

Exploring the attributes of hotel service quality in Florianópolis-SC, Brazil: An analysis of tripAdvisor reviews

Clérito Kaveski Peres & Edson Pacheco Paladini |

To cite this article: Clérito Kaveski Peres & Edson Pacheco Paladini | (2021) Exploring the attributes of hotel service quality in Florianópolis-SC, Brazil: An analysis of tripAdvisor reviews, Cogent Business & Management, 8:1, 1926211, DOI: [10.1080/23311975.2021.1926211](https://doi.org/10.1080/23311975.2021.1926211)

To link to this article: <https://doi.org/10.1080/23311975.2021.1926211>



© 2021 The Author(s). This open access article is distributed under a Creative Commons Attribution (CC-BY) 4.0 license.



Published online: 19 May 2021.



Submit your article to this journal [↗](#)



Article views: 595



View related articles [↗](#)



View Crossmark data [↗](#)



Received: 6 November 2020
Accepted: 29 April 2021;

*Corresponding author: Clérito Kaveski Peres, Department of Industrial and Systems Engineering, Federal University of Santa Catarina, Florianópolis, Brazil
Email: cleritokp@gmail.com

Reviewing editor:
Iman Rahimi, Young Researchers and Elite Club, Malaysia

Additional information is available at the end of the article

MANAGEMENT | RESEARCH ARTICLE

Exploring the attributes of hotel service quality in Florianópolis-SC, Brazil: An analysis of tripAdvisor reviews

Clérito Kaveski Peres* and Edson Pacheco Paladini

Abstract: This study explores user-generated content (UGC) of the hotel sector in the city of Florianópolis-SC, Brazil, to identify the quality attributes of services and determine the polarity of the expressed feelings in the reviews of each attribute. The analysis is based on the latent topics and the polarity of feelings expressed in the reviews. UGC was collected using a crawler, resulting in a text corpus comprising 68,558 reviews. The polarity of feelings, positive or negative, was identified using sentiment analysis techniques and Latent Dirichlet Allocation (LDA) was used to identify latent topics in the corpus associated with the attributes of hotel service quality. This study found that “room,” “location,” “ambience,” “staff,” “breakfast,” “parking,” “reservation,” and “cost-benefit” were the attributes most frequently assessed by consumers in their reviews. The attributes that generate the most negative reviews were “room,” “parking,” and “reservation.” The attributes “location,” “ambience,” “staff,” “breakfast,” and “cost-benefit” were the attributes that generated most of the positive reviews. When comparing the results of this study to those of previous studies, two attributes demonstrated greater prominence: the attribute “room” that attracted a high number of negative comments and the attribute “parking” that had not presented itself with the same level of relevance in other studies.

Subjects: service quality; hotel industry; artificial intelligence;

Keywords: hospitality; Latent Dirichlet Allocation; sentiment analysis

ABOUT THE AUTHOR

Clérito Kaveski Peres is a PhD student in Production Engineering at the Federal University of Santa Catarina. His research interests include Management, Quality Evaluation and Artificial Intelligence.

Edson Pacheco Paladini is Professor of the Department of Production and Systems Engineering at the Federal University of Santa Catarina. Titular member of the Brazilian Quality Academy. His research interests include Engineering, Management and Quality Evaluation.

PUBLIC INTEREST STATEMENT

The presence of a growing volume of user-generated content (UGC) on online platforms in the hospitality sector, due to the advancement of information technologies, has proved to be a new source of useful data for the interpretation of customer needs. However, given the environmental and cultural differences in the contexts in which this content is generated, its interpretation becomes a challenge for managers and researchers interested in the subject. In this sense, this study is dedicated to exploring UGC, using natural language processing techniques (NLP), to identify the quality attributes addressed by customers and measure the polarity of the feelings expressed in the customer evaluations to provide consistent results and contribute to better decision-making in hotel management.

1. Introduction

Tourism is one of the most important sectors in the world economy, accounting for 10.4% of gross domestic product (GDP) and responsible for 9.9% of all jobs, according to the most recent data (WTTC, 2018b). Despite the importance of international tourism because of export revenues, domestic tourism investment represents 73% of the sector's total revenue (WTTC, 2018a).

In Brazil, tourism represents 8.5% of GDP. Brazil is the most dependent on domestic tourism, which accounts for 94% of the sector's total revenue (WTTC, 2018a). In 2017, Brazil had 60 million domestic tourists (Brasil, 2018).

Investing in domestic tourism has been an important strategy adopted by several countries to eliminate local poverty, generate jobs and economic growth, modernize infrastructure, relieve the pressure of overcrowding, and regulate the seasonality of regional visitors (WTTC, 2018a).

Moreover, the tourism industry and other economic sectors have been undergoing a transformation driven by advances in information technologies. Recommendation systems, online reservations, dynamic pricing, and interactive service review platforms have transformed the way individuals consume tourism products (Sanchez-Franco et al., 2019).

Today, customers are connected, critical in their evaluation process, have new demands, and expect to be surprised by each new experience. In this way, continuing to focus on the customer by monitoring their needs is necessary to navigate changes to survive in the market.

Previous studies conducted in different contexts in the hospitality field, such as on hotels in New York (Lee et al., 2020), Malaysia (Padma & Ahn, 2020), and in different countries (Guo et al., 2017), and on luxury hotels in Spain have shown significant aspects of the quality of services.

These studies are based on the user-generated content (UGC) on digital platforms and on websites focused on online reviews, which have allowed for a fast growth in the volume of data on the hospitality sector. In the case of hotels, customer reviews through spontaneous feedback on hotels' goods and services and UGC can be freely accessed and collected for analysis (Guo et al., 2017).

UGC enables the identification of significant structures between various aspects and attributes related to hospitality and tourism products, and is a promising area of research in social media analysis in the field of hospitality and tourism (Xiang et al., 2017). However, previous studies have not established a consensus regarding the relevance level of quality attributes about hotels. Studies have identified distinct sets of attributes that vary in relation to the level of relevance given by customers, and according to the environment and profile of the traveler (Sann & Lai, 2020a, 2020b; Sun et al., 2017; Ying et al., 2020).

Most of the previous studies also seek to determine the polarity of the feelings expressed in the reviews using the score scales suggested by the platforms (Lee et al., 2020; Sann & Lai, 2020a; Ying et al., 2020), where low scores are identified as negative reviews and high scores are identified as positive reviews. Nevertheless, a reviewer may express positive and negative feelings regarding different quality attributes in the same sentence (B. Kim et al., 2016), which leads the way to a more in-depth analysis of sentiments in UGC.

Thus, this study explores UGC in the hotel sector in the city of Florianópolis-SC, Brazil, to identify the quality attributes of services and determine the polarity of the reviews, positive or negative, regarding each attribute. Additionally, this study seeks to expand the knowledge base about UGC

in Brazilian Portuguese, since the literature shows a predominance of studies that analyze reviews written in English compared to other languages (Ma et al., 2018).

This study is structured as follows: after this introduction, section 2 presents the concepts and results of the UGC approach in previous studies. Section 3 describes the methodology for collecting, cleaning, and analyzing the data. Section 4 presents the results. Section 5 presents the discussion and explores the polarity of feelings around the identified quality attributes. Section 6 presents the study's conclusions and limitations.

2. Literature review

In the hospitality and tourism industry, customer satisfaction is seen as a general emotional response to any intangible service. Whether the service meets or exceeds expectations, customers are generally satisfied (Li et al., 2013). Moreover, customer satisfaction is a factor that positively influences customer loyalty (Malik et al., 2020).

Just as people have different levels of experience and knowledge, they have different expectations and perceptions about the same product. It is up to organizations to identify and equalize the offer and delivery of products to their customers. To achieve this, it is necessary to identify the attributes of product quality that are valued by the customer and determine the level at which they are being satisfied. This is a difficult task that requires the formulation of models, which we discuss below.

2.1. Service quality models

There is a consensus in the development of service quality models that the abstract dimensions of perception, expectation, and satisfaction are defined by the customer and not by the service provider (Martin, 1995).

Thus, the correct identification of the main attributes of service quality is crucial to increase customer satisfaction. Several studies have been conducted to characterize service quality based on the subjective perceptions of clients and to identify the main factors that determine what is considered good service (Ju et al., 2019).

Instruments such as SERVQUAL and SERVPERF have been used for these purposes. SERVQUAL evaluates customer satisfaction and characterizes the gap between expectation and performance through a series of pre-established dimensions (Parasuraman et al., 1985). SERVPERF, created as an alternative to SERVQUAL, is based only on the perception of service performance (Cronin & Taylor, 1992).

