CLUSTERING ASSIGNMENT

SUBMITTED BY

INDU R

PROBLEM STATEMENT

HELP International is an international humanitarian NGO that is committed to fighting poverty and providing the people of backward countries with basic amenities and relief during the time of disasters and natural calamities. It runs a lot of operational projects from time to time along with advocacy drives to raise awareness as well as for funding purposes.

After the recent funding programmes, they have been able to raise around $ 10 million. Now 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.

PROCEDURE FOLLOWED

1. Understanding the data- inspecting dataset details
2. EDA
3. Performing clustering

    a. Data preparation for clustering

        i. Outlier treatment Clustering Assignment

        ii. Hopkins check

    b. Clustering

        i. K-MEANS

    1. Run K-Means and choose K using both Elbow and Silhouette score

    2. Run K-Means with the chosen K

    3. Visualise the clusters

    4. Clustering profiling using “gdpp, child_mort and income”

.

ii. Hierarchical Clustering

1. Use both Single and Complete linkage

2. Choose one method based on the results

3. Visualise the clusters Clustering Assignment

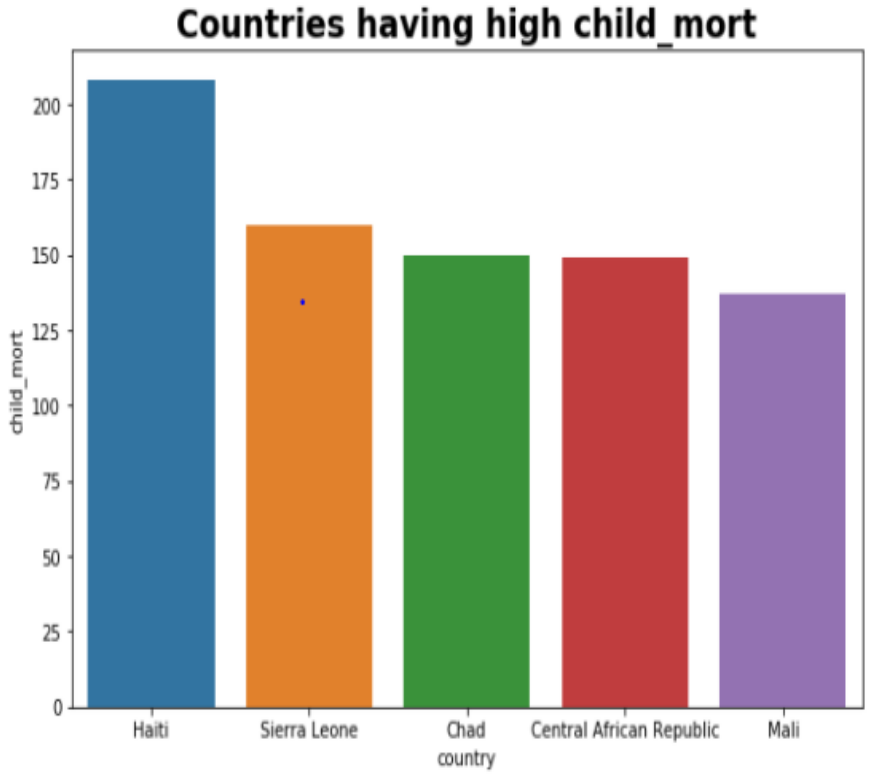
4. Clustering profiling using “gdpp, child_mort and income”

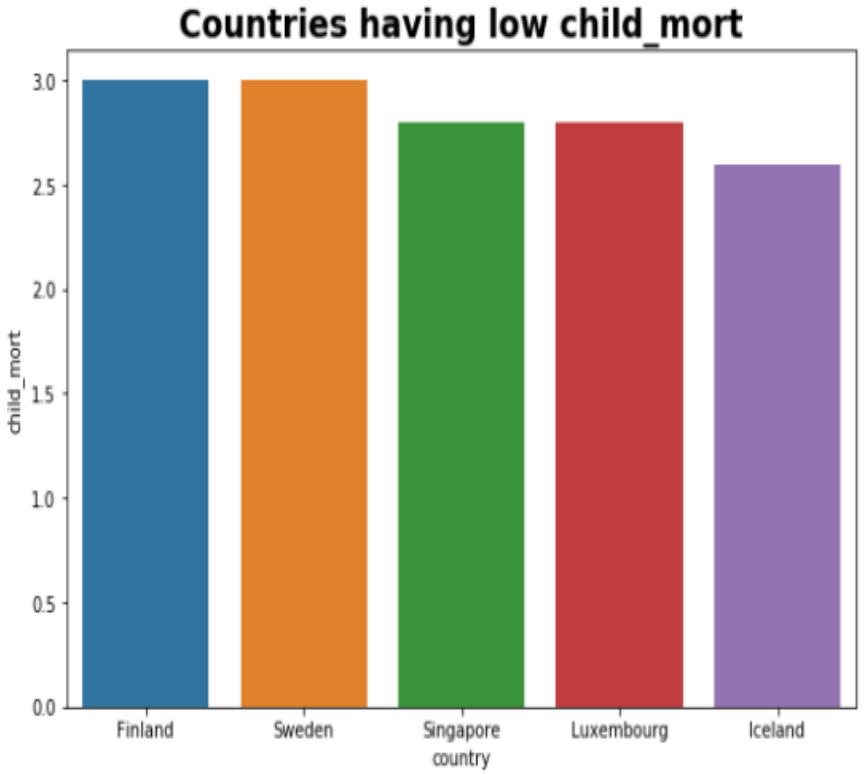
4. Country Identification

a. Based on the analysis, choose the countries that are in need for the aid.

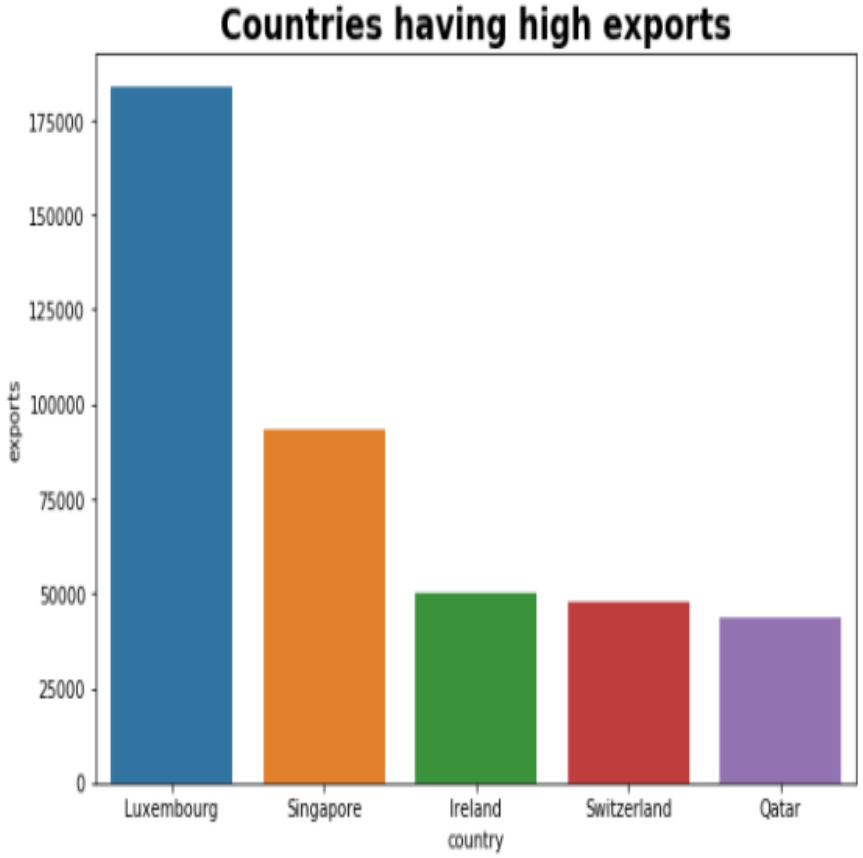
b. Choose the countries based on some socio-economic and health factors

