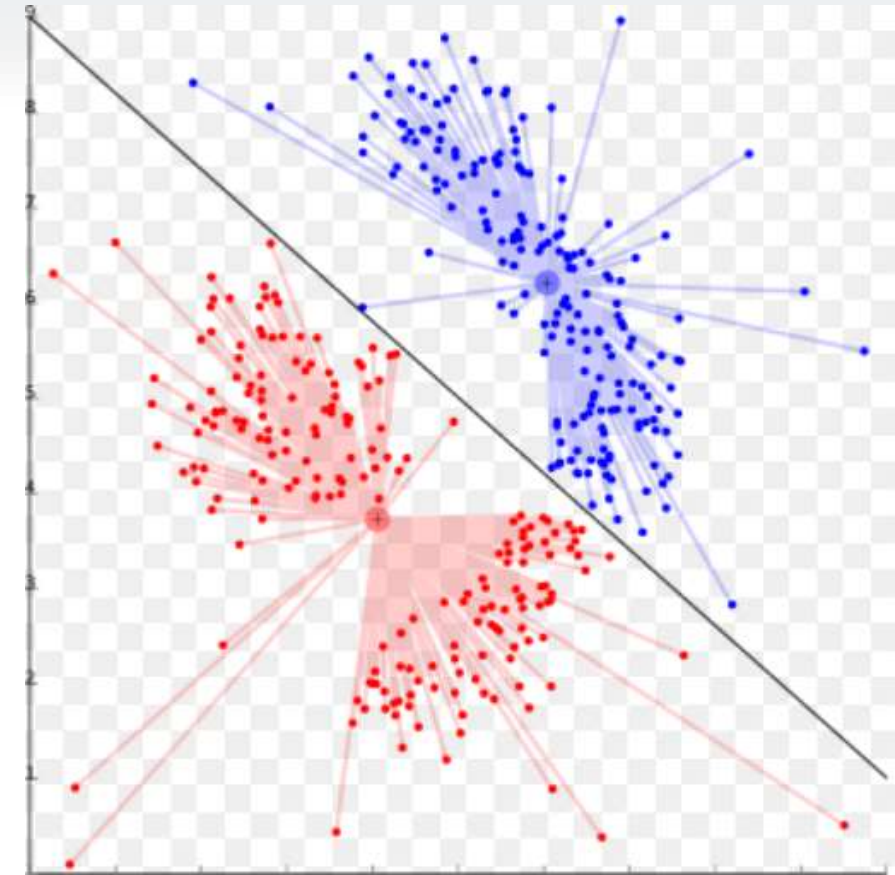
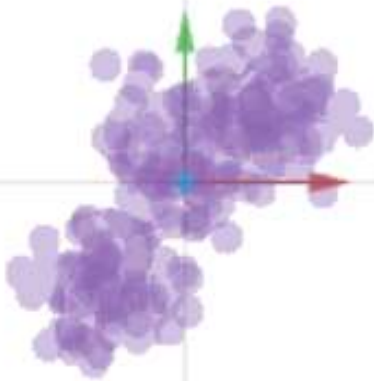


