American Journal of Social and Humanitarian Research

Vol. 6 Issue 8 | **pp.** 1937-1943 | **ISSN:** 2690-9626
Available online @ https://globalresearchnetwork.us/index.php/ajshr

Article

The Application of Clustering as a Statistical Analysis Method in Assessing The Socio-Economic Development of Regions

Rajabov Alibek Xushnudbekovich[*1]

1. PhD, "Mamun University" Associate Professor of the Department of Economics

* Correspondence: alibek.rajabov@gmail.com

Abstract: The application of the clustering statistical analysis method in assessing the socio-economic development of regions has gained significant relevance amidst global economic instability and disparities in resource distribution. This study examines the clustering of cities and districts in the Khorezm region of the Republic of Uzbekistan based on socio-economic development indicators from 2017 to 2024. The clustering of cities and districts was conducted using the "Cluster Analysis" method. According to the research findings, the cities and districts of the Khorezm region were divided into five clusters. This clustering was performed based on the "Euclidean distance" measure. The Euclidean distances reflect the degree of similarity, where a larger distance indicates a lower level of similarity, and a smaller distance signifies a higher level of similarity. Elements identified as the most similar were grouped into the same cluster.

Keywords: Region, Socio-Economic Development, Cluster Analysis, Euclidean Metric, Similarity Matrix, Set, Elements

Citation: Xushnudbekovich R. A. The Application of Clustering as a Statistical Analysis Method in Assessing The Socio-Economic Development of Regions. American Journal of Social and Humanitarian Research 2025, 6(8), 1937-1943.

Received: 15th Jun 2025
Revised: 30th Jun 2025
Accepted: 20th Jul 2025
Published: 02nd Aug 2025

1. Introduction

The increasingly rapid pace of globalization, coupled with global instability, particularly the shift of risks to global economic growth toward a negative trajectory, continues to contribute to uncertainties in the future trajectory of the world economy. According to the forecasts of the International Monetary Fund, "global economic growth is expected to decline from 3.3 percent in 2023 to 3.2 percent in 2024 and remain at 3.2 percent in 2025" [1]. In turn, this situation underscores the necessity of implementing effective regional policies to address disparities in per capita income across countries and to overcome imbalances in achieving sustainable socio-economic development.

Ensuring alignment between macroeconomic stability and structural reforms, as well as implementing an effective system and mechanisms for managing economic cycles influenced by external and internal factors, based on modern forecasting models, requires increasing attention. "For the development of the economy, it is essential to promote the comprehensive and balanced socio-economic development of regions, districts, and cities, while effectively and optimally utilizing their potential. At present, it is necessary to establish a systematic and continuous approach to assessing the socio-economic development of regions based on specific criteria" [2]. From this perspective, the efficient use of regional potential and the objective evaluation of the sustainable socio-economic development of regions are considered among the most critical issues of today.

The application of the clustering method in statistical analysis for assessing regional socio-economic development has gained significant relevance amidst global economic instability and disparities in resource distribution. In recent years, the acceleration of urbanization processes, digital transformation, and the demands for ecological sustainability have further deepened regional developmental disparities, increasing the need for precise and multidimensional analytical methods in policy formulation. Clustering, by grouping data based on similarities, enables a systematic evaluation of regions' socio-economic indicators (e.g., GDP, employment rate, access to education, and healthcare), serving as a key tool for efficient resource allocation and the development of policy measures.

In recent years, significant attention has been paid in the scientific works of foreign researchers to the issues of assessing the socio-economic development of regions. For instance, Cui, D., Yu, Y., and Song, Z. [3] select ten statistical indicators to evaluate the development level of cities in Hebei Province, employing factor analysis and cluster analysis methods for data analysis. Following the analysis, the authors categorize the cities into four levels and subsequently summarize the regional economic characteristics of Hebei Province.

Andreea-Ionela Puiu and Necula Marian [4] identify several reasons for the regional convergence of research and innovation activities, taking into account the existing economic disparities across European regions. Their study focuses on assessing the total factor productivity indicators related to knowledge production in the European region, proposing the identification of several clusters in European countries based on their innovation potential and total factor productivity.

