

Cars4U

Nibu Kuriakose
Apr 2021

Content

No	Item	Slide #
1.	Introduction and Proposed Approach	3
2.	Dataset Information	4
3.	Analysis 3.1 Univariate Analysis 3.2 Factors Impacting Price 3.3 Others 3.4 Multivariate Analysis	5 6-9 10-11 12
4.	ML Model 4.1 Predicting Price& Influencing Factors 4.2 Performance Metrics and Assumptions	13 14
5.	Key Insights	15
6.	Recommendations	16

1. Introduction and Proposed Approach

Background

- There is high demand for used cars in the Indian market. In 2018-19, used car sales (4m units) was more than new car sales (3.6m units)
- Supply and pricing for used cars are uncertain. Hence, pricing becomes important to grow and be profitable in this market.

Purpose

- Perform exploratory data analysis to derive insights and recommendations
- Create a liner regression pricing model to predict market price to allow differential pricing which drives higher profit and growth.

Proposed Approach

1. Build a liner regression model to predict the price using the relevant data attributes in the dataset provided. Ensure this model provides reasonable prediction results using the test data which is separate from the data used to build the model.
2. Use this model in a small scale initially in few locations to predict the price.
3. More data is gathered to further enhance the model.
4. Once the enhanced model is proving successful in selected location, roll-out the model more widely to other locations.
5. As more data is gathered and the model is enhanced, it is expected to predict the price more accurately.

2. Dataset Information

- There are 7,253 samples
- Each sample has 14 attributes

Data Cleansing and Feature Engineering

- The following manipulation had to be done to the data.
 - The units of the following fields had to be removed to make it numeric- Engine, Power, New_Price, Mileage
 - The first name of the cars were extracted into a single column- 33 brand names
 - New column created to capture where the car design is from (Europe, Japan etc.)

Data Manipulation

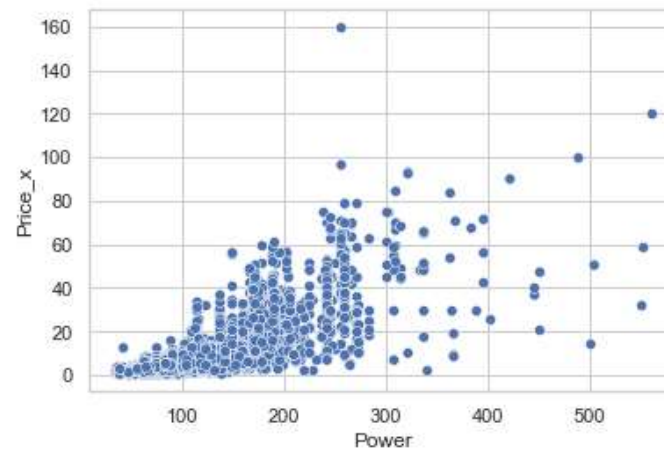
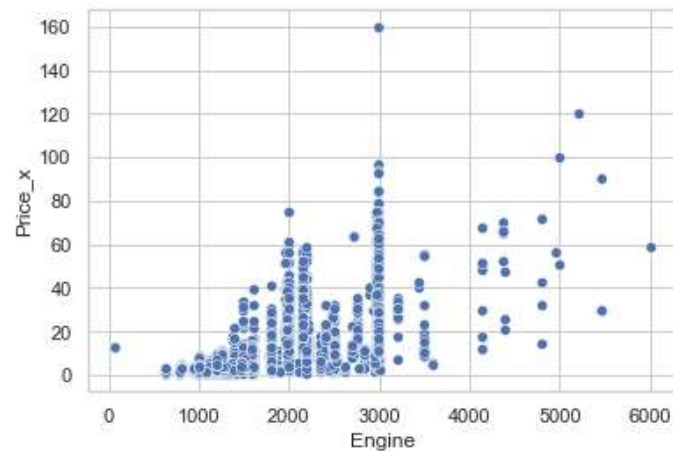
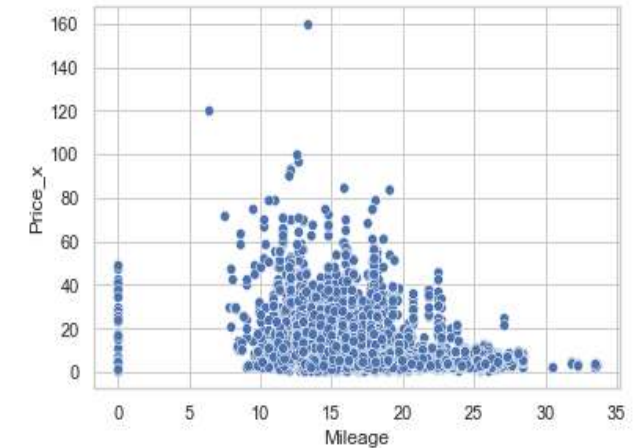
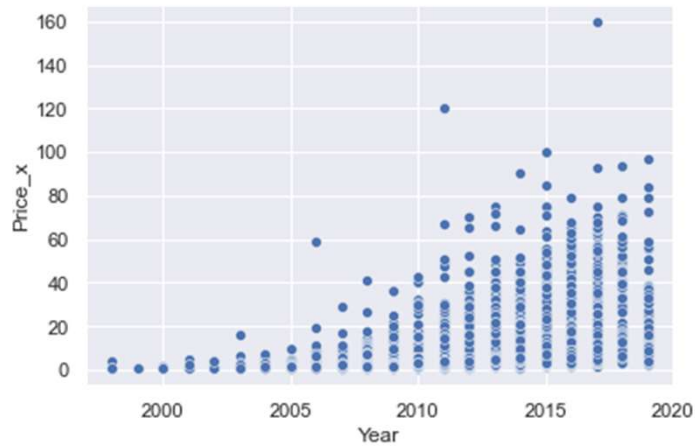
- For cars whose 'New Price' or 'Price' are not present in the dataset, it was filled with mean 'New Price' or 'Price' respectively of cars with same name, year, fuel_type and transmission. This allowed 555 'Price' fields and 108 'New_Price' fields to be filled.
- Median imputation was done to fill null values in 'Engine', 'Power', 'New_Price', 'Mileage'

