UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Faculty of Business Administration and Management

Analysis and segmentation of the hotel offers in Valencia using TripAdvisor's key terms

Master's Thesis

Master's Degree in Business, Product and Service Management

AUTHOR: Tu, Yue

Tutor: Trull Domínguez, Óscar

ACADEMIC YEAR: 2023/2024

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Faculty of Business Administration and Management

Analysis and segmentation of the hotel offers in Valencia using

TripAdvisor's key terms

Master's Thesis

Master's Degree in Business, Product and Service Management

AUTHOR: Tu, Yue

Tutor: Trull Domínguez, Óscar

ACADEMIC YEAR:

Table of Content

Yue Tu

Abstract

In today's digital age, user-generated comments on the Internet play an important role in shaping potential customers' perceptions of hotels and influencing their booking decisions. This trend presents both challenges and opportunities for the hotel industry, especially in the post-pandemic era where increased uncertainty and evolving customer expectations have intensified competition. To gain insights into the customer satisfaction levels and customer focus in Valencia's star-rated hotels during this new era, the study utilizes online customer rating scores and reviews written in English about the star-rated hotels in Valencia, sourced from the TripAdvisor portal. This thesis conducts a descriptive analysis of customer satisfaction levels based on scores and to perform text mining analysis of customer reviews. The results show that 3-star hotels have the highest customer satisfaction levels, followed by 5-star, 4-star and 2-star hotels. Satisfaction levels tend to peak during off-peak travel seasons, while lower levels are observed during peak periods. In terms of customer types, couple and family generally exhibit the highest satisfaction scores, followed by friends, business and solo groups. Regarding customer focus, 2-star category prioritizes location, while customer experience is highly valued in 3-star, 4-star, and 5-star hotels. Across all hotel categories, recurring terms such as "hotel," "room," "staff," "stay," "location," and "great" consistently emerge in customer reviews, indicating the core needs that require attention from hoteliers. The findings empower hoteliers and tourism professionals to enhance competitiveness through tailored services that meet customer demands effectively.

Key words: Customer satisfaction, star-rated hotel, text mining, LDA, Reviews

List of figures

1. Introduction

The COVID-19 pandemic has significantly altered various aspects of people's lives. Travel restrictions implemented during the pandemic have had a substantial impact, particularly on the tourism industry, including the hotel sector. In 2020, the number of tourists in the Valencian Community reached its lowest point in nearly five years, with a 73.3% decrease in international visitors compared to 2019 and a 41.3% decrease in domestic tourists [1].

The onset of the post-pandemic era is characterized by the implementation of various measures and policies aimed at combating the pandemic, including periods of lockdown, subsequent reopening, and strategic initiatives designed to stimulate economic recovery. This is a transitional period when the global and national economy will return to normal and full-scale business operations after the devastating disruption of workplace and businesses by the COVID-19 pandemic [2].

Starting from 2021, the tourism industry in the Valencian Community has shown signs of recovery, with tourism demand increasing annually. On 20/09/2022, the Spain Travel Health (SpTH) portal, both the website and the mobile applications, ceased operations. It was unnecessary to fill in the health control form to travel to Spain nor to present the SpTH QR code when boarding the plane or upon arrival at the airport. This measure canceled the restrictions and made it more convenient for tourists to visit Spain. It further promoted the recovery of the tourism in post-pandemic era. In 2022, tourism contributed 15.8% to the Valencian Community's GDP, reaching levels comparable to the pre-pandemic year of 2019 (15.7%) [1].

On 04/07/2023, the Spanish government declared an end to the health crisis caused by

the COVID-19 pandemic, and announced that people no longer had to wear masks in health care centers, hospitals and chemists. These measures signified a more favorable socioeconomic environment for the advancement of tourism in Spain. Based on the report presented by Valencian Community Tourism at Fitur 2024, It is quite likely that, for the first time, the contribution of tourism to the CV's GDP may have exceeded 16% in 2023 [1].

This thesis focuses on post-pandemic era when tourism is gradually recovering from pandemic's impact. During this stage, hotel management undergoes adjustments to comply with health measures. Travelers' preferences about choices of hotels and evaluations of hotels also shift compared to the pre-pandemic period. In order to offer a thorough understanding of customer satisfaction levels in Valencia's star-rated hotels from the customers' views in post-pandemic era, the thesis gathers and analyzes the hotels' score data and reviews written in English from TripAdvisor portal. Simultaneously, it examines the differences in hotel attribute topics across various star categories. The aim is to provide valuable insights to guide future improvements in the hospitality sector of Valencia.

Additionally, the motivation behind this thesis stems from the author's pursuit of a master's degree in Valencia. The intention is to contribute to the local academic researches of Valencia. Furthermore, the thesis aligns with the courses in big data, operation management, and strategic management, aiming to apply knowledge practically to actual business by providing strategic and decision analysis. Finally, the author's personal interest in browsing others' reviews to assess the value of products, restaurants, and hotels, combined with previous studies in tourism management during undergraduate studies, inspires the theme of this thesis.

2. Overall Environment of the Hotel Industry

The hotel industry is vulnerable and susceptible to threats such as pandemics, natural disasters, and terrorist attacks. Compared to previous incidents like natural disasters or pandemics SARS, the impact of COVID-19 is unprecedented. The pandemic and severe travel restrictions have taken a toll on the global hospitality industry. As Rivera [3] commented, the pandemic is affecting the DNA of hospitality at its core. Hospitality is about "host and guest", "coming together", "tangible and intangible", and "providing security, psychological, and physiological comfort" [4]. But these characteristics have been weakened and mitigated during the pandemic. Pillai et al. [5] think hygiene, cleanliness, and safety protocols are the most important measures taken by businesses in the hospitality and tourism industry, both during and post epidemic/pandemic times. Besides, technologies also play an important role in effective disaster management. In the future, guests are even more likely to be conscious of personal hygiene and social distancing. In the post-pandemic era, while the hotel industry is experiencing rapid recovery and development, it is also marked by uncertainty due to changing customer demands and the evolving environment. To tackle these challenges, the hotel industry must thoroughly comprehend the shifts in the landscape, identify shortcomings, and enhance strategies and operations to stand resilient in the face of uncertainties.

In 2023, the results of tourism businesses across Spain are growing due to increased sales and efforts in cost containment, despite a new rise in costs: financial costs (+18.9%) and labor costs (+7.3%) are on the rise, while supply costs (+9.6%) are moderating, and energy costs are falling (-2.8%) [1]. The issue of costs is not to be underestimated in the hotel industry.

2.1 Valencia Hotel Industry Situation

Valencia city enjoys a unique and advantageous geographical climate, that allows visitors to enjoy its natural and cultural offer at any time of the year. It has abundant natural and cultural tourism resources, particularly with its wealth of cultural heritage, such as the renowned Fallas festival, which attracts visitors from around the world. With well-developed international and domestic transportation networks, cruise tourism and MICE (Meetings, Incentives, Conferences and Exhibitions) tourism also hold significant positions.

In 2022, the tourism sector in Valencia confirmed its recovery after two very complicated years. The city reached figures for overnight stays and travelers close to pre-pandemic records 5,325,710 and 2,197,133 respectively. The recovery was largely supported by the domestic market, with growth of 19.2% over 2019. Although international overnight stays still showed slightly lower figures compared to 2019 (-10.6%), they evolved positively compared to 2021. Besides, Valencia made progress in two key areas: the commitment to sustainability and the creation of value by strengthening an offer closely linked to the city's most identifying resources. Also, during this year, Valencia held the prestigious titles of World Design Capital 2022 and European Capital of Smart Tourism. Furthermore, it retained its recognition by InterNations as the best city in the world to live in [6].

In 2023, tourist numbers reached new heights, with the total number of visitors reaching 2,319,546, representing a 5.1% increase from the previous year. International tourists showed a significant growth of 10.9%, while domestic tourists experienced a slight decline of 1%. Hotel accommodations saw a positive trend in hosting both international and domestic visitors, totaling 2,072,575, marking a 7.2% increase compared to the previous year. Particularly noteworthy was the international market, which saw a surge

of 12.4%. The overall overnight stays for the year reached 5,493,065, showing a marginal growth of 1.4% from 2022. In this period, the international market experienced a notable increase of 7.9%, whereas the domestic market decreased by 8.4%. The total overnight stays in hotels for both international and domestic visitors amounted to 4,684,432, reflecting a 5.4% increase. The international market witnessed a remarkable growth of 10.2%, while the domestic market experienced a 2% decline. The year 2023 witnessed further recovery in the international market [7].

Besides, the increases in hotel main indicators are also obvious. the hotel room occupancy rate for the entire year reached a new high in nearly five years, hitting 79.5%, surpassing the pre-pandemic level of 2019 (76.9%). The average room price and RevPAR (Revenue per Available Room) for all hotels and lodgings showed an upward trend. The average room price was €112.18, representing a 14.4% increase from the previous year, while the RevPAR stood at €90.33, marking an 18.8% rise. Notably, 3 and 4-star hotels exhibited particularly significant growth in RevPAR, reaching 20.8% and 21.3%, respectively, indicating further strengthening of financial performance in these star-rated hotels [7]. Star-rated hotels have shaken off the shadow of the pandemic and are experiencing a vibrant resurgence.

2.2   Expectations and Trends in 2024

Based on the report presented by Valencian Community Tourism at Fitur 2024, business expectations for 2024 foresee another year of significant improvement in sun and beach destinations, with the presence of domestic and European holiday demand, especially along the coast of the Valencian Community, Andalusia, and the Balearic archipelago. This growth is driven by enhanced offerings and already elevated booking states from key markets.

As for the tendencies, Increasingly, travel is focused on the experiences that will be enjoyed. Travelers seek to discover new things, experience novel sensations, and project them into lasting memories. Interest in gastronomy and local products also cannot be overlooked.

Trends more relevant to hotels include: travelers demonstrate a greater willingness to invest in luxury and wellness experiences. This encompasses trips to exclusive resorts, spa treatments, yoga or meditation classes, etc. Valencia offers an increasingly relevant luxury and wellness tourism offering, with high-quality hotels, resorts, and spas. This represents a significant opportunity.

Besides, an increase in sustainability and social responsibility in supply and demand poses higher demands for the hotel industry. Hotels need to intensify efforts to adapt to and mitigate the impact of climate change, enhancing competitiveness.

As a disruptive element throughout the tourism value chain, technology is changing the way travelers book and pay for their trips, and it is becoming increasingly relevant in enhancing the travel experience. AI will enable greater personalization and more detailed market segmentation. The tourism sector is characterized by a gradual integration of technology into its internal and external processes. In the contemporary tourism landscape, there's a notable shift towards a data-driven environment, where tourism businesses and destinations are aware that information and knowledge management are crucial factors for their competitiveness [1]. These trends in information technology bring both significant opportunities and challenges to the hotel industry. Hotel enterprises need to enhance their digital transformation, leverage big data to understand customer expectations and preferences, and meet customer experiences effectively, thereby enhancing their competitiveness.

3   Objectives

This thesis conducts a descriptive analysis of customer satisfaction levels based on scores and performs text mining analysis of customer reviews. Its objective is to assess similarities and differences in customer satisfaction levels and attention to various aspects of star-rated hotels in Valencia. The study utilizes online scores and online travel reviews (OTRs) written in English about the star-rated hotels in Valencia, sourced from the TripAdvisor portal.

The findings provide a more comprehensive understanding of the hotel industry's current situation in Valencia and internal variations from customers' views in the post-pandemic era, helping government officials from the tourism authority gain a more informed perspective. Additionally, understanding the evolving expectations and preferences of customers in the new era enables hoteliers and tourism industry professionals to meet the demands more effectively and enhance competitiveness.

As a guideline, the following questions will be answered in the thesis:

- What are the differences in TripAdvisor scores among hotels in Valencia depending on factors such as hotel star category, travel month, and customer type?
- What are the differences in the topics customers focus on in reviews of different star-rated hotels?
- What are the most frequent key terms in hotel reviews?

3.1 State of the Art

The influence of online platforms as mass influencers is steadily growing as users continue to express their opinions through Web 2.0 applications. Online review websites, a prevalent type of Web 2.0 application, enable users to efficiently share their post-purchase experiences with products and services within online communities at minimal cost.

TripAdvisor dedicated to the rating of hotels and restaurants, stands as one of the most renowned online review platforms with immense popularity. According to Similarweb, a free online platform that provides website ranking and competitive data analysis, in September 2023, Tripadvisor is the world's largest travel guidance platform. It assists hundreds of millions of people monthly in exploring activities, hotels, and restaurants based on insights from previous visitors. In Dec 2023, tripadvisor.es was ranked 64 among the most visited websites in Spain, demonstrating its importance and influence.

TripAdvisor solicits ratings from travelers in two parts: an evaluation score typically ranging from one (Terrible) to five (Excellent), and a written description. Rating patterns can be further influenced by travelers' expectation-experience congruence. When their experience exceeds their expectation, it often manifests as satisfaction reflected in their ratings. Conversely, when their experience fails to meet their expectation, their discontent may be conveyed through their ratings [8].

In the digital era, Word-of-mouth (WOM) behavior attracts attention and is closely linked to loyalty. According to Anderson [9], WOM is an information communication between private parties regarding the evaluation of products and services. Substantial empirical evidence exists in the tourism context revealing that online WOM consultations influence choice decisions [10]. User-generated content (UGC) on

Yue Tu

TripAdvisor is influencing potential customers' assessment of hotels and booking decisions greatly, presenting both challenges and opportunities for a hotel. Through analysis of the scores and reviews left by customers, hotels can get valuable information about customers' preferences, strengths, and weaknesses in hotel management, enabling them to improve their operations, and ultimately establish a positive reputation and image among customers.

Due to the importance of UGC, many researches about consumer psychology, behaviors, customer rating patterns, and content analysis have developed. Studies are focusing on the psychological aspects of comments on social media, as well as variations in ratings among different customer types and regions. Regarding user textual data, employing techniques such as topic modeling and programming helps identify the focal points and pain points of hotel guests. These studies greatly enrich the literature on tourism, enhancing the connection between hotels and customers, and fully leveraging the characteristics of the big data era to achieve a customer-centric approach in the hotel industry.

4   Literature Review

This section first introduces the primary research topics and trends in the Spanish hotel industry, followed by an examination of the impact of the pandemic on the tourism industry. Subsequently, it delves into studies on customer satisfaction, and finally, it addresses research related to Electronic Word-of-Mouth (eWOM).

4.1 Research Topics and Trends in the Spanish Hotel Industry

In the star-rated hotel literature concerning Valencia, studies are scarce, with a focus primarily on green sustainable practices [11]. As for research on star-rated hotels in Spain, there is a concentration on green initiatives, quality, human resource management, big data analytics, and revenue economics (e.g. [12,13,14]). There is also a considerable amount of analysis concerning online rating patterns and customer reviews. Sánchez-Vargas et al. [15] evaluated the titles of the reviews published on TripAdvisor about 3 and 4-star hotels of a World Heritage City (Cáceres, Extremadura, Spain) to outline the attributes most valued by tourists. Marrero et al. [16] researched hotels' response strategies for online reviews and found that a positive relationship existed between hotel communication policies and the ratings given by reviewers.

These studies are predominantly centered around popular destinations such as the Canary Islands, Barcelona, Mallorca, and Madrid. Research on 3, 4, and 5-star hotels, especially luxury accommodations, is more prevalent, while there is a dearth of studies on 1 and 2-star establishments. This thesis combines big data analysis of user reviews to fill the gap in Valencia's star-rated hotel social website review analysis, enriching tourism literature.

4.2 Pandemic Impact on the Tourism Industry

The tourism sector serves as both a channel for the transmission of viruses and a casualty of such transmission. Consequently, it stands as one of the industries most profoundly impacted by pandemics and crises [17]. In recent decades, the tourism sector has faced significant pandemics, including outbreaks such as severe acute respiratory syndrome (SARS) from 2002 to 2004, novel influenza A (H1N1) from 2009 to 2010, avian influenza A (H7N9) from 2013 to 2017, and Ebola Virus Disease (EVD) from 2014 to 2016. The impact of COVID-19 has been more severe compared to previous pandemics due to its widespread geographical reach and heightened uncertainty. It's essential to research the impact and enhance the management of the pandemic crisis in anticipation of potential future risks.

Many studies researched travel flows, travel intentions, and revenue patterns during the pandemic. Teng et al. [18] analyzed the impact of perceived risk associated with COVID-19, perceptions of travel risk, and hotel hygiene and safety practices on the intention to stay in those hotels by using questionnaires. Moreno-Izquierdo et al. [19] conducted a spatiotemporal and economic analysis of Madrid and Valencia and found the pandemic had a significant impact on both occupancy and accommodation prices of Airbnb's properties. There is limited research literature on customer expectations during the pandemic. Hu et al. [17] investigated the changes in travelers' expectations and perceptions of hotel services during different stages of COVID-19. This master thesis intends to enrich the study of customer expectations within the context of the post-pandemic era.

4.3 Customer Satisfaction

Oliver [20] introduced the expectancy confirmation theory (ECT) broadly used for

understanding customer satisfaction and service optimization. Per ECT, customer satisfaction is determined by the evaluations customers make, which result from a thorough comparison (confirmation) between their expectations and perceptions regarding product/service attributes. Customers' expectations and evaluations of products or service attributes vary based on their own experiences, profiles, and cultural backgrounds. Galati and Galati [21] found that Italian, Chinese, and American travelers placed different emphasis on hotel attributes. Many studies suggest that travelers' assessments of hotel services may vary depending on the specific service contexts like different star ratings of hotels or varied visiting times. Similarly, the global pandemic may serve as a contextual factor influencing travelers' evaluations of accommodation services [17]. Moreover, the hotel sector delivers uniform services, nurturing a heightened competitive environment. In such a fiercely competitive environment, coupled with the multitude of factors stemming from customer backgrounds and the uncertainty of changing circumstances, customer expectations and evaluations continue to evolve. Hotels must pay attention to the changes in customer satisfaction to better meet their expectations, thereby enhancing customer loyalty and cultivating a positive reputation.

4.4 eWOM

eWOM represents a contemporary mode of informal communication whereby messages are shared with others via social media and various electronic platforms. Compared with traditional WOM methods, customers have access to a wide array of content created by strangers or peers on the Internet. It takes various forms, such as online reviews, social media posts, and emails.

The importance of eWOM in the digital era leads to a significant amount of research.

Researches include the source and the participants involved in this behavior, customers' psychological motivations for consulting eWOM, and the influences of eWOM. Some studies have demonstrated a strong positive correlation between online reputation and the rise in ADR, occupancy, and RevPAR in the hotel industry [22]. Besides, Zhao et al. [23] noted positive causal relationships between online review features (usefulness, reviewer expertise, timeliness, volume, and comprehensiveness) and respondents' online booking intentions. A significantly negative relation between negative online reviews and online booking intentions was identified. The significant impact of online feedback underscores the necessity for hotel managers to pay close attention to and monitor it diligently. This is crucial for enhancing competitiveness and building a positive reputation.

Recent literature in the tourism field has displayed a growing inclination towards searching rating patterns on websites and utilizing text mining technology and sentiment analysis to extract travelers' expectations and perceptions regarding hotel attributes from online reviews. Banerjee and Chua [8] found that travelers' rating patterns differed substantially between independent and chain hotels across various profiles and regions. Business travelers tended to be the most stringent in their rating patterns compared to other profiles. Additionally, travelers with families sometimes exhibited strict rating patterns. Zhang and Verma [24] used regression analysis and text analytics to uncover common consumer concerns and their effects on ratings. Their research concluded that service and rooms were overwhelmingly the most important aspects of high-end properties. Alrawadıeh and Law [25] adopted a hybrid approach of inductive and deductive content analysis of TripAdvisor reviews about the best-rated hotels in Istanbul and found that the quality and size of rooms, along with the service quality from staff, mainly determined guest satisfaction. Hu et al. [17] revealed shifts in the importance travelers place on different hotel attributes during COVID-19 through text mining of online reviews. A few attributes (e.g., breakfast, location) lost importance

in the pandemic periods. On the contrary, certain attributes such as experience, cleanliness, service, and front desk gained significance during the pandemic phases due to the rising demand for indoor experiences and enhanced security and hygiene measures. This master thesis makes an additional contribution to the existing literature on eWOM and sheds light on the concerns of hotel customers in the post-pandemic era.

