Clustering Assignment

SUBMITTED BY:

ABHISHEK VERMA

Problem Statement:

Categorise the countries using some socio-economic and health factors (determining the overall development of the country) and analyse the clusters to identify the ones which are in dire need of aid so that the international humanitarian NGO named "HELP" which is committed to fighting poverty can providing the people of backward countries with basic amenities and relief during the time of disasters and natural calamities can utilise its funds efficiently.
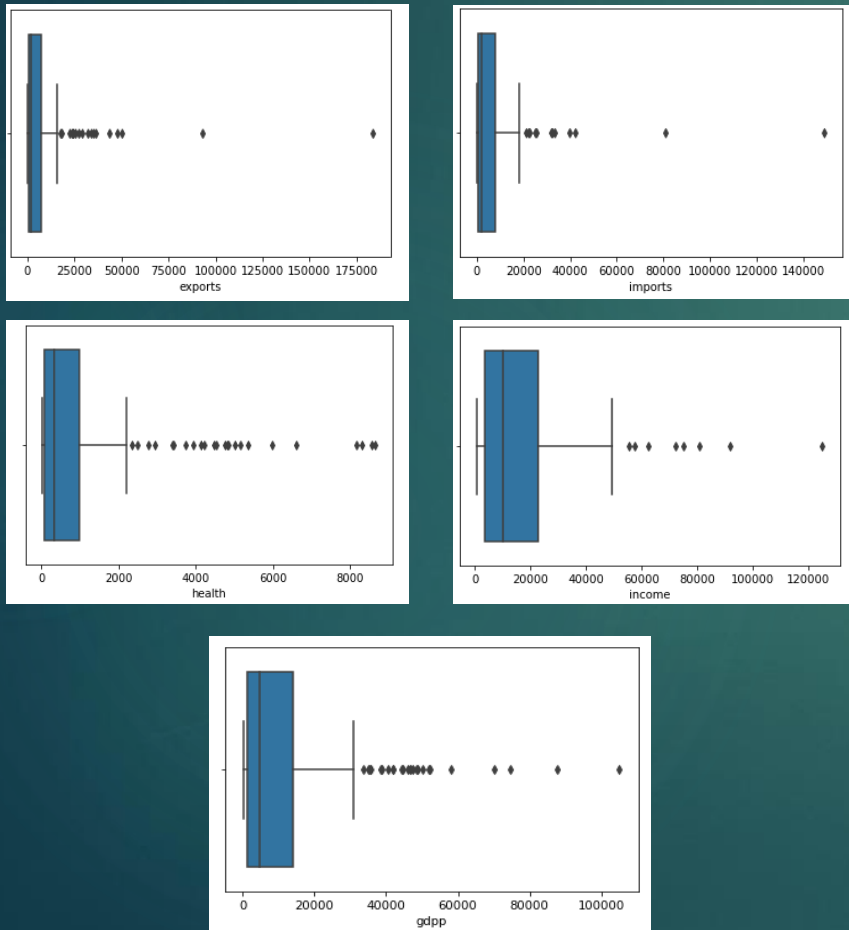
Methodology:

The metric behind selecting the top countries is based on that the countries have comparatively low GDP per capita and income, has high child mortality rate and has low expenditure towards health (per capita).
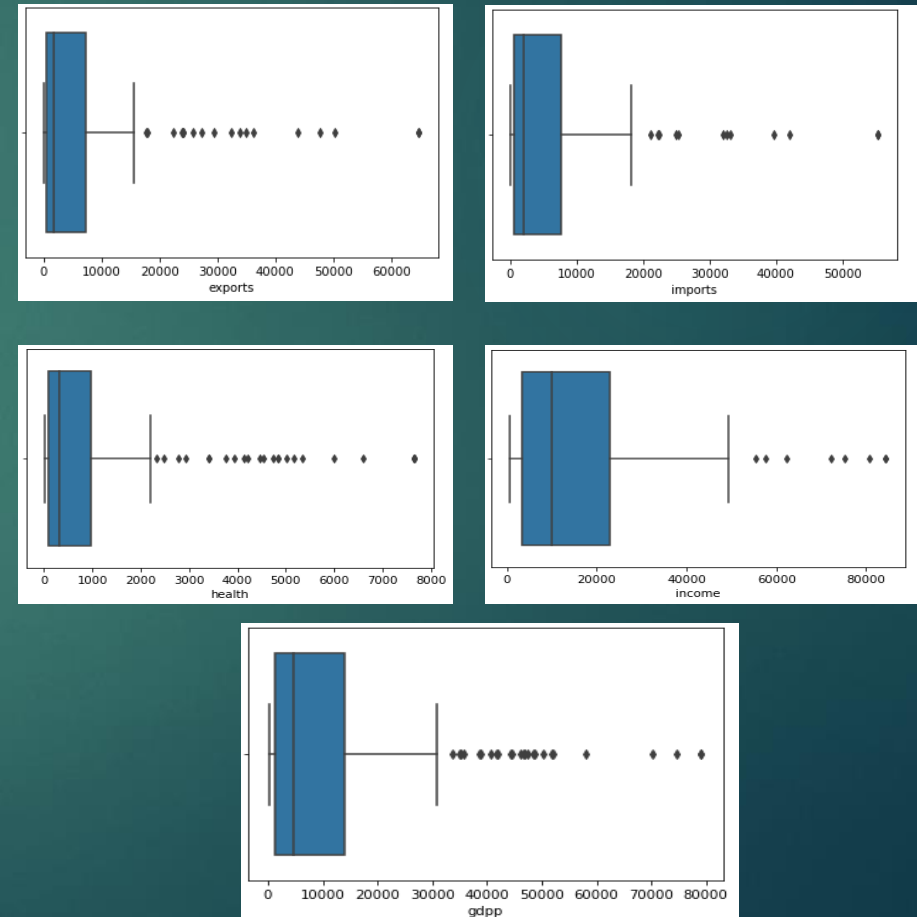
Data Sanity Check

