Journal of Engineering Science *Vol. XXXI, no. 4 (2024), pp. 33 - 54*
Fascicle *Industrial Engineering* *ISSN 2587-3474*
Topic *Applied Engineering Sciences and Management* *eISSN 2587-3482*

https://doi.org/10.52326/jes.utm.2024.31(4).03
UDC 656.121.072:303.72(478-25)

MEASURING PASSENGER SATISFACTION IN URBAN PUBLIC TRANSPORT IN CHISINAU

Igor Rotaru [1] *, ORCID: 0000-0003-3636-4896,
Andrei Rusu [2, 3], ORCID: 0000-0002-0259-3060

[1] *Technical University of Moldova, 168 Ștefan cel Mare Blvd., Chisinau, Republic of Moldova*
[2] *Moldova State University, VAIMCS, 5 Academiei str., Chisinau, Republic of Moldova*
[3] *Ovidius University of Constanta, 124 Mamaia Blvd., Constanta, Romania*
* Corresponding author: Igor Rotaru, *igor.rotaru@tran.utm.md* Andrei Rusu, *andrei.rusu@math.md*

Received: 10. 24. 2024
Accepted: 11. 22. 2024

Abstract. The quality of services provided by the urban public transport system plays an essential role in determining the level of satisfaction of a city's inhabitants and in improving their quality of life. The analysis of the methods used to assess the quality of urban transport services reveals the need to adopt new concepts that more realistically reflect the various aspects of their quality. A critical aspect identified is the fact that traditional methods for assessing the quality of urban passenger transport fail to fully integrate the subjective perceptions of users. They usually focus on technical, operational or economic parameters, neglecting the importance of the personal experience and satisfaction felt by passengers. The integration of subjective perceptions and the use of advanced evaluation methods contribute to the development of a public transport system that better meets passengers' expectations, stimulating the use of public transport and reducing the negative impact on the urban environment.This article represents an application of the principal component analysis (PCA) method to assess the quality of services provided by the urban public transport system using the data collected during the study.This approach will allow for a more balanced assessment focused on the real needs of passengers, thus providing more solid support for decision-making and the implementation of measures to improve urban public transport.

Keywords: *urban public transport system, quality indicators, Principal Component Analysis, case study.*

Rezumat. Calitatea serviciilor oferite de sistemul de transport public urban joacă un rol esențial în determinarea nivelului de satisfacție al locuitorilor unui oraș și în îmbunătățirea calității vieții acestora. Analiza metodelor utilizate pentru evaluarea calității serviciilor de transport urban relevă necesitatea adoptării unor concepte noi care să reflecte mai real diversele aspecte ale calității acestora. Un aspect critic identificat este faptul că metodele tradiționale de evaluare a calității transportului urban de pasageri nu reușesc să integreze complet percepțiile subiective ale utilizatorilor. Acestea se concentrează, de obicei, pe parametrii tehnici, exploataţionali sau economici, neglijând importanța experienței personale și a satisfacției resimțite de călători. Integrarea percepțiilor subiective și utilizarea unor metode avansate de evaluare contribuie la dezvoltarea unui sistem de transport piblic care răspunde mai bine așteptărilor pasagerilor, stimulând utilizarea transportului public și reducând impactul negativ asupra mediului urban. Acet articol reprezintă o aplicare a metodei analizei componentelor principale (PCA), pentru aprecierea calităţii serviciilor oferite de sistemul de transport public urban folosind datele colectate pe parcursul studiului. Această abordare va permite o evaluare mai echilibrată și orientată spre nevoile reale ale pasagerilor, oferind astfel un suport mai solid pentru luarea deciziilor și pentru implementarea măsurilor de îmbunătățire a transportului public urban.

Cuvinte cheie: *sistem de transport public urban, indicatori de calitate, Analiza Componentelor Principale , studiu de caz.*

1. Introduction

At the end of the 19th century and at the beginning of the 20th century, many scientists, sociologists, economists, and engineers focused on the laws governing population movement, which ultimately led to the organization and subsequent development of transportation of people, and particularly public transportation of people [1,2].

Simultaneously with the development of public transport, the emphasis has been placed on transport quality based on indicators that reflect efficiency, accessibility, sustainability and user satisfaction [3,4].

Often the assessment of public transport quality is carried out through direct indicators, either technical and operational indicators or economic indicators, such as those described in the normative acts of the Chisinau City Hall [5,6].

The goal of the paper is to determine the factors that influence the overall satisfaction of the users of the public urban transport in the municipality Chisinau by using the principle component analysis method. The tools used to conduct the analysis are based on programming language of Python and the appropriate specialized libraries.

The goal of the paper is to use the statistical method known as principal component analysis (PCA) that is a statistical technique which analyze data that contains observations described by some inter-correlated numerical variables. The method is usually used to extract the important information from the statistical data and then represent that data in a set of new orthogonal variables called principal components. From the mathematics perspective the PCA is determined by eigenvectors and eigenvalues. Eigenvectors and eigenvalues are numbers and vectors associated to square matrices. Together they provide the eigen-decomposition of a matrix, which analyzes the structure of this matrix such as correlation, covariance, or cross-product matrices. This method was first described by Pearson [7] in 1901 and Hotelling [8] in 1933. More modern explanations and applications of the PCA method in industry, research areas as chemistry, data analysis, physics, multifactor analysis, etc, can be found in [9-14].

2. Description of the data

We use in our paper the results of the survey conducted in November 2022 in the municipality of Chisinau regarding the satisfaction of the passengers of the urban public transport in Chisinau. The survey was conducted by a team of specialists from the Technical University of Moldova at the request of the Chisinau City Hall. The method used in survey was a face-to-face interview. More details about the the survey can be found in [6]. The analyzed data is presented in the excel file: DB_sondaj_pasageri_initial.xlsx [15]. The file contains information regarding the gender of the respondent, age group of the responder, whether the respondent lives in Chisinau or in the suburbs, the type of public transport used, type of urban transport used, frequency of public transport use, availability of information on routes and there scheduling, the conditions in the stop stations, conditions in the vehicles, and various satisfaction metrics, such as availability, readability of schedules, and general satisfaction with public transport. Actually the data was collected on papers, and the xlsx file is the result of its digitalization. The total number of collected parameters is 47 (see the list of fitures F1, ..., F47 below the Figure 5).

