

HELP funding

- by Prerna Prakash

Problem Statement:

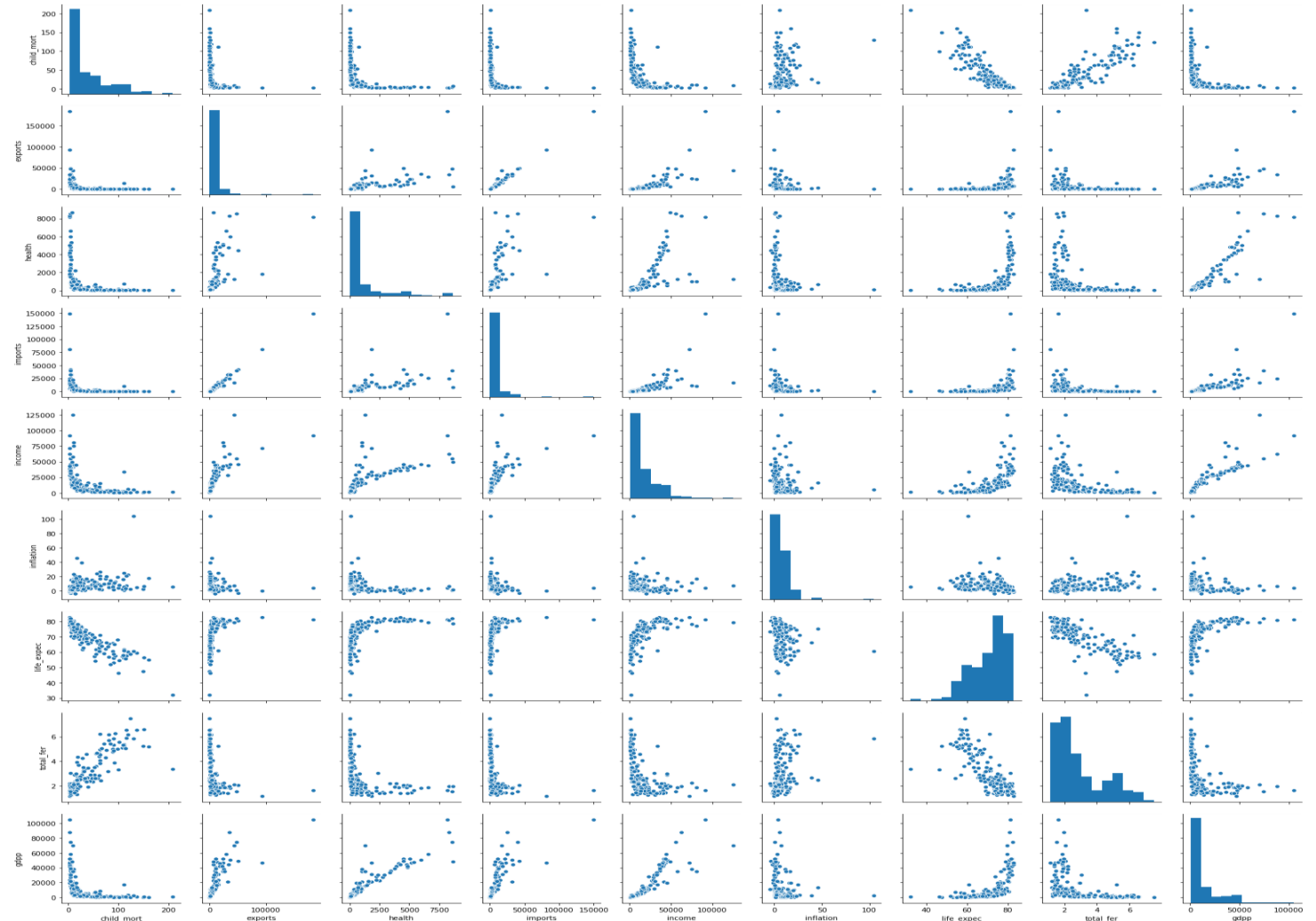
The organisation wants to identify the countries which are in direst need of aid so that it can channel its collected funds in the right direction.

