Clustering Assignment

-- by Anindita Kundu

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 funding programmers, 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.

Objective:

The requisite is:

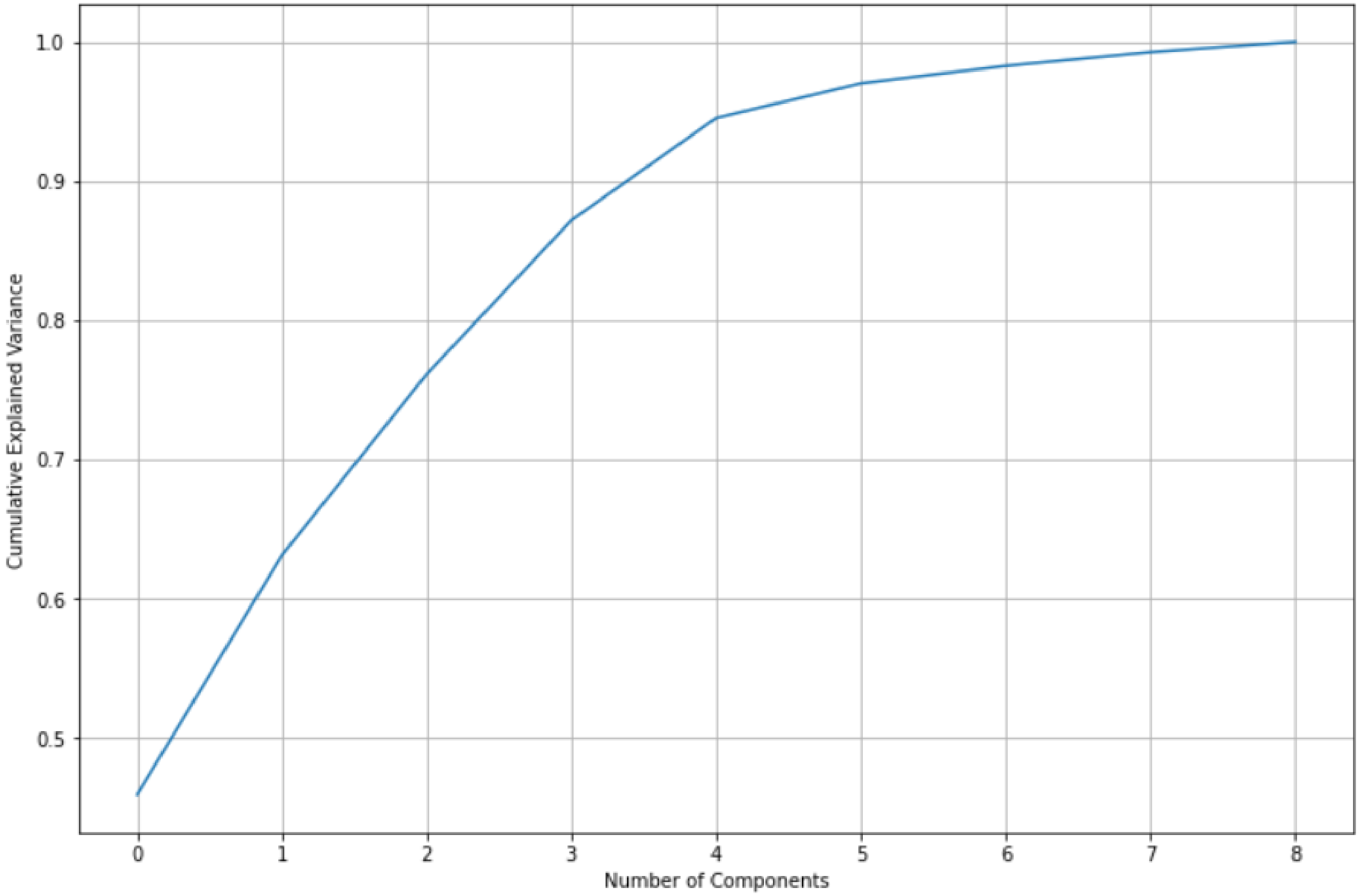
- To categorize the countries using some socio-economic and health factors that determine the overall development of the country.
- To suggest the countries which the CEO needs to focus on the most.

Data Processing:

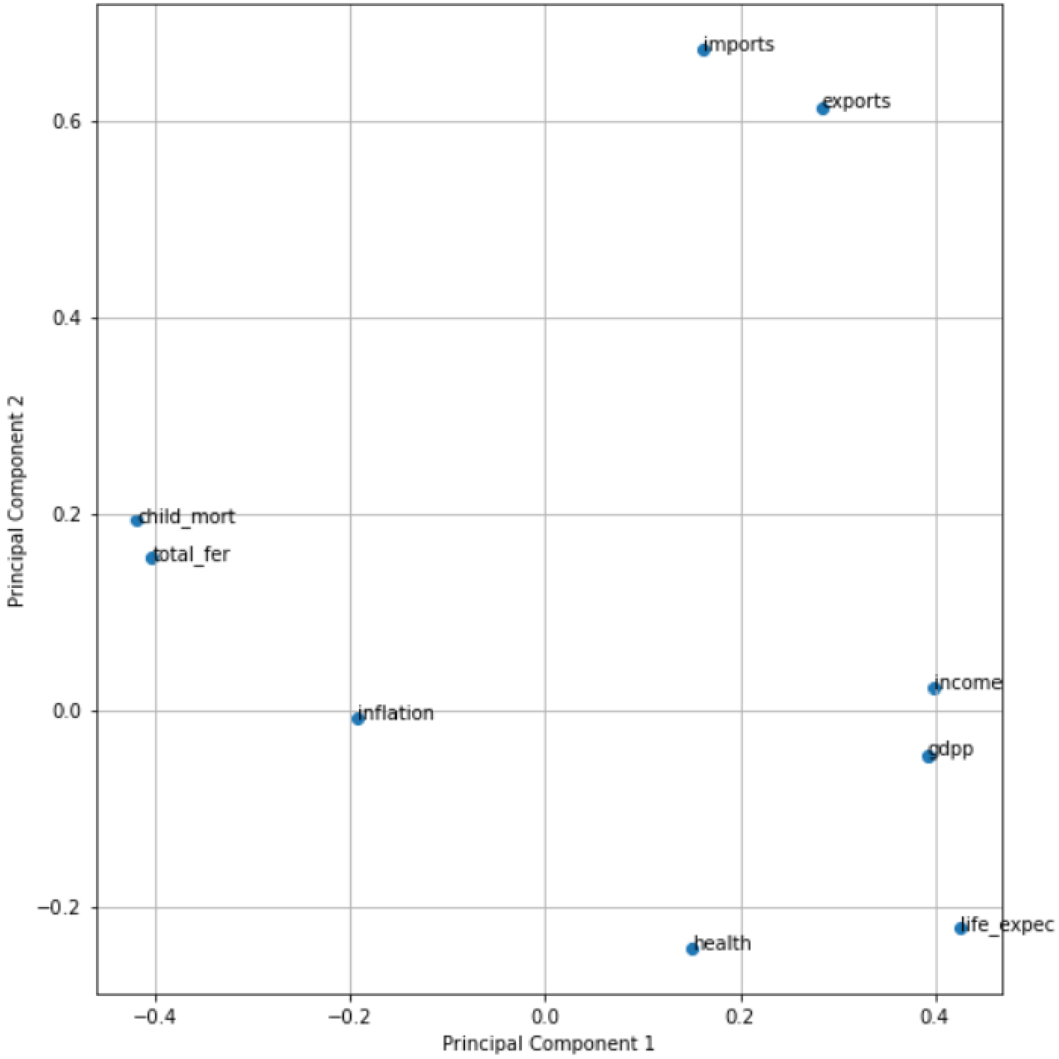
- No Null values were found
- No duplicates for country was found
- There were few outliers and those were handled during Principal Component Analysis (PCA)
- Standardization was done for PCA

PCA Screeplot

Choosing PC as 4 since then there is almost 95% variance



PCA



imports
exports
child_mort
total_fer
income
inflation
gdpp
life_expec
health
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.2
0.0
0.2
0.4
Principal Component 1
Principal Component 2

CLUSTERING

- - Both the methods K means and Hierarchical Clustering was used on the 4 PCA components

    - For K means , K= 3 was taken.

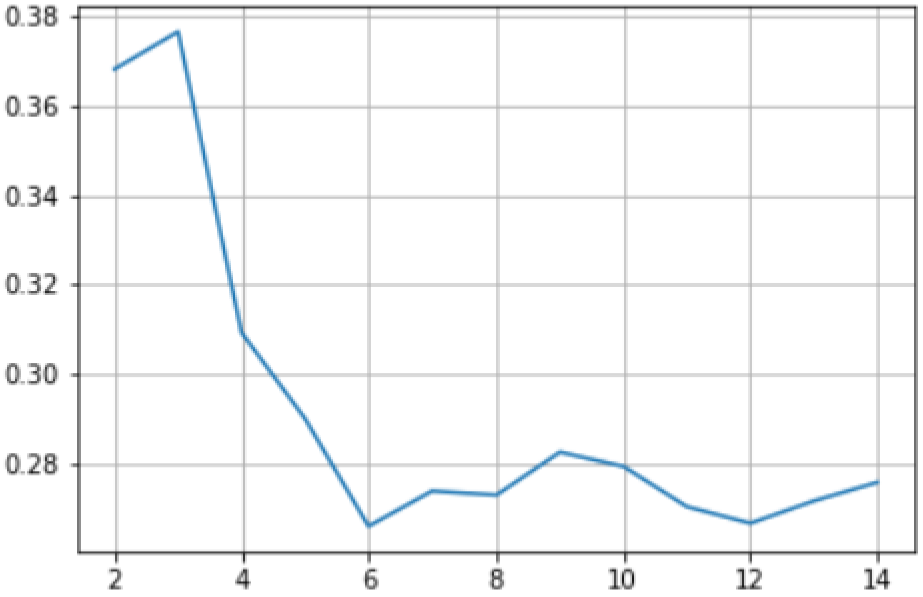
    - While doing the Hopkins Statistics a value of 0.77 was attained.

If the values are:

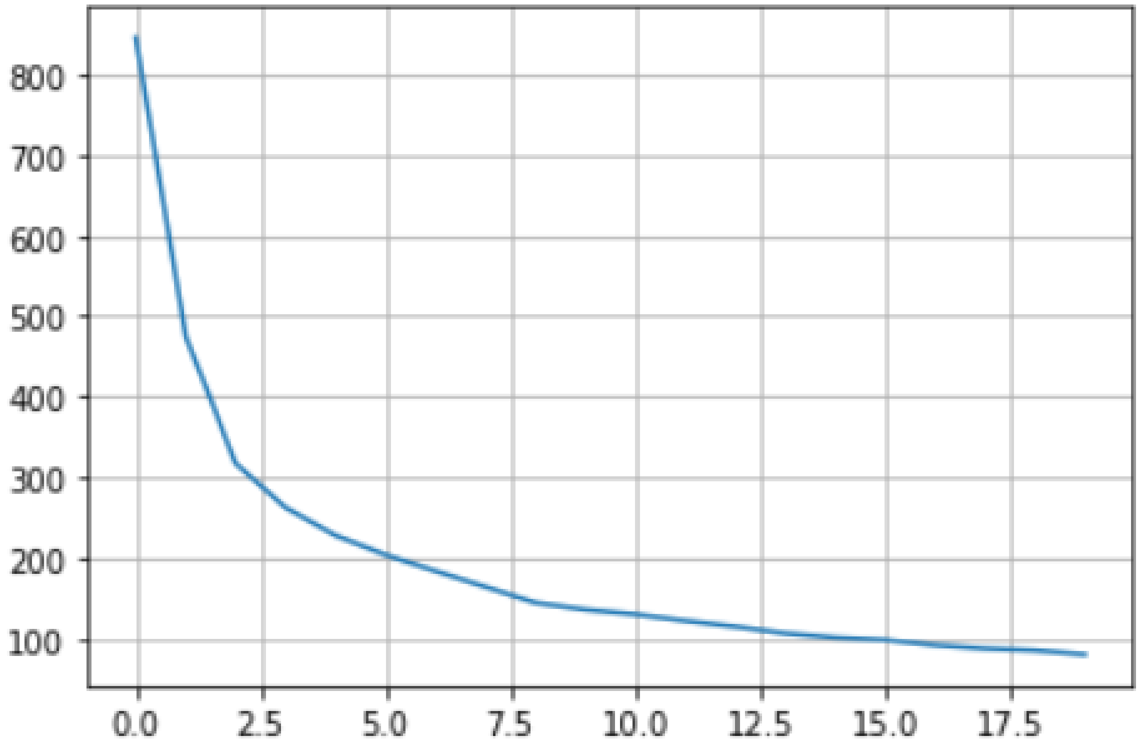
0.01 - 0.3	: Low chase of clustering
Around 0.5	: Random
Between 0.7 - 0.99	: High chance of clustering

SILHOUETTE ANALYSIS (PEAKING AT 3)

- Value of the silhouette score range is in between -1 to 1.
- A score closer to 1 : The data point is very similar to other data points in the cluster
- A score closer to -1 : The data point is not similar to the data points in its cluster.

