



Clustering Assignment

Presentation By

Mahesh Babu R

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.

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.

Goals and Objectives



The requisite is:

- To categorize the countries using some socio-economic and health factors that determine the overall development of the country.
- To suggest the countries which the CEO needs to focus on the most.

Provided Data And Description

| Column Name | Description |
|-------------|--|
| country | Name of the country |
| child_mort | Death of children under 5 years of age per 1000 live births |
| exports | Exports of goods and services per capita. Given as %age of the GDP per capita |
| health | Total health spending per capita. Given as %age of GDP per capita |
| imports | Imports of goods and services per capita. Given as %age of the GDP per capita |
| Income | Net income per person |
| Inflation | The measurement of the annual growth rate of the Total GDP |
| life_expec | The average number of years a new born child would live if the current mortality patterns are to remain the same |
| total_fer | The number of children that would be born to each woman if the current age-fertility rates remain the same. |
| gdpp | The GDP per capita. Calculated as the Total GDP divided by the total population. |

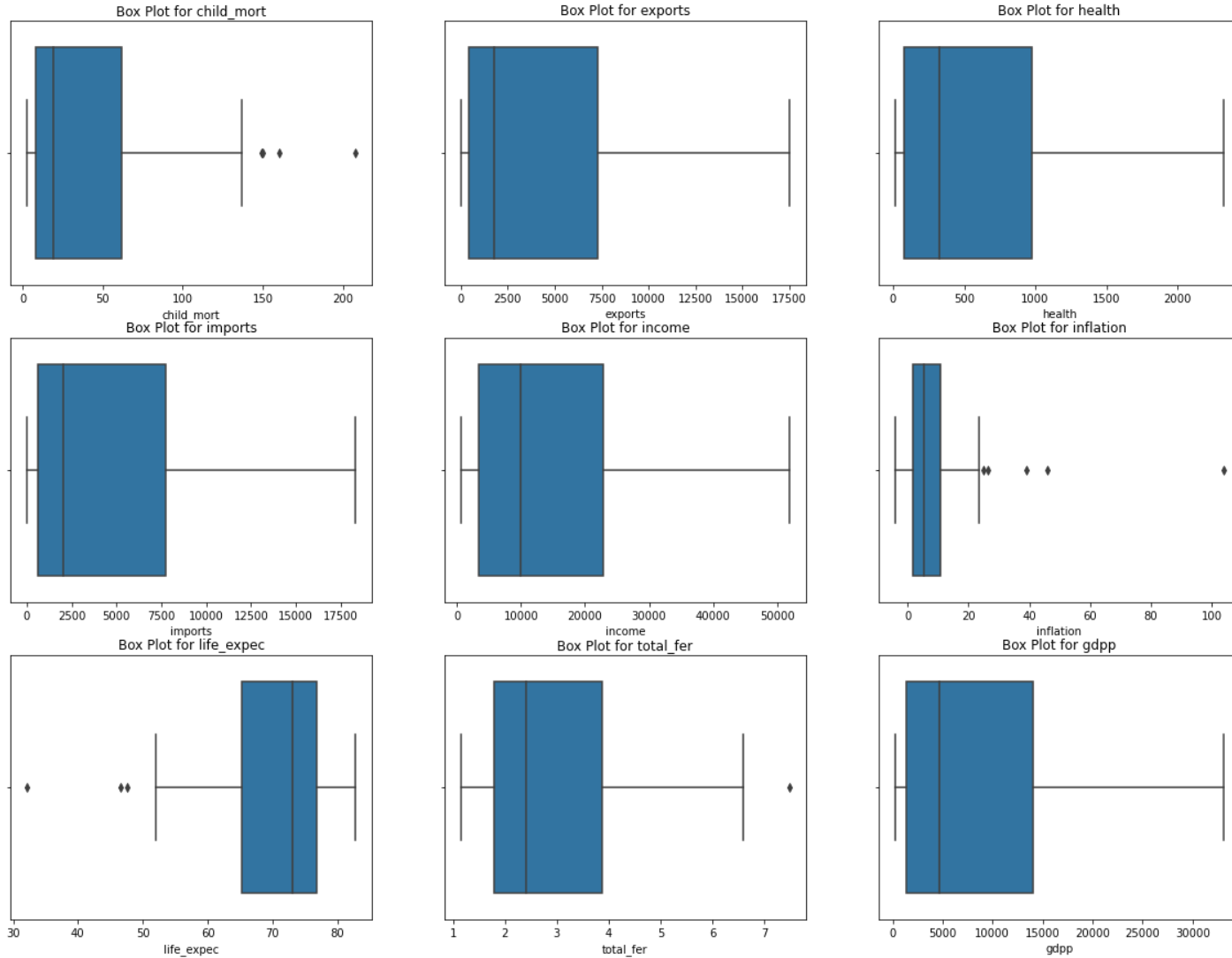
Project Analysis

ANALYSIS



- The Dataset Contains of 167 countries along with Descriptions.
- It was found that there were no null values.
- There were also no duplicate values for country .
- There were a few outliers and they were treated .
- The data was standardized for K-Means and Hierarchical Clustering Analysis

