Clustering of Municipal Self-Government Bodies on the Basis of Statistical Indicators

Rakhimboev Khikmat Jumanazarovich, Ismailov Mirkhalil Agzamovich, Khusinov Khamid Khudayberganovich

Abstract*: The application of cluster analysis by the method of k – means local self-government bodies of the Republic of Uzbekistan is examined from the standpoint of the state of self-government bodies, as well as the level and quality of life of the population. Based on statistical information, the paper presents the results of the division of regions according to key indicators of self-government bodies. In the course of the study, five clusters were identified: with a low, below average, average, above average, and a high state of development of self-government. To characterize the self-governing bodies of the Khorezm region, statistics were used on self-governing bodies (population) and some system of socio-economic indicators of the level and quality of life (average per capita total income of the population, average per capita real income of the population, number of crimes and unemployment, facilities and municipal facilities , living area per inhabitant, etc.).

***Keywords* : level and quality of life, classification of subjects, linear transformation, multivariate statistical analysis, k-means clustering algorithm, hierarchical cluster analysis.**

I. INTRODUCTION

Nowadays, in modern economic conditions, the socio-economic development of regional and district administrative regions, which at one time consists of local municipal self-government bodies (mahalla), is becoming increasingly relevant. According to most experts, it is the solution of socio-economic problems at the level of the district and local governments that largely determines the development of states. The key task at the present stage of development facing public authorities of any level is to ensure a decent standard and quality of life of the population [1]. The level of development of self-government bodies is closely related to the standard of living and the quality of life of the population of the self-government body. The standard of living is determined by the conditions of human existence in the sphere of consumption and is measured through socio-economic indicators of the General welfare of people (for example, income, consumption, etc.).

Revised Manuscript Received on December 15, 2019.

Rakhimboev Khikmat Jumanazarovich, Senior Lecturer, Department of Information Technology, Urgench Branch of Tashkent University of Information Technologies, Uzbekistan.

Ismailov Mirkhalil Agzamovich, Chief Researcher, Research and Innovation Center information and communication technologies, Tashkent University of information technologies, Uzbekistan.

Khusinov Khamid Khudayberganovich, Lecturer, Urgench branch of Tashkent University of Information Technologyies, Uzbekistan.

Quality of life is a broader set of living conditions and includes the standard of living, as well as such components that relate to the territory of residence. The performance of self-government bodies is reflected in the passport of the mahalla and the economic book of the mahalla.

The methodological and theoretical basis of the study was the works of leading foreign scientists on the study of the level and quality of life, regional economy, socio-economic statistics, multivariate statistical methods.

Today, almost all territories are characterized by problems of spatial inequality and socio-economic asymmetry. Territories have differences in levels of economic development, population size and employment, income and quality of life, infrastructure, etc. That is why minimizing inequalities that impede the sustainable development of territories, regions or the country as a whole is one of the most important tasks of the state social and economic policy.

In order to monitor the dynamics of the development of municipal regions, as well as the quality and standard of living of the population in the regions, a quantitative assessment of the relevant indicators is needed, and in order to develop competent management decisions, it is necessary to understand and evaluate the situation of one administrative region (self-government body, region, oblast, country) relative to others regions (self-government body, district, region, country) [2]. The solution to this issue implies their division into groups, i.e. classification of regions (countries) according to one or another sign.

Cluster analysis, which takes into account their existing group and individual differences, helps to solve the problem of classifying self-government bodies. Cluster analysis allows you to break down a set of organs into groups and simultaneously determine the number of such groups. The groups into which the sample is divided are called clusters. The paper used the cluster analysis method.

The performance of the two proposed algorithms with different parameter values is compared in the source [4]. The two proposed algorithms are compared with the two existing algorithms. In addition, to reduce the computational time, the corresponding parallel versions of the two proposed algorithms are developed.

Cluster analysis is an important method in many applications such as biology, medicine, psychology, pattern recognition, image processing, marketing, and data development.

A large number of existing clustering algorithms can be divided into two types: (1) hierarchical and (2) partitioned. Depending on the adopted algorithmic approach, the hierarchical structure starts with N clusters, one per template,

and grows the clustering sequence until all N templates are in the same cluster (agglomerative approach), or starts with a single cluster containing all N templates, and sequentially divides the clusters until N clusters are reached (separation approach) [4].