Although these instruments are extensively used, they have known limitations, such as the lack of temporal resolution, meaning that they do not capture the frequency of visits to an establishment, and the dependence of the interviewees' perceptions on remembering past events (He et al., 2018).

Moreover, the development of empirical measurement scales relies on the prior identification of customer satisfaction dimensions based on the researchers' knowledge. The complexity of this type of research requires researchers to make a trade-off between the cost of collecting samples and the performance of the estimate (Guo et al., 2017).

In the hotel sector, the focus of this study is on other difficulties in implementing this type of methodology, such as the lack of co-operation from hotel managers for reasons related to the security of sensitive data about customers and about the business, and the high costs of performing the research (Malik et al., 2020).

However, new possibilities are now feasible as big data technologies allow a more accurate assessment of UGC (Alaei et al., 2019), as we discuss next.

2.2. Hotel service quality

Technological advances linked to the internet, and particularly social media platforms such as Facebook, Instagram, and Twitter, have dramatically increased UGC during recent years (Haghighi et al., 2018). Other platforms, such as TripAdvisor, Booking.com, Expedia, and Travelocity, were developed specifically for the tourism sector and promote reviews of consumer experiences in this sector (Xiang et al., 2017).

UGC is spontaneous and insightful feedback that is widely available, free or inexpensive, and can easily be accessed anywhere, anytime (Guo et al., 2017). It is an important source of information for researchers and professionals to understand customer preferences and demands in varied domains such as restaurants (Ha & Lee, 2018), accommodation (Ma et al., 2018), public transport (Nisar & Prabhakar, 2018), airports (Martin-Domingo et al., 2019), and traffic (Haghighi et al., 2018).

The greater capacity of data processing, given the advancement big data technologies in recent years, has allowed for considerable advances in the field of natural language processing (NLP). Unsupervised learning algorithms can be used to identify the attributes of quality through topic modeling and grouping. These algorithms enable the identification of aspects that users have focused on in reviews without the need for prior knowledge (Tran et al., 2019).

In the hospitality field, several studies have been conducted using UGC to understand the guests' reviews regarding the quality attributes of hotels and the factors that generate positive and negative comments.

In order to get a representative overview of previous studies in this specific study field, data were collected on scientific articles published over the last five years (2016–2020) in the following journals: *International Journal of Contemporary Hospitality Management*, *International Journal of Culture Tourism and Hospitality Research*, *International Journal of Hospitality Management*, *Journal of Hospitality and Tourism Technology*, *Journal of Hospitality and Tourism Management*, *Journal of Hospitality Marketing & Management*, *Journal of Travel Research*, *Journal of Vacation Marketing*, *Sustainability*, *Tourism Management*, and *Tourism Review*. Twenty-three articles were selected, as shown in Table 1.

The distinct results found in these studies have shown that the relevance given by guests to the quality attributes of the hotel services may vary depending on the environmental and cultural context. For example, a study conducted in five-star hotels in Spain using TripAdvisor's UGC identified several hotel quality attributes and classified them into "staff," "services," "room," and "location" (Rios-Martin et al., 2020).

A study conducted in New York, analyzed UGC on TripAdvisor and identified nine topics related to service quality attributes and classified them as "sensory experience," "brand," "hotel class," "sleep," "location," "room," "service," "value," and "cleaning" (Lee et al., 2020).

In Malaysia, a study analyzed TripAdvisor content for luxury hotels. The study revealed that the main topics related to the quality of luxury hotels' services were related to hotel attributes (restaurant and breakfast), "room," "staff," "trip" (walking, taxi, and business), and possible consequences of the "experience" (recommendation) (Padma & Ahn, 2020).

The main attributes identified in a study conducted in Portugal using data from TripAdvisor's UGC were "experience," "hotel," "learn," "host," "camping," "nature," "food," "ingredients," "different," "eco(logy)," and "yoga" (Brochado & Brochado, 2019).

The results of these studies highlight the need to identify the quality attributes of the services in each context to inform the correct decision-making by the managers. Another important aspect,

Table 1. Attributes of hotel service quality identified in previous studies

Author (year)	Data Source	Place	Hotel Attributes
(Sann & Lai, 2020a)	TripAdvisor	Bangkok, Thailand	Staff, service, room, location, bed, bathroom, sleep, breakfast, cleanliness, and hotel facilities & complementary services
(Dogan et al., 2020)	HolidayCheck.de	Turkey	Hotel condition, location, staff, food-beverage, facilities (sport and pools), and room
(Foris et al., 2020)	Booking.com, TripAdvisor, Directbooking.ro, Travelminit.ro	Romanian	Position, transport, organized tours/courses, accommodation, and disabilities
(Lee et al., 2020)	TripAdvisor	New York, USA	Sensory experience, brand, hotel class, sleep, location, room, service, value, and clean
(Padma & Ahn, 2020)	TripAdvisor	Malaysia	Attributes related to the hotel, room, staff, travel, and possible outcomes
(Sann & Lai, 2020b)	TripAdvisor	UK	Reservation, front-office, front-office, housekeeping, food quality, security, engineering, marketing, and information technology
(Rios-Martin et al., 2020)	TripAdvisor	Spain	Staff, services, room, and location
(Sann et al., 2020)	TripAdvisor	UK	Service, cleanliness, room, sleep quality, location, and value
(Ying et al., 2020)	Ctrip.com and TripAdvisor	China	Room, travel, food, environment, and hotel facility
(Alrawadieh & Law, 2019)	TripAdvisor	Istanbul, Turkey	Rooms, service quality, hotel's characteristics and food, and beverage
(Brochado & Brochado, 2019)	TripAdvisor	Portugal	Experience, hotel, learn, host, camping, nature, food, ingredients, different, eco and yoga
(Hu et al., 2019)	TripAdvisor	New York, USA	Facilities, service, location, value, and general experience
(Köseoglu et al., 2019)	Booking.com	Hong Kong	Location, room, staff, view, service, clean, breakfast, pool, restaurant, food etc.
(Xu, 2018)	Booking.com	USA	Room, staff, location, noise, wi-fi, luggage-carrying, facility, bed, and bathroom
(Fernandes & Fernandes, 2017)	TripAdvisor	Porto, Portugal	Location, sleep quality, rooms, customer care, value, cleanliness, bathrooms, breakfast, bar/restaurant, service, parking, and facilities
(Herjanto et al., 2017)	TripAdvisor	Panama	Physical hotel facilities, bedroom, hotel, staffs, restaurants, location, and price factors
(Liu et al., 2017)	TripAdvisor	China	Rooms, value, location, cleanliness, and service
(Phillips et al., 2017)	TrustYou	Switzerland	Grounds, building, ambiance, rooms, and internet
(Sun et al., 2017)	DaoDao.com and Ctrip. com.	Inter-Continental	Service, rooms and facilities, location, cleanliness, and value
(Tontini et al., 2017)	TripAdvisor	Rio de Janeiro, Brazil	Location, attitudes of employees, quality and variety of food/beverages, comfort, cleanliness, additional facilities, access, hotel value, wi-fi, check-in, check-out time, room, noise, management's knowledge, appearance, and additional facilities.
(Wu et al., 2017)	Agoda	Shanghai, China	Staff, hotel, room, location, shopping, and beautiful
(B. B. Kim et al., 2016)	TripAdvisor	New York, USA	Room, staff, hotel property/appearance, facility, and other
(Xu & Li, 2016)	Booking.com	USA	Location, room, restaurant, facility, parking, bathroom, Wi-Fi, staff, smoking and in-room polluted air, hallway machine, swimming pool, air-condition, and value

evidenced in previous studies, is the behavior variation regarding the quality attributes according to the traveler profile.

Research that used UGC of hotels in China revealed that Chinese guests expect personalized service, while North American guests prefer standardized service (Ying et al., 2020); the behavior of

guests of different nationalities toward quality attributes of hotels in the United Kingdom is influenced by cultural context (Sann & Lai, 2020b); in hotels in the United Kingdom, Asian guests encounter more service failures with respect to the engineering aspect of operations (room equipment issues), while non-Asian guests encountered more service failures in housekeeping (toilets, public areas, cleaning, and bedding) (Sann & Lai, 2020a).

However, the identification of attributes alone is not enough to support decision-making in the hotel industry. It is necessary to identify the sentiment of the reviewer when expressing their opinion about a certain attribute. The analysis of users' feelings can be used, for example, to classify the polarity of sentiments as positive, neutral, or negative, and allow the analysis of many documents as a corpus (Martin-Domingo et al., 2019).

