

CAR PRICE PREDICTION

Submitted by:

Shivam N. Gadekar

**ACKNOWLEDGMENT**

Thanks to cardekho.com, for providing user friendly HTML framework and keeping all tags and class names common. Also thanking them for not giving any frequent popups asking contact details and address.

Also thanks to local car dealer, His domain knowledge helped me to find out key factors playing important role in car price.

**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.

* Conceptual Background of the Domain Problem

Second hand cars sales increasing day by day as people are using new cars/ new technology, our goal is to identify which key aspect need to consider during car price finalization. From customer as well as supplier point of view. This analysis will help to predict car price based on few factors.

* Review of Literature

Based on discussion with local car dealer, he said there are three things which plays important role in defining car price. First is Brand, second is distance travelled and most important appearance, we are not considering colour in this analysis.

Also one more factor, Registration year plays important role.

* Motivation for the Problem Undertaken

Being from mechanical and first problem in real time data, this motivated me to work even harder.

**Analytical Problem Framing**

* Mathematical/ Analytical Modeling of the Problem

Describe the mathematical, statistical and analytics modelling done during this project along with the proper justification.

* Data Sources and their formats

Data fetched from cardekho.com and it was in HTML format. Extracted all data using selenium web driver. Used for loop to remove symbols, for loops and replace functions.

* Data Preprocessing Done

1. For price, Removed Rupee symbol and space. Converted it into int/float for numerical processing.
2. For distance travelled, remove kms and converted it into int and rounded it off.
3. For car name, splatted data with the help of split by space. Then picked second value as name, and keeping first as registration year.
4. For registration year, sliced first four letters of string and converted it into number and created new column giving years\_old value.

* State the set of assumptions (if any) related to the problem under consideration

Considered all data is valid and there are no outliers.

Car colour is not considered.

* Hardware and Software Requirements and Tools Used

Intel i3, HDD-1 TB and 8 GB ram

**Model/s Development and Evaluation**

* Identification of possible problem-solving approaches (methods)

For first step collect few features and keeping all features after VIF as all values as less than 10.

As output variable is numerical, chosen regression models for model development.

Also tried to standardize data, using standard scaler.

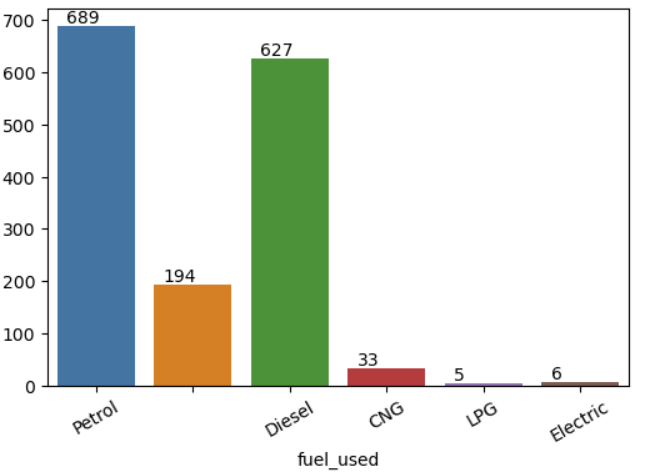
* Testing of Identified Approaches (Algorithms)

Linear Regression, Lasso, Ridge, Random Forest, Gradient Boosting Regressor.

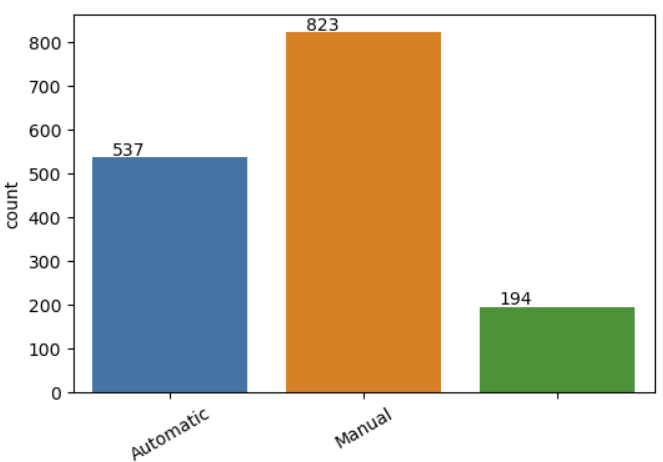
* Run and Evaluate selected models

From all above models, chosen Random Forest Regressor with random state 100. After hyper parameter tuning get r2 as 88.

