HELP International international humanitarian NGO

(Predictive Analysis and Clustering using PCA for choosing the countries that are in the direst need of aid.)

Presented By

-Husn Ara

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. After the recent 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.

Analytics Approach

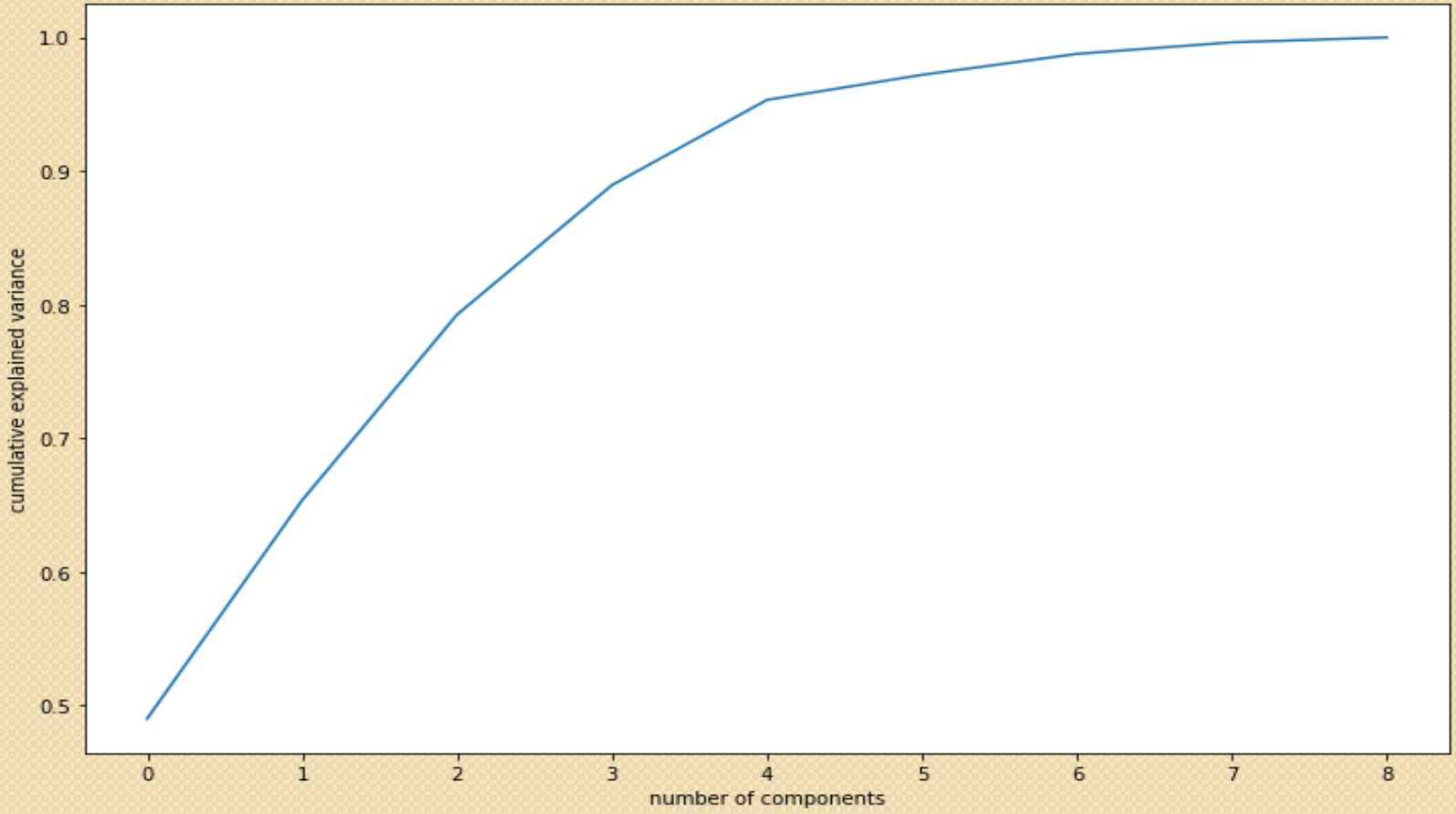
1. Data Cleaning
2. Exploratory Data Analysis
3. Scaling Dataset
4. Principal Component Analysis
5. Transformation
6. K-Means Clustering using Hopkins Statistics & Silhouette Analysis.
7. Hierarchical Clustering

