

Clustering and PCA Assignment

Identifying impoverished countries: Abstract

Business Objective:

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. To identify this, the top 5 most impoverished countries are selected and are divided the \$10 million

Goals:

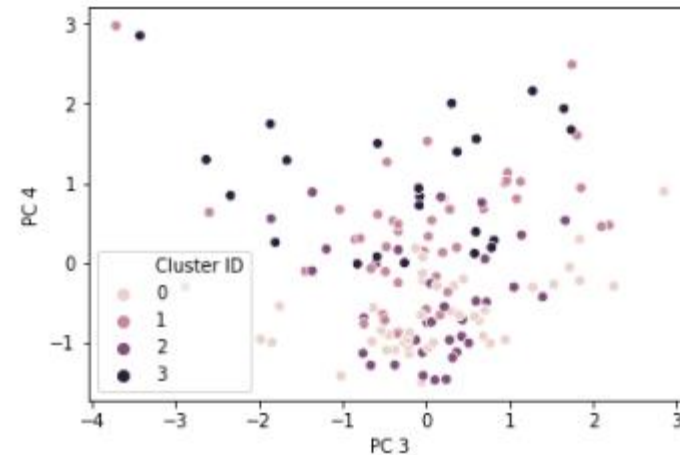
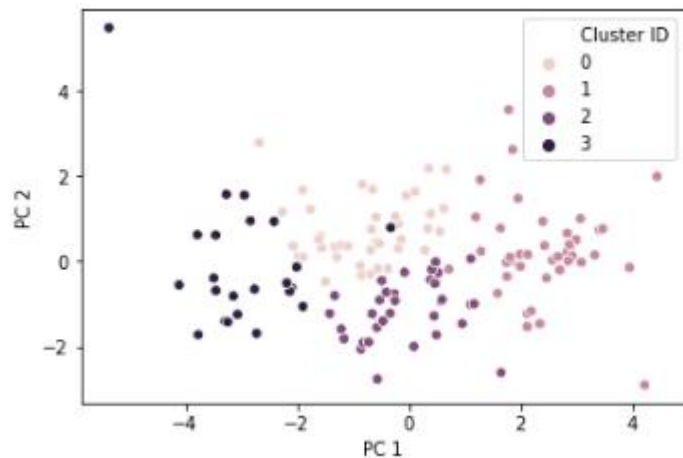
To find the top 5 impoverished countries.

Prediction Technique:

Clustering – K-Means and Hierarchical Clustering using Principal Component Analysis

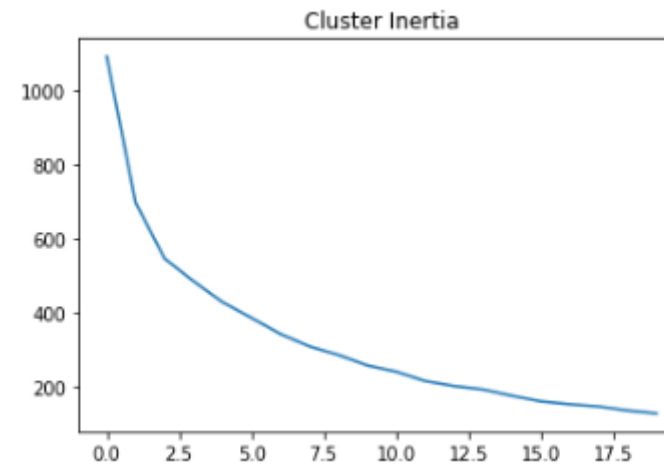
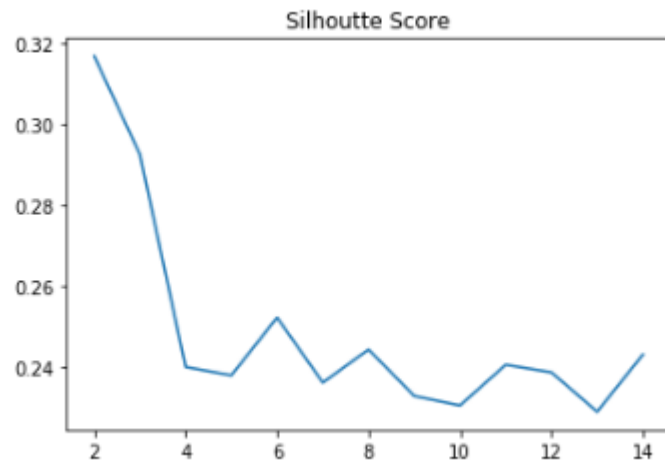
Solution Approach

