



World Countries Clustering

An Unsupervised
Machine Learning Project

About country dataset

Source : [kaggle.com](https://www.kaggle.com)

Child Mortality

Death of children under 5 years of age per 1000 live births

Inflation

The measurement of the annual growth rate of the Total GDP

Health

Total health spending per capita.
Given as %age of GDP per capita

Imports

Imports of goods and services per capita. Given as %age of the GDP per capita

Life Expectancy

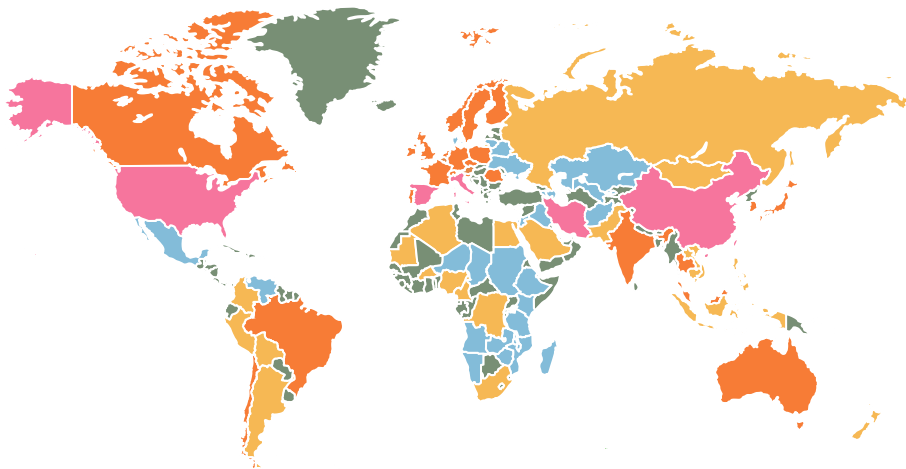
The average number of years a newborn child would live if the current mortality patterns are to remain the same

Exports

Exports of goods and services per capita. Given as %age of the GDP per capita

Total Fertility

The number of children that would be born to each woman if the current age-fertility rates remain the same.



GDPP

The GDP per capita.
Calculated as the Total GDP divided by the total population.

Income

Net income per person

Number of
States: 167

Methodology

1

EDA

Exploring and understanding the dataset with graphs

2

Clustering

Kmeans Method for clustering

3

Characterization

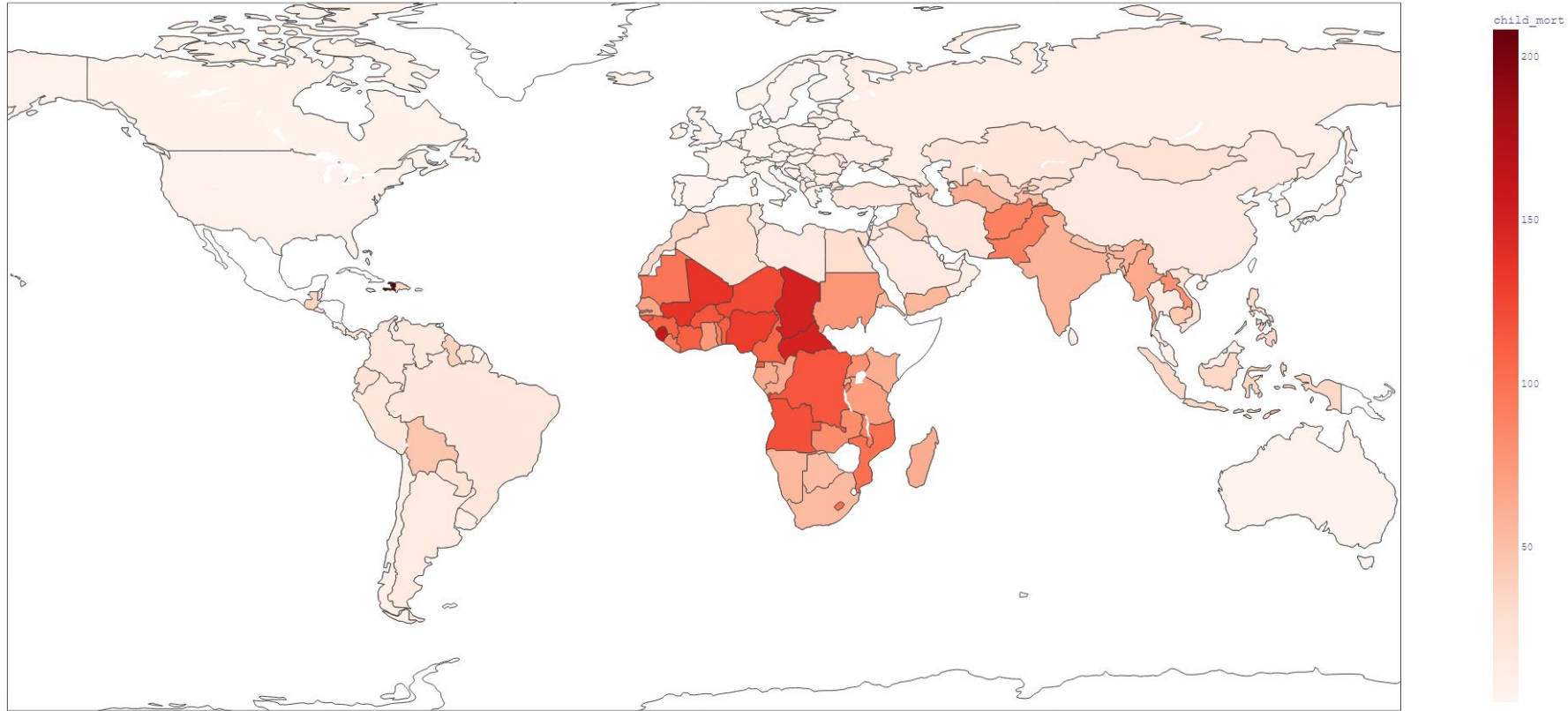
Finding out what characterize each cluster

