MARKETING SCIENCE & INSPIRATIONS

Marketing and Marketing Management Journal

msijournal.com Printed on April 9, 2026

IDENTIFICATION OF SATISFACTION AND DISSATISFACTION FACTORS OF HOTEL CUSTOMERS USING NATURAL LANGUAGE PROCESSING TECHNIQUES

Cherdouh, S., Kebir, S., & Meslem, H. (2025). Identification of satisfaction and dissatis-faction factors of hotel customers using natural language processing techniques. Marketing Science & Inspirations, 20(3), 27–45. https://doi.org/10.46286/msi.2025.20.3.4

This paper proposes a novel approach to identifying the factors that influence satisfaction and dissatisfaction among Algerian hotel customers through the analysis of online customer reviews. Unlike traditional quantitative methods such as questionnaires, this study employs advanced natural language processing techniques to uncover key insights into customer experiences. The study employs natural language processing techniques to extract and analyze data from online customer reviews. This method aims to identify significant concerns and satisfaction factors mentioned by Algerian hotel customers, offering an innovative alternative to conventional survey-based approaches. The analysis revealed that satisfaction factors are specific, tangible aspects of the customer's experience, which can be easily conceptualized. In contrast, dissatisfaction factors are more abstract and challenging to define, which makes them more difficult to comprehend. The paper introduces an innovative approach by leveraging natural language processing to analyze customer reviews, offering a fresh perspective on understanding customer satisfaction and dissatisfaction. This methodology provides valuable insights into customer experiences and highlights the differences in how satisfaction and dissatisfaction are perceived and articulated by customers.

1 INTRODUCTION

The Internet has become one of the main channels for offering and demanding services across all sectors. With the advent of Web 2.0 and the emphasis on competition and e-reputation, it has become vital for businesses to have an exemplary image and offer in order to retain their clients and attract new ones. The tourism industry, and particularly the hospitality sector, is no exception to this rule (Yassin 2022). Indeed, the use of specialized travel platforms such as TripAdvisor, Expedia, and Booking.com has become a deeply ingrained habit among tourists for optimizing their travel planning. However, the use of these platforms is not limited to booking flights and hotel rooms; it also extends to other activities such as sharing reviews and experiences (Sangkaew and Zhu 2020; Xin et al. 2023), as well as recommending hotels (Nilashi et al. 2018). These activities are collectively referred to as Online Customer Reviews (OCR).

OCRs are part of what is commonly referred to as user-generated content (Krumm et al. 2008). This concept, in the context of big data terminology, refers to the phenomenon of engaging the public as active participants in the voluntary creation of subjective online content (Li et al. 2018). This contrasts with the traditional approach, where online content is generated, created, and disseminated solely by companies in a one-sided manner.

OCRs have progressively emerged as an important source of information, significantly influencing consumers' decision-making in hospitality purchases (Sparks and Browning 2011; Vermeulen and Seegers 2009). Due to their open format, OCRs enable customers to thoroughly and accurately capture their consumption experiences and perceptions (Xiang et al. 2015). When travelers write an online review about a hotel they stayed at, they subjectively and explicitly describe their experiences, whether positive or negative, detailing what they liked or disliked during their stay (He et al. 2017). Additionally, they can assign a score, which generally reflects their level of satisfaction or dissatisfaction (Geetha et al. 2017; Zhu et al. 2020). These ratings are often influenced by specific attributes, with some factors tending to increase ratings and others tending to lower them (Gunasekar and Sudhakar 2019). According to Park et al. (2018), feedback from repeat visitors tends to contain longer sentences and express more pronounced positive or negative sentiments compared to one-time visitors. In contrast, reviews from first-time visitors often include more analytical and anxious language, reflecting a different evaluative approach than that of repeat guests.

In the hospitality context, OCRs provide operators with a rich source of information that can be exploited and analyzed in an automated and continuous manner, unlike traditional approaches such as opinion surveys based on questionnaires, where data collection is a time-consuming and resource-intensive task (Fernández et al. 2016). However, with the exponential increase in their volume, it is inconceivable to process OCRs in their raw form. For these reasons, an increasing number of researchers in the hospitality and tourism fields are turning to innovative techniques from the domain of machine learning, particularly Natural Language Processing (NLP) (Kang et al. 2020), which is an area of research and application in artificial intelligence that explores how computers can be used to understand and manipulate natural language text or speech to do useful things (Hirschberg and Manning 2015). Application of NLP includes several fields of studies (Chowdhury 2003) such as sentiment analysis (Medhat et al. 2014), automatic translation (Wang et al. 2022), and topic modeling (Vayansky and Kumar 2020).

Our research problem is framed within this context. In this paper, we propose an NLP-based approach to uncover the factors driving satisfaction and dissatisfaction among customers of Algerian hotels, leveraging online reviews as our primary data source. To ensure a comprehensive and robust analysis, we integrate NLP techniques, including sentiment analysis, text preprocessing (Anandarajan et al. 2019), topic modeling, and keyword extraction (Firoozeh et al. 2020). These techniques enable us to structure and interpret the data effectively, uncovering key themes and terms that shape customer experiences. By combining these techniques, we aim to provide a deeper and more nuanced understanding of the drivers of customer satisfaction and dissatisfaction in the Algerian hospitality context, addressing the following research questions:

Research question 1: Does the sentiment expressed by customers in their online reviews explain their overall satisfaction?

Research question 2: What are the most important factors mentioned by customers in their online reviews?

Research question 3: Is there a relationship between these factors and the customer's overall satisfaction?

2 THEORETICAL BACKGROUND

OCRs have become a critical resource in the hospitality industry for understanding customer perceptions and identifying factors that influence satisfaction and dissatisfaction (Park et al. 2018; Padma and Ahn 2020). Unlike traditional methods such as face-to-face interviews or surveys, OCRs provide a scalable and dynamic means of capturing customer feedback, enabling researchers and practitioners to uncover nuanced insights into guest experiences (Zhao et al. 2019). However, the exponential growth of user-generated content and the advent of Big Data have made manual analysis of OCRs impractical, necessitating the use of advanced Natural Language Processing (NLP) techniques to process and extract meaningful insights from this data (Álvarez-Carmona et al. 2022; Khurana et al. 2023).