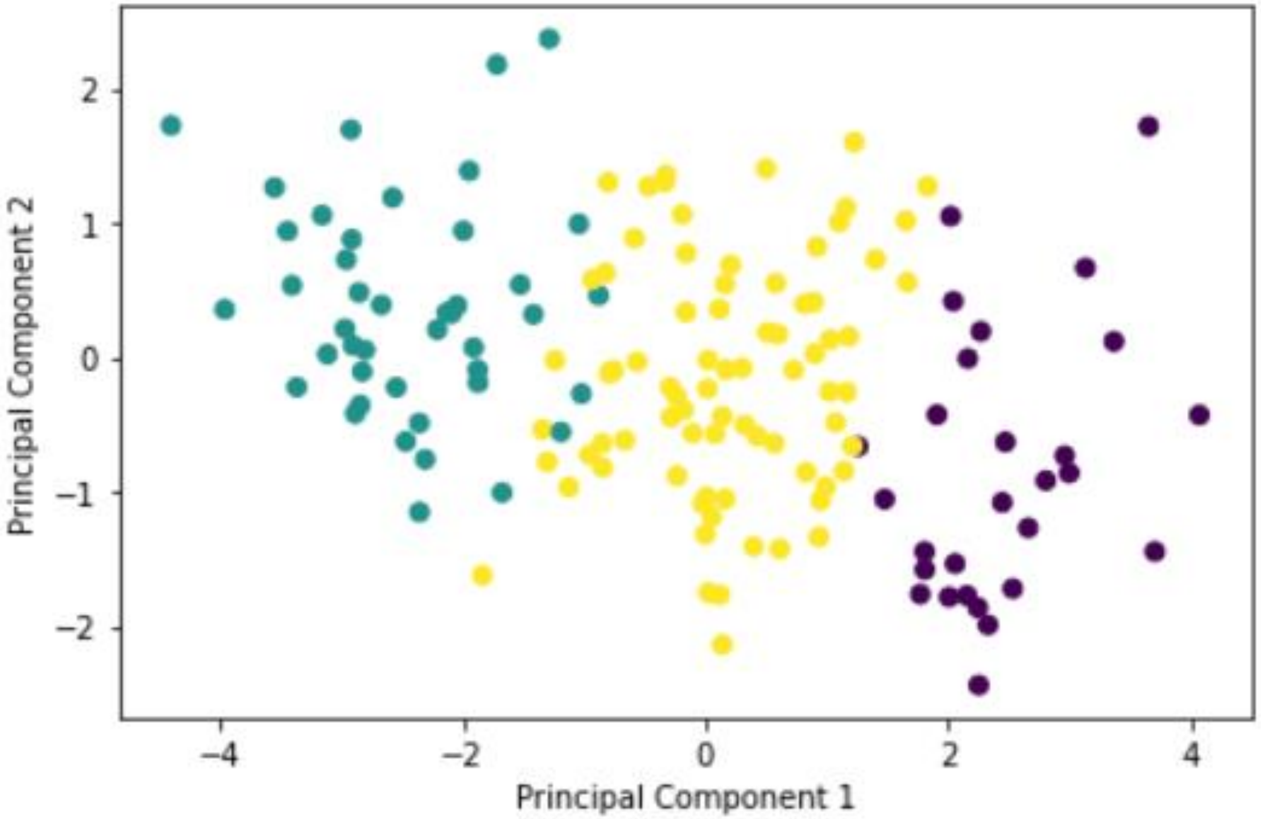


SUM OF SQUARED DISTANCES

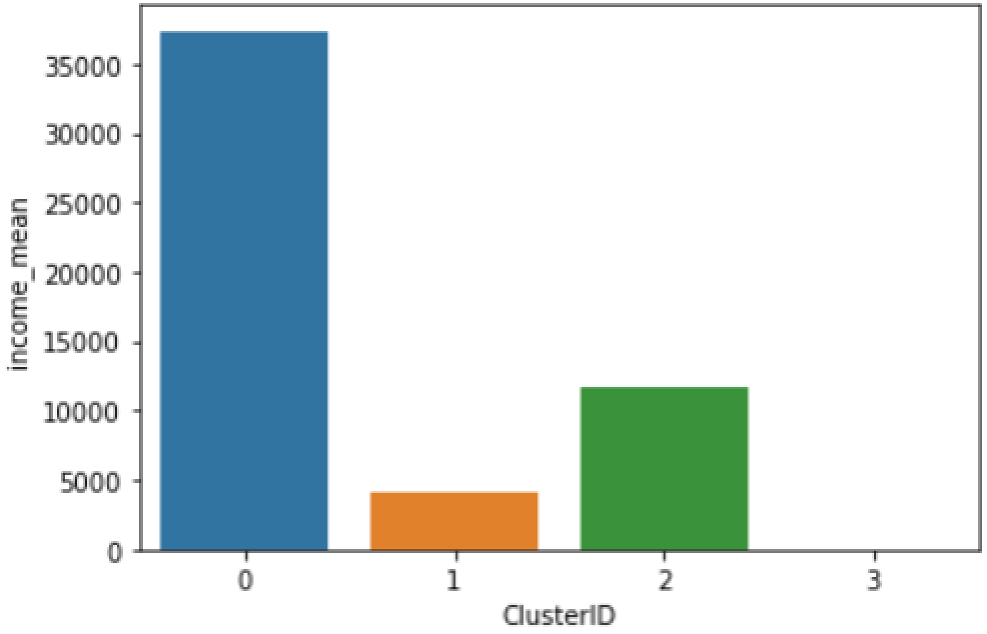


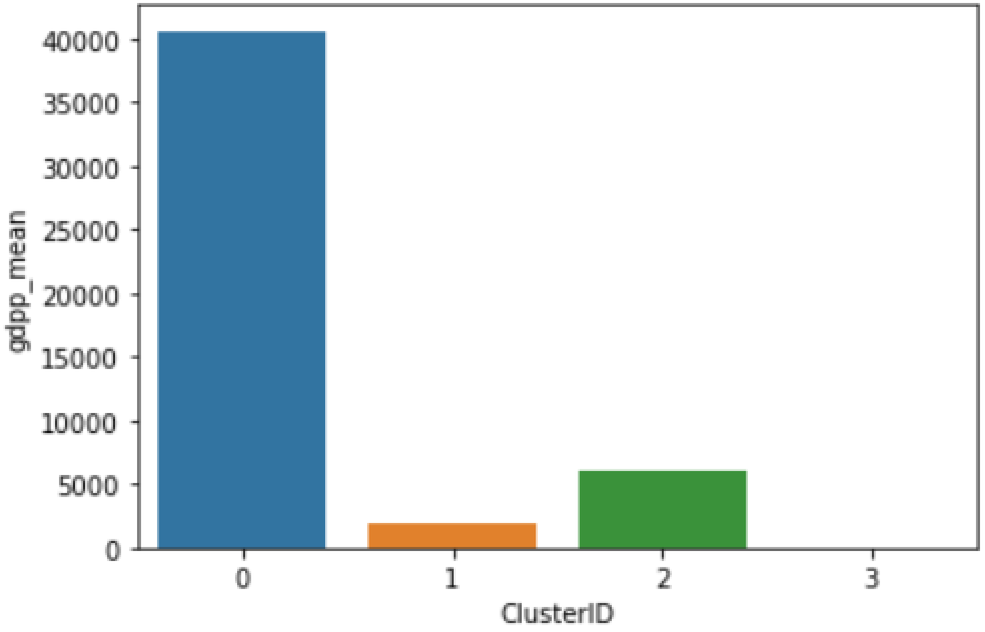
CLUSTERING

Variable Relation with 2PCs



CLUSTERING

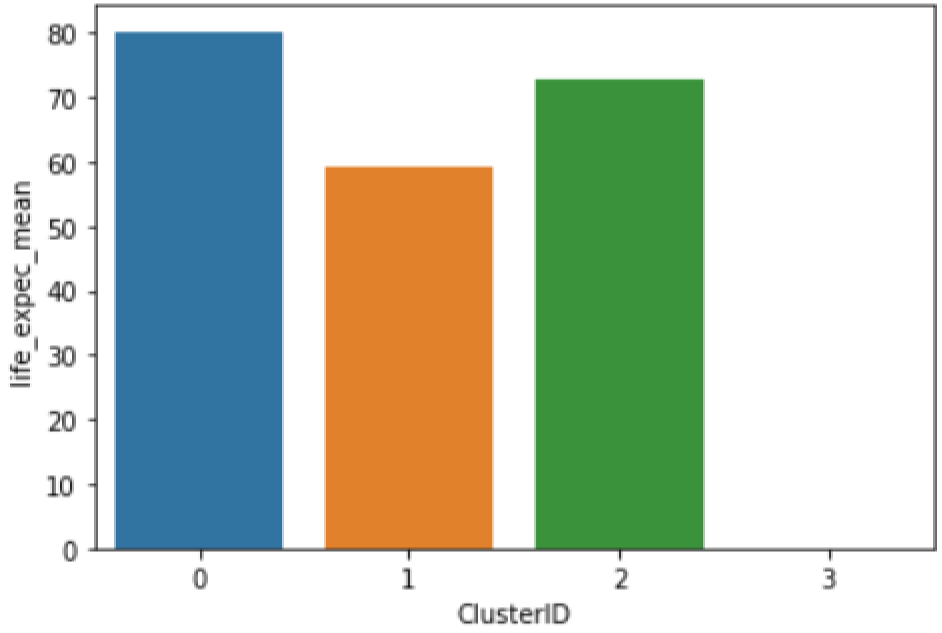


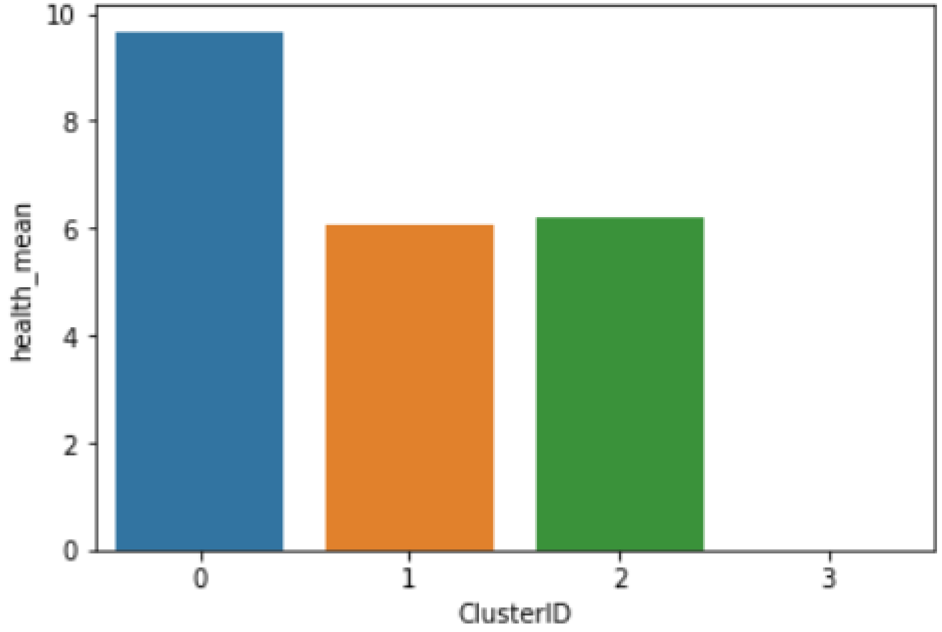


