# Analysing the Impact of Car Features on Price and Profitability Project

## Project Description:

The main aim of this project is to identify patterns that indicate if a customer will have difficulty paying their instalments. 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.

This problem could be approached by analysing 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.

Dataset Description:

The dataset contains information on various car models and their specifications, and is titled "Car Features and MSRP". It was collected and made available on Kaggle by Cooper Union, a private college located in New York City.
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.
- **MSRP:** the manufacturer's suggested retail price of the car.

## Approach:

I went through the Excel data provided by the Trainity Impact of Car Features project and understood that there were columns related to the Car Features in the dataset. Further, I understood the columns and their respective constraints to do the analysis. I was given a set of questions to solve as part of the analysis. By using the Microsoft Excel, I did solve the queries and provided the result as expected.

## Tech-Stack Used:

Microsoft Excel 2021 – To answer the queries with the help of Excel formulas in the tool.

## Insights:

Did the data cleaning like:

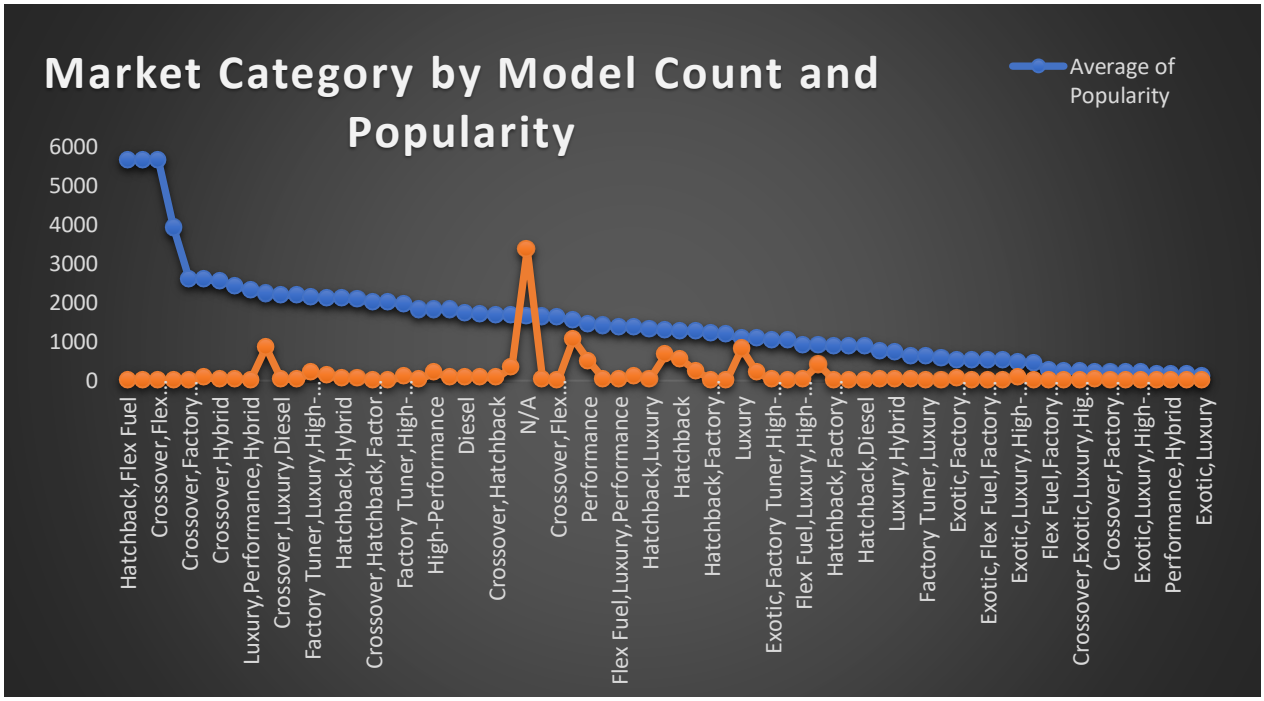
- Removing null values.
- Removed the columns which we don't use for the analysis.
- Removing the Duplicate rows.

Before the Data Cleaning the number of columns in the Excel were:

Car data – 11915, After cleaning now we have 11098 columns.

### Task A – How does the popularity of a car model vary across different market categories?

Consider the columns Car Model, Popularity and Market Categories. Convert the columns into a pivot chart and used a line chart to show the popularity and the count of models across various Market Categories.