To address these challenges, researchers have employed various NLP techniques to analyze customer satisfaction in the hospitality industry. For instance, Arindra et al. (2024) analyzed 12,949 user reviews from TripAdvisor using an NLP approach to identify key factors influencing stay experience and satisfaction. Their findings highlight that ease of booking plays a crucial role in enhancing satisfaction, while issues related to service, facilities/amenities, and the overall stay experience are primary contributors to customer dissatisfaction. Luo et al. (2020) conducted a comprehensive analysis of 363,723 reviews from Chinese economy hotels using deep learning-based sentiment analysis. Their findings revealed that positive sentiments are most frequently associated with location, followed by facilities, service, price, image, and reservation experience. Conversely, negative sentiments were primarily linked to issues such as sound insulation, air conditioning, bedding, toilets, and other hotel amenities. Similarly, Cheng and Jin (2019) found that noise was a major source of dissatisfaction among Airbnb users, highlighting the universal challenge of environmental factors in guest experiences. Complementing these findings, Saraswati et al. (2024) identified room replacement policies as another critical area requiring improvement, while Aakash and Aggarwal (2020) emphasized that high-quality standards in rooms, service, cleanliness, location, and value are essential determinants of overall hotel performance and guest satisfaction.

However, it is important to recognize that the factors influencing customer satisfaction and dissatisfaction are not static and can vary significantly depending on several criteria (Xu and Li. 2016). For example, customer satisfaction and expectations are influenced by factors such as origin (domestic vs. international) and hotel star ratings, which moderate the impact of hotel attributes on satisfaction (Li et al. 2020). Furthermore, satisfaction levels can vary depending on the trip mode, even for the same traveler, as highlighted by Liu et al. (2013). These variations underscore the complexity of guest experiences and the need for tailored approaches to address diverse customer needs and preferences. Expanding on this, Roy (2023) examined online reviews across different hotel tiers using the Theory of Lodging (ToL), revealing that guests in luxury hotels tend to focus on subjective evaluations, such as personalized service and ambiance, whereas guests in low-tier hotels rely more on objective evaluations, such as cleanliness and value for money.

In addition to these factors, geographic and regional variations in customer sentiment have also been explored, offering further insights into the contextual influences on customer satisfaction. The study conducted by Bulkrock and Alsharman (2024) revealed a significant geographic variation in guest sentiment across cities, states, and countries. Similarly, Carvalho et al. (2024) investigated customer satisfaction in mountain hotels within UNESCO's Global Geoparks, analyzing 5,590 online reviews from 20 hotels in the Estrela UNESCO Global Geopark. Their study identified factors such as seasonality, nationality, and travel experience as significant influences on satisfaction, with pool and spa facilities emerging as particularly important determinants of guest satisfaction.

Beyond geographic and contextual factors, recent research have also examined the role of technology and sensory experiences in shaping customer satisfaction, further enriching our understanding of the hospitality landscape. Özen and Katlav (2023) analyzing 12,396 reviews evaluate customer satisfaction with technology-supported products in hotels. Their findings indicated that technology integration positively impacts guest satisfaction, particularly when it enhances basic services like room lighting and bedding at an affordable cost. However, Cherdouh et al. (2022) found that while information and communication technologies (ICT) contribute to customer satisfaction in Algerian hotels, their impact is less significant compared to non-ICT services. Building on these insights, Lee et al. (2019) emphasize the critical role of multisensory experiences in enhancing customer satisfaction. Their findings suggest that multisensory experiences facilitate the evaluation process, with positive multisensory experiences amplifying positive affect, thereby significantly increasing customer satisfaction. In a similar vein, Luo et al. (2021) highlight the growing role of robots and artificial intelligence in the hospitality industry, emphasizing their potential to enhance customer satisfaction. By analyzing online reviews, their research identifies a positive correlation between guests' sentiments toward robotic services and their overall hotel satisfaction.

In our study, we aim to achieve our goals by combining different NLP techniques. First, we seek to determine whether the sentiment expressed by customers in their online reviews explains their overall satisfaction. Second, we aim to identify the most important factors mentioned by customers in their online reviews. Third, we investigate whether there is a relationship between these factors and customers' overall satisfaction.

Our approach differs from traditional topic modeling by clustering words based on their semantic similarity rather than simple co-occurrence, and from ABSA methods by exploring themes and their correlation with satisfaction without relying on automated sentiment analysis tools. Indeed, sentiment analysis tools were only used as a validation tool to verify the consistency between the user's rating and the sentiment expressed in the review.

In the following sections, we will first present the methodology adopted in this study. Subsequently, we will present the key results obtained and discuss their significance. Finally, we will examine the theoretical and practical implications of our findings.

3 METHODOLOGY

3.1 DATA COLLECTION

To conduct our study, we utilized TripAdvisor as our primary data source. Established in the early 2000s, TripAdvisor is one of the largest and most widely used platforms for OCRs. By early 2022, the number of OCRs on TripAdvisor had surpassed one billion (Statista 2022). The platform enables users to post, comment on, and share travel recommendations, as well as rate hotels, restaurants, and destinations. Each review on TripAdvisor includes several key pieces of information, such as the review title, body, publication date, hotel name, star rating, hotel location (city and country), and the customer's rating (on a scale of 1 to 5).

The data collection process in our study consisted of three successive steps. First, we developed a Python program to extract reviews from TripAdvisor. This program takes a list of hotel URLs as input and generates a

file containing all the extracted reviews. We manually collected the URLs of the top 144 Algerian hotels listed on TripAdvisor, sorted in descending order of their ratings. After gathering the reviews from these hotels, we retained only those written in French and English, as reviews in other languages (e.g., Arabic, Italian, Chinese) were extremely limited in number. Including these reviews would have compromised the reliability of our results. Among the collected reviews, the most recent one dates from December 2024, and the oldest one dates from September 2015. Additionally, since most sentiment analysis libraries are optimized for English text, we translated all French reviews into English to ensure consistency and accuracy in our analysis.
The translation was performed automatically using a Python program that employs the T5 translation model developed by Google (Raffel et al. 2020), which is one of the most downloaded text translation models on the Hugging Face Model Hub (HuggingFace 2022). Additionally, recent studies have shown that T5 achieves a significantly lower Translation Error Rate compared to other translation models, indicating excellent performance in multilingual translation tasks (Zhu et al. 2025). This platform is a repository that hosts state-of-the-art machine learning models dedicated to natural language processing (NLP), created and maintained by leading artificial intelligence researchers and major tech companies such as Google, Facebook, and Microsoft (Wolf et al. 2020). Once the translation is complete, all reviews are collected and stored in a single file containing relevant information about each review, such as the title, text, reason for the stay, review URL, and more. Table 1 illustrates the structure of a review.