3. Description of the method used

In the paper we use the Principal Component Analysis method [7-14] to analyse the data. The basic reason to choose the method is that is allows to reduce the dimensionality of the problem (in our our case from 47 parameters to less ones – may be 2, 3, or 5). In the following we will briefly describe the method that involve next steps:

1. Standardization of data.
2. Covariance Matrix Computation.
3. Computation of Eigenvalues and Eigenvectors of the Covariance Matrix.
4. Selection of Principal Components.
5. Project Data onto New Feature Space.
6. Interpret results.

In the following we give brief explanations of the above steps.

3.1. Standardization of data

Since Principal Component Analysis method is sensitive to the scale of the data it is important to standardize the dataset. This ensures that all variables contribute equally to the analysis, regardless of their original scales. Thus for each feature subtract the mean and divide by the standard deviation so that each feature has a mean of 0 and a standard deviation of 1, so use the formula:

$$z = \frac{x-\mu}{\sigma}, \tag{1}$$

where: x is the original data, μ is the mean value, and σ is the standard deviation.

Observation. In order to apply the standardization all data must have numerical values (in order to be able to apply formula (1) above). If some data is not numerical, may be it can be assign numbers to non-numerical data. The method can not be applied if data can not be represented on a scale.

3.2. Covariance matrix computation

It is known that covariance matrix shows how much the variables vary with each other. It's essential for understanding the relationships between the variables in the dataset, so we have to calculate the covariance matrix of the standardized data. If there is a $n \times p$ dataset (with n samples and p variables) the covariance matrix is a $p \times p$ matrix, where each element $C(i, j)$ represents the covariance between variable i and variable j. The covariance between two variables X and Y is computed by the well-known formula:

$$Cov(X,Y) = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \mu_X)(Y_i - \mu_Y) \tag{2}$$

In (2) μ_X represents the mean value of the registered samples X_i for variable X , and μ_Y represents the mean value of the registered samples Y_i for variable Y .

3.3. Computation of eigenvalues and eigenvectors of the Covariance matrix

Eigenvalues and eigenvectors are useful in the process of identification of the principal components. More about eigenvectors and eigenvalues can be found in [16]. The eigenvectors determine the directions of the new feature space, and the eigenvalues show the amount of variance carried in each direction. The eigenvectors are the principal components, and the corresponding eigenvalues give the amount of variance each component explains. To calculate the eigenvalues and eigenvectors from the covariance matrix involve two steps: computation of the eigenvalues, then determine the corresponding eigenvalues λ are determined from the equation:

$$C \cdot v = \lambda \cdot v, \tag{3}$$

where: C is the covariance matrix, v is the eigenvector, and λ is the eigenvalue.

3.4. Selection of Principal Components

The eigenvalues tell us the amount of variance each principal component explains. We typically select the top k eigenvectors corresponding to the largest eigenvalues, which represent the principal components.

The goal is to keep components that explain most of the variance (usually the first few components).

3.5. Project Data onto New Feature Space

Finally, we project the original standardized data onto the new feature space using the selected principal components. The new dataset Z is computed as:

$$Z = X \cdot W, \tag{4}$$

where: W is the matrix of the selected eigenvectors.

3.6. Interpret the results

The transformed data will have fewer dimensions (principal components), and the variance explained by each principal component can help determine how much information is retained after the reduction. We can plot the explained variance or cumulative variance to assess how many principal components to keep.

4. Case study

In the following we apply the steps of PCA method described earlier in the article to the data.

4.1 Standartization of data

In order to apply the PCA method to the raw data we have to standardize it. According to observation in 3.1. “Standardization of Data” initially we have to get all data as numerical values.

Analizing the data we found out that it contains some columns that are not even numerical one. They are related to the gender of the respondent, to the age interval of the responder, whether the responder lives in Chisinau or lives in the surburbs, how often the responder uses the public transport, and the overall satisfaction of the responder with the public transport. We apply next transformation to the data: for each such column with non-numerical data we change the value to an appropriate numerical one as it is noted below:

- Gender of the responder:
  - Female → 0,
  - Male → 1,
- Age of the respondent:
  - 12-18 → 1
  - 18-27 → 2
  - 28-37 → 3
  - 38-47 → 4
  - 48-57 → 5
  - 58-64 → 6
  - > 64 → 7
- Reponder lives in Chisinau or in the surburb:
  - Chisinau → 0
  - Surburb → 1
- How often the respondent uses the public transport:
  - Satisfied, but there are comments → 3,
  - Absolutely unsatisfied → 1,
  - Absolutely satisfied → 4.

In oder to do all these manipulations with the data in the initial excel file we use the following python code. The result is the excel file: DB_sondaj_pasageri_numeric.xlsx. The python code is presented bellow.

```
import pandas as pd

# Load the Excel file
file_path = "DB_sondaj_pasageri_initial.xlsx"
excel_data = pd.ExcelFile(file_path)

# Display sheet names to understand the structure of the file
excel_data.sheet_names

# Load the data from 'Sheet1'
data = pd.read_excel(file_path, sheet_name="Sheet1")

# Display the first few rows to understand the structure
data.head()

# Let's review non-numerical columns in the dataset
non_numerical_columns = data.select_dtypes(include=["object"]).columns

# List the unique values in non-numerical columns to determine how to encode
them
non_numerical_columns_unique_values = {
    col: data[col].unique() for col in non_numerical_columns
```

```
}

non_numerical_columns_unique_values

# Apply transformations to non-numerical columns

# 1. Encode 'Gender of the responder'
data["Gender of the responder "] = data["Gender of the responder "].replace(
    {"Female": 0, "Male": 1}
)

# 2. Encode 'Age of the respondent'
data["Age of the respondent "] = data["Age of the respondent "].replace(
    {
        "12-18": 1,
        "18-27": 2,
        "28-37": 3,
        "38-47": 4,
        "48-57": 5,
        "58-64": 6,
        "> 64": 7,
    }
)

# 3. Encode ' Do you live in Chisinau or in the suburbs '
data["Do you live in Chisinau or in the suburbs "] = data[
    " Do you live in Chisinau or in the suburbs "
].replace({"Chisinau": 0, "Suburb": 1})

# 4. Encode 'What mode of transport do you use most often to travel'
data["What mode of transport do you use most often to travel"] = data[
    "What mode of transport do you use most often to travel"
].replace(
    {
        "Trolleybus": 1,
        "City bus": 2,
        "Minibus": 3,
        "Suburban bus": 4,
    }
)

# 5. Encode 'How often do you use public transport'
data["How often do you use public transport"] = data[
    "How often do you use public transport"
].replace(
    {
```

```
            "6-7 days a week": 6,
            "5 days a week": 5,
            "3-4 days a week": 4,
            "1-2 days a week": 3,
            "About 1-3 times a month": 2,
            "Every 2-3 months": 1,
            "Every 4-6 months": 0.5,
            "Once a year": 0.25,
        }
    )

    # 6. Encode 'General satisfaction with public transport in Chisinau'
    data["General satisfaction with public transport in Chisinau"] = data[
        "General satisfaction with public transport in Chisinau"
    ].replace(
        {
            "Satisfied, but there are comments": 3,
            "Absolutely dissatisfied": 1,
            "Absolutely satisfied": 4,
        }
    )

    # Save the modified data into a new Excel file
    output_path = "DB_sondaj_pasageri_numeric.xlsx"
    data.to_excel(output_path, index=False)

    output_path
```

The code above actually documents by itself the transformation of non-numerical data into a numerical one.

