**Car Price Prediction**

**Submitted By:**

**V.V.S.N.SAI MOHAN - AM.EN.U4CSE20275**

M. VISHAL REDDY - AM.EN.U4CSE20248

S. MANIKANTA KUMAR - AM.EN.U4CSE20268

K. NAVEEN KUMAR - AM.EN.U4CSE20238

B. JAGADANANDA THARAKA - AM.EN.U4CSE20217

**ABSTRACT**

I need a model which predicts the price of a used cars with given constraints.

Accurate car price prediction involves expert knowledge, because price usually depends on many distinctive features and factors.

Typically, the most significant ones are present price, brand and model, age, mileage etc. The fuel type used in the car as well as fuel consumption per mile highly affect the price of a car due to a frequent change in the price of a fuel. Different features like exterior colour, type of transmission, safety, air condition, etc. will also influence the car price.

* ***Task*** *(T): Classify the price of that used car using the previous data knowledge.*
* ***Experience*** *(E): A corpus of data with the features to be learned and predict the price.*
* ***Performance*** *(P): Classification accuracy, the number of car prices predicted correctly out of all car data considered as a percentage.*

**Assumptions:**

Some ML models won’t predict 100% correct there is chance of wrong prediction. What if it predicts wrong…?

A) We should train model to get max correct prediction.

B) Finding the best regression algorithm, among Linear Regression, Lasso Regression, Random Forest Regression, Ridge regression, etc., for our problem would be a challenge.

C) Reconstructing the given dataset. Changing the categorical values into numerical form. And removing the unnecessary features.

**1. INTRODUCTION**

**Motivation:**

Some car buyers will suffer a lot of problems like dealer may charge amount more than the car value.

This model can help them to predict the actual and reasonable price of those used cars.

**Benefits of solution:**

If this model generates more than 0.9 accuracy, then it’ll be more helpful to people. Because when your model accuracy is more than 0.9 then it’ll almost give correct decision. This can be useful for dealers as well as individual people to buy used cars.

For example: Some dealers will set more amount of car value then the individuals may suffer a lot. So by using this individuals can get the car by the reasonable price.

**Solution Use**:

This solution may be a temporary solution for the given data. Because in future there may be some extra features added like for this data set. We have fuel\_type with petrol and diesel and in future electric vehicles may add to this so there must be some extra features considered to set the value of the used cars. The accuracy may drop down in future. So, this solution is for only petrol and diesel vehicles.

This solution is mainly for individuals who want to buy a car from owners directly without dealers. After completing this project we’ll post this in GitHub and LinkedIn so that it’ll be helpful for those who want to do projects and also the actual use like implementing this model for car value.

**-M. Vishal Reddy (20248)**

**-S. Manikanta (20268)**

**2. Dataset finalization**

**80% of data set for training and 20% of data for testing. We collected 3 data sets suitable for our problem, we’ll do on any one of them.**

1.



- It is about the used cars description that given.

- There are 9 features:

Car\_name, year, selling\_price, present\_price, Kms\_driven, fuel\_type, seller\_type, transmission, owner.

- It has been used to predict the used cars price prediction.

**Source: Kaggle.com**

2.



- It is about the used cars description that given.

- There are 9 features,

Brand, price, body, Mileage, EngineV, Engine\_type, Registration, Year, Model.

**Source: Github.com**

3.

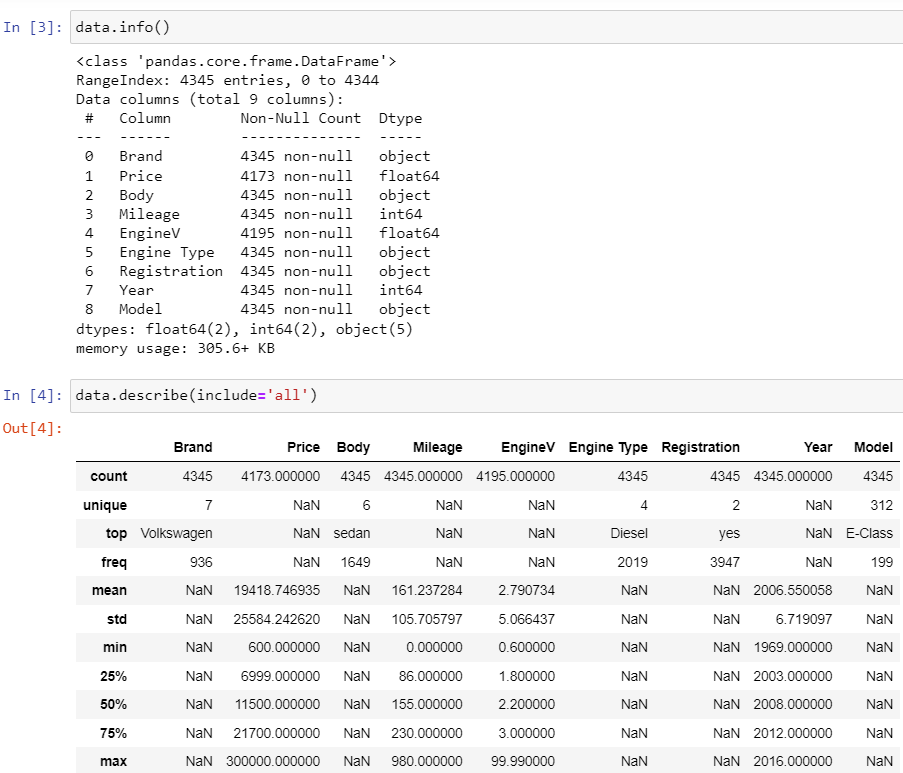


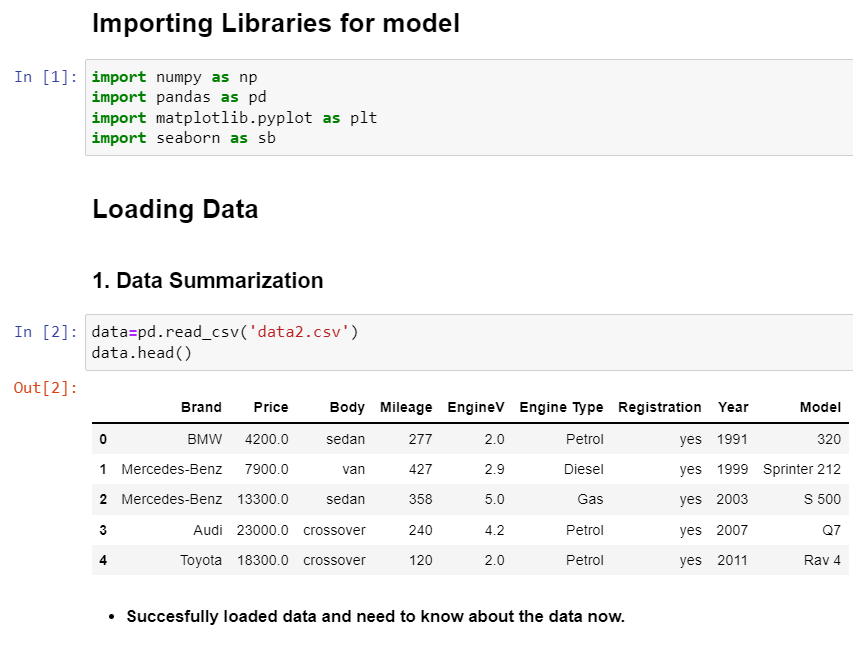
- It is about the used cars description that given.

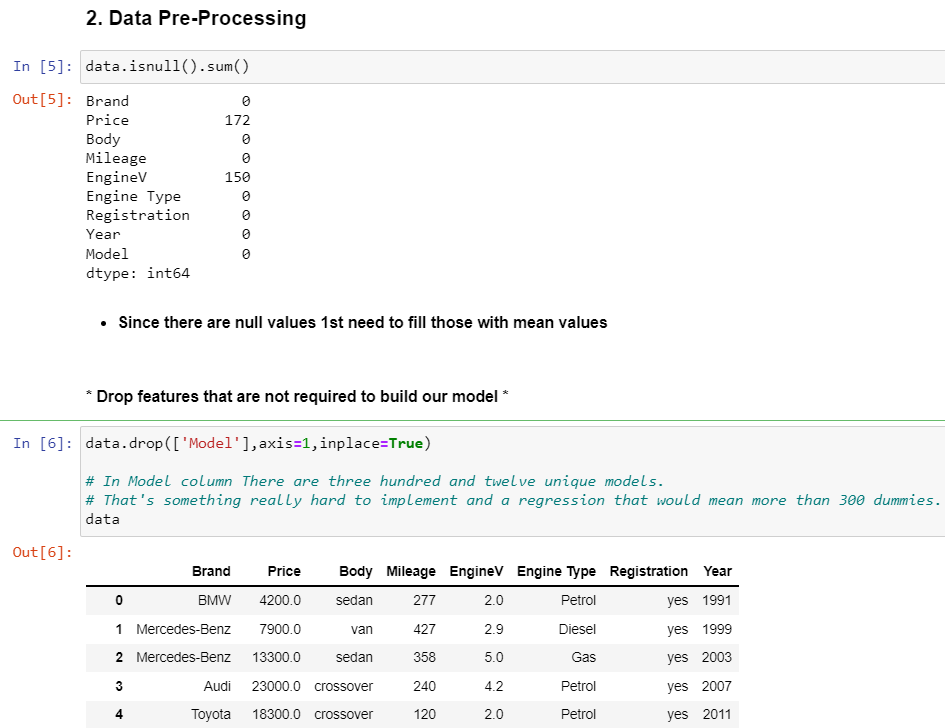
- There are 13 features (with empty cells),

