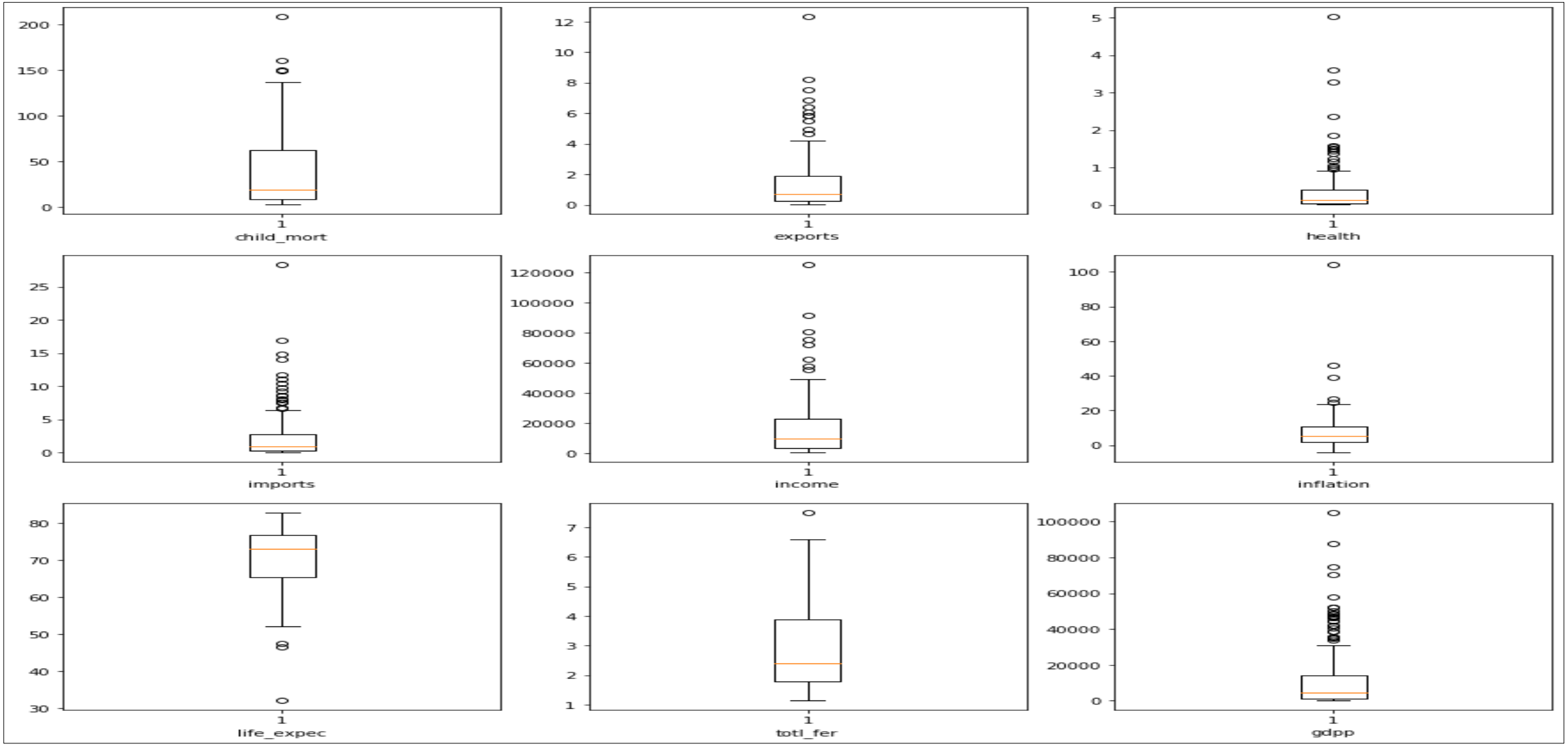
Detecting outlier from the data :

