**Model: Predicting The Used Car Selling Price**

**Aim:** The Main Aim Of this Model is to Predict the Selling Price of the Used Car for the location Mumbai

**Problem Statement:**

**The Prices Of New Cars In The Market Is Generally Fixed By The Manafacturer Company With Some Additional or Hidden Costs Incurred By The Government. So Customers Buying New Car Can Be Assured Of The Money Invest To Worthy.**

**But Due To The Increased Of New Car Prices And Incapability Of Customers To Buy New Cars. So Used Cars Sales Are On a Global Increase, There Is a Need For A Used Car Price Prediction System.**

**THIS PROJECT MODEL AIMS TO PREDICT THE SALES PRICE OF USED CARS**

**For this Model prediction we use 9 Parameters those are the 9 columns for the dataset**

1. **Year:**

The car purchased year is mentioned in the dataset. Year of the car Purchased by the owner.

1. **Purchase Price:**

In this The Purchased Price of the car is mentioned in lakhs in

The location of Mumbai

1. **Kms\_Driven:**

In this how many kilometers driven by the car from purchasing year to till the year

1. **Fuel Type:**

In this which type of Fuel is used by the Car like Petrol,Diesel,CNG etc…

1. **Seller Type:**

In this Who is the Seller for the car..i.e wheather the car is selled by the Dealer or Individual

1. **Transmission:**

A Transmission Changes the gear depend upon the speed either it is a manual..i.e By humans or Automatic..i.e Machine

1. **Owner:**

In this the owner of the Car is mentioned

1. **Location:**

The whole dataset is in the location of Mumbai to predict the selling price of the used car

1. **Selling Price:**

In this we have to Predict the Selling Price of the given used car using given independent parameters

**Steps for Building a Predecting the used car price model:**

1. Collection of data and loading the data.
2. Data processing
3. Visualization
4. Separating training set and testing set
5. Model implementation
6. Model Selection
7. Model Evalution

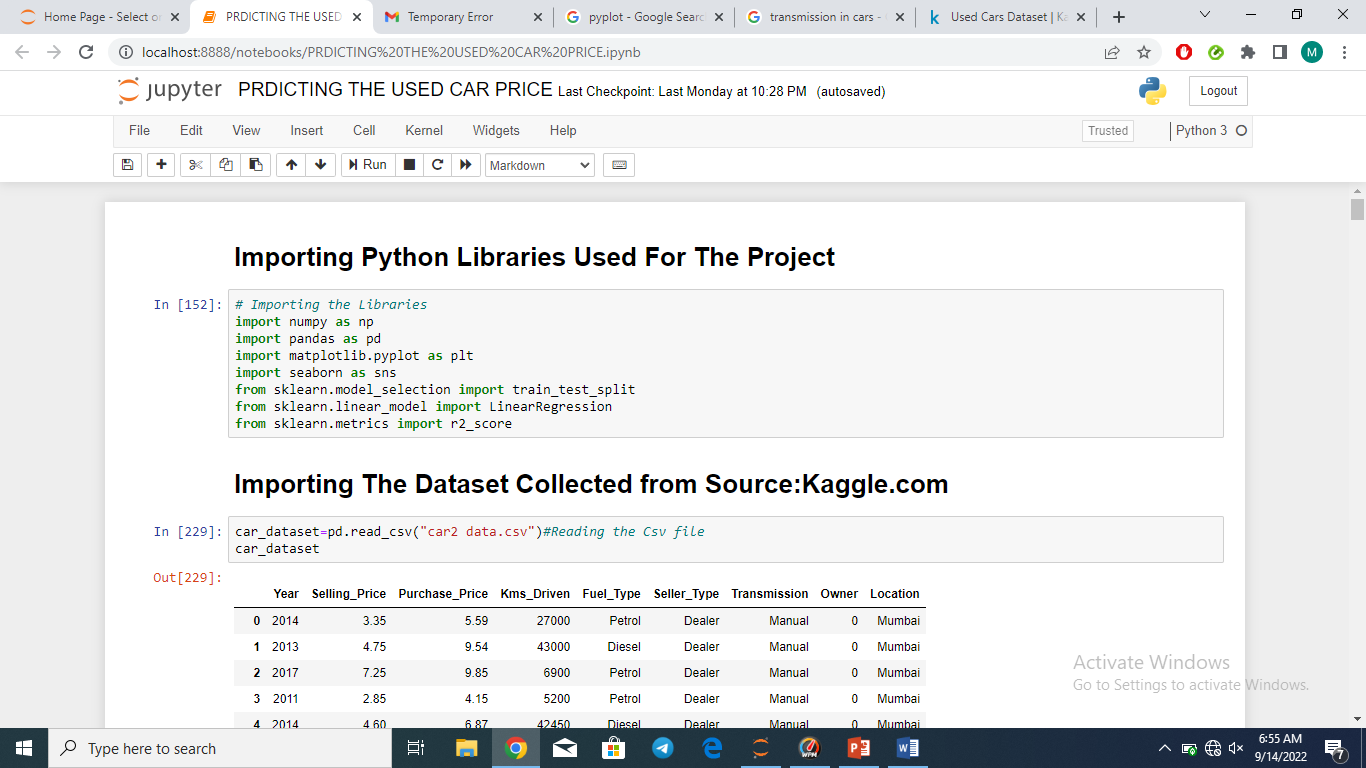
## 1)Collection of data and loading the data:

The dataset used in this model is from Kaggle Website.

Dataset Link:

<https://www.kaggle.com/datasets/austinreese/craigslist-carstrucks-data>

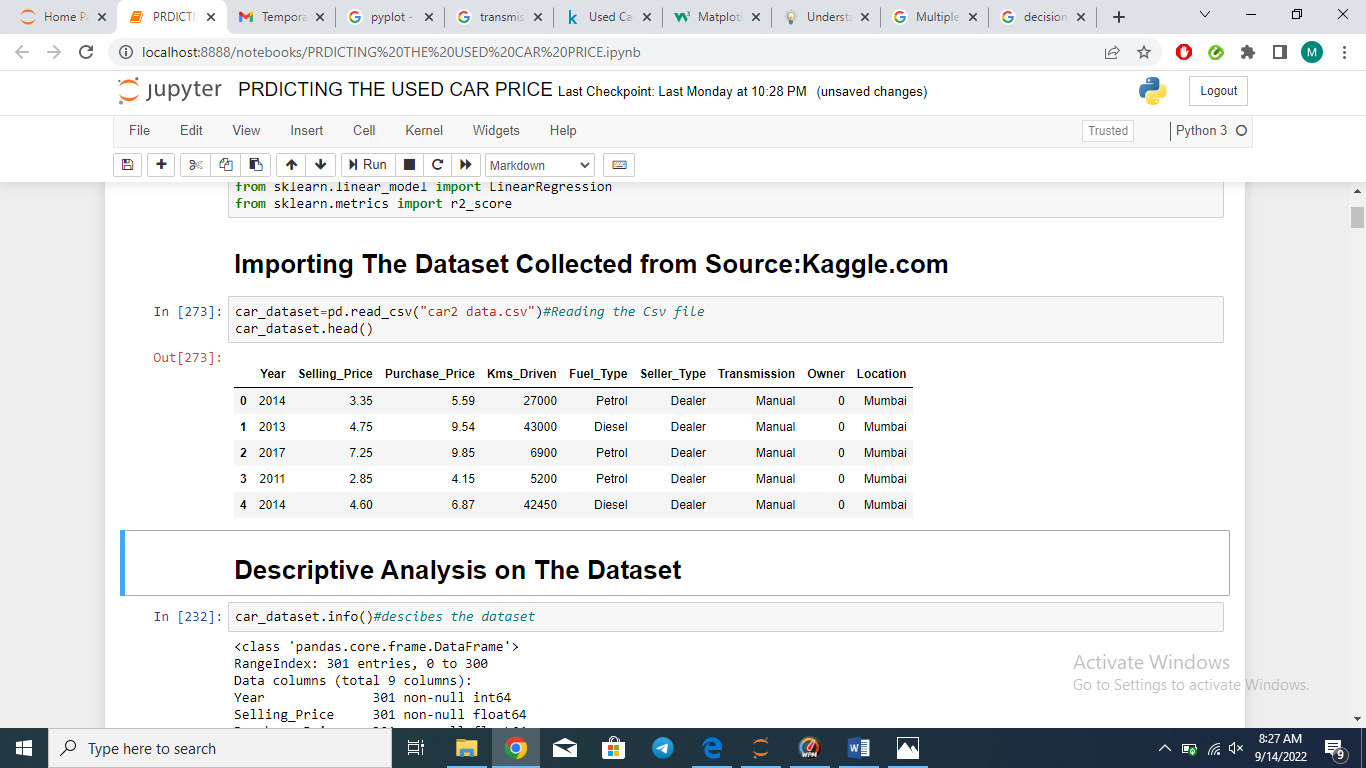
* **IMPORTING THE lIBRARIES:**
* In this We have to Import All the Libraries which are used in the model project

