

Clustering Assignment

Submitted by:

Sawal Malhotra

Introduction

- HELP International is an International humanitarian NGO that aims to fight the poverty in the country by providing the people of backward countries with basic amenities and relief during disaster and natural calamities. They have received a financial fund of \$ 10 million and we need to identify and present the List of Top countries that are socio-economic and financially weak.

Problem Statement

- HELP International have raised financial fund of \$ 10 million and we need to identify and present the List of Top countries that are socio-economic and financially weak to the CEO of the company.

Analysis approach

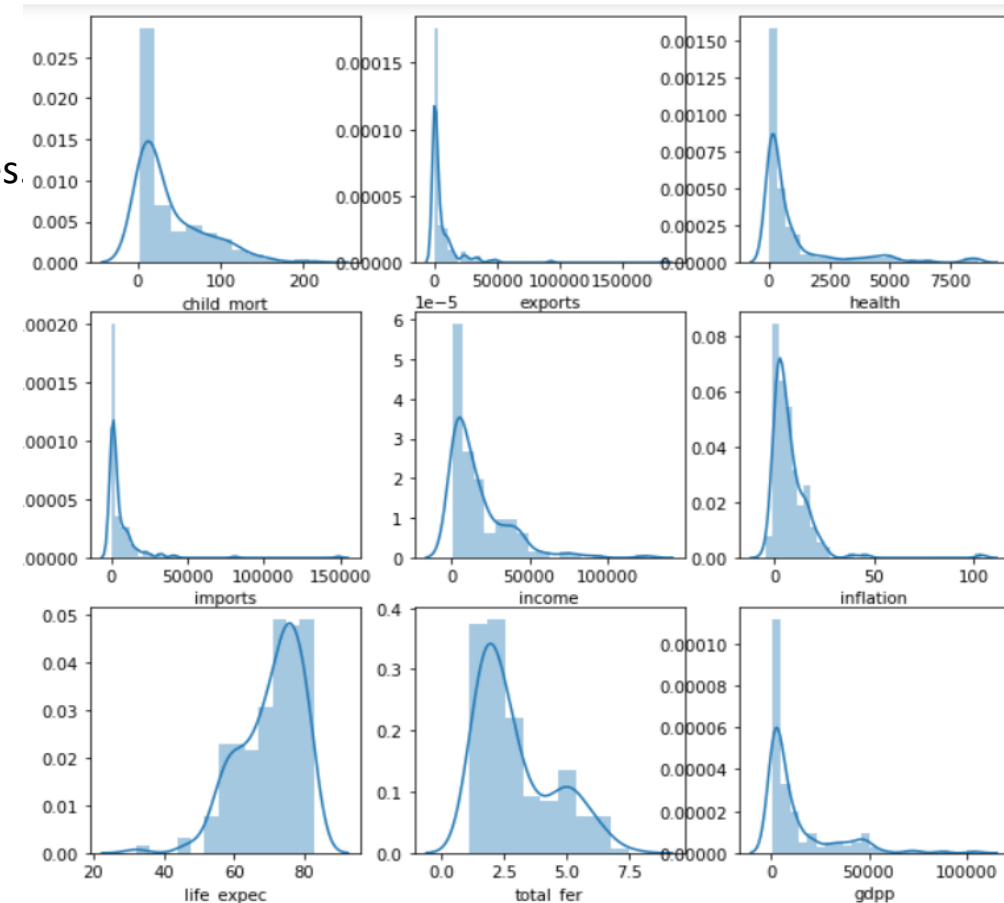
As part of analysis, we would be following below approach to identify hidden pattern and trends followed by List of financial weak countries as per our Business objective.

- Data understanding
- Clustering (including Outlier treatment and Hopkins Check)
- K-means algorithm
- Hierarchical clustering
- Country Identification as final result

