NGO PCA- Clustering Analysis

Name:

1. Alok Kumar Singh

Abstract

HELP International is an international humanitarian NGO has been able to raise around $ 10 million as their fund.

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.

Business Objective –

- The aim of analysis is to categories the countries using some socio-economic and health factors that determine the overall development of the country.
- Provide the suggestion to CEO, the countries which he/she needs to focus on the most.

Analysis approach

1- Use the PCA to get all the principal component

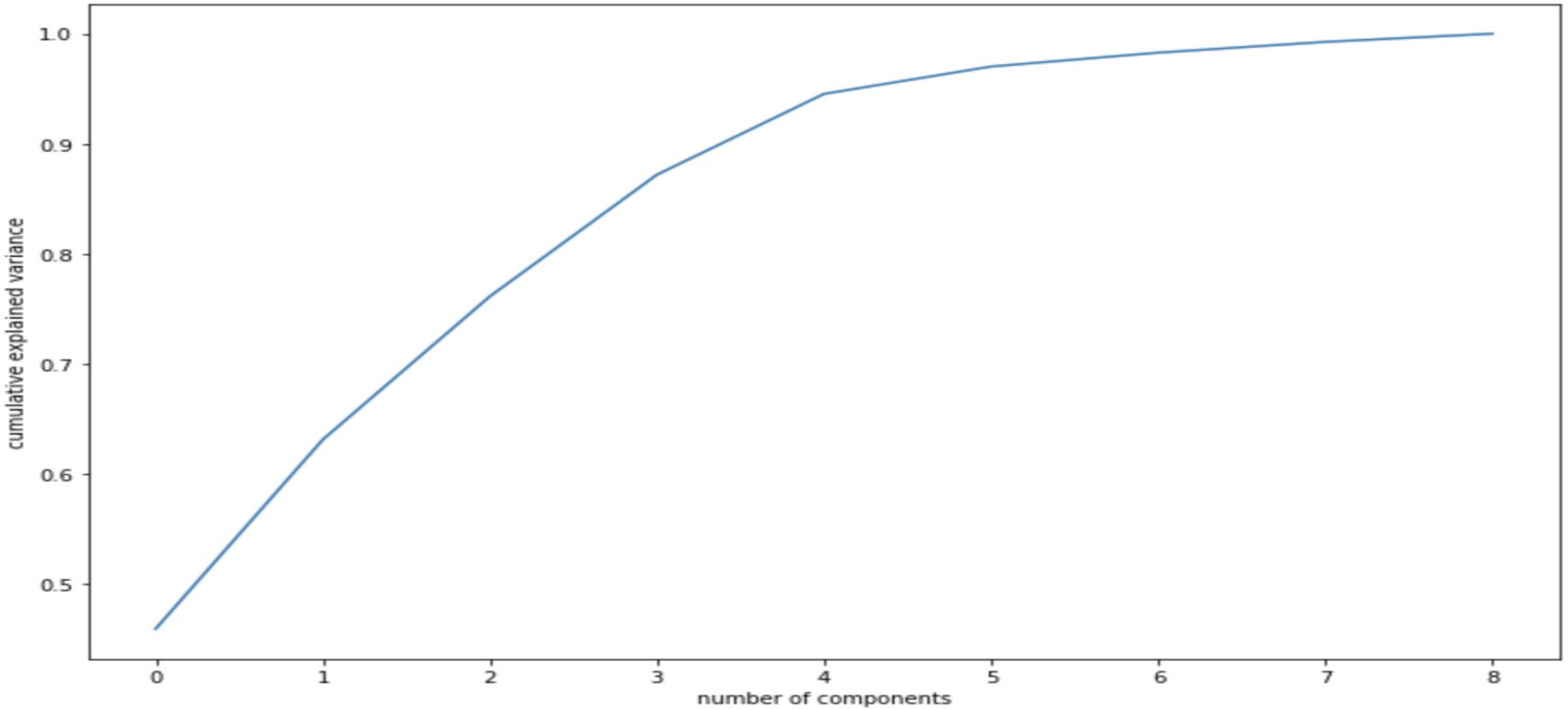
2- Draw the Scree plot to get the optimal no of principal component which covers about 95% of variance.

3- Do the K means clustering to find a cluster with countries which are good candidate of NGO's investment.

