SILESIAN UNIVERSITY OF TECHNOLOGY PUBLISHING HOUSE

SCIENTIFIC PAPERS OF SILESIAN UNIVERSITY OF TECHNOLOGY 2024
ORGANIZATION AND MANAGEMENT SERIES NO. 213

CLASSIFICATION OF LARGE EUROPEAN CITIES BASED ON THE RATINGS OF THEIR INHABITANTS

Adam SOJDA

Silesian University of Technology, Faculty of Organization and Management, Department of Economics and Informatics; adam.sojda@polsl.pl, ORCID: 0000-0002-3021-4451

Purpose: Trends indicate an increasing share of the population living in an urbanized world. It is important to provide them with appropriate living conditions. Most often, we describe living conditions using objective indicators. We hope that they reflect the expectations of the residents well. What if this is not the case? The aim of the research presented in the article is to provide a classification of large European cities based on the opinions of their inhabitants.
Design/methodology/approach: Eurostat provides data from the Perception Survey on the Quality of Life in European Cities. The research is carried out every four years. The last was in 2023.
Findings: The existence of three types of large European cities covered by the study was indicated.
Research limitations/implications: The results of the research are based on subjective assessments of the residents, perhaps they should be verified with objective factors.
Social implications: A work to look at cities through the eyes of their inhabitants. It shows the similarities in their perception of the conditions resulting from the location of the city and the economic conditions of the state.
Originality/value: Application of clustering methods to the analyzed data, determination of groups of similar cities. Defining the dimensions in the PCA analysis that cities can be described.
Keywords: PCA, K-means, Eurostat, European Cities, Quality of Life.
Category of the paper: Research paper.

1. Introduction

In 2023, the degree of urbanization worldwide was 57 percent. North America, Latin America, and the Caribbean had the highest level of urbanisation, about 83 percent. Next was Europe, with 75 percent. North America is the most urbanized continent, but Tokyo-Yokohama in Japan was the largest urban area in the world that year, with 37.7 million inhabitants.

Cities are the future of the population. The possibility of using modern technologies has become the foundation of the Smart City concept. It was quickly noticed that an IT-only approach was not enough. This can be seen by following the development of the Smart City concept: (Cohen, 2015; Svítek et al., 2020; Kinelski, 2022; Kuzior, 2024):

- Smart City 1.0 – the ICT sector offers its products for cities.
- Smart City 2.0 – cities are the initiators of ICT implementation.
- Smart City 3.0 – city citizens take over the initiative to implement ICT solutions.
- Smart City 4.0 – knowledge sharing modern technologies ensure sustainable development of urban areas and their inhabitants.
- Smart City 5.0 – a multi-agent ecosystem of smart services allows for a harmonious balance of various aspects of residents' lives.

The diversity of the approach to Smart City has allowed us to dimension of smart city. The classic ones include (Cohen, 2015; Giffinger, Gudrun, 2010; Marchlewska-Patyk, 2023):

- economy – actions aimed at transforming and strengthening the city's economy,
- environment – environmental management to improve living standards and reduce civilization pollution,
- government – interaction between the city authorities and all stakeholders – citizens, entrepreneurs, civil society organizations,
- living – improving quality of life, social and digital exclusion, safety and care,
- mobility – urban transport services, improving the flow of people, goods and services in the city,
- people – appropriate forms of education, career opportunities in labor markets.

The appearance of dimensions coincided with the appearance of rankings evaluating cities. These tools can be used to plan development, assess your strengths and weaknesses against other similar cities. A high position in the ranking is what the city can boast about in order to attract new investors and develop tourism (Akande et al., 2019; Pangsy-Kania, Kania, 2024; Berger, 2019; Toh, 2022; Vanli, 2024).

The need to take into account the socio-economic dimension has been noted in the development (Jonek-Kowalska, 2019). In this dimension, it is the human being who is most important for the development of the city. His needs are in the foreground. Large cities are the location of higher education centers. Graduates very often stay in the city where they studied, creating a base of people that should play a significant role in the development of the city. The city should meet their living needs as well as further development.

2. Data structure

The Perception Survey on the Quality of Life in European Cities was conducted in 79 European cities. It covered all capitals of the countries studied (except Switzerland) and one to six additional cities in larger countries. About 500 residents were interviewed in each city, and 835 interviews were collected in each city. Targets were set at a minimum of 100 online interviews per city and a maximum of 735 via telephone. In some cities, more online interviews were collected.