Project Analysis



- Outlier Treatment Provided According to Variable Preference and fig above after the Outliers Capped to Extreme limits

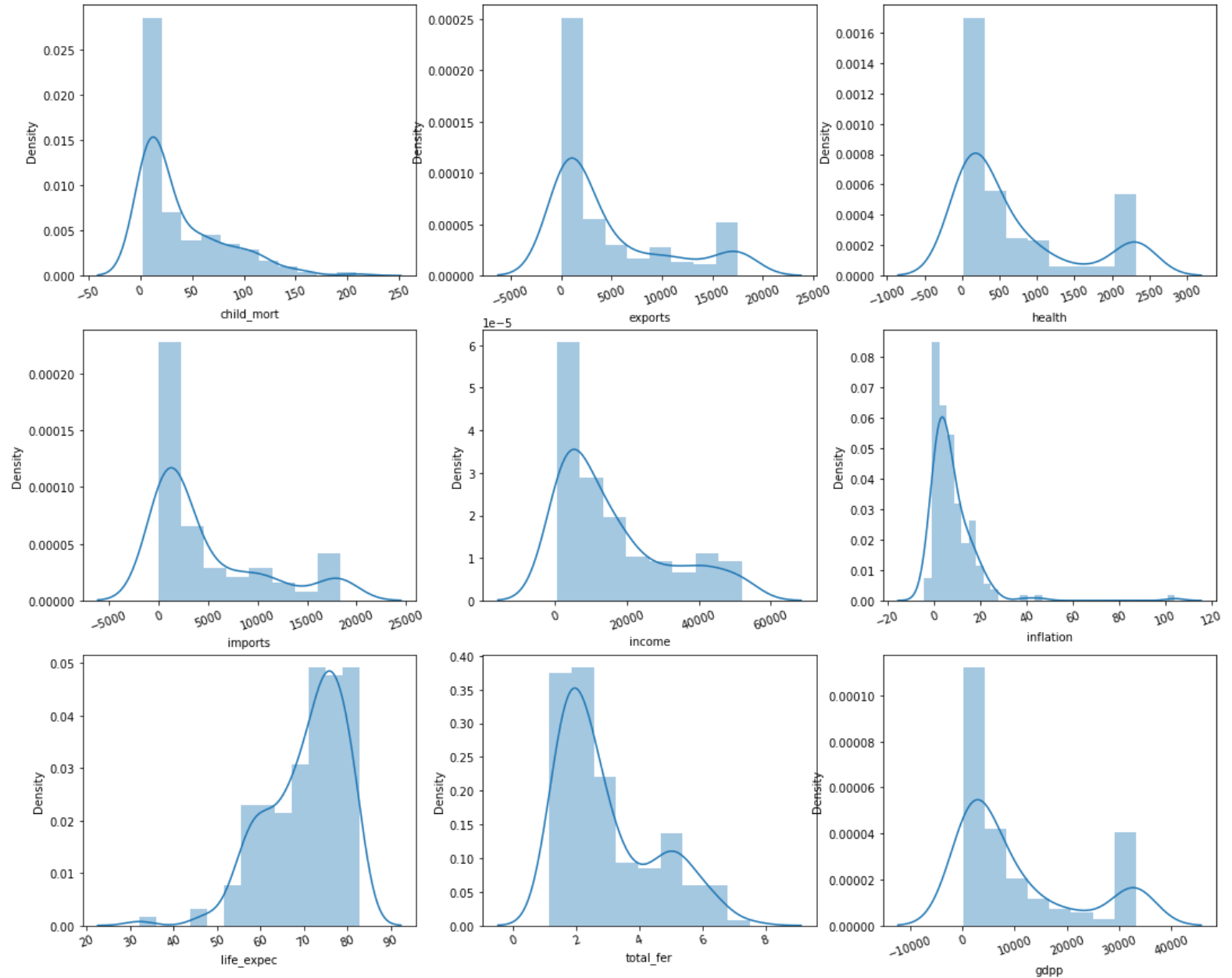
Project Analysis

- The Distribution of Features clearly says there are some groups internally in each feature.



