**1. Classification**

**- This target here is to determines whether the user can be given a loan or denied. To achieve this we have used the Lending clubs dataset which has 2 separate datasets - Approved Loans and denied Loans.**

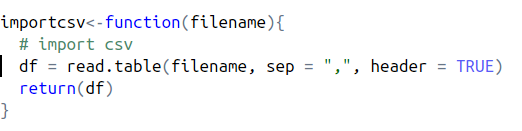
**- We have merged the datasets, cleaned, performed missing value analysis and derived some columns in Part1 of this assignment which will help us in getting the desired result of predicting a flag as Approved and Declined**

**- For this we will perform some classification algorithms as below**

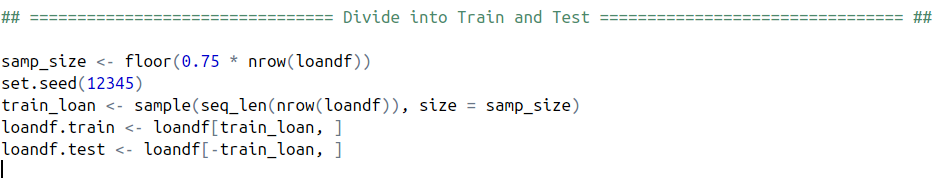
1. **Logistic Regression**
2. **Random Forest**
3. **Neural Network**
4. **SVM**

**Steps :**

- Import the dataset generated from the Part1.



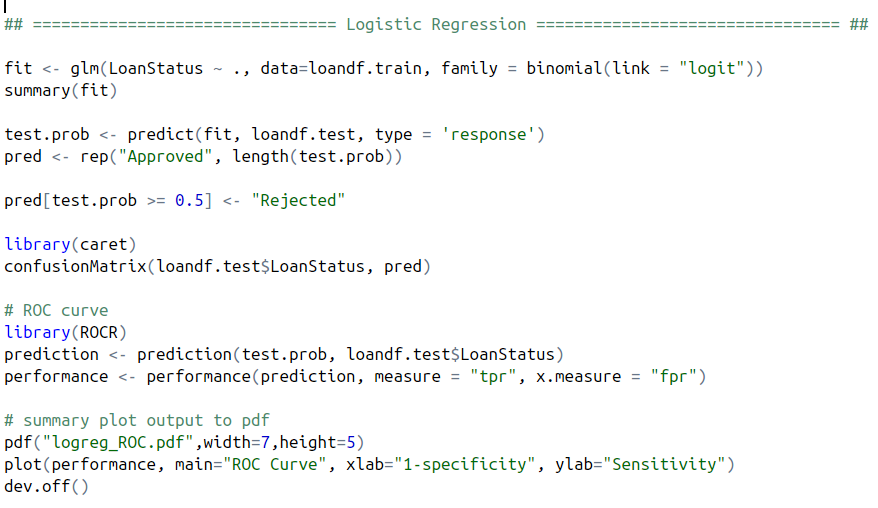
- Once imported, divide the dataset into Train and Test to evaluate and score the models



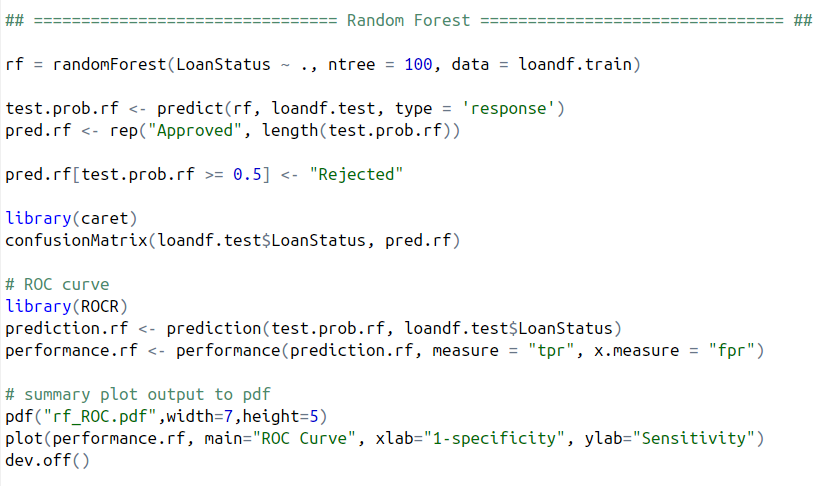
- Now execute the Individual algorithms with the test and train data frames created and evaluate the models