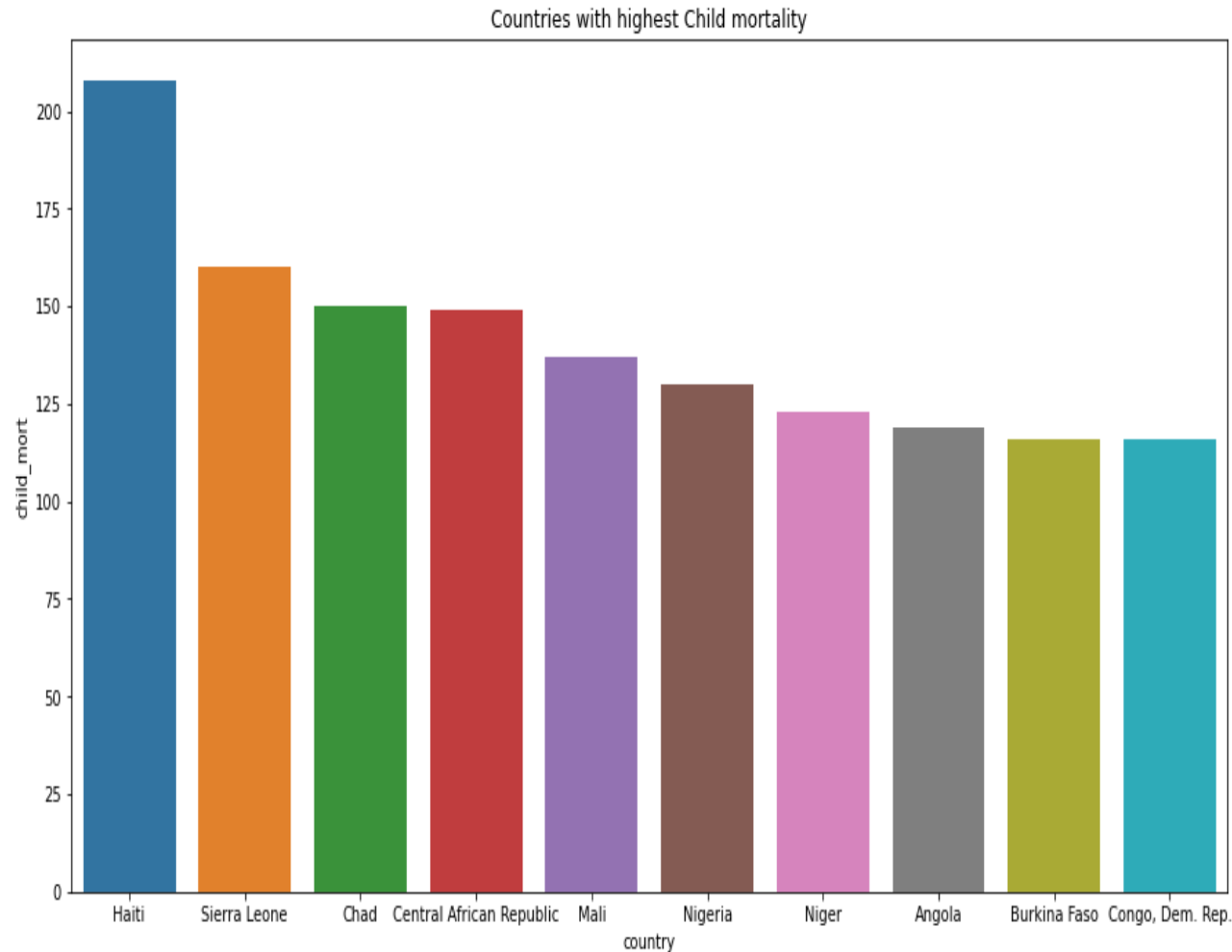
EDA(Exploratory Data Analysis)

- **Univariate analysis:** In this process, a single variable is analysed and visualized from given Country data set using Distribution Plot.

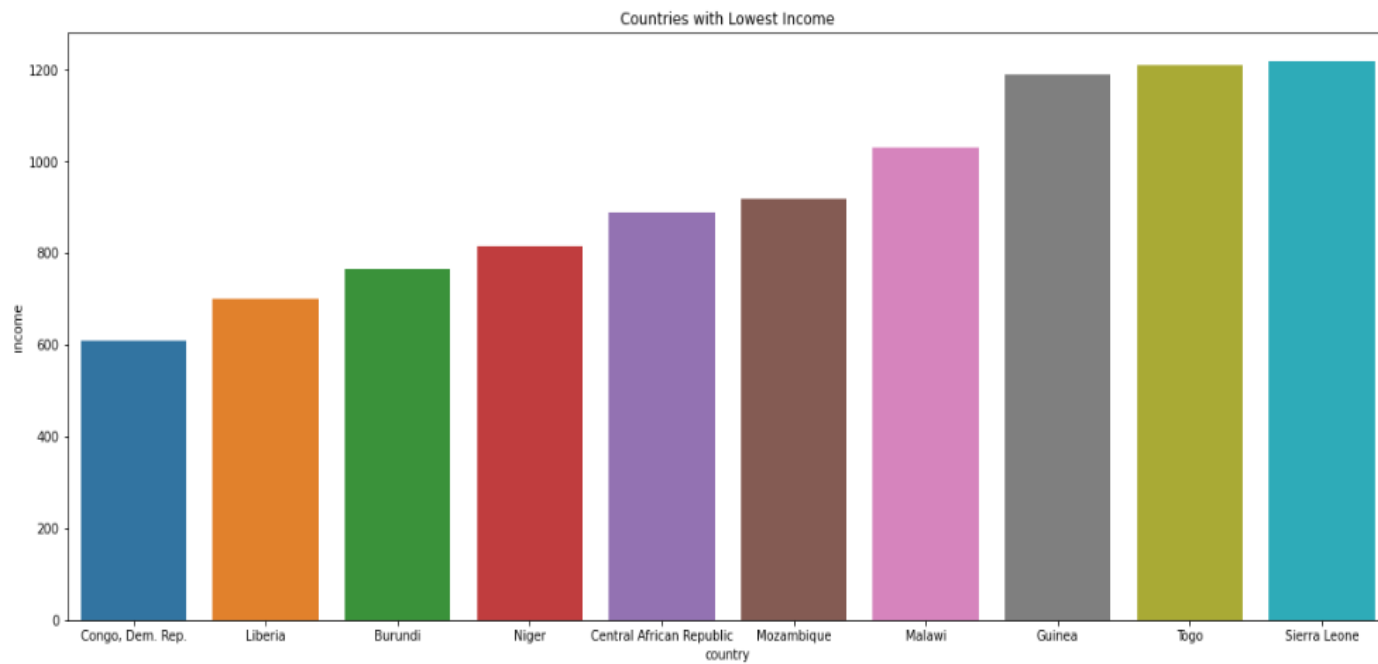
- 1.As per observation , the graphs are not completely Normally distributed. In case of child_mort.income and life_expec,total_fer and gdpp we c et multiple spikes.
- 2.Most of the countries have the Child mortality of less than 50.
- 3.Highest Inflation rate of around 10 was recorded for most of the countries.
- 4.Large number of countries have life expectancy between age of 50 and 80 years
- 5.Most of the countries have a total_fer less than 4.
- 6.Maximum countries have a gdpp less than 20000