Hopkins Statistics help us to decide whether the clusters can be formed easily or not. If the values are:

- 0.01 - 0.3 : Low chance of clustering
- around 0.5 : Random
- 0.7 - 0.99 : High chance of clustering
- By Performing Hopkins Test it came at 0.87 and High Chances of clustering.**

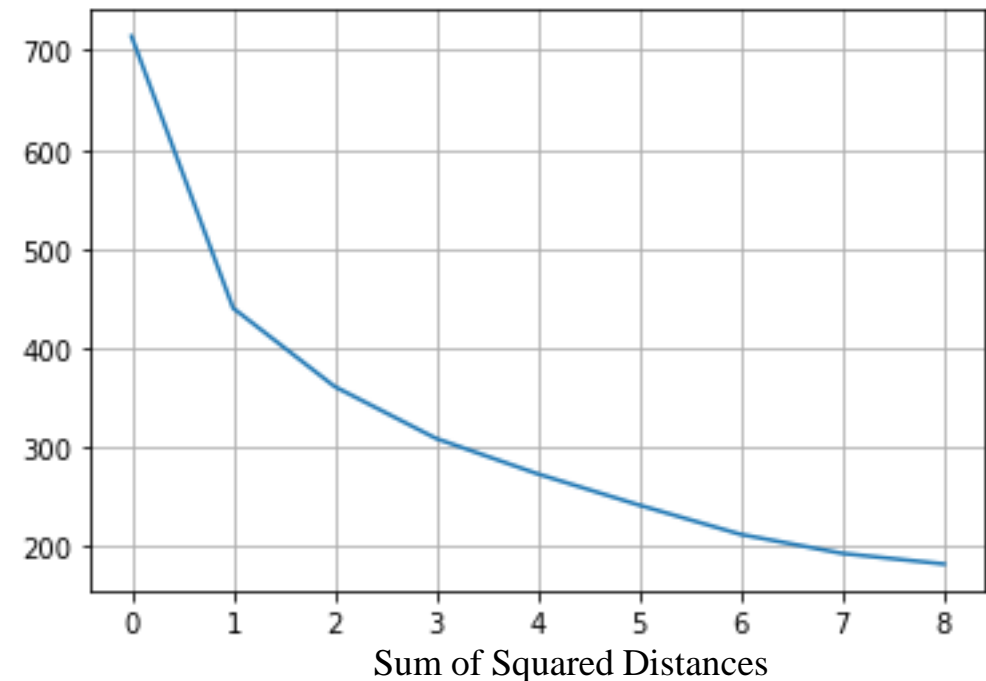
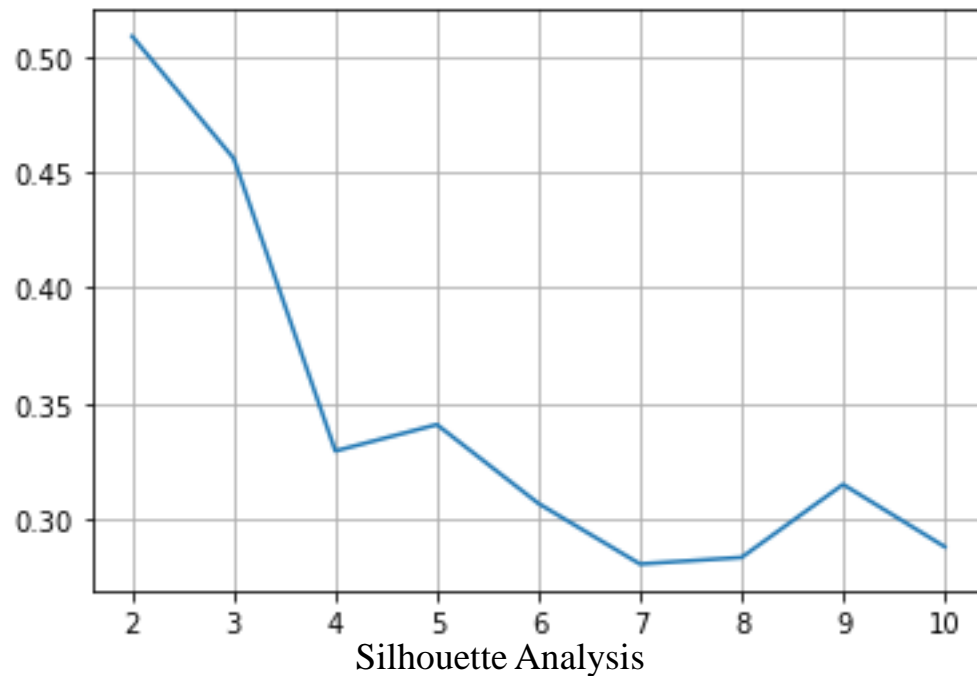


