Cluster Analysis of Ukrainian Regions Regarding the Level of Investment Attractiveness in Tourism

Ganna Kharlamova [1], Andrii Roskladka [2], Nataliia Roskladka [2], Andriy Stavytskyy [1] and Yuliia Zabaldina [2]

[1] *Taras Shevchenko National University of Kyiv, Vasylkivska st. 90A, 03022, Kyiv, Ukraine*
[2] *Kyiv National University of Trade and Economics, Kyoto st. 19, 02156, Kyiv, Ukraine*

Abstract

The article contains a description of the process and results of the implementation of the k-means algorithm in the analytical platform Loginom for the problem of clustering the regions of Ukraine by the level of investment attractiveness in the field of tourism. The selection of tourism clusters and their ranking is a difficult task in the field of data analysis, as there is no single consolidated indicator of investment attractiveness. The conclusion about the affiliation of a particular region to one of the tourist clusters is determined by a set of indicators of the volume of tourist services for different types of economic activity in the field of tourism. The Loginom system has powerful tools for cluster analysis using EM-Clustering, k-means, g-means and others. The tools of statistical and visual analysis of the obtained results deserve special attention: Table, Statistics, Chart, OLAP-Cube, Cluster Profiles. Clustering has made it possible to identify groups of regions that are actively developing the tourism industry (primarily Kyiv city and Odesa region) and are currently formed for tourism investors. Equally important is the selection of problem regions that have a low level of attractiveness for domestic and foreign tourism. It is noted that Ukraine has a huge potential for the development of the tourism industry. The regions that, according to the results of the cluster analysis, are in the problem group have "world-class tourist pearls". The Government of Ukraine and local authorities should pay attention to the insufficient level of development of the tourism industry, provide comprehensive support to the regions that are in the problem cluster, and thus increase their level of investment attractiveness.

Keywords

Tourism industry, regions of Ukraine, investment attractiveness, tourism cluster, cluster analysis, EM-clustering, k-means algorithm, Loginom analytical platform

1. Introduction

Tourism is one of the most important forms of international cooperation, which provides many countries with significant budget revenues and employment growth. The formation and development of the tourism market in Ukraine in recent years has taken place in conditions of a sharp decline in consumption of tourist services, exacerbation of inflation with a corresponding increase in prices, limited demand and declining real incomes. Problems related to the spread of the coronavirus, significant restrictions on the movement of citizens within the country and the almost complete closure of borders have further complicated the development of tourism in Ukraine. Search tools to solve these problems require radical economic transformations and the use of innovative tools for forecasting the development of tourism through mathematical and computer modelling to manage the activities of tourism enterprises at the regional and national levels.

ICTERI-2021, Vol II: Workshops, September 28 – October 2, 2021, Kherson, Ukraine
EMAIL: akharlamova@ukr.net (A. 1); a.roskladka@knute.edu.ua (A. 2); n.roskladka@knute.edu.ua (A. 3); a.stavytskyy@gmail.com (A. 4); y.zabaldina@knute.edu.ua (A. 5)
ORCID: 0000-0003-3614-712X (A. 1); 0000-0002-1297-377X (A. 2); 0000-0001-7333-4050 (A. 3); 0000-0002-5645-6758 (A. 4); 0000-0003-2741-5604 (A. 5)

CEUR Workshop Proceedings (CEUR-WS.org)

The development of tourism is associated with the creation of a tourist product, the development of tourist infrastructure, investment growth, formation and strengthening of the image of the territory, several solutions for urban design, areas and forms of promotion. Each solution requires reliable information support based on the calculation and analysis of quantitative indicators, which are based on statistical data [1]. The authors of the article are part of the authors of the "Tourist Barometer of Ukraine" – an innovative statistical publication, which was presented in late 2020 by the National Tourist Organization of Ukraine. Clear and accessible tour-ism statistics become the basis for strategic decisions by both entrepreneurs and state and local authorities, including those related to investment or financing of tourism projects.

Despite declaring the availability of statistics in the field of tourism, the national system of tourism statistics is characterized by some internal contradictions, which greatly complicates the further use of data for strategic decisions [1]:

- the time lag of information flows from market needs. Thus, some of the data for 2019 were published only at the end of 2020, and some data for 2019 remained unpublished at the time of writing;
- lack of time series for some data groups;
- loss of a significant part of the tourist flow (domestic tourism), as well as producers of certain services outside the relevant statistical observations.
- inconsistency of data obtained from different sources;
- obvious inaccuracy of some administrative data due to the imperfection of the relevant procedures for obtaining them;
- the inconvenience of presenting and searching for information, which is now presented on different platforms and in extremely inconsistent forms.

Another problem that arises after the collection of statistics is their heterogeneity. The task of ranking the regions of Ukraine by the level of their investment attractiveness is complicated by the lack of a clear leader or outsider among all performance indicators. The authors propose to conduct a cluster analysis of the regions of Ukraine to identify tourism clusters of different levels based on the consolidation of statistical indicators.

Cluster analysis or clustering is the task of grouping a set of objects in such a way that objects in the same group (called a cluster) are more similar (in some sense) to each other than to those in other groups (clusters).

As a software implementation of cluster analysis of regions of Ukraine, the authors use Loginom - an analytical platform that allows you to perform all stages of business analysis in a single environment: from data consolidation and model building to visualization and integration into the business process. The main advantage of the Loginom analytical platform is that it has a free version, is a low-code system and is available to almost all users of any skill level. All steps in the clustering process are intuitive. The visual designer allows adjusting all processes of the analysis: integration, data preparation, modelling, visualization.

2. Literature and hypothesis development

The field of tourism is the subject of research by a large number of specialists in various fields. Recently, much attention has been paid to modern information tools for the development of tourist destinations, for example, such as Internet marketing [2]. The authors of the article have considerable experience in the study of tourism [3-6], in particular, in the application of mathematical and computer modelling methods for research.

Clustering, which is the main methodological tool of this research, is one of the most well-known methods of Data Science. General issues of clustering are fully covered in the sources [7, 8]. Also in scientific works, it is possible to meet a sufficient number of specific methods of cluster analysis. Thus, Iyigun C., Türkeş M., Batmaz I., Yozgatligil C., Purutçuoğlu V., Kartal E., Öztürk M. in the article [9] and Sablin K., Kagan E., Chernova E. in the article [10] use methods of hierarchical clustering, Gorbatiuk K., Mantalyuk O., Proskurovych O., Valkov O. in [11] investigate fuzzy clustering methods. The advantages and disadvantages of cluster analysis are well noted Mazur, V. Barmuta, K. Demin, S. Tikhomirov, E. and Bykovskiy, M. [12]. Interesting is the article by Jakobsen S. E., Njos R., [13], which examines the negative role of cluster projects, which are often supported by certain regional industries and sectors.

Cluster analysis is often used in scientific works to study the differentiation of socio-economic development of regions, both domestic and foreign authors. In particular, Sablin K., Kagan E., Chernova E. in [10] study the clustering of regions of Russia, Opmane I. [17] conducts a cluster analysis of the regions of Latvia. Works [15,16,18-21] are devoted to different directions of cluster construction among the regions of Ukraine. The formation of tourist clusters is carefully studied in the monograph of A. Mazaraki [22] and the article by Mudrak R., Moiseeva N. [23].

Different software systems and programming languages are used for software implementation of clustering processes: R [10], Deductor [14], SPSS [15], Statistica [16].

The Loginom system [24, 25], which recently appeared on the market of analytical platforms, also has powerful tools for various types of clustering and effective means of visualizing the results of cluster analysis.

3. Research methodology

The fundamental principles of the research are the application of a systematic approach, analysis and synthesis of information and the reasonable use of information technology to obtain scientific research results.

In particular, the following scientific methods are used in research:

- method of system analysis - to determine the most important and influential indicators of tourism in the regions;
- k-means algorithm – for the formation of tourist clusters based on consolidated information on the volume of tourist services provided by different regions of Ukraine;
- graphical method – to create diagrams of the distribution of average values of services by clusters and to build a map of the regions of Ukraine based on the results of cluster analysis;
- method of quantitative analysis – to study the structure of Cluster profiles, the importance of clusters and determine statistical characteristics for different types of travel services.

The information basis of the study is the statistical data on the volume of tourist services provided in the regions of Ukraine, collected by the authors of the article during the formation of the Tourist Barometer of Ukraine.

This study aims to conduct a cluster analysis of the regions of Ukraine to identify regions with a favourable investment climate in tourism, as well as problem regions that need further tourism development and support at the state and local levels.

Clustering is an effective method of dividing objects into groups according to similar characteristics and separating them from other groups that have different characteristics. To carry out cluster analysis, Loginom is used. It is an analytical platform that provides in-depth analytics and allows you to make management decisions based on accurate and reliable information. The platform has a user interface that does not require special training. Loginom has support for analysis technologies: from simple logic to machine learning.

The obtained results of clustering will contribute to the further study of the development of tourism in Ukraine to increase its investment attractiveness.

4. Results

4.1. Choosing a clustering method

According to the aim of the research, it is necessary to divide the regions of Ukraine into functional groups according to the observed volume of tourist services and to identify hidden patterns. As mentioned in the Introduction, obtaining initial data for clustering is a separate challenge. The authors of the article selected for analysis 13 influential indicators that are the most important in the total volume of tourist services in the region. The indicators are selected in such a way as to comprehensively and objectively represent the provided tourism services and minimize the problems with data in the field of tourism, which the authors mentioned in the introduction to the article. This approach will make it possible to effectively carry out the process of clustering the regions of Ukraine. The initial data are formed by the authors based on [1, 26-28].

The Loginom analytical platform allows clustering using one of three methods:

- EM-Clustering;
- k-means Clustering;
- g-means Clustering.

