Microsoft

MICROSOFT ENGAGE 2022

BY

SOBHIKA MAHAJAN

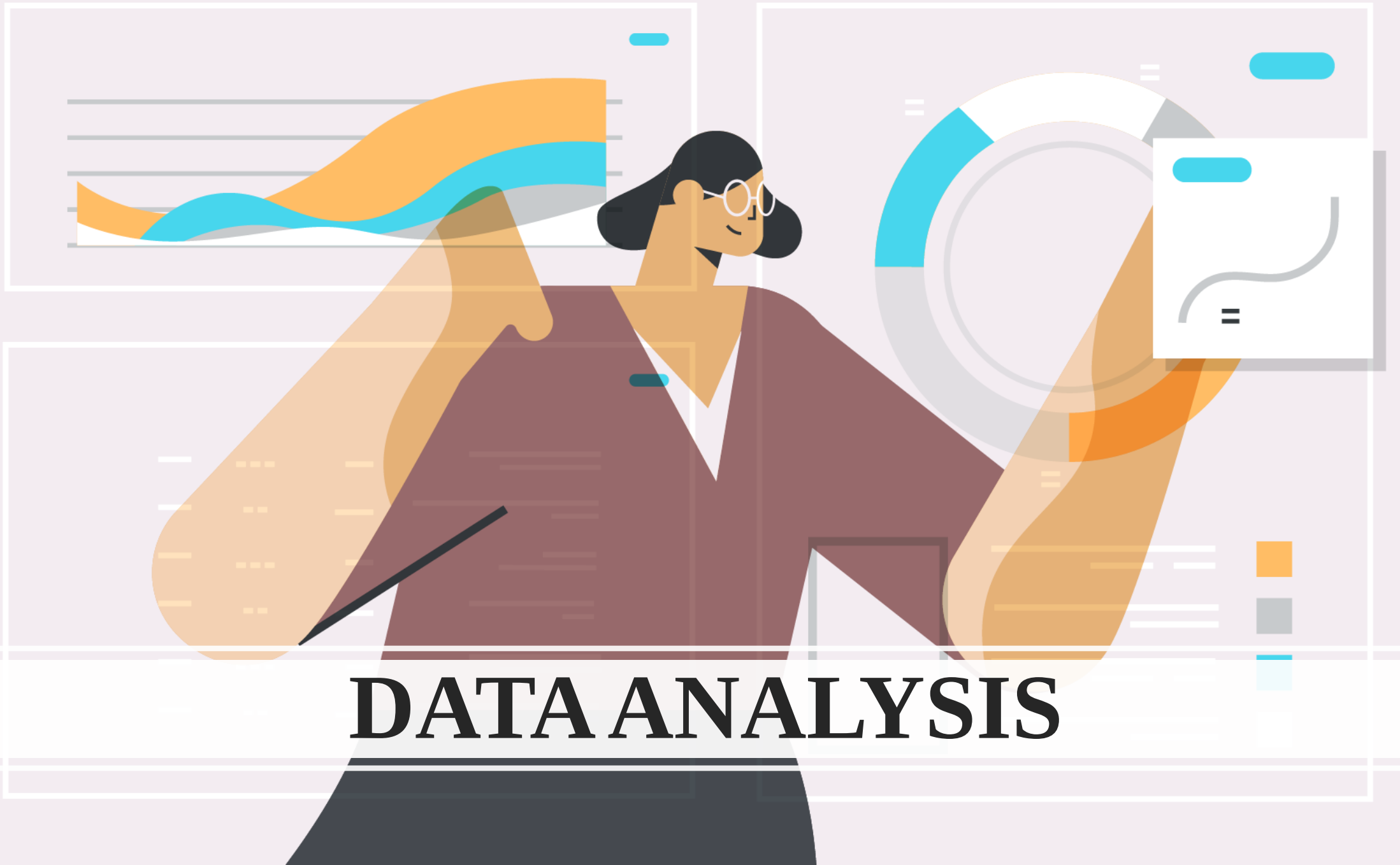
About Me

- My name is **Sobhika Mahajan**. I am currently pursuing B.Tech in **Computer Science** from **Banasthali Vidyapith**, Jaipur.

I love learning new technologies as well contributing my knowledge for betterment of society

LinkedIn:
https://www.linkedin.com/in/sobhika-mahajan-16776b1b8/

GitHub Repo:
https://github.com/SobhikaMahajan/Project



DATA ANALYSIS

About the Data

1. symbolling: -3, -2, -1, 0, 1, 2, 3.

make: alfa-romero, audi, bmw, chevrolet, dodge, honda, isuzu, jaguar, mazda, mercedes-benz, mercury, mitsubishi, nissan, peugot, plymouth, porsche, renault, saab, subaru, toyota, volkswagen, volvo

fuel-type: diesel, gas.

aspiration: std, turbo.

num-of-doors: four, two.

body-style: hardtop, wagon, sedan, hatchback, convertible.

drive-wheels: 4wd, fwd, rwd.

engine-location: front, rear.

wheel-base: continuous from 86.6 120.9.

length: continuous from 141.1 to 208.1.

About the Data

width: continuous from 60.3 to 72.3.	**height**: continuous from 47.8 to 59.8.	**curb-weight**: continuous from 1488 to 4066.	engine-type: dohc, dohcv, l, ohc, ohcf, ohcv, rotor.	**num-of-cylinders**: eight, five, four, six, three, twelve, two.
engine-size: continuous from 61 to 326.	**fuel-system**: 1bbl, 2bbl, 4bbl, idi, mfi, mpfi, spdi, spfi.	bore: continuous from 2.54 to 3.94.	stroke: continuous from 2.07 to 4.17.	compression-ratio: continuous from 7 to 23.
horsepower: continuous from 48 to 288.	**peak-rpm**: continuous from 4150 to 6600.	**city-mpg**: continuous from 13 to 49.	highway-mpg: continuous from 16 to 54.	**price**: continuous from 5118 to 45400.

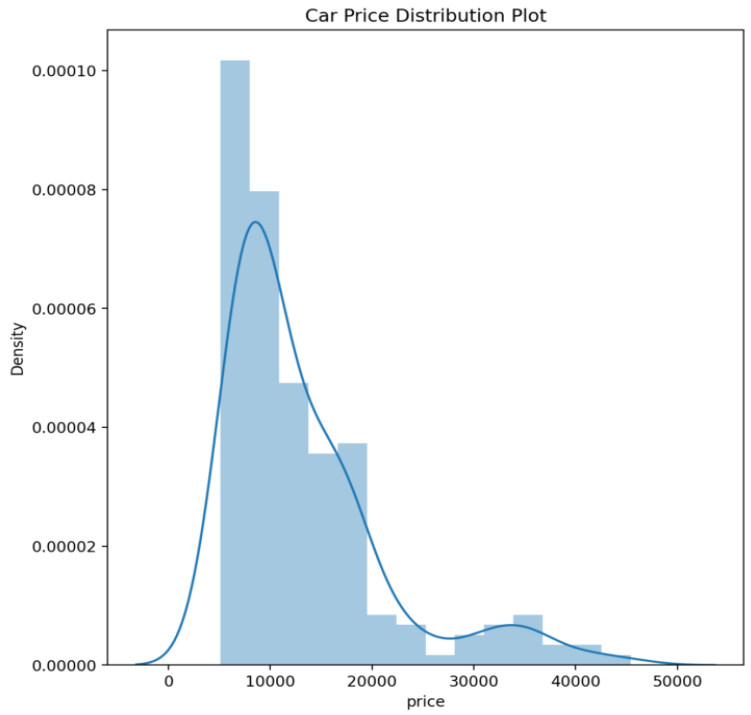
Data Preparation Steps

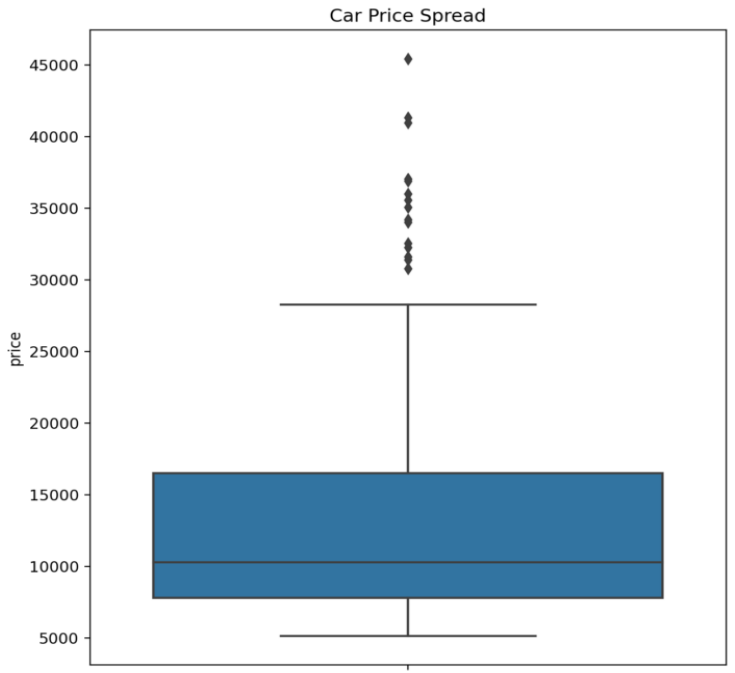
STEP	DETAILS
Variable Identification	Symbolling, Price, and other necessary variables to support initial hypothesis
Analysis	Refer Notebook (GitHub)
Treating missing values	Imputed missing/non-numeric values with mean of the group for continuous and mode of the group for categorical variables
Detecting, analyzing and treating outliers	Engine size and horsepower outliers are kept because they represent real world data
Deriving variables	New variables calculated for the purpose of analysis.

