

Article

Challenge for Planning by Using Cluster Methodology: The Case Study of the Algarve Region

David Bienvenido-Huertas ^{1,*}, Fátima Farinha ², Miguel José Oliveira ², Elisa M. J. Silva ²
and Rui Lança ²

¹ Department of Building Construction II, University of Seville, Av. Reina Mercedes 4A, 41012 Seville, Spain

² University of Algarve, Institute of Engineering, Campus da Penha, 8005-139 Faro, Portugal;
mfarinha@ualg.pt (F.F.); mjolivei@ualg.pt (M.J.O.); esilva@ualg.pt (E.M.J.S.); rlanca@ualg.pt (R.L.)

* Correspondence: jbienvenido@us.es

Received: 31 January 2020; Accepted: 17 February 2020; Published: 18 February 2020



Abstract: This study analyses the most appropriate methodology to make similarity classifications among the cities of the Algarve (Portugal) according to 105 sustainability indicators monitored with the Observatory of Sustainability of the Algarve Region for Tourism (OBSERVE). The methodology used to establish the similarities was the cluster analysis with 4 different approaches which reduced the dimensions of the data set: total approach, pillar approach, subject area approach, and indicator approach. By combining the approaches, a total of 620 different cluster analyses were performed. The results reflected that the data analysis approaches with less dimensions were those that performed the best groups among cities. In this sense, the approaches with a high number of indicators (e.g., the total or the pillar approach) were characterised by misclassifying cities in more than 30% of the indicators. Thus, the most acceptable cluster analysis approach was that with a low number of indicators. Through this approach, it was possible to make correct groups of the sustainability level of the cities of the Algarve. These results provided an appropriate methodology for the decision-making regarding the sustainability of a region and could be extrapolated to other regions to assess sustainability or environmental indicators.

Keywords: cluster analysis; sustainability indicators; the Algarve region; OBSERVE platform; sustainable tourism

1. Introduction

1.1. Tourism and Sustainability Indicators

The sustainable development of societies is an important issue belonging to the governmental policy to achieve a more sustainable development. In this regard, one of the activities of the cities which significantly influences their sustainability level is their tourist activity [1–3]. This activity similarly influences as other activities do, such as urban, agricultural or maritime [4–6], due to the impact of the tourist activity of cities in their social, economic and environmental dimensions. A positive and clear aspect of such activities is the improvement of employability in a region by generating job positions, although some negative aspects such as the alteration of lifestyles, damage in the heritage or the alteration of ecosystems could be a sign of the unsustainable impact of the activity in a region [7]. In addition, tourist activities affect the sustainability of other sectors, such as infrastructures [8]. Without an adequate control, the expansion of the tourism market will increase the pressure on the ecosystems on which the livelihood of local communities depends [7]. Therefore, tourist offers in regions or cities should be managed in a broader sustainability context [9], also guaranteeing the economic advancement of such activities [10]. Improving the sustainability usually also leads to an improvement of competitiveness [11,12].

Therefore, it is necessary to control the sustainability of the tourist activity of a region, thus making possible to guarantee that the tourist offer could contribute to the economy of the region without affecting its inhabitants. For this purpose, the monitoring of sustainability indicators allows the evolution of a region to be determined and, in this way, appropriate policies to be established towards a more sustainable tourism [13,14]. Sustainability indicators are understood as those tools contributing to the analysis and assessment of the information so that managers could make right decisions [15]. As general criterion, sustainability indicators should be quantitative to be assessed [16].

However, there are not defined criteria about which such indicators should be, thus varying among different research studies [17]. Some examples are as follows: (i) Miller [18] established 9 indicators for sustainable tourism through a Delphi survey of tourism researchers; (ii) Liu et al. [19] considered 20 indicators according to the parties interested: tourists, local residents, governmental agencies, and business owners; (iii) Blancas et al. [20] used a set of 32 indicators to assess the sustainability of coastal tourist destinations based on 3 dimensions: social, economic, and environmental. In a subsequent study [21], these authors considered a set of 89 indicators to assess the sustainable tourism based on the same dimensions (social, economic, and environmental); (iv) Nesticò and Maselli [22] established a set of 23 indicators for the economic evaluation of tourism projects in the islands; and (v) Castellani and Sala [23] used 20 indicators concerned with: tourism characteristics of the region under investigation; environmental factors; economic and social conditions of local communities; and demographic dynamics.

In any case, the use of a set of indicators will allow the behaviour of different aspects of tourism to be evaluated in the different regions of each study. Through this evaluation, the necessary corrective policies may be established by the government of the region.

1.2. Data Analysis of Sustainability Indicators

To establish corrective measures, an essential aspect is the treatment of the compiled information of the monitored indicators [24]. Some of these analyses are based on spatial distribution through maps of the sustainability of a region. In this regard, Hély and Antoni [25] developed a grid analysis map of the Besançon region (France) from which the strengths and weaknesses of the territory can be identified in terms of sustainability. In another similar study, Palmisano et al. [26] analysed the spatial distribution of a set of rural sustainability indicators to establish a common Rural Sustainable Development strategy to allocate the European Agricultural Fund for Rural Development Budget. Through these indicator analyses, decisions could be established using different methods, such as decision-making matrix [27] and fuzzy logic [28].

However, the analysis of sustainability indicators could be complex when various typologies are analysed. In this sense, many studies have evaluated the possibility of conducting a cluster analysis. A cluster analysis is a multivariant statistical technique which allows a set of objects to be classified in a way that, on the one hand, similar objects are in the same conglomeration and, on the other hand, different objects are in several groups, resulting in various homogeneous groups among them. Thus, this analysis has been used in previous research studies to make groups between indicators: (i) Akande et al. [29] classified 32 indicators of the Smart level and of sustainability of European cities in 5 components through hierarchical clustering; (ii) Yi et al. [30] assessed the sustainability of the 17 cities of the area of Shandong (China). For this purpose, they used 21 indicators of environmental, social, and economic dimensions and classified the 17 cities in 4 groups through the average values and growth values; (iii) Dang et al. [31] applied a classification analysis of the indicators included in the certifications of China's new Assessment Standard for Green Eco-districts (ASGE) and of Leadership in Energy and Environmental Design for Neighbourhood Development (LEED-ND). To do this, authors used k-means to make groups; and (iv) Neri et al. [32] conducted cluster analyses in an indicator framework of input-state-output to group 83 countries according to their sustainability. A total of 3 indicators (the emergency flow per capita, the Gini index of income distribution and the Gross Domestic Product per Capita (GDP)) and the k-median algorithm were used for the analysis. However,

more research studies analysing new methodologies are required to assess the sustainable tourism with the aim to reduce current existing limitations [33]. Assessment methodologies of regions are included in such new methodologies. Previous research studies did not analyse the limitations of group techniques when the various monitored indicators are individually analysed. Additionally, there are few research works analysing a broad sample of sustainability indicators.

1.3. Aim of This Study

For this reason, this study makes different methodologies to group the cities of a region based on their sustainability indicators. In this way, the strategical decision-making of local governments are contributed to be in accordance with the United Nations Sustainable Development Goals [34]. This study therefore aims to analyse the most appropriate methodology to make similarity classifications among cities of a region according to sustainability indicators. For this purpose, the case study used is the Algarve region due to the importance of tourism in the region [35]. The sustainability indicators in the region are monitored through the Observatory of Sustainability of the Algarve Region for Tourism (OBSERVE) platform [36]. Based on the data compiled by OBSERVE in 105 sustainability indicators between 2011 and 2015, this study conducts different methodological cluster analysis approaches. The results determined the most appropriate methodological approach to group the 105 indicators appropriately and guaranteed consistent classifications of the cities to establish appropriate policies in each indicator. Thus, the results of this study also analysed the possibilities of determining similarities among cities when analysing a large number of sustainability indicators.

