Report on Countries in need for Aid

Prepared by –

Rohit Lal

Who are we?

- 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.

Problem Statement?

- We need to strategically and effectively utilise collected funds of $ 10 million, for helping the countries with direst need of aid.
- As a analyst, we need to analyse and get top 5 countries with direst need of funds.

Analysis Approach

- Data collection and cleaning
  - Import the data
  - Identify and eradicate the data quality issues.
- Outlier Analysis and removal
  - Handling the outliers effectively, so that it does not affect the output of clustering algorithms.
- EDA Analysis on data
  - Perform EDA to analyse the characteristics of data.
  - Identify patterns in data.

Analysis Approach

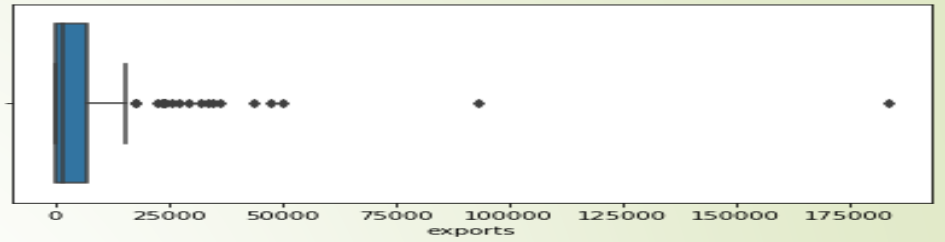
- Scaling the data
  - Scale the data to make it uniform so that it doesn't affects the clustering algorithm output.
- Hopkins statistics
  - To check if the data is suitable for clustering and it is not a complete random distribution.
- K-Means clustering
  - Identify the 'k' by silhouette analysis and elbow curve.
  - Divide the data into k-clusters.
  - Visualize the clusters by plotting.
  - Perform cluster profiling to get information about the clusters.

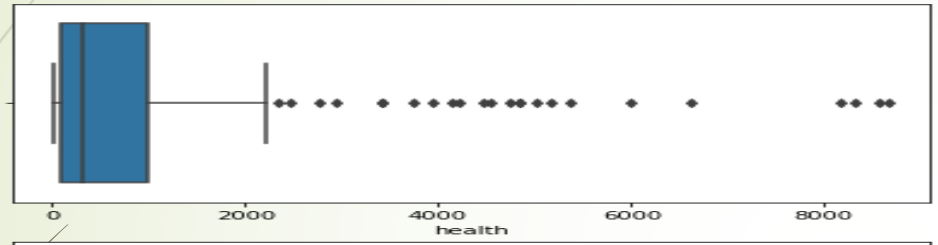
Analysis Approach

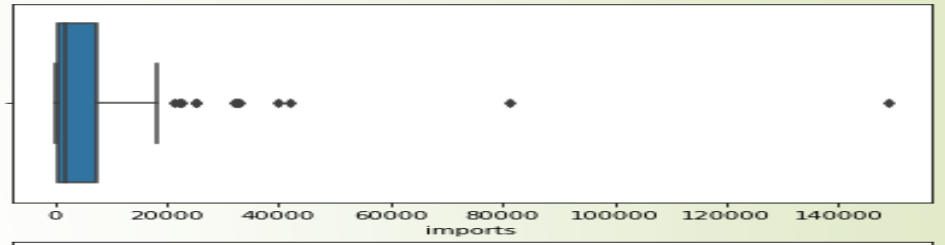
- Hierarchical clustering
  - Identify ‘n’ by dendrogram.
  - Divide the data into n-clusters.
  - Visualize the clusters by plotting.
  - Perform cluster profiling to get information about the clusters
- Final decision making
  - Identify the top 5 countries in dire need for funds by interpreting the results from both K-Mean & Hierarchical clustering.

