DEV – I PROJECT

Submitted by:

LAKSHIT BHATTACHARYA

045028

PGDM-BDA-04

Section-H

Submitted to:

Prof. AMARNATH MITRA

TABLE OF CONTENTS

PROJECT OBJECTIVES

1. To do basic analysis of DataSet and understand it thoroughly.
2. To do categorization of data on the basis of fuel type, year, transmission, model and engine size.
3. To do sorting of data with respect to price, model and mileage.
4. To make a Line graph to check on the number of used cars per year.
5. To make a Bar graph for number of cars with respect to fuel type and check which fuel type has the highest number of cars.
6. To check how each model of the cars is divided with respect to fuel type.
7. To check how different models of the cars are differentiated with respect to transmission type.
8. To check how different car models are differentiated with respect to their average pricing and mpg.
9. To make a heat map and understand the correlation between different variables of the data set.

GENERAL DESCRIPTION OF DATA

I decided to do my analysis on a dataset of used BMW cars over the years. I got my Dataset from Kaggle and then decided to work upon it and get some findings out of it. As I am a Gearhead (car enthusiast) by nature this dataset was perfect for me and doing analysis on it was fun, exciting and filled with a lot of learnings.

So this dataset gives us a basic information of used BMW cars with different models and their specifications. It gave us an idea about the manufacturing year of the car , its model name, engine size , transmission type, fuel type and other useful details. As the dataset was huge, doing analysis on it helped me learn a lot of new things and get a better understanding of different libraries in Python language.

	A	B	C	D	E	F	G	H	I	J
1	model	year	price	transmissi	mileage	fuelType	tax	mpg	engineSize	
2	5 Series	2014	11200	Automatic	67068	Diesel	125	57.6	2	
3	6 Series	2018	27000	Automatic	14827	Petrol	145	42.8	2	
4	5 Series	2016	16000	Automatic	62794	Diesel	160	51.4	3	
5	1 Series	2017	12750	Automatic	26676	Diesel	145	72.4	1.5	
6	7 Series	2014	14500	Automatic	39554	Diesel	160	50.4	3	
7	5 Series	2016	14900	Automatic	35309	Diesel	125	60.1	2	
8	5 Series	2017	16000	Automatic	38538	Diesel	125	60.1	2	
9	2 Series	2018	16250	Manual	10401	Petrol	145	52.3	1.5	
10	4 Series	2017	14250	Manual	42668	Diesel	30	62.8	2	
11	5 Series	2016	14250	Automatic	36099	Diesel	20	68.9	2	
12	X3	2017	15500	Manual	74907	Diesel	145	52.3	2	
13	1 Series	2017	11800	Manual	29840	Diesel	20	68.9	2	
14	X3	2016	15500	Automatic	77823	Diesel	125	54.3	2	
15	2 Series	2015	10500	Manual	31469	Diesel	20	68.9	2	
16	X3	2017	22000	Automatic	19057	Diesel	145	54.3	2	
17	3 Series	2017	16500	Manual	16570	Diesel	125	58.9	2	
18	3 Series	2017	14250	Automatic	55594	Other	135	148.7	2	
19	3 Series	2017	16000	Automatic	45456	Diesel	30	64.2	2	
20	1 Series	2017	15500	Automatic	22812	Diesel	20	68.9	1.5	
21	4 Series	2014	14000	Automatic	47348	Diesel	125	60.1	2	
22	1 Series	2015	9700	Automatic	75124	Diesel	20	70.6	2	

- Dataset in Excel format.

