**ASSIGNMENT 4**

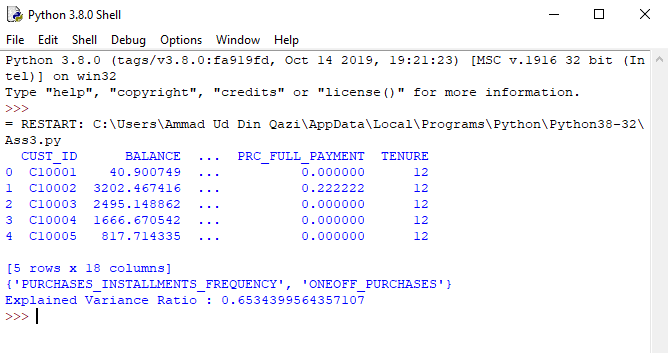
**L16-6343**

**I16-0282**

**L16-6359**

**L16-6344**

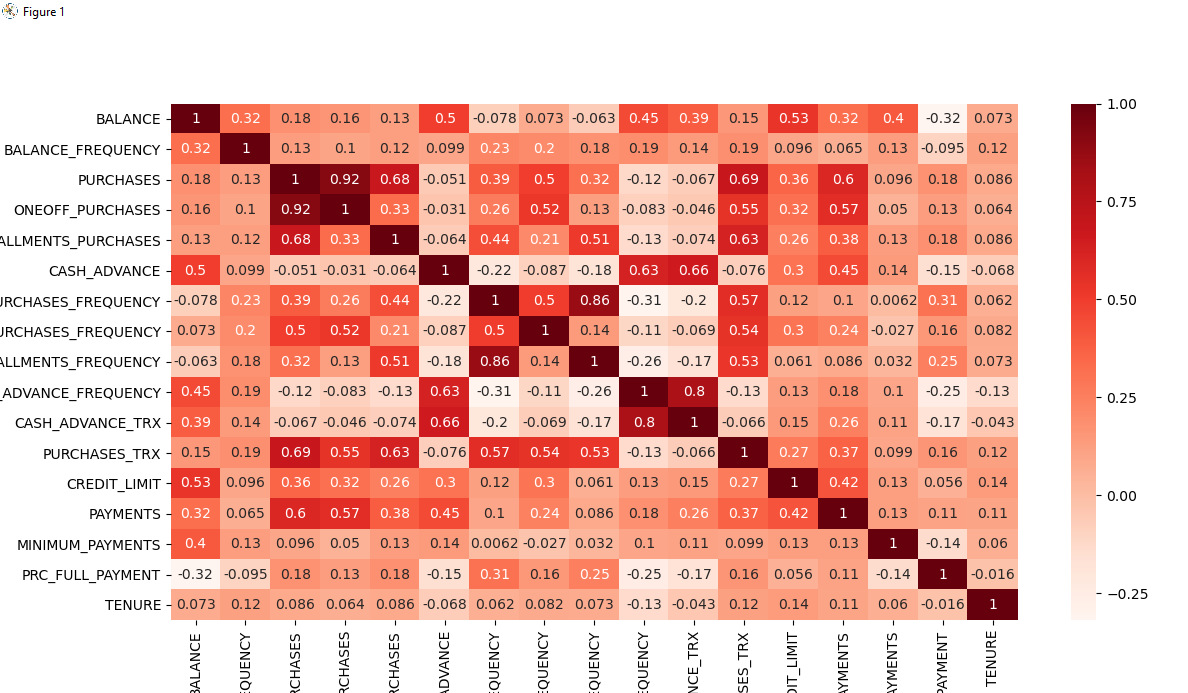
**OUTPUT**

****

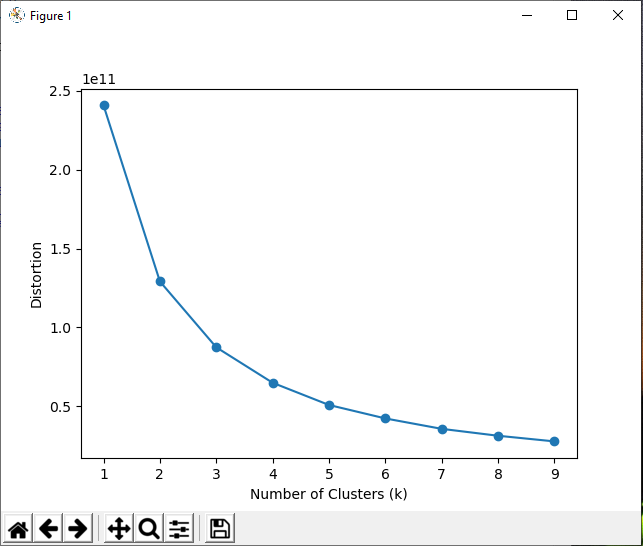