Analysis in Business Aspect

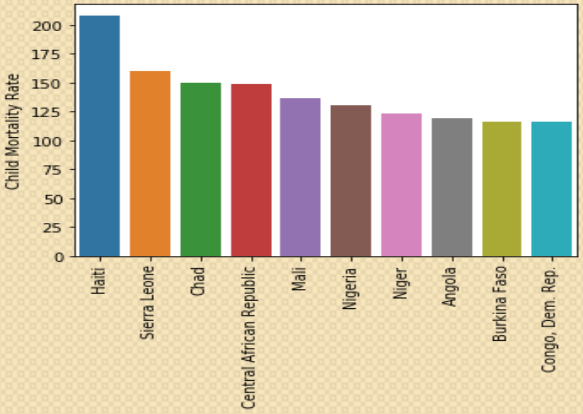
- Find out most backward countries that have low Import, Export, GDP per capita and income per capita.
- Find out inflation (Annual growth rate of GDP), Child mortality, Fertility rate, Life Expectancy average, Health Spending.
- If Country have low income and GDP that means its a Poor Country.
- If Country have high rate of child mortality (Death of children under 5 years of age per 1000 live births), inflation, health spending life expectancy, high fertility rate that means country have no sufficient money to survive well and is in direst need of aid.
- After getting list of most backward countries in chronological order, NGO can easily decide to divide money accordingly to needed countries.

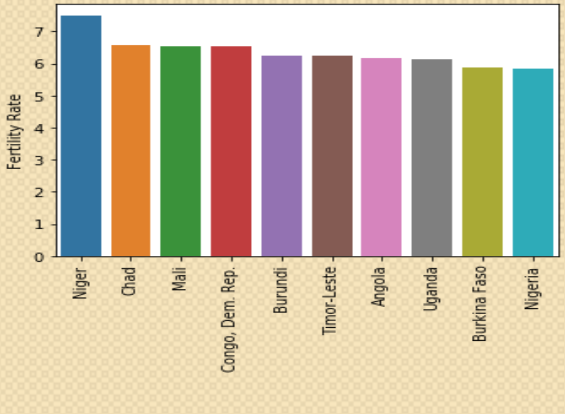
PRINCIPAL COMPONENT ANALYSIS RESULT

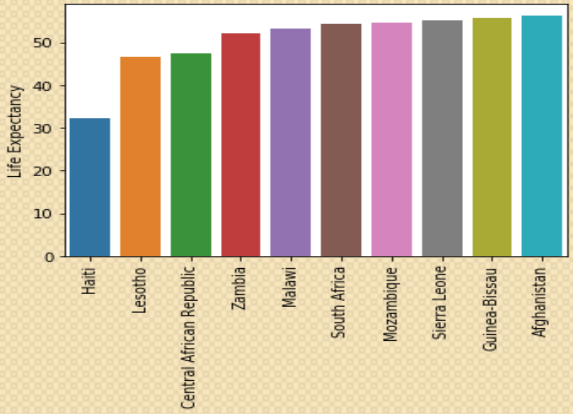
After performing PCA we get result that 4 components are sufficient for 95% variance. It was clearly shown in scree plot-

