HELP INTERNATIONAL – NGO
FIGHTING POVERTY

Author : Jithin Prakash K
Clustering Assignment
DSC June 2020

PROBLEM STATEMENT.

HELP International – NGO, has an objective to eliminate poverty by providing the people of backward countries with basic amenities and relief during the time of disasters and natural calamities.

A sum of $10 Million fund is raised recently. This amount needs to be strategically and effectively used across the globe. An analysis need performed to find the countries that are in direst need of money.
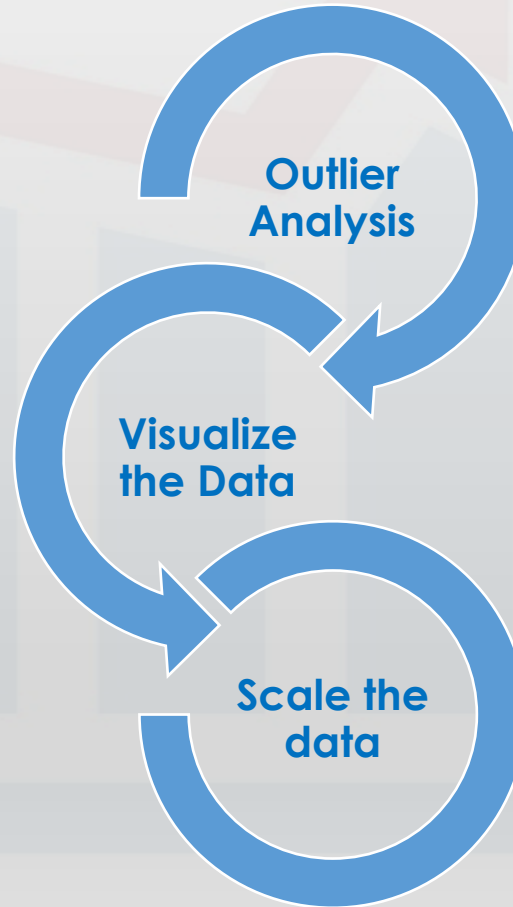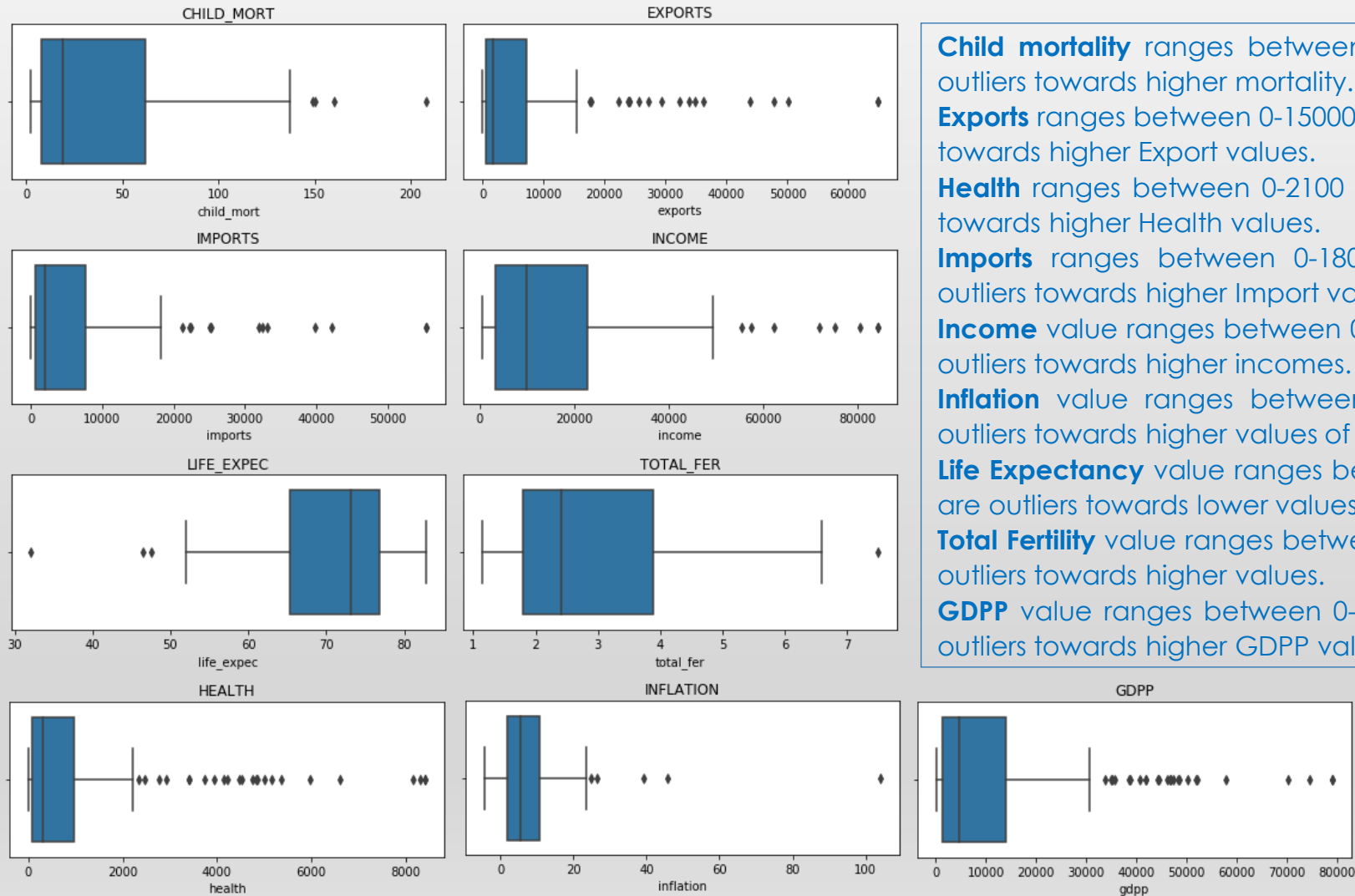
ANALYSIS APPROACH

Step 1

- Read the Data
- Analyse the data
- Prepare the data

Step 2

- Outlier Analysis
- Visualize the Data
- Scale the data

Step 3

- Hopkins Readiness
- K Means Clustering
- Hierarchical Clustering

Step 4

- Decision Making

BOX PLOTS FOR THE FEATURES

Child mortality ranges between 0-145 with median 20 and there are outliers towards higher mortality.

Exports ranges between 0-15000 with median 1700 and there are outliers towards higher Export values.

Health ranges between 0-2100 with median 320 and there are outliers towards higher Health values.

Imports ranges between 0-18000 with median 2050 and there are outliers towards higher Import values.

