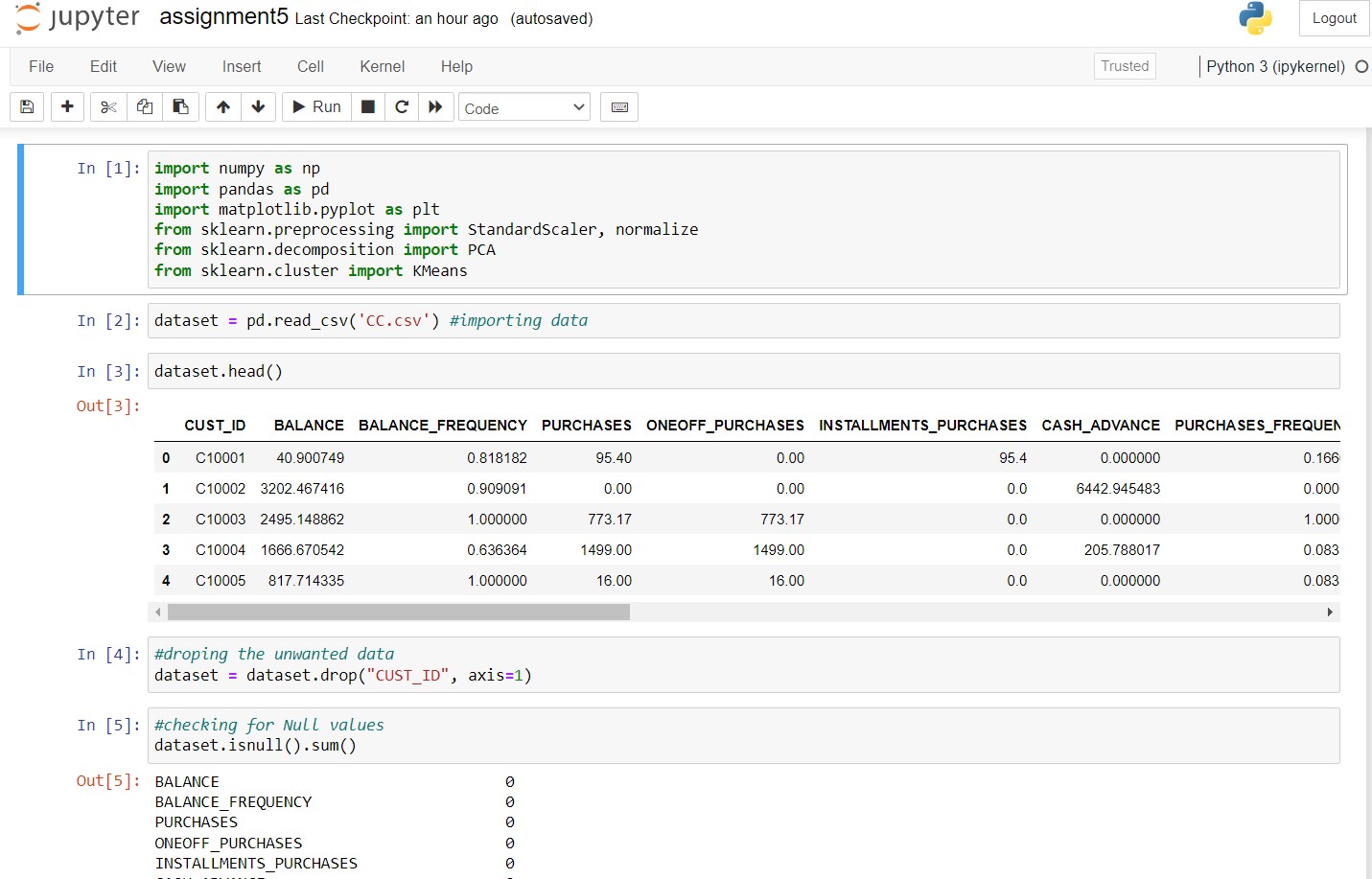
# Assignment 5

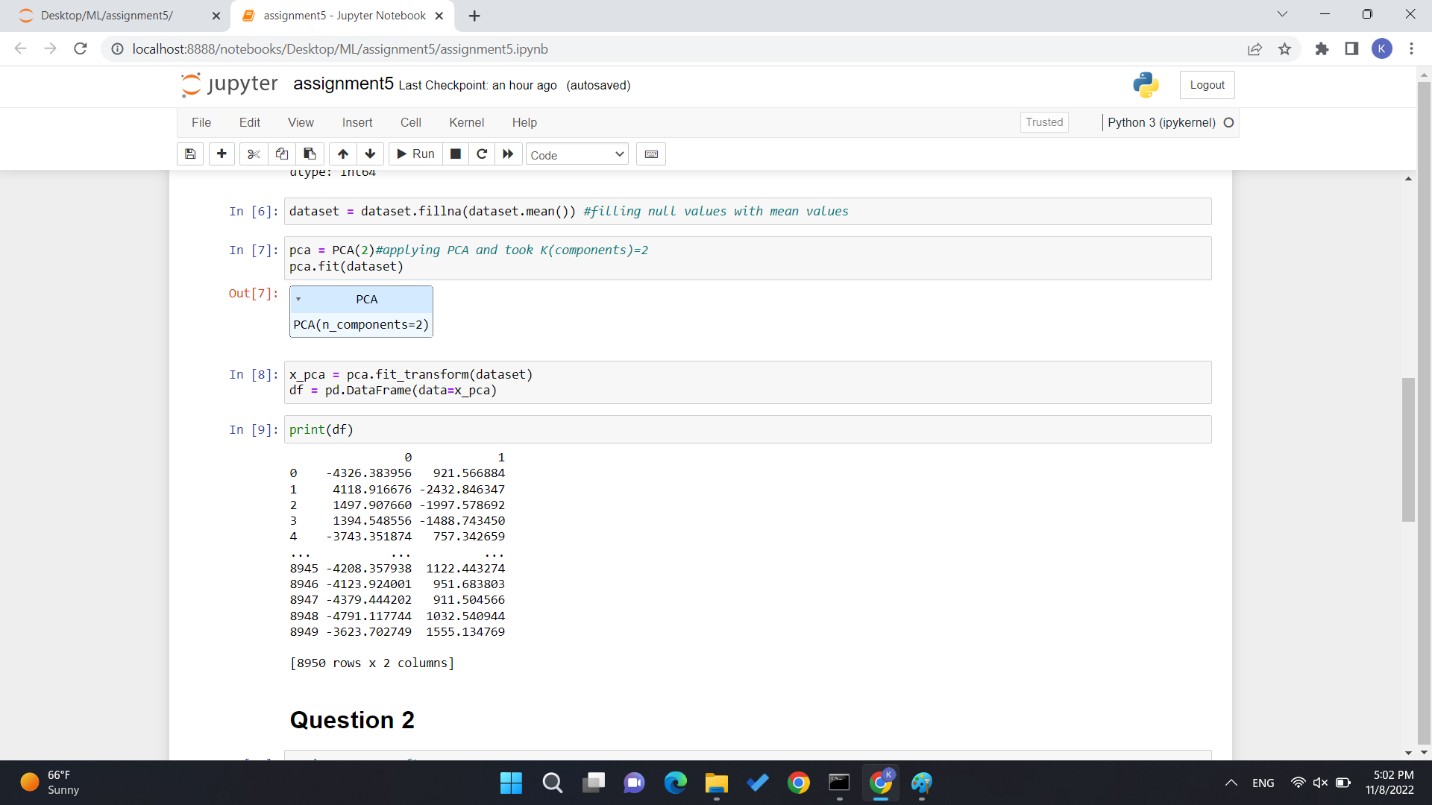
# Pallavi Harishchandra Meher

# 700727681

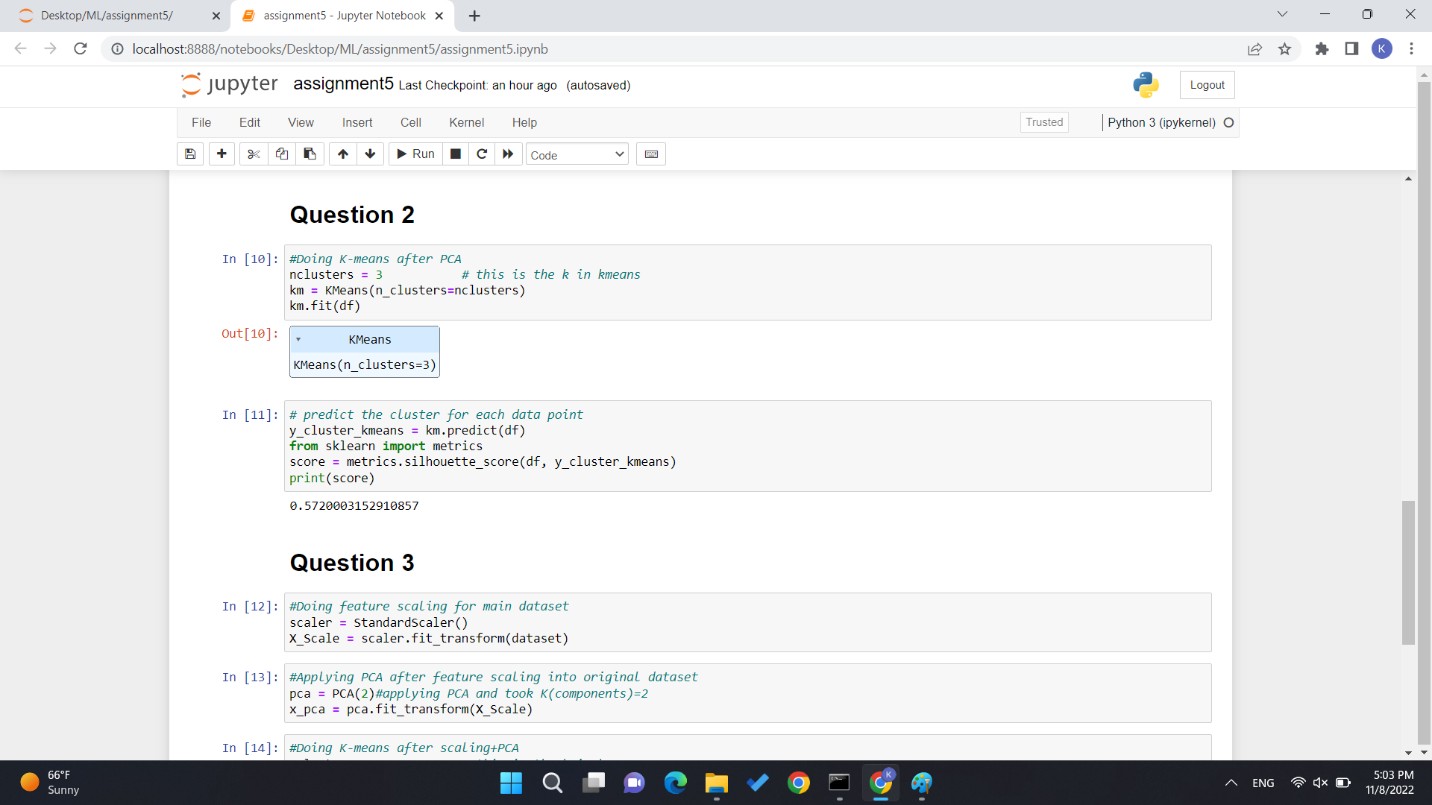
