

# Cars4U

## Used Car Price Predictive Model

---

May 2021

Brian Hall – DSBA – Project 3

03 | Problem & Objective

04 | Data Overview

05 | EDA Overview & Recommendations

06 | Model Performance & Conclusions

Appendix

08 | Statsmodel Final Model Comparison

11 | Uni-Variate Data Analysis

23 | Multi-Variate Data Analysis

# Problem

There is high demand for used cars in the Indian Market today and the market shows rapid growth, in 2018-19 second-hand cars sales were higher than new. Unfortunately, the used car market has huge uncertainty in both pricing and supply which makes it difficult to understand and grow as a business in this market.

# Objective

Cars4U needs a pricing model that can effectively predict the price of used cars. Ideally this model would help the business in devising profitable strategies using differential pricing, thus never selling a used car below market price.

**Question:** Can we build a reliable predictive model that will allow Cars4U to predict the prices of a specific used car (make, model, performance, age, location, etc.) within a reasonable range of certainty?

This will be accomplished through the following activities.

1. Exploration and visualization of the dataset
2. Build a linear regression model to predict the prices of used cars
3. Generate a set of insights and recommendations that will help the business

May 2021



# Data Overview

The data is assumed to be a random subset of used car sales in India from 1996 – 2019

Data has been provided via CSV (used\_cars\_data.csv - 768 kb)

## Variable Description

S.No	Serial Number
Name	Name of the car which includes Brand name and Model name
Location	The location in which the car is being sold or is available for purchase Cities
Year	Manufacturing year of the car
Kilometers_Driven	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	The type of transmission used by the car. (Automatic / Manual)
Owner	Type of ownership
Mileage	The standard mileage offered by the car company in kmpl or km/kg
Engine	The displacement volume of the engine in CC.
Power	The maximum power of the engine in bhp.
Seats	The number of seats in the car.
New_Price	The price of a new car of the same model in INR Lakhs.(1 Lakh = 100, 000) 14 Price : The price of the used car in INR Lakhs (1 Lakh = 100, 000)
Price	The price of the used car in INR Lakhs (1 Lakh = 100, 000)

## The Data Contains:

- 7253 rows and 13 columns (1 duplicate column dropped)
- Column Keys match data description provided
- 9 variables of type object
- 2 variables of type int64
- 2 variable of type float64
- Missing data | 7760 values total
  - New\_Price 86%
  - Price 17%
  - Seats < 1%
  - Engine < 1%
  - Power < 1%
  - Mileage 2 values

## Data Treatment (pre-model):

- New Price missing 86% of values and will be dropped
- New column of Age (age of the car from today) will be added based on Years
- Name will be split into Make, Model & Sub\_Class
- Units will be dropped from Fuel\_Type, Mileage, Engine & Power

# EDA Overview

7253 observations

Variable	Mean	Range
Price (Lakh)	9.50	0.44 - 160
Mileage (kmpl)	18.18	1 – 33.5
Engine (CC)	1617	72 - 5998
Power (bhp)	111	18 – 616
KM Driven	58,634	171 – 650,000

Variable	Unique	Top
Location	11	Mumbai
Make	33	Maruti
Transmission	2	Manual
Owner Type	4	First
Seats	10	10
Fuel Type	5	Diesel

Year (Top 2015)	1996 -2019
-----------------	------------

## Key Observations

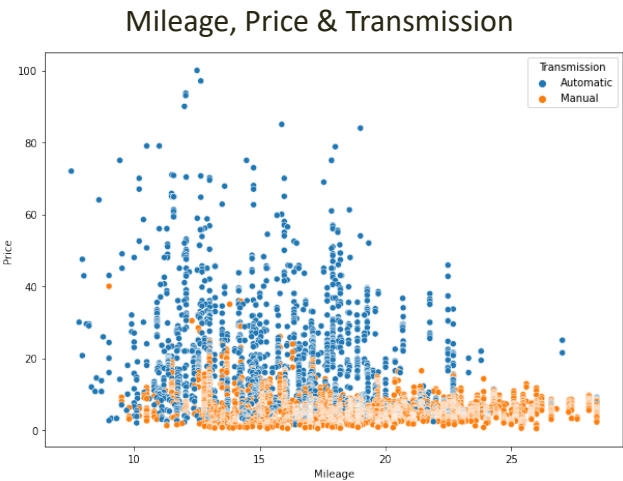
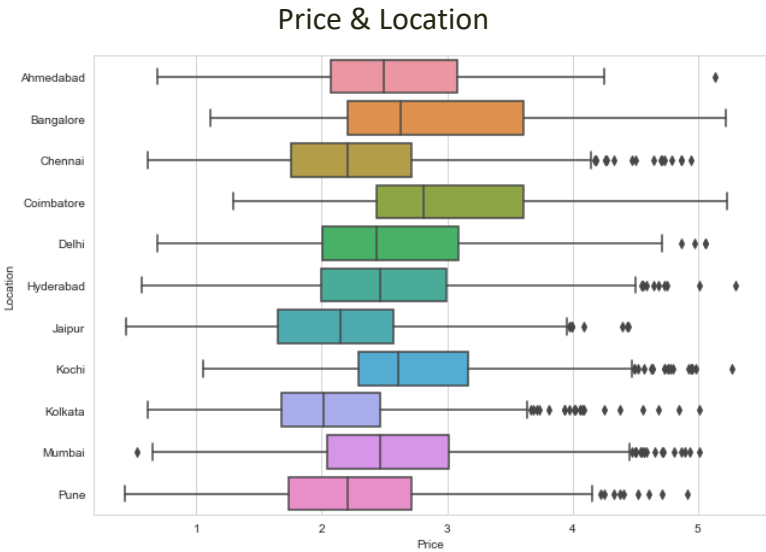
- **Maruti and Hyundai dominate the used car market with Honda and Toyota the closest with less than half the number of cars**
- **Most higher power and larger engines seem to be automatic transmission**
- For Engine there 62 outliers above 3000 CC, the highest of which are in high performing models. Keeping these in the model.
- There are 71 power values below 40 and are attributed to "lower end" cars, & 59 power values above 300 with the majority belonging to be to "higher end cars and engines," we will leave these in the model as they seem to fit
- For Prices above 100 Lakh the actual highest was 120 Lakh Lanbroghini, 1 Land Rover with 160 Lakh was imputed.
- Age has mild correlation to KM Driven (0.4)
- Automatic transmission make up the bulk of higher KM Driven values
- **The highest power and priced cars are Petrol fuel**
- There is a spike in Engines around 3000 cc and most are automatic and range across the price spectrum
- The majority of higher mileage cars (> 20) have a manual transmission and a lower cost
- Median age across locations range from 5 to ~7 years.
- Kochi and Coimbatore are showing the 'youngest' cars with a median age of 5.
- Engine sizes are fairly stable across locations at a mean around 7.25

## Recommendations

Mumbai and Hyberdad are where most of the used cars are located and are fairly evenly distributed across attributes. **These locations should be investigated in more detail to determine approach**

Combatore is 4th in the number of cars, more 'younger' cars, the widest range of engine sizes and the highest prices. Kolkata & Jaipur have the lowest prices. **Consider finding cars in Koldata & Jaipur and selling in Combatore.**

**Investigate why Combatore has such an interesting dynamic and how to capitalize on it.**



# Model Performance

By developing a Machine Learning model that is based on historical data we can predict with a range of certainty what the prices of a specific car (make, model, performance, age, location, etc.) is in the current market.

12 Models were tested, most were iterations looking at outlier treatment, transformations and investigating problems with assumptions. Most of these are not shown, however they are documented in the accompanying Jupyter Notebook

**OLSRES4** was the highest performing model that met linear regression assumptions and showed a good overall fit with the training data

There were 13 columns that resulted in 56 features going into modeling

The final model contains 49 features plus the target variable of Price

Surprisingly:

- Kilometers\_Driven and Location showed very little significance in the overall performance of the model. However, these should not be removed from the model as more data could increase their significance and they could be important for ongoing insights.
- Engine, Power & Mileage showed high correlation initially, however not in the model. Removing any combination only impacted the model by 1-2%

**Although this model shows good results some assumptions such as normality need to be further worked through adding additional data and other methods to increase the reliability of making inferences from the model.**

## Conclusion

Overall **OLSRES4** Shows very good results in both  $R^2$  and MAPE and will be able to:

1. Explain **92%** of the data's variation
2. Predict within **8%** of the used car Price

MAE	0.177	<b>Key Variables for the model:</b>
MAPE	8.163	Mileage
RMSE	0.242	Engine
R2	0.923	Power
Adj R2	0.921	Age

**Utilizing this model could help Cars4U be successful by:**

1. Ensuring that cars are not selling below the market price
2. Targeting cars to bring into inventory, ie. Buy for lower than predicted market cost and sell above predicted cost
3. Set competitive pricing & messaging for marketing purposes
4. Begin to determine which types of cars could bring the most profit in specific locations. This could be for shifting of inventory, optimizing operations, marketing, etc.

See Appendix for full Statsmodel comparison of training and test results

**Note:** Initially Make and Fuel\_Type were excluded from the final model due to VIF values being infinite. Results from that model were  $R^2$  (86%) & MAPE (10.8%), while these results are not as good as OLSRES4 it used only 16 features & linear regression assumptions could be assumed to be met. Make and Fuel\_Type were added back into OLSRES4 as a result of guidance from Great Learning and there was no time to further explore assumptions to achieve better results.