Since all the data in the table already have a numeric value, we can apply the standardization procedure to it in order to get for every feature (column) in the table the mean value to be equal to 0, and the standard deviation to be equal to 1. To do this we use the following snippet of python program:

```
    # Standardize the data
    scaler = StandardScaler()
    standardized_data = scaler.fit_transform(data)

    # Convert the standardized data back to a DataFrame for easier manipulation and
export
    standardized_df = pd.DataFrame(standardized_data, columns=data.columns)

    # Save the standardized data to a new Excel file
```

```
standardized_output_path = 'DB_sondaj_pasageri_standardized_final.xlsx'
standardized_df.to_excel(standardized_output_path, index=False)

standardized_output_path
```

4.2 Covariance Matrix Computation

Next step on the road to the goal of finding out the principal components of the data is to calculate the covariance matrix. This is achieved by the help of the following program:

```
import pandas as pd

# Load the dataset
data = pd.read_excel('DB_sondaj_pasageri_standardized_final.xlsx')

# Calculate the covariance matrix
covariance_matrix = data.cov()

# Save the covariance matrix to a CSV file
covariance_matrix.to_csv('DB_sondaj-Covariance-matrix-of-standardized-data.csv')
```

4.3 Computation of Eigenvalues and Eigenvectors of the Covariance Matrix

Let us calculate now the eigenvalues and the corresponding eigenvectors of the covariance matrix obtained at the previous step. Again we use a specialized python library and the corresponding functions in order to get it.

```
# Importing necessary libraries for eigenvalue calculation
import numpy as np

# Calculate the eigenvalues and eigenvectors of the covariance matrix
eigenvalues, eigenvectors = np.linalg.eig(covariance_matrix)

# Creating a DataFrame to show both eigenvalues and their corresponding
eigenvectors
eigen_details = pd.DataFrame(eigenvectors, columns=[f'Eigenvector_{i+1}' for i in
range(eigenvectors.shape[1])])
eigen_details.insert(0, 'Eigenvalue', eigenvalues)

# Display the details of the calculation to the user
tools.display_dataframe_to_user(name="Eigenvalues and Eigenvectors",
dataframe=eigen_details)
```

According to the results the explained variance can be calculated by:

```
# Calculating the explained variance for each eigenvalue
total_variance = np.sum(eigenvalues)
explained_variance_ratio = [(i / total_variance) * 100 for i in eigenvalues]

# Creating a DataFrame to display the explained variance
explained_variance_df = pd.DataFrame({
    'Eigenvalue': eigenvalues,
    'Explained Variance (%)': explained_variance_ratio
})

# Displaying the explained variance to the user
tools.display_dataframe_to_user(name="Explained Variance of Each Principal Component", dataframe=explained_variance_df)
```

The result is in the table below which contains the first 10 eigenvalues and the corresponding percent of the explained variance. The eigenvalues are usually ordered in descending way:

Table 1

The first 10 eigenvalues of the covariance matrix ant there importance in the explanation of the variance in data

Order	Value of eigenvalue	Explained variance, %	Cumulative explained variance, %
1	10.27404491	21.85	21.85
2	3.589577401	7.63	29.48
3	2.973504444	6.32	35.80
4	2.286543124	4.86	40.66
5	2.150963809	4.57	45.24
6	1.84612014	3.93	49.16
7	1.81918762	3.87	53.03
8	1.583324077	3.37	56.40
9	1.337555732	2.84	59.24
10	1.151848884	2.45	61.69

The explained variance gives insight into how much information (or variance) each principal component (associated with each eigenvalue) captures from the dataset. Here's how to interpret the results:

- **Eigenvalue** - represents the amount of variance captured by the corresponding principal component (a direction in feature space). A higher eigenvalue indicates that the corresponding principal component captures a larger portion of the variance in the data.
- **Explained Variance (%)** - tells us the proportion of the total variance in the dataset that is explained by each principal component. For example, the first principal

component in your data captures 21.85% of the total variance, and the second component captures 7.63%.

According to the data we get the following interpretations of the data:

- **Reduction of the dimension**. The principal components with the highest explained variance are the most significant. If you're looking to reduce dimensionality, you might retain only the components that explain most of the variance. For example, the top 5 components in your case explain over 45% of the variance, so you could potentially reduce the dimensionality to 5 components while still preserving a significant amount of information.
- **Cumulative Variance**. You can sum the explained variance percentages to see how much of the total variance is captured by a given number of components. This helps to decide how many components to keep in a dimensionality reduction task. If you're satisfied with, say, 70% of the variance explained, you can retain as many components as it takes to reach that threshold. In the present case the first 14 features are responsible for the at least 70% of the variance, and first 6 features explains a bit more than 49% of the variance.
- **Insights into Data Structure**. The first few components represent the directions in your data that have the most variability. If the first few components explain most of the variance, it means your data has some underlying structure that can be described by fewer variables.

