The theory-practice research gains from big data: evidence from hospitality loyalty programs

Paulo Rita (NOVA Information Management School (NOVA IMS), Universidade Nova de Lisboa, Lisbon, Portugal)

Maria Teresa Borges-Tiago (School of Business and Economics, University of the Azores, Ponta Delgada, Portugal)

Joana Caetano (NOVA Information Management School (NOVA IMS), Universidade Nova de Lisboa, Lisbon, Portugal)

This is the accepted author *manuscript of the following article published by EMERALD* **in** *International Journal of Contemporary Hospitality Management:*

Rita, P., Tiago, M. T. B., & Caetano, J. (2023). The theory-practice research gains from big data: Evidence from hospitality loyalty programs. International Journal of Contemporary Hospitality Management. https://doi.org/10.1108/IJCHM-05-2022-0646

Funding: This work was supported by national funds through FCT (Fundação para a Ciência e a Tecnologia), under the project - UIDB/04152/2020 - Centro de Investigação em Gestão de Informação (MagIC)/NOVA IMS, research grants UIDB/04521/2020 of the Advance/CSG, ISEG – Lisbon School of Economics and Management; and UIDB/00685/2020 of the Centre of Applied Economics Studies of the Atlantic, School of Business and Economics of the University of the Azores

The theory-practice research gains from big data: evidence from hospitality loyalty programs

Journal:	*International Journal of Contemporary Hospitality Management*
Manuscript ID	IJCHM-05-2022-0646.R3
Manuscript Type:	Original Article
Keywords:	Loyalty, Hotel loyalty programs, Customer segmentation, clustering, k-means, Big data

The theory-practice research gains from big data: evidence from hospitality loyalty programs

Structured Abstract

Purpose

The hospitality industry values segmentation and loyalty programs, but there is limited research on new methods for segmenting loyalty program members, so managers often rely on conventional techniques. This study seeks to use big data-driven segmentation methods to cluster customers and provide a new solution for customer segmentation in hotel loyalty programs.

Design/methodology/approach

Using the k-means algorithm, the study examined 498,655 profiles of guests enrolled in a multinational hotel chain's loyalty program. The objective was to cluster guests according to their consumption behavior and monetary value, and compare data-driven segments based on brand preferences, demographic data, and monetary value with loyalty program tiers.

Findings

The study shows that current tier-based loyalty programs lack features to improve customer segmentation, and some high-tier members generate less revenue than low-tier members. Therefore, more attention should be given to truly valuable customers.

Implications

Hotels can segment LP members to develop targeted campaigns and uncover new insights. This will help to transform LPs to make them more valuable and profitable and use differentiated rewards and strategies.

Originality

Since not all guests or hotel brands benefit equally from loyalty programs, additional segmentation is required to suit varying guest behaviors. Hotel managers can use data mining techniques to develop more efficient and valuable loyalty programs with personalized strategies and rewards.

Keywords

Loyalty; Hotel loyalty programs; Big data; Customer segmentation; clustering; k-means

INTRODUCTION

Increasing competition, new disruptive business models (Dolnicar, 2020b; Xu *et al.*, 2022), and customer-buying process modifications (Garrigos-Simon *et al.*, 2017) have reshaped the tourism and hospitality industry. In early 2020, these challenges were exacerbated by the COVID-19 pandemic, leading Sigala (2020, p.312) to state that "COVID-19 tourism impacts will be uneven in space and time" and impel researchers and practitioners to search for new tools to enhance tourism reignition. The transformative opportunity created by this scenario requires more than reimagining and creating new tools (Gretzel *et al.*, 2020); it demands looking at all available data from a customer-oriented perspective that takes full advantage of the segmentation processes (Dolnicar, 2020a).

According to Stylos et al. (2021), agility and market intelligence are essential for businesses in the tourism and hospitality industries to develop value propositions that attract and retain guests, as also noted by Su and Reynolds (2017). As market competitiveness increases, customers become key elements for achieving success (Law *et al.*, 2018; Rahimi *et al.*, 2017), impelling firms to adopt different approaches that enhance the customer-firm relationship (Afaq *et al.*, 2022).

Owing to their positive impact on occupancy rates and profitability, hotels widely use loyalty programs (LPs) as customer relationship management (CRM) tools to boost client loyalty (Chen *et al.*, 2021; Hua *et al.*, 2018; Koo *et al.*, 2020). However, most LPs are obsolete (Nastasoiu and Vandenbosch, 2019) and focus on tier levels and financial benefits that clients do not value (Gandomi and Zolfaghari, 2017; Shoemaker and Lewis, 1999; Tanford, *et al.*, 2016). Reward-centric LPs appeal more to customers with higher loyalty motivations than tier-based ones (Kreis and Mafael, 2014). Nonetheless, LPs are still commonly used, leading to the question of whether new insights and strategies for improving CRM can emerge given the available data. The primary purpose of this study is to provide evidence that big data underlying tier-based LPs are underexplored and can provide valuable insights to target customers.

Big data methods are required to address challenges such as data complexity and volume, to obtain actionable information to support decision-making for these strategies (Zarezadeh *et al.*, 2022). Data mining techniques have been employed to investigate hotel guest segmentation based on expenditure and demographics (Moro *et al.*, 2018; Talón-Ballestero *et al.*, 2018). However, there is limited research on improving LP strategies through a refined customer segmentation strategy. Tanford *et al.* (2016) recommended that future research should focus on segmentation studies that identify customer groups with similar characteristics to create sophisticated reward tier structures. Dursun and Caber (2016) suggested that research should evaluate customer profiles according to hotel type and location. However, Zarezadeh *et al.* (2022) noted that big data need to be treated carefully to support management decision processes.

In line with the above discussion, there is a need for an empirical attempt to operationalize this approach using a more agile, innovative, and data-driven approach that takes advantage of already available LP data to recreate hotel segmentation and promote customer-centric strategies. Increasingly available big data linked to LPs may offer an opportunity to address this research gap. Thus, this study assesses whether tier segments provide enough information to pursue valuable loyalty strategies compared to the real insights derived from a loyalty member database. A refined segmentation approach was developed using big data, and a double-segmentation process was implemented to identify distinct and meaningful customer profiles hidden behind the current loyalty-tier segments. The findings show that LP confines a great level of information that can be used to conduct fine-tuned targeting that considers customer behavior, regardless of the tier level. Thus, managers can use this information to customize their offers within tier levels and improve customer relationships and loyalty.

this paper is structured in the following manner. First, the research questions addressed in the subsequent section frame the relevant literature. Then, the two adopted data treatment processes are shown, followed by the findings. The last section provides the theoretical and managerial implications, suggestions for future research and the limitations.

LITERATURE REVIEW

Firms face a new reality in which brands are co-created with customers (Borges-Tiago *et al.*, 2021) in spaces with wider power. This scenario challenges firms to enhance their firm-customer relationships and target increasing customer loyalty (Shin *et al.*, 2020; Shin *et al.*, 2021).

Loyal customers are less sensitive to promotional offers, but more committed to the service or product, and consequently expect a reciprocal relationship; thus, they are stricter when there is a failure (Wu *et al.*, 2021). Hence, loyal customers have higher switching costs to move from one brand or product because of their favorable relationship or attachment to a service provider (Evanschitzky *et al.*, 2012), and they expect a higher level of service and recognition. Therefore, firms invest considerable effort in pursuing customer loyalty and understanding all the dimensions (Kandampully *et al.*, 2015).

Initially, it explored the frequency and amount of purchases, and it was designed as a reward system. Subsequently, these reward systems were improved to form exchange systems, allowing users to exchange points earned for their past purchases. A new tiered program type was established with airline frequent flyer programs, in which loyal customers were rewarded with free products, discounts, and upgrades that increased with higher tiers. Technology and social media enhance client-firm interactions over different platforms and facilitate better understanding of the attitudinal and behavioral responses that an LP ignites.

Hierarchical LPs or tier-based programs are standard instruments used in relationship marketing; they are traditionally associated with frequency. Customers are awarded tiers according to their expenditure patterns and if they exceed certain spending levels (Tanford and Malek, 2015). They earn status points, and their tiers are upgraded as they shop. Therefore, firms with traditional frequency programs tend to adopt metrics directly linked to tier growth when segmenting markets. The literature criticizes loyalty-tier-based strategies because they do not consider program members' individual consumer behaviors, leading to less effective segmentation (Voorhees *et al.*, 2011; Zeng *et al.*, 2022).