**Income** value ranges between 0-50000 with median 9950 and there are outliers towards higher incomes.

Inflation value ranges between -4-25 with median 5 and there are outliers towards higher values of inflation.

Life Expectancy value ranges between 51-84 with median 72 and there are outliers towards lower values.

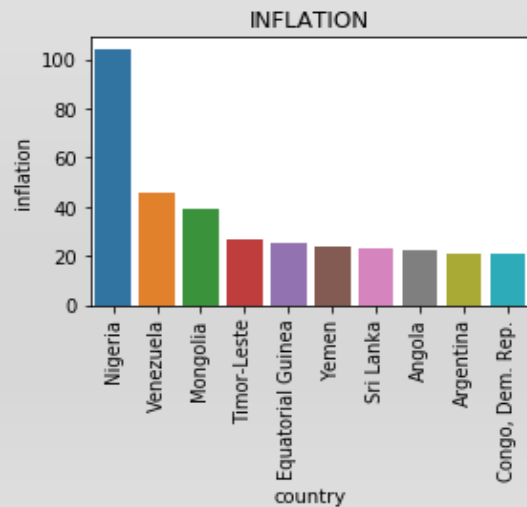**Total Fertility** value ranges between 1-6.5 with median 2.4 and there are outliers towards higher values.

GDPP value ranges between 0-30000 with median 5000 and there are outliers towards higher GDPP values.