Car Price – Data Distribution

Distribution of Price:

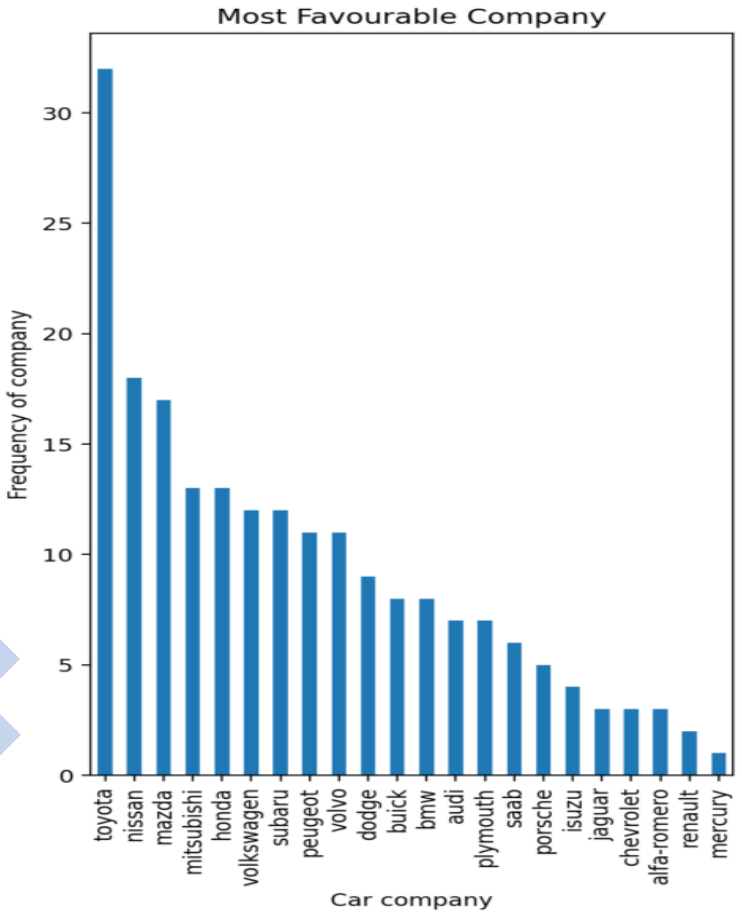
Majority of cars belong to the lower price brackets (< 15K) even though there are cars that go up to 45K

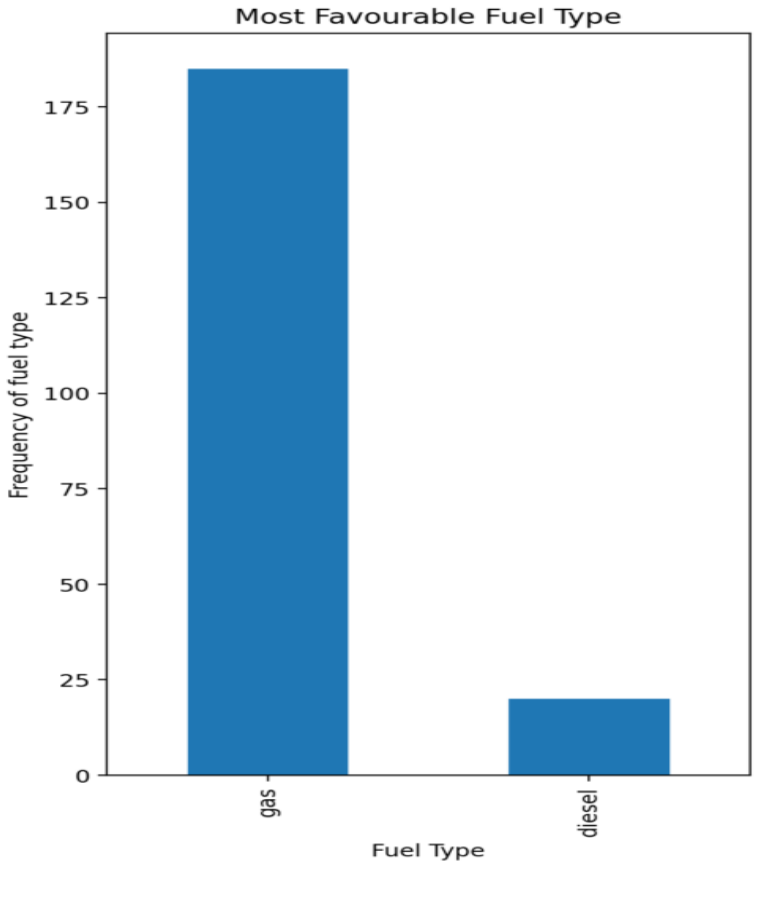


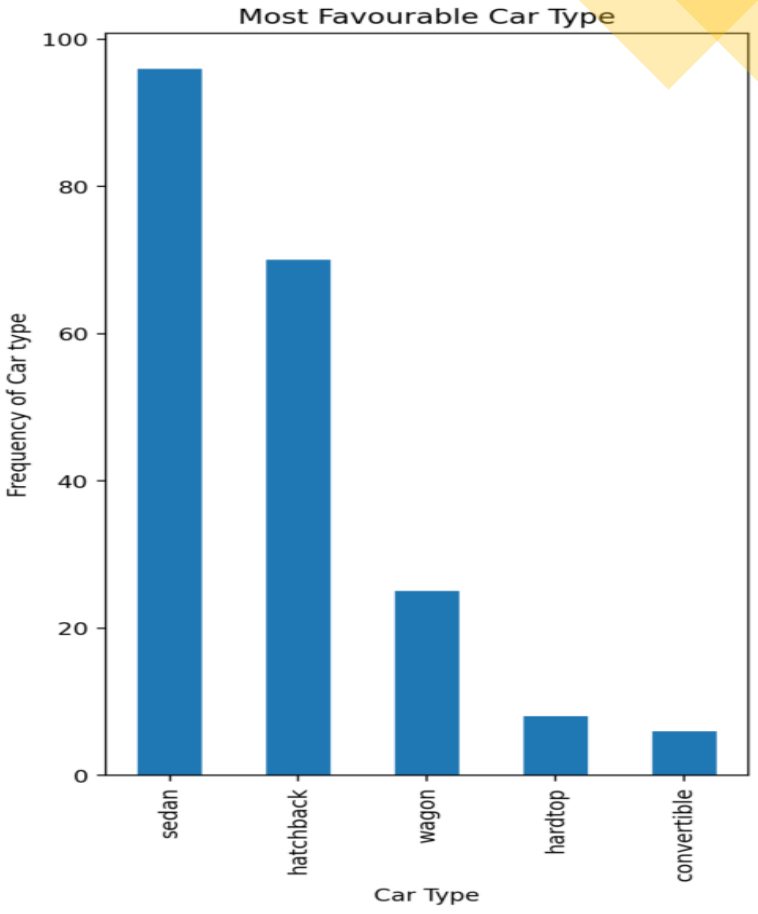


Favorable

Toyota seemed to be favored car company.
2. Number of gas fueled cars are more than diesel.
3. sedan is the top car type preferred.

