

Article

An Application of Ordered Weighted Averaging Operators to Customer Classification in Hotels

Pere Josep Pons-Vives ¹, Mateu Morro-Ribot ², Carles Mulet-Forteza ^{1,*} and Oscar Valero ^{3,4}

¹ Departament d'Economia de l'Empresa, Universitat de les Illes Balears, 07122 Palma de Mallorca, Illes Balears, Spain; p.pons2@estudiant.uib.eu

² Business Intelligence & Data Analytics Department, Hotelbeds, 07007 Palma de Mallorca, Illes Balears, Spain; mmorro@hotelbeds.com

³ Departament de Ciències Matemàtiques i Informàtica, Universitat de les Illes Balears, 07122 Palma de Mallorca, Illes Balears, Spain; o.valero@uib.es

⁴ Institut d' Investigació Sanitària Illes Balears (IdISBa), Hospital Universitari Son Espases, 07120 Palma de Mallorca, Illes Balears, Spain

* Correspondence: carles.mulet@uib.es

Abstract: An algorithm widely used in hotel companies for demand analysis is the so-called K-means. The aforementioned algorithm is based on the use of the Euclidean distance as a dissimilarity measure and this fact can cause a main handicap. Concretely, the Euclidean distance provides a global difference measure between the values of the descriptive variables that can blur the relative differences in each component separately and, hence, the cluster algorithm can assign a custom to an incorrect cluster. In order to avoid this drawback, this paper proposes an application of the use of Ordered Weighted Averaging (OWA) operators and an OWA-based K-means for clustering customers staying at a real five-star hotel, located in a mature sun-and-beach area, according to their propensity to spend. It must be pointed out that OWA-based distance calculates relative distances and it is sensitive to the differences in each component separately. All experiments show that the use of the OWA operator improves the performance of the classical K-means up to 21.6% and reduces the number of convergence iterations up to 48.46%. Such an improvement has been tested through a ground truth, designed by the marketing department of the firm, which states the cluster to which each tourist belongs. Moreover, the customer classification is achieved regardless of the season in which the customer stays at the hotel. All these facts confirm that the OWA-based K-means could be used as an appropriate tool for classifying tourists in purely exploratory and predictive stages. Furthermore, the new methodology can be implemented without requiring radical changes in the implementation of the classical methodology and in data processing which is crucial so that it can be incorporated into the control panel of a real hotel without additional implementation costs.

Keywords: Euclidean distance; OWA; clustering; K-means; hotel performance; customer classification

MSC: 03B52; 54E35; 62H30; 91B86



Citation: Pons-Vives, P.J.; Morro-Ribot, M.; Mulet-Forteza, C.; Valero, O. An Application of Ordered Weighted Averaging Operators to Customer Classification in Hotels. *Mathematics* **2022**, *10*, 1987. <https://doi.org/10.3390/math10121987>

Academic Editor: Gia Sirbiladze

Received: 3 May 2022

Accepted: 7 June 2022

Published: 9 June 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Sun-and-beach hotel performance has been widely studied in tourism literature [1]. However, among the two possible growth profit strategies, researchers have focused on expansion, leaving aside diversification [2]. Expansion strategies imply income growth through the addition of hotel establishments or rooms. Different expansion strategies, which involve the use of property, leasing, and franchise and management contracts, represent different levels of effort in terms of management and investment [2]. Diversification takes advantage of underused resources and economies of scope to obtain resources and create synergies between departments [3]. However, diversification strategies have received less interest. A chief reason may be that the industry growth model is based on

Fordism (see Section 2) until the destination reaches the maturity stage [4]. Nevertheless, as destinations reach a certain degree of maturity, these growth strategies are no longer viable [5]. Additionally, the sun-and-sea Fordism model assumes that tourists traveling to sun-and-beach destinations only look for sunny weather and idyllic beaches ([6,7]).

Motivation is the starting point of consumption, the basis of the consumers' behavioral analysis. That is, the research field on how and why different groups of consumers behave as they do [8]. The other capstone of behavioral analysis is the individuals' characteristics. In this way, the different consumption behaviors of different segments are referred to in tourism literature as tourism consumption patterns [9]. Tourist consumption patterns are analyzed from different scales and perspectives: macro-, micro-, and nano-scale [7]. Although there is already a vague border between scales it can be stated that the 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 [10]. Finally, the nano-scale comprises 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 behavior of tourists lodged at their hotels to allocate resources to satisfy the needs of the most profitable segment. Under the assumption of the stability of tourism consumption patterns, its study has been set aside. Recently, tourists' preferences and demands have become more complex [11]. In addition, Mediterranean sun-and-beach destinations are in an advanced maturity state which, among other factors, implies a high degree of competence between hotels. These two factors may affect the hotels' performance ([5,12]). Tourists do not only want to stay all day at the beach and then go back to the hotel room, but they also seek other activities beyond sunbathing. They seek hotels offering rooftops, spas, restaurants, bars, or sky bars. Therefore, via nano-scale analysis, hotel managers may develop strategies focused on differentiating their product from the large number of competitors in mature destinations. In fact, there is a high degree of competence derived from the number of establishments that compete for the same segment with the same productive model [4,13].