Table 1: Online customer reviews structure
Source: TripAdvisor

3.2 DATA CLEANING AND PREPROCESSING

A total of 11,957 reviews were initially collected from users across various hotels. However, to conduct a reliable statistical analysis of the data, we only retained 11,310 reviews concerning 3, 4, and 5-star hotels and containing more than three words, from the 11,957 user reviews. The 1 and 2-star hotels were excluded from the study because the number of their reviews represented only 5% of the total number of reviews collected. This small proportion was deemed insufficient to provide meaningful insights or to significantly influence the overall analysis. By concentrating on higher-rated hotels, we aimed to capture a more representative and consistent sample that would allow for a robust examination of user feedback. Table 2 describes the characteristics of the sample of hotel reviews that we collected.

Table 2: Profile characteristics
Source: Authors

4 RESULTS

4.1 RESEARCH QUESTION 1

To address research question 1, which examines the sentiment expressed by the customer in their review and its effect on the rating they gave to the hotel, we defined and calculated the following functions for each collected review:
• score(OCR): indicates the score assigned to the hotel by the client. Its value is an integer ranging from 1 (very dissatisfied) to 5 (very satisfied). This score reflects the overall satisfaction of the client with regard to the hotel.
• sentimentlib(OCR): denotes the polarity of the sentiment in the OCR text, measured using the lib library. Its value is a real number within the interval , where: -1.0 indicates a very negative sentiment and 1.0 indicates a very positive sentiment.

To provide a comprehensive answer to research question 1, we used four different sentiment analysis libraries: TextBlob (Loria 2020), Vader (Hutto and Gilbert 2014), Flair (Akbik et al. 2019), and Transformers (Wolf et al. 2020). TextBlob and Vader are both Python sentiment analysis libraries based on a lexicon, meaning that for these two libraries, the sentiment of a given text is an aggregate of weights assigned to the words in that text. For example, the words „good," „great," and „happy" have a positive weight, while the words „horrible," „difficult," and „unhappy" have a negative weight. Flair and Transformers, on the other hand, are two Python libraries based on machine learning for sentiment analysis. That is, both use supervised learning models trained on large text corpora. Machine learning-based sentiment analysis libraries generally offer better accuracy than lexicon-based libraries because they operate not directly on the text itself, but on a tree representation of the text that captures the intensity of the relationships between words. However, due to the computational and memory requirements for their implementation, machine learning-based libraries require significantly more execution time than lexicon-based libraries.
To examine the potential influence of customer sentiment expressed in their online reviews (OCR) on their overall satisfaction with the hotel, we visually analyzed the distribution of the sentimentlib(OCR) function values across the five levels of overall satisfaction, as measured by the score(OCR) function. Figure 1 presents this analysis using box plot charts for each sentiment analysis library.

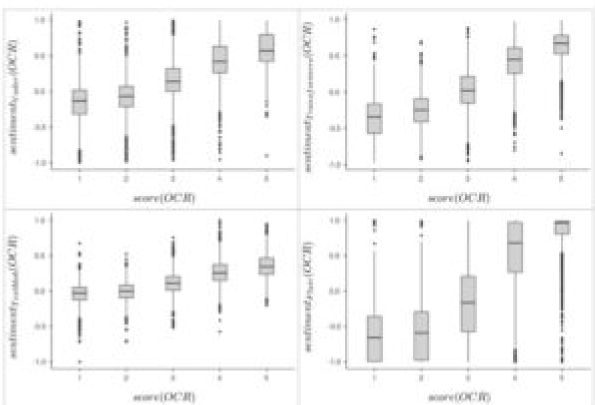

Figure 1: Boxplots describing the distribution of sentiment scores
Source: Authors

At first glance, the four charts indicate that there is a positive correlation between the sentiment expressed by the client in their OCR and the score they assigned to the hotel. However, considering the respective width of

the boxes, which indicates data dispersion, it is surprisingly noted that lexicon-based sentiment analysis libraries appear to be more accurate than those based on machine learning.
To validate our response to research question 1, we conducted a correlation analysis between the values of the two functions, sentimentlib(OCR) and score(OCR), by measuring the Pearson correlation coefficient . Table 3 presents the results of the correlation test for the four sentiment analysis libraries.

Notes: ***
Table 3: Correlation test
Source: Authors

The results of the correlation test in Table 3 validate our initial finding and show a significant positive correlation (p-value <0.001) between sentimentlib(OCR) and score(OCR) for all four sentiment analysis libraries, with r=0.634 for TextBlob, r=0.680 for Vader,r=0.771 for Flair, and r=0.802 for Transformers. Furthermore, it is worth noting the superiority of machine learning-based libraries over lexicon-based ones in terms of accuracy.
In conclusion, based on the results obtained, we can answer research question 1 and assert that the sentiment expressed by the client in their OCR explains their overall satisfaction.

4.2 RESEARCH QUESTION 2

Research question 2 focuses on identifying the most important factors mentioned by clients of Algerian hotels in their OCRs. To address this question, we performed a lexical analysis of the text from all collected OCRs to identify the key themes around which client concerns are centered. For this purpose, we utilized the Python natural language processing library NLTK (Bird et al. 2009) to extract a list of all words and their frequency of occurrence from the text of the collected OCRs.
It is important to note that our program was configured to retain only common nouns. Specifically, we excluded proper nouns (e.g., „Sonia," „Hilton," „Algiers"), as well as verbs, adjectives, adverbs, and stop words such as „a," „the," „is," „then," and „of." Additionally, all plural common nouns were converted to their singular forms. The resulting list contains 7,892 unique words, ranging from highly frequent terms like „hotel" (appearing in 8,817 OCRs) and „room" (appearing in 7,161 OCRs) to words that occur only once, such as „clandestine" and „millimeter." Figure 2 illustrates the distribution of these words in descending order of frequency.

