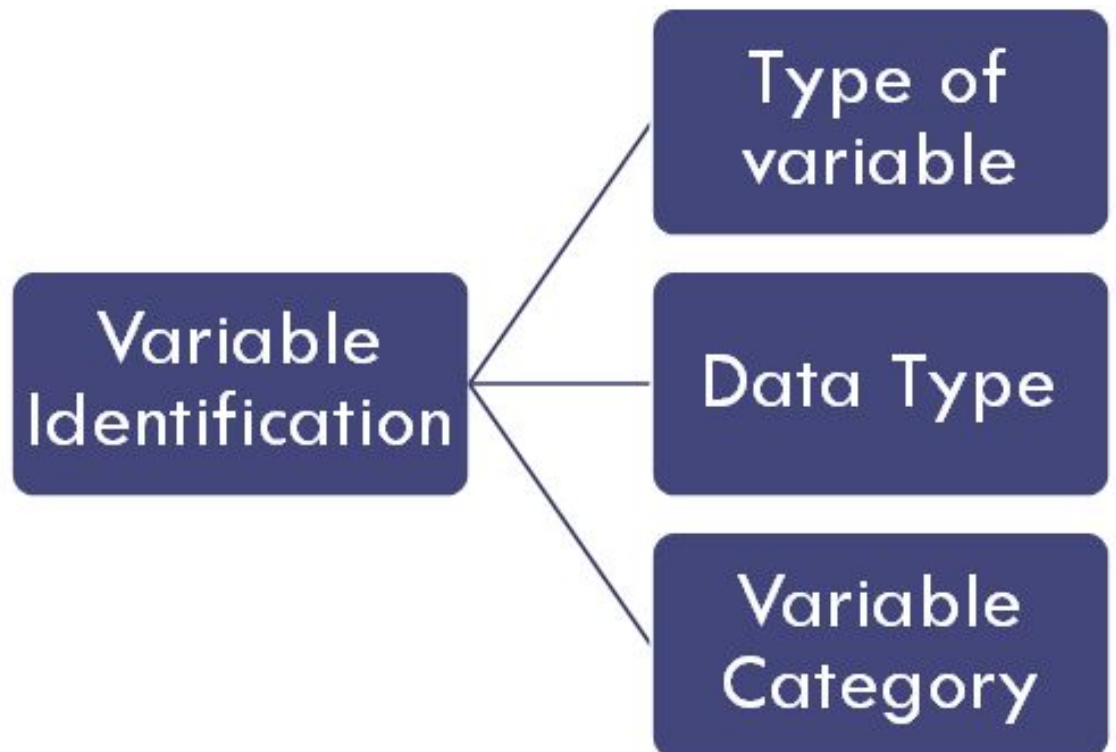


- **Name** - Mahaveer Rulaniya
  - **The language used** - Python
  - **Hardware and Software requirements**- GPU, Jupyter Notebook
  - **Detailed explanation of the code is present in Github-**  
[https://github.com/mahaveer-rulaniya/Data\\_Analysis\\_Challenge](https://github.com/mahaveer-rulaniya/Data_Analysis_Challenge)
- 