The Eurostat database includes 82 European cities with at least 500,000 inhabitants. Not all of them are included in the quality of life survey. By limiting the original collection of cities the following cities were received: Amsterdam, Ankara, Antalya, Antwerpen, Athens, Barcelona, Berlin, Bordeaux, Brussel, Bucharest, Budapest, Copenhagen, Diyarbakir, Dortmund, Dublin, Essen, Geneve, Glasgow, Hamburg, Helsinki, Istanbul, Krakow, Leipzig, Lille, Lisbon, London, Madrid, Malaga, Manchester, Marseille, Munchen, Napoli, Oslo, Palermo, Paris, Prague, Riga, Rome, Rotterdam, Sofia, Stockholm, Stuttgart, Torino, Vilnius, Warsaw, Wien, Zurich. Not all cities can be considered strictly European, but it was decided that they should be left in the study due to their membership in the EU or NATO community.

Due to the specificity of the available data, new areas of city assessment have been defined.

- **ECONOMY** (ECO) – questions about work, real estate prices, the financial and material situation of the household.
- **ENVIRONMENT** (ENV) – here are the assessments of green space, noise and air quality.
- **GOVERNANCE** (GOV) – assessment of satisfaction with solving local problems, procedures applied by the city authorities, information and administrative services, corruption of local authorities.
- **HUMAN CAPITAL** (HUC) – sports and cultural activities as well as facilities offered by the city, education.
- **QUALITY OF LIFE** (QLI) – trust in other residents, satisfaction with life in the city, assessment of the city as a place to live.
- **SOCIAL COHESION INDICATORS** (SCI) – health care, safety, friendliness towards immigrants, minorities, LGBT communities, non-material help.
- **TRANSPORT** (TRN)- diverse evaluation of urban transport.

Based on the Eurostat database, the following set of indicators has been proposed (table 1).

Table 1.
Indicators for the assessment of cities

ID	Questions with answer variants and weights
ECO_01	In this city it is easy to find a good job: [2, 1, -1, -2, 0]
ECO_02	In this city, it is easy to find good housing at a reasonable price: [2, 1, -1, -2, 0]
ECO_03	The financial situation of your household: [2, 1, -1, -2, 0]
ECO_04	If you needed material help (e.g. money, loan or an object) you could receive it from relatives, friends, neighbours or other persons you know: [1, 0, 0]
ECO_05	Your personal job situation: [2, 1, -1, -2, 0]
ECO_06	Within the last 12 months, would you say you had difficulties to pay your bills at the end of the month: [-2, -1, 1, 0]
ENV_01	Green spaces such as public parcs or gardens: [2, 1, -1, -2, 0]
ENV_02	The quality of the air in the city: [2, 1, -1, -2, 0]
ENV_03	The noise level in the city: [2, 1, -1, -2, 0]
ENV_04	The cleanliness in the city: [2, 1, -1, -2, 0]
GOV_01	I am satisfied with the amount of time it takes to get a request solved by my local public administration: [2, 1, -1, -2, 0]
GOV_02	The procedures used by my local public administration are straightforward and easy to understand: [2, 1, -1, -2, 0]
GOV_03	Information and services of my local public administration can be easily accessed online: [2, 1, -1, -2, 0]
GOV_04	There is corruption in my local public administration: [-2, -1, 1, 2, 0]
GOV_05	The fees charged by my local public administration are reasonable: [2, 1, -1, -2, 0]
HUC_01	Sports facilities such as sport fields and indoor sport halls in the city: [2, 1, -1, -2, 0]
HUC_02	Cultural facilities such as concert halls, theatres, museums and libraries in the city: [2, 1, -1, -2, 0]
HUC_03	Schools and other educational facilities: [2, 1, -1, -2, 0]
QLI_01	Generally speaking, most people in this city can be trusted: [2, 1, -1, -2, 0]
QLI_02	Most people in my neighbourhood can be trusted: [2, 1, -1, -2, 0]
QLI_03	I'm satisfied to live in this city: [2, 1, -1, -2, 0]
QLI_04	The neighbourhood where you live: [2, 1, -1, -2, 0]
QLI_05	For people in general: a good place to live; not a good place to live; don't know/no answer/refuses: [2, 1, 0]
QLI_06	Public spaces in this city such as markets, squares, pedestrian areas: [2, 1, -1, -2, 0]
QLI_07	The life you lead: [2, 1, 1, -2, 0]
QLI_08	How is your health: [2, 1, 0, -1, -2, 0]
QLI_09	How much of the time, during the past 4 weeks, have you been feeling lonely: [-5, -4, -4, -2, 0, -1, 0]
QLI_10	How much of the time, during the past 12 months, have you been feeling lonely: [-5, -4, -3, -2, 0, -1, 0]
QLI_11	Compared to five years ago quality of life in your city or area has: [-2, 0, 2, 0]
SCI_01	Health care services, doctors and hospitals: [2, 1, -1, -2, 0]
SCI_02	I feel safe walking alone at night in my city: [2, 1, -1, -2, 0]
SCI_03	For racial and ethnic minorities: [-1, -2, 0]
SCI_04	For gay or lesbian people: [2, 1, 0]
SCI_05	For immigrants from other countries: [-1, -2, 0]
SCI_06	Confidence in the local police force: [1, 0, 0]
SCI_07	Money or property stolen from you or another household member in your city the last 12 months: [0, 1, 0]
SCI_08	Being assaulted or mugged in your city the last 12 months: [0, 1, 0]
SCI_09	If you needed non material help (e.g. somebody to talk to, help with doing something or collecting something) you could receive it from relatives, friends, neighbours or other persons you know: [1, 0, 0]
SCI_10	I feel safe walking alone at night in my neighbourhood: [2, 1, -1, -2, 0]
SCI_11	For young families with children: [2, -2, 0]
SCI_12	For elderly people: a good place to live: [2, -2, 0]
TRP_01	Public transport in the city, for example bus, tram or metro: [2, 1, -1, -2, 0]

