Sentiment Analysis in Guest Reviews: An NLP Approach Applied to Portuguese Hospitality

Eduardo Albuquerque, Daniel Azevedo, Joel Fernandes and Teresa Pataco
CiTUR, ESHT, Polytechnic of Porto, Rua D. Sancho I, 4480-876 Vila do Conde, Portugal

ealbuquerque@esht.ipp.pt
dazevedo@esht.ipp.pt
joelfernandes@esht.ipp.pt
tpataco@esht.ipp.pt

Abstract: Guest reviews represent a valuable resource for strategic decision-making in the hospitality industry. By analysing these evaluations, managers can identify market trends, adjust management practices, reshape marketing strategies, and implement improvements in infrastructure and services. In this way, adopting a data-driven approach based on feedback directly from customers enhances the competitiveness and differentiation of hotels in the tourism market. This study aims to evaluate not only the level of sentiment—dissatisfaction, neutrality, or satisfaction—expressed by each guest regarding a specific product or service, but also how personal and opinionated the published message is. Highly subjective texts are expected to contain emotions, opinions, and individual perceptions, whereas texts with low subjectivity tend to resemble more factual and objective descriptions. To achieve this, a Natural Language Processing (NLP) approach was employed, involving the extraction, translation, and pre-processing of 11,810 reviews from both Portuguese and foreign visitors. The dataset is organised by hotel rating (3, 4, and 5 stars), followed by the application of two complementary sentiment analysis techniques: VADER, to determine polarity (negative, neutral, or positive), and TextBlob, to measure the degree of subjectivity. The results show that the majority of reviews (77%) carry a strongly positive connotation, allowing the conclusion that Portuguese hospitality enjoys a high approval rate among guests. However, by identifying dissatisfaction points and satisfaction patterns specific to different customer segments, managers can better target investments and strategies to further enhance the guest experience.

Keywords: Sentiment Analysis, Natural Language Processing (NLP), Guest Reviews, Portuguese Hospitality, Subjectivity, VADER

1. Introduction

The hospitality industry increasingly relies on user-generated content (UGC), especially online reviews, as a valuable resource for gaining insights into customer experiences(Nikolić, Stojanović and Marjanović, 2025). These reviews are an indispensable guide for consumer decision-making (Ku, Chang and Wang, 2024) and significantly influence customer choices, providing businesses with actionable feedback on service quality and areas for improvement (Nikolić, Stojanović and Marjanović, 2025), (Krishnan et al., 2024) . Online booking sites have become essential for collecting customer feedback data (Le et al., 2025). This feedback is crucial for assessing and enhancing service quality (Shafiezad and Mostofi, 2024), (Erdoğan et al., 2025) as it allows businesses to understand customers' emotional states and needs (Nguyen and Nguyen, 2023).

The rapid growth of the internet and social media has led to an explosion of UGC, which plays a vital role in shaping consumer behaviour across various industries, including tourism and hospitality (Dang and Nguyen, 2023). AI methods can continuously learn and improve from input data and predict customer behaviour (Suman, Vignjevic and Car, 2023). Analysing online reviews through techniques like text mining and sentiment analysis offers actionable insights for hospitality providers to align their offerings with customer demand, build virtual brands, and improve their service quality measures (Krishnan et al., 2024). Data analytics supports data-driven decision-making (Nguyen and Nguyen, 2023) and enables hotels to rapidly and accurately capture customers' emotions, expectations, and perceptions (Erdoğan et al., 2025).

Traditionally, assessing service quality relied on methods like guest comment cards and questionnaires (Jeong and Lee, 2024) or surveys (Vargas-Calderón et al., 2021), which are time- and money-consuming (Vargas-Calderón et al., 2021) and limited by sample size (Mojoodi et al., 2025). The sheer volume of UGC generated on popular review platforms can be overwhelming for researchers and practitioners to handle manually (Le et al., 2025) (Dang and Nguyen, 2023). Furthermore, text data is often unstructured, noisy, and contains various sentiment levels, making it challenging to extract meaningful insights using traditional quantitative methods (Dang and Nguyen, 2023). These conventional approaches lack scalability and automation (Mojoodi et al., 2025), often failing to provide the nuanced, real-time insights needed in a rapidly evolving service sector.

The primary objective is to evaluate how guests feel and what they express in their online hotel reviews, determining whether their comments convey a positive, negative, or neutral emotional tone. This process aims to pinpoint which specific elements of hotel service significantly influence guests' experiences and perceptions. By doing so, the study seeks to thoroughly understand the factors contributing to both customer satisfaction and dissatisfaction, providing a clear and insightful overview of their views on various service components.

