# Clustering

Humanitarian Aid

Objective is to evaluate and identify the countries into categories basis the socio-economic factors and help the CEO with the list of countries to lay focus on

Outlier detection:

Provided data comprise 167 records

	country	child_mort	exports	health	imports	income	inflation	life_expec	total_fer	gdpp
0	Afghanistan	90.2	10.0	7.58	44.9	1610	9.44	56.2	5.82	553
1	Albania	16.6	28.0	6.55	48.6	9930	4.49	76.3	1.65	4090
2	Algeria	27.3	38.4	4.17	31.4	12900	16.10	76.5	2.89	4460
3	Angola	119.0	62.3	2.85	42.9	5900	22.40	60.1	6.16	3530
4	Antigua and Barbuda	10.3	45.5	6.03	58.9	19100	1.44	76.8	2.13	12200

- For backward countries child mortality can be high, life expectancy can be low so, we don't modify these values

- It looks like there are outliers present in exports, imports, income, inflation, total fertility and GDP on the higher range of the values.

- So, we set all the values that are greater than 99th percentile value to the 99th percentile value
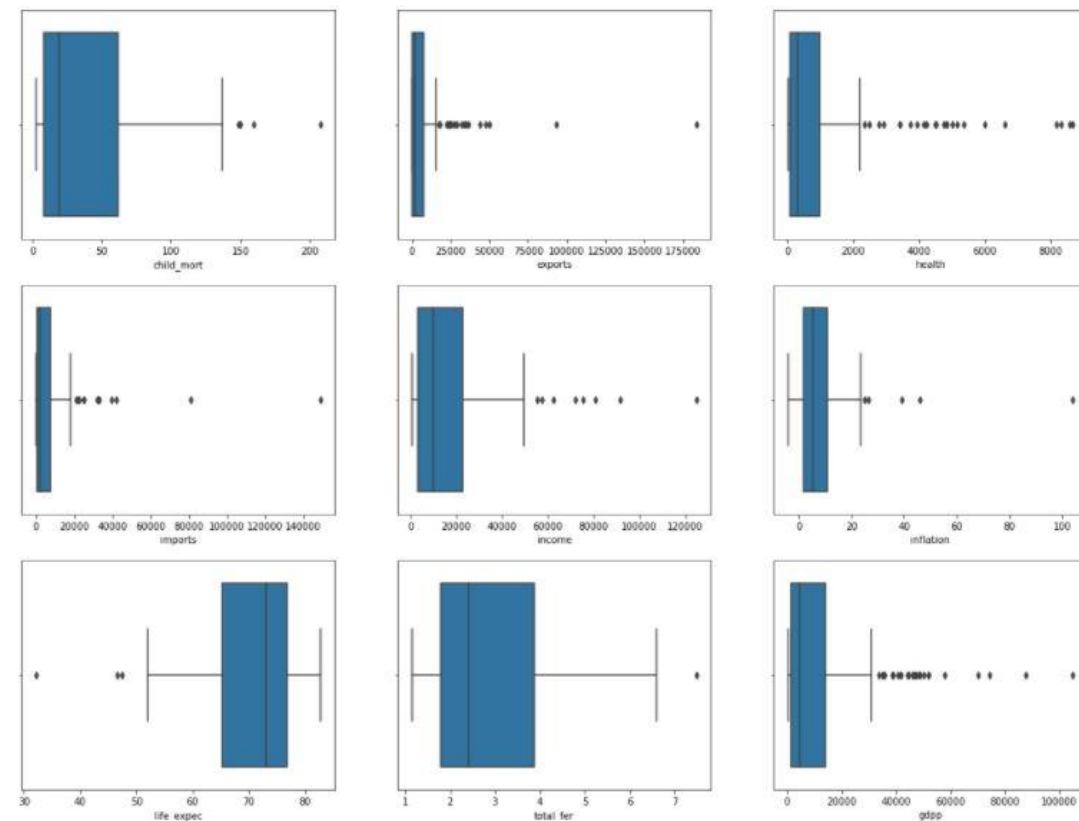
Exploratory Data Analysis:
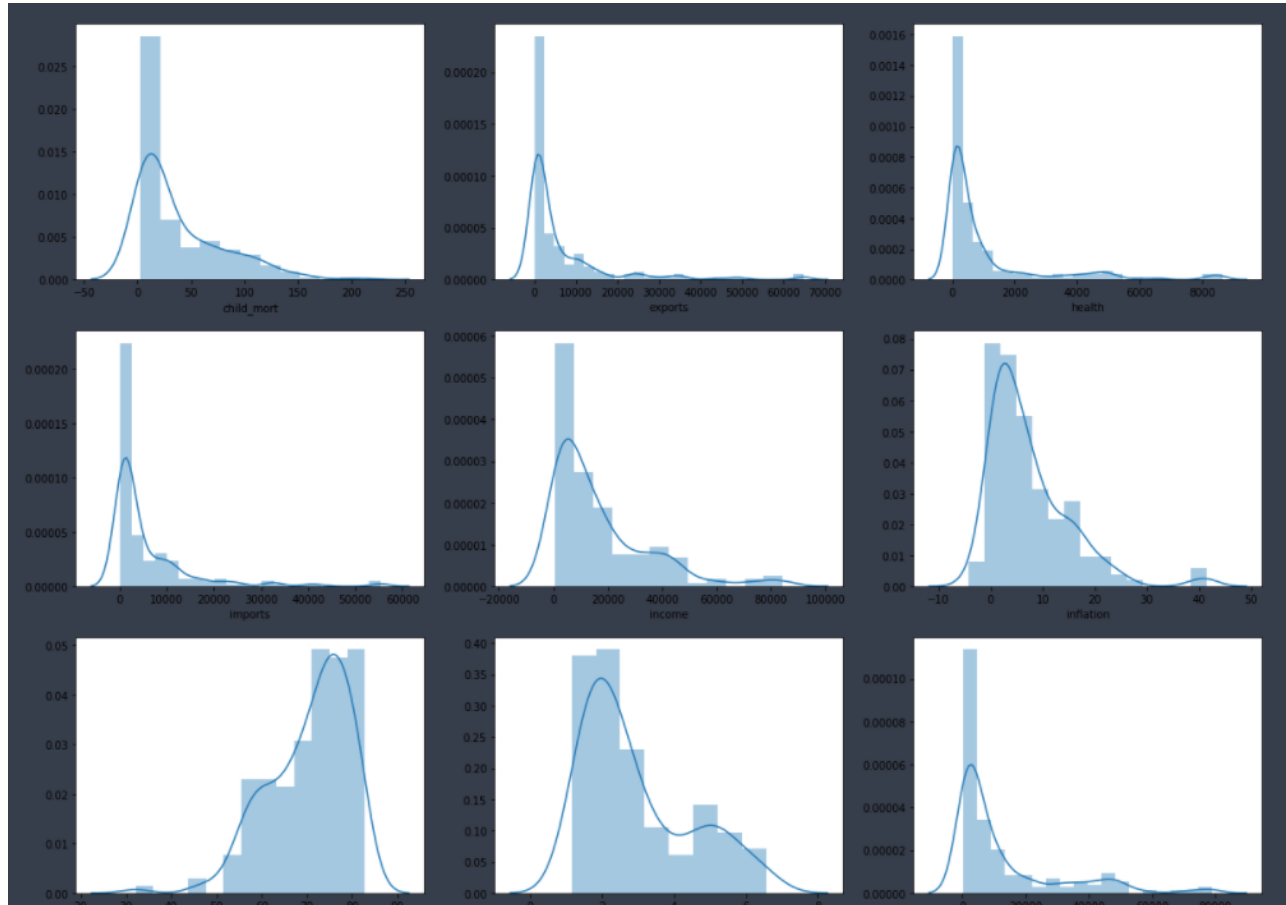
- We can see that the distribution of data for the features child mortality, income and gdp is not uniform, the data has values distributed from low range to high range, which will help us to find countries with low income, low gdp and high child mortality rate.

- So we can use these features to form good clusters and find out backward countries.

