Problem Statement and the analysis approach:

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

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.

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

Analysis approach:

To find the countries which are direst need of aid, here I used K – means clustering and Hierarchical clustering approach after treating the outliers in the data by capping method and also I have chosen the country based on low gdpp, low income and high child mort.

Result of clustering model:

K-Means Clustering:

After performing outlier treatment in the dataset and scaling the data with standard scaling mechanism then I have found the K – means clustering K value is 3 using Silhouette curve method and elbow curve method.

Performed the K means clustering in the scaled dataset by k value equal to 3 and added the cluster id in the original data set and performed cluster profiling using the gdpp, child_mort and income column.

Found the countries which are in need of funding based on low gdpp, low income and high child mort in the cluster 1.
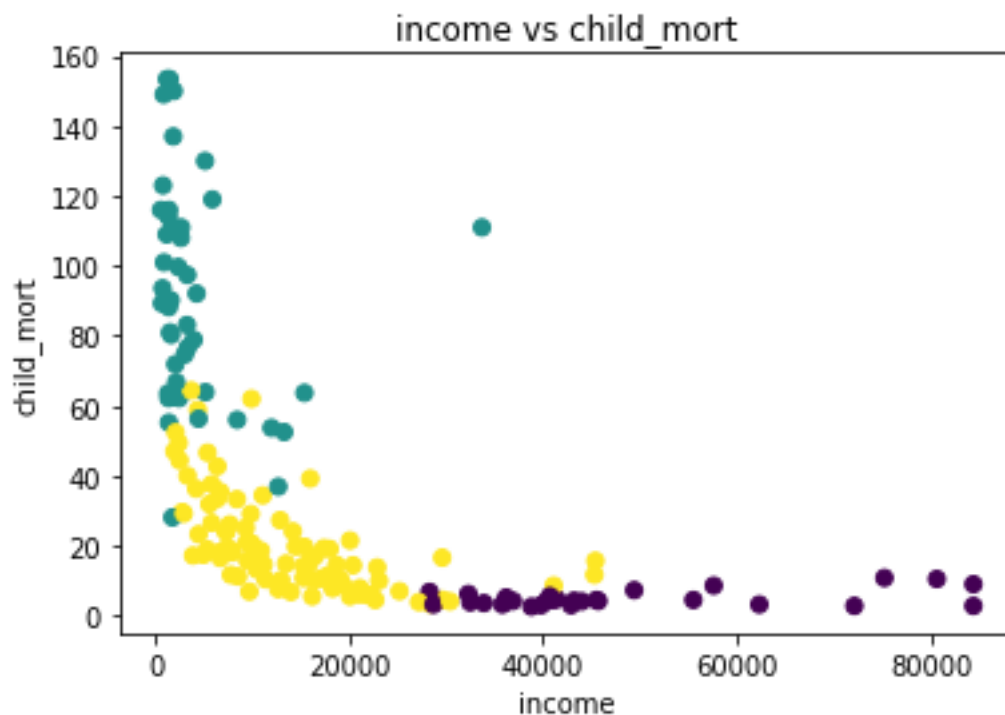
Hierarchical Clustering:

After performing outlier treatment in the dataset and scaling the data with standard scaling mechanism then I have used both single linkage and complete linkage and chosen complete linkage method clustering.

