HELP International

Shortlisting of countries that need aid

Introduction

Context

- Funds have been raised from a fundraising program.
- The funds have to be used to help the countries.
- There are 167 countries that are on the list
- The countries have to be shortlisted based on socio economic and health data

Solution

Grouping the countries based on socio economic and health data

An unsupervised machine learning algorithm is used for this case scenario .

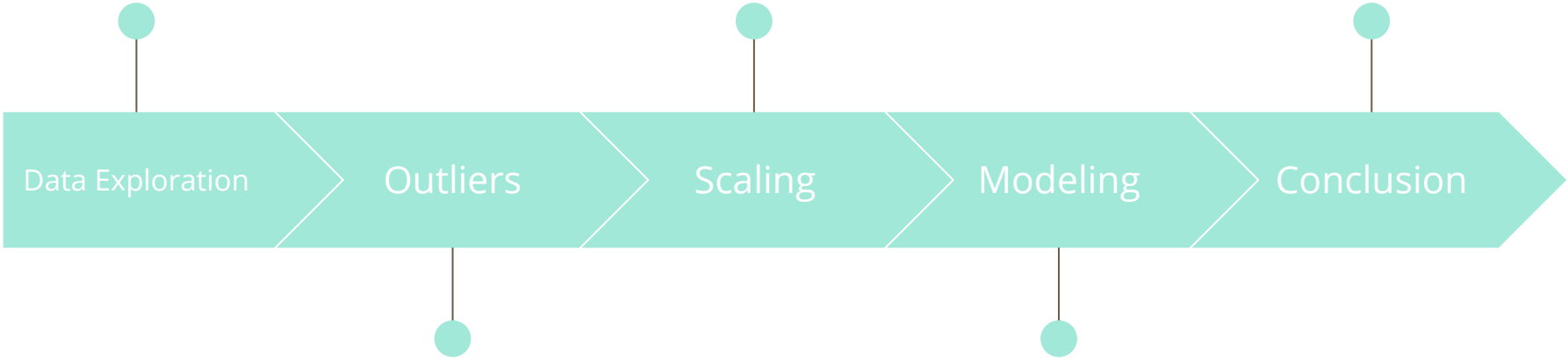
Implementation

Analysis Approach

Data Exploration reveals there are 167 countries

Standard scaling is done to the numerical columns

Countries who need aid are shortlisted



The outliers in columns like gdp,income are soft capped. Child mortality has outliers but is retained.

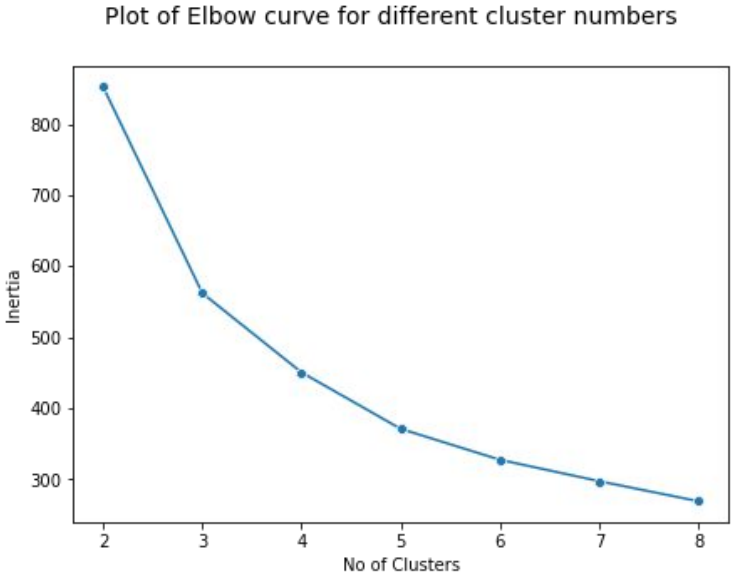
KMeans and Hierarchial clustering is done