- Based on the results generated from the algorithms we will determine which algorithm to deploy on the Azure ML Studio

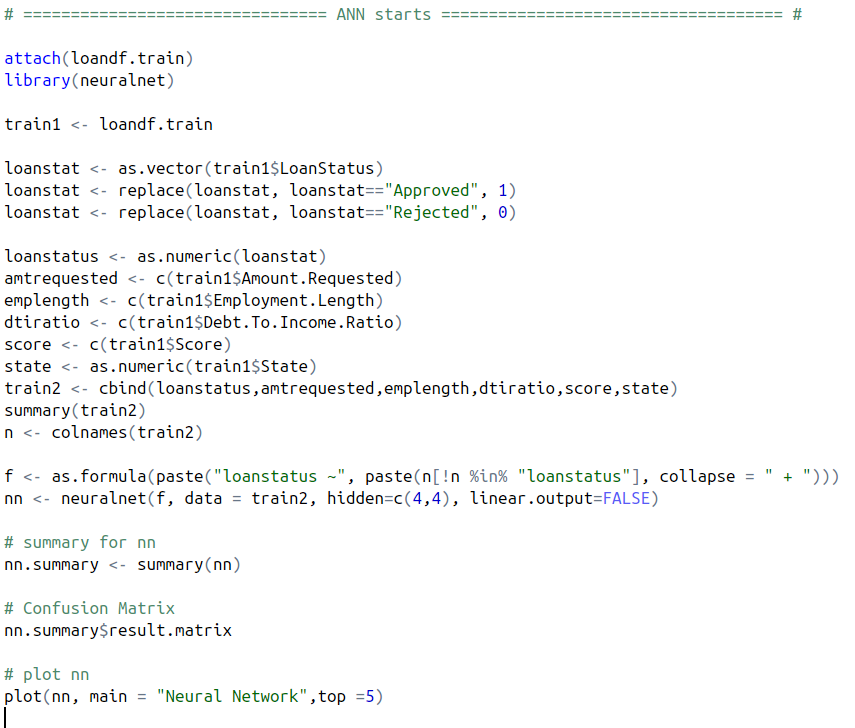
**Logistic Regression :**



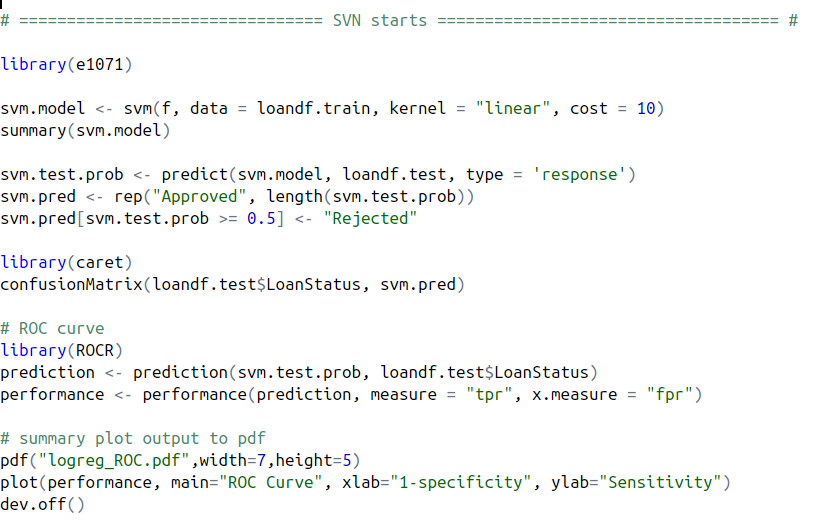
**Random Forest :**



**Neural Network :**



**SVM :**

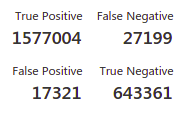
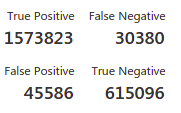


**The results of all the algorithms are as below :**

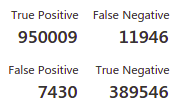
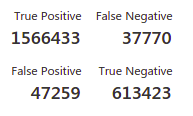
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Logistic Regression** | **Random Forest** | **Neural Network** | **SVM** |
| **Accuracy** | **0.98** | **0.986** | **0.966** | **0.962** |

