**METRO INTERSTATE TRAFFIC VOLUME**

**PROJECT REPORT**

**Group Members:**

**Vansh Mehta (20csu177)**

**Nishit Bhatnagar (20csu175)**

**Prithish Soni (20csu176)**

**Table Of Contents**

**1.Introduction**

1.1 Background

1.2 Objective

1.3 Scope

**2.Methodology**

2.1 Data Collection

2.2 Data Analysis

2.3 Statistical Techniques

2.4 Model Development

**3.Data Collection**

3.1 Data Sources

3.2 Variables Considered

3.3 Data Pre-processing

**4.Data Analysis**

4.1 Descriptive Analysis

4.2 Exploratory Analysis

4.3 Correlation Analysis

**5.Statistical Techniques**

5.1 Time Series Analysis

5.2 Regression Analysis

5.3 Feature Importance

**6.Model Development**

6.1 Model Selection

6.2 Model Training

6.3 Model Evaluation

**7.Results**

7.1 Analysis of Traffic Volume Patterns

7.2 Factors Influencing Interstate Traffic Volume

7.3 Model Performance Evaluation

**8.Conclusion**

8.1 Summary of Findings

8.2 Implications and Recommendations

**9.References**

**Introduction**

**1.1 Background:**

Interstate traffic volume in metropolitan areas plays a crucial role in transportation planning and infrastructure development. Understanding the factors affecting traffic volume can help in optimizing transportation systems, reducing congestion, and improving overall efficiency.

**1.2 Objective:**

The objective of this project is to analyse the metro interstate traffic volume and identify the key factors influencing it. Additionally, a predictive model will be developed to forecast future traffic volume based on these factors.

**1.3 Scope:**

The project focuses on a specific metropolitan area and its interstate highways. The analysis will consider various variables such as time of day, day of the week, weather conditions, holidays, and other relevant factors affecting traffic volume.

**Methodology**

**2.1 Data Collection:**

Data will be collected from various sources, including transportation departments, traffic monitoring systems, and weather databases. The data will cover a specific time period, capturing traffic volume at regular intervals along with associated variables.

**2.2 Data Analysis:**

The collected data will undergo thorough analysis to identify patterns, trends, and correlations. Descriptive and exploratory analysis techniques will be employed to gain insights into the dataset.

**2.3 Statistical Techniques:**

Time series analysis will be utilized to study traffic volume patterns over time. Regression analysis will be employed to identify the relationship between traffic volume and various independent variables. Feature Importance has been done with the help of PyCaret.

**2.4 Model Development:**

Based on the analysis results, a predictive model will be developed using appropriate modelling techniques. The model will be trained using historical data and validated to assess its performance and accuracy.

**Data Collection**

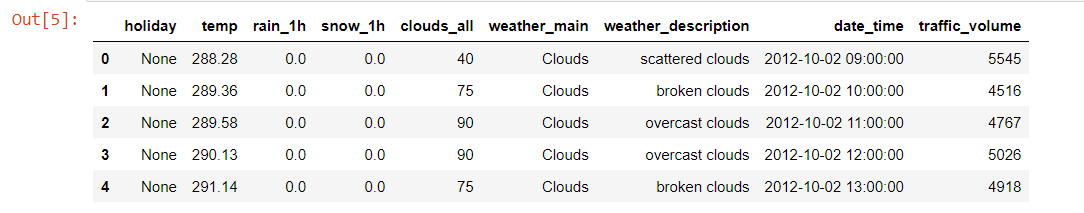
**3.1 Data Sources:**

The main source of data is from Kaggle. The link to dataset is given as follows-

<https://www.kaggle.com/code/ramyahr/metro-interstate-traffic-volume/input>

Data will be collected from transportation agencies, traffic monitoring stations, and meteorological databases. Historical traffic volume data, weather conditions, and relevant variables will be obtained.

**3.2 Variables Considered:**

****

traffic\_volume (Target)

holiday

temp

rain\_1h

snow\_1h

clouds\_all

weather\_main

weather\_description

date\_time

**3.3 Data Pre-processing:**

The collected data will undergo pre-processing steps, such as data cleaning, outlier detection, and missing value imputation. The data will be organized and prepared for further analysis.

**Data Analysis**

**4.1 Descriptive Analysis:**

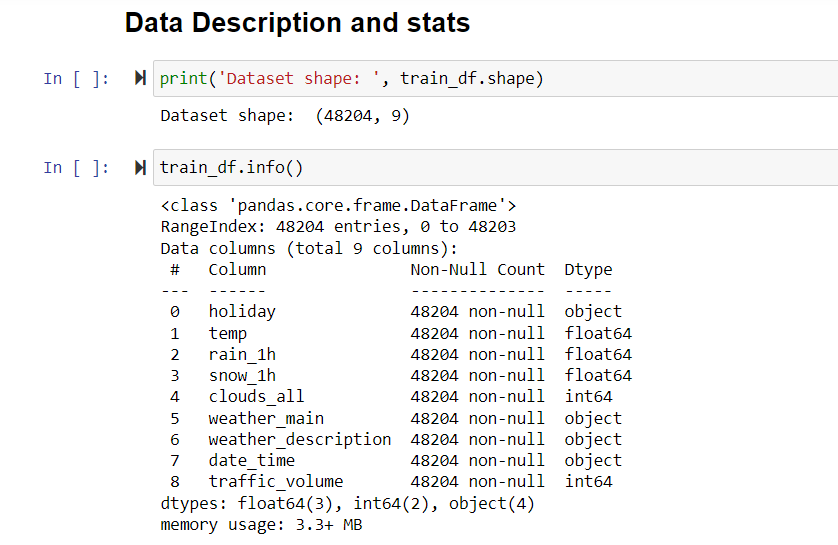
Descriptive statistics will be used to summarize and describe the dataset. This will include measures of central tendency, dispersion, and graphical representations of the data.

**4.2 Exploratory Analysis:**

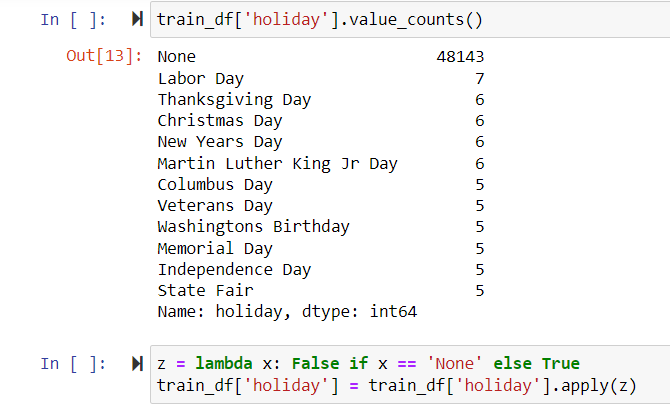
Exploratory analysis techniques, such as data visualization, will be employed to identify patterns and relationships within the dataset. This will provide initial insights into the factors influencing metro interstate traffic volume.

**4.3 Correlation Analysis:**

Correlation analysis will be conducted to determine the strength and direction of relationships between traffic volume and independent variables. This will help in understanding the impact of each variable on traffic volume.