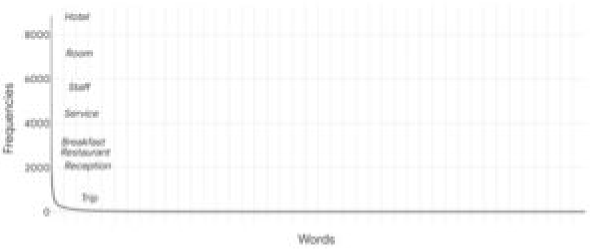

Figure 2: Word frequency distribution
Source: Authors

Table 4: Most frequent words in the collected OCR
Source: Authors

To identify the most important factors mentioned by clients in their OCRs, we used a hierarchical clustering algorithm to further reduce the number of words by grouping those with very similar semantic fields. To measure the similarity between the semantic fields of two words, we used the Topic modeling Python library Gensim (Rehurek and Sojka 2011).
The hierarchical clustering algorithm is an unsupervised classification algorithm (meaning that the number of groups to be formed is not known in advance) that is based on the notion of similarity and proceeds incrementally at each iteration by either grouping the two most semantically similar words together and/or including a word into one of the already formed groups that is closest to it semantically. The result obtained is called a dendrogram. It is a hierarchical structure where each level provides a candidate classification.
As we move up each level, the number of groups decreases and the number of words per group increases. It should be noted that the choice of the level to retain for classification can be guided by identifying large increases in the fusion level, as such jumps indicate that dissimilar clusters are being merged and that the preceding level represents a meaningful partition of the data (Everitt et al. 2011). We chose the classification illustrated in Figure 3 and manually assigned an appropriate theme to each formed word group, namely:
• Food: the quality and price of the dining, the free breakfast.
• Staff: the helpfulness and friendliness of the employees and managers.
• Room: the cleanliness, layout, amenities, and quality of the room.
• Location: the area where the hotel is located and its proximity to points of interest
• Family: the hotel’s suitability for a family setting.
• Stay: the overall stay experience.
• Work: the suitability of the hotel for a family setting
• Service: the overall quality of the service.
We can then answer research question 2 and conclude, based on the previous results, that these eight themes constitute the most important factors mentioned by clients of Algerian hotels in their OCRs.

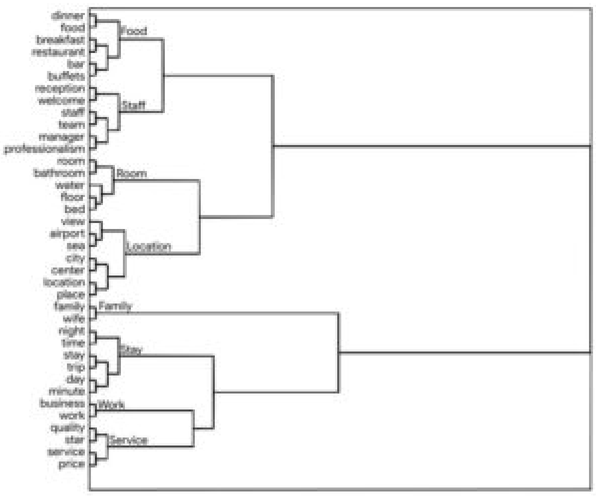

Figure 3: Dendrogram obtained from the hierarchical clustering algorithm

Source: Authors

4.3 RESEARCH QUESTION 3

After identifying the most concerning factors for clients of Algerian hotels, this section focuses on examining whether there is a relationship between these factors and the score the client assigns to the hotel in their OCR. To answer this question, we first used the results from research question 2 to associate to each OCR the list of relevant factors based on the words it contains. Then, we generated a heatmap to visualize the distribution of these factors across different scores. The heatmap illustrates the frequency of each factor in OCRs corresponding to specific scores, as shown in Figure 4.
Given the color scale used (ranging from dark red to dark green) to emphasize differences in factor frequencies, only the cells with colors ranging from light green to dark green are of interest to us. The green color indicates the dominance of a specific factor compared to the others. It is important to note that the heatmap should be read vertically, column by column, to identify the most dominant factors for each score level. However, since the aim of this paper is to identify the factors of satisfaction and dissatisfaction among Algerian hotel customers, the left side of the heatmap highlights the dominant factors contributing to client dissatisfaction, namely: the room, the stay, and the service. On the other hand, the right side of the heatmap reveals the dominant factors associated with client satisfaction, namely: the room, the staff, the food, and the location.
Although it is evident from the previous paragraph that there is a relationship between the factors mentioned by clients in their feedback and the scores they assign to the hotel, it is necessary to confirm this using a statistical test. In our context, given the categorical nature of the two variables being analyzed, a contingency table analysis accompanied by a chi-square test of independence is the most appropriate approach, as illustrated in Table 5.

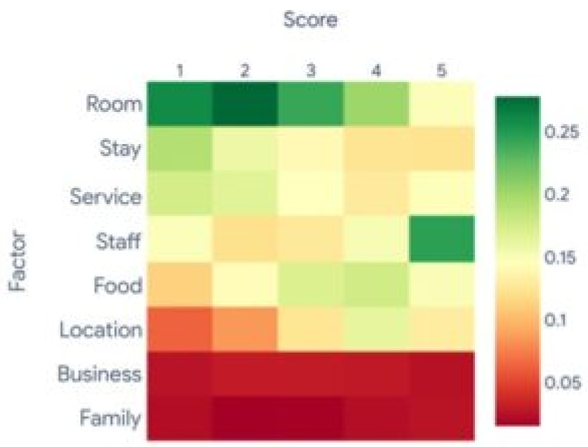

Figure 4: Heatmap of the distribution of factors on the score
Source: Authors

Table 5: Contingency table between factors and scores
Source: Authors

Although the effect size was relatively small (Cramer's V=0.10, 95% CI , referring to the chi-square distribution table, the results suggest the presence of a weak but meaningful dependence as we can observe that the chi-square value obtained in our analysis exceeds the critical chi-square value (p-value <0.001). As a result, we reject the null hypothesis which states that the two variables are independent and conclude that there is a statistically significant relationship between the factors mentioned by clients in their feedback and the score they assign to the hotel.

5 DISCUSSION

The results of this study demonstrate that when clients write online reviews, they implicitly express sentiments, either positive or negative, toward the hotel. These sentiments are not only evident in the tone and language used but are also strongly correlated with the final score that the client assigns to the hotel. These findings are consistent with (Geetha et al. 2017), who identified a clear alignment between customer ratings and expressed sentiments across both premium and budget hotel categories. This correlation highlights the importance of analyzing both the quantitative scores and the qualitative content of reviews to gain a comprehensive understanding of customer satisfaction.

