

# CLUSTERING – GRADED ASSIGNMENT



NGO HUMANITARIAN AID – CLUSTERING

**Bhaveshkumar Thaker**

UPGRAD – CU – M.SC. IN DATA SCIENCE

# 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.
- NGO have been able to raise around \$10 million.
- The CEO of the NGO needs to decide how to use this money strategically and effectively.
- The significant issues CEO is facing is related to choosing the countries that are in the direst need of aid.
- **Business Objective:**
  - CEO of HELP International NGO wants to know the top 5 or 10 countries which are in direst need of aid.
  - For that they want to build a Model which identifies the countries in need of Aid.

# Analysis Approach

- Load and Clean the Data
  - Check for Null values and convert percentage values to normal values
- Perform Exploratory Data Analysis (EDA)
  - Perform Univariate analysis and Bivariate analysis to understand the data
- Data Preparation
  - Standardize the data
- Assess Clustering Tendency
  - Perform Hopkins' Statistic test on data to find clustering tendency of the data provided.
- Apply Principal Component Analysis (PCA)
  - Make scree plot
  - Apply PCA and identify number of components which doesn't have correlation between them
- Perform Clustering using Kmeans
  - Find optimal number of clusters using Elbow Curve, and Silhouette Analysis
  - Perform KMeans clustering with optimal number of clusters
  - Identify Countries which are in the direst need of aid.
- Perform Clustering using Hierarchical Clustering
  - Perform complete linkage Hierarchical Clustering and cut the tree with optimal number of clusters
  - Identify Countries which are in the direst need of aid.
- Conclusions and provide recommendations



# Data Preparation

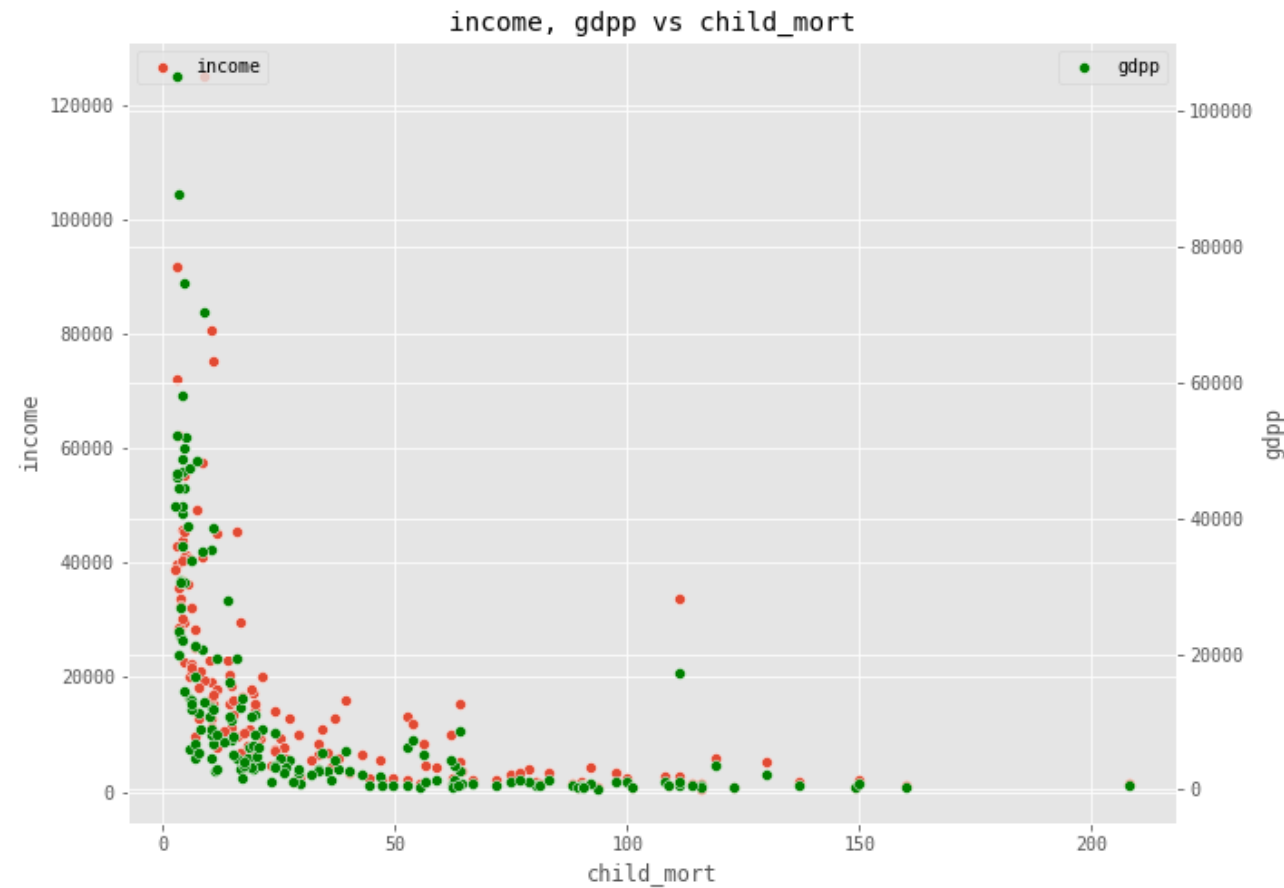
- Convert percentage values to normal values for various features
- Standardize all the features



