

Clustering & PCA Assignment

Abstract

- **Business Objective:**

HELP International is a global humanitarian NGO that is focused on battling destitution and furnishing the general population of in backward nations with essential comforts and alleviation amid the season of debacles and normal calamities.

After the ongoing financing programs, they have raised around \$ 10 million. The critical issues that come while settling on this choice are generally identified with picking the 5 nations that are in the direst need of help.

- **Strategy:**

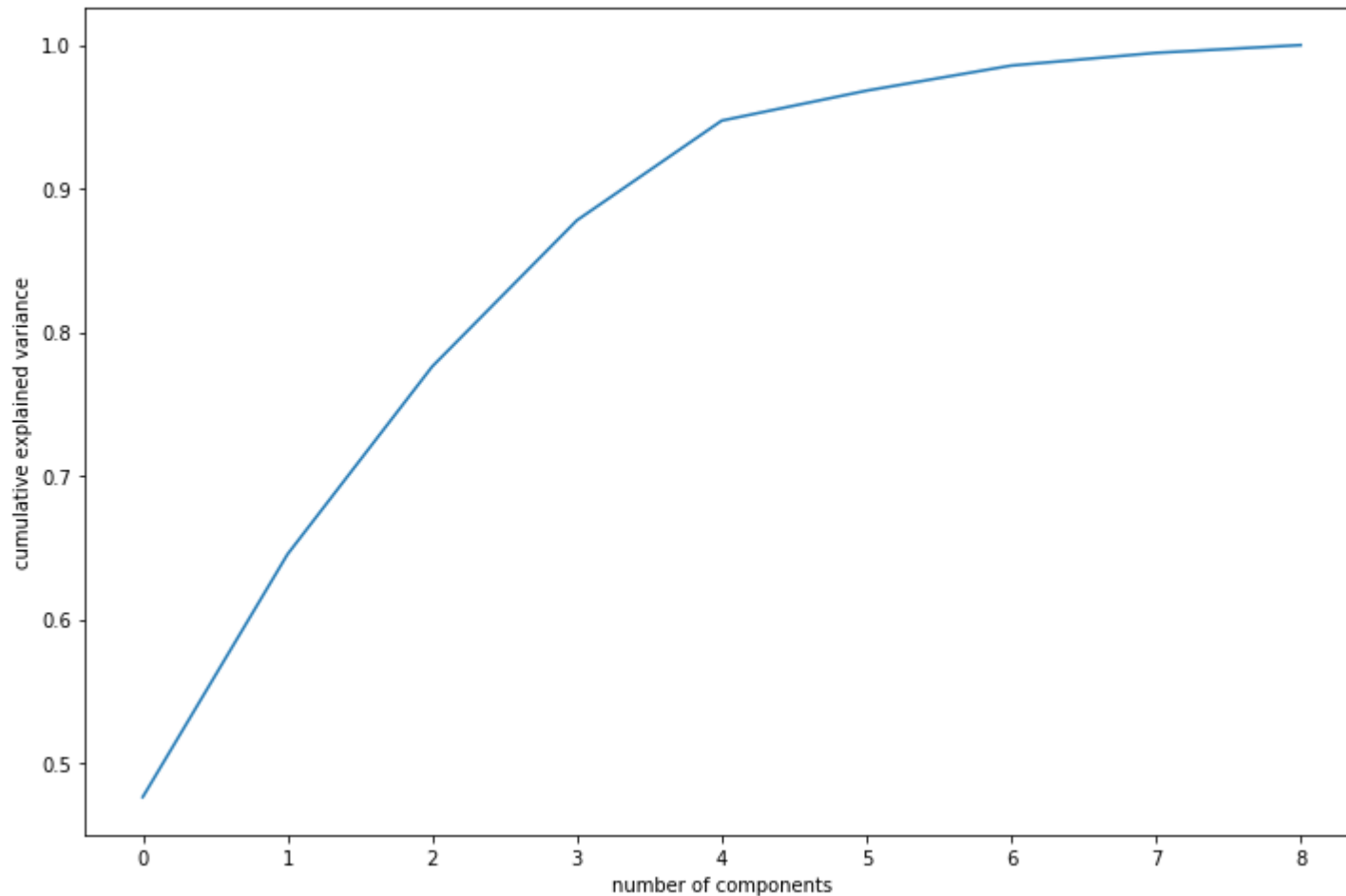
Identify parameters/variables that significantly indicates the need for nations of funds. Through these parameters we can identify 5 countries

Data Preparation

- Remove all Rows having Z^* score less than 3 for outlier analysis
- Delete “country” Column
- Use Standard Scalar
- Assume all relevant information about PC will be between 5% to 95%. We can remove outliers
- No Null values detected
- Columns' data type are correct.

