## PROBLEM STATEMENT:

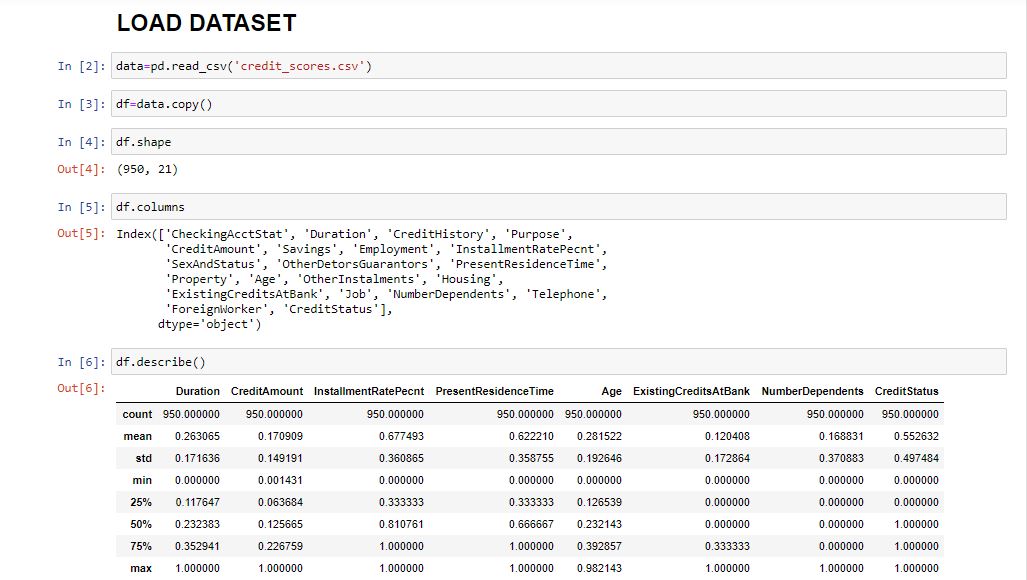
To predict the creditworthiness of the customers and to propose a clustering for these customer.

### INTRODUCTION AND MOTIVATION FOR CREDIT SCORING:

Credit is a very important product in banking and financial institutions. There is always a customer in need of a loan. Since Loans are always accompanied by risks, it is important to identify suitable applicants, and there have to be a means to determine and separate the good applicants from the bad. To solve this issue, financial institutions such as banks started developing credit scores. Using the customer’s credit scores lenders can define the risk of loan applicants. By calculating the credit score, lenders can make a decision as to who gets credit, would the person be able to pay off the loan and what percentage of credit or loan they can get (Lyn, et al., 2002). Lenders generally use “historical” data gathered from customers to build the scorecard for the applicants. They did this by gathering valuable information about candidates like the applicant’s income, type of work, working current place, residual status, financial asset, time with the bank, credit history, if he/she had default or problem with payment. Credit scoring became widely used after the 1980s (Lyn, et al., 2002). In the past, only banks used credit scoring, but then it was extensively used for issuing credit cards, as another kind of loan. Currently, credit scoring is used in credit cards, club cards, mobile phone companies, insurance companies and government departments. Credit scoring is beneficial from both the lenders and customers’ point of view. From the bank’s perspective, it helps them in evaluating potential clients and setting a credit limit based on their credit score. This helps the banks to avoid credit risk. Credit scoring is also a faster process in determining the credit worthiness of a customer, as compared to the traditional method which is time-consuming. From the Page 11 of 64 perspective of the client, they can keep on improving their credit score and extend their credit limit (Mester, 1997). Thus, credit scoring can help avoid unnecessary credit risk to both lender and customer. As per (Mester, 1997), there are three main benefits of credit scoring. The main advantage of credit scoring is that each client is evaluated quickly. Also, since this system is automated, it results in a lot of cost savings to the lenders. As customers need to provide only the information used in the scoring system, applying for credit becomes easy to the customers. Also, this helps lenders to implement the same criteria in making credit decisions to all customers regardless of their gender, race, or other factors. Thus, this process is more objective for all customers and avoids discrimination in any form.

