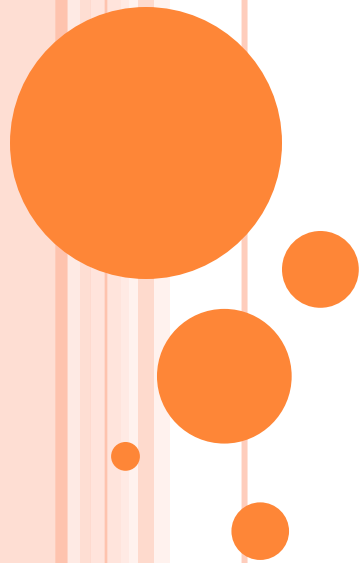# CLUSTERING OF COUNTRIES

By: Seema Rani

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 programmes, 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.
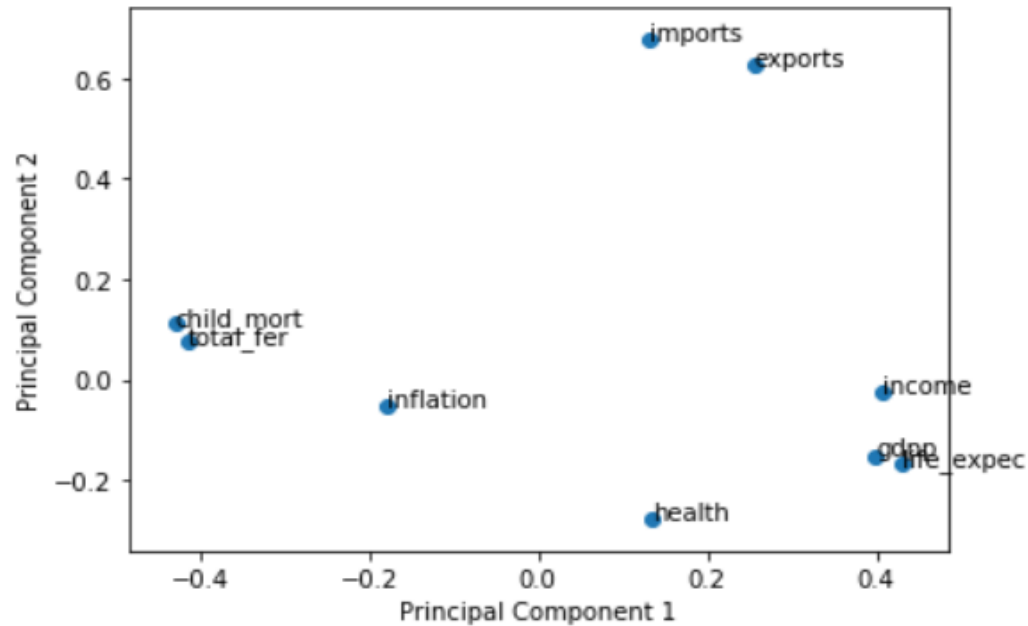
Correlation in the data

