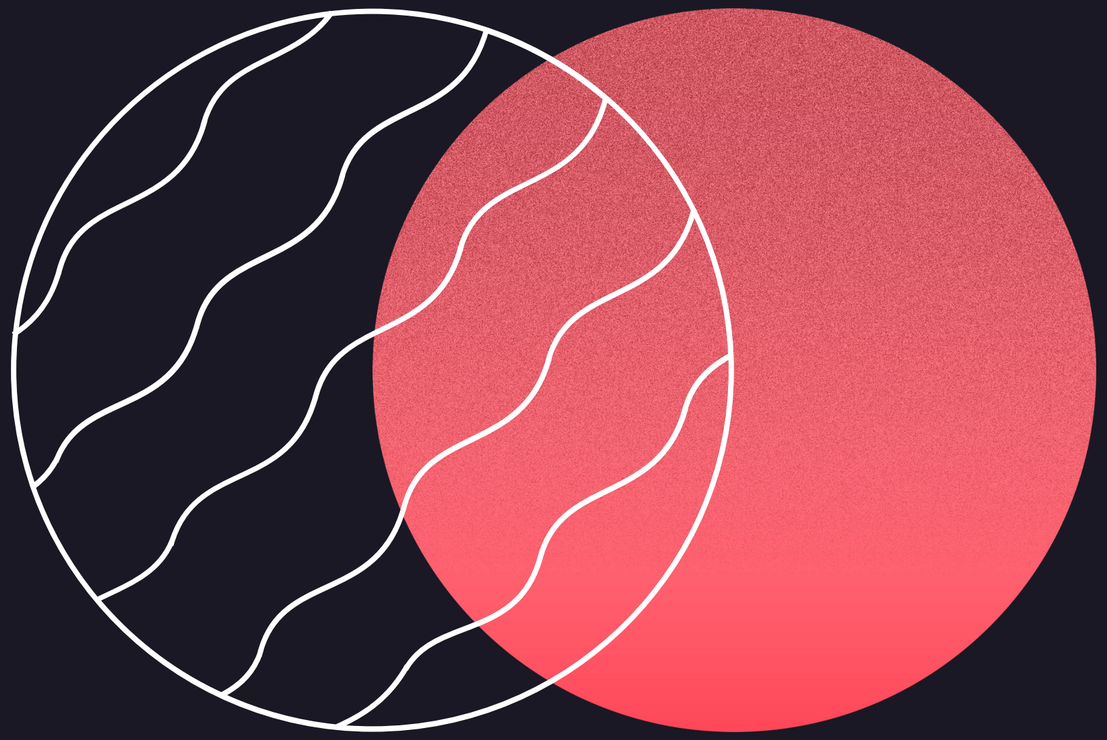
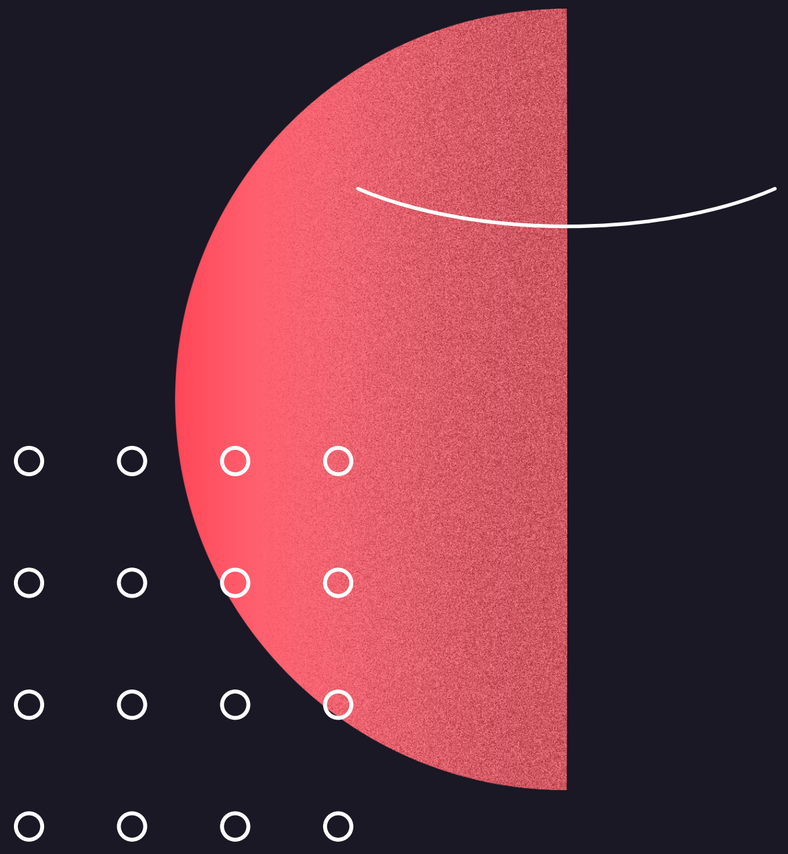


CLUSTERING ASSIGNMENT

BY :

