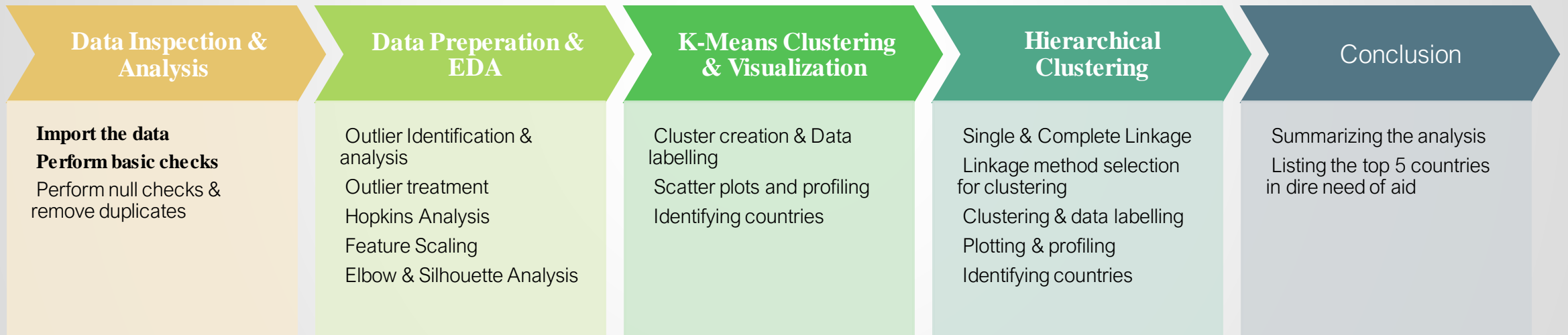# CLUSTERING ASSIGNMENT

BY:

NIRBHAY TANDON

PROBLEM STATEMENT & BUSINESS OBJECTIVE

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. After the recent funding programmes, 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.

Business Objective: Your job is to categorise 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.

METHODOLOGY

Data Inspection & Analysis	Data Preperation & EDA	K-Means Clustering & Visualization	Hierarchical Clustering	Conclusion
Import the data **Perform basic checks** Perform null checks & remove duplicates	Outlier Identification & analysis Outlier treatment Hopkins Analysis Feature Scaling Elbow & Silhouette Analysis	Cluster creation & Data labelling Scatter plots and profiling Identifying countries	Single & Complete Linkage Linkage method selection for clustering Clustering & data labelling Plotting & profiling Identifying countries	Summarizing the analysis Listing the top 5 countries in dire need of aid

EXPLORATORY DATA ANALYSIS

Heatmap of Data before(L) & After(R) Outlier Treatment

Analysis

Image: Heatmap Before Outlier Treatment

Image: Heatmap AfterOutlier Treatment
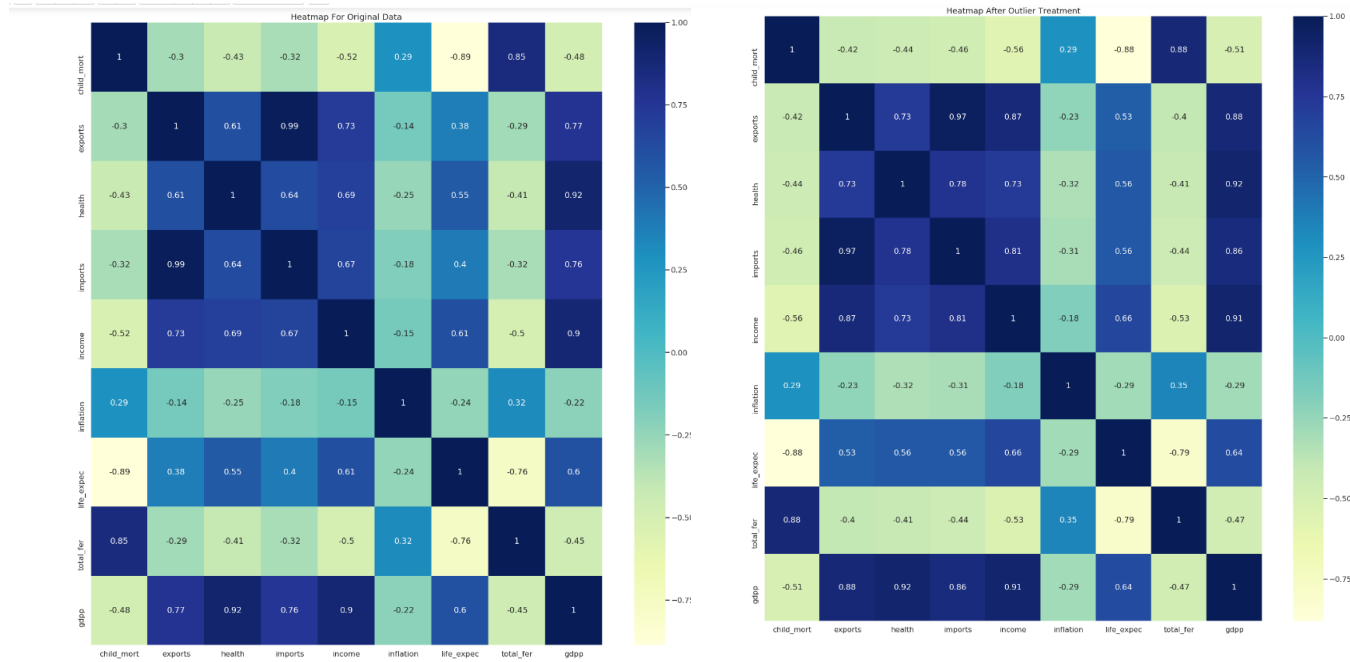
- As we can clearly see from the heatmaps, the removal of outliers has changed the way data correlates in certain features.

