HUMANITARIAN AID

(APFE19805132)-Saily Shah

CONTENTS

- Problem Statement
- Objectives
- Analysis approach
- Observations
- Recommendations

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.

- Planning the utilisation of these funds, will help achieve the goal of helping the countries in need – strategically as well as effectively.

OBJECTIVE

- Laying out priorities amongst all the countries requiring help, for a need based utilisation of the available funds. Country in direst need coming in at Priority – 1.

- To lay out priorities, various socio–economic factors shall be considered, namely:
  - *Child mortality rate*
  - *Income per capita*
  - *Gdpp of the country*

ANALYSIS APPROACH

- Data understanding and its treatment :
  - *The dataset has the record of 167 countries explaining their socio–economic status.*
  - *The columns such as imports, exports and health are expressed in terms of percentage of total GDP. These columns are converted to absolute values for better understanding.*

- The performance of countries with respect to various socio–economic factors is plotted and relation between these factors is observed. From these relations, it is evident that factors such as child mortality, income & gdpp of a country are some of the key factors for deciding a country as under-developed.

- Principal component analysis is carried out to achieve dimensionality reduction and reduce correlation among the achieved components.

ANALYSIS APPROACH

- Hopkins test is carried out to observe the clustering tendency of the PCA.

- Clustering is done by two popular methods:
  - *Kmeans clustering technique*
  - *Hierarchical clustering technique*

- The clusters formed by both the methods are observed by focusing on 3 socio-economic factore:
  - *Child mortality*
  - *Income per capita*
  - *Gdpp*

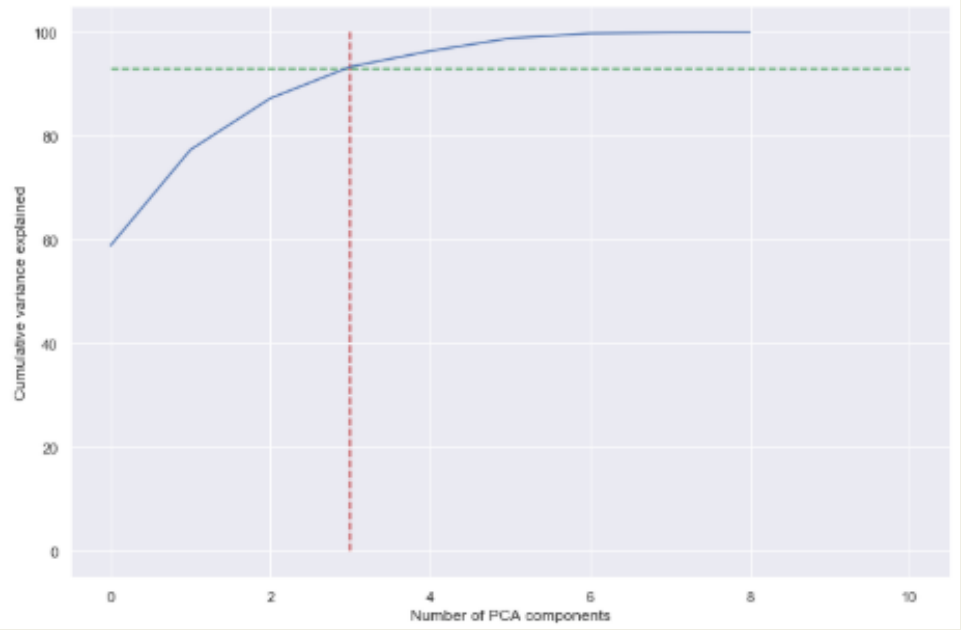
- The clusters obtained by ‘Kmeans’ are clearer and hence used for the study.