Let us visualize the cumulative variance and the explained variance by principal components:

```
# Re-importing the required pandas library
import pandas as pd

# Redefining the file path and reloading the data
file_path = 'DB_sondaj_pasageri_standardized_final.xlsx'

# Reload the dataset and compute the covariance matrix
data = pd.read_excel(file_path)
covariance_matrix = data.cov()

# Recalculate the eigenvalues and explained variance
eigenvalues, eigenvectors = np.linalg.eig(covariance_matrix)
total_variance = np.sum(eigenvalues)
explained_variance_ratio = [(i / total_variance) * 100 for i in eigenvalues]

# Plotting the explained variance
plt.figure(figsize=(10, 6))
plt.bar(range(1, len(explained_variance_ratio) + 1), explained_variance_ratio,
alpha=0.7, align='center', color='blue', label='Individual Explained Variance')
```

```
        plt.step(range(1, len(explained_variance_ratio) + 1),
np.cumsum(explained_variance_ratio), where='mid', color='red', label='Cumulative Explained
Variance')

        plt.xlabel('Principal Component')
        plt.ylabel('Explained Variance (%)')
        plt.title('Explained Variance by Principal Components')
        plt.legend(loc='best')
        plt.grid(True)
        plt.show()
```

The generated corresponding Figure 1 is:

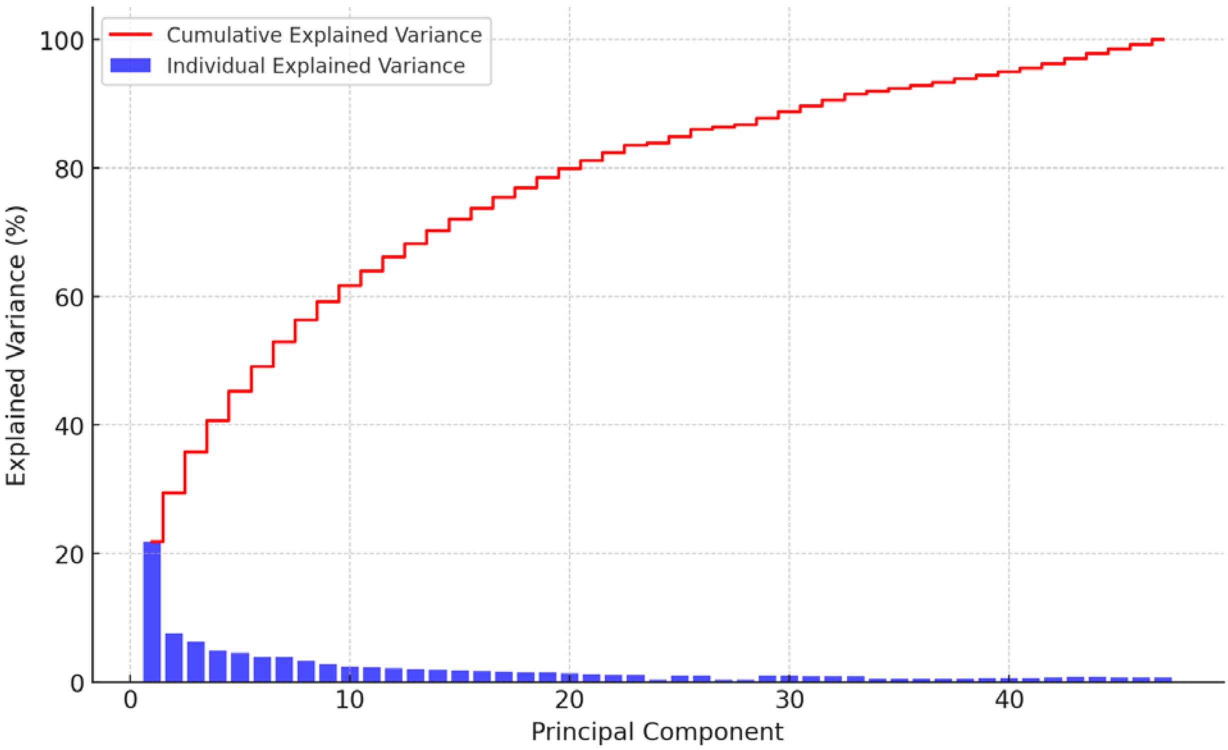

Figure 1. Explained Variance by Principal Components.

Now it is time to select the principal components and analyze them.

4.4 Selection of Principal Components

It is clear from the obtained data that the most important component is the first one which explains almost 21% of the variance, but it is not sufficient to explain the major change in variance. The first 5 principal components determines almost 50% of the variability in the variance. In the following we present the visualizations of the principal components. The next code uses a standard python module for data analysis.

4.5. Project data onto new feature space and interpretation of results

```
        from sklearn.decomposition import PCA
```

```
        from sklearn.preprocessing import StandardScaler

        # Step 1: Standardize the data (as PCA requires standardized data)
        # Imputing missing values with the mean of each column
        data_filled = data.fillna(data.mean())

        # Standardizing the data after filling NaN values
        standardized_data_filled = scaler.fit_transform(data_filled)

        # Step 2: Perform PCA
        pca = PCA()
        pca_components = pca.fit_transform(standardized_data_filled)

        # Step 3: Capture the explained variance ratio
        explained_variance_ratio = pca.explained_variance_ratio_

        # Creating a DataFrame to show the principal components and explained variance
        pca_results_df = pd.DataFrame(pca_components, columns=[f'PC{i+1}' for i in
range(pca_components.shape[1])])

        # Adding explained variance information
        explained_variance_df = pd.DataFrame({
            'Principal Component': [f'PC{i+1}' for i in range(len(explained_variance_ratio))],
            'Explained Variance (%)': explained_variance_ratio * 100
        })

        # Displaying the PCA results to the user
        tools.display_dataframe_to_user(name="Principal Components",
dataframe=pca_results_df)
        tools.display_dataframe_to_user(name="Explained Variance per Principal
Component", dataframe=explained_variance_df)

        # Visualizing the first two principal components
        plt.figure(figsize=(10, 6))
        plt.scatter(pca_components[:, 0], pca_components[:, 1], alpha=0.7, edgecolors='k')
        plt.title('PCA: First Two Principal Components')
        plt.xlabel('Principal Component 1')
        plt.ylabel('Principal Component 2')
        plt.grid(True)
        plt.show()
```

