

# Impact of Car Features

**Submitted by:**

**Nadendla Dharani**

**Hyper Link to the Excel Sheet:**

<https://docs.google.com/spreadsheets/d/1EkUggN86DEk6vFpkXON2NG9V-UZPbDhl/edit?usp=sharing&ouid=111691645789497796027&rtpof=true&sd=true>

## **Project Description**

This project aims to empower a car manufacturer to make informed decisions in the dynamic automotive industry by optimizing pricing and product development strategies. Focused on maximizing profitability while meeting consumer demand, the project utilizes a comprehensive dataset encompassing car features, market categories, pricing, and sales data. Through meticulous data cleaning and preprocessing, including handling missing values and feature engineering, the team ensures the dataset's reliability. The approach combines analytical methods such as exploratory data analysis, regression analysis, and market segmentation to uncover key insights into consumer preferences and profitability drivers. The project's success lies in providing actionable recommendations based on a nuanced understanding of the relationships between car features, market dynamics, and pricing, ultimately guiding the manufacturer towards enhanced competitiveness in the market.

## **Approach**

The approach for this project revolves around leveraging Microsoft Excel as the primary tool for comprehensive data analysis, pricing optimization, and product development decision-making in the automotive industry. Starting with data cleaning and preprocessing within Excel, we ensure the dataset's integrity by handling missing values and performing necessary transformations. Descriptive statistics and visualization techniques are employed to gain insights into data distributions, correlations, and trends. The core analytical method involves using Excel for regression analysis, effectively modeling the relationships between car features, market categories, and

pricing. This enables us to uncover key insights into consumer preferences and profitability drivers. While the analytical methods are confined to Excel, the focus is on a meticulous and robust exploration of the dataset to provide actionable recommendations. The simplicity and accessibility of Excel make it an ideal platform for the project, allowing us to deliver valuable insights that will inform pricing strategies and guide future product development efforts for the car manufacturer.

### **Tech-Stack Used**

The project leverages Microsoft Excel as the primary tool for data analysis and visualization. Excel's versatile functions and features facilitate data cleaning, computation of summary statistics, and the creation of various visualizations, including pie charts, bar graphs, and histograms. The familiar interface of Excel allows for efficient handling of tasks such as handling missing values, clubbing categories, and outlier detection. Additionally, standard statistical functions within Excel contribute to deriving meaningful insights, making it a powerful and accessible tech-stack for this data analytics project.

### **Insights**

- Flex fuel and Crossover are popular market categories.
- From the chart, we can conclude that Engine HP is directly related to car prices.
- Engine cylinder is the most important feature in determining a car's price.
- Maybach, Rolls Royce and Lamborghini have the highest average prices.
- Obtained moderate negative correlation between the number of cylinders and fuel efficiency.
- Brands like Chevrolet, Mercedes-Benz and Ford have higher car prices.
- Brands like Maybach, Rolls Royce and Lamborghini have higher average car prices.
- Cars with AUTOMATED\_MANUAL and DIRECT\_DRIVE transmission types tend to have higher prices, and Vehicle Styles like Convertible and Coupe have higher prices.

- Fuel efficiency is increasing over time.
- As engine HP increases, fuel efficiency decreases, and car prices increase.

## Results

### Data Cleaning

- There are blank cells in the Engine Fuel Type Column.

1	Make	Model	Year	Engine Fuel Type	Engine HP	Engine Cylinders	Transmission Type	Driven_Wheels	Number of Doors
11323	Suzuki	Verona	2004		155	6	AUTOMATIC	front wheel drive	4
11324	Suzuki	Verona	2004		155	6	AUTOMATIC	front wheel drive	4
11325	Suzuki	Verona	2004		155	6	AUTOMATIC	front wheel drive	4

Engine Fuel Type	Count of Engine Fuel Type
diesel	154
electric	66
flex-fuel (premium unleaded recommended/E85)	26
flex-fuel (premium unleaded required/E85)	54
flex-fuel (unleaded/E85)	899
flex-fuel (unleaded/natural gas)	6
natural gas	2
premium unleaded (recommended)	1523
premium unleaded (required)	2006
regular unleaded	7175
<b>Grand Total</b>	<b>11911</b>

So the Mode of Fuel Engine Type=regular unleaded.

So I replaced all the blank cells with 'regular unleaded'.

- Blanks cells in the Engine HP column.