3.1 Univariate Analysis

Attribute	Description
Year	<ul style="list-style-type: none"> Left skewed with 50% of cars between 2011 and 2016
Kilometers Driven	<ul style="list-style-type: none"> Slightly left skewed with 50% of cars between 34k and 73k
Engine	<ul style="list-style-type: none"> Right skewed with 50% of data between 1,198 and 1,968
Power	<ul style="list-style-type: none"> Right skewed with 50% of data between 75 and 138
Seat	<ul style="list-style-type: none"> 84% of cars have 5 cars
New_Price	<ul style="list-style-type: none"> Right skewed and 50% of cars are priced between 7.9L and 25.2L
Price	<ul style="list-style-type: none"> Right skewed and 50% of cars are priced between 3.5L and 9.9L
Fuel_Type	<ul style="list-style-type: none"> 5 types of fuels, but majority either Diesel or Petrol. Most cars are diesel.
Name	<ul style="list-style-type: none"> Maruti, Hyundai, Honda are the top three type of cars
Transmission	<ul style="list-style-type: none"> Majority are manual
Owner_Type	<ul style="list-style-type: none"> Majority are first hand.

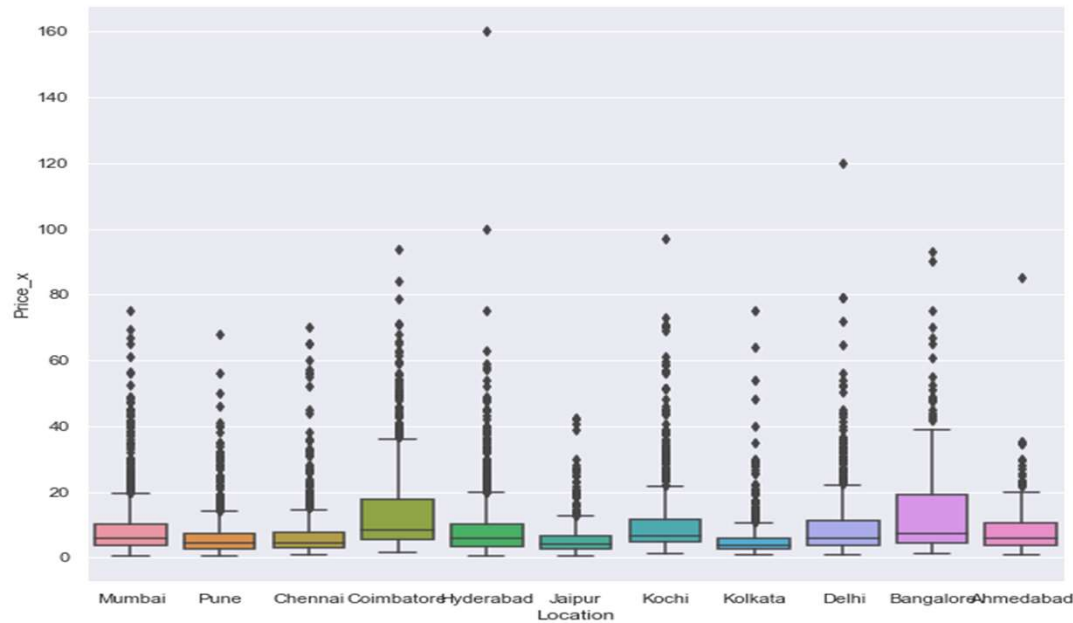
	Year	Kilometers_Driven	Mileage	Engine	Power	Seats	New_Price	Price
count	7253	7253	7251	7207	7078	7200	1114	6698
mean	2013.4	58699.1	18.1	1616.6	112.8	5.3	22.2	9.4
std	3.3	84427.7	4.6	595.3	53.5	0.8	26.9	11.0
min	1996	171	0	72	34.2	0	3.91	0.44
25%	2011	34000	15.17	1198	75	5	7.94	3.5
50%	2014	53416	18.16	1493	94	5	11.455	5.5875
75%	2016	73000	21.1	1968	138.1	5	25.27	9.9
max	2019	6500000	33.54	5998	616	10	375	160