jupyter BMW.csv✔ 19 hours ago

File Edit View Language

```
model,year,price,transmission,mileage,fuelType,tax,mpg,engineSize
 5 Series,2014,11200,Automatic,67068,Diesel,125,57.6,2
 6 Series,2018,27000,Automatic,14827,Petrol,145,42.8,2
 5 Series,2016,16000,Automatic,62794,Diesel,160,51.4,3
 1 Series,2017,12750,Automatic,26676,Diesel,145,72.4,1.5
 7 Series,2014,14500,Automatic,39554,Diesel,160,50.4,3
 5 Series,2016,14900,Automatic,35309,Diesel,125,60.1,2
 5 Series,2017,16000,Automatic,38538,Diesel,125,60.1,2
 2 Series,2018,16250,Manual,10401,Petrol,145,52.3,1.5
 4 Series,2017,14250,Manual,42668,Diesel,30,62.8,2
 5 Series,2016,14250,Automatic,36099,Diesel,20,68.9,2
 X3,2017,15500,Manual,74907,Diesel,145,52.3,2
 1 Series,2017,11800,Manual,29840,Diesel,20,68.9,2
 X3,2016,15500,Automatic,77823,Diesel,125,54.3,2
 2 Series,2015,10500,Manual,31469,Diesel,20,68.9,2
 X3,2017,22000,Automatic,19057,Diesel,145,54.3,2
 3 Series,2017,16500,Manual,16570,Diesel,125,58.9,2
 3 Series,2017,14250,Automatic,55594,Other,135,148.7,2
 3 Series,2017,16000,Automatic,45456,Diesel,30,64.2,2
 1 Series,2017,15500,Automatic,22812,Diesel,20,68.9,1.5
 4 Series,2014,14000,Automatic,47348,Diesel,125,60.1,2
 1 Series,2015,9700,Automatic,75124,Diesel,20,70.6,2
 3 Series,2015,12600,Automatic,78957,Diesel,30,62.8,2
 3 Series,2016,15100,Automatic,29213,Diesel,30,64.2,2
 1 Series,2016,9400,Manual,44498,Diesel,0,83.1,1.5
 1 Series,2016,14300,Automatic,22461,Diesel,20,67.3,2
 1 Series,2016,11200,Manual,23005,Petrol,125,53.3,1.5
 3 Series,2019,17800,Automatic,22310,Diesel,145,64.2,2
```

- Dataset in CSV format uploaded in Jupyter Notebook.

ANALYSIS

So analysis starts with knowing how to read a csv file which is first done by importing pandas and numpy. They help in working with datasets and gathering information out of them. Then I used "pfd.read_csv('file name')" to read the csv file of used BMW cars. After that I used head function to get the first few values of the dataset as loading all the values would not be feasible and it is actually not required.

Following this it is important that we know how big is our dataset. This would help us get a basic idea on what we are working on and this is achieved by using the "shape()" function.

After getting this brief idea about the dataset I tried to get a count of the number of cars differentiated with respect to fuel type. This was done by the use of "value_counts()" function.

```
bm['fuelType'].value_counts()
```

```
Diesel      7027
Petrol      3417
Hybrid       298
Other         36
Electric       3
```

```
bm.loc[3,['price','mileage','fuelType','mpg']]
```

```
price        12750
mileage      26676
fuelType    Diesel
mpg           72.4
```

I repeated this process to differentiate cars in terms of Transmission, Engine size, model type, year of manufacturing.

Then I played with data by accessing the descriptions of different cars by using the " iloc" function and putting in the numerical value inside [ ]. This helped me get the details of a specific car at a particular numerical value. One more addition to this feature is using 'loc' function to get only specified descriptions of a particular car. Lets say that I only want to see the price and mileage of a particular car and not any other feature then I would use this function.

After this I tried Indexing and sorting of data with respect to a particular description. For Ex. : If I want to sort data in terms of price then I would use "sort_values(by="price")" to sort this data in terms of price and sorting gives us the data in ascending or descending order of the value. We can also sort data with respect to two variables like model and mileage . This would give us a better understanding of the data we want to see.

```
sorted_bm = bm.sort_values(by="price") # arranging cars on the basis of
sorted_bm
```

	model	year	price	transmission	mileage	fuelType	tax	mpg	engineSize
9744	3 Series	1999	1200	Automatic	82000	Petrol	245	31.0	2.8
9696	3 Series	2004	1445	Manual	162000	Diesel	205	49.6	2.0
9554	1 Series	2007	1500	Manual	167000	Diesel	125	57.6	2.0
10006	3 Series	2000	1550	Automatic	93000	Petrol	270	29.7	2.5
7284	5 Series	2002	1595	Automatic	115000	Petrol	325	28.5	2.2
...	...	...	...	...	...	...	...	...	...
1813	8 Series	2019	88980	Semi-Auto	88	Petrol	145	24.4	4.4
4776	M5	2019	89900	Semi-Auto	2269	Petrol	145	24.1	4.4
2909	M4	2017	89990	Semi-Auto	1336	Petrol	145	33.2	3.0
5362	M4	2016	99950	Automatic	771	Petrol	300	33.2	3.0
3638	2 Series	2015	123456	Semi-Auto	33419	Diesel	20	68.9	2.0

10781 rows × 9 columns

```
selected_col = bm[["model","mileage"]]
selected_col
```

	model	mileage
0	5 Series	67068
1	6 Series	14827
2	5 Series	62794
3	1 Series	26676
4	7 Series	39554
...	...	...
10776	X3	40818
10777	5 Series	42947

Following this I tried to differentiate different models with respect to transmission type i.e. automatic or Manual. This helped me get a better picture of how many cars were automatic and how many were Manual as it divided the cars in a tabular format and put the models in whichever category they belonged to.