Cont. table 1.

TRP_02	Means of transport most often used: car; motorcycle; bicycle; foot; train; urban public transport; other; do not commute; don't know / no answer / refuses: [0, 0, 0, 0, 0, 1, 0, 0, 0]
TRP_03	Public transport affordable: [2, 1, -1, -2, 0]
TRP_04	Public transport safe: [2, 1, -1, -2, 0]
TRP_05	Public transport easy to get: [2, 1, -1, -2, 0]
TRP_06	Public transport frequent (comes often): [2, 1, -1, -2, 0]
TRP_07	Public transport reliable (comes when it says it will): [2, 1, -1, -2, 0]

Source: Eurostat, The Perception Survey on the Quality of Life in European Cities – questions.

3. Methods

The answers to the questions are predominantly on a 5-point Likert scale. Each question is accompanied by the weights applied to each answer. The database shows the percentage of citizens choosing individual categories. Weighted answers to each question were determined, thus obtaining the indicator's value. The scales have been selected so that a larger value of the indicator shows the greater importance of the indicator.

Then, the obtained values were normalized using a formula that took into account the worst and best assessments in each of the studied periods.

The formula used is min-max normalisation:

$$norm(x_i) = (100 - 50)\frac{x_i - \min\limits_k(x_k)}{\max\limits_k(x_k) - \min\limits_k(x_k)} + 50 \tag{1}$$

After the first stage of normalization and determination of the mean value of the criterion, the criterion was normalized again.

$$u(x_i) = \frac{x_i - mean(x)}{sd(x)} \tag{2}$$

where:

$$mean(x) = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$sd(x) = \frac{\sum_{i=1}^{n}(x_i - mean(x))^2}{n-1}$$

3.1. PCA

PCA allows you to reduce the number of variables in your data set, simplifying data analysis and visualization without losing important information. With PCA, you can identify and remove interdependent variables that are redundant and do not add additional value to your analysis. This method creates new, independent variables, called principal components, which are linear combinations of the original variables. This is a technique commonly used in

exploratory data analysis. It is used in many fields(Osborne, 2014a, 2014b; Tsoulfidis, Athanasiadis, 2022; Watkins, 2018).

The application of PCA is a multi-step process.

- Step1. Standardize the range of continuous initial variables.
- Step 2. Compute the covariance matrix to identify correlations.
- Step 3. Compute the eigenvectors and eigenvalues of the covariance matrix to identify the principal components.
- Step 4. Create a feature vector to decide which principal components to keep.
- Step 5. Recast the data along the principal component's axes.

