#700743770

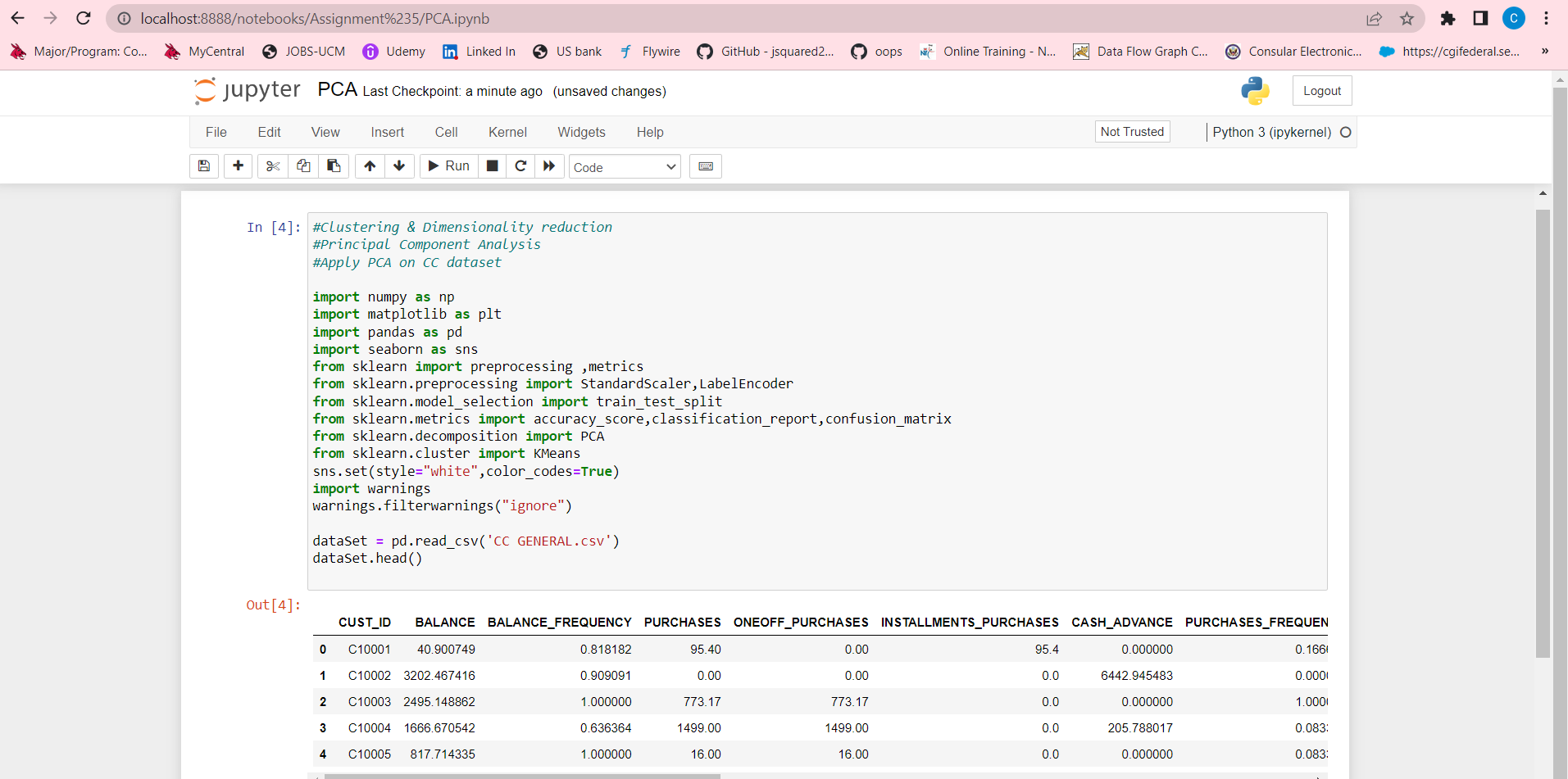
Sravanti Cherukuri Assignment #5

**Git hub link :** <https://github.com/sxc37701/ML_Assignments>

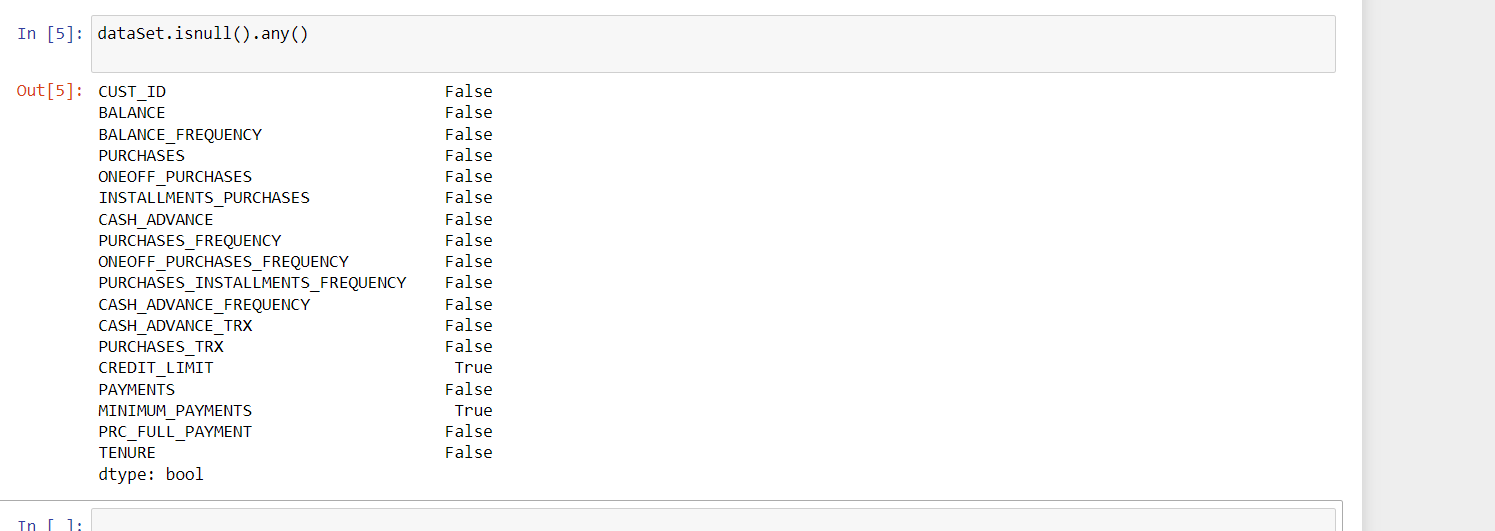
**Video Link:**https://drive.google.com/drive/folders/1CtEjxD967psXs6P3PzpRjSIn1j51gX8p