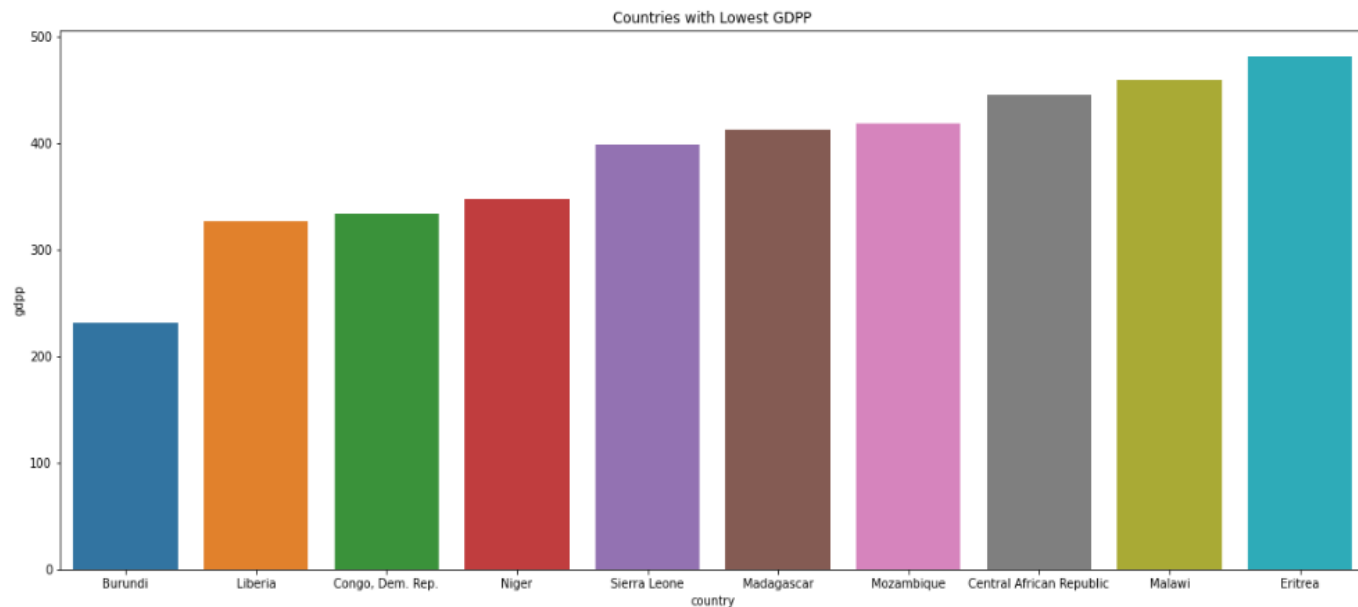
Bivariate Analysis: In this process, two variables are analysed and visualized as Categorical and Continuous variables. Scatter and Bar graphs were used for analysis.