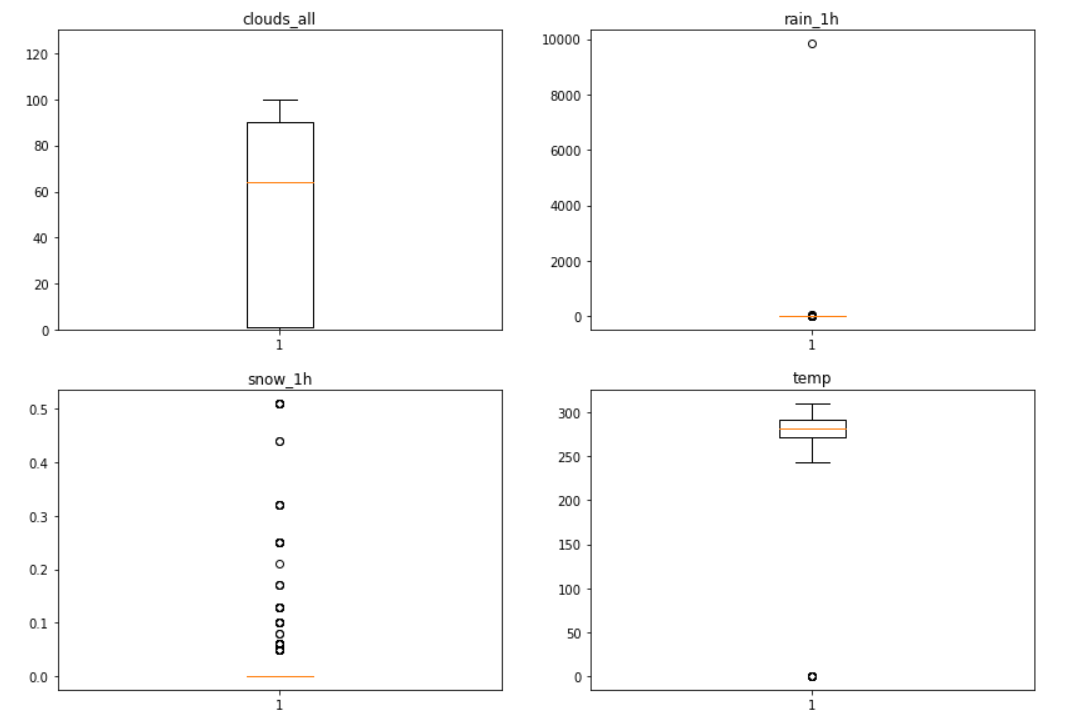


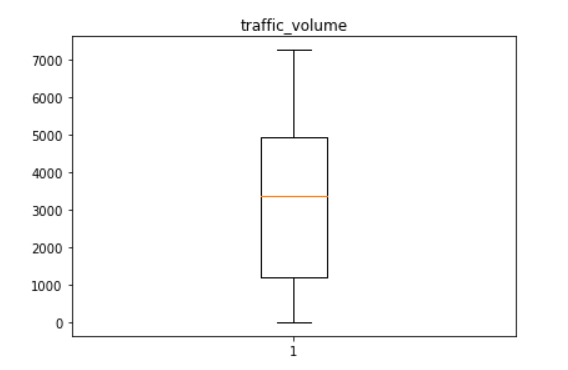


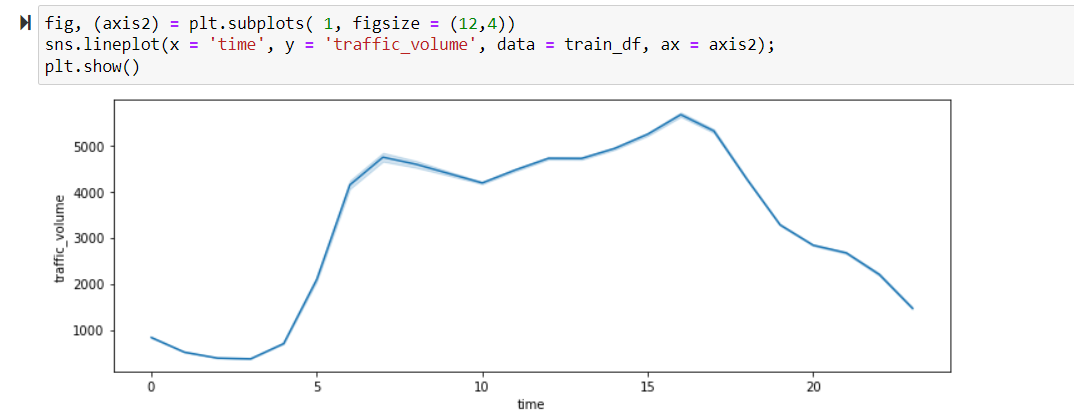


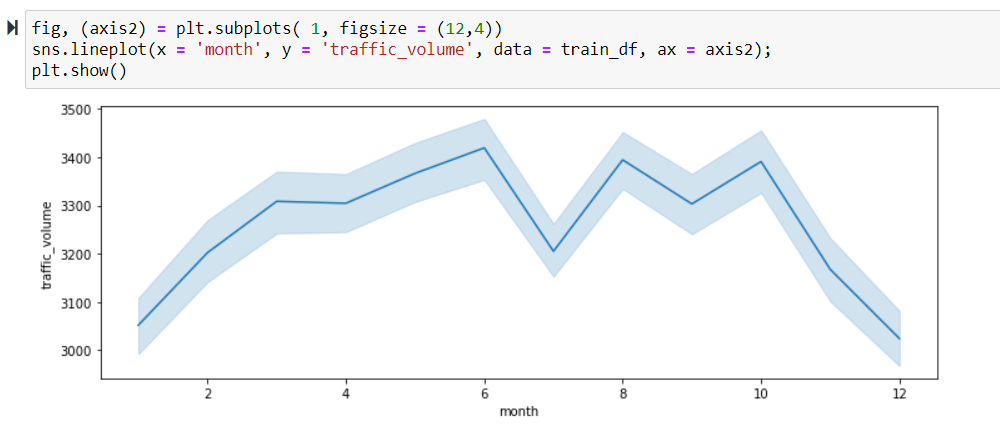


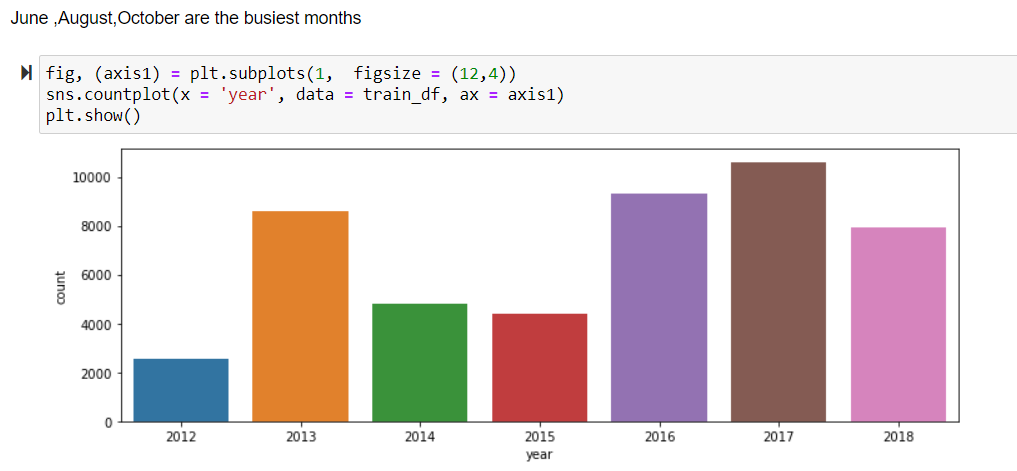


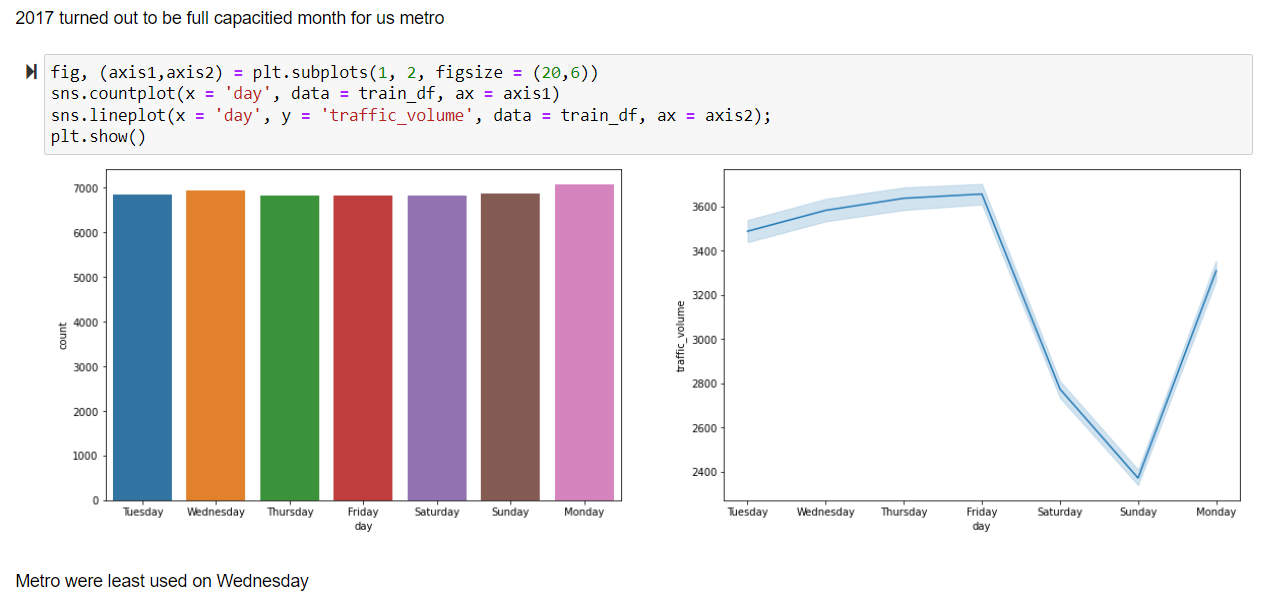


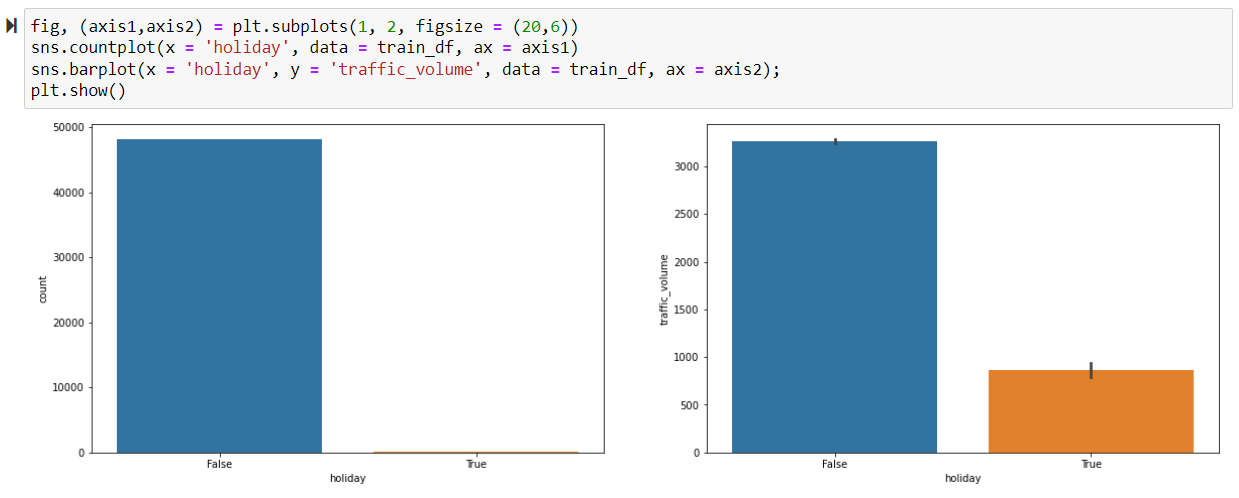


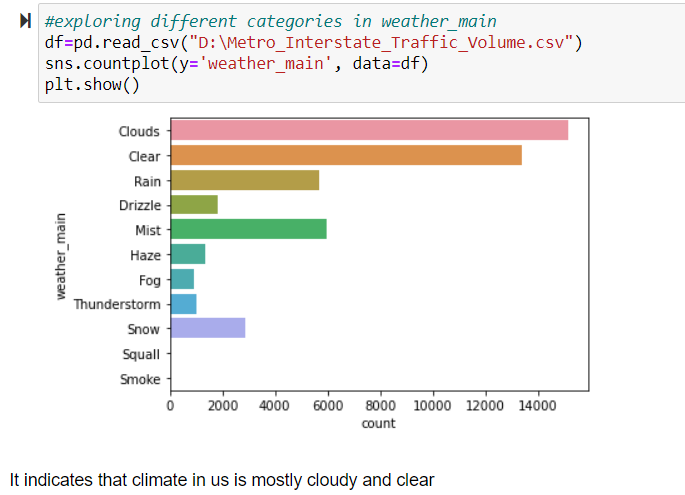


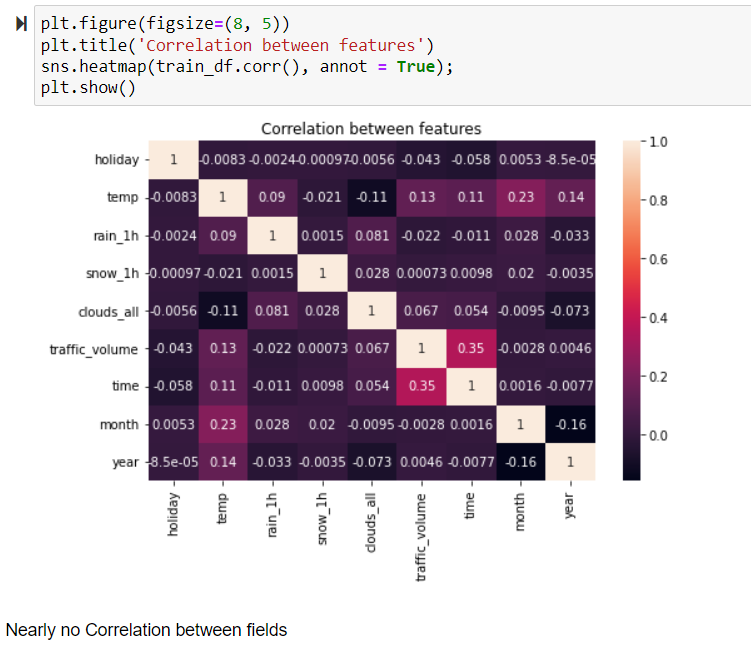


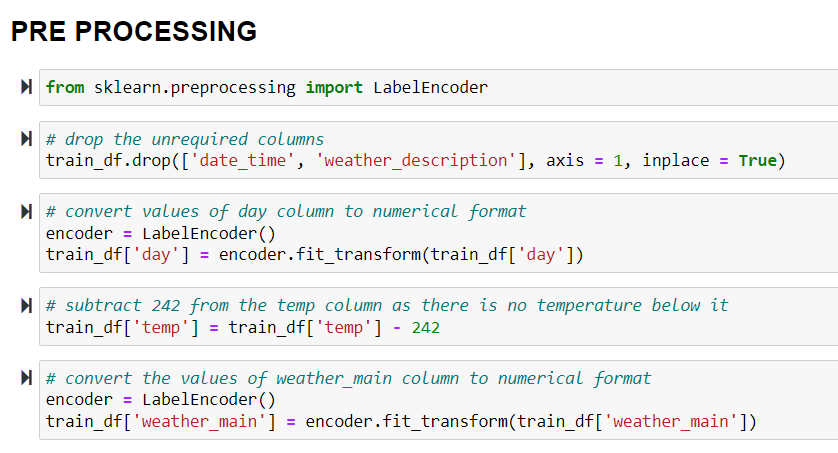












**Statistical Techniques**

**5.1 Time Series Analysis:**

Time series analysis will be used to examine the temporal patterns and trends in metro interstate traffic volume. Techniques such as decomposition, trend analysis, and seasonality detection will be employed.