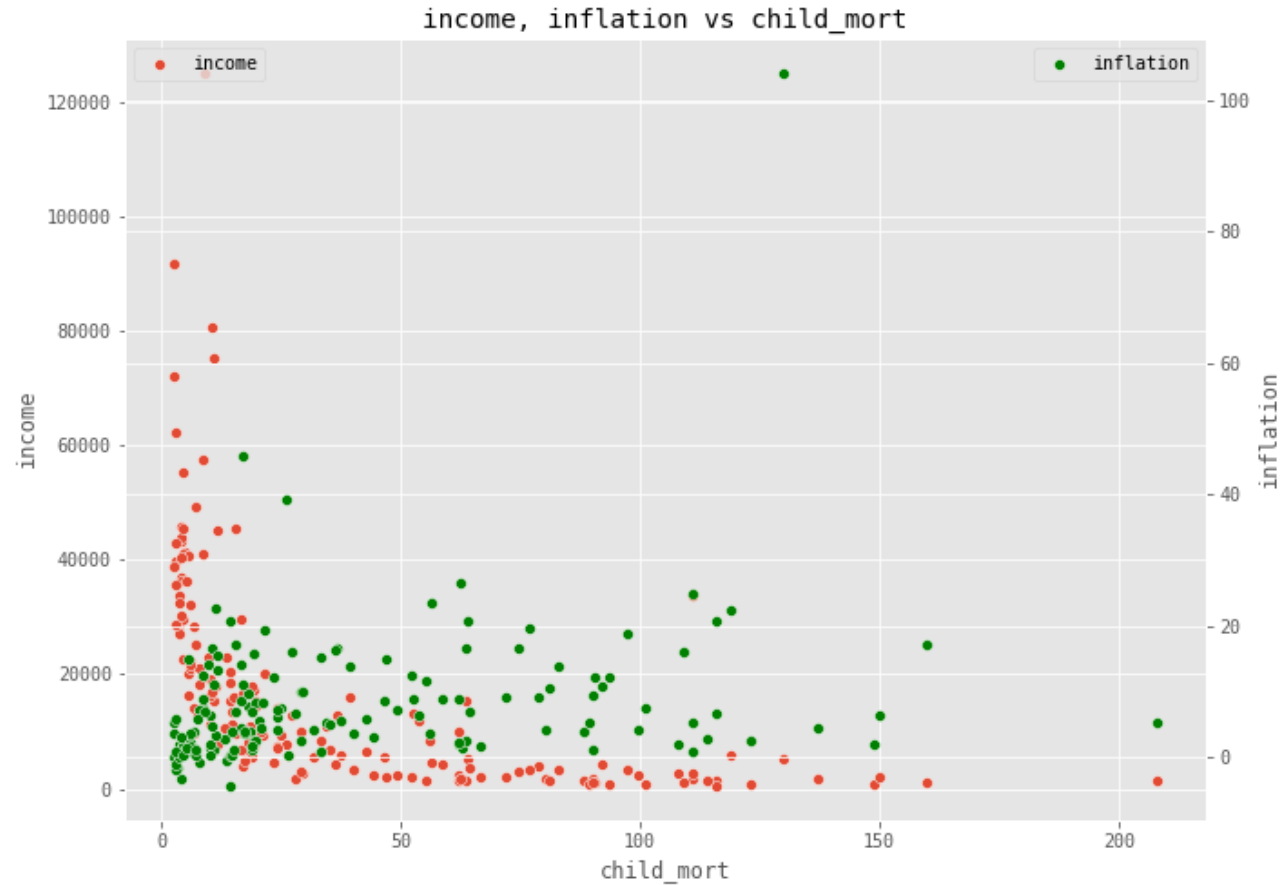
# VISUALIZATIONS

Summary

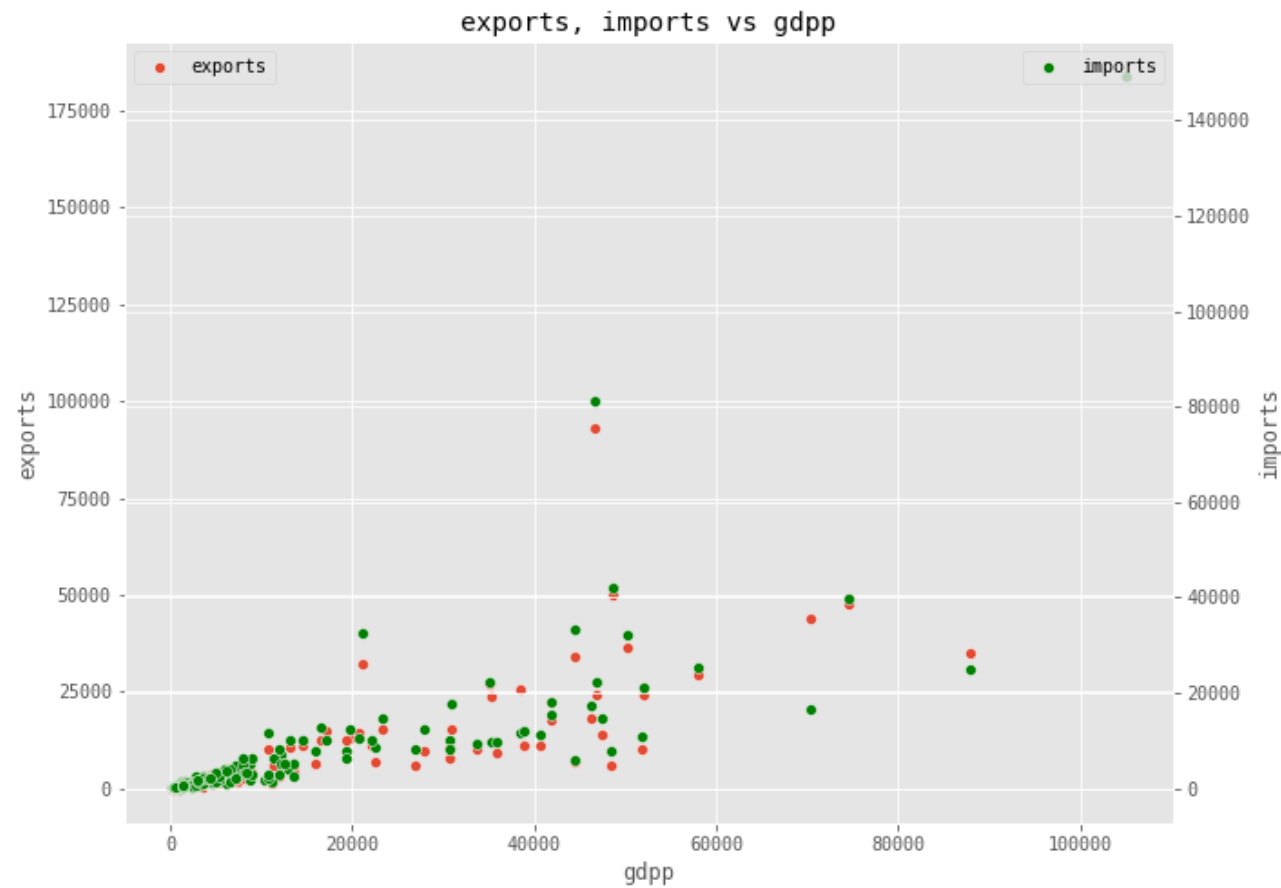
# Income, GDP vs Child Mortality



# Income, Inflation