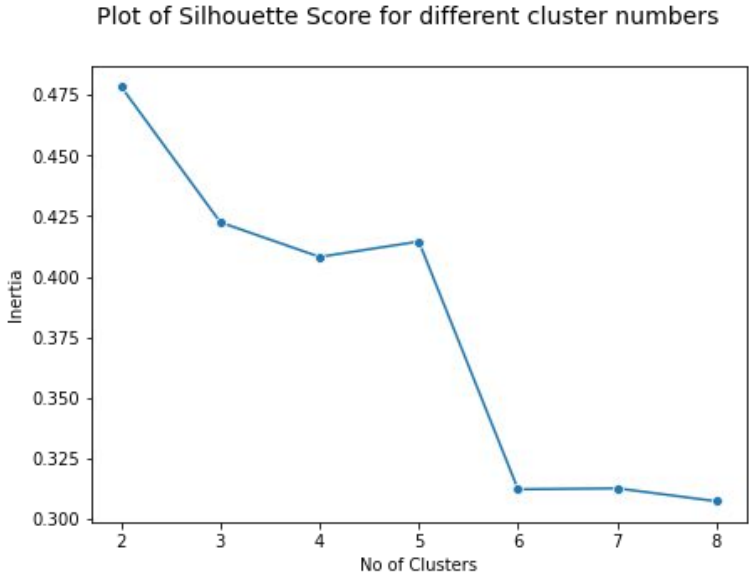
Data Exploration and Preparation

- The data exploration reveals there are 167 countries on the list .
- There are no null values
- Income,exports and import columns are converted to % of gdpp
- There are outliers in the data. The columns for which outliers have been addressed are "income","gdpp","exports","imports","inflation". The others like child mortality, life expectancy are retained.
- 'child_mort','income', 'gdpp','life_expec' are the columns chosen to profile the model.
- Standard scaler is used to scale the features chosen.

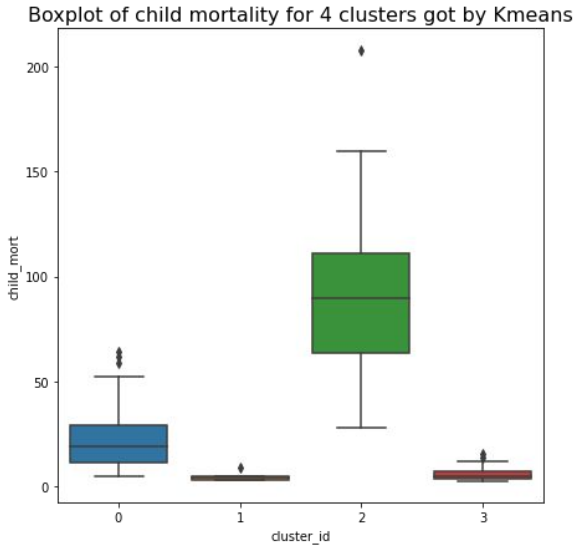
Data Modelling

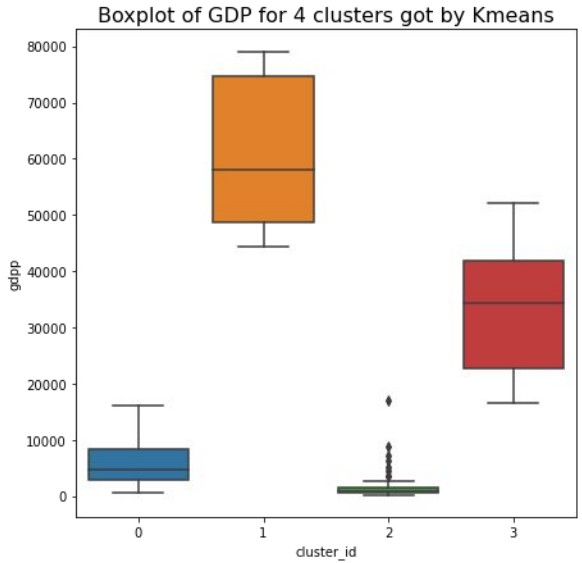
- Hopkins Measure is 89% for the data with chosen features indicating clustering can be done.
- KMeans algorithm is used to model the data after comparing it with Hierarchical method.
- The elbow curve and silhouette analysis is done to come to the conclusion of 4 clusters.