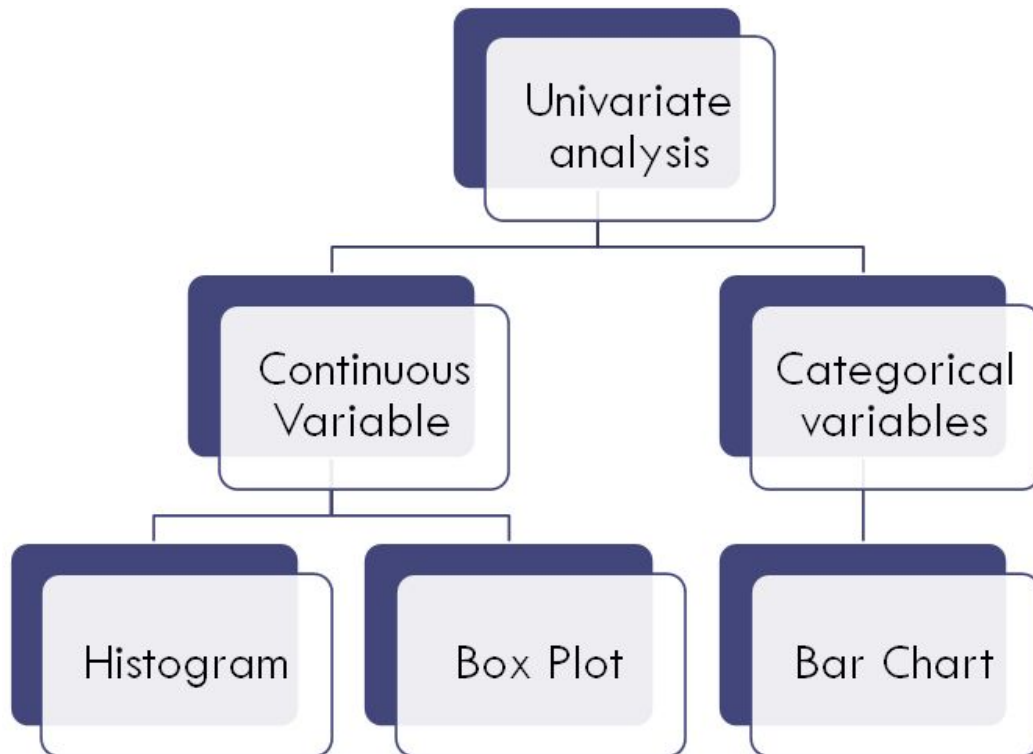
## Analysis of Automobile Prices

- ❑ **Aim and Objective** - Exploring the data by calculating summary and descriptive statistics, and creating visualizations of the data, several potential relationships between automobile characteristics and their price.
- ❑ **Executive Summary** - There are many factors that contribute to the major inclinations in deciding the price of automobiles. Some of them are -
  - Make
  - Cylinders
  - Horsepower
  - City MPG
  - Drive Wheels
- ❑ **Introduction to Exploratory Data Analysis-**

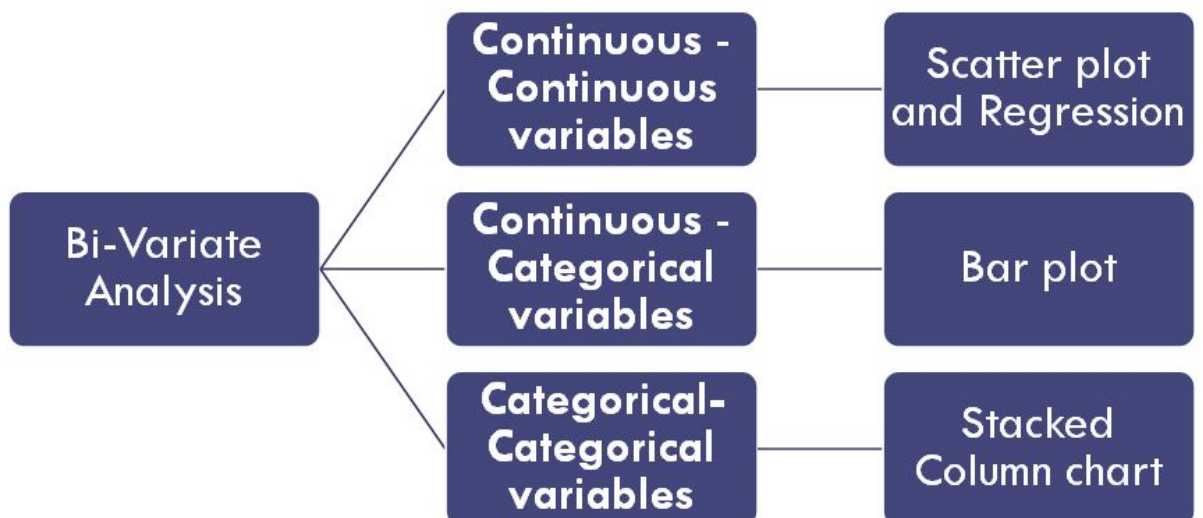
### 1. Variable identification



## 2. Univariate Analysis



## 3. Bivariate Analysis



Now we will explore all these steps in-depth and also perform statistical analysis.

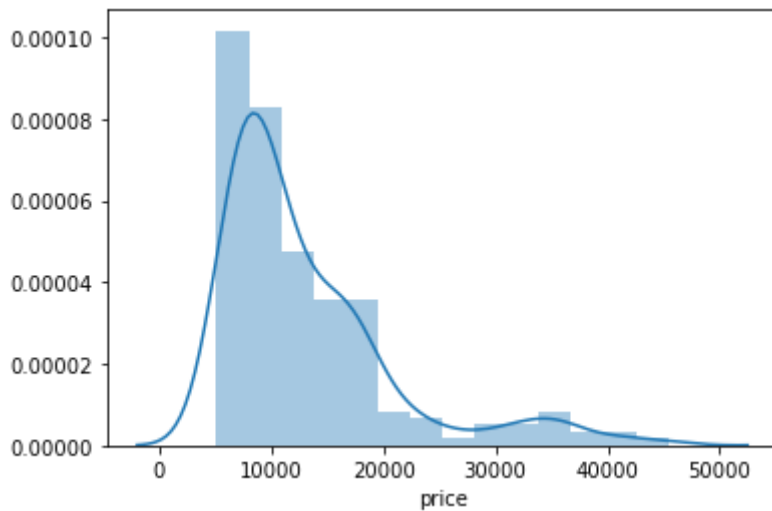
# Data Analysis and Visualization

## ◆ Individual Feature Statistics-

Summary statistics for minimum, maximum, mean, median, standard deviation, and distinct count were calculated for numeric columns.

