A Project On

Analyzing the impact of Car features on Price and

Profitability

By

Swastik Kumar Pal

A student of

Jalpaiguri Government Engineering College

Submitted to

Trainity

PROJECT DESCRIPTION

## Problem Statement

1. 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.

2. 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. So it is very important to know which energy source will be profitable and will have more demand in the market for the manufacturer.

3. It is very important to know that which feature of the cars has more demand among the consumers. Without any proper insights about the driving factors of the consumer, the manufacturers can face losses in the market.

To find the solution of the problems and identifying the important feature for pricing and profitability, I will perform a data analysis from the provided data.
This analysis will help the manufacturer to decide which feature they need to choose for car price and more profitability. It will also help the teams to have better visualizations on every driven factors for impact on car's price and profitability.

# Data set Description

The data set contains information on various car models and their specifications, and is titled "Car Features and MSRP".

Here is a brief overview of the data set:

- **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.

## Goals of this analysis

### Analysis Part – 1

1. To know the the popularity of a car model vary across different market categories.

2. Relationship between a car's engine power and its price.

3. Which car features are most important in determining a car's price?

4. To know the average price of the car vary across different manufacturers.

5. To know the relationship between the fuel efficiency and the number of cylinders in cars engine.

### Analysis Part – 2

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

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

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

4. How does the fuel efficiency of cars vary across different body styles and model years?

5. How does the car's horsepower, MPG, and price vary across different Brands?

Approach to the Analysis

1. At first I will do some data overview of the provided data.

2. Then I will deal with the missing data and data cleaning to start my analysis.

3. After that, I will do an outlier treatment in my data if needed.

4. After preparing the data for the analysis, I will try to show the given tasks to the client by making an interactive dashboard which will help to have better visualization in analysis.

5. To perform the given tasks I used the pivot table feature of the excel and made a interactive dashboard in excel by using the slicers and filters.

6. I also performed market segment analysis and regression analysis in the given tasks as needed.

7. I tried to provide the charts for better visualization and analysis.

8. I will also describe the reasons why I came up with the solutions.

9. During the project, in **second part of task 3** it was not possible to plot the chart in scatter way. So I plotted the graph in scatter column for visualization.

USED  TECH  STACKS

I used **Microsoft office Excel and Tableau** to perform the given tasks.

Reasons behind using **Microsoft office Excel and Tableau**.

1. Calculation: Excel is also a powerful calculator. It can perform complex calculations, such as financial calculations or statistical analysis. It can also be used to create formulas and functions to automate repetitive calculations.

2. Data analysis: Excel also provides a range of functions and tools for analyzing data. These include charts and graphs, pivot tables, and conditional formatting. With these tools, you can quickly visualize and make sense of large amounts of data.

3. Reporting: Excel can be used to create professional-looking reports and presentations. You can format data, add graphics and charts, and create tables to summarize your findings. This can be useful for presenting data to colleagues, clients, or stakeholders
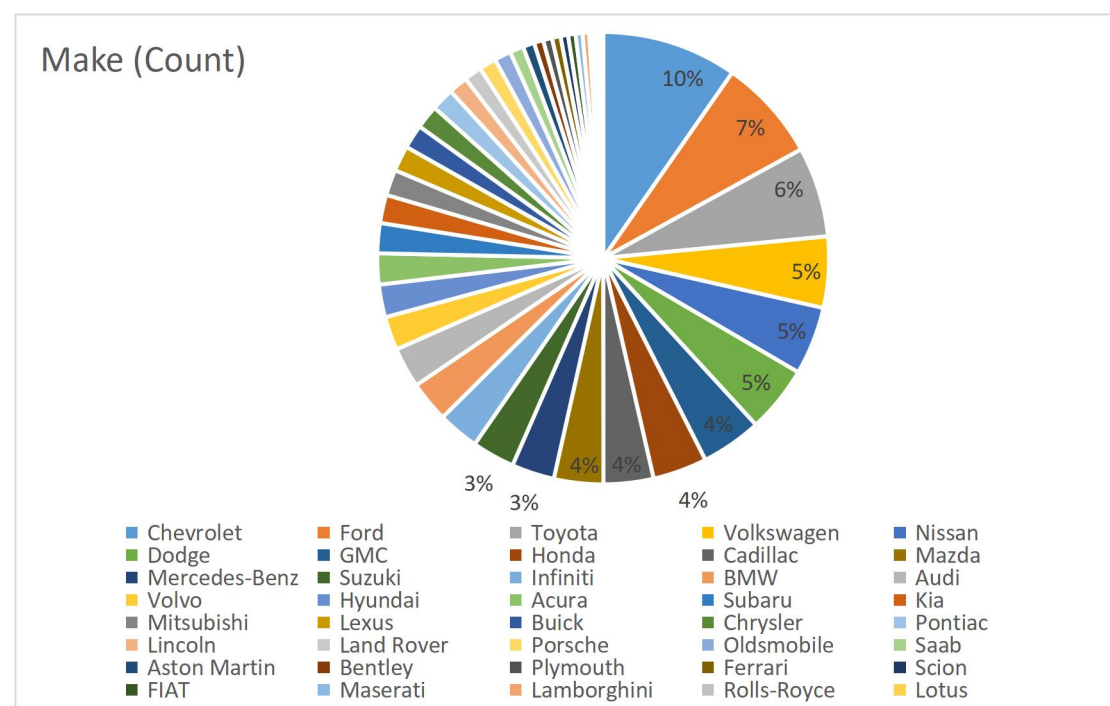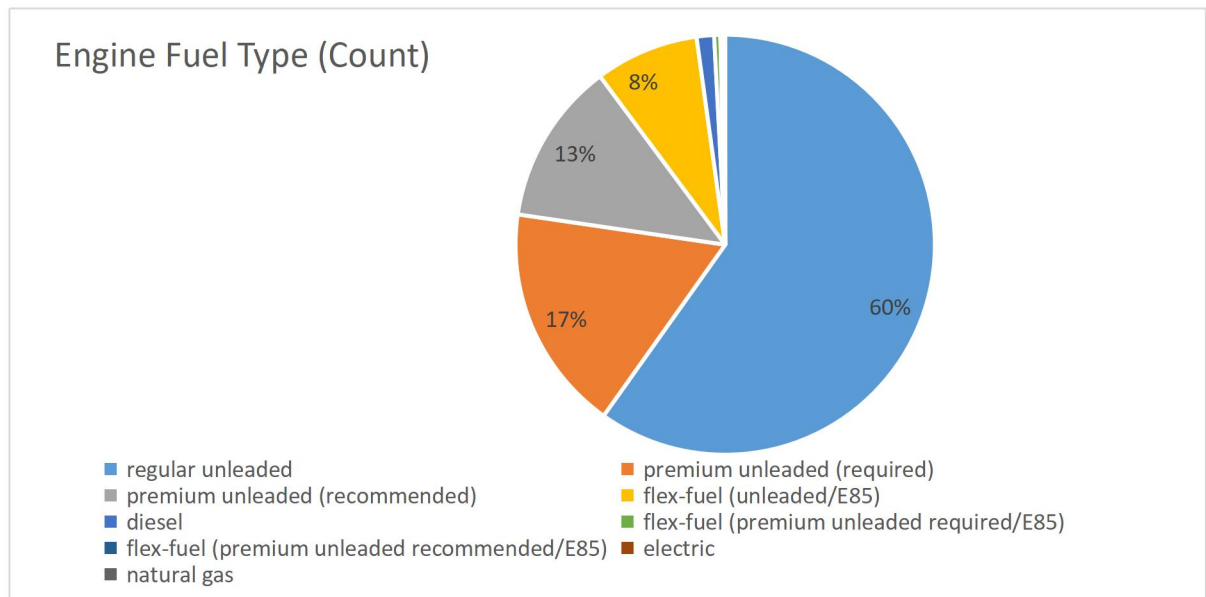
**Insights**

# Data Overview

To analyse the impact of car features on the price and profitability we will do some data overview at first like