Cluster Profiling using KMeans

The silhouette plot suggests that we get optimal performance for 3 clusters
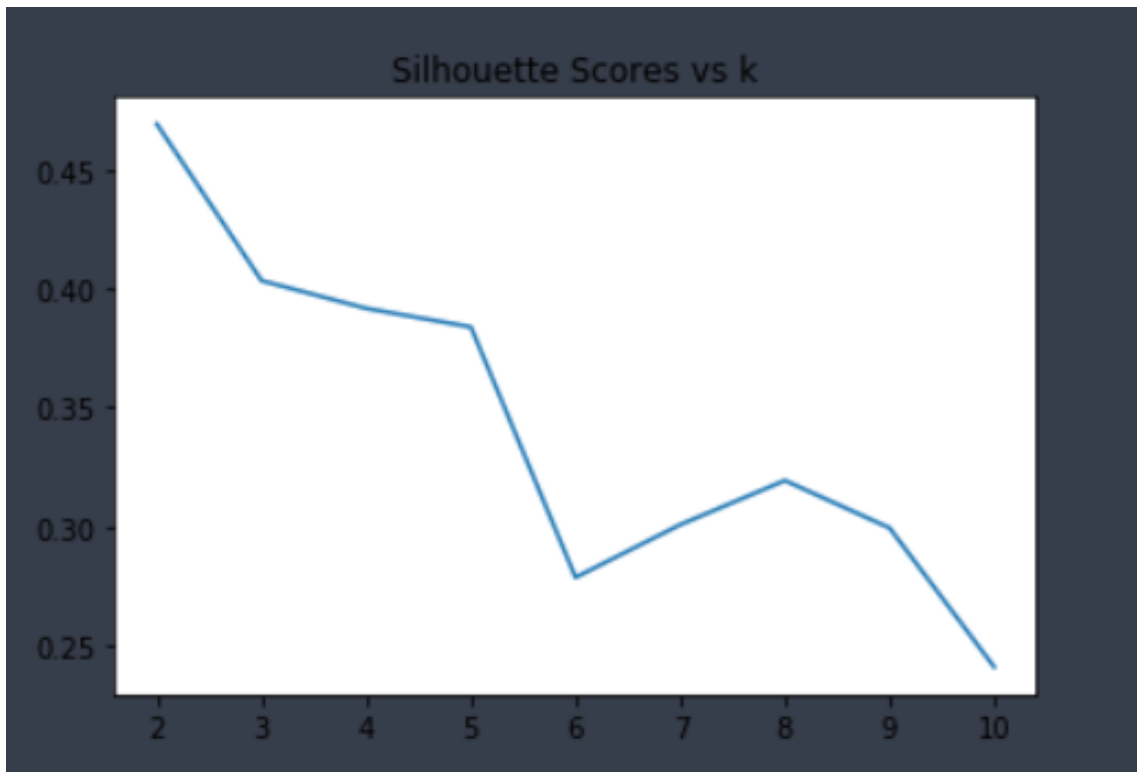
- Clusters show clear difference how child mortality, income and gdp values differ