I repeated this step with fuel type also as it had more categories i.e. Diesel, Petrol, Hybrid, Electric and others. This helped me differentiate car models in a better way and get a glimpse of how cars with different models or for that matter same models were differentiated w.r.t fuel type.

```
automatic_models = bm[bm["transmission"] == "Automatic"]#
manual_models = bm[bm["transmission"] == "Manual"]

print("Car models with Automatic Transmission:") # To Dis
print(automatic_models[["model", "transmission"]])

print("\nCar models with Manual Transmission:") # To Disp
print(manual_models[["model", "transmission"]])
```

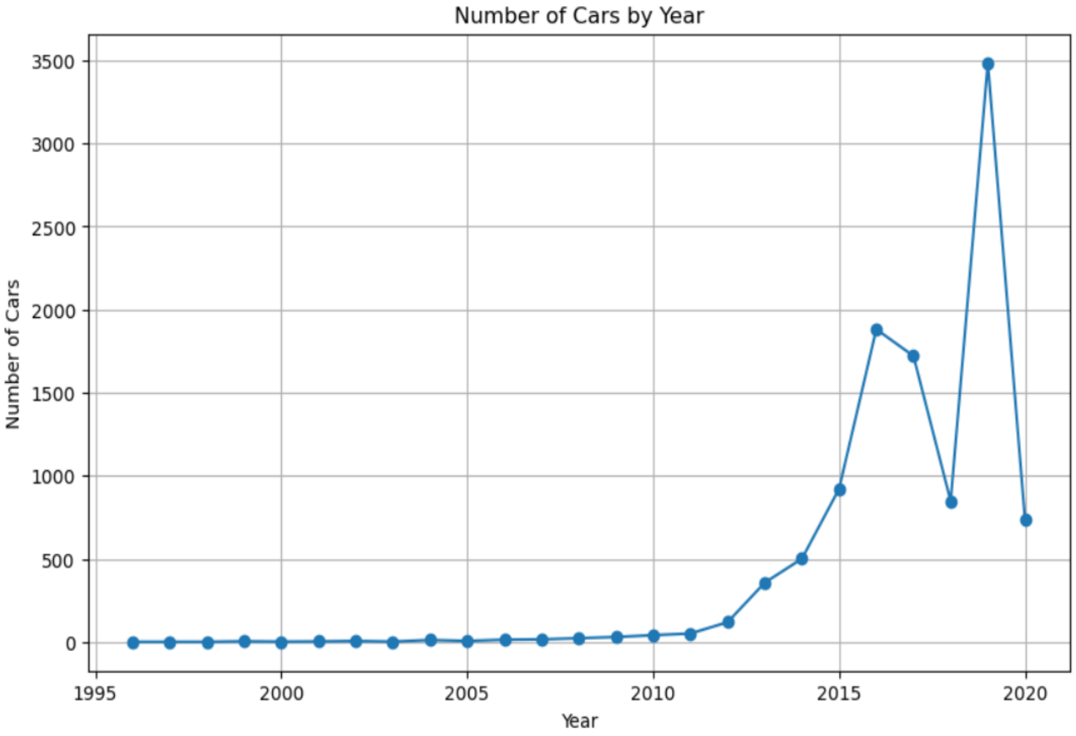
```
Car models with Automatic Transmission:
          model transmission
0      5 Series    Automatic
1      6 Series    Automatic
2      5 Series    Automatic
3      1 Series    Automatic
4      7 Series    Automatic
...         ...          ...
10775        X4    Automatic
10776        X3    Automatic
10777  5 Series    Automatic
10779  1 Series    Automatic
10780        X1    Automatic

[3588 rows x 2 columns]
```

After all these segmentations and sorting of data I tried visualization of data by making line graphs , Bar graphs and Heat map to get the correct understanding of the data. A important function in visualizing data is “ matplotlib.pyplot”. By using this function we can construct any type of graph and different models. For Heatmap we need to import seaborn and get the required results.