The source [5] made an analysis and comparison of the demographic situation in the NUTS2 regions from 28 EU countries based on various demographic criteria. The data is primarily the subject of analysis of the main components in order to identify possible correlations between them. Only those indicators are used that explain a substantial part of the variations. Consequently, cluster analysis is used to group states with a similar situation in a particular year (2004, 2012). Hierarchical clustering using the Ward method with square Euclidean distances grouped the regions into 5 groups. This allows you to highlight those who have the potential for favorable development, because they have a high birth rate and the number of live births. Regions were also identified (in cluster 5), where the situation in 2004 was better than in 2012, as their potential worsened [5].

In the source [6] the determination of the optimal number of clusters is investigated. Cluster analysis is a multidimensional method that aims to classify objects. In modern literature, there are many methods and measures of distance that can be mutually combined. There is no manual or rule that clearly defines the appropriate combined method and distance measures during clustering. At the same time, in cluster analysis, it is often necessary to determine the optimal number of clusters to which objects should be assigned. The aim of this paper is to illustrate the possibilities of the cluster number determination process and to evaluate the selected coefficients for determining the number of clusters in combination with clustering by different methods and with different distance measures. For example, the CHF coefficient is more suitable for use in conjunction with the Mahalanobis distance, where success is higher compared to the Euclidean distance. For example, when using the average layout method, the success rate is 21.88% higher. On the other hand, the d-B coefficient is more successful when using Euclidean distance measures. In the case of ward's method, success is higher by 15.63% [6

II. DATA AND METHODS

The initial information for cluster analysis is obtained from the Khorezm regional Council for coordination of activities of self-government bodies of citizens, official statistical collections of the state statistics Committee of the Republic of Uzbekistan and other official sources [8, 9, 10]. The self-government bodies of the Khorezm region act as observation units, i.e. the source data array contains information on 445 self-government bodies of the Khorezm region. The characteristic of self-government bodies is determined by the system of socio-economic indicators for each. In table 1 the system of indicators which was applied for clustering of self-government bodies on a situation of development, and also on a level and quality of life is presented. KNIME Analytics Platform package is used for processing and analysis of statistical information.

Table 1. The indicators system assessing the state of self-government bodies and the level of socio-economic development.

Group of indicators	Designation and name of indicators
Names	x1 - Names of the self-government body; x2 - district;
Demographic situation	x3 - number of households; x4 is the number of families; x5 is the resident population; x6 is the number of men;
	x7 is the number of women; x8 - children, adolescents and youth (up to 30 years old); x9 - children under 7 years old; x10 - children and adolescents from 7 to 16 years old; x11 - minors from 16 to 18 years old; x12 - youth from 18 to 30 years old; x13 - the number of families who have lost the bread-winner; x14 is the number of single elderly people; x15 - the number of people with disabilities; x16 - the number of pensioners;
Economic situation	x17 - the number of people employed in entrepreneurship; Of them: x18 is the population engaged in trade; x19 - the number of people employed at home; x20 - the number of people employed in crafts; x21 - population of cattle breeding, poultry farming and beekeeping; x22 - the number of people involved in other areas of entrepreneurship; x23 - the number of families engaged in family business; x24 - the number of retail stores x25 - the number of objects holding mass events; x26 - the number of catering points (including the number of teahouses) x27 - number of farmers

Social status	x28 - the number of college graduates (2018/2019 academic year); x29 - the number of graduates of academic lyceums (academic year 2018/2019); x30 - the number of crimes committed; x31 - preschool educational institutions; x32 - the amount of school; x33 - the number of students per teacher; x34 - percentage of male teachers%; x35 - percentage of female teachers%; x36 - the number of teachers with the highest category%; x37 - the number of teachers in the 1st category%; x38 - number of teachers of the 2nd category%; x39 - the number of specialist teachers; x40 - number of colleges; x41 - the number of children attending preschool educational institutions; x42 - the number of students; x43 - the number of students attending college; x44 - the number of attending lyceum; x45 - the number of university students; x46 - number of non-students; x47 - the number of students abroad; x48 - the number of low-income families; x49 - the number of recipients of material assistance; x50 - The number of recipients of benefits up to 14 years; x51 - The number of recipients of benefits up to 2 years; x52 - the number of recipients of disability benefits;
	x53 - the number of family divorces; x54 - the number of dysfunctional families registered with preventive care x55 - the number of people who use narcotic and psychotropic substances registered in preventive care
	x56 - the number of persons belonging to the category of religious extremist movement, registered on a preventive basis x57 - number of unemployed x58 - number of mosques x59 - number of cemeteries
	x60 - repair and repair facilities (shoes, blacksmithing, television and radio equipment, sewing)
The standard of living indicators.	x61 - total income per capita (thousand soums) x62 - real total income per capita (thousand soums) x63 - the volume of gross domestic (regional) product per capita);
Housing conditions of the population	x64 - housing stock (average per inhabitant, m2)