TOP 10 DEVELOPED COUNTRIES - EDA
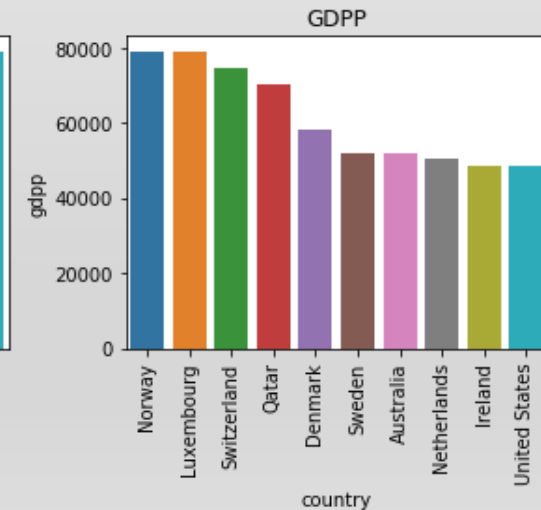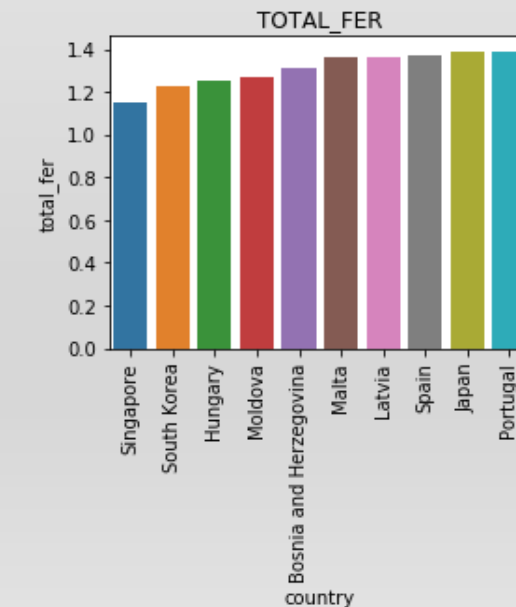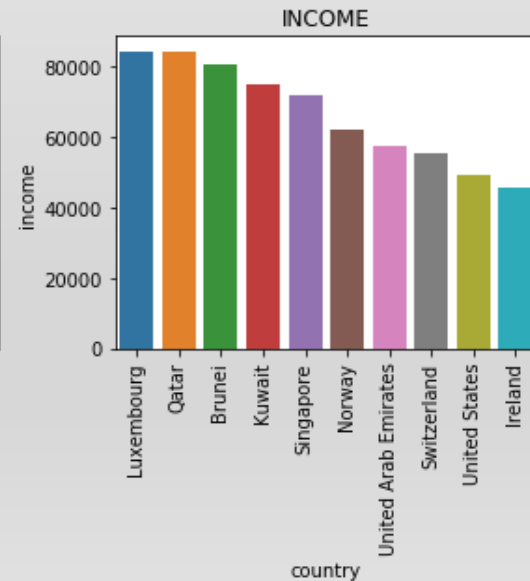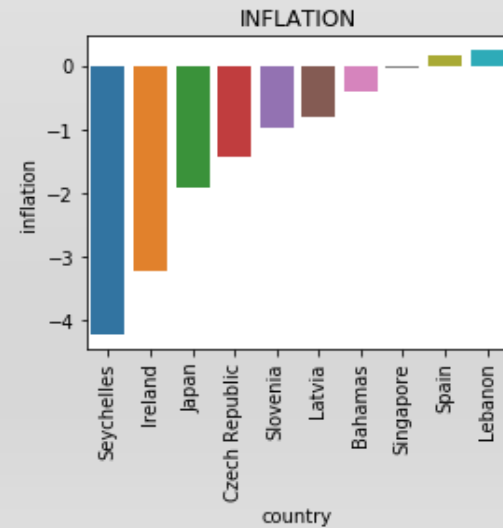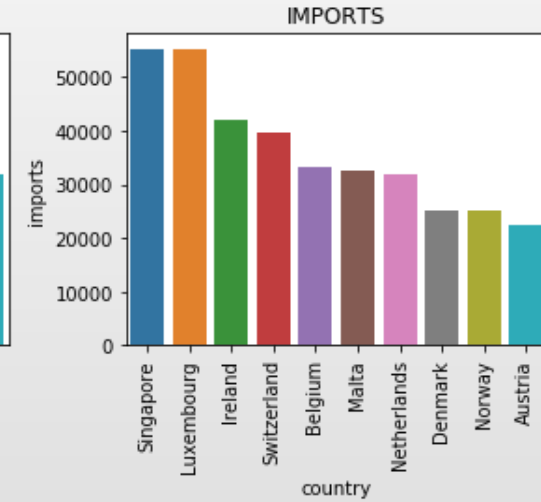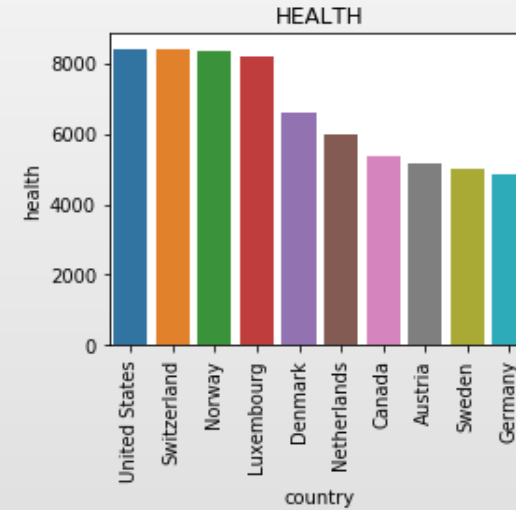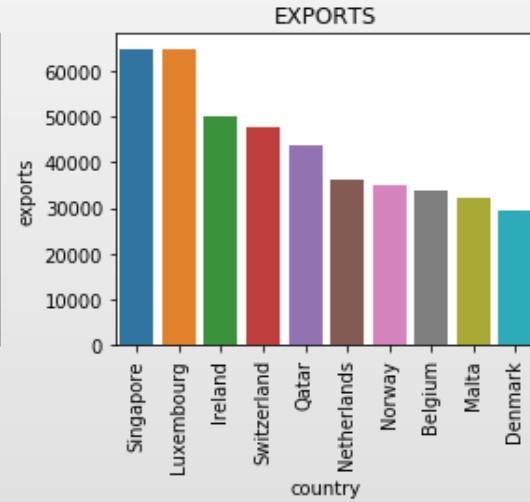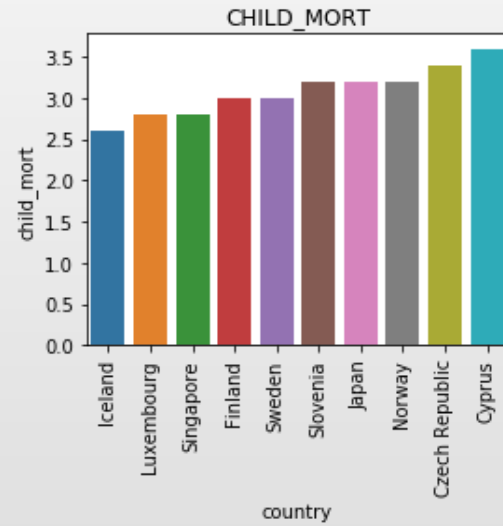
TOP 10 UNDER DEVELOPED COUNTRIES - EDA