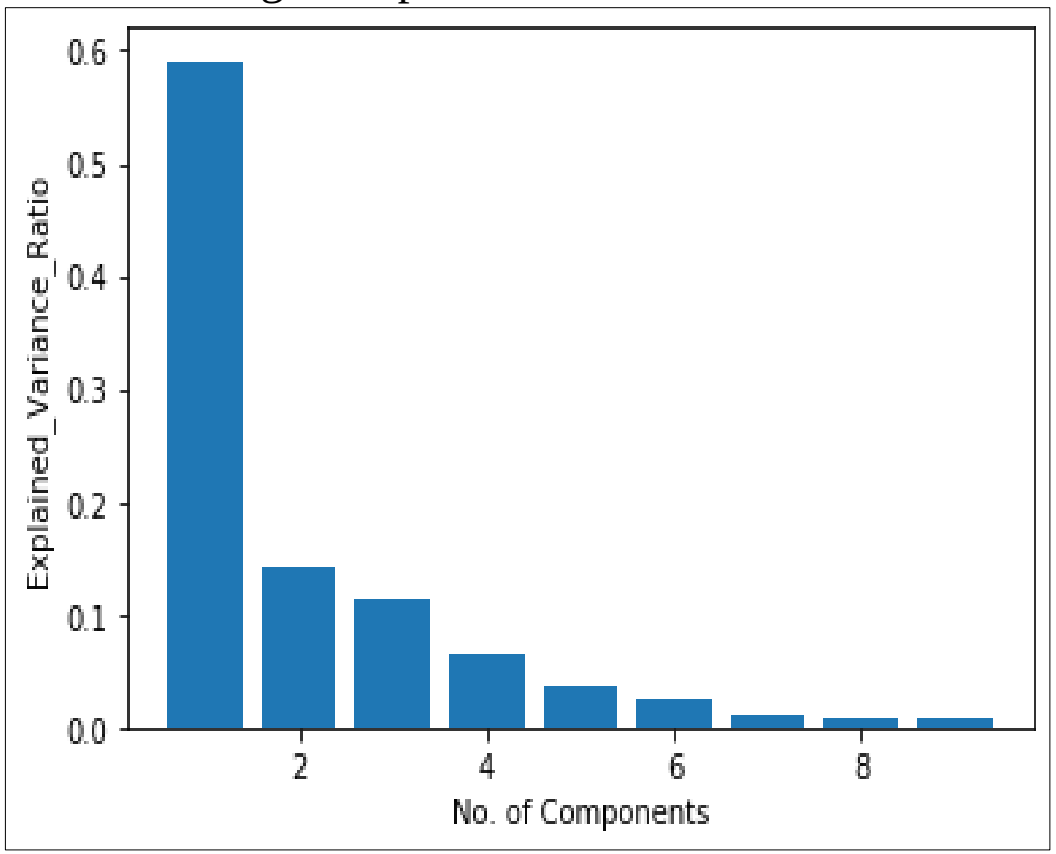
For detecting outliers in data we use boxplot as it shows maximum variance in the data



After Performing PCA :

Checked the explained variance ratio between the PCA components to get the exact number of components.