4

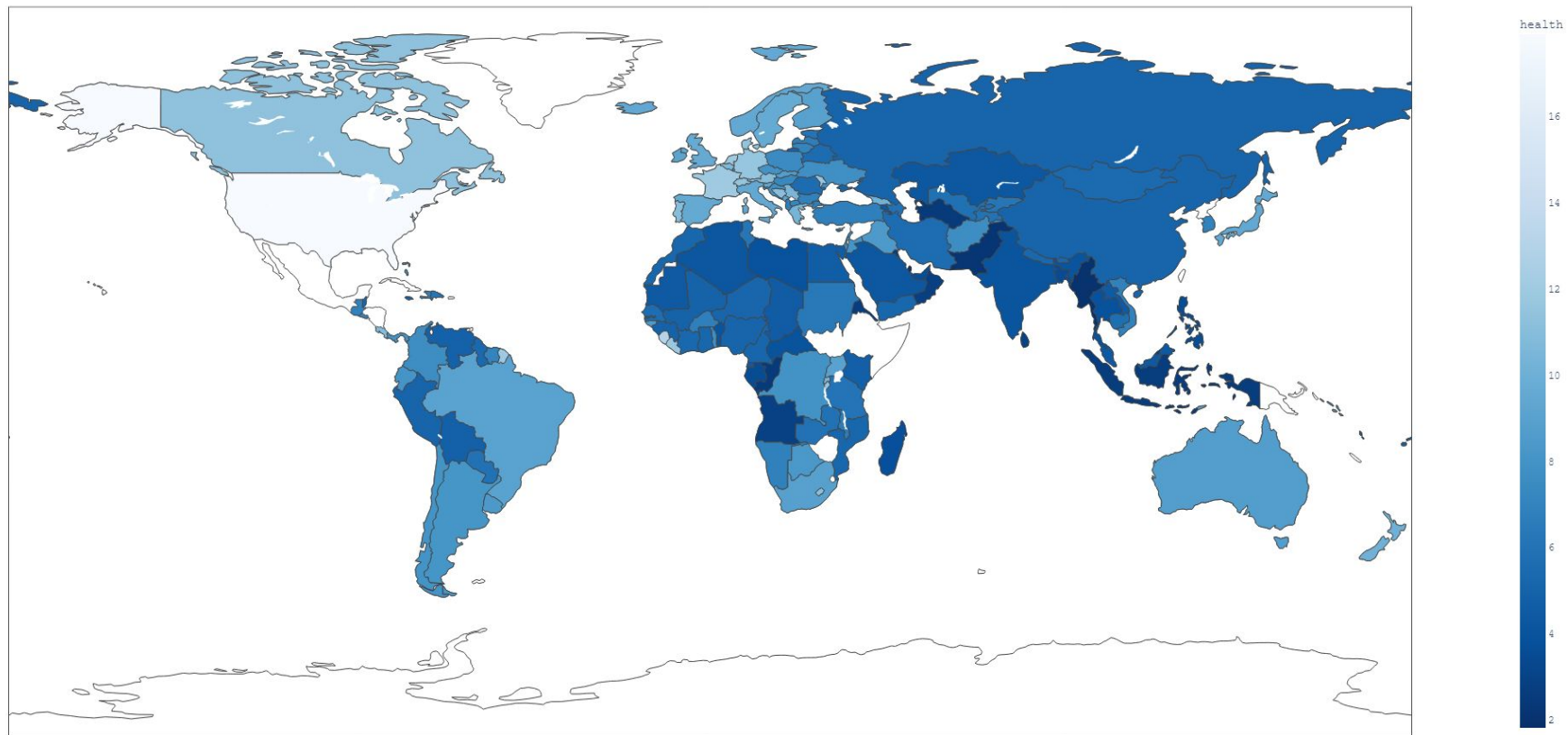
Conclusion

Questions and discussion

