

A  
CAPSTONE PROJECT  
IN  
DATA SCIENCE

---

CAR PRICE PREDICTION

---

*By:*  
GANESH BHANDARI

@



JANUARY 2021

## 1. Problem, background and data:

A manager of a car/motor vehicle dealer company having dealership showrooms in New York and Washington, DC wants to know actual price of the cars based on the features of the cars provided by the car manufacturer company. The manager believe that if he know the price based on the features, he can make great deals with the manufacturers and can make more profit for his company.

I would like to help the manager to predict price of the cars. The manager gave me a data set of cars sold from his both showrooms in NY and DC. The data set contains 205 observations with 26 features.

## 2. Methods used:

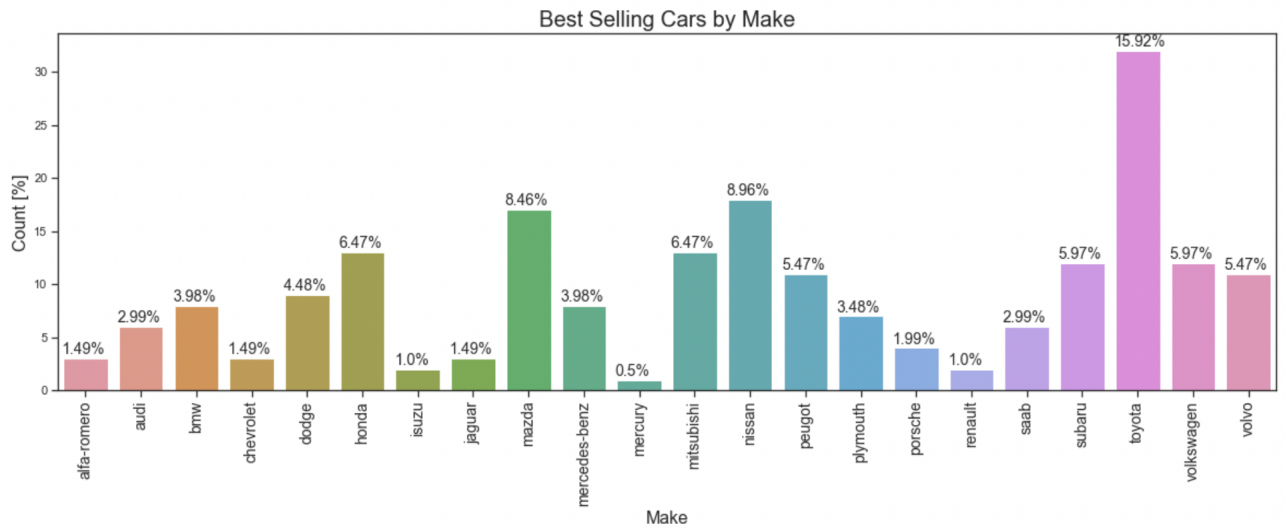
This problem is a regression problem. So to solve this problem I have created a model based on regression algorithms . In this project, I used regression algorithms like Multiple Linear Regression, Ridge Regression, Polynomial Regression, Random Forest Regressor, Extra Trees Regressor, XGBRegressor and chose the best model based on their accuracy performance. I used the best selected model to predict the house prices. The steps that I used to get the result are as follows:

- Loading data
- Cleaning and organizing data
- Performing Exploratory Data Analysis (EDA) to understand the data and important features.
- Creating model and training the model
- Using suitable metric to check the performance of the model
- Deploying the model to get the results

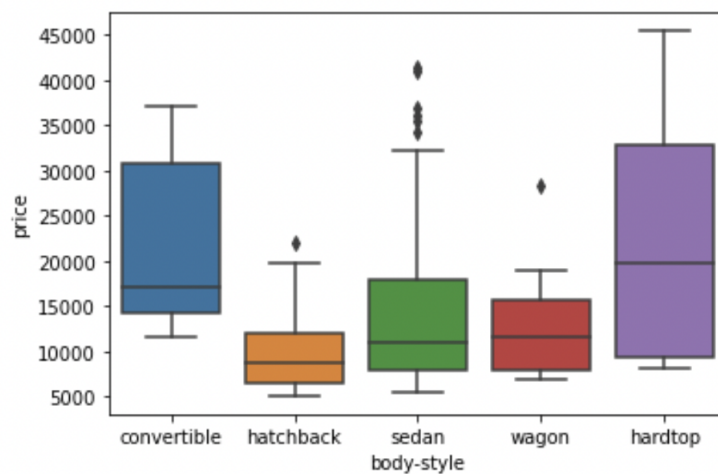
## 3. Current Situation:

The top 5 selling cars by manufacturers (or by brand) are as follows:

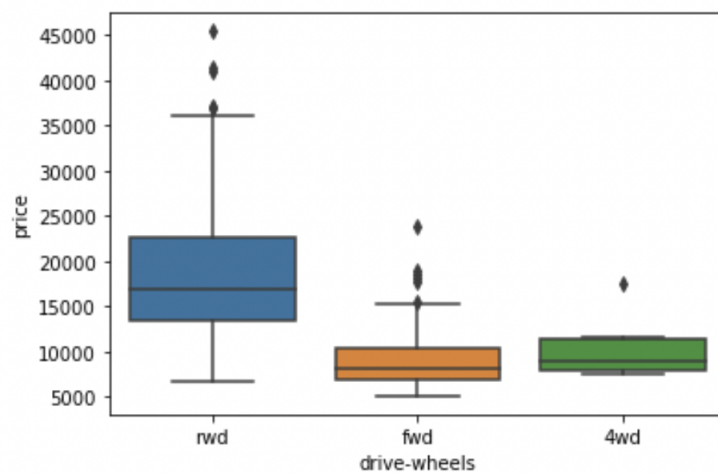
	make
<b>toyota</b>	15.92
<b>nissan</b>	8.96
<b>mazda</b>	8.46
<b>honda</b>	6.47
<b>mitsubishi</b>	6.47



The data shows that Hardtop and Convertible Cars are more expensive in comparison with other body styles



Also the RWD Cars are more expensive.

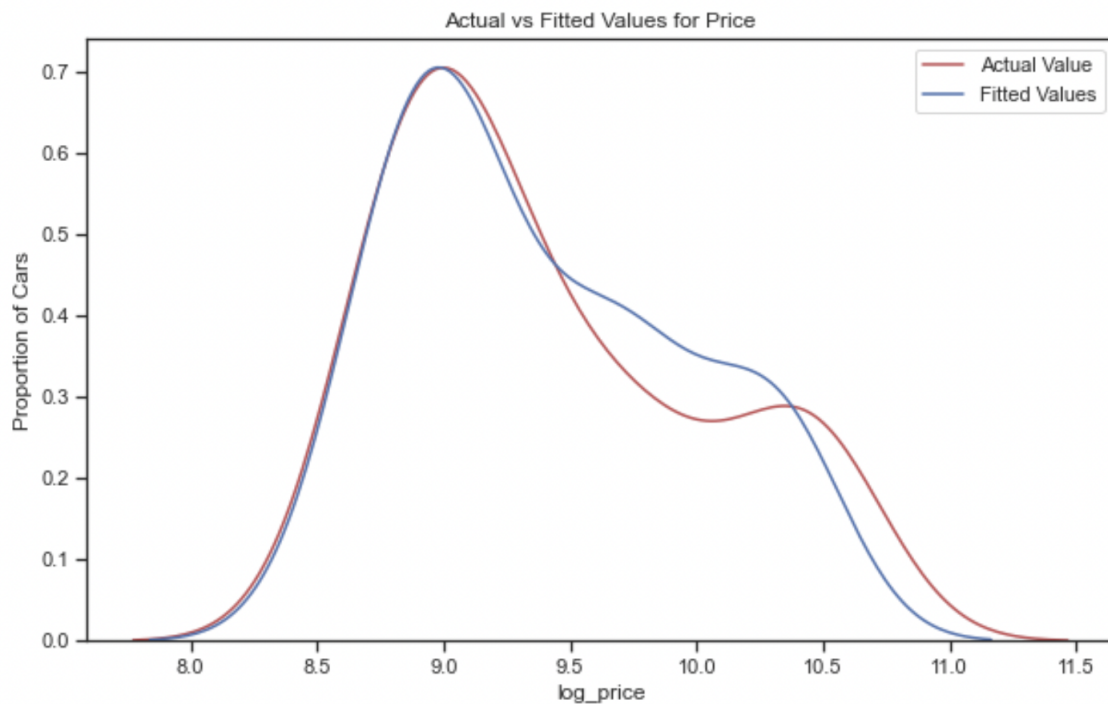


## 4. Model:

In this project, I used regression algorithms like Multiple Linear Regression, Ridge Regression, Polynomial Regression, Random Forest Regressor, Extra Trees Regressor, XGBRegressor to create a suitable model. The performance report of these model on the data is as follows:

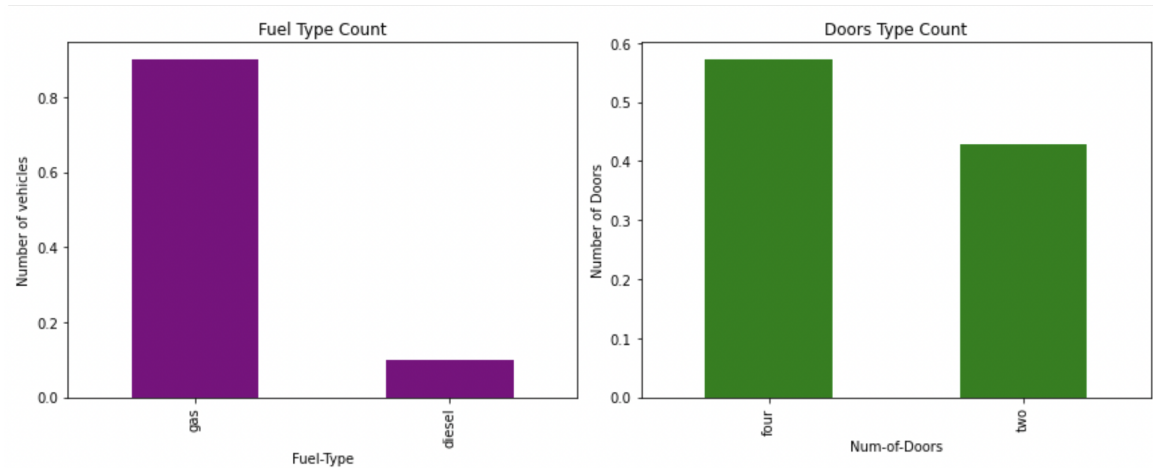
Algorithm	R-Squared	MSE
Multiple Linear Regression	9.000000e-01	0.19
Pipe Line with Linear Regression	-3.801465e+15	37455456.90
Ridge Regression	9.000000e-01	0.19
Polynomial Regression of degree 2	8.100000e-01	0.27
Polynomial Regression of degree 3	-1.160000e+00	0.89
Random Forest Regressor	9.500000e-01	0.14
Extra Tree Regressor	9.500000e-01	0.13
XGB Regressor	9.500000e-01	0.14

The report table shows that, among all these algorithm, the Extra Trees Regressor is more consistent and works better than the others. So, I chose the Extra Trees Regressor model as the final model for my project. The fitting of the model is given below:

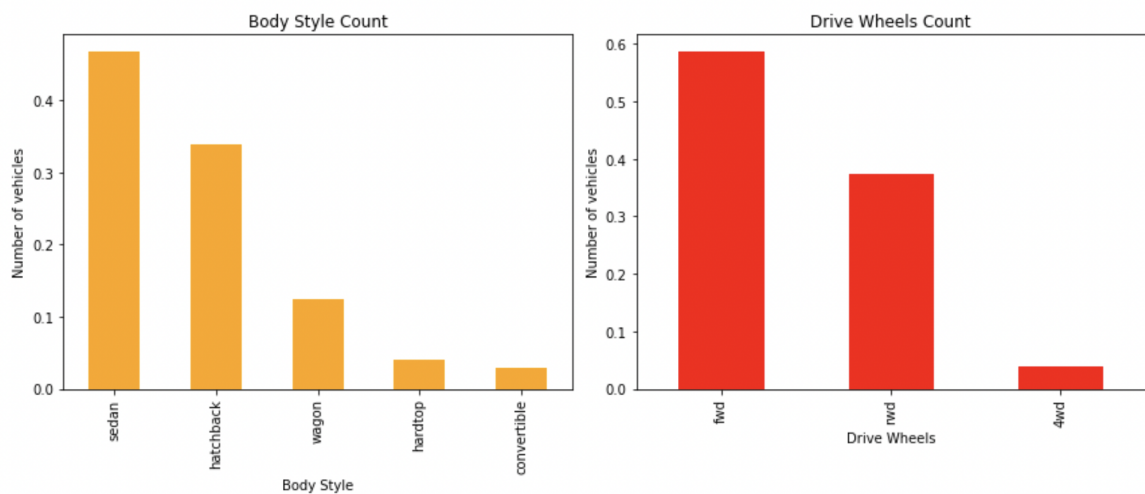


## 5. Findings:

- Most selling car is of brand Toyota. This brand has 16% sales of total sales.
- More than 90% of the vehicle have fuel type gas.
- More than 57% of the cars have four door.



- Mostly produced vehicles have body style sedan around 48% then hatchback 32%.
- 59% of the cars have fwd, 37% have rwd and 4% have 4wd.
- 99% of the cars have engine at front.
- More than 70 % of the vehicle has Ohc type of Engine



## **6. Recommendations:**

Based on the findings obtained during data analysis process and deploy of the model, I have some recommendations for the manager:

- Emphasize on selling cars of brand Toyota.
- Reduce the number of 2 doors cars
- Reduce the number of diesel engine cars in the showroom.
- Focus on selling the cars having engine Ohc.

## **7. Conclusion:**

The Extra Trees Regressor is more consistent and works better than the others. It has more than 95 % of accuracy, which is very good level of accuracy. One can apply this model to predict the price of carss.

## **8. Acknowledgments:**

I am very grateful to my mentor for his valuable suggestions while completing this project. Also I am thankful to the bank authority for providing valuable information.