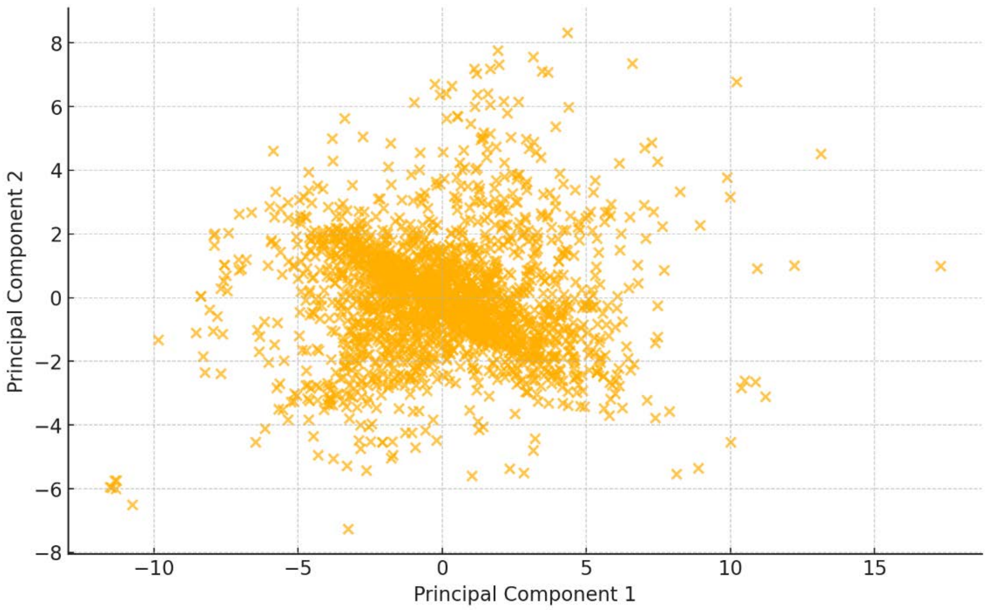

Figure 2. PCA – first two principal components.

The Figure 3 below shows the first 3 principal components.

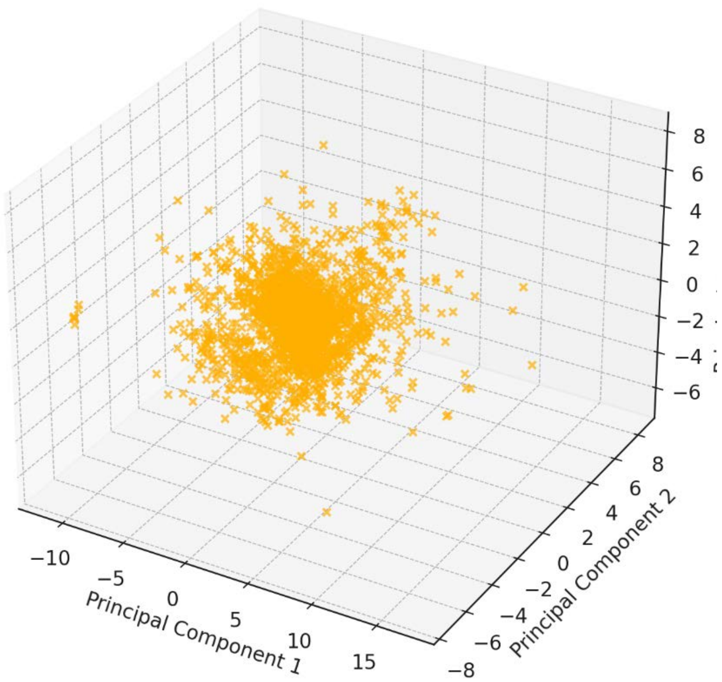

Figure 3. PCA – first three principal components

The Figure 3 above is obtained using the code:

```
from mpl_toolkits.mplot3d import Axes3D

# Visualizing the first three principal components in 3D
fig = plt.figure(figsize=(10, 8))
ax = fig.add_subplot(111, projection='3d')
```

```
    # Plotting the first three PCs
    ax.scatter(pca_components[:, 0], pca_components[:, 1], pca_components[:, 2],
alpha=0.7, edgecolors='k')

    # Setting labels and title
    ax.set_title('PCA: First Three Principal Components')
    ax.set_xlabel('Principal Component 1')
    ax.set_ylabel('Principal Component 2')
    ax.set_zlabel('Principal Component 3')

    plt.show()
```

In order to visualize the first 5 principal components we use slices, or different views, which are based on showing different important pairs of principal components. To get them use the code:

```
    # Visualizing the first five principal components in 2D scatter plots

    fig, axs = plt.subplots(2, 3, figsize=(15, 10))

    # Plotting the first five principal components pairwise
    pairs = [(0, 1), (0, 2), (0, 3), (1, 2), (1, 3), (2, 3)]
    for i, (pc1, pc2) in enumerate(pairs[:5]):
        row, col = divmod(i, 3)
        axs[row, col].scatter(pca_components[:, pc1], pca_components[:, pc2], alpha=0.7,
edgecolors='k')
        axs[row, col].set_xlabel(f'PC{pc1 + 1}')
        axs[row, col].set_ylabel(f'PC{pc2 + 1}')
        axs[row, col].grid(True)

    plt.suptitle('Scatter Plots of Top 5 Principal Components', fontsize=16)
    plt.tight_layout(rect=[0, 0, 1, 0.96])
    plt.show()
```

The result is in the Figure 4 below:

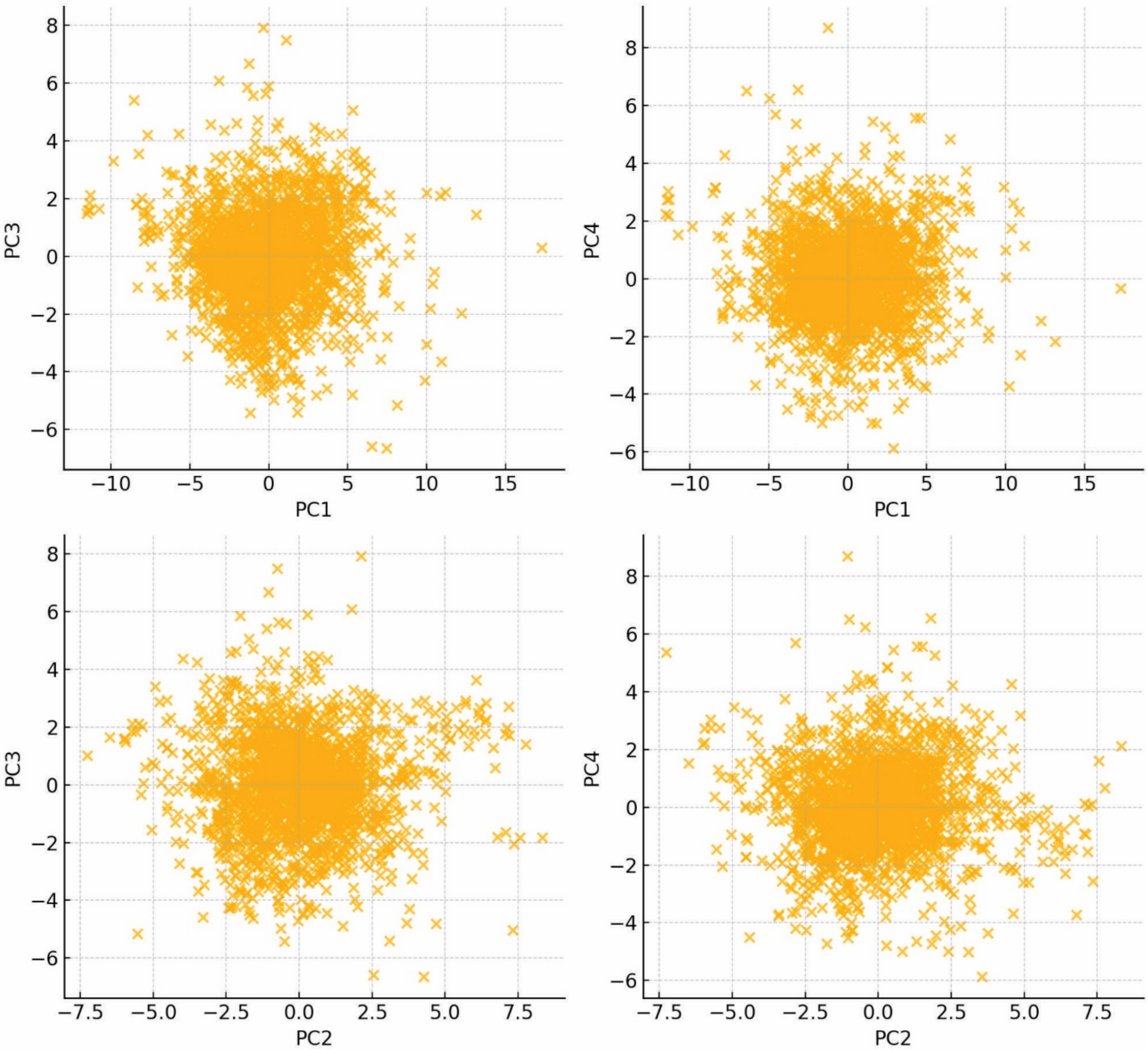

Figure 4. PCA – next two by two principal components

To determine the most important factors (or features) in the first three principal components, we need to analyze the **loading scores** or **eigenvectors** associated with each principal component. These loading scores show the contribution of each original feature to a principal component.

Let's extract the top contributing features for each of the first three principal components. I'll calculate and display the features with the highest loading scores for the first three principal components: PC1, PC2, and PC3.

```
# Extracting the eigenvectors (loadings) of the first 3 principal components
loading_scores = pd.DataFrame(pca.components_.T, columns=[f'PC{i+1}' for i in range(pca.n_components_)], index=data.columns)

# Getting the most important features (highest absolute loadings) for the first three principal components
top_features_pc1 = loading_scores['PC1'].abs().sort_values(ascending=False).head()
top_features_pc2 = loading_scores['PC2'].abs().sort_values(ascending=False).head()
```

```
top_features_pc3 = loading_scores['PC3'].abs().sort_values(ascending=False).head()

# Combining results into a single DataFrame for display
important_factors_df = pd.DataFrame({
    'PC1 Top Features': top_features_pc1.index,
    'PC1 Loading': top_features_pc1.values,
    'PC2 Top Features': top_features_pc2.index,
    'PC2 Loading': top_features_pc2.values,
    'PC3 Top Features': top_features_pc3.index,
    'PC3 Loading': top_features_pc3.values
})

# Display the important factors
tools.display_dataframe_to_user(name="Important Features in PC1, PC2, and PC3",
dataframe=important_factors_df)
```

The result is in the Table 2 below.

Table 2

Important features in first three principal components

No.	PC1 Top Features	PC1 Loading	PC2 Top Features	PC2 Loading	PC3 Top Features	PC3 Loading
1	Satisfaction with public transport stations in general	0.2248	Satisfaction with the minibus route network - general satisfaction	0.3383	Satisfaction with the trolleybus route network - general satisfaction	0.2590
2	General satisfaction with public transport in Chisinau	0.2232	Satisfaction with the minibus route network - time you spend approaching (leaving) a stopping point	0.3253	Available mobile apps	0.2590
3	Experience of being on public transport while traveling on the route you most often use - It is easy for you to get in and out of the public transport (applying physical force)	0.1939	Satisfaction with the minibus route network - the ability to get anywhere and at any time to any part of the city	0.3161	Given each of these aspects, how satisfied are you with providing information about the public transport timetable as a whole	0.2523

Continuation Table 2

4	Given each of these aspects, how satisfied are you with providing information about the public transport timetable as a whole	0.1904	Satisfaction with the minibus route network - number of available routes	0.3082	Information shown on electronic scoreboards at some stops about the time of arrival of public transport	0.2395
5	Satisfaction with public transport stations - possibility to sit	0.1902	Satisfaction with the minibus route network - distance to the nearest public transport stop	0.2778	Satisfaction with the trolleybus route network - distance to the nearest public transport stop	0.2375

We have identified the most important features for the first three principal components (PC1, PC2, and PC3). For each component, the features with the highest loading scores contribute the most to that principal component. Here are the top contributing factors for each:

PC1: Primarily influenced by general satisfaction factors related to transport stations and the ease of entry and exit during travel.

PC2: Heavily related to the satisfaction with microbus routes, including the availability of routes, travel time, and distance to stations.

PC3: Largely driven by satisfaction with trolleybus routes and available mobile applications, along with the availability of public information.

Let us also get a heatmap for the loading of the first 5 principal components. The color intensity indicates the contribution (positive or negative) of each feature to the respective principal component. This visualization helps to see which features strongly influence each component.

```
import seaborn as sns

# Plotting a heatmap of the loadings for the first 5 principal components
plt.figure(figsize=(12, 8))
sns.heatmap(loading_scores.iloc[:, :5], annot=True, cmap='coolwarm', center=0)
plt.title('Heatmap of Loadings for the First 5 Principal Components')
plt.xlabel('Principal Components')
plt.ylabel('Features')
plt.show()
```

The heating map is the presented in Picture 5.

