



# Cars4U – Project 3

---

Sarah A. Thomas

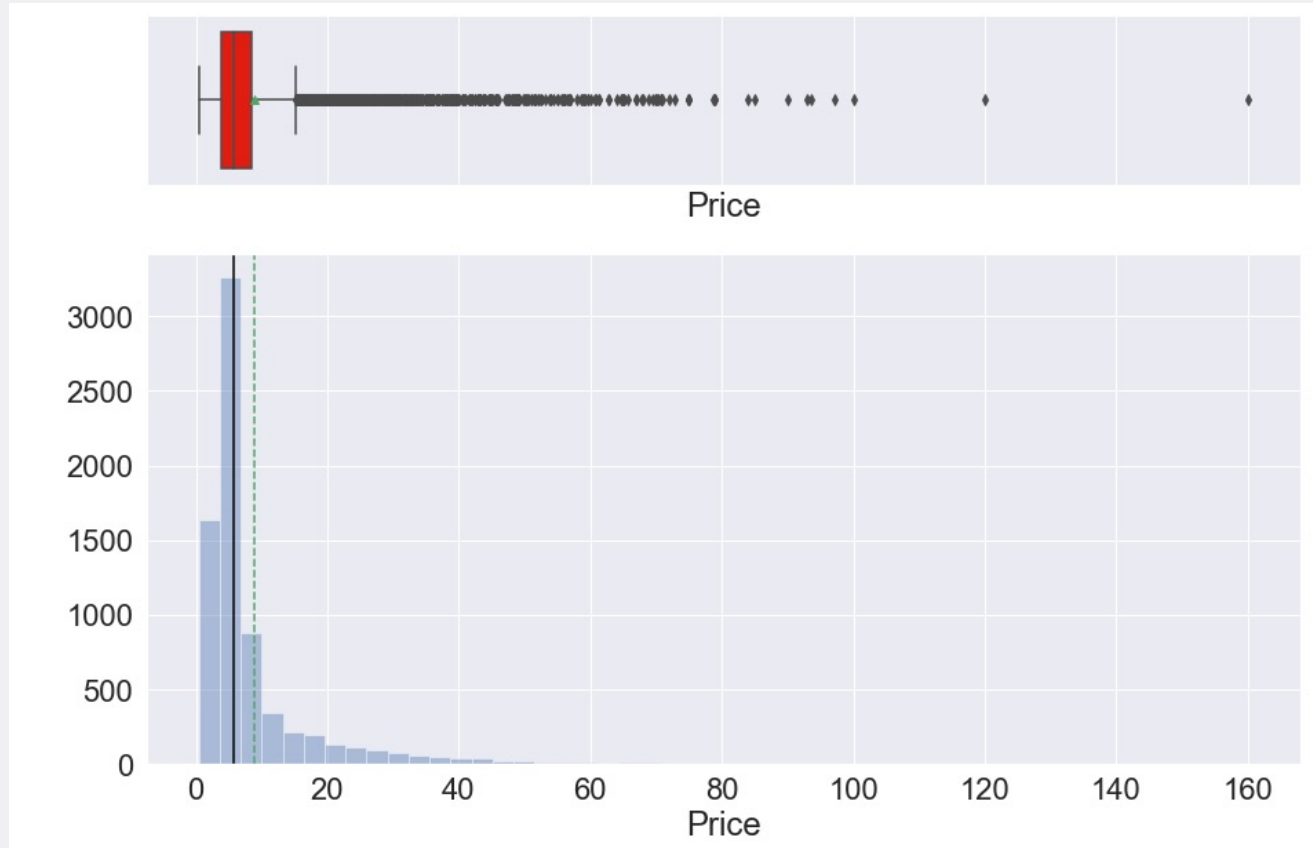
# Business Problem Overview and Solution Approach

- The Indian Market is experiencing increased demand for used cars while sales of new cars have slowed down.
- We know that in 2018-19 new car sales were at 3.6 million units while used cars were at 4 million in terms of number sold. Pricing scheme of the latter is of the utmost importance in growing the market.
- In order to construct a pricing scheme, we first need to create an accurate predictive model using data science, supervised learning techniques.

