**ASSIGNMENT6**

**RISHITHA REDDY NALLAKALVA 700742428**

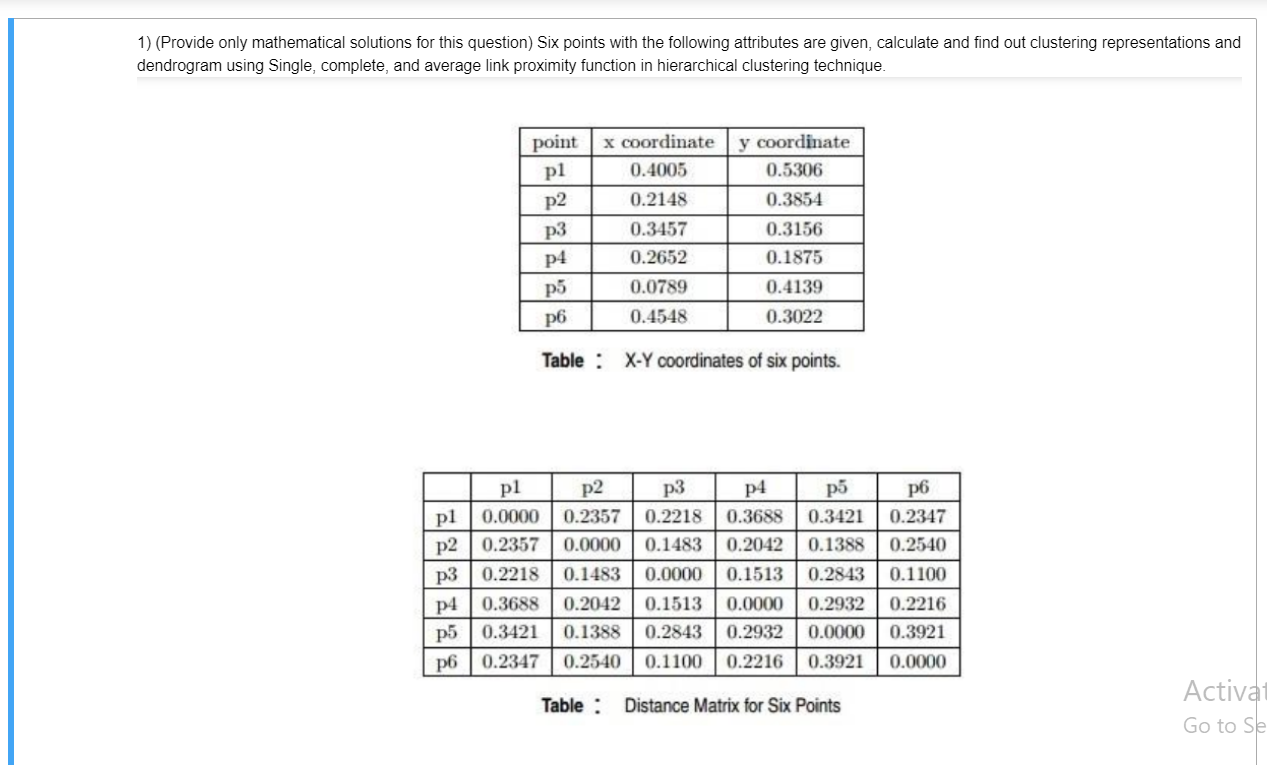
**Source code:**

[**https://github.com/rishithareddy3339/rishi\_assignment/blob/master/ML\_ASS6\_700742428.ipynb**](https://github.com/rishithareddy3339/rishi_assignment/blob/master/ML_ASS6_700742428.ipynb)

