Machine Learning

Assignment-05

NAME: Charithasri Vadlamudi

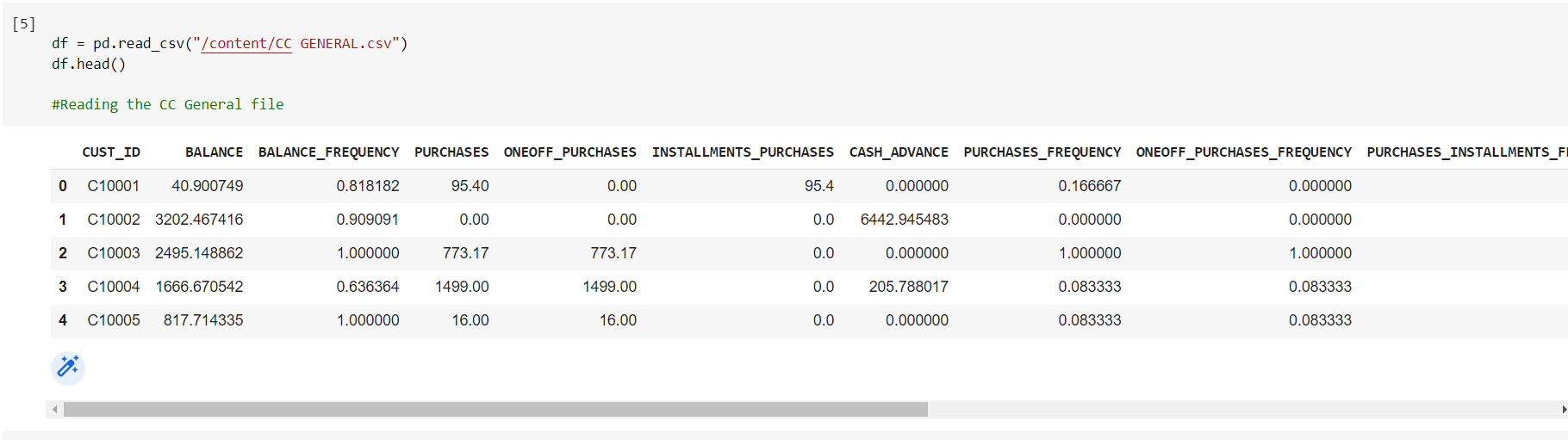
STUDENT ID:700745127

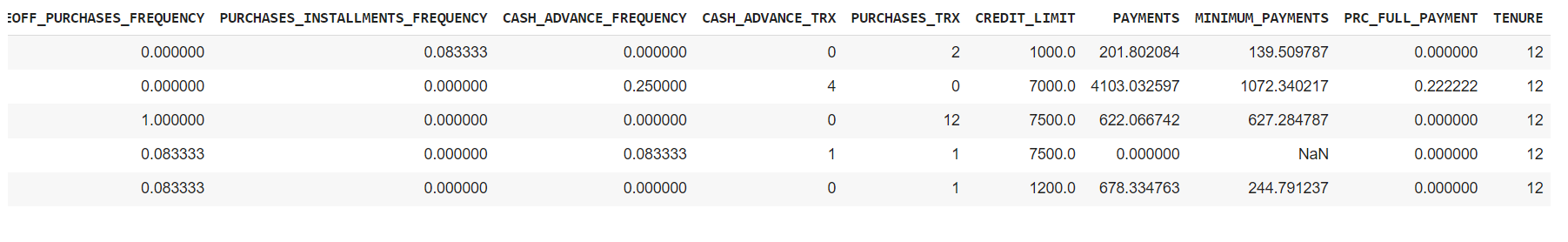
CRN:23921

COURSE ID:CS 5710

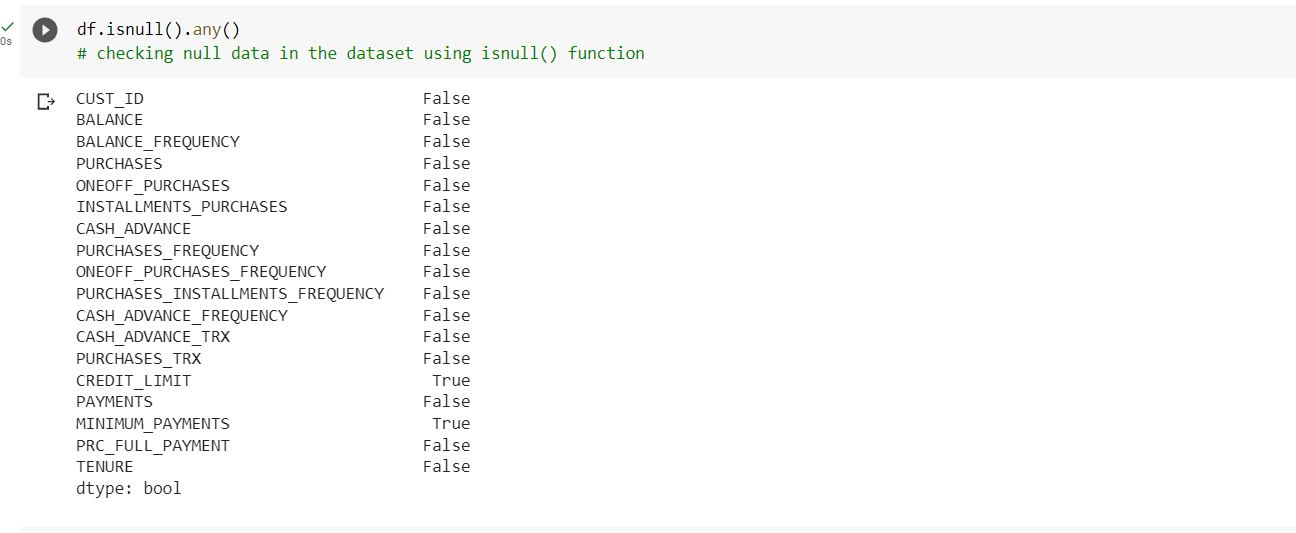
**Github Link**: https://github.com/Charithasri1/ML\_Assignment5