In order to select significant variables, a correlation matrix is constructed (Fig. 1).

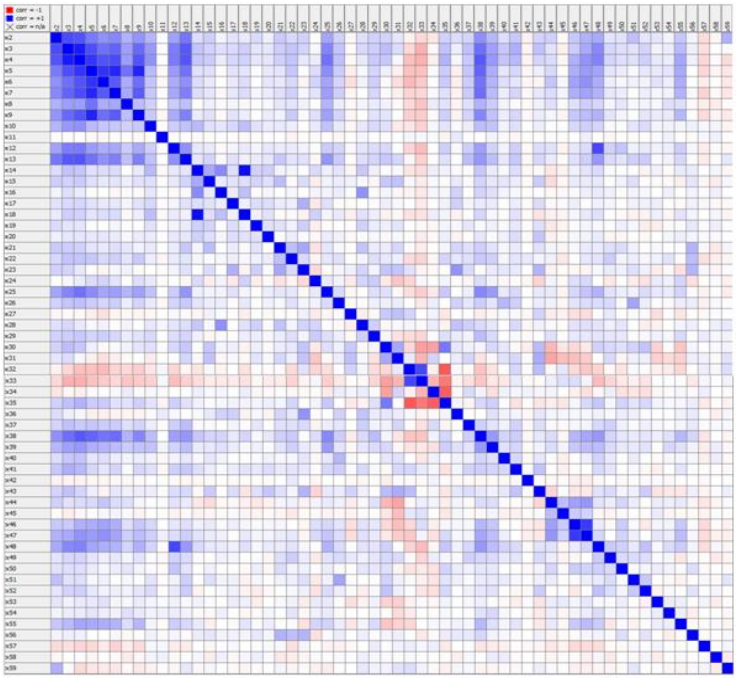

Fig. 1. The correlation matrix for the variables

As a result of the analysis of the variables for collinearity, the variables "x6 - the number of men" and "x7 - the number of women" were excluded due to their strong correlation with the variable "x5 - number of the resident population" where $r > 0.98$; the variable "x34 is the proportion of male teachers" is a strong correlation with the variable "x35 is the proportion of female teachers", where r = -1; the variable "x61 is the total per capita income" has a correlation with the variable "x62 is the real total per capita income", where r> 0.93. Here r is the correlation coefficient. As a result, 60 variables remained.

Each self-government body has a set of n numbers. N is a characteristic of an object (in our case, these are indicators of the state of organs, level and quality of life), each object is a numerical vector in n-dimensional space. Similar objects are located close to each other in n-dimensional space, their characteristics differ little and, accordingly, the distance between the points is small. The cluster of points located close to each other will be a cluster. It is important to note that the indicators selected for the study are incommensurable with each other (units of measurement - millions, thousands, percentages, etc.), so there was a need to normalize them. In order to normalize the initial data, a linear transformation is applied, where the minimum value is 0 and the maximum value is 1.

The remaining variables about self-government bodies, the number of which is exactly 445, are subjected to cluster analysis by the method. Consider the algorithm for cluster analysis of k-means. Initially, each object is a cluster, i.e. in one cluster is one subject. You need to know in advance how many clusters you need to distribute the data. This is a significant disadvantage of this method, but this problem is solved by improved implementations of the algorithm. You need to select the initial cluster centers.

To select data containing n records (objects), the number of clusters k, which must be formed, is specified. Then the algorithm splits all the objects of the sample into k sections ($k < n$), which are clusters. The k-mean s algorithm is one of the simplest and most effective clustering algorithms. It consists of four steps; first, the number of clusters k, which should be formed from the objects of the initial sample, is specified.

