HELP INTERNATIONAL - Clustering Assignment

Submission By:

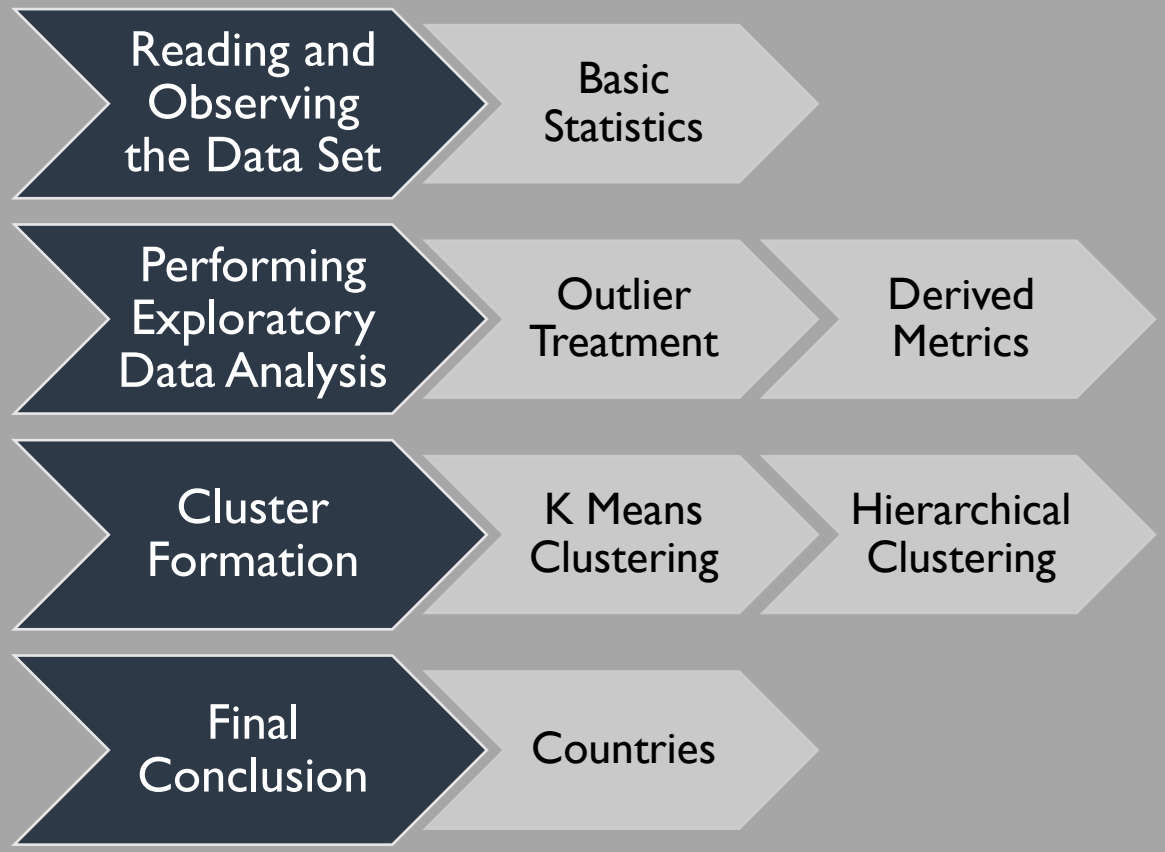
Gaurav Khare

Business Objective

The aim of this analysis is to enable HELP International to make a strategic investment decision to choose countries that are direst need of aid on basis of Socio-Economic And Health factors that are the indicators of overall development of the country

The categorization will suggest us the clusters thus formed by countries which required from Help International.