PCA

- Scree Plot

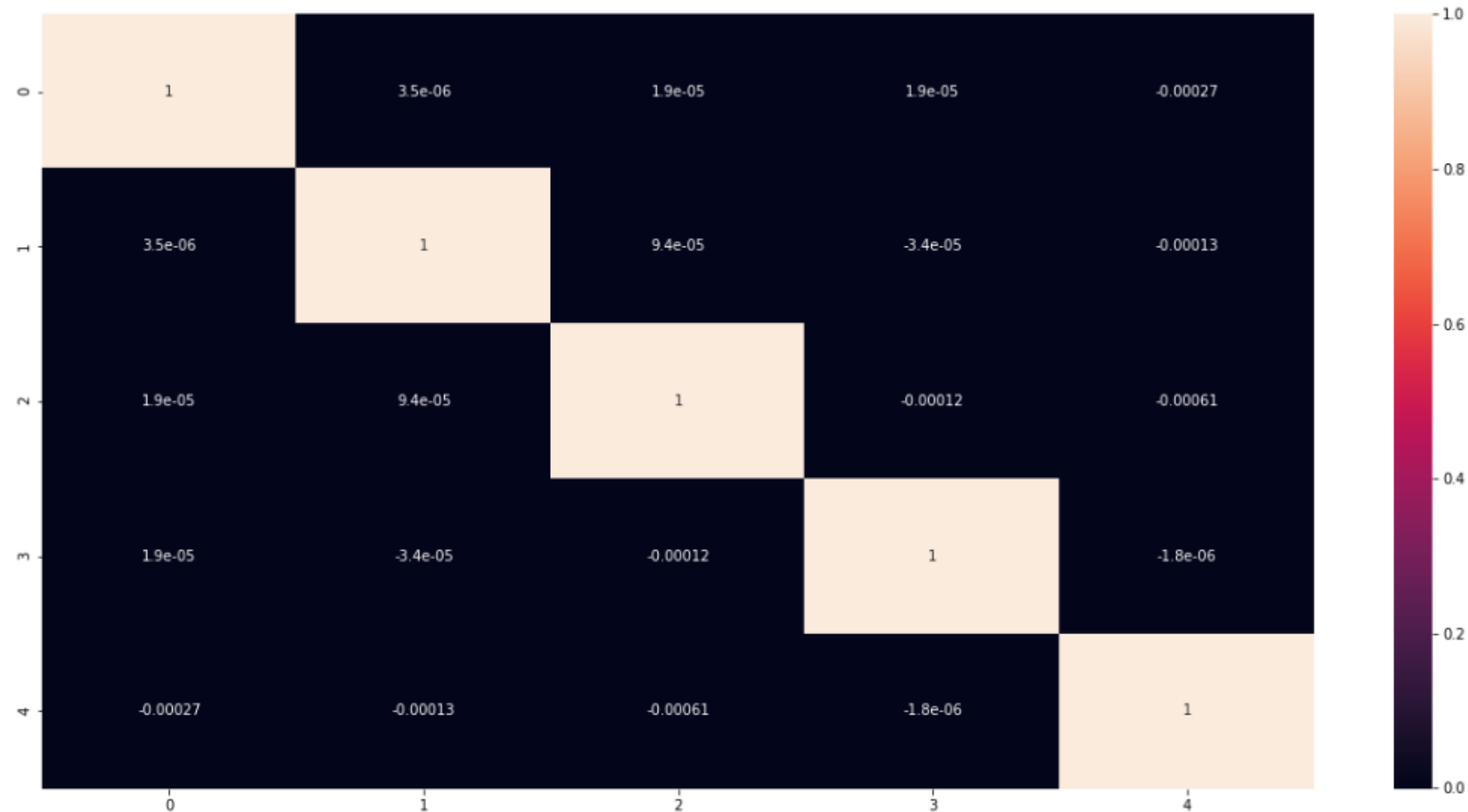


From this Scree Plot, we can assume that with 5 components, 90% Variance will be covered.

