RIAT
Revista Interamericana de Ambiente y Turismo
Interamerican Journal of Environment and Tourism
Print version ISSN - 0717 - 6651, Online version ISSN - 0718 - 235X
http://dx.doi.org/10.4067/S0718-235X2024000200126
http://riat.utalca.cl
www.fen.utalca.cl

Research paper

Beyond the Reviews: Unveiling the Hotel Experience Through Text Mining

Más allá de las reseñas: Desvelando la experiencia hotelera con minería de texto

Sebastián Araya-Pizarro[1*] and Nando Verelst[2]

[1] Doctor in Education, Faculty of Social, Business and Legal Sciences, Universidad de La Serena, La Serena, Chile. https://orcid.org/0000-0002-5857-8441
[2] Master's in Strategic Leadership, Management, and Communication, Faculty of Administration and Economy, Universidad Diego Portales, Santiago, Chile. https://orcid.org/0000-0002-0957-767X
[*] Corresponding author: saraya@userena.cl

Received: 2024-08-04 Accepted for publication: 2024-09-08 Published: 2024-12-30

ABSTRACT

In hotel management, it is crucial to understand the factors that are essential to customer satisfaction and loyalty. This study aims to determine the attributes influencing tourist experiences in hotels in Santiago, Chile. The methodology consists of text mining, sentiment analysis, and binary logistic regressions were employed to analyze 1,467 online reviews from 10 hotels extracted from the Booking.com platform. Results highlight those comments on infrastructure, location, food, and staff, with similar sentiments. Additionally, it was found that the likelihood of a positive experience significantly increases when location, amenities, and price-quality ratio are adequate. It is concluded that data mining and sentiment analysis are practical tools to comprehend key attributes of the hotel experience, whose enhanced management could substantially enhance guest satisfaction and loyalty.

Keywords: *Data mining, hospitality, sentiment analysis, touristic experiences, Chile.*

RESUMEN

En la gestión hotelera, es crucial comprender los factores que resultan esenciales para la satisfacción y fidelización de los clientes. Este estudio tiene como objetivo determinar los atributos que influyen en las experiencias de los turistas que hospedan en hoteles de Santiago, Chile. La metodología consiste en técnicas de minería de textos, análisis de sentimientos y regresiones logísticas binarias para analizar 1,467 reseñas en línea de 10 hoteles extraídas desde la plataforma Booking.com. Los resultados destacan que los comentarios se concentran en la infraestructura, ubicación, comida y personal, con sentimientos similares. Además, se halló que la probabilidad de una experiencia positiva aumenta significativamente cuando la localización, las amenidades y la relación precio-calidad son adecuadas. Se concluye que la minería de datos y el análisis de sentimientos son herramientas efectivas para comprender los atributos clave de la experiencia hotelera, cuya gestión mejorada podría aumentar sustancialmente la satisfacción y la lealtad de los turistas.

Palabras clave: Minería de datos, hotelería, análisis de sentimiento, experiencia turística, Chile.

INTRODUCTION

In the realm of hotel management, it is crucial to acknowledge the unique characteristics of the service and guest experience offered by the establishment, as they play a pivotal role in establishing a competitive advantage and fostering customer loyalty. Within a constantly evolving global market, the ability to differentiate oneself becomes indispensable in order to distinguish amidst the diverse range of lodging options available. Hence, comprehending the evolving demands and expectations of clientele, and continuously adapting to emerging market trends, are critical factors in sustaining relevance and competitiveness within the hotel industry (Millán-García & Gómez-Díaz, 2018).

In this regard, when designing lodging services, it is essential to consider a multitude of factors that shape the visitor's experience throughout their stay. These factors encompass aspects such as price, infrastructure, location, staff, food, hygiene, amenities, and other crucial elements (Bi et al., 2020; Jawabreh et al., 2022; Mondaca-Marino et al., 2019; Nilashi et al., 2023; Stefko et al., 2020; Verma & Thakur, 2022; Xue & Chen, 2021). Overall, previous research has consistently shown that possessing positive and relevant attributes directly influences user satisfaction (Bi et al., 2020; Manolitzas et al., 2022; Spoerr, 2021), consequently fostering customer loyalty (Darmawan et al., 2021; Paulose & Shakeel, 2022).

Moreover, in the current landscape of continual technological advancement, the increasing use and analysis of big data present themselves as invaluable assets for hotel management (Carneiro et al., 2023). This field provides robust tools for scrutinizing vast datasets (both organized and disorganized) and extracting significant insights to gain a deeper understanding of tourists' preferences and requirements. Through sophisticated data analysis methodologies like text mining and sentiment analysis, hotels can discern patterns and trends within guests' feedback and reviews (Chatterjee, 2020; Khamphakdee & Seresangtakul, 2021; Zarezadeh et al., 2022). This empowers hotel establishments not only to pinpoint areas for enhancement in their services and operations but also to identify critical attributes that influence customer satisfaction and booking decisions. By adeptly harnessing data mining techniques, hotels can customize their marketing approaches, elevate the personalization of customer experiences, and carve out a distinct niche in the market.

In Chile, the hotel industry stands out as a dynamic, competitive, and essential sector within the lodging market (Castro et al., 2023). Santiago, the capital of the country and the epicenter of the domestic tourism industry, leads the nationwide hotel supply. Renowned for its urban and business tourism, Santiago maintains activity year-round and showcases a counter-cyclical dynamic comparison to sun and beach destinations. Furthermore, this metropolis is acknowledged as an emerging tourist hub, emblematic of others in Latin America (Mondaca-Marino et al., 2019).

In this context, the objective of the present article is to ascertain the factors that impact the experiences of tourists lodging in hotels in Santiago, Chile. To accomplish this objective, a blend of advanced methodologies including text mining, sentiment analysis, and binary logistic regressions is utilized to analyze feedback from tourists registered on the accommodation platform Booking.com. The choice of this platform is grounded in its significance and extensive adoption as an evaluation source (Mondaca-Marino et al., 2019; Moro et al., 2018; Özen & Özgül, 2023).

It's important to highlight that there is a scarcity of studies analyzing customer perceptions of the lodging experience in Chilean hotels using extensive datasets, despite the availability of freely accessible information sources like Booking.com. Additionally, recent scientific research on customer satisfaction among hotel guests in the country is limited. Therefore, the objective is to provide valuable insights from open platforms regarding how users perceive various hotel attributes. This, in turn, contributes to improving the quality of lodging services and delivering more satisfying experiences for visitors. Ultimately, this feedback aids in elevating quality standards across the entire hotel industry, fostering continuous improvement within the territorial tourism ecosystem.

