

# CLUSTERING ASSIGNMENT

Submitted by: Deepthi Yalavarthi

# **BUSINESS PROBLEM**

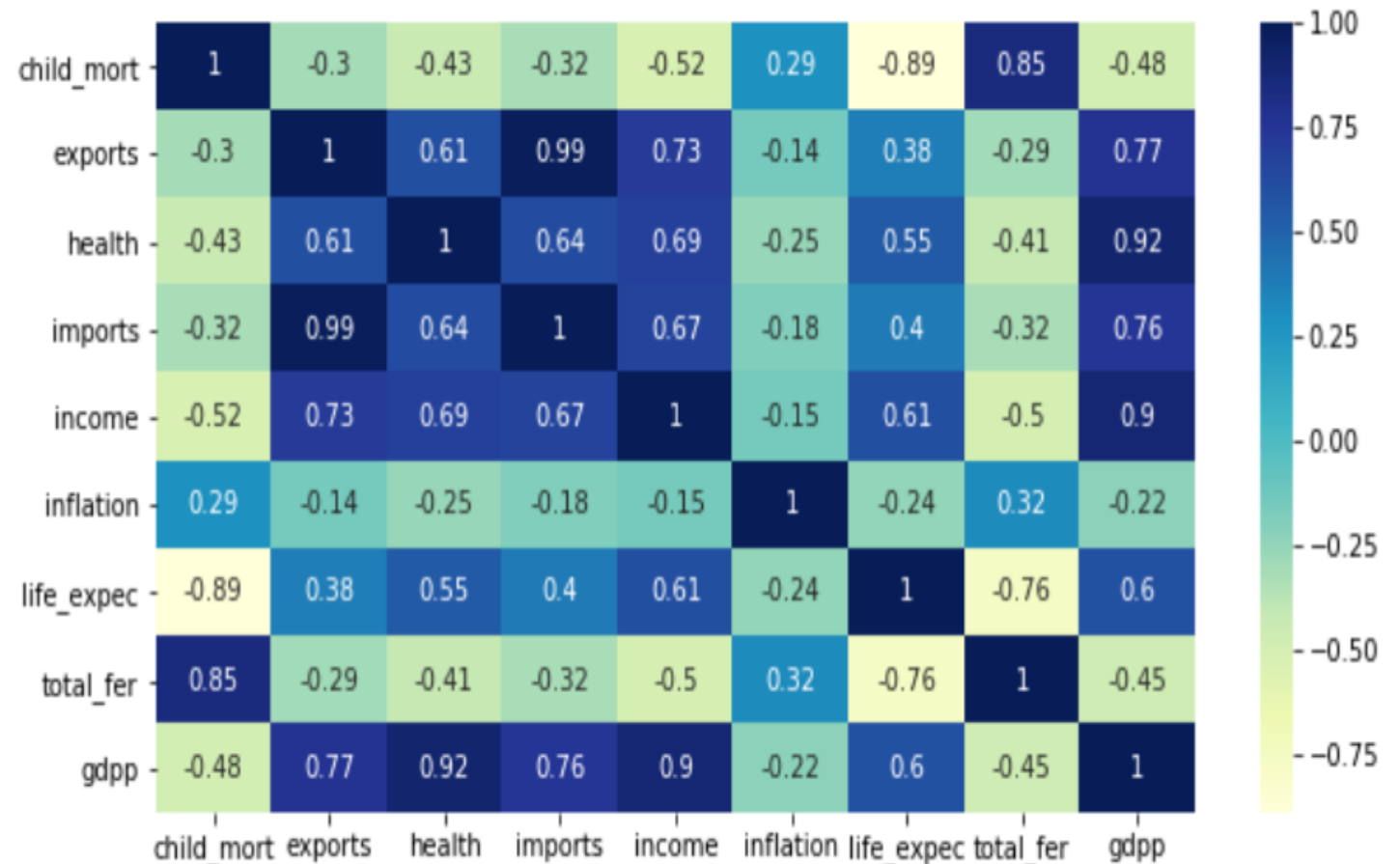
An international NGO organization CEO wants to use the money strategically and effectively on the categorized list of countries, who are in the direct need of aid, using some socio-economic and health factors that determine the overall development of the country.