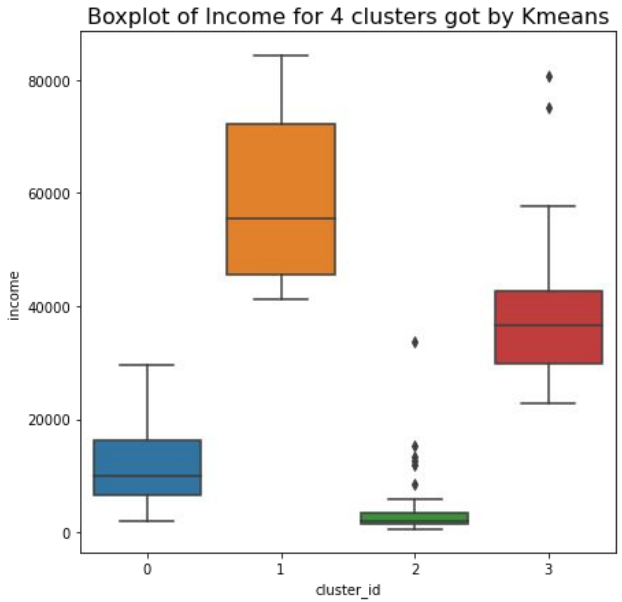




Clusters obtained by Kmeans method

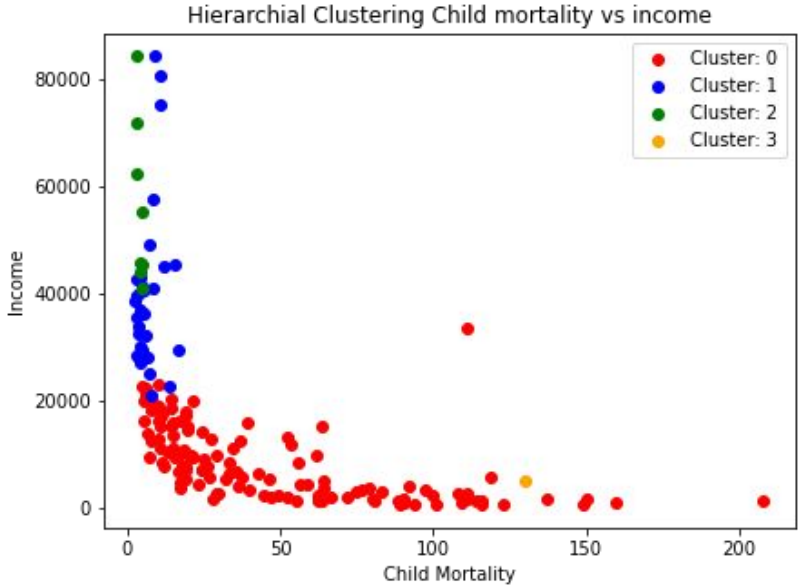


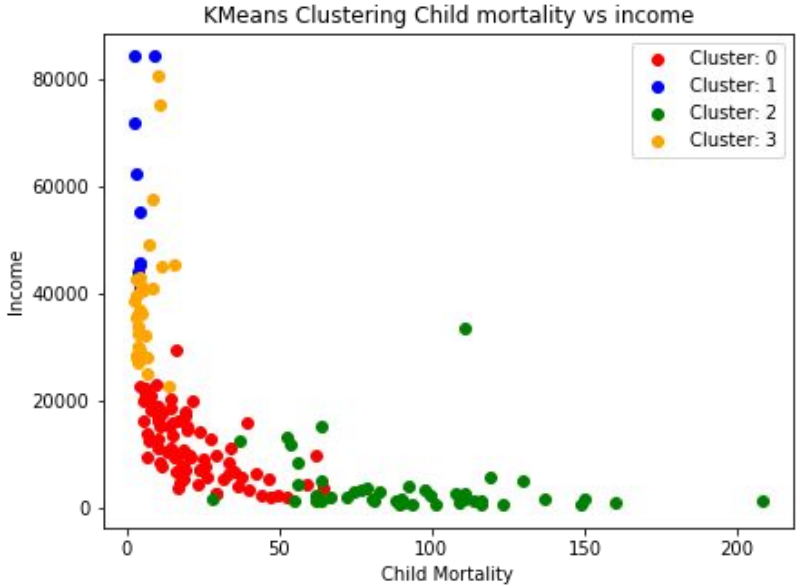




Comparing the hierarchial vs KMeans

ChildMortality vs income

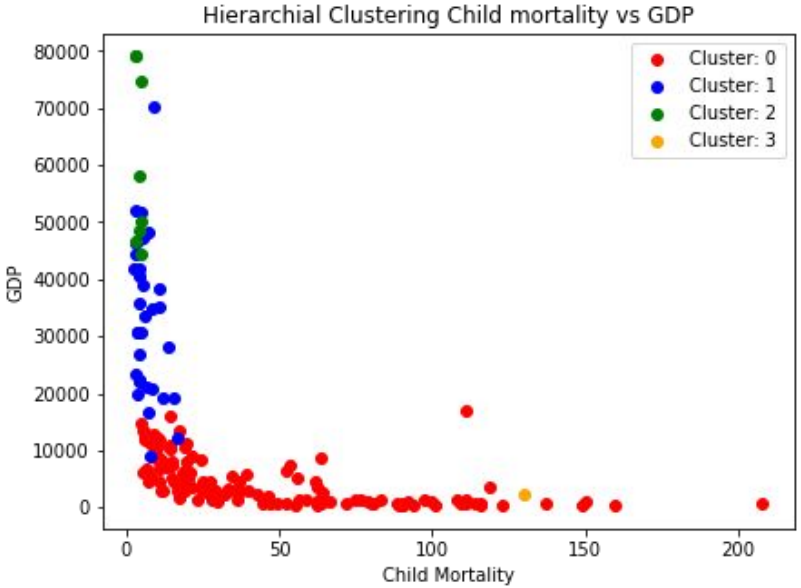