Furthermore, the findings reveal that there are eight main factors of concern for clients of Algerian hotels: food, staff, room, the hotel's location, its suitability for a family setting, the overall stay experience, its suitability for work purposes, and, finally, the quality of service. These factors collectively shape the client's perception of their stay, but their relative importance varies. Some of these factors are more important than others in explaining satisfaction and dissatisfaction. Indeed, we found that the main factors contributing to client satisfaction are the room, the staff, the food, and the location. For example, clients often praise spacious and well-maintained rooms, attentive and friendly staff, delicious and varied food options, and convenient locations close to tourist sites or business districts. On the other hand, the main factors contributing to dissatisfaction are the room, the stay experience, and the service. Dissatisfied clients frequently mention issues such as uncomfortable beds, poor cleanliness, unprofessional staff behavior, or a lack of responsiveness to their needs. Interestingly, the „room" factor appears in both categories, suggesting that it plays a dual role in shaping the client's overall experience.

We can observe that, with the exception of the room factor, the dissatisfaction factors are more abstract than the satisfaction factors. Indeed, stay experience and service relate to broad aspects of the client's time at the hotel, making them harder to define or visualize. On the other hand, the reasons for satisfaction are more concrete. Staff, food, and location refer to specific, tangible parts of the customer's experience that are easy to picture. In comparison, the reasons for dissatisfaction are much harder to visualize, as they often reflect a general sense of disappointment rather than a specific issue. These results are consistent with those of (Kim et al. 2016), who found that most satisfiers in the full-service hotel segment were associated with tangible features, while most dissatisfiers tended to be linked to intangible features.

This suggests that, in general, when clients are unhappy with their stay, they tend to express their dissatisfaction using vague or general terms. This could be due to the emotional nature of negative experiences, which often lead to broader, less specific complaints. On the other hand, when clients are satisfied, they often use more detailed and descriptive language to highlight the specific things they enjoyed. This difference in language reflects the way positive experiences are more likely to be associated with specific, memorable details, while negative experiences are often summarized in broader terms.

6 CONCLUSION

6.1 THEORETICAL CONTRIBUTIONS AND IMPLICATIONS

In this paper, we explored how innovative techniques from machine learning, particularly NLP, can be used to analyze customer satisfaction in Algerian hotels. By leveraging advanced algorithms, we were able to extract meaningful insights from unstructured text data, such as online reviews, which traditional methods often struggle to process efficiently. Like other studies, our approach shows that using Big Data is not only a viable alternative to traditional data collection methods but also offers a more scalable and cost-effective solution. From a theoretical perspective, our work contributes to the growing body of research on evaluating customer satisfaction in the hospitality industry. Specifically, it highlights the potential of NLP techniques to uncover hidden patterns in customer feedback, which can lead to more accurate and actionable insights. We hope this approach will provide a strong basis for future studies, encouraging researchers to explore new ways of integrating machine learning into customer experience analysis.

Moreover, the approach we used and the way we combined different NLP libraries to analyze online hotel

reviews can be applied more broadly. For instance, the framework we developed is not limited to the hospitality sector; it can be adapted to other industries where customer feedback plays a critical role, such as retail, healthcare, or even education. It can serve as a foundation for other researchers in the field, offering a step-by-step guide on how to preprocess, analyze, and interpret textual data. Additionally, our method can be adapted to other emerging contexts in developing countries, such as renting houses, apartments, private rooms, or other properties. This flexibility makes it particularly valuable for regions where traditional data collection methods are less feasible due to resource constraints. Furthermore, it can help to evaluate satisfaction factors for different customer segments, such as families, solo travelers, or business professionals, providing tailored insights for each group.
From a managerial perspective, the results of this research offer valuable insights for hotel managers in Algeria about their clients' preferences. For example, by identifying the most frequently mentioned factors in positive and negative reviews, managers can prioritize areas for improvement, such as enhancing the quality of food or training staff to deliver better service. They can also help managers better understand what clients expect, enabling them to design more targeted marketing campaigns and personalized experiences. Moreover, these findings can be useful for policymakers and hotel managers in other developing countries with tourism potential similar to Algeria's. By adopting a data-driven approach, they can make informed decisions about infrastructure development, service standards, and customer engagement strategies. Ultimately, this research not only benefits the hospitality industry but also contributes to the broader goal of promoting sustainable tourism growth in developing regions.

6.2 LIMITATIONS

Algeria is a country where most tourist attractions and hotels are located in the northern region, operating within a cultural and sometimes religious context unique to the country. As a result, the findings of this study should be interpreted with caution, taking into account the specific context of Algeria. Additionally, due to the limited number of online reviews for 1-star and 2-star hotels, our analysis focused solely on 3-, 4- and 5-star hotels. This limitation arises because lower-category hotels are less likely to be reviewed online, either because their clients are less inclined to share feedback or because these establishments are less visible on digital platforms. Consequently, the results of this study may not fully represent the experiences of clients staying in budget accommodations. Therefore, it would not be appropriate to assume that these results apply to other hotel categories or types of accommodations, such as guest houses, hostels or eco-lodges, which may cater to different customer segments with distinct priorities.
It is also important to note that more than half of the online hotel reviews analyzed were written by business travelers, whose needs and expectations differ from those of other customer segments (Zhang et al. 2018; Kim et al. 2020). In the context of Algeria, where tourism is still emerging compared to other destinations, online reviews predominantly reflect the experiences of business travelers, especially in major cities and commercial hubs. This imbalance in the dataset could skew the results, making them less representative of the broader population of hotel guests. Future studies could address this limitation by collecting a more balanced sample of reviews from diverse customer segments. Moreover, reviews may exhibit seasonal or temporal variation, with business travel peaking during weekdays or certain months, while leisure travel may concentrate during holidays and summer periods, further affecting the representativeness of the dataset. Future studies could address this limitation by collecting a more balanced sample of reviews from different customer segments and across various regions and seasons, ensuring a more comprehensive understanding of hotel satisfaction in the Algerian context.
Finally, the textual nature of online reviews and the languages in which they are written present certain limitations. Sentiment analysis remains a complex field, as machines still struggle to fully grasp nuances of natural language, such as irony, humor, and sarcasm. Additionally, reviews in Algerian hotels are often written in multiple languages, including French, Arabic, and English, each with its own linguistic subtleties. This multilingual aspect adds another layer of complexity to the analysis, as sentiment analysis models trained on one language may not perform equally well on others. These challenges highlight the need for continued advancements in NLP to improve the accuracy and reliability of sentiment analysis tools.