# STEPS FOLLOWED

- Data Inspection, Analysis and Cleaning
- EDA Analysis and Visualization
- Outlier Analysis
- Hopkins Statistics
- SSD Elbow Curve
- Silhouette Score
- K-Means clustering
- Final list of countries
- Hierarchical clustering : Single Linkage, Complete Linkage

# EDA ANALYSIS

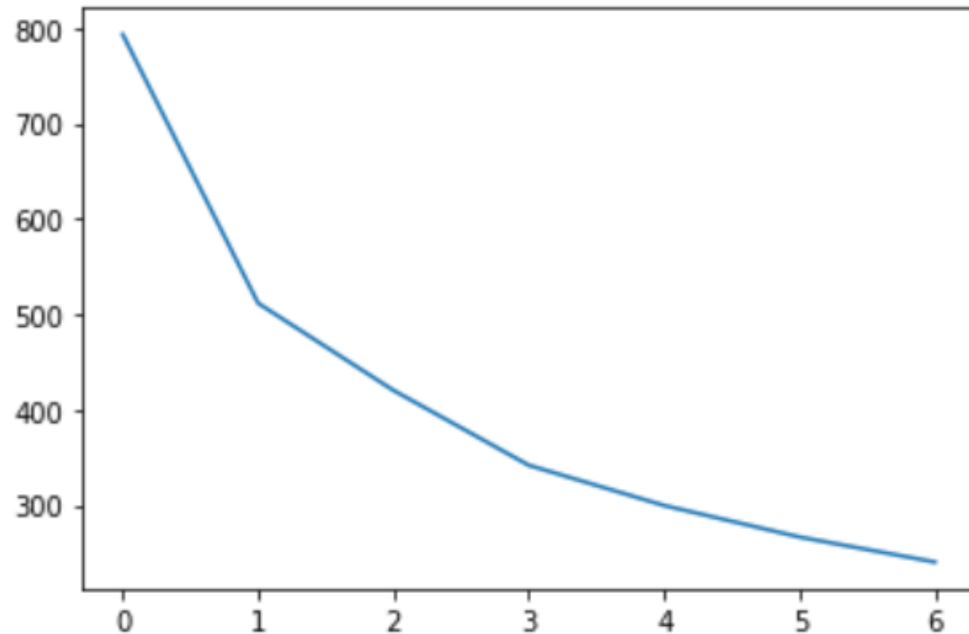
- Univariate and Multivariate analysis was done by making Distribution plots, Box plots and Correlation matrix on single and multiple variables.
- All the variables are Normally distributed in Univariate Analysis
- Correlation among other variables are represented using heatmap



# HOPKINS STATISTICS

From the dataset, we got greater than 90% of the statistical value which is a very good value. The more the dataset is different from the random distribution, the better candidate it is for performing Clustering.