**5.2 Regression Analysis:**

Regression analysis will be applied to establish the relationship between traffic volume and independent variables. Multiple regression models will be developed to quantify the influence of each variable and predict traffic volume.

**5.3 Feature Importance:**

To check which feature is contributing how much, which one is contributing most and which one is contributing least.

**Model Development**

**6.1 Model Selection:**

Based on the analysis results and objectives, an appropriate modelling technique will be selected. This may include linear regression, multiple regression, time series forecasting models (e.g., ARIMA, Auto ARIMA), or machine learning algorithms (e.g., random forests, gradient boosting).

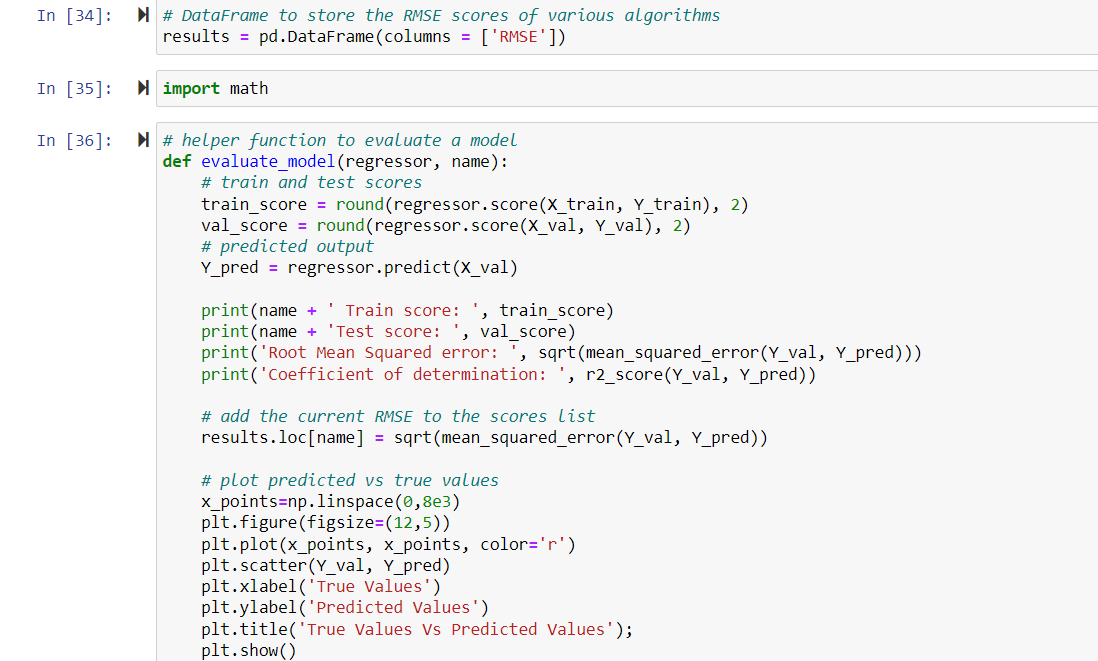
**6.2 Model Training:**

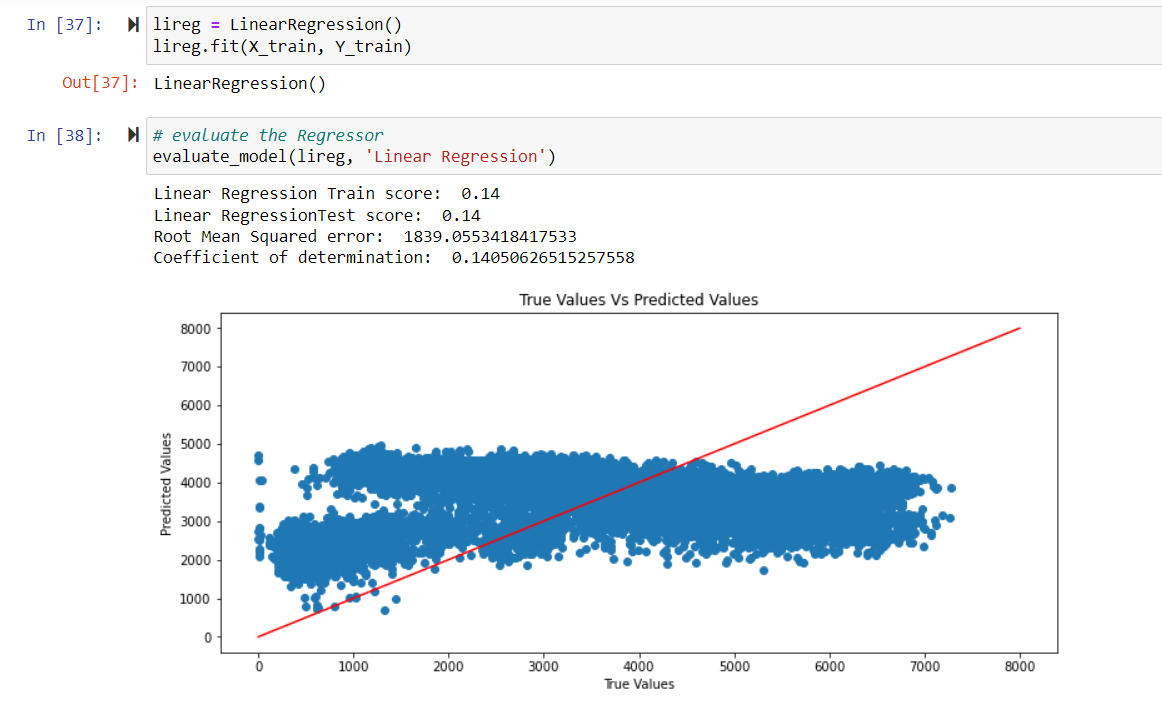
The selected model will be trained using historical data, with traffic volume as the target variable and other relevant variables as predictors. The training process will involve parameter estimation and model fitting.