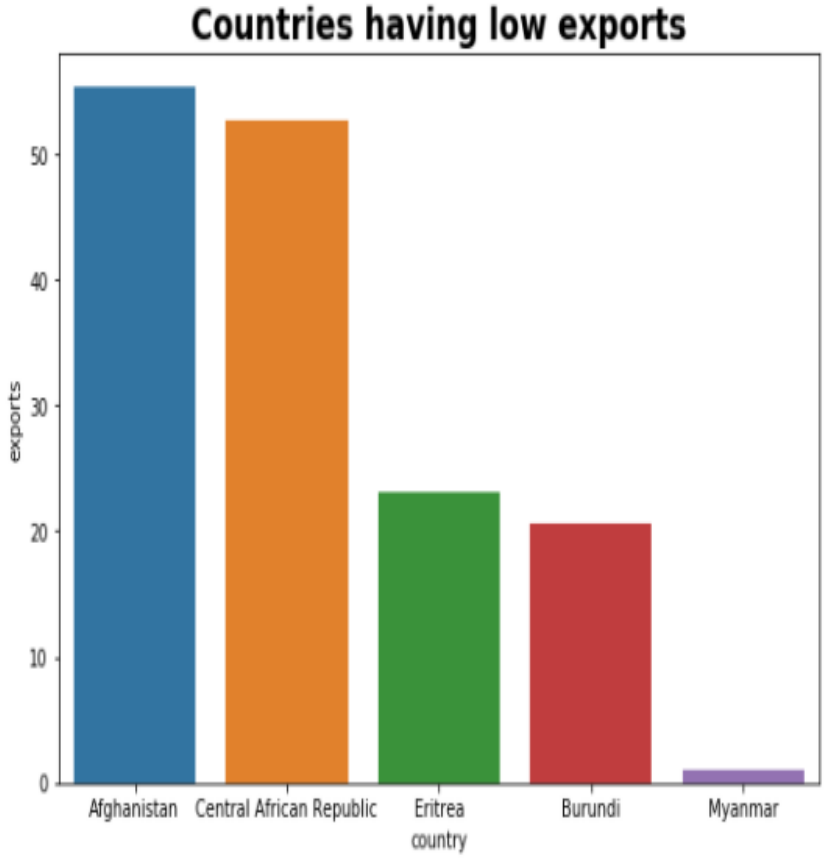
COUNTRIES WITH HIGH AND LOW CHILD MORTALITY



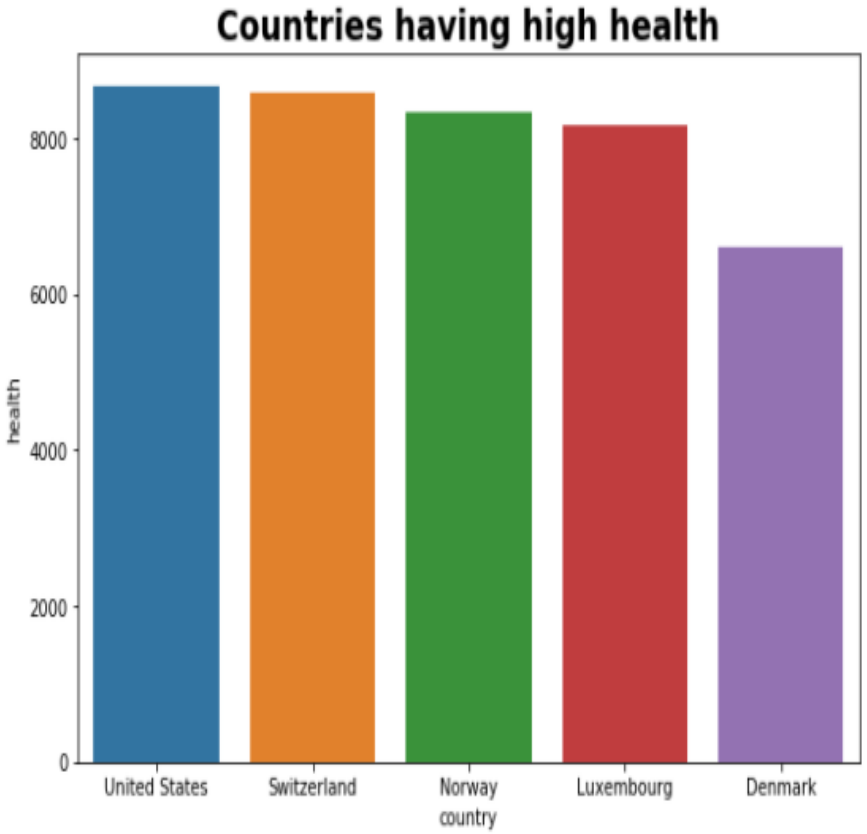


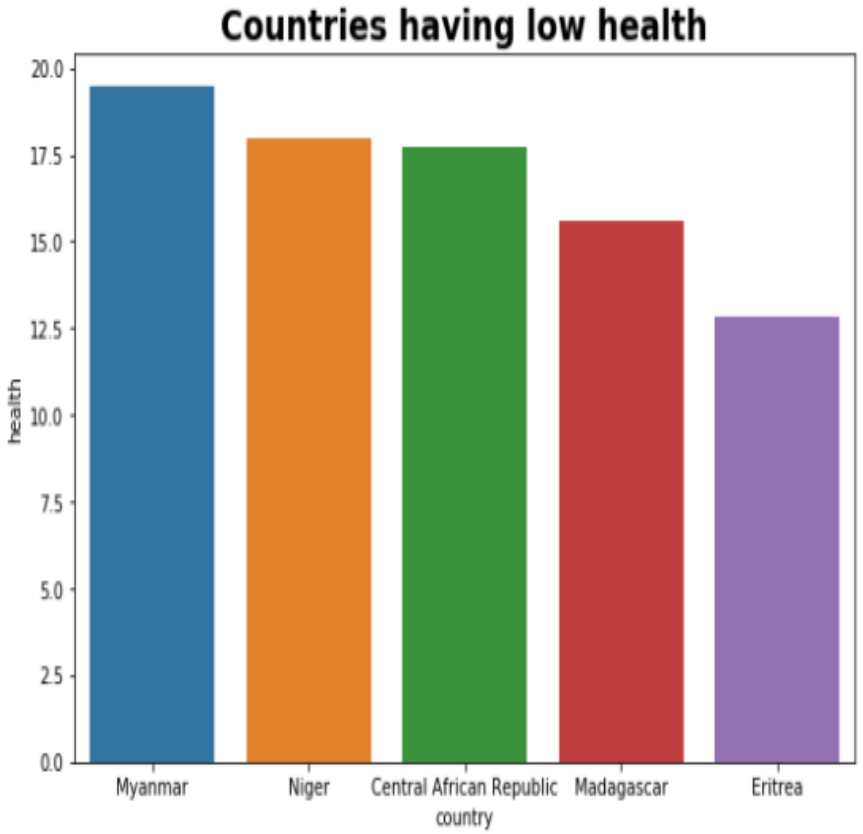
COUNTRIES WITH HIGH AND LOW EXPORTS



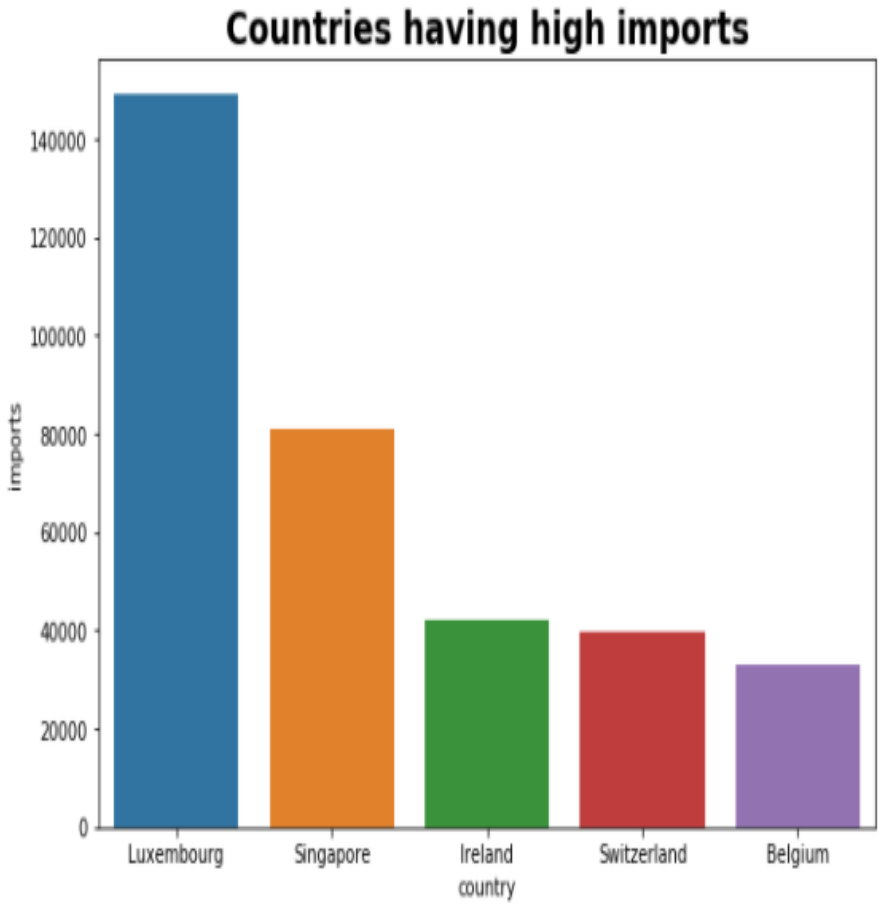


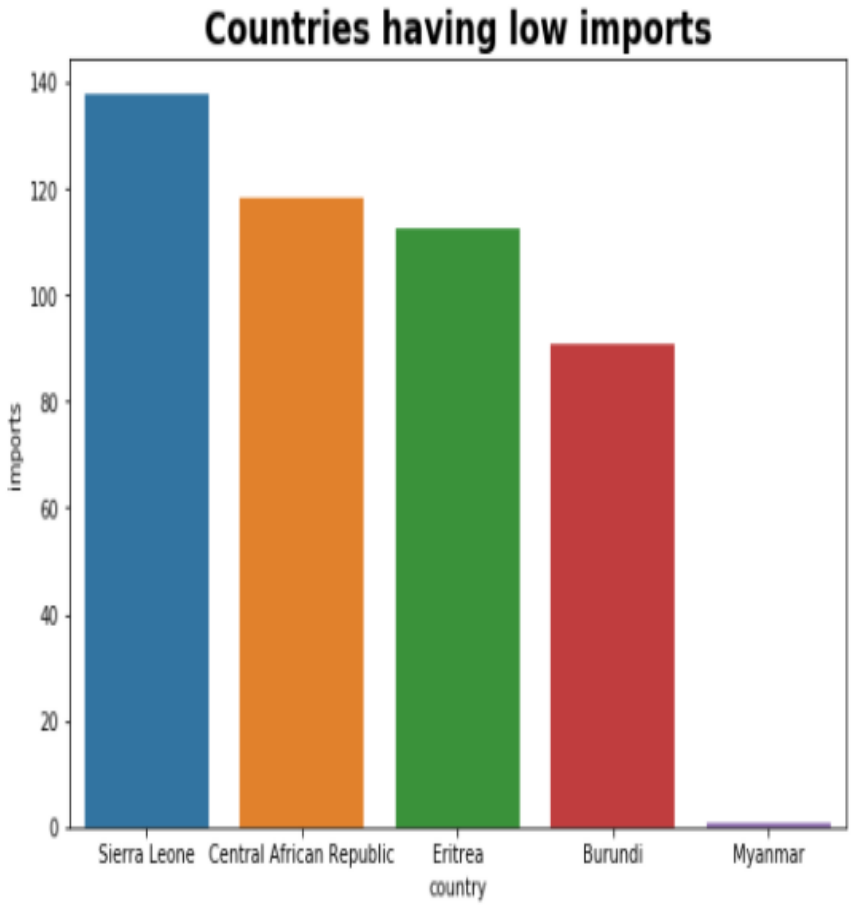
COUNTRIES WITH HIGH AND LOW HEALTH SPENDING PER CAPITA



