

# Indian Automobile Market Analysis



# Group 3

Srujana B

Kuldeep G S

Lakshmi Keerthi B





## Agenda:

Specification Requirement

Python for Analysis

Excel

Power BI visualization

Git Links

Reference

# Requirement Specification :

## ***Hardware Requirement:***

Windows11 Os' with i5 processor and above.

## ***Software Requirement :***

*Python Compiler*

*Excel*

*Power BI*



# Python for Analysis:

## *Exploratory Data Analysis(EDA):*

- EDA is a method of analyzing data sets to understand their main characteristics
- It involves summarizing data features, detecting patterns and uncovering relationship through visual and statistical techniques.

### **1. Import python libraries**

- Import all libraries which are required for our analysis. Such as Data Loading , Statistical analysis etc
- Pandas and Numpy have been used for data manipulation and numerical calculation.
- Matplotlib and seaborn have been used for data visualization.

#### ***Syntax:***

```
import pandas as pd  
import numpy as np  
import matplotlib.pyplot as plt  
import seaborn as sns
```



## 2. Reading Dataset:

- The pandas library offers a wide range of possibilities for loading data into the pandas data frame from files like JSON, .csv, .xlsx etc
- Most of the data are available in a tabular format of csv files.
- Using the read.csv() function, data can be converted to a pandas data frame.



```
[ ] df=pd.read_csv("/content/cars_ds_final (1).csv")
```

	Unnamed: 0	Make	Model	Variant	Ex-Showroom_Price	Displacement	Cylinders	Valves_Per_Cylinder	Drivetrain	Cylinder_Configuration
0	0	Tata	Nano Genx	Xt	Rs. 2,92,667	624 cc	2.0	2.0	RWD (Rear Wheel Drive)	Ir
1	1	Tata	Nano Genx	Xe	Rs. 2,36,447	624 cc	2.0	2.0	RWD (Rear Wheel Drive)	Ir
2	2	Tata	Nano Genx	Emax Xm	Rs. 2,96,661	624 cc	2.0	2.0	RWD (Rear Wheel Drive)	Ir

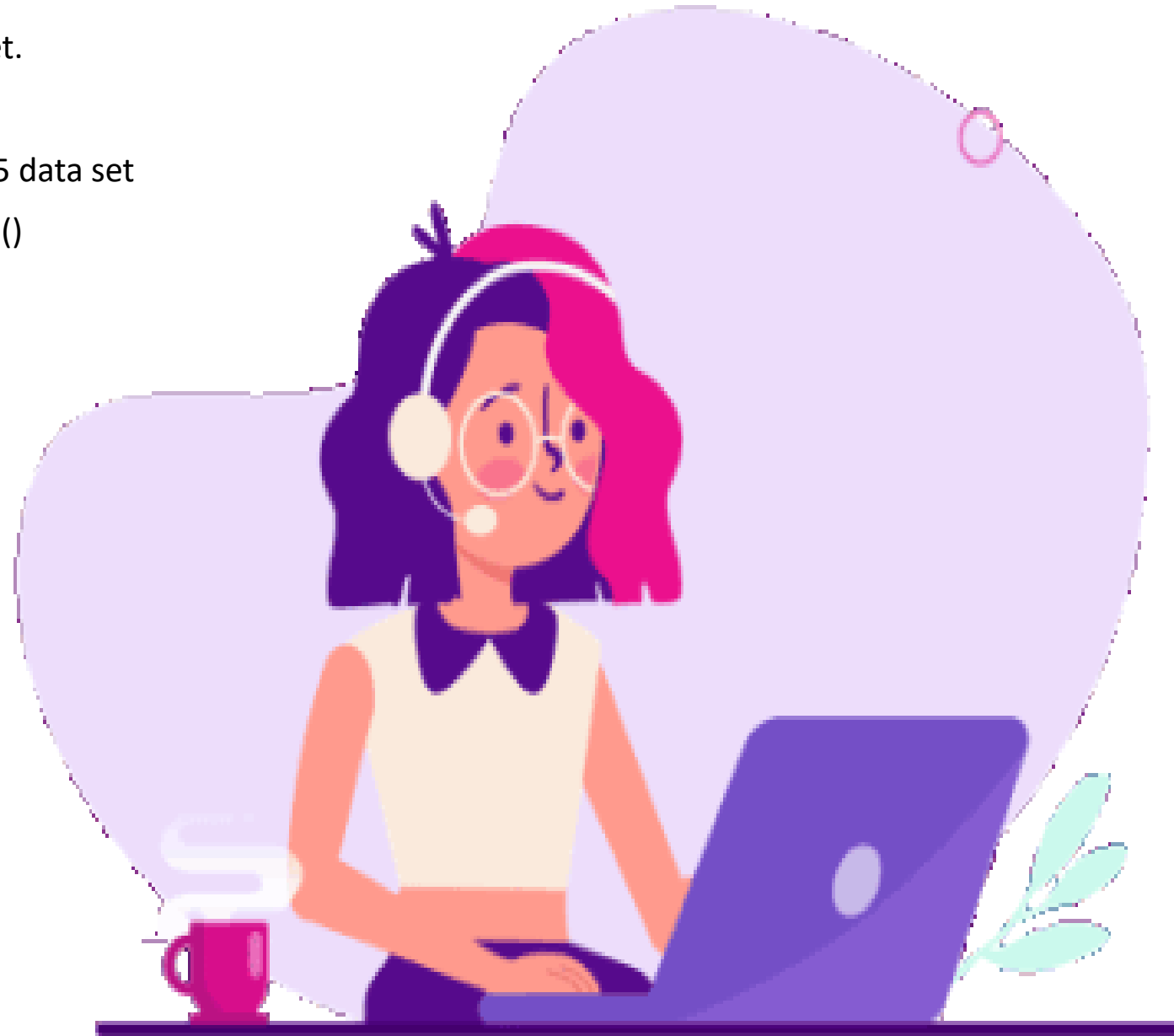
RWD (Rear

### 3.Analyzing the Data:

- df.Shape displays the rows and columns in the dataset.
- We have 1276 rows and 141 columns.
- df.head() displays top 5 dataset, df.tail() displays last 5 data set
- For Checking duplications we use df.duplicated().sum()
- Unwanted columns were removed
- 1 duplicate column were found
- For Checking missing values we use df.isnull().sum()

```
plt.show() #display
```

Unnamed: 0	0.000000
Make	5.877743
Model	0.000000
Variant	0.000000
Ex-Showroom_Price	0.000000
...	
USB_Ports	97.727273
Heads-Up_Display	96.003135
Welcome_Lights	94.592476
Battery	98.981191
Electric_Range	98.667712
Length: 141, dtype: float64	





## Descriptive Statistics:

It describes categorical data within a DataFrame using the `describe()` method in pandas. By specifying `include=object`, it focuses on object-type columns, such as strings or categorical variables, and provides summary statistics like count, unique values, top value, and frequency of the top value.