PCA

- Heat Map

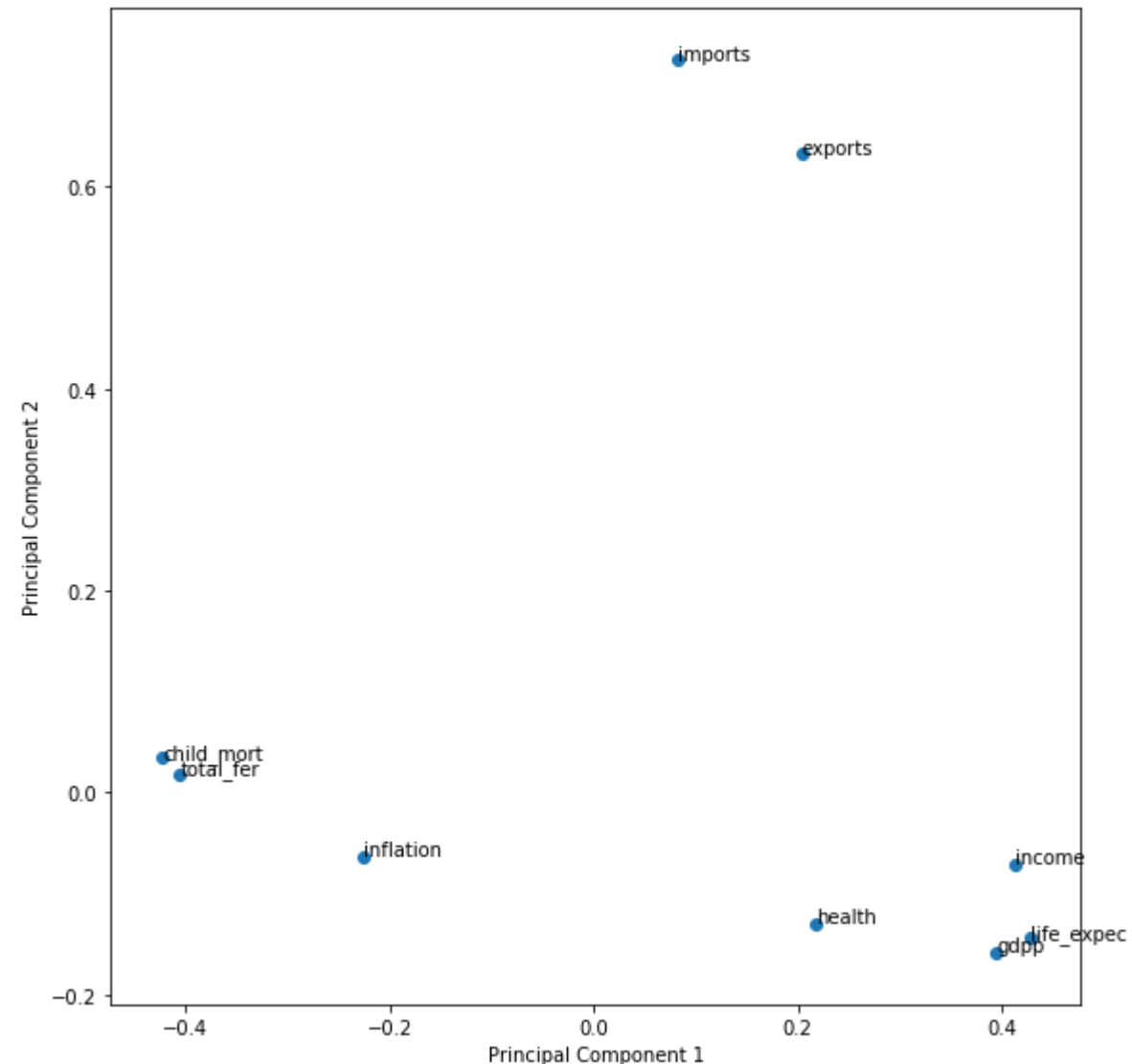
max corr: 9.441162254718294e-05 , min corr: -0.00061081039695825



Heat Map also confirms that 5 components are enough to capture the useful information