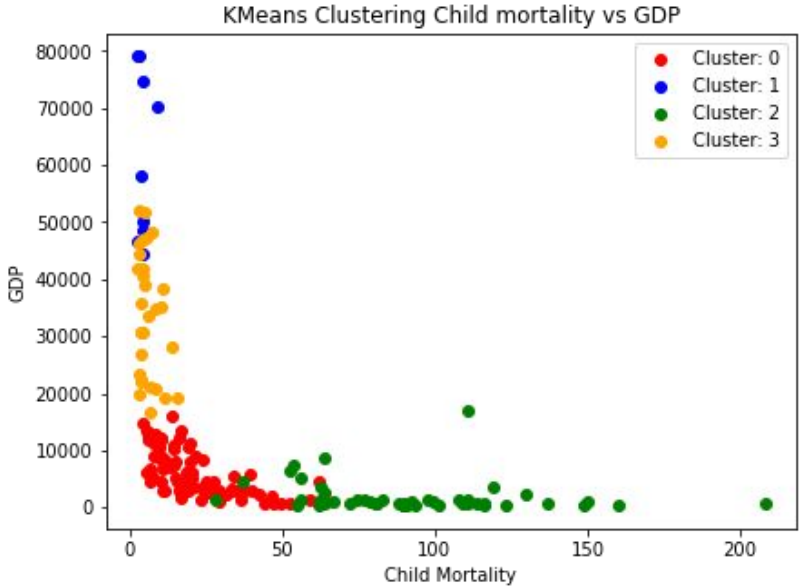


The clustering done by Kmeans method seems to be having a better segregation of the clusters compared to the hierarchical method. Cluster 2 in Kmeans has the highest child mortality rate and low income.