Nastasoiu and Vandenbosch (2019) mapped different loyalty strategies and positioned them according to how easily competitors could replicate them. Inexpensive and simple-to-replicate incentives often resulted in expensive price wars that eroded market value. Conversely, exclusive and customized incentives provided a competitive edge and generated value for the organization. Most original LPs rely on creating value sources, challenging emulations by competitors, and identifying valuable customers. Research has addressed value-added benefits in three areas: financial, psychological, and brand awareness (Shoemaker, 2003; Tanford *et al.*, 2016). In addition, program value perceptions are crucial for engaging with the program and avoiding the loss of clients to competitors (Nastasoiu and Vandenbosch, 2019; Tanford, 2013).

Nonetheless, Lacey and Sneath (2006) stated that firms must consider customer value to avoid misplacing efforts and resources on less valuable customers while under satisfying others. According to Hu and Yeh (2014), a valuable customer is one who has recently and repeatedly spent large amounts of money on a brand.

Design or structure also play a critical role in customers' value perception and motivation to belong to an LP. Kreis and Mafael (2014) concluded that customers' motivation to belong to a program highly influences the LP's perceived value. In addition, the LP design also affects the reasons for being a member. Reward-centered LPs attract customers with higher loyalty intentions (Kreis and Mafael, 2014). Recent studies point to the need to analyze the processes of tier demotion. Studies have shown that when customers in the top tier are demoted, they increase their willingness to buy and their loyalty intentions in the short run to restore their lost status (Chang, 2020).

Communication determines the acquisition of new customers and retention of existing ones in the LP (Berezan *et al.*, 2016). The communication process strengthens the bond between the customer and brand. With web and social media developments, firms have explored communication as a promotion, interaction, and information-gathering tool (Raab *et al.*, 2016). These data can enhance customer understanding and help develop tailored offers and communication, which is crucial for the overall program value. Although efforts have been made to create campaigns and offers based on customer characteristics, programs still rely primarily on traditional tier structures with point schemas (Tanford *et al.*, 2016).

All dimensions of an LP are equally important for enhancing its value. Although many studies have analyzed LP performance, research on solutions to improve customer use and satisfaction with LPs is limited. However, as noted by Meyer-Waarden (2007), consumers generally have cards from different LPs, which implies that belonging to a program does not equate to their loyalty toward the brand or firm. Moreover, these customers tend to assess the differences between firms' value propositions, and when there are no perceived differences, they tend not to behave loyally (Kim *et al.*, 2021). Within an LP, several unique segments can be identified based on perceived value and loyalty behavior during consumers' life stages (Allaway *et al.*, 2014). Thus, in addition to inducing perceived value, from a firms' perspective, LPs are also relevant in assessing customer value across their lifecycles, refining LPs, and focusing effectively on retaining valid and valuable customer segments.

Hotels and loyalty programs

Loyalty schemes and incentives are widely used in the hospitality industry to pursue loyalty strategies and stimulate customer frequency and retention (Koo *et al.*, 2020; Shin *et al.*, 2021). Hotels use LPs to collect customer data and reward clients based on their expenditures and frequency. The number of LP members who claimed that membership influenced their choice of hotel has increased (Barsky, 2011). Shoemaker and Lewis (1999) distinguished between two types of programs: frequency programs that foster repeat business and real LPs that focus on developing an emotional bond with the brand (see, Supplementary_material_appendix_1). These authors developed the loyalty triangle, which classifies hotel programs into three categories: financial, procedural, and psychological. Shoemaker (2003) expanded upon his prior research to establish the loyalty circle, a model utilized to assess hospitality loyalty. This framework encompasses three equally significant dimensions: communication, value, and process.

The process aspect involves tier structures, program regulations, benefits, and redemption policies. It pertains to the structured systems that impact customer satisfaction and expectations. Additionally, the kind, amount, and timing of rewards provided by a program significantly influence its perceived value and, consequently, guest loyalty (Hu *et al.*, 2010). Generally, hotel LPs are designed to reward customers through point accumulation and redemption systems, along with tier progression. Tier structure design is a management strategy to reduce costs and is an easy way to segment members according to their spending levels (McCall and Voorhees, 2010). A challenge associated with program design in multi-branded hotel chains, is that the same program features may not be equally efficient across different brands and customer groups.

Although higher-tier members have higher behavioral loyalty (Tanford, 2013) and loyal customers are more willing to pay (Evanschitzky *et al.*, 2012), their customer value may change. Certain high-tier guests may generate less revenue for a hotel due to an ill-suited loyalty program design that does not align its regulations with customer value (Voorhees *et al.*, 2011). Therefore, hotels may be overinvesting in these customers, suggesting that current tier-based programs are not optimal (Gandomi and Zolfaghari, 2017). Because creating, implementing, and monitoring customer-driven and cross-functional strategies require significant operational effort, only a small number of hotel managers have adopted these approaches to facilitate effective customer communication (Sarmaniotis *et al.*, 2013). Therefore, hotels mostly position their LPs as frequency programs instead of actual LPs that offer customer-centric programs.

Hotel LPs should be designed according to customer preferences and behavior because an LP is not a one-size-fits-all solution (Xie and Chen, 2014). As Chang (2020) posited, there is a need to reassess the client's profile over time using available data, since LPs only prove effective in specific contexts. Moreover, a well-structured tiered LP must be designed based on the natural segmentation of hotel clients (Nastasoiu and Vandenbosch, 2019). This approach allows hotels to identify the most valuable clients, recognize their value through tier levels, and address their needs and preferences to foster an emotional bond. This emotional bond makes a difference and strengthens loyalty toward the brand (Evanschitzky *et al.*, 2012). Since the earliest programs, the rule of rewarding loyal customers has remained unchanged (Chen *et al.*, 2021). This challenges hotels to classify guests according to their loyalty behavior (Hansen *et al.*, 2010) and reinvent customer experiences to increase LP efficiency (Chen *et al.*, 2021).

Various classification techniques using big data have been employed to compare guest behaviors over a period. Although big data plays a crucial role in revolutionizing hospitality research and practice, its availability alone does not guarantee better decision-making for managers and researchers (Mariani & Baggio, 2022; Zarezadeh *et al.*, 2022). Talón-Ballestero *et al.* (2018) and Tanford and Malek (2015) employed data mining techniques to study hotel guest segmentation based on guests' monetary value and demographics. The first criterion–financial or monetary value, is linked to the price, perceived value of money, and value of points. Therefore, the monetary value of a program is easily compared between different programs and is easily replicable by competitors (Nastasoiu and Vandenbosch, 2019). Demographics, as the second set, are frequently utilized. The results highlight the need for different segmentation approaches because traditional tier-based segmentation cannot capture consumer behavior within a tier-based LP.

Min *et al.* (2002) used data mining techniques, specifically decision trees, to classify guest data and form hotel customer profiles according to profitability, travel purpose, and demographics. Although Min *et al.* (2002) assessed customer segmentation using data mining techniques, their findings were limited to luxury hotels located in a single country. Chung *et al.* (2004) presented a similar outcome using three segmentation approaches with data collected from guest surveys of 12 deluxe hotels in Seoul. Unlike these studies, Tanford and Malek (2015) focused only on segmenting LP members from several hotels in the United States. Their hierarchical cluster analysis produced six clusters, and the guests were profiled according to their loyalty and green concerns. The results showed that other segmentation solutions, driven by behavioral and psychographic data, enhanced LPs, apart from tiers. Talón-Ballester *et al.* (2018) developed a more complete and extensive study on guest profiling using data from nearly four million guests.

The evolution of these studies clearly shows the growing importance of using big data methods to obtain meaningful and actionable customer insights (Talón-Ballestero *et al.*, 2018). Descriptive and exploratory analytics generally strive to enhance efficiency, streamline processes, and uncover new knowledge (Mariani and Baggio, 2022). In the hospitality ecossystem, big data has three primary sources: devices, users, and operations (transaction data) (Zarezadeh *et al.*, 2022). The use of big data gathered on social media allows for the unveiling of consumer behavioral patterns (Liu and Beldona, 2021), and for these reasons, it has been the focus of research (Rita *et al.*, 2022). In addition to external data, firms have transactional data that have not been fully explored; however, only a few studies in the literature have explored data mining capabilities to conduct segmentation within a customer LP to reveal guest behaviors, attitudes, and preferences.

A common trait of big data methods is the use of data-driven analysis. Dolnicar and Leisch (2010) drew attention to misinterpretation concerns related to data-driven market segmentation analysis. According to them, segmenting is an exploratory analysis, thus it can produce different outcomes if applied repeatedly or by using a different distance-based method. Therefore, it is advisable to consider the stability of the segments, which consist of having similar final solutions regarding size, number, or main attributes, when repeating the analysis over time or with different clustering techniques.

Although hotels have already adopted the technology required to collect and process customer data, most hotel management do not take full advantage of these tools (Sarmaniotis *et al.*,2013), and use traditional approaches such as tier-based programs to segment clients (Voorhees *et al.*, 2011). Thus, Voorhees *et al.* (2011) and Tanford *et al.* (2016) pointed out the need to enhance loyalty members' segmentation to sustain personalized communication and value strategies, which are difficult to emulate by competitors. Considering these scenarios, this research addressed the following questions about hotel LPs and the information concealed therein:

1. Do guest segments match the tier-program layers?

2. To what extent can data-driven segmentation unveil value segments that differ from current tier-level segments to foster real loyalty relationships?