We have categorised the countries using some socio-economic and health factors that determine the overall development of the country and suggested the countries the NGO can to focus on the most.

Exploratory Data Analysis:

From the pair plot of the dataset we can make some inferences like:

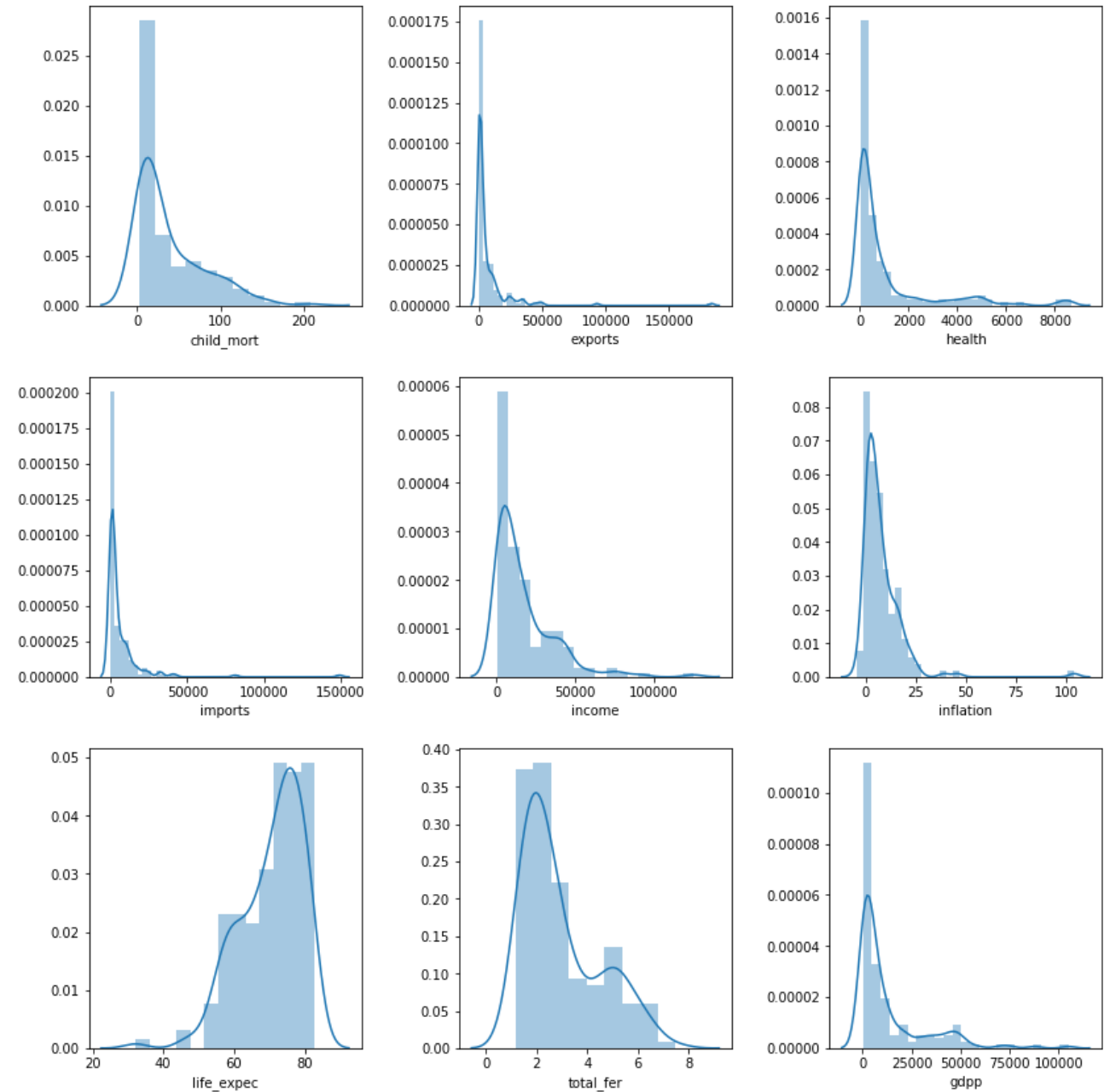
- Countries with higher net income per person observe higher expenditure on factors like health, imports and exports.
- Low inflation rate is observed for countries with lower as well as higher gdp, net income per person, imports of goods and services per capita and exports of goods and services per capita



By checking the distribution of each column we can figure out which columns will be beneficial to us while profiling the clusters.

