total	38351	2208	2153	1299	195	140	78	43

TABLE 2.3. Filtering steps for the second research expression

DB	No filter	First filter	Second filter	Third filter	Fourth filter	Fifth filter	Sixth filter	Seventh filter
Scopus	142132	1261	2505	1207	217	168	70	23
IEEE Xplore	8031	94	94	40	40	10	6	0
Web of Science	4449	3408	3310	1654	114	33	25	2
Total	154612	6133	5909	2901	371	211	101	25

articles, as provided by scientific article and conference databases, were manually assessed to understand the distribution of the final sample across these three criteria.

Figure 2.1 reveals that approximately 88% of the articles are sourced from scientific journals, making it the predominant information source for the articles essential to this investigation.

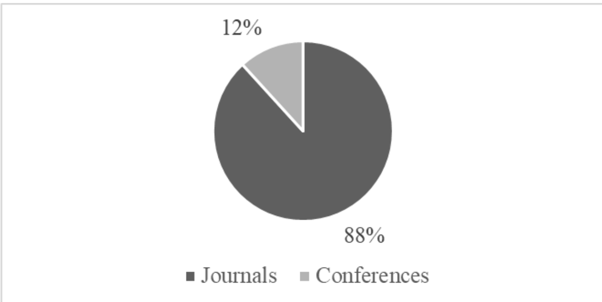

FIGURE 2.1. Articles distribution by publication type

Figure 2.2 illustrates the distribution of the number of articles by publication year. In the obtained sample of articles, 2021 had the highest number of publications, while 2019 had the fewest. It is noteworthy that the count for the year 2023 includes articles published up to December 13th, the date of the article review.

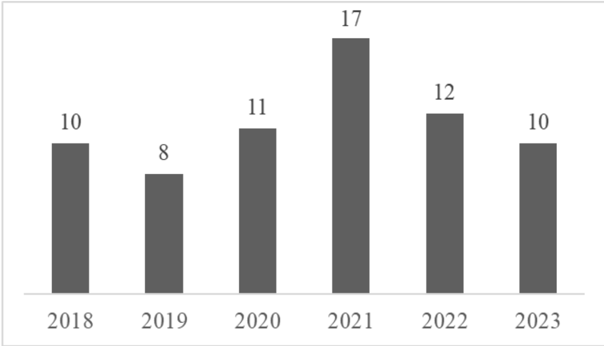

FIGURE 2.2. Articles distribution by publication year

Finally, Figure 2.3 depicts how the final number of articles is distributed based on the quality rank of their sources. Notably, most articles from scientific journals are categorized as Q2, while most conference papers are ranked as A according to the ERA rating scale.

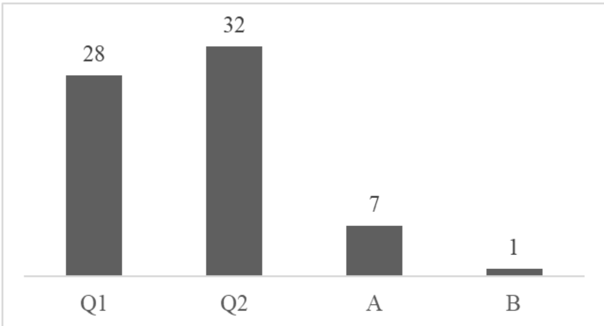

FIGURE 2.3. Articles distribution by source quality rank

2.3. Reporting the review

This stage of the systematic literature review is aimed at presenting and analyzing the results obtained from the review. Accordingly, Section 2.3.1 presents the prevalent research themes identified within the shared accommodation economy and links these themes to the conclusions drawn by the studies. In turn, Section 2.3.2 focuses on outlining the key methods used in the analyzed studies and establishing a connection between the identified themes and the employed methods.

2.3.1. Most prominent themes identified in the review

The manual assessment of the final sample of articles revealed that, in general, each article explores a specific theme. Table 2.4 presents the distribution of the filtered sample of articles across the various identified themes. This study focuses on understanding the four predominant themes in the filtered sample (customer experience, customer purchase behavior, price, and trust), highlighting and synthesizing the key findings from the articles within each theme.

Table 2.4. Most prominent themes, based on manual assessment

Theme	**Number of articles**	**Articles**
Customer experience	22	[24]–[45]
Customer purchase behaviour	13	[14], [46]–[57]
Price	9	[13], [58]–[65]
Trust	6	[66]–[71]
Discrimination & inequality	4	[9], [72]–[74]
Impact of Covid-19 pandemic	3	[75]–[77]
Impact of regulation and policy	3	[10], [78], [79]
Urban environment	3	[12], [80], [81]
Impact of accommodation sharing on traditional accommodation	1	[82]
Pros and cons of accommodation sharing	1	[83]
Sustainability	1	[84]
Other	2	[11], [85]

2.3.1.1 Customer experience

Under the topic "Customer experience", the focus lies on scrutinizing the experiences of consumers during their stays in specific accommodations. Typically, this involves an examination of the reviews posted by users on dedicated accommodation rental platforms like Airbnb, where users can freely share their opinions and perceptions.

In examining various studies, a comprehensive understanding emerges regarding diverse facets of customer experiences and preferences within the sharing economy, particularly focusing on Airbnb.

Lee *et al.* [40] investigation into Airbnb customer reviews spanning 2011-2015 reveals recurring themes in guest priorities. Attributes such as amenities, cleanliness, homeliness, host qualities, location, and transport connectivity consistently shape guest evaluations. Furthermore, seasonal shifts in preferences are notable, with heightened emphasis on the neighbourhood during summer and specific amenities like hot showers, fast internet, and television in winter.

Quattrone *et al.* [35] study on Airbnb dynamics highlights a significant shift in guest priorities towards business-related aspects over social considerations. The research emphasizes the growing importance of practical factors such as location, property type, and interior features. Interestingly, despite the shift towards business considerations, interpersonal interactions and host characteristics retain their significance in shaping the overall guest experience. Moreover, the study explores the nuanced relationship between neighbourhood adoption rates and social scores, revealing intriguing dynamics in areas with varying Airbnb penetration.

Zhang and Fu [25] analysis of accommodation experiences for domestic and foreign guests identifies distinct dimensions shaping their preferences. Foreign guests prioritize recommendation and booking flexibility, emphasizing personalized suggestions and reservation options. Conversely, domestic guests place importance on revisit and cleanliness, showcasing a preference for familiarity and hygiene. The study highlights cultural variables as influential factors shaping these varied preferences, emphasizing the impact of culture on guest expectations in the hospitality sector.

Bai *et al.* [30] investigation into Airbnb customer reviews delves into trending topics mentioned by guests. General experience emerges as the predominant topic, encompassing both overall and special experiences. The study notes fluctuations in topic proportions across different periods, with significant shifts during the early years of Airbnb and the COVID-19 pandemic. Notably, there is a decreasing proportion of topics related to customers' special experiences, attributed to the rise of professional hosts and service standardization.

The studies on customer experiences within the sharing economy, with a focus on Airbnb, reveal dynamic and evolving preferences. Key attributes consistently shaping guest evaluations include amenities, cleanliness, homeliness, host qualities, location, and transport connectivity. Seasonal shifts and a notable transition toward business-related aspects underscore the complexity of guest priorities. Despite the shift, interpersonal interactions and host characteristics remain pivotal. Cultural variables significantly influence preferences for domestic and foreign guests, emphasizing the need for cultural awareness in the hospitality sector. The evolving topics in Airbnb reviews, particularly the decreasing emphasis on special experiences, highlight the impact of industry changes. In this rapidly changing landscape, recognizing and adapting to these dynamics is crucial for delivering exceptional customer experiences.

2.3.1.2 Customer purchase behaviour

In the realm of "Customer purchase behaviour", the articles delve into studying the purchasing patterns of users across various online accommodation booking platforms. This behaviour primarily encompasses the process of renting accommodations through these platforms. A common metric used to analyse this behaviour is the number of reviews left on platforms such as Airbnb, often regarded as indicative of the minimum number of bookings for a particular accommodation. This practice stems from the fact that users can only leave a review after completing a reservation. Consequently, these reviews offer valuable insights into customers' purchasing habits and their overall satisfaction with their lodging experiences.

The influence of property characteristics on Airbnb reviews is a common thread in studies by Nash [46], Biswas, Sengupta, and Chatterjee [47] and Zhang, Lu, and Lu [48].

The authors emphasize the importance of property and room type, indicating that more exclusive and private stays generally lead to more reviews. However, consensus is lacking regarding the impact of the number of rooms and maximum capacity on review counts. Nash [46] suggests a positive contribution, while Biswas, Sengupta, and Chatterjee [47] and Zhang, Lu, and Lu [48] argue the opposite. Biswas, Sengupta, and Chatterjee [47] additionally highlight the positive effect of unique amenities on review counts, emphasizing the role of diverse amenities in enhancing guest experiences.

In terms of host-related factors, the possession of a superhost badge emerges as a positive factor in all three studies [46]–[48]. Verified host identities also contribute positively to review counts, as found in Nash [46] and Biswas, Sengupta, and Chatterjee [47]. The importance of a quick host response time in influencing review behaviour is emphasized by Nash [46] and Zhang, Lu, and Lu [48]. Further, Biswas, Sengupta, and Chatterjee [47] explore additional factors such as hosts having a profile picture and being the longest-serving on the platform, which positively influence review counts. In contrast, Zhang, Lu, and Lu [48] provide insights into the impact of host characteristics like facial attractiveness and self-introduction diversity on room bookings, revealing the complex interplay between host traits and guest behaviour.

In terms of review and rating characteristics, the positive influence of ratings on review counts is a consistent finding across Nash [46], Biswas, Sengupta, and Chatterjee [47] and Zhang, Lu, and Lu [48]. Biswas, Sengupta, and Chatterjee [47] delve deeper into user-generated content, finding that positive sentiments boost review counts, while negative sentiments have a corresponding negative impact. Additionally, Liu *et al.* [51] explore electronic word-of-mouth (eWOM) and its impact on sales performance. Both host-based and listing-based eWOM positively affect sales, with host-based eWOM exerting a stronger influence. The positive effect of listing-based eWOM is moderated by the listing price, while host responsiveness strengthens the impact of host-based eWOM on sales performance.

Finally, in booking characteristics, the studies by Biswas, Sengupta, and Chatterjee [47] and Zhang, Lu, and Lu [48] reveal a significant negative effect of the price per night on customer reviews or bookings. Higher pricing is shown to deter guests from leaving reviews or making bookings. Biswas, Sengupta, and Chatterjee [47]) further explore website attributes, with positive effects observed for attributes like summary length, readability, and the use of superlatives. Guest phone verification and a 90-day availability are also associated with more positive customer reviews. However, a strict cancellation policy is linked to a negative impact on review counts. Zhang, Lu, and Lu [48] introduce instant bookable options, showing a significant positive effect on room bookings, indicating the appeal of properties that allow instant bookings.

2.3.1.3 Price

Several key factors play a crucial role in predicting pricing within the accommodation sector. The number of bedrooms, bathrooms, guests, and beds are identified as significant contributors to price prediction. Additionally, the number of listings per host, host response rate, cancellation policy, and response time are essential elements influencing the pricing dynamics [64].

Room type, city, and location are recognized as primary price determinants, indicating that the type of accommodation, its geographical setting, and specific location within a city significantly impact pricing. Moreover, the number of pictures posted, and the range of amenities provided emerge as influential determinants affecting pricing decisions.

An interesting observation is that the relationship between the number of pictures posted, and prices varies across room types. Prices for entire homes and private rooms tend to increase with the number of pictures, while for shared rooms, more pictures result in decreased prices. Additionally, customer preferences for the same attribute may differ across different cities, introducing a nuanced layer to the pricing dynamics in the accommodation sharing market.

These findings are supported by Engin and Vetschera [65] who assert that both premises information and host-related information have a substantial impact on accommodation prices in the shared economy. Notably, premises information exerts a more pronounced influence than host-related information.

Considering the standardization phenomenon observed in this economy, it becomes crucial to comprehend the composition of hosts within the Airbnb market. Contrary to assumptions, having a high ratio of professional decision-makers does not ensure maximized transaction values, supply sizes, or platform profits. Jia and Wang [61] suggest that, rather than focusing on recruiting a large proportion of professional hosts, emphasizing the distinctive features of various host types can enhance customer experiences and effectively set apart market positioning, ultimately impacting market outcomes.

2.3.1.4 Trust

In the dynamic landscape of the sharing economy, trust plays a pivotal role in shaping consumer behaviour. This section explores the intricate relationship between perceived trust, host attributes, and guest behaviours within the Airbnb platform.

Zhang, Yan, and Zhang [66] emphasizes the factors influencing perceived trust in Airbnb hosts. Analyses reveal a positive correlation between the number of reviews and perceived trust, underscoring the significance of social validation. Achieving a superhost status emerges as a key factor, positively impacting perceived trust. The responsiveness of hosts, both in terms of response time and rate, is found to significantly influence perceived trust. Additionally, the number of verifications, positive self-descriptions, and facial expressions in photos all contribute to building trust.

Moving to Zhang, Yan, and Zhang [67], the focus shifts to the content of host self-descriptions. Readability and perspective-taking in self-descriptions are identified as significant contributors to trust perception. Sentiment intensity exhibits an inverted U-shaped relationship with trust perception, highlighting the nuanced interplay between positive sentiments and trust. Notional word count in self-descriptions is positively associated with trust perception. Beyond perception, trust is shown to have a positive relationship with purchase behaviour.

Both Zhang, Yan, and Zhang [66] and Zhang, Yan, and Zhang [67] pinpoint aspects within hosts' descriptions that can either positively or negatively shape trust perception. Both underscore that details concerning the host's profession, personality, and hobbies carry a significant negative impact on trust perception. Zhang, Yan, and Zhang [67] further notes that information about the host's origin also contributes to this unfavourable effect. Conversely, Zhang, Yan, and Zhang [66] and Zhang, Yan, and Zhang [67] highlight that information related to communication, service, and local expertise positively influences trust perception. Moreover, Zhang, Yan, and Zhang [67] accentuates that insights into the host's family and travel experiences serve to improve trust perception.

Alsamani [68] examines the link between trust and purchase behaviour. Verified status, response rate, and superhost badge are positively associated with purchase behaviour, underscoring the role of trust in influencing consumer decisions. Interestingly, the type of listing, host response time, and pricing also play crucial roles. Listings offering an entire room receive more purchases than those for private and shared rooms. Stricter cancellation policies positively influence purchase behaviour, while the maximum number of customers accommodated is associated with higher purchase behaviour. Hosts requiring guest identification verification experience a higher rate of successful transaction completion, highlighting the importance of identity verification.

2.3.2. Methods identified in the review

The present investigation identified a diverse range of methods that are used to achieve the objectives of the articles covering the four most frequently addressed themes in the final sample, including Regression, Deep Learning, Classical Supervised Machine Learning, Unsupervised Machine Learning, Statistical Analysis, and Natural Language Processing (NLP).

Examples of Regression methods include Ordinary Least Squares, Multiple Linear Regression, and Negative Binomial Regression, among others. Deep Learning techniques cited include Long Short-Term Memory (LSTM), Back Propagation Neural Networks (BPNN), and General Regression Neural Networks (GRNN), among others. Classical Supervised Machine Learning Methods comprise Logistic Regression, Bayesian Methods, Support Vector Machine, Decision Trees, and Ensemble Methods. Unsupervised Methods include Topic Modelling and Clustering. Examples of Statistical Analysis methods cited include ANOVA Analysis and CONCOR Analysis, among others. NLP techniques mentioned are Sentiment Analysis and Term Frequency-Inverse Document Frequency (TF-IDF), among others.

Studies related to the theme of customer experience predominantly employ unsupervised machine learning methods, as well as natural language processing and statistical analysis, to draw their conclusions, which aligns with their focus on analyzing customer reviews that are posted on shared accommodation platforms. In contrast, the articles on customer purchase behavior, as well as those concerning accommodation pricing, primarily utilize regression methods to derive their conclusions, a preference that is understandable given that these studies often involve models in which the dependent variable is numerical. Finally, with respect to the articles addressing trust within the context of the accommodation sharing economy, there is no predominant method observed, which can be explained by the more comprehensive and diverse nature of the studies included in this theme, requiring varied analytical approaches to address the multifaceted aspects of trust (Table 2.5).

The systematic literature review reveals a significant gap in research on the impacts of the Covid-19 pandemic on the shared accommodations sector. Among the sixty-eight articles analyzed, only three specifically address this topic. This scarcity of research limits understanding of the pandemic's true effects on this sector, making it challenging to identify the nuances and specifics that emerge during a global crisis. Consequently, managers and business owners lack the information needed to develop effective strategies for mitigating negative impacts and seizing emerging opportunities. Without a thorough analysis, the shared accommodations sector remains vulnerable to future crises, as essential lessons and adaptations are not adequately documented and understood. This research aims to address this gap by providing valuable insights for hosts of shared accommodations.