1	Make	Model	Year	Engine Fuel Type	Engine HP	Engine Cylinders	Transmission Type	Driven_Wheels	Number of Doors	Market Category
4205	Ford	Escape	2017	regular unleaded			4 AUTOMATIC	front wheel drive	4	Crossover
4206	Ford	Escape	2017	regular unleaded			4 AUTOMATIC	all wheel drive	4	Crossover
4207	Ford	Escape	2017	regular unleaded			4 AUTOMATIC	all wheel drive	4	Crossover
4208	Ford	Escape	2017	regular unleaded			4 AUTOMATIC	front wheel drive	4	Crossover
4916	Ford	Freestar	2005	regular unleaded			6 AUTOMATIC	front wheel drive	4	N/A
4917	Ford	Freestar	2005	regular unleaded			6 AUTOMATIC	front wheel drive	4	N/A
4918	Ford	Freestar	2005	regular unleaded			6 AUTOMATIC	front wheel drive	4	N/A
4919	Ford	Freestar	2005	regular unleaded			6 AUTOMATIC	front wheel drive	4	N/A
4920	Ford	Freestar	2005	regular unleaded			6 AUTOMATIC	front wheel drive	4	N/A
4921	Ford	Freestar	2005	regular unleaded			6 AUTOMATIC	front wheel drive	4	N/A
6910	Lincoln	MKZ	2017	regular unleaded			4 AUTOMATIC	front wheel drive	4	Luxury,Hybrid
6912	Lincoln	MKZ	2017	regular unleaded			4 AUTOMATIC	front wheel drive	4	Luxury,Hybrid
6918	Lincoln	MKZ	2017	regular unleaded			4 AUTOMATIC	front wheel drive	4	Luxury,Hybrid
6920	Lincoln	MKZ	2017	regular unleaded			4 AUTOMATIC	front wheel drive	4	Luxury,Hybrid

Sum of Engine HP	Average of Engine HP
2953978	249

Replacing the blank cells with the average of Engine HP.

- The Engine Cylinders column has blank cells.

1	Make	Model	Year	Engine Fuel Type	Engine HP	Engine Cylinders
8697	Mazda	RX-7	1993	regular unleaded	255	
8698	Mazda	RX-7	1994	regular unleaded	255	
8699	Mazda	RX-7	1995	regular unleaded	255	
1016						

Engine Cylinders	Sum of Engine Cylinders
0	0
3	90
4	19008
5	1125
6	27114
8	16248
10	680
12	2760
<b>Grand Total</b>	<b>67025</b>

Replacing all the blank cells with the Mode of Engine Cylinders, which is 6.

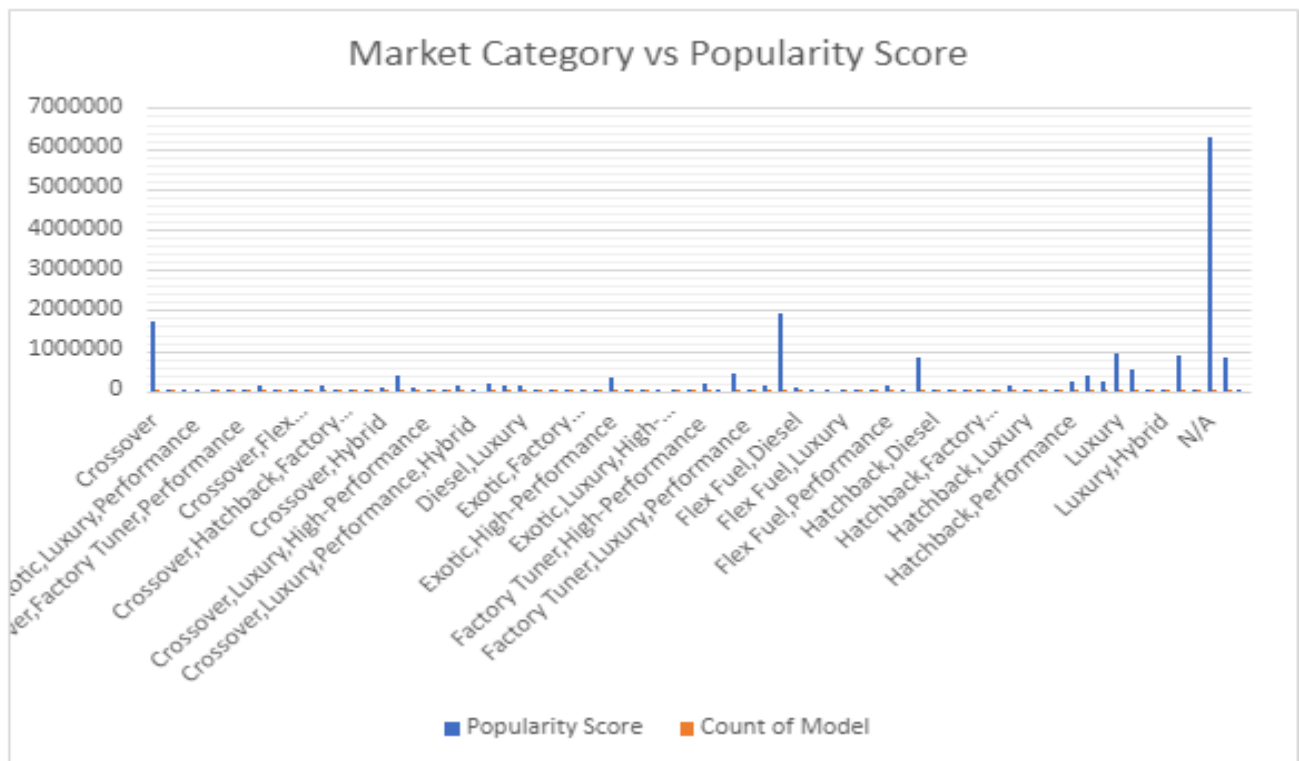
- The Number of Doors column has blank cells. Replacing them with Mode=4