AISHWARYA GYANJYOTI

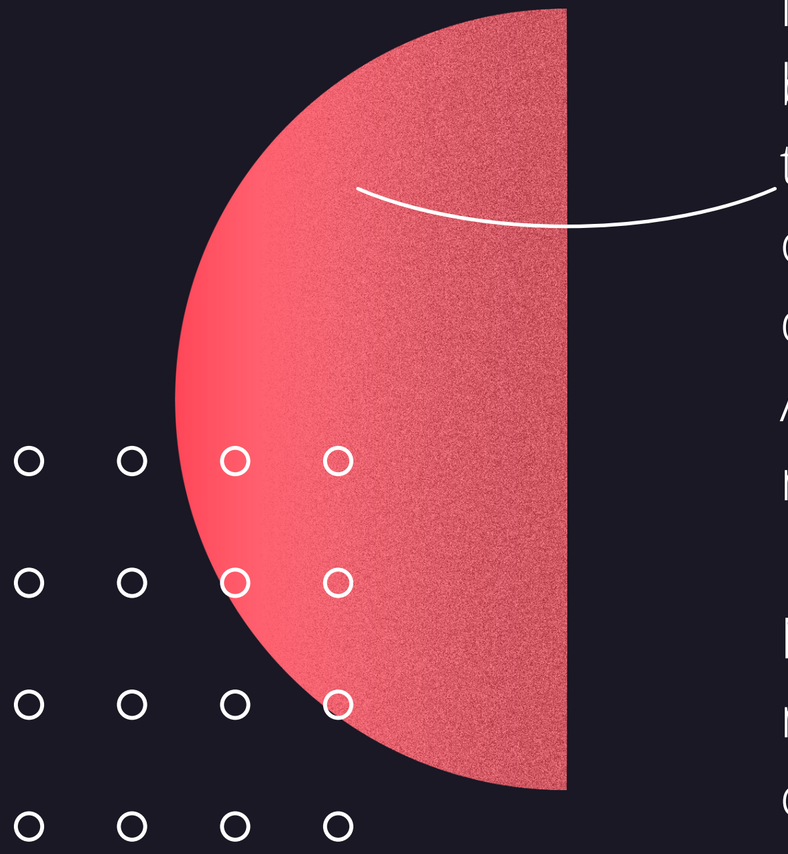
OBJECTIVE



· To mention the countries that are in direst need of aid on the basis of socio-economic and health factors determining the overall development of the country.

· Suggesting the countries to CEO at least 5 countries which are in direst need of aid from the analysis work that you perform.

PROBLEM STATEMENT

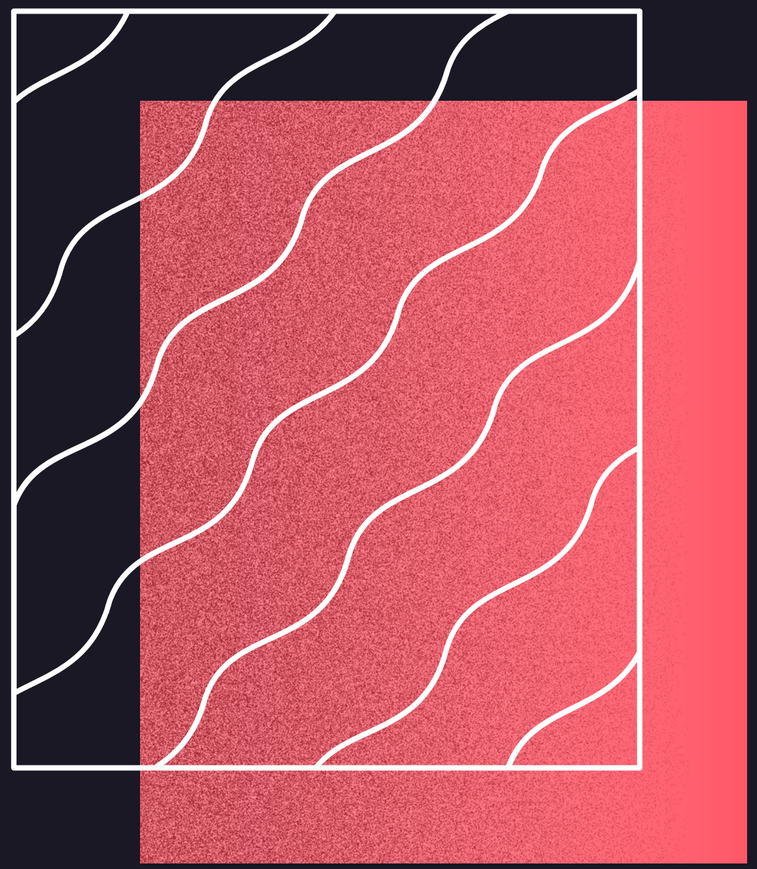


HELP International is an international humanitarian NGO that is committed to fighting poverty and providing the people of backward countries with basic amenities and relief during the time of disasters and natural calamities. It runs a lot of operational projects from time to time along with advocacy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