Sruthi Krishnan V and Mohammed Firoz C [5] propose an assessment of environmental quality using cluster analysis within the framework of sustainable regional development. In this study, settlements are divided into five clusters based on the homogeneity of environmental quality, and the associated challenges, prospects, as well as the key directions for policy interventions for each of these clusters are elucidated.

Roger R. Stough and Junbo Yu [6] have assessed the influence of geographic factors on the formation, development, and future dynamics of the processed food industry cluster by employing cluster analysis methods, while accounting for the competitiveness of Northeast Asia.

In their research, M. Çağlar and C. Gürler [7] classified global countries according to sustainable development goals (SDGs) using progress data compiled from the Sustainable Development Report 2019. This classification aimed to identify disparities among 110 countries and to establish priorities for action. The K-Means algorithm, a non-hierarchical clustering method, was applied to categorize the countries. Following the formation of homogeneous country groups, each cluster was examined based on the socio-economic and politico-cultural structures of the constituent nations. The results of the cluster analysis demonstrate that countries can be partitioned into five distinct clusters. Countries within each cluster exhibit substantial similarities not only in terms of SDG progress but also with respect to their socio-economic and politico-cultural frameworks. Overall, clusters characterized by advanced socio-economic structures and superior politico-cultural systems demonstrate higher levels of SDG achievement. Socio-economic and politico-cultural indicators are positively correlated with the majority of SDG indices.

In the study by L. M. Akimova et al. [8], the primary aspects of planning the socio-economic development of regions are thoroughly examined, using the example of European Union member states with transitional economies. Research methods such as comparison and grouping were employed to investigate the characteristics of socio-economic development planning in the regions of Bulgaria, Estonia, Latvia, Lithuania, Poland, Romania, Slovakia, Hungary, Croatia, and Czechia. The theoretical aspects of planning the socio-economic development of regions were reviewed. The practical aspects of planning the socio-economic development of regions in Bulgaria, Estonia, Latvia, Lithuania, Poland, Romania, Slovakia, Hungary, Croatia, and Czechia were analyzed, presented through the following indicators: life expectancy; employment rate; economically active population; number of active enterprises; and labor productivity in agriculture. Based on the results of the conducted research and the main directions of the practical aspects of planning the socio-economic development of regions, it was determined that specific measures for regional development are being implemented in European Union countries such as Bulgaria, Estonia, Latvia, Lithuania, Poland, Romania, Slovakia, Hungary, Croatia, and Czechia, including in rural areas.

In their study, M. C. Nogueira and M. Madaleno [9] examine the Global Competitiveness Index (GCI), Human Development Index (HDI), Ease of Doing Business (EDB), Environmental Performance Index (EPI), and Global Entrepreneurship Index (GEI). A country's performance on these indices is associated with economic growth, particularly given that several empirical studies have identified evidence supporting these associations, as the indices are constructed based on the scientific literature on economic growth. By constructing a database of these indices for European Union countries over the period from 2007 to 2017, the authors empirically investigate the relationship between the aforementioned indices and economic growth in European Union countries during this period through panel data evaluation using panel data methodology and subsequently two-stage least squares (2SLS) to address endogeneity issues. However, since the European Union comprises diverse countries with varying economic and social realities, the researchers divide the countries into six clusters and provide individual interpretations for each. The researchers find that human development and competitiveness play a significant role in economic growth, and that entrepreneurship also influences this growth; furthermore, with regard to income distribution, employing the Gini index, they determine that only human development mitigates inequalities.

The research by S. E. Barykin et al. [10] investigates the impact of the steady growth of public debts on average debt sustainability in 11 regions of Russia from 2010 onward, spanning approximately a decade. The present study is focused on evaluating the debt sustainability of Russian regional budgets through the identification of Euclidean distance-based budget constraints and cluster analysis. This investigation is grounded in hierarchical cluster analysis methodology, which enables the extraction of object clusters from aggregate data and their consolidation into homogeneous segments. The central hypothesis of this study posits that employing this method can enhance the precision of values that constrain budgetary limitations within the regional financial system. The results indicate that regions with high debt sustainability include the city of Saint Petersburg, Leningrad Oblast, and Kaliningrad Oblast. From the perspective of public debt policy, the findings provide additional evidence supporting debt reduction measures for the Republic of Komi, Republic of Karelia, Arkhangelsk Oblast, and Pskov Oblast.