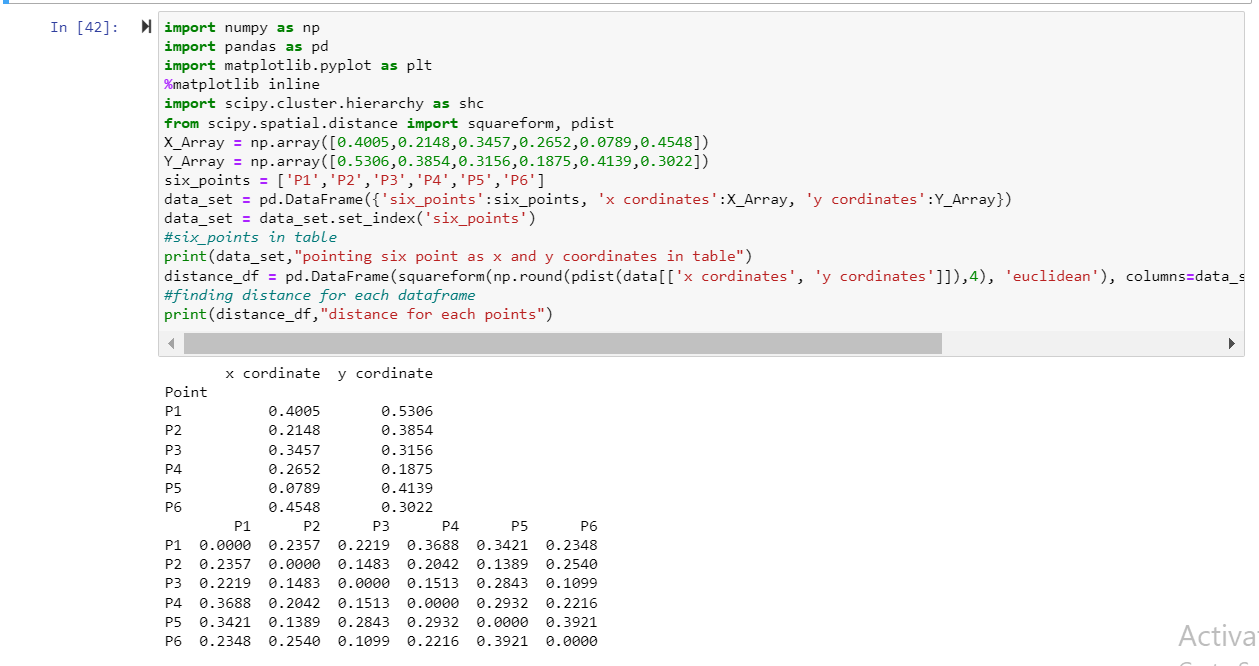
**video recording:**

[**https://github.com/rishithareddy3339/rishi\_assignment/blob/master/Untitled19%20-%20Jupyter%20Notebook%20-%20Google%20Chrome%202022-11-15%2023-40-56.rar**](https://github.com/rishithareddy3339/rishi_assignment/blob/master/Untitled19%20-%20Jupyter%20Notebook%20-%20Google%20Chrome%202022-11-15%2023-40-56.rar)

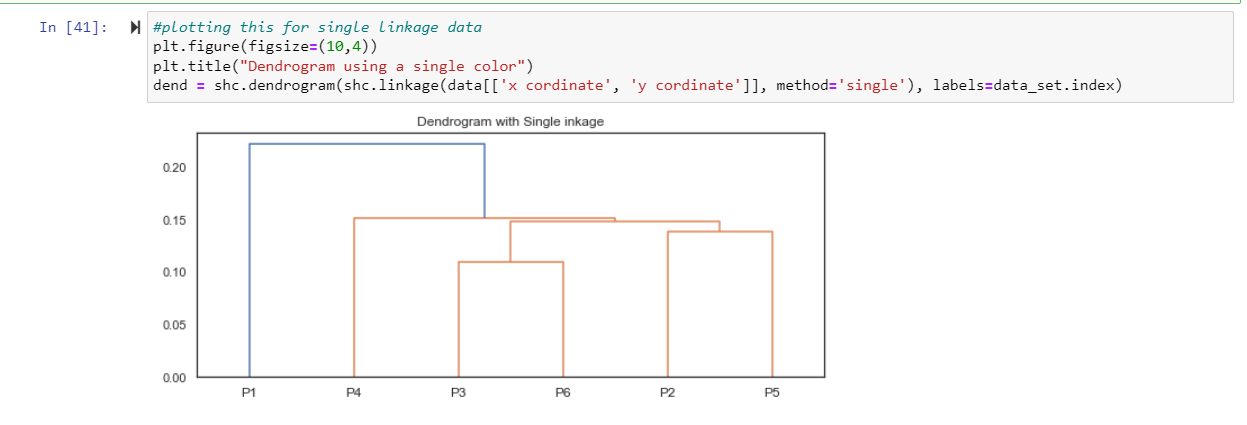
1. (Provide only mathematical solutions for this question) Six points with the following attributes are given, calculate and find out clustering representations and dendrogram using Single, complete, and average link proximity function in hierarchical clustering technique.

****

• In Single Linkage, the separation between two clusters corresponds to the shortest separation between their constituent members.