I would talk about these graphs and their finding in detail in the next segment of my report which is inferences and findings.

<u>FINDINGS AND INFERENCES</u>

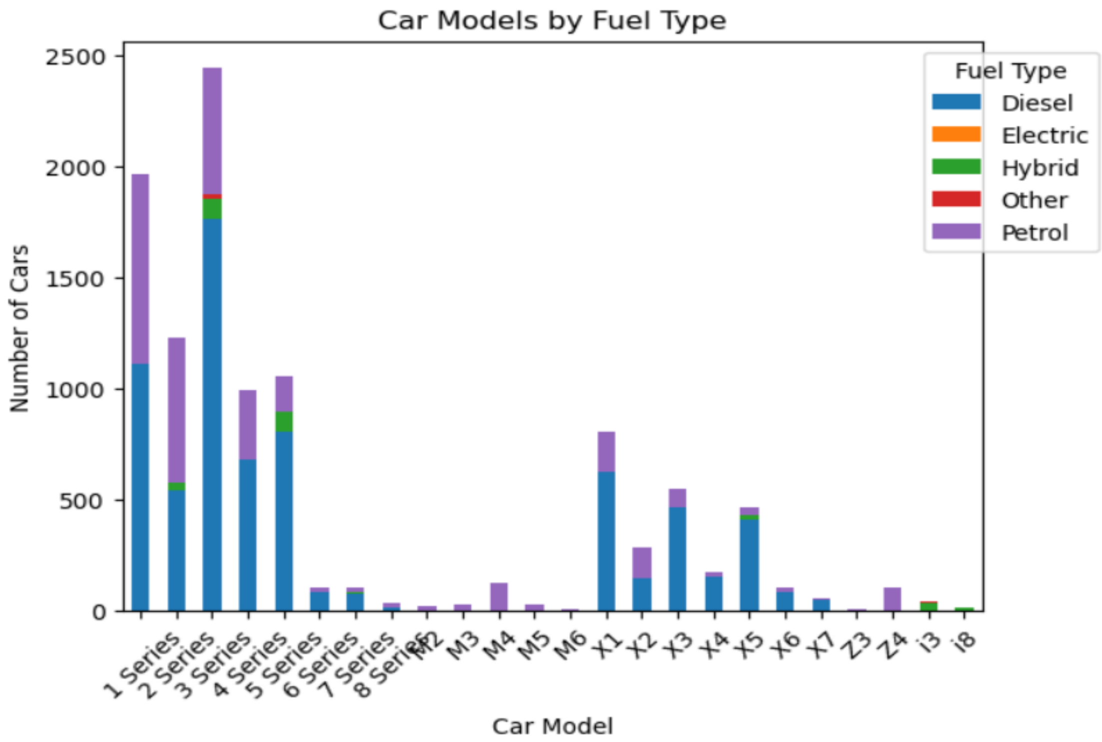
- So the first graph that I created was a line graph which showed me the number of cars per year. So it gave me a basic idea of what was the number of the cars and their manufacturing year. It was a very straightforward graph which gave quiet a plain picture of the dataset.



So the findings from this graph was that after the year 2012 the manufacturing of BMW cars increased and peaked. It is quite evident that for the year 2019 most number of cars were produced.

This graph gave me a basic idea about the number of cars produced per year.

- The second graph that I created was a graph which showed the fuel type of different models of cars listed in the dataset. Fuel types that were listed were Diesel, Petrol, Hybrid, Electric and others. This graph gave a broader picture of which car model has what kind of fuel type. This would help us in dissecting which car to go for and which model to choose.



