

ANALYSIS ON COUNTRY DATASET

Submittedby: -

Ankit Gupta

Content

1. Introduction
2. Analysis
 - 2.1 Analysis on Outliers
 - 2.2 Analysis on Child Mortality Rate
 - 2.3 Analysis on Fertility Rate
 - 2.4 Analysis on Lowest Life Expectancy Rate
 - 2.5 Analysis on Lowest Health Spending
 - 2.6 Analysis on Lowest Per Capita Income
 - 2.7 Analysis on Highest Inflation Rate
 - 2.8 Analysis on Exports/Imports
 - 2.9 Data Preparation
 - 2.10 Silhouette Score
 - 2.11 K-Means Clustering
 - 2.12 Hierarchical Clustering
3. Conclusion

Introduction

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. It runs a lot of operational projects from time to time along with advocacy drives to raise awareness as well as for funding purposes.

After the recent funding program, they have been able to raise around \$ 10 million. Now the CEO of the NGO needs to decide how to use this money strategically and effectively. The significant issues that come while making this decision are mostly related to choosing the countries that are in the direst need of aid.

Objective: -

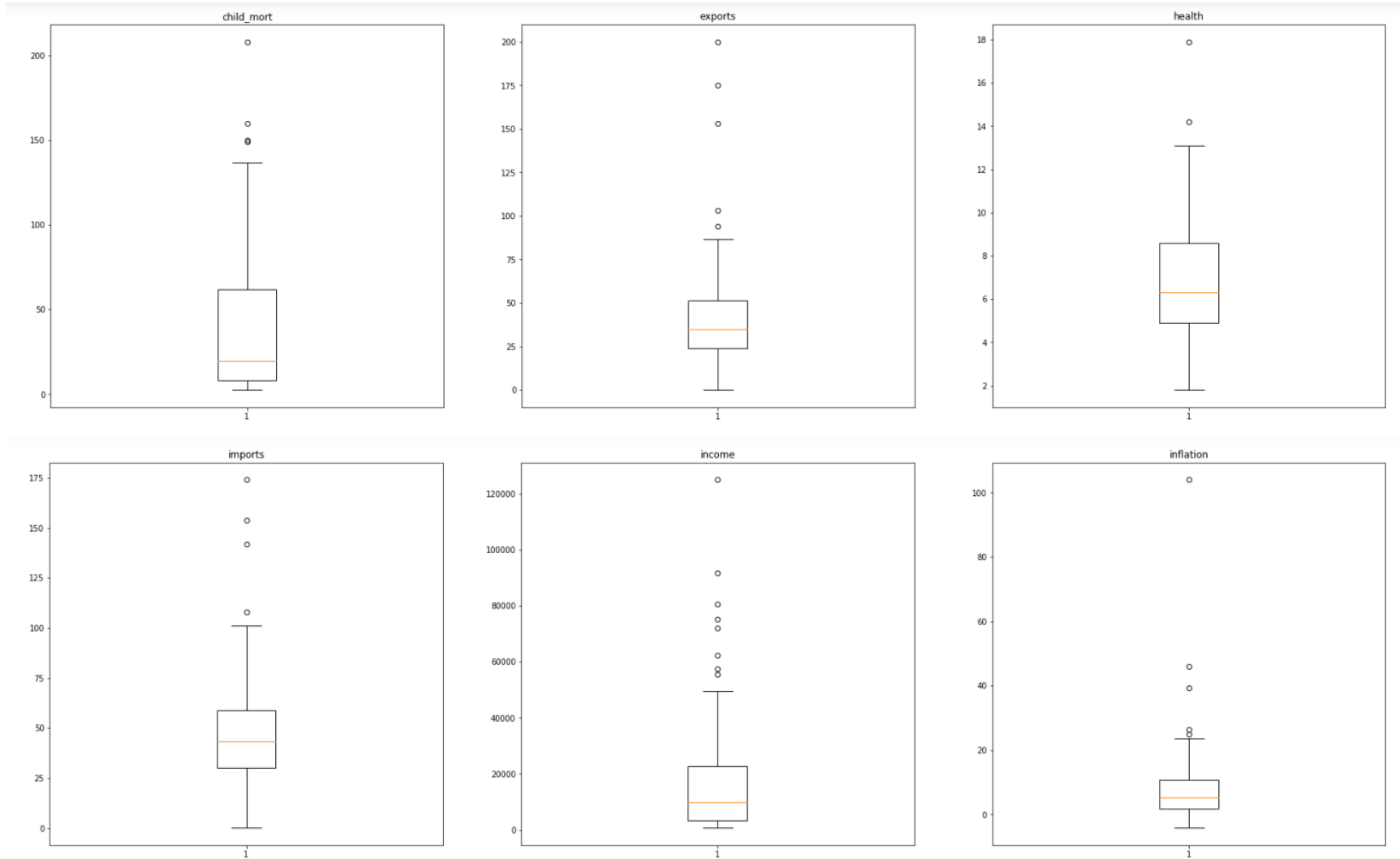
Our job is to categorize the countries using some socio-economic and health factors that determine the overall development of the country. Then we need to suggest the countries which the CEO needs to focus on the most.

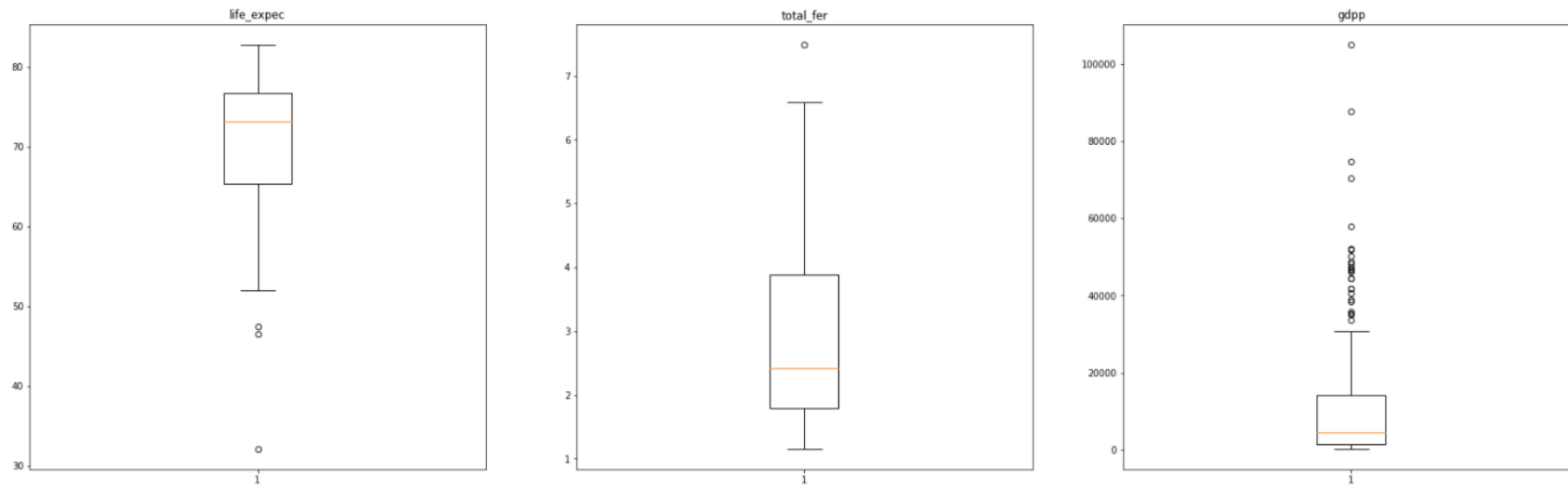
To achieve this objective, we need to analyze the data but before doing so, we need to increase the quality of the dataset for better understanding of the data, by doing the following steps: -