Net income per person is the lowest in ClusterID = 0

The GDP per capita is the lowest in ClusterID = 0

CLUSTERING

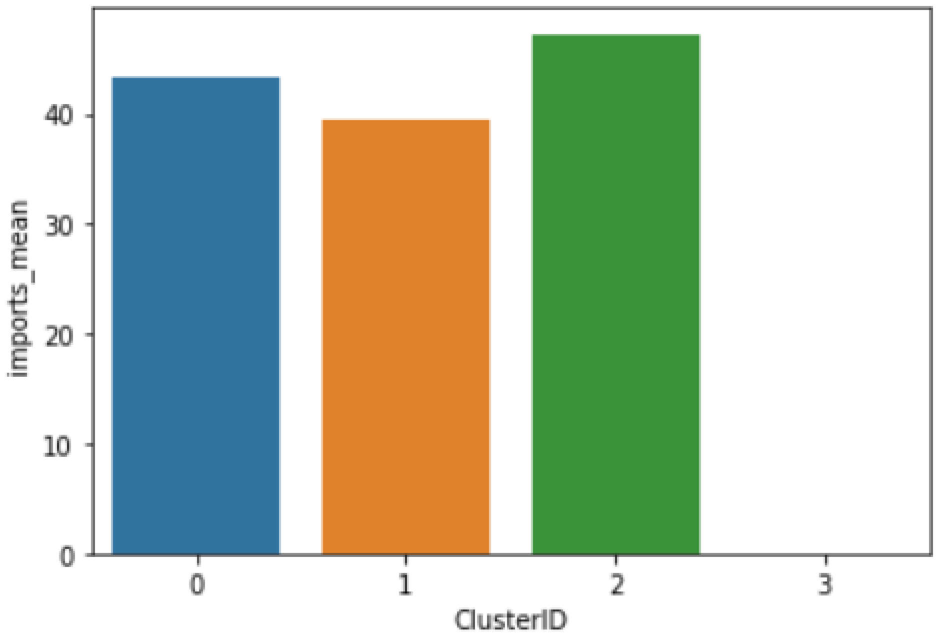


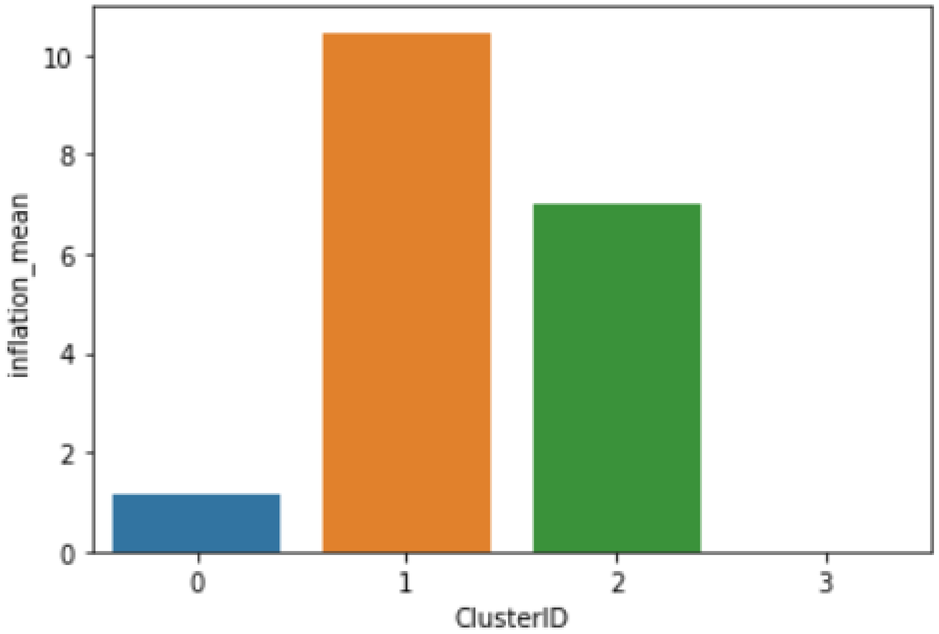


The average number of years a new born child would live is the lowest in ClusterID = 0