1. Car Manufacturers.
2. Engine  Fuel type.
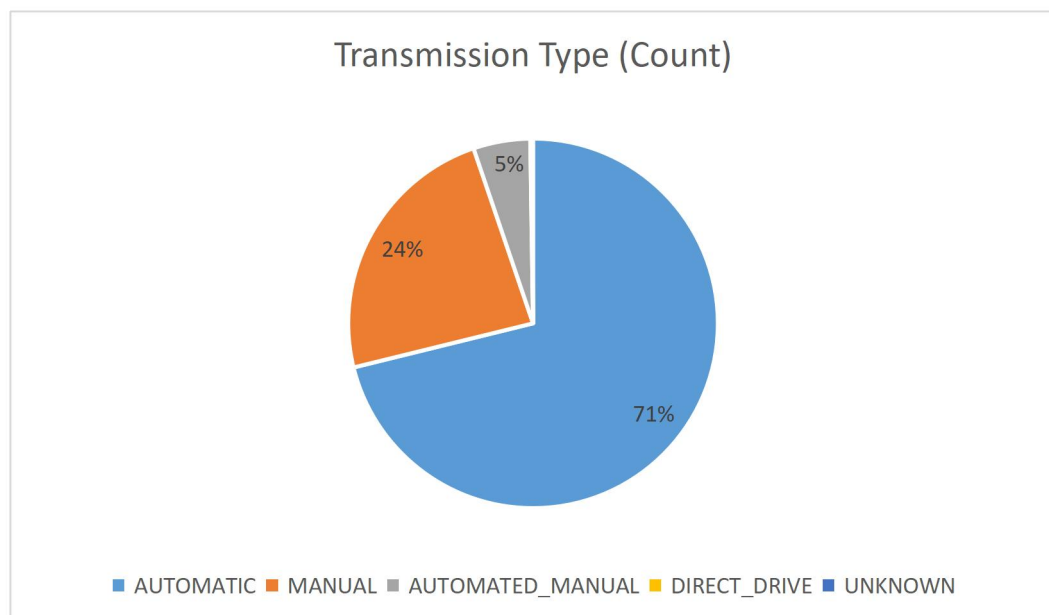3. Transmission Type.
4. Driven wheels
5. Vehichle size

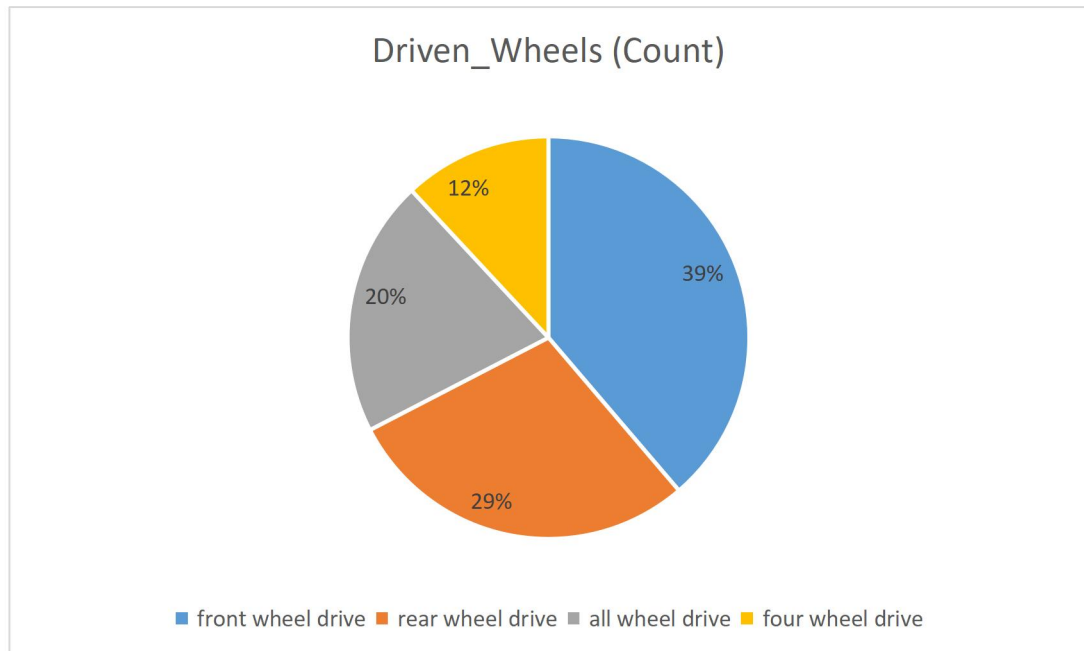So we will look into the **car manufacturers in the marke**t in the given data set



Make (Count)

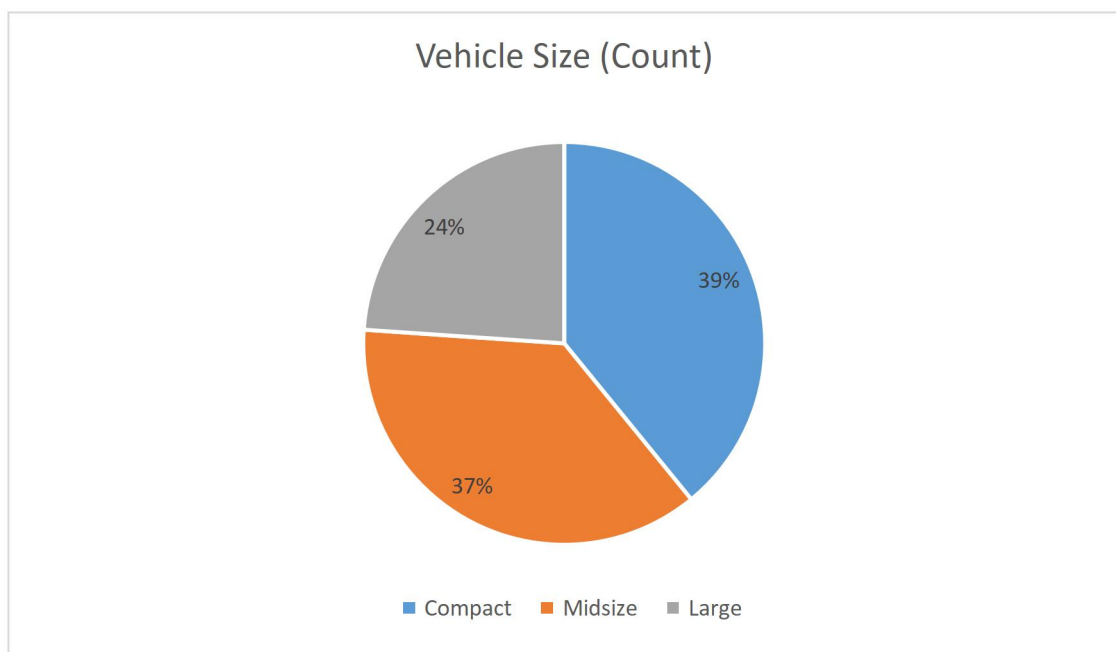## 2. Engine Fuel type

### Engine Fuel Type (Count)



- regular unleaded
- premium unleaded (required)
- premium unleaded (recommended)
- flex-fuel (unleaded/E85)
- diesel
- flex-fuel (premium unleaded required/E85)
- flex-fuel (premium unleaded recommended/E85)
- electric
- natural gas

## 3. Transmission Type

### Transmission Type (Count)



- AUTOMATIC
- MANUAL
- AUTOMATED_MANUAL
- DIRECT_DRIVE
- UNKNOWN

## 4. Driven Wheels



Driven_Wheels (Count)

39%

29%

20%

12%

front wheel drive    rear wheel drive    all wheel drive    four wheel drive

## 5. Vehicle Size



Vehicle Size (Count)

39%

37%

24%

Compact    Midsize    Large

## Data Cleaning

**Feature distribution:**

- Car Characteristics such as Engine hp , Engine cylinders, Transmission type, vehicle size are the important feature to decide the impact on the car's prices and the profitability.
- Some variables like year, Market category, Manufacturers, MSRP will help to judge the corresponding profitability and the car's prices on the market.