2. Methodology

2.1. Area of Study: Algarve Region

The Algarve Region is in the south of Portugal and is made up of 16 municipalities whose names come from the capital city of each (see Figure 1). This region has a coast of about 200 km long and is the most important tourist destination in Portugal [35] (with 43.8% of the total overnight stays [37]) and in Europe [38]. This aspect becomes important as the incomes from tourism in Portugal corresponded to 7% of the GDP of the country and 6.3% of employment in the year 2016 [39]. Sustainability is therefore one of the 10 challenges proposed by the Portuguese tourism policy for the next 10 years [40]. It is important to note that tourism has evolved continuously since the construction of the Faro Airport in 1965 [41], thus implying a large number of visitors in the present days. In this sense, the region received around 2.7 million international visitors in 2015 [42].



Figure 1. The Algarve region and municipalities (capital cities are drawn with a circle).

The beaches of the region are the main tourist attraction [35] and traditionally the most valued attribute by tourists [43]. Therefore, this destination attracts both national and international tourists [44,45]. It is also important to highlight the relevance of golf facilities in the Algarve region. In this regard, the region has expanded its range of golf since 1990 [46], thus making this type of tourism one of the best counterweights to the strong seasonality of the region [47].

In addition, there is a wide range of existing accommodation, including luxury hotels and hostels [48]. Consequently, the Algarve is the region of Portugal with the greatest tourist activity.

2.2. OBSERVE Platform

Controlling the sustainability level of a region through observatories is essential to guarantee the correct development of urban environments [49]. For this reason, the sustainability of the Algarve region can be measured by the OBSERVE platform [36], whose objective is to monitor various sustainability indicators classified in 4 dimensions (also known as pillars) (see Figure 2): environmental, institutional, economic, and sociocultural. The indicators for each of these pillars were chosen based on the consensus of various public bodies and institutions, such as the Algarve Hotels Association (in Portuguese, Associação dos Hotéis e Empreendimentos Turísticos do Algarve) and the Algarve Regional Coordination and Development Commission (in Portuguese, Comissão de Coordenação e Desenvolvimento Regional do Algarve) [36]. Thanks to meetings and surveys, the indicators to be monitored by OBSERVE were determined [36]. Table 1 summarises the OBSERVE platform's sustainability indicators [36].

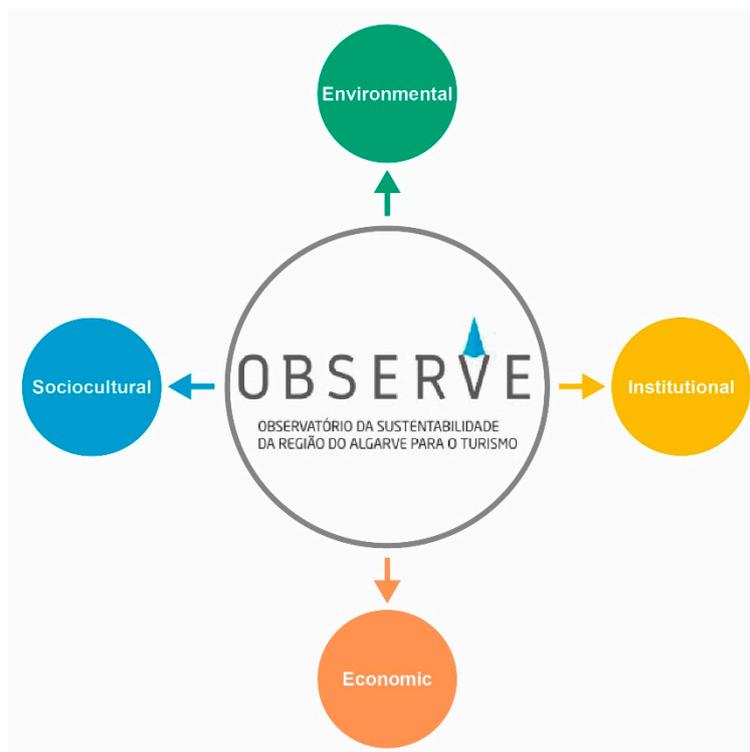


Figure 2. The 4 dimensions of the OBSERVE platform's sustainability indicators.

Table 1. The OBSERVE platform's sustainability indicators [36].

| Pillar | Subject Area | Indicator |
|---|---|--|
| Environmental | Air quality index | Air quality index |
| | Energy management | Consumption of electric energy by inhabitant |
| | | Consumption of motor fuel by inhabitant |
| | Environmental management | Production of electric energy using renewable sources |
| | | Environmental expenditure of municipalities by 1000 inhabitants |
| | | Number of blue flags, beaches and marinas |
| | Materials and waste management | Number of bathing waters and quality classes |
| | | Percentage of municipal waste prepared for reuse and recycling |
| | Mobility | Urban waste selectively collected per inhabitant |
| | | Urban waste collected per capita |
| Number of embarked and disembarked passengers in Faro Airport | | |
| Number of passengers per kilometre carried by enterprises exploring inland transportation | | |
| Natural capital management | Movement of passengers in inland waterways | |
| | Number of embarked and disembarked passengers of cruise ships in Portimão Port | |
| | Charging stations for electric vehicles | |
| Water cycle management | Average daily traffic on main roads and secondary roads | |
| | Walking and cycling routes | |
| | Burnt area | |
| Institutional | Governance and citizenship | Investments on protection of biodiversity and landscapes of municipalities |
| | | Percentage of safe water |
| | Innovation and knowledge | Fresh water supplied per inhabitant |
| | | Wastewater sewerage per capita |
| | | Abstention rate |
| | Percentage of capital expenditure | |
| | Broadband internet access per 100 inhabitants | |
| | Gross expenditure on research and development of institutions and enterprises | |
| | Gross expenditure on research and development as percentage of gross domestic product | |

Table 1. Cont.

| Pillar | Subject Area | Indicator | |
|---------------|---|---|--|
| Economic | Economic impact | Gross value added of enterprises | |
| | | Apparent labour productivity in establishments, food and beverage service activities | |
| | | Inflation | |
| | | Per capita purchasing power | |
| | | Number of establishments and economic activity | |
| | | Persons employed of establishments and economic activity | |
| | | Turnover of establishments and economic activity | |
| | | Relative contribution of establishments, food and beverage service activities to the Algarve Economy (GVA of Enterprises) | |
| | | Job | Employment by gender and economic sector |
| | | | Seasonal employees |
| | | Seasonality | Establishments open all year |
| | | | Lodging capacity in hotel establishments |
| | | Tourist occupation | Nights in hotel establishments |
| | | | Revenue per available room (Rev Par) of hotel establishments |
| Accessibility | Average stay in hotel establishments | | |
| | Number of accessible beaches | | |
| Culture | Number of cultural properties | | |
| | Municipalities' expenditures on cultural heritage | | |
| Sociocultural | Demography | Annual population balances: natural and migratory | |
| | | Resident population | |
| | Education | Foreign population with status of resident | |
| | | Population education level with 15 and more years | |
| | Health care | Health Care | |
| | | Lodging capacity in hotel establishments by 1000 inhabitants | |
| | Pressure | Tourist intensity | |
| | | Regional tourist density | |
| | | Municipal tourist density | |
| | Safety | Crime rate | |
| | | Number of registered crimes | |
| | Social cohesion | Regional development composite index (cohesion) | |
| | | Beneficiaries of social integration income, of social security per 1000 inhabitants in active age | |
| | | Number of secondary houses per 100 houses | |

2.3. Group Approaches

There is much information compiled by the OBSERVE platform. However, it is necessary to establish appropriate procedures to analyse indicators and, in this way, to establish the most appropriate performance patterns to mitigate possible unsustainable values in some zones of the Algarve. One of the first steps should be the classification of the cities of the region according to the values recorded in each sustainability indicator.