# Data Overview

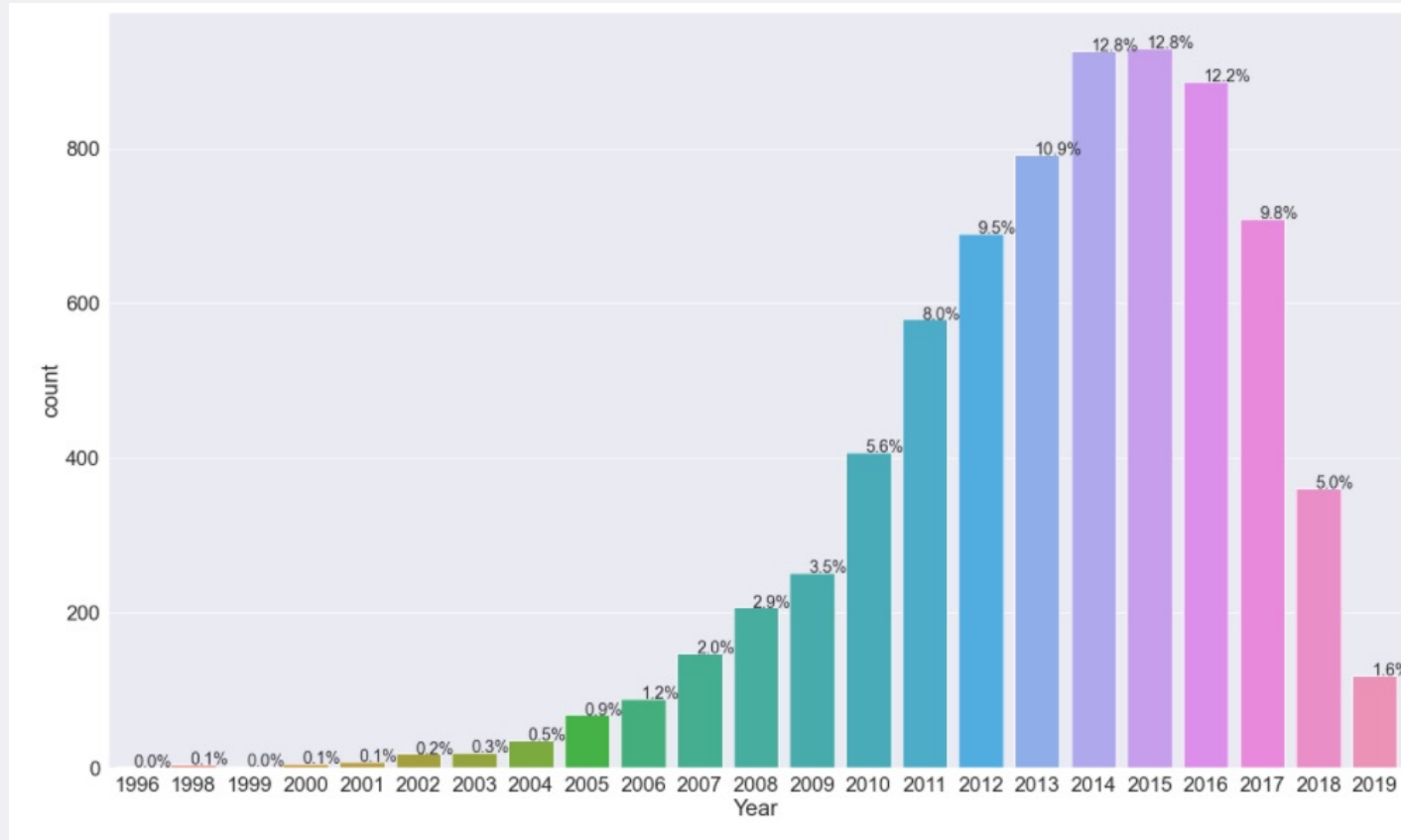
Variable	Description
S.No. - variable dropped	Serial Number
Name - split into 2 variables, Brand and Model (Model dropped in the final model)	Name of the car which includes Brand name and Model name
Location	The location/city in which the car is being sold or is available for purchase
Year	Year the car was manufactured
Kilometers_Driven - put on log scale	The total kilometers driven in the car by the previous owner(s) in km
Fuel_Type	The type of fuel used by the car (Petrol, Diesel, Electric, CNG, LPG)
Transmission	Type of transmission (Automatic or Manual)
Owner	Type of Ownership (ex. First-time owner)
Mileage	Standard mileage offered by the car company in kmpl or km/kg
Engine - variable dropped in final model	Displacement volume of the engine in CC
Power - put on log scale	Maximum power of the engine in bhp
Seats	Number of seats in the car
New_Price - variable dropped	Price of a new car of the same model in INR Lakhs
Price (dependent variable)	Price of the used car in INR Lakhs