The current research by Z. Arshad et al. [11] evaluates the impact of information and communication technology (ICT), trade, economic growth, financial development, and energy consumption on carbon emissions in the South and Southeast Asia (SSEA) region over the period from 1990 to 2014. Furthermore, the study endeavors to validate the environmental Kuznets curve (EKC) hypothesis between per capita GDP and CO_2 emissions. Cluster analysis was employed to delineate two groups (potential and advanced countries) based on social development indicators. The findings reveal that financial development and ICT have degraded environmental quality in the SSEA region, indicating that ICT goods and services are not energy-efficient in both potential and advanced countries, and that a substantial portion of financial investments in potential countries is allocated to environmentally unfriendly projects. Conversely, in developed countries, financial development mitigates CO_2 emissions. Moreover, the results substantiate inverted U-shaped relationships across all three examined panels-namely, potential, advanced, and full country panels-thereby confirming the EKC hypothesis.

The objective of the research by R. Huseynov et al. [12] is to develop a comprehensive approach to the clustering of agricultural enterprises, which enables the diagnosis of their market position and the identification of integration opportunities aimed at increasing production volumes. Cluster analysis is employed as the methodology in this study. The investigation analyzed 44 agricultural enterprises in Russia. Based on cluster analysis, four types of agricultural enterprises were identified, each possessing diverse activity profiles that are significant within a specific economic market segment; for these, the average and threshold values of clustering indicators were calculated. Through clustering, we were able to delineate the characteristics of each group of companies. A leaders' cluster was identified, characterized by a high number of employees and performance indicators exceeding the average. Enterprises in this cluster are regarded as industry leaders; they encompass extensive territories across Russia; they engage in technological, innovative, logistical, and international activities; and they exhibit a high level of strategic planning and management. This generalized description allowed for a comparison of the primary criteria across clusters and facilitated the determination of strategic action domains to enhance production outcomes.

2. Materials and Methods

In our research, we employed the cluster analysis method. Cluster analysis is regarded as a multivariate technique, wherein similar entities are partitioned into multiple groups based on the sets of variables considered in this analysis.

"Cluster analysis" is a statistical method employed to classify elements within a given dataset into groups known as clusters, based on similarities across various features. These similarities are quantified in the form of distance metrics, wherein a greater distance signifies a lower degree of similarity, while a smaller distance indicates a higher degree of similarity. The elements deemed most similar are assigned to the same cluster. Each cluster possesses well-defined centers, specifically the point representing the arithmetic mean of all data points within the cluster. Consequently, each cluster is represented by its centroid.

The cluster analysis method is implemented through the following stages:

1. Let the distance between clusters i and j be denoted by d_{ij}, and let n_i represent the number of elements in cluster i. The set of all remaining d_{ij} values is denoted by D, with a total of N elements present in the cluster. The distance d_{ij} (Euclidean distance) is calculated based on the following formula:

$$d_{ij} = \sqrt{\sum_{i=1}^{N}(x_i - y_i)^2/N} \qquad (1)$$

2. The minimal distance d_{ij} in the set D is determined;
3. Clusters i and j are merged into a single new cluster k;
4. Using the following formula, a new set of distances d_{km} is computed.:

$$d_{km} = \alpha_i * d_{im} + \alpha_j * d_{jm} + \beta * d_{ij} + \gamma * \left|d_{im} - d_{jm}\right| \qquad (2)$$

Here, m denotes any cluster distinct from k, while d_{im} and d_{jm} represent the new distances that are incorporated into the set D. Furthermore, $n_k = n_i + n_j$

$$\alpha_i = \frac{n_i}{n_k}\alpha_j = \frac{n_j}{n_k}\beta = -\alpha_i * \alpha_j$$