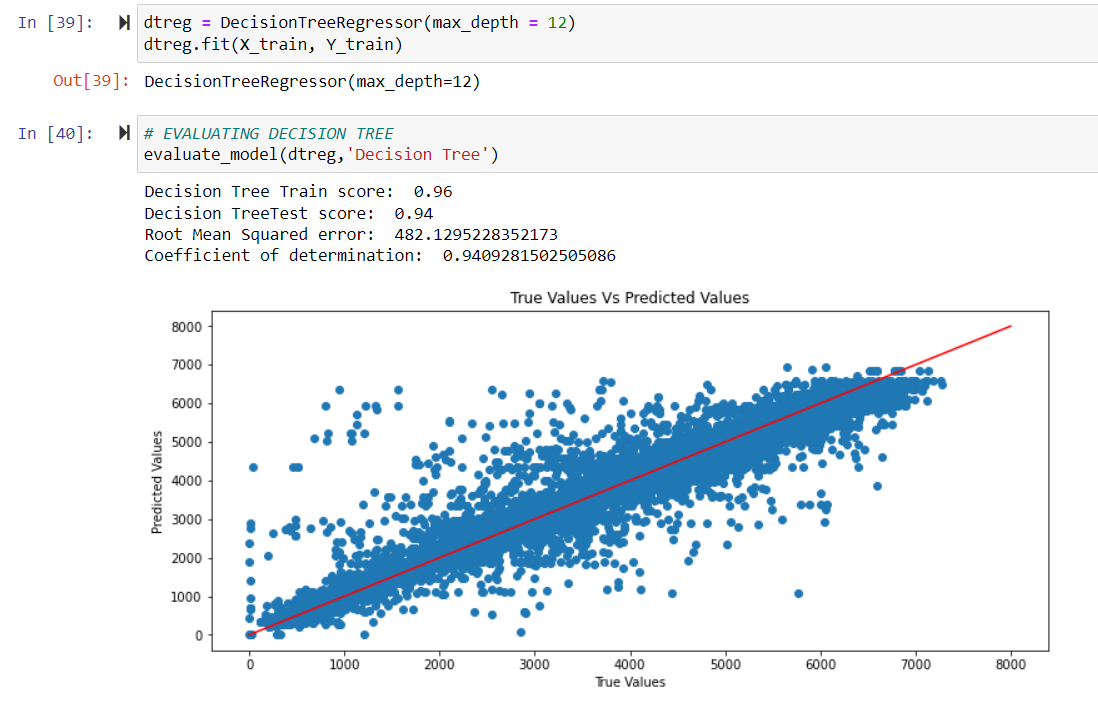
**6.3 Model Evaluation:**

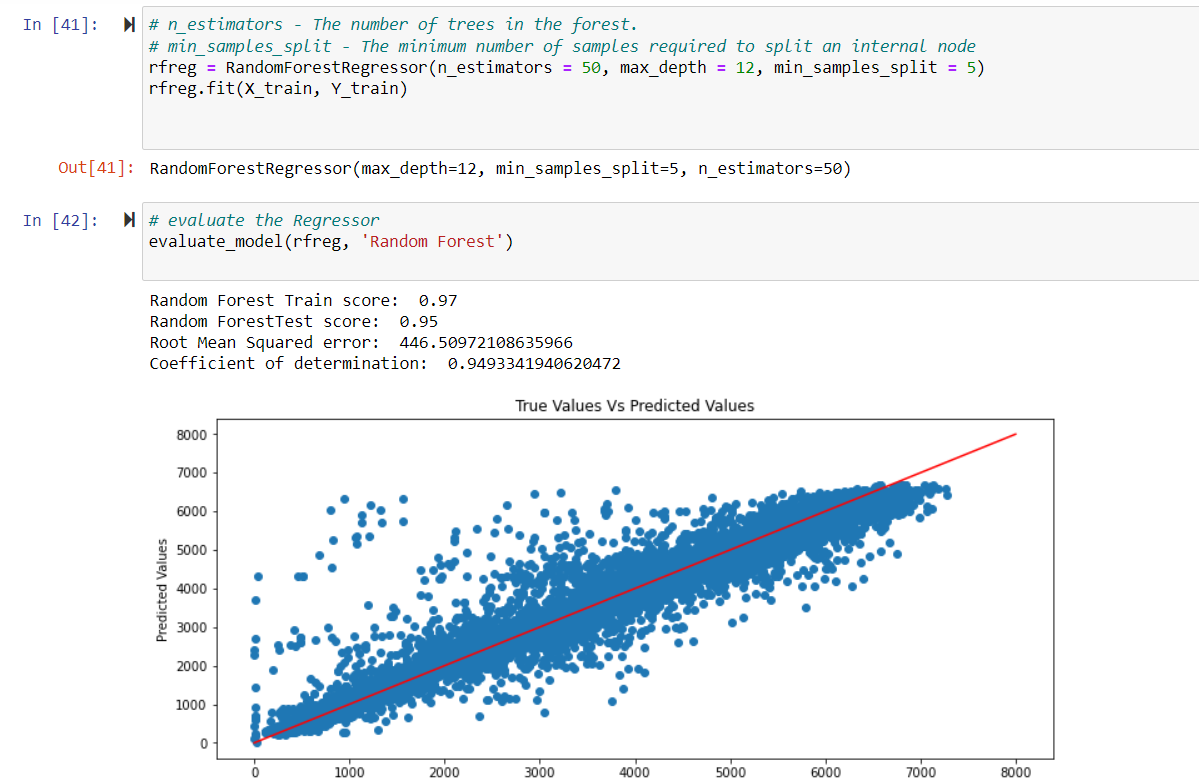
The developed model will be evaluated using appropriate evaluation metrics, such as mean squared error (MSE), mean absolute error (MAE), and R-squared. The performance of the model will be assessed to determine its accuracy and predictive capabilities.