The article is structured into four main sections, comprising this introduction, the study methodology, the results, and discussions, culminating in conclusions.

THEORETICAL BACKGROUND

The choice of lodging within the tourism sector entails considering numerous critical attributes that impact visitors' experiences throughout their stay, influencing their satisfaction and loyalty (Yoon & Uysal, 2005). In this regard, models of tourist behavior are grounded in the concept of consumer limited rationality (Conlisk, 1996), which posits

that each person, in their role as a decision-maker, is conditioned by their previous experience, cognitive abilities, time constraints, and other contextual factors. Rational choice theory assumes that consumers are homo economicus, meaning rational decision-makers seeking to maximize their utility or satisfaction (Jung & Kim, 2016). Thus, the consumer's decision regarding lodging is not only based on objective criteria such as price or location but also on a series of subjective factors that influence their perception of service quality and overall experience.

Therefore, understanding the market targeted by each company is crucial, especially in the tourism and hospitality industry, where the Lancasterian nature of tourist products is recognized. According to Lancaster's Theory, goods are composed of a series of attributes that are evaluated individually and contribute to the overall assessment of the customer experience. Thus, people's decisions focus more on product characteristics than on the products themselves (Boz et al., 2020). This complexity increases the difficulty in choosing lodging and underscores the importance of understanding the factors that influence this decision, especially considering that each tourist has unique expectations influenced by diverse motivations (Sánchez-Oro et al., 2021).

In the specialized literature, various attributes of hotels are highlighted, subject to extensive debate and analysis. Among these are food quality (Nilashi et al., 2023; Tager et al., 2021), hotel and room cleanliness (Stefko et al., 2020), infrastructure (Xue & Chen, 2021), location (Bi et al., 2020; Latinopoulos, 2020), pricing (Mondaca-Marino et al., 2019), staff quality of service (Jawabreh et al., 2022; Xin & Choi, 2020), administrative processes and services (Nunkoo et al., 2020; Wu et al., 2021), amenities (Shah & Bhatt, 2020; Verma & Thakur, 2022), and equipment (Stefko et al., 2020; Xue & Chen, 2021). When these aspects are positively evaluated, they become essential strategic elements in ensuring visitor satisfaction and loyalty.

On the other hand, data mining refers to a set of techniques and tools that allow the extraction of large volumes of data available in repositories or websites. Its applicability is wide-ranging, as it proves useful in various fields such as finance, retail, medicine, computer science, among others. Its primary objective is to collect data that initially do not exhibit a defined pattern and, through analysis and interpretation, utilize them for decision-making (Gupta & Chandra, 2020). In the literature review on Artificial Intelligence and Big Data in tourism conducted by Samara et al. (2020), it is evident that the use of Big Data Analysis can create value and advantages, such as increased efficiency, productivity, and profitability, in addition to offering a highly enriched and personalized experience for travelers. Particularly, text mining, focused on knowledge discovery from text-based databases, has been applied since the late 1990s, although its use in studies related to hospitality and tourism began in the mid-2000s (Huang et al., 2022). In hotel management, various applications of data mining techniques have been carried out. For instance, a study conducted by Kampalpour et al. (2017) explored the potential use of Data Mining and Web Mining techniques in the tourism industry to uncover underlying knowledge in hotel visitor information. More recently, Zarezadeh et al. (2022) determined that variables utilized by data mining techniques related to online reviews often cluster into two main aspects: guest-staff interaction and physical aspects of the hotel.

Additionally, sentiment analysis stemming from data mining of online comments is highlighted in the literature (Chatterjee, 2020; Khamphakdee & Seresangtakul, 2021). This technique involves assigning values based on emotions, feelings, personality, and other aspects to specific text (Ray et al., 2021), proving useful for categorizing comments or evaluations about a particular good or service and drawing conclusions from positively (positive sentiment) and negatively (negative sentiment) evaluated aspects. Such analysis is of paramount importance for businesses as it enables a summarized understanding of areas for improvement and management, thereby facilitating informed and strategic decision-making (Wankhade et al., 2022).

Building upon the preceding information, the hypothesis posits that certain attributes (infrastructure, staff, equipment, food, hygiene, location, amenities, price-quality ratio, and administrative processes and services) influence the generation of positive sentiments towards hotels.

METHODOLOGY

The research was based on a quantitative approach, employing a non-experimental, descriptive-explanatory design with a cross-sectional scope. The purpose was to determine the attributes influencing tourists' experiences in hotels in Santiago, Chile, through the application of data mining techniques and multivariate statistical methods.
To conduct the text mining process and sentiment analy-

sis, the methodological flow proposed by Wankhade et al. (2022) was followed, consisting of three primary phases: defining the level of analysis, selecting the data collection technique, and defining sentiment concepts.

In the first stage, phrases from user comments, both positive and negative, referring to one or more attributes were analyzed. Subsequently, in the second phase, a web scraping tool was employed to extract comments from the top ten hotels in Santiago, Chile, from the Booking.com platform. The web scraping process was conducted using code developed in the R programming language, gathering a total of 1,467 comments for the period between 2021 and 2023. Finally, in the third phase, relevant sequences of contiguous texts, such as bigrams (two words) and trigrams (three words), reflecting positive or negative sentiments in the comments were identified. These labels were manually applied using the ATLAS.ti software.

Table 1 presents examples of these text sequences used for labeling (categorizing) the units of analysis (phrases). However, it was necessary to conduct a detailed review of each review, especially regarding the metadata associated with each one. For instance, some comments classified as positive in the Booking.com system only mentioned the "location" attribute, so they were adjusted to "good location." Similarly, spelling corrections were made in several messages. Additionally, in the process of coding sentiments, labels associated with gender (male, female, or other), hotel number (1 to 10), establishment attributes, and type of company (friends, family, couple, or other) were assigned.

Furthermore, a binary logistic regression analysis was conducted to statistically evaluate the probability that attributes (price-quality ratio, administrative processes and services, amenities, equipment, food, hygiene, infrastructure, location, and staff) could predict sentiment towards hotels. The response variable was encoded with values of 0 to represent a negative sentiment and 1 for a positive sentiment. In turn, the predictor variables were coded with values of 0 and 1 to represent negative and positive evaluations, respectively, for each attribute. Additionally, the final classification table of the overall percentage of correct predictions, Cox and Snell R-squared coefficients, Nagelkerke R-squared coefficients were evaluated, and the Hosmer-Lemeshow (H-L) test was conducted to demonstrate the quality of the proposed model.