**Approach for the data cleaning :**

- So at first  I removed all the duplicate values present in the given data set.
- After that I will remove those columns which has blank values greater than the 40%.
- Then I will do the outlier treatment in the given data.

**In the given dataset there was no column which has blank values greater than 40%. So I did not remove any column from the data set.**

# Outlier Treatment

Now we need to identify the outliers for some feature columns which has numerical values.
For the outlier analysis the feature columns we will focus on

1. Engine HP
2. Engine Cylinders
3. Popularity
4. Highway MPG
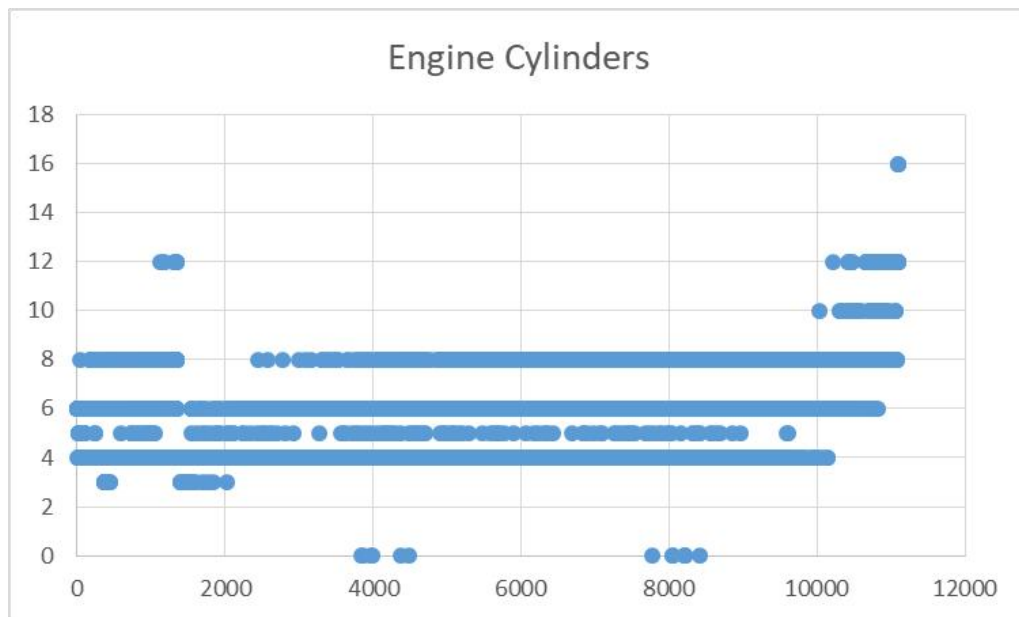5. MSRP

## Outlier analysis for Engine HP

I used a scattered chart plot for the visualization of the outliers.
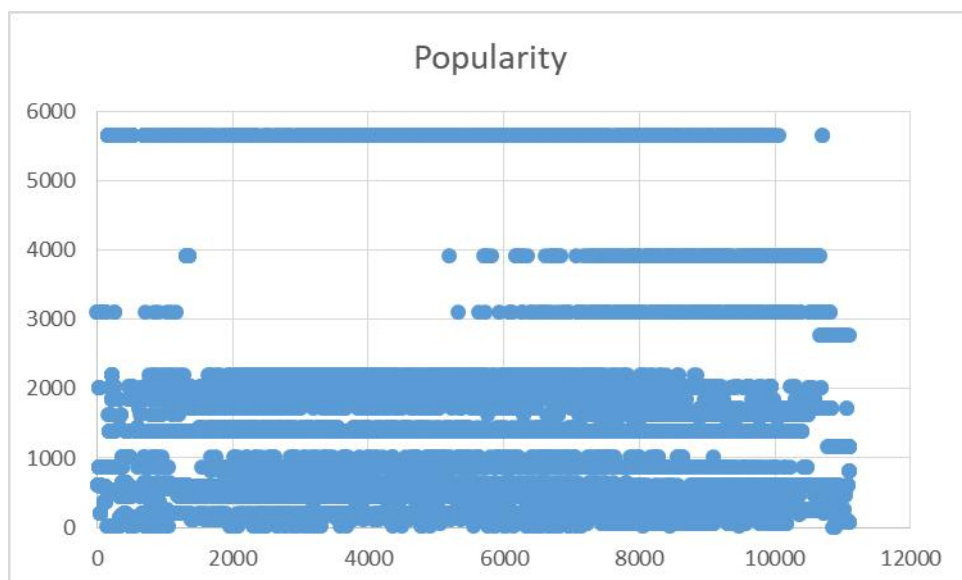To treat the outliers I used the upper limit technique.
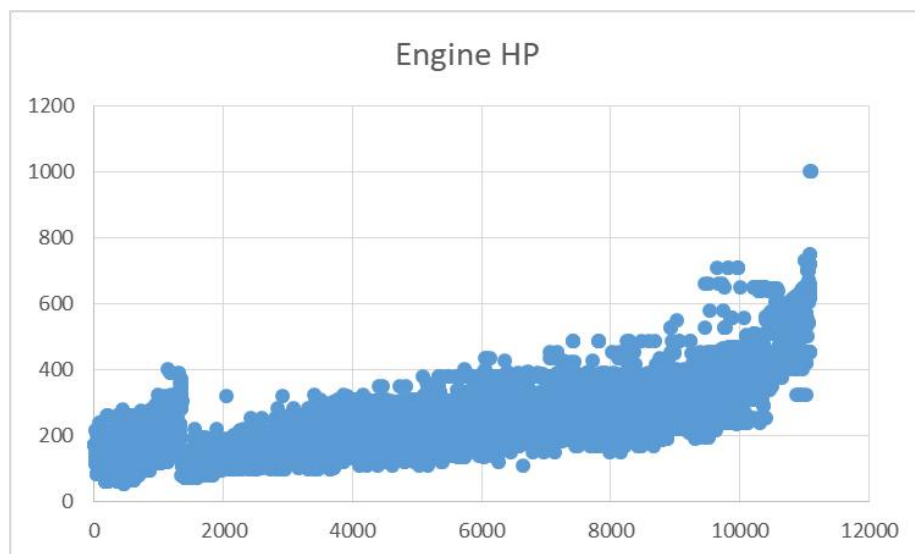


In this chart I did not consider any outlier because Engine Hp is gradually increasing , as a important feature of the vehicle.

Engine Cylinders

In this feature there are also no outlier is present to consider to remove from the data.

Popularity

In the Popularity segment there is no outlier is present.
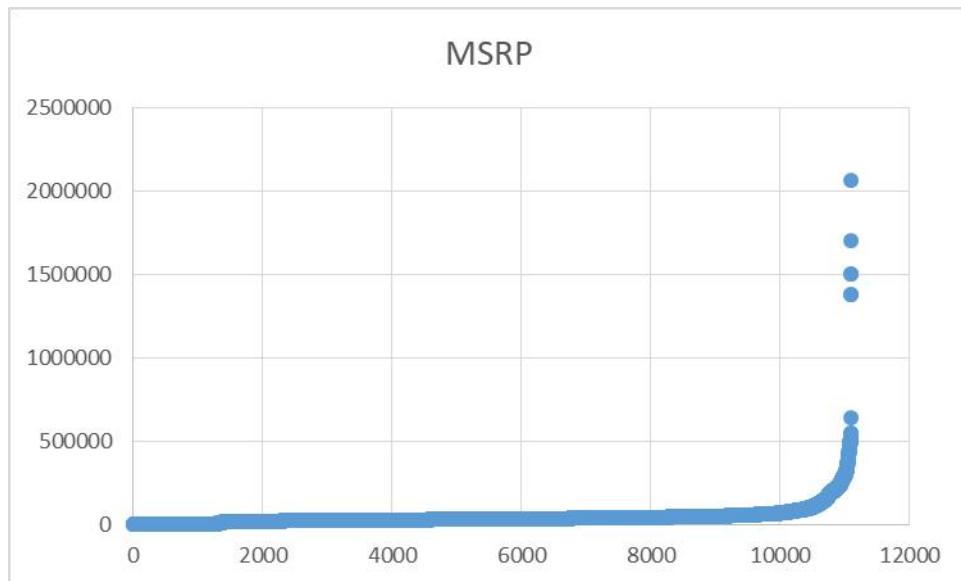
As we can see there is a pure increasing trend in the price. There are some values which are too much but as we know some car brands has more price in their luxury segment. So I did not remove this values from the data set for better analysis in every market category.