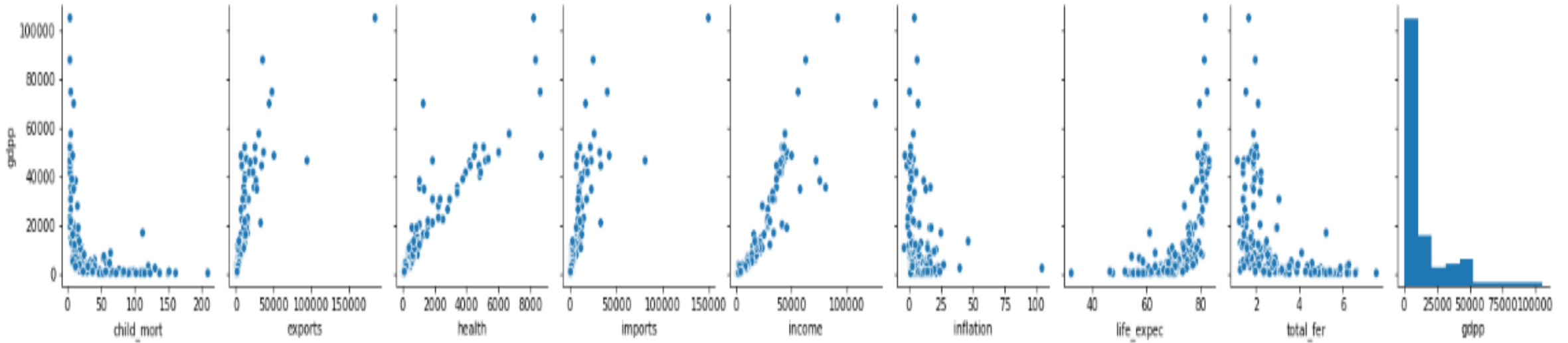
- Country : Haiti and Sierra Leone have the Highest Child mortality in the Country data set.
- Display the list of Top 10 countries with Highest Child mortality.



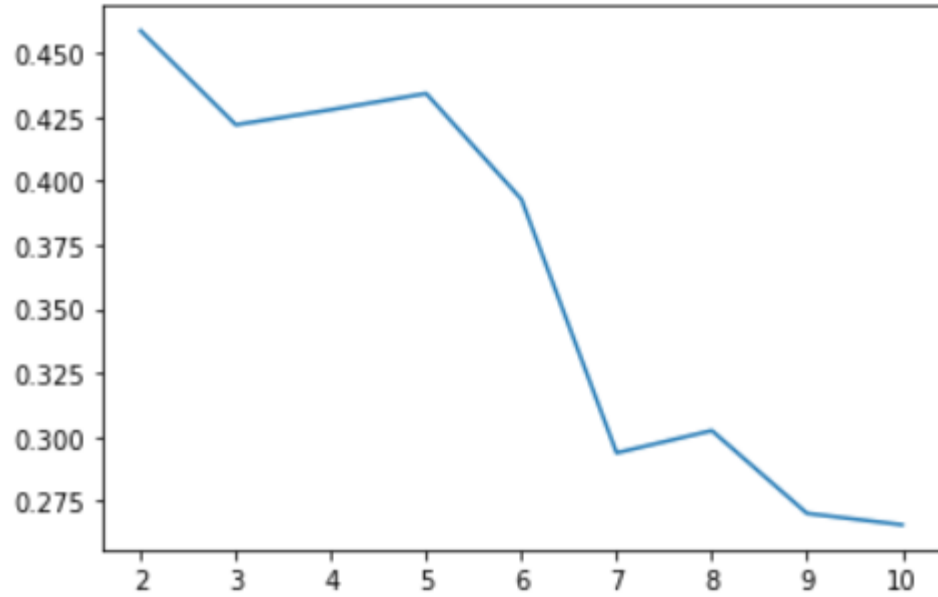
- Country : Congo.Dem.Rep have the Lowest Income amongst all the countries.
- Display the list of Top 10 countries with Lowest Income.



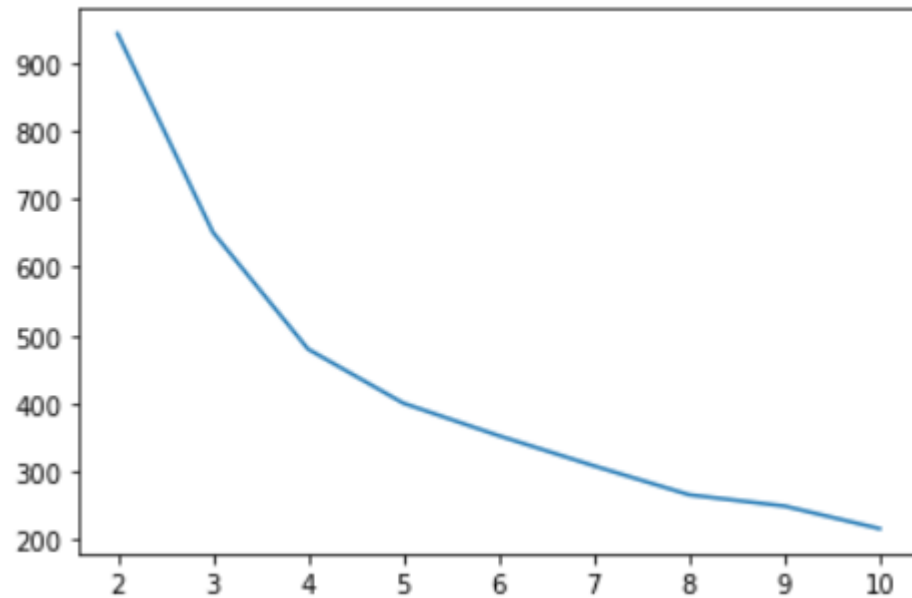
- Country : burundi have the Lowest GDPP amongst all the countries.
- Display the list of Top 10 countries with Lowest GDPP.