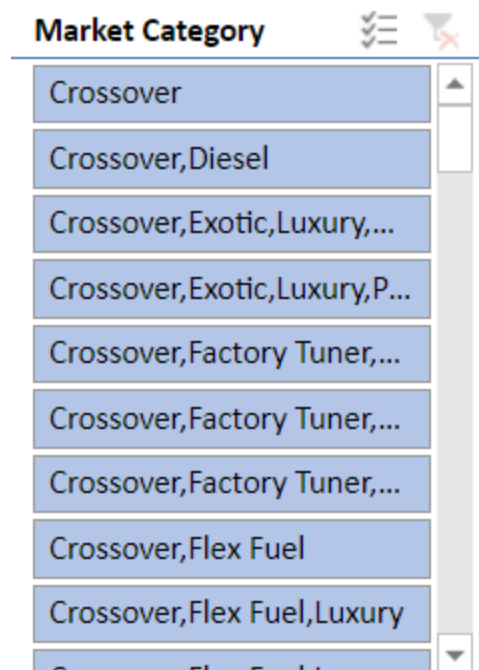
Number of Doors	Count of Number of Doors
2	3157
3	395
4	8359
Grand Total	11911

- There are outliers in the Engine HP column with values greater than 17000000. So I deleted the rows containing those values.

**Insight Required:** 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.
- **Task 1.B:** Create a combo chart that visualizes the relationship between market category and popularity.