First we have to import the packages which are required to our model. The packages are: 

* LOADING THE DATAFRAME:

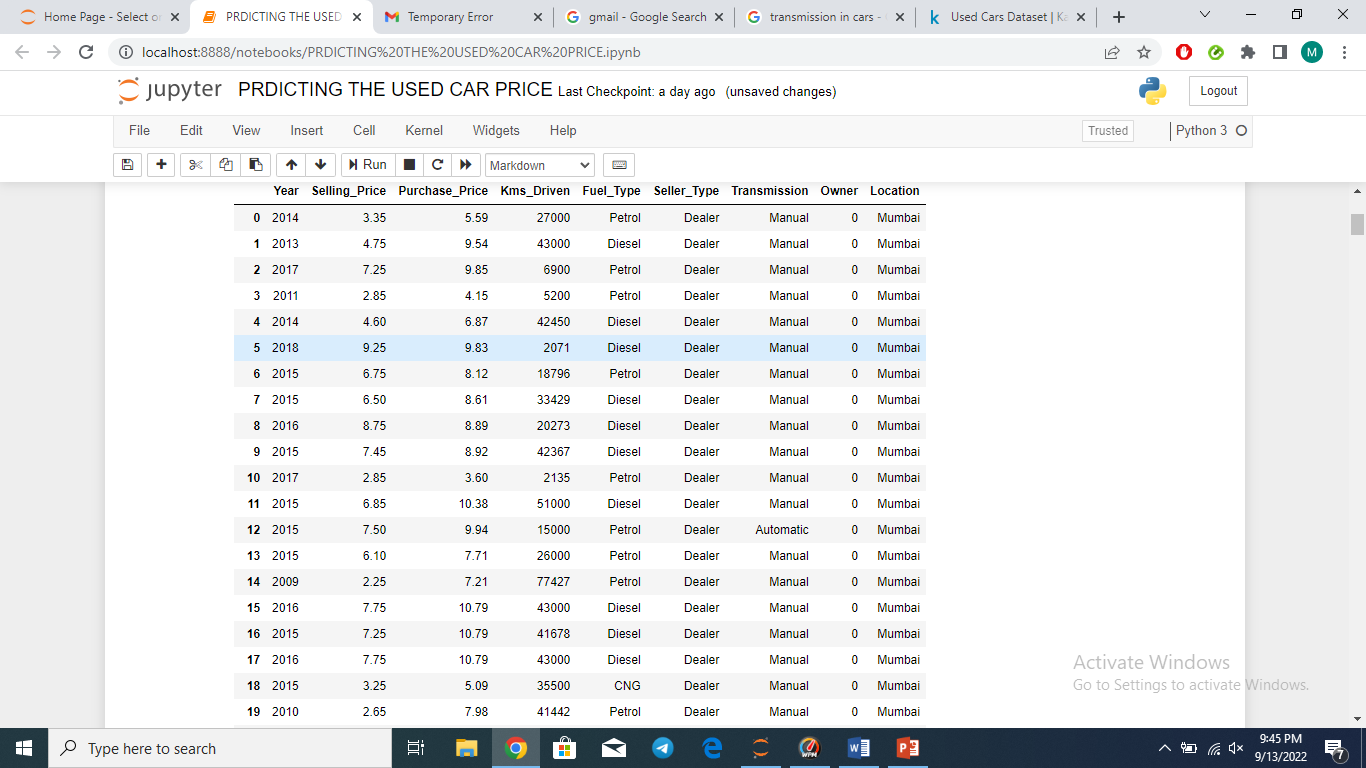
Car\_dataset=pd.read\_csv(“car2 data.csv”)

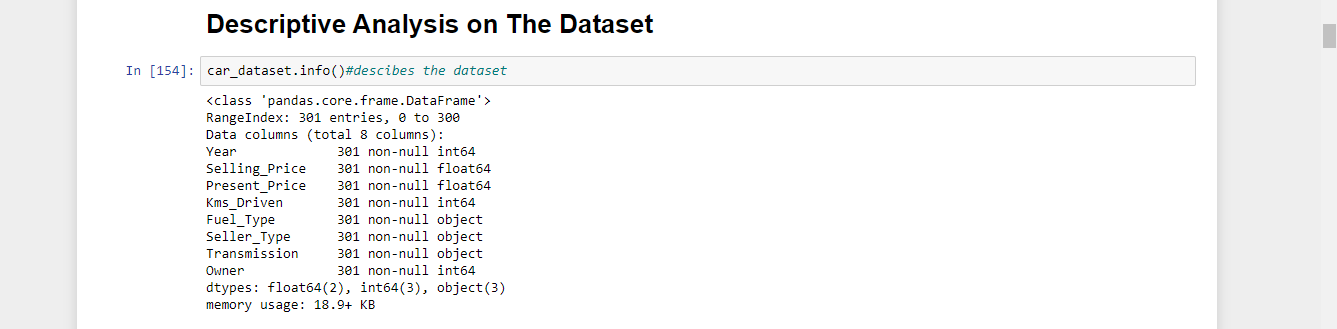
## It converts the csv file into dataframe



* **LOADING THE DATA IN A CSV FILE TO DATAFRAME:**

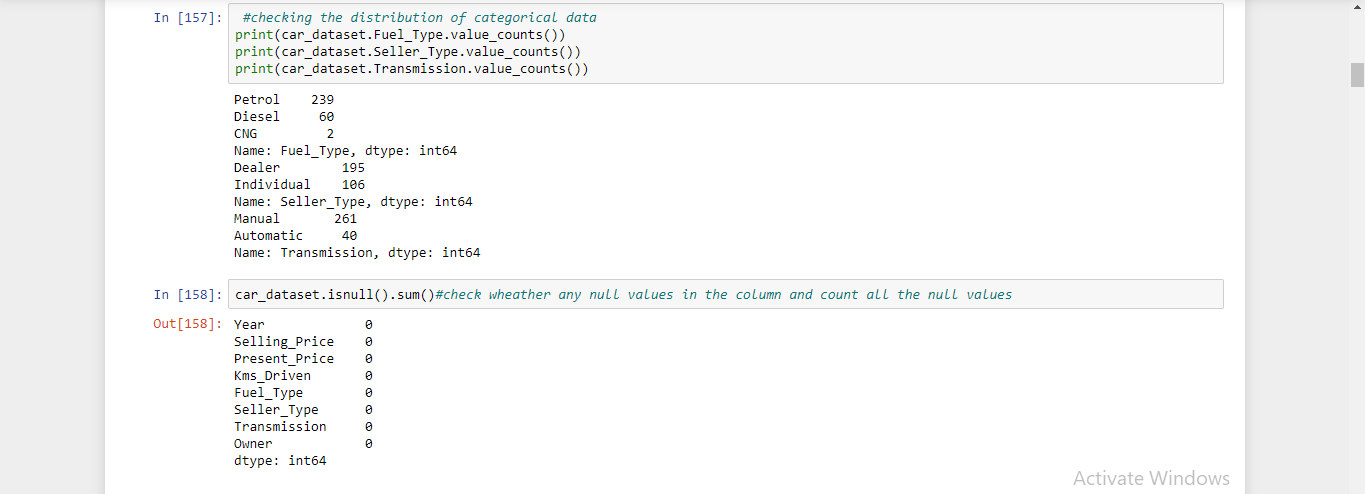
As I have loaded the data into “dataset.csv” file, by using pandas dataframe and we can easily import this file.



