

Car Price Prediction Project

Submitted by:

Joshua Paul

ACKNOWLEDGMENT

Thanks for giving me the opportunity to work in FlipRobo Technologies as Intern and I would like to express my gratitude to Data Trained Institute as well for training me in Data Science Domain. This helps me to do my projects well and understand the concepts.

Dataset – FlipRobo Tech

Resources used – Google, GitHub, Blogs for conceptual referring.

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.

* Conceptual Background of the Domain Problem –

With the change in market due to covid 19 impact, our client is facing problems with their previous car price valuation machine learning models.

Some cars are in demand and making them costly and some are not in demand, and it will be cheaper. This will help our client to do better trade and help in his/her business to grow.

* Motivation for the Problem Undertaken

Car is one of the most needed in everyone lives, and all people cannot afford to buy a new one and people who want to buy can exchange their old car in a good rate.

Our Prediction will help the client to sell the car in a smart way.

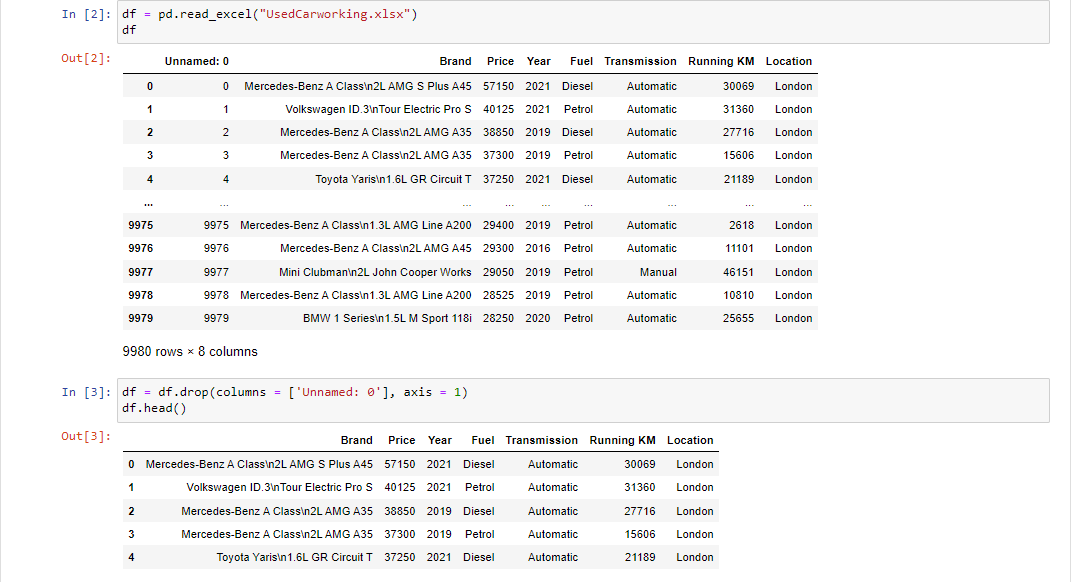
Analytical Problem Framing

* Mathematical/ Analytical Modeling of the Problem

Here our target variable is price and as the data is having continuous variables, hence this is Regression Problem.

