STATISTICAL ANALYSIS OF THE EU COUNTRIES FROM A TOURISTIC POINT OF VIEW

T. DEMETER[1]   G. BRĂTUCU[1]

Abstract: *The purpose of this paper is to study the EU countries by taking into account major characteristics that effect tourism, such as the number of tourists that visited each country, the total number of nights that were spent in the local accommodation units and the total number of accommodation units available. These characteristics will be analysed both individually, using the graphical method, and collectively, using the cluster analysis, having as main objective the development of homogeneous groups that contain these countries. The results show that the optimal solution is to create 3 clusters, allowing a suitable differentiation between them, while keeping the homogeneity among the countries that form it.*

Key words: *Tourism, Cluster analysis, Graphical analysis, Euclidian distances, EU countries*

1. Introduction

Tourism has made a significant contribution to the economies of many communities around the world because of its ability to create income, taxes, hard currency and jobs [4], [7].

According to a tourism report [1], four European Union (EU) member countries, Austria, Germany, France and Spain, are ranked among the top countries that continue to lead the way in travel and tourism.[12] To them a number of countries in Central and Eastern Europe (CEE) can be added, which have emerged from an environment of one-party communist governments and centralised planning, now following models closer to the Western European ones. [11] These CEE countries now set out to attract tourists from the main generating countries.

However, they are not a uniform entity, but are diverse in terms of location, topography, climate, history, culture and economic development, each willing to identify its own comparative advantage with respect to tourism [8]

The aim of this paper is to study the EU countries, both developed and less developed, by taking into account major characteristics that effect tourism. These characteristics will be analysed both individually and simultaneously, having as main objective the development of a few homogeneous groups, based on the similarities existing between the countries.

This will be carried out using the cluster analysis methods, which represent the most frequently used unsupervised market segmentation technique in the literature. [6], Cluster analysis consists of grouping

---

[1] Dept. of Marketing, Tourism and International Relations, *Transilvania* University of Braşov.

similar objects according to their degree of similarity. Objects within each cluster are more closely related to one another than objects assigned to different clusters and each cluster can be distinguished from the others [3]. This method comprises a set of different techniques, which can be broadly divided into partitioning and hierarchical methods that have also been used in tourism research. For example: to identify strategic groups of UK hotels [5]

Out of the above mentioned types, the most common one is hierarchical clustering, [16] a multivariate approach that encompasses many separation/agglomeration techniques which aim to identify natural groupings, amongst objects in a dataset, through minimisation of the within-cluster variance and maximisation of the between-cluster variance [13]. Therefore, each case starts in a separate cluster and joins up to the other clusters as the linkage distance grows, and only one cluster remains in the end. [14]

This technique presents two main advantages with respect to more traditional clustering techniques [15]: it is not necessary to impose the number of clusters in advance; the final solution is less dependent on the initialization of the algorithm. [6]

The aim of this paper is to analyse the countries from the European Union from the touristic point of view, taking into consideration the similarities between them.

2. Methodology

In order to perform the analysis of the countries, 7 criteria that are relevant from the touristic point of view were taken into consideration. Namely, the number of tourists for 2012, the total number of nights that were spent in the local accommodation units, the total number of accommodation units in the country, the rate of tourism growth, the number of airports, the length of the roads, and the surface of protected areas that represent a tourist attraction.

All the data were collected from the website of the European Statistical Institute and are valid for the year 2012.

The data were analysed using graphical methods and hierarchical cluster analysis in SPSS.

The hierarchical method was chosen due to the fact that the volume of data is not very big, so there was no need for the K-means analysis. The cluster was performed for classes using the Between-groups linkage method, while the measure between the intervals was calculated with the Euclidean distance.

3. Results and interpretation

The analysis was divided into two parts, first using the graphical method which studies the elements that influence tourism individually and the multivariate analysis, which takes into account all the factors at the same time, in order to group the EU countries based on their similar characteristics, forming homogenous clusters.

3.1. Graphical analysis

When analysing tourism, it is of great importance to know the percentage that each country holds on the entire European market; this is illustrated in Fig 3.1.1

According to the graph, the country with the highest number of tourists is Germany ,with a percentage of 19 from the total, followed by France with 17% and Italy with 14%. The countries that have the most developed tourism are clearly visible, and Great Britain together with Austria are also added to the list.