LITERATÚRA/LIST OF REFERENCES

Aakash, A. and Gupta Aggarwal, A., 2022. Assessment of hotel performance and guest satisfaction through eWOM: big data for better insights. In: International Journal of Hospitality & Tourism Administration. 2022, 23(2), 317-346. ISSN 1525-6480. Available at: <https://doi.org/10.1080/15256480.2020.1746218>
Akbik, A., Bergmann, T., Blythe, D., Rasul, K., Schweter, S. and Vollgraf, R., 2019. FLAIR: An easy-to-use framework for state-of-the-art NLP. In: Ammar, W., Louis, A., Mostafazadeh, N. (Eds.), 2019. Proceedings of

the 2019 conference of the North American chapter of the association for computational linguistics (demonstrations). 2019, 54-59. Available at: <https://doi.org/10.18653/v1/N19-4010>
Álvarez-Carmona, M. Á., Aranda, R., Rodríguez-Gonzalez, A. Y. et al. 2022. Natural language processing applied to tourism research: A systematic review and future research directions. In: Journal of king Saud university-computer and information sciences. 2022, 34(10), 10125-10144. ISSN 1319-1578. Available at: <https://doi.org/10.1016/j.jksuci.2022.10.010>
Anandarajan, M., Hill, C. and Nolan, T., 2019. Text preprocessing. In: Sharda, R. (Ed.), 2019. Practical Text Analytics. Advances in Analytics and Data Science, 2, 45-59. Springer, Cham. ISBN 978-3-319-95663-3. Available at: <https://doi.org/10.1007/978-3-319-95663-3_4>
Arindra, M., Li, J., Sengupta, P. and Oztekin, A., 2024. NLP-Driven insights on boutique hotel satisfaction. In: Journal of Computer Information Systems. 2024, 1-16. ISSN 0887-4417. Available at: <https://doi.org/10.1080/08874417.2024.2362824>
Bird, S., Klein, E. and Loper, E., 2009. Natural language processing with Python: analyzing text with the natural language toolkit. O'Reilly Media, Inc., 2009. ISBN 978-0-596-51649-9.
Bulkrock, O. and Alsharman, N., 2024. A natural language processing approach for sentiment analysis of hotel reviews. In: International Journal of Advances in Soft Computing & Its Applications. 2024, 16(3). ISSN 2074-8523. Available at: <https://doi.org/10.15849/IJASCA.241130.02>
Carvalho, F., Ramos, R. F. and Fortes, N., 2024. Customer satisfaction in mountain hotels within UNESCO Global Geoparks: an empirical study based on sentiment analysis of online consumer reviews. In: Tourism & Management Studies. 2024, 20(1), 35-47. ISSN 2182-8458. Available at: <https://doi.org/10.18089/tms.20240103>
Chowdhury, G. G., 2003. Natural language processing. In: Annual Review of Information Science and Technology. 2003, 37, 51-89. ISSN 0066-4200. Available at: <https://doi.org/10.1002/aris.1440370103>
Cheng, M. and Jin, X., 2019. What do Airbnb users care about? An analysis of online review comments. In: International Journal of Hospitality Management. 2019, 76, 58-70. ISSN 0278-4319. Available at: <http://dx.doi.org/10.1016/j.ijhm.2018.04.004>
Cherdouh, S., Kherri, A., Abbaci, A. and Kebir, S., 2022. Using sentiment analysis of online hotel reviews to explore the effect of information and communication technologies on hotel guest satisfaction. In: Journal of Tourismology. 2022, 8(1), 49-67. ISSN 2459-1939. Available at: <https://doi.org/10.26650/jot.2022.8.1.1038566>
Everitt, B., Sabine, L., Morven, L. and Daniel, S., 2011. Cluster analysis. Wiley Series in Probability and Statistics, 2011. ISBN 978-0-470-74991-3.
Fernández, M. O., Martínez-Torres, M. R. and Marín, S. L., 2016. Harvesting big data in social science: A methodological approach for collecting online user-generated content. In: Computer Standards & Interfaces. 2016, 46, 79-87. ISSN 0920-5489. Available at: <https://doi.org/10.1016/j.csi.2016.02.003>
Firoozeh, N., Nazarenko, A., Alizon, F. and Daillel, B., 2020. Keyword extraction: Issues and methods. In: Natural Language Engineering. 2020, 26(3), 259-291. ISSN 1351-3249. Available at: <https://doi.org/10.1017/S1351324919000457>
Geetha, M., Singha, P. and Sinha, S. R., 2017. Relationship between customer sentiment and online customer ratings for hotels – an empirical analysis. In: Tourism Management. 2017, 61, 43-54. ISSN 0261-5177. Available at: <https://doi.org/10.1016/j.tourman.2016.12.022>
Gunasekar, S. and Sudhakar, S., 2019. Does hotel attributes impact customer satisfaction: A sentiment analysis of online reviews. In: Journal of Global Scholars of Marketing Science. 2019, 29(2), 180-195. ISSN 2163-9159. Available at: <https://doi.org/10.1080/21639159.2019.1577155>
He, W., Tian, X., Tao, R., Zhang, W., Yan, G. and Akula, V., 2017. Application of social media analytics: a case of analyzing online hotel reviews. In: Online Information Review. 2017, 41(7), 921-935. ISSN 1468-4527. Available at: <https://doi.org/10.1108/OIR-07-2016-0201>
Hirschberg, J. and Manning, C. D., 2015. Advances in natural language processing. In: Science. 2015, 349(6245), 261-266. ISSN 0036-8075. Available at: <https://doi.org/10.1126/science.aaa8685>
HuggingFace, 2022. Translation models-hugging face huggingface.co. 2022. . . Available at: <https://huggingface.co/models?pipe line_tag=translation&sort=downloads>
Hutto, C. and Gilbert, E., 2014. VADER: A parsimonious rule-based model for sentiment analysis of social media text. In: Proceedings of the International AAAI Conference on Web and Social Media, 8(1), 216-225. ISSN 2334-0770. Available at: <https://doi.org/10.1609/icwsm.v8i1.14550>
Kang, Y., Cai, Z., Tan, C. W., Huang, Q. and Liu, H., 2020. Natural language processing (NLP) in management research: A literature review. In:Journal of Management Analytics. 2020, 7(2), 139-172. ISSN 2327-0012. Available at: <https://doi.org/10.1080/23270012.2020.1756939>

