Capstone Project #3

Clustering of Countries

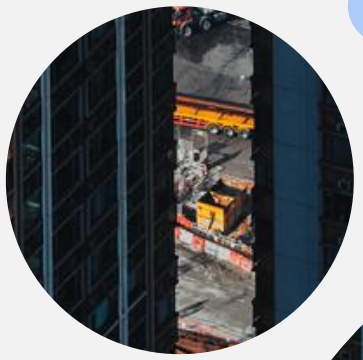
Using Unsupervised Learning Models

October 16, 2020

Leon Felipe Galindo Barragan

Agenda

- Introduction
- Preliminary view of the data
- Analyzing model results
- Conclusions



Introduction

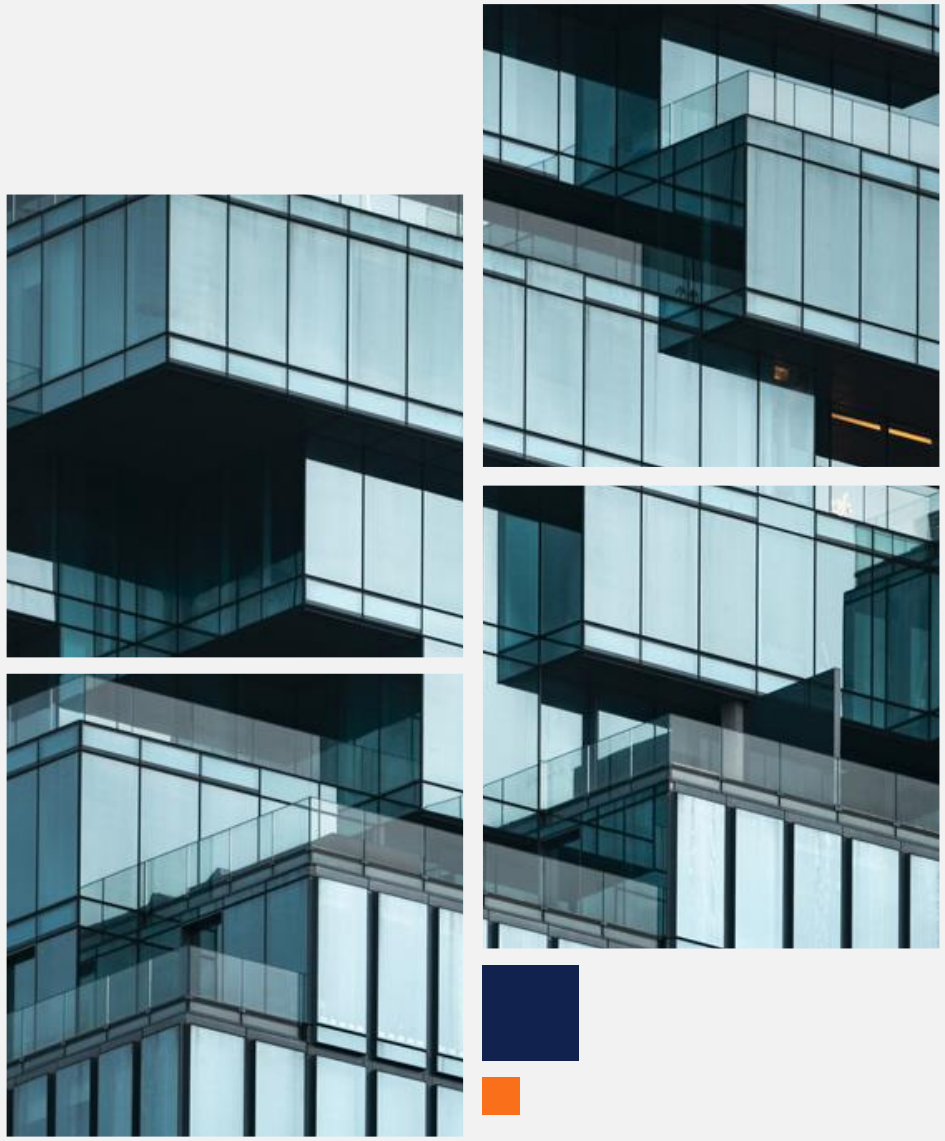
Objective:

Categorize the countries using socio-economic and health factors that determine the overall development of the country.

