

CAR PRICE PREDICTION

Submitted by:

NITIN GUPTA

**ACKNOWLEDGMENT**

I wish to express our gratitude to our mentor .Sapna Verma for providing me an opportunity to do my research project work on "CAR PRICE PREDICTION". Under her guidance I have completed my project report and tried my best to implement what I had learnt till now .She also helped me by updating me about the information of what to do and not to do during my research project and help me with all

**ABSTRACT**

This dataset consists information about used car listed on cars24.com. It has 8 columns each columns consists information about specific features like Car\_Name gives information about car company.Which Year the brand new car has been purchased. Last\_price the price at which car is being sold this will be target label for further prediction of price .Driven\_KM number of kilometre car has been driven. Fuel this feature the fuel type of car (CNG ,petrol, diesel etc). No. of Owners tells whether the Owner is 1st owner second or third owner .Location gives information about the Location of car where it is

**INTRODUCTION**

* Business Problem Framing

With the covid 19 impact in the market, we have seen lot of changes in the car market. Now some cars are in demand hence making them costly and some are not in demand hence cheaper. One of our clients works with small traders, who sell used cars. With the change in market due to covid 19 impact, our client is facing problems with their previous car price valuation machine learning models. So, they are looking for new machine learning models from new data. We have to make car price valuation model. This project contains two phase-

**Data Collection Phase:**

I have collected data from Cars24.com.

I have collected data from Cars 24.com using Data Scraping using Selenium technique.

**Analytical Problem Framing**

* Data Sources and their formats

Data scrap from Cars24.com using selenium.

* Data Preprocessing Done

In Pre-processing step I changed feature data as ML friendly to remove rupee sign and removing commas, Make data columns Last Price and Driven\_Km in Float type. Checking counts as pr countplots for object data type data distribution among variables. Checking Outliers there are few outliers present in data which was removed .There is no skewness in data for continuous columns.

* Data Inputs- Logic- Output Relationships

We have split data into features and target columns

* Hardware and Software Requirements and Tools Used

I have used Jupyter NoteBook to run the code for both Data Scraping as well as model Building.

**Model/s Development and Evaluation**

* Identification of possible problem-solving approaches (methods)

I have imported 5 models K-Neighbors Regressor, Decision Tree Regressor, Random Forest Regressor, Adda Boost Regressor and Ridge Regressor for if overfitting exist. While I checked R2 score on different random states for different models I found Random forest regressor gives the best result on Random\_state 11

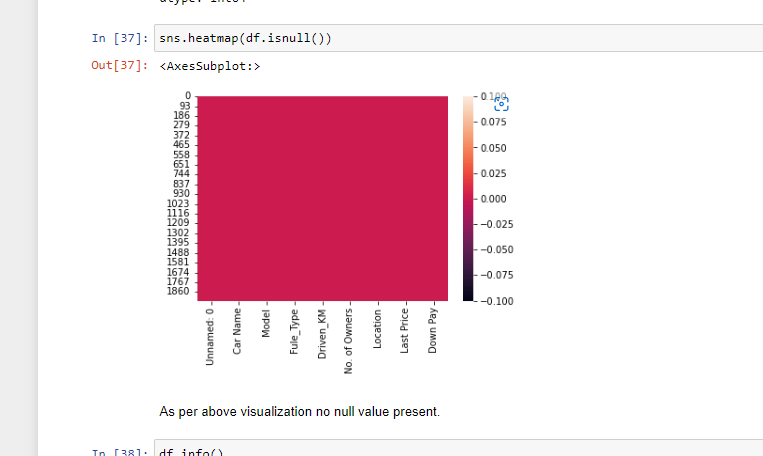
* Run and Evaluate selected models

So I have selected Random Forest model is best model as per random state=11.