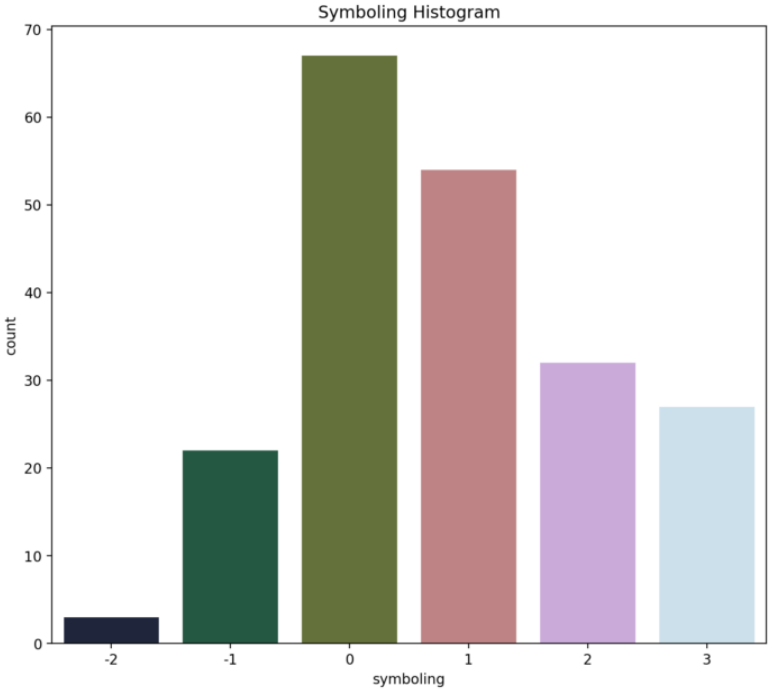


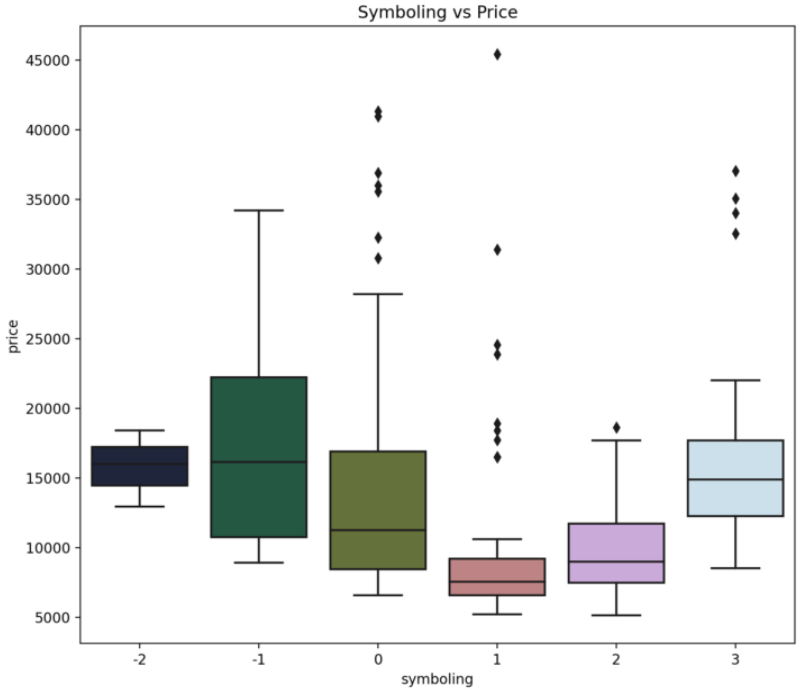




About Symbolling

- Symbolling value shows how risky or safe a vehicle is, from an insurer's perspective. It can range from -3 to +3.
- -3 indicates a safe car while +3 denotes a risky one.

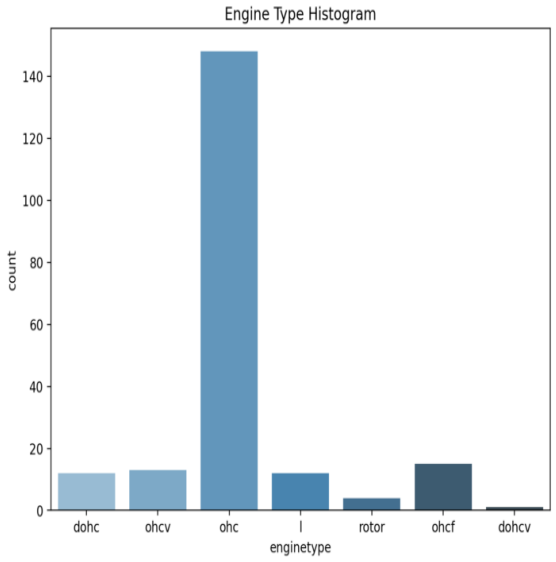


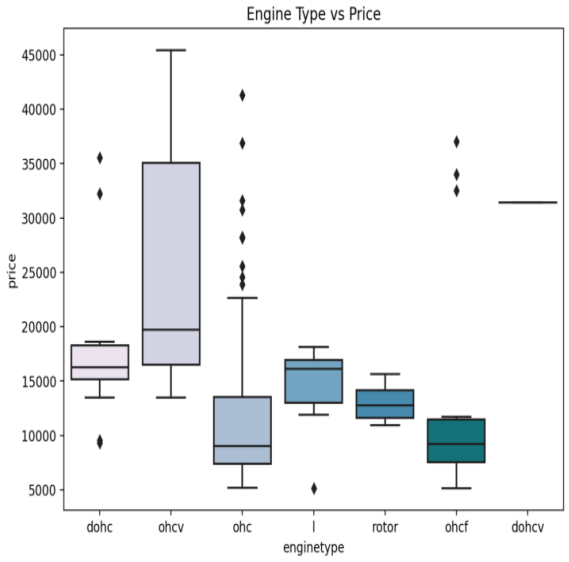


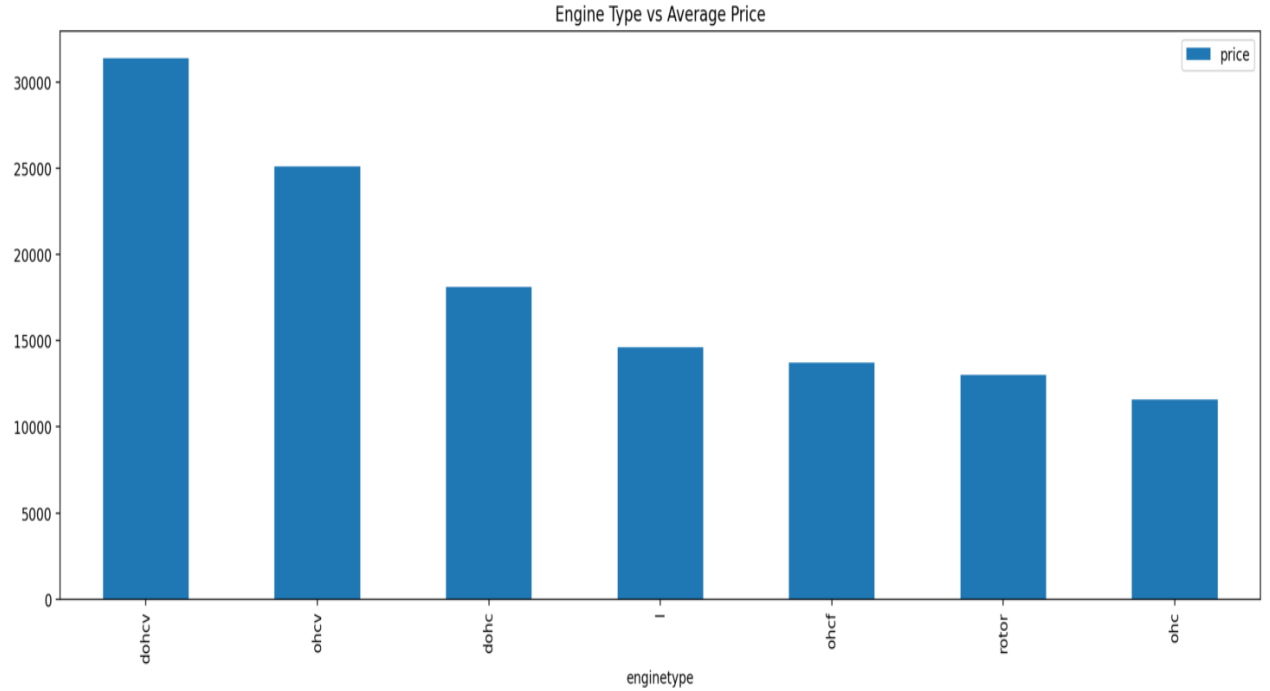
ENGINE SIZE

Car pricing maintains strong positive correlation with its engine size.

1. ohc Engine type seems to be most favored type.
2. ohcv has the highest price range (While dohcv has only one row), ohc and ohcf have the low price range.