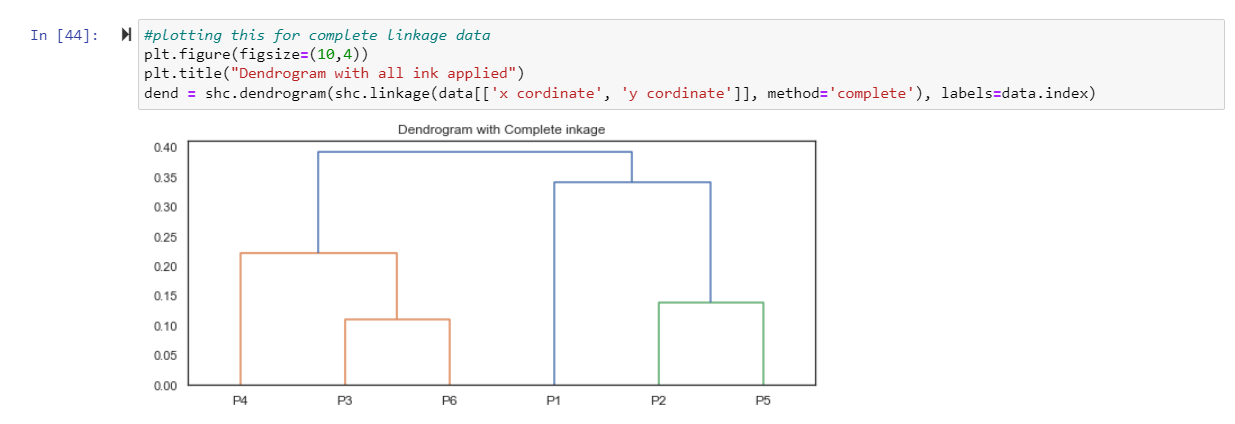
****

**plotting this for single linkage data**

****

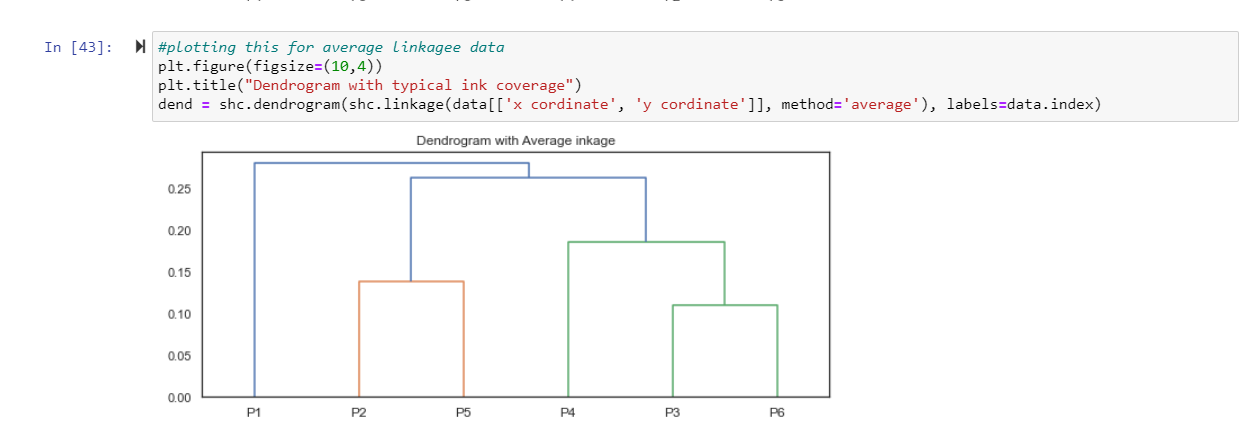
**plotting this for complete linkage data**

**In Complete Linkage, the separation between two clusters is equal to the greatest separation between their individual members.**

****

**#plotting this for average linkage data**

**The distance between two clusters in average linkage is equal to the average of all distances between its members.**

****

**2) Use CC\_GENERAL.csv given in the folder and apply:**

**a) Preprocess the data by removing the categorical column and filling the missing values.**

**b) Apply StandardScaler() and normalize() functions to scale and normalize raw input data.**

**c) Use PCA with K=2 to reduce the input dimensions to two features.**

**d) Apply Agglomerative Clustering with k=2,3,4 and 5 on reduced features and visualize**

**result for each k value using scatter plot.**

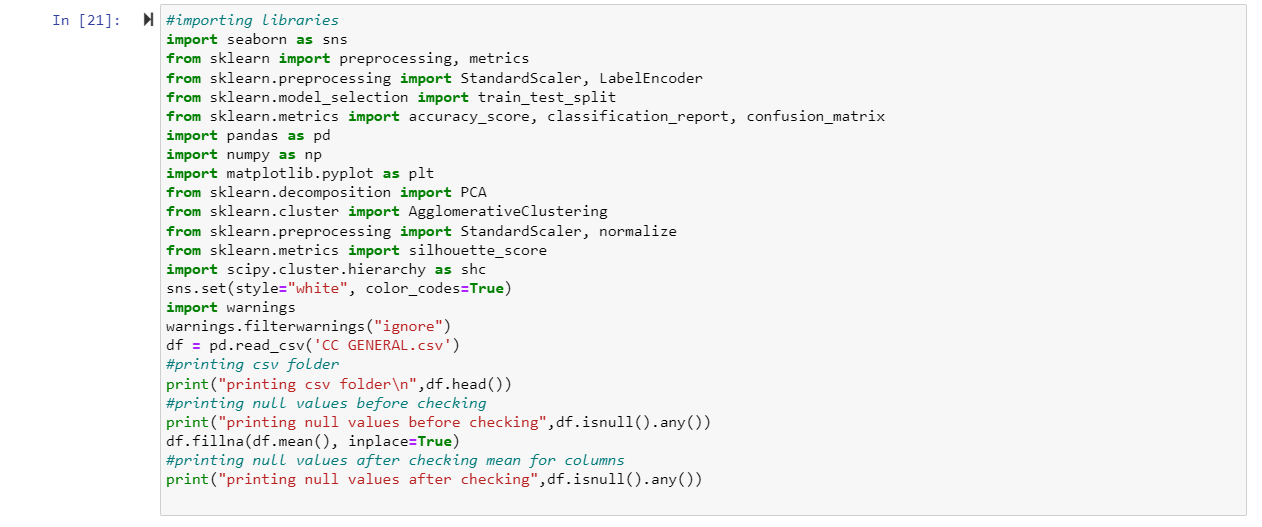
**e) Evaluate different variations using Silhouette Scores and Visualize results with a bar chart.**

