

# HELP INTERNATIONAL

---

PROJECT REPORT

VISHAKHA CHAVAN

# ABSTRACT

---

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.

And this is where you come in as a data analyst. Your job is to categorise the countries using some socio-economic and health factors that determine the overall development of the country. Then you need to suggest the countries which the CEO needs to focus on the most.

# APPROACH

---

- Exploratory Data Analysis
- Outlier Analysis and Outlier Treatment
- Data Pre-processing
- Hopkins Statistic and Elbow Curve
- Silhouette Score Analysis
- K-Means Clustering
- Hierarchical Clustering
- Final Model Selection

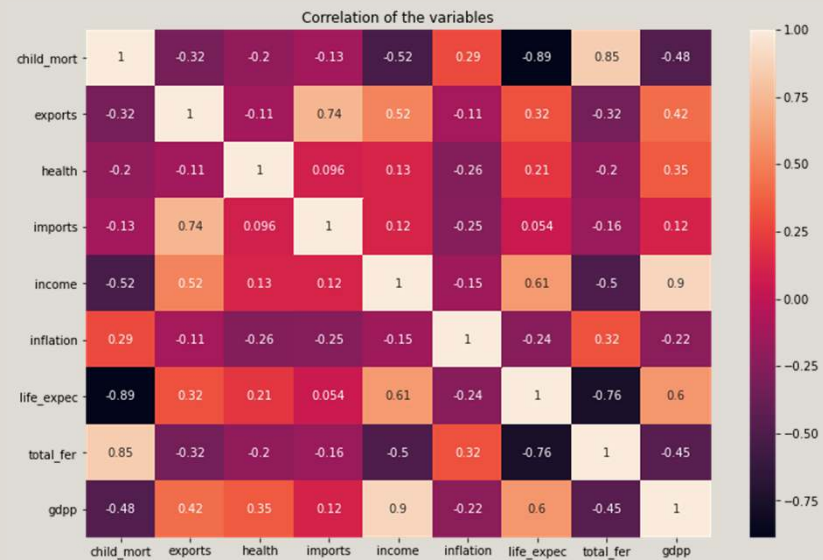
# DATA EXPLORATION

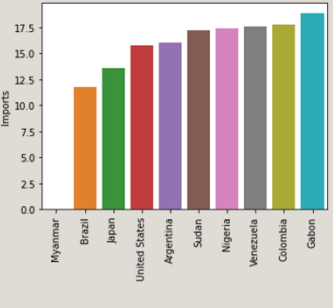