Project Analysis

IMPLEMENT



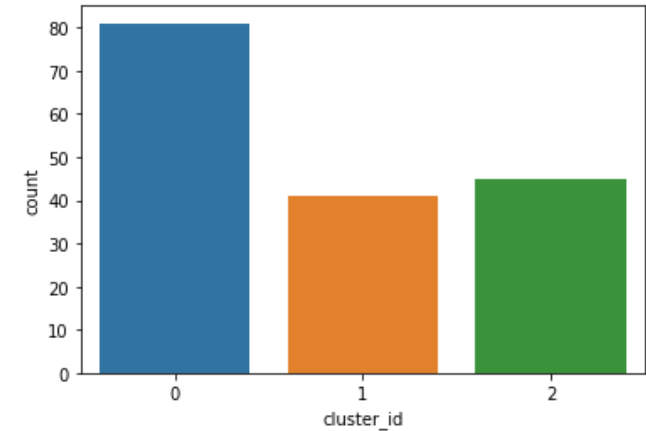
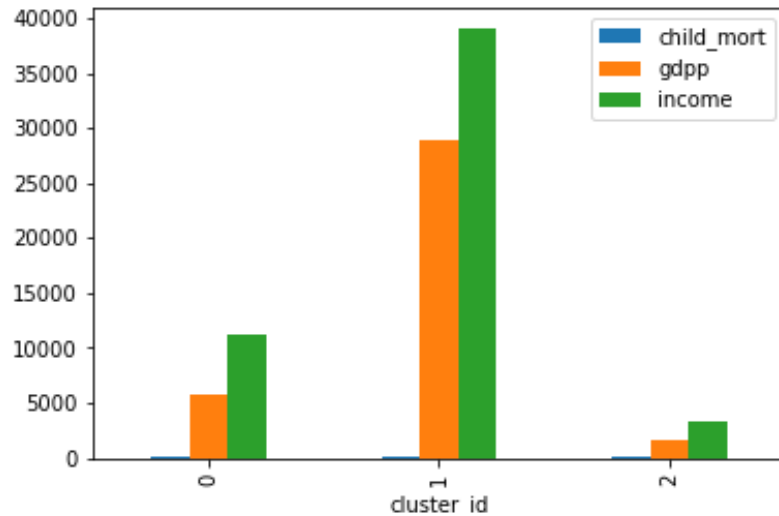
- To get a feasible no of clusters we performed silhouette Analysis and Sum of Squared distances
- There are some high chances of maximum clusters but if we go for more no.of clusters we are falling in below 5 records in some clusters that don't give proper evaluation so we concluded with 3,2 clusters respectively for K-means and Hierarchical Clustering using silhouette and sum of squared distances analysis.



K-Means

- It shows the three groups are clearly segregated according to their fall in locations.
- 81,41 and 45 are the respective no.of counts in clusters 0 , 1 and 2.

| | child_mort | gdpp | income |
|------------|------------|--------------|--------------|
| cluster_id | | | |
| 0 | 23.520988 | 5770.395062 | 11250.493827 |
| 1 | 5.651220 | 28969.512195 | 38988.780488 |
| 2 | 94.537778 | 1633.600000 | 3312.733333 |



- If we compare the mean values with respect to child_mort, gdpp and income we can conclude the cluster 2 is having high child mortality low gdpp and income.