# Appendix

---

**Cars4U**

May 2021

## Statsmodel Final Model – OLSRES4

### Train & Test Results

---

**Cars4U**

May 2021



# OLSRES4 Statsmodel – Train Data

```
olsmod4 = sm.OLS(y_train, X_train5)
olsres4 = olsmod4.fit()
```

```
print(olsres4.summary())
```

executed in 62ms, finished 17:47:36 2021-05-21

## OLS Regression Results

```
=====
Dep. Variable:      Price      R-squared:      0.932
Model:              OLS      Adj. R-squared:    0.931
Method:             Least Squares      F-statistic:    1273.
Date:               Fri, 21 May 2021      Prob (F-statistic):    0.00
Time:               17:47:36      Log-Likelihood:    397.61
No. Observations:   4132      AIC:      -705.2
Df Residuals:       4087      BIC:      -420.5
Df Model:           44
Covariance Type:    nonrobust
=====
```

	coef	std err	t	P> t	[0.025	0.975]
const	-1.8564	0.135	-13.752	0.000	-2.121	-1.592
Kilometers_Driven	-0.0009	7.02e-05	-12.741	0.000	-0.001	-0.001
Mileage	-0.0145	0.002	-8.388	0.000	-0.018	-0.011
Engine	0.8046	0.029	27.416	0.000	0.747	0.862
Power	0.2148	0.015	14.738	0.000	0.186	0.243
Seats	0.0131	0.007	1.968	0.049	5.28e-05	0.026
Age	-0.1121	0.002	-68.527	0.000	-0.115	-0.109
Make_Audi	0.3908	0.021	18.750	0.000	0.350	0.432
Make_BMW	0.3924	0.021	18.827	0.000	0.352	0.433
Make_Chevrolet	-0.5977	0.025	-24.108	0.000	-0.646	-0.549
Make_Datsun	-0.6789	0.066	-10.331	0.000	-0.808	-0.550
Make_Fiat	-0.4716	0.050	-9.495	0.000	-0.569	-0.374
Make_Ford	-0.4038	0.018	-23.017	0.000	-0.438	-0.369
Make_Honda	-0.2956	0.016	-18.948	0.000	-0.326	-0.265
Make_Hyundai	-0.2786	0.013	-21.735	0.000	-0.304	-0.253
Make_Isuzu	1.169e-15	6.78e-17	17.244	0.000	1.04e-15	1.3e-15
Make_Jaguar	0.4990	0.044	11.376	0.000	0.413	0.585
Make_Jeep	-0.0051	0.060	-0.084	0.933	-0.123	0.113
Make_Land_Rover	0.5635	0.035	16.218	0.000	0.495	0.632
Make_Mahindra	-0.4876	0.022	-22.665	0.000	-0.530	-0.445
Make_Maruti	-0.2377	0.014	-17.220	0.000	-0.265	-0.211
Make_Mercedes-Benz	0.3831	0.019	20.075	0.000	0.346	0.421
Make_Mini_Cooper	0.7366	0.049	14.968	0.000	0.640	0.833
Make_Mitsubishi	-0.1855	0.052	-3.573	0.000	-0.287	-0.084

Make_Nissan	-0.3971	0.028	-14.408	0.000	-0.451	-0.343
Make_Porsche	0.8582	0.071	12.035	0.000	0.718	0.998
Make_Renault	-0.3144	0.024	-13.079	0.000	-0.362	-0.267
Make_Skoda	-0.2975	0.022	-13.323	0.000	-0.341	-0.254
Make_Smart	7.318e-16	7.76e-17	9.429	0.000	5.8e-16	8.84e-16
Make_Tata	-0.7159	0.021	-34.715	0.000	-0.756	-0.675
Make_Toyota	-0.1486	0.019	-7.703	0.000	-0.186	-0.111
Make_Volkswagen	-0.3523	0.017	-20.401	0.000	-0.386	-0.318
Make_Volvo	0.1878	0.056	3.356	0.001	0.078	0.298
Location_Bangalore	0.1329	0.023	5.834	0.000	0.088	0.178
Location_Chennai	0.0184	0.022	0.847	0.397	-0.024	0.061
Location_Coimbatore	0.0830	0.021	3.953	0.000	0.042	0.124
Location_Delhi	-0.0910	0.021	-4.310	0.000	-0.132	-0.050
Location_Hyderabad	0.0968	0.020	4.743	0.000	0.057	0.137
Location_Jaipur	-0.0498	0.022	-2.235	0.025	-0.093	-0.006
Location_Kochi	-0.0348	0.021	-1.667	0.096	-0.076	0.006
Location_Kolkata	-0.2570	0.021	-12.029	0.000	-0.299	-0.215
Location_Mumbai	-0.0686	0.020	-3.350	0.001	-0.109	-0.028
Location_Pune	-0.0477	0.021	-2.277	0.023	-0.089	-0.007
Fuel_Type_Diesel	-0.8291	0.072	-11.558	0.000	-0.970	-0.688
Fuel_Type_Petrol	-1.0274	0.064	-16.153	0.000	-1.152	-0.903
Owner_Type_Fourth & Above	0.1182	0.129	0.916	0.360	-0.135	0.371
Owner_Type_Second	-0.0671	0.010	-6.665	0.000	-0.087	-0.047
Owner_Type_Third	-0.0810	0.026	-3.088	0.002	-0.132	-0.030
Transmission_Manual	-0.1227	0.011	-10.815	0.000	-0.145	-0.100

```
=====
Omnibus:                222.122      Durbin-Watson:                1.963
Prob(Omnibus):           0.000      Jarque-Bera (JB):              855.228
Skew:                    -0.074      Prob(JB):                      1.95e-186
Kurtosis:                 5.224      Cond. No.                      4.64e+16
=====
```

### Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.  
 [2] The smallest eigenvalue is 1.1e-25. This might indicate that there are strong multicollinearity problems or that the design matrix is singular.

# OLSRES4 Statsmodel – Test Data

```
olsmod4 = sm.OLS(y_test, X_test)
olsres4 = olsmod4.fit()
print(olsres4.summary())
```

executed in 60ms, finished 17:47:41 2021-05-21

## OLS Regression Results

```
=====
Dep. Variable:      Price      R-squared:      0.923
Model:              OLS      Adj. R-squared:    0.921
Method:             Least Squares      F-statistic:    468.4
Date:              Fri, 21 May 2021      Prob (F-statistic): 0.00
Time:              17:47:40      Log-Likelihood: 30.659
No. Observations:  1772      AIC:          28.68
Df Residuals:      1727      BIC:          275.3
Df Model:          44
Covariance Type:    nonrobust
=====
```

```
=====
               coef      std err          t      P>|t|      [0.025      0.975]
-----
const          -2.0536      0.231      -8.902      0.000      -2.506      -1.601
Kilometers_Driven -0.0010      0.000      -8.430      0.000      -0.001      -0.001
Mileage         -0.0097      0.003      -3.341      0.001      -0.015      -0.004
Engine          0.7752      0.049     15.727      0.000      0.679      0.872
Power           0.2854      0.026     11.179      0.000      0.235      0.335
Seats           0.0242      0.011      2.208      0.027      0.003      0.046
Age            -0.1068      0.003     -40.167      0.000      -0.112      -0.102
Make_Audi        0.3627      0.034     10.806      0.000      0.297      0.429
Make_BMW         0.3922      0.036     11.045      0.000      0.323      0.462
Make_Bentley    -1.259e-15      2.42e-16      -5.212      0.000      -1.73e-15      -7.85e-16
Make_Chevrolet  -0.5695      0.044     -13.009      0.000      -0.655      -0.484
Make_Datsun     -0.5758      0.166      -3.459      0.001      -0.902      -0.249
Make_Fiat       -0.4654      0.096      -4.852      0.000      -0.653      -0.277
Make_Force     -4.802e-16      5.67e-17      -8.473      0.000      -5.91e-16      -3.69e-16
Make_Ford       -0.3878      0.031     -12.656      0.000      -0.448      -0.328
Make_Hindustan  6.769e-16      1.84e-16      3.671      0.000      3.15e-16      1.04e-15
Make_Honda     -0.3134      0.027     -11.727      0.000      -0.366      -0.261
Make_Hyundai    -0.2974      0.022     -13.389      0.000      -0.341      -0.254
Make_ISUZU      1.03e-16      8.28e-17      1.245      0.213      -5.93e-17      2.65e-16
Make_Isuzu      1.93e-16      9.14e-17      2.111      0.035      1.37e-17      3.72e-16
Make_Jaguar      0.5221      0.068      7.666      0.000      0.389      0.656
Make_Jeep       -0.0335      0.165      -0.203      0.839      -0.358      0.291
Make_Lamborghini 2.783e-16      3.13e-16      0.888      0.375      -3.36e-16      8.93e-16
Make_Land_Rover 0.7179      0.068     10.562      0.000      0.585      0.851
Make_Mahindra   -0.5236      0.038     -13.770      0.000      -0.598      -0.449
Make_Maruti     -0.2654      0.024     -10.997      0.000      -0.313      -0.218
Make_Mercedes-Benz 0.3846      0.032     11.957      0.000      0.322      0.448
Make_Mini_Cooper 0.6674      0.097      6.868      0.000      0.477      0.858
Make_Mitsubishi -0.1906      0.081      -2.351      0.019      -0.350      -0.032
=====
```

```
Make_Nissan      -0.3500      0.052     -6.676      0.000      -0.453      -0.247
Make_OpelCorsa   3.384e-16      1.41e-16      2.395      0.017      6.13e-17      6.16e-16
Make_Porsche      0.4265      0.089      4.768      0.000      0.251      0.602
Make_Renault     -0.3247      0.037     -8.882      0.000      -0.396      -0.253
Make_Skoda       -0.2496      0.036     -7.029      0.000      -0.319      -0.180
Make_Smart       -2.902e-17      5.99e-17      -0.485      0.628      -1.46e-16      8.84e-17
Make_Tata        -0.6703      0.039     -17.316      0.000      -0.746      -0.594
Make_Toyota      -0.1755      0.033     -5.374      0.000      -0.240      -0.111
Make_Volkswagen  -0.3678      0.027     -13.400      0.000      -0.422      -0.314
Make_Volvo       0.2333      0.098      2.388      0.017      0.042      0.425
Location_Bangalore 0.1633      0.039      4.204      0.000      0.087      0.239
Location_Chennai 0.0471      0.036      1.293      0.196      -0.024      0.119
Location_Coimbatore 0.1300      0.035      3.690      0.000      0.061      0.199
Location_Delhi   -0.0522      0.036     -1.451      0.147      -0.123      0.018
Location_Hyderabad 0.1432      0.035      4.147      0.000      0.075      0.211
Location_Jaipur  -0.0253      0.037     -0.677      0.499      -0.099      0.048
Location_Kochi   -0.0210      0.035     -0.591      0.555      -0.091      0.049
Location_Kolkata -0.2131      0.036     -5.893      0.000      -0.284      -0.142
Location_Mumbai  -0.0707      0.034     -2.074      0.038      -0.137      -0.004
Location_Pune    -0.0375      0.036     -1.051      0.293      -0.107      0.032
Fuel_Type_Diesel -0.9407      0.122     -7.681      0.000      -1.181      -0.701
Fuel_Type_Electric 0          0          nan          nan          0          0
Fuel_Type_LPG    0          0          nan          nan          0          0
Fuel_Type_Petrol -1.1129      0.109     -10.232      0.000      -1.326      -0.900
Owner_Type_Fourth & Above -0.0913      0.110     -0.829      0.407      -0.307      0.125
Owner_Type_Second -0.0658      0.018     -3.744      0.000      -0.100      -0.031
Owner_Type_Third -0.1158      0.050     -2.313      0.021      -0.214      -0.018
Transmission_Manual -0.1197      0.019     -6.187      0.000      -0.158      -0.082
=====
```

```
Omnibus:          893.429      Durbin-Watson:          2.020
Prob(Omnibus):    0.000      Jarque-Bera (JB):      34668.379
Skew:             -1.687      Prob(JB):              0.00
Kurtosis:         24.405      Cond. No.              7.27e+16
=====
```

### Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.  
 [2] The smallest eigenvalue is 1.93e-26. This might indicate that there are strong multicollinearity problems or that the design matrix is singular.