Several studies, as identified in Table 1, use the Likert scale (ranging from 1: terrible to 5: excellent) defined on the platforms to determine a negative (dissatisfaction) or positive (satisfaction) review from the guest regarding the quality attributes of the service platforms (Lee et al., 2020; Sann & Lai, 2020a; Ying et al., 2020).

However, a more detailed analysis of the sentiments expressed in reviews may present more striking results (Alaei et al., 2019). This is explained by the fact that some reviewers who rated their overall experience as excellent also made negative comments (B. Kim et al., 2016).

By associating the identification of topics related to quality attributes with sentiment analysis techniques allows one to transform qualitative data into quantitative data, measure consumer attitudes toward products, and compare its position with that of competitors (Ma et al., 2018).

3. Methodology and development of the study

The methodology comprises data collection and pre-processing—including data cleaning and transformation, the modeling of topics to discover the attributes of hotel service quality, and sentiment analysis to detect the polarity of the reviews.

3.1. Data set

The municipality of Florianópolis-SC was selected for this study because it is one of the most visited tourist destinations in Brazil. It is a destination recognized worldwide for its natural beauty—it has more than 100 beaches—and for the quality of life it provides. Florianópolis-SC is also the Brazilian state capital with the highest score on the Human Development Index (HDI). Its economy is heavily based on information technology, tourism, and services.

Reviews from foreign tourists were not considered in this study. Brazil receives only 0.47% of foreign tourism, which represents about 6% of national tourism. Foreign tourists are mainly from Argentina, United States, Chile, Paraguay, Uruguay, France, Germany, Italy, United Kingdom, Spain, and Portugal (Brasil, 2020). Therefore, the low volume of UGC and its distribution in different languages made it impossible to apply any of the topic analysis tools.

The data for analysis was downloaded from the TripAdvisor website, which is the largest aggregator of reviews of tourist products in the world and is visited monthly by 460 million users. The TripAdvisor website contains 830 million reviews and opinions on 8.6 million accommodations, restaurants, attractions, tours, airlines, and cruises. TripAdvisor operates in 49 markets in 28 languages (Tripadvisor, 2020).

TripAdvisor was chosen as the data source for this study because it has a few advantages over other platforms. This platform has been widely used in previous scientific studies (Taecharungroj & Mathayomchan, 2019; Xiang et al., 2017), which makes it easier to compare results. TripAdvisor also shows data reliability, since the platform has maintained its reputation by policing the system to avoid false reviews (Taecharungroj & Mathayomchan, 2019). The use of TripAdvisor is

widespread in Brazil, where platforms such as Yelp and Ctrip are rarely used. Last, an earlier study proved that TripAdvisor has the best overall data quality, when compared to platforms such as Expedia and Yelp (Xiang et al., 2017).

The data were collected using a crawler developed by the authors in the Python programming language. As a result, a data set of 68,558 reviews provided between 2005 and 2019 was obtained, and included reviews of 922 hotels and inns listed on the platform. As this study focused on domestic tourists, only reviews written in Brazilian Portuguese were collected.

The dates of the evaluations and the ratings of each property—defined using metrics from the TripAdvisor platform and shown as several circles—were also collected. The circle scores were determined by user ratings on a scale of 1 to 5, where 3 was considered average, and 5 excellent. Some hotels had not yet been evaluated and were thus represented by zero. The distribution of data by year and by classification in circles is shown in Figure 1.

The first online reviews were recorded in 2005. However, since 2011, there has been a consistent growth in the number of reviews. This growth may be associated with the increasing use of online platforms by travelers and by the occurrence of the 2014 FIFA World Cup tournament and the 2016 Olympic Games (both events were hosted in Brazil).

3.2. Pre-processing

The text was pre-processed to organize, clean, and standardize the textual data to be subsequently used by the natural language processing and machine learning algorithms. The purpose of this procedure was to remove unnecessary content, thus reducing processing time and improving the performance of the models (Sarkar, 2019).

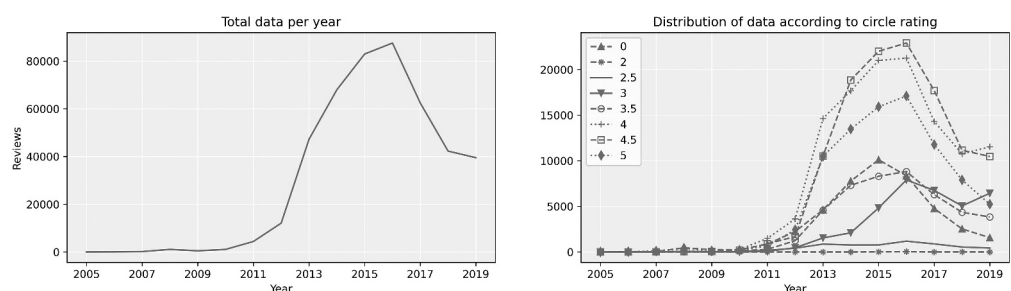
The techniques used for pre-processing the data set included removing special characters, converting capital letters to lower case, removing words without semantic value (stop words), and removing other unnecessary terms. Other techniques included applying stemming, which refers to the removal of affixes (prefixes and suffixes) from the words, and applying tokenization, which refers to the division of the text into words, sentences, or paragraphs, as necessary. These tasks were performed using the Python language and the Natural Language Toolkit (NLTK) library.

After conducting the initial tests, we decided to perform tokenization at a sentence level, as users often express conflicting feelings within the same review when addressing different aspects of their travel experience. Thus, judgment at a sentence level tends to increase the accuracy of the classifier. This process yielded a data set of 202,790 sentences for further analysis.

3.3. Topic modeling

The topics were modeled to discover the quality attributes of hotel services that influence customer satisfaction. For this, the Latent Dirichlet Allocation (LDA) method was used.

Figure 1. Distribution of data by year.



The LDA is an algorithm that belongs to the class of unsupervised methods for the automatic detection of topics in texts. It is based on the idea that every document includes multiple topics and it seeks to discover latent (hidden) patterns to understand relationships between documents and words. Thus, words present in related documents are grouped into topics (Blei et al., 2003). The LDA method was implemented using the Machine Learning for Language Toolkit (MALLET) package.

The data were then transformed into vectors for later inclusion in the algorithms using the Bag of Words (BoW) N-grams attribute engineering method. The BoW model represents each text document as a numerical vector in which each dimension is a specific word in the corpus and the vector value is represented by the number of times it occurs in the document. The adding of N-grams causes a fusion of words preceding and following the position of each word according to the N value (Sarkar, 2019).

In this study, the value two was used for the parameter N of the BoW N-grams, creating a bi-gram. Other values were tested but a loss of precision in the model was observed as the N value increased. The model was also adjusted to remove terms with fewer than 20 occurrences in all documents and terms occurring in more than 60% of documents, consequently reducing the total number of terms from 25,704 to 4,295 terms.

This practice increases the accuracy of the model. It reduces the computational effort required to run the model because unique or rare terms have a low semantic contribution to the model and because recurring terms interfere with the interpretation of the context (Sarkar, 2019).

The optimal number of topics was defined based on the coherence score. A coherence score is a measure of the degree of semantic similarity among high scoring words in each topic. To this end, models were generated with the number of topics ranging from 2 to 50. Figure 2 shows the variation in coherence scores according to the number of topics.

As shown in the Figure 2, the greatest coherence was observed in 14 topics. Therefore, this was the value used in this study. Thus, all sentences in the data set were classified into one of the 14 topics revealed by the LDA according to the highest probability of belonging to each of the topics.

3.4. Sentiment analysis

The polarity of the users' opinions expressed in the reviews was identified through sentiment analysis. This technique allowed us to determine whether the opinion expressed in each text carried positive sentiments (e.g., "Awesome bed, complete, and delicious breakfast," "Attentive and friendly service," and "The service is excellent, and the location is great"); or negative ones

Figure 2. Variation of model coherence according to the number of topics.