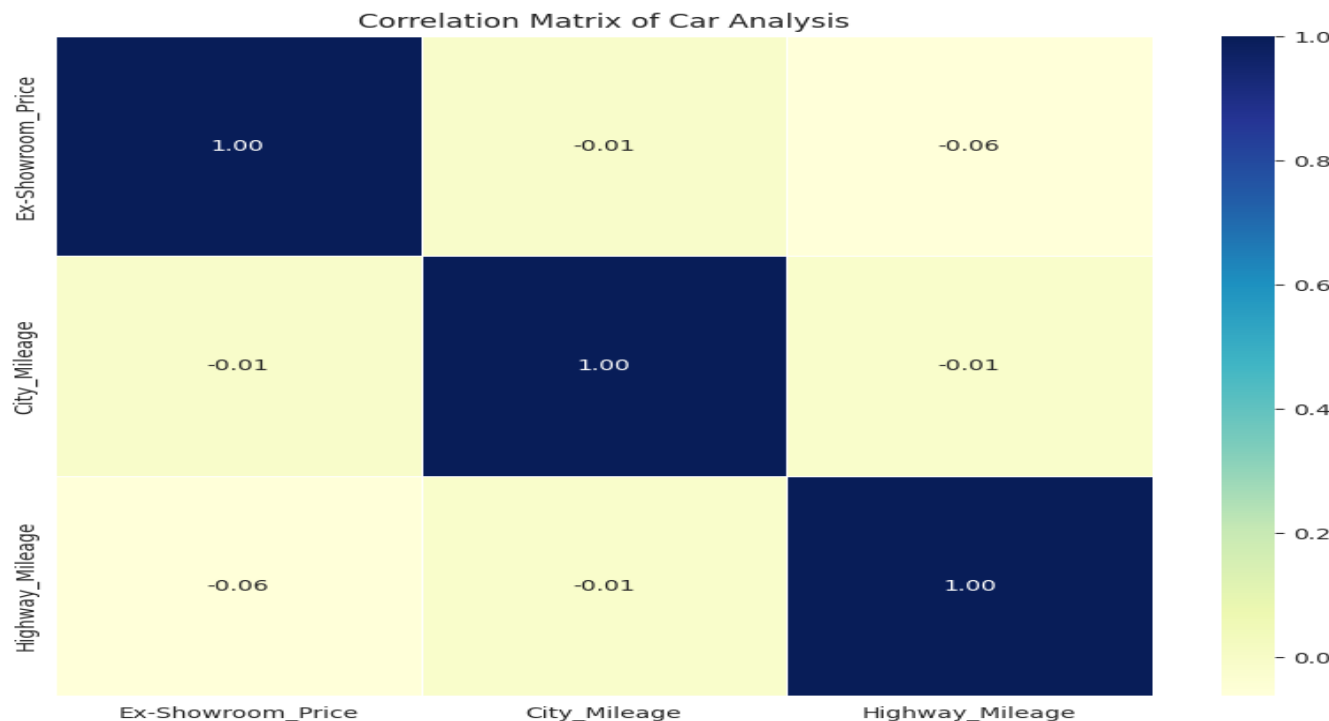
	count	mean	std	min	25%	50%	75%	max
NA	1276.0	6.375000e+02	3.684938e+02	0.0	318.75	637.5	956.25	1275.0
Ex-Showroom_Price	1276.0	4.596538e+06	1.214735e+07	236447.0	743876.00	1060064.5	2979827.75	212155397.0
Cylinders	1210.0	4.380992e+00	1.660957e+00	2.0	4.00	4.0	4.00	16.0
Valves_Per_Cylinder	1174.0	3.977853e+00	8.337633e-01	1.0	4.00	4.0	4.00	16.0
Doors	1272.0	4.550314e+00	7.478161e-01	2.0	4.00	5.0	5.00	5.0
City_Mileage	1277.0	2.931722e+01	7.282939e+02	0.0	0.00	8.5	16.00	26032.0
Seating_Capacity	1270.0	5.270079e+00	1.145231e+00	2.0	5.00	5.0	5.00	16.0
Number_of_Airbags	1141.0	3.787029e+00	2.522399e+00	1.0	2.00	2.0	6.00	14.0





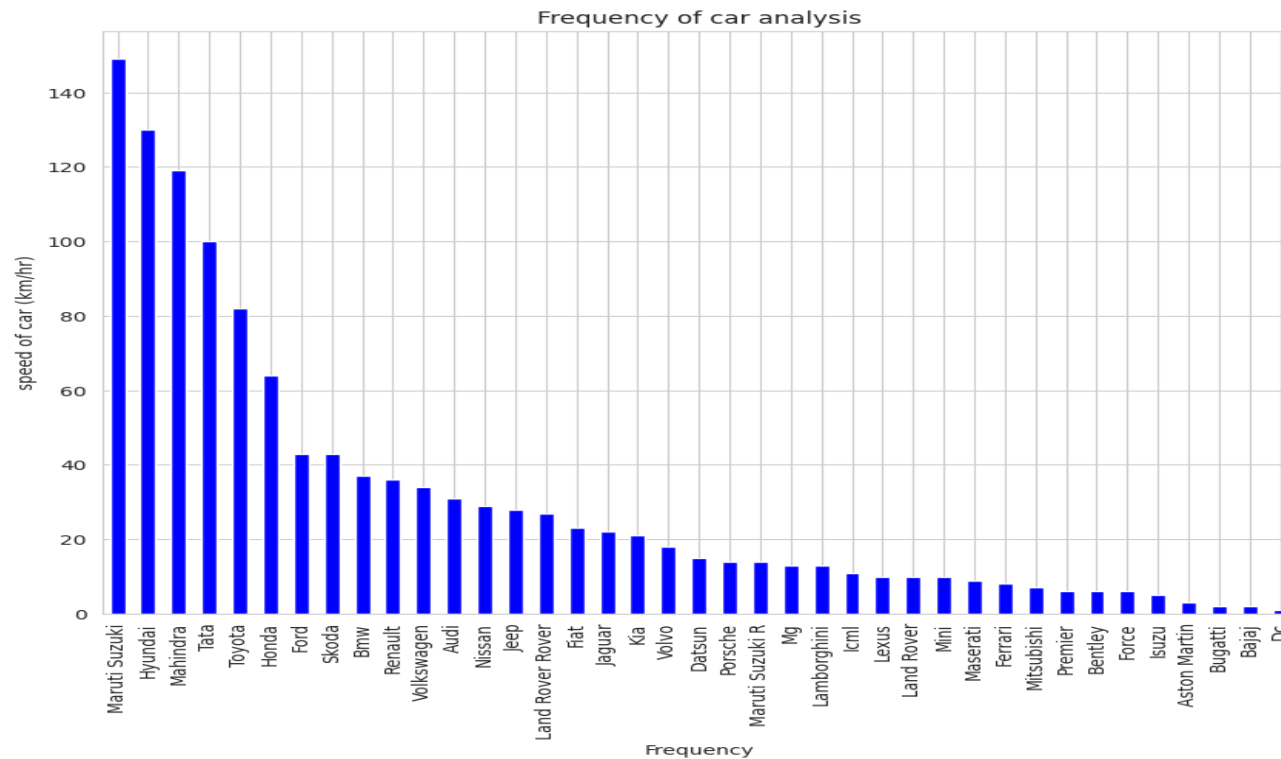
## Correlation matrix:

- Correlation measures the strength and direction of a linear relationship between two variables.
- A correlation matrix is a table showing correlation coefficients between sets of variables.
- The color coding helps visualize the strength and direction of correlations:
  - Dark blue represents strong positive correlation (closer to 1).
  - Light colors (yellow to green) represent weaker correlations, both positive and negative.
  - Since most of your correlations are close to zero, the cells are light-colored.