5 Methodology

The research methods primarily consist of quantitative analysis and qualitative analysis. Quantitative analysis uses descriptive analysis to present the overall situations and differences in customer satisfaction levels among star-rated hotels in Valencia based on TripAdvisor's bubble scores. It also examines satisfaction differences across travel months and among various customer types. Qualitative analysis primarily employs text mining techniques mainly like Latent Dirichlet Allocation (LDA) to reveal hotel attribute topics that customers focus on across different star-rated hotels. It then compares these topics to uncover similarities and differences. Moreover, the Clustering method is used to uncover more details in similar review groups in each star tier, while FP-Growth identifies the frequent items in the reviews to offer insights.

5.1 Data Collection

Based on the official Valencia community tourism website, this thesis collects hotel data including hotel names, and star categories, and registers the time of all the hotel establishments in Valencia municipality for the research regardless of hostels and pensions ending on 12/11/2023.

Subsequently, customer data from TripAdvisor for each hotel abovementioned, with posting dates ranging from 20/09/2022 to 19/09/2023, is documented. Starting from 20/09/2022, the SpTH portal ceased operations, making the health control form and SpTH QR code unnecessary for travel to Spain. On 04/07/2023, the Spanish government declared an end to the health crisis caused by the COVID-19 pandemic. This chosen timeframe specifically targets the period of recovery and development in the Valencia tourism industry post pandemic.

Moreover, the thesis records customer names, rating scores, reviews, travel months, and customer types of 2-star, 3-star, 4-star, and 5-star hotels, except for 1-star hotels. This exclusion is due to the absence of reviews and scores after the onset of COVID-19. Additionally, within the 1-star category, hotels primarily consist of fully furnished apartments.

Thus, the thesis obtains a total of 1,995 user data entries, comprising 62 entries from 2-star hotels, 604 entries from 3-star hotels, 957 entries from 4-star hotels, and 372 entries from 5-star hotels.

5.2   Research Method

5.2.1 Descriptive Analysis

A descriptive statistic is a summary statistic that quantitatively describes or summarizes features from a collection of information [26]. Some measures that are commonly used to describe a data set are frequency distribution analysis, measures of central tendency, and measures of dispersion. Frequency distribution analysis involves generating frequency tables or bar charts to illustrate the occurrence frequency or proportion for each category within categorical variables. Measures of central tendency include the mean, median, and mode, while measures of dispersion include the standard deviation (or variance) to examine the dispersion or variability of the data [27].

The thesis mainly conducts frequency distribution analysis of Valencia star-rated hotel overview and sample sources from each star tier hotel. Next, mean values are utilized to evaluate customer satisfaction levels overall, across different travel months, and among various customer types. Standard deviation is also used to measure the

magnitude of inter-group differences. The final step involves comparing the differences in values to illustrate the differences in the levels of customer satisfaction across different classifications.

5.2.2 Text Mining

Customer reviews are textual and unstructured data. Traditional methods of analyzing customer reviews involve researchers reading through and subjectively interpreting and coding the text, which is impractical and difficult due to the large amount of data in the thesis. To address this challenge, this thesis employs text mining, a process aimed at extracting valuable and meaningful insights from unstructured text, thus overcoming information overload [28]. In this study, different text mining techniques are used to reveal customer insights from the reviews. Initially, the unstructured texts of these reviews are changed to a structured form amenable to analysis, and then natural language processing algorithms are applied.

In this section, this thesis briefly reviews some of the popular analytical techniques associated with text mining. The purpose is to lay the groundwork for the study, as well as to describe the methodology and introduce terms used in the thesis. The techniques utilized involve preprocessing and text representation, LDA, Topic Modeling Tool, Clustering, and FP-growth.

Preprocessing and Text Representation

Arguably, the most crucial stage in analyzing unstructured text documents involves converting free-form text into a structured format suitable for analysis. One commonly used transformation is the "bag of words" representation. This approach disregards linguistic features such as word order, grammar, and other textual attributes for statistical analysis purposes [29]. This process entails representing the documents to be

analyzed, commonly known as a "corpus," using a document-term matrix. This matrix includes a column for each word found in the corpus and a row for every document (in this case, each customer review). Each entry in the matrix indicates the frequency of a specific word occurring in each document.

The document-term matrix is a structured numerical table that can theoretically be analyzed using natural language processing algorithms.

In case the matrix is overly redundant, text mining analyses need preprocessing of the corpus to retain meaningful words while removing uninformative ones. Standard preprocessing steps are as follows: (1) all of the letters in these documents are transformed into lowercase, (2) the documents are tokenized with non-letter separators, (3) the "stop words" (including trivial words such as "and," "the," "is," "are," "a," and "an,") in the identified word list are removed,

(4) all tokens that are less than two letters (e.g., "s," "x") are deleted, (5) any words or tokens that only appear in one document are eliminated, (6) term-stemming techniques are applied to a word list. Term-stemming identifies the word's root and regards all words with the same root as one token, and (7) an n-gram algorithm is applied to identify repeated phrases (e.g., friendly staff, good location, or poor Wi-Fi) in the documents [30].

LDA

As outlined by Blei and Lafferty [31], topic modeling algorithms automatically summarize large archives of text by discovering hidden topics or themes found within a set of documents. **Figure 1** shows a workflow of the typical theme model. LDA stands as a robust and extensively utilized algorithm for topic modeling. It deduces the concealed topic structure from the original texts employing a probabilistic framework.

LDA is based on probabilistic modeling and the observed variables are the bags of words per article whereas hidden random variables are the topic distribution per article. The main goal of LDA is to compute the posterior of the hidden variables given the value of the observed variables [32]. The assumptions of LDA for the study are (1) articles with similar topics will use similar groups of words, (2) articles are a probability distribution over latent topics, and (3) topics are probability distributions over words [33].
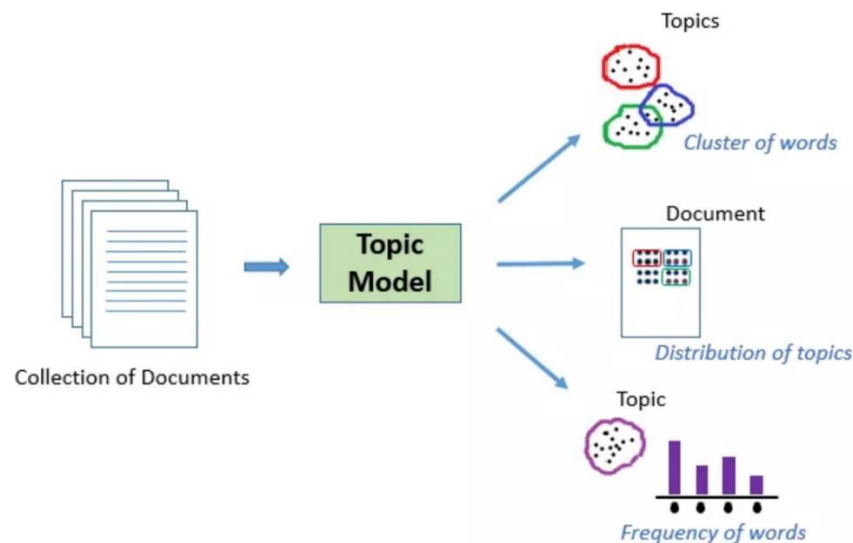
Figure 1 A workflow of the typical theme model [34]

The advantage of LDA is that it is unsupervised machine learning without relying on manually annotated training data. In other words, the only inputs are the set of documents and the number of topics. In addition, LDA can always find representative words for each topic [35]. Moreover, LDA yields more informative summaries of textual data compared to alternative approaches. This is mainly because the LDA model is based on more realistic assumptions and offers greater flexibility. For instance, LDA can represent each text using multiple topics rather than a single one.

The underlying concept of this method is that all documents share the same set of topics, but each document demonstrates a unique probabilistic blend of these topics. Put simply,

certain words are more probable to be associated with certain topics. LDA utilizes a Bayesian estimation framework to infer these topics (word distributions) from the text and decomposes each document into a mixture of topics. In this thesis, the outputs of LDA are two probability distributions: P(topic|review), which is the probability distribution of topics for a given review, and P(word|topic), which is the probability distribution of words for a given topic [29].

LDA has been widely utilized in research related to user review studies, and there are also studies in the hotel industry that have employed LDA. Han et al. [29] analyzed hotel reviews in Moscow using the LDA method and found that different hotel tiers prioritized specific attributes, with high-tier hotels emphasizing guest experience and mid-tier motels focusing on amenities and location. Guo et al. [36] utilized LDA to uncover 19 key dimensions of customer service from a dataset of 266,544 online reviews spanning 25,670 hotels across 16 countries.

Topic modeling tool (TMT)

Topic Modeling Tool is a graphical interface tool designed for conducting LDA topic modeling. It is powered by Java and serves as a user-friendly application for topic modeling, utilizing the MALLET toolkit as its back-end [37]. All the star-rated reviews are first converted into text format and then processed using TMT by star ratings. In the toolkit, the following parameters are fixed for the study: (1) Number of iterations: 200, (2) Number of topic words printed: 5, and (3) Topic proportion threshold: 0.05. The output files are then analyzed to extract insights and categorize the group of words into relevant topics. Moreover, representative reviews are concurrently reviewed to accurately label the topics.

Clustering

Clustering involves dividing unlabeled data or data points into distinct clusters, where similar data points are grouped while dissimilar ones are separated into different clusters. Simply put, the goal of clustering is to identify and group data with similar characteristics into clusters. Centroid Models represent one of the widely used clustering techniques. These clustering algorithms iterate, deriving similarity from the proximity of a data point to the centroid or cluster center. The k-means clustering algorithm, a popular example, falls into this category. Determining the number of clusters is essential for these models, necessitating a prior understanding of the dataset. They repeatedly execute to identify local optima [38].

The k-means algorithm identifies k clusters and assigns each example (in this case, customer review) to exactly one cluster based on their similarity. Each cluster includes examples with similar attributes, determined by the centroid's position in the n-dimensional space of the example set. Examples are then assigned to the nearest cluster, and centroids are recalculated by averaging over all examples in the cluster. This process repeats until centroids no longer move or the maximum optimization steps are reached. The algorithm is run multiple times with different start points to find the set of clusters with the minimal sum of squared distances to their centroids [39].

In the thesis, k-means clustering is conducted by RapidMiner Studio, a leading data mining tool. Through clustering, more details are revealed to understand each star hotel's reviews' key aspects.

FP-Growth

The FP-Growth algorithm is a widely used technique for mining frequent patterns in data mining. It operates by creating a condensed representation of the dataset known as a frequent pattern tree (FP-tree). This tree captures the frequency and correlation details of items within the dataset.

Initially, the algorithm examines the dataset, associating each transaction (in this case, customer review) with a path in the FP-tree. Items within transactions are arranged based on their occurrence frequency, prioritizing the most frequent items. Subsequently, through recursive exploration of the FP-tree, frequent item sets are derived. This process entails traversing the tree from the bottom upwards, identifying all combinations of item sets that meet the minimum support threshold [40]. In this case, support = (Number of times an item or item sets appear in the database) / (Number of reviews in the database). In general, the concept of "minimum support" creates a cutoff, defining what is meant by frequent or not-so-frequent occurrences of an itemset [41].

In the thesis, FP-Growth is used to illustrate the most frequent words and frequent phrases occurring together in customers' reviews in each star-rated hotel by RapidMiner Studio.

6. Results

6.1 Descriptive Analysis

6.1.1 Valencia Star-rated Hotels Overview

There are 384 star-rated hotels in total in Valencia municipality until 12/11/2023. As for the star categories shown in **figure 2**, nearly half of them are 4-star followed by 3-star, 2-star, 5-star, and 1-star. 1-star only accounts for 2%.
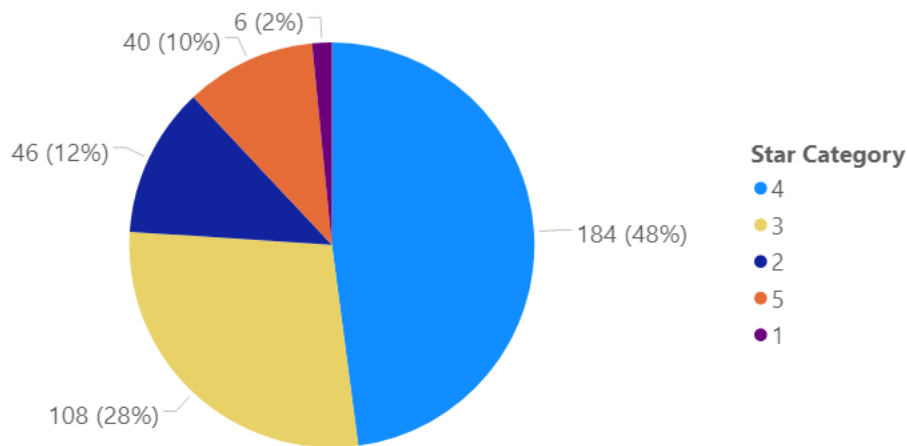
Figure 2 Proportion of star-rated hotels in Valencia

According to the register year in **figure 3**, from 1990 to 2007, there was steady growth, reaching a peak of 9 new hotels in 2007. Between 2007 and 2012, the growth rate slowed down. After 2012, there was fluctuating but overall increasing growth. In 2023, the number of new hotels added was 8, showing a positive trend in growth.
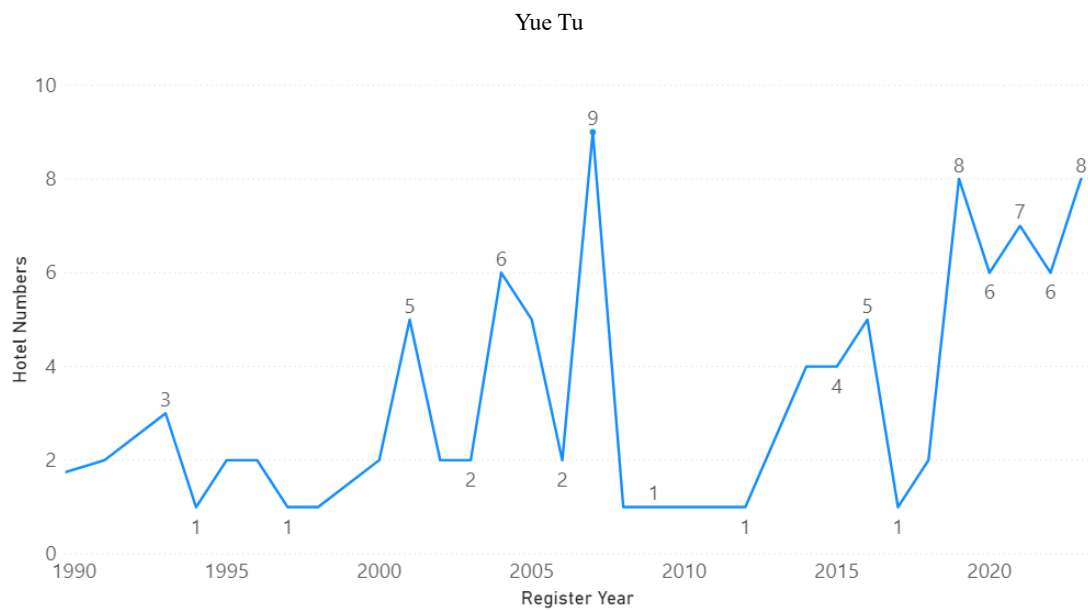
Figure 3 The number of star-rated hotels in Valencia by year of registration

6.1.2 Sources and Proportions of Scores in Star-rated Hotels

This section unveils the sources and distribution of customer scores of each star category, providing insights into the individual contributions of specific hotels to the overall scores in each category.

According to the chosen one-year timeframe in the TripAdvisor portal, 1995 scores in total are collected for the research. **Figure 4** shows the distribution of customer scores across each star category. In general, 4-star hotels have the highest number of scores among all the hotels and take nearly half of the total scores, followed by 3-star, 5-star, and 2-star.
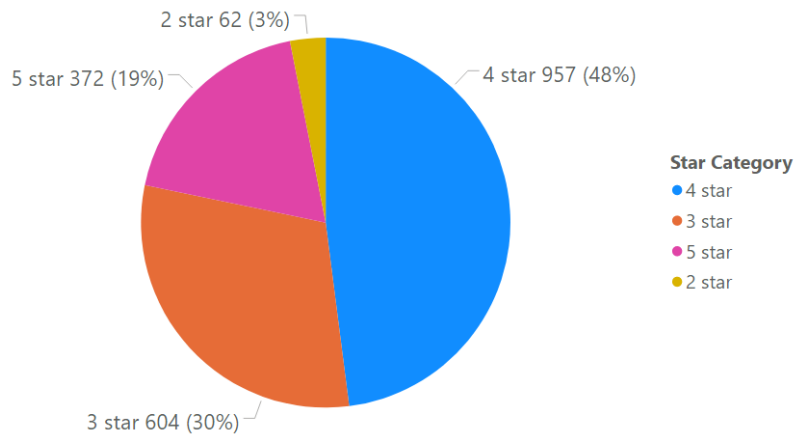
Figure 4 The number of samples for each star category

Breaking it down for each star rating, in figure 5, 2-star hotels have accumulated 62 customer scores for 12 hotels in total. The top three hotels with the highest number of scores are Casual Vintage Valencia, Soho Valencia, and Casual del Cine Valencia, accounting for 24%, 18%, and 13% respectively.
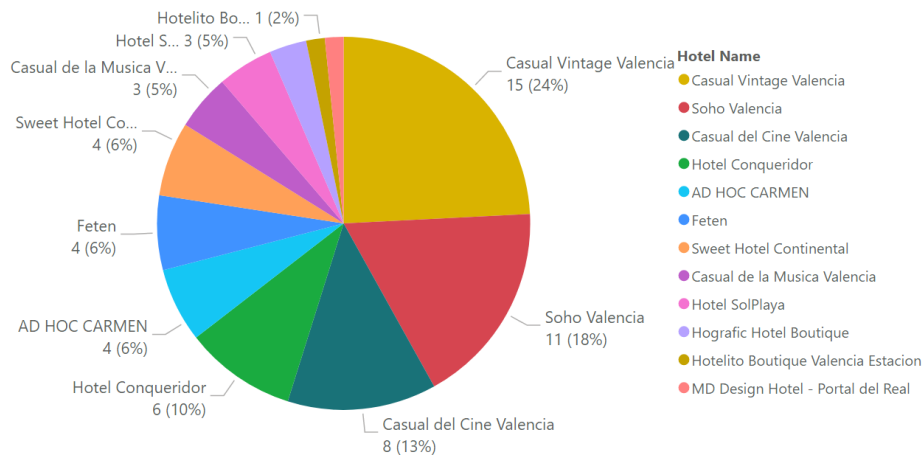
Figure 5 The composition of 2-star hotels

For 3-star hotels in **figure 6**, 604 customer scores for 32 hotels are collected in total. The top five hotels with the highest count of scores are Hotel Malcom and Barret, Cosmo Hotel & Bar, Hotel Kramer, Petit Palace Plaza de la Reina, and Hotel Ilunion