4- Do the Hierarchical clustering to get a cluster on the basis of social/economic, health metrices where NGO can invest

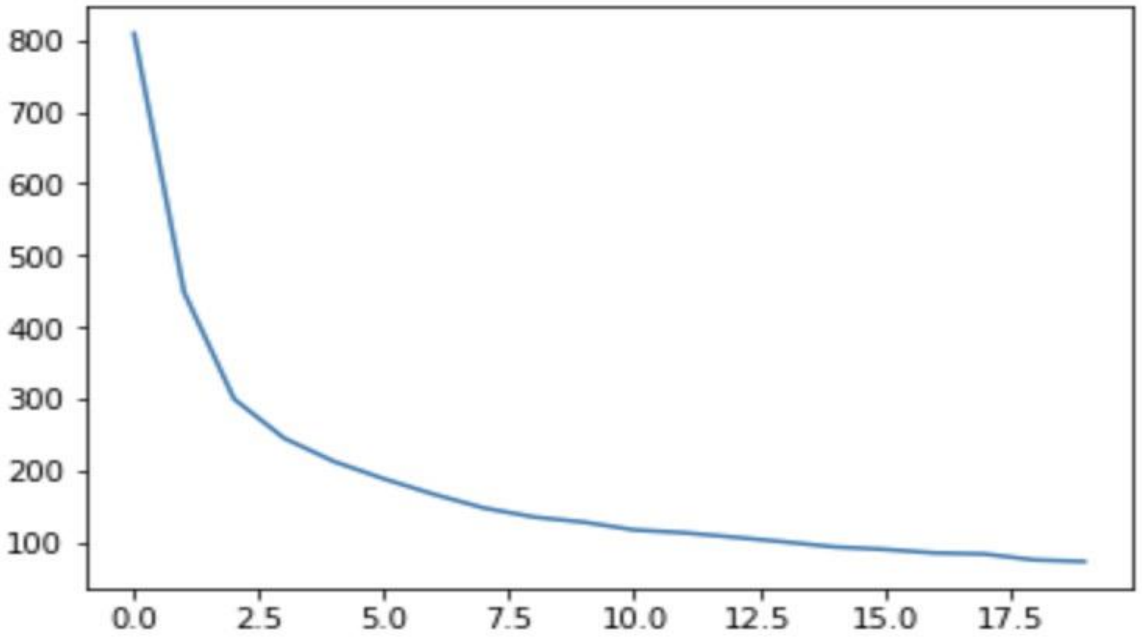
PCA

By using the Scree plot we got 4 principal component which covers maximum variance in given data frame.

