New Insights into Hotel Customers' Revisiting Intentions, Based on Big Data

Journal:	*International Journal of Contemporary Hospitality Management*
Manuscript ID	IJCHM-06-2022-0719.R3
Manuscript Type:	Original Article
Keywords:	Revisit Intention, Market segmentation, Marketing strategy, Marketing Mix, Feature Extraction, User Generated Content

New Insights into Hotel Customers' Revisiting Intentions, Based on Big Data

Abstract

Purpose: This research employs big data analysis and sheds light on key hotel features that play a role in the revisit intention of customers. In addition, we endeavor to highlight hotel features for different customer segments.

Design/Methodology/Approach: We employ a machine learning method and analyze around 100,000 reviews of customers of one hundred selected hotels around the world where they had indicated on Trip Advisor their intention to return to a particular hotel. The important features of the hotels are then extracted in terms of the 7Ps of the marketing mix. We have then segmented customers intending to revisit hotels, based on the similarities in their reviews.

Findings: 71 important hotel features are extracted using text analysis of comments. The most important features are the room, staff, food, and accessibility. Also, customers are segmented into fifteen groups, and key hotel features important for each segment are highlighted.

Originality/Value: By employing text mining analysis, we identify and classify important hotel features that are crucial for the revisit intention of customers based on the 7Ps. Methodologically, we suggest a comprehensive method to describe the revisit intention of hotel customers based on customer reviews.

Practical implications: This study highlights key hotel features that are crucial for customers' revisit intention and identifies related market segments that can support managers in better designing their strategies and allocating their resources.

Limitations: In this research, the number of repetitions of words was employed to identify key hotel features, while sentence-based analysis or group analysis of adjacent words can be employed.

Keywords

Revisit Intention, Segmentation, Marketing Strategy, Marketing Mix, Feature Extraction, User-Generated Content

1 Introduction

Consumer-generated data on social media and review websites have grown rapidly (Ahani et al., 2019). Many individuals visit these websites to write a review or select and book a hotel based on existing reviews (Filieri et al., 2021a). According to statistics, the number of reviews has grown and on TripAdvisor.com alone, there are 411 million monthly visitors who search in about 700 million user-generated comments about hotels (Filieri et al., 2021a). With a proliferation of these unstructured text data, new methodological approaches are necessary (Xiang et al., 2015) to better understand the behavior of consumers and provide more reliable and valid criteria for the decision-making of hotel industry managers (Ahmad and Sun, 2018).

In this research, we endeavor to analyze the big data generated by customers' reviews to provide new insights about revisit intention, which is one of the most important topics in the hotel industry (Peng et al., 2015) as it can considerably improve hotels' financial performance (Jang and Feng, 2007). Research shows that higher repurchase intention can attract existing customers at a lower cost compared to new customers; according to statistics, a 5% increase in customers' return rate can lead to a 25% to 85% growth in profitability (Jang and Feng, 2007).

A plethora of previous studies (Liu and Beldona, 2021, Wu et al., 2021) has attempted to untangle factors contributing to hotel revisit intention. However, the majority of existing research has relied on conventional methods such as qualitative interviews. Only recently, a few studies have employed a big data analysis approach to provide new insight into the revisit intention of customers. For example, Park et al. (2020) attempted to predict customer revisit intention behaviors according to online hotel comments and found relationships between the structure and sentiment of comments with revisit intention. Similarly, Liu and Beldona (2021) examined different machine learning methods to better classify hotel customers' comments into re-visitors and non-re-visitors.

According to technology experts, customer experiences have been changed by the evolving use of new technologies such as social media and mobile technologies (Zhang et al., 2017). These changes have led to an urgent need to implement technology-based approaches such as big data analysis in hospitality management to detect real-time opportunities in the industry and optimize processes and decision-making in hotel management (Cheng et al., 2023). This research builds on these developments and employs online customer reviews which have been shown to provide some of the most important content generated by consumers (Zarezadeh et al., 2022). Online reviews impact the behavior and decisions of customers (Bortoluzzi et al., 2020). They allow customers to gain new insights about a hotel and consequently play a pivotal role in encouraging customers to visit a hotel (Lo and Yao, 2019). Prior studies have corroborated that analysis of the big data generated by customers can provide new and valuable insights and support managers in better designing their strategies (Ranjbari et al., 2020, Ying et al., 2020).

We have selected hotel features as the best descriptors of the behavioral intention to return (Um et al., 2006). Many studies have employed traditional survey methods and have collected a small number of responses from one country to identify key hotel features suggested by travelers. They have shed light on some aspects such as location, price, facilities and cleanliness as important hotel features (Lockyer, 2005). We complement this strand of literature by not only collecting large numbers of reviews from the TripAdvisor website but also collecting customer reviews on the most popular hotels in several countries around the world.

In addition, we provide simultaneous analysis of these features which can help to identify the relationships between these features and to learn about important segments of customers based on special features (Francesco and Roberta, 2019). Market segmentation highlights that hotel customers are not the same and have different needs and desires (Hajibaba et al., 2020). Data-driven market segmentation has proved to be a necessary method to provide better market insights (Ernst and Dolnicar, 2018). However, there is scant research on market segmentation based on big data (Han et al., 2021). One of the few studies on this area is related to the work of Ahani et al. (2019) who employed machine learning techniques to segment the customers of spa hotels.

While the research of Ahani et al. (2019) contributes to the development of the literature, their focus is only on spa hotel features and related market segments. In this research, we expand their findings by not limiting hotels to spa hotels, and by providing a more comprehensive understanding of key hotel features and related market segmentations. In addition, in order to segment markets, we employ a different perspective, a marketing mix approach, to classify the key features of hotels and segment customers that are willing to revisit hotels. A marketing mix has been used to regulate and manage marketing activities and has seven sub-categories of *product*, *price*, *promotion*, *place*, *people*, *physical evidence* and *processes* (Wilson et al., 2016).

To summarize, in this research, we aim to answer the following two questions: 1) What are the most important features of a hotel that shape customers' revisit intention? 2) For different segments of hotel customers, which category of hotel features together leads to the desire to return to the hotel? To answer the research questions, we have collected 98,201 customer reviews from the top 100 hotel destinations around the world. We have utilized machine learning text analysis to analyze the data.

Our findings contribute to the hospitality and marketing literature in several ways. First, we employ machine learning methodology to analyze big data and introduce a new approach to developing marketing strategies based on user-generated content. In contrast to prior studies, our analysis of the reviews provides a more reliable and comprehensive understanding of the intended behavior of customers. Second, we highlight the key hotel features that are critical to customers' decisions to intend to return to a hotel in more detail, compared to previous studies. Moreover, we explain that even considering these key features, they are not equally important for all types of customers. Instead, hotel managers should pay particular attention to the market segments that they want to serve and develop and emphasize those features for their targeted customers.

2 Literature review

2.1 Digital technology, big data, and automated text analysis in the hotel industry

Digital technologies have enabled hotels to reach higher levels of efficiency, better organizational performance and co-creation with customers (Buhalis and Leung, 2018). Internet-based systems are one of the most important technologies that shape this industry (Zhang et al., 2017). In particular, the importance of online reviews in shaping customer behavior (Liu et al., 2022), and advances in online review platforms in hospitality social media, have led to the growth in user-generated content and big data analysis in the hotel industry (Mariani and Borghi, 2021b). Implementation of real-time analysis methods (e.g., online review-based big data analysis) can help businesses to better evaluate customer behavior trends and take rapid action (Stylos et al., 2021).

Technologies have also been instrumental in carrying out machine learning analysis (Filieri et al., 2021b). With the development of machine learning methods and an increase in their accuracy, the use of these methods in identifying hotel features has also become common (Francesco and Roberta, 2019). Zhong et al. (2023) analyzed Chinese travelers' comments and divided hotel attributes into different segments. They found that landscape, traffic, food and attractions are core attributes of the hotels from travelers' point of view.

Some researchers have also developed models to extract revisit intention from user feedback texts. As mentioned before, the study of Park et al. (2020) applies a sentiment analysis on 105,126 customer reviews to compare the reviews of one-time visitors and re-visitors. They found out that reviews of re-visitors included more words in each sentence and had more positive/negative emotions relative to first-time visitors. Liu and Beldona (2021) analyzed TripAdvisor.com comments in order to detect revisit intention in hotel customers, alongside a sentiment analysis of the comments. They proposed and compared multiple methods of sentiment detection, and the performance of these methods in detecting the sentiment behind hotel customers' comments. Sun et al. (2021) implemented automated text mining methods to analyze big data collected from TripAdvisor.com and concluded that there are significant differences in customer desires before and after the Covid-19 pandemic.