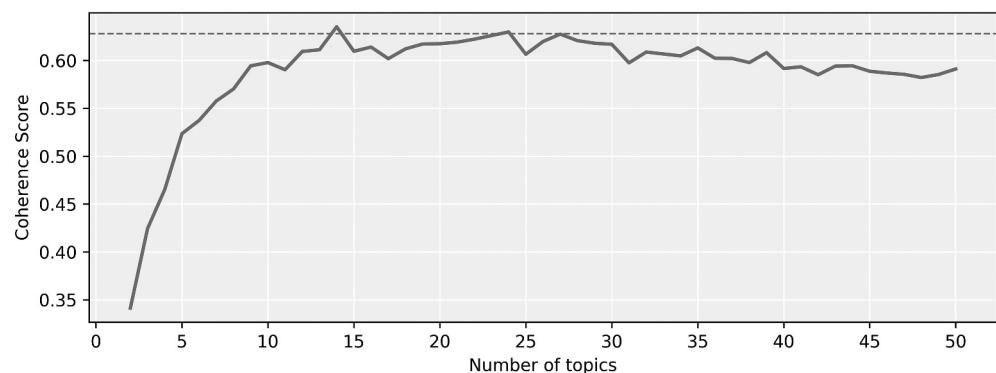


Table 2. Comparison of model accuracy

Model	Naive Bayes	Logistic Regression	SVM	Random Forest	Gradient Boosting
CV Score (BoW)	0.803642	0.781635	0.774397	0.75643	0.666955
Test Score (BoW)	0.8	0.791216	0.772973	0.754054	0.684459
CV Score (TF-IDF)	0.75413	0.773818	0.792065	0.761644	0.667537
Test Score (TF-IDF)	0.760135	0.787838	0.79527	0.77027	0.67973

(e.g., “Unbelievably bad breakfast,” “It is a bit tight and charges a little more than the competitors,” and “There was never anyone at the reception”).

In this study, feelings were classified using a supervised learning algorithm. The authors manually tagged a random sample of 5,000 sentences from the data set, resulting in 66.19% positive reviews, 23.72% negative reviews, and 10.09% neutral reviews. In the training stage of the model, the data set was divided, with 70% used for training and 30% for testing. Thus, algorithm learning took place in 70% of the data, and the level of error was measured in the other 30%.

The data were transformed into vectors for later inclusion in the algorithms using the following formats: BoW and the Term Frequency-Inverse Document Frequency (TF-IDF). The TF-IDF model assumes that if a word is important for a document, it must have many occurrences in the document and few in the rest of the data set (Sarkar, 2019).

Term frequency (TF) means the same as BoW, the number of occurrences of a word in a given document, while the inverse document frequency (IDF) is the number of occurrences of the word in the set of documents (D. Kim et al., 2019).

The criterion used for choosing the classification model was accuracy. For that criterion, we evaluated the Multinomial Naive Bayes, Logistic Regression, Support Vector Machines (SVM), Random Forest, and Gradient Boosting Machine models. The Scikit-Learn library available for machine learning in Python was used in this step.

The accuracy of the models was evaluated through cross-validation (CV) and application to a battery of tests. While performing CV, the data set of each classification model was divided into 10 subsets, and consecutive runs were performed by alternating the test subset in each run, to obtain the mean accuracy value (CV score).

The models were applied to the test set to evaluate the behavior of the model in a data set not yet analyzed by it, that is, the 30% of the data set previously reserved for testing the model. The results of the application of the models to the test set (Test Score) and CV (CV Score) are shown in Table 2.

After comparing the accuracy of the models, the classifier we selected for this study was the Naive Bayes with the BoW attribute engineering method, which showed an accuracy of 80.36% in the CV Score, and 80% in the Test Score, both superior to the other models.

This classifier was applied on the rest of the data set, resulting in the classification of 77.89% sentences as positive and 21.28% of sentences as negative. The results of the sentiment analysis are discussed below, together with the results of the analysis of topics performed with the LDA.

4. Results

In this study, 14 topics were extracted from the corpus using the LDA method. Table 3 shows the main terms of each topic. As each sentence was previously disposed in a topic by the LDA, the

Table 3. Topic modelling by LDA

Topic	% Docs*	Relevant Keywords
1	9.70	beach, stands, location, quiet, foot, go, island, place, access, path, where, far
2	8.69	room, clean, comfortable, bed, space, cosy, spacious, large, new, pleasant
3	9.23	service, staff, attentive, receptivity, personal, polite, helpful, friendly, cleanliness
4	6.28	parking, vacancies, little, lack, need, car, problems, payment, fee, distance
5	6.58	host, day, spent, stayed, there, stay, night, first, this, during, year, two, ness, travel
6	5.93	everywhere, everything, family, people, home, children, lots, makes, perfect
7	6.11	reception, problem, slow, check-in, check-out, reservation, early, late, arrive
8	7.52	pool, view, sea, wonder, front, beautiful, has, beautiful, heated, beauty, leisure
9	6.97	location, downtown, next, restaurant, stay, close, city, side, bar, main
10	5.90	bad, better, price, option, cost benefit, worth, daily, accommodation, work, value
11	7.32	recommend, come back, improve, others, right, stay, well, still, being, choice
12	4.97	bath, room, small, rain, water, towel, noise, outlet, door, inside, night, sleep
13	6.35	room, wi-fi, for, some, need, missing, old, air conditioning, bad, works, renovation
14	8.45	breakfast, simple, quality, options, service, tasty, reasonable, variety, varied

*Percentage of sentences by topic

importance of the topics can be determined by the percentage of sentences disposed in each topic, as shown in the second column of the [Table 3](#).

The topics were manually labeled based on a combination of human judgment of each topic's main terms, as shown in [Table 3](#), and the literature that described the quality attributes of hospitality services. Since topic 5, which corresponded to 6.23% of total stay-related reviews and group reviews, and topic 11, which corresponded to 7.32% of total reviews and group recommendations for future guests, did not have the characteristics of quality attributes, they were disregarded.

The results were validated by analyzing the words with the highest weightings disposed in each topic. In this analysis process, topics 1 and 9 were related to the "location" attribute, topics 2, 12 and 13 were related to the "room" attribute, and topics 6 and 8 were related to the "ambience" attribute. This procedure resulted in eight attributes, which are shown in [Table 4](#).

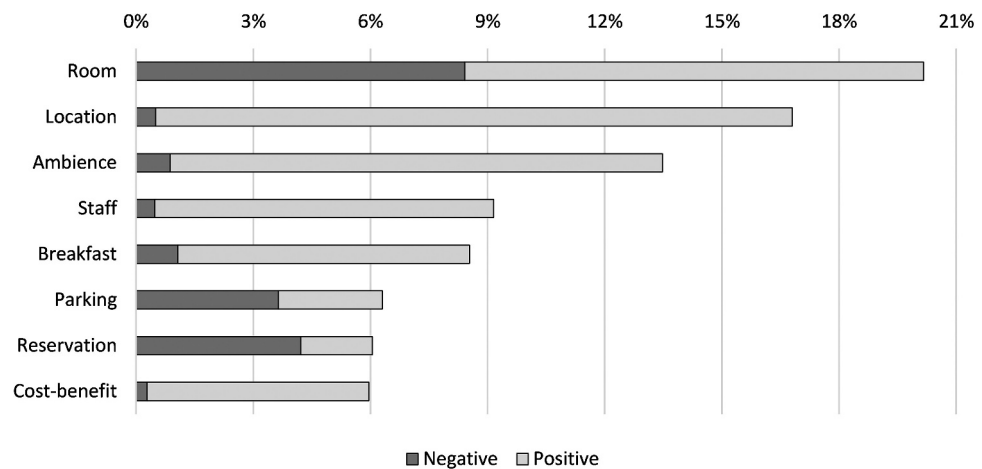
After defining the attributes, they were combined with the results of the sentiment analysis. The combined information allowed us to determine the polarity of feelings expressed by customers concerning the attributes of service quality. Thus, besides identifying the attributes, the proportion

Table 4. Attributes of hotel service quality in Florianópolis-SC

n*	% Docs**	Attributes	Description
1 e 9	16.67	Location	Ease of access and position in relation to tourist points.
2, 12 e 13	20.01	Room	Indoor room facilities.
3	9.23	Staff	Staff responsible for direct customer service.
4	6.28	Parking	Availability and space to park cars.
6 e 8	13.45	Ambiance	Related to subjective aspects of comfort.
7	6.11	Reservation	Scheduling and guest reception.
10	5.90	Cost-benefit	Ratio between the amount spent and the service received in return.
14	8.45	Breakfast	Quality and variety of menu items and of customer service

*Topic number; ** Percentage of sentences by attribute

Figure 3. Feelings expressed by customers in relation to each attribute of service quality.



of positivity and negativity ascribed to the attributes could be measured, based on the polarity attributed to each sentence (see Figure 3).

