HELP NGO CASE STUDY

By
NAME: Shivaprasad A
ROLL NO: DDS1910473

INTRODUCTION

Problem Statement :

- 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 project that included a lot of awareness drives and funding programs, 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.
- And this is where you come in as a data analyst. Your job is to categorize 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.

Objective :

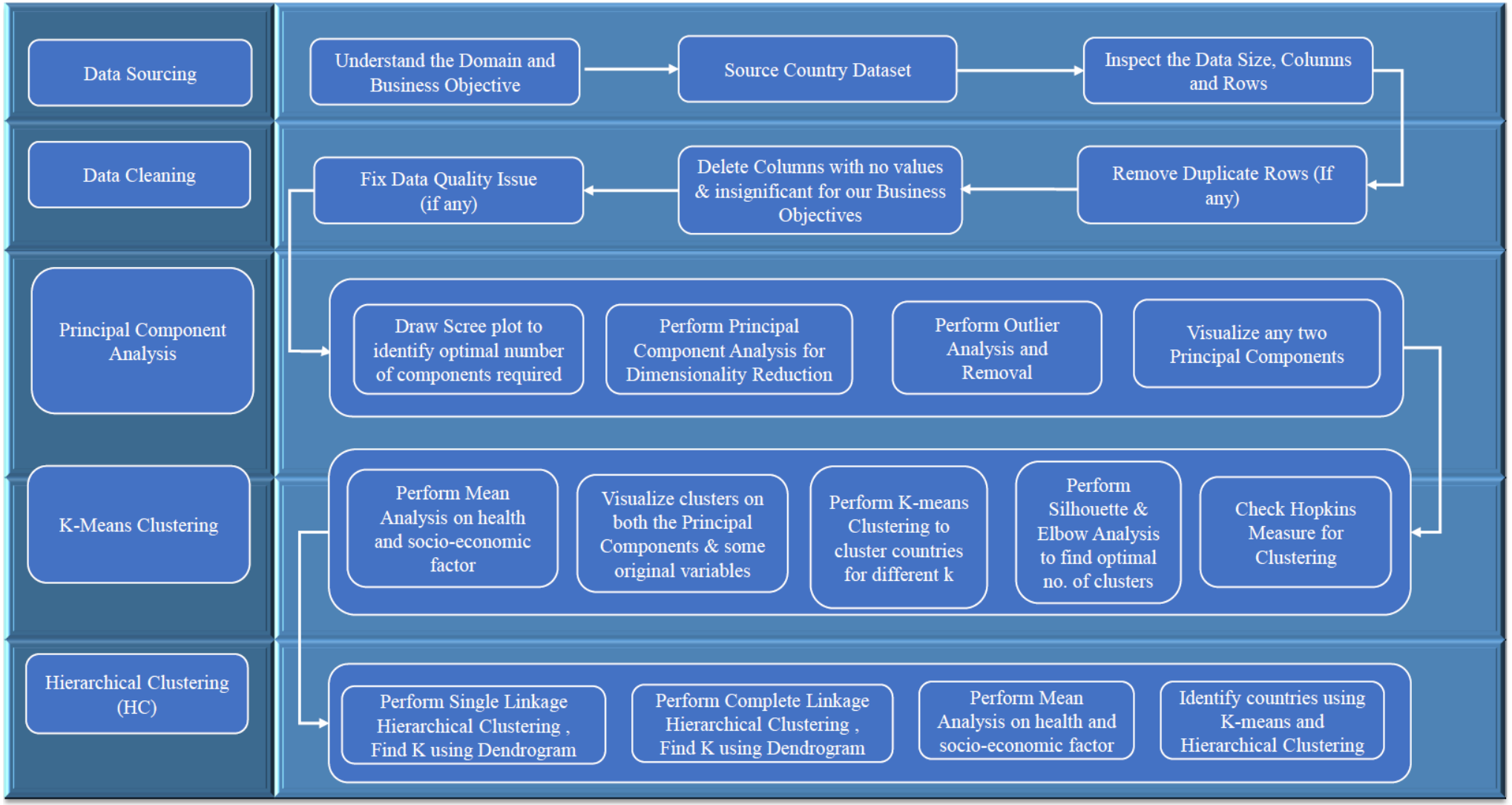
- The main objective is to form clusters of the given countries and analyze based on socio-economic factors to present the recommendations to the CEO so that decision can be taken for which countries that needs immediate attention.