- Countries that belong to cluster 2 have high child mortality rate, low income and low gdp. These countries should be given aid
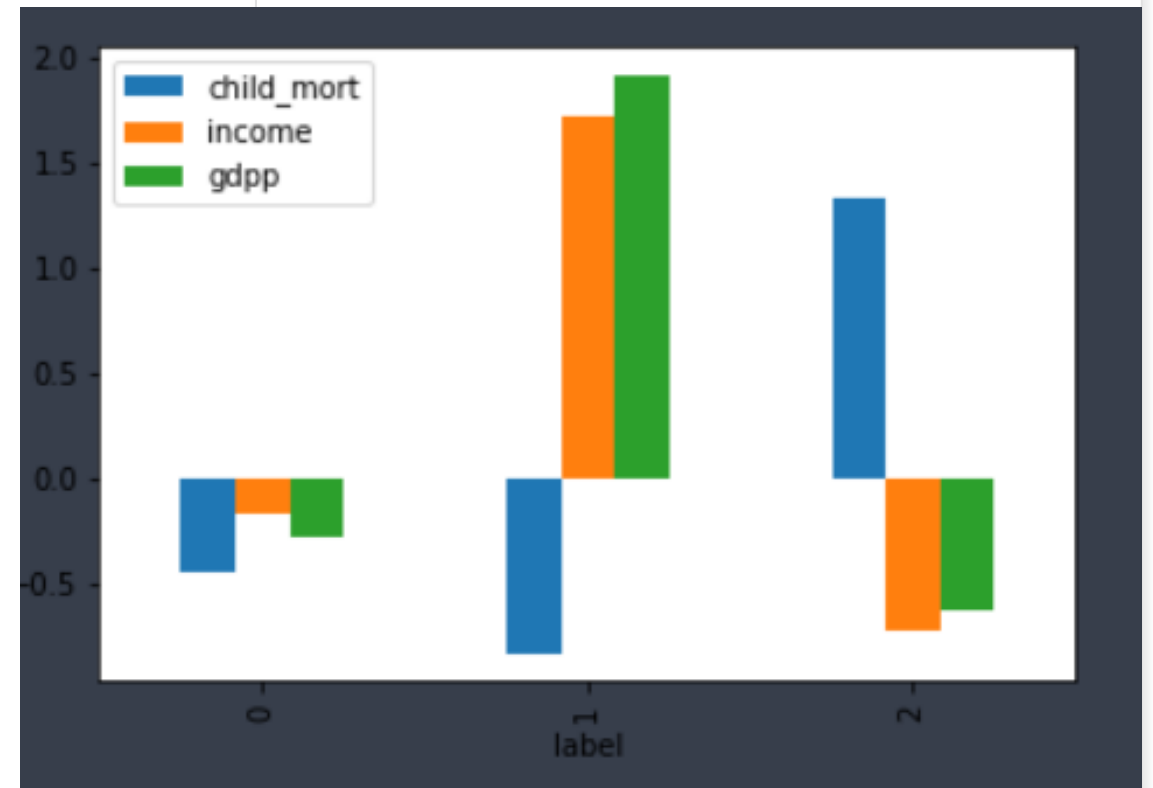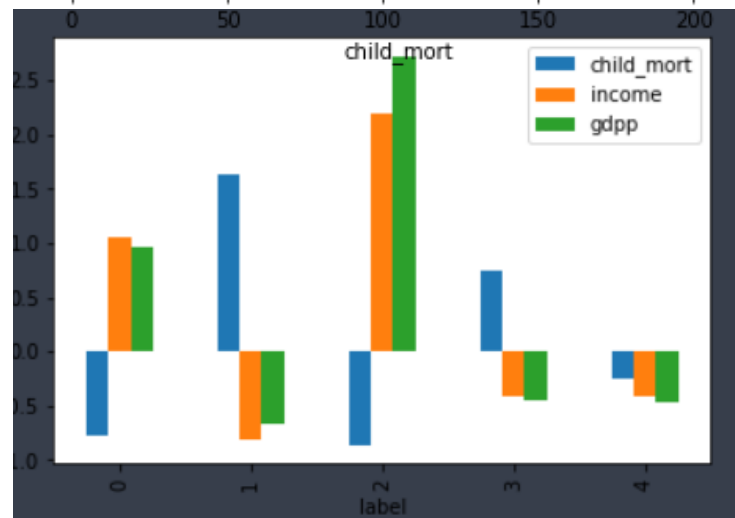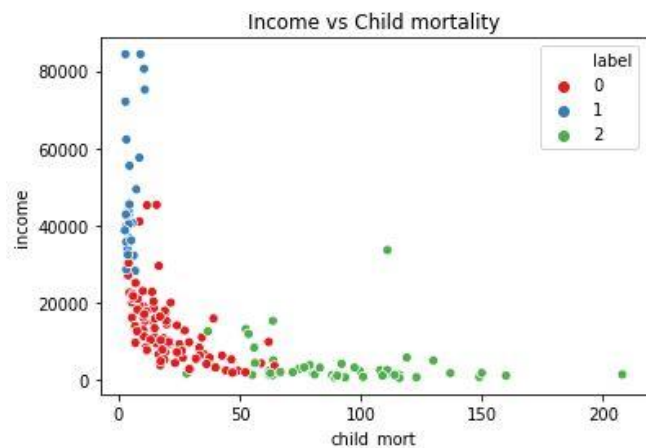
Top 5 countries using KMeans