# EDA - Price



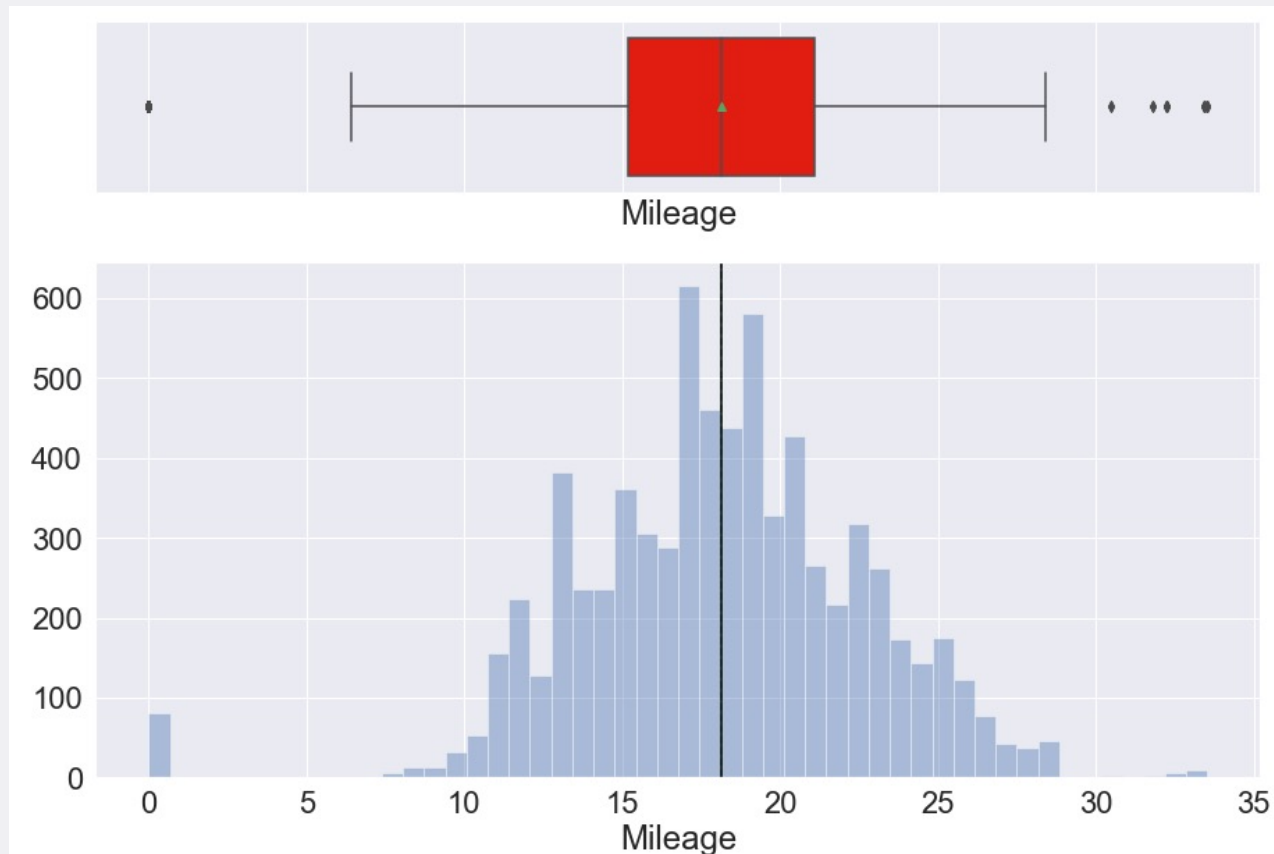
- Range: 0.44-160.0 Lakh
- Mean: 8.83 Lakh
- Median: 5.64 Lakh
- Positively Skewed

# EDA - Years



- The top 5 years are 2014, 2015, 2016, 2013, and 2017.
- The years range from 1996 to 2019.