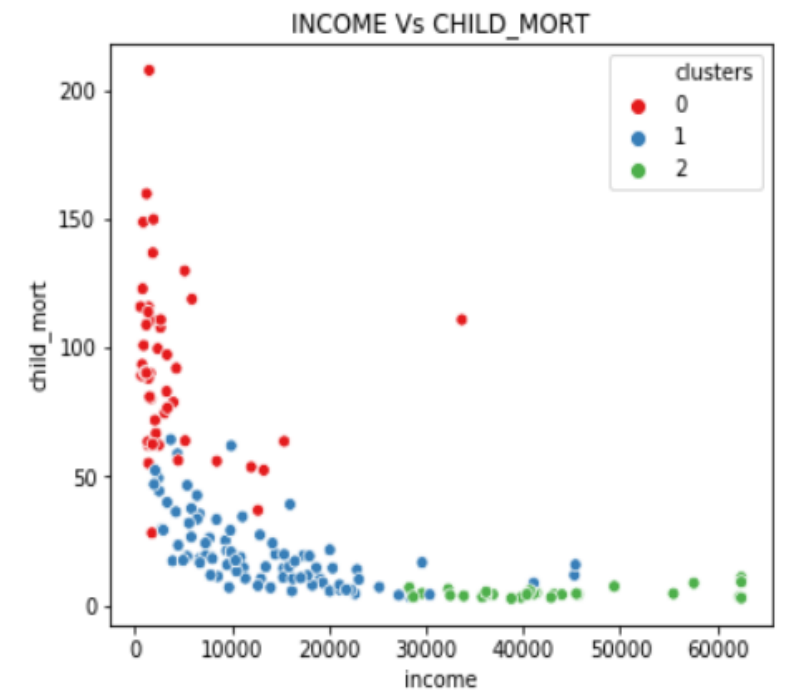
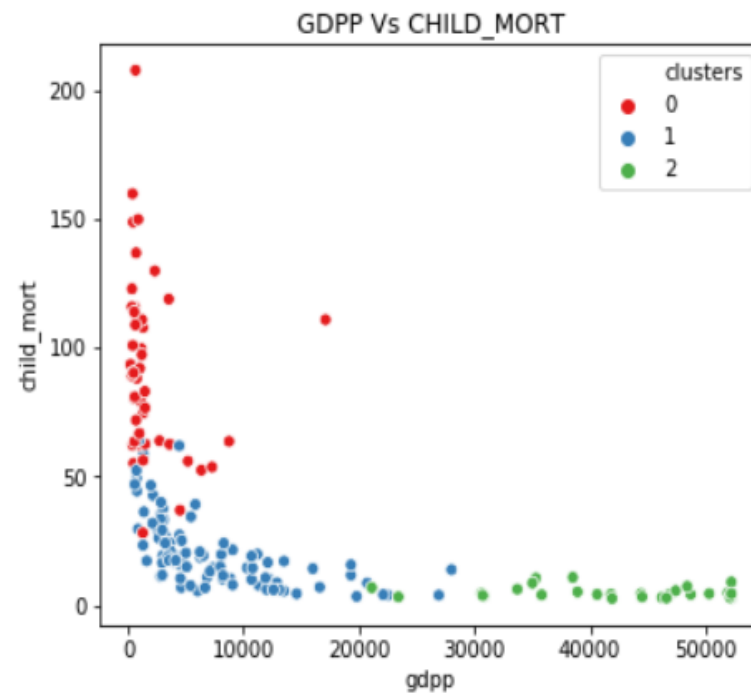
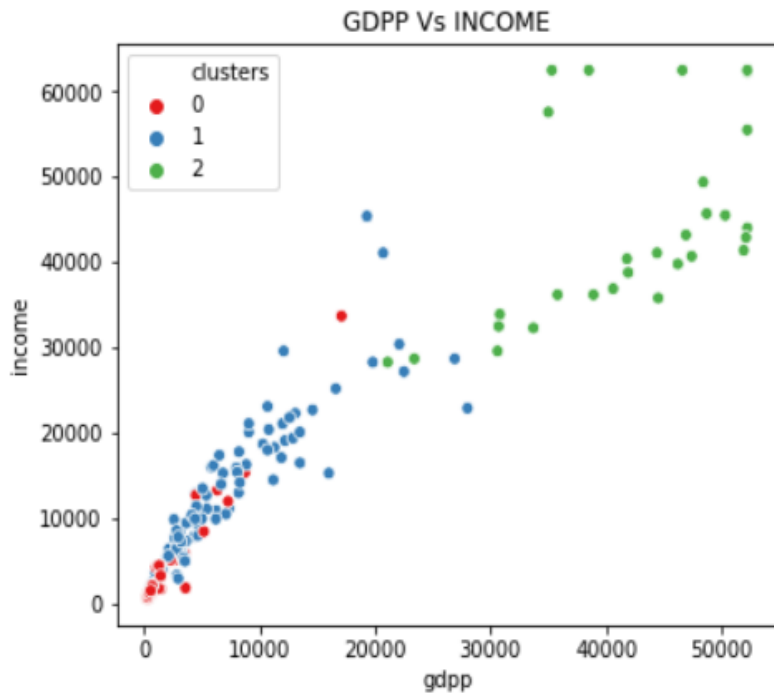
Choosing the number of Clusters by using K Means and Silhouette Score Methods. Both gives the 3 as the good number of clusters



