



Clustering Assignment

By Krunal Nagda

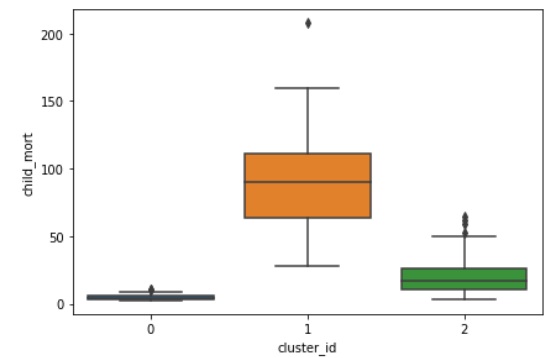
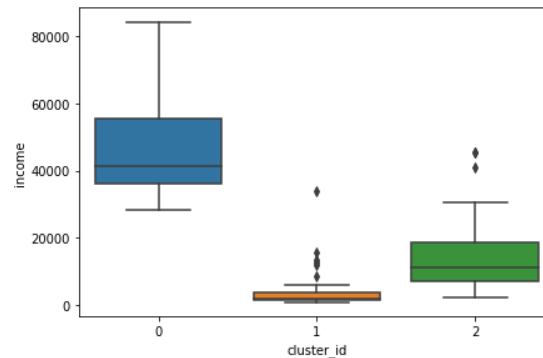
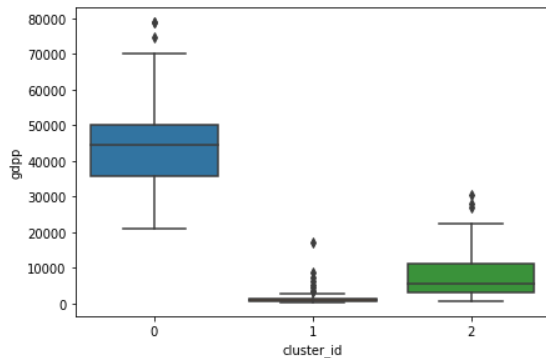
Problem Statement

- Allocating \$10 million to countries that are in direst need of financial aid based on socio-economic factors
- This could be done by fitting a clustering model to form clusters based on the most prominent socio-economic factors

Outlier Treatment

- All fields have outliers which could affect the clustering exercise so we need to treat them accordingly
- We do soft capping on the upper outliers by imputing the values greater than the 99th percentile by this value
- Soft capping is done so that we do not lose too much information
- We ignore the lower outliers since they could result in our model not giving correct results

K-Means Clustering

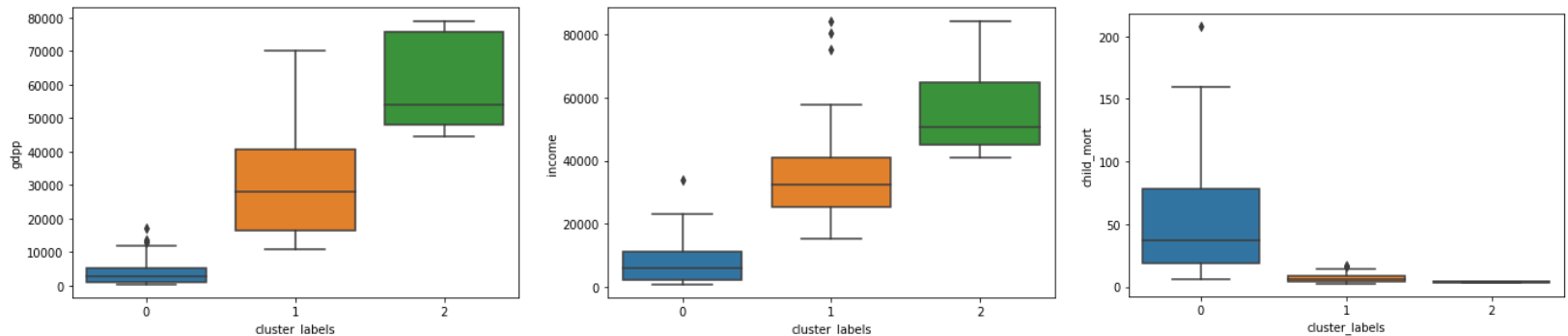


- The clusters formed are:
 - 0 - Low child mortality, high income and high gdp
 - 1 - High child mortality, low income and low gdp
 - 2 - Moderate child mortality, income and gdp
- Here, we are only interested in cluster 1, since it complies with our problem statement

K-Means Clustering

- Out of the 48 countries belonging to cluster 1, the countries that are in direst need of financial aid are the following:
 - Burundi
 - Liberia
 - Democratic Republic of Congo
 - Niger
 - Sierra Leone
 - Madagascar
 - Mozambique
 - Central African Republic
 - Malawi
 - Eritrea

Hierarchical Clustering



- The clusters formed are:
 - 0 - High child mortality, low income and low gdp
 - 1 - Moderate child mortality, income and gdp
 - 2 - Low child mortality, high income and high gdp
- Here, we are only interested in cluster 0, since it complies with our problem statement

Hierarchical Clustering

- Out of the 118 countries belonging to cluster 0, the countries that are in direst need of financial aid are the following:
 - Burundi
 - Liberia
 - Democratic Republic of Congo
 - Niger
 - Sierra Leone
 - Madagascar
 - Mozambique
 - Central African Republic
 - Malawi
 - Eritrea

Final List of Countries

- As we can see, both K-Means and Hierarchical Clustering give the same countries which are in direst need of financial aid
- The three most important factors have been summarized country-wise below:

	country	child_mort	income	gdpp
0	Burundi	93.6	764.0	231.0
1	Liberia	89.3	700.0	327.0
2	Congo, Dem. Rep.	116.0	609.0	334.0
3	Niger	123.0	814.0	348.0
4	Sierra Leone	160.0	1220.0	399.0
5	Madagascar	62.2	1390.0	413.0
6	Mozambique	101.0	918.0	419.0
7	Central African Republic	149.0	888.0	446.0
8	Malawi	90.5	1030.0	459.0
9	Eritrea	55.2	1420.0	482.0

where,

child_mort : Death of children under 5 years of age per 1000 live births

income : Net income per person

gdpp : The GDP per capita. Calculated as the Total GDP divided by the total population.

Conclusion

- After the entire clustering exercise, we get a list of countries on which the NGO should ideally focus to provide financial need
- The allotment of the \$10 million need not be equal, but rather should be proportionate to the GDP, average household income, child mortality rate and the population of the countries



Thank You.