Clustering & PCA Assignment

CASE STUDY



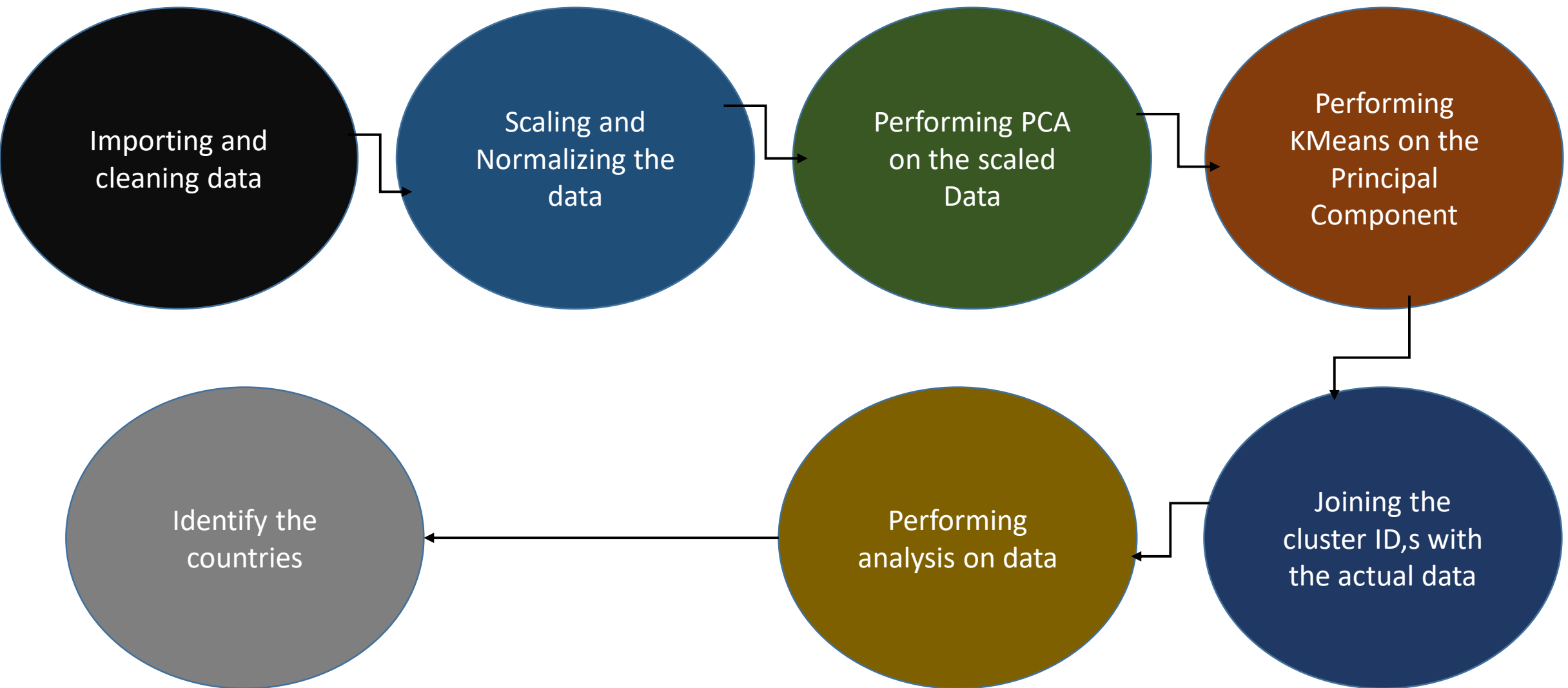
Abstract

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

Goals:

- *As a analyst we need to categories the countries using some socio-economic and health factors that determine the overall development of the country. Then you need to suggest the countries which the CEO needs to focus on the most.*

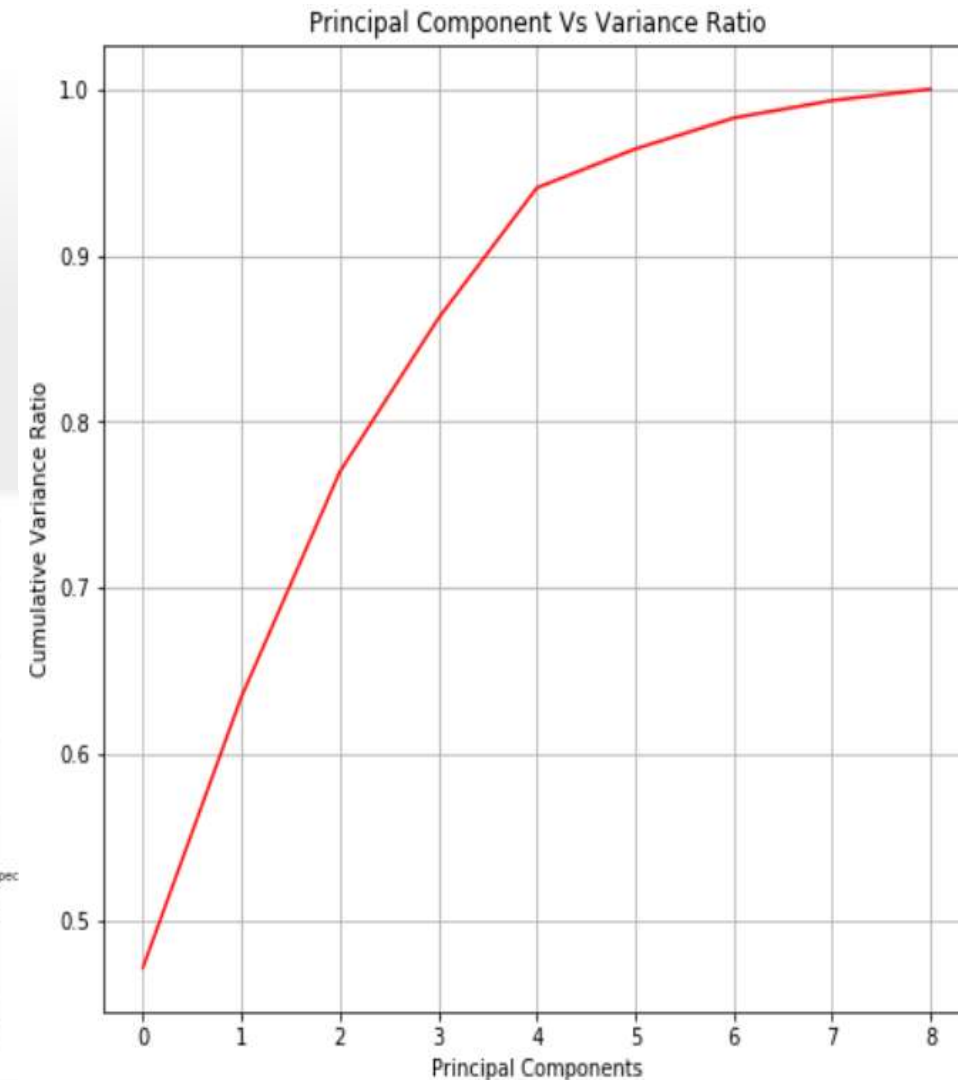
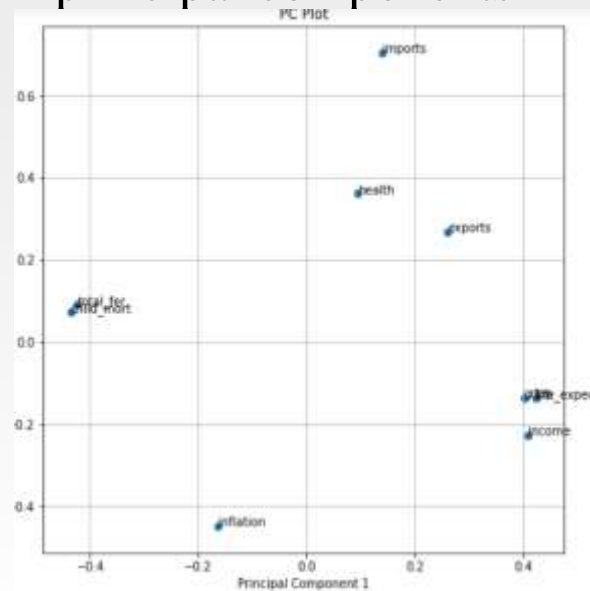
Problem solving methodology