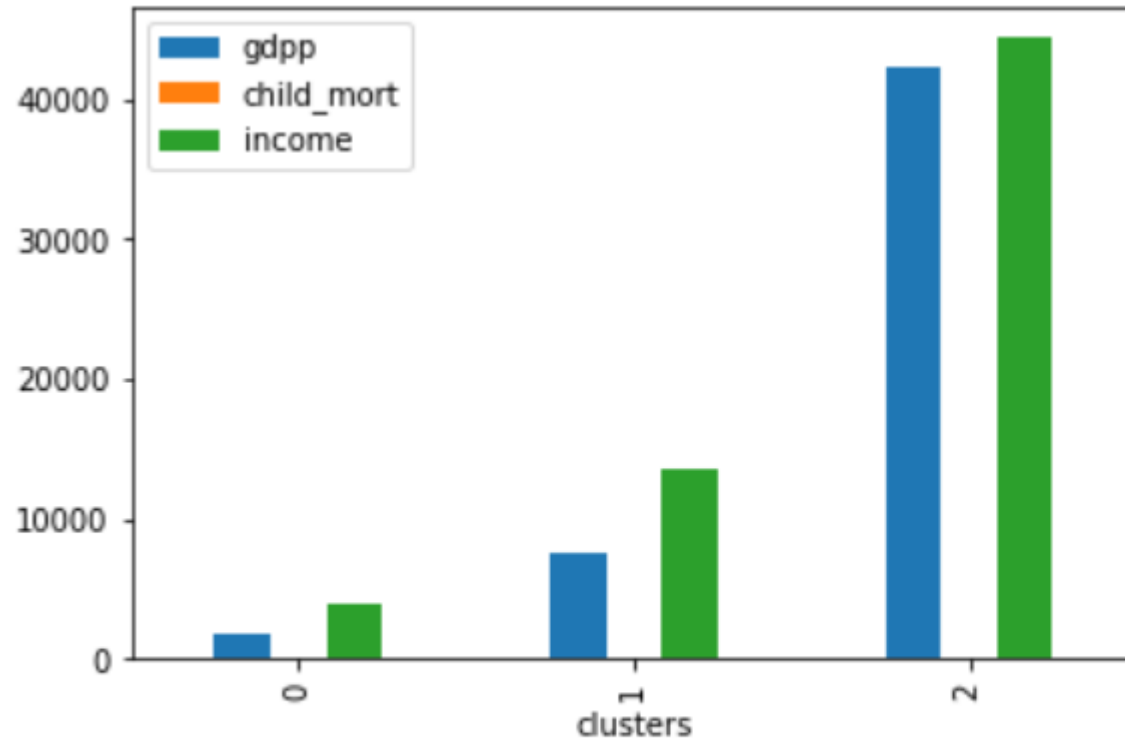
```
n_clusters = 2, the siloutte score is 0.4942188045326538
n_clusters = 3, the siloutte score is 0.4271232295006371
n_clusters = 4, the siloutte score is 0.4091741186271785
n_clusters = 5, the siloutte score is 0.41502573859482306
n_clusters = 6, the siloutte score is 0.3198221355201532
n_clusters = 7, the siloutte score is 0.28779468203946734
n_clusters = 8, the siloutte score is 0.2911356808708908
n_clusters = 9, the siloutte score is 0.3168710640879151
n_clusters = 10, the siloutte score is 0.257422152419032
```