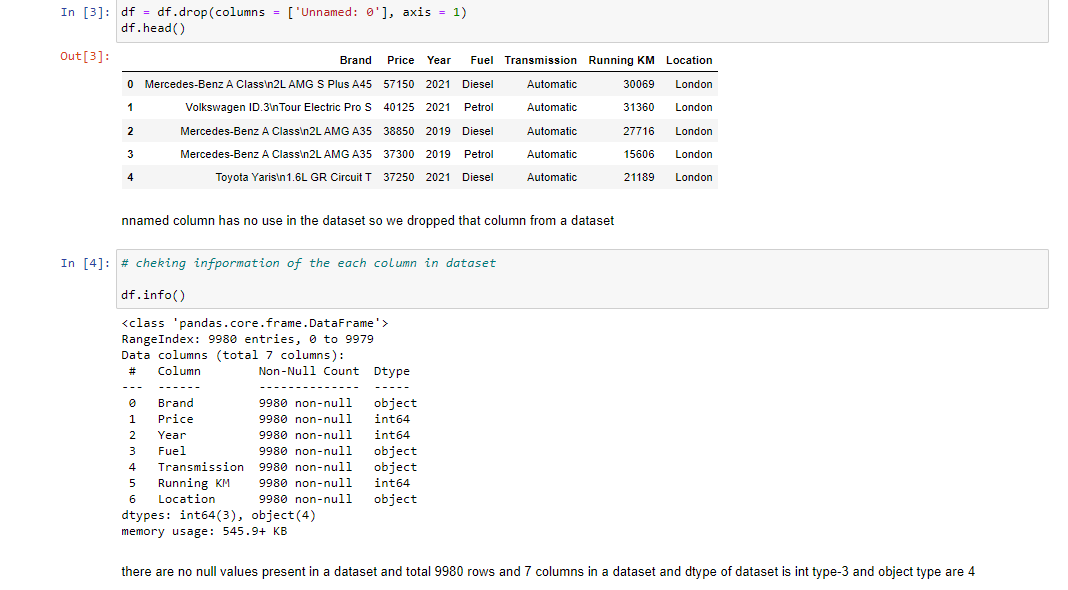
* Data Sources and their formats

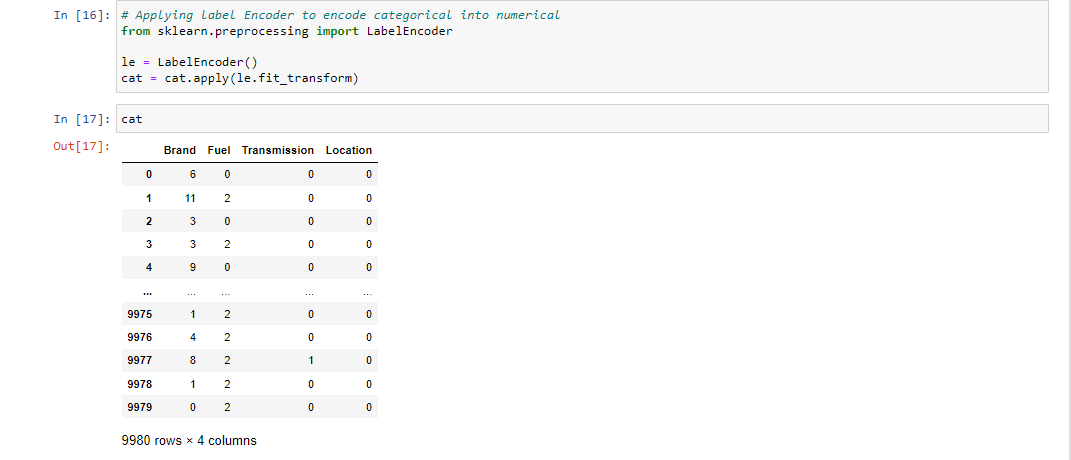
The data is collected from One of the famous websites for used cars and price are in Euros and it has 8 columns and 9980 rows.



Data is not having any null values and we are good to pre-process the data further.