1.Principal Component Analysis

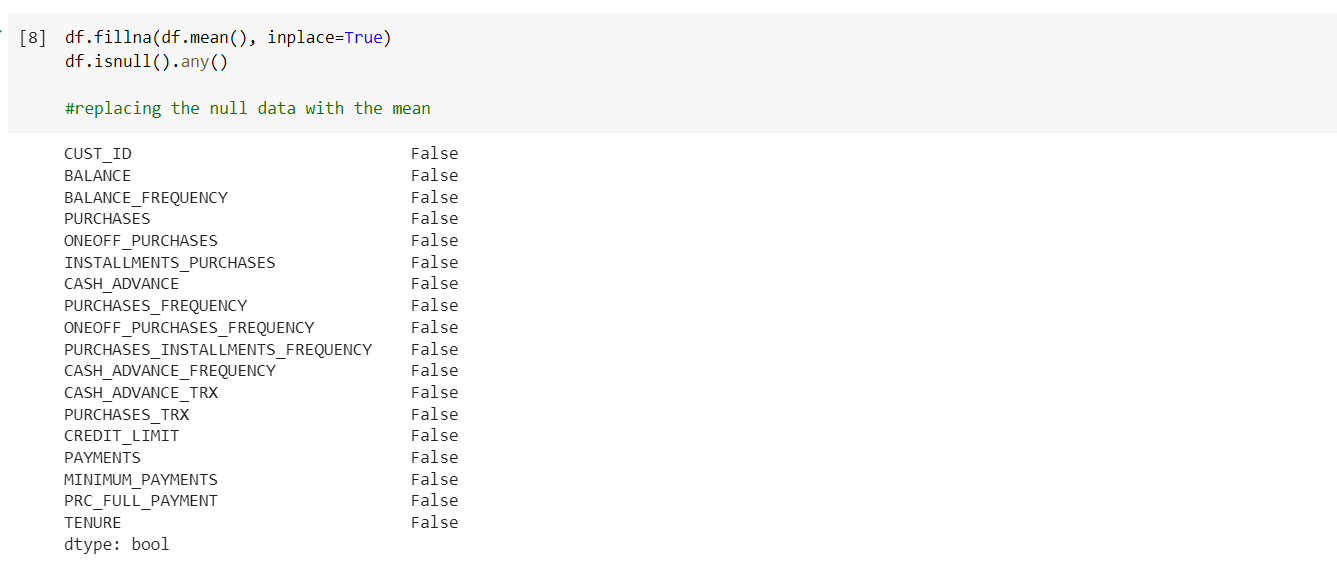


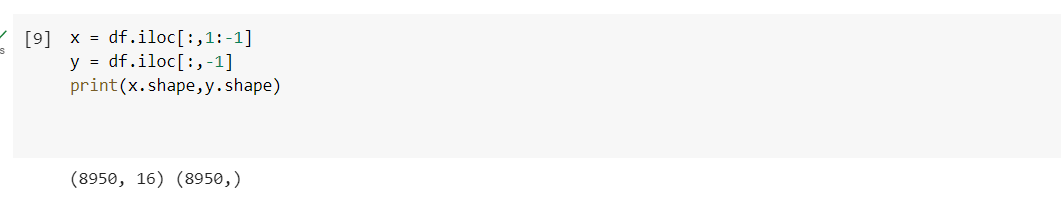


* Reading the data from dataset and displaying it using df.head()

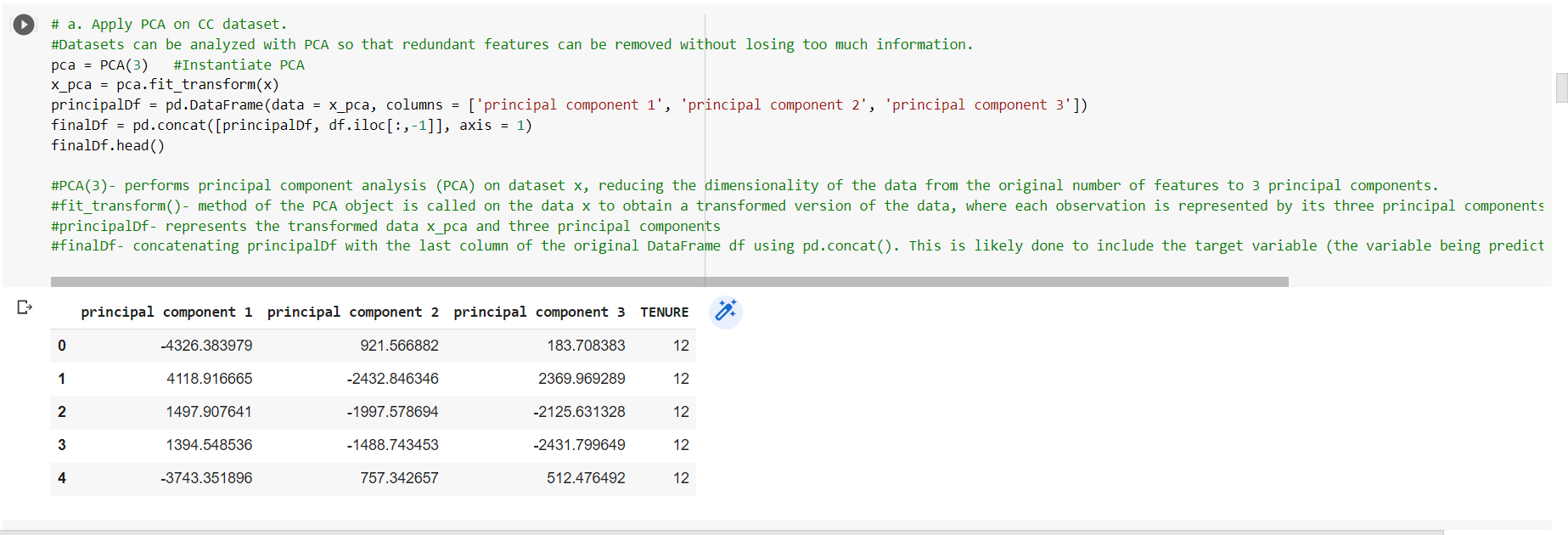


* Checking the data if it has null using isnull()



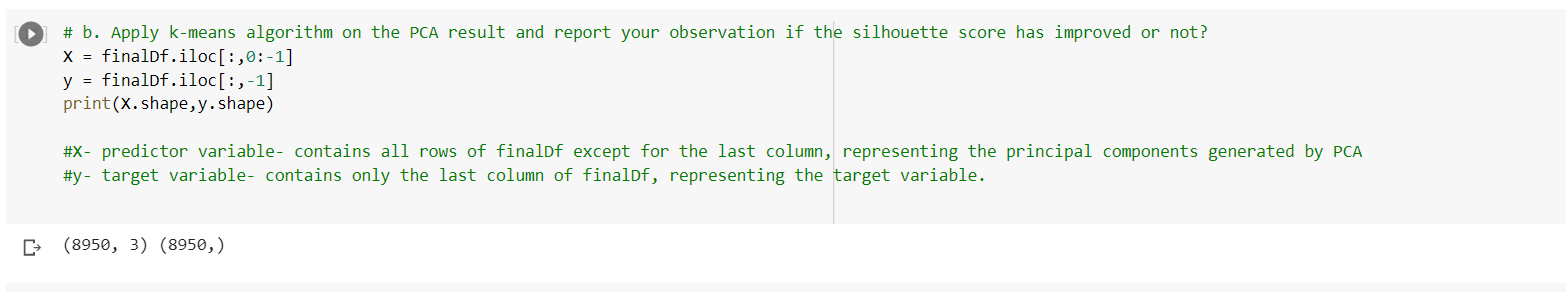
* Replacing the null with mean of the data