Based on the polarity of feelings expressed in the reviews, the quality attributes of hotel services were divided into two groups: a set of attributes with a predominance of negative reviews and another set with a predominance of positive reviews.

4.1. Quality attributes with a predominance of negative reviews

The attributes “room,” “parking,” and “reservation” had predominantly negative reviews, with 32.51% of the reviews being negative. Negative reviews about these attributes accounted for 16.28% of total reviews. The words used most frequently in the reviews and classified as negative in the sentiment analysis, are shown in Table 5.

The most frequently reviewed attribute was “room,” which accounted for 20.16% of total reviews and 41.76% of negative reviews. As shown in Table 5, the reviews referred to aspects such as: lack of space in the room, the poor condition of the furniture, the slow speed and intermittence of internet service, linen that was stained or in poor condition, poorly functioning doors, general uncleanliness, an uncomfortable bed, problems with the temperature and quality of the bath water, the lack of in-room furnishings such as a desk to support professional activities, and problems with the functioning of air conditioning, the television, and the minibar.

The attribute “parking,” which appeared in 6.31% of the reviews, received negative reviews in 66.89% of its occurrences, and received the highest proportion of negative reviews among the attributes. For this attribute, as shown in Table 4, customers reported problems related to space,

Table 5. Main terms in the topics related to each attribute with negative highlight

Attributes	Terms
Room	room (5674), bathroom (1996), cleaning (1154), bath (930), television (870), wi-fi (866), noise (866), old (836), bed (777), water (699), air conditioner (682), service (596), towel (562), mold (539), smell (521), maintenance (413), reform (403), box (402), port (380), size (353), walk (347), corridor (331), elevator (293), window(284), floor (276), mini fridge (271), walls (245), guests (224), mattress (198)
Parking	parking (1400), places (1258), pay (1237), price (829), daily (741), service (702), value (512), vehicle (198), valet (191), time (170), location (141), problem (135), street (111), outsourcing (66), reservation (61), structure (55)
Reservation	reception (2125), check-in (1862), reservation (1150), complaint (918), problem (855), price (602), late check-out (584), early check-in (569), wait (563), check- out (507), system (287), information (207), card (127), phone (127), email (125), bags (113), key (110)

Table 6. Main terms in the topics related to each attribute with a positive highlight

Attributes	Terms
Location	beaches (11,765), location (6768), restaurants (4578), central region (4141), tourist attractions (4058), shopping (2027), commerce (1822), access (1657), city (1617), bars (1121), supermarkets (951), tours (487), bus (480), pharmacies (405), road (398), airport (365), taxi (205), parking (167), conventions (106)
Ambiance	infrastructure (9237), view (5176), beauty (4333), nature (4000), perfection (3676), familiar (2648), recreation (2170), environment (2160), receptivity (1779), gastronomy (1376), suitable children (1366), comfort (766), cleaning (655), leisure (436), pleasant (386), decoration (290), rest (201), relaxing (172), tranquility (105)
Staff	staff (7240), attention (3023), customer service (2865), sympathy (1526), education (1129), helpfulness (784), owner (712), cordiality (676), kindness (505), efficiency (309), agility (168), receptivity (100), courtesy (89)
Breakfast	breakfast (6967), variety (3536), products (2473), quality (1731), flavor (1309), service (964), environment (869), simplicity (574), coffee (533), cleaning (341), buffet (237), price (218), freshness (176), kitchen (114)
Cost benefit	satisfaction (4198), cost benefit (865), quality (593), price (551), worth (222), prices (139), liked (85)

vehicle incidents, distance between the parking area and the hotel, time restrictions, exorbitant prices or prices undisclosed at the time of booking, and the lack of parking at certain establishments.

The third attribute with a predominance of negative reviews was “reservation,” which received 6.05% of total reviews, 69.7% of which were negative. As shown in [Table 5](#), in the “reservation” attribute, customers reported problems such as scheduling errors, incompatibility of room specifications with those displayed on online platforms, wait time, inflexibility with schedules, and lack of receptivity from frontline staff.

4.2. Quality attributes with a predominance of positive reviews

The attributes “location,” “ambiance,” “staff,” “breakfast,” and “cost-benefit” received a predominance of positive reviews. These attributes accounted for 47.97% of the reviews, and of this total, 45.05% were positive and only 2.92% were negative. The words used most frequently in reviews and classified as negative in the sentiment analysis, are shown in [Table 6](#).

In this group, the most frequently reviewed attribute was the “location” of the hotel relative to the places of interest to the traveler, such as restaurants, bars, beaches, the city’s shopping district, the airport, tourist attractions, and event venues, among others, as shown in [Table 6](#). This attribute was frequently mentioned in the reviews of hotels in Florianópolis-SC, with 16.8% of the reviews including this attribute, of which 97.03% were positive.

Next, the attribute “ambiance” was mentioned in 13.48% of the reviews. It was associated with subjective aspects, such as tranquility, hospitality, receptivity, rest, and the views from the property. So, a good ambiance makes a space more welcoming and conducive to socializing, as shown in [Table 6](#). This attribute received 93.54% positive reviews.

Another attribute with a predominance of positive reviews was “staff,” which was mentioned in 9.15% of the reviews. It was associated with aspects such as service, attention, reception, promptness, friendliness, education, helpfulness, receptivity, kindness, and friendliness, as shown in [Table 6](#). This attribute received 94.74% positive reviews.

The attribute “breakfast” was also associated with positive evaluations by travelers. It was mainly associated with the variety, availability, and flavor of products such as coffee, fruits, and

Figure 4. Longitudinal distribution of attributes with a predominance of negative reviews according to the classification in circles.

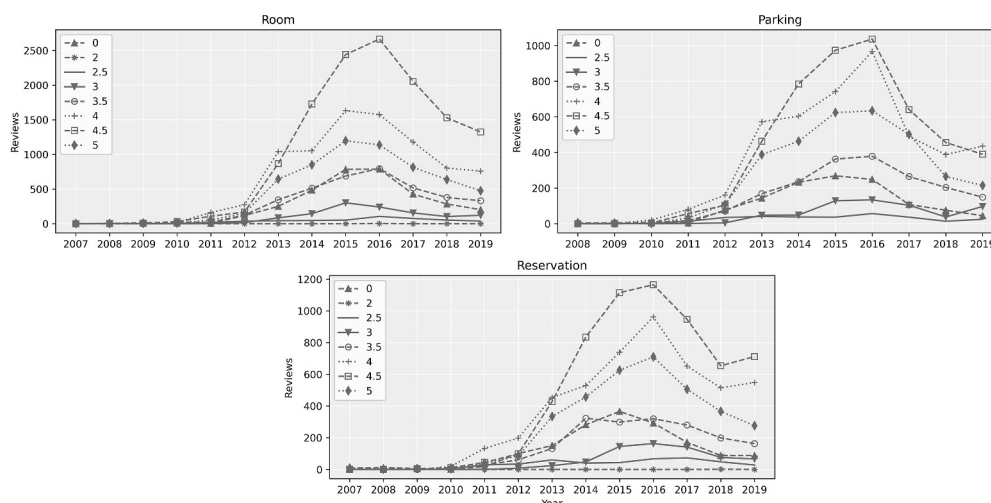
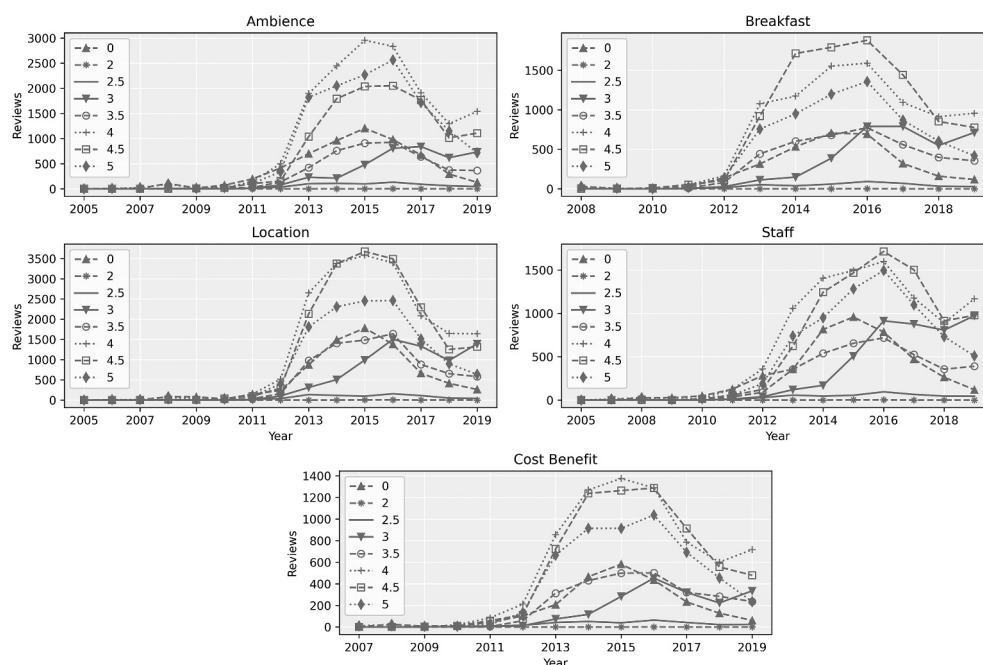


Figure 5. Longitudinal distribution of attributes with a predominance of positive reviews according to the classification in circles.



cakes, and the quality of customer service, as shown in Table 6. This attribute was reviewed in 8.54% of total reviews and received 87.5% positive reviews.

