

A statistical assessment of tourism development disparities at the district level: The case of Serbia

Milan Stamenković^{1*}, Marina Milanović¹

¹ University of Kragujevac, Faculty of Economics, Kragujevac, Serbia

Abstract: Tourism is a sector of immense importance and its advancement plays a crucial role in improving national and promoting balanced regional socio-economic development. This study presents a complex multivariate methodological approach for categorization of 25 districts in Serbia into internally-more similar and externally-more dissimilar clusters by implementing hierarchical agglomerative clustering procedure and analysis of present interdependencies between selected indicators of tourism demand. The statistical validity and quality of the extracted optimal clustering structure are evaluated and confirmed based on the adequate optimality criteria and corresponding results of the non-hierarchical clustering procedure. The proposed categorization of districts clearly and unambiguously confirms the presence of significant tourism development asymmetries between NUTS 3 territories in Serbia, and the existence of intra-regional tourism activity polarization with the developed east and south-west (including the city of Belgrade) on the one end of the spectrum and the less developed north and central areas of Serbia.

Keywords: cluster analysis, tourism development, regional disparities, Serbia

JEL classification: C38, L83, R11

Statistička ocena turističkih razvojnih dispariteta na nivou okruga: Studija slučaja Srbija

Sažetak: S obzirom na značajnu ulogu sektora turizma i njegovog razvoja u promovisanju nacionalnog i podsticanju ravnomernog regionalnog socio-ekonomskog razvoja, u ovom radu, kompleksan multivarijacioni metodološki pristup za klasifikaciju 25 okruga u Srbiji u interno-homogene / eksterno-heterogene klasterne, zasnovan primarno na hijerarhijskoj proceduri grupisanja i ispitivanju prisutnih međuzavisnosti između odabranih indikatora turističke aktivnosti / tražnje je primenjen i prezentovan. Statistička validnost i kvalitet izdvojene optimalne klasifikacione strukture dodatno su evaluirani i potvrđeni na osnovu vrednosti adekvatnih kriterijuma optimalnosti i rezultata nehijerarhijske procedure grupisanja. Predložena kategorizacija okruga jasno i nedvosmisleno potvrđuje prisustvo izraženih nejednakosti u razvijenosti sektora turizma između NUTS 3 teritorijalnih jedinica u Srbiji i postojanje unutar-regionalne polarizacije turističke aktivnosti / tražnje, primarno u pravcu: razvijeni istočni / jugozapadni deo, sa gradom Beogradom – manje razvijeni severni / centralni deo Srbije.

Cljučne reči: klaster analiza, razvijenost sektora turizma, regionalni dispariteti, Srbija

JEL klasifikacija: C38, L83, R11

* m.stamenkovic@kg.ac.rs



This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In a global context, tourism, as “an area of the economy, which includes and connects various industries, branches and activities aimed at providing services that enable tourists to meet their needs” (Petrović et al., 2020, p. 167), represents “the largest service industry in the world” (Roman et al., 2020, p. 1) and “one of the fastest growing and most profitable, constantly developing, sector of the economy” (Morozova et al., 2016, p. 2). According to the World Travel and Tourism Council (WTTC, 2022), prior to the pandemic, the travel and tourism sector accounted for 10.3% of global gross domestic product (GDP), i.e. USD 9.2 trillion, with a growth rate of 3.5%, and 333 million jobs, or 10.4% of total employment in the world in 2019. With such global economic statistics, it is not surprising that the following metaphors are often used by scholars when describing this sector of economy: “an accelerator of the socio-economic development” (Gabbrakmanov et al., 2016, p. 5291), “the largest generator of wealth and employment in the world, the economic engine for developed and developing economies around the world” (Rita, 2000, p. 434), “a driving force for economic development” (Petrović et al., 2020, p. 169), etc. However, the effects of the COVID-19 crisis and significant restrictions on tourism mobility emphasized “the tremendous importance and positive contribution of tourism industry, causing the decline of its global contribution to GDP by 50.4% in 2020, and decrease of employment by 18.6% across this sector globally” (WTTC, 2022, p. 2). Confirming its resilience and ability to bounce back, the tourism sector began its recovery in 2021, although slower than expected, but with a positive future outlook, increasing its share in global GDP from 5.3% in 2020 to 6.1% in 2021, and “gaining 18.2 million jobs, representing an increase of 6.7%, compared to previous, 2020” (WTTC, 2022, p. 2).

Viewed from the perspective of national economies, tourism unequivocally represents a significant economic activity, a specific, well-paid export product (Gajić et al., 2014), which “contributes, to a higher or lower degree, to the country’s overall economic development” (Roman et al., 2022, p. 1). According to the modern concept key postulates called tourism-driven development, the tourism sector plays an important role in solving economic and social problems in a country. Actually, by providing new employment opportunities, additional tax revenues and foreign exchange reserves for the governments, tourism development directly contributes to the GDP increase and residents welfare in a host country. Additionally, its indirect contribution to the national economy development is reflected in stimulating many tourism-related economic activities, such as agriculture, gastronomy, transport, trade, construction, etc. (Gabbrakmanov et al., 2016; Petrović et al., 2020; Roman et al., 2022). Owing to these positive multiplier effects, the tourism industry is one of the essential tools for the “revival of many economic and non-economic activities, the development of underdeveloped areas” (Gajić et al., 2014, p. 113), and thus achieving sustainable economic growth in most countries (Petrović et al., 2020).

Finally, when it comes to the position of a regional economy within the country, the tourism sector and its development must be considered from a slightly different, more specific angle. Namely, the regional potential for development of tourism, the dominant type of tourism, and intensity of tourist traffic largely depend on nature-given factors (i.e. geographical location and climatic features of the regions, spatial distribution of natural resources and attractions, etc.) as well as human-created conditions (i.e. accessibility, development of road infrastructure and sports–recreational facilities, development of supporting services sector, etc.) (Bećirović et al., 2011; Gorina et al., 2020). Inequality in the regional distribution of these factors inevitably leads to tourism development disparities among regions. Given the important role of “tourism as a catalyst in national and sustainable regional socio-economic development” (Gall, 2019, p. 452), it is not surprising that tourism sector is increasingly being regarded as “a savior of the countryside, with many governments recognizing its potentials in fostering regional economic

development” (Jackson & Murphy, 2006, p. 1018). Accordingly, the quantitative assessment of the development level of tourism in individual regions represents the first analytical step, which is necessary in order to create suitable conditions for balanced regional economic development through the reduction of the existing disparities. It provides useful and reliable informational input for formulating programs and strategies to foster regional tourism development and thus achieve balanced regional and overall economic development.