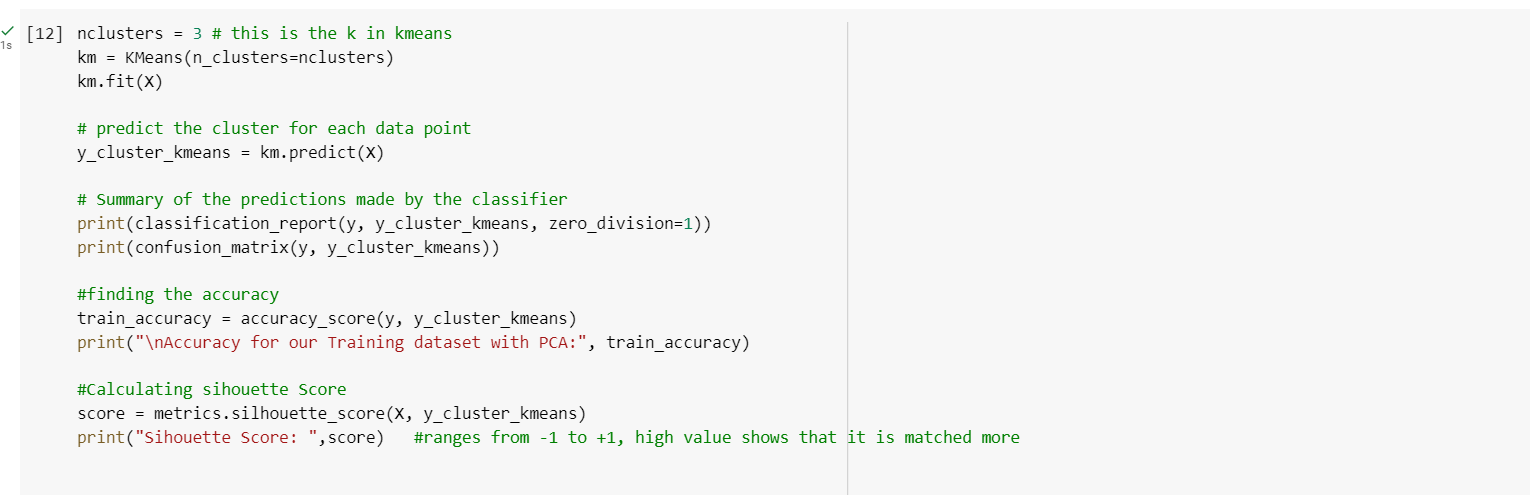
a. Apply PCA on CC dataset

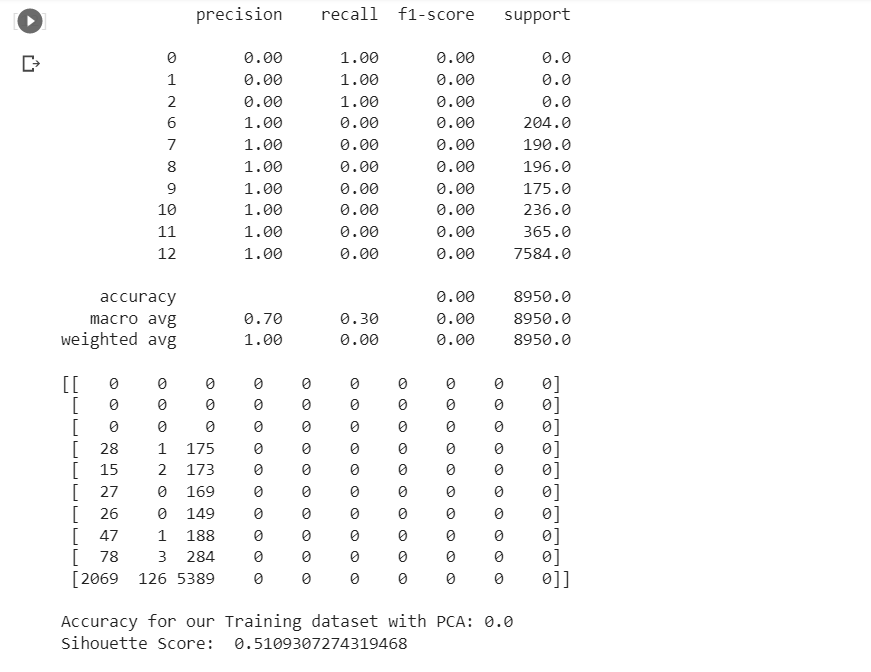


* Applying the PCA on the CC dataset which was given

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



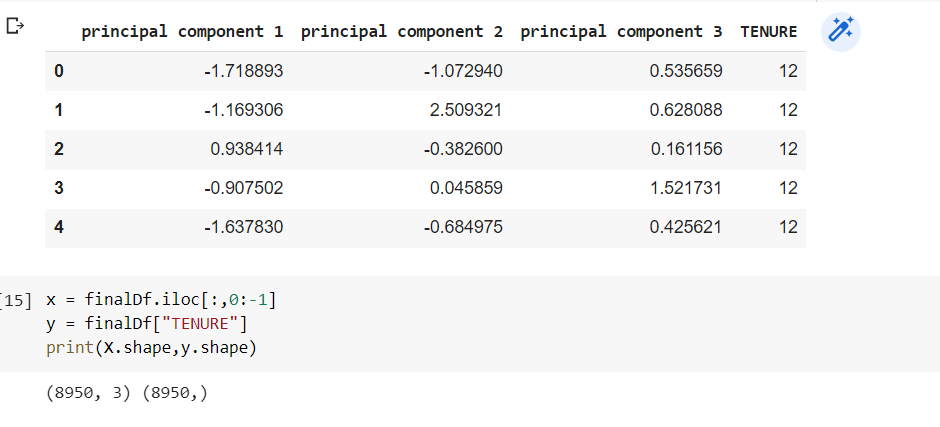