1)**Principal Component Analysis**

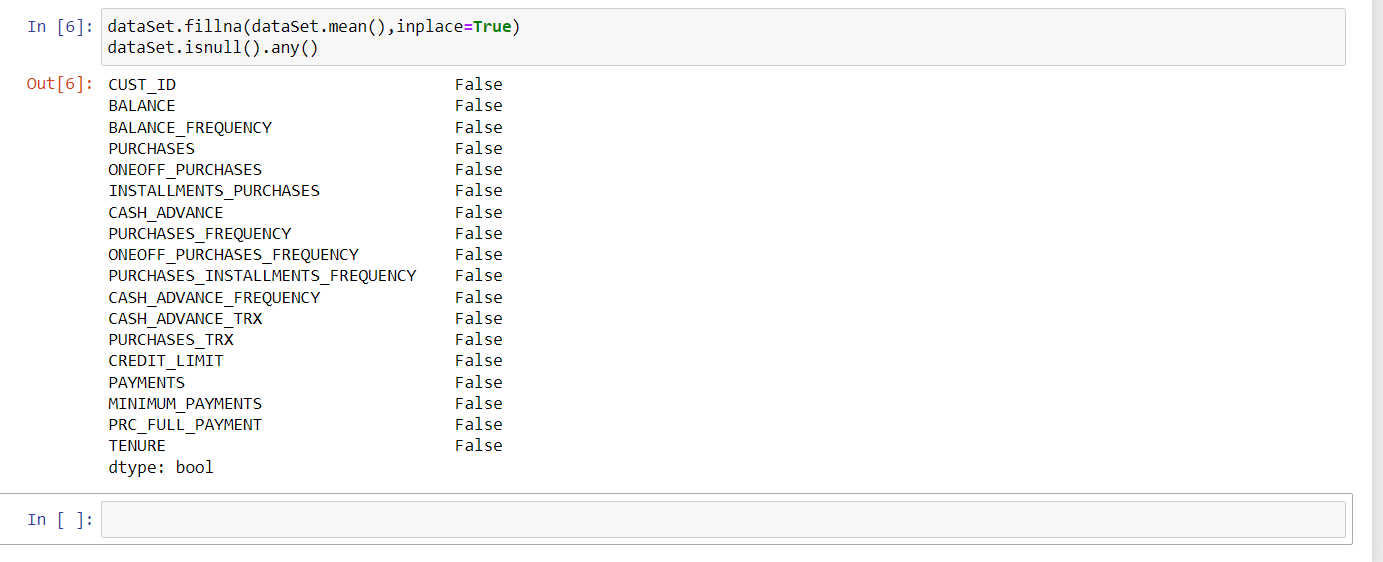
**a) Apply PCA on CC dataset**

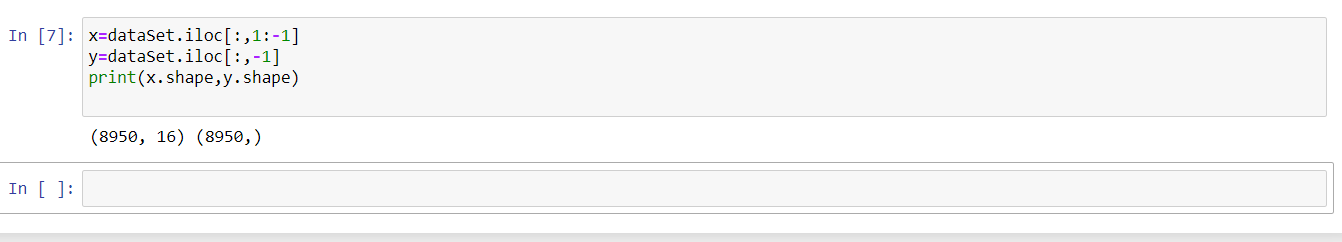


Checked if any data in the csv is null or not



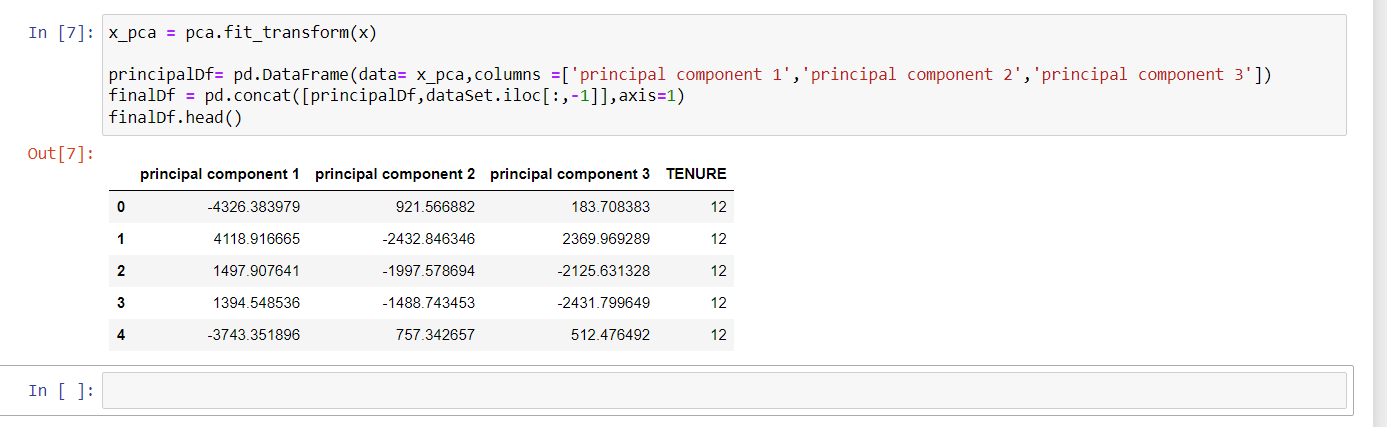
So fill the nulls with mean value





* transform() & Dataframe functions used to get the principal components in a

tabular column and the final Df .