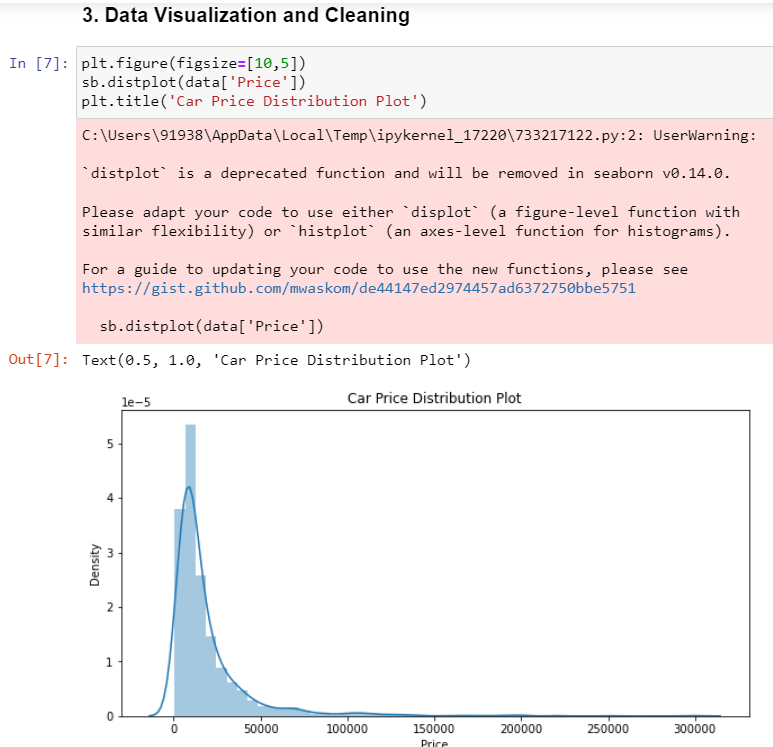
Name, Location, Year, Kilometers\_Driven, Fuel\_type, Transmission, Owner\_type, Mileage,Engine, power, seats, New\_price, price.