Consequently, this study presents the analysis of tourism development in Serbia at regional level (i.e. at the level of districts, or territorial units of NUTS 3 level), and regards it as a specific (multi-dimensional) tourism-economic phenomenon. The following research objectives were formulated: (1) presentation and popularization of implementation possibilities of cluster analysis methods in the field of regional tourism development; (2) creation of a statistically based and evaluated categorization of selected NUTS 3 territories into internally-more similar and externally-more dissimilar clusters, according to the representative indicators of tourism demand; and (3) analysis of profiles of identified groups of districts in Serbia, in terms of tourism development achieved in 2019. The practical contributions of the conducted research are reflected in providing: (1) a clear and thorough demonstration of statistically valid application of cluster analysis in the domain of regional tourism development research; (2) an informative snapshot of the situation, regarding the recorded level of tourism activity and development of districts in Serbia in 2019. The proposed classification can serve as a useful tool for identifying the extent of regional tourism development disparities among the observed territorial units, as well as a suitable basis for decision makers and experts in the field of planning and implementation of appropriate regional (and national) tourism (and overall economic) development activities and strategies aimed at mitigating the identified disparities. Finally, this paper adds to the existing literature by filling a specific research gap, elaborated in Section 2.

2. Research background

Since tourism plays a significant part in national economies, it is hardly surprising that numerous scholars conduct research related to tourism development achieved by territorial units at different NUTS level. More precisely, previous analyses use different methodologies in order to empirically examine and / or verify the possible economic benefit and impact of tourism sector, that is, its growth and employment potential, competitiveness, present disparities, and particular regions or countries development levels. The scope of official tourism development indicators (for details: Eurostat, 2014) and multidimensional nature of related activities make the multivariate statistical methods a suitable tool to analyze the aspects discussed above (for details: Dwyer et al., 2012). One of the most applicable multivariate statistical methods in previously conducted studies was the cluster analysis that was used for identifying the extent of regional tourism development disparities between territorial units at different NUTS levels. The Table 1 presents selected relevant research papers with similar research objectives, which applied various clustering methods on territorial units at different NUTS levels, using a diverse set of tourism sector development indicators.

In spite of the obvious similarities among the studies above, their results cannot be seen as comparable with ours due to the differences in terms of the used tourism variables, spatial-temporal scope of analysis, and certain methodological specificities, which directly influence the objectivity and reliability of the conclusions. In this regard, although there are clearly numerous scientific papers dealing with the research topic related to tourism development and territorial classification using cluster analysis (CA), it is important to point out that, according to the authors' knowledge, there is no research on regional tourism development disparities between NUTS 3 territories in Serbia.

Table 1: Selected studies on territorial classification according to different regional tourism development levels

Author(s) / (publication year)	Temporal coverage	Spatial units [NUTS / LAU level]	Spatial coverage [Country(ies) Area]	Methodology approach
Duarte-D. et al. (2021)	2018	87 Municipalities	Santander Dep. (Colombia)	HCA
Chalupa et al. (2013)	2010	77 Districts (LAU-1)	Czech Republic	HCA
Li et al. (2021)	2018	16 Cities in Anhui	Anhui Province (PRC)	HCA & FA
Batista et al. (2021)	2018	1165 NUTS 3 regions	EU-27 Countries	HCA & GIS
Morozova et al. (2016)	2014	65 NUTS 3 regions	Russian Federation	HCA & FA
Gall (2019)	2018	21 Regions of tourism	Slovakia	HCA
Qiao (2018)	2016	12 Provinces (Cities)	PR of China (West)	HCA
Lascu et al. (2018)	2016	17 NUTS 2 regions	Spain	two-step CA
Kolvekova et al. (2019)	2014	54 NUTS 2 regions	10 CEE Countries	HCA
Gorina et al. (2020)	2018	25 NUTS 2 regions	Ukraine	HCA
Vysochan et al. (2021)	2019	25 NUTS 2 regions	Ukraine	Non-HCA

Notes: HCA (Hierarchical cluster analysis), FA (Factor Analysis), Non-HCA (Non-Hierarchical cluster analysis), GIS (Geographical Information Systems)

Source: Authors' tabular representation

In addition, through a detailed analysis of methodological approaches used in the aforementioned 11 studies, the following methodological specificities were identified: (a) Except in the case of the research conducted by Gall (2019), in 8 out of 9 studies in which the hierarchical clustering procedure was applied, Ward's method was used by default, solely based on the subjective assessment of the author(s), without clear statistical justifications. (b) The use of statistical criteria in determining the final number of groups was recorded in approximately 55% of analyzed studies. (c) The quality examination of the obtained final clustering solution, based on the application of various statistical criteria, was conducted only by Morozova et al. (2016) and Vysochan et al. (2021). Also, the use of Non-HCA for the purpose of checking the quality of the derived HCA classification was not recorded in any of the mentioned works. Consequently, compared to the studies in Table 1, the research presented in this paper provides a detailed presentation of a statistically valid implementation of CA in the field of tourism, with the previously observed methodological specificities eliminated, thus providing a triple contribution from a methodological perspective.

3. Materials and methods

In this section, we present detailed descriptions of used numerical indicators, sources of data, temporal-spatial scope of the conducted research, and the applied statistical methodology.

3.1. Variables, data sources and spatial-temporal coverage of research

Secondary data used for calculating the values of the three representative indicators of the achieved level of tourism activity and development were collected and analyzed for each of the 24 official districts and the Belgrade area (i.e. NUTS 3 territorial units) in the Republic of Serbia (RS). Table 2 contains the list of formed tourism variables, supplemented by a detailed procedure used for determination of their values. Data were obtained from the thematic publication of the Statistical Office of the Republic of Serbia (acronym, SORS), *Municipalities and Regions of the RS* (SORS, 2020). Districts belonging to the Autonomous Province of Kosovo and Metohija are not included in the research because the SORS has not provided any information about these territories since 1999. Although the latest available data refer to 2021 (the second year of the COVID-19 pandemic), when tourist activity worldwide was drastically reduced as a result of measures for mitigating and preventing the spread of the pandemic, for this study, we collected and used the data from 2019, since it was the last pre-pandemic year.