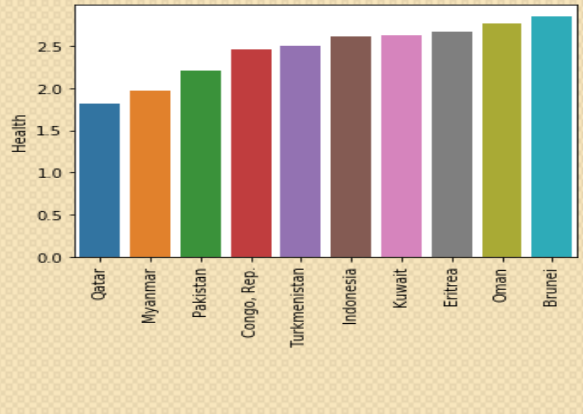


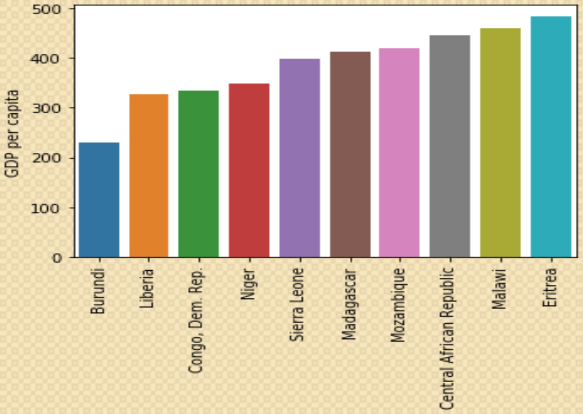
BAR PLOTS FOR ALL FACTS

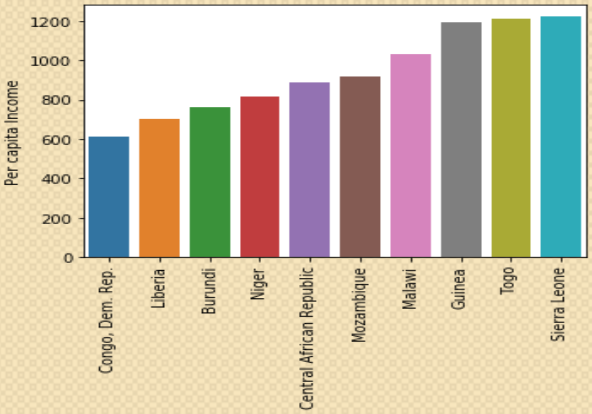


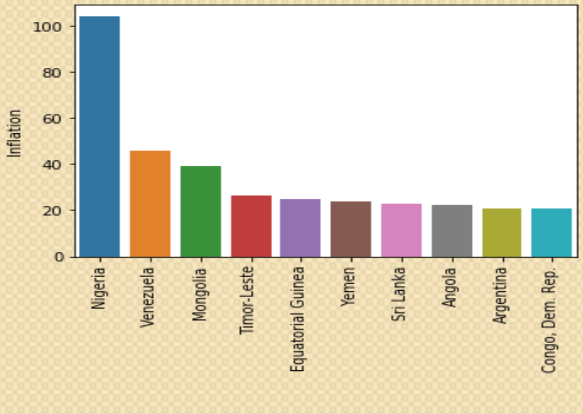


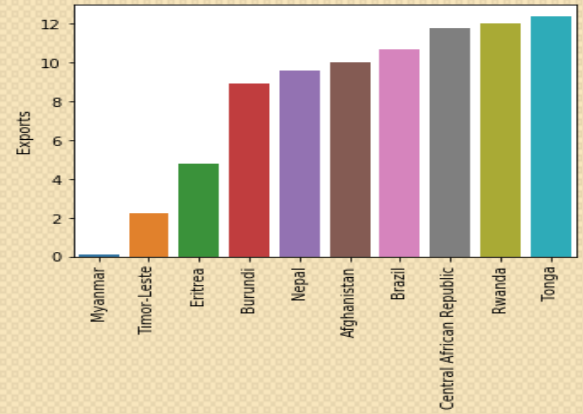