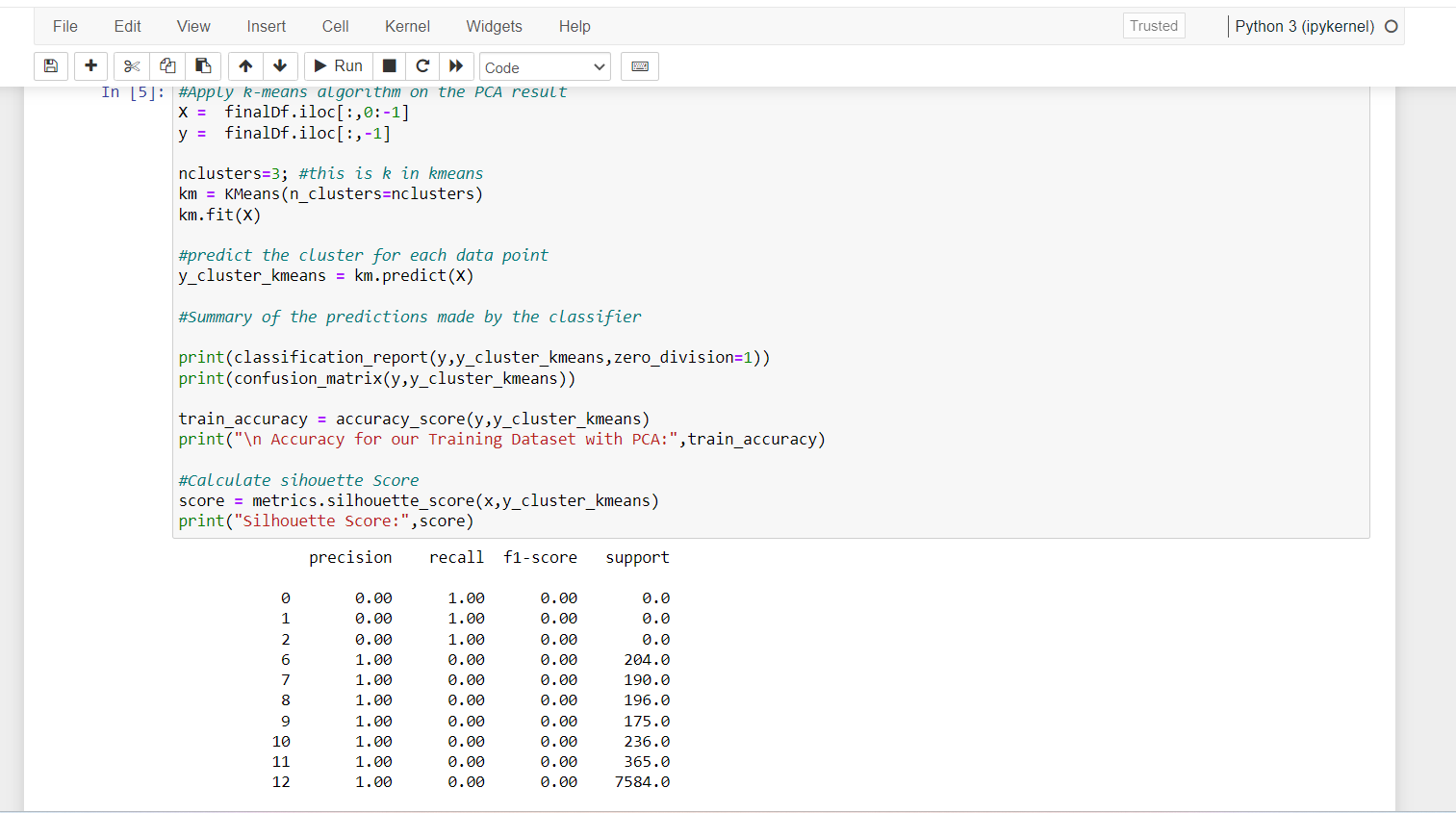


**b. Apply k-means algorithm on the PCA result and report your observation if the silhouette score has improved or not?**

Used Kmeans function to get the clusters and fix() to fix the data, iloc() function

to get the location, predict() to predict the cluster for each data point and print the predictions

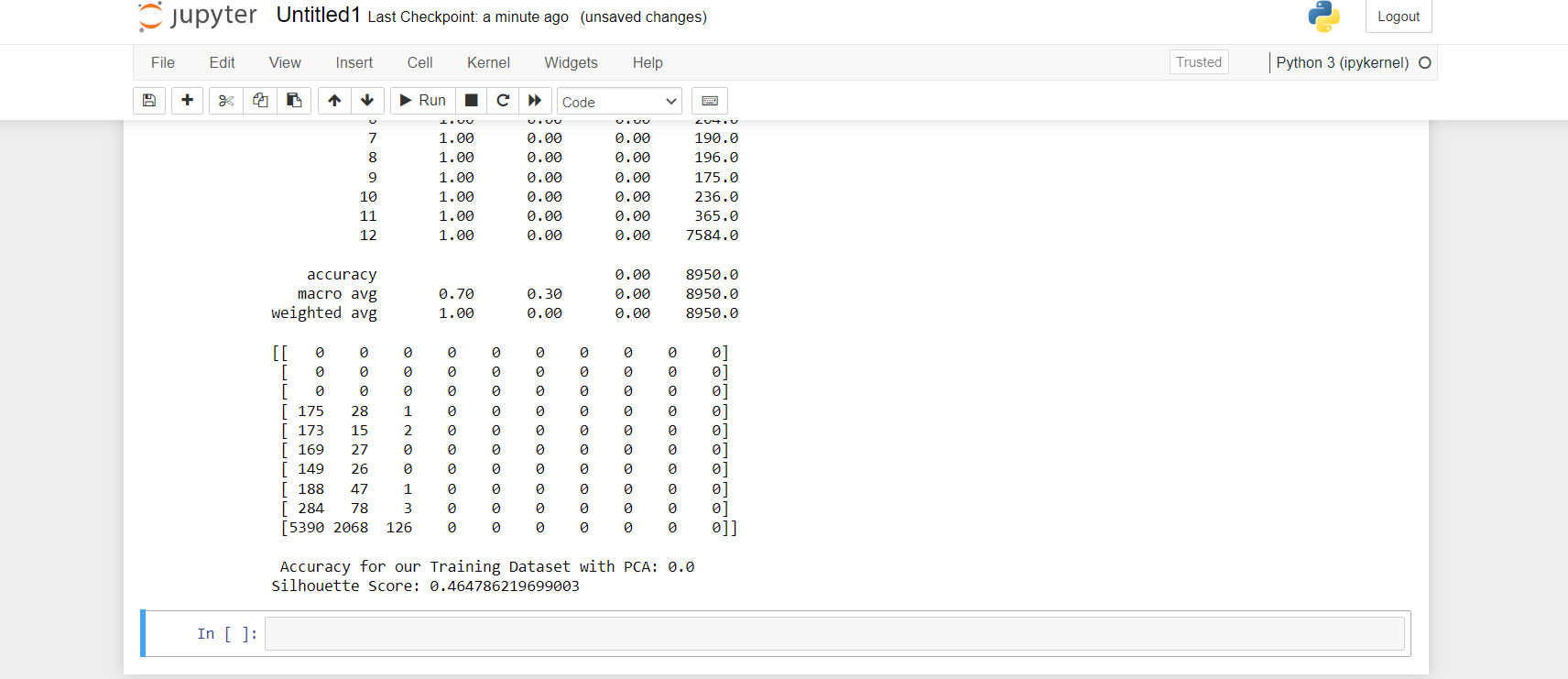
made by the classifier. Silhouette to get the silhouette score

