Question 1: Assignment Summary

- In EDA, the summary of data was shown along with graphical representation.
- Later converted the exports, imports and health feature to the values as it was earlier represented in terms of percentage
- Univariate and bivariate analysis to understand the correlation between the features of the data
- Then the outlier treatment is done to remove the column with higher gdpp and income.
- Later carried out Hopkins statistics to understand if caring out clustering is a good option. The output of Hopkins range between 87%-95% which is very good for clustering
- Standardised scaling of data
- Silhouette score & Elbow curve to understand K value in K-means which was later selected as K = 3
- Clustering the data with labels 0,1,2.
- Representing the data in graphical form to understand the clusters well
- Cluster profiling to understand which cluster is in direst need for funding
- Top 5 direst need countries are Sierra Leone, Chad, Central African Republic, Mali according to K-means
- Performing single and complete linkage for hierarchical clustering
- Defining cluster value to 3
- Performing graphical representations for the clusters to analyse which cluster is with low income, gdpp group and high mortality rate
- Top 5 direst need countries are Sierra Leone, Chad, Central African Republic, Mali according to Hierarchical clustering similar to K-means.