Aqua 3, constituting 26%, 24%, 13%, 6%, and 4% respectively.
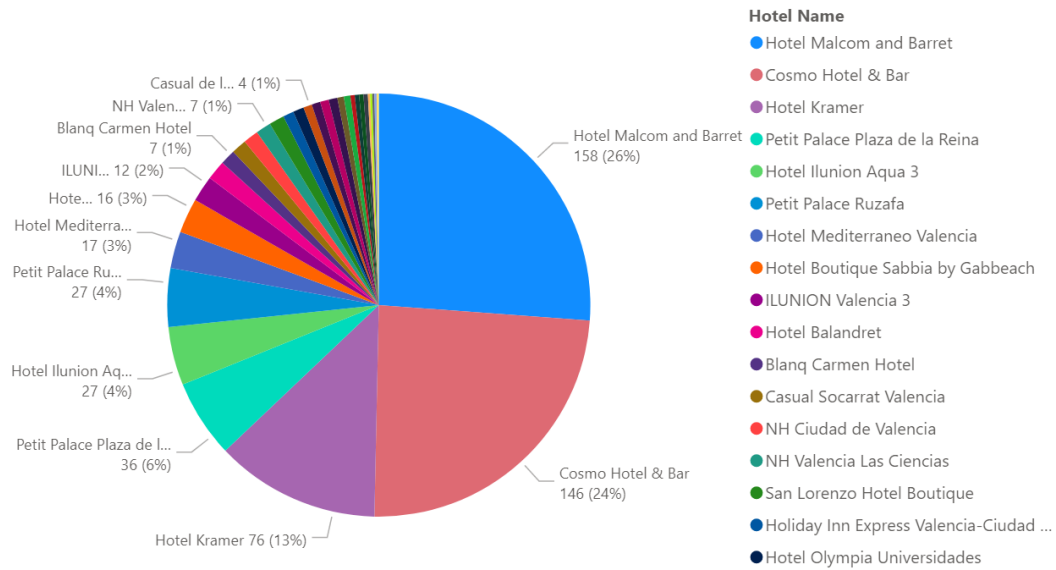
Figure 6 The composition of 3-star hotels

For 4-star hotels in **figure 7**, 957 customer scores for 45 hotels are collected in total. The top five hotels with the greatest quantity of scores are Helen Berger Boutique Hotel, Barcelo Valencia, Vincci Mercat, Hotel Melia Valencia, and SH Ingles, accounting for 18%, 7%, 5%, 5%, and 4% respectively. Compared with other star categories, 4-star hotels exhibit relatively dispersed hotel score data, not concentrated in a few representative hotels.
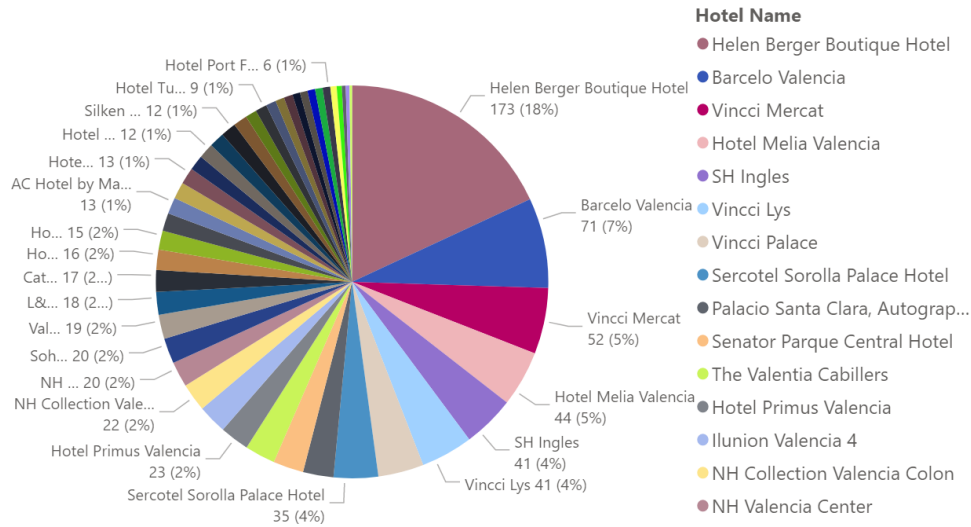
Figure 7 The composition of 4-star hotels

For 5-star hotels in **figure 8**, 372 customer scores for 8 hotels are collected in total. The top three hotels with the highest number of scores are Hotel Las Arenas Balneario Resort, Only YOU Hotel Valencia, and SH Valencia Palace, accounting for 26%, 20%, and 16% respectively.
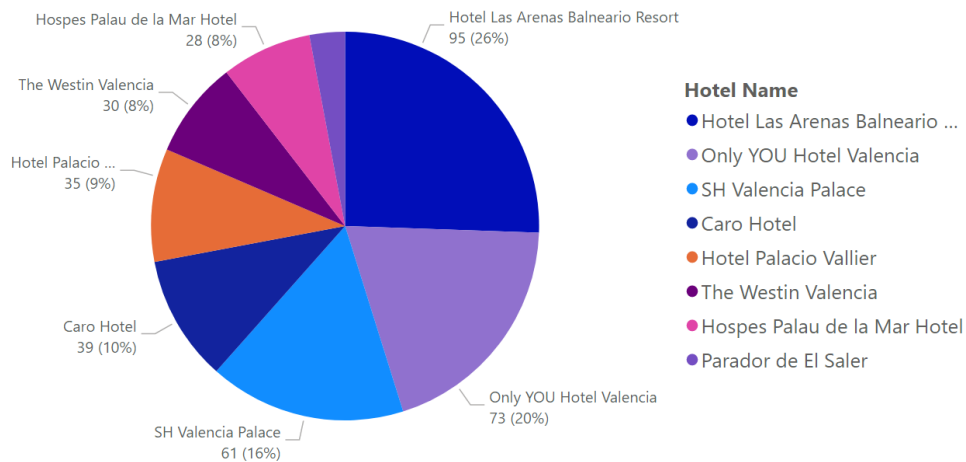
Figure 8 The composition of 5-star hotels

To sum up, 4-star hotels emerge as the most frequently rated category in the study, showing a relatively dispersed and varied composition of hotels. Rest star-rated hotels consist of more representative hotels with greater influence.

6.1.3 Customer Satisfaction Levels across Star-rated Hotels

In this section, average values and standard deviations are used to showcase the customer satisfaction levels across star-rated hotels. By comparative analysis, the thesis identifies the satisfaction level differences among hotels. The analysis is conducted considering overall scores, scores based on travel months, and scores categorized by customer types.

Overall Satisfaction Levels across Star-rated Hotels

The average score of all the hotels is 4.43 out of 5, indicating a generally high level of satisfaction in the hotel sector of Valencia. Among them, the average score of 3-star hotels is the highest, followed by 5-star, 4-star and 2-star hotels in **figure 9**. The scores of 3-star and 5-star hotels are higher than the overall average score showing a higher performance than the rest stars. 4-star and 2-star are lower than the overall average score. Especially, 2-star hotel's average score is 3.87 far below the average level. **This indicates a higher incidence of issues, potentially impacting customer satisfaction.**
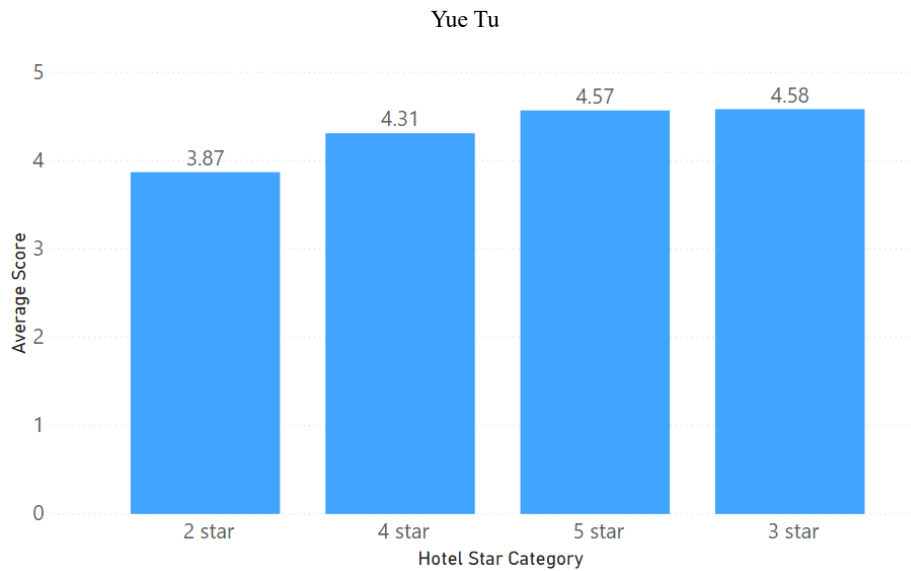
Figure 9 Customer satisfaction levels of hotels in Valencia by star category

Besides, the standard deviation is used to measure the magnitude of inter-group differences, assessing whether there are significant differences within each group of star-rated hotels.

From **figure 10**, for 3-star and 5-star hotels, standard deviations are lower, indicating fewer score differences among individual hotels inside the two groups. 2-star hotels have the highest standard deviation, showing the greatest variation in customer satisfaction scores among hotels within this group. The internal performance levels are more uneven.
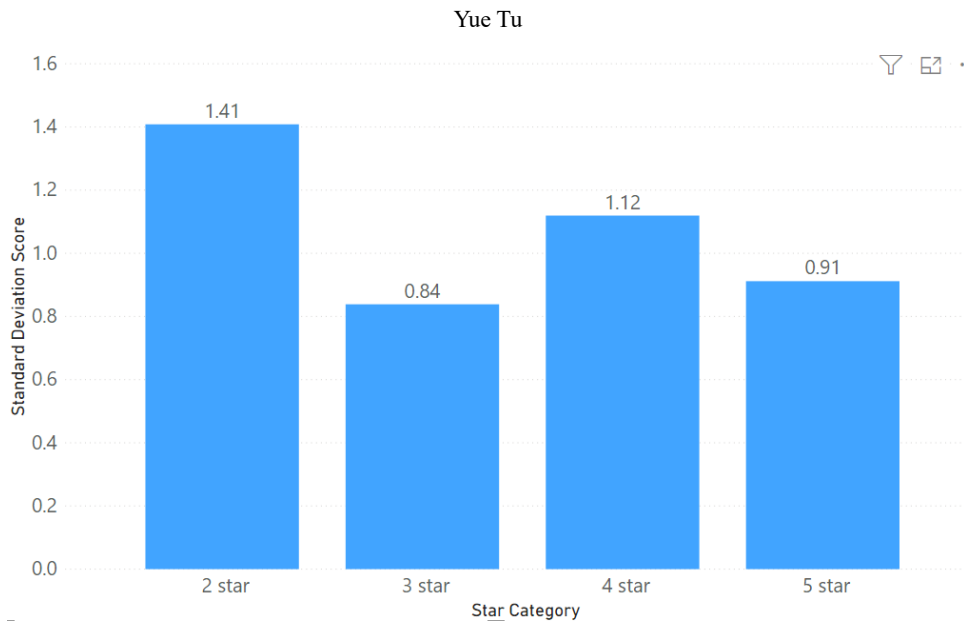
Figure 10 Standard deviations of hotels in Valencia by star category

Hotel Satisfaction Level based on Travel Month

Figure 11 shows the distribution of the number of customer scores in each month. The highest number of customer scores is recorded in July, followed by September, May, April, and March. The months from November to February exhibit relatively lower figures in the dataset, which aligns with the fact that it is the off-peak tourism season in Valencia during winter.
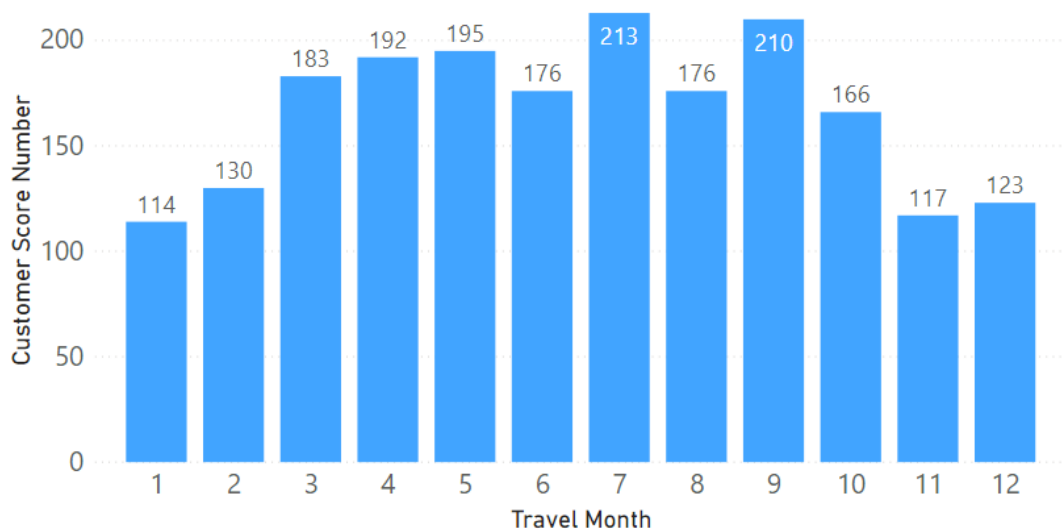
Figure 11 Monthly distribution of customer review counts

Figure 12 reveals customer average scores in each month. Overall, the average score for all the months is 4.43 out of 5. January has the highest customer satisfaction level, followed by February, April, July, May and June. The rest months have lower scores than the overall average score especially October, which is the least satisfactory month. In general, customer satisfaction peaks from December to February, followed by March to May and June to August, with the lowest scores observed from September to November. **More measures can be taken during this period to enhance the performance.**
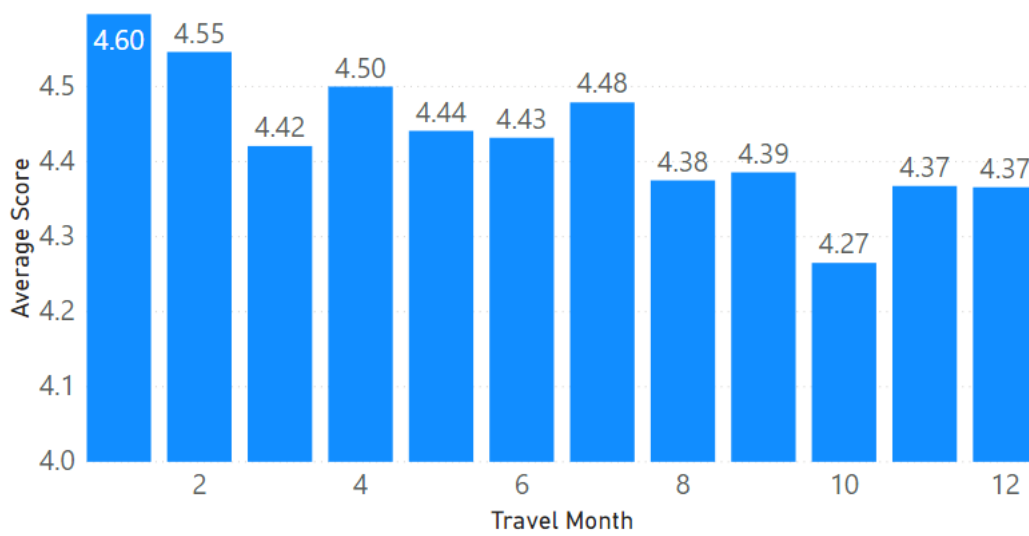
Figure 12 Average customer scores by month

Figure 13 shows the trend of satisfaction scores for each star category according to the month. **The 2-star category exhibits the most significant fluctuations throughout the year, indicating considerable variability.** In contrast, 4-star hotels maintain the most consistent scores across the whole year, followed by 3-star and 5-star hotels.
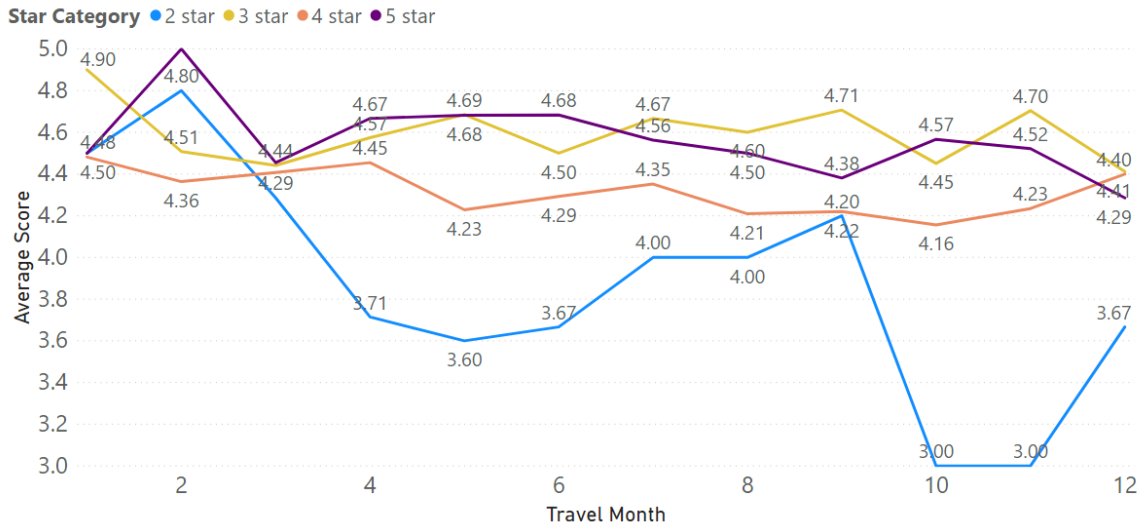
Figure 13 Satisfaction score trends by star category and month

Hotel Satisfaction Level based on Customer Type

Figure 14 illustrates the distribution of the number of customer scores in each customer type Including all the hotels. In general, couple constitutes a significant portion, representing 51% of customer types, followed by family, friends, business, and solo.
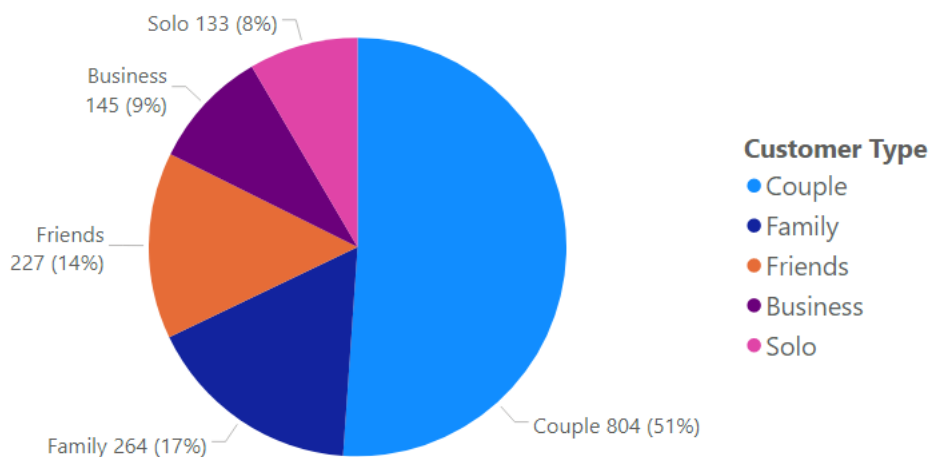
Figure 14 Composition of hotel customer types

As for the overall customer satisfaction levels revealed in **figure 15**, couple and family have the highest score, followed by friends, business, and solo.
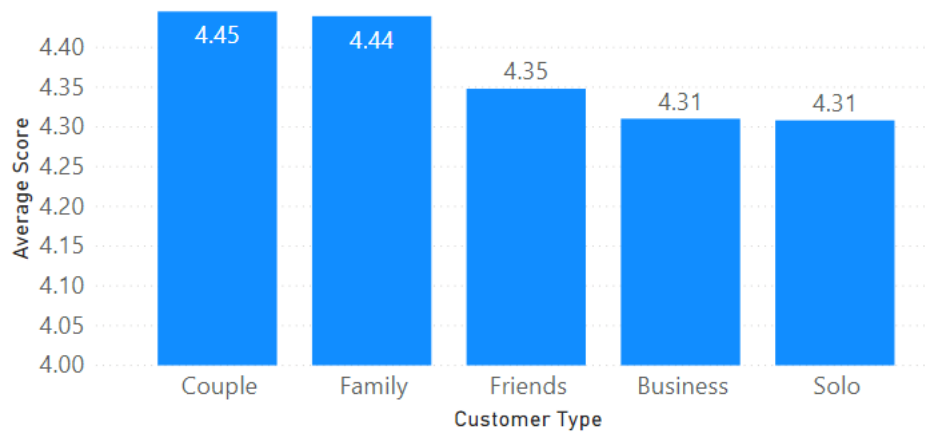
Figure 15 Overall satisfaction levels of different customer types

The situation varies between each star-rated hotel as shown in **figure 16** and **17**. For 2-star hotels, couple, friends, and solo are the most common customer types. Family has the highest satisfaction level followed by couple, friends, solo, and business. **The scores are lower than 4, except for the family type. The scores of solo and business are far below the average level of all the hotels.**

For 3-star hotels, couple, family, and friends are the most common customer types. Family has the highest satisfaction level followed by couple, friends, business, and solo. The solo group has the lowest satisfaction level but is still higher than the average score of all the hotels. **The satisfaction gap among different customer types in this star category is not significant, and the satisfaction level of each type is relatively high.**

For 4-star hotels, couple, family, and friends are the most common customer types. Couple has the highest satisfaction level followed by solo, family, business, and friends. Couple group has the highest satisfaction level but still lower than the average score of all the hotels. The satisfaction gap among different customer types in this star category is not significant, and the satisfaction level of each type is relatively low. **Friends group**

also needs to be paid more attention.

