Journal of Tourism Analysis: Revista de Análisis Turístico Vol. 29 No. 1, Año 2022
How to cite this work: Pons Vivies, P.J., Morro Ribot, M., Mulet Forteza, C., & Valero, Ó. (2022). How to increase hotel performance through complimentary services. A comparison between machine learning and classical segmentation techniques. Journal of Tourism Analysis, (29), 1, 131–159. https://doi.org/10.53596/jta.v29i1.403

How to increase hotel performance through complimentary services. A comparison between machine learning and classical segmentation techniques

Pere Josep Pons Vives[1]

Mateu Morro Ribot[2]

Carles Mulet Forteza[3]

Óscar Valero[4]

[1] Universitat de les Illes Balears. p.pons2@estudiant.uib.eu

[2] Hotelbeds. mateumorroribot@gmail.com

[3] Universitat de les Illes Balears. carles.mulet@uib.es

[4] Universitat de les Illes Balears. o.valero@uib.es

Abstract

The consumption patterns of tourists in hotels, how they consume the different services offered, have been extensively analyzed through traditional cluster analysis. However, the changes in consumption preferences have forced hotel firms search for proper techniques. This study compares the classic K-means cluster analysis with the Random Forest technique. A comparison is conducted by segmenting 2,256 bookings, showing that proposed technique provides much higher accuracy and more profitable clusters. This is due to its greater flexibility when segmenting, because the K-means is based on an extremely rigid membership function. From a theoretical point of view, this study contributes to the literature by introducing diversification strategies in sun-and-sea luxury hotels to overcome

***Reception:** 26.09.2021   Revision: 07.02.2022   **Acceptance:** 11.02.2022   **Publicationn:** 22.05.2022*

the decrease in profitability caused by destination maturity. From a practical point of view, it allows hotel managers to effectively implement diversification strategies because Random Forest technique can explain the consumption pattern of 98.47% of the customers.

Keywords: Random Forest; hotel performance; diversification; consumption pattern; hotel management.

Resumen

Las pautas de consumo de los turistas en los hoteles, como consumen los distintos servicios ofrecidos, han sido ampliamente analizadas a través de los clústeres de análisis tradicional. Así, las empresas hoteleras, mediante un adecuado análisis de los patrones de consumo pueden mejorar su rentabilidad. Este artículo compara los resultados obtenidos por el K-means clásico con el Random Forest a la hora de explicar las pautas de consumo de los clientes. Para comparar estas dos técnicas se usa una muestra de 2,256 reservas. Como resultado se obtiene que el Random Forest ofrece una mayor precisión, además de encontrar unos segmentos más rentables. Esto se debe a su mayor flexibilidad a la hora de segmentar, ya que el K-means se basa en una función de pertenencia extremadamente rígida. Desde un punto de vista teórico, este artículo contribuye a la literatura introduciendo estrategias de diversificación en los hoteles de lujo de sol y playa para superar la disminución de la rentabilidad provocada por la madurez del destino. Desde un punto de vista práctico, permite a los gerentes de hoteles implementar estrategias de diversificación de manera efectiva porque la técnica Random Forest puede explicar la pauta de consumo del 98,47% de los clientes.

Palabras clave: Random Forest; rentabilidad hotelera; diversificación; pautas de consumo; gestión hotelera.

1 Introduction

Hotel performance has been widely studied in tourism literature (Hua et al., 2020). Researchers have focused on two complementary strategies to improve hotels performance: expansion and diversification. Expansion strategies imply income growth through the addition of hotel establishments or rooms. Different expansion strategies, which involve the use of property, leasing, franchise, and management contracts, represent different levels of effort in terms of management and investment (Martorell & Mulet-Forteza, 2010). Diversification takes advantage of underused resources and economies of scope to obtain resources and create synergies between departments (Hsu & Liu, 2008).

The relationship between hotel performance and diversification strategies has received less interest in the hotel industry. A chief reason is that the industry based its business model on Fordism until the maturity phase of the business (Butler, 1980). The main characteristics of the model are a lack of product differentiation together with rigidity and high standardisation. It assumes that tourists travelling to sun-and-beach destinations only looked for sunny weather and idyllic beaches (Aguiló & Juaneda, 2000; Bujosa, Riera, & Pons, 2015). Under this approach, product differentiation is not necessary to maintain hotels' performance as the number of arrivals increases, and the room price remains at least stable. Therefore, researchers and practitioners continue to use classical segmentation techniques as K-means with variables constructed from personal opinions/judgments, or other psychographic variables to segment tourists. Therefore, it is necessary to have certain degree of destination maturity and competence between establishments to allow hotels to look for diversification (Horner & Swarbrooke, 2016).

Nowadays tourists' preferences and demands have become more complex. In addition, Mediterranean sun-and-beach is in an advanced maturity state which, among other factors, implies a high degree of competence between hotels. These two factors may treat hotels performance (Poon, 1993; Deyà & Tirado, 2011). Tourists not only want to stay all day at the beach and then go back to the hotel room, but they also seek for other activities beyond sunbathing. They look for gastronomic experiences, sports, and culture. Vacations have become multi-purpose trips, tourists combine beaches with visits to other attractions (Bujosa et al., 2015). In addition, tourists share and look for activities using social networks and review platforms. In this sense, they get and provide information to tourists and business managers that can be used for product adaptation (Lee et al., 2020; Moro et al., 2019).

In this way, Wilkins et al. (2007) using a mixed-method approach find that Food & Beverage (F&B) services are a key factor when determining customer satisfaction as well as hotel performance. Similarly, Xu and Li (2016) applied a latent semantic analysis to luxury hotel reviews found that F&B services highly influence tourists' satisfaction. Mun, Woo, and Paek

(2019) find that not only F&B services at luxury hotels increase revenue and customers' satisfaction. This diversification may provide better performance of the rooms department in terms of occupancy, average daily rate (ADR), revenue per available room (RevPAR), and gross operating profit per available room (GOPPAR). Therefore, a better understanding of tourist patterns propitiates the implementation of complementary and better-managed services, as well as generating synergies with existing departments. In addition, it contributes to better use of resource-offering services in unused spaces. However, this implies financial and operational risks. If the company is not capable of satisfying the needs of the customers of the core business and the new one as well as create synergies, the cost may be higher than the revenue from the diversification. Therefore, understanding consumption patterns is crucial.