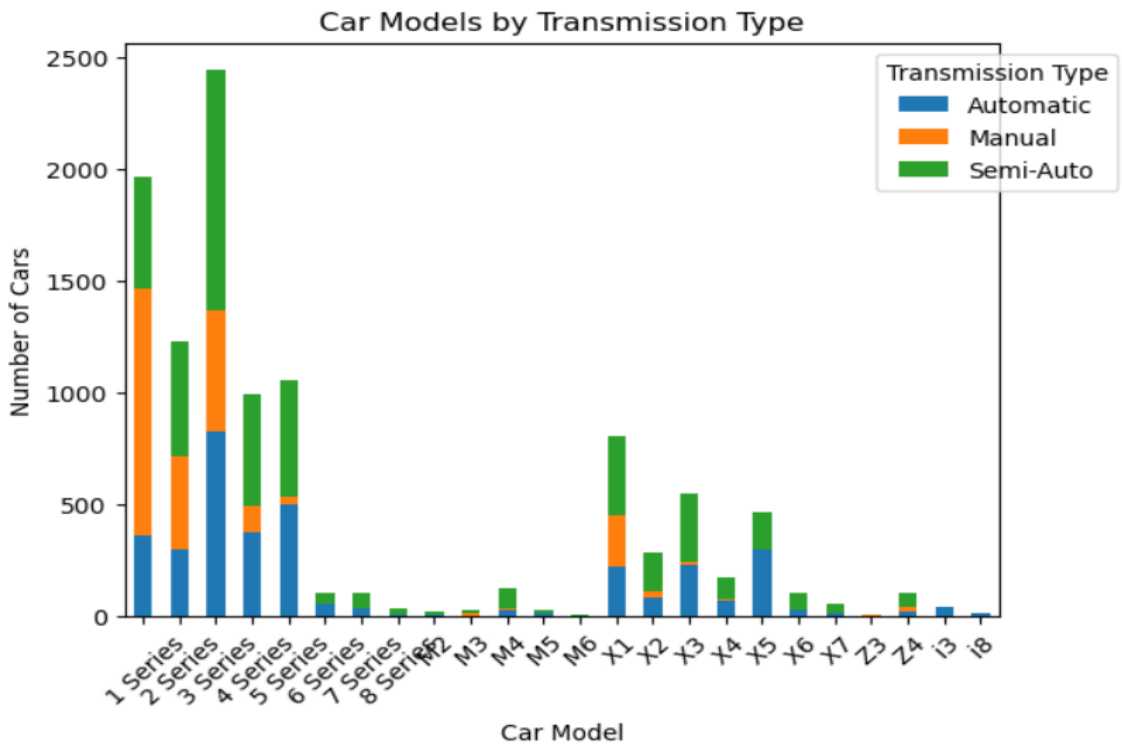
The findings from this graph was that BMW 3 series had the most number of Diesel cars among the lot.

It also shows BMW 2 series was the first car with hybrid mode of fuel type. And BMW 3 series and 5 series have almost equal number of hybrid models.

It shows that BMW 1 series was the car with most number of petrol engine.

This graph also shows that BMW 3 series is also the most manufactured car among others and therefore it stand out in every graph.

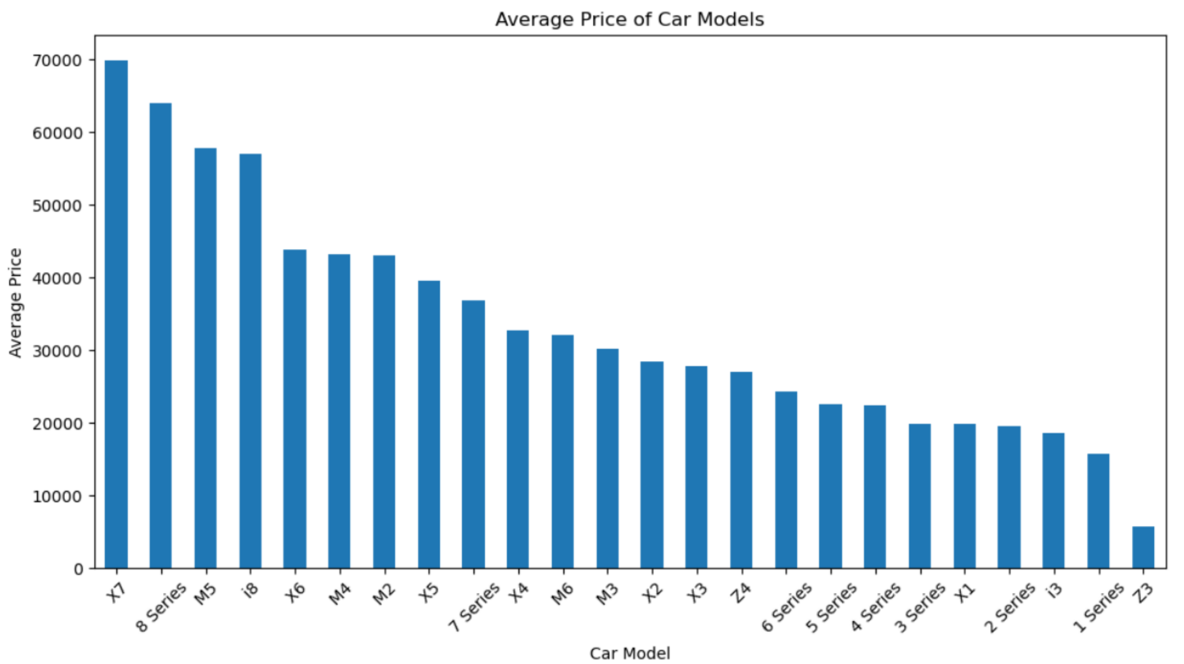
- The third graph that I created was a graph which showed the transmission type of different models of cars listed in the dataset. Transmission types that were listed were automatic, manual and semi automatic. This graph gave a broader picture of which car model has what kind of transmission type. This would help us in dissecting which car to go for and which model to choose.



