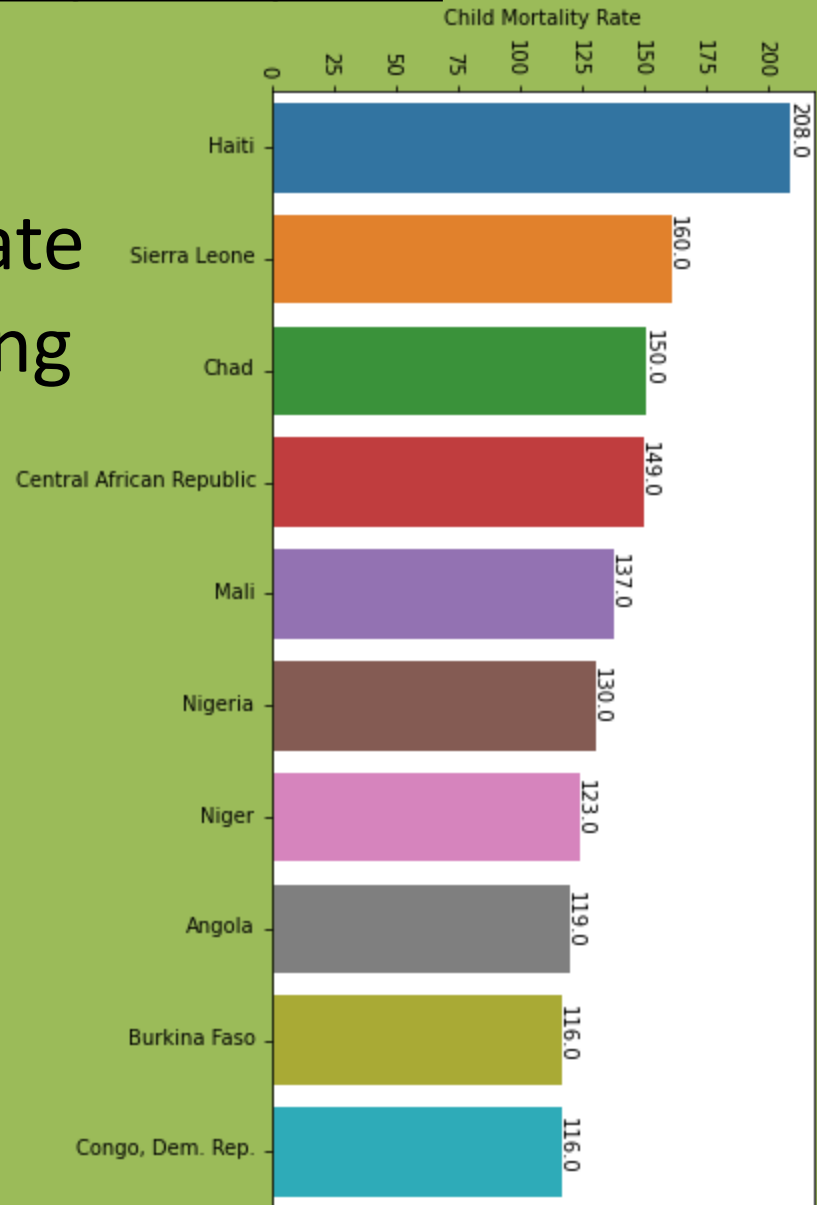