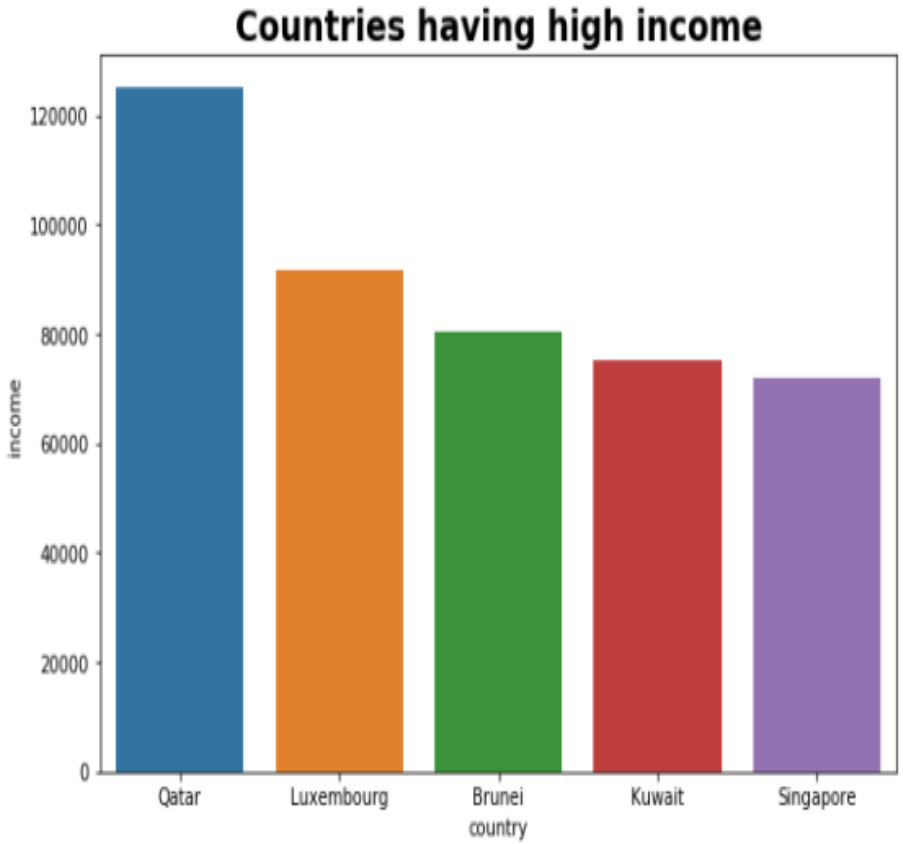


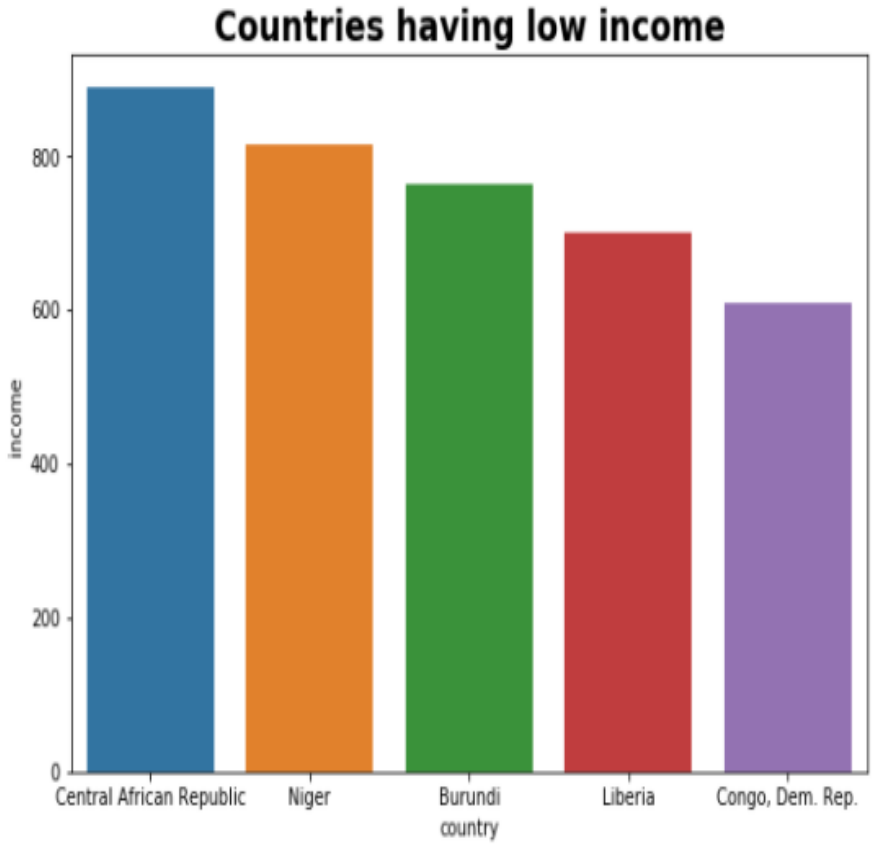
COUNTRIES WITH HIGH AND LOW IMPORTS



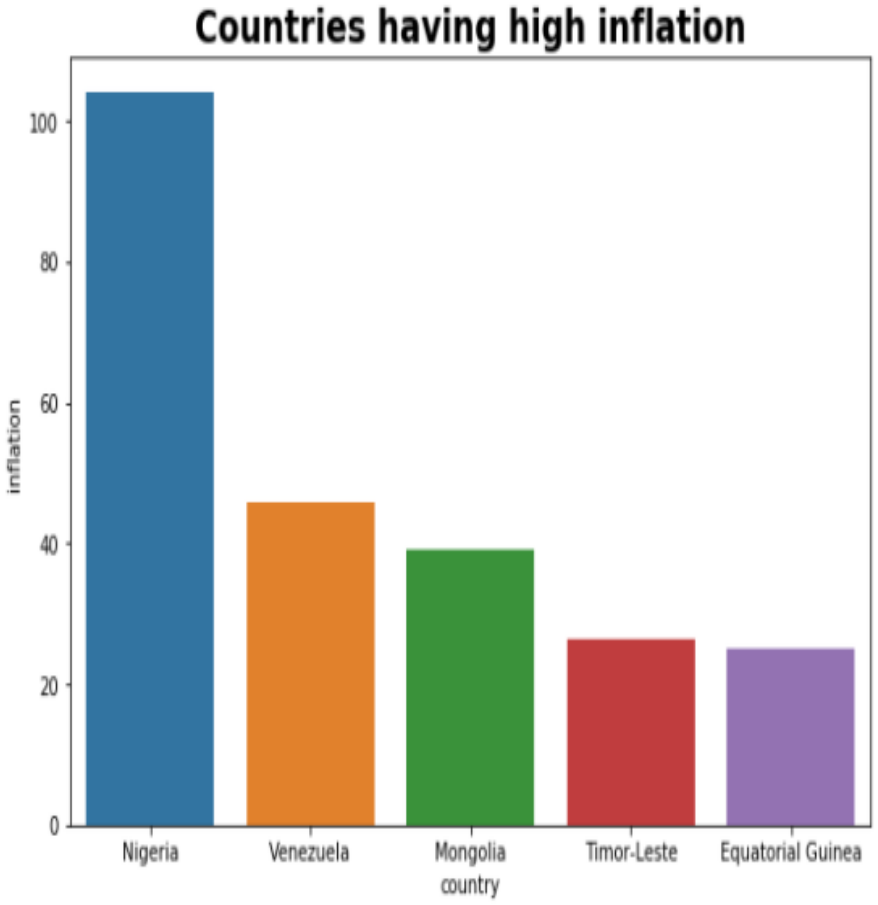


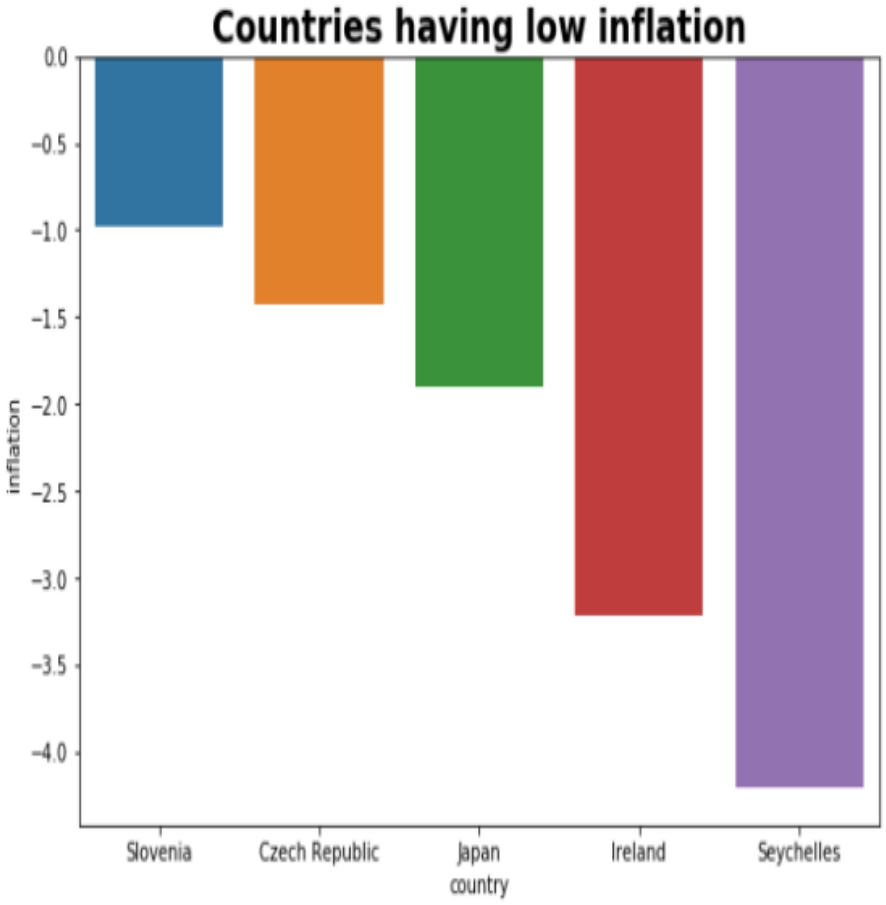
COUNTRIES WITH HIGH AND LOW INCOME



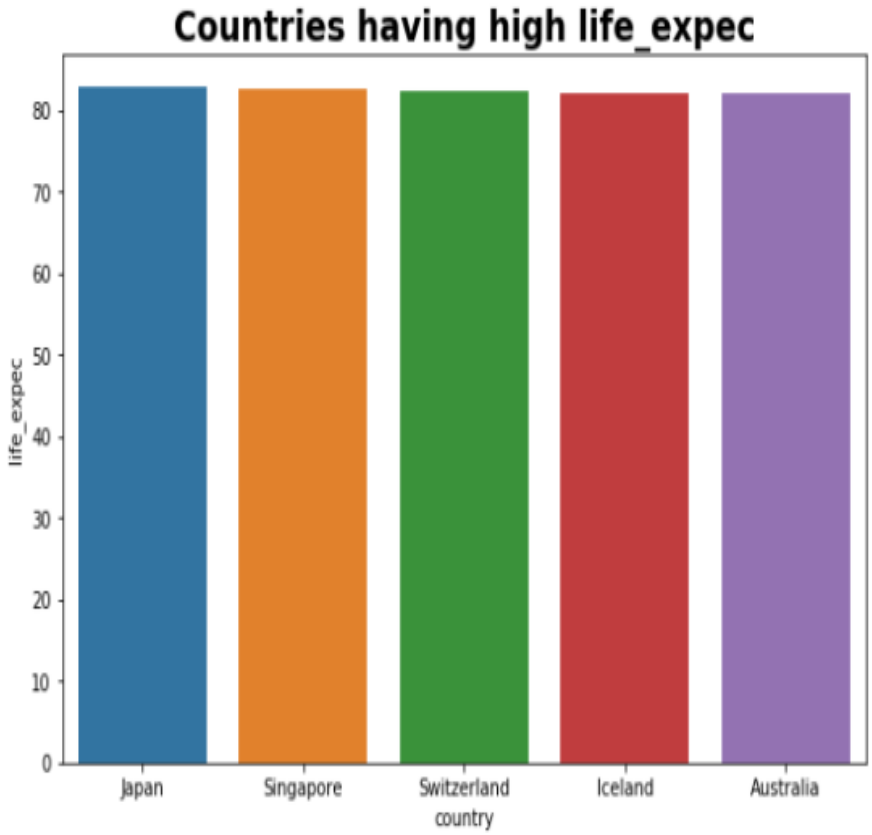


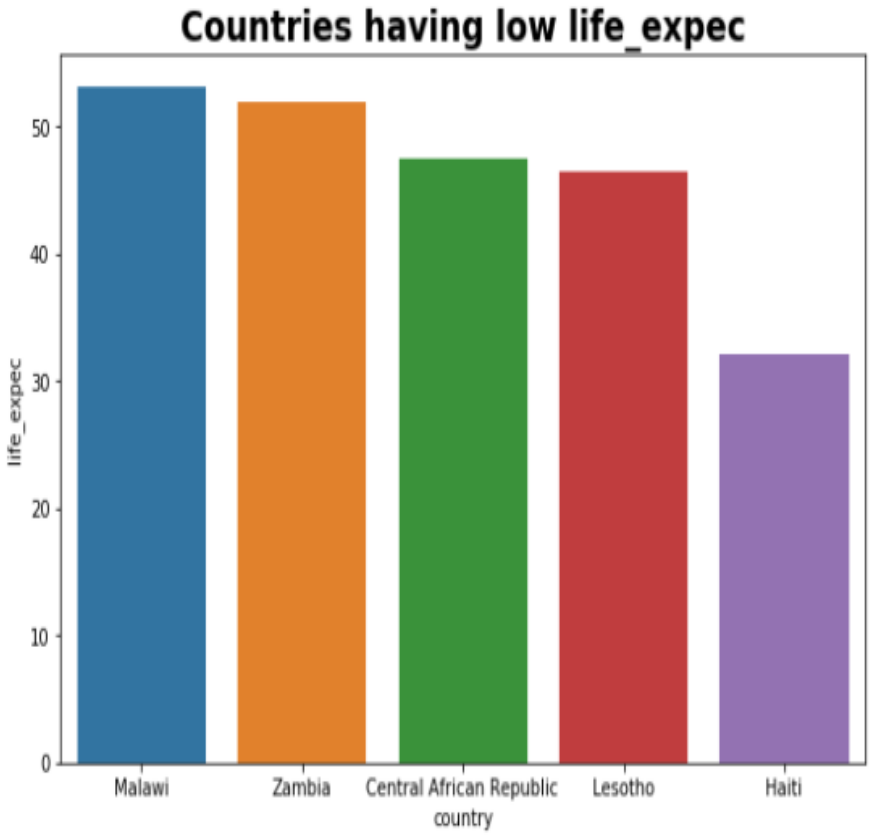
COUNTRIES WITH HIGH AND LOW INFLATION