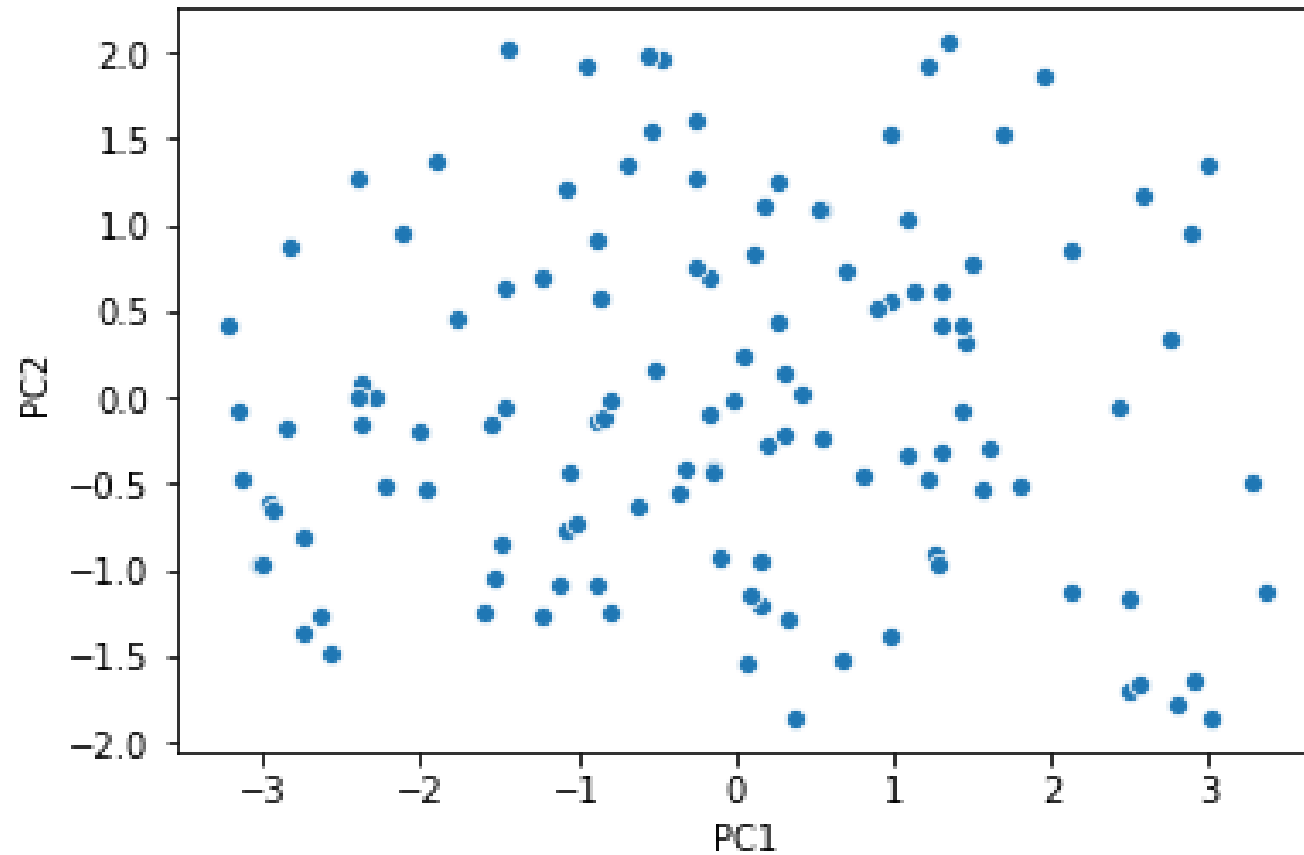
PCA

- Distribution of columns across PCS



PCA

- Plotting values on scatterplot to visualize plots



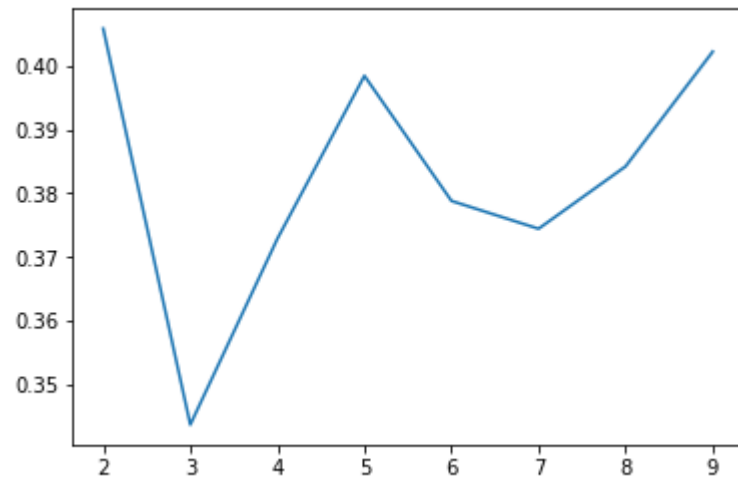
Values are equally spread out. Its little hard to make out right. Lets go to Clustering