Based on the Elbow curve method, it is clearly seen that  $n\_clusters = 3$

# VISUALIZATION ON THREE MAIN VARIABLES



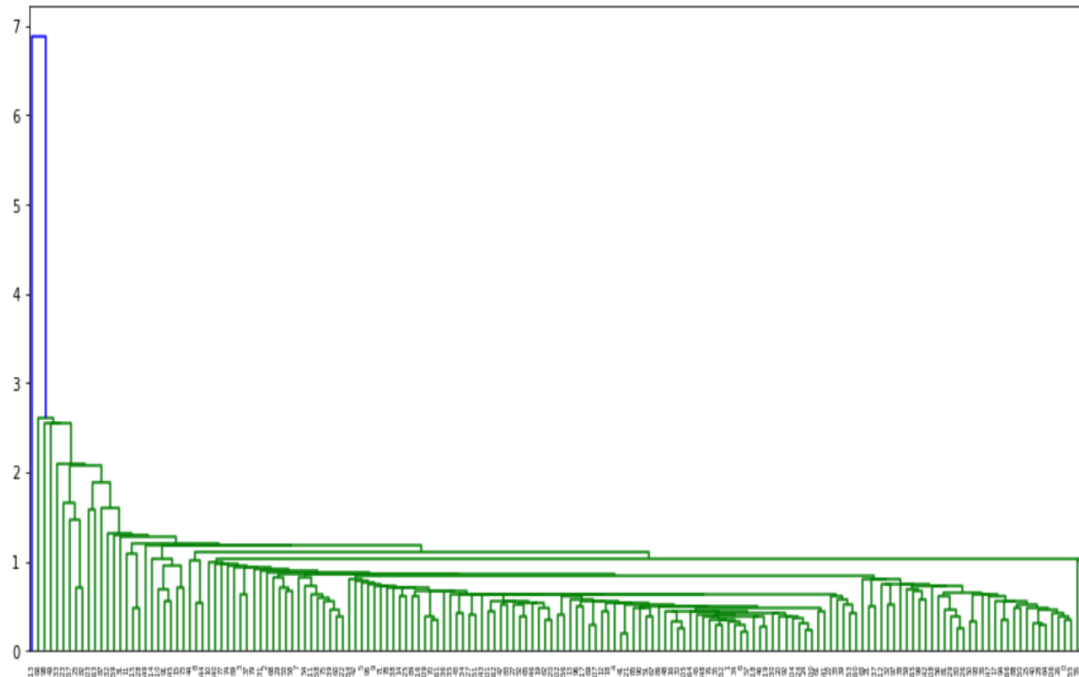
# CLUSTER PROFILING



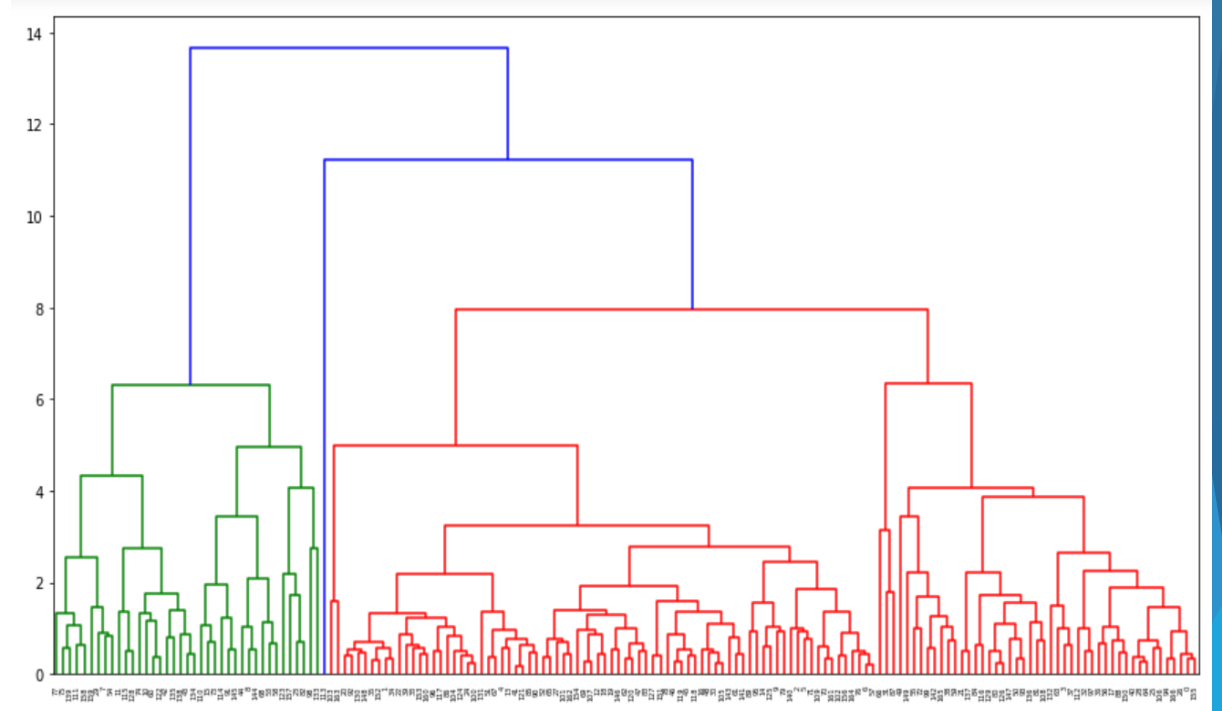
From the bar plot, Cluster 0 is having LOW GDPP, LOW INCOME and HIGH CHILD\_MORT.



# HIERARCHICAL CLUSTERING



Single Linkage



Complete Linkage

# FINAL LIST OF COUNTRIES

Top 10 List of Countries  
who are in dire need of aid

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