In this context, at mature destinations the number of luxury hotels has increased ([14,15]). These kinds of establishments can satisfy the needs and demands. However, this wide range of services ranging from a spa to a gastronomic restaurant causes different costs and contributions to the profitability of the hotel establishment, which is why proper customer segmentation becomes essential to increase profitability. In [15], it is stated that luxury hotels have unique operational characteristics such as person-to-person interactions, diverse amenities, and high staff–customer ratios. Therefore, to maintain service quality while maintaining performance, luxury hotels may focus on diversification, operational efficiency, and diversification strategies [16]. To set the proper diversification, it is essential for hotels to implement an accurate segmentation technique that identifies those services that can be more profitable. In this way, following [17], in superior hotel establishments, Food and Beverage (F&B) services, beyond contributing to their Gross Operating Profit (GOP), favor room sales, which translates into improvements in occupancy ratios, Average Daily Rate (ADR), Revenue Per Available Room (RevPAR), and Gross Operating Profit Per Available Room (GOPPAR). Therefore, understanding the new consumption patterns of tourists entails the incorporation of new services aimed at improving the service offered to the hotel establishment's customers, with the aim of improving its profitability ([18–20]). With this aim, a classical segmentation technique widely used in hotel companies for consumption patterns of tourists is the segmentation algorithm known as K-means. However, classical segmentation techniques, among them K-means, seem to be limited in their ability to segment luxury hotel consumption patterns. Among other reasons, the aforementioned algorithm is based on the use of the Euclidean distance as a dissimilarity measure and this fact can cause a main handicap. Concretely, in many situations the Euclidean distance is a measure that is insensitive to the coordinate-to-coordinate differences of the variables involved in the measure because it is able to produce a compensation between the differences in different coordinates even when data are normalized. This may result in individuals

being identified with the wrong centroid. Thus, the Euclidean distance provides a global difference measure between the values of the descriptive variables when dissimilarities between centroids and objects are measured. This can blur the relative differences in each component separately and, hence, the cluster technique can assign an object (a tourist) to an incorrect cluster. In order to avoid this drawback, in [21] the Euclidean distance has been replaced by a new distance constructed by means of the use of an Ordered Weighted Averaging (OWA) operator in the sense of [22]. It must be stressed that such a distance does not require normalization of the data because it calculates relative distances and, in addition, it is sensitive to the coordinate-to-coordinate differences (see Section 3).

The aim of this paper is to apply the OWA-based K-means in order to cluster customers staying at a real five-star hotel, located in a mature sun-and-beach area, according to their propensity to spend. The experimental results are obtained from real data provided by a luxury hotel located in the city of Palma in Mallorca. The obtained results show that the use of the OWA operator provides better segments than classical K-means, improving its performance up to 21.6%, and reduces the number of convergence iterations up to 48.46%. Such an improvement has been tested through a ground truth, designed by the marketing department of the firm, which states the cluster to which each tourist belongs. Moreover, the customer classification is achieved regardless of the season in which the customer stays at the hotel. All these facts confirm that the OWA-based K-means could be used as an appropriate tool for classifying tourists in purely exploratory and predictive stages. Furthermore, the novelty of the OWA-based methodology is given by the fact that it can be implemented without requiring radical changes in the implementation of the classical methodology (an easy modification of the classical K-means) and in data processing which is crucial so that it can be incorporated into the control panel of a real hotel without additional implementation costs, which can allow improvement in the performance of the hotel establishment significantly, both in the short and medium term, and its profitability. In addition, this is seen without having to create models for low, high, and mid-season.

The remainder of the paper is organized as follows. In Section 2 the notion of the nano-scale is introduced and the need for such a scale in the analysis of consumption patterns at the hotel level is justified. Section 3 is devoted to recalling the basics about aggregation functions and the OWA operators. Moreover, the construction of those distances based on OWAs that will play a central role for our target is described. Furthermore, the OWA-based K-means algorithm is also shown. In Section 4, the data description is provided, i.e., the variables considered in order to describe the tourists to be classified. In Section 5, the obtained experimental results are described in detail. In addition, an illustrative numerical example that allows us to show the functionality of the OWA-based clustering technique and its advantage with respect to that based on the use of the Euclidean distance is given. Finally, conclusions and further work are given in Section 6.

2. The Need for Consumption Patterns at the Nano-Scale: The Hotel-Level Case

Hotel establishments, like destinations, go through different product phases [4]. Up to the maturity stage, they base their revenues and profitability on the Fordist model. Under this approach, performance efficiency is measured by comparing observed and optimal costs and revenues subject to price and quality constraints [23]. Since the 1960s, this has allowed the creation of multinational hotel chains that are extremely efficient in cost and quality management [2]. However, in mature destinations, where cost reduction is difficult and tourist preferences have evolved towards multipurpose travel, expansion and cost control strategies can be combined with service diversification. In this scenario, hotels may choose to develop new businesses that are to a greater or lesser extent related to their existing lines of business. However, few studies consider product diversification strategies in a hotel [24].

In [25], diversification of F&B services was identified as a key variable of hotel performance. In [26], it was highlighted that the effect of diversification on performance is based on the combined effect of synergies and the possibility of sharing resources and knowledge

between different business units that can lead to higher performance. However, the costs may outweigh the benefits generated by synergies at some level of diversification (see [27]). In [18], the effects of Taiwanese hotels' diversification in F&B strategies on their growth and earnings stability were examined. In particular, it was found that hotels with total revenue generated mainly from F&B service tend to have higher profit margin growth, but also suffer from higher instability. Along these lines, in [28] a trend towards service diversification was also found when examining data from the hotel sector in Turkey, an example of a mature sun-and-sea destination. Concretely, the firm size and sector-specific knowledge (intra-industry investments and experience of hotel workers) are shown to be important variables in determining the success of diversification strategies.

Taking a holistic view, in [29] it was evidenced using stochastic frontier analysis that revenue diversification in the rooms and F&B department and the efficiency of other services are explained by the overall structure, technological efficiency, workers' capabilities, and hotel characteristics. One step further, in [30] non-linearity in the profitability analysis of diversification was introduced. In particular, it was found that unrelated diversification increases profitability up to a certain level. However, beyond that level, unrelated diversification decreases profitability, implying that at high levels of unrelated internal diversification there is a loss of control and effort due to distance from the core business. They also found that at low levels of related diversification, the synthetic related business risk is larger than the risk reduction effect. This means that at low levels of related diversification, the synthetic related business risk is higher than the risk reduction effect.

A better understanding of customer preferences and behavior can be key for a hotel when implementing diversification strategies. Therefore, fuzzy segmentation techniques can provide a better understanding of their consumption patterns that prevent internal transaction costs from being greater than the synergies created between departments [30]. This is especially important in the initial aspects of diversification as the learning curve appears to exhibit diminishing marginal returns [19].