PCA

We have performed the PCA on the scaled and normalized Data ,First I have performed PCA to the entire dataset and we Have got 9 vectors and 9 variance ratio in descending order , In order to find the appropriate number of PC which can explain Maximum variance we have used scree plot which uses cumulative Sum of variance and we have noticed that 4 principal components Can explain 95% of variance.

We have also plotted the PC1 and PC2 With features to get a good understanding Of data.

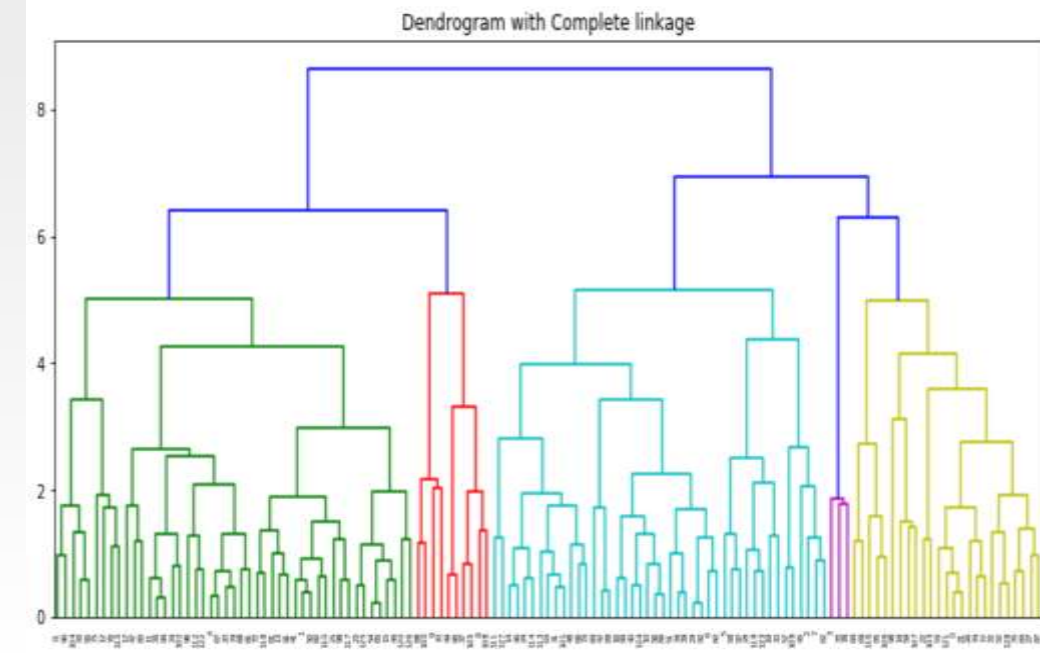
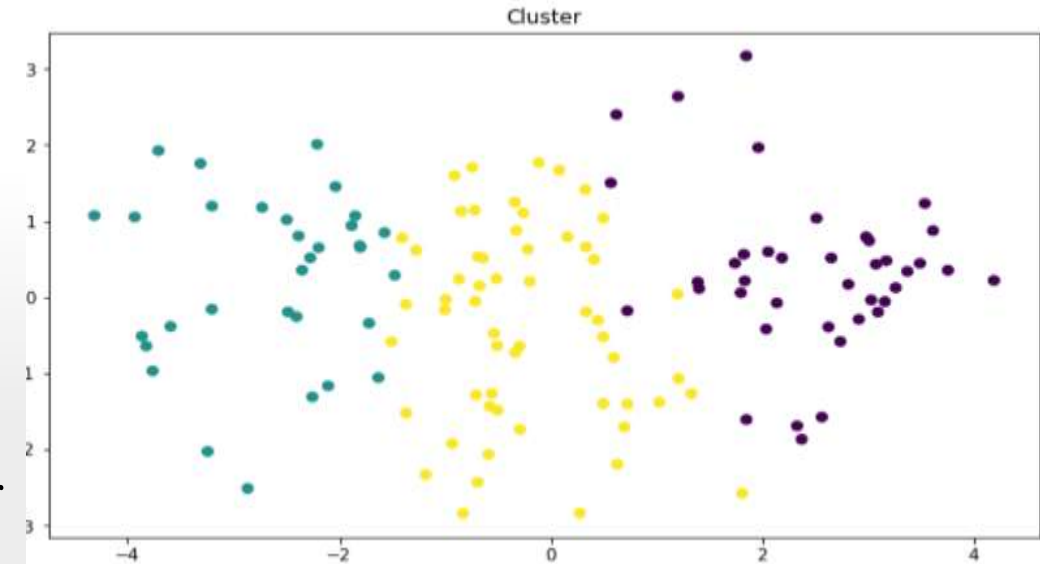