The output shell first shows the data read.

It then shows columns which were co-related and therefore not contributing hence removed.

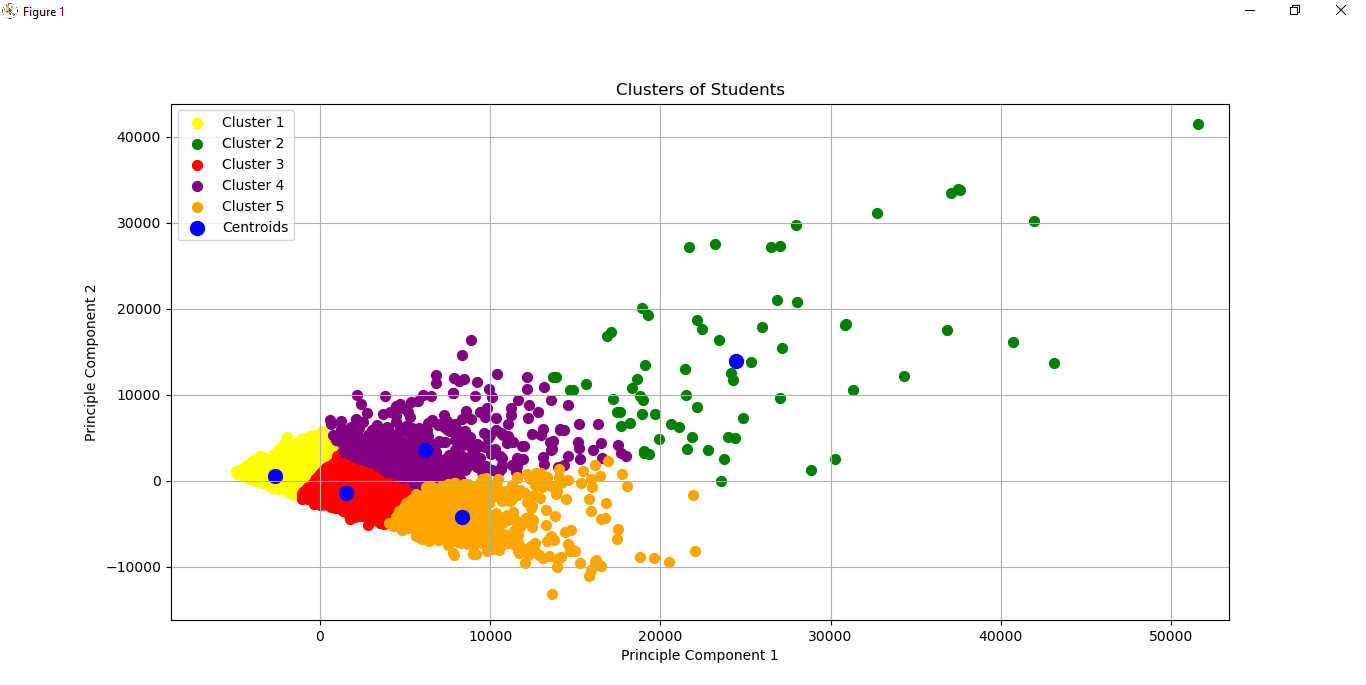
The explained variance ratio for PCA is also displayed