However, there is a gap in the literature regarding the analysis of tourists' consumption patterns at the hotel where they stay. The existing literature has focused on tourism product choice, routes, itinerant cognition, spatio-temporal distribution, and mental maps. This may be due to the fact that in order to obtain meaningful results it is necessary to obtain large and precise datasets on tourist behavior. In this sense, in [31] tourists were segmented in terms of their behavior. The authors combined traditional interviews with socio-demographic questions (age, gender, size of travel group, etc.) with information from Global Positioning Systems (GPSs) (length of trip, duration of trip, number of attractions visited, average speed, etc.). By 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.

New technologies allow hotel managers to interact with their customers almost immediately. This is why the aggregated use of information on tourist preferences and characteristics at the hotel level allows for a better understanding of how the customer interacts with the hotel establishment [32], allowing the company to develop and communicate targeted and differentiated diversification strategies for each customer typology ([28,32]). These more precise segmentation techniques make it possible to adapt the classic segmentation models, resulting in more precise models, without having to make radical changes in data processing (such as normalization processes). All this will result in an improvement in the performance of hotel establishments, while allowing them to better understand their customers and diversify their sources of revenue. In fact, in [33] it was already indicated, in terms of the profitability that the restaurant service of a hotel establishment can provide, that the average satisfaction with the service provided to these customers positively influenced the performance of the hotel establishment.

Based on all of the above and given the few works that analyze consumption patterns within a hotel, this paper focuses on advancing the analysis on this scale (nano-scale). As already indicated, the nano-scale can be defined as the interaction between the tourist

and a local business offering different services. In this sense, new technologies currently allow hotel managers to interact with their customers almost immediately. This is why the aggregate use of information on tourist preferences and characteristics at this level allows for a better understanding of the way in which a customer interacts with the hotel establishment [32], enabling the company to develop and communicate differentiated strategies focused on each type of customer ([28,34]). Regarding the use of aggregation methodologies applied to segmentation and classification and taking the fact that companies prefer to advance by adapting classic segmentation models without having to make radical changes in data processing and, thus, without incurring many implementation costs, we introduce OWA-based K-means, which allows adapting the classical K-means while avoiding some of its shortcomings in classification mentioned in Section 1 in [21]. This adaptation can be carried out without having to make radical changes in the implementation of classical methodologies on the one hand and, on the other hand, in data processing (in particular, there is no need to normalize incoming data).

In the light of the information above, the objective is to apply the aforementioned methodology to the nano-scale analysis of a hotel establishment of a superior category in a mature sun-and-beach destination, for which it is desired to segment customers according to their consumption potential regardless of the time of year in which they visit the establishment. On the one hand, we aim to achieve an improvement in the performance of hotel establishments and, on the other hand, allow them to get to know their customers better and, thus, diversify their sources of income.

3. The OWA-Based K-Means

In the market segmentation literature, partitional clustering algorithms are used to find the patterns of customers in such a way that those assigned to the same group (cluster) are more similar to each other than the patterns of those customers contained in the other clusters ([35,36]). Among these algorithms, K-means is one of the most popular in the social sciences (see, for instance, [35,37,38]). However, K-means has a significant probability of not converging to a solution. Moreover, it is very sensitive to outliers and statistical noise ([39]). Furthermore, another disadvantage of the aforementioned algorithm was shown in [21]. Specifically, the Euclidean distance does not allow, in general, one to obtain a dissimilarity measure that takes into account the information provided by the explanatory variables coordinate-to-coordinate. In fact, this distance dilutes the aforementioned information by providing, in a way, a measure that is insensitive to the coordinate-to-coordinate differences of the variables involved in the measure because it is able to produce a compensation between the differences in different coordinates. This may result in objects (individuals/customers) being identified with the wrong centroid (for a deeper discussion, we refer the reader to [21]).

According to [40], a function $A : [0, 1]^n \rightarrow [0, 1]$ is an aggregation function provided that it is monotone ($A(x) \leq A(y)$ if $x, y \in [0, 1]^n$ and $x_i \leq y_i$ for all $i = 1, \dots, n$) and it satisfies the so-called boundary conditions $A(0, \dots, 0) = 0$ and $A(1, \dots, 1) = 1$. Of course, Euclidean distance can be understood as a measure of dissimilarity obtained by means of the aggregation of the information coming from each coordinate whose numerical value is normalized. In fact, such a measure aggregates a collection of squared distances computed coordinate-to-coordinate. Following [40], aggregation functions play a crucial role in decision-making processes. Thus, if $A : [0, 1]^n \rightarrow [0, 1]$ is an aggregation function and A_i is the i th-coordinate function of A with $i \in \{1, \dots, n\}$, then each A_i can be interpreted as the different criteria to be taken into account for decision making. Indeed, if $X \subseteq [0, 1]^n$ represents the (non-empty) set of alternatives, then for every $x \in X$ the value of $A_i(x) \in [0, 1]$ can be interpreted as the degree to which x satisfies the criterion represented by A_i . Thus, the aggregation function A can be understood as a tool to produce an overall degree to which alternative x satisfies at the same time the n criteria under consideration.

A special case of aggregation function is the so-called Ordered Weighted Averaging (OWA) operator which was introduced in [22] (see also [40]). As mentioned before, in [21]

a new distance was introduced as a possible replacement for the Euclidean distance in the K-means algorithm which is obtained by means of aggregation of distances computed coordinate-to-coordinate and merged via an OWA operator and it is called Ordered Weight Distance Relative (OWDr for short). It must be stressed that such a distance generalizes the ordered weighted distance introduced in [41]. In order to introduce its constructions, let us recall that, given a weighting vector $W \in [0, 1]^n$ such that $\sum_{i=1}^n w_i = 1$, an OWA aggregation operator of dimension n is an aggregation function $OWA : [0, 1]^n \rightarrow [0, 1]$ such that

$$OWA(x_1, \dots, x_n) = \sum_{i=1}^n w_i x_{(i)},$$

where $x_{(i)}$ denotes the i th largest element in the collection $\{x_1, \dots, x_n\}$. Based on the notion of the OWA operator, given a weighting vector $W \in [0, 1]^n$ such that $\sum_{i=1}^n w_i = 1$, the OWDr $OWDr : \mathbb{R}_+^n \times \mathbb{R}_+^n \rightarrow [0, 1]$ is given as follows ([21]):