OBSERVATIONS

- Principal component analysis
- Hopkins statistics
- Clustering techniques
  a) *Kmeans clustering*
  b) *Hierarchical clustering*
- Results from both the clustering methods

PRINCIPAL COMPONENT ANALYSIS

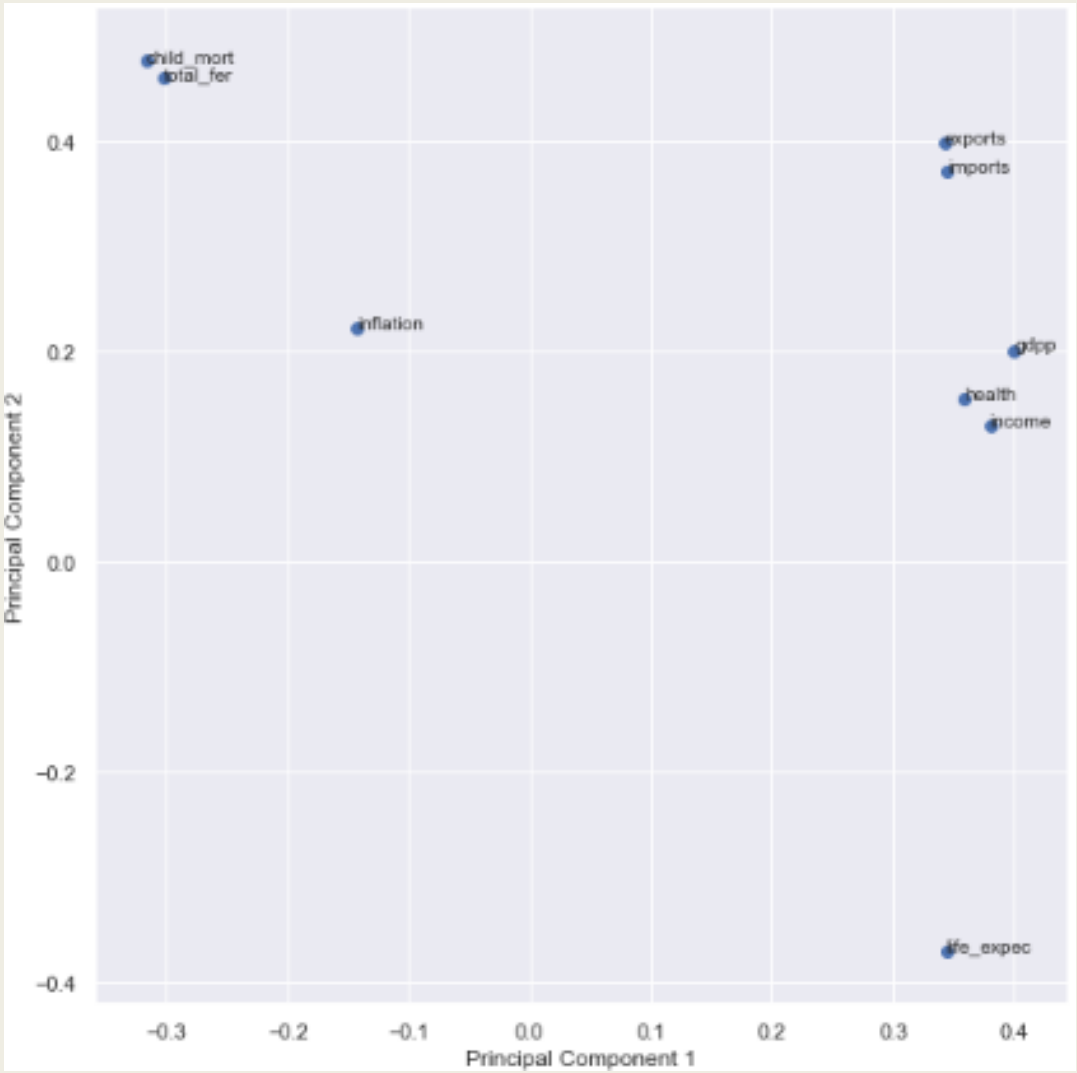
- The first step before applying PCA was to scale and standardise the data so that the variables of differing magnitude fall in the same range.
- A scree plot was made to understand the variance explained.



- From the scree plot it is evident that 3 out of the 8 components are enough to describe 93% of the variance in the dataset. So 3 components are chosen.

PRINCIPAL COMPONENT ANALYSIS

- The first two principal components are plotted against all the socio – economic factors to observe how well they are explained by them.