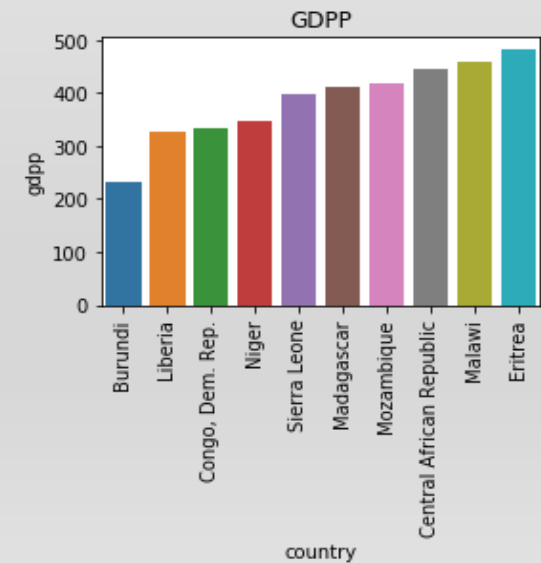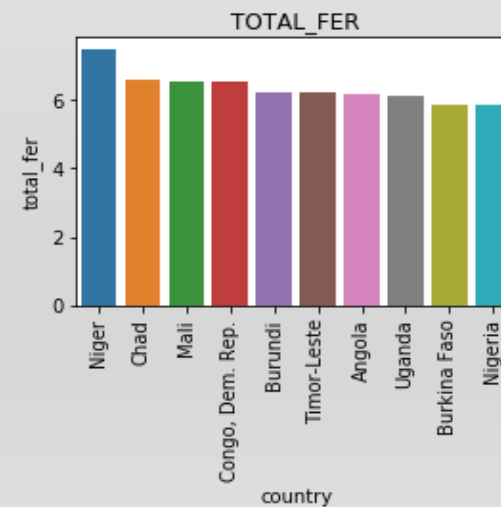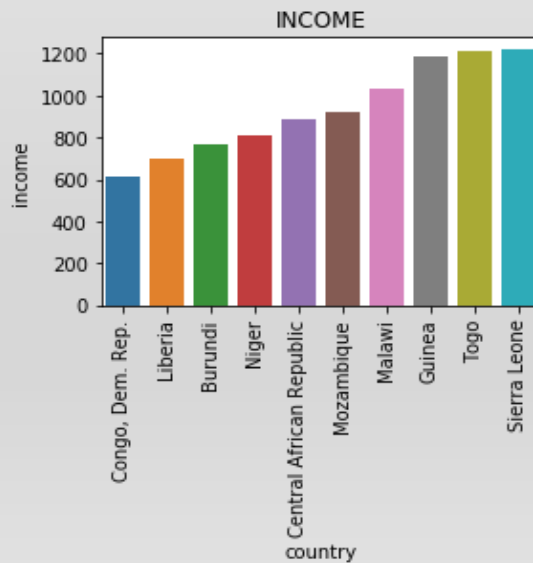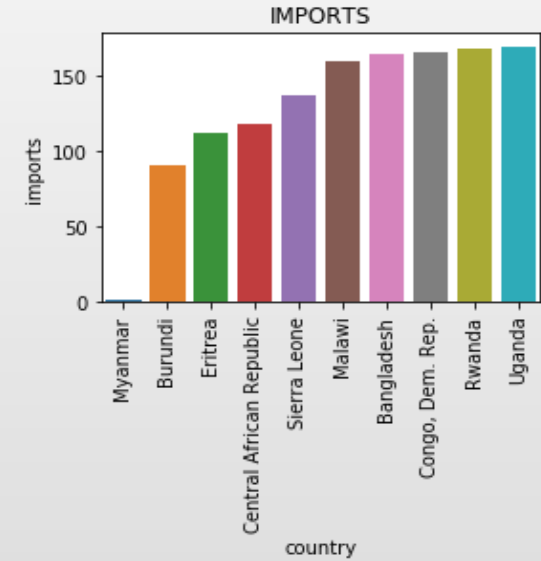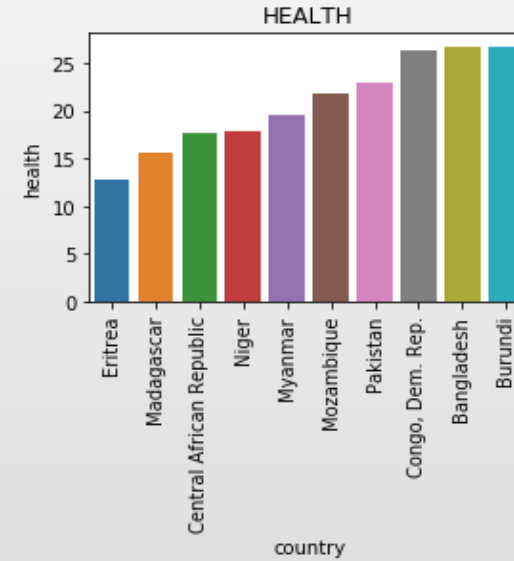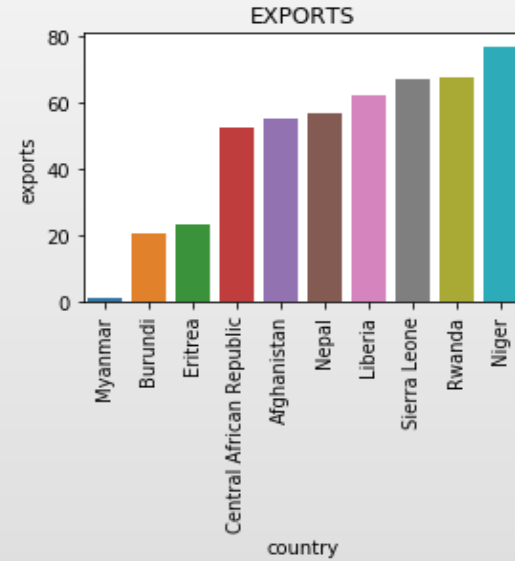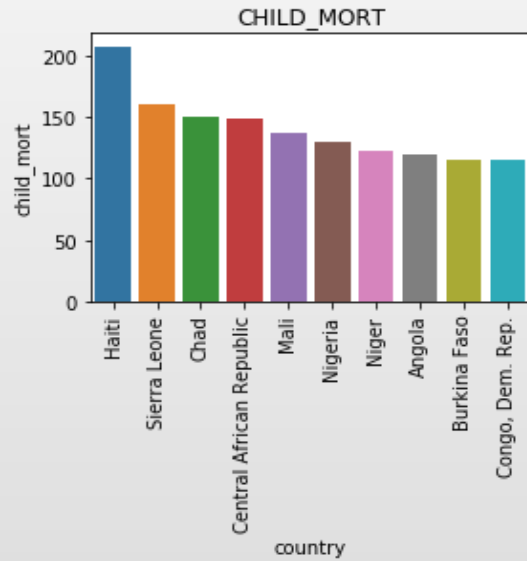
CORRELATION HEATMAP
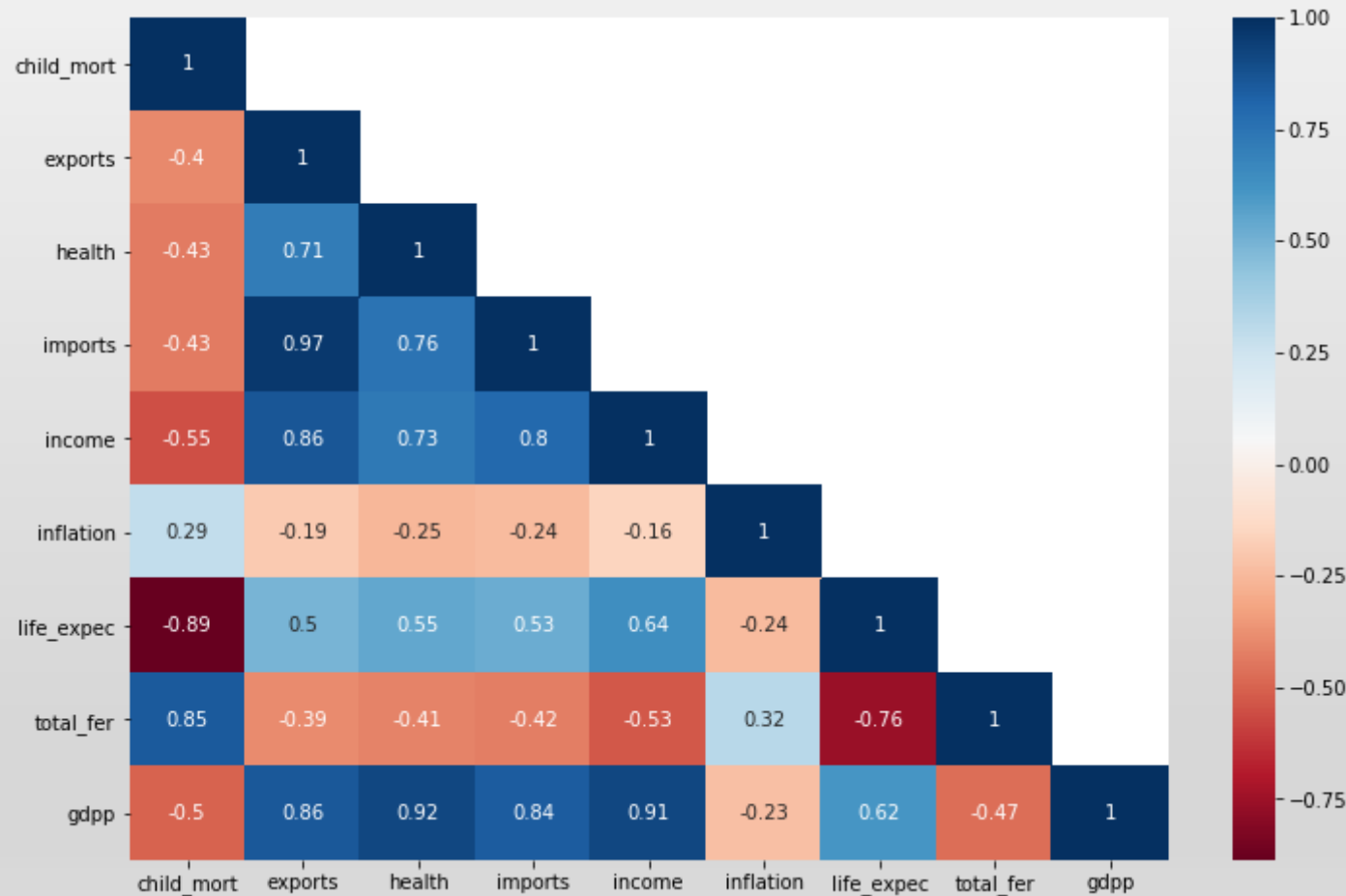
Life expectancy and **child mortality** are highly negatively correlated.