Clustering

First we have calculated Hopkins statistics to check if the dataset is feasible to perform clustering.

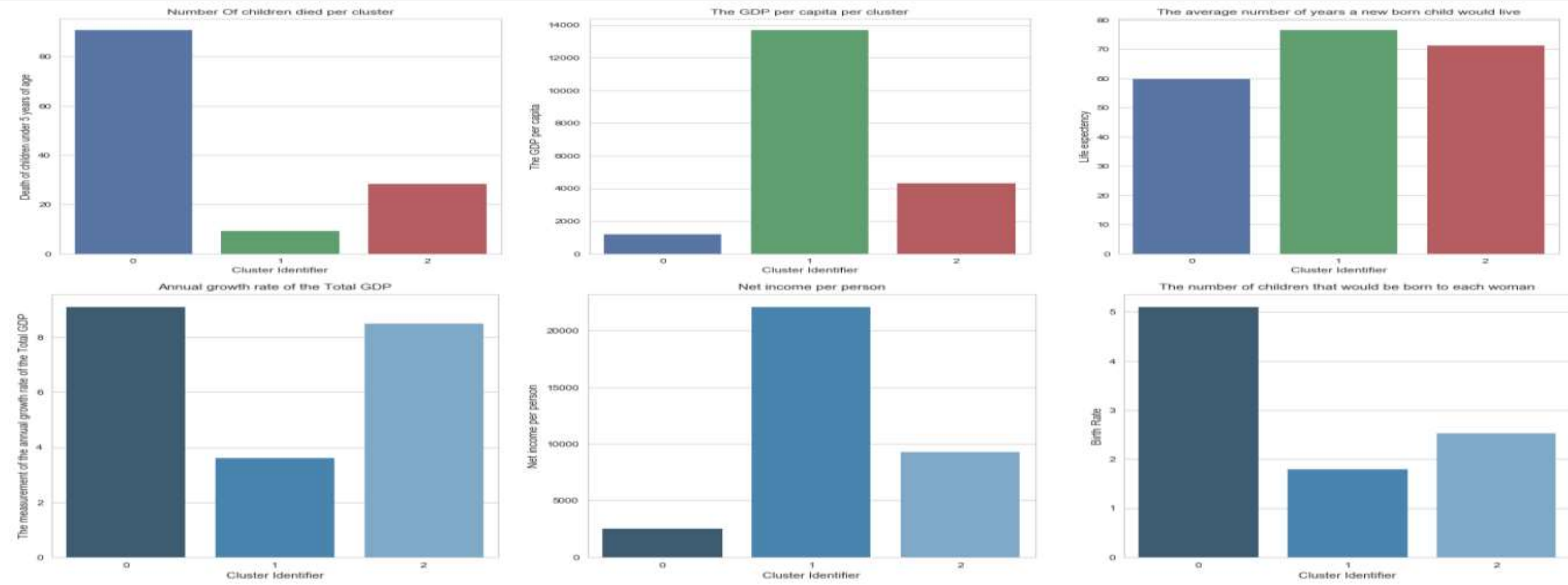
In the next step we have created silhouette score graph and inertia graph to find out the number of cluster present in the data set for current dataset I have identified a total number of three clusters, After identifying the clusters I have created a scatter plot to check how the clusters are differentiated .

Next we have used same dataset to perform a hierarchical clustering where I have use single and complete linkage to draw the dendrogram where I have noticed that the plot where I have used complete linkage is properly arranged.