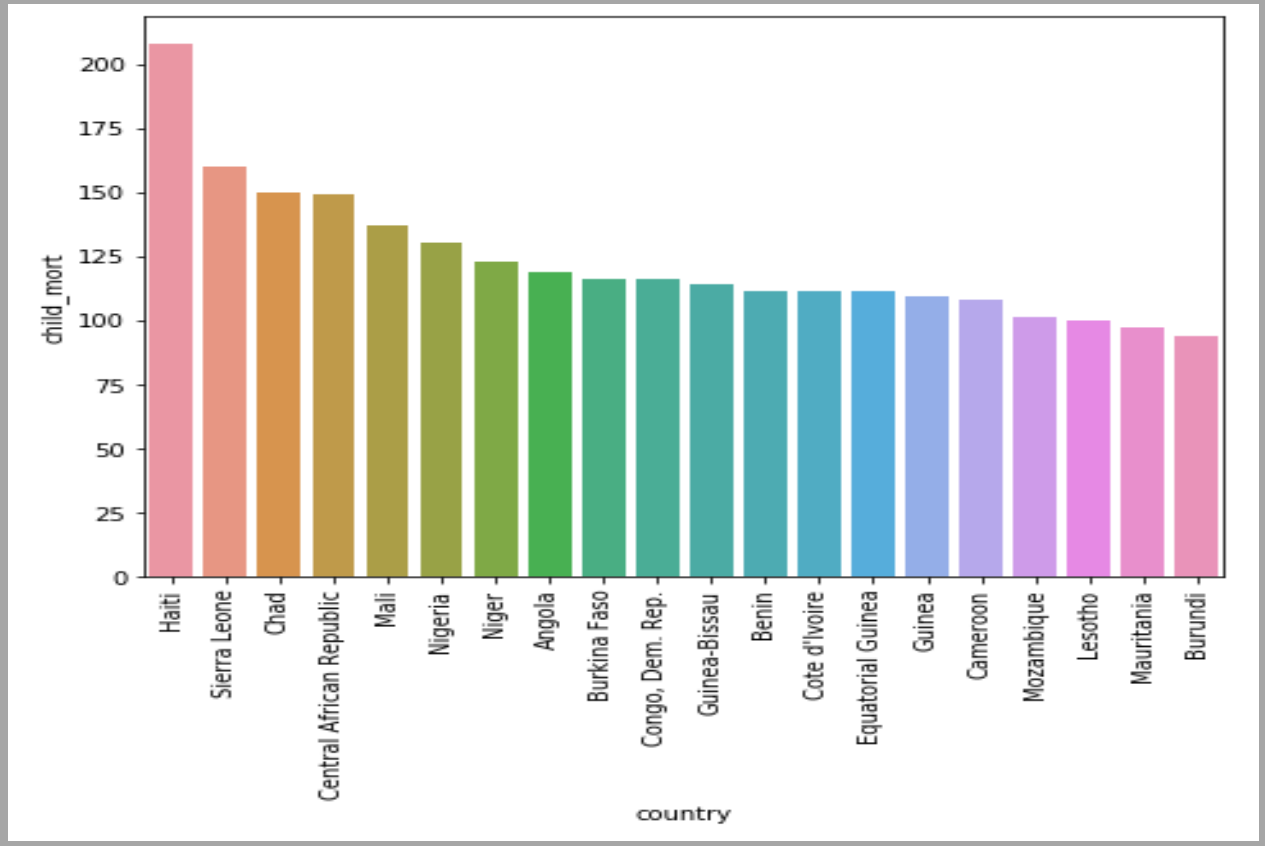
Approach to the Problem Statement



Country vs Child Mortality

The Top 5 Countries with High Child Mortality

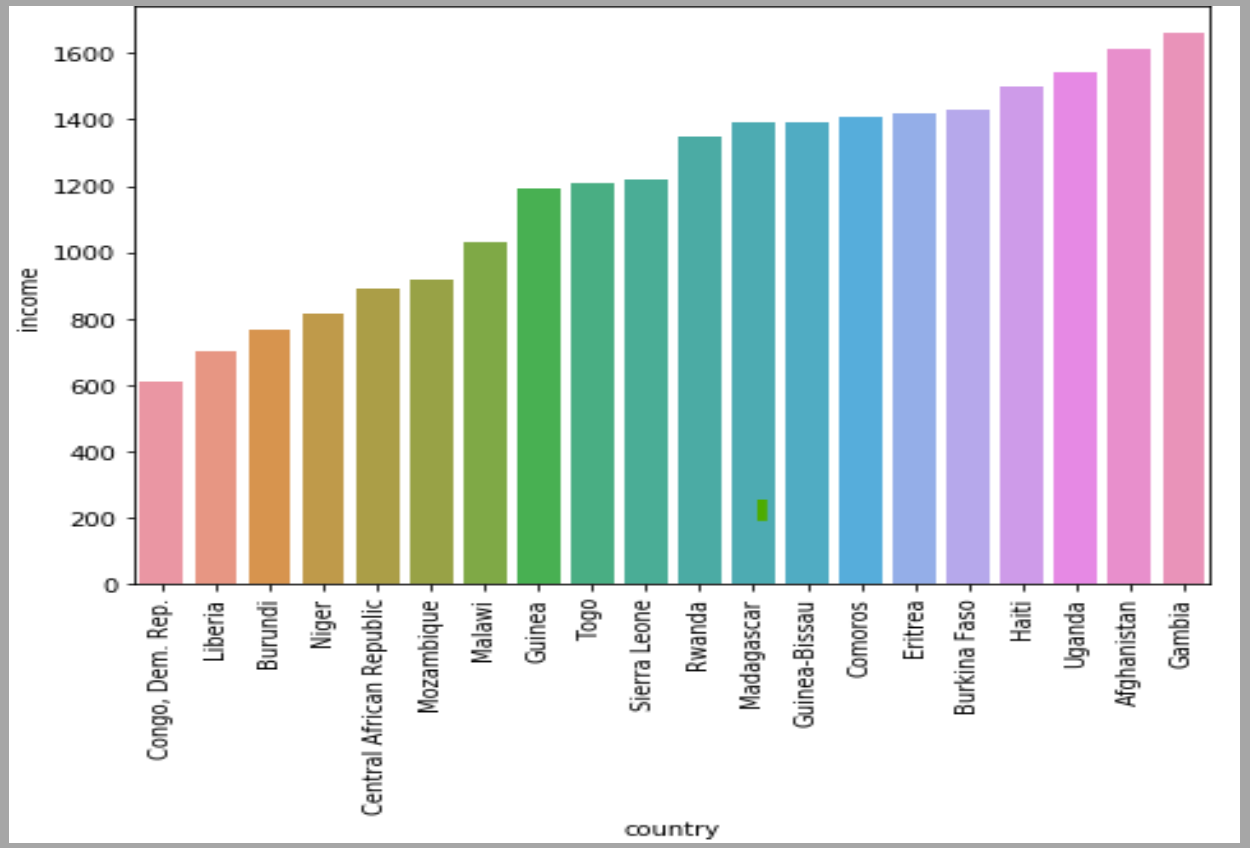
- Haiti
- Sierra Leone
- Chad
- Central African Republic
- Mali



