Introduction

It is, indeed, a great pleasure to be a part of the noble cause the HELP is dong and be able to contribute by able to guide them in focusing on the right countries which are in direst need of aid. The analysis ensures that the right countries (based on their socio economic and health factors) are being aided and the fund raised is utilize in the right direction.

Problem Statement and Objective

Objective

An international humanitarian NGO 'HELP' 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.

After the recent funding programs, they have been able to raise around $ 10 million and wanted to utilize this money strategically and effectively.

Problem Statement

While the CEO of help wants to ensure the amount raised from funding program is effectively utilized, the significant challenge come while making this decision is mostly related to choosing the countries that are in the direst need of aid.

Hence as a data analyst, we have to is to categorize the countries using some socio-economic and health factors that determine the overall development of the country. We need to suggest the CEO the right countries he should focus most

Analysis Methodology
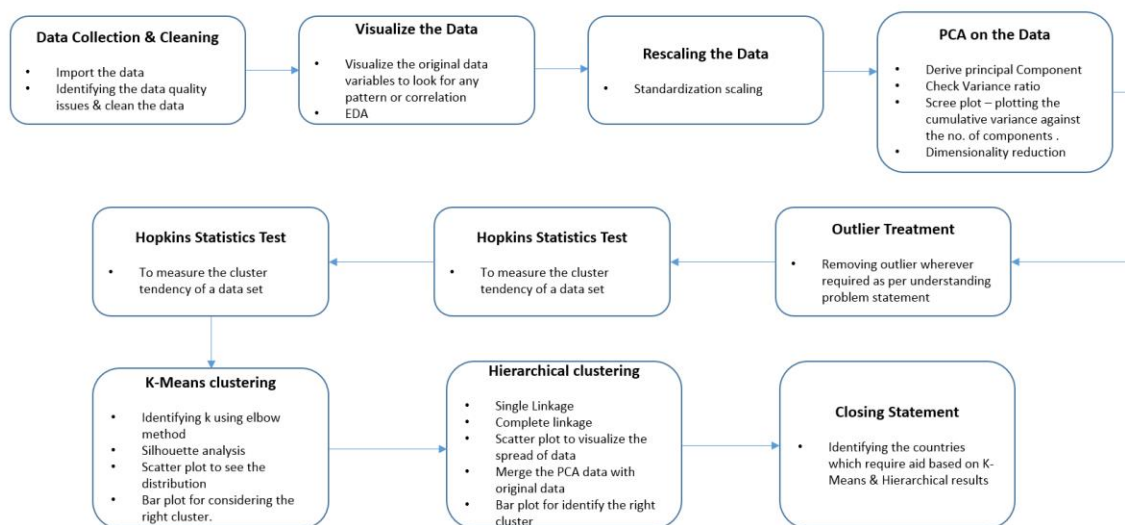
Data Provided

Based on the data provided , here are the different factors/variables provided for each country:

|	country	child_mort	exports	health	imports	income	inflation	life_expec	total_fer	gdpp
0	Afghanistan	90.2	10.0	7.58	44.9	1610	9.44	56.2	5.82	553
1	Albania	16.6	28.0	6.55	48.6	9930	4.49	76.3	1.65	4090
2	Algeria	27.3	38.4	4.17	31.4	12900	16.10	76.5	2.89	4460
3	Angola	119.0	62.3	2.85	42.9	5900	22.40	60.1	6.16	3530
4	Antigua and Barbuda	10.3	45.5	6.03	58.9	19100	1.44	76.8	2.13	12200

Changing the % values to absolute values

Three important factors- '**Exports', 'Health' and 'imports**' are provided in % values. Hence this might not give the clear picture of spending by the country.

For example two countries (Afghanistan & Albina) have similar import % but not necessarily have the same gdpp which doesn't give accurate of idea of country being develop or under develop.