Table 2: List of used indicators of tourism activity at district level

Symbols	Tourism activity variables
X_1	Number of domestic tourist arrivals (DTAs) per 1,000 inhabitants
X_2	Number of foreign tourist arrivals (FTAs) per 1,000 inhabitants
X_3	Number of nights spent by tourists (domestic & foreign) per capita

Notes regarding the way of determining the values of selected tourism activity indicators:

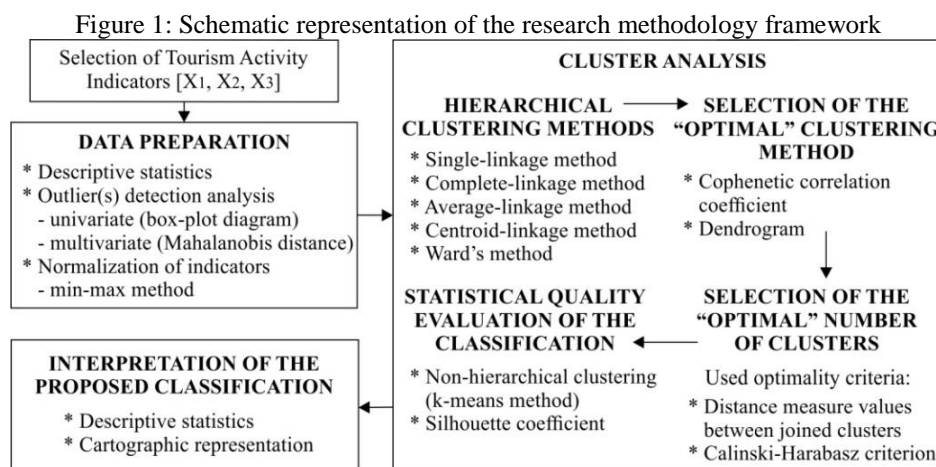
- The values of X_1 are calculated as ratio of total number of domestic tourist arrivals in 2019 and corresponding total number of inhabitants for particular district, multiplied by 1,000;
- The values of X_2 are calculated as ratio of total number of foreign tourist arrivals in 2019 and corresponding total number of inhabitants for particular district, multiplied by 1,000;
- The values of X_3 are calculated as ratio of total number of nights spent by domestic & foreign tourists in 2019 and corresponding total number of inhabitants for particular district.

Source: Authors' tabular representation

The selected core variables (i.e. the number of DTAs, FTAs and total nights spent) are the most commonly used indicators by the World Travel & Tourism Council (WTTC) for measuring the volume of tourism expansion and level of tourism activity. In addition, as can be seen in Table 2, instead of total absolute values, their values are expressed per 1,000 inhabitants, or per capita for individual districts. These calculations were performed, according to the suggestions made by Kolvekova et al. (2019), Li et al. (2021) and Morozova et al. (2016), with the aim of neutralizing or mitigating the effects of population size of individual districts on the selected tourism indicators' values and, consequently, the clustering results. This approach enables the creation of a comparable database, suitable for providing the best insight into the actual tourism attractiveness of the analyzed NUTS 3 territories.

3.2. Research methodology framework

Figure 1 presents the methodological framework primarily based on the implementation of cluster analysis (acronym, CA), one of the most commonly used non-parametric multivariate statistical methods. As a specific, unsupervised learning classification method, CA enables the simultaneous examination of interdependencies between selected tourism activity indicators at NUTS 3 level RS territories in 2019 and, consequently, the discovery of natural, not so obvious, classification structure within the described set of multivariate observations.



Source: Authors' representation

The CA application is supported by appropriate methods of one-dimensional statistical analysis, mainly exploited in the domain of initial data preparation and interpretation of CA results. As can be seen from the given framework, after conducting one-dimensional and multivariate outlier analysis, the normalization of values of selected tourism indicators is performed using the *min–max* method, considering the different measurement units in which they are expressed. Also, by extending the initial range of normalized values (i.e. from 0 to 1) to a new scale ranging from 1 to 10, a more precise comparative basis is provided. This conversion is done using the following expression (Stamenković & Savić, 2017, p. 110):

$$x'_{ij} = 9 \times \frac{x_{ij} - x_j^{\min}}{x_j^{\max} - x_j^{\min}} + 1. \quad (1)$$

Within this expression, the symbols denote the following: x'_{ij} represents normalized and x_{ij} original value of j^{th} tourism indicator for i^{th} district (for $i = 1, 2, \dots, 25$, and $j = 1, 2, 3$), while x_j^{\max} and x_j^{\min} are the largest and smallest original values of j^{th} tourism variable.

In order to obtain a classification solution of the highest statistical quality, the selection of the most suitable hierarchical CA method was made based on the comparison of cophenetic coefficient values, determined for different HCA procedures. Contrary to the default application of Ward's method, the cophenetic-based approach is considered more objective, thus ensuring the necessary scientific basis and confirmation of the CA results. It is application of this approach that presents the first methodological advantage of the present research in comparison to the studies in Table 1. In addition, the selection of the optimal HCA classification, in terms of the final (a priori unknown) number of mutually exclusive, internally-more similar and externally-more dissimilar groups of districts, was performed using two optimality criteria. Their use represents another methodological advantage of this research, since it ensures objectivity in the selection of the final clustering solution, compared to the (highly subjective) approach where the most interpretable classification is chosen on the basis of the researcher's opinion. A comprehensive statistical quality evaluation of the proposed classification of NUTS 3 territories in RS was conducted using the silhouette coefficient values and Non-HCA procedure, based on the application of the *k*-means clustering method. This step in the research framework represents the third methodological advantage of this empirical analysis. For the realization of presented steps of statistical analysis, the software package *IBM SPSS Statistics*, version 20, and *Microsoft Office Excel* were used.