In order to establish the Paretto diagram, the population was divided into 6 intervals using the Sturge's method with the value of M=5.5
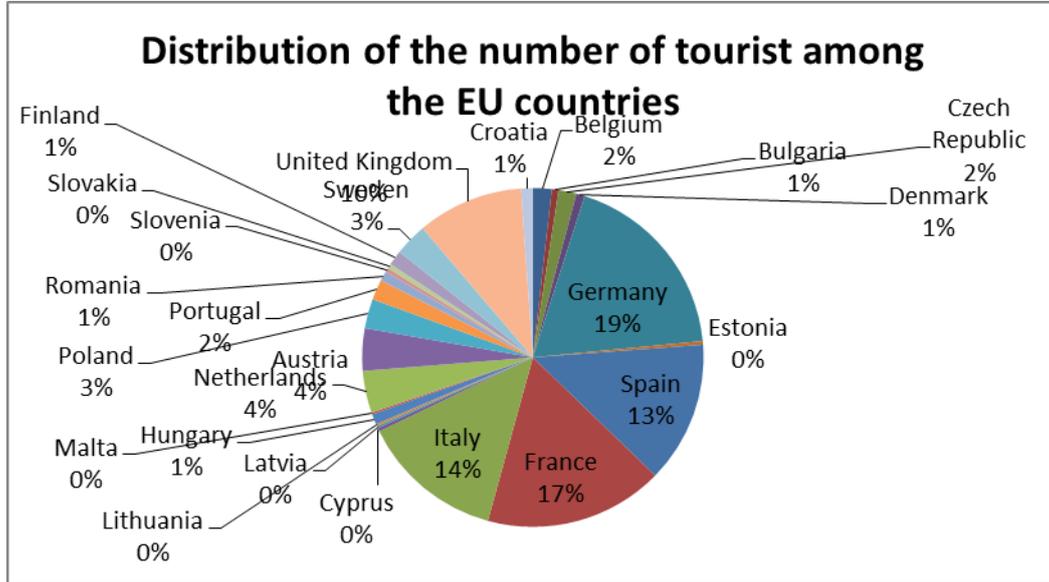
Fig. 1. *Distribution of the number of tourist*

According to the Histogram, it is notable that most of the countries have an average of 29.4 million tourists.
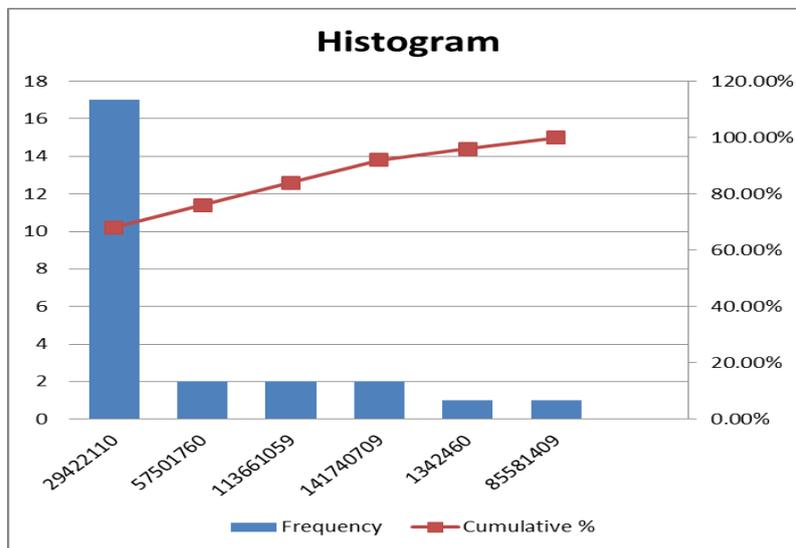
Fig. 2. *Distribution of tourists*

Analysing the number of nights spent in the local accommodation units, Graph 3.1.2 shows that Germany is the first one, with the highest rate, and is followed by Spain, which has a smaller share when it comes to the number of tourists, meaning that the tourists visiting Spain spend more nights there than the ones visiting other countries except for Germany.