Country vs Income

Top 5 Countries with low Income

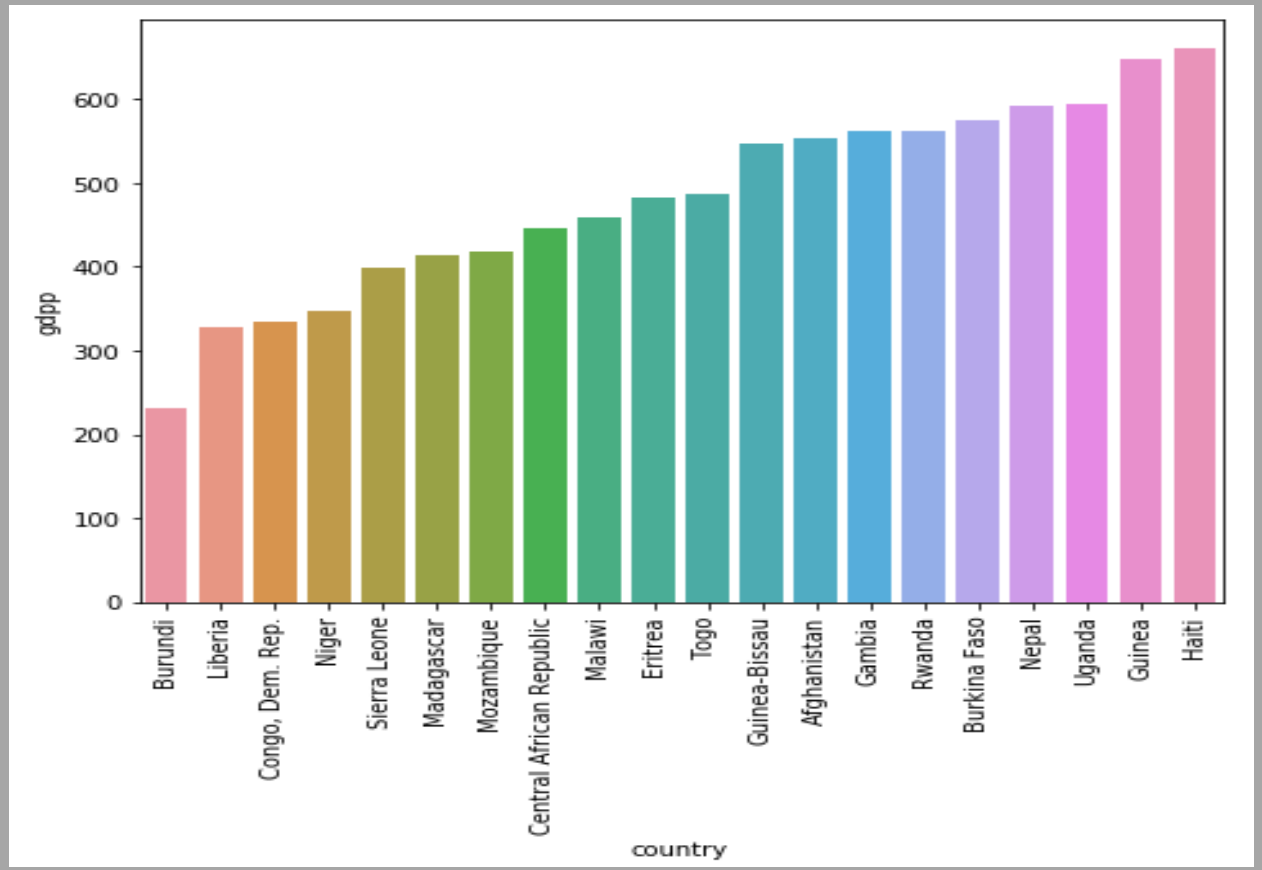
- Congo. Dem. Rep.
- Liberia
- Burundi
- Niger
- Central African Republic