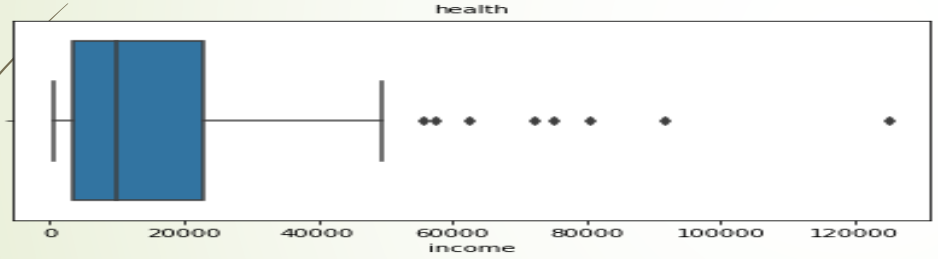
Detailed Approach Analysis

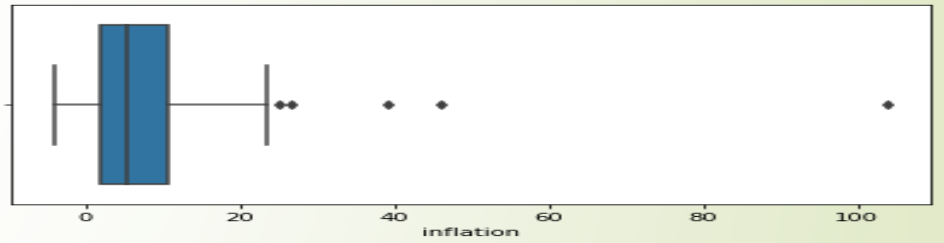
Outlier Analysis & Handling

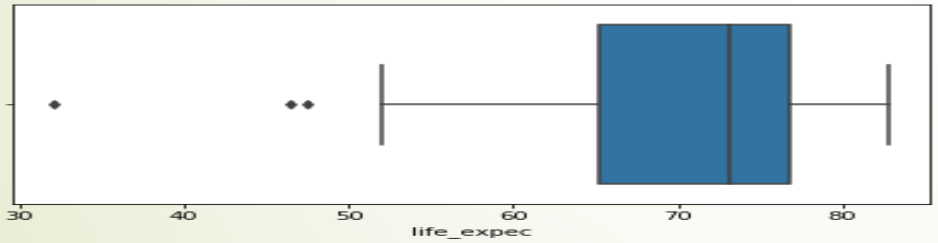