Most of the countries have exports in the range of 0-50000, and as we see no variation there, this can not be considered a good attribute to segregate the clusters. Similarly there are other attributes with this characteristics like health, imports, inflation, life_expec and total_fer

Whereas attributes like gdp, child_mort and income show variation in the distribution of values hence making them useful for profiling the clusters.

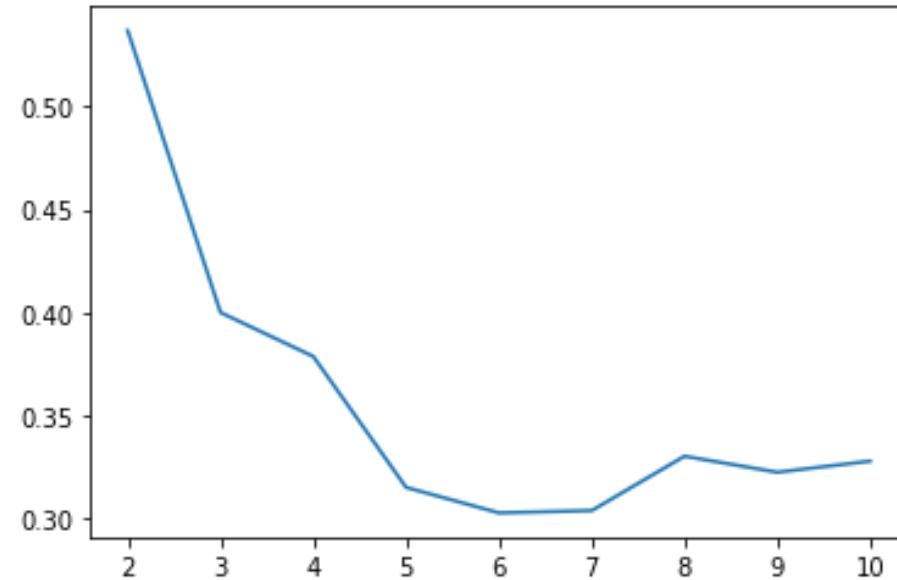


K-Means clustering:

The first figure displays:

number of clusters(x- axis) against the silhouette score(y-axis)

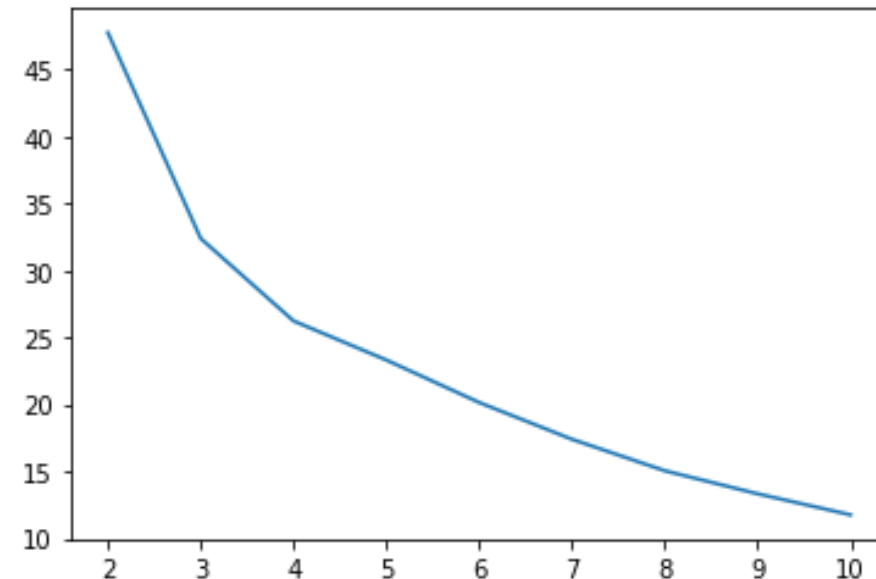
Silhouette score of clusters starting from 2 to 10 are plotted in the graph.