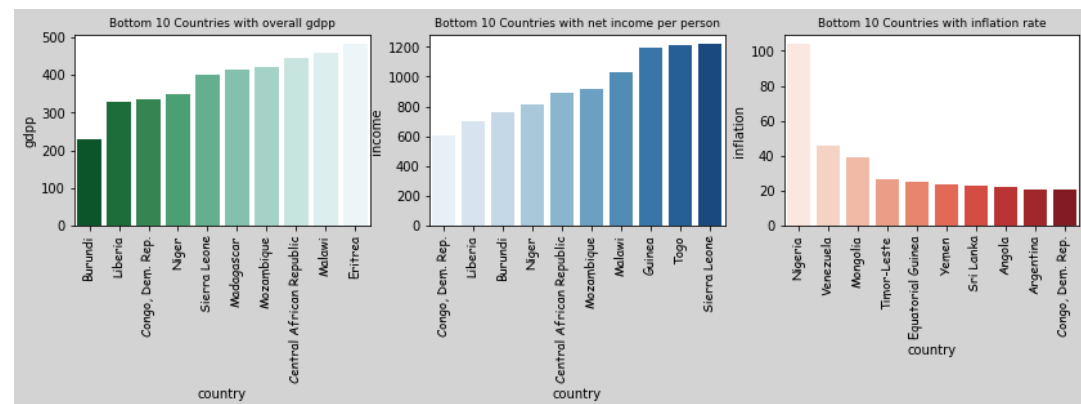Hence we need to derive the actual value of this variable.