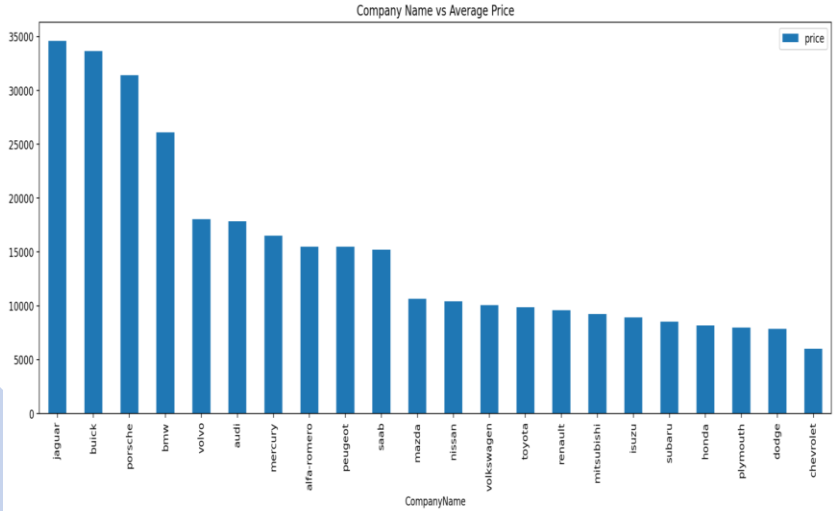


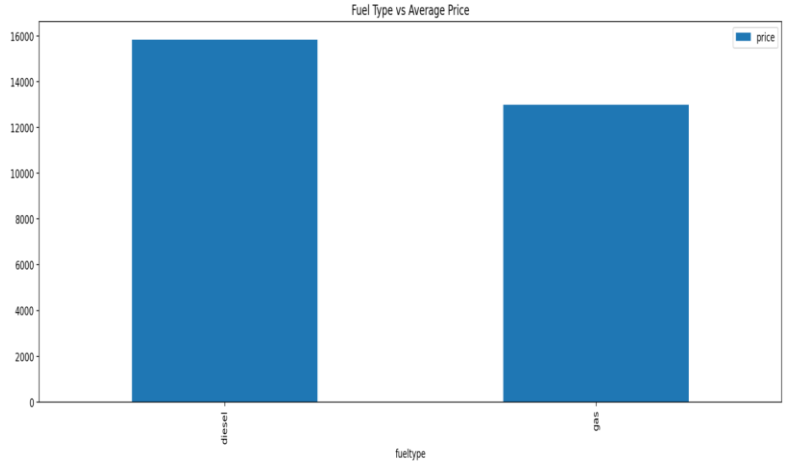


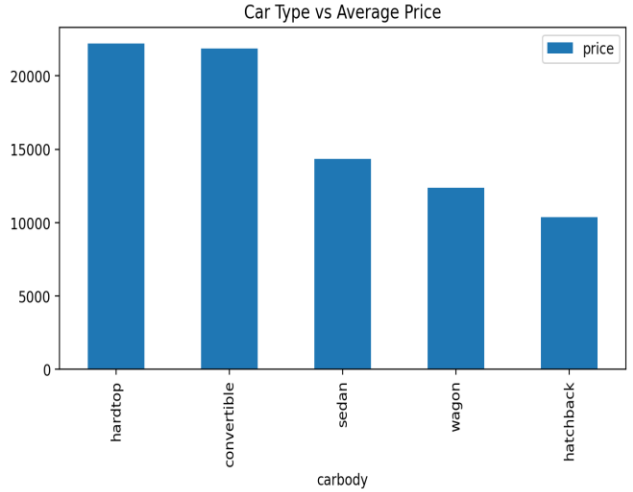


AVERAGE SIZE

- 1. Jaguar and Buick seem to have highest average price.
- 2. diesel has higher average price than gas.
- 3. hardtop and convertible have higher average price.

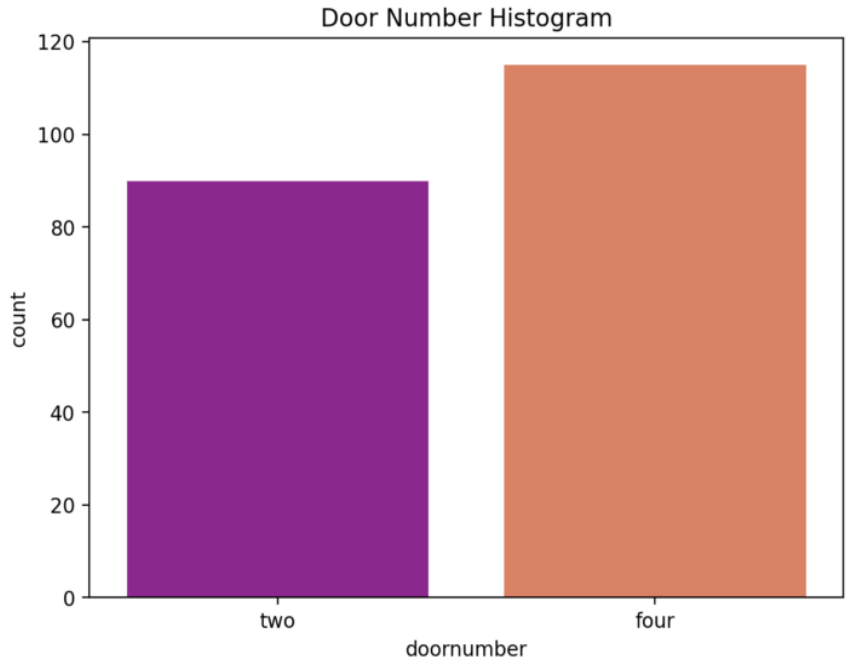


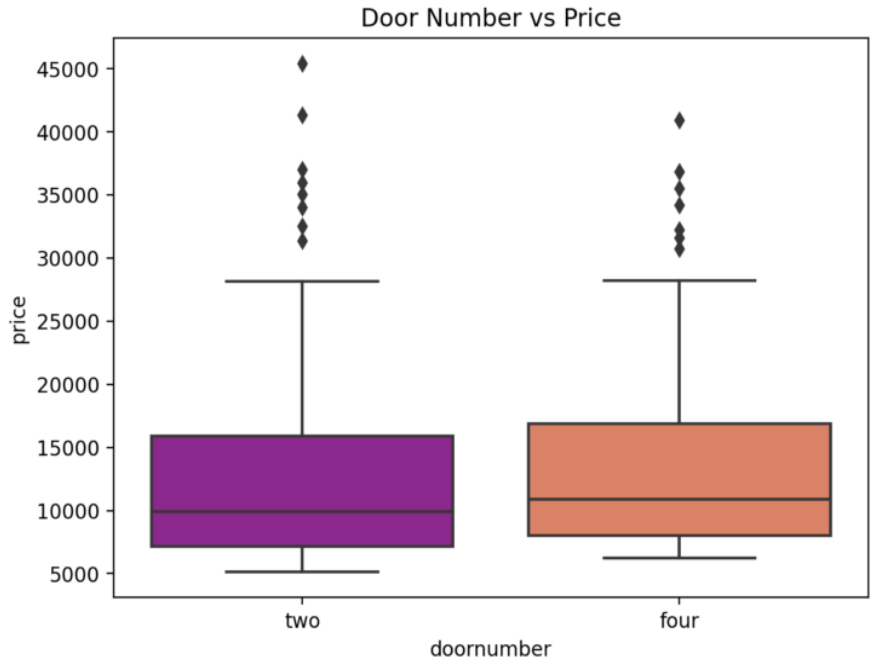




Doors

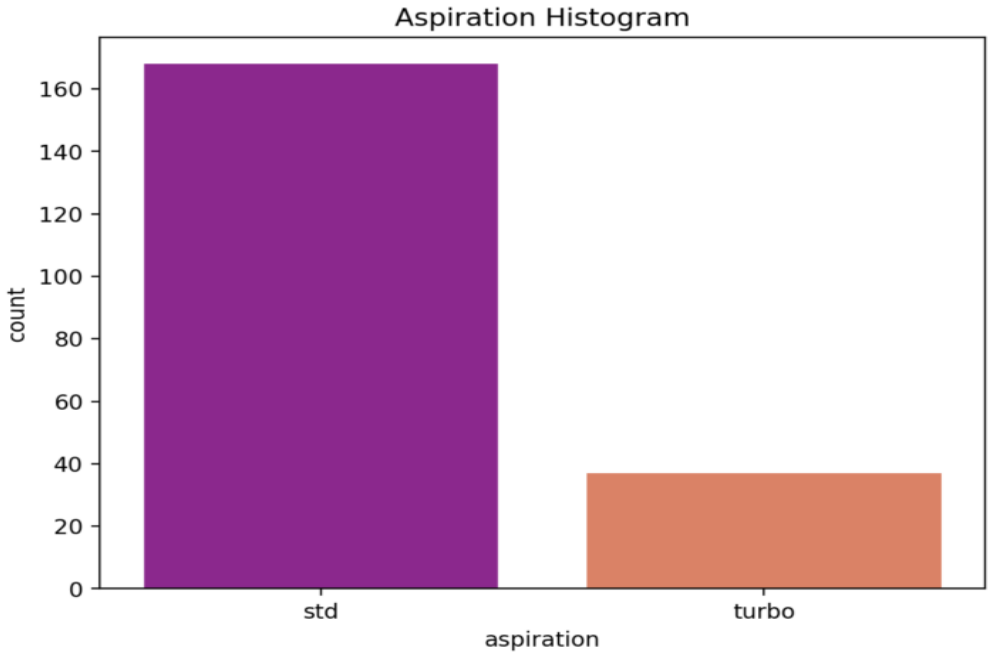
1. Vast Majority of safe cars have four doors.
2. Door number variable is not affecting the price much. There is no significant difference between the categories in it.