# Question 1:



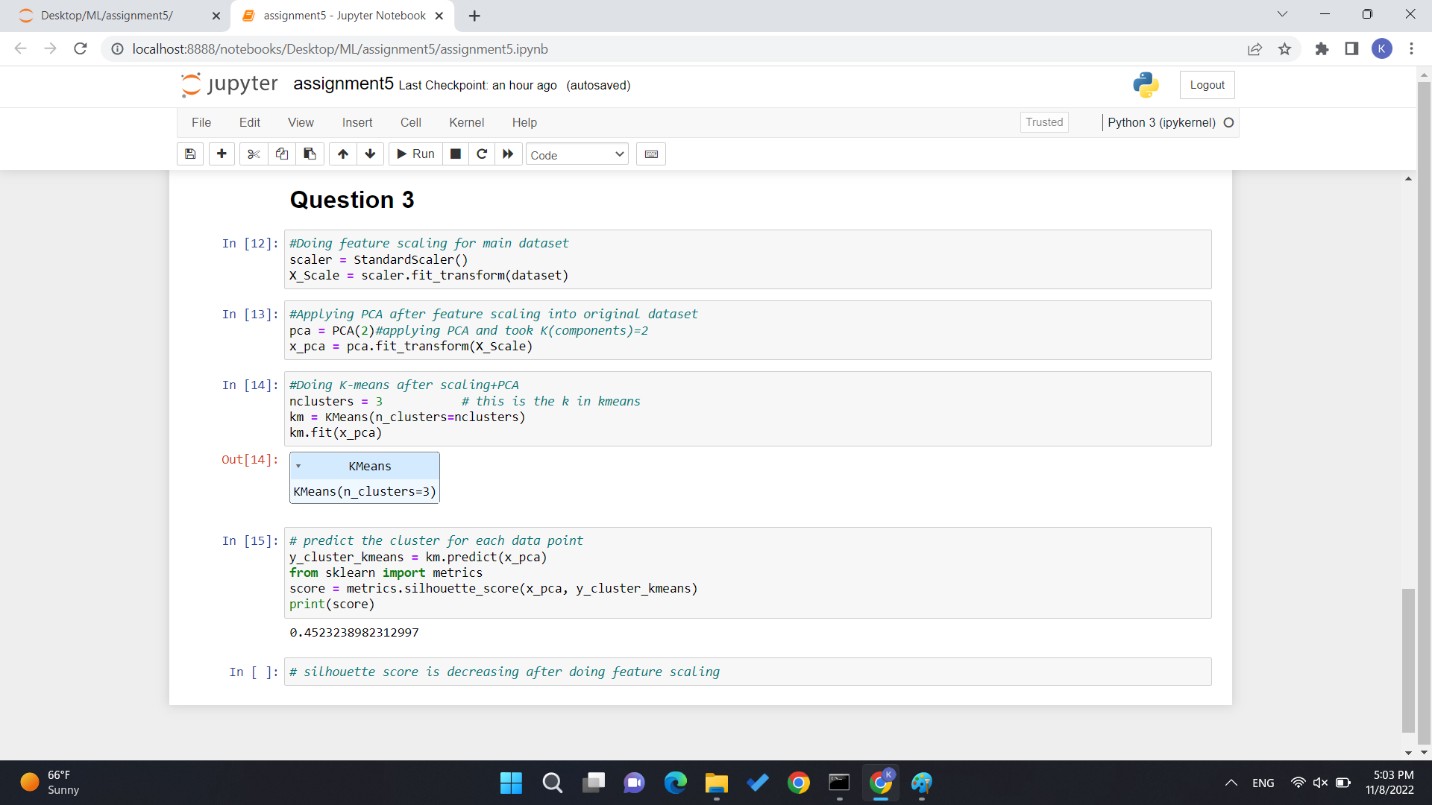
First, I have imported the required libraries then imported the dataset “cc.csv”. After that displayed the first few rows of the dataset. I have dropped the CUST\_ID column from the dataset. And then checked if there are any null values in the dataset.