2.2 Feature extraction in the hotel industry

To identify significant hotel features, research based on traditional methods such as questionnaires has long been used to find these features from the perspective of travelers. For example, Peng et al. (2015) have tried to examine the effect of different hotel features on the behavioral intention to return to a hotel. In their research, they have divided the characteristics of hotels into two main and auxiliary groups. The results of their study show that the room and its equipment are among the most important main features, and breakfast is one of the most important auxiliary features. The use of the interview method has also been commonly used to extract hotel features. Ren et al. (2016) first extracted a number of hotel characteristics based on interviews with customers, and then carried out a survey completed by 205 people. The results of their research show that the behavior and performance of hotel employees are of great importance from the point of view of travelers. Cleanliness, visual appeal inside and outside the hotel, color combinations and facilities close to the hotel are also important features.

2.3 Marketing mix

'Marketing mix' was first used by McCarthy et al. (1979), and includes product, place, promotion, and price. *Product* is the company's offer to the customer to meet his/her needs; *place* refers to making a service or product available in the right position and quantity; *promotion* is the exchange of information between the seller and the potential buyer or other people in the sales channels, which is designed to shape their behavior and attitude; and the *price* is the amount of money that the customer pays to achieve the value (Perreault Jr et al., 2013). Over the years, researchers have concluded that the four elements proposed by McCarthy et al. (1979) did not well capture the characteristics of the service industry. Therefore, Fisk et al. (1993), by adding three elements of people, physical evidence, and processes, tried to make the marketing mix more usable for the service industry.

People refer to all those who have been involved in the customer service process and have had an impact on customer perception; *physical evidence* is related to the environment in which the service is provided, where the customer and the service provider meet, along with all the visible components that help improve service performance and communication; and finally, mechanisms and flows of activity through which the customers' desired service is provided, consumed and produced with their cooperation, along with all existing operating systems, is called the *process*

(Fisk et al., 1993). In this study, we have used these definitions to identify and group each of the hotel features.

2.4 Segmentation in the hotel industry

Market segmentation helps to divide the market into smaller segments, create value for them according to the needs of these segments, gain a long-term competitive advantage, and reduce marketing costs (Dolnicar, 2020). Prior researchers have also examined and proposed different segmentations of customers in the hotel industry. Santos et al. (2020), for example, implemented segmentation for travelers of a region based on their food priorities. They collected data with questionnaires and clustered travelers into three segments: no interest in food, interest in particular types of food, and interest in local food.

Alongside these methods, the development of big data has shed light on new potentials in customer segmentation with the help of user-generated content. For example, Ahani et al. (2019) segmented the customers of spa hotels, and their analysis reveals 9 customer segments. One of these customer segments is, for instance, those travelers with a health-related attribute (e.g., stress) who have mentioned a hotel feature such as a steam room or sauna to be important for them. Another segment comprises customers who mention face treatments as a health requirement and includes those who value hotel attributes such as a mineral bath or healing water. Han et al. (2021) segmented Italian tourism destinations according to travelers' destination selection patterns using their location data. They found that cities which geographically were near to each other had been clustered into the same segments. Some studies have also segmented customers' travel destinations with the help of text-mining methods. Gour et al. (2021), for example, implemented segmentation on Indian nature tourism hotels and divided them into four segments based on their comment ranks. After that, they implemented text analysis to extract the most important features in each segment and classified all customer segments as ‘satisfied’ or ‘dissatisfied’.

3 Materials and Methodology

The purpose of this study is to identify the characteristics of hotels, and ultimately, the segmentation of hotel customers, based on the reviews of those who have mentioned that they intend to revisit a hotel. With the fast growth in user-generated content on online review platforms such as Tripadvisor.com, there is an intense need for better and faster analysis of travelers’ intentions. Prior studies have analyzed hotel revisits based on travelers’ reservation history (Park, 2019), but the proposed four-step methodology in this paper can help researchers to extract insights about travelers’ behavioral intentions more quickly and efficiently with the help of machine learning methods.

To this end, four main steps have been taken. In the first stage, after selecting the target countries and hotels, the opinions of hotel users and travelers were downloaded from Tripadvisor.com and the initial cleaning was conducted (Ahani et al., 2019). Secondly, the comments expressed by users indicating that they will return to the hotel were selected from the general data set using machine learning methods (Sánchez-Franco et al., 2019). In the third stage, the important and determining features of hotels from the perspective of hotel customers were extracted using computer text analysis and divided according to the 7Ps model of the marketing mix (Moro et al., 2019). Finally, hotel customers were segmented based on these features mentioned in the comments according to the 7Ps of the marketing mix (Bondzi-Simpson and Ayeh, 2019). The methodological process of this study is presented in Figure 1. Further details of each of these steps are provided below.

****Please insert Figure 1 here****

3.1 Step 1: data collection

Among various websites, Tripadvisor.com was chosen as it is one of the most popular online review platforms in the tourism sector and has millions of users (Ahani et al., 2019). Since in this research, we are going to use comments to segment customers, we followed the rule suggested by Dolnicar et al. (2014) to decide the number of comments to be collected. According to them, the result of the segmentation procedure will be statistically acceptable when the sample population is about seventy times bigger than the number of features used in segmentation. Therefore, we collected around 100,000 comments.

We also had to select countries, cities and hotels from which to extract comments. In order to achieve a high diversity of features (Francesco and Roberta, 2019), hotels from different geographical areas were selected. Considering the purpose of the present study, the generalizability of data was very important (Lee and Kim, 2021). To achieve this goal, we first selected at least one country from each continent. Then, we selected more countries from those continents that had higher shares, considering their share in the tourism industry (UNWTO, 2019). This led to the selection of 17 countries (see Supplementary Material, Appendix 1).

Following Mariani and Borghi (2021a), the most popular city in each country according to Tripadvisor.com data was selected for the study. The USA has four cities in our selected countries due to its highest share (15.5% of total tourism receipts) in the tourism industry (UNWTO, 2019).

This resulted in the selection of 20 cities for data collection (see Supplementary Material, Appendix 1) for the list of selected hotels, countries, and cities). At this stage, following the method that Mariani and Borghi (2021a) implemented in selecting hotels, based on the ranking of the hotels on the TripAdvisor.com website, the top five hotels from each of the selected cities were identified as the source of data collection. It was crucial to choose highly-ranked hotels because these hotels have a wide range of features and the output of analyzing this data could be more generalizable (Mariani and Borghi, 2021b). Finally, with the required number of comments and hotel listings, 1,000 comments were downloaded from each of the 100 selected hotels' pages on TripAdvisor.com. After data cleansing, 98,201 comments remained for analysis at the next stage. All these comments were posted in February 2020. This period of time was selected as it was before the start of the Covid-19 pandemic (Belete, 2021), after which the hotel industry and travelling patterns were harshly affected (Sun et al., 2021). For initial cleansing, comments written in a language other than English were removed.

3.2 Step 2: comment selection

After downloading the comments, at the first stage, we had to select the comments in which the users said that they would return to the hotel. We looked at the selection of comments as a text classification problem in which we could classify comments into two groups (re-visitors and non-re-visitors). We employed the Naïve Bayesian method, which has been one of the commonly used methods in hospitality literature, to classify texts and separate the comments (Sánchez-Franco et al., 2019). This method is simple, easy to use, and has a low probability of overfitting (Mehraliyev et al., 2022). To do this, we used the Scikit-learn package for the Python programming language. First, we selected *training set* comments (to learn the machine learning model); for each of them in which customers had claimed the intention to return, we chose the STAY tag, and for those comments without any phrase showing the intention to return, we chose the DO NOT STAY tag. Examples of some phrases from comments with revisit intention are shown in Table I.

****Please insert Table I here****

The process of tagging data was implemented by one author and double-checked with another author to enhance the reliability of identified tags (Liu and Beldona, 2021). We then divided tagged data into training and test sets which were implemented automatically by Python Scikit-learn package algorithms on the rest of the data. For the existing unstructured data to be interpretable, we had to convert each comment to an n-dimensional vector by separating all the words in the texts (Xu and Li, 2016). Finally, all the comments were converted into n-dimensional vectors using computer code, which included all the words in all the comments. We used the accepted Wrappers method (Kohavi and John, 1997) to reduce the dimensions used. Using this method, we could reduce the number of features or increase their number from zero to the highest possible accuracy, and select the desired number of features (Kohavi and John, 1997). Following Liu and Beldona (2021), with about 50 features, we reached the highest accuracy (98.2%) with a precision equal to 1 and a recall approximately equal to 0.96 of the model in classifying comments. These numbers demonstrate that the model can detect comments with revisit intention with good performance because of high recall; with its high precision, the model is also good in the classification of comments. So, with the same number of features, we trained the model and applied the obtained model to all comments. Finally, out of 98,201 comments, 23,996 comments were detected in which customers had acknowledged that they would return to the hotel.

3.3 Step 3: feature extraction

Feature selection has always been one of the issues of interest to researchers in the hotel and tourism literature. Researchers in this field have always tried to detect important features from the perspective of users through quantitative and qualitative methods (Xu and Li, 2016). To extract features, we employed the 'term frequency' method which is based on counting the number of repetitions of features, which is an acceptable and commonly used method in the literature (Yu et al., 2017). To do this with the Scikit-learn package for Python programming language, at first all the characters were changed to lowercase letters. Then, using the stop words prepared in this package, common spelling words, their verbs and suffixes, and items such as pronouns and prepositions were removed from the word list. This code provided us with an output list of these 3,000 words, along with a number of repetitions.