**Confusion Matrices**

**Logistic Regression Neural Network**

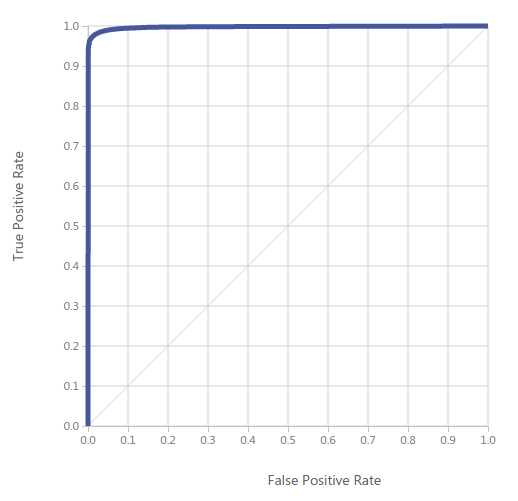
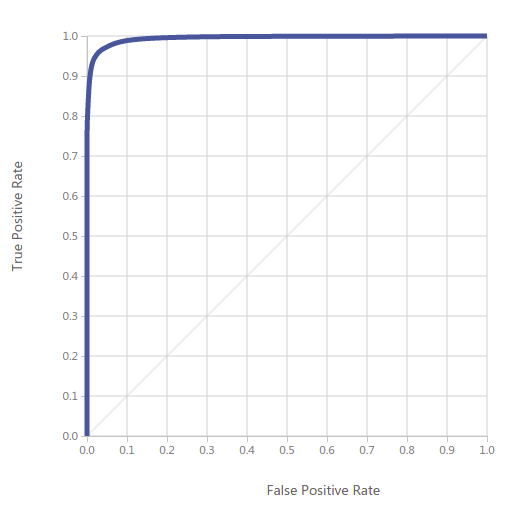
 

**Random Forest SVM**

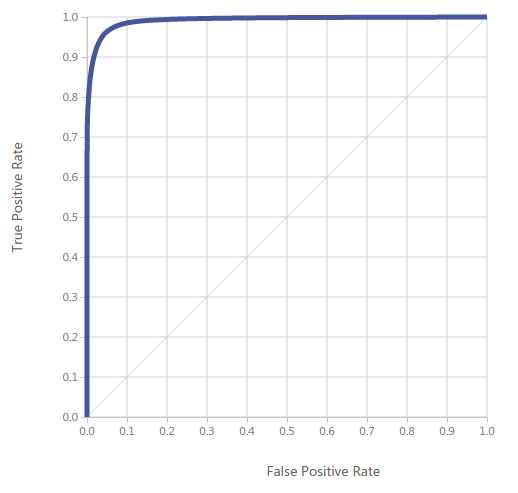
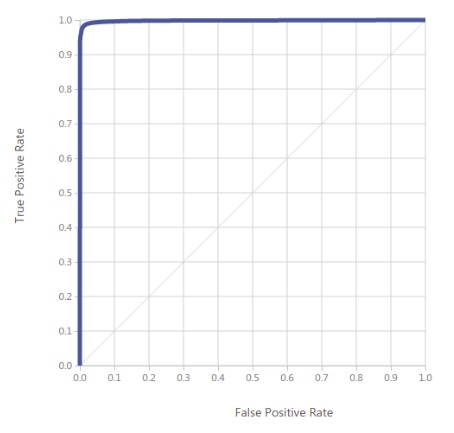
 