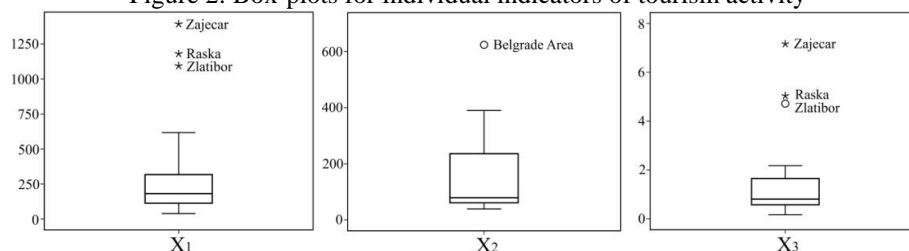
4. Results and discussion

This section presents preprocessing of input multivariate observations and CA classification of 25 districts in RS into internally similar clusters according to the recorded level of tourism activity in 2019. Also, this section focuses on corresponding interpretation of the proposed classification, the comparison of clusters' profiles, and the discussion of results.

4.1. Classification of districts in Serbia by tourism activity indicators

Before the application of CA, during the data preparation phase, the outlier detection analysis was performed in a set of univariate and multivariate observations, based on the adequate graphical representations (i.e. box plots) and Mahalanobis distance values (determined for individual observation units), respectively.

Figure 2: Box-plots for individual indicators of tourism activity



Source: Authors' research

Box-plot diagrams in Figure 2, constructed for each of the three tourism indicators, clearly reveal the presence of several atypical observations (i.e. stars) within the values of variables X_1 and X_3 , while X_2 contains only one value identified as a suspected outlier (i.e. circle). Substantial differences between average and median values, as well as those between the highest and lowest values of analyzed indicators (Table 3), unequivocally confirm previously made statements.

Table 3: Descriptive statistical measures for selected tourism demand indicators

Tourism activity indicators		Average	Median	Minimum	Maximum
Domestic tourist arrivals per 1000 inhabitants	X_1	322.00	181.00	39 [Podunavlje]	1393 [Zaječar]
Foreign tourist arrivals per 1000 inhabitants	X_2	154.88	79.00	39 [Kolubara]	624 [Belgrade area]
Total nights spent by tourists per capita	X_3	1.50	0.81	0.17 [Podunavlje]	7.16 [Zaječar]

Source: Authors' research

In addition, observation units corresponding to the Belgrade area and Zaječar district were marked as multivariate outliers, since their values of Mahalanobis distance measure (i.e., $MD_{BELGRADE} = 17.973$ and $MD_{ZAJEČAR} = 13.931$) are significantly above the 97.5 percentile of chi-square distribution, defined as critical threshold (i.e., $\chi^2_{(3; 0.975)} = 9.348$). Regardless of the sensitivity of CA to the presence of outliers, they were not excluded from further analysis, due to the relatively small size of the data set and the fact that these observations contain information valuable for comparison and creation of a comprehensive classification map of the achieved level of tourism activity in RS at NUTS 3 level territories. It is expected that these districts will form single-member clusters, or perhaps be identified as members of so-called outlier-clusters, together with units of similar tourism properties. After the *min-max* normalization of tourism indicators, following the methodology guidelines given in Figure 1, five different HCA agglomeration methods were implemented on the pre-processed multivariate data set, using the squared Euclidean distance, as an adequate distance measure of their mutual proximity. In order to ensure statistically-based and therefore more objective selection of the most appropriate HCA method for classifying the analyzed territories, for each of the obtained clustering solutions, the corresponding values of cophenetic correlation coefficient (r_{cp}) were determined. Calculated r_{cp} values, representing a specific measure of the overall quality of obtained clustering solutions, are presented in Table 4.

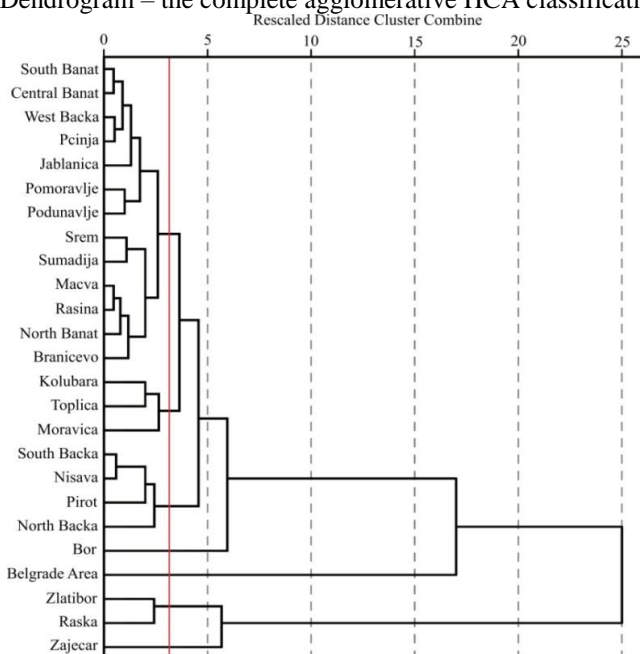
Table 4: Values of cophenetic coefficient for the obtained hierarchical CA solutions

Applied HCA methods	Single-linkage	Complete-linkage	Average-linkage	Centroid-linkage	Ward's method
Cophenetic values	0.9078	0.9175	0.9254	0.9251	0.8954

Source: Authors' research

Although some clustering solutions are characterized by highly approximate cophenetic values (e.g. average- and centroid-linkage methods), the hierarchical classification obtained by applying the average-linkage method was identified as optimal, since it has, generally, the highest r_{cp} value. In addition, a value this high ($r_{cp} = 0.9254 \approx 1$) indicates the presence of an almost perfect correlation between the corresponding values of the initial (Euclidean) and derived (i.e. obtained by average-linkage method) distance matrices, and therefore very high quality of the singled out clustering results. It is also interesting to note that the classification results obtained by Ward's method, according to the presented r_{cp} values, are characterized by the worst quality in this case, even though it represents a HCA method that is most frequently used in CA, but mainly as a consequence of the subjective will of researchers. Precisely from this methodological observation follows justification of the implementation of cophenetic coefficient. The complete (step-by-step) classification structure obtained as a result of selected hierarchical (average-linkage) clustering of 25 NUTS 3 territories in RS, according to the values of analyzed tourism activity indicators in 2019, is given in Figure 3 in the form of specific multilayer tree-based graphical representation, called dendrogram.