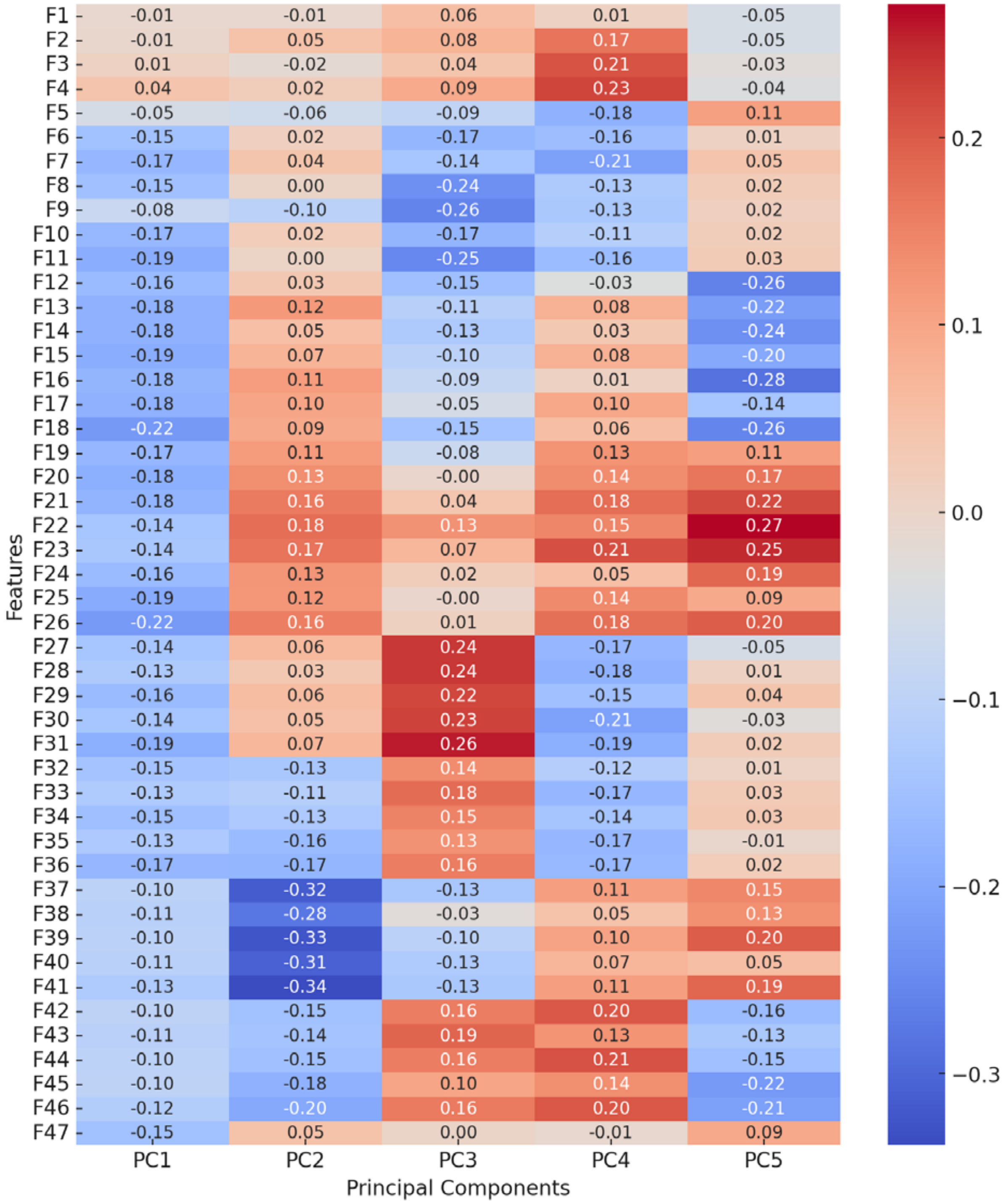

Figure 5. Heatmap of loadings for the first 5 principal components.

The features F1, …, F47 are given in the next list:

- F1 = Gender of the responder
- F2 = Age of the respondent
- F3 = Do you live in Chisinau or in the suburbs.
- F4 = What mode of transport do you use most often to travel
- F5 = How often do you use public transport
- F6 = Availability of the correct public transport timetable at the stops
- F7 = Easy to read and understand a public transport timetable at stops

- F8 = Information shown on electronic scoreboards at some stops about the time of arrival of public transport
- F9 = Available mobile apps
- F10 = Actual compliance with established public transport timetable
- F11 = Given each of these aspects, how satisfied are you with providing information about the public transport timetable as a whole
- F12 = Satisfaction with public transport stations - lighting
- F13 = Satisfaction with public transport stations - cleanliness and the amount of garbage
- F14 = Satisfaction with public transport stations - the opportunity to hide from rain and wind
- F15 = Satisfaction with public transport stations - possibility to sit
- F16 = Satisfaction with public transport stations - pavement condition
- F17 = Satisfaction with public transport stations - congestion of passengers at a time when you usually drive
- F18 = Satisfaction with public transport stations in general
- F19 = Experience of being on public transport while traveling on the route you most often use - cabin occupancy
- F20 = Experience of being on public transport while traveling on the route you most often use - softness of public transport travel
- F21 = Experience of being on public transport while traveling on the route you most often use - comfort inside the cabin
- F22 = Experience of being on public transport while traveling on the route you most often use - temperature inside the cabin in winter
- F23 = Experience of being on public transport while traveling on the route you most often use - temperature inside the cabin in summer
- F24 = Experience of being on public transport while traveling on the route you most often use - quality and level of illumination of the cabin
- F25 = Experience of being on public transport while traveling on the route you most often use - It is easy for you to get in and out of the public transport (applying physical force)
- F26 = Satisfaction with the experience of being in public transport cabin in general
- F27 = Satisfaction with the trolleybus route network - the ability to get anywhere and at any time to any part of the city
- F28 = Satisfaction with the trolleybus route network - distance to the nearest public transport stop
- F29 = Satisfaction with the trolleybus route network - time you spend approaching (leaving) a stopping point
- F30 = Satisfaction with the trolleybus route network - number of available routes
- F31 = Satisfaction with the trolleybus route network - general satisfaction
- F32 = Satisfaction with the city bus route network - the ability to get anywhere and at any time to any part of the city
- F33 = Satisfaction with the city bus route network - distance to the nearest public transport stop

- F34 = Satisfaction with the city bus route network - time you spend approaching (leaving) a stopping point
- F35 = Satisfaction with the city bus route network - number of available routes
- F36 = Satisfaction with the city bus route network - general satisfaction
- F37 = Satisfaction with the minibus route network - the ability to get anywhere and at any time to any part of the city
- F38 = Satisfaction with the minibus route network - distance to the nearest public transport stop
- F39 = Satisfaction with the minibus route network - time you spend approaching (leaving) a stopping point
- F40 = Satisfaction with the minibus route network - number of available routes
- F41 = Satisfaction with the minibus route network - general satisfaction
- F42 = Satisfaction with the suburban bus route network - the ability to get anywhere and at any time to any part of the city
- F43 = Satisfaction with the suburban bus route network - distance to the nearest public transport stop
- F44 = Satisfaction with the suburban bus route network - time you spend approaching (leaving) a stopping point
- F45 = Satisfaction with the suburban bus route network - number of available routes
- F46 = Satisfaction with the suburban bus route network - general satisfaction
- F47 = General satisfaction with public transport in Chisinau

It follows from the heatmap that (see PC2) some data related to microbuses have great negative impact, the accessibility and the diversity of trolleybus routes have a positive impact (PC3). The PC1 shows that all factors have a light negative impact on general satisfaction, while other principal components have higher diversity in negative and positive perception.