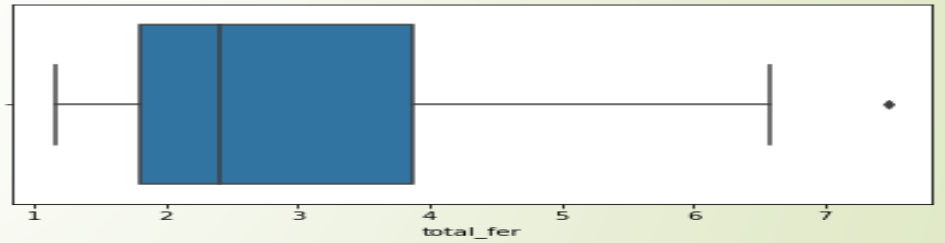


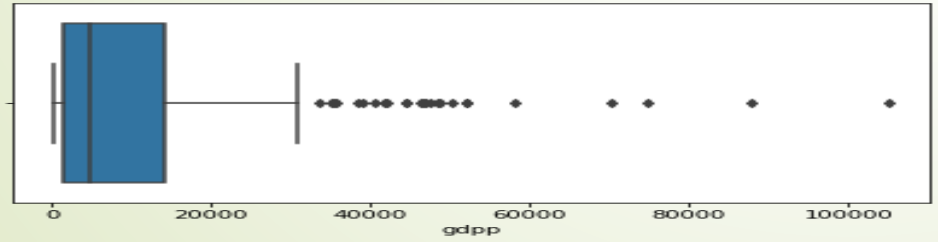




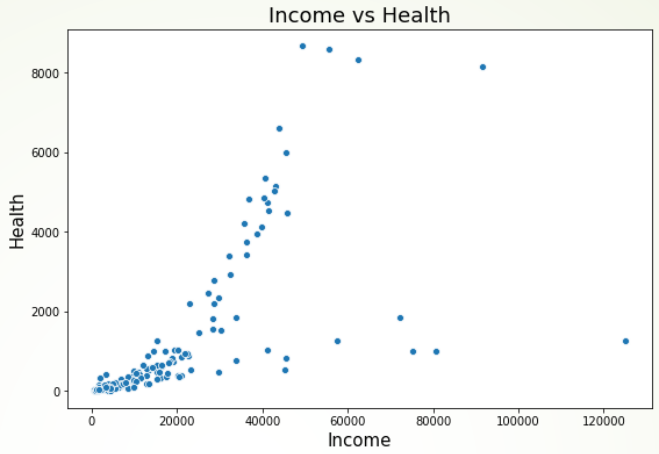


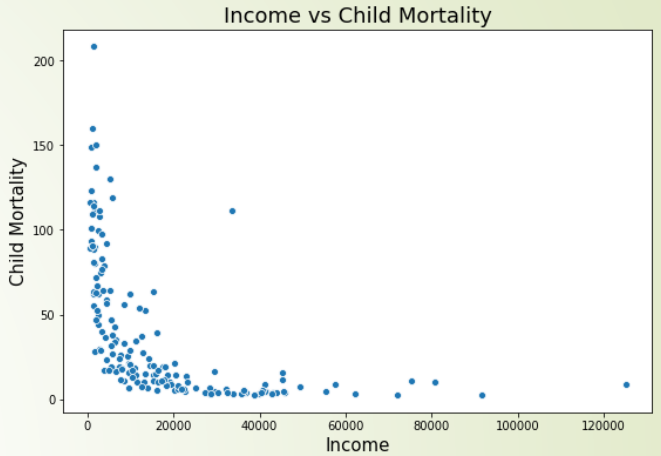