The equation for binary logistic regression can be expressed as follows:

$$ln\left(\frac{p}{1-p}\right) = b_0 + b_1x_1 + b_2x_2 + \cdots + b_kx_k \qquad (1)$$

Thus, it is possible to calculate the probability of the event of interest occurring, considering the values of the predictor variables, using the following formula:

Table 1. Examples of positive and negative bigrams and trigrams

Attribute	Positive comments	Negative comments
Quality-Price ratio	• Good option • Price-quality ratio • Meets expectations	• Didn't like it • Expensive • Nothing extraordinary
Administrative processes	• Quick check-in • Connectivity service • Good timing	• Slow check-in • Delay in check-in • Administrative issues
Amenities	• No towels • No replenishment • Poor Wi-Fi	• Perfect bedding • Appreciated beverages • Complimentary ware
Equipment	• Comfortable bed • Good heating • Good Workplace	• Small bed • Noisy air conditioning • Lack of heating
Food	• Great breakfast • Good bar • Good restaurant	• Disastrous restaurant • Bad food • Breakfast flexibility

Hygiene	• Room hygiene • General cleanliness • Impeccable hotel	• Sewer smell • Unclean bathroom • Musty smell
Infrastructure	• Good facilities • Spacious areas • Spacious room	• Noise from outside • Few parking spaces • Hotel needs maintenance
Location	• Near subway stations • Excellent location • Excellent views	• Unsafe neighborhood • Non-central location • Poor exterior view
Staff service	• Attentive staff • Good service • Friendly staff	• No one answered • No assistance provided • Lack of friendliness

Source: Author's own survey data.

$$\hat{p} = \left(\frac{e^{\hat{b}_0+\hat{b}_1x_1+\hat{b}_2x_2+\cdots+\hat{b}_kx_k}}{1-e^{\hat{b}_0+\hat{b}_1x_1+\hat{b}_2x_2+\cdots+\hat{b}_kx_k}} \right) \quad (2)$$

Where p represents the probability of experiencing a positive sentiment towards hotels, while the predictor variables refer to the nine evaluated attributes. A total of eleven models were created: one general model for the entire sample dataset, and the others segmented by gender (male, female, and other), type of companion (friends, family, and couple), and hotel star rating (two, three, four, and five). It is worth mentioning that all calculations were performed using IBM SPSS Statistics 24.

RESULTS

Sample characterization

The sample comprised 1,467 comments in total, with 58% expressing positivity and 42% negativity. Most comments were derived from hotels rated four or five stars (67%), while the minority stemmed from two-star hotels (11%) (Table 2).

Table 2. Sample characteristics

Variable	Segment	Frequency	Percentage
Type of comment	Positive	846	57.7
	Negative	621	42.3
Hotel star rating	2 stars	155	10.6
	3 stars	330	22.5
	4 stars	549	37.4
	5 stars	433	29.5
Guest gender	Male	696	47.4
	Female	600	40.9
	Other	171	11.7
Travel companions	Friends	347	23.7
	Family	360	24.5
	Couple	359	24.5
	Other	401	27.3

Source: Author's own survey data.

In terms of the guests' profile, it was noted that 41% of the reviews were from women, 47% from men, and 12% were unspecified gender. Additionally, 24% of tourists mentioned traveling with friends, 25% with family members, and another 25% as couples, while 27% indicated traveling with various types of companions.

Regarding the studied hotels, the majority (4) have four stars, followed by three five-star hotels, two three-star hotels, and one two-star hotel. The average overall score is 8.3 (on a scale of 1 to 10), with a standard deviation of 0.53, indicating some consistency in the good ratings given by guests (Table 3).

Table 3. Hotel details and statistics

Hotel	Stars	Score	Hotel	Stars	Score
1	3	7.2	6	4	8.7
2	2	8.7	7	4	8.2
3	3	7.9	8	4	8.4
4	4	9.1	9	5	8.9
5	5	8.2	10	5	8.1

Mean Score	8.30
Confidence Interval for the mean score (95%)	[8.27 - 8.33]
Standard Deviation of the score	0.53

Source: Author's own survey data.

Analysis of comments

Most of the comments focus on infrastructure (429), location (375), food (344), and staff service (306), as shown in Table 4. In contrast, a smaller number of reviews mention hygiene (174), price-quality ratio (159), amenities (105), and other aspects (21) (Table 4).

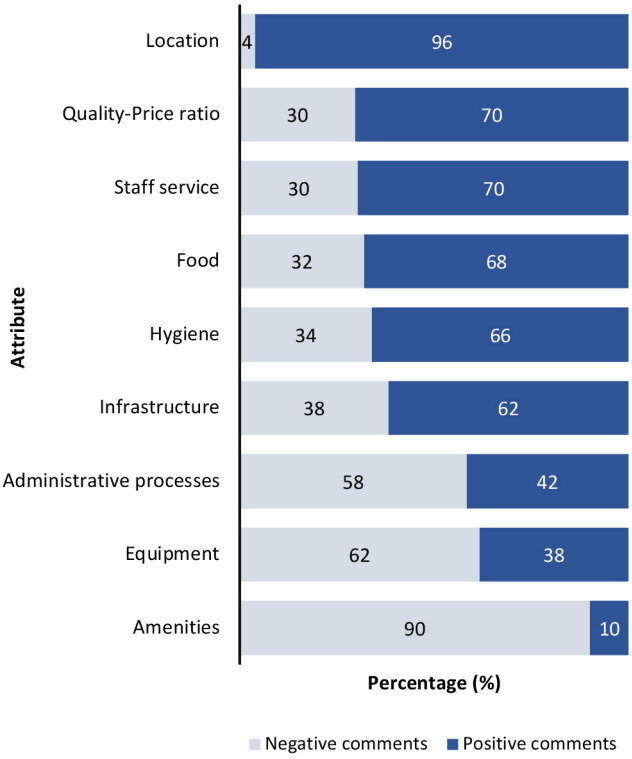
Table 4. Quantity and type of comments by attribute

Attribute	Total	Positive		Negative	
		Case	%	Case	%
Quality-Price ratio	159	112	70.4	47	29.6
Administrative processes	206	86	41.7	120	58.3
Amenities	105	10	9.5	95	90.5
Equipment	177	68	38.4	109	61.6
Food	344	234	68.0	110	32.0
Hygiene	174	114	65.5	60	34.5
Infrastructure	429	265	61.8	164	38.2
Location	375	362	96.5	13	3.5
Staff service	306	213	69.6	93	30.4
Others	21	5	23.8	16	76.2

Note: Percentage (%) calculated based on the total number of comments per attribute.