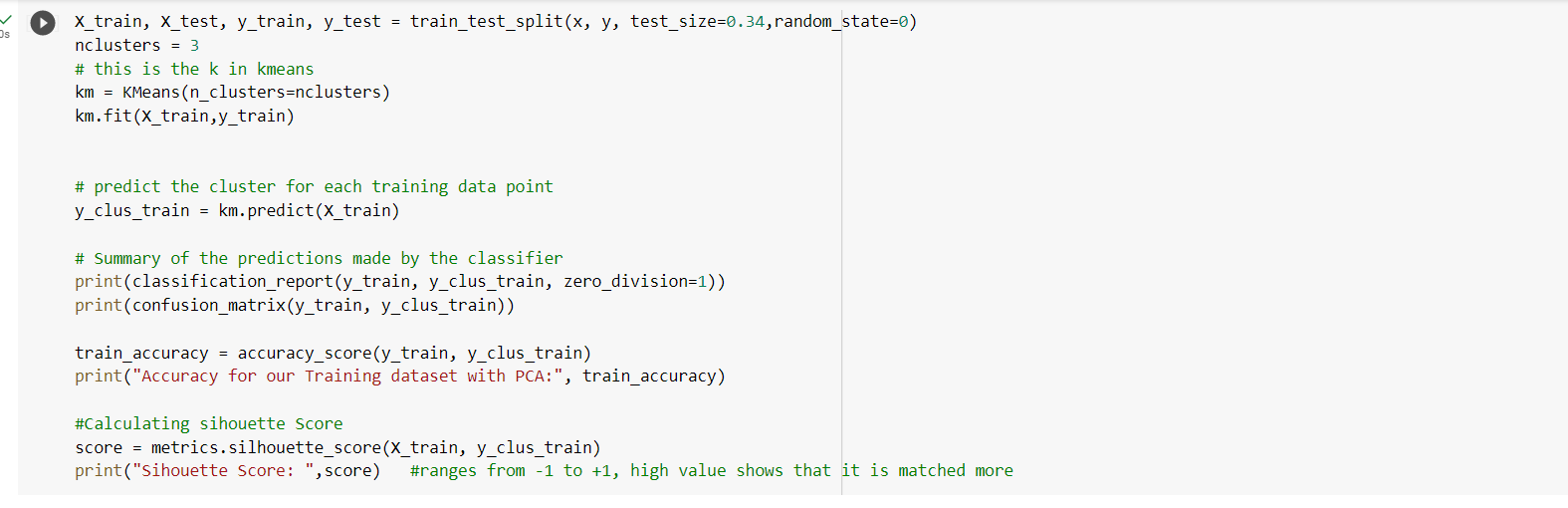


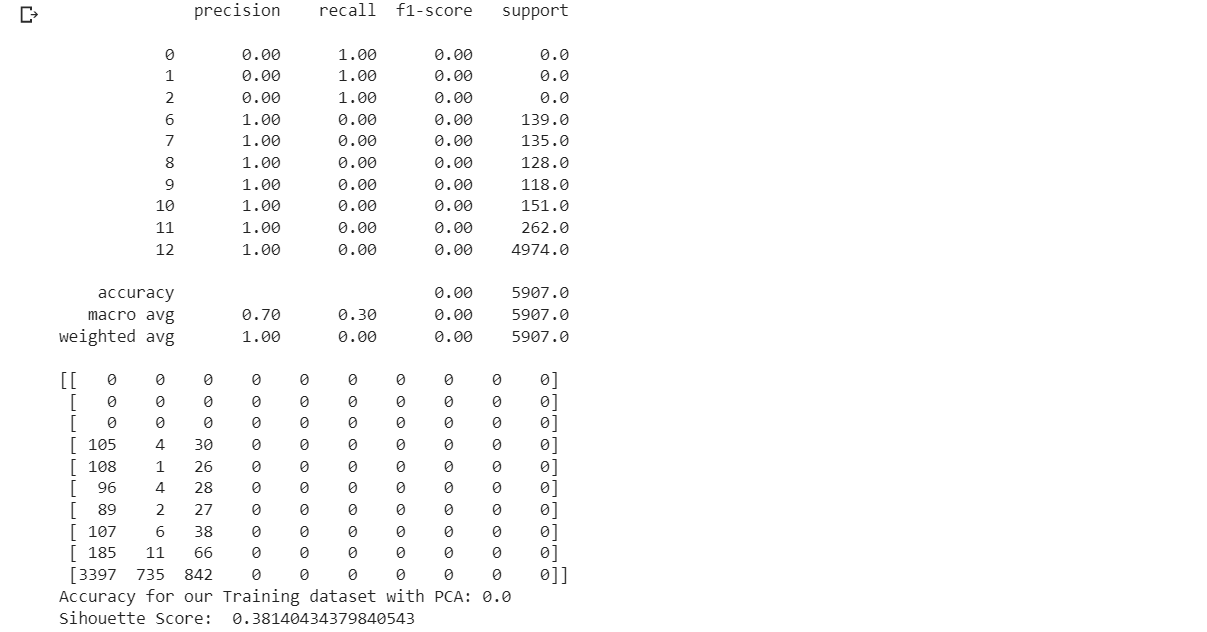
* Here the evaluation is done and observed for accuracy and silhouette score

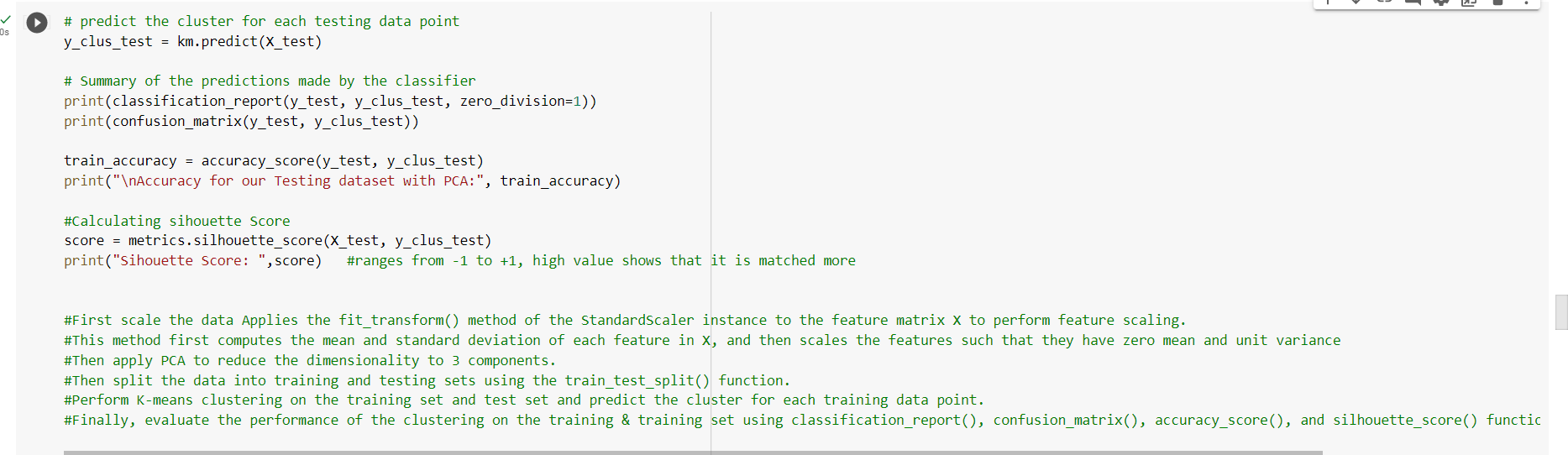
C.Perform Scaling+PCA+K-Means and report performance.