The second figure displays:

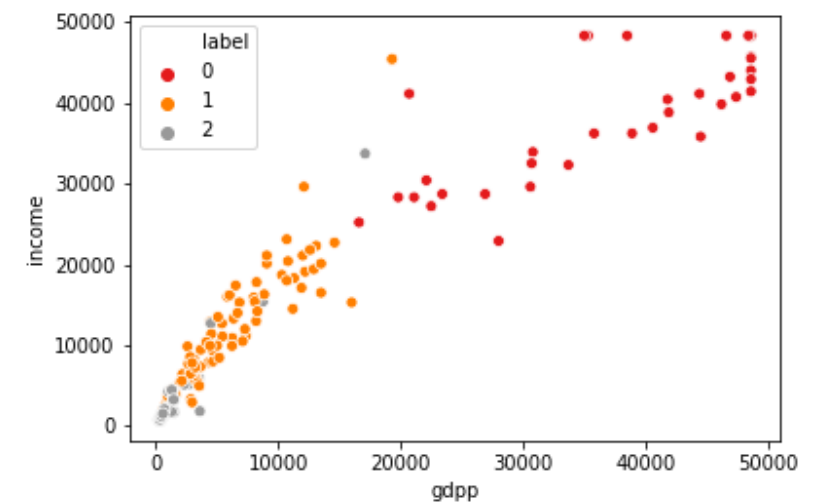
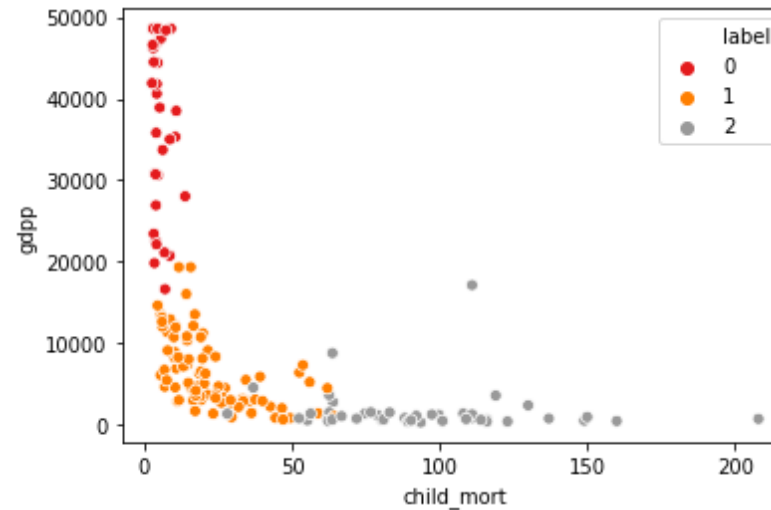
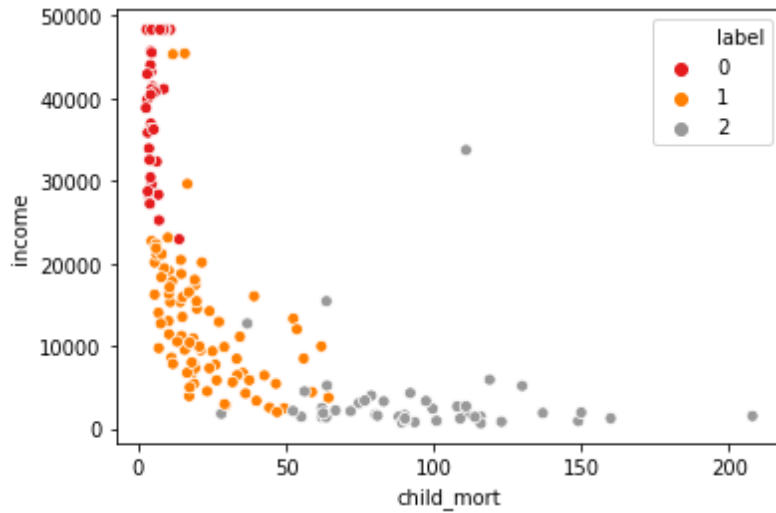
number of clusters(x- axis) against the inertia(y-axis)

Inertia of clusters starting from 2 to 10 is plotted in the graph.