Researchers in the literature have chosen different procedures to select and examine a limited number of words. Following Stepchenkova et al. (2009), we selected one percent of features to be analyzed in the next steps. Finally, in these 23,976 comments, we reached 900 words that were repeated more than 240 times in the existing texts which were selected. It should be mentioned that the code used counted similar words as different words. For example, the words 'room' and 'rooms' were on a separate list, despite the high number of repetitions. Also, there were many words that had synonymous meanings, or words that were in the text and had the same meaning as one of the selected attributes. All these words were merged with the relevant attribute using the existing list and literature, and each of these words in the text became a synonymous attribute. At this stage, all features, according to the definitions provided in the literature (Perreault Jr et al., 2013), were placed in the relevant category of the marketing mix of the service industry. The final list of extracted features sorted by marketing mix is presented in Table II.

****Please insert Table II here****

3.4 Step 4: segmentation

One of these accepted and widely used segmentation methods in marketing literature is K-means (Zhong et al., 2023). Segmentation with the K-means method gives researchers the freedom to choose an appropriate number of segments according to the interpretability of results and managerial purposes (Park and Yoon, 2009). We used the Scikit-learn package for Python programming language to perform the clustering procedure. Before using the K-means method for clustering, the desired number of segments must be specified. In this step, we used the elbow method (Scheuffelen et al., 2019). Following Han et al. (2021), we plotted the clustering inertia diagram for the number of divisions for 2 to 20 divisions (see Supplementary Material, Appendix 2). The selection of the appropriate number of segments in the elbow diagram shows itself at the breaking point of the diagram. The change in inertia from 15 to 16 segments is approximately equal to 0.5%, which is less than the numbers calculated for differences in the number of segments less or more than 15 (Scheuffelen et al., 2019). Finally, 15 segments were selected as the desired number of final segments.

At this stage, similar to the comment selection step, the existing comments had to be converted into vectors with dimensions equal to the selected features, and the Tf-Idf method was used in the formation of vectors (Xu and Li, 2016). In calculating the number of each feature in one comment using this method, the number of repetitions of that word in the comment was considered, along with the total number of repetitions of that word in all comments. If the word was repeated more in that comment and less in all texts, it was considered as a more important one and had a bigger

Tf-Idf number (Xu and Li, 2016); this makes a certain word more important if that word was repeated in a text or comment, but is not mentioned in many comments (Netzer et al., 2012). For example, a word like ‘room’, which is mentioned in many comments, is less important due to its total high frequency and because it does not have the power to create a high distinction. Finally, using the tools in the Python Scikit-learn package, the comments were transformed into a vector with dimensions containing the Tf-Idf numbers of each feature. Then, K-means segmentation was applied to the data. After clustering, the accuracy of this step was measured by statistical tests. Following Zhong et al. (2023), we performed the ANOVA test separately on all features. The results of this test showed that for all properties, the P-value is less than 0.001 and that there is a significant difference between the different partitions, so it can be concluded that the partitions are different.

4 Results

4.1 Identifying important features from the travelers' point of view

According to the explanations given in the methodology section, 71 important features were extracted from all the words in 23,996 selected comments. Among the various features, a reference to the hotel features like room, staff, location and access, and food, respectively, were identified as the most important features in the comments of those users who had acknowledged the intention to return to the hotel. Table III shows the percentage of comments that each of the attributes contained.

****Please insert Table III here****

4.1.1 *Extracted features grouped by marketing mix*

After classifying extracted features based on the marketing mix 7Ps' framework, the results show that the room, with a percentage of 68%, is the most important *product*-related feature from the users' point of view. Food, beverages, and overall quality of services are next. Among meals, breakfast, dinner, and lunch had the highest number of mentions, respectively, which is not implausible given the type of hotel services. Interestingly, the free services provided in addition to the hotel cost are very popular with travelers. Among these services, the Internet, the quality of sleep, the gym, the tourist service, the laundry, and the sauna were important to the users.

Travelers look at the *place* component from three perspectives. The most important issue from their point of view is the location of the hotel and its accessibility. Another important issue is the geographical location of the hotel, for example, whether the hotel is located by the sea, in the city center, or by the river. The third group also talked about the convenience of using public transportation and the proximity to the airport and railways. Users have also talked about the two *promotion* channels: user reviews and hotel websites. Identifying these channels and examining them along with other features can help marketing experts perform better in planning marketing strategies. *Price* was one of the most important issues for travelers. This hotel feature (in the form of all keywords such as 'expensive', 'cheap', and 'paid') was mentioned in 20% of texts, which shows that it is an important issue for users. The users' position on price has not been discussed, and the mere mention of related words has been seen just as an indication of the importance of the financial issue for the traveler, and the existence of such concerns.

Travelers have spoken about the *people* component in two ways: the first group talked about the behavioral characteristics of employees such as being friendly, helpful, professional, and polite. Among these, the friendly behavior and helpfulness of the staff, with 39% and 33% of the total comments, were of importance. The second group also talked about special work positions in the hotel such as the hotel management team, tour leaders, the waiter, and the restaurant chef.

The features presented under the *physical evidence* consist of two parts: the first category is related to the physical equipment and facilities available in the hotel. Like the product section, in this section, a large and spacious room has been the most important feature from the users' point of view. Facilities and equipment in the room, swimming pool, new furniture and buildings in general, bathroom, beds, lobby, decor, balcony and terrace, rooftop, garden, and parking are important features of this category. The second group of customers referred to the qualities they experienced in the hotel. The most important of these items for travelers was the favorable scenery, the cleanliness of the hotel, the silence, the peace and comfort, the proper ventilation, the security, and the pleasant music. Some *processes* of hotels have been discussed and considered by users in reviews. A general reference to the hotel entry process and related issues, with 10,355 repetitions, is one of the most important processes from the users' point of view. The check-in process is their second consideration, and how to book a hotel is the next step. Cargo and luggage handling have also been mentioned by passengers, with 1,631 repetitions. A total of 5,704 passengers also spoke about the speed or slowness of the process, indicating the importance of this issue.

4.2 Identified segments

We have also attempted to segment customers to better understand the similarities and differences between each market segment in relation to the 7Ps of the marketing mix. The output of our segmentation process divides hotel customers into 15 different segments. The lowest value of opinions is related to four segments with a share of 5%, and the highest number is in the possession

383 of segments, with a share of 11% of total comments. The numbers obtained from the segmentation for different characteristics were normalized between two numbers, 0 and 100 so that in addition to the possibility of interpretation between the segments, it is possible to compare the characteristics of a component of the marketing mix in each segment. The results related to the importance of each feature in each of the 15 segments can be seen in Figure 2. As mentioned earlier, the numbers obtained from the clustering have been normalized to better understand the importance of each feature in the respective segment. Also, all the features of the hotel are included in the relevant marketing mix.

****Please insert Figure 2 here****

As outlined in Figure 2, people in the *first segment* are looking for a suitable and quiet room with enough facilities. All the important features for them, such as the size of the room, the bathroom, the Internet and the ventilation system, are somehow related to living in the room. Hotel managers can identify people in this group, provide them with special services based entirely on the hotel room, and achieve a high rate of return at no extra cost in other respects. For those in the *second segment,* hotel managers can still emphasize key features inside the hotel, but the difference between them and the first group is that they are looking for enjoyment inside the hotel (music, lobby, rooftop, drink). By offering a complete package of suitable entertainment in the hotel itself, in addition to earning more profit, hotel managers can increase the desire of customers to return to the hotel.

The *third segment* seeks to experience great food and good restaurants. The best package offered to these people can be a mix of a good restaurant and a cafe. Members of the *fourth segment* travel to the hotel to relax and exercise. Spa facilities and sports equipment are very important to them, and by providing this package along with providing a comfortable and carefree sleeping service, they can be encouraged to return to the hotel. The *fifth* and *sixth segments* mainly seek professional services from hotel staff. They can be kept satisfied by the provision of timely and appropriate services such as taking care of the rooms, along with proper staff training. It seems that for the members of the *seventh segment,* sleeping in a clean bed has the highest priority. Providing the necessities of a restful and comfortable sleep can shape their desire to return to the hotel. They also look for the least hassle in the process and often choose hotels located in urban centers.

For those in the *eighth segment* who are likely to travel with their children, all meals are important. They tend to receive good out-of-hotel tourism services. In addition, the tranquility and attractive environment of the hotel can be influential in their desire to return to the hotel. The *ninth segment* is related to economic travelers who mentioned financial concerns a lot in their comments. The hotel website is important to them, and this channel can be used as a good promotion channel by which to target them. The presence of a casino in the hotel can greatly help to make them want to return; Internet, parking, security, and cleanliness of the building are also other features that can be provided to improve the formation of targeted behavioral tendencies. The behavioral tendencies of the members of the *tenth segment* can be formed by high speed and efficiency in carrying out hotel-related processes, along with proximity to the airport and proper room maintenance.