Country vs GDPP

Top 5 Countries with low GDPP

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



Correlation between variables

Child Mortality & Life Expectancy

Child Mortality & Total Fertility

Correlation

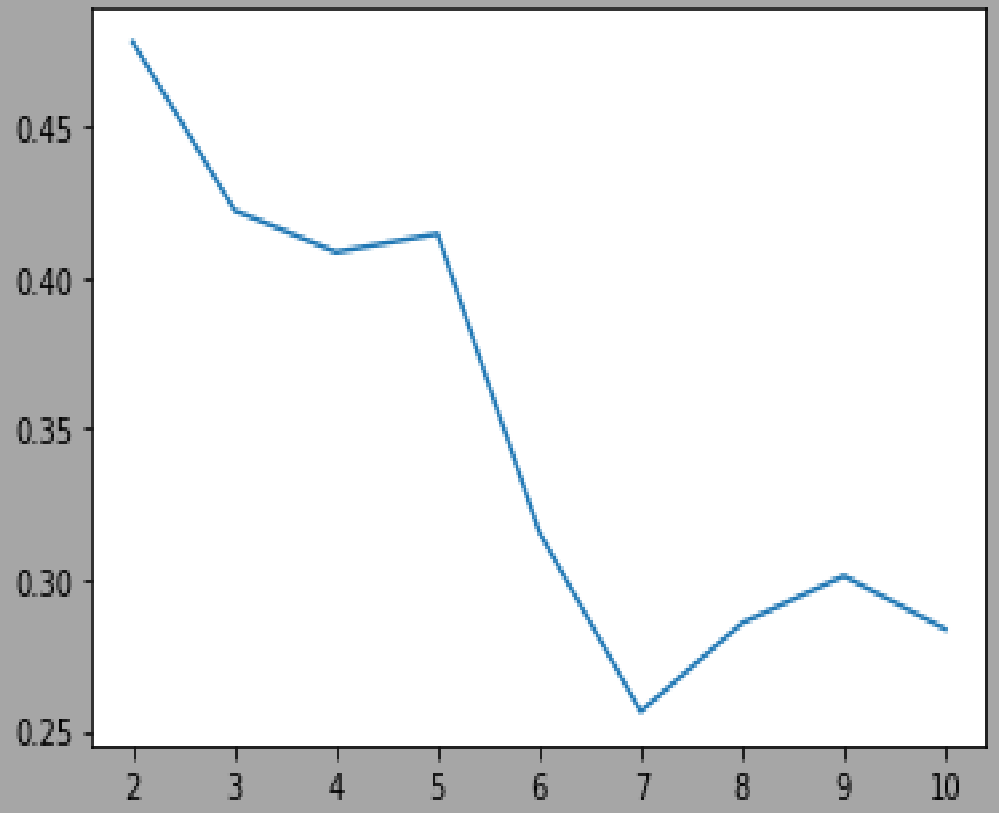
Imports & Exports

Life Expectancy & Total Fertility

The Silhouette Method

As per the business requirement

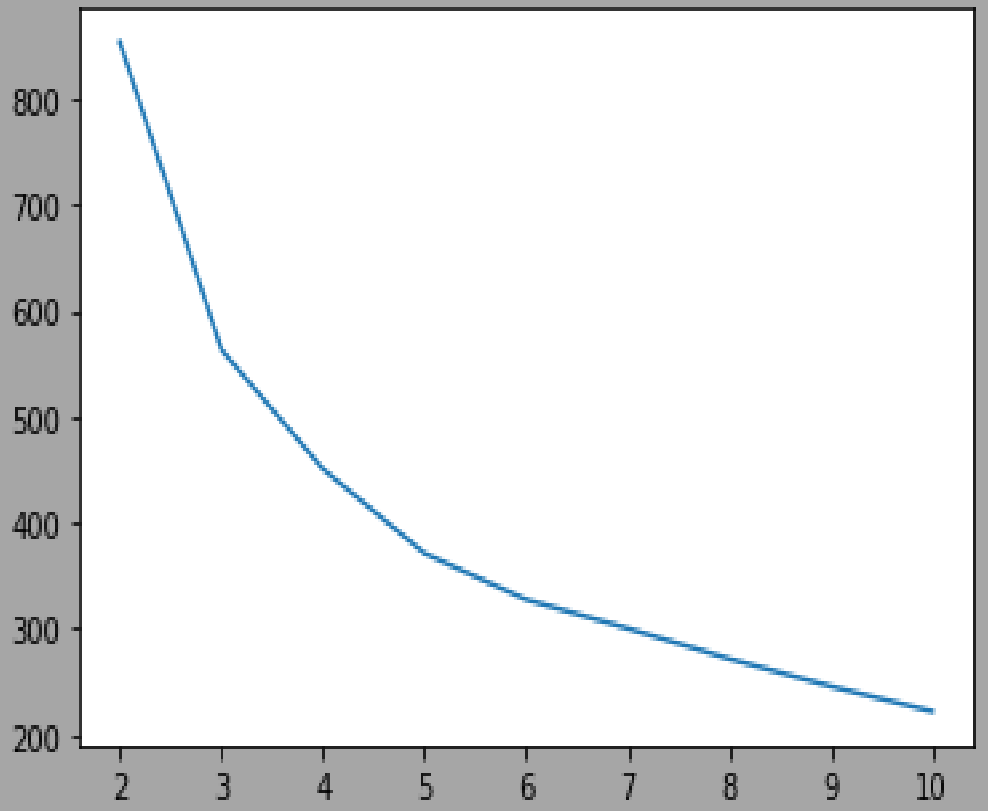
The k value is taken as 3



The Elbow Method-SSD

As per the business requirement

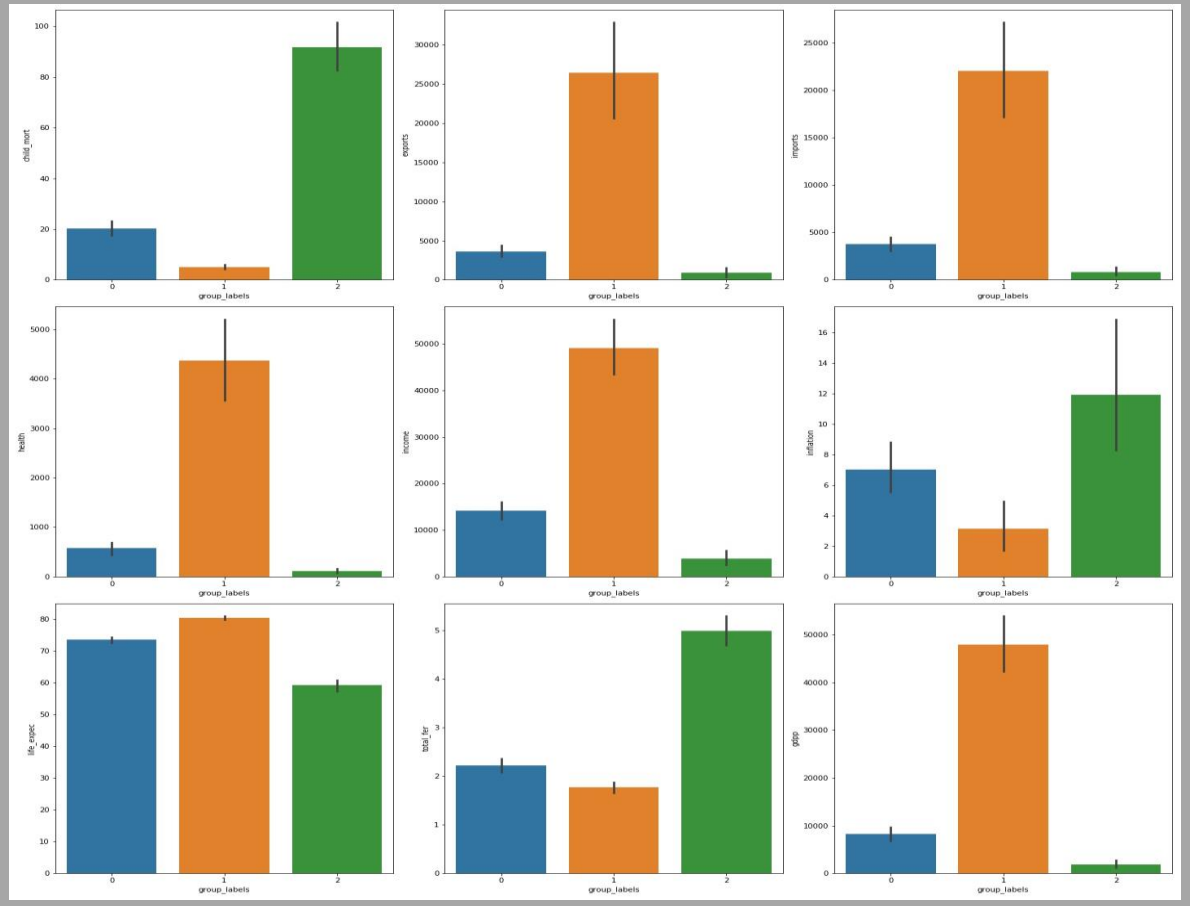