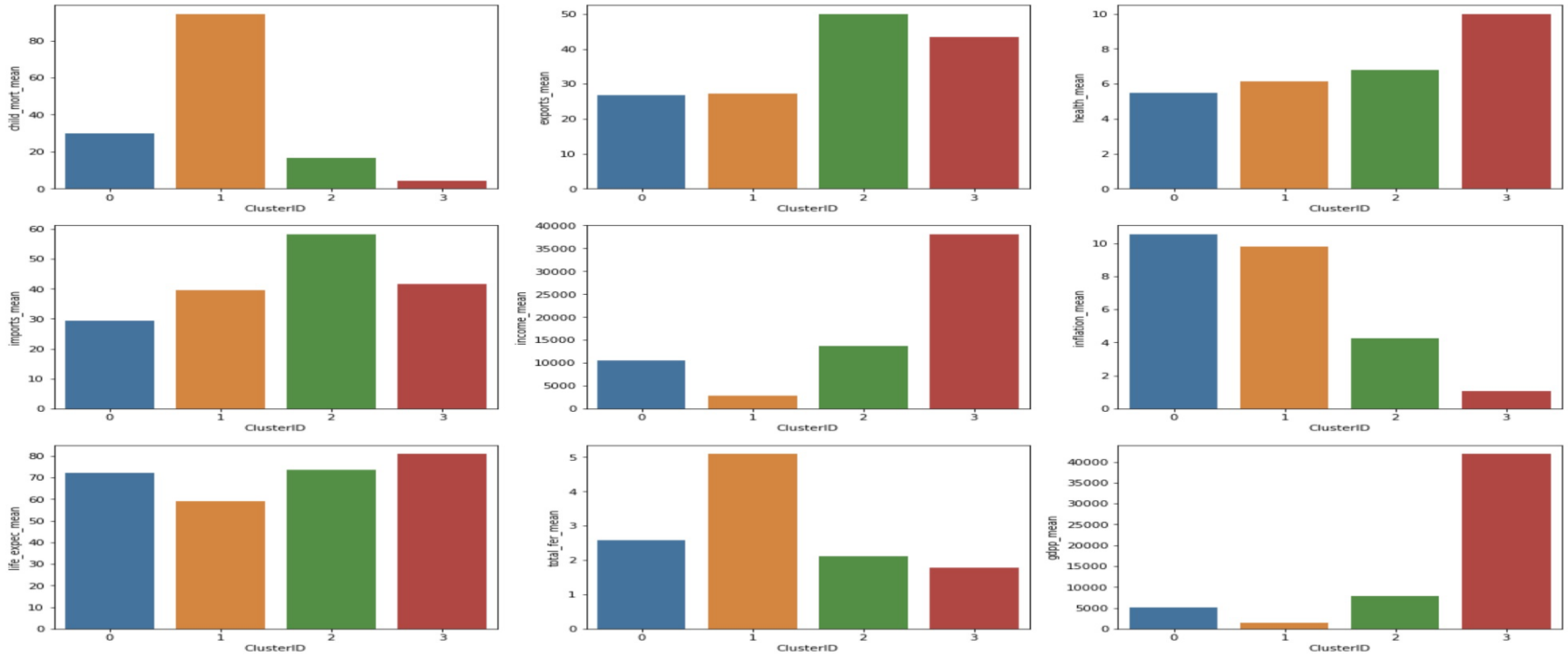


K Means clustering

1- By using elbow analysis, we found optimal no of cluster is 4.



K means analysis result



We can see cluster 1 has highest moratality rate, lowest gdpp,lowest life expectancy. So we choose to invest in cluster 1.

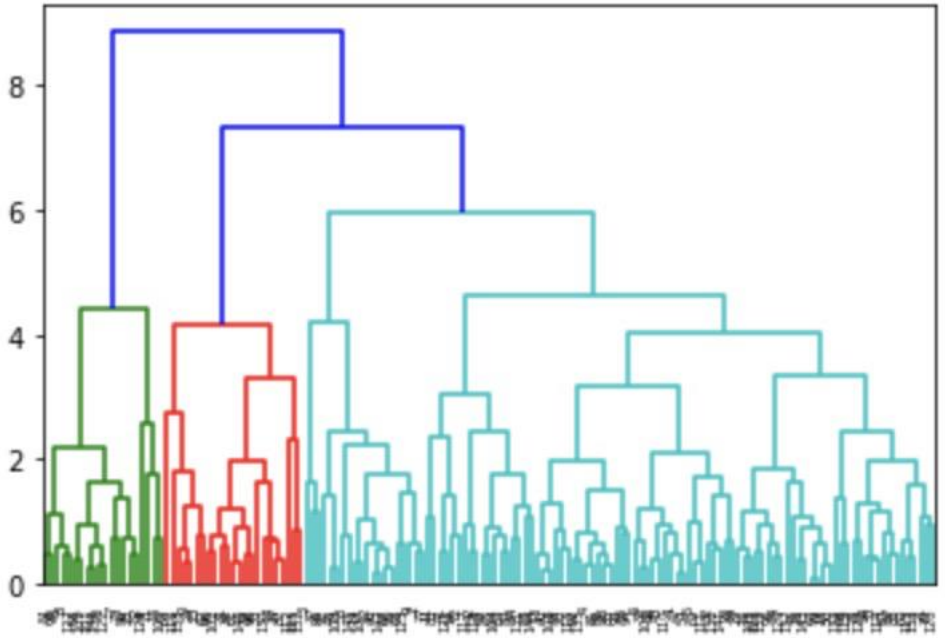
List of countries in cluster 1

- We found 40 countries where CEO can invest on the basis of child mortality, gdpp, life_expect etc.