****

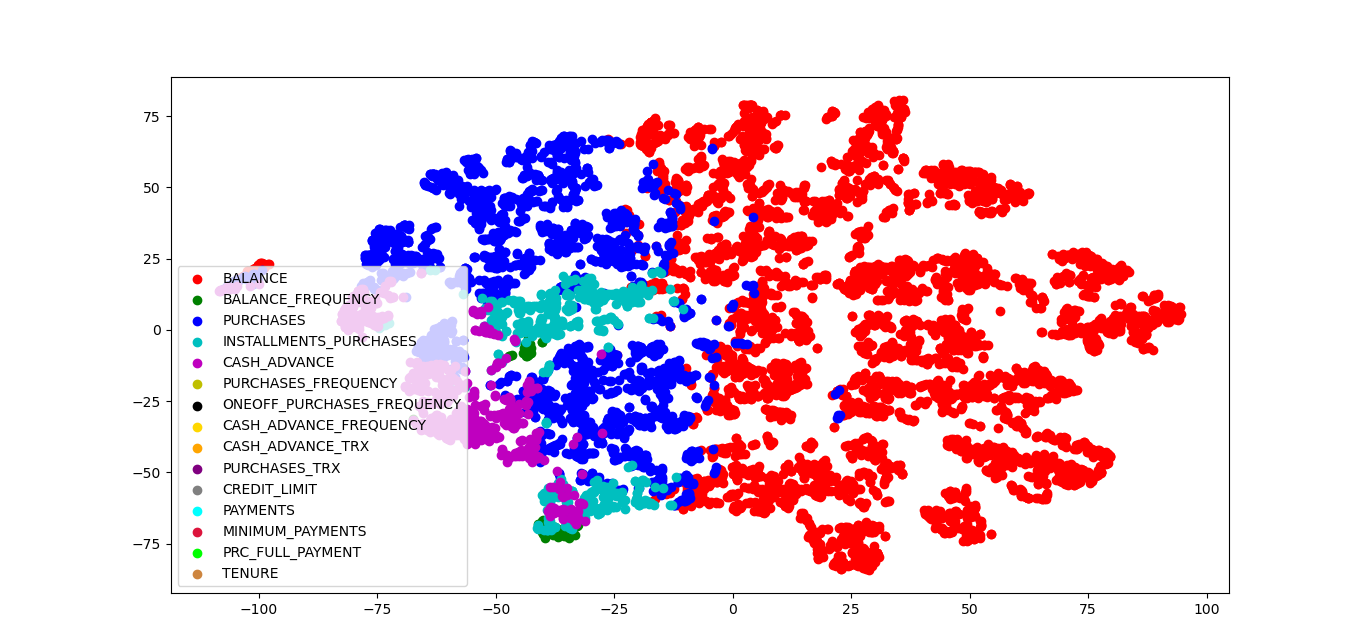
This is the heat map which shows co-relation values of each column. We have set the threshold for co-relation at > then 0.8. The co-related features are then removed from our code.

****

This is the elbow plot for our k-means after removing those features which were not contributing. We use this elbow plot to determine exact value of k which we have used as 5 since decrease is linear after it. After this we again calculate k means using the ideal number of clusters from our elbow plot.

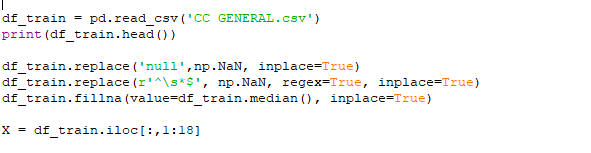


This is our data plotted using the princinpal components. The clusters are highlighted by various colours and the centroids for each cluster have also been displayed.