Motivation is the starting point of consumption, which is the basis of consumers' behavioural analyses. That is, the research field on how and why different groups of consumers behave as they do (Eagles, 2016; Rita et al., 2018). The other capstone of behavioural analysis is individuals' characteristics. In this way, the different consumption behaviours of different segments are referred to in tourism literature as tourism consumption patterns (Horner & Swarbrooke, 2016). Tourist consumption patterns are analysed from different scales and perspectives: macro, micro, and nanoscale (Bujosa, Riera, & Pons, 2015). Although there is already a fuzzy border between scales it can be stated that while macro-scale comprises consumption choices that tourists make at the origin country before traveling. The micro-scale focuses on the different tourist choices between destinations or within a destination (Chang et al., 2018). Finally, the consumption patterns of tourists of a specific attraction or local business, such as a beach or hotel. Thus, hotel managers must focus on the behaviour of tourists lodged at their hotels to allocate resources to satisfy the needs of the most profitable segment.

Customer segmentation and clustering have been widely studied in the hospitality industry. In this sense, the data structure and input variables are not suitable for finding new segments of customers because of their computational limitations. Machine learning techniques may provide a more flexible tool can help practitioners to better implement diversification strategies.

This study analyses the F&B consumption patterns of tourists lodged at a luxury hotel located in a mature sun-and-beach destination with the objective that a better understanding of such consumption patterns results in an improved profitability of the hotel establishment. To the best of our knowledge, classical techniques do not provide suitable tools owing to their rigidity when incorporating new information which differs from the aggregated mean. Therefore, Machine Learning (ML) techniques may be able to better predict the complexity of tourists' behavior. Considering this, the study provides empirical evidence on the following:

- Can machine learning techniques overcome K-means when explaining complementary services consumption patterns of tourists?
- Can machine learning techniques help hotel chains implement effective diversification strategies that result in an improvement in their profitability?
- Which are the characteristics of different customer segments?

In this study, the analysis of F&B consumption patterns of clients lodged at a luxury hotel located in Mallorca, which have visited their Sky Bar, were analysed. The data were provided directly by a hotel chain. Thereby furnishing individualized and confidential information not available in any database.

2 Literature review

Based on the Fordist Model, hotel chains increased their performance through expansion and cost control. Under this approach, performance efficiency is measured by comparing observed and optimum costs and revenue subject to constraints on prices and qualities (Grosskopf, 1993). This leads to the creation of extremely efficient multinational hotel chains (Martorell, 2012). However, in mature destinations, where cost reduction is difficult and tourist preferences have evolved towards a multi-purpose trip, an efficient strategy may be combined with service diversification. Under this scenario, hotels may choose to develop new businesses that are either related or unrelated to their current businesses. However, few studies consider product diversification strategies hotel (Oltean & Gabor, 2016). Thus, Chesters (2017) identified F&B service diversification as a key variable in hotel performance.

Chari et al. (2019) highlights that the effect of diversification on performance is based on the combined effect of synergies and the possibility of sharing resources and knowledge between the different business units that may lead to higher performance. However, costs may be greater than benefits generated by synergies at some diversification level (Grant et al., 2017). The number of studies in the hotel sector is limited, and the results are mixed. Chen and Chang (2012) examined the effects of Taiwanese hotel diversification on F&B strategies on their growth and profit stability. Researchers found that hotels with total revenue generated mostly from F&B services tend to have higher growth in profit margins, but also suffer higher instability. In this regard, Erkuş-Öztürk (2016) also found a tendency toward service diversification when examining sector data of hotels in Turkey, an example of a sun-and-sea mature destination. The authors state that company size and sector-specific knowledge (intra-industry investments and experience of hotel workers) are important variables in determining the success of diversification strategies.

Applying a holistic perspective, Lei (2019), uses a stochastic frontier analysis and finds that revenue diversification across the room, F&B, and other services efficiency is explained by general structure, technological efficiency, workers' capabilities, and hotel characteristics. Park and Jang (2012) introduce non-linearity in diversification profitability analysis. They find that unrelated diversification increases profitability up to a certain level. However, beyond that level, unrelated diversification decreases profitability, which implies that at high levels of unrelated diversification, there is a loss of control and effort due to the distance from the primary business. They also find that at low levels of related diversification, synthetic-related business risk is larger than the risk reduction effect. This means that at low levels of related diversification, synthetic-related business risk is larger than the risk reduction effect. However, the number of studies in the hotel sector is limited, and the results are mixed (Lee & Jang, 2007).

A better understanding of customer preferences and behaviour may be key for hotels when implementing diversification strategies. Therefore, machine learning techniques may provide better customer information that avoids the negative effects derived from internal transaction costs, allowing the implementation of stronger synergies between departments (Park & Jang, 2012).

However, there is a gap in the literature analysing the consumption patterns of tourists in hotels where they are lodged. The existing literature has focused on the choices of tourism products, routes, travelling cognition, spatio-temporal distribution, and mental maps. This may be because obtaining significant results is necessary to obtain large and accurate data sets on tourist behaviour. In this sense, De Cantis et al. (2016) segment cruise tourists visiting Palermo based on their consumption patterns. They combine traditional interviews with sociodemographic questions (age, gender, travel group dimension, etc.) with global positioning system (GPS) information (length of tour, duration of tour, number of visited attractions, average speed, etc.). Delivering a GPS device that tracks tourists' movements during their visit to the city, researchers obtain a higher response rate than using a travel diary as well as more accurate data.

The main motivation for this study is the need for hotel managers to better understand the consumption patterns of their clients related to their additional services in order to increase the profitability of their hotel establishments.

3 Methodology

There are two main approaches in market segmentation literature: supervised and unsupervised market segmentation. The difference between them is that supervised segmentation requires the researcher to first choose the variables of interest and then classify individuals according to that designation, for example, dividing the age variable in c ranges and trying to explain the characteristics of the individuals assigned to this group (Qu et al., 2016). In this regard, common practices of supervised market segmentation are expenditure market, motivation, and visitation (Mok & Iverson, 2000; Sung et al., 2015).

In contrast, unsupervised segmentation only requires that researchers choose a range of interrelated variables and then cluster individuals into groups. These groups must satisfy the criterion that the similarity measure between individuals belonging to a group is high and the between-group similarity is low (Ernst & Dolnicar, 2017; Wierzchoń & Kłopotek, 2018). In other words, supervised segmentation does not require any previous analysis, whereas unsupervised segmentation requires an explorative analysis of the variables to determine the different segments.