Fig 2. Explained Variance Ratio



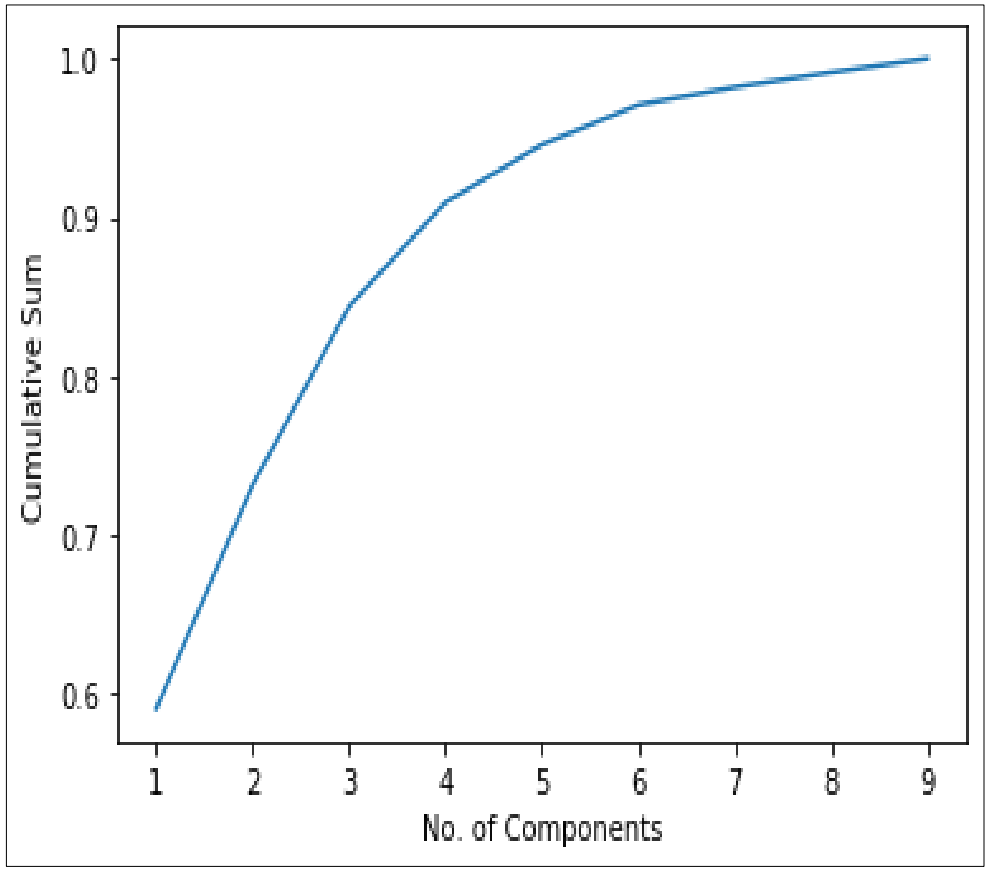


Fig 3. Scree Plot

The above fig 3. Scree plot which helps us understand how many components are to be taken after performing PCA