Table 2.5. Identified analysis methods

Theme / Method	Nr. of articles	Articles
Customer purchase behaviour		
Regression	9	[14], [46], [47], [48], [49], [51], [53], [54], [56]
Deep Learning	1	[54]
Classical Supervised Machine Learning	3	[52], [54], [55]
Unsupervised Machine Learning	2	[46], [47]
Statistical Analysis	3	[53], [54], [57]
Natural Language Processing	0	
Customer experience		
Regression	4	[27],[36], [39], [43]
Deep Learning	0	
Classical Supervised Machine Learning	2	[33], [38]
Unsupervised Machine Learning	7	[25], [30], [31], [32], [33], [34], [41]
Statistical Analysis	6	[27], [36], [38], [39], [41], [43]
Natural Language Processing	7	[31], [32], [34], [38], [40], [42], [44]
Price		
Regression	5	[13], [58], [60], [63], [64]
Deep Learning	2	[62], [64]
Classical Supervised Machine Learning	3	[13], [58], [64]
Unsupervised Machine Learning	1	[62]
Statistical Analysis	1	[65]
Natural Language Processing	0	
Trust		
Regression	2	[66], [67]
Deep Learning	1	[66]
Classical Supervised Machine Learning	1	[66]
Unsupervised Machine Learning	1	[69]
Statistical Analysis	1	[68]
Natural Language Processing	1	[69]
Total		
Regression	20	
Deep Learning	4	
Classical Supervised Machine Learning	9	
Unsupervised Machine Learning	11	
Statistical Analysis	12	
Natural Language Processing	8	

CHAPTER 3

Data

The data for this study were collected for the years 2018 and 2023, which correspond to the years preceding and following the pandemic period, respectively, and were sourced from Inside Airbnb (http://insideairbnb.com), a platform that provides a range of datasets on Airbnb operations. These datasets cover four primary areas:

- Listings: Contains detailed information about available accommodations, including features such as property type, price, amenities, capacity, and more;
- Reviews: Includes data on reviews for accommodations, such as the accommodation ID, review ID and date, reviewer ID and name, and the review text;
- Calendar: Records daily data on price, availability, and minimum and maximum stay requirements for each accommodation;
- Neighbourhoods: Stores information about accommodation locations, linking neighborhoods to municipalities.

Among the available datasets, the one related to listings was chosen for analysis in this study due to its comprehensive range of data on various aspects of accommodations, providing the necessary foundation for the research.

3.1. Data understanding and preparation

In terms of record counts, the 2018 dataset includes 19889 records, while the 2023 dataset has 20097 records. Since each record represents a unique accommodation in the Lisbon district, it can be concluded that the number of accommodations listed on Airbnb in the Lisbon region in 2023 was 1.05% higher compared to 2018.

A preliminary analysis of the variables in the datasets for 2018 and 2023, aided by the data dictionary provided by Inside Airbnb, revealed a change in the structure of the datasets between the two years. The 2018 dataset consists of ninety-six variables, whereas the 2023 dataset contains seventy-five variables. Despite the reduction in the number of variables from 2018 to 2023, the information provided remains broadly similar. This reorganization has reduced redundancy, making the dataset more concise and user-friendly. The datasets cover eight aspects of the accommodations:

(1) General accommodation information;
(2) Host information;
(3) Accommodation location details;
(4) Physical characteristics of the accommodation;
(5) Pricing information;
(6) Reservation policies;

(7) Reviews and ratings;
(8) Legal requirements.

The variables included in each of the mentioned aspects are detailed in Table B.1 of Appendix B for both years under analysis.

Among the various aspects included in the listings dataset, two stand out as particularly relevant to the research questions and objectives defined in the scope of this investigation: accommodation location data and physical characteristics data. For location data, the variable '`neighbourhood_group_cleansed`' (henceforth referred to as '16_regions') was selected, which indicates the municipality of each accommodation and provides the appropriate level of granularity for the analysis. Regarding physical characteristics, the variable 'amenities' was chosen, which enlists the features expected to show more significant changes due to the pandemic, in contrast to attributes like the number of bedrooms, bathrooms, or accommodation capacity, that are intrinsic to the property and less influenced by external factors.

3.1.1. Location

The distribution of the total number of shared accommodations across various geographic locations defined by the '16_regions' variable is shown in Table 3.1 for the years under analysis. It can be observed that, in both 2018 and 2023, Lisbon has the highest number of shared accommodations, while Arruda dos Vinhos has the fewest available. Between 2018 and 2023, several municipalities experienced increases in the number of available shared accommodations. The most notable percentage increase was around 116%, observed in Vila Franca de Xira, where the number of accommodations rose from 25 to 54. Arruda dos Vinhos also saw its number of accommodations double over the period. In Alenquer, the growth rate was approximately 78%, with the number of accommodations increasing from 40 to 71. Loures also recorded a significant increase, from 120 to 197 accommodations, representing a growth rate of about 64%. In Amadora, the number of accommodations grew from 110 to 154, a rise of approximately 40%. Another municipality with a notable increase was Lourinhã, where the number of accommodations grew from 311 to 409, an increase of around 31.5%. The municipalities with the smallest increases in the number of accommodations were Sintra, Odivelas, Torres Vedras, and Cascais, with growth rates of 9%, 7%, 4%, and 2%, respectively.

On the other hand, some municipalities experienced decreases in the number of accommodations. Azambuja saw the largest percentage drop, from 26 to 13 accommodations, a reduction of 50%. Oeiras experienced a small decrease, with the number of accommodations falling from 356 to 342, a reduction of 3.9%. Mafra also had a slight decline, with the number of accommodations decreasing from 1213 to 1187, a drop of 2.1%. In Lisbon, there was a modest reduction in accommodations, decreasing from 14241 in 2018 to 14052 in 2023, which represents a decrease of approximately 1.3%.

The municipalities of Cadaval and Sobral de Monte Agraço maintained a constant number of accommodations between 2018 and 2023, with 52 and 19 accommodations, respectively.

The analysis of the data in Table 3.1 reveals a diverse landscape of the shared accommodations market in the Lisbon district. This suggests the need for tailored policies for each municipality to promote balanced growth in tourism and accommodation availability.

TABLE 3.1. Distribution of accommodations across the regions defined by '16_regions'

Region	**Number of accommodations**	
	2018	**2023**
Lisboa	14241	14052
Cascais	1881	1924
Mafra	1213	1187
Sintra	1177	1285
Oeiras	356	342
Lourinhã	311	409
Torres Vedras	240	250
Loures	120	197
Amadora	110	154
Odivelas	73	78
Cadaval	52	52
Alenquer	40	71
Azambuja	26	13
Vila Franca de Xira	25	54
Sobral de Monte Agraço	19	19
Arruda dos Vinhos	5	10

Considering the geographical location of accommodations is crucial to address the objectives of this study. However, utilizing the '16_regions' variable is impractical, as it would lead to the construction of a classification model with 16 classes, and this level of detail may be excessively granular, potentially resulting in overfitting or an overly complex model that does not effectively capture meaningful patterns in the data. Therefore, a dictionary was created to group neighborhoods based on that field, establishing two main categories: Lisbon and outside Lisbon. Beyond these broad categories, a more detailed segmentation was implemented, grouping neighbourhoods into six distinct classes (Table 3.2). As a result of this segmentation, two new columns were added to the datasets: the '6_regions' column, which contains the detailed classification into six classes, and the

‘2_regions’ column, a boolean indicator marking whether the accommodation is located in Lisbon. This enhancement of the datasets allows for a more in-depth analysis of spatial variables and their relationship with the amenities of the accommodations.

Table 3.2. Neighbourhood grouping into six classes

‘16-regions’	**‘6-regions’**
Lisboa	Lisbon
Amadora	Metropolitan Lisbon
Odivelas	
Sintra	Lisbon Coast
Cascais	
Oeiras	
Mafra	Metropolitan West
Loures	
Vila Franca de Xira	
Arruda dos Vinhos	Rural West
Torres Vedras	
Sobral de Monte Agraço	
Alenquer	Peripheral Rural West
Azambuja	
Cadaval	
Lourinhã	

The distribution of the total number of accommodations across the various geographic zones defined by ‘6_regions’ is illustrated in Table 3.3.

Table 3.3. Distribution of accommodations across the regions defined by ‘6_regions’

‘6_regions’	**Number of accommodations**	
	2018	**2023**
Lisbon	14241	14052
Metropolitan Lisbon	183	232
Lisbon Coast	3414	3551
Metropolitan West	1358	1438
Rural West	264	279
Peripheral Rural West	429	545

3.1.2. Amenities

To understand the impact of the COVID-19 pandemic on how hosts of shared accommodations offer amenities in their properties, the first step was to identify the unique amenities available in both 2018 and 2023. This task involved using the "amenities" field from each dataset, which lists all amenities for each accommodation, as strings or lists. Initially, 121 unique amenities were identified for 2018, and 3010 unique amenities for 2023, which highlights a drastic shift rather than a trivial fluctuation. A closer examination revealed that the apparent discrepancy in the number of amenities between the two periods is exacerbated by changes in how amenities were recorded. Specifically, the amenities data for 2023 show a much higher level of detail compared to 2018. In 2023, hosts not only indicate the presence or absence of an amenity but also provide additional information that characterizes these amenities, allowing guests to make more informed choices based on their needs and preferences.

The inclusion of additional information about amenities is a key factor in the substantial (and misleading) increase in the number of unique amenities identified between 2018 and 2023. The same amenity is counted multiple times if it has different characteristics. To address this, it was necessary to aggregate all records for a given amenity in the 2023 dataset, regardless of its characteristics, to facilitate the initial task. Each grouped record for the same amenity was manually assigned the name of the most generic record (the one without additional characteristics), referred to as the "normalized label".

Table 3.4 illustrates the aggregation performed for three distinct normalized labels: “Coffee maker”, “Shampoo”, and “TV”.

Through the aggregation process, it became evident that different characteristics are recorded depending on the amenity. For the normalized label “Coffee maker”, the following characteristics are recorded: category (e.g., “expresso machine”, “drip coffee”, “pour-over coffee”, “french press”) and brand (e.g., Keurig and/or Nespresso). For the normalized label “Shampoo”, only the brand of the shampoo available in the accommodation is recorded, with between 30 and 40 brands available across all accommodations. Lastly, for the normalized label “TV”, characteristics such as the type of television (e.g., TV or HDTV), cable type (e.g., standard and/or premium), screen size (in inches), and available streaming services (e.g., Netflix, Amazon Prime Video, HBO Max, Apple TV, Disney+, Chromecast, FireTV, Hulu, and/or Roku) are recorded. Detailed information about the characteristics present in each standardized label can be found in Table C.1 of Appendix C.

To ensure consistency between the 2018 and 2023 datasets, the manual assignment of normalized labels was also applied to the 2018 dataset.

With all records for the same amenity properly aggregated and renamed, it was then possible to determine the correct number of normalized labels in the datasets. A total of 100 normalized labels were identified for 2018, and 128 normalized labels were identified for 2023. Since the initial task involves comparing data from two distinct periods, it was

Table 3.4. Examples of amenity groupings

Amenity	**Normalized label**
Coffeee maker: Nespresso	Coffee maker
Coffee maker: espresso machine	
Coffee maker: drip coffee maker, Keurig coffee machine	
Coffee maker: drip coffee maker, espresso machine, french press, Nespresso	
Coffee maker: drip coffee maker, espresso machine, Keurig coffee machine, pour-over coffee	
Dove shampoo	Shampoo
Lousani shampoo	
Aroma de Portugal shampoo	
Rituals shampoo	
Castelbel shampoo	
TV with standard cable	TV
HDTV with Netflix	
HDTV with Amazon Prime Video, Disney+, HBO Max, Netflix, premium cable	
24" HDTV with Amazon Prime Video, Apple TV, Disney+, Netflix, standard cable	
32" HDTV	

necessary to check if all amenities recorded in 2018 were also present in 2023 (and vice versa). It was determined that between 2018 and 2023, 32 normalized labels were no longer recorded, 60 new normalized labels were added, and 68 amenities were common to both years. The set of normalized labels common to 2018 and 2023, listed in Table 3.5, primarily includes essential amenities that are commonly expected by guests across most accommodations. Some of these amenities for comfort and convenience include air conditioning, hair dryer, washing and drying machines, and heating. Family-oriented amenities offered include crib, high chair, children's toys, dining utensils for children, and bathtub. The list also covers kitchen items such as coffee maker, microwave, stove, and basic utensils. Safety amenities include fire extinguisher, first aid kit, fireplace guards, and carbon monoxide detector. Additionally, amenities for entertainment and relaxation such as TV, fireplace, pool, and jacuzzi are included. The set of normalized labels also features amenities like wired internet, electric vehicle charger, and private entrance, reflecting a balance between comfort, safety, and convenience.

Table 3.5. Normalized labels common to 2018 and 2023

Normalized label	2018	2023	Normalized label	2018	2023
Access	574	1314	Host greets you	6269	5919
Air conditioning	4682	7927	Hot tub	491	371
Baby bath	803	1222	Hot water	9065	15583
Baby monitor	101	96	Indoor fireplace	2144	1933
Babysitter recommendations	780	1109	Iron	13679	15555
Bathtub	1752	3954	Keypad	439	1330
BBQ grill	806	2224	Kitchen	18622	18624
Beach essentials	555	778	Lock on bedroom door	5701	3141
Bed linens	5986	13723	Lockbox	421	3224
Breakfast	1511	795	Long term stays allowed	4958	9739
Carbon monoxide alarm	2553	4756	Luggage dropoff allowed	2921	5119
Changing table	185	242	Microwave	5897	13798
Children's books and toys	1285	1781	Outlet covers	396	530
Children's dinnerware	772	1551	Oven	4969	11473
Coffee maker	5640	14108	Pack 'n play/travel crib	2526	4925
Cooking basics	5649	14500	Patio or balcony	2682	8205
Crib	2758	5995	Pets allowed	3170	2835
Dishes and silverware	6321	15670	Pool	1802	2140
Dishwasher	3457	8377	Private entrance	2984	5330
Dryer	4652	3022	Private living room	1039	409
Elevator	4415	4326	Refrigerator	6391	15314
Essentials	18547	18204	Room-darkening shades	1616	6420
Ethernet connection	593	2384	Self check-in	1404	6287
EV charger	59	282	Shampoo	10263	11132
Extra pillows and blankets	3362	7090	Single level home	787	2119
Fire extinguisher	12251	15216	Smart lock	196	1150
Fireplace guards	172	302	Smoking allowed	3447	1805
First aid kit	12472	15310	Stove	6044	13462
Game console	147	182	Table corner guards	112	123
Gym	228	427	TV	16164	16853
Hair dryer	15410	16834	Washer	13922	9862
Hangers	15410	16133	Waterfront	374	1064
Heating	9605	10569	Wifi	18950	19635
High chair	2644	4784	Window guards	177	338

On the other hand, the set of amenities present in 2018 but not found in 2023, as shown in Table 3.6, is primarily related to accessibility. This set includes normalized labels such as adapted beds and bathrooms, showers with grab bars, step-free entrances, and wheelchair access. It also features amenities like 24-hour check-in, doorman services, and wireless intercom systems, which offer a more convenient and secure arrival and departure experience. Additionally, it includes amenities such as air purifiers and internet access, as well as items for events and meetings, such as event suitability and well-lit spaces. These items reflect a focus on accessibility, security, and convenience, with less emphasis on details and the personalization of the guest experience.

TABLE 3.6. Normalized labels unique to 2018

Normalized label	**2018**	**Normalized label**	**2018**
24-hour check-in	2827	Other	425
Accessible-height bed	1109	Pets live on this property	746
Accessible-height toilet	951	Private bathroom	6
Air purifier	2	Roll-in shower	54
Buzzer/wireless intercom	3858	Safety card	3630
Cleaning before checkout	973	Stair gates	162
Doorman	492	Step-free access	1817
Family/kid friendly	13904	Suitable for events	1225
Firm mattress	80	Toilet	400
Fixed grab bars for shower	2	Well-lit path to entrance	1421
Flat path to front door	672	Wheelchair accessible	1125
Front desk/doorperson	360	Wide clearance to bed	882
Ground floor access	15	Wide clearance to shower	400
Handheld shower head	1076	Wide doorway	1292
Internet	7831	Wide entryway	859
Laptop friendly workspace	11010	Wide hallway clearance	937

In contrast, 2023 introduced a broad range of new amenities that reflect a trend towards greater personalization and comfort, as demonstrated in Table 3.7. This includes child safety items such as baby gates, and leisure and comfort amenities like outdoor spaces, barbecue equipment, and board games. The list also features more advanced kitchen equipment, such as blenders, bread machines, and large refrigerators. Additionally, it includes dedicated spaces like work areas, as well as extra amenities such as cleaning products, outdoor furniture, free parking, and equipment for outdoor activities like kayaking and hammocks. Entertainment items such as pianos, ping-pong tables, pool tables, and sound systems are also included. These amenities reflect a heightened focus on guest comfort, safety, and personalization, with a more diverse and tailored range of offerings.

