HELP INTERNATIONAL NGO

Clustering the Countries

- By Ezazaehmad Paliza

Abstract

Objective :

We, HELP International humanitarian NGO, committed to fight poverty and provide the people of backward countries with basic amenities and relief during the time of disasters and natural calamities. We run a lot of operational projects from time to time, along with advocacy, drives to raise awareness as well as for funding purposes.

Problem Statement :

During the recent funding programs, we have been able to raise around $ 10 million. As an analyst, we have to come up with the countries list that are in the direst need of aid.

Correlation in the data:

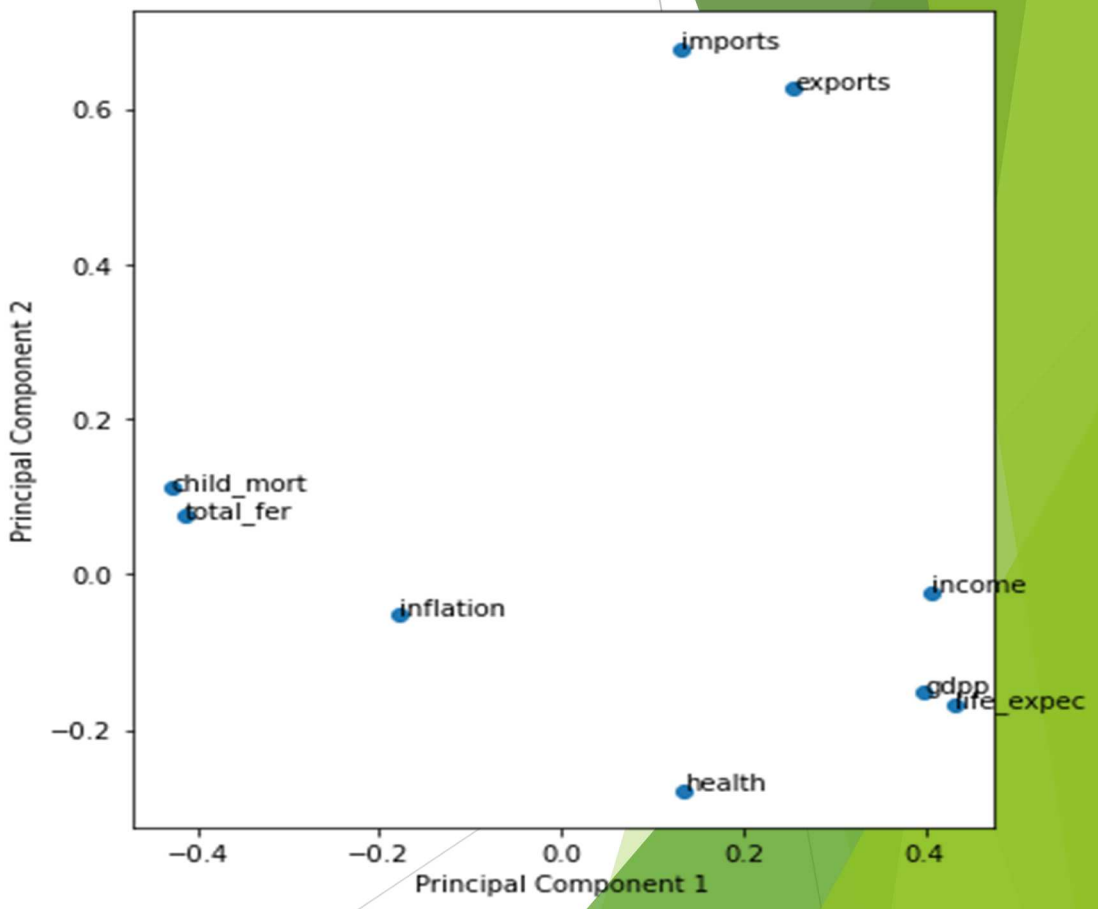
- After data cleaning , we removed outlier from gdpp column because the country with high gdpp would not require any aid as there are already doing good.

- We did standardized scaling to standardize all parameters on cleaned, outlier removed data.