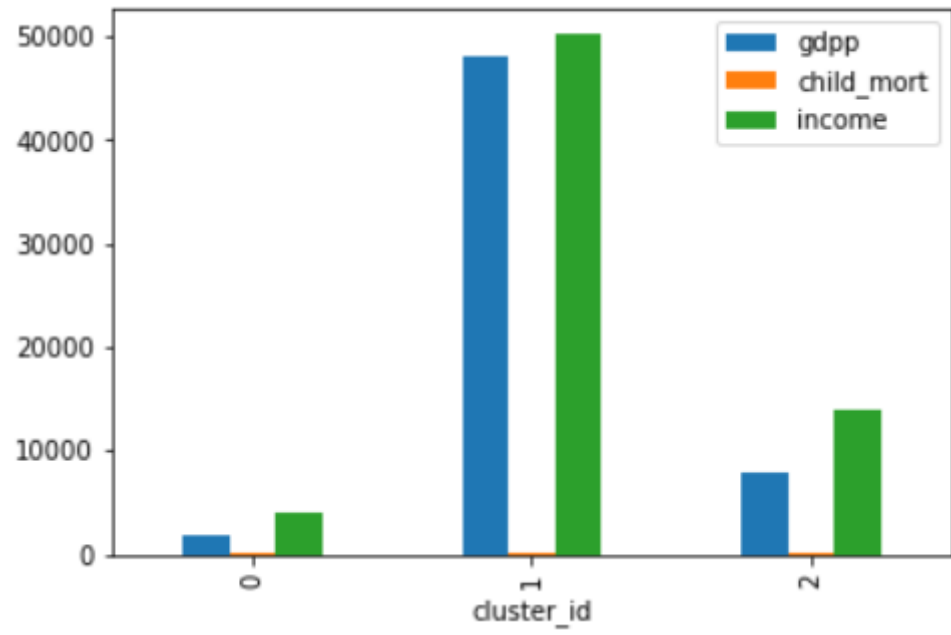
- Exports is increasing in GDP of a country.
- Health expenditure is increasing with increase in GDP of country
- Imports of good and service increases with increase in GDP
- With increasing Income of the employee, the GDP of the country increases.
- With increase in GDP, the inflation rate does not change much.
- With increase in GDP, Life_expec shows an Increasing trend.
- With Increase in GDP, the tota_fer decreases.



➤ It is a Silhouette Curve and value of $K=3$, is considered for Analysis.

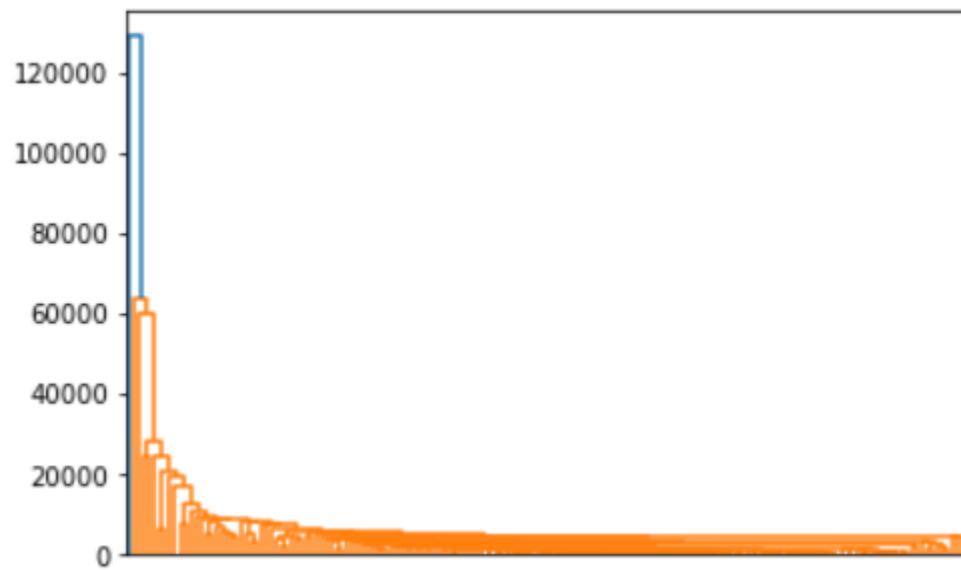


➤ It is an Elbow Curve and value of $K=3$, is considered for Analysis.