vs Child Mortality

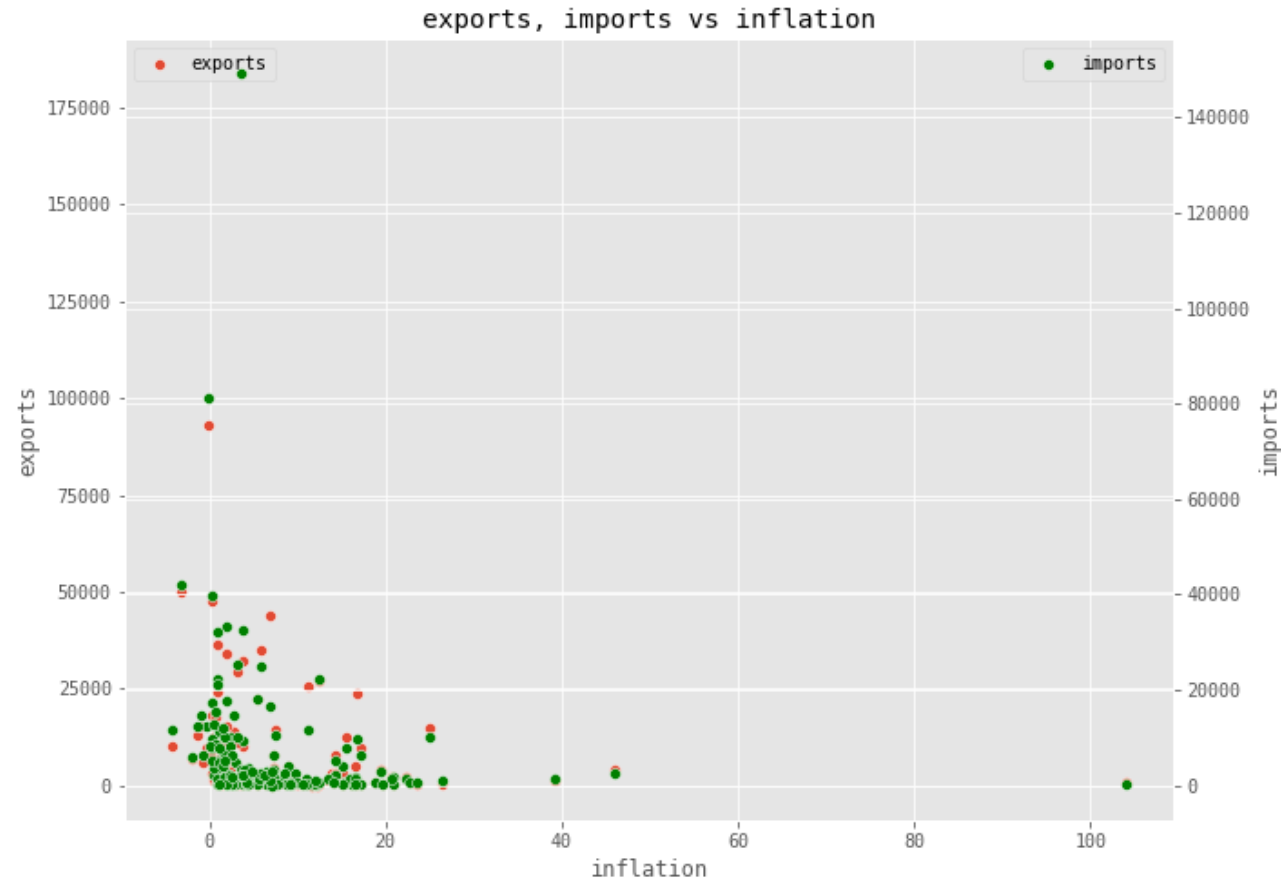


# Imports, Exports vs GDP





# Imports, Exports vs Inflation

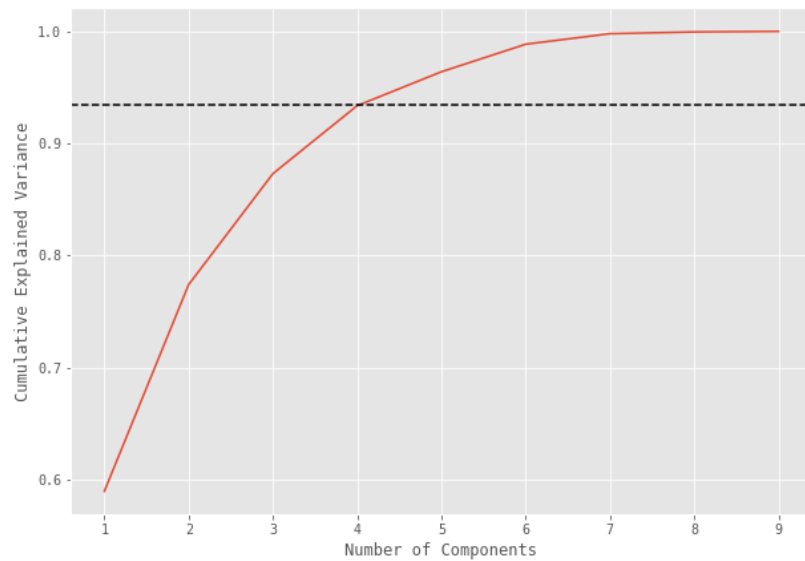