The authors created a model in the Loginom system and solved the problem of clustering using all three methods (Fig. 1)

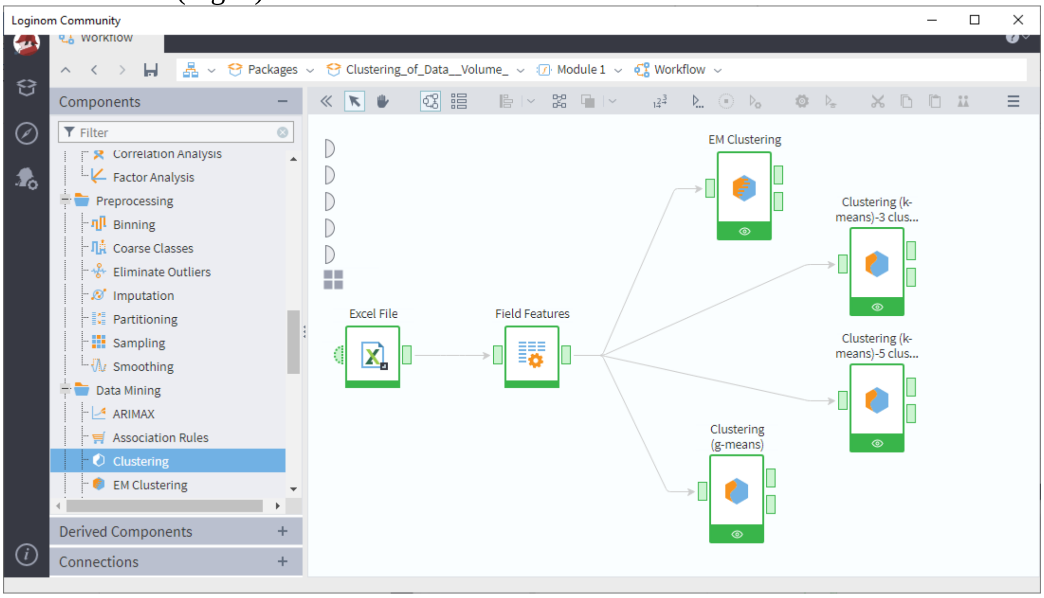


Figure 1: Model of the clustering problem in the Loginom system

Consider briefly the features of the three methods available for clustering based on the Loginom platform.

EM (Expectation-maximization) – a popular clustering algorithm that allows you to work efficiently with large amounts of data. EM-clustering is based on the EM-algorithm, which is based on the assumption that the studied data set can be modelled using a linear combination of multidimensional normal distributions. The aim is to estimate the distribution parameters that maximize the logarithmic likelihood function, which is used as a measure of model quality. In other words, it is assumed that the data in each cluster is subject to a certain distribution law, namely, the normal distribution.

Thus, any observation (object) belongs to all clusters, but with different probabilities. The object should be assigned to the cluster for which this probability is higher.

The field of application of the EM-algorithm is extremely wide: discriminant analysis, clustering, restoration of gaps in data, etc. The EM algorithm is based on the assumption that the clustered data obey a linear combination (mixture) of normal (Gaussian) distributions. The name of the algorithm comes from the words «expectation-maximization». Its purpose is to determine and estimate the distribution parameters - mean and variance, which maximizes the likelihood function used as a measure of the model's quality.

Among the advantages of the EM algorithm are the following:

- effective processing of Big Data;
- resistance to noise and data gaps;
- ability to build the desired number of clusters;
- fast convergence with successful initialization.

The problem of clustering Ukrainian regions according to 13 main parameters can hardly be attributed to Big Data. Thus, our attention will be focused on two other methods, for which the handler performs clustering of objects based on the k-means and g-means algorithms. The main difference between one algorithm and another is whether the number of clusters is known in advance. If the number of clusters is known, then the k-means algorithm is used, otherwise – g-means, which determines this number automatically within a specified interval.

From a formal point of view, all means methods use a single algorithm, which differs in some computational parameters. There are clustering methods that can be viewed as derived from k-means. For example, the k-medians method uses the median rather than the mean to calculate the centroids, which makes the algorithm more robust against anomalous values in the data.

The g-means algorithm builds clusters in which the data distribution tends to the normal (Gaussian) and removes the uncertainty in the choice of the initial clusters. The C-means algorithm uses elements of fuzzy logic, taking into account when calculating centroids, not only distances but also the degree of belonging of observation to a set of objects in a cluster. There is also known Lloyd's algorithm, which uses not a set of vectors, but a region of a vector space as an initial partition. The idea of the k-means method was simultaneously formulated by Hugo Steinghaus and Stuart Lloyd in 1957. The term «k-means» was first introduced by J. McQueen in 1967.

Thus, for the clustering of regions, we will use the basic algorithm k-means.

4.2. Implementation of the clustering process in the Loginom system

In the first module, we download the data shown in Fig. 1 from the Excel file (Fig. 2). For quick registration when replacing the full names of travel services, we leave only the appropriate classification codes.

In the second stage, we will configure the fields (Fig. 3), specifying the types and kinds of data that will be used in the future for clustering.

The second module – Field Features is used to configure additional field parameters, specify input and output parameters of the model for loading them into clustering nodes.

The third module is a direct clustering processor for one of the certain types. In this article, we will consider the application of the k-means algorithm to divide many regions of Ukraine into three and five clusters. The division into three clusters corresponds to the usual system of evaluation, which can be divided into clusters with "high", "medium" and "low" levels of investment attractiveness in the field of tourism. The division of the initial data set into 5 clusters is given to further detail the groups of regions and study the stability of the clustering process. In other words, you need to find out how the clusters will behave as their numbers increase.

Setting the parameters of the third module requires the choice of the method of data normalization, as which we choose the standardized values of the volume of tourist services (Fig. 4).

In the next step of configuring the Clustering processor, it is proposed to set a certain number of clusters or choose the method of auto-detection of such a number. According to the problem statement, we build two separate modules for the implementation of the k-means algorithm. First, select the division into 3 clusters (Fig. 5), and in another module at the same step, select 5 clusters.

The last step for the Clustering processor is to configure the visualizers (Fig. 6), which will be discussed in detail and analyzed below.

In the future, to interpret the results of clustering, we will consider in parallel the visualizers of both modules (for three and five clusters), presented in the model in Fig.1. Clustering by the k-means method gave the following quantitative results - cluster weights. When 25 regions were divided into three clusters, cluster "0" included 18 regions, cluster "1" has a single representative, cluster "3" contains 6 regions (Fig. 7).

When divided into five clusters, cluster "0" includes 3 regions, cluster "1" contains 18 regions, cluster "4" - two and clusters "2" and "3" have one representative (Fig. 8).

The Table visualizer in the Clustering visualization subsystem (fig. 6) gives a qualitative clustering composition. For the case of three clusters, we have the results presented in Fig. 9.

According to the results of clustering, Kyiv city forms a separate cluster, six regions of Ukraine (Dnipropetrovsk, Ivano-Frankivsk, Kyiv, Lviv, Odesa and Kharkiv regions) form cluster "2", the rest of the regions belong to cluster "1".

For the case of five clusters, we have the results presented in Fig. 10.

According to the results of this clustering, Kyiv city and Odesa region form separate clusters according to the numbers "2" and "3". The next two clusters are also small: cluster "4" - Dnipropetrovsk and Kyiv regions, cluster "0" - Ivano-Frankivsk, Lviv and Kharkiv regions. The largest cluster "1" with a capacity of 18 is formed by other regions of Ukraine.