3.2 Factors Impacting Price (I/IV)

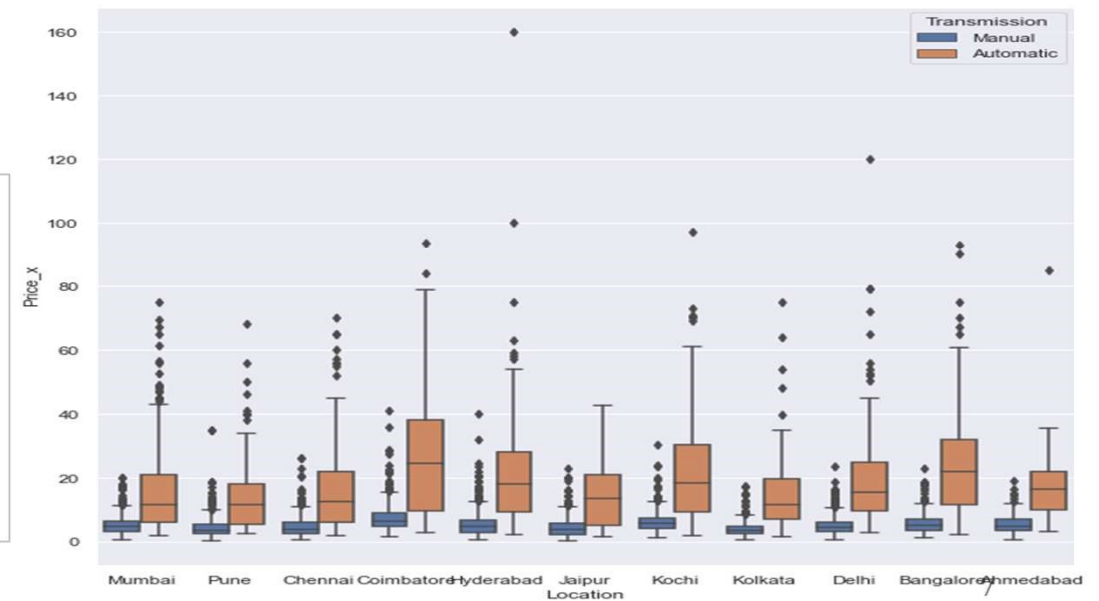
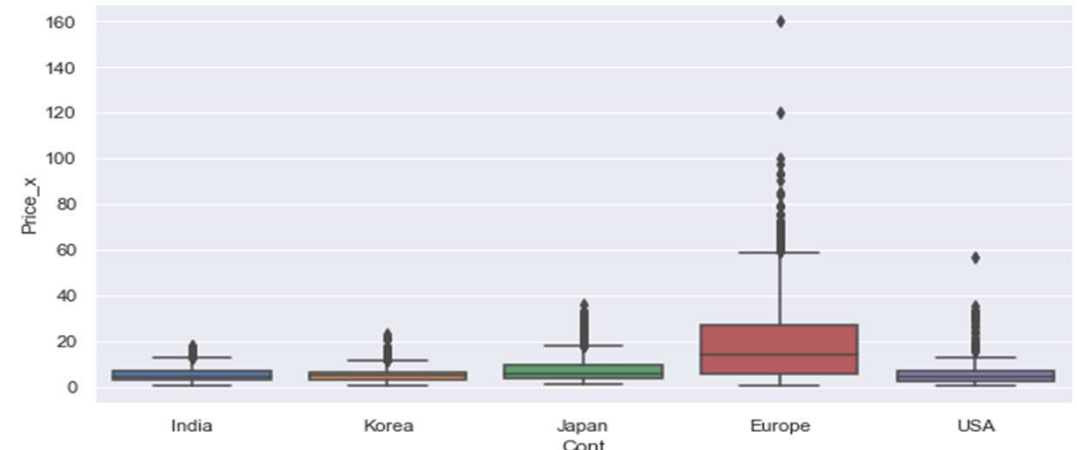


- Older vehicles are of lower price compared to newer ones.
- Price increases if the price of the new vehicle is higher
- Price increases with higher power and bigger engine
- Price of low mileage vehicles are higher. This might be because the bigger vehicles which are expensive have a lower mileage.