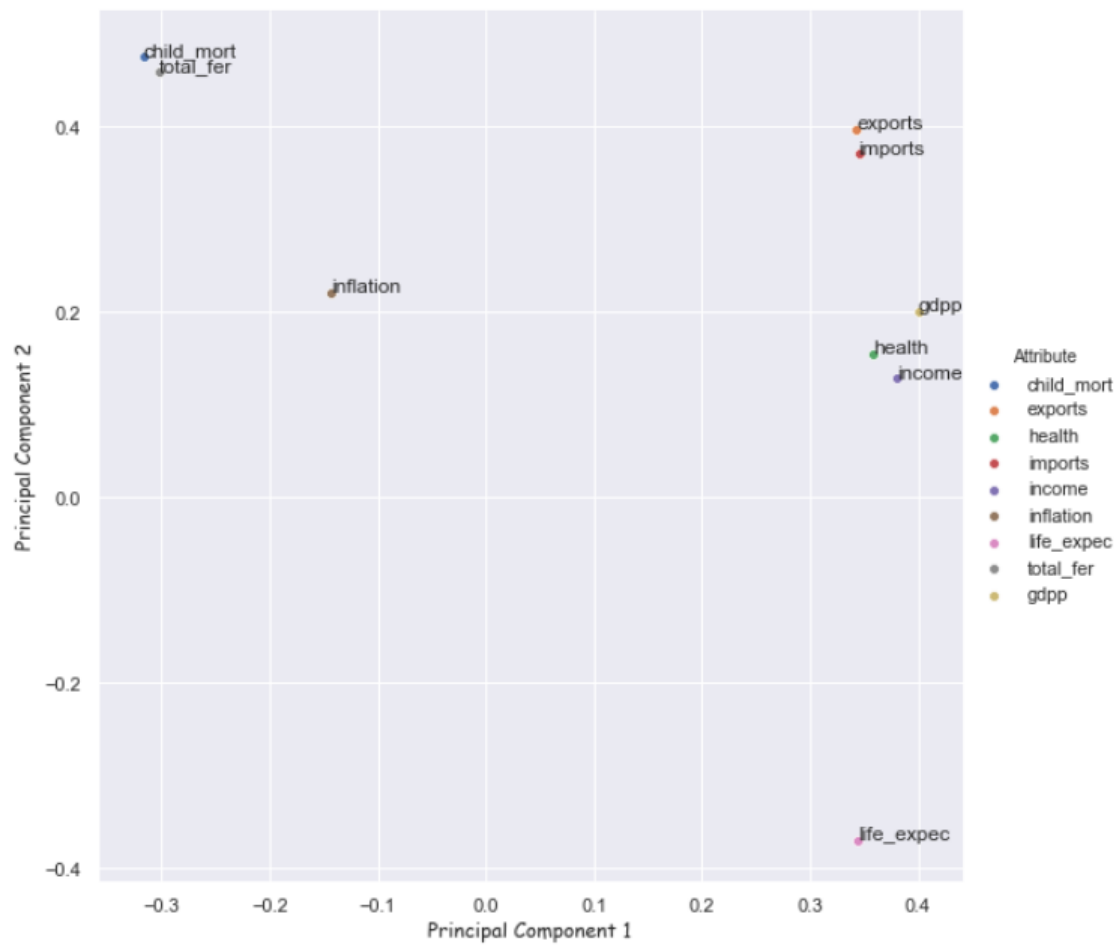# Graphs showing top 10 countries with lowest health, import & export
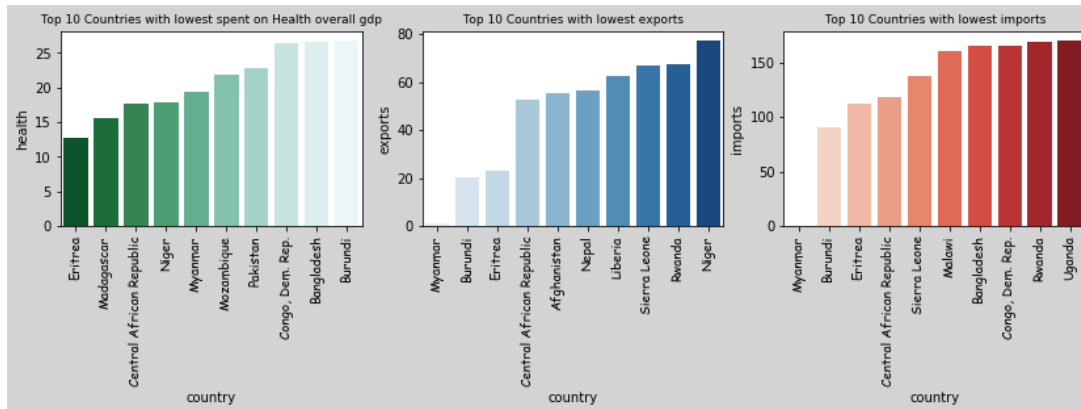
Graphs showing top 10 countries with highest child mortality and fertility rate and lowest life Expectancy
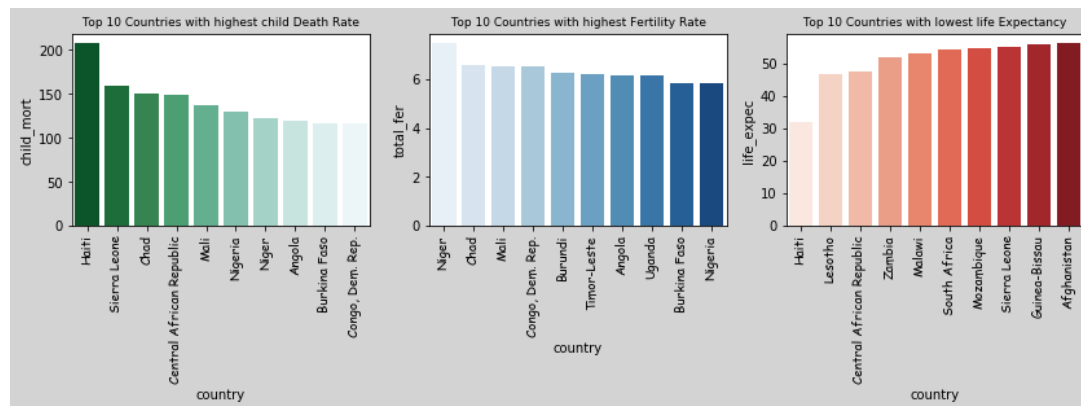
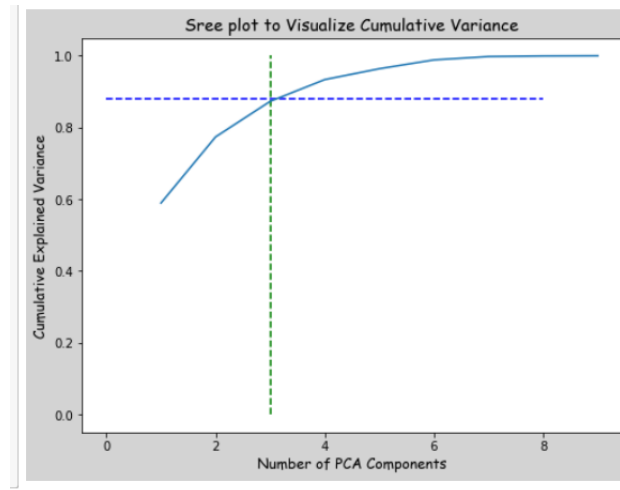# Graphs showing top 10 countries with Overall GDP, net income and inflation rate

Principal Component Analysis

- • We see features like life expectancy, income, gdpp and health are very well explained by PC1.
- • Features like child mortality, inflation and total fertility are well explained by PC2

