Project: Analyzing the Impact of Car Features on Price and Profitability

BY: Muskan Bansal

Problem Statement:

The automotive industry has been rapidly evolving over the past few decades, with a growing focus on fuel efficiency, environmental sustainability, and technological innovation. With increasing competition among manufacturers and a changing consumer landscape, it has become more important than ever to understand the factors that drive consumer demand for cars.
In recent years, there has been a growing trend towards electric and hybrid vehicles and increased interest in alternative fuel sources such as hydrogen and natural gas. At the same time, traditional gasoline-powered cars remain dominant in the market, with varying fuel types and grades available to consumers.

For the given dataset, as a Data Analyst, the client has asked How can a car manufacturer optimize pricing and product development decisions to maximize profitability while meeting consumer demand?

This problem could be approached by analyzing the relationship between a car's features, market category, and pricing, and identifying which features and categories are most popular among consumers and most profitable for the manufacturer. By using data analysis techniques such as regression analysis and market segmentation, the manufacturer could develop a pricing strategy that balances consumer demand with profitability, and identify which product features to focus on in future product development efforts. This could help the manufacturer improve its competitiveness in the market and increase its profitability over time.

Introduction:-

This project involves conducting an exploratory data analysis (EDA) in a real business scenario related to car data. The goal is to develop a basic understanding of how car features impact the pricing and profitability of vehicles in the automotive industry. By analyzing patterns present in the data, EDA can help identify the car features that significantly affect the vehicle's price and profitability, which can be used by the automotive industry to make data-driven decisions on production and pricing strategies. The project includes a dataset consisting of car features such as the make, model, year, engine size, fuel type, mileage, and other attributes, as well as the price and profitability of each vehicle. The aim of EDA in this case study is to identify patterns indicating which car features are correlated with higher prices and profitability, which can lead to actions such as prioritizing production of these features, adjusting pricing strategies, or changing marketing approaches. This will help ensure that the automotive industry is producing and pricing vehicles in a way that maximizes profitability while meeting consumer demand, and the project's goal is to identify such patterns using EDA.

Understanding the Dataset:

The dataset contains information on over 11,000 car models and their specifications, including details on the car's make, model, year, fuel type, engine power, transmission, wheels, number of doors, market category, size, style, estimated miles per gallon, popularity, and manufacturer's suggested retail price (MSRP).
This dataset could be useful for a variety of data analysis tasks, such as:

- Exploring trends in car features and pricing over time
- Comparing the fuel efficiency of different types of cars
- Investigating the relationship between a car's features and its popularity
- Predicting the price of a car based on its features and market category

However, it's important to note that the dataset was last updated in 2017, so it may not reflect current trends or prices in the automotive industry.
Here is a brief overview of the dataset:

- **Number of observations:** 11,159
- **Number of variables:** 16
- **File type:** CSV (Comma Separated Values)

The variables in the dataset are:

- **Make:** the make or brand of the car
- **Model:** the specific model of the car
- **Year:** the year the car was released
- **Engine Fuel Type**: the type of fuel used by the car (gasoline, diesel, etc.)
- **Engine HP:** the horsepower of the car's engine
- **Engine Cylinders:** the number of cylinders in the car's engine
- **Transmission Type**: the type of transmission (automatic or manual)
- **Driven_Wheels:** the type of wheels driven by the car (front, rear, all)
- **Number of Doors:** the number of doors the car has
- **Market Category:** the market category the car belongs to (Luxury, Performance, etc.)
- **Vehicle Size:** the size of the car
- **Vehicle Style:** the style of the car (Sedan, Coupe, etc.)
- **Highway MPG:** the estimated miles per gallon the car gets on the highway
- **City MPG:** the estimated miles per gallon the car gets in the city
- **Popularity:** a ranking of the popularity of the car (based on the number of times it has been viewed on Edmunds.com)
- **MSRP:** the manufacturer's suggested retail price of the car

Tech used:-

- Microsoft Excel 2019
- Microsotf Word

Functions Used: COUNTIFS, AVERAGEIFS, Pivot Tables, Regression Analysis, Ancoding Categorical variables, MIN, MAX, Conditional Formatting, CORREL, VBA, etc.

Approach:-

To implement this project, I utilized Microsoft Office Excel. Firstly, I imported the given datasets into the software and carefully analyzed each column and its attributes. I also checked the connections between different columns. During the analysis, I identified null and duplicate values and removed them to improve the accuracy of the results. Pivot tables were very helpful in performing the project tasks as they allowed me to connect different columns and perform various analyses.

Exploratory Data Analysis:-

A. Cleaning:-

➢ For cleaning the data I have to calculate the null values.

➢ Here in this data set I have multiple types of data present related to car, I have 16 columns and 11159 rows each columns represents the multiple type of data.

➢ I used countblank(range) function to calculate the null values present in the dataset.

➢ Some of the formulas I used to calculate to find the percentage of the blank values is : **=countblank(range)/counta(range)*100**

➢ we have to calculate the **OUTLIERS** using the various techniques.

➢ To find the outliers I used the box plot and after finding the box plot I will remove those outliers. Some of the outliers are given below.

OUTLIERS:
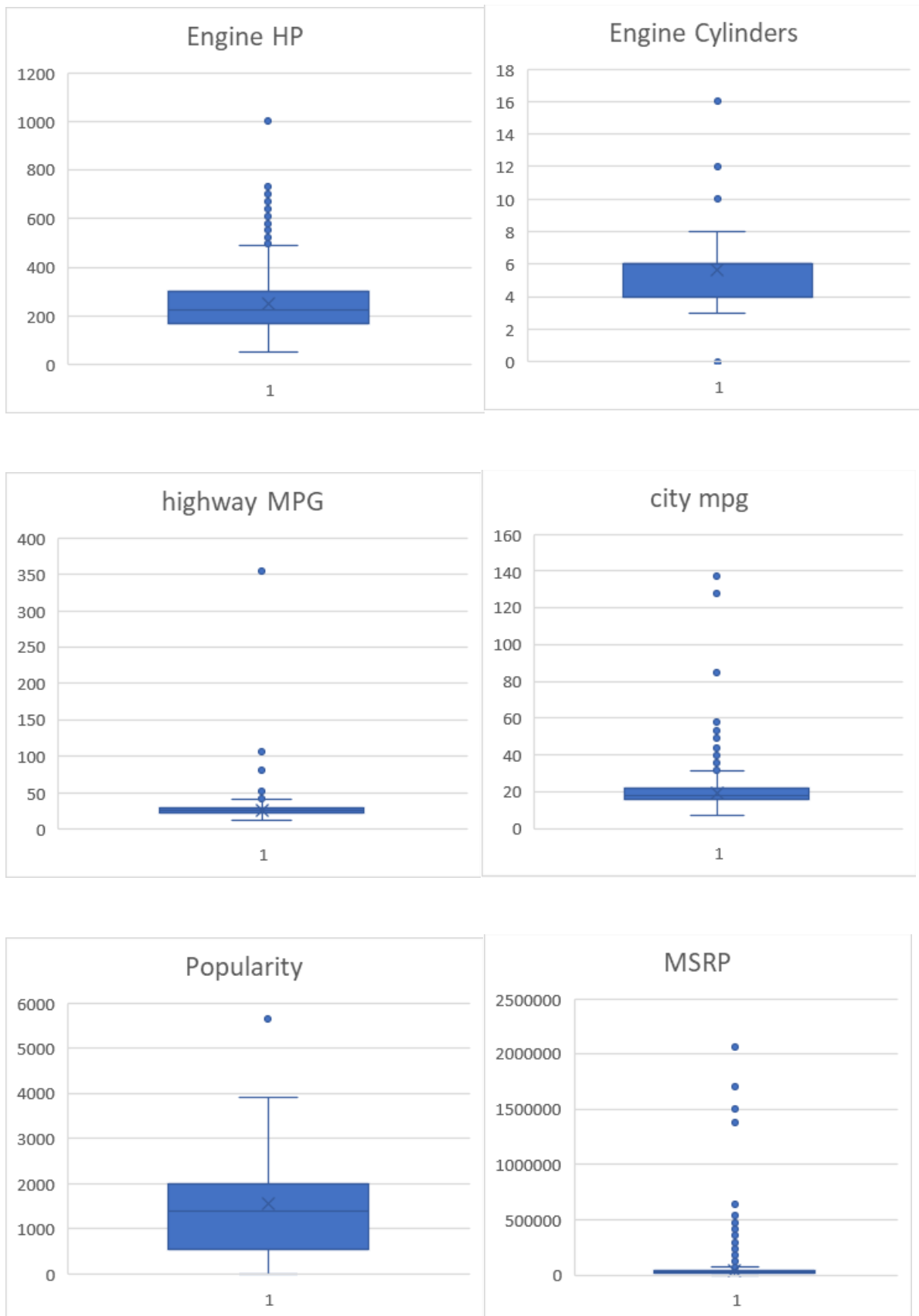
Tasks: Analysis

<u>Insight Required:</u> How does the popularity of a car model vary across different market categories?

- **Task 1.A:** Create a pivot table that shows the number of car models in each market category and their corresponding popularity scores.