Added cluster lables in the original data set and performed cluster profiling using the gdpp, child_mort and income column.
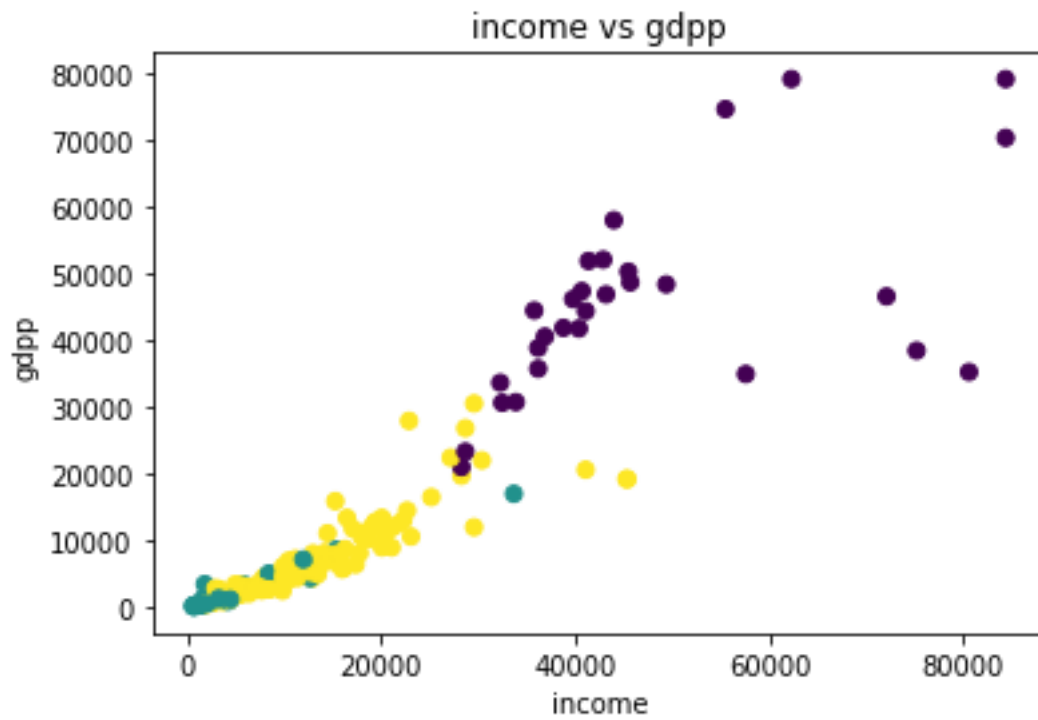
Found the countries which are in need of funding based on low gdpp, low income and high child mort in the cluster 0.
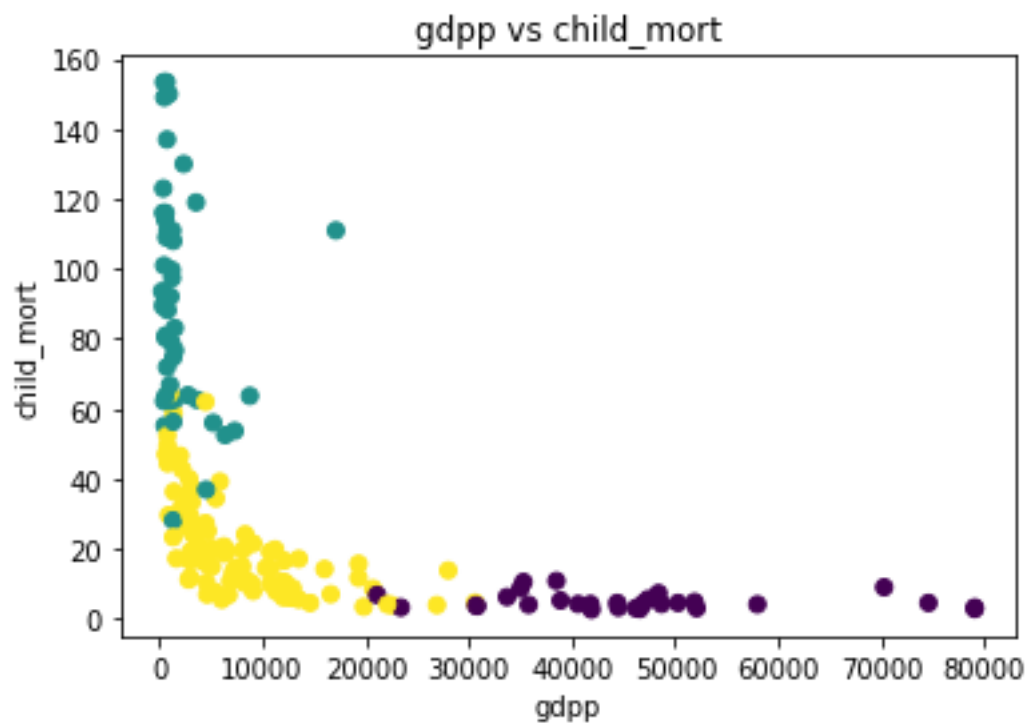
Visualisations:

K – Means:

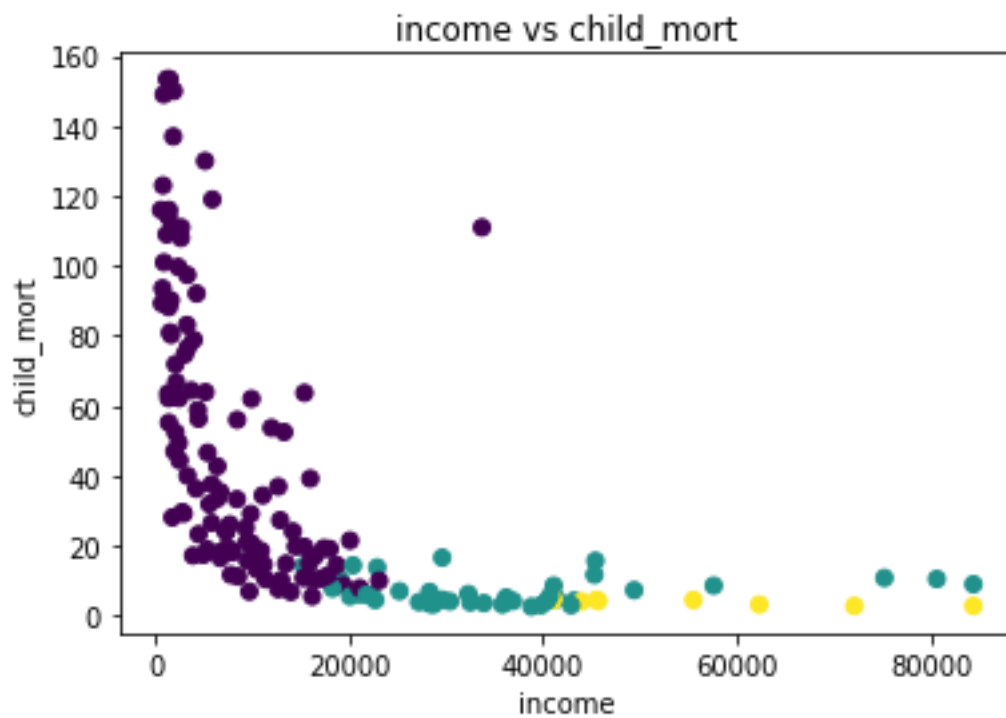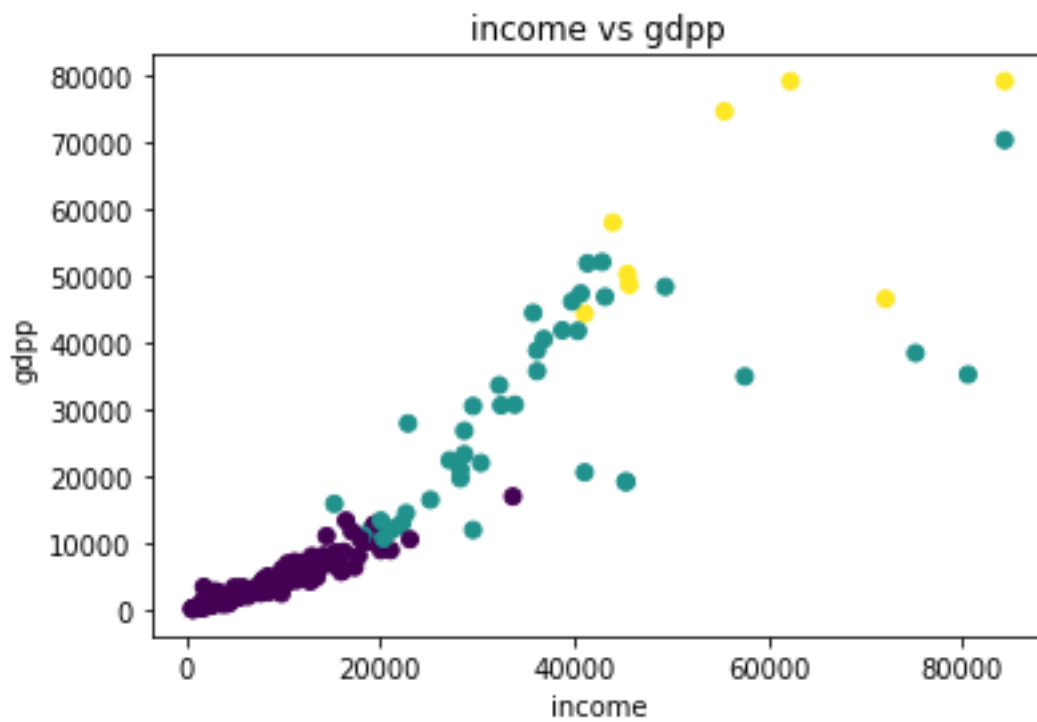Child Mort is increasing when there is less income.