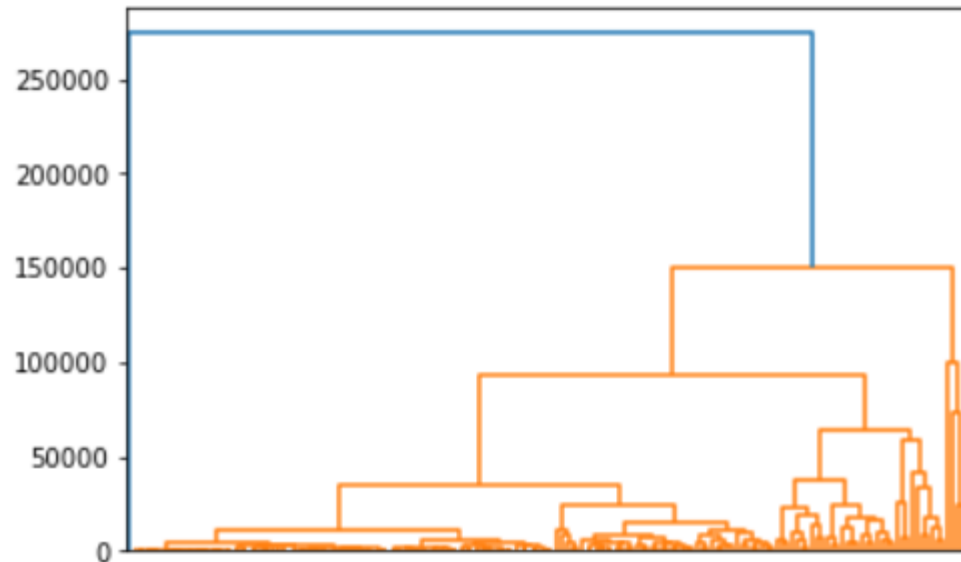


- Cluster with Cluster_Id=0 satisfies our Business Objective of Low GDPP, Low INCOME and High CHILD_MORT.

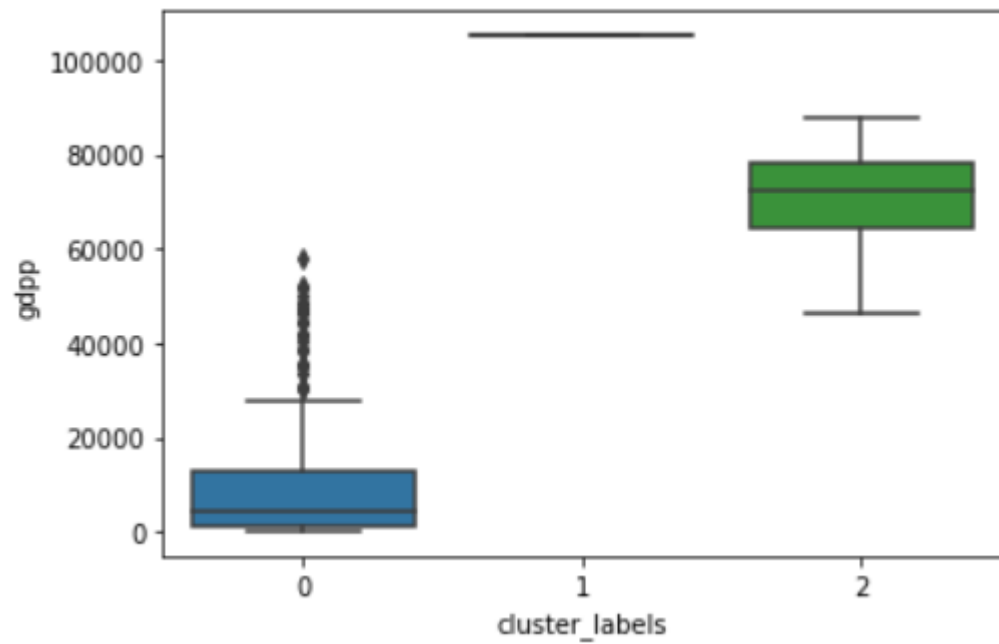
Hierarchical Clustering



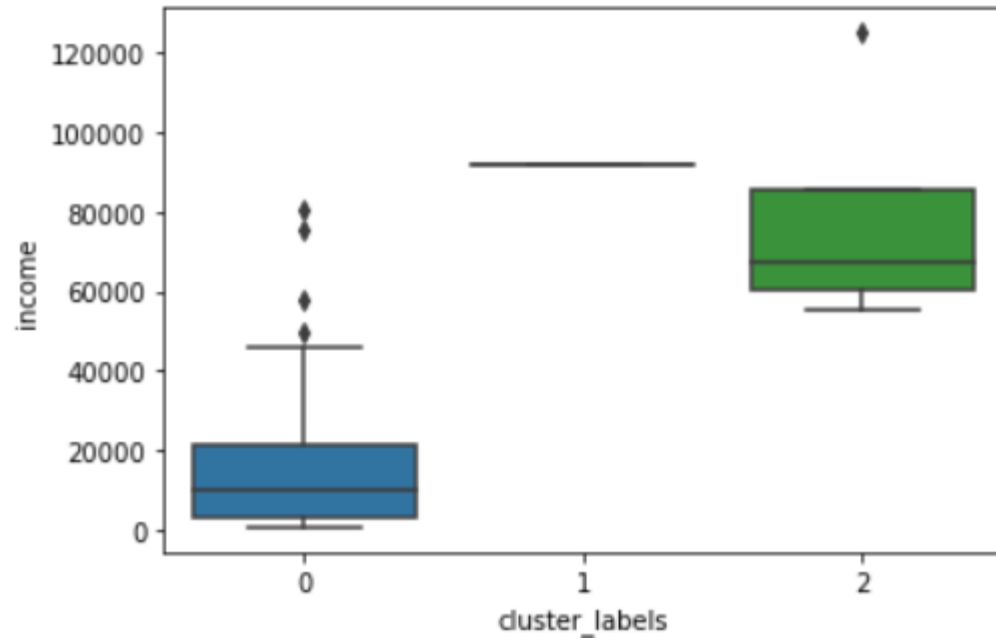
➤ As we can see, it is difficult to read the above Dendrogram, as we have used Single Linkage in this case.