	country	child_mort	exports	health	imports	income	inflation	life_expec	total_fer	gdpp	label
26	Burundi	93.6	20.6052	26.7960	90.552	764.0	12.30	57.7	6.2600	231.0	2
88	Liberia	89.3	62.4570	38.5860	302.802	700.0	5.47	60.8	5.0200	327.0	2
37	Congo, Dem. Rep.	116.0	137.2740	26.4194	165.664	609.0	20.80	57.5	6.5400	334.0	2
112	Niger	123.0	77.2560	17.9568	170.868	814.0	2.55	58.8	6.5636	348.0	2
132	Sierra Leone	160.0	67.0320	52.2690	137.655	1220.0	17.20	55.0	5.2000	399.0	2

The results obtained from KMeans clustering suggest that these 5 countries should be given priority in providing aid:
1.Burundi
2.Liberia
3.Congo,Dem. Rep.
4.Niger
5.Sierra Leone

Cluster Profiling using Hierarchical clustering (Complete Linkage)

The silhouette plot suggests that the optimal performance can be obtained when we use 5 clusters
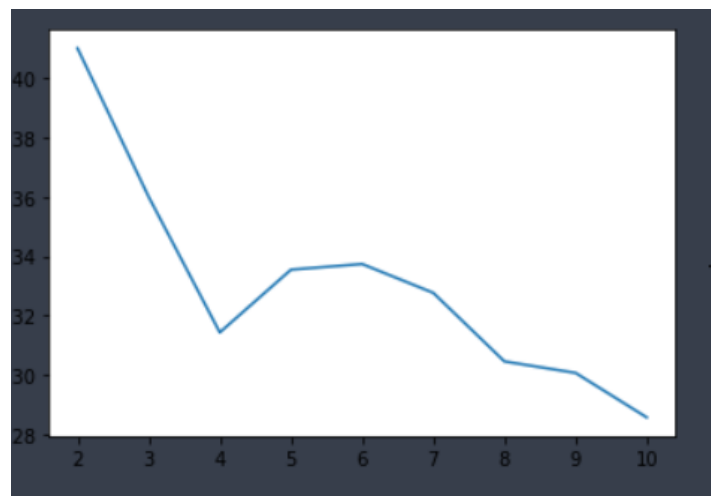
- From the plot we can see how the child mortality, income and gdp values differe from each other