3.2 Factors Impacting Price (II/IV)

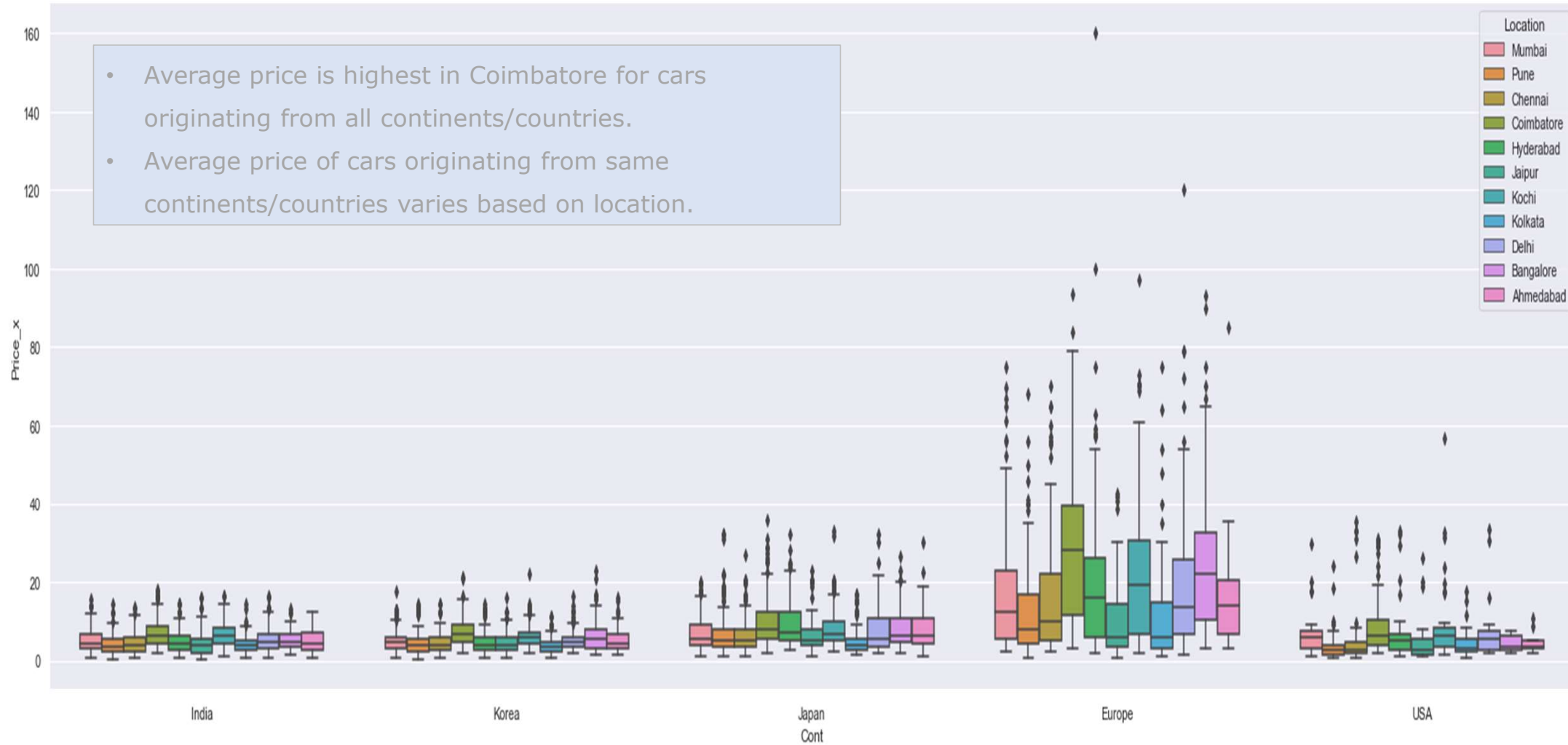


- Average price is highest in Coimbatore, followed by Bangalore, Kochi and Hyderabad. The lowest is in Kolkata.
- European cars have highest price followed by Japanese
- Automatic cars have higher price than manual transmission



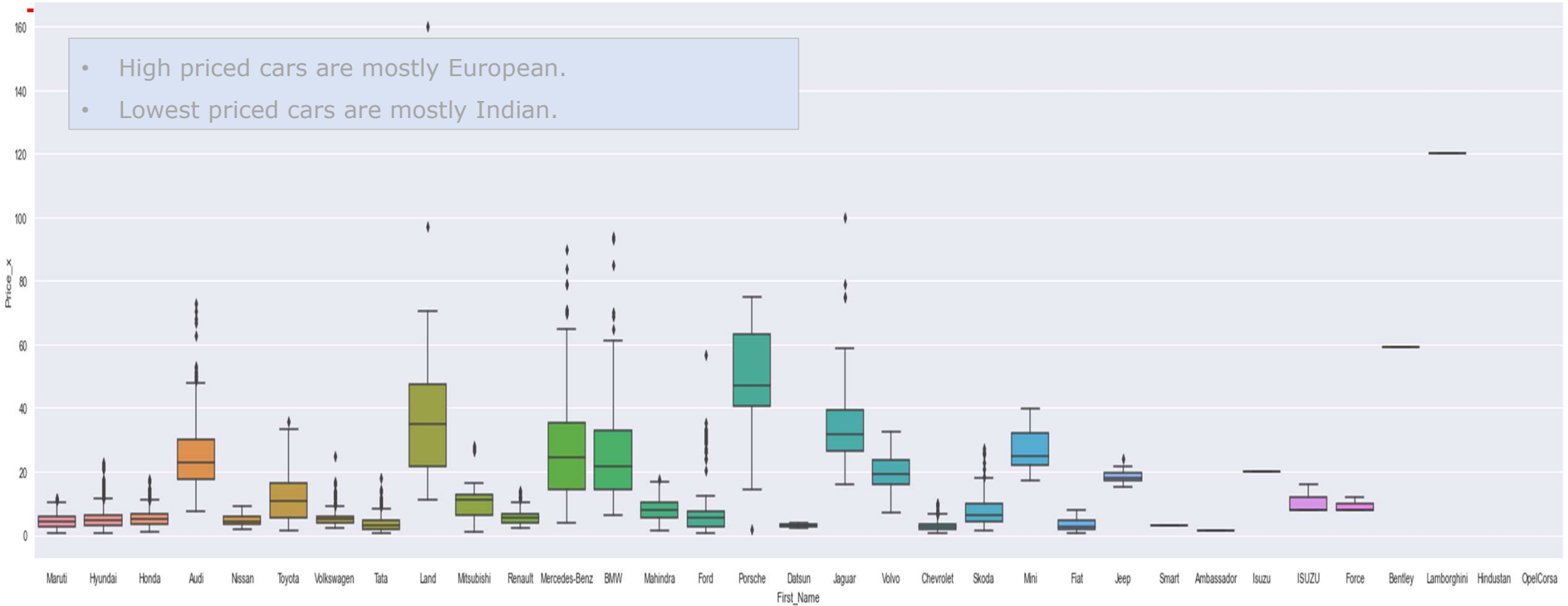
3.2 Factors Impacting Price (III/IV)

