HELP INTERNATIONAL

AN ANALYSIS TO SHORTLIST BACKWARD COUNTRIES TO HELP WITH BASIC AMENITIES AND RELIEF

Chandra Shekhar Banerjee

OBJECTIVE AND APPROACH

- Objective:

The objective of the above analysis was to cluster or make a group similar countries which are in need of help. Figure out a group of countries where the accumulated monetary resource can be utilized in terms of fighting poverty and providing with basic amenities and relief during the time of disasters and natural calamities

- Approach:

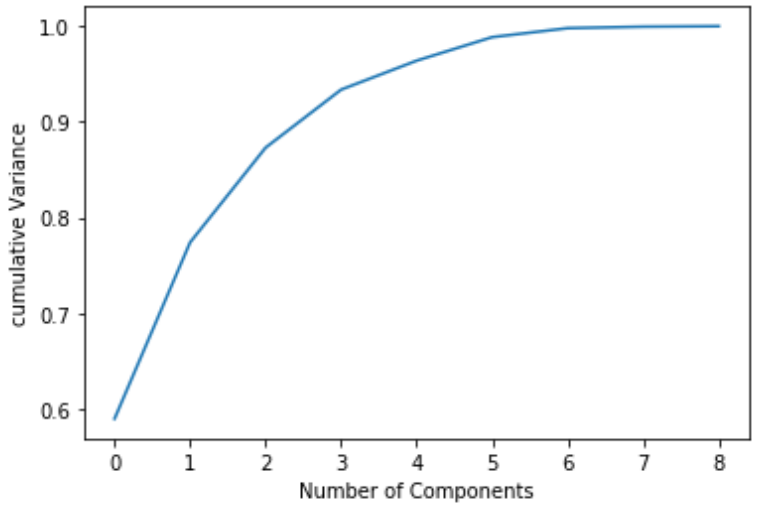
To conduct this analysis we have performed the Cluster Analysis to do the same in 2 ways.

  - K-means Clustering (Where k is considered as 3 as well as 5)
  - Hierarchical Clustering (Where number of clusters are considered as 2)

DIMENSIONALITY REDUCTION WITH HELP OF PCA

- **Principal component analysis** (PCA) is one of the most commonly used dimensionality reduction techniques in the industry. By converting large data sets into smaller ones containing fewer variables. It helps generally model performance and visualizing complex dataset
  - By utilizing this technique and with help of Scree Plot, we were able to reduce

from 9 features to 3 Principle Components. However, prior to conduct this Step we had to perform standardization of the data or bring the features of the dataset into a similar scale

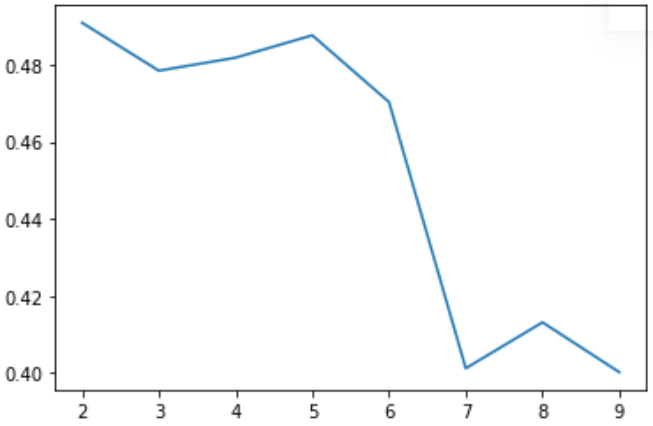


K MEANS CLUSTERING

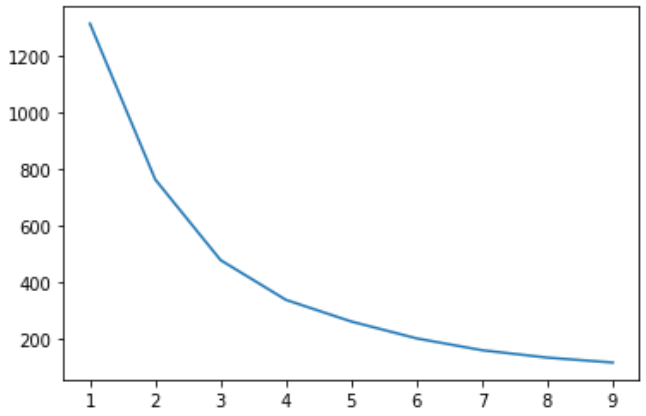
K Means clustering is a well known unsupervised learning technique in the industry