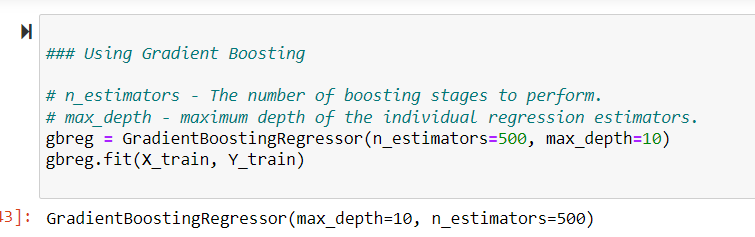


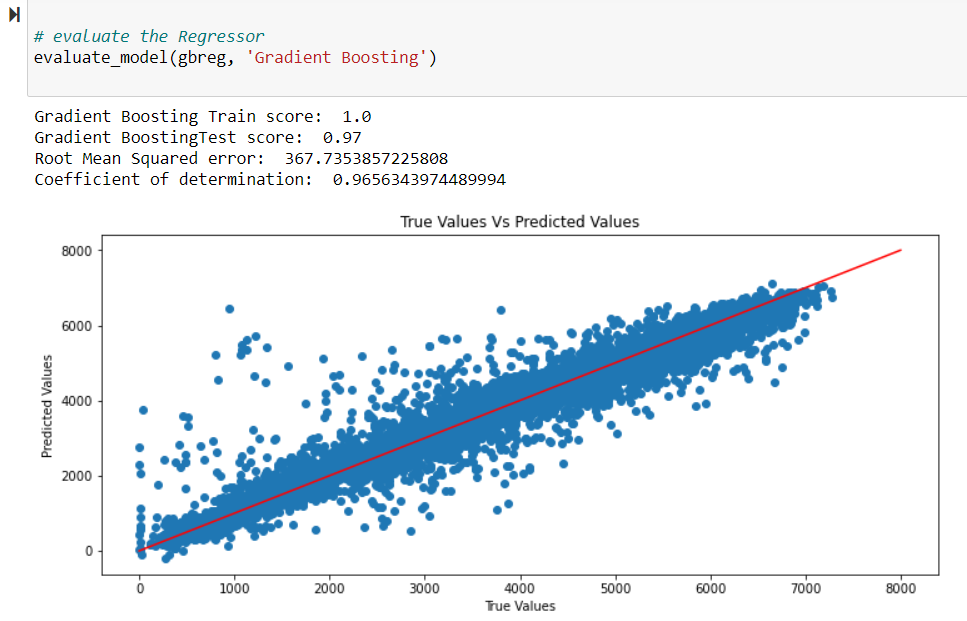


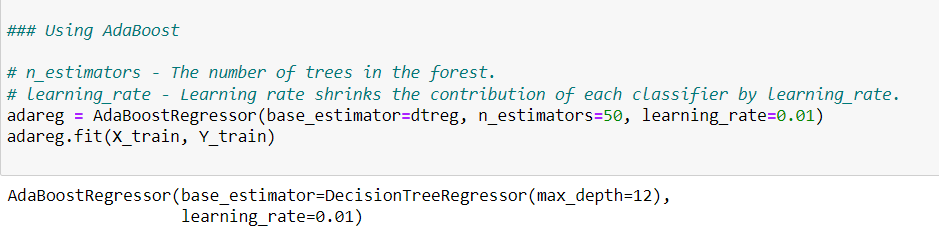


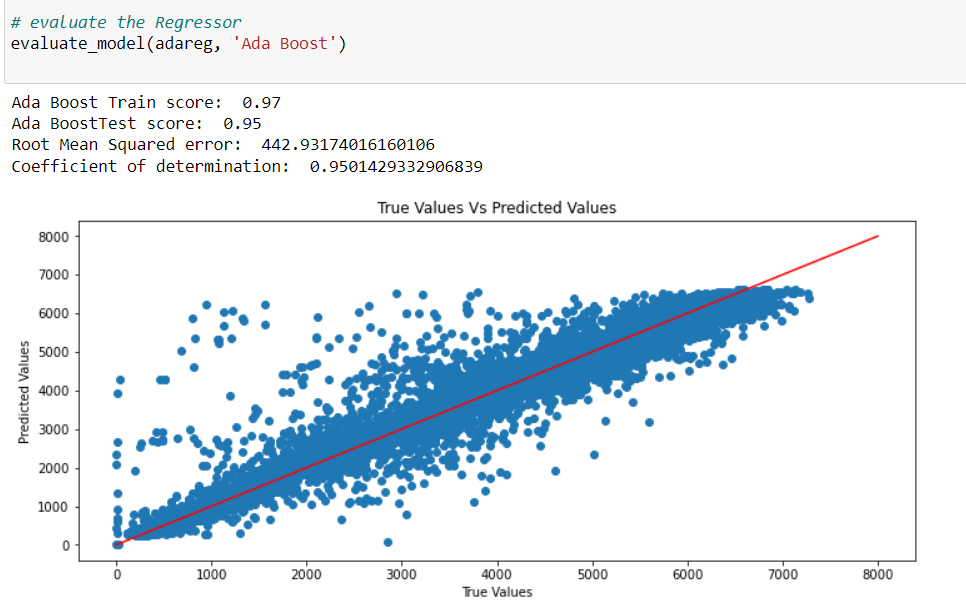


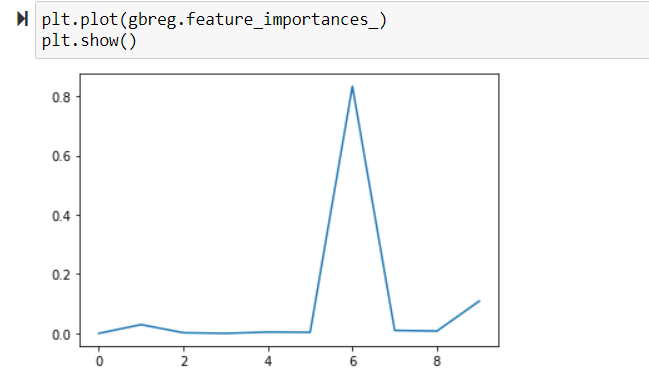


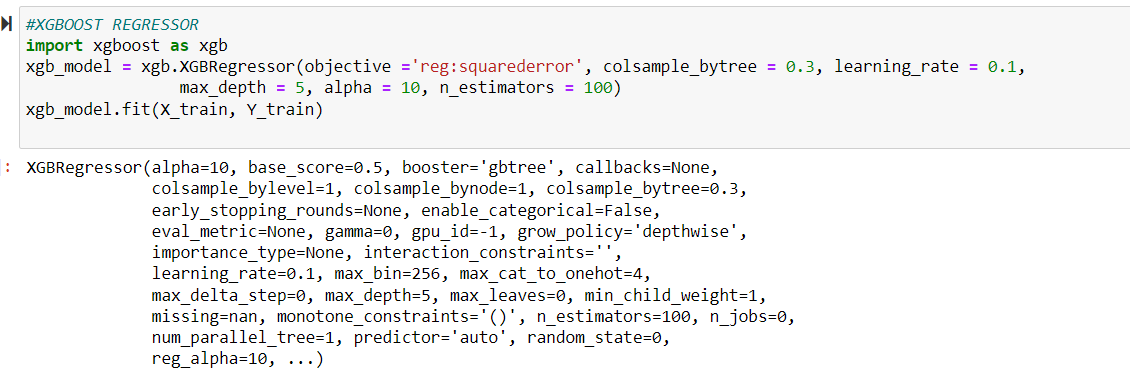


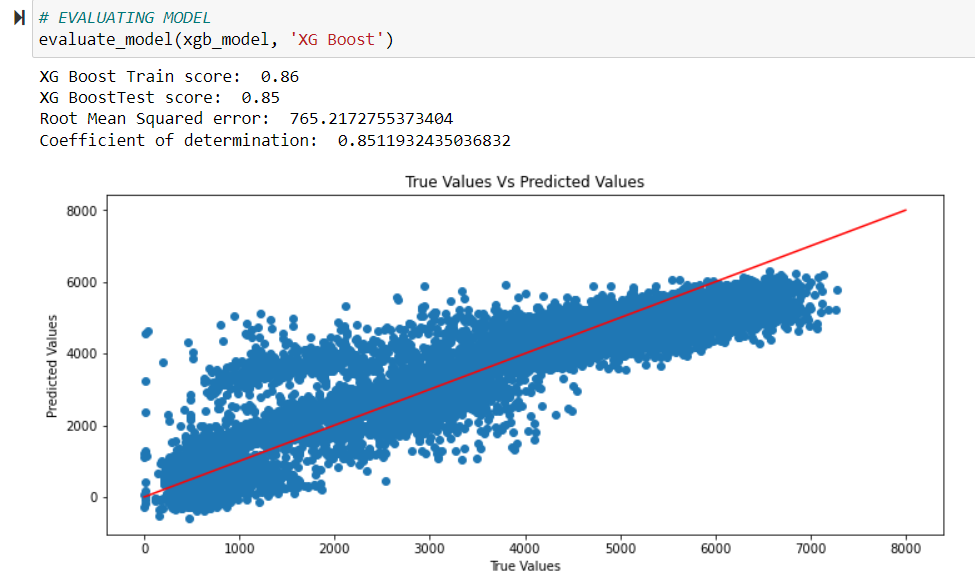


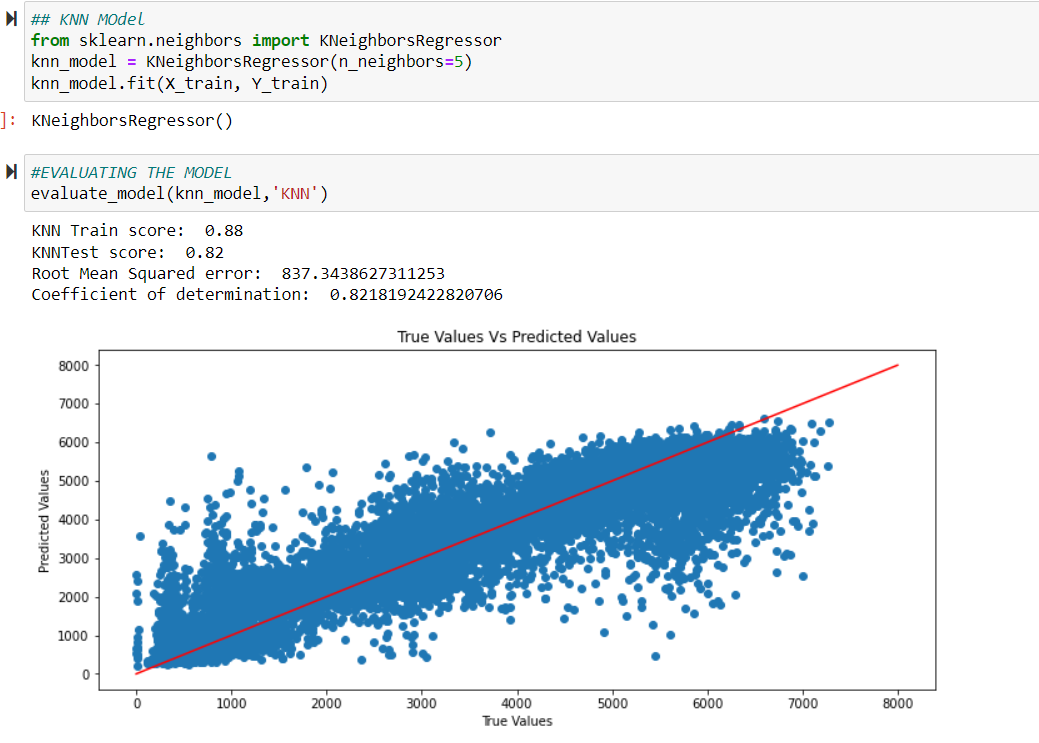


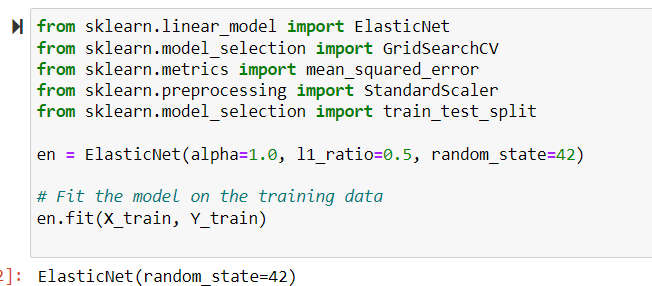


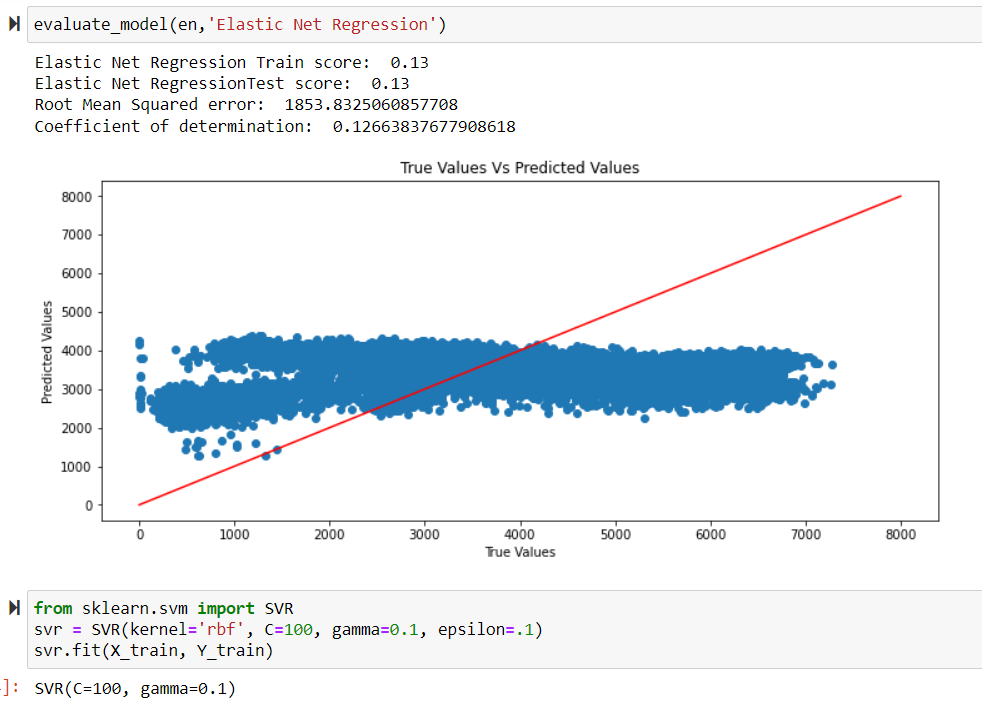


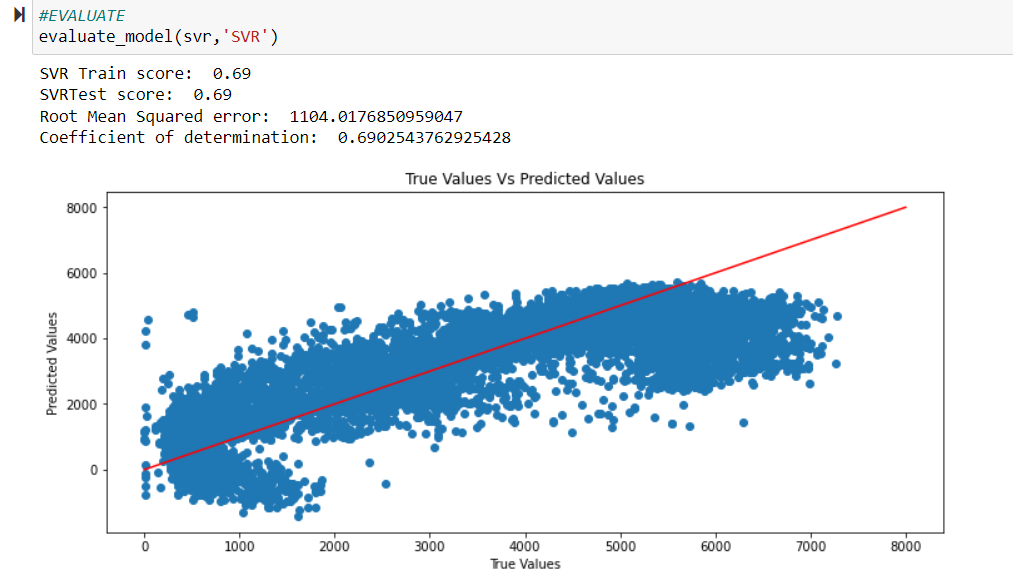


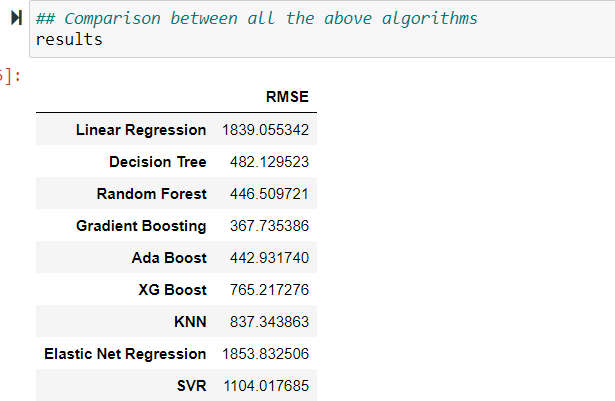




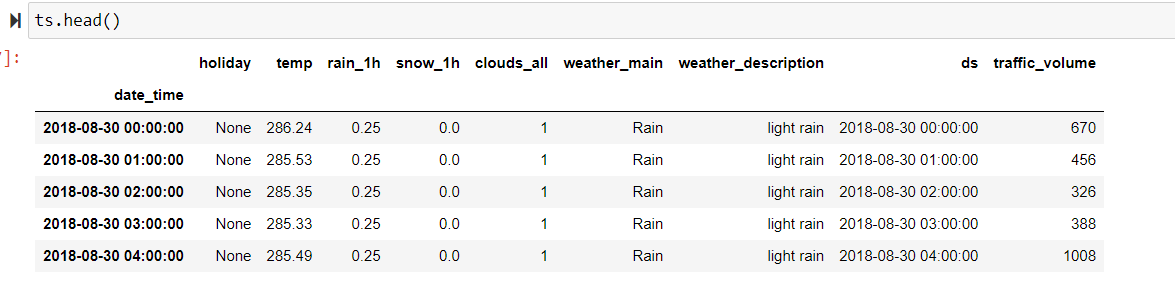


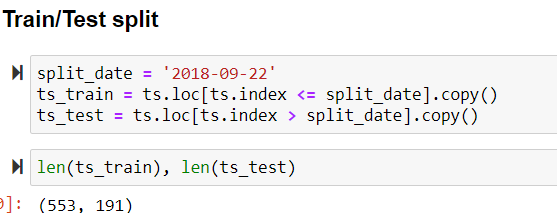


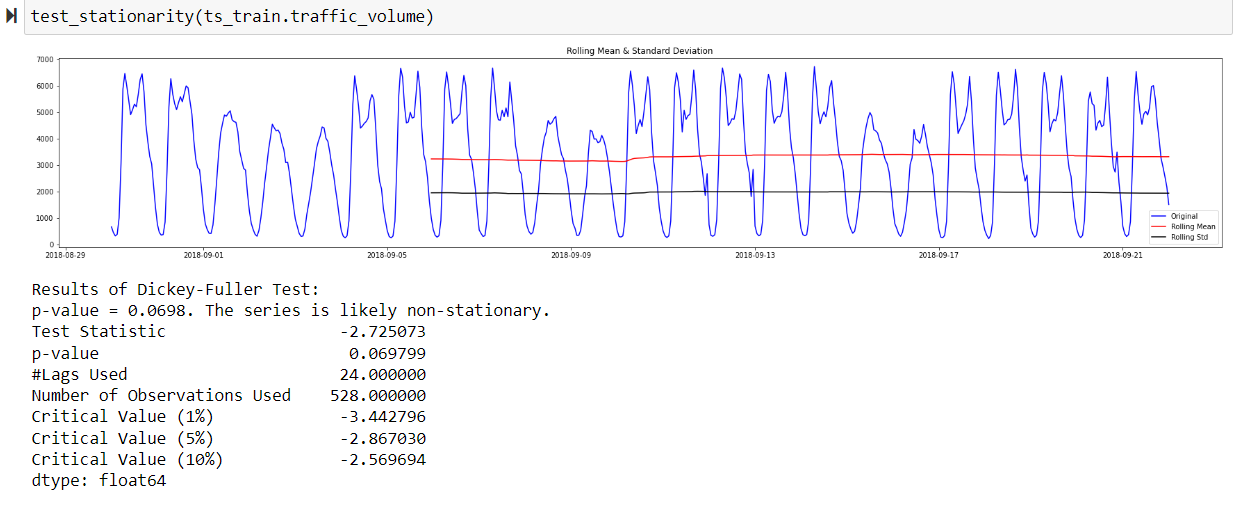


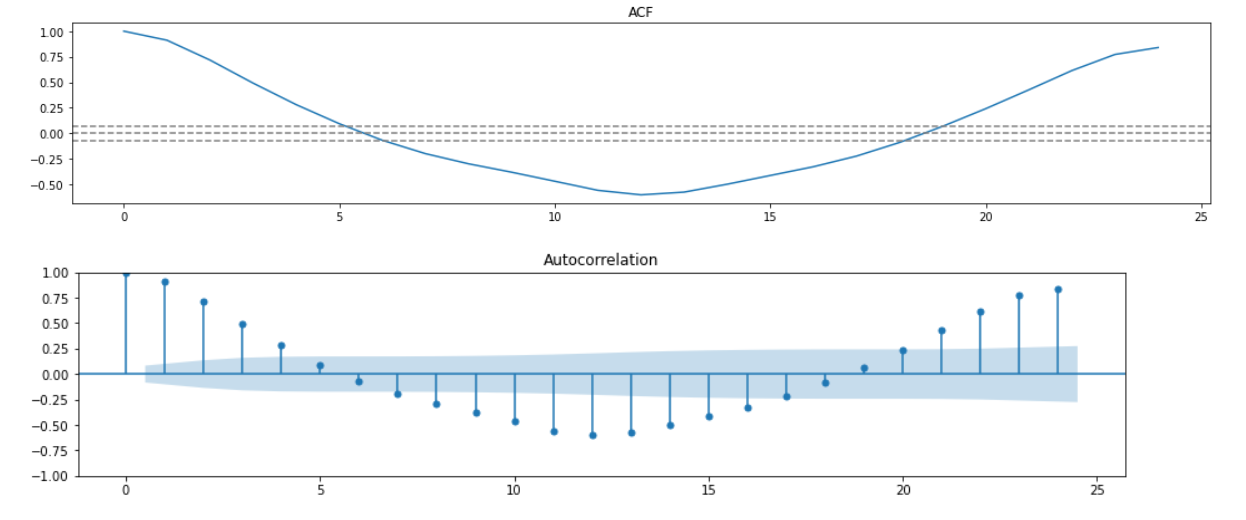


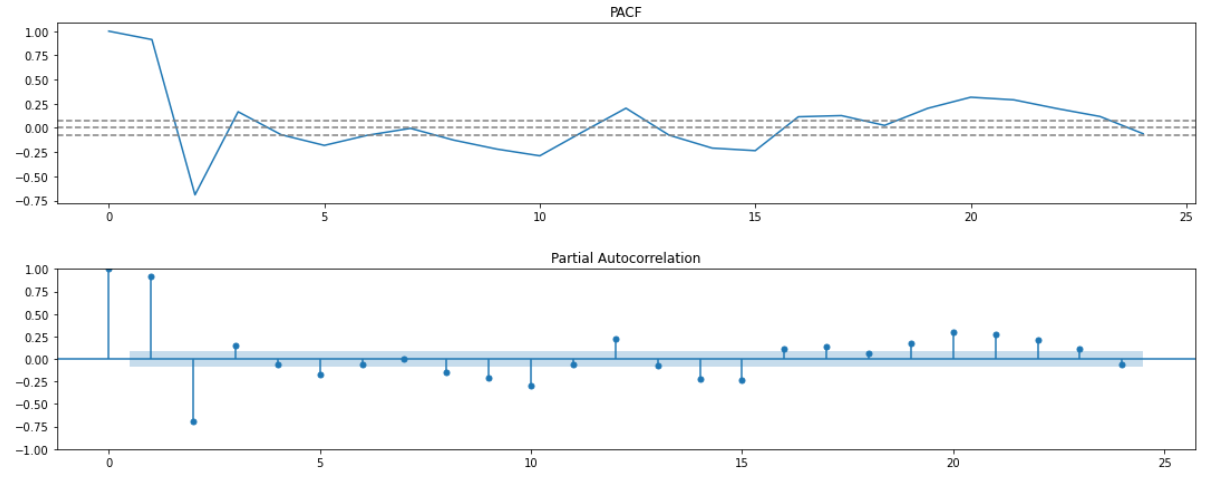


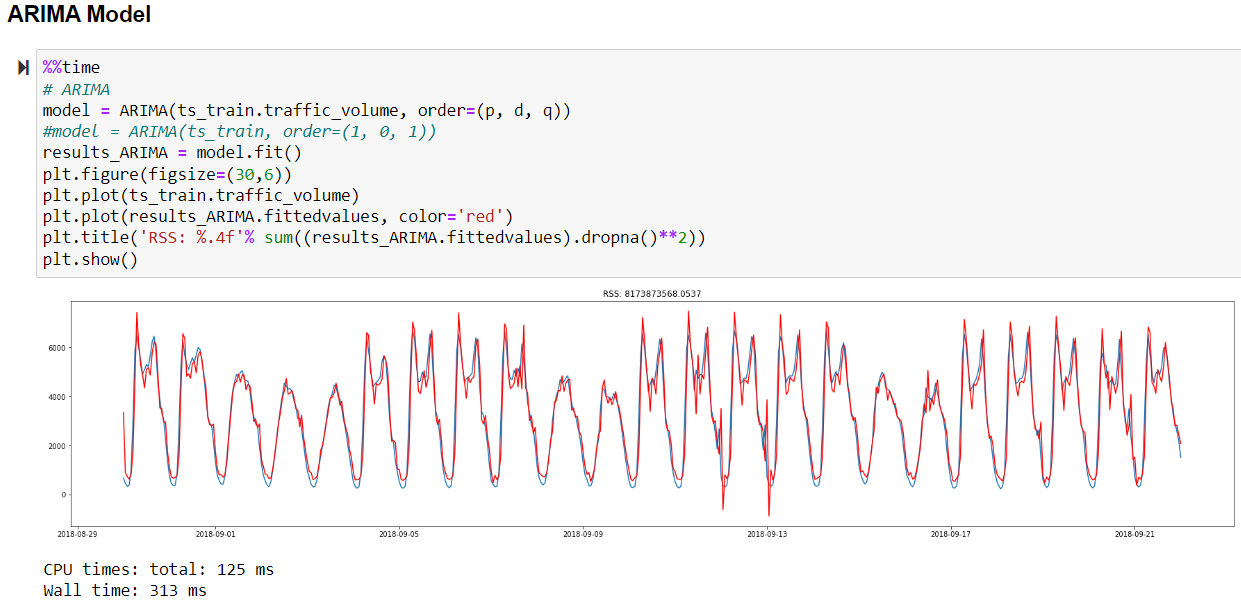


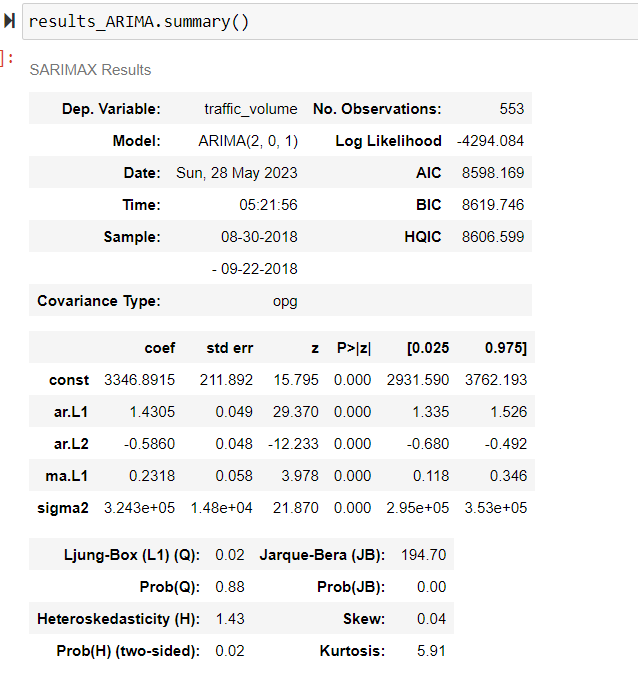


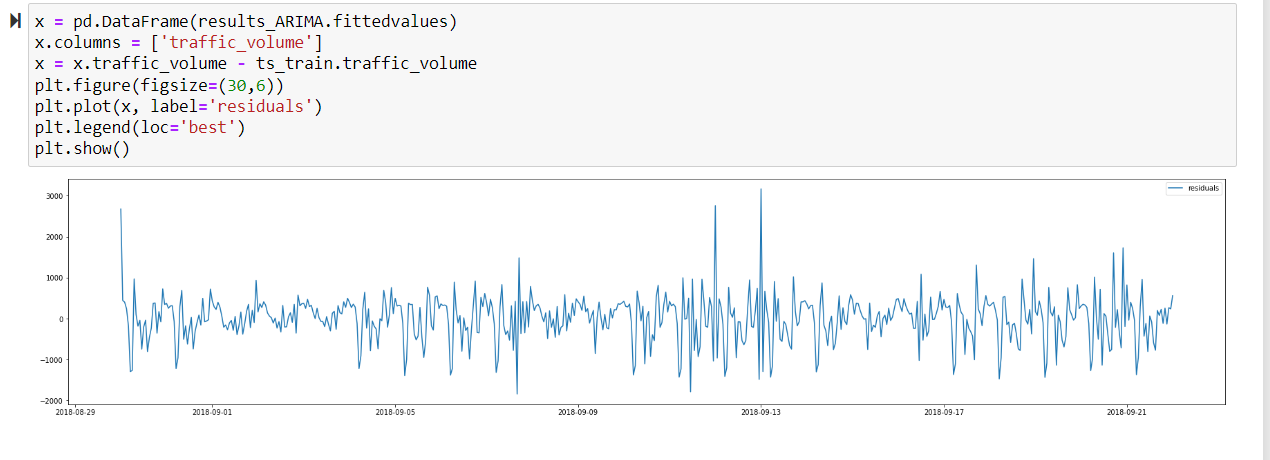


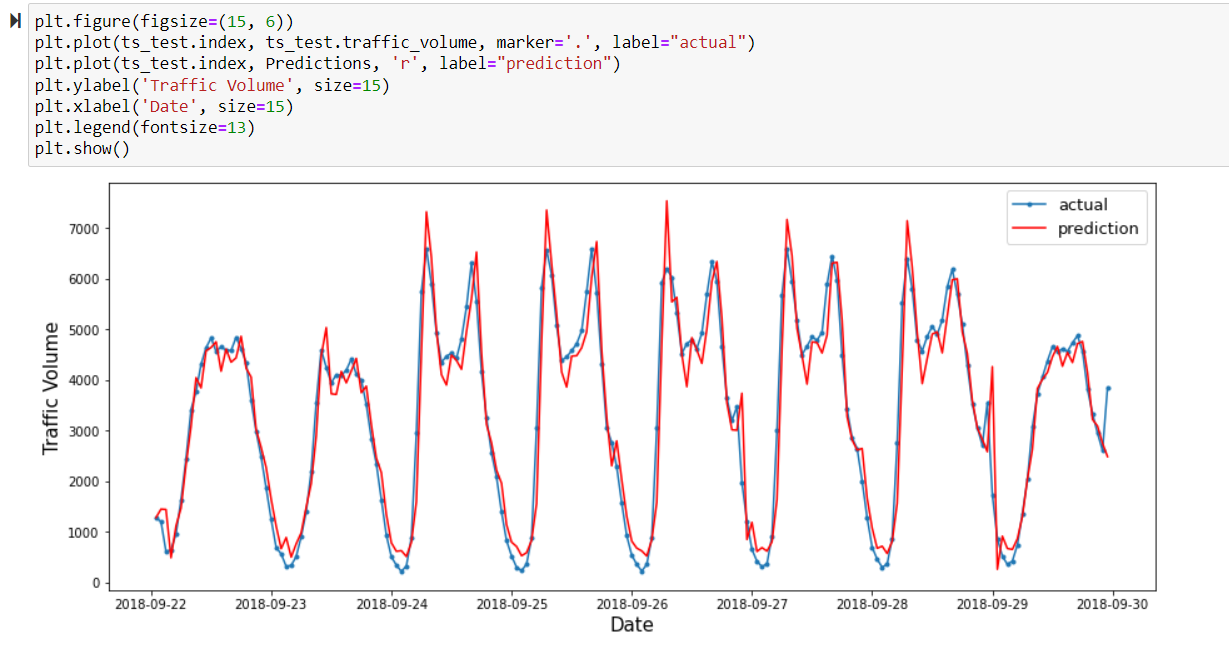