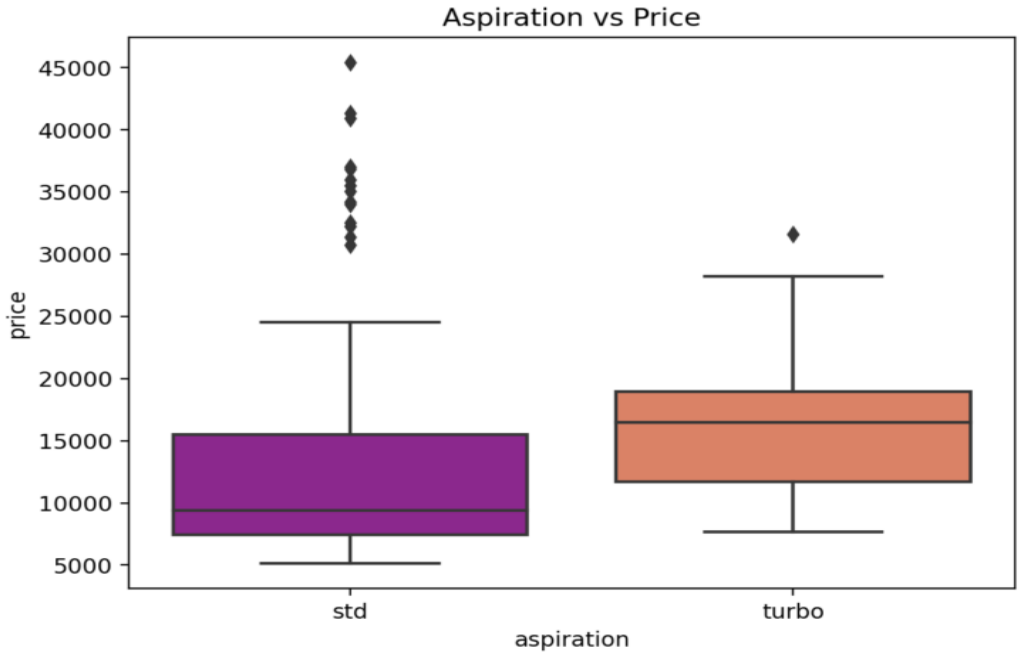




Aspiration

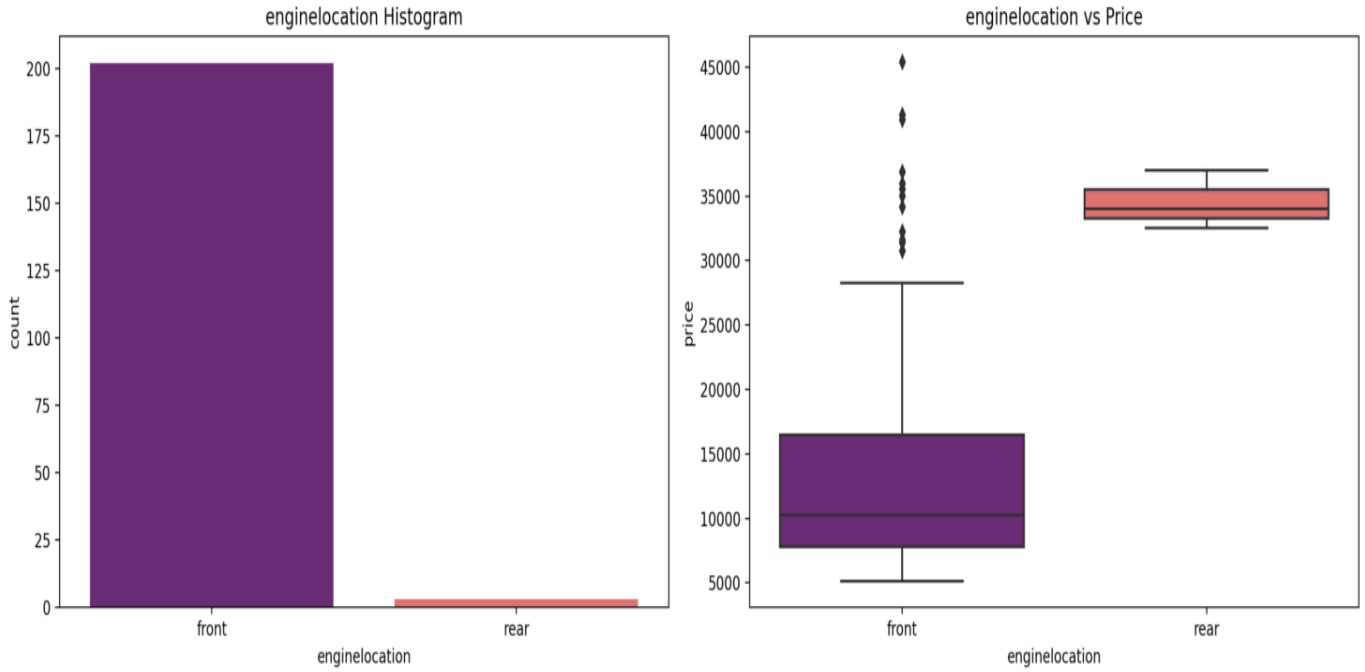
It seems aspiration with turbo have higher price range than the std(though it has some high values outside the whiskers.)





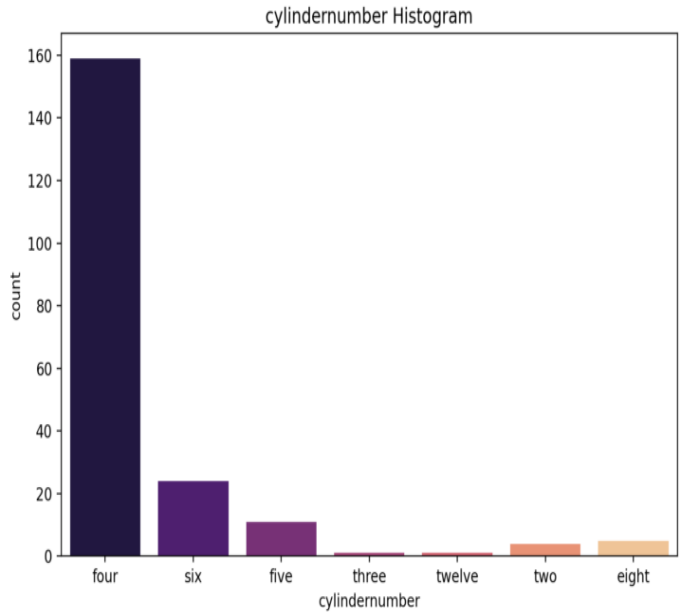
Engine Location

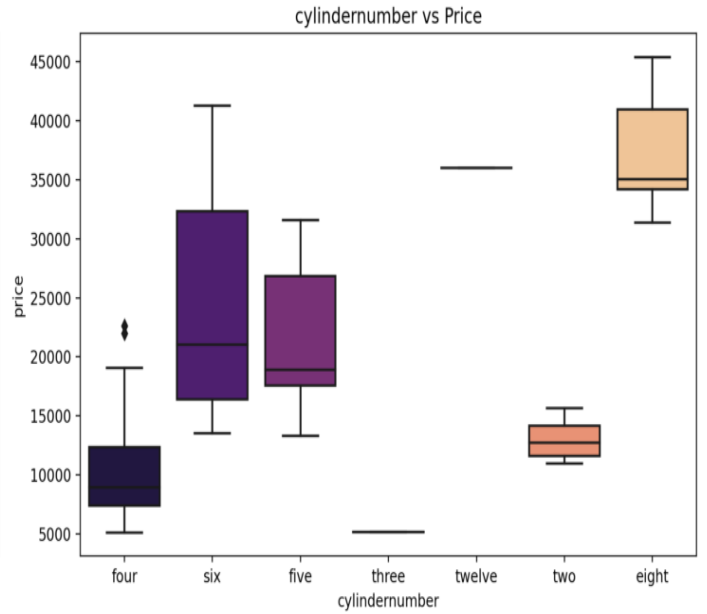
- Very few datapoints for Engine Location categories to make an inference.