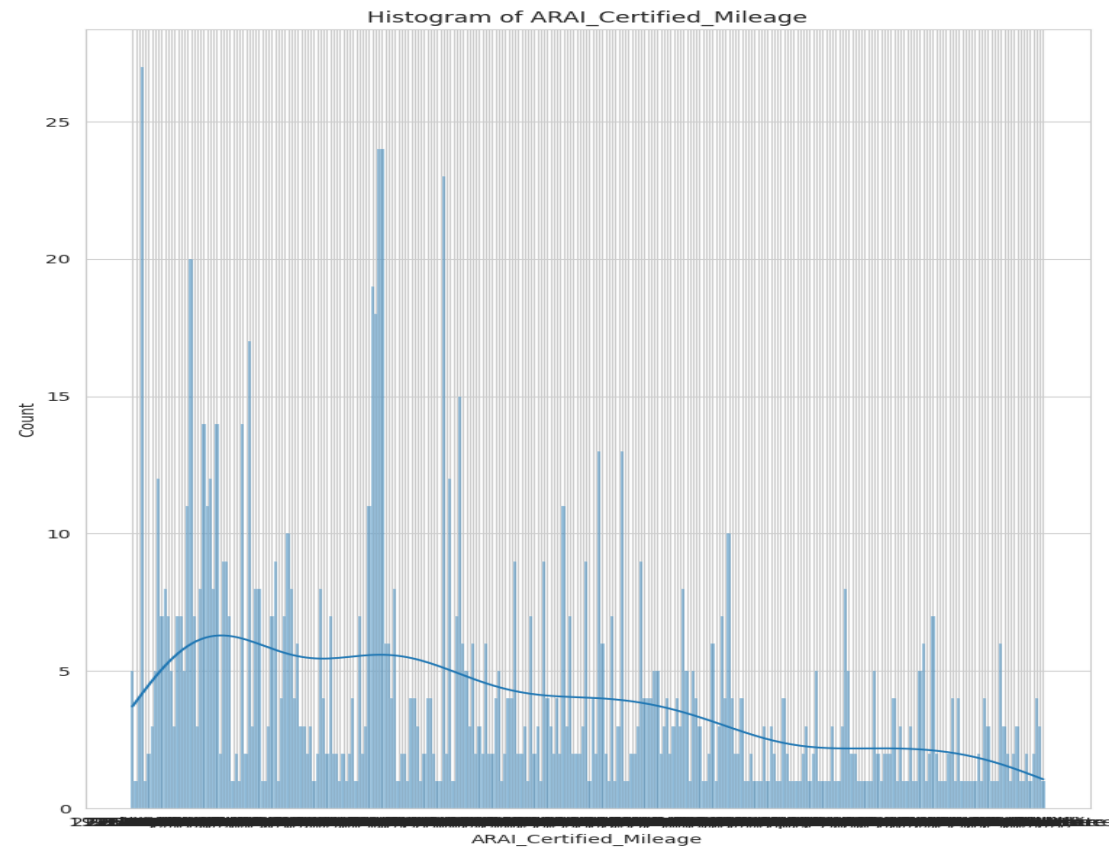
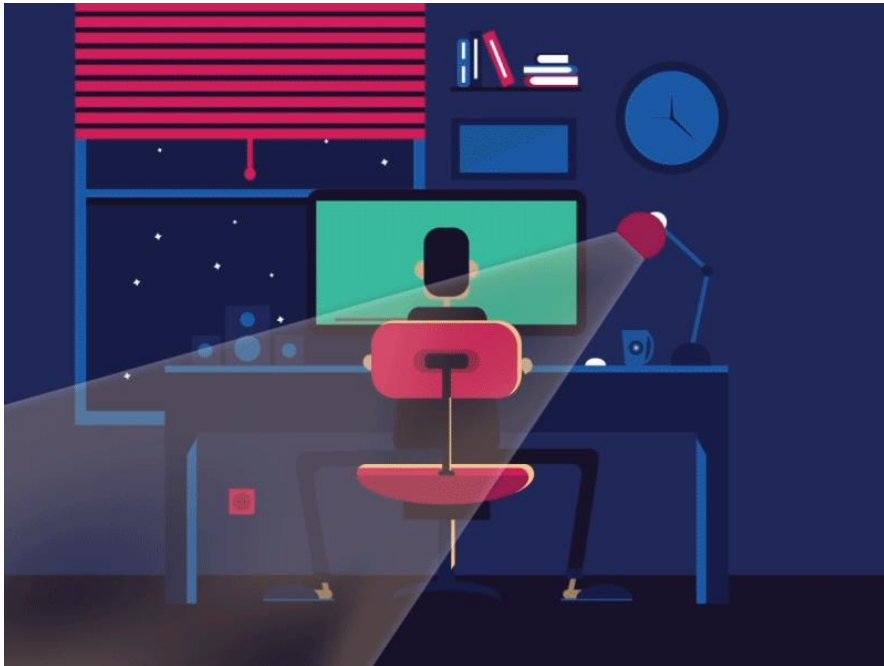
## Frequency of car analysis:

- The frequency of cars from various brands based on their speed in km/hr, highlighting which brands have the most and least cars recorded at certain speeds.
- Brands like Maruti Suzuki, Hyundai, and Mahindra have the highest frequencies of cars with recorded speeds, while brands like Bajaj and DC have the lowest.



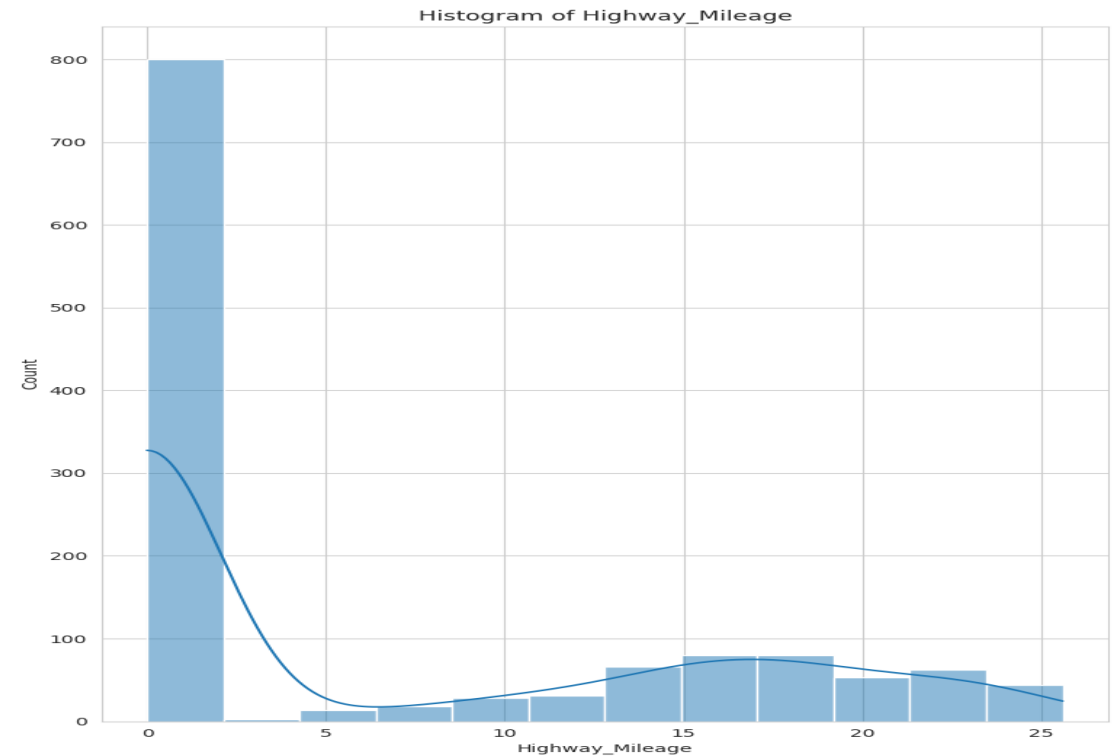
## ***Distribution of ARAI Certified Mileage:***

- There are many distinct mileage values with varying frequencies.
- The distribution appears to be right-skewed, with a higher frequency of cars having lower ARAI-certified(Automotive Research Association of India Research Institute of India) mileage values.
- Most cars have lower certified mileage.



## ***Distribution of Highway Mileage:***

- The distribution shows a large number of cars with very low highway mileage.
- The majority of cars have a highway mileage around 0-5 km/l, with fewer cars achieving higher highway mileage values.
- The distribution is heavily right-skewed, with a sharp drop in the number of cars as the highway mileage increases.
- most cars are not highly fuel-efficient on highways.