Several techniques are used in tourism market segmentation, ranging from elementary ones, such as statistically dividing the sample into more complex multivariate ones (Wierzchoń & Kłopotek, 2018; Mok & Iverson, 2000). However, the most commonly used techniques include factor analysis (Sharma & Nayak, 2020), cluster analysis, latent class analysis (Ferrer-Rosell et al., 2016), (Maingi et al., 2016; Pivčević et al., 2020), random forest (Brida et al., 2018), and more recently, neural networks (Bigné et al., 2019; Moral-Cuadra et al., 2021). In the remainder of this paper, we focus on K-means cluster analysis and random forest tourism segmentation.

The first step of cluster analysis is to determine the number of clusters in which the demand is divided. There are two ways to determine the number of clusters: based on information provided by practitioners and experts or through rigorous mathematical-based techniques (Aguiló & Rosselló, 2005; D. Xu & Tian, 2015). Among such methodologies, the following should be highlighted: Schwarzs Bayesian information criterion (BIC), Akaike information criterion (AIC), and elbow method indicators (Wierzchoń & Kłopotek, 2018). When determining the number of clusters, the researcher must seek interpretable and coherent results (Bujosa et al., 2015). Therefore, it is useful to check the results obtained by the mathematical-based technique with the opinions provided by experts.

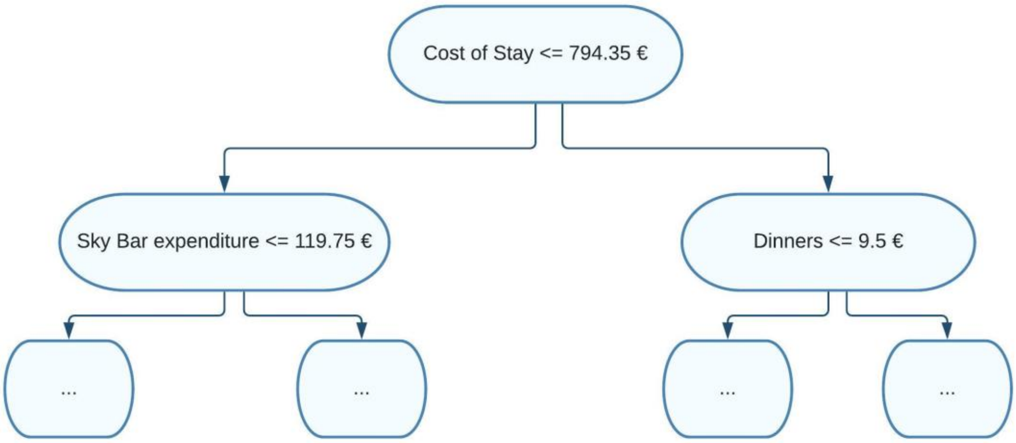
Among the segmentation algorithms, K-means is the most popular in the social sciences (Jain, 2008; Wierzchoń & Kłopotek, 2018). However, it has several handicaps that facilitate the adaptation of other algorithms in social sciences, which are derived from their rigid membership function. It is sensitive to outliers and noise, and it cannot guarantee convergence to a solution (Jain, 2008). In addition to this limitation, the availability of teras of data on tourist behaviour, as well as the rise in computational power, makes it easy to use more efficient classification algorithms. In the next phases, our purpose is to apply both K-means and Random Forest to nanoscale tourism segmentation in order to verify the superiority of the accuracy of the Random Forest.

The K-means hierarchical clustering algorithm aims to partition n observations into k clusters. Each observation belongs to the cluster with the nearest centroid. In the initial stage, the centroid is located randomly, and then the algorithm starts. It moves the centroids until there is a minimum distance between the observations of the cluster and its centroids (Jain, 2008).

The Random Forest technique consists of numerous individual decision trees that operate as a whole (Breiman, 1999). Each tree in the random forest splits out a class prediction, and the class with the most votes become the model prediction. A low correlation between tree structures operating as a committee is the key to performance (Xuan et al., 2018). Random Forest is a supervised machine learning algorithm that can be used for both classification and regression (Brida et al., 2018; Dieu et al., 2020). A decision tree is a predictor h: X→Y, which predicts the label associated with an instance x by travelling from a root node (where the tree starts) of a tree to a leaf every one of the splits. At each node on the root-to-leaf path, the successor child is chosen based on the splitting of the input space (Shalev-Shwartz & Ben-David, 2014).

Splitting is usually based on one of the features of x or a predefined set of splitting rules. The leaf contained a specific label. An example of a decision tree for expenditure propensity is given as follows:

Figure 1. Decision Tree example



Source: Author elaboration

As shown in Figure 1, a decision tree is a flowchart-like structure in which each node represents a test on an attribute, depending on the results on each level a tourist is classified as Non-Spender, Standard, or Spender. For example, extracting a single decision tree from the Random Forest applied in the analysis, the first variable that it assesses is Cost of Stay. If it is lower than 794,35 € then it is assessed using the Sky Bar expenditure, and if it is higher, it uses the number of dinners. The process continues *N* times for the *N* trees of the forest.

That is, it moves to the right, centre, or left child of the node based $1_{[x_{i<\theta}]}$ where $i \in [d]$ is the index of the relevant feature and θ ∈ is the threshold. In such cases, a decision tree can be considering as a splitting of the instance space, $X = R^d$, into cells, where each leaf of the tree corresponds to one cell. It follows that a tree with k leaves can shatter a set of k instances. Hence, if we allow decision trees of arbitrary size, we obtain a hypothesis class of infinite dimensions. To avoid overfitting, the minimum description length (MDL) is used (Shalev-Shwartz & Ben-David, 2014). The most frequently used attribute selection measures in decision tree induction are the information gain ratio criterion and the Gini Index (Breiman, 1999). For a given training set T, selecting one case at random and saying that it belongs to some class C_i, the Gini index can be written as:

$$\sum \sum_{i \neq j} (f(c_i, T)/|T|)(f(c_j, T)/|T|)$$

The Random Forest classifier used in this study consists of using randomly selected features or a combination of features at each node to grow a tree. Bagging, a method to generate a training dataset by randomly drawing with replacement N examples, where N is the size of the original training set, was used for each feature-feature combination selected (Ketkar, 2017).

Although K-means is still very popular among academics and managers, it presents a series of limitations that arise from its restrictive membership function. Thus, this rigidity is overcome by the Random Forest which can handle missing values, automatically process outliers, and work well with both continuous and categorical data (Chattopadhyay and Mitra, 2020). For this reason, researchers assume that Random Forest may be a suitable option for academics and practitioners.