There are two statistical measures to assess the factor ability of the data: Kaiser-Meyer-Olkin (KMO) and Bartlett's test of Sphericity. Kaiser-Meyer-Olkin (KMO) Measure of Sampling Adequacy KMO test is a measure that has been intended to measure the suitability of data for factor analysis. The KMO values between 0.8 to 1.0 indicate the sampling is adequate. KMO values between 0.7 to 0.79 are middling and values between 0.6 to 0.69 are mediocre. KMO values less than 0.6 indicate the sampling is not adequate and the remedial action should be taken. If the value is less than 0.5, the results of the factor analysis undoubtedly won't be very suitable for the analysis of the data. Bartlett's Test of Sphericity Bartlett's Test of Sphericity tests the variables are orthogonal. A p-value of < 0.05 indicates that factor analysis can be performed on the data.

3.2. Hierarchical Clustering Analysis

Hierarchical grouping is performed to identify natural groups. Groups are formed based on the similarity between data points. The most common method is to perform the clustering method using the Ward method, which minimizes the variance in each cluster (group). It is important to use an appropriate metric to determine the distance between points. For the distance measure, the standard Euclidean distance is used.

A dendrogram is a key tool for interpreting and visualizing the data structure in the context of clustering. It displays the cluster layout as a nested grouping of objects. It starts with a single cluster with only one object, showing how the clusters connect to each other step by step. Horizontal lines or branches illustrate relationships between or distances between clusters. The height of these lines reflects the difference, the distance between the clusters. A dendrogram is designed to help you understand the relationships between clusters (Gan et al., 2007).

3.3. K-means clustering

K-mean is a cluster analysis method that divides a dataset into (K) groups, clusters. Each data point is assigned to the cluster with the closest centroid. A centroid is a point that is the average of the values of points in a set. This method is one of the most popular clustering methods in machine learning(Gan et al., 2007; Morissette, Chartier, 2013).

The precedent of algorithm is as follows:

- Step 1. Randomly assign K objects from the dataset as cluster centres.
- Step 2. (Reassign) Assign each object to which object is most similar based upon mean values.
- Step 3. Update Cluster means, i.e., recalculate the mean of each cluster with the updated values.
- Step 4. Repeat Step 2 until no change occurs.

3.4. Tools for calculations

IT tools are crucial to carry out analyses and present results, whether in a tabular or graphical version. The R programming language was used to perform analyses and graphs. Additionally, libraries such as eurostat, tidyverse, ggrepel, fpc, ggpolt2, DataExplorer, FactoMineR, factoextra were used.

4. Results and discussion

After downloading data from the Eurostat database using the "*eurostat*" library, cities and variables describing the areas were selected. Due to the presentation of the values of the variables in the Likert scale, the weighted values of the variable were determined according to the weights presented in Table 1. These values were then standardised according to formula (1) and an average value was determined for each criterion. Then, each of the criteria was standardized according to formula (2) using R language procedures.

To assess the suitability of the data for factor analysis, the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy was calculated. The KMO value obtained was 0.83. This suggests that the data is adequate for factor analysis.

Based on the given Measures of Sampling Adequacy (MSA) for each item, here's the interpretation for each variable: ECO (0.76), ENV (0.84), GOV (0.81), HUC (0.86), QLI (0.77), SCI (0.89), TRP (0.87). Overall, all your variables show adequate to excellent sampling adequacy for factor analysis, with MSA values well above the minimum threshold.

Bartlett's Test of Sphericity was performed to examine whether the correlation matrix is an identity matrix, suggesting that the variables are unrelated and unsuitable for structure detection. The test was significant, (χ^2 (21) = 298.94, p < .001), indicating that the variables do indeed share common factors and are suitable for factor analysis. This result supports the appropriateness of proceeding with such an analysis.

The correlation coefficients between the individual criteria were determined, the results are presented in Figure 1.

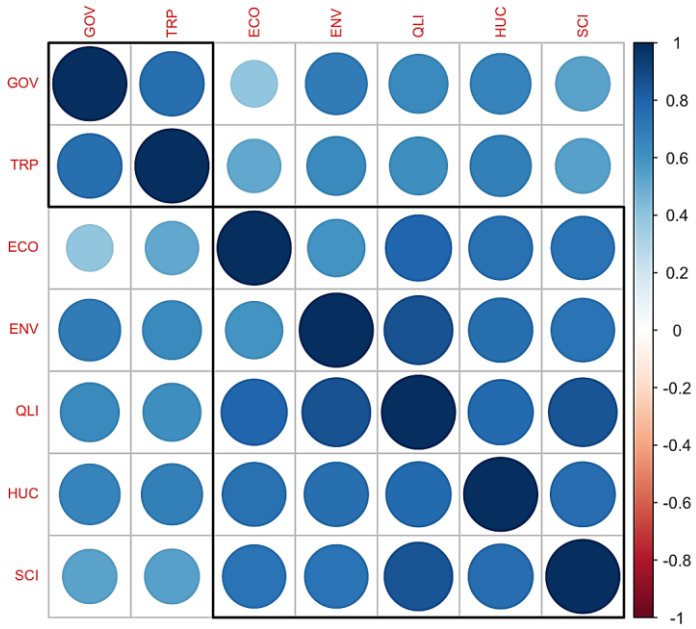

Figure 1. Criteria and correlation coefficients between them grouped hierarchically with two groups selected.

