

# FORMATION OF THE FEATURE SPACE AFFECTING THE LEVEL OF URBANIZATION USING PRINCIPAL COMPONENT ANALYSIS

**Ilxom Tursunbayevich Ismailov**

Senior Lecturer, Department of Information Systems and Technologies, Tashkent State Agrarian University

[ilxomismailov1988161287@gmail.com](mailto:ilxomismailov1988161287@gmail.com)

**Shoxsanam Ermamatovna Ermamatova**

Assistant Professor, Department of Digital Economy, Tashkent State University of Economics

[shoxsanameramatova04@gmail.com](mailto:shoxsanameramatova04@gmail.com)

## ABSTRACT

This article is dedicated to studying the factors influencing the level of urbanization using the example of Samarkand region. The study analyzed 96 socio-economic, demographic, and infrastructural indicators for the period 2000-2022. Using the Principal Component Analysis (PCA) method, these factors were reduced to 25 main components. The results identified that economic development, demographic changes, and social infrastructure have the strongest impact on the urbanization process. The article emphasizes the need for a comprehensive approach to effectively manage urbanization processes in Samarkand region. The research findings have not only scientific but also practical significance and can be applied in developing urban development strategies.

**Keywords:** Urbanization, Samarkand region, PCA, urbanization, demography.

## INTRODUCTION

The process of urbanization is one of the most significant socio-economic phenomena in the modern world, profoundly affecting all aspects of society [1]. This complex process is characterized by the concentration of population and economic activities in urban areas. Urbanization involves not only changes in the distribution of population between urban and rural areas but also a comprehensive transformation that impacts all spheres of society [2].

Particularly in developing countries, including Uzbekistan, urbanization processes are occurring at a rapid pace. While this process stimulates economic growth and innovation on one hand, it also poses serious challenges for the environment, infrastructure, and social systems

on the other. To quantitatively assess the level of urbanization, the following formula is typically used:

$$\text{Urbanization\_level} = \frac{\text{Urban\_population}}{\text{Total\_population}} * 100\%. \quad (1).$$

This indicator, determined by this formula, reflects the degree of development of urbanization processes in a country or region and allows for comparison of urbanization status across different areas [3].

Samarkand region, one of the most ancient and culturally significant areas of Uzbekistan, has also been experiencing active urbanization processes in recent years [4]. An in-depth study of the characteristics of these processes and their determining factors is relevant not only at the local but also at the national level. The experience of Samarkand region can provide valuable lessons for other regions as well.

The factors influencing urbanization processes are numerous and diverse. Among them, socio-economic [5], demographic, and infrastructural indicators [6] hold a special place. However, analyzing such a large number of factors simultaneously is a complex task. Therefore, the use of modern statistical methods, particularly Principal Component Analysis (PCA), to reduce the number of factors and identify the most significant components is of great importance [7].

The main objective of this study is to identify factors affecting the level of urbanization using the example of Samarkand region and to determine the most important components by reducing them through PCA. To achieve this goal, the following tasks have been set:

1. Collect and analyze socio-economic, demographic, and infrastructural indicators affecting urbanization processes in Samarkand region from 2000 to 2022;
2. Reduce the collected 96 factors to 25 main components using the PCA method;
3. Assess and analyze the impact of the identified main components on the level of urbanization;
4. Develop recommendations for effective management of urbanization processes in Samarkand region based on the obtained results.

The results of this study have not only scientific but also practical significance and can be applied by local and central government bodies in developing urban development strategies. Moreover, this research methodology can be applied in other regions and provinces, allowing for a deeper understanding and more effective management of urbanization processes in Uzbekistan.

## LITERATURE REVIEW

In their co-authored article, Jolliffe, I. T., and Cadima provide a comprehensive analysis of the fundamentals of PCA, its



recent developments, and applications in various fields. They emphasize the importance of PCA in reducing data dimensionality and simplifying complex datasets. The article presents practical applications of PCA in various fields such as biology, economics, and social sciences [8].