5. Conclusions

The results shows that there are no unique set of factors which has a high influence on the general satisfaction of the users of the public transport in in the municipality of Chisinau. At the same time the PCA method helps to identify among the 47 factors under consideration the mixtures that have a significant impact on general satisfaction. These mixtures are collected in first 5 identified principal components. This can help de decision factors in the administration of the municipality to consider the most adequate policies in order to increase the general satisfaction of the users.

The conducted case study shows that the first principal component is responsible for about 20% of the users identified satisfaction. At the same time the first principal component has a weight as the sum of the next three principal components. Since the first 5-6 princiaple components explains about 50% in the variance of the level of satisfaction of the users we can conclude that almost all examined factors are relatively important, and the general satisfaction of the users depends on all of them, so this implies an action polity regarding almost all of them.

At the same time the analysis of the first principal component (Table 2) tell us that the factors that are relatively important in this component are:

- satisfaction with public transport stations in general;
- satisfaction with the experience of being in public transport cabin in general;
- experience of being on public transport while traveling on the route you most often use - it is easy for you to get in and out of the public transport (applying physical force);
- given each of these aspects, how satisfied are you with providing information about the public transport timetable as a whole;
- satisfaction with public transport stations - possibility to sit.

On the other hand, the same Table 2 and Figure 5 the second principal component have brought to light the general insatisfaction with any aspect related to minibuses:

- satisfaction with the minibus route network - general satisfaction;
- satisfaction with the minibus route network - time you spend approaching (leaving) a stopping point;
- satisfaction with the minibus route network - the ability to get anywhere and at any time to any part of the city;
- satisfaction with the minibus route network - number of available routes;
- satisfaction with the suburban bus route network - distance to the nearest public transport stop.

The third principal component shows us that in general, travelers are satisfied with the trolleybus network, the distance to the trolleybus stations, the level of information through mobile applications and other information sources.

Conflicts of interest: The authors declare no conflict of interests.

Acknowledgments: The work carried out by Rusu Andrei was partially supported within the Institutional Project 011301 "Computer Science Systems based on Artificial Intelligence" (strategic priority "V. Innovative technologies, sustainable energy, digitalization"; strategic direction "Innovative technologies and products; Information and communication technology and digital development"; carried out at State University of Moldova, Vladimir Andrunachievici Institute of Mathematics and Computer Science).

References

1. Ballon, H.; Jackson, K. T. *Robert Moses and the modern city: the transformation of New York*. W.W. Norton & Co, NY, SUA, 2007, 336 p.
2. Wardrop, J.G. *Some Theoretical Aspects of Road Traffic Research*. Institution of Civil Engineers, Northwestern University, 1952, 54 p.
3. Vuchic, V. *Transportation for Livable Cities*. Taylor and Francis, Milton, UK, 2017, 376 p.
4. Spirin, I. V., *Organization and management of passenger transportation*. Academy, Moscow, 2020, 398 p. [in Russian].
5. "Regulations on road transport of passengers and luggage on the territory of Chisinau municipality." approved by the decision of the Chisinau Municipal Council no. 4/22 of May 5, 2023. [in Romanian].
6. Rotaru, I.; Ceban,V. Quality of public transport services in Chisinau. *Journal of Engineering Science* 2024, 31(2), pp. 7−16, doi: 10.52326/jes.utm.2024.31(2).01.
7. Pearson, K. LIII. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin philosophical magazine and journal of science* 1901, 2(11), pp. 559−572.
8. Hotelling, H. Analysis of a complex of statistical variables into principal components. *Journal of educational psychology* 1933, 24(6), pp. 417-441, doi: https://doi.org/10.1037/h0071325.

9. Jolliffe, I. T. *Principal component analysis for special types of data*. pp.338–72. Springer Series in Statistics. New York: Springer-Verlag, 2002. https://doi.org/10.1007/0-387-22440-8_13.
10. Greenacre, M.P.; Groenen, J. F.; Hastie, T.; D'Enza, A. I.; Markos, A.; Tuzhilina, E. Principal component analysis. *Nature Reviews Methods Primers* 2022, 2(1), 100, doi: 10.1038/s43586-022-00184-w.
11. Wold, S.; Esbensen, K.; Geladi, P. Principal component analysis. *Chemometrics and intelligent laboratory systems*, 1987, 2(1–3), pp. 37–52.
12. Bro, R.; Smilde, A. K. Principal component analysis. *Analytical methods* 2014, 6(9), pp. 2812–2831.
13. Abdi, H.; Williams, L. J. Principal component analysis. *Wiley interdisciplinary reviews: computational statistics* 2010, 2(4), pp. 433–459.
14. Jolliffe, I. T.; Cadima, J. Principal component analysis: a review and recent developments. *Philosophical transactions of the royal society A: Mathematical, Physical and Engineering Sciences* 2016, 374, 2065.
15. Rotaru, I. *Survey on the satisfaction of public transport users in Chisinau municipality*. Nov. 28, 2022. [Excel]. Available online: https://docs.google.com/spreadsheets/d/1m0An_EEqww6dfLvN9Ts0GnVn58zweFYY/edit?usp=drive_link&ouid=110108346911593091702&rtpof=true&sd=true (accessed on 12.10.2024).
16. Strang, G. *Introduction to Linear Algebra.* Wellesley-Cambridge Press, Massachusetts, SUA, 2023, 440 p.

Citation: Rotaru, I.; Rusu, A. Measuring passenger satisfaction in urban public transport in Chisinau. *Journal of Engineering Science* 2024, XXXI (4), pp. 33-54. https://doi.org/10.52326/jes.utm.2024.31(3).03.

Submission of manuscripts: jes@meridian.utm.md