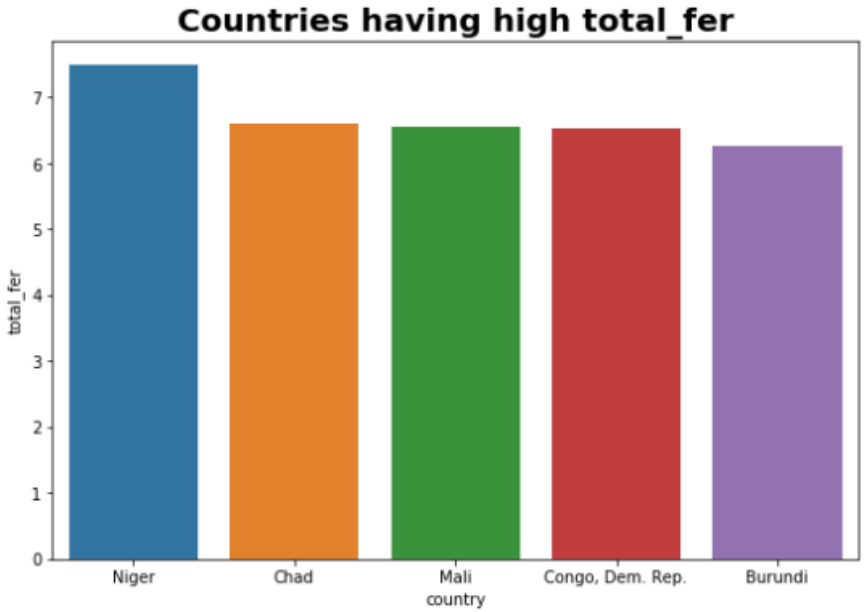


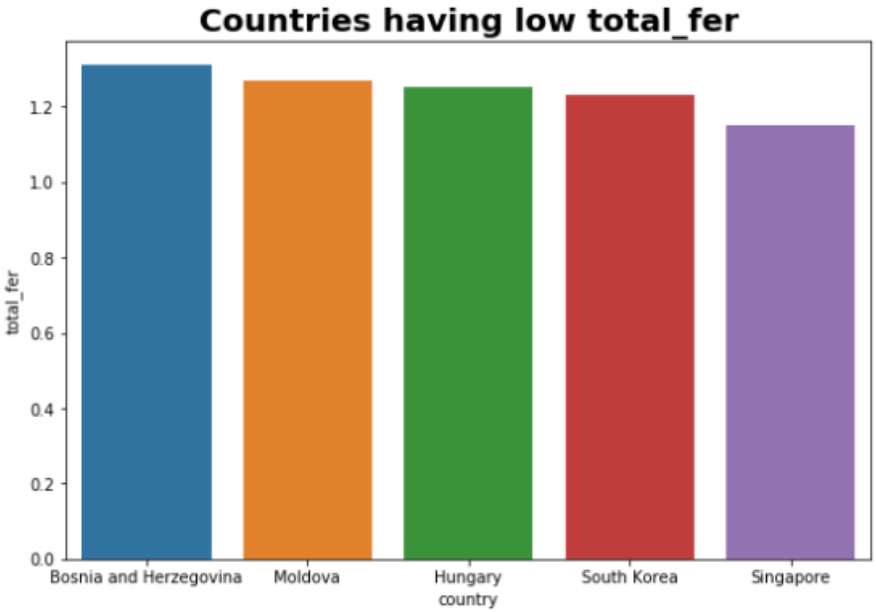
COUNTRIES WITH HIGH AND LOW LIFE EXPECTANCY



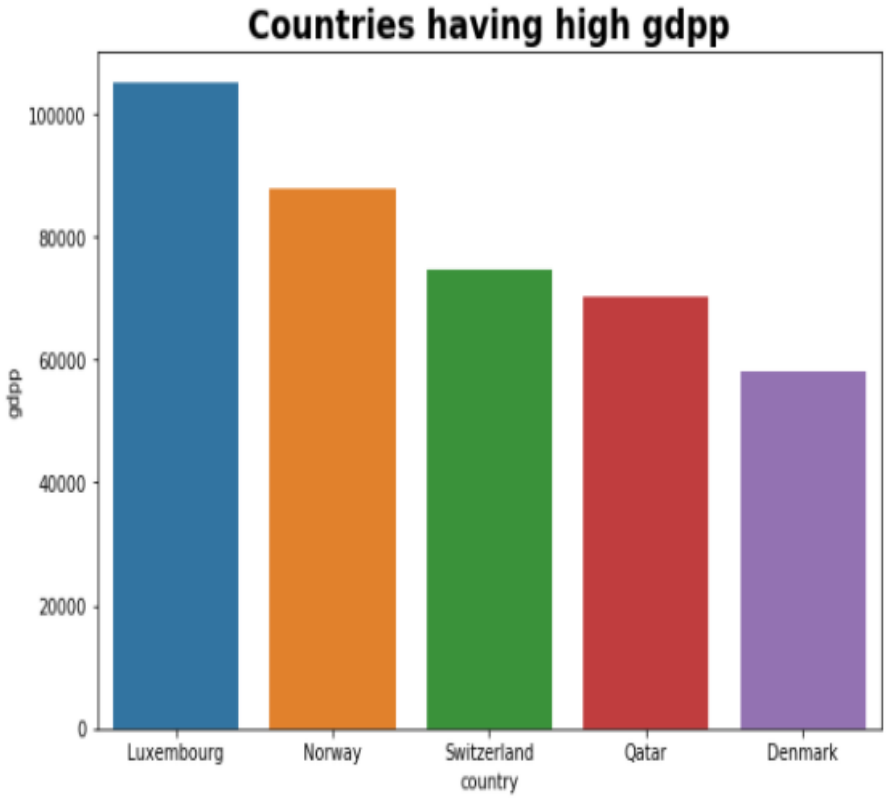


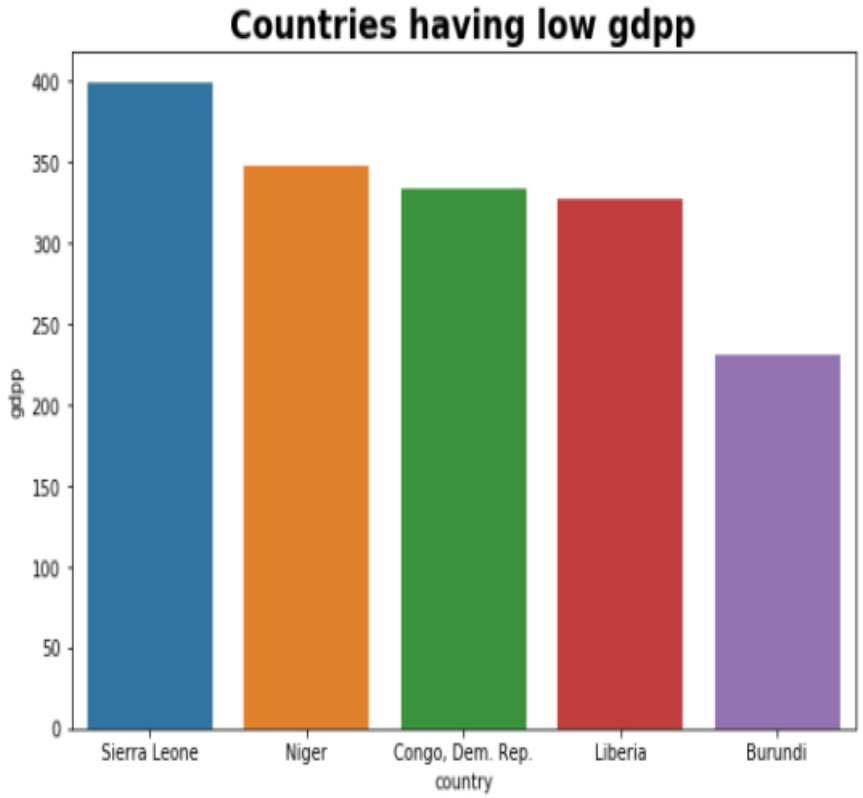
COUNTRIES WITH HIGH AND LOW FERTILITY



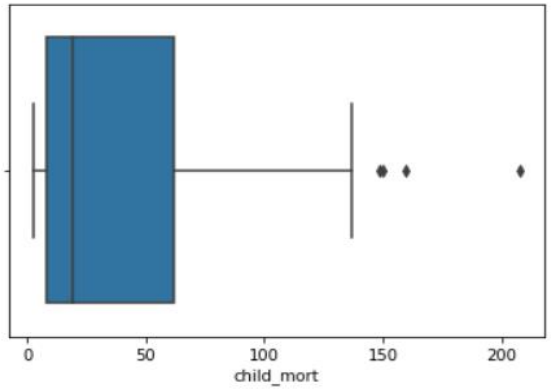


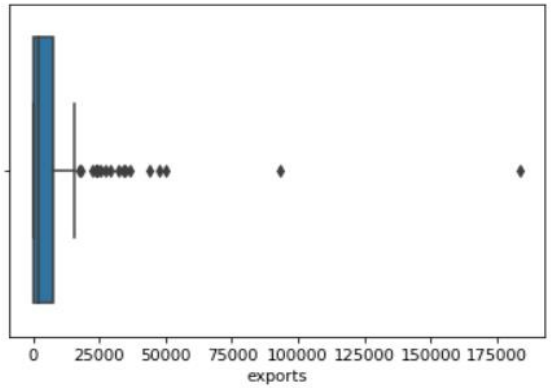