Total health spending is the lowest in ClusterID = 0

CLUSTERING

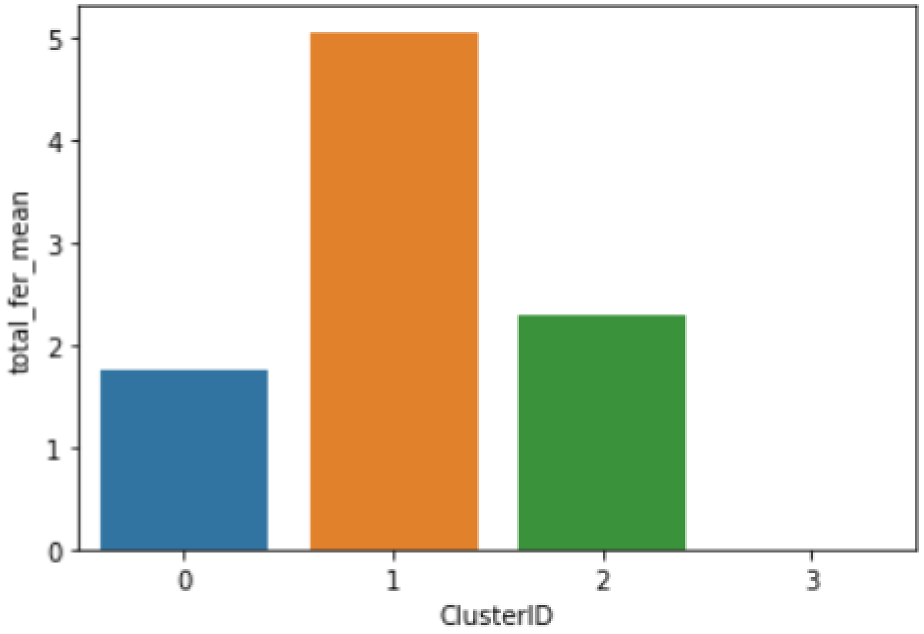




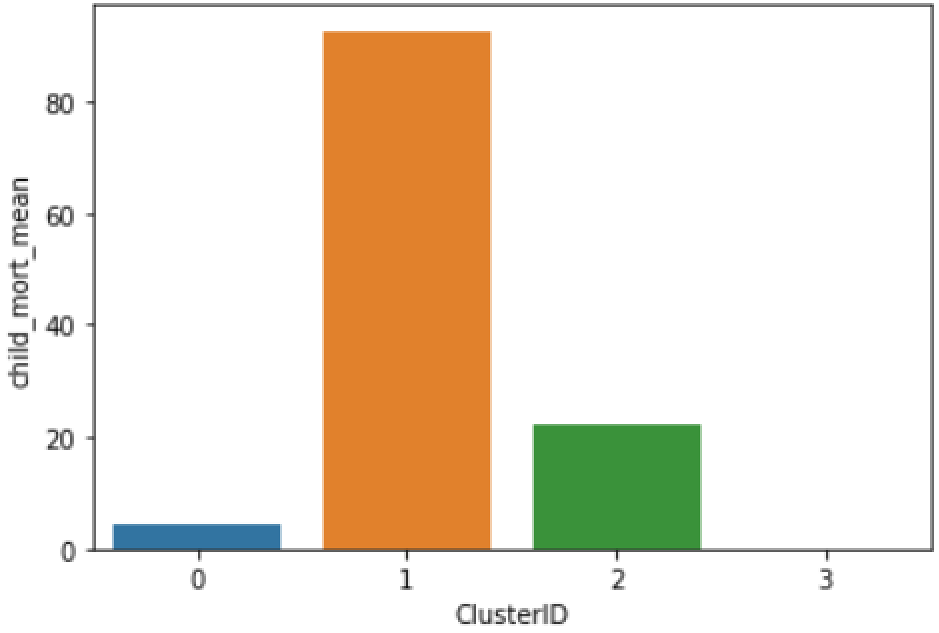
Imports of goods and services is the lowest in ClusterID = 0