- Remove the outlier data, about 85% of the data is retained, this contains the higher range outliers only as the lower range outliers are critical.
- Principal Component Analysis :
Features are transformed into components as linear combinations of the original variables.
- While performing PCA, the K components of K-means algorithm is arrived at 4 and the its total variance is explained by 95%
- Perform Clustering on these PCA's to get clear clusters:



K- Means Clustering

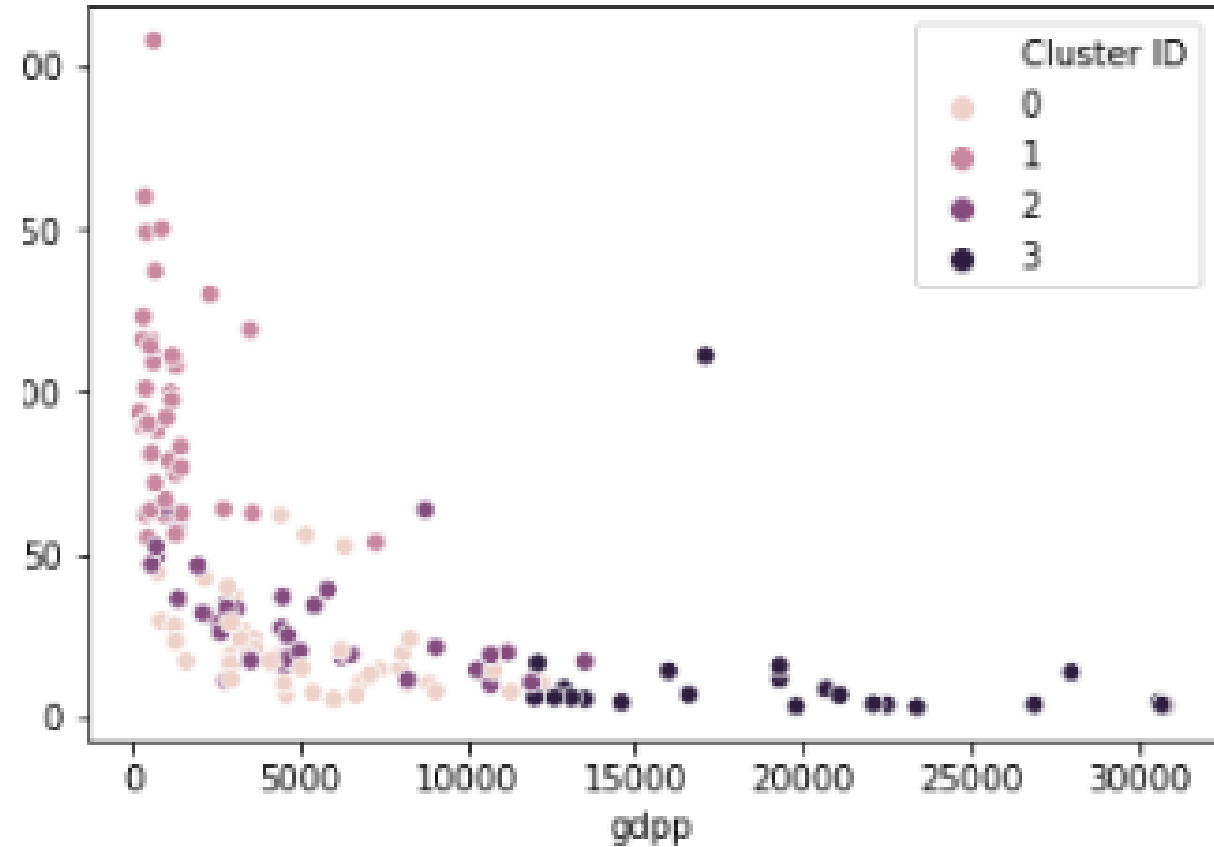
Ideal Value of K is found by Silhouette Score and Elbow Curve:



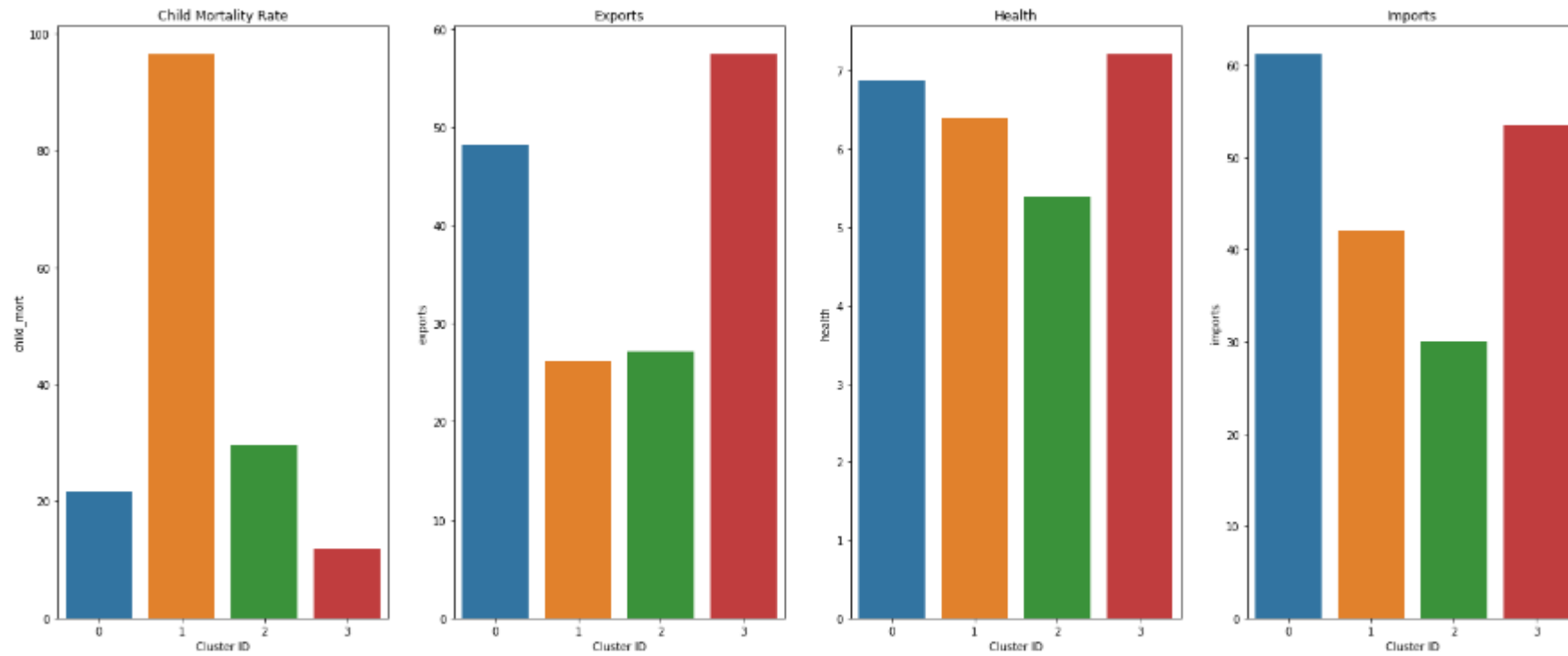
With the output of these two, we can arrive at a conclusion that 4 is an optimum number for clustering, these are confirmed with the equal distribution of clusters also.