# PRINCIPAL COMPONENT ANALYSIS

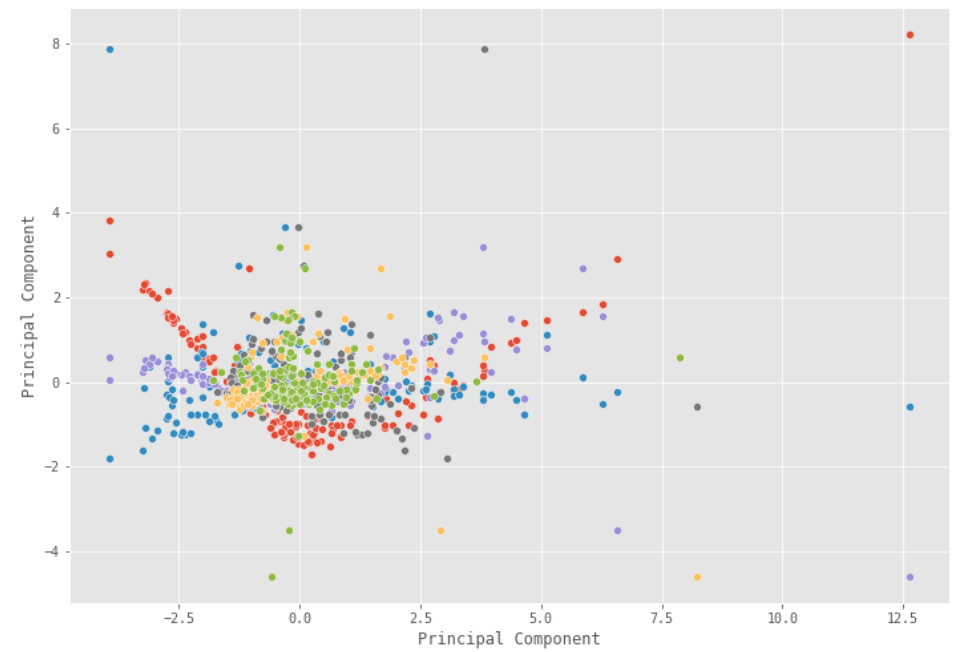


# PCA

## Scree Plot



## PCA Features

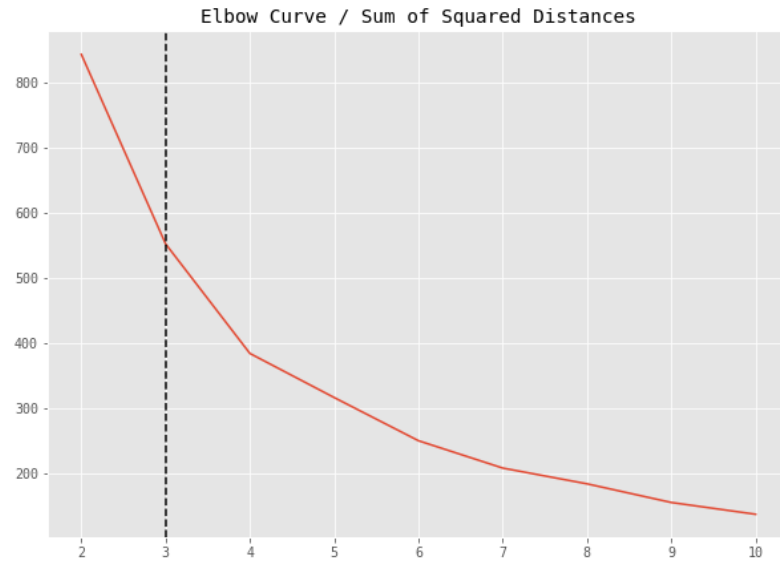