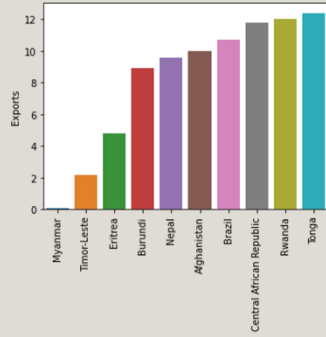
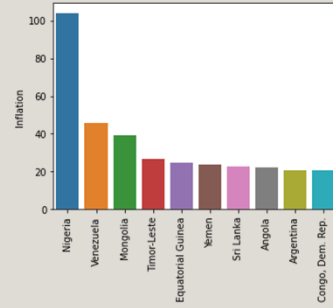
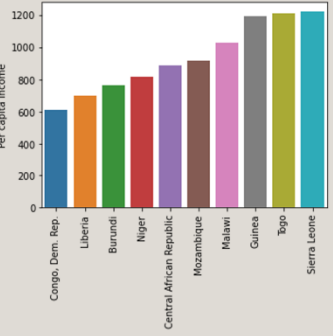
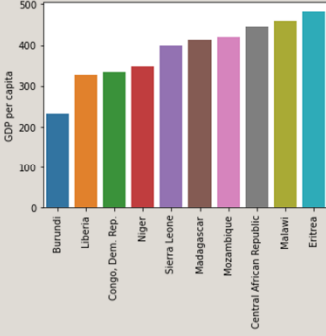
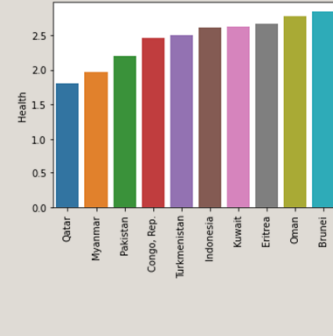
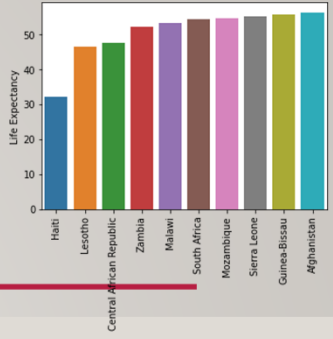
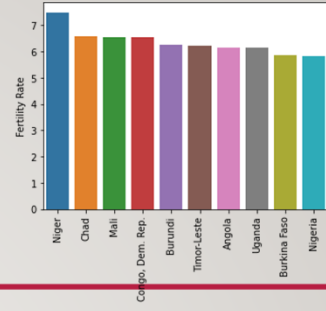
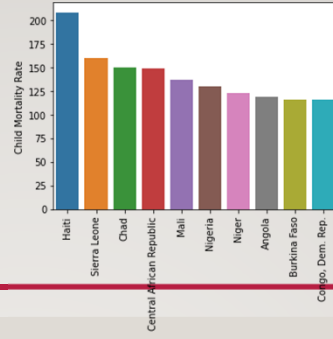
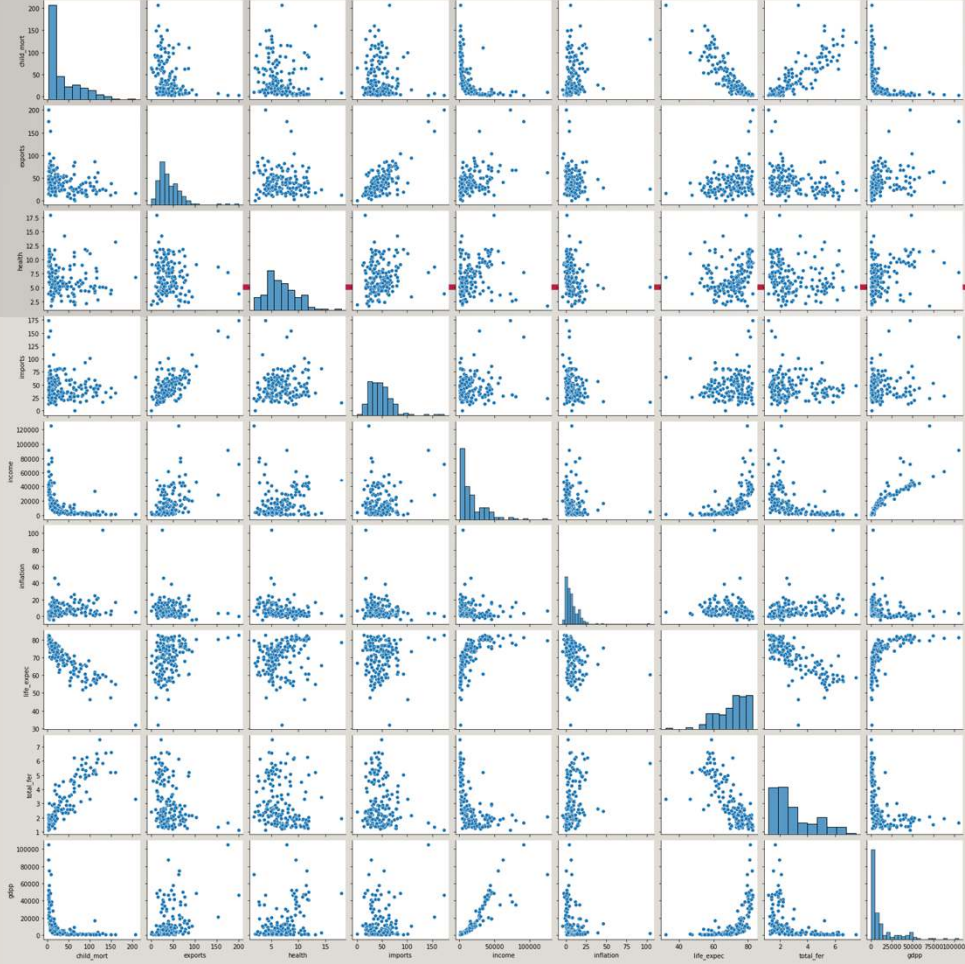
---

- We do not have any missing values here.
- The data here is right-skewed (except life span), which indicates the presence of outliers.
- The outliers in our data are completely acceptable from a Business perspective, as we have a variety of poor countries to highly developed countries.
- To get rid of skewness and make our data normal, the variables have been transformed using Power Transformer from SKLearn.