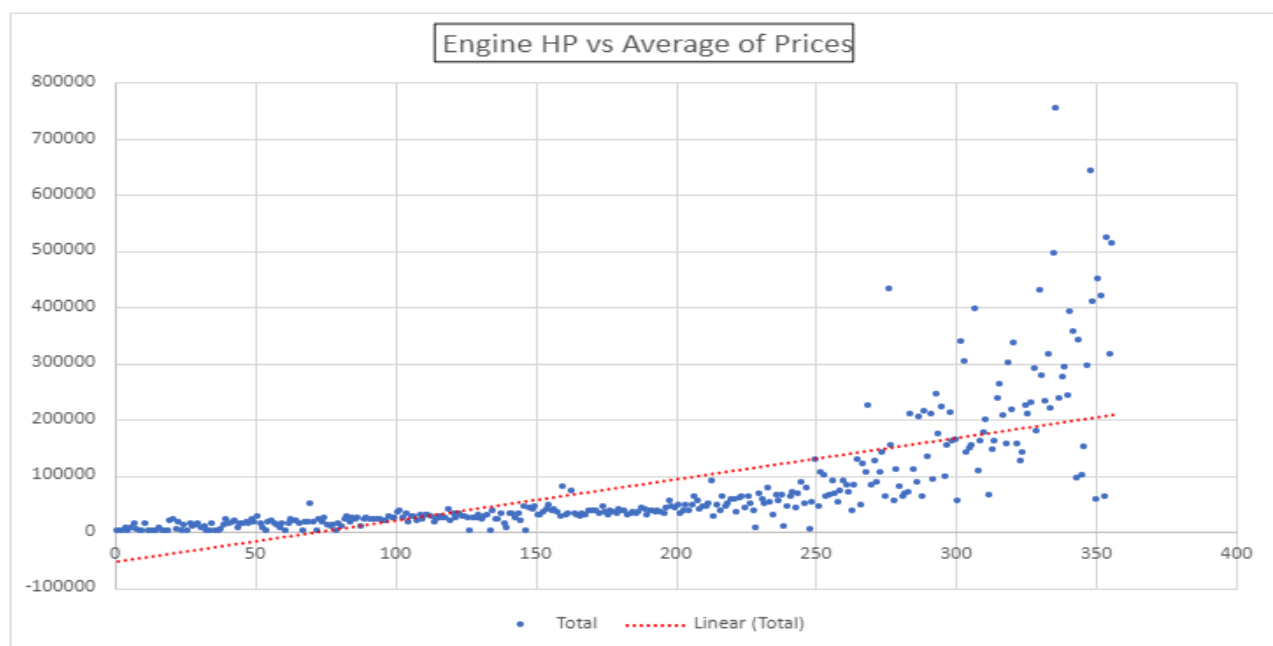




I am using MS-Office Web Version. So creating Combo charts in the web version is not possible. So I simply created Bar charts, but I provided slicers for the Pivot table, so it will still be understandable to the readers.

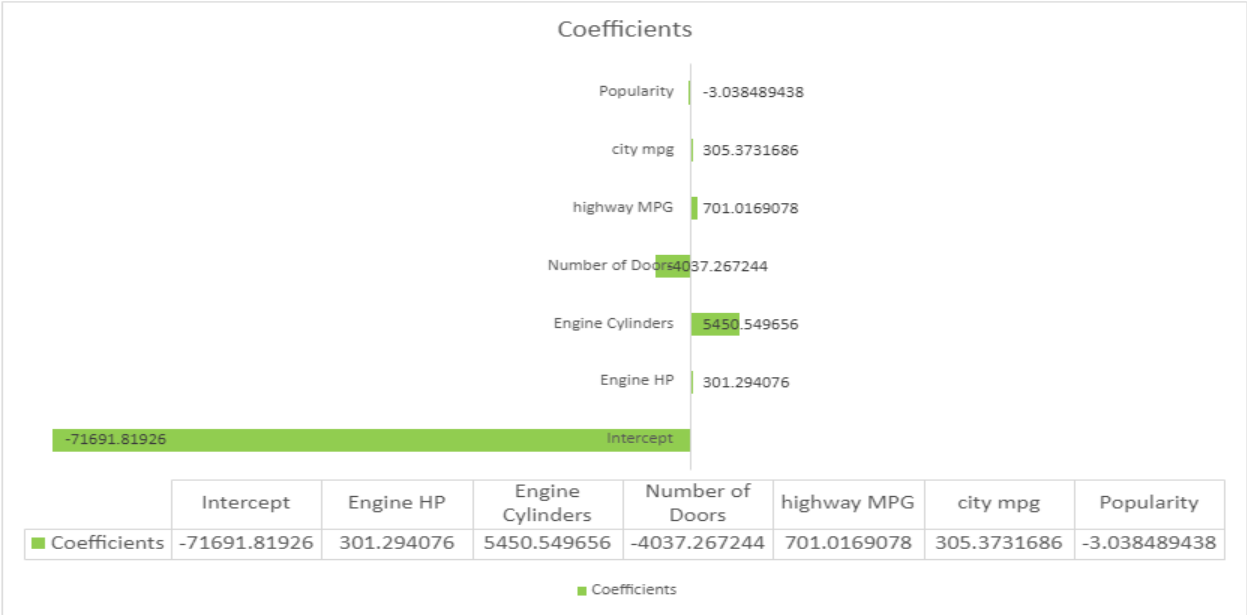
**Insight Required:** What is the relationship between a car's engine power and its price?

- **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.