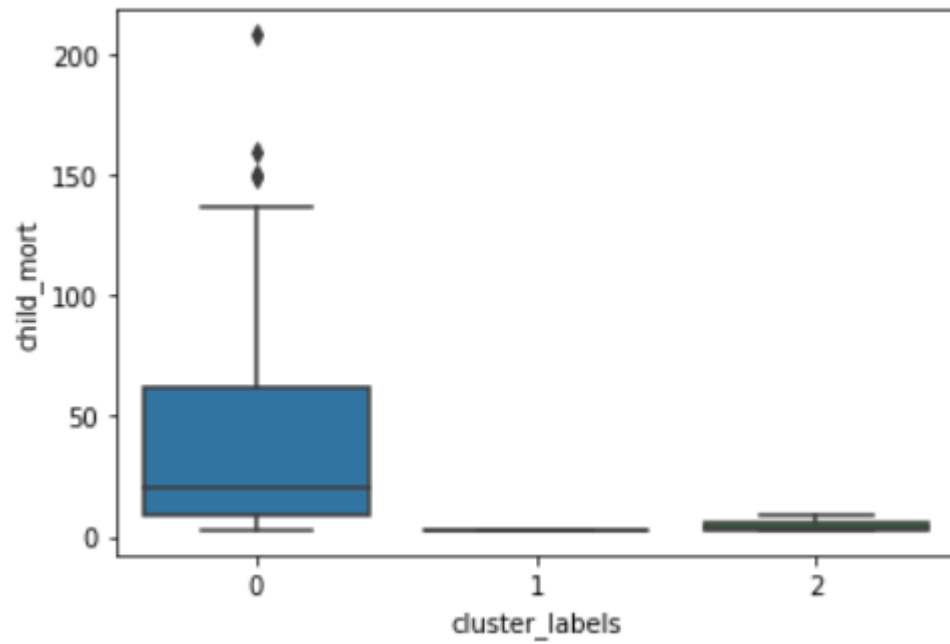
- Based on the Cut at different points in Dendrogram, we may get different number of Clusters.
- We can take value of $K=3$