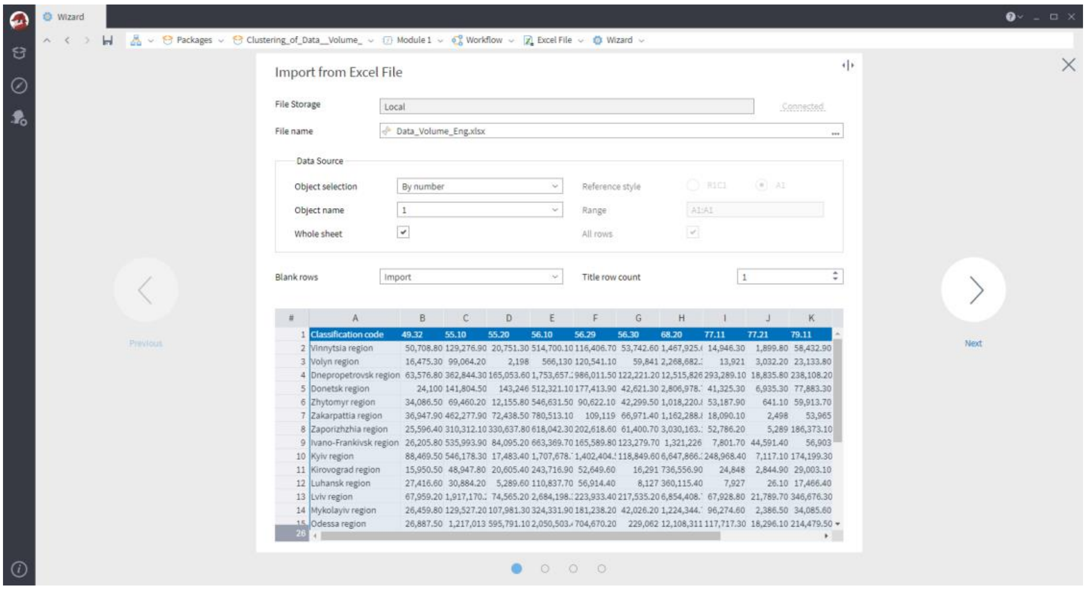


Figure 2: Data Import from Excel file in Loginom

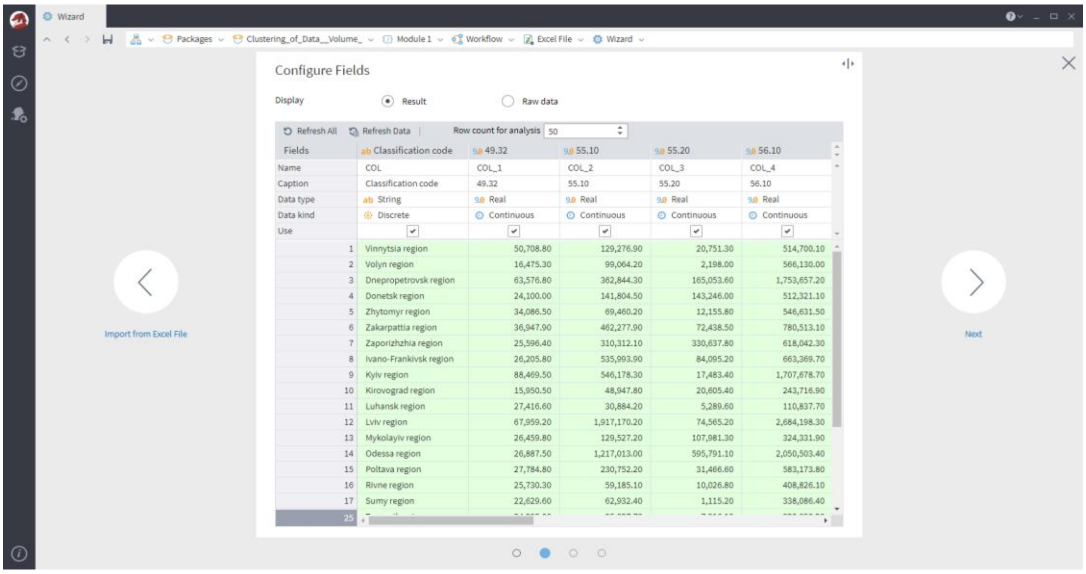


Figure 3: Configure fields in Loginom

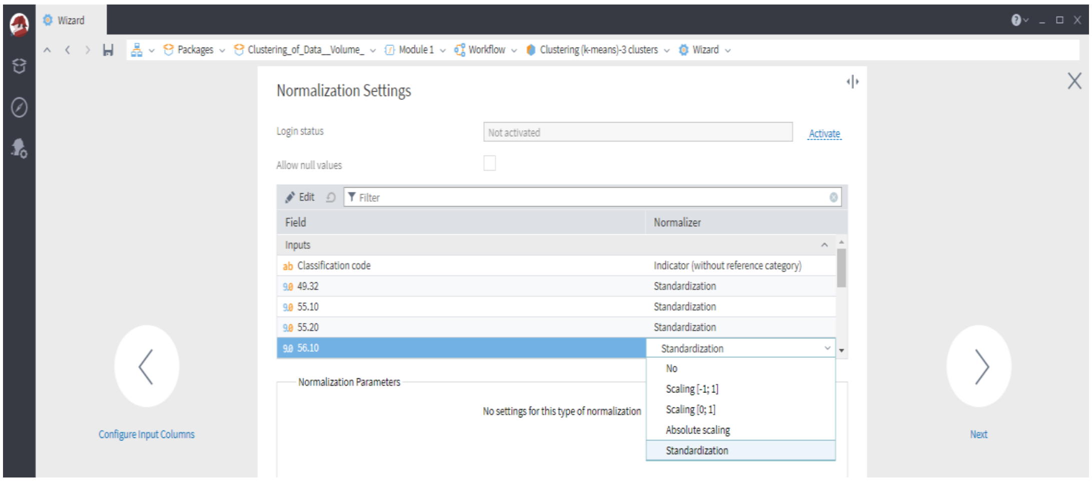


Figure 4: Normalization settings in Loginom

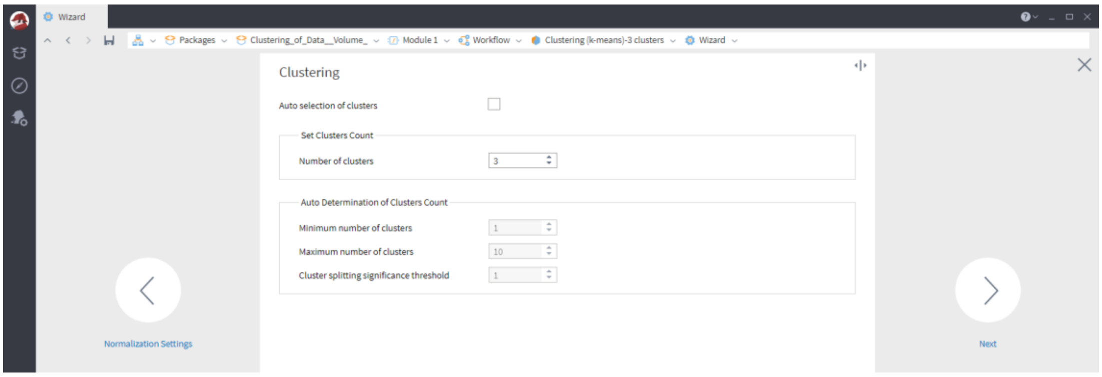


Figure 5: Determining the number of clusters in Loginom

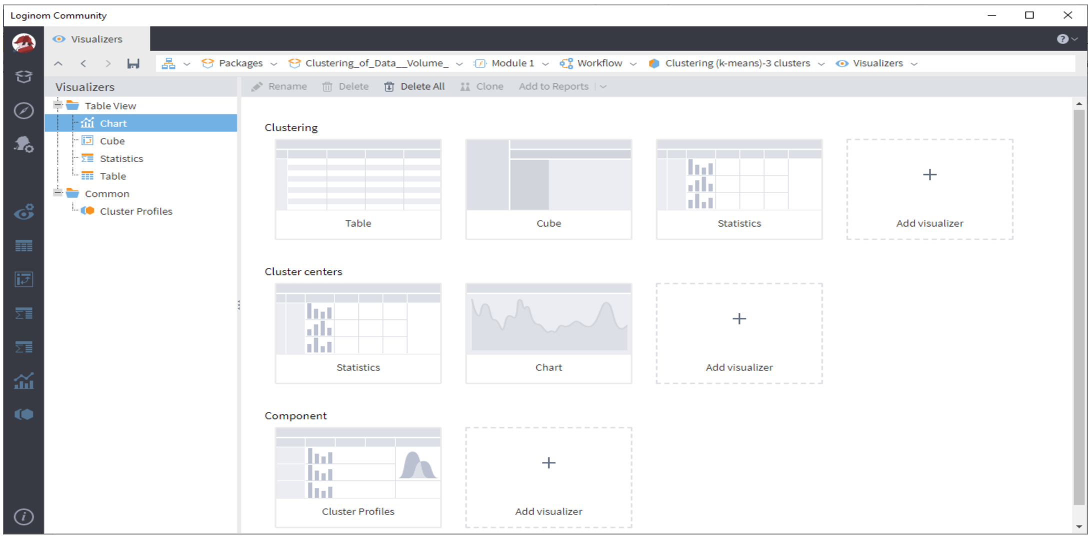
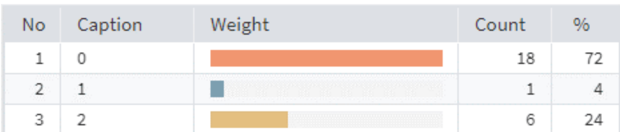
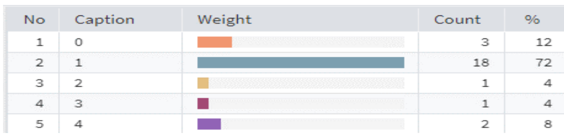


