**Abstract**

The goal of this project was to use linear regression models to know if customers would be willing to pay for a used vehicle based up the vehicles characteristics? To Recommendations for Business.

**Design**

This project originates from the data of car selling price and car specifications. The data is provided by Kaggle website to help the businessman to production price depending on car specifications, ​cars that are not sold makes will result in less flooring, reducing capital investment

**Data**

A data frame containing 8,128 observations on 13 variables, it is contained name, year, selling price, km driven, fuel, seller type, transmission, owner, mileage, engine, max power, torque, seats. Dataset is from Kaggle website the last updated of this data is in 2020-10-24

Reference of Dataset: <https://www.kaggle.com/nehalbirla/vehicle-dataset-from-cardekho?select=Car+details+v3.csv>

**Algorithms**

1. Creating car company from the name of car taking a cut of the first word of name without model of the car.
2. Converting string value to integer.
3. Delete all null value and cut all shortcut ex.CC and keep it as numbers.
4. Fix linearity of linear Regression.

*Model Evaluation*

The entire training dataset of 8,128 records was split into 70/30 train vs. testing.

**Final Linear** **Regression Model scores:** Accuracy 92%

* R-squared = 0.923
* Mean absolute error = 0.17

*Models*

### Conclusion and Recommendations for Business

Analysis of trends in used car vehicle sales indicates that:

1) Maruti, Hyundai, Honda, Toyota, Ford, Mahindra and Tata are sold in higher volume than other vehicle makes, and the preferred makes are lower in price.

2) The Price (lower rather than higher) is a significantly factor in customer purchasing decisions.

3) The customers prefer vehicles with mileage between 15-25 kmpl.

4) The customers overwhelmingly prefer manual transmissions over automatic.

5) The customers tend to purchase used vehicles 2 -5 years old.

##### *Based upon the above observations, it is recommended that focuses its inventory to match such preferences. Since the higher in demand makes are lower in price than other makes, focusing on stocking these makes will result in less flooring, reducing capital investments.*

##### *The model that has been provided in this analysis is a 92% accurate pricing model for determining what customers would be willing to pay for a used vehicle based up the vehicle's characteristics. It has been shown that the model is reliable, meeting the assumptions for linear regression and demonstrating a low error rate.*[*¶*](http://localhost:8888/notebooks/PLR.ipynb#The-model-that-has-been-provided-in-this-analysis-is-a-92%-accurate-pricing-model-for-determining-what-customers-would-be-willing-to-pay-for-a-used-vehicle-based-up-the-vehicles-characteristics.-It-has-been-shown-that-the-model-is-reliable,-meeting-the-assumptions-for-linear-regression-and-demonstrating-a-low-error-rate.)

**Tools**

* Numpy and Pandas for data manipulation.
* Matplotlib and Seaborn for plotting.
* PyLab is a procedural interface to the Matplotlib object-oriented plotting library.
* Scikit-learn for modeling.
* Import warnings to hide all warnings in python

**Communication**

In addition to the slides will be embedded on my GitHub <https://github.com/Malzaid1/BootCamp_projec> .