Indian Automobile Market Analysis

Group 3

Srujana B
Kuldeep G S
Lakshmi Keerthi B

eduJournal

Agenda:

Specification Requirement (ok)

Python for Analysis

Excel

Power BI visualization (ok)

Git Links(ok)

Reference(ok)

Requirement Specification :

Hardware Requirement:

Windows11 Os' with i5 processer and above.

Software Requirement :

Python Compiler: Google Colab

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")
```

```
df
```

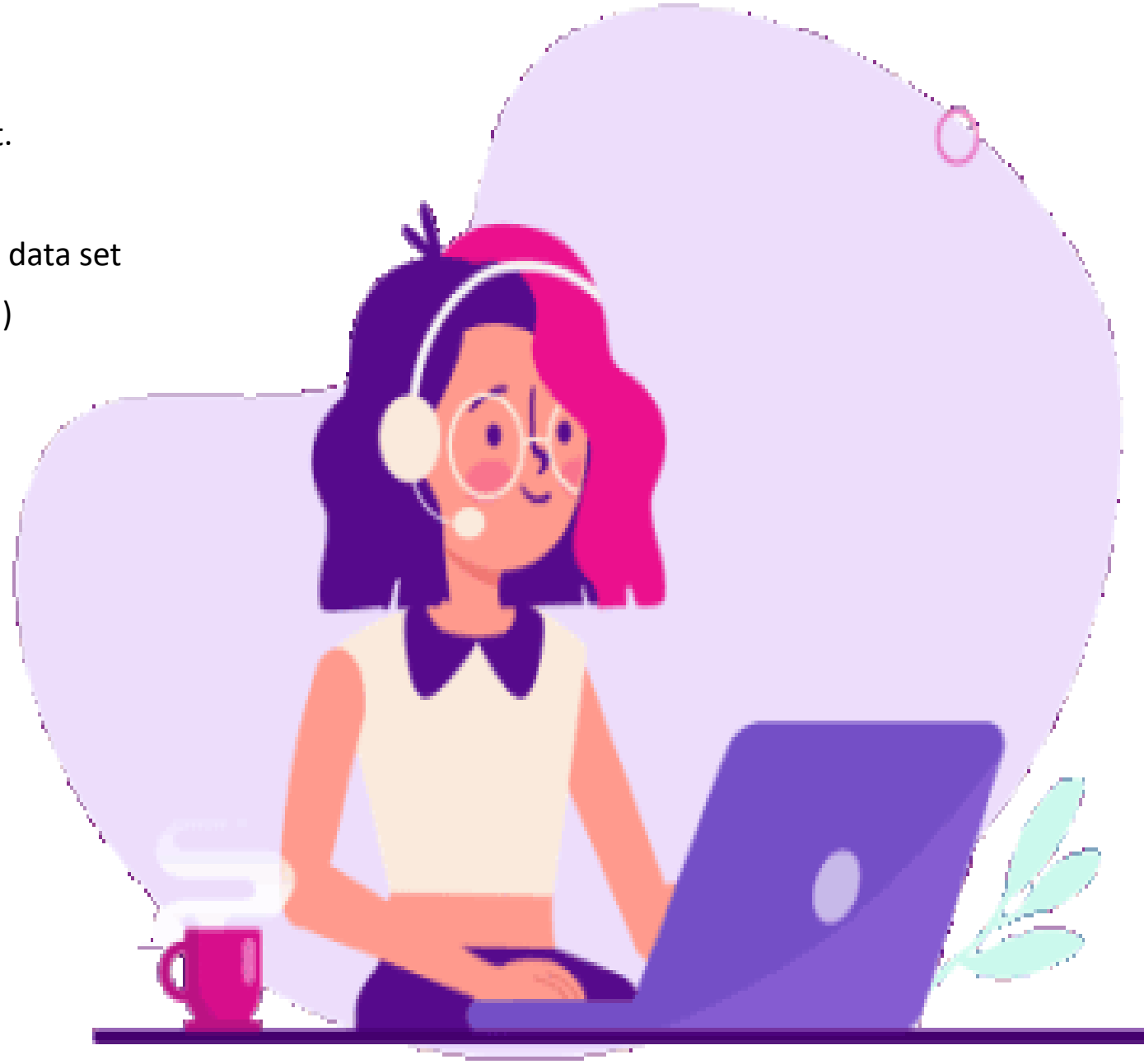
	Unnamed: 0	Make	Model	Variant	Ex-Showroom_Price	Displacement	Cylinders	Valves_Per_Cylinder	Drivetrain	Cylinder_Configura
0	0	Tata	Nano Genx	Xt	Rs. 2,92,667	624 cc	2.0	2.0	RWD (Rear Wheel Drive)	In
1	1	Tata	Nano Genx	Xe	Rs. 2,36,447	624 cc	2.0	2.0	RWD (Rear Wheel Drive)	In
2	2	Tata	Nano Genx	Emax Xm	Rs.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