- After completing PCA we consider Outlier treatment. However, we do not exclude any country here so that we do not loose any of them who are actual in need
- We check Hopkins Score to check if the dataset is ready for the K Means Clustering or not
- We check Silhouette Score test for number of Cluster should be considered. Here we considered 3 and 5 for initial phase
- We consider Elbow Curve to recheck the above point

Silhouette Curve



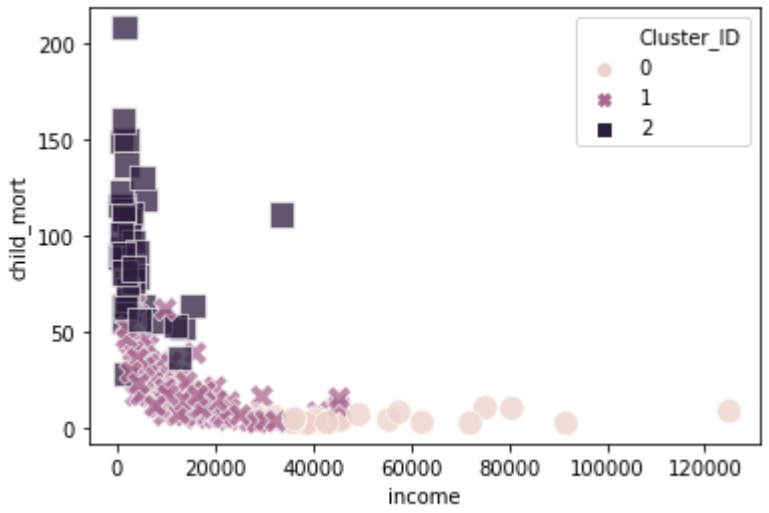
Elbow Curve



K MEANS CLUSTERING [K=3]

- With k=3, we get 3 set of clusters which varies from each other. When we merge our original data set with our clustering dataset and visualize it, we get 3 groups of countries as mentioned in the picture:

Income vs. child mortality



gdpp vs. child mortality

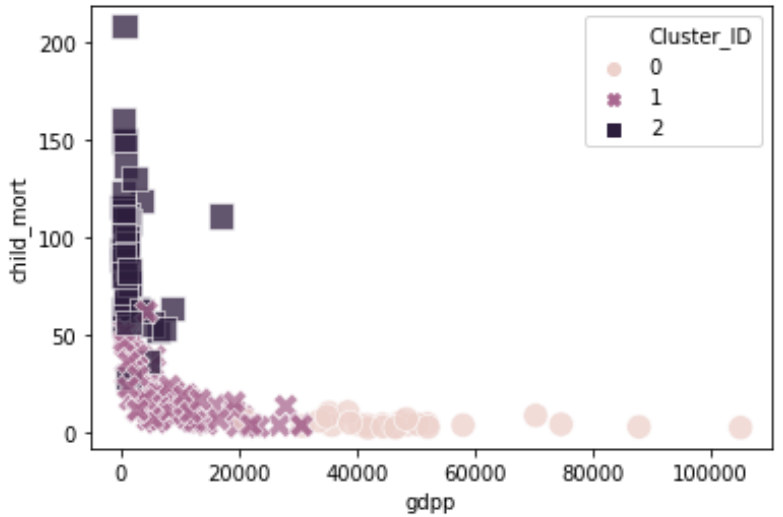


Table:1

Cluster ID	Count
1	92
2	48
0	27

Table:2

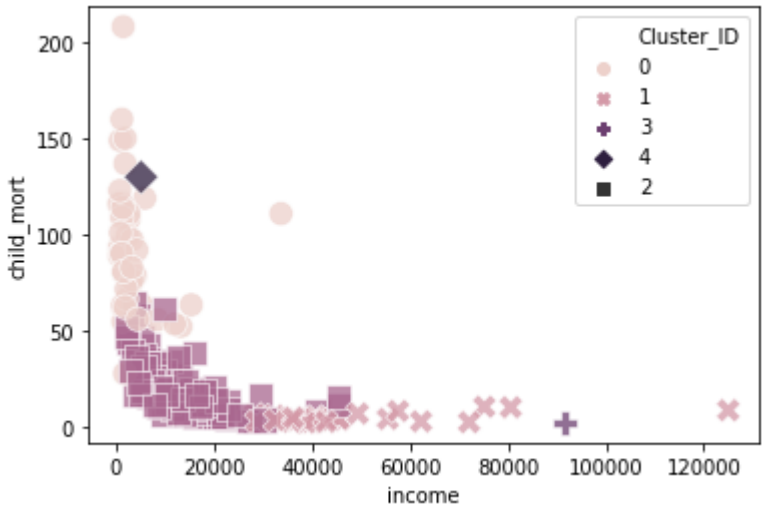
Cluster_ID	gdpp	child_mort	income
0	48759.259259	5.092593	50833.333333
1	8226.869565	20.177174	14169.456522
2	1909.208333	91.610417	3897.354167