For 5-star hotels, couple, family, and friends are the most common customer types. Friends group has the highest satisfaction level followed by family, couple, business, and solo. Solo group has the lowest satisfaction level and lower than the average score of all the hotels. **The satisfaction gap among the rest customer types in this star category is not significant and the satisfaction level of each type, except for the solo group, is relatively high.**
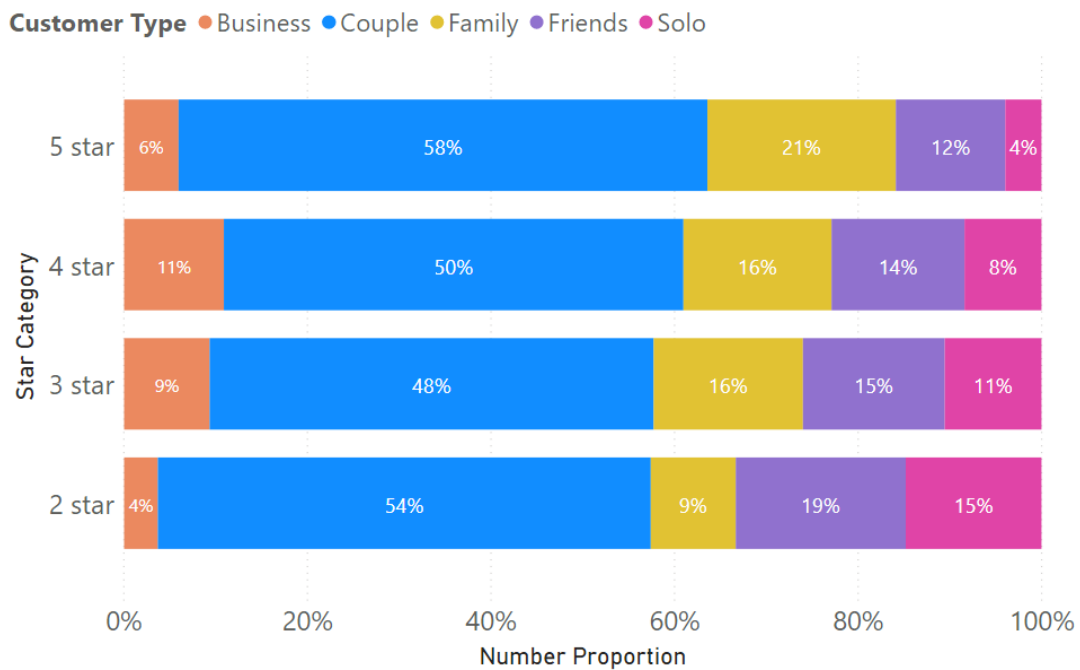
Figure 16 Distribution of customer types across star categories
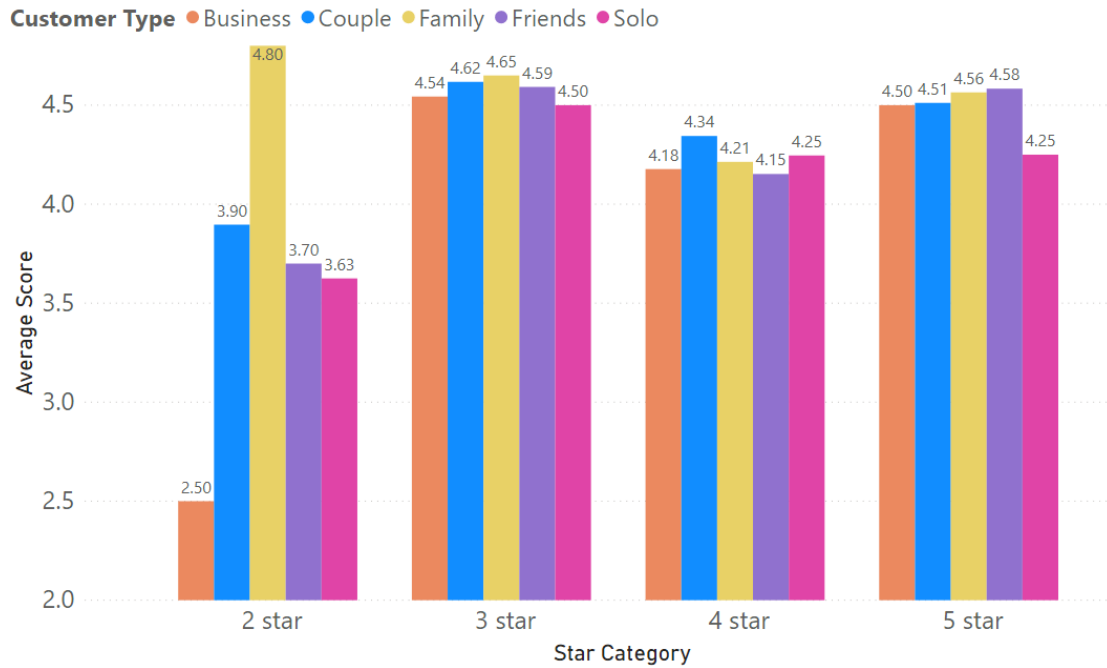
Figure 17 Satisfaction levels of various star categories across customer types

In summary, couple, family, and friends constitute the three most common customer types in the scores, while solo and business have a lower proportion. The 2-star hotel is the only exception, where couple, friends, and solo travelers are the top three types, and family and business have the lowest proportions.

Family group generally exhibits a relatively high satisfaction level across various hotels, except in 4-star hotels, where their satisfaction is lower than the average and slightly lower than that of couples and solo travelers.

Couple and friends groups both have high satisfaction levels in 3-star and 5-star hotels, followed by 4-star and 2-star hotels. Notably, there is a significant gap in satisfaction, particularly in 2-star hotels when compared to other hotel categories. Friends group tends to have lower satisfaction level compared with couple, except for 5-star hotels.

Business travelers share a similar satisfaction level across different star ratings with the couple and friends groups. However, its scores are generally lower than those of the

friends group, except for a slightly higher level in 4-star hotels.

Solo group shows the highest satisfaction level in 3-star hotels, followed by 4-star and 5-star hotels, where the satisfaction levels are identical. Conversely, 2-star hotels exhibit the least satisfactory score.
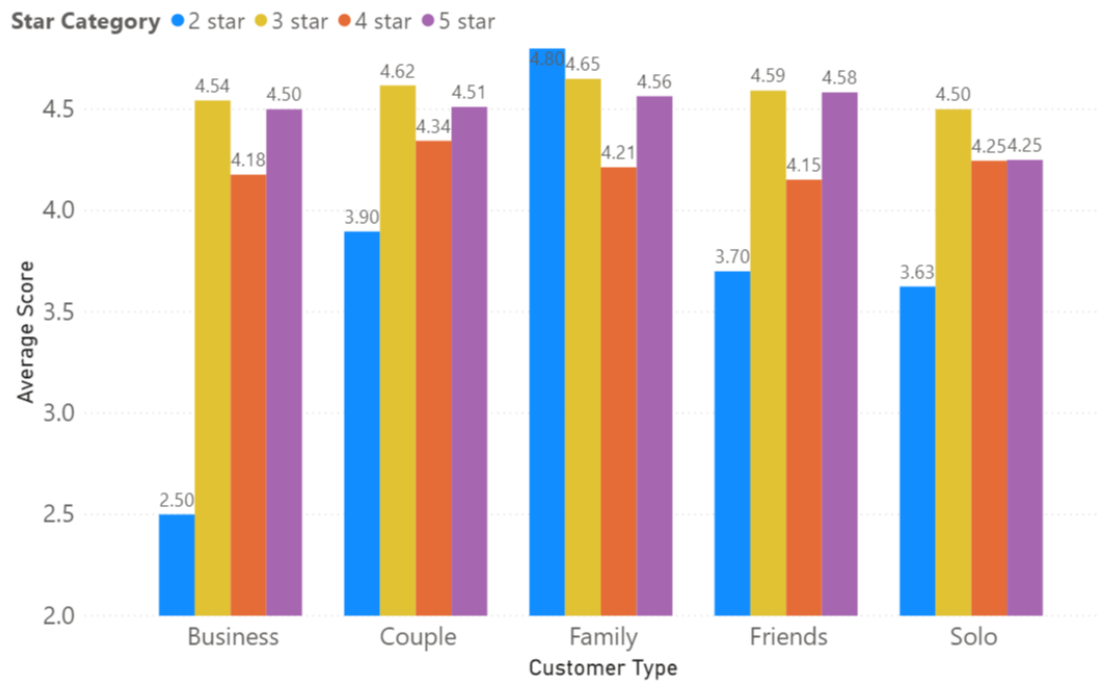
Figure 18 Satisfaction levels of various customer types across star categories

6.2 Text Mining Results

This section reveals the topics hidden in customer reviews in star-rated hotels by using the LDA method. At the same time, other text mining techniques like FP-tree and Clustering are used to uncover more details to help understand the important aspects and problems in customer reviews.

6.2.1 Topic Analysis

The output files in RapidMiner Studio are analyzed to derive insights. The group of words are categorized into appropriate topics by applying topic modeling to the corpus of customer reviews from each star category. These topics are presented as a series of words, requiring human intervention to define them. In addition to the groups of words, representative reviews are also consulted at the same time to label the topics properly. The representative reviews are identified as the top 10 core reviews based on the highest proportion percentage of the modeled topic.

To define the topics, commonly used hotel attributes are referred to in previous literature to help label them. Zhang and Verma [24] used 4 main topics: service, room, location and building, amenities, and facilities, which were rated most commonly by customers in hotel post-stay surveys and OTA sites. Han et al. [29] chose 5 topics: amenities, location, transactions, value, and experience for their review analysis. Drawing on the summarization and explanation provided in previous literature, this thesis aims to refine its topics more effectively and clearly.

Table 1 displays the labeled topics extracted from reviews across star-rated hotels. They are sorted in descending order (A through E) based on the frequency with which they appear as the top topic in reviews. Besides, the top 5 keywords with the highest loading are also listed under each topic. Next, the topics for each star rating are elaborated in detail.

Star rating	2 Star (N=62)	3 Star (N=604)	4 Star (N=957)	5 Star (N=372)
Topic A	**Location**	**Experience**	**Experience**	**Experience**(service)
words	great	good	good	great
	location	room	nice	hotel
	helpful	shower	staff	Valencia
	breakfast	location	clean	friendly
	stay	coffee	breakfast	location
Topic B	**Room**	**Transaction**	**Location** (nearby transportation)	**Facilities**
words	room	hotel	hotel	hotel
	stay	stay	city	stay
	city	bed	helpful	city
	small	excellent	breakfast	pool
	walk	restaurant	great	area
Topic C	**Experience**	**Location**	**Staff**	**Staff**
words	Valencia	great	staff	good
	clean	city	Valencia	staff
	night	breakfast	great	service
	bathroom	location	location	breakfast
	street	room	excellent	helpful
Topic D	**Amenities**	**Amenities** (especially breakfast)	**Amenities**	**Transaction**
words	good	hotel	room	room
	friendly	nice	bed	day
	room	Valencia	bathroom	bed
	bed	breakfast	shower	night
	time	walk	didn(didn't)	stay

Analysis and segmentation of the hotel offers in Valencia using TripAdvisor's key terms

Topic E	Staff	Staff	Transaction (issue communication, refund issue, before, check-in)	Hotel overall assessment
words	hotel	staff	hotel	hotel
	staff	clean	service	nice
	nice	friendly	day	lovely
	perfect	helpful	time	excellent
	day	lovely	stay	clean

Table 1 Topics and high loading terms across star-rated hotels

2 Star

5 topics are modeled for the 2-star-rated hotel, where number of reviews is 62 in total. Topic A is about **location** including positive descriptive terms like "great" and terms related to specific features such as "location," and "stay." Topic A is identified as the top topic in 17 reviews, emphasizing the importance of location in 2-star hotels. **Room** stands out as a key factor influencing customers' preferences, as indicated by 13 reviews. The reviews of Topic B encompass room overall evaluation like room size. Keywords include "room," "small," and "stay." Furthermore, topic C centers on **experience** during the stay, as well as their feelings and actions afterward, which is considered as the top topic by 12 reviews. The typical keywords include "clean," "night," and "bathroom." The following is Topic D, which concentrates on **amenities** and is cited as the main topic in 10 reviews, equaling the number of reviews where Topic E **staff** is considered the top topic. Keywords under topic amenities include "room," "bed," and "time." Meanwhile, the core words associate with the topic staff include "staff," "nice," and

"perfect."

3 Star

5 topics are modeled for the 3-star-rated hotel, where number of reviews is 604. Topic A is about the overall **experience** including positive descriptive terms like "good," and terms related to specific features such as "room", "shower," "location," and "coffee." In addition, value mainly about price is mentioned too under this topic. Topic A is identified as the top topic in 149 reviews, emphasizing the importance of guest experience in 3-star hotels. **Transaction** also significantly influences customers' preferences, as regarded as the main topic by 130 reviews. The reviews of Topic B encompass preparations before arrival, interactions during the stay, and check-out. Keywords include positive adjectives like "excellent," and mechanics of a guest's stay like "stay," "bed," and "restaurant." Equally significant to Topic B is Topic C, which focuses on **location** and is considered the top topic by 128 reviews. The typical keywords include "city," "location," and "breakfast." The following is Topic D, which focuses on **amenities** and is cited as the main topic in 100 reviews. In this case, breakfast is frequently discussed in the amenities. The core words associate with the topic include "nice," "hotel," and "breakfast." Lastly, the topic E is **staff**, which is regarded as the primary focus by 97 customers. High-loading terms encompass positive adjectives such as "clean," "friendly," "helpful," and "lovely."

4 Star

5 topics are modeled for the 4-star-rated hotel, where number of reviews is 957. Topic A is about **experience** including positive descriptive terms like "good," "nice," and terms related to specific features such as "staff" and "breakfast." Topic A is identified

as the top topic in 248 reviews, highlighting the significance of the guest experience in 4-star-rated hotels. **Location**, especially the nearby transportation, also plays an important role in customers' focus, as regarded as the main topic by 200 reviews. The reviews of Topic B refer to "hotel," "city," and "breakfast." Equally significant to Topic B is Topic C, which focuses on **staff** and is considered the top topic by 199 reviews. The typical keywords include "staff," and some positive adjectives related to service like "great" and "excellent." The following is Topic D, which focuses on **amenities** and is cited as the primary subject in 180 reviews. The core words associate with the amenities provided in the hotel room including "room," "bed," "bathroom," and "shower." The last topic E is **transaction**, which is considered as the primary focus by 130 customers. In this case, staff communication, refund issues, preparations before arrival, and check-in are the main subjects under transaction. Typical words include terms connected to the mechanics of a guest's stay such as "hotel," "service," "day," "time," and "stay."

5 Star

5 topics are modeled for the 5-star-rated hotel, where number of reviews is 372. Topic A is about **experience**. In this context, experience is mainly about service. Additionally, location is also mentioned under this topic. Positive descriptive terms related to service like "great," "friendly," and terms related to location such as "Valencia" and "location" are listed. Topic A is identified as the top topic in 116 reviews, showing the significance of the guest experience in 5-star category. **Facilities** are noteworthy in customers' focus in luxury hotels, as regarded as the main topic by 77 reviews. The reviews of Topic B mention places that fulfill guest's needs like "pool" and "area." Topic C emphasizes **staff service**, which is considered the top topic by 63 reviews. The typical keywords include "staff," "service," "breakfast," and some adjectives related to staff like "good"

and "helpful." The following is Topic D, which focuses on **transaction** and is cited as the primary subject in 62 reviews. It is significant to point out that in this category, the transaction aspect involves the following areas like issue communication, booking issues, check-in, stay, and check-out. Value is identified in terms of membership, star category, guests' perceived value, and money too. The core words associate with the transaction including "room," "bed," "day," "night," and "stay." Lastly, **hotel overall evaluation** is the primary focus for 54 customers. Typical words include "hotel," "nice," "lovely," "excellent," and "clean."

Summary of Topic Analysis

After the analysis of each star category's topics, **table 2** shows the summary of all the generated topics in all star-rated hotels and the frequency of each topic. Experience and staff are the 2 most commonly discussed topics in the hotels, followed by location, amenities and transaction. **These are the attributes that receive the highest attention from customers and require careful consideration by hotel management.** Room, facilities, and hotel assessment are each mentioned once in the hotel reviews.

Num	Topics	Number of occurrences	2 Star	3 Star	4 Star	5 Star
1	Location	3	√	√	√	
2	Room	1	√			
3	Experience	4	√	√	√	√
4	Amenities	3	√	√	√	
5	Staff	4	√	√	√	√
6	Transaction	3		√	√	√

Analysis and segmentation of the hotel offers in Valencia using TripAdvisor's key terms

| 7 | Facilities | 1 | | | | √ |
| 8 | Hotel overall assessment | 1 | | | | √ |

Table 2 Topic occurrences in star-rated hotels

Topic Importance Difference across Star-rated Hotels

Although the mentioned attributes are important in the hotel industry, each attribute has a different importance ranking across star categories. **Table 3** reveals the rankings of the aspects in each star-rated hotel.

For the attribute **experience**, there is no doubt that it's the most important in both middle or high-tier hotels. Experience also holds significant weight in 2-star hotels. This attribute encompasses a customer's overall impression, judgment of the elements during the stay, sentiment, and action afterward. For instance, some reviews express the intention to return or recommend the hotel to other people.