Detecting Outlier in PCA Column

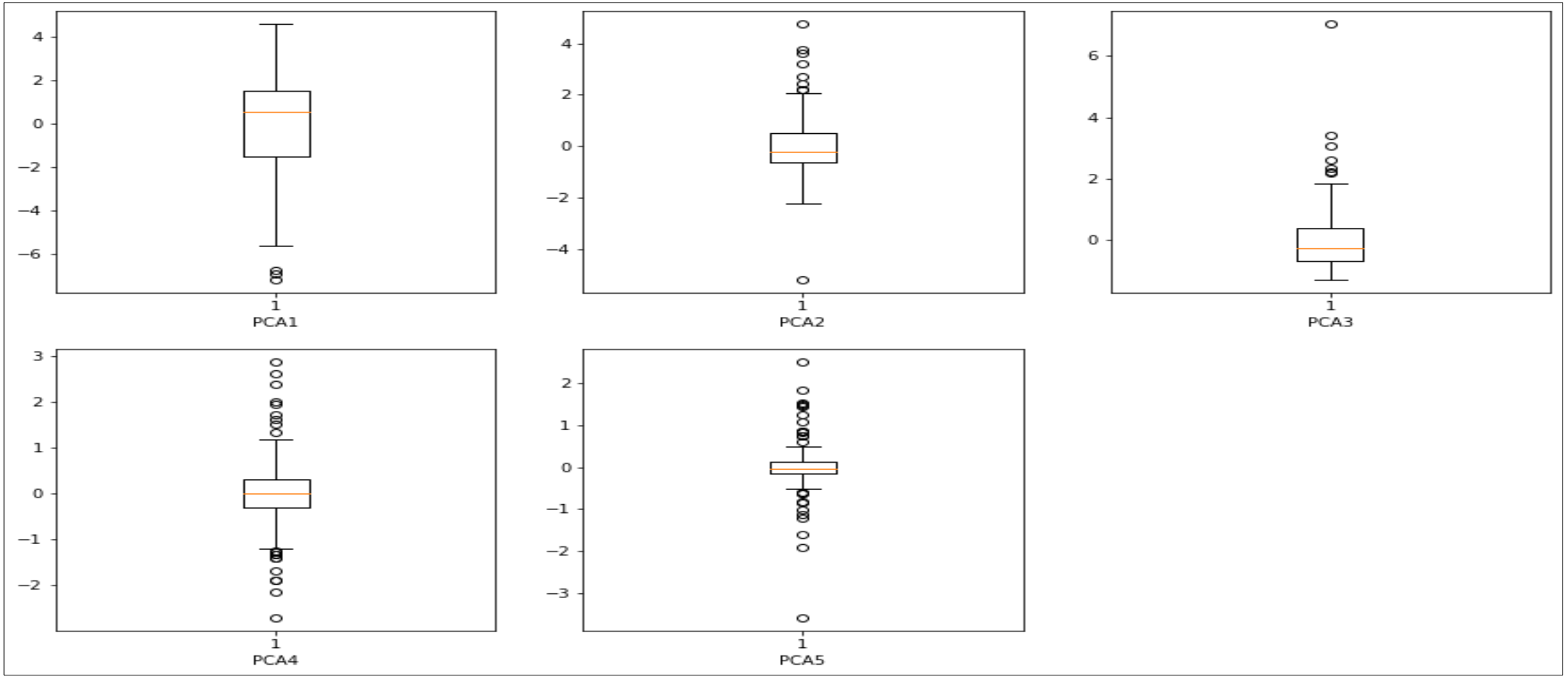
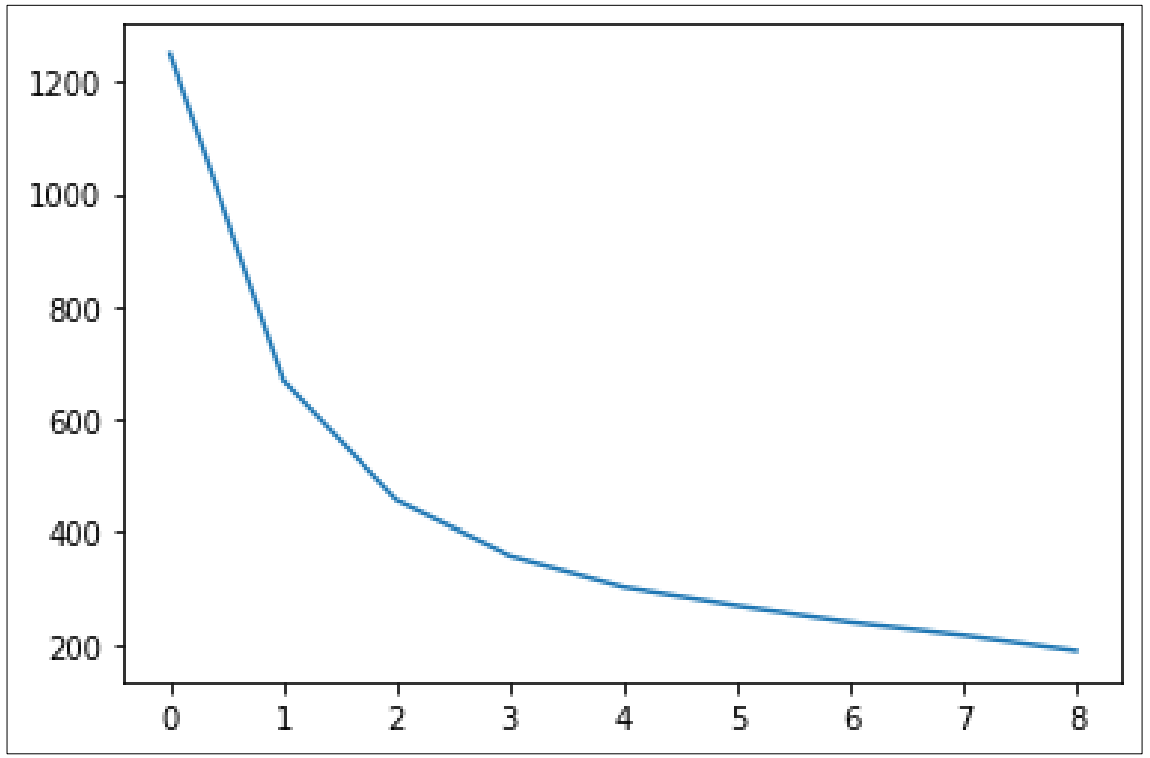


Fig 4. Outlier detection in PCA Columns

Silhouette Score : Is used to choose ‘k’ in K –Means Clustering
Elbow Curve : There is a popular method known as **elbow method** which is used to determine the optimal value of K to perform the K-Means Clustering Algorithm.