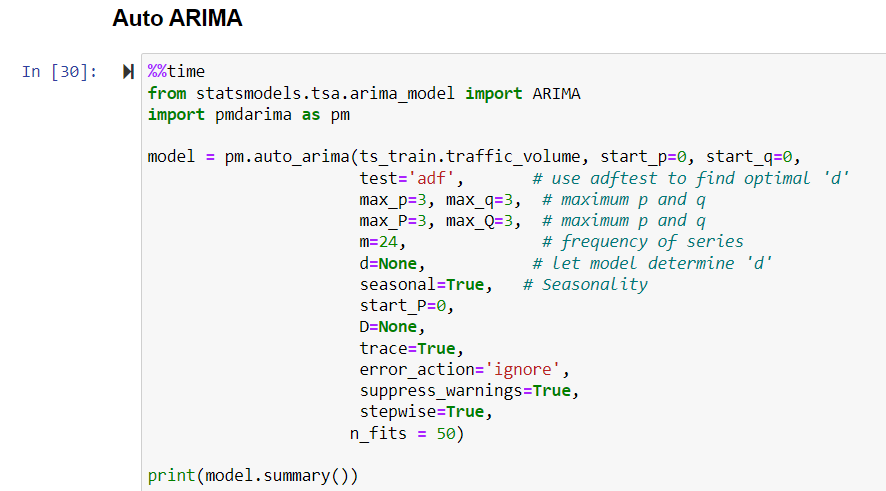


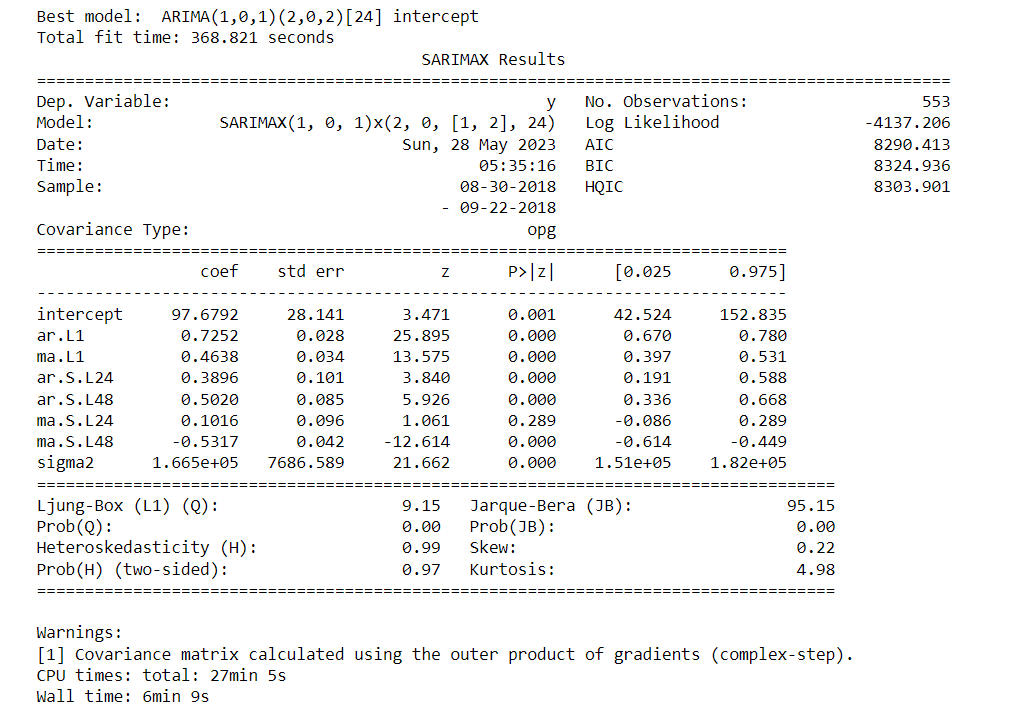


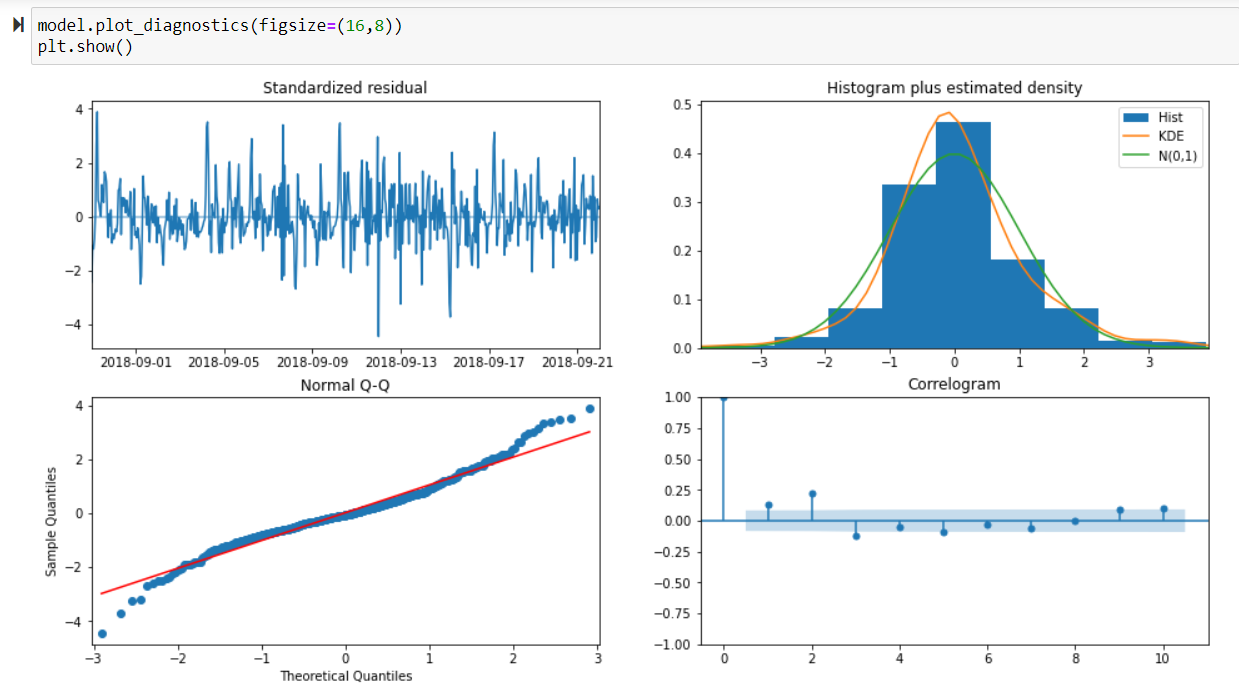


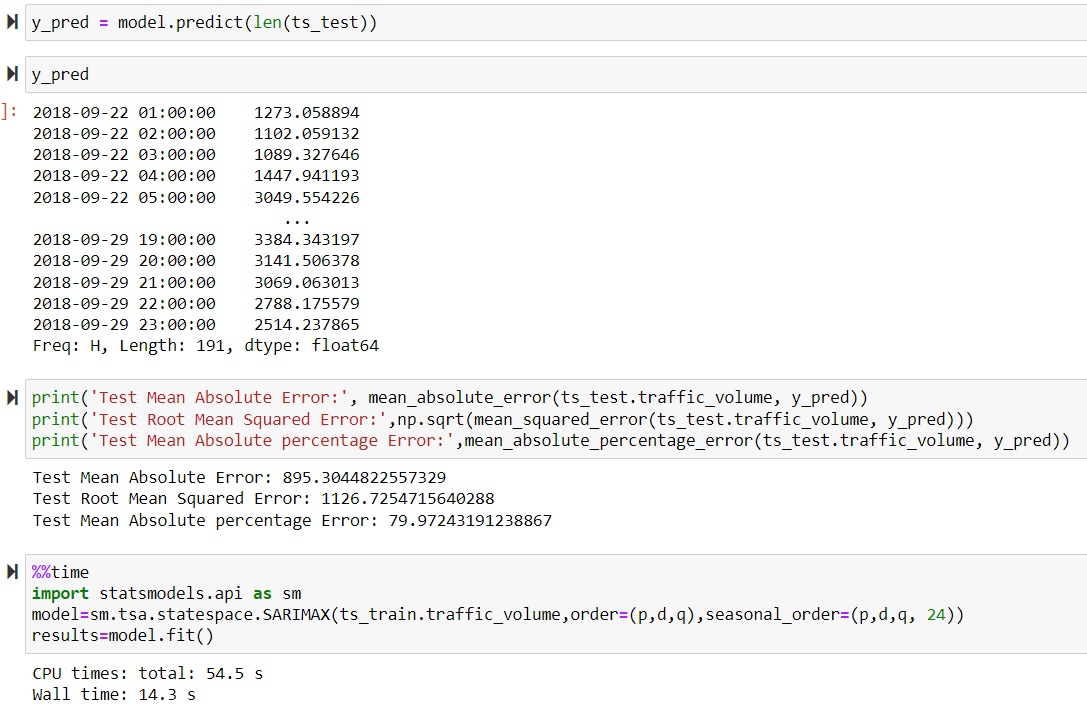










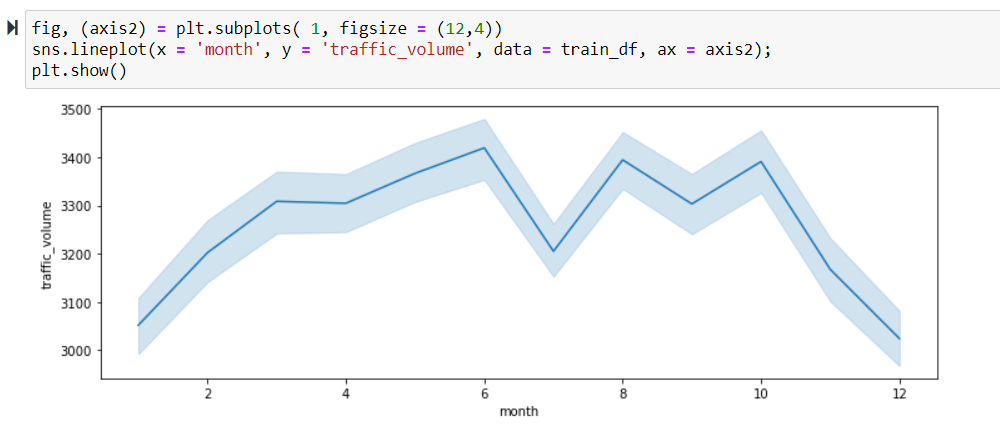


**Results and Discussion**

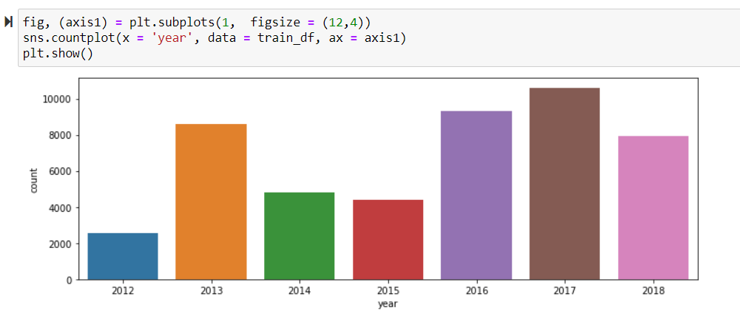
**7.1 Analysis of Traffic Volume Patterns:**

The analysis will reveal the patterns and trends in metro interstate traffic volume. It will provide insights into peak hours, busy days, and seasonal variations.

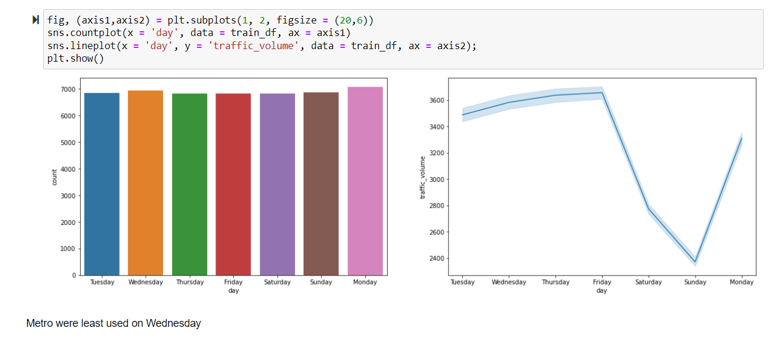
Like:



June, August, October are the busiest months.



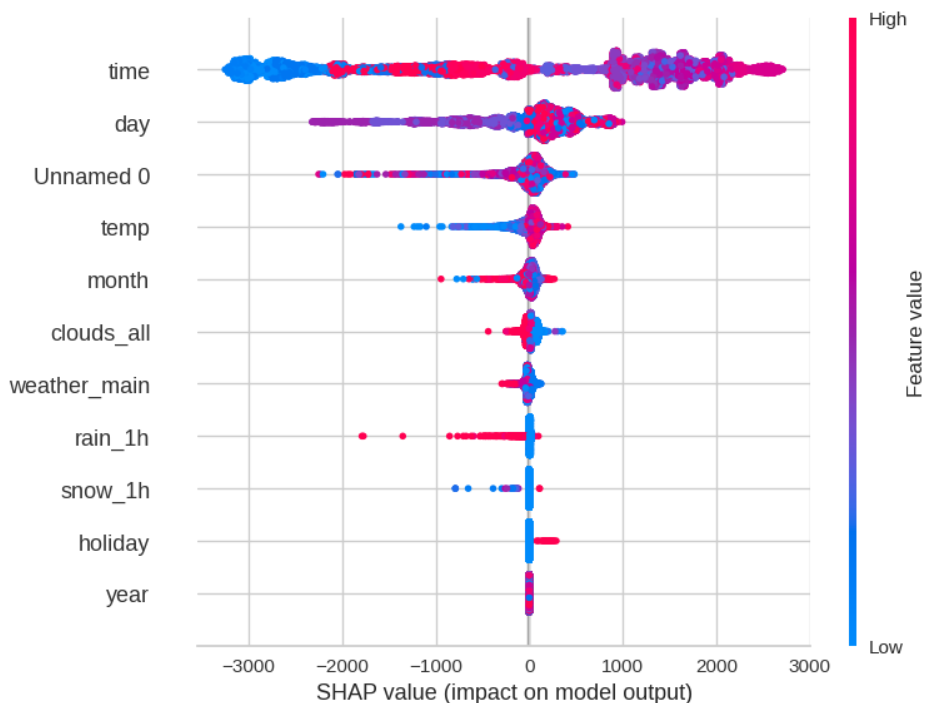
2017 was the busiest year



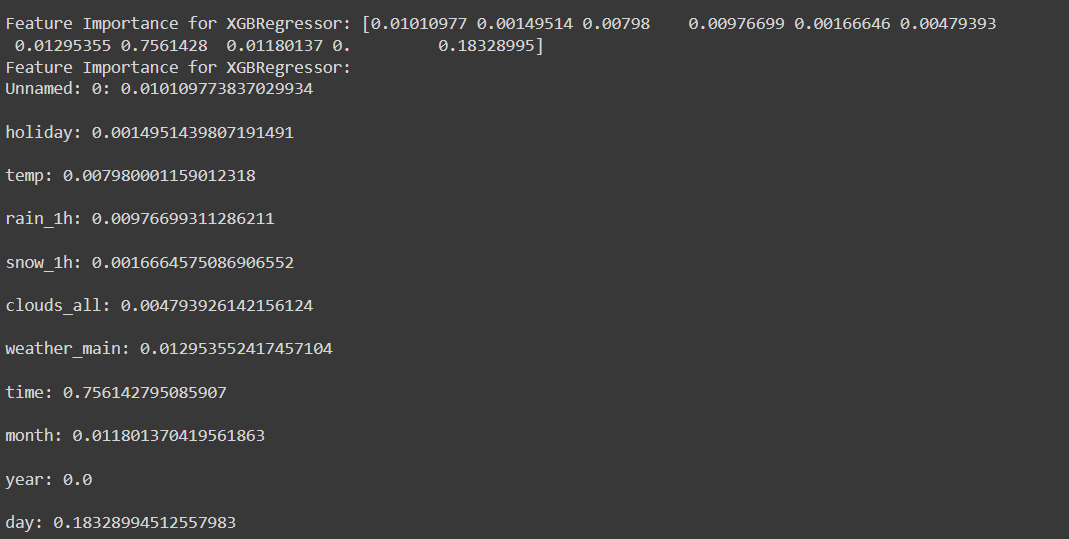
Metro was least used on Wednesday.

**7.2 Factors Influencing Interstate Traffic Volume:**

The correlation and regression analysis will identify the key factors influencing traffic volume. It