Row Labels	Average of Popularity	Count of Model
Crossover	1184.357564	1018
Crossover,Diesel	873	7
Crossover,Exotic,Luxury,High-Performance	238	1
Crossover,Exotic,Luxury,Performance	238	1
Crossover,Factory Tuner,Luxury,High-Performance	2270.428571	7
Crossover,Factory Tuner,Luxury,Performance	2607.4	5
Crossover,Factory Tuner,Performance	210	4
Crossover,Flex Fuel	1246.846154	52
Crossover,Flex Fuel,Luxury	1173.2	10
Crossover,Flex Fuel,Luxury,Performance	1624	6
Crossover,Hatchback	1675.694444	72
Crossover,Hatchback,Factory Tuner,Performance	2009	6
Crossover,Hatchback,Luxury	204	7
Crossover,Hatchback,Performance	2009	6
Crossover,Hybrid	1299	21
Crossover,Luxury	884.5487805	410
Crossover,Luxury,Diesel	2237.1875	32
Crossover,Luxury,High-Performance	843.5714286	7
Crossover,Luxury,Hybrid	737.0666667	15
Crossover,Luxury,Performance	1334.754545	110
Crossover,Luxury,Performance,Hybrid	3916	2
Crossover,Performance	1658.849057	53
Diesel	960.7241379	58
Diesel,Luxury	1915.641026	39
Exotic,Factory Tuner,High-Performance	2774	2
Exotic,Factory Tuner,Luxury,High-Performance	497.25	20
Exotic,Factory Tuner,Luxury,Performance	520	3
Exotic,High-Performance	1109.549451	91
Exotic,Luxury	303	2
Exotic,Luxury,High-Performance	487.4181818	55
Exotic,Luxury,Performance	520	11
Factory Tuner,High-Performance	1305.131148	61
Factory Tuner,Luxury,High-Performance	2242.602941	136
Factory Tuner,Luxury,Performance	1330	30
Factory Tuner,Performance	1179.730769	78
Flex Fuel	1265.468521	683
Flex Fuel,Hybrid	155	2
Flex Fuel,Luxury	746.5384615	39
Flex Fuel,Luxury,High-Performance	1624	15
Flex Fuel,Luxury,Performance	1380.071429	28
Flex Fuel,Performance	1014.594203	69
Flex Fuel,Performance,Hybrid	155	2

Hatchback	1062.332024	509
Hatchback,Diesel	873	3
Hatchback,Factory Tuner,High-Performance	834.1666667	12
Hatchback,Factory Tuner,Luxury,Performance	886.8888889	9
Hatchback,Factory Tuner,Performance	847.3125	16
Hatchback,Luxury	1178	39
Hatchback,Luxury,Performance	1566.131579	38
Hatchback,Performance	974.1538462	247
High-Performance	1671.048387	186
Hybrid	1059.875	32
Luxury	1106.932704	847
Luxury,High-Performance	1662.189781	274
Luxury,High-Performance,Hybrid	568.8333333	12
Luxury,Hybrid	937.0625	32
Luxury,Performance	1285.561086	663
Luxury,Performance,Hybrid	2333.181818	11
N/A	1217.983696	3312
Performance	1068.845725	538
Performance,Hybrid	155	1
Grand Total	**1213.427662**	**10057**

- **Task 1.B:** Create a combo chart that visualizes the relationship between market category and popularity.
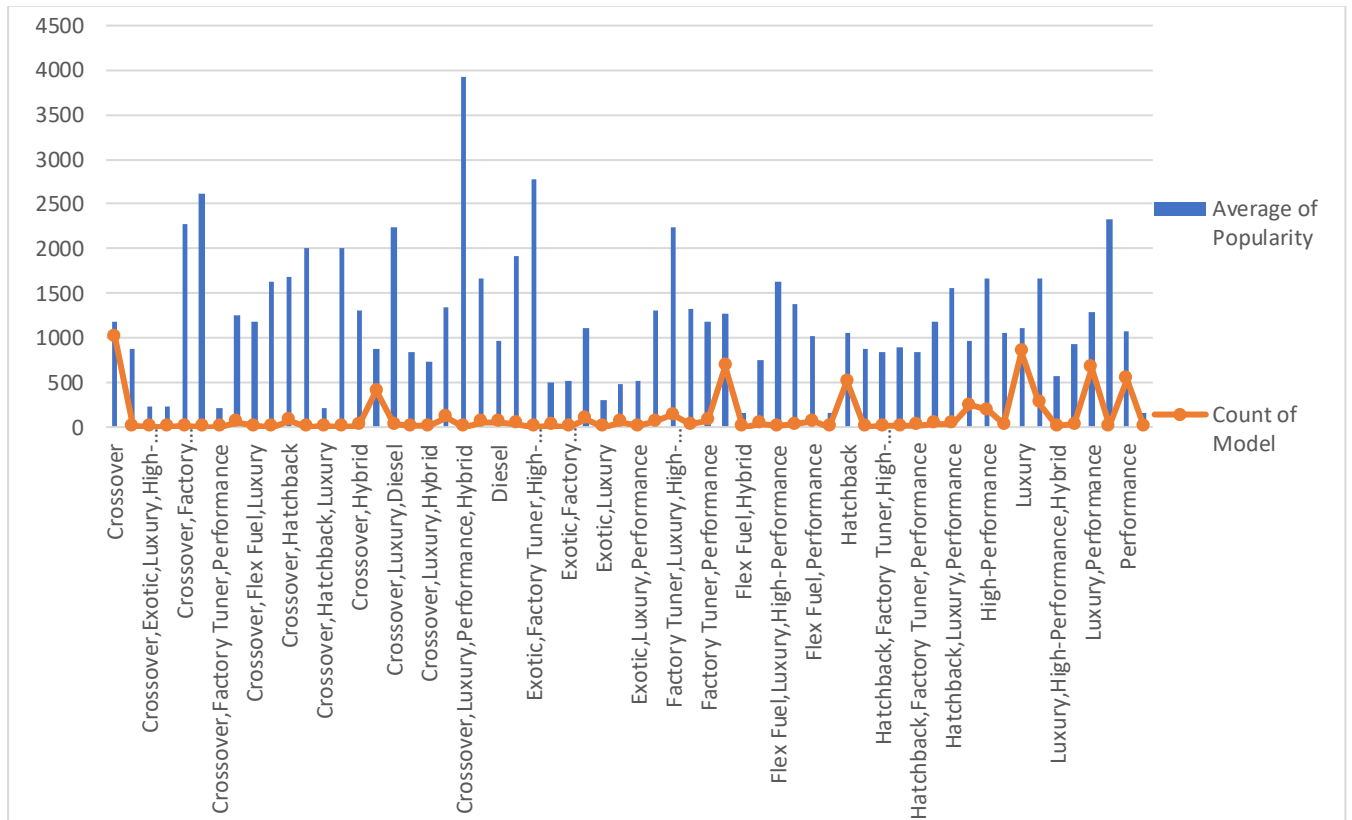
INSIGHT:
From the above table and visualisation, croassover is the most popular followed by exotic and luxury.
Also, crossover seems to have the most number of models.

Insight Required: What is the relationship between a car's engine power and its price?
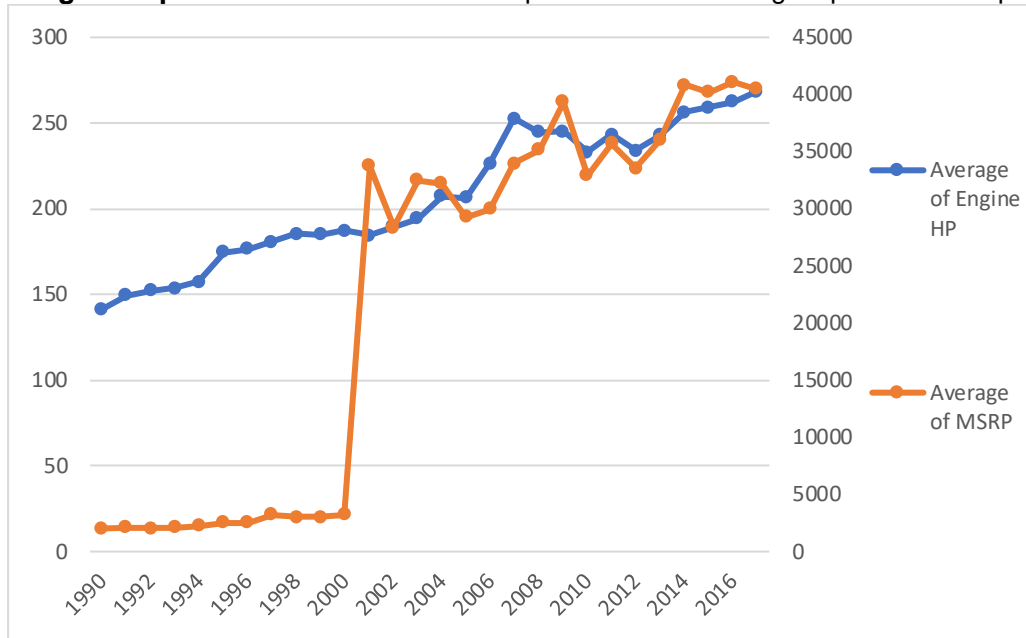
INSIGHT:

There was a huge rise of price after 2000, and since then the rates has fluctuated a lot. It is also observed that with increase of Engine's HP, the price has increased too.
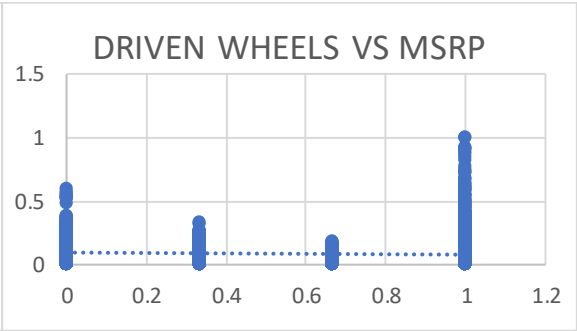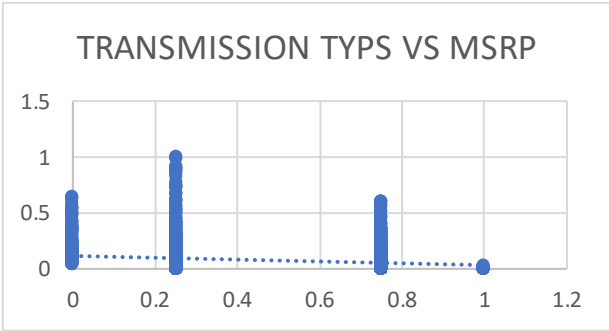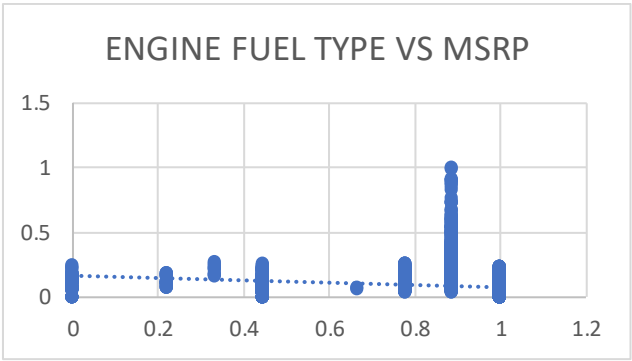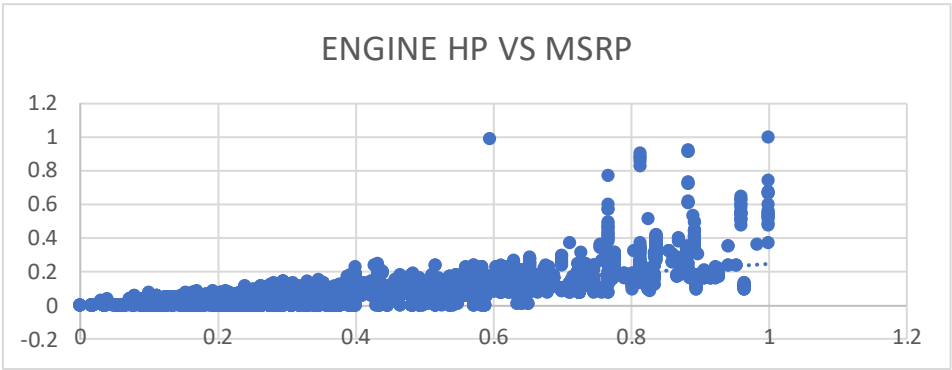
- **Task 2:** Create a scatter chart that plots engine power on the x-axis and price on the y-axis. Add a trendline to the chart to visualize the relationship between these variables.
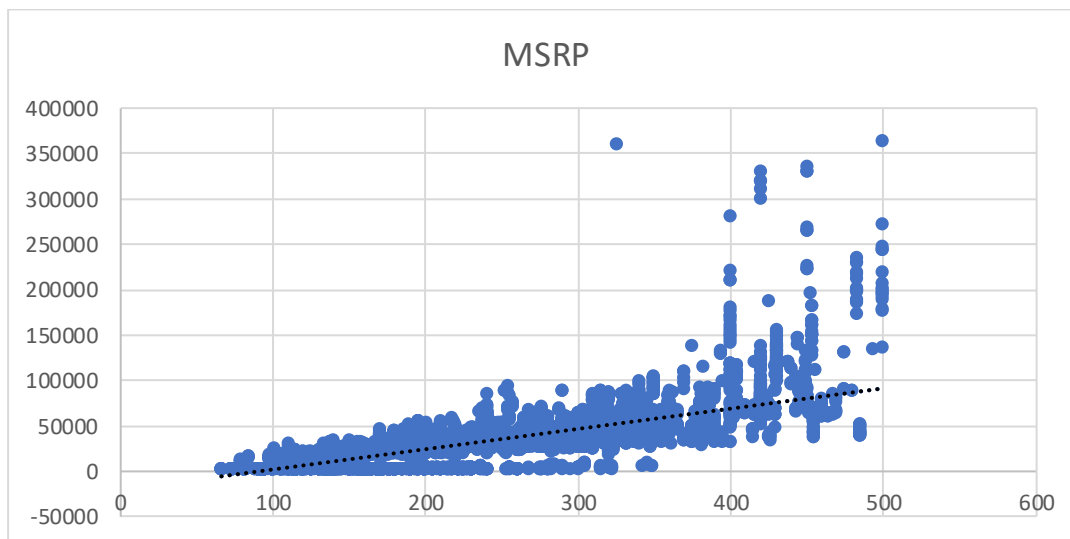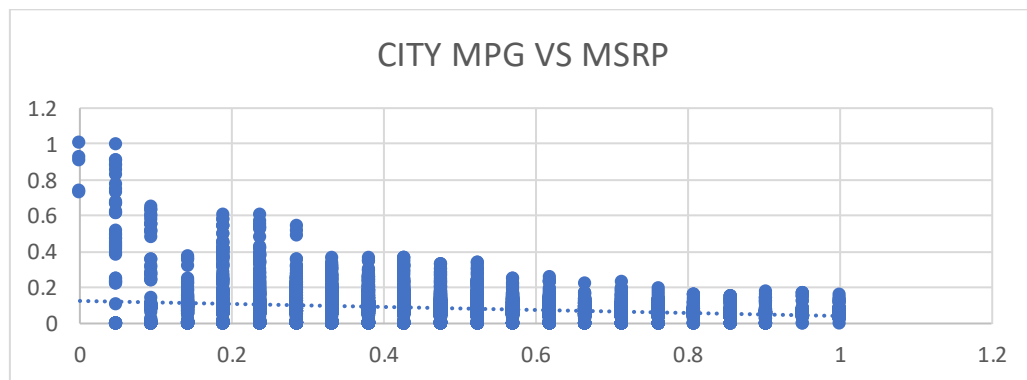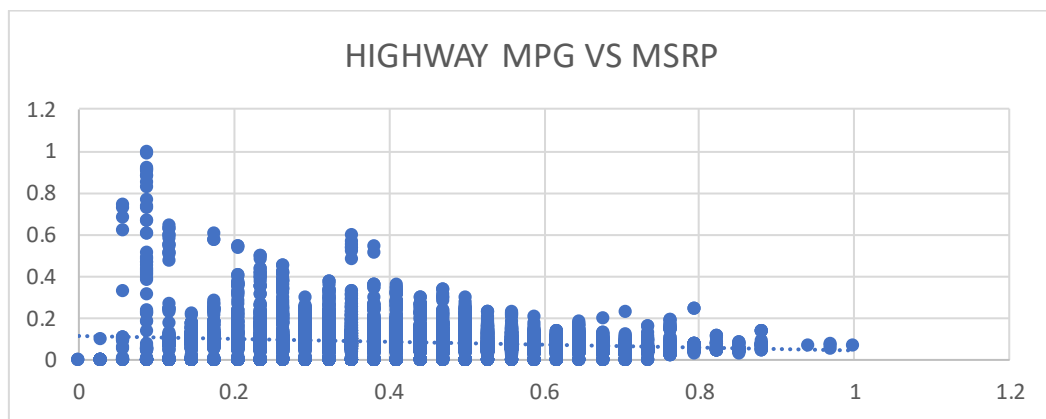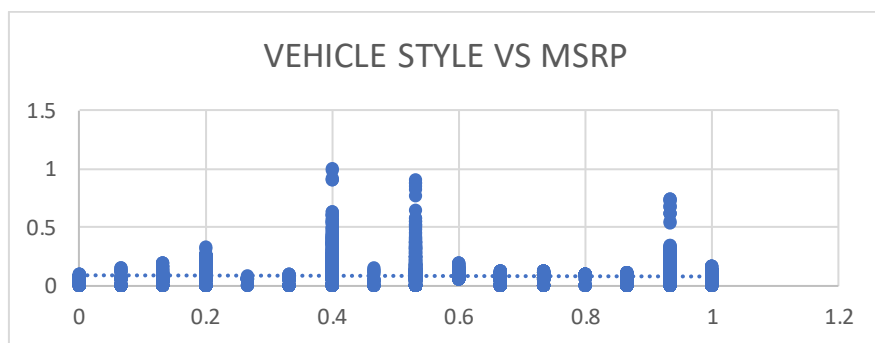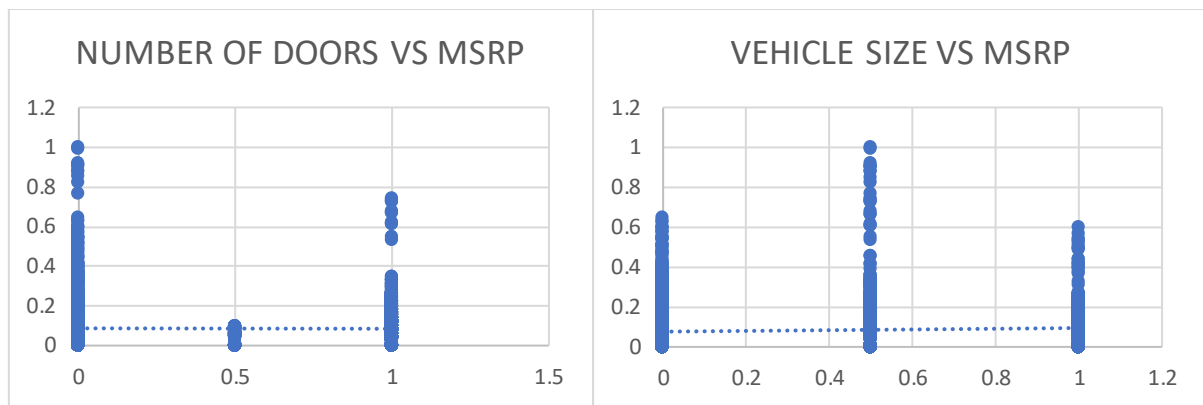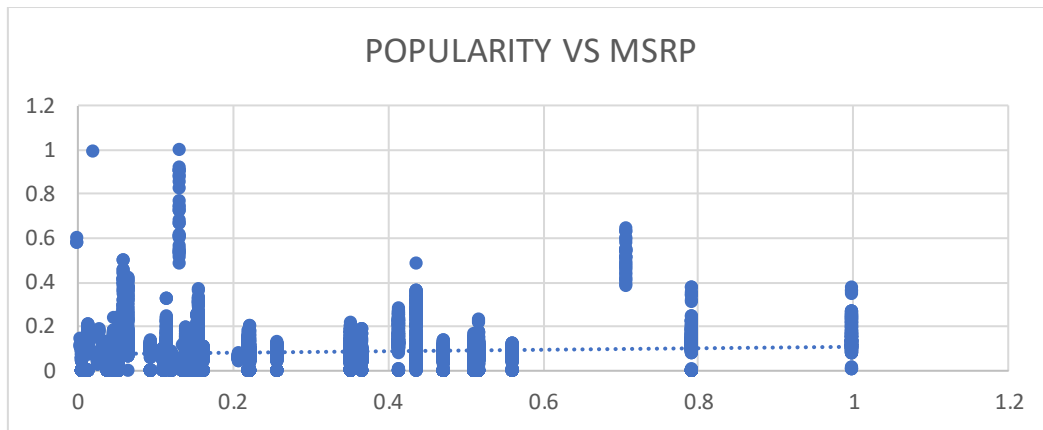

INSIGHT:

I created a scatter chart and also added a trend line, and with the help this visualisation it is visible that with the increase of the HP, price also increases.