Fig 5. Elbow Curve



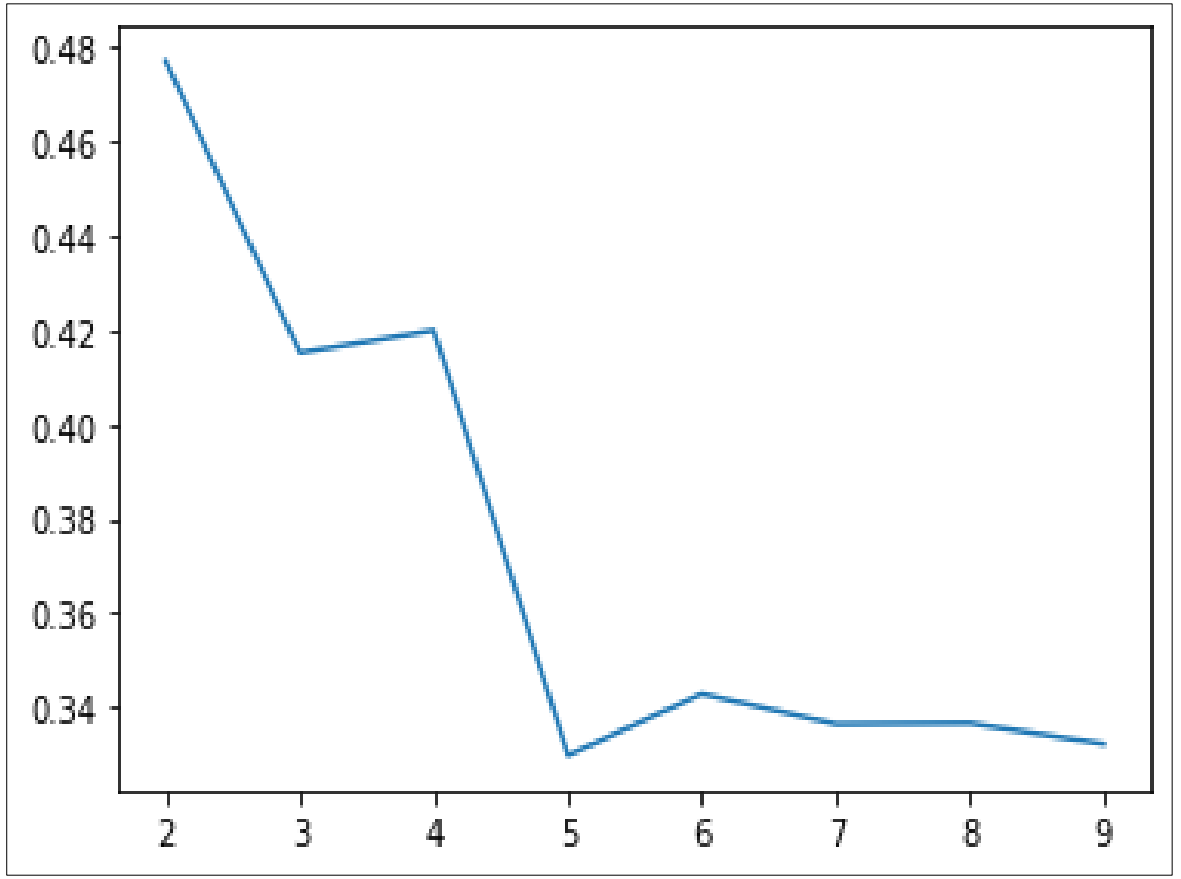


Fig 6. Silhouette Score

After Performing K-Means Clustering plotting the Clusterid wrt PCA components

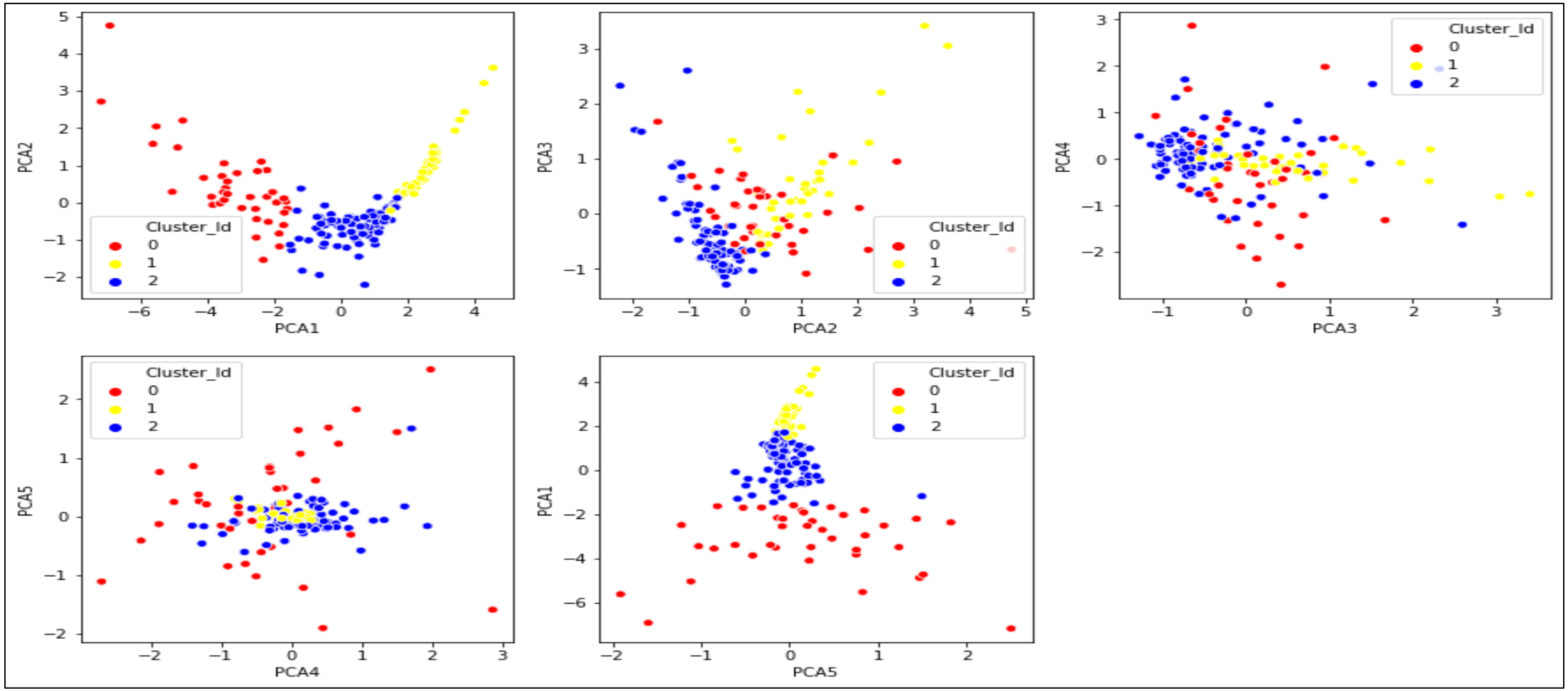


Fig 7. Scatter plot of Cluster Id wrt PCA Components

Plotting Clusters with original Data