Total fertility and **child mortality** are highly positively correlated.

Life expectancy and **total fertility** are negatively correlated.

Imports and **exports** are positively correlated.

Income and **exports** are highly positively correlated.

GDPP is highly positively correlated with **Income, Imports, Health and Exports**

K-MEANS CLUSTERING (1/4)…

Elbow-curve / Sum of Squared Distance   **Silhouette Score**
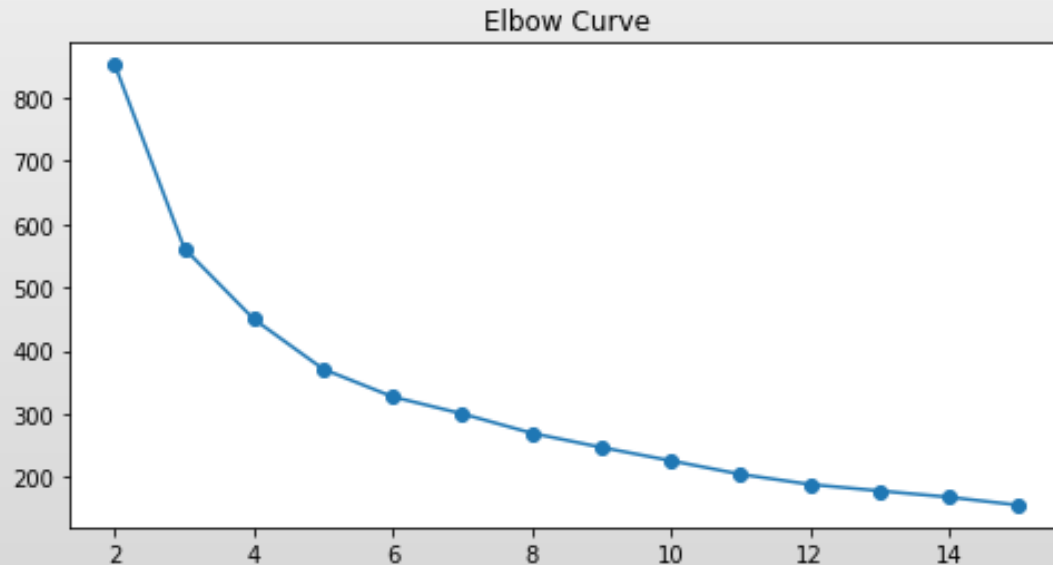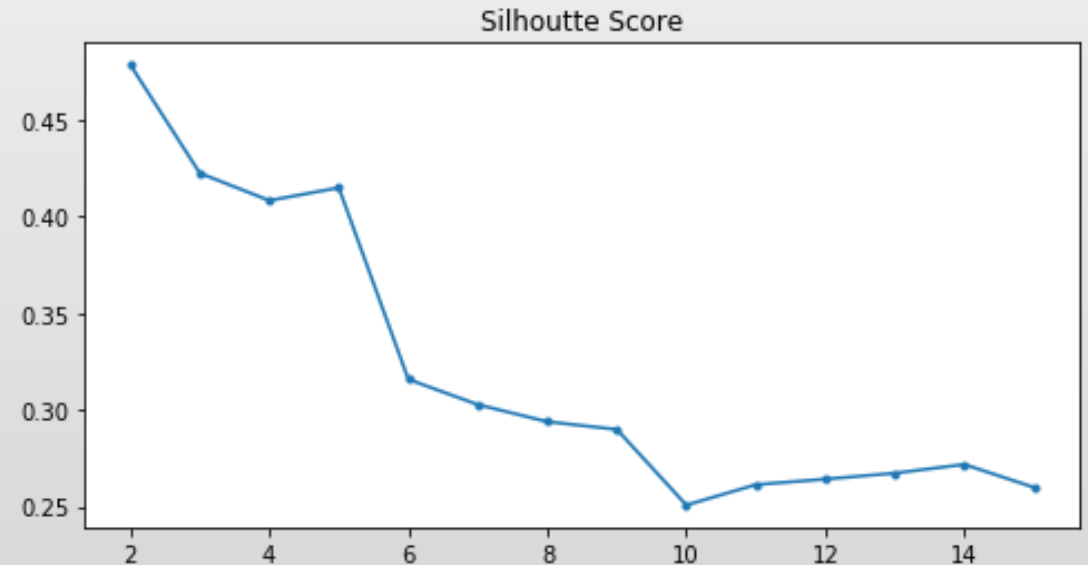
Both Elbow curve and Silhouette Score shows that number of clusters (k) = 3 is apt number for the clusters.

K-MEANS CLUSTERING (2/4)…
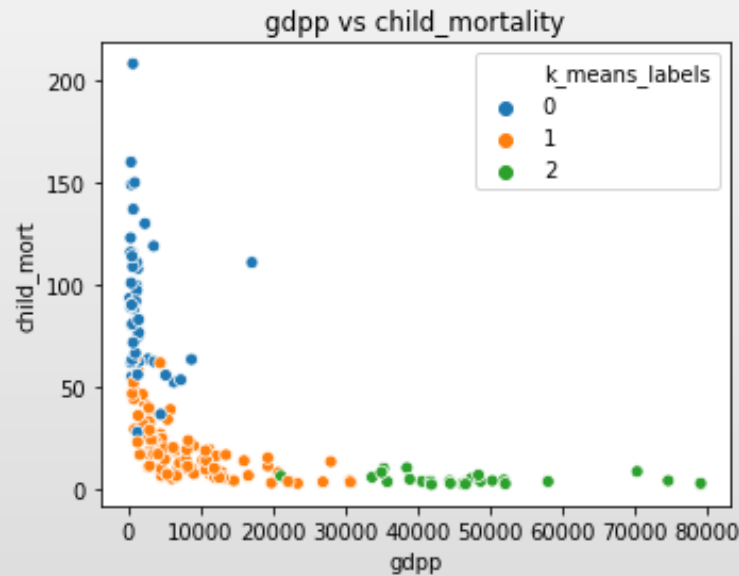
Cluster 0 is distributed in low GDPP and high child mortality.

Cluster 1 is distributed in low GDPP and low child mortality.

Cluster 2 is distributed in high GDPP and low child mortality.
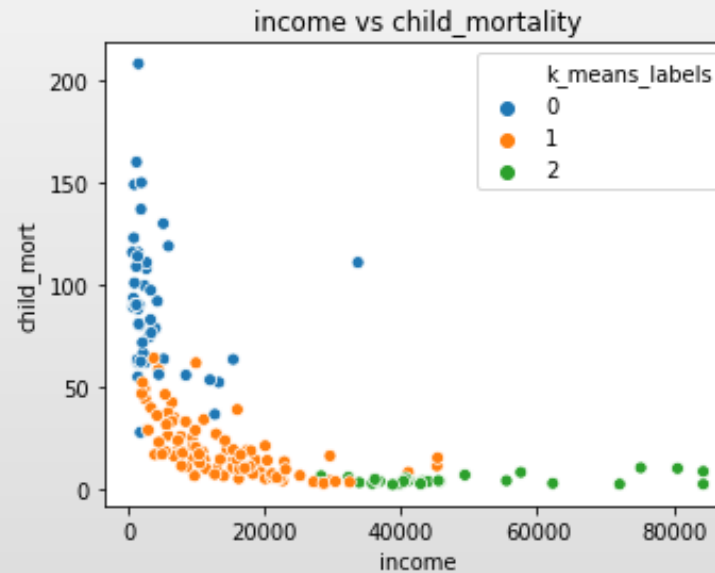
Cluster 0 is distributed in low income and high child mortality.

