**Assignment – 6**

**Name: Chethan Kacham**

**Username: CXK34890**

**Student ID: 700743489**

**Question - 1:**

I have done all the calculations and plotted the dendrograms in question1.docx file. Please refer to the question1.docx/question1.pdf file for question 1.

For first question, I have performed all the mathematical calculations for Single link, Complete link and Average link. Plotted the three dendrograms as shown below

For the Single link, I have used minimum distance to do the calculations and plot the dendrogram.

For the Complete link, I have used maximum distance to do the calculations and plot the dendrogram.

For the Average link, I have used average distance to do the calculations and plot the dendrogram.

Dendrogram of Single link:

Chart, box and whisker chart

Description automatically generated

Dendrogram of Complete link:

Chart, box and whisker chart

Description automatically generated

Dendrogram of Average link:

Chart, box and whisker chart

Description automatically generated

**Question - 2:**

In the second question as explained in the class the followed the below steps

* Imported all the required modules and suppressed warnings.
* Read the data file and replaced the null values with mean and categorized the columns.
* Later applied StandardScaler and normalize functions.
* Used PCA(n=2) to reduce the input dimensions to two features.
* Applied Agglomerative Clustering with k=2,3,4 and 5 on finaldf.
* Visualized the results of Agglomerative clustering for k=2,3,4 and 5 using scatter plot.
* Evaluated the Silhouette scores for each k value i.e., 2,3,4 and 5 and plotted the results in a bar graph with respective to number of clusters.

I have commented the tasks as a heading in the code which are given to the question 2.

Adding the screenshots of the code and output for Question - 2

Code & Output:

Graphical user interface, text, application

Description automatically generated

Imported all the required modules and suppressed warnings, read the data file CC GENERAL.csv and replaced the null values with mean and categorized the columns in the above screenshot.

Text

Description automatically generated

Applied StandardScaler() function and normalize() function to the input data and then applied PCA(components=2) for reducing the data to two features in the above screenshot.

Chart, scatter chart

Description automatically generated

Chart, scatter chart

Description automatically generated

Chart, scatter chart

Description automatically generated

Chart, scatter chart

Description automatically generated

Applied Agglomerative Clustering with k=2,3,4 and 5 on finaldf and visualized the results using scatter plot in the above 4 screenshots.

Graphical user interface, text, application

Description automatically generated

Evaluated the Silhouette Scores for k=2,3,4 and 5 in the above screenshot.

Chart, bar chart

Description automatically generated

Visualized the Silhouette Scores for different number of clusters using bar graph in the above screenshot.

In GitHub Machine Learning repository, we have an Assignment 6 folder where I have uploaded document(Word & PDF) for question 1 and I have uploaded the code for question 2 in question2.ipynb file. I have uploaded screenshots of dendrograms of single link, complete link and average link for question1 and screenshots of code and output for question 2 in Screenshots folder.

As I have explained the code for question 2, I haven’t recorded any video for this assignment. I have uploaded word and pdf document of Assignment-6 in Word & PDF file of Assignment 6. The following is the link of the GitHub repository

**GitHub link:** <https://github.com/ChethanKacham/MachineLearning.git>

**Video link:** Explained the code in the class