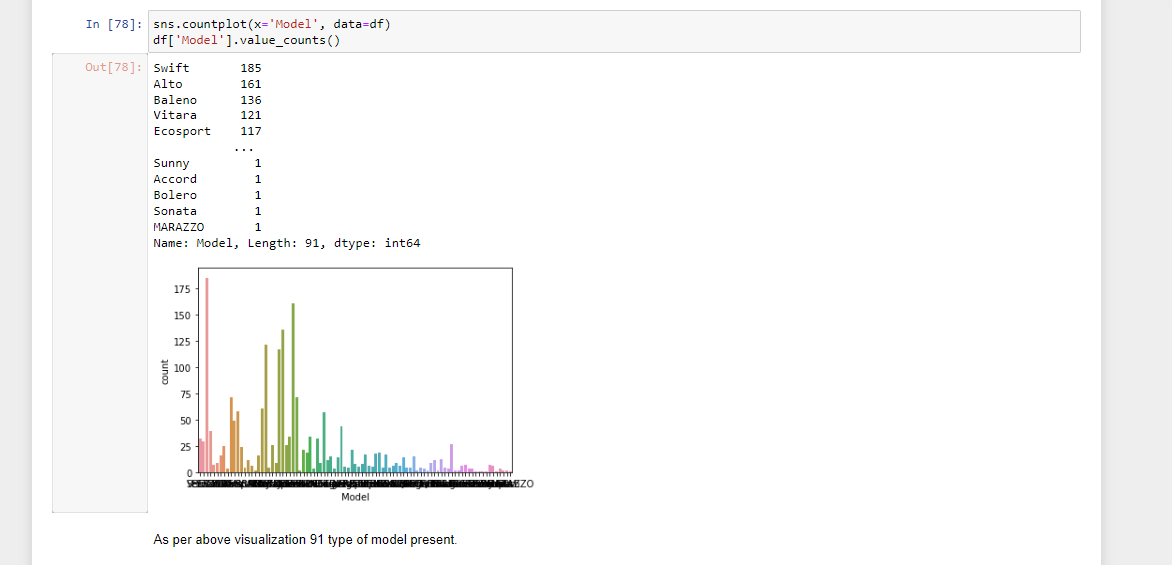
* Visualizations

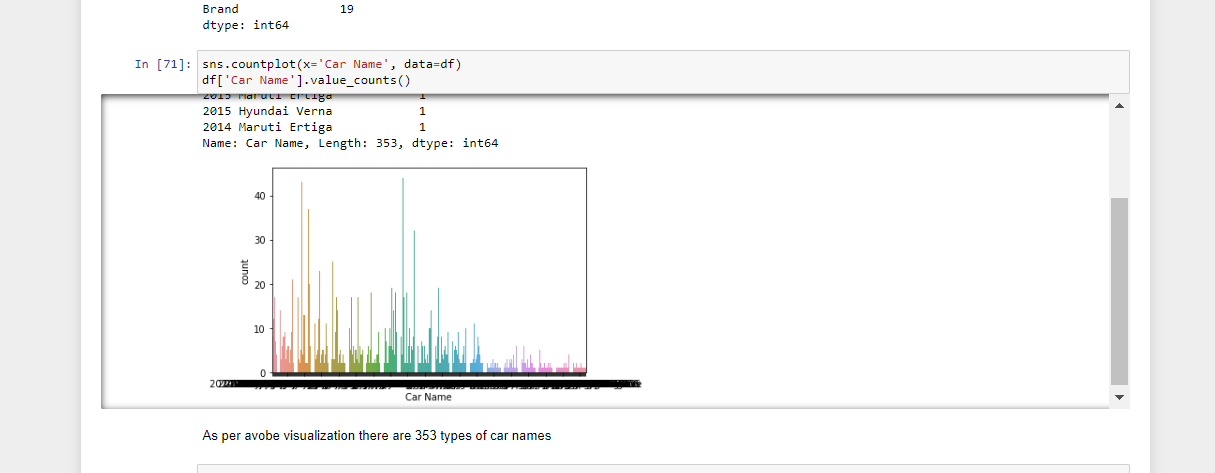
I have visualized the data as per

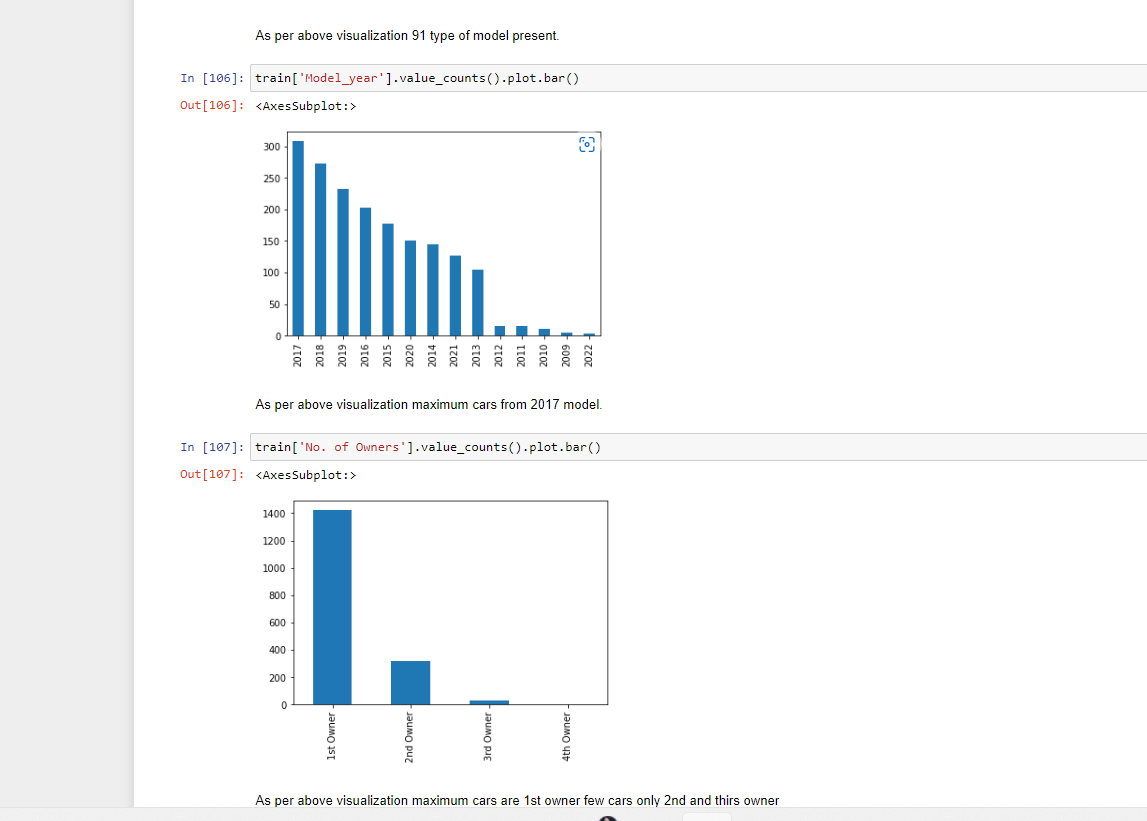
For checking null values