Randomly selected k records of the original sample, which will serve as the initial centers of clusters μ. Such starting points, from which the cluster then "grows", are often called "seeds" (from the English. seeds - seeds, crops). Each such record will be a kind of "embryo" cluster consisting of only one element. For each centroid μ_k, calculate the distances $l_k^{(i)}$ to all $x^{(i)}$ points $l_k^{(i)} = \rho(x^{(i)}, \mu_k)$. The key point of the K-means algorithm is to calculate at each iteration the distance between the records and the cluster centers, which is necessary to determine which of the clusters this record belongs to. The rule by which the distance is calculated in a multidimensional feature space is called a metric. The most commonly used metrics in practical clustering problems are Euclidean distance (metric L2), Manhattan distance (metric L1), Chebyshev distance, etc.

We have a set of objects $X = \{x_i\}$, $i = \overline{1,n}$, where x^j - objects, in our case it is statistical data on the state of local self-government (mahalla), adopted in the form of indicators, recorded at a certain time in all civil gatherings of Khorezm region. That is, given a data set $\{X\}$, which is obtained from the Department of statistics in Khorezm region and Khorezm regional Council on coordination of activity of local authorities, as follows:

Objects	1	2	3	…	n
X_1	x_{11}	x_{12}	x_{13}	…	x_{1n}
X_2	x_{21}	x_{22}	x_{23}	…	x_{2n}
X_3	x_{31}	x_{32}	x_{33}	…	x_{3n}
…	…	…	…	…	…
X_m	x_{m1}	x_{m2}	x_{m3}	…	x_{mn}

Here X_i $(i = \overline{1,m})$ are objects (self-governing bodies) and $x_{ij}(i = \overline{1,m}, j = \overline{1,n})$ the j-th sign of the i-th object.

Let X be a set of objects, S be a set of identifiers (labels) of clusters. On the set X, a distance function between objects $\rho(x,x')$ is given. A finite training set of objects is given $X_k = \{x_{k1}, \ldots, x_{mk}\} \subset X$. It is necessary to divide the sample into subsets (clusters), that is, to each object $x_i \in X$ we associate a label $s_i \in S$, so that the objects inside each cluster are close relative to the metric ρ, and objects from different clusters differed significantly [7]. To determine the distances between the clusters, the full coupling method was used:

$$R_{max}(U,V) = max_{u \in U, v \in V} \rho(u,v); \quad (1)$$

Clusters show the degree of proximity of organs and clusters [11]. According to the substantive and statistical criteria, the breakdown of self-government bodies of 5 clusters turned out to be the best (Table 2), since their further unification is already impractical due to their dissimilarity to each other (this is indicated by a too high jump in the distance between the clusters). The clustering of regions was carried out according to the data for 2018.

III. RESULTS

As a result of data analysis by the selected method, 5 clusters were selected (table 2). The first and second cluster has a smaller accumulation of self-government bodies, respectively 75 and 53 of the subject. Next in terms of the number of self-government bodies in the cluster are clusters No. 3, No. 4 and No. 5, which unite 103, 105 and 109 local authorities, respectively. In table 2, you can see the affiliation of self-government bodies to a particular cluster in the context of the districts of the Khorezm region.

Table 2. The results of clustering of local self-government bodies of the Khorezm region according to the main socio-economic indicators (indicators of the level and quality of life of the population)

Cluster	Districts of the Khorezm region	The number of local governments accumulated in the cluster
Cluster 1	Total:	75 subjects
	Bagat district:	8
	Gurlen district:	6
	Kushkupir district :	8
	Urgench district:	12
	Urgench:	8
	Khazarasp district:	15
	Khiva district:	1
	Khiva:	1
	Khanka district:	8
	Shavat district:	2
	Yangiarik district:	2
	Yangibazar district:	4
Cluster 2	Total:	53 subjects
	Bagat district:	3
	Gurlen district:	3
	Kushkupir district :	1
	Urgench district:	7
	Urgench city:	10
	Khazarasp district:	13
	Khiva district:	6
	Khiva city:	10
	Khanka district:	0
	Shavat district:	0
	Yangiarik district:	0
	Yangibazar district:	0
Cluster 3	Total:	103 subjects
	Bagat district:	8
	Gurlen district:	12
	Kushkupir district :	20
	Urgench district:	16
	Urgench city:	7
	Khazarasp district:	5
	Khiva district:	3
	Khiva city:	5
	Khanka district:	11
	Shavat district:	4
	Yangiarik :	8
	Yangibazar district:	4
Cluster 4	Total:	105 subjects
	Bagat district:	16
	Gurlen district:	16
	Kushkupir district :	10
	Urgench district:	9
	Urgench city:	7
	Khazarasp district:	7
	Khiva district:	3
	Khiva city:	4
	Khanka district:	7
	Shavat district:	3
	Yangiarik district:	13