Cluster 1 is distributed in low income and low child mortality.

Cluster 2 is distributed in high income and low child mortality.
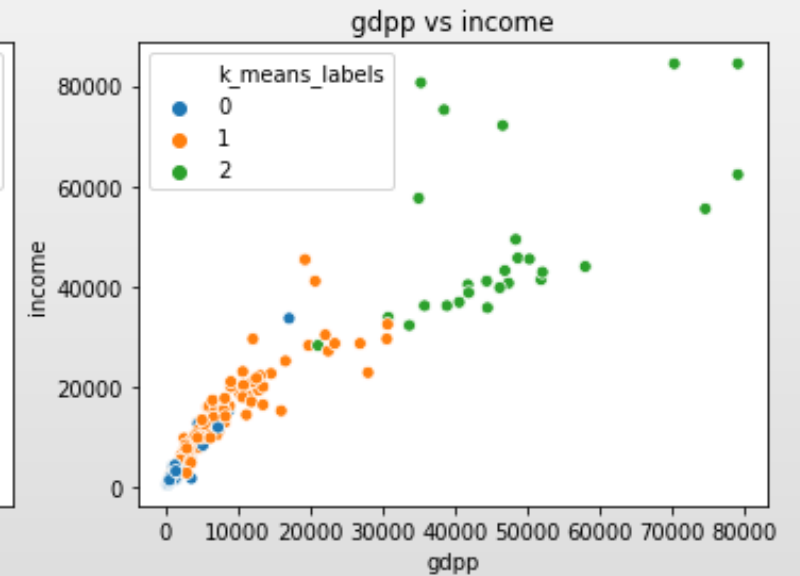
Cluster 0 is distributed in low income and low GDPP.

Cluster 1 is distributed in average income and average GDPP.

Cluster 2 is distributed in high income and high GDPP.

K-MEANS CLUSTERING (3/4)…
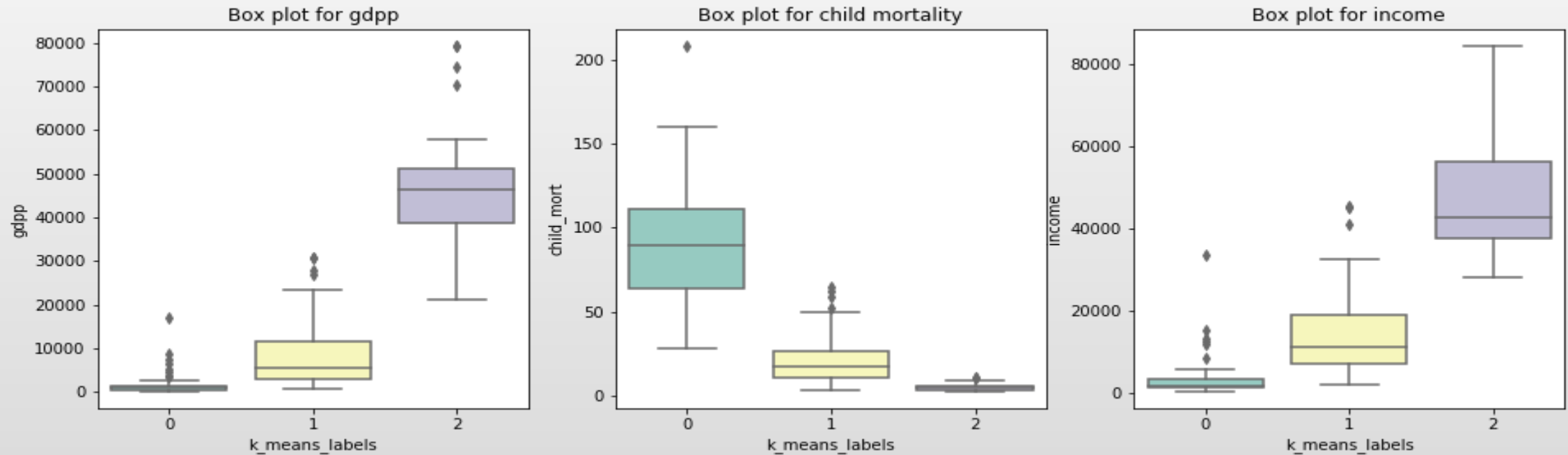
Cluster	GDPP	Income	Child Mortality
0	Low	Low	High
1	Low	Low	Low
2	High	High	Low

Cluster-0 has low median and less distribution for GDPP and income and higher median and spread for child morality.

Cluster-1 has median at lower side for GDPP , income and child morality, however the distribution and mean are higher than cluster-0 and lower than cluster-2 for income and GDPP.

Cluster-2 has high median and higher spread for GDPP and income, however the spread and median is low for child morality

K-MEANS CLUSTERING (4/4)

Standardized gdpp, income, child mortality for all clusters

Cluster 0 - Low GDPP, Low Income and High Child mortality
- *These are the countries who are in need of aid.*

Cluster 1 - Low GDPP, Low Income and Low Child Mortality

Cluster 2 - High GDPP, High Income and Low Child Mortality.
– *These are the developed Countries*

Sl No	Country	GDPP	Child mortality	Income
1	Burundi	231.0	93.6	764.0
2	Liberia	327.0	89.3	700.0
3	Congo, Dem. Rep.	334.0	116.0	609.0
4	Niger	348.0	123.0	814.0
5	Sierra Leone	399.0	160.0	1220.0
6	Madagascar	413.0	62.2	1390.0
7	Mozambique	419.0	101.0	918.0
8	Central African Republic	446.0	149.0	888.0
9	Malawi	459.0	90.5	1030.0
10	Eritrea	482.0	55.2	1420.0