APPROACH

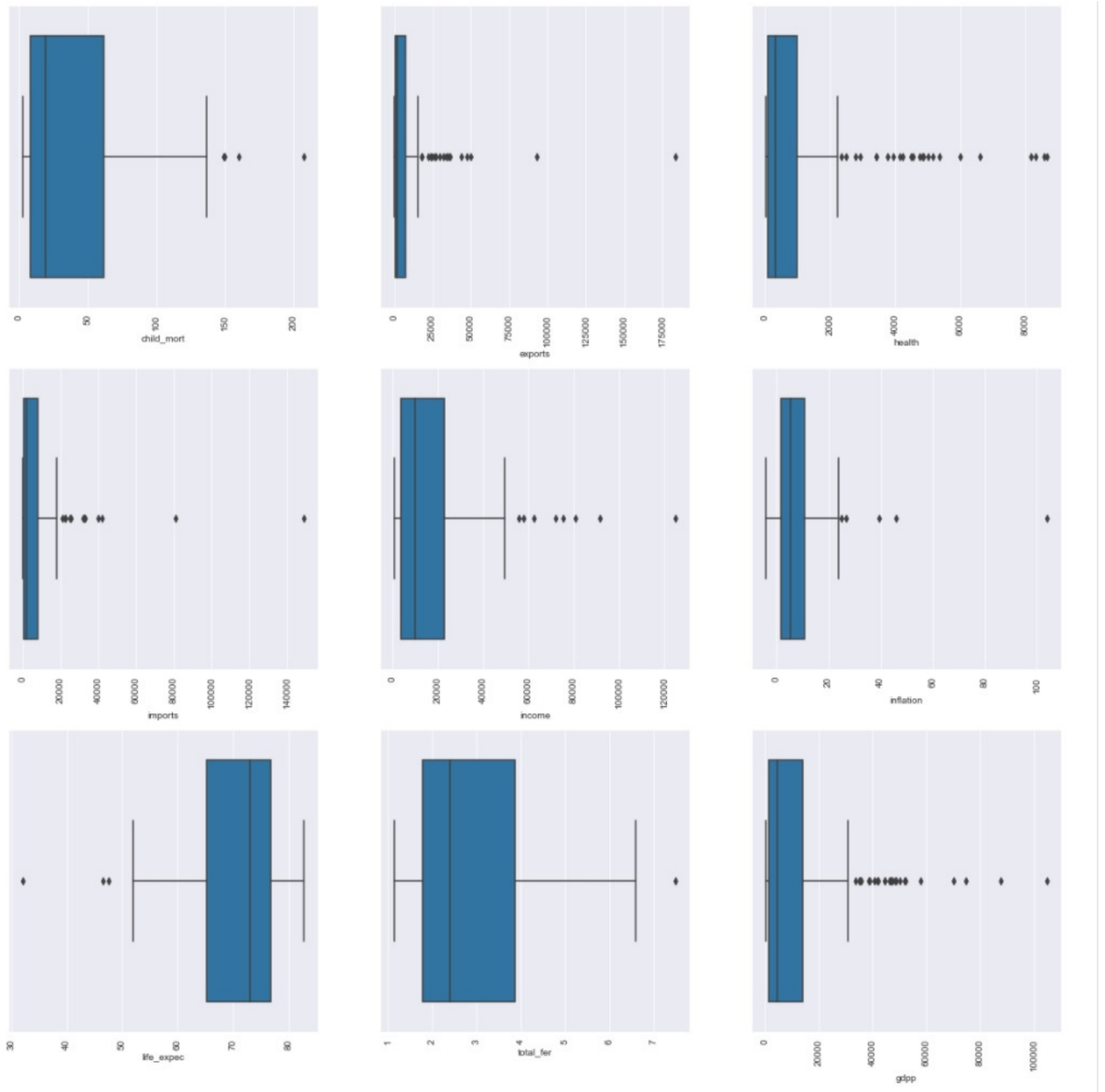
- Data Collection, Inspection and Cleaning
- EDA: Univariate, Bivariate Analysis & Heatmap
- Outlier Treatment
- Hopkins Statistics
- Data Scaling
- Finding best value of k using SSD & Silhoutte Score Analysis.
- Performing final KMeans Analysis using final value of k.
- Visualize the cluster using a scatter plot.
- Cluster profiling: GDPP, CHILD_MORT, and INCOME using boxplot.
- Hierarchical Clustering Single & Complete Linkage: Dendrogram. Complete Linkage is used for final clusters.
- Visualize using the scatter plot
- Cluster profiling: GDPP, CHILD_MORT, and INCOME using boxplot
- Deciding the top-5 countries that are in direst need of aid.



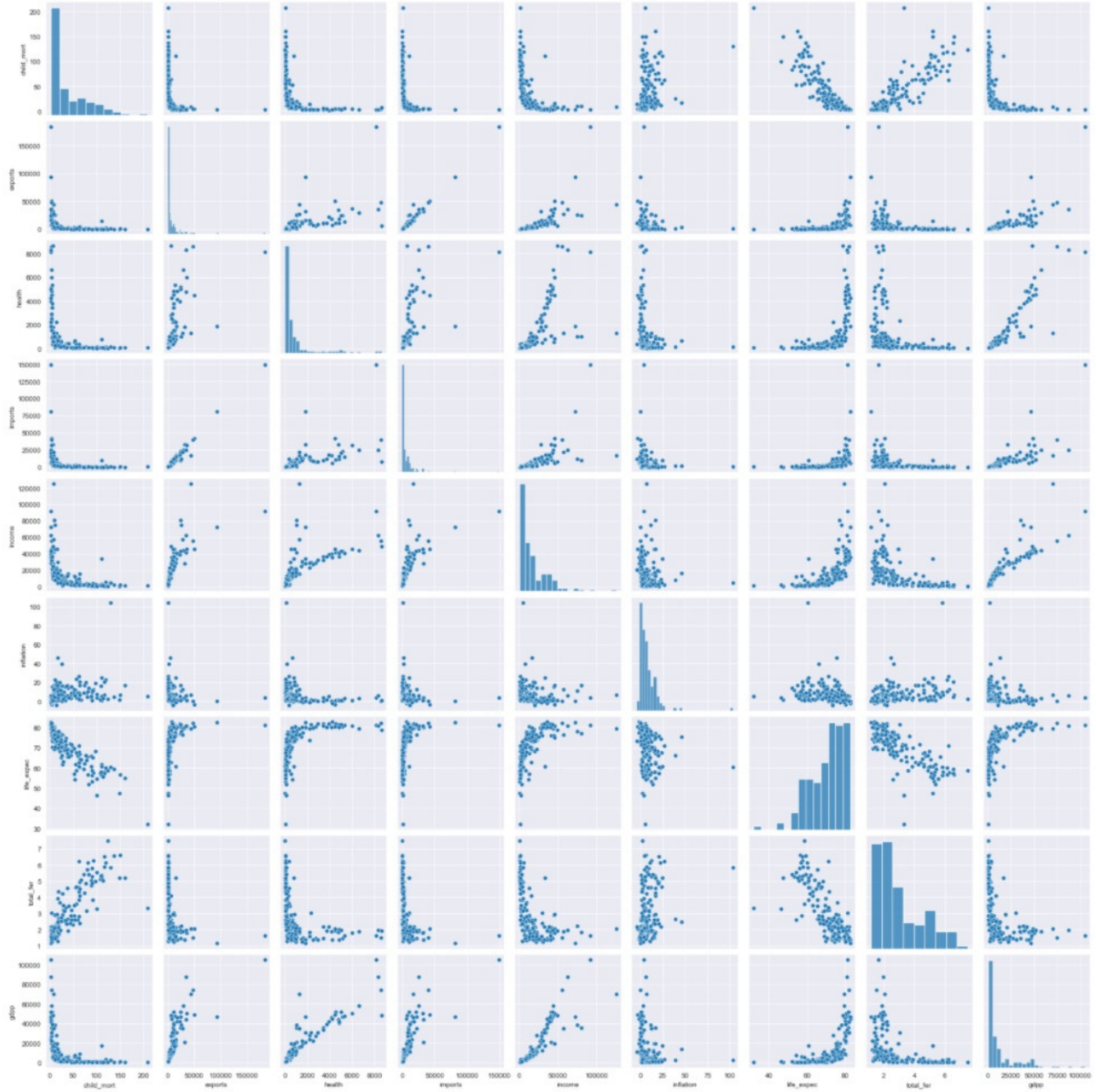
VISUALIZATION

UNIVARIATE ANALYSIS

As per the boxplot,
- Visibility of outliers in every columns.
- In comparison to other boxplots, columns like imports, exports and inflation has very thin size of quartiles.