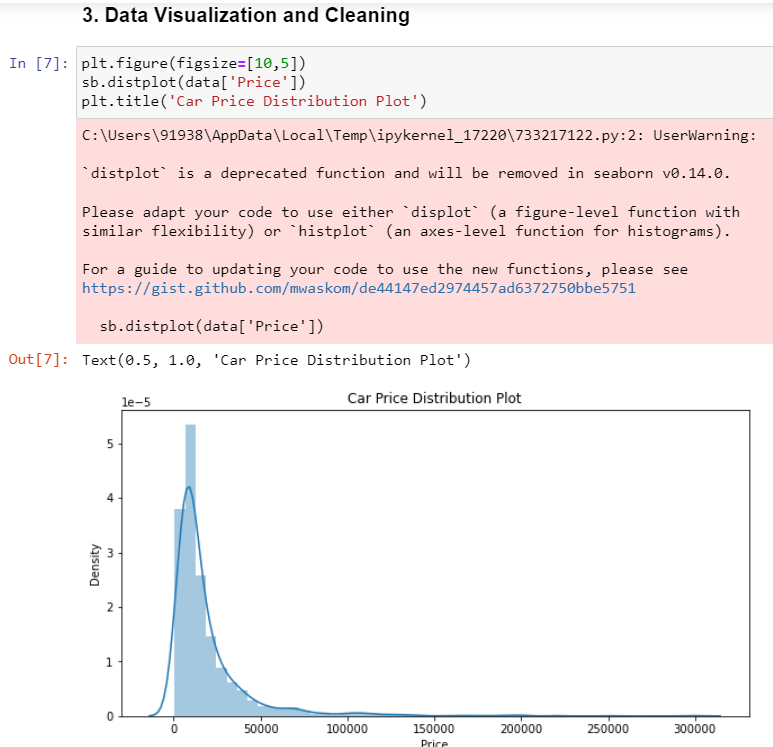
**Source: Github.com**

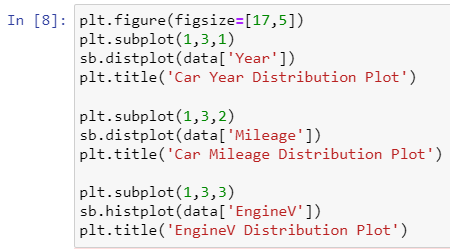
**3. Data Pre-processing:**

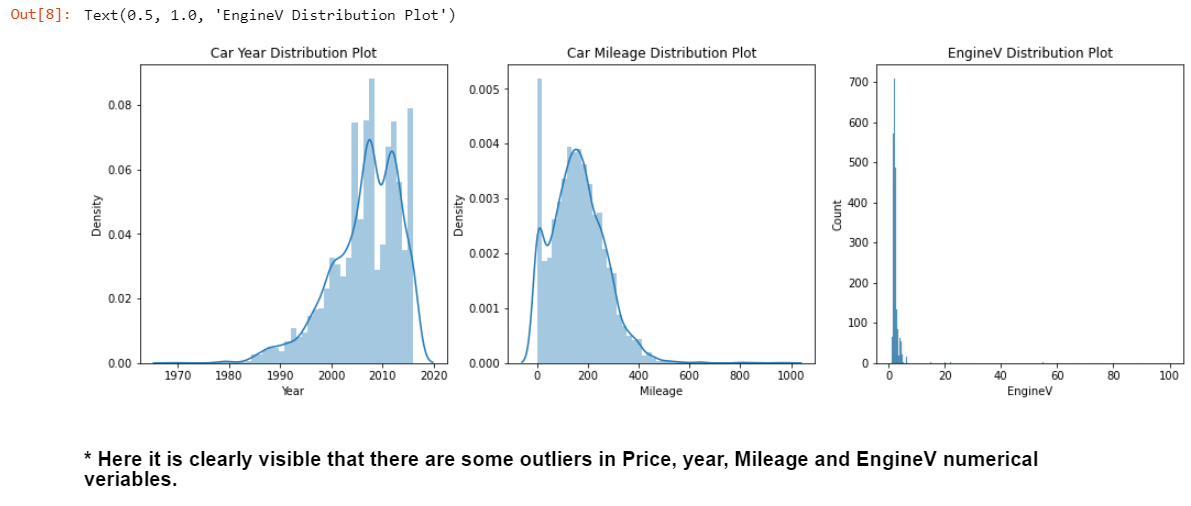
****

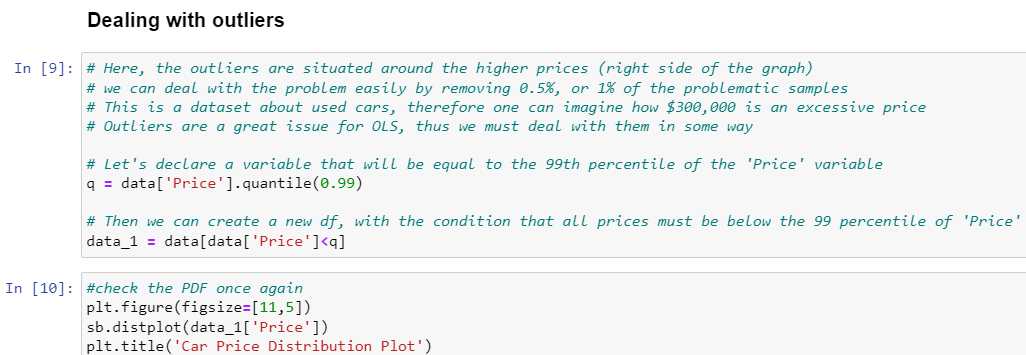
****

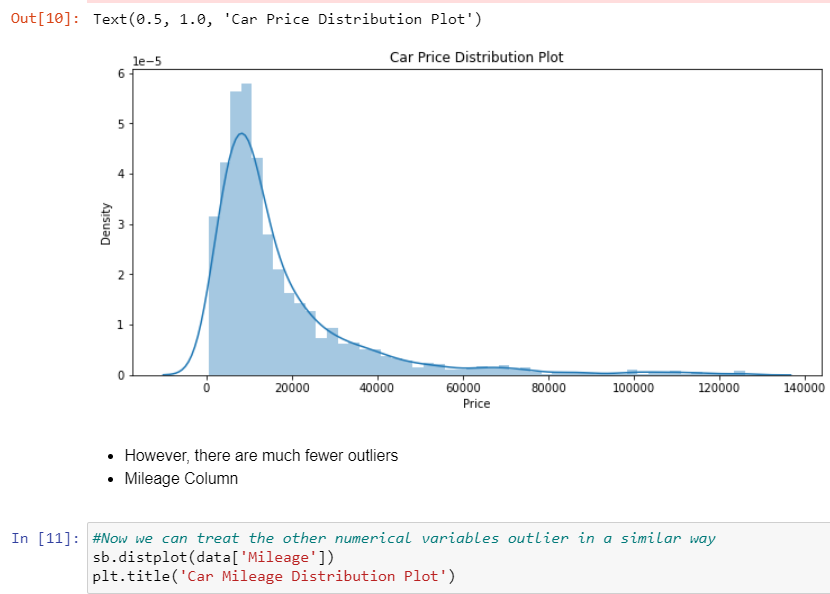
****

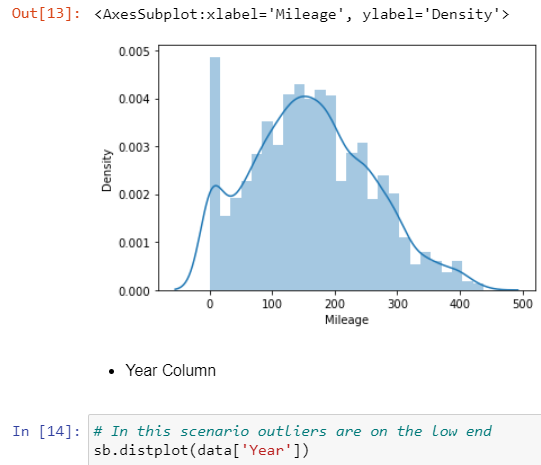
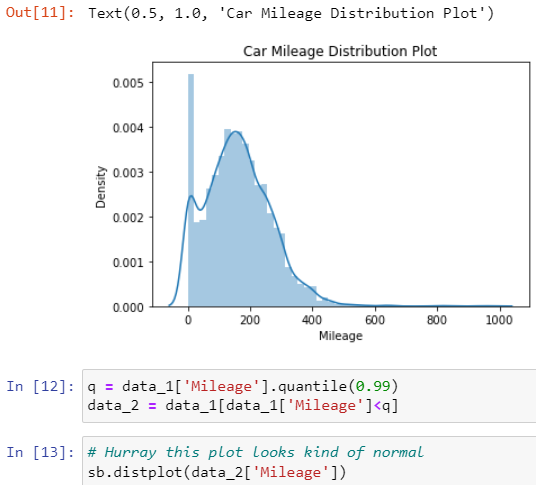
****