4 Data description

This study compares the segmentation capacity of classical K-means and Random Forest. To do so, 2,256 room nights between March 1, 2019, and October 31, 2019 were gathered from the hotel company database. Based on the opinion of the hotel chain marketing experts, three groups have been pre-defined: Non-Spenders, Standard, and Spenders. Regarding the 780 visits to the Sky Bar, 34.60% are considered to have a low propensity to spend (Non-Spenders), 42% (Standard) have a medium propensity, and 23.4% have a high propensity. Selected tourists had to be lodged for at least one night at the hotel.

Table 1 describes the variables used as input in both segmentation techniques.

Table 1. Variable description

Variable name	Variable description
Length of Stay	Number of days the tourist stays at the hotel
Cost of stay	Cost of the accommodation per room
Dinners	Number of customers per visit to the Sky Bar
Sky Bar expenditure	Total amount spent at the Sky Bar by each reservation
Sky Bar visits	Total number of visits made to the Sky Bar during the stay

Source: author elaboration

Table 2 shows that the average length of stay is 3.5 nights. This result is consistent with the average length of stay for this hotels category reported by official institutions (IBESTAT, 2020). The minimum length of stay is one night, and the maximum is fifteen nights corresponding to a long stay reservation. Table 2 shows that the average cost of the stay is 1,008.97 €. The minimum of this variable is zero because on occasions in which there has been an error, being a luxury hotel, the company does not charge the accommodation to the client. Regarding the consumption pattern of the establishment's Sky Bar customers, on average it corresponds to reservations of 5 people who spend an average of € 93.58 and who visit the bar about three times during their stay.

Table 2. Data description

Variable name	Minimum	Maximum	Mean	Standard deviation
Sky Bar visits	1	27	3.42	3.08
Length of Stay	1	15	3.56	2.43
Cost of stay	0	11,050 €	1,008.97 €	759.26
Dinners	1	47	5.14	4.96
Sky Bar expenditure	2 €	830 €	93.58 €	91.73 €

5 Results

The segmentation presented in this study provides a model that works independently from the season (Fleischer & Pizam, 1997). To obtain significant results, 1,500 iterations of each model, Random Forest and K-means, were run. Following the guidelines of Jain (2010), the data set was randomly divided into training and testing (30% of the sample) using Python 3 (Van Rossum & Drake, 2009).

To compare the results obtained, the accuracy was used according to the literature. This metric is especially used to evaluate the classification models. Thus, we can define it as the quotient between correct predictions over total predictions (Visa et al., 2011). literature, the accuracy has been used. This metric is especially used to evaluate classification models. Thus, we can define it as the quotient between correct predictions over total predictions (Visa et al., 2011).

$$Accuracy = \frac{Correclty\ predicted}{Totally\ predicted}$$

Accuracy allows comparing the ability to compare the predictions of the two models even if the models have different statistical natures. The following table (Table 3) shows the results of the 1,500 iterations made with each of the classification models.

Table 3. Results

Segmentation technique	Iterations	Average of accuracy	Standard Deviation of accuracy
Random Forest	1,500	0.9847	0.0028

K - Means	1,500	0.5965	0.0014

Source: author elaboration

It can be appreciated that the Random Forest technique has better accuracy than the classic K-means. Moreover, a very low standard deviation is observed, which indicates a consistent average of accurate results. From the 1,500 experiments, Random Forests achieved an average accuracy of 98.47%. Although surprisingly high, it has been demonstrated that this algorithm is a superior classifier; it is even used for crime prediction (Lin et al., 2017; Xia et al., 2021). In this regard, Xuan et al. (2018) used a Random Forest to classify Internet transactions with a credit card and obtained an accuracy of 98.67%.

In the same way that Guo et al. (2019), the confusion matrix from Table 4 shows, in one of the 1,500 iterations, how 98.47% of the predicted labels are correct.

Table 4. Confusion matrix

		Real Label		
		Non-Spenders	**Spenders**	**Standards**
Predicted Label	Non-Spenders	239	0	1
	Spenders	0	148	1
	Standards	2	1	285

Source: Author elaboration

As shown in Table 4, true positives (correctly predicted values: 239, 148, and 185) were dominant in the confusion matrix of the Random Forest. For instance, Random Forest only classifies two Non-Spenders as Standards.

Table 5. Random Forest clusters' description

Cluster	**Minimum**	**Maximum**	**Mean**	**Standard deviation**
Non-Spenders				
Sky Bar visits	1	5	1.6	0.9
Length of Stay	2	4	2	0.8
Cost of Stay	0 €	792 €	436.8 €	206.50 €

Dinners	1	5	2.3	1.3
Sky Bar expenditure	2 €	119.5 €	39.8 €	28.7 €
Standards				
Sky Bar visits	1	8	3.1	1.8
Length of Stay	1	6	3.5	1.1
Cost of Stay	0 €	1,770.0 €	1,016.9 €	329.4 €
Dinners	1	12	4.6	2.8
Sky Bar expenditure	2.0 €	199.0 €	79.3 €	50.3 €
Spenders				
Sky Bar visits	1	27	6.4	4.5
Length of Stay	1	33	5.8	3.3
Cost of Stay	396.9 €	11,050 €	1,918.9 €	1,230.6 €
Dinners	1	47	6.3	7.7
Sky Bar expenditure	5.5 €	609.5 €	187.5 €	122.6 €

Source: author elaboration

Table 5 describes the three clusters generated by the Random Forest: Non-Spenders (34,6% of the sample), Standard (42.0%), and Spenders (23.4%). Clusters are coherent with the expectations of researchers. There is an ascending scale in the mean of the variables from Non-Spenders to Spenders.

Non-spenders, on average, visit the sky bar 1.6 times during their two-day stay at the hotel. On average, the expenditure on accommodation is 436.8 € with a maximum of 790, while their expenditure at the sky bar is only 39.8 €, where they usually go in groups of two people. The Standards almost doubles the number of sky bar visits and doubles the cost of the stay as well as the number of visitors to the sky bar. The total expenditure in F&B Sky Bar is 79.3 € in a maximum of eight visits per stay.

The Spenders visit the Sky Bar 6.4 times each time they are lodged at the hotel, spending 1,918.9€ on accommodation, four times more than the Non-Spenders and almost double the Standards’ expenditure. They spend, on average 187.5 € on the Sky Bar in groups of 6.3. In terms of expenditure, it is 50.01% more than Non-Spenders and more than double the Standards.

Table 6 shows the results of the results of the K-means segmentation.