## Outliers Detection:

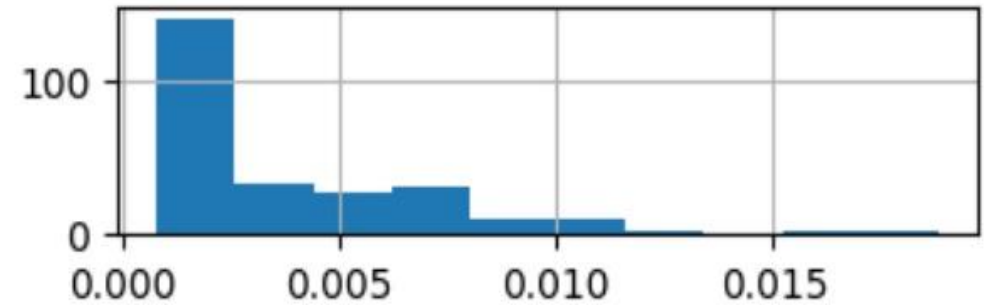
- Getting the outlier using frequency method
- We found very few outliers at the end

Outlier in the Model:

```
Model
Nexon      0.018809
Kuv100 Nxt 0.017241
Xuv500     0.016458
Compass    0.016458
Amaze      0.015674
...
Quattroporte 0.000784
M5           0.000784
S60 Cross Country 0.000784
Mercedes-Benz Amg Gt 4-Door Coupe 0.000784
A5           0.000784
Name: proportion, Length: 263, dtype: float64
```

Series: Model\_frequency

proportion float64



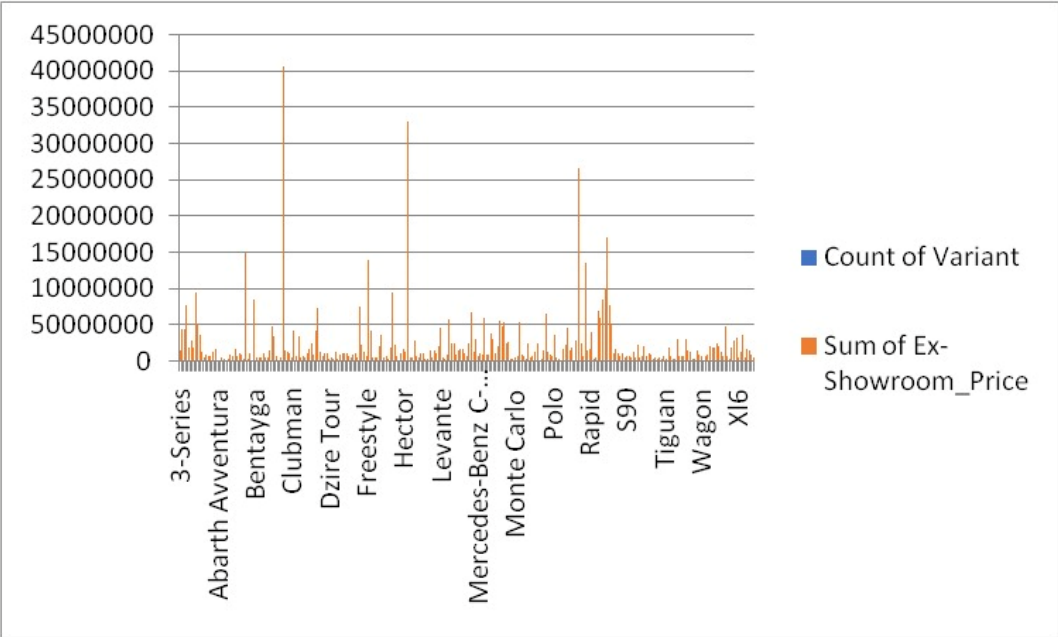
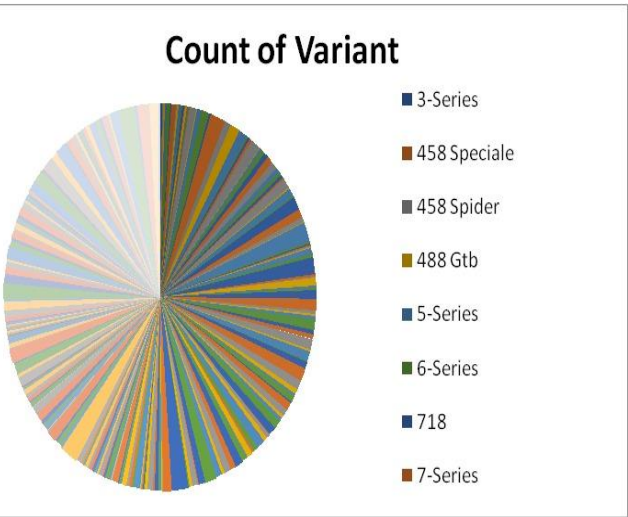
```
count    263.0
min       0.0
25%       0.0
```



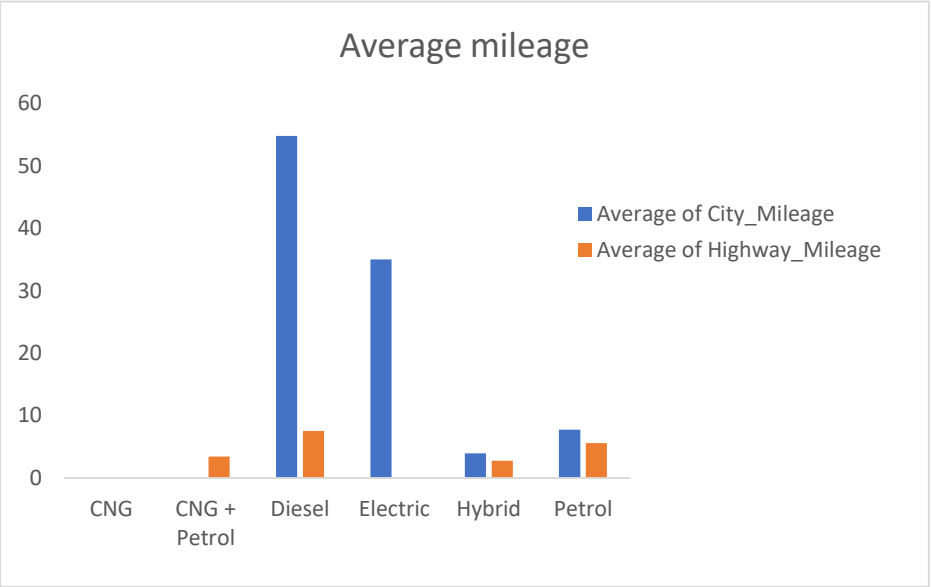
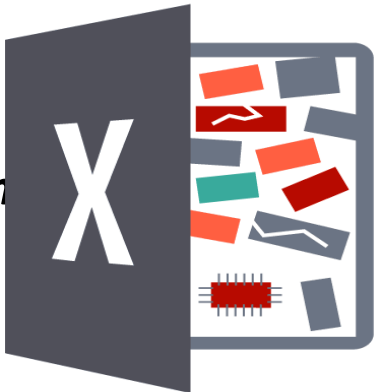
# Excel:

## 1. What is the distribution of ex-showroom prices across different car models and variants?

The distribution of ex-showroom prices across different car models and variants is highly twisted, with a significant number of models falling in the INR 500,000 to INR 15,000,000 range, but also featuring some high-end luxury models with significantly higher prices.