COUNTRIES WITH HIGH AND LOW GDPP

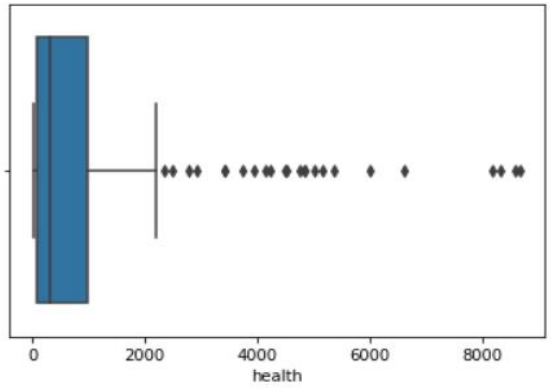


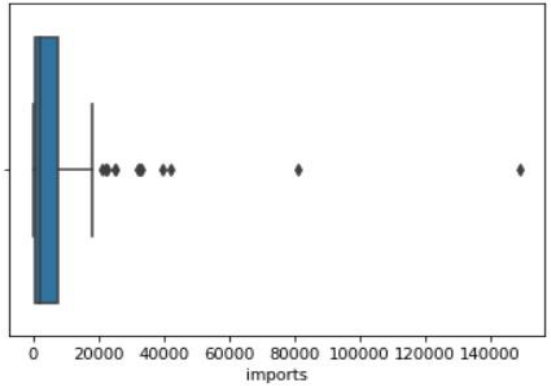


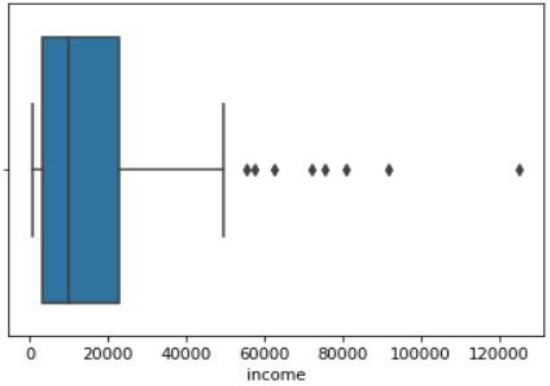
OUTLIER ANALYSIS

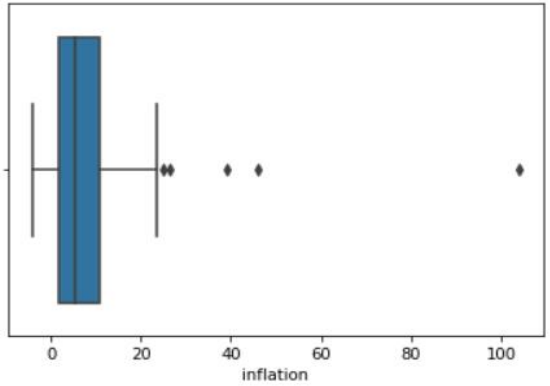


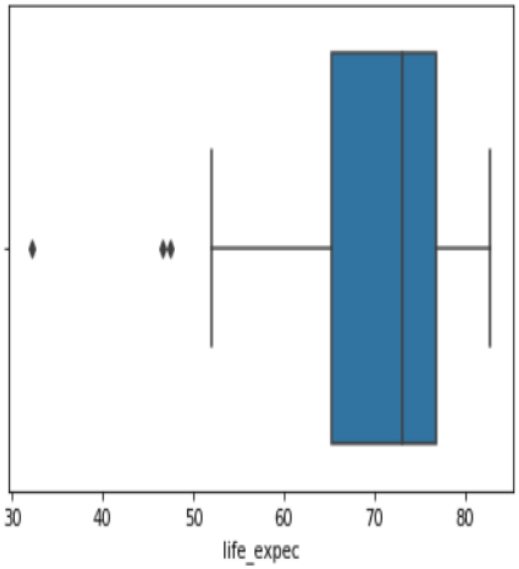


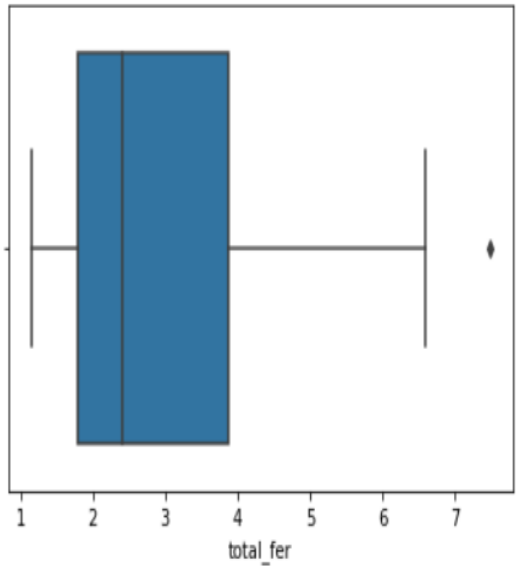


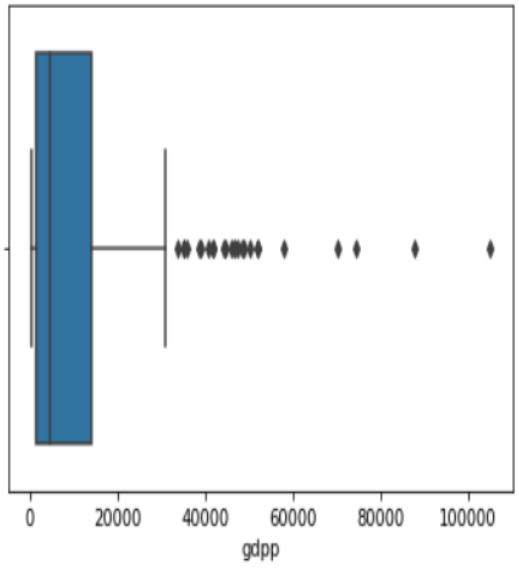