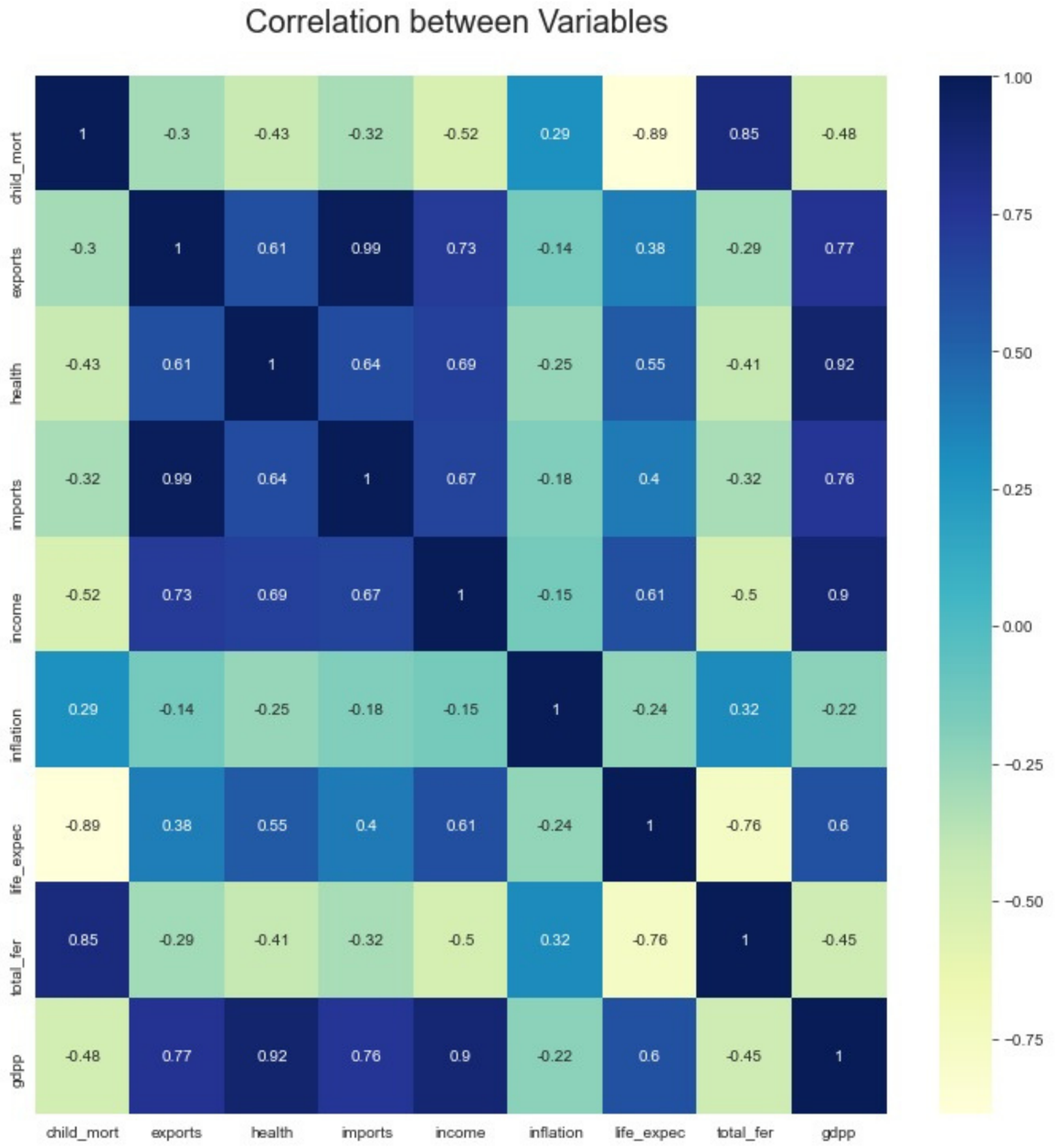


BIVARIATE ANALYSIS

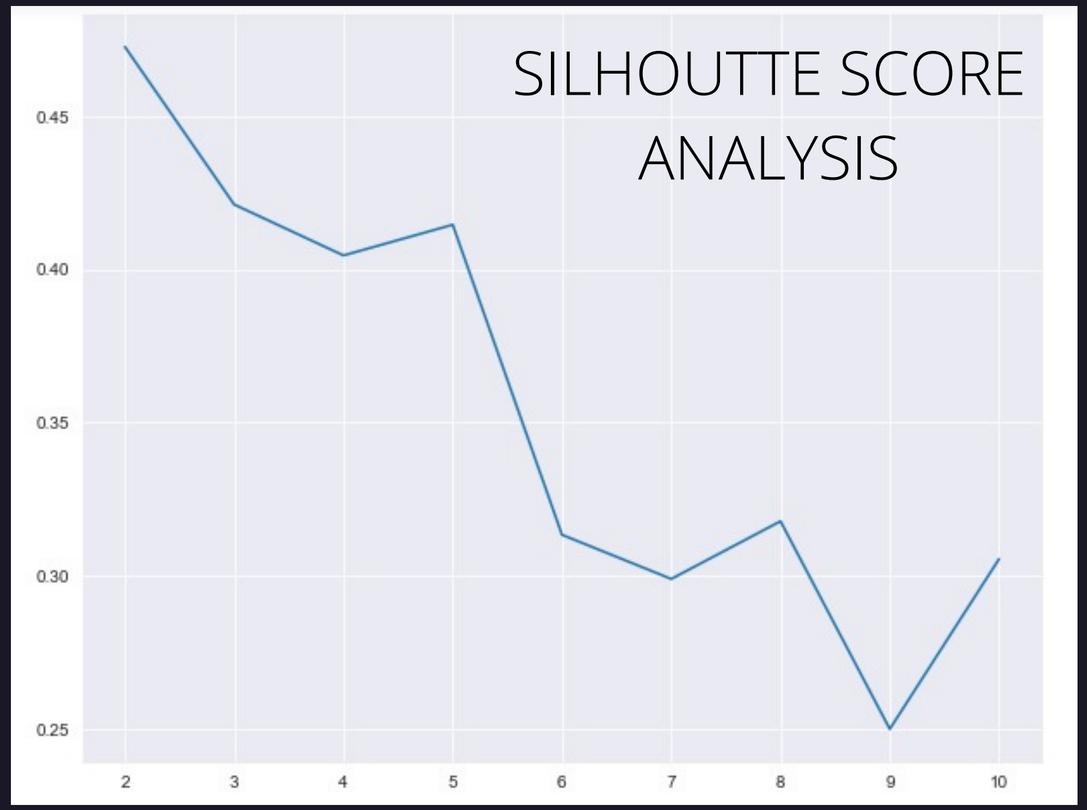


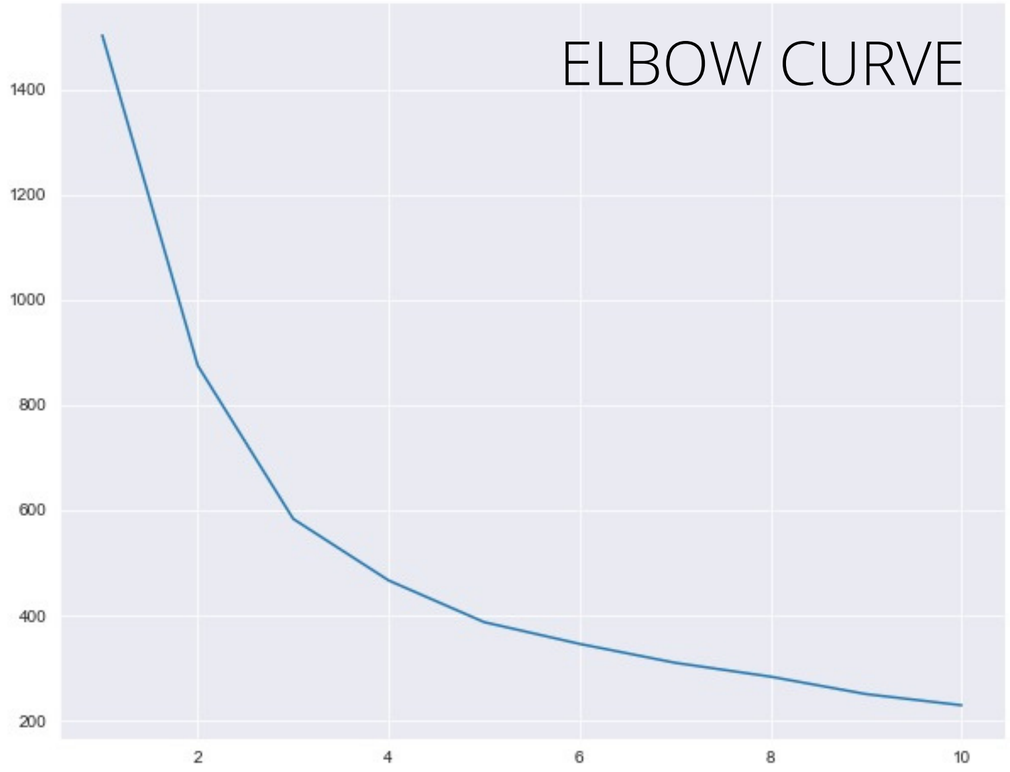
CORRELATION BETWEEN VARIABLES USING HEATMAP

As per the heatmap,
High correlation is observed in every variables (positive and negative values).



K- MEANS CLUSTERING USING ELBOW CURVE AND SILHOUTTE SCORE ANALYSIS