1. What is the average mileage (city and highway) for different fuel types and configurations?



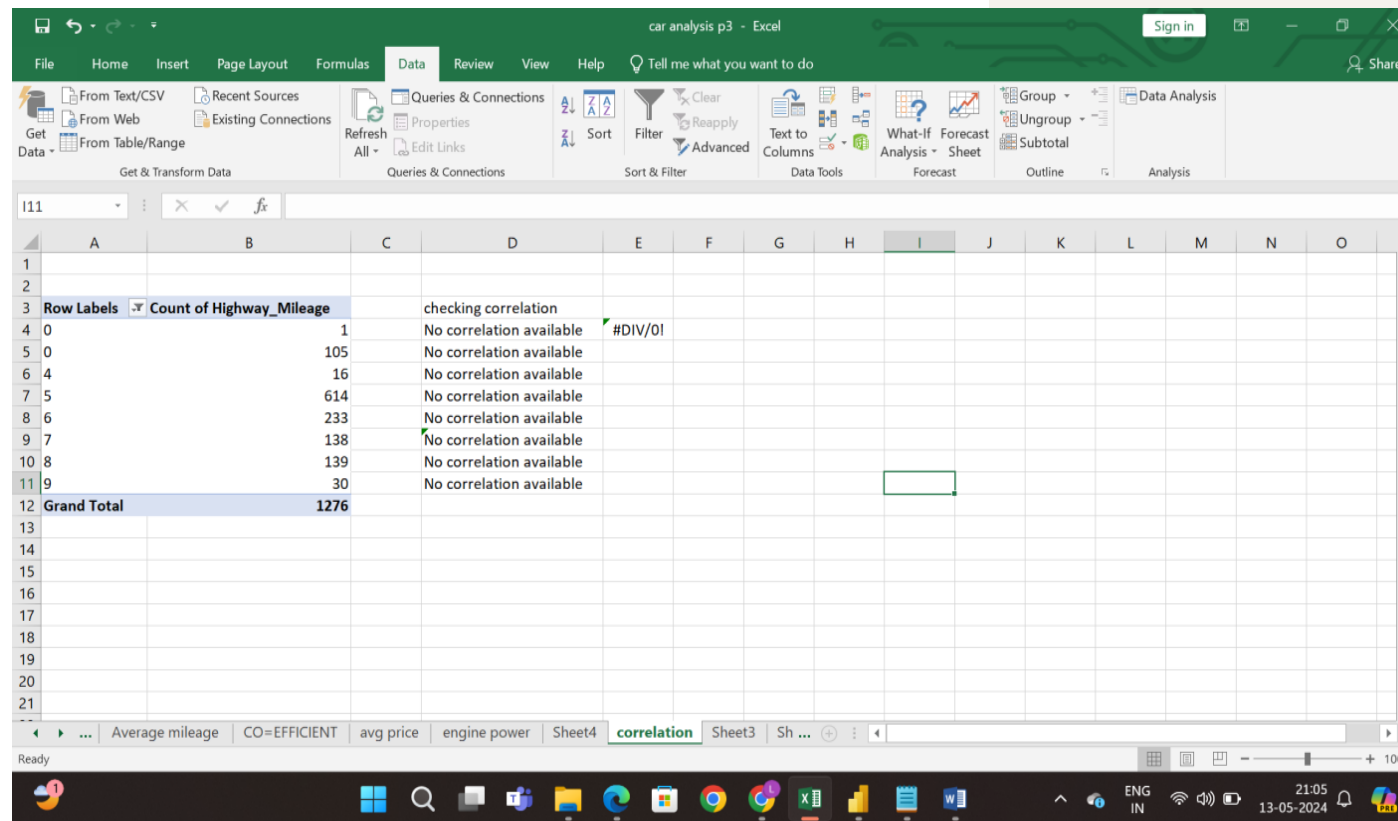
The average City mileage& Highway Mileage for different fuel types and engine configurations is as follows:

Fuel Type	Avg City Mileage	Avg Highway Mileage
CNG	0	0
CNG+ Petrol	0	04
Diesel	55	08
Electric	35	0
Hybrid	6	4
Petrol	9	8



## 2. Is there any correlation between the number of gears and fuel efficiency?

To establish a correlation, we would need continuous data with a larger sample size, where each data point represents a single vehicle with its corresponding number of gears and highway mileage.



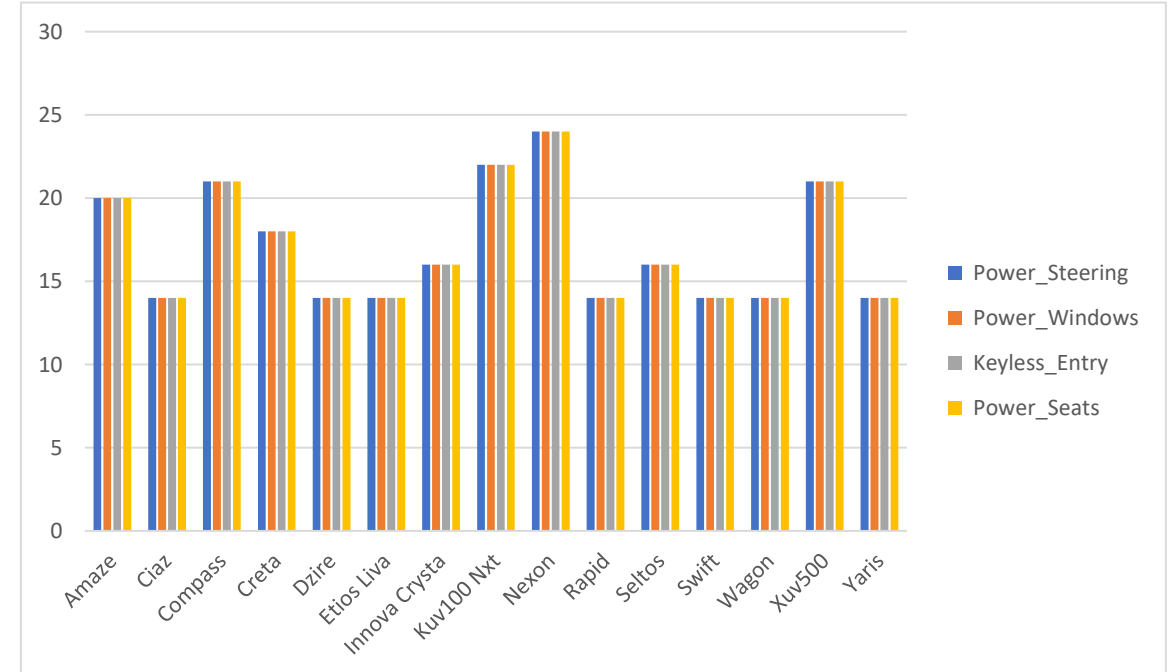
Row Labels	Count of Highway_Mileage	checking correlation
0	1	No correlation available
0	105	No correlation available
4	16	No correlation available
5	614	No correlation available
6	233	No correlation available
7	138	No correlation available
8	139	No correlation available
9	30	No correlation available
Grand Total	1276	



### 3. Which car models offer the most features such as power steering, power windows, power seats, keyless entry, etc.?

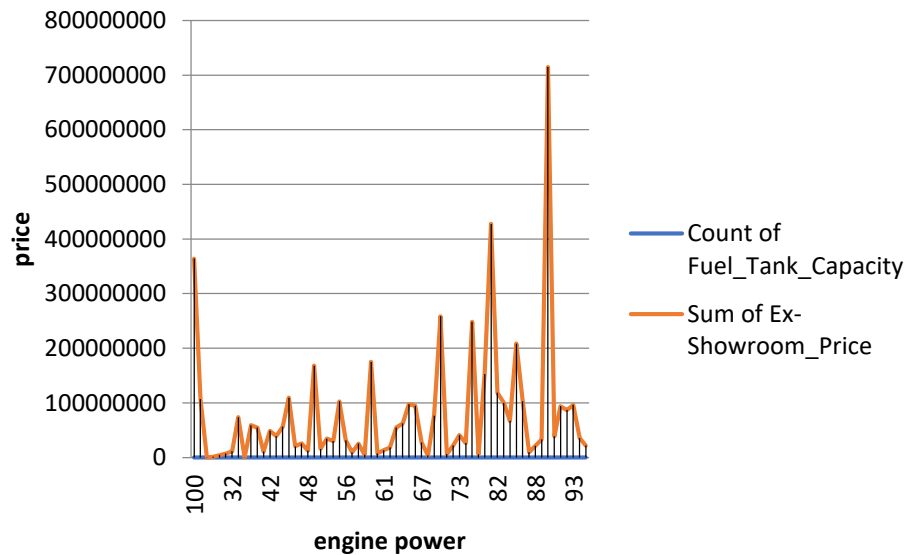
The following car models offer the most features such as power steering, power windows, power seats, and keyless entry:

- KUV100 Nxt: 22 points (22 for power steering, power windows, power seats, and keyless entry)
- Nexon: 24 points (24 for power steering, power windows, power seats, and keyless entry). These two models have the highest scores in the table, indicating that they offer all the mentioned features.