In general, location is one of the best-rated aspects (96%), followed by price-quality ratio (70%), staff (70%), food (68%), hygiene (66%), and infrastructure (62%). Meanwhile, the least well-rated aspects are amenities (10%), equipment (38%), and administrative processes and services (58%). See Figure 1.

Figure 1. Percentage of positive and negative comments, by attribute



Source: Author's own survey data.

The application of the chi-square independence test between comments and sample characteristics revealed a statistically significant association only between the type of visitor comment and the number of hotel stars ($\chi 2$ (3, N = 1,467) = 13.53, $p < .01$). This demonstrates that positive ratings are more frequent in higher-rated hotels (four and five stars), while more negative opinions are found in two-star hotels. See Figure 2.

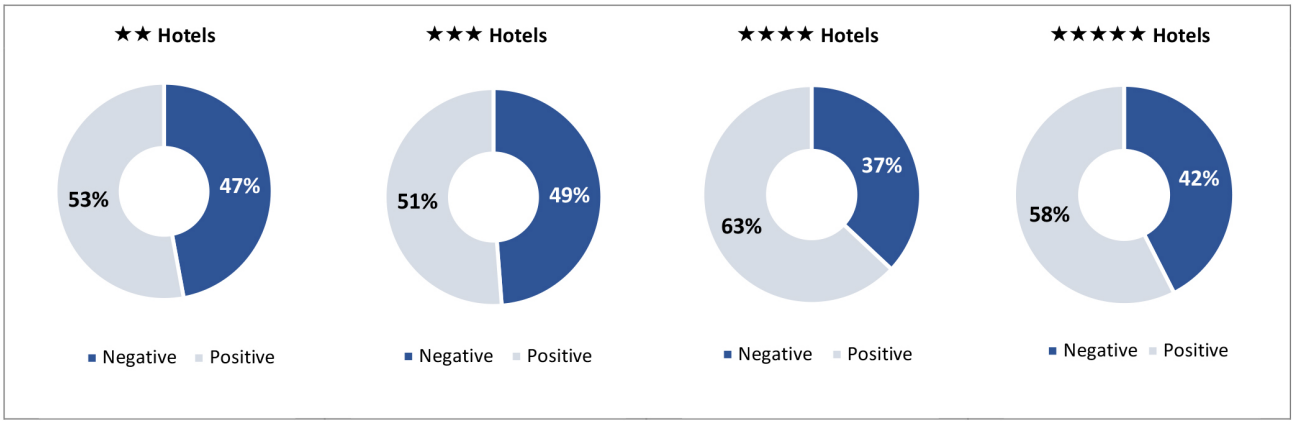
Binary logistic regression models

Table 5 summarizes the results of the logistic regressions by study segment. Overall, the most relevant attributes for hotel guests are location, amenities, price-quality ratio, staff, hygiene, food, and infrastructure. It is noteworthy that a visitor is more likely to experience a positive sentiment towards the hotel when satisfied with the location (56 times more likely), price-quality ratio (5.7 times), staff (3.3 times), hygiene (2.8 times), food (2.7 times), and infrastructure (2.2 times).

Therefore, it can be observed that the effects of the odds ratios (ORs) vary in intensity according to the hotel attribute. According to Chen et al. (2010) classification, the most relevant factors are location and amenities, which have a large impact (OR > 6.7). Secondly, in terms of importance, the price-quality ratio is moderate (3.47 < OR < 6.71). On the other hand, staff, hygiene, food, and infrastructure show a relatively smaller impact (between 1.68 and 3.47).
In the analysis of attributes by sample characteristics, disaggregated by gender, it is highlighted that for both men and women, location (b=3.8 and b=4.2), amenities (b=1.9 and b=2.6), price-quality ratio (b=1.3 and b=2.5), staff (b=1.2 and b=1.3), food (b=0.9 and b=1.1), and infrastructure (b=0.8 and b=0.9) are prioritized aspects. Comparatively, it is notable that for women, the price-quality ratio and hygiene are more relevant than for men.

Regarding the type of companions, it is evident that location emerges as the primary attribute for all evaluated segments, i.e., for those traveling with friends (b=4.3), as a couple (b=4.2), or with family (b=3.8). Additionally, although with different relative importance, the significant influence of price-quality ratio, amenities, staff, and food is noted, with beta coefficients ranging between 0.69 and 2.8. Furthermore, the examination of differences between types of companionship shows that infrastructure is important for those traveling with friends (b=1.2) and as a couple (b=0.9), while administrative processes and services gain importance for those traveling with family members (b=1.4), and hygiene is crucial for those traveling as a couple (b=1.1). Additionally, it is noted that location, price-quality ratio, and staff are more relevant for those traveling with friends or as a couple.

Figure 2. Percentage of positive and negative comments, by hotel star rating



Source: Author's' own survey data.

Regarding categorization by the number of hotel stars, it is revealed that the priority factors for guests, overall, are location, price-quality ratio, staff, and hygiene. Additionally, the significance of amenities is highlighted in most hotels, except for two-star ones. On the other hand, for guests in two-star hotels, food (b=2.5) and hygiene (b=2.1) are more significant, while for tourists choosing luxury establishments, administrative processes and services are essential (b=1.5). It is interesting to note that the price-quality ratio becomes more relevant for guests in two-star hotels (b=4.9), while location is established as a key factor for users of five-star (b=4.5) and four-star (b=4.2) hotels.

Finally, it is important to highlight that the models demonstrated a good fit according to the Hosmer-Lemeshow test (p > .05) and had a moderate level of explanatory power, with a Nagelkerke R-squared of approximately 50%, ranging between 43.4% and 68.6%. The models managed to explain a global percentage of between 74.3% and 83.3% of cases overall. Additionally, it is relevant to mention that the models related to the classification by number of hotel stars exhibited higher explanatory power, with an average Nagelkerke R-squared of 54%.