There is high correlation between total_fer & Child_mortality , Gdpp & income and in between exports & imports.
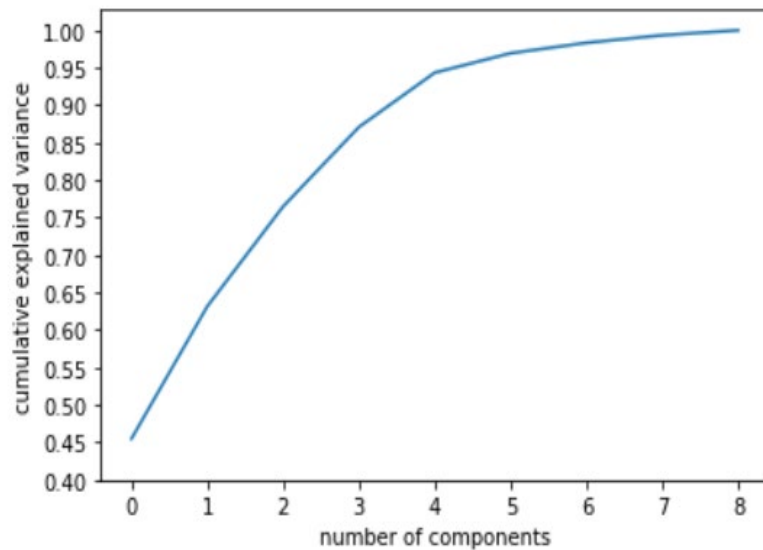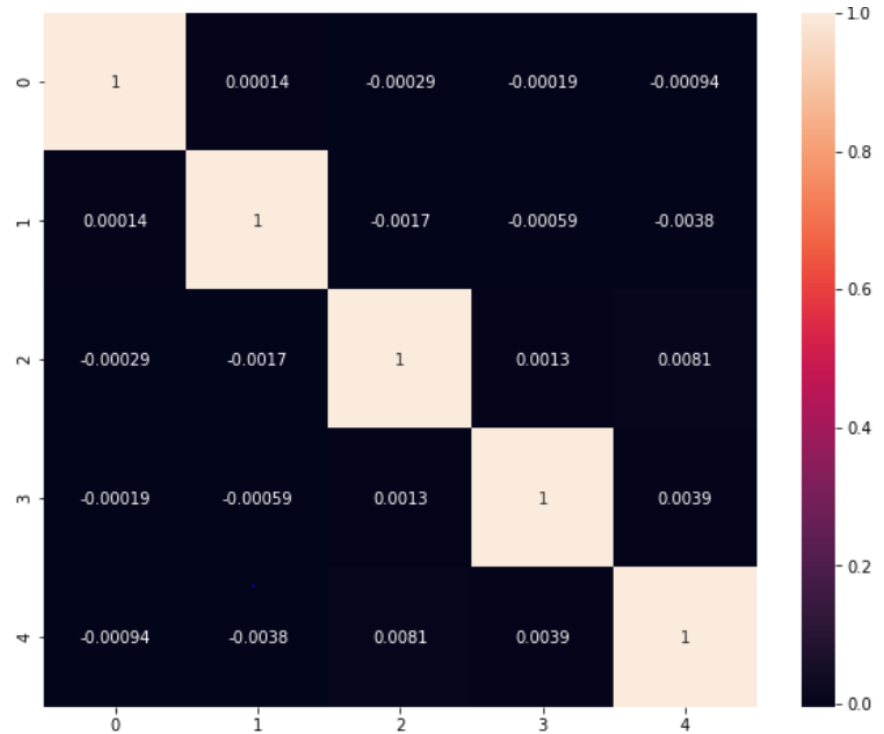
PRINCIPAL COMPONENT

Visualise the feature taking PC1 along PC2

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.

In Screeplot, 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
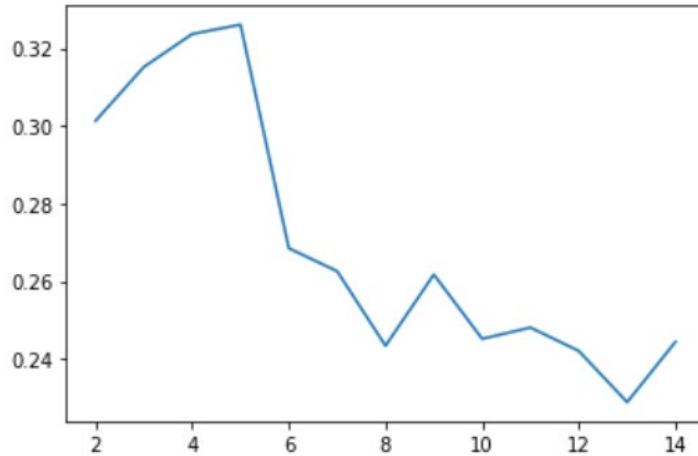
Silhouette Analysis