Figure 6: Visualizers settings in Loginom



No	Caption	Weight	Count	%
1	0		18	72
2	1		1	4
3	2		6	24

Figure 7: Power for 3 clusters



No	Caption	Weight	Count	%
1	0		3	12
2	1		18	72
3	2		1	4
4	3		1	4
5	4		2	8

Figure 8: Power for 5 clusters

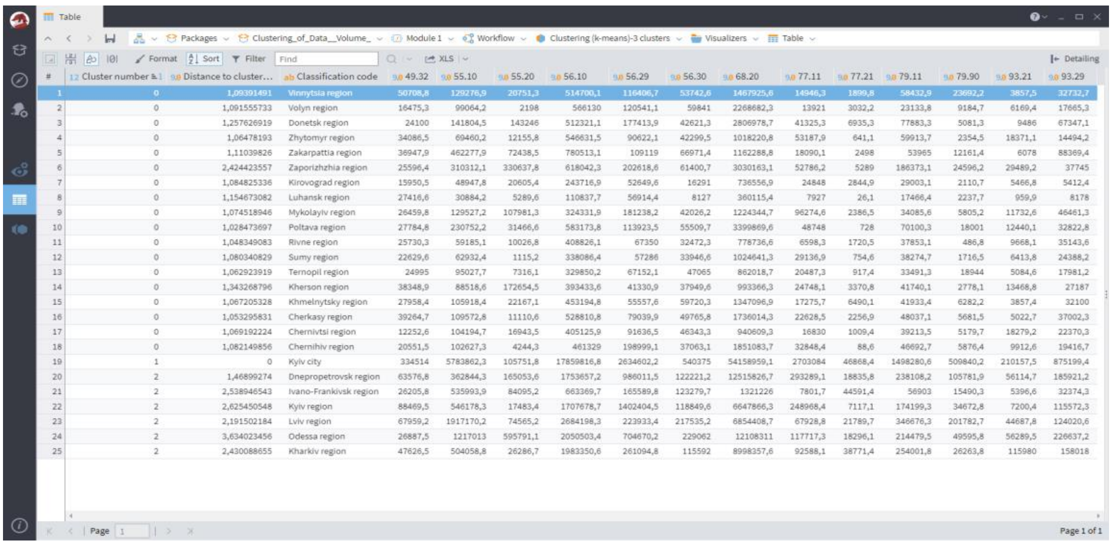
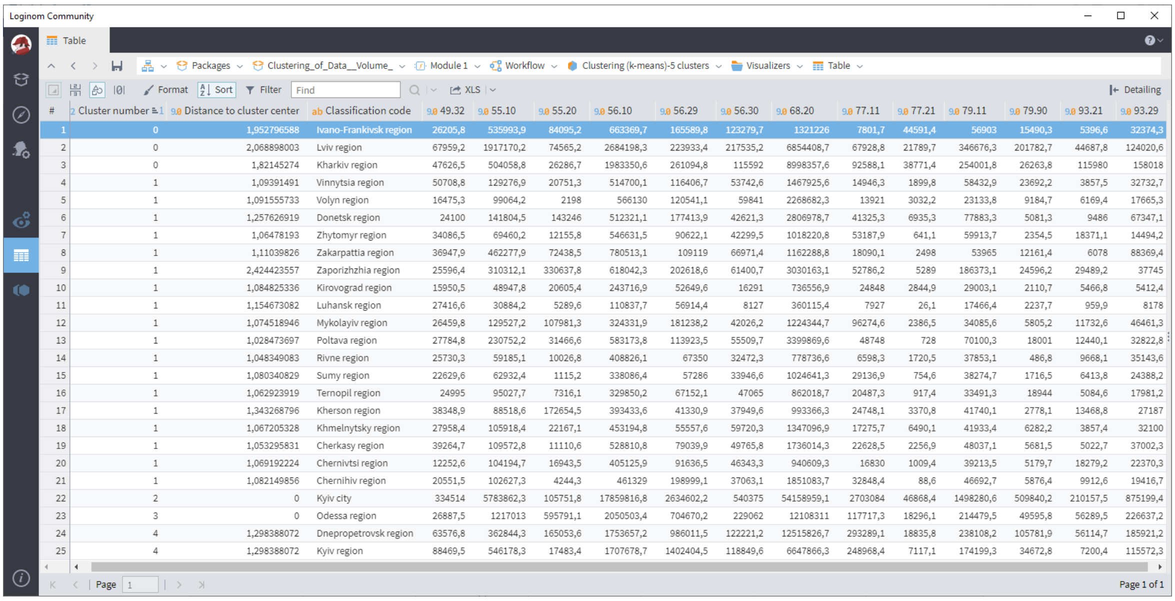


Figure 9: Clustering results for problem 3 clusters



#	Cluster number	Distance to cluster center	Classification code	49.32	55.10	55.20	56.10	56.29	56.30	68.20	77.11	77.21	79.11	79.90	93.21	93.29
1	0	1,952796588	Ivano-Frankivsk region	26205,8	535993,9	84095,2	663369,7	165589,8	123279,7	1321226	7801,7	44591,4	56903	15490,3	5396,6	32374,3
2	0	2,068898003	Lviv region	67959,2	1917170,2	74565,2	2684198,3	223933,4	217535,2	6854408,7	67928,8	21789,7	346676,3	201782,7	44687,8	124020,6
3	0	1,82145274	Kharkiv region	47626,5	504058,8	26286,7	1983350,6	261094,8	115592	8998357,6	92588,1	38771,4	254001,8	26263,8	115980	158018
4	1	1,09391491	Vinnytsia region	50708,8	129276,9	20751,3	514700,1	116406,7	53742,6	1467925,6	14946,3	1899,8	58432,9	23692,2	3857,5	32732,7
5	1	1,091555733	Volyn region	16475,3	99064,2	2198	566130	120541,1	59841	2268682,3	13921	3032,2	23133,8	9184,7	6169,4	17665,3
6	1	1,257626919	Donetsk region	24100	141804,5	143246	512321,1	177413,9	42621,3	2806978,7	41325,3	6935,3	77883,3	5081,3	9486	67347,1
7	1	1,06478193	Zhytomyr region	34086,5	69460,2	12155,8	546631,5	90622,1	42299,5	1018220,8	53187,9	641,1	59913,7	2354,5	18371,1	14494,2
8	1	1,11039826	Zakarpattia region	36947,9	462277,9	72438,5	780513,1	109119	66971,4	1162288,8	18090,1	2498	53965	12161,4	6078	88369,4
9	1	2,424423557	Zaporizhzhia region	25596,4	310312,1	330637,8	618042,3	202618,6	61400,7	3030163,1	52786,2	5289	186373,1	24596,2	29489,2	37745
10	1	1,084825336	Kirovograd region	15950,5	48947,8	20605,4	243716,9	52649,6	16291	736556,9	24848	2844,9	29003,1	2110,7	5466,8	5412,4
11	1	1,154673082	Luhansk region	27416,6	30884,2	5289,6	110837,7	56914,4	8127	360115,4	7927	26,1	17466,4	2237,7	959,9	8178
12	1	1,074518946	Mykolayiv region	26459,8	129527,2	107981,3	324331,9	181238,2	42026,2	1224344,7	96274,6	2386,5	34085,6	5805,2	11732,6	46461,3
13	1	1,028473697	Poltava region	27784,8	230752,2	31466,6	583173,8	113923,5	55509,7	3399869,6	48748	728	70100,3	18001	12440,1	32822,8
14	1	1,048349083	Rivne region	25730,3	59185,1	10026,8	408826,1	67350	32472,3	778736,6	6598,3	1720,5	37853,1	486,8	9668,1	35143,6
15	1	1,080340829	Sumy region	22629,6	62932,4	1115,2	338086,4	57286	33946,6	1024641,3	29136,9	754,6	38274,7	1716,5	6413,8	24388,2
16	1	1,062923919	Ternopil region	24995	95027,7	7316,1	329850,2	67152,1	47065	862018,7	20487,3	917,4	33491,3	18944	5084,6	17981,2
17	1	1,343268796	Kherson region	38348,9	88518,6	172654,5	393433,6	41330,9	37949,6	993366,3	24748,1	3370,8	41740,1	2778,1	13468,8	27187
18	1	1,067205328	Khmelnytsky region	27958,4	105918,4	22167,1	453194,8	55557,6	59720,3	1347096,9	17275,7	6490,1	41933,4	6282,2	3857,4	32100
19	1	1,053295831	Cherkasy region	39264,7	109572,8	11110,6	528810,8	79039,9	49765,8	1736014,3	22628,5	2256,9	48037,1	5681,5	5022,7	37002,3
20	1	1,069192224	Chernivtsi region	12252,6	104194,7	16943,5	405125,9	91636,5	46343,3	940609,3	16830	1009,4	39213,5	5179,7	18279,2	22370,3
21	1	1,082149856	Chernihiv region	20551,5	102627,3	4244,3	461329	198999,1	37063,1	1851083,7	32848,4	88,6	46692,7	5876,4	9912,6	19416,7
22	2	0	Kyiv city	334514	5783862,3	105751,8	17859816,8	2634602,2	540375	54158959,1	2703084	46868,4	1498280,6	509840,2	210157,5	875199,4
23	3	0	Odessa region	26887,5	1217013	595791,1	2050503,4	704670,2	229062	12108311	117717,3	18296,1	214479,5	49595,8	56289,5	226637,2
24	4	1,298388072	Dnepropetrovsk region	63576,8	362844,3	165053,6	1753657,2	986011,5	122221,2	12515826,7	293289,1	18835,8	238108,2	105781,9	56114,7	185921,2
25	4	1,298388072	Kyiv region	88469,5	546178,3	17483,4	1707678,7	1402404,5	118849,6	6647866,3	248968,4	7117,1	174199,3	34672,8	7200,4	115572,3

Figure 10: Clustering results for problem 5 clusters

Table 1 presents the average values of tourist services for each of the three clusters in terms of economic activities.

Table 1

Average values of tourist services by 3 clusters

	49.32	55.10	55.20	56.10	56.29	56.30	68.20	77.11	77.21	79.11	79.90	93.21	93.29
0	27,63	132,24	55,13	451,06	104,43	44,06	1500,48	30,14	2,38	52,09	8,45	9,76	31,49
1	334,51	5783,86	105,75	17859,82	2634,60	540,38	54158,96	2703,08	46,87	1498,28	509,84	210,16	875,20
2	53,45	847,21	160,55	1807,13	623,95	154,42	8074,33	138,05	24,90	214,06	72,26	47,61	140,42

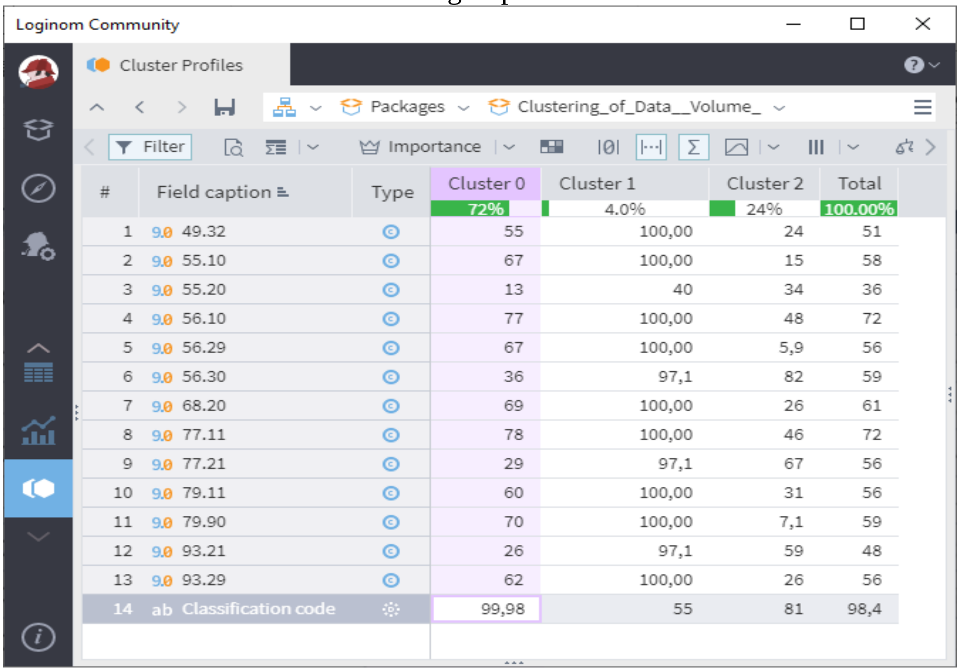
Source: formed by the authors.