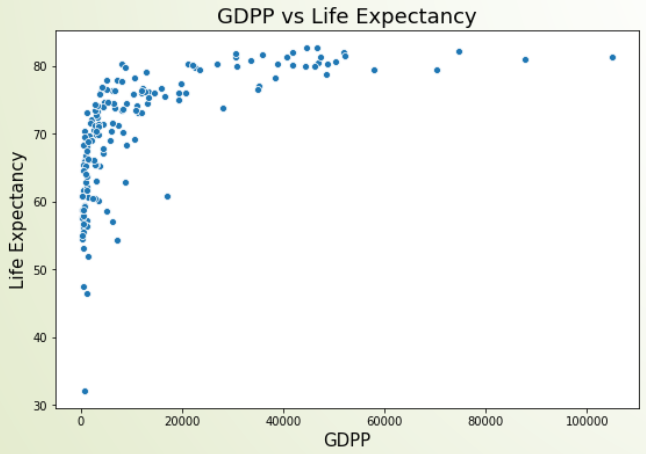


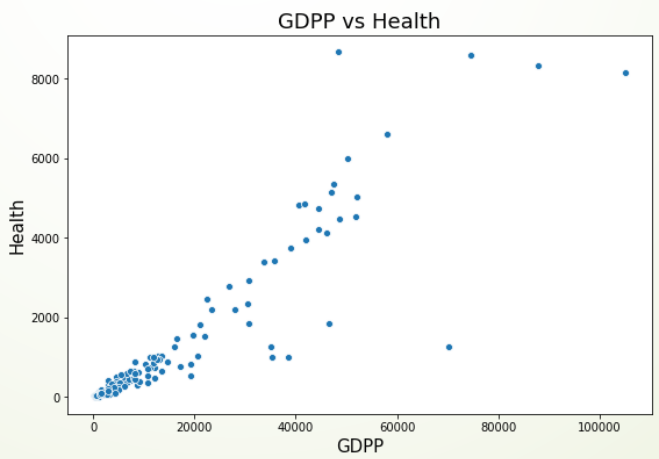


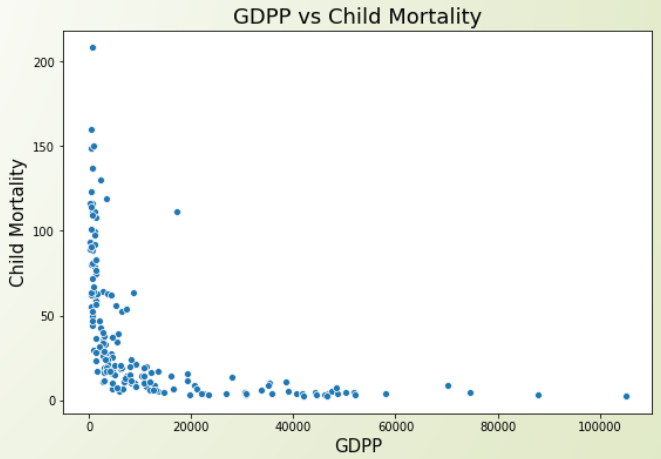
EDA Analysis



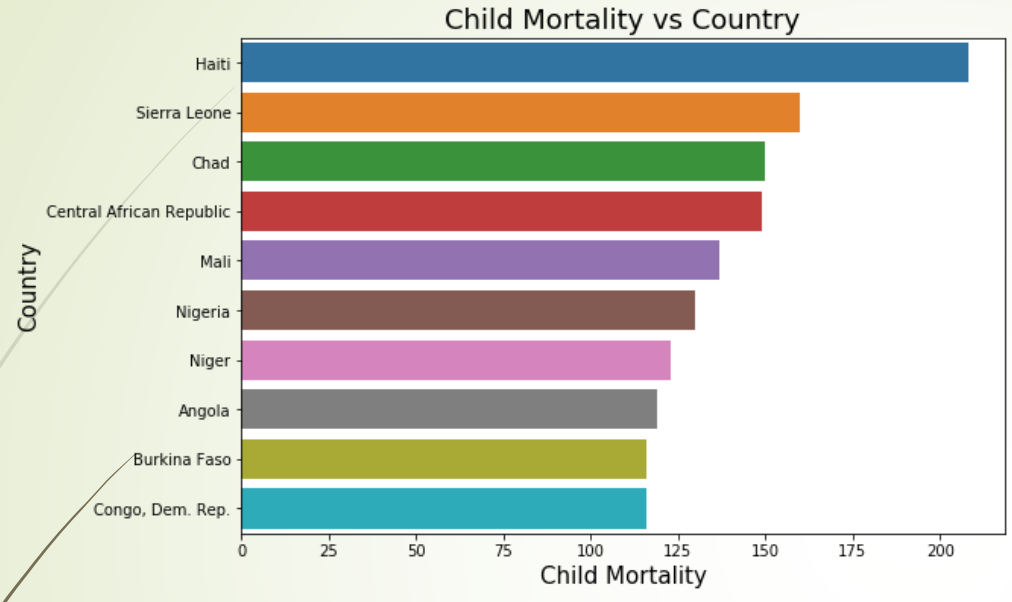