**2) Data Processing:**

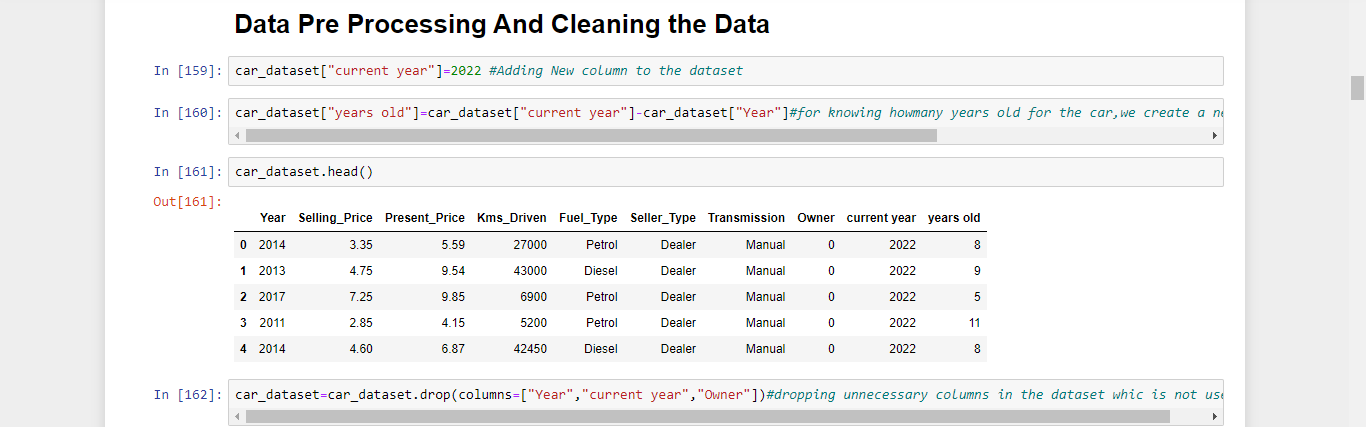
In this we are getting the information about the dataset by using the info()

Before implementing a model, first we have check whether the data is completed with non-null values are not.



Here, we did not get any null values, So we can proceed to build a model.

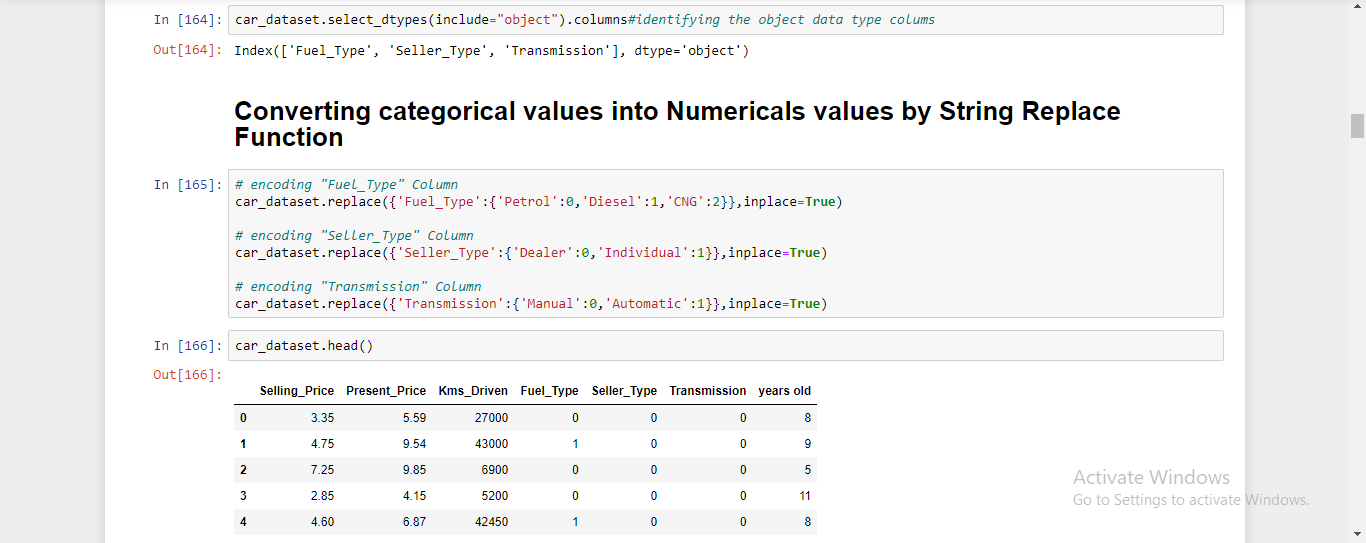
* **Data Pre Processing And Cleaning The Data:**

****

In this We are Dropping the Unnecessary Columns for Predicting the Selling price of Used car those columns are:

* YEAR
* CURRENT YEAR
* OWNER
* LOCATION

# Converting categorical values into Numericals values by String Replace Function:



**In this Fuel type, Seller type, Transmission are the categorical values and these are converted into Numerical values by string replace function**

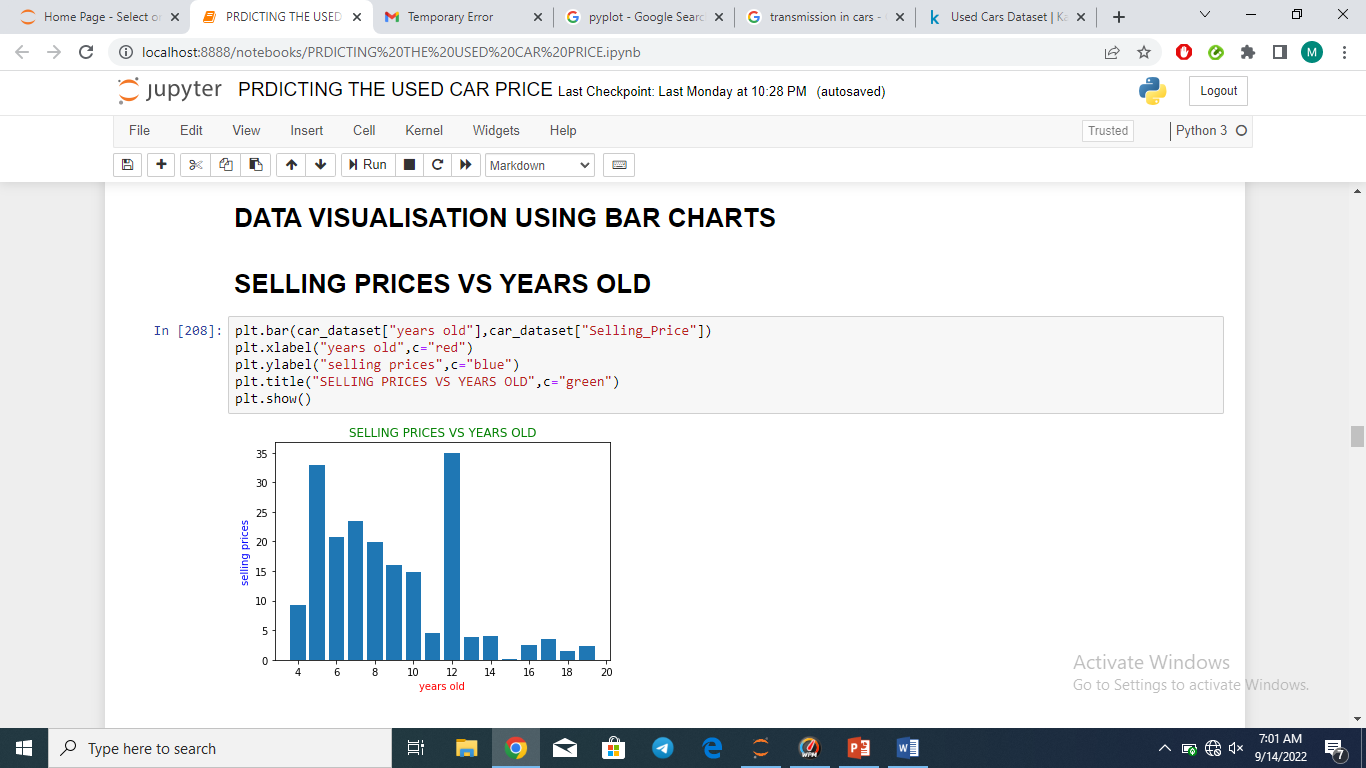
## 4) Visualization:

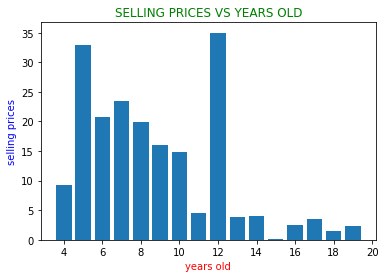
By using the Matplotlib.pyplot library, we plot the Data

**Pyplot:**

Pyplot is a collection of functions that make matplotlib work like MATLAB. Each pyplot function makes some change to a figure: e.g., creates a figure, creates a plotting area in a figure, plots some lines in a plotting area, decorates the plot with labels, etc. In matplotlib.