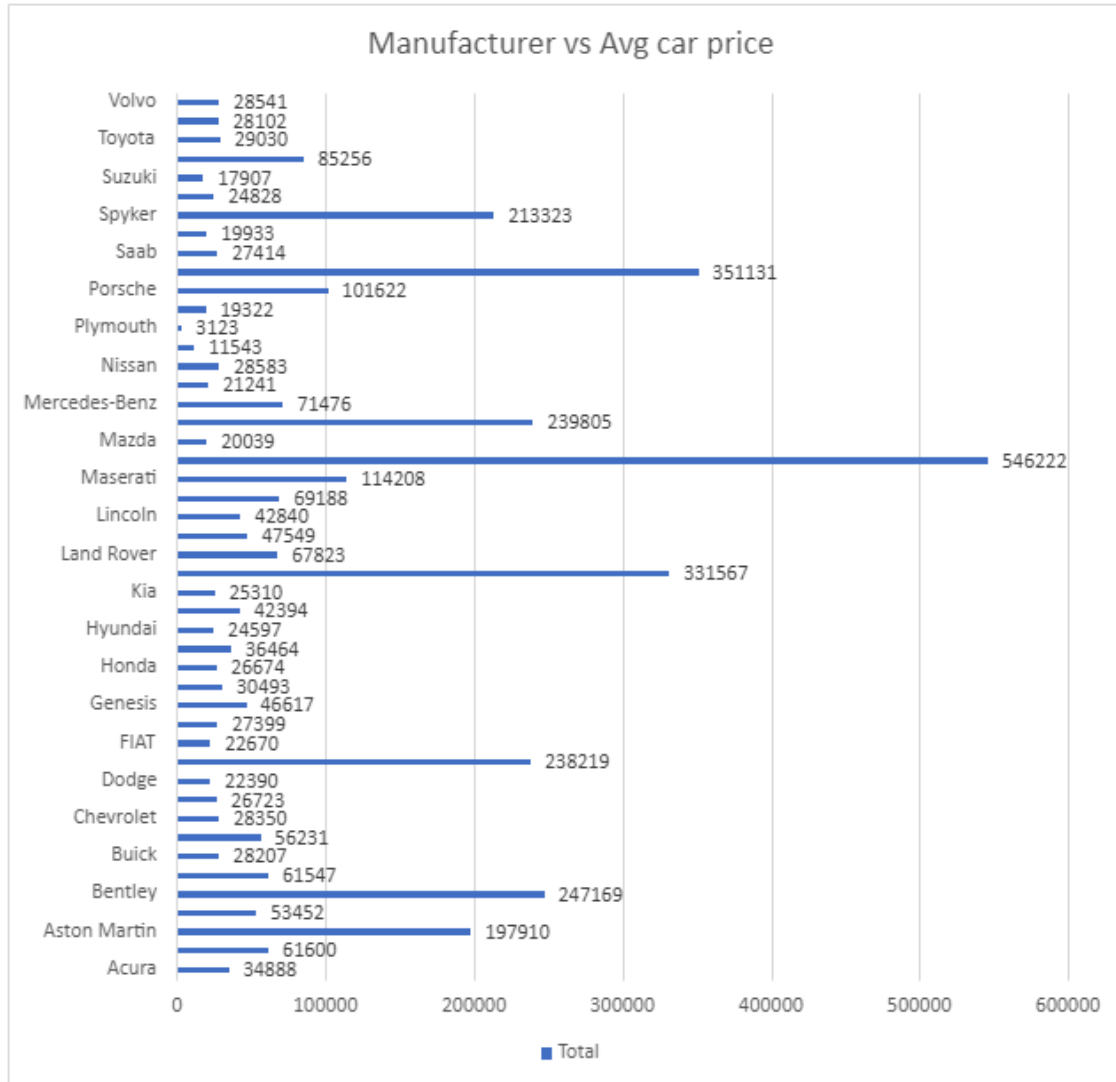
**Insight Required:** Which car features are most important in determining a car's price?

- **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?

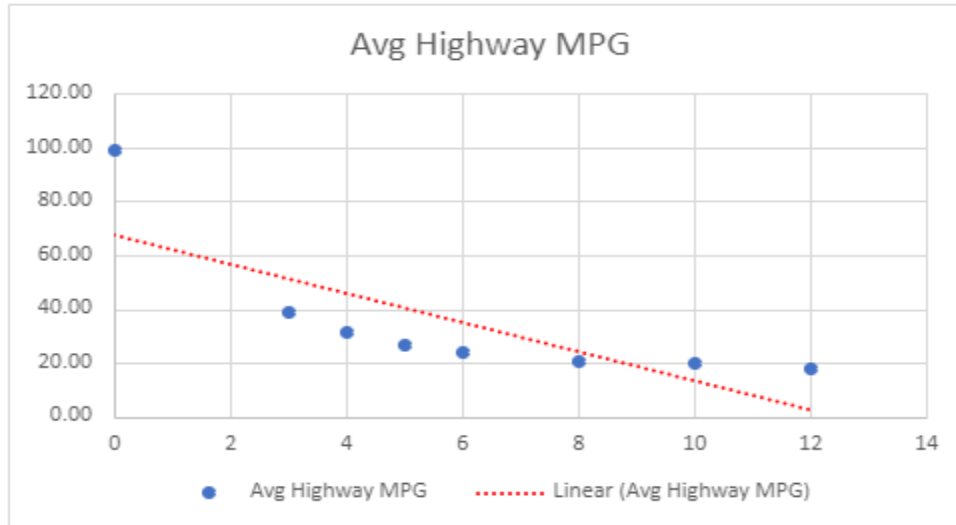
- **Task 4.A:** Create a pivot table that shows the average price of cars for each manufacturer.
- **Task 4.B:** Create a bar chart or a horizontal stacked bar chart that visualizes the relationship between manufacturer and average price.