# Uni-Variate Analysis

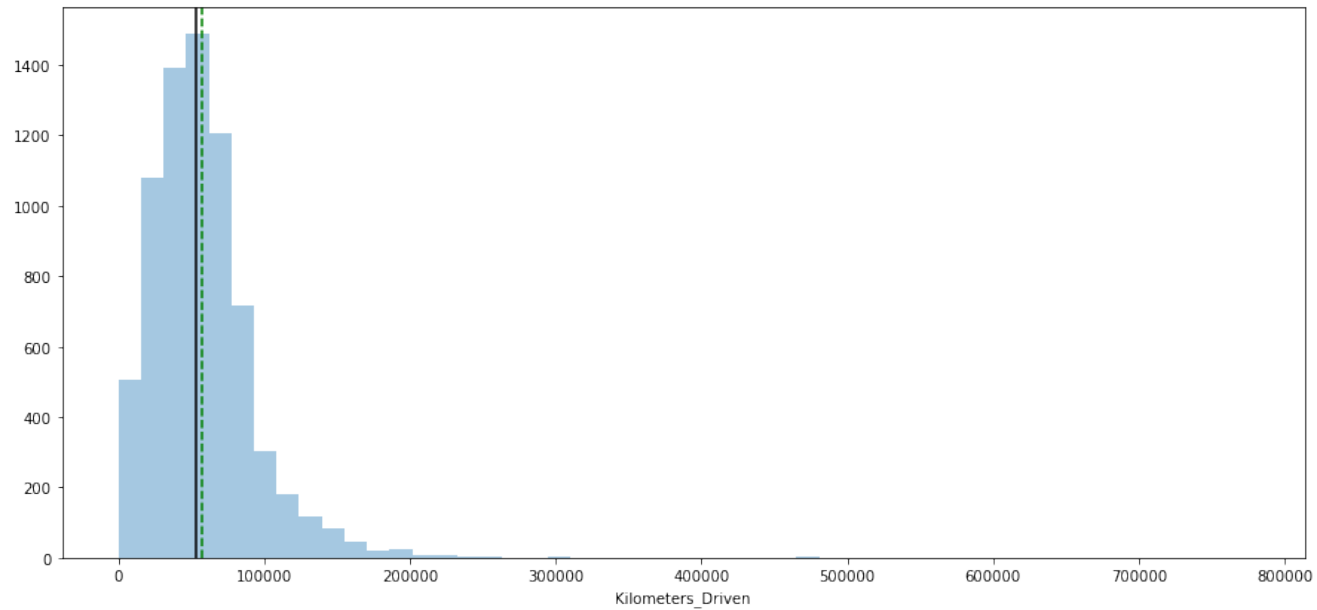
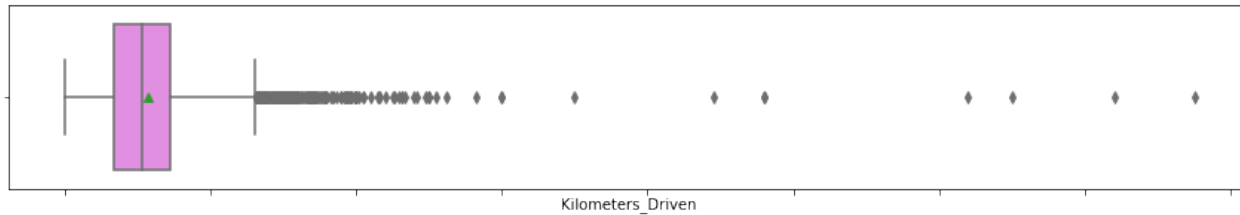
---

**Cars4U**

May 2021

# Univariate Analysis

## Kilometers\_Driven

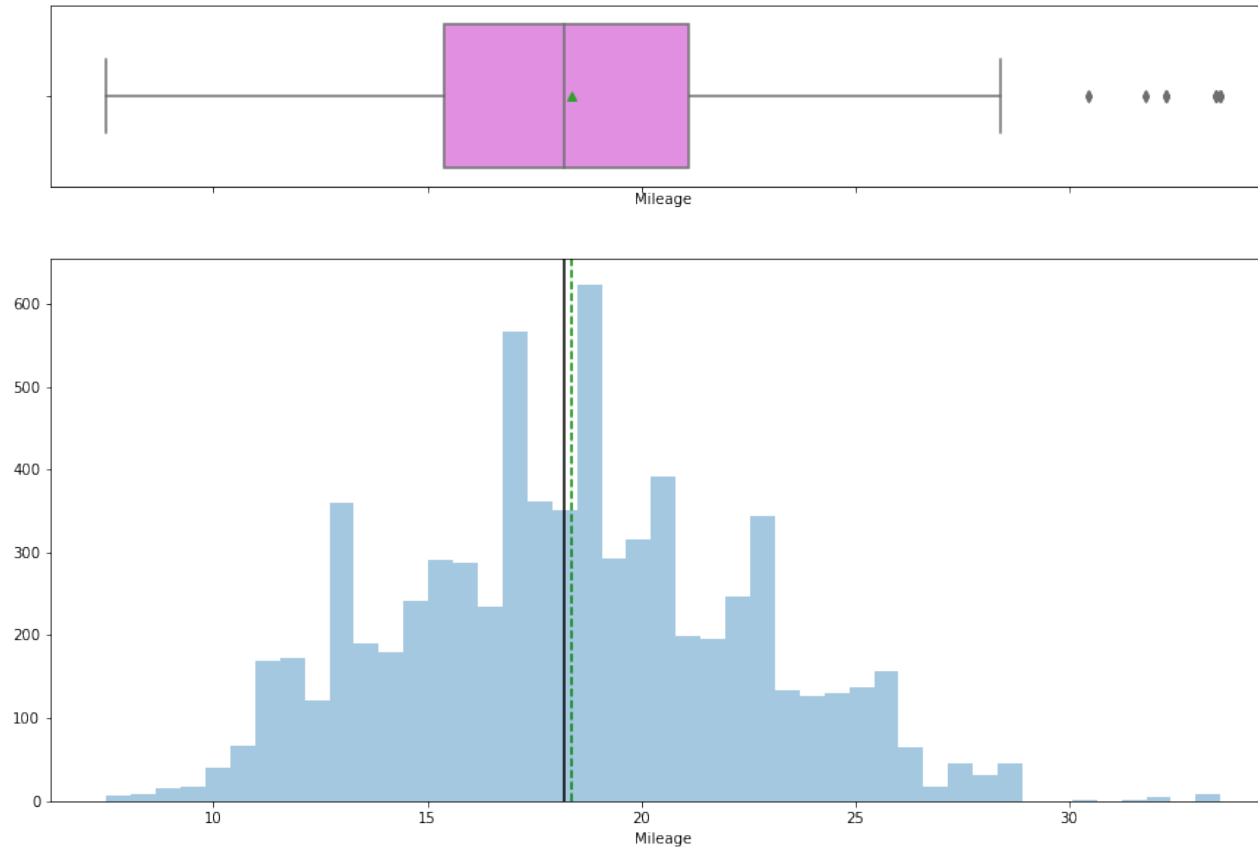


## Observations

- KM Driven is highly right skewed with many high outliers
- One outlier at 6.5 million changed to 650,000
- Most values are between 20,000 & 100,000

# Univariate Analysis

## Mileage

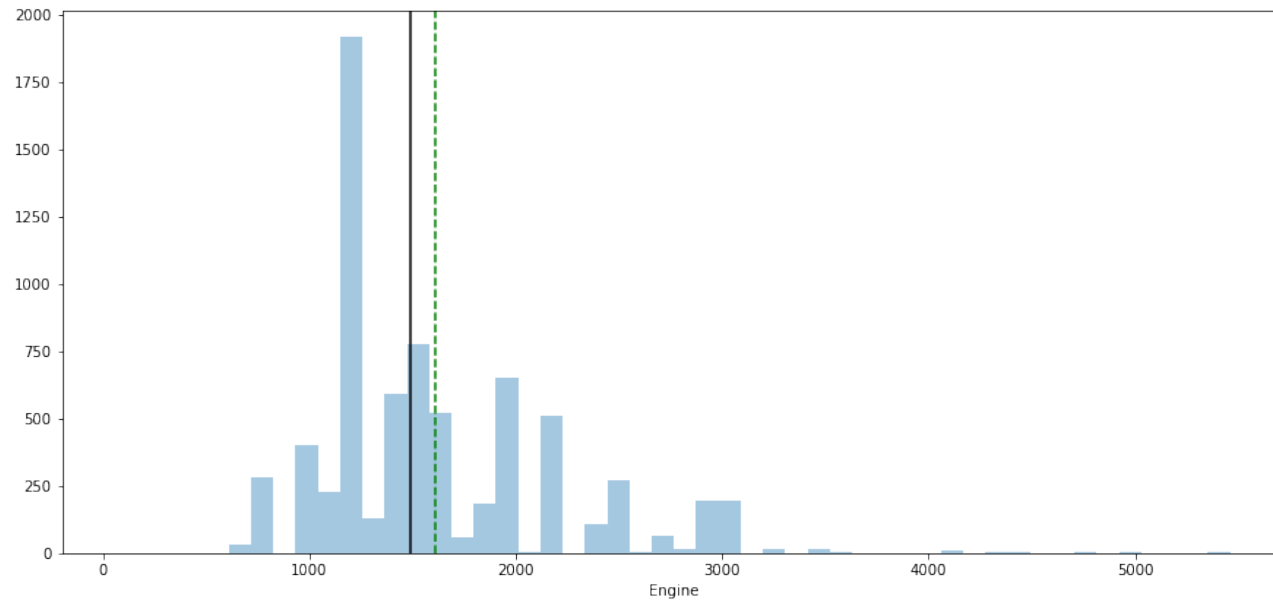
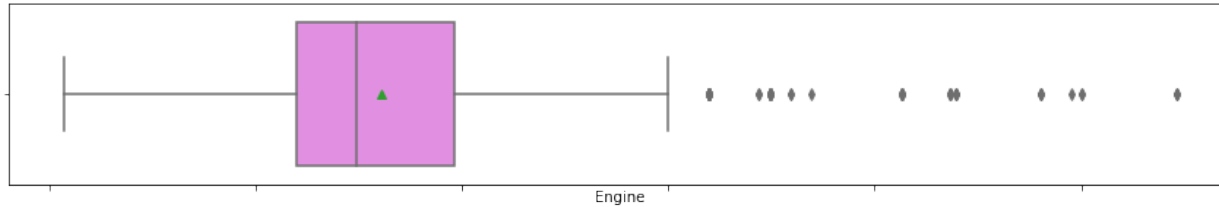