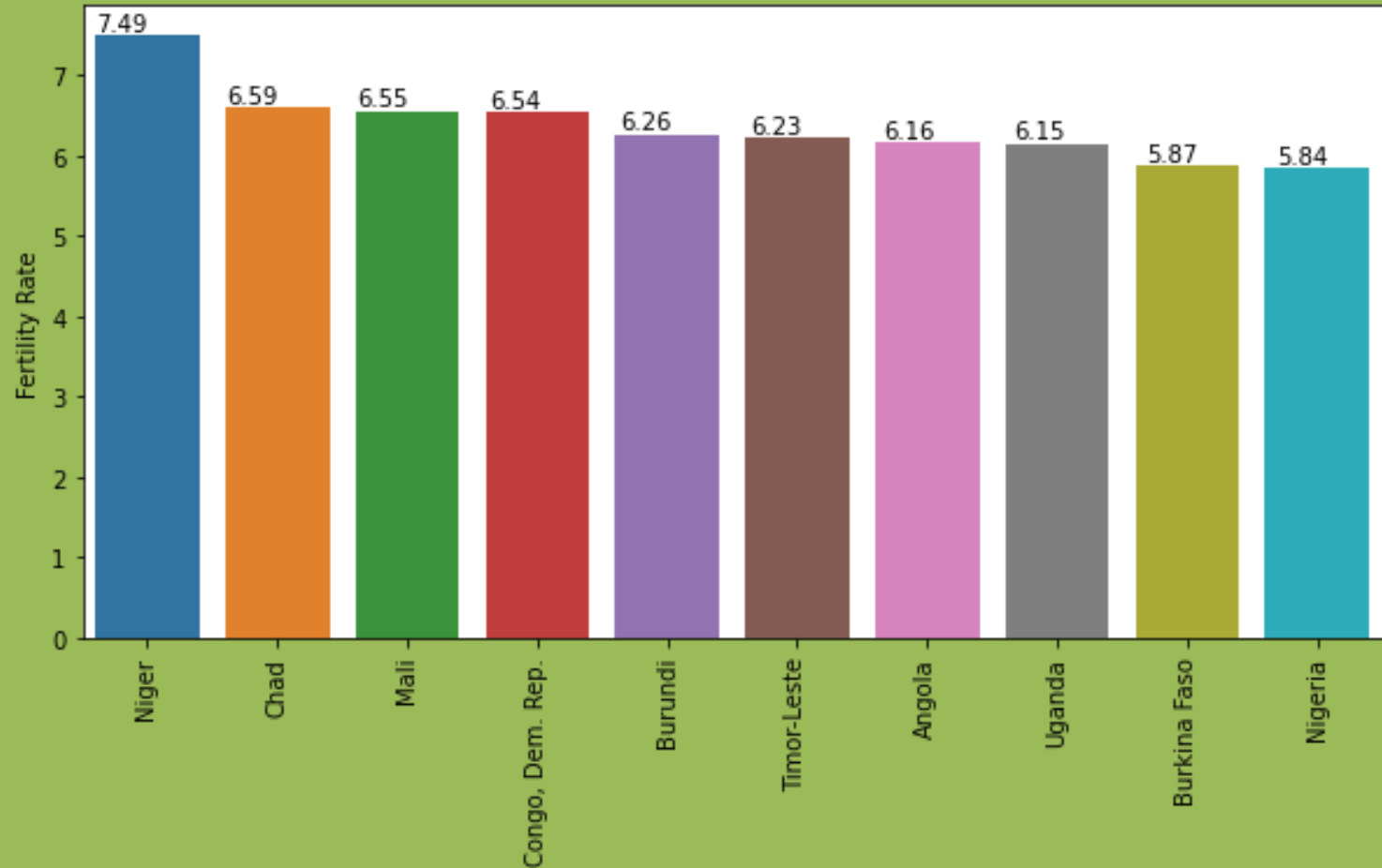# international humanitarian NGO