**ROC Curves**

**Neural Network Logistic Regression**



**Random Forest** **SVM**



1. **Prediction**

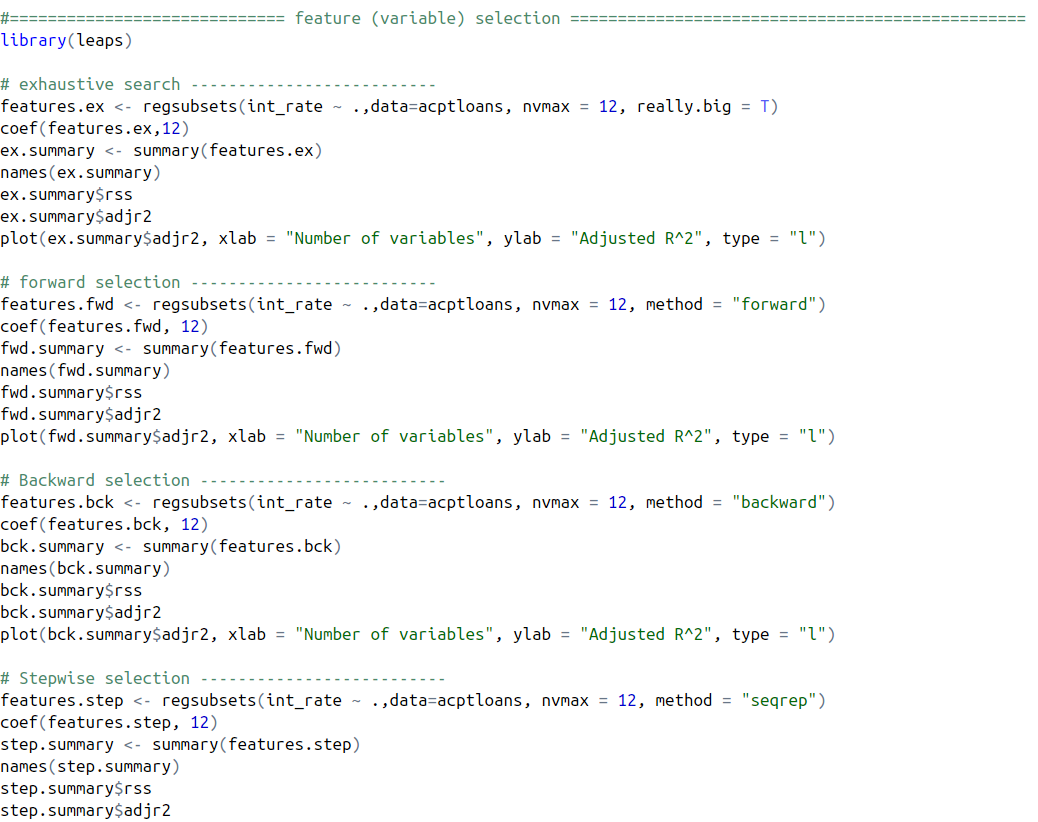
- This step will be executed as a part of the clustering as well. So have documented before the clustering step.

**Steps :**

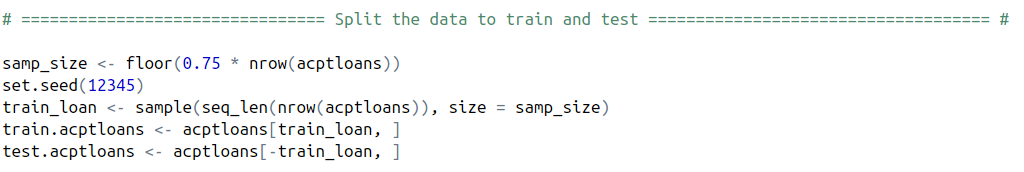
- Import the dataset



- Excute the feature selection process using the exhaustive, forward, backward and stepwise selection process.



- We can the then sample the data for training and testing the model



- This data can now be used on all the clustering methods and we can observe the results

**2. Clustering**

- This part comes into play when we have decide to grant a loan and we need to determine what would be the interest rates.

- To achieve this we have implemented 3 methods for clustering :

A. Manual Clustering

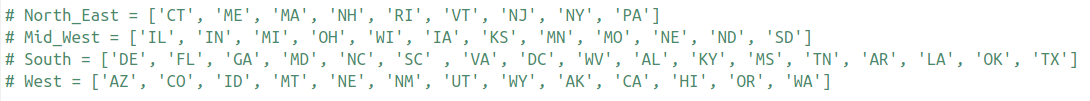
B. Clustering by Algorithm (k-means)