Scree plot
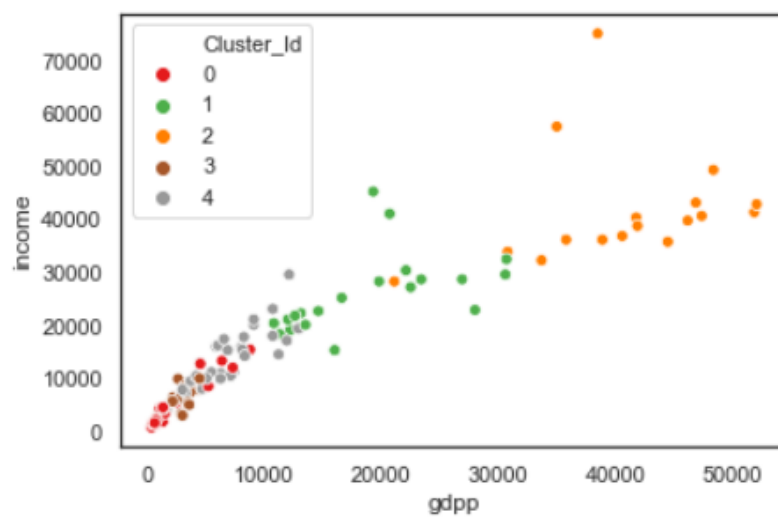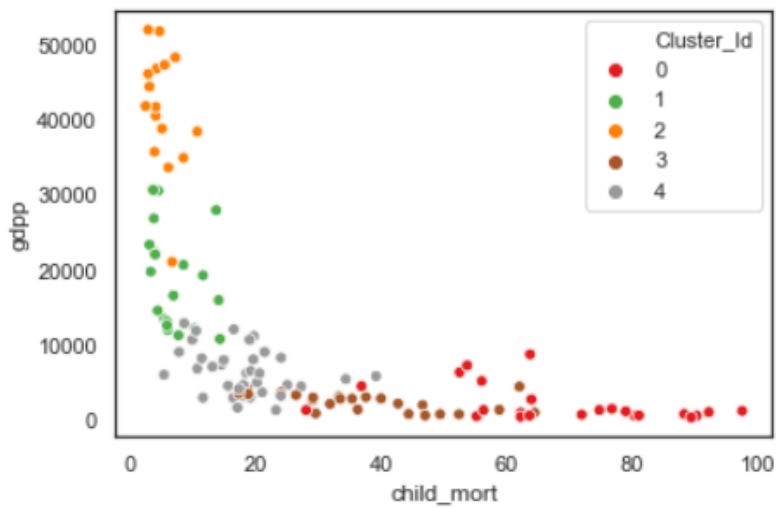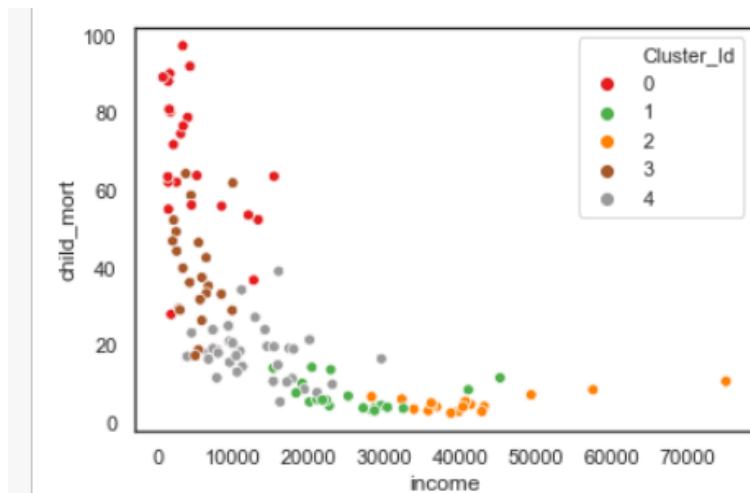