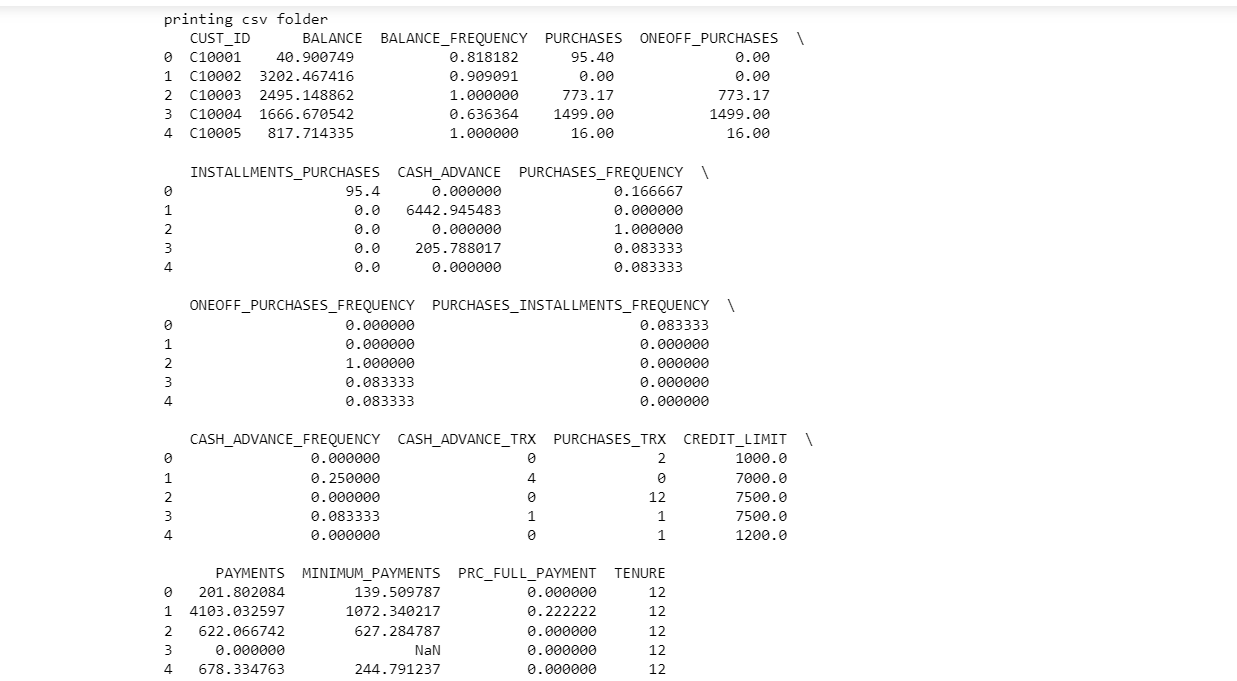
**Importing libraries for graphs and some using libraries**