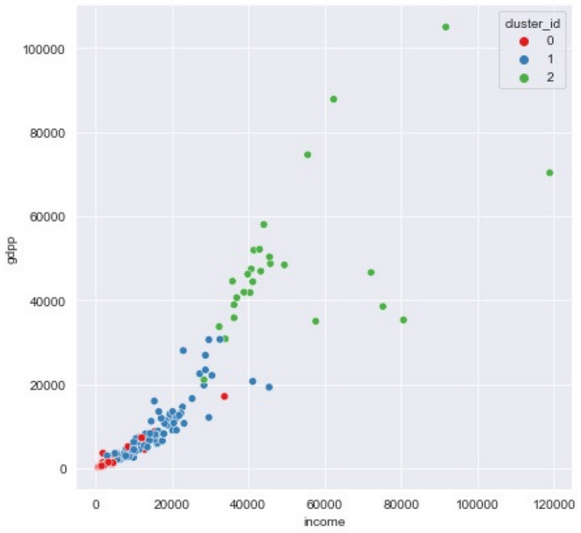
In elbow curve, highest peak is at 2 .
While in the plot of Silhoutte Score Analysis, no. of clusters aren't suitable for further clustering process and analysis.

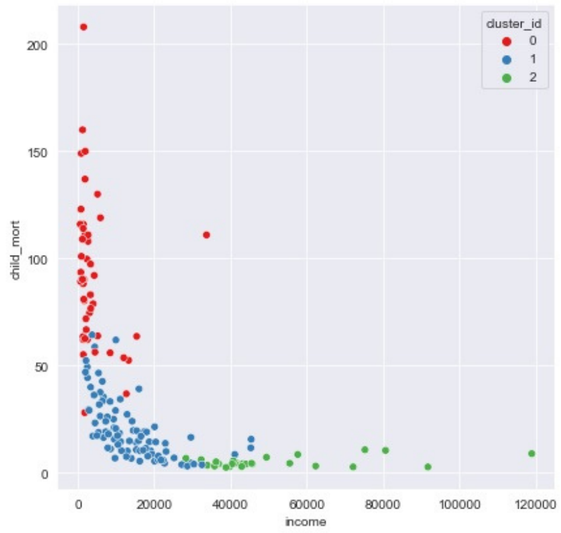
So,the optimal number of clusters is been taken as 3.