2. Literature Review

2.1 Guest Reviews as Strategic Resources

Online guest reviews have emerged as a vital channel for hotel guests to share their travel experiences, significantly influencing booking decisions and providing valuable content for prospective customers (Kazakov et al., 2025). They are increasingly read and shared by the travel community, serving as a digital extension of traditional word-of-mouth (eWOM) (Kazakov et al., 2025), (Le et al., 2025).

Hotel managers can leverage these reviews to obtain actionable feedback regarding service quality, satisfaction, and areas ripe for improvement (Nikolić, Stojanović and Marjanović, 2025). Reviews directly reflect guests' service experiences and interactions (Oshriyeh and Egger, 2025), influencing brand perception and booking decisions and directly impacting a business's online reputation (Erdoğan et al., 2025).

The careful analysis of this feedback provides important insights for improving service processes, understanding customer needs, and gaining a competitive advantage (Erdoğan et al., 2025). Hotels can consistently gather genuine consumer feedback to pinpoint both their strengths and limitations (Mojoodi et al., 2025), and managers can even use these insights to evaluate competitors' performance (Mojoodi et al., 2025). Ultimately, online reviews serve as a crucial resource for gaining deep insights into customer experiences within the hospitality sector (Nikolić, Stojanović and Marjanović, 2025).

2.2 The Role of Customer Feedback in Hospitality Innovation

Customer feedback is a primary source of information in the service sector, driving both experience-sharing and decision-making processes (Erdoğan et al., 2025). It is fundamental for enhancing the overall customer experience in the hospitality industry (Dang and Nguyen, 2023) and is crucial for hoteliers to build a strong reputation and attract new customers (Suman, Vignjevic and Car, 2023).

Sentiment Analysis (SA) allows managers to systematically collect and analyse past consumer experiences, informing service improvements that contribute significantly to customer satisfaction and service recovery(Ameur, Hamdi and Ben Yahia, 2024). Insights derived from opinion mining can help hospitality providers align their offerings with customer demand, build virtual brands, and enhance service quality measures (Krishnan et al., 2024).

By analysing online comments and the critical elements revealed within them, hotel managers can formulate and implement strategies to boost customer satisfaction and loyalty (Mojoodi et al., 2025). Analysing customer reviews helps to pinpoint services that require improvement and identify unmet needs, thereby fostering service innovation (Le et al., 2025). Focusing on the main interests of customers allows hotel managers to achieve efficient and profitable outcomes (Le et al., 2025). Furthermore, addressing varied customer demands, particularly the differences in preferences between international and local guests, helps hoteliers to meet actual needs more effectively (Le et al., 2025).

The findings from advanced analytical models have direct implications for managers and homestay owners in devising appropriate marketing strategies and improving overall guest experience (Krishnan et al., 2024). Crucially, timely and appropriate managerial responses to online feedback are vital for maintaining strong customer relationships and preventing value co-destruction (Oshriyeh and Egger, 2025). This feedback enables hotels to identify specific assets and limitations (e.g., room service), guiding targeted enhancements (Mojoodi et al., 2025). Such responsiveness demonstrates a hotel’s commitment to consumer satisfaction, which in turn enhances consumer loyalty (Mojoodi et al., 2025), and cultivates a strong, positive brand image that attracts new clientele (Mojoodi et al., 2025).

Text mining and Natural Language Processing (NLP) are indispensable for extracting meaningful insights from large volumes of unstructured text, providing hotels with a deeper understanding of their customers (Huang, Liang and Choi, 2022). The application of Artificial Intelligence (AI) methods in this context offers the capability for rapid analysis and extraction of valuable information from guest feedback (Suman, Vignjevic and Car, 2023).

2.3 Sentiment and Subjectivity in Texts

Sentiment analysis (SA) is a computational technique that helps to automatically discover hotel customers' satisfaction from their shared experiences and feelings on social media(Ameur, Hamdi and Ben Yahia, 2024). It involves identifying opinions within reviews and classifying them as positive, negative, or neutral based on the emotions expressed(Ameur, Hamdi and Ben Yahia, 2024). A sentiment can represent a view, opinion, thought, or attitude rooted in emotions (Krishnan et al., 2024), and eWOM sentiment specifically indicates the level of satisfaction articulated by online reviewers (Krishnan et al., 2024).