- Check necessary data inspection and EDA tasks suitable for the dataset
- Perform the Outlier Analysis on the dataset
- Try both K-means and Hierarchical clustering (both single and complete linkage) on the dataset to create the clusters.
- Analyze the clusters and identify the ones which are in dire need of aid
- Visualizations on the clusters that have been formed

Analysis

Analysis on Outliers: By placing the box plots

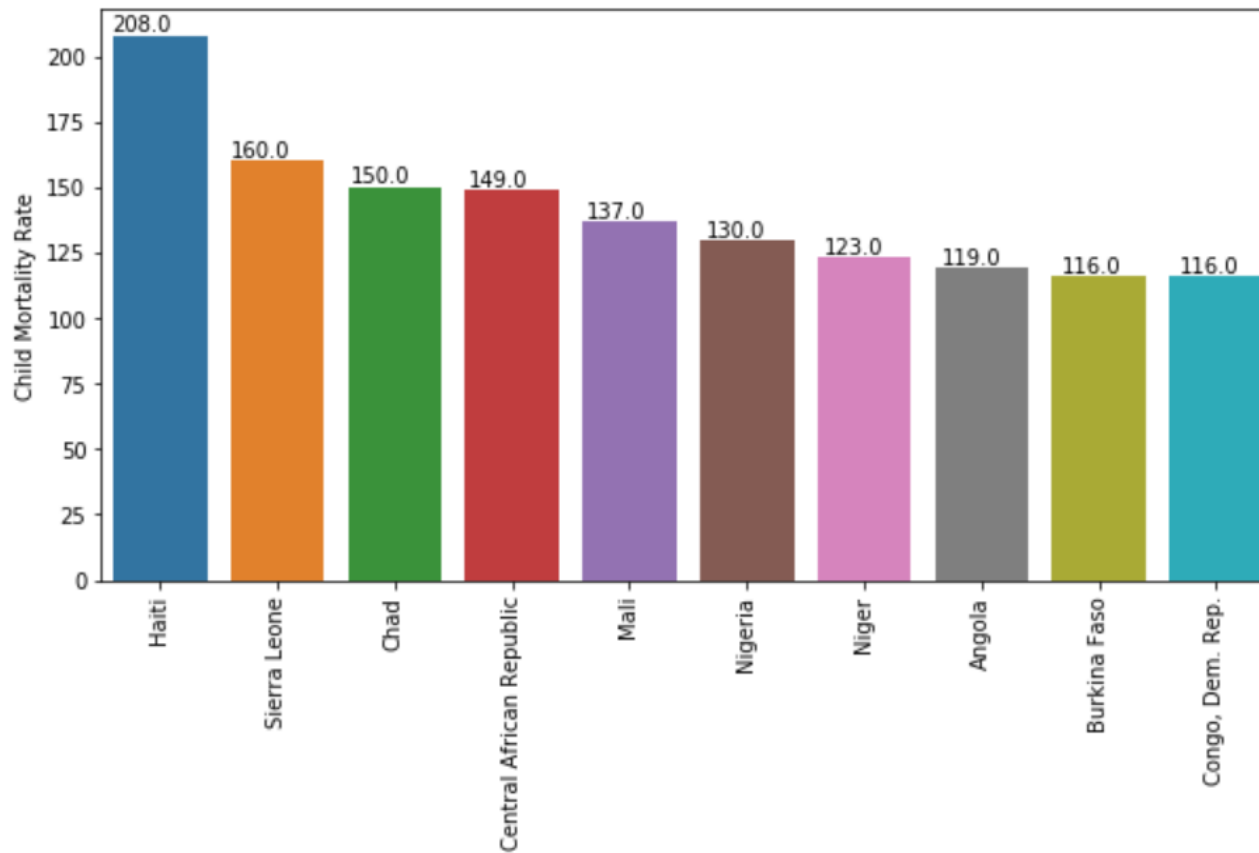




Insights: -

1. We can see that some of the features like Income, gdpp have outliers.
2. To handle the outliers, we must do capping on the upper limits.

