



CAR PRICE PREDICTION PROJECT

Submitted by:

ANISH ANTONY

ABSTRACT:

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. To predict this model, we need to scrape the used car details from online websites using selenium web driver, and the data have preprocessed, trained and tested using the regression algorithms. Then it is hypertuned and best algorithm with best parameters is obtained and finally the selling price of the car is predicted.

Keywords: Cars ,Selenium ,Data cleaning ,Selling Price ,Regression

The features used in the dataset are:

- ❖ Registration Year
- ❖ Make
- ❖ Model
- ❖ KMs Driven
- ❖ No of Owners
- ❖ Transmission
- ❖ Fuel Type
- ❖ Mileage
- ❖ Location
- ❖ Selling Price

Data Cleaning:

- ❖ Registration Year – (Converted to object)
- ❖ Make – (Extracted from Name)
- ❖ Model – (Extracted from Name)
- ❖ KMs Driven – (Removed the string and converted to integer)
- ❖ No of Owners - (Made as unique variables for all websites)
- ❖ Transmission – If not available in some websites, then give the data as not available
- ❖ Fuel Type - (Made as unique variables for all websites)
- ❖ Registration – If not available in some websites, then give the data as not available
- ❖ Location – (Based on the location give the location)
- ❖ Selling Price – (Removed the string and converted to integer)

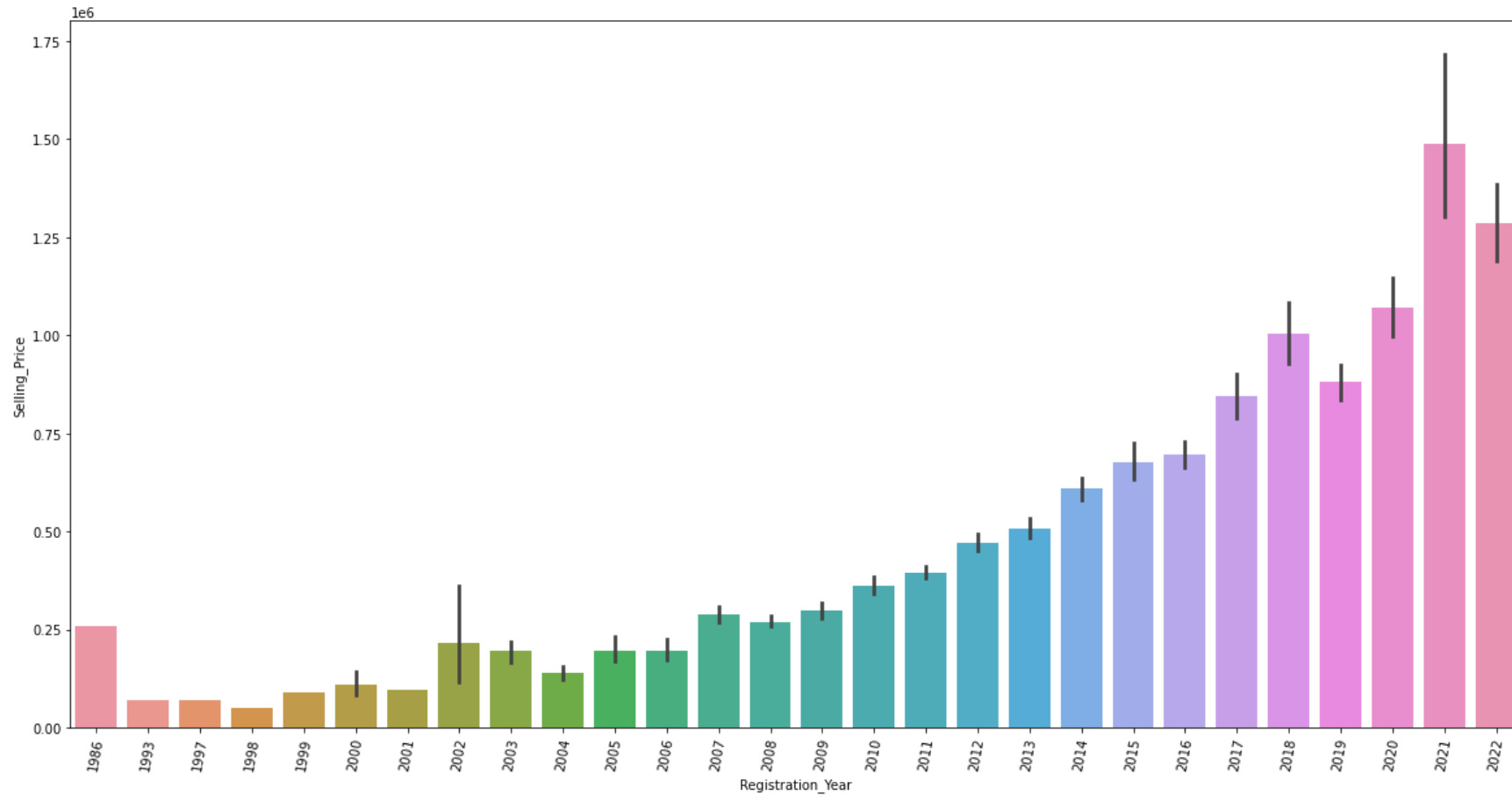
Dataset Description:

S.NO	Columns	Datatype	Unique values	Mode/Mean
1	Registration Year	int64	28	2013.56283
2	Make	object	29	Maruti
3	Model	object	1155	Maruti Suzuki Swift VXi
4	KMs Driven	int64	629	77207.429351
5	No of Owners	object	4	First
6	Transmission	object	3	Manual
7	Fuel Type	object	9	Petrol
8	Location	object	6	Coimbatore
9	Selling Price	int32	805	607882.439353