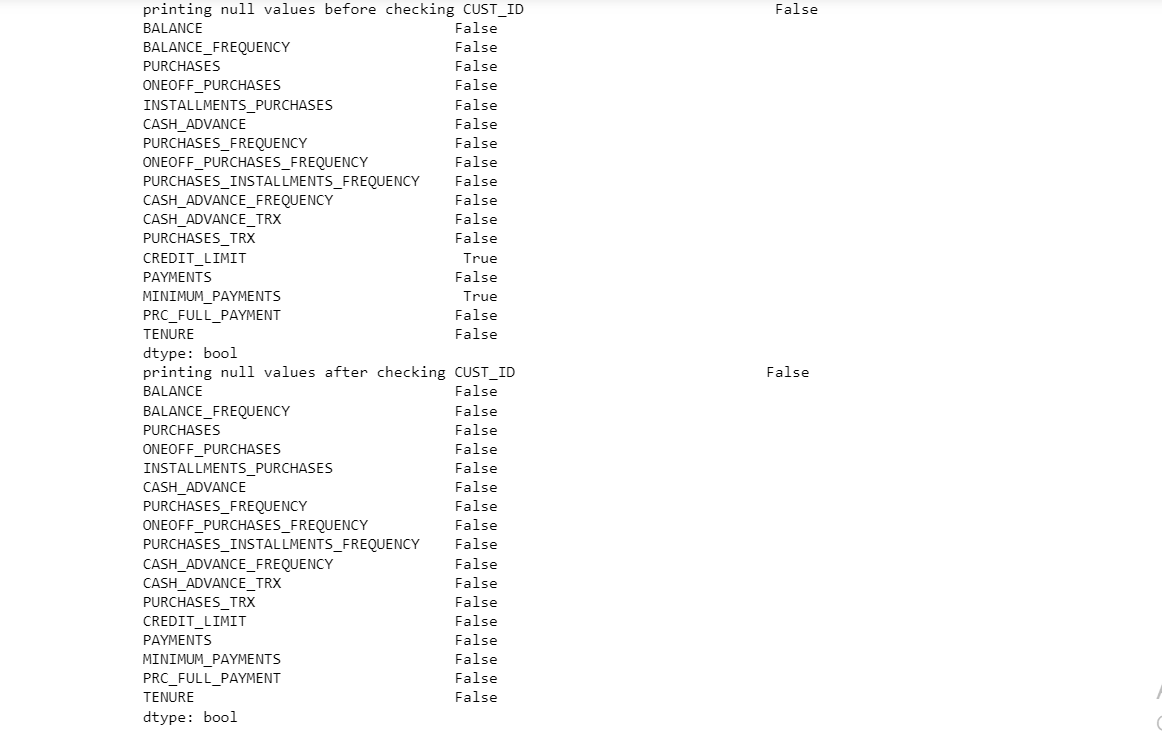
**printing csv folder**

**printing null values before checking**

**printing null values after checking mean for columns**

****

****

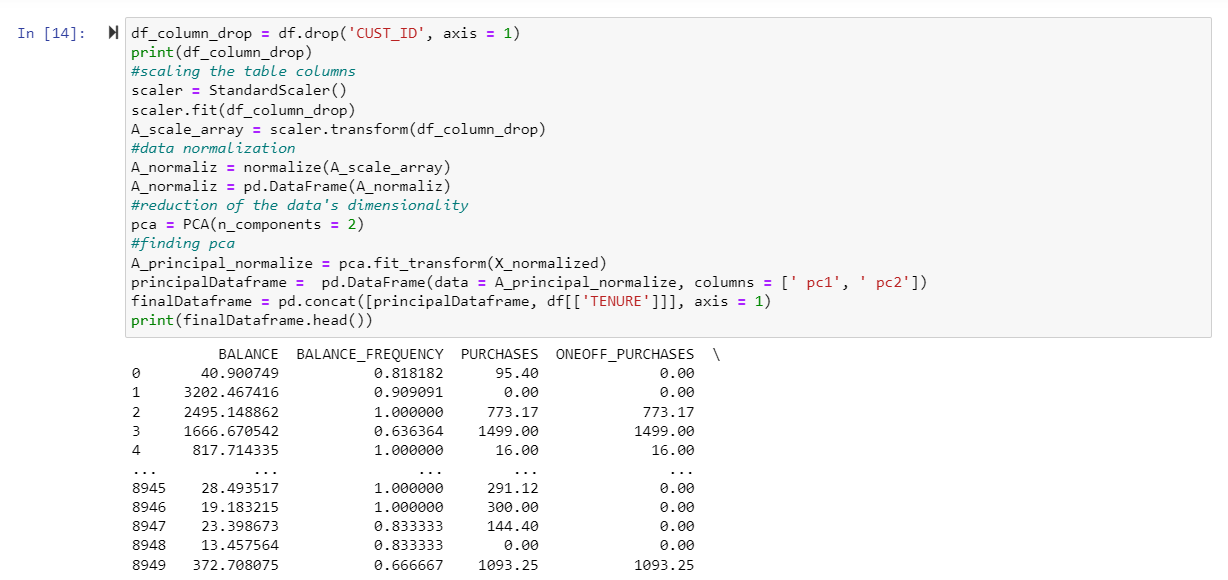
****

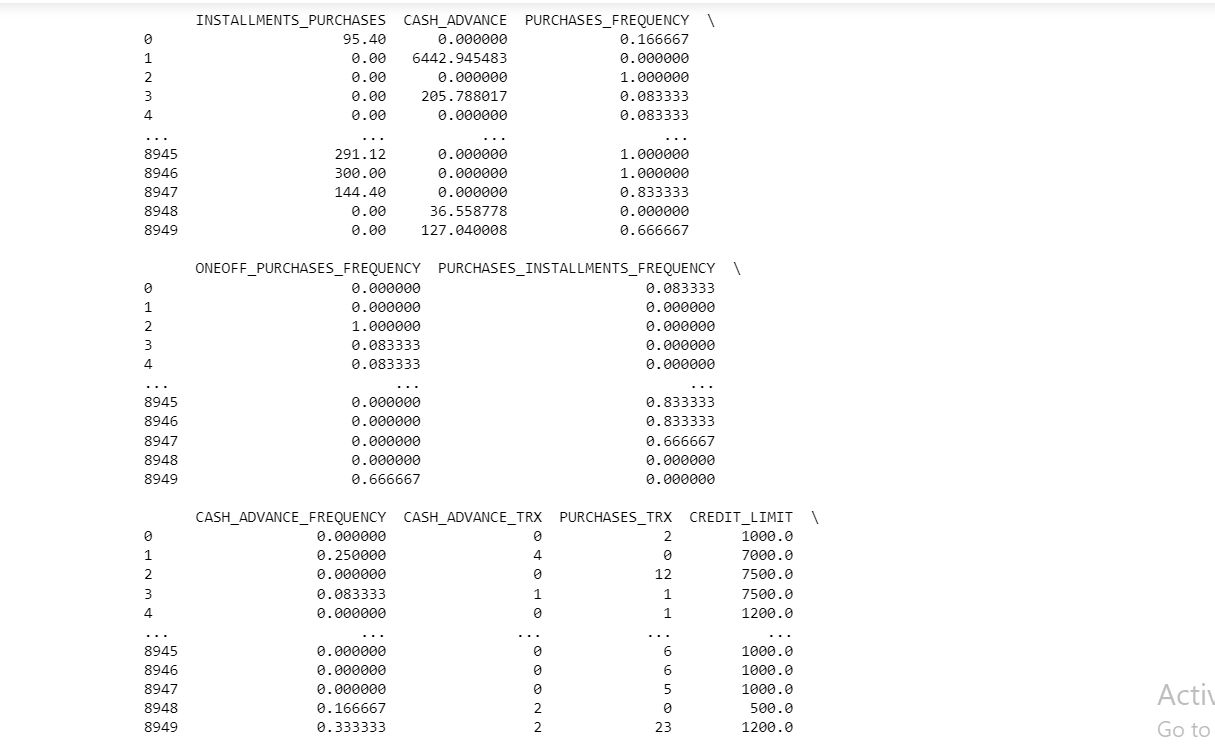