# EDA

- Pairplots, heatmap and barplots have been plotted to study the relationship between various variables.



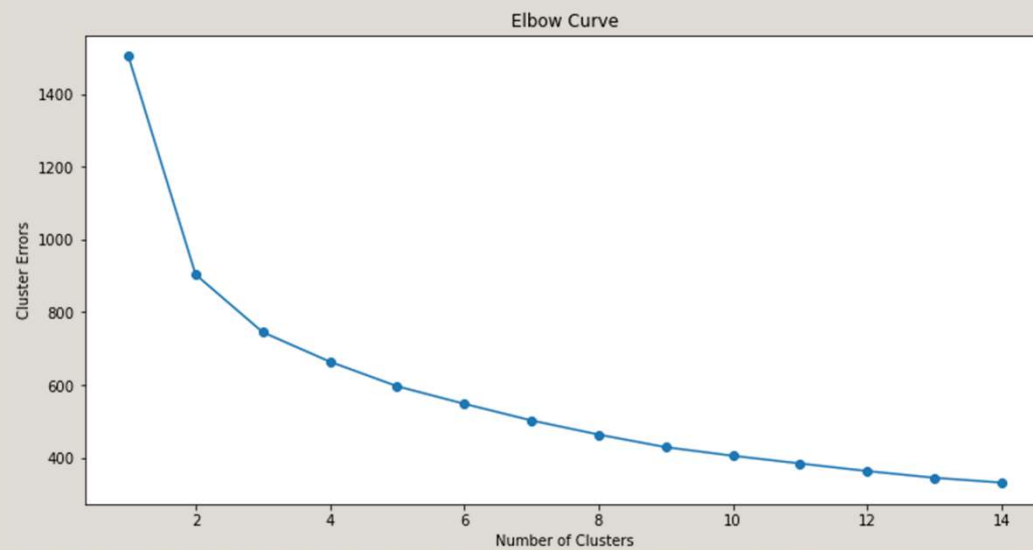




# HOPKINS STATISTIC & ELBOW CURVE

---

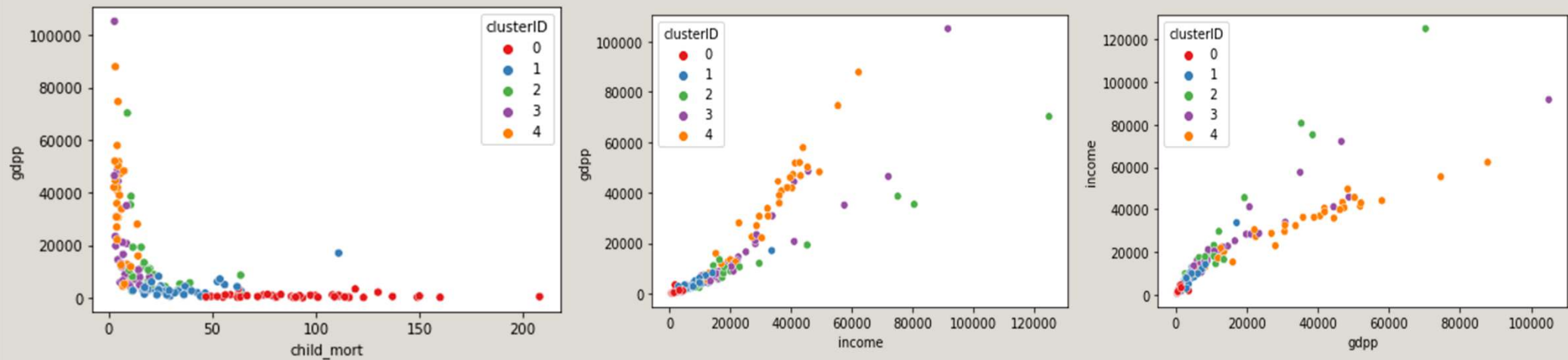
- From elbow curve, we can see that the optimal number of clusters are 3.
- Also from Hopkins statistic (0.78) we can say that there is a high chance of clustering to be formed.



# K-MEANS MODELLING

---

- Using 3 optimal clusters and init method as 'K-Means++', we have built the model using fit method.
- The clustering for child mortality rate, gdp, and income is as follows.