## ENGINE POWER:

- The chart displays engine power values ranging from 54 PS to 639 PS
- Analyze the distribution of engine power values there are several engine power values clustered around the 100-200 PS range, with fewer instances of higher power values
- Look for trends or patterns: that engine power values tend to increase as RPM values increase, suggesting a positive correlation between engine power and RPM
- Compare engine power values: engine power at 4000 RPM ranges from 70 PS to 190 bhp, with a significant increase in power between these two values



## AVERAGE PRICE:

The average price of cars from these manufacturers seems to vary quite a bit. Here's a breakdown of the average prices for each manufacturer, from highest to lowest:

Aston Martin: 1.4 billion €

Bentley: 1.2 billion €

Premier: 1 billion € (note: it's unclear from the context whether "Premier" is a manufacturer or a car model)

Land Rover: 800 million €

Volvo: 600 million €

Maruti Suzuki: 400 million €

Honda: 200 million €

Kia: 160 million €

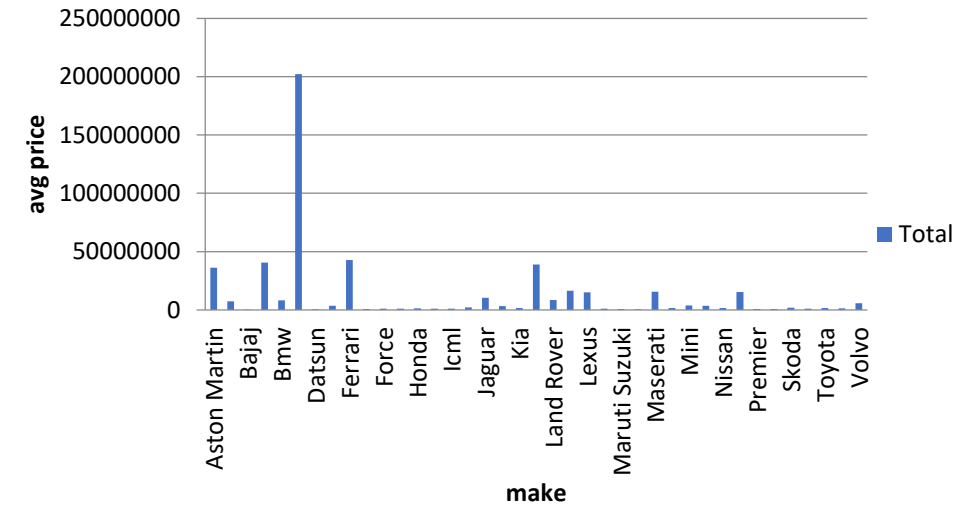
Isuzu: 120 million €

Fiat: 100 million €

Datsun: 60 million €



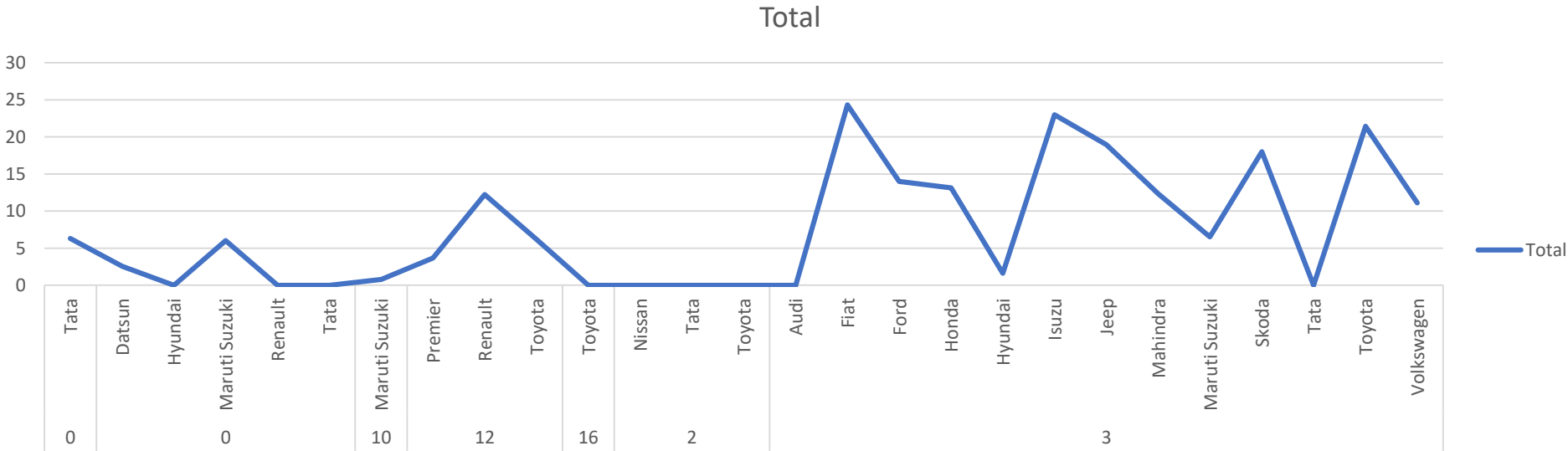
Avg price of car across different manufacture