For this reason, this study assessed the possibilities to group the 16 cities of the Algarve (see Figure 1). A total of 105 indicators from the OBSERVE platform related to the sustainability of the region were selected. Some of the indicators correspond to developments of a particular indicator. For example, the crime rate indicator was divided into different subcategories, such as crimes against people and crimes against heritage. Likewise, some subject areas included in Table 1 were not considered as data of each city were not available (e.g., the subject area of mobility). Additionally, indicators were intended to be analysed in a long temporary period. For this study, the period 2011–2015 was considered as a wide sample of indicators with data for where the research is available.

A total of 4 various approaches were used for cluster analyses. Such approaches were based on the structure used by the OBSERVE platform to classify indicators: Total approach (TA), Pillar approach (PA), Subject area approach (SAA), and Indicator approach (IA). These approaches suppose that the dataset used in the cluster analysis has a lower number of variables from left to right, so that the TA groups all indicators (i.e., it corresponds to a multidimensional group), whereas the IA corresponds to individual analysis of each indicator (i.e., it corresponds to a 1D cluster).

The analysis was carried out for the period 2011–2015, and the approaches were independently analysed in each year. Table 2 includes the cluster analysis per approach and year. The results of this research are based on a total of 620 clusters.

Table 2. Cluster analyses carried out in the research.

| Year | Approach | | | |
|------|----------|----|-----|-----|
| | TA | PA | SAA | IA |
| 2011 | 1 | 4 | 14 | 105 |
| 2012 | 1 | 4 | 14 | 105 |
| 2013 | 1 | 4 | 14 | 105 |
| 2014 | 1 | 4 | 14 | 105 |
| 2015 | 1 | 4 | 14 | 105 |

2.4. *k*-Means

The algorithm *k*-means was used for cluster analyses. *k*-means is an iterative clustering algorithm based on the centroid concept of a group of individuals [50]. The method is based on an X sample of n individuals classified in k groups, for which a W partition of such sample with $W = (w_1, \dots, w_a, \dots, w_b, \dots, w_k)$ is considered, thus achieving that the total sum of the sums of squares of the Euclidean distances within each group is minimum:

$$\operatorname{argmin}_W \sum_{a=1}^k \sum_{x_i \in w_a} \sum_{r=1}^p (x_{ir} - \mu_{ar})^2 \quad (1)$$

At the performance level, the *k*-means algorithm's steps are as follows:

- Step 1: the number of k groups is identified to carry out the analysis.
- Step 2: k individuals from the dataset are randomly selected, constituting the initial centroids.
- Step 3: by using the association measurement chosen, the distance of each individual to each k centroid is calculated.
- Step 4: k groups are created by allocating each individual to the closest centroid.
- Step 5: the new centroids of each k group are identified.
- Step 6: steps 3 and 4 are repeated. This step could lead to two situations: (i) going to step 5 if in step 4 some of the individuals change the group, thus repeating the cycle; and (ii) the cluster analysis process is finished when no individual changes the group in step 4.

The method is sensitive to initial centroids, so different results could be given by varying the initial values of k . In this sense, the greater the k used in the algorithm, the lower the variation within groups (i.e., more individual groups are usually created, thus losing the main potential of the analysis: to detect similarity patterns among individuals). If the variables have various units (as in this research), a pre-processing to normalise data should be conducted before the cluster analysis (i.e., the variables are rescaled between 0–1 by using a min-max normalisation).

To optimally select the number of clusters, a total of 3 different analyses were used in this research. Such analyses were based on the Elbow method, the silhouette index ($s(i)$), and the ratio between the sum of squares and the total sum of squares (BSS/TSS).

The Elbow method consists in selecting the optimal number of k by minimising the total within-cluster sum of squares (WSS) [51]. The Elbow method is made up of 4 phases:

1. *k*-means is applied for different values of k .

2. For each k , WSS is calculated:

$$WSS = \sum_{k=1}^K \sum_{i \in S_k} \sum_{j=1}^p (x_{ij} - \bar{x}_{kj})^2 \quad (2)$$

where S_k is the set of instances grouped in the k -th cluster, and \bar{x}_{kj} is the j -th variable of the cluster center for the k -th cluster.

3. The WSS curve is plot with respect to the number of k groups.
 4. The location of the elbow in the graphic is generally considered as an indicator of the optimal number of groups (see Figure 3).

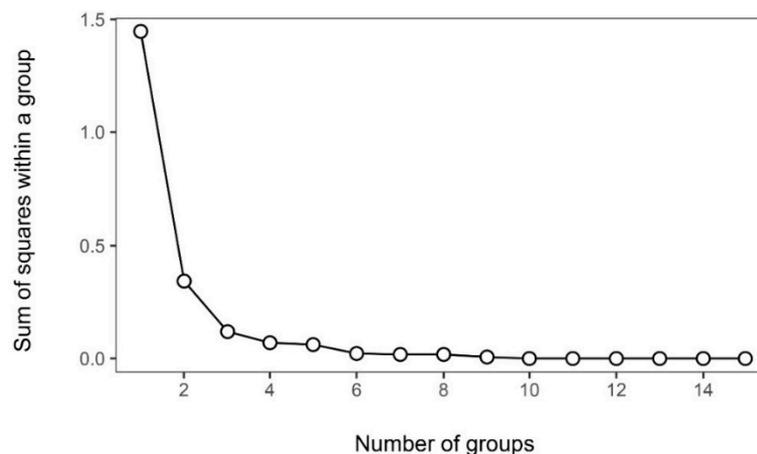


Figure 3. Example of the total within-cluster sum of squares of each cluster conducted in the study.

The elbow of the graphic can be clearly seen [51]. This characteristic especially takes place in cases in which there is a gradual and continuous data transition. For these cases, the method does not provide a unique possible solution, but several possible solutions which should be inspected to determine the best. For this reason, this study combines the Elbow method with two indicators: $s(i)$ and BSS/TSS.

The BSS/TSS ratio is a relation of the cluster compactness. It is a percentage relation, with values between 0 and 100%. The greater the ratio value, the greater the compactness of individuals within a group. Likewise, given that $TSS = BSS + WSS$, by having a greater BSS, WSS will be lower. The ratio formulation is as follows:

$$\frac{BSS}{TSS} = \frac{\sum_{k=1}^K \sum_{j=1}^p (\bar{x}_{kj} - \bar{x}_G)^2}{\sum_{k=1}^K \sum_{j=1}^p (\bar{x}_{kj} - \bar{x}_G)^2 + \sum_{k=1}^K \sum_{i \in S_k} \sum_{j=1}^p (x_{ij} - \bar{x}_{kj})^2} \quad (3)$$

where \bar{x}_G is the grand mean of the means of each group.