For greater clarity, we present a comparative analysis in the form of diagrams of the distribution of average values of services in thousands of UAH by clusters (Fig. 11)

Figure 11: Diagrams of distribution of average values of volumes of services on three clusters

Analysis of diagrams in Fig. 11 shows the advantage of the first cluster (Kyiv - green) in the absolute majority of indicators of the volume of tourist services. The only exception is the indicator with the code 55.20 "Accommodation activities for the period of vacation and other temporary residences", for which cluster "2" has an advantage.

Loginom allows you to quantify the degree of importance of clusters for each type of service provided (Fig. 12) and present the contribution of each cluster in the total volume of provided tourist services of a particular type (Fig. 13). In the Cluster profiles visualizer, you can see the general structure of the formed clusters. It reflects all the considered indicators together with the nature of their impact on the composition of the cluster. The main factor determining the composition of the cluster is the importance of the properties, expressed as a percentage. The total importance of the considered field is determined by the variability of its considered parameters. The importance of continuous and discrete fields is defined differently. For continuous fields, it is set depending on the deviation of the average value of this group of clusters from the total average of the entire sample. The more pronounced this deviation, the greater its significance. The importance of discrete fields is determined by the presence of individual differences between the groups.



#	Field caption	Type	Cluster 0 (72%)	Cluster 1 (4.0%)	Cluster 2 (24%)	Total (100.00%)
1	49.32		55	100,00	24	51
2	55.10		67	100,00	15	58
3	55.20		13	40	34	36
4	56.10		77	100,00	48	72
5	56.29		67	100,00	5,9	56
6	56.30		36	97,1	82	59
7	68.20		69	100,00	26	61
8	77.11		78	100,00	46	72
9	77.21		29	97,1	67	56
10	79.11		60	100,00	31	56
11	79.90		70	100,00	7,1	59
12	93.21		26	97,1	59	48
13	93.29		62	100,00	26	56
14	Classification code		99,98	55	81	98,4

Figure 12: Visualizer *Cluster profiles*

Table 2 presents the average values of tourist services for each of the clusters in terms of economic activities in the division of the regions of Ukraine into 5 clusters.

Table 2

Average values of tourist services by 5 clusters

	49.32	55.10	55.20	56.10	56.29	56.30	68.20	77.11	77.21	79.11	79.90	93.21	93.29
0	47,26	985,74	61,65	1776,97	216,87	152,14	5724,66	56,11	35,05	219,19	81,18	55,35	104,80
1	27,63	132,24	55,13	451,06	104,43	44,06	1500,48	30,14	2,38	52,09	8,45	9,76	31,49
2	334,51	5783,86	105,75	17859,82	2634,60	540,38	54158,96	2703,08	46,87	1498,28	509,84	210,16	875,20
3	26,89	1217,01	595,79	2050,50	704,67	229,06	12108,31	117,72	18,30	214,48	49,60	56,29	226,64
4	76,02	454,51	91,27	1730,67	1194,21	120,54	9581,85	271,13	12,98	206,15	70,23	31,66	150,75

Source: formed by the authors.

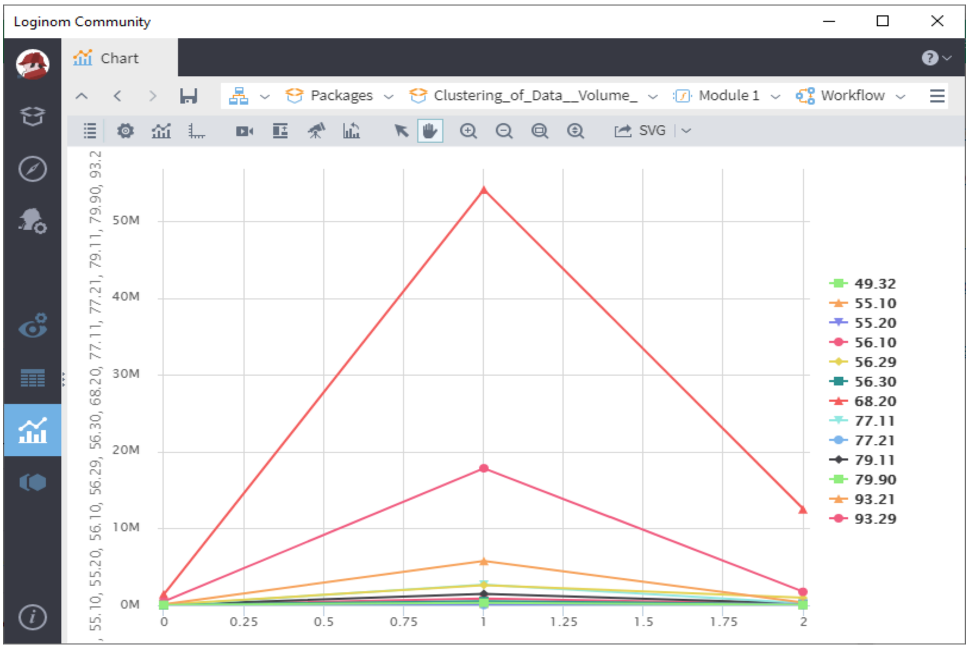


Figure 13: The contribution of each of the three clusters to the total volume of tourist services provided

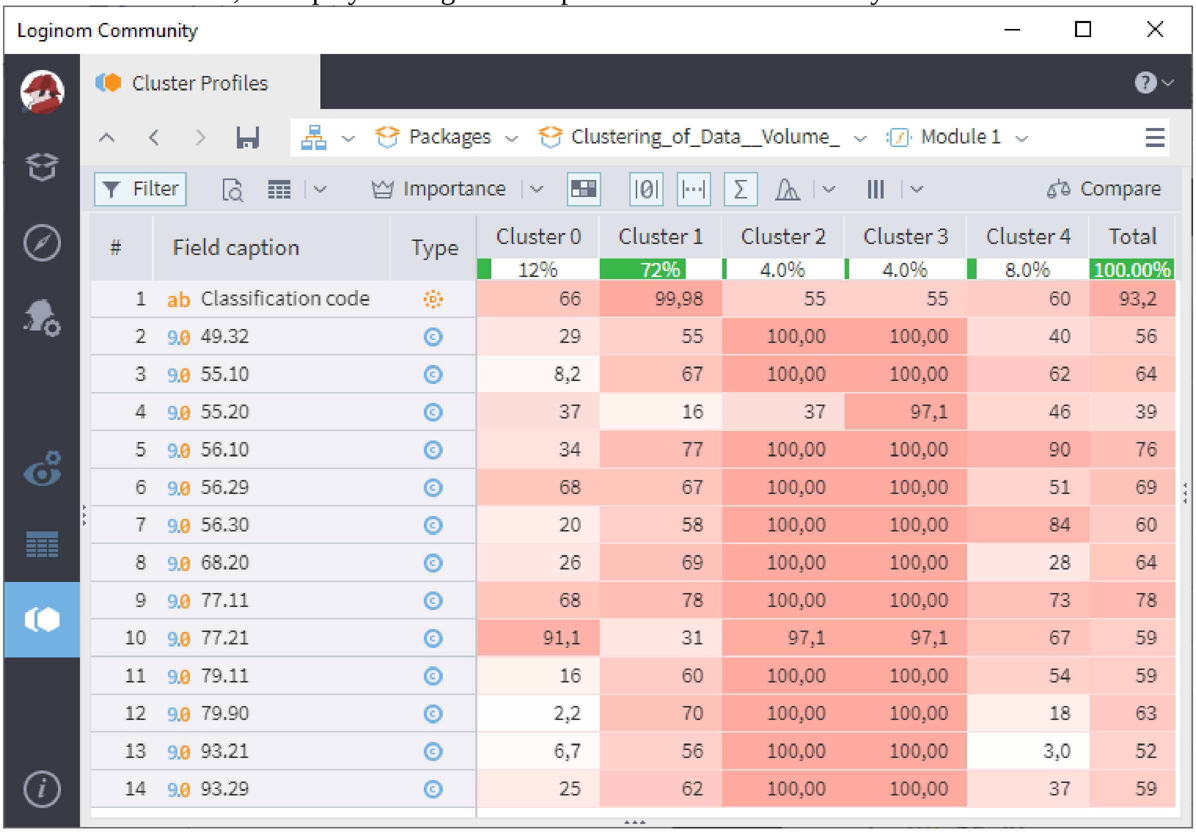
Diagrams of distribution of average values of volumes of services in thousand UAH for five clusters are presented in Fig. 14.

Figure 14: Diagrams of distribution of average values of volumes of services on five clusters

The distribution shown in these diagrams is structurally similar to the case of the three clusters. The Kyiv cluster wins in all services, except 55.22, where the leadership belongs to the blue cluster of Odesa. By the way, Odessa was part of the winning cluster "2" on this indicator and in the previous version of clustering. This similarity indicates the stability of clusters in detail. However, we also have differences. For example, for the service 56.29 "Supply of other ready meals" the cluster "4" (Dnipropetrovsk and

Kyiv regions) was in the second position, and for the service with code 77.21 "Rental of goods for sports and recreation" - the cluster "0" (Ivano- Frankivsk, Lviv and Kharkiv regions), not significantly inferior to Kyiv.

In fig. 15 you can see the distribution of the five clusters by importance. Loginom allows, in addition to numerical values, to display the degree of importance of colour intensity.



