Cars4U Project Business Presentation

Contents

- Business Problem Overview

- Data Overview

- Exploratory Data Analysis (EDA)

- Model Performance Summary

- Business Insights and Recommendations

# Business Problem Overview and Solution Approach

- Cars4U is a tech startup that wants to find a foothold in the market of used cars, given the rising demand for used cars in the current Indian market.

- New car sales (around 3.6 million units) captured a lower market share compared to second-hand car sales (around 4 million units) in 2018-19, indicating higher relative importance for pre-owned cars in the Indian car market.

- Used cars show significantly different behavior compared to new cars, with huge uncertainty in both pricing and supply. This calls for a proper setup in the pricing scheme for these used cars in order to grow in the market.

- We will develop a linear regression model by analyzing data related to the various attributes of a car (mileage, type of fuel, power, etc.) to help set up a differential pricing scheme and use it to devise profitable strategies.

Data Overview

- The data contains information about 7253 cars and their characteristics.

- The characteristics include car name, the location in which the car is being sold or is available for purchase, manufacturing year of the car, power of the car, kilometers driven in the car, car mileage, fuel type used by car, transmission type used by car, used model price, new model price, and more.

- Many of the columns have values with units attached, which we will remove before proceeding with the analysis.

- Some columns (like Kilometers_Driven and Mileage) have some extreme and irregular values, which warrants a quick outlier and anomaly check.

- We will extract brand and model names from the car name for our analysis.

- We will also apply a log transform to our target variable (Price) to remove the skewness, making it better suited for linear regression.
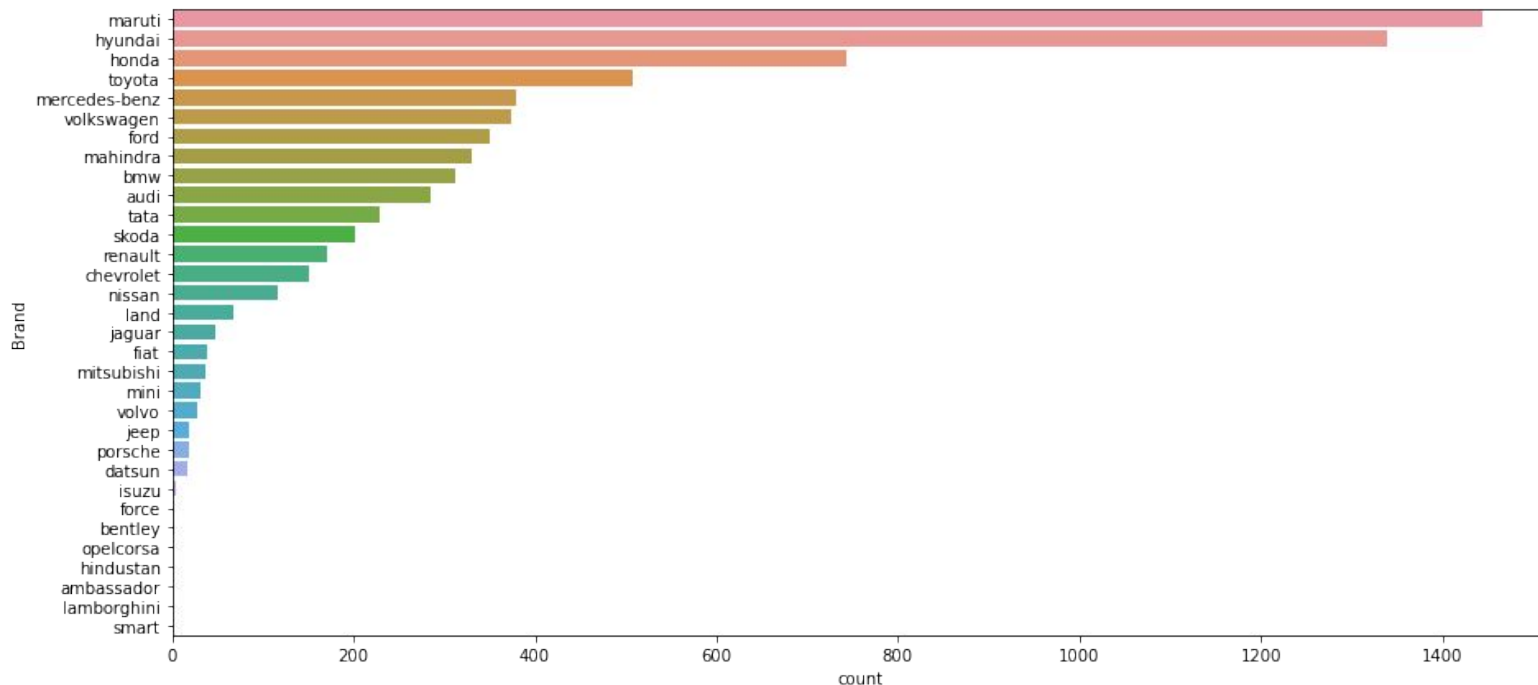
Exploratory Data Analysis