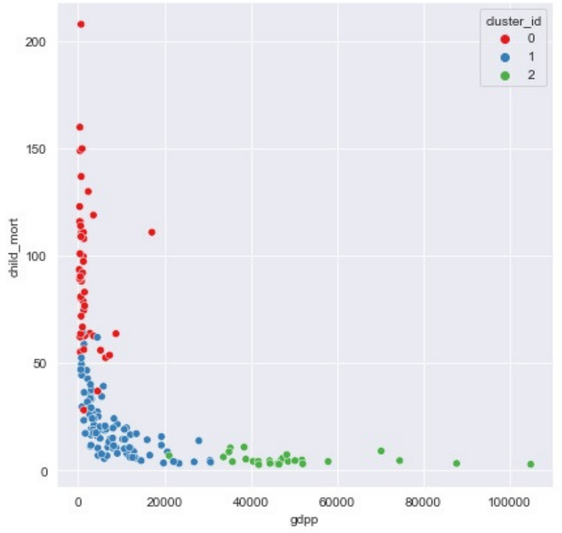
VISUALIZATION OF K MEANS CLUSTER USING SCATTER PLOT

In this scatter plot,

- gdpp and income is low in 0, while it's high in cluster_id =2.
- There's a rise in child mortality rate and income for cluste_id =0. 1 and 2 seems to be in decreasing form.
- Same scenario is with gdpp also.