The k value is taken as 3



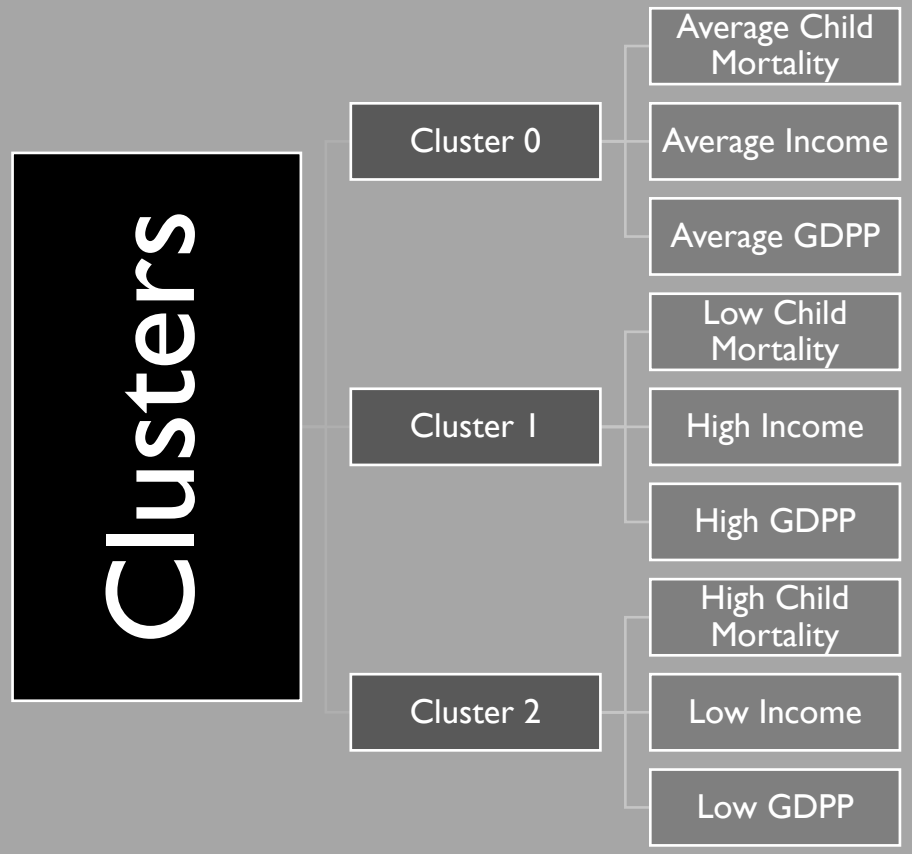
K Means Clustering- All Variables

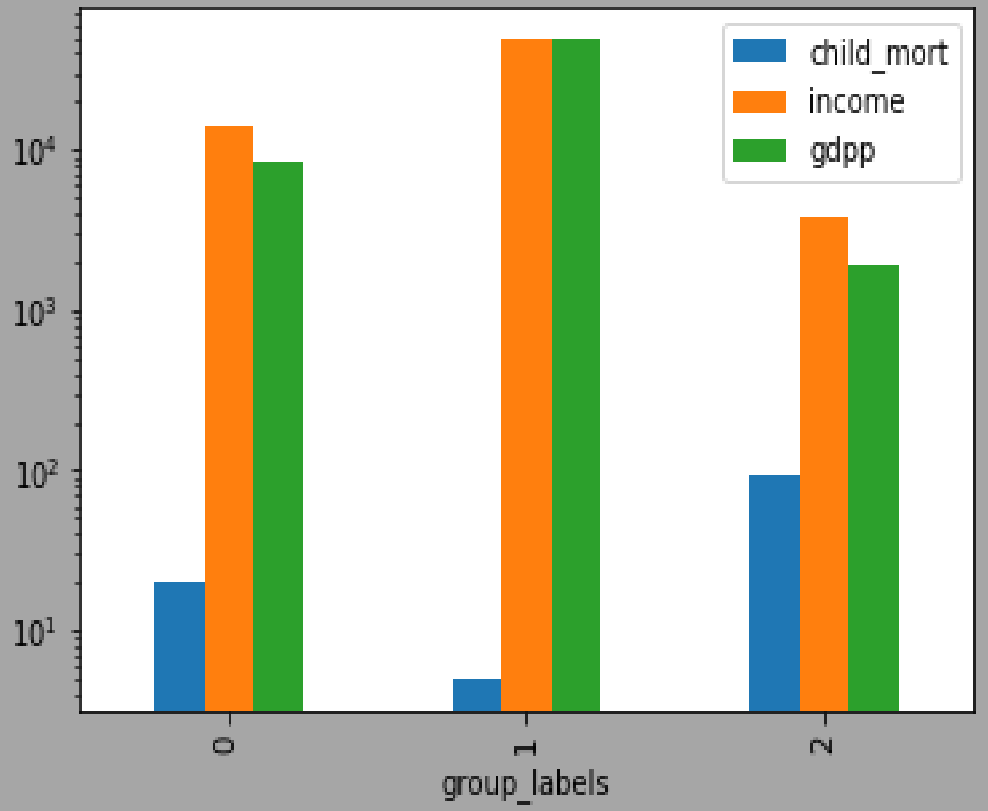
Based on above plots we recommend countries belonging to Cluster no 2 because of following reasons:

- low value of gdpp & income
- high value of child_mort and total_fer
- Life_expec is low in these contries as compared to other countries
- Also Inflation is high in these contries as compared to other countries
- Exports,imports & health is low in these countries.