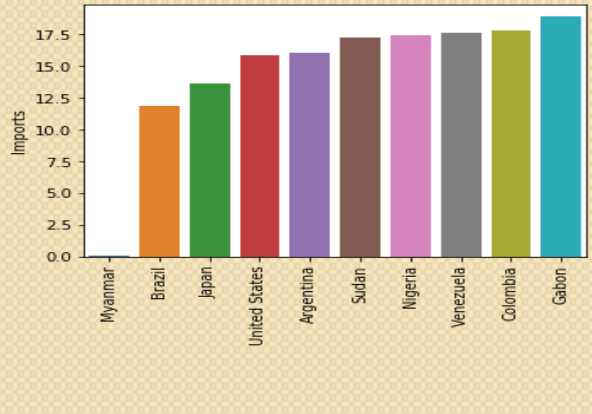




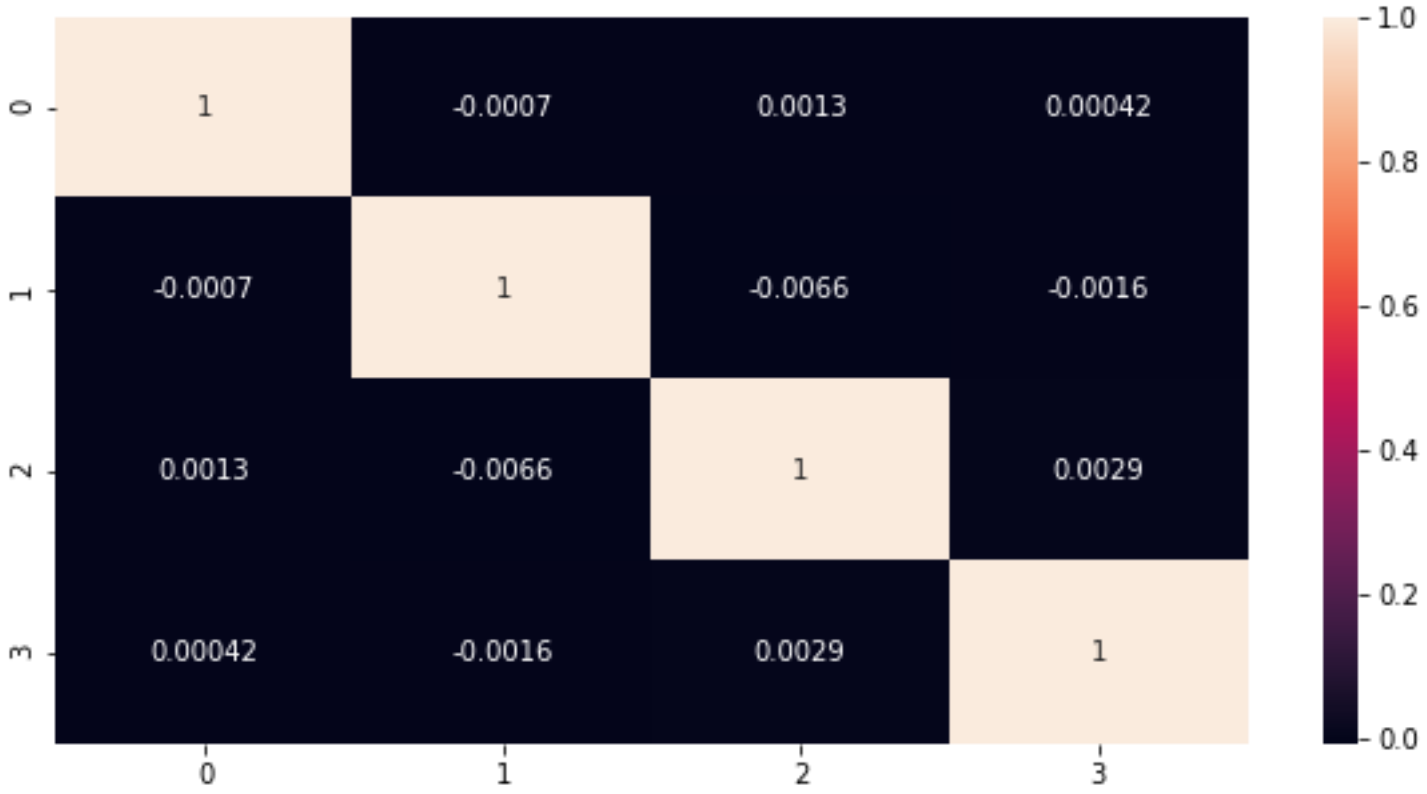








After that we performed transformation and get correlation matrix that have principal components with low correlation.