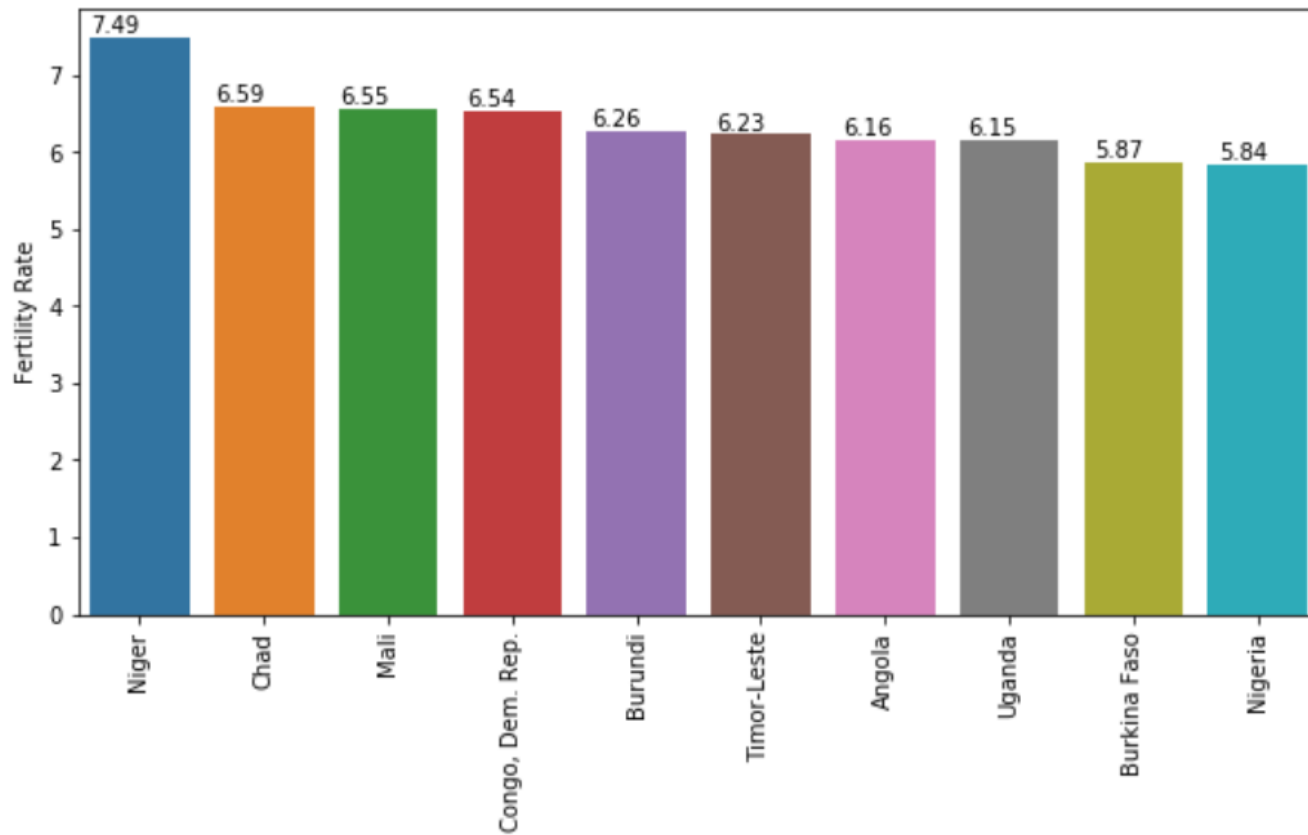
Analysis on Child Mortality Rate



Insights: -

1. Countries like Haiti, Sierra Leone, Chad, Central African Republic, Mali etc. having highest Child Mortality Rate are present in Africa having poor healthcare facilities.