Sree plot to Visualize Cumulative Variance

From the above it is clear that first 3 principal components can well explain around 90% varaiance. Hence we will use them clustering process
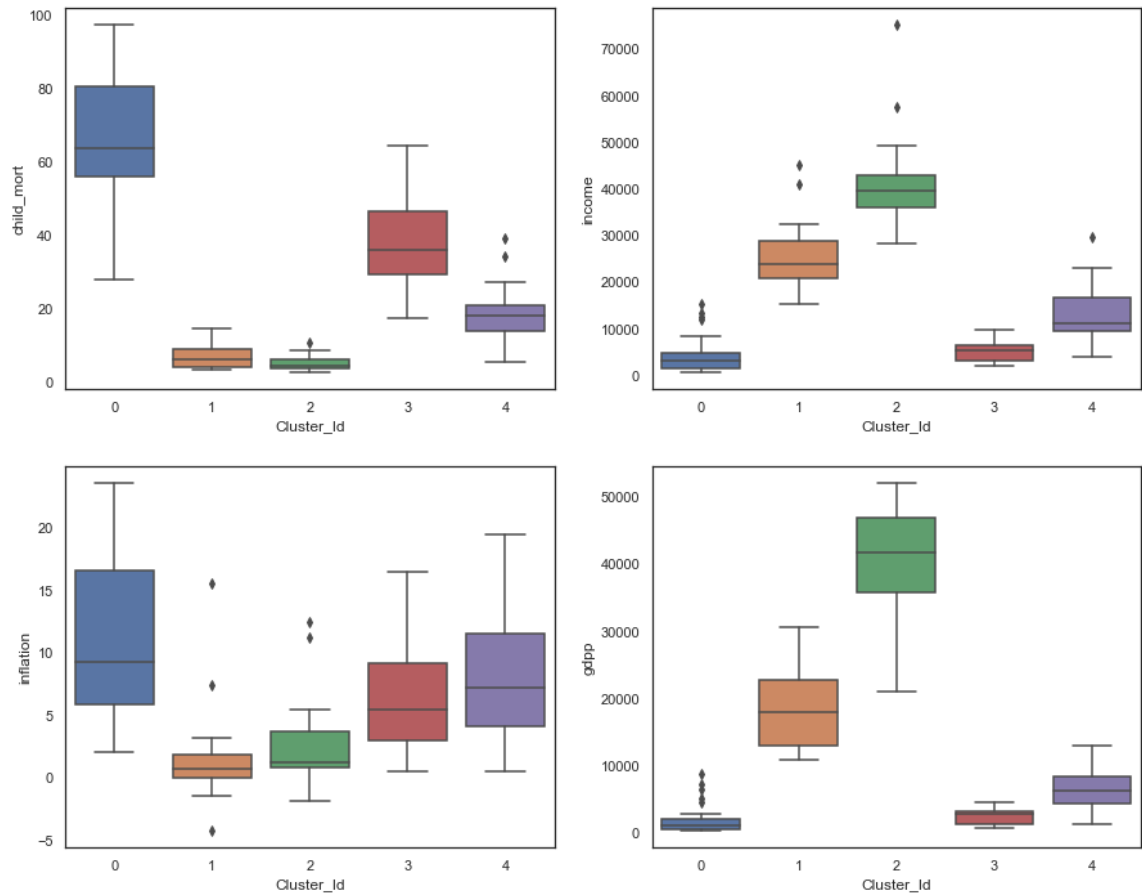
K-Means Analysis

1. Elbow Curve - Looking at the below elbow curve it looks good to proceed with either 4 or 5 clusters
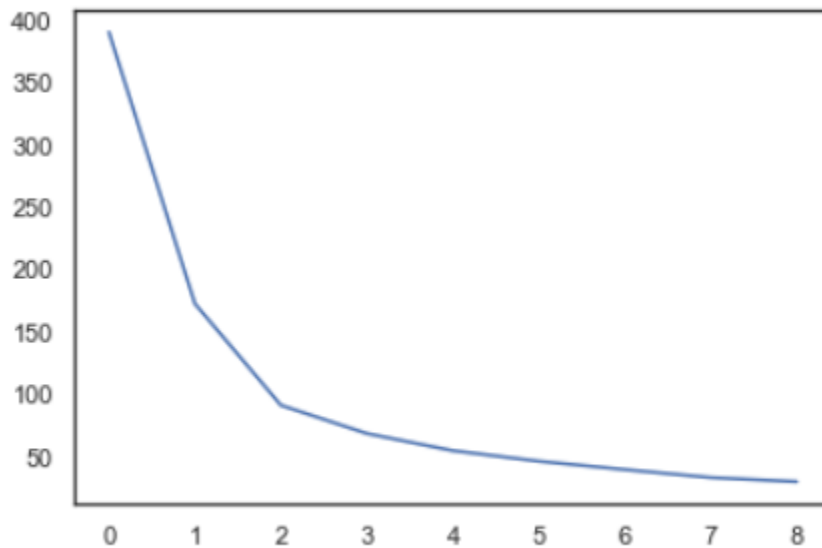