#	Field caption	Type	Cluster 0 (12%)	Cluster 1 (72%)	Cluster 2 (4.0%)	Cluster 3 (4.0%)	Cluster 4 (8.0%)	Total (100.00%)
1	ab Classification code		66	99,98	55	55	60	93,2
2	9.0 49.32		29	55	100,00	100,00	40	56
3	9.0 55.10		8,2	67	100,00	100,00	62	64
4	9.0 55.20		37	16	37	97,1	46	39
5	9.0 56.10		34	77	100,00	100,00	90	76
6	9.0 56.29		68	67	100,00	100,00	51	69
7	9.0 56.30		20	58	100,00	100,00	84	60
8	9.0 68.20		26	69	100,00	100,00	28	64
9	9.0 77.11		68	78	100,00	100,00	73	78
10	9.0 77.21		91,1	31	97,1	97,1	67	59
11	9.0 79.11		16	60	100,00	100,00	54	59
12	9.0 79.90		2,2	70	100,00	100,00	18	63
13	9.0 93.21		6,7	56	100,00	100,00	3,0	52
14	9.0 93.29		25	62	100,00	100,00	37	59

Figure 15: Cluster profiles visualizer for the case of 5 clusters

In fig. 16 shows diagrams of the distribution of the contribution of clusters in terms of different types of tourist services.

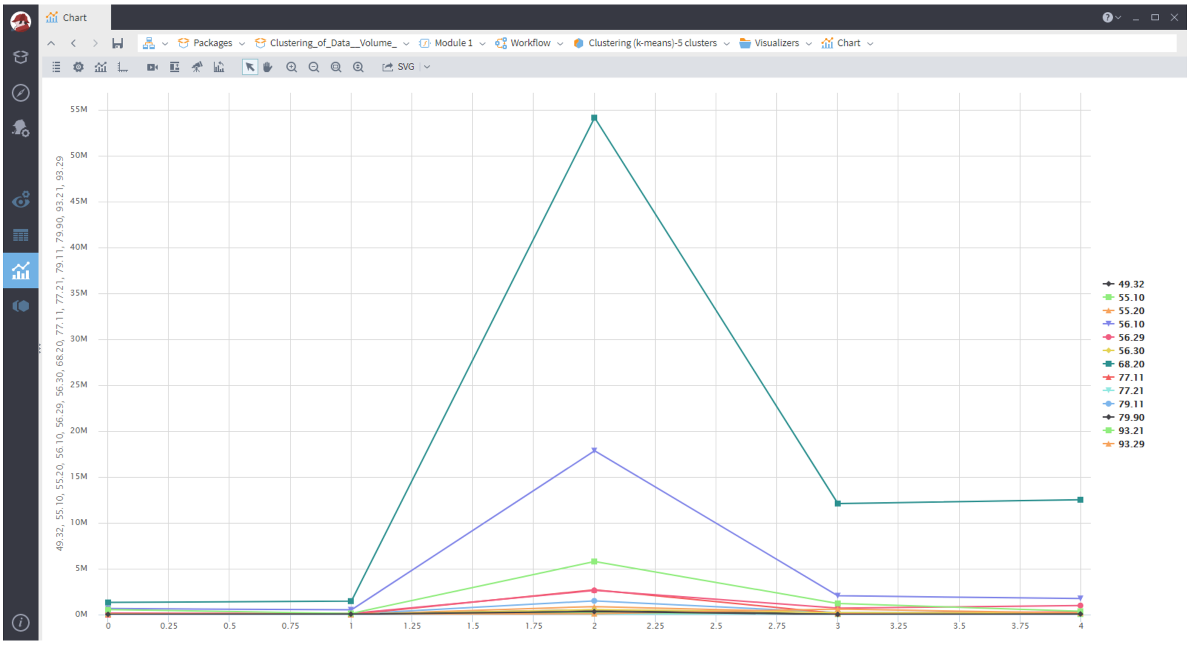
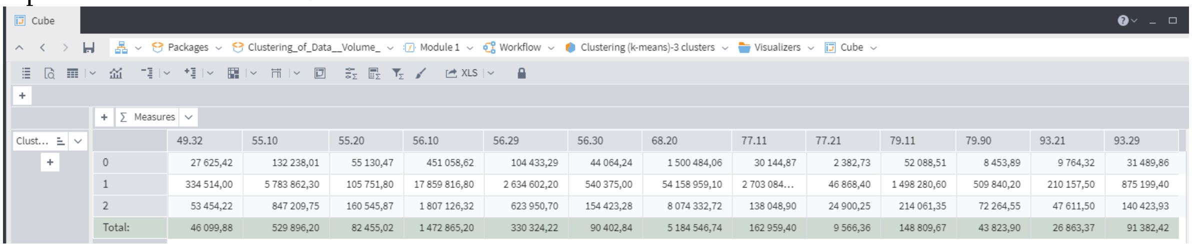


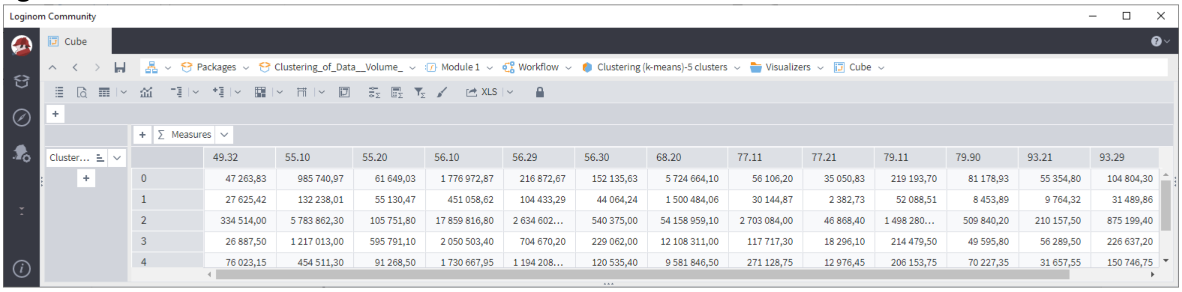
Figure 16: The contribution of each of the five clusters to the total amount of tourist services provided

Fig. 17 and 18 show the «Cube» visualizer, which is one of the common methods of complex multidimensional analysis of OLAP (OnLine Analytical Processing). It is based on the representation of data in the form of multidimensional cubes (OLAP-cubes). A cube is a convenient tool for visualizing multidimensional data and obtaining the necessary report forms. It contains measurements and facts that are determined during construction. The main feature of the cube is that its structure is not rigidly defined. By manipulating the measurement headers, the user can achieve the most informative representation of the cube.



Clust…	49.32	55.10	55.20	56.10	56.29	56.30	68.20	77.11	77.21	79.11	79.90	93.21	93.29
0	27 625,42	132 238,01	55 130,47	451 058,62	104 433,29	44 064,24	1 500 484,06	30 144,87	2 382,73	52 088,51	8 453,89	9 764,32	31 489,86
1	334 514,00	5 783 862,30	105 751,80	17 859 816,80	2 634 602,20	540 375,00	54 158 959,10	2 703 084…	46 868,40	1 498 280,60	509 840,20	210 157,50	875 199,40
2	53 454,22	847 209,75	160 545,87	1 807 126,32	623 950,70	154 423,28	8 074 332,72	138 048,90	24 900,25	214 061,35	72 264,55	47 611,50	140 423,93
Total:	46 099,88	529 896,20	82 455,02	1 472 865,20	330 324,22	90 402,84	5 184 546,74	162 959,40	9 566,36	148 809,67	43 823,90	26 863,37	91 382,42

Figure 17: OLAP-cube visualizer for the case of three clusters

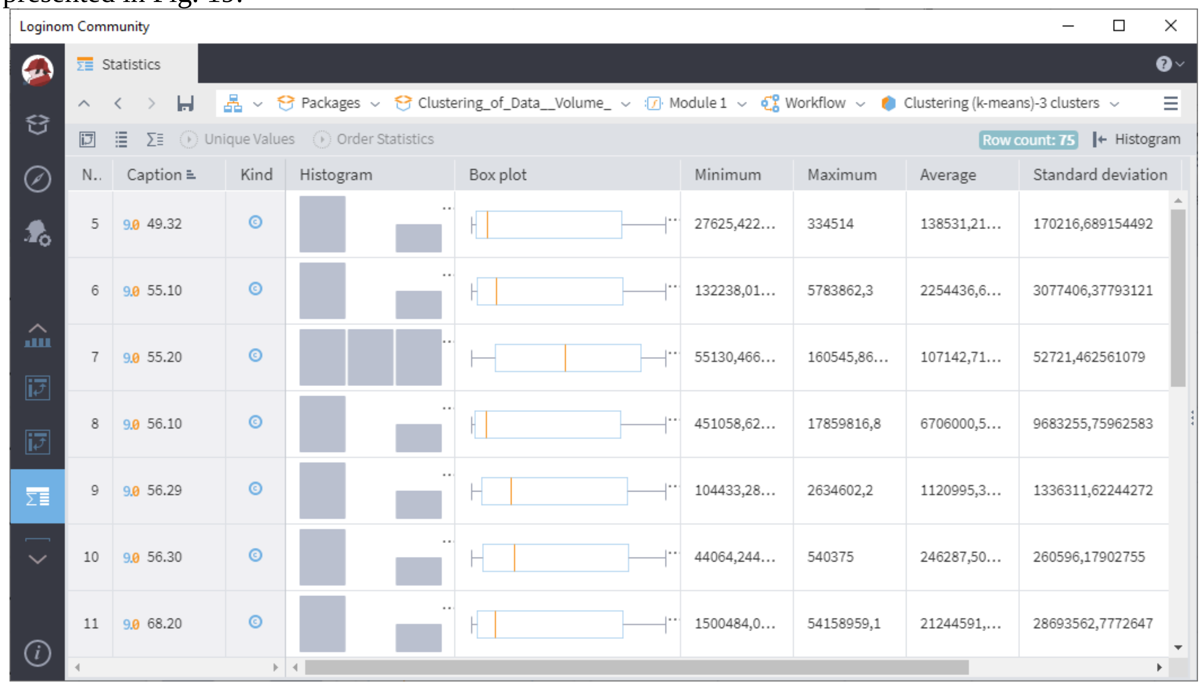


Cluster…	49.32	55.10	55.20	56.10	56.29	56.30	68.20	77.11	77.21	79.11	79.90	93.21	93.29
0	47 263,83	985 740,97	61 649,03	1 776 972,87	216 872,67	152 135,63	5 724 664,10	56 106,20	35 050,83	219 193,70	81 178,93	55 354,80	104 804,30
1	27 625,42	132 238,01	55 130,47	451 058,62	104 433,29	44 064,24	1 500 484,06	30 144,87	2 382,73	52 088,51	8 453,89	9 764,32	31 489,86
2	334 514,00	5 783 862,30	105 751,80	17 859 816,80	2 634 602…	540 375,00	54 158 959,10	2 703 084,00	46 868,40	1 498 280…	509 840,20	210 157,50	875 199,40
3	26 887,50	1 217 013,00	595 791,10	2 050 503,40	704 670,20	229 062,00	12 108 311,00	117 717,30	18 296,10	214 479,50	49 595,80	56 289,50	226 637,20
4	76 023,15	454 511,30	91 268,50	1 730 667,95	1 194 208…	120 535,40	9 581 846,50	271 128,75	12 976,45	206 153,75	70 227,35	31 657,55	150 746,75