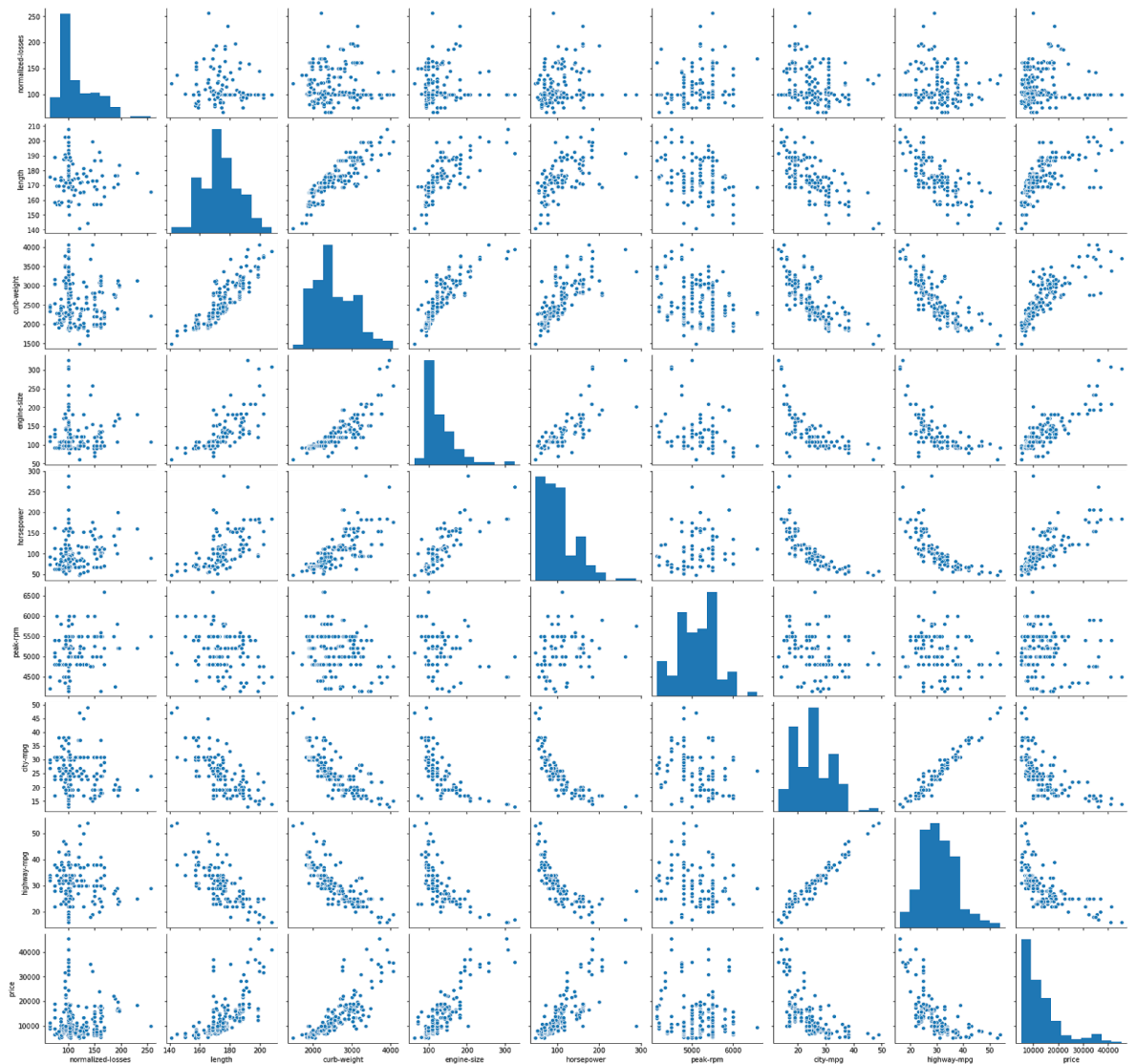
Column	mean	std	min	max
symboling	0.886364	1.079372	-1.00	3.00
normalized-losses	113.840909	22.407156	89.00	161.00
wheel-base	98.315909	6.300593	88.60	113.00
length	174.786364	11.480316	155.90	199.60
width	65.872727	2.096126	63.60	71.10
height	53.425000	2.140080	48.80	56.70
curb-weight	2596.840909	564.805432	1874.00	4066.00
engine-size	129.863636	43.006141	90.00	258.00
bore	3.369545	0.276930	2.92	3.78
stroke	3.271136	0.336353	2.64	4.17
compression-ratio	9.213636	1.379198	7.60	14.90
horsepower	104.750000	38.625106	68.00	207.00
peak-rpm	5217.045455	352.658810	4425.00	5900.00
city-mpg	24.477273	5.819665	15.00	38.00
highway-mpg	30.136364	6.067442	19.00	43.00
price	14860.840909	8880.370439	6295.00	35550.00

A histogram of the Price column shows that the price values are right-skewed – in other words, most cars are priced at the lower end of the price range, as shown here:



### ◆ Numeric Relationships-

The following scatter-plot matrix was generated initially to compare numeric features with one another.