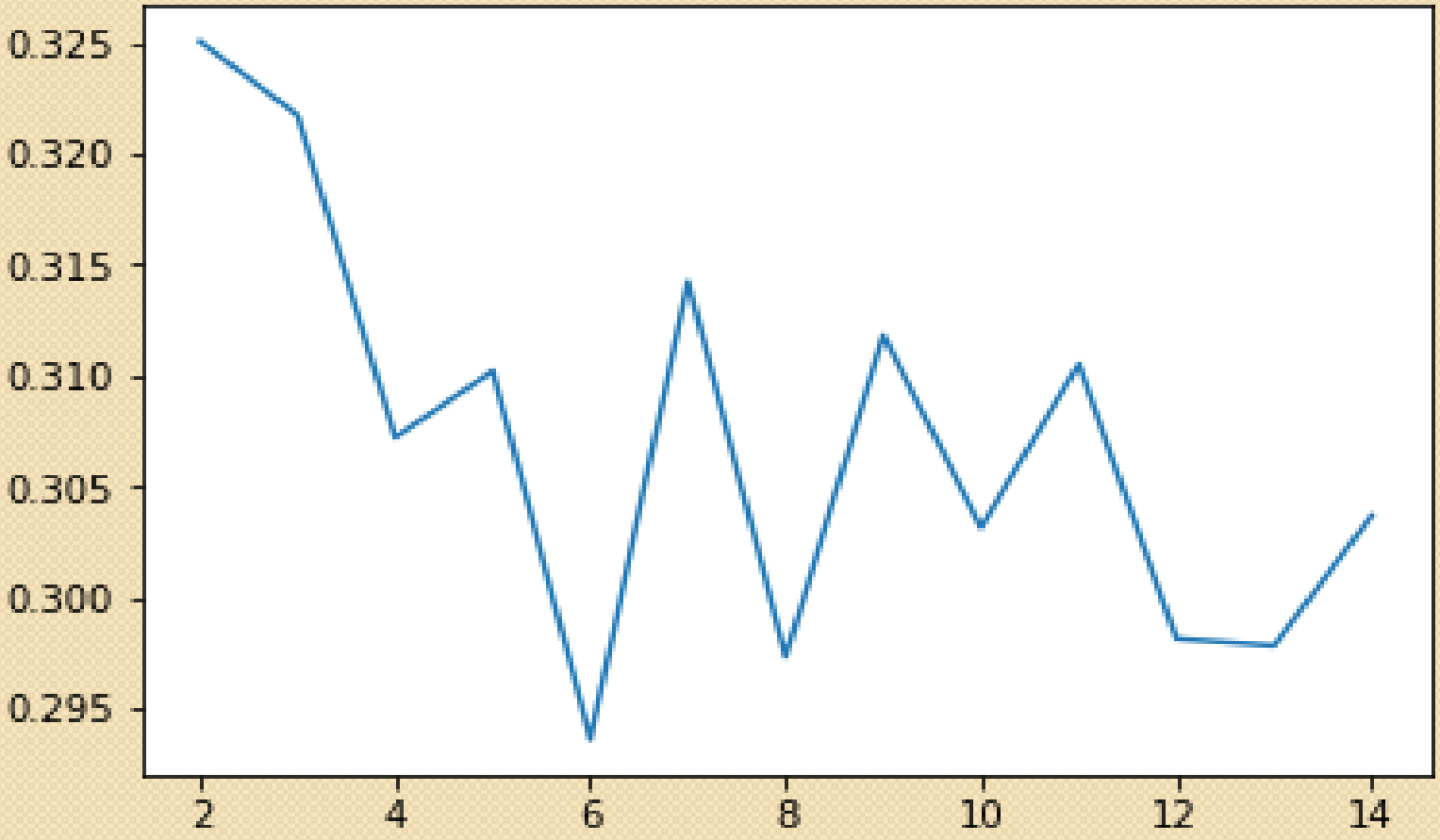
K-Means Clustering (Hopkins Statistics)

- Hopkins statistics Rule-
- The Hopkins statistic, is a statistic which gives a value which indicates the cluster tendency, in other words: how well the data can be clustered.
- If the value is between {0.01, ...,0.3}, the data is regularly spaced.
- If the value is around 0.5, it is random.
- If the value is between {0.7, ..., 0.99}, it has a high tendency to cluster.
- After Performing Hopkins statistics on PCs dataset we got result 0.7311293910509467 that clearly state that It has High Tendency to Cluster.