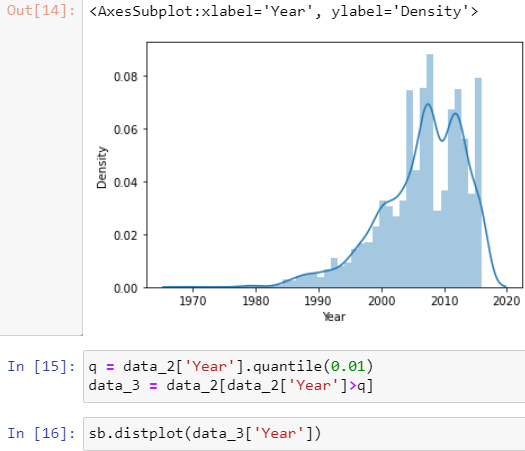
****

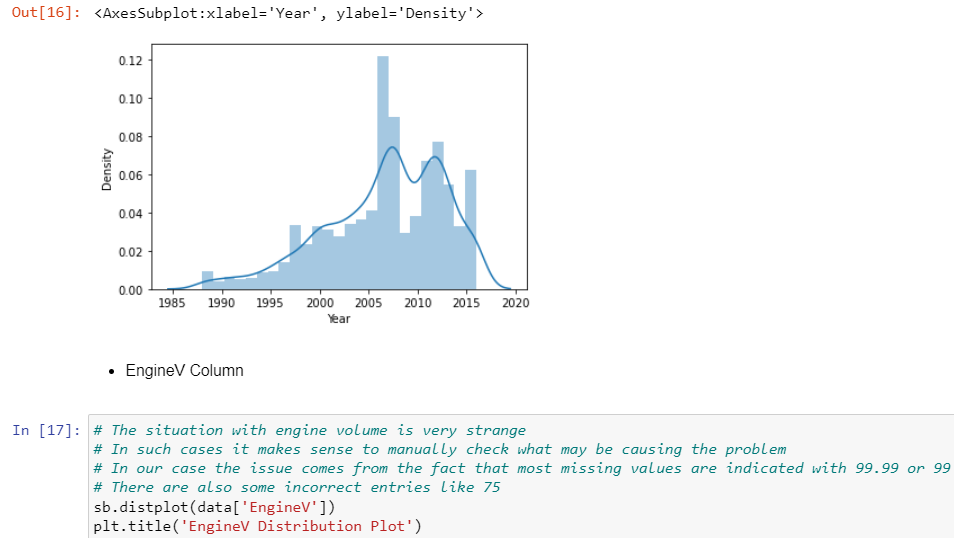
****

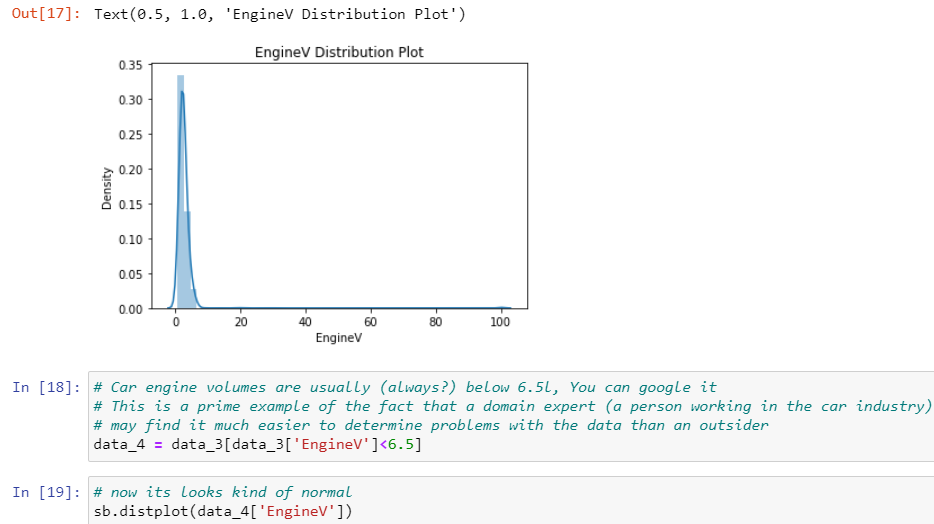
****

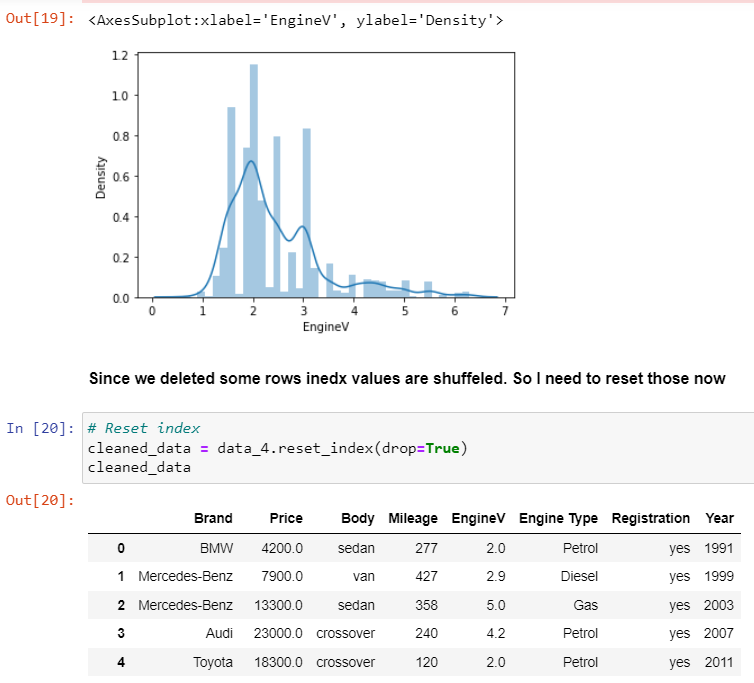
****

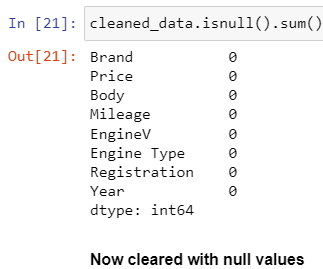
****

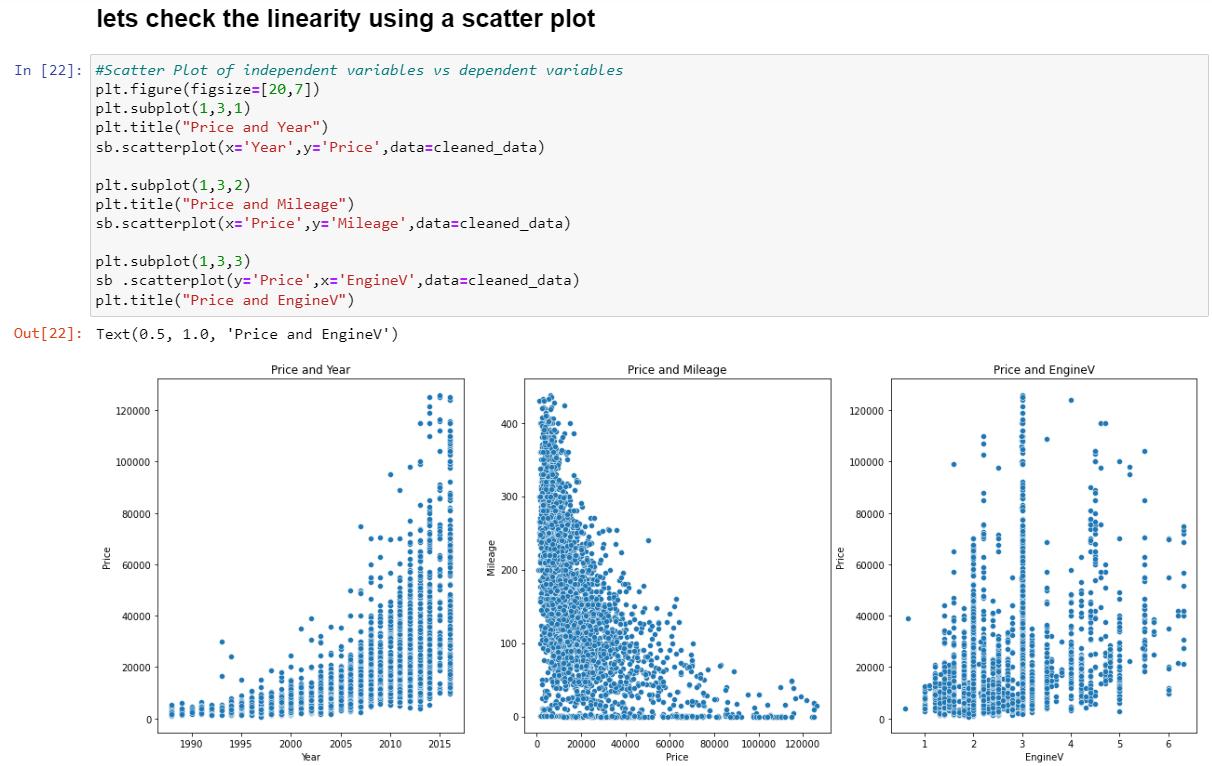
****

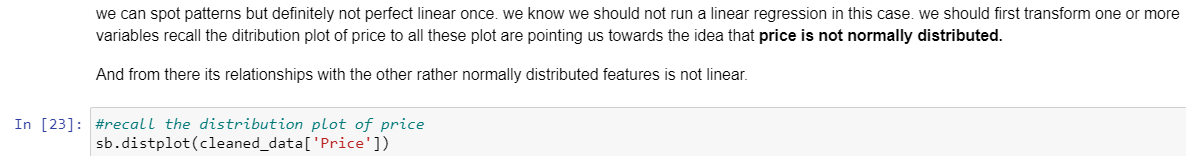
****

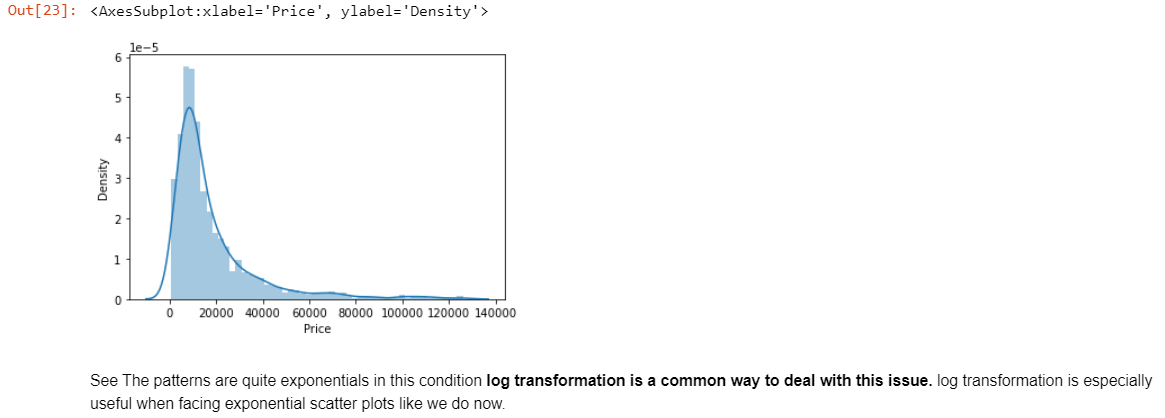
****

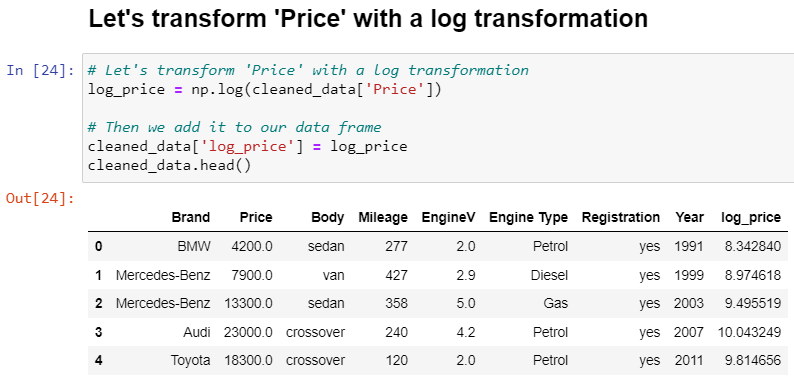
****

****

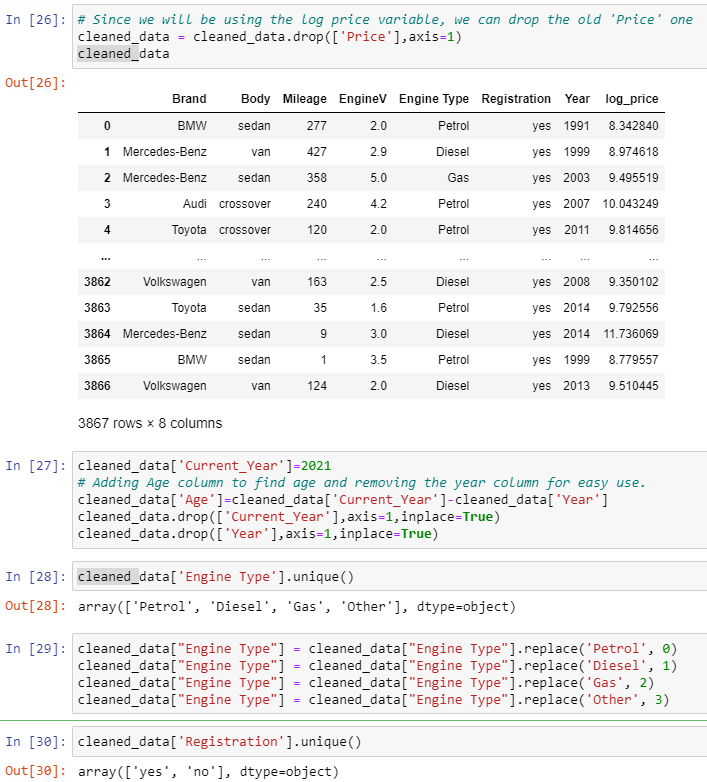
****

****

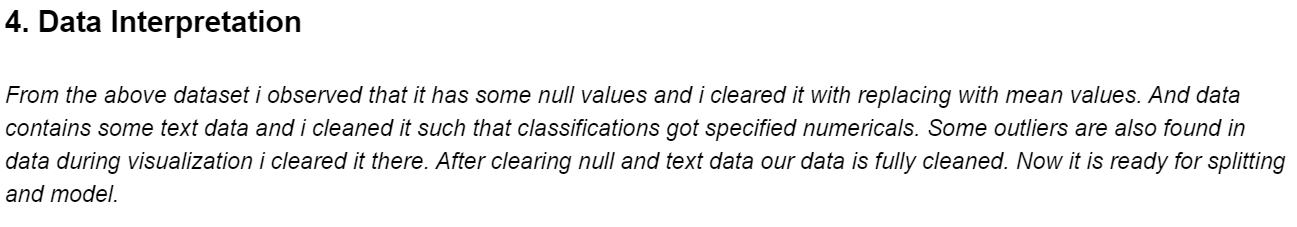
****

****

****

****

****

****

**Dataset:**

****

**Cleaned Dataset:**

****

**PHASE-2**

**- M. Vishal Reddy (20248)**

**- S. Manikanta (20268)**

**1. Problem Definition: -**

In this Project, developing a model which predicts the price of a used cars with different variables. Some car buyers will suffer a lot of problems like dealer may charge amount more than the car value. This model can help them to predict the actual and reasonable price of those used cars.