Following experience, **location** is the second notable topic, which is quite reasonable. Customers from 2-star and 4-star hotels refer location frequently compared with other stars. However, it's less evident in 5-star hotels. A good location often entails easy access to transportation, tourism spots, supermarkets, and restaurants. The pleasant surrounding environment and scenic views from the hotel are also discussed in customer's comments.

Moreover, **Amenities** also play a significant role in 2, 3, and 4-star hotels, whereas they are less emphasized in 5-star category. Under this topic, bed, shower, towel, breakfast,

etc. are frequently discussed and have a great influence on evaluations of the room and guest experiences.

Additionally, **staff** performance is another influential aspect, especially in higher-tier hotels. Staff serves as a crucial link between customers and hotels, and their interactions have a significant impact on guests' overall experience. Some reviews comment that staff's behaviors make the guest feel like home or special. In higher-tier hotels, there is greater emphasis on staff communication. Besides, exclusive services like membership benefits, complimentary room upgrades, welcome gifts, and birthday surprises are mentioned more frequently than lower and middle-tier hotels, indicating higher service standards and guest expectations based on the hotel's high value.

Next, **transaction** aspect includes various stages such as the booking stage, email communications, check-in, stay experience, and check-out stage. It holds greater importance in middle and high-tier establishments, particularly in 3-star hotels. However, it's not as prominently mentioned in 2-star hotels. Under this subject, website information, room description, advance payment, and refund policies are noteworthy and should be transparent and reliable.

Lastly, **facilities** receive special attention in 5-star hotels, especially its pools. Besides, **hotel overall assessment** is mentioned. In 2-star hotels, **room** evaluation is noteworthy, particularly room size and cleanliness.

Topics	Ranking			
	2 Star	**3 Star**	**4 Star**	**5 Star**
Location	1	3	2	N/A
Room	2	N/A	N/A	N/A

Analysis and segmentation of the hotel offers in Valencia using TripAdvisor's key terms

Experience	3	1	1	1
Amenities	4	4	4	N/A
Staff	5	5	3	3
Transaction	N/A	2	5	4
Facilities	N/A	N/A	N/A	2
Hotel overall assessment	N/A	N/A	N/A	5

Table 3 Topic importance rankings in star-rated hotels

6.2.2 Clustering Analysis

The reviews are clustered using the k-means method by RapidMiner Studio. This section presents the clustering results for hotel reviews in each star category. It provides more details inside similar reviews and sheds light on important aspects of customer focus and noteworthy issues in the operations.

2 Star

The reviews in 2-star hotels are segmented into 5 clusters. **Figure 19** shows the details in each cluster. **Figure 20** heat map shows the important terms. Deeper colors mean higher values, indicating more importance. The term "Noisy night window" stands out with the highest value, highlighting the importance of **noise** in customer's focus on room conditions. Besides, **location** attribute such as "Valencia walk," "attraction," and "tourist" is also important. Details regarding **amenities** are captured by terms like "comfortable bed," "bathroom old," and "air conditioning."

Number of Clusters: 5

Cluster 0 16

cost is on average **287.50%** larger, **air_con** is on average **287.50%** larger, **bathroom_old** is on average **287.50%** larger

Cluster 1 15

hotel_friend is on average **313.33%** larger, **comfortable_b** is on average **313.33%** larger, **feb** is on average **313.33%** larger

Cluster 2 21

ok is on average **195.24%** larger, **sound** is on average **195.24%** larger, **i_go** is on average **195.24%** larger

Cluster 3 4

noisy_night is on average **1,450.00%** larger, **noisy_night_window** is on average **1,450.00%** larger, **night_window** is on average **1,450.00%** la

Cluster 4 6

valencia_walk is on average **933.33%** larger, **tourist** is on average **750.61%** larger, **attract** is on average **698.35%** larger

Figure 19 Cluster details in 2-star hotels

Figure 20   Cluster heat map in 2-star hotel

3 Star

The reviews in 3-star hotels are segmented into 10 clusters. **Figure 21** shows the details in each cluster. **Figure 22** heat map reveals the key terms with darker colors. Attributes such as location, staff, value, breakfast, experience are highlighted in the clusters.

Location is illustrated by terms like "terrace nice view," "beach city art," "arts y ciencia," "access to city," and "old town." **Amenities** are also mentioned like terms

"bigger room," and "bed perfect." **Breakfast** is the most impressive subject indicated by many terms such as "breakfast tasty," "breakfast selection," "breakfast fantastic," and "book breakfast."

Besides, **staff** performance is praised in comments like "great helpful staff," "gave map," "desk staff friendly," "amazing hospitality," and "cheerful." Guests also share positive **experiences**, describing the hotel as "good nice," and " cool hotel," with some expressing a desire to extend their stay. A comment wrote, "Valencia is a wonderful city - shall come back to the city and to the hotel/bar." Finally, **value** is expressed in the comments like term "Excellent price/value ratio."

Number of Clusters: 10

Cluster 0 `83`

breakfast_fantast is on average 627.71% larger, **s_old_town** is on average 627.71% larger, **valencia_s_old** is on average 627.71% larger

Cluster 1 `50`

room_beauti is on average 1,108.00% larger, **stay_want** is on average 1,108.00% larger, **breakfast_tasti** is on average 1,108.00% larger

Cluster 2 `102`

enjoy_valencia is on average 492.16% larger, **amazing_hospit** is on average 492.16% larger, **city_com** is on average 492.16% larger

Cluster 3 `49`

desk_staff_friend is on average 1,132.65% larger, **helpful_accommod** is on average 1,132.65% larger, **partner** is on average 1,132.65% larger

Cluster 4 `54`

artes_y_ciencia is on average 1,018.52% larger, **bed_perfect** is on average 1,018.52% larger, **breakfast_select** is on average 1,018.52% larger

Cluster 5 `59`

bigger_room is on average 923.73% larger, **book_breakfast** is on average 923.73% larger, **i_stayed_long** is on average 923.73% larger

Cluster 6 `45`

beach_city_art is on average 1,242.22% larger, **cheer** is on average 1,242.22% larger, **city_sci** is on average 1,242.22% larger

Cluster 7 `45`

access_c is on average 1,242.22% larger, **check_fast** is on average 1,242.22% larger, **gave_map** is on average 1,242.22% larger

Cluster 8 `41`

good_nic is on average 1,373.17% larger, **terrace_nice_view** is on average 1,373.17% larger, **cool_hotel** is on average 1,373.17% larger

Cluster 9 `76`

excellent_pric is on average 694.74% larger, **great_helpful_staff** is on average 694.74% larger, **ivo** is on average 694.74% larger
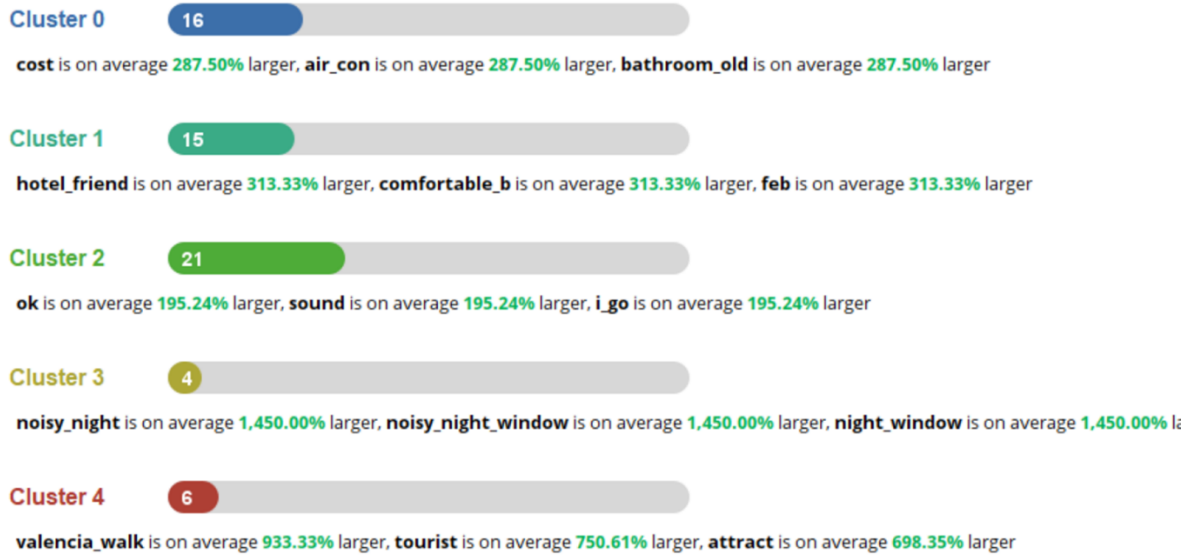
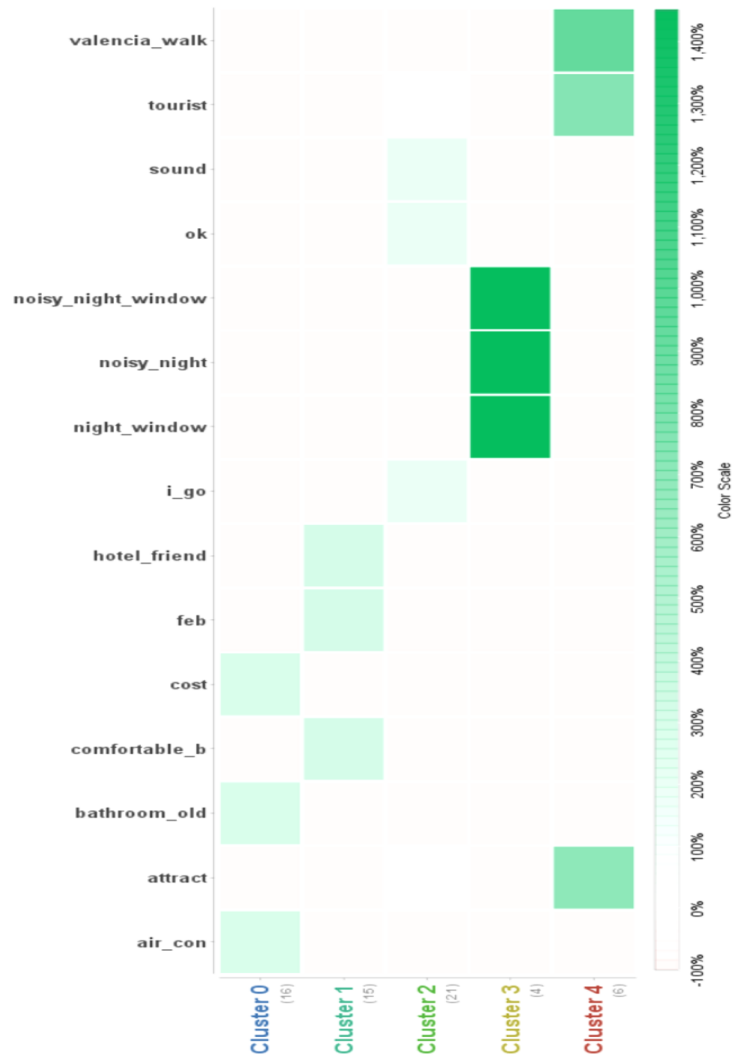**Figure 21 Cluster details in 3-star hotels**
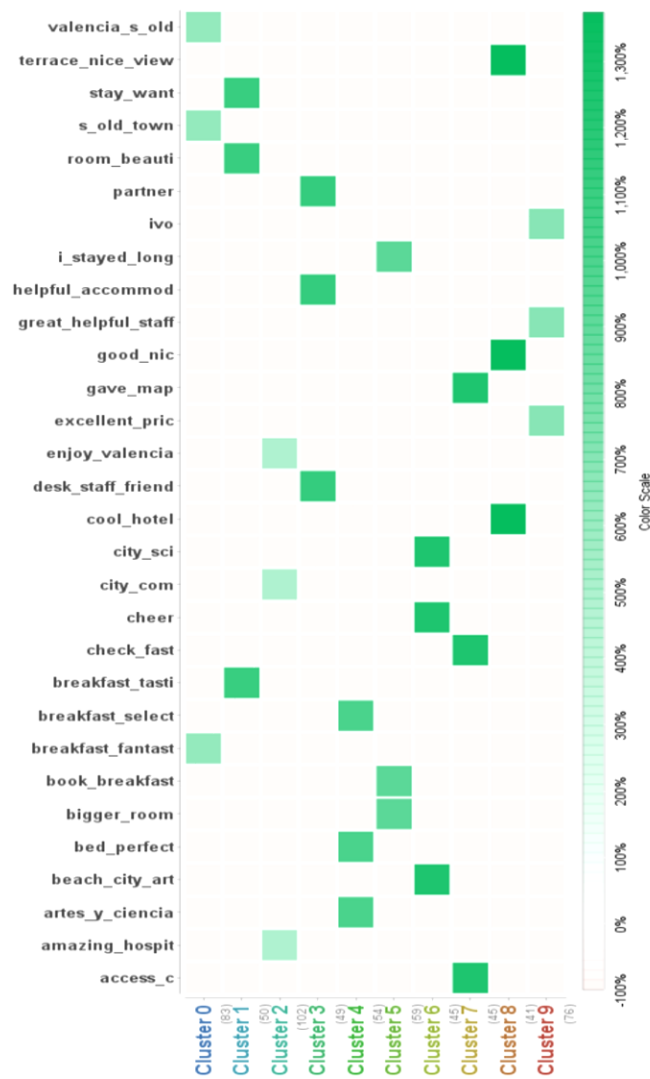
Figure 22 Cluster heat map in 3-star hotels

4-star

The reviews in 4-star hotels are divided into 15 clusters. **Figure 23** shows the details in each cluster. **Figure 24** heat map illustrates key terms with deeper colors indicating higher significance. Pool, staff attitude and service, transaction, value, location are notable topics in the reviews.

Pool is greatly highlighted in 4-star hotel reviews. Some customers expressed

disappointment after booking the hotel specifically for its pool, only to find it closed during their stay. Except for the facility, **staff attitude** is also emphasized, with discussions on poor customer service being similarly prevalent. Negative feelings are expressed in the comments. For instance, one guest wrote, "Some of the staff are very rude and spoilt our stay. I wouldn't stay here again due to the staff attitude." Another one remarked, "This property is the perfect example of a staff attitude letting the property down which is otherwise was a good one." However, positive experiences with exceptional staff members can significantly improve overall ratings, as evidenced by comments like, "This would have been a 3-star review if it weren't for the absolutely wonderful hostess in the breakfast/bar." These reviews reveal the great importance of customer service and staff attitude in shaping guests' perceptions of a hotel. **Hoteliers should pay more attention to the staff training.**

Additionally, **transaction** is noteworthy. Guests report problems such as booking issues, order cancellations, and unclear refund policies, especially when booking through agent websites. Moreover, credit card self-deduction and room upgrade fee communication are discussed. Payment is frequently mentioned in the reviews mainly about invoice and billing amounts which should be taken into great account since it's directly related to customer's money and interest. Cancellation and refund policies should be reasonable and clearly communicated to customers when booking in order to avoid future complications. When problems do arise, it is crucial for hotel staff to demonstrate a willingness to address and resolve such issues promptly and professionally, rather than resorting to personal attacks or neglect.

Besides, comments about **value** highlight positive experiences, such as "great value price" about the massage and spa pool session in the case. Many brand chain names are frequently mentioned in this star category. For example, one review noted, "I chose the hotel because of the Marriott membership, of which I am a Titan member - I got

absolutely everything I needed, and my loyalty was recognized." This aligns with the high brand value and allure associated with upscale hotels.

Last but not least, **location** remains a popular topic indicated by terms like "central area walk." While other aspects like experience, and breakfast receive less attention compared to the previously mentioned factors.

Analysis and segmentation of the hotel offers in Valencia using TripAdvisor's key terms

Cluster 10 28
arrival_receptionist is on average 3,317.86% larger, cancel_t is on average 3,317.86% larger, customer_service_poor is on average 3,317.86% larger

Cluster 11 80
absolutely_delight is on average 1,096.25% larger, absolutely_wond is on average 1,096.25% larger, asked_mainten is on average 1,096.25% larger

Cluster 12 67
nice_place_stay is on average 1,328.36% larger, hotel_feel is on average 1,328.36% larger, made_breakfast is on average 1,328.36% larger

Cluster 13 10
staff_attitud is on average 9,470.00% larger, catch_flight is on average 9,470.00% larger, payment is on average 9,470.00% larger

Cluster 14 75
hotel_old_town is on average 1,176.00% larger, amount_street is on average 1,176.00% larger, amount_street_nois is on average 1,176.00% larger

Figure 23 Cluster details in 4-star hotels

Figure 24 Cluster heat map in 4-star hotels

5-star

The reviews in 5-star hotels are divided into 9 clusters. **Figure 25** shows the details in each cluster. **Figure 26** heat map shows key terms with deeper colors indicating higher significance. Location, service, experience are notable aspects in the reviews.

Topic **location** is mentioned frequently using terms "walking distance hotel," "easy to get", And "transportation link." **Service experience** is highlighted with terms like "service experience," "evening restaurant," "manager jose," and "cava chocolate." In 5-star hotels, evaluations of services are more prevalent compared to lower-starred hotels. Front desk, housekeeping, gym, breakfast, and dinner services are commonly referenced. This is partly due to the broader range of services and facilities offered in 5-star establishments. Personal services like birthday surprises, welcome gifts, room upgrades are discussed more, consistent with the higher standards expected in luxury accommodations.