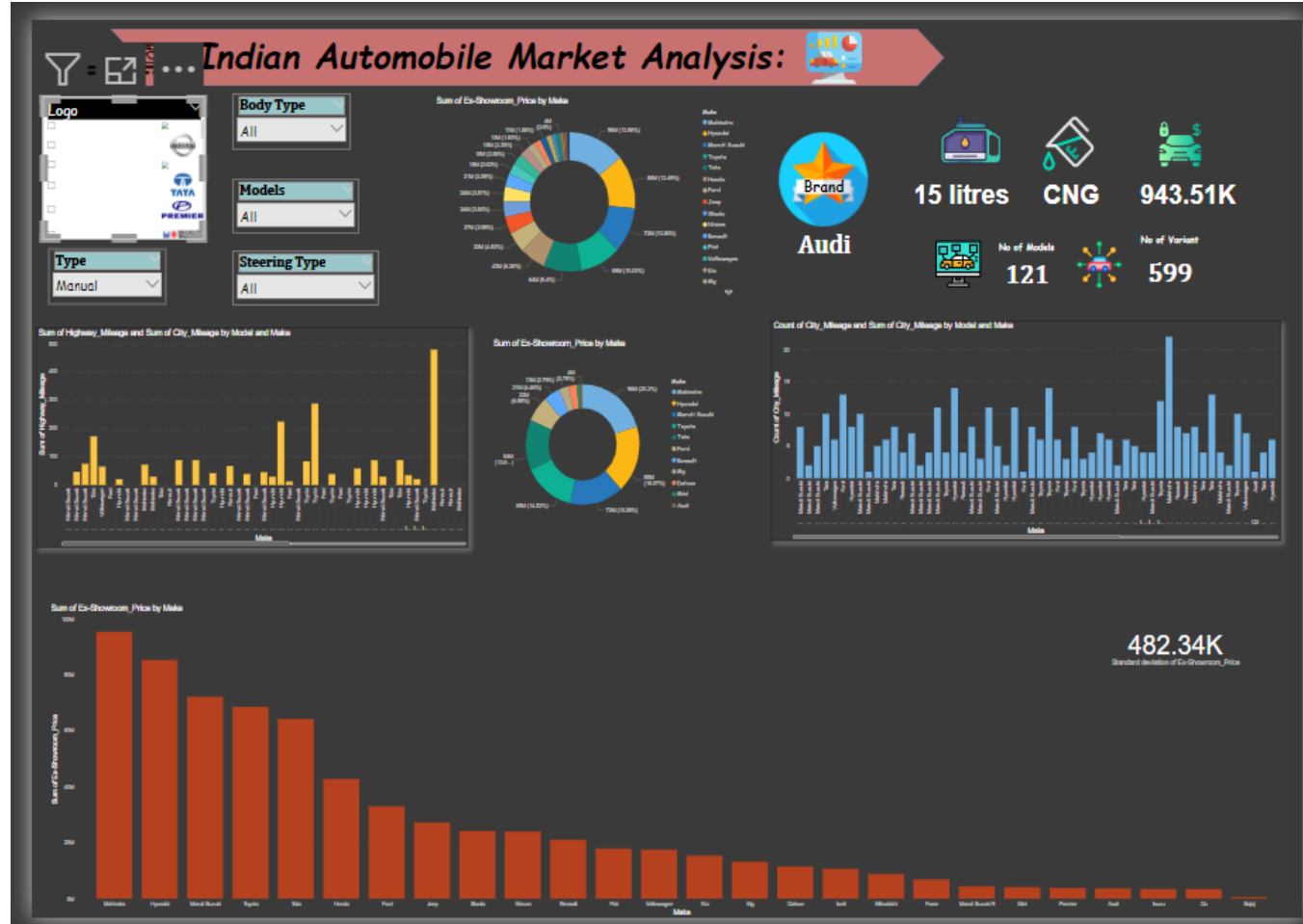
***Relation between no.of cylinders and Mileage :***

➤Car Mileage vs. Cylinder Count

➤**Axes:** Label x-axis "Number of Cylinders" (units) and y-axis "Highway Mileage (MPG)".



# Power BI Visualization:

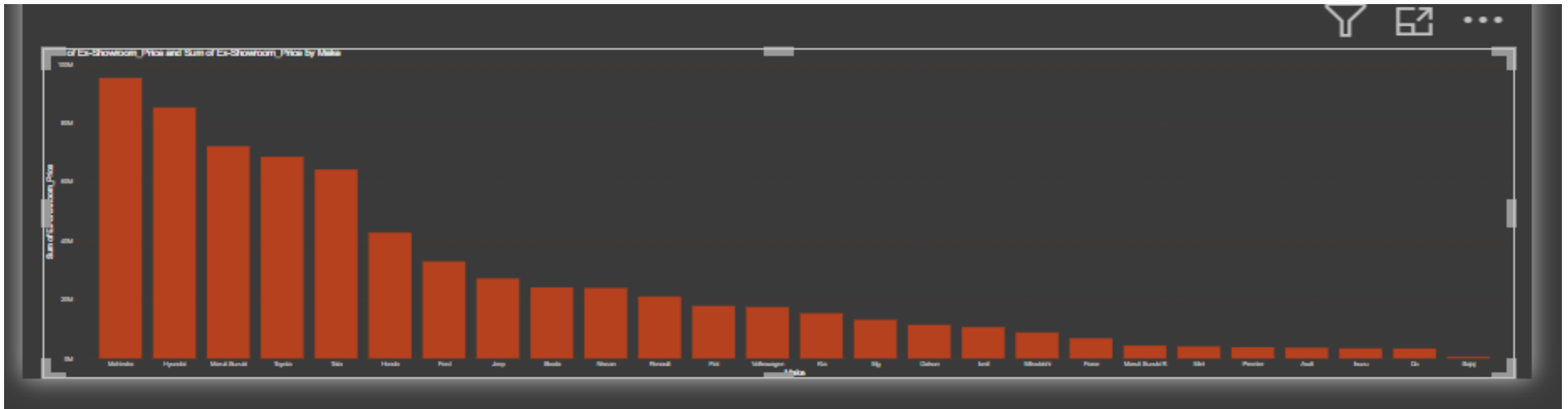


The dashboard provides a summary of the Indian car dataset, including information about ex-showroom prices, city mileage, highway mileage, fuel capacity and fuel type for each car model and make.

Here, we used Pi chart, Cards, Slicer, Stacked Area chart, Stacked Column chart for understandable Visualization according to Dataset.