PCA

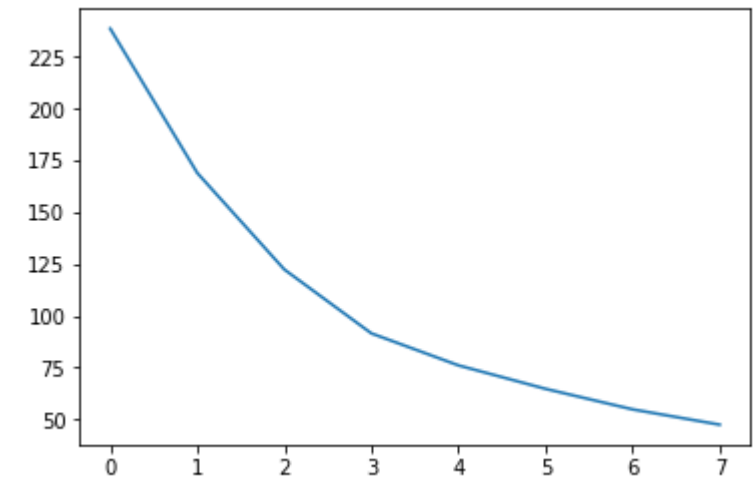
- Now that we have found 5 vectors which represent all data, we can use them for clustering to derive further insights

Clustering

- K-Means



Silhouette curve

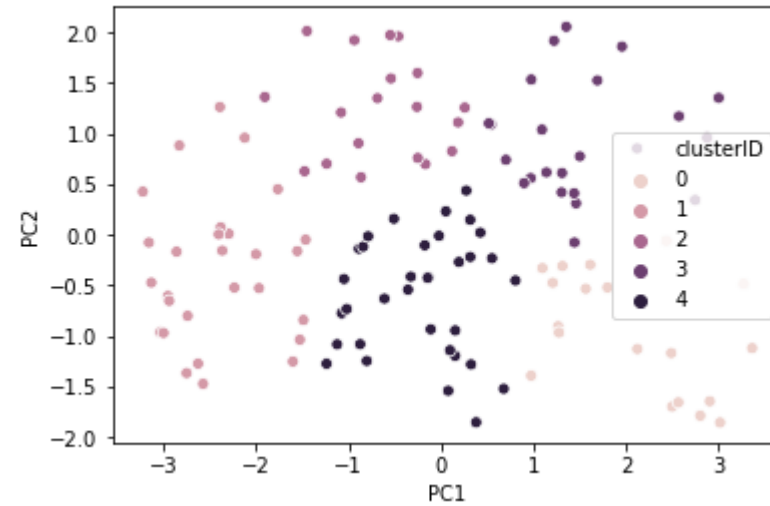


Elbow Curve

From both curves, we can assume that 5 clusters will be enough to give all Information about country's needs