Although Italy is in the third place taking into consideration the number of tourists/year, and fourth taking into consideration the number of nights spent there, it has the greatest number of accommodation units from Europe, having a share of 35% of the total.
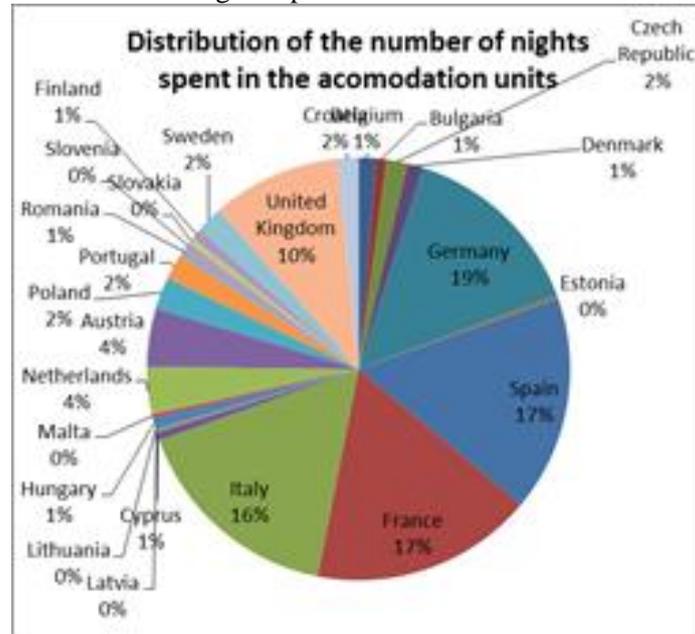
Fig. 3. *Distribution of nights spent in the accommodation units*

Germany maintains the same percentage it has had until now, having the same value as Great Britain 19%. France has only 6% from the total of the accommodation units in the European Union; however it has a high percentage considering the number of tourists and the number of nights spent there.
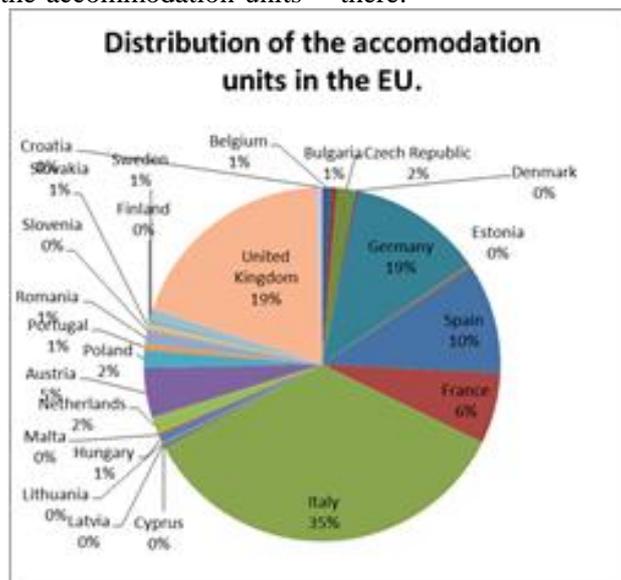
Fig. 4. *Distribution of accommodation*

The distribution of airports is best illustrated using the Pareto diagram, showing that most of the countries, more exactly 17, have a number of 16 airports.

This is essential information, knowing that the tourist flow in directly influenced by the transportation methods available in that particular country.
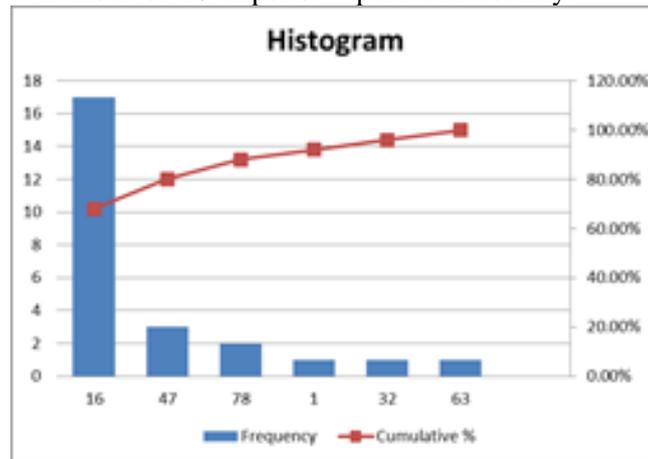