Finally, $s(i)$ is among the most used indexes in the cluster analysis [52]. The index shows the similarity of an individual with the rest of the individuals of a same group. So, it measures the quality of a group. For this purpose, the following equation is used:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (4)$$

where $a(i)$ is the average distance between the individual (i) and the rest of points of the same group; and $b(i)$ is the minimum distance between the individual and the rest of groups. The silhouette index could obtain values between -1 and 1. The meaning of such values determines the suitability of the

cluster analysis: (i) if the value is between 0 and 1, the observation is correctly grouped, obtaining optimal values those groups closer to 1; (ii) if the value is 0, the individual is between two groups, thus meaning that either the individual has very different characteristics from the rest which could not be grouped with the others or that the cluster analysis has excessively classified the individual groups; and (iii) if the value is between -1 and 0, the individual is placed in the incorrect group. Figure 4 shows an example of one analysis of the silhouette index followed in the research.

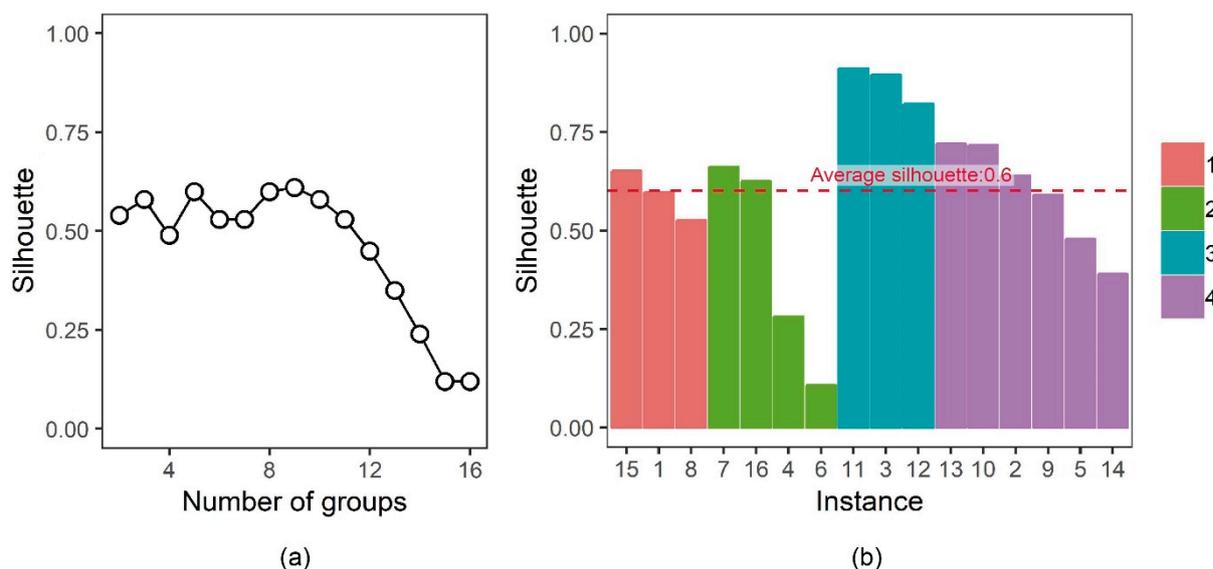


Figure 4. Example of the silhouette index of each cluster conducted in the study: (a) evolution of the average silhouette with variations of the number of groups; and (b) distribution of the silhouette in each town.

Thus, the control of this value leads us to know whether individuals are correctly grouped. It is important to stress that, in a multidimensional cluster analysis, the silhouette index obtained is an average of the various dimensions. Although the average silhouette value is high, there are erroneous similarity patterns among the different variables.

3. Results and Discussion

Firstly, the optimal number of k for each approach was determined in the cluster analysis. For this purpose, the elbow method and the analysis of $s(i)$ and of BSS/TSS were used. Through this assessment, the optimal number of k was determined in the 620 clusters conducted in the research.

After determining the optimal number of k for each approach, the statistical parameters obtained from $s(i)$ and BSS/TSS were analysed to assess the most appropriate approach for the cluster analysis of the Algarve region's sustainability indicators. Figure 5 includes the distributions of BSS/TSS obtained among the different approaches in the 5 years analysed with violin-plots. Violin-plots are an evolution of box-plots by including information of the kernel density and rotating them to both sides of the box [53]. As can be seen, the values of BSS/TSS obtained were high as all groups obtained ratios greater than 70% due to the process of optimal selection of k followed in the research. However, the use of approaches with a lower number of dimensions in the cluster analysis increases the BSS/TSS ratio. In this regard, the use of the 1D approach of indicators allowed average values of BSS/TSS greater than 94% to be obtained, with an increase with respect to the TA between 3.75 and 12.84% in all the years analysed (see Table 3). This approach to reduce the dimensions of the cluster analysis was the only approach obtaining better results in all years, as in the PA and SAA, the behaviour was different depending on the year: (i) in the PA, better results were obtained in 2013 and 2014, whereas in the other years, the BSS/TSS ratio was lower than that of the TA; and (ii) in the SAA, better results were

obtained in 3 years, whereas results were worse in the other 2. These results show the great variability that the BSS/TSS indicator could present in the cluster analyses carried out with a high number of variables. In general terms, the reduction of dimensions of the dataset of sustainability indicators used in the cluster analysis could improve group compactness, although this aspect is only guaranteed by 1D approaches. In addition, despite that the average values of indicator groups were better, most of the distribution was in higher values with respect to the TA (see Figure 5). Depending on the year, between 80 and 98% of clusters of the approach of indicators obtained better values in BSS/TSS.

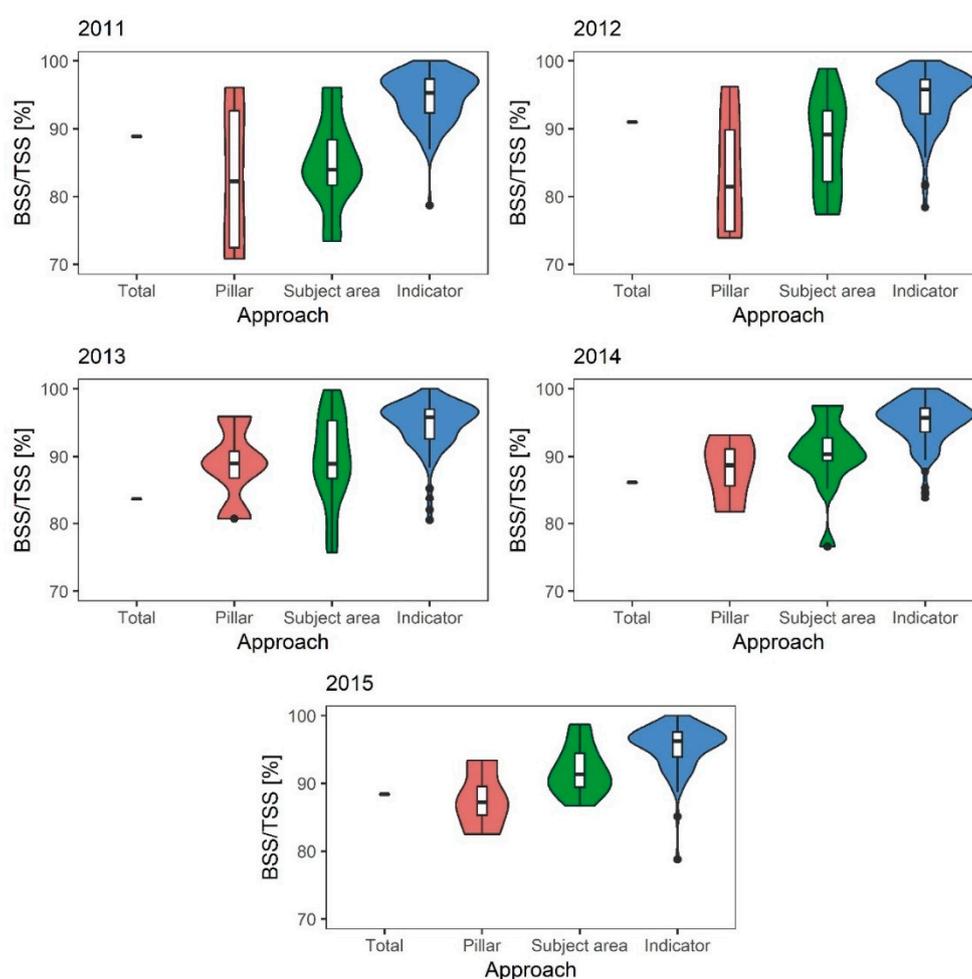


Figure 5. Violin- and box-plots with the distributions of BSS/TSS obtained in the different cluster analysis approaches.