DATA UNDERSTANDING

The dataset comprises of 167 rows and 10 columns. Below are the information for each columns :

- **country** : Name of the country
- **child_mort** : Death of children under 5 years of age per 1000 live births
- **exports** : Exports of goods and services. Given as %age of the Total GDP
- **health** : Total health spending as %age of Total GDP
- **imports** : Imports of goods and services. Given as %age of the Total GDP
- **income** : Net income per person
- **inflation** : The measurement of the annual growth rate of the Total GDP
- **life_expec** : The average number of years a new born child would live if the current mortality patterns are to remain the same
- **total_fer** : 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.

PROBLEM ANALYSIS METHODOLOGY



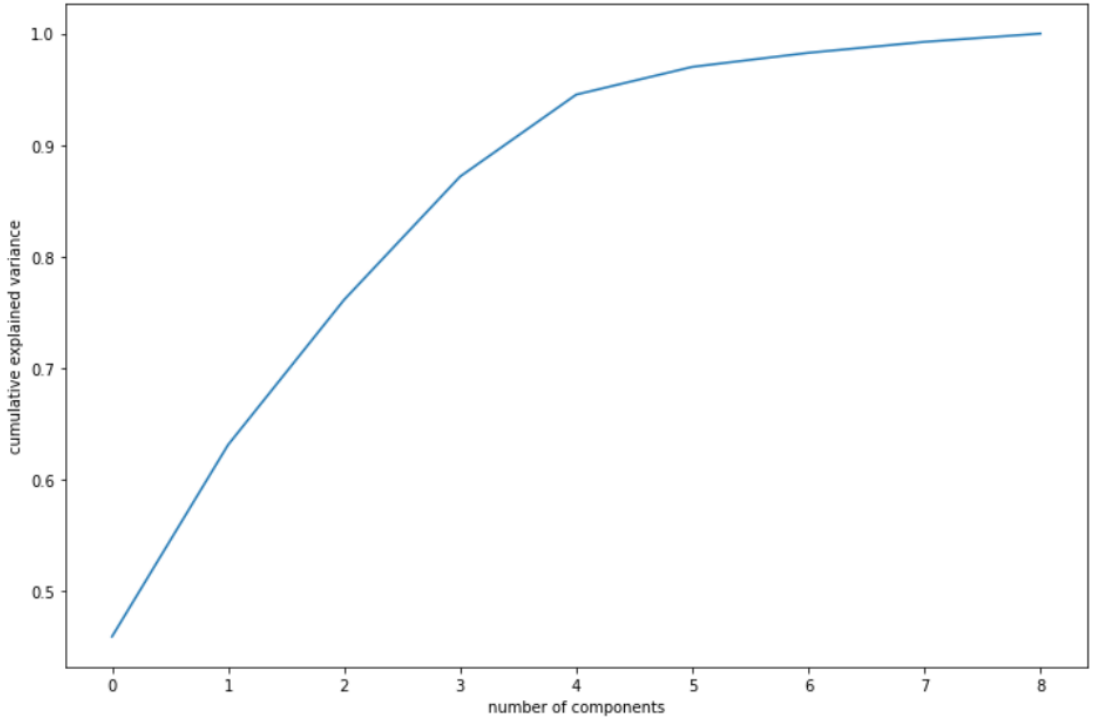
Principal Component Analysis

Principal Component Analysis (PCA)

- Technique used for identification of a smaller number of uncorrelated variables known as **principal components** from a larger set of data.
- It is used to emphasize variation and capture strong
- patterns in a data set
- It is a Dimensionality Reduction method.