**Insight Required:** Which car features are most important in determining a car's price?

ENGINE HP VS MSRP

ENGINE FUEL TYPE VS MSRP

TRANSMISSION TYPS VS MSRP

DRIVEN WHEELS VS MSRP

ENGINE CYLINDERS VS MSRP

NUMBER OF DOORS VS MSRP

VEHICLE SIZE VS MSRP

VEHICLE STYLE VS MSRP

HIGHWAY MPG VS MSRP

CITY MPG VS MSRP
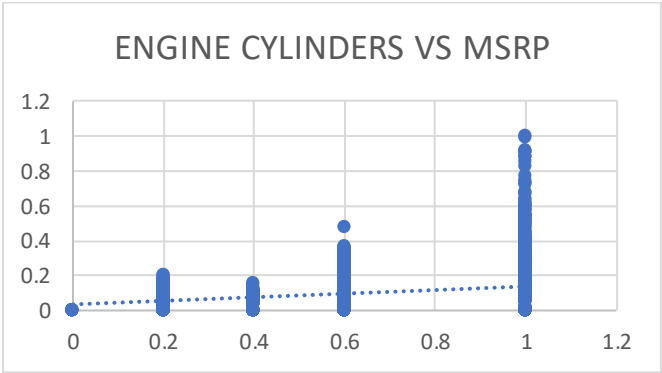
INSIGHT:

Engine Hp, Engine Cylinders and Popularity seems to follow a positive trend line signifying these are the factors that are important in determining the car's prises.

- **Task 3:** Use regression analysis to identify the variables that have the strongest relationship with a car's price. Then create a bar chart that shows the coefficient values for each variable to visualize their relative importance.

Insight Required: How does the average price of a car vary across different manufacturers?
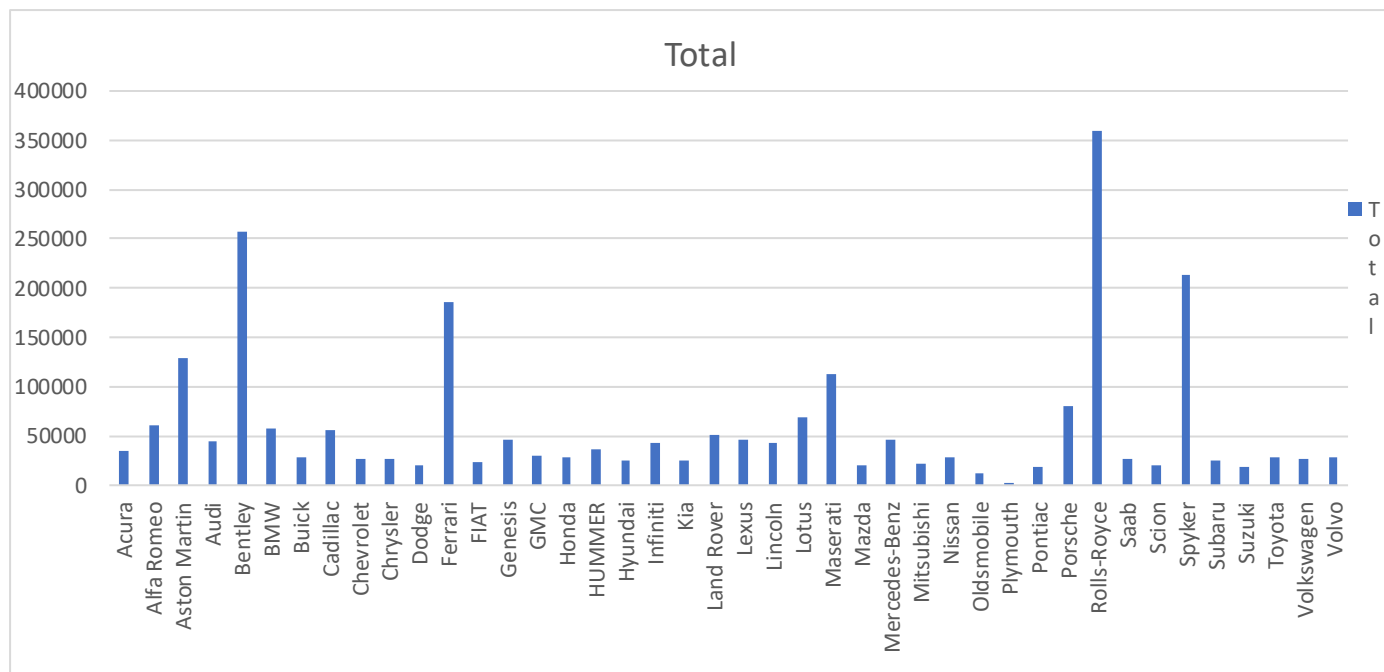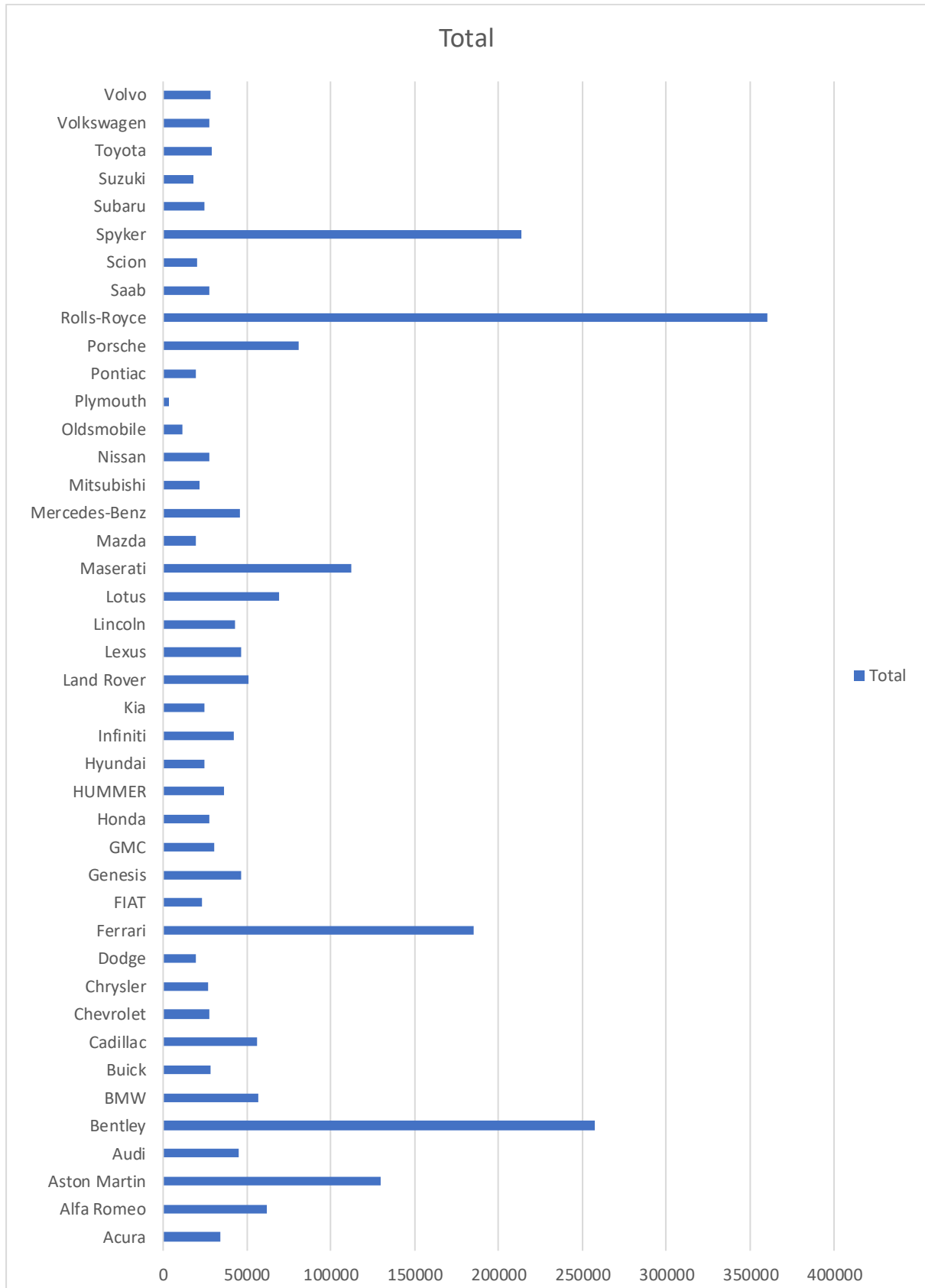

INSIGHT:

Rolls-Royace is the most expensive manufacturer followed by Bentley, Spyker and Ferrari respectively.

- **Task 4.A:** Create a pivot table that shows the average price of cars for each manufacturer.

Row Labels	Average of MSRP
Acura	34426.39357
Alfa Romeo	61600
Aston Martin	129535
Audi	45228.31525
Bentley	257094.8276
BMW	56994.63793
Buick	28206.61224
Cadillac	55936.85052
Chevrolet	27387.8153
Chrysler	26722.96257
Dodge	19649.94518
Ferrari	185189.1034
FIAT	23008.69565
Genesis	46616.66667
GMC	30493.29903
Honda	27711.86857
HUMMER	36464.41176
Hyundai	24504.59864
Infiniti	42394.21212
Kia	24972.91284
Land Rover	50930.28037
Lexus	46534.98333
Lincoln	42602.23179
Lotus	69188.27586
Maserati	112056.4259
Mazda	19630.15038
Mercedes-Benz	45542.8037
Mitsubishi	21795.25389
Nissan	27668.66472
Oldsmobile	11542.54
Plymouth	3122.902439
Pontiac	19321.54839
Porsche	80830.69725
Rolls-Royce	359990
Saab	27413.5045
Scion	20347.68519
Spyker	213323.3333
Subaru	24827.50391
Suzuki	18087.27405
Toyota	29018.68053
Volkswagen	27528.49804
Volvo	28541.16014
Grand Total	**32867.48941**

- **Task 4.B:** Create a bar chart or a horizontal stacked bar chart that visualizes the relationship between manufacturer and average price.