****

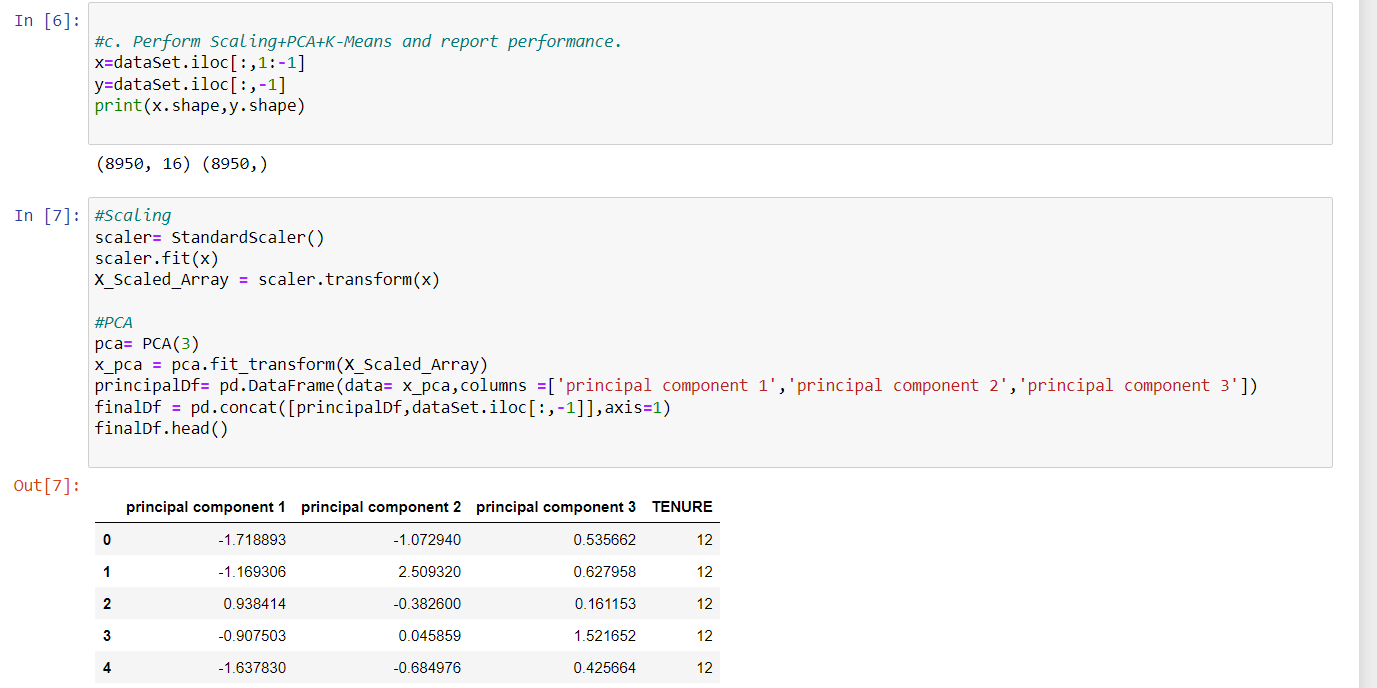
* Used iloc() to get the location, standard scalar and fit() function to get

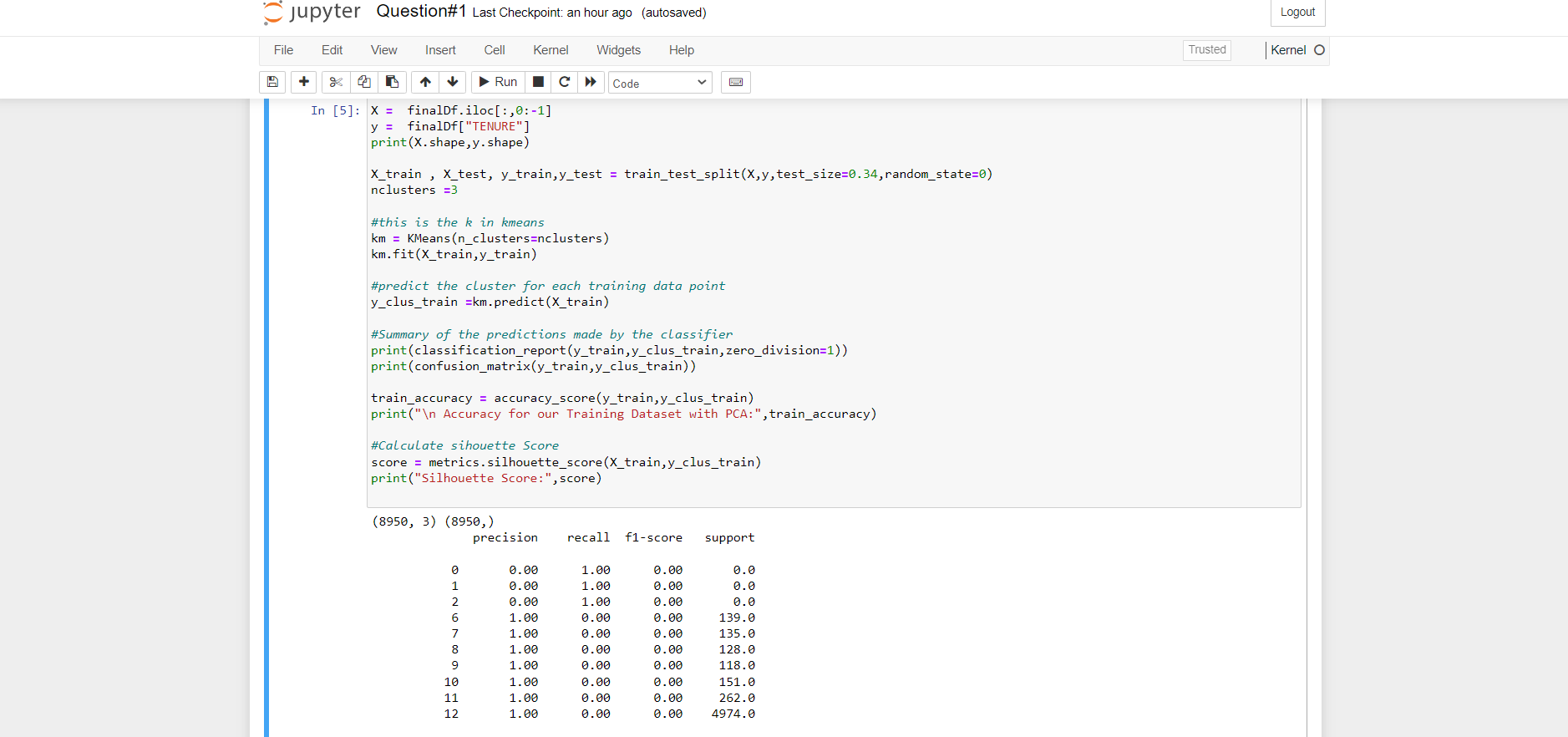
the data. predict() to predict the cluster for each data point and classification and train accuracy