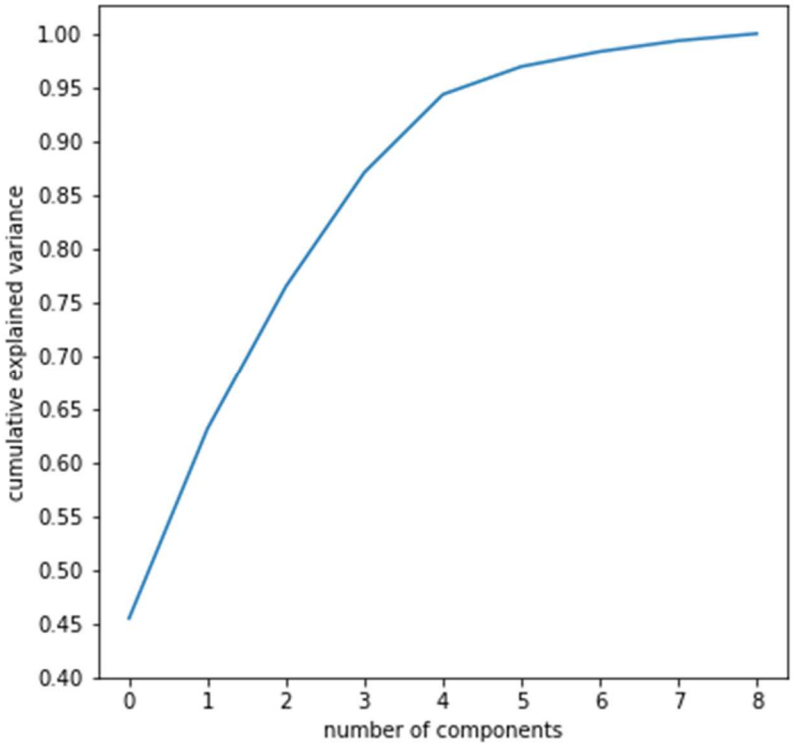
- Looking at the heatmap, we see that few variables like (total fertility, child mortality) , (income , gdpp) and (imports and exports) have high correlation.

PCA

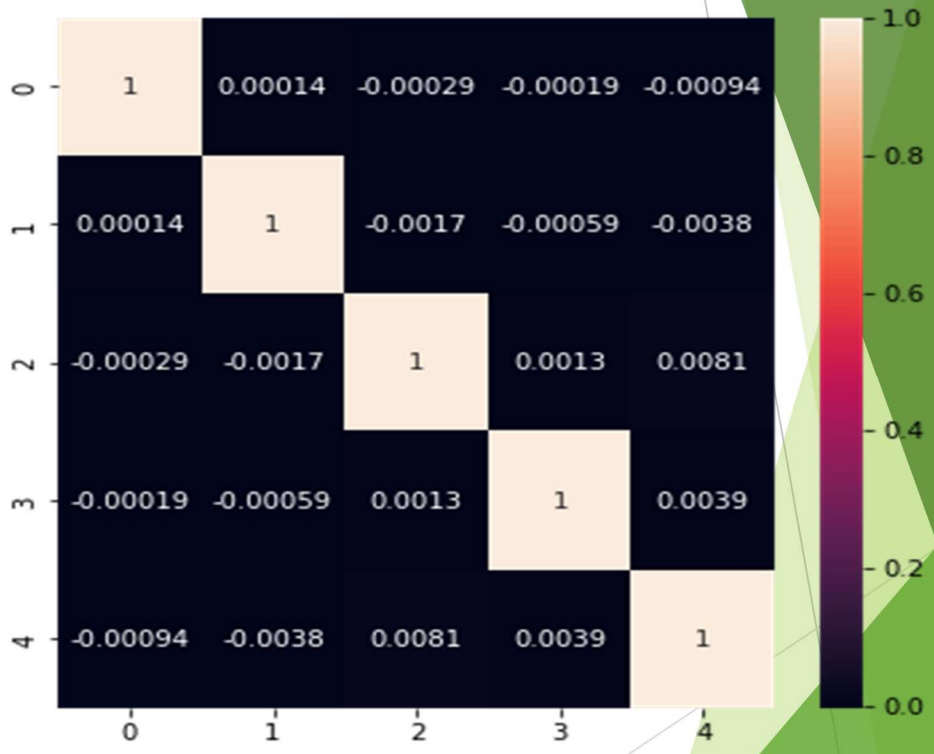
We see that features like gdpp, life expectancy and income are along the direction of PC1 and other features like total fertility and child mortality are along PC2 direction.