For those present in the *eleventh segment*, it is valuable to provide the necessities for visiting the tourist destinations around the city. In their reviews, they talked about tourism services, hotel access, shopping malls and attractions, and providing facilities for these series of activities that can lead to a desire to return to the hotel. Customers in the *twelfth segment* are attracted by clean hotels with laundry services, and the provision of good access to the airport, transportation, and town. The *thirteenth segment* is for travelers who travel to coastal areas to experience relaxing moments. The quiet surroundings and the beach, along with rooms that have a beautiful view, are enough to make them want to return to the hotel. The members of the *fourteenth segment*, like the first segment, attach great importance to the room, with the difference that they consider the room as a product. The main emphasis of this segment in their comments was on the decoration, the facades of the hotel and rooms, the modernity and novelty of the building and its equipment. It seems that with impressive decor, these customers can be attracted to return to the hotel. The *fifteenth segment* includes people who pay special attention to the quality of services provided, and this issue plays the most fundamental role in determining their behavioral intentions.

5 Discussion and conclusions

5.1 Conclusions

In this research, we attempt to answer two critical questions about the revisit intention of hotel customers. First, to answer which key hotel features shape customers' revisit intentions, we shed light on 71 important hotel features and classify them based on the 7Ps (see Table II). To answer the second research question of this study and identify different segments of hotel customers that intend to return to a hotel, we reveal fifteen segments and highlight the key hotel features that are important for customers of each segment (see Figure 3).

5.2 Theoretical implications

The findings of this research thus offer several methodological, theoretical**,** and practical implications. First, we contribute to hospitality management and marketing literature by relying on digital technologies and employing big data analytics techniques which allow a richer interpretation of customers' experience (Zarezadeh et al., 2022). Regarding the urge to apply new methods in order to reach smart hospitality, it seems necessary to implement machine learning methods and big data-based technologies to have a customer-centric service (Buhalis et al., 2023). While prior studies in hospitality literature have shed light on some of the features impacting the revisit intention of customers, they were mainly based on traditional methods (Cheng et al., 2023). In this research, we have analyzed big data from customer reviews to complement previous knowledge and provide more generalizable insights. The methodology introduced in this paper can help researchers and hotel managers to implement customer insights analysis based on user-generated content in online review platforms, and to find revisit intention trends. In addition, although in the tourism literature, unstructured customer reviews have been analyzed to answer other key topics such as feature extraction (Taecharungroj and Mathayomchan, 2019), there are limited studies examining the revisit behavior of customers based on big data (Park et al., 2020).

In this research, we add to this strand of literature and highlight the key hotel features that should be considered and emphasized when encouraging customers to revisit hotels. In addition, we have employed the comment selection process to find the customers with revisit intention. Other studies have used the real revisit history of travelers to analyze customer reviews (Park, 2019, Park et al., 2020). The method introduced in this study can help researchers use big data and user-generated content to detect revisit intention in hotel customers when they have no access to precise historical revisit data.

In addition, the marketing mix framework has been used to describe the relationship between extracted features and the marketing strategy of the hotels. This methodology paves the way for researchers to not only consider the extraction of features and segmentation but also contemplate the marketing strategy to find more practical outputs. This study also introduces a comprehensive methodology that starts from big data collection and continues with text classification, feature extraction, and customer segmentation, to find insights about hotel customers' behavioral intentions. This four-step methodology empowers technology-based big data analysis, and introduces a new approach for a better understanding of consumer behavior and consequently better strategic planning in a wide variety of hotels, from small hotels to chain businesses (Buhalis and Leung, 2018). This big data based methodology can help researchers to better predict customer behaviors and identify new trends (Stylos et al., 2021). It also paves the way to find better solutions to enhance revenues and competitiveness by predicting real-time customer desires (Cheng et al., 2023).

The findings of this research also contribute to the hospitality and marketing literature by classifying the key hotel features playing a role in the revisit intentions of hotel customers which are studied, based on the 7Ps marketing mix literature. In terms of *product*, our findings illustrate that the most important feature of hotels from the point of view of these users is the main value proposition of the hotel, namely the room. Prior studies have also corroborated that rooms are one of the most important features from the hotel customers' point of view (Xu and Li, 2016). According to our results, the second most important factor is the quality of food and restaurants, which can result in higher revisit intention in travelers (Ding et al., 2022).

Hotel *location* is the next important factor. Many studies have talked about the accessibility of the hotel, the location of the hotel building and the importance of this issue from the perspective of travelers (Francesco and Roberta, 2019). The findings of our research also shed light on some key features in this category that are very important in behavioral intention to return to the hotel, but which are less discussed in the literature. These include features like access to public transportation and access to the airport and the railway. Regarding *promotion*, our research supports the current understanding in the hospitality and marketing literature about the importance of customer reviews for a hotel (Liu et al., 2020). Although a review of the literature illustrates that *price* and payment issues are among the most important features in some studies (Chen and Tsai, 2007), our findings indicate that despite the importance of price, it is a crucial factor in revisiting decisions from the perspective of only one segment of customers.

Research in recent years has also demonstrated that *employees* and issues related to them have always been important for travelers (Xu and Li, 2016). In this study, based on big data, we tried to prioritize the behavioral characteristics of staff in terms of importance from the perspective of travelers. The results show that friendly behavior, helpfulness, professional attitude and polite behavior are the most critical behavioral characteristics of employees, from the perspective of passengers, respectively (Ren et al., 2016).

The *physical evidence* of the hotel has also been examined in this study. Existing equipment, new equipment and facilities, bathroom, bed, ventilation system, hotel lobby, decoration and other similar items have been identified as important physical evidence from the perspective of travelers. Some of these features have been previously studied in the literature in this field (Peng et al., 2015, Ren et al., 2016), and some were added to the list of important features in this category, such as safety, hotel balcony, rooftop, garden, and music. The results of our research illuminate that the *processes* of booking, hotel arrival, reception and cargo are the most important process characteristics of the hotel from the customers' point of view. The significance of reservation, check-in and check-out has been mentioned in previous studies (Ren et al., 2016).

Another contribution of this research is to the segmentation literature in the hospitality and marketing domains. Prior studies have provided some segmentation of customers based on their food preferences (Santos et al., 2020), what customers want when visiting rural destinations (Park and Yoon, 2009), or travelers' destination selection patterns (Han et al., 2021). However, not only have some of these studies employed different criteria for the segmentation of customers but also, they have not necessarily focused on segmenting customers with revisit intention. In addition, only a few studies have suggested segmentation based on machine learning and text mining (Ahani et al., 2019, Gour et al., 2021). Moreover, the focus of those studies has either been on customer segmentation of a specific type of customer like spa hotel customers (Ahani et al., 2019) or segmentation of destinations based on customers' reviews (e.g., Gour et al. (2021). Customer segmentation based on the hotel features extracted from their opinions has been a step towards filling the gap in the literature in this field and trying to add to the body of this issue in the literature in the field of marketing in the hospitality industry (Ahani et al., 2019, Francesco and Roberta, 2019).

5.3 Practical implications

The findings of this research, which are based on big data analysis of customers' reviews with revisit intentions, offer several important implications for managers. First, this research sheds light on 71 hotel features that play a role in customers' decision to return to a hotel (see Table 2). Since hotel managers might have limited budgets and resources, it is crucial for them to understand the most important hotel features that they should invest in. In particular, we show that in terms of product, they should give priority to rooms, food, breakfast and service. Regarding the *place*, the hotel's geographical location and its access to the sea or the city center can be important. To enhance the promotion, particular attention should be paid to customer reviews.

Price is also an important issue, though our analysis illustrates it is a critical issue for returning to a hotel for one customer segment, but not all customers. Moreover, attention should be paid to the *people* aspect of the marketing mix. In particular, the professional and friendly behavior of the staff is critical. Finally, in terms of *physical evidence*, it is important that rooms are spacious, clean, and have a good view. Hotel facilities also seem to be a pivotal feature for many customers. Finally, hotel managers should endeavor to improve the arrival and chec- in *processes* as they seem to play a role in revisiting decisions, too.

Another important managerial implication of this research involves shedding light on different segments of customers with a revisit intention. According to our findings, there are 15 main segments of customers (see Figure 3) intending to visit a hotel again. It is important that hotel managers learn about these segments and the key hotel features that are important for each of these segments. They should then decide, based on the features of their hotels, which customer segments they want to target and how they should design their strategies to highlight those features in the eyes of their customers, to encourage a return to their hotels. Technologies such as big data analysis and machine learning methods play an important role by enabling managers to reach real-time insights about customers and decide quickly about their resource allocation and process enhancement. By learning about the key hotel features for each segment, managers can focus on highlighting the key facilities they need in a targeted manner, and achieve a high degree of customer willingness to return at a lower cost.

