

# Clustering and PCA Assignment

Bindhu B

# Problem Statement

- We have a data of 167 countries - rated on socio economic factors like - Child Mortality , exports, health, imports, income, inflation, life-expectancy, total\_fertility, gdpp.
- We have to analyze the data and cluster the countries - based on the given data and suggest the NGO- HELP International - which countries deserve to get their 10 million dollar funding.

Approach used:

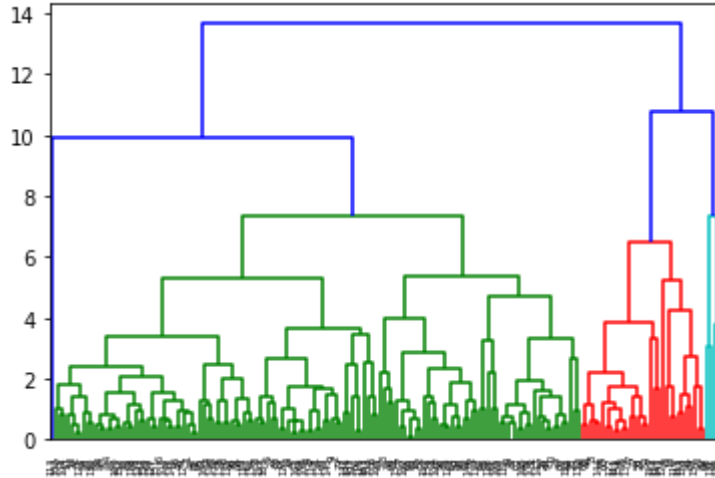
## PCA : Principal Component Analysis

**Principal component analysis (PCA)** is a statistical procedure that uses an [orthogonal transformation](#) to convert a set of observations of possibly correlated variables (entities each of which takes on various numerical values) into a set of values of [linearly uncorrelated](#) variables called **principal components**.

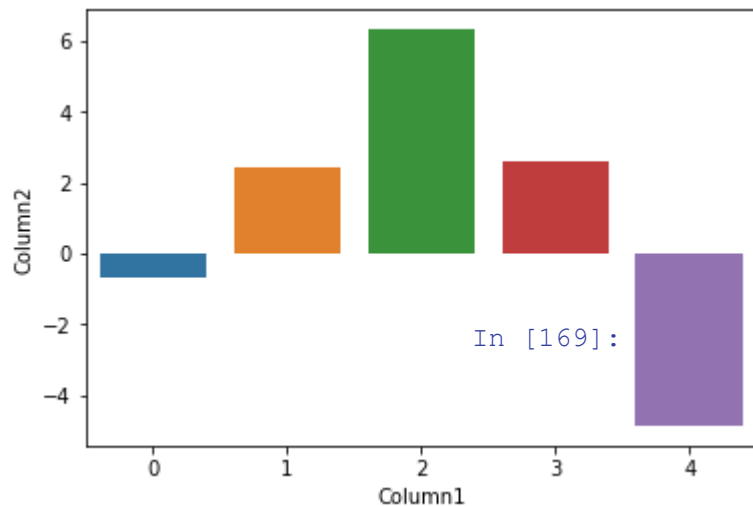
- Identify optimal number of PC using scree plot
- Check if k-means can be performed using hopkins measure
- Once confirmed, perform silhouette and elbow analysis to determine optimal clusters
- Perform clustering with first K to obtain cluster id
- Do crosscheck/analysis of few individual countries selected from each cluster selected for funding
- Perform hierarchical clustering with single and complete linkage on PC dataset and obtain cluster id

# Dendrogram – cut at $K = 6$

In [190]:



# Results of Hierarchical Clustering



# Countries That need help : that falls in Cluster4

1.Haiti

8. Angola

2.Sierra Leone

9 . Burkina Faso

3.Chad

10. Congo, Dem. Rep.

4.Central Africa

5.Mali

6.Nigeria

7.Niger