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

- **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.





### Correlation

-0.782324016

### Building the Dashboard:

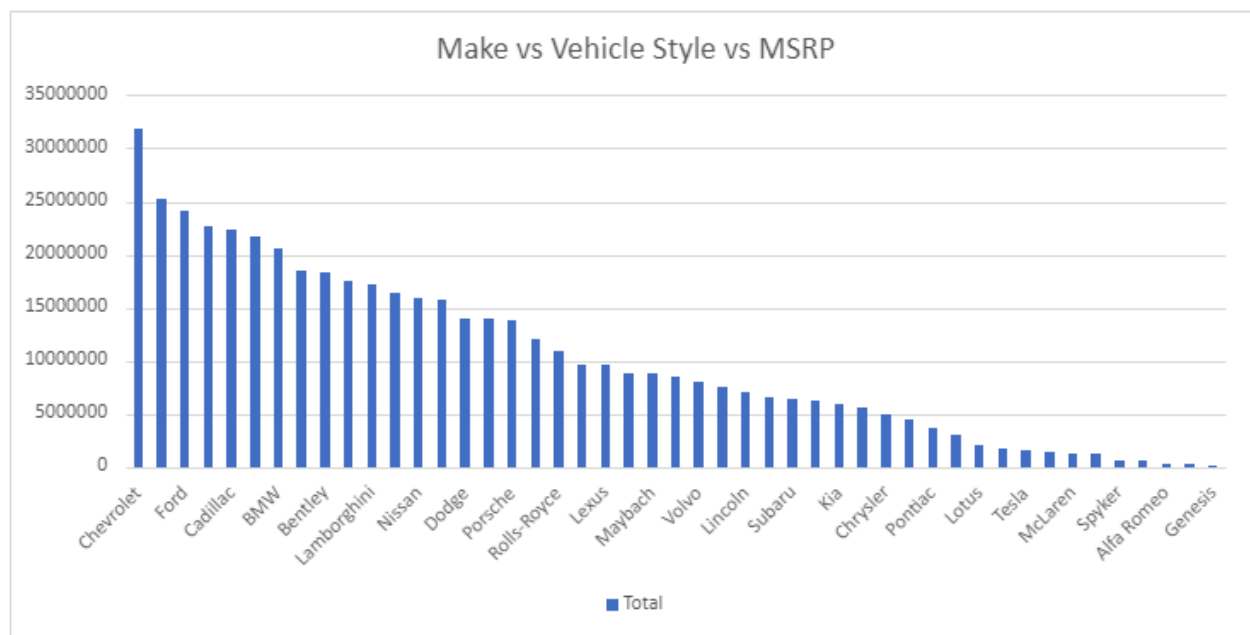
Now for the Next portion of the Project, you need to create the Interactive Dashboard. Use filters and slicers to make the chart interactive. The client has requested these questions given below:

**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.

I am using the MS-Excel web version, and there are limitations to using filters for Pivot Charts. So I created the stacked column as usual and used slicers for the pivot table instead.

Make	Vehicle Style
Acura	2dr Hatchback
Alfa Romeo	2dr SUV
Aston Martin	4dr Hatchback
Audi	4dr SUV
Bentley	Cargo Minivan
BMW	Cargo Van
Buick	Convertible
Cadillac	Convertible SUV
Chevrolet	Coupe