Cylinder Number

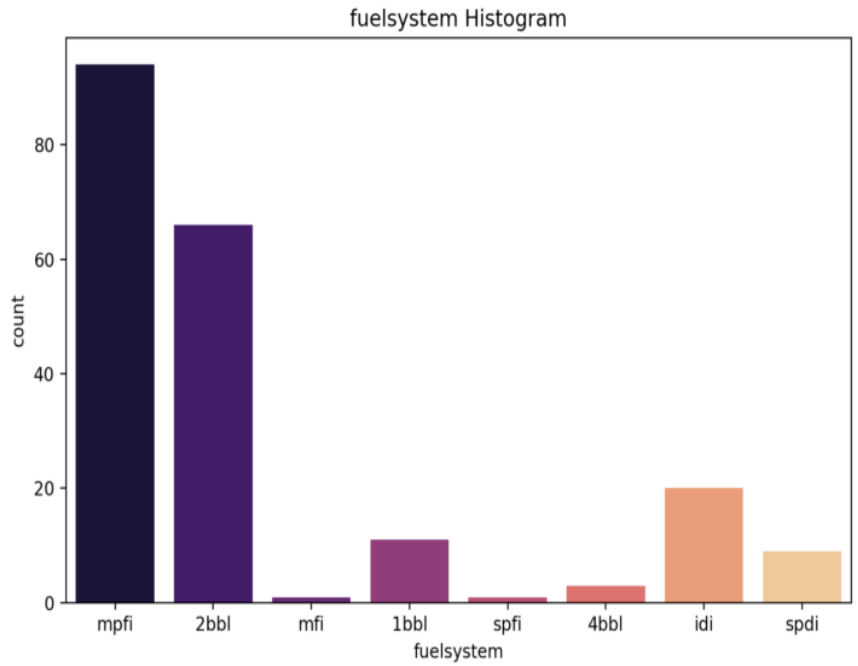
Most common number of cylinders are four, six and five. Though eight cylinders have the highest price range.

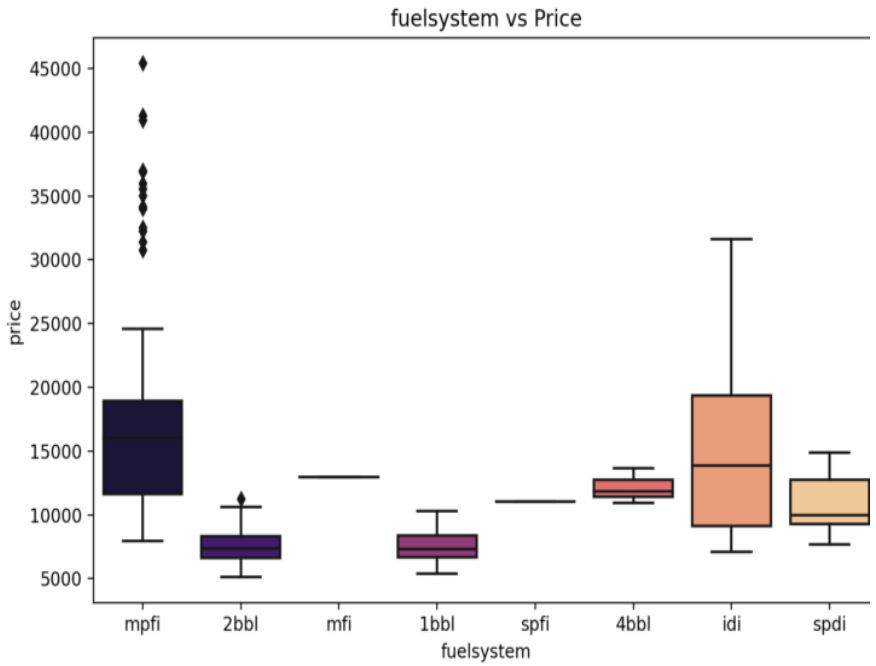




Fuel System

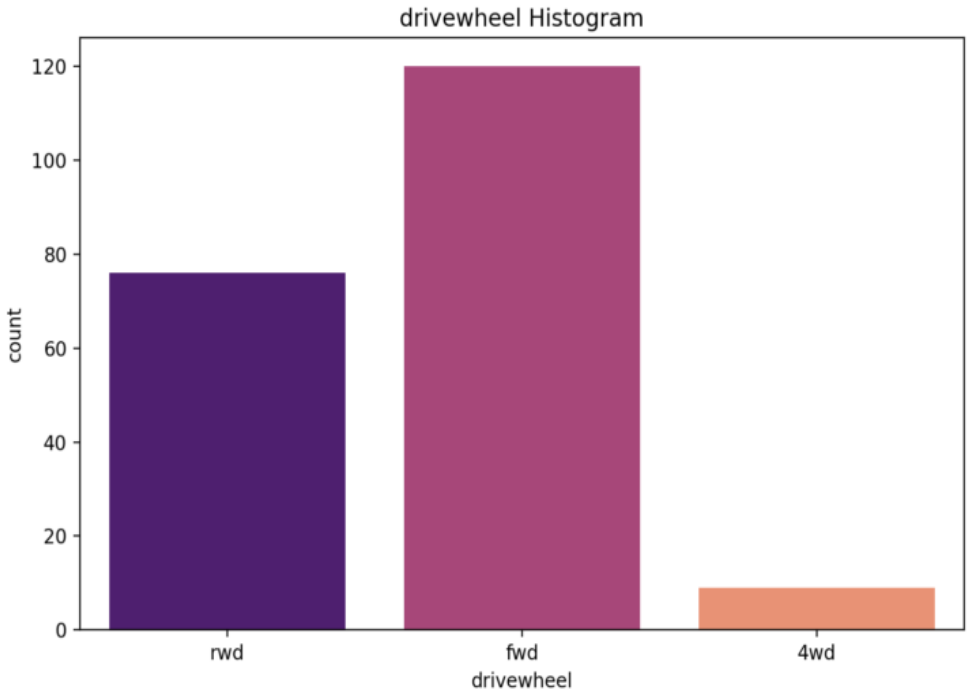
- mpfi and 2bbl are most common type of fuel systems. mpfi and idi having the highest price range. But there are few data for other categories to derive any meaningful inference