**2. Data Sets: -**

We are chosen 3 datasets for this project and each with different variables and different shapes.

1.



* It is about the used cars description that given.
* There are 9 features: Car\_name, year, selling\_price, present\_price, Kms\_driven, fuel\_type, seller\_type, transmission, owner.
* It has been used to predict the used cars price prediction.

**Source: Kaggle.com**

2.



* It is about the used cars description that given.
* There are 9 features: Brand, price, body, Mileage, EngineV, Engine\_type, Registration, Year, Model.

**Source: Github.com**

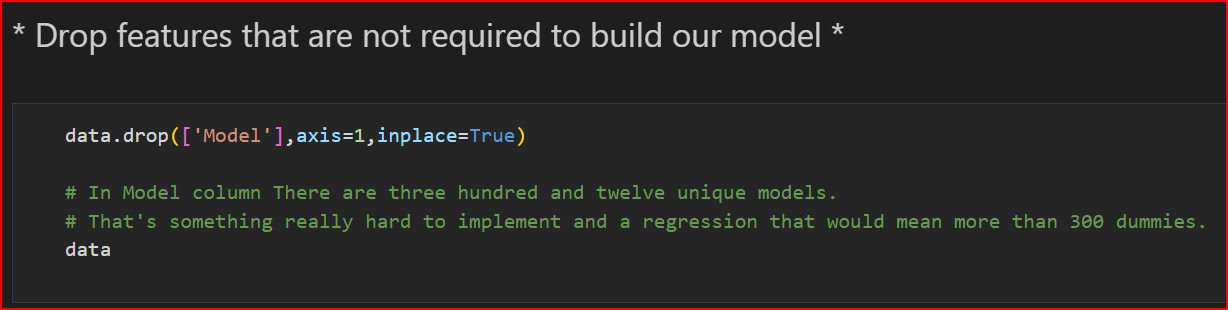
3.



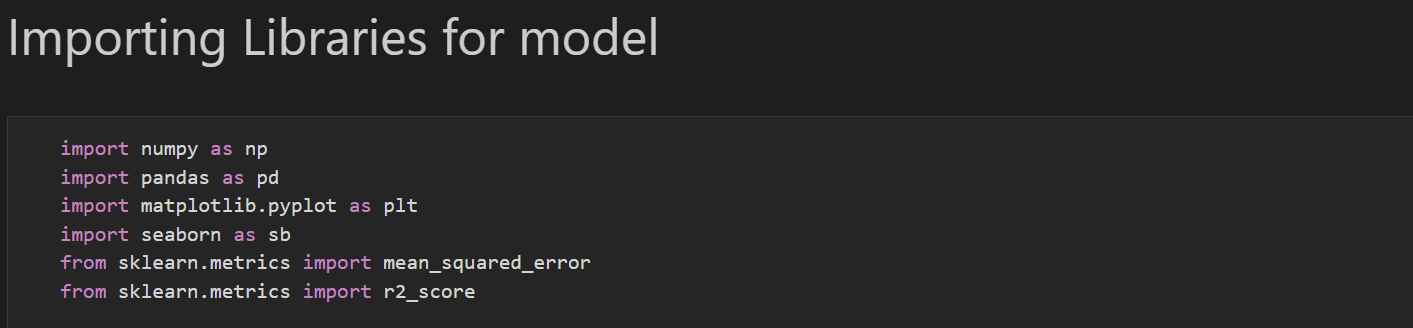
* It is about the used cars description that given.
* There are 13 features (with empty cells): Name, Location, Year, Kilometers\_Driven, Fuel\_type, Transmission, Owner\_type, Mileage,Engine, power, seats, New\_price, price.