5.4 Limitations and future research

This research is not without limitations. Text analysis of comments could be conducted with other methods such as sentence-based analysis. The criterion for identifying and recognizing the importance of attributes in this research was the number of repetitions of words, while sentence-based analysis or group analysis of adjacent words can be employed to provide more accurate insights. Another limitation of this research is related to its method of dealing with synonyms. Future researchers can replicate the methodology by implementing other methods of word cluster detection. In addition, demographic information related to users has not been examined along with text analysis, and the loading and use of this data can be useful and effective in future research studies, especially in the process of segmentation and interpretation of the various sections obtained.

Due to limitations in the demographic information on TripAdvisor.com users, it was not possible to access a wide range of this information. This study detects customers with revisit intentions based on their comments, and researchers can replicate this study by collecting data from customers who have revisited hotels. In addition, in collecting data from the Tripadvisor.com platform, we have chosen comments on only the top five hotels from selected cities, which may lead to a bias in the results. Future studies can strengthen the findings of this research by collecting all the comments on all hotels in a larger number of cities around the world.

6 References

Ahani, A., Nilashi, M., Ibrahim, O., Sanzogni, L. & Weaven, S. (2019), "Market Segmentation and Travel Choice Prediction in Spa Hotels through Tripadvisor's Online Reviews", *International Journal of Hospitality Management*, Vol. 80, pp. 52-77.

Ahmad, W. & Sun, J. (2018), "Modeling Consumer Distrust of Online Hotel Reviews", *International Journal of Hospitality Management*, Vol. 71, pp. 77-90.

Belete, T. M. (2021), "Review on up-to-Date Status of Candidate Vaccines for Covid-19 Disease", *Infection and Drug Resistance*, Vol. 14, pp. 151.

Bondzi-Simpson, A. & Ayeh, J. K. (2019), "Assessing Hotel Readiness to Offer Local Cuisines: A Clustering Approach", *International Journal of Contemporary Hospitality Management*, Vol. 31 No. 2, pp. 998-1020.

Bortoluzzi, D. A., Lunkes, R. J., dos Santos, E. A. & Mendes, A. C. A. (2020), "Effect of Online Hotel Reviews on the Relationship between Defender and Prospector Strategies and Management Controls", *International Journal of Contemporary Hospitality Management*, Vol. 32 No. 12, pp. 3721-3745.

Buhalis, D. & Leung, R. (2018), "Smart Hospitality—Interconnectivity and Interoperability Towards an Ecosystem", *International Journal of Hospitality Management*, Vol. 71, pp. 41-50.

Buhalis, D., O'Connor, P. & Leung, R. (2023), "Smart Hospitality: From Smart Cities and Smart Tourism Towards Agile Business Ecosystems in Networked Destinations", *International Journal of Contemporary Hospitality Management*, Vol. 35 No. 1, pp. 369-393.

Chen, C.-F. & Tsai, D. (2007), "How Destination Image and Evaluative Factors Affect Behavioral Intentions?", *Tourism Management*, Vol. 28 No. 4, pp. 1115-1122.

Cheng, X., Xue, T., Yang, B. & Ma, B. (2023), "A Digital Transformation Approach in Hospitality and Tourism Research", *International Journal of Contemporary Hospitality Management*, doi: 10.1108/IJCHM-06-2022-0679.

Ding, L., Jiang, C. & Qu, H. (2022), "Generation Z Domestic Food Tourists' Experienced Restaurant Innovativeness toward Destination Cognitive Food Image and Revisit Intention", *International Journal of Contemporary Hospitality Management*, Vol. 34 No. 11, pp. 4157-4177.

Dolnicar, S. (2020), "Market Segmentation Analysis in Tourism: A Perspective Paper", *Tourism Review*, Vol. 75 No. 1, pp. 45-48.

Dolnicar, S., Grün, B., Leisch, F. & Schmidt, K. (2014), "Required Sample Sizes for Data-Driven Market Segmentation Analyses in Tourism", *Journal of Travel Research*, Vol. 53 No. 3, pp. 296-306.

Ernst, D. & Dolnicar, S. (2018), "How to Avoid Random Market Segmentation Solutions", *Journal of Travel Research*, Vol. 57 No. 1, pp. 69-82.

Filieri, R., Acikgoz, F., Ndou, V. & Dwivedi, Y. (2021a), "Is Tripadvisor Still Relevant? The Influence of Review Credibility, Review Usefulness, and Ease of Use on Consumers' Continuance Intention", *International Journal of Contemporary Hospitality Management*, Vol. 33 No. 1, pp. 199-223.

Filieri, R., D'Amico, E., Destefanis, A., Paolucci, E. & Raguseo, E. (2021b), "Artificial Intelligence (Ai) for Tourism: An European-Based Study on Successful Ai Tourism Start-Ups", *International Journal of Contemporary Hospitality Management*, Vol. 33 No. 11, pp. 4099-4125.

Fisk, R. P., Brown, S. W. & Bitner, M. J. (1993), "Tracking the Evolution of the Services Marketing Literature", *Journal of Retailing*, Vol. 69 No. 1, pp. 61-103.

Francesco, G. & Roberta, G. (2019), "Cross-Country Analysis of Perception and Emphasis of Hotel Attributes", *Tourism Management*, Vol. 74, pp. 24-42.

Gour, A., Aggarwal, S. & Erdem, M. (2021), "Reading between the Lines: Analyzing Online Reviews by Using a Multi-Method Web-Analytics Approach", *International Journal of Contemporary Hospitality Management*, Vol. 33 No. 2, pp. 490-512.

Hajibaba, H., Grün, B. & Dolnicar, S. (2020), "Improving the Stability of Market Segmentation Analysis", *International Journal of Contemporary Hospitality Management*, Vol. 32 No. 4, pp. 1393-1411.

Han, Q., Novais, M. A. & Zejnilovic, L. (2021), "Toward Travel Pattern Aware Tourism Region Planning: A Big Data Approach", *International Journal of Contemporary Hospitality Management*, Vol. 33 No. 6, pp. 2157-2175.

Jang, S. S. & Feng, R. (2007), "Temporal Destination Revisit Intention: The Effects of Novelty Seeking and Satisfaction", *Tourism Management*, Vol. 28 No. 2, pp. 580-590.

Kohavi, R. & John, G. H. (1997), "Wrappers for Feature Subset Selection", *Artificial Intelligence*, Vol. 97 No. 1-2, pp. 273-324.

Lee, Y. & Kim, D.-Y. (2021), "The Decision Tree for Longer-Stay Hotel Guest: The Relationship between Hotel Booking Determinants and Geographical Distance", *International Journal of Contemporary Hospitality Management*, Vol. 33 No. 6, pp. 2264-2282.

Liu, J., Yu, Y., Mehraliyev, F., Hu, S. & Chen, J. (2022), "What Affects the Online Ratings of Restaurant Consumers: A Research Perspective on Text-Mining Big Data Analysis", *International Journal of Contemporary Hospitality Management*, Vol. 34, No. 10, pp. 3607-3633, doi: 10.1108/IJCHM-06-2021-0749.

Liu, J., Zhang, H., Sun, J., Li, N. & Bilgihan, A. (2020), "How to Prevent Negative Online Customer Reviews: The Moderating Roles of Monetary Compensation and Psychological Compensation", *International Journal of Contemporary Hospitality Management*, Vol. 32 No. 10, pp. 3115-3134.

Liu, Y. & Beldona, S. (2021), "Extracting Revisit Intentions from Social Media Big Data: A Rule-Based Classification Model", *International Journal of Contemporary Hospitality Management*, Vol. 33 No. 6, pp. 2176-2193.

Lo, A. S. & Yao, S. S. (2019), "What Makes Hotel Online Reviews Credible? An Investigation of the Roles of Reviewer Expertise, Review Rating Consistency and Review Valence", *International Journal of Contemporary Hospitality Management*, Vol. 31 No. 1, pp. 41-60.

Lockyer, T. (2005), "Understanding the Dynamics of the Hotel Accommodation Purchase Decision", *International Journal of Contemporary Hospitality Management*, Vol. 17 No. 6, pp. 481-492.

Mariani, M. & Borghi, M. (2021a), "Are Environmental-Related Online Reviews More Helpful? A Big Data Analytics Approach", *International Journal of Contemporary Hospitality Management*, Vol. 33 No. 6, pp. 2065-2090.

Mariani, M. & Borghi, M. (2021b), "Customers’ Evaluation of Mechanical Artificial Intelligence in Hospitality Services: A Study Using Online Reviews Analytics", *International Journal of Contemporary Hospitality Management*, Vol. 33 No. 11, pp. 3956-3976.

McCarthy, E. J., Shapiro, S. J. & Perreault, W. D. 1979. *Basic Marketing*, Irwin-Dorsey, Ontario.

Mehraliyev, F., Chan, I. C. C. & Kirilenko, A. P. (2022), "Sentiment Analysis in Hospitality and Tourism: A Thematic and Methodological Review", *International Journal of Contemporary Hospitality Management*, Vol. 34 No. 1, pp. 46-77.