**scaling the table columns**

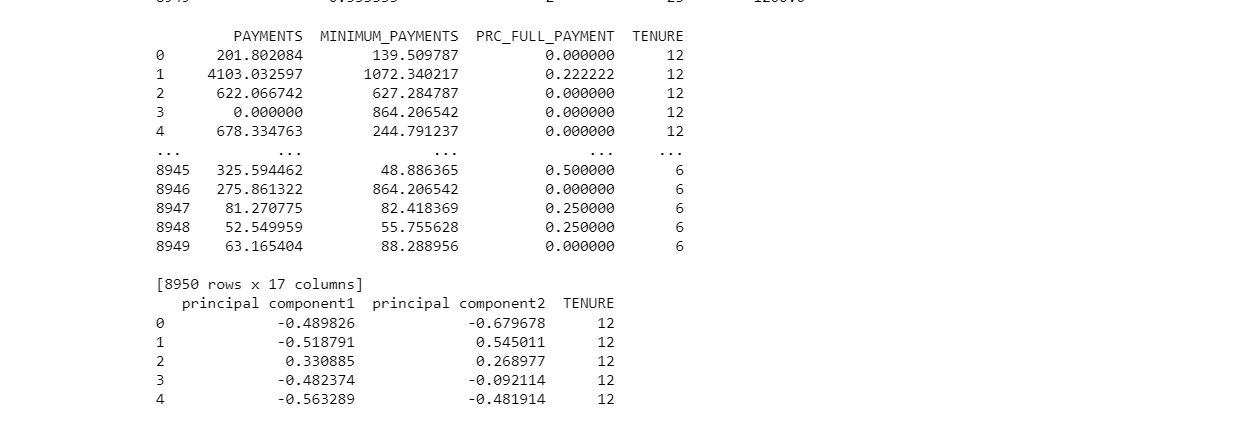
**data normalization**

**reduction of the data's dimensionality**

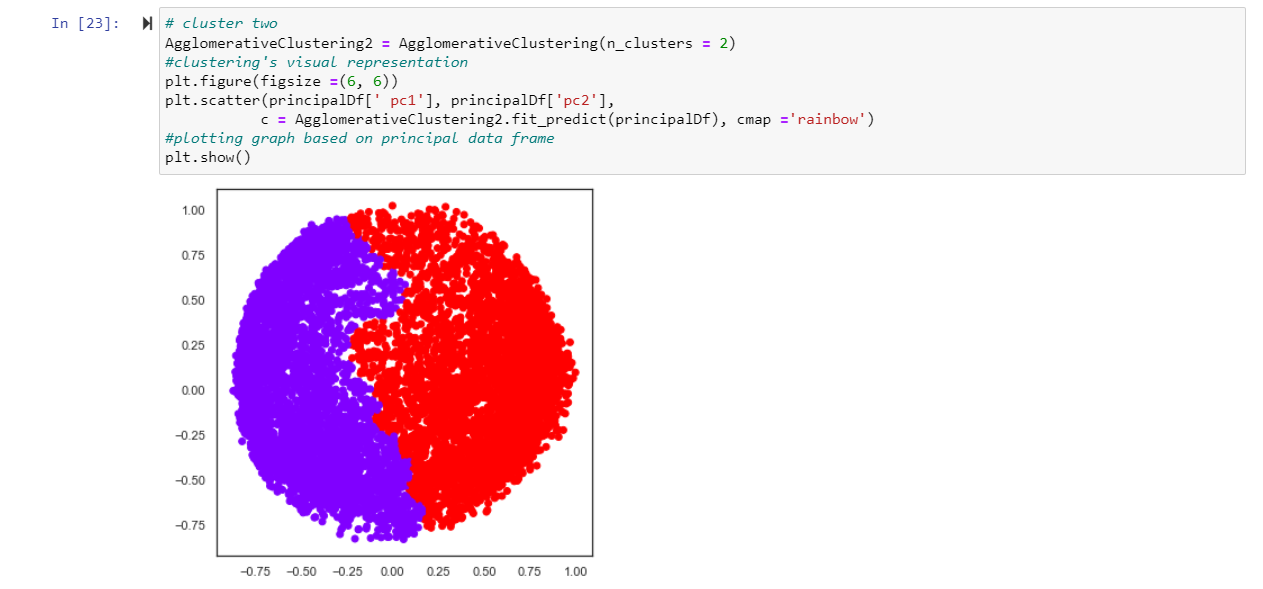
**finding pca**

****

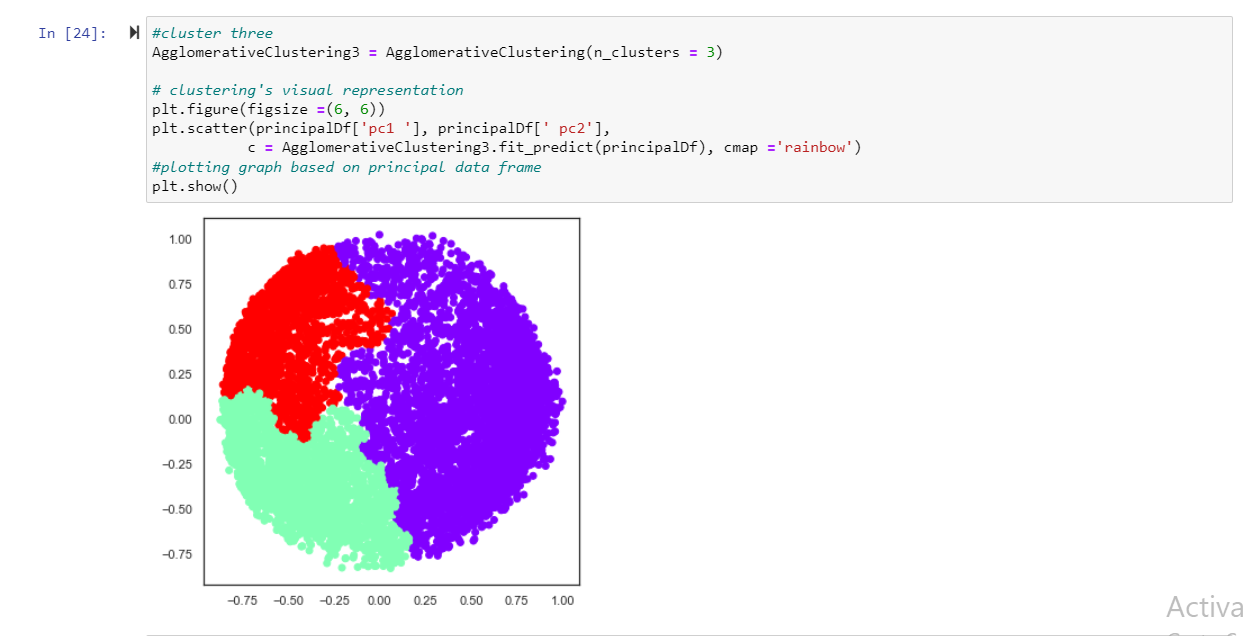
****

****

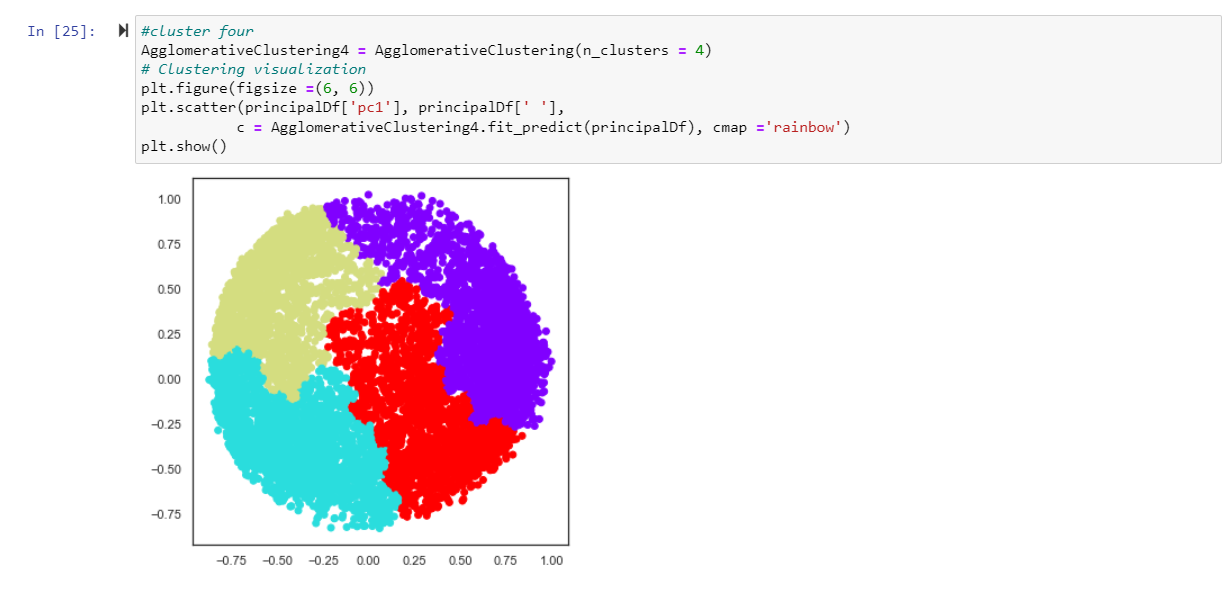