In the context of hotels and homestays, positive sentiments reflect customer satisfaction and are often associated with emotions such as enthusiasm, happiness, or excitement (Krishnan et al., 2024). Conversely, negative sentiments signify dissatisfaction and are linked to emotions like annoyance and frustration(Krishnan et al., 2024). Neutral sentiments convey satisfaction without explicitly expressed emotions (Krishnan et al., 2024).

Polarity defines the extent to which a text leans towards either a positive or negative valence (Chatterjee, 2020). Reviews exhibiting high polarity are considered "extreme reviews"—either overwhelmingly positive or negative. However, such extreme reviews are generally perceived as untrustworthy by consumers, and high polarity tends to reduce the helpfulness of a review (Chatterjee, 2020).

Subjectivity analysis aims to determine whether a text expresses personal feelings (subjective) or factual details (objective) (Le et al., 2025). Objective information in reviews typically includes concrete facts such as amenities and location (Le et al., 2025), whereas subjective information reflects personal opinions on aspects like staff friendliness and value for money (Le et al., 2025). Subjective reviews are often associated with customer complaints, while objective reviews are typically more logical and compare experiences against expectations (Le et al., 2025). Subjectivity is measured on a scale from 0 (completely objective) to 1 (completely subjective), with a lower score indicating a greater reliance on factual service words rather than personal feelings or opinions (Le et al., 2025).

Emotional tone is a crucial aspect, as emotions significantly influence consumer decision-making and service evaluation processes (Chatterjee, 2020). Emotional expressions serve as key indicators of how service experiences are evaluated (Oshriyeh and Egger, 2025). Specific emotional tones, such as Joy and Confidence, are often associated with value co-creation, while tones like Fear, Anger, Sadness, along with Analytical and Tentative language, tend to result in value co-destruction (Oshriyeh and Egger, 2025). Negative emotions have a more pronounced impact on customer behaviour than positive emotions in eWOM (Oshriyeh and Egger, 2025), and customer perceptions are strongly influenced by the level of emotionality within review texts, irrespective of their overall valence (Oshriyeh and Egger, 2025). Emotions can influence review ratings and sentiments: Joy generally has a positive impact, while Anger, Fear, and Sadness typically lead to negative impacts (Oshriyeh and Egger, 2025). Satisfied customers are prone to express positive comments, whereas unsatisfied customers tend to articulate negative feedback (Oshriyeh and Egger, 2025).

2.4 Applications of NLP in Tourism

Text analytics has been extensively applied to online tourism reviews, employing a variety of analytical methods such as sentiment analysis, trend analysis, text mining, topic modelling, cluster analysis, and predictive analytics (Krishnan et al., 2024) . NLP fundamentally automates the analysis of textual data, enabling the efficient extraction of meaningful insights (Jeong and Lee, 2024). This technology facilitates the identification of specific aspects of customer feedback, such as service quality or room cleanliness, by utilizing targeted keywords, followed by the assessment of keyword frequency and the analysis of associated sentiments (positive or negative) (Jeong and Lee, 2024). NLP is also critical for analysing review content and generating high-quality responses, particularly through fine-grained aspect-based sentiment analysis (Katsiuba, Dolata and Schwabe, 2025).

Topic modelling, a family of unsupervised machine learning techniques including LDA, NMF, and LSA, is widely used to uncover latent themes within extensive collections of text data (Dang and Nguyen, 2023). Text mining methods are applied to discover the underlying structure and implicit meanings present in textual data (Huang, Liang and Choi, 2022) (Kozłowski and Korzeniewski, 2024). The overarching goal of these NLP applications in tourism is to analyse sentiment, identify key topics, track trends, optimize services, and compare performance against competitors, all with the aim of improving service quality and supporting management and marketing strategies (Le et al., 2025). NLP methods are actively used to extract insights from user-generated content across diverse contexts, including hotel, restaurant, and destination reviews (Malik and Bilal, 2024).

Existing studies using VADER, TextBlob, and other tools.

Numerous tools and algorithms have been deployed in NLP-driven sentiment analysis within the hospitality and tourism sectors:

- VADER (Valence Aware Dictionary and sEntiment Reasoner): This lexicon-based tool is commonly used for sentiment analysis in social media text (Kazakov et al., 2025). It computes positive, negative, neutral, and a comprehensive compound score from reviews (Shafiezad and Mostofi, 2024), (Nguyen and Nguyen, 2023). The compound score, which ranges from -1 (most negative) to +1 (most positive), is then used to classify reviews into positive (score > 0.05), neutral (score between -0.05 and 0.05), or negative (score < -0.05) (Shafiezad and Mostofi, 2024). VADER has been successfully applied in studies analyzing tourists' food quality perceptions (Shafiezad and Mostofi, 2024), investigating hotel service micro-elements, and analyzing Vietnamese hotel reviews failures (Kazakov et al., 2025).
- TextBlob: This Python library is utilized for sentiment analysis, enabling the efficient extraction of both polarity and subjectivity from hotel customer reviews (Le et al., 2025). It was notably adopted in a study to investigate the subjectivity and polarity of customer reviews (Le et al., 2025).