Finally, “cost-benefit” ratio was considered, and accounted for 5.96% of the reviews. This attribute was related to customers’ perception of the ratio between money spent and service received, as shown in Tables 6; 95.3% of the reviews for this attribute were positive.

5. Discussion

Priority actions should be taken on attribute sets with a predominance of negative reviews, because negative content on travel platforms directly influences the purchase decision, and directly affects hotels’ revenue. Around 35% of travelers change their hotel booking decisions

after browsing social media, 53% say they do not book a hotel that has no reviews to view, and 87% say that reviews make them feel more confident when deciding to purchase accommodation (Nicoli & Papadopoulou, 2017).

The attribute related to the internal environment of the “room” is the one that managers should give the most attention, besides being most frequently reviewed by travelers, it is the one with the highest percentage of negative reviews. As shown in Figure 4, hotels with 4.5 circles received the highest proportion of negative reviews for this attribute. A longitudinal plot of the data also shows that there are no significant changes in the proportion of the categories over the period evaluated.

The attribute “room” was presented in most of the studies analyzed during article reviews, as shown in Table 1. Many of these studies identified this attribute as a positive comment generator (Alrawadieh & Law, 2019; Lee et al., 2020; Sann & Lai, 2020a), and linked it to customer satisfaction (Herjanto et al., 2017; Liu et al., 2017; Padma & Ahn, 2020; Wu et al., 2017). On the other hand, there were also studies that corroborated the results of this research, by observing the attribute “room” as a negative comment generator (Hu et al., 2019; Köseoglu et al., 2019).

According to the most commented words of each topic, according Table 5, actions to address this attribute should include routine protocols for checking the functioning of equipment and furniture, so that customers are not confronted by malfunctions. There should be a protocol for cleaning and replacing items that are in a poor condition. Criteria should be established to evaluate the adequacy of the furniture and the layout of the guest room to ensure free movement in the environment. Additionally, the room’s layout should also support work-related activities, a rising trend in current times.

The “parking” attribute attracted similar reviews, as shown in Figure 4, with no changes in the order of categories over the period. This attribute appeared to be more important for guests in the area covered by this study, as it was not frequently found in related studies. When observed in other studies, this attribute was identified as a negative review generator (Fernandes & Fernandes, 2017; Xu & Li, 2016). The great number of reviews related to “parking” was another remarkable finding of the present study.

Actions related to this attribute should include offering adequate parking proportional to the demand and adapt parking spaces, while considering the technical criteria of architecture. The distance traveled by a guest is a critical factor and should be reduced whenever possible. Service personnel should be trained to serve the customer and quickly resolve potential problems.

The attribute “reservation” received the same proportion of reviews throughout the period, as shown in Figure 4. Longitudinally, this attribute was found in a higher proportion in hotels that received 4.5 circles, followed by those that received 4 and 5 circles. The attribute “reservation” also received the highest proportion of negative comments, with 69.7% of total reviews regarding this attribute. These findings confirm the results of previous studies (Hu et al., 2019; Sann & Lai, 2020b).

Actions on this attribute should include service staff training, avoiding errors, and improving customer treatment. Hotels should invest in technology to integrate internal software with external scheduling and billing platforms. Hotels should also examine the adequacy of their marketing policies for greater transparency to the customer to avoid creating disproportionate expectations of the actual service delivered. They should also adopt a clear policy regarding early check-ins and late check-outs, with flexible hours, and with no additional charges whenever possible.

Figure 5 shows the distribution of the attributes with a predominance of positive reviews according to the classification using circles. Organizations with a low performance in these

attributes could use these results to improve the quality of their relationship with customers by improving valued attributes and thus add value to their product. Organizations with a good performance record can enhance positive features to generate customer loyalty.

Positive ratings for the “ambience” attribute predominate in hotels rated with 4 circles, followed by hotels rated with 5 and 4.5 circles. Other studies have also identified this attribute related to positive comments on ecotourism experiences (Brochado & Brochado, 2019), and related to negative hotel reviews (Fernandes & Fernandes, 2017).

Low-performing organizations in this attribute can consider reviewing the infrastructure, improve aspects such as views from the hotel, interior decoration, comfort, and contact with nature. Actions to improve this attribute may include a discussion between users and workers to adapt architectural features. Moreover, actions can be taken to improve cleaning services, food, and recreation, and making the environment more familiar and suitable for children.

Reviews on the attributes “breakfast” and “staff” were more frequent among hotels that received 4.5 circles, followed by those that received 4 and 5 circles. Regarding the “location” attribute, hotels that received 4 and 4.5 circles predominated and had similar values, followed by hotels that received 5 circles. An important observation in this set of positive attributes is the growth in the number of hotel reviews receiving 3 and 3.5 circles, indicating that there may be significant improvements in these attributes in these categories of hotels.

Previous studies have also identified a significant occurrence of the attribute “breakfast” in reviews (Köseoglu et al., 2019; Xu & Li, 2016). As in this study, the “breakfast” attribute was linked to positive reviews (B. Kim et al., 2016). However, its greatest occurrence was related to customer satisfaction (Alrawadieh & Law, 2019; Padma & Ahn, 2020; Sann & Lai, 2020a), and services flaws (Fernandes & Fernandes, 2017; Sann & Lai, 2020b).

Actions to improve the attribute “breakfast” may include standardizing the menu, adapting it to dietary restrictions, and using opinion polls to discover what products suit their guests’ tastes. Quality programs to ensure the availability of products throughout attendance and adequate services, with extensions to suppliers, to ensure a standard flavor and punctuality in delivery.

This study’s results regarding the attribute “staff” were supported by previous studies. This attribute was also identified as an important subject by several studies (B. Kim et al., 2016; Köseoglu et al., 2019; Padma & Ahn, 2020; Sann & Lai, 2020a; Xu & Li, 2016; Ying et al., 2020) and, as in this paper, its occurrence was mostly associated with positive reviews (Alrawadieh & Law, 2019; Brochado & Brochado, 2019; Wu et al., 2017).

Regarding the attribute “staff,” it is important to select employees according to the hotel’s target audience and to maintain a training program that aids in adapting to new technologies and valuing employees. Customers consider and value features such as friendliness, education, kindness, warmth, efficiency, agility, and courtesy. Another critical factor is the standardization of processes, which allows defining metrics to evaluate the results.

The attribute “cost-benefit,” linked to an overview of the costumers’ perception of the ratio of costs and services received, has not been directly pointed out in previous studies. Despite being related to a high percentage of positive comments, organizations with negative ratings on the “cost-benefit” attribute should consider revising their pricing policy. Actions in this attribute should consider hotel occupancy rates, which is an element outside the scope of this study.

Finally, the “location” of the hotel is related to the proximity of places of interest for travelers to visit. Most of the previous studies confirmed the results of this research when associating the attribute “location” with a higher volume of positive comments or customer satisfaction (Alrawadieh & Law, 2019; Hu et al., 2019; B. Kim et al., 2016; Sann & Lai, 2020a; Wu et al., 2017).

Hotels with poor performance in this attribute could develop collaboration programs with other establishments and tourism agents. Such programs would allow services to be offered in the form of packages, with the possibility of monitoring the quality of the services offered to the traveler throughout the travel experience and correcting identified deficiencies.

6. Conclusion

This study has explored the UGC in the hotel sector of the city Florianópolis-SC, Brazil to identify the quality of attributes and determine the polarity of feelings based on reviews. For this purpose, a combination of analysis methods of topics and a sentiment analysis was used.