2. Scatter plot

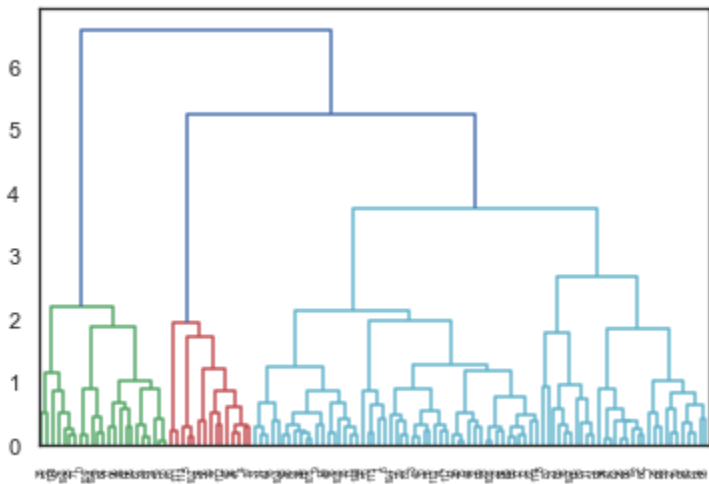3. Boxplot

Based on the scatter and boxplot –

- Child Mortality is highest for Cluster 0 and Cluster 3.These clusters need some aid.
- Income and Gdpp are measures of development. Higher the per capita income and gdpp better is the country's development. Income per capita and gdpp seems lowest for countries in clusters 0 and 3. Hence, these countries need some help.
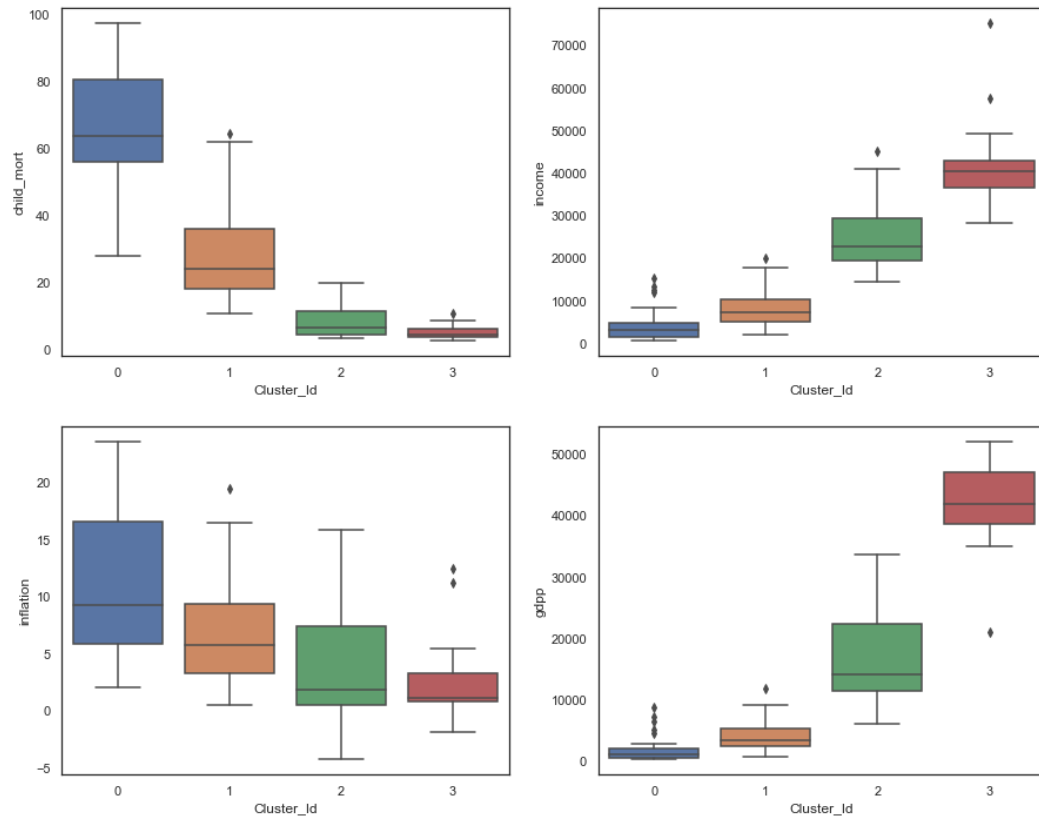
Hierarchical Clustering