Fig. 5. *Distribution of airports*

3.2. Cluster analysis

The graphical analysis offers important information about the tourism in the European Union; however, the analysis of all the criteria simultaneously offers a greater relevance for this topic, and compares the countries between each other based on them. This purpose is best served using the cluster analysis, obtaining the following tables.

Case summary   Table 1

Cases						
Valid		Missing		Total	
N	Percent	N	Percent	N	Percent
25	100.0%	0	.0%	25	100.0%

a. Euclidean Distance used

Knowing that the data were collected based on previous research, all of them are valid and there is no value missing.

The *Proximity Matrix* comprises the Euclidean distances between each case. The diagonal represents the distance of each case from itself, therefore having the value of 0. It is also clear that it is a symmetrical Matrix having the same values under the diagonal as above it.

The lowest value 0.223 is for the pair Latvia − Lithuania, as they are the two countries that present most similarities, while the most different are Lithuania and France with the Euclidian distance of 7.221.

The *Agglomeration Schedule* shows how each case was grouped during the groping phases. The second column, Cluster Combined, contains the grouped cases; the Coefficients contain the distance coefficients of the grouped cases; the Stage cluster shows the stages in which the

elements appeared previously; while the Next stage shows the stage in which the case will later appear again, and the phase in which it will be modified.

down within square brackets, in the head of

Agglomeration Schedule   Table 2

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	11	12	.223	0	0	7
2	2	6	.273	0	0	7
3	1	4	.298	0	0	8
4	10	20	.333	0	0	11
5	3	18	.346	0	0	8
6	13	21	.348	0	0	9
7	2	11	.405	2	1	13
8	1	3	.447	3	5	9
9	1	13	.519	8	6	12
10	15	16	.537	0	0	16
11	10	25	.559	4	0	12
12	1	10	.608	9	11	14
13	2	19	.801	7	0	21
14	1	14	.803	12	0	16
15	22	23	.910	0	0	17
16	1	15	1.062	14	10	20
17	17	22	1.091	0	15	20
18	5	8	1.545	0	0	19
19	5	7	2.107	18	0	23
20	1	17	2.161	16	17	21
21	1	2	2.509	20	13	24
22	9	24	2.849	0	0	23
23	5	9	3.967	19	22	24
24	1	5	5.726	21	23	0

In the first phase, Latvia and Lithuania are grouped, as established before, based on the Proximity Matrix, the distance between them being the smallest. It can be also noted that none of them appeared in previous stages and the next stage in which Latvia will appear again will be the seventh, and it will be modified during this stage. These chainges can also be observed in the Stage Cluster First Appears column, under the Cluster 2 title where there is a reference to stage 1 for case 11. The whole clustering process contains a total of 24 stages. And a total of 10 clusters containing 2 countries, which are later joined with other clusters.

Therefore, the countries having most similarities when it comes to grouping into 2 are: Latvia-Lithuania, Bulgaria-Estonia, Belgium-Denmark, Cyprus-Slovenia, Czech Republic-Portugal and Hungary-Slovakia.

The next table, Cluster membership, gives more details about the clusters that can be formed.

If all the studied countries were grouped into just *2 clusters*, they would be the following:

Cluster Membership   Table 3

Case	4 Clusters	3 Clusters	2 Clusters
1:Belgium	1	1	1
2:Bulgaria	1	1	1
3:Czech Re	1	1	1
4:Denmark	1	1	1
5:Germany	2	2	2
6:Estonia	1	1	1
7:Spain	2	2	2
8:France	2	2	2
9:Italy	3	3	2
10:Cyprus	1	1	1
11:Latvia	1	1	1
12:Lithuani	1	1	1
13:Hungary	1	1	1
14:Malta	1	1	1
15:Netherla	1	1	1
16:Austria	1	1	1
17:Poland	1	1	1
18:Portugal	1	1	1
19:Romania	1	1	1
20:Slovenia	1	1	1
21:Slovakia	1	1	1
22:Finland	1	1	1
23:Sweden	1	1	1
24:United K	4	3	2
25:Croatia	1	1	1

First cluster: Belgium, Bulgaria, the Czech Republic, Denmark, Estonia, Cyprus, Latvia, Lithuania, Hungary, Malta, Netherlands, Austria, Poland, Portugal, Romania, Slovenia, Slovakia, Finland, Sweden, Croatia.