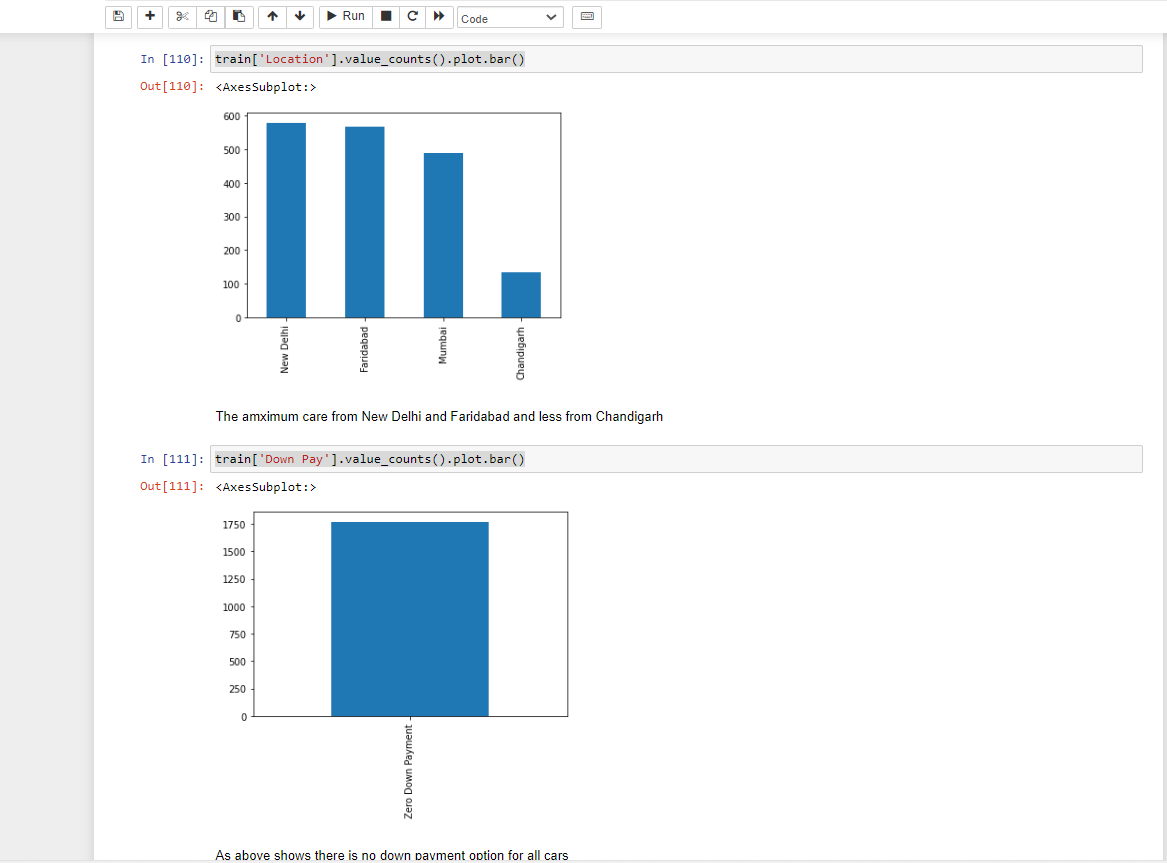


For count plot for object data

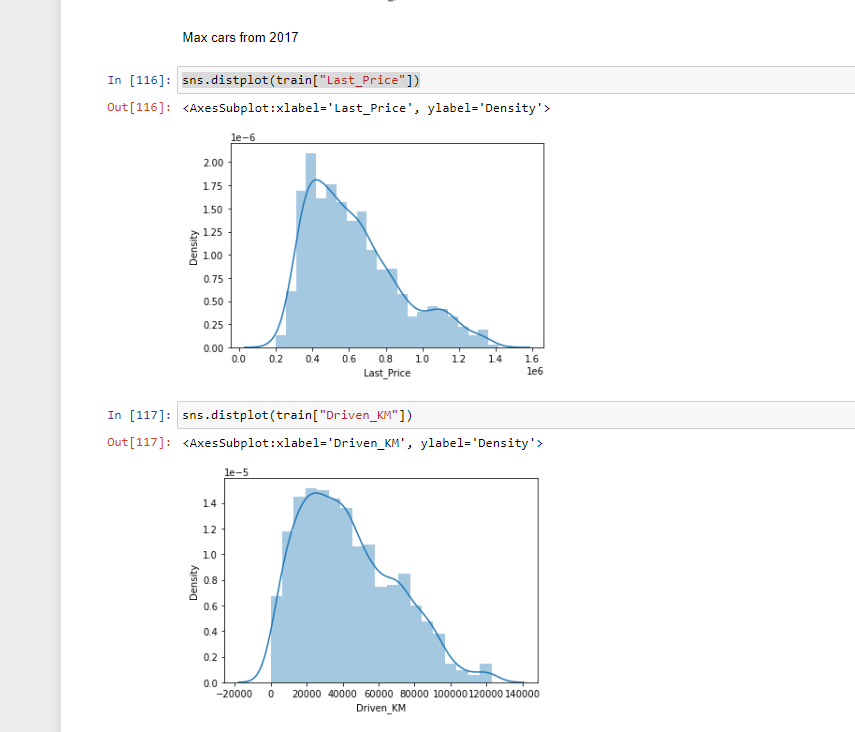


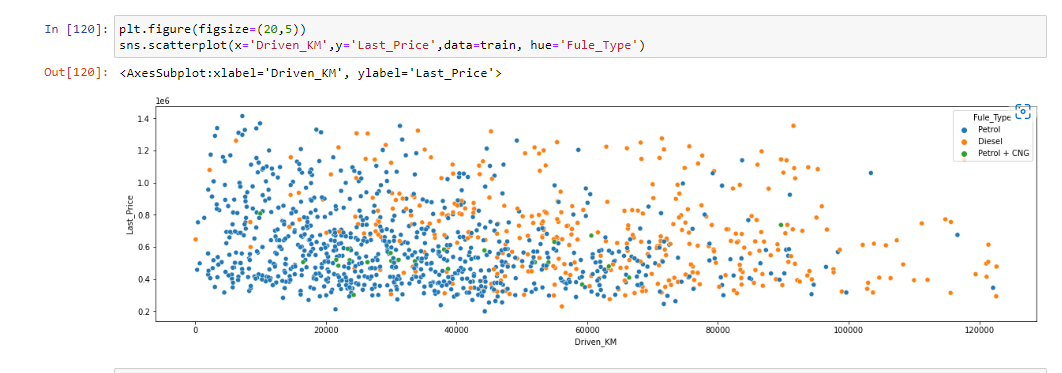




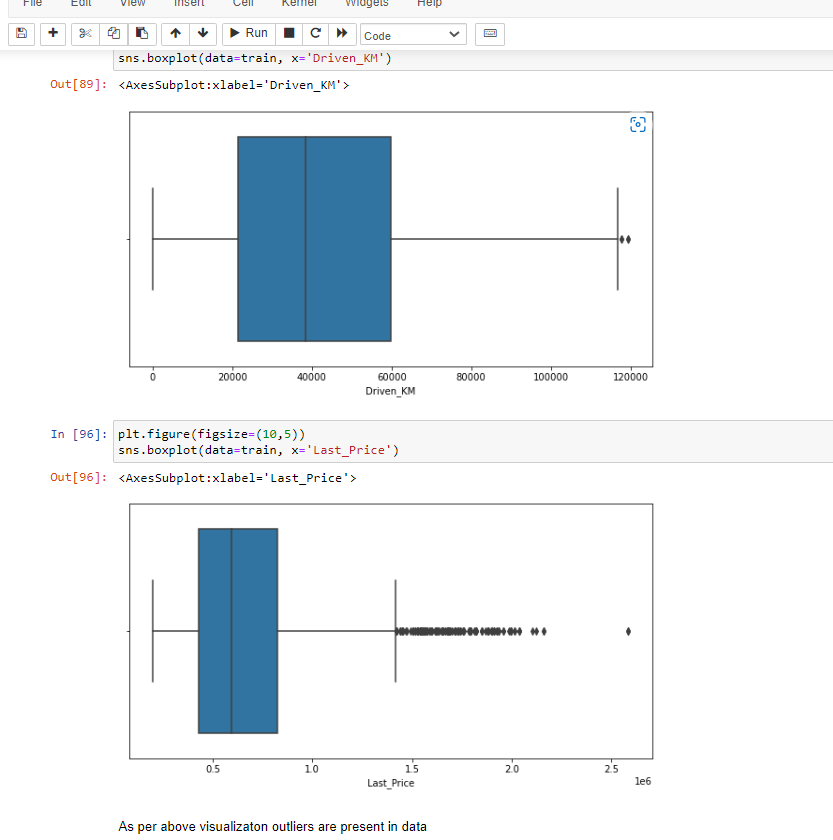


For Continuous Data