Viewing plots in the bottom row or the right-most column of this matrix shows an apparent relationship between price and other numeric features. Specifically, as length, curb-weight, engine size, and horsepower increase, so does price; and as city-mpg increases, the price reduces.

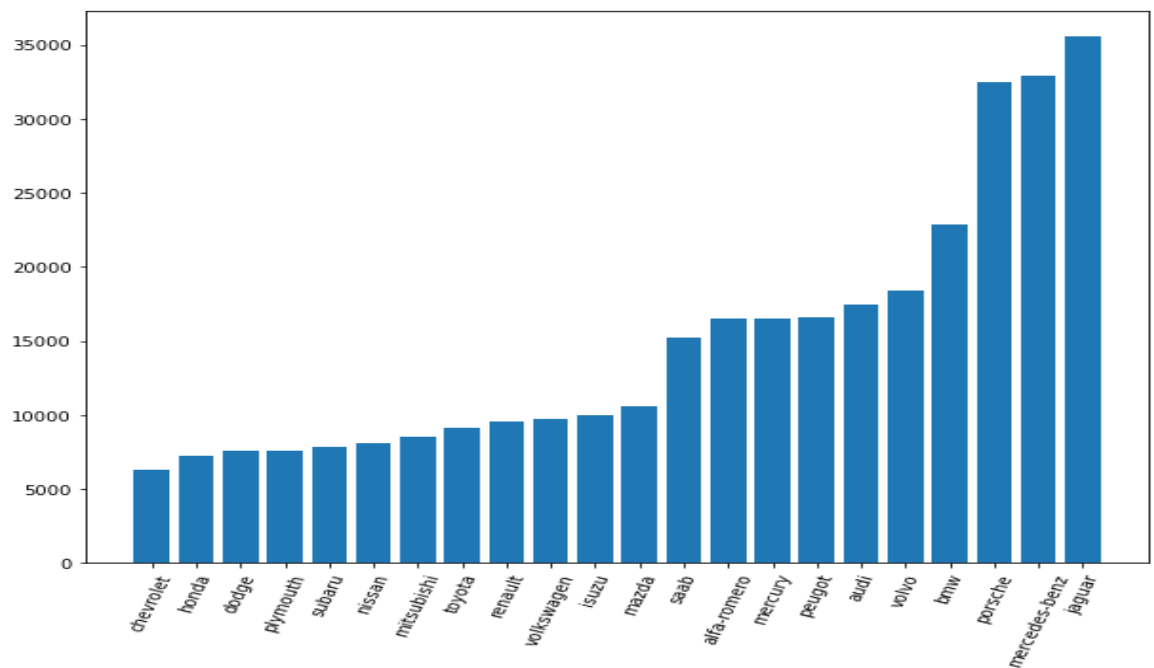
**Correlation**-The correlation between the numeric columns was then calculated with the following results:

Columns	normalized-losses	length	curb-weight	engine-size	horsepower	peak-rpm	city-mpg	highway-mpg	price
normalized-losses	1.000000	-0.060433	-0.006289	-0.008657	0.088777	0.242206	-0.124516	-0.086070	0.014377
length	-0.060433	1.000000	0.877728	0.683360	0.551990	-0.286362	-0.670909	-0.704662	0.686829
curb-weight	-0.006289	0.877728	1.000000	0.850594	0.750505	-0.266358	-0.757414	-0.797465	0.819595
engine-size	-0.008657	0.683360	0.850594	1.000000	0.809321	-0.244383	-0.653658	-0.677470	0.860074
horsepower	0.088777	0.551990	0.750505	0.809321	1.000000	0.130119	-0.800797	-0.770269	0.749343
peak-rpm	0.242206	-0.286362	-0.266358	-0.244383	0.130119	1.000000	-0.114230	-0.054195	-0.107847
city-mpg	-0.124516	-0.670909	-0.757414	-0.653658	-0.800797	-0.114230	1.000000	0.971337	-0.668862
highway-mpg	-0.086070	-0.704662	-0.797465	-0.677470	-0.770269	-0.054195	0.971337	1.000000	-0.693190
price	0.014377	0.686829	0.819595	0.860074	0.749343	-0.107847	-0.668862	-0.693190	1.000000