PRINCIPAL COMPONENT ANALYSIS

- Following observations are made:
  - *Life expectancy, income, gdpp and health are very well explained by PC1.*
  - *Imports and exports are well explained by both the components PC1 and PC2.*
  - *Child mortality , Inflation and total fertility are well explained by PC2.*

HOPKINS STATISTICS

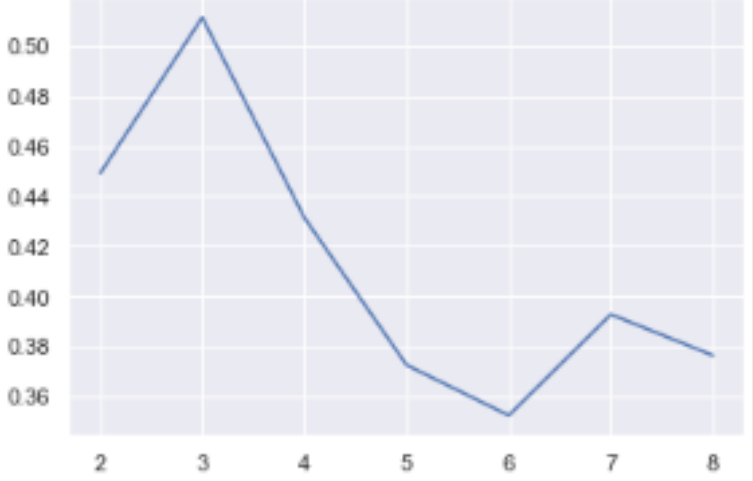
- The Hopkins statistic is a way of measuring the cluster tendency of a data set.
- If this value is close to 1, it indicates that data has high tendency to form clusters.
- Hopkins score achieved was than 0.75, which proved that the treated dataset has good tendency to form clusters.

CLUSTERING TECHNIQUES

Two popular clustering techniques are adopted:

1) Kmeans Clustering : Optimal number of clusters are calculated with help of 2 methods –