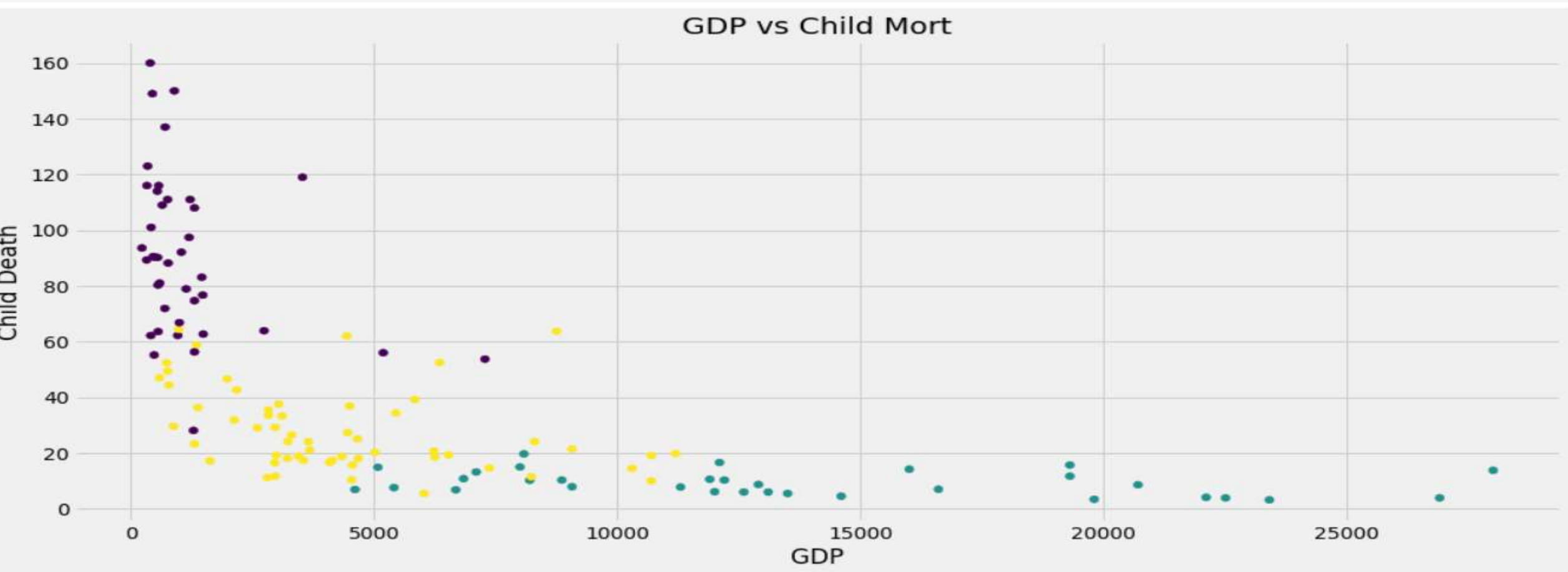


Mean analysis for clusters

From the below graphs we can easily identify that the cluster 0 is our area of interest as the children death rate is high, GDP is low , Net income is low and Birth rate is high(lack of education)