To ensure uniformity and facilitate the analysis of the various normalized labels present in accommodations, a dictionary was created to map each amenity to its corresponding normalized label. This dictionary enables data standardization and ensures that different terms for the same feature are correctly grouped together.

Table 3.7. Normalized labels exclusive to 2023

Normalized label	**2023**	**Normalized label**	**2023**
Baby safety gates	191	Kettle	6931
Backyard	2036	Laundromat nearby	3868
Baking sheet	1588	Microwave/oven	2
Barbecue utensils	920	Mini fridge	1211
Bidet	2381	Mosquito net	239
Bikes	223	Outdoor dining area	3053
Blender	1334	Outdoor furniture	3086
Board games	677	Outdoor shower	662
Boat slip	14	Parking	12398
Body soap	4181	Piano	94
Books and reading material	1841	Ping pong table	107
Bread maker	132	Pool table	104
Building staff	585	Portable fans	2759
Ceiling fan	255	Portable heater	1459
Cleaning available during stay	2153	Record player	93
Cleaning products	5027	Rice maker	102
Clothing storage	6028	Safe	1908
Coffee	3375	Sauna	53
Conditioner	1073	Security cameras on property	1686
Dedicated workspace	6553	Shampoo/shower gel	32
Dining table	6182	Shower gel	6376
Drying rack for clothing	4529	Ski-in/ski-out	8
Exercise equipment	358	Smoke alarm	6906
Fire pit	230	Sound system	783
Free carport on premises	13	Sun loungers	323
Free residential garage on premises	106	Toaster	6109
Freezer	6973	Trash compactor	337
Hammock	438	View	4048
Ironing board	4	Window AC unit	254
Kayak	11	Wine glasses	5921

Based on the normalized labels defined in the dictionary, boolean columns were added to the 2018 and 2023 datasets. Each column represents a normalized label and indicates the presence or absence of the corresponding feature for each accommodation. This approach simplifies subsequent analysis, allowing for quick and clear identification of each amenity's availability across various accommodations.

To avoid including data that might be unrepresentative and potentially noisy, all normalized labels present in fewer than 150 accommodations were removed. This filtering ensures that only the most common and relevant normalized labels are considered, enhancing the robustness and reliability of the conclusions drawn.

CHAPTER 4

Knowledge Extraction

With the necessary data properly prepared, data models were constructed to assess whether it is possible to distinguish the different regions of the Lisbon district using the normalized labels defined in Section 3.1.2. Two classification models were developed: one for two geographic regions and another for six geographic regions.

The first model is a binary classification model with the target variable '2_regions', which has two classes. This model represents the initial phase of data modeling, distinguishing between the municipality of Lisbon and the other municipalities in the Lisbon district. For a more detailed and specific analysis of each municipality within the Lisbon district, a second model was created, using the target variable '6_regions', which has six classes, resulting in a multiclass model. It is important to note that both models were applied to data from both years under analysis, resulting in a total of four models.

For this study, several classification algorithms were applied to the data, including k-Nearest Neighbors (kNN), Decision Tree, Random Forest, Logistic Regression, Linear SVM, RBF SVM, Perceptron, and MLP (Multi-Layer Perceptron). Although the primary goal is to identify the normalized labels that best characterize each municipality in the Lisbon district, algorithms that do not estimate class probabilities were also included to provide a comprehensive analysis.

Four metrics were selected to evaluate the models: Accuracy, Precision, Recall, and F1-score. Additionally, Precision, Recall, and F1-score were assessed in two different contexts — weighted and unweighted classes —, taking into account the unequal distribution of accommodations across the defined geographic zones.

4.1. Distinguishing between two regions: shift between 2018 and 2023

4.1.1. Comparing classifiers

The results from applying the classification algorithms to the data model targeting '2_regions' for the years 2018 and 2023 are presented in Tables 4.1 and 4.2, respectively.

Although there are differences in the performance of the various classifiers applied to the data models, all algorithms were effective in distinguishing between the two regions of the Lisbon district (Lisbon and Outside Lisbon) using normalized labels, both in 2018 and 2023. This suggests that the classification problem was handled adequately by all methods, regardless of the year or classifier used. However, certain classifiers stood out compared to others.

In 2018, RBF SVM and Logistic Regression proved to be the most effective, achieving the best results in terms of accuracy and F1-score. These models not only managed to

Table 4.1. Binary classification performance on the 2018 dataset

Classifier	**Acc**	**Macro**			**Weighted**		
		Prec	**Rec**	**F1**	**Prec**	**Rec**	**F1**
kNN (k=3)	0.817	0.805	0.718	0.742	0.813	0.817	0.802
Decision Tree	0.791	0.743	0.750	0.746	0.794	0.791	0.792
Random Forest	0.847	0.821	0.791	0.803	0.843	0.847	0.844
Logistic Regression	**0.848**	0.841	0.769	**0.793**	0.846	0.848	**0.839**
Linear SVM	0.829	0.825	0.733	0.759	0.828	0.829	0.815
RBF SVM	**0.852**	0.857	0.766	**0.795**	0.854	0.852	**0.842**
Perceptron	0.804	0.770	0.718	0.736	0.795	0.804	0.794
MLP (1 hidden)	0.839	0.811	0.777	0.791	0.834	0.839	0.834
MLP (2 hidden)	0.832	0.799	0.772	0.783	0.827	0.832	0.828

Table 4.2. Binary classification performance on the 2023 dataset

Classifier	**Acc**	**Macro**			**Weighted**		
		Prec	**Rec**	**F1**	**Prec**	**Rec**	**F1**
kNN (k=3)	0.835	0.836	0.754	0.778	0.835	0.835	0.824
Decision Tree	0.813	0.776	0.782	0.779	0.815	0.813	0.814
Random Forest	**0.865**	0.846	0.823	**0.833**	0.862	0.865	**0.862**
Logistic Regression	0.858	0.853	0.796	0.817	0.857	0.858	0.852
Linear SVM	0.842	0.828	0.779	0.797	0.838	0.842	0.835
RBF SVM	**0.867**	0.874	0.801	**0.826**	0.869	0.867	**0.860**
Perceptron	0.794	0.756	0.735	0.744	0.788	0.794	0.789
MLP (1 hidden)	0.849	0.828	0.800	0.812	0.845	0.849	0.845
MLP (2 hidden)	0.841	0.813	0.798	0.805	0.838	0.841	0.839

separate the regions more precisely but also demonstrated greater consistency across the evaluated metrics, indicating their robustness for this type of problem.

Similarly, in 2023, Random Forest and RBF SVM emerged as the top performers. Random Forest achieved an accuracy of 0.865 and an F1-score of 0.833, while RBF SVM slightly outperformed it with an accuracy of 0.867 and an F1-score of 0.826. Both models displayed precision and consistency across all metrics, confirming their status as the most robust options for the 2023 classification task.

It is also evident that between 2018 and 2023, there was an overall improvement in metric values, which may be related to various factors. Notably, the increased number of records in 2023 provided a more comprehensive dataset for training the models, enhancing their ability to learn from diverse examples. Additionally, the distribution of classes became more balanced, which often alleviates the challenges posed by imbalanced datasets and allows classifiers to perform better across both majority and minority classes. Furthermore, the introduction of more independent variables in the 2023 model potentially contributed to capturing more nuanced patterns in the data, ultimately leading to higher accuracy and F1-scores. These factors collectively underscore the significance of dataset quality and composition in achieving robust classification results.

While classifiers like RBF SVM and Random Forest have demonstrated superior performance in terms of accuracy and consistency, Logistic Regression offers a different approach that emphasizes interpretability and probability distribution. This probabilistic model, which operates by maximizing entropy, allows for the application of the softmax function to generate probabilities for each class based on the given features. This capability enables the identification of the most probable classes, providing a nuanced understanding of model predictions. Therefore, the remainder of this section will utilize Logistic Regression to derive the probability distribution of the classes, which is informed by the relevant features.

4.1.2. The best classifier – Logistic Regression

Table 4.3 specifies the performance of the Logistic Regression algorithm in classifying the two classes defined by the variable ‘2_regions’ — "Outside Lisbon" and "Lisbon" — for the year of 2018. The algorithm exhibits good precision in identifying instances of the "Outside Lisbon" class; however, it struggles to capture all actual cases, resulting in lower recall and, consequently, a reduced F1-score for this class. In contrast, for the "Lisbon" class, Logistic Regression demonstrates notably stronger performance. The metrics indicate that the model not only maintains high precision in its predictions but also excels in identifying all positive instances, leading to a significantly higher F1-score. Specifically, the F1-score for "Lisbon" is markedly superior to that for "Outside Lisbon", suggesting that the model is more reliable in identifying instances of Lisbon.

TABLE 4.3. Logistic Regression performance for two classes, using 2018 data

Class	**Logistic Regression**		
	Prec	**Rec**	**F1**
Outside Lisbon	0.828	0.587	0.687
Lisbon	0.853	0.952	0.900

The confusion matrix for the Logistic Regression algorithm, presented in Table 4.4, offers a detailed breakdown of the model’s predictions and corroborates the findings from

Table 4.3. It reveals that for the "Outside Lisbon" class, the model accurately identified 2647 instances but misclassified 1866 instances as "Lisbon", which contributes to the lower recall and F1-score noted previously. Conversely, for the "Lisbon" class, the model correctly classified 10852 instances, while misclassifying only 546 as "Outside Lisbon". The high number of true positives aligns with the strong precision and recall metrics for the "Lisbon" class, further emphasizing the model's effectiveness in accurately identifying instances of this class.

Table 4.4. Logistic Regression confusion matrix for two classes, using 2018 data

		Predicted	
		Outside Lisbon	Lisbon
Actual	Outside Lisbon	**2647**	1866
	Lisbon	548	**10850**

Moving to the 2023 model, the Logistic Regression algorithm performed similarly, as shown in Table 4.5. For the "Outside Lisbon" class, there is a slight improvement in precision and F1-score. However, the recall for this class remains somewhat low, indicating that the model still struggles to capture all instances of "Outside Lisbon". For the "Lisbon" class, the performance remains robust, with precision, recall, and F1-score maintaining high values.

The confusion matrix for the 2023 Logistic Regression model, presented in Table 4.6, provides further clarity on these metrics. For "Outside Lisbon", the model correctly identified 3084 instances but misclassified 1709 as "Lisbon", which again contributes to the lower recall for this class. Meanwhile, for "Lisbon", the model correctly classified 10709 instances while misclassifying only 575 as "Outside Lisbon", reaffirming the high performance for the "Lisbon" class.

Overall, while the Logistic Regression algorithm continues to perform well for the "Lisbon" class in both 2018 and 2023, its ability to correctly classify instances of "Outside Lisbon" remains a challenge.

Table 4.5. Logistic Regression performance for two classes, using 2023 data

Class	**Logistic Regression**		
	Prec	**Rec**	**F1**
Outside Lisbon	0.843	0.643	0.730
Lisbon	0.862	0.949	0.904

4.1.3. Most important amenities

The Tables 4.7 and 4.8 present the normalized labels with the highest probabilities of belonging to the classes defined by the variable 2_regions for the years 2018 and 2023.

Table 4.6. Logistic Regression confusion matrix for two classes, using 2023 data

		Predicted	
		Outside Lisbon	Lisbon
Actual	Outside Lisbon	**3084**	1709
	Lisbon	575	**10709**

Given that this is a two-class model, the labels with the highest probability for one class inherently have lower probabilities for the other.

In 2018, six of the top ten labels were associated with the "Outside Lisbon" class, while four corresponded to "Lisbon".

Features such as "Indoor fireplace", "Pool", "Access", "Parking", "BBQ grill", and "Backyard" exhibited high probabilities (ranging from 0.697 to 0.919) for accommodations outside Lisbon. These labels are more prevalent in suburban and rural areas, where larger spaces allow for outdoor features and leisure facilities. For instance, pools and BBQ grills cater to guests seeking outdoor experiences, which are common in accommodations with more land. In contrast, these features showed lower probabilities in Lisbon (between 0.081 and 0.303), reflecting the urban density and limited outdoor space typical of city accommodations.

Conversely, labels such as "Changing table", "Waterfront" and "Wifi" had higher probabilities of occurring in Lisbon, ranging from 0.751 to 0.816. Features like "Changing table" indicate family-friendly accommodations, which are often found in urban settings with a higher concentration of families. The "Waterfront" designation suggests proximity to scenic areas along the Tagus River, enhancing the appeal of urban listings for both residents and visitors. Additionally, "Wifi" reflects the demand for connectivity in a bustling urban environment, where modern amenities are essential for guests.

In 2023, these trends continued, with seven labels showing higher probabilities for the "Outside Lisbon" class and only three for "Lisbon".

For the "Outside Lisbon" class, characteristics such as "Access", "Pool", and "Indoor fireplace" stand out with the highest probabilities, ranging between 0.893 and 0.916. These features indicate a focus on amenities that provide comfort and convenience, typically associated with accommodations located in less urbanized areas. The high probability of "Access" suggests that properties outside Lisbon often offer direct access to beaches, lakes, or resorts due to their more rural locations. The presence of "Pool" (0.907), "Parking" (0.809), and "BBQ grill" (0.778) reinforces the idea that these accommodations provide various amenities aimed at leisure and outdoor use, which is typical of properties situated in more spacious and open areas. Additionally, "Indoor fireplace" (0.893) is also a notable feature, highlighting its role in offering extra comfort to guests and reflecting a concern for guest well-being in rural environments.

TABLE 4.7. Normalized label probabilities for two classes in 2018 data

Normalized label	**Outside Lisbon**	**Lisbon**
Indoor fireplace	**0.919**	0.081
Pool	**0.913**	0.087
Access	**0.880**	0.120
Parking	**0.822**	0.178
Changing table	0.184	**0.816**
Waterfront	0.224	**0.776**
Wifi	0.249	**0.751**
BBQ grill	**0.708**	0.292
Backyard	**0.697**	0.303
Other	0.309	**0.691**

Other significant features outside Lisbon included "Beach essentials" (0.706), indicating that accommodations in these areas frequently provide items for coastal activities, suggesting a focus on outdoor leisure opportunities. The prominence of "Game console" (0.677) further emphasizes the availability of modern entertainment options, catering to guests seeking contemporary amenities in more spacious rural settings.

In Lisbon, the most prominent characteristics are "Kitchen", "Portable fans", and "Window AC unit". The high probability of "Kitchen" (0.714) suggests that having a well-equipped kitchen is an important factor for urban accommodations. "Portable fans" (0.699) and "Window AC unit" (0.676) indicate that providing practical climate control solutions is a priority for hosts in this region, where space may be more limited.

Generally, the distribution of probabilities between "Outside Lisbon" and "Lisbon" underscores the distinct characteristics of urban versus rural accommodations in the sharing economy.

4.2. Distinguishing between six regions: shift between 2018 and 2023

4.2.1. Comparing classifiers

Table 4.9 presents the results of various classification algorithms applied to the model with the target variable '6_regions' for the year of 2018, while Table 4.10 shows the corresponding results for the year of 2023.

In both years, Random Forest and RBF SVM emerge as the standout classifiers, achieving similarly high levels of accuracy. Their strong performance underscores their effectiveness in handling the classification task across multiple regions, demonstrating their ability to manage the varying complexities of the six classes.

A comparison of the macro and weighted metrics reveals that, in 2018 and 2023 alike, the classifiers perform better in the weighted metrics, which take into account the

TABLE 4.8. Normalized label probabilities for two classes in 2023 data

Normalized label	Outside Lisbon	Lisbon
Access	**0.916**	0.084
Pool	**0.907**	0.093
Indoor fireplace	**0.893**	0.107
Parking	**0.809**	0.191
BBQ grill	**0.778**	0.222
Kitchen	0.286	**0.714**
Beach essentials	**0.706**	0.294
Portable fans	0.301	**0.699**
Game console	**0.677**	0.323
Window AC unit	0.324	**0.676**

class distribution. This pattern highlights that models are more proficient at classifying the majority classes, as the weighted metrics are influenced by the proportion of each class in the dataset. In contrast, the lower macro F1-scores in 2018 and 2023 suggest that classifiers face difficulties when dealing with less-represented or minority classes. This challenge is a common issue in multi-class classification, where models may excel at handling the dominant classes but struggle with the minority ones, which can obscure performance issues in imbalanced datasets.

Despite slight variations in performance across algorithms from year to year, the overall trends remain consistent, with Random Forest and RBF SVM consistently demonstrating their reliability.

Given the similarity of results among the classifiers, Logistic Regression was adopted to study the probability distribution of each class, as was done previously in Section 4.1.