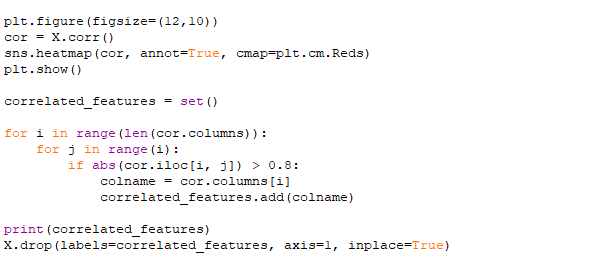
****

This our TSNE Graph. Each of the target is assigned a specific color. You can zoom in on our image to see all the targets values exactly. The key has also been displayed.

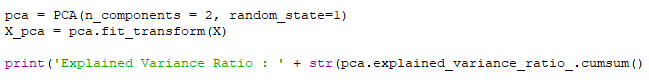
**CODE**

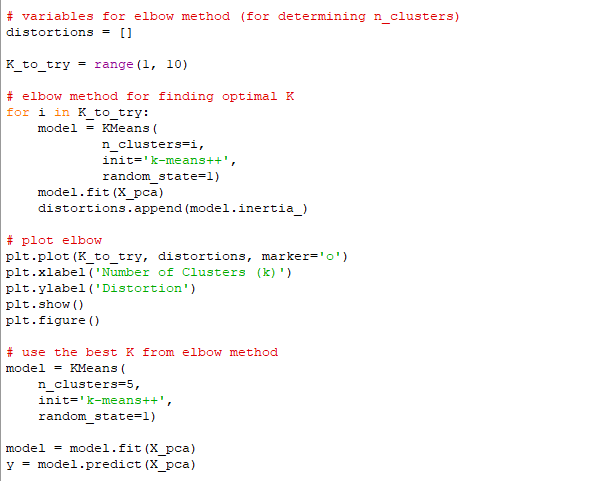


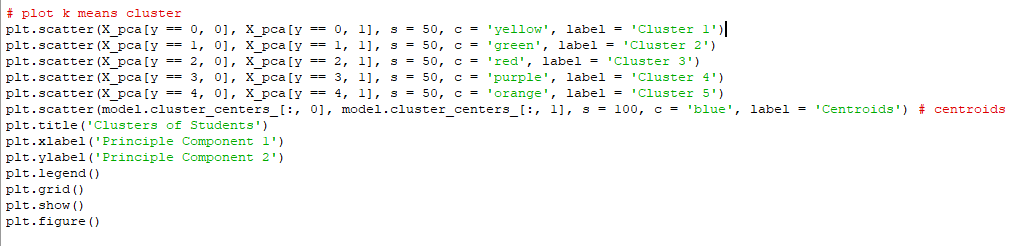
This is the code for reading the data. Since the first column of data customer id is not be used since it is not a feature we have picked data from first column to last column using df\_train.iloc [:,1:18].



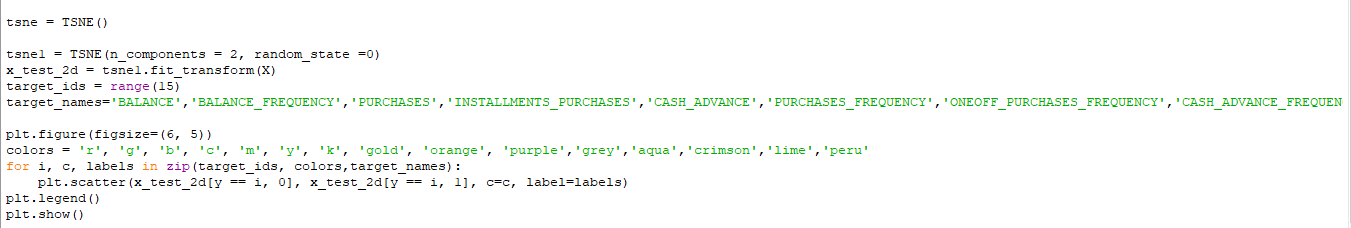
This is the code for displaying the heat map and then finding out columns which are not contributing i.e are co-related. The co-related features are then dropped.







Code for finding principal components and using them to display k-means clusters. For k-means we first use elbow plot method to determine best value of clusters and then use this to calculate k-means.



Code for plotting k-means clusters using TSNE.