## Observations

- Distribution after 80 values of were imputed
- Mileage seems to have a fairly normal distribution with a few outliers.
- Mean and median are approximately 18
- Most values fall between 13 & 23

# Univariate Analysis

## Engine

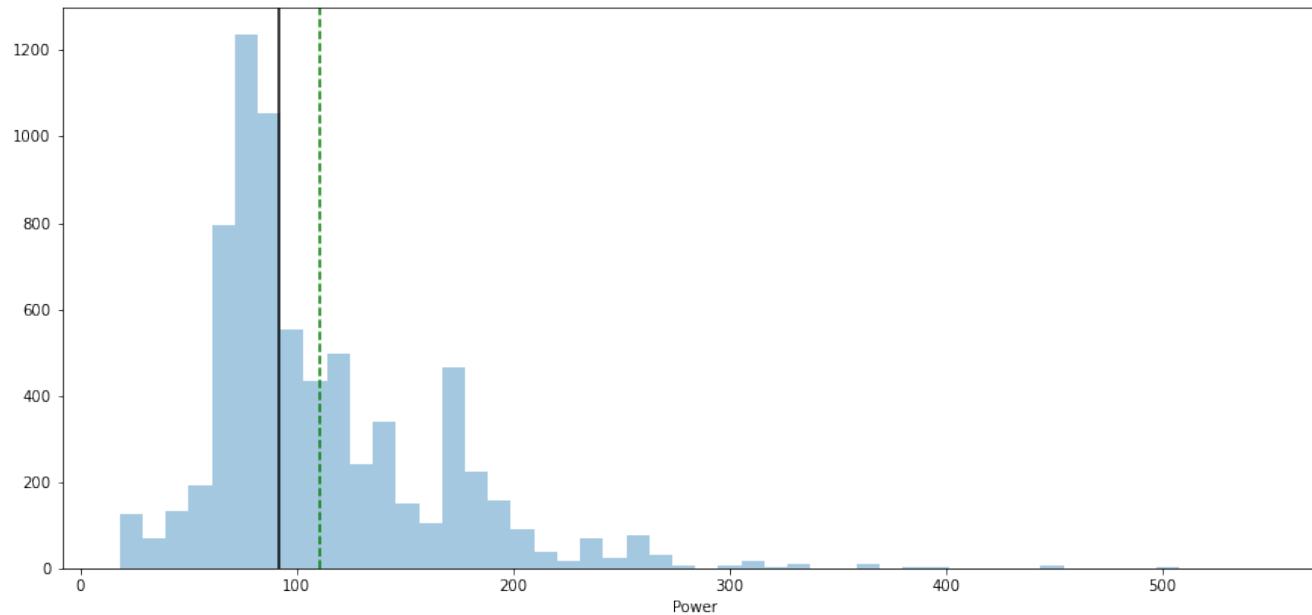
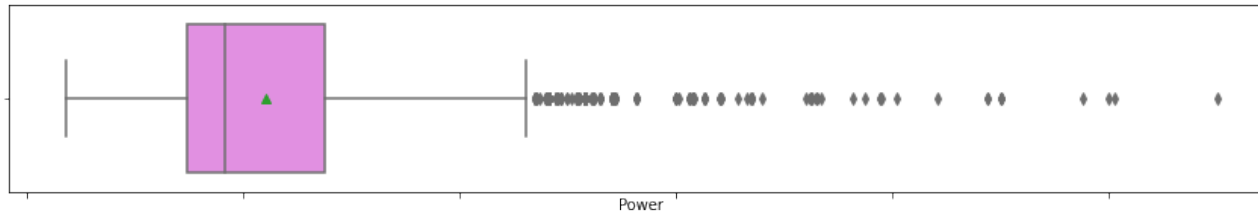


## Observations

- Engine is slightly normal with many outliers above 3000.
- Check the higher outliers for patterns and treatment
- Median and mean are relatively close around 1500
- There is a spike in engine CC's around 1200

# Univariate Analysis

## Power

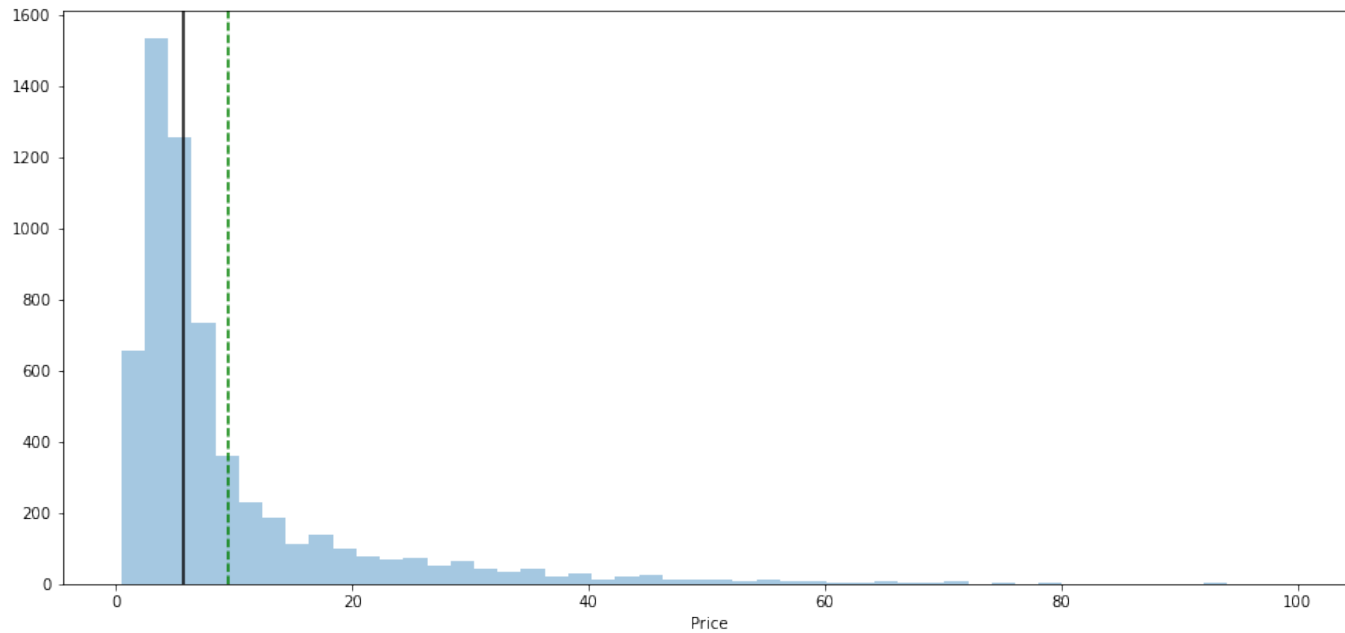
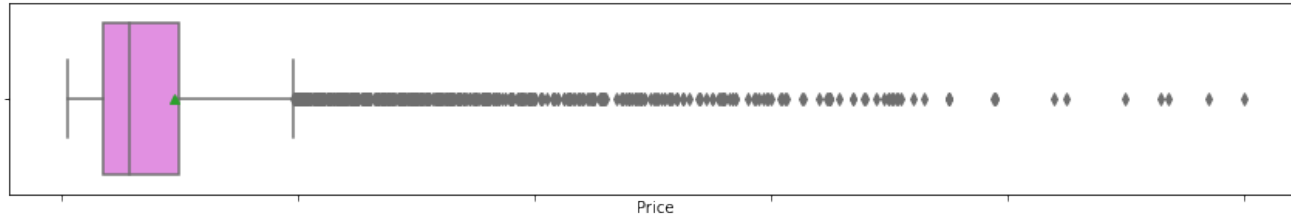


## Observations

- Power is right skewed with many outliers above ~225 which are pulling the mean higher
- Mean and median are relatively close and there appears to be close to a normal distribution
- There are a number of power values that are below 50, this should be checked
- The majority of power values are between 75 and 175

# Univariate Analysis

## Price



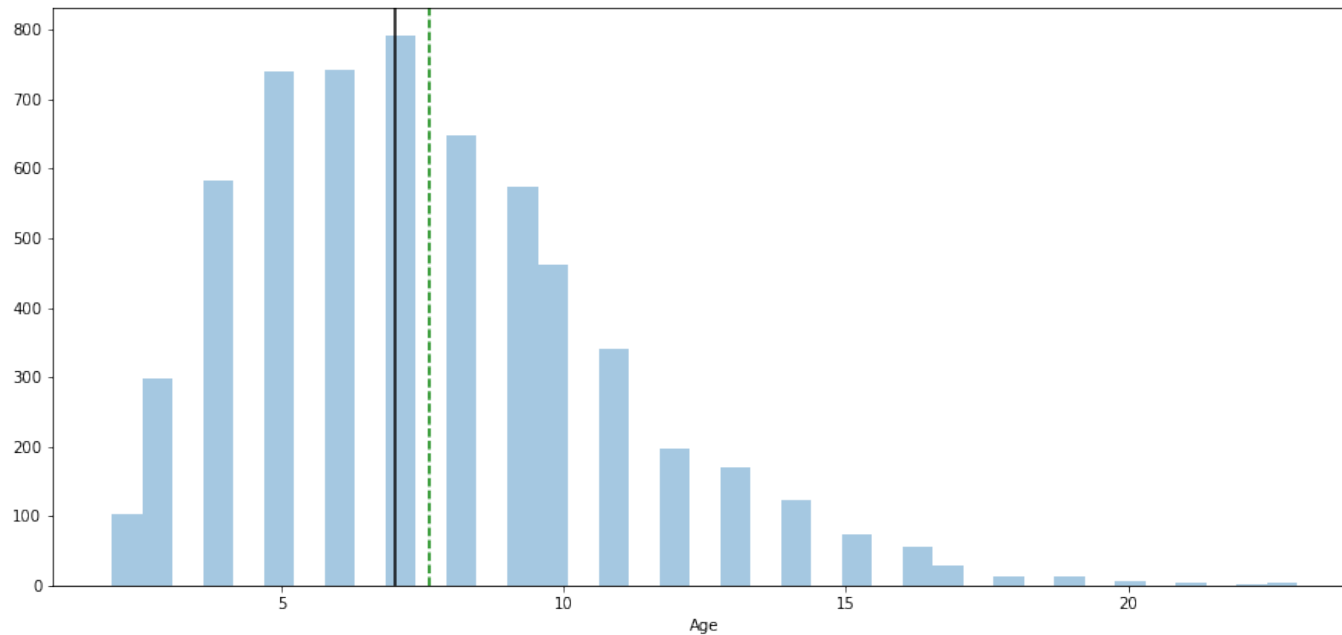
## Observations

- Price is highly right skewed with a huge amount of high outliers above 20 Lakh
- One outlier at 160 Lakh for a Land Rover is likely an entry mistake and was set to the median
- There are no other prices above 100 Lakh (7 observation were removed, 1 was a 120 Lakh Lanbroghini)
- There are 75 prices below 1 Lakh - none with 0
- There are 1221 prices that are missing with no apparent pattern