4.2.2. The best classifier – Logistic Regression

Table 4.11 highlights key differences in the performance of the Logistic Regression model across six geographical classes in 2018. The model performs exceptionally well for the "Lisbon" class, demonstrating strong predictive accuracy and balance in identifying true positives. However, it struggles significantly with other classes, particularly "Metropolitan Lisbon", "Rural West", and "Peripheral Rural West", where the model fails to effectively detect instances. These classes show poor recall, indicating that many true cases are missed, and overall weak performance.

The confusion matrix (Table 4.12) further illustrates the challenges faced by the model. While it accurately identifies instances of the "Lisbon" class, it misclassifies many instances from other classes, especially "Lisbon Coast" and "Metropolitan West". For instance, "Lisbon Coast" demonstrates a significant number of misclassifications, with 1446 instances

Table 4.9. Six-class classification performance on the 2018 dataset

Classifier	Acc	Macro			Weighted		
		Prec	Rec	F1	Prec	Rec	F1
kNN (k=3)	0.769	0.624	0.347	0.404	0.739	0.769	0.739
Decision Tree	0.699	0.356	0.367	0.361	0.711	0.699	0.705
Random Forest	**0.783**	0.594	0.352	**0.403**	0.756	0.783	**0.759**
Logistic Regression	0.780	0.435	0.301	0.330	0.734	0.780	0.746
Linear SVM	0.770	0.419	0.257	0.263	0.724	0.770	0.725
RBF SVM	**0.784**	0.669	0.280	**0.306**	0.763	0.784	**0.743**
Perceptron	0.716	0.311	0.297	0.287	0.698	0.716	0.696
MLP (1 hidden)	0.772	0.381	0.313	0.333	0.731	0.772	0.746
MLP (2 hidden)	0.770	0.356	0.301	0.318	0.725	0.770	0.743

Table 4.10. Six-class classification performance on the 2023 dataset

Classifier	Acc	Macro			Weighted		
		Prec	Rec	F1	Prec	Rec	F1
kNN (k=3)	0.786	0.642	0.380	0.443	0.763	0.786	0.760
Decision Tree	0.720	0.388	0.394	0.391	0.728	0.720	0.724
Random Forest	**0.797**	0.691	0.377	**0.434**	0.778	0.797	**0.774**
Logistic Regression	0.779	0.444	0.316	0.345	0.735	0.779	0.748
Linear SVM	0.770	0.415	0.267	0.266	0.731	0.770	0.730
RBF SVM	**0.793**	0.679	0.301	**0.325**	0.779	0.793	**0.755**
Perceptron	0.727	0.323	0.278	0.283	0.691	0.727	0.702
MLP (1 hidden)	0.770	0.432	0.341	0.366	0.735	0.770	0.748
MLP (2 hidden)	0.771	0.403	0.340	0.357	0.738	0.771	0.751

incorrectly predicted as "Lisbon". Similarly, "Metropolitan West" suffers from misclassifications, with 389 instances misidentified as "Lisbon Coast" and 449 as "Lisbon". The "Rural West" and "Peripheral Rural West" classes also exhibit notable confusion, indicating a pattern where the model often confuses these classes with others.

In contrast, "Metropolitan Lisbon" is particularly problematic; it is rarely identified correctly, with only one instance accurately predicted out of 153 actual occurrences. This highlights a critical weakness in the model's ability to distinguish this class from others.

TABLE 4.11. Logistic Regression performance for six classes, using 2018 data

Class	Logistic Regression		
	Prec	**Rec**	**F1**
Lisbon Coast	0.524	0.386	0.444
Lisbon	0.836	0.971	0.899
Metropolitan Lisbon	0.143	0.007	0.012
Metropolitan West	0.465	0.211	0.290
Rural West	0.167	0.029	0.049
Peripheral Rural West	0.472	0.201	0.282

TABLE 4.12. Logistic Regression confusion matrix for six classes, using 2018 data

Actual Class/ Predicted Class	Lisbon Coast	Lisbon	Metropolitan Lisbon	Metropolitan West	Rural West	Peripheral Rural West
Lisbon Coast	**1040**	1446	2	171	8	28
Lisbon	292	**11067**	4	29	1	5
Metropolitan Lisbon	14	137	**1**	1	0	0
Metropolitan West	389	449	0	**236**	13	31
Rural West	107	51	0	33	**6**	12
Peripheral Rural West	144	81	0	37	8	**68**

Turning to the results for 2023 presented in Table 4.13, the performance of the Logistic Regression model shows some improvements for the "Lisbon" class, maintaining strong predictive accuracy and high recall. However, similar challenges persist for other classes. The model continues to struggle with "Metropolitan Lisbon," achieving a notably low recall, correctly identifying only 2 instances out of 183 actual occurrences. This underscores the ongoing difficulty in differentiating this class from others.

The confusion matrix for 2023 further emphasizes the misclassification issues faced by the model (Table 4.14). While it again accurately identifies a large number of instances for the "Lisbon" class, it misclassifies numerous instances from other classes. For example, "Lisbon Coast" sees 1,270 instances misclassified as "Lisbon," indicating persistent confusion between these two classes. The "Metropolitan West" class continues to struggle, with 439 instances misclassified as "Lisbon Coast" and a substantial number as "Lisbon".

TABLE 4.13. Logistic Regression performance for six classes, using 2023 data

Class	Logistic Regression		
	Prec	**Rec**	**F1**
Lisbon Coast	0.542	0.452	0.493
Lisbon	0.845	0.967	0.902
Metropolitan Lisbon	0.105	0.011	0.019
Metropolitan West	0.438	0.212	0.286
Rural West	0.326	0.065	0.108
Peripheral Rural West	0.408	0.192	0.261

TABLE 4.14. Logistic Regression confusion matrix for six classes, using 2023 data

Actual Class/ Predicted Class	Lisbon Coast	Lisbon	Metropolitan Lisbon	Metropolitan West	Rural West	Peripheral Rural West
Lisbon Coast	**1264**	1270	5	177	15	67
Lisbon	307	**10915**	6	49	0	7
Metropolitan Lisbon	11	170	**2**	4	0	1
Metropolitan West	439	426	5	**244**	8	30
Rural West	108	47	0	31	**14**	17
Peripheral Rural West	201	94	1	52	6	**84**

Overall, the performance of the Logistic Regression model across 2018 and 2023 reveals a consistent pattern. While the model maintains strong predictive accuracy for the "Lisbon" class in both years, significant misclassification issues persist for other classes.

4.2.3. Most important amenities

To complement the analysis conducted in Section 4.1.3, an examination of the features with the highest and lowest probability of belonging to the six-class model was performed. This analysis was conducted in two phases: first, on a class-by-class basis, and then focusing on the features with the highest overall probability.

4.2.3.1 Lisbon Coast

Tables 4.15 and 4.16 present the normalized labels with the highest and lowest probabilities of belonging to the "Lisbon Coast" region, which includes the municipalities of Sintra, Cascais, and Oeiras, for the years 2018 and 2023, respectively.

In 2018, the most prominent features in the "Lisbon Coast" region include "Babysitter recommendations" (0.256), "Smart lock" (0.247), and "Backyard" (0.246), indicating a strong focus on family-oriented accommodations with particular attention to security and outdoor space. The high prevalence of "Babysitter recommendations" suggests that properties cater to families with young children, offering services or advice that make stays more convenient for parents. "Smart lock" highlights a focus on security, with hosts ensuring that guests have access to safe and convenient locking systems. The inclusion of "Backyard" suggests that outdoor space is a priority, likely because the region's suburban and coastal areas provide ample room for properties with gardens or outdoor areas, appealing to families and leisure travelers who value extra space for activities.

On the other hand, features like "Suitable for events" (0.098) and "Front desk/doorperson" (0.093) have much lower probabilities, indicating that accommodations in "Lisbon Coast" are generally more private and independent, with less emphasis on services like event hosting or staffed reception areas. Other features, such as "Stove" (0.092) and "Lockbox" (0.090), while present, are not as prominent, pointing to a more practical approach to amenities that still cover basic needs without being a primary focus. The label with the lowest percentage, "Changing table" (0.073), further emphasizes that while accommodations may cater to families, there is less emphasis on amenities for infants, contrasting with the relatively higher presence of "Babysitter recommendations.

By 2023, there is a noticeable shift in the types of amenities offered in "Lisbon Coast". While some family-oriented and convenience-based features, such as "Babysitter recommendations" (0.282), remain relevant, new trends emerge, reflecting the evolving needs of guests. Notably, the feature "Building staff" (0.355) becomes the most prominent, indicating that accommodations increasingly provide on-site personnel to support and assist guests during their stay. This shift suggests that properties in the region are adapting to offer more comprehensive guest services, catering to travelers seeking higher levels of convenience and support.

The feature "Access" (0.338) also gains importance, likely reflecting the region's coastal geography, with accommodations providing easy access to beaches, lakes, or resorts. This aligns with the region's role as a leisure destination, where guests might prioritize proximity to natural attractions. Additionally, the growing prominence of "Mini fridge" (0.247) highlights a focus on practical amenities that offer convenience during shorter stays or for guests who may not require full kitchen facilities.

A significant new feature in 2023 is the "EV charger" (0.242), signaling the region's adaptation to the rise in electric vehicle usage. This suggests that "Lisbon Coast" is becoming more attuned to sustainability trends, with properties recognizing the need to support guests with eco-friendly infrastructure.

In contrast, features such as "Trash compactor" (0.099) and "Kitchen" (0.084) have lower probabilities in 2023, indicating that certain amenities, particularly those related to waste management and full kitchen facilities, are not as emphasized in this region. The lower focus on "Kitchen" might suggest that guests are increasingly looking for shorter stays or preferring to eat out.

Table 4.15. Normalized label probabilities for "Lisbon Coast" in 2018 data

Normalized label	**Lisbon Coast**	Lisbon	Metropolitan Lisbon	Metropolitan West	Rural West	Peripheral Rural West
Babysitter recommendations	**0.256**	0.228	0.119	0.217	0.122	0.059
Smart lock	**0.247**	0.259	0.119	0.169	0.117	0.089
Backyard	**0.246**	0.092	0.101	0.139	0.237	0.185
(...)	**(...)**	(...)	(...)	(...)	(...)	(...)
Suitable for events	**0.098**	0.089	0.117	0.177	0.195	0.324
Front desk/ doorperson	**0.093**	0.165	0.193	0.178	0.200	0.172
Stove	**0.092**	0.109	0.256	0.143	0.224	0.175
Lockbox	**0.090**	0.124	0.152	0.233	0.185	0.216
Changing table	**0.073**	0.359	0.272	0.054	0.152	0.089

Table 4.16. Normalized label probabilities for "Lisbon Coast" in 2023 data

Normalized label	**Lisbon Coast**	Lisbon	Metropolitan Lisbon	Metropolitan West	Rural West	Peripheral Rural West
Building staff	**0.355**	0.146	0.167	0.155	0.085	0.092
Access	**0.338**	0.025	0.024	0.234	0.198	0.182
Babysitter recommendations	**0.282**	0.290	0.073	0.192	0.088	0.076
Mini fridge	**0.247**	0.206	0.086	0.134	0.135	0.192
EV charger	**0.242**	0.227	0.171	0.134	0.156	0.070
(...)	**(...)**	(...)	(...)	(...)	(...)	**(...)**
Trash compactor	**0.099**	0.217	0.183	0.263	0.137	0.101
Kitchen	**0.084**	0.269	0.156	0.146	0.106	0.239

4.2.3.2 Lisbon

The results for the "Lisbon" region are presented in Table 4.17 for 2018 and in Table 4.18 for 2023, highlighting the trends in accommodation features over these two years.

In 2018, the labels with the highest percentages in Lisbon accommodations were "Wifi" (0.451) and "Waterfront" (0.446), underscoring the priorities of connectivity and scenic locations for hosts. The strong emphasis on "Wifi" highlights its importance as an essential service for guests, while the prominence of "Waterfront" suggests that many accommodations boast picturesque views by the river, appealing to those seeking a vibrant atmosphere.

Other notable features included "Keypad" (0.372) and "Other" (0.364), indicating a focus on secure entry systems and a diverse array of additional amenities. The presence of "Changing table" (0.359) further suggests that many accommodations cater to families with young children.

By 2023, the most prominent features shifted slightly. "Window AC unit" took the lead with a probability of 0.349, reflecting the growing need for air conditioning in urban settings where natural ventilation may be limited. "Self check-in" (0.304) became increasingly important, highlighting the demand for flexible and automated check-in options. The presence of "Portable fans" (0.293) indicates a continued need for effective cooling solutions, complementing air conditioning systems. Additionally, "Babysitter recommendations" emerged as a significant feature at 0.290, reinforcing the family-friendly focus of accommodations in the area.

Despite these shifts, certain features remained notably rare in Lisbon's accommodations. "Parking" is infrequent, largely due to high urban density and the limited availability of dedicated spaces. Similarly, "Access" is low, as many properties lack direct routes to beaches, reflecting the city's riverside location. The absence of "Pool" underscores the constraints of urban environments on outdoor space, while the low occurrence of "Indoor fireplace" suggests a preference for compact climate control solutions. Lastly, the very limited presence of "BBQ grill" indicates that the compact nature of these properties restricts outdoor cooking facilities.

TABLE 4.17. Normalized label probabilities for "Lisbon" in 2018 data

Normalized label	Lisbon Coast	**Lisbon**	Metropolitan Lisbon	Metropolitan West	Rural West	Peripheral Rural West
Wifi	0.195	**0.451**	0.109	0.151	0.046	0.049
Waterfront	0.132	**0.446**	0.059	0.153	0.179	0.032
Keypad	0.225	**0.372**	0.188	0.065	0.091	0.058
Other	0.189	**0.364**	0.099	0.134	0.098	0.116
Changing table	0.073	**0.359**	0.272	0.054	0.152	0.089
(...)	(...)	**(...)**	(...)	(...)	(...)	(...)
Parking	0.124	**0.030**	0.082	0.173	0.285	0.305
Access	0.168	**0.025**	0.037	0.284	0.209	0.277
Pool	0.173	**0.018**	0.030	0.252	0.150	0.377
Indoor fireplace	0.197	**0.017**	0.044	0.201	0.371	0.170

Table 4.18. Normalized label probabilities for "Lisbon" in 2023 data

Normalized label	Lisbon Coast	**Lisbon**	Metropolitan Lisbon	Metropolitan West	Rural West	Peripheral Rural West
Window AC unit	0.216	**0.349**	0.113	0.068	0.053	0.201
Self check-in	0.137	**0.304**	0.189	0.141	0.085	0.144
Portable fans	0.135	**0.293**	0.321	0.070	0.088	0.094
Babysitter recommendations	0.282	**0.290**	0.073	0.192	0.088	0.076
(...)	(...)	**(...)**	(...)	(...)	(...)	**(...)**
BBQ grill	0.131	**0.045**	0.147	0.184	0.206	0.287
Parking	0.127	**0.034**	0.203	0.145	0.252	0.239
Indoor fireplace	0.222	**0.025**	0.055	0.205	0.320	0.173
Access	0.338	**0.025**	0.024	0.234	0.198	0.182
Pool	0.153	**0.017**	0.014	0.230	0.240	0.347

4.2.3.3 Metropolitan Lisbon

The normalized labels with the highest and lowest presence in the "Metropolitan Lisbon" region, including Amadora and Odivelas, are shown in Table 4.19 for 2018 and in Table 4.20 for 2023, offering insights into the region's most and least common accommodation features over time.

In 2018's context, accommodations in "Metropolitan Lisbon" placed a high emphasis on practical features that cater to the needs of an urban environment. The presence of "Elevator", with a probability of 0.470, highlights the multi-story nature of buildings in this densely populated region, where accessibility is crucial. Similarly, the "Handheld shower head", with a probability of 0.354, indicates a preference for flexible and practical bathroom amenities, enhancing guest comfort. Another significant feature in 2018 was the "Microwave", with a probability of 0.350, which suggests that many accommodations provided this appliance for guests who prefer the convenience of preparing or heating meals during their stay. Additionally, the presence of "Outlet covers", at 0.341, pointed to a focus on safety, particularly important for families with young children staying in urban accommodations.

By 2023, some features remained consistent, while new priorities emerged. Elevators continued to be a key amenity, with a high probability of 0.422, indicating that verticality remains a defining characteristic of buildings in this urban area. However, the most notable change in 2023 was the rise of "Bikes" as a prominent feature, with a probability of 0.438. This reflects an increased focus on sustainable mobility and outdoor activities, particularly in a region where cycling has become more common for commuting and leisure. Another significant feature in 2023 was fireplace guards, with a probability of 0.379, showing a heightened concern for safety in accommodations that do include fireplaces.