Insight Required: What is the relationship between fuel efficiency and the number of cylinders in a car's engine?
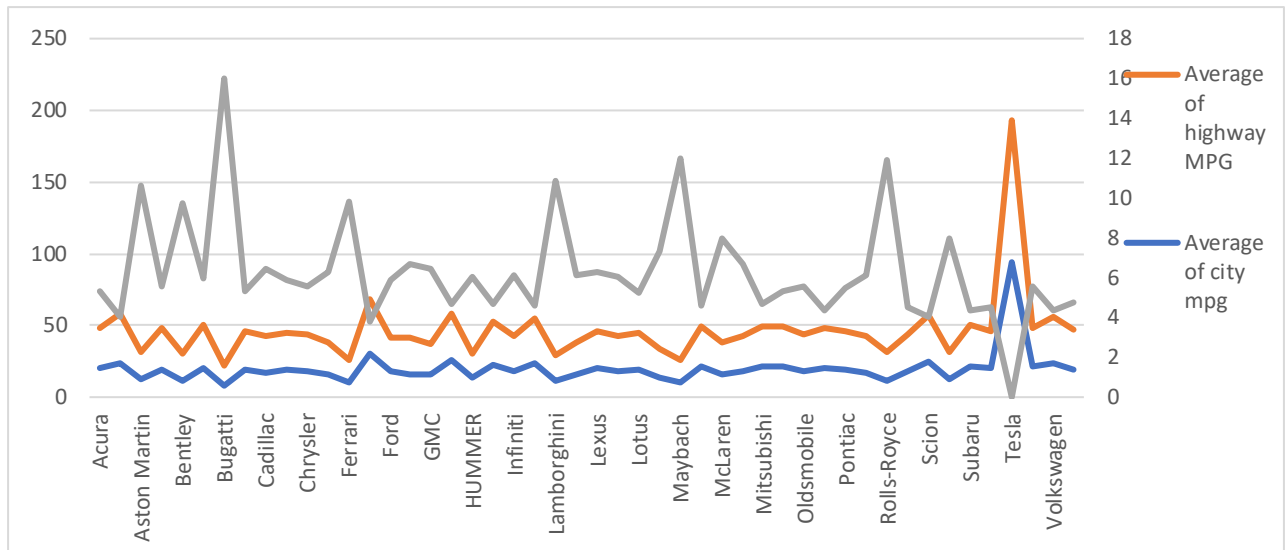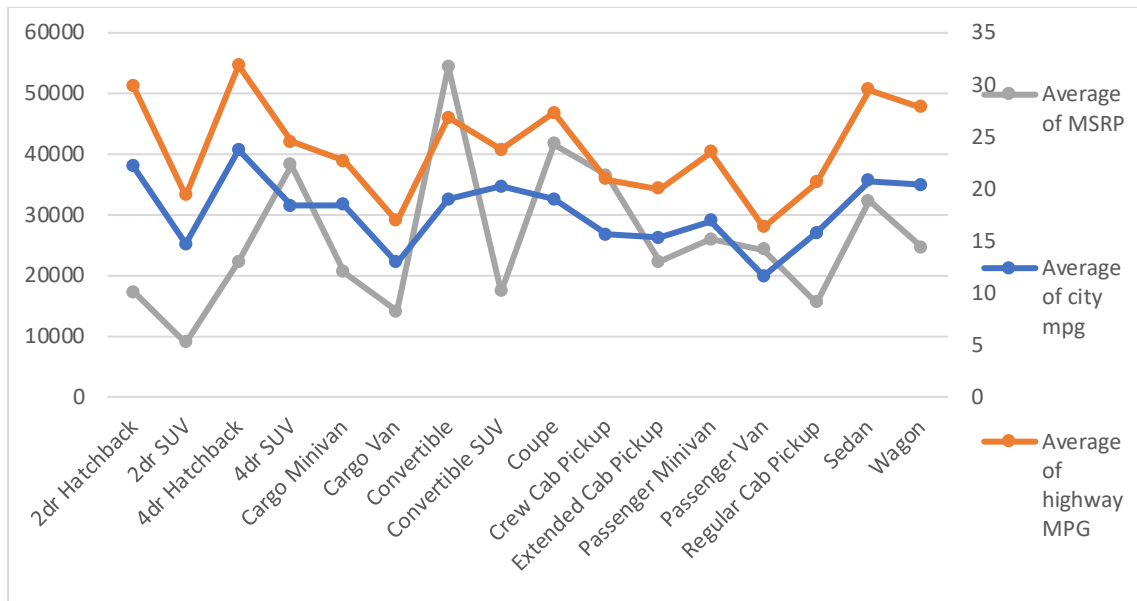

INSIGHT:
Mostly, you can see that with the increase in the number of engine cylinders, the fuel efficiency also increases, however, in case of Tesla, even with significant low number of Engine Cylinders compared to others, the fuel efficieny is the highest.
Also, Bughati has the most number of average cylinders but still the efficiency is pretty low.
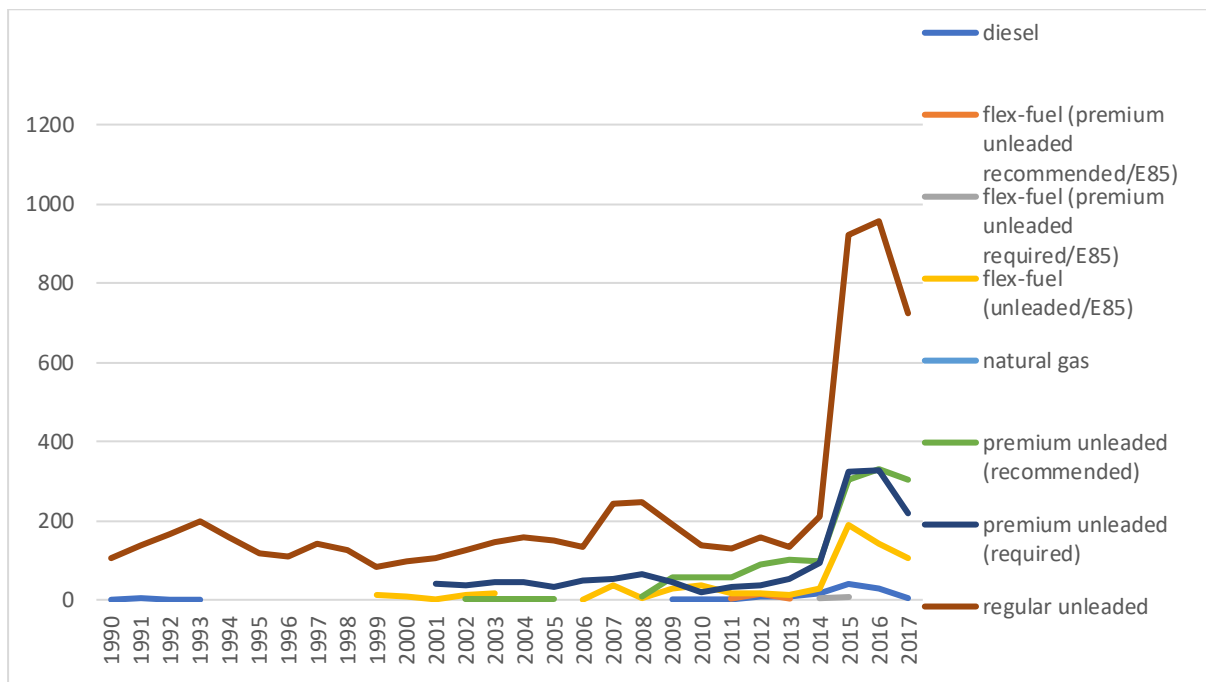
INSIGHT:
Most commonly regular unleaded fuel is used followed by premium unleaded and flex-fuel.

INSIGHT:

Convertible is the most expensive, while 2 dr Hatchback, 4Ha Hatchback, Sedan and Wagon seems most desirable, Since, it has lower price and better avg city and Highway MPG.

- **Task 5.A:** Create a scatter plot with the number of cylinders on the x-axis and highway MPG on the y-axis. Then create a trendline on the scatter plot to visually estimate the slope of the relationship and assess its significance.
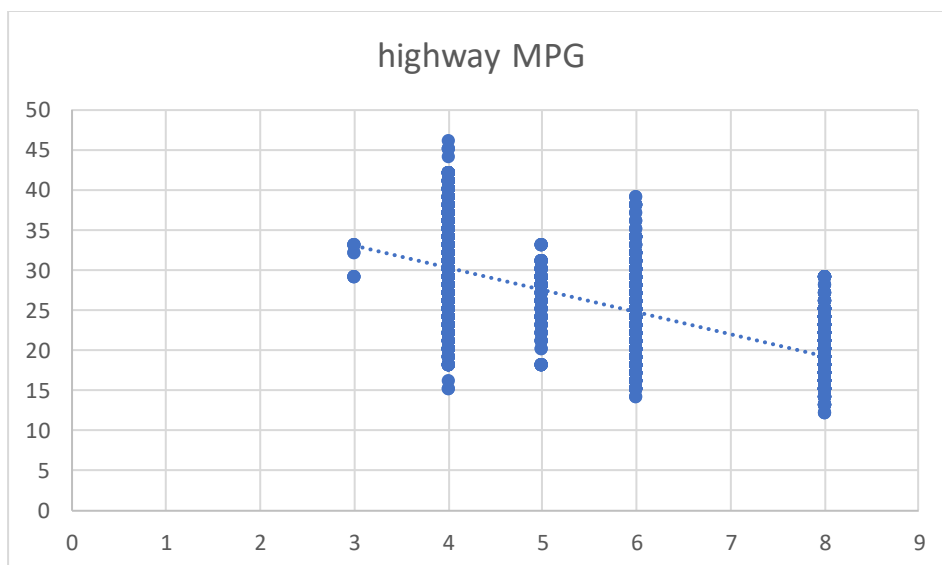
- **Task 5.B:** Calculate the correlation coefficient between the number of cylinders and highway MPG to quantify the strength and direction of the relationship.