Table 3. Average value of BSS/TSS and deviation percentage of the approaches with less dimensions (PA, SAA, and IA) with respect to the TA. Positive values in the deviation percentage indicate an increase of the ratio, and negative values imply a reduction.

| Year | TA | | PA | | SAA | | IA | |
|------|---------|---------|---|---------|---|---------|---|--|
| | BSS/TSS | BSS/TSS | Percentage Deviation with Respect to the TA [%] | BSS/TSS | Percentage Deviation with Respect to the TA [%] | BSS/TSS | Percentage Deviation with Respect to the TA [%] | |
| 2011 | 88.9 | 82.85 | −6.81% | 84.84 | −4.57% | 94.59 | 6.40% | |
| 2012 | 91 | 83.25 | −8.52% | 87.64 | −3.69% | 94.41 | 3.75% | |
| 2013 | 83.7 | 88.65 | 5.91% | 89.53 | 6.97% | 94.45 | 12.84% | |
| 2014 | 86.2 | 88.08 | 2.18% | 90.34 | 4.80% | 95.08 | 10.30% | |
| 2015 | 88.4 | 87.6 | −0.90% | 91.97 | 4.04% | 95.39 | 7.91% | |

However, the BSS/TSS ratio is not the only aspect determining which approaches with less dimensions allow better classifications of the sustainability indicators to be obtained. This aspect was also reflected in $s(i)$. Furthermore, this index also assesses the degree of correct classification conducted by the analysis, as low values of the silhouette index could mean that either the cities have not been places in the correct group or that the cluster analysis has generated too many individual groups. Figure 6 represents the violin-plots with the distributions of the average silhouette index obtained by each cluster. For the silhouette index, the use of approaches with less variables lead to a better classification of cities. The TA was that obtaining the lowest values of $s(i)$. Likewise, $s(i)$ did not get worse in this case, unlike BSS/TSS (see Table 4). So, the reduction of dimensions in the cluster analysis progressively improved the increase of $s(i)$ with respect to the TA: in the PA, there was an increase between 42.11 and 244.44%; in the SAA, between 115.79 and 388.89%; and in the IA, between 263.16 and 566.67%. Figure 6 also shows that most concentrations of $s(i)$ values of the different clusters were in the most upper sides of the distribution of values, except in the PA. However, and despite the important improvement, only SAA and IA obtained $s(i)$ values greater than 0.5. As seen in Section 2, the $s(i)$ values closer to 1 show that the city has been placed in the correct group. Based on the analysis of average values of clusters, the percentage of groups with a $s(i)$ greater than 0.5 was between 14.29 and 21.43% in the SAA, and between 95.24 and 100% in the IA.

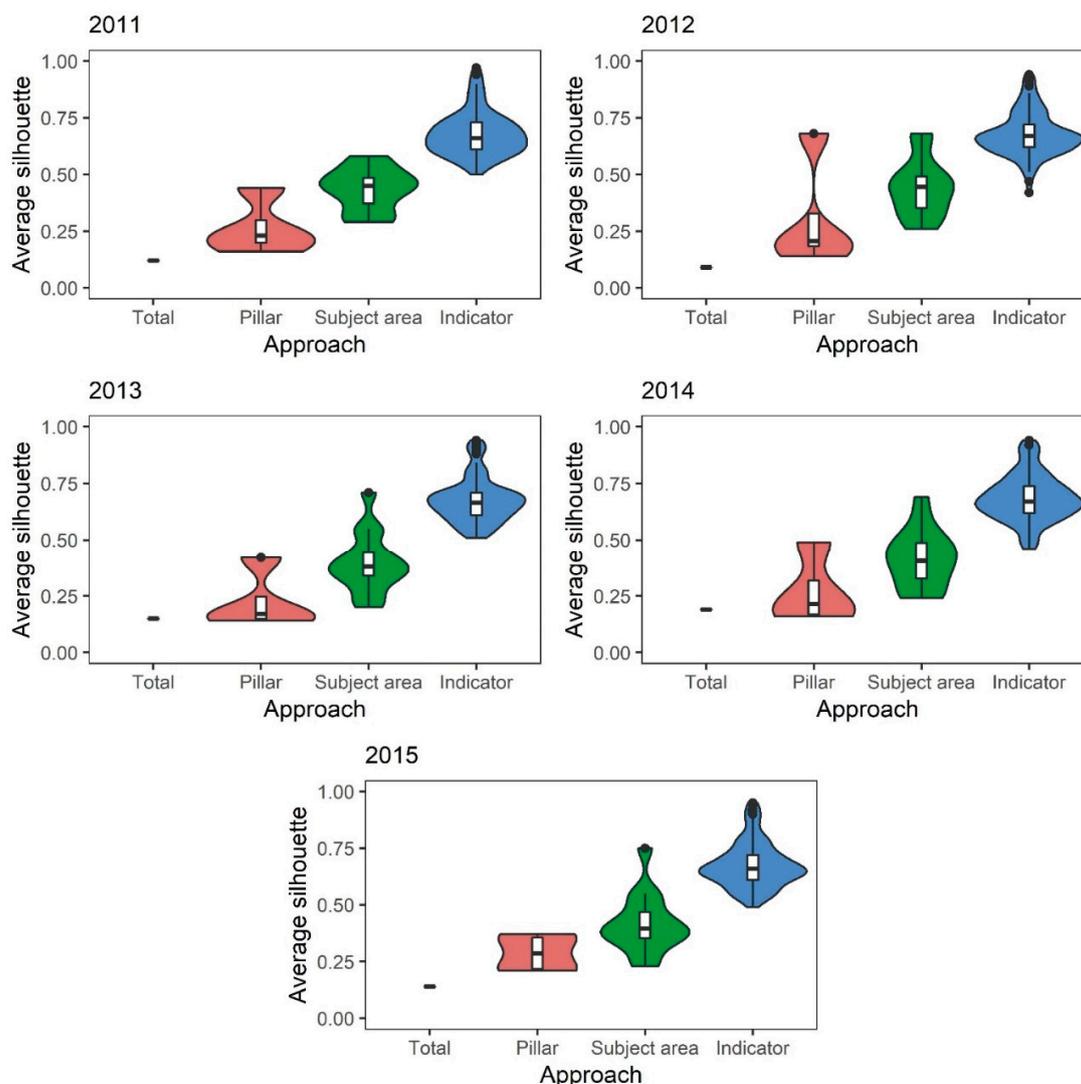


Figure 6. Violin- and box-plots with the average silhouette distributions obtained in the different cluster analysis approaches.