Clustering

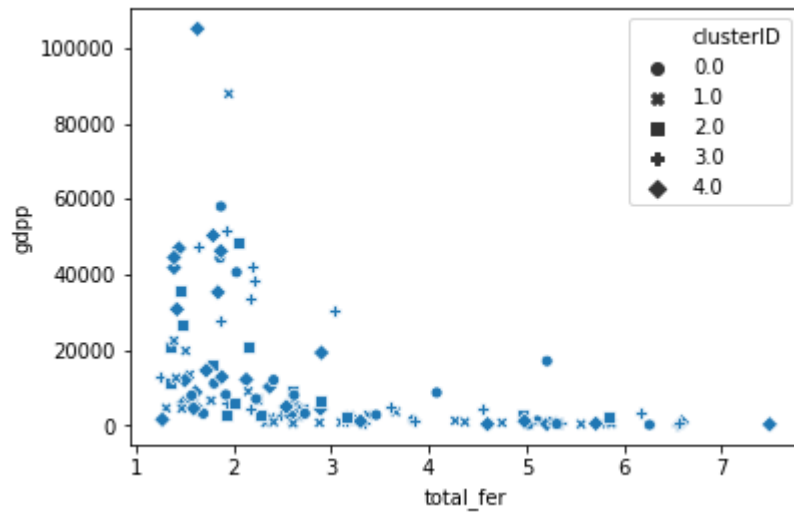
- K-Means



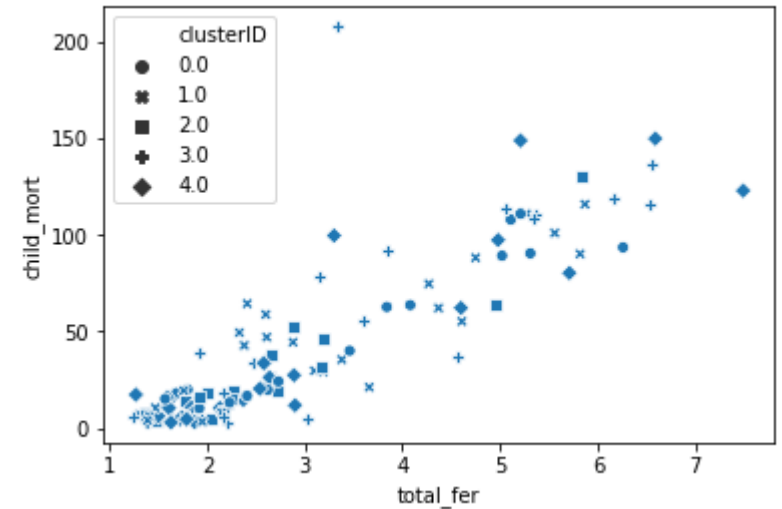
Visualizing the clusters

Clustering

- K-Means



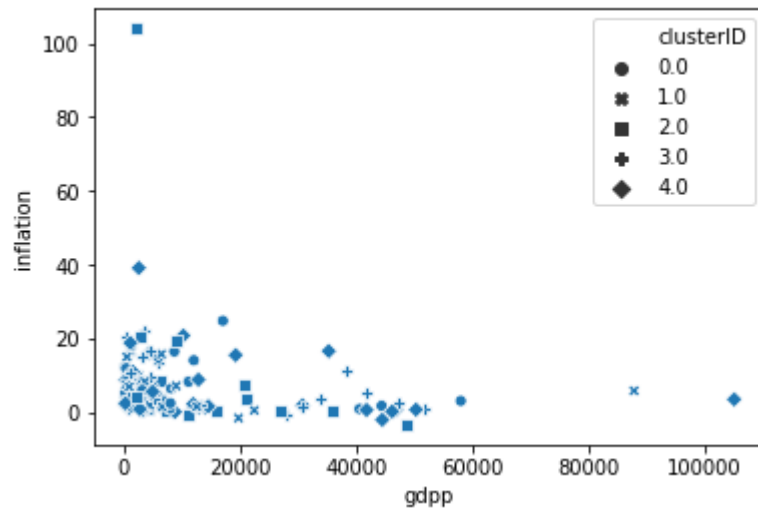
Fertility rate is inversely proportional to GDP w.r.t. each cluster



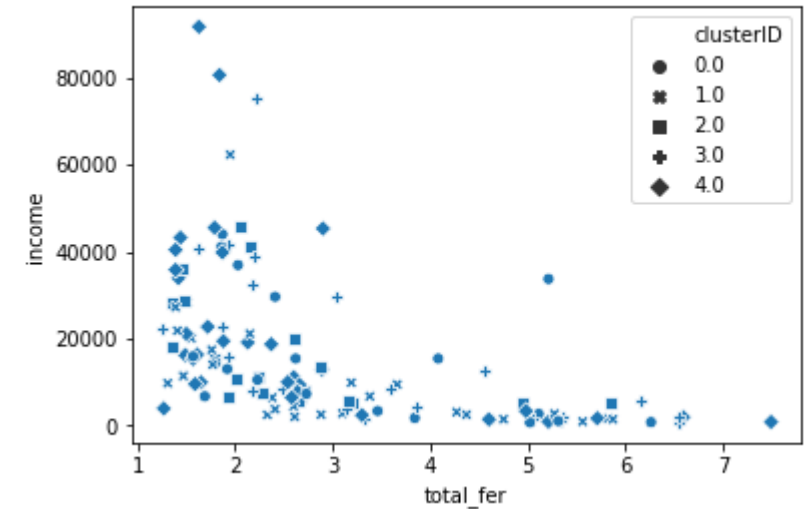
Fertility rate is directly proportional to child Mortality w.r.t. each cluster