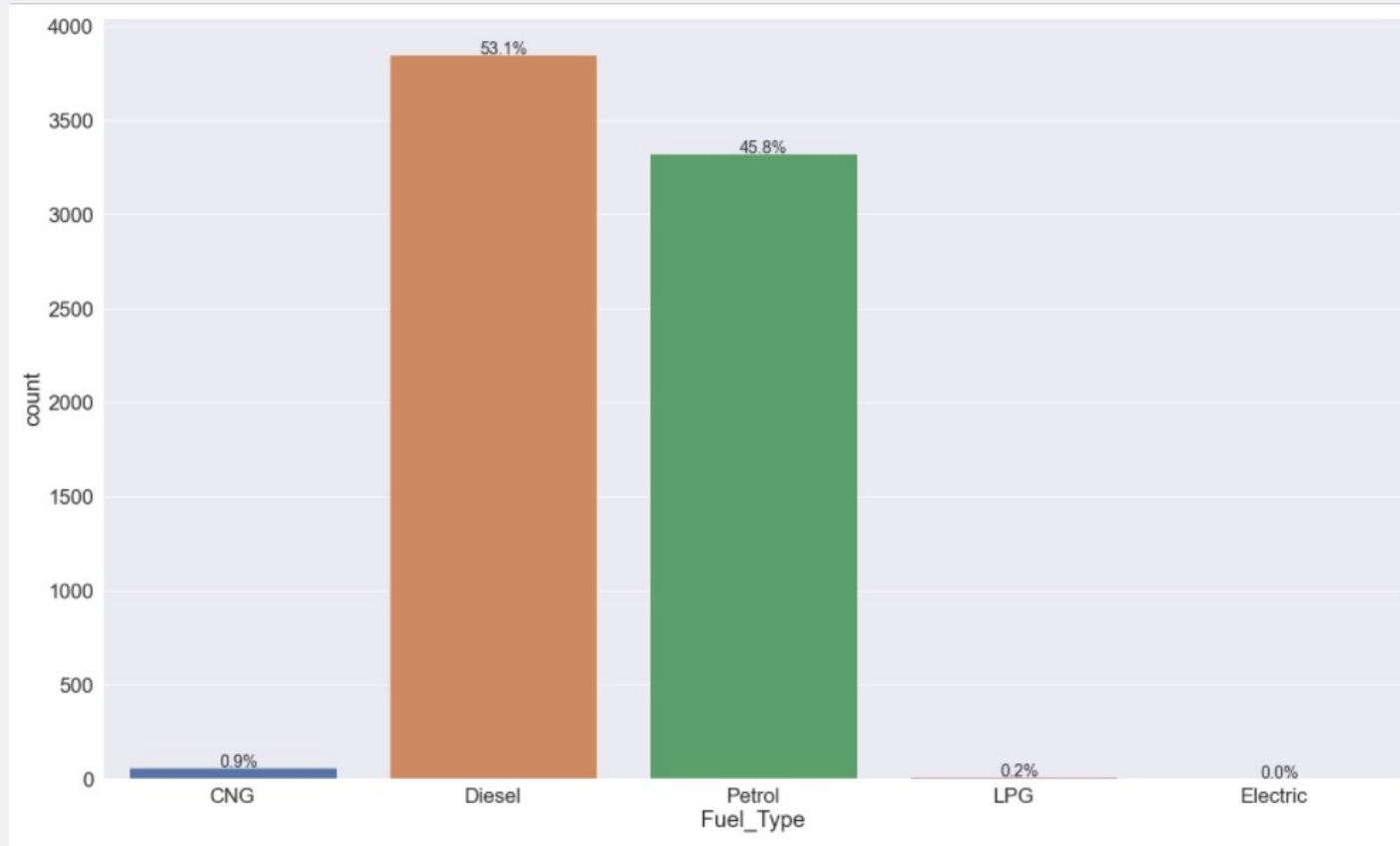
# EDA - Mileage



Mileage has outliers on both the left and right, but more so on the right. Distribution