Table 5. Results of binary logistic regression by study segment, beta, and Odds Ratio (OR)

Attribute	Total (n=1467)	Male (n=696)	Female (n=600)	Other (n=171)	Friends (n=347)	Family (n=360)	Couple (n=359)	2-star (n=155)	3-star (n=330)	4-star (n=549)	5-star (n=433)
Quality-Price	1.73** 5.7	1.3** 3.7	2.51** 12.3	1.65* 5.2	2.29** 9.9	1.08* 2.9	1.83** 6.3	4.9** 134.4	1.47* 4.3	1.27** 3.6	1.85** 6.3
Admin. Serv.	0.16 1.2	0.03 1.0	0.13 1.1	1.02 2.8	-0.58 0.6	1.38** 4.0	-0.1 0.9	0.26 1.3	1.93 6.9	-0.22 0.8	1.54** 4.7
Amenities	2.27** 9.7	1.87** 6.5	2.63** 13.9	3.02* 20.4	2.18* 8.8	2.56** 12.9	2.76** 15.7	1.07 2.9	2.13** 8.4	3.68** 39.7	2.23** 9.3
Equipment	0.16 1.2	0.48 1.6	-0.08 0.9	-0.07 0.9	-0.3 0.7	0.5 1.6	0.42 1.5	-1.47 0.2	2.2** 9.0	-0.51 0.6	0.75* 2.1
Food	0.98** 2.7	0.87** 2.4	1.09** 3.0	1.11* 3.0	0.91* 2.5	0.83* 2.3	0.69* 2.0	2.49** 12.1	0.17 1.2	0.53 1.7	1.55** 4.7
Hygiene	1.02** 2.8	0.74* 2.1	1.32** 3.7	1.03 2.8	0.69 2.0	0.39 1.5	1.1* 3.0	2.05** 7.8	0.33 1.4	0.92** 2.5	0.79 2.2
Infrastructure	0.81** 2.2	0.8** 2.2	0.91** 2.5	0.54 1.7	1.16** 3.2	0.59 1.8	0.89* 2.4	1.41* 4.1	-0.83* 0.4	1.56** 4.8	0.77* 2.2
Location	4.03** 56.1	3.83** 46	4.15** 63.5	4.86** 128	4.31** 74.1	3.84** 46.7	4.15** 63.7	2.73** 15.4	21.82 M	4.23** 69	4.47** 87.4
Staff service	1.19** 3.3	1.19** 3.3	1.28** 3.6	0.82 2.3	1.48** 4.4	0.99** 2.7	1.53** 4.6	1.93** 6.9	0.9* 2.5	0.91** 2.5	1.49** 4.4
Constant	-3.75**	-3.47**	-4.04**	-4.79**	-2.86**	-5.03**	-4.33**	-2.46*	-6.81**	-4.01**	-5.66**
HL Test	.143	.209	.853	.509	.920	.245	.765	.841	.279	.464	.183
Cox & Snell R2	.343	.324	.376	.358	.341	.350	.372	.390	.515	.322	.382
Nagelkerke R2	.461	.434	.506	.485	.459	.470	.501	.521	.686	.439	.513
% Global	.755	.743	.778	.760	.755	.772	.783	.761	.833	.769	.762

Note: *p<5%, **p<1%. Dependent variable = positive sentiment. M = very large number. Effect size of the OR: insignificant (OR <1.68), small (1.68< OR< 3.47), moderate (3.47< OR <6.71), and big (OR>6.7).

DISCUSSION

The study findings confirm the Lancasterian conception of hotel lodging services, demonstrating that tourists' feelings towards establishments, whether positive or negative, are based on the unique value attributed to a proposition or combination of accommodation attributes, such as location, infrastructure, amenities, and staff service, among other factors (Boz et al., 2020). These preferences, in turn, vary according to the behavioral and sociodemographic characteristics of each individual (Laesser et al., 2019). Furthermore, the findings align with models of bounded rationality of the consumer, which suggest that people (in this case, tourists) make decisions based on heuristics that are strongly influenced by their cognitive capacity, availability of time, and resources (Conlisk, 1996).

In this context, a series of predominant attributes have been identified that influence tourists' perceptions and feelings about hotels. Among the positive aspects, location, value for money, staff, food, hygiene, and infrastructure stand out. In contrast, amenities were negatively highlighted. These attributes, together, define the overall experience of the tourist during their stay at a hotel, emphasizing the importance of considering both favorable and unfavorable aspects to drive continuous improvement and differentiation in an increasingly competitive global market (Mondaca-Marino et al., 2019).

The regression models analysis confirms that location is pivotal in generating positive emotions among tourists towards hotels, serving as a key strategic factor for competitive advantage (Bi et al., 2020). In this regard, Latinopoulos (2020) concluded that location is highly significant for guest satisfaction, considering factors such as proximity to points of interest, convenience for travel within and outside the city, neighborhood ambiance (hospitality), availability of recreational activities, and the option to enjoy panoramic views from the room. Moreover, it was observed that location is more relevant for those traveling with friends or partners, possibly due to the greater availability of options and experiences. For instance, families often prefer renting vehicles for more independent travel, which implies they may opt for less central locations but more convenient in terms of comfort and access. Additionally, this preference is accentuated in three-star hotels and attenuated in two-star ones, which could reflect a classic trade-off between less convenient placement and a more economical price. These findings carry significant implications for entrepreneurs and policymakers in determining where to build hotels to ensure an attractive offer for tourists.

Another crucial aspect is the amenities, as they represent a minimum standard expected by guests during their stay. These complementary services, such as toiletries, gifts, internet access, gym, pool, among others, significantly contribute to the guest's well-being. For example, Shah and Bhatt (2020) revealed that inadequate provision of amenities can impact tourist satisfaction and their willingness to repeat the visit. Verma and Thakur (2022), on the other hand, while confirming the relevance of this attribute, argue that the adoption of technological amenities can be a differentiating element that positively influences guest sentiment. Along these lines, it is interesting to note the differences found among different visitor segments regarding the importance attributed to hotel amenities. While for female tourists, or those traveling as couples or choosing higher-category hotels, this aspect is more relevant, for male travelers or those opting for lower-category hotels, its impact is lower. These discrepancies underscore the need to consider demographic and behavioral aspects when evaluating the tourist experience, as supported in previous studies (Mak et al., 2012; Tian et al., 2023; Wong et al., 2016).

The price-quality ratio also emerges as a crucial aspect for guests during their stay, significantly impacting their level of satisfaction and loyalty (Ahmed et al., 2023). The results indicate that this attribute becomes more relevant for female tourists (Hong et al., 2020) or those traveling with friends (Xu, 2018), as well as for guests staying in lower-star-rated hotels (Razavi & Israeli, 2019). Consequently, these segments would exhibit higher price sensitivity in terms of elasticity. In contrast, for men and those traveling with family, the effect of the price-quality ratio becomes relatively less significant.