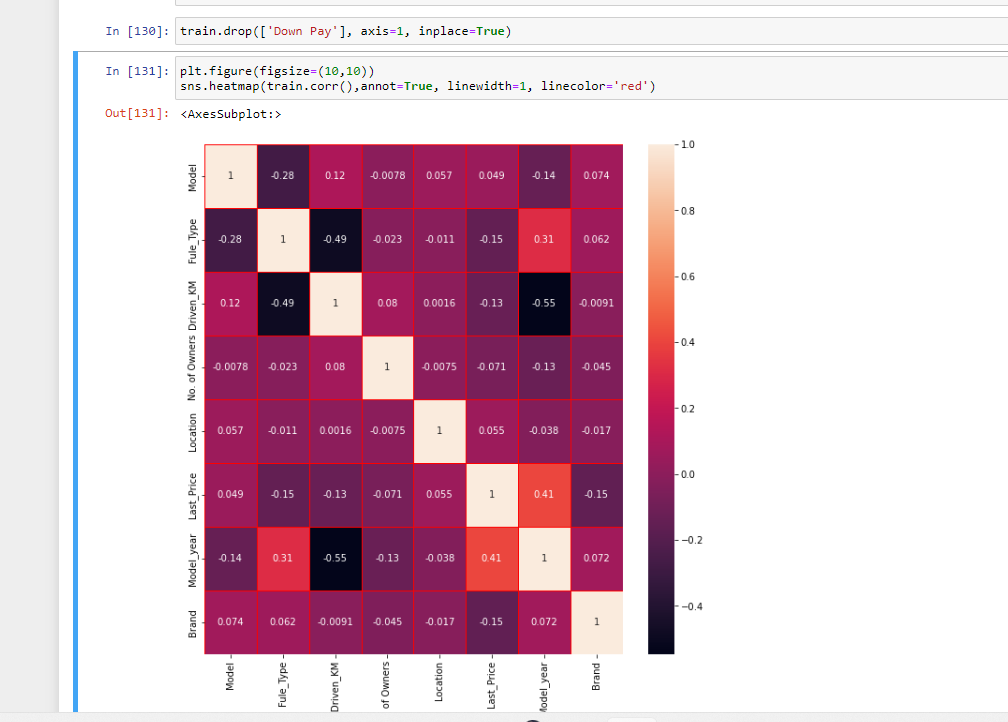




Checking for Outliers

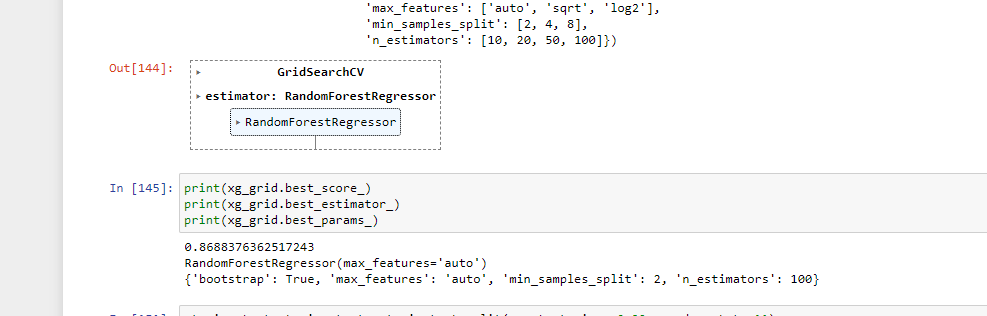


Coorelation matrix



* Interpretation of the Results

I have selected the Random forest is best model now using gris search cv Hyperparameter tunning done.

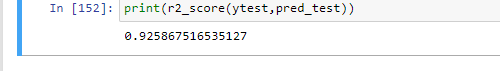


We have found the best parameters to fit under model and gives the best result.

**CONCLUSION**

* Key Findings and Conclusions of the Study

We have found that the model is running good and not biased and it gives the r2\_score below.



* Learning Outcomes of the Study in respect of Data Science

The conclusion is the model is running best.