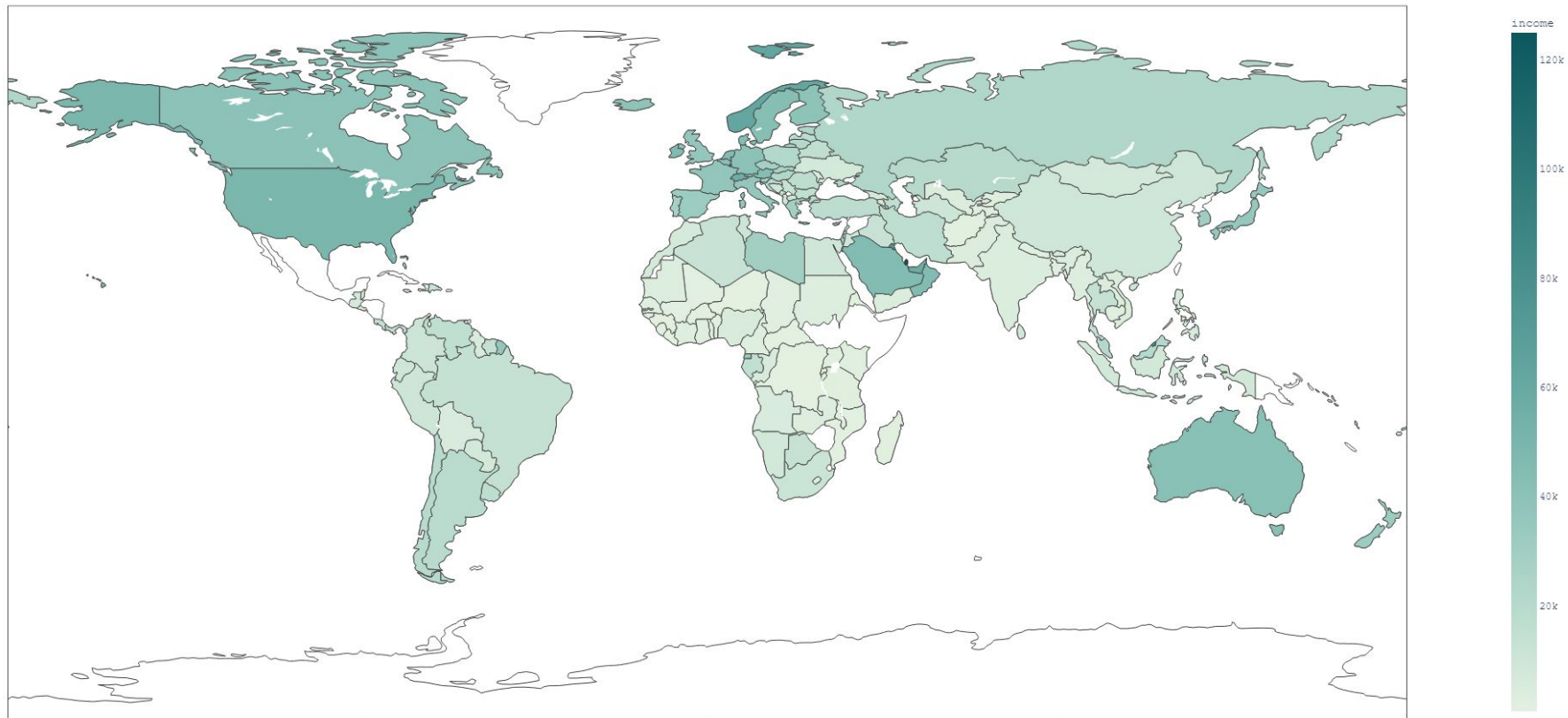
Death of children under 5 years of age per 1000 live births