- There is increased correlation in GDPP-Imports, GDPP-Exports etc.

EXPLORATORY DATA ANALYSIS CONTD...

Silhouette Analysis

- From the Silhouette analysis we can see that there is a significant change at 3.

- There is also one at 5, however, the value increases in the next iteration at 6, so we shall discard that value.
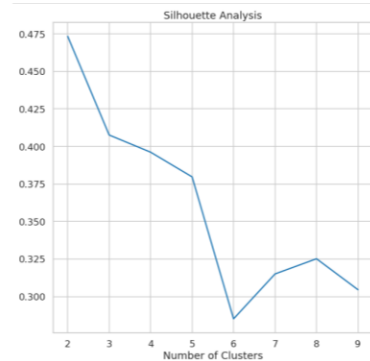
Image: Silhouette Analysis

Image: Elbow Analysis

Elbow Analysis
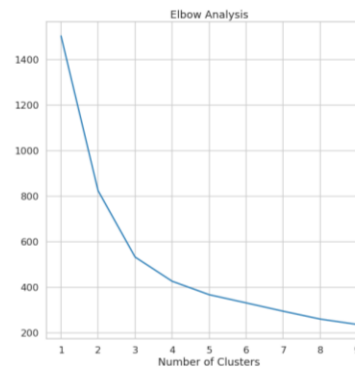
- Similarly from the Elbow Analysis we see that the value 3 is significant

K-MEANS CLUSTER VISUALIZATIONS

Inferences Drawn From Adjacent Plots

Based on the scatter plots adjacent we can see that:

- Cluster 2 contains countries with low income and gdpp but a very high child mortality rate

- Cluster 1 has countries with the highest income and gdpp but a vert low child mortality rate

- We should primarily focus on countries in cluster 2 which can identify countries in dire need

- Some countries in cluster 0 might also need help but our primary concern should focus on cluster 2 countries
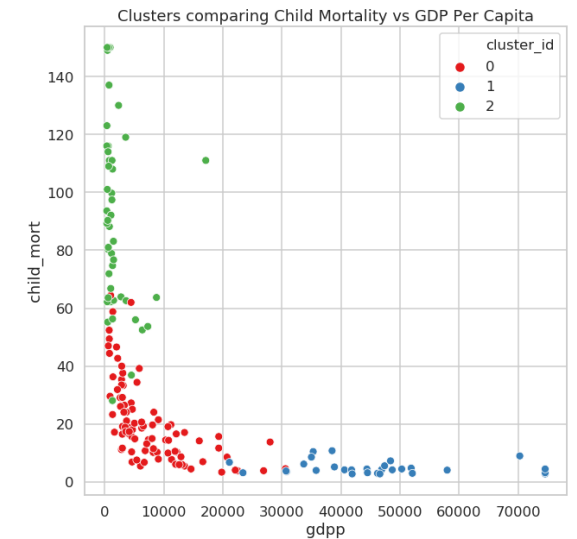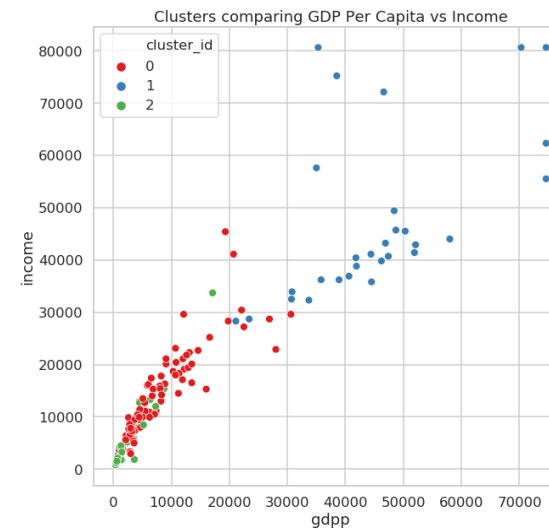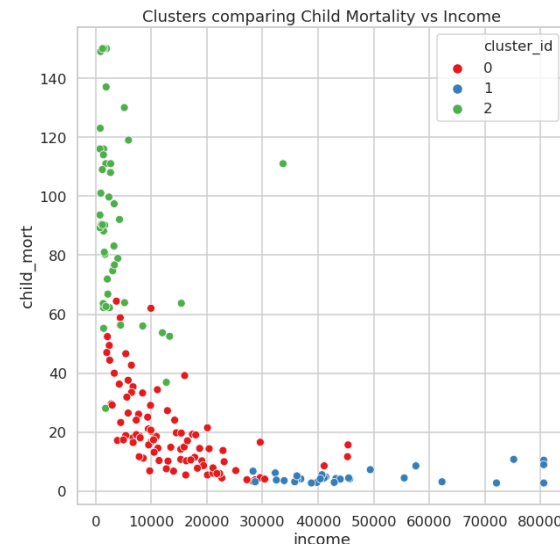
HIERARCHICAL CLUSTERING