Project Analysis

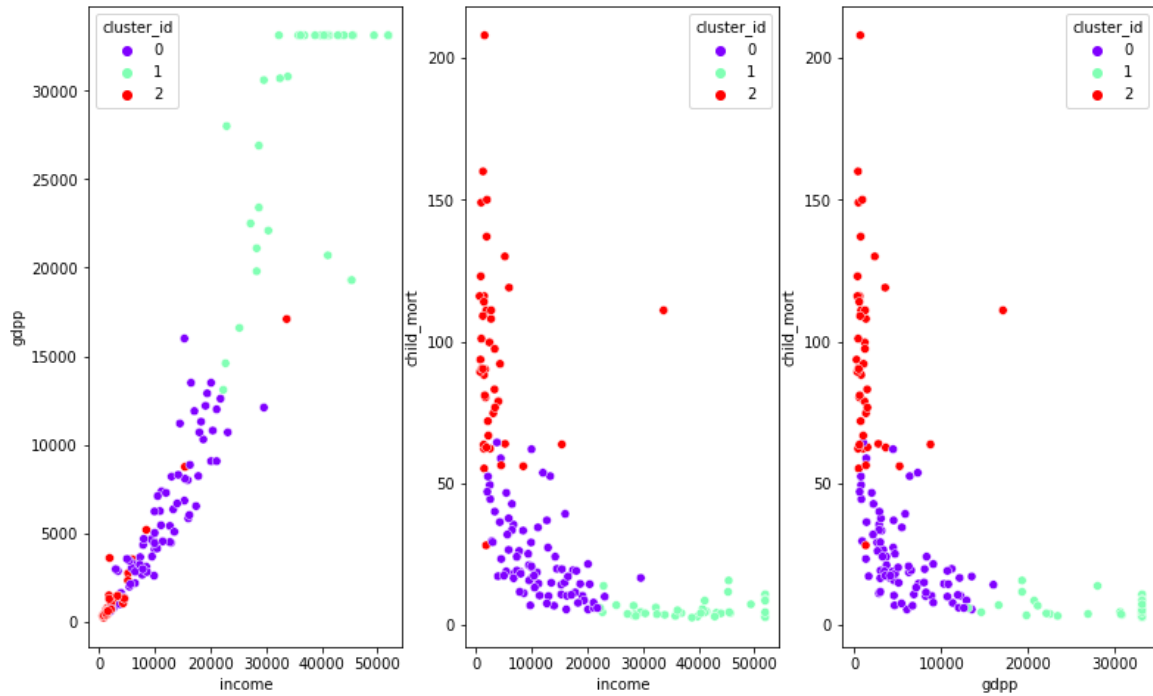


Fig 1

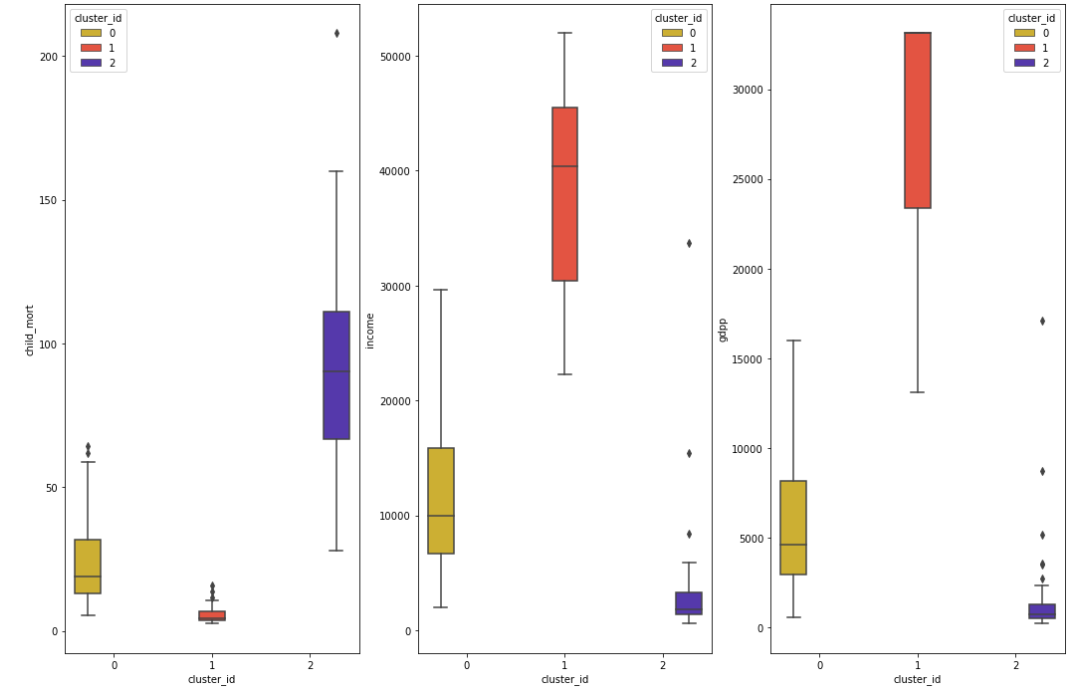


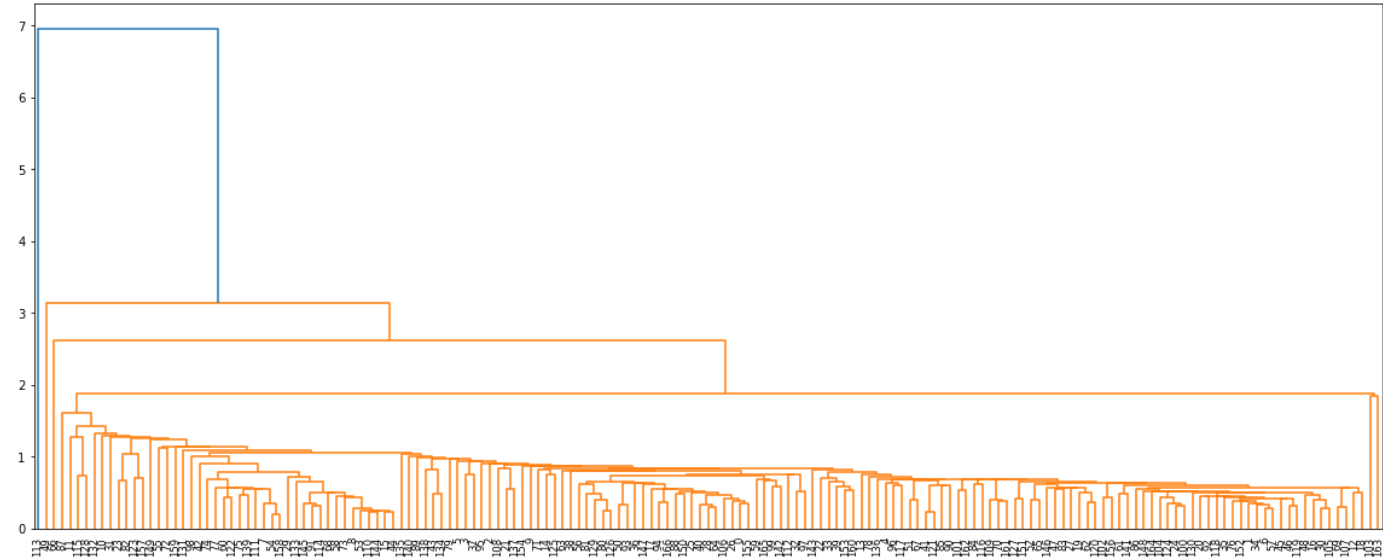
Fig 2

- By looking at Fig 1 and 2 gdpp and income low in cluster 2 , same time high child mortality is present is those countries so we mainly focused on cluster 2
- Sorted the countries from cluster 2 to get top 10 countries which are Aid needed