In above screenshot I have filled the null values with the mean values of the column. After that applied the PCA for components =2 and printed the output. This is the question 1.



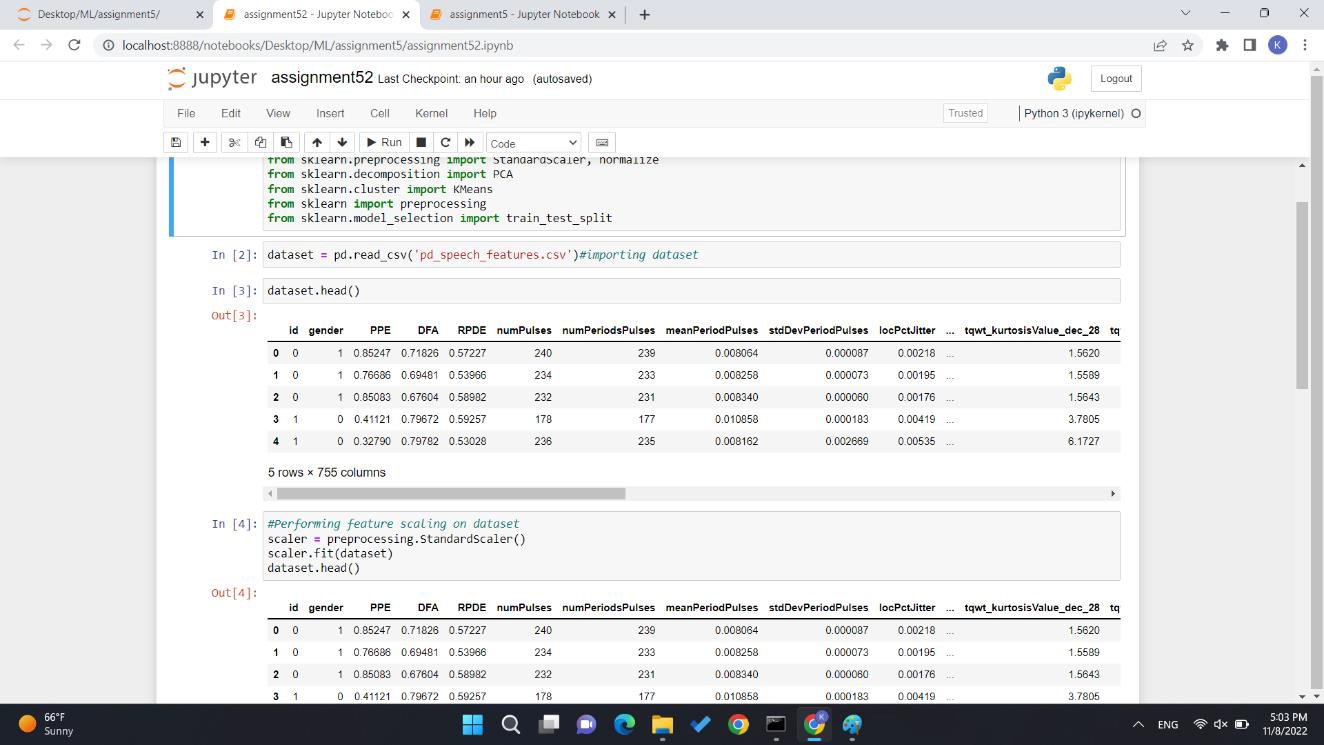
For question 1.2 first I have applied the k-means (clusters = 3) on the output from question 1. Then predicted the number of clusters for each datapoints and computed the silhouette score. Which is 0.572000.