	country	PC1	PC2	PC3	PC4	ClusterID	child_mort	exports	health	imports	income	inflation	life_expec	total_fer	gdpp
0	Afghanistan	-2.913025	0.095621	-0.718118	1.005255	1	90.2	10.00	7.58	44.9	1610	9.440	56.2	5.82	553
3	Angola	-2.932423	1.695555	1.525044	0.839625	1	119.0	62.30	2.85	42.9	5900	22.400	60.1	6.16	3530
17	Benin	-2.672314	0.418172	-0.257368	0.278672	1	111.0	23.80	4.10	37.2	1820	0.885	61.8	5.36	758
21	Botswana	-0.882088	0.457368	-0.584633	0.406161	1	52.5	43.60	8.30	51.3	13300	8.920	57.1	2.88	6350
24	Burkina Faso	-3.122053	0.038775	-0.455751	1.080918	1	116.0	19.20	6.74	29.6	1430	6.810	57.9	5.87	575
25	Burundi	-2.898971	-0.422663	-1.347413	1.821474	1	93.6	8.92	11.60	39.2	764	12.300	57.7	6.26	231
27	Cameroon	-2.807909	0.078649	-0.342961	0.543557	1	108.0	22.20	5.13	27.0	2660	1.910	57.3	5.11	1310
30	Central African Republic	-3.964964	0.386619	-0.323557	0.918301	1	149.0	11.80	3.98	26.5	888	2.010	47.5	5.21	446
31	Chad	-3.557555	1.289128	-0.103015	1.208406	1	150.0	36.80	4.53	43.5	1930	6.390	56.5	6.59	897
35	Comoros	-2.093556	0.343601	-0.328381	-0.118947	1	88.2	16.50	4.51	51.7	1410	3.870	65.9	4.75	769
36	Congo, Dem. Rep.	-3.173370	1.050382	-0.001219	1.574210	1	116.0	41.10	7.91	49.6	609	20.800	57.5	6.54	334
37	Congo, Rep.	-1.725676	2.176349	1.492719	-0.138593	1	63.9	85.10	2.46	54.7	5190	20.700	60.4	4.95	2740
39	Cote d'Ivoire	-2.581706	1.207873	-0.233066	0.665293	1	111.0	50.60	5.30	43.3	2690	5.390	56.3	5.27	1220
48	Eritrea	-2.474696	-0.618025	0.803497	-0.488256	1	55.2	4.79	2.66	23.3	1420	11.600	61.7	4.61	482
54	Gambia	-2.213670	0.223496	-0.393111	0.289914	1	80.3	23.80	5.69	42.7	1660	4.300	65.5	5.71	562
57	Ghana	-2.054167	0.380034	0.419654	0.090971	1	74.7	29.50	5.22	45.9	3060	16.600	62.2	4.27	1310
61	Guinea	-2.969529	0.728534	0.391039	0.666549	1	109.0	30.30	4.93	43.2	1190	16.100	58.0	5.34	648
62	Guinea-Bissau	-2.833616	-0.091128	-1.205174	1.133045	1	114.0	14.90	8.50	35.2	1390	2.970	55.6	5.05	547
64	Haiti	-4.409717	1.742230	-1.462809	1.793592	1	208.0	15.30	6.91	64.7	1500	5.450	32.1	3.33	662
78	Kenya	-1.874702	-0.171030	-0.249557	-0.285555	1	62.2	20.70	4.75	33.6	2480	2.090	62.8	4.37	967
80	Lao	-1.527770	0.545787	0.137159	-0.415582	1	78.9	35.40	4.47	49.3	3980	9.200	63.8	3.15	1140
86	Madagascar	-2.136009	0.342733	0.234842	-0.317469	1	62.2	25.00	3.77	43.0	1390	8.790	60.8	4.60	413
87	Malawi	-2.979885	0.216622	-0.204596	0.839108	1	90.5	22.80	6.59	34.9	1030	12.100	53.1	5.31	459
90	Mali	-3.412255	0.561469	-0.248615	1.073364	1	137.0	22.80	4.98	35.1	1870	4.370	59.5	6.55	708
91	Mauritania	-1.953927	1.383385	0.731462	0.195413	1	97.4	50.70	4.41	61.2	3320	18.900	68.2	4.98	1200
96	Mozambique	-2.920541	0.893270	-0.258704	0.633835	1	101.0	31.50	5.21	46.2	918	7.640	54.5	5.56	419
98	Namibia	-1.043375	1.002841	-0.732535	0.091483	1	56.0	47.80	6.78	60.7	8460	3.560	58.6	3.60	5190
102	Niger	-3.450168	0.969922	-0.581452	1.126581	1	123.0	22.20	5.16	49.1	814	2.550	58.8	7.49	348
103	Pakistan	-2.360347	-0.479400	1.013555	-0.421710	1	92.1	13.50	2.20	19.4	4280	10.900	65.3	3.85	1040
112	Rwanda	-1.679704	-1.001629	-1.468436	0.673221	1	63.6	12.00	10.50	30.0	1350	2.610	64.6	4.51	563
114	Senegal	-1.912170	0.091560	-0.506959	0.027625	1	66.8	24.90	5.66	40.3	2180	1.850	64.0	5.06	1000
116	Sierra Leone	-3.381625	-0.236302	-1.364254	2.581976	1	160.0	16.80	13.10	34.5	1220	17.200	55.0	5.20	399
120	South Africa	-1.191837	-0.556757	-0.717966	0.550206	1	53.7	28.60	8.94	27.4	12000	6.350	54.3	2.59	7280
125	Sudan	-2.319424	-0.769407	0.732740	0.492083	1	76.7	19.70	6.32	17.2	3370	19.600	66.3	4.88	1480
130	Tanzania	-2.554049	-0.215028	-0.088490	0.462226	1	71.9	18.70	6.01	29.1	2090	9.250	59.3	5.43	702
132	Timor-Leste	-2.371970	-1.177513	0.398923	1.158045	1	62.6	2.20	9.12	27.8	1850	26.500	71.1	6.23	3600
133	Togo	-1.997642	0.958362	-1.218037	0.560917	1	90.3	40.20	7.65	57.3	1210	1.180	58.7	4.87	488
138	Uganda	-2.854836	-0.352082	-0.700859	1.296271	1	81.0	17.10	9.01	28.6	1540	10.600	56.8	6.15	595
145	Yemen	-1.887451	-0.109453	1.109752	0.056257	1	56.3	30.00	5.18	34.4	4480	23.600	67.5	4.67	1310
146	Zambia	-2.864064	0.485998	0.223167	0.816364	1	83.1	37.00	5.89	30.9	3280	14.000	52.0	5.40	1460