Inferences Drawn From Adjacent Plots

- From the Complete Linkage Hierarchical Clustering we can see that we have 2 visible clusters that we can explore

- Upon exploring 2 clusters in complete linkage we obtained the adjacent plots

- From the clusters adjacent we can see that cluster 0 has the lowest income & gdpp but the highest child mortality.

- This behavior is similar to cluster 2 from K-means.
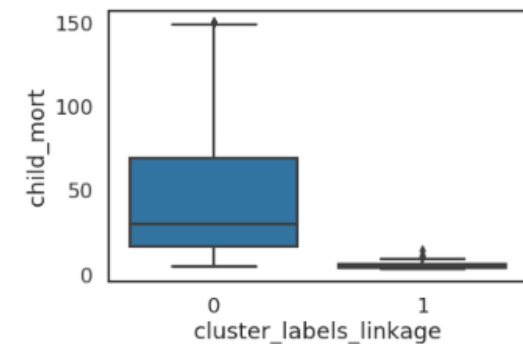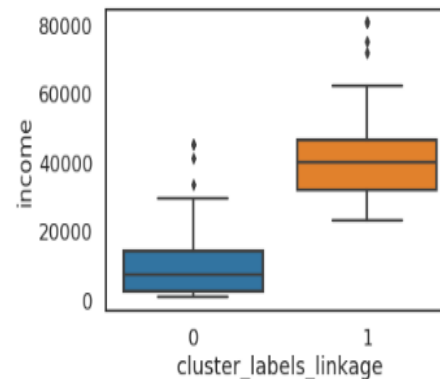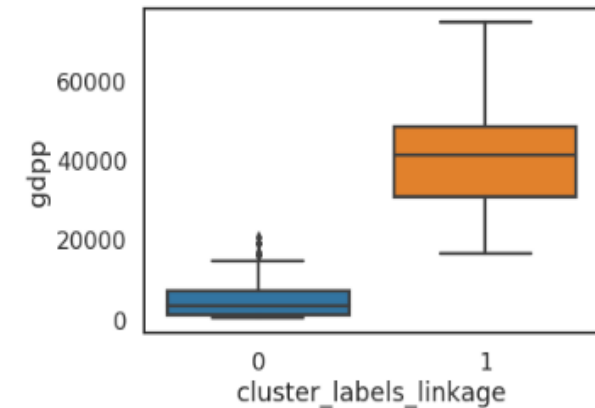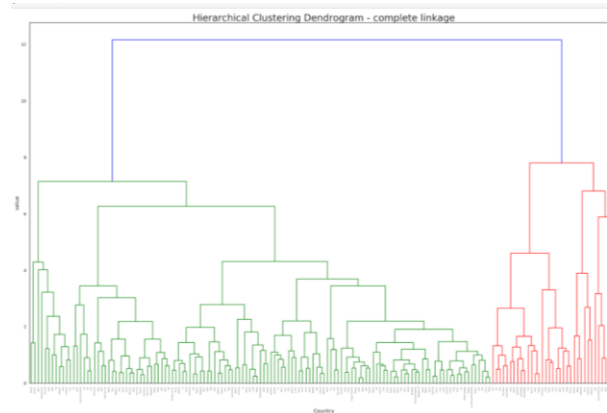
CONCLUSION

Based on the analysis we have carried out, we can conclusively say that the HELP International NGO should focus on the 5 countries outlined below for immediate aid:

- Burundi
- Liberia
- Congo, Dem. Rep.
- Niger
- Sierra Leone

- From the analysis we have carried out from the provided countries data we can conclusively say that based on the following factors:

- The countries in cluster 2(K-Means) & cluster 0(Hierarchical) have the lowest GDP Per Capita(GDPP)

- These countries have the lowest income as well as the highest child mortality rate

- The fertility rate for these countries is very high but due to low incomes the child mortality rate is also high, which leads to the assumption that the children are born but are unable to be fed due to low family incomes.