# Univariate Analysis

## Age



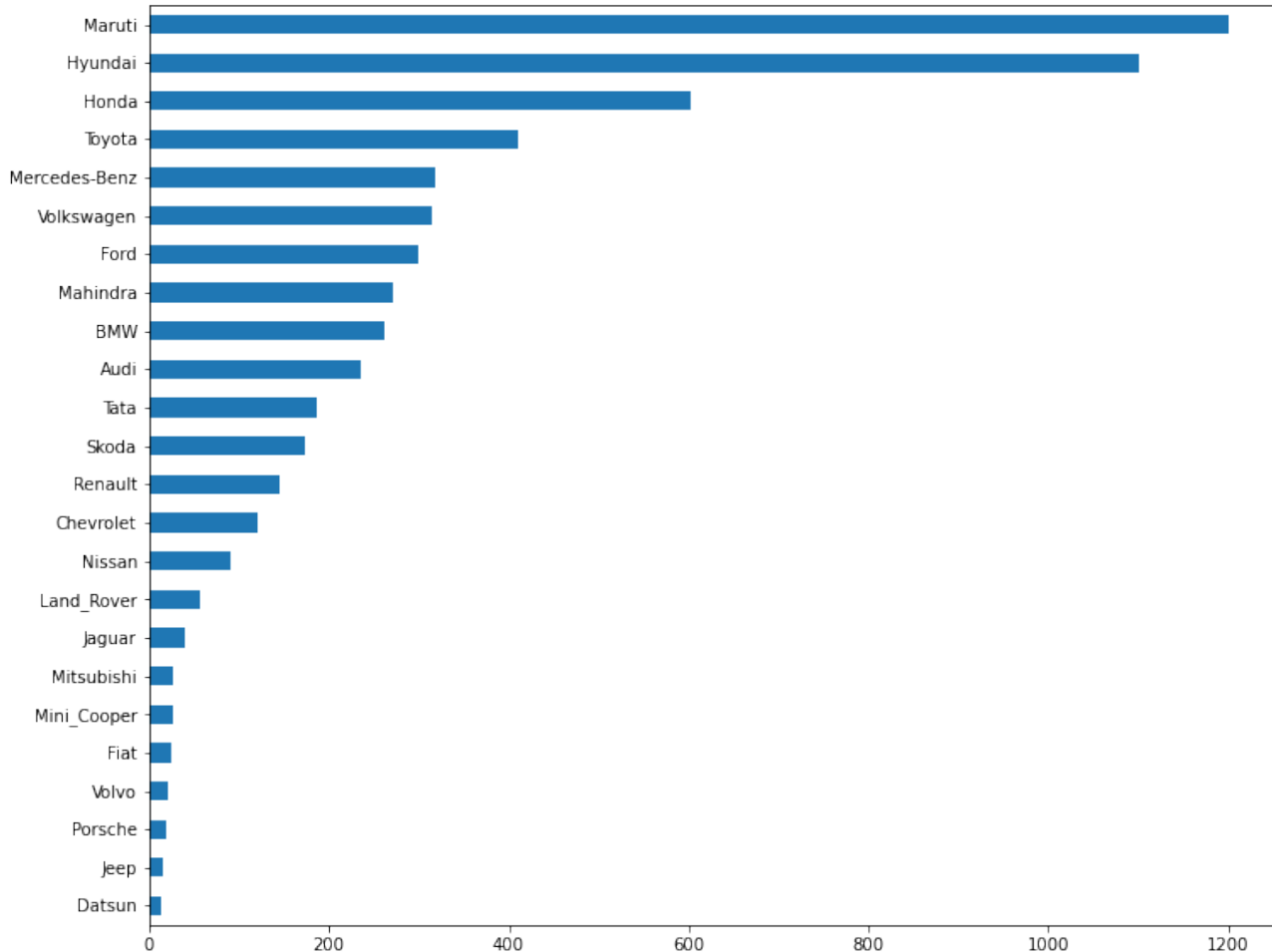
## Observations

- Age is normally distributed with a slight right skew
- Age has few outliers all above 18 years
- Age range is 2 to 23 years with an average of 7.6 years
- Age has no missing values

# Univariate Analysis

Make

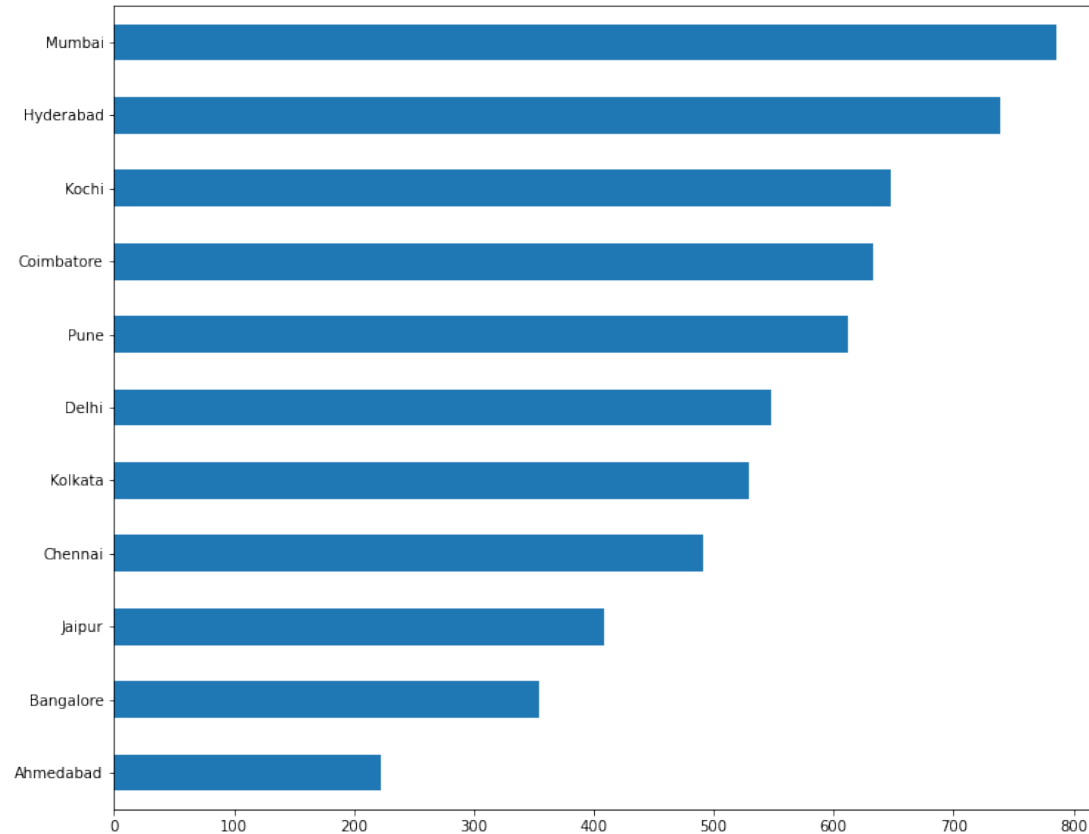
Observations



- There are 24 unique Makes of cars on the market (15 rows on 9 makes dropped, makes with  $\leq 3$  observations)
- Maruti and Hyundai dominate the used car market with Honda and Toyota the closest with less than half the number of cars
- There are no missing values

# Univariate Analysis

## Location

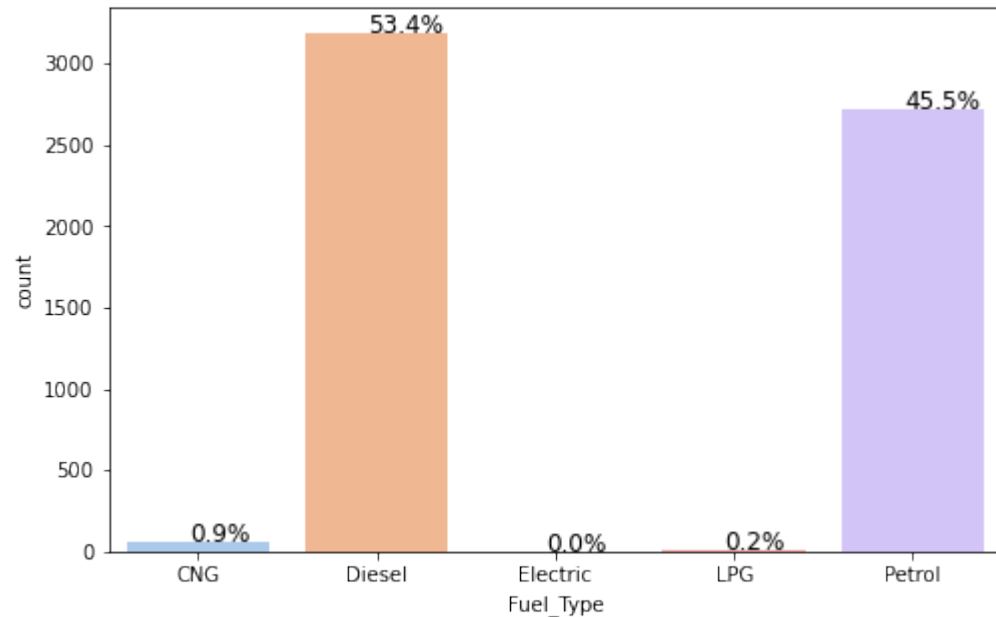


## Observations

- There are 11 unique locations with used cars for sale
- Mumbai is the location with the most used cars (945)
- Location has no missing values