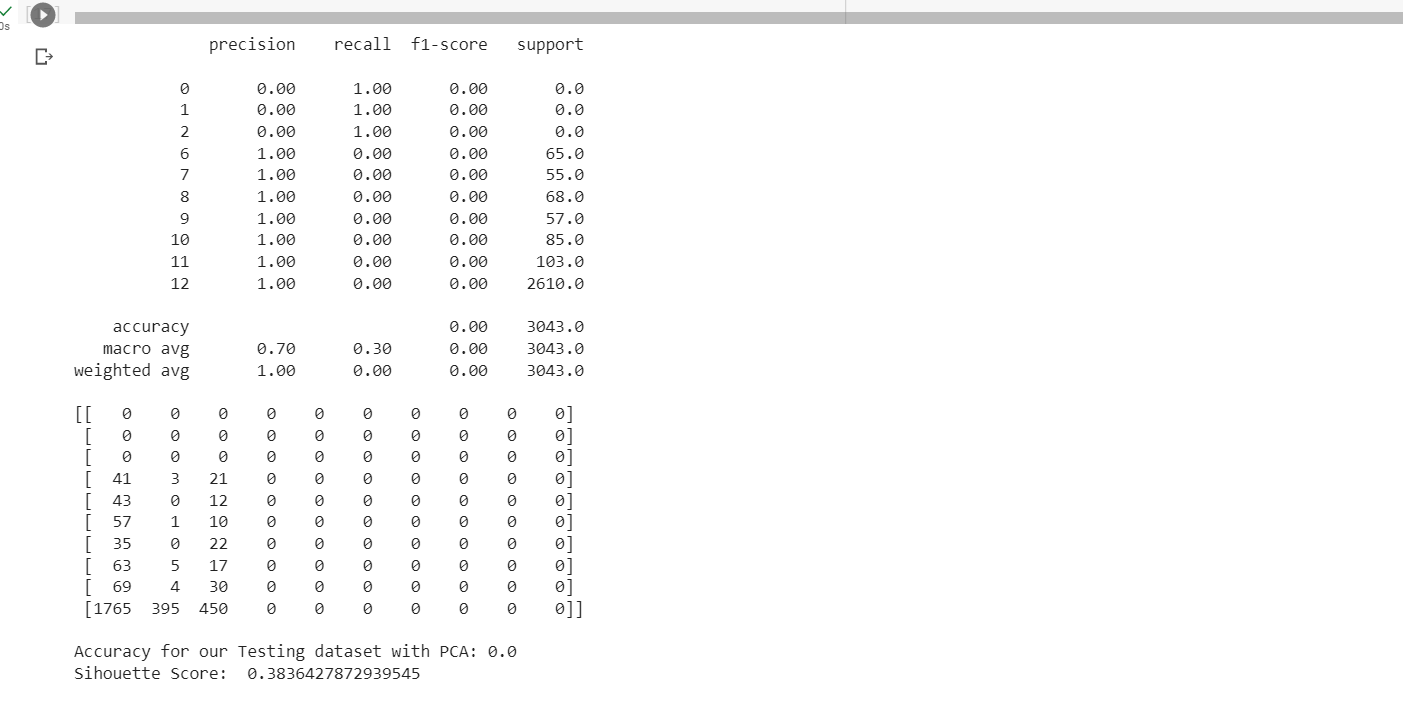




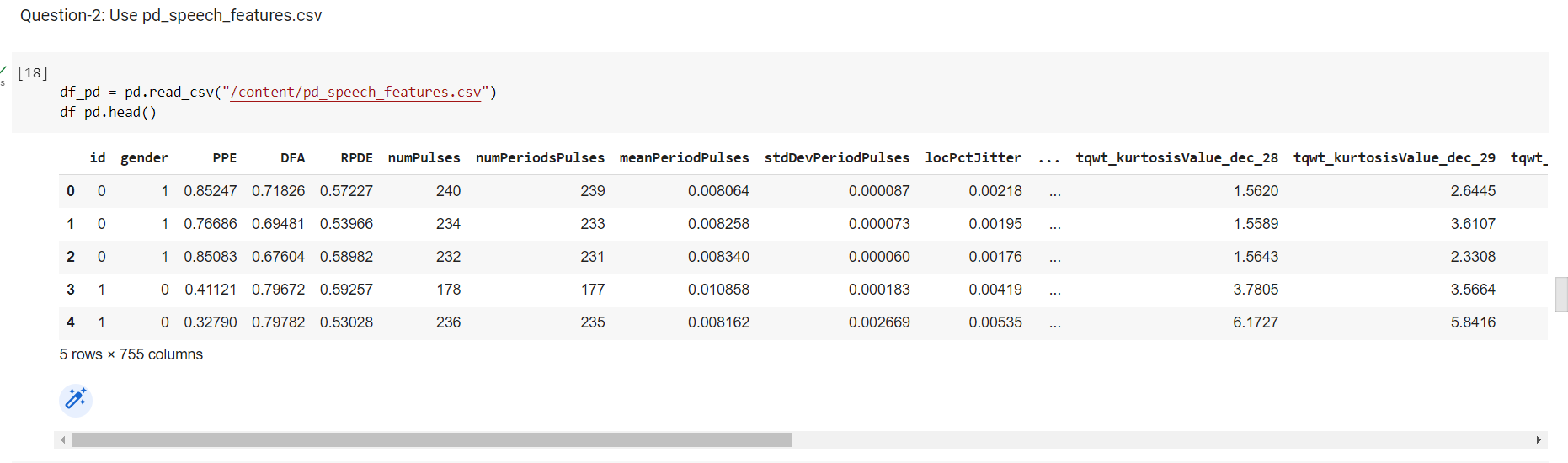


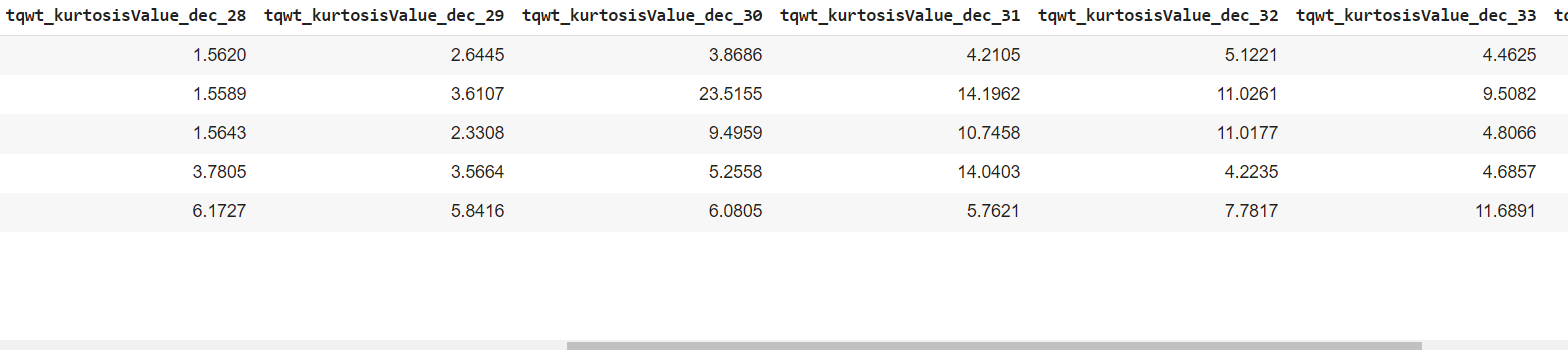