2.5 Gap in Literature

Despite significant advancements in text analysis within hospitality and tourism, several gaps remain. Generally, there is a need for more comprehensive review studies on Sentiment Analysis in hospitality (Ameur, Hamdi and Ben Yahia, 2024) . Many existing analyses on value co-destruction often overlook the impact on brand value and the rich textual and emotional insights available, focusing instead on basic sentiment metrics, which necessitates more sophisticated analytical approaches (Oshriyeh and Egger, 2025). Prior research has often been confined to specific destinations, limiting the generalizability of findings and leaving a gap in understanding negative sentiments (Kazakov et al., 2025). Traditional research methodologies like interviews and surveys, while valuable, are limited by small sample sizes, data collection biases, and inability to handle large-scale data (Mojoodi et al., 2025). Furthermore, many studies rely on data collected before the COVID-19 pandemic (Le et al., 2025), and most focus on single platforms, indicating a need for multi-source studies (Le et al., 2025). There is also a recognized lack of research leveraging longitudinal data for service failure and recovery analysis (Huang, Liang and Choi, 2022).

The literature on hotel reviews frequently emphasizes sentiment analysis based solely on text polarity, which often fails to capture the full significance of customer emotions and review valence for brand value (Oshriyeh and Egger, 2025). While machine learning is increasingly used for online customer review processing, there is a limited focus on assessing overall quality of service in the hospitality sector through these methods (Vargas-Calderón et al., 2021). Additionally, given that most studies use data from developed countries and are primarily in English, there is a general call for more research from non-English speaking countries to account for diverse language structures and cultural perspectives (Vargas-Calderón et al., 2021).

Regarding Portuguese hospitality specifically, while some studies touch upon aspects of tourism and hotel reviews in Portugal or neighbouring regions (e.g., small and medium-sized hotels in Portugal (Erdoğan et al., 2025) or reviews from Lisbon, the sources do not explicitly state a specific, identified gap for large-scale sentiment and subjectivity studies focused on Portuguese hotels. However, the general gaps in the literature concerning the depth of sentiment analysis, the exploration of subjectivity, the use of diverse datasets, and the impact of cultural contexts, underscore a broader opportunity. Therefore, there is an implied need for more extensive, large-scale sentiment and subjectivity studies specifically focused on Portuguese hotels, which would significantly contribute to a nuanced understanding of guest experiences and service quality within this particular geographical and cultural context.

3. Methodology

The analysed sample consists of 11.810 comments obtained from Booking.com, made by Portuguese and foreign guests about hotels located in Portugal. The distribution of the comments was based on different hotel categories (3, 4 and 5 stars) and on the guests' continent of origin. This characterisation makes it possible to understand the profile of the reviews and identify patterns in visitors' experiences, contributing to a better understanding of the perceived quality of hotels in Portugal.

Table 1: Guest Reviews Distribution by Continent and Hotel Category

Continent of Origin / Category	3-star	4-star	5-star	Total
Africa	33	83	38	154
America	171	267	112	550
Asia	20	32	38	90
Europe	341	578	402	1.321
Oceania	19	50	28	97
Portugal	2.956	5.080	1.562	9.598
TOTAL	**3.540**	**6.090**	**2.180**	**11.810**

The distribution of comments by hotel category shows that most reviews are concentrated in 4-star hotels, which account for 51.58% of the total, with 6,090 comments. Three-star hotels gather 3,540 comments (29.97%), while five-star hotels account for 2,180 comments (18.45%). This pattern suggests that 4-star hotels are the most popular choice among guests in Portugal, either due to the market offering or the balance between price and quality. The smaller proportion of comments for 5-star hotels may be linked to a more restricted and selective clientele, whereas the relatively high percentage of reviews for 3-star hotels indicates that this category remains in demand, possibly among travellers more sensitive to accommodation costs.