1. Single linkage -
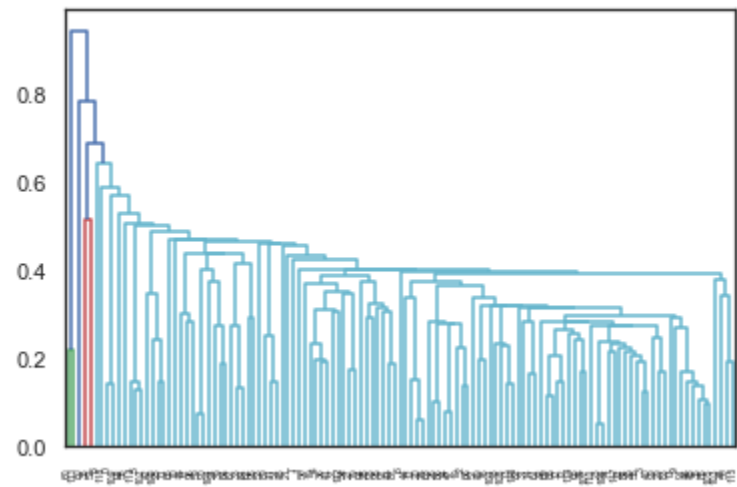
2. Complete linkage

3. Boxplot

- We have analyzed both K-means and Hierarchial clustering and found clusters formed are identical. The clusters formed in both the cases are not that great but gives some idea about countries which need of aid.

Closing statement

We have used PCA above to reduce the variables involved and then done the clustering of countries based on those Principal components and then later we identified few factors like child mortality, income etc which plays a vital role in deciding the development status of the country and builded clusters of countries based on that. Based on those clusters we have identified the below list of countries which are in dire need of aid. The list of countries are subject to change as it is based on the few factors like Number of components chosen, Number of Clusters chosen, Clustering method used etc.which we have used to build the model.

Final list of Countries that HELP should focus on

1. Afghanistan
2. Botswana
3. Comoros
4. Congo, Rep.
5. Eritrea
6. Gabon
7. Gambia
8. Ghana
9. Iraq
10. Kenya
11. Lao
12. Liberia
13. Madagascar
14. Mauritania
15. Namibia
16. Pakistan
17. Rwanda
18. Solomon Islands
19. South Africa
20. Sudan
21. Tanzania
22. Uganda
23. Yemen