Analysis on Fertility Rate

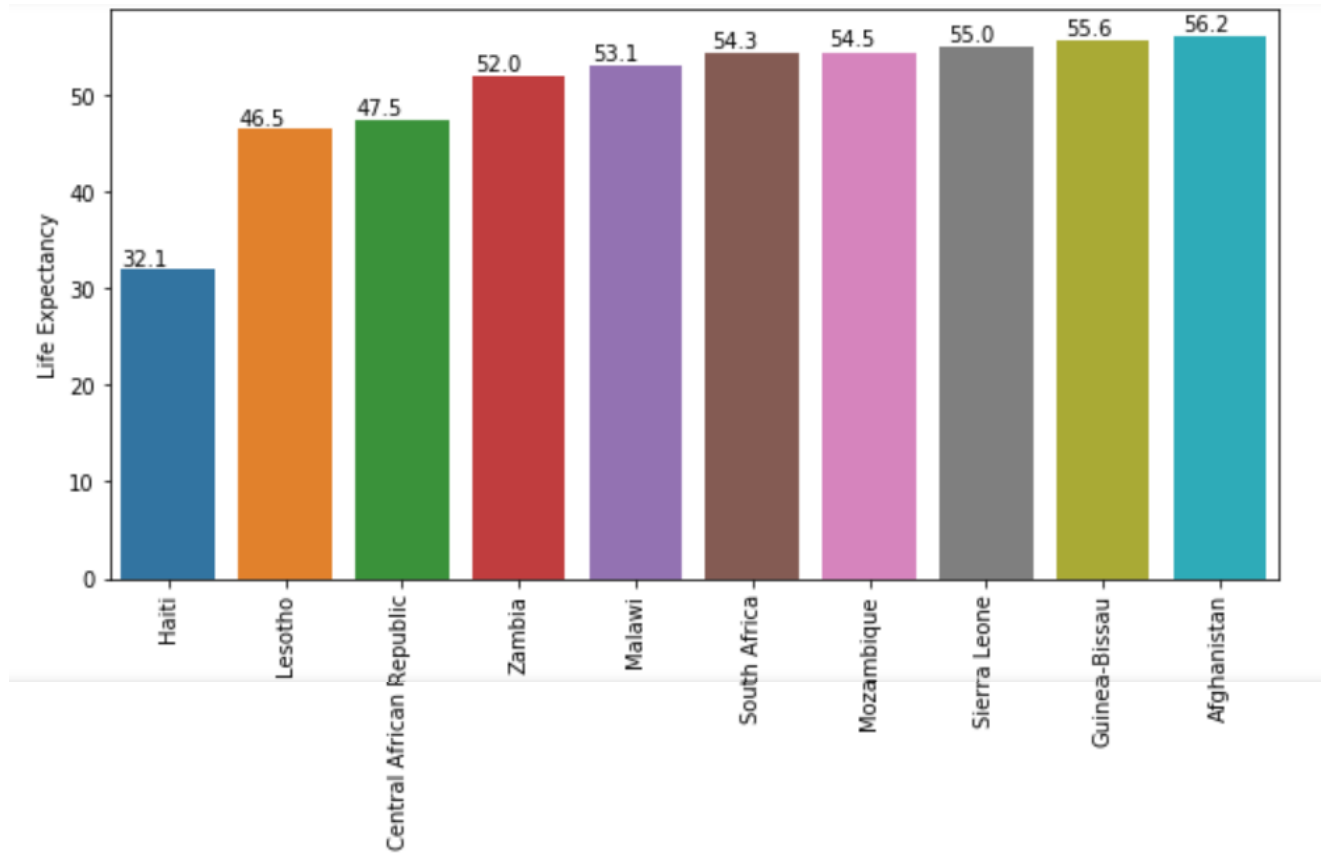


Insights: -

From the above graph we can observe: -

1. Countries like Niger, Chad, Mali etc. having highest Fertility Rate are the places where people are poorest in all.

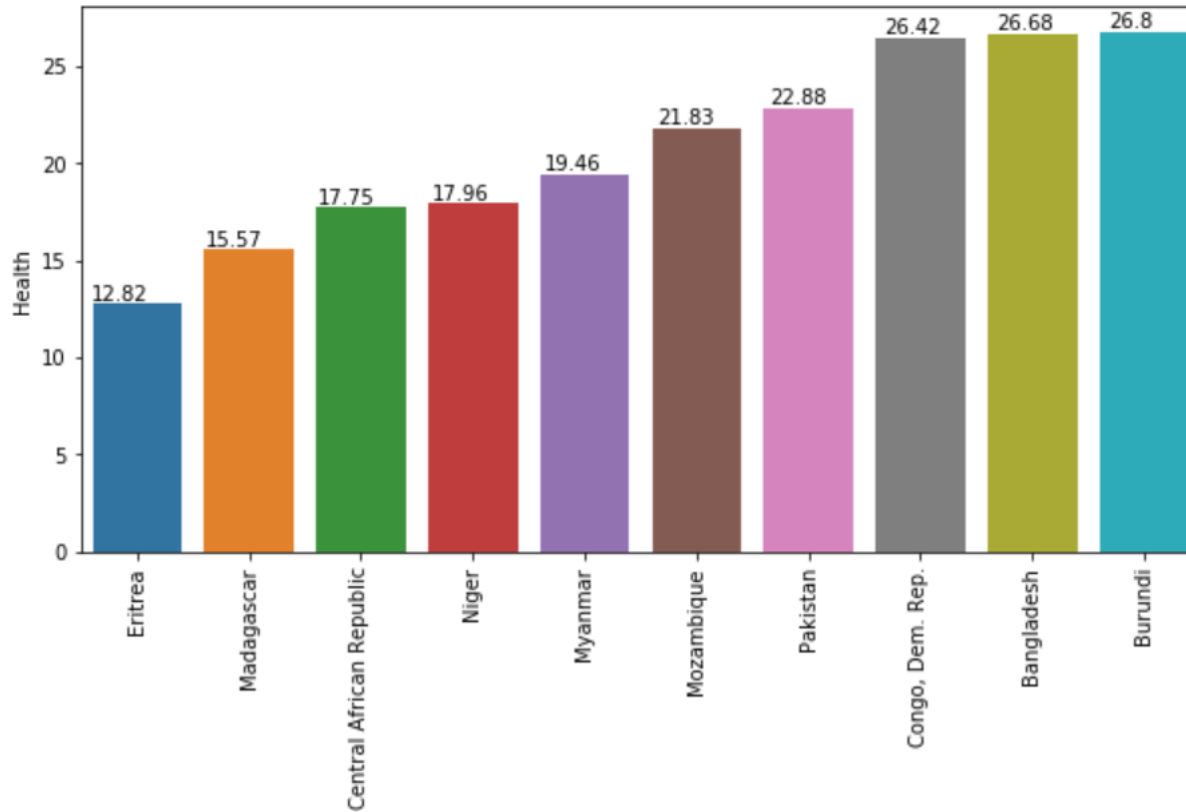
Analysis on Life Expectancy



Insight: -

1. Countries like Haiti, Central African Republic, Lesotho etc having lowest Life Expectancy are places where healthcare system is not available or efficient.

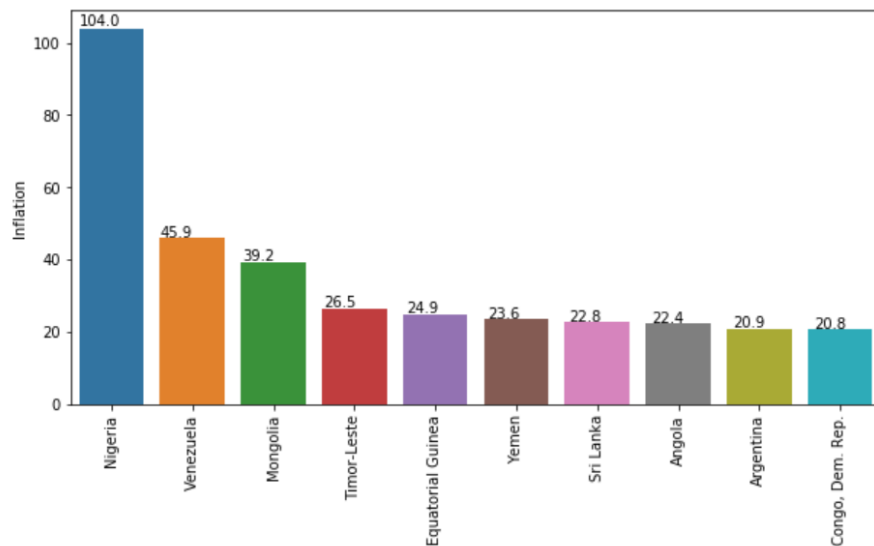
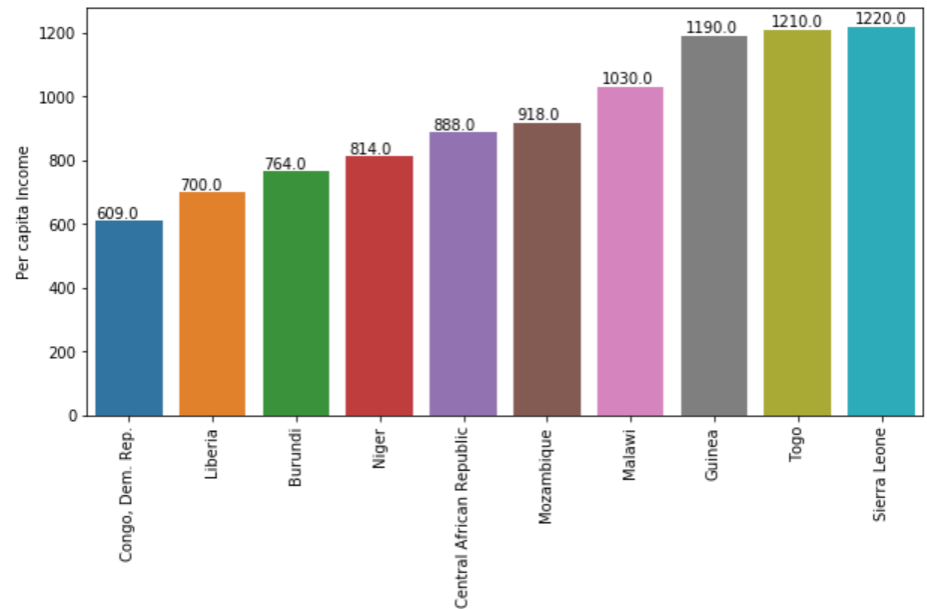
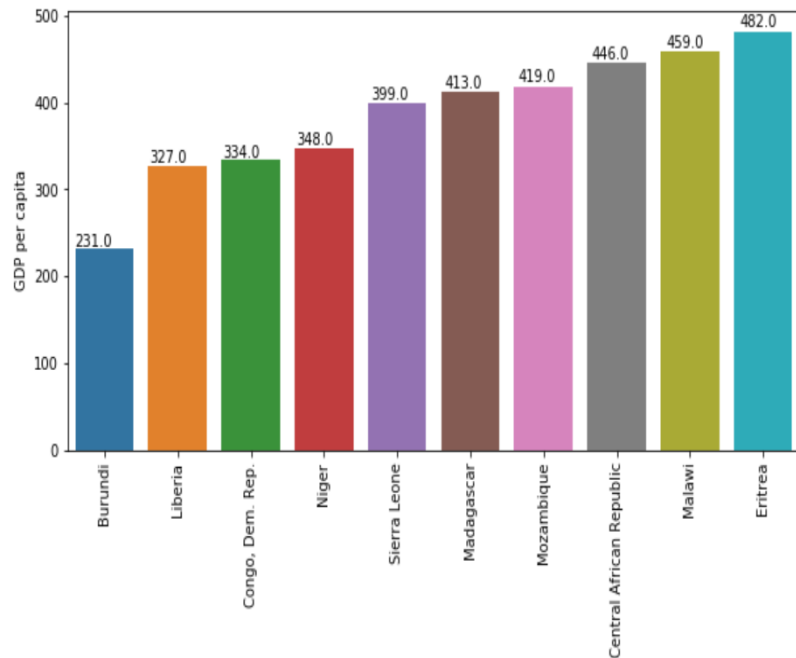
Analysis on Health