looks close to normal.

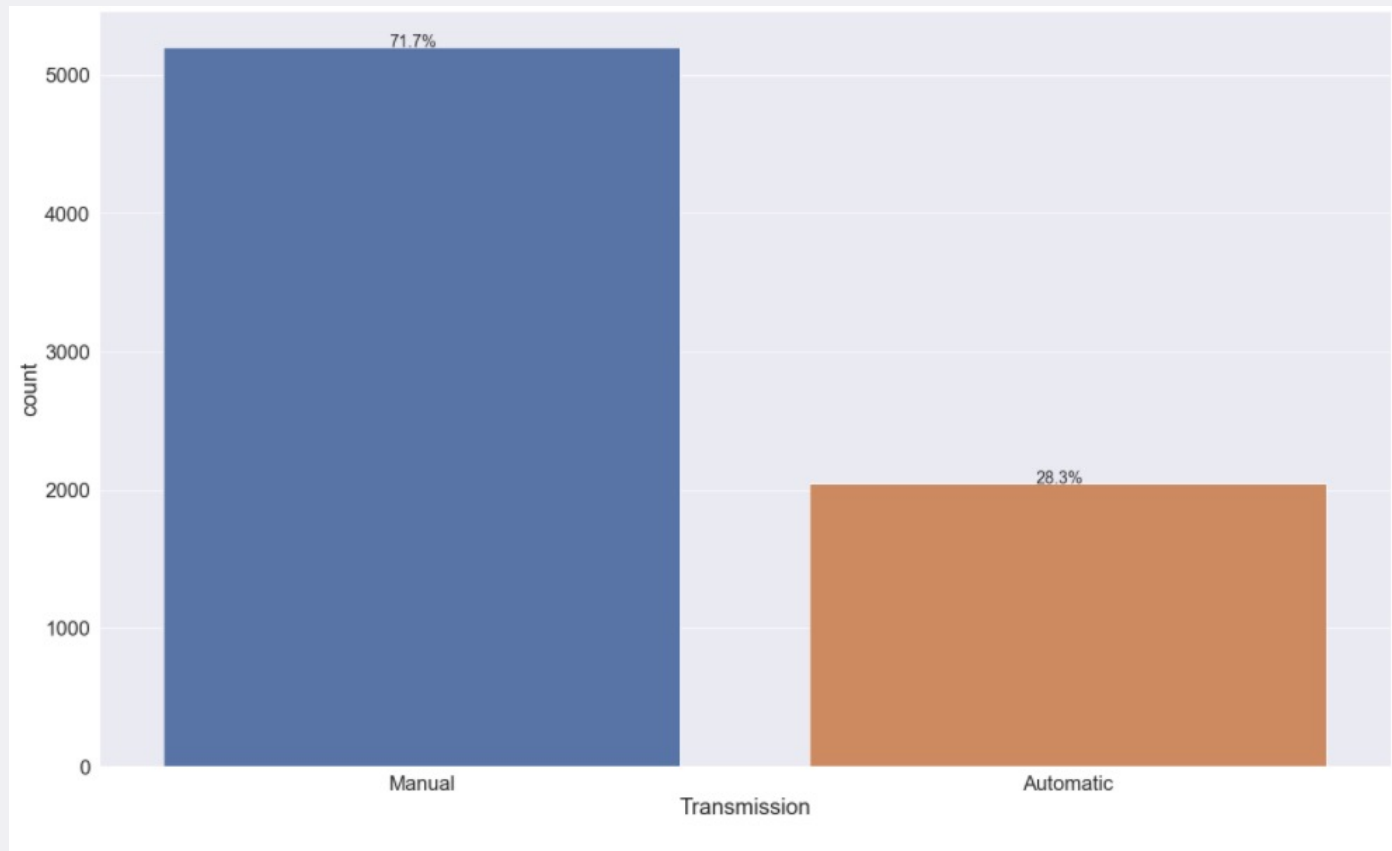
- Range: 0 to 33.54 kmpl
- Mean: 18.14 kmpl
- Median: 18.16 kmpl