5. Steps 1–3 are iterated until the set D encompasses the complete ensemble comprising all elements.

3. Results and Discussion

Based on the aforementioned formulas, a proximity matrix was constructed for the cities and districts of the Khorezm region using indicators of sustainable socio-economic development, see Table 1. Upon analyzing the data in the table, it is evident that Urgench district is positioned at the closest distance to Urgench city (0.329 units), whereas Khiva city is at the farthest distance (0.918 units). Furthermore, Urgench city is somewhat more distant from Yangiariq and Gurlan districts, with distances amounting to 0.729 and 0.700 units, respectively. It is notably distant from Qushkupir district (i.e., 0.609 units). This indicates that Urgench city holds a leading position in sustainable socio-economic development compared to other cities and districts in the region.

The Yangiariq district exhibited the closest proximity to the city of Khiva, with a distance metric of 0.248, whereas the Tuproqqala district was the most remote, registering a value of 0.643. Additionally, the Khiva district demonstrated substantial separation from the Urgench district, as indicated by a metric of 0.623.

The Tuproqqala district is the farthest from the Bagat district (distance: 0.550), while the closest districts are Shavat (0.074), Khazarasp (0.090), and Khiva (0.093). This indicates that the Bagat, Shavat, Khazarasp and Khiva districts belong to the same cluster. It can be inferred that the Bagat district is in close proximity to the Qushkupir (0.140), Urgench (0.140), and Honka (0.183) districts.

The districts of Gurlan, Yangiariq and Yangibazar are also positioned in close proximity to one another (with respective distances of 0.031 and 0.089), thereby indicating that these districts constitute a single cluster. The districts located farthest from Gurlan are Khazarasp (0.466), Urgench (0.461) and Khiva (0.442).

In terms of proximity to Qushkupir district, the nearest districts are Khiva (0.098), Khazarasp (0.131), and Honka (0.163). The most distant districts are Tuproqqala district at 0.609 units and Yangibazar district at 0.439 units. Analyzing the distances relative to Urgench district, the closest districts are Shavat (0.071), Khazarasp (0.203) and Khiva (0.222). The farthest districts are Tuproqqala (0.506), Yangibazar (0.487), and Yangiariq (0.485).

Table 1. Proximity matrix among the cities and districts of Khorezm region based on socio-economic development indicators (computed using “Euclidean distance”).

Names of Cities and Districts	Urgench city	Khiva city	Bagat	Gurlan	Qushkupir	Urgench	Khazarasp	Tuproqqala	Honka	Khiva	Shavat	Yangiariq	Yangibazar
Urgench city	0.000												
Khiva city	0.918	0.000											
Bagat	0.469	0.510	0.000										
Gurlan	0.700	0.276	0.401	0.000									
Qushkupir	0.609	0.402	0.140	0.368	0.000								
Urgench	0.329	0.623	0.140	0.461	0.280	0.000							
Khazarasp	0.523	0.532	0.090	0.466	0.131	0.203	0.000						
Tuproqqala	0.550	0.643	0.550	0.368	0.609	0.506	0.640	0.000					
Honka	0.570	0.353	0.183	0.221	0.163	0.272	0.245	0.454	0.000				
Khiva	0.546	0.499	0.093	0.442	0.098	0.222	0.033	0.635	0.223	0.000			
Shavat	0.398	0.556	0.074	0.410	0.211	0.071	0.153	0.506	0.208	0.164	0.000		
Yangiariq	0.729	0.248	0.418	0.031	0.376	0.485	0.479	0.394	0.236	0.454	0.432	0.000	
Yangibazar	0.688	0.352	0.450	0.089	0.439	0.487	0.524	0.292	0.281	0.504	0.446	0.107	0.000