Table 4. Average value of $s(i)$ and deviation percentage of the approaches with less dimensions (PA, SAA, and IA) with respect to the TA. Positive values in the deviation percentage indicate an increase of the index, and negative values imply a reduction.

| Year | TA | | PA | | SAA | | IA | |
|------|--------|--------|---|--------|---|--------|---|--|
| | $s(i)$ | $s(i)$ | Percentage deviation with respect to the TA [%] | $s(i)$ | Percentage deviation with respect to the TA [%] | $s(i)$ | Percentage deviation with respect to the TA [%] | |
| 2011 | 0.12 | 0.25 | 108.33% | 0.44 | 266.67% | 0.68 | 466.67% | |
| 2012 | 0.09 | 0.31 | 244.44% | 0.44 | 388.89% | 0.60 | 566.67% | |
| 2013 | 0.15 | 0.23 | 53.33% | 0.40 | 166.67% | 0.68 | 353.33% | |
| 2014 | 0.19 | 0.27 | 42.11% | 0.41 | 115.79% | 0.69 | 263.16% | |
| 2015 | 0.14 | 0.29 | 107.14% | 0.42 | 200.00% | 0.68 | 385.71% | |

Due to these values obtained, the number of cities incorrectly classified in each sustainability indicator used in cluster analyses was analysed in detail. For this purpose, according to each approach, the centroid of each indicator was determined in the various groups (i.e., the correct classification of the cities in the 105 sustainability indicators used in the research was assessed in each approach). According to this centroid, it was assessed whether cities were grouped among cities with a similar similarity degree depending on each indicator. As a total of 105 indicators and 16 cities were used, the number of cases assessed per approach was 1680. Table 5 indicates the percentage of cases in which a city was grouped incorrectly. Similarly to $s(i)$, the reduction of dimensions of the cluster analysis reduced the number of cases incorrectly grouped. In this regard, in the TA, the percentage of cases incorrectly grouped ranged between 37.03% and 43%; in the PA, between 30.90% and 36.38%; in the SAA, between 25.06% and 33.36%; and in the IA, 0% was always the percentage. So, only the IA correctly grouped all cities in each indicator. This is very important when assessing the evolution tendencies of the relation between indicators-cities throughout the years, as each indicator have their own characteristics which are required to be assessed, and the use of cluster analysis approaches with dimensions greater than 1D could lead to erroneous groups. This aspect can be seen in the clusters of two indicators from the year 2011, which constitute an example included in Figure 7. The clusters obtained by each approach are represented in Figure 7. Both the TA and PA have the same clusters, as both indicators are of dimension Environmental (see Figure 1). Therefore, only clusters from SAA and IA are different between both indicators, thus showing the limitations of the TA by grouping cities with different values in their indicators.

Table 5. Percentage of cases incorrectly grouped in the cluster analysis using the various approaches of the study.

| Year | Percentage of Cases Incorrectly Grouped [%] | | | |
|------|---|-------|-------|------|
| | TA | PA | SAA | IA |
| 2011 | 43.00 | 33.04 | 28.01 | 0.00 |
| 2012 | 39.56 | 30.90 | 28.87 | 0.00 |
| 2013 | 37.21 | 36.38 | 33.63 | 0.00 |
| 2014 | 39.43 | 31.01 | 30.61 | 0.00 |
| 2015 | 37.03 | 31.19 | 25.06 | 0.00 |

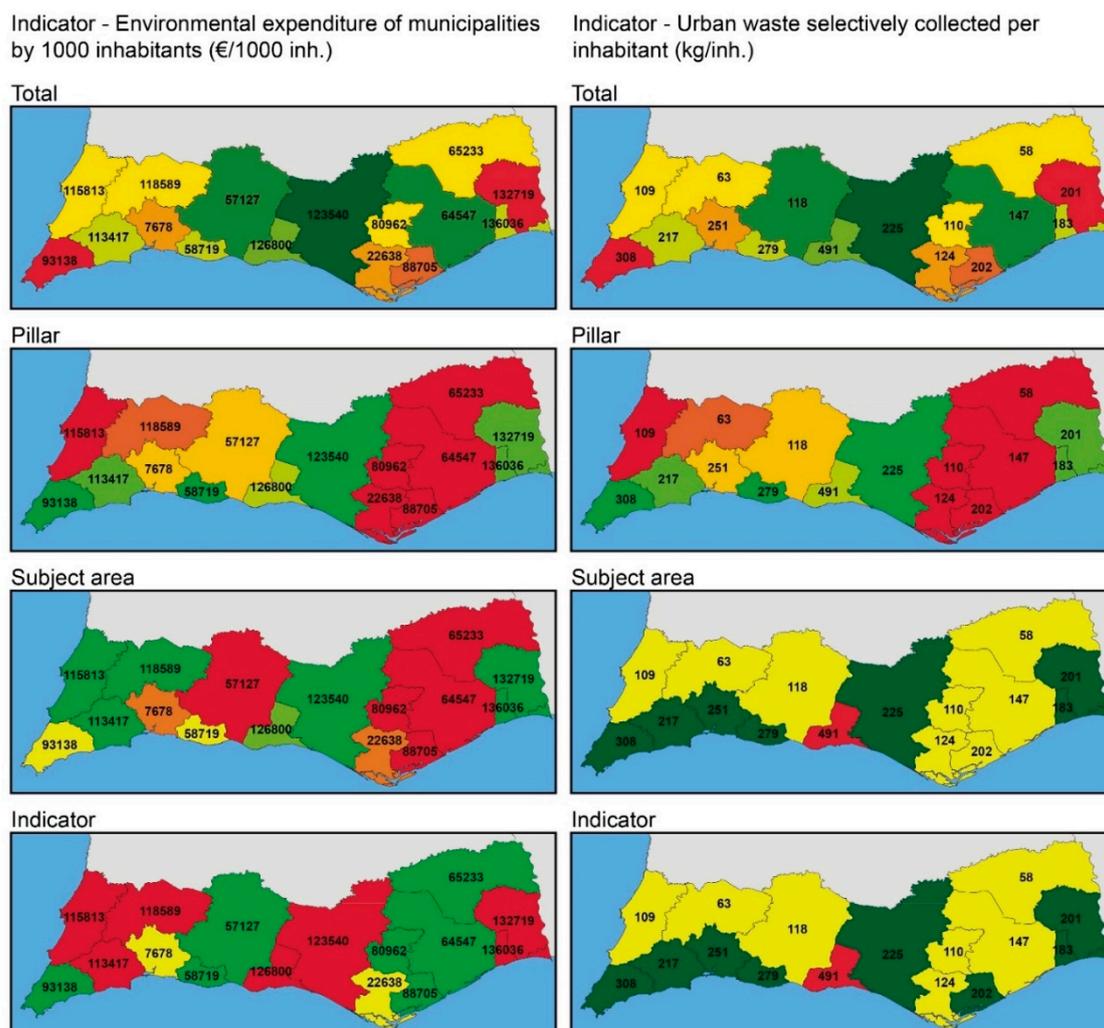


Figure 7. Example of the variations in the groups obtained with the different approaches in two environmental indicators in 2011. Each colour corresponds to a cluster. Colours among different squares are not related (i.e., the colours of the clusters from the PA are not related to that from the SAA).

In the TA, groups with cities with a greater degree of similarity with cities from other groups were obtained. In this regard, in the environmental expenditure indicator, Alcoutim was grouped with the cities of Aljezur and Monchique, whose expenditure difference is greater than 50,580 €/inhab, whereas other cities with a very low difference (e.g., Tavira) were grouped with other cities. This aspect can also be seen in the urban waste selectively collected per inhabitant, as Portimão was grouped with Faro, whereas cities with closer values, such as Loulé, Olhão and Lagoa, were placed in different groups. In the PA (with the same classification in both indicators), there were also erroneous classifications, such as Silves with Portimão in the environmental expenditure and in the urban waste selectively collected. Regarding the SAA, there was an almost correct classification, and the classification was erroneous only in some cases, such as Olhão in urban waste selectively collected and Lagoa in expenditure. Finally, the most appropriate classification was obtained with the 1D approach.