- Replaced Nan values with 0

  df.fillna(0 , inplace = True)

Statistical analysis:

df.describe()

	NA	Ex-Showroom_Price	Cylinders	Valves_Per_Cylinder	Doors	City_Mileage	Seating_Capacity	Number_of_Airbags	USB_Ports
count	1277.000000	1.277000e+03	1277.000000	1277.000000	1277.000000	1277.000000	1277.000000	1277.000000	1277.00000
mean	637.000783	4.592938e+06	4.151135	3.657009	4.532498	29.317220	5.241190	3.383712	0.04072
std	368.781101	1.214327e+07	1.889136	1.346588	0.798657	728.293882	1.206602	2.655210	0.29077
min	0.000000	0.000000e+00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00000
25%	318.000000	7.435040e+05	4.000000	4.000000	4.000000	0.000000	5.000000	2.000000	0.00000
50%	637.000000	1.060000e+06	4.000000	4.000000	5.000000	8.500000	5.000000	2.000000	0.00000
75%	956.000000	2.977000e+06	4.000000	4.000000	5.000000	16.000000	5.000000	6.000000	0.00000
max	1275.000000	2.121554e+08	16.000000	16.000000	5.000000	26032.000000	16.000000	14.000000	3.00000

Fixed target Variable:

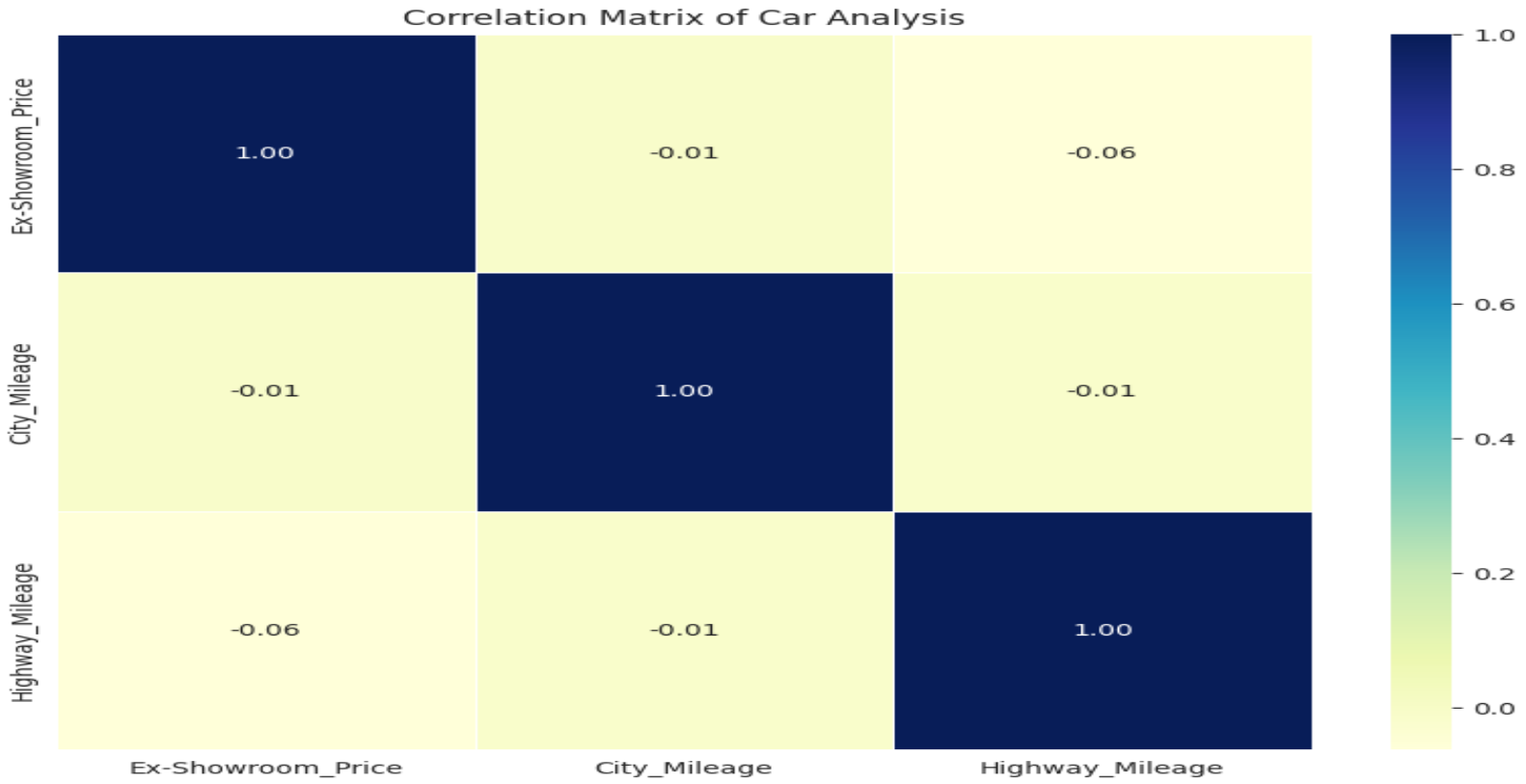
- On our observation, based on inbuilt features of cars(make), ARAI certified mileage, High end accessories and model will showcase the ex-showroom price
- As we consider ex-showroom price is our fixed target variable in the project

Statistical analysis for fixed target variable:

- Mean=4592938
- median =1060000
- Mode=999900

Correlation matrix:

- Correlation measures the strength and direction of a linear relationship between two 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 gold) represent weaker correlations, both positive and negative.