In both 2018 and 2023, accommodations in the "Metropolitan Lisbon" region consistently showed low percentages for features associated with leisure and outdoor environments, which is indicative of the urban nature of the area. The presence of "Waterfront" remained minimal, with probabilities of 0.059 in 2018 and 0.030 in 2023, while "Access" to beaches, lakes, or resorts was similarly scarce, dropping from 0.037 to 0.024 over the years, reflecting the region's inland location and lack of proximity to such natural features. "Pool" was also uncommon, with a probability of 0.030 in 2018 and further declining to 0.014 in 2023, emphasizing the spatial constraints typical of densely populated urban settings. Additionally, the low presence of "Indoor fireplace", at 0.044 in 2018, suggests this feature was rare, likely due to the impracticality of such amenity in urban accommodations.

Table 4.19. Normalized label probabilities for "Metropolitan Lisbon" in 2018 data

Normalized label	Lisbon Coast	Lisbon	**Metropolitan Lisbon**	Metropolitan West	Rural West	Peripheral Rural West
Elevator	0.175	0.165	**0.470**	0.096	0.070	0.025
Handheld shower head	0.144	0.161	**0.354**	0.131	0.090	0.119
Microwave	0.191	0.135	**0.350**	0.114	0.113	0.097
Outlet covers	0.117	0.064	**0.341**	0.133	0.192	0.153
(...)	(...)	(...)	**(...)**	(...)	(...)	(...)
Waterfront	0.132	0.446	**0.059**	0.153	0.179	0.032
Indoor fireplace	0.197	0.017	**0.044**	0.201	0.371	0.170
Access	0.168	0.025	**0.037**	0.284	0.209	0.277
Pool	0.173	0.018	**0.030**	0.252	0.150	0.377

4.2.3.4 Metropolitan West

The normalized labels with the highest and lowest presence in the "Metropolitan West" region, which includes the municipalities of Mafra, Loures, and Vila Franca de Xira, are detailed in Table 4.21. Among the most prevalent features, "Stair gates" (0.305) and "High chair" (0.297) highlight a strong emphasis on child safety and family-friendly accommodations. The significant percentage of "Stair gates" reflects a commitment to safety for young children, while the high occurrence of "High chair" indicates a focus on catering to families' needs for additional amenities. The notable presence of "Access" (0.284) in the Metropolitan West region is particularly interesting, given that only Mafra is

Table 4.20. Normalized label probabilities for "Metropolitan Lisbon" in 2023 data

Normalized label	Lisbon Coast	Lisbon	**Metropolitan Lisbon**	Metropolitan West	Rural West	Peripheral Rural West
Bikes	0.123	0.075	**0.438**	0.079	0.148	0.137
Elevator	0.159	0.191	**0.422**	0.115	0.076	0.037
Fireplace guards	0.171	0.120	**0.379**	0.119	0.138	0.074
(...)	(...)	(...)	**(...)**	(...)	(...)	(...)
Waterfront	0.132	0.156	**0.030**	0.294	0.198	0.191
Access	0.338	0.025	**0.024**	0.234	0.198	0.182
Pool	0.153	0.017	**0.014**	0.230	0.240	0.347

coastal; Loures and Vila Franca de Xira are located inland. However, since Mafra accounts for 90% of the accommodations in this area, this helps explain why "Access" is prominently featured. Another important characteristic is "Wide hallway clearance" (0.287), suggesting many accommodations prioritize spacious corridors for enhanced mobility, especially for individuals with reduced mobility or families using strollers.

Conversely, the low percentage of "Keypad" indicates that few accommodations provide code-entry systems, suggesting that added security from these devices is not a priority in the region. Additionally, the low presence of "Doorman" services reflects a lesser emphasis on such amenities. Similar to the "Lisbon Coast" region, "Changing table" appears as the feature with the lowest percentage, indicating limited availability for accommodating families with infants compared to other child-friendly features.

Table 4.21. Normalized label probabilities for "Metropolitan West" in 2018 data

Normalized label	Lisbon Coast	Lisbon	Metropolitan Lisbon	**Metropolitan West**	Rural West	Peripheral Rural West
Stair gates	0.153	0.151	0.161	**0.305**	0.056	0.175
High chair	0.186	0.193	0.107	**0.297**	0.065	0.151
Wide hallway clearance	0.112	0.144	0.136	**0.287**	0.190	0.131
Access	0.168	0.025	0.037	**0.284**	0.209	0.277
(...)	(...)	(...)	(...)	**(...)**	(...)	(...)
Keypad	0.255	0.372	0.188	**0.065**	0.091	0.058
Doorman	0.217	0.246	0.185	**0.054**	0.162	0.136
Changing table	0.073	0.359	0.272	**0.054**	0.152	0.089

The analysis of Table 4.22, detailing the predominant characteristics for the "Metropolitan West" class in 2023, provides further insights. In 2023, "Game console" emerges as a standout feature with a probability of 0.333, indicating a strong focus on modern entertainment options for guests. Additionally, the feature "Waterfront" has a probability of 0.294, suggesting that while the region is not primarily coastal, many accommodations are located near waterfront or riverine areas. The presence of "Trash compactor" (0.263) also reflects a commitment to efficient waste management.

On the other hand, features with the lowest probabilities include "Bikes" (0.079), "Portable fans" (0.070), and "Window AC unit" (0.068). The low probability for "Bikes" suggests that the Metropolitan West is less conducive to bicycle commuting, likely due to its more suburban characteristics. Furthermore, the minimal presence of portable fans and window AC units indicates that cooling options are not prioritized in accommodations within Mafra, Loures, or Vila Franca de Xira.

Table 4.22. Normalized label probabilities for "Metropolitan West" in 2023 data

Normalized label	Lisbon Coast	Lisbon	Metropolitan Lisbon	**Metropolitan West**	Rural West	Peripheral Rural West
Game console	0.134	0.079	0.127	**0.333**	0.231	0.096
Waterfront	0.132	0.156	0.030	**0.294**	0.198	0.191
Trash compactor	0.099	0.217	0.183	**0.263**	0.137	0.101
(...)	(...)	(...)	(...)	**(...)**	(...)	(...)
Bikes	0.123	0.075	0.438	**0.079**	0.148	0.137
Portable fans	0.135	0.293	0.321	**0.070**	0.088	0.094
Window AC unit	0.216	0.349	0.113	**0.068**	0.053	0.201

4.2.3.5 Rural West

Table 4.23 details the normalized labels with the highest and lowest presence in the "Rural West" region, which includes the municipalities of Arruda dos Vinhos, Torres Vedras, and Sobral de Monte Agraço.

In 2018, the most prominent features in this region were "Indoor fireplace" (0.371) and "BBQ grill" (0.318), which highlight the rural focus on traditional and outdoor amenities. Unlike the urban regions like "Lisbon" and "Metropolitan Lisbon," where "Indoor fireplace" is less common, its significant presence in the "Rural West" region reflects the lifestyle and larger outdoor spaces available in this area. The high presence of "BBQ grill" further supports this, suggesting that rural accommodations emphasize outdoor dining experiences suited to the environment.

On the other hand, features such as "High chair" (0.065), "Stair gates" (0.056), and "Fireplace guards" (0.053) were less common, indicating that accommodations catering to families with young children are rarer. "24-hour check-in" (0.062) was also low, reflecting the less hurried pace typical of rural accommodations. Additionally, the low presence of "Hot tub" (0.059) suggests that such luxury amenities are less emphasized, with more focus on traditional comforts like fireplaces and BBQ grills. The lowest percentage was observed for "Wifi" (0.046), indicating that internet connectivity is limited or not a priority, further reinforcing the idea that accommodations in this region focus on providing a rural experience rather than urban conveniences.

TABLE 4.23. Normalized label probabilities for "Rural West" in 2018 data

Normalized label	Lisbon Coast	Lisbon	Metropolitan Lisbon	Metropolitan West	**Rural West**	Peripheral Rural West
Indoor fireplace	0.197	0.017	0.044	0.201	**0.371**	0.170
BBQ grill	0.105	0.054	0.073	0.222	**0.318**	0.228
(...)	(...)	(...)	(...)	(...)	**(...)**	(...)
High chair	0.186	0.193	0.107	0.297	**0.065**	0.151
24-hour check-in	0.175	0.188	0.305	0.164	**0.062**	0.106
Hot tub	0.158	0.168	0.258	0.127	**0.059**	0.229
Stair gates	0.153	0.151	0.161	0.305	**0.056**	0.175
Fireplace guards	0.165	0.141	0.328	0.158	**0.053**	0.155
Wifi	0.195	0.451	0.109	0.151	**0.046**	0.049

By 2023, as shown in Table 4.24, some trends had evolved, but the emphasis on traditional rural features remained. "Indoor fireplace" continued to dominate with the highest probability (0.320), indicating its persistent role as a key amenity in the region's accommodations. The presence of "Safe" (0.280) highlights an increased focus on guest security during their stays.

In contrast, the features with the lowest probabilities remained centered around urban conveniences. "Wifi" (0.058) and "Window AC unit" (0.053) were the least common amenities in 2023, further underscoring the region's focus on offering a more rustic and authentic experience. The low presence of "Window AC unit" aligns with the continued preference for traditional heating options like the "Indoor fireplace," which suits the region's accommodation style and environment.

Table 4.24. Normalized label probabilities for "Rural West" in 2023 data

Normalized label	Lisbon Coast	Lisbon	Metropolitan Lisbon	Metropolitan West	**Rural West**	Peripheral Rural West
Indoor fireplace	0.222	0.025	0.055	0.205	**0.320**	0.173
Safe	0.123	0.114	0.162	0.162	**0.280**	0.158
(...)	(...)	(...)	(...)	(...)	**(...)**	(...)
Wifi	0.229	0.234	0.135	0.220	**0.058**	0.124
Window AC unit	0.216	0.349	0.113	0.068	**0.053**	0.201

4.2.3.6 Peripheral Rural West

Table 4.25 presents the normalized labels with the highest and lowest presence in the "Peripheral Rural West" region, which includes the municipalities of Azambuja, Alenquer, Cadaval, and Lourinhã.

Among the most prevalent features, "Pool" stands out with a high percentage of 0.377, reflecting the presence of swimming pools in this more remote area from the capital, which typically offers larger and more spacious properties. The significant presence of "Cleaning before checkout" (0.370) is also notable, indicating that hosts in this region place a strong emphasis on cleanliness before guests depart, ensuring a high-quality experience in less urban settings. Other important features include "TV" (0.341) and "Suitable for events" (0.324). The high percentage of "TV" suggests that televisions are a valued amenity in accommodations, likely due to the need for entertainment in more rural areas. The label "Suitable for events" indicates that many properties are designed to accommodate events, reflecting a range of versatile spaces that can host various types of gatherings and celebrations. "Parking" (0.305) is also prominent, indicating that many accommodations in the region offer parking spaces for guests. Combined with the high prevalence of "Pool" (0.377), it is clear that accommodations in the Peripheral Rural West often feature larger spaces, allowing for these amenities and providing a higher level of comfort and convenience.

In contrast, the presence of features such as "Babysitter recommendations" (0.059), "Keypad" (0.058), "Children's dinnerware" (0.057), "Wifi" (0.049), and "Waterfront" (0.032) is notably lower. The low percentage of "Babysitter recommendations" and "Children's dinnerware" suggests that accommodations in this region are less focused on amenities for families with young children. The reduced prevalence of "Keypad" and "Wifi" indicates that additional security and digital connectivity are not priorities for hosts in this area, possibly reflecting a greater emphasis on disconnection and natural surroundings. The low percentage of "Waterfront" (0.032) reveals that, despite Lourinhã being located in a coastal area and having the highest number of accommodations compared to other municipalities in the region, there is no predominant presence of properties with direct access to waterfront areas. This suggests that even properties in Lourinhã are relatively distant from the coast. Additionally, the low presence of "Elevator" (0.025) indicates that, in this predominantly rural region with many properties in less urbanized areas, elevators are not a priority, reflecting the prevalence of properties with fewer floors that do not require elevators.

TABLE 4.25. Normalized label probabilities for "Peripheral Rural West" in 2018 data

Normalized label	Lisbon Coast	Lisbon	Metropolitan Lisbon	Metropolitan West	Rural West	**Peripheral Rural West**
Pool	0.173	0.018	0.030	0.252	0.150	**0.377**
Cleaning before checkout	0.177	0.123	0.103	0.114	0.113	**0.370**
TV	0.133	0.078	0.151	0.129	0.167	**0.341**
Suitable for events	0.098	0.089	0.117	0.177	0.195	**0.324**
Parking	0.124	0.030	0.082	0.173	0.285	**0.305**
(...)	(...)	(...)	(...)	(...)	(...)	**(...)**
Babysitter recommendations	0.256	0.228	0.119	0.217	0.122	**0.059**
Keypad	0.225	0.372	0.188	0.065	0.091	**0.058**
Children's dinnerware	0.169	0.216	0.195	0.173	0.190	**0.057**
Wifi	0.195	0.451	0.109	0.151	0.046	**0.049**
Waterfront	0.132	0.446	0.059	0.153	0.179	**0.032**
Elevator	0.175	0.165	0.470	0.096	0.070	**0.025**

Moving to 2023 (Table 4.26) , "Ceiling fan" stands out with a probability of 0.437, which suggests that accommodations in this area frequently offer ceiling fans as cooling solution well-suited to the needs of rural and spacious environments. Additionally, "Pool" and "Breakfast" are also prominent features, with probabilities of 0.347 and 0.341, respectively. The significant presence of pools indicates that many accommodations in the Peripheral Rural West value providing outdoor leisure areas, making use of the available space. The inclusion of breakfast services reflects an emphasis on offering a more complete and convenient experience for guests, which can be particularly appealing in rural areas.

Consistent with 2018, the feature "Elevator" has the lowest probability, at 0.037, which suggests that elevators are uncommon in this region, aligning with the profile of accommodations that are generally less urban and situated in areas with lower population density, where the need for elevators is less prevalent.

TABLE 4.26. Normalized label probabilities for "Peripheral Rural West" in 2023 data

Normalized label	Lisbon Coast	Lisbon	Metropolitan Lisbon	Metropolitan West	Rural West	**Peripheral Rural West**
Ceiling fan	0.113	0.135	0.071	0.170	0.074	**0.437**
Pool	0.153	0.017	0.014	0.230	0.240	**0.347**
Breakfast	0.124	0.107	0.077	0.157	0.194	**0.341**
(...)	(...)	(...)	(...)	(...)	(...)	**(...)**
Elevator	0.159	0.191	0.422	0.115	0.076	**0.037**

4.2.3.7 General analysis

To summarize the class-by-class analysis, a further evaluation was conducted to identify the characteristics with the highest probability of belonging to a class in a more general sense. The results for 2018 are presented in Table 4.27. Among the eleven selected characteristics with the highest probability of belonging to a class, five are associated with "Lisbon", three with "Metropolitan Lisbon", two with "Peripheral Rural West", and only one with "Rural West". The classes "Lisbon Coast" and "Metropolitan West" do not feature any of the top characteristics with the highest probability of belonging.

Among the analyzed characteristics, "Elevator" stands out with the highest probability of 0.470 in the "Metropolitan Lisbon" area. The characteristics "Wifi" and "Waterfront" have similar probabilities of 0.451 and 0.446, respectively, in the "Lisbon" area. Following this, "Pool" shows a probability of 0.377 in the "Peripheral Rural West" region, while "Keypad" has a probability of 0.372 in "Lisbon". "Indoor fireplace" and "Cleaning before checkout" have their highest probabilities in the "Rural West" and "Peripheral Rural West" regions, respectively. Finally, "Other" and "Changing table" are prominent in "Lisbon" with probabilities of 0.364 and 0.359, while "Handheld shower head" and "Microwave" show higher values in "Metropolitan Lisbon", with probabilities of 0.354 and 0.350, respectively.

A more in-depth analysis of Table 4.27 reveals some relevant patterns that help complement the class-by-class analysis. For the characteristic "Wifi", the probability of belonging is 0.455 when considering the combination of "Lisbon Coast", "Metropolitan Lisbon", and "Metropolitan West". This value is higher than the probability for "Lisbon" alone, suggesting that "Wifi" may be more relevant in areas that encompass a greater diversity of urban and suburban zones. For "Pool", the probability of belonging is 0.575 when considering the areas "Lisbon Coast", "Metropolitan West", and "Rural West". This value is higher than the one recorded for "Peripheral Rural West", indicating that the presence of a pool is more common or valued in areas that combine urban and rural zones compared to more peripheral rural regions. Regarding "Cleaning before checkout", although it has the highest probability of belonging in the "Peripheral Rural West" region, this characteristic shows relatively well-distributed probability values across other classes. This reflects that

"Cleaning before checkout" is an important feature in all analyzed areas, with particular significance in the "Peripheral Rural West" region. A similar pattern is observed for "Handheld shower head" and "Microwave", which also show a distribution of probability indicating significant relevance across various classes, though with a greater emphasis in the "Metropolitan Lisbon" region.