Figure 3: Dendrogram – the complete agglomerative HCA classification structure



Source: Authors' research

The presented dendrogram contains and shows 24 different clustering solutions to the analyzed classification problem, regarding the possible number of clusters. The decision about the optimal number of groups was specified, based on the comparative analysis of values of adequate statistical criteria of optimality, calculated for individual classification alternatives, consisting of two to nine groups (Table 5).

By analyzing the evident and expected growing tendency of distance measure values under which single districts or clusters of districts merge within selected consecutive steps of agglomeration process, as well as the size of its corresponding absolute changes, it can be noticed that their first large, sudden increase occurs at the moment of creating a classification alternative with 6 groups of districts. The magnitude of the mentioned increase is even more evident if the fact that the increment of distance value achieved in this step of agglomeration

(2.85) is nearly 7 times higher than the value recorded in the previous step (0.41) is taken into account. In addition, the tendency of pseudo F statistic' values suggests an identical conclusion. In fact, the previously emphasized agglomeration step that results in the classification of districts in 6 different groups, is recognized, by this criterion as well, as one within which the most pronounced value decrease of pseudo F statistic (i.e. from 100.88 to 62.12) occurred. This decrease (-38.76) is the largest recorded among the analyzed solutions. Having in mind the fact that observed drastic changes in optimality criteria values occur mainly as a consequence of merging two highly dissimilar clusters, the classification solution consisting of 7 clusters is selected as the optimal since it precedes the previously described, less desirable, hierarchical agglomeration results. Viewed from a graphical perspective, the moment of achieving the optimal CA classification of districts, during the process of agglomeration, is marked on the constructed dendrogram with a red vertical line (Figure 3).

Table 5: Optimality indicators' values for different clustering solutions

Used optimality indicators		Possible classifications with different numbers of clusters							
		9	8	7	6	5	4	3	2
Distance between joined clusters	value	2.06	2.14	2.55	5.40	7.94	12.78	13.37	66.68
	change	0.65	0.08	0.41	2.85	2.54	4.84	0.89	53.01
Pseudo- F statistic	value	95.05	96.09	100.88	62.12	39.24	44.69	55.65	46.08
	change	-	1.04	4.79	-38.76	-22.88	5.45	10.96	-9.58

Source: Authors' research

For the purpose of a statistical evaluation of the quality of the obtained classification, in terms of the specified structure of extracted 7 clusters, a non-HCA approach (i.e. k -means clustering method) was applied to normalized values of observed three tourism variables. The main reason for doing so lies in the fact that non-HCA approach, unlike HCA agglomeration, represents a reversible classification process, since it allows reallocation of individual observations during the clustering. As a result of this activity, the identical allocation was obtained for all districts, in terms of their membership within previously identified 7 clusters, thus confirming the statistical quality of the formed classification. The final step in statistical quality evaluation of obtained 7-cluster classification results is based on the interpretation of the silhouette coefficient values, calculated for the overall solution and individual clusters within its structure. Representing a comprehensive statistical measure of the achieved level of internal homogeneity and external heterogeneity at the mentioned levels of observation, these values are presented in Table 6.

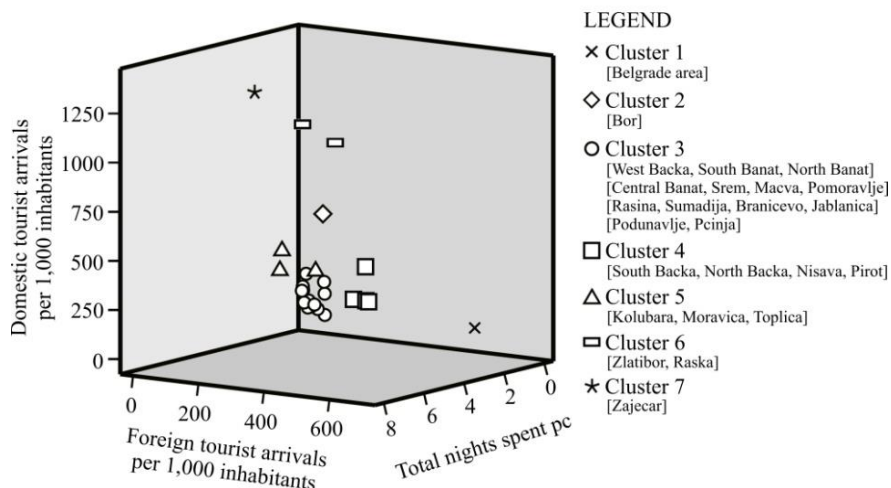
Table 6: Silhouette coefficient values for the obtained CA classification of districts

Coefficient	Individual clusters' code							Overall solution
	C-1	C-2	C-3	C-4	C-5	C-6	C-7	
Silhouette values	1.00	1.00	0.79	0.76	0.58	0.83	1.00	0.85

Source: Authors' research

The obtained overall silhouette value (0.85), since it is within the range from 0.70 to 1.00, suggests that the extracted 7-clusters classification is of very high quality. According to guidelines given in [Izenman \(2008\)](#), the proposed clustering structure can be defined as strong. The validity of this conclusion is confirmed by individual clusters' silhouette values, because for 6 groups the same level of quality as the overall was recorded. The exception is only cluster marked as C-5, characterized by a moderate level of quality.

Figure 4: 3D Scatter diagram of classification of districts in RS by tourism activity indicators



Source: Authors' research

A graphical representation of the obtained tourism activity classification of 25 districts in RS into 7 clusters is given in Figure 4. It should be noted that the identified real multivariate (i.e. Belgrade, Zaječar) or the one-dimensional outliers (i.e. Zaječar, Raška and Zlatibor) were isolated, as expected during the data preparation phase, as members of so-called outlier clusters, created in a form of single-member clusters (i.e. C-1, C-7) or very small size group (i.e. C-6). The reason why the Bor district represents the only member of C-2 lies in the fact that this unit was identified as a suspected atypical multivariate observation.

4.2. Interpretation of the obtained classification and discussion

The interpretation of the proposed multivariate typology of districts in RS, together with the corresponding values of selected indicators of tourism activity and demand in 2019, is presented below. The overview of main numeric characteristics for extracted clusters is given in Table 7, together with appropriate cartographic representation (Figure 5).