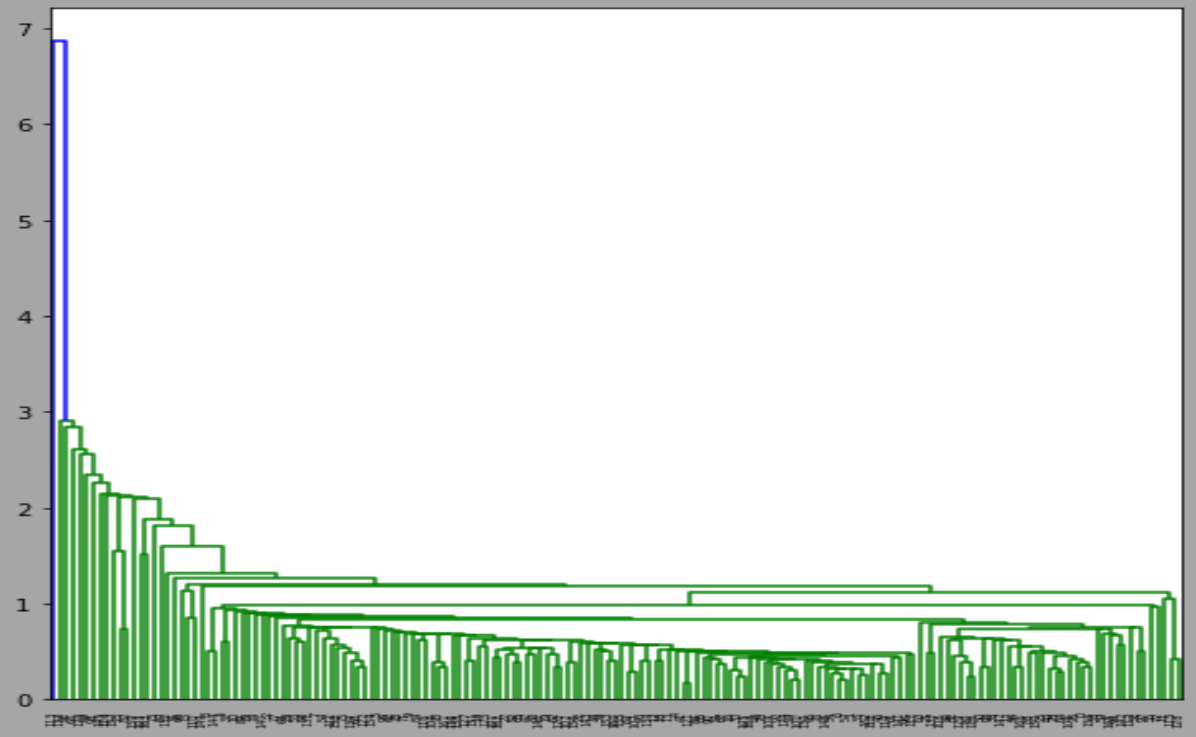
Cluster Formation





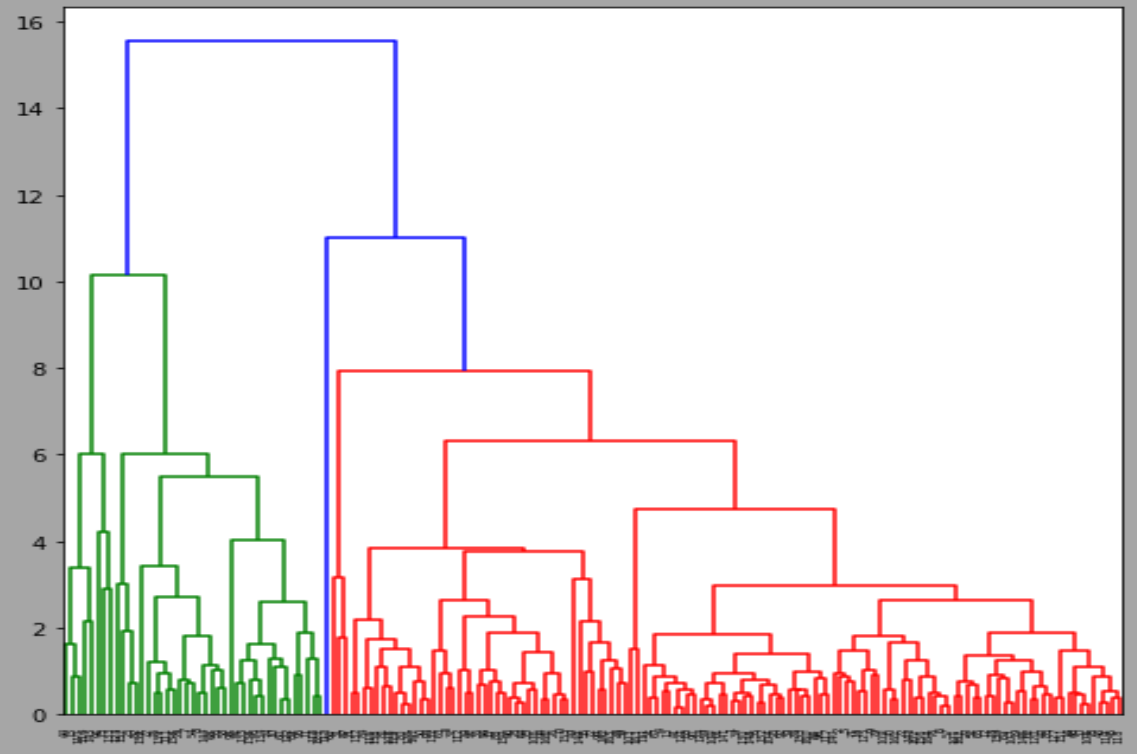
Hierarchical Clustering

SINGLE LINKAGE



- Difficult to get insights from Single Linkage

- COMPLETE LINKAGE



- Go by Complete Linkage, and consider the value of n clusters as 3

Different Clusters formed by K Means

Select Top 5 Countries

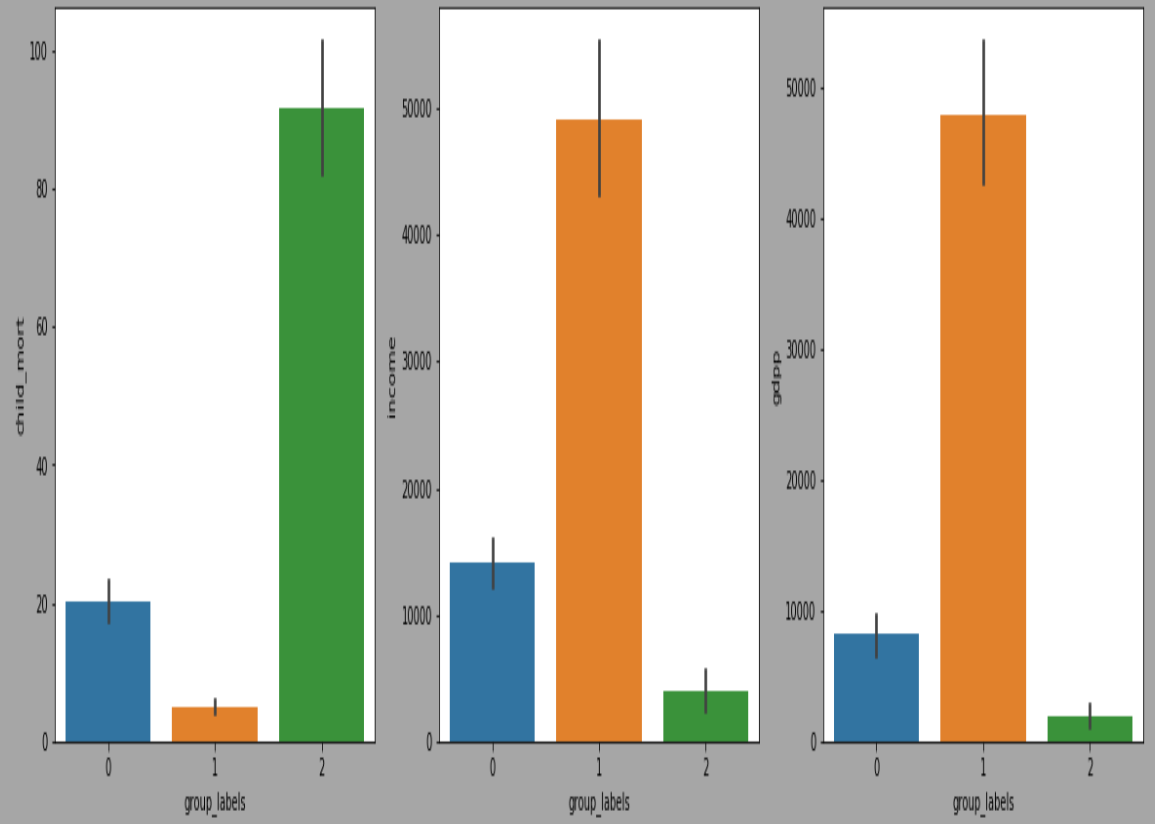
Socio-Economic factors

Health Factors

GDPP

Income

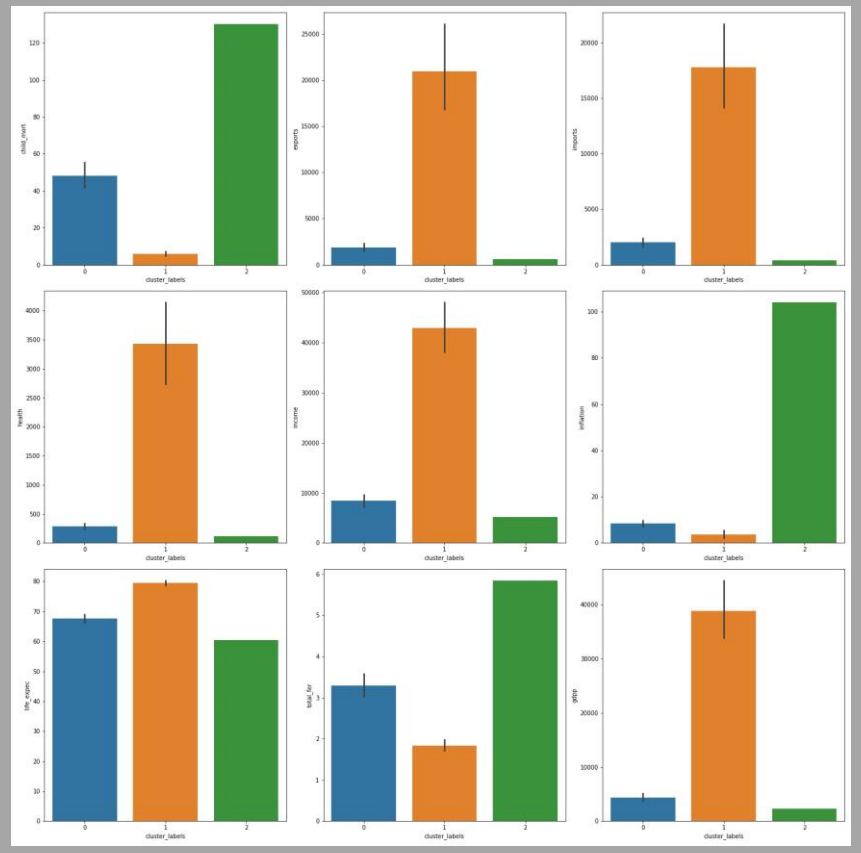
Child Mortality



Hierarchical Clustering- All Variables

Based on above plots we recommend countries belonging to Cluster no 0 because of following reasons:

- low value of gdpp & income
- high value of child_mort total_fer
- Life_expec is low in these contries as compared to other countries
- Also Inflation is high in these contries as compared to other countries
- Exports,imports & health is low in these countries.



Visualizing Clusters for Socio-Economic Factors – Hierarchical Clustering

From all the three clusters, we select the top 5 countries which are common across all clusters, which match the criteria