  - Since most of your correlations are close to zero, the cells are light-colored



Outliers Detection:

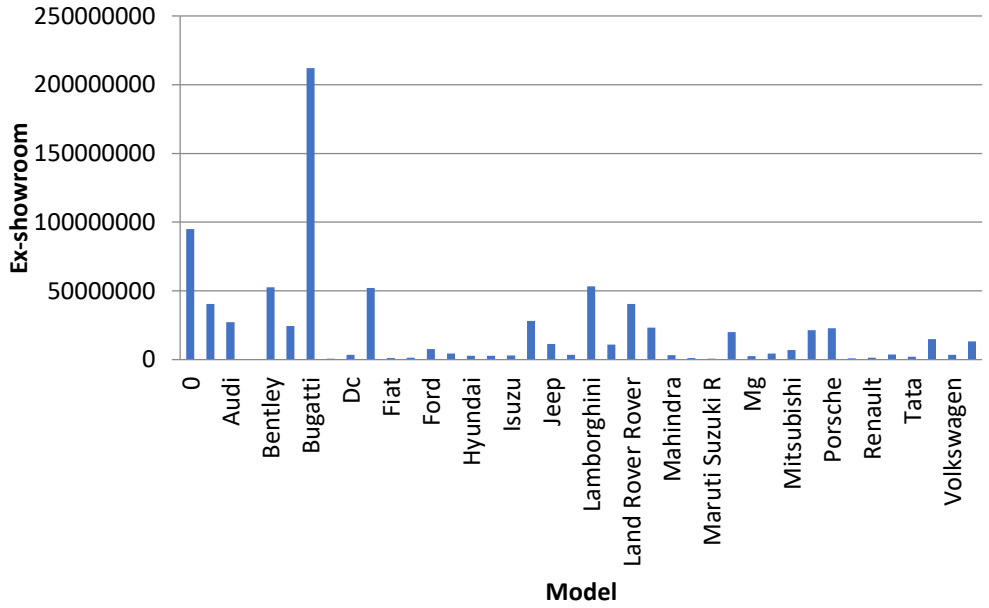
- Outlier is a data point that differs significantly from other observation
- Using frequency method we get outlier detection by considering target variable (Ex-showroom) inside cleaned data(EDA)

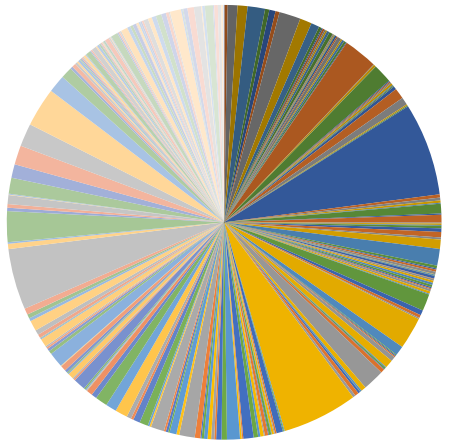
```
Outliers in the Model:
Ex-Showroom_Price
999900     0.010963
999990     0.003132
745000     0.002349
925000     0.002349
989000     0.002349
              ...
875358     0.000783
673261     0.000783
1102689    0.000783
919689     0.000783
0          0.000783
Name: proportion, Length: 1180, dtype: float64
```

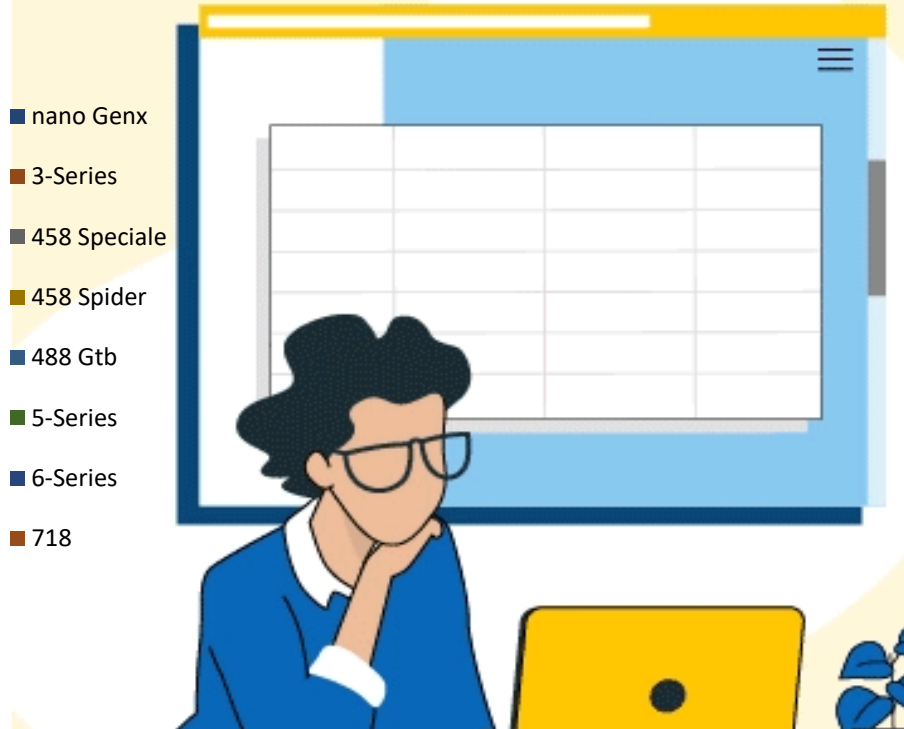
Excel:

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