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

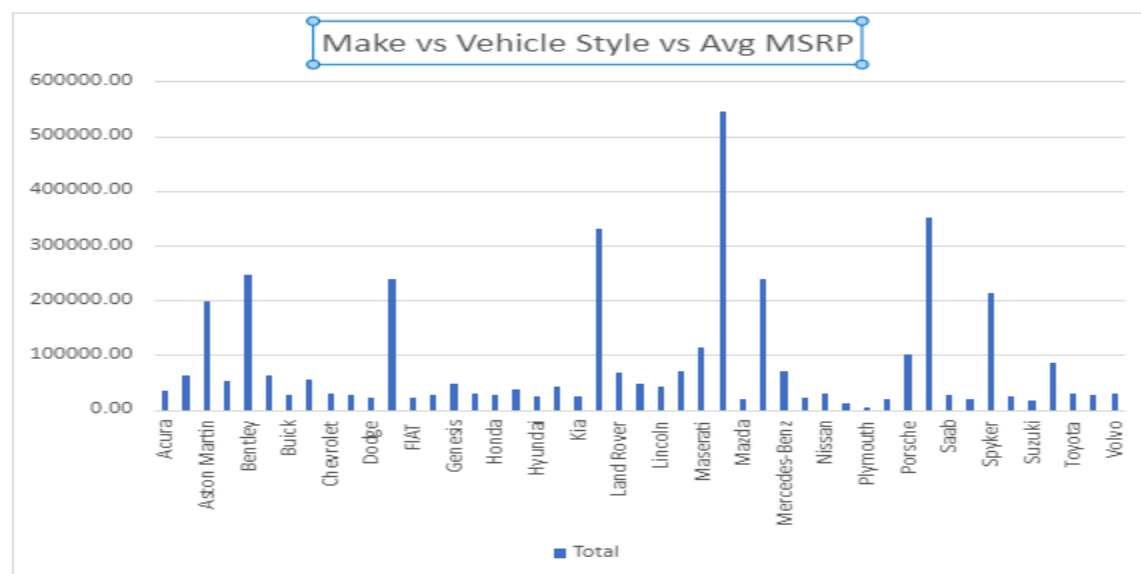
- **Hints:** Clustered column chart to compare the average MSRPs across different car brands and body styles. Calculate the average MSRP for each brand and body style using AVERAGEIF or Pivot Tables.

**Make**

- Acura
- Alfa Romeo
- Aston Martin
- Audi
- Bentley
- BMW
- Buick
- Cadillac
- Chevrolet

**Vehicle Style**

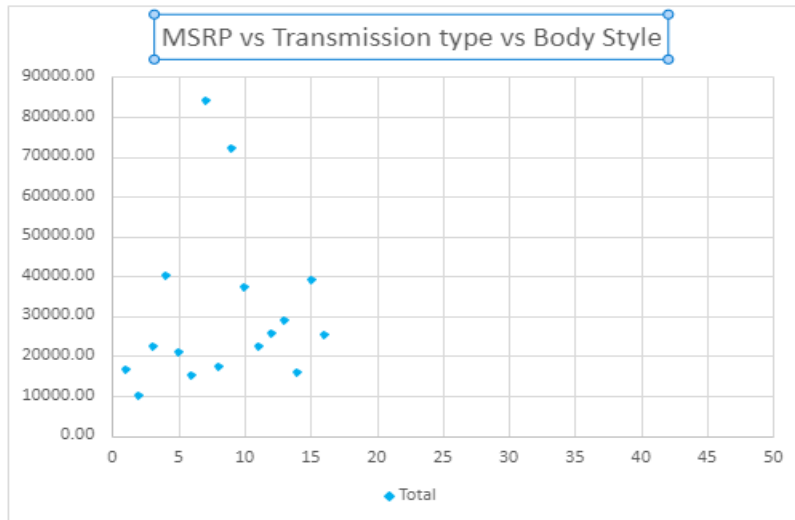
- 2dr Hatchback
- 2dr SUV
- 4dr Hatchback
- 4dr SUV
- Cargo Minivan
- Cargo Van
- Convertible
- Convertible SUV
- Coupe