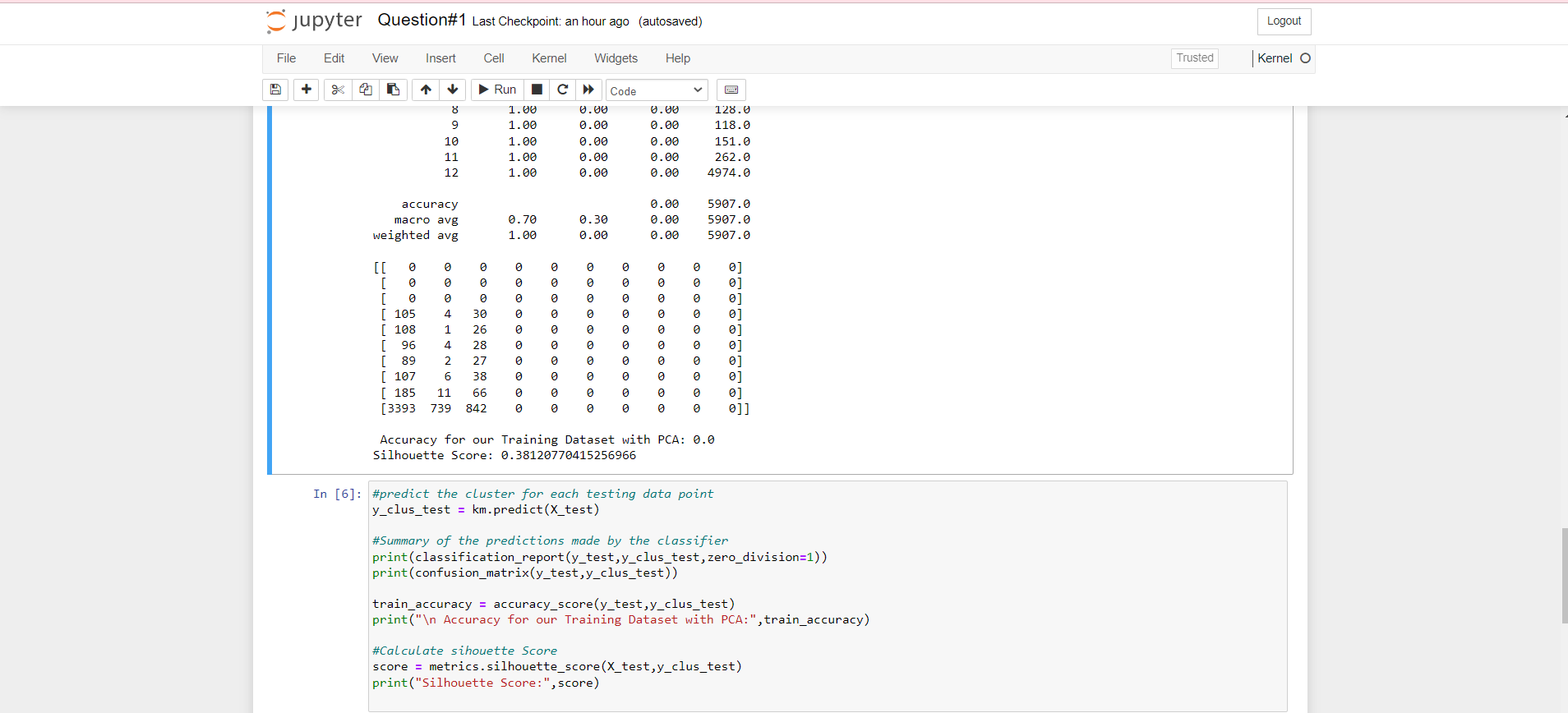
to get the accuracy score and silhouette to get the silhouette score

