

**Problem Statement:**

HELP International 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. So, CEO has to make decision to choose the countries that are in the direst need of aid.

**About organization:**

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.

**Objective:**

To categorise the countries using socio-economic and health factors that determine the overall development of the country.

**Data:**

We have data of 167 countries. Which includes child mortality rate, life expectancy, health, import, export, income, total fertility rate and GDP per capita.

**Data Pre-processing:***Missing values:*

There were no missing values.

### *EDA:*

- There is a very high negative trend between child mortality rate and life expectancy. A country where life expectancy is low, child mortality rate must be pretty high.
- There is high positive trend between child mortality rate and total fertility rate. there are some countries where female give birth to more number of children and expecting some of them will survive. due to high mortality rate.
- Income having high positive correlation with export, import, health. countries with more income will spend more on health, imports and sending more exports.
- GDP per capita having high positive correlation with export, health, import, income. countries with high gdpp must be developed countries with better socio-economic conditions.

### *Outlier treatment:*

There were some very rich countries whose income, gdpp, health, life expectancy was very high and there were some very poor countries also whose gdpp, income, life expectancy was very low. Both kind of countries are valuable for the analysis. So we can't remove them, I capped them at 5<sup>th</sup> & 95<sup>th</sup> percentile.

### *Scaling data:*

We use standard scalar to scale the data.

### **Cluster tendency:**

Performed Hopkins test to check the data is having clustering tendency or not. Hopkins score was 0.80 which is grater than 0.5, concludes data have pretty high clustering tendency.

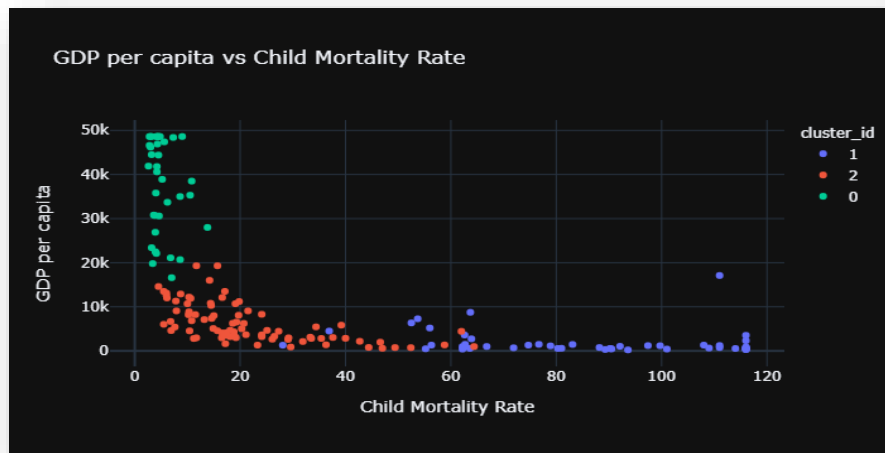
### **K-means Clustering:**

#### ***Choosing value of k:***

From elbow curve method and silhouette analysis decided to form 3 clusters.

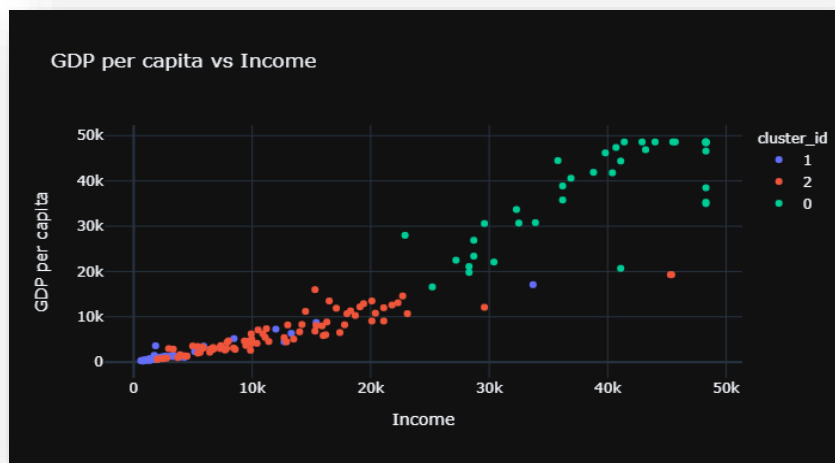
## Cluster Profiling:

### GDP per capita vs Child Mortality Rate:



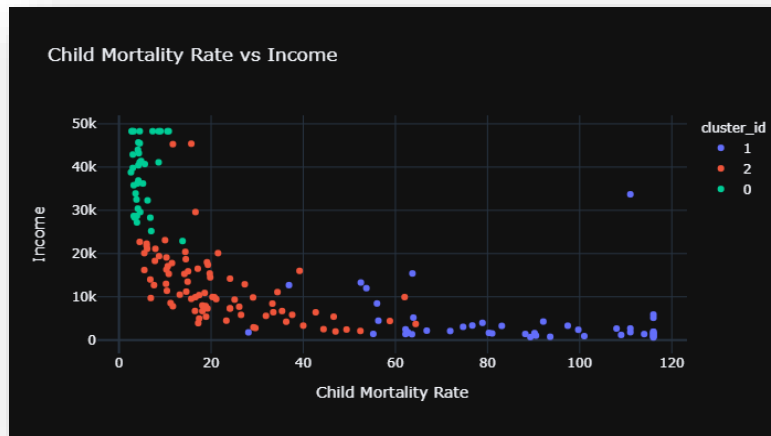
- cluster 2: countries with medium level or average level of GDP per capita and medium or average level of Child mortality rate.
- cluster 0: countries with very high GDP per capita and very low Child Mortality Rate.
- cluster 1: countries with very low GDP per capita and very high Child Mortality Rate.

### GDP per capita vs Income:



- cluster 1 and cluster 2 are overlapping. but cluster 2 looks better in comparison with cluster 1 in both GDP per capita and Income terms.
- cluster 0 is clearly looks most better with High GDP per capita and High Income in comparison with cluster 1 and cluster 2.

## Child Mortality Rate vs Income:



- cluster 2: countries with medium level or average level of Income and medium or average level of Child mortality rate.
- cluster 0: countries with very high Income and very low Child Mortality Rate.
- cluster 1: countries with very low Income and very high Child Mortality Rate.

Based on analysis:

- Cluster 0: Developed countries
- Cluster 1: Underdeveloped Countries
- Cluster 2: Developing countries

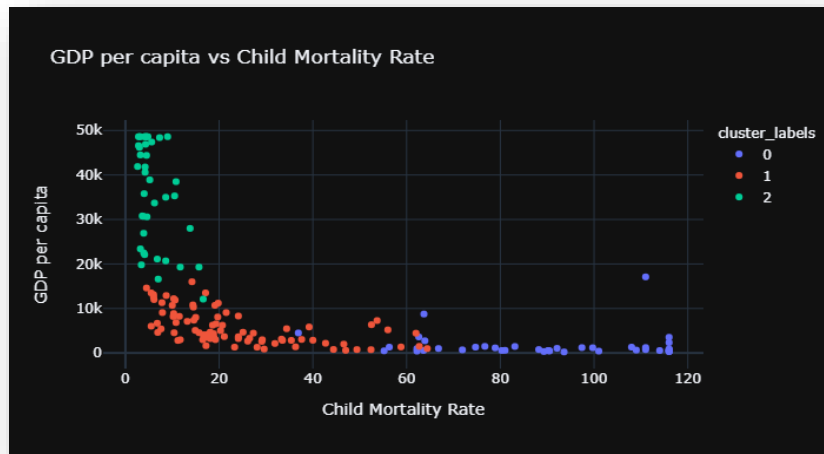
As per the K-means algorithm, countries which are in direst need are as follow:

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

## Hierarchical Clustering:

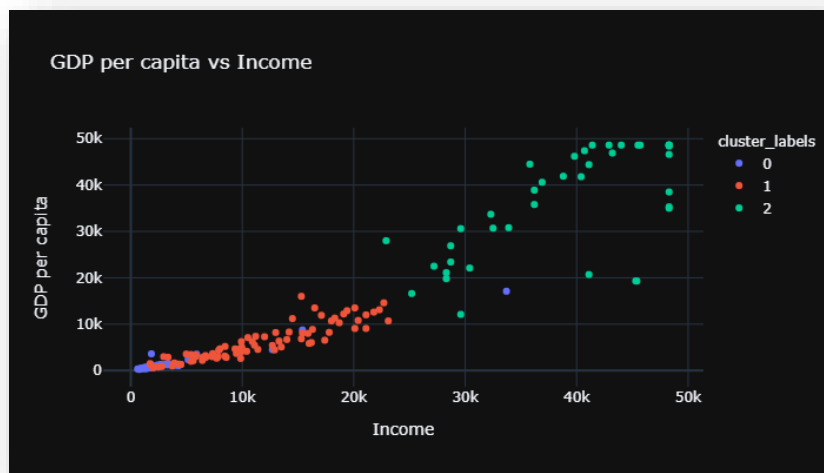
Used single linkage and complete linkage methods, found out complete linkage method clearly works better. And cut the tree at 3.

## GDP per capita vs Child Mortality Rate:



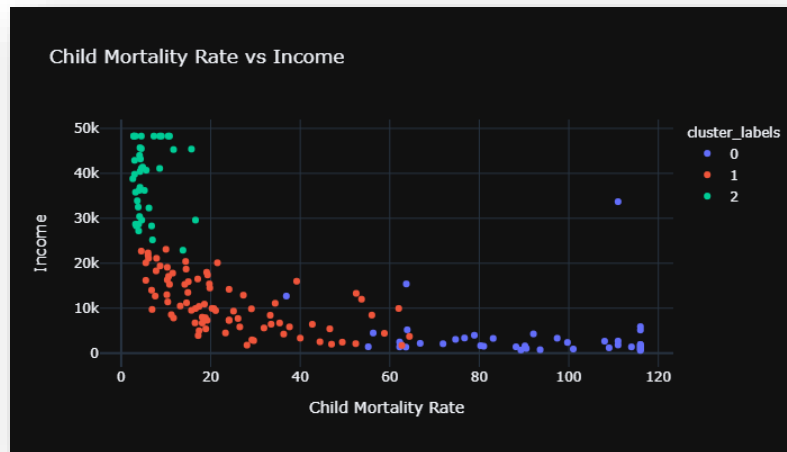
- cluster 1: countries with medium level or average level of GDP per capita and medium or average level of Child mortality rate.
- cluster 2: countries with very high GDP per capita and very low Child Mortality Rate.
- cluster 0: countries with very low GDP per capita and very high Child Mortality Rate.

## GDP per capita vs Income:



- cluster 1 and cluster 0 are overlapping. but cluster 1 looks better in comparison with cluster 0 in both GDP per capita and Income terms.
- cluster 2 is clearly looks most better with High GDP per capita and High Income in comparison with cluster 0 and cluster 1.

## Child Mortality Rate vs Income:



- cluster 1: countries with medium level or average level of Income and medium or average level of Child mortality rate.
- cluster 2: countries with very high Income and very low Child Mortality Rate.
- cluster 0: countries with very low Income and very high Child Mortality Rate.

Based on analysis:

- Cluster 2: Developed countries
- Cluster 0: Underdeveloped Countries
- Cluster 1: Developing countries

As per the Hierarchical Clustering algorithm, countries which are in direst need are as follow :

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

## Country Identification:

we choose top 10 countries by using both k-means algorithm and Hierarchical Clustering algorithm and we got exact same result.

here first identify the cluster of underdeveloped countries, then based on GDP per capita, child mortality rate and net income per person we identify 10 countries.

- lowest GDP per capita among underdeveloped countries means, these countries are really poor.
- high child mortality rate means, healthcare structure of these countries is in very bad state, maybe here child under age 5, where its immunity system is not developed yet did not have access to different vaccination.
- low net income per person means, person do not have enough income to spend for its own betterment and since the GDP per capita is also low that's why government also not able to spend on its people.
- we can also observe that selected countries in the final list, all of them are African countries.

Based on these criteria we choose following 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**