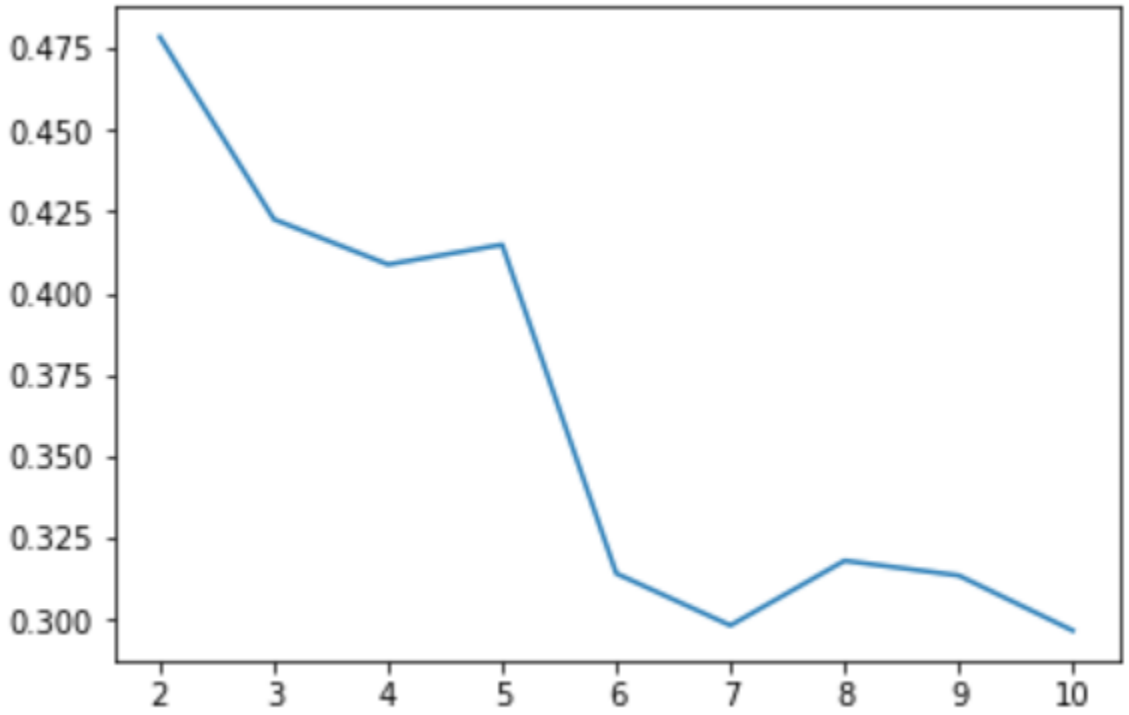




The numerical columns has outliers and soft capping was done to handle these values.

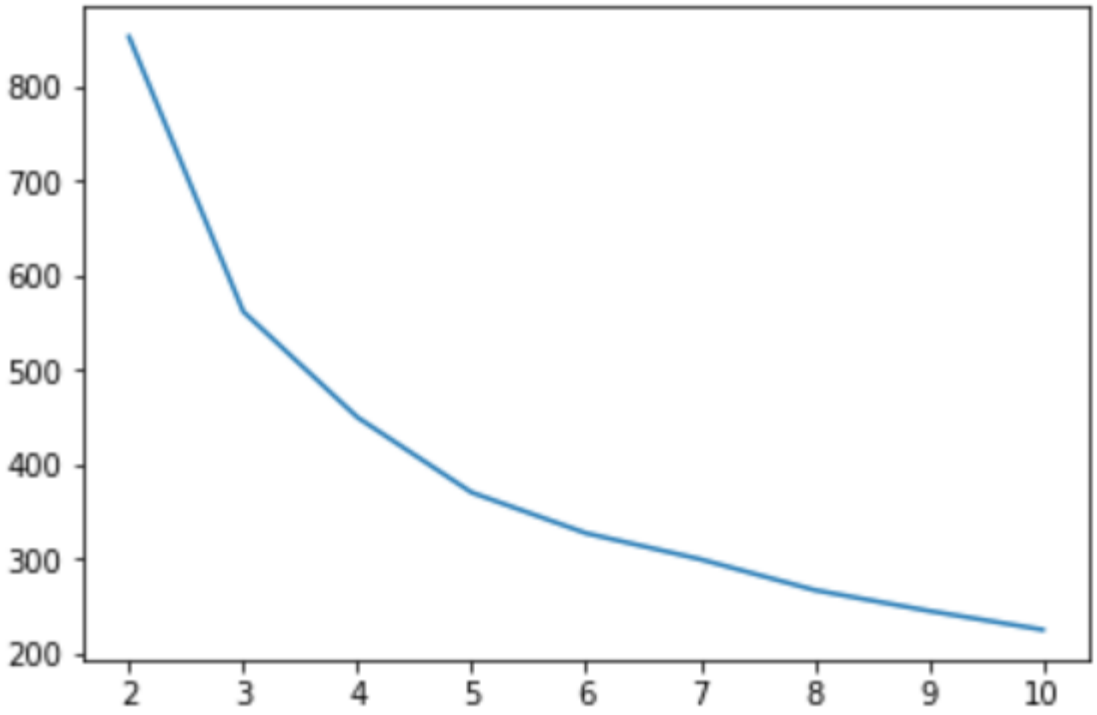
K MEANS CLUSTERING

FINDING THE VALUE OF K USING SILHOUETTE SCORE



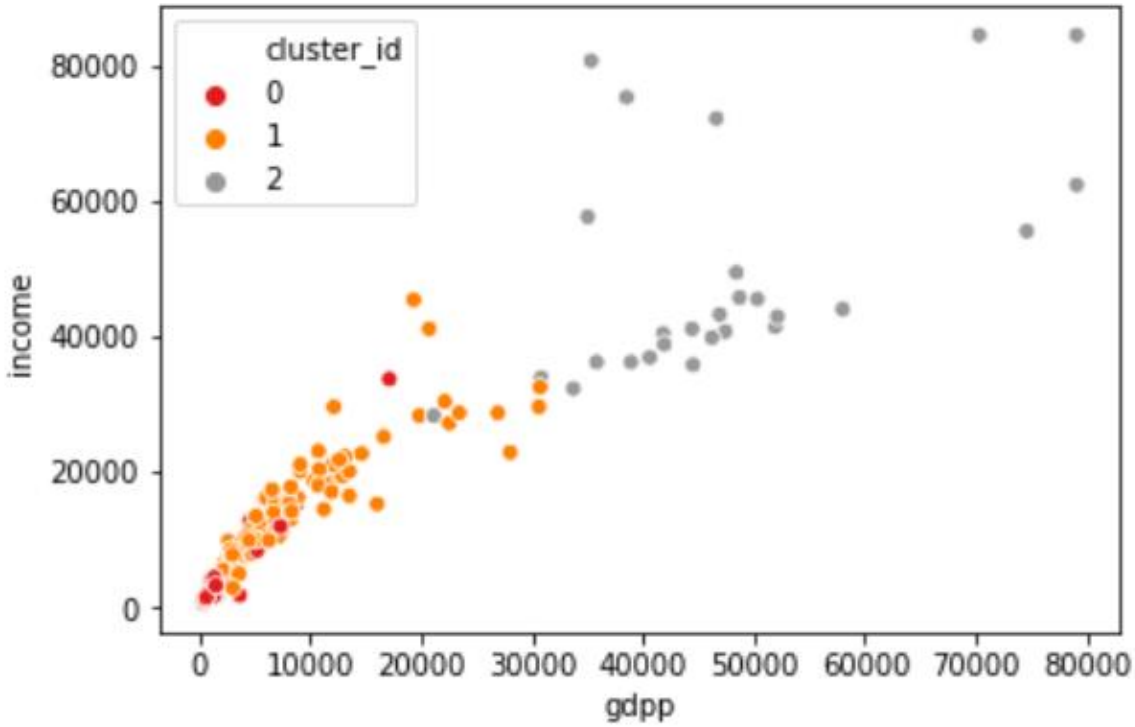
K value was selected as 3 since it was the second highest value after 2