Figure 18: OLAP-cube visualizer for the case of five clusters

In the problem of clustering of regions of Ukraine OLAP-cube allows in a convenient and compact form to present the results of clustering and the corresponding average values of services by each type and each cluster.

For a more in-depth analysis of the results of clustering use the visualizer «Statistics». For each considered property in the cluster is calculated: confidence interval, mean, standard deviation and standard error. For the problem with three clusters, the statistical characteristics of clustering are presented in Fig. 19.



N..	Caption	Minimum	Maximum	Average	Standard deviation
5	49.32	27625,422…	334514	138531,21…	170216,689154492
6	55.10	132238,01…	5783862,3	2254436,6…	3077406,37793121
7	55.20	55130,466…	160545,86…	107142,71…	52721,462561079
8	56.10	451058,62…	17859816,8	6706000,5…	9683255,75962583
9	56.29	104433,28…	2634602,2	1120995,3…	1336311,62244272
10	56.30	44064,244…	540375	246287,50…	260596,17902755
11	68.20	1500484,0…	54158959,1	21244591,…	28693562,7772647

Figure 19: Statistics visualizer for a problem with three clusters

The Statistics visualizer is designed to view different statistics for each field in a data set and is a table in which the names of the fields in the data set are arranged in rows and the names of the statistical indicators are arranged in columns. At the intersection, in the cells of the table, are the values of statistical indicators of the respective fields. The "Statistics" visualizer for the clustering problem in the form of five clusters is shown in Fig. 20.

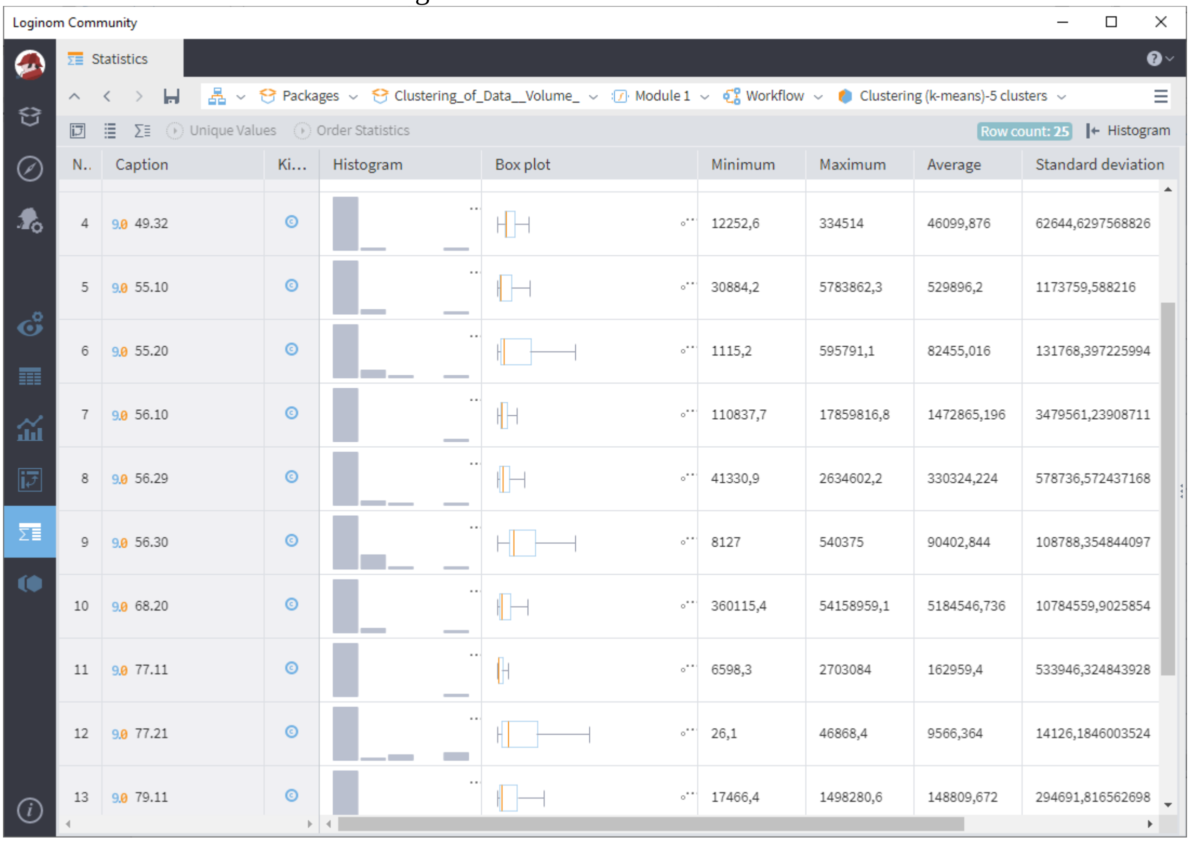


Figure 20: Statistics visualizer for a problem with five clusters

In addition to statistical characteristics and histograms, which are analogues of the diagrams shown in Fig. 11 and 14, note the scale diagrams - Box plot, which is an excellent graphical tool for a compact presentation of information on the distribution values within a single indicator of the volume of tourist services provided.

4.3. Interpretation of clustering results

Data clustering allows you to solve such important problems of Data Science:

- *Data study*. Dividing a set of objects into similar groups helps to identify the structure of data, increase the clarity of their representation, put forward new hypotheses, understand how informative the properties of objects are.
- *Facilitate analysis*. With the help of clustering, you can simplify further data processing and model construction: each cluster is processed individually and the model is created for each cluster separately.
- *Data compression*. When the data is large (hundreds of thousands and millions of rows), clustering reduces the amount of data stored, leaving one most typical representative from each cluster.
- *Forecasting*. Clusters are used not only to briefly describe existing objects but also to identify new ones. Each new object belongs to the cluster to which the join best satisfies the clustering quality criterion.
- *Detection of anomalies*. Clustering is used to select atypical objects that do not join any of the clusters

Thus, fig. 21 presents a map of Ukraine with the results of clustering of the first type.

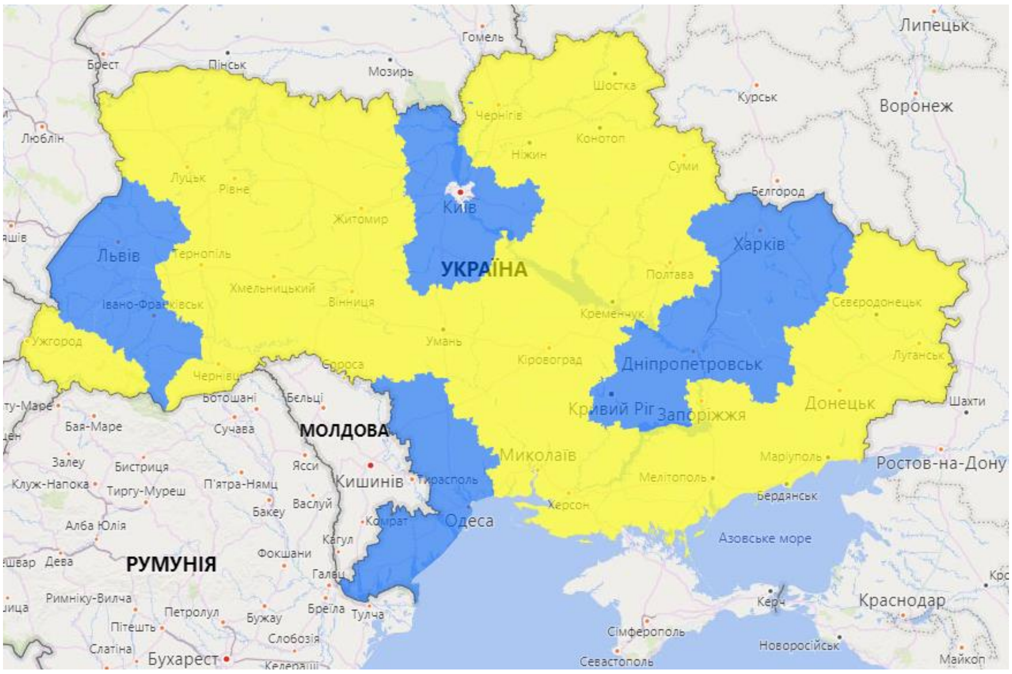


Figure 21: Results of clustering of regions of Ukraine into 3 clusters

The solution to the problem of clustering the regions of Ukraine into three clusters according to the level of investment attractiveness predictably separated the capital of Ukraine into a separate cluster. Kyiv city is not only the largest city in the country but also a major tourist centre. Cluster number 2 includes Dnipropetrovsk, Ivano-Frankivsk, Kyiv, Lviv, Odesa and Kharkiv regions. This cluster ranks second in the ranking. It includes four regions corresponding to the largest cities of Ukraine, as well as Prykarpattia, which is one of the prominent tourist centres of the country (the highest peak in Ukraine - Hoverla, the world-famous resort "Bukovel", etc.).