METHODS

To answer these questions, LP data from an international hotel group were used as the data source and ethical issues regarding guests and identified units were safeguarded. This allowed us to investigate the behavior of multitier LPs and client segments using big data.

This hotel group has operated globally for more than 40 years. It has a portfolio of almost 100 hotels worldwide, consisting of four brands with different hotel features, services, locations, and prices. Hence, an understanding of each brand's characteristics is critical for the interpretation of the results. One brand explores historical and iconic properties such as castles, palaces, and monasteries. With four- to five-star hotels, this brand is well known for its monumental, historical, and small-luxury environment (H&IP). The group's premium brand includes seven five-star hotels in Europe in luxurious buildings in prime locations offering luxurious services (PB). The hotels and resorts (H&R) brand have more than 60 four- and five-star hotels spread throughout Europe, Africa, and the Americas. This includes city hotels, beach hotels, nature hotels, and family resorts. Its' youngest brand, linked to a brand ambassador, is present in a small number of vibrant, urban, and unique environmental units (YB). In general, the first two brands are pricier compared to the latter two.

The database consisted of transactional and sociodemographic data of nearly half a million (498,655) loyalty members with at least one reservation during the time window from 2016 to 2019 (first semester). Transactional data are extremely reliable and valid, as they relate to real customer purchases within the hotel group and are automatically recorded on the company's IT systems at the time of consumption. Sociodemographic data can be less accurate because customers and employees provide data manually. The data gathered were only related to customers enrolled in the LP. The LP was a multi-tier program created over ten years ago; it involved a point accumulation and redemption system. It had four tiers (silver, gold, platinum, and corporate), was free of charge, and offered benefits throughout the group's four brands. The first three tiers were based on client expenditure and reservations. The corporate level was assigned via contracts to corporate customers and discounts differed according to the contract.

In the present study, a double-segmentation process was implemented as data-driven market segmentation was conducted. However, several clustering models have previously been used to segment hotel guests, such as the decision tree classifier CHAID (Chung *et al.*, 2004), C5.0 decision tree algorithm (Min *et al.*, 2002), RFM (Dursun and Caber, 2016), and MapReduce algorithm combined with statistical techniques (Talón-Ballestero *et al.*, 2018). However, k-means was the first clustering algorithm employed in this study because of its simplicity and ability to detect patterns (Davidson, 2002; Tripathi *et al.*, 2018). In big data treatment, applying k-means algorithms can be challenging because of resource consumption (CPU and time) (Jiang, et al., 2001; Jain, 2010). Therefore, K-means++ was used as a randomized seeding technique, as proposed by Arthur and Vassilvitskii (2006). This version of k-means has improved speed and accuracy of the k-means results by using advanced seeding of the initial cluster centers.

In addition, a two-step cluster analysis procedure was adopted to minimize potential misinterpretation of the optimal cluster solution. Compared to conventional clustering methods, this algorithm possesses unique features and scalability, enabling it to handle both categorical and continuous variables. This algorithm allows the handling of both categorical and continuous variables and has scalability and distinctive features compared to traditional clustering techniques. Moreover, according to Jiang *et al.* (2001), two-step clustering is a straightforward method for data mining. Therefore, a second clustering approach was considered that decreased time and space complexity and aimed to unveil the similarities and robustness of the initial cluster solutions. Thus, a two-path process was adopted for data analysis and algorithm implementation: (i) the Jupyter Notebook, an open-source web application, was utilized alongside the Python programming language, and (ii) IBM's SPSS 23.0 was used.

Past studies have emphasized the relevance of diverse customer profiling aspects, such as buying behaviors (Kim et al., 2006) and socio-demographic, psychographic, and behavioral traits (Tasci, 2017). Demographic factors such as nationality, sex, and age significantly influence consumer loyalty (Min *et al.*, 2002). Expenditure, frequency, and length of stay were also deemed relevant for classifying hotel guests according to their behavior (Dursun and Caber, 2016; Talón-Ballestero *et al.*, 2018). Brand customer preferences were considered using the number of reservations made and the number of nights booked by each customer for each hotel brand as measurement variables.

From the available variables (see Supplementary_material_appendix_2), new variables were computed for analysis purposes, such as "age" and "country." The variable "country" had more than 100 distinct values and comprised of the top ten nationalities representing 82.6% of the total population: Portuguese (23.4%), British (19.3%), German (10.7%), Brazilian (6.4%), North American (5.8%), Spanish (5.4%), French (5.2%), Dutch (2.4%), Chinese (2.3%), and Belgian (1.7%).

The information used for further evaluation was derived from 383,180 loyalty members; the difference from the original number (498,655) was the result of removing outliers, inconsistencies, and the top ten countries' filter. The average age of LP members was 55 years, and approximately 71% of them were male. Although the highest percentage of guests were from Portugal, customers who spent the most money were from Great Britain. Customers from Great Britain, Portugal, and Germany accounted for 68% of the total expenditure. Assessing the behavior of tiers per brand, gold and platinum customers stayed at H&R establishments most of the time. Corporate members preferred to stay at either H&R or H&IP establishments. With regards to silver members, the percentage of stays per brand was more diversified: 56% preferred H&R properties, 30% preferred H&IP establishments, and 11% preferred to stay with the premium brand. The percentage of stay per brand was partly defined by the intrinsic characteristics of each brand. The H&R brand accounted for 57% of hotels in the group.

RESULTS

Brand Preference Clustering

The clustering process was conducted using brand preference as the partition variable. According to the elbow method, the points where the marginal effect of adding one more cluster had a lower effect on the SSE had three, five, and six clusters (see Figure 1). When evaluating the cluster silhouette coefficient, values closer to 1 were achieved for six, seven, and eight clusters. The average distance of the nearest cluster was greater than the average distance from the observations within the cluster. Higher values indicated more separated and denser clusters, with a higher number of clusters regarding the CH score.

Insert Figure 1 around here

Finally, the number of clusters that indicated a model with better separation between clusters was two, six, and seven, where the DB index was lower, suggesting fewer similarities between clusters. Therefore, six clusters were used in the brand-preference clustering algorithm.

After implementing the k-means algorithm on the normalized dataset, six clusters were defined. An introductory analysis of cluster centroids suggested that cluster 0 had a higher number of stays at premium brand hotels and cluster 5 had a higher number of stays at H&R establishments (see Supplementary_material_appendix_3).

Using the two-step clustering method, the following auto-clustering solution was found using variables related to brand preference as partition criteria. In this case, the clustering criterion was BIC, and it was calculated for every possible number of clusters. A desirable solution is characterized by a significant ratio of BIC changes, a substantial ratio of distance measures, and a lower BIC value. In this study, the two-cluster solution met these criteria (Supplementary_material_appendix_4).

The solution obtained through the two-step clustering procedure did not match that obtained in the initial clustering procedure regarding clustering agglomeration. This might be because the final decision regarding the optimal cluster number was made in the first analysis using business knowledge instead of statistical criteria. This result reinforces the idea presented by Dolnicar and Leisch (2017) that clusters within suboptimal solutions are sometimes discarded. This could be interesting for a specific organization, such as the present case, since the suboptimal cluster solution enhances business knowledge and brand loyalty capabilities.

Monetary Value Clustering

In the case of monetary value clustering, the elbow method had a lower effect on the SSE related to the marginal effect of adding one more cluster with three, four, and seven clusters. Higher coefficients were achieved for the silhouette coefficients with two, three, and four clusters. However, when analyzing the CH score and DB index, the best values were obtained with three, seven, and eight clusters. Thus, for the monetary value clustering algorithm, three clusters were the most suitable choice for the number of clusters to be used according to all validation measures.

The centroids that resulted from the algorithm implementation revealed that observations belonging to cluster 0 had the highest expenditure and number of stays from the clustering process. Regarding enrollment year in the LP, cluster 1 had observations with an older enrollment date.

By applying the same technique, the obtained solution suggested three clusters (Supplementary_material_appendix_5). The cluster distribution demonstrates the occurrence rate of each cluster, indicating a clear predominance of cluster 3. In this case, the number of clusters achieved is the same for both solutions, reinforcing the significance of the clustering solution obtained.