FINDING THE VALUE OF K USING ELBOW CURVE METHOD



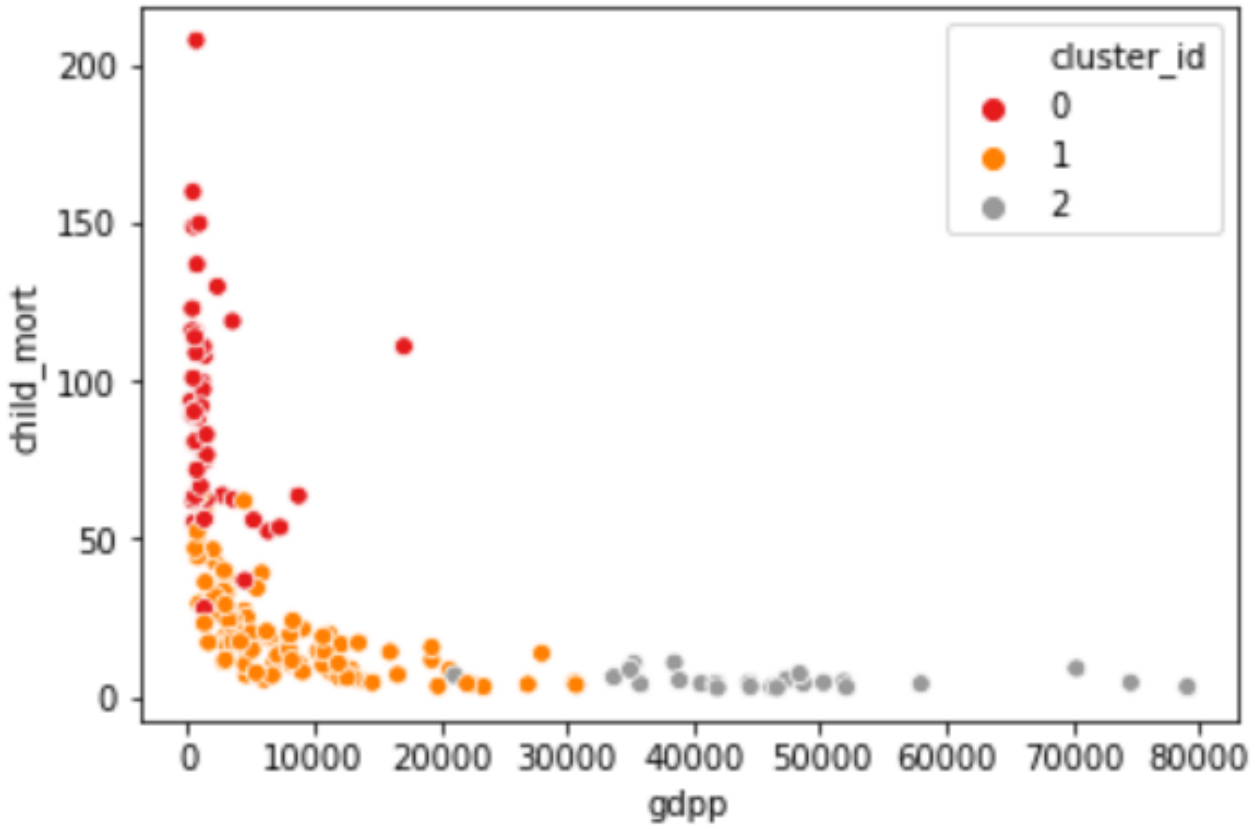
K value was selected as 3 since there was a sudden change in the curve at that point

GDPP Vs INCOME



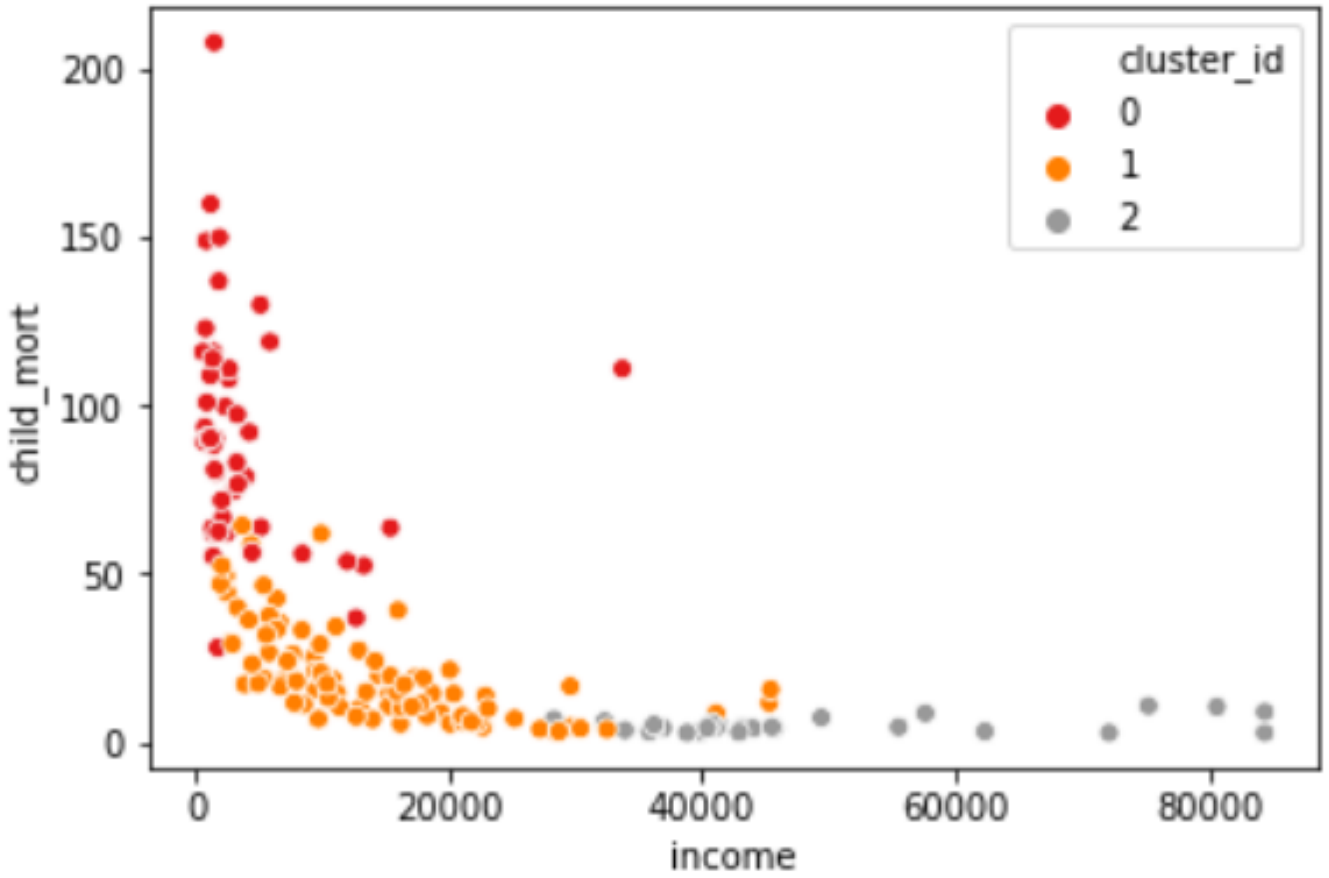
Cluster 0 shows the countries with low gdpp and low income

GDPP Vs CHILD MORTALITY



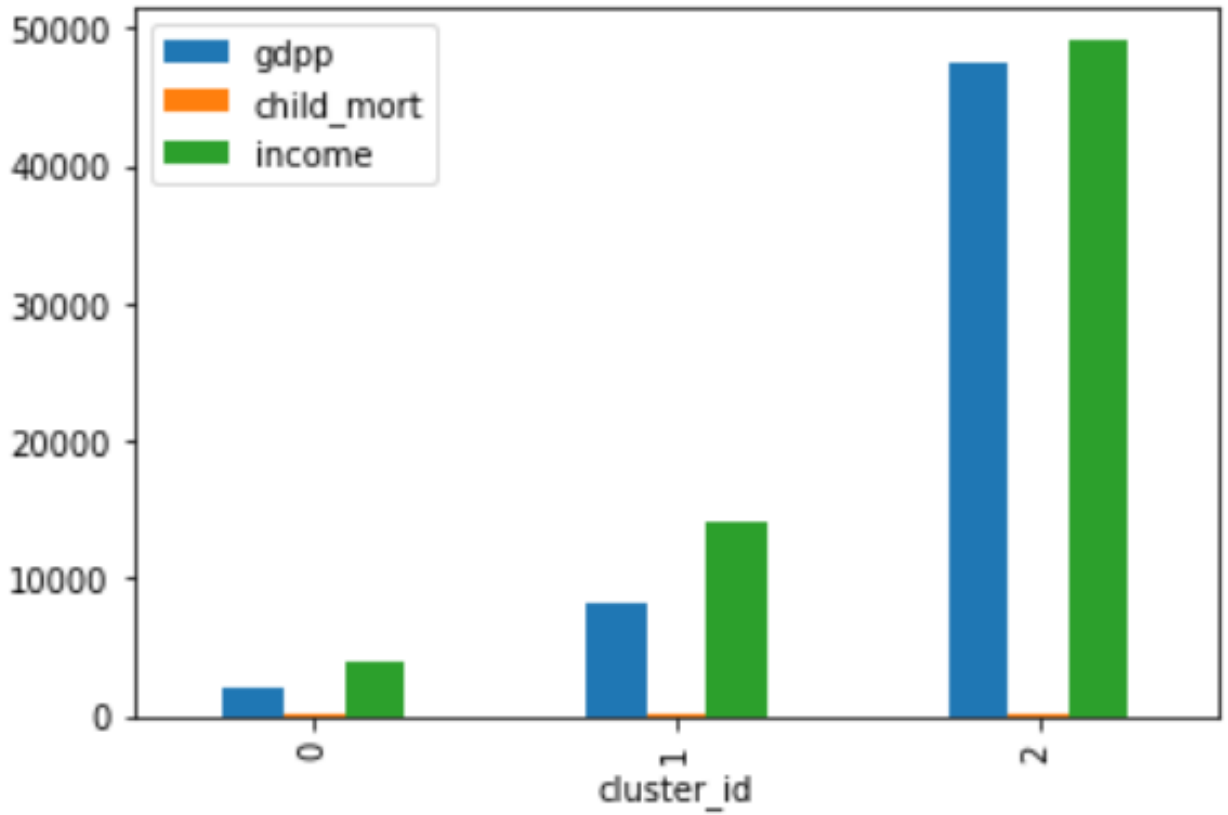
Cluster 0 shows the countries with low gdpp and high child mortality