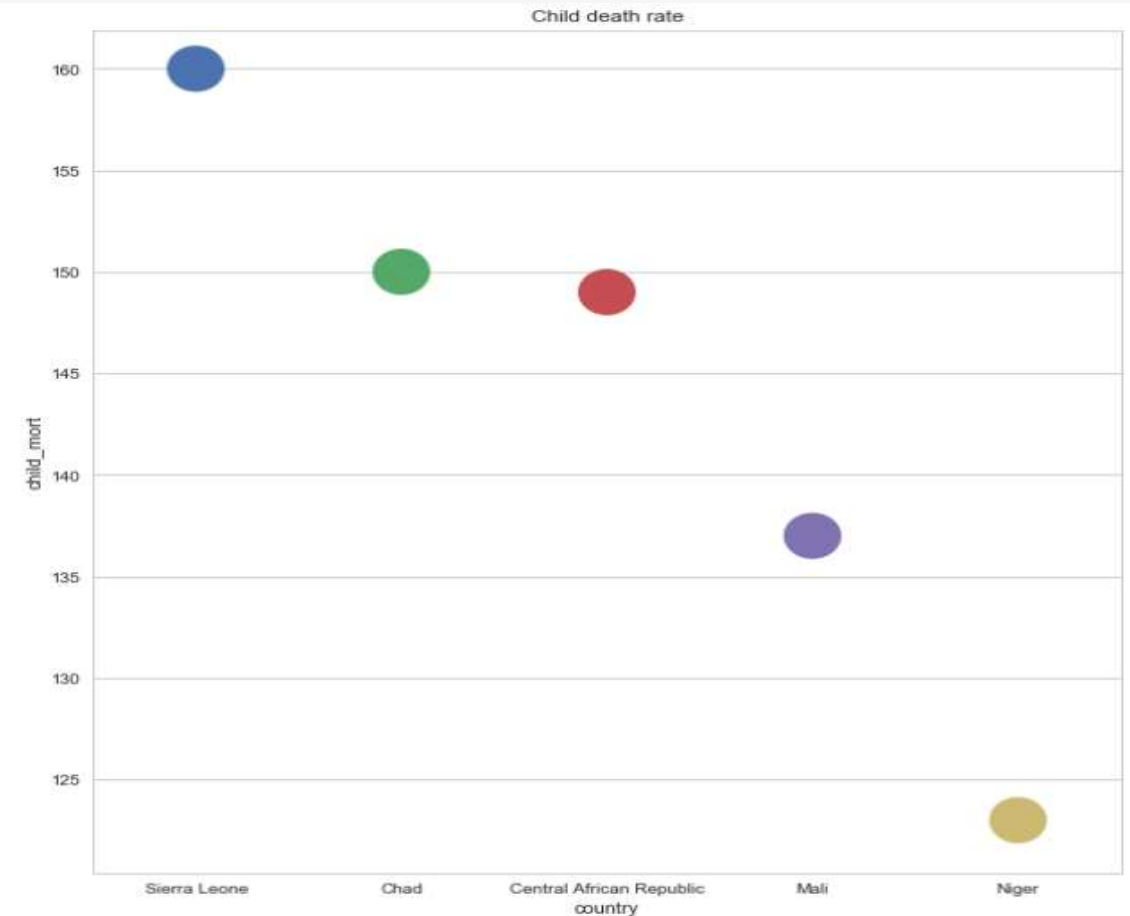
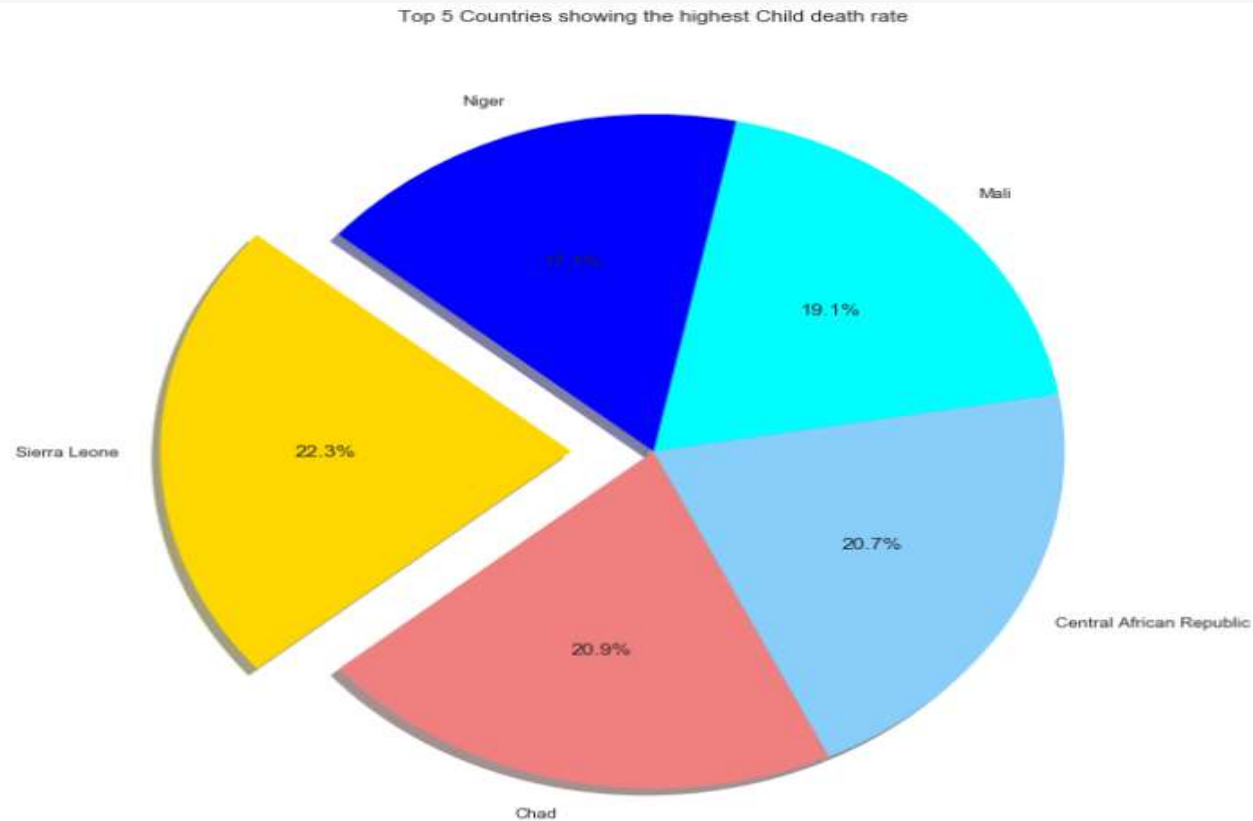


We can see the graph GDP vs Child Mort where you can easily identify the different clusters also we can conclude that the lower GDP countries are having a high death rate for children's