In the above 2 scatter plots, each Cluster ID represents a country from the dataset. Table:1 represents number of countries in each cluster and Table:2 represents mean of gdpp, income and child mortality of them. Hence, it can be seen that Cluster ID 2 is those countries where gdpp and income is low but child mortality is very high

K MEANS CLUSTERING [K=5]

- With k=5, we get 5 set of clusters which varies from each other. When we merge our original data set with our clustering dataset and visualize it, we get 5 groups of countries as mentioned in the picture:

Income vs. child mortality



gdpp vs. child mortality

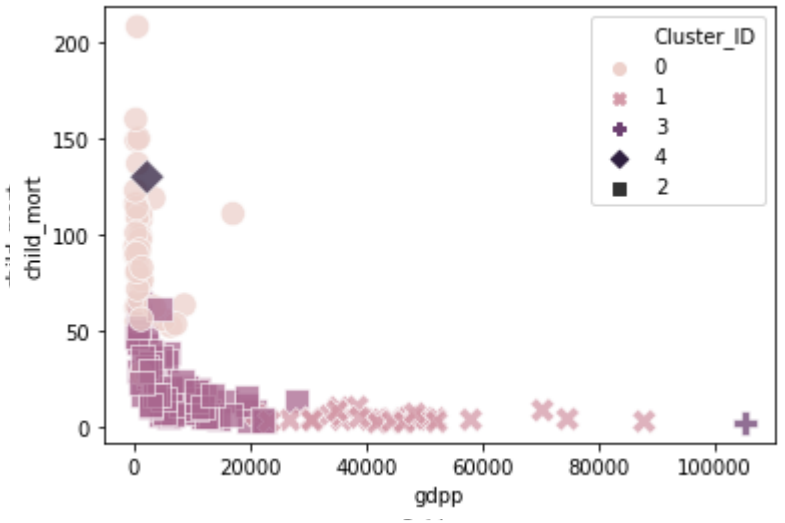


Table:1

Cluster ID	Count
2	89
0	46
1	30
4	1
3	1

Table:2

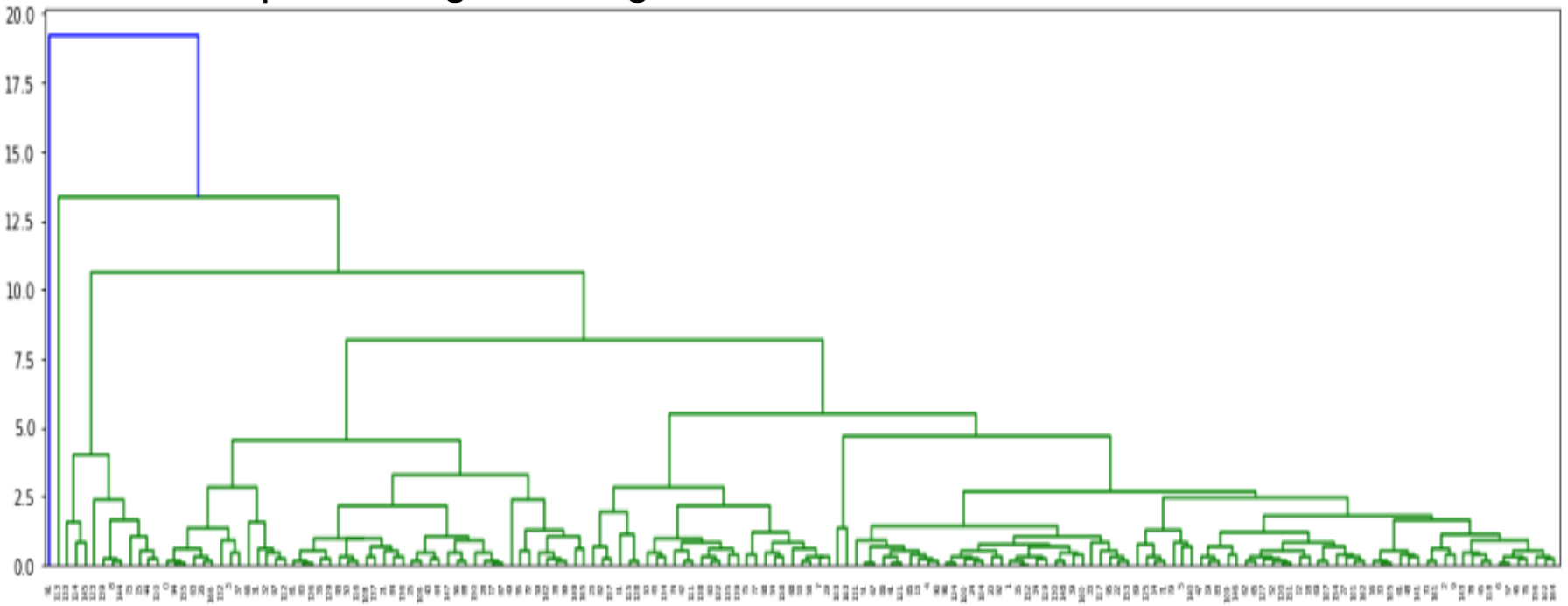
Cluster_ID	gdpp	child_mort	income
0	1843.739130	91.965217	3678.760870
1	44103.333333	5.006667	46676.666667
2	7300.808989	21.097753	13447.078652
3	105000.000000	2.800000	91700.000000
4	2330.000000	130.000000	5150.000000