$$OWDr(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^n w_i d_r(x_{(i)}, y_{(i)}), \tag{1}$$

where $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ (features vectors or centroids) with $\mathbf{x} = (x_1, \dots, x_n)$, $\mathbf{y} = (y_1, \dots, y_n)$, $d_r(x_{(i)}, y_{(i)})$ denotes the i th largest element in the collection of distances

$$\{d_r(x_1, y_1), \dots, d_r(x_n, y_n)\},$$

and $d_r : \mathbb{R} \times \mathbb{R} \rightarrow [0, 1]$ is the distance given by

$$d_r(x, y) = \begin{cases} 0 & x = y = 0 \\ \frac{|x-y|}{\max\{x,y\}} & otherwise \end{cases} . \tag{2}$$

Observe that the input vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ represent the data instances involved in the clustering process (centroids/features vectors). Moreover, given input vectors \mathbf{x}, \mathbf{y} and assigning specific weighting vector W , the associated OWDr generates an overall degree of dissimilarity $OWDr(\mathbf{x}, \mathbf{y})$ in such a way that the information from the different scales is all incorporated into that measure via the values $d_r(x_1, y_1), \dots, d_r(x_n, y_n)$, that measure differences only coordinate-to-coordinate, in such a way that the weighting vector is able to intensify (or diminish as appropriate) the most notable differences.

The OWDr was shown to be sensitive to differences in the scales of the variables involved in the measurement coordinate-to-coordinate, avoiding the aforementioned possible drawbacks between these differences. The most salient feature of the OWDr is that it can diminish (or intensify) the influence of excessively large or excessively small deviations in the data to be aggregated by assigning them low (or high) weights.

Consider $X = [z_1, \dots, z_m]$, a dataset with n dimensions ($z_i \in \mathbb{R}^n$) for all $i = 1, \dots, m$, to be divided into k clusters. The objective of K-means is to obtain a partition of the data in which the mean square error between the cluster centroid and the cluster points is minimum. The process of the OWA-based K-means is as follows:

1. An initial partition with k clusters is selected and k initial clusters c_1^0, \dots, c_k^0 are set.
2. For each step $t \geq 0$, each $x_i \in X$ is in the cluster C_l^t such that its distance, measured via the OWDr, from the centroid c_l^t of C_l^t is minimum for all $l \in \{1, \dots, k\}$.
3. The centroid of each cluster is recalculated for the next step $t + 1$ by calculating the arithmetic mean of each cluster until step t .
4. If the algorithm does not converge, then repeat step 2.

It must be stressed that the algorithm is considered convergent when in one step the centroids of the clusters, after recalculation, remain unchanged. Thus, such a fact is considered as a stopping criterion.

In the light of the exposed facts, it is worth mentioning that the introduction of the OWD_r allows fixing greater weights to be assigned to those observations that are further apart, causing those distances with a larger value to contribute more to the overall measure and, hence, providing enough quantitative information in order to be able to discriminate if the datum differs enough from the centroid to be discarded from the cluster.

4. Data Description

As mentioned before, the objective is to segment the customers staying at a hotel according to their consumption potential regardless of the time of the year when they visit the establishment. Following the clustering, the bookings of the marketing department of the firm were divided into three segments depending on their propensity to spend: low, 780 (34.6%), medium, 946 (42%), and high, 529 (23.4%). Under this approach, the company takes into account the vagueness of the subjective opinion of the reception managers and the F&B manager. The preceding segmentation is used as ground truth in order to confirm or reject the tourists' classification that both algorithms provide and, thus, to check their performance by means of the accuracy metric detailed in Section 5.

As can be seen in Table 1, the variables considered in order to describe the tourists are days of stay, the price paid by the customer for the accommodation (cost of stay), number of customers per visit to the rooftop bar (number of diners), total amount spent at the bar for each booking (expenditure per reserve), and total number of visits made to the rooftop bar during the stay (number of visits).

Table 1. Description of variables that are incorporated into the model.

Days of stay	Number of days that tourist has been hosted in hotel
Cost of stay	Price paid by the customer for the accommodation
Number of diners	Number of customers per visit to the rooftop bar
Expenditure per reserve	Total amount spent at the bar for each booking
Number of visits	Total number of visits made to the rooftop bar during the stay

Source: own work.

Tourism literature has recognized that length (days) of stay arguably is a key determinant of the success of a destination as well as its firm's success [42]. The price paid by the guest for the room (cost of stay) is a primary filter as well as the main source of income of hotels. Higher room prices exclude low purchasing power segments [43,44]. The expenditure per reserve and the number of diners and the number of visits determine the success of the diversification strategy because no diversification can be conducted if there is not enough revenue to, at least, cover its costs [17,45]. The understanding of the interaction of the customer with the different outlets is the core of the analysis of consumption patterns at the nano-scale.

In this way, we have a database with information on 2256 bookings that spent at least one night in the hotel between 1 March 2019 and 31 October 2019. This information has been obtained directly from the database of the hotel chain.

As can be seen in Table 2, the average stay per booking is 3.56 nights, which usually coincides with the weekend. Thus, there are tourists who spend one night or tourists who spend their entire holiday in the hotel, reaching a maximum of 33 nights. Similar to the average cost provided by the Spanish Instituto Oficial de Estadística (INE), the average cost of the stay for this period is just over a thousand euros, with some cases where the cost is zero due to commercial or operational reasons (see [46]). The number of diners per visit at the rooftop bar is about five people and they spend an average of EUR 93.58. Moreover, they visit the bar about 3 times during their stay.

Table 2. Description of data.

	Minimum	Maximum	Mean	Standard Deviation
Days of stay	1	33	3.56	2.43
Cost of stay	EUR 0	EUR 11,050	EUR 1008.97	EUR 759.26
Number of diners	1	47	5.14	4.96
Expenditure per reserve	EUR 2	EUR 830	EUR 93.58	EUR 91.73
Number of visits	1	27	3.42	3.08

Source: own work.