# EDA - Fuel Type



Diesel and Petrol are by far the most prevalent fuel types.

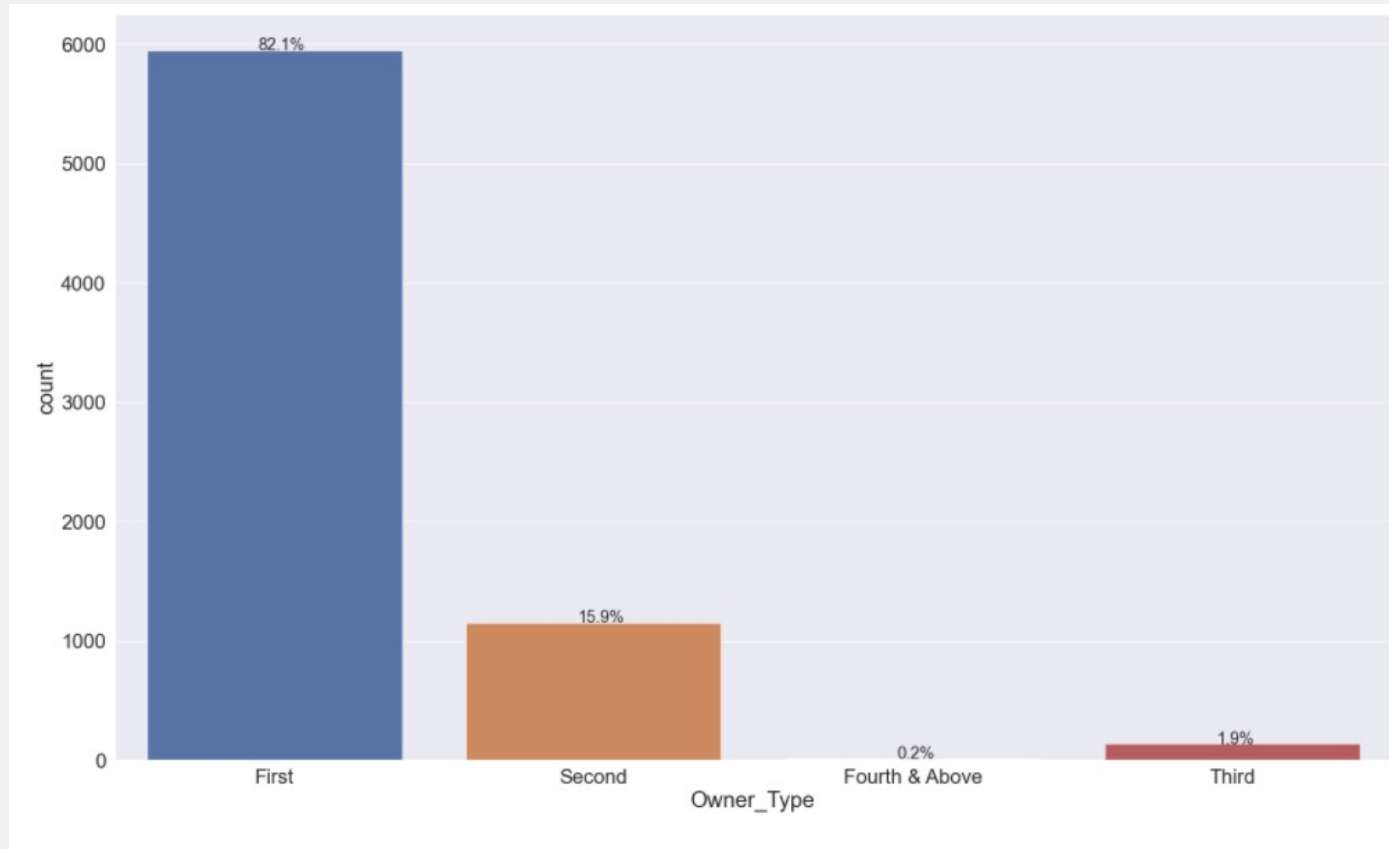
# EDA - Transmission



Manual transmissions are most prevalent.