**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.

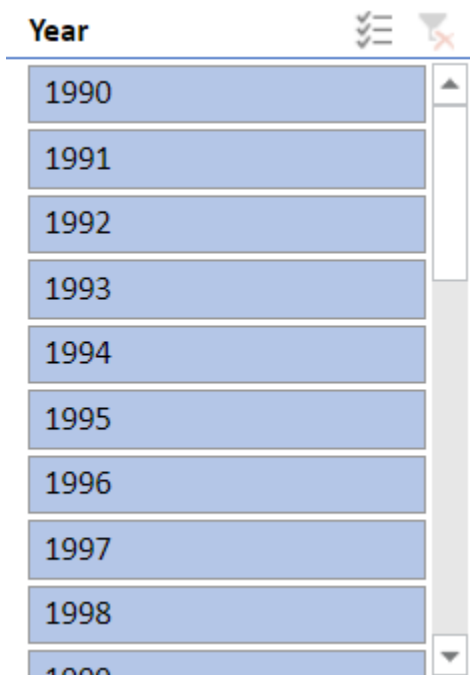
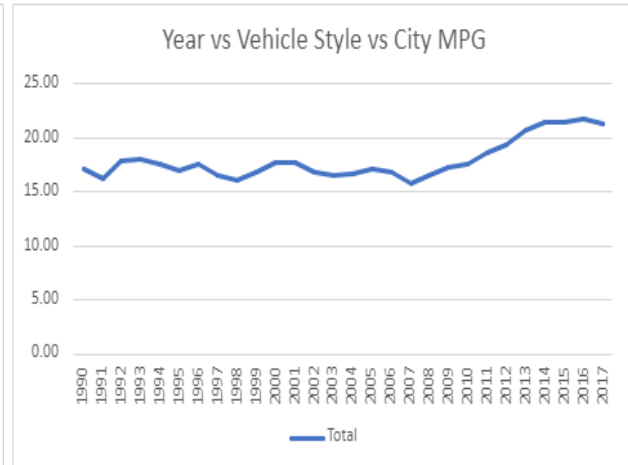
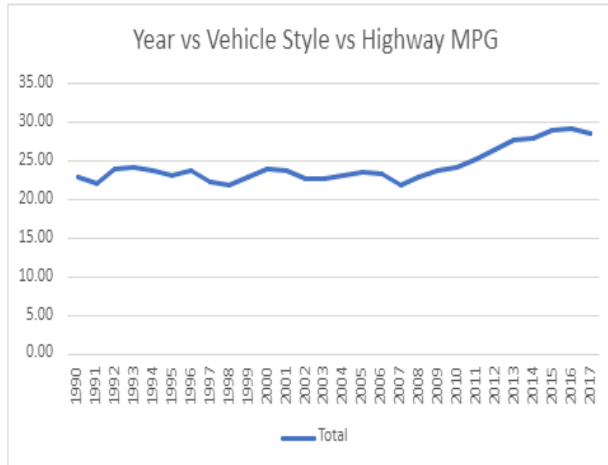
I am using the MS-Excel web version. In the web version using Scatter Plot with different symbols for each type is not available. So I used the same symbols for each body style.



**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.

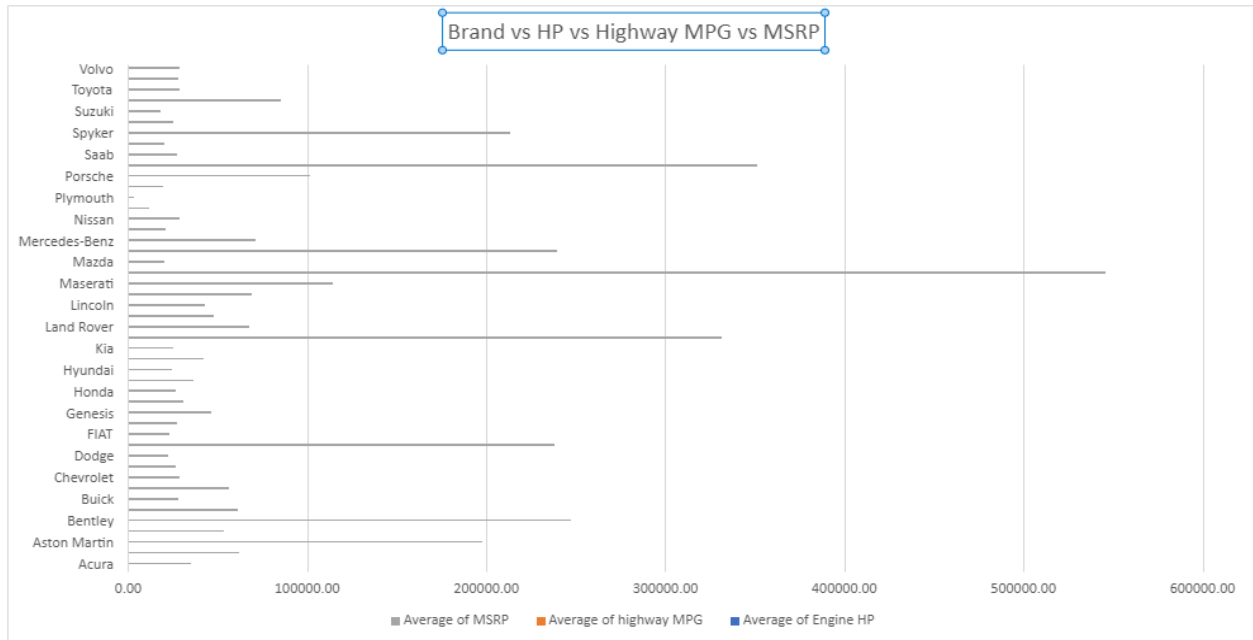
Again in the Web version of Excel there are limitations to Line charts as well. So to make it understandable, I added slicers.





**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.

Bubble charts don't exist in the web version of Excel. So instead, I used bar charts along with slicers.



Make	
Acura	 
Alfa Romeo	
Aston Martin	
Audi	
Bentley	
BMW	
Buick	
Cadillac	
Chevrolet	
Chrysler	