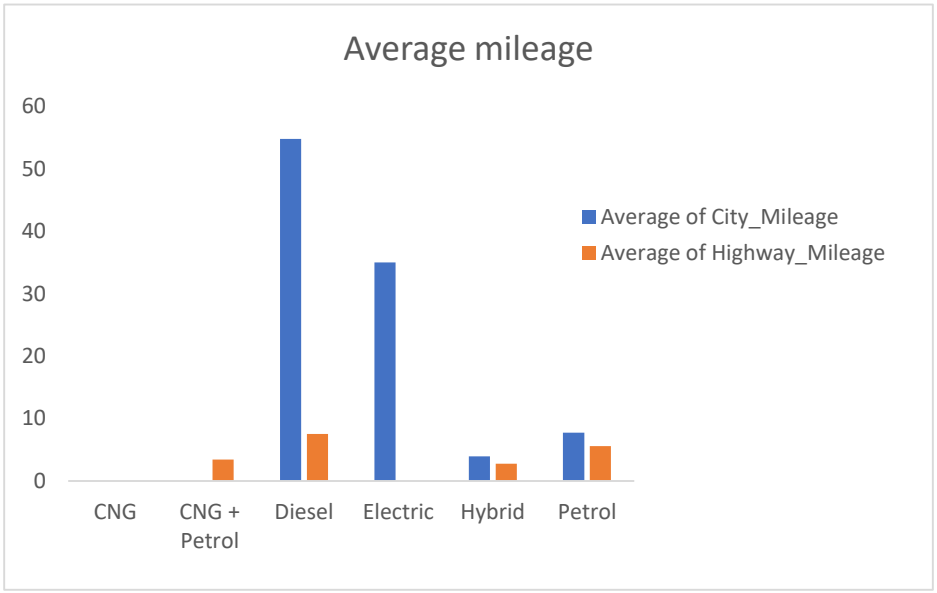
- Pie chart represents the relationship between Models and variants of the cars . From the below pie chart 458Gtb model has more variants in car
- Below bar graph indicates the distribution of ex-showroom price and model. It describes bugatti has average ex-showroom price







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



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

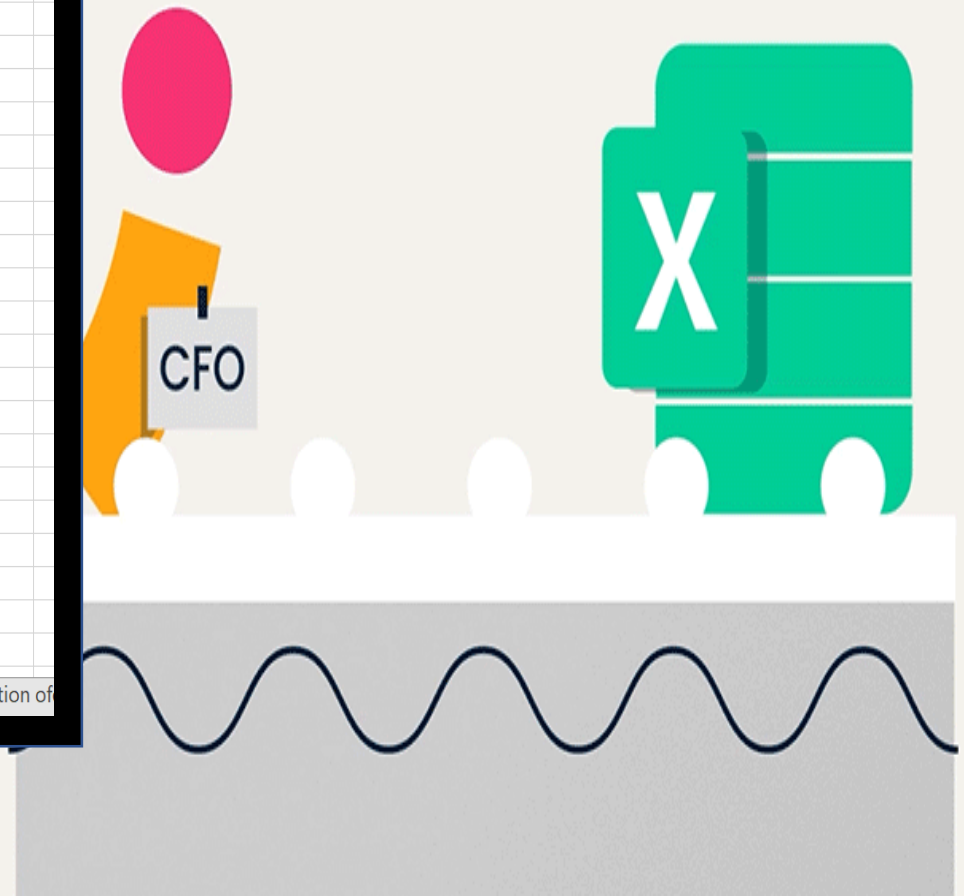
Fuel Vehicle Type	Avg City Mileage	Avg Highway Mileage
CNG	0	0
CNG+ Petrol	0	04
Diesel	55	08
Electric	35	0
Hybrid	6	4
Petrol	9	8

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.

Gears	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**	

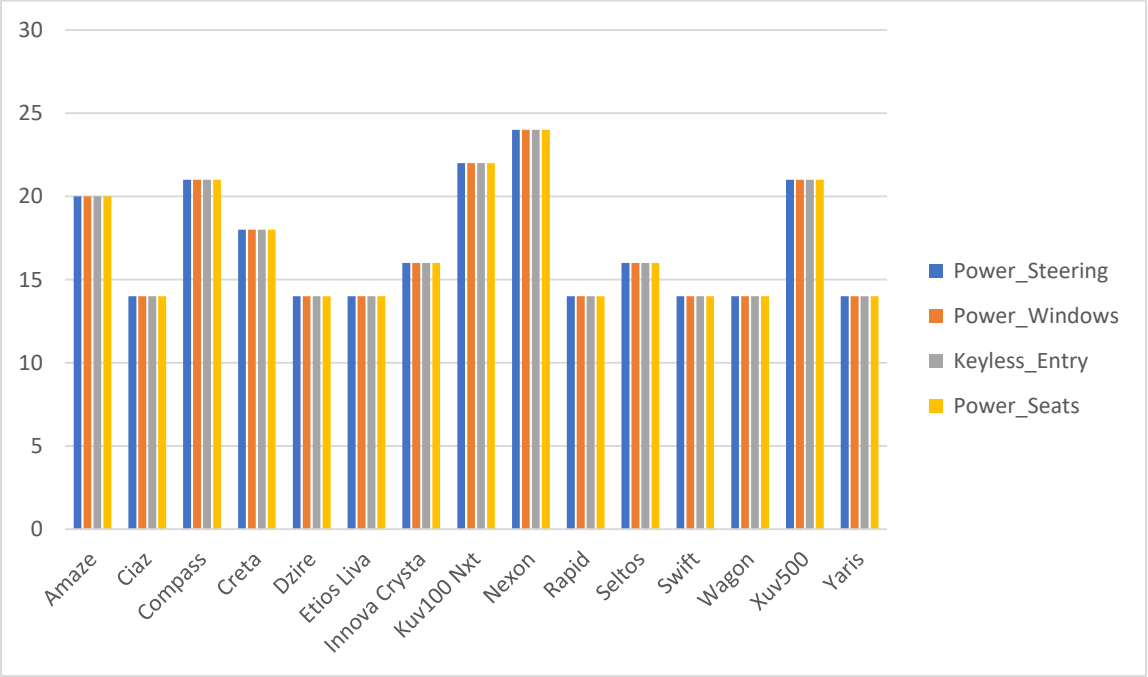