Table 7: Min / max / average (or individual) values of tourism activity indicators per clusters

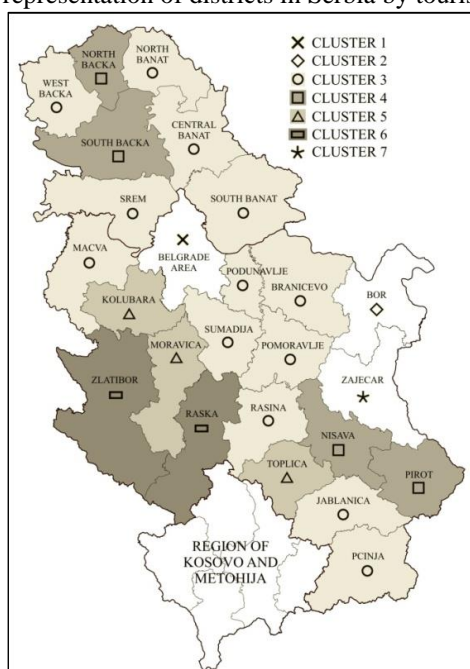
Clusters of districts		Number of DTAs per 1,000 inhabitants			Number of FTAs per 1,000 inhabitants			Number of total nights spent per capita		
Code	Size	min	max	average	min	max	average	min	max	average
C-1	1	-	-	119	-	-	624	-	-	1.6
C-2	1	-	-	617	-	-	178	-	-	1.9
C-3	13	39	261	140	41	114	68	0.2	1.2	0.6
C-4	4	151	342	201	192	257	234	0.6	1.1	0.8
C-5	3	313	395	342	39	149	84	1.7	2.2	1.9
C-6	2	1094	1180	1137	309	390	350	4.7	5.0	4.9
C-7	1	-	-	1393	-	-	293	-	-	7.2
National average		322			155			1.5		

Source: Authors' research

Four, out of a total of 7 extracted groups of districts, previously identified as outlier-clusters, represent single-member or two-member clusters. Together, they comprise only 5 out of a total of 25 observed districts, or 20% of the sample. By comparing the presented individual / average values of tourism activity indicators with the corresponding country's mean values,

it can be seen that these districts are characterized mainly by the numbers of DTAs and FTAs per 1,000 inhabitants, as well as the nights spent per capita that are significantly above the corresponding averages determined at the national level, and even for most of other 3 (multi-member) groups of districts. Therefore, it can be stated that these 5 districts, in proportion to the size of their populations, are characterized by the highest tourism sector development level, compared to the districts allocated within the remaining 3 clusters. Consequently, it would be justified and logical to merge them into one common cluster, whose descriptive name can be formulated as: a *high level of tourism sector development*. However, regardless of this common characteristic, each of these clusters is distinguished by the corresponding specificities in terms of the considered indicators of tourism activity.

Figure 5: Cartographic representation of districts in Serbia by tourism activity classification



Source: Authors' research

More precisely, Zaječar district, as the only member of C-7, stands out as an extremely attractive and desirable destination for the largest number of DTAs (i.e. $\approx 1,393$ per 1,000 inhabitants), with the highest average number of overnight stays per capita (i.e. ≈ 7.2). These observations are fully expected, since this district is considered very rich in terms of the available tourist offer, comprising numerous natural attractions and resources (e.g. Rtanj and Stara Planina mountain with nature park and ski-center, Sokobanja and Gamzigrad spa resorts, Bogovina Cave – the longest cave in RS, Ripaljka waterfall, etc.), cultural-historical sites and monuments (e.g. Felix Romuliana – the archaeological site of the ancient Roman complex, Zaječar National Museum, etc.), medieval monasteries (e.g. Suvodol, Grlište, etc.), and popular musical events (e.g. Guitariada, the rock manifestation with the longest tradition in Europe). The wide and diverse range of tourist sights that this district offers to visitors, manifested in the form of different types of tourism, from spa, mountain, adventure, countryside, urban, archeological to cultural-entertainment tourism, stimulates the arrival of a large number of tourists throughout the year. As the most receptive tourist area within this district, Sokobanja, the oldest spa in Serbia, stands out, since it was visited by 108,151

domestic and 16,726 foreign tourists in 2019, which makes up 73.2% of the total number of DTAs and 53.9% of the total FTAs recorded in Zaječar district that year.

Cluster 6 includes only Zlatibor and Raška districts. It is also characterized by extremely favorable values of tourism activity indicators, in proportion to the population size. Out of a total of 1.84 million domestic and 1.85 million foreign tourists who visited Serbia in 2019, 35.3% and 10.7% of them, respectively, visited some of the tourist attractions located in these districts. With a population of about 570,000 inhabitants, nearly 1137 DTAs and 350 FTAs per 1,000 inhabitants, and ≈ 5 overnight stays per capita, these two districts can be classified as the most visited tourist areas in RS. Compared to the C-7, and taking into account the population ratio of 5.4:1 in favor of C-6, the position of this cluster, in terms of the tourism development level, can be considered more favorable, due to the larger number of FTAs, regardless of the slightly lower values of the other two indicators. Also, mountain and countryside tourism plays a dominant role in the tourist offer of C-6 districts, thanks to their numerous mountains and ski centers famous for their beauty, intact nature and various recreational facilities (Kopaonik, Zlatibor, Tara, Zlatar and Golija). Although this specificity reflects one of the key differences between C-6 and C-7, Zlatibor and Raška districts are also characterized by highly developed spa tourism, given the huge tourism potential of their numerous spa resorts, such as: Jošanička, Mataruška, Bogutovačka spa, and of course Vrnjačka Banja spa, officially the most visited spa in Serbia.

The district of Bor is the only member of cluster C-2 and is characterized by nearly double number of DTAs, and slightly higher number of FTAs per 1,000 inhabitants and nights spent per capita compared to the country's averages. These values are twice, or even more, lower than the comparable values in C-6 and C-7, indicating the lower level of tourism sector development, in spite of the great potential. The dominant types of tourism are recreational, adventure, countryside, religious and archeological-historical tourism, since the central part of tourist offer of this district includes the Djerdap National Park, numerous monuments of nature and beauty spots (e.g. Bor Lake, Rajko's Cave, Vratna Gates, the Great Kazan gorge on Danube), Lepenski vir (one of the largest archeological sites from the Stone Age), and cultural-historical monuments of exceptional importance (e.g. Rajac and Rogljevo wine cellars).