Source: Author estimations

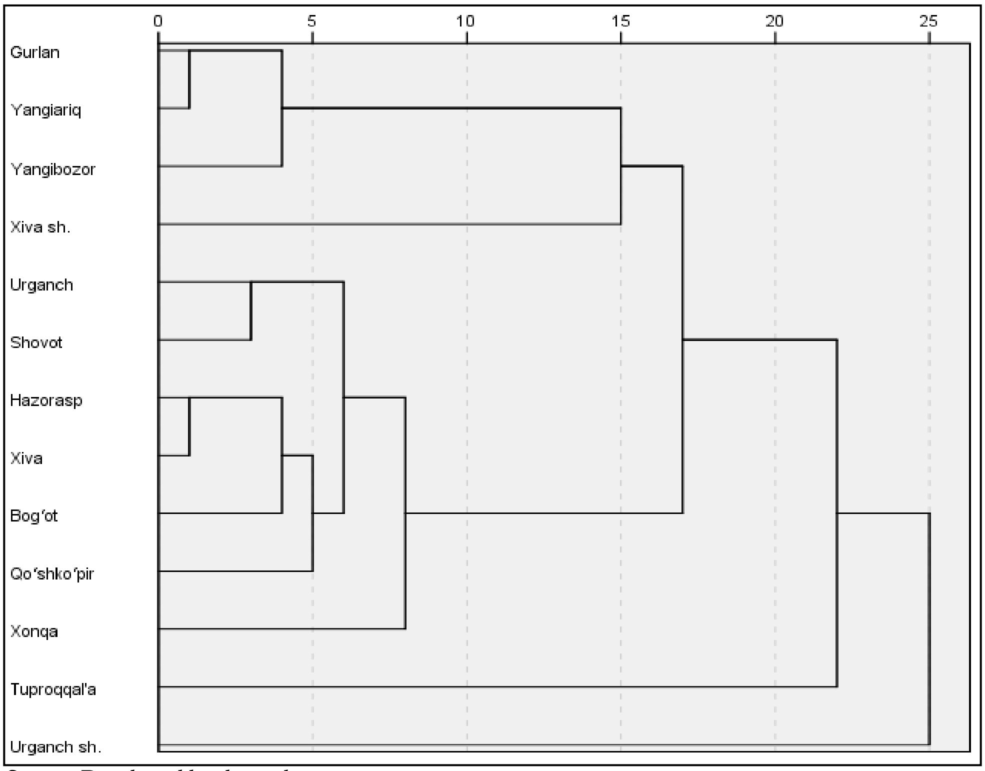
Regarding the Khazarasp district, the most distant districts are Tuproqqala (0.640), Yangibazar (0.524) and Yangiariq (0.479), whereas the nearest districts are Khiva (0.033), Shavat (0.153) and Honka (0.245).

The analysis concerning Tuproqqala district yielded the following results. The nearest districts in terms of distance are Yangibazar (0.292) and Yangiariq (0.394), whereas the farthest are Khiva (0.635), Shavat (0.506) and Honka (0.454). As evident from this configuration, it indicates that Tuproqqala district forms a distinct cluster (class) by itself.

Based on the analysis of distances relative to Honka, the most distant districts are Yangibazar (0.281) and Yangiariq (0.236), whereas the closest distances were recorded in the Shavat (0.208) and Khiva (0.223) regions.

It is evident that the districts of Urgench and Shavat (0.071), Qushkupir-Khiva (0.098), and Khazarasp-Khiva (0.033) are located in close spatial proximity to one another.

Based on the aforementioned circumstances, the clustering of cities and districts in the Khorezm region, delineated according to indicators of sustainable socio-economic development, is illustrated, see Figure 1.



Source: Developed by the author

Figure 1. Dendrogram representation of cities and districts in the Khorezm region based on indicators of sustainable socio-economic development.