# KMEANS CLUSTERING

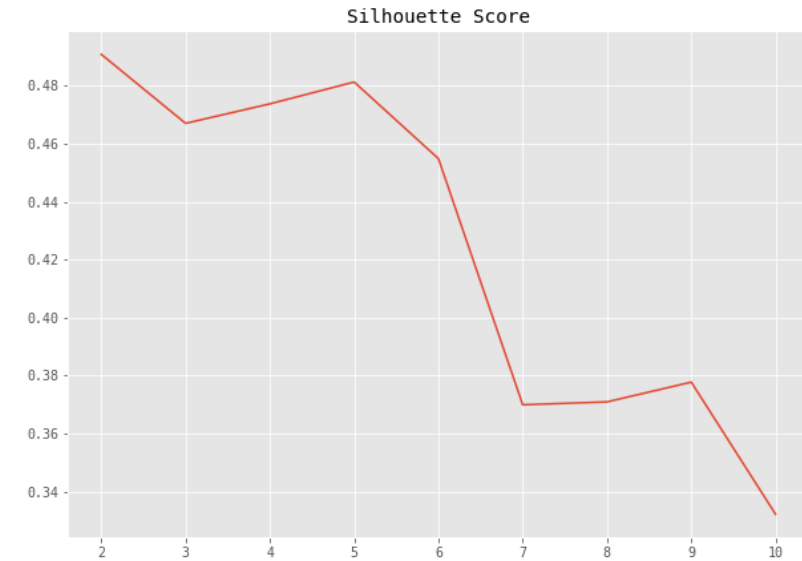


# Optimal Number of Clusters

## Elbow Curve

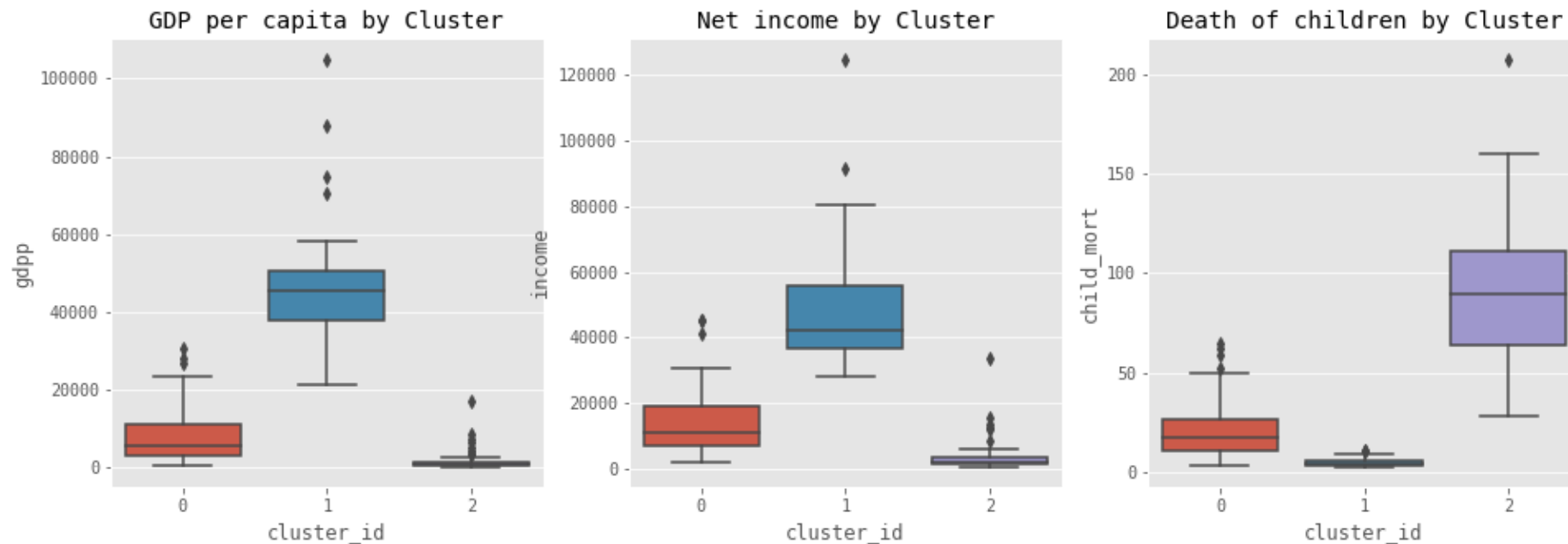