Source: own elaboration.

A scree diagram was made (Figure 2) to determine the number of dimensions used in the analysis. As you can see in the graph, the number of dimensions that will explain more than 90% of the variance is three. Two dimensions explain 86.3% of the variance.

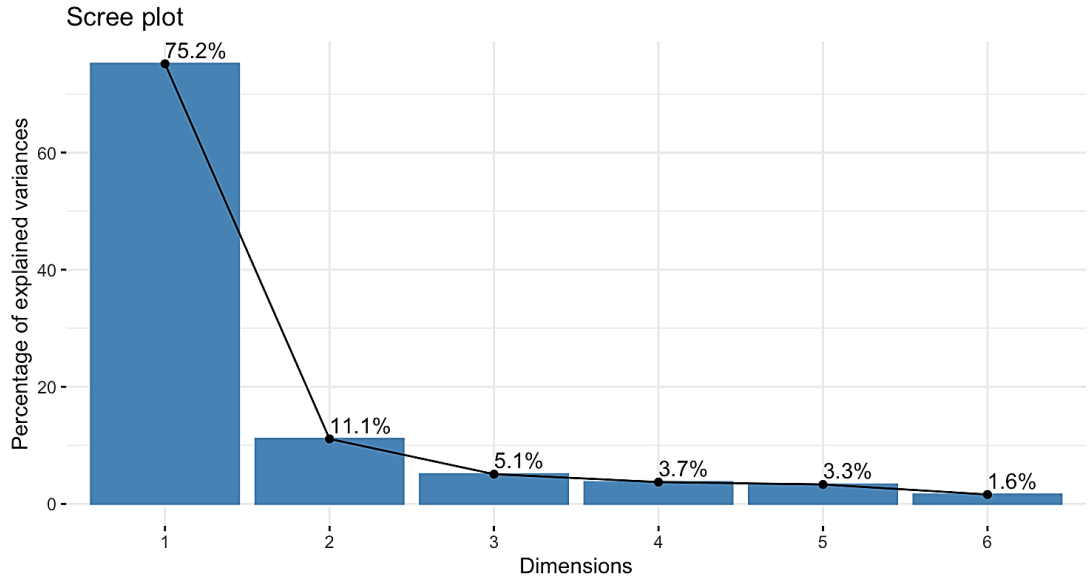

Figure 2. Scree plot - FactoMineR library.

Source: own elaboration.