About organization:

HELP International is an international humanitarian NGO that is committed to fighting poverty and providing the people of backward countries with basic amenities and relief during the time of disasters and natural calamities.

https://help-international.org/

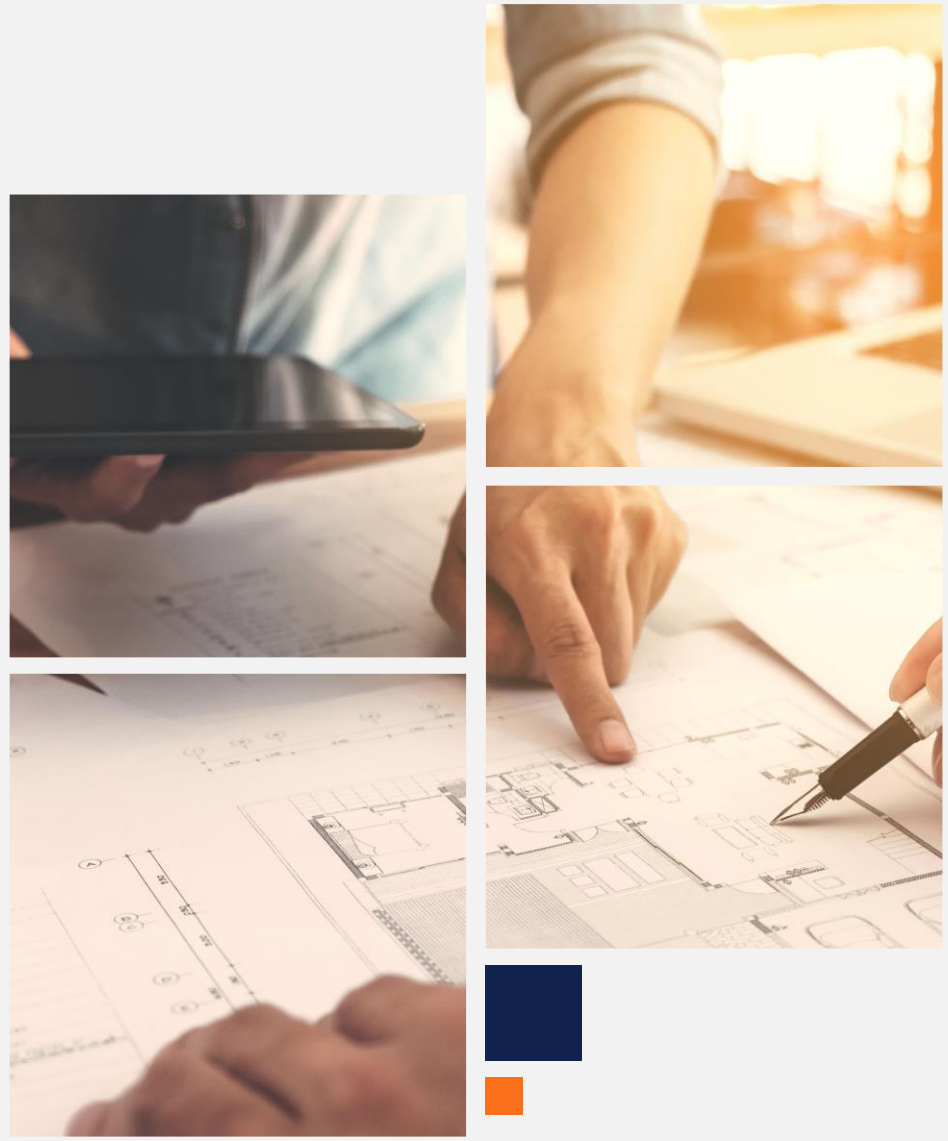


Problem Statement

HELP International have been able to raise around $ 10 million so they need to decide how to use this money strategically and effectively.

Our job is to categorize the countries using some socio-economic and health factors that determine the overall development of the country.

Then we need suggest the countries which HELP International needs to focus on the most.



