



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



Countries' Data



Download



Data Dictionary



Download

## 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.

## Results Expected


1. A well-commented Jupyter notebook containing the Clustering Models(both K-means and Hierarchical Clustering) and the final clusters of countries. The clusters must be visualised on both the Principal Components and some of the original variables. Basically, you need to perform PCA on the dataset in order to obtain the Principal Components, perform Clustering on the countries using a suitable number of components and show the clusters on those components as well as a few selected original variables(the socio-economic factors mentioned above).

2. Present the overall approach of the analysis in a presentation

- Mention the problem statement and the analysis approach.
- Explain the results of Principal Component Analysis and Clustering briefly.
- Include visualisations and summarise the most important results in the presentation.

You need to submit the following two components

- **Python notebook:** Should include detailed comments and should not contain unnecessary pieces of code
- **PPT:** Make a PPT to present your analysis to the CEO (and thus you should include both the technical and the business aspects). The PPT should be concise, clear, and to the point. **Submit the PPT after converting into the PDF format. The visualisations mentioned above must be present in this file.**

 [Report an error](#)

Next  
Evaluation Rubric →