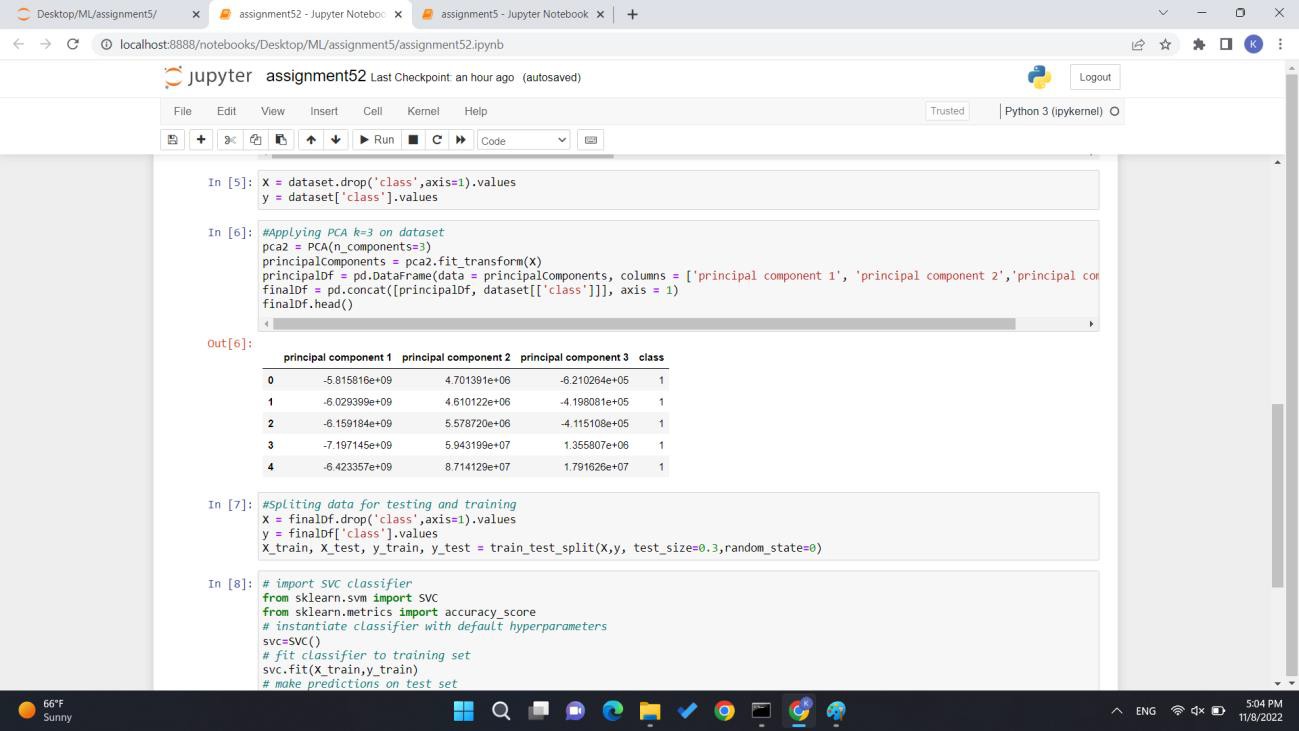
Question 1.3: first I have applied the feature scaling using standardscaler() on the original dataset. Then again applied the PCA and k-means (clusters = 3). And calculated silhouette score which is 0.45232.

Silhouette score after feature scaling is less compared to first silhouette score.