TABLE 4.27. Best normalized label probabilities for six-class model using 2018 data

Normalized label	Lisbon Coast	Lisbon	Metropolitan Lisbon	Metropolitan West	Rural West	Peripheral Rural West
Elevator	0.175	0.165	**0.470**	0.096	0.070	0.025
Wifi	0.195	**0.451**	0.109	0.151	0.046	0.049
Waterfront	0.132	**0.446**	0.059	0.153	0.179	0.032
Pool	0.173	0.018	0.030	0.252	0.150	**0.377**
Keypad	0.225	**0.372**	0.188	0.065	0.091	0.058
Indoor fireplace	0.197	0.017	0.044	0.201	**0.371**	0.170
Cleaning before checkout	0.177	0.123	0.103	0.114	0.113	**0.370**
Other	0.189	**0.364**	0.099	0.134	0.098	0.116
Changing table	0.073	**0.359**	0.272	0.054	0.152	0.089
Handheld shower head	0.144	0.161	**0.354**	0.131	0.090	0.119
Microwave	0.191	0.135	**0.350**	0.114	0.113	0.097

Similar to the analysis conducted for 2018, an examination of the normalized labels with the highest probabilities of occurrence, regardless of class, was carried out for 2023. The results, shown in Table 4.28, reveal that "Metropolitan Lisbon" and "Peripheral Rural West" are the regions with the most normalized labels, each having three. They are followed by "Lisbon Coast", with two normalized labels, and finally Lisbon and Metropolitan West, each with one label among those with the highest probability of belonging to a class. Notably, none of the eleven normalized labels with the highest probability are associated with the "Rural West" class.

Among the analyzed labels, "Bikes" has the highest probability, at 0.438 in the "Metropolitan Lisbon" region. The label "Ceiling fan" has a very close probability of 0.437 in the "Peripheral Rural West" region. Following these, "Elevator" and "Fireplace guards" have probabilities of 0.422 and 0.279, respectively, in the "Metropolitan Lisbon" region. "Building staff" has a probability of 0.355 in the "Lisbon Coast" region, and "Window AC unit" has a probability of 0.349 in the "Lisbon" region. "Pool" and "Breakfast" show the highest probabilities of occurrence in the "Peripheral Rural West" region, with values of 0.347 and 0.341, respectively. Finally, "Access" stands out in the "Lisbon Coast" region with a probability of 0.338, while "Game console" is notable in the "Metropolitan West" region with a probability of 0.333.

The analysis of Table 4.28 also reveals some interesting patterns. The normalized labels "Elevator" and "Building staff" show the lowest probabilities of belonging to a class in the "Rural West" and "Peripheral Rural West" regions. This indicates a lower presence of accommodations in multi-story buildings or large-scale properties (such as hotels), which would require elevators or dedicated staff. These features are therefore more common in urban areas. Conversely, the labels "Pool" and "Access" have very low probabilities in the "Lisbon" and "Metropolitan Lisbon" regions, reinforcing the idea that in densely populated urban areas, there is limited space for installing pools and less common proximity to beaches, lakes, or resorts. This contrasts with more rural areas, where these features are more prevalent.

Table 4.28. Best normalized label probabilities for six-class model using 2023 data

Normalized label	Lisbon Coast	Lisbon	Metropolitan Lisbon	Metropolitan West	Rural West	Peripheral Rural West
Bikes	0.123	0.075	**0.438**	0.079	0.148	0.137
Ceiling fan	0.113	0.135	0.071	0.170	0.074	**0.437**
Elevator	0.159	0.191	**0.422**	0.115	0.076	0.037
Fireplace guards	0.171	0.120	**0.379**	0.119	0.138	0.074
Building staff	**0.355**	0.146	0.167	0.155	0.085	0.092
Window AC unit	0.216	**0.349**	0.113	0.068	0.053	0.201
Pool	0.153	0.017	0.014	0.230	0.240	**0.347**
Breakfast	0.124	0.107	0.077	0.157	0.194	**0.341**
Access	**0.338**	0.025	0.024	0.234	0.198	0.182
Game console	0.134	0.079	0.127	**0.333**	0.231	0.096

4.3. Discussion

This study explored the effects of the COVID-19 pandemic on the accommodation sharing economy, specifically analyzing how Airbnb hosts in the Lisbon district adjusted their listings before and after the pandemic. The study's findings reveal significant trends and transformations in the characteristics and classification of accommodations, informed by advanced data classification models.

The analysis demonstrated that all classifiers effectively addressed classification tasks for both the '2_regions' and '6_regions' categories, with RBF SVM and Random Forest consistently emerging as top performers. The observed improvements in accuracy and F1-scores from 2018 to 2023 can be attributed to a larger and more balanced dataset, along with the inclusion of additional independent variables. However, the models faced challenges in classifying minority classes, a common issue in multi-class classification tasks.

Focusing specifically on Logistic Regression, the findings indicated that while it performed well in identifying the "Lisbon" class, it struggled significantly with the minority classes. This pattern persisted across both years, highlighting ongoing difficulties in accurately classifying less-represented classes. Confusion matrices revealed a notable bias toward the "Lisbon" class, leading to frequent misclassifications.

An examination of normalized labels illustrated distinct differences between urban and rural accommodations. In both years, features associated with "Outside Lisbon" prominently included amenities like pools and outdoor spaces, while urban accommodations leaned toward essentials such as kitchens and climate control solutions. These trends reflect the unique characteristics of the sharing economy in Lisbon and underscore the varying demands of urban versus rural settings.

The analysis revealed a significant shift in "Lisbon Coast" accommodations from 2018 to 2023, transitioning from a focus on family-oriented amenities—like security and outdoor spaces—to an increased emphasis on modern conveniences, including on-site support staff and sustainable features. This evolution illustrates the region's adaptability to changing guest preferences and its appeal as a destination balancing relaxation with contemporary hospitality trends.

In the "Metropolitan Lisbon" region, the evolution of accommodation features from 2018 to 2023 demonstrated a growing focus on comfort and convenience, particularly family-friendly amenities. While the availability of bicycles increased and elevators remained important for urban mobility, the persistent scarcity of features like parking and outdoor spaces highlights the challenges posed by urban density.

Accommodations in the Metropolitan West region have prioritized child safety, accessibility, and spaciousness, while placing less emphasis on security systems and reception services. The evolution from 2018 to 2023 indicates a growing focus on modern entertainment and effective waste management, although cooling options and sustainable transportation remain less emphasized.

In the "Rural West" region, accommodations in both 2018 and 2023 continued to emphasize traditional comforts, such as fireplaces and outdoor amenities, while urban conveniences like internet access and modern cooling systems were less prioritized. This region caters to guests seeking a rural experience focused on comfort and space, with family-oriented amenities and high-tech conveniences remaining less emphasized.

Finally, the "Peripheral Rural West" region has seen an evolution in accommodation features from 2018 to 2023, with an increasing emphasis on comfort and leisure amenities such as ceiling fans and breakfast offerings. While pools remain significant, family-oriented features and elevators continue to be uncommon, reflecting the area's rural character and preference for spacious properties.

This study highlights the effectiveness of various classification algorithms and reveals significant trends in accommodation features across the Lisbon district. The findings underscore the need for continued attention to minority class performance in classification tasks and provide insights into the evolving preferences of guests in urban and rural settings.

CHAPTER 5

Conclusions and future work

The present investigation begins by characterizing the regions within the Lisbon district based on the distribution of accommodations across its 16 municipalities. To streamline the analysis, the study consolidated the '16_regions' variable into two primary categories — Lisbon and outside Lisbon — and further divided these into six distinct classes. This reclassification resulted in the addition of two new columns to the datasets, thereby enhancing the spatial analysis of accommodation amenities.

Regarding amenities, the study found a drastic increase in the number of unique amenities recorded, from 121 in 2018 to 3,010 in 2023. This increase was mainly due to more detailed documentation by hosts in 2023. In the scope of this study, we have manually normalized and aggregated similar amenities into standardized labels, resulting in 100 labels for 2018 and 128 for 2023.

The analysis identified 68 common amenities while noting that 32 amenities from 2018 were absent in 2023, and 60 new ones were introduced, reflecting a trend towards personalising the guest experience.

To standardize amenities, a dictionary was created to map each amenity to its normalized label, facilitating clearer data interpretation. Boolean columns indicating the presence or absence of these amenities were added to the datasets, simplifying further analysis. Finally, the study filtered out amenities found in fewer than 150 accommodations, ensuring focus on the most relevant features.

After preparing all the data, the analysis examined the performance of various classifiers based on the groupings into two and six geographic regions. The analysis demonstrated that all classifiers effectively addressed classification tasks for both the '2_regions' and '6_regions' categories, with RBF SVM and Random Forest consistently emerging as top performers. However, Logistic Regression was selected as the most suitable algorithm for obtaining the probabilities of amenities belonging to the respective classes. Finally, the study identified the normalized labels that best characterize each region of the Lisbon district according to the models of two and six classes.

This study has provided valuable insights into the transformation of the accommodation sharing economy in the Lisbon district following the COVID-19 pandemic. The findings reveal that Airbnb hosts adapted their listings significantly in response to changing guest expectations and market dynamics. The pandemic acted as a catalyst for hosts to incorporate modern amenities and sustainable practices, reflecting a shift toward greater emphasis on comfort, safety, and eco-friendliness.

The analysis highlighted distinct characteristics across various regions of Lisbon, demonstrating that urban and rural accommodations cater to different consumer preferences. Urban listings prioritized practicality, with features that support longer stays, while rural accommodations maintained traditional comforts that appeal to those seeking spacious and serene environments. This divergence illustrates how local market conditions and guest demographics shape the offerings in the accommodation sector.

The adaptability of hosts during this period not only underscores their resilience but also indicates a broader trend toward innovation within the sharing economy. These developments suggest that as the market continues to evolve, there will be ongoing opportunities for hosts to align their strategies with shifting consumer demands.

While this study has laid a foundation for understanding these changes, there are several avenues for future research that could enhance our comprehension of the accommodation sharing economy. First, a longitudinal analysis extending beyond 2023 would provide deeper insights into the long-term trends in the market, helping to identify how host strategies and guest preferences continue to evolve over time. Second, improving the classification of minority classes within accommodation listings is essential. Future studies could explore advanced techniques such as ensemble learning or data augmentation to capture these less-represented segments more effectively, leading to a richer understanding of the diverse offerings in the market. Finally, conducting cross-regional comparisons with other cities or countries would yield insights into how different markets respond to similar disruptions. This would enhance the generalizability of findings and contribute to a broader understanding of the accommodation sharing economy's trajectory amidst ongoing global changes.

References

[1] E. Ert, A. Fleischer, and N. Magen, "Trust and reputation in the sharing economy: The role of personal photos in airbnb," *Tourism Management*, vol. 55, 2016, ISSN: 02615177. DOI: 10.1016/j.tourman.2016.01.013.

[2] D. Guttentag, "Airbnb: Disruptive innovation and the rise of an informal tourism accommodation sector," *Current Issues in Tourism*, vol. 18, 12 2015, ISSN: 13683500. DOI: 10.1080/13683500.2013.827159.

[3] P. M. Lin, D. X. Fan, H. Q. Zhang, and C. Lau, "Spend less and experience more: Understanding tourists' social contact in the airbnb context," *International Journal of Hospitality Management*, vol. 83, 2019, ISSN: 02784319. DOI: 10.1016/j.ijhm.2019.04.007.

[4] M. A. Camilleri, "The planning and development of the tourism product," in 2018. DOI: 10.1108/978-1-78756-291-220181001.

[5] D. Guttentag, *Progress on airbnb: A literature review*, 2019. DOI: 10.1108/JHTT-08-2018-0075.

[6] V. C. Medina-Hernandez, E. Marine-Roig, and B. Ferrer-Rosell, "Accommodation sharing: A look beyond airbnb's literature," *International Journal of Culture, Tourism, and Hospitality Research*, vol. 14, 1 2020, ISSN: 17506182. DOI: 10.1108/IJCTHR-07-2019-0130.

[7] S. Kuhzady, S. Seyfi, and L. Béal, *Peer-to-peer (p2p) accommodation in the sharing economy: A review*, 2020. DOI: 10.1080/13683500.2020.1786505.

[8] M. Fuchs, W. Höpken, and M. Lexhagen, "Big data analytics for knowledge generation in tourism destinations - a case from sweden," *Journal of Destination Marketing and Management*, vol. 3, 4 2014, ISSN: 2212571X. DOI: 10.1016/j.jdmm.2014.08.002.

[9] V. Koh, W. Li, G. Livan, and L. Capra, "Offline biases in online platforms: A study of diversity and homophily in airbnb," *EPJ Data Science*, vol. 8, 1 Dec. 2019, ISSN: 21931127. DOI: 10.1140/epjds/s13688-019-0189-5.

[10] R. Mousavi and K. Zhao, "Examining the impacts of airbnb review policy change on listing reviews," *Journal of the Association for Information Systems*, vol. 23, pp. 303–328, 1 2022, ISSN: 15583457. DOI: 10.17705/1jais.00720.

[11] E. Ndaguba and C. V. Zyl, "Professionalizing sharing platforms for sustainable growth in the hospitality sector: Insights gained through hierarchical linear modeling," *Sustainability (Switzerland)*, vol. 15, 10 May 2023, ISSN: 20711050. DOI: 10.3390/su15108267.

[12] N. Stors and S. Baltes, "Constructing urban tourism space digitally: A study of airbnb listings in two berlin neighborhoods," *Proceedings of the ACM on Human-Computer Interaction*, vol. 2, CSCW Nov. 2018, ISSN: 25730142. DOI: 10.1145/3274435.

[13] C. Chang and S. Li, "Study of price determinants of sharing economy-based accommodation services: Evidence from airbnb.com," *Journal of Theoretical and Applied Electronic Commerce Research*, vol. 16, pp. 584–601, 4 Jul. 2021, ISSN: 07181876. DOI: 10.3390/jtaer16040035.

[14] D. Kim, S. P. Park, and S. Yi, "Relevant and rich interactivity under uncertainty: Guest reviews, host responses, and guest purchase intention on airbnb," *Telematics and Informatics*, vol. 65, Dec. 2021, ISSN: 07365853. DOI: 10.1016/j.tele.2021.101708.

[15] R. Situmorang and A. Japutra, "Foreign versus local managers: Finding the perfect leaders for multinational hotel subsidiaries," *International Journal of Hospitality Management*, vol. 78, 2019, ISSN: 02784319. DOI: 10.1016/j.ijhm.2018.11.015.

[16] S. Zenker and F. Kock, "The coronavirus pandemic – a critical discussion of a tourism research agenda," *Tourism Management*, vol. 81, 2020, ISSN: 02615177. DOI: 10.1016/j.tourman.2020.104164.

[17] S. Gössling, D. Scott, and C. M. Hall, "Pandemics, tourism and global change: A rapid assessment of covid-19," *Journal of Sustainable Tourism*, 2020, ISSN: 17477646. DOI: 10.1080/09669582.2020.1758708.

[18] J. I. Taylor, "Airbnb hosts send 40k+ letters to congress pleading for financial relief from coronavirus," *Florida Politics (FLAPOL)*, 2020. [Online]. Available: https://floridapolitics.com/archives/324696-airbnb-hosts-send-40k-letters-to-congress-pleading-for-financial-relief-from-coronavirus.

[19] K. Y. Wang, M. L. Ma, and J. Yu, "Understanding the perceived satisfaction and revisiting intentions of lodgers in a restricted service scenario: Evidence from the hotel industry in quarantine," *Service Business*, vol. 15, 2 2021, ISSN: 18628508. DOI: 10.1007/s11628-021-00445-w.

[20] A. Azevedo and M. F. Santos, "Kdd, semma and crisp-dm: A parallel overview," 2008.

[21] C. Okoli and K. Schabram, "Working papers on information systems a guide to conducting a systematic literature review of information systems research," *Working Papers on Information Systems*, 2010, ISSN: 1535-6078. DOI: 10.2139/ssrn.1954824.

[22] R. Mallett, J. Hagen-Zanker, R. Slater, and M. Duvendack, "The benefits and challenges of using systematic reviews in international development research," *Journal of Development Effectiveness*, vol. 4, 3 2012, ISSN: 19439342. DOI: 10.1080/19439342.2012.711342.

[23] B. Kitchenham, "Procedures for performing systematic reviews," *Keele University,UK and National ICT Australia*, vol. 33, 2004, ISSN: 13537776. DOI: 10.1.1.122.3308.

[24] L. Guo, J. Li, J. Wu, W. Chang, and J. Wu, "A novel airbnb matching scheme in shared economy using confidence and prediction uncertainty analysis," *IEEE Access*, vol. 6, pp. 10 320–10 331, Feb. 2018, ISSN: 21693536. DOI: 10.1109/ACCESS.2018.2801810.

[25] Z. Zhang and R. J. Fu, "Accommodation experience in the sharing economy: A comparative study of airbnb online reviews," *Sustainability (Switzerland)*, vol. 12, pp. 1–11, 24 Dec. 2020, ISSN: 20711050. DOI: 10.3390/su122410500.