The second cluster: Germany, Spain, France, Great Britain, and Italy. It is clear that this cluster contains the most developed countries from the European Union, from the touristic point of view.

For a more accurate classification, *3 clusters* can be used, grouping the countries in the following way:

First cluster: Belgium, Bulgaria, Czech Republic, Denmark, Estonia, Cyprus, Latvia, Lithuania, Hungary, Malta, Netherlands, Austria, Poland, Portugal, Romania, Slovenia, Slovakia, Finland, Sweden, Croatia.

The second cluster: Germany, Spain and France.

The third cluster: Great Britain, and Italy.

When grouping into *4 clusters,* the elements that are further divided are the developed countries, obtaining this way:

First cluster: Belgium, Bulgaria, Czech Republic, Denmark, Estonia, Cyprus, Latvia, Lithuania, Hungary, Malta, Netherlands, Austria, Poland, Portugal,

Romania, Slovenia, Slovakia, Finland, Sweden, Croatia.

The second cluster: Germany, Spain and France.

The third cluster: Italy

The fourth cluster: Great Britain.

For the illustration of the clusters the Dendogram was used.

It is very visible that the 5 most developed countries (Germany, France, Spain, Italy and Great Britain) group together separately from the others until the last stage, where they are joined together, underlining again the different characteristics they have in comparison to the other 20.
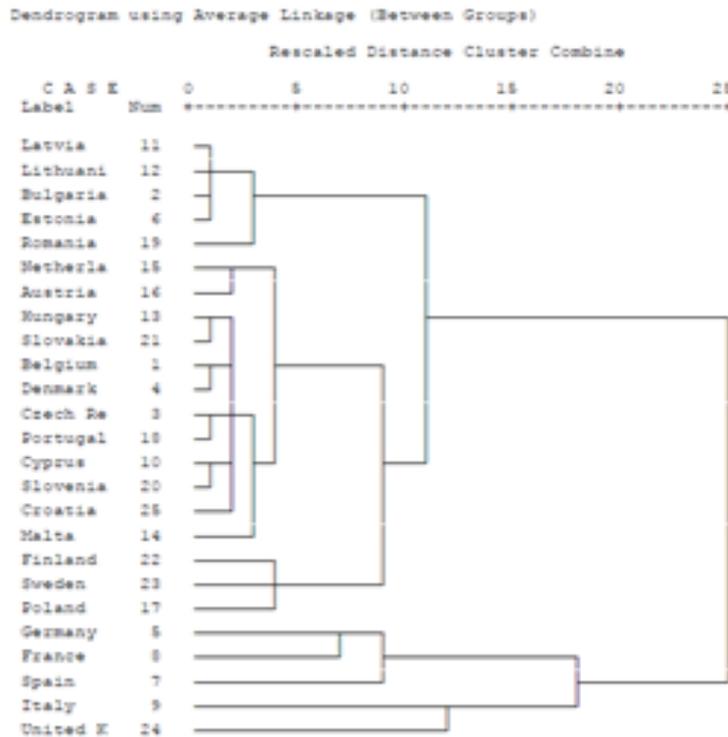

Fig. 6.

4. Conclusion

From the graphical analysis, it can be concluded that the most visited country in the European Union is Germany, having the greatest number of tourists/year, and the most nights spent in the accommodation units, although it does not possess the greatest number of accommodation units from the Union. Spain however ranks 3rd when it comes to the number of tourist, but occupies the 2nd place in the number of nights spent there.

