**Report**

Introduction The aim of this project is to analyze country data using three statistical methods: linear ordering, cluster analysis, and multidimensional scaling. The project is conducted in the R Studio environment, and the data analysis is based on a dataset containing information on fertility rates, gross domestic product per capita, female life expectancy, the percentage of the urban population, and infant mortality rates.

As part of the project, the following analyses will be conducted:

1. Preliminary statistical analysis of the data, including basic descriptive statistics, coefficient of variation analysis, correlation between variables, and identification of outliers.
2. Linear ordering, in which two methods will be applied:
   * Template method: Hellwig’s method
   * Non-template method: standardized sums The results will be presented in the form of rankings and compared.
3. Cluster analysis using the k-medoids method to divide countries into groups. The optimal number of clusters will be determined using the Elbow and Profile methods, and the formed groups will be characterized based on charts and statistics.
4. Multidimensional scaling, which will be used to visualize the results of cluster analysis and assess the quality of the applied approach.

The goal of the analysis is not only to conduct the necessary calculations but, most importantly, to interpret the results and draw conclusions regarding the studied countries and the effectiveness of the applied methods. The report will focus mainly on data analysis and interpretation, following the project guidelines.

Description and Preliminary Analysis of the Data Used The data include 39 European countries and contain five variables describing demographic and economic aspects:

* fertility (fertility rate)
* ppgdp (GDP per capita)
* lifeExpF (female life expectancy)
* pctUrban (percentage of the urban population)
* infantMortality (infant mortality rate)

Obraz zawierający tekst, zrzut ekranu, numer, Czcionka

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.

**Basic Statistics**

Obraz zawierający tekst, zrzut ekranu, Czcionka, numer

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.

**Skewness**



**Kurtosis**



**Frequency histogram**

**Obraz zawierający diagram, zrzut ekranu, Plan, linia

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.**

**Box plot**

**Obraz zawierający diagram, tekst, Plan, zrzut ekranu

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.**

**Scatter plot matrix**

**Obraz zawierający tekst, diagram, linia, Równolegle

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.**

The average fertility rate is 1.59, with a median of 1.506. The lowest value is 1.134 (Bosnia and Herzegovina), while the highest is 2.098 (Iceland). It is worth noting that for most countries, the value oscillates around the median.

GDP per capita has a very wide range—from $1,625.8 (Moldova) to $105,095.4 (Luxembourg), indicating significant differences in economic development across Europe. The average value is $27,394; however, the high skewness (1.28) and kurtosis (4.59) indicate that the distribution is highly asymmetric and concentrated around lower values, with some extreme outliers.

Female life expectancy averages 80.69 years and exhibits low variability (CV = 3.87%). The shortest life expectancy is 73.48 years (Moldova), while the longest is 84.90 years (France).

The percentage of the urban population ranges from 48% (Moldova) to 97% (Belgium). The distribution of this variable is relatively symmetrical (skewness = 0.19), suggesting no significant deviations.

Infant mortality shows the highest variability (CV = 61.97%). The lowest values are observed in Iceland (2.057 per 1,000 births), while the highest is in Albania (16.561).

**Mode**

****

**Range (min-max)**

**Obraz zawierający tekst, Czcionka, zrzut ekranu, linia

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.**

**Interquartile range (IQR)**



The most frequently occurring value for fertility rate is 1.501, while for GDP per capita, it is $3,677.2. The range of values for individual variables indicates significant diversity among countries.

IQR analysis shows that the highest variability occurs in GDP per capita ($32,899), while the lowest is in female life expectancy (4.74 years).

**Coefficient of Variation (CV)**

****

The analysis of variable variability shows that the most stable indicator is lifeExpF (CV = 3.87%), whereas the highest variability is observed in ppgdp (CV = 86.95%) and infantMortality (CV = 61.97%). This indicates significant disparities in economic development and healthcare quality.

**Correlation Matrix and Correlation Heatmap**

**Obraz zawierający tekst, zrzut ekranu, Czcionka, linia

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.**

**Obraz zawierający zrzut ekranu, Wielobarwność, kwadrat, Prostokąt

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.**

The correlation matrix indicates several significant relationships between variables:

* GDP per capita and female life expectancy: A strong positive correlation (0.71) suggests that in countries with higher GDP, women live longer.
* GDP per capita and infant mortality: A strong negative correlation (-0.67) indicates that in wealthier countries, infant mortality is lower.
* Female life expectancy and infant mortality: The strongest negative correlation (-0.77), meaning that in countries where women live longer, infant mortality is significantly lower.
* Fertility and GDP per capita: A positive correlation (0.55), but not very strong—suggesting that in countries with higher GDP, women tend to have more children.
* Fertility and urbanization: A moderate correlation (0.57) suggests that in more urbanized countries, the fertility rate is slightly higher.