K means Clustering Results for countries

Clear Clusters are formed

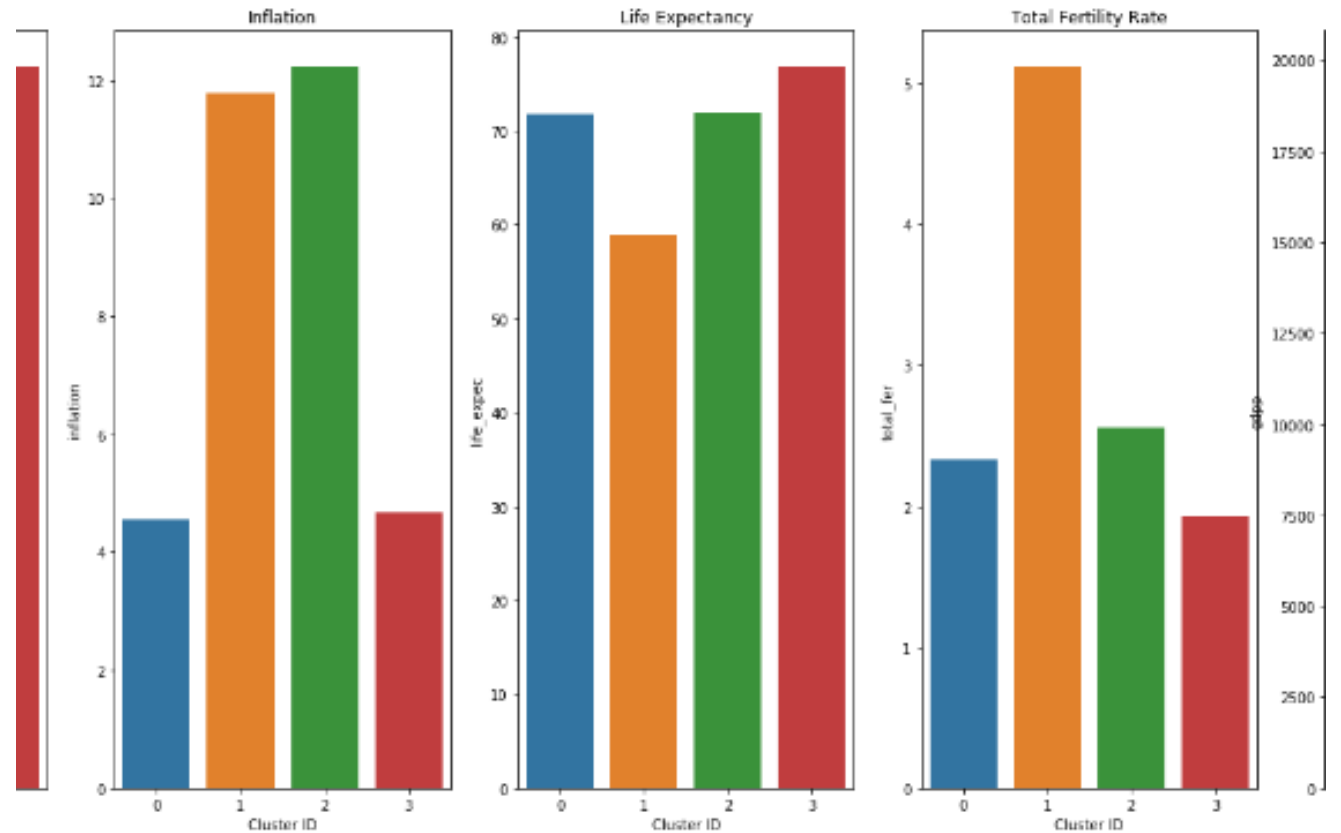


K means Clustering Results



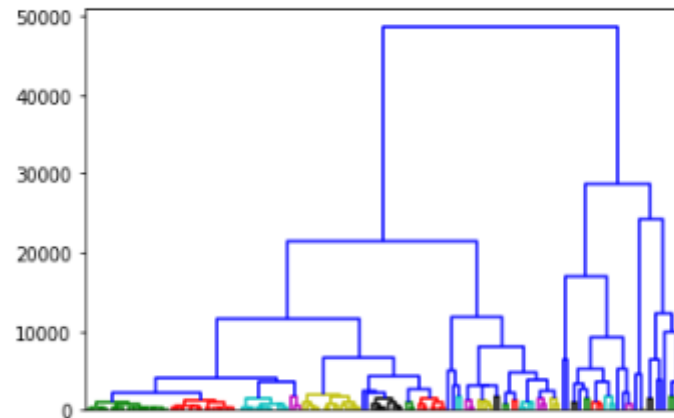
K means Clustering Results

