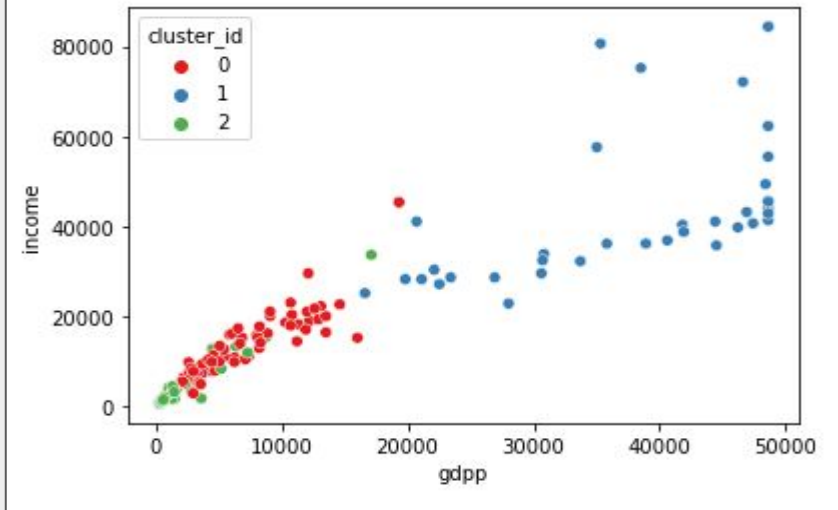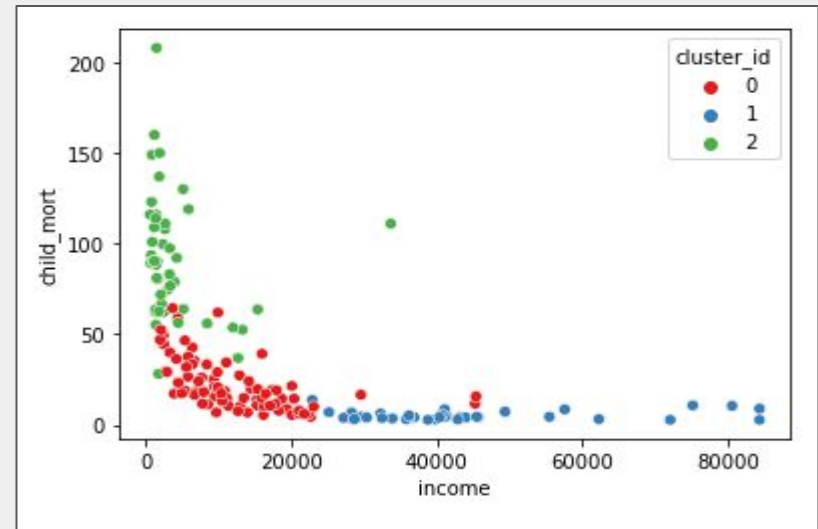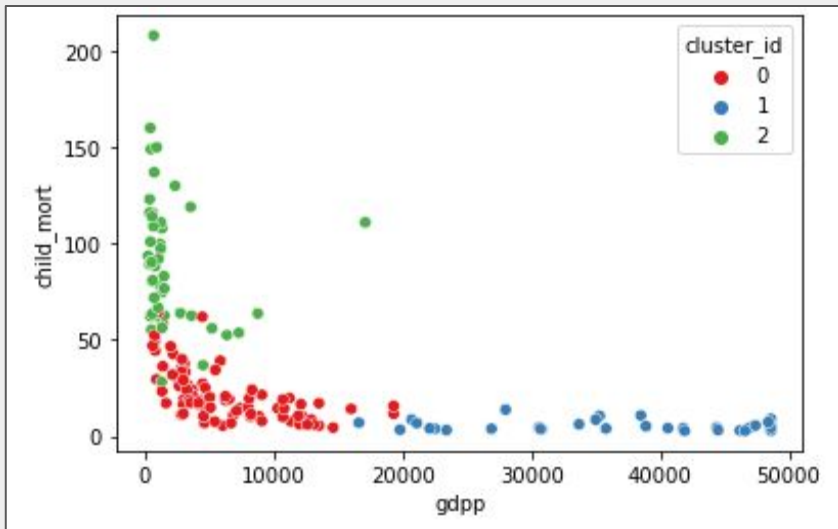# Clustering Assignment