- Both Price and Kilometers_Driven are heavily skewed.

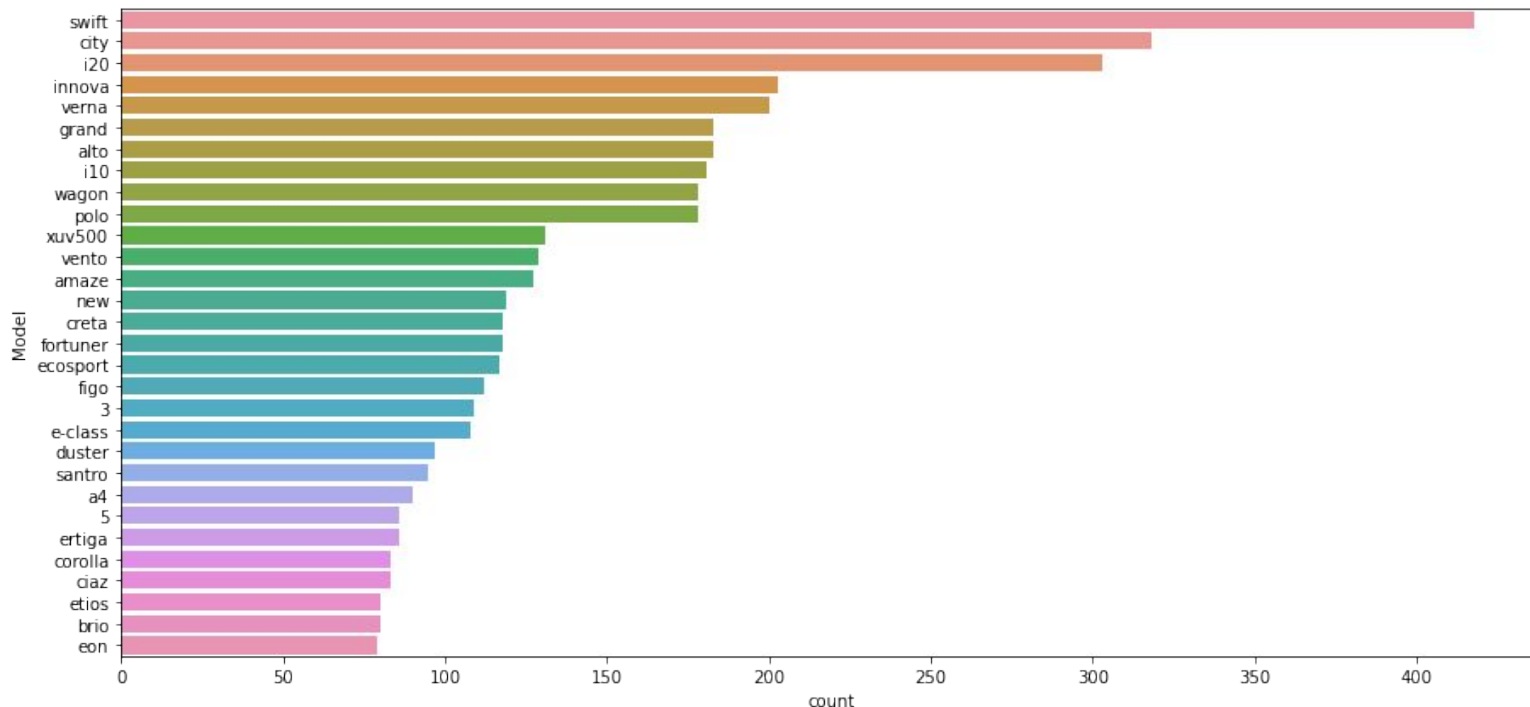- Log transform has been applied to both to reduce the extreme skewness.
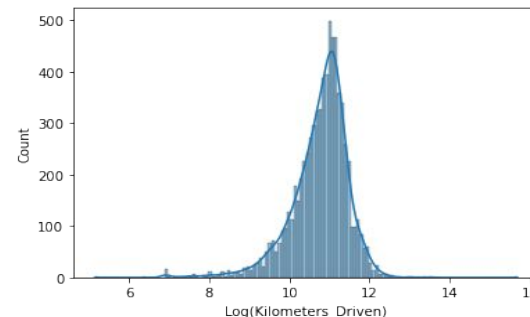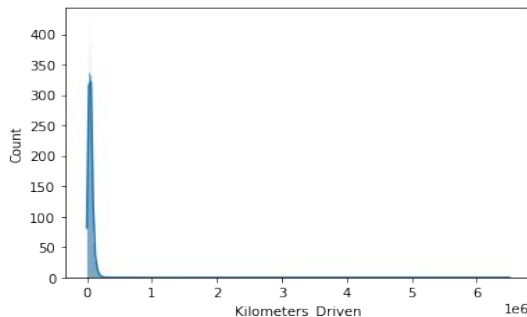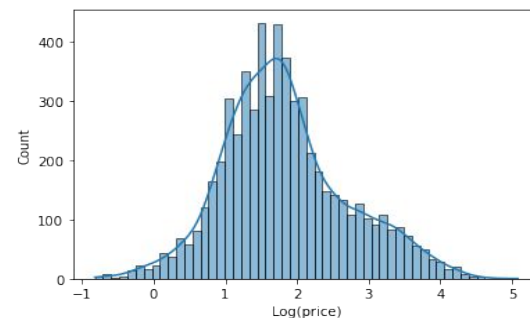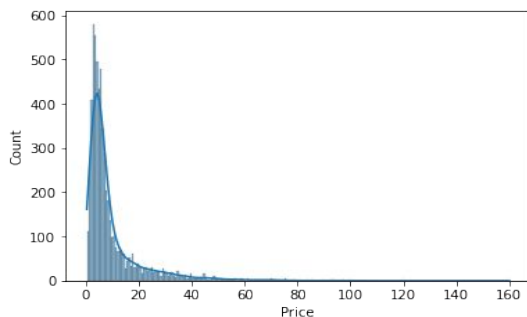
Exploratory Data Analysis