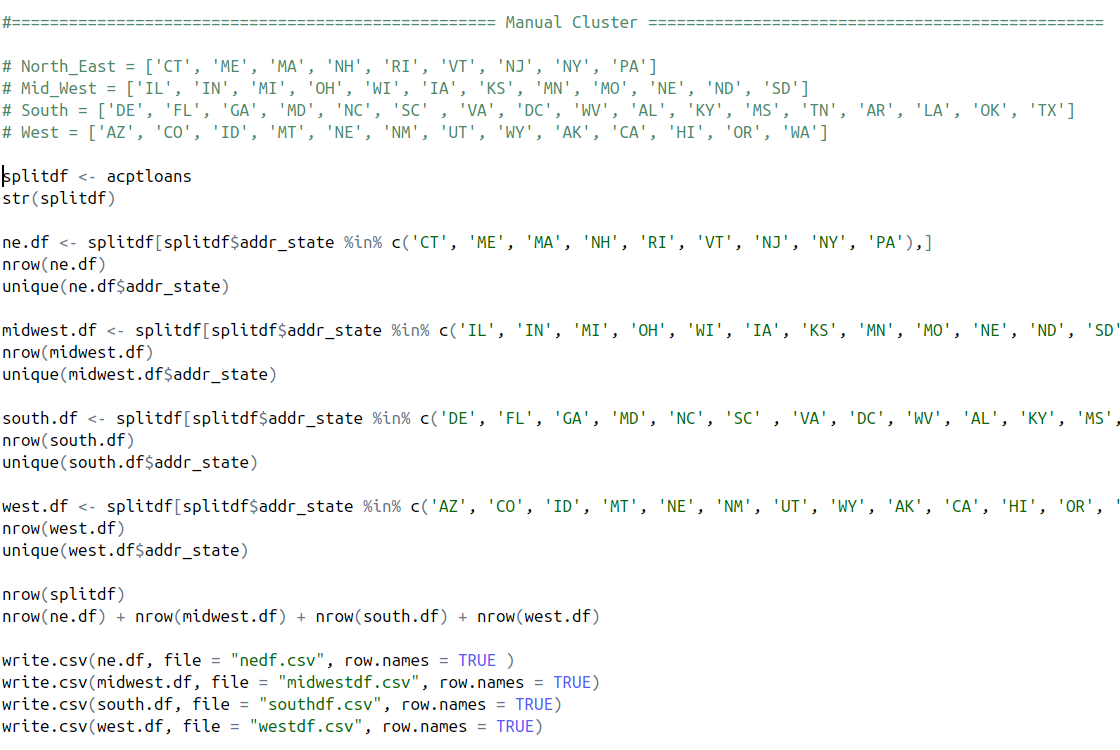
C. Entire Dataset as entire Cluster

**A. Manual Clustering :**

**A.1 - Steps :**

**In this method we have divided the dataset into regions as : Northeast, Mid\_West, South and West**





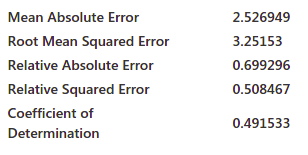
- Once these clusters are ready, it can be treated as an individual dataset and can be split into train and test. We can then run our prediction algorithms to determine which model suits the best and provides us the highest interest rate with a good level of prediction accuracy.

- Below are the results of the Algorithms executed on the individual clusters created manually