* Data Preprocessing Done





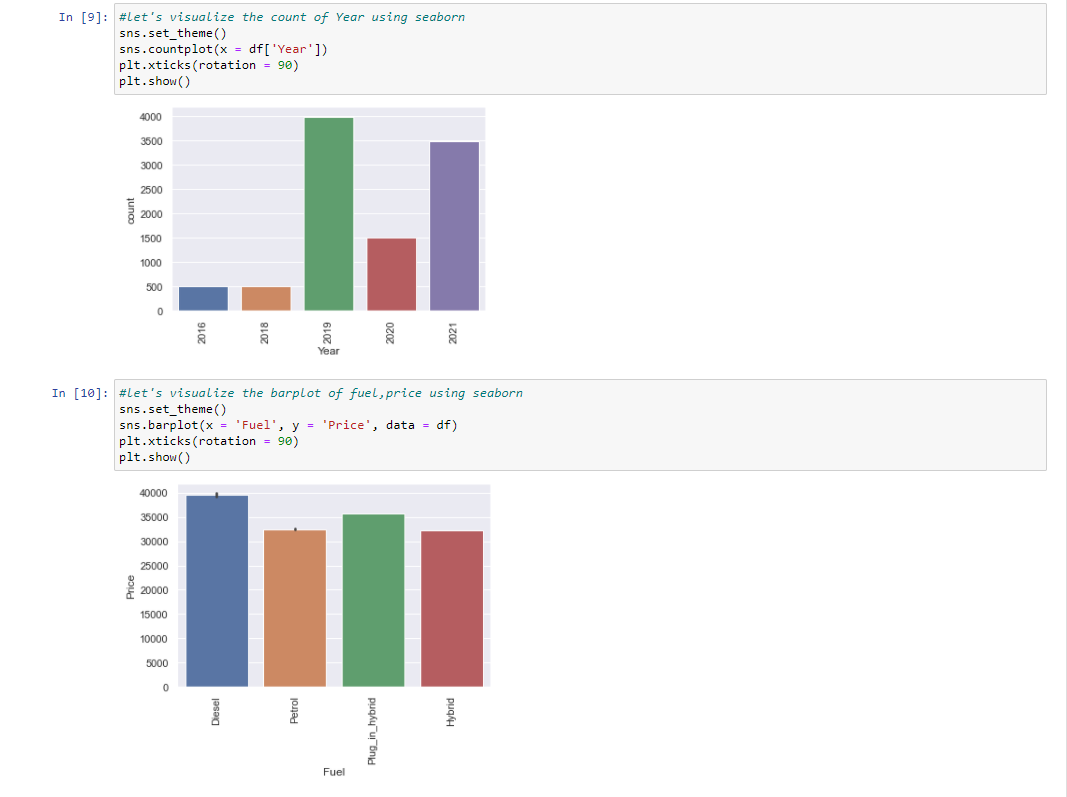
* Data Inputs- Logic- Output Relationships

Almost all car brands are in demand but among all popular brands, Mercedes and Volkswagen are having high price than other brand cars.



We can see from the below plot that most of the car registered year is from 2019, 2020 and 2021.

Also, most of the cars are having fuel type is Diesel, Electric and petrol.



Hardware and Software Requirements and Tools Used -Model training was done on Jupiter notebook. Kernel Version is python 3.

Libraries- Scikit Learn, Pandas, Numpy.

Model Pre-process – Standard Scaler for normalize the ranges from 0-1.

Label Encoder to encode the categorical values and convert into Numerical values.

Metrics - MSE, RMSE, R2 Score

Model Selection – Train\_Test\_split for spilitting the data into train and test dataset. CV score to check the model is over fit or under fit. Gridsearch CV for hyper parameter tuning the model.

Model/s Development and Evaluation

* Testing of Identified Approaches (Algorithms)