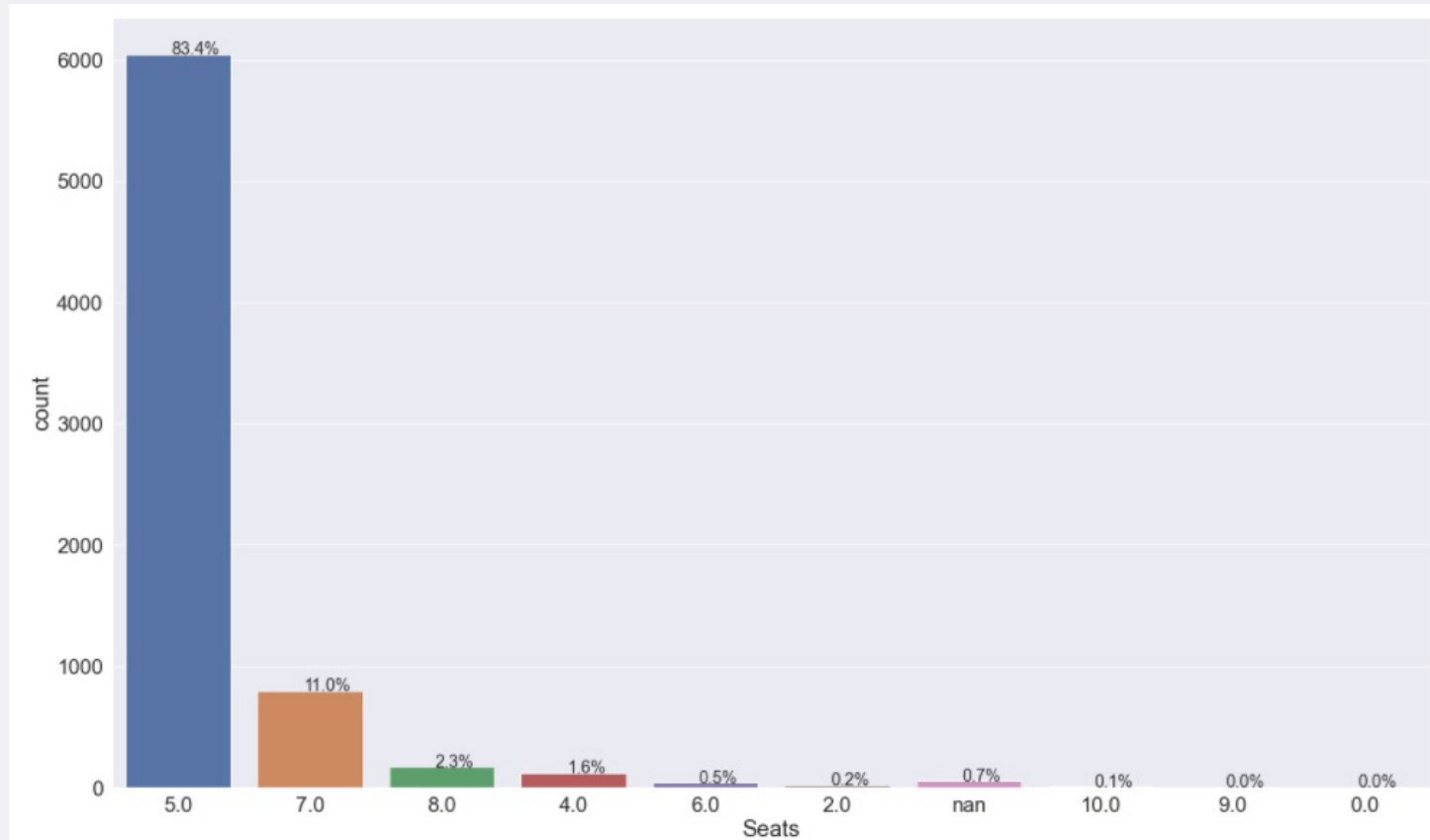


## EDA - Owner Type



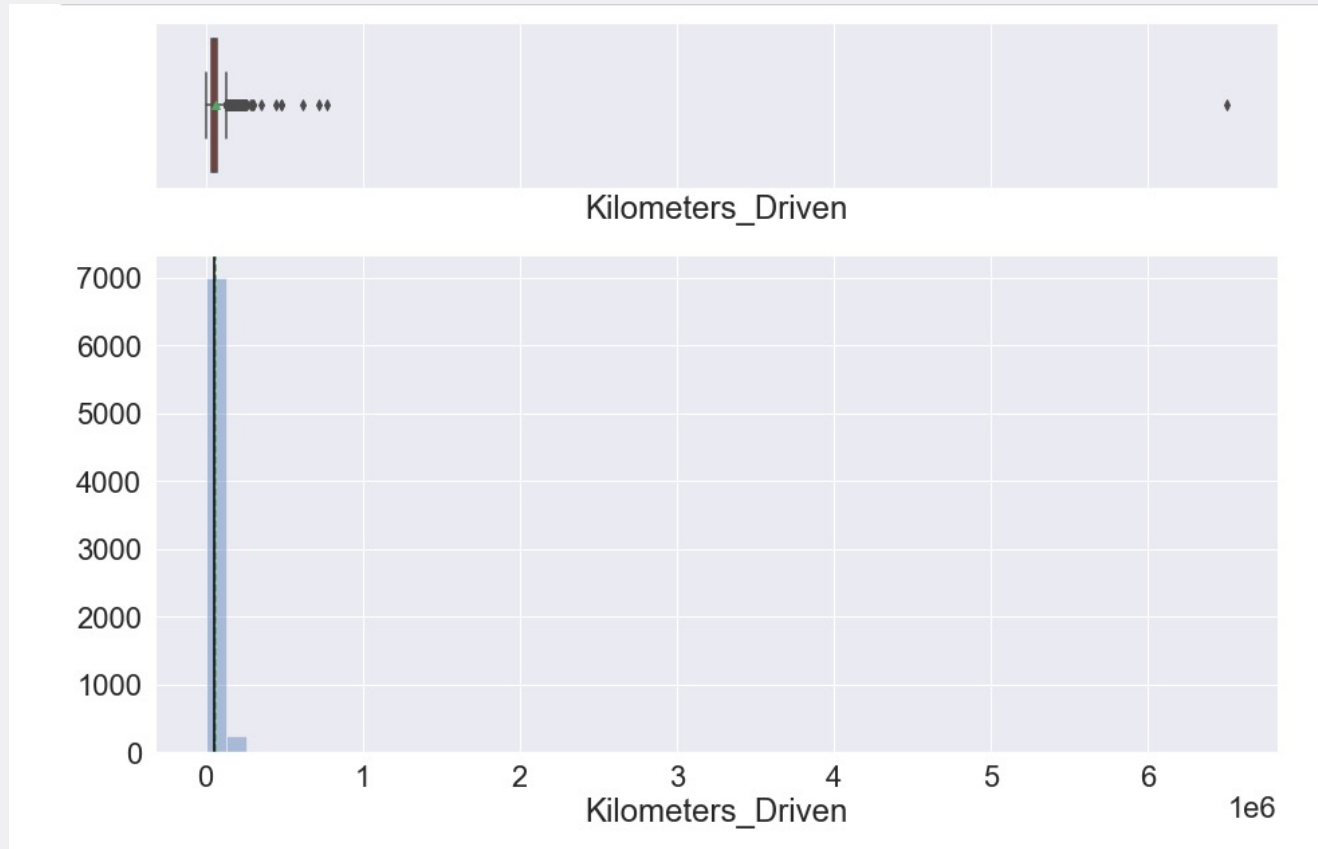
Most owners are first-time car owners.

## EDA - Seats



5.0 is by far the prevalent number of seats.