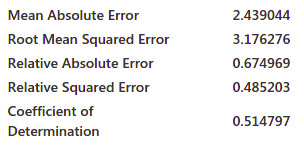
**A.2 - Execution Results :**

**For region Midwest :**

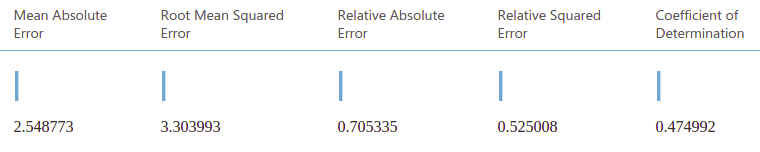
**Linear Regression**



**Neural Network**

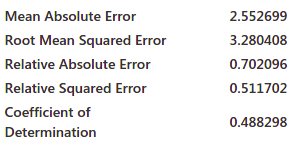


**Random Forest**

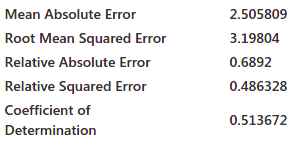


**For region Northeast :**

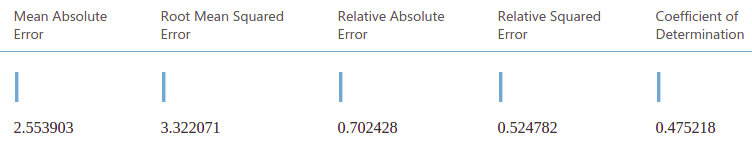
**Linear Regression**



**Neural Network**

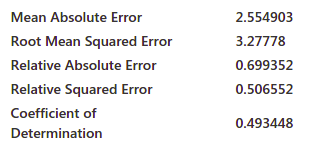


**Random Forest**

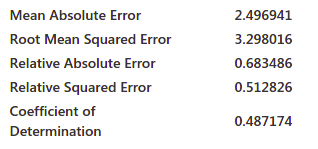


**For region South :**

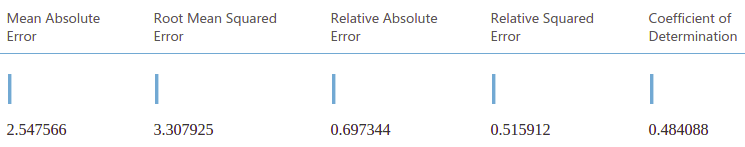
**Linear Regression**



**Neural Network**

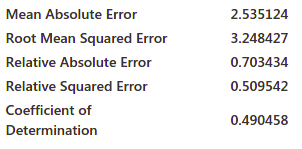


**Random Forest**

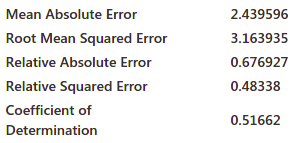


**For region West :**

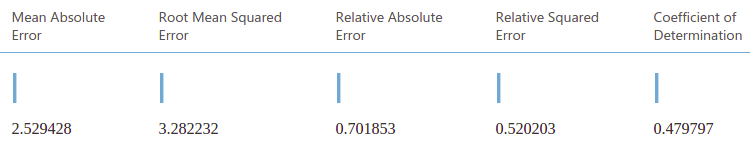
**Linear Regression**



**Neural Network**



**Random Forest**



**To Summarize the Above Results**

- Based on our results of the Algorithms we have deployed the below models individually for the clusters created :

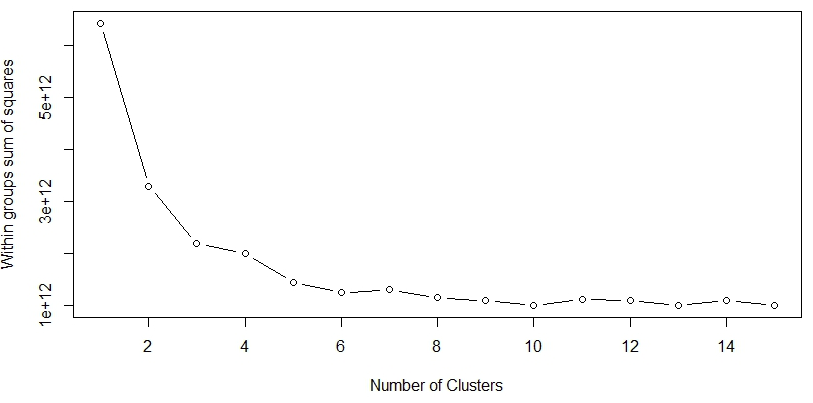
|  |  |
| --- | --- |
| **Region** | **Model (Algorithm)** |
| Mid West Region | Neural Network |
| Northeast Region | Neural Network |
| South Region | Linear Regression |
| West Region | Neural Network |