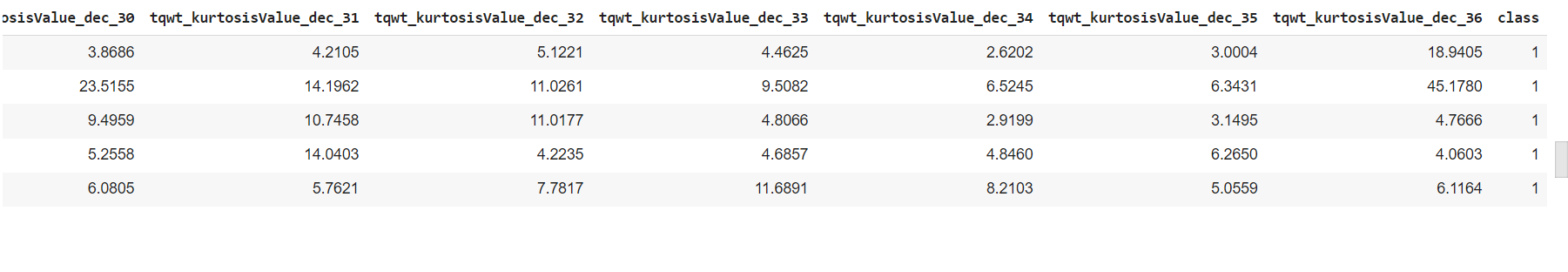




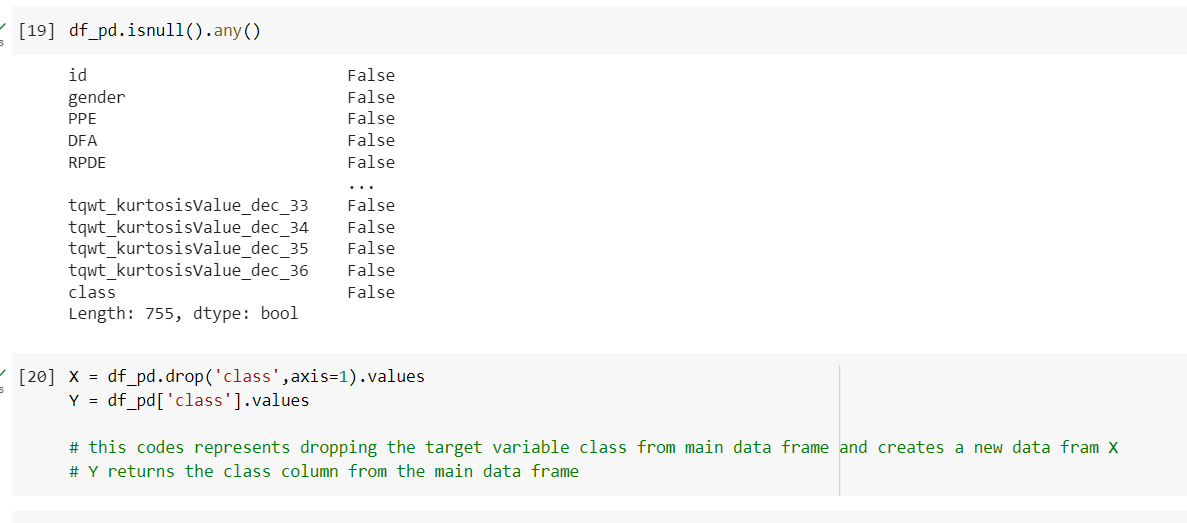
2.