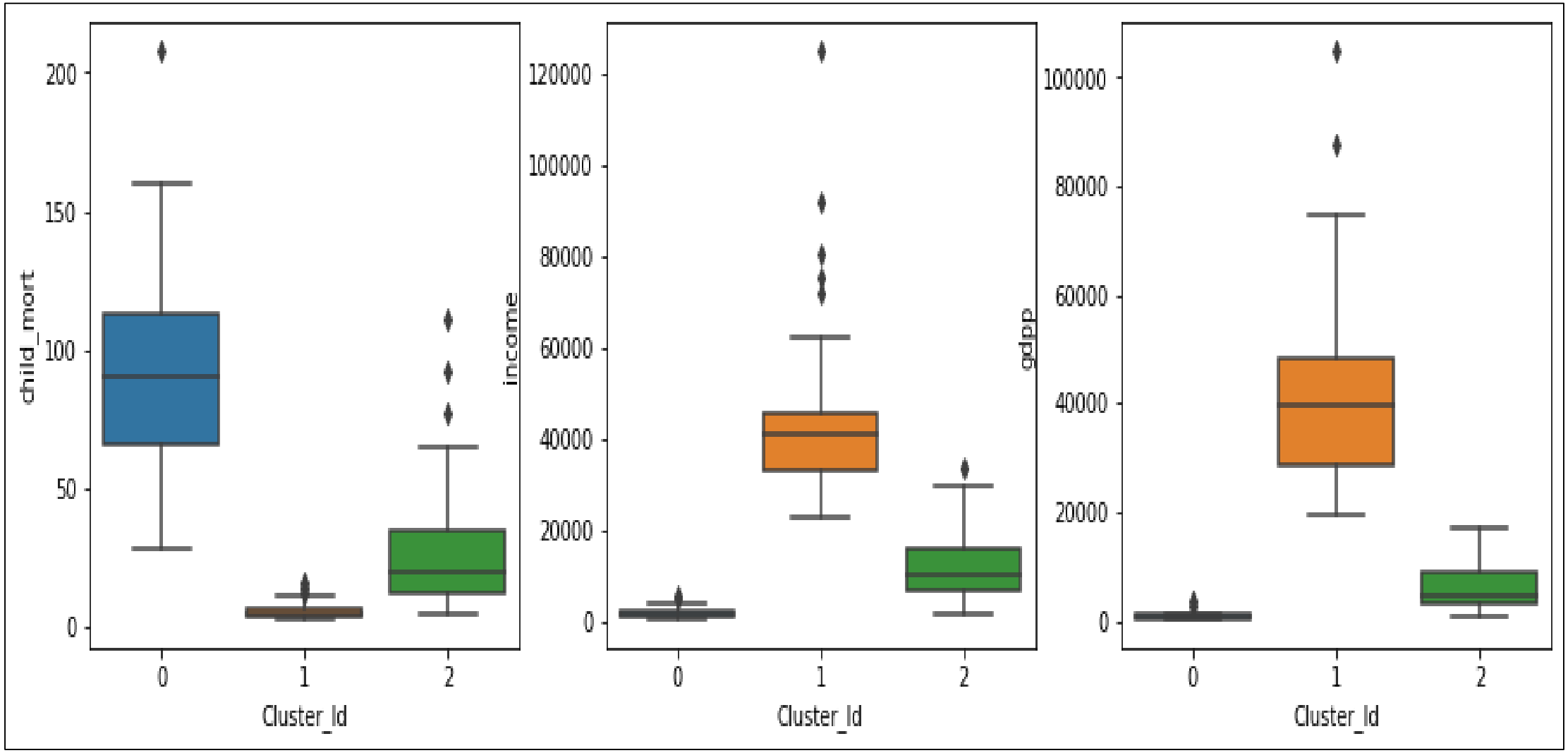


Fig 8. Box plot of Cluster Id wrt Original 3 Important columns

Recommendation :

Here are the top 5 countries according to the clustering done.

This countries should be provided with help

- Haiti
- Sierra Leone
- Chad
- Central African Republic
- Mali

Hierarchical Clustering : 2 Types

1) Single Linkage : Here, the distance between 2 clusters is defined as the shortest distance between points in the two clusters

2) Complete Linkage : Here, the distance between 2 clusters is defined as the maximum distance between any 2 points in the clusters