Moro, S., Pires, G., Rita, P. & Cortez, P. (2019), "A Text Mining and Topic Modelling Perspective of Ethnic Marketing Research", *Journal of Business Research*, Vol. 103, pp. 275-285.

Netzer, O., Feldman, R., Goldenberg, J. & Fresko, M. (2012), "Mine Your Own Business: Market-Structure Surveillance through Text Mining", *Marketing Science*, Vol. 31 No. 3, pp. 521-543.

Park, D.-B. & Yoon, Y.-S. (2009), "Segmentation by Motivation in Rural Tourism: A Korean Case Study", *Tourism Management*, Vol. 30 No. 1, pp. 99-108.

Park, E. (2019), "Motivations for Customer Revisit Behavior in Online Review Comments: Analyzing the Role of User Experience Using Big Data Approaches", *Journal of Retailing and Consumer Services*, Vol. 51, pp. 14-18.

Park, E., Kang, J., Choi, D. & Han, J. (2020), "Understanding Customers' Hotel Revisiting Behaviour: A Sentiment Analysis of Online Feedback Reviews", *Current Issues in Tourism*, Vol. 23 No. 5, pp. 605-611.

Peng, J., Zhao, X. & Mattila, A. S. (2015), "Improving Service Management in Budget Hotels", *International Journal of Hospitality Management*, Vol. 49, pp. 139-148.

Perreault Jr, W., Cannon, J. & McCarthy, E. J. 2013. *Basic Marketing*, McGraw-Hill Higher Education, London.

Ranjbari, M., Esfandabadi, Z. S. & Scagnelli, S. D. (2020), "A Big Data Approach to Map the Service Quality of Short-Stay Accommodation Sharing", *International Journal of Contemporary Hospitality Management*, Vol. 32 No. 8, pp. 2575-2592.

Ren, L., Qiu, H., Wang, P. & Lin, P. M. (2016), "Exploring Customer Experience with Budget Hotels: Dimensionality and Satisfaction", *International Journal of Hospitality Management*, Vol. 52, pp. 13-23.

Sánchez-Franco, M. J., Navarro-García, A. & Rondán-Cataluña, F. J. (2019), "A Naive Bayes Strategy for Classifying Customer Satisfaction: A Study Based on Online Reviews of Hospitality Services", *Journal of Business Research*, Vol. 101, pp. 499-506.

Santos, J. A. C., Santos, M. C., Pereira, L. N., Richards, G. & Caiado, L. (2020), "Local Food and Changes in Tourist Eating Habits in a Sun-and-Sea Destination: A Segmentation Approach", *International Journal of Contemporary Hospitality Management*, Vol. 32 No. 11, pp. 3501-3521.

Scheuffelen, S., Kemper, J. & Brettel, M. (2019), "How Do Human Attitudes and Values Predict Online Marketing Responsiveness?: Comparing Consumer Segmentation Bases toward Brand Purchase and Marketing Response", *Journal of Advertising Research*, Vol. 59 No. 2, pp. 142-157.

Stepchenkova, S., Kirilenko, A. P. & Morrison, A. M. (2009), "Facilitating Content Analysis in Tourism Research", *Journal of Travel Research*, Vol. 47 No. 4, pp. 454-469.

Stylos, N., Zwiegelaar, J. & Buhalis, D. (2021), "Big Data Empowered Agility for Dynamic, Volatile, and Time-Sensitive Service Industries: The Case of Tourism Sector", *International Journal of Contemporary Hospitality Management*, Vol. 33 No. 3, pp. 1015-1036.

Sun, S., Jiang, F., Feng, G., Wang, S. & Zhang, C. (2021), "The Impact of Covid-19 on Hotel Customer Satisfaction: Evidence from Beijing and Shanghai in China", *International Journal of Contemporary Hospitality Management*, Vol. 34 No. 1, pp. 382-406.

Taecharungroj, V. & Mathayomchan, B. (2019), "Analysing Tripadvisor Reviews of Tourist Attractions in Phuket, Thailand", *Tourism Management*, Vol. 75, pp. 550-568.

Um, S., Chon, K. & Ro, Y. (2006), "Antecedents of Revisit Intention", *Annals of Tourism Research*, Vol. 33 No. 4, pp. 1141-1158.

UNWTO 2019. International Tourism Highlights 2019 Edition. UNWTO Madrid, Spain.

Wilson, A., Zeithaml, V., Bitner, M. J. & Gremler, D. 2016. *Services Marketing: Integrating Customer Focus across the Firm*, McGraw-Hill Education, London.

Wu, J. S., Ye, S., Zheng, C. J. & Law, R. (2021), "Revisiting Customer Loyalty toward Mobile E-Commerce in the Hospitality Industry: Does Brand Viscosity Matter?", *International Journal of Contemporary Hospitality Management*, Vol. 33 No. 10, pp. 3514-3534.

Xiang, Z., Schwartz, Z., Gerdes Jr, J. H. & Uysal, M. (2015), "What Can Big Data and Text Analytics Tell Us About Hotel Guest Experience and Satisfaction?", *International Journal of Hospitality Management*, Vol. 44, pp. 120-130.

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*, Vol. 55, pp. 57-69.

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*, Vol. 32 No. 10, pp. 3249-3269.

Yu, Y., Li, X. & Jai, T.-M. C. (2017), "The Impact of Green Experience on Customer Satisfaction: Evidence from Tripadvisor", *International Journal of Contemporary Hospitality Management*, Vol. 29 No. 5, pp. 1340-1361.

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*, Vol. 34 No. 6, pp. 2320-2336.

Zhang, T. C., Omran, B. A. & Cobanoglu, C. (2017), "Generation Y's Positive and Negative Ewom: Use of Social Media and Mobile Technology", *International Journal of Contemporary Hospitality Management*, Vol. 29 No. 2, pp. 732-761.

Zhong, L., Liu, J., Morrison, A., Dong, Y., Zhu, M. & Li, L. (2023), "Destination Image: A Consumer-Based, Big Data-Enabled Approach", *International Journal of Contemporary Hospitality Management*, doi: 10.1108/IJCHM-12-2021-1557.