Comparing the hierarchial vs KMeans

ChildMortality vs GDP

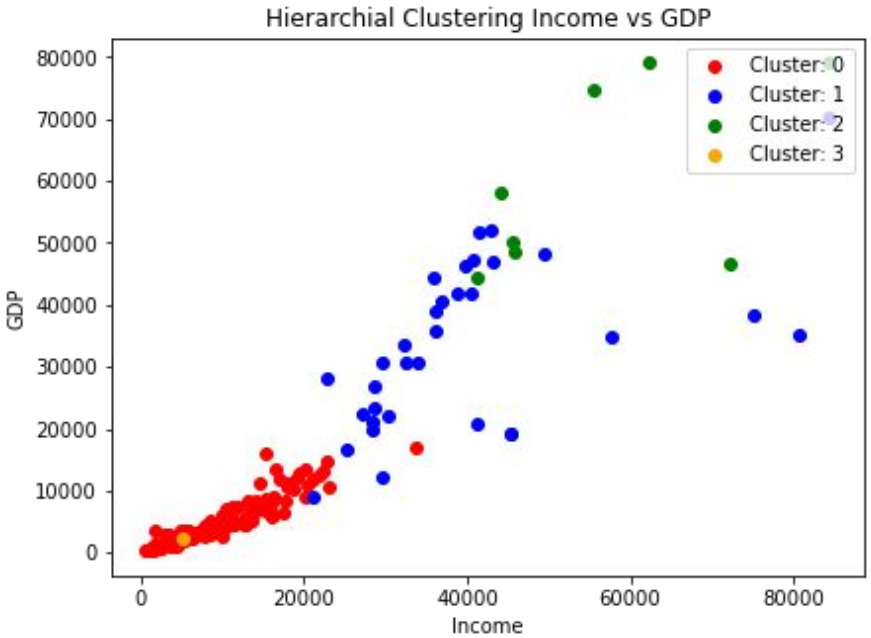


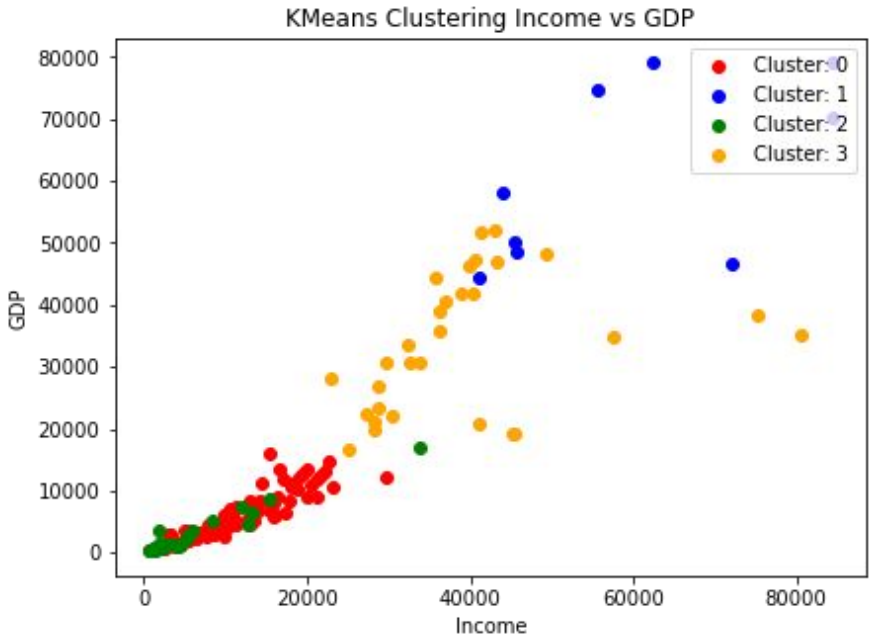


The clustering done by Kmeans method seems to be having a better segregation of the clusters compared to the hierarchical method. Cluster 2 in Kmeans method has the highest child mortality rate and low GDP.

Comparing the hierarchial vs KMeans

Income vs GDP





The clustering done by Kmeans method seems to be having a better segregation of the clusters compared to the hierarchical method. Cluster 2 in Kmeans method has the lowest income and low GDP.

List of 5 countries that have been shortlisted

The countries were sorted by high child mortality rate, low income and low gdp.

1. Haiti,
2. Sierra Leone
3. Chad
4. Central African Republic
5. Mali