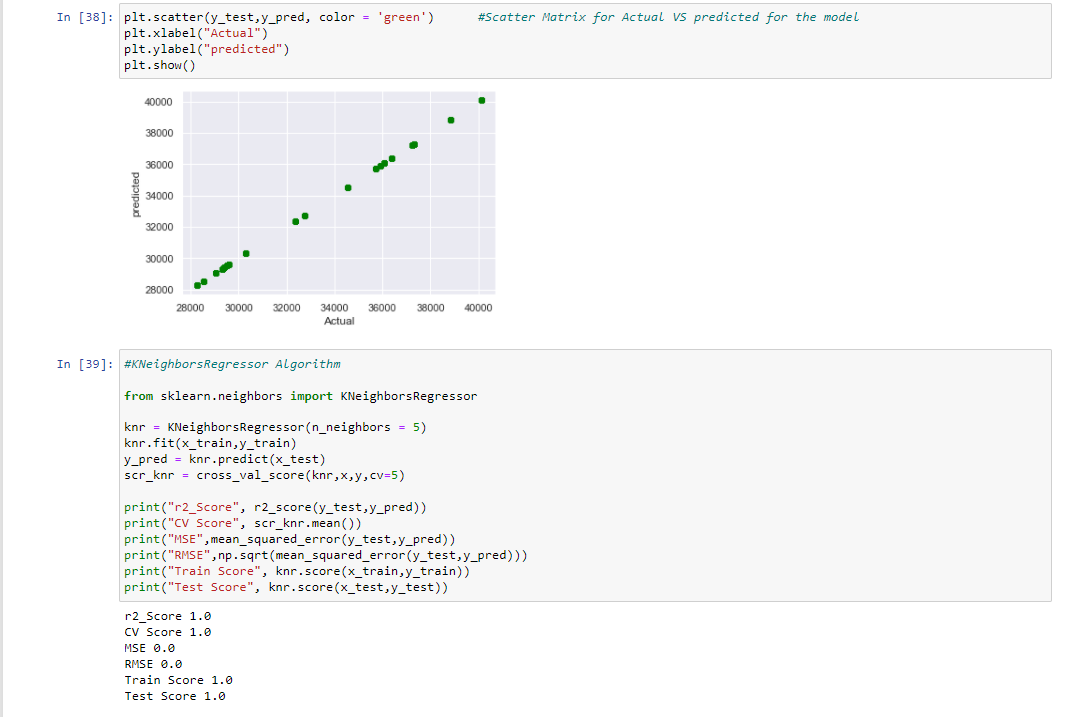
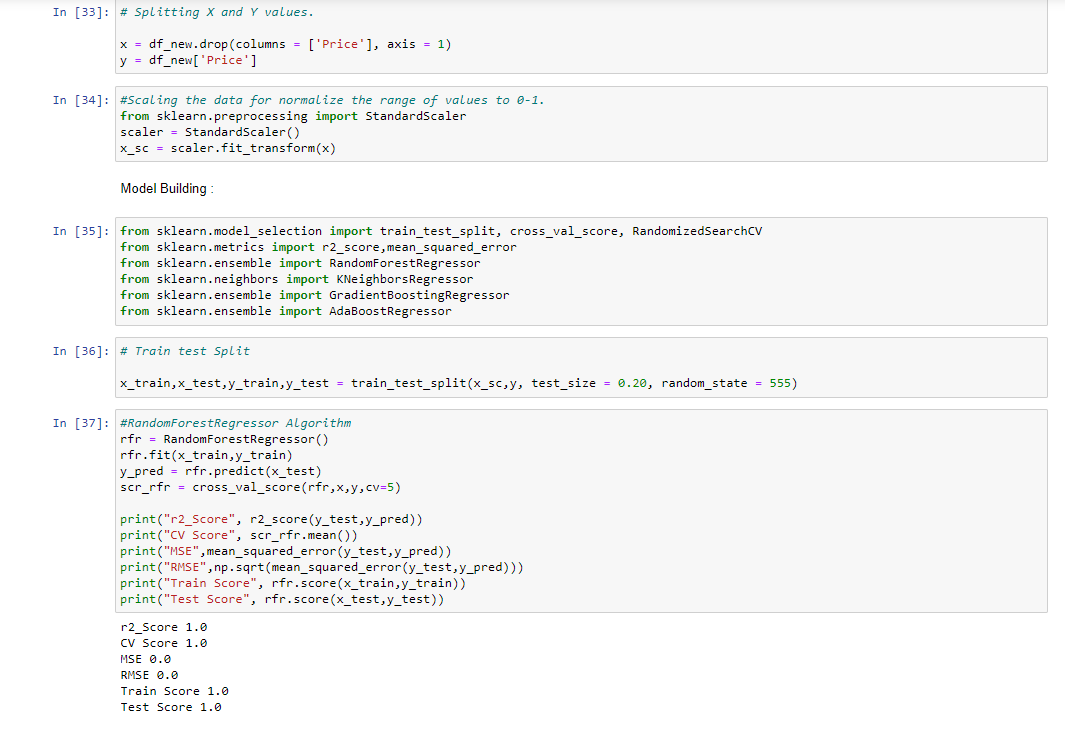
-Random Forest Regressor