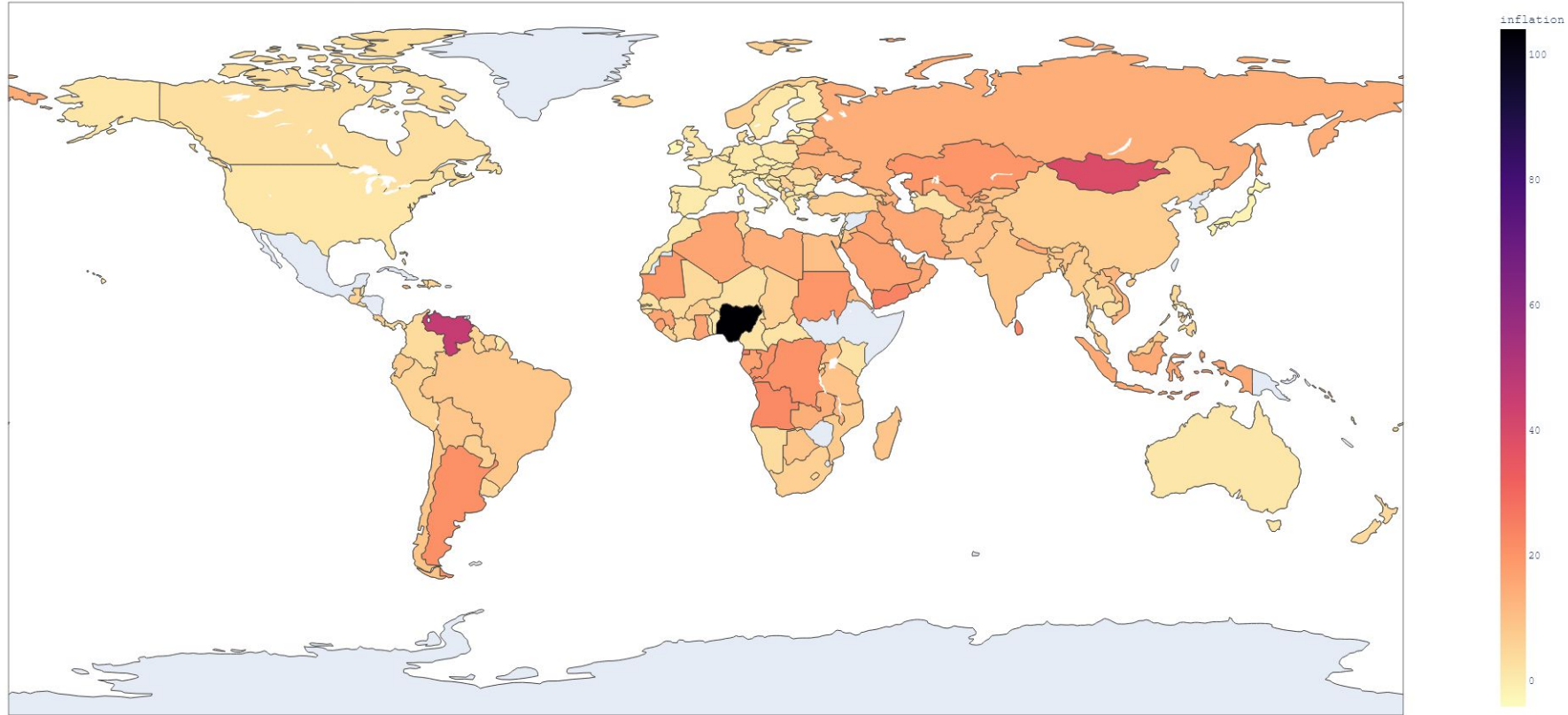
Total health spending per capita



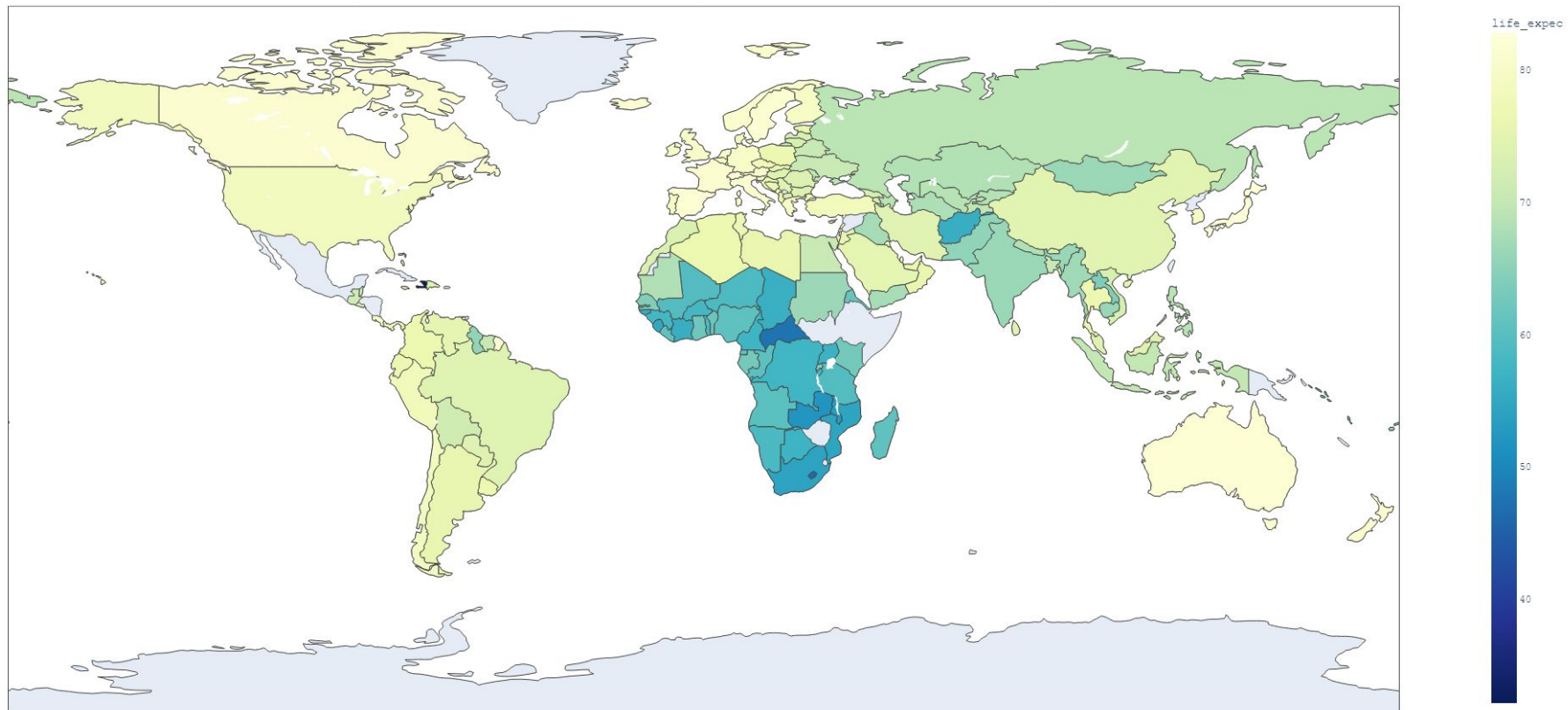
Net income per person



Inflation – The measurement of the annual growth rate of the Total GDP