**Source: Github.com**

**3. Prepare Data: -**

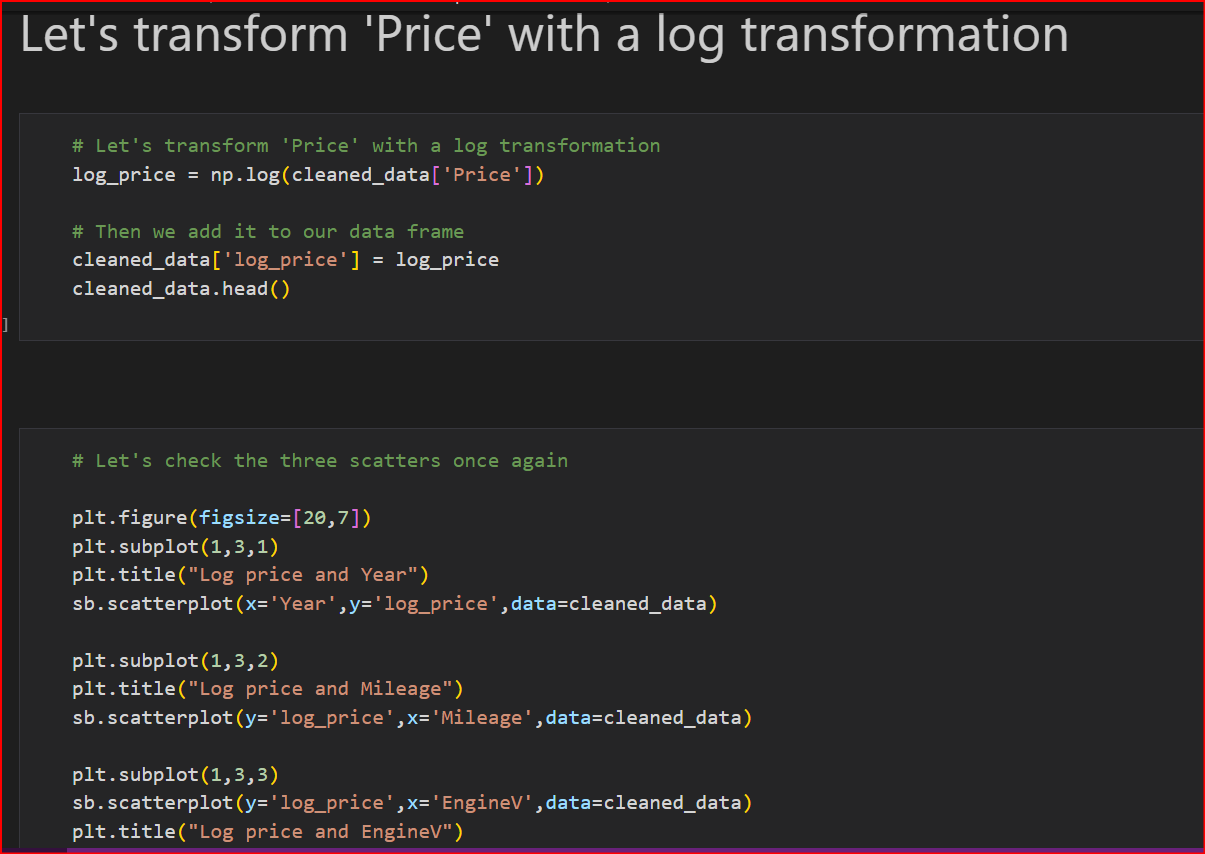
****

**Here we are dropping the data**

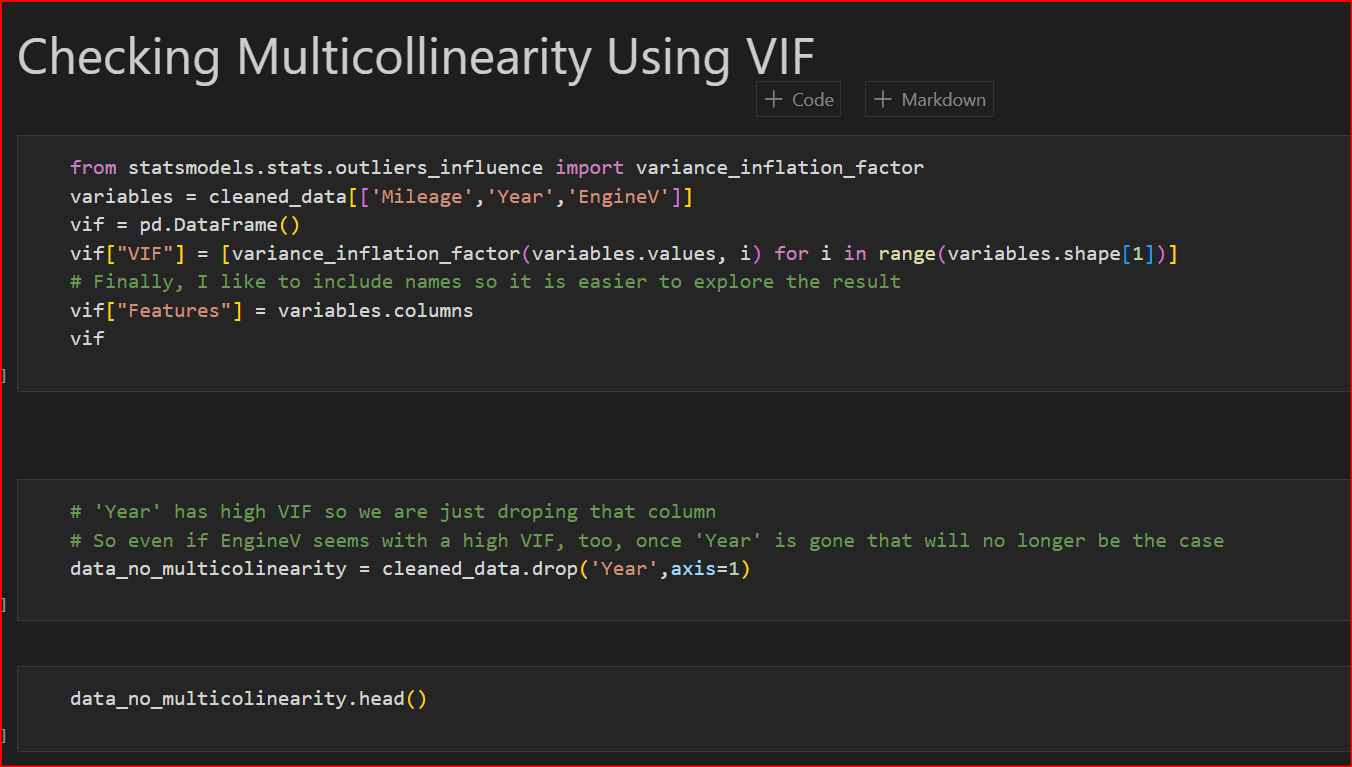
****

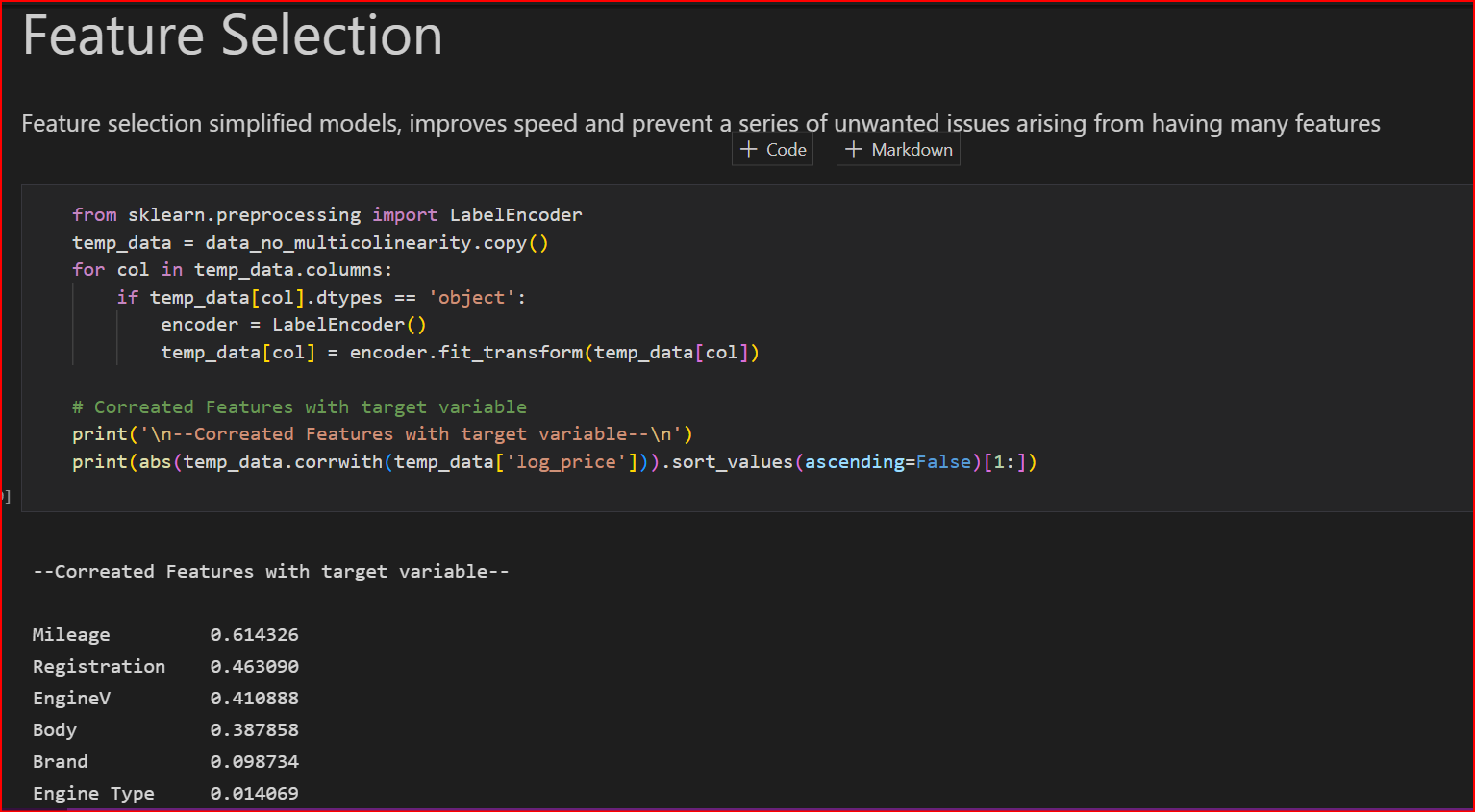
**4. Package Description:**

* Numpy and pandas were imported as to perform the basic operations.
* Matplotlib and seaborn was imported to perform the visualization.
* mean\_squared\_error and r2\_score were imported from the sklearn module to calculate the error found in the dataset and to calculate the accuracy.
* We imported variance\_inflation\_factor because it measures how much the behavior (variance) of an independent variable is influenced, or inflated, by its interaction/correlation with the other independent variables.
* LabelEncoder is imported from the sklearn module as it labels with a value between 0 and n\_classes-1 where n is the number of distinct labels.
* ExtraTreesRegressor was imported to as this class implements a meta estimator that fits a number of randomized decision trees.
* StandardScaler was imported to scale the dataset.
* Train\_test\_split was imported from the sklearn to divide the dataset into train and test
* LinearRegression is imported as we have to perform the Linear regression for the dataset
* RandomForestRegressor was imported from the sklearn module to perform the random forest algorithm
* GradientBoostingRegressor was imported from sklearn module to decrease the Bias error.
* Metrics is imported from the sklearn to depict the error and the performance of the model.