## Silhouette Score



# Clusters identified using KMeans

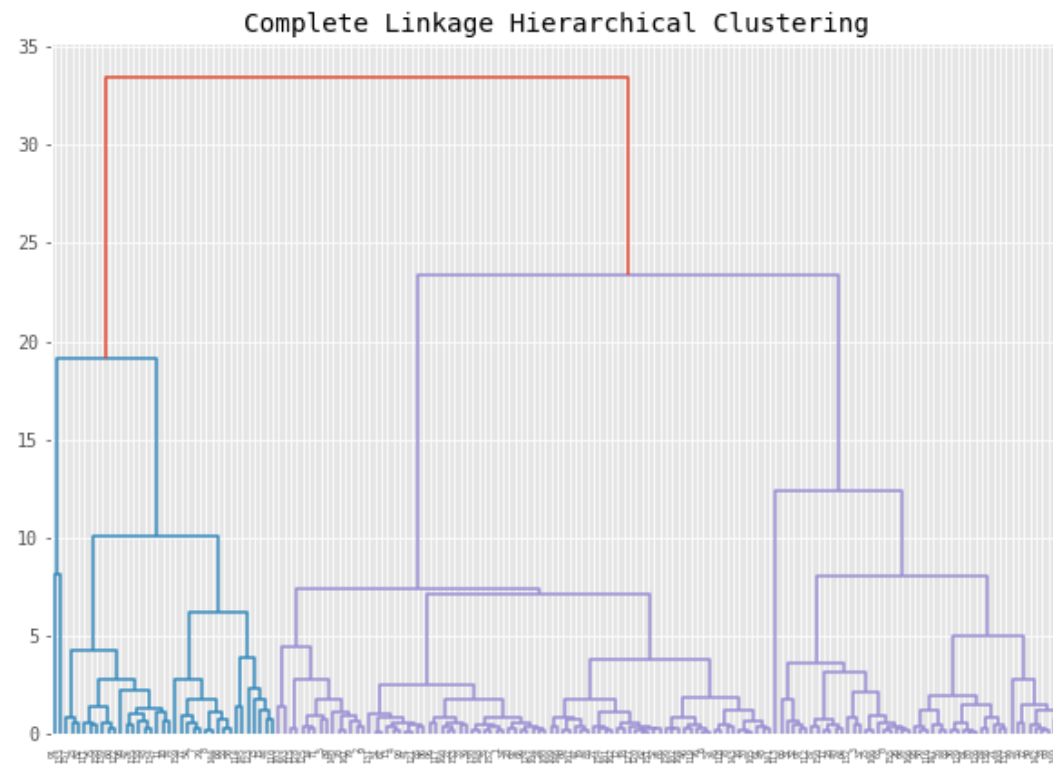
- Cluster 0 - Has medium child mortality, medium GDP and medium income (Developing)
- Cluster 1 - Has low child mortality, high GDP and high income (Developed)
- Cluster 2 - Has high child mortality, low GDP and low income (Under Developed)