**B. K - Means Clustering Method**

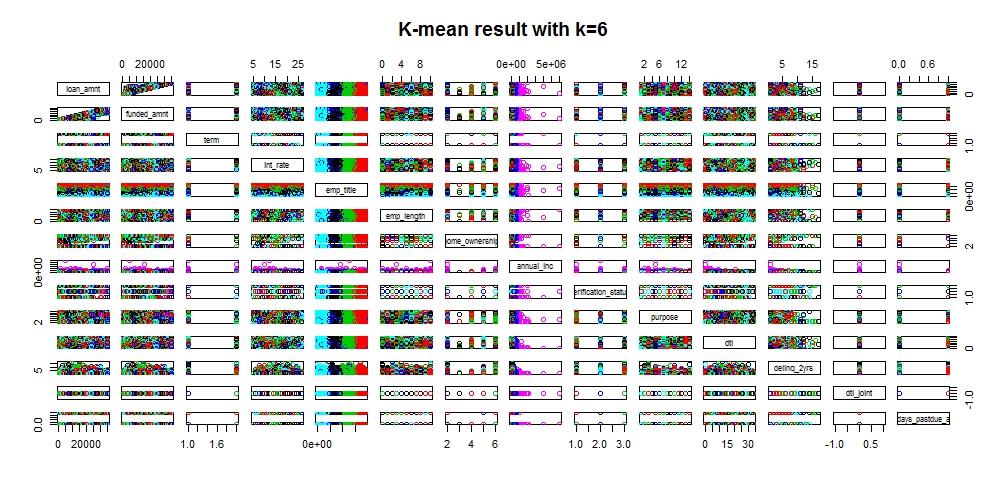
**B.1 - Steps :**

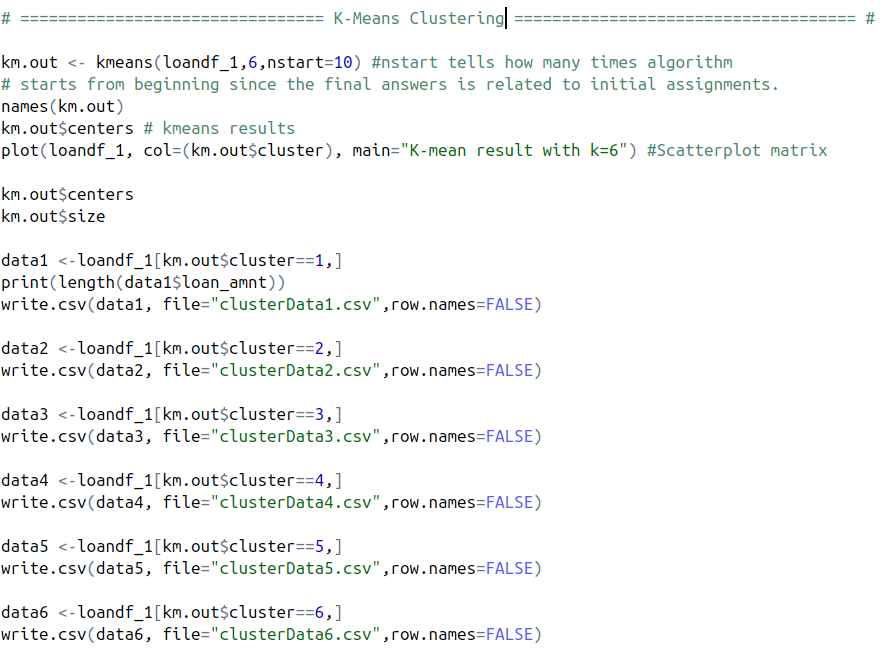
**-** We will generate a Bend graph to determine how many clusters can possibly be generated from the dataset using the k-means clustering algorithm

- The below graph shows that we can create 6 clusters after which the graph flattens out



**-** We execute the k-means clustering algorithm and have generated 6 clusters of data on which now we can individually execute the prediction algorithms to determine the best model for the interest prediction