Sum of Squared Distances

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

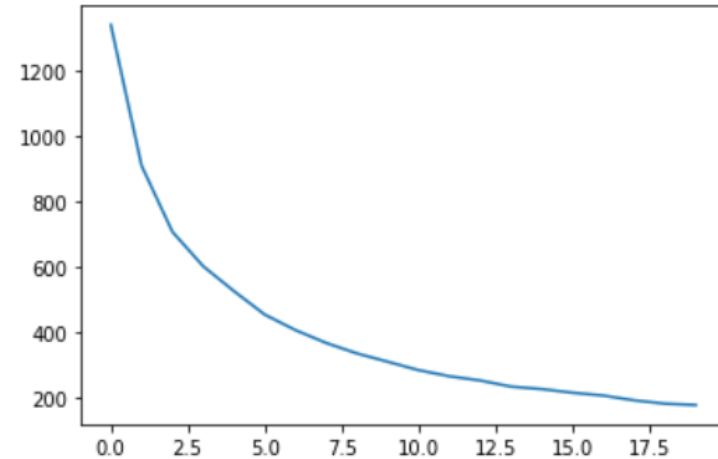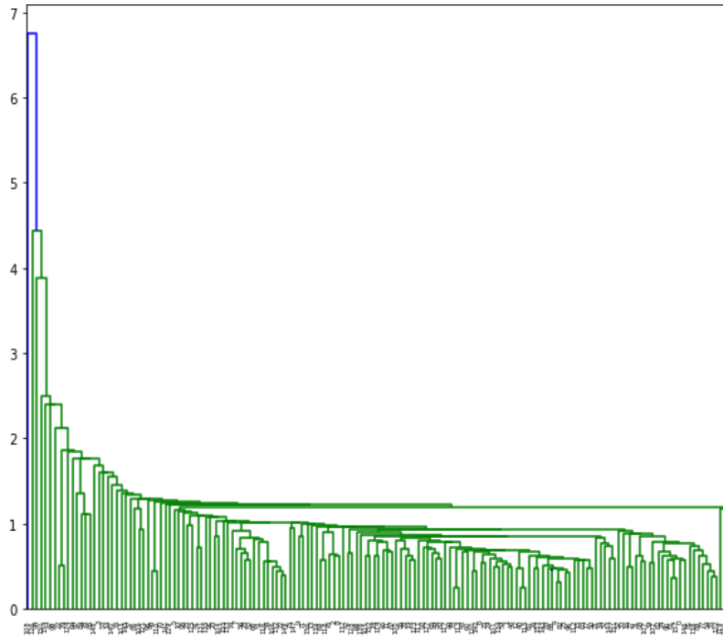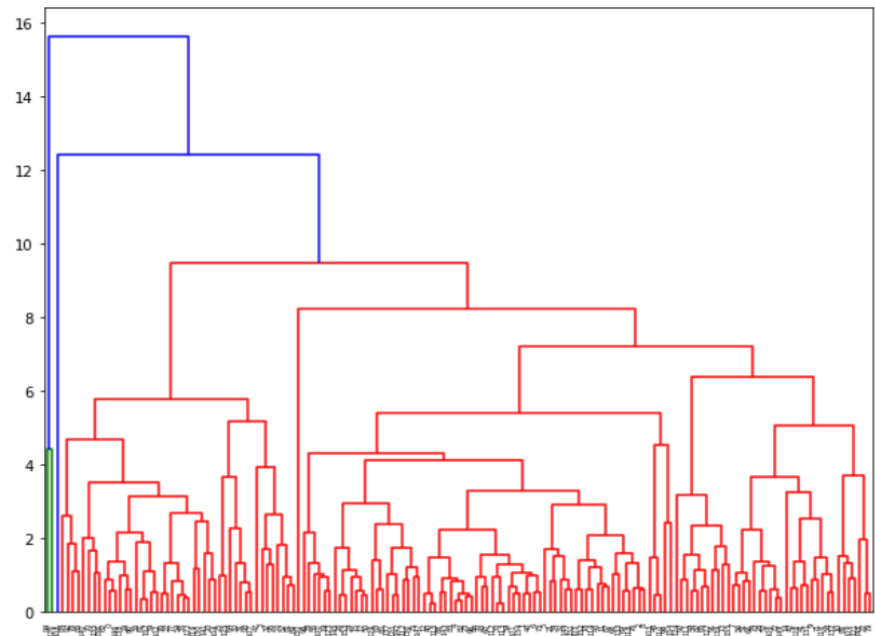| country	child_mort	exports	health	imports	income	inflation	life_expec	total_fer	gdpp	ClusterID
Burundi	93.6	8.92	11.60	39.2	764	12.30	57.7	6.26	231	1
Liberia	89.3	19.10	11.80	92.6	700	5.47	60.8	5.02	327	1
Congo, Dem. Rep.	116.0	41.10	7.91	49.6	609	20.80	57.5	6.54	334	1
Niger	123.0	22.20	5.16	49.1	814	2.55	58.8	7.49	348	1
Sierra Leone	160.0	16.80	13.10	34.5	1220	17.20	55.0	5.20	399	1
Madagascar	62.2	25.00	3.77	43.0	1390	8.79	60.8	4.60	413	1
Mozambique	101.0	31.50	5.21	46.2	918	7.64	54.5	5.56	419	1
Central African Republic	149.0	11.80	3.98	26.5	888	2.01	47.5	5.21	446	1
Malawi	90.5	22.80	6.59	34.9	1030	12.10	53.1	5.31	459	1
Eritrea	55.2	4.79	2.66	23.3	1420	11.60	61.7	4.61	482	1