Regarding guests' origin, 9,598 comments (81.3%) were made by Portuguese guests, while foreign guests contributed 2,212 comments (18.7%). Within the foreign group, Europe accounts for the majority of reviews, with 1,321 comments (11.2%), reflecting the strong presence of European tourists in the country. Next are guests from the Americas, with 550 comments (4.7%), followed by tourists from Africa (154 comments, 1.3%), Oceania (97 comments, 0.8%) and Asia (90 comments, 0.8%). These figures show that the Portuguese tourism market is predominantly national and European, while visitors from other continents, although present, represent a smaller share of the sample.

The predominance of comments from Portuguese guests may be related to the growing practice of domestic tourism, reinforced in recent years by factors such as the COVID-19 pandemic, which encouraged Portuguese travellers to explore their own country. Meanwhile, the significant participation of European tourists confirms Portugal's importance as a well-established destination within Europe. On the other hand, the lower representation of guests from Asia, Africa and Oceania suggests that these markets still have a limited impact on the overall volume of tourists reviewing Portuguese hotels.

This analysis provides a clearer understanding of the profile of guests reviewing Portuguese hotels and how different hotel categories are perceived by tourists from various regions. Knowledge of these trends can be crucial for the hospitality sector to adjust its strategies and enhance the experience offered to visitors, ensuring services are better aligned with the expectations of different guest profiles.

The present study aims to analyse the polarity and subjectivity of comments from guests of Portuguese hotels using Natural Language Processing (NLP) techniques. To this end, the ten most recent comments were extracted from a sample of 1,181 Portuguese hotels, allowing for a focused analysis of updated reviews. As most comments were written in Portuguese, it was necessary to translate them into English to ensure better compatibility with the sentiment analysis tools employed. For this task, the Deep Translator library was used, a tool that enables automatic text translation through various APIs, such as Google Translate, DeepL, Microsoft Translator, MyMemory and Yandex. The use of this library facilitated text conversion. To guarantee semantic integrity and achieve more accurate sentiment analysis, validation was carried out by a professional in translation and linguistics.

After extracting and translating the comments, the data were stored in an Excel file, ensuring the preservation of the original structure and compatibility with subsequent processing steps. Next, text pre-processing was performed, an essential step to ensure analysis quality. This phase included the removal of special characters, numbers and unnecessary punctuation, as well as text normalisation by converting all words to lowercase. Additionally, in specific cases, language detection was carried out to confirm the adequacy of the translations.

Sentiment analysis was conducted using two complementary methods. First, the VADER (Valence Aware Dictionary and sEntiment Reasoner) model was applied, a lexicon-based model developed specifically for analysing informal texts such as comments and online reviews. This method assigns a polarity score to the text on a continuous scale from -1 (extremely negative) to 1 (extremely positive), enabling a quantitative assessment of the emotional tone of the comment. In addition, TextBlob was used, a tool that measures the degree of subjectivity in the comment, distinguishing between more objective opinions and more subjective expressions.

Based on the polarity score obtained through the VADER method, comments were classified into five distinct categories, allowing for a more detailed analysis of their emotional distribution. Comments with polarity less than or equal to -0.6 were classified as very negative, reflecting extremely unsatisfactory evaluations. Comments between -0.6 and -0.2 were categorised as negative, expressing lower-intensity dissatisfaction. Texts with polarity ranging from -0.2 to 0.2 were considered neutral, as they did not show a clear bias towards positive or negative sentiments. Comments with scores between 0.2 and 0.6 were classified as positive, revealing favourable opinions but without a high degree of enthusiasm. Finally, comments with polarity equal to or greater than 0.6 were categorised as very positive, indicating great satisfaction on the part of guests.

4. Results

The analysis of the polarity of guest reviews for Portuguese hotels demonstrates a widely positive perception of the quality of services provided. Based on the segmentation by hotel category (3, 4, and 5 stars), the results clearly show that 77% of the reviews are highly positive, 17% are positive, and only 6% are neutral, whilst negative and highly negative reviews account for a mere 1% of the total.

5-star hotels exhibit the highest proportion of highly positive reviews (78%), followed by 4-star (77%) and 3-star (75%) establishments, confirming that higher-category establishments tend to provide more satisfactory experiences. However, the high satisfaction rate in 3- and 4-star hotels indicates that many of these establishments are capable of meeting or even exceeding guest expectations.

The proportion of neutral and negative reviews is negligible, reinforcing the perception that the Portuguese hospitality sector is highly regarded by tourists. Furthermore, the results suggest that investment in service improvement, especially within lower-category hotels, could generate a positive impact on guest perception. The high consistency of positive reviews underscores the importance of excellence in service, comfort, and the overall customer experience as essential factors for public satisfaction.