Li, X., Zhou, and others use the PCA method to study urbanization processes and predict the growth of future urban areas. They analyzed numerous socio-economic and demographic indicators to identify the main factors influencing urbanization [9].

Abdi, H., and Williams, L. present a deep theoretical foundation and practical application of the PCA method in their article "Principal component analysis. Wiley interdisciplinary reviews: computational statistics". The authors explain in detail the mathematical foundations of PCA, its geometric interpretation, and applications in various fields. They also discuss important points to consider when applying PCA and the limitations of the method. This source can be useful for correctly applying the PCA method in our study and properly interpreting the results [10].

## METHODOLOGY

For this study, 96 factors affecting urbanization processes in Samarkand region between 2000 and 2022 were collected and analyzed. This data was obtained from the official website "www.stat.uz" of the State Statistics Committee of the Republic of Uzbekistan. The collected data includes comprehensive indicators reflecting the multifaceted nature of the urbanization process [11].

The factors in the dataset are mainly divided into four main categories: demographic, social, economic, and infrastructural indicators. Demographic factors include population size, birth and death rates, and migration indicators. These factors allow observing the dynamics of urban and rural populations [12].

Social factors include indicators reflecting the living standards of the population. These include income levels, employment rates, achievements in education and healthcare. The number and quality of social infrastructure objects are also taken into account [13].

Among economic factors are indicators such as GDP, volume of industrial and agricultural products, and amount of investments. This data reflects the economic development trends of Samarkand region and shows their impact on urbanization processes [14].

Infrastructural factors include indicators from the transport system (length of roads, passenger traffic volume), communication means (number of internet users), electricity supply, and construction sector. This



data reflects factors that directly influence the expansion of urban areas and the growth of urbanization levels [15].

This comprehensive dataset allows for in-depth analysis of the dynamics of urbanization processes in Samarkand region and identification of key factors influencing this process. Such an approach creates a solid foundation for comprehensive study of urbanization processes and forecasting future trends.

Principal Component Analysis (PCA) is a statistical method used to analyze multidimensional data and reduce its dimensionality [16]. This method is particularly important in studying urbanization processes as the factors influencing urbanization are numerous and diverse.

Urbanization is a complex socio-economic process influenced by demographic, economic, social, and infrastructural factors. We examined 96 different factors using the example of Samarkand region. Analyzing such a large number of factors simultaneously and determining their relative importance is a challenging task. This is where the PCA method demonstrates its effectiveness.

The analysis of factors influencing urbanization using the PCA algorithm consists of the following stages:

Stage 1: Data Preparation

$$X = \begin{bmatrix} x_{11}, x_{1,2}, \dots, & x_{1,96} \\ x_{2,1}, x_{2,2}, \dots, & x_{2,96} \\ \dots & \dots \\ x_{23,1}, x_{23,2}, \dots, & x_{23,96} \end{bmatrix}$$

Where X represents 96 indicators of Samarkand region from 2000 to 2022

Stage 2: Data Standardization

$$Z = \frac{(X - \mu)}{\sigma}$$

Where Z is the standardized matrix,  $\mu$  is the mean value vector for each column,  $\sigma$  is the standard deviation vector for each column

Stage 3: Calculating the Covariance Matrix

$$C = \frac{1}{22} * (Z^T) * Z$$

Where C is the covariance matrix,  $Z^T$  is the transpose of matrix Z

Stage 4: Calculating Eigenvalues and Eigenvectors

$$C * v = \lambda * v$$

Where v is the eigenvector,  $\lambda$  is the eigenvalue

Stage 5: Identifying Principal Components

$$\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \dots \geq \lambda_{96}$$

Eigenvectors represent the principal components. We order them by their eigenvalues