Insight: -

1. Top 10 Countries having lowest health spending are mix bag of countries where per Capita income is very high.

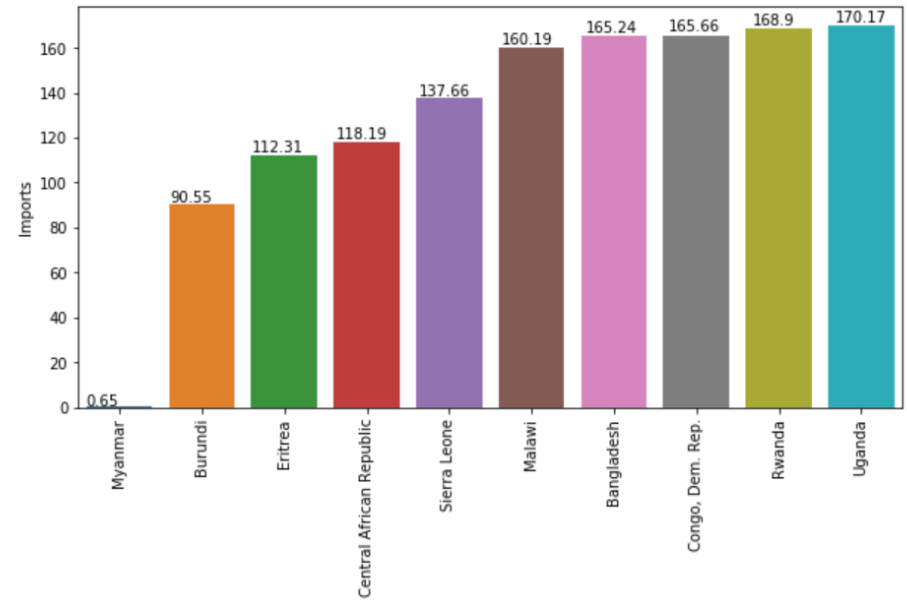
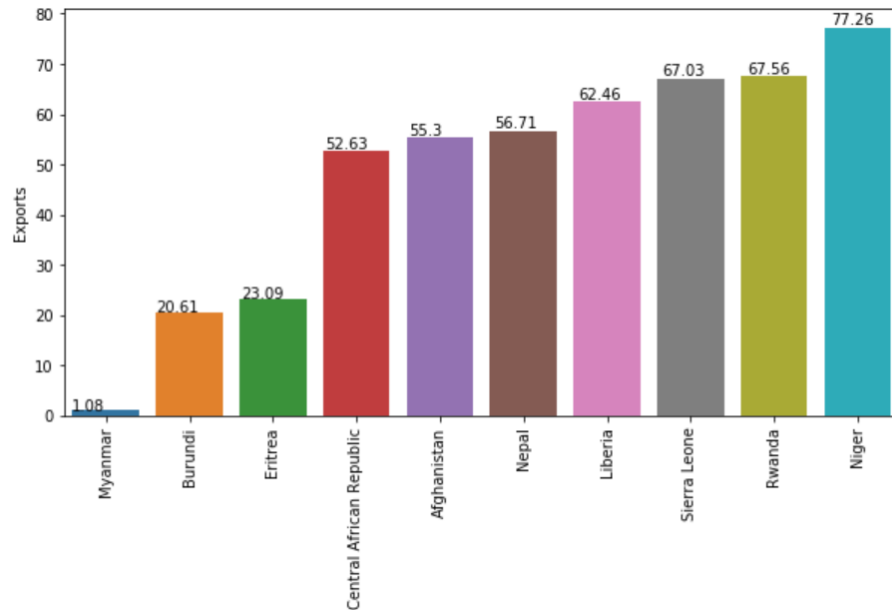
Analysis on GDP, Income per capita



Insight: -

- Top 10 Countries having lowest GDP per capita
- Top 10 Countries having lowest income per capita
- Top 10 Countries having highest inflation rate