PCA

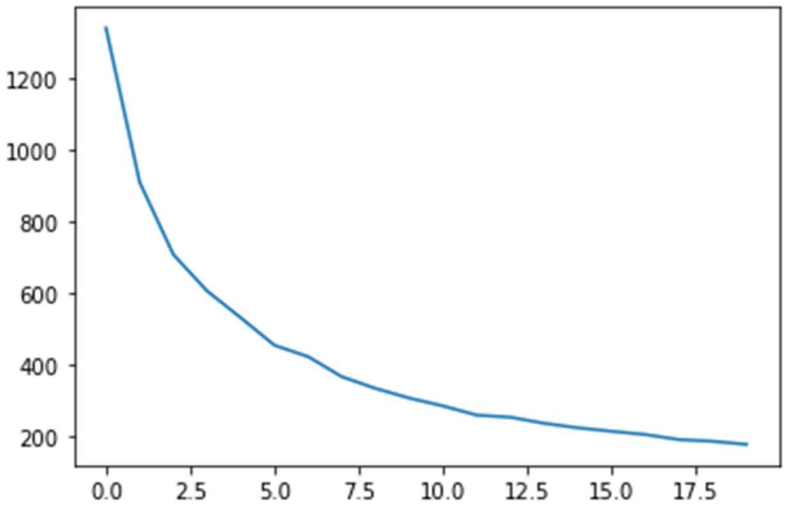


- Looking at the screeplot , we see that 95% of variance in the data is explained by 5 components.

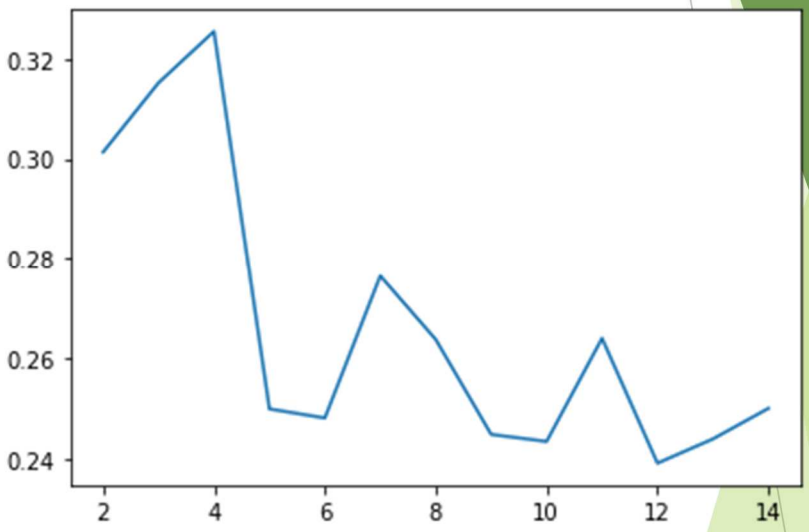


- After doing dimensionality reduction via incremental PCA by taking 5 components, we see that the correlation in the data has almost reduced to zero.