With nearly 4 times the number of FTAs, approximately 3 times lower number of DTAs per 1,000 inhabitants, and a slightly higher number of nights spent per capita, compared to the values of national averages, the Belgrade area (C-1), stands out. Regarding the tourism indicators' values and observed specificities, it can be stated that C-1 shows similar tendencies as C-2, but in the opposite direction. In fact, in terms of the number of FTAs, this territory holds a record value per 1,000 inhabitants, which is especially evident if the size of population (nearly 1.7 million inhabitants, i.e., $\approx 25\%$ of the population in RS) is taken into account. A completely opposite tendency is present in the case of the number of DTAs. The numerical characteristics of C-1 are not a surprise, since it covers the territory of Belgrade city, the capital of the RS. The dominant types of tourism are city, business, urban-adventure, archeological-historical, shopping, health tourism, etc.

The remaining 20 districts, or nearly 80% of the observed NUTS 3 territories, were placed into three other clusters. The average values of selected tourism indicators determined for these clusters (Table 7) are significantly below the corresponding national averages, but also below the average values of the previously interpreted 4 outlier-clusters. Therefore, it is logical to conclude that these 3 clusters are characterized by the level of tourism development that is lower than that of clusters C-1, C-2, C-6 and C-7.

The fact that tourist destinations located in the territories of 5 districts (i.e. Belgrade, Bor, Raška, Zlatibor and Zaječar), distributed within clusters characterized by a high level of

tourism activity and development (C-1, C-2, C-6, C-7), recorded 57.9% and 70.7% of the total number of registered DTAs and FTAs in 2019, respectively, with more than 64% of the total registered overnight stays in RS, unequivocally indicates the supremacy of these clusters of NUTS 3 territories, in terms of tourism activity, compared to the remaining 20 districts. The observed differences in the level of tourism activity are even more striking in the light of the fact that these 5 districts together cover only 23% of the total territory and include almost 35.7% of the total population of RS. Also, they are responsible for $\approx 55.6\%$ of total regional gross added value in the Wholesale / retail trade, transportation / storage, and accommodation / catering services sector (SORS, 2022). It is also interesting to note that 4 clusters of high tourism activity, in general, differ in terms of the dominant type of tourism, namely: mountain (C-6), spa (C-7), adventure / natural beauties (C-2) and urban / business tourism (C-1). Consequently, it can be stated that Serbia, unfortunately, is characterized by very noticeable and emphasized interregional inequalities, present among the analyzed NUTS 3 territories, regarding the recorded level of tourism activity and development level in 2019. This observation is fully consistent with the statement made by Gajić et al. (2014) about the significant disproportions between regions in the RS in terms of the volume of tourist traffic, as well as by Bećirović et al. (2011) who point out the small dispersion of tourist activity at the regional level, highlighting the city of Belgrade, Zlatibor and Raška districts as the main bearers of tourism activity. In this context, it should be noted that a further and more detailed comparison and discussion of our research results with these or other (similar) empirical analyses conducted in the field of tourism on the territory of Serbia is not generally feasible, due to very noticeable differences in terms of selected temporal scope of research, observation units' NUTS level, used tourism activity indicators, applied statistical methodology framework, etc. Regardless of these notes, the proposed classification of NUTS 3 territories in Serbia can be used as a useful analytical basis for further quantitative analysis, and as an additional source of potentially valuable information for creators of policies and strategies for (regional) development of this economic activity.

5. Conclusion

Given the indisputable importance of information regarding the tourism sector development at different NUTS level territories for planning and efficient implementation of appropriate regional (and national) tourism (and overall economic) development strategies and measures aimed at mitigating (eventually) present regional disparities, in this paper, according to the formulated research objectives, a statistically demanding multivariate methodology framework (based on combined implementation of non-hierarchical and hierarchical agglomerative clustering procedures), intended for classification of districts in Serbia, in terms of the selected tourism activity indicators' values in 2019, is applied and presented. Based on a thorough statistical examination of its validity and quality, the proposed optimal classification, composed of seven internally-homogeneous / externally-heterogeneous clusters of districts, unambiguously verifies the existence of noticeable and large tourism development disparities between NUTS 3 territories, and intra-regional tourism activity disproportion in Serbia, present in the following direction: developed east / south-west, with the city of Belgrade – less developed north / central part of Serbia.

The presented multivariate statistical approach, intended for the analysis of regional tourism development inequalities, is characterized by certain methodological advantages, in comparison to similar studies, namely: (a) it enables simultaneous examination of interdependencies between representative tourism indicators, unlike the approach based on evaluation and analysis of individual indicators' values and separate interpretation of numerous one-dimensional classifications; (b) the selection of the optimal HCA method and the best quality classification structure is conducted by using the appropriate, statistically

based criteria rather than the researcher's subjective decision; and (c) statistical validity of interpreted classification was additionally verified by non-hierarchical allocation results, thus ensuring objectivity and scientific justification of the obtained classification results.

Finally, due to the methodological specificities of our analysis, the findings of this study may be insufficiently comparable to the results of other scholars. This, together with the impossibility to use a larger number of tourism variables, due to the unavailability of useful data for the territories of selected NUTS level in Serbia, can be singled out as key limitation of this empirical research. On the other hand, the conducted analysis is highly applicable and the applied methodology is flexible, which suggests that in future research, it can be used with different spatial (i.e. territories' NUTS level / a country) and / or temporal (i.e. year) data coverage, and the same or an expanded list of representative tourism indicators. In addition, further studies may also incorporate other statistical methods.

Conflict of interest

The authors declare no conflict of interest.