# Univariate Analysis

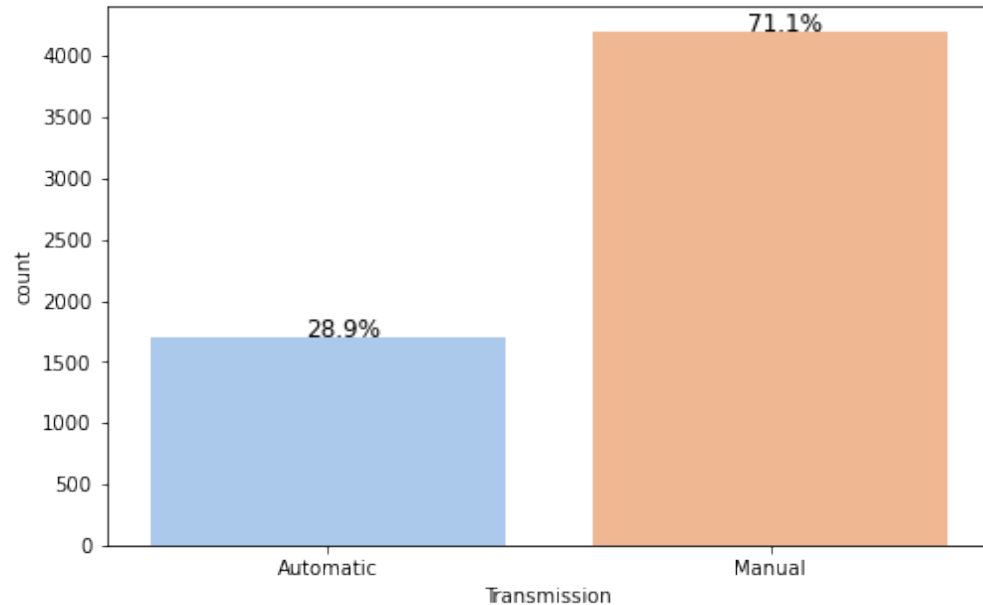
## Fuel\_Type



## Observations

- There are 4 unique Fuel Types
- Diesel is the largest Fuel Type at 53%, Followed by Petrol at 46%
- CNG, Electric & LPG combined make up about 1%

## Transmission

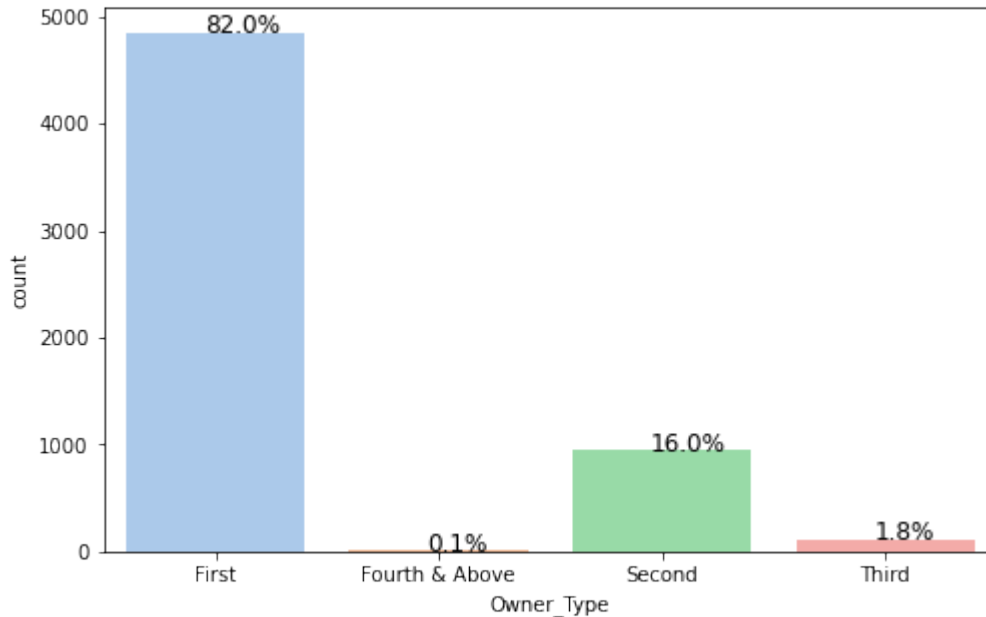


## Observations

- There are 2 types of transmission - manual and automatic
- Manual is the largest type at 71.5%
- There are no missing values

# Univariate Analysis

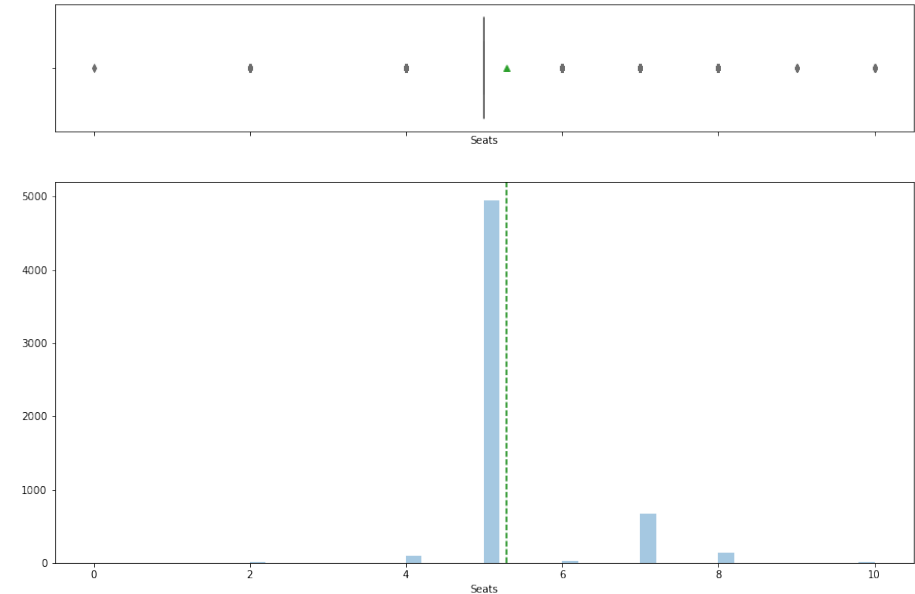
## Owner\_Type



## Observations

- There are 4 types of owner
- First (owned once) is by far the most prevalent with 82.3%
- There are no missing values

## Seats

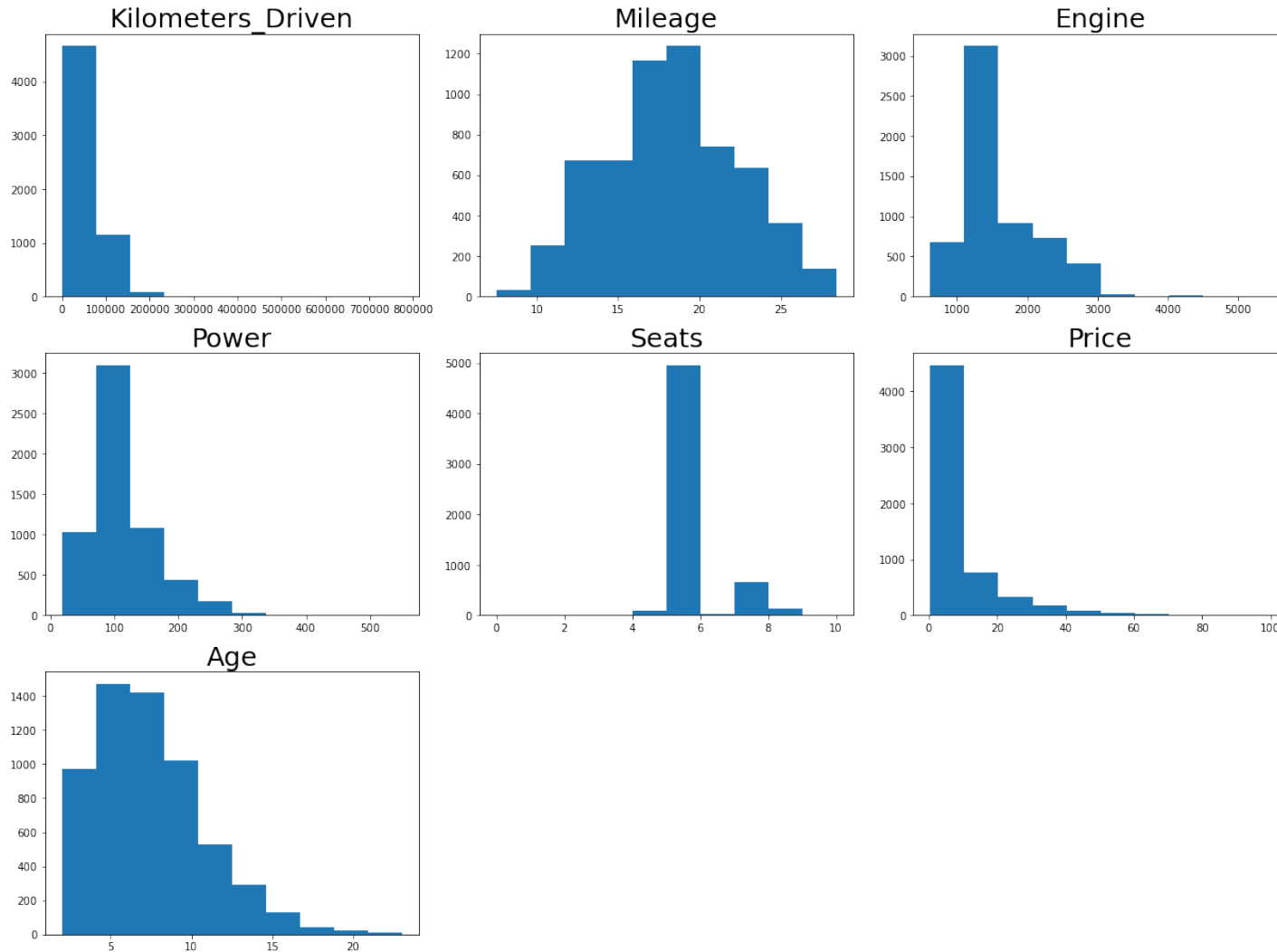


## Observations

- There are 9 types of seats
- 5 Seats is by far the most common with 84% with 7 seats following with 11%
- 0 & 9 seats have 0 observations
- There are 7 missing values that need to be imputed to 5 seats

# Univariate Analysis

## Distribution Overview



## Observations

- Kilometers Driven has many high outliers, even with removing some extreme. CAP OUTLIERS and recheck distribution. Rescale.
- Mileage appears normal and scale is likely OK
- Engine has some high outliers, even after removing some extreme values. CAP OUTLIERS and recheck distribution. Rescale.
- Power has some high outliers, CAP OUTLIERS and recheck distribution. Rescale.
- Price has many outliers, CAP OUTLIERS at a higher IQR. Rescale
- Age is fairly normal and scale should be OK