According to the dendrogram illustrated in Figure 1, the initial formation involved the Gurlan-Yangiariq-Yangibazar cluster, followed by the Khazarasp-Khiva-Bagat cluster, the Urgench-Shavat cluster, Qushkupir, Honka, Tuproqqala, Urgench city and Khiva city, resulting in a total of 8 clusters. Subsequently, based on similarity levels, the Khazarasp-Khiva-Bagat cluster, the Urgench-Shavat cluster, Qushkupir and Honka clusters merged to form a single new cluster. In this manner, at the final stage, five clusters of the cities and districts in the Khorezm region were established and arranged in the following order:

1. Urgench city;
2. Khiva city;
3. Urgench, Bagat, Khazarasp, Shavat, Honka, Khiva, Qushkupir;
4. Gurlan, Yangibazar, Yangiariq;

5. Tuproqqala.

4. Conclusion

The analyses conducted in the present research endeavor have facilitated the formulation of the following conclusions. In particular:

1. Identifying disparities between urban and rural districts based on contemporary approaches to exploring emerging trends in the sustainable socio-economic development of the region, and devising scientifically substantiated measures to mitigate these disparities, constitutes a pertinent issue. Undoubtedly, this process necessitates an in-depth analysis of socio-economic development indicators across the region's cities and districts.
2. Based on the indicators of sustainable socio-economic development in the region, the similarity levels in cluster analysis were computed using the “Euclidean metric”.
3. Based on the similarity measures computed using the “Euclidean metric”, the region was partitioned into the following clusters: (Urgench city), (Khiva city), (Urgench, Bagat, Khazarasp, Shavat, Honka, Khiva, Qushkupir), (Gurlan, Yangibazar, Yangiariq), and (Tuproqqala).

REFERENCES

[1] International Monetary Fund, *World Economic Outlook: Policy Pivot, Rising Threats*, Washington, DC, Oct. 2024, pp. 9–10.

[2] Sh. M. Mirziyoyev, *Yangi O‘zbekiston Strategiyasi*, Toshkent: O‘zbekiston, 2021, pp. 131–170.

[3] D. Cui, Y. Yu, and Z. Song, "Cluster Analysis of Regional Economic Development in Hebei Province," in *Engineering Education and Management*, L. Zhang and C. Zhang, Eds., Lecture Notes in Electrical Engineering, vol. 111, Springer, Berlin, Heidelberg, 2012.

[4] I.-A. Popescu and M. Neagu, "Cluster analysis of regional research and development disparities in Europe," *Studies in Business and Economics*, vol. 15, no. 3, pp. 303–312, 2020.

[5] S. K. V. Krishnan and M. F. C. Mohammed, "Assessment of regional environmental quality using cluster analysis," *Environmental Development*, vol. 45, 2023.

[6] R. R. Stough and J. Yu, "Industrial Cluster Analysis, Entrepreneurship and Regional Economic Development," in *Food Security and Industrial Clustering in Northeast Asia*, L. Kiminami and T. Nakamura, Eds., New Frontiers in Regional Science: Asian Perspectives, vol. 6, Springer, Tokyo, 2016.

[7] M. Çağlar and C. Gürler, "Sustainable Development Goals: A cluster analysis of worldwide countries," *Environment, Development and Sustainability*, vol. 24, no. 6, pp. 8593–8624, 2022.

[8] L. M. Akimova, N. L. Khomiuk, I. M. Bezena, I. L. Lytvynchuk, and O. Petroye, "Planning of socio-economic development of the territories (experience of European Union)," *International Journal of Management*, vol. 11, no. 4, 2020.

[9] M. C. Nogueira and M. Madaleno, "Are international indices good predictors of economic growth? Panel data and cluster analysis for European Union countries," *Sustainability*, vol. 13, no. 11, p. 6003, 2021.

[10] S. E. Barykin, A. A. Mikheev, E. G. Kiseleva, Y. E. Putikhin, N. S. Alekseeva, and A. Mikhaylov, "An empirical analysis of Russian regions’ debt sustainability," *Economies*, vol. 10, no. 5, p. 106, 2022.

[11] Z. Arshad, M. Robaina, and A. Botelho, "The role of ICT in energy consumption and environment: an empirical investigation of Asian economies with cluster analysis," *Environmental Science and Pollution Research*, vol. 27, no. 26, pp. 32913–32932, 2020.

[12] R. Huseynov, N. Aliyeva, V. Bezpalov, and D. Syromyatnikov, "Cluster analysis as a tool for improving the performance of agricultural enterprises in the agro-industrial sector," *Environment, Development and Sustainability*, vol. 26, no. 2, pp. 4119–4132, 2024.