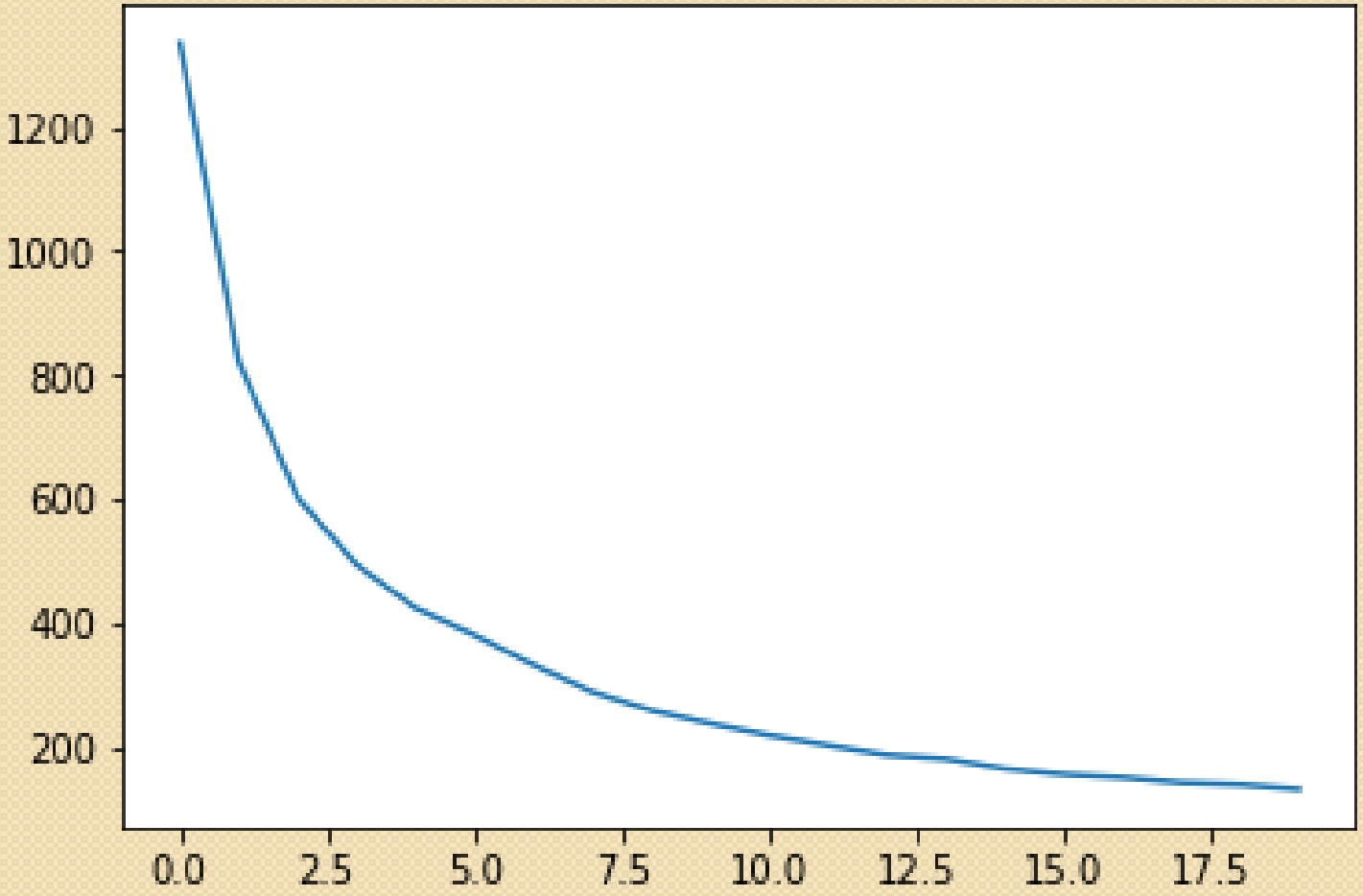
K-MEANS CLUSTERING

(SILHOUETTE ANALYSIS)