The final groups found using the monetary value clustering algorithm, presented higher stability because both solutions showed k = 3. Nonetheless, brand preference clustering results were adequate from a business perspective. Thus, starting with analyzing the brand preference clustering results, it was impossible to assess clusters with different behaviors corresponding to guests' consumption patterns. However, the initial analysis implies this (Supplementary_material_appendix_6). One segment had a high value, representing 2.5% of the total number of members considered in this study. These customers were "Old in the Loyalty Program" and had an average expenditure per customer of €4,213. The brand where the guests spent the most was H&R, where tourists stayed for extended periods and had a broader hotel portfolio. Moreover, this segment presented a more balanced tier distribution, with 44% of the guests being silver members, 33% gold members, 22% platinum members, and 1% corporate members. The segment "Recent in LP and Low Value" corresponds to 85% of customers and presents a similar average expenditure pattern to the cluster with “Oldest in LP & Low-Value” clients. However, the average enrollment year in the LP was much more recent. “Recent in LP & Low Value” were recent clients, mainly from the silver tier. However, guests with fewer stays but with longer stay lengths or stays in more expensive hotels would probably be considered valuable even though they do not present with frequent behavior.

By cross-tabbing the two clustering solutions and looking at customer tiers in the LP, the cluster with the higher proportion of guests was characterized by a preference for H&R hotels, never staying at the youngest brand, and having the lowest frequency (see Figure 2).

Insert Figure 2 around here

Guests in this segment were mostly from Great Britain, Portugal, and Germany; 95% of these clients were silver members, and 4% were gold members. In contrast, the “H&R Lovers and Higher Frequency” segment was the smallest segment in the sample, mainly with silver members and a high proportion of British guests. The two segments identified with a specific brand prefer to distinguish themselves by having the most frequent and older enrollment guests that belong to the LP for a longer time. However, the tier distributions of these two segments were notably different. In the “H&R Lovers” case, 43% of the guests were gold members, and 20% were platinum members. In addition, silver members who belonged to the “Old in LP and High Value” may have had higher average expenditure per year since enrollment than gold members. On the other hand, in the “Iconic properties lovers, explorers and frequent” segment, 79% of guests were silver members, 13% were gold members, and 7% were platinum members. Although frequent guests comprised majority of this segment, a significant portion belonged to the lower tier, and a high proportion of these guests came from the Netherlands or Portugal.

The “H&R, Never YB, Lowest Freq” segment comprised Portuguese clients followed by American and British with a low-frequency use rate and silver tier. The “Iconic Properties or H&R Beginner Lovers” segment was the youngest, averaging 51 years. Some of the silver-tier guests assigned to this cluster exhibited repetitive behavior in top-brand hotels and came from various countries. The segment with the most guests with no explicit brand preference, the “Multi Branded and Low Frequency” segment, had many silver members from the United States of America.

The clustering results showed that other characteristics could better define and segment loyalty members than tiers. Three main findings were obtained by comparing the segments based on guest data with the current tier segments. First, lower-tier members were more valuable or frequent users than higher-tier members. Second, guests showed certain consumption behaviors presented in all tiers, which leads to the belief that there are groups of customers with distinct brand preferences, stay durations, and destination patterns that are not being correctly grouped because of tier segmentation. Third, some low-tier customers did not have frequent behavior but spent more than higher-tier members during their stays. Furthermore, when segmenting guests based on their monetary value and brand preference, valuable insights were extracted to better understand them.

These findings support those of Voorhees *et al.* (2011) and Tanford and Malek (2015), who argued that LP segmentation should comprise descriptive foundations, such as transactional and demographic data, to differentiate guests effectively. The findings suggest that certain high-tier members generate less revenue for the hotel, highlighting the need for further segmentation to cater to different guest behaviors. As seen in the previous results and analysis, customers belonging to the same tier have different countries of residence, preferences, and consumption behaviors. The findings demonstrate that current tier-based LPs do not adequately differentiate between guests in terms of behavior patterns and idiosyncratic characteristics.

CONCLUSION

Nowadays, hoteliers and brand managers focus on guest behavior (Rahimi *et al.*, 2017), to achieve loyalty by using different strategies to attract and retain guests (Su and Reynolds, 2017). With social CRM, reward programs have been designed as guest retention strategies using tiers with standard benefits as guest segmentation criteria (Gandomi and Zolfaghari, 2017; Shoemaker and Lewis, 1999; Tanford *et al.*, 2016). However, several other dimensions apart from tiers should be used to segment LP members.

Our study introduces a double segmentation process to classify and regroup clients within LP members, taking advantage of the large amounts of transactional data available on hotels that tend to be neglected. This enables dynamic updates of segments and the adoption of new loyalty and promotion strategies. The results were obtained by applying two different data-driven segmentation procedures and unveiled value segments that differed from the current tier-level segments. For instance, it shows that to foster genuine loyalty relationships, hoteliers need to look beyond the top tiers, since long-term engagement is not obtained at the highest levels.

These findings question LP design and rules, which consider a tier-based structure and mainly adopt the measurement of transactions and sales growth over time. This study shows that loyalty to a hotel brand can be achieved by measuring individual lifetime values. However, clients with high lifetime values may not correspond to those in higher program tiers. Thus, hotel managers must rethink LP design and strategies by considering three dimensions: brand preference, individual lifetime value, and program tier. This new perspective arises from the already available data and shows different segments unrelated to the tier concept. Moreover, the different profiles within the tiers permitted a narrower segmentation (e.g., mature, younger), and consequently, will allow the design of several targeting communicational and promotional strategies.

The unveiled segments stand out from existing tier classifiers by considering additional variables, given the new purpose of the data. This approach is also suitable for real-time processing in large and dynamic data environments. Recent trends can emerge within client databases and used to offer personalized and differentiated services. Furthermore, the additional segmentation of LP members can be used to detect trends.

Theoretical Implications

This research aimed to challenge hotels' traditional segmentation strategies based on LPs, which are outdated, easily replicable by competitors, and do not take advantage of the large volume of data available concerning their loyalty members. Hotel LPs usually segment customers based on tier levels (Voorhees *et al.*, 2011), primarily on the frequency and amount spent. However, this type of segmentation does not truly reflect customer characteristics. Therefore, our research focused on using big data techniques, specifically clustering processes, to extract meaningful knowledge from customer demographics and transactional data. The main goal was to compare the current segmentation strategy used by hotel LPs based on tiers with groups of customers formed through autonomous clustering processes. A double-segmentation process was conducted to guarantee the quality of the obtained solutions. The results reinforce the notion presented by Tanford and Malek (2015) and led to the conclusion that guest segments do not match the tier-program layers (see Figure 2). The outcomes suggest that LPs, although ranking members into groups based on specific purchase amounts and frequencies, do not fully align with customer behavior. It is expected that while customers move upward through a program tier, the amount spent in the hotel will increase; however, our results show that lower-tier members present higher spending patterns during their stay.

Furthermore, the results show that an LP may be inefficient if designed to cover several types of hotel brands, because it does not consider guests' brand preferences, customer value, or profitability potential. Thus, different segmentation levels are required to suit the behavior of distinct guest behaviors and hotel characteristics. Similar findings were reported by Voorhees *et al.* (2011) and Tanford and Malek (2015).

Chen *et al.* (2021) reported the need to add a cultural perspective to LP research. Although not fully embracing this challenge, the results showed the predominance of certain nationalities in some clusters. As in other research fields, cultural differences are considered drivers of guest behavior (Liu *et al.*, 2020). Nonetheless, more recent studies have demonstrated that different consumer behaviors are not only connected to country-cultural differences but also reflect individual responses to situational variables that affect consumer-firm interactions (Liu *et al.*, 2020). Furthermore, with this consideration in mind, further research is necessary from customer and firm perspectives to posit that cultural differences can be found within these sub-segments.

Practical Implications

Along with previous studies, this research demonstrates the need for an alternative segmentation of LP members to match guests' idiosyncratic behaviors and features with the hotel brands' intrinsic characteristics. As tiers create natural opportunities to connect and engage with customers, it is time for hoteliers to use the additional data available in these programs and effectively unveil segments within every tier by monitoring and analyzing member spending activities, preferences, and behaviors. Hotels can use effective segmentation when adopting big data treatment methods to extract hidden patterns and relevant knowledge that are currently not being explored. For example, the information collected in LP programs and usually not treated for the tier process hides different guests' profiles that can arise from their values and preferences. This information can be used to develop communication and promotional campaigns to nurture specific groups or to understand which customers would respond better to gamification or exclusivity rewards as bounding strategies. Nevertheless, it also allows the creation of specific tiers for brand ambassadors, influencers, or clients identified as valuable brand endorsers. All these efforts can help uncover new consumer insights and enhance specific segments' perceived value without massively increasing hotels' costs to provide it, which will drive loyalty to the segments of guests.

From data to value is the path to follow, which allows hotel managers to identify different consumer spending patterns within their LP and even establish their value for the hotel. Moreover, to better pinpoint high-value customers and tailor strategies that fine-tune the LP based on the features and benefits that consumers value the most and do not increase costs. In addition to creating a new model, this study provides insights that unveil the endless possibilities available within LP data related to customers' transactions and engagement.