Khurana, D., Koli, A., Khatter, K. and Singh, S., 2023. Natural language processing: state of the art, current trends and challenges. In: Multimed Tools Applications. 2023, 82, 3713-3744. ISSN 1380-7501. Available at: <https://doi.org/10.1007/s11042-022-13428-4>
Kim, B., Kim, S. and Heo, C. Y., 2016. Analysis of satisfiers and dissatisfiers in online hotel reviews on social media. In: International Journal of Contemporary Hospitality Management. 2016, 28(9), 1915-1936. ISSN 0959-6119. Available at: <https://doi.org/10.1108/IJCHM-04-2015-0177>
Kim, D., Hong, S., Park, B. J. and Kim, I., 2020. Understanding heterogeneous preferences of hotel choice attributes: Do customer segments matter? In: Journal of Hospitality and Tourism Management. 2020, 45, 330-337. ISSN 1447-6770. Available at: <https://doi.org/10.1016/j.jhtm.2020.08.014>
Krumm, J., Davies, N. and Narayanaswami, C., 2008. User-generated content. In: IEEE Pervasive Computing. 2008, 7(4), 10-11. ISSN 1536-1268. Available at: <https://doi.org/10.1109/MPRV.2008.85>
Lee, M., Lee, S. A. and Koh, Y., 2019. Multisensory experience for enhancing hotel guest experience: Empirical evidence from big data analytics. In: International Journal of Contemporary Hospitality Management. 2019, 31(11), 4313-4337. ISSN 0959-6119. Available at: <https://doi.org/10.1108/IJCHM-03-2018-0263>
Li, H., Liu, Y., Tan, C. W. and Hu, F., 2020. Comprehending customer satisfaction with hotels: Data analysis of consumer-generated reviews. In: International Journal of Contemporary Hospitality Management. 2020, 32(5), 1713-1735. ISSN 0959-6119. Available at: <https://doi.org/10.1108/IJCHM-06-2019-0581>
Li, J., Xu, L., Tang, L., Wang, S. and Li, L., 2018. Big data in tourism research: A literature review. In: Tourism management. 2018, 68, 301-323. ISSN 0261-5177. Available at: <https://doi.org/10.1016/j.tourman.2018.03.009>
Liu, S., Law, R., Rong, J., Li, G. and Hall, J., 2013. Analyzing changes in hotel customers' expectations by trip mode. In: International Journal of Hospitality Management. 2013, 34, 359-371. ISSN 0278-4319. Available at: <https://doi.org/10.1016/j.ijhm.2012.11.011>
Loria, S., 2020. Textblob documentation. 2020. . . Available at: <https://buildmedia.readthedocs.org/media/pdf/textblob/latest/textblob.pdf>
Luo, J., Huang, S. and Wang, R., 2020. A fine-grained sentiment analysis of online guest reviews of economy hotels in China. In: Journal of Hospitality Marketing & Management. 2020, 30(1), 71-95. ISSN 1936-8623. Available at: <https://doi.org/10.1080/19368623.2020.1772163>
Luo, J. M., Vu, H. Q., Li, G. and Law, R., 2021. Understanding service attributes of robot hotels: A sentiment analysis of customer online reviews. In: International Journal of Hospitality Management. 2021, 98, 103032. ISSN 0278-4319. Available at: <https://doi.org/10.1016/j.ijhm.2021.103032>
Medhat, W., Hassan, A. and Korashy, H., 2014. Sentiment analysis algorithms and applications: A survey. In: Ain Shams engineering journal. 2014, 5(4), 1093-1113. ISSN 2090-4479. Available at: <https://doi.org/10.1016/j.asej.2014.04.011>
Nilashi, M., Ibrahim, O., Yadegaridehkordi, E., Samad, S., Akbari, E. and Alizadeh, A., 2018. Travelers decision making using online review in social network sites: A case on TripAdvisor. In: Journal of computational science. 2018, 28, 168-179. ISSN 1877-7503. Available at: <https://doi.org/10.1016/j.jocs.2018.09.006>
Özen, İ. A. and Özgül Katlav, E., 2023. Aspect-based sentiment analysis on online customer reviews: a case study of technology-supported hotels. In: Journal of Hospitality and Tourism Technology. 14(2), 102-120. ISSN 1757-9880. Available at: <https://doi.org/10.1108/JHTT-12-2020-0319>
Padma, P. and Ahn, J., 2020. Guest satisfaction & dissatisfaction in luxury hotels: An application of big data. In: International Journal of Hospitality Management. 2020, 84, 102318. ISSN 0278-4319. Available at: <https://doi.org/10.1016/j.ijhm.2019.102318>
Park, E., Kang, J., Choi, D. and Han, J., 2018. Understanding customers' hotel revisiting behaviour: a sentiment analysis of online feedback reviews. In: Current Issues in Tourism. 2018, 23(5), 605-611. ISSN 1368-3500. Available at: <https://doi.org/10.1080/13683500.2018.1549025>
Raffel, C., Shazeer, N., Roberts, A. et al., 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. In: Journal of machine learning research. 2020, 21(140), 1-67. ISSN 1532-4435. Available at: <https://doi.org/10.48550/arXiv.1910.10683>
Rehurek, R. and Sojka, P., 2011. Gensim-python framework for vector space modelling. NLP Centre, Faculty of Informatics, Masaryk University, Brno, Czech Republic, 3(2).
Roy, G., 2023. Travelers' online review on hotel performance. Analyzing facts with the theory of lodging and sentiment analysis. In: International Journal of Hospitality Management. 2023, 111, 103459. ISSN 0278-4319. Available at: <https://doi.org/10.1016/j.ijhm.2023.103459>
Sangkaew, N. and Zhu, H., 2020. Understanding tourists' experiences at local markets in Phuket: An analysis of TripAdvisor reviews. In: Journal of Quality Assurance in Hospitality & Tourism. 2020, 23(1), 89-114. ISSN 1528-008X. Available at: <https://doi.org/10.1080/1528008X.2020.1848747>