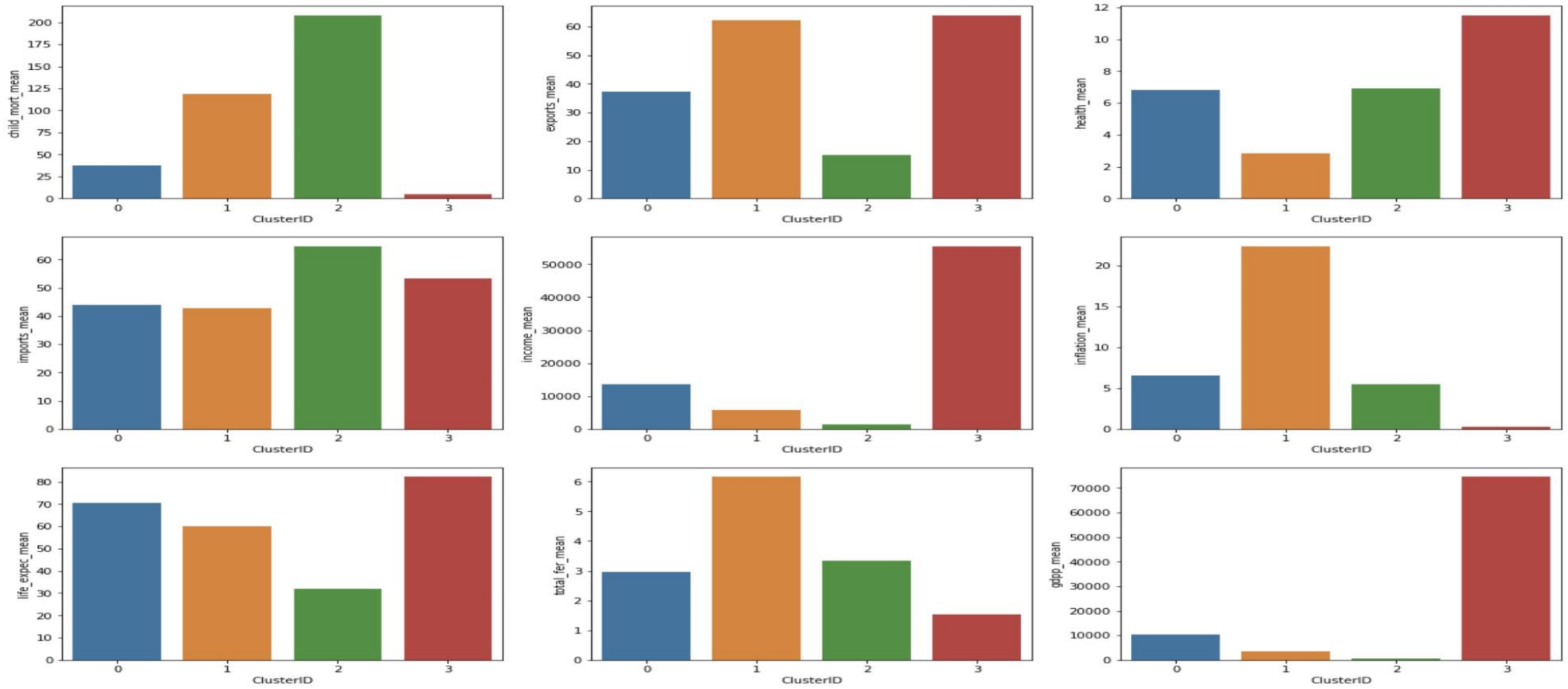
Hierarchical Clustering

1- By using complete linkage I found below dendogram to decide the no of optimal cluster to be taken into account.



2- we can easily see that 4 clusters are optimal choice.

Hierarchical Cluster analysis



We can again see, Inflation mean is very high and child mortality is significant. Cluster 1 is also having low income. So its better choice to invest in.

Outlier analysis

1- We found 20 countries which seems outliers on the basis of our algorithm.

2- When we compare behaviour of these countries. They are either developed countries which has higher gdpp OR having good income OR having less child mortality rate. So CEO should not invest in these countries.

	country	PC1	PC2	PC3	PC4	ClusterID	child_mort	exports	health	imports	income	inflation	life_expec	total_fer	gdpp
147	Brunei	NaN	NaN	NaN	NaN	NaN	10.5	67.4	2.84	28.0	80600	16.700	77.1	1.84	35300
148	Equatorial Guinea	NaN	NaN	NaN	NaN	NaN	111.0	85.8	4.48	58.9	33700	24.900	60.9	5.21	17100
149	Kiribati	NaN	NaN	NaN	NaN	NaN	62.7	13.3	11.30	79.9	1730	1.520	60.7	3.84	1490
150	Kuwait	NaN	NaN	NaN	NaN	NaN	10.8	66.7	2.63	30.4	75200	11.200	78.2	2.21	38500
151	Lesotho	NaN	NaN	NaN	NaN	NaN	99.7	39.4	11.10	101.0	2380	4.150	46.5	3.30	1170