Through the topic analysis it was possible to establish that “room,” “location,” “ambience,” “staff,” “breakfast,” “parking,” “booking,” and “cost-benefit” were the main attributes of the service quality of the hotels evaluated by customers.

The sentiment analysis made it possible to identify that “room,” “parking,” and “reservation” were the attributes that most frequently generated negative reviews. On the other hand, the findings also indicated that “location,” “ambience,” “staff,” “breakfast,” and “cost-benefit” were the attributes that most frequently generated positive reviews from customers.

The results support the initial assumption about the existence of a variation in the level of relevance of attributes given by customers, based on the environmental and cultural context in question. In this study the attribute “room” stands out, which contradicts the findings of previous studies by being associated with a high number of negative comments. The attribute “parking” also stands out by not presenting the same level of relevance in other studies.

The results of the study support previous claims that UGC is a useful source of content for identifying the attributes of quality. The methodological procedure for allocating reviews to topics identified by the LDA and classifying guest feedback as positive and negative made it possible to transform the poorly structured data of customer reviews into a quantitative picture of quality attributes, allowing for replication in other studies. Finally, the results contribute to the field of quality evaluation of hospitality services in the Portuguese language, where UGC studies are still scarce.

6.1. Limitations

This study has some limitations. For instance, it only addressed attributes considered by domestic tourists, in Portuguese, and for only one Brazilian tourist destination. Further research could expand this scope by addressing the content generated by foreign tourists and consider tourist destinations in other Brazilian regions.

The data were obtained from a platform where the reviews refer to attributes of hotels. However, aspects outside the control of hotels too can influence the decision to visit or return to a tourist destination. Thus, future research should address this issue using the available content and other platforms related to tourism products, and include reviews on social networks such as Twitter, Instagram, Facebook, among others.

Last, this study does not relate the polarity of the expressed feelings to the reviews, or regard the quality attributes of the services, or general customer satisfaction. Further research could introduce this approach, including a correlation with the traveler profile.

Funding

The authors received no direct funding for this research.

Author details

Clérito Kaveski Peres
 E-mail: cleritokp@gmail.com
 ORCID ID: <http://orcid.org/0000-0002-5643-6888>
 Edson Pacheco Paladini
 ORCID ID: <http://orcid.org/0000-0002-8651-0970>
 Department of Industrial and Systems Engineering,
 Federal University of Santa Catarina, Florianópolis, Brazil.

Citation information

Cite this article as: Exploring the attributes of hotel service quality in Florianópolis-SC, Brazil: An analysis of TripAdvisor reviews, Clérito Kaveski Peres & Edson Pacheco Paladini, *Cogent Business & Management* (2021), 8: 1926211.

References

- Alaei, A. R., Becken, S., & Stantic, B. (2019). Sentiment analysis in tourism: capitalizing on big data. *Journal of Travel Research*, 58(2), 175–191. <https://doi.org/10.1177/0047287517747753>
- Alrawadie, Z., & Law, R. (2019). Determinants of hotel guests satisfaction from the perspective of online hotel reviewers. *International Journal of Culture Tourism and Hospitality Research*, 13(1), 84–97. <https://doi.org/10.1108/IJCTHR-08-2018-0104>
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, 3 (4–5), 993–1022. <https://doi.org/10.1016/b978-0-12-411519-4.00006-9>
- Brasil. Ministério do Turismo. (2018). Plano nacional de turismo - 2018–2022: mais emprego e renda para o Brasil. <https://cultura.rs.gov.br/upload/arquivos/carga20180322/28162245-pnt-2018-2022.pdf>
- Brasil. Ministério do Turismo. (2020). Anuário Estatístico de Turismo 2020 - Ano Base 2019–2ª Edição (Vol. 47, Issue 2). <http://www.dadosefatos.turismo.gov.br/2016-02-04-11-54-03/demanda-turistica-internacional.html>
- Brochado, A., & Brochado, F. (2019). What makes a glamping experience great? *Journal of Hospitality and Tourism Technology*, 10(1), 15–27. <https://doi.org/10.1108/JHTT-06-2017-0039>
- Cronin, J., . J., & Taylor, S., . A. (1992). Measuring service quality: A re-examination and extension. *Journal of Marketing*, 56(3), 55–68. <https://doi.org/10.1177/002224299205600304>
- Dogan, S., Basaran, M. A., & Kantarci, K. (2020). Determination of attributes affecting price-performance using fuzzy rule-based systems: Online ratings of hotels by travel 2.0 users. *Journal of Hospitality and Tourism Technology*, 11(2), 291–311. <https://doi.org/10.1108/JHTT-07-2018-0067>
- Fernandes, T., & Fernandes, F. (2017). Sharing dissatisfaction online: analyzing the nature and predictors of hotel guests negative reviews. *Journal of Hospitality Marketing & Management*, 27(2), 127–150. <https://doi.org/10.1080/19368623.2017.1337540>
- Foris, D., Crialmean, N., & Foris, T. (2020). Exploring the environmental practices in hospitality through booking websites and online tourist reviews. *Sustainability*, 12(24), 1–18. <https://doi.org/10.3390/su122410282>
- Guo, Y., Barnes, S. J., & Jia, Q. (2017). Mining meaning from online ratings and reviews: tourist satisfaction analysis using latent dirichlet allocation. *Tourism Management*, 59, 467–483. <https://doi.org/10.1016/j.tourman.2016.09.009>
- Ha, E. Y., & Lee, H. (2018). Projecting service quality: the effects of social media reviews on service perception. *International Journal of Hospitality Management*, 69 (September 2017), 132–141. <https://doi.org/10.1016/j.ijhm.2017.09.006>
- Haghighi, N. N., Liu, X. C., Wei, R., Li, W., & Shao, H. (2018). Using twitter data for transit performance assessment: A framework for evaluating transit riders' opinions about quality of service. *Public Transport*, 10(2), 363–377. <https://doi.org/10.1007/s12469-018-0184-4>
- He, W., Tian, X., Hung, A., Akula, V., & Zhang, W. (2018). Measuring and comparing service quality metrics through social media analytics: A case study. *Information Systems and E-Business Management*, 16 (3), 579–600. <https://doi.org/10.1007/s10257-017-0360-0>
- Herjanto, H., Erickson, E., & Calleja, N. F. (2017). Antecedents of business travelers' satisfaction. *Journal of Hospitality Marketing & Management*, 26 (3), 259–275. <https://doi.org/10.1080/19368623.2017.1234954>
- Hu, N., Zhang, T., Gao, B., & Bose, I. (2019). What do hotel customers complain about? text analysis using structural topic model. *Tourism Management*, 72, 417–426. <https://doi.org/10.1016/j.tourman.2019.01.002>
- Ju, Y., Back, K. J., Choi, Y., & Lee, J. S. (2019). Exploring airbnb service quality attributes and their asymmetric effects on customer satisfaction. *International Journal of Hospitality Management*, 77(July 2018), 342–352. <https://doi.org/10.1016/j.ijhm.2018.07.014>
- Kim, B., Kim, S., & Heo, C. Y. (2016). Analysis of satisfiers and dissatisfiers in online hotel reviews on social media. *International Journal of Contemporary Hospitality Management*, 28(9), 1915–1936. <https://doi.org/10.1108/IJCHM-04-2015-0177>
- Kim, D., Seo, D., Cho, S., & Kang, P. (2019). Multi-co-training for document classification using various document representations: TF-IDF, LDA, and Doc2Vec. *Information Sciences*, 477, 15–29. <https://doi.org/10.1016/j.ins.2018.10.006>
- Köseoglu, M. A., Mehraliyev, F., Altin, M., & Okumus, F. (2019). Competitor intelligence and analysis (CIA) model and online reviews: Integrating big data text mining with network analysis for strategic analysis. *Tourism Review*, ahead-of-print(ahead-of-print). <https://doi.org/10.1108/TR-10-2019-0406>
- Lee, M., Cai, Y. J., DeFranco, A., & Lee, J. (2020). Exploring influential factors affecting guest satisfaction big data and business analytics in consumer-generated reviews. *Journal of Hospitality and Tourism Technology*, 11(1), 137–153. <https://doi.org/10.1108/JHTT-07-2018-0054>
- Li, H., Ye, Q., & Law, R. (2013). Determinants of customer satisfaction in the hotel industry: an application of online review analysis. *Asia Pacific Journal of Tourism Research*, 18(7), 784–802. <https://doi.org/10.1080/10941665.2012.708351>
- Liu, Y., Teichert, T., Rossi, M., Li, H., & Hu, F. (2017). Big data for big insights: investigating language-specific drivers of hotel satisfaction with 412,784 user-generated reviews. *Tourism Management*, 59, 554–563. <https://doi.org/10.1016/j.tourman.2016.08.012>
- Ma, E., Cheng, M., & Hsiao, A. (2018). Sentiment analysis – A review and agenda for future research in hospitality contexts. *International Journal of Contemporary Hospitality Management*, 30(11), 3287–3308. <https://doi.org/10.1108/IJCHM-10-2017-0704>
- Malik, S. A., Akhtar, F., Raziq, M. M., & Ahmad, M. (2020). Measuring service quality perceptions of customers in the hotel industry of Pakistan. *Total Quality Management and Business Excellence*, 31(3–4),