**Outliers indicates**

Obraz zawierający tekst, Czcionka, zrzut ekranu

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.

The analysis of outliers indicates that:

* GDP per capita has one extremely high value—$105,095.4 for Luxembourg, which may distort the analysis.
* Infant mortality shows several outliers: 16.561 (Albania), 14.344 (Moldova), 13.063 (North Macedonia), suggesting that infant healthcare in these countries is significantly worse than in others.
* Fertility rate, female life expectancy, and percentage of the urban population do not contain outliers.

**Linear Ordering**

Division into Stimulants and Destimulants

* Stimulants (higher value = better situation):
  + PPGDP
  + LifeExpF
  + PctUrban
* Destimulants (higher value = worse situation, lower value = better):
  + InfantMortality
* Variable with an ambiguous nature - Nominal:
  + Fertility – its impact depends on context and interpretation. In developed countries, lower fertility may indicate social stability, but in the long term, it may lead to demographic problems.

**Hellwig's Method**

**Obraz zawierający tekst, zrzut ekranu, dokument, Czcionka

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.**

**Obraz zawierający zrzut ekranu, tekst, Wykres, linia

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.**

Hellwig’s method involves determining a synthetic measure that quantifies the distance of a given country from an ideal benchmark. In this case, countries with the highest Hellwig measure values are considered the most developed based on the analyzed indicators.

**Top-ranked countries:**

1. Luxembourg (0.763)
2. Switzerland (0.677)
3. Norway (0.578)
4. Germany (0.541)
5. Netherlands (0.536)

These countries exhibit a high GDP per capita, a long life expectancy for women, and low infant mortality rates. A high level of urbanization also contributes to their strong ranking.

**Lowest-ranked countries:**

* Moldova (-0.031)
* Bosnia and Herzegovina (0.045)
* Albania (0.072)
* Macedonia (0.119)
* Ukraine (0.130)

These countries have low GDP per capita, shorter life expectancy, and higher infant mortality rates. Additionally, their lower level of urbanization negatively impacts their quality of life.

**Standardized Sum Method**

Obraz zawierający tekst, Czcionka, zrzut ekranu, numer

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.

Obraz zawierający tekst, zrzut ekranu, Wykres, diagram

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.

This method sums the standardized values of variables to create a ranking of countries. Similar to Hellwig’s method, a higher value indicates a better position.

**Top-ranked countries:**

1. Luxembourg (5.825)
2. Switzerland (4.831)
3. Spain (3.758)
4. Norway (3.336)
5. Belgium (3.001)

Luxembourg and Switzerland dominate again, but Spain ranks significantly higher in this method than in Hellwig’s method. This results from the different impacts of individual variables on the final score.

**Lowest-ranked countries:**

* Moldova (-6.633)
* Bosnia and Herzegovina (-6.285)
* TFYR Macedonia (-4.552)
* Albania (-4.239)
* Romania (-3.794)

The order at the bottom of the ranking is similar to Hellwig’s method, confirming that countries with the lowest values tend to have weaker socio-economic indicators.

**Comparsion of Both Methods**

Both methods yield similar results, particularly in terms of the highest- and lowest-ranked countries. Luxembourg and Switzerland dominate both rankings, while Moldova, Bosnia, Albania, and Macedonia consistently rank at the bottom.

**Key differences:**

* Spain ranks 3rd in the Standardized Sum Method but only 8th in Hellwig’s Method due to the different weighting of variables.
* Some countries shift positions, e.g., France ranks higher in Hellwig’s Method than in the Standardized Sum Method.

**Conclusion**

* Both methods identify the same groups of countries as the best and worst, indicating their stability in assessing quality of life.
* Hellwig’s Method favors countries with stable and high economic development, while the Standardized Sum Method highlights countries with specific strong indicators.
* The choice of method can impact detailed analyses, but the overall trend remains the same.Obraz zawierający zrzut ekranu, linia, tekst, diagram

  Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.