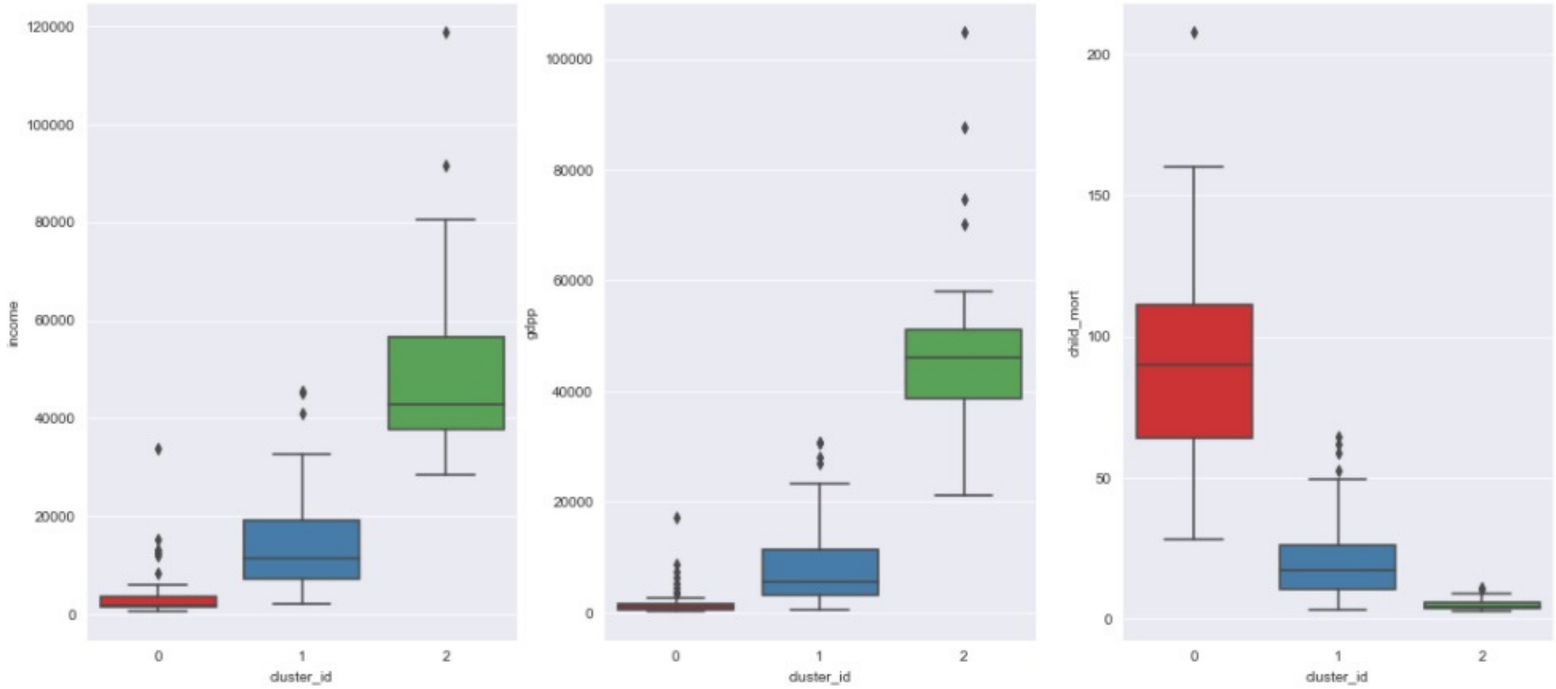






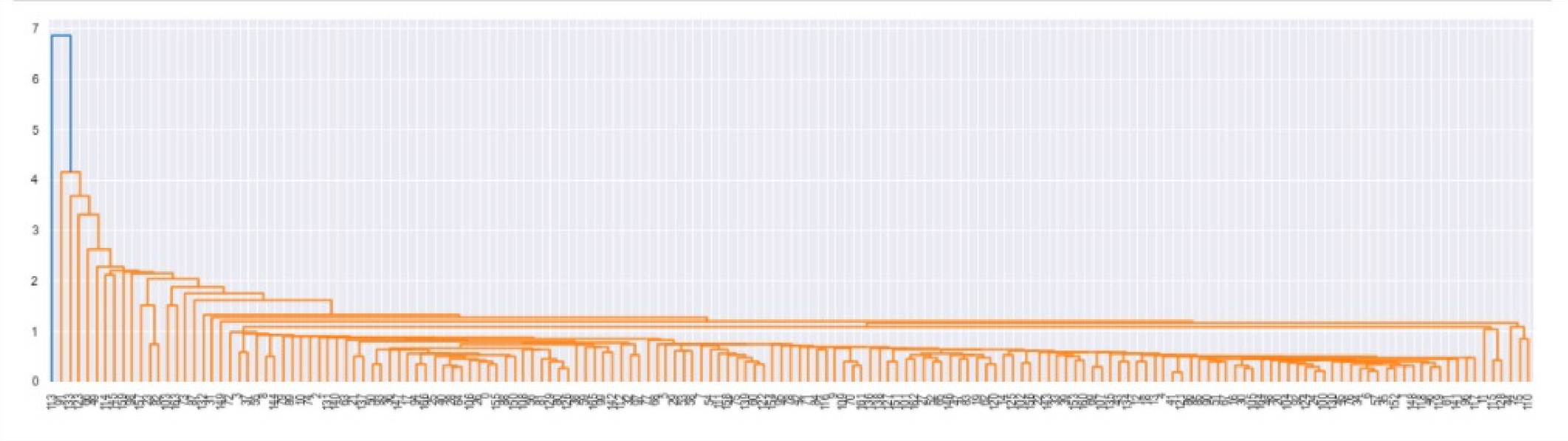
K-MEANS CLUSTER PROFILING

From the boxplot, cluster_id = 0 has low income,low gdpp and high child mortality. In cluster_id = 2, it seems to have high income and gdpp with low child mortality rate.