K-Means Clustering



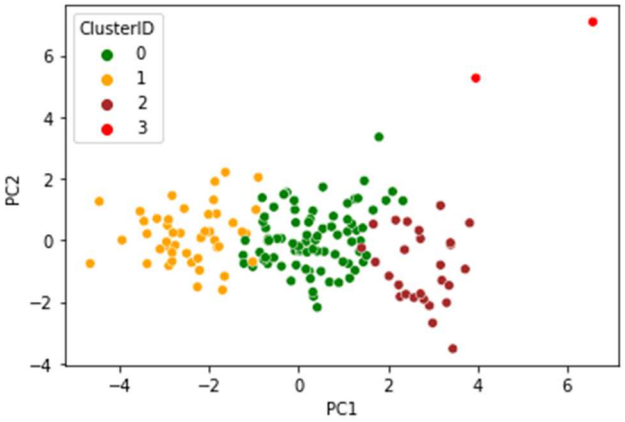
Sum of Squared Distances

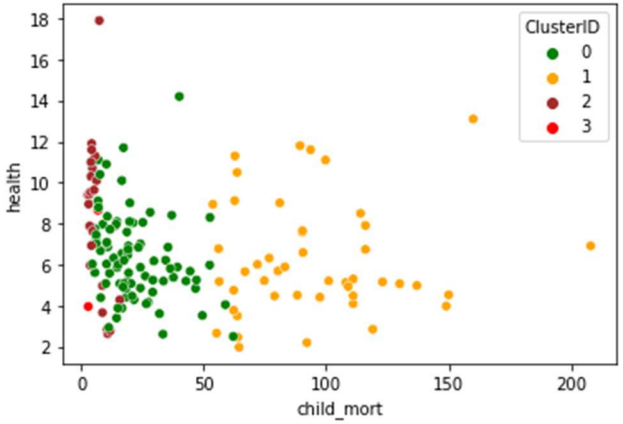


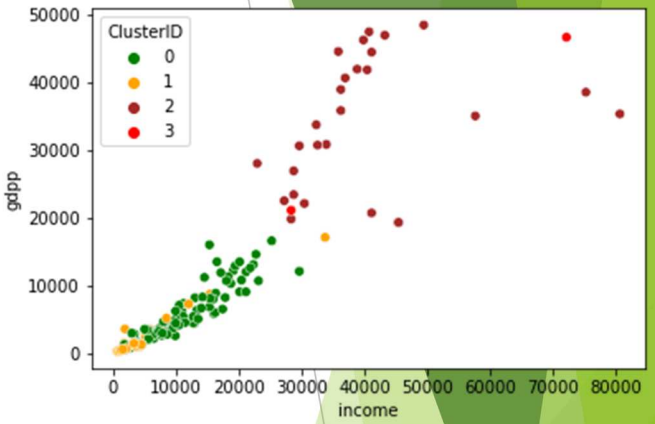
Silhouette Analysis

- By looking silhouette analysis, we see the highest peak is at k =4 and in sum of squared distances graph , we see that the elbow is in the range of 3 to 5 , so we are going ahead with k as 4.

K-means clustering







- Scatter plot of PC1 , PC2 for various clusters. We see the formation of the cluster.

