PROJECT NAME :-

“CAR PRICE PREDICTION USING MACHINE MODEL”

SUBMITTED BY:-

VARSHA VASANT SHINDE

INDEX

INTRODUCTION

- I will be trying to predict the price of used cars based on their features.
- As it would help the people to decide whether the used car is worth the pasted price by different online used-car sites.
- It would also help people when they plan selling their cars.

DATA COLLECTION

- For accurate and real time analysis, data is prepared from scratch.
- Data is scrapped from ‘Cars24 website’.

PROBLEM STATEMENT

Predicting the price of used cars given the features.

Three step approach to understand the problem and the approaches used:

- ✓ Step 1: What is the problem/Task?

What is the price of car given its features.

- ✓ Step 2: Why does the problem need to be solved?

It would help people to determined the best price by comparing the prices against different online used-car sites.

- ✓ Step 3: How would I solve problem?

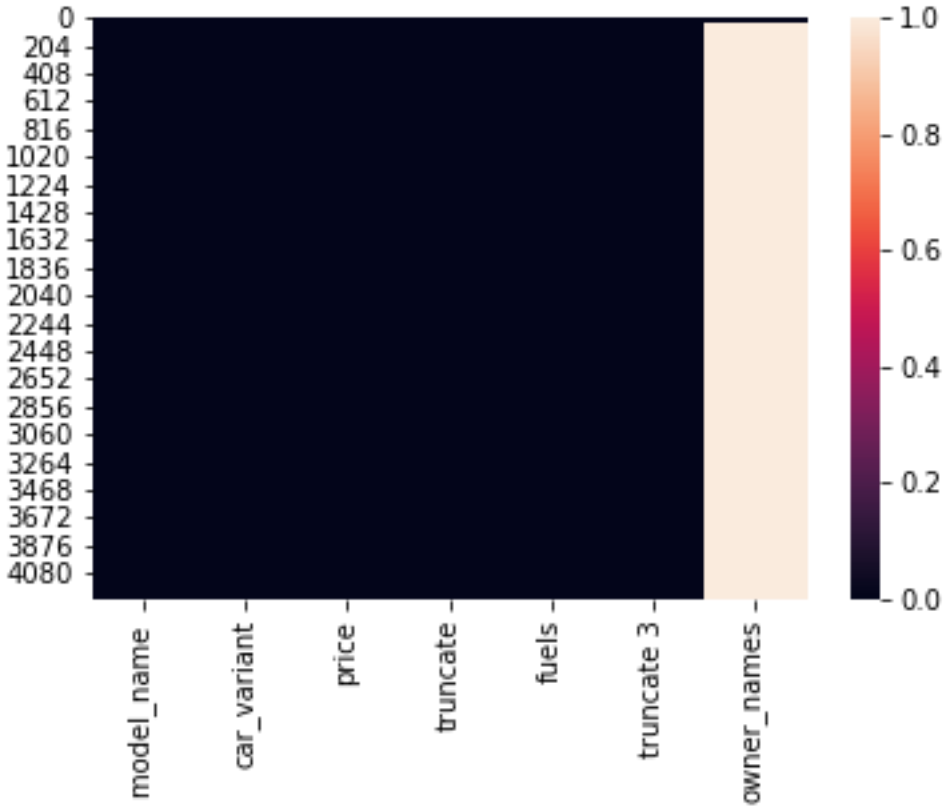
The problem is solved using machine learning techniques/models.

RANDOM FOREST REGRESSION RESULTS

- ✓ Models are run using python in jupyter notebook
- ✓ Dropped the irrelevant columns and null values in ....
- ✓ There are many categorical columns in the data. So I have done the one hot encoding to all the object data type columns.
- ✓ The test set is 30% of overall dataset.
- ✓ After running the base model, the parameters are tuned using random search method and best parameters are selected.

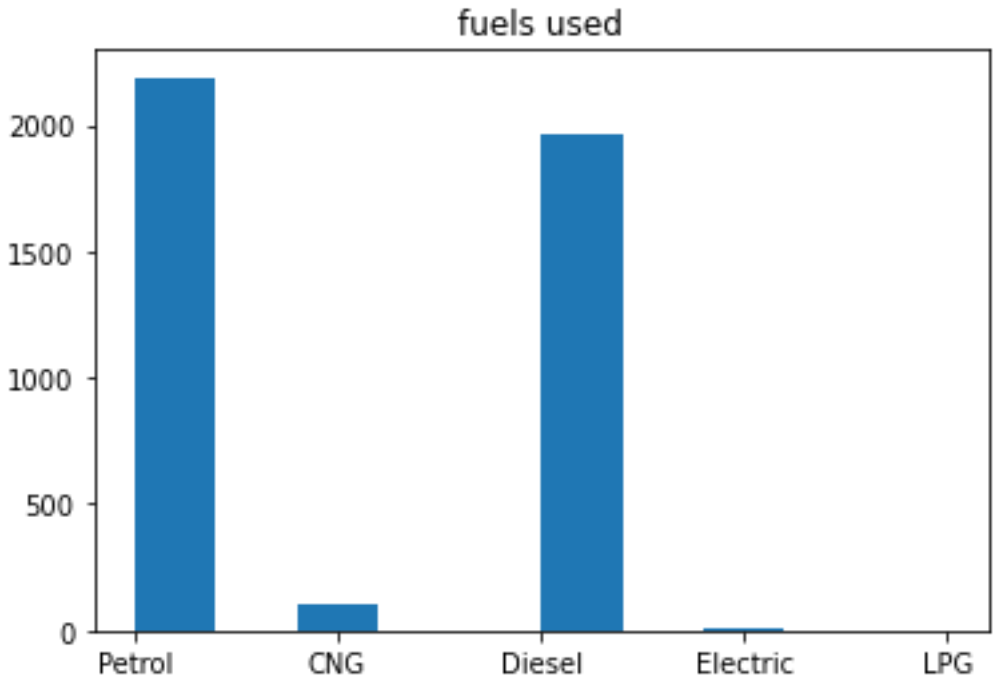
EXPLORATORY DATA ANALYSIS

Summary Statistics



EXPLORATORY DATA ANALYSIS

Bar graph of fuel used data



SPITTING DATA INTO TEST AND TRAIN

- The dataset was split into training data and testing data.
- The training data is the data that the model would leverage for learning.
- The test data is the data that would be leveraged to measure the performance of our models on unseen data.

TRAINING MODELS

There are two types of supervised machine learning algorithms including Regression and Classification.

- Predicting the price of a used cars is a regression problem.
- Different types of regression models can be implemented using python Scikit-Learn.
- The Scikit-Learn is a python machine learning library.
- The training dataset would be trained or fitted using 4 different regression models including Linear Regression, Decision Tree Regression, Random Forest Regression and Ensemble (Voting Regressor).

ACCURACY PREDICTIONS

- Get the predictions by providing the test data to the Linear Regression and Ensemble Voting Regressor models.
- This would give us the prediction accuracy score.
- The prediction accuracy score from each model would be used as one of the basis to determine the best model.

CONCLUSION

- Random Forest Regression Model provides the best prediction accuracy out of the four models.
- Random Forest also provide the best RMSE metrics for predicting the response(y).
- Hence, Random Forest Regression Model is the selected model for predicting the price of used car.