There are certain countries (probably developed) with extremely high exports, imports, income, health expenditure and GDP with respect to others. So that they do not overshadow the others, such attributes values were soft capped (1-2%, as most of the extreme outliers were in this range only).
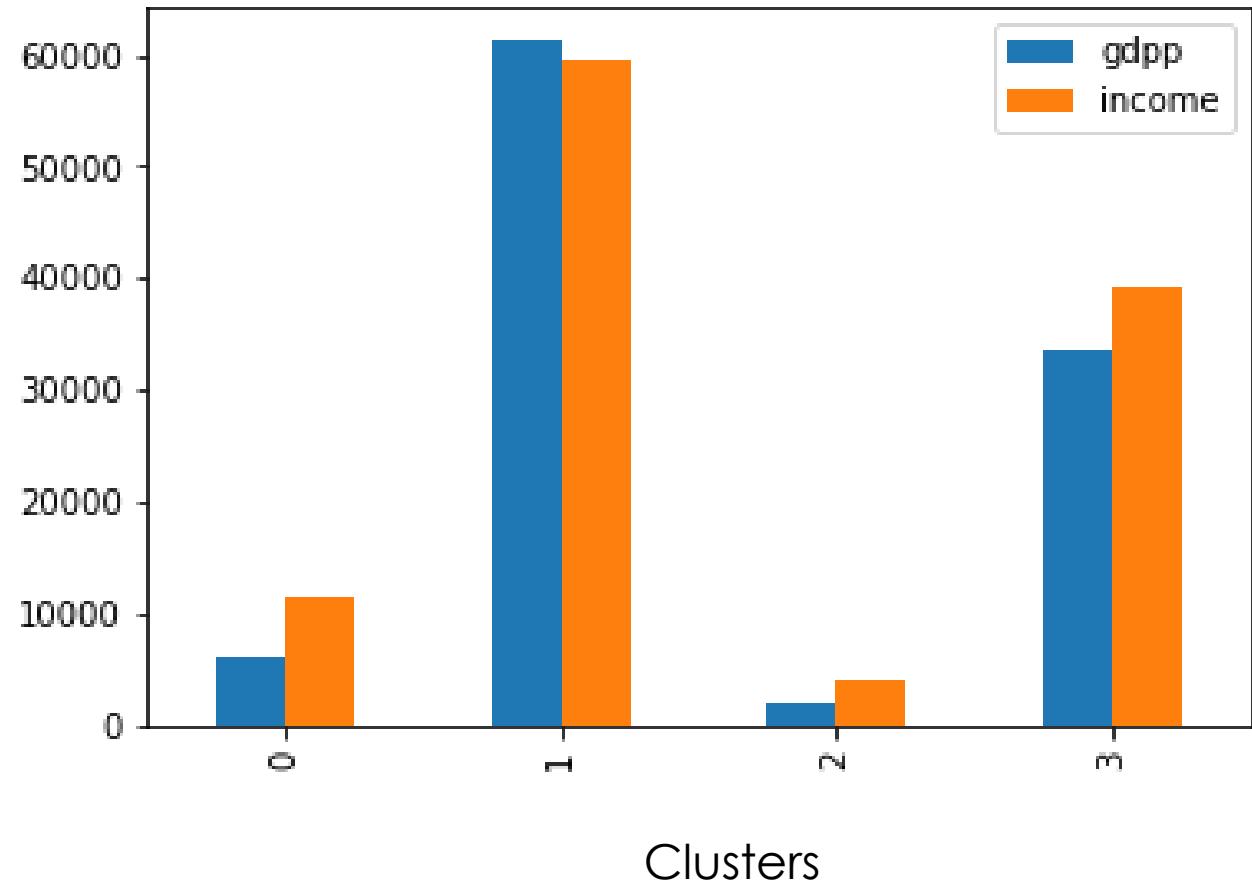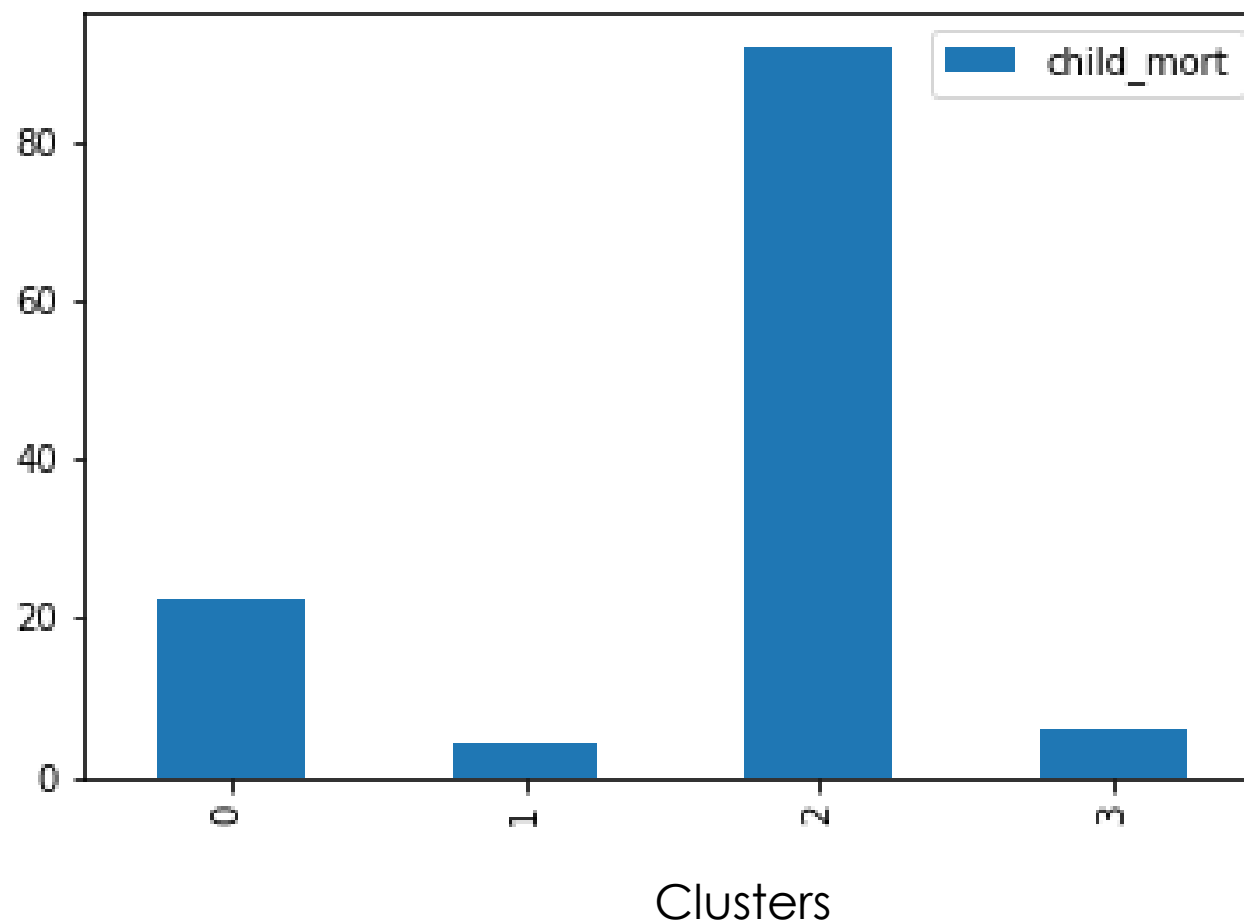
Original data:

After soft capping:

Result of Clustering Model

Countries belonging to cluster 2 had really low GDP and Income as compared to others.
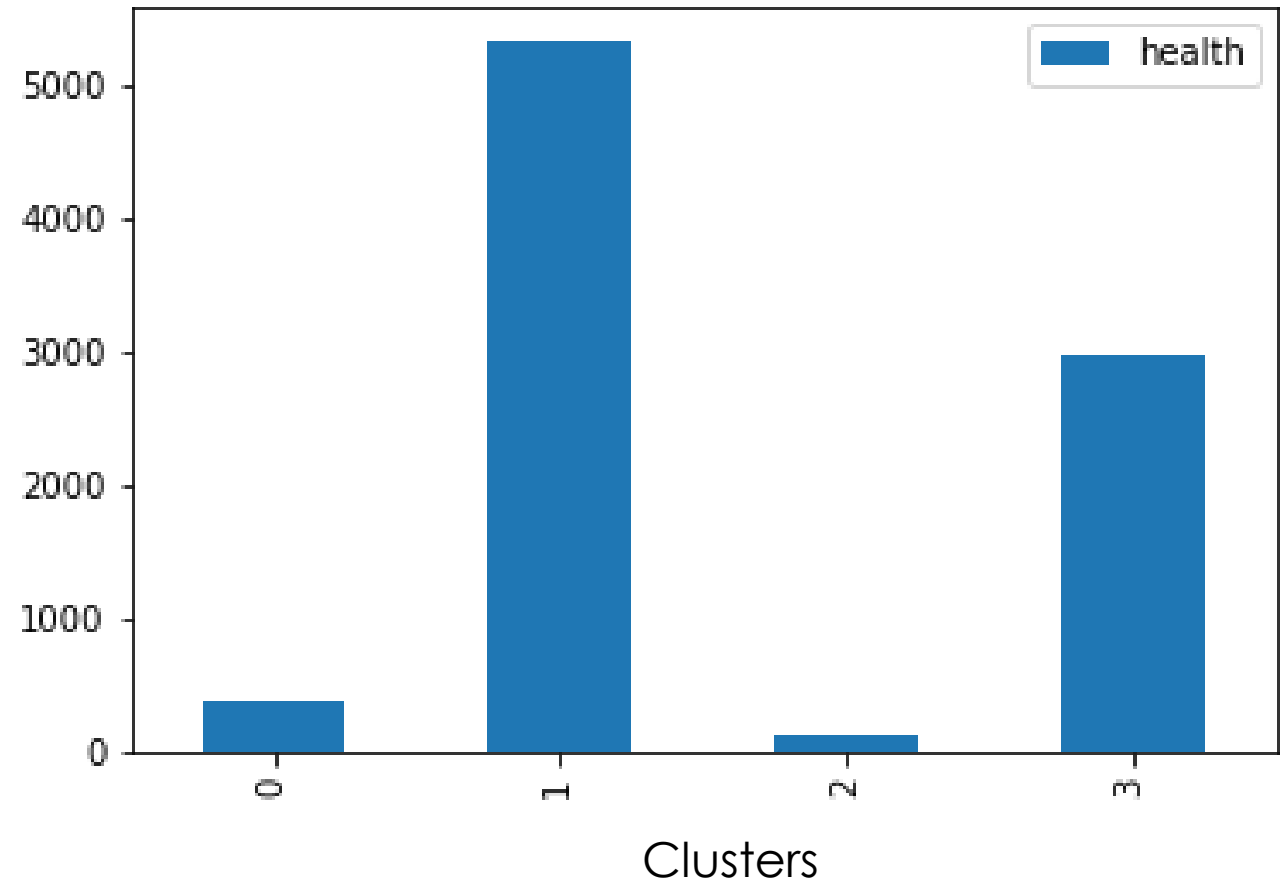
Result of Clustering Model