ones with respect to Child_mort in decreasing and gdpp, income in increasing order.
- Stored to finalize the countries ones after getting the list from hierarchical clustering too.

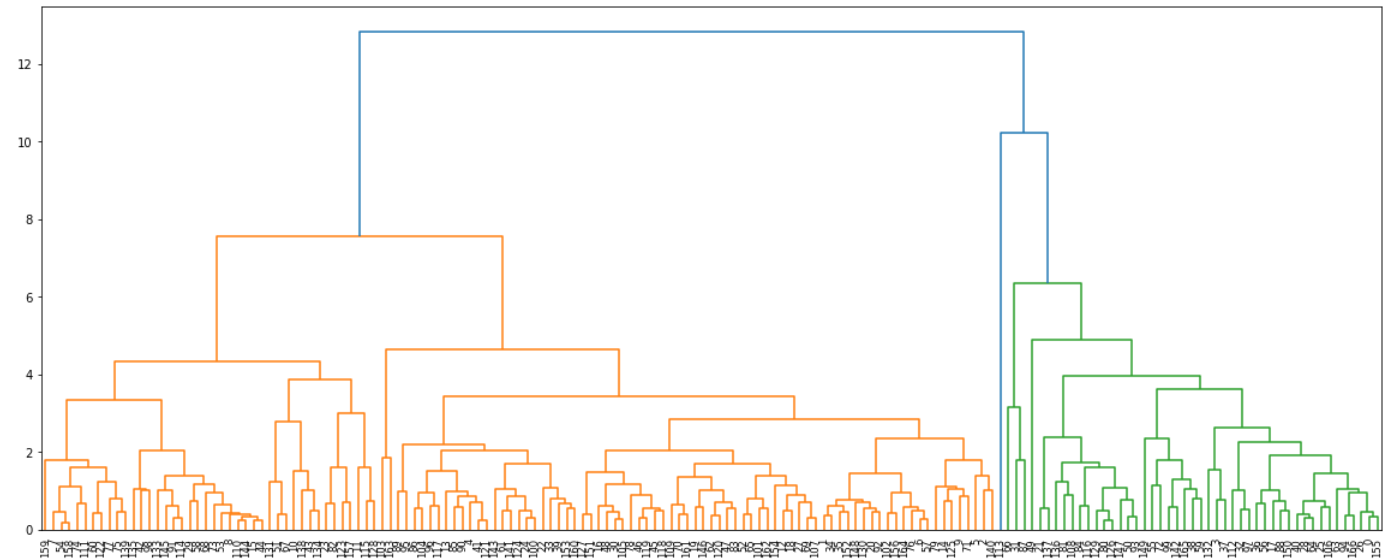


Hierarchical clustering

In hierarchical clustering the complete linkage is giving some good clustering compare with single linkage so we conclude with complete and 2 clusters