Additionally, **value** is noteworthy in the comments. Terms like "5 star," "standard," "level," "member" are frequently used. For example, a customer wrote, "this is NOT the level of service we expected and had paid for in this level of Hotel." Another commented, "Check in experience was less you would normally expect for a 5- star hotel - less than friendly and no information of any kind was offered regarding the hotels facilities." These reviews reveal the perceived value and high expectations from customers. Some guests mentioned choosing the hotel due to their membership, demonstrating loyalty: "I chose the hotel because of the Marriot membership and got absolutely everything I needed there." It emphasizes the importance for high-tier hotels to focus on service experience to meet customer high expectations and maintain loyalty.
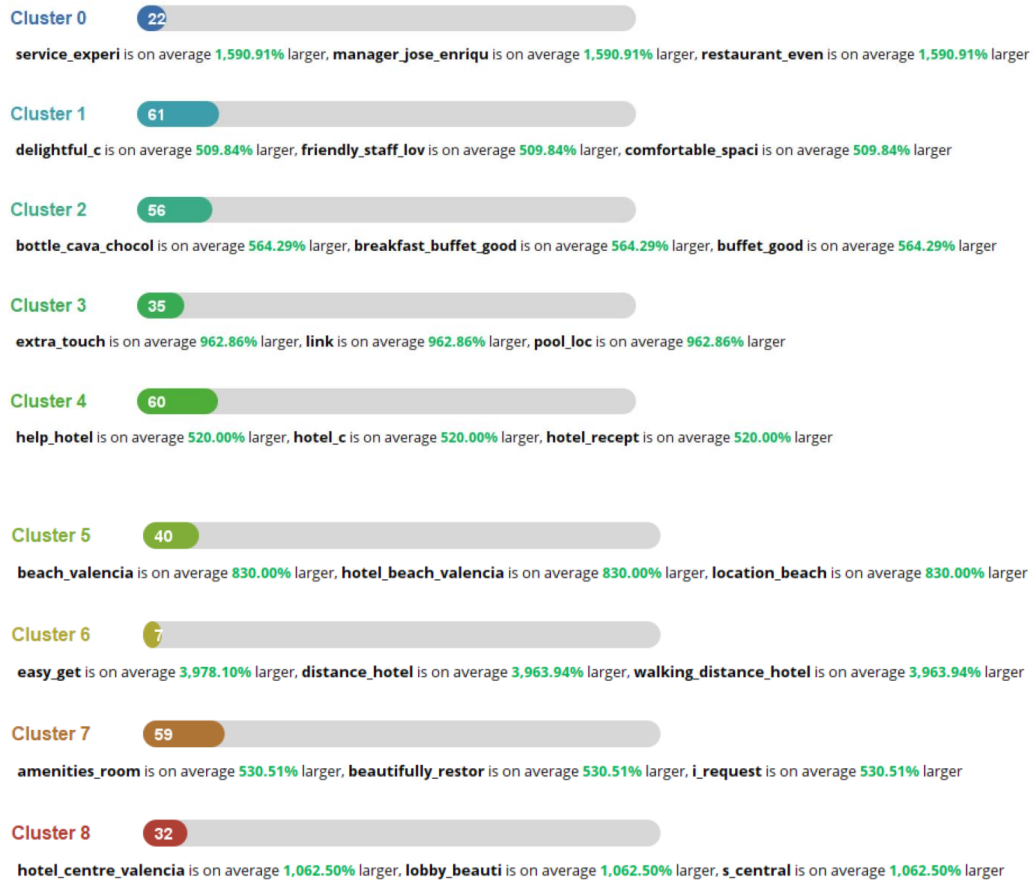
Number of Clusters: 9

Cluster 0 `22`

service_experi is on average **1,590.91%** larger, manager_jose_enriqu is on average **1,590.91%** larger, restaurant_even is on average **1,590.91%** larger

Cluster 1 `61`

delightful_c is on average **509.84%** larger, friendly_staff_lov is on average **509.84%** larger, comfortable_spaci is on average **509.84%** larger

Cluster 2 `56`

bottle_cava_chocol is on average **564.29%** larger, breakfast_buffet_good is on average **564.29%** larger, buffet_good is on average **564.29%** larger

Cluster 3 `35`

extra_touch is on average **962.86%** larger, link is on average **962.86%** larger, pool_loc is on average **962.86%** larger

Cluster 4 `60`

help_hotel is on average **520.00%** larger, hotel_c is on average **520.00%** larger, hotel_recept is on average **520.00%** larger

Cluster 5 `40`

beach_valencia is on average **830.00%** larger, hotel_beach_valencia is on average **830.00%** larger, location_beach is on average **830.00%** larger

Cluster 6 `7`

easy_get is on average **3,978.10%** larger, distance_hotel is on average **3,963.94%** larger, walking_distance_hotel is on average **3,963.94%** larger

Cluster 7 `59`

amenities_room is on average **530.51%** larger, beautifully_restor is on average **530.51%** larger, i_request is on average **530.51%** larger

Cluster 8 `32`

hotel_centre_valencia is on average **1,062.50%** larger, lobby_beauti is on average **1,062.50%** larger, s_central is on average **1,062.50%** larger

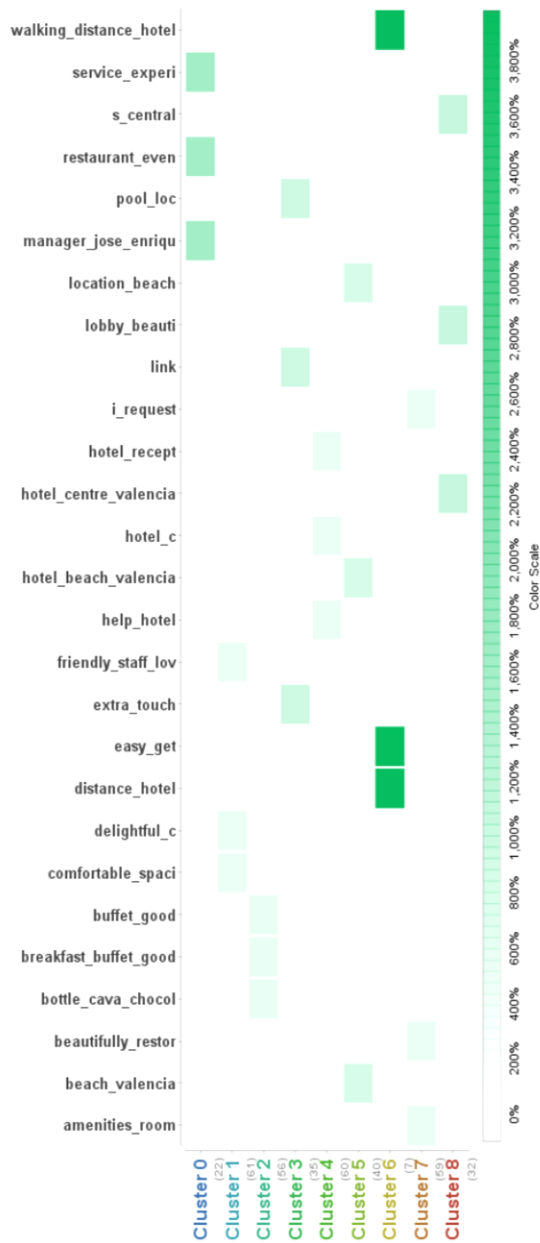**Figure 25 Cluster details in 5-star hotels**

Figure 26 Cluster heat map in 5-star hotels

6.2.3 FP-Growth Analysis

This section mainly introduces the frequent key terms found in the reviews across various star-rated hotels. FP-Growth algorithm illustrates co-occurring terms depicted by the FP-tree structure. Furthermore, the analysis provides the insights of similarities

and differences in the terminology used across different star-rated establishments.
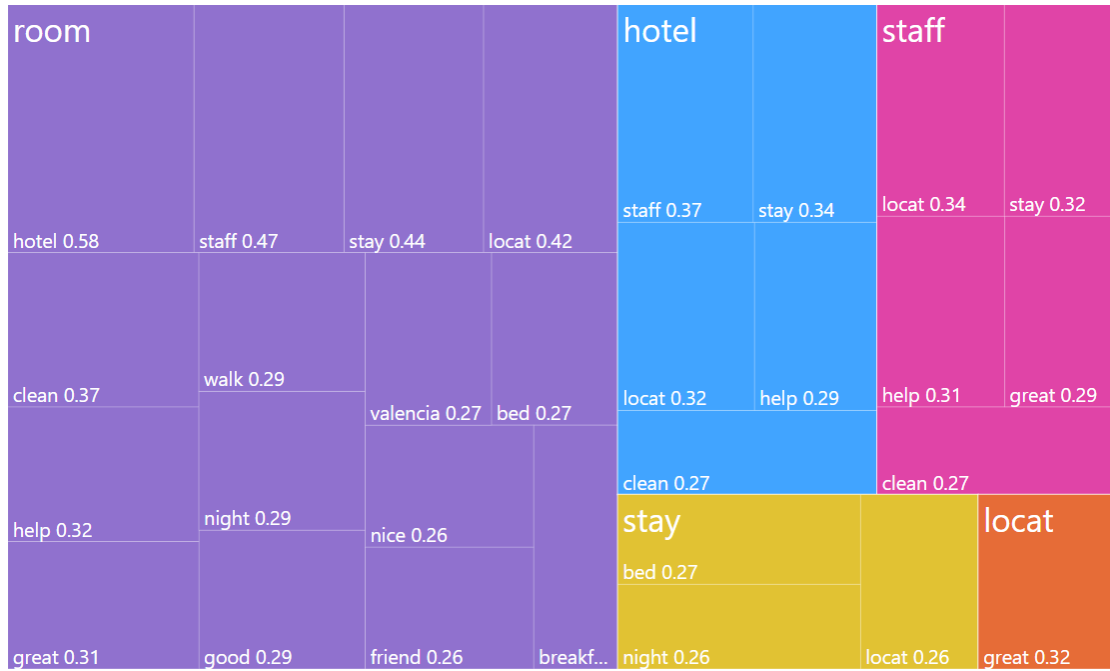
2 star

Figure 27 displays the top 30 most frequently utilized terms based on review frequency. The top 10 words include "room," "hotel," "staff," "stay," "location," "clean," "helpful," "great," "night" and "Valencia." Additionally, amenities such as "bed," "window," and "bathroom" are noteworthy.
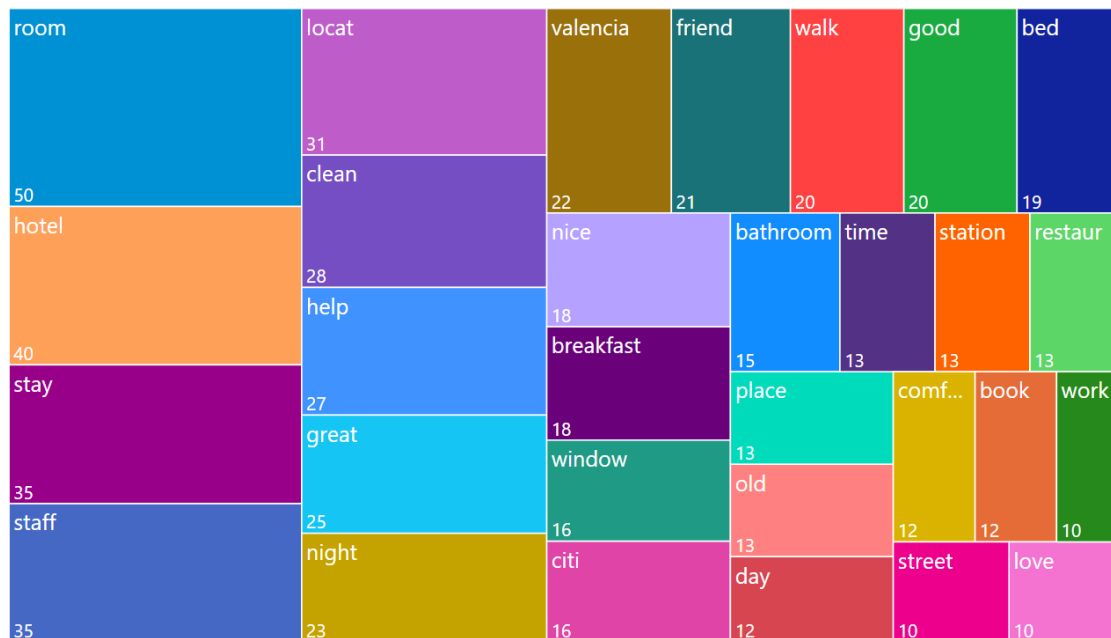
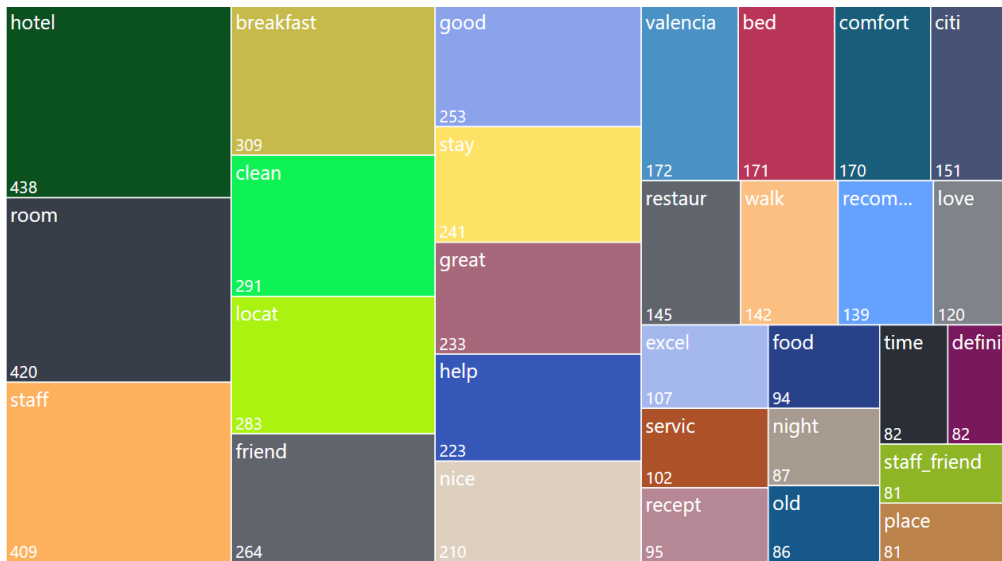**Figure 27 Top 30 most frequently utilized terms based on review frequency in 2-star hotels**

Figure 28 shows the frequently co-occurring terms arranged in a tree structure, with support values above 0.25 chosen for presentation. "Room," "hotel," "staff," "stay," and "location" are mainly featured as they are the most frequent. In addition, these 5 words often appear together and frequently co-occur with positive descriptors like "great."
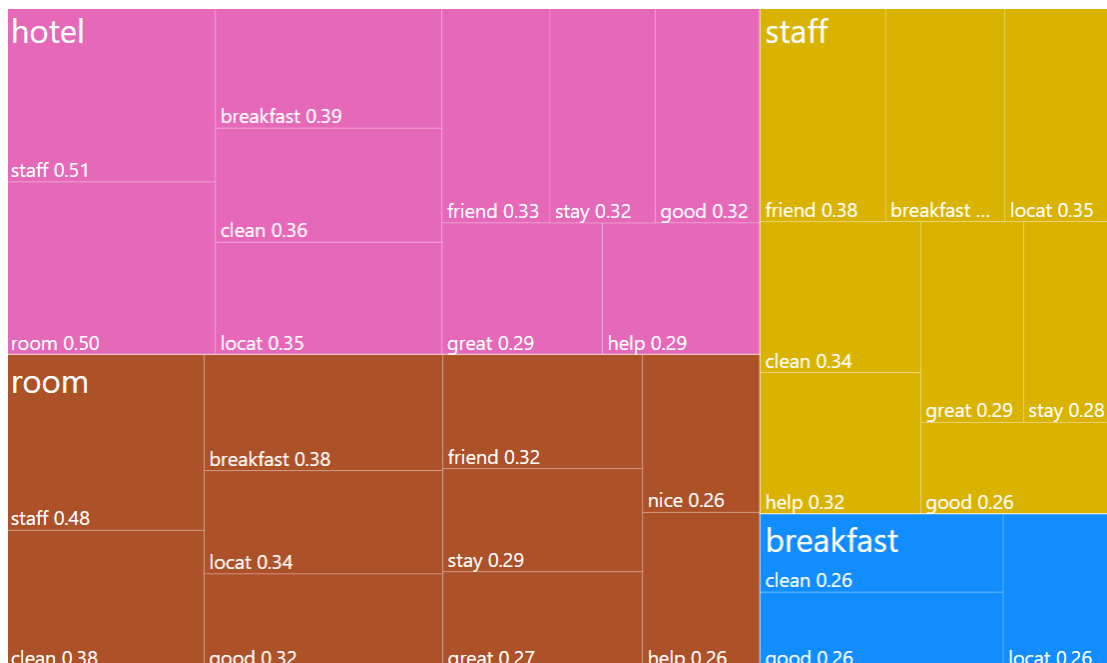
Figure 28 FP-tree in 2-star hotels

3 star

Figure 29 displays the top 30 most frequently utilized terms based on review frequency. The main aspects include "hotel," "room," "staff," "breakfast," "clean," "location," "stay," and some positive sentiment adjectives such as "friendly," "good," "great," and "helpful." In addition, Valencia city is mentioned frequently, along with service and reception.

Figure 29 Top 30 most frequently utilized terms based on review frequency in 3-star hotels
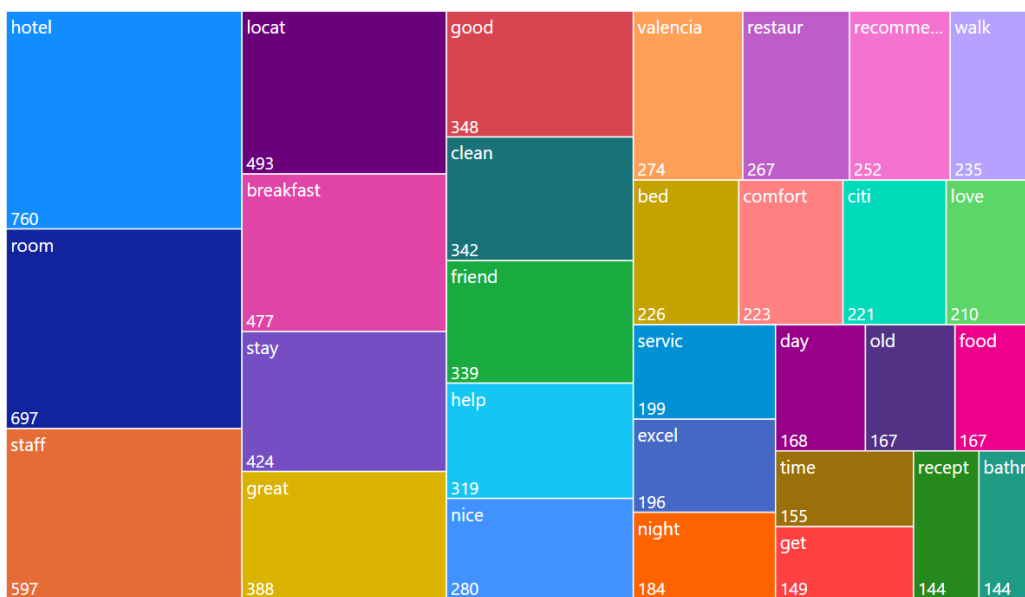
Figure 30 shows the frequently co-occurring terms arranged in a tree structure. "Hotel," "room," "staff," and "breakfast" are prominently featured as they are the most frequent terms. In addition, these 4 words often appear together, they all co-occur frequently with terms "location," "stay," and positive descriptors like "clean," and "good."

Figure 30 FP-tree in 3-star hotels

4 Star

Figure 31 displays the top 30 most frequently used terms in reviews, including "hotel," "room," "staff," "location," "breakfast," "stay," and positive adjectives such as "great," "good," "clean," and "friendly." The mention of Valencia city is also prominent. Additionally, "service" and "reception" are noted, along with important amenities like "bed," "food," and "bathroom."

Figure 31 Top 30 most frequently utilized terms based on review frequency in 4-star hotels

Figure 32 shows the frequently co-occurring terms arranged in a tree structure. "Hotel," "room," "staff," and "location" are mainly featured as they are the most important terms. In addition, these 4 words are interconnected and co-occur frequently with terms such as "breakfast," and positive terms like "great."

Figure 32 FP-tree in 4-star hotels
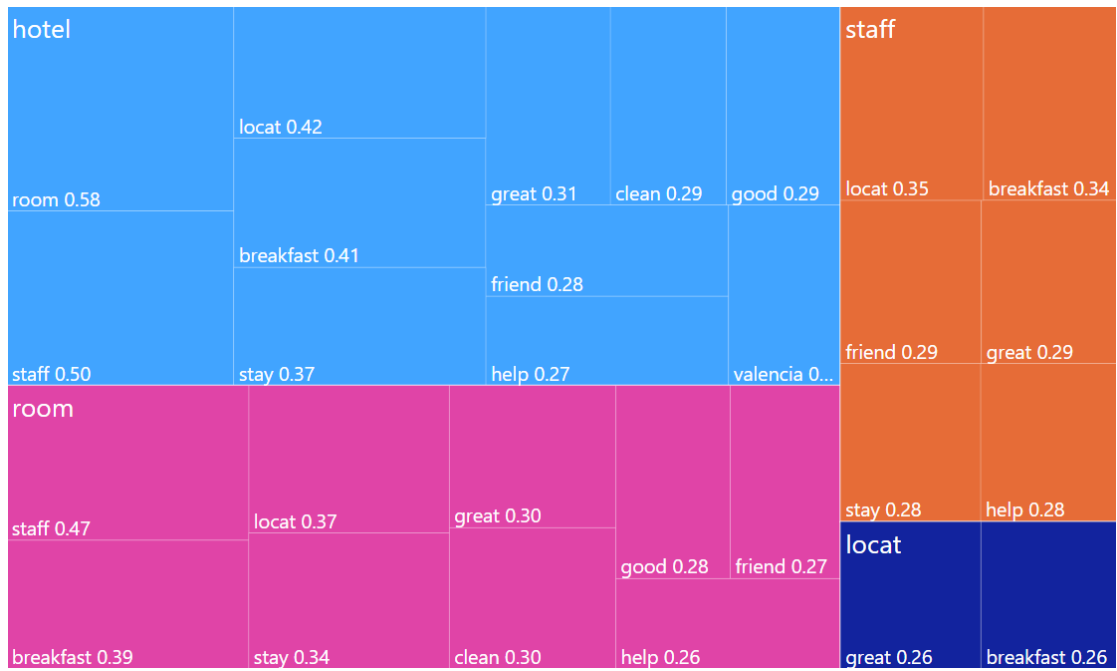
5 Star

Figure 33 displays the top 30 most commonly used terms in reviews, ranked by their frequency. The leading 10 terms include "hotel," "room," "staff," "stay," "breakfast," "location," "great," "Valencia, " "helpful," and "friendly." Service is also highly ranked. Notably, facilities such as "restaurant" and "pool" are discussed, along with references to the "beach."