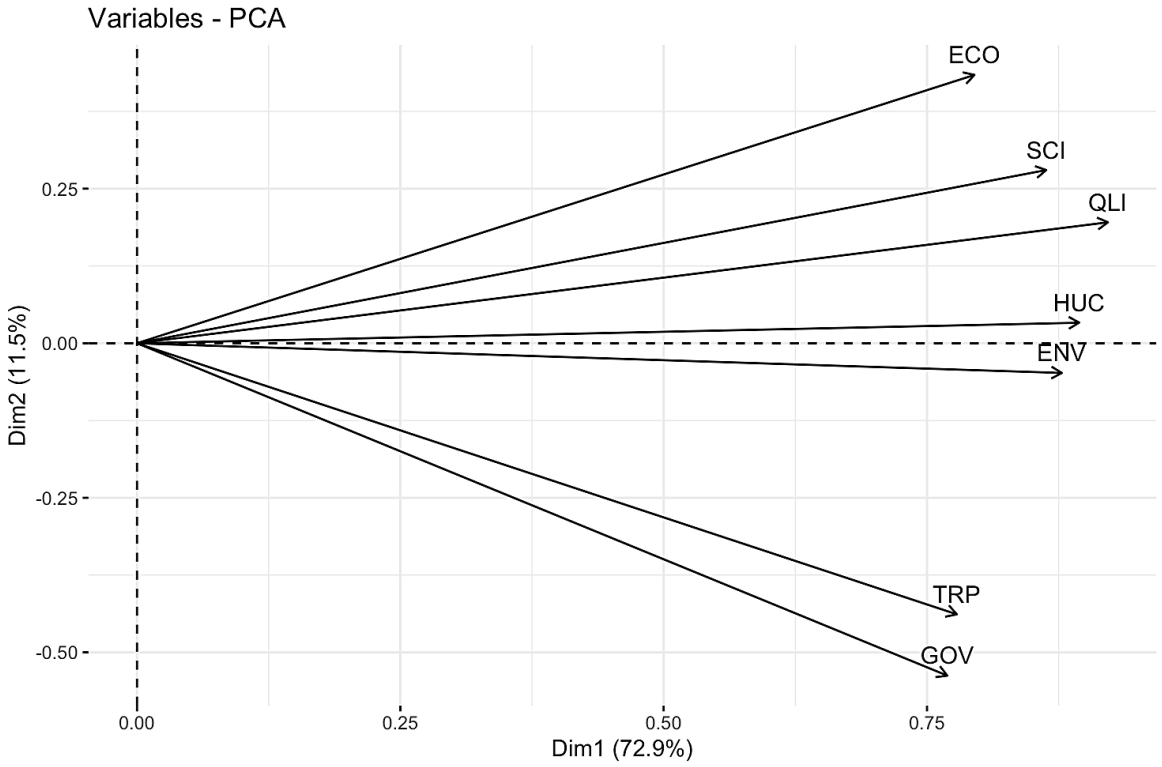

Figure 3. The criteria for the two most important dimensions.

Source: own elaboration.

The relative position of the criteria relative to the two dimensions is shown in Figure 3. The first component, in the light of the questions asked, concerns those elements that relate to matters that you can influence and are close. The second component is matters that you have little influence on, and you can't solve problems easily. The first component can be identified with the quality of one's own life, the second with the quality of management in the city.

A dendogram was performed to confirm the results obtained (Figure 4). At least tree groups of similar cities were marked.

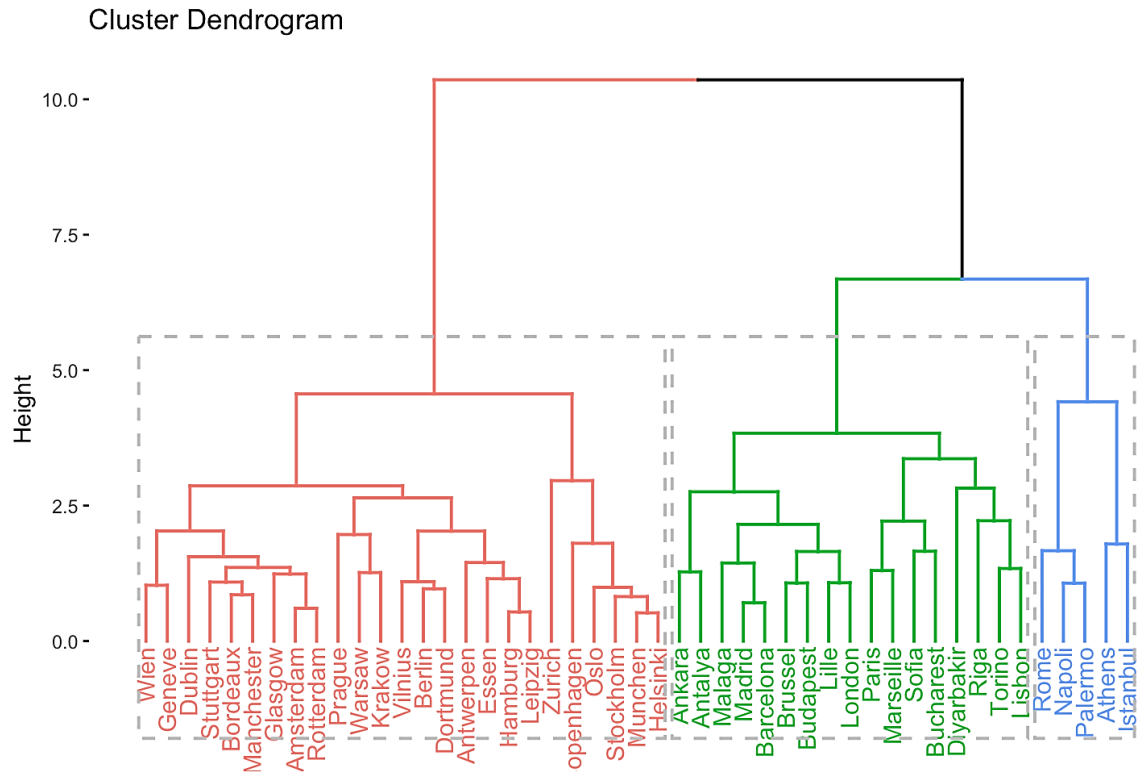

Figure 4. Dendogram with marked groups.