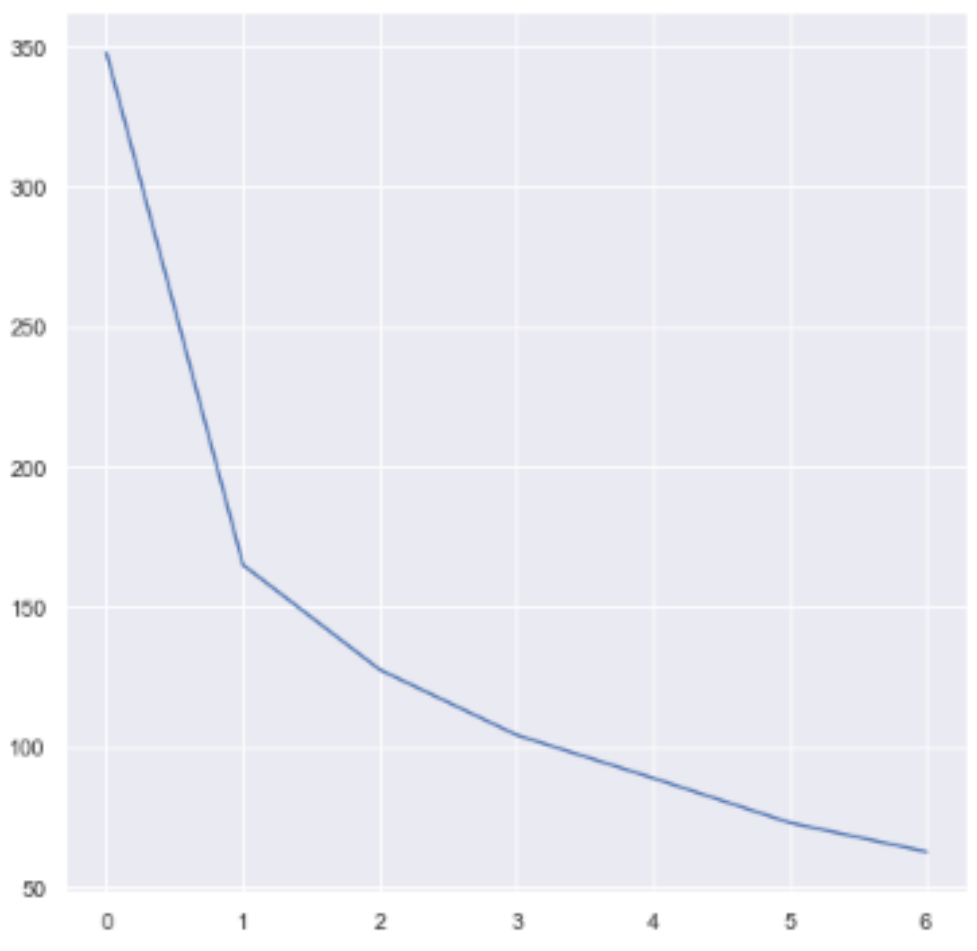
   a) Silhouette curve method



From the graph we can see that with 3 clusters, the silhouette score is higher compared to other number of clusters. The same is verified by elbow curve method.

CLUSTERING TECHNIQUES

b) Elbow curve method

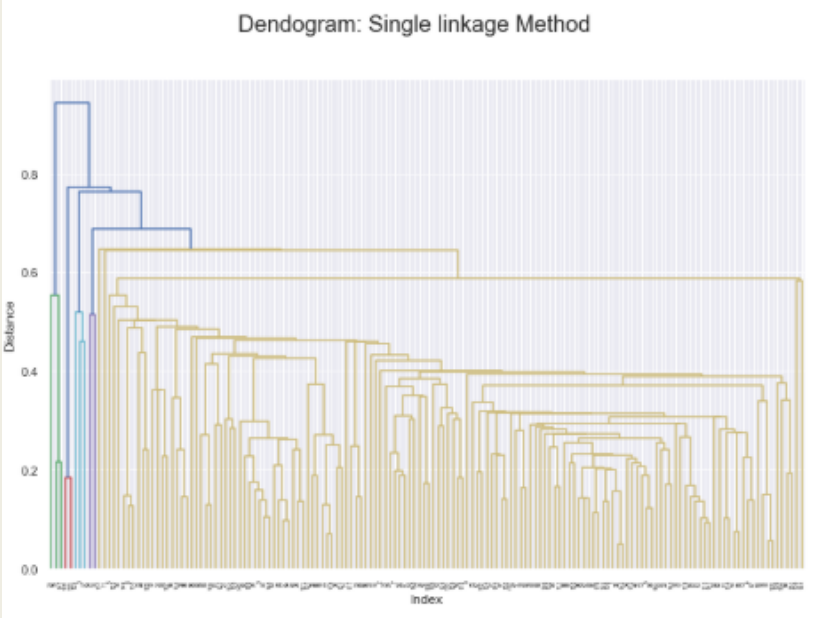


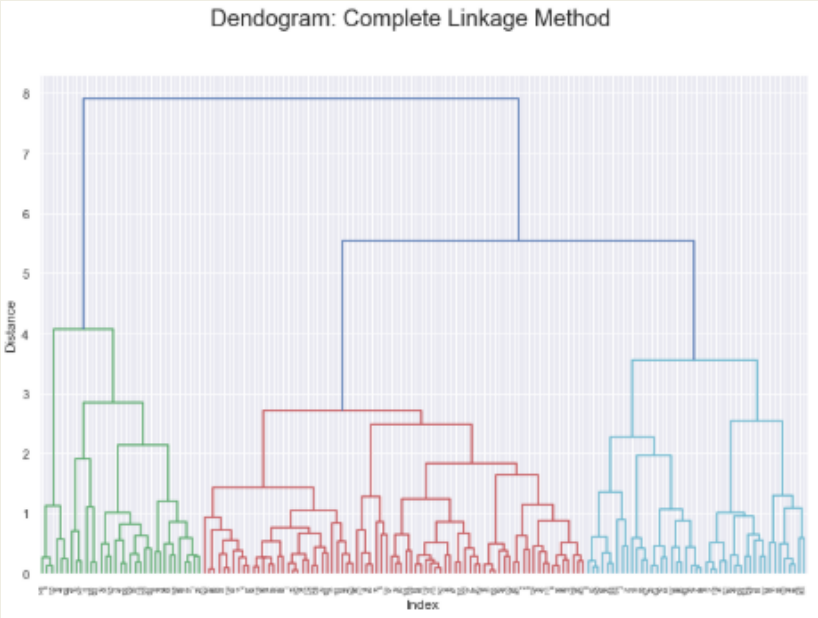
Above elbow curve confirms that either 3 or 4 clusters can be used.