Prepared by
Pankaj Verma

- Objective :
- The 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 :Data inspection and EDA tasks suitable for this dataset - data cleaning, univariate analysis, bivariate analysis etc.
- **Outlier Analysis:** You must perform the Outlier Analysis on the dataset. However, you do have the flexibility of not removing the outliers if it suits the business needs or a lot of countries are getting removed. Hence, all you need to do is find the outliers in the dataset, and then choose whether to keep them or remove them depending on the results you get.
- Create model using both K-means and Hierarchical clustering(both single and complete linkage) on this dataset to create the clusters.
- Analyse the clusters and identify the ones which are in dire need of aid. You can analyse the clusters by comparing how these three variables - [gdpp, child_mort and income] vary for each cluster of countries to recognise and differentiate the clusters of developed countries from the clusters of under-developed countries.
- Perform visualisations on the clusters that have been formed using the features selected for building the clustering model
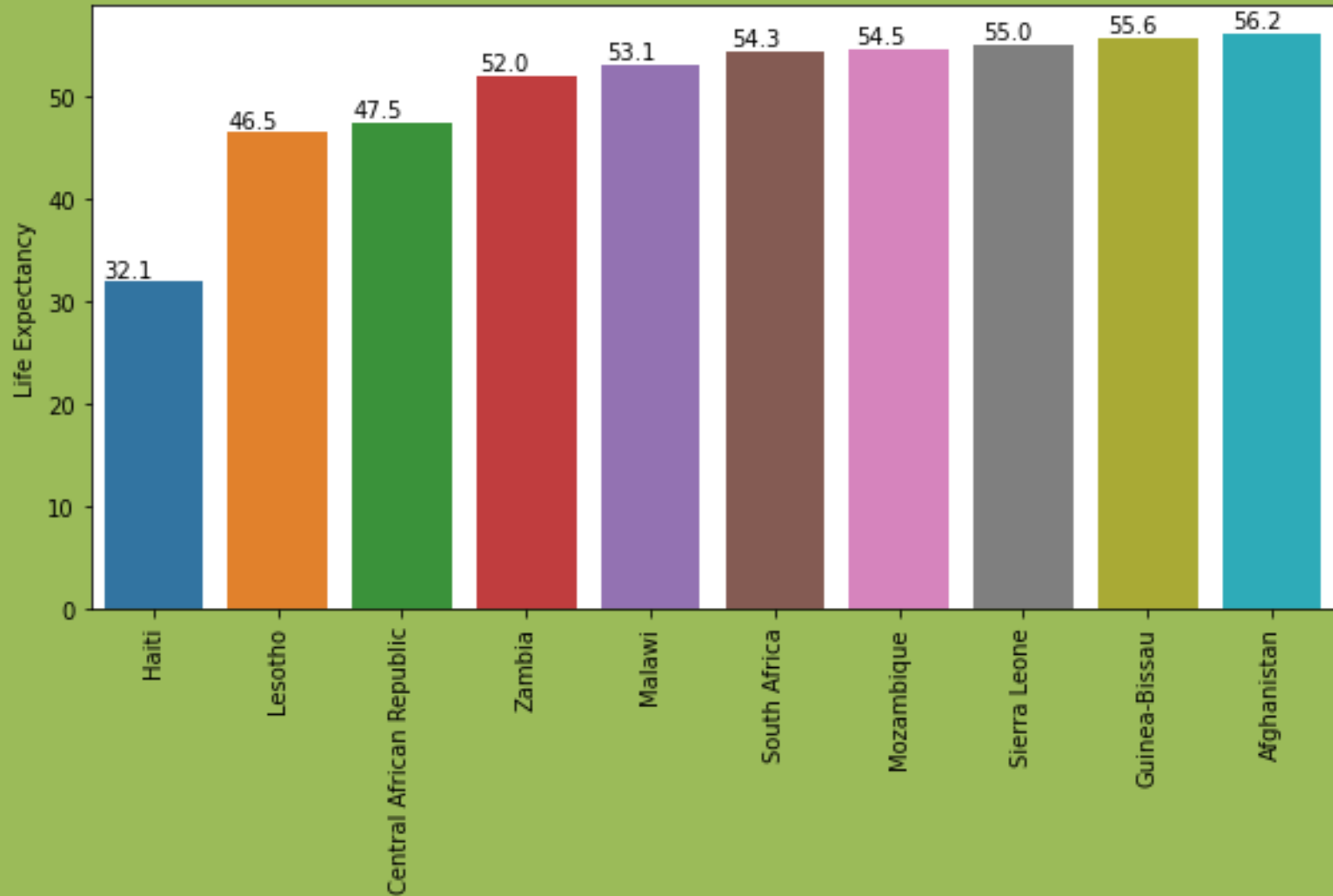-