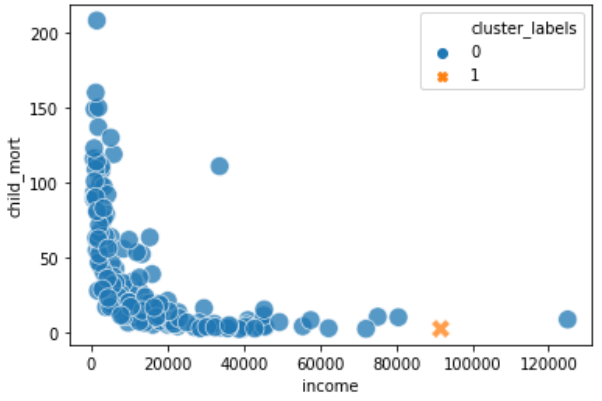
In the above 2 scatter plots, each Cluster ID represents a country from the dataset. Table:1 represents number of countries in each cluster and Table:2 represents mean of gdpp, income and child mortality of them. Hence, it can be seen that Cluster ID 0 (along with id 4) is those countries where gdpp and income is low but child mortality is very high

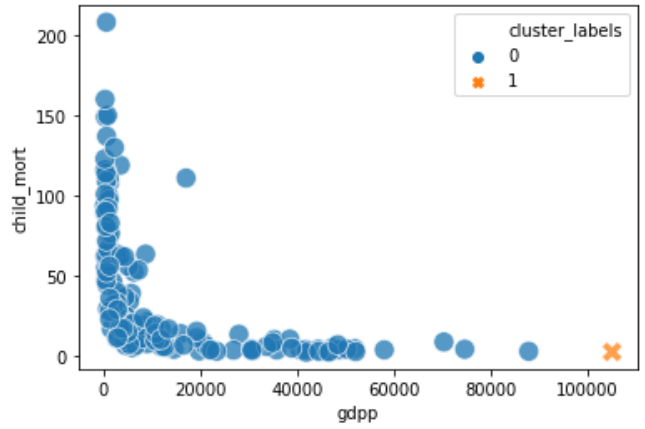
HIERARCHICAL CLUSTERING

- We also conducted the same analysis with the help of Hierarchical Clustering Algorithm. However, we did not received much informative output from this analysis as almost all the countries are falling into same cluster

Complete Linkage Dendrogram







FINAL OUTPUT : RECOMMENDATION

- Hence from the K Means cluster analysis [k=3], we finally get total 48 countries which are recommended. These are the countries with low gdpp , low income and high child mortality. Below are the list of countries mentioned sorted by gdpp in ascending order:
  - Burundi, Liberia, Congo, Dem. Rep., Niger, Sierra Leone, Madagascar, Mozambique, Central African Republic, Malawi, Eritrea, Togo, Guinea-Bissau, Afghanistan, Gambia, Rwanda, Burkina Faso, Uganda, Guinea, Haiti, Tanzania, Mali, Benin, Comoros, Chad, Kenya, Senegal, Pakistan, Lao, Lesotho, Mauritania, Cote dIvoire, Solomon Islands, Yemen, Cameroon, Ghana, Zambia, Sudan, Kiribati, Nigeria, Congo, Rep., Angola, Timor-Leste, Iraq, Namibia, Botswana, South Africa, Gabon, Equatorial Guinea

Thank You