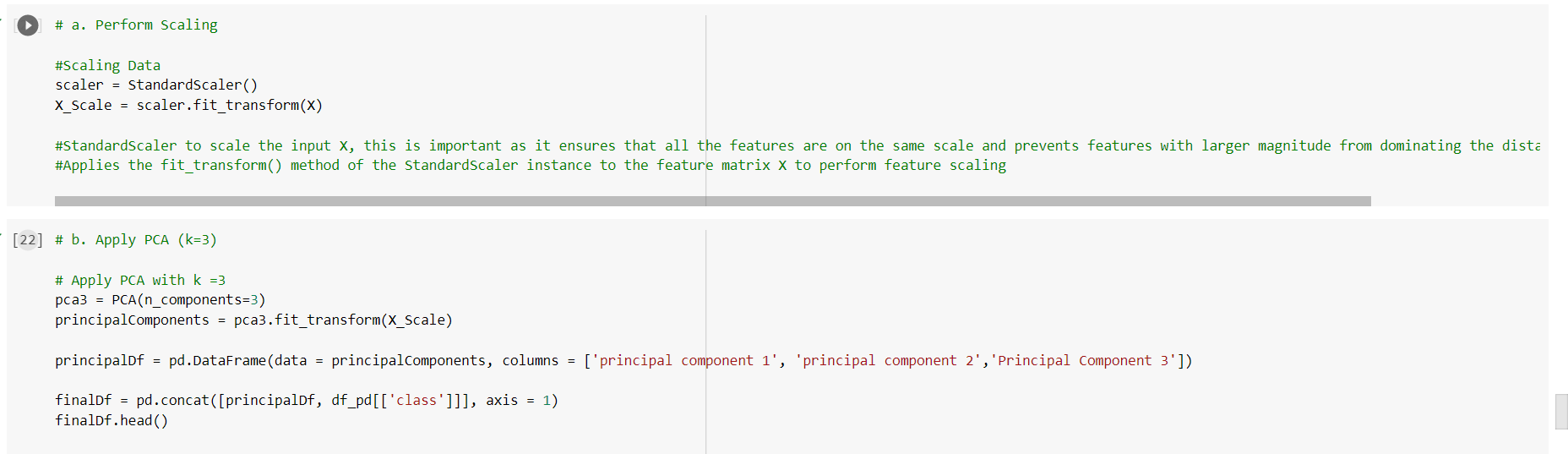


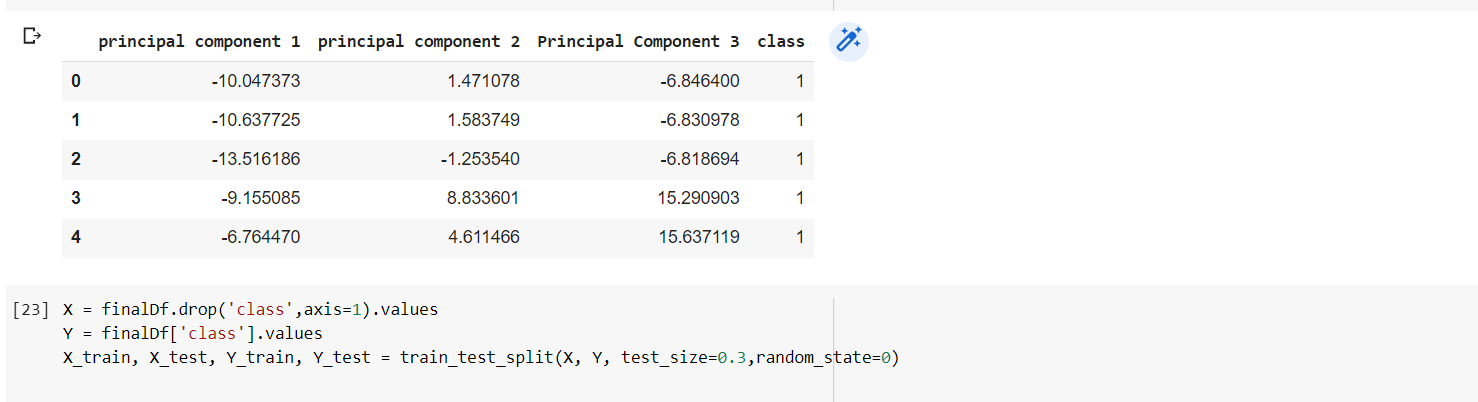


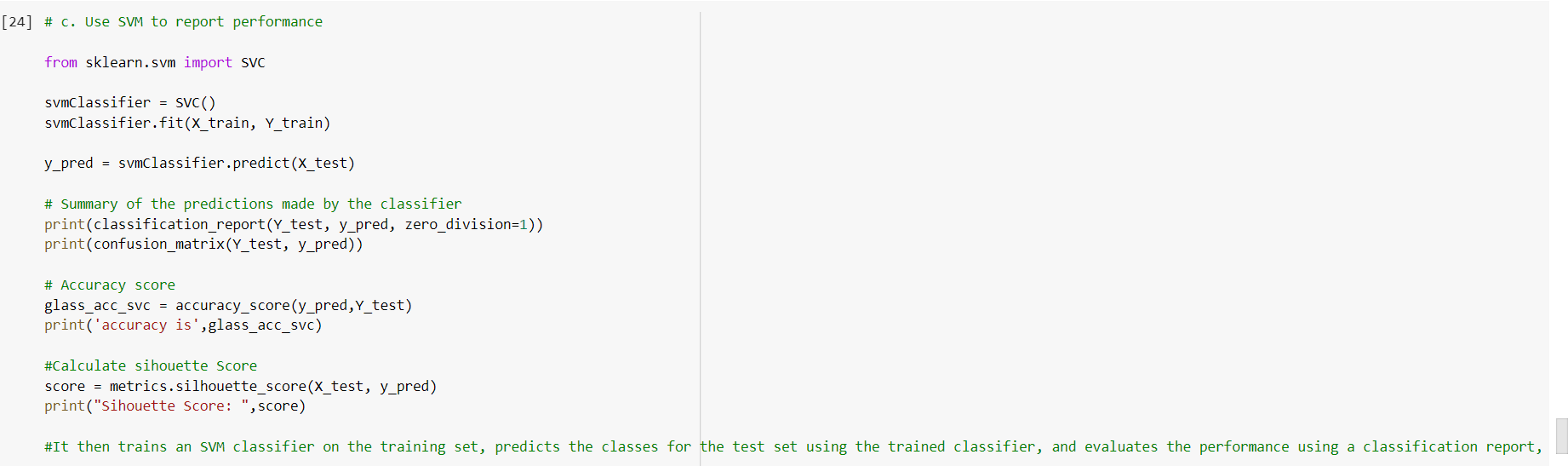


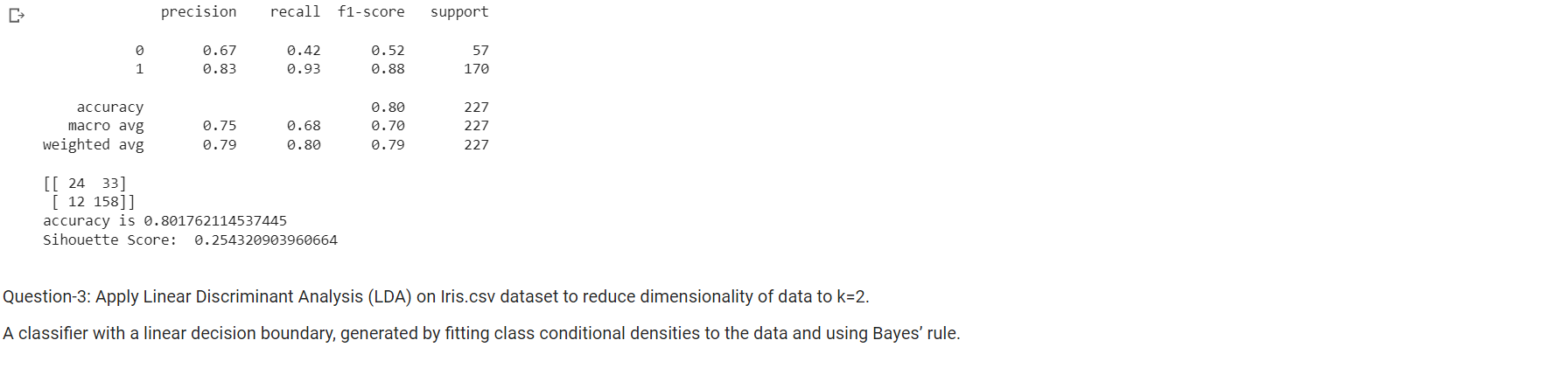
* Reading the data from the dataset and displaying the attributes