FORMULA USED: =CORREL(A:A,B:B)
correlation coefficient between the number of cylinders and highway MPG IS: -0.70134

Building the Dashboard:

Task 1: How does the distribution of car prices vary by brand and body style?

- **Hints:** Stacked column chart to show the distribution of car prices by brand and body style. Use filters and slicers to make the chart interactive. Calculate the total MSRP for each brand and body style using SUMIF or Pivot Tables.
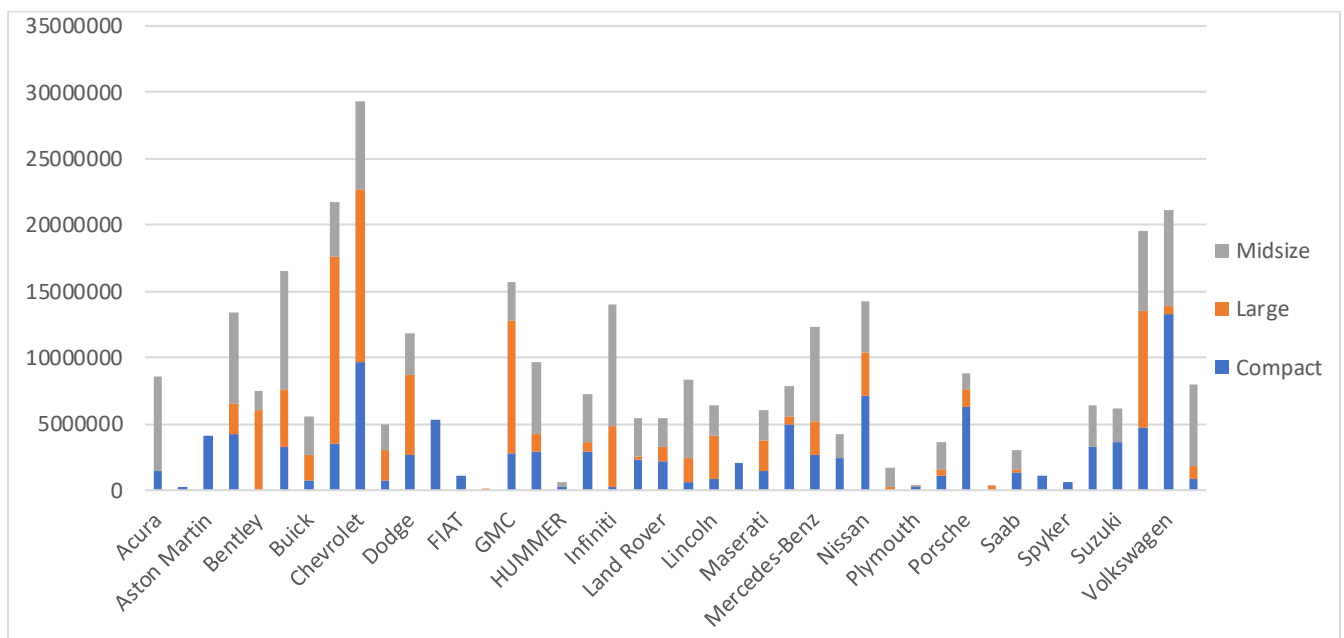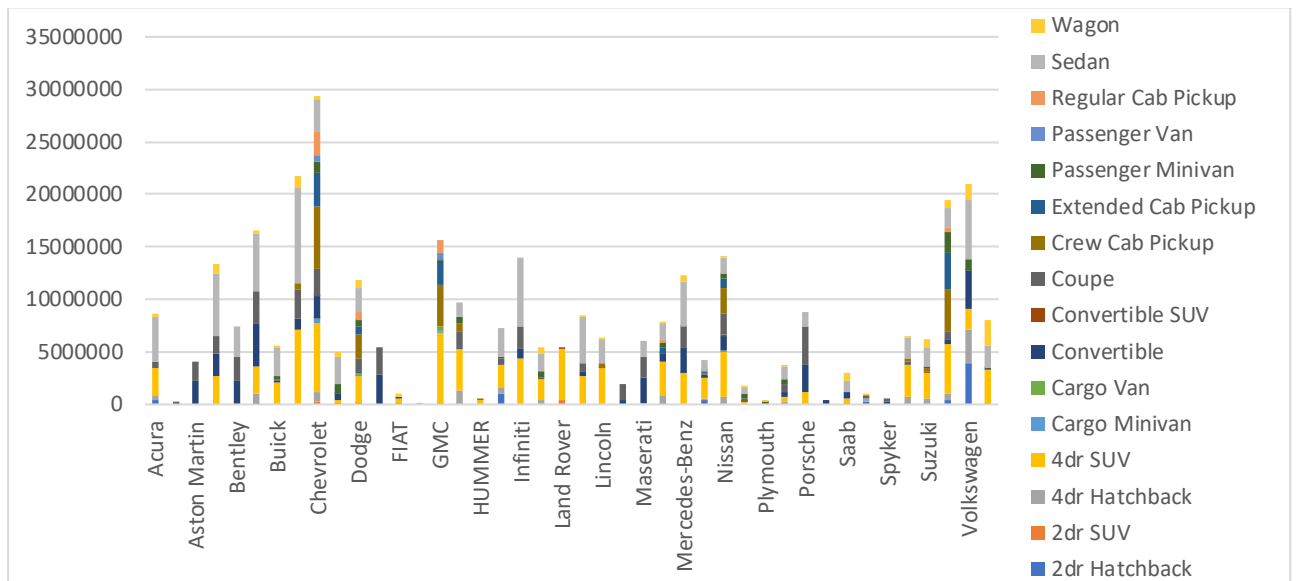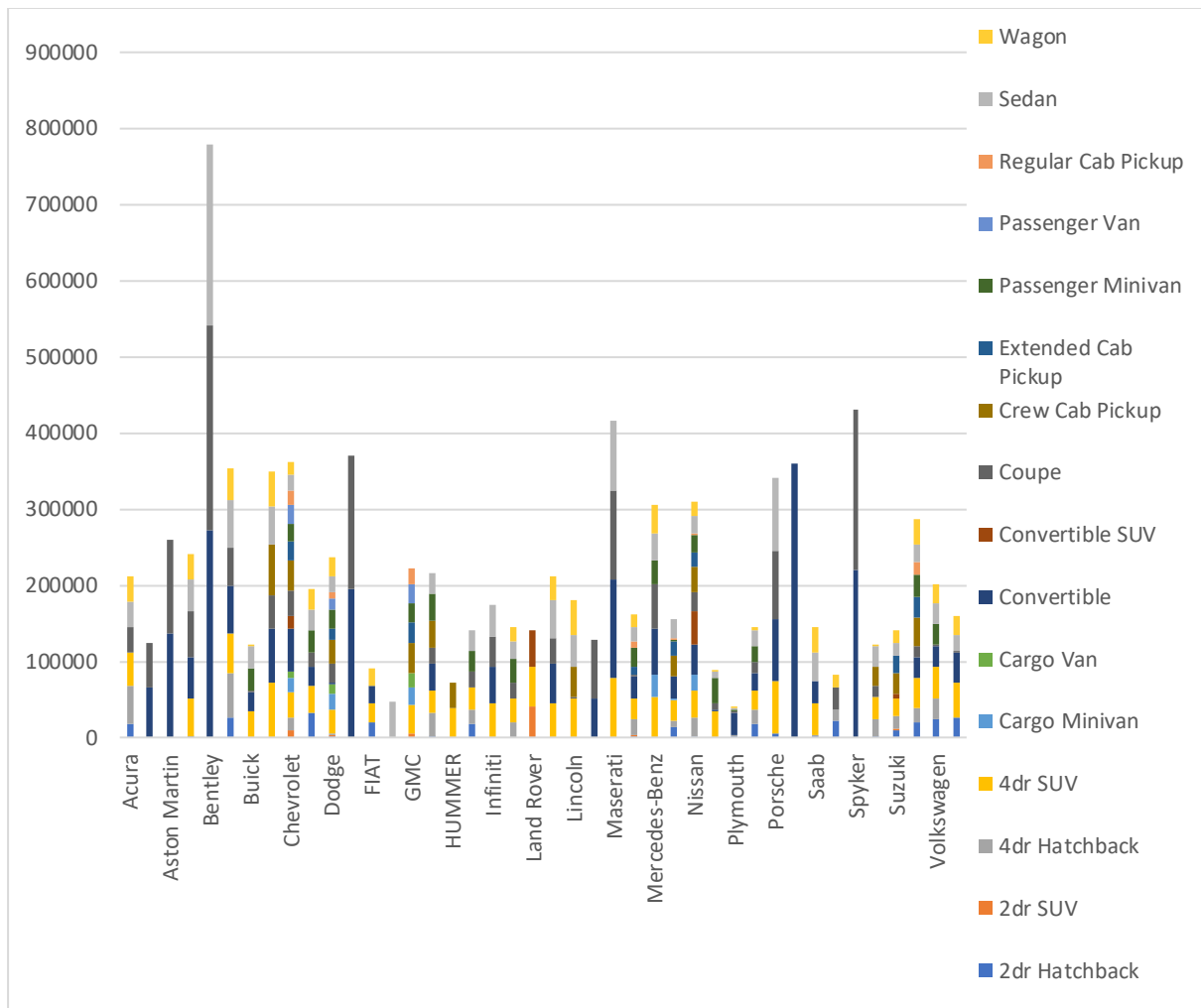

INSIGHT:

Chevrolet is the most expensive followed by Cadillac followed by Volkswagon followed by Toyota.
Mostly, large and compact are preferred over Midsize.

Task 2: Which car brands have the highest and lowest average MSRPs, and how does this vary by body style?