5. Experimental Results

The segmentation has been carried out independently of classical categories such as seasonality and nationality which add complexity to the process and do not always provide useful information.

In order to make the results meaningful, 1500 experiments have been executed for both the Euclidean distance and each choice of weights of the OWDr measure. All experiments have been executed in Python. As a measure to compare the goodness-of-fit between the two distances applied to the K-means, we used the accuracy, a standard measure in the literature ([47]). Observe that, as pointed out before, we rely on a ground truth and, thus, each tourist has been previously classified by the marketing department of the firm and, hence, after the execution of both algorithms we can analyze whether the K-means can be useful for the segmentation that we want to do by comparing the marketing department classification against that given by the algorithms. Hence, following [47], in order to evaluate the performance of the algorithms we have used a confusion matrix. Such a matrix includes, on the one hand, the True Positive (TP) and the True Negative (TN) that correspond to well-made classifications and, on the other hand, the False Negative (FN) and the False Positive (FP) that correspond to those classifications that are incorrect. Therefore, the accuracy has been calculated using $\frac{(TP+TN)}{(TP+TN+FP+FN)}$. In addition, in order to optimize applicability, the average number of iterations that each measure needed to converge (centroids of the clusters remain unchanged after recalculation) to a solution was calculated. The weighting vectors in all experiments have been selected heuristically.

As can be seen in Table 3 the K-means with the OWDr measure outperforms the Euclidean distance in all cases. The accuracy using the Euclidean distance is the same in all cases (all rows), this is because the weights defined have only affected the OWDr considered, therefore the K-means with the Euclidean distance remains constant both in the average number of iterations and in effectiveness. Additionally, except for the vector weight $W = (0.6, 0.2, 0.2, 0, 0)$ (last row), in all combinations (vector weights chosen), it can be seen that apart from having better effectiveness, the OWDr also has a faster convergence.

Table 3 shows the obtained results, on average, of the accuracy achieved and the number of iterations need to converge after 1500 experiments for each weight choice. As can be seen in Table 3, the weight vector that work best is: $(0.35, 0.25, 0.2, 0.1, 0.1)$. Concretely, it provides both the best effectiveness and at the same time the best convergence. When the weight vector chosen is $(0.6, 0.2, 0.2, 0, 0)$ (last row of Table 3), we can see that the performance of the OWA-based K-means algorithm is far from those that heuristically have obtained the best goodness-of-fit and in this case it provides a better performance than the classical K-means.

Table 3. Accuracy and iterations (both as averages) needed to convergence for K-means (Euclidean) and OWA-based K-means (OWDr) after 1500 experiments.

Experimental Results				
Weights	Iterations with OWDr	Iterations with Euclidean	Accuracy with OWDr	Accuracy with Euclidean
(0.2, 0.2, 0.2, 0.2, 0.2)	23.012	29.352	0.712	0.596
(0.3, 0.2, 0.2, 0.2, 0.1)	29.352	29.923	0.596	0.596
(0.35, 0.25, 0.2, 0.1, 0.1)	24.566	29.762	0.714	0.596
(0.4, 0.3, 0.2, 0.1, 0)	29.923	29.570	0.597	0.596
(0.4, 0.2, 0.2, 0.1, 0.05)	20.047	30.231	0.726	0.596
(0.5, 0.3, 0.1, 0.1, 0)	29.762	29.616	0.596	0.596
(0.6, 0.2, 0.2, 0, 0)	78.304	30.079	0.687	0.596

We end this section with an illustrative numerical example that allows us to show the functionality of the two segmentation techniques used and their differences. For this purpose, three reservations have been taken into consideration (see Table 4). In general, the first two, despite having a more different price of stay than the second from the third, it is observed that customers have a more similar behavior from the point of view of propensity to consume. In this sense, this is precisely the aim of this article, that is, to incorporate in K-means a measure that takes into account how far apart the values of the descriptive variables are from each other coordinate-to-coordinate (not only globally as the Euclidean distance does) in such a way that great coordinate-to-coordinate differences cannot be compensated for with other differences when the measure aggregates, providing the global difference. This was carried out without preprocessing the information normalization.

Table 4. Description of the reserves where DS, CS, ND, EPE, NV mean days of stay, cost of stay, number of diners, expenditure per reserve, and number of visits, respectively.

Reservation	DS	CS	ND	EPE	NV
r_1	3	EUR 755.62	11	EUR 100	8
r_2	10	EUR 2900	16	EUR 275.3	8
r_3	8	EUR 3000	6	EUR 80	11

In the following, we show that in effect, the new proposed distance, the OWDr, on the one hand achieves this and, on the other hand, it also has a relativizing effect, since it takes into account the scale of each of the descriptive variables in order to know whether the differences between them are significant or not. At the same time, it is illustrated that this is not the case for the Euclidean distance.

According to Table 5, when the Euclidean distance is applied to measure we obtain that the most similar reservations are r_1 and r_3 . Indeed, $d_{Euclidean}(r_1, r_2) = 4.6$, $d_{Euclidean}(r_2, r_3) = 3.8$ and $d_{Euclidean}(r_1, r_3) = 4.4$. This could imply that the K-means classifies them in the same cluster. However, it is clear that the spending pattern of the customers is well differentiated (they should not belong to the same cluster). It must be pointed out that values of the variables that describe the reserves were previously normalized before computing the Euclidean distance. This done because the classical implementation of K-means with Euclidean distance is carried out in such a way that the data to be analyzed are those previously normalized in order to minimize scale differences.

Table 5. Distances between reserves (normalized data) measured with Euclidean distance.

	r_1	r_2	r_3
r_1	0	4.6	4.4
r_2	4.6	0	3.8
r_3	4.4	3.8	0

However, using the OWDr with vector weight, for instance, $W = (0.4, 0.3, 0.2, 0.1, 0)$, the most similar reserves are r_1 and r_2 which is in line with expectations as the expenditure patterns are similar. Indeed, as shown in Table 6, $OWDr(r_1, r_2) = 0.66$, $OWDr(r_2, r_3) = 0.774$, and $OWDr(r_1, r_3) = 0.75$.