4.1 Analysis of Review Polarity by Continent and Hotel Category

This analysis examines the distribution of polarity in guest comments from different continents regarding Portuguese hotels, segmenting the results by hotel category (3, 4, and 5 stars).

The data were analysed based on the VADER_Polarity metric, which measures the degree of positivity or negativity of the comments, thereby enabling a quantitative assessment of guest satisfaction.

4.2 Distribution of Polarity by Continent

The analysis by continent reveals that guests from Asia (0.80) and Oceania (0.82) present the most positive reviews, whilst guests originating from Africa (0.71) and Portugal (0.72) record the lowest averages. Tourists from the Americas (0.75) and Europe (0.75) are situated at an intermediate level of satisfaction.

These findings suggest that tourists hailing from Asia and Oceania tend to have a more favourable assessment of Portuguese hotels, possibly due to different cultural expectations, comparison standards, or even the way they express satisfaction. Conversely, African and Portuguese guests exhibit slightly lower averages, which may indicate a greater degree of demandingness or higher expectations.

4.3 Analysis by Hotel Category

The average polarity also varies according to the hotel category, confirming that higher-category hotels tend to generate more positive reviews:

- 3-star hotels present the lowest average polarity (0.71), indicating that guests tend to be slightly more critical in this category.
- 4-star hotels record an average polarity of 0.73, showing an improvement in guest perception compared to 3-star hotels.
- 5-star hotels achieve the highest average polarity (0.74), suggesting an experience of greater satisfaction, possibly due to a superior standard of quality and service.
- These results reflect the expected trend that higher-category hotels receive more favourable reviews, as guests anticipate an elevated level of comfort and service.

Table 2: Average Score and Guest Reviews Distribution by Continent and Hotel Category

	3*	4*	5*	Total
Africa				
No. of Comments	33	83	38	**154**
Average Score	0,67	0,70	0,76	**0,71**
America				
No. of Comments	171	267	112	**550**
Average Score	0,72	0,76	0,80	**0,75**
Asia				
No. of Comments	20	32	38	**90**
Average Score	0,76	0,86	0,78	**0,80**
Europe				
No. of Comments	341	578	402	**1.321**
Average Score	0,69	0,78	0,77	**0,75**
Oceania				
No. of Comments	19	50	28	**97**
Average Score	0,81	0,81	0,84	**0,82**
Portugal				
No. of Comments	2956	5080	1562	**9.598**
Average Score	0,72	0,72	0,73	**0,72**
Total Comments	**3.540**	**6.090**	**2.180**	**11.810**
Overall Average Score	**0,71**	**0,73**	**0,74**	**0,73**

4.4 Comparison Between Continents Within Each Hotel Category

The detailed analysis of polarity by continent within each hotel category allows for the identification of more specific patterns:

- In 3-star hotels, guests from Oceania (0.81) and Asia (0.76) present the most positive reviews, whilst tourists from Africa (0.67) and Europe (0.69) demonstrate greater reservation in their assessments.
- In 4-star hotels, Asia (0.86) records the highest polarity, standing out from the other continents. This value indicates a high level of satisfaction among Asian guests with hotels in this category.
- Finally, in 5-star hotels, Oceania (0.84) maintains the highest average, suggesting that guests from this region significantly value the services offered.

The behaviour of Asian guests in 4-star hotels is particularly interesting, as the average (0.86) exceeds the average for 5-star hotels (0.78) for that same region. This finding may indicate that Asian guests consider 4-star hotels to be an excellent value-for-money option, whilst expectations in 5-star hotels might be more demanding.

4.5 Distribution of the Number of Comments

The count of reviews reveals that the majority of comments originate from Portuguese guests (9,598 reviews, 81.3% of the total), reflecting a strong presence of domestic tourism. Among foreign guests, Europe (1,321 comments, 11.2%) represents the largest proportion, followed by the Americas (550 comments, 4.7%) and Asia (90 comments, 0.8%). Oceania (97 comments, 0.8%) and Africa (154 comments, 1.3%) record the lowest participation in the sample.

The predominance of Portuguese guests may be associated with the growth of domestic tourism, driven by factors such as seasonal promotions and international travel restrictions in recent years. The significant participation of European tourists confirms the importance of Portugal as a consolidated destination in the European market.

5. Conclusion

The results indicate that the perception of the quality of Portuguese hotels is broadly positive, regardless of the guests' origin or the hotel category. However, there are significant differences in the average review polarity between continents, suggesting variations in expectations and the way tourists express their satisfaction.