HIERARCHICAL CLUSTERING (1/4)…

Single Linkage
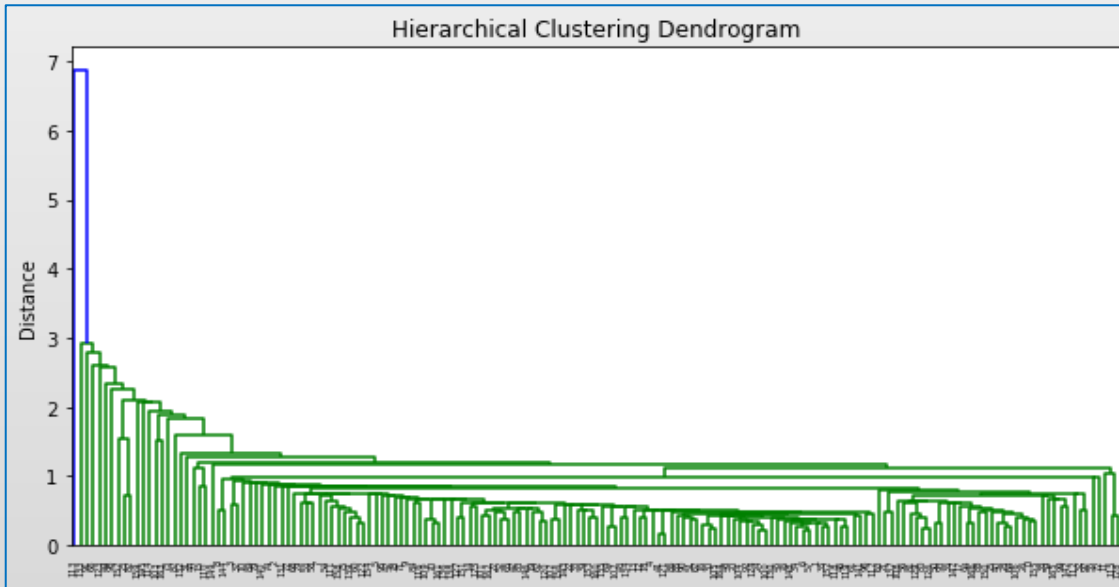
Complete Linkage
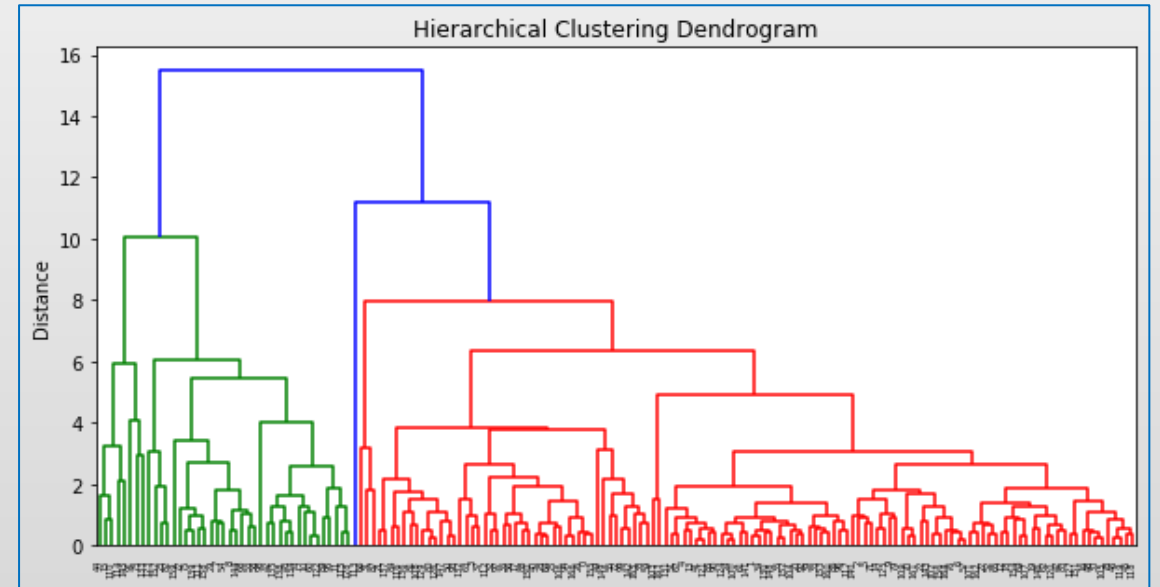
The dendrogram is not proper for single Linkage.

Dendrogram shows proper distinction for complete linkage.

Number of cluster (k) = 4 seems to be a good option. (cut tree distance is : 8-10)

HIERARCHICAL CLUSTERING (2/4)…
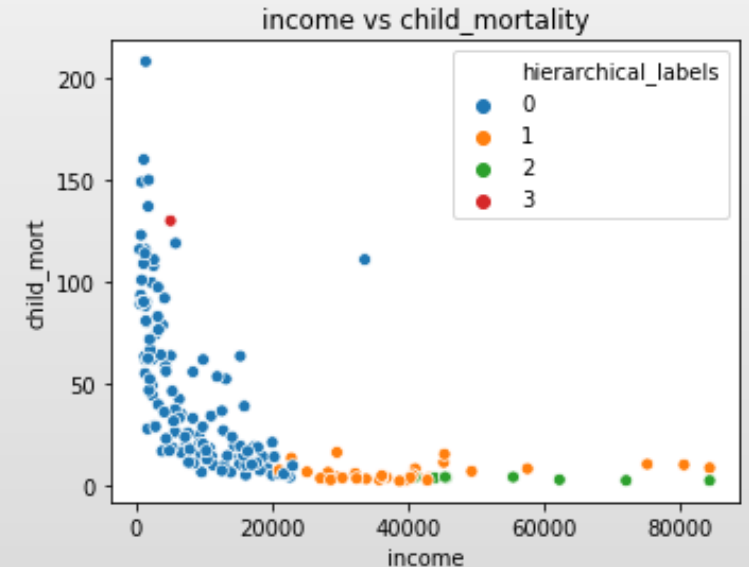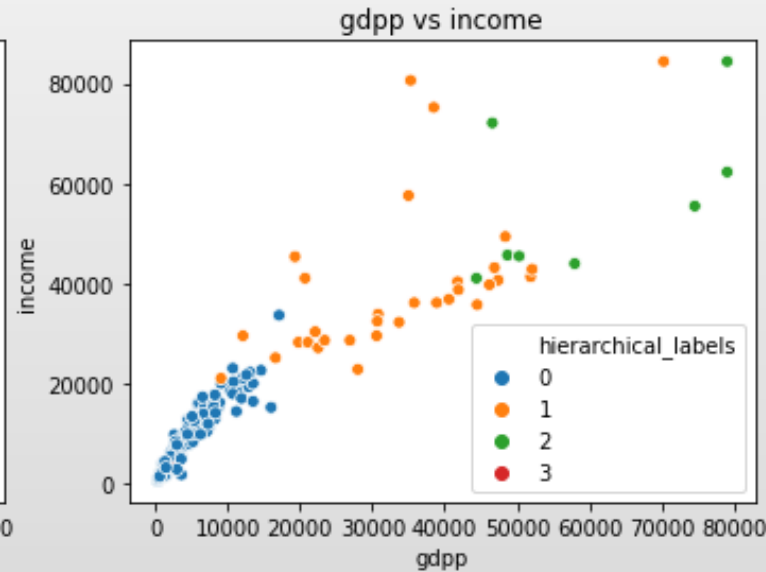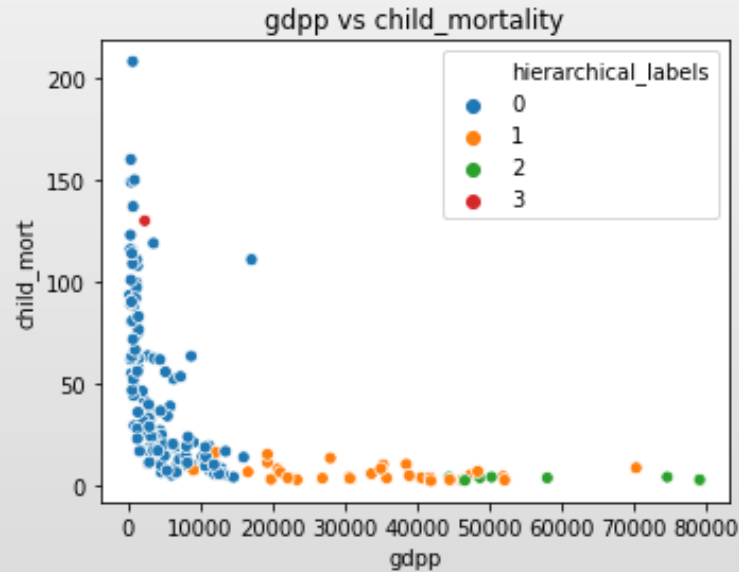
Cluster 0 has low GDPP, low income and high child mortality.