Table 6. K-means clusters’ description

Cluster	Minimum	Maximum	Mean	Standard deviation
Non-Spenders				
Sky Bar visits	1.00 €	5.00 €	1.60 €	0.90 €
Length of Stay	1.00	4.00	2.00	0.90
Cost of Stay	- €	800 €	446.50 €	213.70 €
Dinners	1.00	5.00	2.20	1.30
Sky Bar expenditure	2.00 €	119.50 €	41.60 €	30.00 €
Standards				
Sky Bar visits	1 €	8.00 €	3.10 €	1.80 €
Length of Stay	1.00	6.00	3.60	1.10
Cost of Stay	- €	1.80 €	1,020.30 €	344.60 €
Dinners	1	12	4.60	2.80
Sky Bar expenditure	2.00 €	199.60 €	81.40 €	51.50 €
Spenders				
Sky Bar visits	1 €	27 €	6.60 €	4.80 €
Length of Stay	1.00	33.00	5.80	3.60
Cost of Stay	151.20 €	11.050 €	1.818 €	1.431.40 €
Dinners	1.00	47.00	10.40	8.30

Cluster	Minimum	Maximum	Mean	Standard deviation
Sky Bar expenditure	2.00 €	830.20 €	192 €	126.00 €

Source: author elaboration

The three clusters are composed of 35. 60% Non-Spenders, 42.39% Standards, and 22.01% of Spenders. The percentage difference in the size of the clusters is small, 1.4% in the case of Spenders. However, in the assignment of individuals to each cluster, the superiority of the Random Forest is seen. Thus, the standard deviation of the variables in each cluster is lower in the Random Forest than in the K-means (Tables 5 and 6). The significant difference between the results is on Spenders group. Although both the K-means and Random Forest Spenders groups have a similar size, Random Forest is capable of grouping customers with a higher expenditure in accommodation (1.918.90 €) and Sky Bar expenditure with equal average length of stay (5.8 days).

Through machine learning segmentation techniques, managers can adapt pricing and packages to the preferences of the segment that better fit the hotel strategy. In this way, an increase in revenue at the Sky Bar, cost optimisation, the generation of synergies with other departments, and the increase in tourist satisfaction operate together to boost hotel performance.

6 Conclusion

This study aimed to propose a new technique for segmenting tourists to improve the performance of the establishment. In addition, researchers have shown that the Random Forest works regardless of whether the season is low, medium, or high. This application is an example of how hotels can improve their performance through diversification strategies. They can increase revenue with new services, increase customer satisfaction with the establishment, and optimise resources.

This study answers the questions presented in the introduction. Regarding the first question, this study confirms that machine learning techniques can explain complementary consumption patterns better than classical K-means. Regarding the second question, Random Forest has a higher accuracy and lower standard deviation of the variables and better describes the tourist' consumption patterns. In fact, beyond providing better accuracy, the individuals belonging to the cluster Spenders of the Random Forest have an average expenditure in accommodation and, in Sky Bar expenditure is higher than K-means. Therefore, a better understanding of customers' consumption patterns facilitates hotel chains in the implementation and monitoring of an effective diversification strategy.

From a theoretical perspective, this article contributes to the literature considering the sun-and-beach hotel not only where tourists rest before a day of beach, considering that it is a leisure centre with multiple business units and spaces (nodes), which are interconnected as an interrelated complex network and generate synergies between them (Jackson, 2010). This article proposes and demonstrates new machine-learning techniques. In this sense, for business managers of a hotel in a mature destination, it is more important to understand how tourists interact with hotel complimentary offers in the Sky Bar than to know the number of tourist arrivals. The number of arrivals is more or less stable according to the consolidation stage of the destination (Deyà & Tirado, 2011). This article also provides a new tool for managers to make diversification away from the core business easier. That is, a long period of experience is not necessary in the new business to understand the behaviour of customers.

From a theoretical perspective, this study contributes to the literature on hotel performance because it can achieve high accuracy with a few explanatory variables as well as introduce a technique that does not require a strict process of data cleaning because of its capacity to handle missing values and continuous and categorical variables at the same time (Xiang et al., 2017). Beyond that, Random Forest can take relations between variables and observations into account, that other algorithms, besides machine learning, cannot (Pal, 2007). However, it has the advantage of avoiding overfitting. Another contribution is the optimisation of resources. When customers' consumption patterns are well known, services can be reinforced at peak times or only offer certain products that are better for the interests of the customer and the company, such as a reduced menu.

This work has important practical implications for hotel managers, as the study results can help them, directly and indirectly, compensate for the RevPAR drop. First, it provides a tool that can accurately segment F&B service consumers. Second, higher and better usage of unexploited hotel spaces, such as terraces, directly increases the company's revenue (Yeh et al., 2012). Finally, although the F&B and rooms divisions are separate parts of the business, high customer satisfaction with the F&B service may lead to higher room prices (Xu & Li, 2016). Therefore, as indicated by Ribeiro et al. (2019) hotel managers must have a global vision of the organization, as the performance of the parts fully impacts the performance of the hotel.

In future research, it may be interesting to take a holistic perspective and consider all of the business units of a hotel (Rooms Division, Rooftop, Spa, Restaurant, Bar, Pool Bar, etc.) and investigate how the consumption patterns of tourists interact with all of the outlets. More specifically, future research must not only segment customers by how they utilize an outlet but to segment them by how they utilize all outlets and how the utilization of an outlet is related to that of others. In addition, information from opinions and customer experiences published on social networks and review platforms may be incorporated.

Bibliographic references

Books

Breiman, L. (1999). *Random forests-Random features* (UC Berkeley TR567, Ed.).

Horner, S., & Swarbrooke, J. (2016). Consumer Behaviour in Tourism. Routledge. Retrieved from https://www.taylorfrancis.com/books/9781315795232

Martorell, O. (2012). The growth strategies of hotel chains: Best business practices by leading companies. In *The Growth Strategies of Hotel Chains: Best Business Practices by Leading Companies*. https://doi.org/10.4324/9780203820810

Poon, A. (1993). *Tourism, technology and competitive strategies.* CAB international.

Pöyhönen, P. (1963). A Tentative Model for the Volume of Trade between Countries. *Weltwirtschaftliches Archiv*, *90*, 93–100. Retrieved from https://www.jstor.org/stable/40436776

Van Rossum, G., & Drake, F. L. (2009). Python 3 Reference Manual. Scotts Valley, CA: CreateSpace.