Bar plots of countries against child mortality
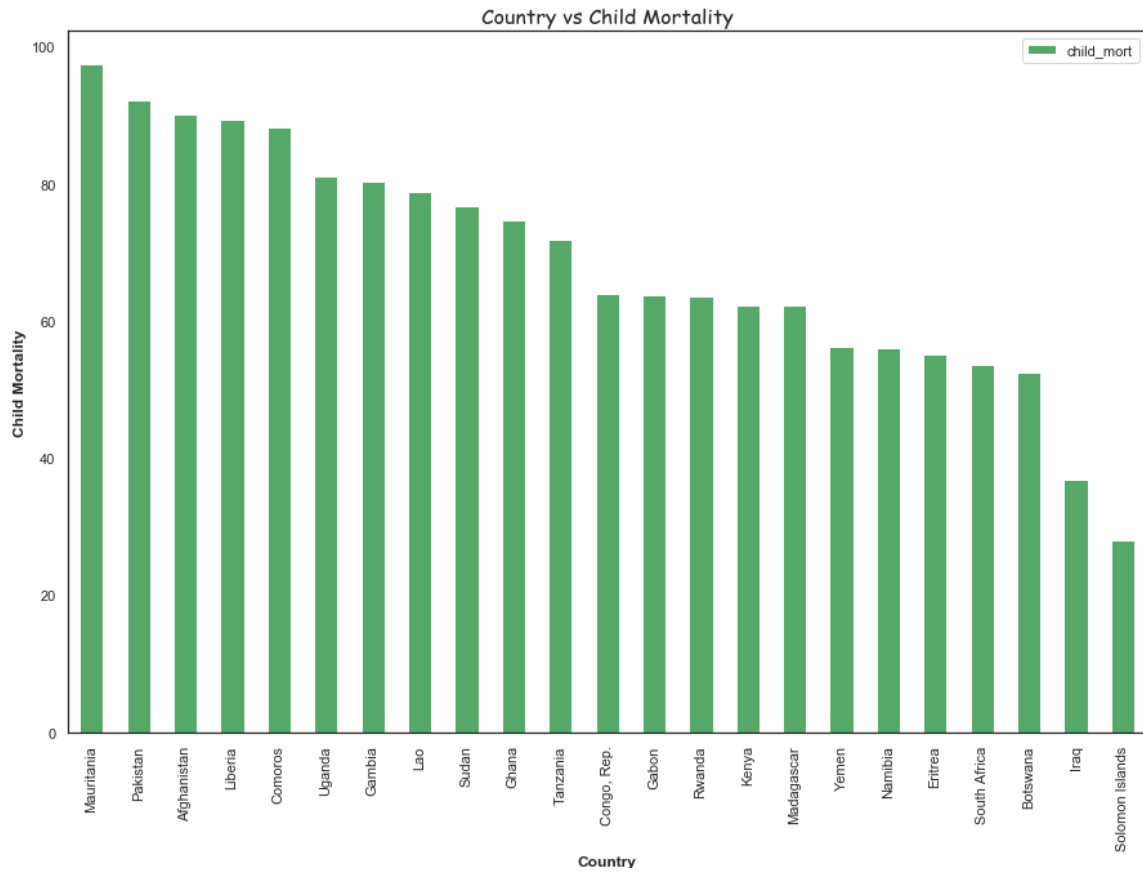
Bar plot of Country vs GDP per Capita
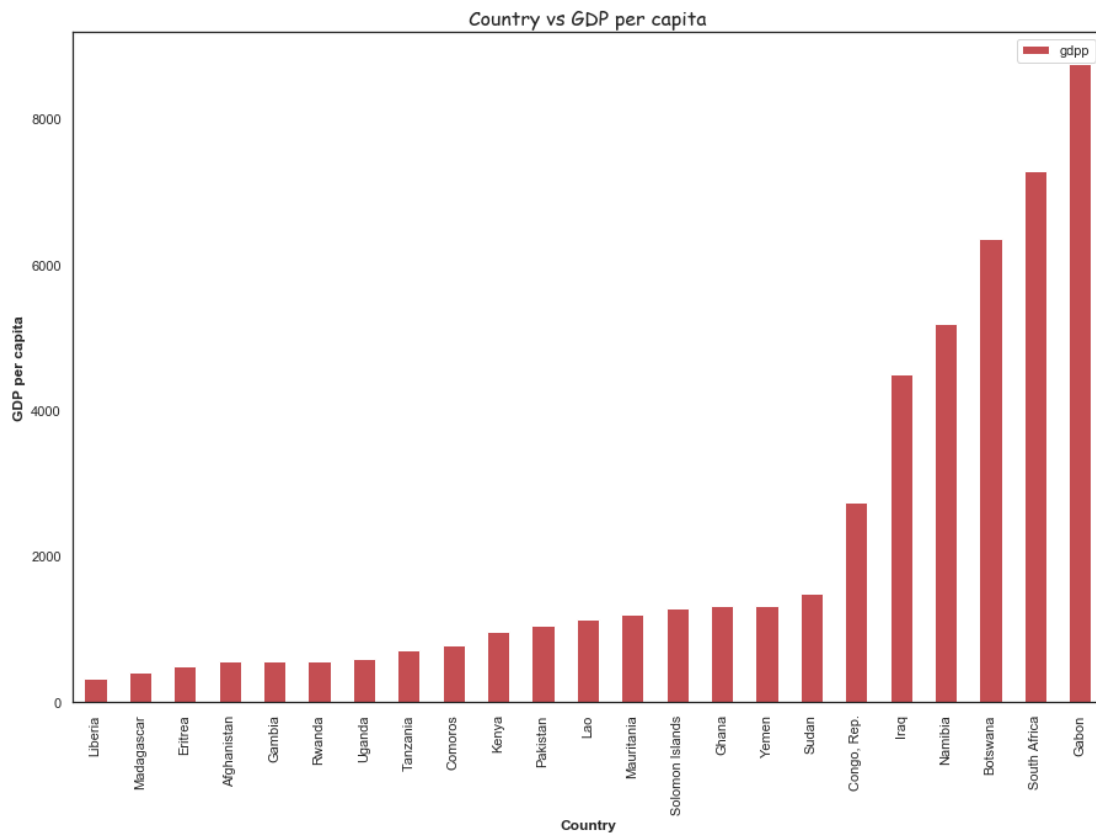
From the above plots we can coslude that countries like Libenia which has high child mortality rate and has least gdp per Capita should be considered as the top most countries for providing aid