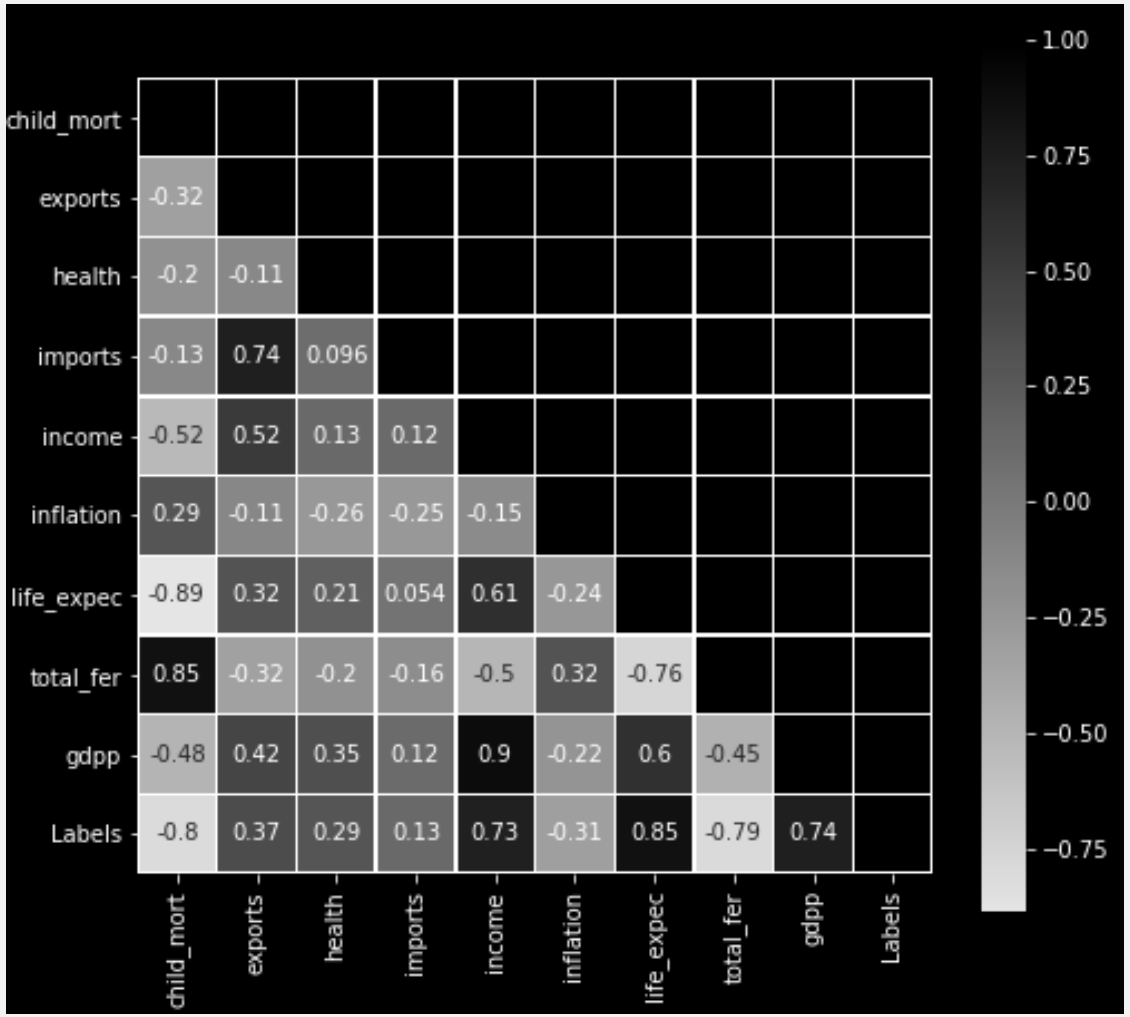
Preliminary view of the data

Let’s dive in

About the features

Column Name	Description
Country	Name of the country
Child Mortality	Death of children under 5 years of age per 1000 live births
Exports	Exports of goods and services per capita. Given as %age of the GDP per capita
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
Income	Net income per person
Inflation	A general increase in prices and fall in the purchasing value of money
Life Expectancy	The average number of years a newborn child would live if the current mortality patterns are to remain the same
Total Fertility	The number of children that would be born to each woman if the current age-fertility rates remain the same
GDP	The GDP per capita. Calculated as the Total GDP divided by the total population

Exploring correlation between features



Based on the heatmap plot we can observe the following:

- Imports have a strong correlation with exports
- Income and GDP have the strongest correlation
- Child mortality and life expectancy also show very strong correlation

Exploring the model results

Let’s dive in

Exploring the zero cluster

Cluster 0:

Underdeveloped economies, these are countries in need of help.