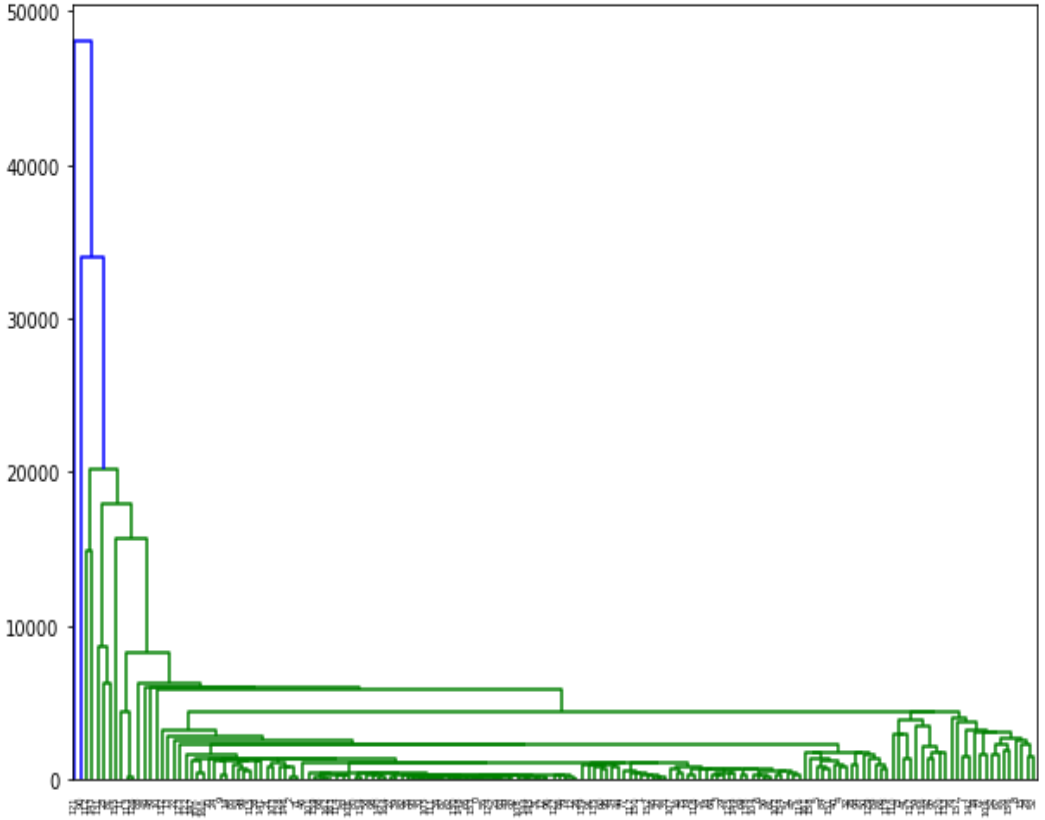


Fig 9. Single Linkage

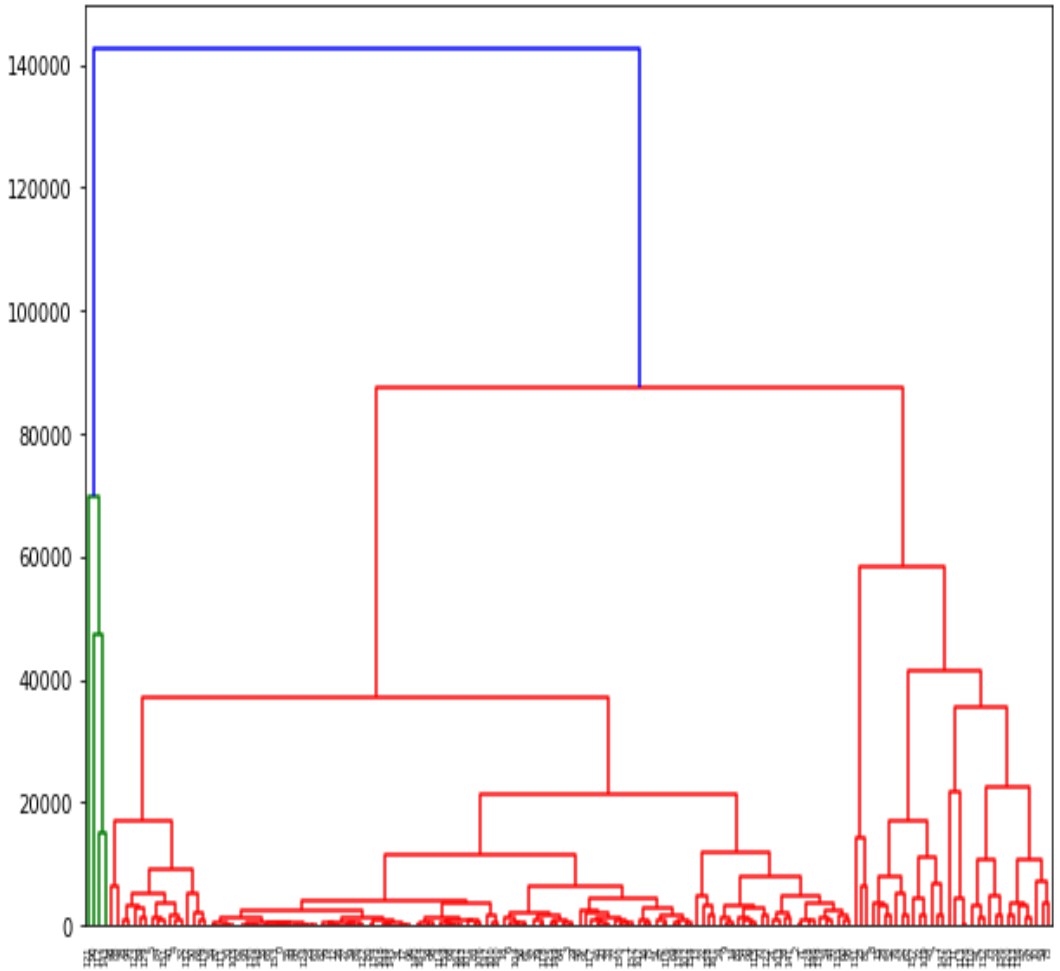


Fig 10. Complete Linkage

Plotting graph on Clusters and Original data after performing Hierarchical Clustering