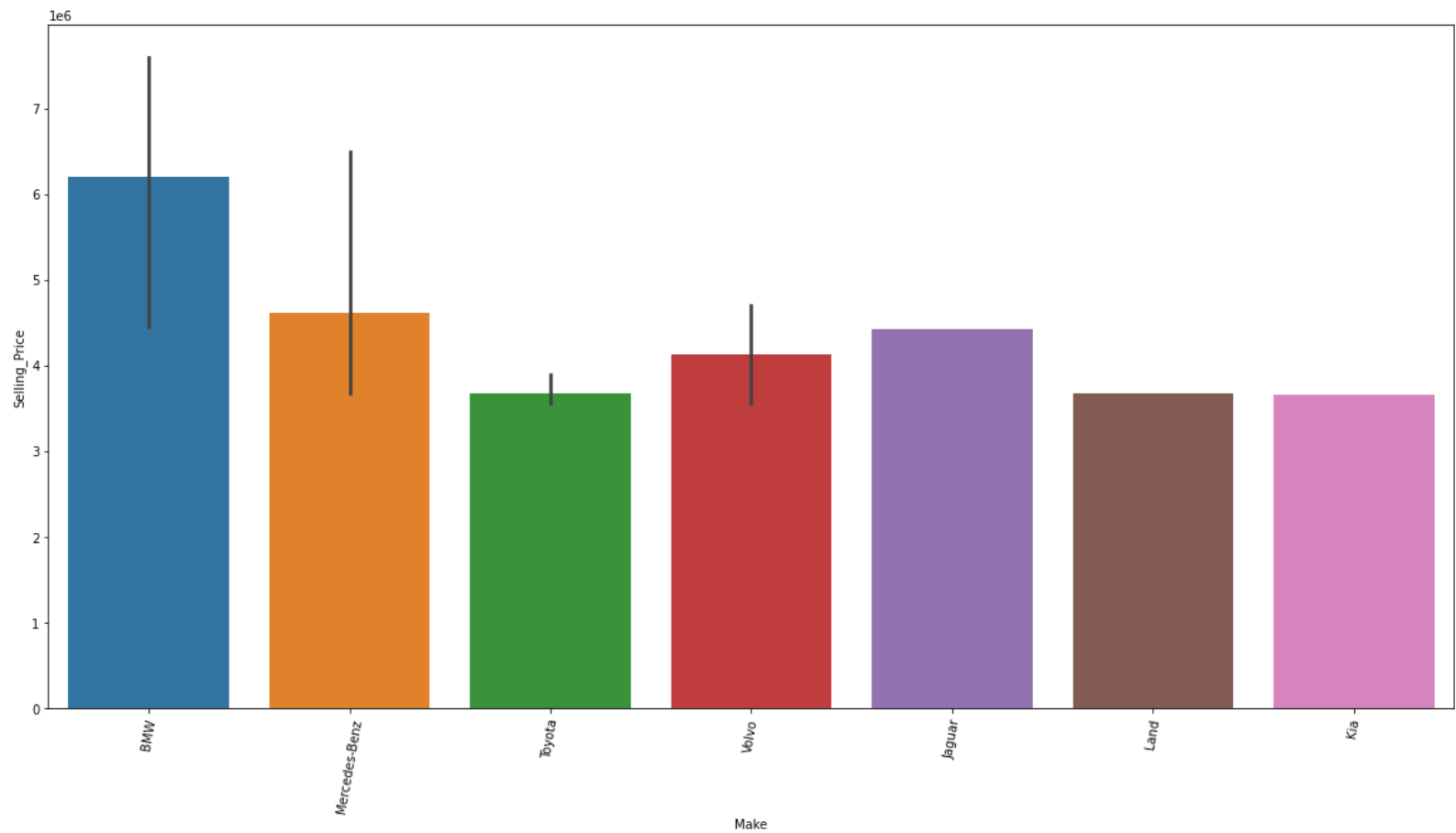
Hardware and Software Requirements and Tools Used:

- Hardware – PC Windows 10, 4 GB Ram
- Software – Google chrome, MS Excel, Python, Selenium webdriver
- Libraries – Pandas, NumPy, Matplotlib, Seaborn, sklearn, SciPy. Stats
 - ☐ Browsing – Google Chrome
 - ☐ Webscraping – Python, Selenium webdriver
 - ☐ Data cleaning – Python, Pandas, NumPy & SciPy. Stats
 - ☐ Data visualization – Matplotlib & Seaborn
 - ☐ Machine learning – Sklearn

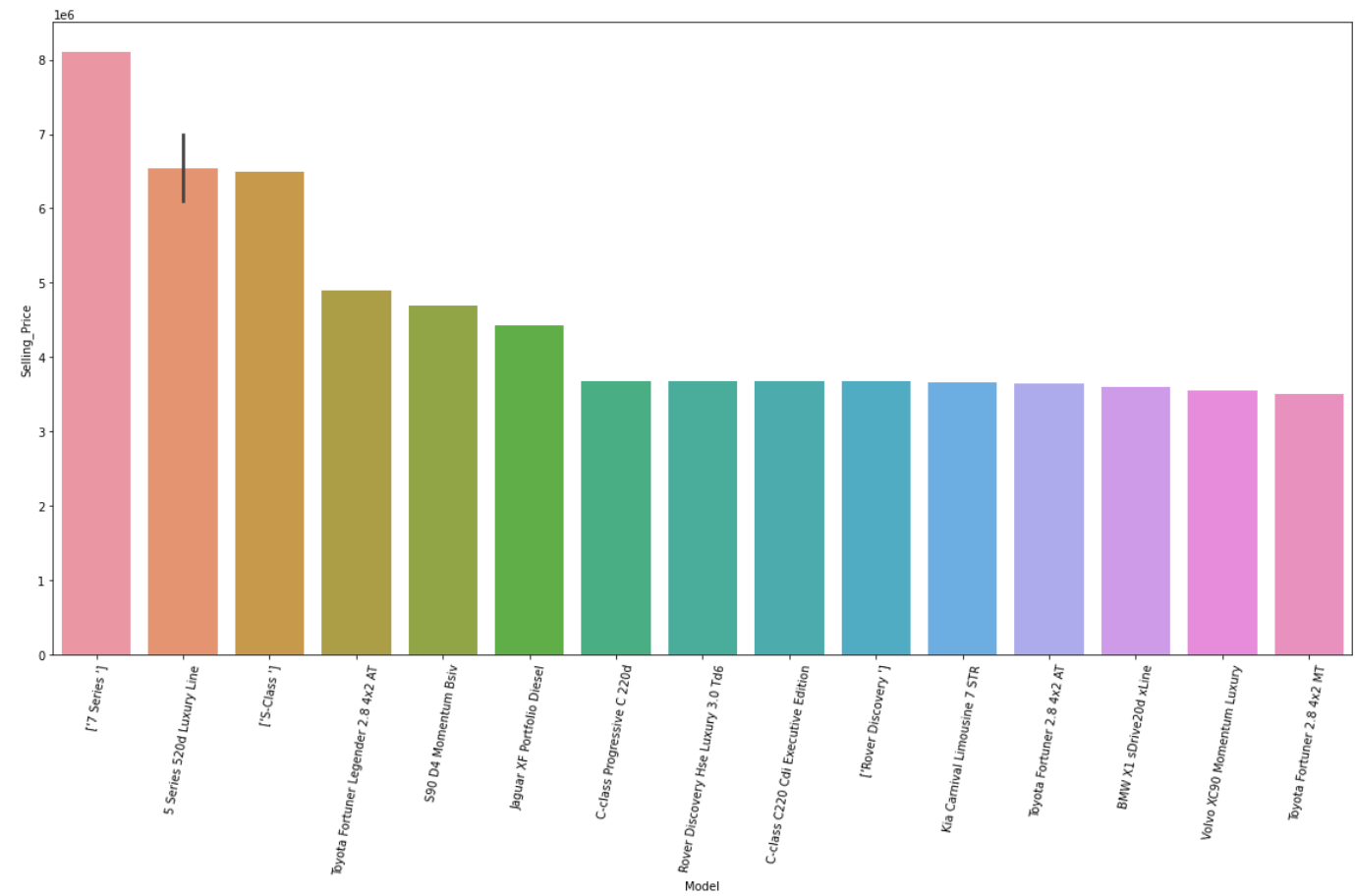
Registration Year:



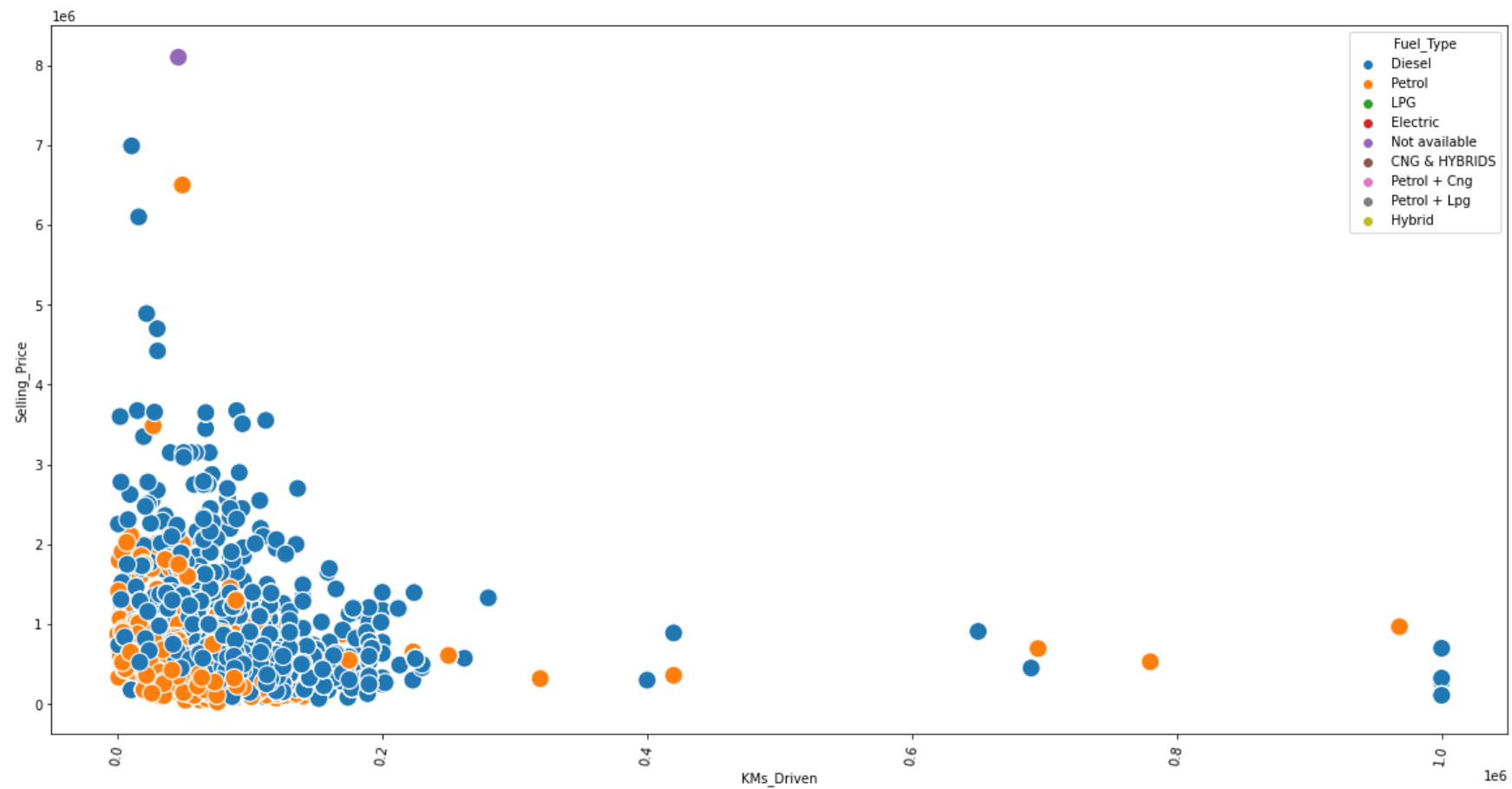
Make:



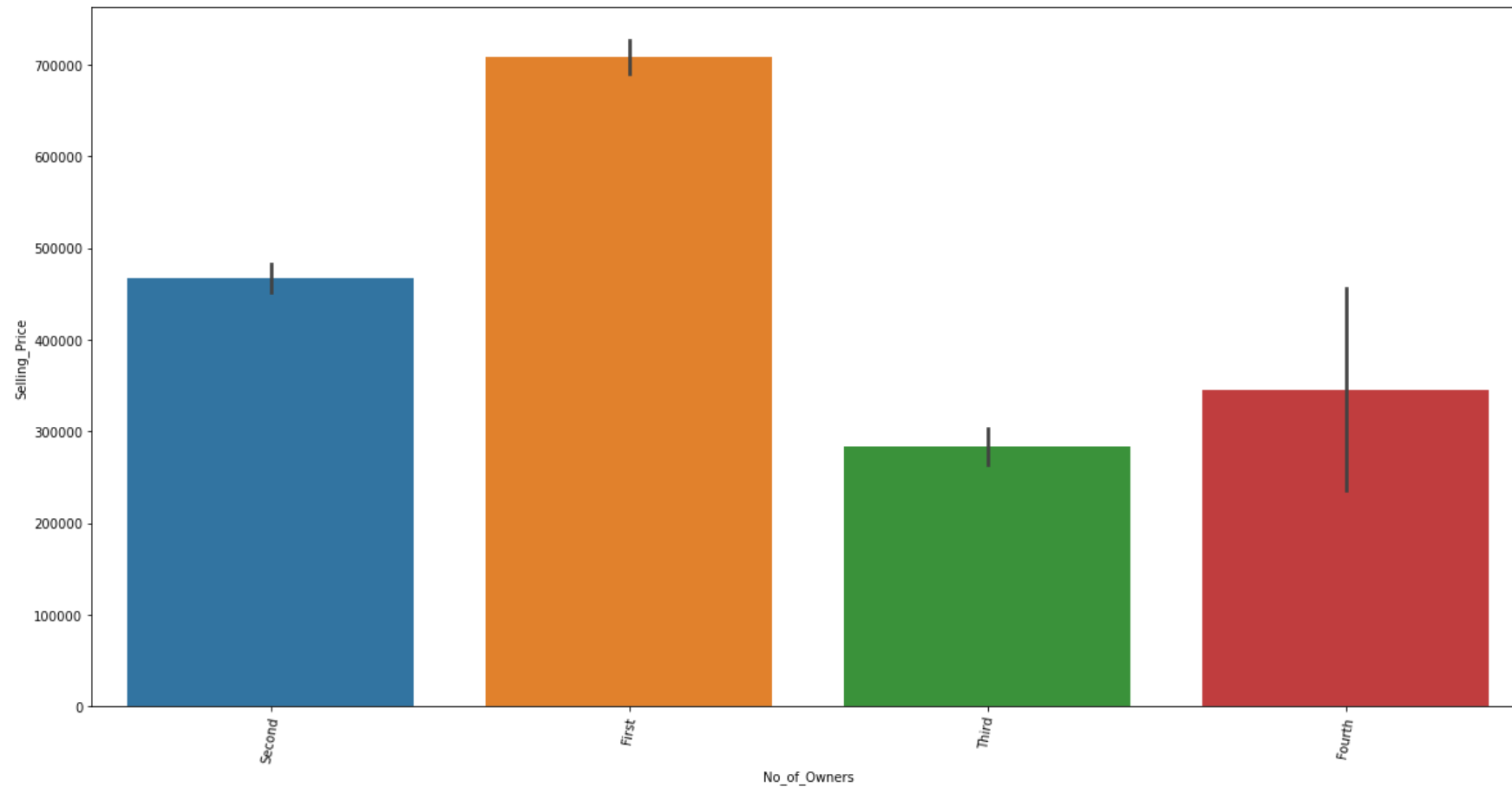
Model:



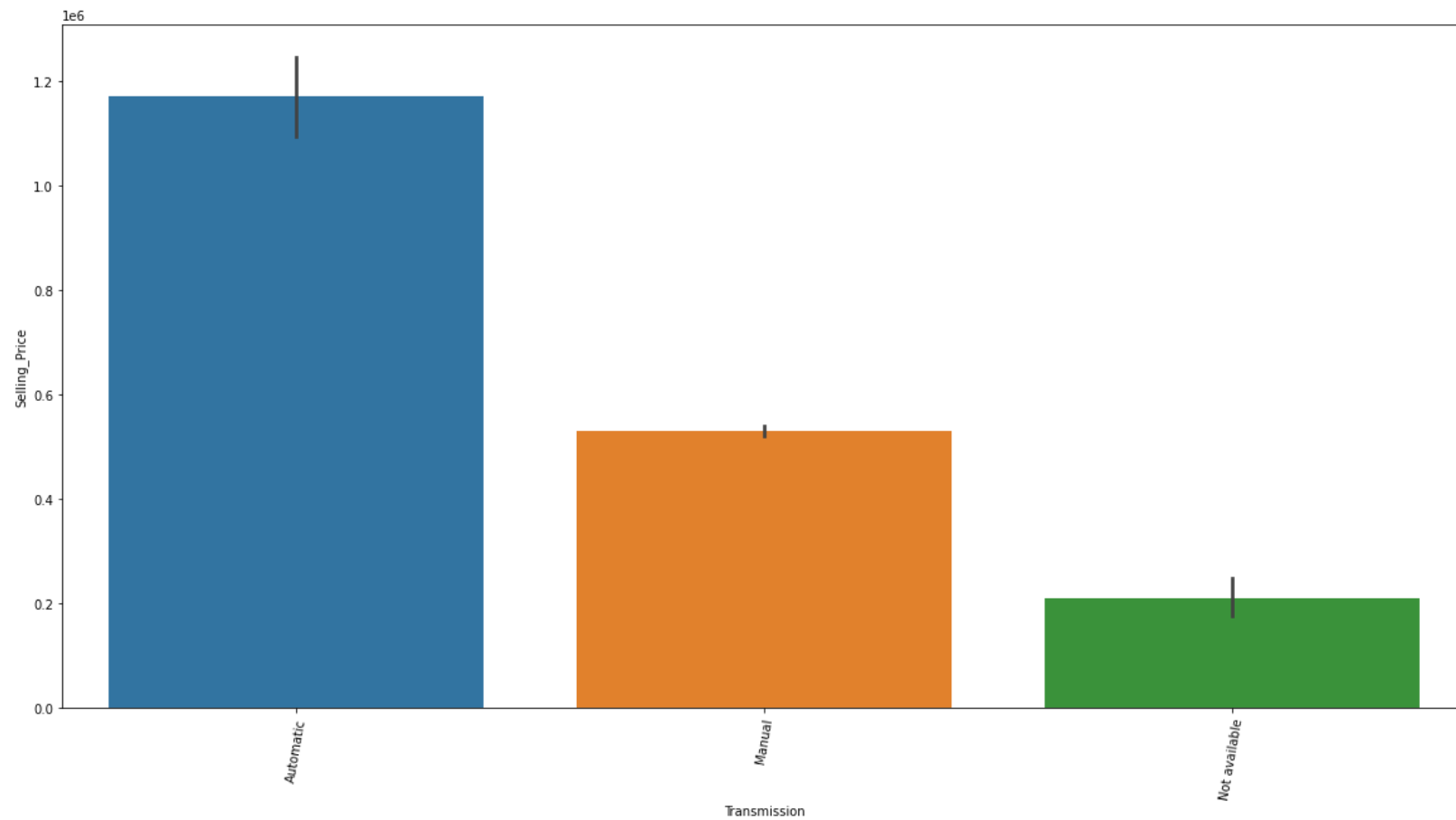
KM's driven:



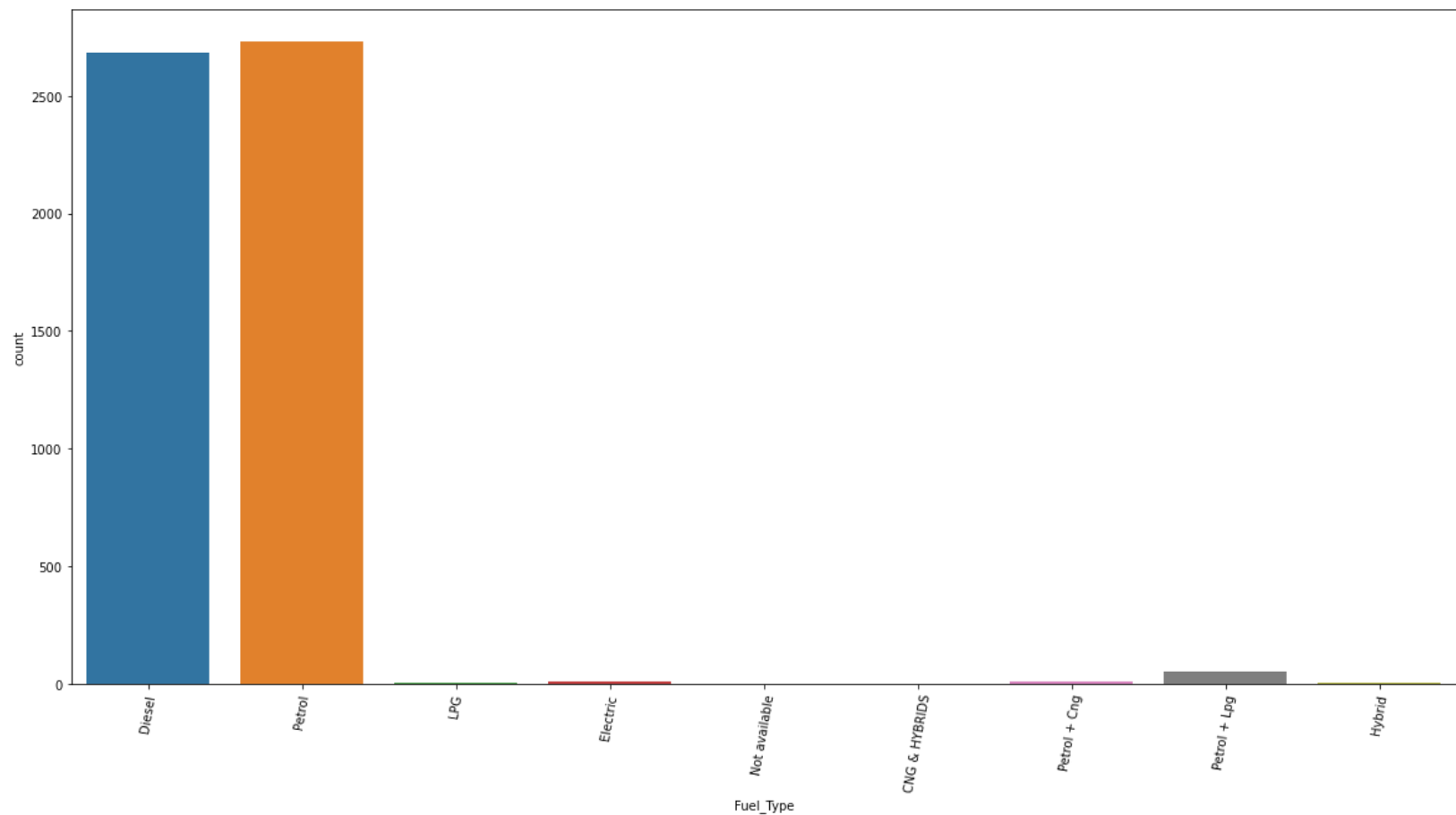
No of owners:



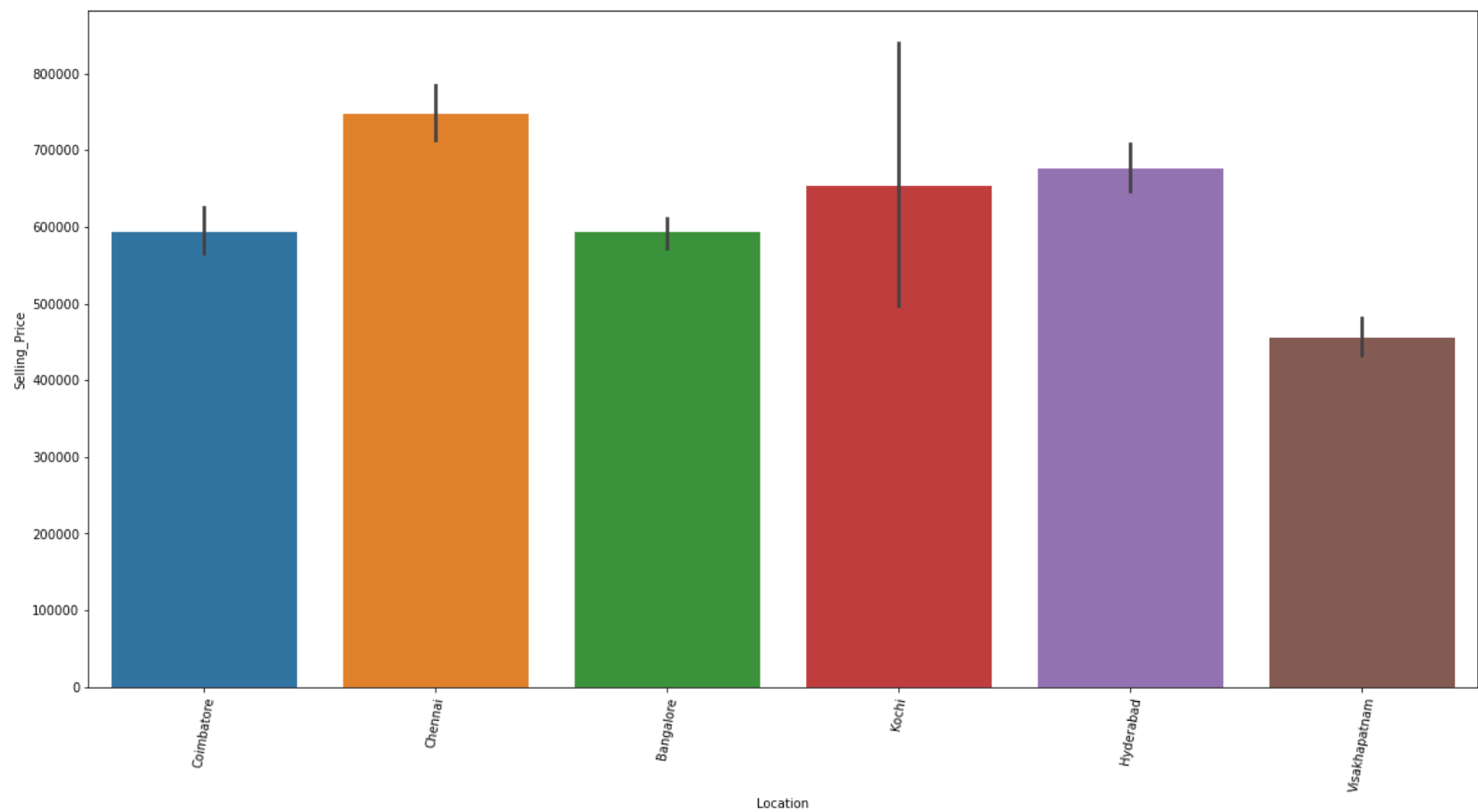
Transmission:



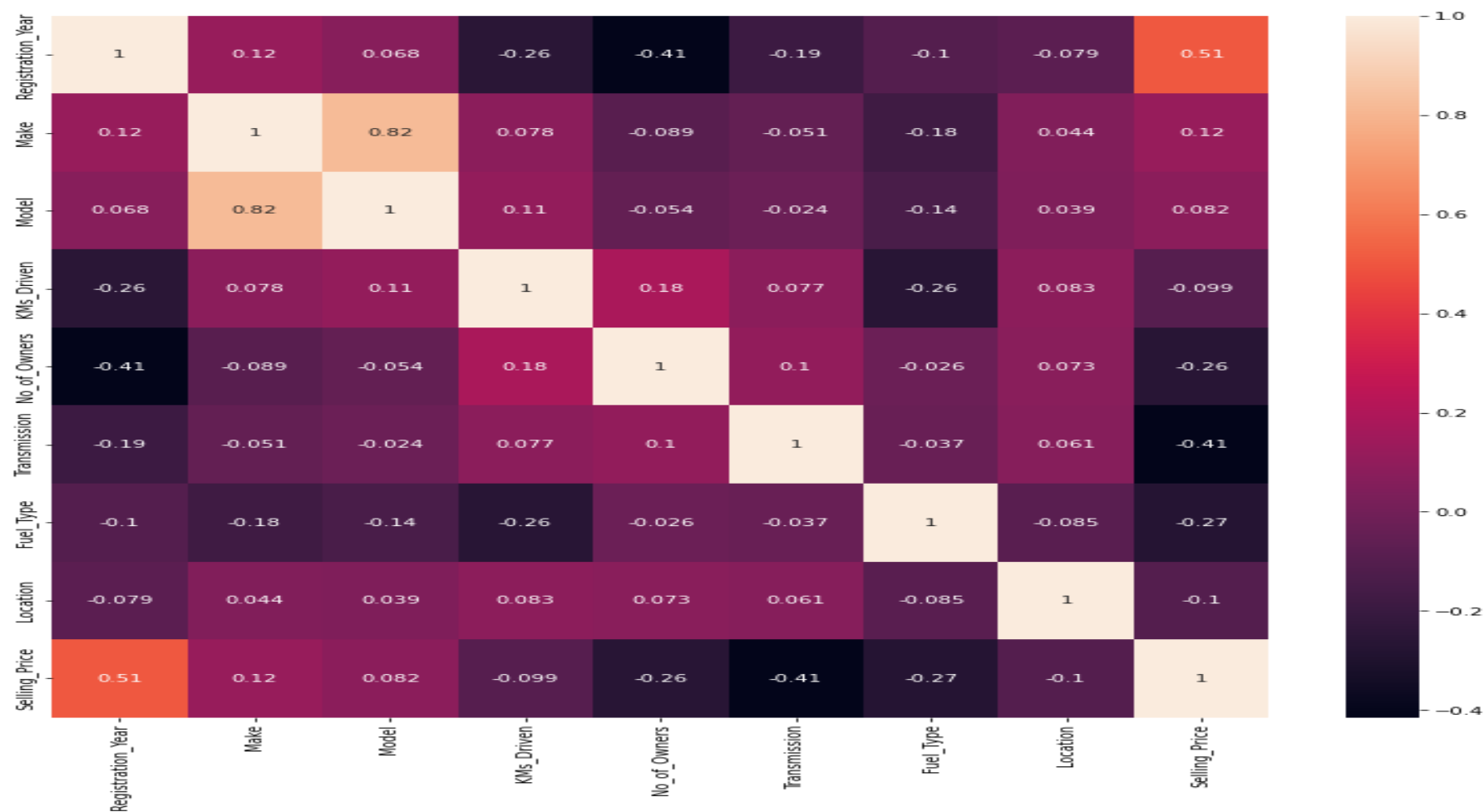
Fuel Type:



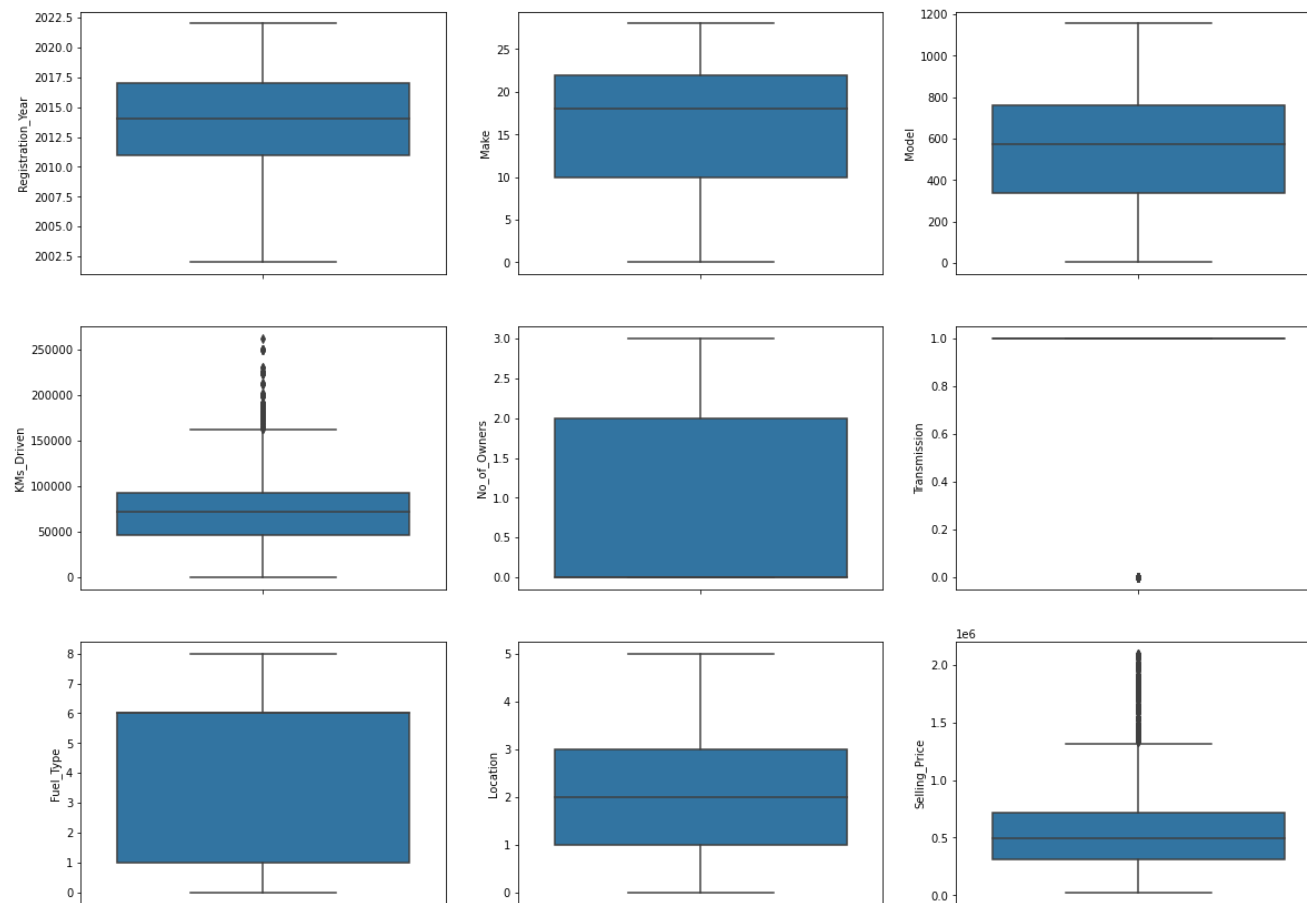
Location



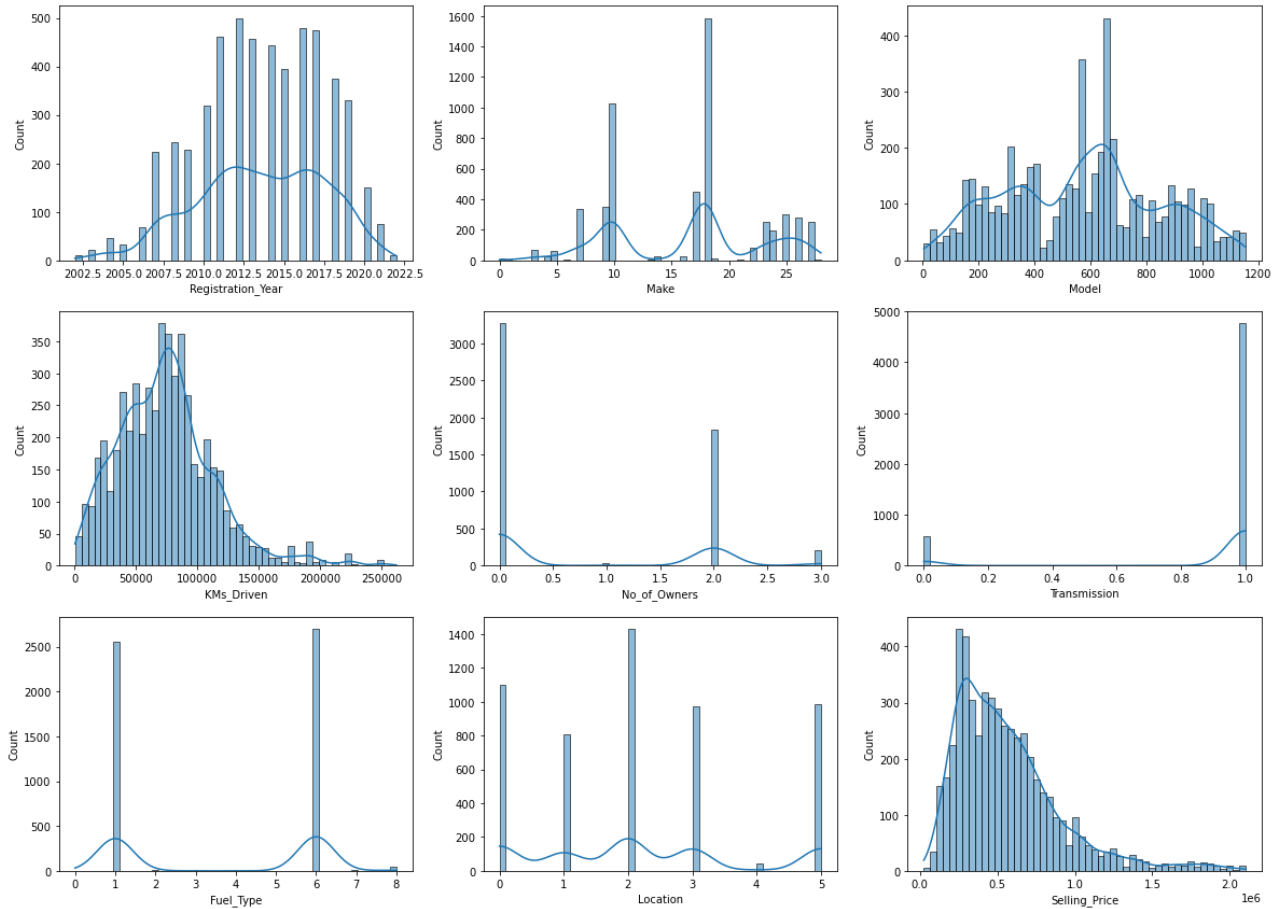
Heatmap:



Boxplot:



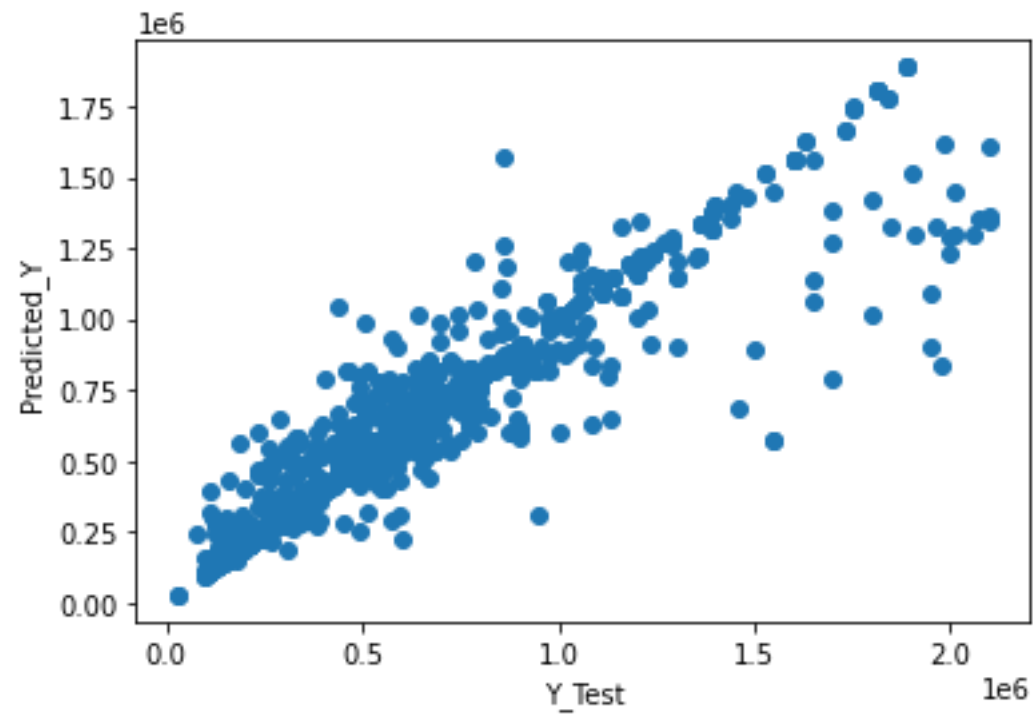
Histogram:



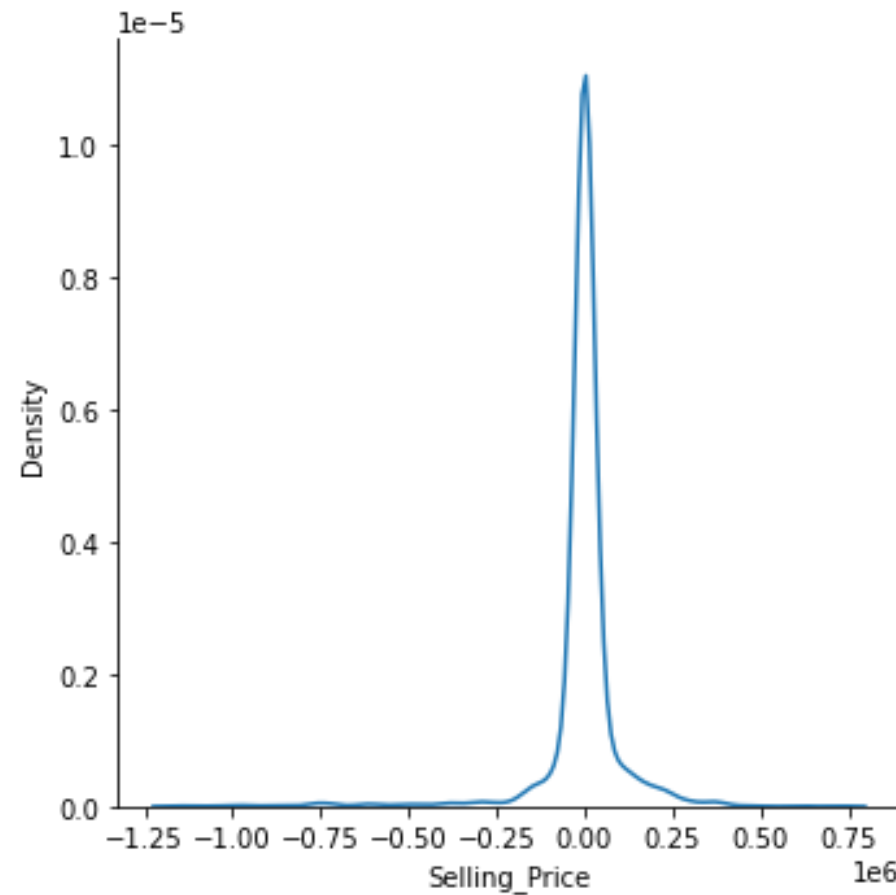
Testing of Identified Approaches (Algorithms):

- Linear Regression
- Gradient Boosting Regressor
- AdaBoost Regressor
- Decision Tree Regressor
- KNeighbors Regressor
- Extra Trees Regressor
- Random Forest Regressor

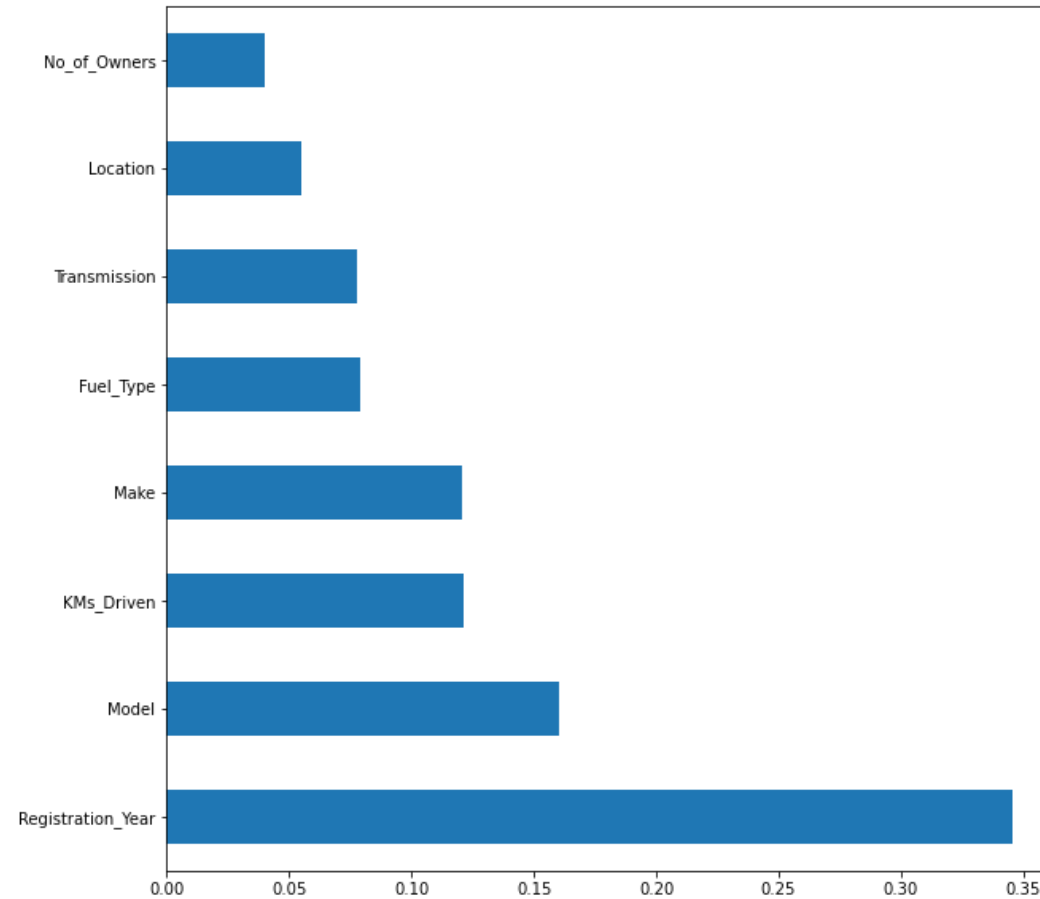
Y test plot:



Distribution of Predicted Values:



Feature importance's



Accuracy Parameter:

- R2 Score: 88.5718027742514
- Mean Absolute Error: 43667.17472188806
- Mean squared Error: 14141070734.25445
- Root Mean Absolute Error: 208.96692255447527

The best model is obtained by Hypertuning the existing models.

Key Findings and Conclusions of the Study:

- This dataset has been taken from 3 websites of this, Droom website constitutes the majority of data
- Since the target feature is continuous data, this problem can be solved by regression algorithms
- Random Forest Regression gives an R^2 score 0.88
- Registration year dominates the selling price more.