- **Hints:** Clustered column chart to compare the average MSRPs across different car brands and body styles. Calculate the average MSRP for each brand and body style using AVERAGEIF or Pivot Tables.
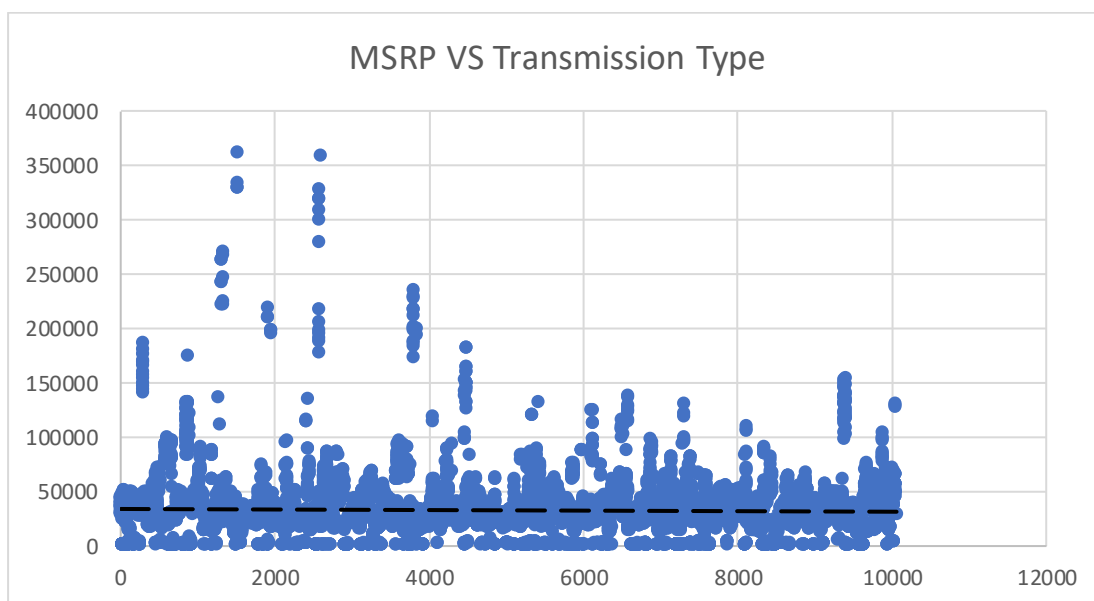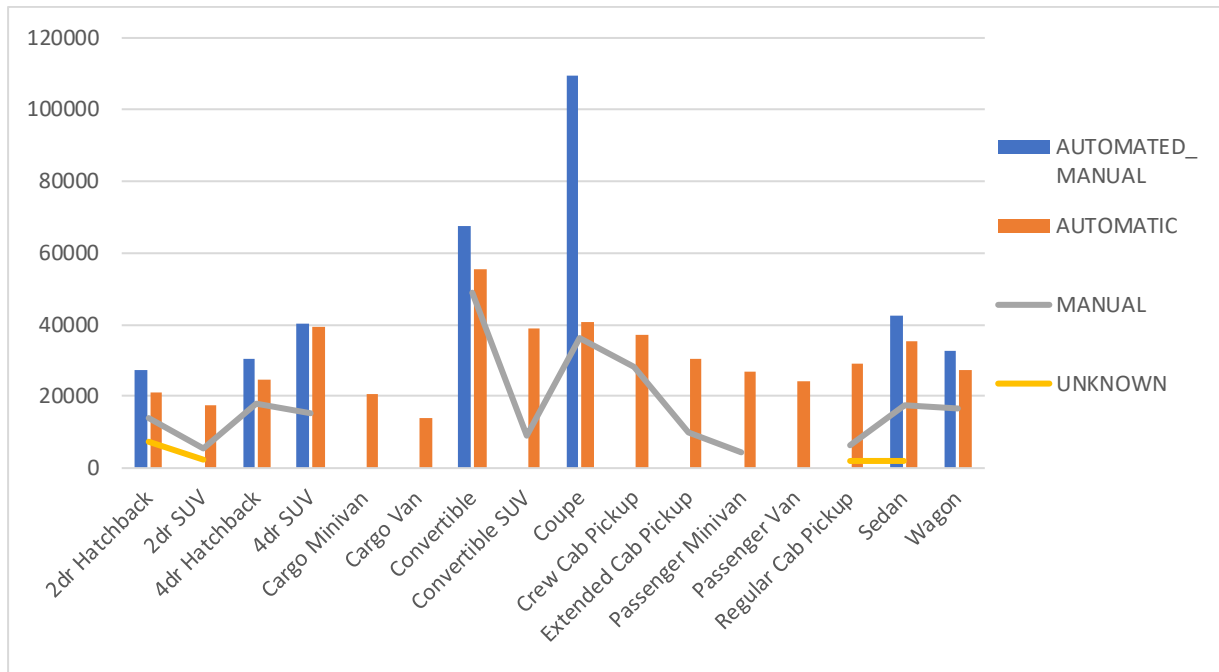

INSIGHT:

Bentley is the most pricey followed by Spyker and Maserati while least expensive ones are Plymouth followed by Genesis and Hummer, FIAT, Skion, Oldsmobile.

For Bentlay,Coupe, Convertible and Sedan, almost all three are equally popular.
While, in case of Plymouth, Convertible is more popular than Sedan.

Task 3: How do the different feature such as transmission type affect the MSRP, and how does this vary by body style?

- **Hints:** Scatter plot chart to visualize the relationship between MSRP and transmission type, with different symbols for each body style. Calculate the average MSRP for each combination of transmission type and body style using AVERAGEIFS or Pivot Tables.


INSIGHT:
So it is visible from the above 2 visualisations that there is no definite relationship between Transmission Type and the MSRP, even with the help of a trendline, there does not seem to have any relation.
So, either there is no relationship betweep the two (which IS the case), or we might need some more data.

Task 4: How does the fuel efficiency of cars vary across different body styles and model years?
- **Hints:** Line chart to show the trend of fuel efficiency (MPG) over time for each body style. Calculate the average MPG for each combination of body style and model year using AVERAGEIFS or Pivot Tables.
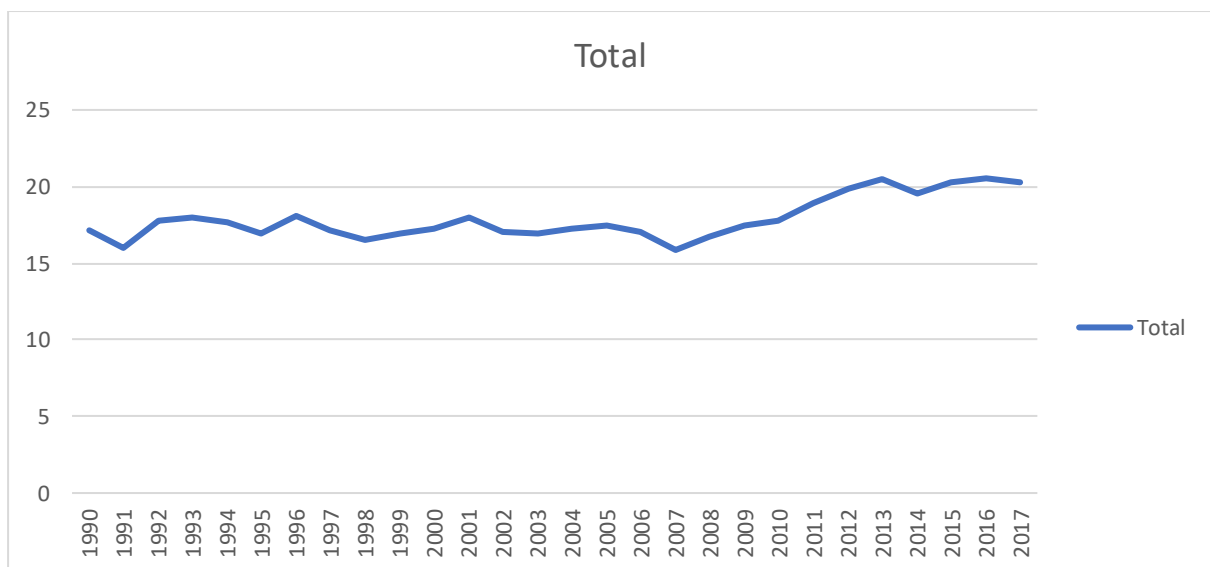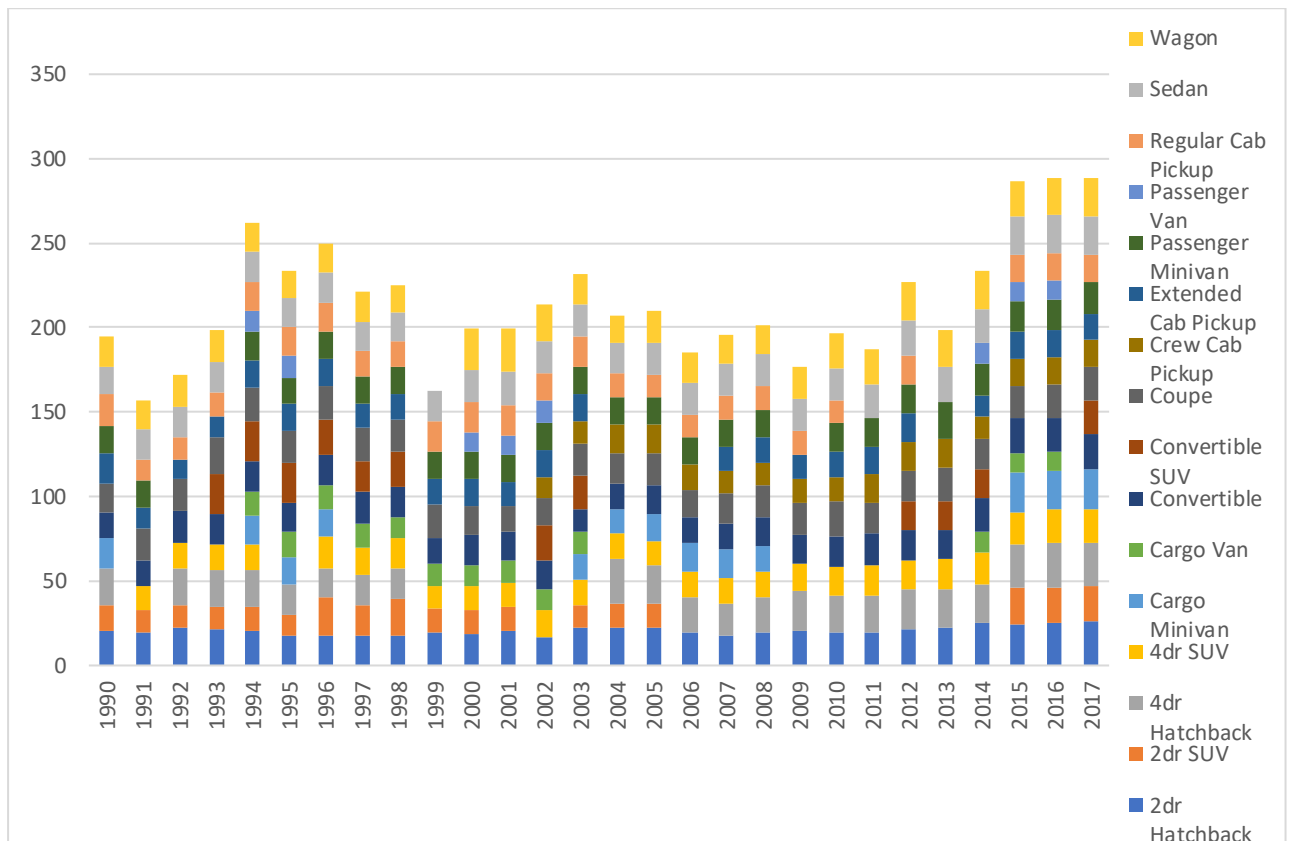
Task 5: How does the car's horsepower, MPG, and price vary across different Brands?