- Maruti and Hyundai dominate the Indian car market, followed by other popular brands like Honda and Toyota.
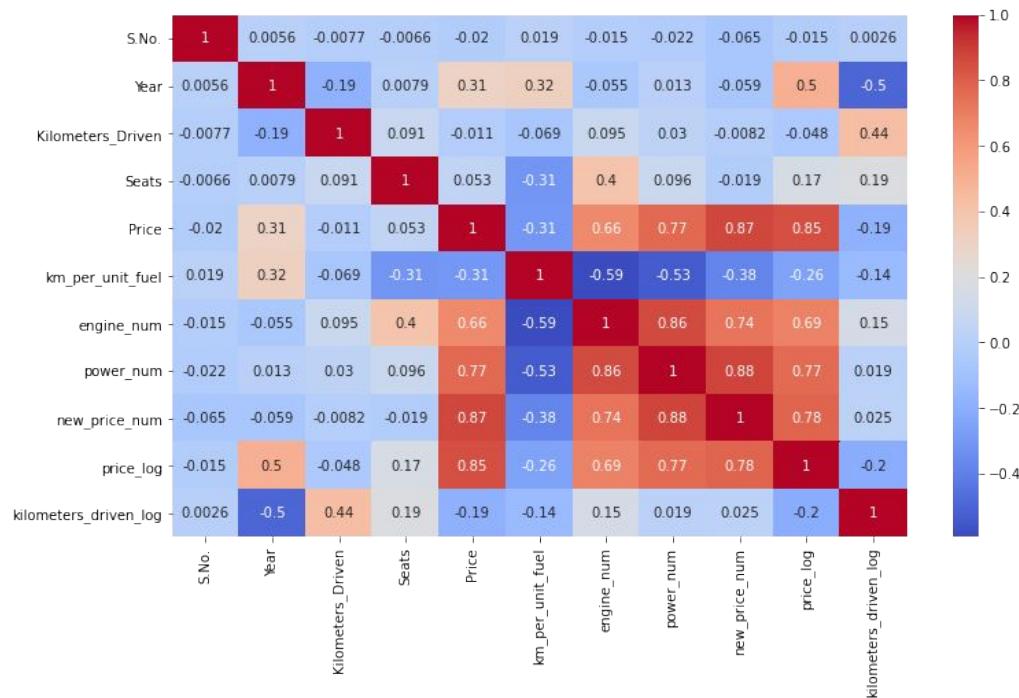
Exploratory Data Analysis

- Maruti Swift seems to be the leader when it comes to models. Honda City and Hyundai i20s take the second and third spots.

Exploratory Data Analysis

- Contrary to intuition Kilometers Driven does not seem to have a relationship with price.

- Price has a positive relationship with Year. Newer the car, the higher the price.