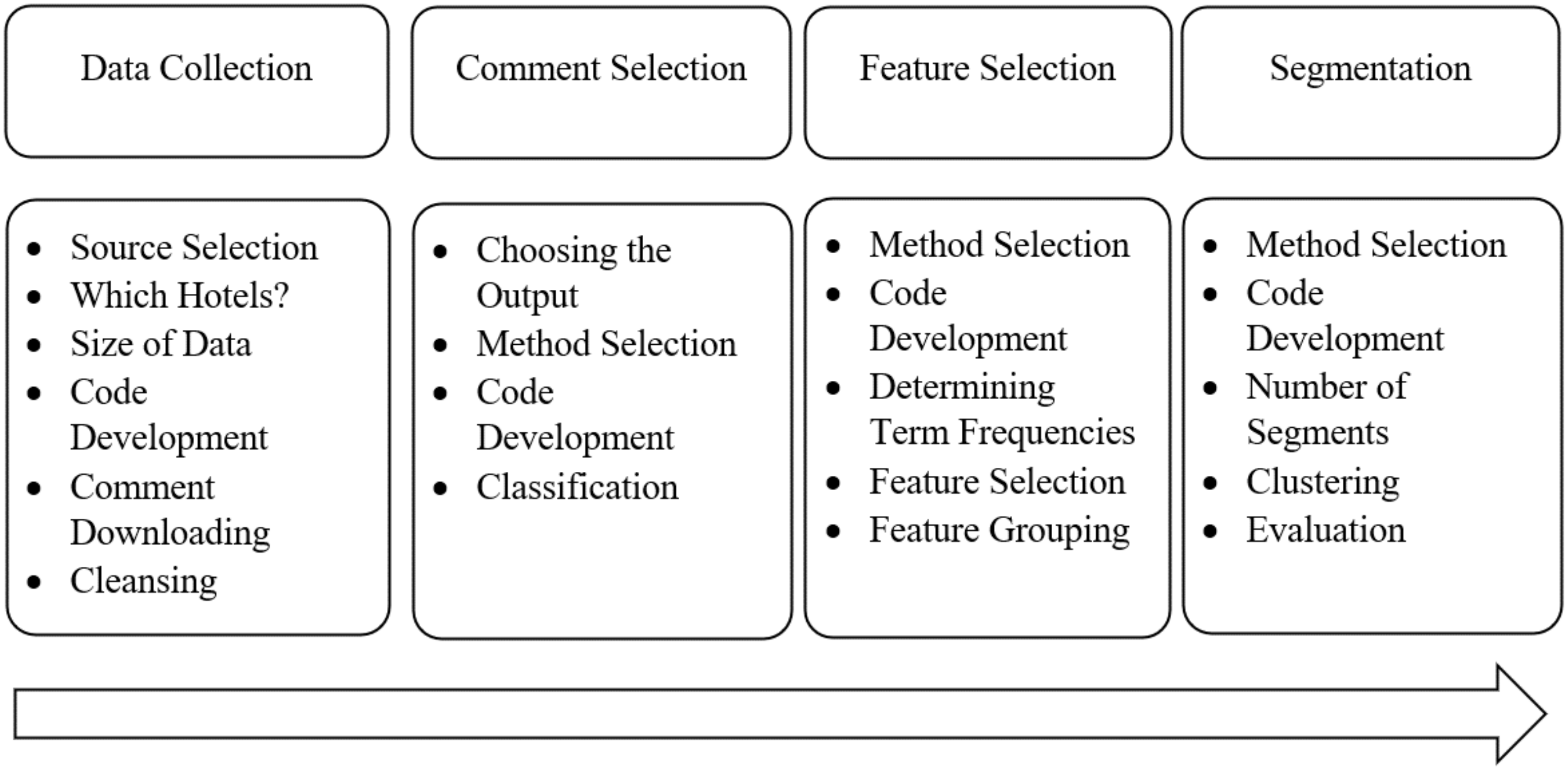


Figure 1: The methodological process of this study

7P	people									physical																					price	process				
Feature	chef	concierge	friendly	helpful	management	polite	professional	staff	waiter	air	balcony	bathroom	bed	building	clean	convenience	modern	decoration	facility	garden	housekeeping	lobby	music	parking	pool	quiet	rooftop	safety	spacious	view	price	arrival	booking	check	fast	luggage
Segment 1	16	51	20	17	23	46	6	31	18	100	76	100	33	67	83	22	55	96	100	43	79	95	38	25	8	79	24	100	100	41	12	50	42	14	28	44
Segment 2	17	32	29	20	49	37	8	39	83	57	93	7	12	45	43	15	42	94	47	38	47	100	96	13	9	72	100	56	37	51	10	47	35	9	24	23
Segment 3	100	60	31	24	72	55	11	44	100	49	63	5	8	40	58	14	35	71	38	41	35	43	33	14	9	56	17	36	33	42	8	37	32	6	19	13
Segment 4	21	51	26	20	46	40	11	41	22	56	67	6	12	34	58	10	43	57	53	100	44	40	100	19	100	100	57	50	41	54	7	33	26	8	21	11
Segment 5	21	100	100	100	64	100	5	100	24	29	31	2	2	24	97	16	24	58	30	18	54	39	19	18	3	51	13	67	26	25	3	42	27	4	13	15
Segment 6	39	60	43	29	85	92	100	60	74	25	16	2	3	15	51	12	33	60	29	20	35	36	43	12	5	40	9	39	21	35	4	38	29	7	14	12
Segment 7	15	37	30	23	26	68	7	41	5	31	81	14	100	47	100	100	44	88	62	24	38	58	12	20	5	82	25	58	77	30	7	38	40	7	21	26
Segment 8	36	37	29	20	68	34	10	38	30	39	61	6	12	39	56	8	49	29	50	80	42	39	24	15	30	66	7	59	46	28	9	32	34	8	22	17
Segment 9	3	35	18	13	47	30	6	33	13	36	28	8	12	61	73	12	49	39	53	20	28	56	28	100	9	48	16	81	42	33	100	40	72	12	21	19
Segment 10	11	73	26	24	89	49	10	40	19	38	18	8	13	48	66	13	52	46	43	21	86	53	24	36	6	49	15	63	33	30	9	100	71	100	100	100
Segment 11	6	33	30	24	14	49	5	39	12	31	39	6	9	76	88	18	43	59	32	32	27	50	8	25	4	64	15	67	47	27	8	26	28	5	17	22
Segment 12	8	42	23	22	12	52	5	33	3	31	25	11	13	90	87	23	38	54	35	28	24	63	3	13	2	62	6	46	60	25	10	42	33	8	19	46
Segment 13	18	50	26	17	32	32	10	36	36	32	100	6	10	33	55	11	53	47	52	69	39	30	79	50	29	79	13	55	35	100	8	26	31	7	22	11
Segment 14	29	75	16	15	100	70	5	34	27	84	72	5	8	100	79	8	100	100	47	33	68	66	36	47	4	66	19	73	43	72	6	70	100	6	23	27
Segment 15	24	73	33	23	80	66	6	46	75	26	21	1	3	18	47	10	22	40	33	13	100	22	31	12	4	35	7	41	16	42	3	34	25	5	20	8

7P	place														product																			promotion	
Feature	access	airport	attraction	cruise	island	location	park	transportaion	river	sea	shop	town	train	sun	birthday	breakfast	cafe	casino	complimentary	dinner	drink	food	gym	internet	kid	laundry	lunch	room	sauna	service	shuttle	sleep	spa	review	website
Segment 1	25	33	23	51	18	59	29	28	62	4	34	51	9	24	28	74	37	23	74	39	19	23	32	100	2	59	20	99	51	10	55	61	31	69	71
Segment 2	18	25	17	50	18	54	21	18	66	5	26	48	4	37	80	65	62	39	100	48	100	30	16	42	2	6	48	55	30	16	27	24	34	64	22
Segment 3	15	19	17	54	34	51	23	16	44	6	33	51	3	36	94	100	100	44	55	100	14	100	29	28	2	21	100	47	28	17	51	14	31	56	35
Segment 4	14	16	9	39	46	55	23	13	22	12	31	36	2	100	37	64	48	35	51	46	18	28	100	18	6	27	47	51	100	16	74	15	100	47	33
Segment 5	9	24	13	43	18	71	28	16	27	3	13	55	1	4	53	64	30	28	26	23	6	13	12	33	1	22	15	47	8	8	43	7	20	46	12
Segment 6	10	14	8	46	12	46	16	8	24	3	18	31	1	18	63	57	51	48	25	57	10	20	17	10	0	11	27	39	30	24	35	16	50	48	6
Segment 7	24	30	23	26	10	65	27	30	55	3	30	53	7	22	32	81	45	29	40	32	10	20	13	60	1	47	12	81	7	9	42	100	28	61	62
Segment 8	13	25	20	32	85	48	32	18	18	14	26	22	3	39	56	65	20	10	57	71	9	27	41	17	100	68	38	47	38	12	82	24	41	48	14
Segment 9	18	28	16	42	50	58	46	22	51	5	25	54	3	16	32	73	53	100	67	40	14	23	22	87	2	45	26	70	29	15	61	24	28	69	96
Segment 10	15	64	15	73	21	43	28	23	24	4	21	32	5	6	70	54	41	55	60	33	12	16	24	46	2	18	38	83	15	14	27	38	19	78	100
Segment 11	100	51	100	100	31	100	59	86	100	4	100	100	6	13	16	71	39	21	49	32	8	25	24	62	1	37	14	55	27	9	82	18	24	52	26
Segment 12	50	100	35	25	11	73	48	100	42	1	66	100	100	2	11	67	73	2	57	22	9	20	20	94	1	100	16	62	7	7	79	31	7	53	52
Segment 13	21	19	6	70	100	55	39	10	7	100	31	52	1	76	29	47	57	79	37	49	13	28	31	20	2	31	45	53	11	15	45	18	67	38	9
Segment 14	11	38	17	98	46	53	100	20	92	3	29	79	2	27	100	70	54	85	78	52	8	12	30	77	1	40	35	100	38	9	100	37	33	100	93
Segment 15	7	8	10	21	13	51	14	7	10	3	14	43	1	7	77	49	97	15	27	64	6	24	16	16	0	35	29	47	6	100	71	9	24	45	8

Figure 2: The importance of each feature in each of the 15 segments categorized by marketing mix

Fig 3 - Segments, their percentage of all comments and most important hotel features

Revisit Phrases
Would definitely come back
I would definitely stay again
Would definitely return
Will definitely try this hotel again
Will be visiting again
We will certainly be booking this hotel again for our trip next year
Will definitely be back again
We hope for a return visit

Table I - Phrases in user comments with intention to revisit hotels

People		***Product***		***Physical Evidence***	
staff	14,672	room	16,259	spacious	8,005
friendly	9,272	food	12,746	clean	7,194
helpful	7,858	breakfast	8,229	view	6,587
professional	4,779	service	7,834	facility	6,090
polite	1,280	drink	7,175	pool	5,111
management	1,191	complimentary	2,696	convenience	5,091
concierge	1,159	dinner	1,999	modern	5,060
waiter	459	kid	1,511	bed	4,585
chef	309	spa	1,255	quiet	4,480
Place		internet	1,038	bathroom	4,069
location	13,428	sleep	1,029	air	2,662
access	9,380	birthday	1,023	building	1,822
transportation	3,899	gym	899	lobby	1,707
sea	3,671	shuttle	790	decoration	1,528
town	3,314	lunch	555	safety	1,182
train	3,201	casino	405	balcony	1,131
shop	3,066	cafe	371	housekeeping	906
airport	2,024	sauna	236	rooftop	899
attraction	1,112	laundry	231	garden	753
river	1,008	***Process***		music	546
park	981	arrival	6,832	parking	491
island	510	check	4,624	***Promotion***	
sun	472	fast	4,278	review	1,258
cruise	330	booking	3,629	website	213
Price		luggage	1,291		
price	4,890				