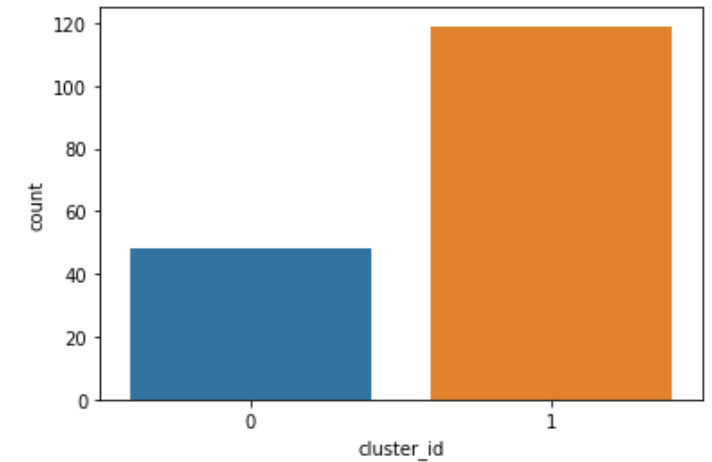
Single linkage



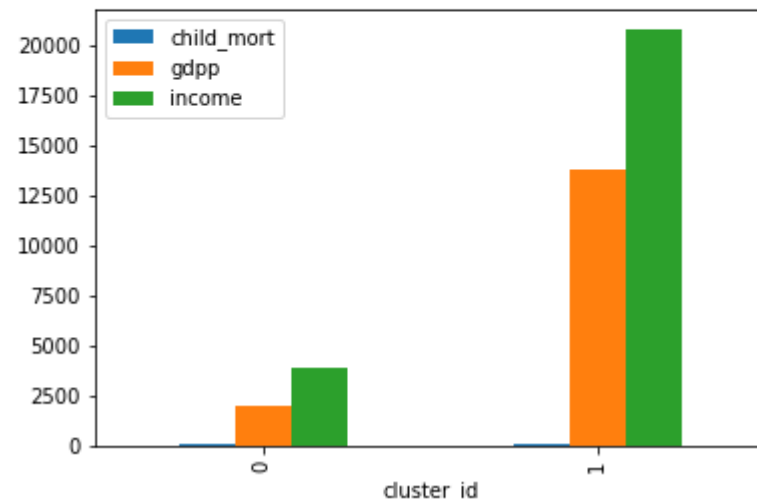
Complete linkage

Project Analysis

- It shows the two groups are clearly segregated according to their fall in locations
- 48 and 119 are the respective no.of counts in clusters 0 and 1



| | child_mort | gdpp | income |
|------------|------------|--------------|--------------|
| cluster_id | | | |
| 0 | 91.610417 | 1909.208333 | 3897.354167 |
| 1 | 16.754622 | 13756.487395 | 20771.680672 |



- If we compare the mean values with respect to child_mort, gdpp and income we can conclude the cluster 0 is having high child_mortality low gdpp and income.