cleaned data | most model | Sheet13 | Average mileage | Sheet6 | avg price | engine power | **correlation** | Sheet1 | Sheet2 | Distribution of



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

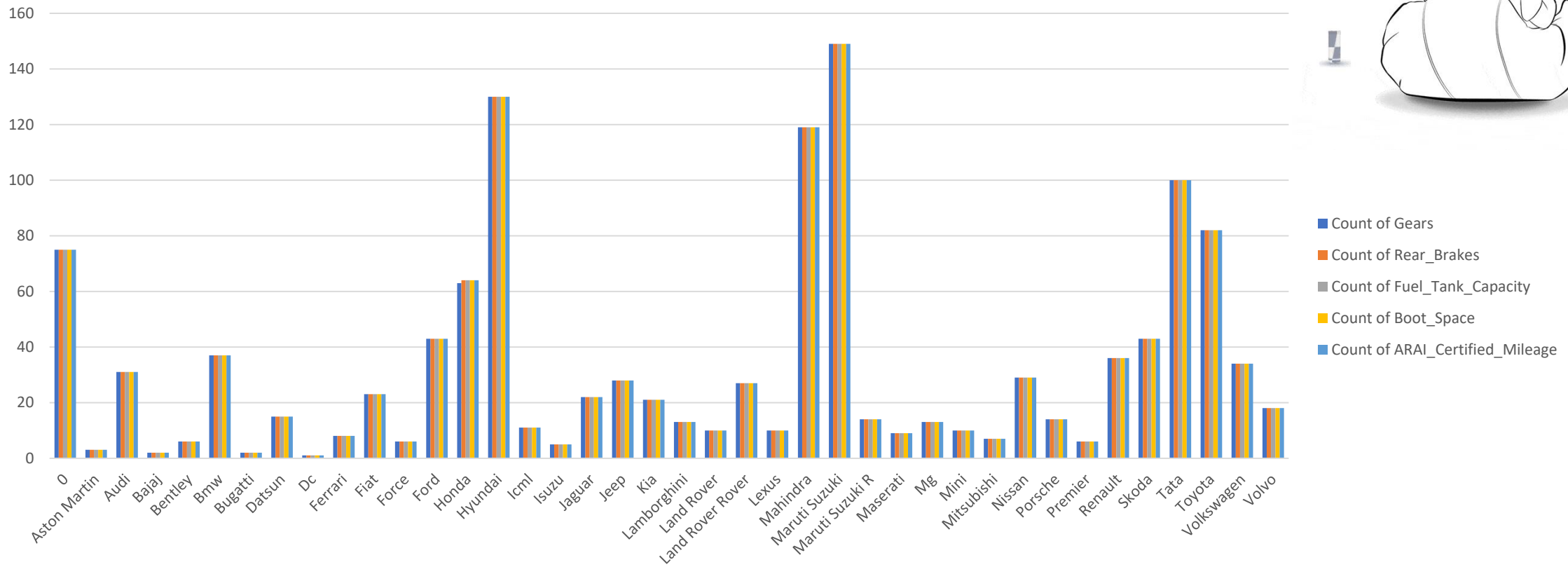
The following car models offer the most features such as power steering, power windows, power seats, and keyless entry:

- 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.



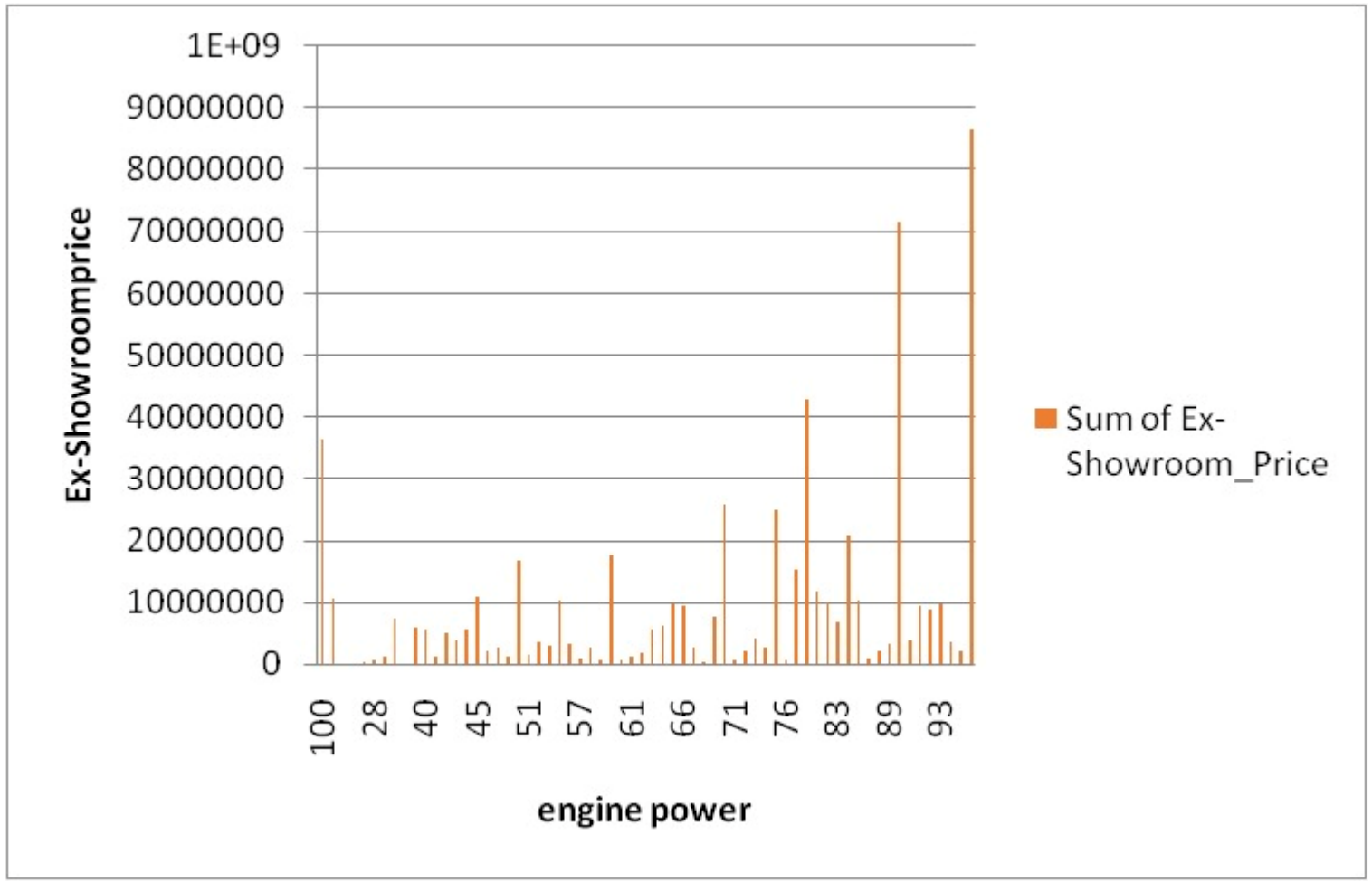
5. Are there any emerging trends in the Indian automotive market, such as the popularity of certain car types, features, or brands?

➢ Maruti Suzuki dominate in terms of the number of models or data points considered, while the average city mileage varies across brands, with some luxury brands performing less efficiently in city driving condition

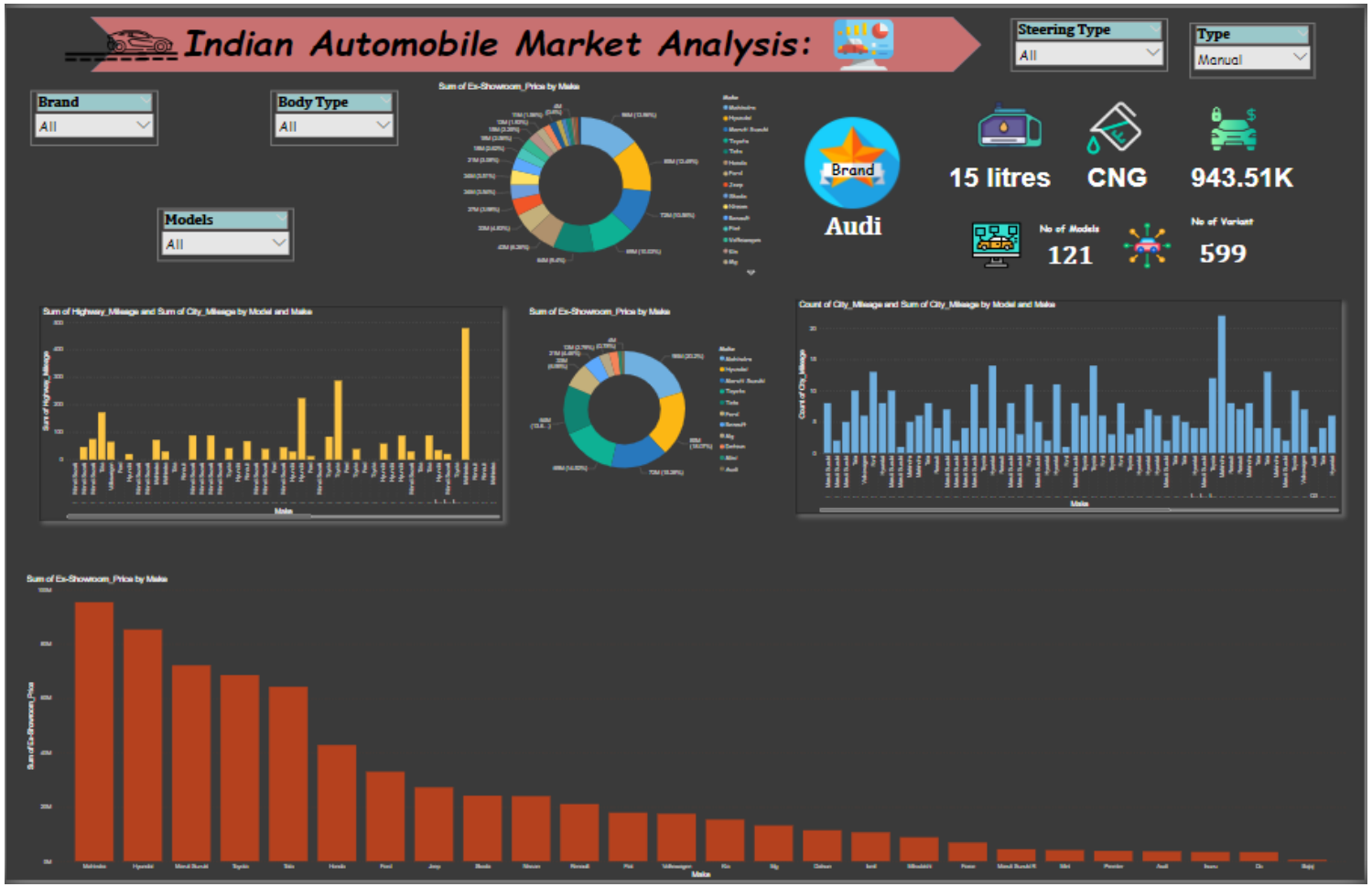


ENGINE POWER:

- Cars with higher engine power values (e.g., 83, 93, and 100) have significantly higher cumulative ex-showroom prices.
- This could be due to higher-priced models or a greater number of vehicles in these categories.



Power BI Visualization:

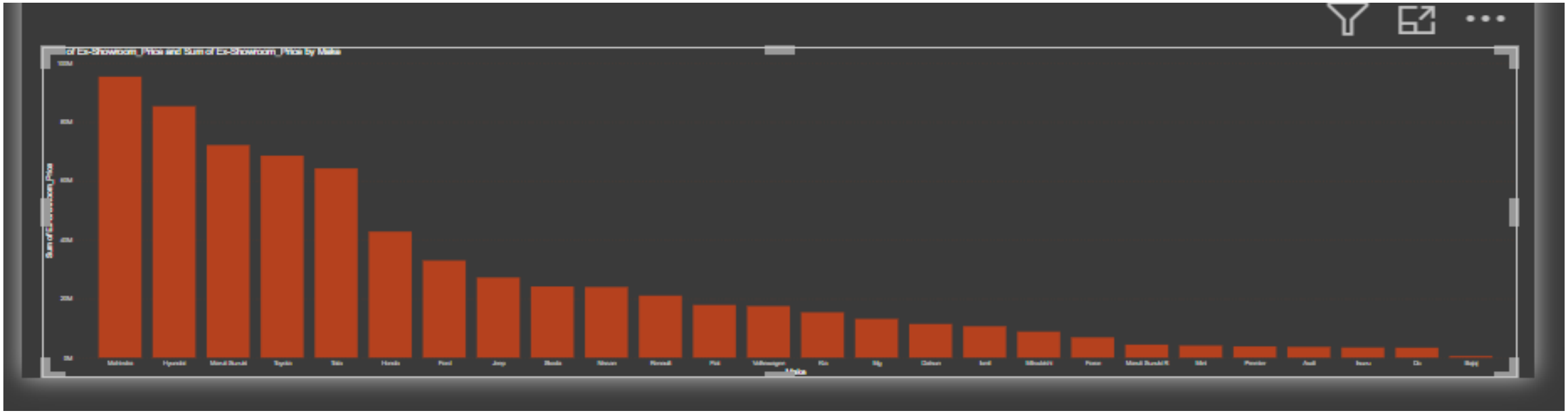


The dashboard provides a summary of the Indian car dataset, including information about ex-showroom prices, city mileage, highway mileage, fuel capacity and fuel type for each car model and make.

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

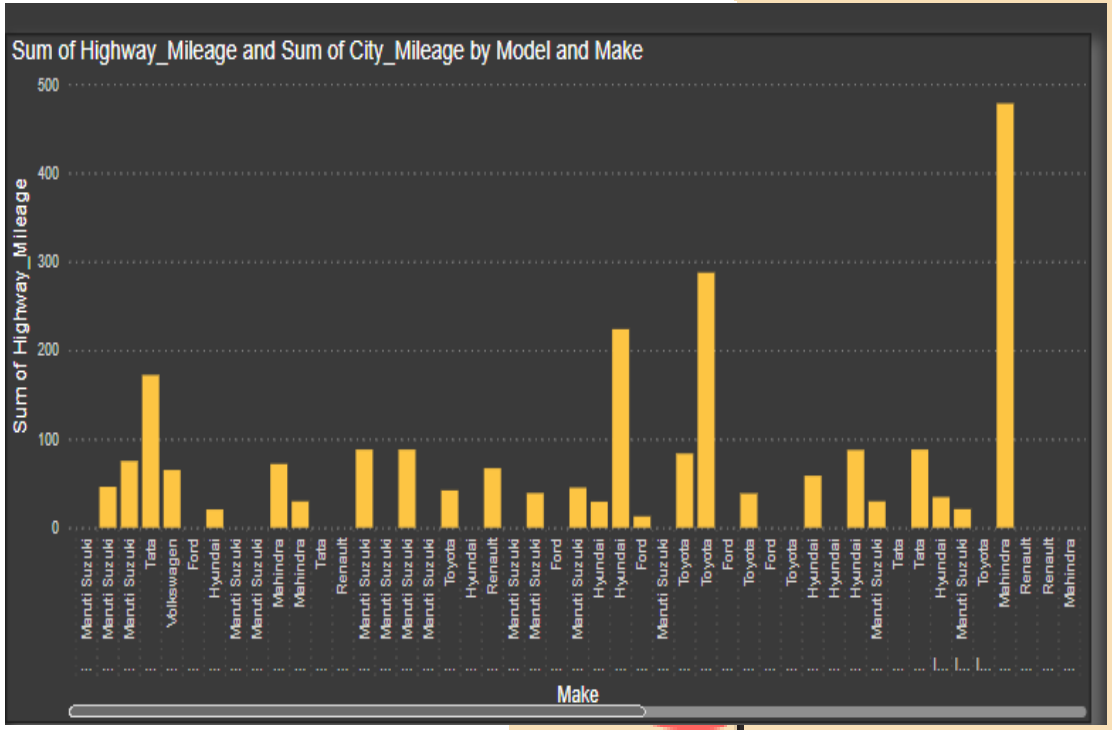
Popular car Ex Showroom Price and brands:

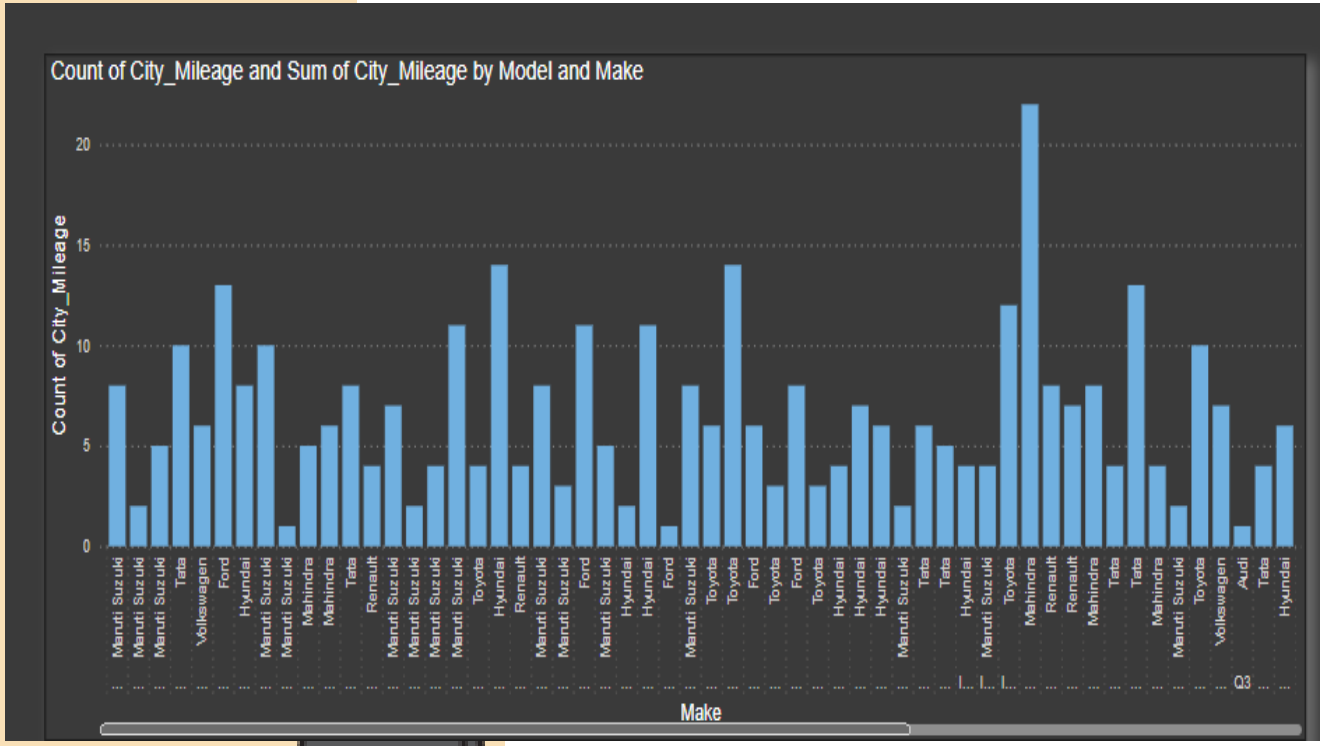
Displaying the sum of EX showroom pricing in relation to the "Make" variable, which specifies the brand name of the car firm. Here is the "Line and stacked column chart" for this Visualization.