Scree Plot

- A Scree Plot is a simple line segment plot that shows the fraction of total variance in the data as explained or represented by each PC.
- The PCs are ordered, and by definition are therefore assigned a number label, by decreasing order of contribution to total variance.
- Used to identify optimal number of Principal Components.

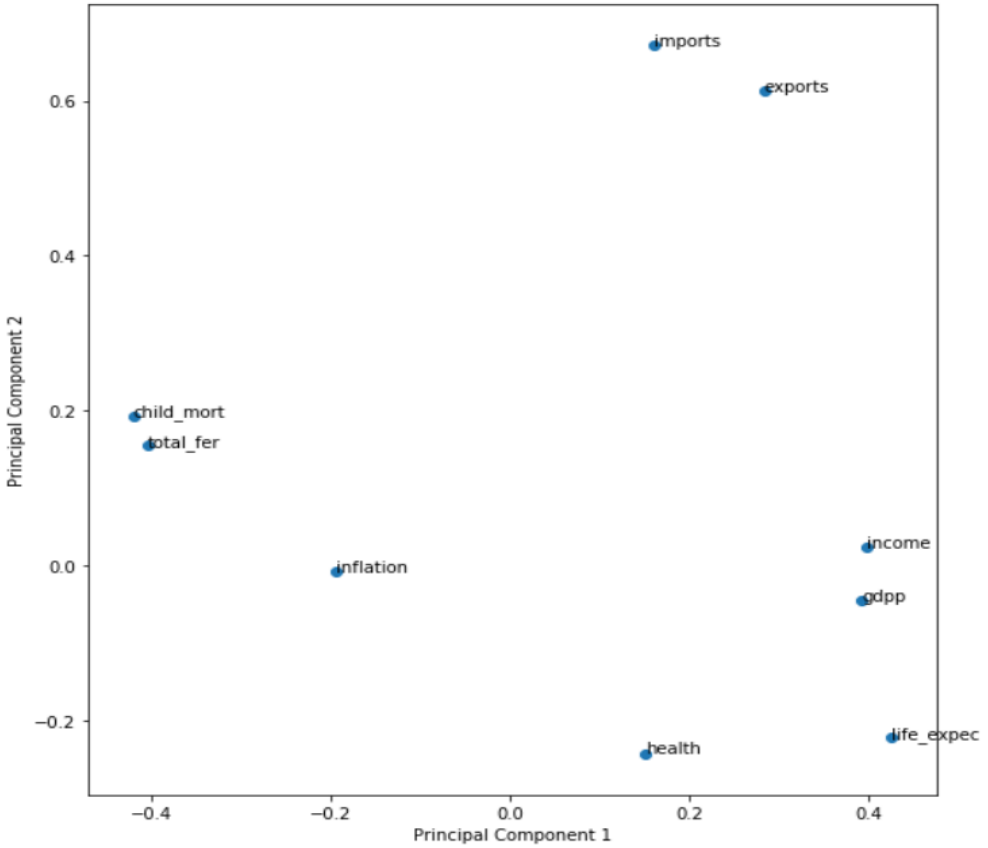


Scree Plot

- From the Scree Plot we can see 5 Principal Component can explain more than 95% of variance.
- So chosen no. of Principal Component = 5

PCA with No. of Principal Component = 5

- Using Number of Principal Component = 5, Principal Component Analysis is performed.
- We obtained five Principal Component namely PC1, PC2, PC3, PC4 and PC5.
- Visualization is performed for two Principal Component PC1 and PC2 on original variables using scatter plot.
- Scatter plot shows the distribution of different original variables in PC1 and PC2.
- We can also analyze from plot about the original variables clustering tendency.