Moreover, the findings suggest that hotels over-invest in some high-tier customers and underinvest in low-tier members who have a higher potential to turn into valuable customers.

Hence, hotel management should transform LPs into more valuable programs for guests and make them more profitable for hotels. To conclude, LPs should guarantee guests' retention and stimulate future behavior and commitment through customized strategies that generate more value for the company. Differentiated rewards, campaigns, and program rules could be applied to acknowledge guests' preferences and generate more value for the LP and company.

Limitations and future research

This research has some limitations. The data collected were limited to only one hotel group and active LP members. In addition, data were only collected for loyalty members who had at least one reservation from 2016 onwards, meaning that other insights might still be unexplored in other past loyalty behaviors. Future research should also evaluate other LP segmentation solutions with more data granularity, such as room or food and beverage preferences. Finally, as the findings highlight some differences driven by the country of residence, future research should also address the cultural differences that influence the adoption and engagement of customers with LP programs.

References

Afaq, A., Gaur, L., and Singh, G. (2022). Social CRM: linking the dots of customer service and customer loyalty during COVID-19 in the hotel industry. *International Journal of Contemporary Hospitality Management*, Vol. 35 No. 3, pp. 992-1009. https://doi.org/10.1108/IJCHM-04-2022-0428

Allaway, A. W., D'Souza, G., Berkowitz, D., and Kim, K. K. (2014). Dynamic segmentation of loyalty program behavior. *Journal of Marketing Analytics*, Vol. 2, No. 1, pp. 18-32. https://doi.org/10.1057/jma.2014.2

Arthur, D., and Vassilvitskii, S. (2006). *k-means++: The advantages of careful seeding*. Stanford.

Barsky, J. (2011). Hotel reward programs attracting more consumers. *Hotel Management*, 226(12), pp. 12.

Berezan, O., Yoo, M., and Christodoulidou, N. (2016). The impact of communication channels on communication style and information quality for hotel loyalty programs. *Journal of Hospitality and Tourism Technology*, Vol. 7, No.1, pp. 100-116. https://doi.org/10.1108/JHTT-08-2015-0031

Borges-Tiago, M. T., Arruda, C., Tiago, F., and Rita, P. (2021). Differences between TripAdvisor and Booking.com in branding co-creation. *Journal of Business Research*, Vol. 123, pp. 380-388. https://doi.org/10.1016/j.jbusres.2020.09.050

Chang, W. (2020). Different status reevaluation period and communication styles for top-tier and bottom-tier customers in multi-tier loyalty programs. *European Journal of Marketing*, Vol. 54, No.12, https://doi.org/10.1108/EJM-10-2018-0713

Chen, Y., Mandler, T., and Meyer-Waarden, L. (2021). Three decades of research on loyalty programs: A literature review and future research agenda. *Journal of Business Research*, Vol.124, pp.179-197. https://doi.org/10.1016/j.jbusres.2020.11.057

Chung, K. Y., Oh, S. Y., Kim, S. S., & Han, S. Y. (2004). Three representative market segmentation methodologies for hotel guest room customers. *Tourism Management*, 25(4), 429-441. doi:10.1016/S0261-5177(03)00115-8

Davidson, I. (2002). *Understanding K-means non-hierarchical clustering*. Computer Science Department of State University of New York (SUNY), Albany.

Dolnicar, S. (2020a). Market segmentation analysis in tourism: a perspective paper. *Tourism Review*, Vol. 75, No.1, pp.45-48. https://doi.org/10.1108/TR-02-2019-0041

Dolnicar, S. (2020b). Sharing economy and peer-to-peer accommodation - a perspective paper. *Tourism Review*,Vol. 76, No. 1, pp.34-37. https://doi.org/10.1108/tr-05-2019-0197

Dolnicar, S., and Leisch, F. (2010). Evaluation of structure and reproducibility of cluster solutions using the bootstrap. *Marketing Letters*, Vol. 21, No.1, pp.83-101. https://doi.org/10.1007/s11002-009-9083-4

Dolnicar, S., and Leisch, F. (2017). Using segment level stability to select target segments in data-driven market segmentation studies. *Marketing Letters*, Vol. 28, No.3, pp. 423-436. https://doi.org/10.1007/s11002-017-9423-8.

Dursun, A., & Caber, M. (2016). Using datamining techniques for profiling profitable hotel customers: An An application of RFM analysis. *Tourism Management Perspectives*, 18, 153-160. doi:10.1016/j.tmp.2016.03.001

Evanschitzky, H., Ramaseshan, B., Woisetschläger, D. M., Richelsen, V., Blut, M., & Backhaus, C. (2012). Consequences of customer loyalty to the loyalty program and to the company. *Journal of the Academy of Marketing Science*, Vol. 40, No.5, pp. 625-638. https://doi.org/10.1007/s11747-011-0272-3

Gandomi, A., and Zolfaghari, S. (2017). To tier or not to tier: an analysis of MultiTier loyalty programs '. *Omega*, Vol.74, pp. 20-36. https://doi.org/10.1016/j.omega.2017.01.003

Garrigos-Simon, F. J., Galdon, J. L., and Sanz-Blas, S. (2017). Effects of crowdvoting on hotels: the Booking.com case. Retrieved from Booking.com. case. *International Journal of Contemporary Hospitality Management*, Vol. 29, No.1, pp. 419-437. https://doi.org/10.1108/IJCHM-08-2015-0435

Gretzel, U., Fuchs, M., Baggio, R., Hoepken, W., Law, R., Neidhardt, J., . . . Xiang, Z. (2020). e-tourism beyond COVID-19: a call for transformative research. *Information Technology and Tourism*, Vol. 22, No.2, pp. 187-203. https://doi.org/10.1007/s40558-020-00181-3

Hansen, J. D., Deitz, G. D., and Morgan, R. M. (2010). Taxonomy of service-based loyalty program members. *Journal of Services Marketing*, Vol. 24, No.4, pp. 271-282. https://doi.org/10.1108/08876041011052980

Hu, H.-H. S., Huang, C.-T., and Chen, P.-T. (2010). Do reward programs truly build loyalty for lodging industry? *International Journal of Hospitality Management*, Vol. 29, No.1, pp.128-135. https://doi.org/10.1016/j.ijhm.2009.07.002

Hu, Y. H., & Yeh, T. W. (2014). Discovering valuable frequent patterns based on RFM analysis without customer identification information. *Knowledge-Based Systems*, 61, 76-88.

Hua, N., Wei, W., DeFranco, A. L., and Wang, D. (2018). Do loyalty programs really matter for hotel operational and financial performance?. *International Journal of Contemporary Hospitality Management*, Vol. 30, No.5, pp. 2195-2213. https://doi.org/10.1108/IJCHM-12-2016-0643

Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, Vol.31, No.8, pp. 651-666. https://doi.org/10.1016/j.patrec.2009.09.011

Jiang, M. F., Tseng, S. S., and Su, C. M. (2001). Two-phase clustering process for outliers detection. *Pattern Recognition Letters*, Vol. 22, No.6-7, pp. 691-700. https://doi.org/10.1016/S0167-8655(00)00131-8

Kandampully, J., Zhang, T., and Bilgihan, A. (2015). Customer loyalty: a review and future directions with a special focus on the hospitality industry. *International Journal of Contemporary Hospitality Management*, Vol. 27, No. 3, pp. 379-414. https://doi.org/10.1108/IJCHM-03-2014-0151

Kim, S.-Y., Jung, T.-S., Suh, E.-H., & Hwang, H.-S. (2006). Customer segmentation and strategy development based on customer lifetime value: A case study. *Expert Systems with Applications*, 31(1), 101-107. doi:10.1016/j.eswa.2005.09.004

Kim, J. J., Steinhoff, L., and Palmatier, R. W. (2021). An emerging theory of loyalty program dynamics. *Journal of the Academy of Marketing Science*, Vol. 49, No.1, pp.71-95. https://doi.org/10.1007/s11747-020-00719-1

Koo, B., Yu, J., and Han, H. (2020). The role of loyalty programs in boosting hotel guest loyalty: impact of switching barriers. *International Journal of Hospitality Management*, Vol. 84, pp. 102328. https://doi.org/10.1016/j.ijhm.2019.102328

Kreis, H., and Mafael, A. (2014). The influence of customer loyalty program design on the relationship between customer motives and value perception. *Journal of Retailing and Consumer Services*, Vol. 21, No.4, pp. 590-600. https://doi.org/10.1016/j.jretconser.2014.04.006

Lacey, R., & Sneath, J. Z. (2006). Customer loyalty programs: are they fair to consumers?. *Journal of consumer marketing*, 23(7), 458-464.