## Analysis Part- I

### Task-1

**How does the popularity of a car model vary across different market categories?**

- At first , I created the pivot table to do the analysis.

- I placed the models in rows and count of models & max popularity scores in the value section. After that I placed the market category in the filter section and also made a slicer with the Market categories.

Here for example I will do two analysis in different market categories.

At first , in **market category of Exotic, Factory tuner and high performance:**
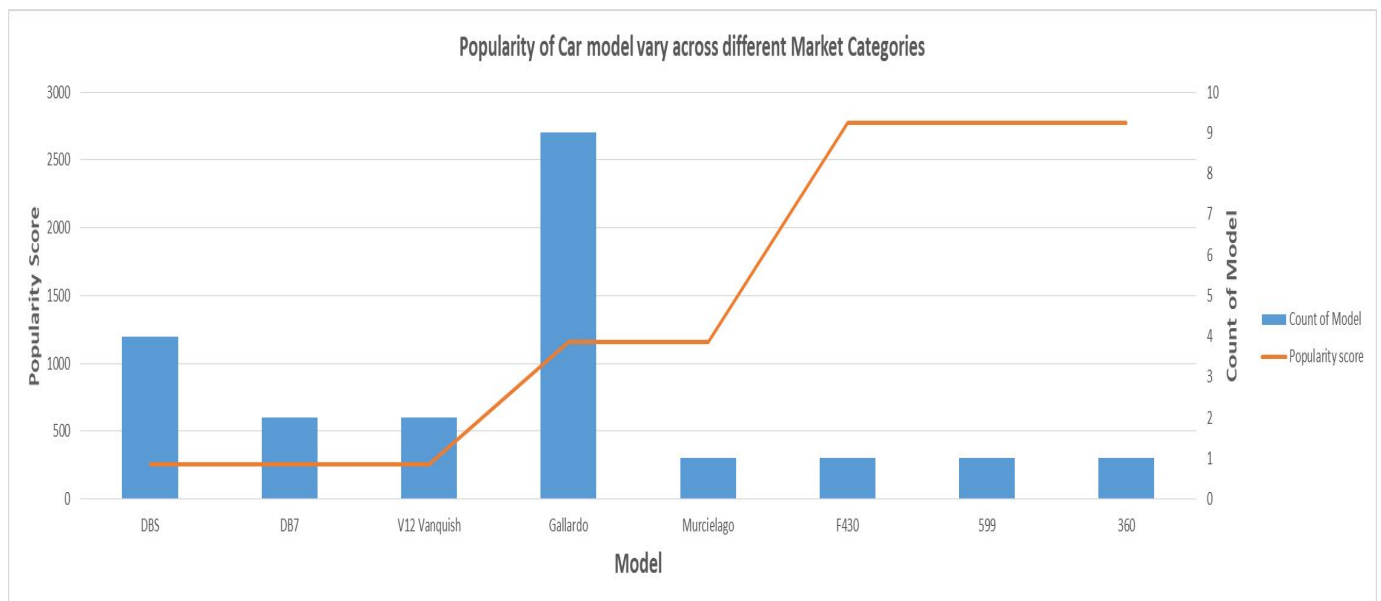
**Market Category -- Diesel, Luxury**



Popularity of Car model vary across different Market Categories
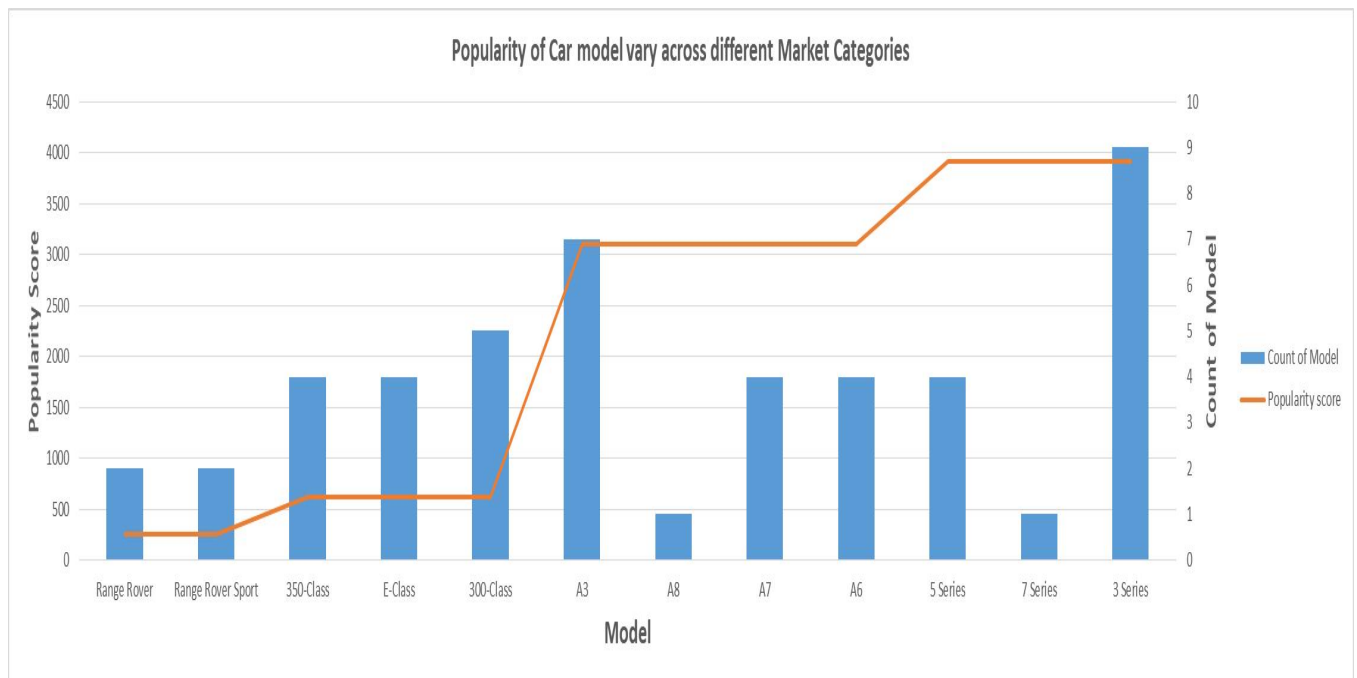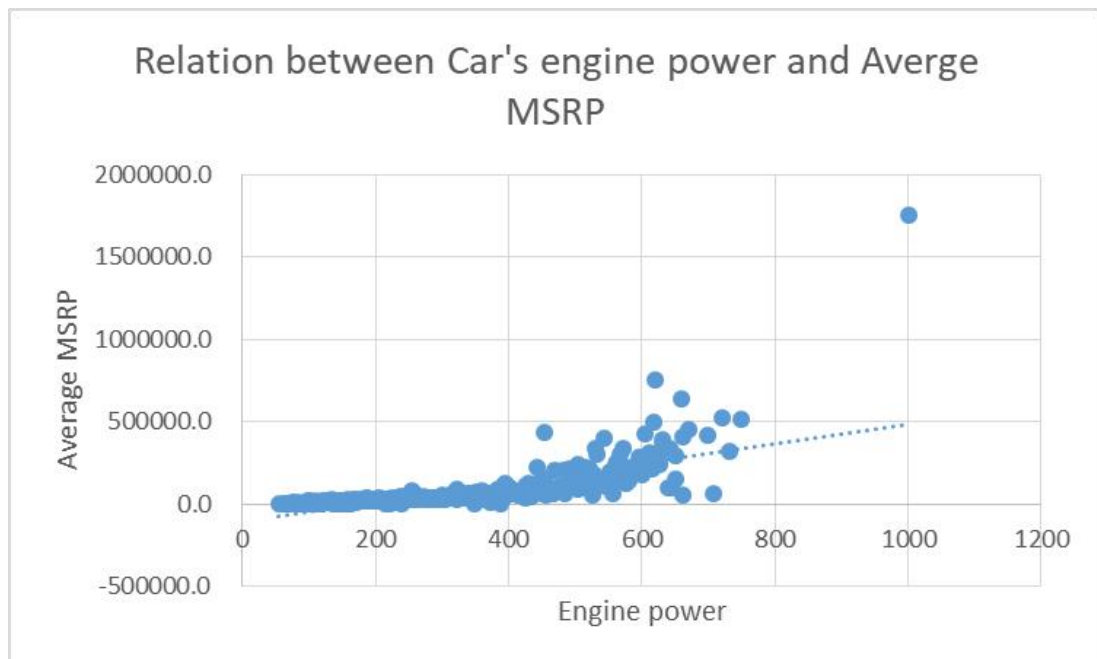
## Key Interpretation :