Stage 6: Selecting Principal Components

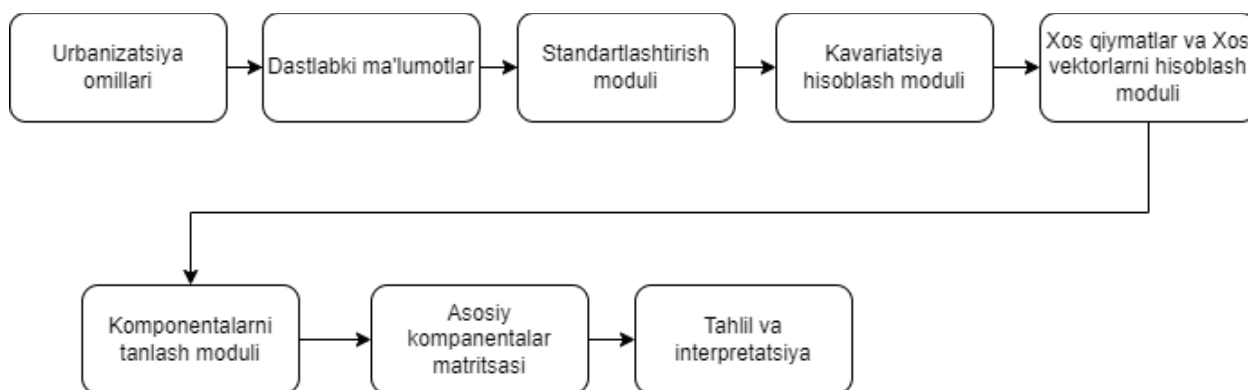
$$EVR = \frac{\lambda_i}{\sum \lambda_j}$$

Where  $\sum \lambda_j$  is the sum of all eigenvalues

Stage 7: Forming the Principal Component Matrix

$$W = [\lambda_1, \lambda_2, \dots, \lambda_{25}]$$

Thus, the PCA method allows for efficient analysis of large volumes of data in studying urbanization processes, identifying the most important factors, and understanding their interactions. This, in turn, is crucial in developing urban development strategies and managing urbanization processes. The architecture for identifying factors affecting the level of urbanization using the PCA method is as follows:



Architecture for reducing factors affecting the level of urbanization using PCA

## RESULTS AND DISCUSSIONS

A deep analysis was conducted on 96 factors influencing urbanization processes in Samarkand region. This analysis revealed several important trends.

Among demographic factors, it was observed that the growth rate of urban population is higher than that of rural population. This, in turn, indicates an intensification of rural-to-urban migration processes. Additionally, the increase in the proportion of young population emerged as a factor accelerating urbanization processes. Analysis of economic factors showed that the share of service sector and industrial production is increasing in Samarkand region. This is leading to the development of urban economy and creation of new job opportunities. The increase in investment volume is also stimulating the development of urban infrastructure and construction of new residential complexes.

Among social factors, a significant increase in the quality and coverage of education and healthcare services was observed.



These factors are increasing the attractiveness of urban life and attracting rural population. Analysis of infrastructural factors showed improvement in the transport system, especially connections with surrounding areas. Moreover, the expansion of internet and mobile communication services is also contributing to the improvement of urban life quality.

As a result of applying the PCA method, the initial 96 factors were reduced to 25 main components. These components explain 85% of the variation in the initial data, ensuring high accuracy of the analysis. The first principal component is associated with overall economic development, including indicators such as GDP, industrial production, and investment volume. The second component reflects demographic changes and combines population size, birth rates, and migration indicators. The third component is related to social infrastructure development, including the number and quality of educational and healthcare institutions.

Analysis of the impact of principal components on the level of urbanization showed that the economic development component has the strongest influence. This indicates that urbanization processes in Samarkand region are mainly driven by economic factors. The demographic change component ranks second, confirming the important role of population dynamics. Although the social infrastructure component ranks third, its importance is increasing year by year. This trend shows that cities need to be attractive not only economically but also socially.

These results demonstrate the need for a comprehensive approach in managing urbanization processes in Samarkand region. Along with stimulating economic development, it is important to improve social infrastructure and take into account demographic changes. This approach will help ensure sustainable and balanced urban development.

## CONCLUSION

This study allowed for an in-depth analysis of urbanization processes in Samarkand region. Using Principal Component Analysis (PCA), we reduced 96 different factors to 25 main components, identifying the factors that have the strongest impact on the level of urbanization.