- Average price is highest in Coimbatore for cars originating from all continents/countries.
- Average price of cars originating from same continents/countries varies based on location.

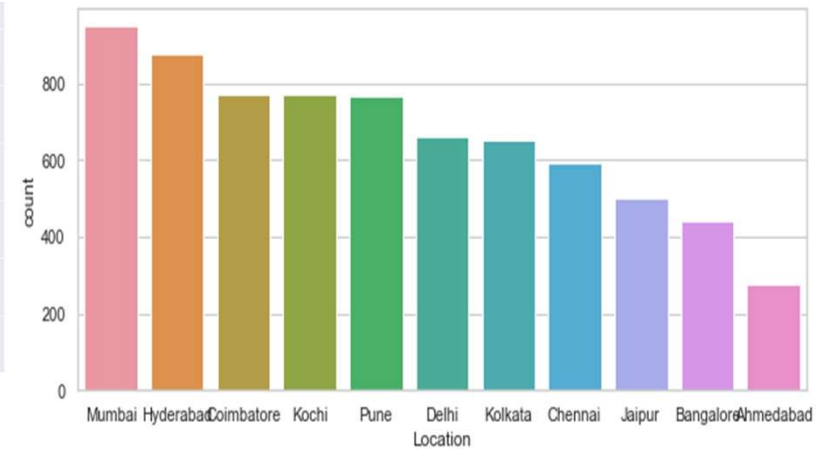
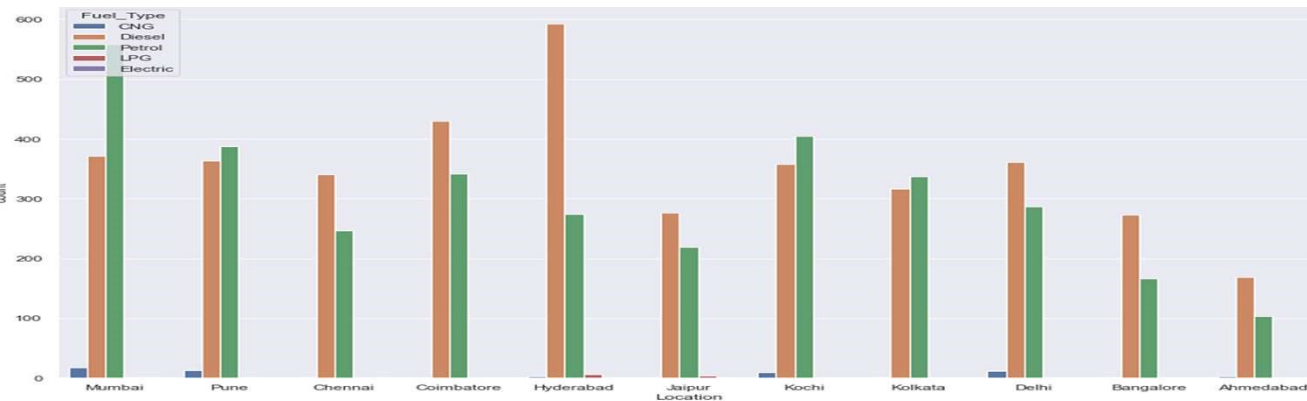
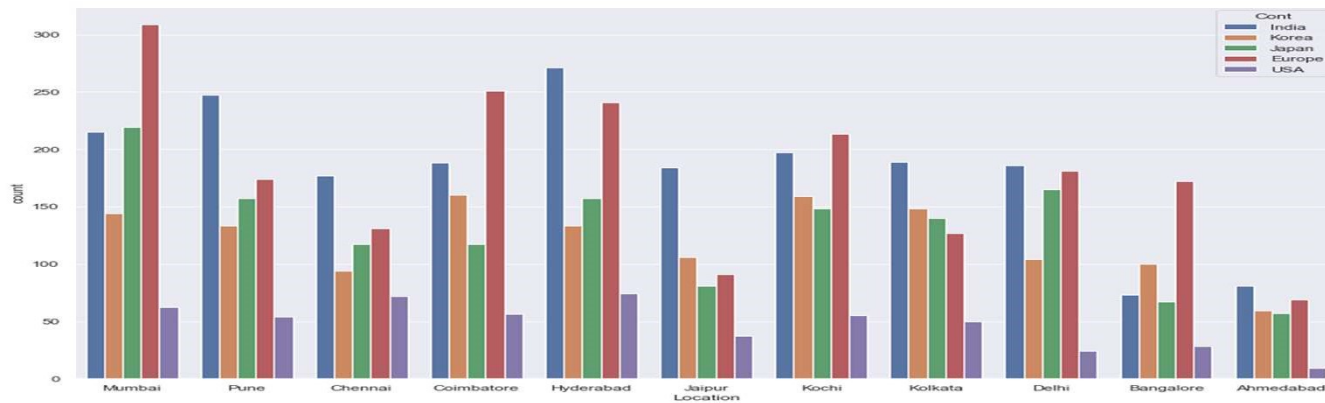


3.2 Factors Impacting Price (IV/IV)

- High priced cars are mostly European.
- Lowest priced cars are mostly Indian.