# Multi-Variate Analysis

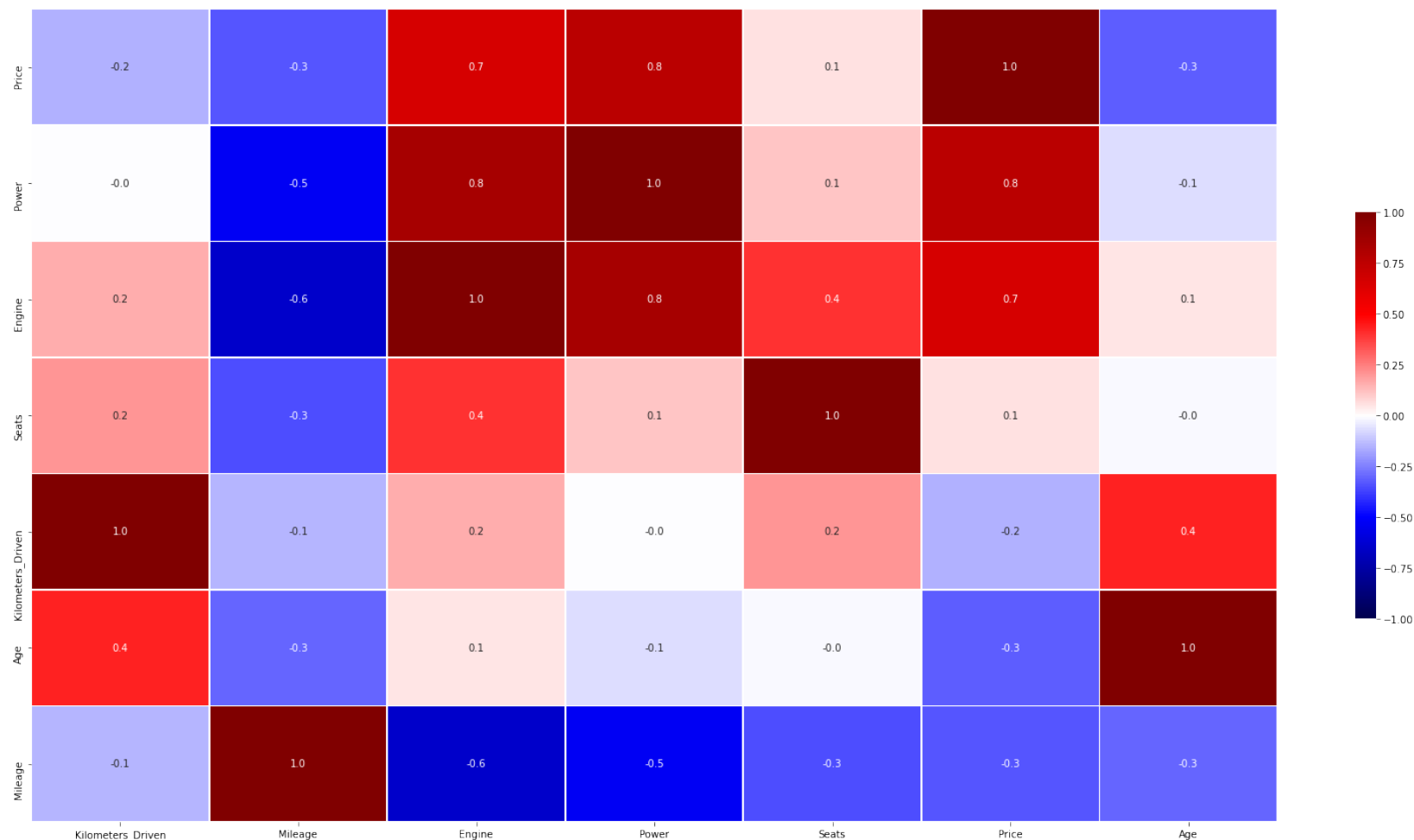
---

**Cars4U**

May 2021

# Multi-Variate Analysis

## Correlation Heat Map



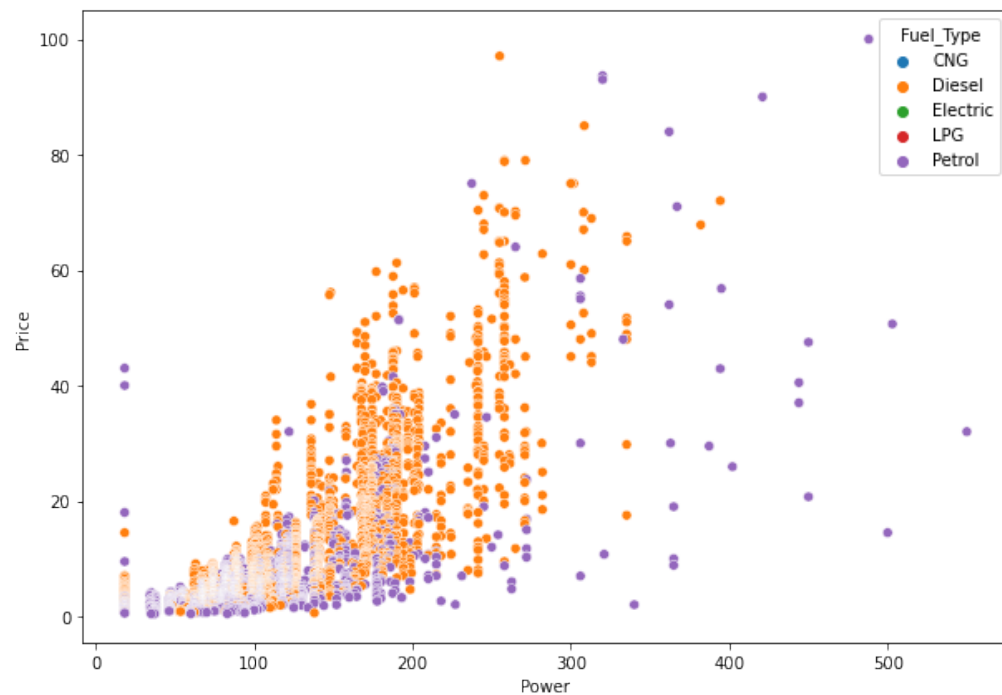
## Observations

- Power (0.8) and Engine (0.7) have the highest correlation to price. All other numeric variables are  $< 0.3$  or  $-0.3$
- KM Driven and Age are mildly correlated
- Engine, Power & Mileage have high correlation as would be expected. These are top candidates for dropping once building models
- Seats have very low correlation to any other variable
- Age has mild correlation to KM Driven (0.4) - A higher correlation was expected, see how they perform in the model
- Transmission seems to play a role in mileage and power, with mileage being somewhat negatively impacted by manual trans
- Most higher power and larger engines seem to be automatic transmission
- Automatic transmission make up the bulk of higher KM Driven values



# Multi-Variate Analysis

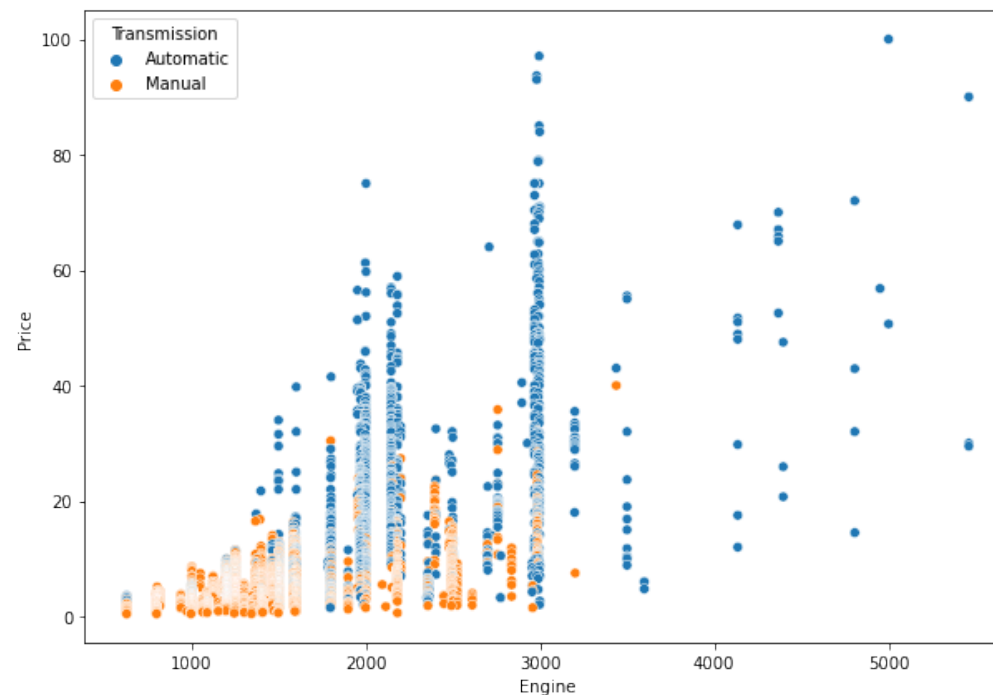
## Price & Power



## Observations

- The highest power and priced cars use Petrol fuel

## Price & Engine

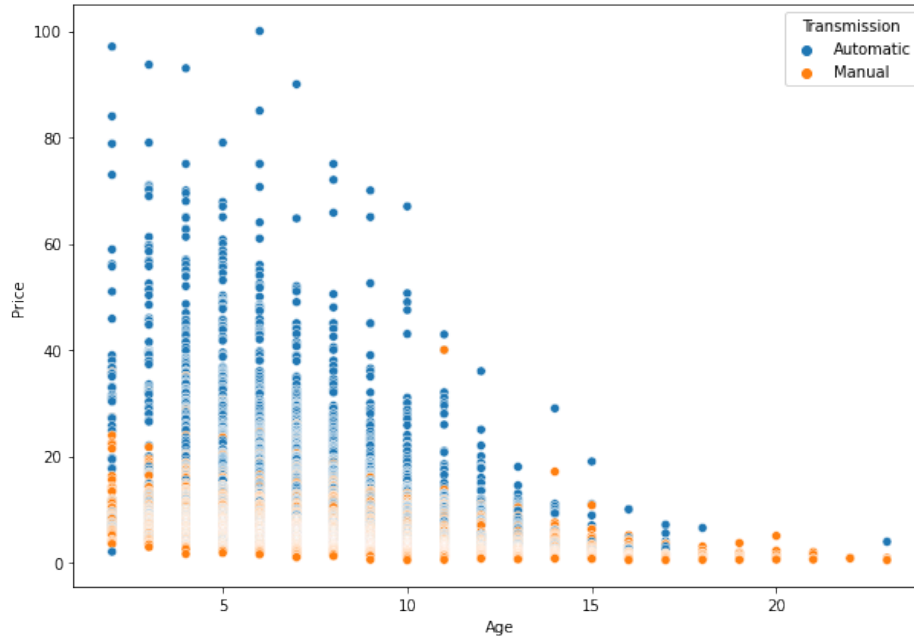


## Observations

- Engines around 3000 cc are mostly automatic and range across the price spectrum