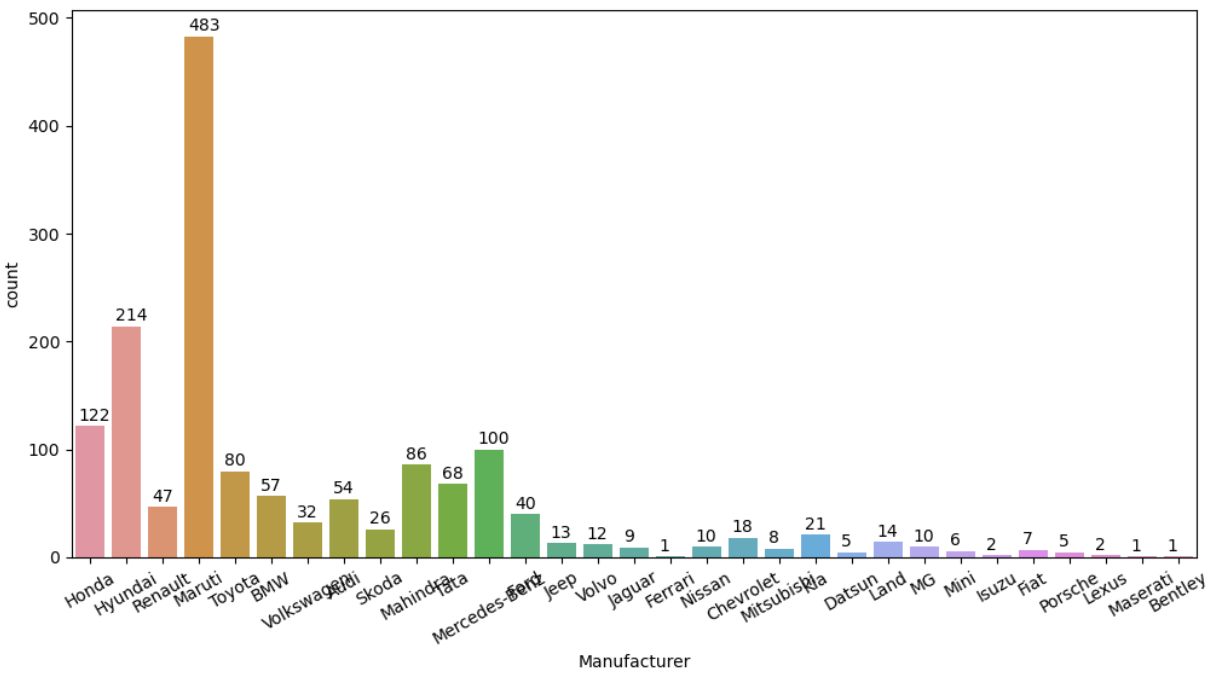
* Visualizations



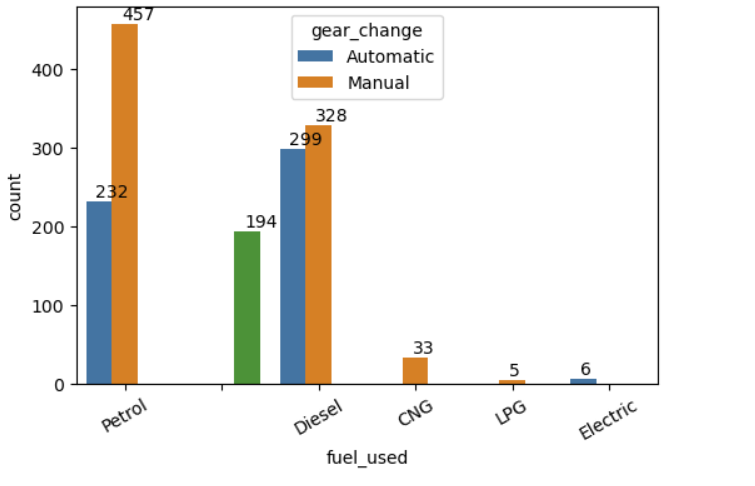
Petrol and diesel cars are widely used.



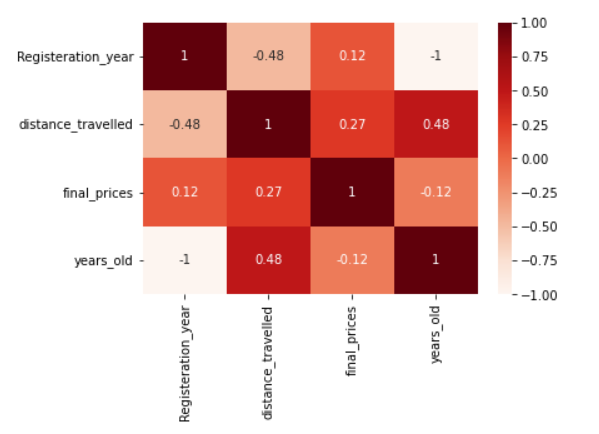
There are more cars with Manual Transmission.



Maruti being leading second hand cars seller and Hyundai being second followed by Honda and Benz.



Almost all 50-50 diesel cars are Manual and Auto transmission. Petrol cars having more manual transmission cars.



Distance travelled showing strong relationship with Price.

* Interpretation of the Results

1. Petrol cars are widely used.
2. Need more scope for electrical vehicles.
3. Maruti Is leading in extracted data set.

**CONCLUSION**

* Key Findings and Conclusions of the Study

Petrol cars are widely used. Adding to that Maruti is highest shareholder, with manufacturing petrol cars. With increasing price, distance travelled gets reduced. Also value is more, if car is of reputed manufacturer.

* Learning Outcomes of the Study in respect of Data Science

Distance travelled majorly plays important role, almost 50% and remaining other years\_old, colour, brand, fuel used.

* Limitations of this work and Scope for Future Work

We can open each car and extract more description about cars to get more details.