Wierzchoń, S., & Kłopotek, M. (2018). Modern Algorithms of Cluster Analysis. https://doi.org/10.1007/978-3-319-69308-8

Books chapters

Dann, G. (2014). Why, oh why, oh why, do people travel abroad? In *Creating experience value in tourism* (pp. 48–62).

D'urso, P. (2007). Fuzzy Clustering of Fuzzy Data. In *Advances in Fuzzy Clustering and its Applications* (pp. 155–192). https://doi.org/10.1002/9780470061190.ch8

Grosskopf, S. (1993). The measurement of productive efficiency: Techniques and applications. In *Efficiency and productivity* (pp. 160–194).

Jackson, M. O. (2010). Social and Economic Networks. In *Social and Economic Networks*. https://doi.org/10.1093/acprof:oso/9780199591756.003.0019

Ketkar, N. (2017). Deep Learning with Python. In *Deep Learning with Python*. https://doi.org/10.1007/978-1-4842-2766-4

Shalev-Shwartz, S., & Ben-David, S. (2014). Understanding Machine Learning: From theory to alorithms. In Understanding Machine Learning: From Theory to Algorithms (2014th ed., Vol. 9781107057). https://doi.org/10.1017/CBO9781107298019

Articles without DOI

Bechdolt, B. V. (1973). Cross-sectional travel demand functions-us visitors to Hawaii, 1961-70. *The Quarterly Review of Economics and Business*, *13*(4).

Hsu, C.-W., & Liu, H.-Y. (2008). Corporate Diversification and Firm Performance: The Moderating Role of Contractual Manufacturing Model. *Asia Pacific Management Review*, *13(1)*, 345–360.

Tinbergen, J. (1962). An analysis of world trade flows. In *Shaping the world economy*.

New York: Twentieth Century Fund.

Visa, S., Ramsay, B., Ralescu, A. L., & Van Der Knaap, E. (2011). Confusion Matrix-based Feature Selection. MAICS, 710, 120‑127.

Articles with DOI

Aguiló, E., & Rosselló, J. (2005). Host community perceptions. A cluster analysis. *Annals of Tourism Research*, *32*(4), 925–941. https://doi.org/10.1016/j.annals.2004.11.004

Aguiló Perez, E., & Juaneda, S. C. (2000). Tourist expenditure for mass tourism markets. *Annals of Tourism Research*, *27*(3), 624–637. https://doi.org/10.1016/S0160-7383(99)00101-2

Bechdolt, B. V. (1973). Cross-sectional travel demand functions-us visitors to hawaii, 1961-70. *The Quarterly Review of Economics and Business*, *13*(4).

Bigné, E., Oltra, E., & Andreu, L. (2019). Harnessing stakeholder input on Twitter: A case study of short breaks in Spanish tourist cities. *Tourism Management*, *71*, 490–503. https://doi.org/10.1016/J.TOURMAN.2018.10.013

Brida, J. G., Lanzilotta, B., Moreno, L., & Santiñaque, F. (2018). A non-linear approximation to the distribution of total expenditure distribution of cruise tourists in Uruguay. *Tourism Management*, *69*, 62–68. https://doi.org/10.1016/J.TOURMAN.2018.05.006

Bujosa, A., Riera, A. A., & Pons, P. J. (2015). Sun-and-beach tourism and the importance of intra-destination movements in mature destinations. *Tourism Geographies*, *17*(5), 780–794. https://doi.org/10.1080/14616688.2015.1093538

BUTLER, R. (1980), THE CONCEPT OF A TOURIST AREA CYCLE OF EVOLUTION: IMPLICATIONS FOR MANAGEMENT OF RESOURCES. Canadian Geographer / Le Géographe canadien, 24: 5-12. https://doi.org/10.1111/j.1541-

0064.1980.tb00970.x

Chang, K. G., Chien, H., Cheng, H., & Chen, H. i. (2018). The impacts of tourism development in rural indigenous destinations: An investigation of the local residents' perception using choice modeling. *Sustainability*, *10*(12), 4766. https://doi.org/10.3390/su10124766

Chari, M. D. R., David, P., Duru, A., & Zhao, Y. (2019). Bowman's risk-return paradox: An agency theory perspective. *Journal of Business Research*, *95*, 357–375. https://doi.org/10.1016/j.jbusres.2018.08.010

Chattopadhyay, M., & Mitra, S. K. (2020). What Airbnb Host Listings Influence Peer-to-Peer Tourist Accommodation Price? *Journal of Hospitality and Tourism Research*, *44*(4), 597–623. https://doi.org/10.1177/1096348020910211

Cohen, S. A., Prayag, G., & Moital, M. (2014, November). Consumer behaviour in tourism: Concepts, influences and opportunities. *Current Issues in Tourism*, 17, 872–909. https://doi.org/10.1080/13683500.2013.850064

Dann, G. M. S. (1977). Anomie, ego-enhancement and tourism. *Annals of Tourism Research*, *4*(4), 184–194. https://doi.org/10.1016/0160-7383(77)90037-8

De Cantis, S., Ferrante, M., Kahani, A., & Shoval, N. (2016). Cruise passengers' behavior at the destination: Investigation using GPS technology. *Tourism Management*, *52*, 133–150. https://doi.org/10.1016/J.TOURMAN.2015.06.018

Deyà Tortella, B., & Tirado, D. (2011). Hotel water consumption at a seasonal mass tourist destination. The case of the island of Mallorca. *Journal of Environmental Management*, *92*(10), 2568–2579. https://doi.org/10.1016/j.jenvman.2011.05.024

Dieu, O., Schnitzler, C., Llena, C., & Potdevin, F. (2020). Complementing subjective with objective data in analysing expertise: A machine-learning approach applied to badminton, *Journal of Sports Sciences, 38*(17), 1943–1952.

https://doi.org/10.1080/02640414.2020.1764812,

Dolnicar, S. (2002). A review of data-driven market segmentation in tourism. *Journal of Travel and Tourism Marketing*, *12*(1), 1–22. https://doi.org/10.1300/J073v12n01_01

D'Urso, P., Disegna, M., Massari, R., & Osti, L. (2016). Fuzzy segmentation of postmodern tourists. *Tourism Management*, *55*, 297–308. https://doi.org/10.1016/j.tourman.2016.03.018

Eagles, P. F. J. (2016). The Travel Motivations of Canadian Ecotourists. *Journal of Travel Research*, *31*(2), 3–7. https://doi.org/10.1177/004728759203100201