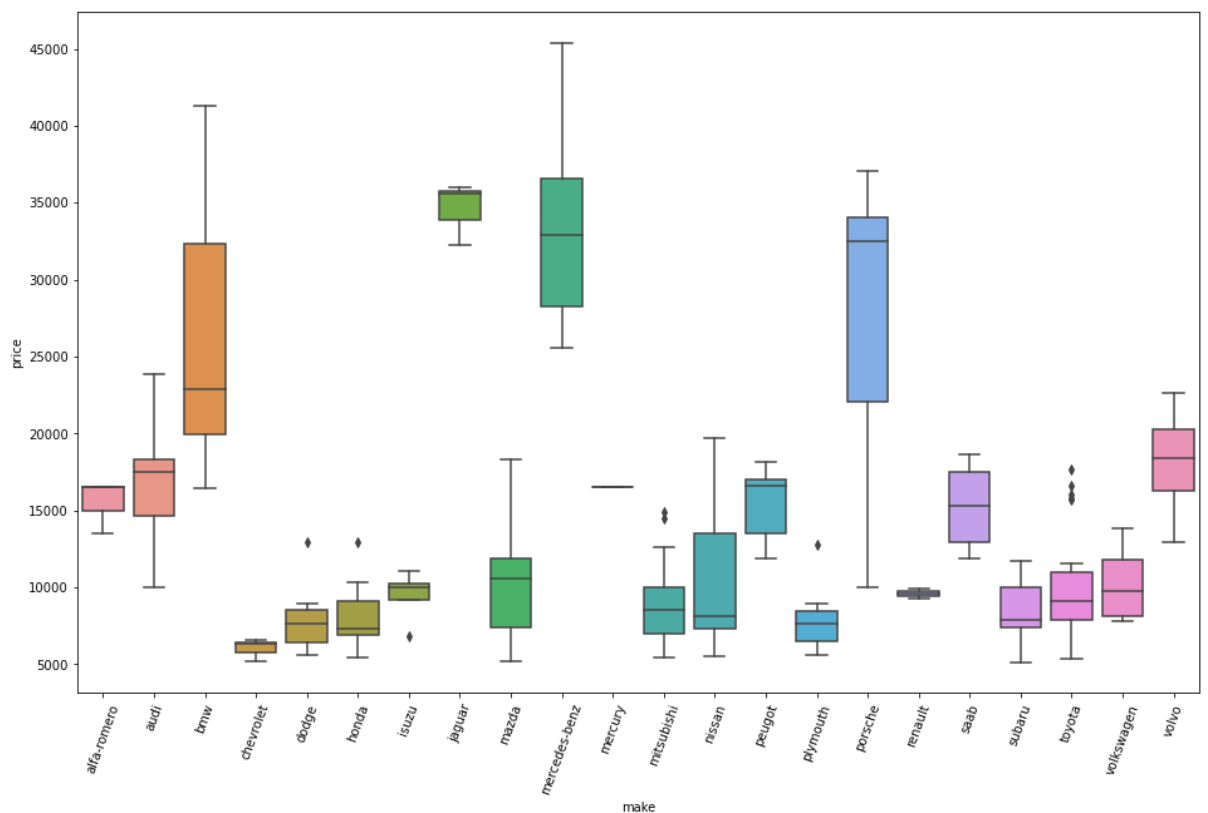
## ❖ Categorical Relationships-

After exploring the Numeric values and price relations of Automobile now let us focus on the Categorical variables and Price relationships through Bar graphs and Box plots.

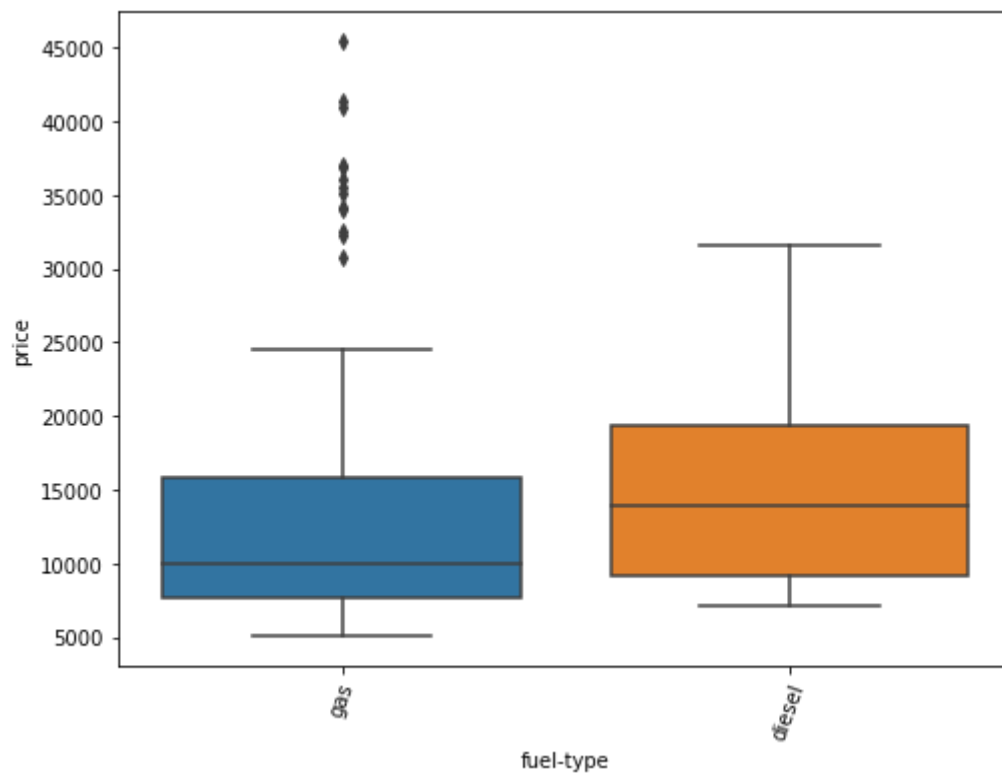
## 1. Make and Price relationship-



There are a few manufacturers that typically create expensive cars, some manufactures with predominantly mid-priced cars, and some that seem to specialize in lower-priced cars