Countries belonging to cluster 2 also have high child mortality rate among the others. This indicates that these countries have problems related to health infrastructure and also the children and mothers might not be getting proper nutrition.
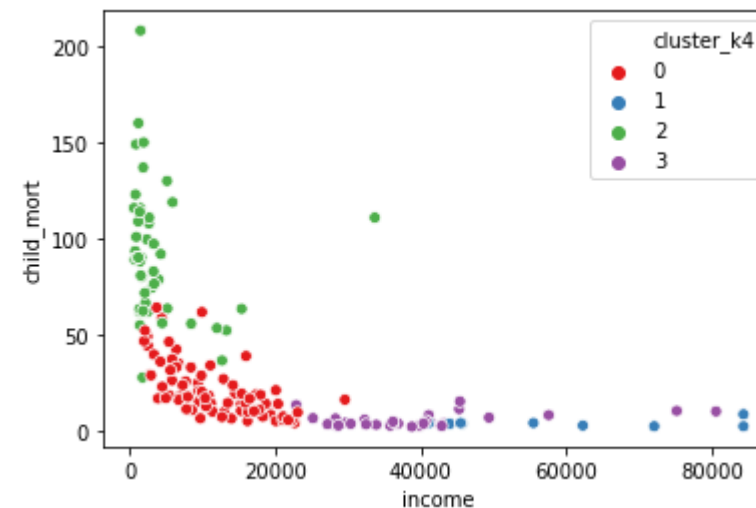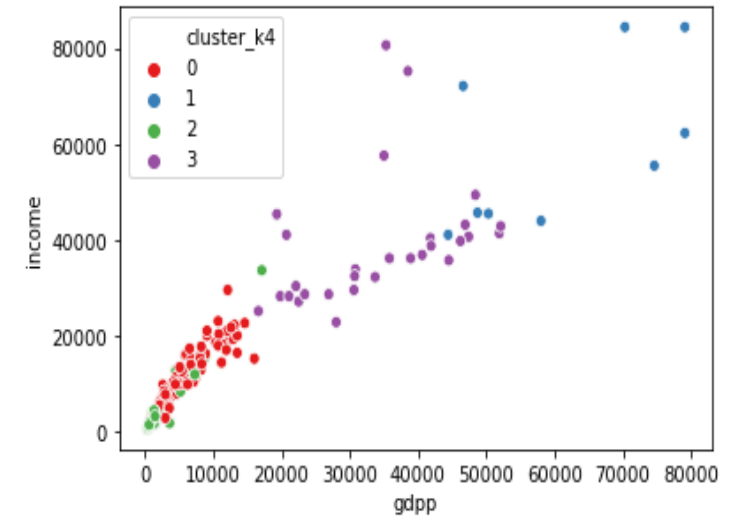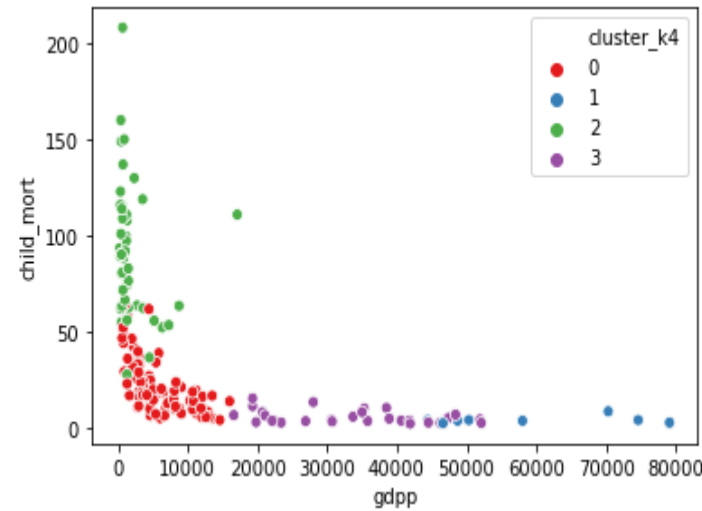
Result of Clustering Model