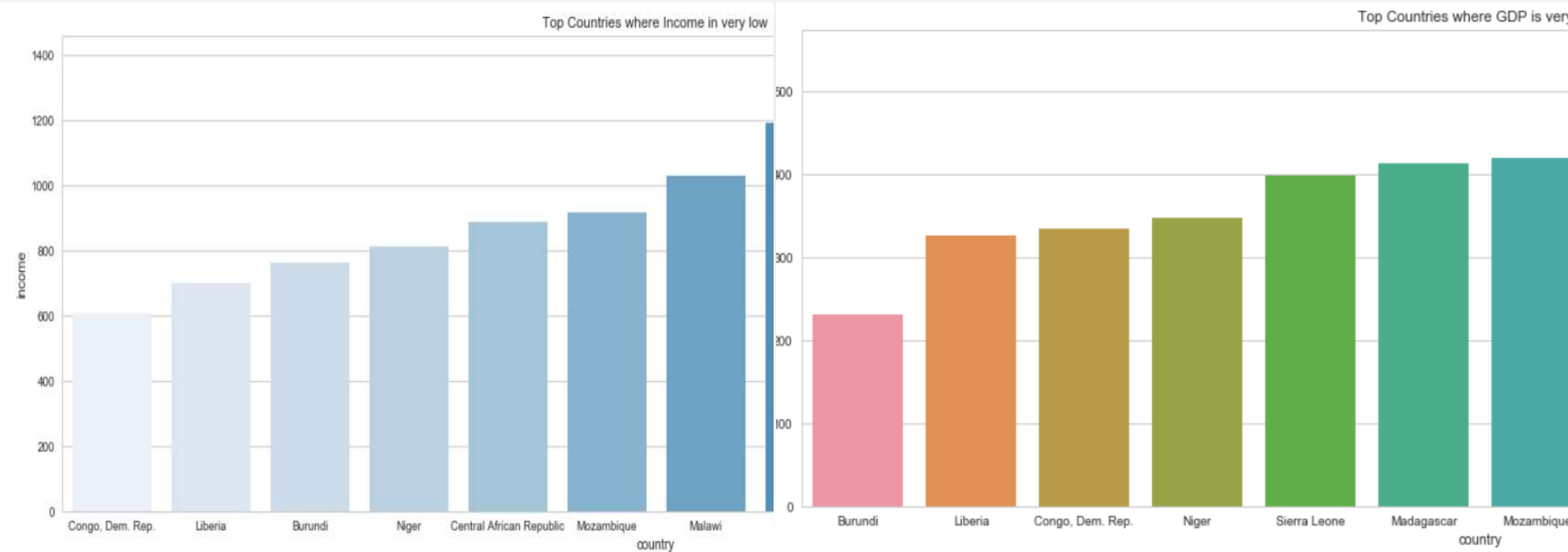
Top 5 countries wise child mort

Below are the top countries where the death of children's are very high



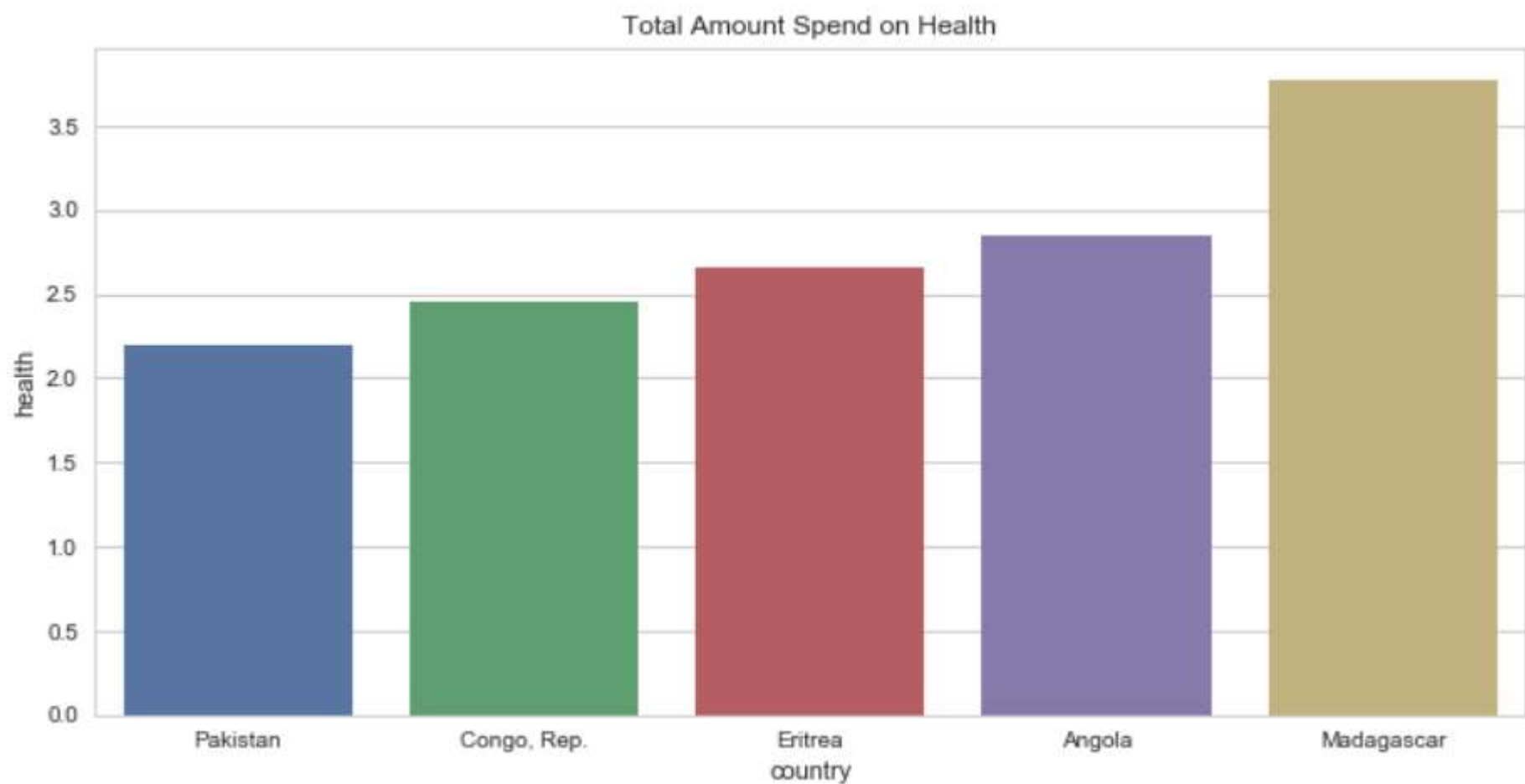
Income Rate and GDP

Below are the countries where income and GDP is very low



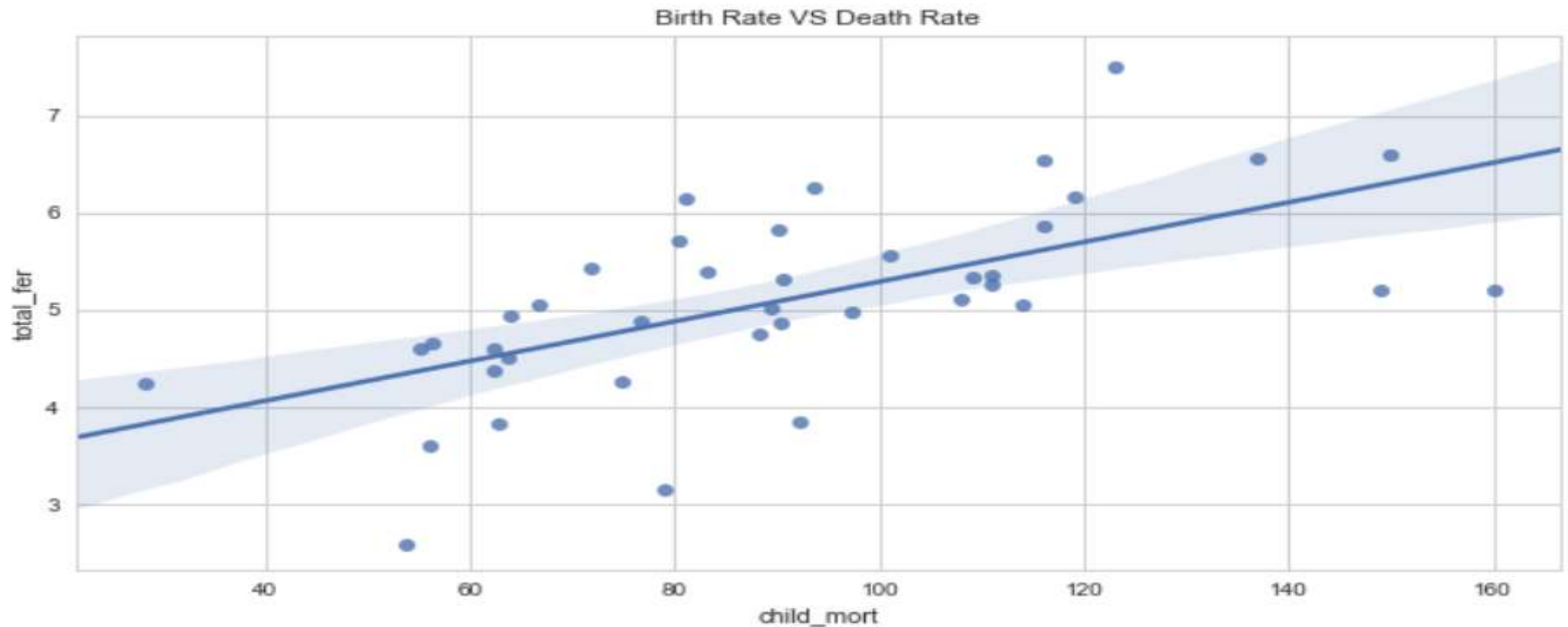
Health wise total spend

Health wise total spending by countries



Linear relation between birth and death rate

We can see with the increase of birth rate death rate also increases there is a linear relationship



On analyzing all the graphs we can find the most repetitive names which are given below. These are the countries which needs help

Hence Countries which are in need are

1 . Sierra Leone 2 . Niger 3 . Congo, Dem. Rep 4 . Chad 5 . Central African Republic