Child mortality top10

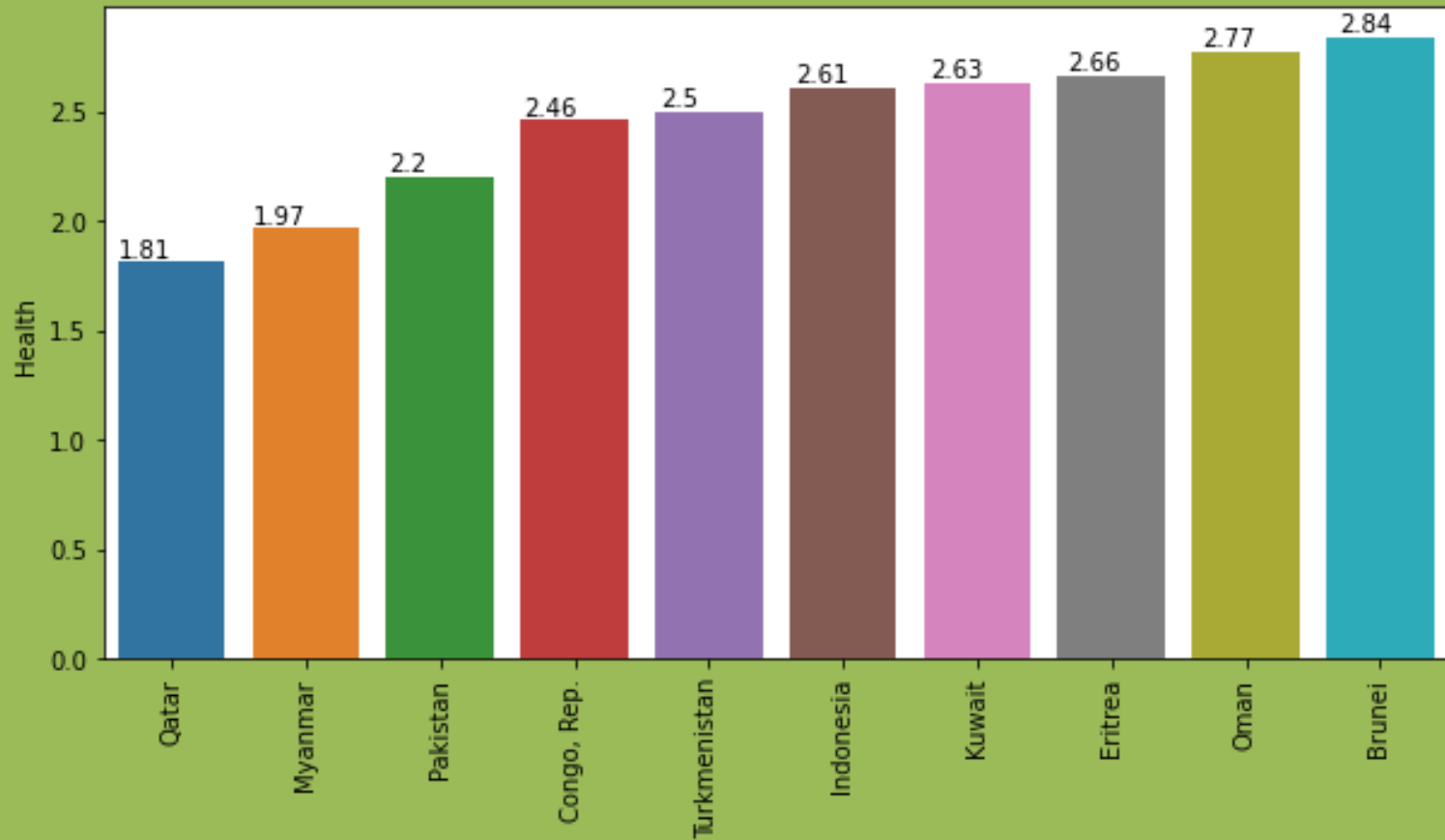- Top 10 Countries having highest Child Mortality Rate are present in **Africa** having poor healthcare facilities.