- So we can see from the upper analysis , In the market category of Exotic, Factory tuner and high performance , 360,599,F430 has more popularity score in this market segment.

- After that, 3 series,7 series has the more popularity score in the market category of Diesel , Luxury.

**What is the relationship between a car's engine power and its price?**

I showed the relation between car's engine power and average MSRP in scattered chart for the visualization



## Key Interpretation :

- From the trend line we can conclude that with increasing Engine power the average MSRP also increases.

## Task-3

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

I used regression analysis to find the solution of it. The features I considered for the correlation with the MSRP are --

1.  Number of cylinders
2.  Engine Hp
3.  Number of doors
4.  Highway MPG
5.  City MPG
6.  Popularity

The respective followings shows the relationship of features with the MSRP

| Features | Correlation |
|---|---|
| cylinder | 55% |
| Engine hp | 66% |
| doors | -15% |
| highway mpg | -21% |
| city mpg | -23% |
| popularity | -5% |

The below chart is the bar chart that shows the coefficient values for each variable to visualize their relative importance

### Relation of features with the MSRP
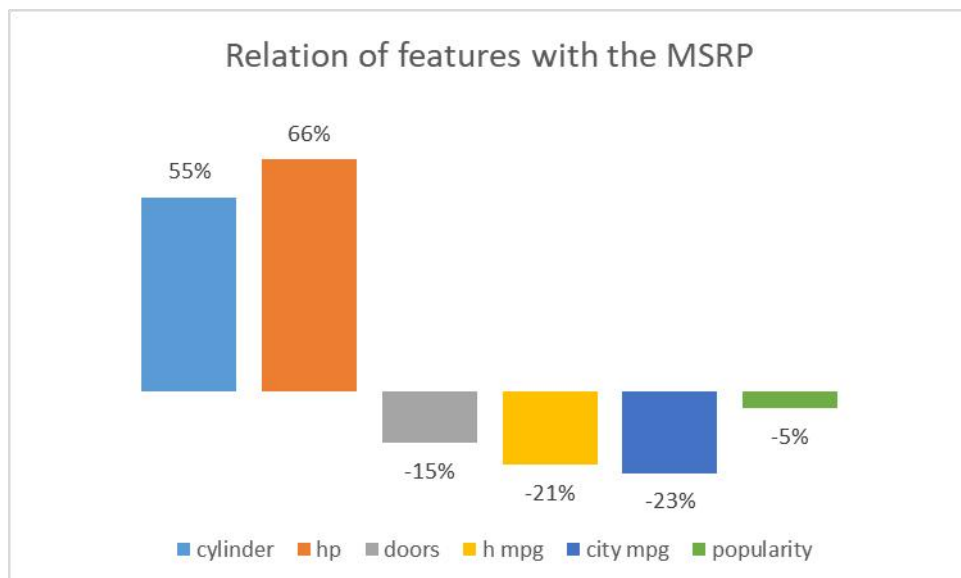


## Key Interpretation :

- From this visualizations and values we can conclude that Engine Hp has the strongest relationship with the Average MSRP.

**How does the average price of a car vary across different manufacturers?**

- I created a pivot chart where I placed the manufacturer in rows and average MSRP in values.

- After that I plotted it in horizontal stacked bar chart for better visualization and also added a manufacture slicer with it for desirable selection .
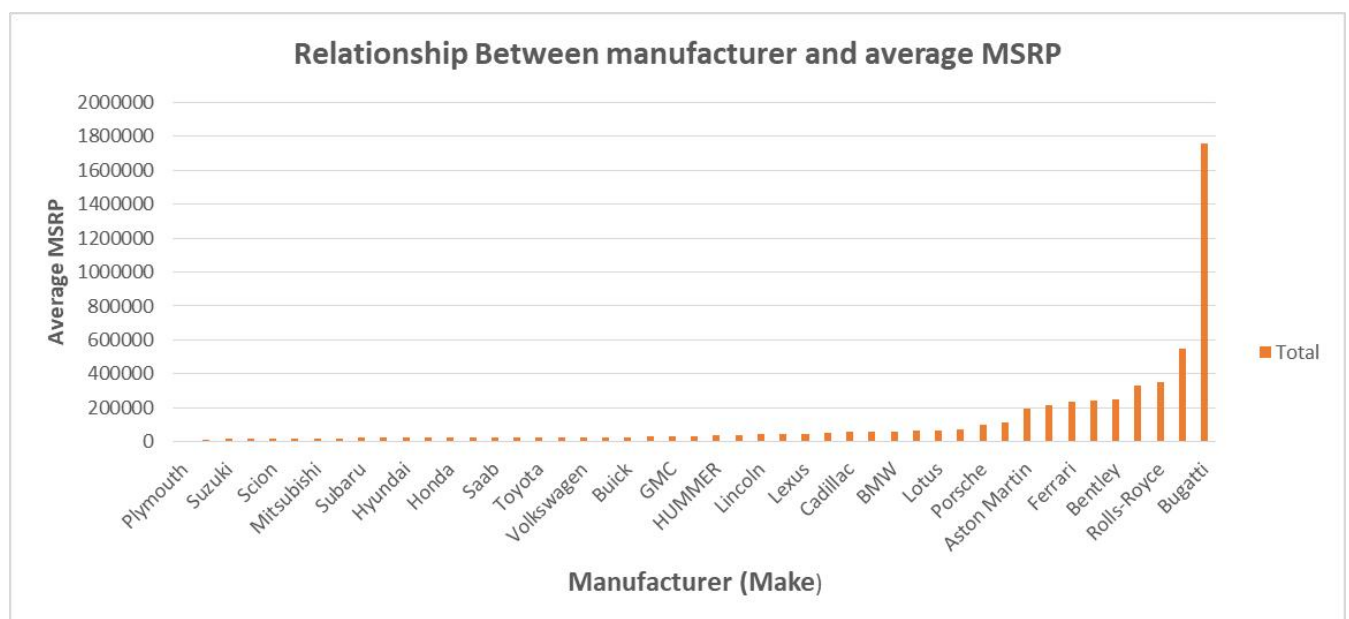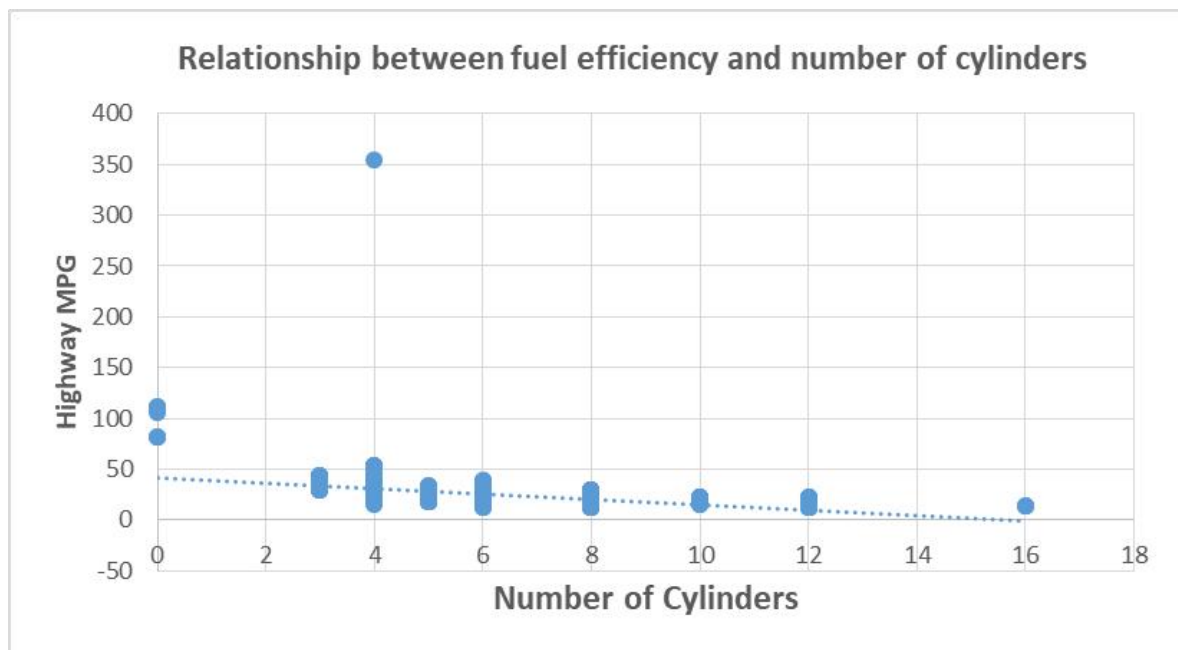