The quality of service provided by the staff also emerged as a fundamental pillar in the hotel experience, highlighting the importance of empathy, reliability, and excellence in service, in line with findings in the specialized literature (Castanha et al., 2023). Jawabreh et al. (2022) emphasize that the quality of attention and the communication skills of the staff are priorities for providing excellent service, emphasizing the need for appropriate recruitment and selection mechanisms for staff and continuous trai-

ning and development processes. Likewise, Xin and Choi (2020) confirm the strategic and positive influence of service quality on guest satisfaction and loyalty. The authors emphasize the importance of maintaining clarity in the functions and responsibilities of each employee through effective communication from the management team.

Hygiene similarly represents a prominent attribute for guests (Hernández et al., 2022). As evidenced in the study by Stefko et al. (2020), hygiene can generate divergent opinions among tourists, making it a fundamental factor that hotel managers should consider to ensure guest satisfaction. Furthermore, the findings showed that it is especially relevant for women, who often value hygiene as an aspect that guarantees comfort and health during their stay (Sürme, 2022). Additionally, it was found to be more valued by those staying in two-star hotels, where hygiene issues are more likely to arise compared to higher-category accommodations. Therefore, it is essential for these types of establishments to implement rigorous measures to maintain high standards of hygiene and cleanliness in their rooms.

Regarding food, the results demonstrate that it is a fundamental component in the overall evaluation of the lodging experience (Nilashi et al., 2023). According to Tager et al. (2021), the quality of food service not only affects guest satisfaction but also influences their willingness to return to the hotel and their propensity to recommend it. It is interesting to note that this aspect showed a significant effect in two- and five-star hotels. In one-star hotels, where services are usually more basic, food may be one of the few aspects where guests expect a satisfactory experience, being essential to compensate for potential deficiencies and ensure a pleasant stay (Boto-Garcia et al., 2021). On the other hand, in five-star hotels, the quality of food becomes a primary factor in the overall luxury experience. Visitors seek refined and memorable gastronomic service that complements the high level of service they expect to receive in all aspects of their stay (Ismail et al., 2022).

Infrastructure also emerges as a critical feature for guests. According to Xue and Chen (2021), maintaining facilities in optimal conditions helps prevent a negative perception of the hotel. Users assume that this minimum standard will be met, and its absence can result in significant dissatisfaction and disinterest in using the establishment's services again. The findings indicate that infrastructure becomes more relevant among female guests or those traveling with friends, which could be linked to requirements sought during the stay such as safety, comfort, accessibility, and overall experience (such as recreational activities), which influence the quality, functionality, and design of the facilities.

On the other hand, administrative processes and services, such as reservations, guest check-in and check-out, payment management, among others, were only relevant for families and guests of luxury hotels (Clauzel et al., 2020), highlighting their importance for those prioritizing a family-oriented and exclusive lodging experience. Additionally, while equipment was significant, it could be considered a secondary aspect compared to other elements that more directly influence the guest experience.

Finally, the study revealed that in higher-rated hotels, especially those with four or five stars, positive reviews are more common, while unfavorable ones predominate in two-star hotels. This phenomenon, linked to the service standard, shows that higher-rated hotels offer a wide range of services and amenities, resulting in more satisfying experiences and, consequently, more positive guest reviews (Lo & Yeung, 2020). In contrast, the most negative reviews in two-star hotels would reflect perceptions of less satisfactory services, such as cleanliness deficiencies or lack of room amenities, and visitors are likely to have lower expectations, increasing the likelihood of expressing negative reviews if these expectations are not met (Boto-Garcia et al., 2021).

CONCLUSIONS

From the study results and their discussion, it is concluded that data mining and sentiment analysis are effective tools for understanding the factors related to tourists' positive and negative emotions, as well as their ability to predict the choice of one hotel over another. These findings have significant implications for hotel management, as understanding the dynamic behavior of tourists becomes a key competitive advantage for any hotel. Especially, since it has been demonstrated that meeting expectations is positively and significantly related to tourist satisfaction and subsequent loyalty (Yoon & Uysal, 2005).

Furthermore, this research provided updated and relevant information regarding the sentiments of tourists staying in hotel establishments in Santiago, Chile, which is useful for managers associated with the hotel industry and for

scholars associated with the study of tourism and hospitality industry. The analysis of the results indicates that there are critical attributes that influence the generation of a positive perception towards hotels, the effective management of which would significantly increase the likelihood of satisfaction and loyalty from tourists.

The location emerged as the most decisive and highly valued aspect by visitors. It is followed in importance by amenities and the price-quality ratio. Additionally, it is indispensable to consider aspects such as having well-trained and friendly staff, adequate infrastructure, quality food, and high standards of hygiene in hotel management. In this sense, the results suggest that lower-category hotels can effectively compete with higher-category ones by focusing on attributes such as price-quality ratio, hygiene, and service personnel, rather than on infrastructure or amenities provision.

Finally, it is recommended to expand this type of study by applying data mining techniques to other accommodation platforms or related websites, and geographical contexts. Additionally, it would be interesting to complement the results through the collection and analysis of primary data, either quantitative or qualitative in nature, which would enrich the understanding of the factors influencing customers' tourism experiences, identify emerging trends, and adapt hotel management strategies more precisely and effectively. Furthermore, based on the study findings, it is relevant to delve into the determinants of guest satisfaction from the perspective of hotel category, considering the socio-demographic and psychographic profile of the visitor. This would provide key insights for personalizing and improving service offerings, addressing the specific preferences and needs of each tourist segment.

REFERENCES

Ahmed, S., Al Asheq, A., Ahmed, E., Chowdhury, U. Y., Sufi, T., & Mostofa, M. G. (2023). The intricate relationships of consumers' loyalty and their perceptions of service quality, price and satisfaction in restaurant service. TQM Journal, 35(2), 519–539. https://doi.org/10.1108/TQM-06-2021-0158

Bi, J.-W., Liu, Y., Fan, Z.-P., & Zhang, J. (2020). Exploring asymmetric effects of attribute performance on customer satisfaction in the hotel industry. Tourism Management, 77, 1–18. https://doi.org/10.1016/j.tourman.2019.104006

Boto-Garcia, D., Escalonilla, M., Zapico, E., & Baños, J. F. (2021). Scale heterogeneity in hotel guests' satisfaction relative to room rates. Applied Economic Analysis, 29(87), 208–225. https://doi.org/10.1108/AEA-07-2020-0093