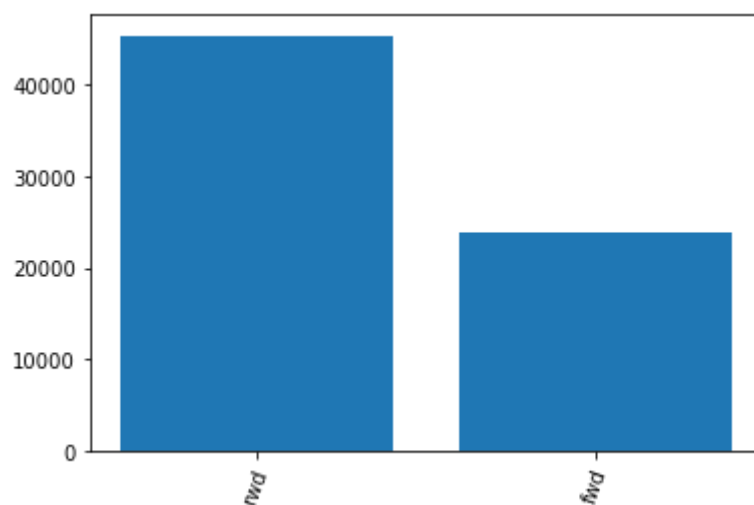


## 2. Fuel-type and price relationship-

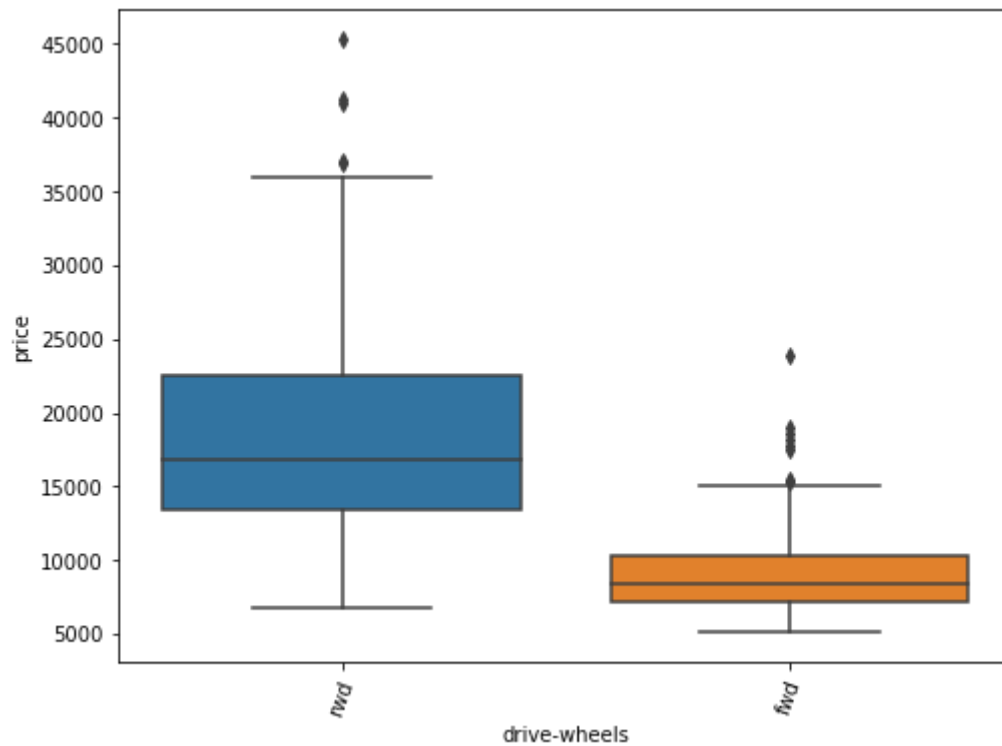


There are a wider range of prices for gas cars than for diesel cars, though the median price is similar for both types