**Clustering Analysis**

**Obraz zawierający linia, zrzut ekranu

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.**

**Obraz zawierający linia, tekst, zrzut ekranu, diagram

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.**

Based on the elbow method and profile analysis, **two** optimal groups were identified, differing in socio-economic indicators.

Obraz zawierający tekst, Czcionka, numer, linia

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.

Obraz zawierający tekst, diagram, Wykres, mapa

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.

**Group 1 (n = 21 countries)**

**Characteristics:**

* Lower socio-economic development (lower GDP per capita, shorter life expectancy, higher infant mortality)
* Less urbanized societies

**Average values for Group 1:**

* Fertility rate: 0.94 (slightly higher than Group 2)
* GDP per capita: 10,380.62 USD (significantly lower than Group 2)
* Female life expectancy: 78.43 years
* Urbanization rate: 62.14%
* Infant mortality rate: -8.34 (worse than Group 2)

**Interpretation:**  
Countries in this group exhibit lower economic and social development levels. They are primarily from Central and Eastern Europe and the Balkans, characterized by higher infant mortality, lower GDP per capita, and shorter female life expectancy. The urbanization level is significantly lower than in the second group.

**Group 2 (n = 18 countries)**

**Characteristics:**

* Higher socio-economic development
* Higher urbanization level

**Average values for Group 2:**

* Fertility rate: 0.79 (slightly lower than Group 1)
* GDP per capita: 47,242.81 USD (almost 5 times higher than Group 1!)
* Female life expectancy: 83.33 years (significantly higher than Group 1)
* Urbanization rate: 80.17% (considerably higher)
* Infant mortality rate: -3.53 (better health conditions)

**Interpretation:**  
This group consists of highly developed countries in Western and Northern Europe. They feature a higher standard of living, greater urbanization, and significantly better health indicators (lower infant mortality and higher female life expectancy).

**Comparison of Groups**

* **Group 1** consists of less developed countries with lower incomes, shorter life expectancy, and higher infant mortality.
* **Group 2** includes highly developed countries with higher GDP, longer life expectancy, and better health indicators.

The most distinguishing factor is **GDP per capita**, which in Group 2 is nearly **5 times higher** than in Group 1! This is a key indicator separating developed from developing countries.

**Multidimensional Scaling (MDS) Analysis  
Obraz zawierający tekst, zrzut ekranu, diagram, linia

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.** **Obraz zawierający diagram, Plan, tekst, mapa

Zawartość wygenerowana przez sztuczną inteligencję może być niepoprawna.**

**Interpretation of Plots**

he **Multidimensional Scaling (MDS)** method was used to visualize the division of countries into two cluster groups. Each point represents a country, and its position on the plot reflects its similarity to other countries (**closer together = more similar, farther apart = less similar**).

**Observations**

1. **Group 1 (Lower development level – black points)**
   * **Close positioning:**
     + **Lithuania and Latvia** → very similar in terms of analyzed variables.
     + **Hungary and Slovakia** → also have close values.
     + **Czech Republic and Greece** → their similarity may result from a moderate level of GDP and urbanization.
   * **Outliers within the group:**
     + **Estonia, Slovenia, Moldova, Bosnia and Herzegovina** → these countries are far from the others in this group, suggesting their characteristics differ from the average values of this cluster.
2. **Group 2 (Highly developed countries – red points)**
   * **Countries positioned closely together** → suggests a high level of similarity, which is expected for developed countries with stable economies and strong health indicators.
   * **Exception:** **Luxembourg** → stands out on the plot, indicating its unique position, likely due to its exceptionally high **GDP per capita**.

**Assessment of MDS Scaling Quality**

1. **Effectively represents cluster structure**
   * Countries within the same group are mostly positioned near each other, indicating that MDS accurately captures their similarities.
2. **Distinct clusters and outliers**
   * Clear groupings and deviations confirm that the method effectively reveals structures and differences between countries.

**Summary**

* **MDS effectively represents cluster structures** – countries within the same group are mostly close together.
* **Interesting relationships** are observed (e.g., **Lithuania-Latvia, Hungary-Slovakia**).
* **Luxembourg and a few countries in Group 1 stand out** as atypical cases.
* **The method is useful for visualizing cluster analysis results**, though it does not fully capture the multidimensional structure of the data.