The findings from this graph are that BMW 1 Series was the only car which had the most number of manual cars and as the time progressed change was visible as BMW 3 series had the most number of automatic and semi-automatic cars.

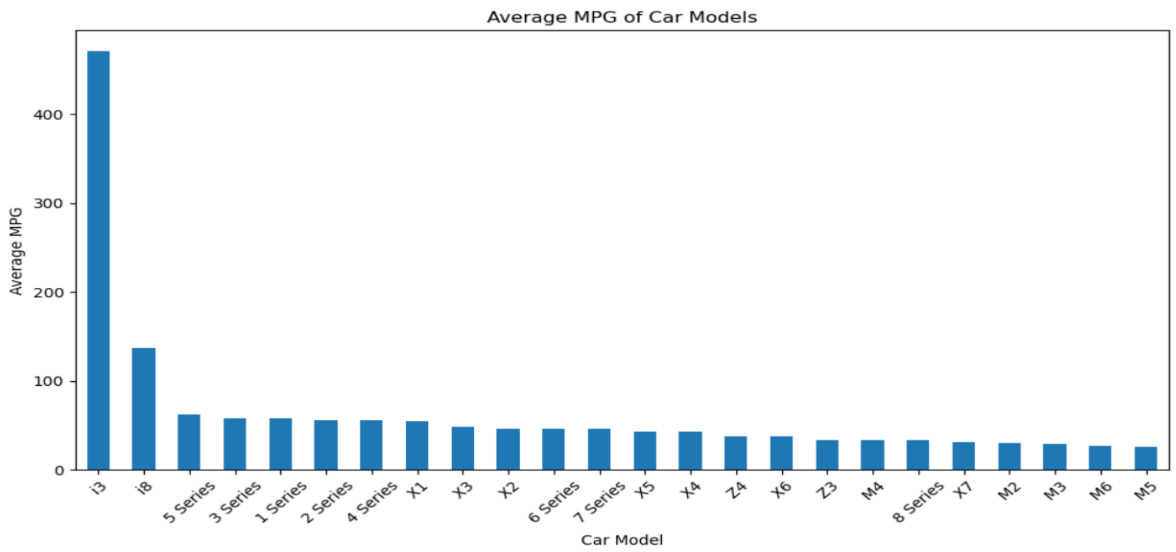
One thing is prominent that among the used cars which are on the dataset BMW 1-5 series fill up about 60-70% of the cars. So we can introspect that may be the owners of these cars want to shift to a better version of their cars.

- The fourth graph that I created was a graph which showed the Average prices of different models of cars listed in the dataset. This graph gave a broader picture of which car model has what average price . This would help us in dissecting which car to go for and which model to choose.