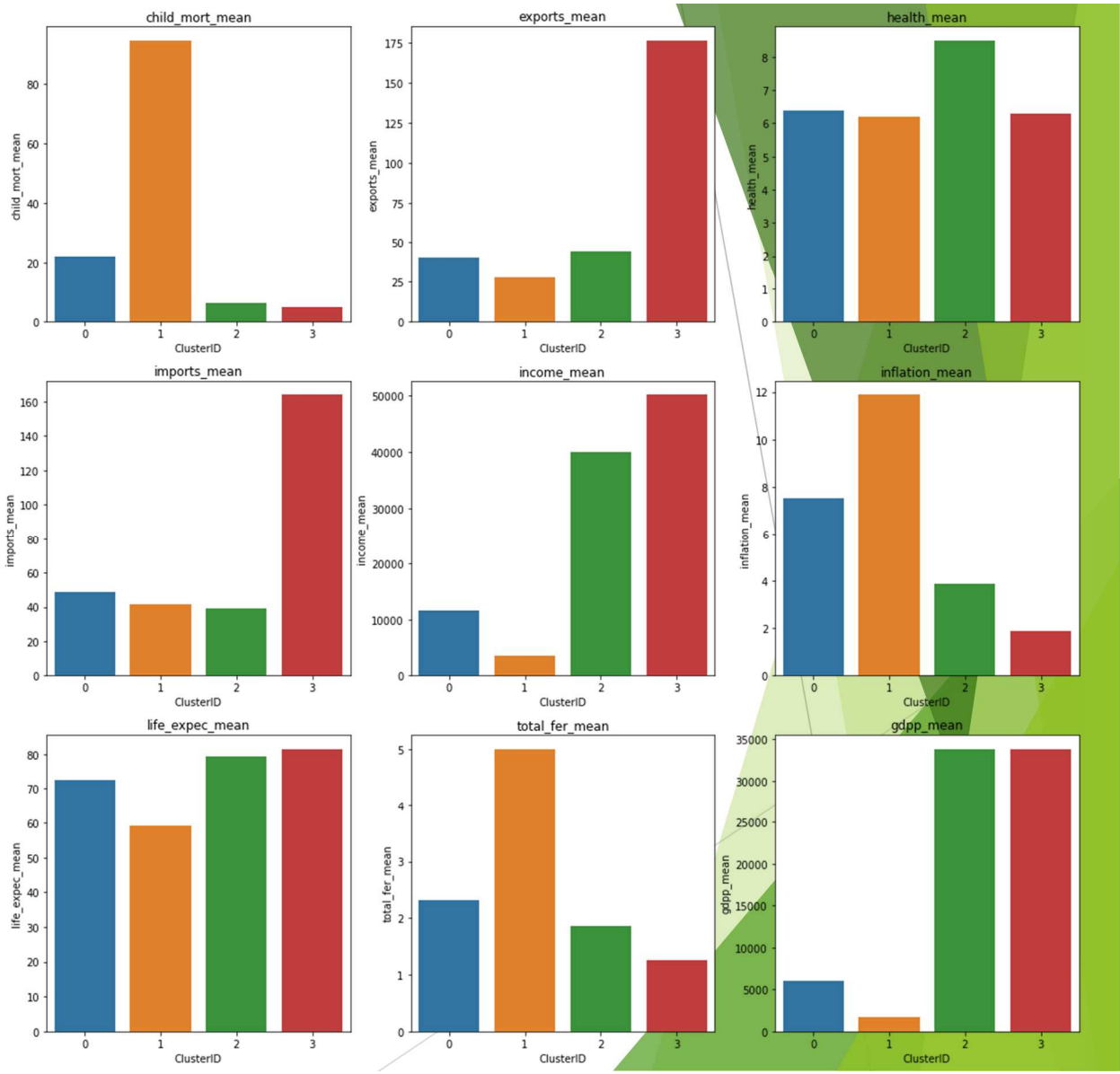
- Scatter plot of health spending, child mortality for various clusters. We see that for cluster 1, the health spending as % of gdp is lower and at the same time child mortality is very high.

- Scatter plot of gdpp , income for various clusters. We see that for cluster 1 , both gdpp and net income per person are very low.

K-means clustering

As per our K- means clusters-
Cluster 1 is area of concern due to :

- Low gdpp
- Low income
- High child mortality
- High inflation
- High total fertility