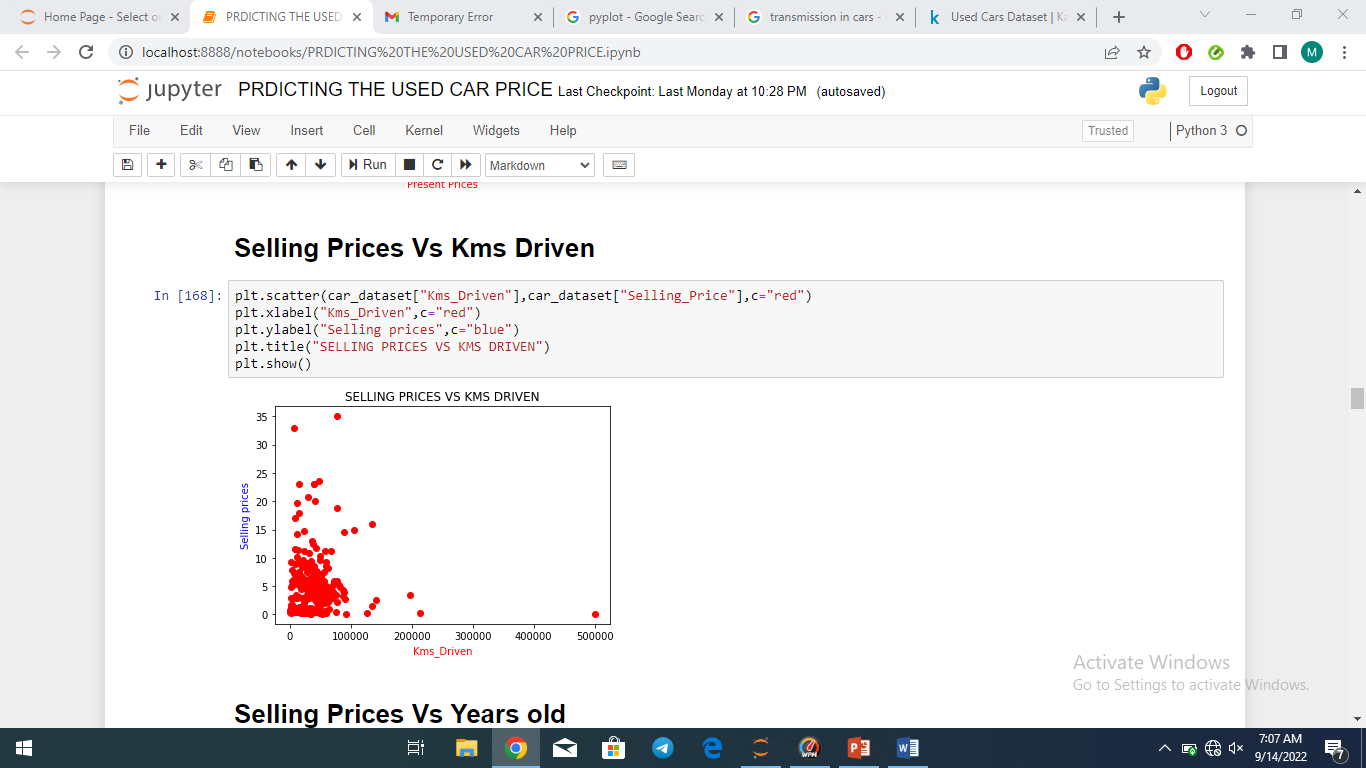
**Visualisation of Selling Prices Vs Years Old by using the bar charts:**

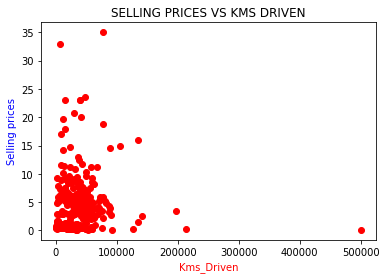
****

****

By Observing the above bar chart we know that if the years are increasing gradually the selling price of the car is decreasing… But some cases the selling price increases because remaining attributes or independent variables most of them cars selling price gradually decreasing when purchasing years are increasing.

**Visualisation of Selling Prices Vs Kms Driven by using the Scatter plot :**

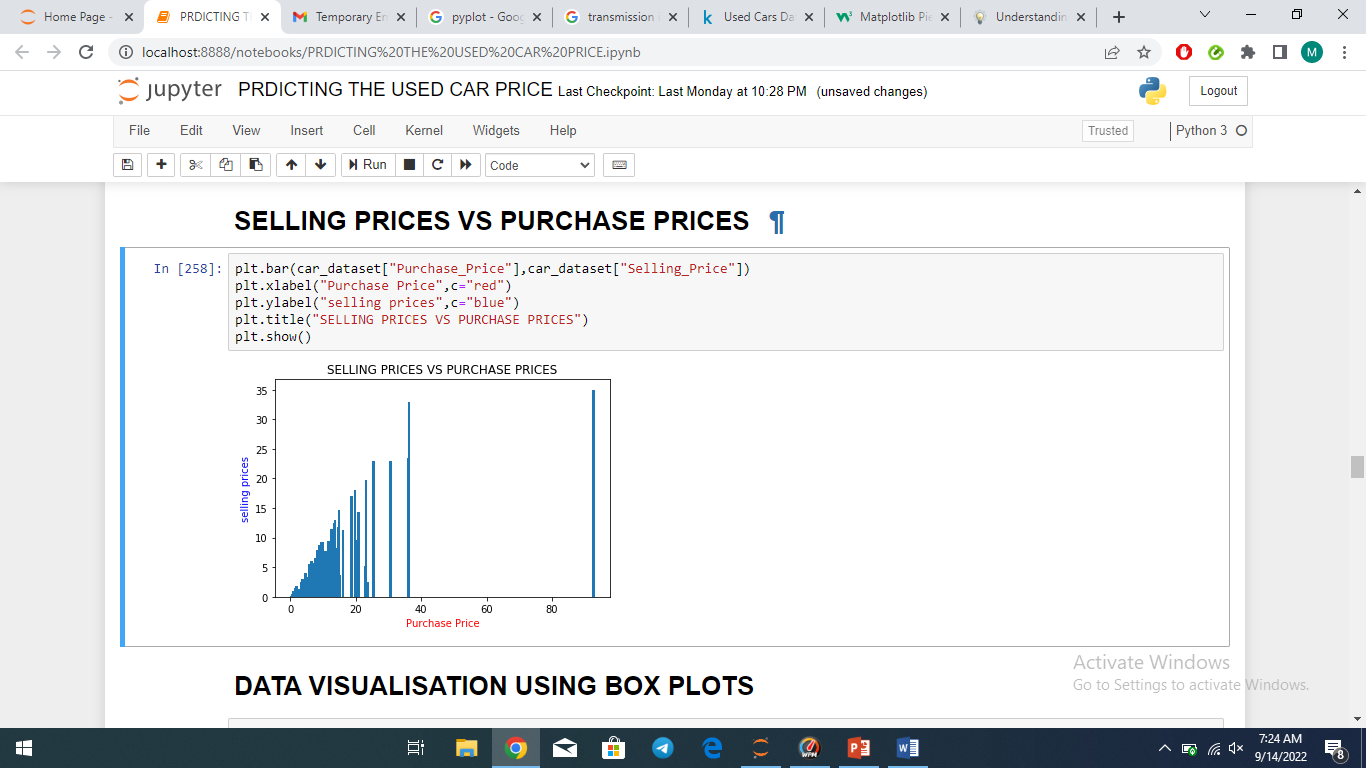
****

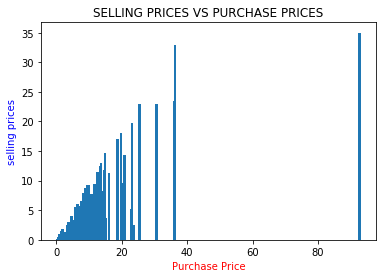
****

By the Above Scatter Plot we are observing that there is a gradually decreasing of selling price when kms driven is increasing accordingly. Most of the Cars selling price is depend upon the kms driven by the car from Purchasing to till now. We can easily identify the relation between the selling price of the used car and the Kms driven.