Distribution of Prices and Mileage:

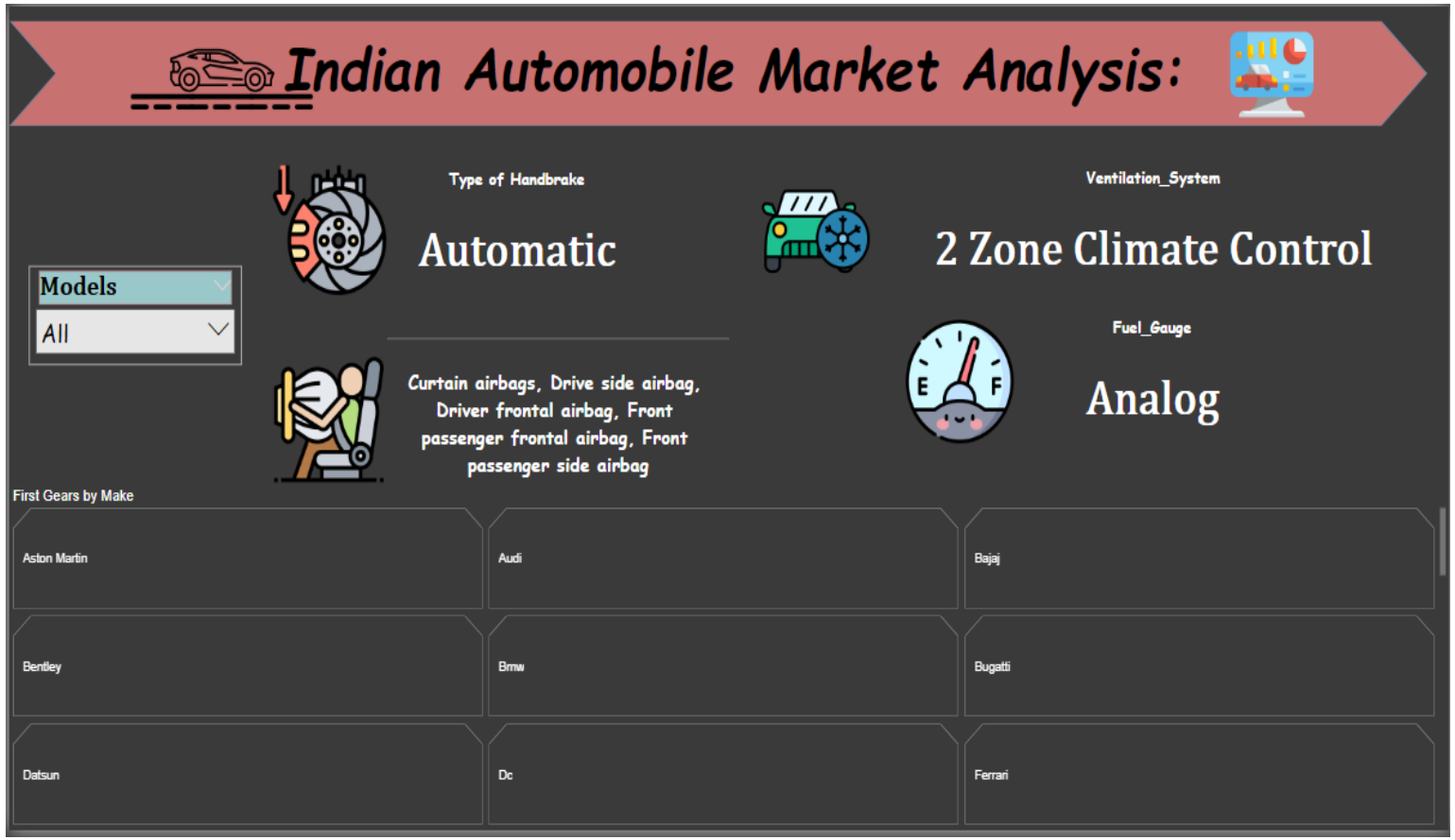
This Graph indicates distribution of price of the car and it’s both mileages i.e. city mileage & highway mileage for a quick and better understanding of the car feature . Here is the "Line and clustered column chart" is used for Visualization.





Availability of amenities and safety features across different models

- Give the information of car parking features and information about the airbag safeties , information about AC's , Fuel Gauge and also shows select brands with the help of Logo "slicer" as shown in the above picture.
- We used "New Slicer" which presents brands.



Git Links:

Indian Analysis Git Link

Reference:

- You Tube
- Geeks vs Geeks
- Udemy


Thank you!