Law, R., Fong, D. K. C., Chan, I. C. C., and Fong, L. H. N. (2018). Systematic review of hospitality CRM research. *International Journal of Contemporary Hospitality Management*, Vol. 30, No.3, pp.1686-1704. https://doi.org/10.1108/IJCHM-06-2017-0333

Liu, Y., and 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. https://doi.org/10.1108/IJCHM-06-2020-0592

Liu, Y., Hultman, M., Eisingerich, A. B., and Wei, X. (2020). How does brand loyalty interact with tourism destination? Exploring the effect of brand loyalty on place attachment. *Annals of Tourism Research*, Vol. 81, pp. 102879. doi:https://doi.org/10.1016/j.annals.2020.102879

Mariani, M., and Baggio, R. (2022). Big data and analytics in hospitality and tourism: a systematic literature review. *International Journal of Contemporary Hospitality Management*, Vol. 34, No. 1, pp. 231-278. https://doi.org/10.1108/IJCHM-03-2021-0301

McCall, M., and Voorhees, C. (2010). The drivers of loyalty program success. *Cornell Hospitality Quarterly*, Vol. 51, No.1, pp.35-52. https://doi.org/10.1177/1938965509355395

Meyer-Waarden, L. (2007). The effects of loyalty programs on customer lifetime duration and share of wallet. *Journal of Retailing*, Vol. 83, No. 2, pp. 223-236. doi:https://doi.org/10.1016/j.jretai.2007.01.002

Min, H., Min, H., & Emam, A. (2002). A data mining approach to developing the profiles of hotel customers. *International Journal of Contemporary Hospitality Management*, 14(6), 274-285. doi:10.1108/09596110210436814

Moro, S., Rita, P., and Oliveira, C. (2018). Factors influencing hotels' online prices. *Journal of Hospitality Marketing & Management*, Vol. 27, No. 4, pp. 443-464. https://doi.org/10.1080/19368623.2018.1395379

Nastasoiu, A., and Vandenbosch, M. (2019). Competing with loyalty: how to design successful customer loyalty reward programs. *Business Horizons*, Vol. 62, No.2, pp. 207-214. https://doi.org/10.1016/j.bushor.2018.11.002

Raab, C., Berezan, O., Krishen, A. S., and Tanford, S. (2016). What's in a word? Building Program Loyalty through Social Media communication. *Cornell Hospitality Quarterly*, Vol. 57, No.2, pp. 138-149. https://doi.org/10.1177/1938965515619488

Rahimi, R., Köseoglu, M. A., Ersoy, A. B., and Okumus, F. (2017). Customer relationship management research in tourism and hospitality: a state-of-the-art. *Tourism Review*, Vol. 72, No.2, pp.209-220. https://doi.org/10.1108/TR-01-2017-0011

Rita, P., Ramos, R., Borges-Tiago, M. T., and Rodrigues, D. (2022). Impact of the rating system on sentiment and tone of voice: A Booking. com and TripAdvisor comparison study. *International Journal of Hospitality Management*, Vol. 104, pp. 103245. https://doi.org/10.1016/j.ijhm.2022.103245

Sarmaniotis, C., Assimakopoulos, C., and Papaioannou, E. (2013). Successful implementation of CRM in luxury hotels: determinants and measurements. *EuroMed Journal of Business*, Vol. 8, No.2, pp.134-153. https://doi.org/10.1108/EMJB-06-2013-0031

Shin, M., Back, K.-J., Lee, C.-K., and Lee, Y.-S. (2020). Enhancing customer-brand relationship by leveraging loyalty program experiences that foster customer-brand identification. *International Journal of Contemporary Hospitality Management*, Vol. 32, No.12, pp. 3991-4016. https://doi.org/10.1108/IJCHM-06-2020-0550

Shin, M., Back, K.-J., Lee, C.-K., and Lee, Y.-S. (2021). The loyalty program for our self-esteem: the role of collective self-esteem in luxury hotel membership programs. *Cornell Hospitality Quarterly*, Vol. 63, No.1, pp.19-32. https://doi.org/10.1177/19389655211017449

Shoemaker, S. (2003). The future of pricing in services. *Journal of Revenue and Pricing Management*, Vol.2, No.3, pp.271-279. https://doi.org/10.1057/palgrave.rpm.5170074

Shoemaker, S., and Lewis, R. C. (1999). Customer loyalty: the future of hospitality marketing. *International Journal of Hospitality Management*, Vol. 18, No.4, pp.345-370. https://doi.org/10.1016/S0278-4319(99)00042-0

Sigala, M. (2020). Tourism and COVID-19: impacts and implications for advancing and resetting industry and research. *Journal of Business Research*, Vol. 117, pp. 312-321. https://doi.org/10.1016/j.jbusres.2020.06.015

Stylos, N., Zwiegelaar, J., and 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. https://doi.org/10.1108/IJCHM-07-2020-0644

Su, N., and Reynolds, D. (2017). Effects of brand personality dimensions on consumers' perceived self-image congruity and functional congruity with hotel brands. *International Journal of Hospitality Management*, Vol. 66, pp.1-12. https://doi.org/10.1016/j.ijhm.2017.06.006

Talón-Ballestero, P., González-Serrano, L., Soguero-Ruiz, C., Muñoz-Romero, S., and Rojo-Álvarez, J. L. (2018). Using big data from customer relationship management information systems to determine the client profile in the hotel sector. *Tourism Management*, Vol. 68, pp. 187-197. https://doi.org/10.1016/j.tourman.2018.03.017

Tanford, S. (2013). The impact of tier level on attitudinal and behavioral loyalty of hotel reward program members. *International Journal of Hospitality Management*, 34, 285-294. doi:10.1016/j.ijhm.2013.04.006

Tanford, S., and Malek, K. (2015). Segmentation of reward program members to increase customer loyalty: The role of attitudes towards green hotel practices. *Journal of Hospitality Marketing & Management*, Vol. 24, No.3, pp. 314-343. https://doi.org/10.1080/19368623.2014.907759

Tanford, S., Shoemaker, S., and Dinca, A. (2016). Back to the future: progress and trends in hotel loyalty marketing. *International Journal of Contemporary Hospitality Management*, Vol. 28, No.9, pp. 1937-1967. https://doi.org/10.1108/IJCHM-05-2015-0237

Tasci, A. D. (2017). A quest for destination loyalty by profiling loyal travelers. *Journal of Destination Marketing & Management*, 6(3), 207-220. doi:10.1016/j.jdmm.2016.04.001

Tripathi, A.K., Sharma, K. & Bala, M. Dynamic frequency based parallel k-bat algorithm for massive data clustering (DFBPKBA). Int J Syst Assur Eng Manag 9, 866–874 (2018). https://doi.org/10.1007/s13198-017-0665-x

Voorhees, C., McCall, M., and Calantone, R. (2011). Customer loyalty: a new look at the benefits of improving segmentation efforts with rewards programs, *Cornell Hospitality Report*, Vol. 11, No. 11, pp.1-18. https://hdl.handle.net/1813/71156

Wu, J. S., Ye, S., Zheng, C. J., and 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. https://doi.org/10.1108/IJCHM-11-2020-1348

Xie, L. K., and Chen, C.-C. (2014). Hotel loyalty programs: how valuable is valuable enough? *International Journal of Contemporary Hospitality Management*, Vol. 26, No. 1, pp. 107-129. https://doi.org/10.1108/IJCHM-08-2012-0145.

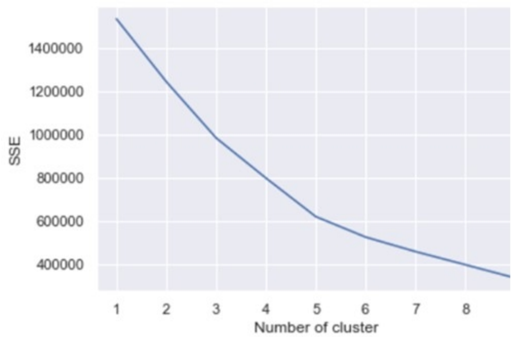
Xu, Y., Gong, C., and Law, R. (2022). Face the competition and take proactive actions: How does neighborhood competition affect hotel online effort?. *International Journal of Hospitality Management*, Vol. 100, pp. 103092. https://doi.org/10.1016/j.ijhm.2021.103092

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. https://doi.org/10.1108/IJCHM-10-2021-1293

Zeng, K. J., Irina, Y. Y., Yang, M. X., and Chan, H. (2022). Communication strategies for multi-tier loyalty programs: The role of progress framing. *Tourism Management*, Vol. 91, pp. 104460. https://doi.org/10.1016/j.tourman.2021.104460