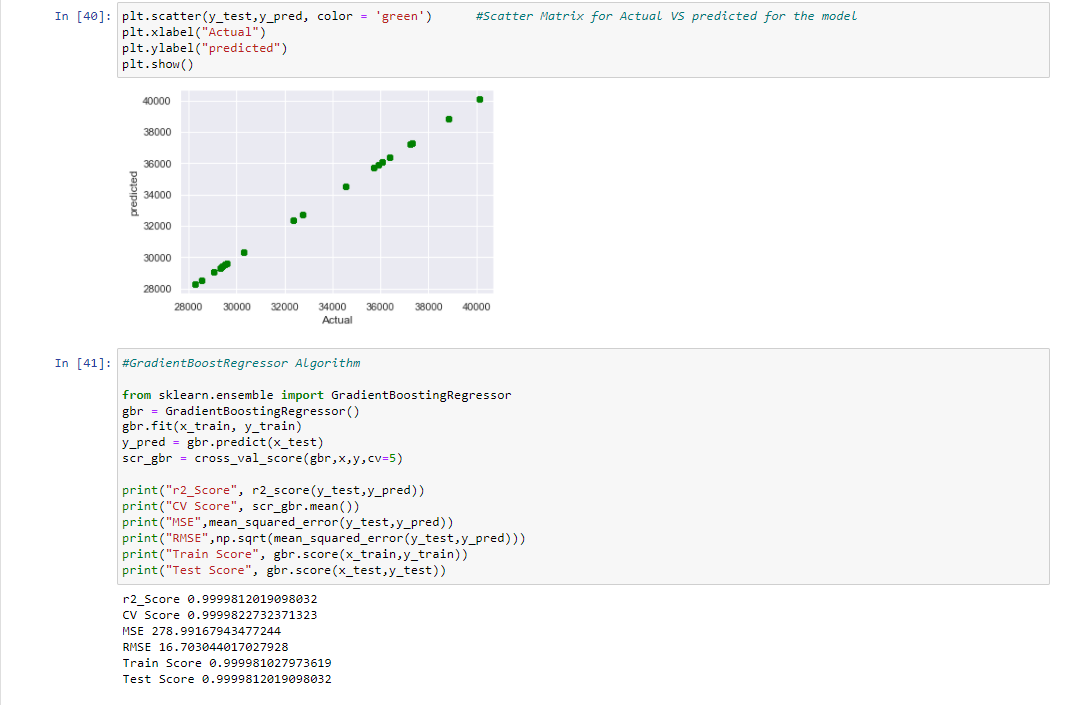
-K Neighbors Regressor

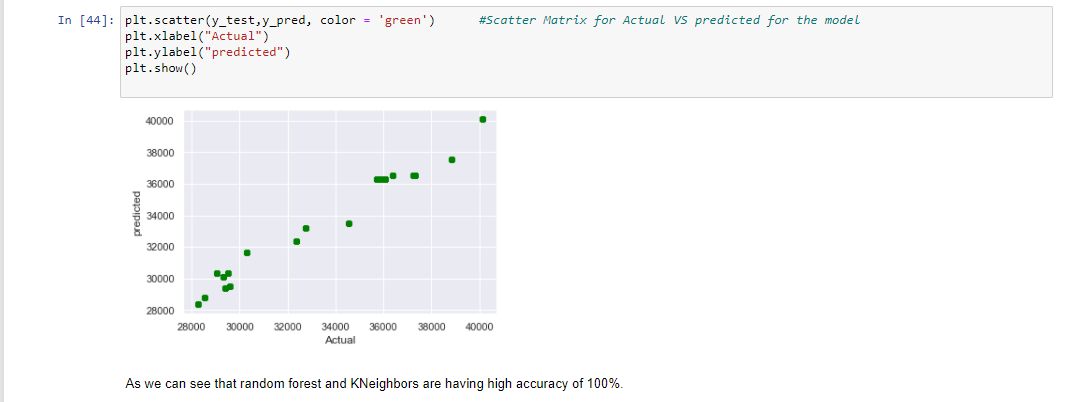
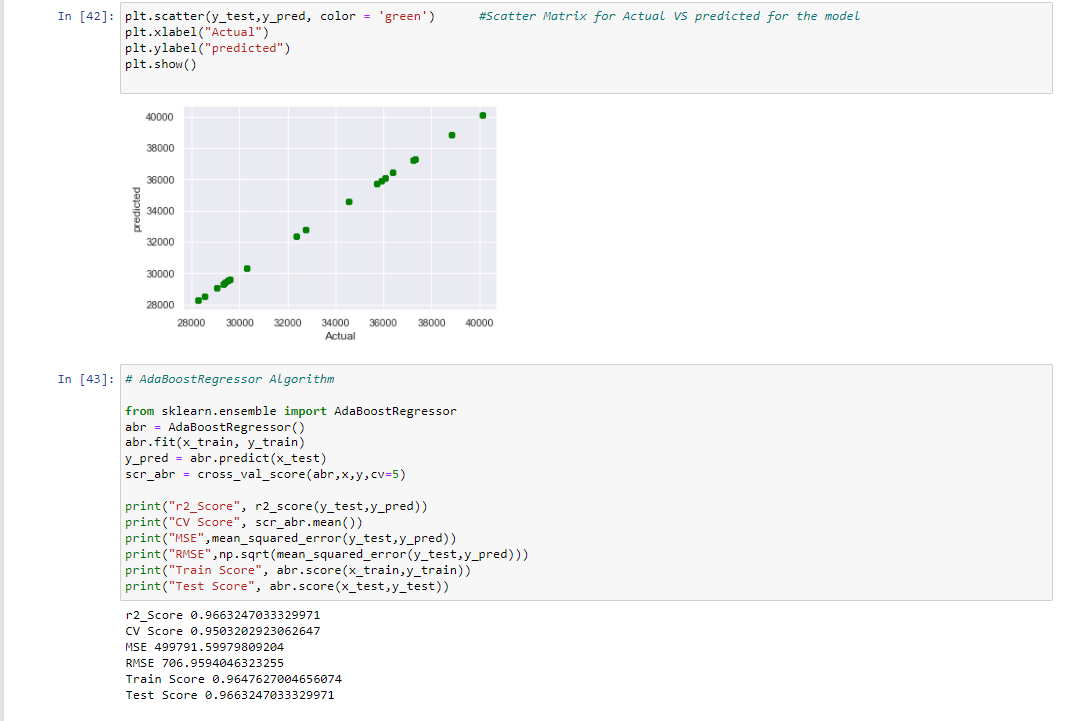
-Gradient Boosting Regressor