Figure 33 Top 30 most frequently utilized terms based on review frequency in 5-star hotels

Figure 34 presents the co-occurring terms in a tree structure. "Hotel," "room," "staff," and "stay" are mainly featured as they are the most frequent terms. In addition, these 4 words frequently appear together and co-occur frequently with terms such as "breakfast." Besides, the first 3 words, especially, often combine with "location," "service," "Valencia," and positive descriptors such as "great," "helpful," and "friendly."
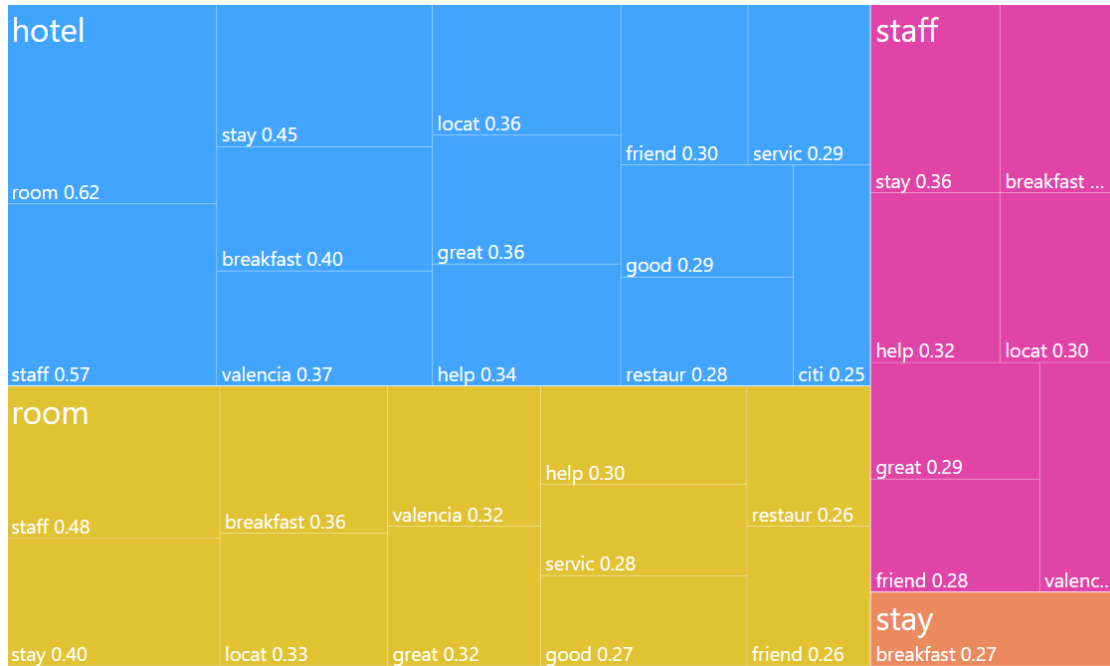
Figure 34 FP-tree in 5-star hotels

Summary of Term Frequency Analysis

As for the similarities, **figure 35** summarizes the top 10 frequent terms in the reviews of each star category. They all share the same terms **like "hotel," "room," "staff," "stay," "location," and "great." These aspects are crucial in customers' views and require attention from hoteliers.** In addition, "breakfast" and "friendly" are discussed in 3, 4, and 5-star hotels, while "clean" is highlighted in 2, 3, and 4-star categories. Given that all the hotels are located in Valencia, the hotel experience is closely connected to the city. Rest terms are positive terms describing the staff, amenities, and experiences, showing a good overall impression among customers in all the star categories.

Regarding the differences, there are subtle distinctions among the star categories. Amenities are more obvious in 2-star category as for terms such as "bed," "window," and "bathroom" with higher ranks. "Breakfast" is highlighted in 3-star hotels. For 3, 4, and 5-star categories, "service" stands out, while it holds a lower rank in the 2-star

category. These observations align with the results obtained from the topic analysis.
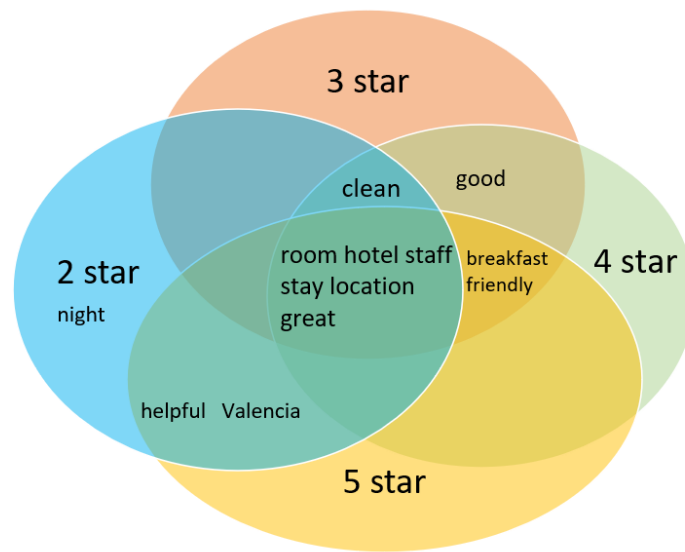
Figure 35 Top 10 frequent terms in each star category

7. Discussion

This section further explains the practical significance for hotels of the descriptive analysis results and the findings from text mining, then outlines the implications in the fields of theory and management.

7.1   Discussion of Analysis Results

7.1.1 Descriptive Analysis Result

The thesis employs TripAdvisor rating scores to conduct a descriptive analysis to evaluate customer satisfaction levels. Overall, the average score of 3-star hotels is the highest, followed by 5-star, 4-star and 2-star hotels. The standard deviations of 3-star and 5-star hotels are lower, indicating relatively consistent performances among the hotels in these categories. Conversely, hotels in the 2-star and 4-star categories exhibit greater variability, particularly 2-star hotels. From the results, the conclusion can be drawn that 3-star hotels generally achieve the highest satisfaction level, with 5-star hotels performing well. However, the satisfaction levels of the rest star categories are comparatively lower, indicating the presence of more issues, with particular emphasis on the 2-star category.

As for the satisfaction levels in different travel months, January has the highest customer satisfaction level, followed by February, April, July, May, and June. Conversely, the rest months have lower scores than the overall average score especially October, which is the least satisfactory month. In general, customer satisfaction peaks from December to February, followed by March to May and June to August, with the lowest scores observed from September to November. It can be concluded that

satisfaction levels are generally higher during the off-peak travel season, with lower levels observed during peak periods. This could be attributed to ample room availability and satisfactory service during the off-peak season. **However, additional measures are necessary during the summer and autumn periods to enhance the performance.** When examining performance by star category and month, 4-star hotels maintain the most consistent scores across the whole year, followed by 3-star and 5-star hotels. On the contrary, the 2-star category exhibits the most significant fluctuations throughout the year, indicating considerable variability in satisfaction level. Tourism officials should prioritize assisting the 2-star category, as it faces the most severe issues and shortcomings in Valencia

In terms of **customer type**, in general, couple constitutes a significant portion, representing 51% among customer types, followed by family, friends, business, and solo. As for the overall customer satisfaction levels, couple and family have the highest scores, followed by friends, business, and solo.

Specifically for each star category, in 2-star hotels, couple, friends, and solo travelers are the most common customer types. However, their scores are all below the average level. Similar situations are also observed in 4-star hotels, where couple, family, and friends are the most common customer types, but they all score below average level, especially the friends segment. **2-star and 4-star hotels need to acknowledge these shortcomings and focus on catering to their main target customers at present.**

For 3-star hotels, couple, family, and friends are the most common customer types. Their satisfaction levels are relatively high, indicating that 3-star hotels adequately meet their main customers' demands. Moreover, the scores of business and solo groups are relatively lower. In 5-star hotels, the main customer segments and satisfaction levels are similar. Especially, the satisfaction score of the solo traveler group is lower than the

average level. **In the future, 3-star and 5-star hotels can implement better improvement measures tailored to business and solo groups to further enhance their satisfaction, especially for solo group.**

7.1.2 Text Mining Result

The thesis employs TripAdvisor customer text reviews to perform text mining analysis to find hotel attributes, operational details, and potential issues that customers emphasize. LDA method is used to extract the main topics from the reviews. Additionally, Clustering is conducted to delve into the details and issues in similar reviews. Lastly, FP-Growth algorithm is used to illustrate patterns of term frequency in hotel reviews, providing further insights into similarities and preferences in terms used in reviews across star-rated hotels.

LDA Result

Based on LDA method, 8 topics including **location**, **room**, **experience**, **amenities**, **staff**, **transaction**, **facilities**, and **hotel overall assessment** are extracted in total. These topics can be categorized into strategic or operational choices based on the types of decisions they represent. **Location** is clearly related to strategic choices, involving an upfront strategic choice about where the hotel is situated. Besides, **experience, staff, transaction, and hotel overall assessment** represent operational choices. Experience and hotel overall assessment aim to engage with the guest's expectations and needs during the stay, while transactions can either create or reduce frictions in the mechanics of the stay. Moreover, staff plays a key role in the interactions with customers and their experience, influencing their impressions greatly. Finally, room, amenities, and facilities are related to both strategic and operational choices, as they relate to a hotel's market positioning and competitive strategy, but also depend on how well they are

maintained and shown each day. Given these connections, reviews can serve as a source of valuable information for hoteliers to evaluate the effects of the choices that they make. Moreover, this categorization of topics can provide managers with a structured framework for better understanding the nature of this assessment [29]. In conclusion, strategic decisions are crucial, as they determine the geographical significance and market positioning of a hotel, and they are difficult and costly to change once established. In contrast, operational choices offer greater room for improvement and require ongoing attention and enhancement from the hotel.

Table 3 shows the topic rankings in each star category. For 2-star category, customers focus **location** most, followed by **room**, **experience**, **amenities**, **staff**. In the initial strategic phase, 2-star hotels need to carefully consider the hotel location, along with appropriate amenities and room standards, which also require consistent maintenance. For established hotels with less favorable locations, improvement measures can include offering shuttle bus services and providing detailed maps of nearby transportation routes to assist guests in touring and exploring the area. Moreover, staff training is crucial for creating a friendly customer experience. In low-tier hotels, the emphasis is not on luxurious amenities but on maintaining cleanliness and essential infrastructure to meet basic accommodation needs. Expanding on this groundwork, enhancing staff service levels can elevate the hotel's satisfaction levels.

For 3-star category, customer **experience** is highly valued, followed by **transaction**, **location**, **amenities,** and **staff**. In addition to considering location and amenities in the initial stage, 3-star hotels should focus on improving breakfast services as it is frequently mentioned in amenities. In terms of transaction, booking through third-party websites like booking.com, and lastminute.com is commonly mentioned, so hotels must ensure the accuracy of room descriptions and the clarity of refund policies to avoid negative reviews from customers with unmet expectations. In case of issues, proactive

communication with customers and third-party booking websites is essential. Furthermore, there should be a strong emphasis on staff training to foster a positive experience.

For 4-star category, **experience** receives the highest focus from customers, followed by **location**, **staff**, **amenities,** and **transaction**. In the initial stage, location is very important for 4-star category hotels. In addition, there is frequent mention of hotel chain brands and loyalty memberships in the reviews, highlighting the importance of brand recognition and customer loyalty. Hotels should capitalize on their brand reputation, particularly in nurturing relationships with loyal members. Regarding the guest experience, attention should be paid to the cleanliness and odors within the hotel and rooms, as references to unpleasant smells, such as "stench of sewage," can significantly impact guests' perceptions. Therefore, hotels should prioritize the maintenance of fundamental amenities and facilities, particularly in rooms and bathrooms, including plumbing and ventilation systems. Indeed, staff performance is of paramount significance in high-tier hotels, as it plays a critical role in shaping brand reputation and ensuring customer satisfaction. This emphasis on staff training is integral to upholding the hotel's value proposition and maintaining its position in the market.

For 5-star category, **experience** is emphasized, followed by **facilities**, **staff**, **transaction,** and **hotel overall assessment**. Due to the premium pricing and elevated expectations, it's reasonable that discussions about value are more prevalent compared with previous categories. On the one hand, renowned hotel chain brands can leverage their existing brand reputation more effectively for promotion and maintenance of loyal customer base. On the other hand, luxury facilities are necessary for this category to provide extra value compared to other star categories. During the initial stages of hotel construction, there should be a strong emphasis on planning and equipping luxurious facilities, particularly the pool. In the operational stage, careful maintenance is required

to ensure their optimal performance.

Service experience takes precedence in 5-star hotels. Hotels should take into great account of each aspect of services including booking, website information service, front desk service, housekeeping, breakfast, lunch, dining service, and check-out service to make their customers feel like home, which is discussed in reviews. Staff performance is closely observed, with positive comments praising their performance, which helps create a welcoming atmosphere. For instance, one comment wrote, "Overall, Only You Hotel Valencia staff not only is professional and friendly but they make you feel like you're at home."

In conclusion, regardless of the star category, staff training remains essential for hotels. Upholding a people-centric approach is crucial for ensuring guest satisfaction and creating a welcoming environment. Additionally, hotels of different star-rated have distinct areas of focus and details of attention in the management.

Clustering Result

The clustering method sheds light on critical aspects of customer preferences and noteworthy issues in operations. For 2-star hotels, **location** remains a key emphasis. Ideally, hotels should be situated near popular attractions o linked to transportation hubs. Regarding operational decisions, specific components in **amenities** are revealed, such as windows that offer noise isolation, comfortable beds, bathroom conditions, and the functionality of air conditioning. Hoteliers can prioritize meeting these fundamental needs to enhance customer satisfaction.

In the case of 3-star hotels, more details about **location**, **amenities**, **staff,** and **value** are

highlighted. Popular attractions such as beach, arts and science city, and old town are frequently mentioned, indicating that hotels can benefit from being closer and more convenient to these locations strategically. The terrace view is discussed too, suggesting that hotels should consider this aspect in their strategic location and design to increase extra experience for guests. Additionally, the discussion on price-value ratio underscores the importance of hotels carefully considering their pricing strategies. In terms of amenities, factors such as room size and bed quality are taken into account.

As for operational choices, breakfast selections and quality are emphasized for enhancement, indicating the need for hotels to offer a diverse and satisfying breakfast experience. Moreover, staff can enhance the guest experience by providing maps in the check-in stage, bike rental services, and touring information during their stay if requested.

Regarding 4-star hotels, there is a higher prevalence of negative issues, which aligns with the lower customer satisfaction scores revealed in the descriptive analysis section. As for **location** selection, hotels can benefit from being close to popular spots like the beach and central area. Some comments mentioned the need to catch a flight. It's preferable to have convenient taxi services or access to transportation to the airport. Additionally, operational challenges are notable in areas such as **pool** maintenance, **staff** attitude, and **transactions** about payment. Luxury facilities such as spas, pools, and gyms should be well-maintained. If they can't be used as usual, information must be updated on the booking page. For high-tier hotels, the premium facilities and services such as massage, spas, and meditation courses can greatly enhance the guest experience, aligning with the consumer trend of a greater inclination towards investing in luxury and wellness experiences.

Moreover, there are numerous complaints regarding staff attitude, indicating an urgent

need for standardized customer service training with key performance indicators to evaluate service delivery. Additionally, frequent complaints were related to hotel online booking and refund procedures. First, the address, facilities, and room information on the hotel booking website must be clear, genuine, and accurate. False advertising and outdated information can mislead customers, resulting in more harm than benefit. It's necessary to maintain regularly and update in time. Clear and consistent terms regarding refund timelines and amounts must be presented on both hotel's own websites and third-party platforms during the booking phase to avoid subsequent disputes. Similarly, rules regarding room upgrade services for members should be uniformly defined in the chain group to maintain customer loyalty and relationships.

Clarity in billing service amounts during the guest's stay, especially for customers using different currencies, is essential. Procedures for managing sensitive customer data such as credit card information should be developed in collaboration with legal departments to ensure adherence to regulations and fairness, thus preventing unauthorized charges. In case of any issues, it is important for hotel staff to demonstrate a willingness to address and resolve them promptly and professionally, rather than resorting to personal attacks or neglect.

As for 5-star hotels, **location**, **service**, **experience** are emphasized in detail. Though location is not presented individually in topic analysis result in 5-star category, it remains a crucial consideration during strategic planning. Service, however, emerges as the paramount aspect. An illustrative example is the frequent mention of the customer experience manager's name, indicating customers' acknowledgment of his outstanding service. This is evidenced by thoughtful preparations for birthday surprises and compassionate handling of unexpected incidents. The insight for hotels of this tier is to further refine standards for various services such as front desk, room service, and dining, regularly monitor service levels, and address any shortcomings, as service levels are

inconsistent—some being positively reviewed while others are negatively evaluated.

Additionally, services should be sufficiently personalized to anticipate and fulfill customer needs. For instance, some studies mention hotels tracking real-time social media comments and stories to identify customer needs, such as providing yoga blankets for a blogger who enjoys indoor yoga or preparing champagne for a couple celebrating a successful proposal [42].

FP-Tree Analysis

This analysis offers insights into the frequent terms in the reviews. It concludes that similar terms are utilized across different star-rated hotels. They all share the same terms like "hotel," "room," "staff," "stay," "location," and "great." These terms are not only frequently mentioned individually but also often occur together, indicating the core and fundamental needs in customers' views and requiring attention from hoteliers.

While some terms rank higher in certain star categories. For instance, "service" stands out in 3, 4, and 5-star categories, while it ranks lower in 2-star category. These observations align with the results of the topic analysis and clustering analysis. The frequent use of positive adjectives describes various attributes of the hotels, demonstrating the overall high level of customer satisfaction in Valencia's star-rated hotels, consistent with the descriptive analysis of the overall satisfaction score.

Due to that the thesis focuses on the post-epidemic era, there are 5 reviews including terms about the pandemic. They mainly shared their experiences about delays in visiting plans and returning to the same hotel after the pandemic. 1 review commented on the pandemic control measures implemented in spas. The decrease in discussions about the

pandemic also indicates a gradual reduction in its impact on travel and the hotel industry. Besides, "clean" is emphasized in the top ten words in star-rated hotels, especially after the pandemic. Hotels should take into account of the cleanness management, learn from the pandemic control measures, and accumulate experiences for future needs.

7.2 Theoretical Implications

Customer reviews are increasingly being recognized for their widespread impact on consumer behavior, highlighting their growing importance in business decision-making and researches. Schuckert et al. [43] concluded 5 main directions of studies about online reviews in the tourism industry. They were opinion mining/sentiment analysis, online reviews and online buying, satisfaction and management, motivation, and the role of reviews. The study of the thesis belongs to opinion mining.