Ernst, D., & Dolnicar, S. (2017). How to Avoid Random Market Segmentation Solutions. *Journal of Travel Research*, *57*(1), 69–82. https://doi.org/10.1177/0047287516684978

Erkuş-Öztürk, H. (2016). Diversification of hotels in a single-asset Tourism City. *Advances in Culture, Tourism and Hospitality Research*, *12*, 173–185. https://doi.org/10.1108/S1871-317320160000012013/FULL/XML

Ferrer-Rosell, B., Coenders, G., & Martínez-Garcia, E. (2016). Segmentation by tourist expenditure composition: An approach with compositional data analysis and latent classes. *Tourism Analysis*, *21*(6), 589-602. https://doi.org/10.3727/108354216X14713487283075

Fleischer, A., & Pizam, A. (1997). Rural tourism in Israel. *Tourism Management*, *18*(6), 367–372. https://doi.org/10.1016/S0261-5177(97)00034-4

Gu, Q., Zhang, H., Huang, S. (Sam), Zheng, F., & Chen, C. (2021). Tourists' spatiotemporal behaviors in an emerging wine region: A time-geography perspective. *Journal of Destination Marketing & Management*, *19*, 100513. https://doi.org/10.1016/J.JDMM.2020.100513

Guo, S., Jiang, Y., & Long, W. (2019). Urban tourism competitiveness evaluation system and its application: Comparison and analysis of regression and classification methods. *Procedia Computer Science*, *162*, 429–437. https://doi.org/10.1016/j.procs.2019.12.007

Grant, R. M., Jammine, A. P., & Thomas, H. (2017). Diversity, Diversification, and Profitability Among British Manufacturing Companies, 1972–1984. *Academy of Management Journal*, *31*(4), 771–801. https://doi.org/10.5465/256338

Han, H., Kim, S. (Sam), & Otoo, F. E. (2018). Spatial movement patterns among intra-destinations using social network analysis. *23*(8), 806–822. https://doi.org/10.1080/10941665.2018.1493519

Hua, N., Huang, A., Medeiros, M., & DeFranco, A. (2020). The moderating effect of operator type: the impact of information technology (IT) expenditures on hotels' operating performance. *International Journal of Contemporary Hospitality Management*, *32*(8), 2519–2541. https://doi.org/10.1108/IJCHM-09-2019-0753

Ibragimov, K., Perles-Ribes, J. F., & Ramón-Rodríguez, A. B. (2021). The economic determinants of tourism in Central Asia: A gravity model applied approach. *Tourism Economics*, https://doi.org/10.1177/13548166211009985

Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, *31*(8), 651–666. https://doi.org/10.1016/j.patrec.2009.09.011

Keshavarzian, P., & Wu, C. L. (2017). A qualitative research on travellers' destination choice behaviour. *International Journal of Tourism Research*, *19*(5), 546–556. https://doi.org/10.1002/jtr.2128

Klenosky, D. B. (2002). The "Pull" of Tourism Destinations: A Means-End Investigation. *Journal of Travel Research*, *40*(4), 396–403. https://doi.org/10.1177/004728750204000405

Konu, H., Laukkanen, T., & Komppula, R. (2011). Using ski destination choice criteria to segment Finnish ski resort customers. *Tourism Management*, *32*(5), 1096–1105. https://doi.org/10.1016/J.TOURMAN.2010.09.010

Leask, A., Fyall, A., & Barron, P. (2014). Generation Y: An Agenda for Future Visitor Attraction Research. *International Journal of Tourism Research*, *16*(5), 462–471. https://doi.org/10.1002/JTR.1940

Lee, M., Hong, J. H., Chung, S., & Back, K.-J. (2020). Exploring the Roles of DMO's Social Media Efforts and Information Richness on Customer Engagement: Empirical Analysis on Facebook Event Pages. *Journal of Travel Research*, *60*(3), 670–686. https://doi.org/10.1177/0047287520934874

Lei, C. K. (2019). The influences of revenue diversification and incoming tourists on the performance of star-rated hotels in China. *Tourism Analysis*, *24*(4), 483–495. https://doi.org/10.3727/108354219X15652651367488

Maingi, S. W., Ondigi, A. N., & Wadawi, J. K. (2016). Market Profiling and Positioning of Park Brands in Kenya (Case of Premium and Under-Utilized Parks). *International Journal of Tourism Research*, *18*(1), 91–104. https://doi.org/10.1002/JTR.2036

Martorell, O, & Mulet-Forteza, C. (2010). The franchise contract in hotel chains: A study of hotel chain growth and market concentrations. *Tourism Economics*, *16*(3), 493–515. https://doi.org/10.5367/000000010792278446

Mok, C., & Iverson, T. J. (2000). Expenditure-based segmentation: Taiwanese tourists to Guam. *Tourism Management*, *21*(3), 299–305. https://doi.org/10.1016/S0261-5177(99)00060-6

Moore, R. S. (1995). Gender and alcohol use in a Greek tourist town. *Annals of Tourism Research*, *22*(2), 300–313. https://doi.org/10.1016/0160-7383(94)00078-6

Moral-Cuadra, S., Solano-Sánchez, M. Á., López-Guzmán, T., & Menor-Campos, A.

(2021). Peer-to-Peer Tourism: Tourists' Profile Estimation through Artificial Neural Networks. *Journal of Theoretical and Applied Electronic Commerce Research 2021, 16*(4), 1120–1135. https://doi.org/10.3390/JTAER16040063

Morley, C. L. (1992). A microeconomic theory of international tourism demand. *Annals of Tourism Research*, *19*(2), 250–267. https://doi.org/10.1016/0160-7383(92)90080-9

Morley, C., Rosselló, J., & Santana-Gallego, M. (2014). Gravity models for tourism demand: Theory and use. *Annals of Tourism Research*, *48*, 1–10. https://doi.org/10.1016/j.annals.2014.05.008

Moro, S., Batista, F., Rita, P., Oliveira, C., & Ribeiro, R. (2019). Are the States United? An Analysis of U.S. Hotels' Offers Through TripAdvisor's Eyes: *43*(7), 1112–1129. https://doi.org/10.1177/1096348019854793

Mun, S. G., Woo, L., & Paek, S. (2019). How important is F&B operation in the hotel industry? Empirical evidence in the U.S. market. *Tourism Management*, *75*, 156–168. https://doi.org/10.1016/j.tourman.2019.03.010

Park, K., & Jang, S. C. S. (2012). Effect of diversification on firm performance: Application of the entropy measure. *International Journal of Hospitality Management*, *31*(1). https://doi.org/10.1016/j.ijhm.2011.03.011