➤ Cluster 0 have the Lowest GDPP



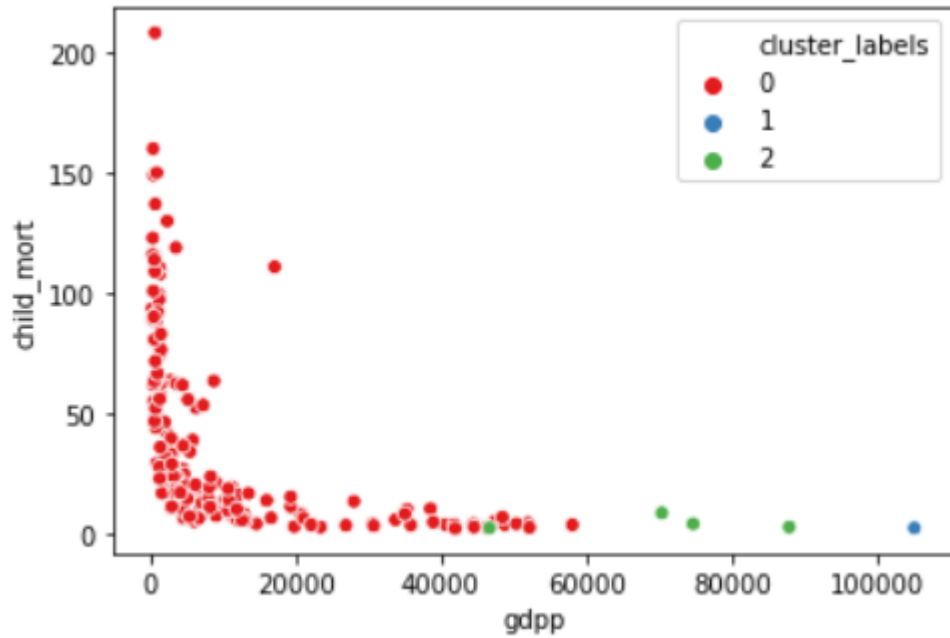
➤ Cluster 0 have the Lowest Income



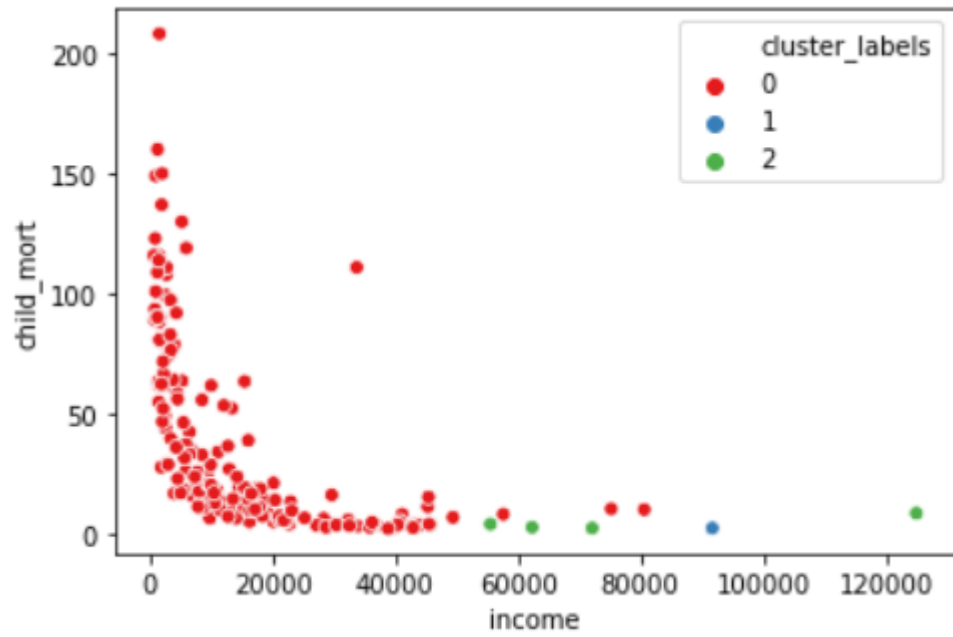
➤ Cluster 0 have the Highest Child mortality

➤ Overall, Cluster 0 satisfies all our Business objective of Low GDPP, low Income and High Child mortality

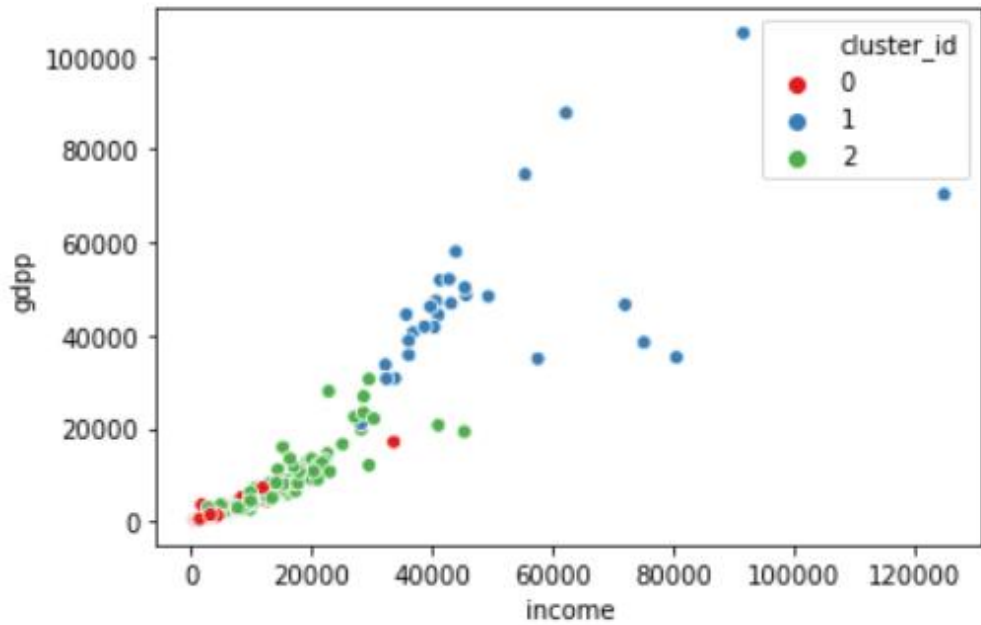
SCATTER PLOTS: Visualization



➤ Child mortality decreases with increase in GDPP



➤ Child mortality decreases with increase on Income



➤ GDPP increases with increase in Income

Final List of Countries that need financial AID

- Country names

country	child_mort	exports	health	imports	income	inflation	life_expec	total_fer	gdpp	cluster_id	cluster_labels
Burundi	93.6	20.61	26.80	90.55	764	12.30	57.7	6.26	231	0	0
Liberia	89.3	62.46	38.59	302.80	700	5.47	60.8	5.02	327	0	0
Congo, Dem. Rep.	116.0	137.27	26.42	165.66	609	20.80	57.5	6.54	334	0	0
Niger	123.0	77.26	17.96	170.87	814	2.55	58.8	7.49	348	0	0
Sierra Leone	160.0	67.03	52.27	137.66	1220	17.20	55.0	5.20	399	0	0

Thank You