# EDA - Kilometers Driven



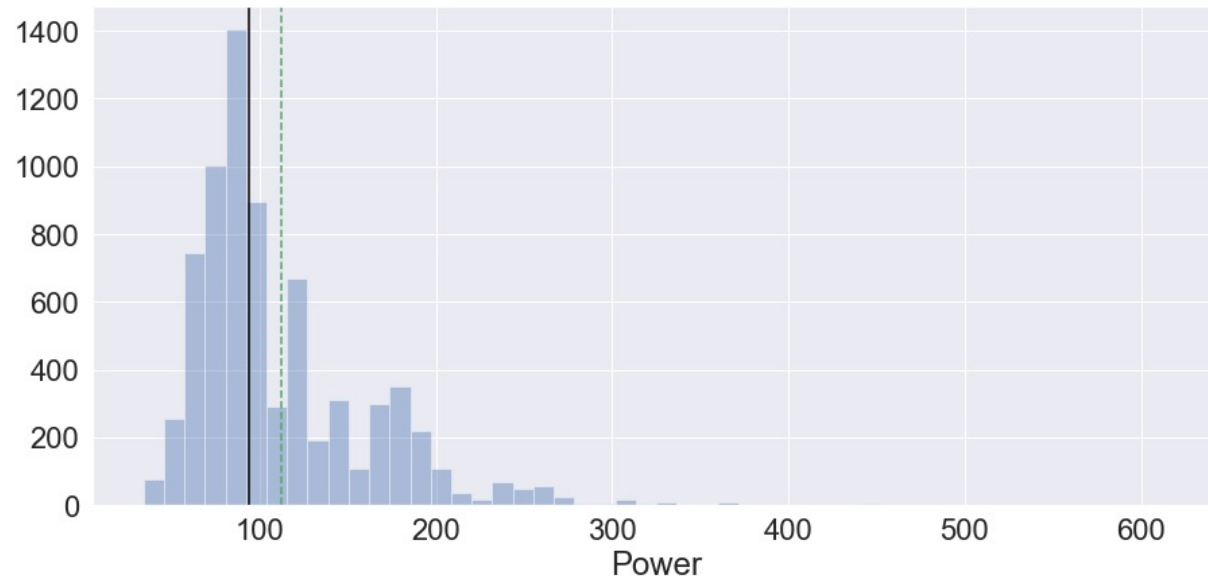
Kilometers\_Driven is right-skewed and there are many outliers.

- Range: 171-6,500,000 km
- Mean: 58,699.06 km
- Median: 53,416 km

# EDA - Power



Power



Power is positively skewed.

- Range: 34.2-616 bhp
- Mean: 112.31 bhp
- Median: 94 bhp

# Model Performance Summary

- Test data consisted of 30% of the data while Training data consisted of 70%.
- In final model, the following variables of statistical significance were included: Mileage, Year (categorical), Fuel Type (Categorical), Transmission (Categorical), Seats (Categorical), Brand (Categorical), Location (Categorical), Kilometers Driven (log scale), Power (log scale).
- Values for MAE and RMSE are comparable for Train and Test data.
- The model is able to explain 66% of the variation on the test set.

	Train	Test
MAE	3.31	3.19
MAPE	52.40	54.14
RMSE	6.23	5.71
R <sup>2</sup>	0.65	0.66

# Conclusions – Interpretation of Statistical Model

- Price is negatively correlated to years 2013 and prior. Price is positively correlated to years after 2013. This makes sense that the newer the model the higher the price.
- An increase of 1 kmpl makes the price decrease by 0.1396 Lakhs.
- Electric increases the price more than LPG and Diesel by, respectively, 13.7883 Lakhs, 2.7677 Lakhs, and 1.5415 Lakhs.
- A small increase in Price exists with Manual Transmission (0.1271 Lakhs).
- Being a second or third-time owner results in a decrease in price, but fourth+-time owners result in an increase in price by 0.8311 Lakhs.
- 2 or 4 seats result in an increase in price (with 2 seats being significant 17.2624 Lakhs), while the other numbers of seats result in a decrease in price.
- 18 brands result in a decrease in Price (Isuzu, Mahindra, Force, Tata, Datsun, Mitsubishi, Fiat, Honda, Renault, Jeep, Volkswagen, Hyundai, Chevrolet, Skoda, Ford, Toyota, Nissan, Hindustan). The range is 0 to -7.2518 Lakh.
- 3 brands result in a small increase in Price ranging from 0 to 2.952 Lakhs (Smart, OpelCorsa, and Volvo).
- 5 brands result in a larger increase in Price ranging from 15.0948 to 84.601 Lakhs (Jaguar, Land, Porsche, Bentley, and Lamborghini).
- An increase in 1 unit of log of Kilometers\_Driven decreases the price by 1.2520 Lakhs.
- An increase in 1 unit of log of Power increases the price by 7.6364 Lakhs. Customers are willing to pay for higher power.

# Recommendations

- Positive impact on price occurs if model year is 2013 or later. Invest in cars with these years of manufacture.
- Positive impact on price occurs with Electric fuel type. Research the feasibility of increasing the number of electric cars in your fleet, especially considering the environmental sustainability factor.
- Invest in car brands that increase price and minimize investment in car brands that decrease the price (see Conclusions above).
- Customers are willing to pay for cars with higher power.