Fig. 22 presents a map of Ukraine with the results of clustering for a problem of five clusters.

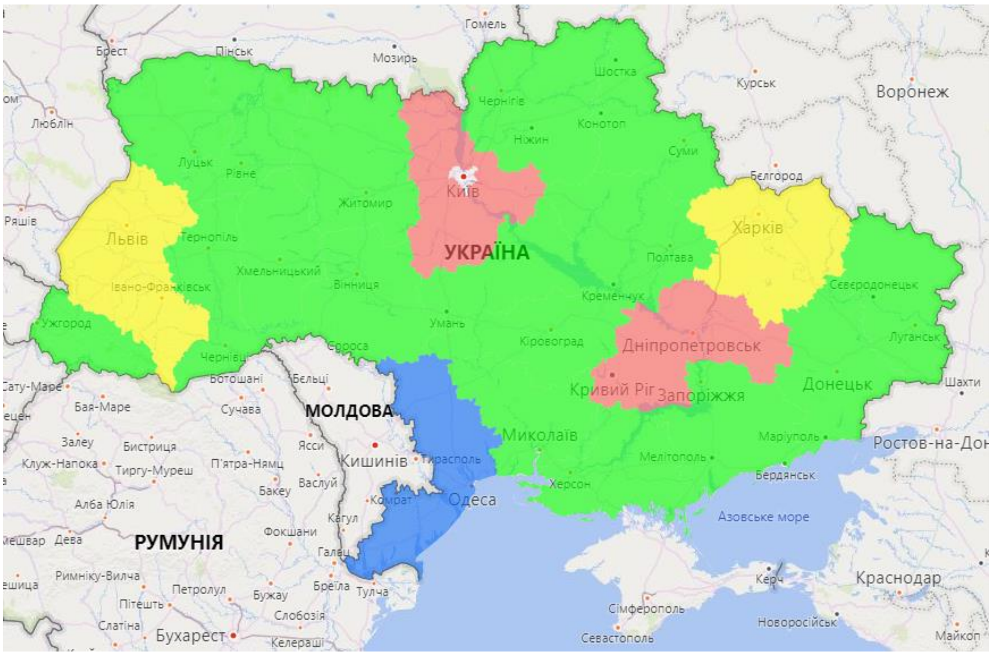


Figure 22: Results of clustering of regions of Ukraine into 5 clusters

Kyiv city is again in a separate cluster, which indicates a stable position as the leader of the tourism industry of Ukraine. Also, a separate cluster is the Odesa region, which after the annexation of Crimea has become an undisputed favourite of maritime tourism for Ukrainians. The increase in the number of clusters has not changed the general trend towards the allocation of levels of investment attractiveness of regions in the field of tourism services. The allocation of Odesa region, Ivano-Frankivsk, Lviv and Kharkiv regions, Dnipropetrovsk and Kyiv regions into separate clusters are the result of high performance, which demonstrates these regions for certain types of tourist services, which can be seen in the diagrams of Figs. 14.

The rest of the regions are inferior to the level of the considered clusters and to attract investors in the field of tourism in the future should work on the full disclosure of their tourism potential.

5. Conclusion

The authors of the article based on the implementation of the k-means algorithm in the analytical platform Loginom obtained the results of clustering the regions of Ukraine by the level of investment attractiveness in the field of tourism.

Loginom system has powerful tools for cluster analysis by EM-Clustering, k-means, g-means and others, statistical and visual analysis of the results. Clustering has made it possible to identify groups of regions that are actively developing the tourism industry and are currently formed for tourism investors. Equally important is the selection of problem regions that have a low level of attractiveness for domestic and foreign tourism.

Ukraine has a big potential for the development of the tourism industry. The regions that, according to the results of the cluster analysis, found themselves in the problem group have world-class "tourist pearls". Ostrog, Baturyn, Velyki Sorochyntsi, Khotyn, Kamyanets-Podilsky, Khust, Chersonesos, Khortytsia - this list of sights can be extended and supplemented.

The Government of Ukraine and local authorities should pay attention to the insufficient level of development of the tourism industry, provide comprehensive support to regions that are in a problem cluster, and thus increase the level of their investment attractiveness.

6. References

[1] Tourism barometer of Ukraine // National Tourism Organization of Ukraine, 2020. URL: http://www.ntoukraine.org/assets/files/ntou-barometer-2020.pdf.

[2] O. Prokopenko, Y. Larina, O. Chetveryk, S. Kravtsov, N. Rozhko, I. Lorvi Digital-toolkit for promoting tourist destinations. International Journal of Innovative Technology and Exploring Engineering, 8(12), (2019): 4982-4987. doi: 10.35940/ijitee.L3745.1081219

[3] A. Roskladka, N. Roskladka, G. Kharlamova, A. Karpuk, A. Stavytskyy The data science tools for research of emigration processes in Ukraine. Problems and Perspectives in Management, 18 (1) (2020): 70-81.

[4] A. Roskladka, N. Roskladka, O. Dluhopolskyi, G. Kharlamova, M. Kiziloglu Data analysis and forecasting of tourism development in Ukraine. Innovative Marketing, 14 (4) (2018): 19-33.

[5] N. Roskladka, A. Roskladka Computer modeling of tourism flows in Ukraine. Ukraine and the World: the tourism system. Prague : Eatern European Center of the Fundamental Researchers (EECFR), 2019, pp. 41-55.

[6] N. Sagalakova, A. Roskladka Innovative approaches to the researching of variability of the tourism industry indicators. Studia i Materiały, 2 (10) (2015): 75-82.

[7] H. Cuesta, S. Kumar Practical Data Analysis.Birmingham, Packt Publishing Ltd, 2016.

[8] Data Science & Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data, EMC Education Services. Indianapolis, John Wiley & Sons, Inc, 2015.

[9] C. Iyigun, M. Türkeş, I. Batmaz, C. Yozgatligil, V. Purutçuoğlu, E. Kartal, M. Öztürk Clustering current climate regions of Turkey by using a multivariate statistical method. Theoretical and Applied Climatology, 114 (2013): 95-106.

[10] K. Sablyn, E. Kahan, E. Chernova Clustering of coal mining regions of Russia: investment and innovation activity. Journal of New Economy, 21 (1) (2020): 89-106.

[11] K. Gorbatiuk, O. Mantalyuk, O. Proskurovych, O.Valkov Application of Fuzzy Clustering to Shaping Regional Development Strategies in Ukraine, Proceedings of the 6th International

Conference on Strategies, Models and Technologies of Economic Systems Management (SMTESM 2019), 2019, pp. 271-276.
[12] V. Mazur, K. Barmuta, S. Demin, E. Tikhomirov, M. Bykovskiy Innovation Clusters: Advantages and Disadvantages, International Journal of Economics and Financial Issues, 6 (2016): 270-274.
[13] R. Njos, S.-E. Jakobsen Cluster policy and regional development: scale, scope and renewal. Regional Studies, Regional Science, 3 (1) (2016): 146-169.
[14] Y. Zolotareva, Y. Pavlenko Deductor analytical platform in assessing the degree of environmental safety of the region. Business analytics. Questions of theory and practice. Use of the analytical platform Deductor in the activities of educational institutions, Ryazan, Database laboratory (2010): 89-97.
[15] T. Paianok, Y. Vazhaliuk Cluster analysis of labor potential of Ukraine. Economy and State, 12 (2019): 109-114.
[16] S. Behun Application of cluster analysis to study the demographic situation in the region. Economic Journal of Lesya Ukrainka East European National University, 2 (2016): 122-128.
[17] I. Opmane Use of Cluster Analysis in Exploring Economic Indicator Differences among Regions: The Case of Latvia. Journal of Economics, Business and Management, 1 (1) (2013): 42-45.
[18] S. Synytsia, O. Vakun Clustering of regions by level of economic potential. Economy and society Mukachevo State University, 12 (2017): 776-784.
[19] N. Potapova Clustering of economic regions of Ukraine in terms of innovation and research. Lviv Polytechnic National University Institutional Repository (2010): 33-39.
[20] O. Riadno, O. Berkut Research of structure and dynamics of differentiation of social and economic development of regions of Ukraine on the basis of the cluster analysis. Economic bulletin of Donbass, 1(43) (2016): 60-67.
[21] L. Zomchak, Y. Dobrotii Clustering of regions of Ukraine by competitiveness. Proceedings of the International scientific-practical conference «Administrative-territorial vs economic-spatial borders of regions», KNEU, 2020, pp. 328-332.
[22] A. Mazaraki, M. Bosovska, O. Hladkyi, S. Kravtsov, H. Mykhailichenko, Y. Zabaldina Tourism clusters, Kyiv, KNUTE, 2018.
[23] R. Mudrak, N. Moisieieva Clustering as a direction of effective management of tourist destinations of the region. Ukrainian Journal of Applied Economics. 3 (3) (2018): 126-132.
[24] Loginom. URL:https://loginom.com.
[25] V. Yakovlev Data analysis in the Loginom analytical platform, LAP LAMBERT Academic Publishing, 2020.
[26] Statistical bulletin «Business activities – 2019», Kyiv, State Statistics Service of Ukraine, 2020.
[27] Statistical bulletin «Collective accommodation facilities in Ukraine (legal entities, separate subdivisions of legal entities) – 2019», Kyiv, State Statistics Service of Ukraine, 2020.
[28] Statistical bulletin «Transport of Ukraine – 2019», Kyiv, State Statistics Service of Ukraine, 2020.