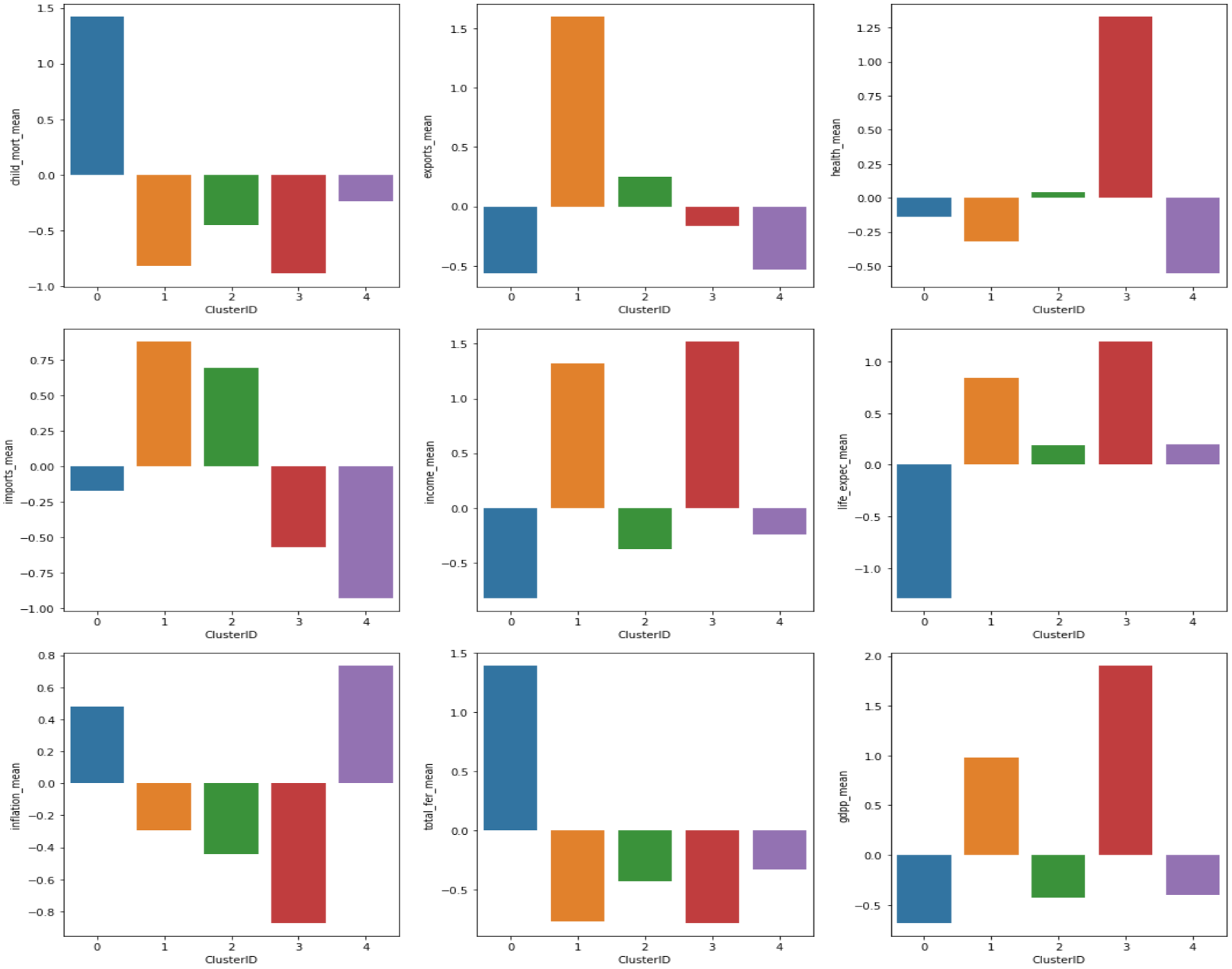
Plot Generated after Silhouette Analysis-