In order help the reader, we compute step-by-step the distance $OWDr(r_1, r_2)$. The remaining distances can be computed analogously.

Table 6. Distances between reserves measured with OWDr.

	r_1	r_2	r_3
r_1	0	0.66	0.75
r_2	0.66	0	0.774
r_3	0.75	0.774	0

Step 1. Given $r_1 = (3; 755.62; 11; 100; 8)$ and $r_2 = (10; 2900; 16; 275.3; 8)$, we compute, applying (2), the collection of relative distances $\{d_r(r_{11}, r_{21}), \dots, d_r(r_{15}, r_{25})\}$, where

$$d_r(r_{11}, r_{21}) = \frac{|3-10|}{\max(3,10)} = 0.7,$$

$$d_r(r_{12}, r_{22}) = \frac{|2,900-755.62|}{\max(2,900;755.62)} = 0.74,$$

$$d_r(r_{13}, r_{23}) = \frac{|11-16|}{\max(11,16)} = 0.31,$$

$$d_r(r_{14}, r_{24}) = \frac{|100-275.3|}{\max(100,275.3)} = 0.64,$$

$$d_r(r_{15}, r_{25}) = \frac{|8-8|}{\max(8,8)} = 0.$$

Step 2. We sort the elements of collection $\{d_r(r_{11}, r_{21}), \dots, d_r(r_{15}, r_{25})\}$ obtained in Step 1 in such a way that the first component will be the 1th-largest element of the collection, etc. This gives the new vector $(0.74; 0.7; 0.64; 0.31; 0)$.

Step 3. Applying (1), we obtain the global value

$$OWDr(r_1, r_2) = 0.4 \cdot 0.74 + 0.3 \cdot 0.7 + 0.2 \cdot 0.64 + 0.1 \cdot 0.31 + 0 \cdot 0 = 0.66.$$

Note that the computing methodology of OWDr itself calculates relative distances (normalized distances) and therefore does not require preprocessing (normalizing) of the data to be treated. Moreover, unlike the Euclidean distance, the new methodology prioritizes those larger differences in the overall computation of the dissimilarity measure.

In the light of the exposed computations, the two procedures detect significantly different patterns in the data and, in addition, the normalization of the data does not manage to avoid the aforementioned problems in relation to the use of the Euclidean distance.

6. Conclusions and Further Work

The main objective of this study was to understand the consumption pattern of tourists staying in a five-star hotel located in a mature sun-and-beach area, as a good understanding of customer demands allows for improving hotel performance [25]. Specifically, this involves being able to classify customers staying at the hotel according to their propensity to

spend. The objective is achieved regardless of the season and in a way that the methodology used can be implemented without requiring radical changes in the implementation of classical methodologies on the one hand, and in data processing on the other hand.

From a theoretical point of view, this study contributes to the literature by providing a method for the categorization of expenditure of tourists visiting the hotel. Moreover, it considers the hotel as a consumption center beyond being where tourists go to rest, introducing the concept of the nano-scale. Thus, the analysis of the consumption pattern at the hotel level is essential for managers of establishments located in mature sun-and-beach destinations that are in an advanced stage of consolidation. In this sense, it is more important for entrepreneurs to understand how tourists interact with hotel offers than to know the number of tourist arrivals, which is more or less stable depending on the consolidation stage of the destination. Methodologically, this article shows how the introduction of a distance based on the use of OWDr improves the performance (accuracy) of classical K-means up to 21.6% and reduces the number of iterations needed for the algorithm to converge up to 48.46%. Moreover, OWDr provides some more relevant advantages with respect to the Euclidean distance. It does not require normalization of the data or calculating relative distances and, in addition, it is sensitive to the coordinate-to-coordinate differences. These improvements in customer segmentation and the cost of implementation lead to an improvement in the profitability of the establishment directly, adapting prices and services to each segment in order to increase prices. In addition, there are indirect benefits in terms of knowing the consumption pattern of tourists, optimizing spaces and services such as terraces, and generating synergies between departments, all of which increase customer satisfaction.

For hotel managers, a better understanding of customer consumption patterns can facilitate the implementation of strategies that increase hotel performance. It can be achieved by an increase in revenue or by improving efficiency through synergies between departments [48–50]. However, the cost of implementing new technologies can be high and time-consuming, making it difficult to recover the investment made [51]. For this reason, being able to incorporate the OWA operators and the aforementioned distances based on them into widely used algorithms not only improves their segmentation capacity but also reduces the implementation costs with respect to other machine learning techniques, since with a modification of K-means, a technique used in most hotel companies, the profitability is obtained. Although the method has been applied to a higher category hotel due to the availability of data, the technique is applicable to any tourist establishment.

All experiments have been executed in Python and they have shown that the OWA-based K-means generally outperforms the classical K-means endowed with the Euclidean distance. Such an improvement has been tested through a ground truth, designed by the marketing department of the firm, which states the cluster to which each tourist belongs. Therefore, it seems that the OWA-based K-means could be used as an appropriate tool for classifying tourists in purely exploratory and predictive stages. However, the weight vectors defining the OWDr have been selected heuristically. A future line of research is the determination of the optimal weight vectors that define the OWDr for the customer data to be used in experiments. In this direction, a comparison of OWA-based K-means and K-means will be made in terms of the running time of computing, also taking into account the time taken to select the aforementioned optimal weights by the former algorithm. Furthermore, the OWA-based algorithm will be tested on a wider selection of datasets coming from different hotels sharing similar characteristics to the hotel considered in the present work. However, in an early rejuvenation phase of a mature destination, few hotels are conceptualized to offer other complementary services than breakfast.

