Summary

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 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.

Solution Methodology: As our main objective is to find the countries which are in dire need of funding we will be forming clustering using k-means and hierarchical algorithms and will see the clusters which are having low income, low GDP and high Child mortality and we will identify countries based on clusters which are in need of funding.

Solution: We first performed Exploratory data analysis and found that exports, health, and imports are in percentage of GDP and we need to convert them to their actual values to form better clusters. We also saw few outliers in the data we did upper capping of values above 99 percentiles for exports, health, imports, income, total_fer and GDP columns. We also performed univariate and bivariate analysis to see the relation among variables in the data.

After that we checked hopkins score to see if there is any cluster tendency in the given data as we are getting score above 0.85 which is good we continued with forming clusters using KMeans algorithm for finding optimal k value we used elbow curve and silhouette score methods and got to know that k = 3 and 5 are best values. Performed KMeans for both k = 3 and k = 5 and got 'Burundi', 'Liberia', 'Congo, Dem. Rep.', 'Niger', 'Sierra Leone' these countries with least GDP, least Income, highest child mort rate for both k = 3 and k = 5.

Then we started with Hierarchical clustering initially we tried to used single linkage method since this is difficult to intuitively say the number of clusters need to be formed we used complete linkage method and formed Dendrogram using cut tree function we got classes for num_clusters = 3 and then found the cluster which is having low GDP, low Income and high Child_mort. We got Congo,Dem. Rep., Liberia, Burundi, Niger and Central African Republic these countries with low income, GDP and high child mort .

Finally, from both clustering methods and countries we got with both cluster we conclude that

- Burundi
- Liberia
- Congo,Dem. Rep.
- Niger
- Sierra Leone
- Central African Republic

These are the countries that are in the direst need of aid.