	Yangibazar district:	10
Cluster 5	Total:	109 subjects
	Bagat district:	7
	Gurlen district:	15
	Kushkupir district :	7
	Urgench district:	16
	Urgench city:	9
	Khazarasp district:	9
	Khiva district:	3
	Khiva city:	5
	Khanka district:	6
	Shavat district:	10
	Yangiarik district:	15
	Yangibazar district:	7

Now let's move on to the analysis of the clusters themselves. The most convenient way to describe the self-government bodies included in the cluster is to calculate the arithmetic mean values of the signs (indicators of the level and quality of life of the population) characterizing the self-government bodies included in this cluster. The analysis of the averaged indicators allows to estimate the nature of the self-government bodies that are part of a certain cluster. Thus, the method of cluster analysis made it possible to divide the data into the following clusters.

In cluster 1 - 75 self-government bodies from different districts of Khorezm region are accumulated. This cluster on 7 indicators has a maximum of values and on 5 indicators has a minimum of average values in comparison with average values of indicators of other clusters. It can be seen in table 3, for this cluster is characterized by a relatively low GRP per capita (average – 7949,4 thousand. sum.), the population engaged in entrepreneurship, trading, cattle breeding, poultry farming and beekeeping is higher than the rest of the clusters and it can be said that the maximum value of these indicators has a positive effect on the number of recipients of financial assistance. The average number of beneficiaries is 3.6 and this is the lowest value among the other clusters. And the cluster has a high rate of the number of shops and farms, which have a positive impact on the economic development of local governments. Another important two indicators for this cluster is the number of mosques for believers and has a maximum value, and the number of crimes committed has a minimum value than the other clusters.

Table 3. Average values of indicators of the level and quality of life in clusters for 2018

Показатели	Кластеры				
	1	2	3	4	5
x3 - number of households;	776,7	*1121,7	626,9	591,5	**514,3
x4 - family number;	1136,9	*1746,7	853,3	835,9	**723,6
x5 - resident population;	4161,6	*6513,1	3137,7	3117,4	**2606
x6 – children, adolescents and youth (up to 30 years);	2197,5	*3454,6	1664,1	1614,7	**1345,3
x7 – children under 7 years old;	519,5	*749,6	368,8	373,1	**322,6
x8 – children and adolescents from 7 to 16 years old;	652,8	*952,5	522,3	510,9	**395,8
x9 – minors from 16 to 18 years old;	243,7	*421,6	154,2	143,3	**125,9
x10 – youth from 18 to 30 years old;	781,4	*1330,9	618,8	587,4	**501
x11 – the number of families who have lost the bread-winner;	11,9	*15,5	9,7	9,4	**8,5
x12 – number of single elderly people;	0,6	0,5	0,7	**0,4	*3,2
x13 – number of disabled people;	57,9	*93	43,3	43,9	**39,4
x14 – number of pensioners;	369,8	*508,7	279,3	276,1	**238
x15 – the number of people employed in business;	*239	215,7	**160,1	207,9	174,4
x16 – population engaged in trade;	36,6	*46,1	31	**29,3	31,1
x17 –;number of population engaged in homework	4,5	*12,5	**3,4	3,6	6,6
x18 – the number of people employed in crafts;	*6,6	6	**2,9	3,2	3,3
x19 – population of cattle breeding, poultry farming and beekeeping;	*165,2	121	**99,8	148,7	106,2
x20 – the number of people involved in other areas of entrepreneurship;	17,7	*21,2	17,1	**15,7	20,3
x21 – number of families engaged in family business;	8,4	*8,9	**5,8	7,4	6,9
x22 - number of shops	*10,2	9,9	6,8	**5,7	7
x23 - number of objects holding mass events;	0,6	*1	0,6	0,5	**0,4
x24 - number of catering points (including the number of teahouses)	1,4	*1,5	1,3	**0,9	1,3