Table II: Extracted features categorized according to marketing mix framework with the number of comments containing each feature

Attribute	Percentage	Attribute	Percentage	Attribute	Percentage
room	68%	bathroom	17%	internet	4%
staff	61%	transportation	16%	sleep	4%
location	56%	sea	15%	birthday	4%
food	53%	booking	15%	river	4%
access	39%	town	14%	park	4%
friendly	39%	train	13%	housekeeping	4%
breakfast	34%	shop	13%	gym	4%
spacious	33%	complimentary	11%	rooftop	4%
helpful	33%	air	11%	shuttle	3%
service	33%	airport	8%	garden	3%
clean	30%	dinner	8%	lunch	2%
drink	30%	building	8%	music	2%
arrival	28%	lobby	7%	island	2%
view	27%	decoration	6%	parking	2%
facility	25%	kid	6%	sun	2%
pool	21%	luggage	5%	waiter	2%
convenience	21%	polite	5%	casino	2%
modern	21%	review	5%	cafe	2%
price	20%	spa	5%	cruise	1%
professional	20%	management	5%	chef	1%
check	19%	safety	5%	sauna	1%
bed	19%	concierge	5%	laundry	1%
quiet	19%	balcony	5%	website	1%
fast	18%	attraction	5%		

Table III: percentage of comments that each of the attributes contained

Supplementary Material, Appendix 1: List of selected hotels by their city and country (sorted alphabetically)

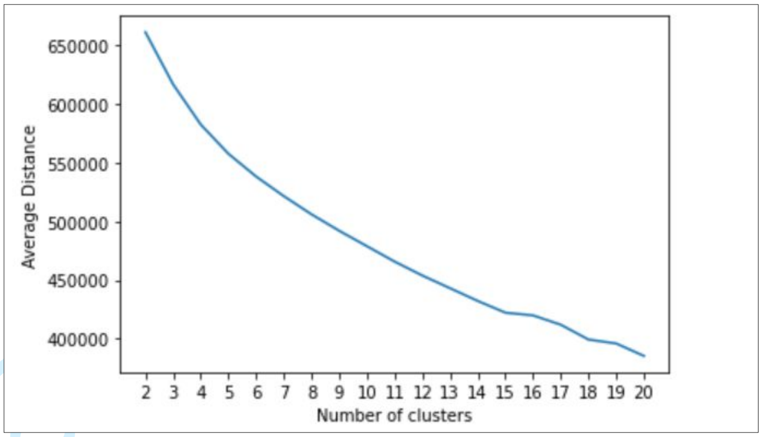
Hotel Name	**Country**	**City**
Adina Apartment Hotel Berlin Hackescher Markt	Germany	Berlin
Alila Seminyak	Indonesia	Bali
Amba Hotel Grosvenor	United Kingdom	London
Ambassade Hotel	Netherlands	Amsterdam
Amora Hotel Jamison	Australia	Sydney
Anantara Riverside Bangkok Resort	Thailand	Bangkok
Anantara The Palm Dubai Resort	United Arab Emirates (UAE)	Dubai
Banyan Tree Bangkok	Thailand	Bangkok
Basileus Hotel	Turkey	Istanbul
Caesars Palace	United States	Las Vegas
Caribe Hilton	United States	Puerto Rico
Casa Camper Hotel Barcelona	Spain	Barcelona
Chatrium Residence Sathon Bangkok	Thailand	Bangkok
citizenM Paris Gare de Lyon	France	Paris
Conrad Istanbul Bosphorus	Turkey	Istanbul
Crowne Plaza Times Square Manhattan	United States	New York
DoubleTree by Hilton Hotel London -Tower of London	United Kingdom	London
El Conquistador Resort	United States	Puerto Rico
FIVE Palm Jumeirah Dubai	United Arab Emirates (UAE)	Dubai
Four Seasons Hotel	Australia	Sydney
Four Seasons Hotel Seoul	South Korea	Seoul
Four Seasons Resort Oahu at Ko Olina	United States	Hawaii
Golden Nugget Hotel & Casino	United States	Las Vegas

Grand Hyatt Bali	Indonesia	Bali
Grand Hyatt Kuala Lumpur	Malaysia	Kuala Lumpur
Grand Visconti Palace	Italy	Milan
H10 Marina Barcelona Hotel	Spain	Barcelona
Harbour Marriott Hotel at Circular Quay	Australia	Sydney
Hilton Istanbul Bomonti Hotel & Conference Center	Turkey	Istanbul
Hilton Kuala Lumpur	Malaysia	Kuala Lumpur
Hilton Milan	Italy	Milan
Holiday Inn Express Paris-Canal de la Villette	France	Paris
Hotel 1898	Spain	Barcelona
Hotel Adlon Kempinski	Germany	Berlin
Hotel Berna	Italy	Milan
Hotel Century Southern Tower	Japan	Tokyo
Hotel Edison	United States	New York
Hotel Eiffel Seine	France	Paris
Hotel Estherea	Netherlands	Amsterdam
Hotel Jen Orchardgateway Singapore by Shangri-La	Singapore	Singapore
Hotel Niwa Tokyo	Japan	Tokyo
Hotel PJ Myeongdong	South Korea	Seoul
Hotel Sultania	Turkey	Istanbul
Hotel Sunroute Plaza Shinjuku	Japan	Tokyo
Ibis Amsterdam Centre	Netherlands	Amsterdam
ibis Milano Centro	Italy	Milan
Keio Plaza Hotel Tokyo	Japan	Tokyo
La Concha Renaissance San Juan Resort	United States	Puerto Rico
Lagoon Beach Hotel & Spa	South Africa	Cape town
Le Royal Meridien Beach Resort & Spa	United Arab Emirates (UAE)	Dubai
Lotte Hotel Seoul	South Korea	Seoul

Luxor Hotel & Casino	United States	Las Vegas
Mandalay Bay Resort & Casino	United States	Las Vegas
Mandarin Oriental, Kuala Lumpur	Malaysia	Kuala Lumpur
Marina Bay Sands	Singapore	Singapore
Marriott's Ko Olina Beach Club	United States	Hawaii
Mercure Hotel MOA Berlin	Germany	Berlin
Mercure Paris Centre Eiffel Tower Hotel	France	Paris
Meriton Suites Kent Street	Australia	Sydney
Millennium Hilton Bangkok	Thailand	Bangkok
NH City Centre Amsterdam	Netherlands	Amsterdam
NH Collection Amsterdam Grand Hotel Krasnapolsky	Netherlands	Amsterdam
NH Collection Milano President	Italy	Milan
Nikki Beach Resort & Spa Dubai	United Arab Emirates (UAE)	Dubai
Nine Tree Hotel Myeong-dong	South Korea	Seoul
Novotel Berlin Mitte	Germany	Berlin
Novotel Paris Les Halles	France	Paris
Padma Resort Ubud	Indonesia	Bali
Park Grand London Kensington	United Kingdom	London
Park Hotel Tokyo	Japan	Tokyo
Park Lane Hotel	United States	New York
Park Plaza Westminster Bridge London	United Kingdom	London
Puri Santrian	Indonesia	Bali
Radisson Blu Hotel Waterfront, Cape Town	South Africa	Cape town
Radisson Blu Hotel, Berlin	Germany	Berlin
Roda Al Murooj	United Arab Emirates (UAE)	Dubai
Row NYC Hotel	United States	New York
San Juan Marriott Resort & Stellaris Casino	United States	Puerto Rico
Shangri-La Hotel	Australia	Sydney

Shangri-La Hotel, At The Shard, London	United Kingdom	London
Shangri-La Hotel, Kuala Lumpur	Malaysia	Kuala Lumpur
Shangri-La Hotel, Singapore	Singapore	Singapore
Shangri-La's Rasa Sentosa Resort & Spa	Singapore	Singapore
Silverton Hotel and Casino	United States	Las Vegas
Swissotel The Bosphorus, Istanbul	Turkey	Istanbul
Taj Cape Town	South Africa	Cape town
The Corner Hotel	Spain	Barcelona
The Fullerton Hotel Singapore	Singapore	Singapore
The New Otani Kaimana Beach Hotel	United States	Hawaii
The President Hotel	South Africa	Cape town
The Shilla Seoul	South Korea	Seoul
The Sukhothai Bangkok	Thailand	Bangkok
The Table Bay Hotel	South Africa	Cape town
The Westin Princeville Ocean Resort Villas	United States	Hawaii
Traders Hotel, Kuala Lumpur	Malaysia	Kuala Lumpur
W Bali - Seminyak	Indonesia	Bali
W Barcelona	Spain	Barcelona
Waikiki Sand Villa Hotel	United States	Hawaii
Wyndham Grand Rio Mar Puerto Rico Golf & Beach Resort	United States	Puerto Rico
YOTEL New York	United States	New York

Supplementary Material, Appendix 2: Elbow diagram


650000
600000
550000
500000
450000
400000
Average Distance
2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Number of clusters