- Mileage does not seem to show much relationship with the price of used cars.

- Engine displacement and Power of the car have a positive relationship with the price.

- New Price and Used Car Price are also positively correlated, which is expected.

Exploratory Data Analysis

- Price varies with location, with Coimbatore and Bangalore showing comparatively higher prices for used cars than the rest of the locations.
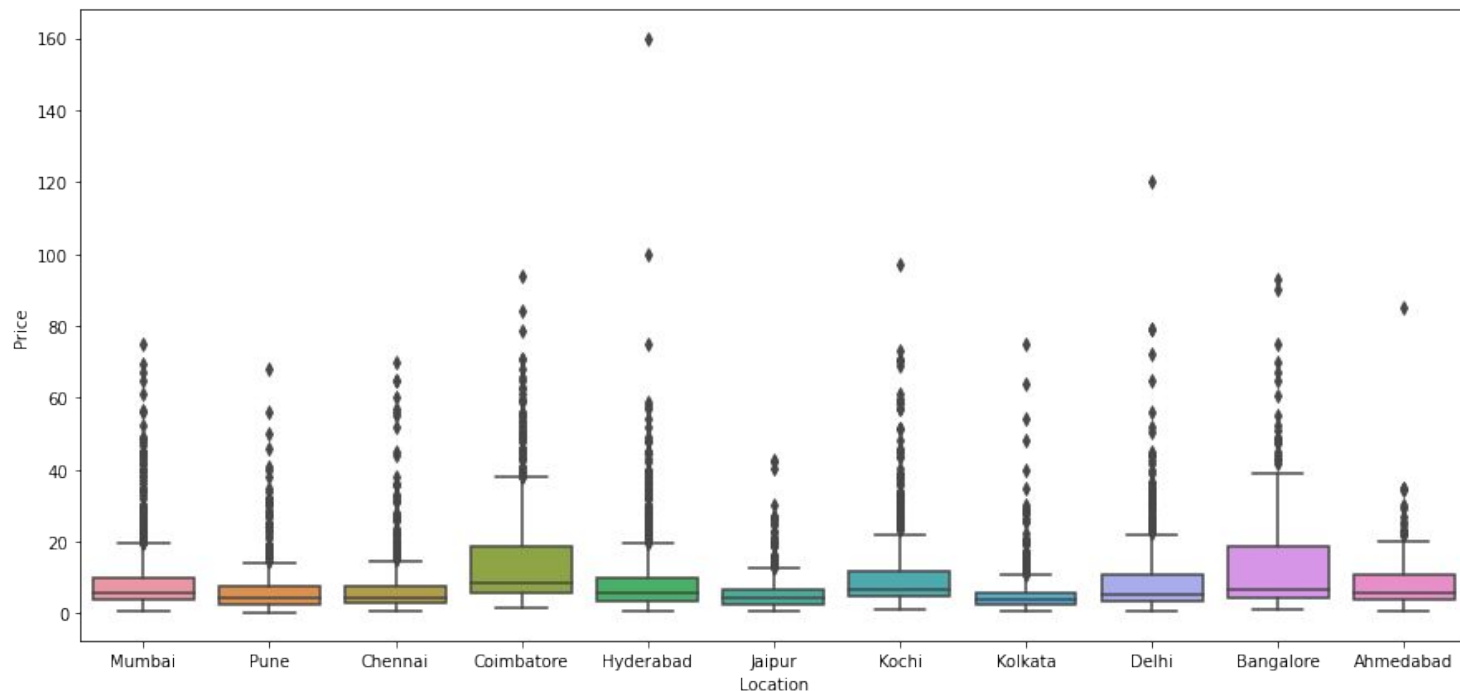
Model Performance Summary