**Visualisation of Selling Prices Vs Purchase Prices by using the**

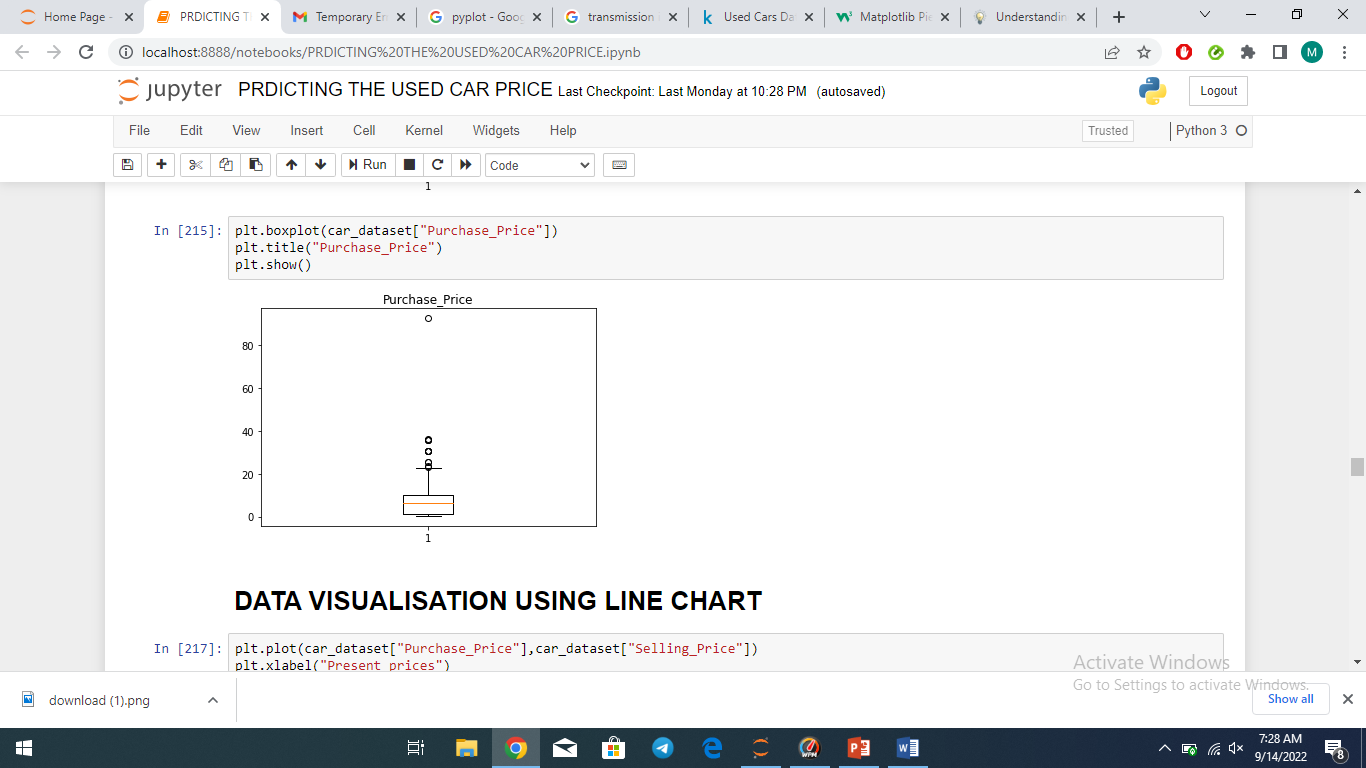
**Bar chart :**

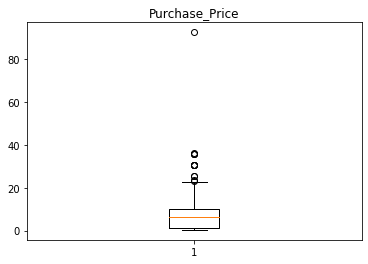




By Above bar chart we observe that the car used selling prices and Purchase prices and selling prices are in between the 35-40 lakhs. By this we observe that there is not much differenece but it depends upion the various independent variables.

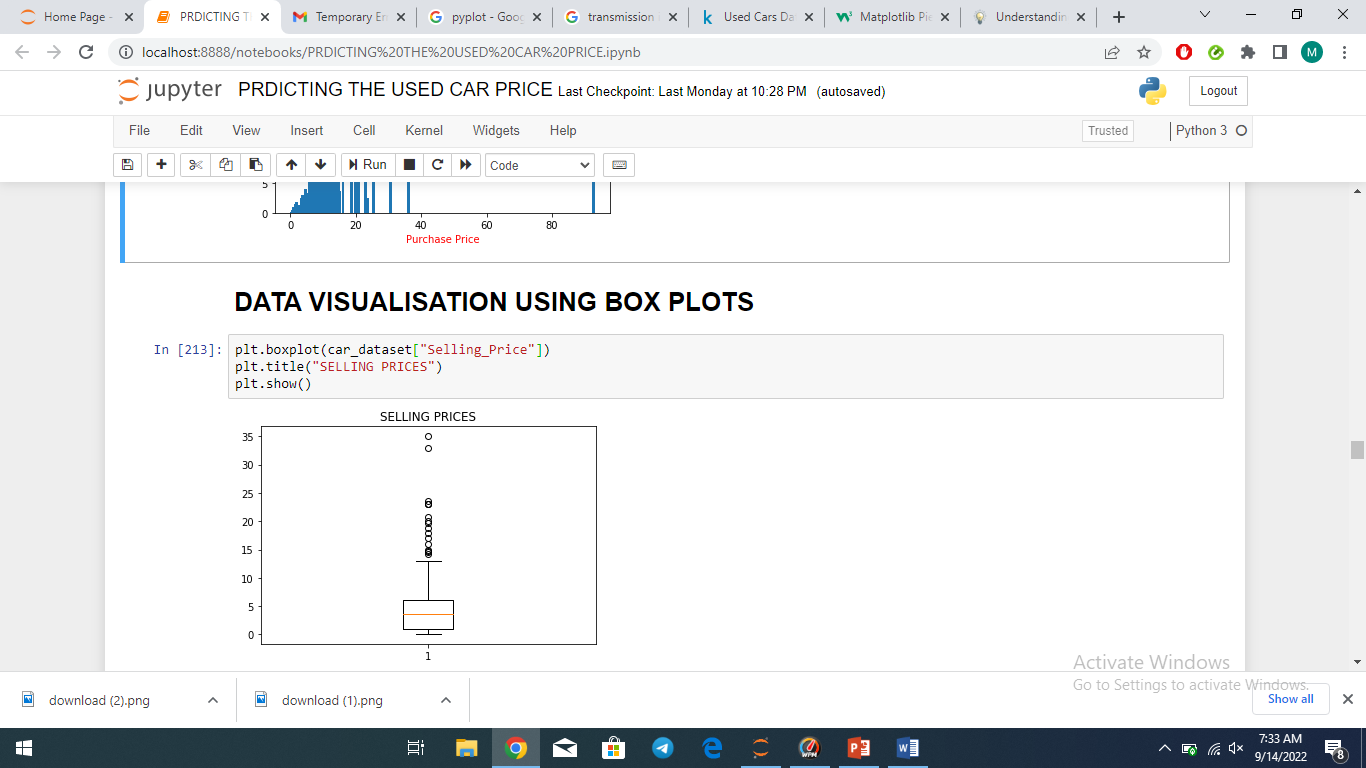
**Visualisation of Purchase Prices by using the Box Plot:**

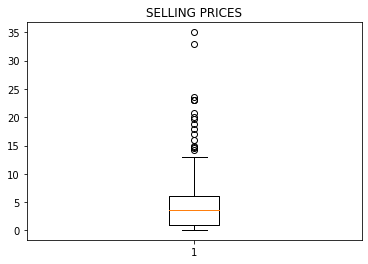
****

****

By this Box Plot we can observe that the Maximum Purchase price of the is nearly 30 lakhs and Minimum Price is 2 lakhs..and some outliers are present.. and remaining used car purchase prices in betwwn them i.e… Maximum and mininmum..