Scatter Plot of PC1 vs PC2 on original variables

Hopkins Statistics:

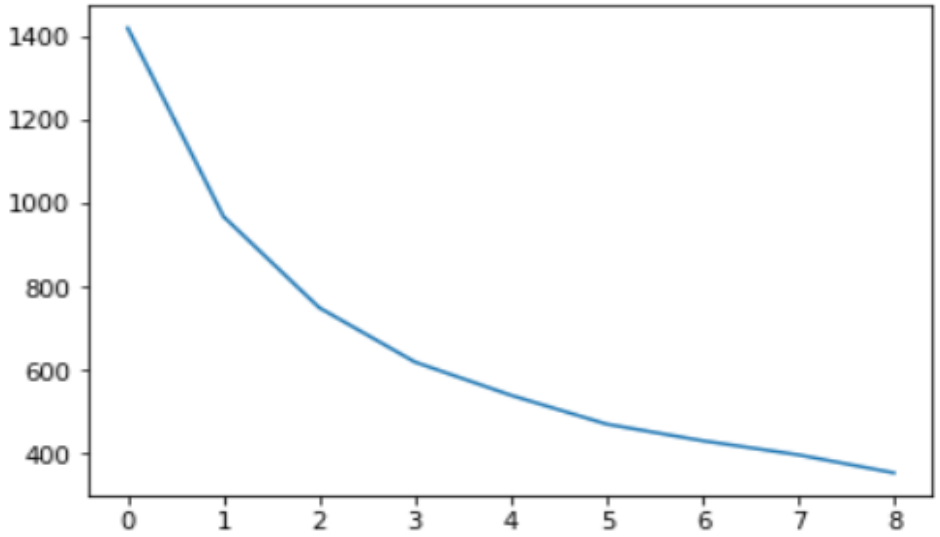
The Hopkins statistic, is a statistic which gives a value which indicates the cluster tendency, in other words: how well the data can be clustered.

- If the value is between {0.01, ...,0.3}, the data is regularly spaced.
- If the value is around 0.5, it is random.
- If the value is between {0.7, ..., 0.99}, it has a high tendency to cluster.

- When we performed this we got Hopkins measure more than **0.8,** that means the data we have has high tendency to cluster.

Finding optimal number of clusters(k)

- Elbow analysis
- The graph looks like elbow of human being
- As the number of cluster increases sum of squared distance decreases