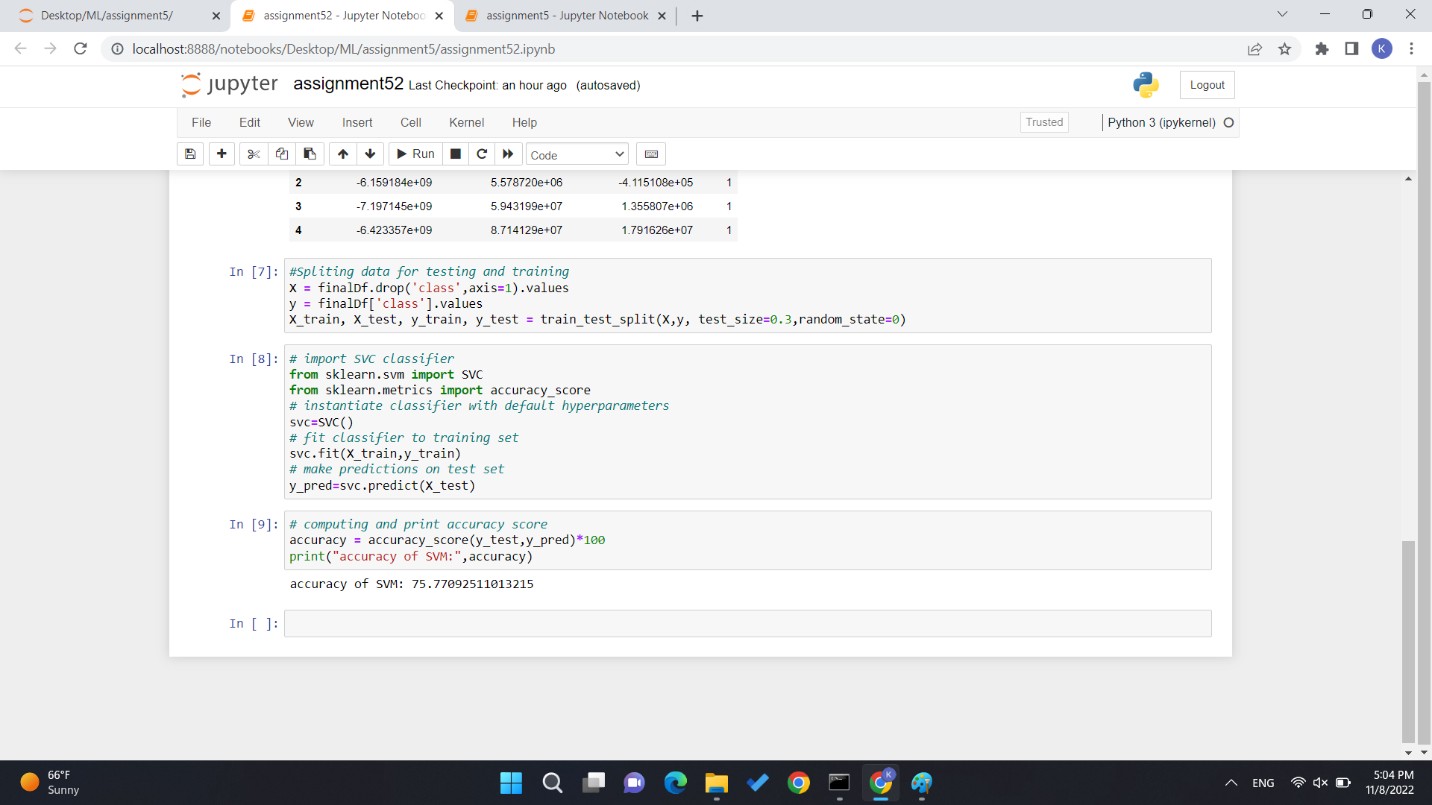
# Question 2:



For question 2 first I have imported the required libraries. Imported the dataset “pd\_speech\_feature.csv” then displayed the first few rows of the dataset. Then applied the feature scaling on the dataset and displayed the first few rows of the dataset.