The analysis reveals that tourists from Asia and Oceania are the most positive in their assessment of Portuguese hotels, whilst guests from Africa and Portugal demonstrate greater demandingness. Furthermore, satisfaction increases as the hotel category rises, being most pronounced in 5-star hotels.

These results have relevant implications for the Portuguese hospitality sector, as they indicate that personalised service and marketing strategies could be developed for different guest profiles. Special attention to the African and Portuguese markets, where satisfaction tends to be slightly lower, could contribute to elevating the customer experience and further improving the perception of hotel services in Portugal.

The data analysis reveals different aspects of the guest experience concerning the hotel's cleanliness, dining, services, and amenities.

With regard to cleanliness, the terms "clean" (1,424 mentions, impact of 2.94) and "cleaning" (1,206 mentions, impact of 2.49) indicate that hygiene is frequently mentioned but has a relatively low impact. This suggests that while it is a relevant factor, it does not stand out as a major strength of the hotel. Furthermore, "rooms" (850 mentions, impact of 1.75) reinforces the perception that the guest rooms are commented upon but are not among the most praised aspects, while "comfortable" (1,992 mentions, impact of 4.11) suggests the stay was considered comfortable, possibly due to the organisation and condition of the rooms.

Regarding dining, breakfast emerges as the main highlight, with "breakfast" (4,272 mentions, impact of 8.82) leading the reviews, followed by terms such as "good" (3,723 mentions, impact of 7.68) and "excellent" (3,033 mentions, impact of 6.26). This clearly indicates that the quality of the food is a highly positive factor in the guest experience.

Concerning services, words such as "staff" (2,419 mentions, impact of 4.99) and "employees" (2,713 mentions, impact of 5.60) demonstrate that the personnel play an essential role in the guest experience, being frequently described as "friendly" (2,112 mentions, impact of 4.36) and "niceness" (2,002 mentions, impact of 4.13). However, the perception of the service in general ("service" with 1,171 mentions and impact of 2.41) is less prominent than that of dining, suggesting opportunities for improvement in efficiency and personalisation of the service delivery.

In relation to amenities, terms such as "pool" (1,024 mentions, impact of 2.11) and "view" (834 mentions, impact of 1.72) indicate that the hotel's infrastructure is acknowledged but does not stand out among the primary factors influencing guest satisfaction.

Thus, it is concluded that breakfast represents the hotel's greatest strength, whilst the perception of cleanliness, services, and amenities could be improved to enhance the overall customer experience.

Acknowledgements

This work is financed by national funds through FCT - Foundation for Science and Technology, IP, under the Projet CiTUR UID/04470/2025

Ethics Declaration

Formal ethical approval was not required for the conduct of this research, as the study relied solely on the analysis of secondary data (guest reviews available publicly/in non-sensitive databases) and did not involve the direct participation, intervention, or interaction with human subjects or animals, nor the collection of confidential personal data. The manipulation and filtering of the data were carried out strictly for the purposes of statistical analysis and visualisation (Word Cloud), with no direct ethical implications in data handling.

AI Declaration

Artificial Intelligence (AI) was employed as a high-productivity tool, providing assistance during the initial and technical phases of this project. Its use primarily centred on brainstorming and conceptual idea generation, and served as a resource for literature search by rapidly identifying and synthesising key information and publications for theoretical contextualisation. Furthermore, the AI proved crucial for Python optimization, offering solutions for debugging and refining the code's logic (including POS Tagging and data manipulation), thus ensuring the analysis's efficiency. Nevertheless, the critical analysis, filter selection, final interpretation of results, and the overall authorship of the work remain the exclusive responsibility of the author.

References

Ameur, A., Hamdi, S. and Ben Yahia, S. (2024) 'Sentiment Analysis for Hotel Reviews: A Systematic Literature Review', ACM Computing Surveys, 56(2). Available at: https://doi.org/10.1145/3605152.

Chatterjee, S. (2020) 'Drivers of helpfulness of online hotel reviews: A sentiment and emotion mining approach'', International Journal of Hospitality Management, 85. Available at: https://doi.org/10.1016/j.ijhm.2019.102356.

Dang, T.D. and Nguyen, M.T. (2023) 'Mining User-Generated Content for Enhanced Customer Experience in Hospitality Industry', in. 2023 3rd International Conference on Emerging Smart Technologies and Applications, eSmarTA 2023. Available at: https://doi.org/10.1109/eSmarTA59349.2023.10293621.