From the Elbow analysis we can identify optimal number of cluster(k) = 3

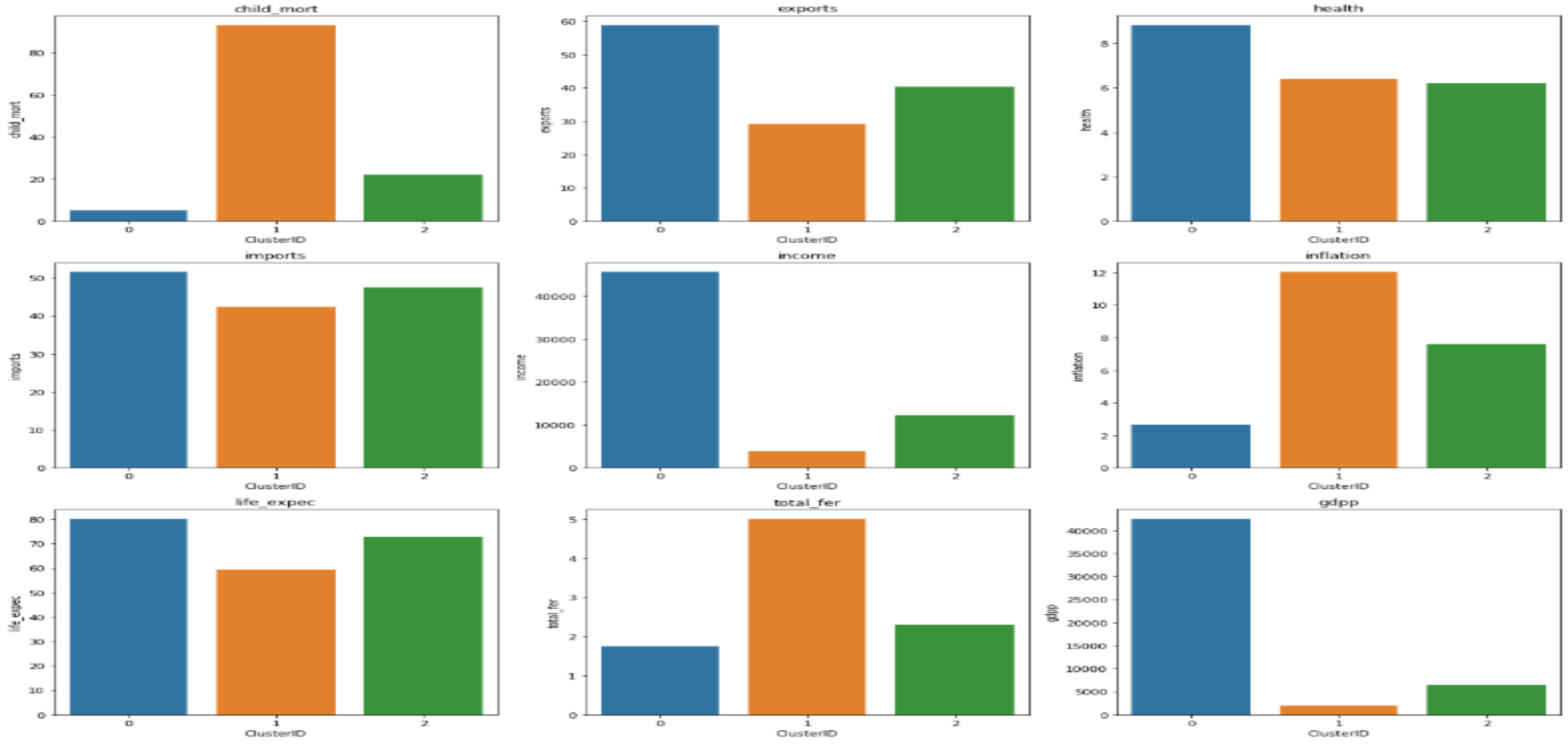
K – Means clustering

K-Means Clustering:

- When we performed clustering with k=3, we found 36 countries belong to cluster 0(first cluster), 47 countries belong to cluster 1(second cluster) and 84 countries belong to cluster 2 (third cluster)

- Let us perform the scio-economic analysis to get the cluster of countries which requires Aide from HELP

Socio-economic Analysis of the Clusters from K-means Method:



- From the above plots of individual variables.
- Its very clear that the countries falling in cluster 1 are the one in need of Fund from the NGO for progress

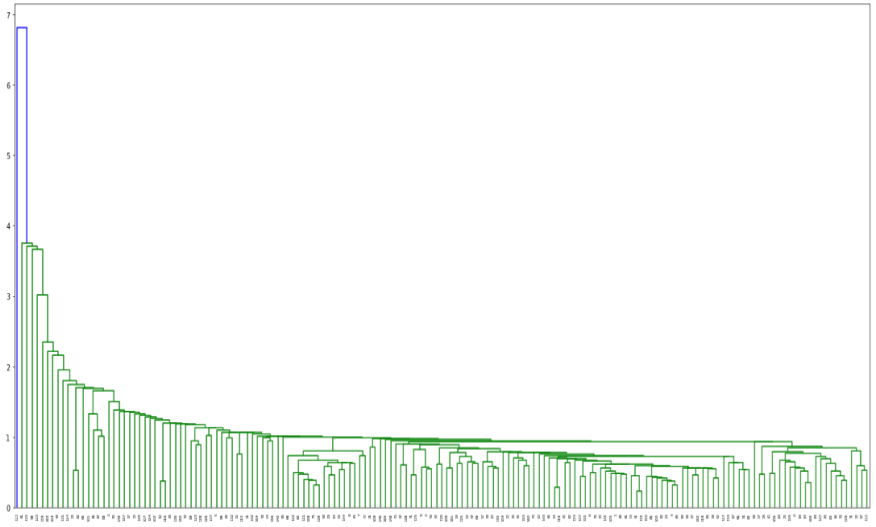
TOP 5 COUNTRIES IN AIDE FROM HELP

ClusterID	country	child_mort	gdpp	income	inflation
1	Burundi	93.6	231	764	12.30
1	Liberia	89.3	327	700	5.47
1	Congo, Dem. Rep.	116.0	334	609	20.80
1	Niger	123.0	348	814	2.55
1	Sierra Leone	160.0	399	1220	17.20