References

1. Batista e Silva, F., Barranco, R., Proietti, P., Pigaiani, C., & Lavallo, C. (2021). A new European regional tourism typology based on hotel location patterns and geographical criteria. *Annals of Tourism research*, 89, 103077. <https://doi.org/10.1016/j.annals.2020.103077>
2. Bećirović, S., Plojović, Š., Ujkanović, E., & Bušatlić, S. (2011). Održivi razvoj turizma u regionima Srbije [Sustainable development of tourism in regions of Serbia]. *Univerzitetska Hronika*, 3(1), 39–47.
3. Chalupa, P., Prokop, M., & Rux, J. (2013). Use of cluster analysis for classification of tourism potential. *Littera Scripta*, 6(2), 59–68.
4. Duarte–Duarte, J., Talero–Sarmiento, L. H., Rodriguez–Padilla, D. C. (2021). Methodological proposal for the identification of tourist routes in a particular region through clustering techniques. *Heliyon*, 7(4), e06655. <https://doi.org/10.1016/j.heliyon.2021.e06655>
5. Dwyer, L., Gill, A., & Seetaram, N. (2012). *Handbook of research methods in tourism—quantitative and qualitative approaches*. Cheltenham, UK: Edward Elgar Publishing Limited.
6. Eurostat (2014). *Methodological manual for tourism statistics*. Retrieved May 6, 2022 from <https://ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/ks-gq-14-013>
7. Gabdrakhmanov, N. K., Rozhko, M. V., & Rubtzov, V. A. (2016). Cluster analysis in tourism development. *International Business Management*, 10(22), 5291–5294.
8. Gajić, T., Vujko, A., & Vugdelija Kočić, V. (2014). Utvrđivanje međuregionalnih dispariteta u razvoju turističke privrede Srbije [Determining interregional disparities in the development of tourism economy in Serbia]. *Ekonomski signali: Poslovni magazin*, 9(1), 113–129.
9. Gall, J. (2019). Determining the significance level of tourist regions in the Slovak Republic by cluster analysis. *Economic Review*, 48(4), 451–462.
10. Gorina, G. O., Barabanova, V. V., Bohatryyova, G. A., Nikolaichuk, O. A., & Romanukha, A. M. (2020). Clustering of regional tourism service markets according to indicators of the functioning of subjects of tourism activity. *Journal of Geology, Geography and Geoecology*, 29(4), 684–692. <https://doi.org/10.15421/112061>
11. Izenman, A. J. (2008). *Modern multivariate statistical techniques*. New York: Springer Science+Business Media, LLC.

12. Jackson, J., & Murphy, P. (2006). Clusters in regional tourism: An Australian case. *Annals of Tourism Research*, 33(4), 1018–1035. <https://doi.org/10.1016/j.annals.2006.04.005>
13. Kolvekova, G., Liptakova, E., Štrba, L., Kršak, B., Sidor, C., Cehlar, M., ... & Behun, M. (2019). Regional tourism clustering based on the three Ps of the sustainability services marketing matrix: An example of Central and Eastern European countries. *Sustainability*, 11(2), 400. <https://doi.org/10.3390/su11020400>
14. Lascu, D.-N., Manrai, L. A., Manrai, A. K., & Gan, A. (2018). A cluster analysis of tourist attractions in Spain: Natural and cultural traits and implications for global tourism. *European Journal of Management and Business Economics*, 27(3), 218–230. <https://doi.org/10.1108/EJMBE-08-2017-0008>
15. Li, X., Zhan, X., & Jiang, J. (2021). Comprehensive evaluation of tourism development potential in Anhui Province based on cluster analysis and factor analysis. *Open Journal of Business and Management*, 9(2), 866–876. <https://doi.org/10.4236/ojbm.2021.92046>
16. Morozova, L., Morozov, V., Havanova, N., Litvinova, E., & Bokareva, E. (2016). Ensuring the development of tourism in the regions of the Russian Federation, with account of the tourism infrastructure factors. *Indian Journal of Science and Technology*, 9(5), 1–13. <https://doi.org/10.17485/ijst/2016/v9i5/87599>
17. Petrović, G., Karabašević, D., Vukotić, S., & Mirčetić, V. (2020). An overview of the tourism economic effect in the European Union member states. *Turizam*, 24(4), 165–177. <https://doi.org/10.5937/turizam24-26469>
18. Qiao, Y. (2018). Cluster Analysis of the development levels of tourism economy in twelve provinces (cities) in China's western regions. In J. Guo & K. L. Teves (Eds.), *Proceedings of 3rd International Conference on Politics, Economics and Law – ICPEL 2018*, (pp. 212–215). Dordrecht, The Netherlands: Atlantis Press. <https://doi.org/10.2991/icpel-18.2018.52>
19. Republički zavod za statistiku [Statistical Office of the Republic of Serbia – SORS] (2020). *Opštine i regioni u Republici Srbiji [Municipalities and regions in the Republic of Serbia]*. Retrieved June 8, 2022 from www.stat.gov.rs/en-us/publikacije/publication/?p=12795
20. Republički zavod za statistiku [Statistical Office of the Republic of Serbia – SORS] (2022). *Regionalni bruto domaći proizvod 2020 [Regional gross domestic product 2020]*. Retrieved May 1, 2023 from <https://www.stat.gov.rs/sr-Latn/oblasti/nacionalni-racuni/regionalni-podaci>
21. Rita, P. (2000). Tourism in the European Union. *International Journal of Contemporary Hospitality Management*, 12(7), 434–436. <https://doi.org/10.1108/09596110010347374>
22. Roman, M., Roman, M., Grzegorzewska, E., Pietrzak, P., & Roman, K. (2022). Influence of the COVID-19 pandemic on tourism in European countries: Cluster analysis findings. *Sustainability*, 14(3), 1602. <https://doi.org/10.3390/su14031602>
23. Roman, M., Roman, M., & Niedziółka, A. (2020). Spatial diversity of tourism in the countries of the European Union. *Sustainability*, 12(7), 2713. <https://doi.org/10.3390/su12072713>
24. Stamenković, M., & Savić, M. (2017). Measuring regional economic disparities in Serbia: Multivariate statistical approach. *Industrija*, 45(3), 101–130.
25. Vysochan, O., Vysochan, O., Hyk, V., & Hryniv, T. (2021). Attributive-spatial tourist clustering of regions of Ukraine. *GeoJournal of Tourism and Geosites*, 35(2), 480–489. <https://doi.org/10.30892/gtg.35228-675>
26. World Travel & Tourism Council (WTTC) (2022). *Travel & tourism economic impact 2022*. Retrieved October 2, 2022 from <https://wtcc.org/Portals/0/Documents/Reports/2022/EIR2022-Global%20Trends.pdf>