x25 - number of farms;	*9,4	8,9	**7,2	8,3	8,1
x26 – number of college graduates (academic year 2018/2019;	49,4	*90,5	41,7	38,1	**34,1
x27 - the number of graduates of academic lyceums (academic year 2018/2019);	6,1	*10,5	5	4,6	**3,6
x28 – number of crimes committed;	**4,8	5,5	*6,3	5,6	5,9
x29 – preschool educational institutions;	0,7	*0,8	**0,6	0,7	0,7
x30 – amount of schools;	1,1	*1,4	**0,8	1	1
x31 – number of students per teacher;	12,1	*12,7	**11,6	11,7	11,9
x32 – share of women teachers%;	75,9	*76,9	76,2	**75	76,2
x33 – the number of teachers with the highest category%;	*4,2	**3,6	4,4	4,6	*4,7
x34 – the number of teachers of the 1st category%;	13,5	**11,9	14	*14,6	14,4
x35 - the number of teachers of the 2nd category%;	*31,5	**29,6	31,1	29,9	30,1
x36 – number of specialist teachers;	50,9	*55,1	**50,7	51	50,8
x37 – number of colleges;	0,2	0,1	*0,3	**0,1	**0,1
x38 – the number of children attending preschool educational institutions;	104	*156,1	99	**86,6	89,1
x39 - number of schoolchildren;	660,3	*1031,2	502,1	495,5	**426,6
x40 - number of students attending college;	77,8	*135,6	61,4	73,4	**59
x41 - number of attending lyceum;	10,7	*19,1	9,6	9,3	**9
x42 - number of university students;	24,1	*26,5	21,9	19,6	**17,9
x43 - number of non-students;	**0	0,5	*4,4	1,2	0,1
x44 - the number of students abroad;	**1,5	*3,8	2	1,8	2,1
x45 - number of low-income families;	26,5	*38,3	26	26,3	**21,6
x46 - the number of recipients of material assistance;	**3,6	3,9	*5	4,4	4,7
x47 - The number of recipients of benefits up to 14 years;	16,8	*21,8	14,5	14	**10,6
x48 - The number of recipients of benefits up to 2 years;	29,4	*42	26,3	24,4	**21,2
x49 - the number of recipients of disability benefits;	52,7	*96	40,9	49,8	**36,7
x50 - number of family divorces;	0,8	*1,4	0,7	0,6	**0,5
x51 - number of dysfunctional preventive families	1,5	*1,9	**1,3	1,4	**1,3
x52 - number of people who use narcotic and psychotropic substances registered	0,3	*0,7	0,3	**0,1	**0,1
x53 - the number of persons belonging to the category of religious extremist movement, registered on a preventive basis	0,5	*1	**0,3	0,4	0,4
x54 - number of unemployed	82	71,6	*82,4	**69,2	76,8
x55 - number of mosques	*0,4	0,3	**0,2	0,3	**0,2
x56 - number of cemeteries	1,2	*1,8	0,9	0,9	**0,8
x57 - repair workshops (shoes, blacksmithing, body radio equipment, sewing)	**0,9	1,2	*1,4	1,2	1,3
x58 - real total income per capita (thousand soums)	8520	6736,7	6824,9	**6452	*8596
x59 - volume of gross domestic (regional) product per capita;	7949,4	8517,9	**7178,7	*9407,8	8744,7
x60 - housing fund (on average per inhabitant, m^2)	111,8	**103,8	123	114,3	*142,3

* is the max value, ** is the min value.

Cluster 2-regions with a high population and a small number of accumulated self-government bodies. The cluster consists of 53 local governments belonging to eight districts of 12 districts of Khorezm region. The cluster has a maximum of 39 values and a minimum of 5 values. The cluster territory is characterized by not high GRP per capita (about 8517,9 thousand sums.). In this cluster, the lowest housing stock per person is 103.8 m^2. In the educational direction, despite the high number of students, the cluster has a minimum value in the proportion of teachers with the highest category, the 1st category and the 2nd category than the other clusters. For teachers, it does not matter-55.1, this is the maximum value among clusters. Despite the high number of populations engaged in entrepreneurship, the number of low-income families is high.