Analysis on Exports/Imports



Insights: -

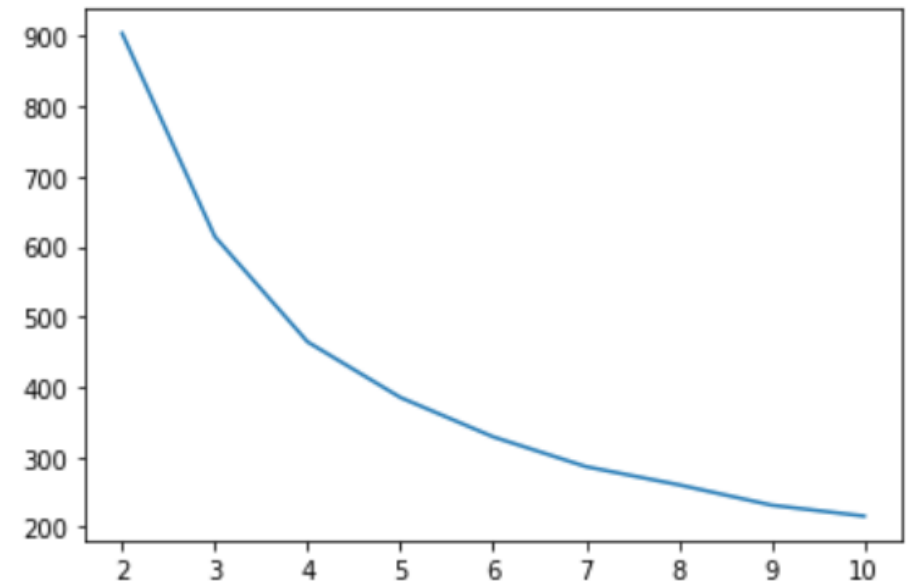
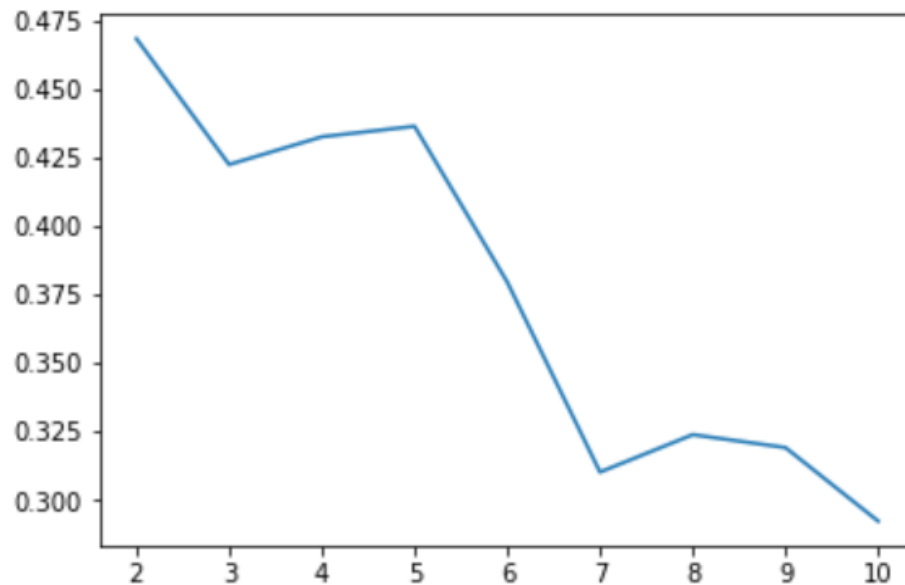
1. Top 10 Countries having lowest Exports are from mostly from underdeveloped / developing countries
2. Top 10 Countries having lowest Imports are from mostly from Developed nations who are Self-reliant, who manufacture locally to support own market or the underdeveloped / developing countries who are poor to afford imports due to political / economic turmoil

Hopkins Statistic

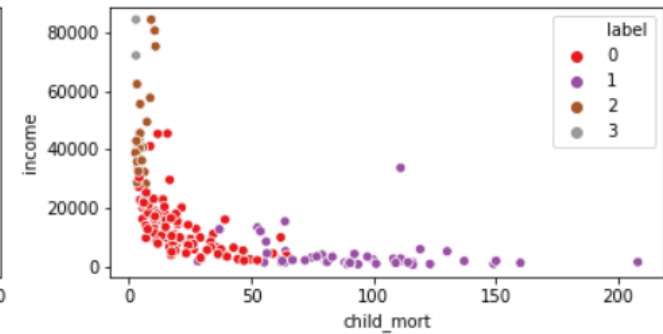
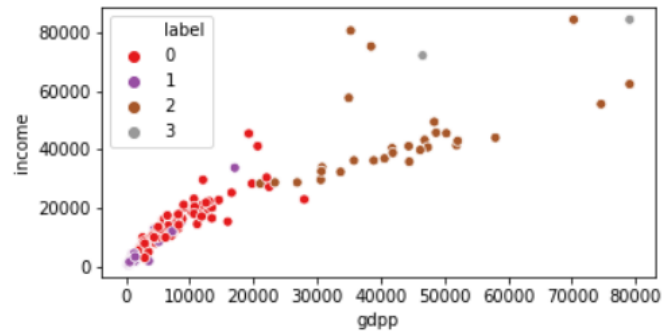
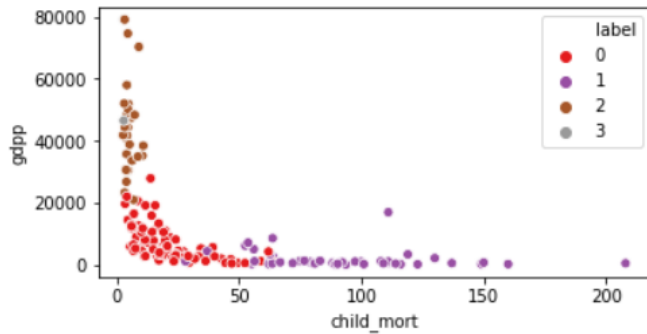
0.9434663900281776

As we can see that Hopkins score is around 94% which is a very good score and hence we can move forward with model building