These results therefore show that the most adequate methodology to assess the similarity in the sustainability of cities is through the 1D cluster analysis of each indicator assessed, thus guaranteeing that the results obtained incorrectly group the cities and assess the variation tendencies that cities could present throughout the time.

4. Conclusions

In this paper, several cluster analyses were used to explore the similarities among cities of the Algarve region based on the monitoring of sustainability indicators. The cluster analysis algorithm used was k-means, and 4 different approaches were used to reduce the number of dimensions of the dataset. The results showed that the use of approaches including a high number of variables in the cluster analysis usually leads to incorrect groups in cities. In this regard, both the silhouette index and the ratio between the sum of squares and the total sum of squares showed that reducing the number of dimensions (i.e., the number of indicators) allowed more appropriate groups to be made, with the individual analysis of each indicator being the optimal case. This same aspect was reflected in the percentage of cases incorrectly grouped, in which only the Indicator approach guaranteed that the group of each indicator put the cities in the correct group, while the other approaches obtained group errors greater than 25%. Thus, the 1D cluster analysis was the best option for an adequate classification of the cities compared to the other approaches. In this sense, the following approach analysed with smaller dimensions (subject area approach) made incorrect groups of cities.

The results of the research therefore show the great influence of the dimensions considered in the cluster analysis. The results could be extrapolated to other regions where sustainability indicators are monitored and the similarity patterns among cities are intended to be assessed. In general terms, the use of an individual analysis approach of each indicator is the most appropriate option. However, this methodology could have limitations when the number of indicators is high. When these situations take place, considering a slightly high dimension (such as the subject area used in the study) would guarantee the obtaining of appropriate values of the silhouette index, although the percentage of cases incorrectly grouped could be high.

To conclude, the results of this research could be of great importance for public bodies and institutions responsible for the proposal of corrective measures with unsustainable behaviour patterns of cities. With the use of the cluster analysis, the zones of a region presenting a similarity in their behaviour could be found (e.g., the number of crimes recorded or the consumption of motor fuel by inhabitants), as well as to propose required performance measures.

Author Contributions: All the authors contributed equally to this work. All the authors participated in preparing the research from the beginning to end, such as establishing the research design, method, and analysis. All the authors discussed and finalized the analysis results to prepare the manuscript in accordance with the research progress. All the authors have read and approved the final manuscript.

Funding: This research was funded by Operational Program CRESC ALGARVE 2020 (ALG-01-0246-FEDER-027503).

Acknowledgments: The authors would like to acknowledge “Erasmus+ traineeship” for financing the international mobility of David Bienvenido-Huertas at the Universidade do Algarve.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Akande, A.; Cabral, P.; Gomes, P.; Casteleyn, S. The Lisbon ranking for smart sustainable cities in Europe. *Sustain. Cities Soc.* **2019**, *44*, 475–487. [[CrossRef](#)]
2. Asmelash, A.G.; Kumar, S. Assessing progress of tourism sustainability: Developing and validating sustainability indicators. *Tour. Manag.* **2019**, *71*, 67–83. [[CrossRef](#)]
3. Cernat, L.; Gourdon, J. Paths to success: Benchmarking cross-country sustainable tourism. *Tour. Manag.* **2012**, *33*, 1044–1056. [[CrossRef](#)]
4. Moussiopoulos, N.; Achillas, C.; Vlachokostas, C.; Spyridi, D.; Nikolaou, K. Environmental, social and economic information management for the evaluation of sustainability in urban areas: A system of indicators for Thessaloniki, Greece. *Cities* **2010**, *27*, 377–384. [[CrossRef](#)]
5. Tao, Y.; Li, F.; Crittenden, J.; Lu, Z.; Ou, W.; Song, Y. Measuring urban environmental sustainability performance in China: A multi-scale comparison among different cities, urban clusters, and geographic regions. *Cities* **2019**, *94*, 200–210. [[CrossRef](#)]

6. Recuero Virto, L. A preliminary assessment of the indicators for Sustainable Development Goal (SDG) 14 “Conserve and sustainably use the oceans, seas and marine resources for sustainable development”. *Mar. Policy* **2018**, *98*, 47–57. [[CrossRef](#)]
7. Borsato, E.; Tarolli, P.; Marinello, F. Sustainable patterns of main agricultural products combining different footprint parameters. *J. Clean. Prod.* **2018**, *179*, 357–367. [[CrossRef](#)]
8. Devereux, P.; Holmes, K. Voluntourism and the sustainable development goals. In *Collaborations for Sustainable Tourism Development*; Goodfellow Publishers: Oxford, UK, 2018; pp. 93–111.
9. Adshead, D.; Thacker, S.; Fuldauer, L.I.; Hall, J.W. Delivering on the Sustainable Development Goals through long-term infrastructure planning. *Glob. Environ. Chang.* **2019**, *59*, 101975. [[CrossRef](#)]
10. Higgins-Desbiolles, F. Sustainable tourism: Sustaining tourism or something more? *Tour. Manag. Perspect.* **2018**, *25*, 157–160. [[CrossRef](#)]
11. Lane, B. Will sustainable tourism research be sustainable in the future? An opinion piece. *Tour. Manag. Perspect.* **2018**, *25*, 161–164. [[CrossRef](#)]
12. Crouch, G.I. Destination competitiveness: An analysis of determinant attributes. *J. Travel Res.* **2011**, *50*, 27–45. [[CrossRef](#)]
13. Pulido-Fernández, J.I.; Cárdenas-García, P.J.; Espinosa-Pulido, J.A. Does environmental sustainability contribute to tourism growth? An analysis at the country level. *J. Clean. Prod.* **2019**, *213*, 309–319. [[CrossRef](#)]
14. Verma, P.; Raghubanshi, A.S. Urban sustainability indicators: Challenges and opportunities. *Ecol. Indic.* **2018**, *93*, 282–291. [[CrossRef](#)]
15. Hermans, F.L.P.; Haarmann, W.M.F.; Dagevos, J.F. Evaluation of stakeholder participation in monitoring regional sustainable development. *Reg. Environ. Chang.* **2011**, *11*, 805–815. [[CrossRef](#)]
16. Manning, T. *What Tourism Managers Need to Know: A Practical Guide to the Development and Use of Indicators of Sustainable Tourism*; World Tourism Organization Pubns: Madrid, Spain, 1996.
17. Michael, F.L.; Noor, Z.Z.; Figueroa, M.J. Review of urban sustainability indicators assessment—Case study between Asian countries. *Habitat Int.* **2014**, *44*, 491–500. [[CrossRef](#)]
18. Kristjánisdóttir, K.R.; Ólafsdóttir, R.; Ragnarsdóttir, K.V. Reviewing integrated sustainability indicators for tourism. *J. Sustain. Tour.* **2018**, *26*, 583–599. [[CrossRef](#)]
19. Miller, G. The development of indicators for sustainable tourism: Results of a Delphi survey of tourism researchers. *Tour. Manag.* **2001**, *22*, 351–362. [[CrossRef](#)]
20. Liu, C.R.; Lin, W.R.; Wang, Y.C.; Chen, S.P. Sustainability indicators for festival tourism: A multi-stakeholder perspective. *J. Qual. Assur. Hosp. Tour.* **2019**, *20*, 296–316. [[CrossRef](#)]
21. Blancas, F.J.; González, M.; Lozano-Oyola, M.; Pérez, F. The assessment of sustainable tourism: Application to Spanish coastal destinations. *Ecol. Indic.* **2010**, *10*, 484–492. [[CrossRef](#)]
22. Blancas, F.J.; Lozano-Oyola, M.; González, M. A European Sustainable Tourism Labels proposal using a composite indicator. *Environ. Impact Assess. Rev.* **2015**, *54*, 39–54. [[CrossRef](#)]
23. Nesticò, A.; Maselli, G. Sustainability indicators for the economic evaluation of tourism investments on islands. *J. Clean. Prod.* **2020**, *248*. [[CrossRef](#)]
24. Castellani, V.; Sala, S. Sustainable performance index for tourism policy development. *Tour. Manag.* **2010**, *31*, 871–880. [[CrossRef](#)]
25. Eustachio, J.H.P.P.; Caldana, A.C.F.; Liboni, L.B.; Martinelli, D.P. Systemic indicator of sustainable development: Proposal and application of a framework. *J. Clean. Prod.* **2019**, *241*. [[CrossRef](#)]
26. Hély, V.; Antoni, J.P. Combining indicators for decision making in planning issues: A theoretical approach to perform sustainability assessment. *Sustain. Cities Soc.* **2019**, *44*, 844–854. [[CrossRef](#)]
27. Palmisano, G.O.; Govindan, K.; Boggia, A.; Loisi, R.V.; De Boni, A.; Roma, R. Local Action Groups and Rural Sustainable Development. A spatial multiple criteria approach for efficient territorial planning. *Land Use Policy* **2016**, *59*, 12–26. [[CrossRef](#)]
28. Huang, L.; Zheng, W.; Hong, J.; Liu, Y.; Liu, G. Paths and strategies for sustainable urban renewal at the neighbourhood level: A framework for decision-making. *Sustain. Cities Soc.* **2020**. [[CrossRef](#)]
29. Ma, J.; Harstvedt, J.D.; Jaradat, R.; Smith, B. Sustainability driven multi-criteria project portfolio selection under uncertain decision-making environment. *Comput. Ind. Eng.* **2020**, *140*, 106236. [[CrossRef](#)]
30. Yi, P.; Dong, Q.; Li, W. Evaluation of city sustainability using the deviation maximization method. *Sustain. Cities Soc.* **2019**, *50*, 101529. [[CrossRef](#)]