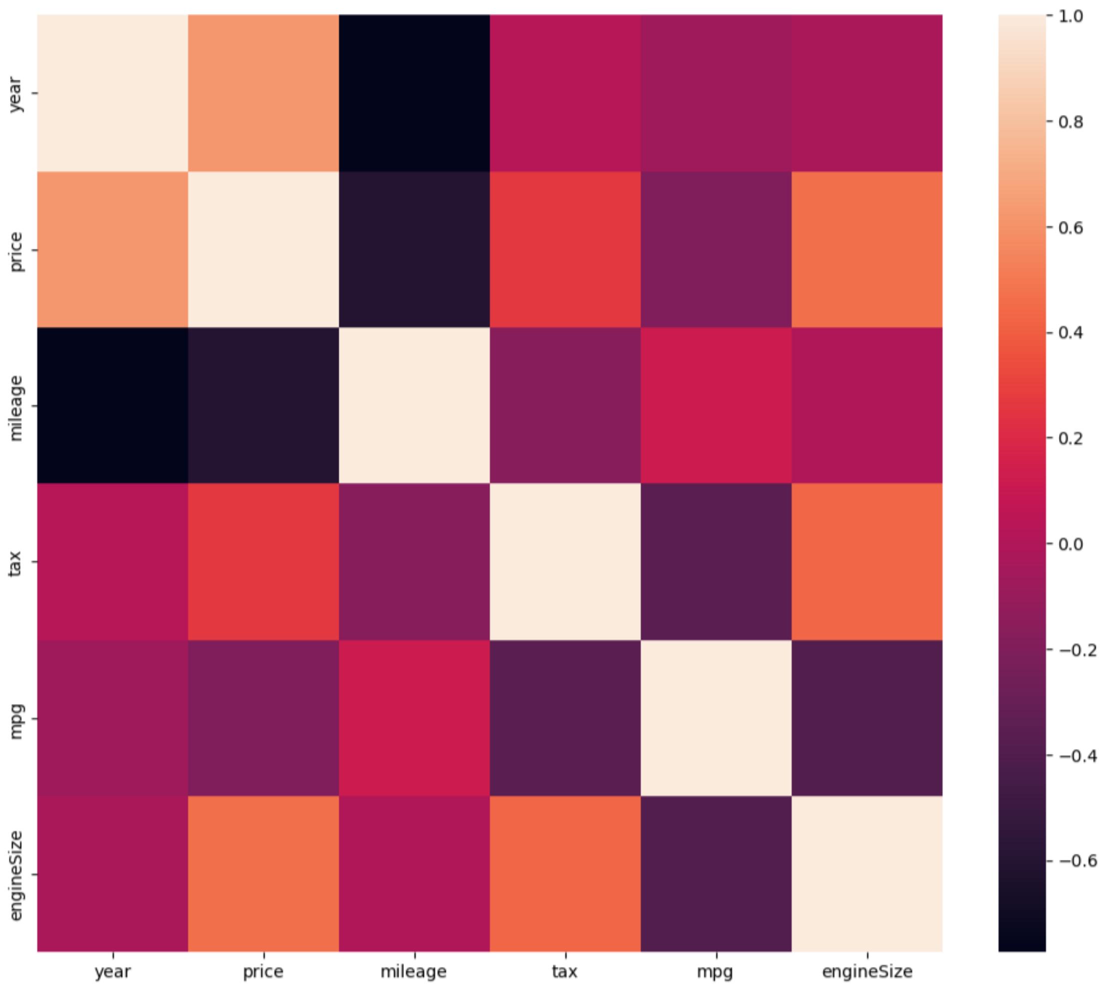
This graph is self explanatory in nature as it clearly states that BMW X7 has the highest average price and BMW Z3 has the lowest average price.

- The fifth graph that I created was a graph which showed the Average mpg(miles per gallon) of different models of cars listed in the dataset. This graph gave a broader picture of which car model has what average mpg . This would help us in dissecting which car to go for and which model to choose.



This graph shows that BMW i3 has the highest average MPG (miles per gallon) so it is actually a good car that can be used again. Higher the average MPG of a car higher would be its efficiency and higher would be its reusable value. So any person would like to buy a used car which has higher MPG value but a car with higher MPG value would have higher average price also so any sensible person has to draw a balance between both.

- Other than this I also created a Heatmap which tells us the correlation between different variables in a dataset.
- A correlation of -1.0 indicates a perfect negative correlation and a correlation of 1.0 indicates a perfect positive correlation.
- In the below Heat map, lighter the color stronger is the correlation between the values/variables


year
price
mileage
tax
mpg
engineSize
year
price
mileage
tax
mpg
engineSize
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6

MANAGERIAL INSIGHTS

1. **Pricing strategy** : This insight can help dealerships maximize their profit margins and remain competitive in the market.

2. **Demand Analysis**: Identify which BMW cars are high in demand in used car market.

3. **Seasonal trends**: Explore whether there are seasonal fluctuations in BMW used cars prices and demand.

4. **Mileage v/s Price:** Understand how mileage impacts the resale value of used BMW cars.

5. **Feature Preferences**: Identify which features have the most significant impact on the resale value of BMW cars.

6. **Competitive Analysis:** Analyse the competitive landscape by comparing the prices and features of used BMW cars to other luxury car brands.

7. **Age v/s Price**: Examine how the age of a BMW car affects its price and try to forecast future values.