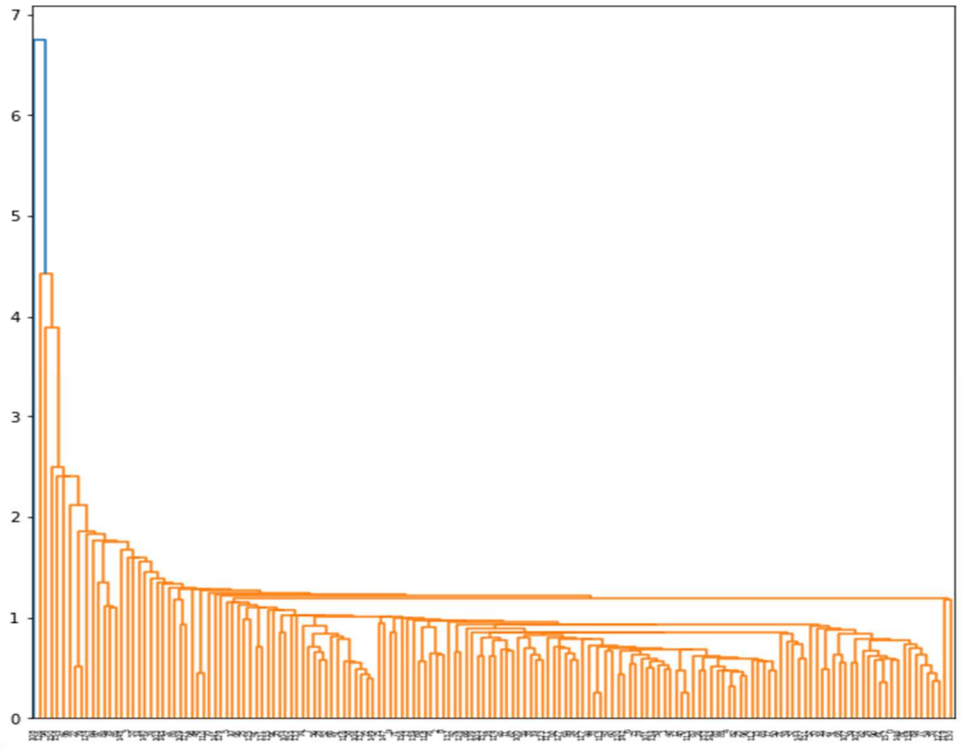


K-means clustering

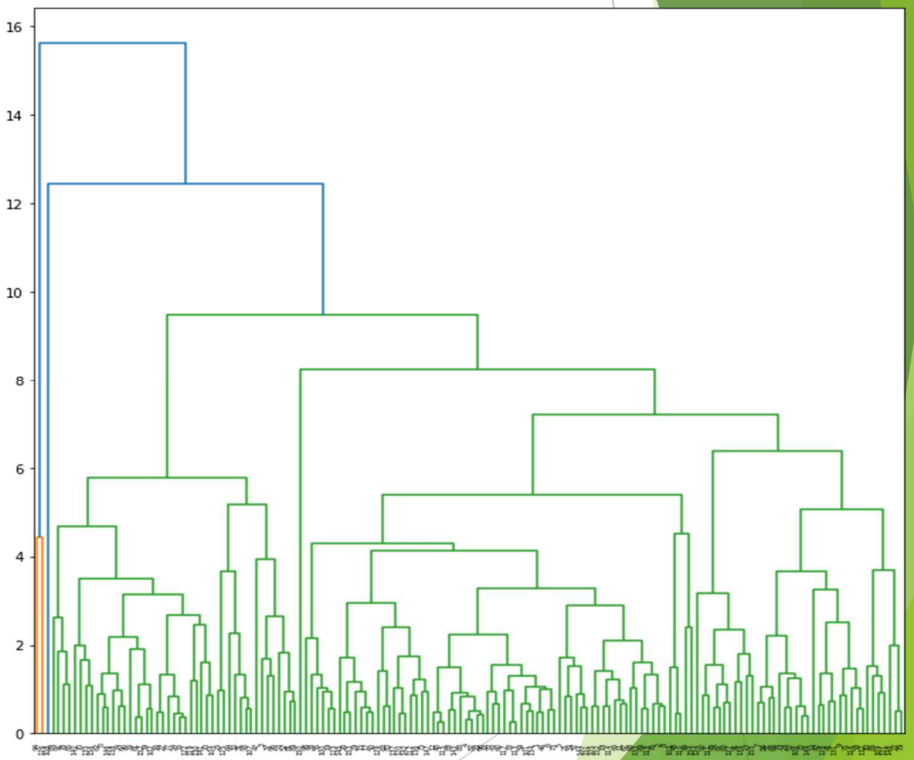
10 countries under cluster 1 are:

1. Burundi
2. Liberia
3. Congo, Dem. Rep
4. Niger
5. Sierra Leone
6. Madagascar
7. Mozambique
8. Central African Republic
9. Malawi
10. Eritrea

