Problem Statement

We need to find out the countries that are in dire need of help to overcome the socio economic problems and since the funds of the NGO is limited we need to target the countries which need our help the most .To find that out we need to analyze the numbers of each country and choose wisely and logically based on the stats visible .

Gayathri Madhira

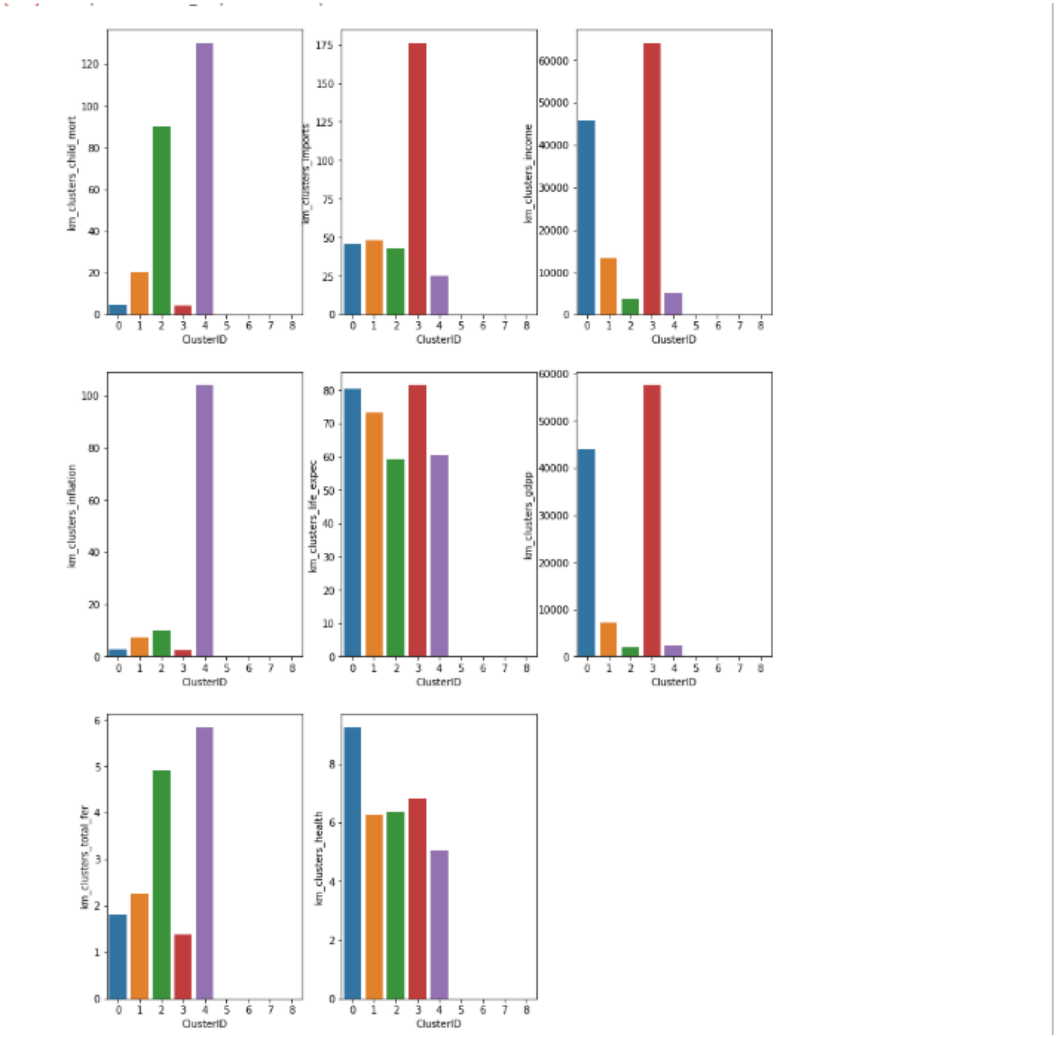
Analysis Approach

- We have used to important techniques here to summarize the data and come to a conclusion .
- We used Principal Component Analysis to reduce the dimensions and make the best combinations of features available in the data .
- We used 2 clustering methods K-Means and Hierarchical clustering to cluster the data .
- We then further summarized the data from those clusters to find top 4 or 5 Countries whose situation is poor .
- We did not remove outliers in this data set as we wanted the least performing country .