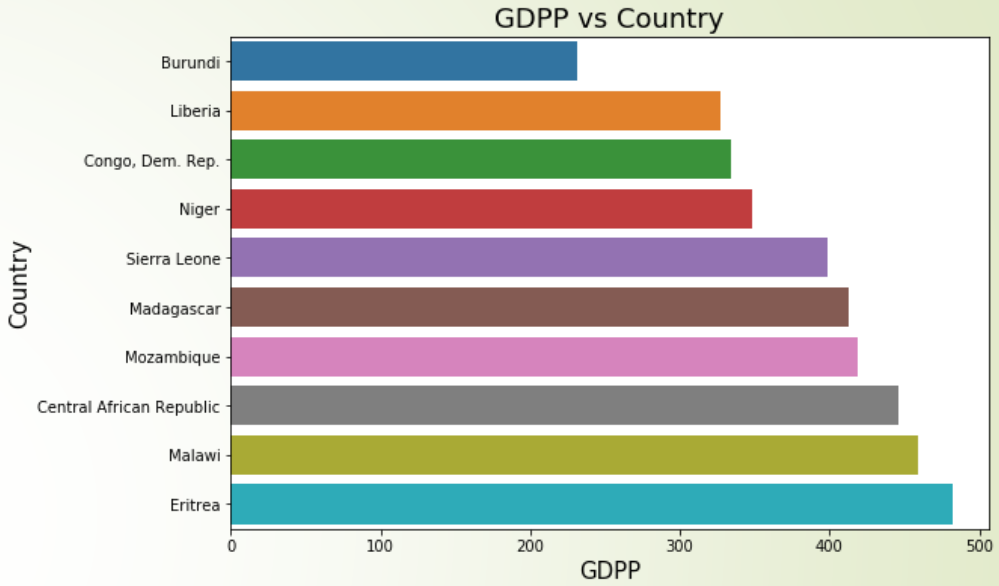


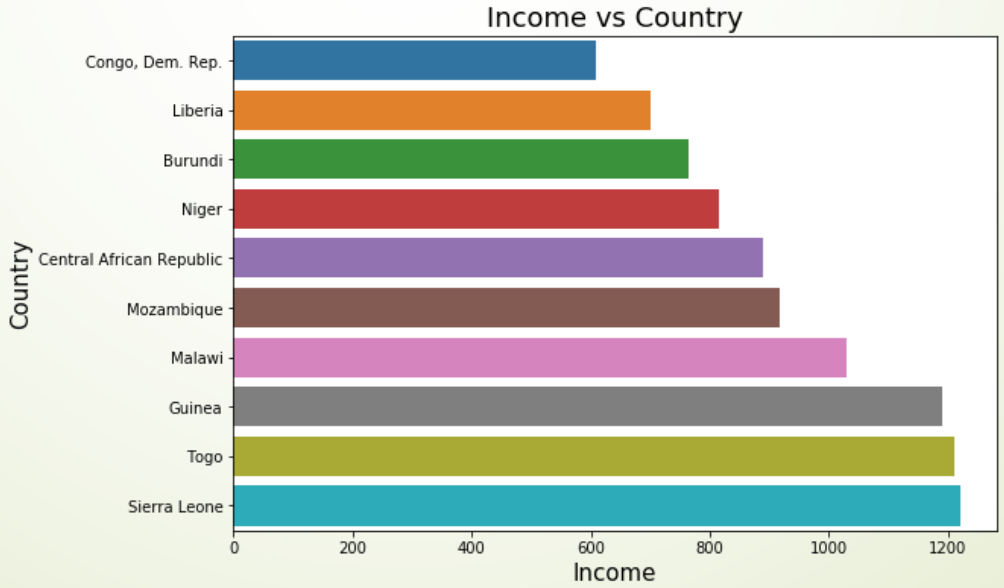




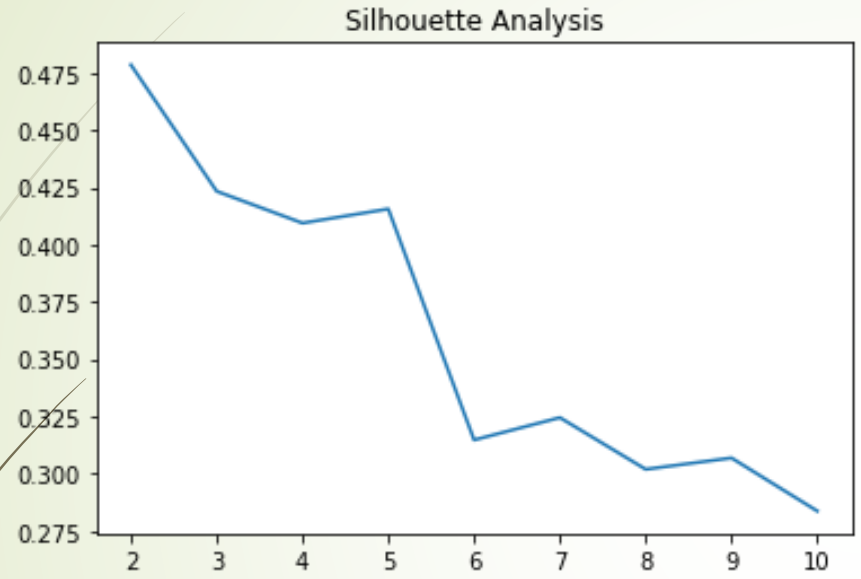
EDA Analysis

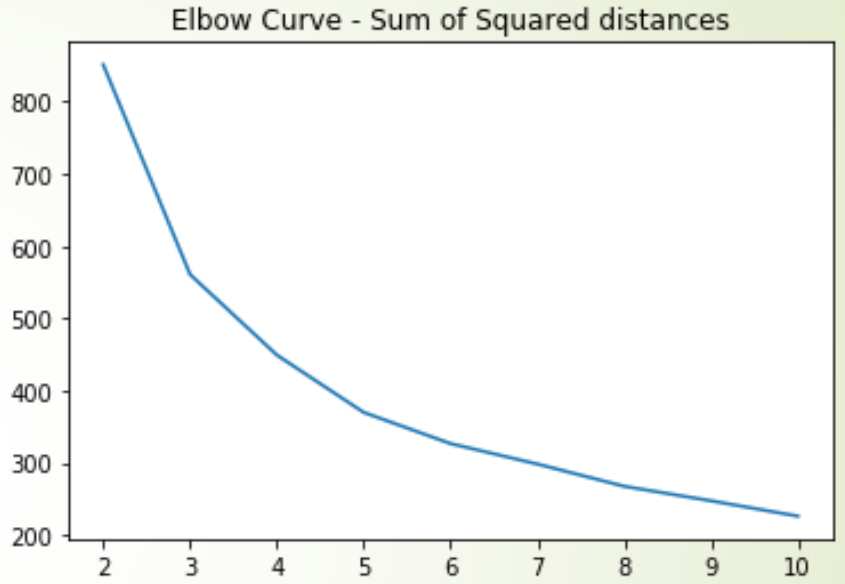