- 263–278. <https://doi.org/10.1080/14783363.2018.1426451>
- Martin, D. W. (1995). An importance/performance analysis of service providers' perception of quality service in the hotel industry. *Journal of Hospitality & Leisure Marketing*, 3(1), 5–17. https://doi.org/10.1300/J150v03n01_02
- Martin-Domingo, L., Martin, J. C., & Mandsberg, G. (2019). Social media as a resource for sentiment analysis of airport service quality (ASQ). *Journal of Air Transport Management*, 78, 106–115. <https://doi.org/10.1016/j.jairtraman.2019.01.004>
- Nicoli, N., & Papadopoulou, E. (2017). TripAdvisor and reputation: A case study of the hotel industry in Cyprus. *EuroMed Journal of Business*, 12(3), 316–334. <https://doi.org/10.1108/EMJB-11-2016-0031>
- Nisar, T. M., & Prabhakar, G. (2018). Trains and twitter: firm generated content, consumer relationship management and message framing. *Transportation Research Part A: Policy and Practice*, 113, 318–334. <https://doi.org/10.1016/j.tra.2018.04.026>
- Padma, P., & Ahn, J. (2020). Guest satisfaction & dissatisfaction in luxury hotels: an application of big data. *International Journal of Hospitality Management*, 84, 1–8. <https://doi.org/10.1016/j.ijhm.2019.102318>
- Parasuraman, A., Zeithaml, V. A., & Berry, L. L. (1985). A conceptual model of service quality and its implications for future research. *Journal of Marketing*, 49(4), 41–50. <https://doi.org/10.1177/002224298504900403>
- Phillips, P., Barnes, S., Zigan, K., & Schegg, R. (2017). Understanding the impact of online reviews on hotel performance: an empirical analysis. *Journal of Travel Research*, 56(2), 235–249. <https://doi.org/10.1177/0047287516636481>
- Rios-Martin, M. A., Folgado-Fernandez, J. A., Palos-Sanchez, P. R., & Castejon-Jimenez, P. (2020). The impact of the environmental quality of online feedback and satisfaction when exploring the critical factors for luxury hotels. *Sustainability*, 12(1), 1–24. <https://doi.org/10.3390/su12010299>
- Sanchez-Franco, M. J., Cepeda-Carrion, G., & Roldán, J. L. (2019). Understanding relationship quality in hospitality services: A study based on text analytics and partial least squares. *Internet Research*, 29(3), 478–503. <https://doi.org/10.1108/IntR-12-2017-0531>
- Sann, R., & Lai, P.-C. (2020a). Do expectations towards thai hospitality differ? the views of english vs Chinese speaking travelers. *International Journal of Culture Tourism and Hospitality Research*, 15(1), 43–58. <https://doi.org/10.1108/IJCTHR-01-2020-0010>
- Sann, R., & Lai, P.-C. (2020b). Understanding homophily of service failure within the hotel guest cycle: applying NLP-aspect-based sentiment analysis to the hospitality industry. *International Journal of Hospitality Management*, 91, 1–13. <https://doi.org/10.1016/j.ijhm.2020.102678>
- Sann, R., Lai, P.-C., & Liaw, S.-Y. (2020). Online complaining behavior: Does cultural background and hotel class matter?. *Journal of Hospitality and Tourism Management*, 43, 80–90. <https://doi.org/10.1016/j.jhtm.2020.02.004>
- Sarkar, D. (2019). *Text Analytics with Python: A Practitioner's Guide to Natural Language Processing*. Apress. <https://doi.org/10.1007/978-1-4842-4354-1>
- Sun, S., Tong, K. T., & Law, R. (2017). Chinese hotel guest perception of international chain hotels under the same hotel brand in different travel destinations: the cases of intercontinental and Sheraton. *Journal of Vacation Marketing*, 23(2), 172–188. <https://doi.org/10.1177/1356766715614344>
- Taecharungroj, V., & Mathayomchan, B. (2019). Analysing TripAdvisor reviews of tourist attractions in Phuket, Thailand. *Tourism Management*, 75, 550–568. <https://doi.org/10.1016/j.tourman.2019.06.020>
- Tontini, G., Bento, G. D., Milbratz, T. C., Volles, B. K., & Ferrari, D. (2017). Exploring the nonlinear impact of critical incidents on customers' general evaluation of hospitality services. *International Journal of Hospitality Management*, 66, 106–116. <https://doi.org/10.1016/j.ijhm.2017.07.011>
- Tran, T., Ba, H., & Huynh, V. N. (2019). Measuring hotel review sentiment: an aspect-based sentiment analysis approach. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 11471(LNAI), 393–405. https://doi.org/10.1007/978-3-030-14815-7_33
- Tripadvisor. (2020). UK press center/about tripAdvisor. <https://tripadvisor.mediaroom.com/uk-about-us>
- WTTC. World Travel & Tourism Council. (2018a). *Domestic tourism: Importance and economic impact*. In World Travel & Tourism Council. <https://www.wttc.org/publications/2018/domestic-tourism/>
- WTTC. World Travel & Tourism Council. (2018b). *Travel & tourism: power and performance*. <https://www.wttc.org/-/media/files/reports/2018/power-and-performance-rankings-2018.pdf>
- Wu, M.-Y., Pearce, P., & Dong, W. (2017). How satisfying are Shanghai's superior hotels? the views of international tourists. *International Journal of Contemporary Hospitality Management*, 29(4), 1096–1115. <https://doi.org/10.1108/IJCHM-01-2015-0014>
- Xiang, Z., Du, Q., Ma, Y., & Fan, W. (2017). A comparative analysis of major online review platforms: implications for social media analytics in hospitality and tourism. *Tourism Management*, 58, 51–65. <https://doi.org/10.1016/j.tourman.2016.10.001>
- Xu, X. (2018). Does traveler satisfaction differ in various travel group compositions? evidence from online reviews. *International Journal of Contemporary Hospitality Management*, 30(3), 1663–1685. <https://doi.org/10.1108/IJCHM-03-2017-0171>
- Xu, X., & Li, Y. (2016). The antecedents of customer satisfaction and dissatisfaction toward various types of hotels: A text mining approach. *International Journal of Hospitality Management*, 55, 57–69. <https://doi.org/10.1016/j.ijhm.2016.03.003>
- Ying, S., Chan, J. H., & Qi, X. (2020). Why are Chinese and North American guests satisfied or dissatisfied with hotels? an application of big data analysis. *International Journal of Contemporary Hospitality Management*, 32(10), 3249–3269. <https://doi.org/10.1108/IJCHM-02-2020-0129>



© 2021 The Author(s). This open access article is distributed under a Creative Commons Attribution (CC-BY) 4.0 license.

You are free to:

Share — copy and redistribute the material in any medium or format.

Adapt — remix, transform, and build upon the material for any purpose, even commercially.

The licensor cannot revoke these freedoms as long as you follow the license terms.

Under the following terms:

Attribution — You must give appropriate credit, provide a link to the license, and indicate if changes were made.

You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.

No additional restrictions

You may not apply legal terms or technological measures that legally restrict others from doing anything the license permits.



***Cogent Business & Management* (ISSN: 2331-1975) is published by Cogent OA, part of Taylor & Francis Group.**

Publishing with Cogent OA ensures:

- Immediate, universal access to your article on publication
- High visibility and discoverability via the Cogent OA website as well as Taylor & Francis Online
- Download and citation statistics for your article
- Rapid online publication
- Input from, and dialog with, expert editors and editorial boards
- Retention of full copyright of your article
- Guaranteed legacy preservation of your article
- Discounts and waivers for authors in developing regions

Submit your manuscript to a Cogent OA journal at www.CogentOA.com