Thus, cluster 2 is the leader in terms of population, the number of disabled people and pensioners, the number of entrepreneurs, the number of students in educational institutions and the number of recipients of material assistance and benefits.

Cluster 3-self-government bodies with the level and quality of life of the population below average. The cluster consists of 103 subjects of Khorezm region. This cluster is characterized by a minimum value of GRP per capita, the average value was less than 7200 thousand soums. In the cluster under consideration, there is a minimum number of employees by type of small business. The peculiarity of the cluster is that in this cluster the maximum or minimum values of indicators that negatively affect the nature of the cluster, such as: the number of crimes, the number of benefits received, the number of unemployed.

Cluster 4-self-government bodies with an average value of GDP per capita of more than 9400 thousand sums. The cluster consists of 105 self-government bodies. Real total income per capita is the lowest among the clusters and is less than 6500 thousand soums. The cluster has a maximum value for the proportion of teachers of the 1st category among the other clusters.

Cluster 5 combines local governments with a small population. In a cluster, very many indicators have a minimum value among clusters. According to the population engaged in entrepreneurship, it occupies an average value. In the educational area, the number of teachers with the highest category is of maximum importance. The number of students in educational institutions is minimal. Real total per capita income will amount to 8600 thousand soums. It is worth noting that the cluster in question is characterized by a rather large living area, which comes to 1 inhabitant of 142 m^2 each, which takes first place among the other clusters. The cluster has minimum values for negative indicators such as the number of recipients of benefits up to 14 years, the number of recipients of benefits up to 2 years, the number of recipients of disability benefits, the number of family divorces, the number of dysfunctional families registered for preventive care. These indicators positively affect the general nature of the cluster under consideration.

To check the quality of clustering, we used the Dunn index and the Davies–Bouldin index [12].

The Dunn index is calculated as follows:

$$I(C) = \frac{min_{i \neq j}\{\partial(C_i C_j)\}}{max\{\Delta(C_l)\}_{1 \leq l \leq k}}, \quad (2)$$

where C is the set of clustered objects with their cluster numbers; k is the maximum number of clusters; ∂ is the distance between the i-th and j-th clusters; Δ - is the diameter of the cluster.

The index numerator calculates the minimum distance between clusters, and the denominator calculates the maximum diameter of the cluster. Thus, the larger the index value, the better the clustering is.

In table. 4 the Dunn index values for different number of clusters are given.

According to the Dunn index, the best partition is achieved by clustering into 5-6 clusters.

Consider another indicator - the Davies–Bouldin index, which is calculated as follows:

$$DB = \frac{1}{k}\sum_{i=1}^{k} R_i\,; \quad (3)$$

$$R_i = \max\left(\frac{s(C_i)+s(C_j)}{\partial(C_i,C_j)}\right), \quad (4)$$

where k is the number of clusters; s is the dispersion within the cluster; d is the distance between the ith and jth clusters.

Table 4. Dunn index for clustering from 2 to 20 self-government bodies

Number of clusters	Dunn index
2	1,9051384
3	1,6358948
4	1,8588234
5	1,9461553
6	1,9453679
7	1,5580750
8	1,4791925
9	1,6303234
10	1,4884319
11	1,4522240
12	1,2925523
13	1,3487938
14	1,2912529
15	1,3206137
16	1,3637968
17	1,4157574
18	1,3587619
19	1,2901372
20	1,2698764

In the numerator of the index, the sum of the variances within the clusters is calculated, and in the denominator, the distance between the clusters.Thus, the target value is the minimum of the index.

In table 5 the values of the Davies–Bouldin index for different number of clusters are given.

The lowest index value is achieved when divided into 4-5 clusters.

Thus, both indicators show that the partition into 5 clusters is reliable and stable, the distance between the clusters is quite large, and the scatter of points within the clusters is small.

Table 5. Davies–Bouldin index for clustering 2 to 20 self-government.

Number of clusters	Davies–Bouldin index
2	0,13490514
3	0,14673232
4	0,07753953
5	0,07753952
6	0,08358717
7	0,08358717
8	0,10507854
9	0,10507854
10	0,12454255
11	0,12454255
12	0,15554586
13	0,15554586
14	0,15554586
15	0,15853574
16	0,15853574
17	0,15853574
18	0,17547404
19	0,17547404
20	0,17547404

IV. CONCLUSIONS

Cluster analysis in assessing the quality and quality of life of the population allows to divide regions into 5 separate clusters: regions with average level and quality of life, regions with low level and quality of life, regions with the level and quality of life below average, regions with the level and quality of life above average, regions with the highest level and quality of life.

