1. Assignment Summary:
    a. I am a data analyst working for an NGO, to identify top 5 countries most in need of aid from a given list of about 200 countries, along with their data like income, GDP, health expenditure per person, child mortality, etc. Using these country metrics, I have to perform analysis and clustering, and then present my solution to the CEO of the NGO.
    b. On analysis, some observations were-
        - Child mortality decreased, life expectancy and health expenditure increased as income per person increased.
        - For total fertility lower than 3, child mortality was low and health expenditure was good. After 3, child mortality starts increasing, and health expenditure keeps dropping.
        - Countries with low trade deficit were also the ones with high GDP and high income. The countries with high trade deficits were low income countries.
    c. Removed some higher range income and GDP countries from dataset. Couldn't remove lower range outliers because those are the countries most in need of aid.
    d. Scaled my data using Standard Scaling, than calculated Hopkins score, which came out to be more than 95%.
    e. K-Means:
        - Calculated optimal number of clusters using both silhouette and elbow method, both suggested value of 3.
        - Performed clustering using k=3.
        - First cluster- countries with high mortality rates, low income and low GDP.
        - Second cluster- countries with medium mortality rates, medium income and medium GDP.
        - Third cluster- countries with low mortality rates, high income and high GDP.
        - All our countries of interest in first cluster.
    f. Hierarchical:
        - Also performed clustering using both single and complete linkage. Both gave similar results.
        - Clusters were poorly formed, but got much the same result as with K-means.
    g. K-means a better clustering method than hierarchical here. Clusters were better formed and easier to read.

2. Compare and contrast K-means Clustering and Hierarchical Clustering.
    a. Number of clusters is needed to be defined beforehand in K-means clustering. Hierarchical clustering will give us a dendrogram of the dataset, and then we can choose based on this dendrogram, how many clusters we want to divide the data in.
    b. K-means clustering will work very well with large amount of data, take less processing power. Hierarchical clustering will take a lot of time to work with the huge datasets usually seen in industry, and takes a lot of processing power to run, because it is a linear algorithm.

3. Briefly explain the steps of the K-means clustering algorithm.

a. First, the algorithm will divide the total data into the number of clusters specified in the code, randomly.
b. Then, it will calculate the mean of all the points in the cluster, and mark that as the center of that particular cluster.
c. Next, it will calculate the distances of all the points from all the cluster points, and re-assign each point to its closest center. Thus, now our clusters have changed from initial position.
d. Next, it will again re-calculate the centers of the clusters, and then re-assign the points to their closest clusters.
e. This process will continue till the centers become constant, or the code has run the maximum times specified. These are our final clusters.

4. How is the value of 'k' chosen in K-means clustering? Explain both the statistical as well as the business aspect of it.
   a. For choosing the value of K-means, we have two ways- silhouette method and elbow score.
   b. When using the silhouette method, we can plot the silhouette scores of different values of K on a line chart, and then choose the K where the silhouette score is maximum.
   c. Elbow score, we choose the value at which the above kind of graph forms an elbow shape, or takes a somewhat sharp curve.
   d. Usually, we will liase with the client to determine the number of clusters that should be taken, because they have the domain knowledge. If our silhouette and elbow curve show 3 as a good number of clusters, but our client asks us to go with 5 clusters, we will form 5 clusters on the data.
   e. An ideal cluster value is between 2-8, because more than these many clusters become difficult to follow up on and manage.

5. Explain the necessity for scaling/standardisation before performing Clustering.
   a. Scaling is performed to bring all our variables to the same scale.
   b. Suppose we have 5 columns in a housing dataset, one of them being number of rooms, and another being area in sq feet. The rooms are going to be in integers up till 5, but area is going to be in thousands.
   c. In such a case, if we perform clustering without scaling, the clustering algorithm will end up giving more importance to certain variables, and lesser importance to other ones, which we don't want.
   d. Scaling will avoid this scenario, and make sure that clustering algorithm gives equal importance to all variables.

6. Explain the different linkages used in Hierarchical Clustering.
   a. Single linkage- the distance between two clusters is defined as the shortest distance between two points in their respective clusters.

b. Complete linkage- the distance between two clusters (inter cluster distance) is defined as the max distance between any two farthest points from each other in two clusters. Complete linkage gives tighter clustering than single linkage.
c. Average linkage- the inter cluster distance is defined as the average of the distances of each point in one cluster to every point in the other cluster.