CLUSTERING TECHNIQUES

2) Hierarchical Clustering:

- *Performed two types of linkages namely single linkage and complete linkage*





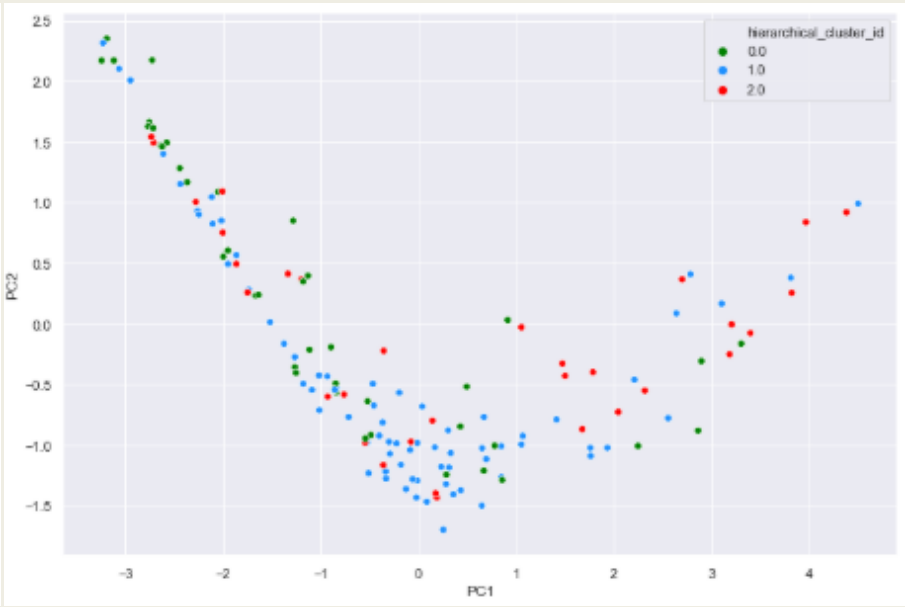
- *We can see that complete linkage tend to yield more balanced, attractive clusters. This dendogram makes much more sense and hence was adopted for the analysis. The dendogram was cut at the height that will yield 3 clusters.*

RESULTS

- Scatter plots are plotted to understand Spread of principal components/variables from PCA against the clusters from Kmeans and hierarchical method:

a) Kmeans Clustering

b) Hierarchical Clustering

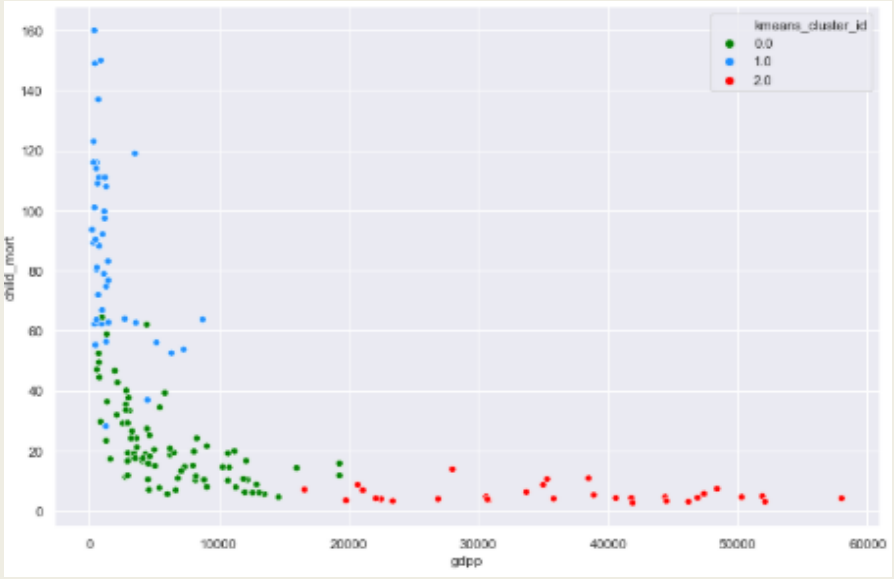


- Clusters obtained from K-Means method seem to be more clear than that of hierarchical method