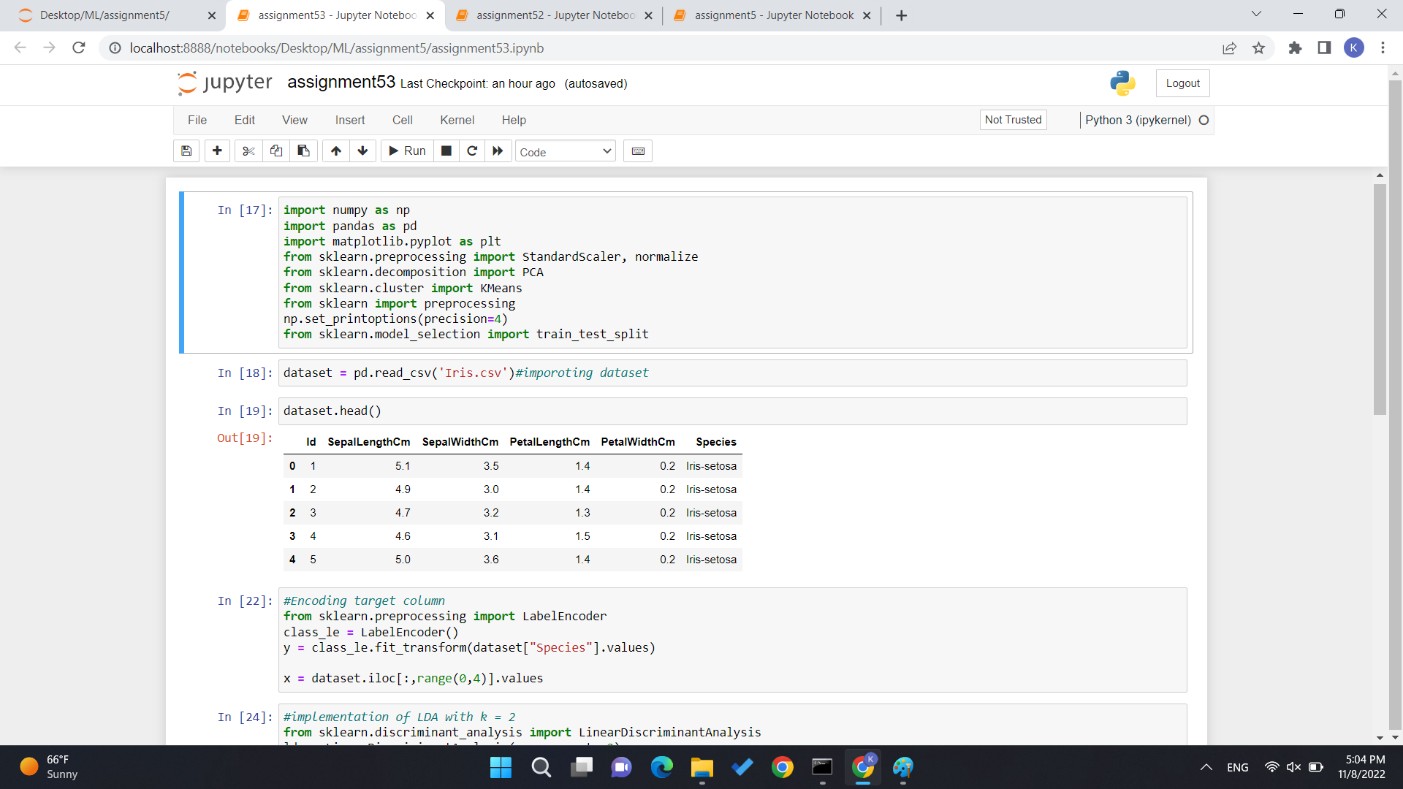


First all the columns but not “class” has been assigned to “X” and column “class” has been assign to “y”. PCA has been applied to “X” where k = 3 which means dataset has been reduced to 3 and displayed the new dataset named as “finaldf”with the class column.