Cluster 1 has low GDPP, low income and low child mortality.

Cluster 2 has high GDPP, high income and low child mortality.

Cluster 3 has low GDPP, low income and high child mortality.

HIERARCHICAL CLUSTERING (3/4)…

Box plot for gdpp

Box plot for child mortality

Box plot for income
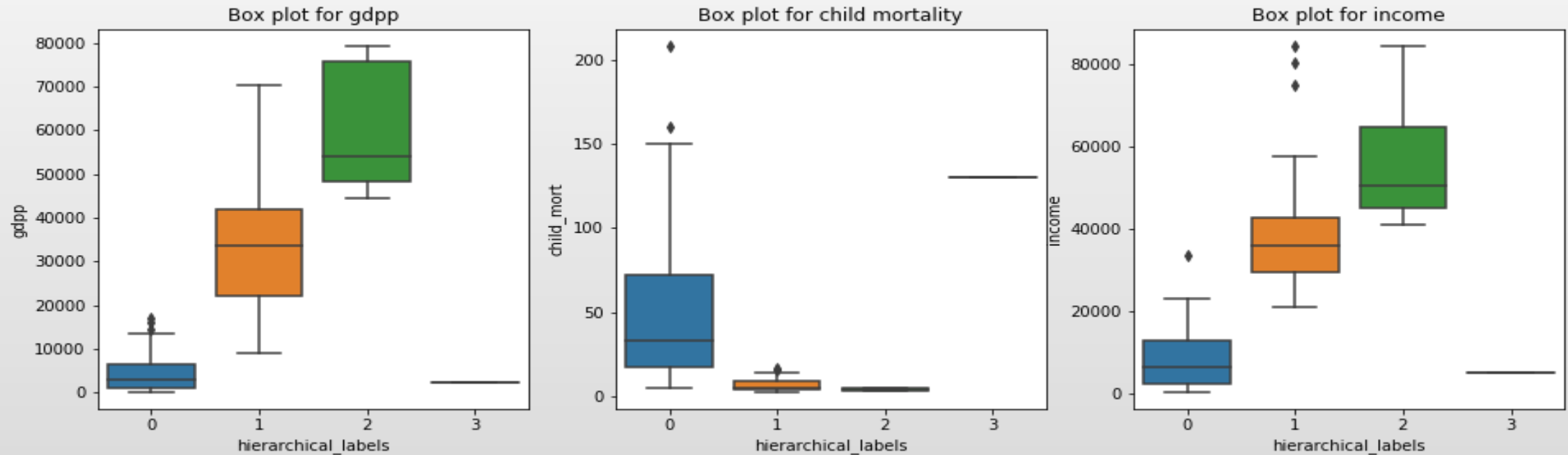
Cluster	GDPP	Income	Child Mortality
0	Low	Low	High
1	High	High	Low
2	High	High	Low
3	Low	Low	High

Cluster 0 has lower distributions for GDPP and income and higher distribution for child morality.

Cluster-1 has higher spread for income and GDPP and lower spread and median for child mortality.

Cluster-2 has higher spread and median for income and GDPP and lower spread and median for child mortality.

Cluster-3 has only 1 data point lower value for GDPP and income and higher value for child morality.

HIERARCHICAL CLUSTERING (4/4)

Standardized gdpp, income, child mortality for all clusters

Cluster 0 and Cluster 3 - Low GDPP, Low Income and High Child mortality - *These are the countries who are in need of aid.*
Cluster 1 - Low GDPP, Low Income and Low Child Mortality
Cluster 2 - High GDPP, High Income and Low Child Mortality.

COUNTRIES THAT REQUIRE AID (UNDER DEVELOPED COUNTRIES)

Sl No	Country	GDPP	Child mortality	Income
1	Burundi	231.0	93.6	764.0
2	Liberia	327.0	89.3	700.0
3	Congo, Dem. Rep.	334.0	116.0	609.0
4	Niger	348.0	123.0	814.0
5	Sierra Leone	399.0	160.0	1220.0
6	Madagascar	413.0	62.2	1390.0
7	Mozambique	419.0	101.0	918.0
8	Central African Republic	446.0	149.0	888.0
9	Malawi	459.0	90.5	1030.0
10	Eritrea	482.0	55.2	1420.0
11	Nigeria	2330.0	130.0	5150.0

COUNTRIES IN THE DIREST NEED OF AID

SI No	Country	GDPP	Child mortality	Income
1	Burundi	231.0	93.6	764.0
2	Liberia	327.0	89.3	700.0
3	Congo, Dem. Rep.	334.0	116.0	609.0
4	Niger	348.0	123.0	814.0
5	Sierra Leone	399.0	160.0	1220.0
6	Madagascar	413.0	62.2	1390.0
7	Mozambique	419.0	101.0	918.0
8	Central African Republic	446.0	149.0	888.0
9	Malawi	459.0	90.5	1030.0
10	Eritrea	482.0	55.2	1420.0

THANK YOU