income vs gdpp

When income increases gdpp increasing

gdpp vs child_mort

Child mort increasing when there is less gdpp in the country

Hierarchical complete linkage clustering:

income vs child_mort

Child Mort is increasing when there is less income.

income vs gdpp

When income increases gdpp increasing

gdpp vs child_mort
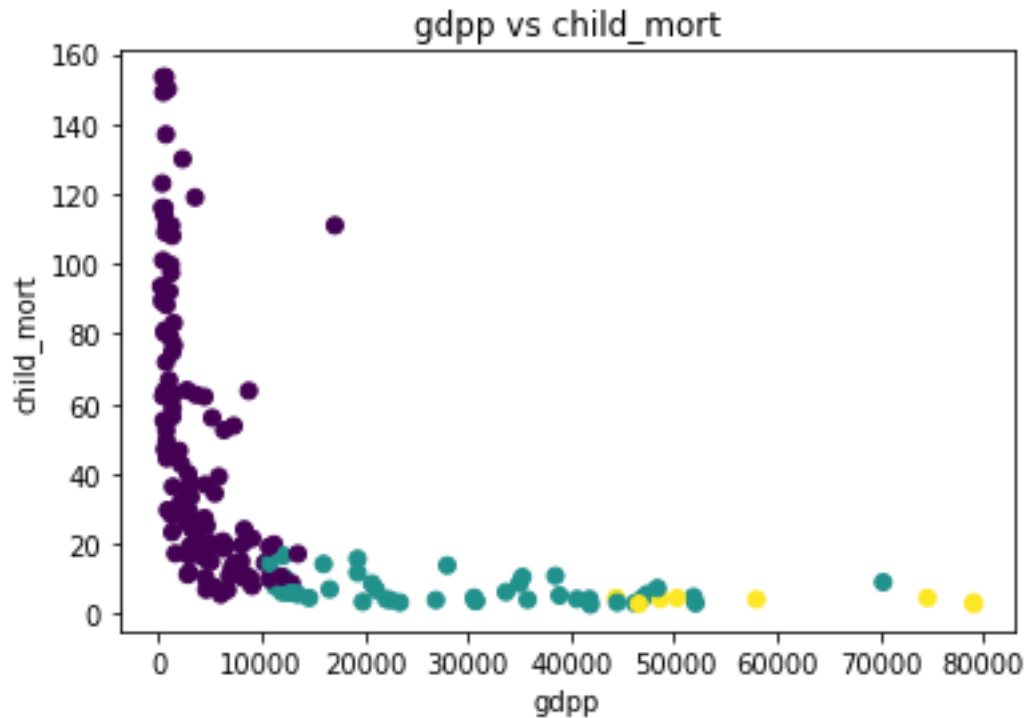
Child mort increasing when there is less gdpp in the country

Analysis:

1.Based on the K-means clustering and Hierarchical Clustering analysis,Found the top 5 countries that are in need for aid

2.Choosing the country based on low gdpp,low income and high child mort

3.Top 5 countries in the both type of clustering analysis is same

4.Top 5 countries that are in need for aid

    1.Burundi

    2.Liberia

    3.Congo, Dem. Rep.

    4.Niger

    5.Sierra Leone