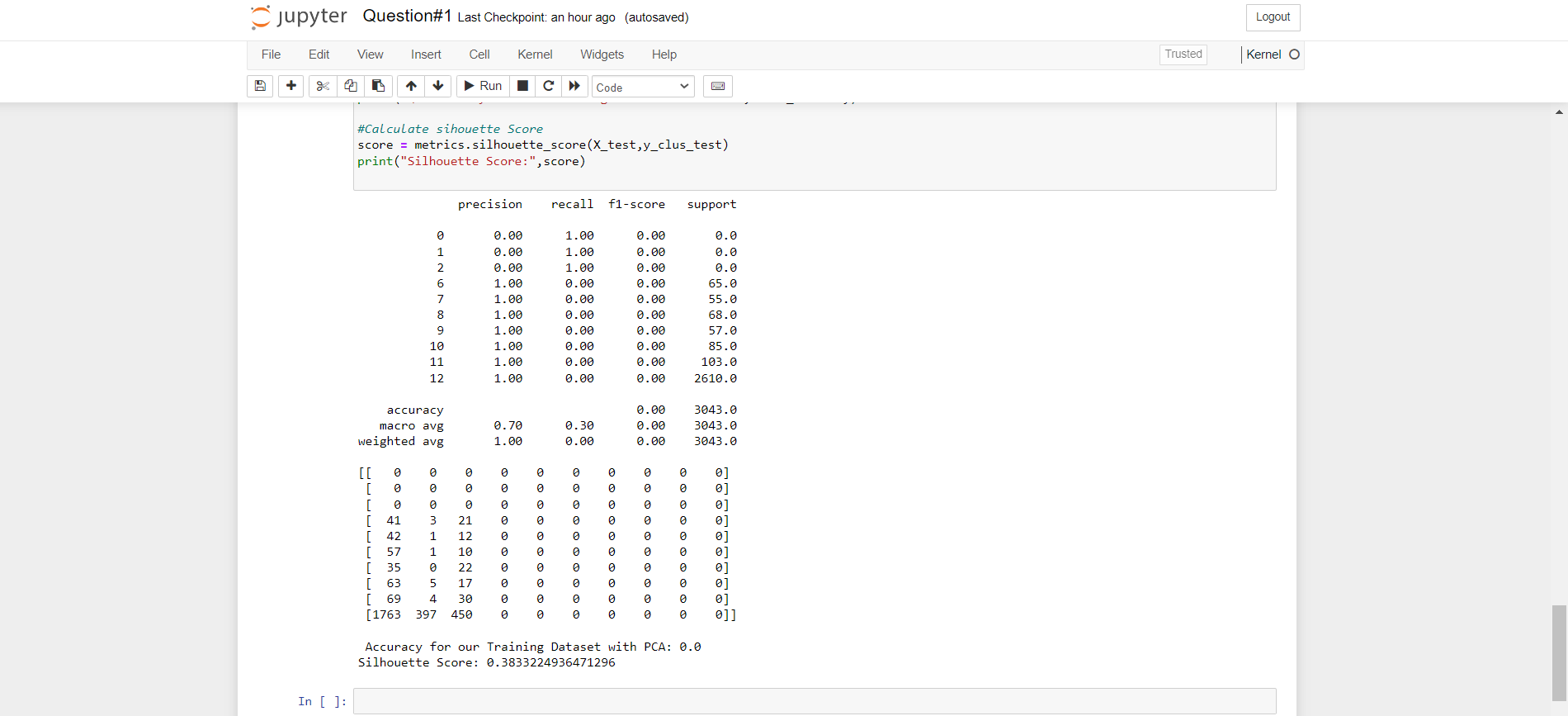
****

**c. Perform Scaling+PCA+K-Means and report performance.**

****

****

****

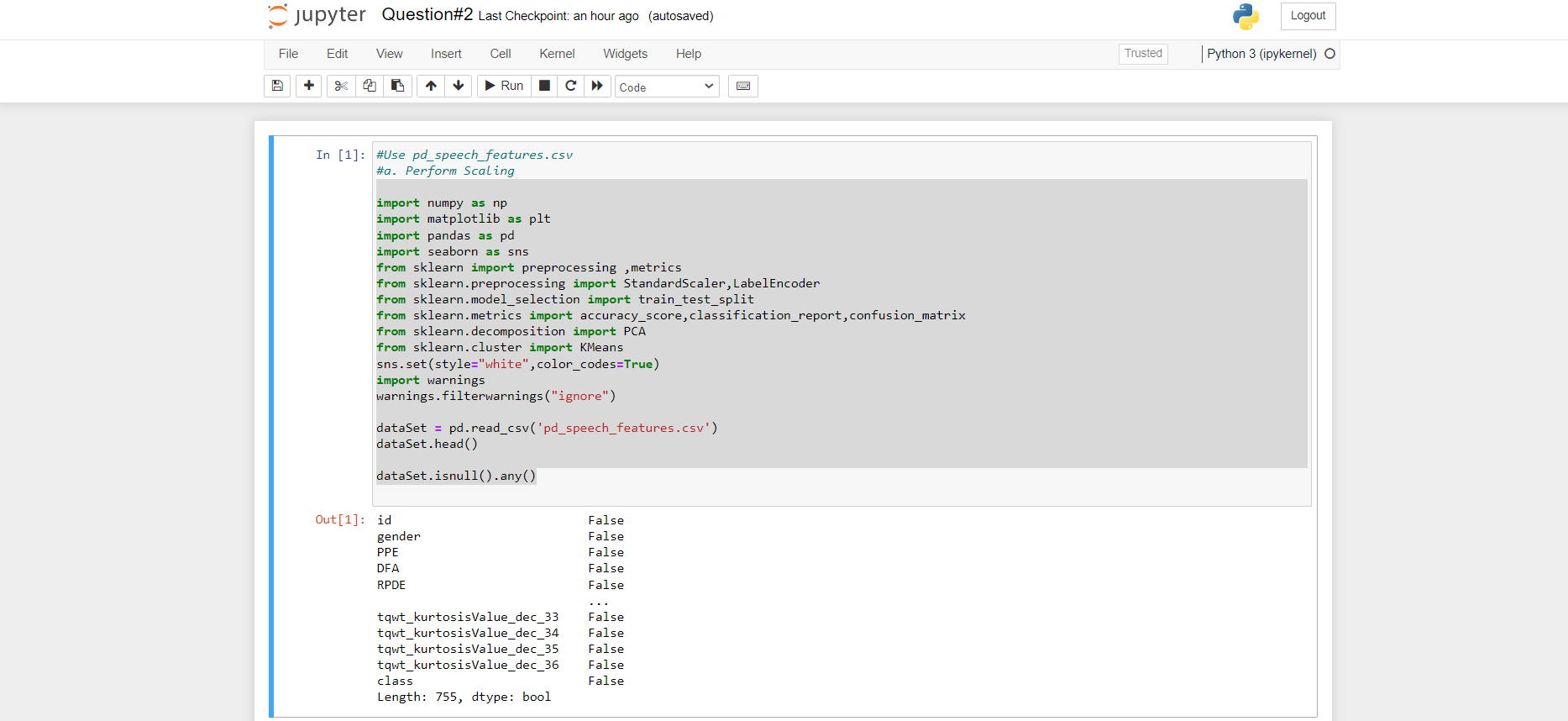