There are high levels of inflation, high levels of child mortality and low levels of life expectancy and economic growth.

Exploring the first cluster

Cluster 1:

These are developing economies.

Mostly middle-income countries where there is higher life expectancy, lower inflation and more economic growth compared to the cluster 0.

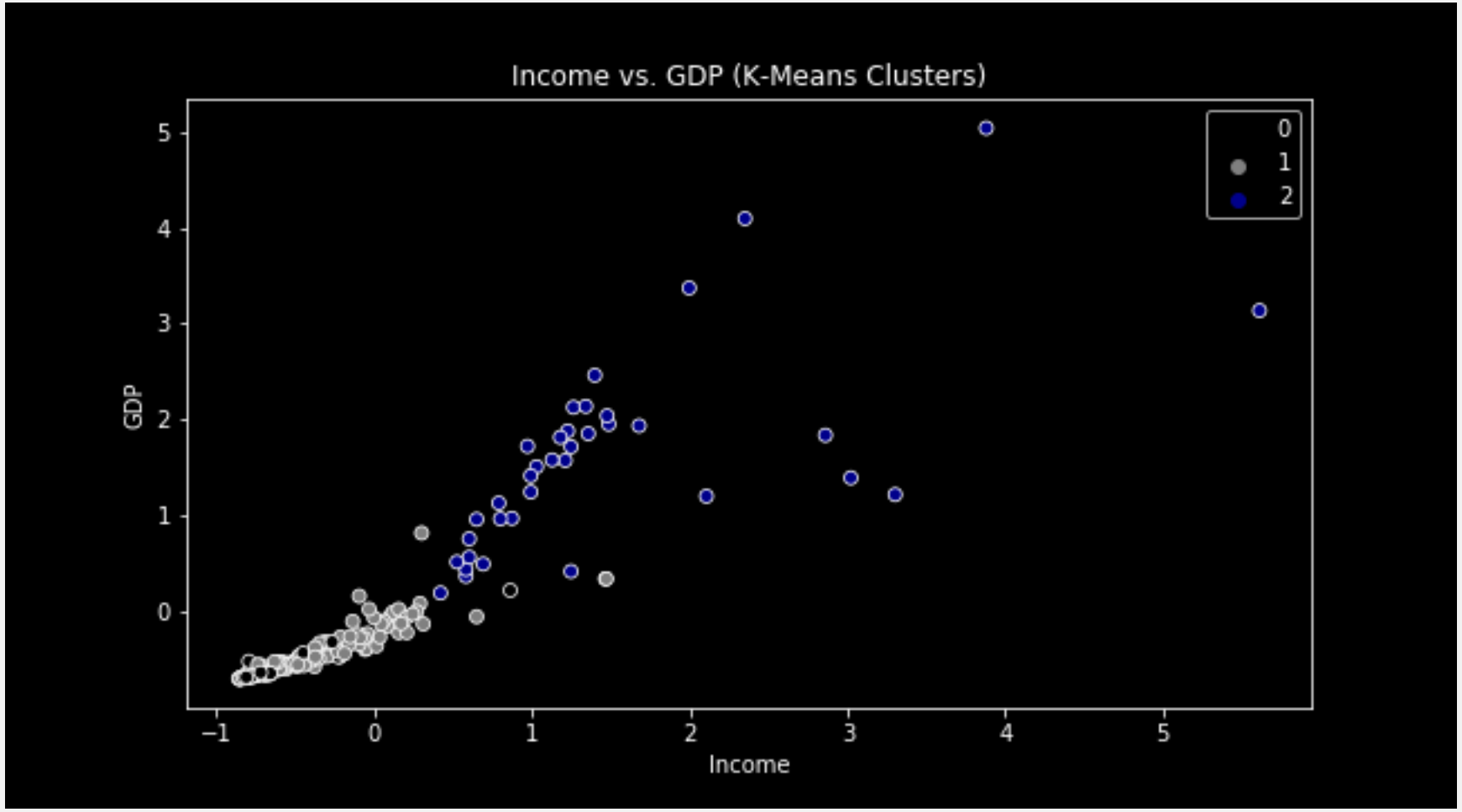
Exploring the third cluster

Cluster 2:

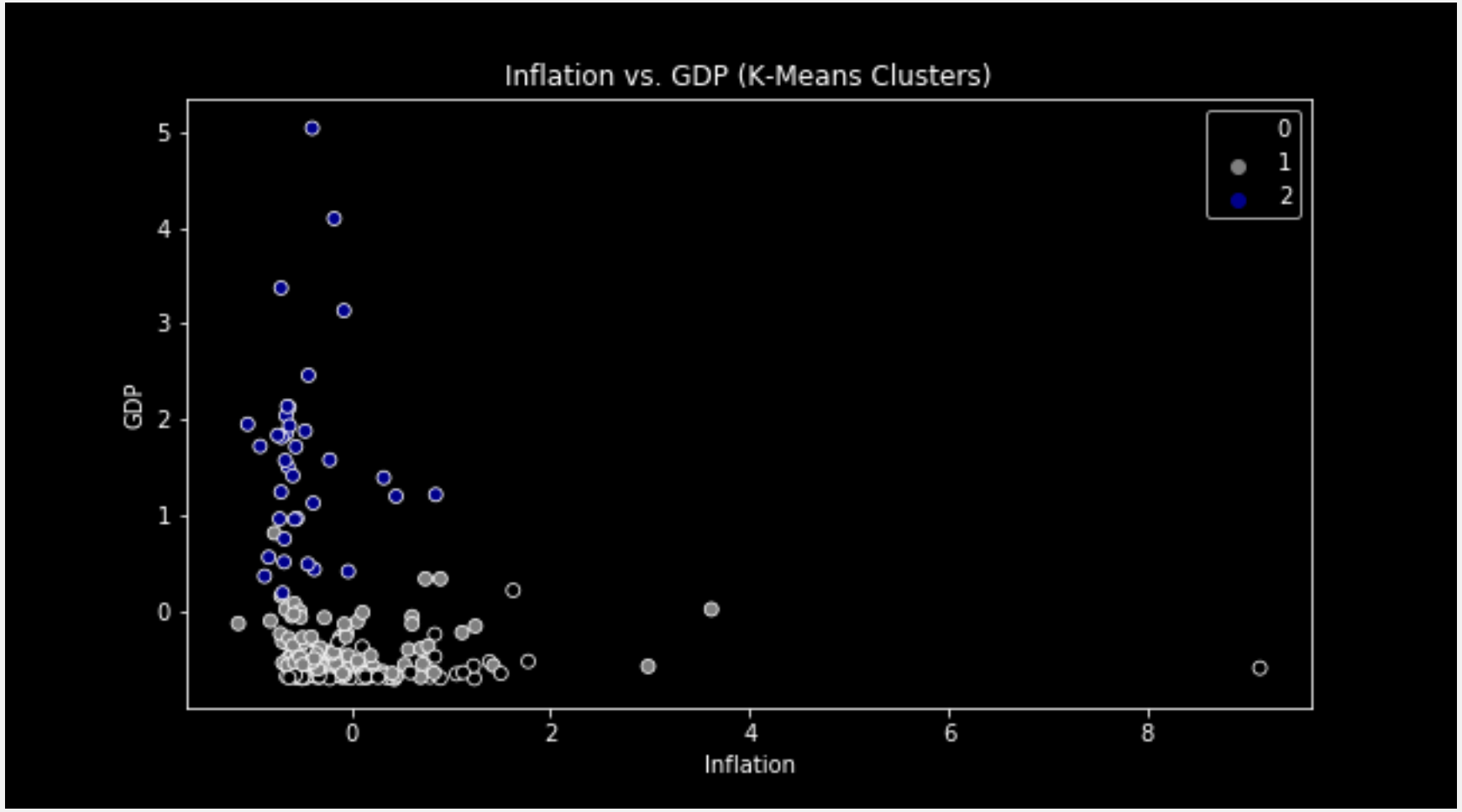
These are developed economies.

Here we can find the biggest economic growth, lowest inflation, higher life expectancy and higher health status.