-Ada Boost Regressor

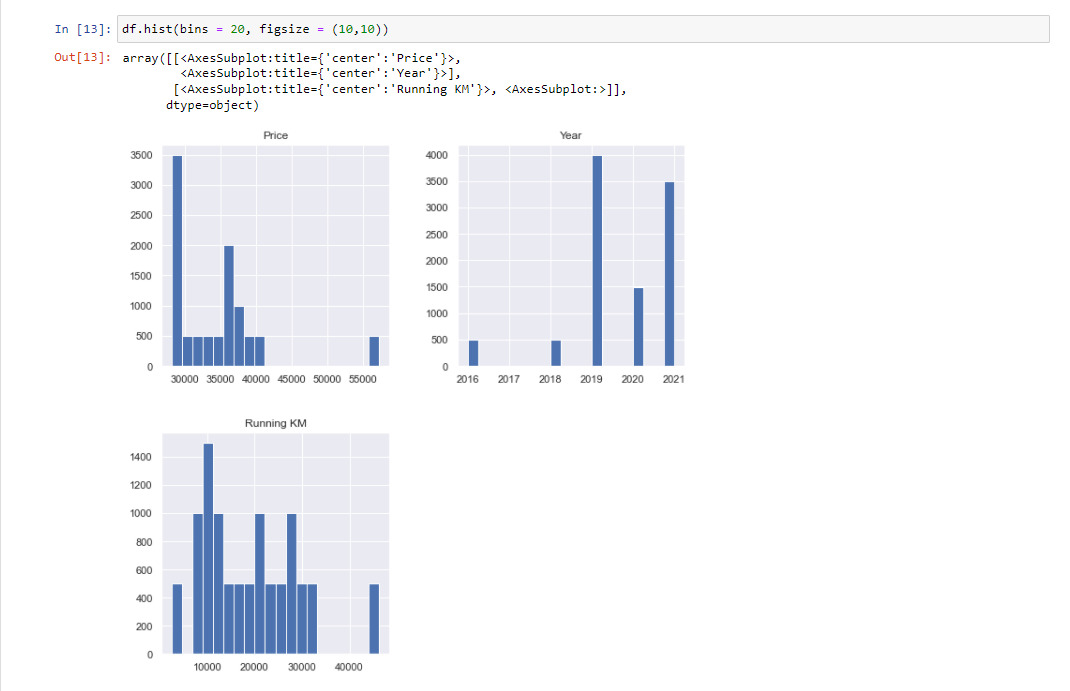
* Run and Evaluate selected models

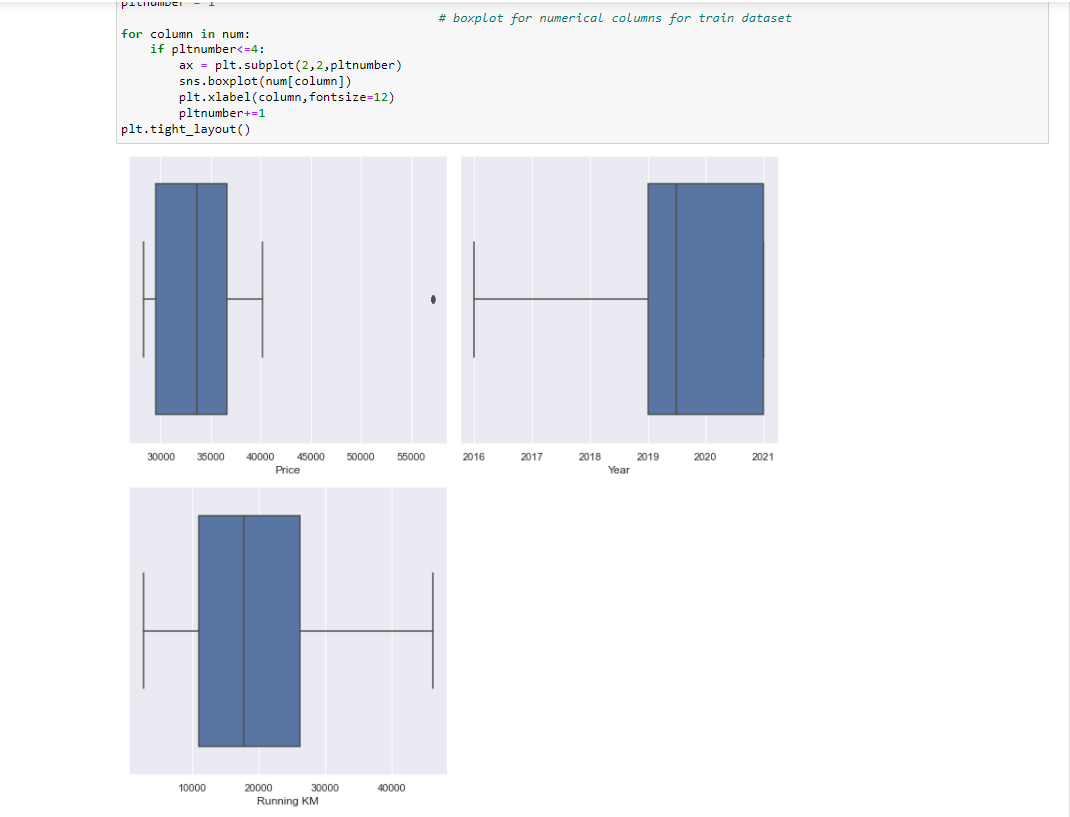




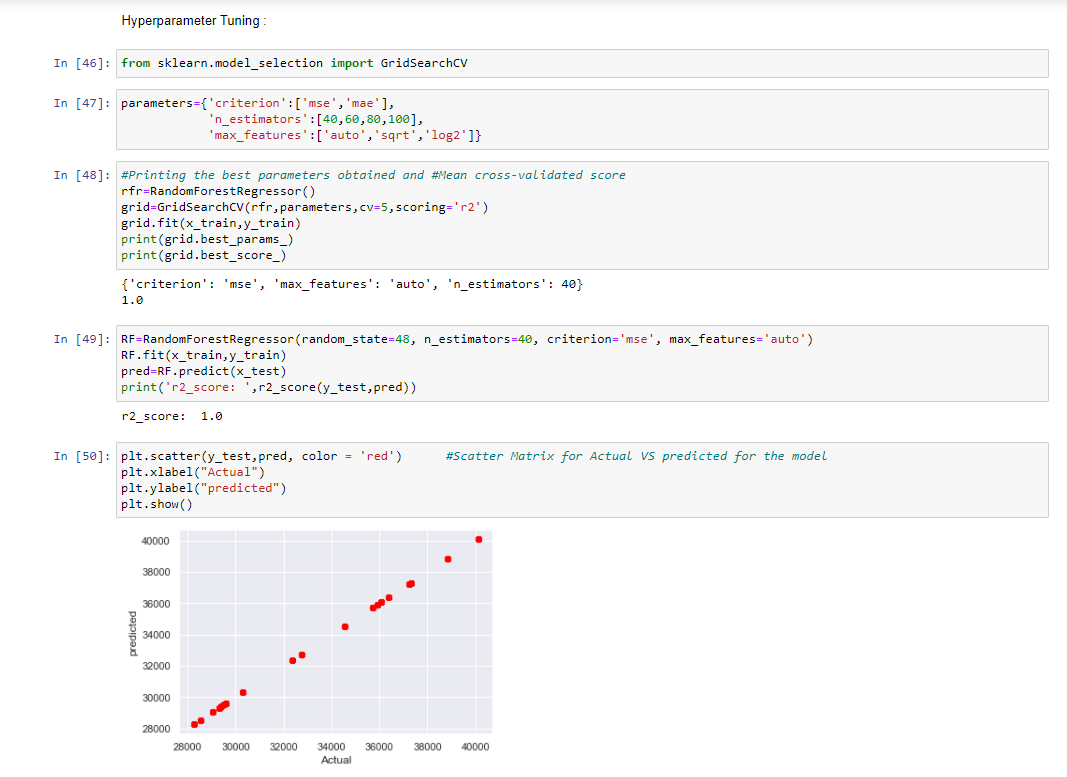


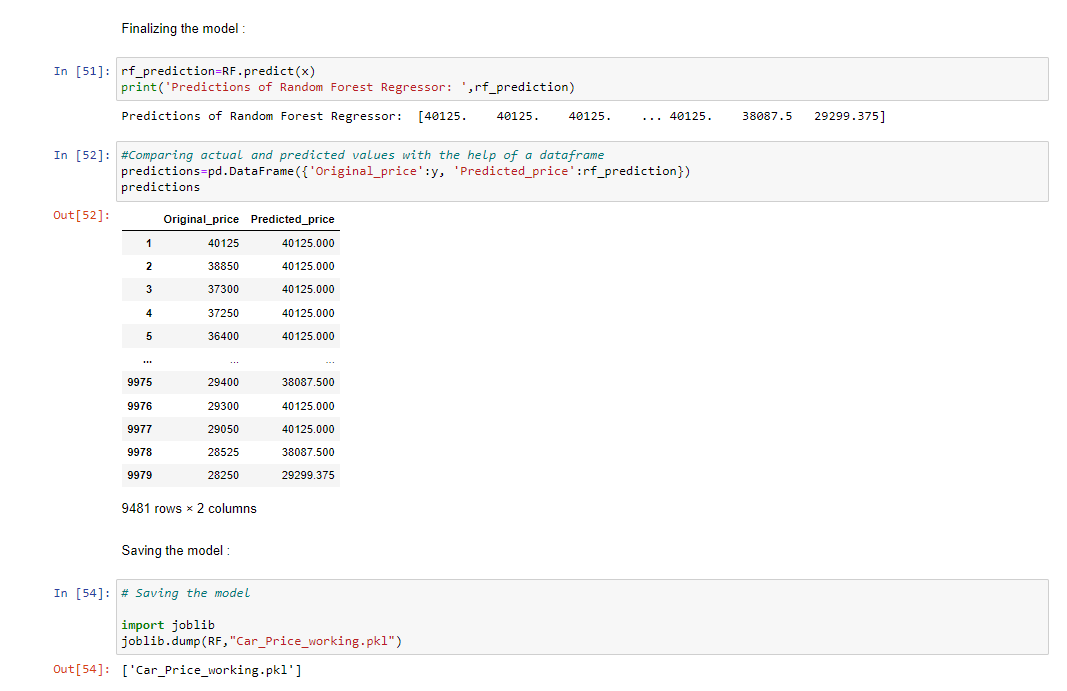
* Visualizations





* Interpretation of the Results





CONCLUSION

* Key Findings and Conclusions of the Study

As this project is about predicting the prices of used cars, it is a regression problem as the target variables are continuous range.

Used r2 score, MSE as a metrics to calculate the model accuracy.

Data is collected by me from theaa.com for used cars. The dataset doesn’t have any null or missing values.

* Learning Outcomes of the Study in respect of Data

Science

Random forest and K Neighbors Algorithm worked as a best model, and which have 100% accuracy and I have used Grid Search CV for hyper parameter tuning.