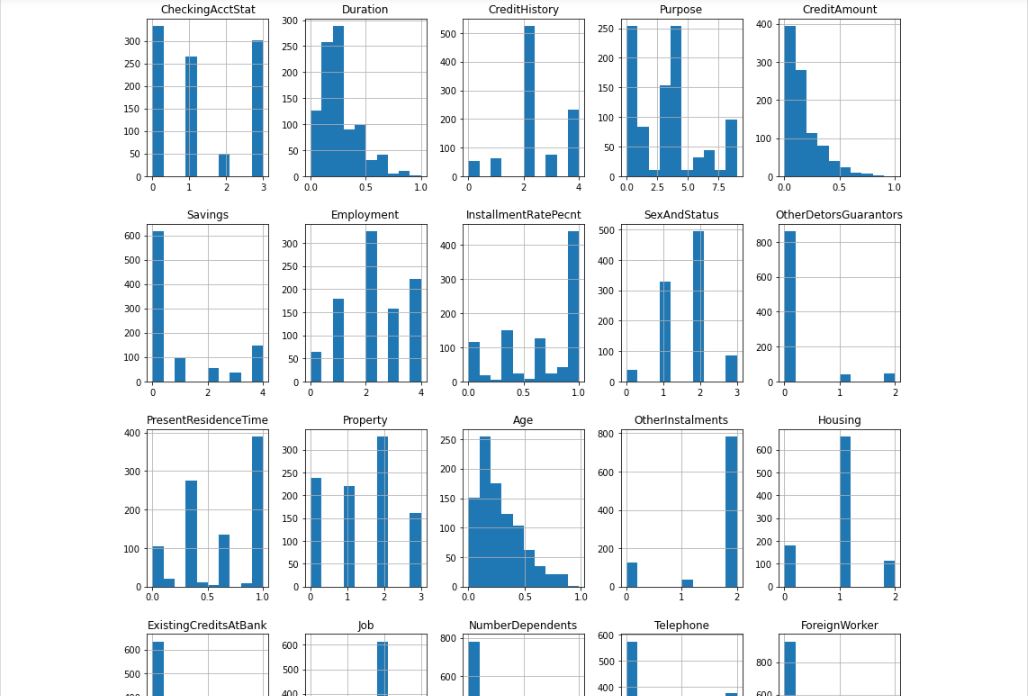
## EXPLORING THE DATA:

Firstly load the csv file into Jupiter.



### DATA VISULAIZATION:

For data visualization I use multiple subplot which is matplotlip pyplot.



Use seaborn for showing heatmap

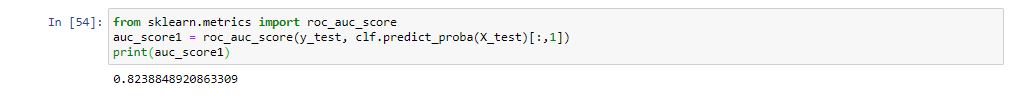


### TEST TRAIN AND ACCURACY:

I use numpy and sklearn model for evaluating the result .The accuracy of the model is 77 percent.



After that the roc accuracy is 82 percent.



### ROC CURVE:

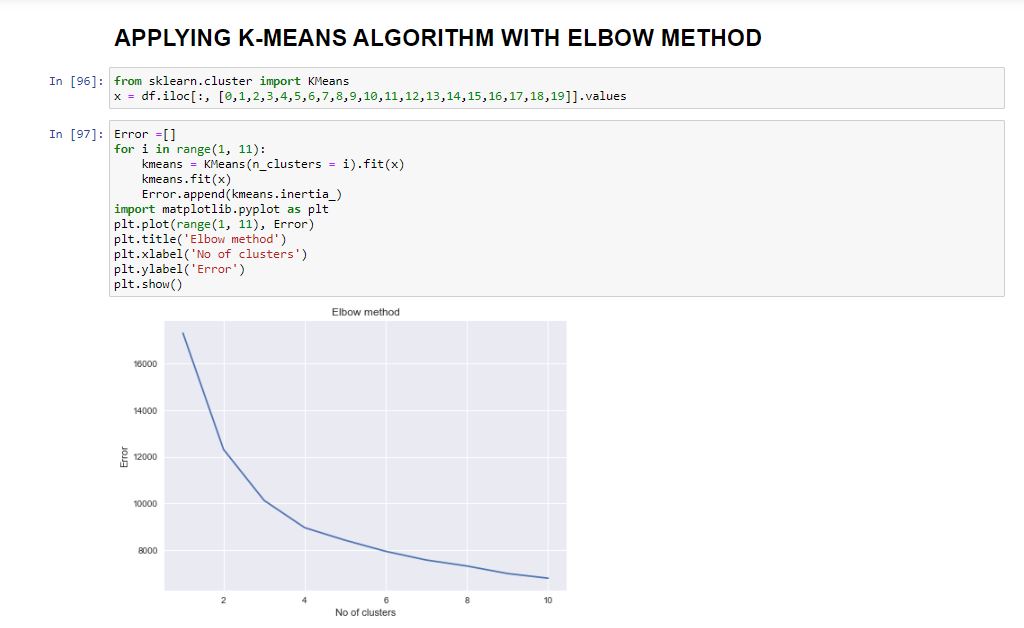
I use Seaborn model for evaluating the graph.