Shilpa Goel

- We have data of 167 countries, and we have to identify countries which require help.
- Countries with low gdp, low income and high child mortality rate would be on our focus.
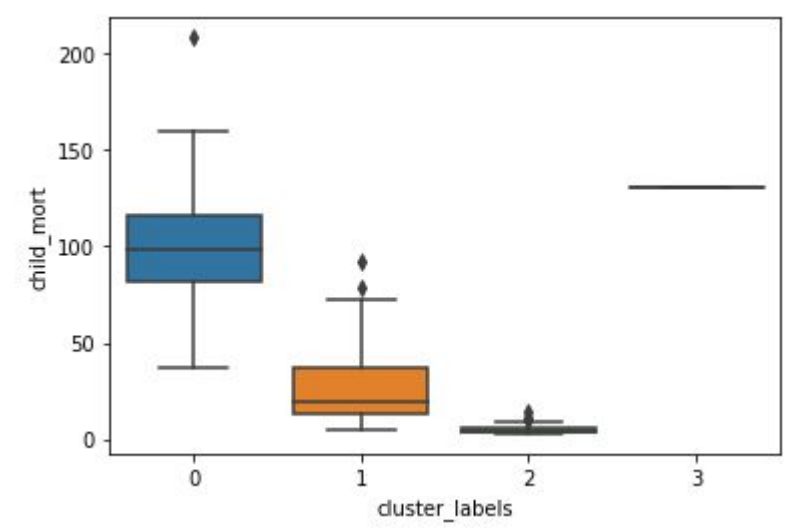
We perform below steps to achieve our goal :

1. Data Quality Check
2. Univariate and Bivariate
3. Outlier
4. Scaling
5. Hopkin's Test
6. Find the best value of k using SSD Elbow, Silhoutte Score
7. Using the final value of k=3, perform the kMeans analysis
8. Visualize it using scatterplot
9. Perform Cluster profiling for columns : gdpp, child_mort and income
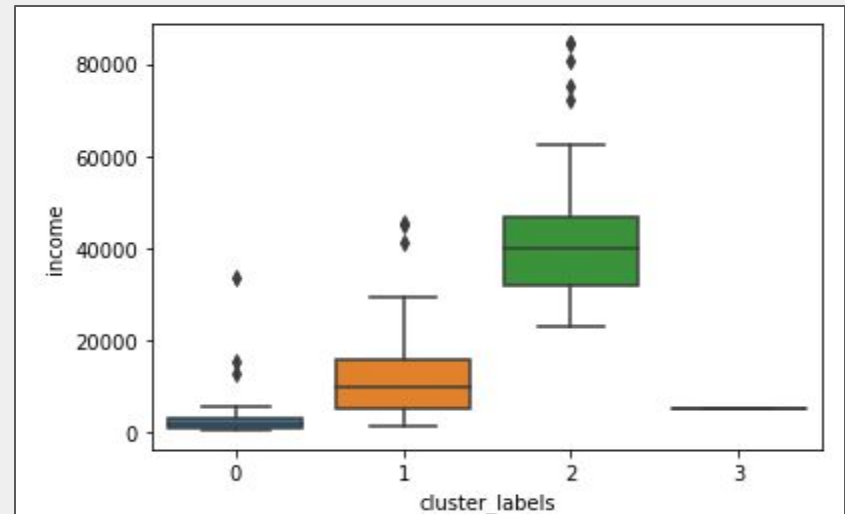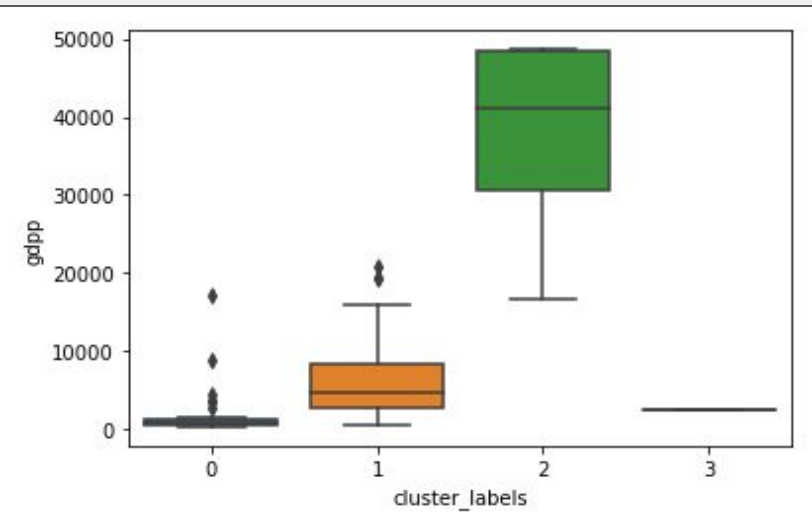10. Hierarchical Clustering: Single linkage, Complete Linkage

- *K-means clustering performed using number of clusters equal to 3*
- *Cluster with id 2 is our required cluster*

- ***Hierarchical clustering performed using number of clusters equal to 4***
- ***Cluster with id 0 is our required cluster***

From both types of clustering, we conclude the 5 countries which are in dire need of aid

1. Burundi
2. Liberia
3. Congo, Dem. Rep.
4. Niger
5. Sierra Leone