- We want to predict the "Price" using the car characteristics initially provided to us and the ones we have engineered.

- We will use the normalized version 'price_log' for building a robust Linear Regression model using the train data and check the performance on test data to understand the predictive power of our model.

- The model indicates that the most significant predictors of the price of used cars are the following:

  - The year of manufacturing

  - Number of seats in the car

  - Kilometers Driven

  - Location

  - Type of fuel

  - Transmission - Automatic/Manual

  - Owner type - First to Fourth and above

  - Car Category - budget-friendly to ultra-luxury

Model Performance Summary

- We have got R-squared and adjusted R-squared as 0.906 and 0.906, which is a clear indication that we have been able to create a very good model that is able to explain variance in the price of used cars up to 90%.

Data	RMSE	MAE	MAPE
Train	5.04	1.94	21.74
Test	3.60	1.77	19.32

- Mean Absolute Error indicates that our current model is able to predict used car prices within a mean error of 1.8 lakhs on test data.

- MAPE is around 20% on the test data, which means that we are able to predict within 20% of the price value.

Business Insights and Recommendations

- Newer cars sell for higher prices and a unit increase in the year of manufacture leads to a 1.13 Lakh increase in the price.

- As the number of seats increases by 1 unit, the price of the car increases by 1.03 Lakhs.

- Mileage is inversely correlated with Price. Generally, high Mileage cars are the lower budget cars.

- Kilometers driven has a negative relationship with the price which is intuitive. A car that has been driven more will have more wear and tear and hence sell at a lower price, everything else being constant.

- All the car_category variables in the dataset have a positive relationship with the Price and the magnitude of this positive relationship increases as the brand category moves to the luxury brands.

Business Insights and Recommendations

- Our final Linear Regression model has a MAPE of 20% on the test data, which means that we are able to predict within 20% of the price value. This is a very good model and we can use this model in production.

- Some southern markets tend to have higher prices. It might be a good strategy to plan growth in southern cities using this information. Markets like Kolkata(coef = -0.2) are very risky and we need to be careful about investments in this area.

- We will have to analyze the cost side of things before we can talk about profitability in the business. We should gather data regarding that.

- The next step post that would be to cluster different sets of data and see if we should make multiple models for different locations/car types.