- **Hints:** Bubble chart to visualize the relationship between horsepower, MPG, and price across different car brands. Assign different colors to each brand and label the bubbles with the car model name. Calculate the average horsepower, MPG, and MSRP for each car brand using AVERAGEIFS or Pivot Tables.
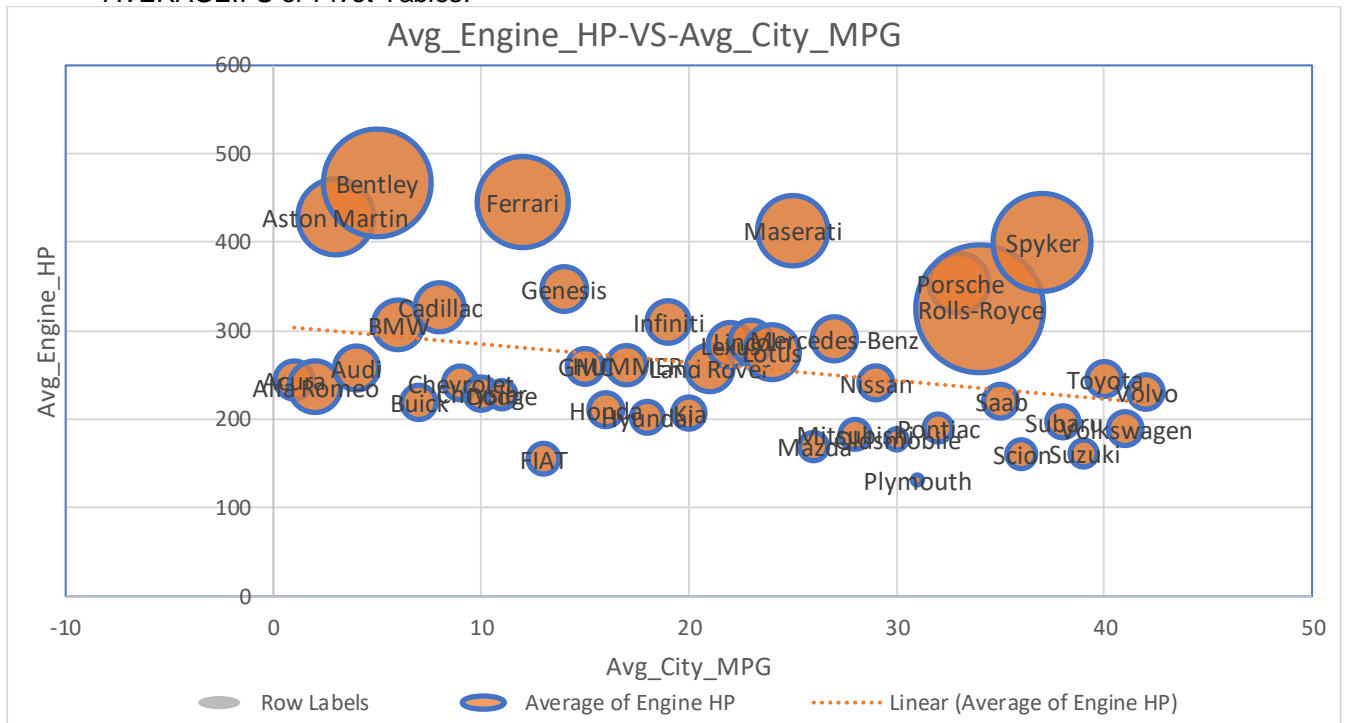
INSIGHTS:

Rolls-Royce is most expensive followed by Bentley and Spyker respectively, while Bentley has low Average City MPG with greatest Engine HP, Spyker has both high City MPG and Engine Hp.
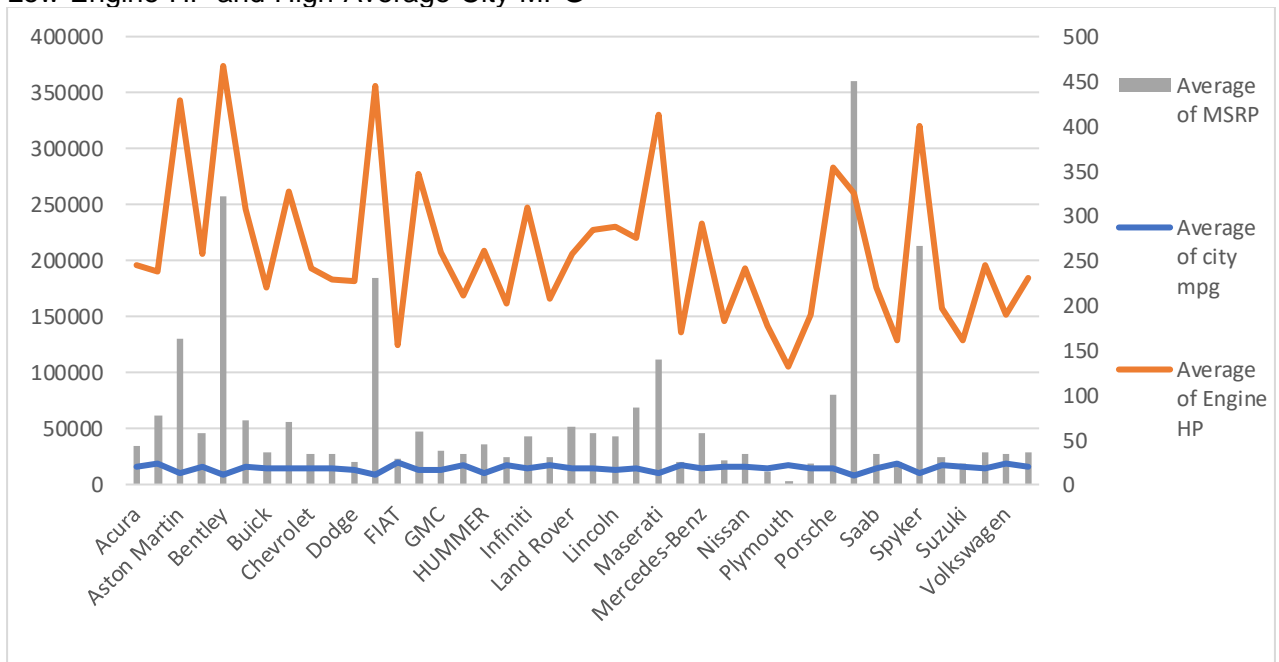
We can divide the chart into 4 equal parts and put them into categories like:

High Engine HP and Low Average City MPG

High Engine HP and High Average City MPG

Low Engine HP and Low Average City MPG

Low Engine HP and High Average City MPG

<u>CONCLUSION:</u>

1. Croassover is the most popular market category followed by exotic and luxury.
2. Also, crossover seems to have the most number of models.
3. With the increase of the HP, price also increases.

4. There was a huge rise of price after 2000, it is also observed that with increase of Engine's HP, the price has increased too.

5. Rolls-Royace is the most expensive manufacturer followed by Bentley, Spyker and Ferrari respectively.

6. Engine Hp, Engine Cylinders and Popularity seems to follow a positive trend line signifying these are the factors that are important in determining the car's prises.

7. Mostly, with the increase in the number of engine cylinders, the fuel efficiency also increases, however, in case of Tesla, even with significant low number of Engine Cylinders compared to others, the fuel efficieny is the highest.
8. Also, Bughati has the most number of average cylinders but still the efficiency is pretty low.

9. Most commonly regular unleaded fuel is used followed by premium unleaded and flex-fuel.
10. Convertible is the most expensive, while 2 dr Hatchback, 4Ha Hatchback, Sedan and Wagon seems most desirable, Since, it has lower price and better avg city and Highway MPG.

11. Chevrolet is the most expensive followed by Cadillac followed by Volkswagon followed by Toyota.
12. Mostly, large and compact are preferred over Midsize.

13. Bentley is the most pricey followed by Spyker and Maserati while least expensive ones are Plymouth followed by Genesis and Hummer, FIAT, Skion, Oldsmobile.

14. For Bentlay,Coupe, Convertible and Sedan, almost all three are equally popular.
15. While, in case of Plymouth, Convertible is more popular than Sedan.

16. Rolls-Royce is most expensive followed by Bentley and Spyker respectively, while Bentley has low Average City MPG with greatest Engine HP, Spyker has both high City MPG and Engine Hp. We can divide the chart into 4 equal parts and put them into categories like:
    - High Engine HP and Low Average City MPG
    - High Engine HP and High Average City MPG
    - Low Engine HP and Low Average City MPG
    - Low Engine HP and High Average City MPG

17. In order to increase the popularity, the above mentioned points should be considered and more focus on the product development using the above features.
18. Since, we don't seem to have enough data about the manufacturing cost of the cars and the features, profitability can't be analysed well.