Assignment no :2

Name : CHOVATIYA JAY MAHESHBHAI
Student ID : 1161980
Course: COMP5313 Artificial Intelligence
Topic: Data Mining via Dimensionality Reduction on Retail Data

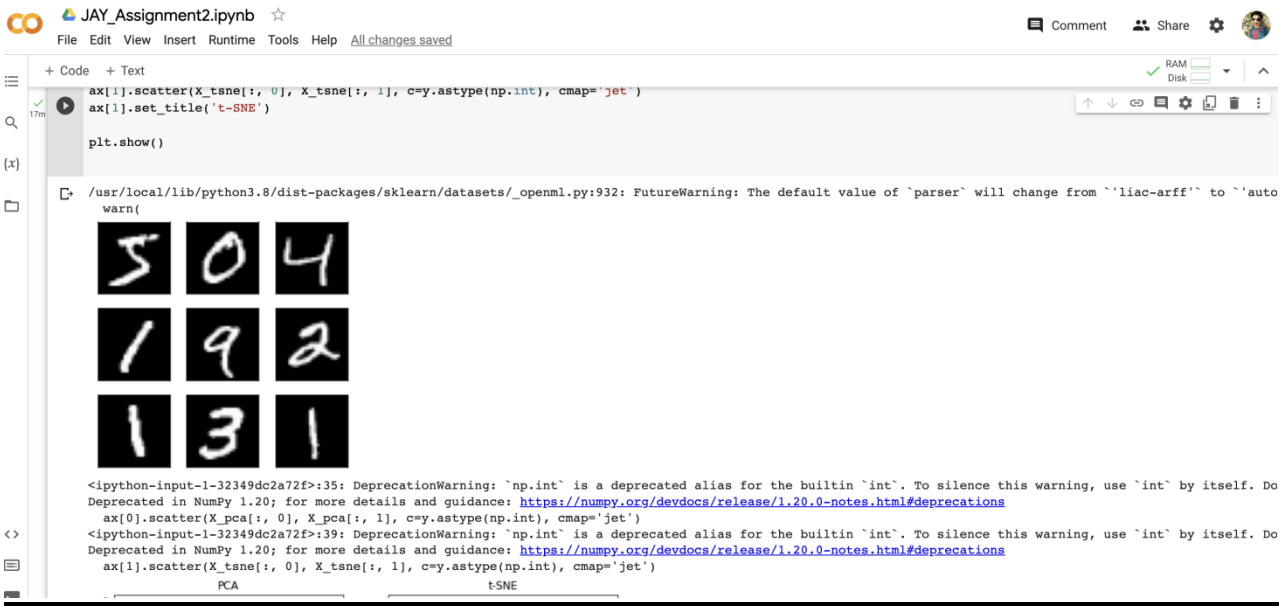
Explanation :

When the question regarding dimensionality reduction is discussed,there are two methods for it.Discussing the principal Component Analysis (PCA),it is a linear dimensionality reduction method which can projects the original high-dimensional information onto a group of orthogonal principal components.In other words,each principal component records the greatest variation in the data along with the number of principal components is set by the user.It can be used on the MNIST dataset to decrease its dimensionality while giving the majority of the data's content.

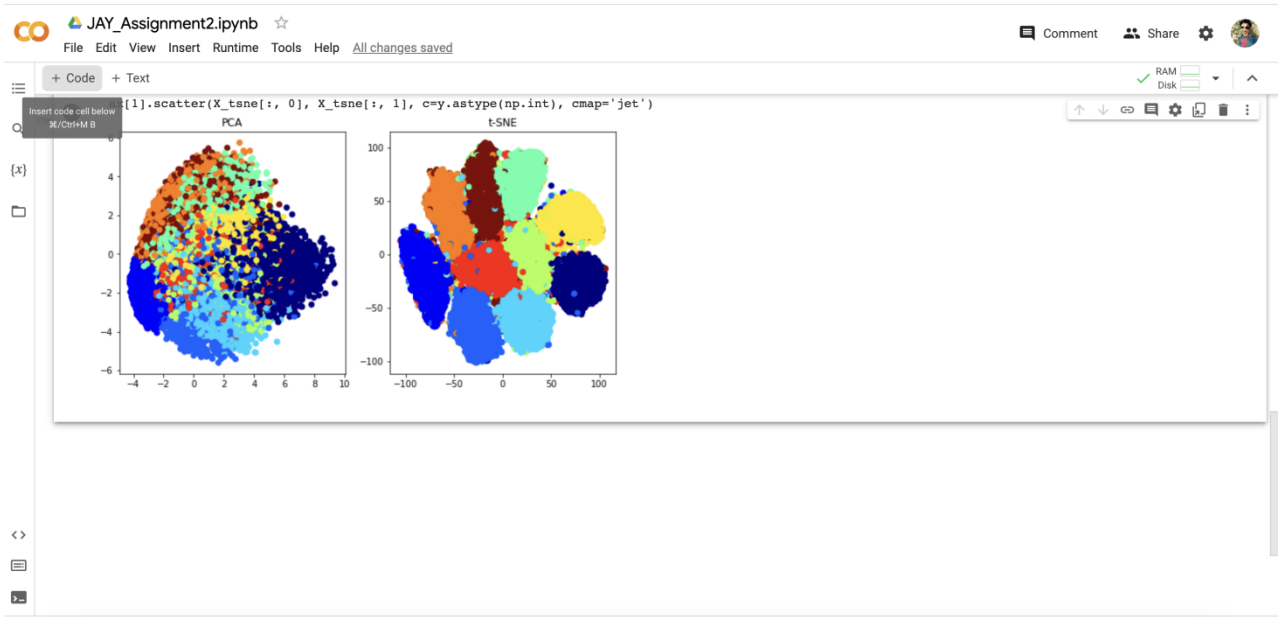
Explaining the (t-SNE) t-distributed Stochastic Neighbor Embedding,it is a nonlinear dimensionality reduction method which maintains pairwise commonalities between the points while mapping the high-dimensional data points to a low-dimensional space.To elaborate it,this method is eminent for retaining local patterns in the data and especially for effective displaying high-dimensional datasets in two or three dimensions.