Oltean, F. D., & Gabor, M. R. (2016). Service diversification – A qualitative and quantitative analysis in Mures county hotels. *Engineering Economics*, *27*(5), 618–628. https://doi.org/10.5755/j01.ee.27.5.14153

Peng, L., Wang, L., Ai, X.-Y., & Zeng, Y.-R. (2020). Forecasting Tourist Arrivals via Random Forest and Long Short-term Memory. *Cognitive Computation*, *13*(1), 125–138. https://doi.org/10.1007/S12559-020-09747-Z

Pivčević, S., Petrić, L., & Mandić, A. (2020). Sustainability of Tourism Development in the Mediterranean—Interregional Similarities and Differences. *Sustainability*, *12*(18), 7641. https://doi.org/10.3390/SU12187641

Pal, M. (2007). Random forest classifier for remote sensing classification. *International Journal of Remote* Sensing, 26(1), 217–222. https://doi.org/10.1080/01431160412331269698

Qu, Y., Qu, H., & Chen, G. (2016). Market segmentation for a leverage revitalization of China's inbound tourism: the case of US leisure tourists. *Current Issues in* Tourism, 21(6), 646–662. https://doi.org/10.1080/13683500.2016.1264054

Richards, G. (1999). Vacations and the Quality of Life: Patterns and Structures. *Journal of Business Research*, *44*(3), 189–198. https://doi.org/10.1016/S0148-2963(97)00200-2

Rita, P., Brochado, A., & Dimova, L. (2018). Millennials' travel motivations and desired activities within destinations: A comparative study of the US and the UK. *Current Issues in Tourism*, *22*(16), 1–17. https://doi.org/10.1080/13683500.2018.1439902

Schuckert, M., Liu, X., & Law, R. (2015). Hospitality and Tourism Online Reviews: Recent Trends and Future Directions. *Journal of Travel and Tourism Marketing*, *32*(5), 608–621. https://doi.org/10.1080/10548408.2014.933154

Sharma, P., & Nayak, J. K. (2020). Understanding the determinants and outcomes of internal reference prices in pay-what–you-want (PWYW) pricing in tourism: An analytical approach. *Journal of Hospitality and Tourism Management*, *43*, 1–10. https://doi.org/10.1016/j.jhtm.2020.02.001

Sung, Y.-K., Chang, K.-C., & Sung, Y.-F. (2015). Market Segmentation of International Tourists Based on Motivation to Travel: A Case Study of Taiwan. *Asia Pacific Journal of Tourism* Research, *21*(8), 862–882. https://doi.org/10.1080/10941665.2015.1080175

Valls, A., Gibert, K., Orellana, A., & Antón-Clavé, S. (2018). Using ontology-based clustering to understand the push and pull factors for British tourists visiting a Mediterranean coastal destination. *Information and Management*, *55*(2), 145–159. https://doi.org/10.1016/j.im.2017.05.002

Wilkins, H., Merrilees, B., & Herington, C. (2007). Towards an understanding of total service quality in hotels. *International Journal of Hospitality Management*, *26*(4), 840–853. https://doi.org/10.1016/J.IJHM.2006.07.006

Xia, Z., Stewart, K., & Fan, J. (2021). Incorporating space and time into random forest models for analyzing geospatial patterns of drug-related crime incidents in a major U.S. metropolitan area. *Computers, Environment and Urban Systems*, *87*, 101599. https://doi.org/10.1016/J.COMPENVURBSYS.2021.101599

Xiang, Z., Du, Q., Ma, Y., & Fan, W. (2017). A comparative analysis of major online review platforms: Implications for social media analytics in hospitality and tourism. *Tourism Management*, *58*, 51–65. https://doi.org/10.1016/J.TOURMAN.2016.10.001

Xie, W., Li, H., & Yin, Y. (2021). Research on the Spatial Structure of the European Union's Tourism Economy and Its Effects. *International Journal of Environmental Research and Public Health*, *18*(4), 1389. https://doi.org/10.3390/IJERPH18041389

Xu, D., & Tian, Y. (2015). A Comprehensive Survey of Clustering Algorithms. *Annals of Data Science*, *2*(2), 165–193. https://doi.org/10.1007/s40745-015-0040-1

Xu, X., & Li, Y. (2016). The antecedents of customer satisfaction and dissatisfaction toward various types of hotels: A text mining approach. *International Journal of Hospitality Management*, *55*, 57–69. https://doi.org/10.1016/j.ijhm.2016.03.003

Yeh, C. Y., Chen, C. M., & Hu, J. L. (2012). Business diversification in the hotel industry: A comparative advantage analysis. *Tourism Economics*, *18*(5), 941–952.

https://doi.org/10.5367/te.2012.0152

Zoltan, J., & McKercher, B. (2015). Analysing intra-destination movements and activity participation of tourists through destination card consumption. *Tourism Geographies*, *17*(1), 19–35. https://doi.org/10.1080/14616688.2014.927523

Conferences

Lin, Y. L., Chen, T. Y., & Yu, L. C. (2017). Using Machine Learning to Assist Crime Prevention. *Proceedings - 2017 6th IIAI International Congress on Advanced Applied Informatics, IIAI-AAI 2017*, 1029–1030. https://doi.org/10.1109/IIAI-AAI.2017.46

Xuan, S., Liu, G., Li, Z., Zheng, L., Wang, S., & Jiang, C. (2018). Random forest for credit card fraud detection. *ICNSC 2018 - 15th IEEE International Conference on Networking, Sensing and Control*, 1-6. https://doi.org/10.1109/ICNSC.2018.8361343

Newspaper article

Benabent Fernández de Córdoba, M., & Mata Olmo, R. (2007, July 13). El futuro de la geografía. El País.https://elpais.com/diario/2007/07/13/opinion/1184277607_850215.html

Web pages

IBESTAT. (2020). INSTITUT BALEAR D'ESTADISTICA. Retrieved February 8, 2020, from Gasto de los turistas con destino principal las Illes Balears por periodo y tipo de alojamiento website: https://ibestat.caib.es/ibestat/estadistiques/f58f0937-c64f-469d-bad5-99f29bbb59ce/755a8af8-2b59-41ee-9c2d-9aa0f4f8509f/es/I208004_n002.px

Chesters, C. (2017). The Business: Redefining Accor's F&B strategy. Retrieved November 20, 2020, from https://www.hotelnewsme.com/catering-newsme/business-redefining-accors-fb-strategy/