Text mining techniques are utilized in the thesis to provide a more objective approach to analyzing the content of reviews due to their mathematical characteristics. The LDA method, known for its efficacy in uncovering latent topics within text, proves instrumental in extracting valuable insights relevant to real-world business applications. This thesis applies this method to the star-rated hotel industry, expanding the scope of application of the theory, and making significant contributions to the practical understanding and management of hotel industry.

Besides, the thesis enhances the existing literature on customer satisfaction in Spanish star-rated hotels. This study addresses a gap in the analysis of hotel rating scores and reviews on social websites in Valencia, covering a comprehensive range of hotels spanning 2 to 5-star, representing various tiers from low to high-end establishments. Historically, research on customer satisfaction has primarily centered on the pre-

pandemic era. The thesis aims to expand the understanding of customer rating scores and preferences in the post-pandemic context, offering updated insights into this evolving landscape.

Moreover, the findings support Oliver's [20] ECT, which assess customer satisfaction by comparing customers' expectations with their perceptions of the service provided. Especially in the 5-star hotels, this contrast between expectations and actual experiences is more evident in the reviews. Besides, the results confirm the multi-attribute theory, emphasizing the different importance of attributes on customer satisfaction [44]. For instance, service experience is highlighted in high-tier hotels, which aligns with findings from prior studies [24,29].

Finally, the study contributes to the knowledge of hotel rating score patterns by examining variations based on star category, travel month, and customer profile. As for the overall customer satisfaction levels, couple and family have the highest scores, followed by friends, business and solo guests. Similarly, Banerjee and Chua [8] found that business travelers tend to be the most stringent in their rating patterns compared to other profiles. Studying the differences among star categories, travel months, and customer profiles helps the hotel industry adopt more targeted marketing strategies and effective measures.

7.3 Managerial Implications

eWOM significantly influences consumer decisions and reputations of hotels, emphasizing the necessity for hotel managers to closely monitor and manage UGC. This is crucial for improvement, enhancing competitiveness, maintaining good customer relationships, and building brand reputation. Besides, during the review

collection process, it was observed that many hotels either did not respond or provided limited responses to customer comments on TripAdvisor. Some even neglect negative comments entirely. These issues indicate that some hotels still don't recognize the value of customer reviews as a powerful resource to improve and lack a proactive attitude. On the contrary, high-scoring hotels tend to promptly address customer feedback, with some even responding to all reviews. Hotels can benefit from implementing real-time monitoring of social media comments to swiftly identify potential customer needs and issues. It is imperative for hotels to respond to feedback, particularly negative feedback, in order to effectively address customer concerns and maintain positive relationships.

Moreover, the thesis provides practical implications for hoteliers regarding resource allocation. During the off-peak travel season, satisfaction levels are generally higher, with lower levels observed during peak periods. Hoteliers should allocate additional resources, such as staffing, during the summer and autumn periods to enhance performance. Besides, the analysis delves deeply into satisfaction levels across 5 main customer profiles, providing tailored recommendations for each star-rated hotel based on their primary customer segments. Finally, the thesis highlights key attributes in customers' minds and specific measures that hotels of different star categories should implement, both strategically and operationally. Moreover, detailed issues and optimization measures are provided in this thesis, especially for 4-star hotels. More details are illustrated in the discussion of text mining results section.

For government officials and tourism agencies, this study presents the customer satisfaction levels of star-rated hotels in Valencia based on TripAdvisor scores. The research reveals that the satisfaction levels of 2-star and 4-star hotels are relatively low, especially for 2-star hotels, while 3-star and 5-star hotels exhibit higher satisfaction levels. Therefore, government officials can take measures to directly understand the operational issues faced by 2-star hotels and provide assistance and training to better

support them.

In conclusion, the research provides an overview of customer satisfaction levels across star-rated hotels in Valencia, offers insights into main hotel attributes and proposes numerous practical measures from a managerial perspective to help hoteliers address shortcomings and better satisfy customer needs across various tiers of hotels.

8. Conclusion

8.1 General Findings

To understand the performance of star-rated hotels in Valencia from the customers' views, the thesis employs TripAdvisor rating scores to conduct a descriptive analysis of customer satisfaction levels. In addition, TripAdvisor reviews are analyzed to identify customer priorities regarding hotel attributes by utilizing text mining techniques. The following are the main conclusions.

Regarding the first research question, the average satisfaction score varies across star categories, with 3-star hotels receiving the highest average score, followed by 5-star, 4-star and 2-star hotels. In addition, the standard deviations of 3-star and 5-star hotels are lower, indicating relatively consistent performances within these categories. Conversely, hotels in the 2-star and 4-star categories exhibit greater variability, particularly 2-star hotels.

As for travel months, January has the highest customer satisfaction level, followed by February, April, July, May and June. Conversely, the rest months have lower scores than the overall average score especially October, which is the least satisfactory month. In general, customer satisfaction peaks from December to February, followed by March to May and June to August, with the lowest scores observed in September to November. It can be concluded that satisfaction levels are generally higher during the off-peak travel season, with lower levels observed during peak periods.

In terms of customer types, couple and family generally exhibit the highest satisfaction scores, followed by friends, business and solo groups. However, specific score differences are observed within each star category, as detailed in the results section.

Regarding the second research question, the analysis reveals distinct areas of focus for customers across different star categories. In 2-star category, customers prioritize location, followed by room, experience, amenities, staff. In 3-star category, customer experience is highly valued, followed by transaction, location, amenities and staff. For 4-star category, experience receives the highest focus from customers, followed by location, staff, amenities and transaction. For 5-star category, experience is emphasized, followed by facilities, staff, transaction and hotel overall assessment. These findings provide valuable insights for hotels to tailor their services according to customer preferences within each star category.

Regarding the third research question, the analysis offers insights into the frequent terms in the reviews. It concludes that similar terms are utilized across different star-rated hotels. Recurring terms such as "hotel," "room," "staff," "stay," "location," and "great" are consistently mentioned across different hotel categories. These terms are not only frequently mentioned individually but also often co-occur, indicating customers' core and fundamental needs and require attention from hoteliers. However, the emphasis on specific terms varies among different star categories, underscoring the necessity of customizing services to cater to the distinct preferences of customers in each category.

8.2 Limitations

From the data source, the thesis collected reviews exclusively from TripAdvisor. Though TripAdvisor is well-known for its extensive review database, it is conceivable that reviews on the site may carry inherent biases. For instance, the number of reviews for 2-star hotels is relatively low. To mitigate this, reviews from other social platforms

such as Booking.com and Google can offer a more comprehensive perspective. Besides, the chosen time frame focuses solely on post-pandemic area, which is a relatively brief time span. Considering the evolution of reviews over time, especially spanning the periods before, during, and after the pandemic, can offer valuable insights into the changes of customer preferences and the epidemic influences. In addition, the thesis chooses only English reviews and there are many reviews in other languages on social platforms, potentially leading to biased or incomplete results.

Regarding the methodological limitation, though LDA is a robust topic modeling tool, manual interpretation and labeling of the generated topics pose challenges. While some topics are fairly straightforward to label, others are more difficult due to the overlapping words across different topics. Additionally, certain hotel attributes are closely correlated, making it challenging to separate them entirely.

8.3 Future Extensions

Future studies could extend the research in the following ways. Firstly, based on the limitations in the data sources, reviews from other social platforms such as Booking.com and Google can be analyzed to mitigate bias. Secondly, a longitudinal study can explore how customers' focus has evolved in response to the pandemic. In addition, comparing results from reviews in different languages can offer insights into cultural differences within the hotel industry.

In addition, this study has analyzed differences in satisfaction levels among different customer types. Future research can further explore the emphasis in reviews among different customer segments to tailor more targeted improvement measures to meet their needs. Besides, incorporating sentiment analysis into the methodology can provide

deeper insights into customer emotions and preferences.

References

[1] Generalitat Valenciana. EL TURISMO DE LA COMUNITAT VALENCIANA 2023-2024. Turisme CV Magazine. Last access (2024, January 23). https://www.turismecv.com/2024/01/23/los-datos-en-el-turismo-de-la-comunitat-valenciana/

[2] Hai-Jew, S. Career Re-Invention in the post Pandemic era (2022). Business Science Reference., p. 26.

[3] Rivera, M. (2020). Hitting the reset button for hospitality research in times of crisis: Covid19 and beyond. International Journal of Hospitality Management, 87, 102528. https://doi.org/10.1016/j.ijhm.2020.102528

[4] King, C. (1995). What is hospitality? International Journal of Hospitality Management, 14(3–4), 219–234. https://doi.org/10.1016/0278-4319(95)00045-3

[5] Pillai, S. G., Haldorai, K., Seo, W. S., & Kim, W. G. (2021c). COVID-19 and hospitality 5.0: Redefining hospitality operations. International Journal of Hospitality Management, 94, 102869. https://doi.org/10.1016/j.ijhm.2021.102869

[6] Fundació Visit València. Estadísticas de Turismo. Last access (2024, January 23). https://fundacion.visitvalencia.com/sites/default/files/media/downloadable-file/files/Folleto_Estadisticas_2022_WEB.pdf.

[7] Fundació Visit València. SIT: Tourism Intelligence System. (2020, November 9). Last access (2024, January 21). https://fundacion.visitvalencia.com/en/sit

[8] Banerjee, S., & Chua, A. Y. K. (2016). In search of patterns among travellers' hotel ratings in TripAdvisor. Tourism Management, 53, 125–131. https://doi.org/10.1016/j.tourman.2015.09.020

[9] Anderson, E. W. (1998). Customer satisfaction and word of mouth. Journal of Service Research, 1(1), 5–17. https://doi.org/10.1177/109467059800100102

[10] Velázquez, B. M., Blasco, M. F., & Saura, I. G. (2021). Antecedents of online word-of-mouth reviews on hotels. Journal of Hospitality and Tourism Insights, 5(2), 377–393. https://doi.org/10.1108/jhti-10-2020-0184

[11] Moise, M. S., Saura, I. G., Šerić, M., & Molina, M. E. R. (2019). Influence of

environmental practices on brand equity, satisfaction and word of mouth. Journal of Brand Management, 26(6), 646–657. https://doi.org/10.1057/s41262-019-00160-y

[12] Tirado, D., Nilsson, W., Deyà-Tortella, B., & García, C. (2019). Implementation of Water-Saving measures in hotels in Mallorca. Sustainability, 11(23), 6880. https://doi.org/10.3390/su11236880

[13] Pérez, F. J. D., Mouhaffel, A. G., Martín, R. D., & Pino-Otín, M. R. (2018). Comparative study of carbon footprint of energy and water in hotels of Canary Islands regarding mainland Spain. Environment, Development and Sustainability, 21(4), 1763–1780. https://doi.org/10.1007/s10668-018-0102-6

[14] Marco-Lajara, B., & Úbeda-García, M. (2013). Human resource management approaches in Spanish hotels: An introductory analysis. International Journal of Hospitality Management, 35, 339–347. https://doi.org/10.1016/j.ijhm.2013.07.006

[15] Sánchez-Vargas, E., Cerro, A. M. C., Prado-Recio, E., Pasaco-González, B. S., & Moreno-Lobato, A. (2022). Exploring the Hotel Experience in a Cultural City through a UGC Analysis. Sustainability, 14(23), 15695. https://doi.org/10.3390/su142315695

[16] Marrero, S. M., Armas-Cruz, Y., Soto, E. G., & Henríquez, J. M. R. (2022). Evolution of communication policies at luxury hotels and their influence on guest ratings. Current Issues in Tourism, 26(23), 3831–3845. https://doi.org/10.1080/13683500.2022.2151421

[17] Hu, F., Teichert, T., Deng, S., Yong, L., & Zhou, G. (2021). Dealing with pandemics: An investigation of the effects of COVID-19 on customers' evaluations of hospitality services. Tourism Management, 85, 104320. https://doi.org/10.1016/j.tourman.2021.104320

[18] Teng, C., Cheng, Y., Yen, W., & Shih, P. (2023). COVID-19 Perceived risk, travel risk perceptions and hotel staying intention: Hotel hygiene and safety practices as a moderator. Sustainability, 15(17), 13048. https://doi.org/10.3390/su151713048

[19] Moreno-Izquierdo, L., Bernabeu-Bautista, Á., Álvarez, E., & Serrano-Estrada, L.

(2023). How did COVID-19 affect tourism occupancy and prices? A spatiotemporal and economic analysis of Madrid and Valencia through Airbnb geospatial data. International Journal of Applied Earth Observation and Geoinformation, 122, 103448. https://doi.org/10.1016/j.jag.2023.103448

[20] Oliver, R. L. (1980). A cognitive model of the antecedents and consequences of satisfaction decisions. Journal of Marketing Research, 17(4), 460. https://doi.org/10.2307/3150499

[21] Galati, F., & Galati, R. (2019). Cross-country analysis of perception and emphasis of hotel attributes. Tourism Management, 74, 24–42. https://doi.org/10.1016/j.tourman.2019.02.011

[22] Chris Anderson, "The Impact of Social Media on Lodging Performance," Cornell Hospitality Report 12, no. 15 (2012): 6–11

[23] Zhao, X., Wang, L., Guo, X. M., & Law, R. (2015). The influence of online reviews to online hotel booking intentions. International Journal of Contemporary Hospitality Management, 27(6), 1343–1364. https://doi.org/10.1108/ijchm-12-2013-0542

[24] Zhang, & Verma. (2017). What Matters Most to Your Guests: An Exploratory Study of Online Reviews. Cornell Hospitality Report, Vol. 17, No. 4. https://www.chr.cornell.edu

[25] Alrawadıeh, Z., & Law, R. (2019). Determinants of hotel guests' satisfaction from the perspective of online hotel reviewers. International Journal of Culture, Tourism and Hospitality Research, 13(1), 84–97. https://doi.org/10.1108/ijcthr-08-2018-0104

[26] Mann, Prem S. (1995). Introductory Statistics (2nd ed.)

[27] Villegas, F. Descriptive analysis: What it is + best research tips. QuestionPro. Last access (2024, January 25). https://www.questionpro.com/blog/descriptive-analysis/

[28] Netzer, O., Feldman, R., Goldenberg, J., & Fresko, M. (2012). Mine your own business: Market-Structure surveillance through text mining. Marketing Science, 31(3), 521–543. https://doi.org/10.1287/mksc.1120.0713

[29] Han, H. J., Mankad, S., Gavirneni, N., & Verma, R. (2016). What guests really think of your hotel: Text analytics of online customer reviews. Cornell Hospitality

Report, 16(2), 3-17.

[30] Xu, X., & Li, Y. (2016). The antecedents of customer satisfaction and dissatisfaction toward various types of hotels: A text mining approach. International Journal of Hospitality Management, 55, 57–69. https://doi.org/10.1016/j.ijhm.2016.03.003

[31] Blei D. and Lafferty J. (2009). Topic Models, in Text Mining: Classification, Clustering, and Applications, ed. Srivastava, A.N., & Sahami M. CRC Press, Boca Raton, FL, 71-94.

[32] Allahyari, M., Pouriyeh, S. A., Assefi, M., Safaei, S., Trippe, E. D., Gutiérrez, J., & Kochut, K. J. (2017). A brief survey of text mining: Classification, clustering and extraction techniques. arXiv (Cornell University). http://export.arxiv.org/pdf/1707.02919

[33] Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. Journal of Machine Learning Research, 3, 993–1022. https://doi.org/10.5555/944919.944937

[34] Ha, H., Prasad, P. W. C., Maag, A., & Alsadoon, A. (2019). Deep Learning for Aspect-Based Sentiment Analysis: A Comparative Review. Expert Systems With Applications, 118, 272–299. https://doi.org/10.1016/j.eswa.2018.10.003

[35] Xue, Y., Kambhampati, C., Cheng, Y., Mishra, N., Wulandhari, N. B. I., & Deutz, P. (2024). A LDA-Based Social Media Data Mining Framework for Plastic Circular Economy. International Journal of Computational Intelligence Systems, 17(1). https://doi.org/10.1007/s44196-023-00375-7

[36] Guo, Y., Barnes, S. J., & Jia, Q. (2017). Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet allocation. Tourism Management, 59, 467–483. https://doi.org/10.1016/j.tourman.2016.09.009

[37] Google Code Archive - Long-term storage for Google Code Project Hosting. (n.d.). Last access (2024, January 25).https://code.google.com/archive/p/topic-modeling-tool/wikis/TopicModelingTool.wiki

[38] Kaushik, S. Clustering | Different methods, and applications (Updated 2024). Analytics Vidhya. Last access (2024, February 23).https://www.analyticsvidhya.com/blog/2016/11/an-introduction-to-clustering-and-different-methods-of-clustering/

[39] GmbH, R. (n.d.). K-Means - RapidMiner Documentation. Last access (2024, February 25). https://docs.rapidminer.com/latest/studio/operators/modeling/segmentation/k_me ans.html

[40] Jodha, R. FP Growth Algorithm in Data Mining - Scaler Topics. Scaler Topics. Last access (2024, February 25). https://www.scaler.com/topics/data-mining-tutorial/fp-growth-in-data-mining/

[41] GmbH, R. (n.d.). FP-Growth - RapidMiner Documentation. Last access (2024, February 25).https://docs.rapidminer.com/latest/studio/operators/modeling/associations/fp _growth.html

[42]. LaTour, K. A., & Brant, A. (2021). Social listening to create bespoke customer experiences: best practices for hospitality operators. Cornell Hospitality Quarterly, 63(3), 320–333. https://doi.org/10.1177/1938965521993087

[43] Schuckert, M., Liu, X., & Law, R. (2015). Hospitality and tourism online reviews: recent trends and future directions. Journal of Travel & Tourism Marketing, 32(5), 608–621. https://doi.org/10.1080/10548408.2014.933154

[44] Ajzen, I. (1991). The theory of planned behavior. Organizational Behavior and Human Decision Process

Appendix

ANEXO I. RELACIÓN DEL TRABAJO CON LOS OBJETIVOS DE DESARROLLO SOSTENIBLE DE LA AGENDA 2030

Anexo al Trabajo de Fin de Grado y Trabajo de Fin de Máster: Relación del trabajo con los Objetivos de Desarrollo Sostenible de la agenda 2030.

Grado de relación del trabajo con los Objetivos de Desarrollo Sostenible (ODS).

Objetivos de Desarrollo Sostenibles	Alto	Medio	Bajo	No Procede
ODS 1. Fin de la pobreza.				√
ODS 2. Hambre cero.				√
ODS 3. Salud y bienestar.			√	
ODS 4. Educación de calidad.				√
ODS 5. Igualdad de género.				√
ODS 6. Agua limpia y saneamiento.				√
ODS 7. Energía asequible y no contaminante.				√
ODS 8. Trabajo decente y crecimiento económico.			√	
ODS 9. Industria, innovación e infraestructuras.			√	
ODS 10. Reducción de las desigualdades.				√
ODS 11. Ciudades y comunidades sostenibles.			√	
ODS 12. Producción y consumo responsables.				√
ODS 13. Acción por el clima.				√
ODS 14. Vida submarina.				√
ODS 15. Vida de ecosistemas terrestres.				√
ODS 16. Paz, justicia e instituciones sólidas.				√
ODS 17. Alianzas para lograr objetivos.				√

Descripción de la alineación del TFG/TFM con los ODS con un grado de relación más alto.

***Utilice tantas páginas como sea necesario.

The thesis mentions the tourism trend in 2024 that travelers demonstrate a greater willingness to invest in luxury and wellness experiences. This encompasses trips to exclusive resorts, spa treatments, yoga or meditation classes, etc. High-tier hotels in valencia can utilize this trend to attract tourists.