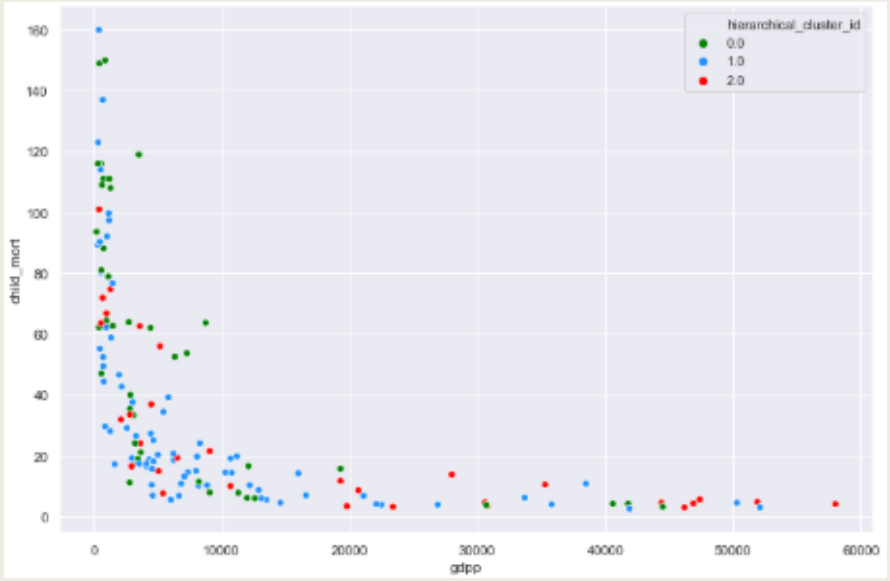
RESULTS

- Plot on original attributes to understand Spread of Principal components:

a) Kmeans Clustering



b) Hierarchical Clustering



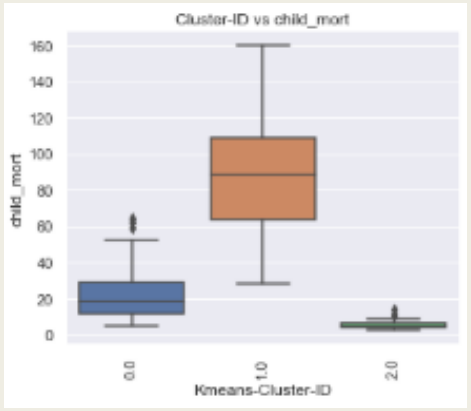
- From the above plots, we can see that the original attributes observe some relation/ pattern when plotted against Kmeans clusters. No pattern is seen when they are observed against Hierarchical clusters.

RESULTS

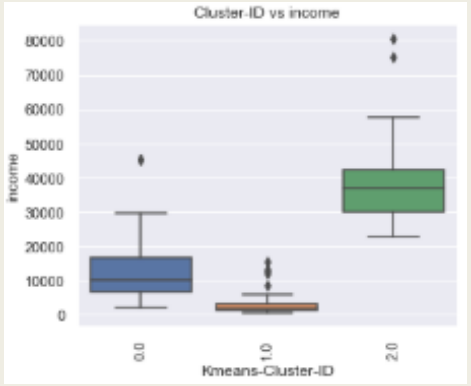
Plot of Cluster-IDs against some socio-economic factors:

a) Kmeans Clustering

- Cluster-ID Vs. Child mortality

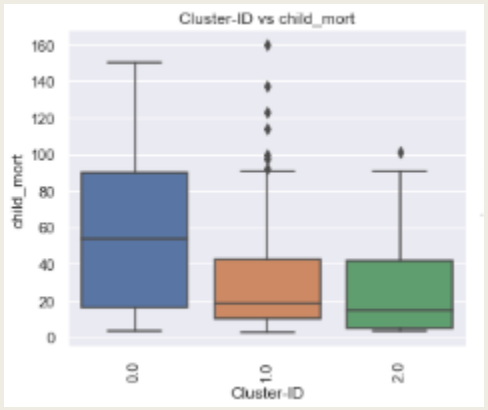


- Cluster-ID Vs. Income

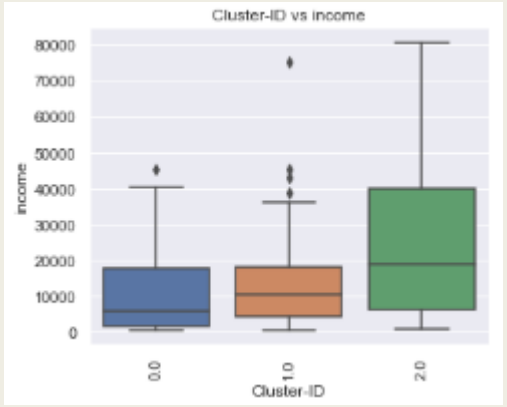


b) Hierarchical Clustering

- Cluster-ID Vs. Child mortality



- Cluster-ID Vs. Income

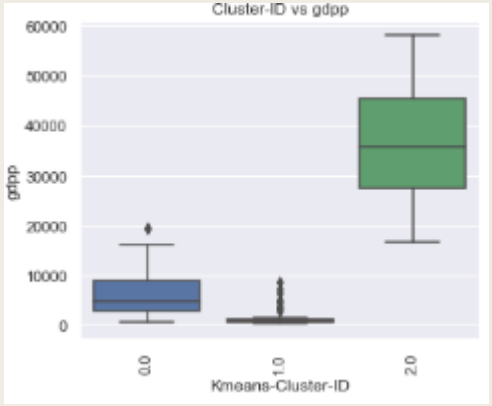


RESULTS

Plot of Cluster-IDs against some socio-economic factors:

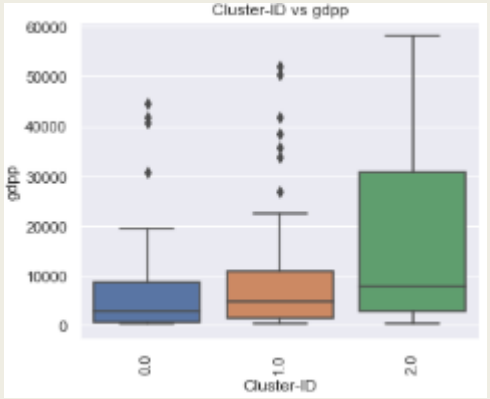
a) Kmeans Clustering

- Cluster-ID Vs. Gdpp



b) Hierarchical Clustering

- Cluster-ID Vs. Gdpp



RESULTS

From the above bxoplots, we can infer following things:

- In Kmeans clustering, the countries lying in cluster-1 have high mortality rate, less income per capita and low gdpp as compared other countries.

- Similarly in hierarchical clustering, countries lying in cluster 0 have high mortality rate, less income per capita and low gdpp as compared other countries. These countries fall under the category of under developed countries.

RECOMMENDATION

- On the basis of extensive research, following countries have shortlisted whose child mortality rate is high and whose income is very low.
- The list has been prioritized based on high mortality rate and income per capita less than 2000, Sierra Leone being the country having maximum child mortality rate and very less per capita income.

1.	Sierra Leone
2.	Chad
3.	Central African Republic
4.	Mali
5.	Niger
6.	Burkina Faso
7.	Congo, Dem. Rep.
8.	Guinea-Bissau
9.	Benin
10.	Guinea

11.	Mozambique
12.	Burundi
13.	Malawi
14.	Togo
15.	Afghanistan
16.	Liberia
17.	Comoros
18.	Uganda
19.	Gambia
20.	Rwanda