**Popular car Ex Showroom Price and brands:**

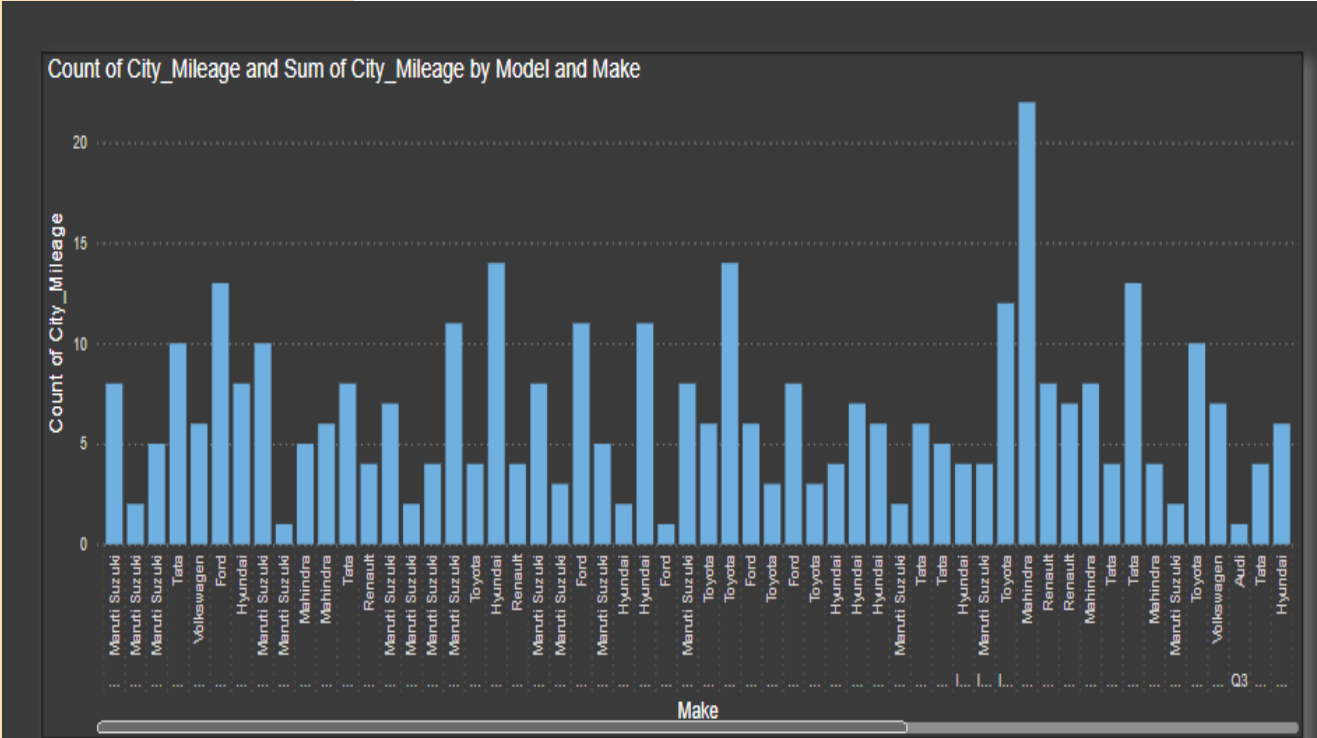
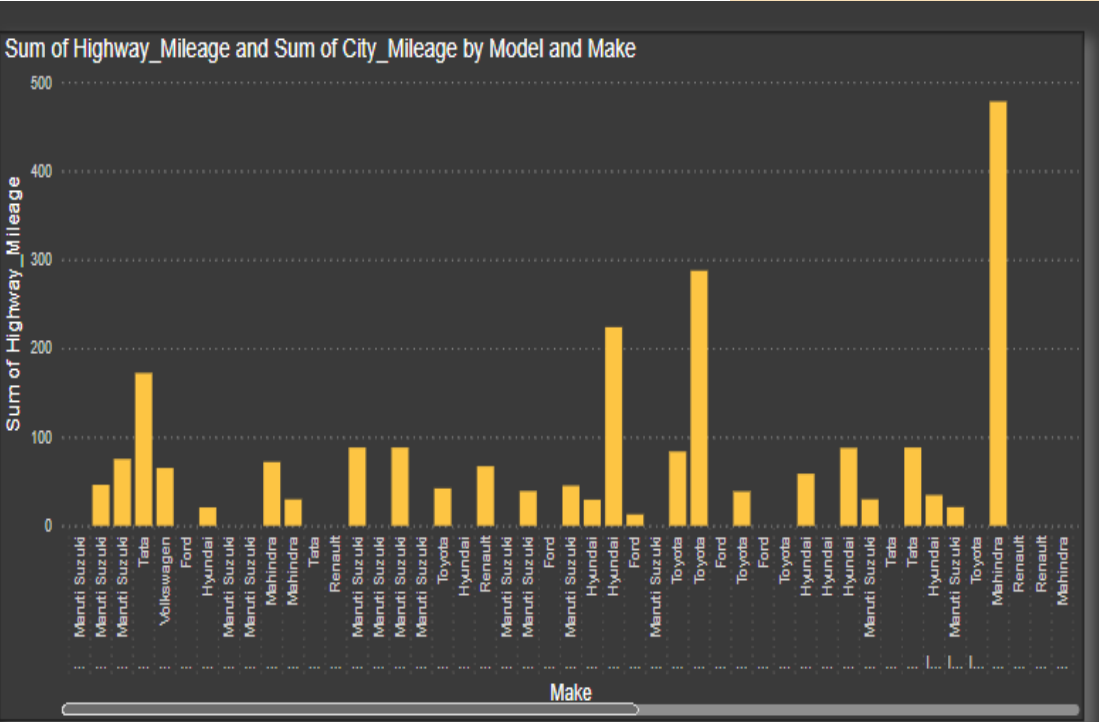
Displaying the sum of EX showroom pricing in relation to the "Make" variable, which specifies the brand name of the car firm. Here is the "Line and stacked column chart" for this Visualization.





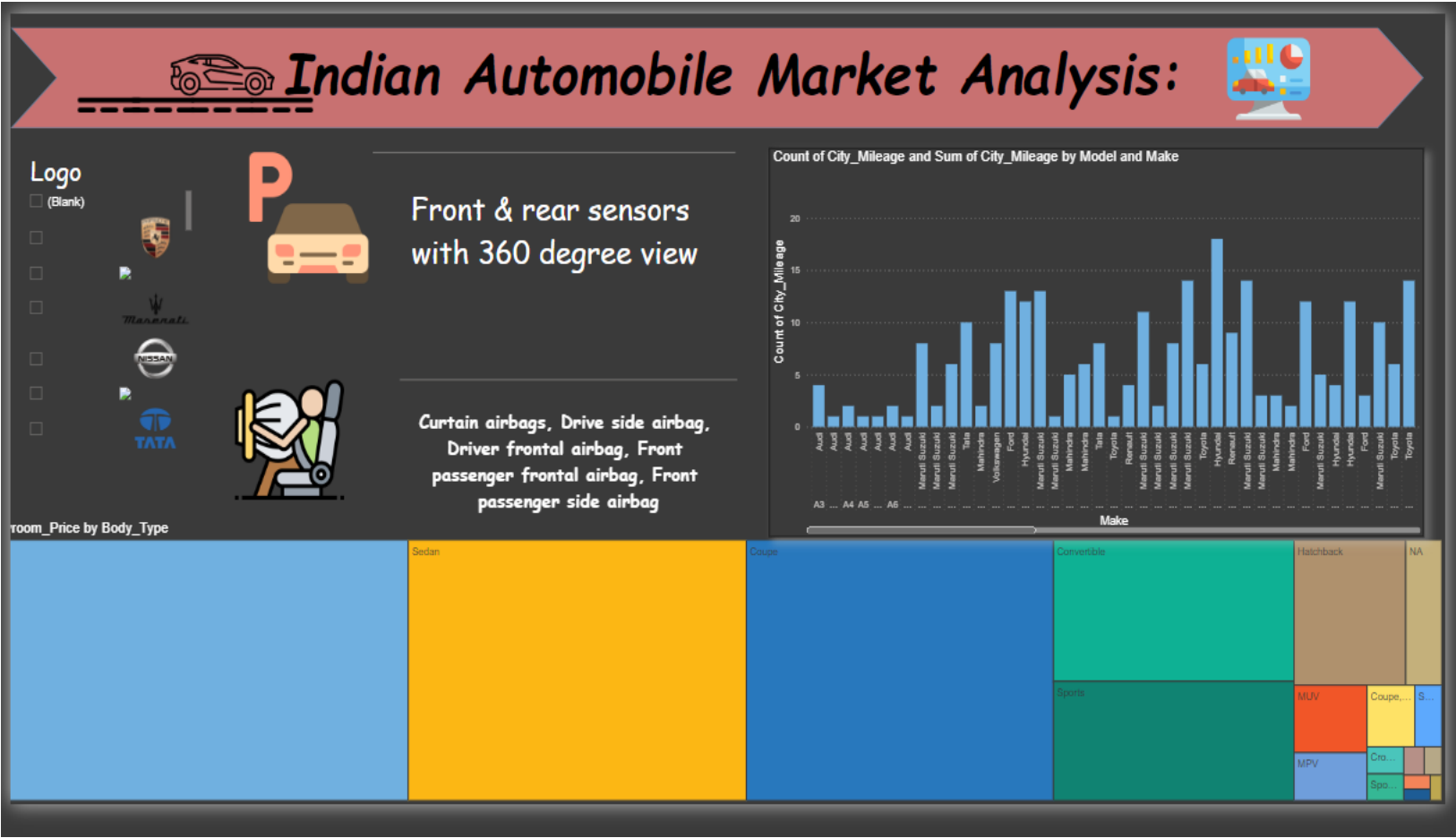
# Distribution of Prices and Mileage:

This Graph indicates distribution of price of the car and it's both mileages i.e. city mileage & highway mileage for a quick and better understanding of the car feature . Here is the "Line and clustered column chart" is used for Visualization.



# Availability of amenities and safety features across different models

Give the information of car parking features and information about the airbag safeties and also shows the information through graphs and can also select brands with the help of Logo “slicer” as shown in the above picture.



# Git Links:

[Lakshmi Keerthi B](#)

[Srujana B](#)

[Kuldeep GS](#)



# Reference:

- You Tube
- Geeks vs Geeks
- Udemy





Thank you!