Furthermore,in terms of comparison between the two methods on the MNIST dataset, t-SNE is a nonlinear method, whereas PCA is a linear method. In addition,PCA is quicker than t-SNE. However, PCA might not be able to retain all of the delicate details in the data, whereas t-SNE is renowned for maintaining local patterns in the data.

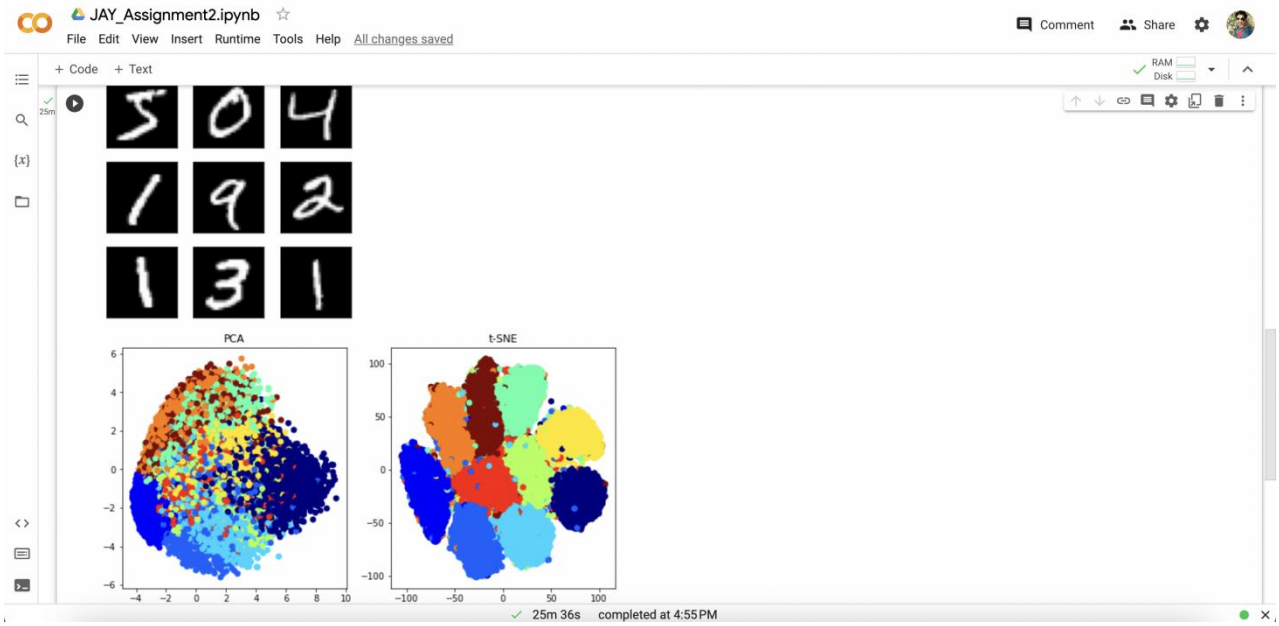
Outputs :1 According to the given question i have attached the output which is display the first nine digits.



Outputs :2



Outputs :3 After removing all the warning output be liked.



Overall, PCA is quicker and better suited to some uses, but t-SNE preserves local data structures and generates high-quality images. The particular needs of the program determine which of the techniques to apply on MNIST dataset.