Hierarchical Clustering

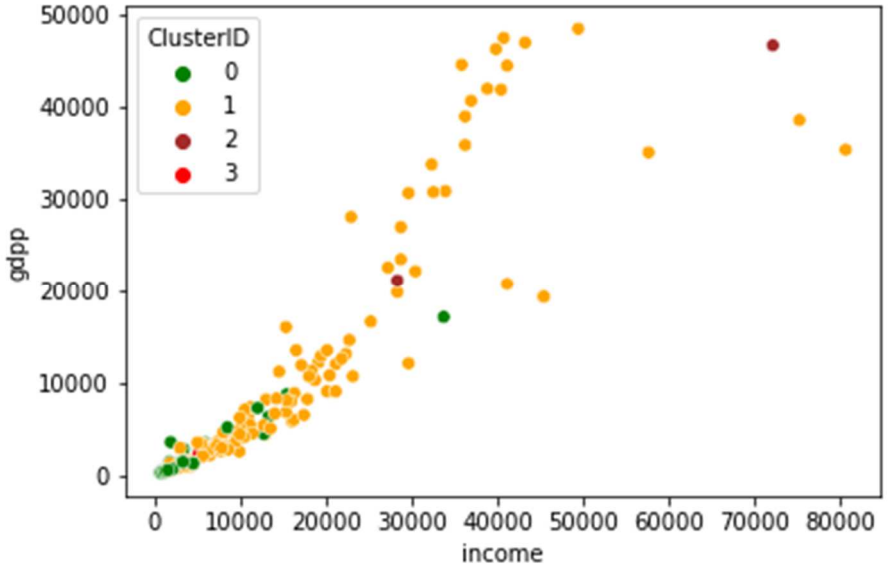


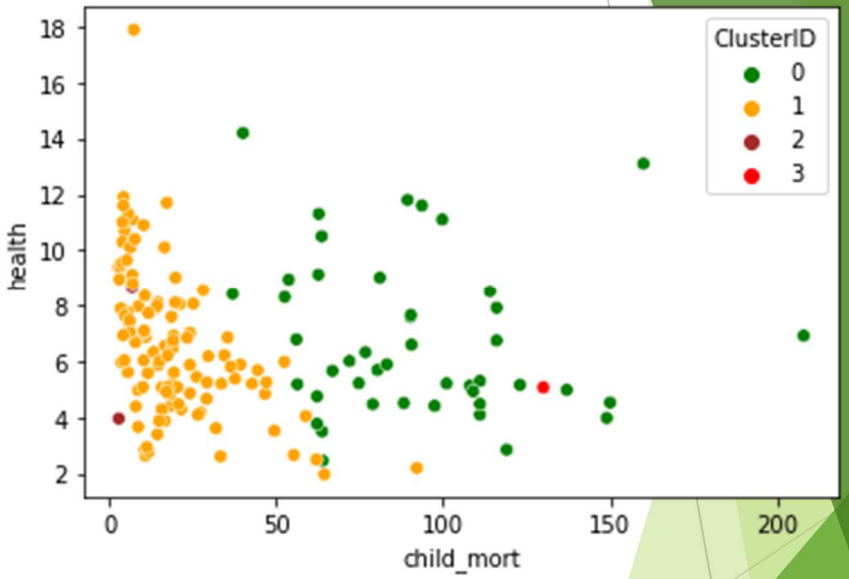
Single method hierarchical clustering



Complete method hierarchical clustering

Hierarchical Clustering





Scatter plot of gdpp , income for various clusters. We see that for cluster 0 , both gdpp and net income per person are very low.

Scatter plot of health spending , child mortality for various clusters. We see that for cluster 0, the health spending as % of gdp of few countries is lower and for those countries -the child mortality is very high.

Hierarchical Clustering

As per our Hierarchical clusters- Cluster 0 is area <u>of concern due to :</u>

- Low gdpp
- Low income
- High child mortality
- High inflation
- High total fertility

<u>10 countries under cluster 0 are:</u>

1. Burundi
2. Liberia
3. Congo, Dem. Rep.
4. Niger
5. Sierra Leone
6. Madagascar
7. Mozambique
8. Central African Republic
9. Malawi
10. Togo

Summary

As by both K means and Hierarchical clustering method - we have got same countries which requires aid.

The following are the countries which are in direst need of aid by considering socio – economic factor into consideration:

1. Burundi
2. Liberia
3. Congo, Dem. Rep.
4. Niger
5. Sierra Leone
6. Madagascar
7. Mozambique
8. Central African Republic
9. Malawi