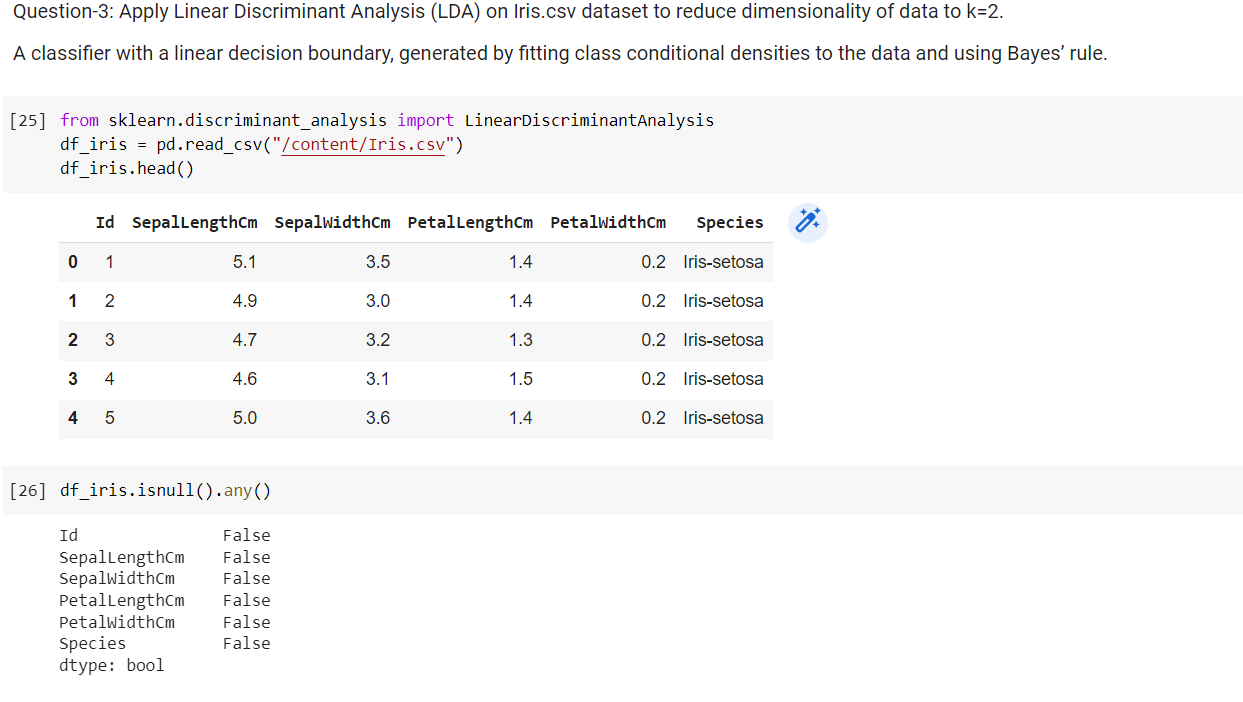


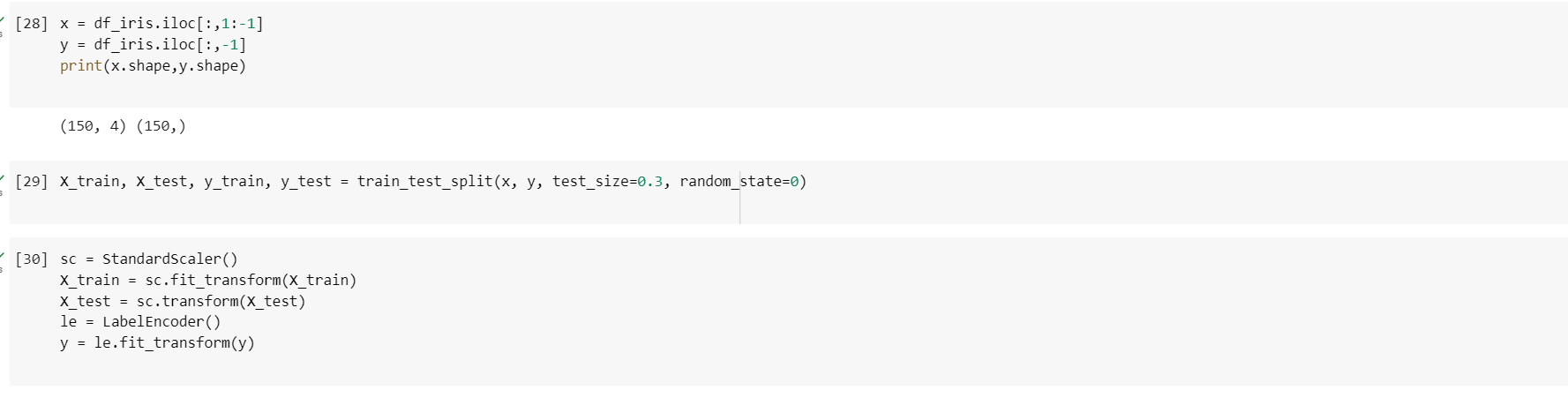






3.Apply Linear Discriminant Analysis (LDA) on Iris.csv dataset to reduce dimensionality of data to k=2.





* Applying the 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

-

PCA (Principal Component Analysis) and LDA (Linear Discriminant Analysis) are both popular techniques in machine learning for dimensionality reduction. However, they have different purposes and methods:

Purpose: PCA is used for unsupervised learning and finds the directions of maximum variance in a dataset. It reduces the number of features by transforming the original dataset into a new coordinate system, where the features are uncorrelated and sorted by their variance. PCA is commonly used for data compression, visualization, and noise reduction. LDA, on the other hand, is used for supervised learning and aims to find the linear combinations of features that best separate the classes. It reduces the number of features by projecting the original dataset onto a lower-dimensional space while maximizing the class separability. LDA is commonly used for feature extraction, pattern recognition, and classification.

Method: PCA operates by finding the eigenvectors and eigenvalues of the covariance matrix of the data. The eigenvectors represent the directions of maximum variance, and the eigenvalues represent the amount of variance explained by each eigenvector. PCA selects the top k eigenvectors, where k is the desired dimensionality of the reduced dataset. LDA, on the other hand, maximizes the between-class scatter and minimizes the within-class scatter of the data. It involves finding the eigenvectors and eigenvalues of the product of two matrices: the between-class scatter matrix and the within-class scatter matrix. LDA selects the top k eigenvectors that correspond to the largest eigenvalues.