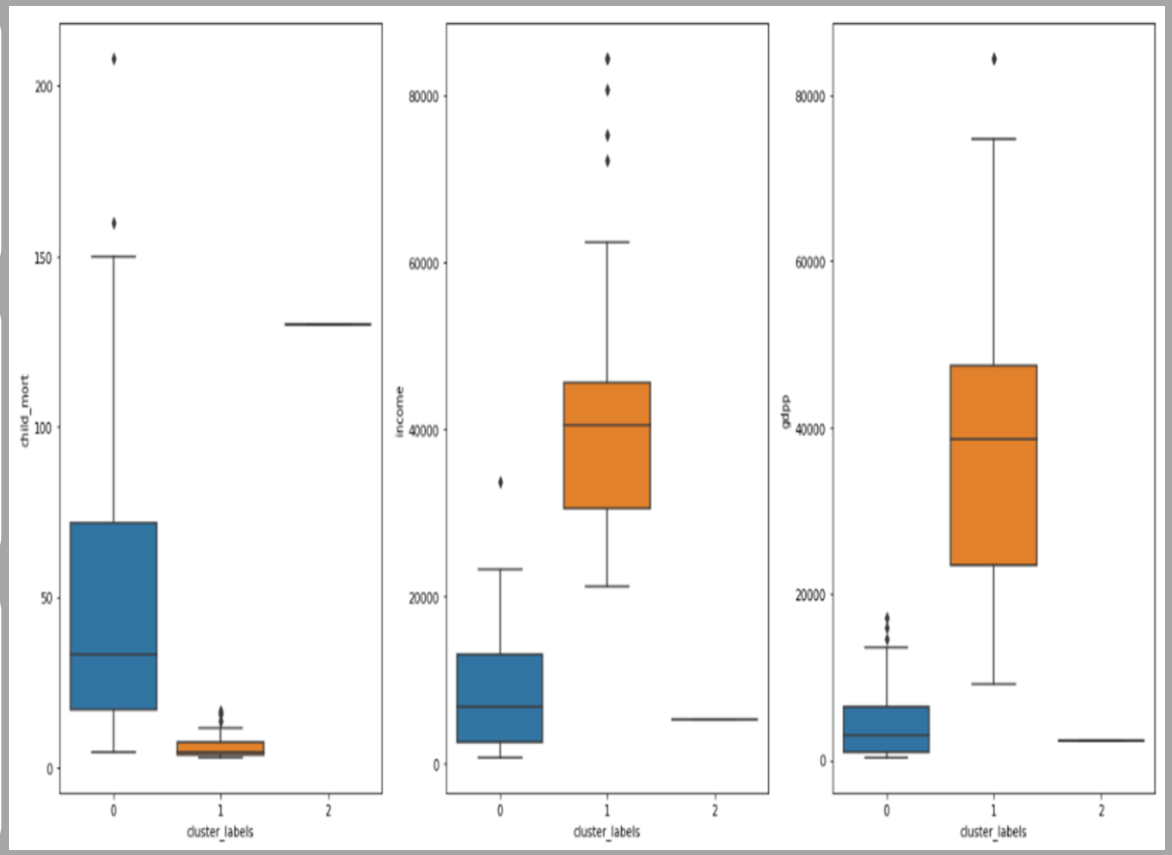
Socio-Economic factors

Health Factors

GDPP

Income

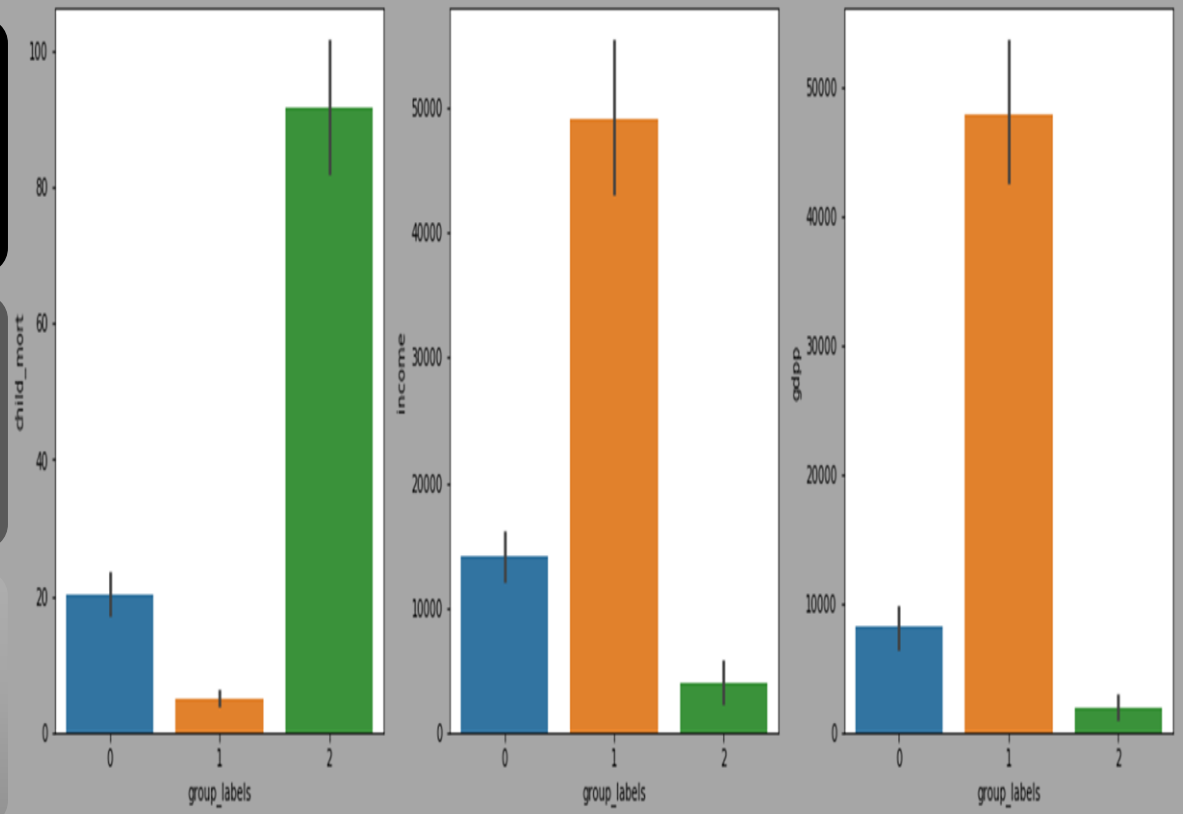
Child Mortality



Visualizing Clusters for Socio-Economic Factors – Hierarchical Clustering

Select Top 5 Countries

Socio-Economic factors:	Health Factors:
GDPP	Income
	Child Mortality



Recommended- Countries Dire need of Aid

K Means Clustering	
S No	Country
66	Haiti
132	Sierra Leone
32	Chad
31	Central African Republic
97	Mali

Hierarchical Clustering	
S No	Country
66	Haiti
132	Sierra Leone
32	Chad
31	Central African Republic
97	Mali