Boz, Z., Korhonen, V., & Sand, C. K. (2020). Consumer considerations for the implementation of sustainable packaging: A review. Sustainability (Switzerland), 12(6), 1–34. https://doi.org/10.3390/su12062192

Carneiro, T., Picoto, W. N., & Pinto, I. (2023). Big Data Analytics and Firm Performance in the Hotel Sector. Tourism and Hospitality, 4(2), 244–256. https://doi.org/10.3390/tourhosp4020015

Castanha, J., Pillai, S. K. B., & Sankaranarayanan, G. G. (2023). Customer Satisfaction Towards Co-Operative Banks-a Systematic Review and Research Agenda. Revista Anais Brasileiros de Estudos Turísticos, 13(Edição Regular), 1–9. https://periodicos.ufjf.br/index.php/abet/article/view/41469/26778

Castro, F., Muñoz, L., Alliende, M., & del Río, M. (2023). Estudio sobre el mercado del hospedaje. https://www.fne.gob.cl/wp-content/uploads/2023/12/Informe_Preliminar_VF-EM08.pdf

Chatterjee, S. (2020). Drivers of helpfulness of online hotel reviews: A sentiment and emotion mining approach. International Journal of Hospitality Management, 85, 1–9. https://doi.org/10.1016/j.ijhm.2019.102356

Chen, H., Cohen, P., & Chen, S. (2010). How big is a big odds ratio? Interpreting the magnitudes of odds ratios in epidemiological studies. Communications in Statistics: Simulation and Computation, 39(4), 860–864. https://doi.org/10.1080/03610911003650383

Clauzel, A., Guichard, N., & Damay, C. (2020). Exploring emotional traces in families' recollection of experiences: A study based on luxury hotel reviews. Qualitative Market Research, 23(1), 21–45. https://doi.org/10.1108/QMR-07-2018-0076

Conlisk, J. (1996). Why bounded rationality? Journal of Economic Literature, 34(2), 669–700. https://www.jstor.org/stable/2729218

Darmawan, D., Purwanto, F., & Infante, A. (2021). The Relationship Between Satisfaction And Loyalty Of Hotel Service Users In The Tourism City Of Batu, Malang. Journal of Engineering and Social Sciences (JESS), 1(1), 49–56. https://journal.unsuri.ac.id/index.php/jess/article/view/111

Gupta, M. K., & Chandra, P. (2020). A comprehensive survey of data mining. International Journal of Information Technology (Singapore), 12(4), 1243–1257. https://doi.org/10.1007/s41870-020-00427-7

Hernández, L. F., Morales, V. H., & Tello, G. (2022). Aspectos de la higiene y seguridad en hotelería y/o gastronomía y su importancia para el desarrollo turístico de Rio Turbio, Santa Cruz. Europub Journal of Education Research, 3(1), 157–177. https://doi.org/10.54745/ejerv3n1-012

Hong, J., Repetti, T., Erdem, M., & Henthorne, T. (2020). Effect of guest demographics on perception of hotel room price. Journal of Hospitality and Tourism Insights, 3(1), 3–16. https://doi.org/10.1108/JHTI-01-2019-0014

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

Ismail, T. A. T., Zahari, M. S. M., Hanafiah, M. H., & Balasubramanian, K. (2022). Customer Brand Personality, Dining Experience, and Satisfaction at Luxury Hotel Restaurants. Journal of Tourism and Services, 13(24), 26–42. https://doi.org/10.29036/jots.v13i24.278

Jawabreh, O., Masa'Deh, R., Fayoumi, T. Al, & Bareya, A. (2022). Customer Satisfaction Towards Communication Skills of Front Office Staff At the Hotel (Case Study Five Stars Hotels in Aqaba). Geojournal of Tourism and Geosites, 44(4), 1312–1318. https://doi.org/10.30892/gtg.44416-948

Jung, O. H., & Kim, H. (2016). Are tourists rational or irrational consumers? Asia Pacific Journal of Tourism Research, 21(11), 1169–1183. https://doi.org/10.1080/10941665.2015.1125378

Kampalpour, M., Aghdam, A., Xu, S., Khani, E., & Baghi, A. (2017). Uncovering hotel guests preferences through data mining techniques. International Journal of Computer Science and Network Security, 17(8), 1–10.

Khamphakdee, N., & Seresangtakul, P. (2021). Sentiment analysis for Thai language in hotel domain using machine learning algorithms. Acta Informatica Pragensia, 10(2), 155–171. https://doi.org/10.18267/j.aip.155

Laesser, C., Luo, J., & Beritelli, P. (2019). The SOMOAR operationalization: a holistic concept to travel decision modelling. Tourism Review, 74(3), 613–631. https://doi.org/10.1108/TR-06-2018-0069

Latinopoulos, D. (2020). Analysing the role of urban hotel location in guests' satisfaction. Anatolia, 31(4), 636–650. https://doi.org/10.1080/13032917.2020.1808489

Lo, A., & Yeung, M. A. (2020). Brand prestige and affordable luxury: The role of hotel guest experiences. Journal of Vacation Marketing, 26(2), 247–267. https://doi.org/10.1177/1356766719880251

Mak, A. H. N., Lumbers, M., Eves, A., & Chang, R. C. Y. (2012). Factors influencing tourist food consumption. International Journal of Hospitality Management, 31(3), 928–936. https://doi.org/10.1016/j.ijhm.2011.10.012

Manolitzas, P., Glaveli, N., Palamas, S., Talias, M., & Grigoroudis, E. (2022). Hotel guests' demanding level and importance of attribute satisfaction ratings: an application of MUltiplecriteria Satisfaction Analysis on TripAdvisor's hotel guests ratings. Current Issues in Tourism, 25(8), 1203–1208. https://doi.org/10.1080/13683500.2021.1915253

Millán-García, C. H., & Gómez-Díaz, M. del R. (2018). Factores e indicadores de competitividad hotelera. Compendium, 21(40). https://www.redalyc.org/articulo.oa?id=88055200021

Mondaca-Marino, C. M., Guala, C., Montecinos-Astorga, A. L., & Concha, C. S. (2019). Factores que Influyen en el Precio de Hoteles en Booking.com. El Caso de Santiago de Chile. Informacion Tecnologica, 30(1), 87–96. https://doi.org/10.4067/S0718-07642019000100087

Moro, S., Rita, P., & Oliveira, C. (2018). Factors Influencing Hotels' Online Prices. Journal of Hospitality Marketing and Management, 27(4), 443–464. https://doi.org/10.1080/19368623.2018.1395379