Life expectancy



Unsupervised Learning

1. KMeans Method

Kmeans method for clustering.
Silhouette Score for 3 clusters : 0.700

2. Standard Scaler

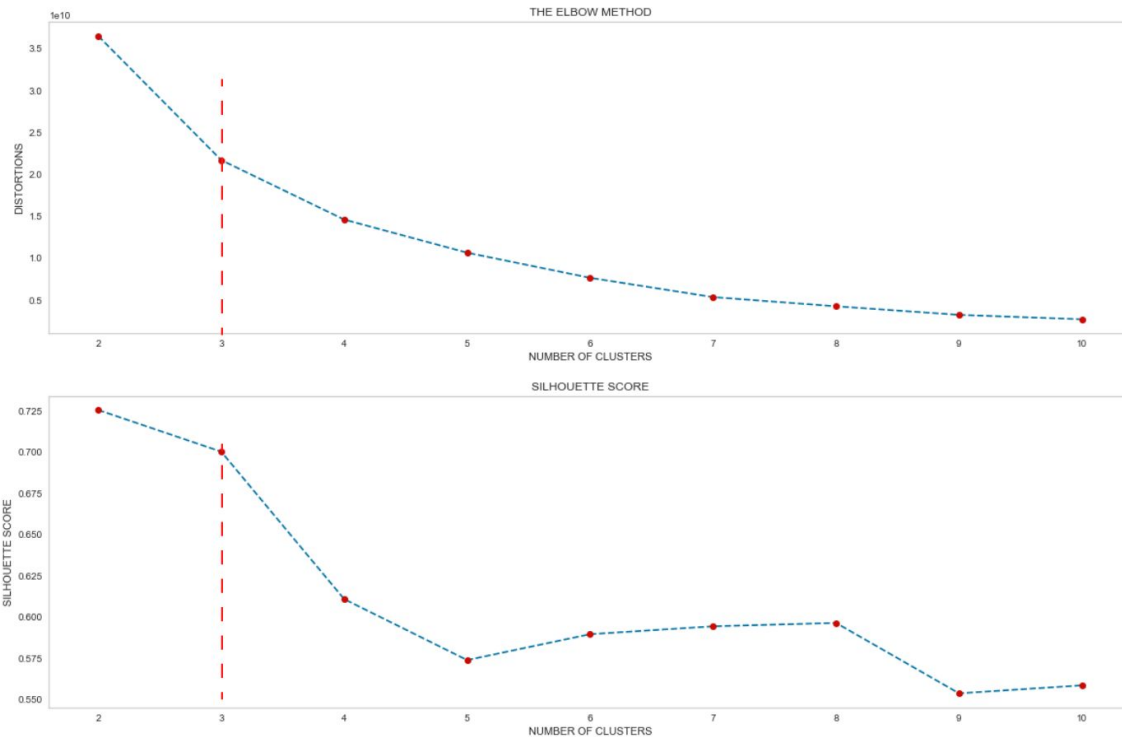
In order to optimize the clustering.