Source: own elaboration.

It was established that three groups of similar cities could be designated. Using the k-means method, cities were assigned to 3 groups. Below, Figure 5 compares the position of cities in relation to two dimensions obtained from PCA. The best city in the PCA assessment in terms of the first component relating to the assessment of the quality of one's own life. In this aspect, Zurich is undoubtedly the best rated by its residents. On the contrary, cities such as Athens, Istanbul and the Italian cities of Palermo, Napoli and Rome are in contrast. The assessment of the city managers shows that the Italian cities of Palermo and Rome are rated the worst. It is puzzling that Turkish cities are the best rated in this respect.

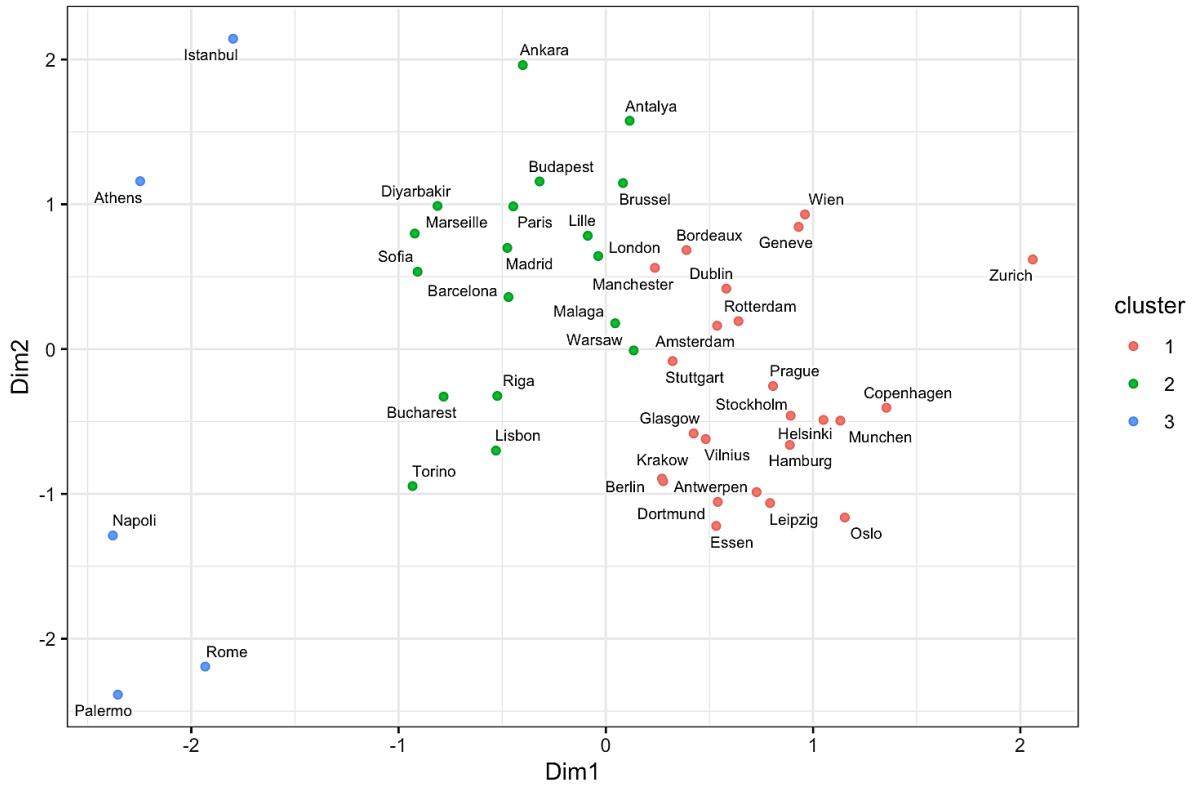

Figure 5. PCA and K-men Combination.

Source: own elaboration.

5. Conclusion

On the basis of the presented analyses concerning large cities in the light of the opinions of their inhabitants, three clearly outlined clusters can be noticed. The first one includes cities such as Zurich, Copenhagen, Munchen, Oslo, Wien. On the opposite side we have Athens, Neapoli, Palermo, Istanbul. Especially the third group clearly stands out from the others. It would be worthwhile to deepen the analysis and add objective indicators to it, such as the income of residents, costs of living, living conditions, etc.