**clustering's visual representation for cluster 1**

****

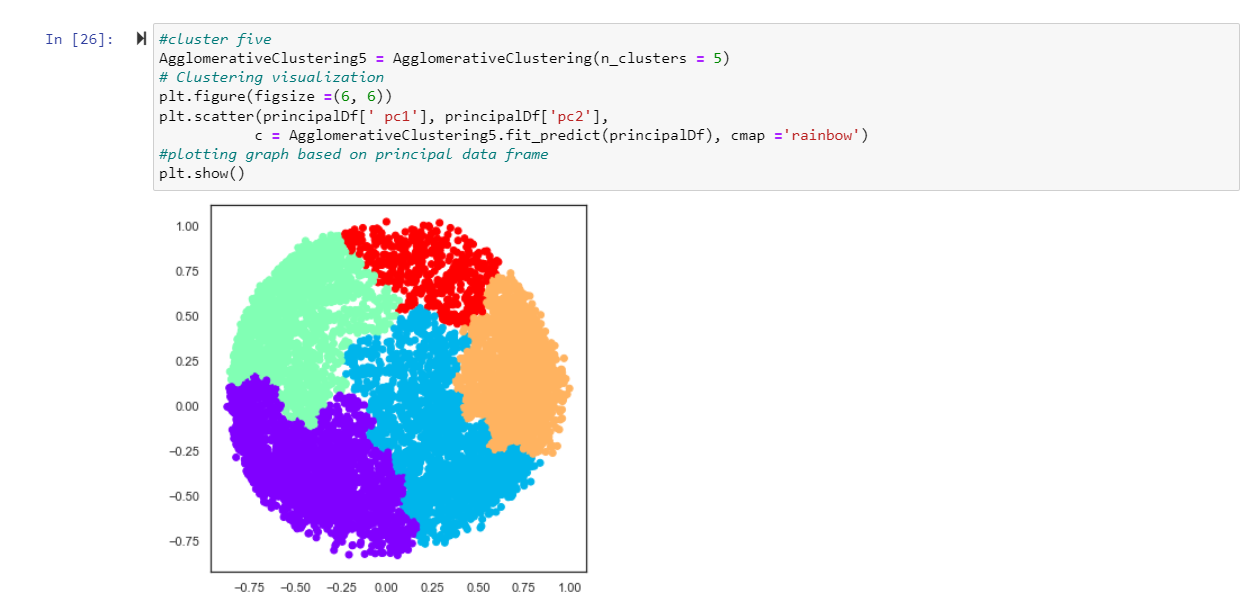
**clustering's visual representation for clusterthree**

****

**clustering's visual representation for cluster four**

****

**clustering's visual representation for cluster five**

****

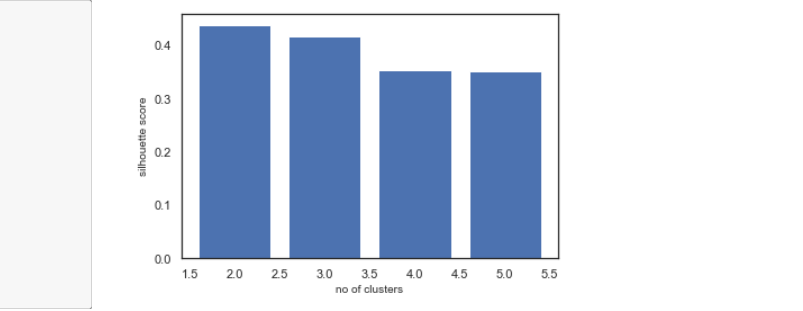
**appending clusters to principal dataframe**

**visualizing the results on bar chart**

****

**making a bar chart to compare the outcomes**

**plotting clusters and silhouette scores by using bar graph**

****