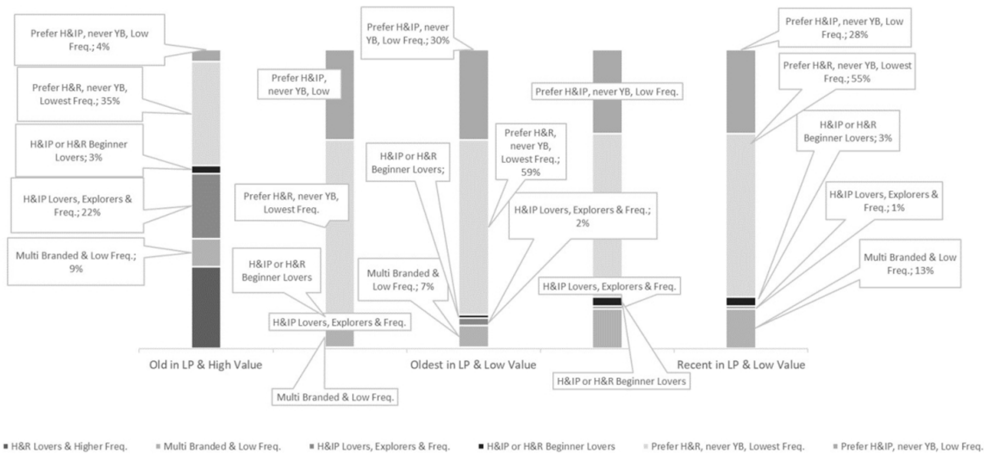
Acknowledgments: This work was supported by national funds through FCT (Fundação para a Ciência e a Tecnologia), under the project - UIDB/04152/2020 - Centro de Investigação em Gestão de Informação (MagIC)/NOVA IMS, research grants UIDB/04521/2020 of the Advance/CSG, ISEG – Lisbon School of Economics and Management; and UIDB/00685/2020 of the Centre of Applied Economics Studies of the Atlantic, School of Business and Economics of the University of the Azores. Editage proofread this work.

Declaration of interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.



Nº of Clusters	**Silhouette Coefficient**	**CH Score**	**DB Index**
2	0.579	89 310	**0.575**
3	0.630	107 039	0.708
4	0.655	116 419	0.655
5	0.679	140 890	0.649
6	**0.680**	**147 229**	**0.632**
7	**0.704**	**149 053**	0.640
8	**0.700**	**157 710**	**0.611**

164x61mm (150 x 150 DPI)



245x113mm (150 x 150 DPI)

Supplementary_material_appendix_1

How it plays out	Traditional frequency	Real loyalty
Objectives	Build traffic, sales, and profits	Build sales, profits, and the brand
Strategy	Offer incentives for repeat transactions	Build personal brand relationships
Focus	A segment's behavior and profitability	An individual's emotional and rational needs and their value
Tactics	Segmented rewards: Transaction status Free/discounted product Collateral product discounts Rewards, such as miles or points Value-added upgrades and add-ons Rewards "menu"	Customer recognition: Individual value, tenure Preferred access, service "Insider information" Value-added upgrades and add-ons Emotional "trophy" rewards Tailored offers/messages
Measurement	Transactions Sales growth Cost structure	Individual lifetime value Attitudinal change Emotional responses

Table 1 – Traditional Frequency Programs versus Real Loyalty Programs (adapted from Shoemaker & Lewis 1999)

Supplementary_material_appendix_2

Variable	Data Type	Description
Card Number	Numeric	Unique identifier of the loyalty member
Card Tier	Factor	Loyalty member tier level: Silver, Gold, Platinum, Corporate
Enrollment Date	Date/time	Date of customer enrollment in the loyalty program
Gender	Factor	Customer gender: Male or Female
Birthday	Date/time	Customer date of birth
Country	Text	Country of customer residence
Total Expenses H&R	Numeric	Customer total expenses within H&R brand
Total Expenses H&IP	Numeric	Customer total expenses within historical and iconic properties brand
Total Expenses PB	Numeric	Customer total expenses within premium brand
Total Expenses YB	Numeric	Customer total expenses within youngest brand
Stay Count H&R	Numeric	Number of reservations made by each customer within H&R Brand
Stay Count H&IP	Numeric	Number of reservations made by each customer within historical and iconic properties brand
Stay Count PB	Numeric	Number of reservations made by each customer within premium brand
Stay Count YB	Numeric	Number of reservations made by each customer within youngest brand
Room Nights H&R	Numeric	Number of room nights stayed by each customer within H&R brand
RoomNights H&IP	Numeric	Number of room nights stayed by each customer within historical and brand
RoomNights PB	Numeric	Number of room nights stayed by each customer within premium brand
RoomNights YB	Numeric	Number of room nights stayed by each customer within youngest brand

Table 2 – Variables used in the analysis

Supplementary_material_appendix_3

Cluster	Stays at H&R	Stays at PB	Stays at H&IP	Stays at YB
0	-0,174	0,251	**5,196**	-0,104
1	**0,241**	-0,315	**-0,417**	-0,148
2	-0,366	-0,19	-0,333	**5,085**
3	-0,424	**2,09**	-0,262	-0,144
4	**-0,456**	-0,309	0,66	-0,148
5	**7,818**	-0,005	-0,0097	-0,082

Table 3 – Brand Preference Centroids

Supplementary_material_appendix_4

Auto-Clustering

Number of Clusters	Schwarz's Bayesian Criterion (BIC)	BIC Change[a]	Ratio of BIC Changes[b]	Ratio of Distance Measures[c]
1	1062504			
2	1052101	-10402,7	1	1,002
3	1029859	-22242,2	2,138	1,109
4	972567,8	-57291,6	5,507	1,218
5	972516,6	-51,209	0,005	1,201
6	972560	43,417	-0,004	1,008
7	972619,8	59,847	-0,006	1,198
8	972631,4	11,598	-0,001	1,506
9	972708,6	77,187	-0,007	1,225
10	972786,1	77,506	-0,007	1,083
11	922363,4	-50422,7	4,847	1,068
12	922447,5	84,031	-0,008	1,118
13	874467,9	-47979,5	4,612	1,024
14	874546,3	78,402	-0,008	1,117

Cluster	N	% of Combined	% of Total
1	45561	11,90%	11,90%
2	337620	88,10%	88,10%
Combined	383181	100,00%	100,00%
Total	383181		100,00%

a. The changes are from the previous number of clusters in the table.

b. The ratios of changes are relative to the change for the two-cluster solution.

c. The ratios of distance measures are based on the current number of clusters against the previous number of clusters.

Table 4 – Brand Preference Auto-Clustering

Supplementary_material_appendix_5

Auto-Clustering

Number of Clusters	Schwarz's Bayesian Criterion (BIC)	BIC Change[a]	Ratio of BIC Changes[b]	Ratio of Distance Measures[c]
1	265626,04			
2	124851,55	-140774,5	1	1,943
3	52416,919	-72434,627	0,515	3,593
4	32273,792	-20143,127	0,143	1,994
5	22183,508	-10090,284	0,072	2,644
6	18383,681	-3799,827	0,027	1,097
7	14922,276	-3461,405	0,025	1,201
8	12043,823	-2878,453	0,02	1,303
9	9841,348	-2202,474	0,016	1,01
10	7660,029	-2181,32	0,015	2,081
11	6625,412	-1034,617	0,007	1,318
12	5846,85	-778,562	0,006	1,167
13	5183,446	-663,404	0,005	1,118
14	4592,559	-590,886	0,004	1,286
15	4138,716	-453,844	0,003	1,195

Cluster	N	% of Combined	% of Total
1	9670	2,50%	2,50%
2	89819	23,40%	23,40%
3	283692	74,00%	74,00%
Combined	383181	100,00%	100,00%
Total	383181		100,00%

a. The changes are from the previous number of clusters in the table.

b. The ratios of changes are relative to the change for the two-cluster solution.

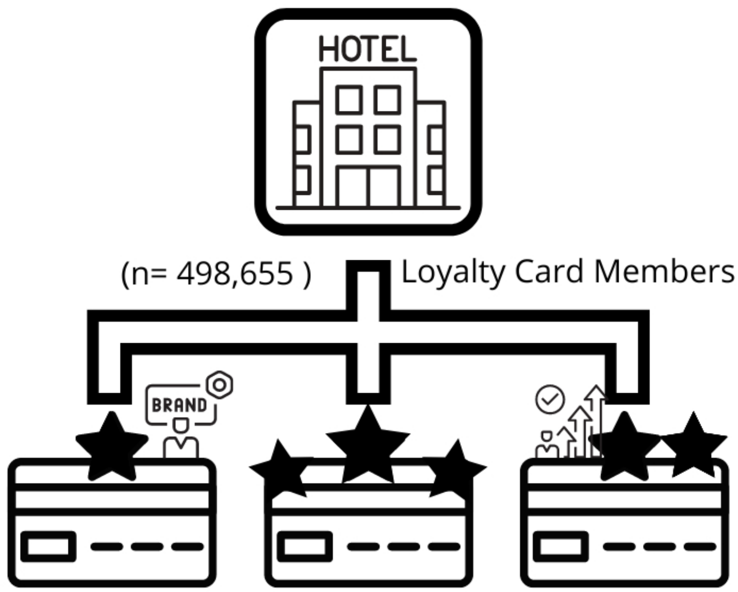
c. The ratios of distance measures are based on the current number of clusters against the previous number of clusters.