Results using K- Means Clustering

- After we performed PCA we found that 3 Principal components were enough to describe the data by the help of scree plot .
- We then moved towards clustering using K-Means and found by Hopkins statistics that the data was good to be clustered.
- We analyzed the data with clusters 2 and 3 and found that the most suitable number of clusters will be five .
- After visualizing the data we found that among all the clusters cluster 3 was the worst but even then we cannot afford to help those many countries so narrowed further using the summarized statistical obtained for each cluster and came to the below conclusion
- Country whose child mortality rate is greater than 90 and exports percentage less than 46 and imports percentage greater than 47 and GDP less than 6000 income less than 4000 and inflation less than 10 are the worst and we have following countries which belong in that category

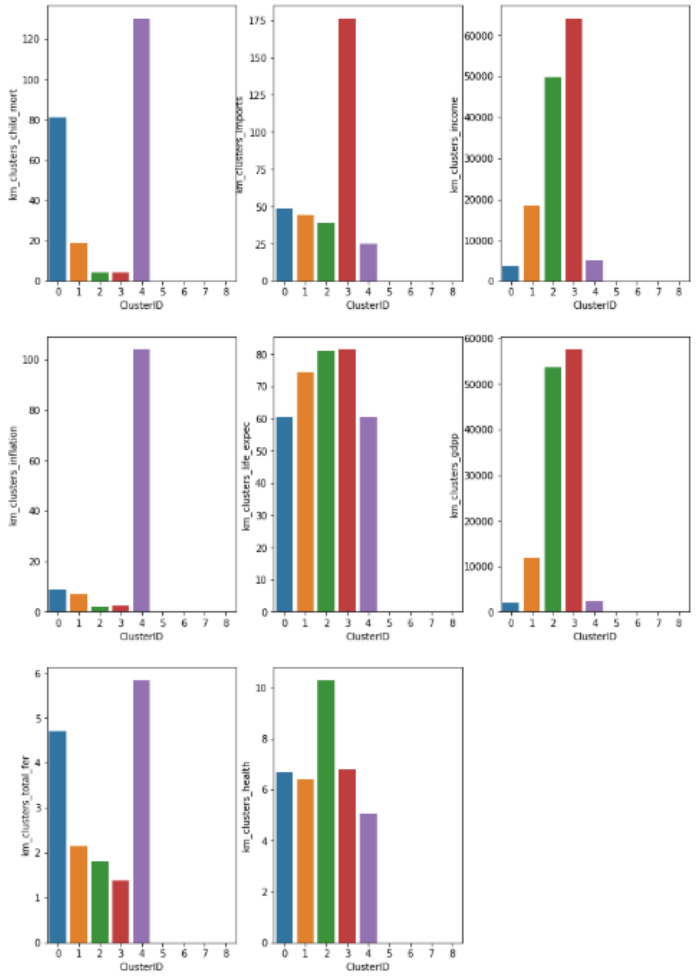
HAITI,LESOTHO,NIGER,TOGO



Results using Hirearchial Clustering

- After having results from K means we moved towards hierarchical clustering
- Since K means had 5 clusters we started with 5 clusters
- We check for 6 clusters as well but cluster 5 suited well
- After visualizing the data we found that among all the clusters cluster 3 was the worst but even then we cannot afford to help those many countries so narrowed further using the summarized statistical obtained for each cluster and came to the below conclusion
- Country whose child mortality rate is greater than 80 and exports percentage less than 42 and imports percentage greater than 47 and GDP less than 3000 income less than 5000 and inflation less than 10 are the worst and we have following countries which belong in that category

HAITI,LESOTHO,NIGER,TOGO, COMOROS, LIBERIA



Conclusion

- After Looking at the results of both K means and Hirearchial clustering it will be our suggestion to focus more on following countries

  HAITI,LESOTHO,NIGER,TOGO