- Countries that belong to cluster 1 have high child mortality rate, low income and low gdp. These countries should be given aid.
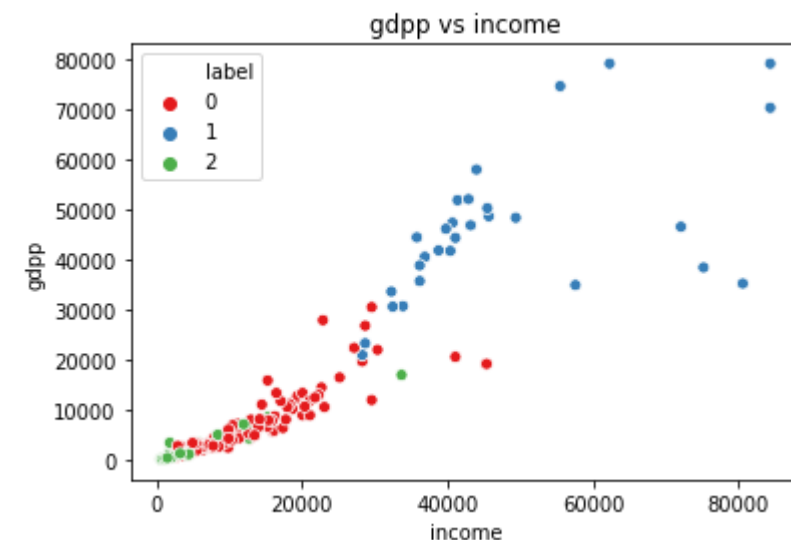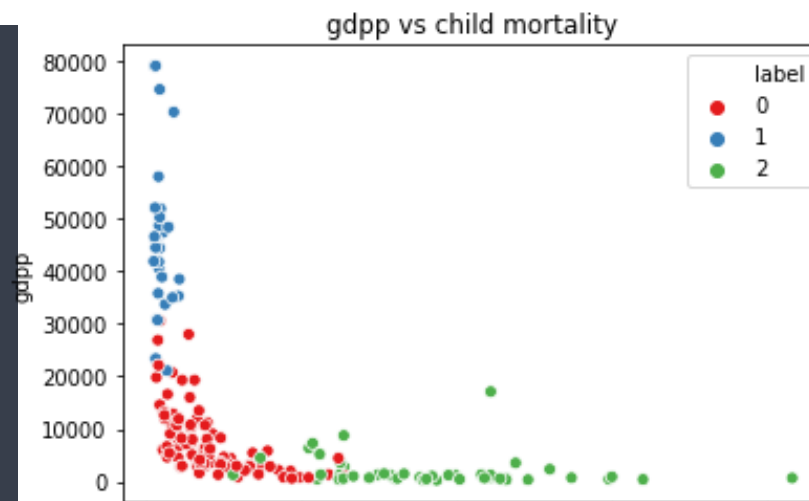
Top 5 countries using complete linkage

- The results obtained from complete linkage hierarchical clustering suggest that these 5 countries should be given priority in providing aid:

- 1.Burundi

- 2.Liberia

- 3.Congo,Dem. Rep.

- 4.Niger

- 5.Sierra Leone

	country	child_mort	exports	health	imports	income	inflation	life_expec	total_fer	gdpp	label
26	Burundi	93.6	20.6052	26.7960	90.552	764.0	12.30	57.7	6.2600	231.0	1
88	Liberia	89.3	62.4570	38.5860	302.802	700.0	5.47	60.8	5.0200	327.0	1
37	Congo, Dem. Rep.	116.0	137.2740	26.4194	165.664	609.0	20.80	57.5	6.5400	334.0	1
112	Niger	123.0	77.2560	17.9568	170.868	814.0	2.55	58.8	6.5636	348.0	1
132	Sierra Leone	160.0	67.0320	52.2690	137.655	1220.0	17.20	55.0	5.2000	399.0	1

<table>
<tr><td></td><td>Conclusion on
clustering methods</td><td>• The results obtained from both the clustering methods are same.

• The countries that need aid are Burundi, Liberia, Congo, Dem. Rep., Niger and Sierra Leone.

• In our analysis, we gave more preference to GDP, followed by child mortality and income to determine the top 5 countries that need aid.</td></tr>
</table>

Countries based on socio economic factors

	country	child_mort	exports	health	imports	income	inflation	life_expec	total_fer	gdpp
26	Burundi	93.6	20.6052	26.7960	90.552	764.0	12.30	57.7	6.2600	231.0
88	Liberia	89.3	62.4570	38.5860	302.802	700.0	5.47	60.8	5.0200	327.0
37	Congo, Dem. Rep.	116.0	137.2740	26.4194	165.664	609.0	20.80	57.5	6.5400	334.0
112	Niger	123.0	77.2560	17.9568	170.868	814.0	2.55	58.8	6.5636	348.0
132	Sierra Leone	160.0	67.0320	52.2690	137.655	1220.0	17.20	55.0	5.2000	399.0

We used the same socio economic and health factors(child mortality rate, income and GDP) that we used in our clustering analysis to find the top 5 countries that need aid and we obtained the same results as our clustering analysis results.

The results obtained from analysis using socio economic and health factors suggest that these 5 countries should be given priority in providing aid:
1.Burundi
2.Liberia
3.Congo,Dem. Rep.
4.Niger
5.Sierra Leone