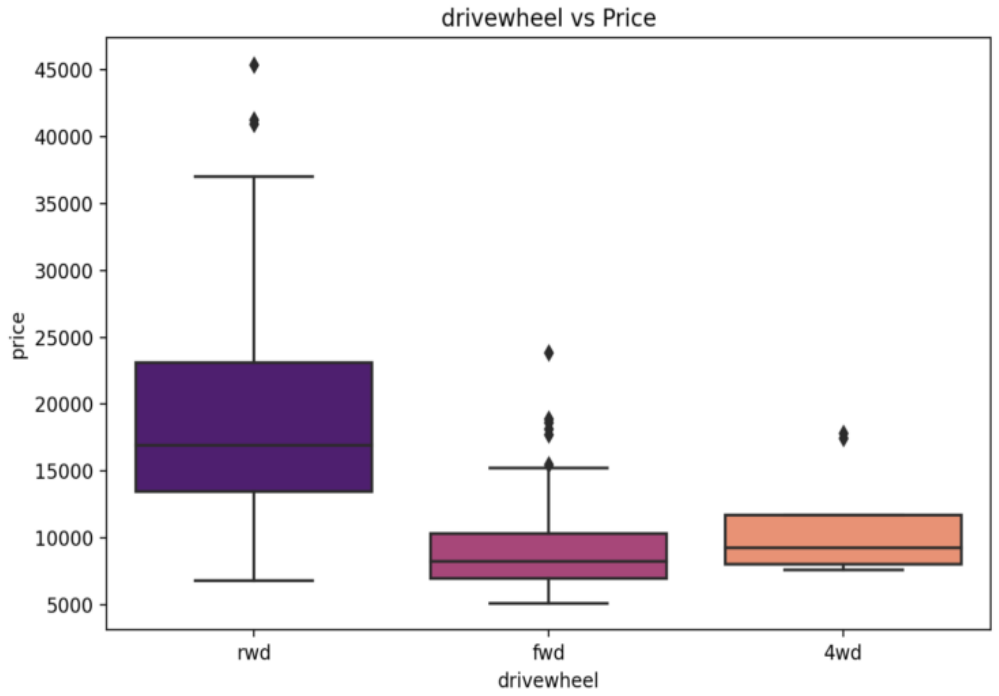




Drive Wheel

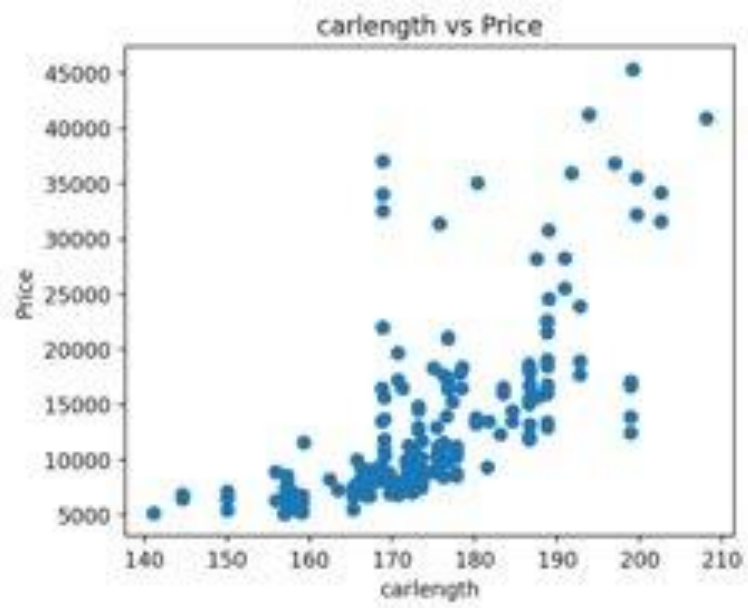
- A very significant difference in drivewheel category. Most high ranged cars seeme to prefer rwd drivewheel.

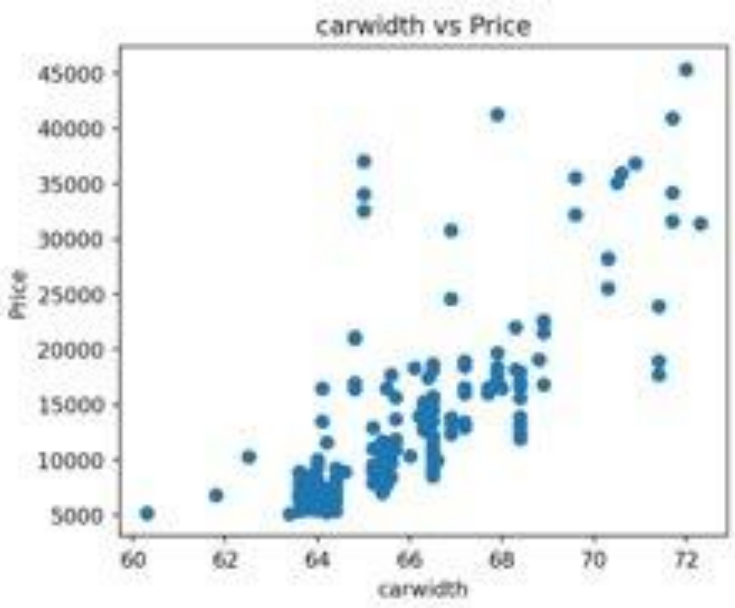


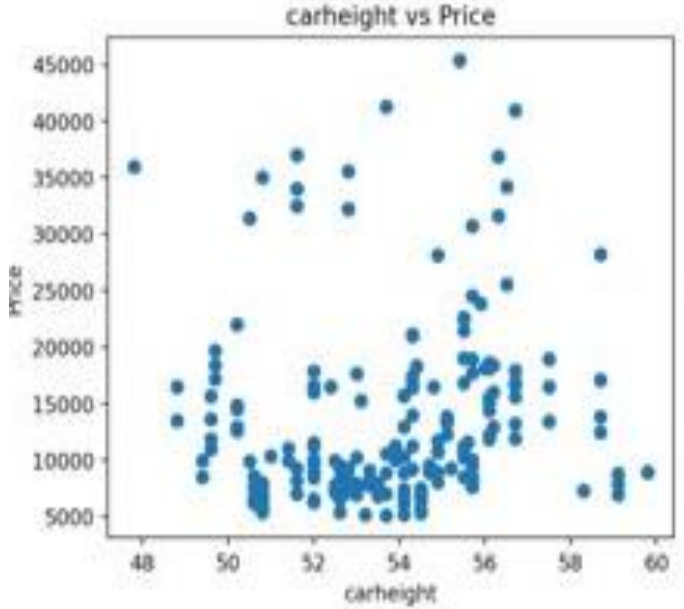


Car Length, Car Width, Car Height

Car Length, Car Width and Car Height seems to have a positive correlation with price.