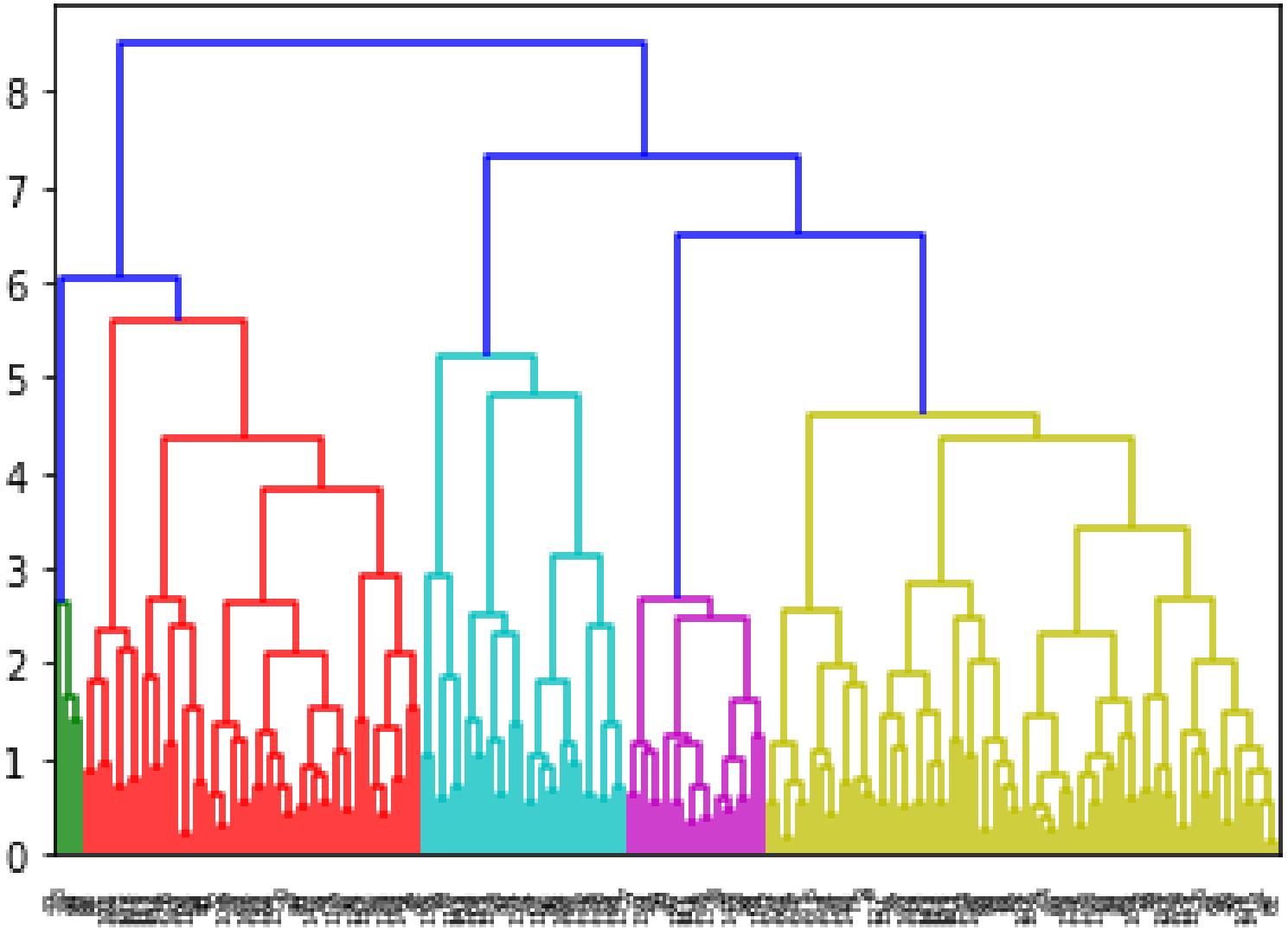
SUM OF SQUARED DISTANCES



Bar Plot after Silhouette Analysis for all Facts



Hierarchical clustering (Dendrogram)



Conclusion and Recommendation

The Resultant countries list indicates that these countries are the most backward country in this dataset, with cluster_id zero.Company need to focus on these countries-

Afghanistan	Botswana	Benin	Burkina Faso
Burundi	Cameroon	Central African	Republic
Chad	Comoros	Congo	Dep. Rep.
Cote d'Ivoire	Eritrea	Guinea-Bissau	Guinea
Madagascar	Liberia	Lesotho	Lao
Malawi	Kiribati	Kenya	Iraq
Haiti	Gabon Gambia	Ghana	Mali
Micronesia	Mozambique	Fed. Sts.	Namibia
Niger	Nigeria	Pakistan	Rwanda
Senegal	Sierra Leone	Solomon Islands	South Africa
Sudan	Tajikistan	Tanzania	Timor-Leste
Togo	Uganda	Yemen	Zambia

Thank You