## Key Interpretation :

- If I consider all the manufactures then Bugatti has the highest average MSRP.Plymoutyh has the lowest average MSRP.

- In case, the client wants to compare between the different manufactures, it is also possible to see in the dashboard.

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

I created a scatter plot with the number of cylinders on the x-axis and highway MPG on the y-axis. Then created a trend line on the scatter plot to visually estimate the slope of the relationship and to assess its significance.



## Key Interpretation :

- From the trend line we can easily conclude that with the increasing number of cylinders in the cars engine the Highway MPG gradually decreases.

- From the scatter chart plotting we can conclude that four number of cylinders can provide relatively high highway MPG than the others.

- Correlation coefficient between the number of cylinders and highway MPG is **-61%**. So the relationship between them is very poor.
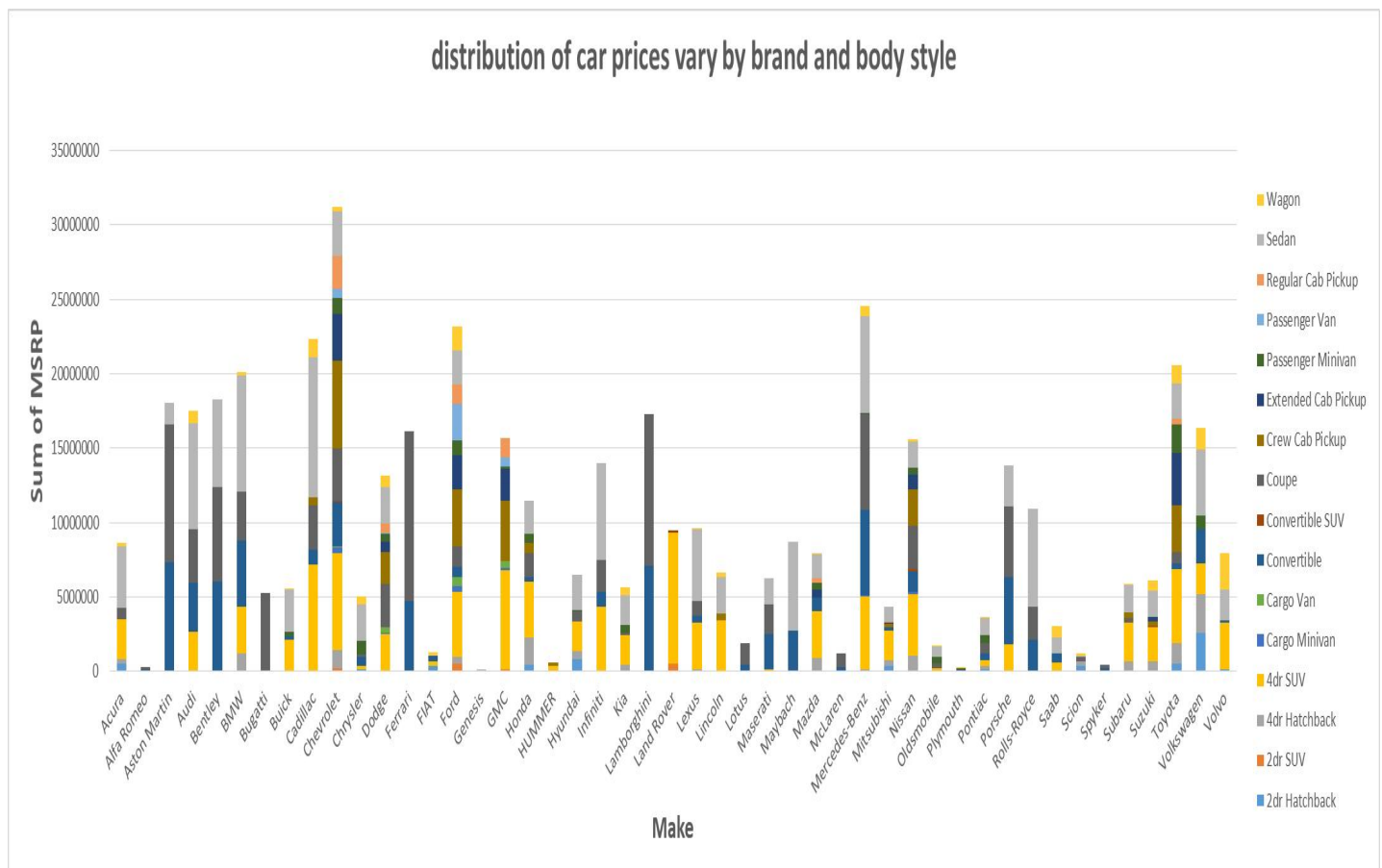
**How does the distribution of car prices vary by brand and body style?**

- I created a stacked column chart by using the pivot table feature
- In the pivot table, I placed **Make** in the rows and the **vehicle style** in the column and placed sum of the MSRP in the values.
- I also added the slicers of them for the flexibility in the analysis which will help the clients.

If I consider all the manufacturers and all the vehicle style then the chart will be



In the dashboard section, clients can see the distribution of car prices vary by brand and body style.