Author Contributions: P.J.P.-V. contributed to formulation of research goals; management activities to annotate, scrub, and maintain research data; application of statistical and mathematical techniques; performing experiments, development of methodology; implementation of code; verification of results and experiments; visualization/data presentation; writing the initial draft. M.M.-R. contributed to formulation of research goals; management activities to annotate, scrub, and maintain research

data; application of statistical and mathematical techniques; performing experiments, development of methodology; implementation of code; verification of results and experiments; visualization/data presentation; writing the initial draft. C.M.-F. contributed to formulation of research goals; development of methodology; verification of results and experiments; coordination responsibility for the research activity planning and execution; critical review of the initial draft. O.V. contributed to formulation of research goals; application of statistical and mathematical techniques; performing experiments, development of methodology; implementation of code; verification of results and experiments; visualization/data presentation; critical review of the initial draft; acquisition of the financial support for the project leading to this publication. All authors have read and agreed to the published version of the manuscript.

Funding: The authors acknowledge financial support from Proyecto PGC2018-095709-B-C21 financiado por MCIN/AEI/10.13039/501100011033 y FEDER “Una manera de hacer Europa” and from project BUGWRIGHT2. The latter project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No. 871260. This publication reflects only the authors’ views and the European Union is not liable for any use that may be made of the information contained therein.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available because they belong to the hotel company.

Conflicts of Interest: The authors declare no conflict of interest, the company and the funders had no role in the design of the techniques and the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

- Hua, N.; Huang, A.; Medeiros, M.; DeFranco, A. The moderating effect of operator type: The impact of information technology (IT) expenditures on “hotels’ ” operating performance. *Int. J. Contemp. Hosp. Manag.* **2020**, *32*, 2519–2541. [[CrossRef](#)]
- Martorell, O.; Mulet, C. The franchise contract in hotel chains: A study of hotel chain growth and market concentrations. *Tour. Econ.* **2010**, *16*, 493–515.
- Hsu, C.W.; Liu, H.Y. Corporate diversification and firm performance: The moderating role of contractual manufacturing model. *Asia Pac. Manag. Rev.* **2008**, *13*, 345–360.
- Butler, R. The concept of a tourist area cycle of evolution: Implications for management of resources. *Can. Geogr.* **1980**, *24*, 5–12. [[CrossRef](#)]
- Tortella, B.D.; Tirado, D. Hotel water consumption at a seasonal mass tourist destination. The case of the island of Mallorca. *J. Environ. Manag.* **2011**, *92*, 2568–2579. [[CrossRef](#)]
- Aguiló, E.; Juaneda, S.C. Tourist expenditure for mass tourism markets. *Ann. Tour. Res.* **2000**, *27*, 624–637. [[CrossRef](#)]
- Bujosa, A.; Riera, A.; Pons, P.J. Sun-and-beach tourism and the importance of intra-destination movements in mature destinations. *Tour. Geogr.* **2015**, *17*, 780–794. [[CrossRef](#)]
- Eagles, P.F. The travel motivations of canadian ecotourists. *J. Travel Res.* **1992**, *31*, 3–7. [[CrossRef](#)]
- Horner, S.; Swarbrooke, J. *Consumer Behaviour in Tourism*; Taylor & Francis: Oxfordshire, UK, 2016.
- Chang, K.G.; Chien, H.; Cheng, H.; Chen, H.-i. The impacts of tourism development in rural indigenous destinations: An investigation of the local residents’ perception using choice modeling. *Sustainability* **2018**, *10*, 4766. [[CrossRef](#)]
- DÚrso, P.; Disegna, M.; Massari, R.; Osti, L. Fuzzy segmentation of postmodern tourists. *Tour. Manag.* **2016**, *55*, 297–308. [[CrossRef](#)]
- Poon, A. *Tourism, Technology and Competitive Strategies*; CAB International: Wallingford, UK, 1993.
- Potter, R.; Phillips, J. The rejuvenation of tourism in Barbados 1993–2003: Reflections on the Butler model. *Geography* **2004**, *89*, 240–247. [[CrossRef](#)]
- Institut Balear D’ Estadística. Gasto de los Turistas con Destino Principal las Illes Balears por Periodo y tipo de Alojamiento. 2020. Available online: https://ibestat.caib.es/ibestat/estadistiques/f58f0937-c64f-469d-bad5-99f29bbb59ce/755a8af8-2b59-41ee-9c2d-9aa0f4f8509f/es/I208004_n002.px (accessed on 2 May 2022).
- Lai, I.K.W.; Hitchcock, M. Sources of satisfaction with luxury hotels for new, repeat, and frequent travelers: A PLS impact-asymmetry analysis. *Tour. Manag.* **2017**, *60*, 107–129. [[CrossRef](#)]
- Kim, J.; Kim, S.I.; Lee, M. What to sell and how to sell matters: Focusing on luxury hotel properties’ business performance and efficiency. *Cornell Hosp. Q.* **2022**, *63*, 78–95. [[CrossRef](#)]
- Mun, S.G.; Woo, L.; Paek, S. How important is F&B operation in the hotel industry? Empirical evidence in the U.S. market. *Tour. Manag.* **2019**, *75*, 156–168.