Saraswati, N. W. S., Putra, I. K. G. D., Sudarma, M. et al., 2024. Revealing the potential of hotel improvements in Bali based on sentiment analysis and tourist characteristics. In: 2024 11th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI) (pp. 722-728). IEEE. Available at: <https://doi.org/10.1109/EECSI63442.2024.10776092>
Sparks, B. A. and Browning, V., 2011. The impact of online reviews on hotel booking intentions and perception of trust. In: Tourism Management. 2011, 32(6), 1310-1323. ISSN 0261-5177. Available at: <https://doi.org/10.1016/j.tourman.2010.12.011>
Statista, 2022. Total number of user reviews and opinions on Tripadvisor worldwide from 2014 to 2021. 2022. . . Available at: <https://www.statista.com/statistics/ 684862/tripadvisor-number-of reviews/>
Tripadvisor, 2018. . . Available at: <https://www.tripadvisor.com/ShowUserReviews-g1071600-d12063561-r630501781-Sheraton_Annaba_Hotel-Annaba_Annaba_Province.html>
Vayansky, I. and Kumar, S. A., 2020. A review of topic modeling methods. In: Information Systems. 2020, 94, 101582. ISSN 0306-4379. Available at: <https://doi.org/10.1016/j.is.2020.101582>
Vermeulen, I. E. and Seegers, D., 2009. Tried and tested: The impact of online hotel reviews on consumer consideration. In: Tourism management. 2009, 30(1), 123-127. ISSN 0261-5177. Available at: <https://doi.org/10.1016/j.tourman.2008.04.008>
Wang, H., Wu, H., He, Z., Huang, L. and Church, K. W., 2022. Progress in machine translation. In: Engineering. 2022, 18, 143-153. ISSN 2096-0026. Available at: <https://doi.org/10.1016/j.eng.2021.03.023>
Wolf, T., Debut, L., Sanh, V. et al. 2020. Transformers: State-of-the-art natural language processing. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations 38-45. Available at: <https://doi.org/10.18653/v1/2020.emnlp-demos.6>
Xiang, Z., Schwartz, Z., Gerdes, J. H. and Uysal, M., 2015. What can big data and text analytics tell us about hotel guest experience and satisfaction. In: International Journal of Hospitality Management. 2015, 44, 120-130. ISSN 0278-4319. Available at: <http://dx.doi.org/10.1016/j.ijhm.2014.10.013>
Xin, L., Qiao, G., Shao, Z., Jiang, T., Wen, C., Zhong, Y. and Li, Z., 2023. Understanding continuous sharing behavior of online travel community users: a case of TripAdvisor. In: Journal of Tourism and Cultural Change. 2023, 21(3), 328-343. ISSN 1476-6825. Available at: <https://doi.org/10.1080/14766825.2023.2170239>
Xu, X. and Li, Y., 2016. The antecedents of customer satisfaction and dissatisfaction toward various types of hotels: A text mining approach. In: International journal of hospitality management. 2016, 55, 57-69. ISSN 02784319. Available at: <https://doi.org/10.1016/j.ijhm.2016.03.003>
Yassin, C. A., 2022. Only memories are captured, and only footprints are left. understanding the perception of eco-friendly hotel and tourist buying behavior. In: Marketing Science & Inspirations. 2022, 17(4). ISSN 1338-7944. Available at: <https://doi.org/10.46286/msi.2022.17.4.3>
Zhang, T., Seo, S. and Ahn, J. A., 2018. Why hotel guests go mobile? Examining motives of business and leisure travelers. In: Journal of Hospitality Marketing & Management. 2018, 28(5), 621-644. ISSN 1936-8623. Available at: <https://doi.org/10.1080/19368623.2019.1539936>
Zhao, Y., Xu, X. and Wang, M., 2019. Predicting overall customer satisfaction: Big data evidence from hotel online textual reviews. In: International journal of hospitality management. 2019, 76, 111-121. ISSN 0278-4319. Available at: <https://doi.org/10.1016/j.ijhm.2018.03.017>
Zhu, L., Lin, Y. and Cheng, M., 2020. Sentiment and guest satisfaction with peer-to-peer accommodation: when are online ratings more trustworthy? In: International Journal of Hospitality Management. 2020, 86, 102369. ISSN 02784319. Available at: <https://doi.org/10.1016/j.ijhm.2019.102369>
Zhu, J., Sun, H. and Kong, B., 2025. Improving multilingual English translation performance through T5 and MAML integration. In: Systems and Soft Computing. 2025, 200394. ISSN 2772-9419. Available at: <https://doi.org/10.1016/j.sasc.2025.200394>

KĽÚČOVÉ SLOVÁ/KEY WORDS

online customer review, natural language processing, satisfaction factors, dissatisfaction factors, hotels
online recenzie zákazníkov, spracovanie prirodzeného jazyka, faktory spokojnosti, faktory nespokojnosti, hotely

JEL KLASIFIKÁCIA/JEL CLASSIFICATION

L83, M31

RÉSUMÉ

Identifikácia faktorov spokojnosti a nespokojnosti hotelových zákazníkov pomocou techník spracovania prirodzeného jazyka

Tento článok navrhuje nový prístup k identifikácii faktorov, ktoré ovplyvňujú spokojnosť a nespokojnosť alžírskych hotelových zákazníkov prostredníctvom analýzy online recenzií zákazníkov. Na rozdiel od tradičných kvantitatívnych metód, ako sú dotazníky, táto štúdia využíva pokročilé techniky spracovania prirodzeného jazyka, aby odhalila kľúčové poznatky o skúsenostiach zákazníkov. Štúdia využíva techniky spracovania prirodzeného jazyka na extrakciu a analýzu údajov z online recenzií zákazníkov. Cieľom tejto metódy je identifikovať významné obavy a faktory spokojnosti, ktoré spomínajú alžírski hoteloví zákazníci, a ponúknuť inovatívnu alternatívu k tradičným prístupom. Analýza odhalila, že faktory spokojnosti sú špecifické, hmatateľné aspekty skúseností zákazníkov, ktoré sa dajú ľahko konceptualizovať. Naopak, faktory nespokojnosti sú abstraktnejšie a ťažšie definovateľné, čo sťažuje ich pochopenie. Článok predstavuje inovatívny prístup, ktorý využíva spracovanie prirodzeného jazyka na analýzu recenzií zákazníkov a ponúka nový pohľad na pochopenie spokojnosti a nespokojnosti zákazníkov. Táto metodika poskytuje cenné informácie o skúsenostiach zákazníkov a zdôrazňuje rozdiely v tom, ako zákazníci vnímajú a vyjadrujú spokojnosť a nespokojnosť.

RECENZOVANÉ/REVIEWED

30. July 2025 / 11. August 2025