Nilashi, M., Abumalloh, R. A., Almulihi, A., Alrizq, M., Alghamdi, A., Ismail, M. Y., Bashar, A., Zogaan, W. A., & Asadi, S. (2023). Big social data analysis for impact of food quality on travelers' satisfaction in eco-friendly hotels. ICT Express, 9(2), 182–188. https://doi.org/10.1016/j.icte.2021.11.006

Nunkoo, R., Teeroovengadum, V., Ringle, C. M., & Sunnassee, V. (2020). Service quality and customer satisfaction: The moderating effects of hotel star rating. International Journal of Hospitality Management, 91, 1–15. https://doi.org/10.1016/j.ijhm.2019.102414

Özen, İ. A., & Özgül, E. (2023). Aspect-based sentiment analysis on online customer reviews: a case study of technology-supported hotels. Journal of Hospitality and Tourism Technology, 14(2), 102–120. https://doi.org/10.1108/JHTT-12-2020-0319

Paulose, D., & Shakeel, A. (2022). Perceived Experience, Perceived Value and Customer Satisfaction as Antecedents to Loyalty among Hotel Guests. Journal of Quality Assurance in Hospitality and Tourism, 23(2), 447–481. https://doi.org/10.1080/1528008X.2021.1884930

Ray, B., Garain, A., & Sarkar, R. (2021). An ensemble-based hotel recommender system using sentiment analysis and aspect categorization of hotel reviews. Applied Soft Computing, 98, 1–18. https://doi.org/10.1016/j.asoc.2020.106935

Razavi, R., & Israeli, A. A. (2019). Determinants of online hotel room prices: comparing supply-side and demand-side decisions. International Journal of Contemporary Hospitality Management, 31(5), 2149–2168. https://doi.org/10.1108/IJCHM-09-2018-0707

Samara, D., Magnisalis, I., & Peristeras, V. (2020). Artificial intelligence and big data in tourism: a systematic literature review. Journal of Hospitality and Tourism Technology, 11(2), 343–367. https://doi.org/10.1108/JHTT-12-2018-0118

Sánchez-Oro, M., Robina-Ramírez, R., Portillo-Fernández, A., & Jiménez-Naranjo, H. V. (2021). Tourist expectations and motivations in visiting rural destinations. The case of extremadura (Spain). Revista Espanola de Investigaciones Sociologicas, 175, 105–128. https://doi.org/10.5477/cis/reis.175.105

Shah, S., & Bhatt, D. K. (2020). A Study on Tourist Satisfaction With the Services Provided By the Hotel Industry of Nainital. International Journal of Tourism & Hotel Business Management (IJTHBM), 2(1), 202–217. https://doi.org/10.13140/RG.2.2.11321.49762

Spoerr, D. (2021). Factor Analysis of Hotel Selection Attributes and Their Significance for Different Groups of German Leisure Travelers. Journal of Quality Assurance in Hospitality and Tourism, 22(3), 312–335. https://doi.org/10.1080/1528008X.2020.1774032

Stefko, R., Fedorko, R., Bacik, R., Rigelsky, M., & Olearova, M. (2020). Effect of service quality assessment

on perception of TOP hotels in terms of sentiment polarity in the Visegrad group countries. Oeconomia Copernicana, 11(4), 721–742. https://doi.org/10.24136/OC.2020.029

Sürme, M. (2022). A Research To Determine The Travel Behaviors Of Women Tourists. Journal of Social, Humanities and Administrative Sciences, 8(49), 168–181. https://doi.org/10.31589/joshas.919

Tager, A., Zaki, M., Megahed, F., & Abbas, T. (2021). The Effect of Food Quality, Service Quality, And Tangibles on Hotel Restaurants Customer Behavioural Intentions: The Mediating Role of Customer Satisfaction. International Journal of Tourism and Hospitality Management, 4(2), 218–242. https://doi.org/10.21608/ijthm.2021.206790

Tian, Z., Wang, Q., Liu, Y., & Wang, Z. (2023). Comparison of the tourist mobility patterns among demographic subgroups with mobile phone data: A case study of Chongqing, China. Applied Geography, 159, 103090. https://doi.org/10.1016/j.apgeog.2023.103090

Verma, V., & Thakur, S. (2022). The rise of the era of technological amenities and their adoption in the hotel industry. World Review of Science, Technology and Sustainable Development, 18(3–4), 379–397. https://doi.org/10.1504/WRSTSD.2022.123782

Wankhade, M., Rao, A. C. S., & Kulkarni, C. (2022). A survey on sentiment analysis methods, applications, and challenges. Artificial Intelligence Review, 55(7), 5731–5780. https://doi.org/10.1007/s10462-022-10144-1

Wong, I. K. A., Fong, L. H. N., & Law, R. (2016). A Longitudinal Multilevel Model of Tourist Outbound Travel Behavior and the Dual-Cycle Model. Journal of Travel Research, 55(7), 957–970. https://doi.org/10.1177/0047287515601239

Wu, X., Wang, J., & Ling, Q. (2021). Managing internal service quality in hotels: Determinants and implications. Tourism Management, 86, 1–13. https://doi.org/10.1016/j.tourman.2021.104329

Xin, G., & Choi, J. G. (2020). The impact of factors formingemployee service attitude on service quality, customer satisfaction and loyalty in the hotel industry. Global Business and Finance Review, 25(3), 84–95. https://doi.org/10.17549/gbfr.2020.25.3.84

Xu, X. (2018). Does traveler satisfaction differ in various travel group compositions? International Journal of Contemporary Hospitality Management, 30(3), 1663–1685. https://doi.org/10.1108/IJCHM-03-2017-0171

Xue, H. J., & Chen, X. M. (2021). Hotel Quality Evaluation and Promotion Strategy Based on IPA Analysis: An Empirical Case of Sofitel Hotel in Xining City. 2021 2nd International Conference on Big Data Economy and Information Management (BDEIM), 168–171. https://doi.org/10.1109/BDEIM55082.2021.00041

Yoon, Y., & Uysal, M. (2005). An examination of the effects of motivation and satisfaction on destination loyalty: A structural model. Tourism Management, 26(1), 45–56. https://doi.org/10.1016/j.tourman.2003.08.016

Zarezadeh, Z. Z., Rastegar, R., & Xiang, Z. (2022). Big data analytics and hotel guest experience: a critical analysis of the literature. International Journal of Contemporary Hospitality Management, 34(6), 2320–2336. https://doi.org/10.1108/IJCHM-10-2021-1293