3. Elbow Method

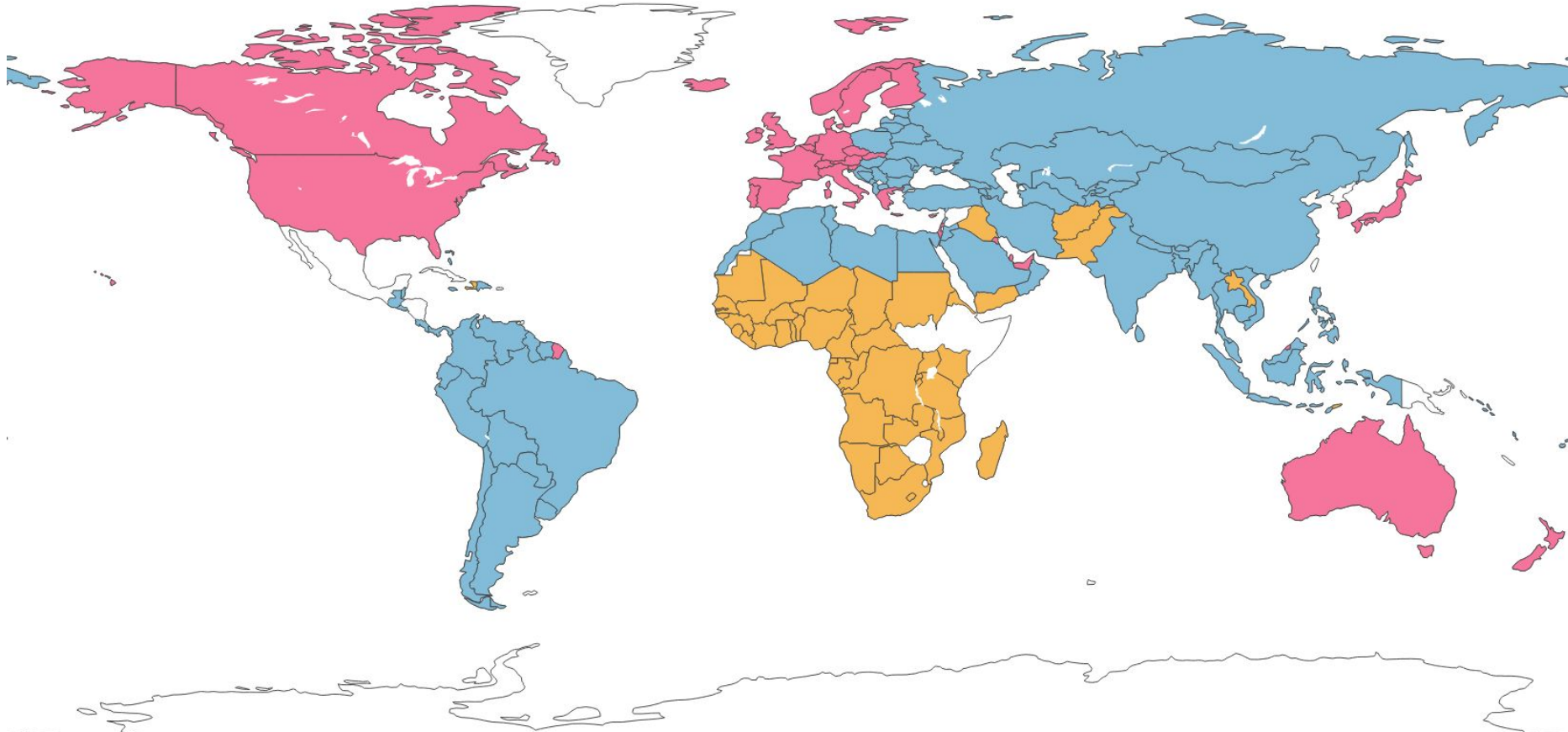
The number of clusters is selected
based on the elbow method and
silhouette score