References

1. Akande, A., Cabral, P., Gomes, P., Casteleyn, S. (2019). The Lisbon ranking for smart sustainable cities in Europe. *Sustainable Cities and Society, 44*, pp. 475-487. Retrieved from: https://doi.org/10.1016/j.scs.2018.10.009.
2. Berger, R. (2019). *Smart city strategy index: Vienna and London leading in worldwide ranking.* Retrieved from: https://www.rolandberger.com/publications/publication_pdf/roland_berger_smart_city_breakaway_1.pdf.
3. Cohen, B. (2015). *The 3 Generations of Smart Cities Inside the Development of the Technology Driven City*. Retrieved from: https://www.fastcompany.com/3047795/the-3-generations-of-smart-cities.
4. Giffinger, R., Gudrun, H. (2010). Smart cities ranking: an effective instrument for the positioning of the cities? *ACE: Architecture, City and Environment, 4(12)*, pp. 7-26. Retrieved from: https://doi.org/10.5821/ace.v4i12.2483.
5. Jonek-Kowalska, I. (ed.) (2019). *Wyzwania i uwarunkowania zarządzania inteligentnymi miastami*. Gliwice: Wydawnictwo Politechniki Śląskiej. Retrieved from: https://www.researchgate.net/publication/340298206.
6. Kaiser, Z.R.M.A. (2024). Smart governance for smart cities and nations. *Journal of Economy and Technology, 2*, pp. 216-234. Retrieved from: https://doi.org/10.1016/j.ject.2024.07.003.
7. Kinelski, G. (2022). Smart City 4.0 as a set of social synergies. *Polish Journal of Management Studies, 26(1)*, pp. 92-106. Retrieved from: https://doi.org/10.17512/pjms.2022.26.1.06.
8. Kozak, A. (2022). The Role of Smart Cities during the COVID-19 Pandemic – the Example of New York City. *Biblioteka Regionalisty, 22*, pp. 23-33. Retrieved from: https://doi.org/10.15611/br.2022.1.03.
9. Kuzior, A. (2024). Smart City Conceptual Framework in the Context of Achieving Sustainable Development Goals. *Management Systems in Production Engineering, 32(2)*, pp. 156-161. Retrieved from: https://doi.org/10.2478/mspe-2024-0015.
10. Lai, C.M.T., Cole, A. (2023). Measuring progress of smart cities: Indexing the smart city indices. *Urban Governance, 3(1)*, pp. 45-57. Retrieved from: https://doi.org/10.1016/j.ugj.2022.11.004.
11. Marchlewska-Patyk, K. (2023.). Ewolucja koncepcji smart city jako wyznacznik rozwoju współczesnych miast-polska perspektywa. *Akademia Zarządzania, 7(1)*. Retrieved from: https://doi.org/10.24427/az-2023-0009.
12. Pangsy-Kania, S., Kania, K. (2024). Key Dimensions in Smart Cities' Rankings - Towards Enhancing the Quality of Life for Smart Cities' Inhabitants. *Ekonomia i Prawo, 23(3)*, pp. 493-518. Retrieved from: https://doi.org/10.12775/EiP.2024.025.

13. *Report on the quality of life in European cities* (2023). Retrieved from: https://doi.org/10.2776/182815.
14. Svítek, M., Skobelev, P., Kozhevnikov, S. (2020). *Smart City 5.0 as an Urban Ecosystem of Smart Services* (pp. 426-438). Retrieved from: https://doi.org/10.1007/978-3-030-27477-1_33.
15. Toh, C.K. (2022). Smart city indexes, criteria, indicators and rankings: An in-depth investigation and analysis. *IET Smart Cities, 4(3*), pp. 211-228. Retrieved from: https://doi.org/10.1049/smc2.12036.
16. Vanli, T. (2024). Ranking of Global Smart Cities Using Dynamic Factor Analysis. *Social Indicators Research, 171(2)*, pp. 405-437. Retrieved from: https://doi.org/10.1007/s11205-023-03259-7.
17. Warszawski, U., Mikucki, J. (2021) Koncepcja smart city a COVID-19. Wykorzystanie nowych mediów w obliczu pandemii. *Media Biznes Kultura, Vol. 2, Iss. 11*, pp. 75 -95.