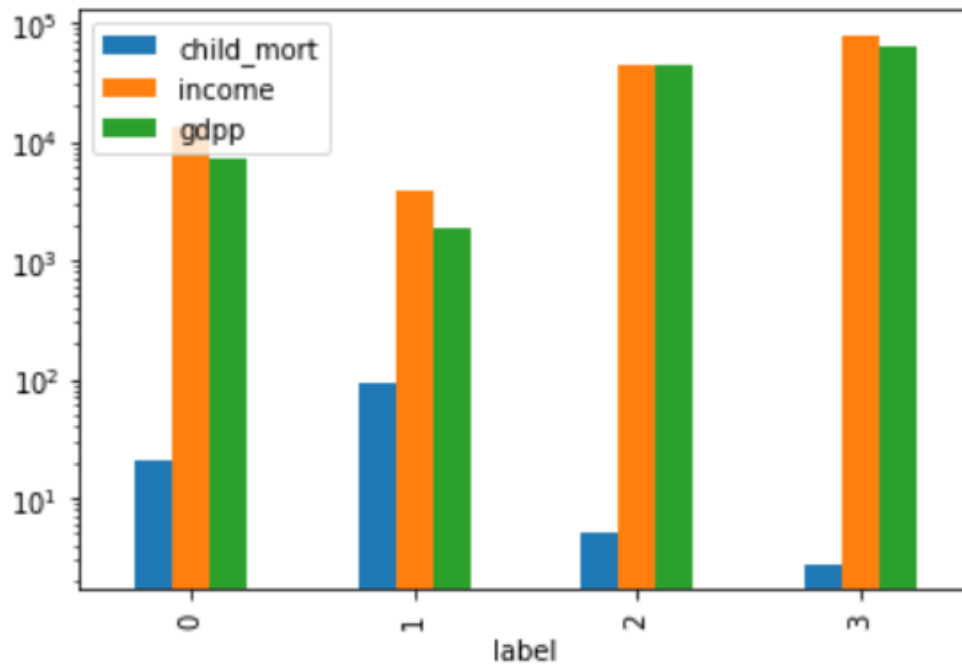
Silhouette Score and Elbow Curve



K-Means Clustering



Cluster Profiling



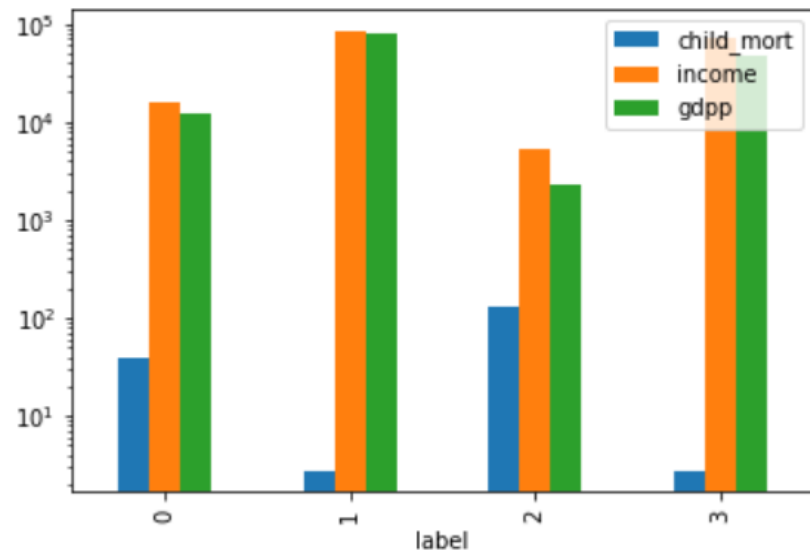
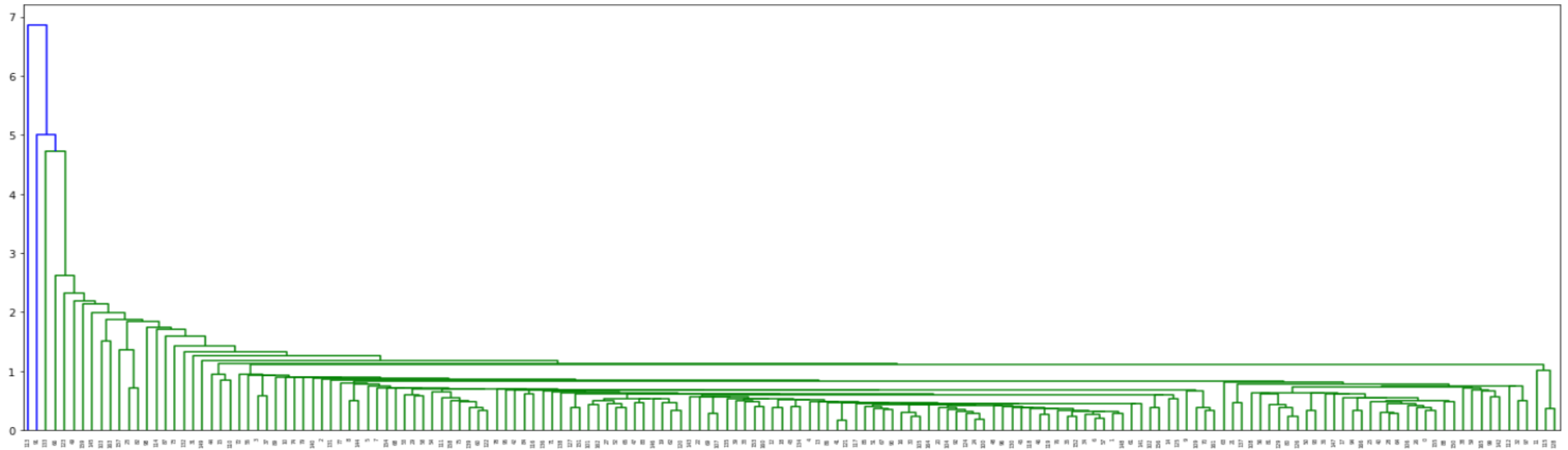
- Taking the cluster with Low income, Low GDP and High Child_mort
- Filter the data for that cluster