Task-2

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

- I created a clustered column chart by using the pivot table feature
- In the pivot table, I placed **Car Brand** in the rows and the **vehicle style** in the column and placed average of the MSRP in the values.
- I also added the slicers of them for the flexibility in the analysis which will help the clients.
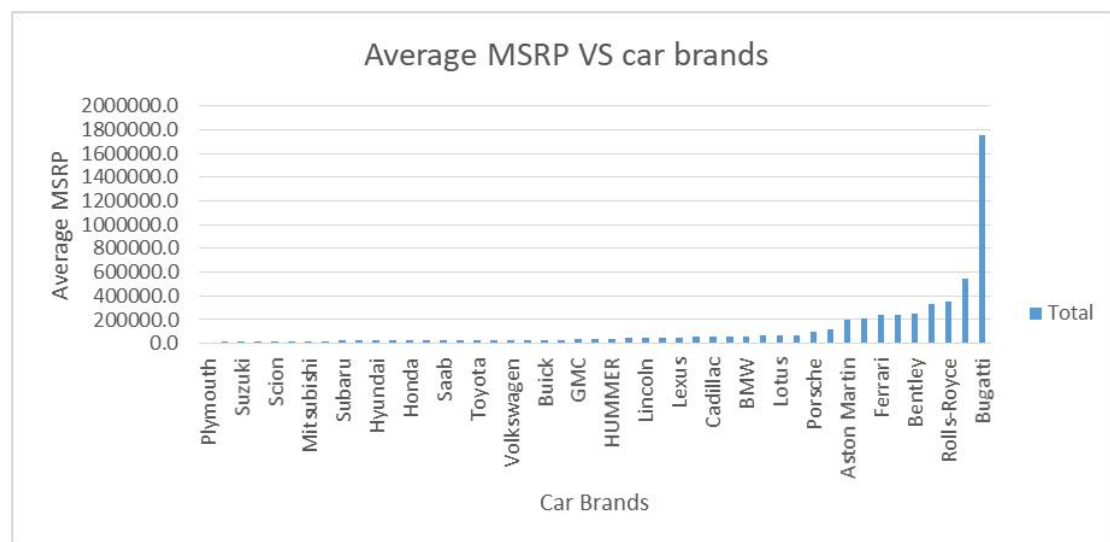


## Key Interpretation :

So we can conclude from the chart that **Bugatti** has the highest average MSRP And the **Plymouth** has the lowest average MSRP.

This Highest and lowest average MSRP can vary with the different body style . It is well represented in our dashboards and in our pivot table as **task2(2 ) sheet name.**

**Task-3**

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

- I created a stacked column chart by using the pivot table feature
- In the pivot table, I placed **Transmission type** in the rows and the **vehicle style** in the column and placed average of the MSRP in the values.
- I also added the slicers of them for the flexibility in the analysis which will help the clients.

Now we will consider two vehicle styles for the multivariate analysis.
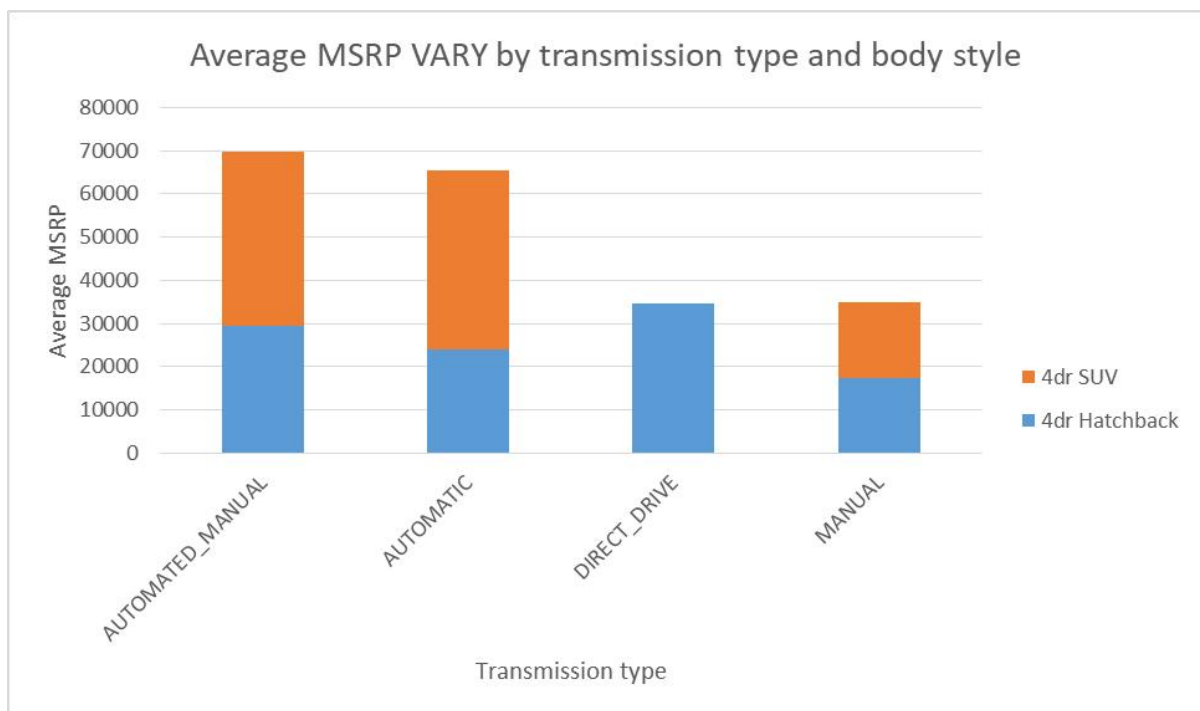


## Key Interpretation :

From the above chart we can say that cars with **vehicle style of 4dr SUV** with **automated_manual transmission type** has the highest average price.

Like this , We can do Lot more multivariate analysis in our dashboard and in the pivot table as needed.

**How does the fuel efficiency of cars vary across different body styles and model years?**

- I created a line chart by using the pivot table feature
- In the pivot table, I placed **vehicle style** in the rows and  placed average of the **highway MPG and city MPG** in the values.
- I also added the slicers of them for the flexibility in the analysis which will help the clients.