[26] M. Abdar and N. Y. Yen, "Analysis of user preference and expectation on shared economy platform: An examination of correlation between points of interest on airbnb," *Computers in Human Behavior*, vol. 107, Jun. 2020, ISSN: 07475632. DOI: 10.1016/j.chb.2018.09.039.

[27] A. Pawlicz, E. Petaković, and A. M. V. Hrgović, "Beyond airbnb. determinants of customer satisfaction in p2p accommodation in time of covid-19," *Sustainability (Switzerland)*, vol. 14, 17 Sep. 2022, ISSN: 20711050. DOI: 10.3390/su141710734.

[28] I. Cenni and P. Goethals, "Business responses to positive reviews online: Face-work on tripadvisor," *Journal of Pragmatics*, vol. 180, pp. 38–50, Jul. 2021, ISSN: 03782166. DOI: 10.1016/j.pragma.2021.04.008.

[29] J. Lyu and S. Fang, "Exploring customers' experiences with p2p accommodations: Measurement scale development and validation in the chinese market," *Sustainability (Switzerland)*, vol. 14, 14 Jul. 2022, ISSN: 20711050. DOI: 10.3390/su14148541.

[30] S. Bai, X. Bi, C. Han, M. Yang, and H. He, "Exploring the over-time variation in customer concerns on sharing economy services," *Journal of Global Information Management*, vol. 31, 1 2023, ISSN: 15337995. DOI: 10.4018/JGIM.321435.

[31] C. K. H. Lee, "How guest-host interactions affect consumer experiences in the sharing economy: New evidence from a configurational analysis based on consumer reviews," *Decision Support Systems*, vol. 152, Jan. 2022, ISSN: 01679236. DOI: 10.1016/j.dss.2021.113634.

[32] C. K. H. Lee and Y. K. Tse, "Improving peer-to-peer accommodation service based on text analytics," *Industrial Management and Data Systems*, vol. 121, 2 2021, ISSN: 02635577. DOI: 10.1108/IMDS-02-2020-0105.

[33] Q. Zhou, Y. Chen, C. Ma, F. Li, Y. Xiao, X. Wang, and X. Fu, "Measurement and analysis of the reviews in airbnb," 2018. DOI: 10.23919/IFIPNetworking.2018.8696543.

[34] G. Santos, V. F. Mota, F. Benevenuto, and T. H. Silva, "Neutrality may matter: Sentiment analysis in reviews of airbnb, booking, and couchsurfing in brazil and usa," *Social Network Analysis and Mining*, vol. 10, 1 Dec. 2020, ISSN: 18695469. DOI: 10.1007/s13278-020-00656-5.

[35] G. Quattrone, A. Nocera, L. Capra, and D. Quercia, "Social interactions or business transactions?what customer reviews disclose about airbnb marketplace," Association for Computing Machinery, Inc, Apr. 2020, pp. 1526–1536, ISBN: 9781450370233. DOI: 10.1145/3366423.3380225.

[36] J. Yu, X. Zhang, and H. S. Kim, "Using online customer reviews to understand customers' experience and satisfaction with integrated resorts," *Sustainability (Switzerland)*, vol. 15, 17 Sep. 2023, ISSN: 20711050. DOI: 10.3390/su151713049.

[37] M. J. Carrasco-Santos, A. Peña-Romero, and D. Guerrero-Navarro, "A luxury tourist destination in housing for tourist purposes: A study of the new airbnb luxe platform in the case of marbella," *Journal of Theoretical and Applied Electronic Commerce Research*, vol. 18, pp. 1020–1040, 2 Jun. 2023, ISSN: 07181876. DOI: 10.3390/jtaer18020052.

[38] N. Čumlievski, M. B. Bakarić, and M. Matetić, "A smart tourism case study: Classification of accommodation using machine learning models based on accommodation characteristics and online guest reviews," *Electronics (Switzerland)*, vol. 11, 6 Mar. 2022, ISSN: 20799292. DOI: 10.3390/electronics11060913.

[39] N. D. Handani, A. L. Riswanto, and H. S. Kim, "A study of inbound travelers experience and satisfaction at quarantine hotels in indonesia during the covid-19 pandemic," *Information (Switzerland)*, vol. 13, 5 May 2022, ISSN: 20782489. DOI: 10.3390/info13050254.

[40] C. K. H. Lee, Y. K. Tse, M. Zhang, and J. Ma, "Analysing online reviews to investigate customer behaviour in the sharing economy: The case of airbnb," *Information Technology and People*, vol. 33, pp. 945–961, 3 Jun. 2020, ISSN: 09593845. DOI: 10.1108/ITP-10-2018-0475.

[41] K. Nessel, S. Kościółek, E. Wszendybył-Skulska, and S. Kopera, "Benefit segmentation in the tourist accommodation market based on ewom attribute ratings," *Information Technology and Tourism*, vol. 23, pp. 265–290, 2 Jun. 2021, ISSN: 19434294. DOI: 10.1007/s40558-021-00200-x.

[42] A. Alsudais and T. Teubner, "Large-scale sentiment analysis on airbnb reviews from 15 cities," 2019.

[43] F. Liu, K. H. Lai, J. Wu, and W. Duan, "Listening to online reviews: A mixed-methods investigation of customer experience in the sharing economy," *Decision Support Systems*, vol. 149, Oct. 2021, ISSN: 01679236. DOI: 10.1016/j.dss.2021.113609.

[44] W. L. Chang and J. Y. Wang, "Mine is yours? using sentiment analysis to explore the degree of risk in the sharing economy," *Electronic Commerce Research and Applications*, vol. 28, pp. 141–158, Mar. 2018, ISSN: 15674223. DOI: 10.1016/j.elerap.2018.01.014.

[45] L. M. Martín-Delgado, J. M. Sánchez-Martín, and J. I. Rengifo-Gallego, "An analysis of online reputation indicators by means of geostatistical techniques-the case of

rural accommodation in extremadura, spain," *ISPRS International Journal of Geo-Information*, vol. 9, 4 2020, ISSN: 22209964. DOI: 10.3390/ijgi9040208.

[46] K. Nash, "Data visualization can shifts our sharing economy perceptions: Austin, texas airbnb landscape," 2021.

[47] B. Biswas, P. Sengupta, and D. Chatterjee, "Examining the determinants of the count of customer reviews in peer-to-peer home-sharing platforms using clustering and count regression techniques," *Decision Support Systems*, vol. 135, Aug. 2020, ISSN: 01679236. DOI: 10.1016/j.dss.2020.113324.

[48] S. Zhang, Y. Lu, and B. Lu, "Shared accommodation services in the sharing economy: Understanding the effects of psychological distance on booking behavior," *Journal of Theoretical and Applied Electronic Commerce Research*, vol. 18, pp. 311–332, 1 Mar. 2023, ISSN: 07181876. DOI: 10.3390/jtaer18010017.

[49] J. Wang and X. Yu, "The driving path of customer sustainable consumption behaviors in the context of the sharing economy-based on the interaction effect of customer signal, service provider signal, and platform signal," *Sustainability (Switzerland)*, vol. 13, 7 Apr. 2021, ISSN: 20711050. DOI: 10.3390/su13073826.

[50] T. Teubner, "The web of host–guest connections on airbnb: A network perspective," *Journal of Systems and Information Technology*, vol. 20, 3 2018, ISSN: 17588847. DOI: 10.1108/JSIT-11-2017-0104.

[51] F. Liu, K. hung Lai, J. Wu, and X. Luo, "How electronic word of mouth matters in peer-to-peer accommodation: The role of price and responsiveness," *International Journal of Electronic Commerce*, vol. 26, pp. 174–199, 2 2022, ISSN: 15579301. DOI: 10.1080/10864415.2022.2050581.

[52] J. Wu, J. Cai, X. (Luo, and J. Benitez, "How to increase customer repeated bookings in the short-term room rental market? a large-scale granular data investigation," *Decision Support Systems*, vol. 143, Apr. 2021, ISSN: 01679236. DOI: 10.1016/j.dss.2021.113495.

[53] Y. Wang, Y. Zheng, and X. Xu, "The impact of discrepancies between offerors' self-disclosure and customers' reviews on online sales of experiences in sharing economy," *International Journal of Electronic Commerce*, vol. 27, pp. 469–499, 4 2023, ISSN: 15579301. DOI: 10.1080/10864415.2023.2255109.

[54] Y. M. Chang, C. H. Chen, J. P. Lai, Y. L. Lin, and P. F. Pai, "Forecasting hotel room occupancy using long short-term memory networks with sentiment analysis and scores of customer online reviews," *Applied Sciences (Switzerland)*, vol. 11, 21 Nov. 2021, ISSN: 20763417. DOI: 10.3390/app112110291.

[55] P. Talón-Ballestero, F. E. García-Muiña, J. J. Rienda-Gómez, and L. González-Serrano, "Repeat consumer behavior on smart p2p tourism platforms," *Sustainability (Switzerland)*, vol. 11, 24 Dec. 2019, ISSN: 20711050. DOI: 10.3390/su11247082.

[56] N. Zhang, R. Liu, X. Y. Zhang, and Z. L. Pang, "The impact of consumer perceived value on repeat purchase intention based on online reviews: By the method of text

mining," *Data Science and Management*, vol. 3, pp. 22–32, Sep. 2021, ISSN: 26667649. DOI: 10.1016/j.dsm.2021.09.001.

[57] C. C. Chen and Y. C. Chang, "What drives purchase intention on airbnb? perspectives of consumer reviews, information quality, and media richness," *Telematics and Informatics*, vol. 35, pp. 1512–1523, 5 Aug. 2018, ISSN: 07365853. DOI: 10.1016/j.tele.2018.03.019.

[58] Z. H. Alharbi, "A sustainable price prediction model for airbnb listings using machine learning and sentiment analysis," *Sustainability (Switzerland)*, vol. 15, 17 Sep. 2023, ISSN: 20711050. DOI: 10.3390/su151713159.

[59] P. Ye, C. H. Wu, J. Qian, Y. Zhou, J. Chen, S. D. Mars, F. Yang, and L. Zhang, "Customized regression model for airbnb dynamic pricing," Association for Computing Machinery, Jul. 2018, pp. 932–940, ISBN: 9781450355520. DOI: 10.1145/3219819.3219830.

[60] M. I. Gordan, E. Peț, G. Popescu, I. Brad, A. I. Milin, T. C. Adamov, R. Ciolac, A. R. Pascariu, and T. Iancu, "Factors influencing the accommodation prices of romanian rural tourism," *Sustainability (Switzerland)*, vol. 15, 1 Jan. 2023, ISSN: 20711050. DOI: 10.3390/su15010191.

[61] R. Jia and S. Wang, "Investigating the impact of professional and nonprofessional hosts' pricing behaviors on accommodation-sharing market outcome," *Sustainability (Switzerland)*, vol. 13, 21 Nov. 2021, ISSN: 20711050. DOI: 10.3390/su132112331.

[62] L. Shen, Q. Liu, G. Chen, and S. Ji, "Text-based price recommendation system for online rental houses," *Big Data Mining and Analytics*, vol. 3, pp. 143–152, 2 Jun. 2020, ISSN: 20960654. DOI: 10.26599/BDMA.2019.9020023.

[63] V. R. Perez-Sanchez, L. Serrano-Estrada, P. Marti, and R. T. Mora-Garcia, "The what, where, and why of airbnb price determinants," *Sustainability (Switzerland)*, vol. 10, 12 Dec. 2018, ISSN: 20711050. DOI: 10.3390/su10124596.

[64] R. Sharma, A. Kumar, and C. Chuah, "Turning the blackbox into a glassbox: An explainable machine learning approach for understanding hospitality customer," *International Journal of Information Management Data Insights*, vol. 1, 2 Nov. 2021, ISSN: 26670968. DOI: 10.1016/j.jjimei.2021.100050.

[65] A. Engin and R. Vetschera, "In whose bed shall i sleep tonight? the impact of transaction-specific versus partner-specific information on pricing on a sharing platform," *Information and Management*, vol. 59, 6 Sep. 2022, ISSN: 03787206. DOI: 10.1016/j.im.2022.103666.

[66] L. Zhang, Q. Yan, and L. Zhang, "A computational framework for understanding antecedents of guests' perceived trust towards hosts on airbnb," *Decision Support Systems*, vol. 115, pp. 105–116, Nov. 2018, ISSN: 01679236. DOI: 10.1016/j.dss.2018.10.002.

[67] L. Zhang, Q. Yan, and L. Zhang, "A text analytics framework for understanding the relationships among host self-description, trust perception and purchase behavior

on airbnb," *Decision Support Systems*, vol. 133, Jun. 2020, ISSN: 01679236. DOI: 10.1016/j.dss.2020.113288.

[68] B. Alsamani, "Better trust between users in sharing economy platforms," Association for Information Systems, 2018, ISBN: 9780996683166.

[69] S. S. Alsheikh, K. Shaalan, and F. Meziane, "Exploring the effects of consumers' trust: A predictive model for satisfying buyers' expectations based on sellers' behavior in the marketplace," *IEEE Access*, vol. 7, pp. 73 357–73 372, 2019, ISSN: 21693536. DOI: 10.1109/ACCESS.2019.2917999.

[70] L. Y. Chaw and C. M. Tang, "Online accommodation booking: What information matters the most to users?" *Information Technology and Tourism*, vol. 21, pp. 369–390, 3 Sep. 2019, ISSN: 19434294. DOI: 10.1007/s40558-019-00146-1.

[71] E. D. Zamani, J. Choudrie, G. Katechos, and Y. Yin, "Trust in the sharing economy: The airbnb case," *Industrial Management and Data Systems*, vol. 119, pp. 1947–1968, 9 Nov. 2019, ISSN: 02635577. DOI: 10.1108/IMDS-04-2019-0207.

[72] G. Quattrone, N. Kusek, and L. Capra, "A global-scale analysis of the sharing economy model – an airbnb case study," *EPJ Data Science*, vol. 11, 1 Dec. 2022, ISSN: 21931127. DOI: 10.1140/epjds/s13688-022-00349-3.

[73] M. Tushev, F. Ebrahimi, and A. Mahmoud, "Digital discrimination in sharing economy a requirements engineering perspective," vol. 2020-August, IEEE Computer Society, Aug. 2020, pp. 204–214, ISBN: 9781728174389. DOI: 10.1109/RE48521.2020.00031.

[74] S. Cho, C. W. Park, and F. Lee, "Homophily and peer-consumer behaviour in a peer-to-peer accommodation sharing economy platform," *Behaviour and Information Technology*, vol. 41, pp. 276–291, 2 2022, ISSN: 13623001. DOI: 10.1080/0144929X.2020.1803403.

[75] M. F. Agina, N. Aliane, O. E. Sawy, H. A. Khairy, and S. Fayyad, "Risks in relation to adopting airbnb accommodation: The role of fear of covid-19," *Sustainability (Switzerland)*, vol. 15, 6 2023, ISSN: 20711050. DOI: 10.3390/su15065050.

[76] M. Roman, M. Roman, E. Grzegorzewska, P. Pietrzak, and K. Roman, "Influence of the covid-19 pandemic on tourism in european countries: Cluster analysis findings," *Sustainability (Switzerland)*, vol. 14, 3 Feb. 2022, ISSN: 20711050. DOI: 10.3390/su14031602.

[77] C. Sun, X. Chai, Q. Fan, and W. Zhang, "The impact of hygiene factors on online hotel consumption in china during the covid-19 pandemic," *Sustainability (Switzerland)*, vol. 15, 4 Feb. 2023, ISSN: 20711050. DOI: 10.3390/su15043537.

[78] Y. Chen, Y. Huang, and C. H. Tan, "Short-term rental and its regulations on the home-sharing platform," *Information and Management*, vol. 58, 3 Apr. 2021, ISSN: 03787206. DOI: 10.1016/j.im.2020.103322.

[79] C. Yu, W. Liao, and D. Margolin, "The differential impacts of blinded online reviews: Comparing socio-emotional features of guest and host reviews on airbnb," *Telematics*

and Informatics, vol. 65, Dec. 2021, ISSN: 07365853. DOI: 10.1016/j.tele.2021.101731.

[80] R. Stewart, "Authenticity for rent? airbnb hosts and the commodification of urban displacement," *Proceedings of the ACM on Human-Computer Interaction*, vol. 6, 2 CSCW Nov. 2022, ISSN: 25730142. DOI: 10.1145/3555594.

[81] S. Jain, D. Proserpio, G. Quattrone, and D. Quercia, "Nowcasting gentrification using airbnb data," *Proceedings of the ACM on Human-Computer Interaction*, vol. 5, CSCW1 Apr. 2021, ISSN: 25730142. DOI: 10.1145/3449112.

[82] X. Zhang and W. Liu, "A comparative analysis between airbnb and hotel industry: The investigation from china," vol. 2020-December, 2020.

[83] M. Alraeeini, Q. Zhong, and E. Antarciuc, "Analysing drivers and barriers of accommodation sharing in dubai using the grey-dematel approach," *Sustainability (Switzerland)*, vol. 11, 20 Oct. 2019, ISSN: 20711050. DOI: 10.3390/su11205645.