**Visualisation of Selling Prices by using the Box Plot:**

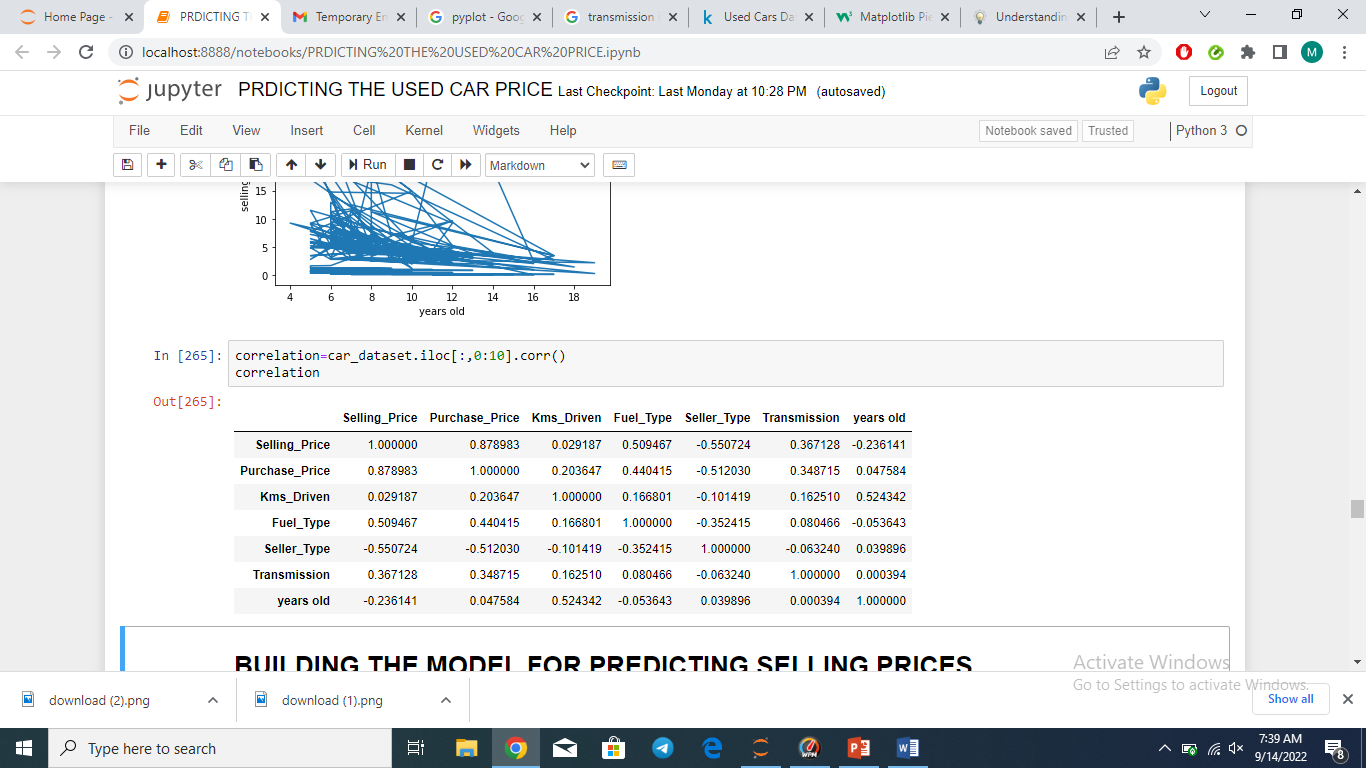




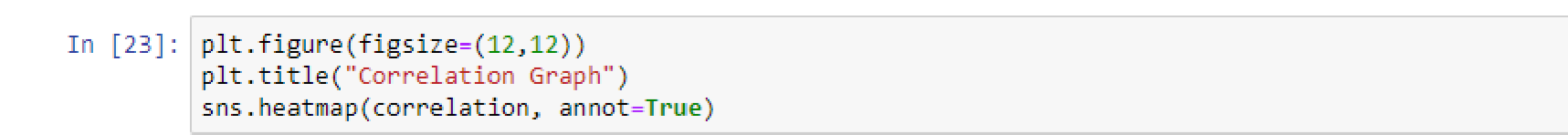
From the above Box Plot we can conclude that the maximum selling price of the car is 12 to 15 and minimum selling price is 1to 3 lakhs and in this also some outliers are present.

**Correlation:**

Correlation refers to the statistical relationship between two entities that quantifies the strength of the relationship. It’s a common tool for describing the simple relationships without making statement about cause and effect.



We observe that for every equivalent column the correlation is equal to 1



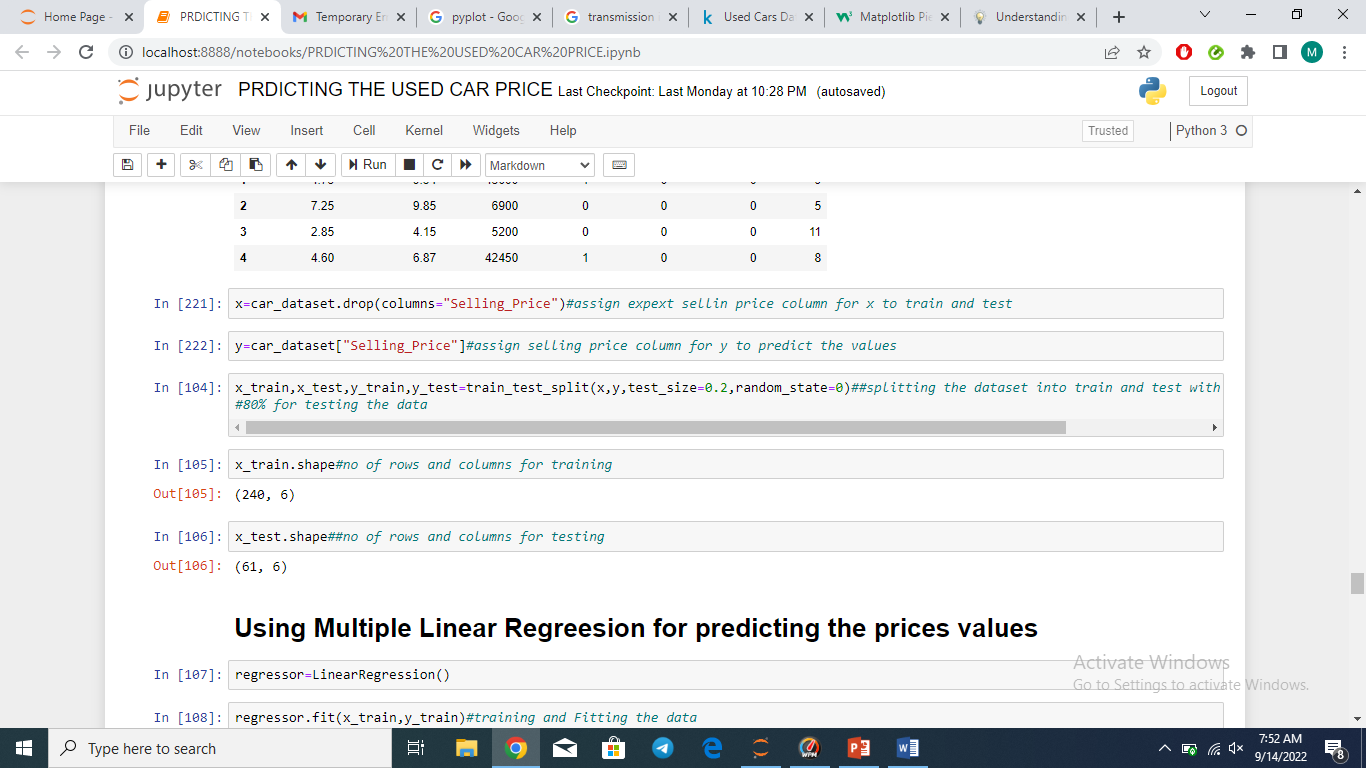


This heat map shows the correlation for each and every column in the dataset

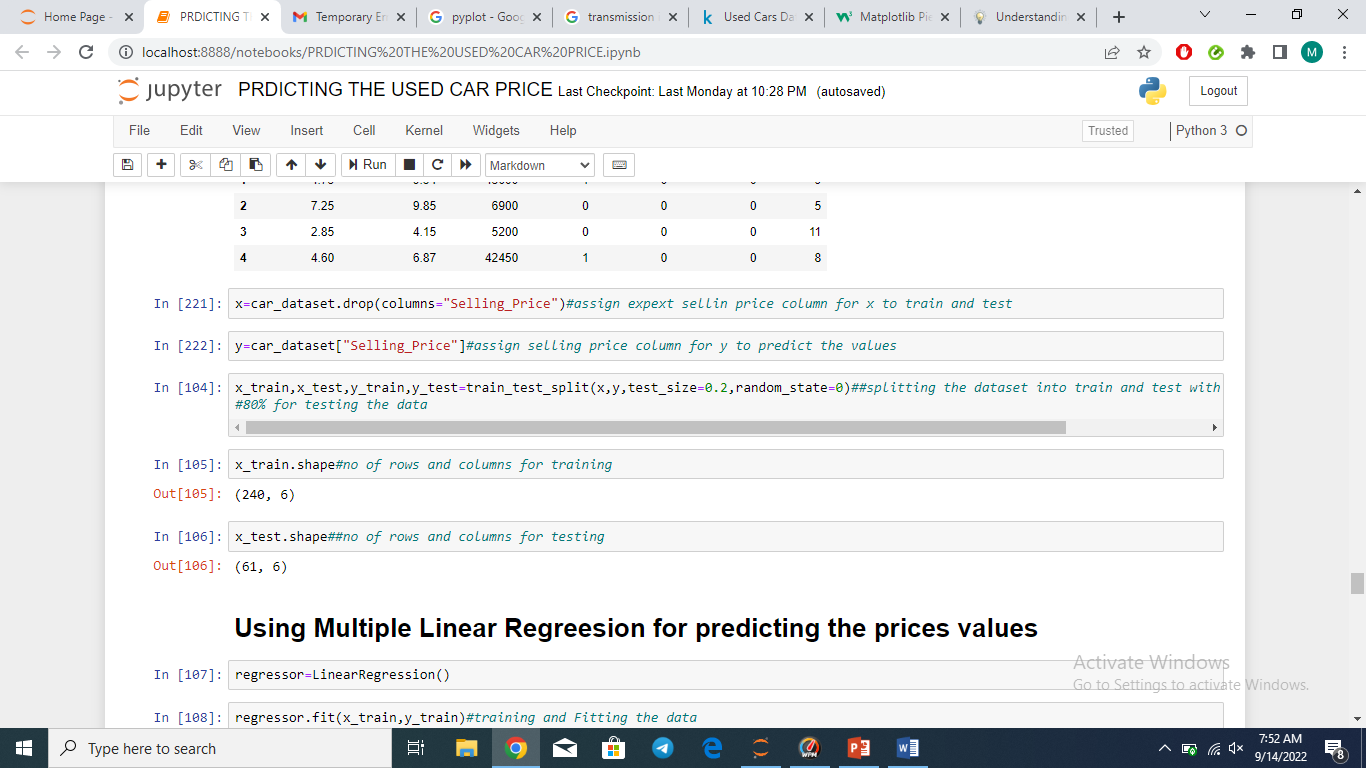
# 4) Separating Training set and testing set:

First, we have to define the x and y values from the dataset.

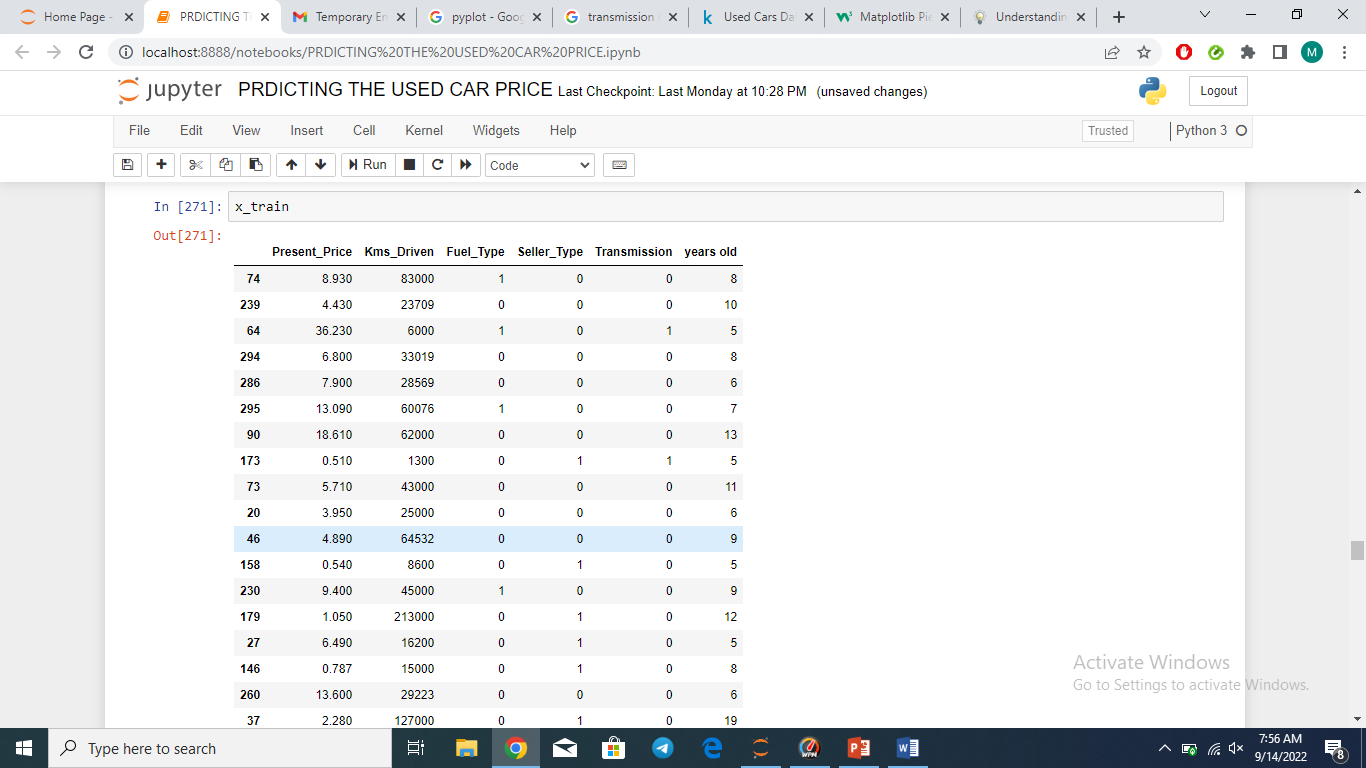
X has the independent variable and it has 9 parameter to check the predictions. Y has only one parameter, which is result of the model



After defining the x and y values, these values are separated under the testing and training by the below code:



X\_train for the dataset is randomly taken….

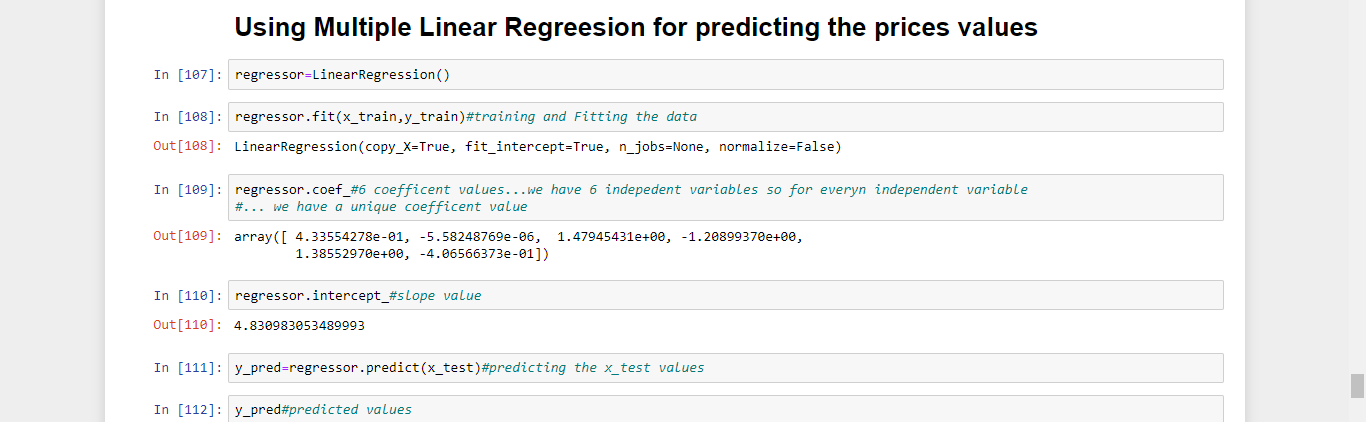


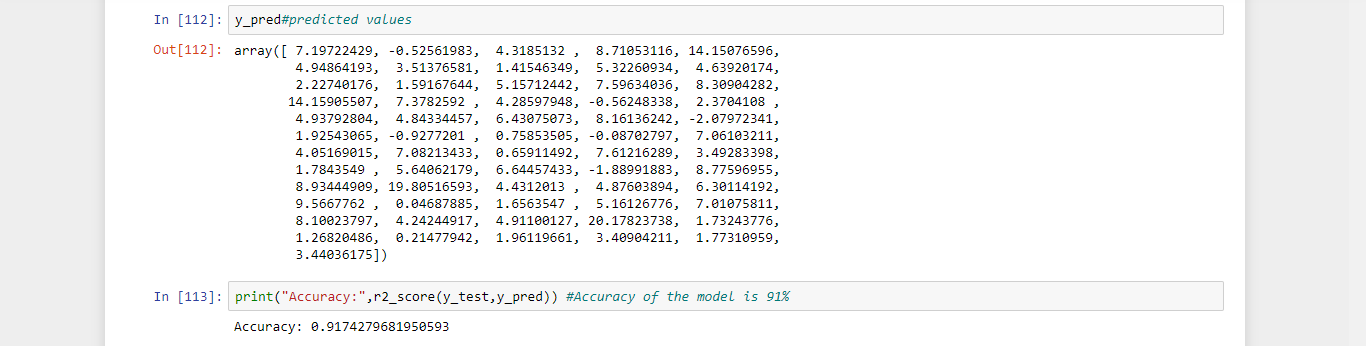
## 5) Model implementation:

For this model implementation, we are used Multiple Linear Regression and Decision tree techniques:

**1.Multiple linear regreesion:**

Multiple linear regression (MLR), also known simply as multiple regression, is **a statistical technique that uses several explanatory variables to predict the outcome of a response variable**. Multiple regression is an extension of linear (OLS) regression that uses just one explanatory variable.