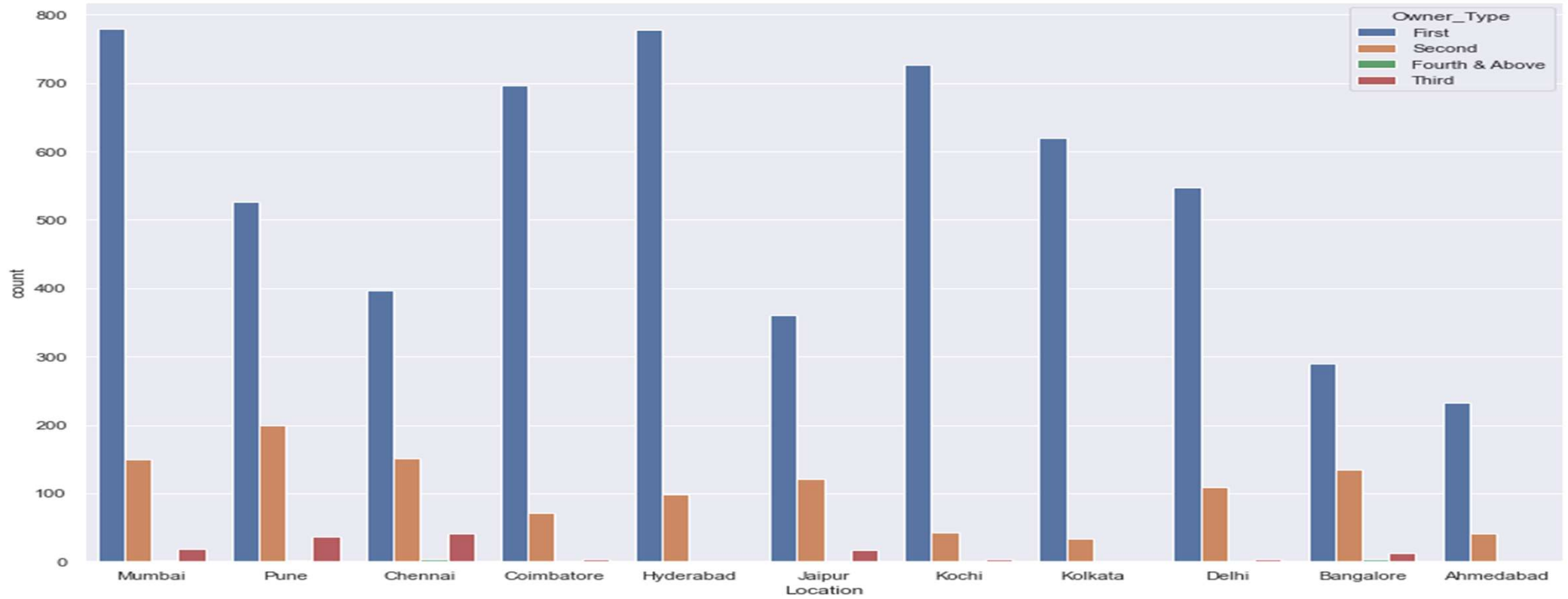


3.3 Others



- Most cars sold in Mumbai and least in Ahmedabad.
- In every location, either European (4 locations) or Indian cars (11 locations) are sold most.
- 7 locations have mostly Diesel vehicles and remaining have Petrol the most.