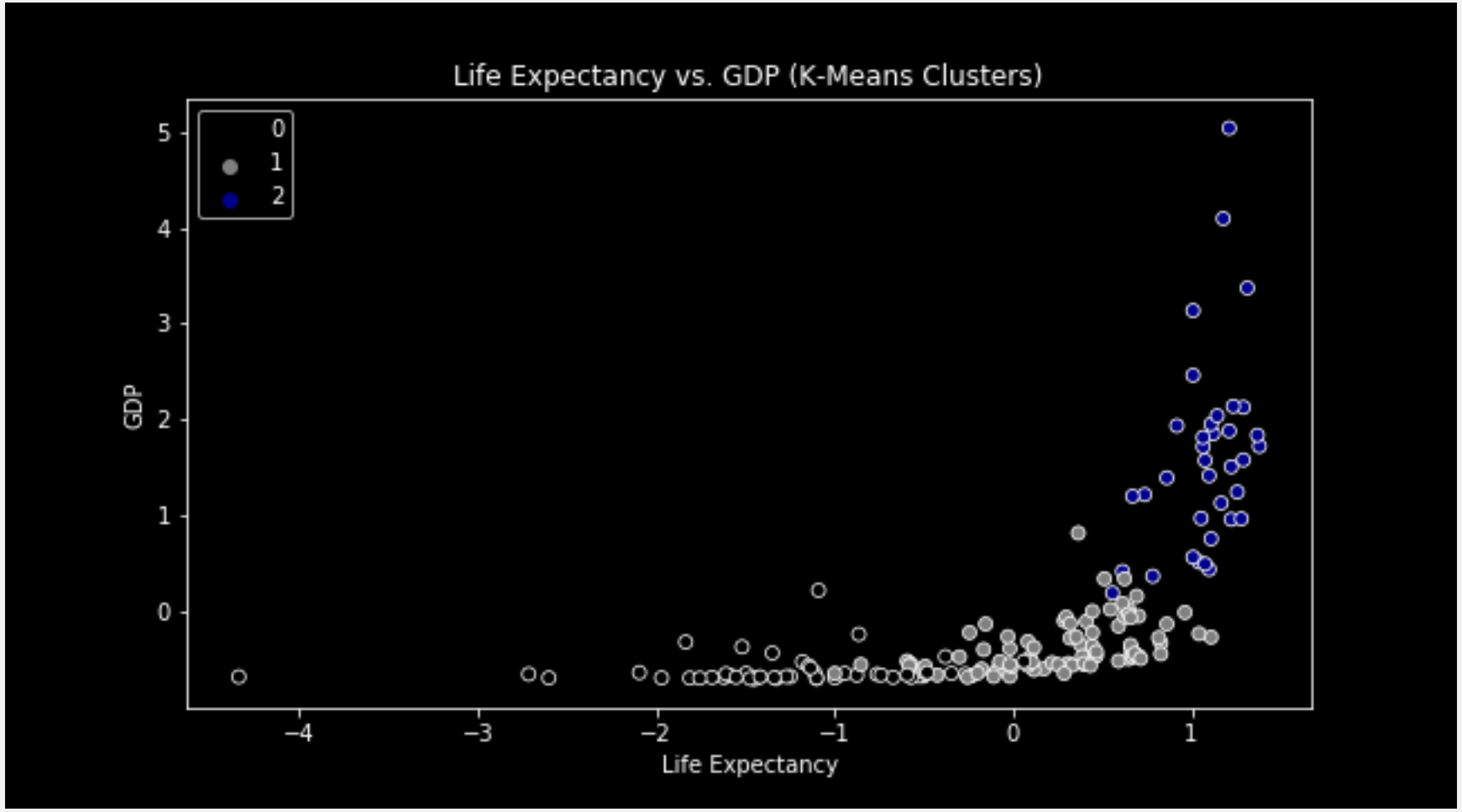
Visualizing the clusters on Income vs. GDP