31. Dang, X.; Zhang, Y.; Feng, W.; Zhou, N.; Wang, Y.; Meng, C.; Ginsberg, M. Comparative study of city-level sustainability assessment standards in China and the United States. *J. Clean. Prod.* **2020**, *251*, 119622. [[CrossRef](#)]
32. Neri, L.; D'Agostino, A.; Regoli, A.; Pulselli, F.M.; Coscieme, L. Evaluating dynamics of national economies through cluster analysis within the input-state-output sustainability framework. *Ecol. Indic.* **2017**, *72*, 77–90. [[CrossRef](#)]
33. Budeanu, A.; Miller, G.; Moscardo, G.; Ooi, C.S. Sustainable tourism, progress, challenges and opportunities: An introduction. *J. Clean. Prod.* **2016**, *111*, 285–294. [[CrossRef](#)]
34. United Nations General Assembly. *Resolution Adopted by the General Assembly on 25 September 2015: Transforming Our World: The 2030 Agenda for Sustainable Development*; United Nations General Assembly: New York, NY, USA, 2015.
35. Do Valle, P.O.; Pintassilgo, P.; Matias, A.; André, F. Tourist attitudes towards an accommodation tax earmarked for environmental protection: A survey in the Algarve. *Tour. Manag.* **2012**, *33*, 1408–1416. [[CrossRef](#)]
36. Estatística, I.N. De Estatísticas. Available online: <https://ine.pt> (accessed on 2 September 2019).
37. Araújo, L. Portuguese tourism strategy 2027 Leading the tourism of the future. *Worldw. Hosp. Tour. Themes* **2017**, *9*, 646–652. [[CrossRef](#)]
38. Farinha, F.; Oliveira, M.J.; Silva, E.M.J.; Lança, R.; Pinheiro, M.D.; Miguel, C. Selection process of sustainable indicators for the Algarve region-OBSERVE project. *Sustainability* **2019**, *11*, 444. [[CrossRef](#)]
39. Andraz, J.M.; Rodrigues, P.M.M. Monitoring tourism flows and destination management: Empirical evidence for Portugal. *Tour. Manag.* **2016**, *56*, 1–7. [[CrossRef](#)]
40. Correia, A.; Kozak, M. Tourists' shopping experiences at street markets: Cross-country research. *Tour. Manag.* **2016**, *56*, 85–95. [[CrossRef](#)]
41. Costa, C. Turismo e cultura: Avaliação das teorias e práticas culturais do sector do turismo (1990–2000). *Análise Soc.* **2005**, *175*, 279–295.
42. Instituto Nacional de Estatística. *Estatísticas do Turismo 2015*; Instituto Nacional de Estatística Lisboa: Lisboa, Portugal, 2016.
43. Barreira, A.P.; Cesário, M.; de Noronha, M.T. Pull attributes of the Algarve: The tourists' view. *Tour. Plan. Dev.* **2017**, *14*, 87–109. [[CrossRef](#)]
44. Oliveira, R.; Pedro, M.I.; Marques, R.C. Efficiency and its determinants in Portuguese hotels in the Algarve. *Tour. Manag.* **2013**, *36*, 641–649. [[CrossRef](#)]
45. Oliveira, R.; Pedro, M.I.; Marques, R.C. Efficiency performance of the Algarve hotels using a revenue function. *Int. J. Hosp. Manag.* **2013**, *35*, 59–67. [[CrossRef](#)]
46. Barros, C.P.; Butler, R.; Correia, A. The length of stay of golf tourism: A survival analysis. *Tour. Manag.* **2010**, *31*, 13–21. [[CrossRef](#)]
47. Pereira, R.L.G.; Correia, A.H.; Schutz, R.L.A. Towards a taxonomy of a golf-destination brand personality: Insights from the Algarve golf industry. *J. Destin. Mark. Manag.* **2015**, *4*, 57–67. [[CrossRef](#)]
48. Lopes, I.C.; Soares, F.; Silva, E.C. e Tourism demand in the Algarve region: Evolution and forecast using SVARMA models. *AIP Conf. Proc.* **2017**, *1836*, 20075. [[CrossRef](#)]
49. Oliveira, M.J.; Farinha, F.; da Silva, E.; Lança, R. Observatory of Sustainability of the Algarve Region for Tourism -Overview and outset. In Proceedings of the 2nd UNWTO World Conferences on Smart Destinations, 25–27 June 2018, Oviedo, Spain; Volume 13, pp. 1237–1244.
50. Hartigan, J.A.; Wong, M.A. Algorithm AS 136: A k-means clustering algorithm. *J. R. Stat. Soc. Ser. C* **1979**, *28*, 100–108. [[CrossRef](#)]
51. Ketchen, D.J.; Shook, C.L. The application of cluster analysis in strategic management research: An analysis and critique. *Strateg. Manag. J.* **1996**, *17*, 441–458. [[CrossRef](#)]
52. Kaufman, L.; Rousseeuw, P.J. *An Introduction to Cluster Analysis*; John Wiley and Sons: Hoboken, NJ, USA, 1990.
53. Hintze, J.L.; Nelson, R.D. Violin Plots: A Box Plot-Density Trace Synergism. *Am. Stat.* **1998**, *52*, 181–184. [[CrossRef](#)]