If we consider all the body styles and all the model years then the line chart will be
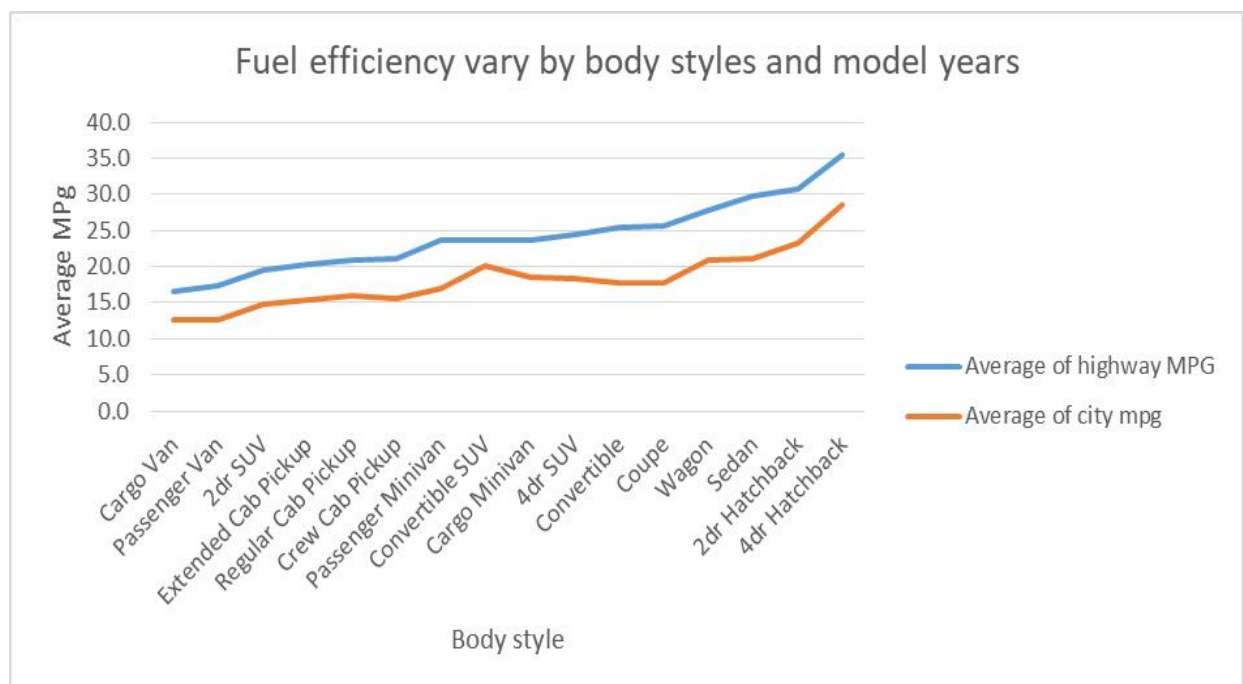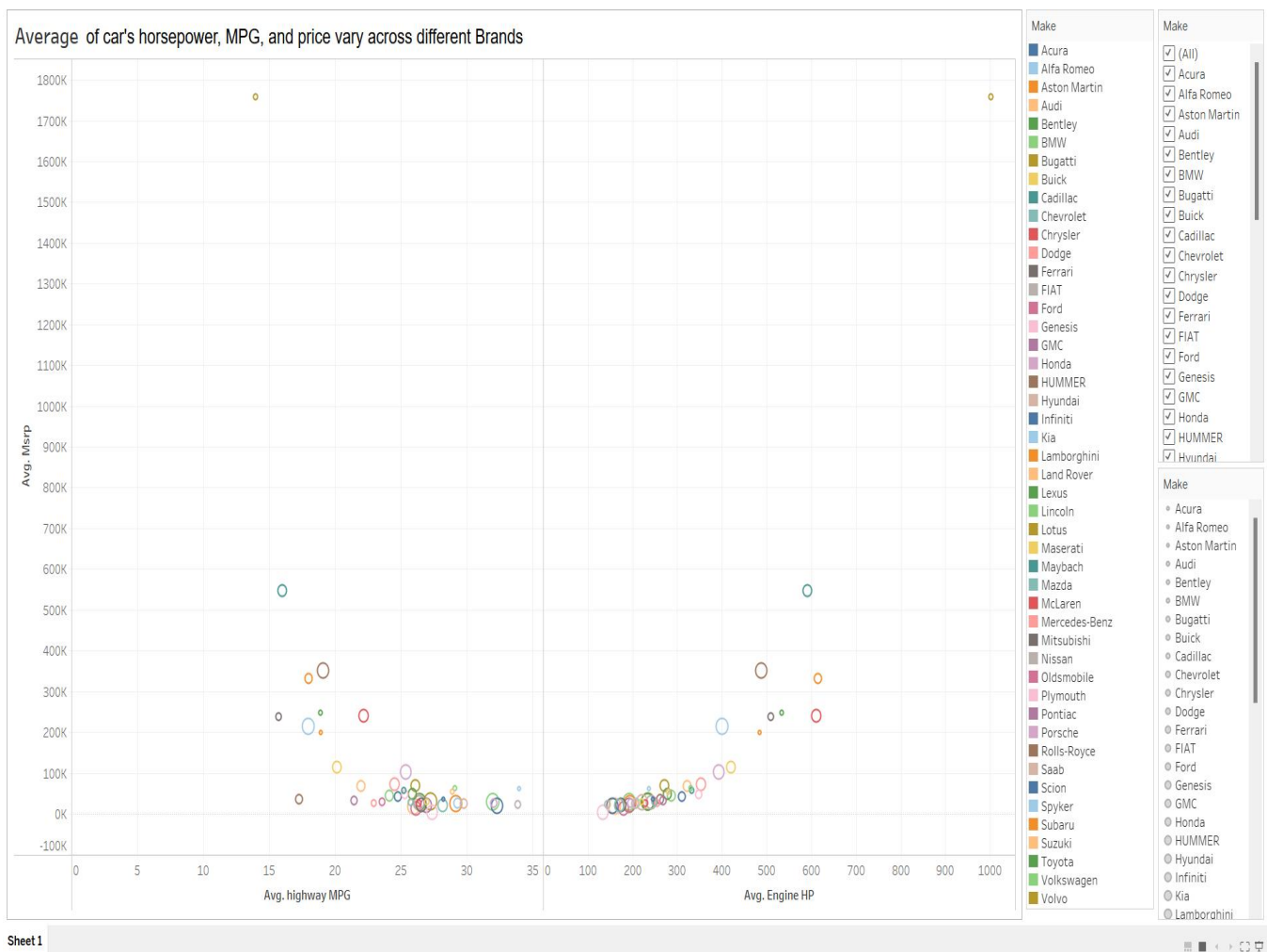


## Key Interpretation :

- From this above line chart we can conclude that 4dr Hatchback can provide the highest city and highway mpg in all model years.

Task-5

**How does the car's horsepower, MPG, and price vary across different Brands?**

To perform this task with better visualization I used the Tableau software for the presentation.
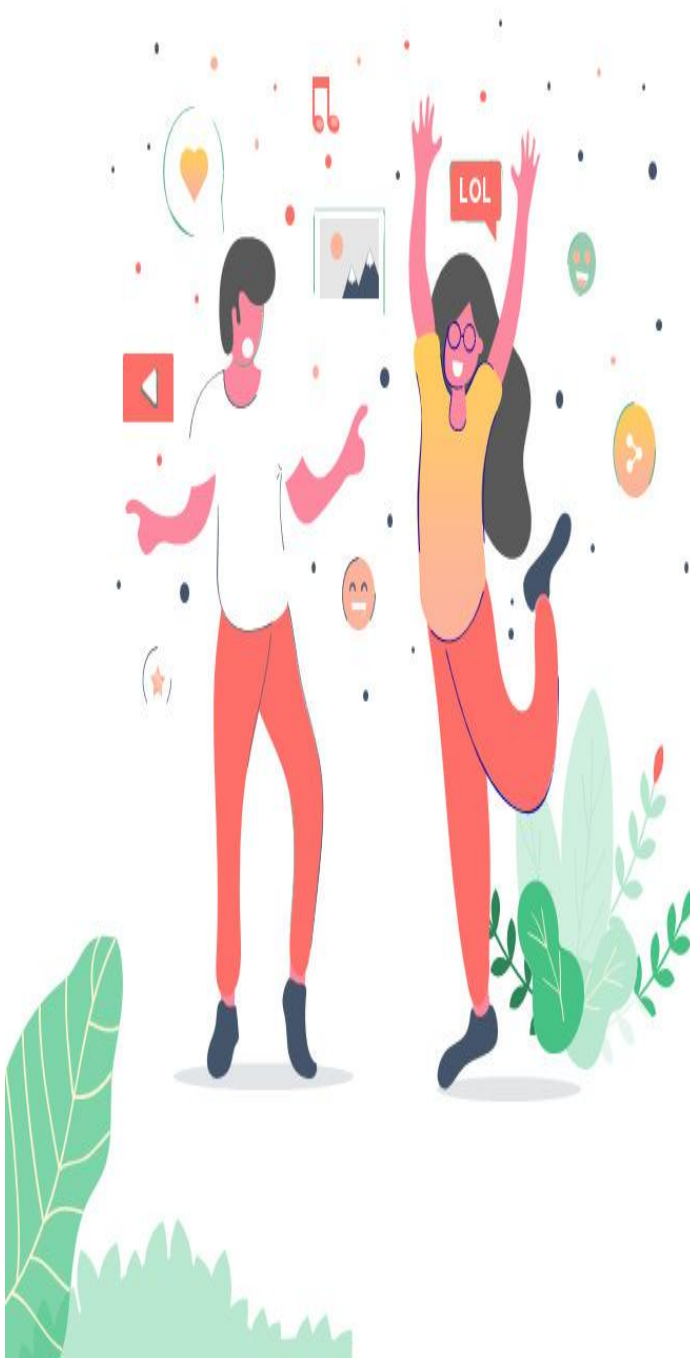
In the Tableau platform, I created a bubble chart for the visualization and also added filter to analyse different brands easily.



## Key Interpretation :

● For example if we consider the brand Toyota then
  Average MSRP =  28,759 Rs
  Average HIGHWAY MPG = 26.426
  Average Engine Power = 234.3 hp

So we can do the analysis like this for different brands also in the dashboard.

# Conclusion

This Analyzing impact of car features on prices and profitability project really helped me in different ways.

1. This project gave me more clarity about the actual industry requirements of a Excel to analyze the data and proper visualization of the large datasets.

2. To do the tasks I need to use some functions and charts/graphs to get the required insights. During that I came to know more about the advanced feature of the excel and more easy way to represent the data to the clients.

3 . It also introduced me with the more advanced features about the Excel I Used.

4.It helped me to have more clear concepts about the concepts and functions of the Excel.