CSE445.3

Assignment Report

Name: Majharul Alam Evan

ID no. 1621460642

Section No.03

Problem No-01

 Model Name:

Decision Tree using bagging classifier

Bagging classifier is a machine learning meta-algorithm designed to improve the stability and accuracy of machine learning algorithms used in statistical classification and regression. That's why we use this model for this problem.
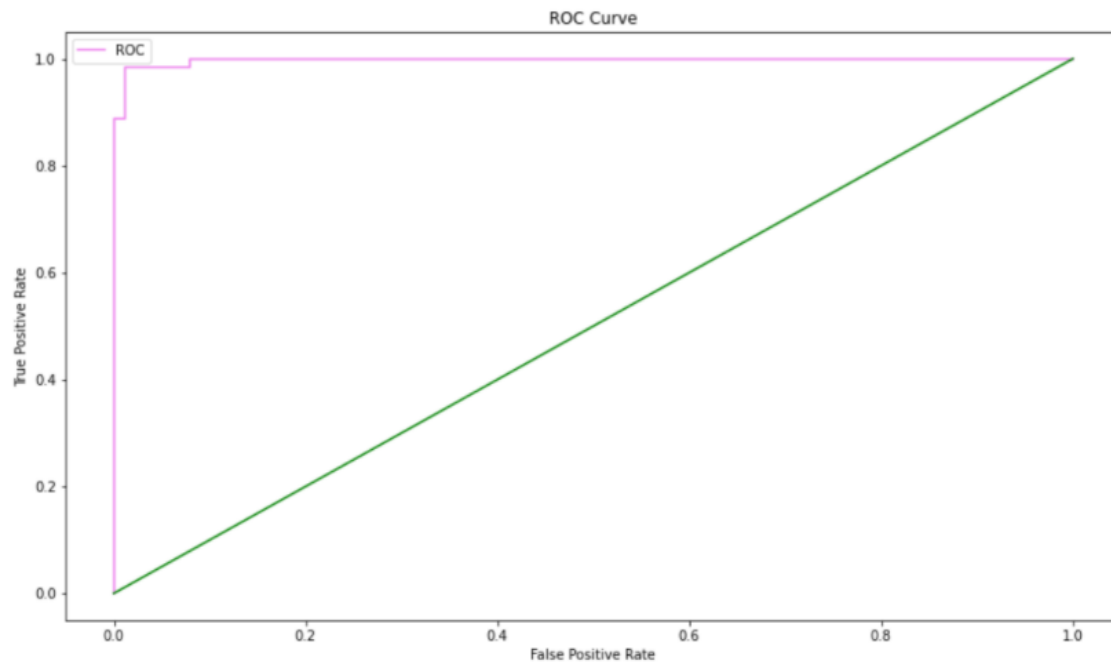
Correlation Matrix:

```
Index(['Risk_A', 'Risk_B', 'TOTAL', 'Score_B.1', 'Risk_C', 'Risk_D', 'RiSk_E',
       'Risk_F', 'Score', 'Marks', 'MONEY_Marks', 'District', 'Loss',
       'LOSS_SCORE', 'History_score', 'Risk'],
      dtype='object')
```

In here, we show the correlation matrix output where we use correlation matrix in columns.

Training and Testing Score:

```
Training Score:   0.9981203007518797
Testing Score:   0.9780701754385965
```

ROC CURVE:

Confusion matrix:

In this problem, There will be no Confusion matrix. Because It has continuous variable and confusion matrix is not for continuous variable.

Graph:

PROBLEM- 02

## Model Name:

Linear Regression(Multivariable)

Linear Regression is a supervised machine learning algorithm with a constant slope and continuous predicted output. Rather than attempting to categorize values, it is used to forecast values within a continuous range That's why linear regression to solve this data.

Data Figure graph:

sell total from 2014 to 2016

Sales Model
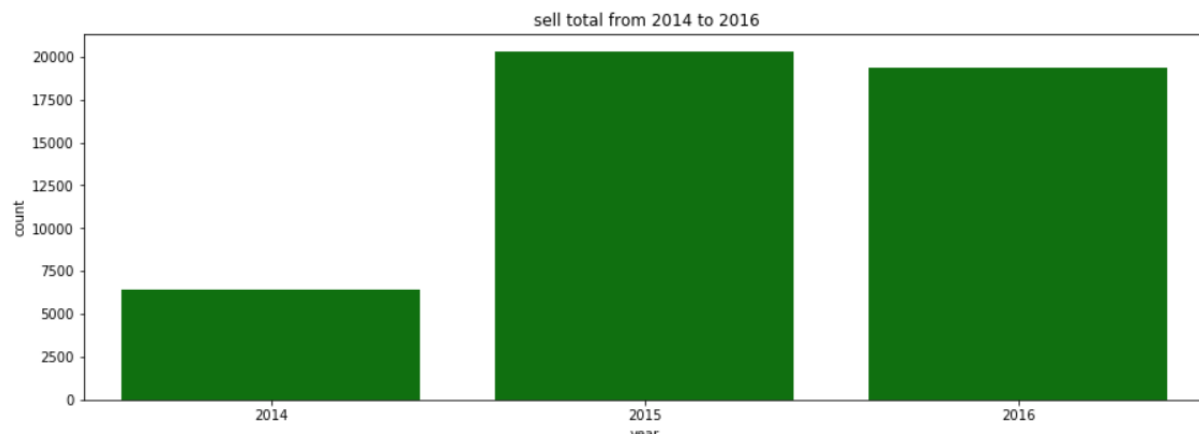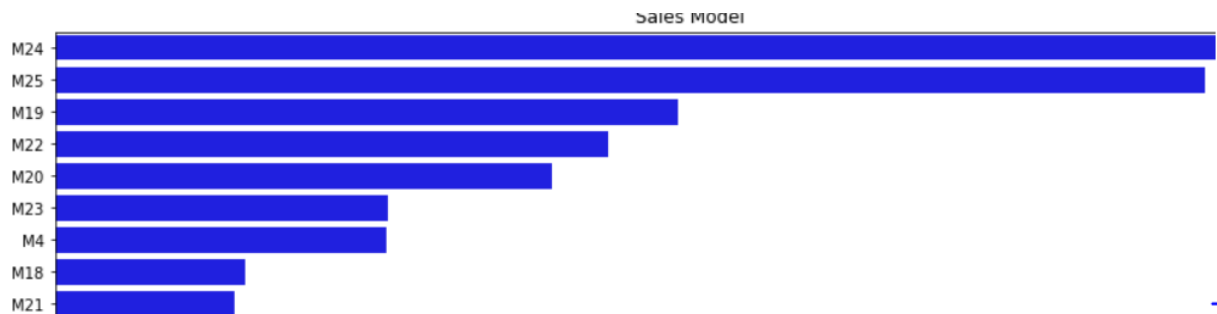
Correlation matrix:
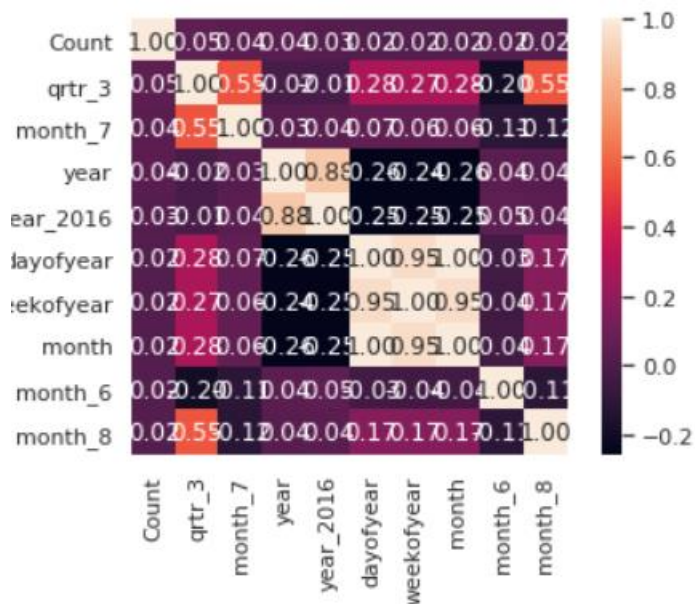
Training Score:

```
MAE 127.82861128429762
R2 0.007170769037891711
RMSE 316.74550691906074
```

Testing Score:

```
MAE 123.78902132567079
R2 0.005584104093325437
RMSE 307.48242895239366
```

ROC Curve:

As it is a linear regression model there will be no Roc curve because The ROC curve represents the trade-off between these two (particularly, the false positive and true positive rates).... There isn't a TP/FP trade-off either. Instead, individuals use a loss function to assess a regression model's performance, which indicates how good or bad a certain quantity of error is.