Project Analysis

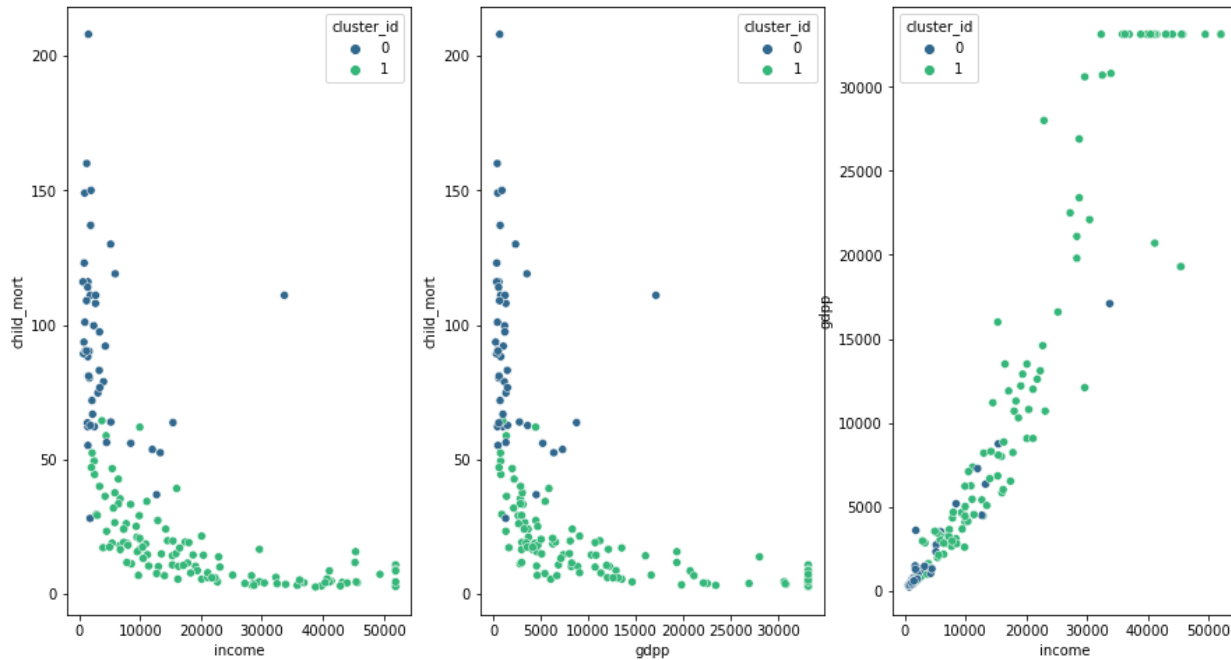


Fig 1

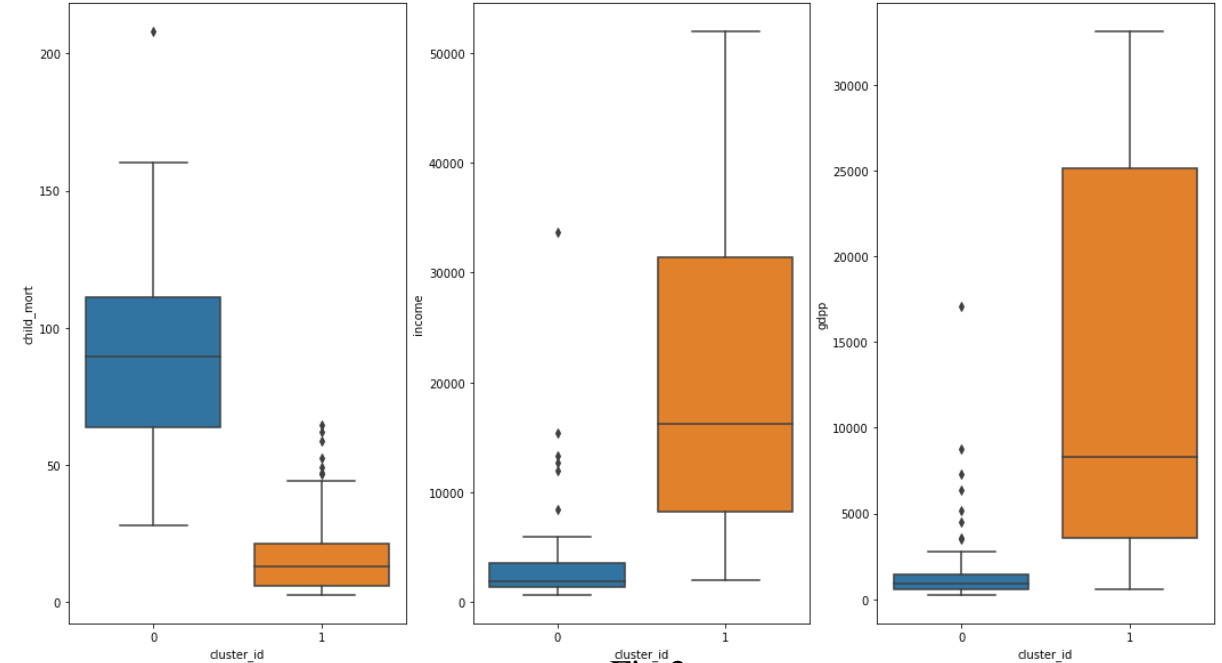


Fig 2

- By looking at Fig 1 and 2 gdp and income low in cluster 0 , same time high child mortality is present is those countries so we mainly focused on cluster 0
- Sorted the countries from cluster 0 to get top 10 countries which are Aid needed ones with respect to Child_mort in decreasing and gdp, income in increasing order.
- The countries from both the analysis are same hence shown below.
- Top 10 countries are **1.Haiti , 2.Sierra Leone , 3.Chad, 4.Central African Republic, 5.Mali, 6.Nigeria, 7.Niger, 8.Angola, 9.“Congo, Dem. Rep”., 10.Burkina Faso.**



The countries that require help the most are listed below(sorted): TOP 5 COUNTRIES

1.Haiti ,2.Sierra Leone , 3.Chad, 4.Central African Republic, 5.Mali, Nigeria, Niger, Angola, “Congo, Dem. Rep”., Burkina Faso, Guinea-Bissau, Benin, Cote d'Ivoire, Equatorial Guinea , Guinea, Cameroon, Mozambique, Lesotho, Mauritania, Burundi, Pakistan, Malawi, Togo, Afghanistan, Liberia, Comoros, Zambia, Uganda, Gambia, Lao, Sudan, Ghana, Tanzania, Senegal, “Congo, Rep.”, Gabon, Rwanda, Kiribati, Timor-Leste, Madagascar, Kenya, Yemen, Namibia, Eritrea, Solomon Islands

These countries have

- very low rate of net income per person, GDP per capita, average number of years a new born child would live, total health spending and imports of goods and services.
- very high rate of measurement of the annual growth rate, number of children that would be born and child mortality rate.
- It is clear that these countries require very quick aid in terms of money, education and services.



Thank You