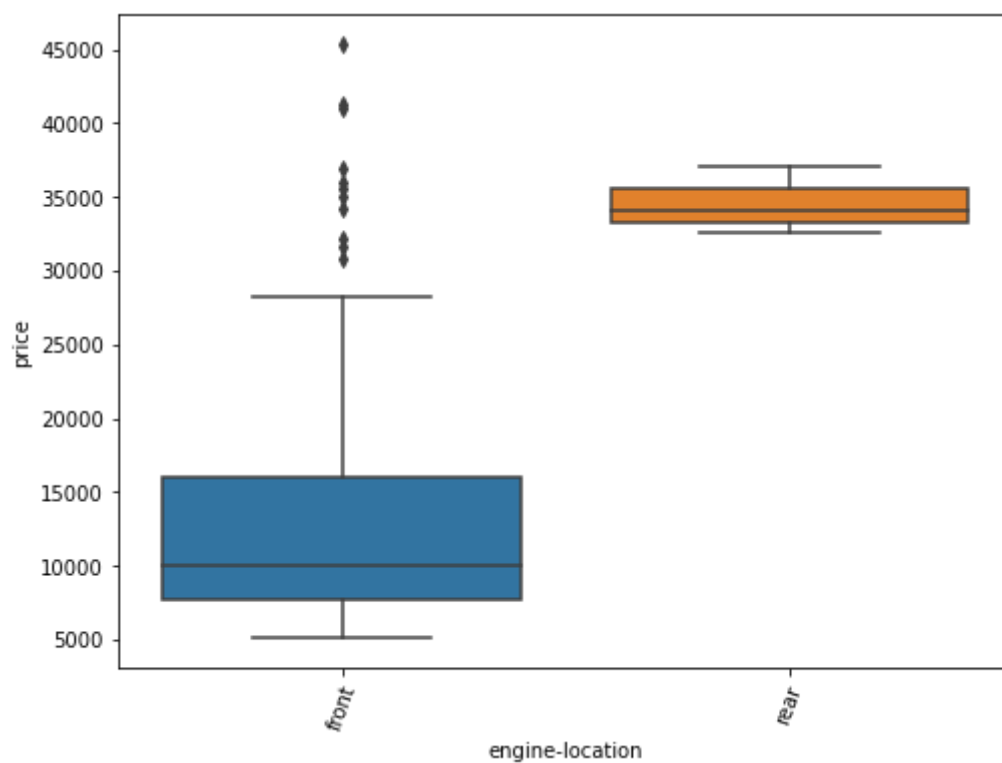
## 3. Drive-wheels and Price relationships-



Rear-wheel-drive cars are typically more expensive than other types of car which can also be seen in the below boxplot:



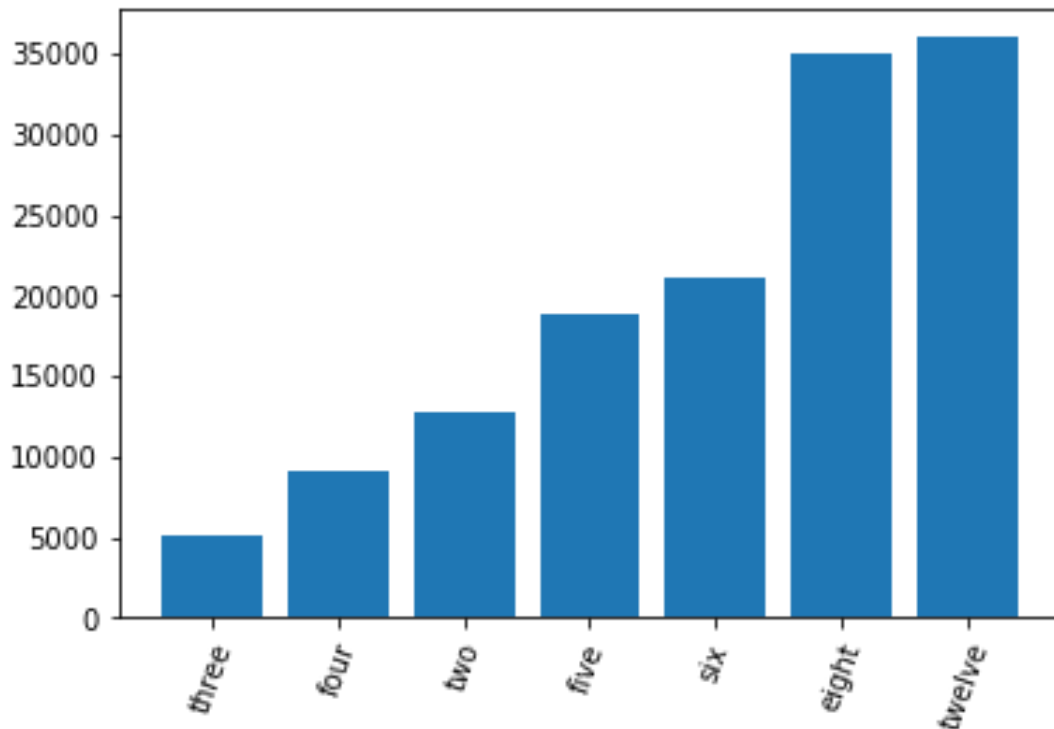
#### 4. Engine Location and Price relationship-



Rear-engine cars are significantly more expensive than front-engine cars; and their prices fall within a smaller range.