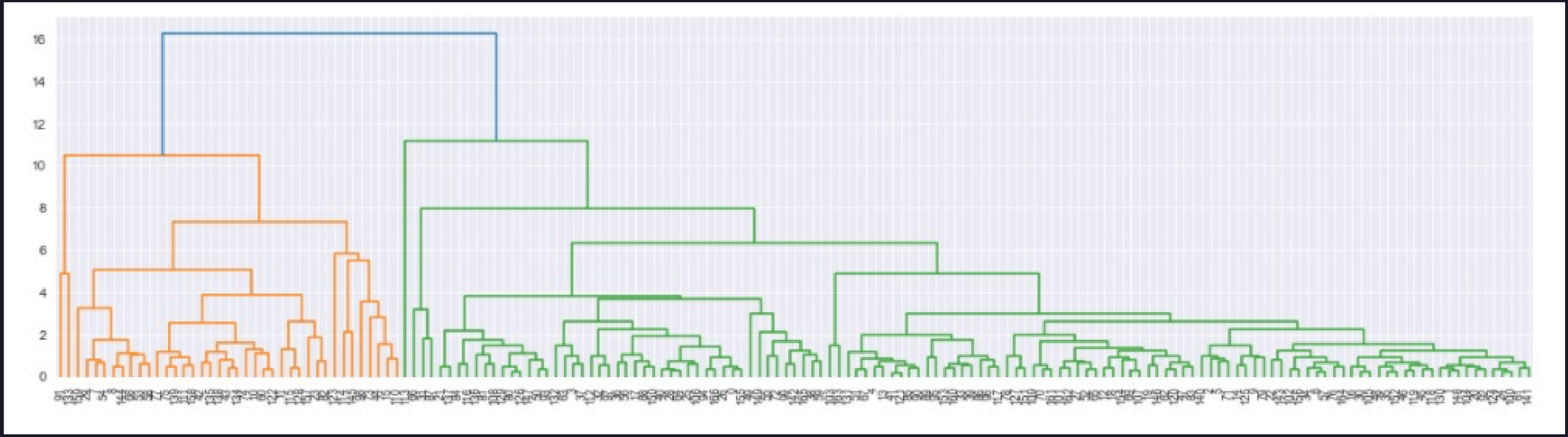
HIERARCHICAL CLUSTERING

SINGLE LINKAGE



HIERARCHICAL CLUSTERING

COMPLETE LINKAGE

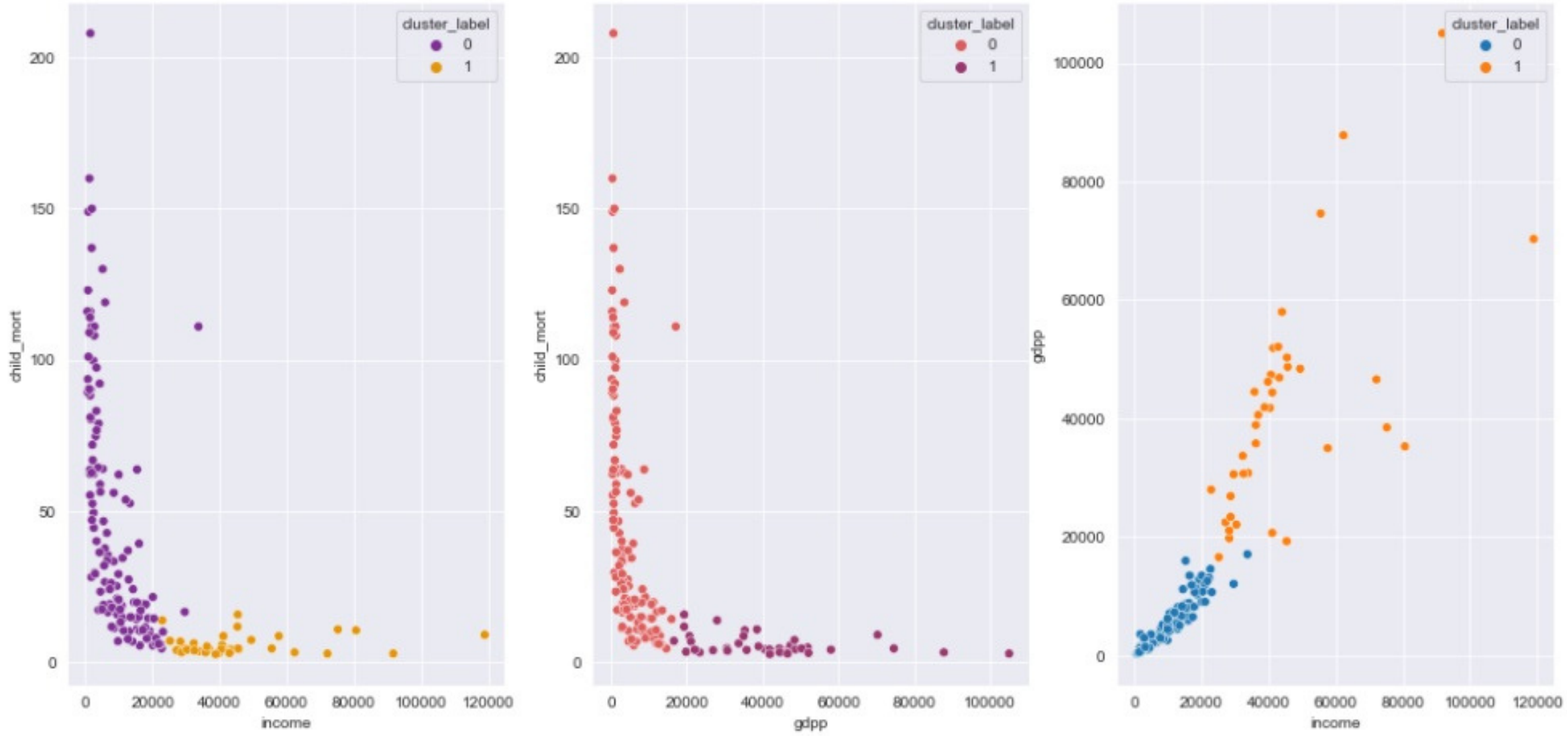


Complete Linkage model will be a good approach for knowing the no. of clusters needed after the horizontal line cuts the dendogram. About the single linkage, it's tough to know the whole hierarchy.

VISUALIZATION OF HIERARCHICAL CLUSTER USING SCATTER PLOT

In this scatter plot ,

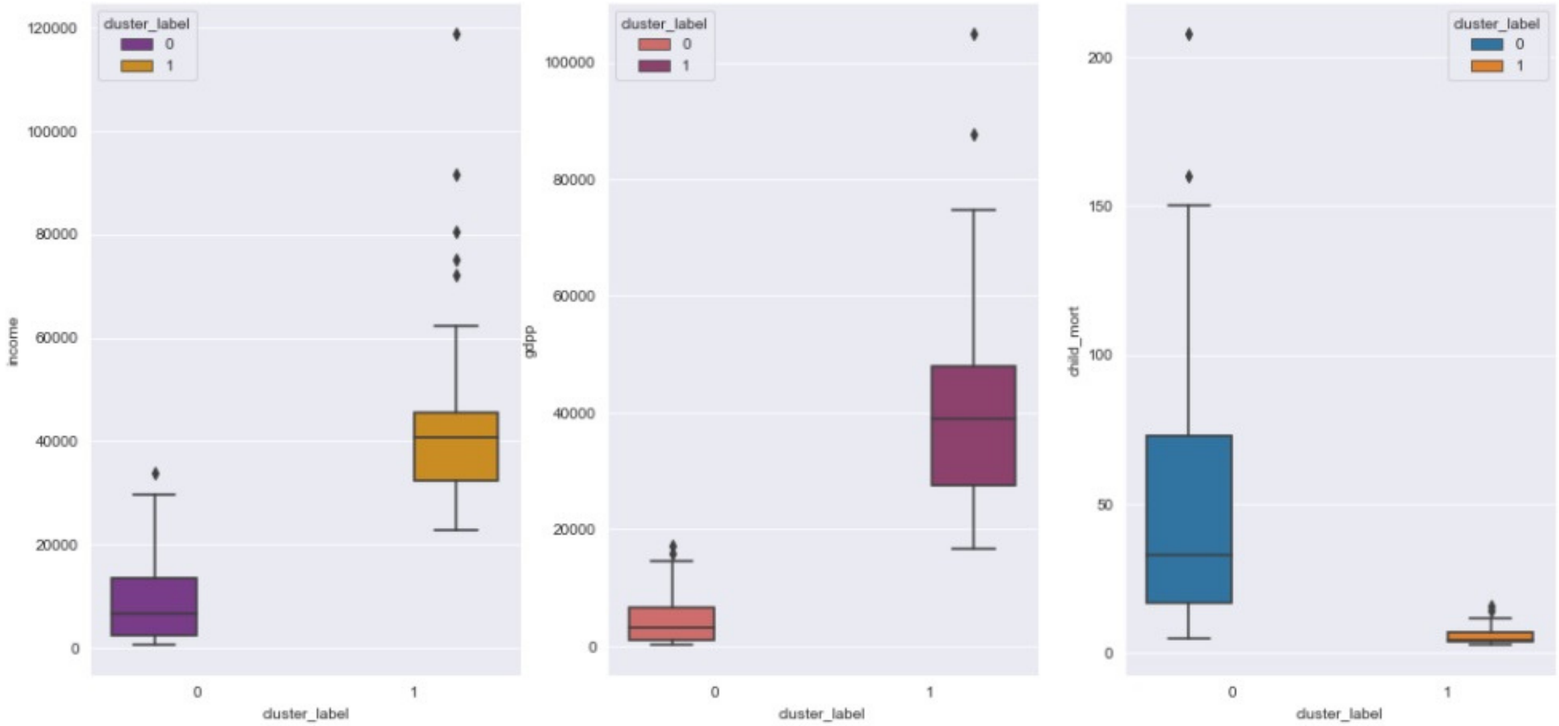
- There seems to be an increment in child mortality rate when its compared with income and gdpp which is low.
- Rise in gdpp and income for cluster_label as 1.



HIERARCHICAL CLUSTER PROFILING

As per the boxplot,

- cluster label= 0 has low income,low gdpp and high child_mort.
- In cluster label=1, gdpp and income seems to be high. Child mortality rate is low.



TOP 10 COUNTRIES THAT ARE IN DIREST NEED OF AID

	country	child_mort	gdpp	income
26	Burundi	93.6	231.0	764.0
88	Liberia	89.3	327.0	700.0
37	Congo, Dem. Rep.	116.0	334.0	609.0
112	Niger	123.0	348.0	814.0
132	Sierra Leone	160.0	399.0	1220.0
93	Madagascar	62.2	413.0	1390.0
106	Mozambique	101.0	419.0	918.0
31	Central African Republic	149.0	446.0	888.0
94	Malawi	90.5	459.0	1030.0
50	Eritrea	55.2	482.0	1420.0

CONCLUSION

The top-5 countries that are in direst need of aid are as follows :

- Burundi
- Liberia
- Congo, Dem. Rep.
- Niger
- Sierra Leone

THANK YOU