We observe that there is a significant change in the value when the number of clusters chosen is 3. Hence, we proceeded with the clustering with number of clusters as 3.

Plotting the cluster formed by running the K-Means algorithm for the attributes chosen for cluster profiling:

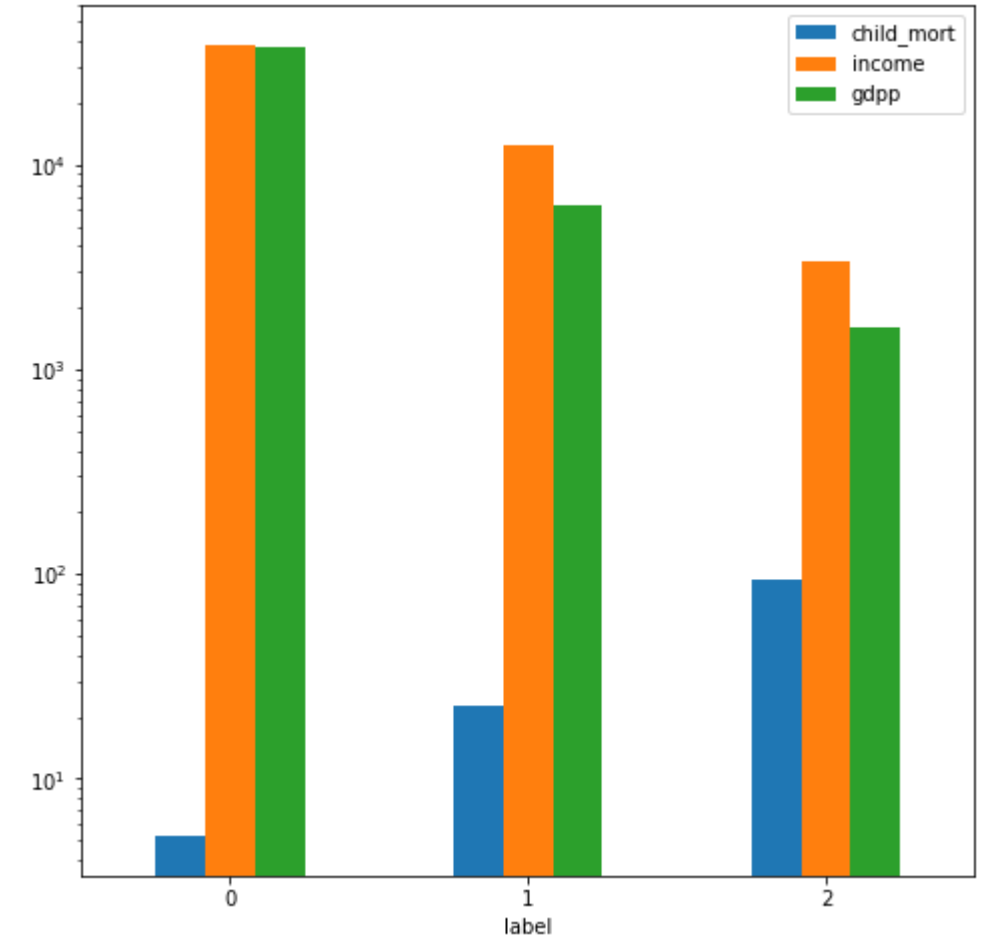


Cluster profiling:

Three clusters are formed – 0, 1 and 2

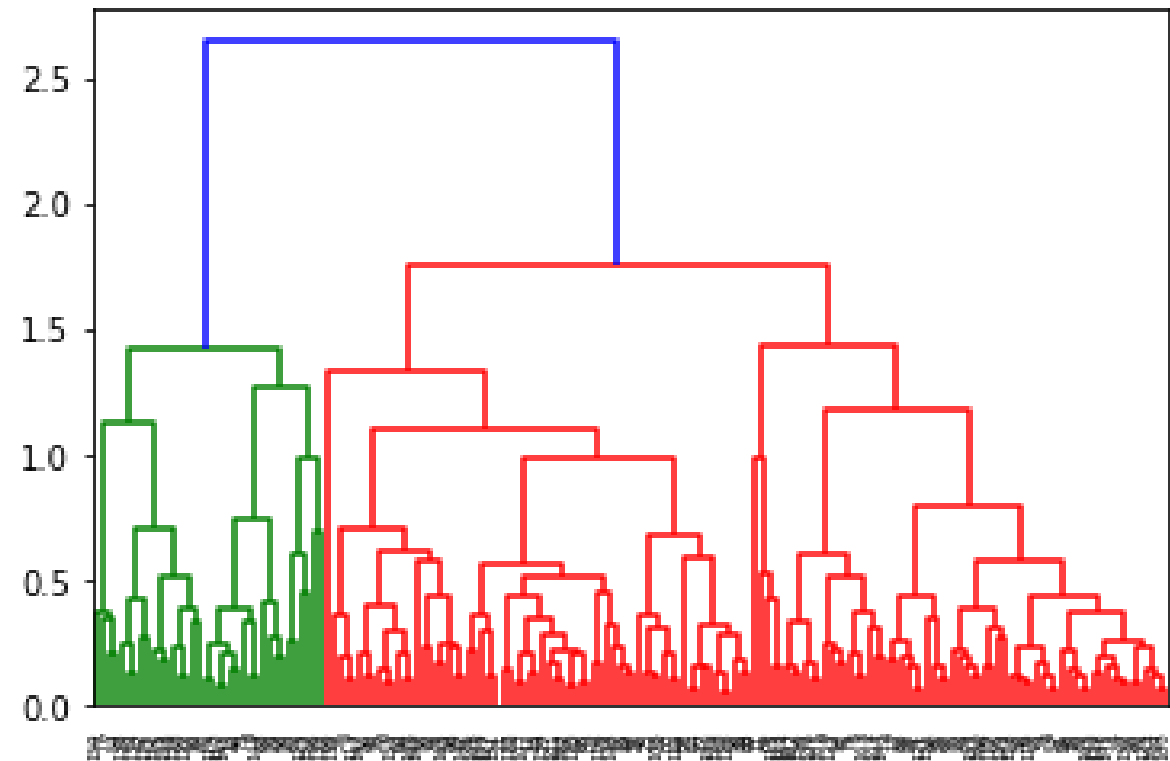
- Cluster 0: Includes the countries with high income and gdpp but with low child mortality rate.
- Cluster 1: Includes the countries with moderate income and gdpp with comparatively higher mortality rate than cluster 0.
- Cluster 2: Includes the countries with low income and gdpp but with highest child mortality rate.

So the countries that we are concerned about fall under cluster 2. We must now choose the top 5 countries from this cluster.



Hierarchical clustering

This dendrogram is created using the complete linkage. We can observe that the dendrogram can be sliced at the level where three clusters are being formed.



Cluster profiling:

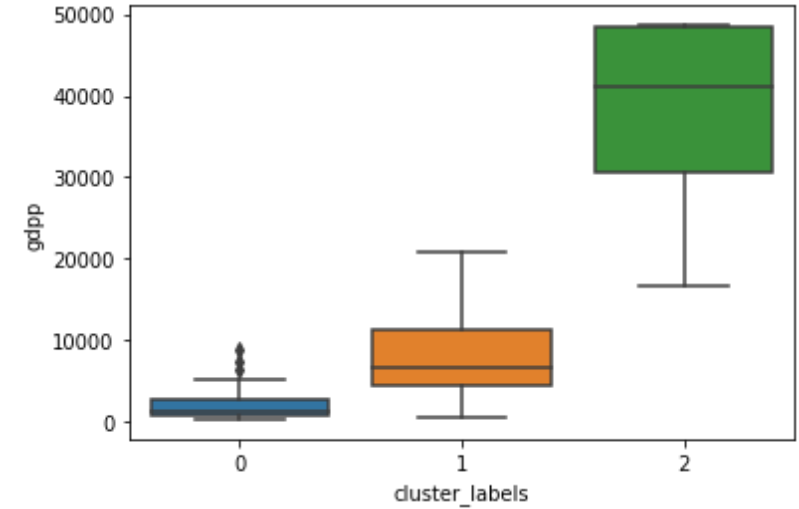
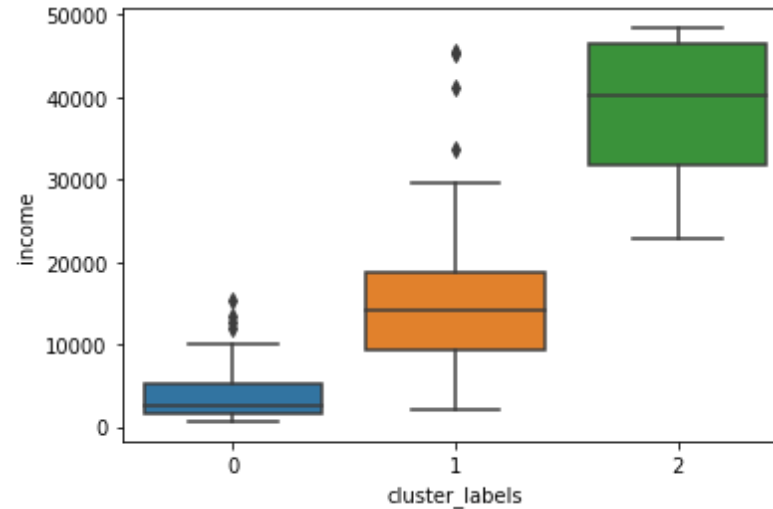
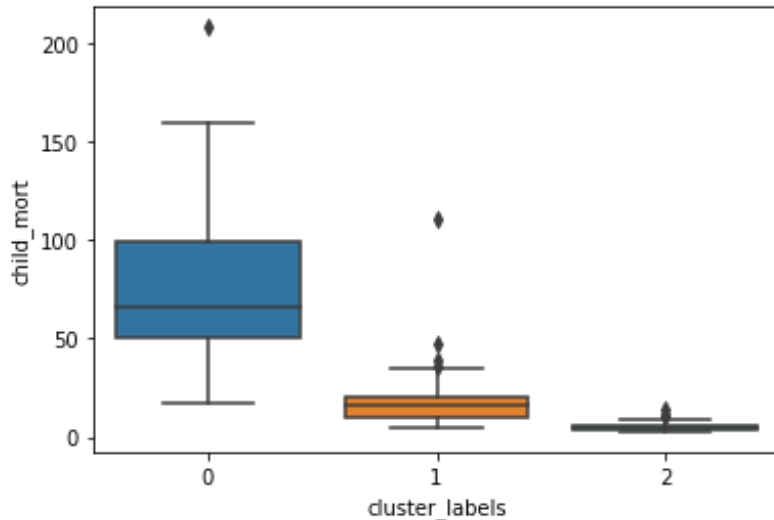
Three clusters are formed – 0, 1 and 2

Cluster 0: Includes the countries with low income and gdpp but with high child mortality rate.

Cluster 1: Includes the countries with moderate income and gdpp with comparatively moderately higher child mortality rate than cluster 0.

Cluster 2: Includes the countries with high income and gdpp but with lowest child mortality rate.

So the countries that we are concerned about fall under cluster 0. We must now choose the top 5 countries from this cluster.



The top 5 countries that are shortlisted for the aid in case of:

K-Means clustering

	country
66	Haiti
132	Sierra Leone
32	Chad
31	Central African Republic
97	Mali

Hierarchical clustering

	country
66	Haiti
132	Sierra Leone
32	Chad
31	Central African Republic
97	Mali

Conclusion:

We observe that both the clustering algorithm return the same result.
Hence, we propose that the countries that need aid are:

	country
66	Haiti
132	Sierra Leone
32	Chad
31	Central African Republic
97	Mali