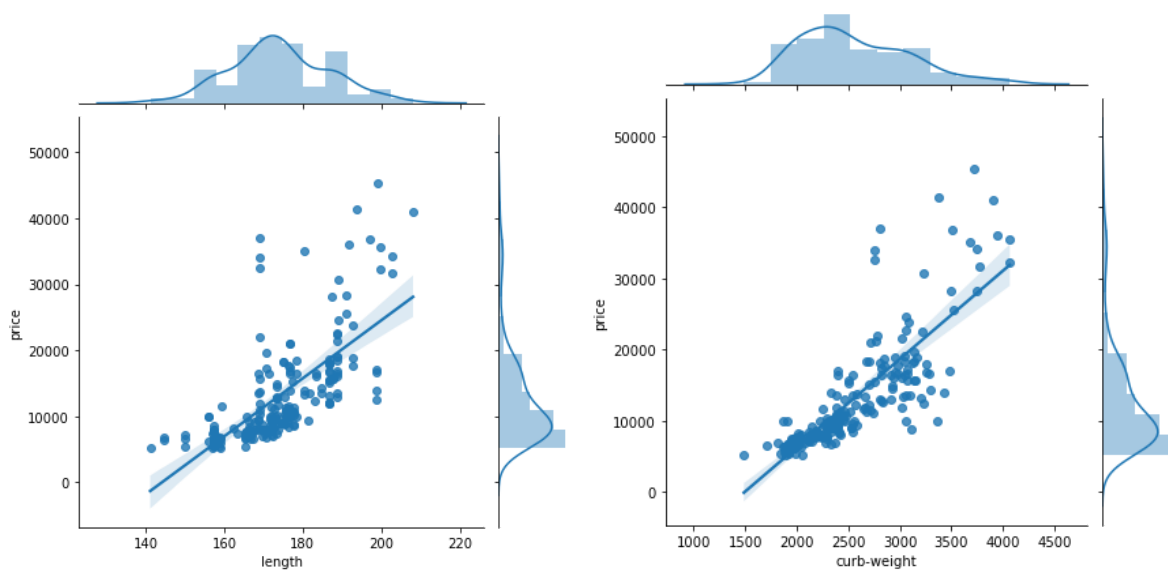
## 5. Number of Cylinders and price relationship-

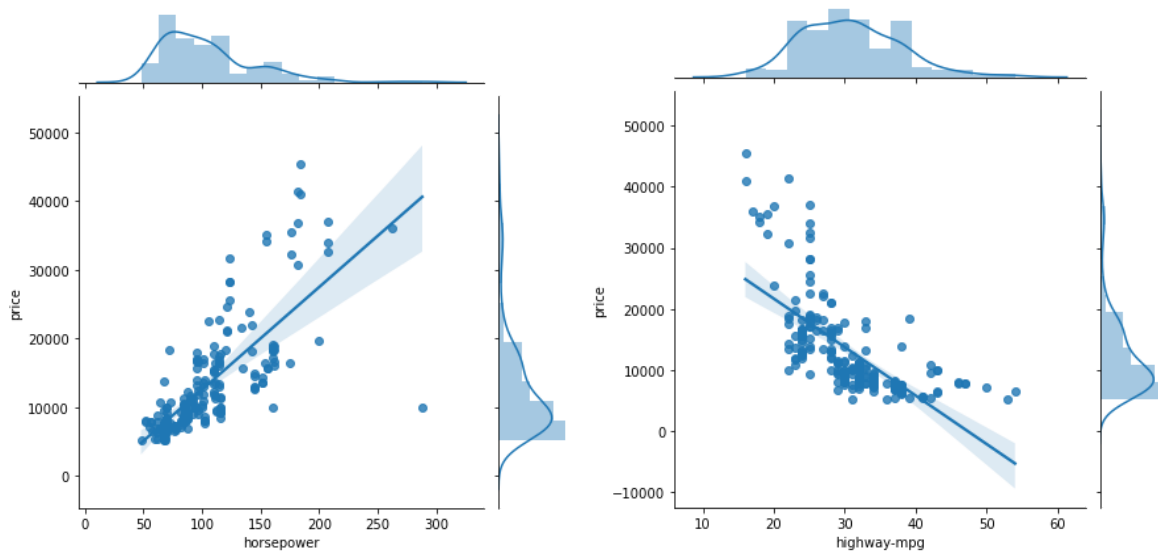


As the number of cylinders increases the price of automobile tend to increase.

## ◆ Regression Analysis-

It is important to see the linear relationships between some of the major factors and the price of the automobile.





Here it can be observed that as length, curb-weight, horsepower increase then the price increases while as Highway-mpg increases the price decreases.

### ❖ Conclusion-

This analysis has shown that the price of an automobile can be predicted from its features. In particular, the manufacturer, number of cylinders, horsepower, city MPG, and drive wheels have a significant effect on the price of an automobile. Secondary features, such as fuel type can help further classify automobiles and determine the price.

---

**Future Scope of my work** - Hello I am third year UG student at IIT Kharagpur. I am a Data Science enthusiast skilled in Python, data analysis, Statistical modeling, Machine Learning, and BI tools and techniques. I am open for the project and work related to Data Science and ML.