Thus Cluster 2 is found to be the most under privileged cluster



Hierarchical Clustering

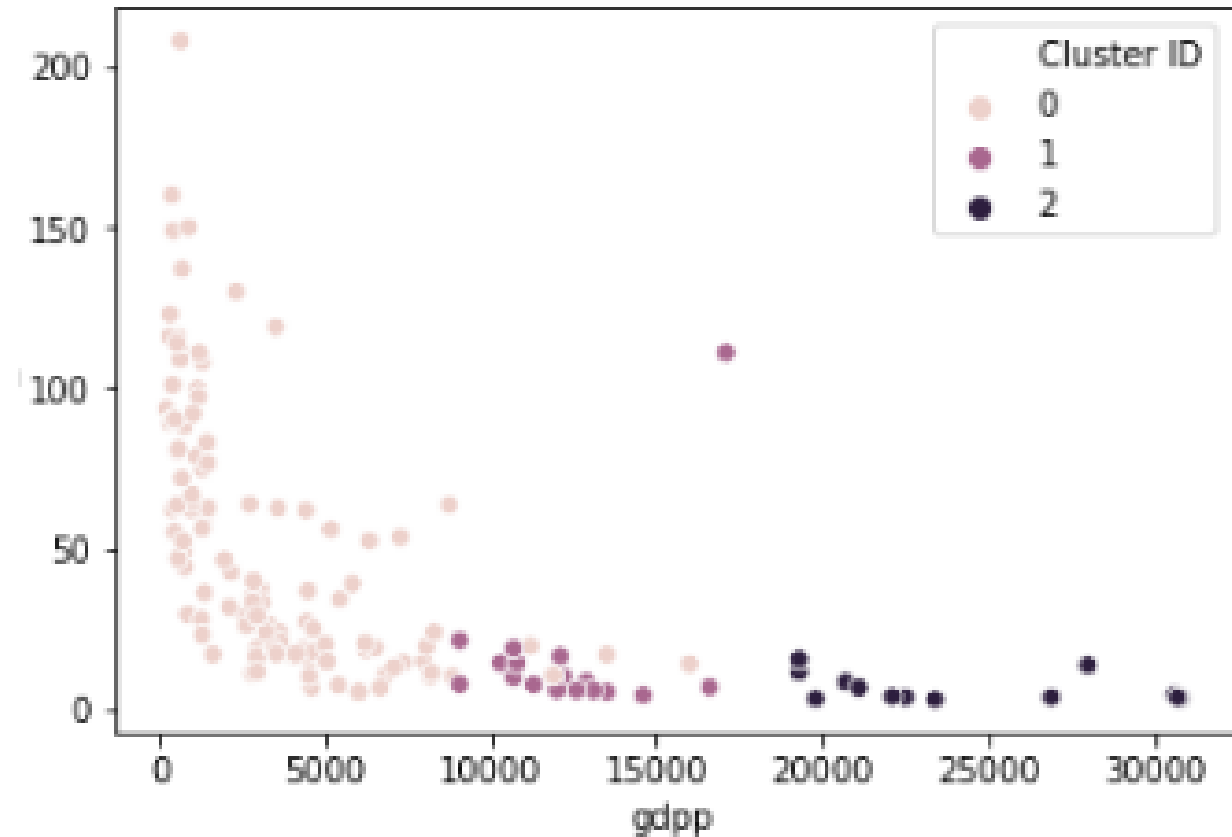
The number of clusters are determined by the point on y-axis which cuts through maximum number of clusters.