Table 5 – Monetary Value Auto-Clustering

Supplementary_material_appendix_6

Clusters	*Segment*	*% guests*	*Average stays at H&R*	*Average stays at PB*	*Average stays at H&IP*	*Average stays at YB*	*Total Average Stays*	*Average Age*	*Average Enrollment Year*
0	Iconic Properties Lovers, Explorers & Freq	2	3,22	1,76	7,46	1,07	8,41	58	2015
1	H&R, never YB, Lowest Freq.	55	1,47	1	1,16		1,48	55	2017
2	Iconic Properties or H&R Beginner Lover	3	2,39	1,5	2,13	1,23	1,68	51	2017
3	Multibrand & Low Freq.	12	2,01	1,26	1,69	1,04	1,62	54	2017
4	Iconic Properties, never YB, Low Freq.	28	1,34	1	1,44		1,53	55	2017
5	H&R Lovers & High Freq	1	16,53	1,94	2,78	1,21	17,14	57	2015
Clusters	*Segment*	*% guests*	*Average expenses at H&R*	*Average expenses at PB*	*Average expenses at H&IP*	*Average expenses at YB*	*Total Average expenses*	*Average Age*	*Average Enrollment Year*
0	Old in LP & High Value	2,5	3 917 €	1 592 €	1 825 €	865 €	4 213 €	59,69	2015
1	Oldest in LP & Low Value	12,9	763 €	562 €	433 €	382 €	697 €	56,2	2013
2	Recent in LP & Low Value	84,6	725 €	620 €	392 €	441 €	626 €	54,2	2017

Table 6 - Clusters' characteristics



628x526mm (38 x 38 DPI)

Manuscript ID IJCHM-05-2022-0646.R2
"The theory-practice research gains from big data: evidence from hospitality loyalty programs"
International Journal of Contemporary Hospitality Management

Response to Editor

I am willing to accept your article contingent upon you making the following important changes/improvements:

1. Respond to one of our associate editors’ and the reviewers' comments and revise your article accordingly.
R: Thank you for allowing us to perform a last revision of our article to make the recommended improvements.

2. Include a structured abstract in page 1 of the main document and make sure that it includes all the required subsections including Purpose, Methodology, Findings, Implications and Originality along with keywords. Implications section is missing.
R: We have included a structured abstract, including implications. We also added a graphical abstract.

3. Make sure to follow IJCHM author guidelines closely: http://emeraldgrouppublishing.com/products/journals/author_guidelines.htm?id=ijchm For example, when there are three or more authors, you need to use Adam et al., XXXX (or Adam et al., XXXX) format for the first time and after. Make sure that you should list references within text in an alphabetical order.
R: We have revised citations and references accordingly.

4. Revisit the Discussion and Conclusions sections one more time to better answer the "So What" question. There should be four sub-sections under this section: (1) Conclusions, (2) Theoretical Implications, (3) Practical Implications and (4) Limitations and Future Research.
R: We have made changes to these sections to improve them.

5. To strengthen your literature review and theoretical implications, you should incorporate more recent and relevant references published in recent months/years.
R: Following your recommendation, we have included more recent and relevant references.

6. Cross check all references within text with your reference list and make sure that all references used in within text are listed in your reference list and remove any uncited references from the reference list. You must also make sure that each reference in your reference list is accurate and complete in terms of authors’ names, title, volume number, issue number, pages, publisher etc.
R: We have cross-checked all references within the text with our reference list.

7. Run your article through iThenticate, Crosscheck or any similar software to check the similarity between your study and previous studies. Try to minimize similarity percentage below 1% with any previous study 9. After you run your article's final version through iThenticate or other similar software, you should upload the similarity report to the system for our records.
R: We ran our article through Turnitin and identified a similarity of 4%, with one source at 1% and the others at <1%. We have uploaded the respective similarity report.

8. Keep your article below 9000 words including references, tables and figures. If your article is longer than 9000 words, you may consider submitting supplementary material alongside your article

as Emerald now publishes supplementary materials.
R: Our article has a length of below 9,000 words.

9. Make sure that the flow of your article is improved. We do not prefer short paragraphs with one-three sentences or long paragraphs longer than half page.
R: Thank you for this remark. We took your recommendation into account when revising and improving our article.

10. Proofread your article one more time and also you may ask a technical writer/copy editor to proofread it for you. After the manuscript is accepted, we will not ask you to proofread it again. In short, after I send you an official acceptance e-mail, you will not be able to make any further changes in your manuscript.
R: The article was proofread by a professional copy-editing organization (Editage). We submit the respective editing certificate.

11. Make sure you have all the coauthors, acknowledgements, tables and figures included in your submission. You must also make sure that order of authors, their full names, their affiliations and email addresses are accurate. We cannot make any changes after the paper is accepted. PLEASE do not thank the editor, guest editors or reviewers in your acknowledgement statement.
R: The coauthors, acknowledgements, tables, and figures are all included in our submission. The authors' order, full names, affiliations, and email addresses are accurate.

12. You cannot add a new coauthor(s) at this stage unless there is a legitimate justification.
R: No new coauthor was added.

13. Submit a clean version of your paper. You don't need to show/highlight all the changes made in the paper. I will read its final version anyway.
R: Following your recommendation, we do not show/highlight all the changes made in the manuscript, thus submitting a clean version of our paper.

14. In addition to responding to the reviewers' comments, you should prepare and submit a brief report showing how you have responded to the above requests as well as the associate editor's comments. When revising your submission, you don't need to show/highlight all the changes made in the paper. I will read its final version anyway.
R: We wish this report and the paper meet your expectations.

Comments from Editor/Associate Editor
This is an interesting study on an important topic. The study can benefit from a strong copy-editing. The theoretical foundation and theoretical implications should be improved further. The following studies can help the authors with this task. Below studies are just suggestions and the authors may find similar relevant and recent studies.

Afaq, A., Gaur, L. and Singh, G. (2022), "Social CRM: linking the dots of customer service and customer loyalty during COVID-19 in the hotel industry", International Journal of Contemporary Hospitality Management, Vol. ahead-of-print No. ahead-of-print. https://doi.org/10.1108/IJCHM-04-2022-0428

Liu, Y. and 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. https://doi.org/10.1108/IJCHM-06-2020-0592

Zarezadeh, Z.Z., Rastegar, R. and 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. https://doi.org/10.1108/IJCHM-10-2021-1293

Buhalis, D., O'Connor, P. and 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. https://doi.org/10.1108/IJCHM-04-2022-0497

Guan, J., Wang, W., Guo, Z., Chan, J.H. and Qi, X. (2021), "Customer experience and brand loyalty in the full-service hotel sector: the role of brand affect", International Journal of Contemporary Hospitality Management, Vol. 33 No. 5, pp. 1620-1645. https://doi.org/10.1108/IJCHM-10-2020-1177

Liu, J., Yu, Y., Mehraliyev, F., Hu, S. and 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. https://doi.org/10.1108/IJCHM-06-2021-0749

Mehraliyev, F., Chan, I.C.C. and 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. https://doi.org/10.1108/IJCHM-02-2021-0132

Jeong, M., Shin, H.H., Lee, M. and Lee, J. (2022), "Assessing brand performance consistency from consumer-generated media: the US hotel industry", International Journal of Contemporary Hospitality Management, Vol. ahead-of-print No. ahead-of-print. https://doi.org/10.1108/IJCHM-12-2021-1516

Mariani, M. and Borghi, M. (2021), "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. https://doi.org/10.1108/IJCHM-06-2020-0548

Mariani, M. and Baggio, R. (2022), "Big data and analytics in hospitality and tourism: a systematic literature review", International Journal of Contemporary Hospitality Management, Vol. 34 No. 1, pp. 231-278. https://doi.org/10.1108/IJCHM-03-2021-0301

Lee, M., Kwon, W. and Back, K.-J. (2021), "Artificial intelligence for hospitality big data analytics: developing a prediction model of restaurant review helpfulness for customer decision-making", International Journal of Contemporary Hospitality Management, Vol. 33 No. 6, pp. 2117-2136. https://doi.org/10.1108/IJCHM-06-2020-0587

R: A professional copy-editing was performed to the whole manuscript by Editage. We improved both the theoretical foundations and theoretical implications of this research. We added some of the suggested studies. Thank you.