18. Chen, C.M.; Chang, K.L. Diversification strategy and financial performance in the Taiwanese hotel industry. *Int. J. Hosp. Manag.* **2012**, *31*, 1030–1032. [[CrossRef](#)]
19. Lee, M.J.; Jang, S.S. Market diversification and financial performance and stability: A study of hotel companies. *Int. J. Hosp. Manag.* **2007**, *26*, 362–375. [[CrossRef](#)]
20. López-Picos, Y.; Otero-González, L.; Lado-Sestayo, R. Efectos de la diversificación en el binomio rentabilidad-riesgo. Un análisis del sector hotelero. *Gran Tour Rev. Investig. Turísticas* **2017**, *16*, 3–22.
21. Miñana, J.; Morro, M.; Valero, O. A modification of OWD aggregation operator and its application to k-means algorithm. In Proceedings of the Simulation and Modelling 2019, Palma, Spain, 28–30 October 2019; pp. 48–52.
22. Yager, R.R. On ordered weighted averaging aggregation operators in multicriteria decisionmaking. *IEEE Trans. Syst. Man Cybern.* **1988**, *18*, 183–190. [[CrossRef](#)]
23. Grosskopf, S. Efficiency and Productivity. In *The Measurement of Productive Efficiency: Techniques and Applications*; Oxford University Press: Oxford, UK, 1993; pp. 160–194.
24. Oltean, F.D.; Gabor, M.R. Service diversification a qualitative and quantitative analysis in Mures county hotels. *Eng. Econ.* **2016**, *27*, 618–628. [[CrossRef](#)]
25. Chesters, C. The Business: Redefining Accor’s Bussiness F&B Strategy. 2017. Available online: <https://www.hotelnewsme.com/catering-news-me/business-redefining-accors-fb-strategy/> (accessed on 1 May 2022).
26. Chari, M.D.; David, P.; Duru, A.; Zhao, Y. Bowman’s risk-return paradox: An agency theory perspective. *J. Bus. Res.* **2019**, *95*, 357–375. [[CrossRef](#)]
27. Grant, R.M.; Jammine, A.P.; Thomas, H. Diversity, diversification, and profitability among British manufacturing companies, 1972/1984. *Acad. Manag. J.* **2017**, *31*, 771–801. [[CrossRef](#)]
28. Erkuş-Öztürk, H. Diversification of Hotels in a Single-Asset Tourism City. In *Tourism and Hospitality Management*; Emerald Group Publishing Ltd.: Bingley, UK, 2016; pp. 173–185.
29. Lei, C.K. The influences of revenue diversification and incoming tourists on the performance of star-rated hotels in China. *Tour. Anal.* **2019**, *24*, 483–495. [[CrossRef](#)]
30. Park, K.; Jang, S.C.S. Effect of diversification on firm performance: Application of the entropy measure. *Int. J. Hosp. Manag.* **2012**, *31*, 218–228. [[CrossRef](#)]
31. Shoal, N.; McKercher, B.; Ng, E.; Birenboim, A. Hotel location and tourist activity in cities. *Ann. Tour. Res.* **2011**, *38*, 1594–1612. [[CrossRef](#)]
32. Kau, A.K.; Lim, P.S. Clustering of Chinese tourists to Singapore: An analysis of their motivations, values and satisfaction. *Int. J. Tour. Res.* **2005**, *7*, 231–248. [[CrossRef](#)]
33. Morey, R.C.; Dittman, D.A. An aid in selecting the brand, size and other strategic choices for a hotel. *J. Hosp. Tour. Res.* **1997**, *21*, 71–99. [[CrossRef](#)]
34. Chen, C.F. Applying the stochastic frontier approach to measure hotel managerial efficiency in Taiwan. *Tour. Manag.* **2007**, *28*, 696–702. [[CrossRef](#)]
35. Najmi, M.; Sharbatoghlie, A.; Jafari, A. Tourism market segmentation in Iran. *Int. J. Tour. Res.* **2010**, *12*, 497–509. [[CrossRef](#)]
36. Jain, A.K. Data clustering: 50 years beyond K-means. *Pattern Recognit. Lett.* **2010**, *31*, 651–666. [[CrossRef](#)]
37. Peña, J.M.; Lozano, J.A.; Larrañaga, P. An empirical comparison of four initialization methods for the K-Means algorithm. *Pattern Recognit. Lett.* **1999**, *20*, 1027–1040. [[CrossRef](#)]
38. Tianyang, W. A K-means Group Division and LSTM Based Method for Hotel Demand Forecasting. *Tech. Gaz.* **2021**, *28*, 1345–1352.
39. DÚrso, P.; Disegna, M.; Massari, R.; Prayag, G. Bagged fuzzy clustering for fuzzy data: An application to a tourism market. *Knowl.-Based Syst.* **2015**, *73*, 335–346. [[CrossRef](#)]
40. Grabisch, M.; Marichal, J.; Mesiar, R.; Pap, E. *Aggregation Functions*; Cambridge University Press: Cambridge, UK, 2009.
41. Xu, Z.; Chen, J. Ordered weighted distance measure. *J. Syst. Sci. Syst. Eng.* **2008**, *17*, 432–445. [[CrossRef](#)]
42. Thrane, C. Analyzing tourists? length of stay at destinations with survival models: A constructive critique based on a case study. *Tour. Manag.* **2012**, *33*, 126–132. [[CrossRef](#)]
43. Abrate, G.; Nicolau, J.; Viglia, G. The impact of dynamic price variability on revenue maximization. *Tour. Manag.* **2019**, *74*, 224–233. [[CrossRef](#)]
44. Soler, I.P.; Gemar, G.; Correia, M.B.; Serra, F. Algarve hotel price determinants: A hedonic pricing model. *Tour. Manag.* **2019**, *70*, 311–321. [[CrossRef](#)]
45. Hendsill, C. Partnerships in dining. *Hotels* **1996**, *30*, 57–60.
46. Spanish Institute of Statistics. Expenditure of International Tourists According to Autonomous Community of Main Destination. 2022. Available online: <https://www.ine.es/jaxiT3/Datos.htm?t=10839> (accessed on 1 May 2022).
47. Visa, S.; Ramsay, B.; Ralescu, A.L.; Knaap, E.V.D. Confusion matrix-based feature selection. *MAICS* **2011**, *710*, 120–127.
48. Chalupa, S.; Petricek, M. Using technology and customer behaviour characteristics to improve hotel sales performance. *TEM J.* **2020**, *9*, 573–577. [[CrossRef](#)]
49. Maier, T.; Johanson, M. An empirical investigation into convention hotel demand and ADR trending. *J. Conv. Event Tour.* **2013**, *14*, 2–220. [[CrossRef](#)]

-
50. Talón-Ballester, P.; González-Serrano, L.; Soguero-Ruiz, C.; Muñoz-romero, S.; Rojo-Álvarez, J. Using big data from customer relationship management information systems to determine the client profile in the hotel sector. *Tour. Manag.* **2018**, *68*, 187–197. [[CrossRef](#)]
 51. Piccoli, G. Information technology in hotel management: A framework for evaluating the sustainability of IT-dependent competitive advantage. *Cornell Hosp. Q.* **2008**, *49*, 282–296. [[CrossRef](#)]