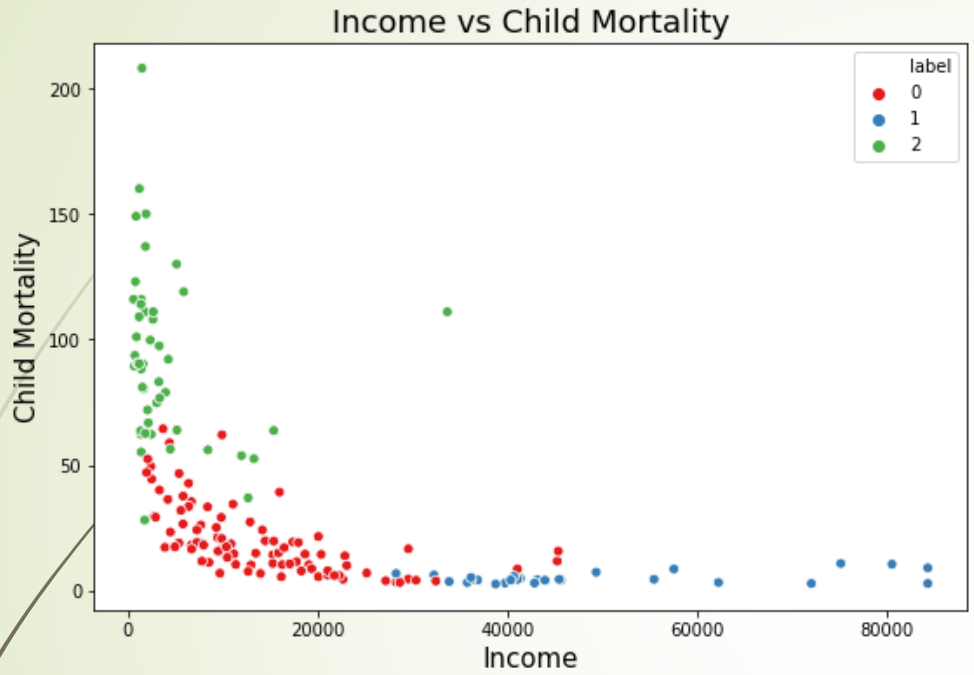
K-Means Clustering

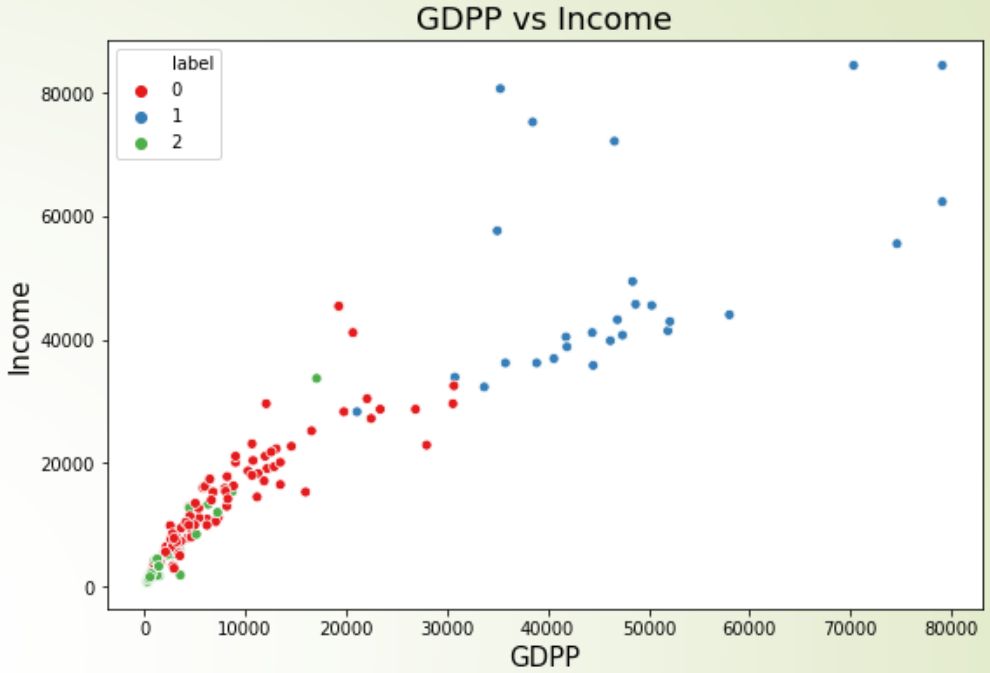




We chose K=3 from the silhouette analysis and elbow curve.