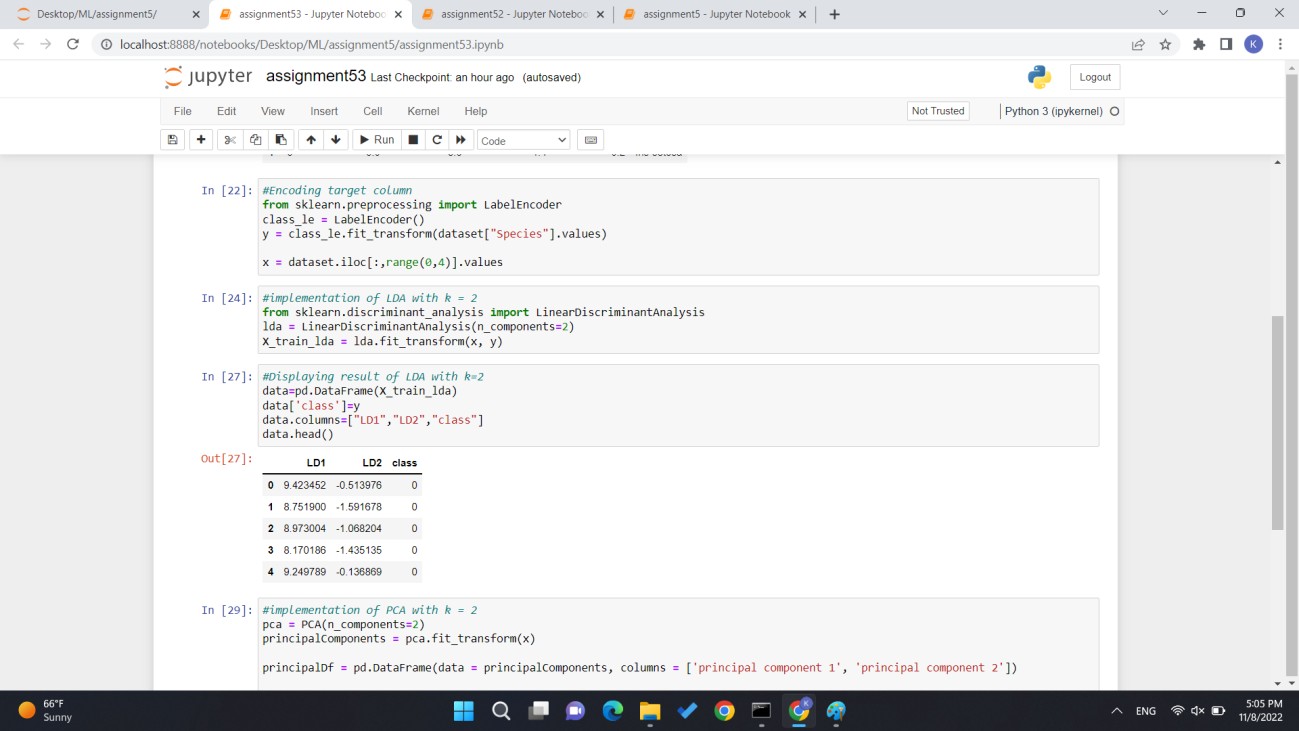


To calculate SVC first new dataset named “finaldf” is assigned to X and y. where column class has been assigned to y and other columns has been assigned to x. Train\_test\_split has been used to split the data for training and testing. Then SVC has been applied on x\_train and y\_train and prediction has been done on x\_test. After that accuracy has been calculated which is 75.77092.

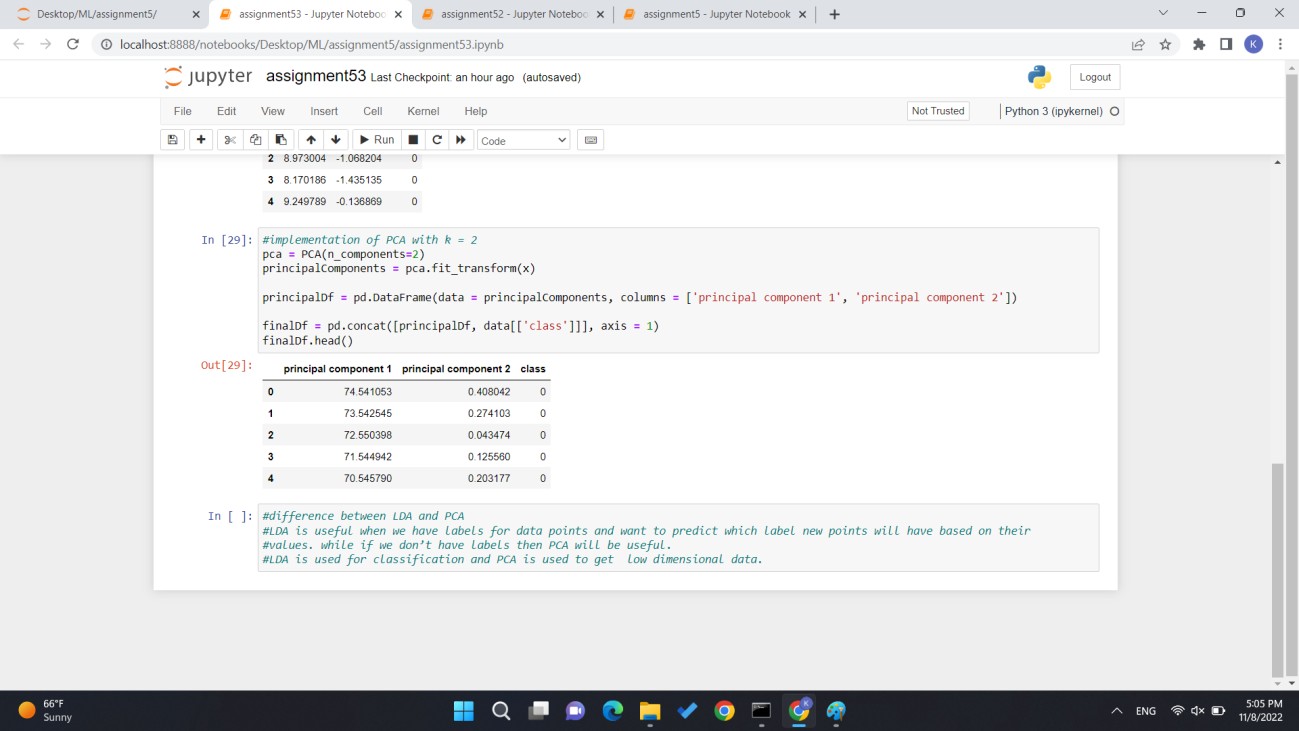
**Question 3:**



First all the libraries and dataset “iris.csv” has been imported and displayed the first few rows of the dataset.



To changed the values of target column labelEncoder() has been used. So it is easy to process the dataset. Values of all columns but the target has been assigned to X and target column has been assigned to y. implementation of LDA has been done and dataset has been reduced to k =2. And result of the new dataset has been displayed.



implementation of PCA has been done where k=2 to compare result of LDA and PCA.

**Question 4:**

difference between LDA and PCA:

LDA is useful when we have labels for data points and want to predict which label new points will have based on their values. while if we don’t have labels then PCA will be useful. LDA is used for classification and PCA is used to get low dimensional data.

GitHub: <https://github.com/pallavi234/ML_Assignment5>

Video: <https://drive.google.com/file/d/1vfKJMk-na1N4_kh3RpsD8RPks2YeMohV/view>