INCOME Vs CHILD MORTALITY



Cluster 0 shows the countries with low income and high child mortality

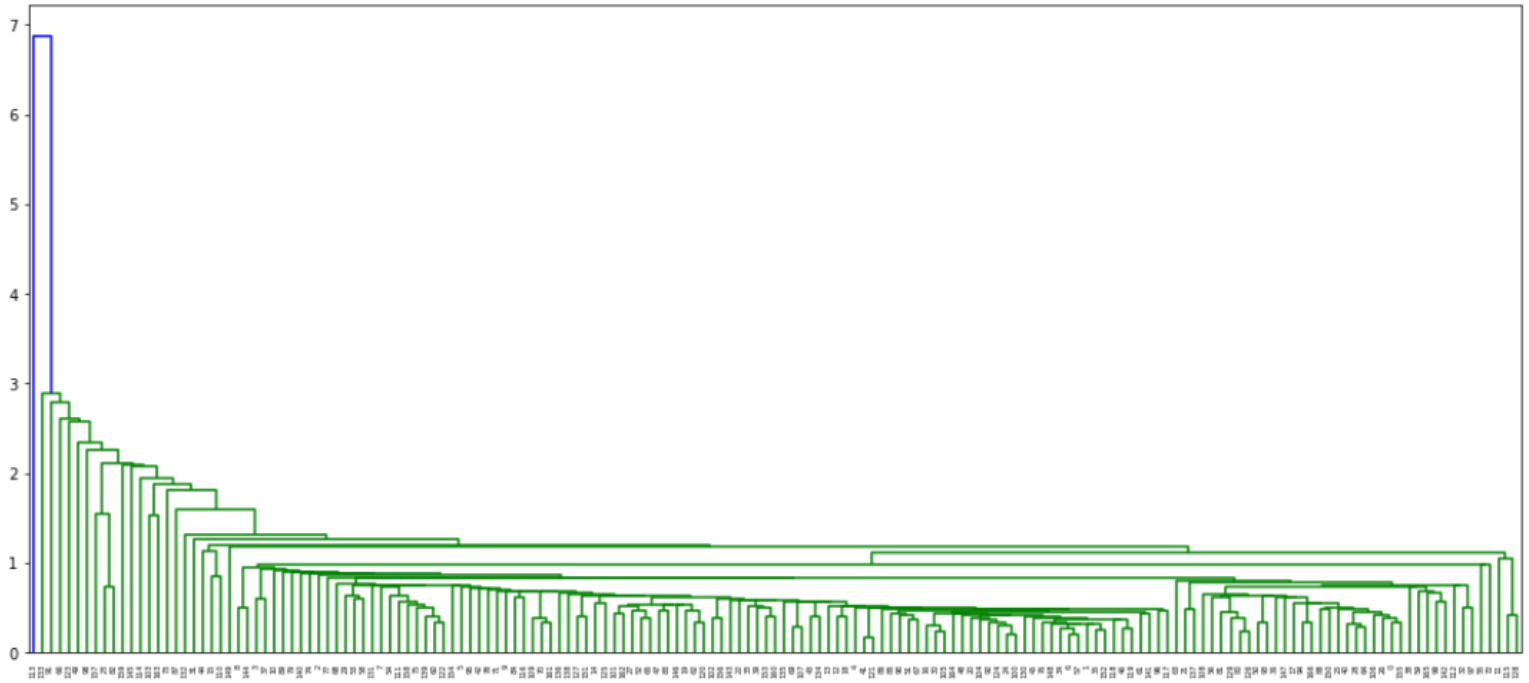
CLUSTERS BY K MEANS METHOD



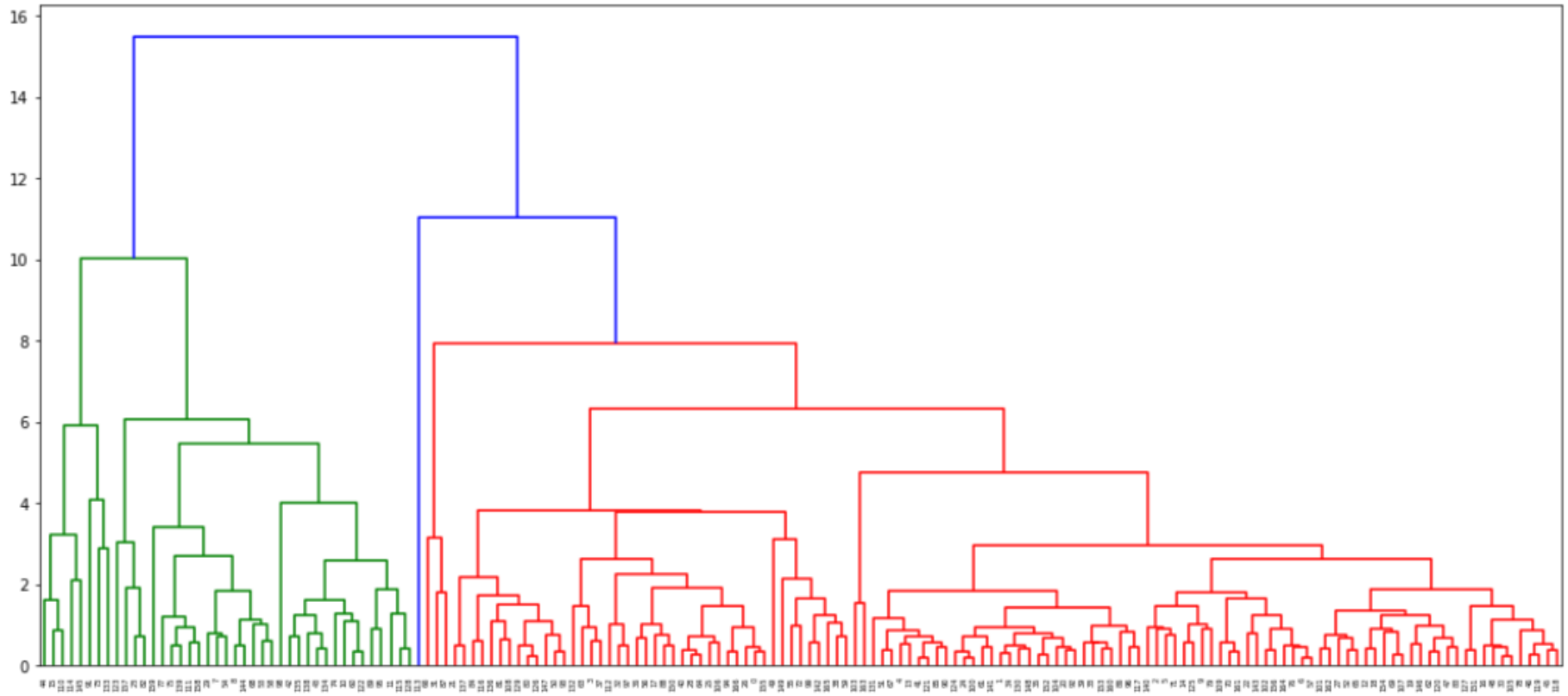
Cluster 0: Underdeveloped countries
Cluster 1: Developing Countries
Cluster 2: Developed countries

HIERARCHICAL CLUSTERING

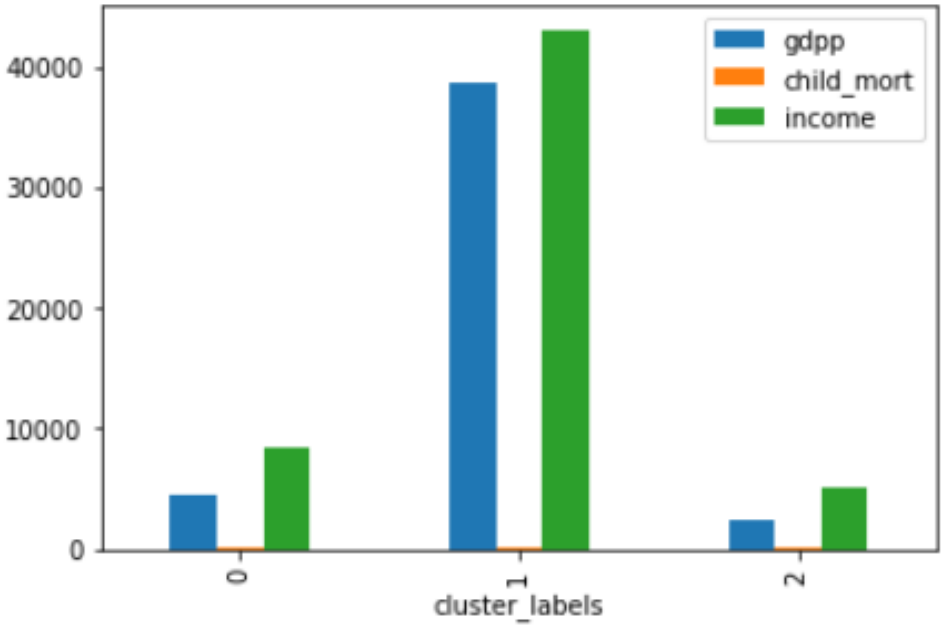
SINGLE LINKAGE METHOD



COMPLETE LINKAGE METHOD



CLUSTERS BY HIERARCHICAL CLUSTERING



- The number of clusters was selected as 3
- Only one data point in cluster 2
- Not a proper form of clustering
- Cluster 0 shows the underdeveloped countries which require direst aid

CONCLUSIONS

The results from Kmeans method can be considered as the final suggested countries

Hierarchical clustering method is not showing proper clusters here. It divides the whole data into two halves

Cluster 0 of the k means method shows the under developed countries with low income, low gdpp and high child mortality

The cluster 0 contains 48 countries, out of this 10 countries are suggested as the top 10 countries which are in direst need of aid

Based on business aspect the number of clusters can be increased as it will give clusters with less number of countries

Countries which need direst aid can be found in a better way by increasing the number of clusters

THE TOP 10 COUNTRIES IN DIREST NEED OF AID

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