Top 5 countries with Low income, Low GDP and High Child_mort

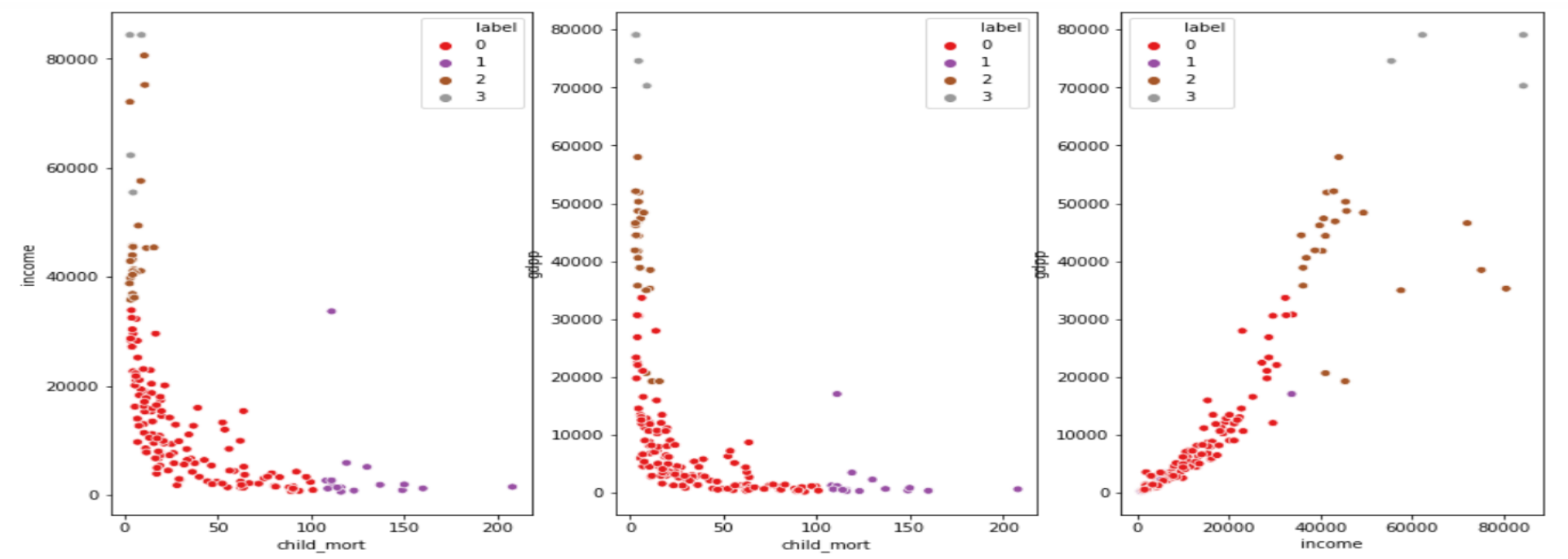
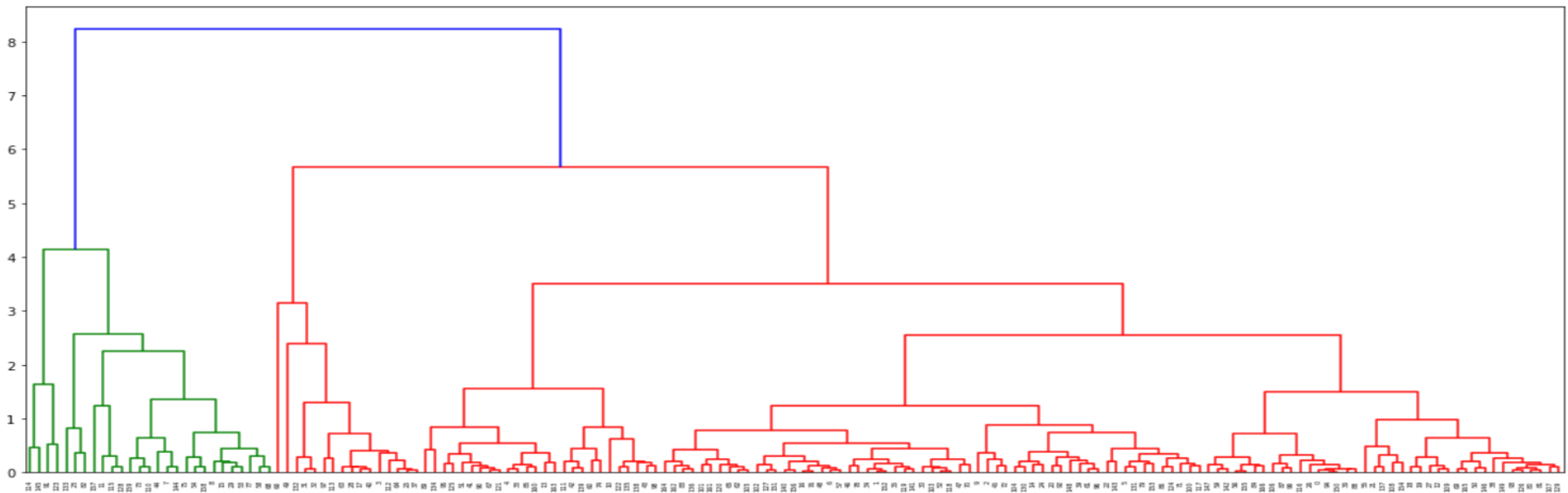
	country	child_mort	exports	health	imports	income	inflation	life_expec	total_fer	gdpp	label
66	Haiti	208.0	101.286	45.7442	428.314	1500.0	5.45	32.1	3.33	662.0	1
132	Sierra Leone	160.0	67.032	52.2690	137.655	1220.0	17.20	55.0	5.20	399.0	1
32	Chad	150.0	330.096	40.6341	390.195	1930.0	6.39	56.5	6.59	897.0	1
31	Central African Republic	149.0	52.628	17.7508	118.190	888.0	2.01	47.5	5.21	446.0	1
97	Mali	137.0	161.424	35.2584	248.508	1870.0	4.37	59.5	6.55	708.0	1

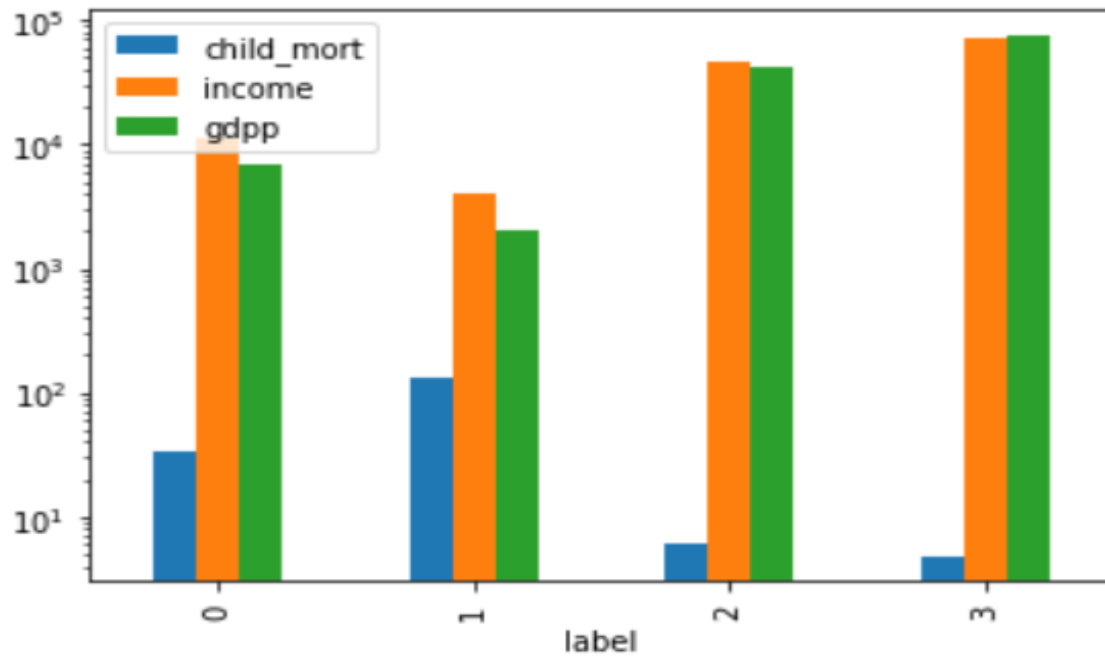
Hierarchal Clustering

Single Linkage



Complete Linkage





Top 5 countries with Low income, Low GDP and High Child_mort

	country	child_mort	exports	health	imports	income	inflation	life_expec	total_fer	gdpp	label
66	Haiti	208.0	101.286	45.7442	428.314	1500.0	5.45	32.1	3.33	662.0	1
132	Sierra Leone	160.0	67.032	52.2690	137.655	1220.0	17.20	55.0	5.20	399.0	1
32	Chad	150.0	330.096	40.6341	390.195	1930.0	6.39	56.5	6.59	897.0	1
31	Central African Republic	149.0	52.628	17.7508	118.190	888.0	2.01	47.5	5.21	446.0	1
97	Mali	137.0	161.424	35.2584	248.508	1870.0	4.37	59.5	6.55	708.0	1

Conclusion

Below is the key finding of the analysis of this case study: -

- Countries having low GDP, Income and Healthcare system having high child mortality rate.
- Countries with high inflation rate having high child mortality rate.
- Countries with high total fertility having high child mortality rate.
- Top 5 countries that are in the direst need of aid with Low income, Low GDP and High Child_mort is below.
 1. Haiti
 2. Sierra Leone
 3. Chad
 4. Central African Republic
 5. Mali