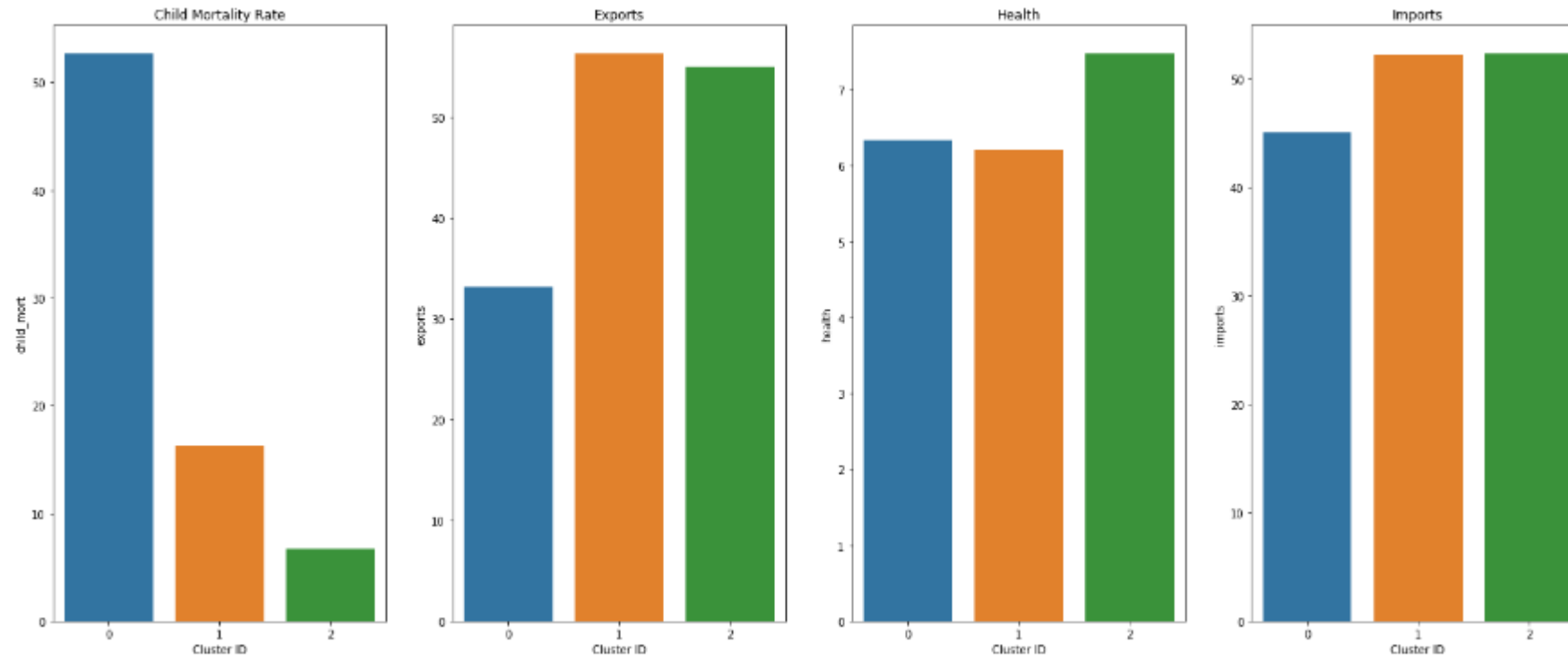
$K = 3$ is optimal for this dendrogram.