152	Liberia	NaN	NaN	NaN	NaN	NaN	89.3	19.1	11.80	92.6	700	5.470	60.8	5.02	327
153	Luxembourg	NaN	NaN	NaN	NaN	NaN	2.8	175.0	7.77	142.0	91700	3.620	81.3	1.63	105000
154	Malta	NaN	NaN	NaN	NaN	NaN	6.8	153.0	8.65	154.0	28300	3.830	80.3	1.36	21100
155	Micronesia, Fed. Sts.	NaN	NaN	NaN	NaN	NaN	40.0	23.5	14.20	81.0	3340	3.800	65.4	3.46	2860
156	Mongolia	NaN	NaN	NaN	NaN	NaN	26.1	46.7	5.44	56.7	7710	39.200	66.2	2.64	2650
157	Nigeria	NaN	NaN	NaN	NaN	NaN	130.0	25.3	5.07	17.4	5150	104.000	60.5	5.84	2330
158	Norway	NaN	NaN	NaN	NaN	NaN	3.2	39.7	9.48	28.5	62300	5.950	81.0	1.95	87800
159	Oman	NaN	NaN	NaN	NaN	NaN	11.7	65.7	2.77	41.2	45300	15.600	76.1	2.90	19300
160	Qatar	NaN	NaN	NaN	NaN	NaN	9.0	62.3	1.81	23.8	125000	6.980	79.5	2.07	70300
161	Saudi Arabia	NaN	NaN	NaN	NaN	NaN	15.7	49.6	4.29	33.0	45400	17.200	75.1	2.96	19300
162	Seychelles	NaN	NaN	NaN	NaN	NaN	14.4	93.8	3.40	108.0	20400	-4.210	73.4	2.17	10800
163	Singapore	NaN	NaN	NaN	NaN	NaN	2.8	200.0	3.96	174.0	72100	-0.046	82.7	1.15	46600
164	United Arab Emirates	NaN	NaN	NaN	NaN	NaN	8.6	77.7	3.66	63.6	57600	12.500	76.5	1.87	35000
165	United States	NaN	NaN	NaN	NaN	NaN	7.3	12.4	17.90	15.8	49400	1.220	78.7	1.93	48400
166	Venezuela	NaN	NaN	NaN	NaN	NaN	17.1	28.5	4.91	17.6	16500	45.900	75.4	2.47	13500