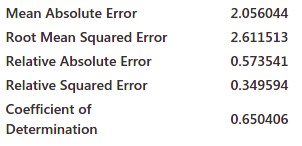
****



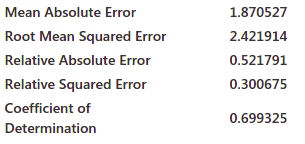
**B.2 - Execution Results :**

**Cluster 1 :**

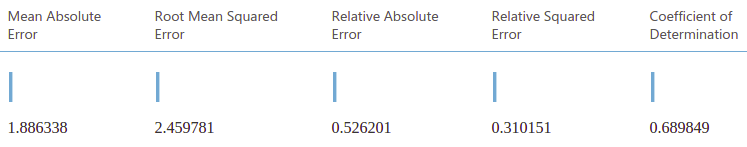
**Linear Regression**



**Neural Network**

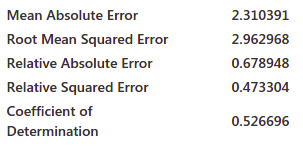


**Random Forest**

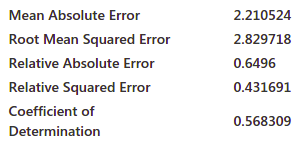


**Cluster 2 :**

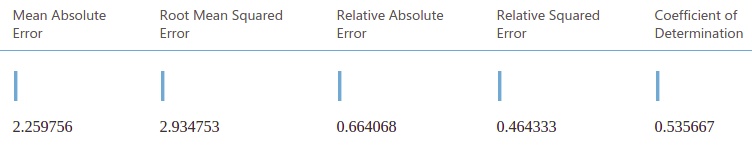
**Linear Regression**



**Neural Network**

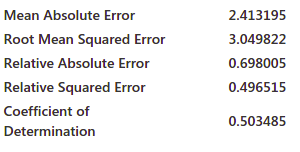


**Random Forest**

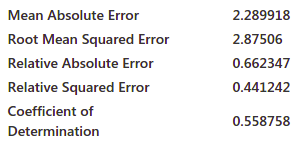


**Cluster 3 :**

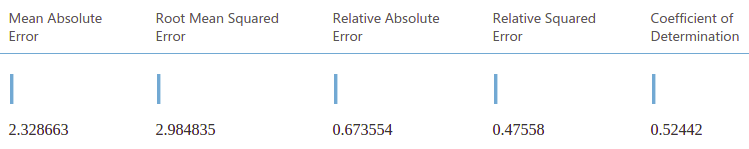
**Linear Regression**



**Neural Network**

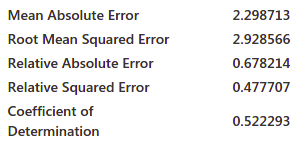


**Random Forest**

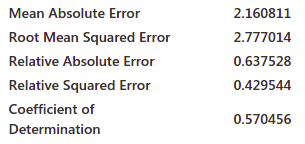


**Cluster 4 :**

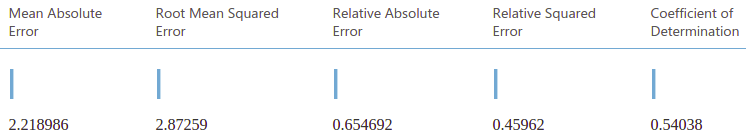
**Linear Regression**



**Neural Network**

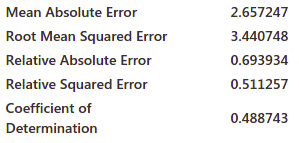


**Random Forest**

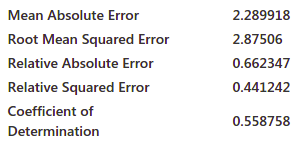


**Cluster 5 :**

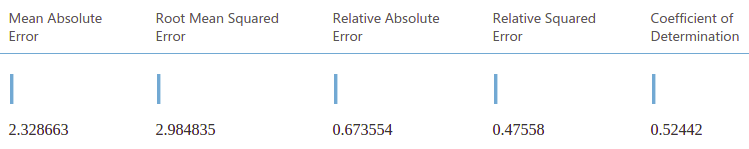
**Linear Regression**



**Neural Network**

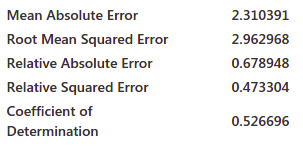


**Random Forest**

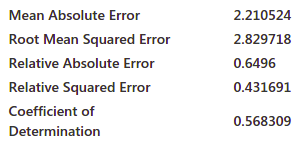


**Cluster 6 :**

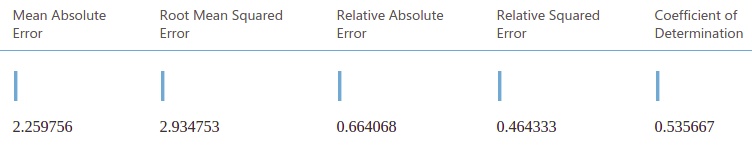
**Linear Regression**



**Neural Network**



**Random Forest**

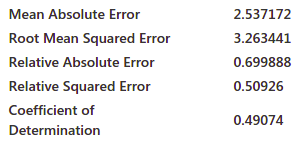


**C. Entire Dataset as entire Cluster :**

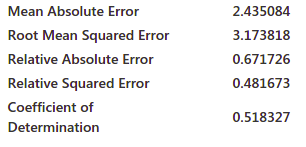
**C.1 - Steps :**

**C.2 - Execution Results :**

**Linear Regression**



**Neural Network**



**Random Forest**