****

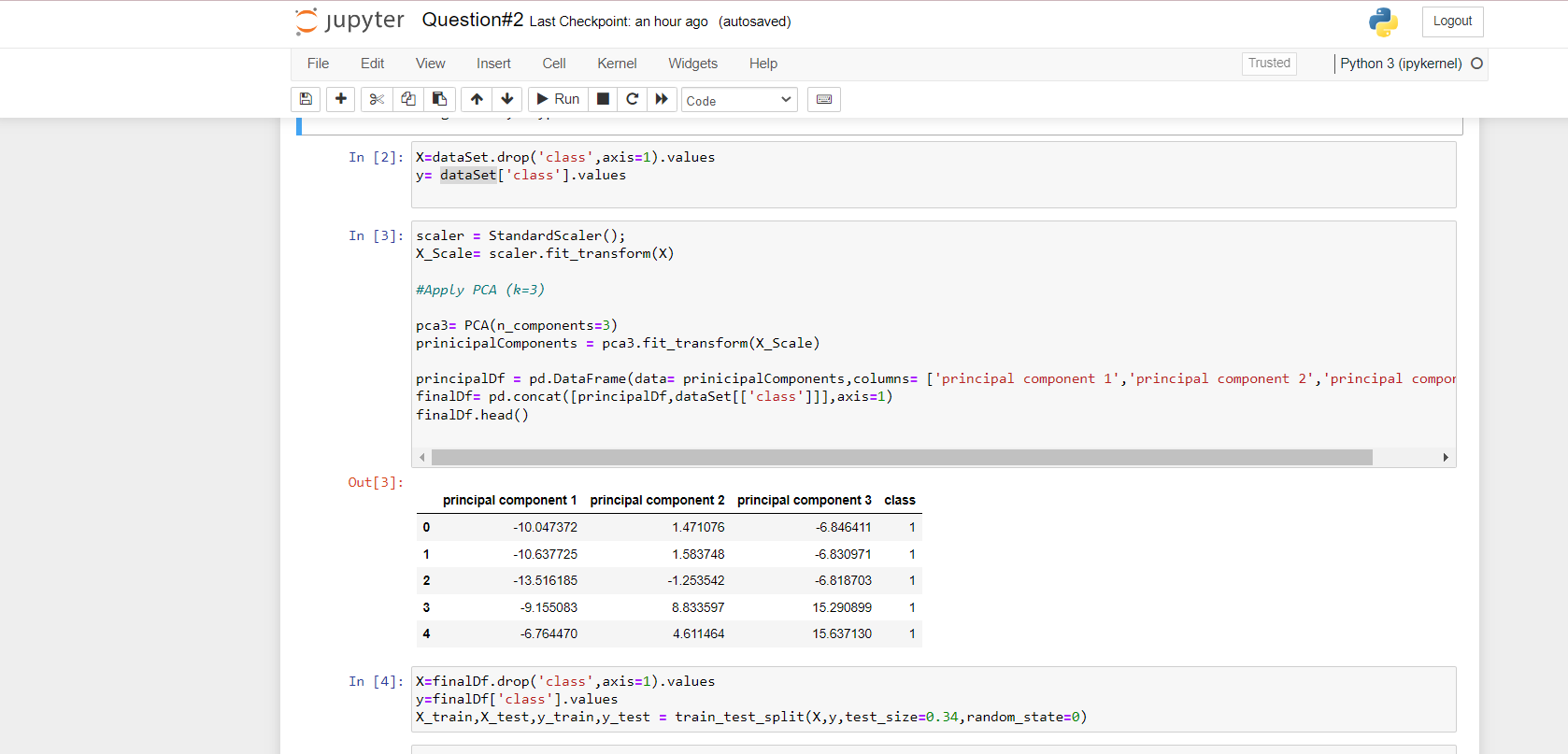
2. Use pd\_speech\_features.csv

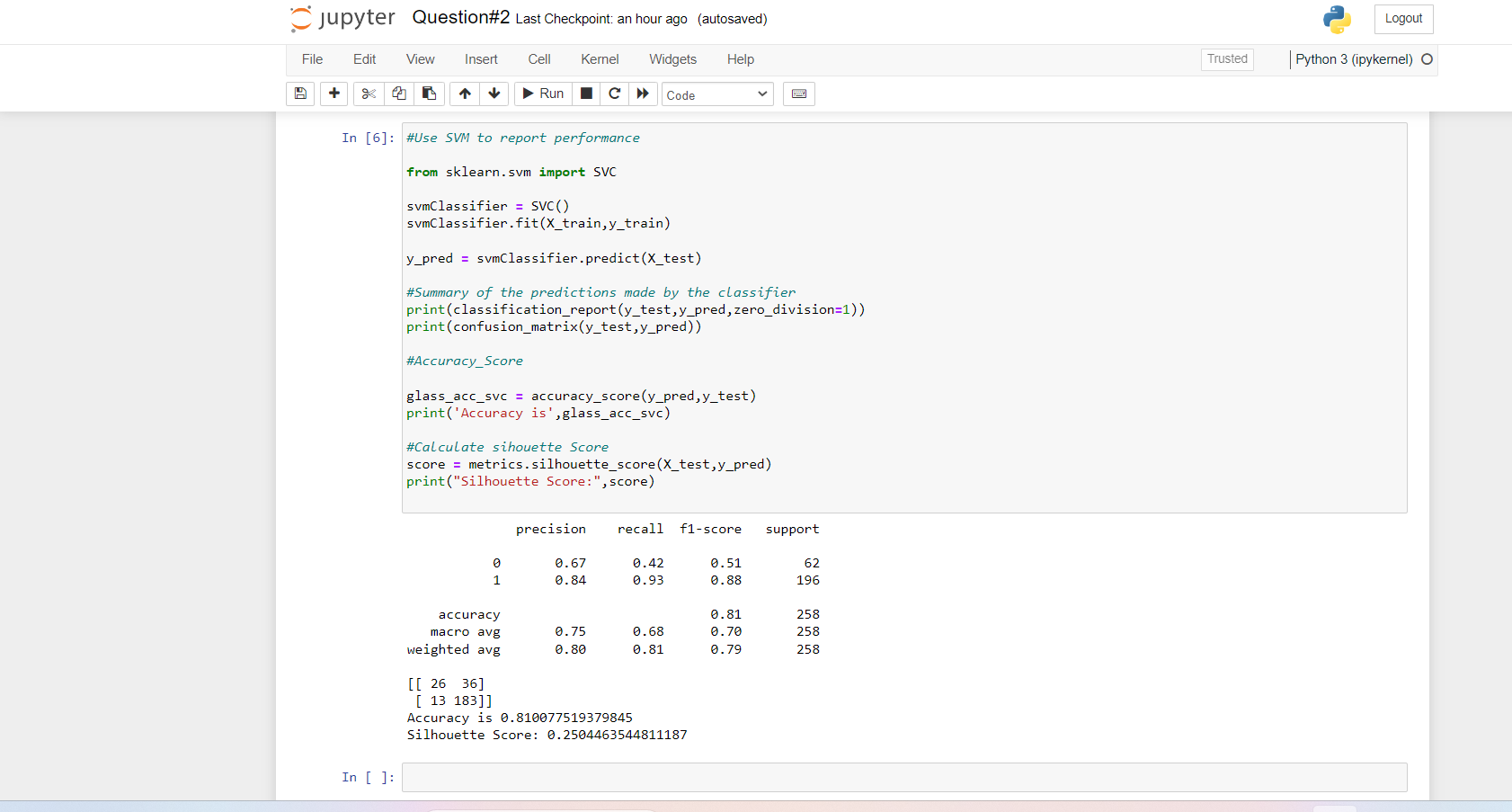
a. Perform Scaling

b. Apply PCA (k=3)