Erdoğan, D. et al. (2025) 'Developing a Deep Learning-Based Sentiment Analysis System of Hotel Customer Reviews for Sustainable Tourism', Sustainability (Switzerland), 17(13). Available at: https://doi.org/10.3390/su17135756.

Huang, S., Liang, L.J. and Choi, H.C. (2022) 'How We Failed in Context: A Text-Mining Approach to Understanding Hotel Service Failures', Sustainability (Switzerland), 14(5). Available at: https://doi.org/10.3390/su14052675.

Jeong, N. and Lee, J. (2024) 'An Aspect-Based Review Analysis Using ChatGPT for the Exploration of Hotel Service Failures', Sustainability (Switzerland), 16(4). Available at: https://doi.org/10.3390/su16041640.

Katsiuba, D., Dolata, M. and Schwabe, G. (2025) 'Joining forces for online feedback management: policy recommendations for human-AI collaboration', Data and Policy, 7. Available at: https://doi.org/10.1017/dap.2025.13.

Kazakov, S. et al. (2025) 'Improving guest satisfaction by identifying hotel service micro-elements failures through Deep Learning of online reviews', Cuadernos de Gestion, 25(1), pp. 71–88. Available at: https://doi.org/10.5295/cdg.242191sk.

Kozłowski, M. and Korzeniewski, J. (2024) 'Studying the effectiveness of investments in hotel services using customer sentiment analysis', Quality and Quantity [Preprint]. Available at: https://doi.org/10.1007/s11135-023-01808-3.

Krishnan, J. et al. (2024) 'Survival strategies for family-run homestays: Analyzing user reviews through text mining', Data Science and Management, 7(3), pp. 228–237. Available at: https://doi.org/10.1016/j.dsm.2024.03.003.

Ku, C., Chang, Y. and Wang, Y. (2024) 'How to strategically respond to online hotel reviews: A strategy-aware deep learning approach'', Information & Management, 61(5). Available at: https://doi.org/10.1016/j.im.2024.103970.

Le, H.T.M. et al. (2025) 'Mining online hotel reviews using big data and machine learning: An empirical study from an emerging country', Annals of Tourism Research Empirical Insights, 6(1). Available at: https://doi.org/10.1016/j.annale.2025.100170.

Malik, N. and Bilal, M. (2024) 'Natural language processing for analyzing online customer reviews: a survey, taxonomy, and open research challenges', PeerJ Computer Science, 10. Available at: https://doi.org/10.7717/PEERJ-CS.2203.

Mojoodi, A. et al. (2025) 'Perceptual mapping and key factors influencing hotel choices: A web mining approach to Booking.com reviews', International Journal of Hospitality Management, 131. Available at: https://doi.org/10.1016/j.ijhm.2025.104308.

Nguyen, H.T.T. and Nguyen, T.X. (2023) 'Understanding customer experience with Vietnamese hotels by analyzing online reviews', Humanities and Social Sciences Communications, 10(1). Available at: https://doi.org/10.1057/s41599-023-02098-8.

Nikolić, M., Stojanović, M. and Marjanović, M. (2025) 'The Power of Words: Leveraging Deep Learning Techniques to Predict Hotel Ratings from User Reviews', in. 2025 24th International Symposium INFOTEH-JAHORINA, INFOTEH 2025 - Proceedings. Available at: https://doi.org/10.1109/INFOTEH64129.2025.10959201.

Oshriyeh, O. and Egger, R. (2025) 'The double-edged sword of user reviews: how customer feedback can build and break hotel brands', Journal of Hospitality Marketing and Management [Preprint]. Available at: https://doi.org/10.1080/19368623.2025.2494567.

Shafiezad, O. and Mostofi, H. (2024) 'Sentiment Analysis of Berlin Tourists' Food Quality Perception Through Artificial Intelligence', Tourism and Hospitality, 5(4), pp. 1396–1417. Available at: https://doi.org/10.3390/tourhosp5040078.

Suman, S., Vignjevic, M. and Car, T. (2023) 'Information extraction and sentiment analysis of hotel reviews in Croatia'', Zbornik Veleučilišta u Rijeci, 11(1), pp. 69–89. Available at: https://doi.org/10.31784/zvr.11.1.5.

Vargas-Calderón, V. et al. (2021) 'Machine learning for assessing quality of service in the hospitality sector based on customer reviews', Information Technology and Tourism, 23(3), pp. 351–379. Available at: https://doi.org/10.1007/s40558-021-00207-4.