The Countries in cluster also have low per capita health expenditure i.e., they have poor health facilities.

Result of Clustering Model

It is certain that countries in cluster 2 are the one who will be in direst need of aid in times of natural disasters and calamities because of the socio-economic and health factors.

Conclusion

It is assumed that the citizens of countries (having low income along with country's low GDP, Exports and Imports) will need the help of our NGO in times of natural calamities and disasters. Therefore, it is recommended that the NGO must prioritize these countries as they are clearly underdeveloped countries. The NGO must spend a significant amount of its funding in helping these countries first. Such top 10 countries are :

- Burundi
- Liberia
- Congo, Dem. Rep.
- Niger
- Sierra Leone
- Madagascar
- Mozambique
- Central African Republic
- Malawi
- Eritrea

country	child_mort	exports	health	imports	income	inflation	life_expec	total_fer	gdpp
Burundi	93.6	20.6052	26.7960	90.552	764.0	12.300	57.7	6.26	231.0
Liberia	89.3	62.4570	38.5860	302.802	700.0	5.470	60.8	5.02	327.0
Congo, Dem. Rep.	116.0	137.2740	26.4194	165.664	609.0	20.800	57.5	6.54	334.0
Niger	123.0	77.2560	17.9568	170.868	814.0	2.550	58.8	7.49	348.0
Sierra Leone	160.0	67.0320	52.2690	137.655	1220.0	17.200	55.0	5.20	399.0
Madagascar	62.2	103.2500	15.5701	177.590	1390.0	8.790	60.8	4.60	413.0
Mozambique	101.0	131.9850	21.8299	193.578	918.0	7.640	54.5	5.56	419.0
Central African Republic	149.0	52.6280	17.7508	118.190	888.0	2.010	47.5	5.21	446.0
Malawi	90.5	104.6520	30.2481	160.191	1030.0	12.100	53.1	5.31	459.0
Eritrea	55.2	23.0878	12.8212	112.306	1420.0	11.600	61.7	4.61	482.0