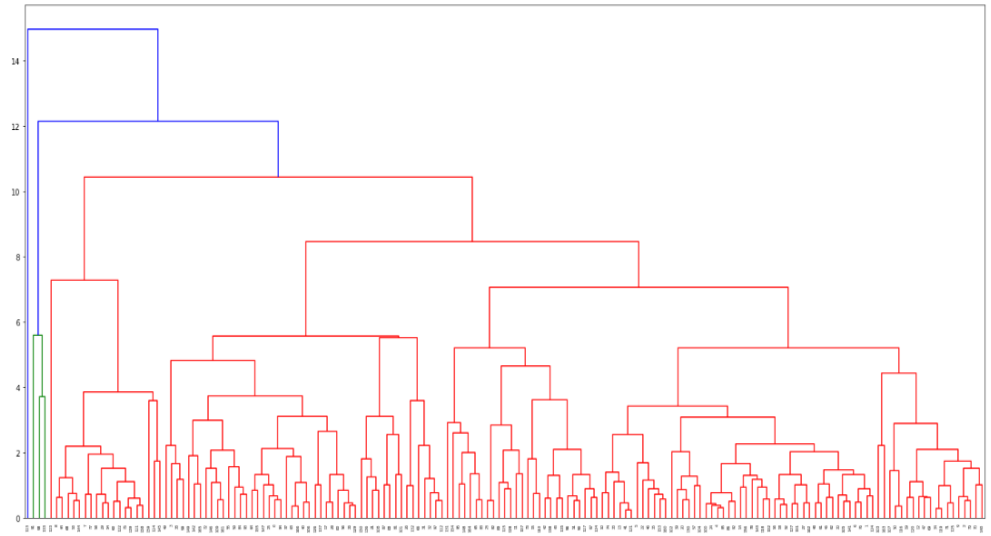
From the above anaylsis and findings the top 5 countries which require fudning from HELP are: Burundi, Liberia, CongoDem.Rep, Niger, Sierra Leone.

HIERARCHICAL CLUSTERING:

- Hierarchical Clustering is performed on the PCA reduced dataset. Both single and complete linkage clustering is performed and cluster_ids are determined. Below are the dendograms obtained from single and complete linkage.