### K-MEANS ALGORITHM AND ELBOW METHOD:

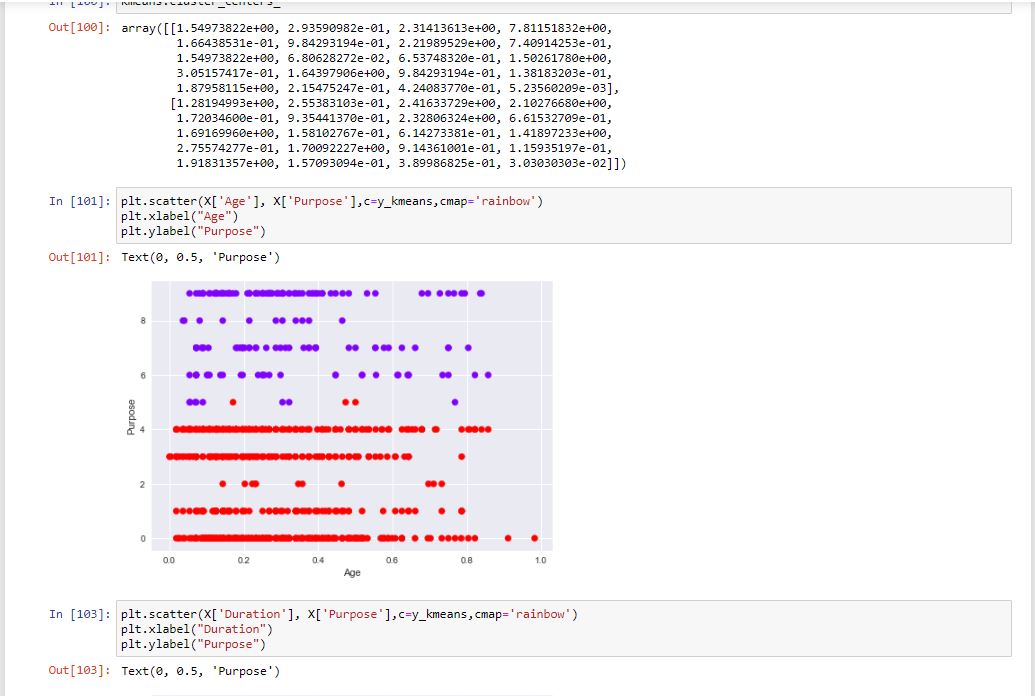
K-means clustering is one of the simplest and popular unsupervised machine learning algorithms. ... In other words, the K-means algorithm identifies k number of centroids, and then allocates every data point to the nearest cluster, while keeping the centroids as small as possible.

Here, K-means is a simple unsupervised machine learning algorithm that groups data into a specified number (k) of clusters. The elbow method runs k-means clustering on the dataset for a range of values for k (say from 1-10) and then for each value of k computes an average score for all clusters.



### K-MEANS CLUSTERING:

K-means clustering is one of the simplest and popular unsupervised machine learning algorithms. ... In other words, the K-means algorithm identifies k number of centroids, and then allocates every data point to the nearest cluster, while keeping the centroids as small as possible.So in this project K-MEANS clustering plot is given below.



Here,

Red - mostly credit status 1 and Blue - credit status 0