Analysis of the quality of clustering using Dunn and Davies–Bouldin indices showed that the grouping of subjects adopted in the work is optimal.

Cluster analysis is a reliable way to divide regions into groups by all indicators, without highlighting the most significant indicators and without ranking them.

Thus, the presented methodology of cluster analysis allows to take into account the existing individual and group differences, which must be taken into account when determining approaches to the formation of a regional strategy of socio-economic development, as well as the whole range of socio-economic measures necessary to improve the level and quality of life of the population.

REFERENCES

1. Karimova P.M., Yapparova P.P. Opit klasterizatsii regionov Rossiyskoy Federatsii po pokazatelyam dezurbanizirovannosti // Vestnik VEGU. 2016. №6 (86). p.12.
2. Kurkudinova E.V. Klasternoy podxod kak texnologiya upravleniya ekonomicheskim razvitiem regiona // Ekonomicheskie nauki. Ekonomika i upravlenie. 2010. № 10. p. 170-172.
3. Mnogomernoy statisticheskiy analiz v ekonomicheskix zadachax: kompyuternoe modelirovanie v SPSS : ucheb. posobie /pod red. I.V. Orlovoy. Moskva : Vuz. uchebnik,2011. 310 p.
4. Perng-Cherng Wang, Jin-Jang Leou. New Fuzzy Hierarchical Clustering Algorithms. Journal of Information Science and Engineering, Vol. 9 No. 3, pp. 461-489.
5. Ondrej Simpach, Marie Pechrova. CLUSTER ANALYSIS OF EU'S REGIONS ACCORDING TO DEMOGRAPHIC CRITERIA. The 10th International Days of Statistics and Economics. Conference Proceedings. September 8–10, 2016; Prague, Czech Republic. pp.1806–1815.
6. Tomas Loster. DETERMINING THE OPTIMAL NUMBER OF CLUSTERS IN CLUSTER ANALYSIS. The 10th International Days of Statistics and Economics. Conference Proceedings. September 8–10, 2016; Prague, Czech Republic. pp. 1078–1090.
7. Scikit-learn Documentation - Clustering. URL: http://scikit-learn.org/stable/user_guide.html
8. Gosudarstvenno'y komitet Respubliki Uzbekistan po statistike. Uroven jizni naseleniya.URL: https://stat.uz/.
9. Upravlenie statistiki Xorezmskoy Oblasti. URL: http://www.xorazmstat.uz/
10. Respublikanskiy sovet po koordinatsii deyatelnosti organov samoupravleniya grajdan. URL: https://mahallakengashi.uz/.
11. Kovanova E.S. Klasterno'y analiz v reshenii zadachi tipologii regionov Rossii po urovnyu i intensivnosti vnutrenney trudovoy migratsii // Vestnik NGUEU.2013. № 4. pp. 171-172.
12. Stein B., Meyer S., Wifi brock F. On Cluster Validity and the Information Need of Users// Jnt. Conference on Articial Intelligence and Applications (AIA03). Benalmadena, Spain, 2013. September. p. 217-218.

AUTHORS PROFILE

Rakhimboev Khikmat Jumanazarovich, received the B.S. and M.S. degrees in Urgench branch of Tashkent University of information technologies. Senior Lecturer, Department of Information Technology, Urgench Branch of Tashkent University of Information Technologies, Uzbekistan. His research interests include intelligent analysis in the management of the municipal executive authorities. E-mail: egov@ubtuit.uz +99897-516-05-22

Ismailov Mirkhalil Agzamovich, Chief Researcher, Research and Innovation Center information and communication technologies at Tashkent University of information technologies, Uzbekistan. Doctor of Technical Sciences. E-mail: mirismail46@mail.ru +99871-242-48-47

Khusinov Khamid Khudayberganovich, received the M.S. degrees in Urgench State University in 1999. He is currently a lecturer in the Urgench branch of Tashkent University of Information Technologyies, Uzbekistan. His research interests include intellectual analysis in the management of social organizations. E-mail: xamid76@mail.ru +99899-743-21-40