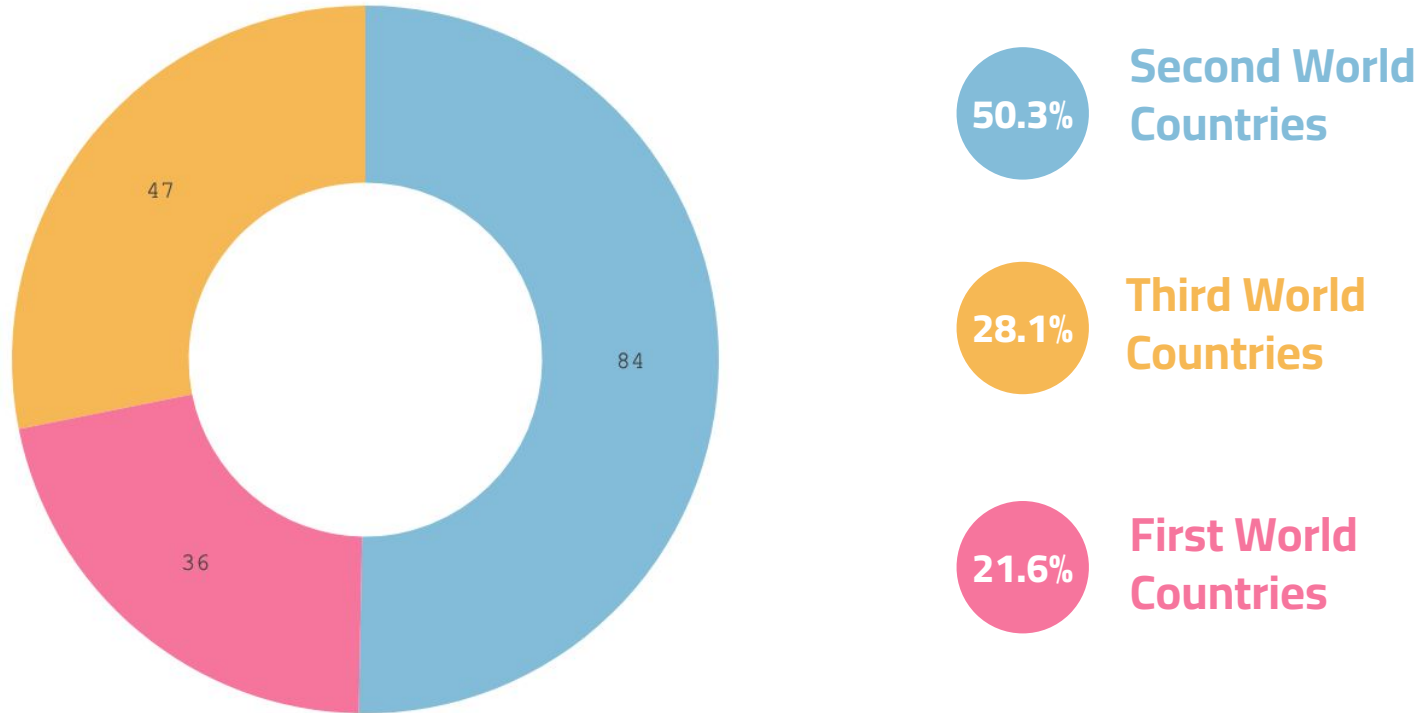
Elbow and Silhouette Score



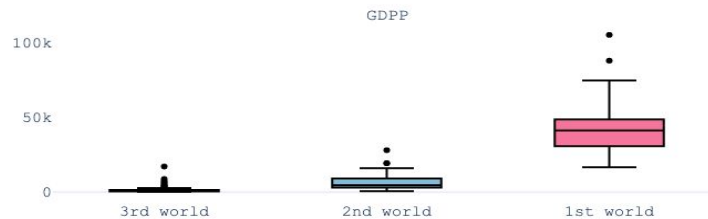
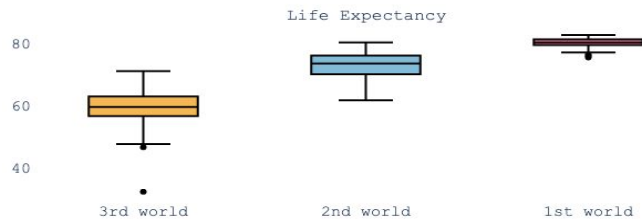
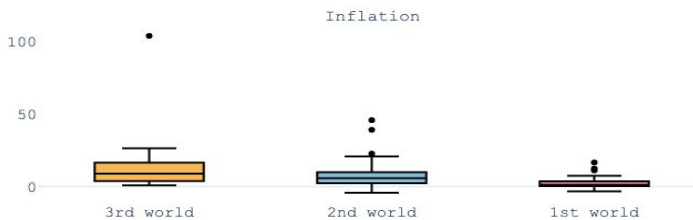
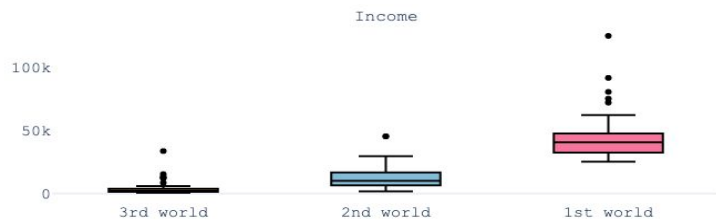
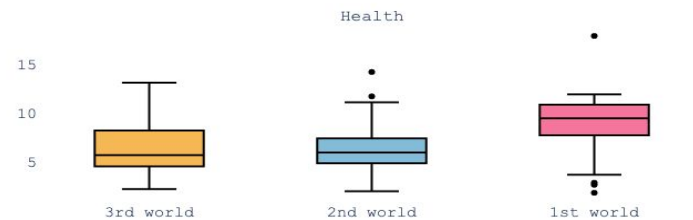
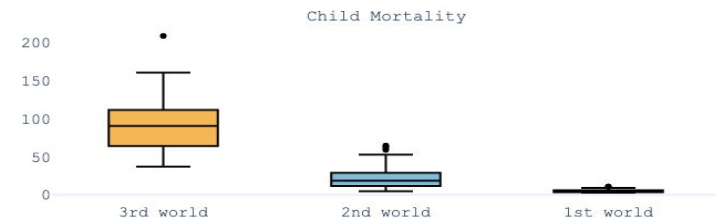
Distribution of countries in groups



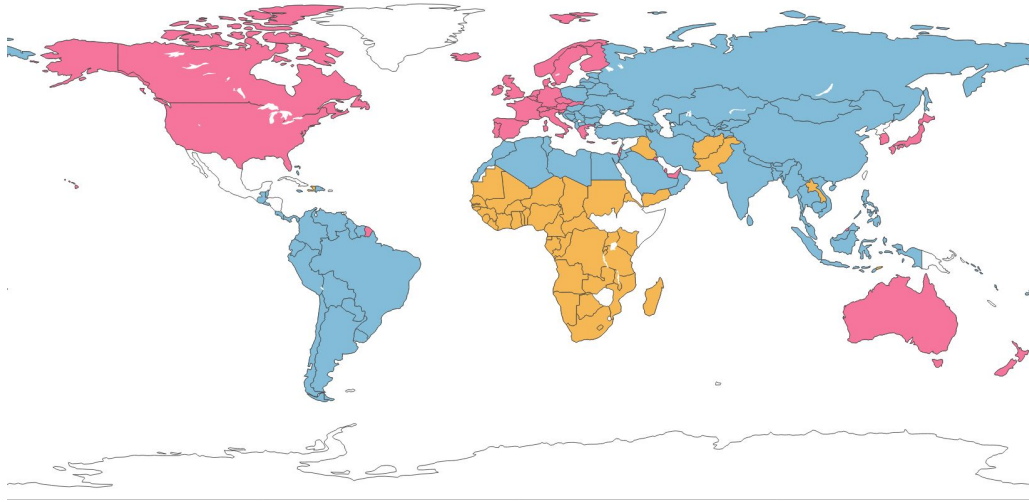
Count of Countries in each cluster



Characteristics of the Clusters



Conclusions



Satisfactory clustering performance of KMeans method



Distinct division into three groups of countries



Fewer failures in clustering (Example: South Africa)



Providing more data may have contributed to a better performance

Any Questions?





Thank you!

Amiran Matsaridis
Konstantina Savvidou