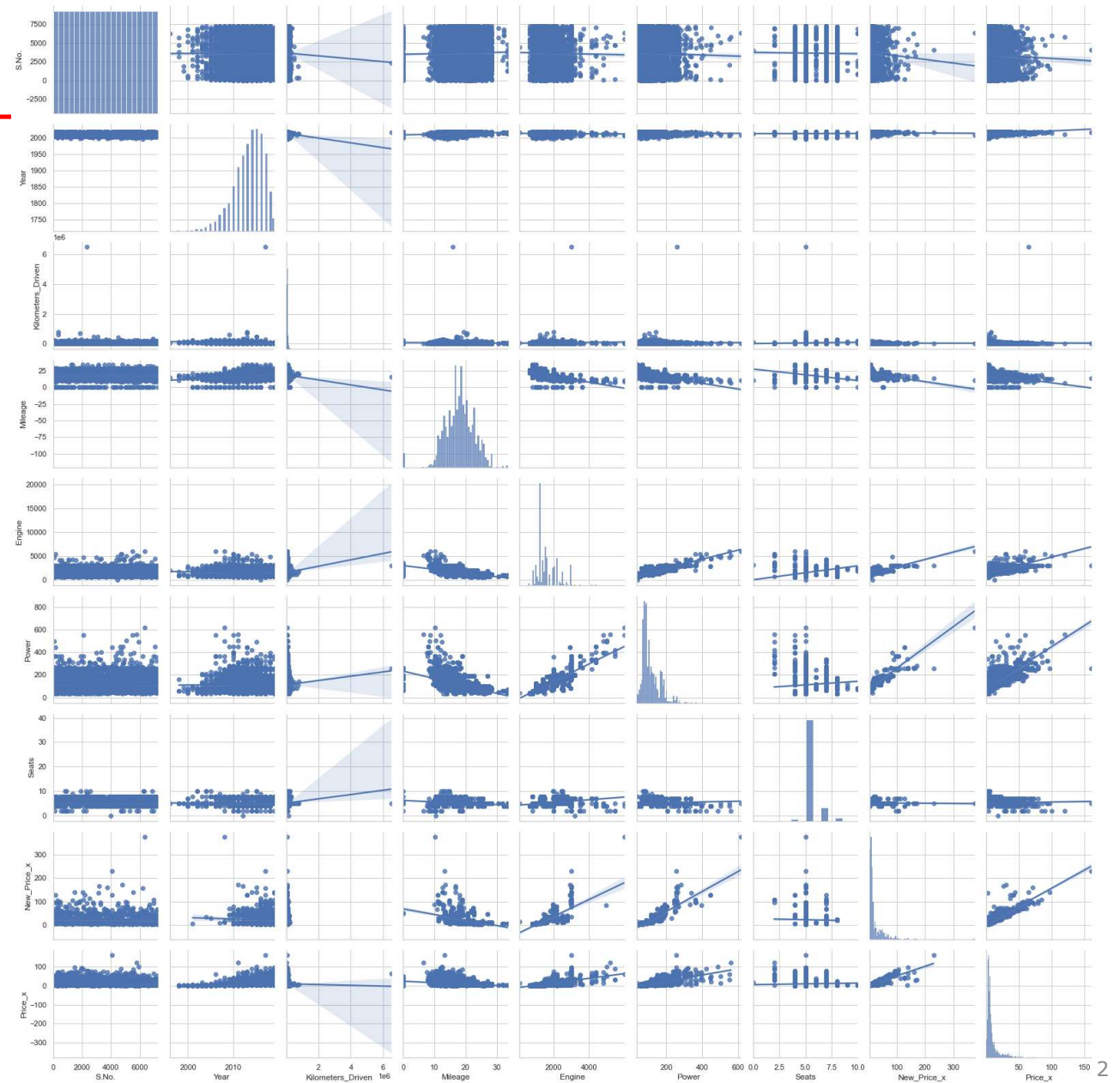
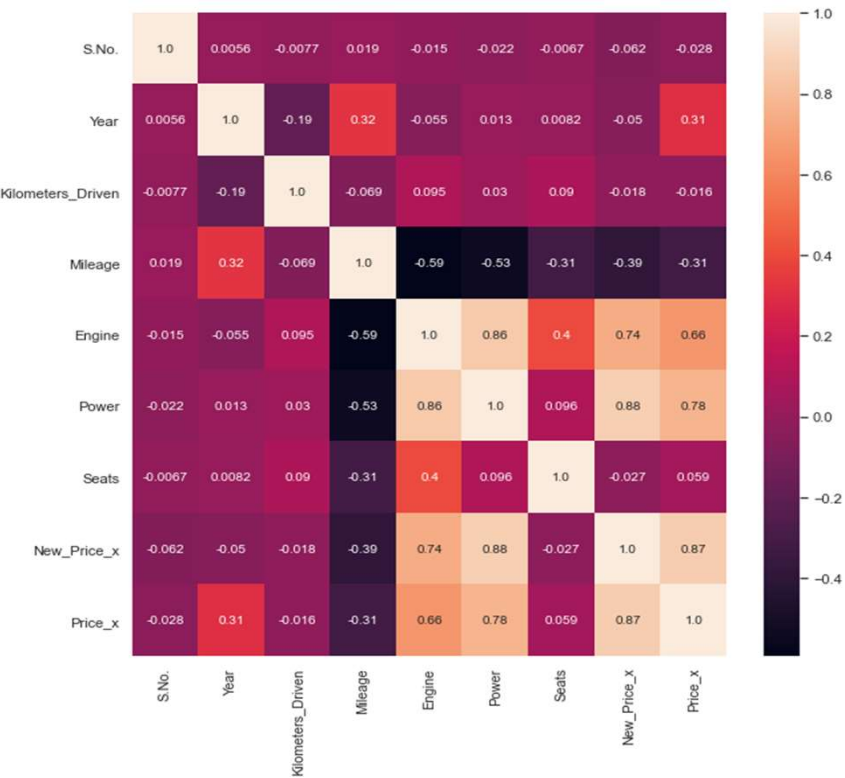
3.3 Others



- Third hand cars sold only in 5 locations.