Clustering

- K-Means



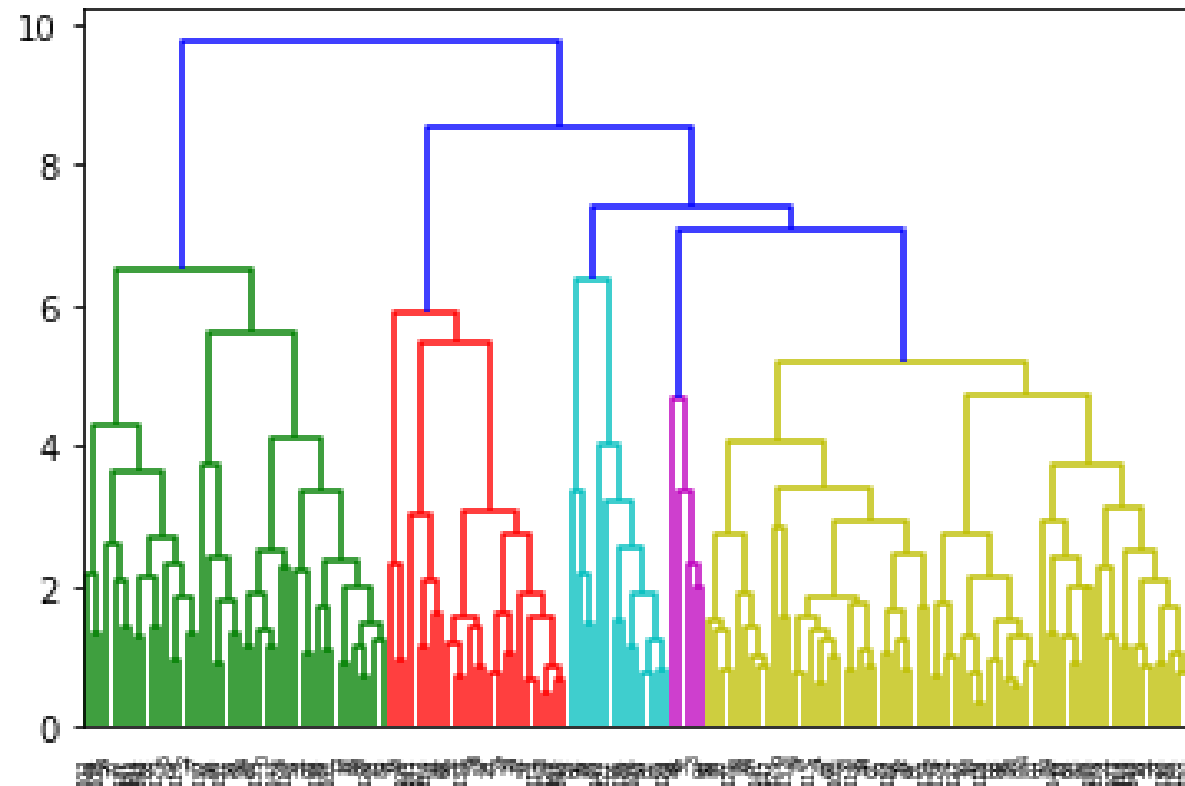
GDP and Inflation are loosely inversely proportional w.r.t. each cluster