# HIERARCHICAL CLUSTERING



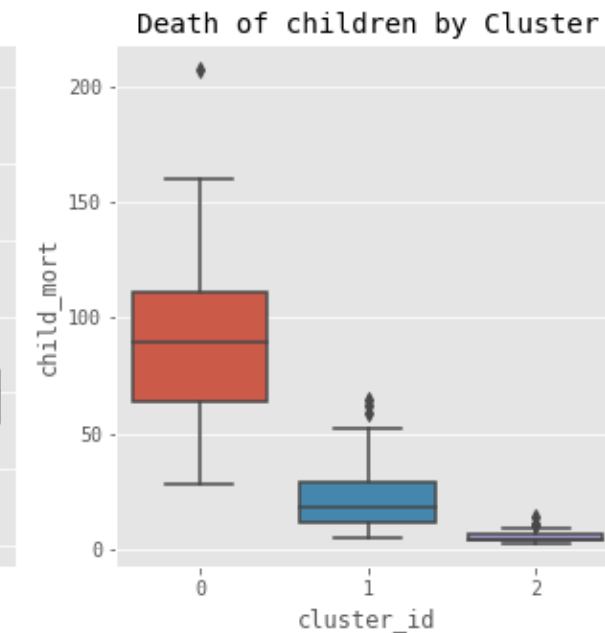
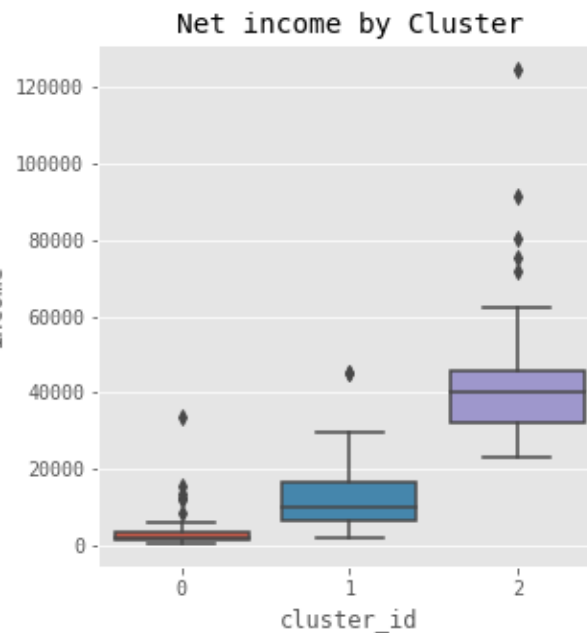
# Complete Linkage Clustering





# Clusters identified using Hierarchical

- Cluster 0 - Has high child mortality, low GDP and low income (Under Developed)
- Cluster 1 - Has medium child mortality, medium GDP and medium income (Developing)
- Cluster 2 - Has low child mortality, high GDP and high income (Developed)





# SUMMARY



# Summary

- CEO of HELP International NGO should provide AID to following countries, which are in direst need of aid/help
  - Haiti
  - Sierra Leone
  - Chad
  - Central African Republic
  - Mali
  - Nigeria
  - Niger
  - Angola
  - Congo, Dem. Rep.
  - Burkina Faso