The research results allow us to draw the following main conclusions:

1. Economic development is the main driving force of the urbanization process. GDP growth, increase in industrial production and investment volumes are directly affecting the rates of urbanization.
2. Demographic changes, particularly the increase in young population and rural-to-urban migration processes, are accelerating urbanization.

3. The development of social infrastructure, especially the improvement in quality and coverage of education and healthcare services, is increasing the attractiveness of urban life.

4. The improvement of transport and communication systems is strengthening the links between urban and rural areas, further activating urbanization processes.

Based on these conclusions, the following recommendations can be made:

1. Apply a comprehensive approach in developing urban development strategies, balancing economic, social, and infrastructural factors.

2. Develop rural development programs to manage urbanization processes and reduce the gap between urban and rural areas.

3. Expand education and employment opportunities for young people, thereby facilitating their adaptation to urban life.

4. Consider ecological aspects in developing urban infrastructure, implementing the concept of "green cities".

5. It is recommended to implement modern information technologies, including "smart city" systems, for future monitoring and management of urbanization processes.

The results and recommendations of this study can be useful not only for Samarkand region but also for other regions of Uzbekistan. In the future, it would be appropriate to continue research in this direction, conduct comparative analyses with other regions, and develop long-term forecasts.

## REFERENCES

1. Chen, M., Zhang, H., Liu, W., & Zhang, W. (2014). The Global Pattern of Urbanization and Economic Growth: Evidence from the Last Three Decades. *PLoS ONE*, 9(8), e103799.
2. Turok, I., & McGranahan, G. (2013). Urbanization and economic growth: the arguments and evidence for Africa and Asia. *Environment and Urbanization*, 25(2), 465-482.
3. Uchiyama, Y., & Mori, K. (2017). Methods for specifying spatial boundaries of cities in the world: The impacts of delineation methods on city sustainability indices. *Science of The Total Environment*, 592, 345-356.
4. Abdullaev, I. (2020). O'zbekistonda urbanizatsiya jarayonlari: hududiy xususiyatlar va rivojlanish tendensiyalari. *O'zbekiston Geografiya jamiyati axboroti*, 55, 14-18.
5. Chen, M., Gong, Y., Lu, D., & Ye, C. (2019). Build a people-oriented urbanization: China's new-type urbanization dream and Anhui model. *Land Use Policy*, 80, 1-9.



6. Jedwab, R., Christiaensen, L., & Gindelsky, M. (2021). Demography, urbanization and development: Rural push, urban pull and... urban push? *Journal of Urban Economics*, 124, 103350.
7. Bro, R., & Smilde, A. K. (2014). Principal component analysis. *Analytical Methods*, 6(9), 2812-2831.
8. 1. Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2065), 20150202.
9. 2. Li, X., Zhou, Y., Eom, J., Yu, S., & Asrar, G. R. (2019). Projecting global urban area growth through 2100 based on historical time series data and future shared socioeconomic pathways. *Earth's Future*, 7(4), 351-362.
10. 3. Abdi, H., & Williams, L. J. (2010). Principal component analysis. *Wiley interdisciplinary reviews: computational statistics*, 2(4), 433-459.
11. O'zbekiston Respublikasi Prezidenti huzuridagi davlat statistika agentligi. (2023). Rasmiy statistika. <https://stat.uz>
12. Farrell, K. (2017). The rapid urban growth triad: a new conceptual framework for examining the urban transition in developing countries. *Sustainability*, 9(8), 1407.
13. Avis, W. R. (2016). *Urban Governance (Topic Guide)*. Birmingham, UK: GSDRC, University of Birmingham.
14. Shatkin, G. (2016). The real estate turn in policy and planning: Land monetization and the political economy of peri-urbanization in Asia. *Cities*, 53, 141-149.
15. Cobbinah, P. B., Erdiaw-Kwasie, M. O., & Amoateng, P. (2015). Africa's urbanisation: Implications for sustainable development. *Cities*, 47, 62-72.
16. Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2065), 20150202.