The measurement of the annual growth rate is the highest in ClusterID = 0

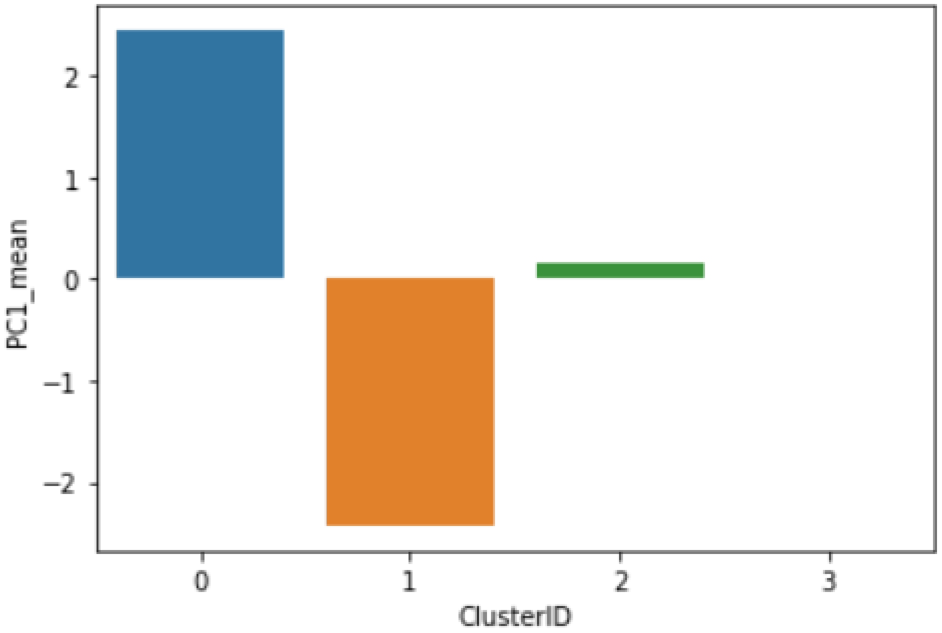
CLUSTERING



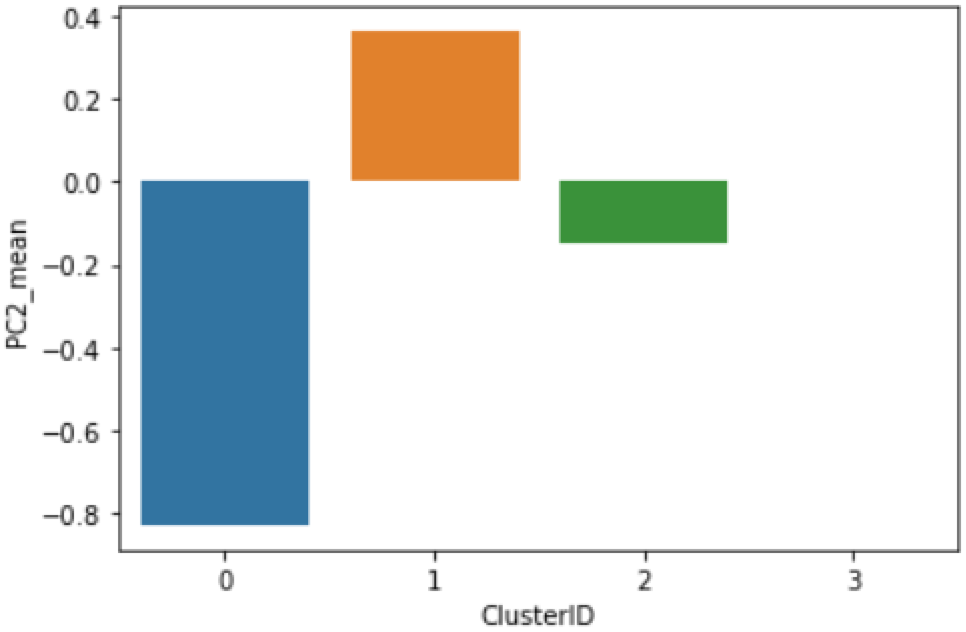
The number of children that would be born is the highest in ClusterID = 0



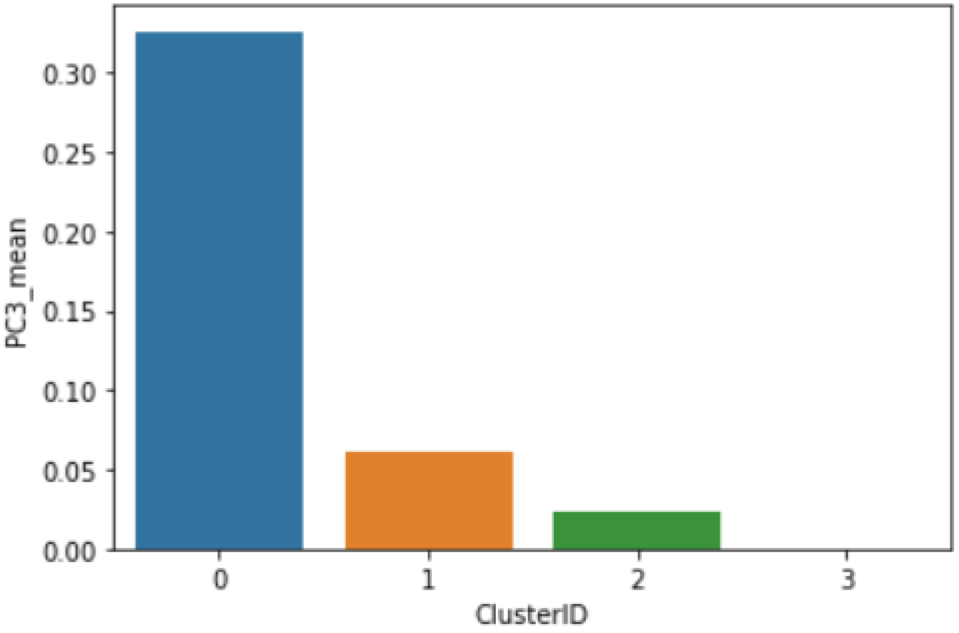
Child Mortality rate is the highest in ClusterID = 0