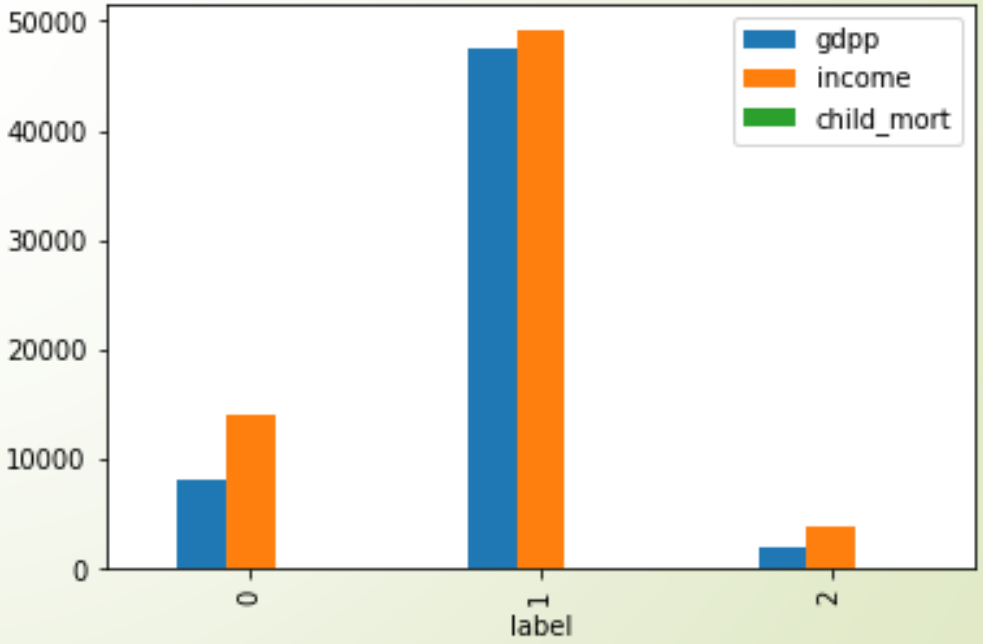
K-Means Clustering



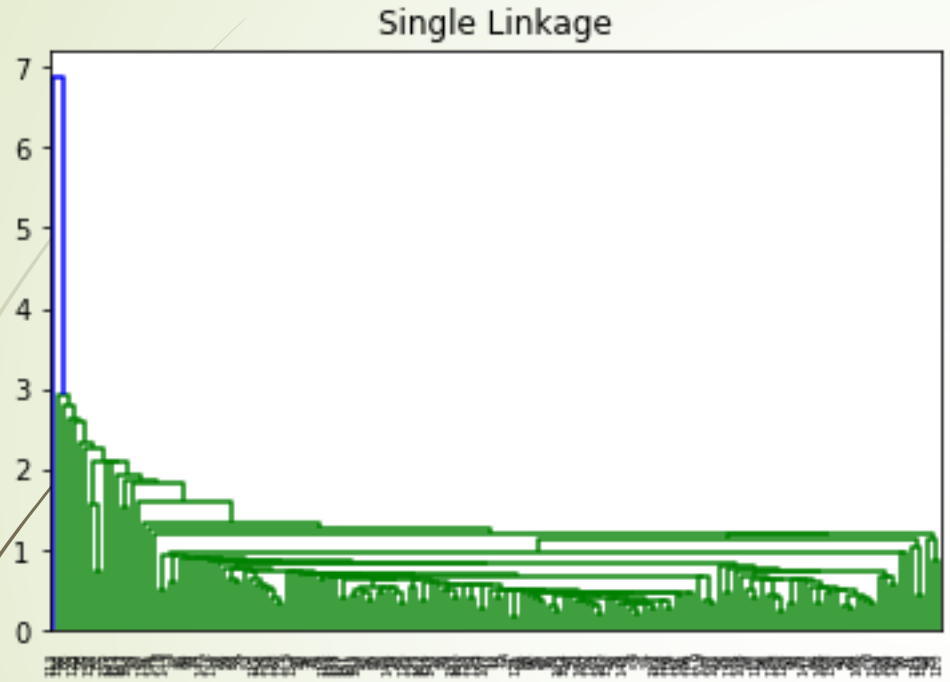


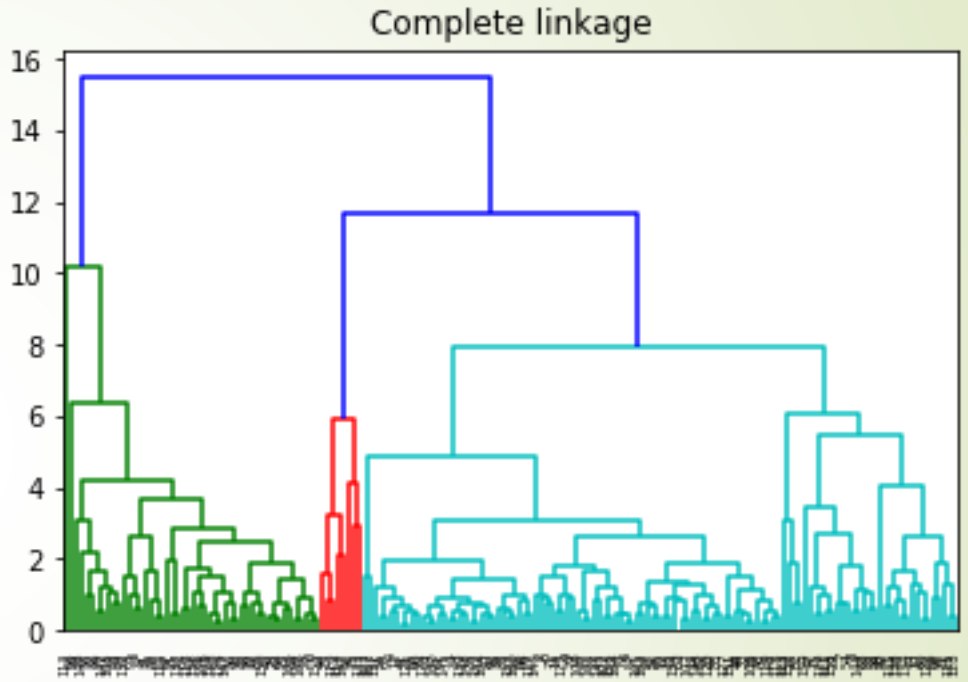
Cluster-2 comes out to be the cluster with lowest income, lowest GDP, and highest child mortality rate.

Top 5 countries for dire need of aid - **Burundi**, **Liberia**, **Congo, Dem. Rep.**, **Niger**, **Sierra Leone**