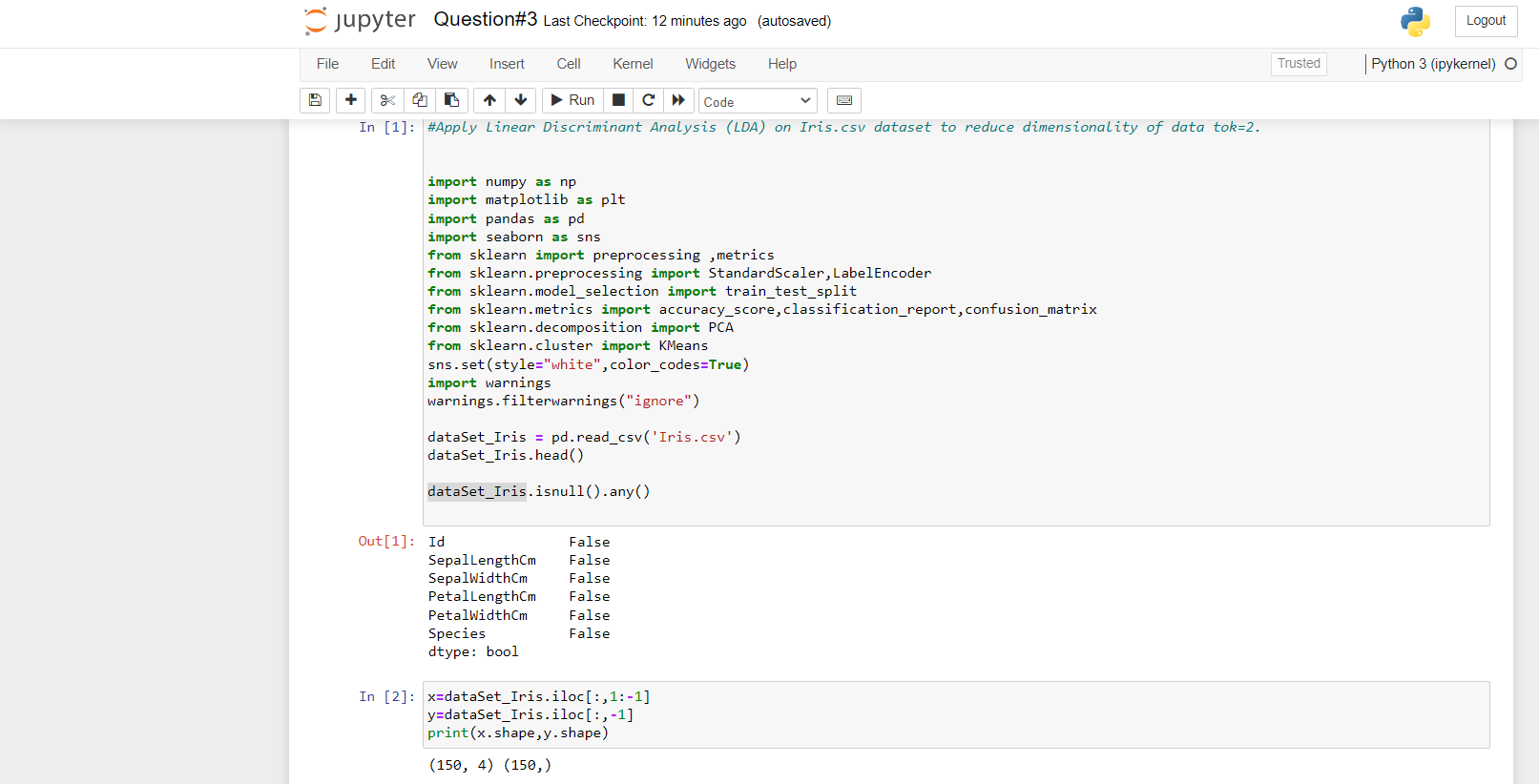
c. Use SVM to report performance

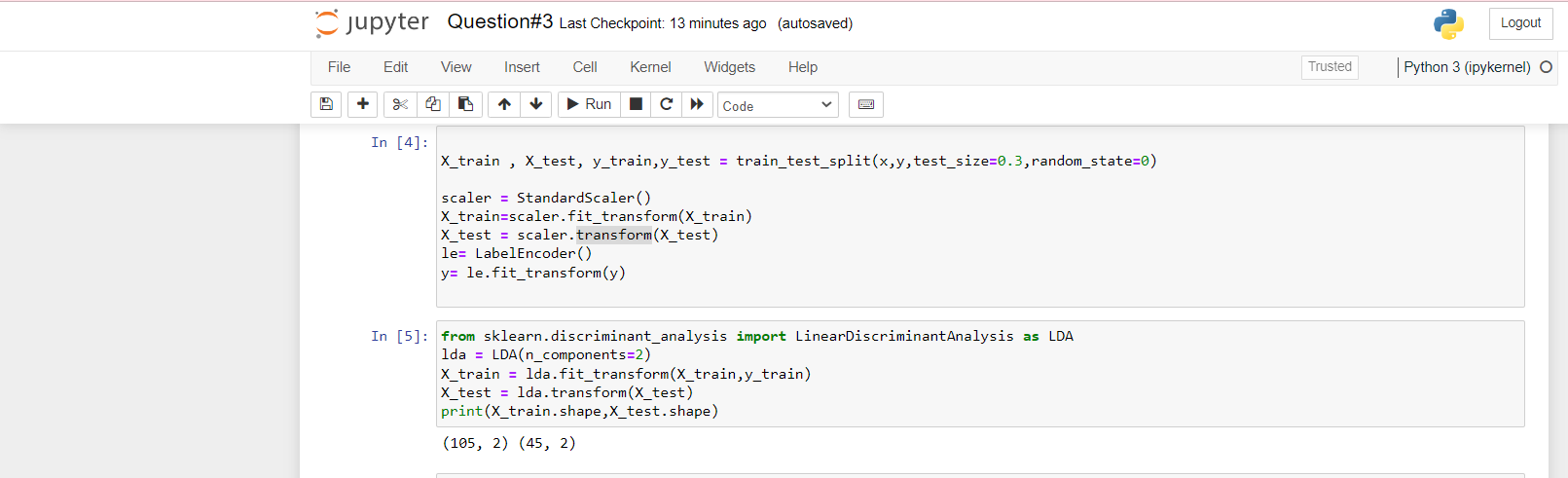






3)Apply Linear Discriminant Analysis (LDA) on Iris.csv dataset to reduce dimensionality of data to k=2





4)Briefly identify the difference between PCA and LDA

Both LDA and PCA rely on linear transformations and aim to maximize the variance in a lower dimension. PCA is an unsupervised learning algorithm while LDA is a supervised learning algorithm. This means that PCA finds directions of maximum variance regardless of class labels while LDA finds directions of maximum class separability. It reduces the features into a smaller subset of orthogonal variables, called principal components -linear combinations of the original variables. The first component captures the largest variability of the data, while the second captures the second large stand so on LDA finds the linear discriminants in order to maximize the variance between the different categories while minimizing the variance within the class