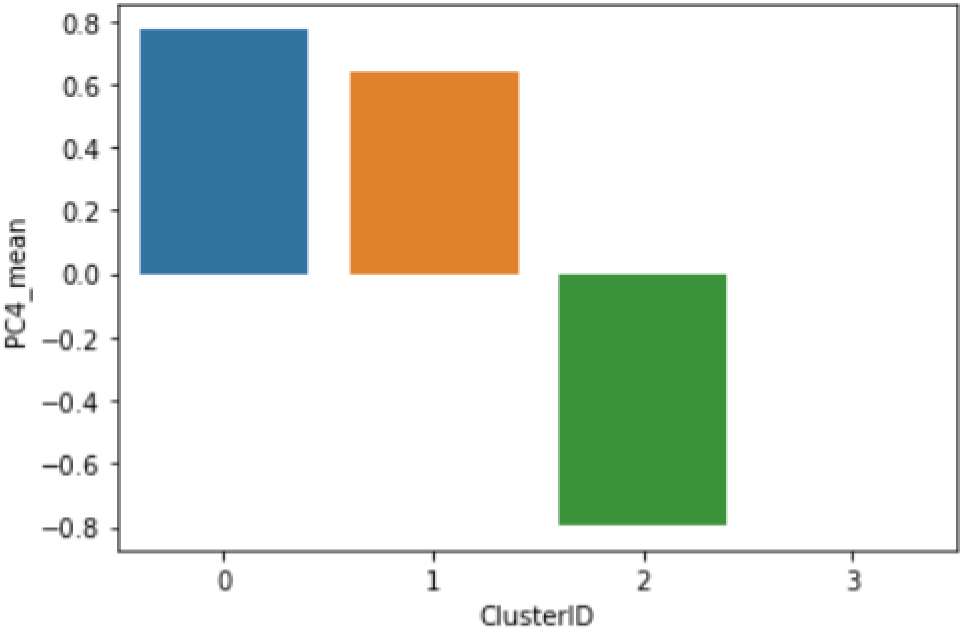


2
1
0
−1
−2
PC1_mean
0
1
2
3
ClusterID



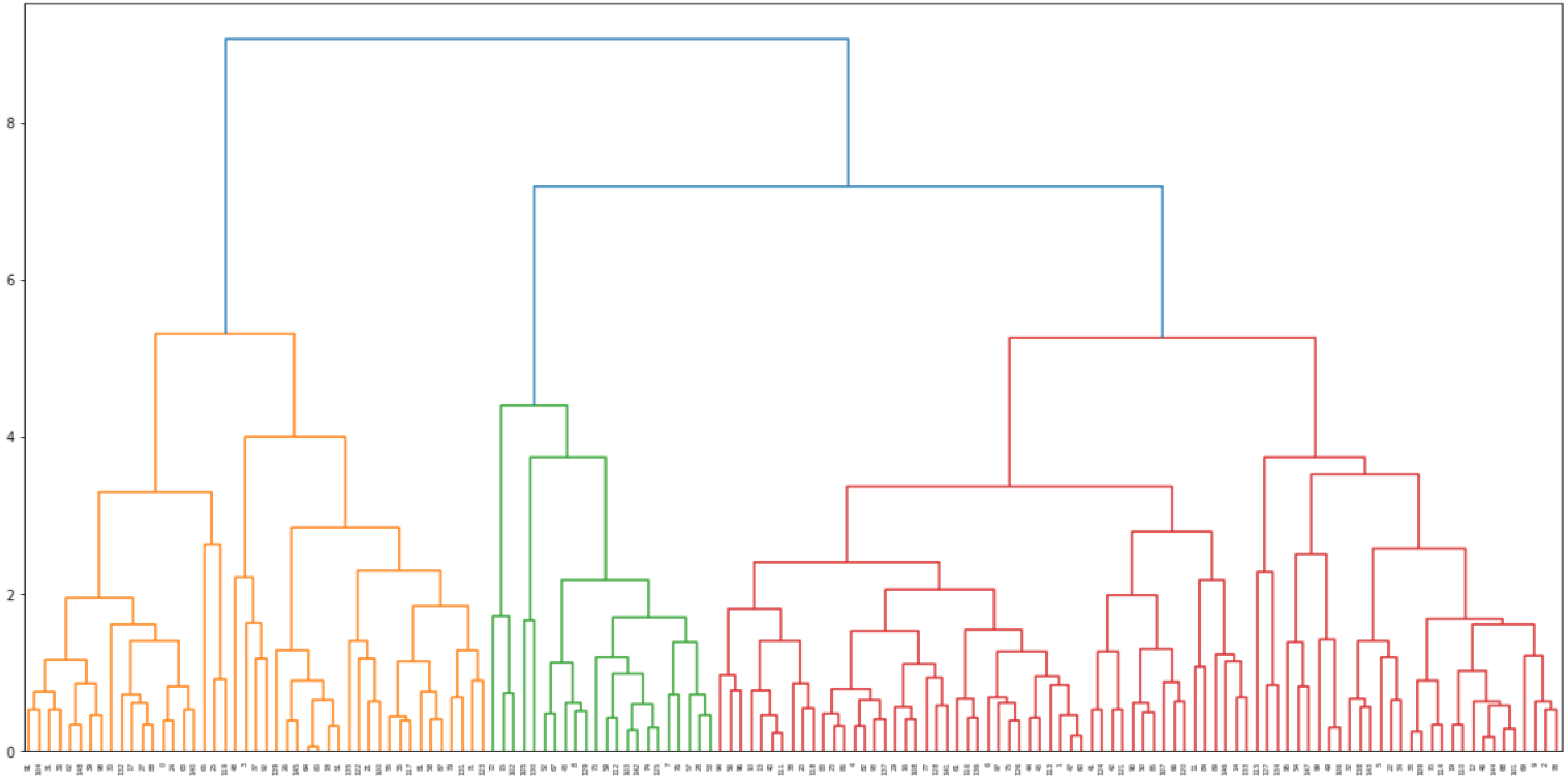
0.4
0.2
0.0
−0.2
−0.4
−0.6
−0.8
PC2_mean
0
1
2
3
ClusterID


PC3_mean
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0
1
2
3
ClusterID


PC4_mean
−0.8
−0.6
−0.4
−0.2
0.0
0.2
0.4
0.6
0.8
0
1
2
3
ClusterID

HIERARCHICAL CLUSTERING

Showing how many clusters can the data be split into



CONCLUSION

The countries that require help the most are listed below:

Afghanistan, Angola, Benin, Botswana, Burkina Faso, Burundi, Cameroon, Central African Republic, Chad, Comoros, Congo, Dem. Rep., Congo, Rep., Cote d'Ivoire, Equatorial Guinea, Eritrea, Gabon, Gambia, Ghana, Guinea, Guinea-Bissau, Haiti, Iraq, Kenya, Lao, Madagascar, Malawi, Mali, Mauritania, Mozambique, Namibia, Niger, Pakistan, Rwanda, Senegal, Sierra Leone, South Africa, Sudan, Tanzania, Timor-Leste, Togo, Uganda, Yemen and Zambia.

These countries have:

- very low rate of net income per person, GDP per capita, average number of years a new born child would live, total health spending and imports of goods and services.
- - very high rate of measurement of the annual growth rate, number of children that would be born and child mortality rate.

It is clear that these countries require very quick aid in terms of money, education and services.