Files:

Dataset: 'Country-data.csv' contains various socio-economic factors of 167 countries. A 'Data Dictionary.xlsx' is provided along with to help understand what those variables in the dataset are.

Analysis: It is carried out in two .ipynb files. First part of analysis is in '1.Analysis - HELP, Clustering of Countries.ipynb'. Second part of analysis is in '2.Analysis - HELP, Clustering of Countries with ClusterID 0.ipynb'.

Presentation: Results and findings are provided in 'Presentation.pdf'.

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 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.

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. The datasets containing those socio-economic factors and the corresponding data dictionary are provided.

Objectives:

Your main task is to cluster the countries by the factors mentioned above and then present your solution and recommendations to the CEO using a PPT. You are also supposed to use dimensionality reduction using PCA to get the visualisations of the clusters in a 2-D form.