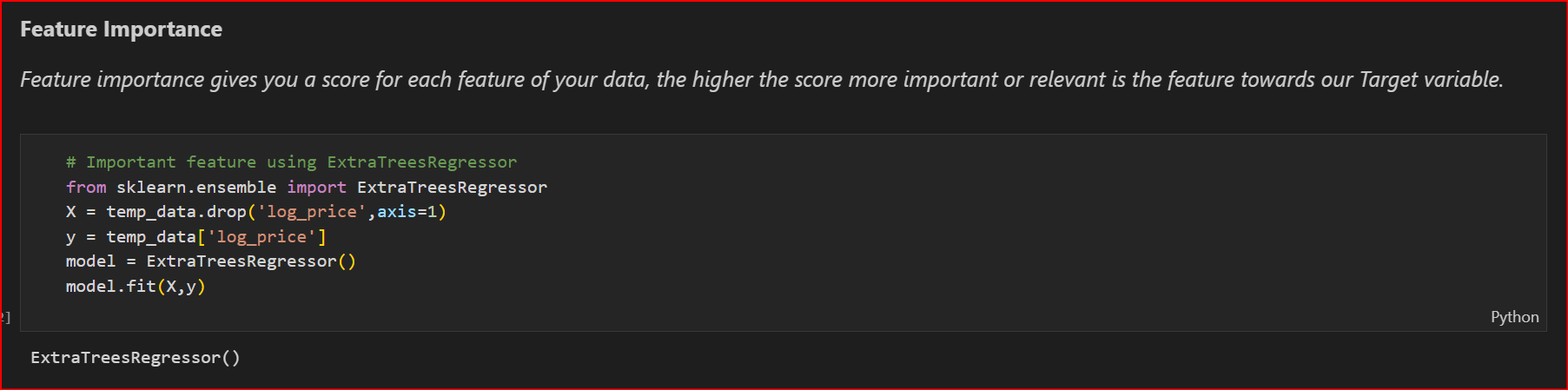
****

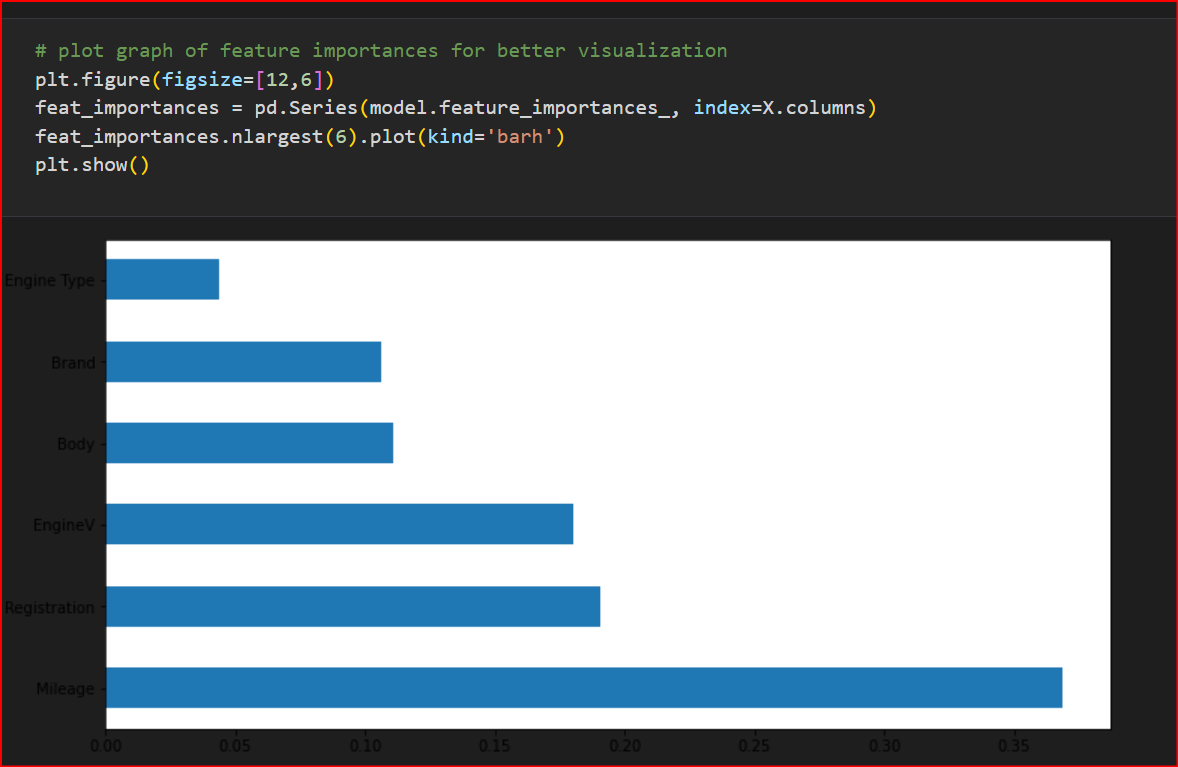
****

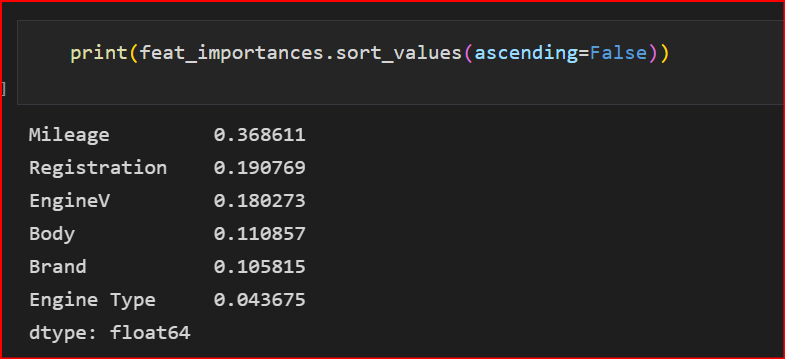
****

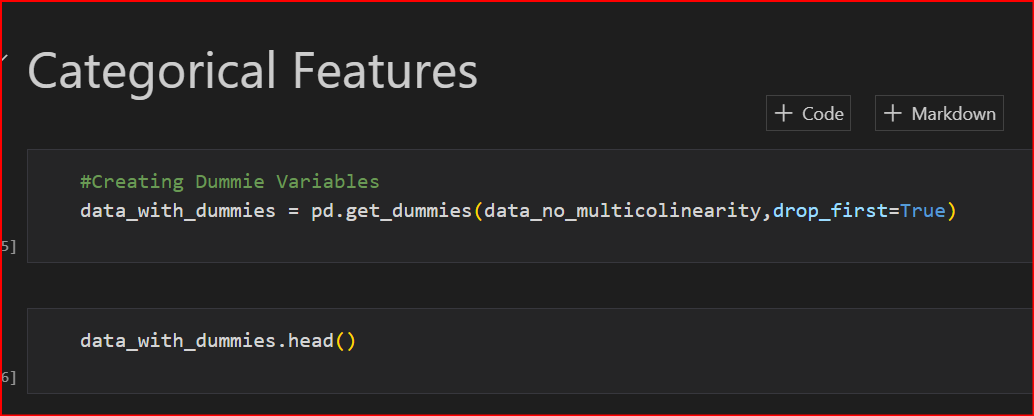
****

****

****

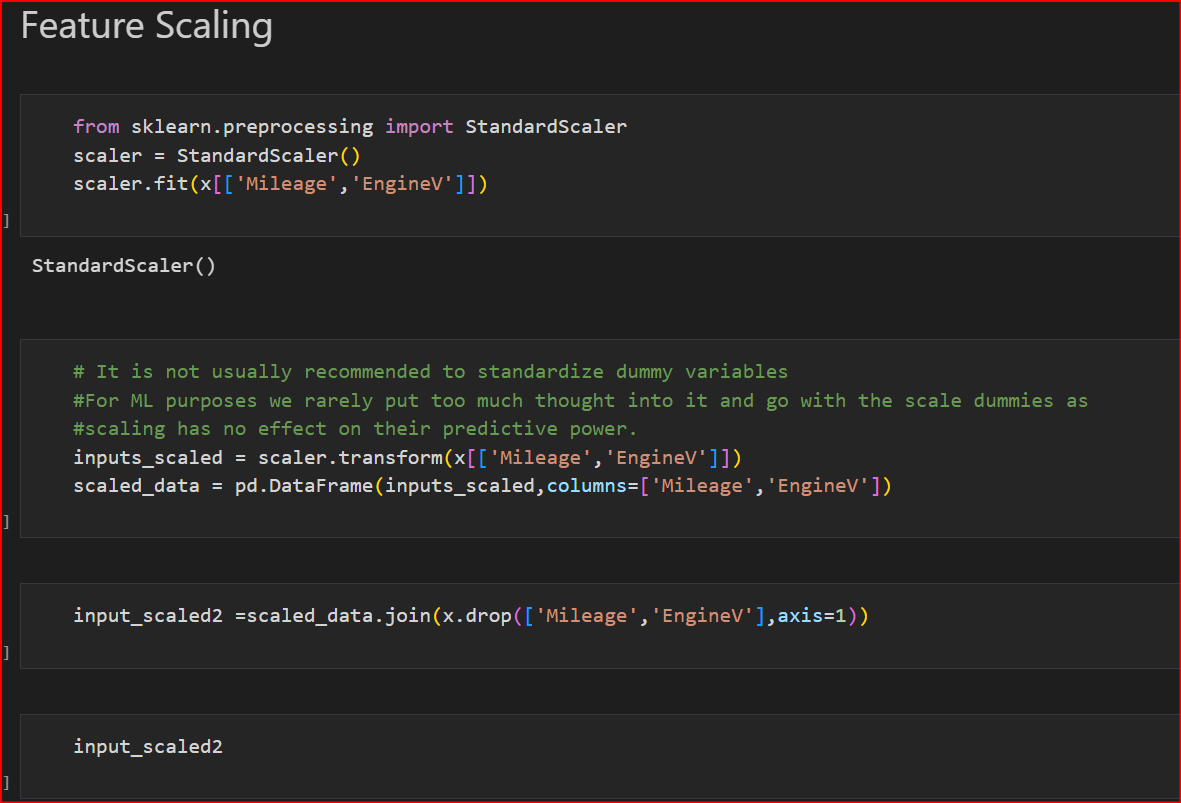
****

****

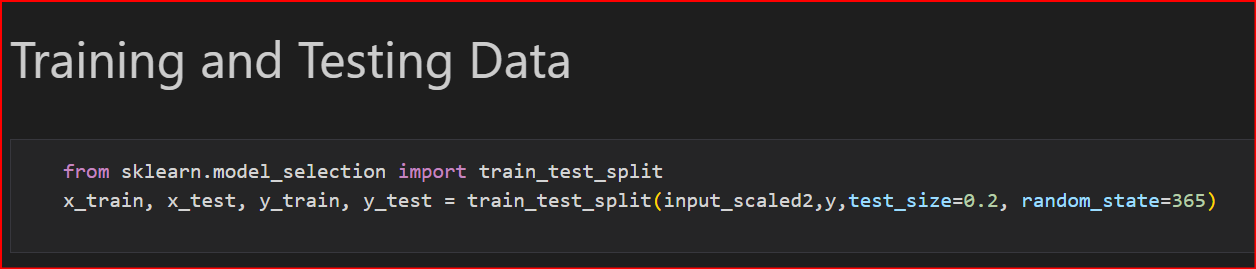
****

**5. Learning Algorithm**

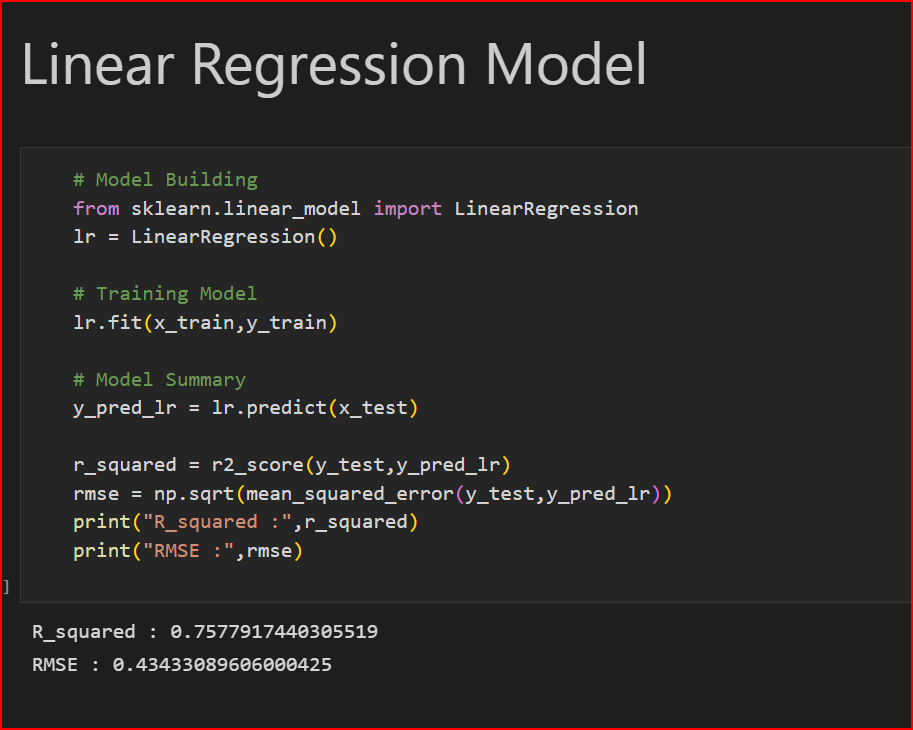
****

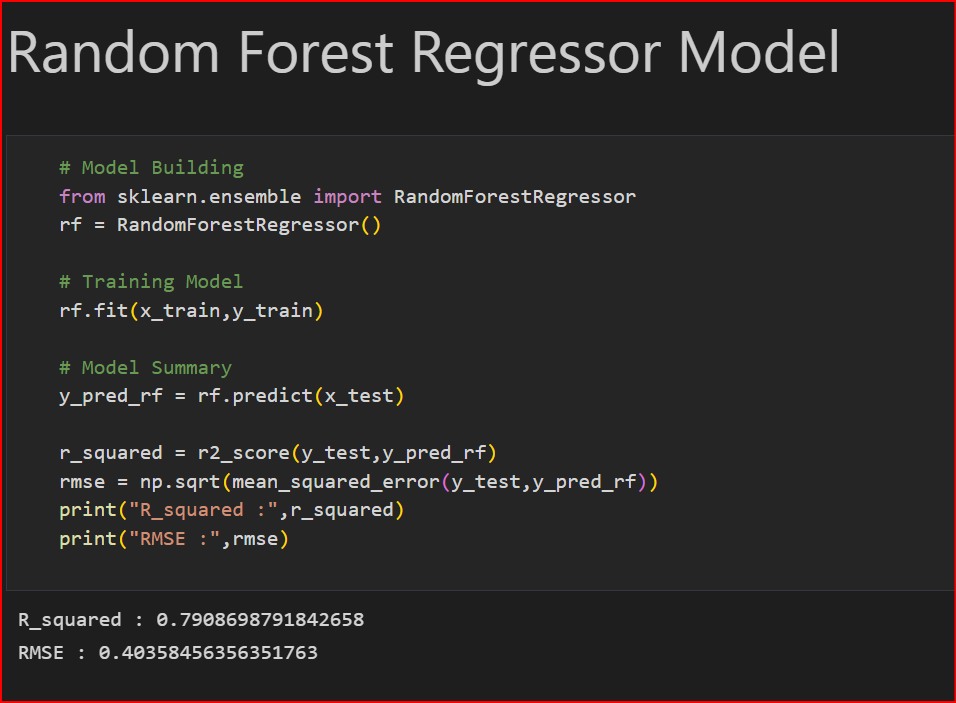
****

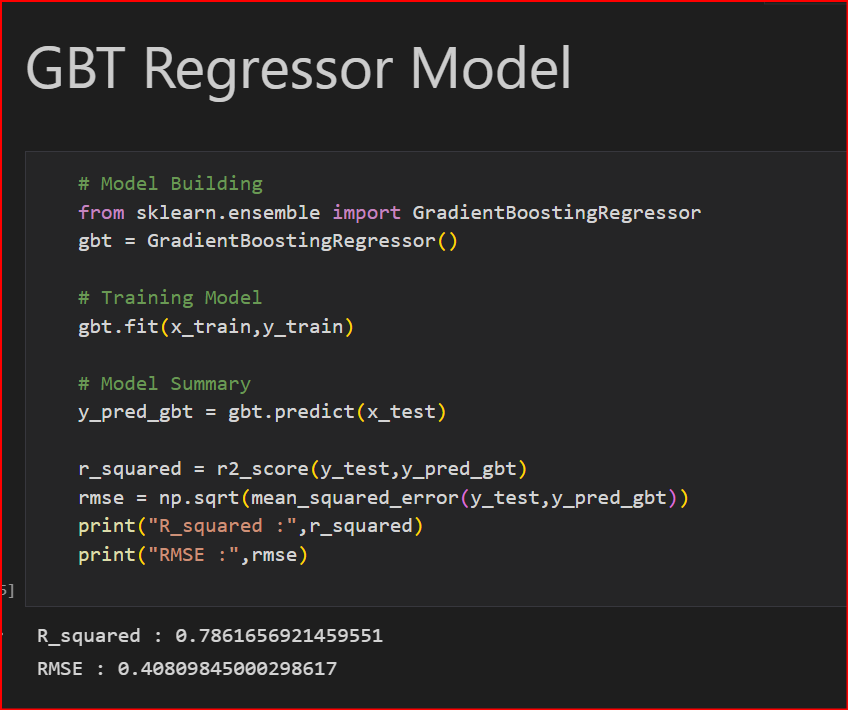
Splitting data into training, validation, and testing:

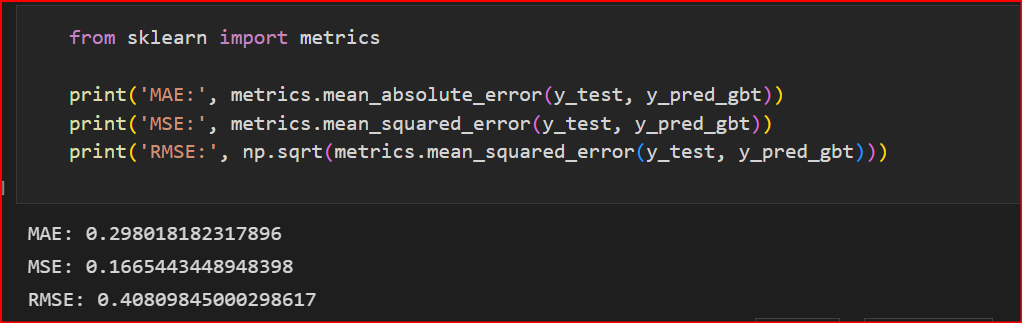
****

Brief on each ML algorithm chosen for creating model from your dataset with proper justification:

****

****

****

****

After comparing all 3 model’s accuracies,

So, as we try different Regression Algorithms and found that "Random Forest Regressor Model" is giving better accuracy compare to others with 0.798 accuracy.

****

**Conclusions:**

* Present price of a car plays an important role in predicting Selling Price, one increases the other gradually increases.
* Car age is affecting negatively as older the car lesser the Selling Price.
* Selling Price of cars with Fuel type Diesel is higher.
* Car of Manual type is of less priced whereas of Automatic type is high.
* Cars sold by Individual tend to get less Selling Price when sold by Dealers.