Income and Fertility are loosely inversely proportional w.r.t. each cluster

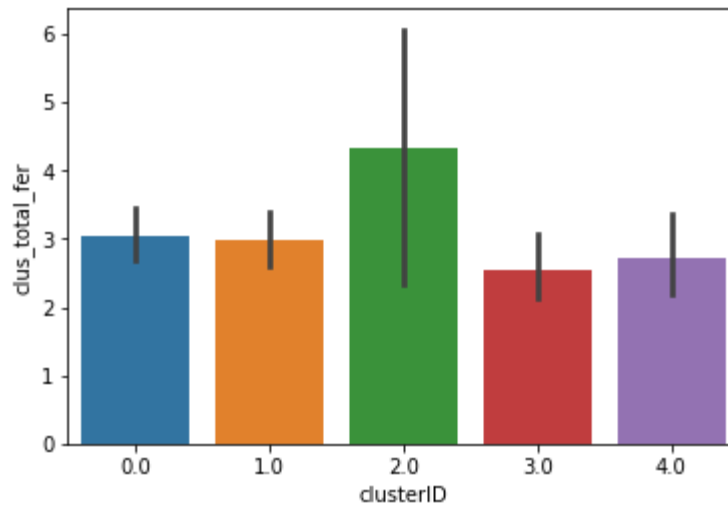
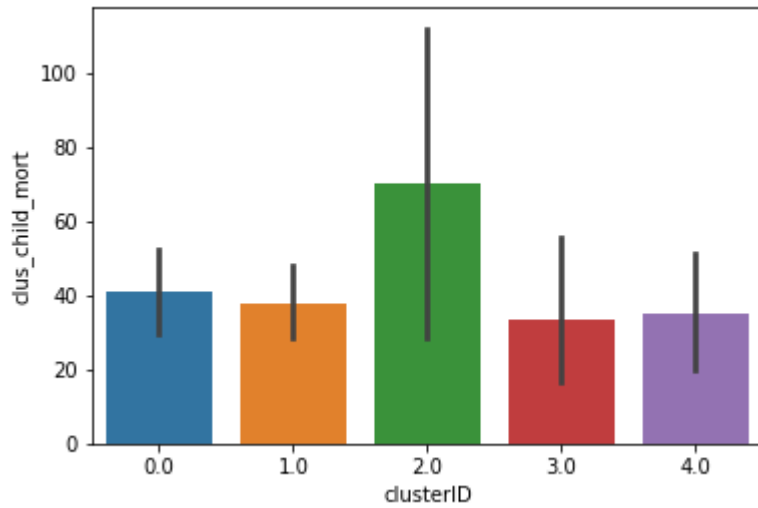
Clustering

- Hierarchical Clustering

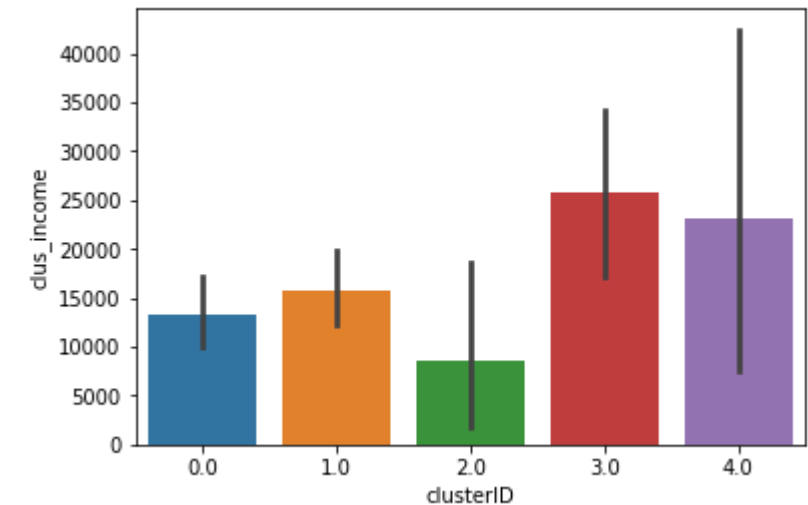
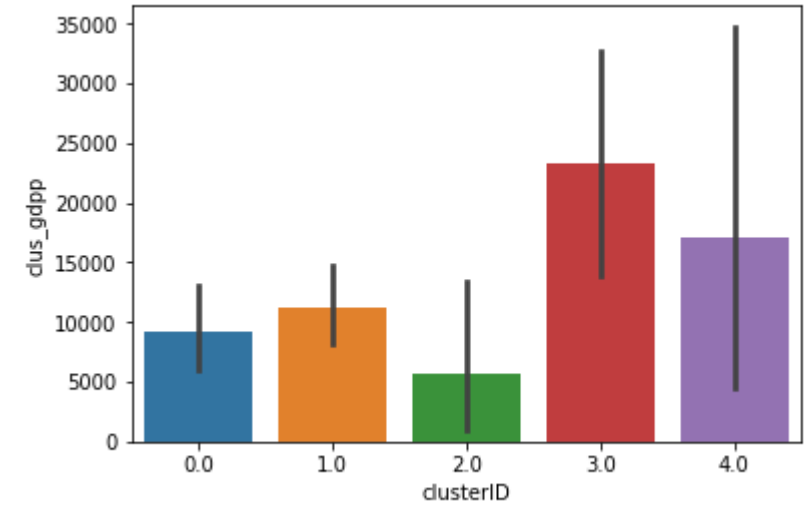


Dendrogram Graph

- Hierarchical Clustering



Analyzing different Bar Plots;
Cluster 2 is performing worst
in all segments





Conclusions for Analysis & Decisions

As per analysis, below are the 5 nations:

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