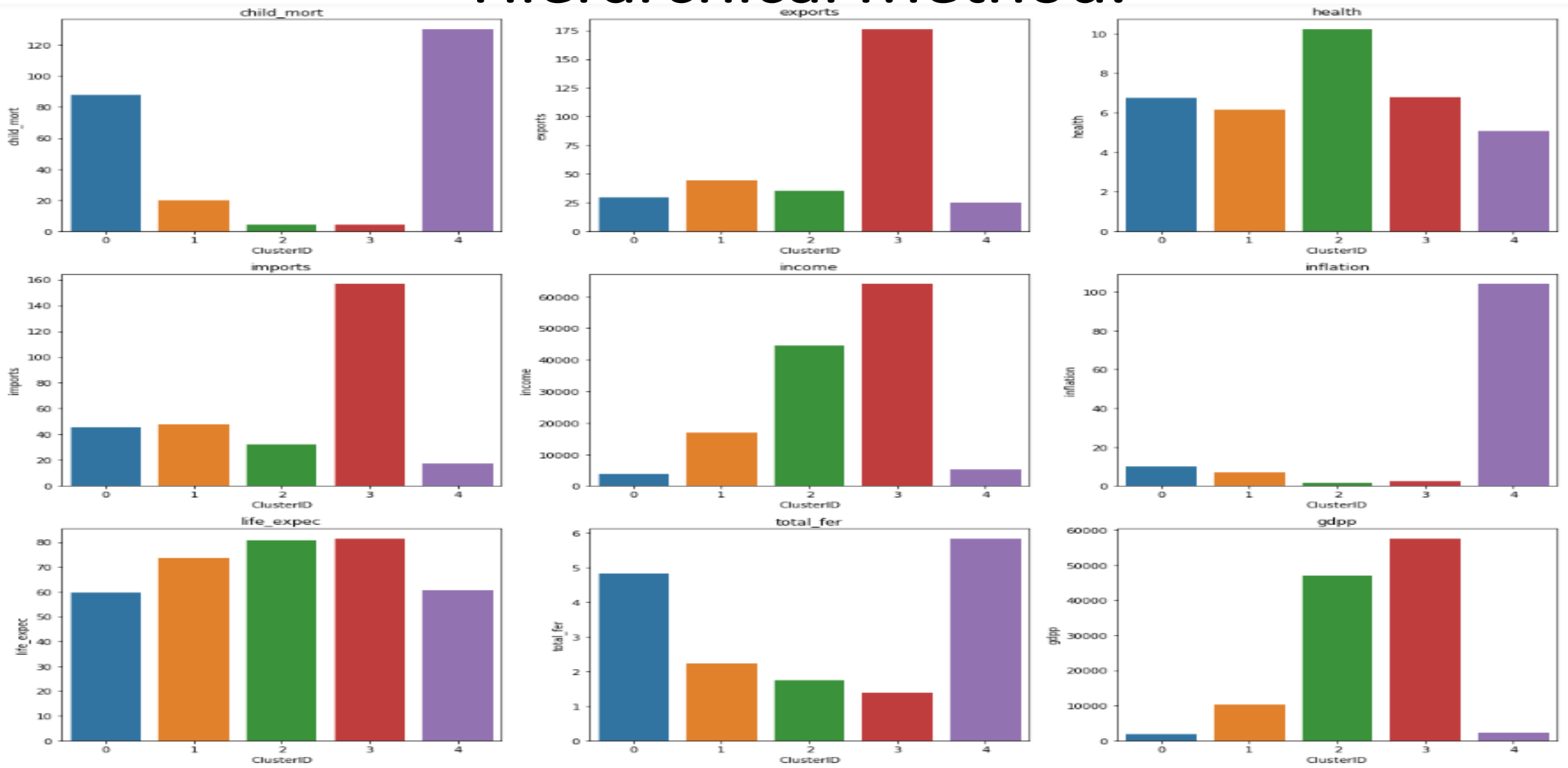
Single Linkage



Complete Linkage

- We will cut the dendrogram at an appropriate level and deciding number of clusters k=5

Socio-economic Analysis of the Clusters from Hierarchical Method:



- From the above plots of individual variables.
- Its very clear that the countries falling in cluster 0 are the one in need of Fund from the NGO for progress

TOP 5 COUNTRIES REQUIRING FUND FROM HELP

ClusterID	country	child_mort	gdpp	income	inflation
0	Burundi	93.6	231	764	12.30
0	Liberia	89.3	327	700	5.47
0	Congo, Dem. Rep.	116.0	334	609	20.80
0	Niger	123.0	348	814	2.55
0	Sierra Leone	160.0	399	1220	17.20

CONCLUSION

- From the plots obtained after K-Means and Hierarchical Clustering, we can arrive at the following conclusion :
- Countries which are part of cluster-1 in K-means clustering and cluster-0 in Hierarchical clustering are the countries which are having in requirement of fund from HELP.
- Because very high child mortality rate and least gdpp and income. Their child fertility rate is also high which means that the population tends to increase and this combined with the poor gdpp and income, imports/exports reflects that these countries are in high need of aide and support.
- The top 5 countries that fall under this cluster are third world countries like **Burundi, Liberia, Congo Dem. Rep., Niger, Sierra Leone**
- NGO Help should direct their funds for the social betterment of these countries.