Hierarchical Clustering

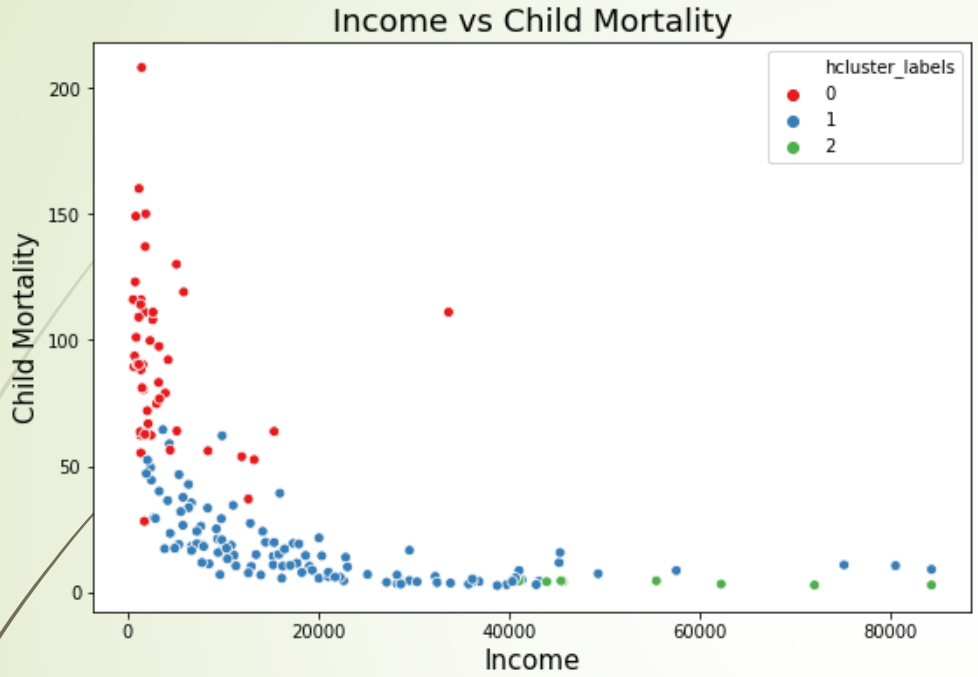


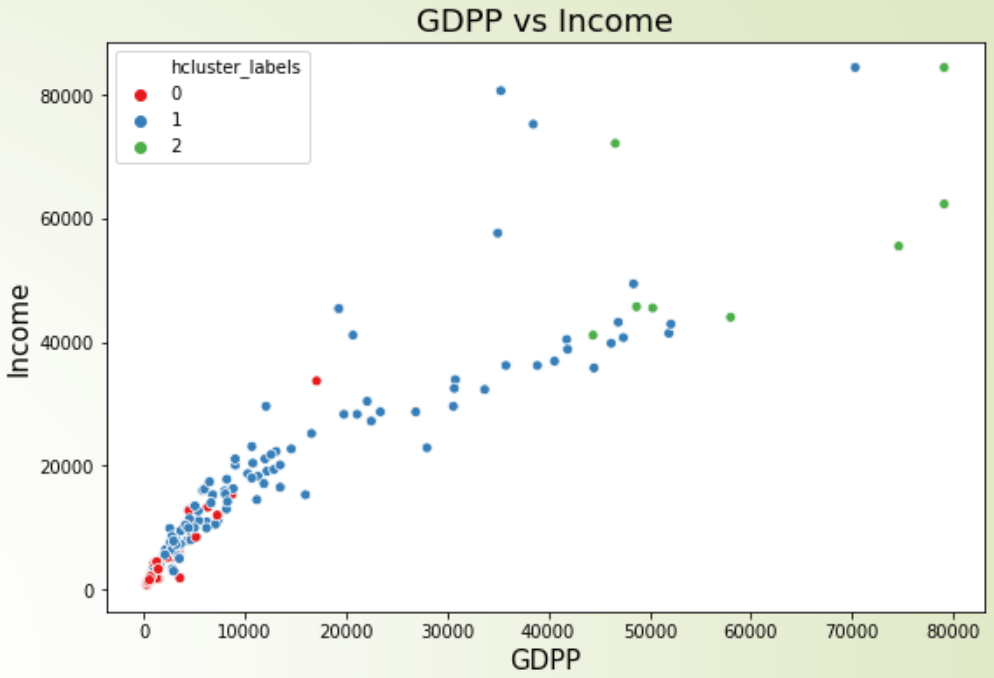


Complete linkage is much more interpretable in the hierarchical clustering.

We chose number of clusters = 3 for cutting the dendrogram in complete linkage.

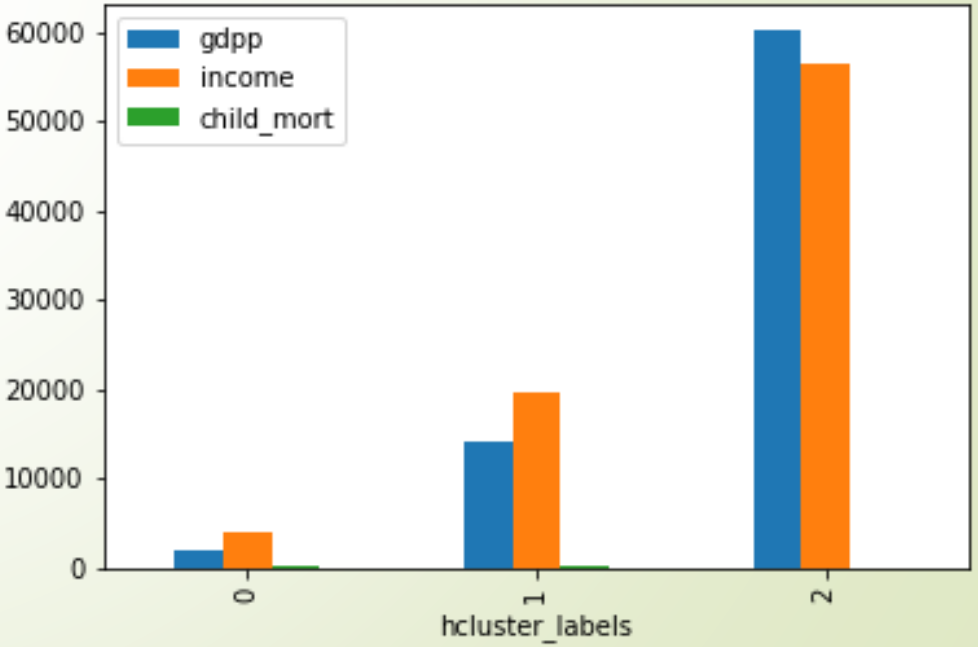
Hierarchical Clustering





Cluster-0 comes out to be the cluster with lowest income, lowest GDP & highest child mortality.

Top 5 countries for dire need of aid - **Burundi**, **Liberia**, **Congo, Dem. Rep.**, **Niger**, **Sierra Leone**



Summary

- By both K-Means & Hierarchical clustering, we got the same set of 5 countries which are in dire need of aid.

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