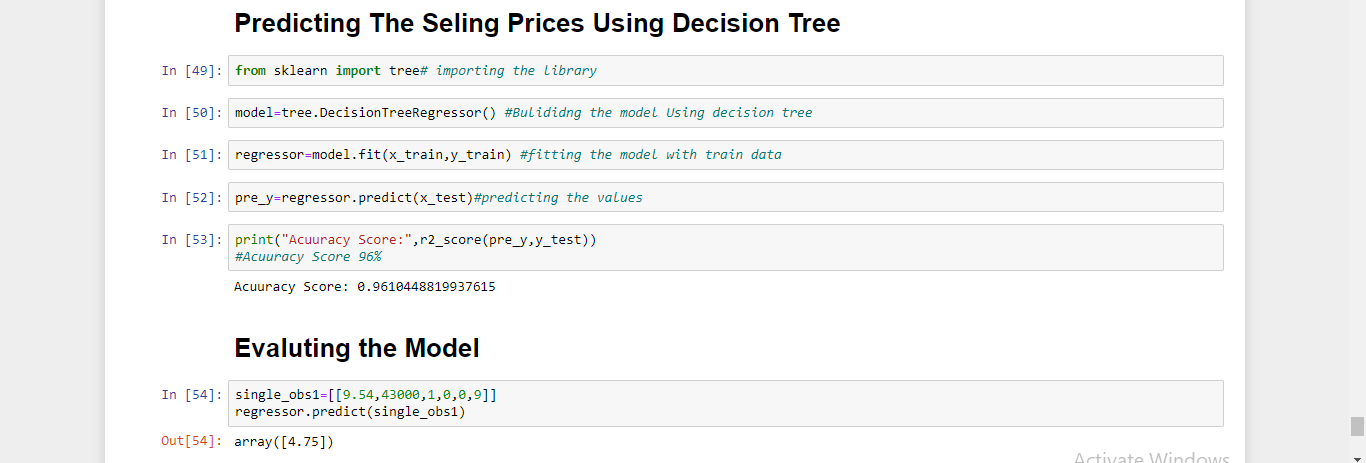




In this y\_pred gives the prediction values of the Selling Prices of the used car and the Accuracy of this is 0.91 but it is for testing accuracy..

**2.Decision Tree:**

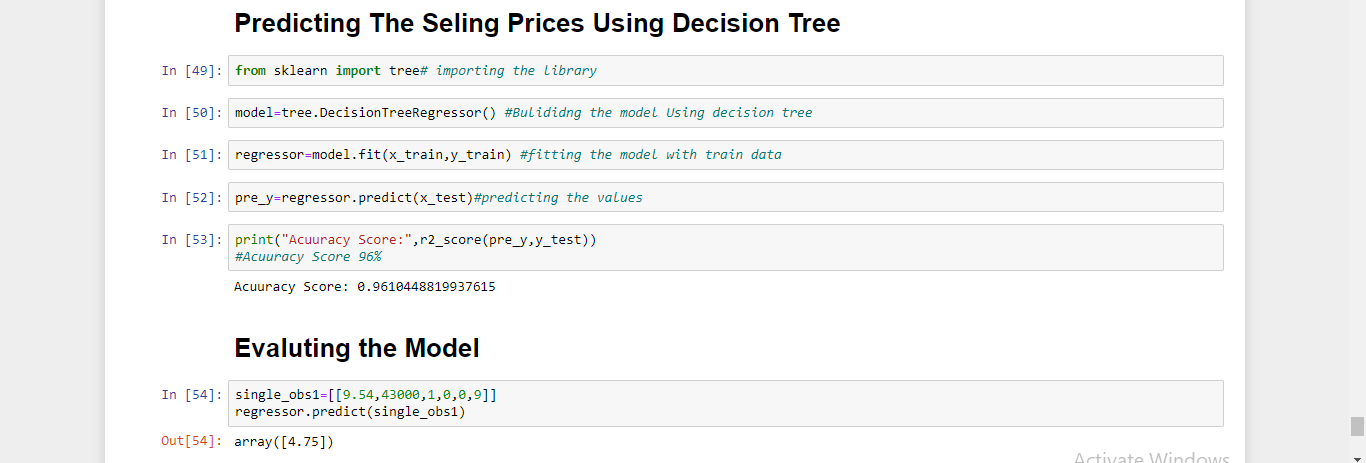
Decision tree regression **observes features of an object and trains a model in the structure of a tree to predict data in the future to produce meaningful continuous output.**



For this pre\_y gives the prediction values of the used car selling prices and the accuracy of the testing is 0.96%

# 6) Model Selection:

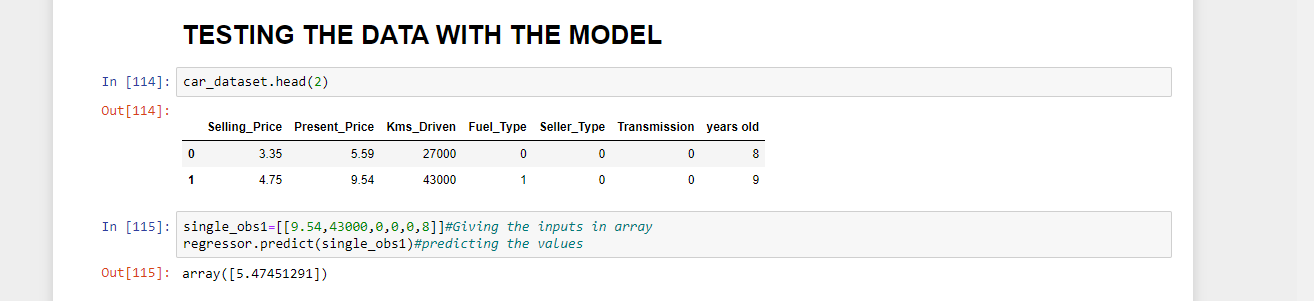
Among these 2 different techniques, we choose the one, which is the highest accuracy value i.e 0.96, and the library used is DecisionTree.



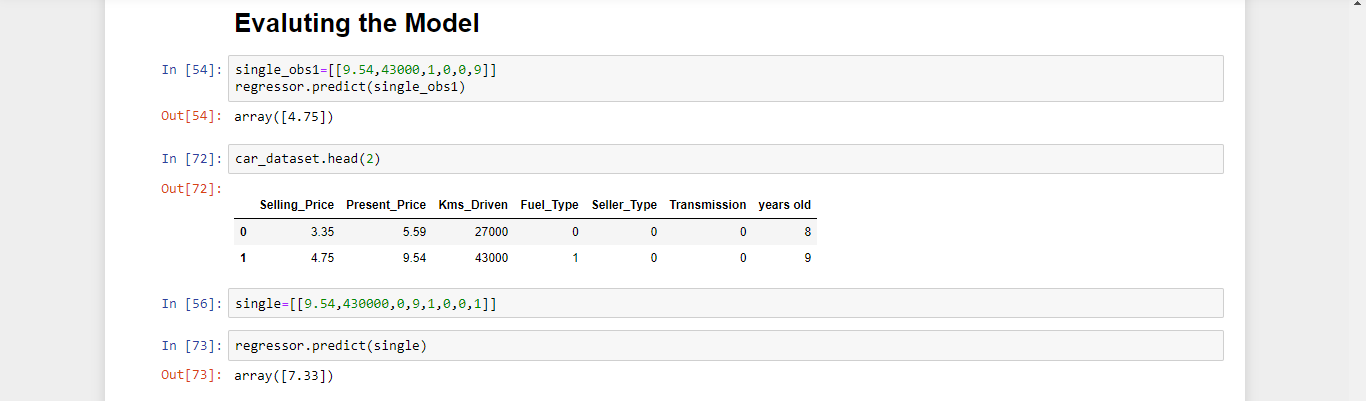
# 7)Model Evalution:

Now, we have to check the model to give the random values for the each library.

1.Multiple Linear Regression:



2.Decision Tree:



For both the models the predicted values are Overfitted they failed while testing the data.. To overcome this we have to train the model with more dataset.

## Conclusion:

We Predict that the Car used Selling Price of the Mumbai Location by using different techniques and we take the highest accuracy technique for this .i.e.. Decision tree