may uncover the impact of variables such as time of day, day of the week, weather conditions, and special events. We can check Feature Importance with the help of pycaret.



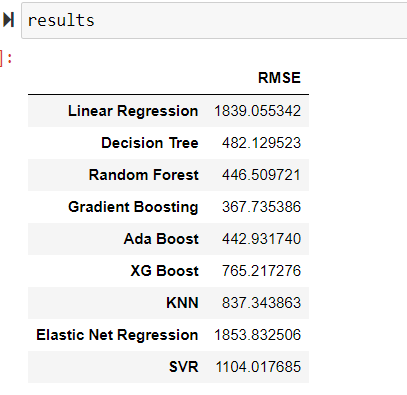




From this we can depict that the ‘time’ is the feature which is contributing most in the prediction and then day and then temperature. And the feature which is contributing least is year.

**7.3 Model Performance Evaluation:**

The predictive model's performance will be evaluated using appropriate metrics. The accuracy and reliability of the model in forecasting future traffic volume will be assessed.





We are getting that the Gradient Boosting is performing best as it has highest accuracy and least RMSE than others, then Ada Boost and then as follows.

**Conclusion**

**8.1 Summary of Findings:**

In this project, we found out various traffic patterns and trends like: June, August and October are the busiest months, 2017 was the busiest year, Wednesday faces least traffic, etc.

We found out that the feature which is contributing most to the model building and prediction is ‘time’ then ‘day’ and then ‘temperature’ and the feature which is contributing least is ‘year’.

We are getting that the Gradient Boosting is performing best as it has highest accuracy and least RMSE than others, then Ada Boost and then as follows.

**8.2 Implications and Recommendations:**

Based on the analysis results, implications for transportation planning, infrastructure development, and traffic management will be discussed. Recommendations for optimizing traffic flow and reducing congestion may also be provided.

**References:**

GitHub:

<https://jpvt.github.io/post/metro_traffic_volume/>

Kaggle:

<https://www.kaggle.com/code/ramyahr/metro-interstate-traffic-volume/notebook>