Hierarchical Clustering Results for countries

Clear Clusters are formed

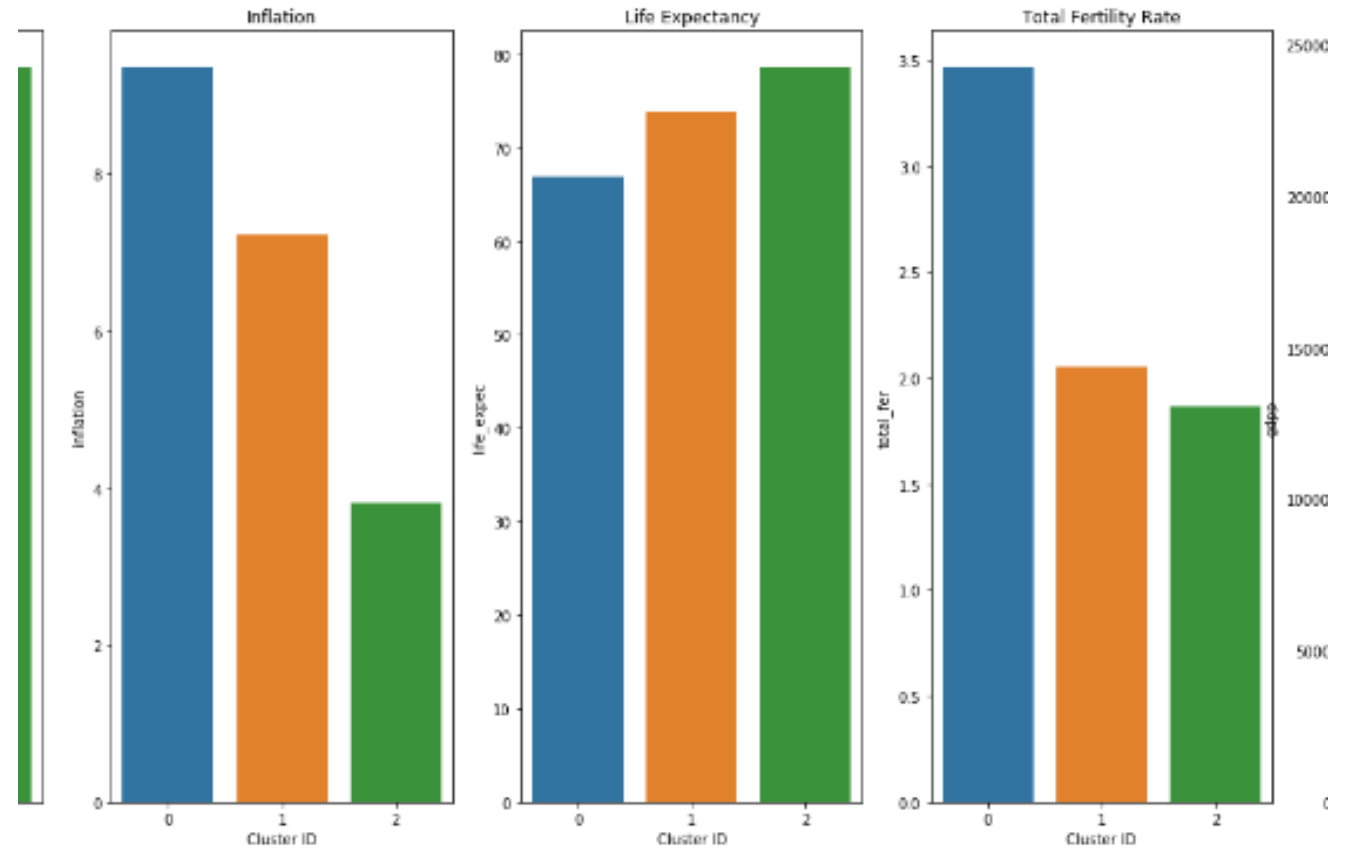


Hierarchical Clustering Results



Hierarchical Clustering Results

Thus Cluster 0 is found to be the most under privileged cluster



Conclusion

Thus, Top 5 countries that are in dire help common to both hierarchical cluster 0 and k-means cluster 2 are :

- 1. Haiti
- 2. Sierra Leone
- 3. Chad
- 4. Central African Republic
- 5. Mali