Child Mortality Rate

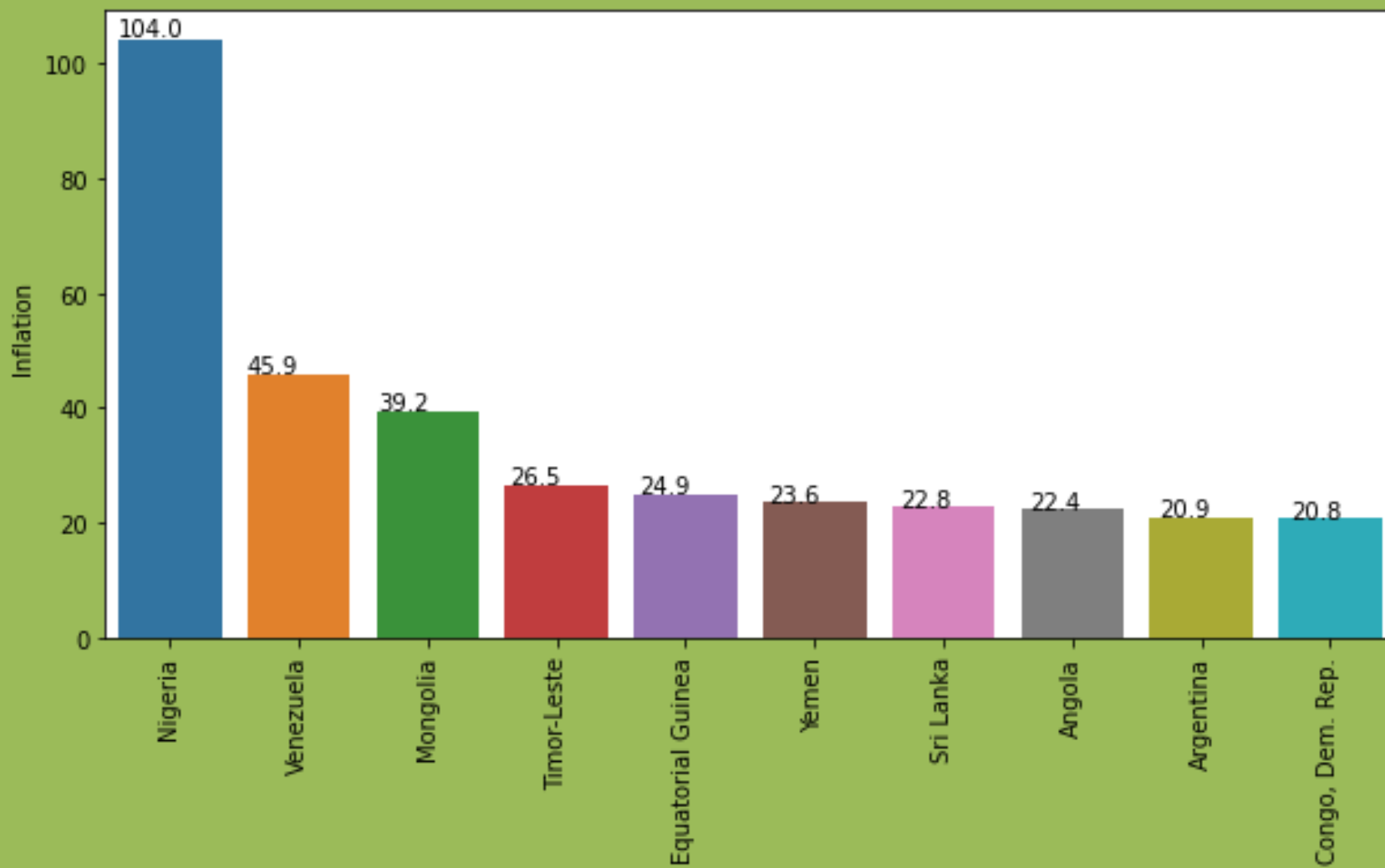| Country	Rate
Haiti	208.0
Sierra Leone	160.0
Chad	150.0
Central African Republic	149.0
Mali	137.0
Nigeria	130.0
Niger	123.0
Angola	119.0
Burkina Faso	116.0
Congo, Dem. Rep.	116.0

Top 10 Fertility Rate

Health

Inflation

- **Outcome:**
  The following 10 countries requires NGO Aid . These falls under Underdeveloped Country list and performing worst among the Under Developed Countries :Burundi
- Liberia
- Congo
- Niger
- Sierra Leone
- Madagascar
- Mozambique
- Central African Republic
- Malawi
- Eritrea
-

  Reasons for Aid:
  High child mortality
- Low Income
- Low GDP
- Low health spent
- High Inflation
- Lower life expectency
- High fertility rate (i.e more number of children per family/woman)

- **Conclusion**
- We have used PCA above to reduce the variables involved and then done the clustering of countries based on those Principal components and then later we identified few factors like child mortality, income etc which plays a vital role in deciding the development status of the country and builded clusters of countries based on that. Based on those clusters we have identified the below list of countries which are in dire need of aid. The list of countries are subject to change as it is based on the few factors like Number of components chosen, Number of Clusters chosen, Clustering method used etc.which we have used to build the model.