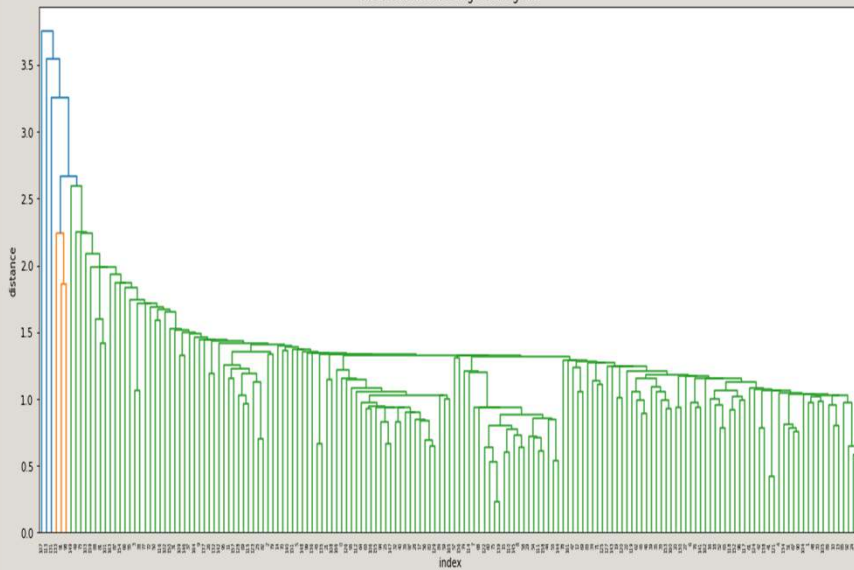




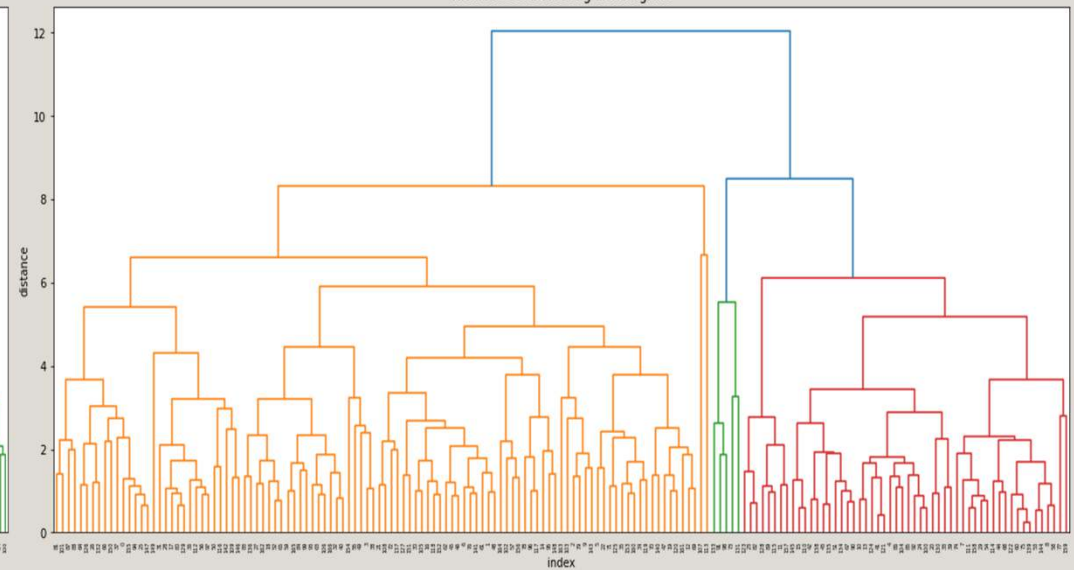
# HIERARCHICAL CLUSTERING

---

Hierarchical Clustering Dendrogram



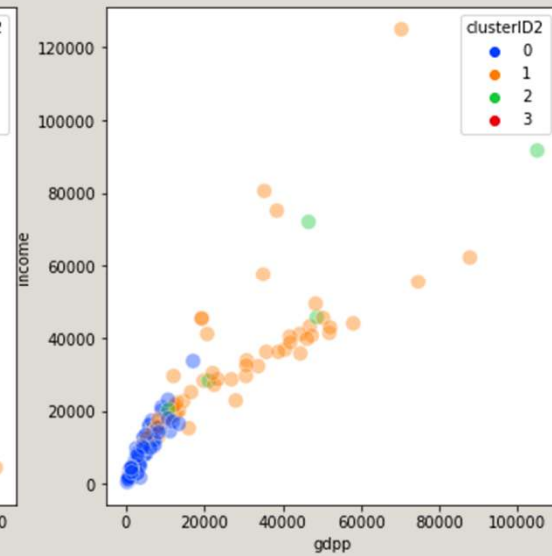
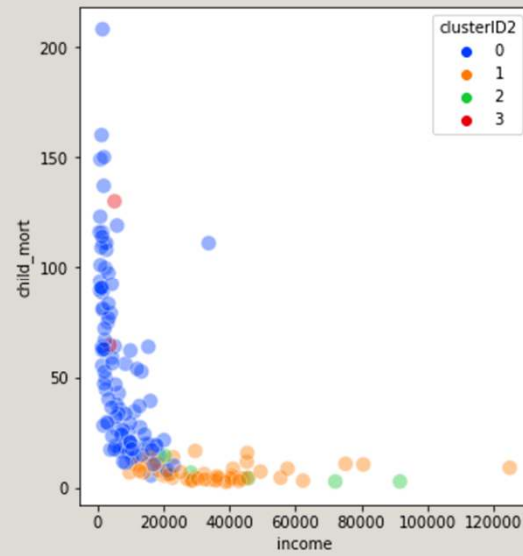
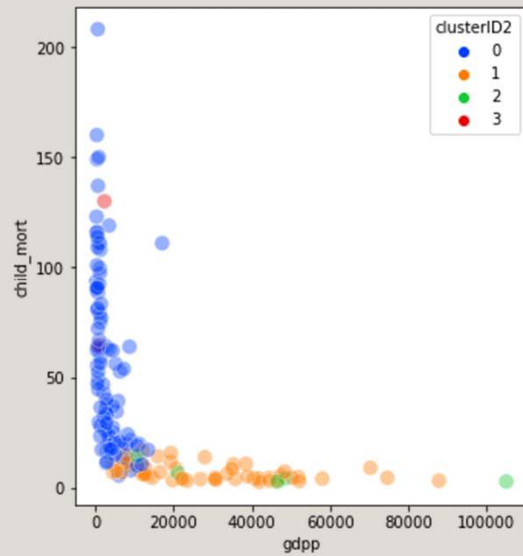
Hierarchical Clustering Dendrogram



# HIERARCHICAL CLUSTERING

---

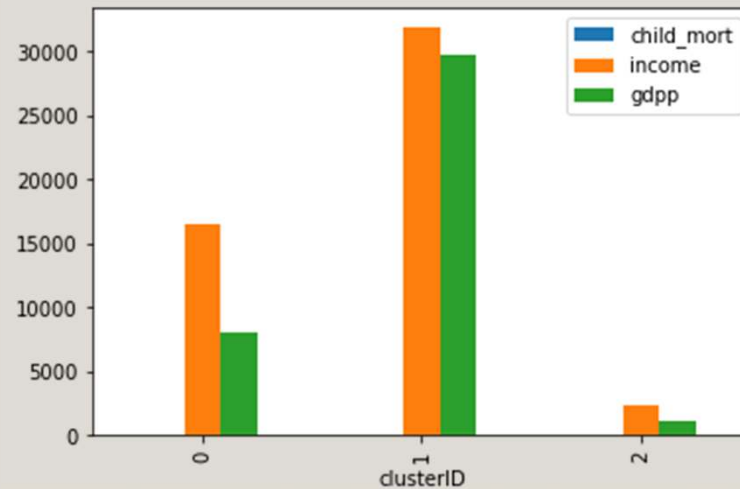
- From hierarchical clustering, the clustering for child mortality rate, gdp, and income is as follows.



# FINAL SELECTION

---

- We have selected the K-Means Algorithm for this particular problem statement as it is more accurate.
- The cluster-2 has been selected from this Algorithm for sorting the countries.



# FINAL LIST OF COUNTRIES

---

1. Burundi
2. Liberia
3. Congo, Dem. Rep
4. Niger
5. Sierra Leone
6. Madagascar
7. Mozambique
8. Central African Republic
9. Malawi
10. Eritrea