[84] A. Khatter, L. White, J. Pyke, and M. McGrath, "Stakeholders' influence on environmental sustainability in the australian hotel industry," *Sustainability (Switzerland)*, vol. 13, pp. 1–17, 3 Feb. 2021, ISSN: 20711050. DOI: 10.3390/su13031351.

[85] A. Fierro and I. Aranburu, "Airbnb branding: Heritage as a branding element in the sharing economy," *Sustainability (Switzerland)*, vol. 11, 1 Jan. 2019, ISSN: 20711050. DOI: 10.3390/su11010074.

APPENDIX A

Studies for each one of the research expressions

Table A.1: Selected studies based on the first research expression

Title	Author(s)	Year
"A computational framework for understanding antecedents of guests' perceived trust towards hosts on Airbnb"	Zhang, Yan, and Zhang	2018
"A Novel Airbnb Matching Scheme in Shared Economy Using Confidence and Prediction Uncertainty Analysis"	Guo *et al.*	2018
"A Sustainable Price Prediction Model for Airbnb Listings Using Machine Learning and Sentiment Analysis"	Alharbi	2023
"A text analytics framework for understanding the relationships among host self-description, trust perception and purchase behavior on Airbnb"	Zhang, Yan, and Zhang	2020
"Accommodation experience in the sharing economy: A comparative study of airbnb online reviews"	Zhang and Fu	2020
"Airbnb branding: Heritage as a branding element in the sharing economy"	Fierro and Aranburu	2019
"Analysis of user preference and expectation on shared economy platform: An examination of correlation between points of interest on Airbnb"	Abdar and Yen	2020
"Authenticity for Rent? Airbnb Hosts and the Commodification of Urban Displacement"	Stewart	2022

Title	Author(s)	Year
“Better trust between users in sharing economy platforms”	Alsamani	2018
“Beyond Airbnb. Determinants of Customer Satisfaction in P2P Accommodation in Time of COVID-19”	Pawlicz, Petaković, and Hrgović	2022
“Business responses to positive reviews online: Face-work on TripAdvisor”	Cenni and Goethals	2021
“Constructing urban tourism space digitally: A study of Airbnb listings in two Berlin neighborhoods”	Stors and Baltes	2018
“Customized regression model for Airbnb dynamic pricing”	Ye *et al.*	2018
“Data visualization can shifts our sharing economy perceptions: Austin, Texas Airbnb landscape”	Nash	2021
“Examining the determinants of the count of customer reviews in peer-to-peer home-sharing platforms using clustering and count regression techniques”	Biswas, Sengupta, and Chatterjee	2020
“Examining the Impacts of Airbnb Review Policy Change on Listing Reviews”	Mousavi and Zhao	2022
“Exploring Customers’ Experiences with P2P Accommodations: Measurement Scale Development and Validation in the Chinese Market”	Lyu and Fang	2022
“Exploring the Effects of Consumers’ Trust: A Predictive Model for Satisfying Buyers’ Expectations Based on Sellers’ Behavior in the Marketplace”	Alsheikh, Shaalan, and Meziane	2019
“Exploring the Over-Time Variation in Customer Concerns on Sharing Economy Services”	Bai *et al.*	2023

Title	Author(s)	Year
"Factors Influencing the Accommodation Prices of Romanian Rural Tourism"	Gordan *et al.*	2023
"How guest-host interactions affect consumer experiences in the sharing economy: New evidence from a configurational analysis based on consumer reviews"	Lee	2022
"Improving peer-to-peer accommodation service based on text analytics"	Lee and Tse	2021
"Investigating the impact of professional and nonprofessional hosts' pricing behaviors on accommodation-sharing market outcome"	Jia and Wang	2021
"Measurement and Analysis of the Reviews in Airbnb"	Zhou *et al.*	2018
"Neutrality may matter: sentiment analysis in reviews of Airbnb, Booking, and Couchsurfing in Brazil and USA"	Santos *et al.*	2020
"Offline biases in online platforms: a study of diversity and homophily in Airbnb"	Koh *et al.*	2019
"Relevant and rich interactivity under uncertainty: Guest reviews, host responses, and guest purchase intention on Airbnb"	Kim, Park, and Yi	2021
"Risks in Relation to Adopting Airbnb Accommodation: The Role of Fear of COVID-19"	Agina *et al.*	2023
"Shared Accommodation Services in the Sharing Economy: Understanding the Effects of Psychological Distance on Booking Behavior"	Zhang, Lu, and Lu	2023
"Short-term rental and its regulations on the home-sharing platform"	Chen, Huang, and Tan	2021
"Social Interactions or Business Transactions?What customer reviews disclose about Airbnb marketplace"	Quattrone *et al.*	2020

Title	Author(s)	Year
“Study of price determinants of sharing economy-based accommodation services: Evidence from airbnb.com”	Chang and Li	2021
“Text-based price recommendation system for online rental houses”	Shen *et al.*	2020
“The differential impacts of blinded online reviews: Comparing socio-emotional features of guest and host reviews on Airbnb”	Yu, Liao, and Margolin	2021
“The driving path of customer sustainable consumption behaviors in the context of the sharing economy-based on the interaction effect of customer signal, service provider signal, and platform signal”	Wang and Yu	2021
“The web of host–guest connections on Airbnb: a network perspective”	Teubner	2018
“The what, where, and why of airbnb price determinants”	Perez-Sanchez *et al.*	2018
“Turning the blackbox into a glassbox: An explainable machine learning approach for understanding hospitality customer”	Sharma, Kumar, and Chuah	2021
“Using Online Customer Reviews to Understand Customers’ Experience and Satisfaction with Integrated Resorts”	Yu, Zhang, and Kim	2023
“How Electronic Word of Mouth Matters in Peer-to-Peer Accommodation: The Role of Price and Responsiveness”	Liu *et al.*	2022
“How to increase customer repeated bookings in the short-term room rental market? A large-scale granular data investigation”	Wu *et al.*	2021
“In whose bed shall I sleep tonight? The impact of transaction-specific versus partner-specific information on pricing on a sharing platform”	Engin and Vetschera	2022

Title	Author(s)	Year
"The Impact of Discrepancies between Offerors' Self-Disclosure and Customers' Reviews on Online Sales of Experiences in Sharing Economy"	Wang, Zheng, and Xu	2023

Table A.2: Selected studies based on the second research expression

Title	Author(s)	Year
"A comparative analysis between Airbnb and hotel industry: The investigation from China"	Zhang and Liu	2020
"A global-scale analysis of the sharing economy model – an AirBnB case study"	Quattrone, Kusek, and Capra	2022
"A Luxury Tourist Destination in Housing for Tourist Purposes: A Study of the New Airbnb Luxe Platform in the Case of Marbella"	Carrasco-Santos, Peña-Romero, and Guerrero-Navarro	2023
"A Smart Tourism Case Study: Classification of Accommodation Using Machine Learning Models Based on Accommodation Characteristics and Online Guest Reviews"	Čumlievski, Bakarić, and Matetić	2022
"A Study of Inbound Travelers Experience and Satisfaction at Quarantine Hotels in Indonesia during the COVID-19 Pandemic"	Handani, Riswanto, and Kim	2022
"Analysing drivers and barriers of accommodation sharing in Dubai using the grey-DEMATEL approach"	Alraeeini, Zhong, and Antarciuc	2019
"Analysing online reviews to investigate customer behaviour in the sharing economy: The case of Airbnb"	Lee *et al.*	2020

Title	Author(s)	Year
"Benefit segmentation in the tourist accommodation market based on eWOM attribute ratings"	Nessel *et al.*	2021
"Digital Discrimination in Sharing Economy A Requirements Engineering Perspective"	Tushev, Ebrahimi, and Mahmoud	2020
"Forecasting hotel room occupancy using long short-term memory networks with sentiment analysis and scores of customer online reviews"	Chang *et al.*	2021
"Influence of the COVID-19 Pandemic on Tourism in European Countries: Cluster Analysis Findings"	Roman *et al.*	2022
"Large-scale sentiment analysis on airbnb reviews from 15 cities"	Alsudais and Teubner	2019
"Listening to online reviews: A mixed-methods investigation of customer experience in the sharing economy"	Liu *et al.*	2021
"Mine is yours? Using sentiment analysis to explore the degree of risk in the sharing economy"	Chang and Wang	2018
"Nowcasting Gentrification Using Airbnb Data"	Jain *et al.*	2021
"Online accommodation booking: what information matters the most to users?"	Chaw and Tang	2019
"Professionalizing Sharing Platforms for Sustainable Growth in the Hospitality Sector: Insights Gained through Hierarchical Linear Modeling"	Ndaguba and Zyl	2023
"Repeat consumer behavior on smart P2P tourism platforms"	Talón-Ballestero *et al.*	2019
"Stakeholders' influence on environmental sustainability in the Australian hotel industry"	Khatter *et al.*	2021

Title	**Author(s)**	**Year**
"The impact of consumer perceived value on repeat purchase intention based on online reviews: by the method of text mining"	Zhang *et al.*	2021
"The Impact of Hygiene Factors on Online Hotel Consumption in China during the COVID-19 Pandemic"	Sun *et al.*	2023
"Trust in the sharing economy: the AirBnB case"	Zamani *et al.*	2019
"What drives purchase intention on Airbnb? Perspectives of consumer reviews, information quality, and media richness"	Chen and Chang	2018
"An analysis of online reputation indicators by means of geostatistical techniques-the case of rural accommodation in extremadura, Spain"	Martín-Delgado, Sánchez-Martín, and Rengifo-Gallego	2020
"Homophily and peer-consumer behaviour in a peer-to-peer accommodation sharing economy platform"	Cho, Park, and Lee	2022

APPENDIX B

Dataset features

Table B.1: Dataset features for 2018 and 2023

Feature	**2018**	**2023**
id	X	X
listing_url	X	X
scrape_id	X	X
last_scraped	X	X
source		X
name	X	X
summary	X	
space	X	
description	X	X
experiences_offered	X	
neighborhood_overview	X	X
notes	X	
transit	X	
access	X	
interaction	X	
house_rules	X	
thumbnail_url	X	
medium_url	X	
picture_url	X	X
xl_picture_url	X	
host_id	X	X
host_url	X	X
host_name	X	X
host_since	X	X
host_location	X	X
host_about	X	X
host_response_time	X	X
host_response_rate	X	X
host_acceptance_rate	X	X
host_is_superhost	X	X

Feature	2018	2023
host_thumbnail_url	X	X
host_picture_url	X	X
host_neighbourhood	X	X
host_listings_count	X	X
host_total_listings_count	X	X
host_verifications	X	X
host_has_profile_pic	X	X
host_identity_verified	X	X
calculated_host_listings_count	X	X
calculated_host_listings_count_entire_homes		X
calculated_host_listings_count_private_rooms		X
calculated_host_listings_count_shared_rooms		X
street	X	
neighbourhood	X	X
neighbourhood_cleansed	X	X
neighbourhood_group_cleansed	X	X
city	X	
state	X	
zipcode	X	
market	X	
smart_location	X	
country_code	X	
country	X	
latitude	X	X
longitude	X	X
is_location_exact	X	
property_type	X	X
room_type	X	X
accommodates	X	X
bathrooms	X	X
bathrooms_text		X
bedrooms	X	X
beds	X	X
bed_type	X	
amenities	X	X
square_feet	X	
price	X	X
weekly_price	X	

Feature	2018	2023
monthly_price	X	
security_deposit	X	
cleaning_fee	X	
guests_included	X	
extra_people	X	
minimum_nights	X	X
maximum_nights	X	X
minimum_minimum_nights		X
maximum_minimum_nights		X
minimum_maximum_nights		X
maximum_maximum_nights		X
minimum_nights_avg_ntm		X
maximum_nights_avg_ntm		X
has_availability	X	X
availability_30	X	X
availability_60	X	X
availability_90	X	X
availability_365	X	X
instant_bookable	X	X
is_business_travel_ready	X	
cancellation_policy	X	
require_guest_profile_picture	X	
require_guest_phone_verification	X	
calendar_updated	X	X
calendar_last_scraped	X	X
number_of_reviews	X	X
number_of_reviews_ltm		X
number_of_reviews_l30d		X
first_review	X	X
last_review	X	X
review_scores_rating	X	X
review_scores_accuracy	X	X
review_scores_cleanliness	X	X
review_scores_checkin	X	X
review_scores_communication	X	X
review_scores_location	X	X
review_scores_value	X	X
reviews_per_month	X	X

Feature	2018	2023
requires_license	X	
license	X	X
jurisdiction_names	X	

APPENDIX C

Characteristics of normalized labels

Table C.1: Characteristics of normalized labels

Normalized label	**Type**	**Free / paid**	**Shared / private**	**Always at the listing / Available upon request**	**Brand**	**Nearby / In building**	**Available all year / Available seasonally**	**Open 24 hours / Open specific hours**	**Outdoor / Indoor**	**Stainless steel**	**Specific attributes of a normalized label**
Access	X (Beach, Lake, Resort)	X	X								Beachfront
Air conditioning	X (Central, Portable, Split type ductless system)										
Baby bath				X							
Backyard			X							Fully fenced / not fully fenced	
BBQ grill	X (Electric, Gas, Wood-burning, Charcoal)		X								
Body soap					X						
Changing table				X							

Normalized label	**Type**	**Free / paid**	**Shared / private**	**Always at the listing / Available upon request**	**Brand**	**Nearby / In building**	**Available all year / Available seasonally**	**Open 24 hours / Open specific hours**	**Outdoor / Indoor**	**Stainless steel**	**Specific attributes of a normalized label**
Children's books and toys	X (0-2 years old, 2-5 years old, 5-10 years old, 10+ years old)										
Clothing storage	X (Closet, Wardrobe, Dresser, Walk-in closet)										
Coffee maker	X (Espresso, Drip, Pour-over, French press)			X							
Conditioner					X						
Crib		X		X							
Dryer		X									

Normalized label	**Type**	**Free / paid**	**Shared / private**	**Always at the listing / Available upon request**	**Brand**	**Nearby / In building**	**Available all year / Available seasonally**	**Open 24 hours / Open specific hours**	**Outdoor / Indoor**	**Stainless steel**	**Specific attributes of a normalized label**
Exercise equipment	X (Free weights, Yoga mat, Stationary bike, Elliptical, Treadmill)										
Game console					X						
Gym							X				
Heating	X (Central, Radiant, Split type ductless system)										
High chair	X (Standalone, Folding / Convertible, Booster seat, Clamp on table)	X	X								

Normalized label	Type	Free / paid	Shared / private	Always at the listing / Available upon request	Brand	Nearby / In building	Available all year / Available seasonally	Open 24 hours / Open specific hours	Outdoor / Indoor	Stainless steel	Specific attributes of a normalized label
Hot tub			X		X	X					
Indoor fireplace	X (Electric, Gas, Wood-burning, Ethanol, Pellet stove)										
Kitchen			X				X				
Microwave / oven					X						
Oven					X				X	Double / single	
Pack 'n Play / travel crib		X		X							
Parking	X (Street, Lot, Garage, Driveway, Valet)	X									On premises / off premises
Patio or balcony			X								

Normalized label	**Type**	**Free / paid**	**Shared / private**	**Always at the listing / Available upon request**	**Brand**	**Nearby / In building**	**Available all year / Available seasonally**	**Open 24 hours / Open specific hours**	**Outdoor / Indoor**	**Stainless steel**	**Specific attributes of a normalized label**
Pool	X (Lap, Olympic)	X			X	X	X		Heated, pool cover, pool toys, rooftop, infinity pool, saltwater		
Refrigerator					X						
Sauna			X								
Shampoo					X						
Shampoo / shower gel					X						
Sound system	X (Bluetooth, Aux, Bluetooth and aux)			X							
Stove	X (Electric, Gas, Induction)							X			

Normalized label	**Type**	**Free / paid**	**Shared / private**	**Always at the listing / Available upon request**	**Brand**	**Nearby / In building**	**Available all year / Available seasonally**	**Open 24 hours / Open specific hours**	**Outdoor / Indoor**	**Stainless steel**	**Specific attributes of a normalized label**
TV	X (HDTV, TV)								Standard cable / premium cable, size, streaming services		

Normalized label	**Type**	**Free / paid**	**Shared / private**	**Always at the listing / Available upon request**	**Brand**	**Nearby / In building**	**Available all year / Available seasonally**	**Open 24 hours / Open specific hours**	**Outdoor / Indoor**	**Stainless steel**	**Specific attributes of a normalized label**
View	X (City skyline, Garden, River, Courtyard, Ocean, Sea, Pool, Mountain, Beach, Valley, Bay, Park, Harbor, Marina, Vineyard, Canal, Lake, Resort, Golf course, Desert)										
Washer		X					X				
Wifi	X (Pocket, Fast)										Mbps