The number of accommodation units is not in a strong relation with the number of tourists that visit a country, and the number of nights spent, considering that Italy has far more accommodation units than necessary for the number of tourists that visit it, or the number of nights that tourists spend in their accommodation units. at the other end of the spectrum is France, which has a relatively small number of accommodation units, but a good level of tourist flow and a high number of nights purchased.

The Cluster analysis offers a more complex perspective, and taking into consideration the aim of the study, it is most appropriate

to group the countries in 3 clusters having the following structure: 1. Belgium, Bulgaria, the Czech Republic, Denmark, Estonia, Cyprus, Latvia, Lithuania, Hungary, Malta, the Netherlands, Austria, Poland, Portugal, Romania, Slovenia, Slovakia, Finland, Sweden, Croatia; 2. Germany, Spain and France; 3. Great Britain and Italy.

Therefore, the similarity between these countries can be easily observed, and the study of the clusters can give a thorough view on their characteristics.

This study represents the starting point for another study that takes into account the strategies and policies that these countries, within the clusters, use in order to establish the ones that are applicable for countries that seek a better tourist development.

Other information may be obtained from the address: timea.demeter@unitbv.com

References

1. Blanke, J., Chiesa, T., Herrera, E. T: *The travel and tourism competitiveness index 2009: measuring sectoral drivers in a downturn*. In J. Blanke, T. Chiesa (Eds.), The travel and tourism competitive report 2009 Geneva, Switzerland: World Economic Forum p. 3-38.
2. Brida, J., G, *et. al: Cruise passengers' experience embarking in a Caribbean home port. The case study of Cartagena de Indias* In: Ocean & Coastal Management 55, 2012, p. 135-145.
3. Brida, J.G., Zapata, S.: *Cruise tourism: economic, socio-cultural and environmental impacts.* In International Journal of Leisure and Tourism Marketing 1 (3), 2010, *p.* 205-226.
4. Choi, H.-C., Sirakaya, E: *Sustainability indicators for managing community tourism*. In: *Tourism Management*, 27(6), 2006, p. 1274-1289.
5. Curry,B.,Davies,F.,Phillips,P.,Evans, M.,Moutinho,L.: *The Kohonen self-organizing map: Anapplication to the study of strategic groups in the UK hotel industry.*In Expert Systems, 18, 2001, p. 19–31.
6. D'Urso, P., DeGiovanni, L., Disegna, M., Massari, R.: *Bagged Clustering and its application to tourism market segmentation* In: Expert Systems with Applications 40, 2013, p. 4944–4956
7. Dwyer, L., Forsyth, P.: *Economic measures of tourism yield: what markets to target?* In: International Journal of Tourism Research, 10(2), 2008, p. 155-168.
8. Hall, D: *Central and Eastern Europe: Tourism, development and transformation.* In A. Williams, G. Shaw (Eds.), Tourism and economic development: European experiences 3rd ed. London: Wiley, 1998, p. 345–373.
9. http://epp.eurostat.ec.europa.eu Accessed 04.02.2013
10. Hughes, H., Allen, D.: *Cultural tourism in Central and Eastern Europe: the views of 'induced image formation agents* In Tourism Management 26, 2005, p. 173–183
11. Jansen-Verbeke, M.: *Cultural tourism in the twenty-first century.* In: *World Leisure and Recreation*, 31(1),1996,p. 6–11.
12. Jung,W.,L., Tantatape, B: *Investigating the influence of tourism on economic growth and carbon emissions: Evidence from panel analysis of the European Union* In: Tourism Management 38, 2013, p.69-76
13. Kaufman, L., Rousseeuw, P. J.: *Finding Groups in Data: An Introduction to Cluster Analysis.* Wiley, New York. 1990

14. Lattin, J.M., Carroll, J.D., Green, P.E.: *Analyzing Multivariate Data*. Duxbury

15. Leisch, F.*: Bagged clustering*. In *Working paper 51SFB adaptive information systems and modelling ineconomics and management science* WUVienna University of Economics and Business, 1999

16. Magyar, N. et. al.,: *Application of multivariate statistical methods in determining spatial changes in water quality in the Austrian part of Neusiedler See* In: Ecological Engineering 55, 2013, p. 82– 92 Press, New York, 2003.