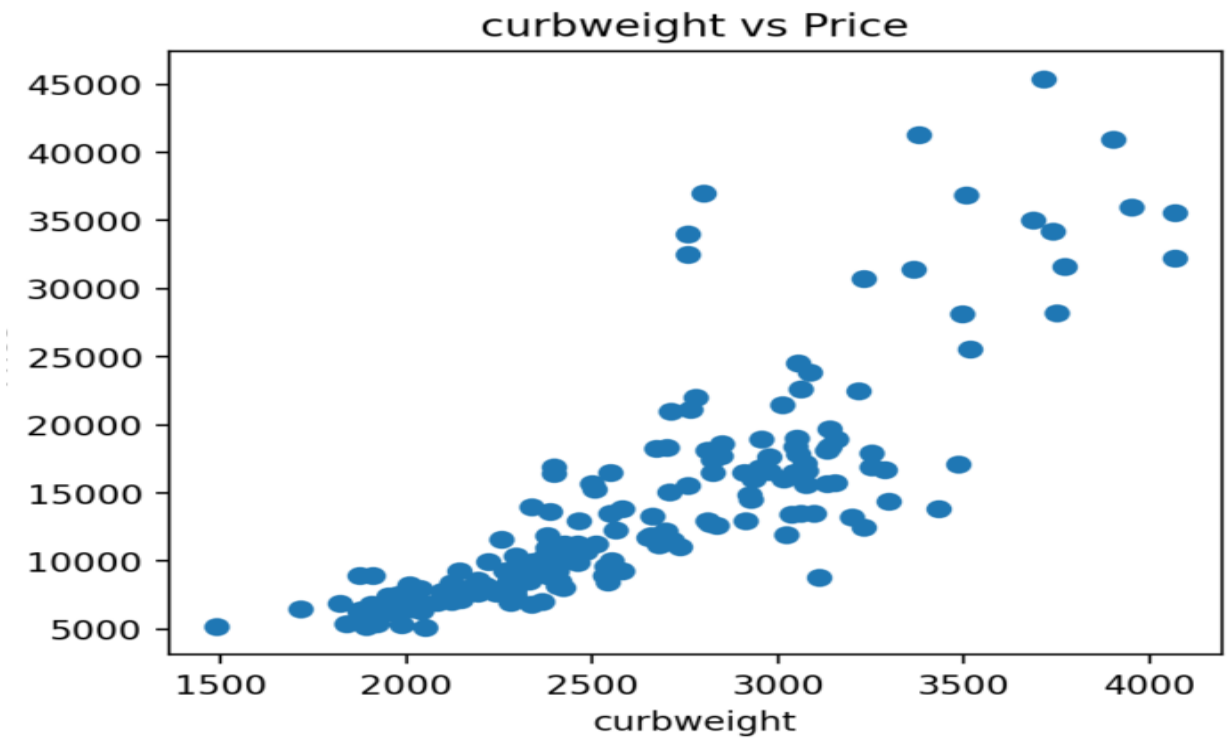






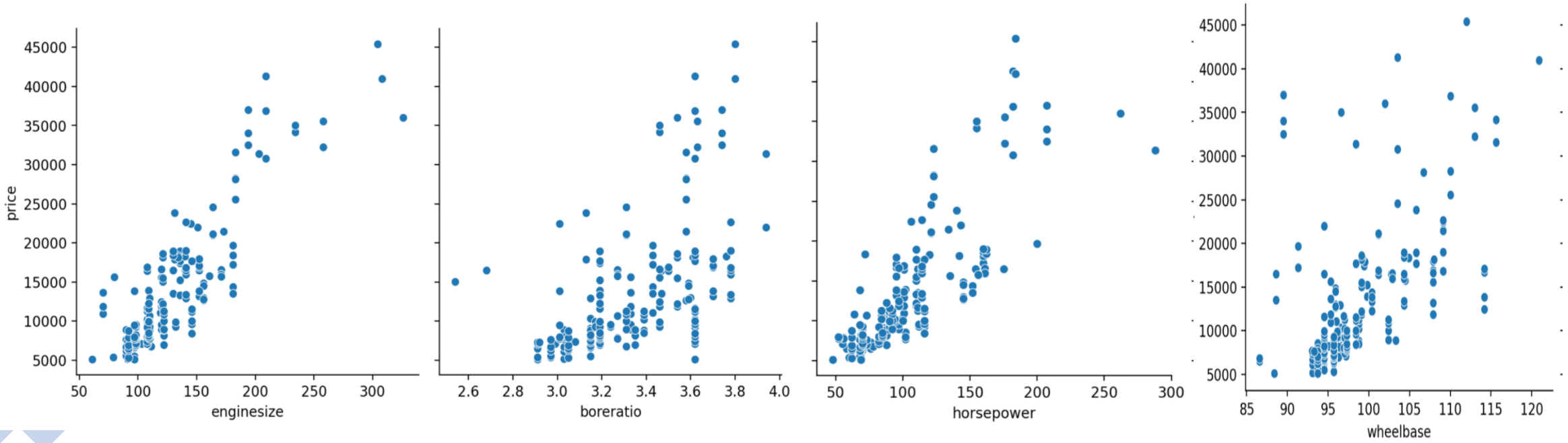
Car Height

Car Height does not show any significant trend with price



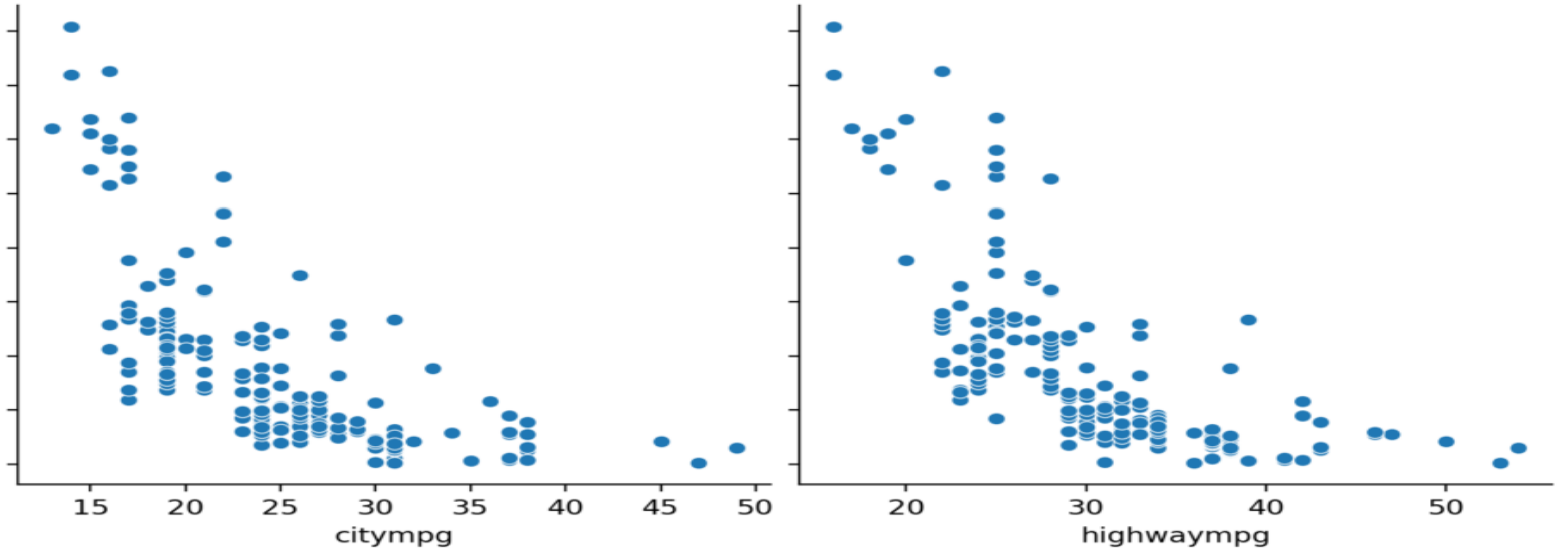
Engine Size, Bore Ratio, Horsepower, Wheelbase

Engine Size, Bore Ratio, Horsepower, Wheelbase - seem to have a significant positive correlation with price.



City mpg, Highway mpg

City mpg, Highway mpg - seem to have a significant negative correlation with price.



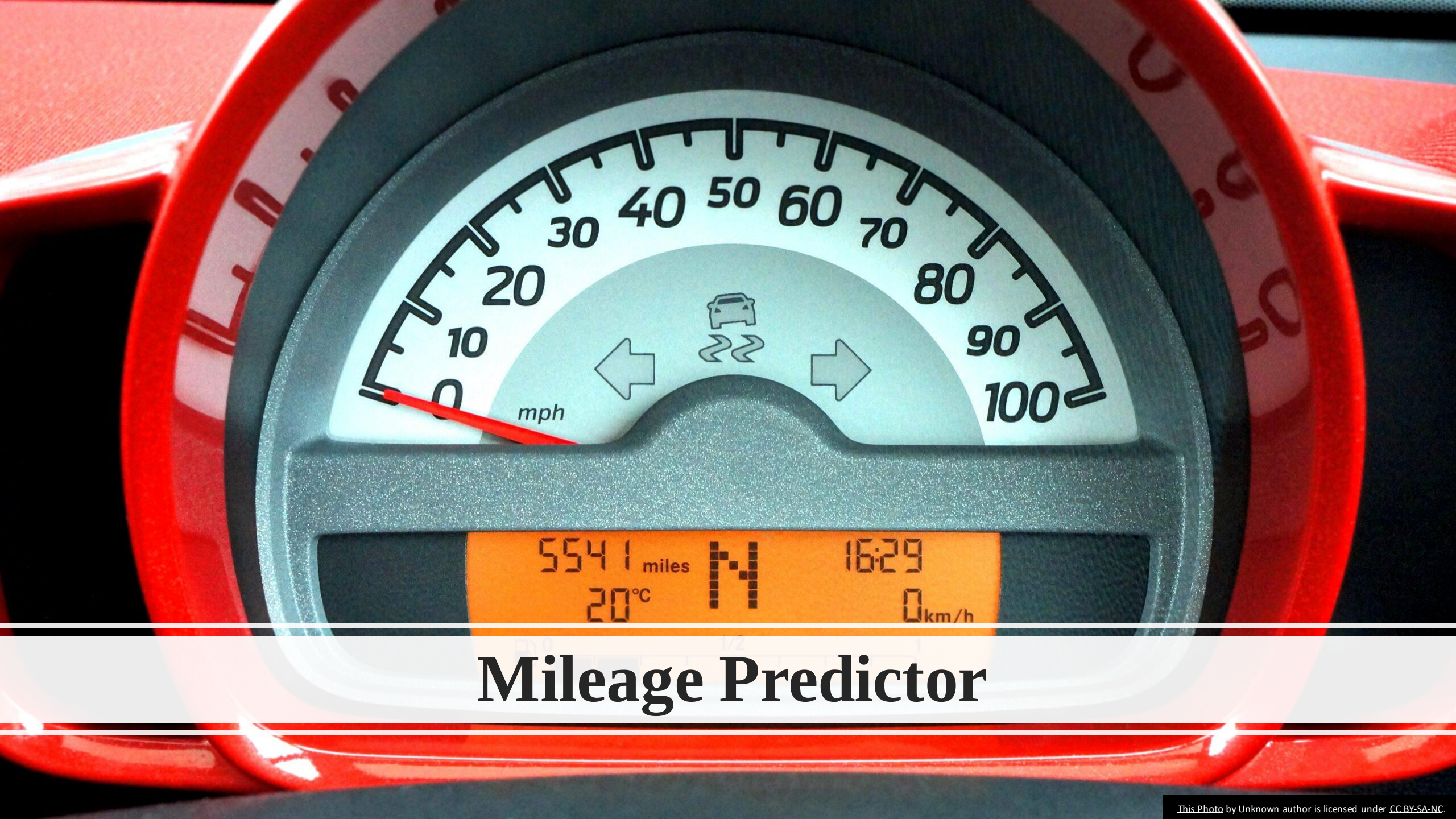
Car Price Prediction Model

About

The car price prediction model takes the following parameters :

- In which year car was purchased ?
- What is the current ex-showroom price of the car ? (In lakhs)
- What is distance completed by the car in Kilometers ?
- The number of owners the car had previously ?
- What is the fuel type of the car ?
- Are you a dealer or an individual ?
- What is the Transmission Type ?

Based on these inputs, the prediction model predicts and shows the estimated price we will get for selling the car



Mileage Predictor

About

The car price prediction model takes the following parameters :

- What is the number of Cylinders in cc?
- What is the Displacement in cc?
- What is the Horsepower in watts?
- What is the Weight in Kg?

Based on these inputs, the prediction model predicts and shows the estimated price we will get for selling the car

thank you