10 Countries come under cluster 1.

HEIRARCHICAL CLUSTERING

Single method

Complete Method
As Single method is not clear.
By looking at this dendrogram will
take n- cluster as 4

HEIRARCHICAL CLUSTERING

country	child_mort	exports	health	imports	income	inflation	life_expec	total_fer	gdpp	ClusterID
Burundi	93.6	8.92	11.60	39.2	764	12.30	57.7	6.26	231	0
Liberia	89.3	19.10	11.80	92.6	700	5.47	60.8	5.02	327	0
Congo, Dem. Rep.	116.0	41.10	7.91	49.6	609	20.80	57.5	6.54	334	0
Niger	123.0	22.20	5.16	49.1	814	2.55	58.8	7.49	348	0
Sierra Leone	160.0	16.80	13.10	34.5	1220	17.20	55.0	5.20	399	0
Madagascar	62.2	25.00	3.77	43.0	1390	8.79	60.8	4.60	413	0
Mozambique	101.0	31.50	5.21	46.2	918	7.64	54.5	5.56	419	0
entral African Republic	149.0	11.80	3.98	26.5	888	2.01	47.5	5.21	446	0
Malawi	90.5	22.80	6.59	34.9	1030	12.10	53.1	5.31	459	0
Togo	90.3	40.20	7.65	57.3	1210	1.18	58.7	4.87	488	0

10 Countries comes under cluster 0.

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:

country	child_mort	exports	health	imports	income	inflation	life_expec	total_fer	gdpp	ClusterID
Burundi	93.6	8.92	11.60	39.2	764	12.30	57.7	6.26	231	0
Liberia	89.3	19.10	11.80	92.6	700	5.47	60.8	5.02	327	0
Congo, Dem. Rep.	116.0	41.10	7.91	49.6	609	20.80	57.5	6.54	334	0
Niger	123.0	22.20	5.16	49.1	814	2.55	58.8	7.49	348	0
Sierra Leone	160.0	16.80	13.10	34.5	1220	17.20	55.0	5.20	399	0
Madagascar	62.2	25.00	3.77	43.0	1390	8.79	60.8	4.60	413	0
Mozambique	101.0	31.50	5.21	46.2	918	7.64	54.5	5.56	419	0
entral African Republic	149.0	11.80	3.98	26.5	888	2.01	47.5	5.21	446	0
Malawi	90.5	22.80	6.59	34.9	1030	12.10	53.1	5.31	459	0
Togo	90.3	40.20	7.65	57.3	1210	1.18	58.7	4.87	488	0