# Multi-Variate Analysis

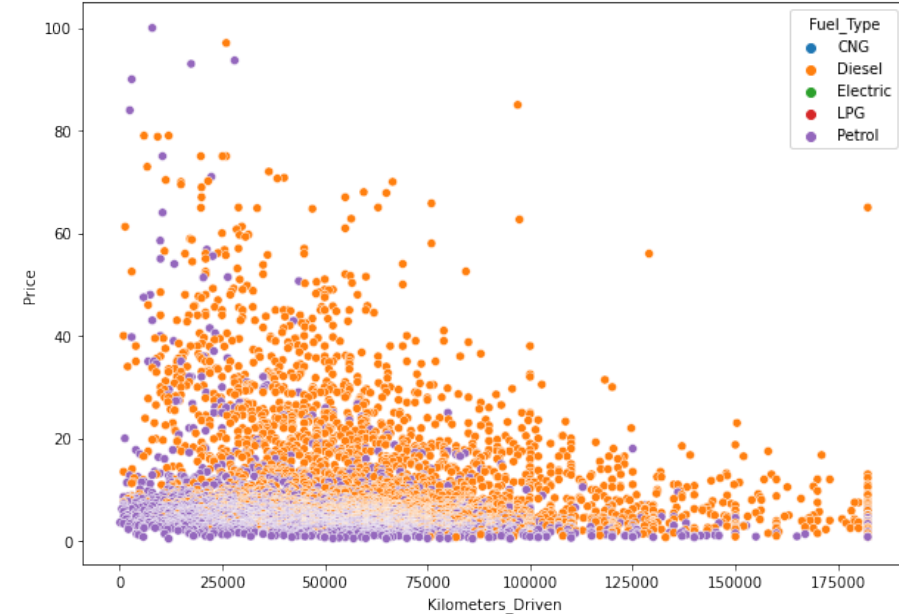
## Price & Age



## Observations

- Price and age have relatively low correlation
- Price drops steadily throughout age
- Most higher priced cars are automatics under the age of 10

## Price & Kilometers\_Driven

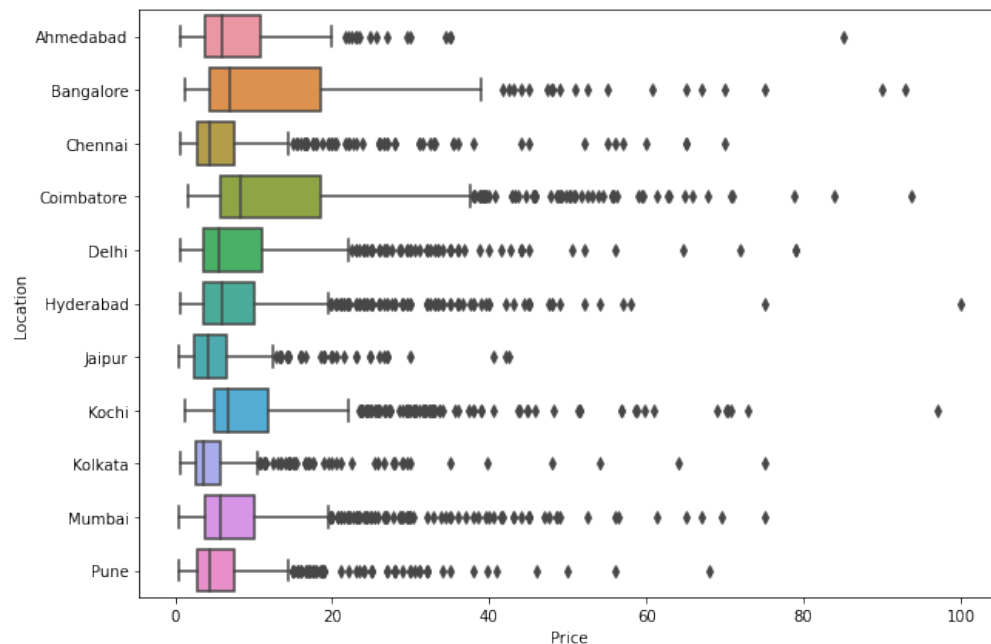


## Observations

- KM driven had still has far too many high outliers and was capped around 190,000
- As would be expected price drops at a steady rate as KM Driven rises
- Diesel fuel appears to account for most higher priced cars, this is counter intuitive given most higher performance cars are petrol. This should be investigated more closely

# Multi-Variate Analysis

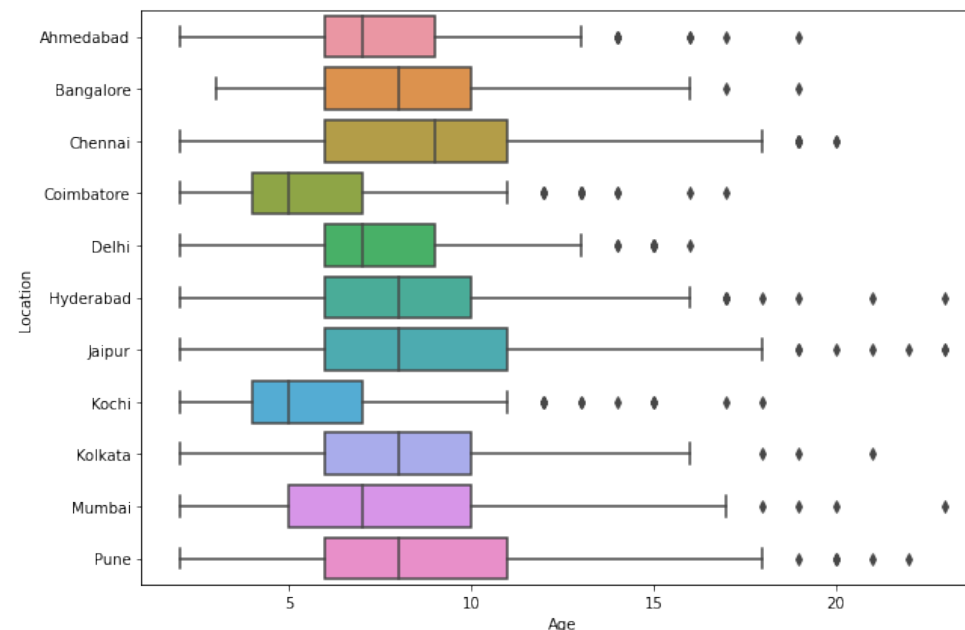
## Price & Location



## Observations

- Price across locations are mostly even with Mumbai, Hyderabad (most cars located here) showing very similar IQR and mean, however Hyderabad is showing fewer outliers so could be a more stable target for predictions.
- Bangalore & Coimbatore are showing the highest mean prices

## Age & Location

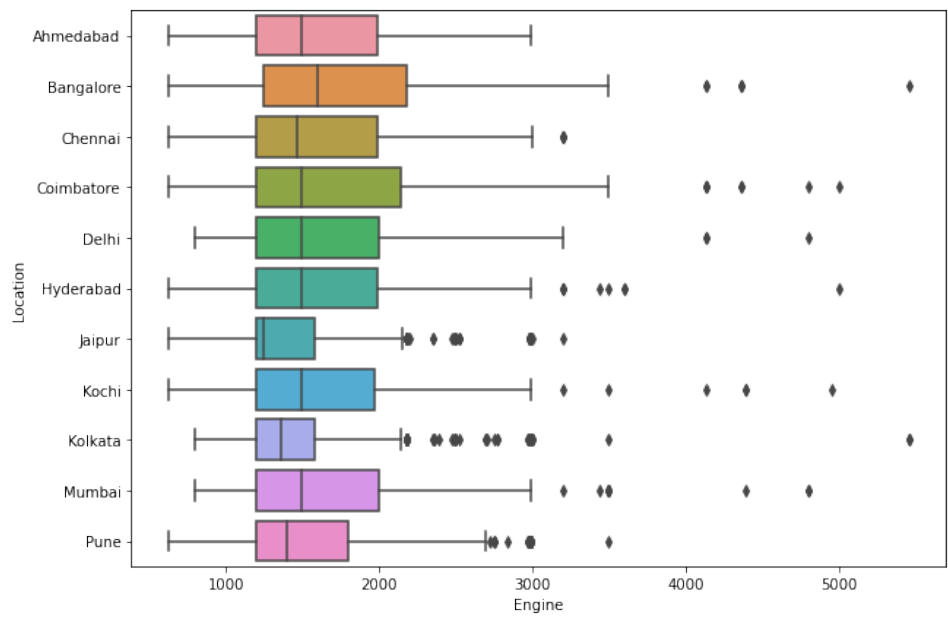


## Observations

- Median age across locations range from 5 to ~7 years
- Kochi and Coimbatore are showing the 'youngest' cars with a median age of 5.

# Multi-Variate Analysis

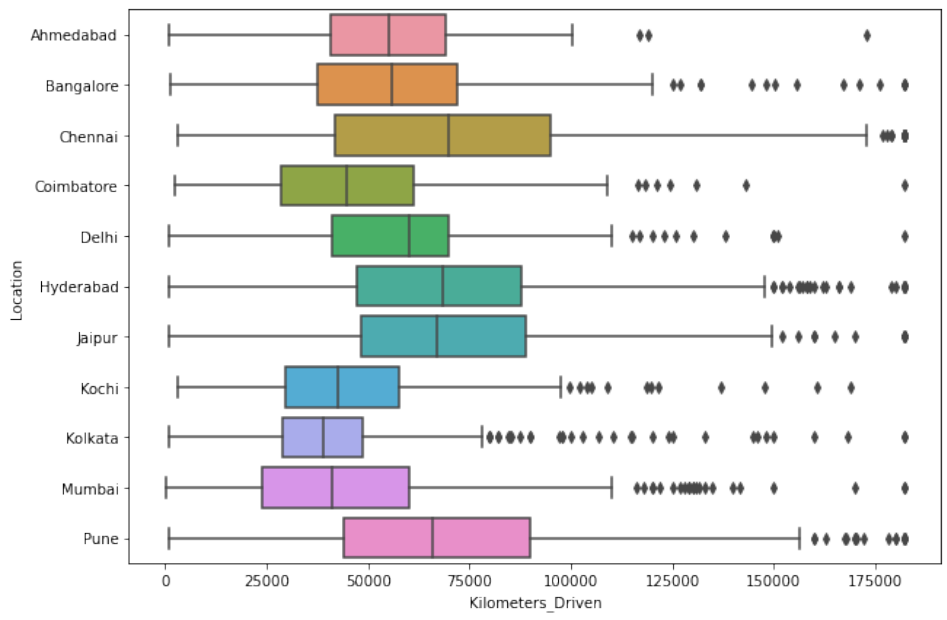
Engine & Location



Observations

- Engine sizes are fairly stable across locations at a mean around 7.25
- Combatore has the larges range of engine sizes

Kilometers Driven & Location



Observations

- Mumbai & Chennai have the widest range of KM Driven with Mumbai having a lower mean by ~80000 KM