3.4 Multivariate Analysis

- New_Price is positively co-related with Price, Power and Engine.
- Engine and Power are positively correlated
- Mileage is negatively correlated with Price, New_Price, Engine, Seats and Power



4.1 ML Model- Predicting Price& Influencing Factors

ML Pricing Model

- Below are the coefficients of the machine learning model which will help predict price of cars with 90% accuracy (based on the data provided).
- There are 10 parameters for the model

	Feature	Coefs
Location	Year	0.6
	Kilometers_Driven	0.0
	Mileage	-0.2
	Power	0.04
	New_Price_x	0.0
	Location_Ahmedabad	0.1
	Location_Bangalore	0.8
	Location_Chennai	0.2
	Location_Coimbatore	0.6
	Location_Delhi	-0.4
Fuel Type	Location_Hyderabad	0.7
	Location_Jaipur	0.2
	Location_Kochi	-0.1
	Location_Kolkata	-1.0
	Location_Mumbai	-0.2
Owner Type	Fuel_Type_CNG	-1.1
	Fuel_Type_Electric	4.4
	Fuel_Type_LPG	-1.7
	Fuel_Type_Petrol	-2.1
	Transmission_Manual	-1.1
	Owner_Type_Fourth & Above	-1.3
	Owner_Type_Second	-0.3
	Owner_Type_Third	-0.3
	Intercept	-1157.9

	Feature	Coefs
Vehicle Brand	First_Name_Ambassador	2.1
	First_Name_Audi	5.2
	First_Name_BMW	4.2
	First_Name_Bentley	9.3
	First_Name_Chevrolet	-1.5
	First_Name_Datsun	-2.0
	First_Name_Fiat	-1.7
	First_Name_Force	0.4
	First_Name_Ford	-0.4
	First_Name_Honda	-0.5
	First_Name_Hyundai	-0.3
	First_Name_ISUZU	-0.2
	First_Name_Isuzu	0.0
	First_Name_Jaguar	5.1
	First_Name_Jeep	4.6
	First_Name_Lamborghini	0.0
	First_Name_Land	5.6
	First_Name_Mahindra	-0.7
	First_Name_Mercedes-Benz	4.3
	First_Name_Mini	8.4
	First_Name_Mitsubishi	1.8
	First_Name_Nissan	-0.9
	First_Name_Porsche	4.1
	First_Name_Renault	-0.8
	First_Name_Skoda	-0.1
	First_Name_Smart	-3.1
	First_Name_Tata	-1.9
	First_Name_Toyota	2.5
	First_Name_Volkswagen	-1.1
	First_Name_Volvo	4.0

Major Factors Impacting Pricing Model

Items marked in green have a major positive impact on the price. Key points to note are:

- Type of vehicle- Bentley, Audi, BMW, Jaguar, Jeep, Land, Mercedes, Mini and Porsche commands high price.
- Year- The more recent the car, the higher the price
- Power- More powerful vehicles command a higher price (50% of cars have power above 94).
- Location- Cars in Coimbatore, Bangalore, Hyderabad have higher price than others
- Electric fuel – Electric cars have a very high premium compared to others

Items highlighted in red have a negative impact on price. Main factors are below:

- Tata, Chevrolet, Datsun, Fiat car brands
- Mileage- Relative impact higher if this is a car with lower price
- Location- Kolkata
- Petrol cars
- Third hand or more

4.2 ML Model- Performance Metrics and Assumptions

Model Performance Metrics

	Training Data	Test data
R ²	0.8965	0.9
RMSE	1.82936	1.82704
MAE	1.33821	1.34755

Following assumptions have been validated

1. No Multicollinearity
2. Mean of residuals close to zero
3. No Heteroscedacity
4. Linearity of variables proved
5. Normality of error terms proved

5. Key Insights

Key Insights

1. Price of cars are highly dependent on the location where the car is sold.
2. As the price of new cars go up, the used car price also goes up.
3. European cars have higher price than others and is highest sold in most of the locations. Remaining locations have the highest sales of Indian cars.
4. Majority of the cars are petrol or diesel, but electric vehicles has a premium price.
5. Brand of the car is a major factor determining the price of the car.
6. High power and bigger engines have higher price
7. Price of low mileage cars are higher.
8. Manufacturing year has major influence on the price. The more recent the car, higher is the price.

6. Recommendations

1. Cars should be sold at higher price in Coimbatore, Bangalore, Kochi and Hyderabad.
2. Cars should be sold in Mumbai, Pune and Delhi at lower prices where volumes are high. Exit markets such as Ahmedabad where volume and price are low.
3. Sell cars not more than 5 years old.
4. Explore possibility of buying cars from lower priced location and selling them in high priced locations.
5. Procure and sell high power cars and European (specifically Betley, Audi, BMW, Jaguar, Jeep, Land, Mercedes, Mini and Porsche) designed cars at higher price.
6. If the price of new cars go up, the price of used cars should be correspondingly increased.

Model improvement

1. New price was available only for around 15% of cars. Obtaining price of new cars for rest will help to improve the model.
2. Gather the data on purchase price of cars so that the profits can be calculated so that the model can be built on maximising the profit rather than the sale price. Some of the high price cars might be low margin which can be confirmed only if the purchase price is known.
3. As this is a dynamic market more data should be gathered and the model should be trained on a regular basis to ensure the model accuracy is kept high.