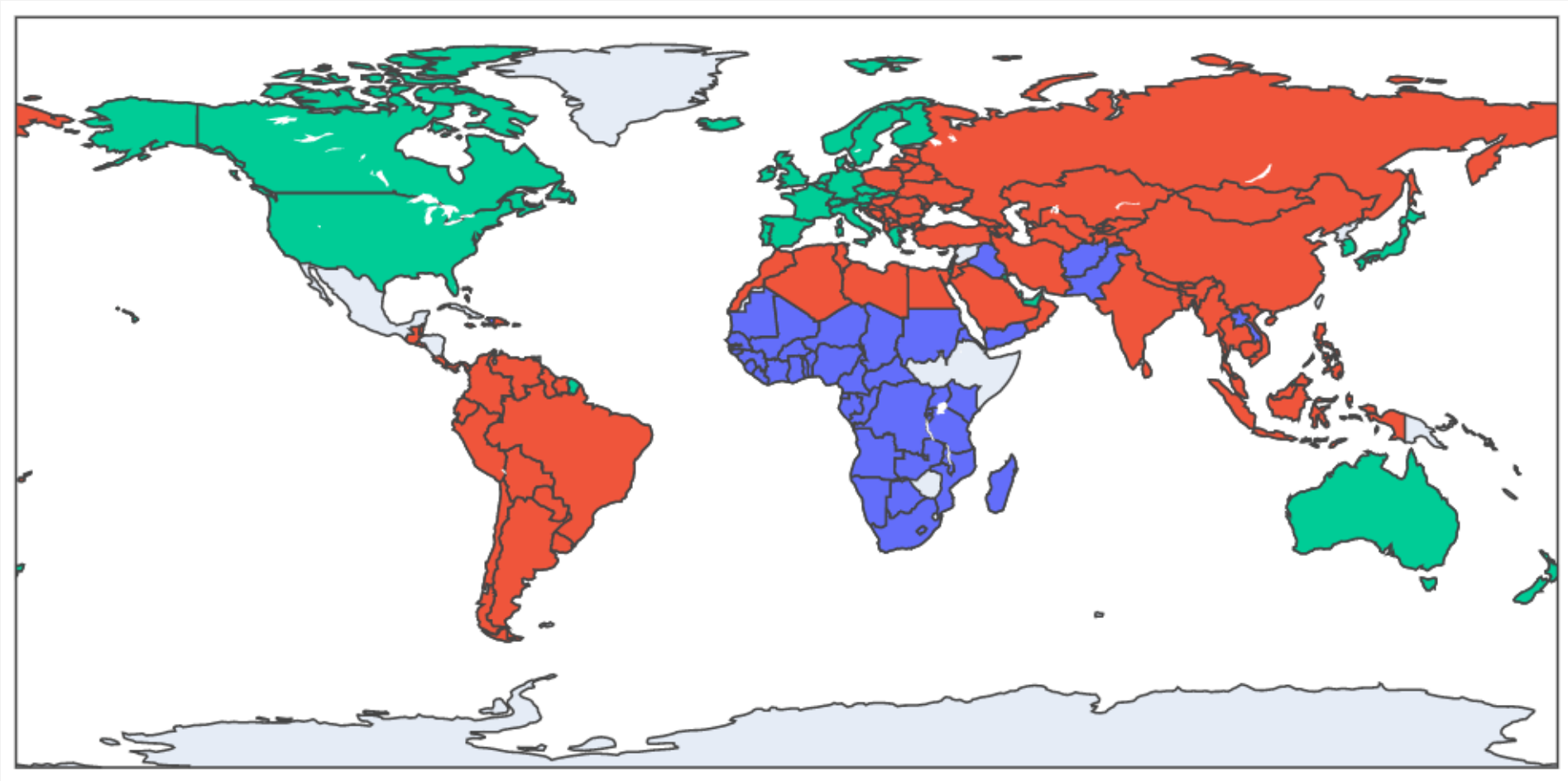
Visualizing the clusters on Inflation vs. GDP



Visualizing the clusters on Life Expectancy vs. GDP



Visualizing the clusters on geographical locations



Conclusions and recommendations

- Most countries located in central Africa are in dire need of help. They have the lowest health expenditure and the highest child mortality.

- Other countries located in the Middle East region, such as Pakistan, Iraq, Afghanistan, also show high levels of child mortality as well as high levels of inflation and low levels of economic growth.

- Our recommendation is to focus the organization resources specifically on those regions, and with special focus on improving the health system of those countries.

Thank you

Leon Felipe Galindo Barragan