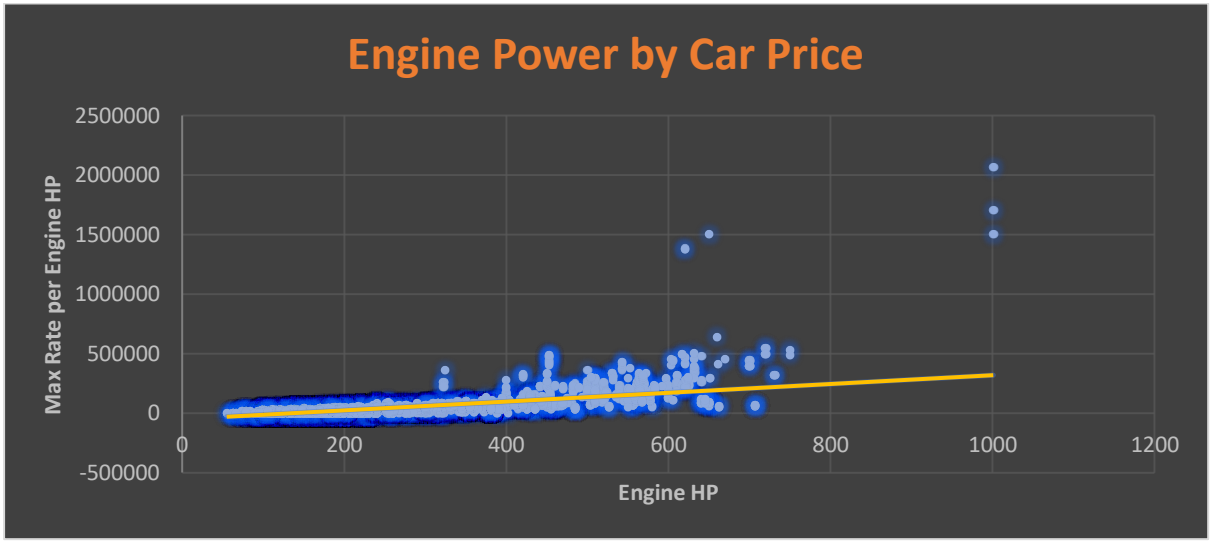


From the above chart it is evident the "**Crossover**" has a greater **number of counts** and with the **popularity** it is: **"Hatchback, Flex Fuel, Crossover, Performance"** is more.

## Task B – What is the relationship between a car's engine power and its price?

For this analysis, I have considered the columns Engine HP and MSRP for the comparison.

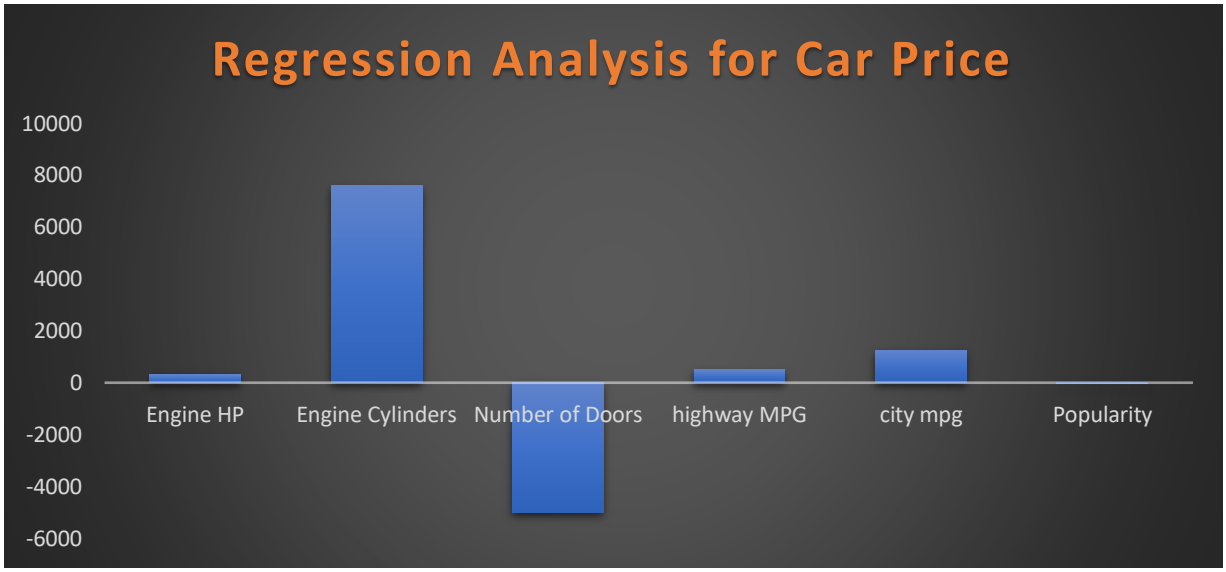
I used the Scatter Plot chart to find out the relationship between them and also added the Trendline in order to find out the trend between them.




Found the scatter plot, it is clear that the **with the increase in the Engine HP**, the **Price of the car also increases**. The Trendline also shows the same.

## Task C – Which car features are most important in determining a car's price?

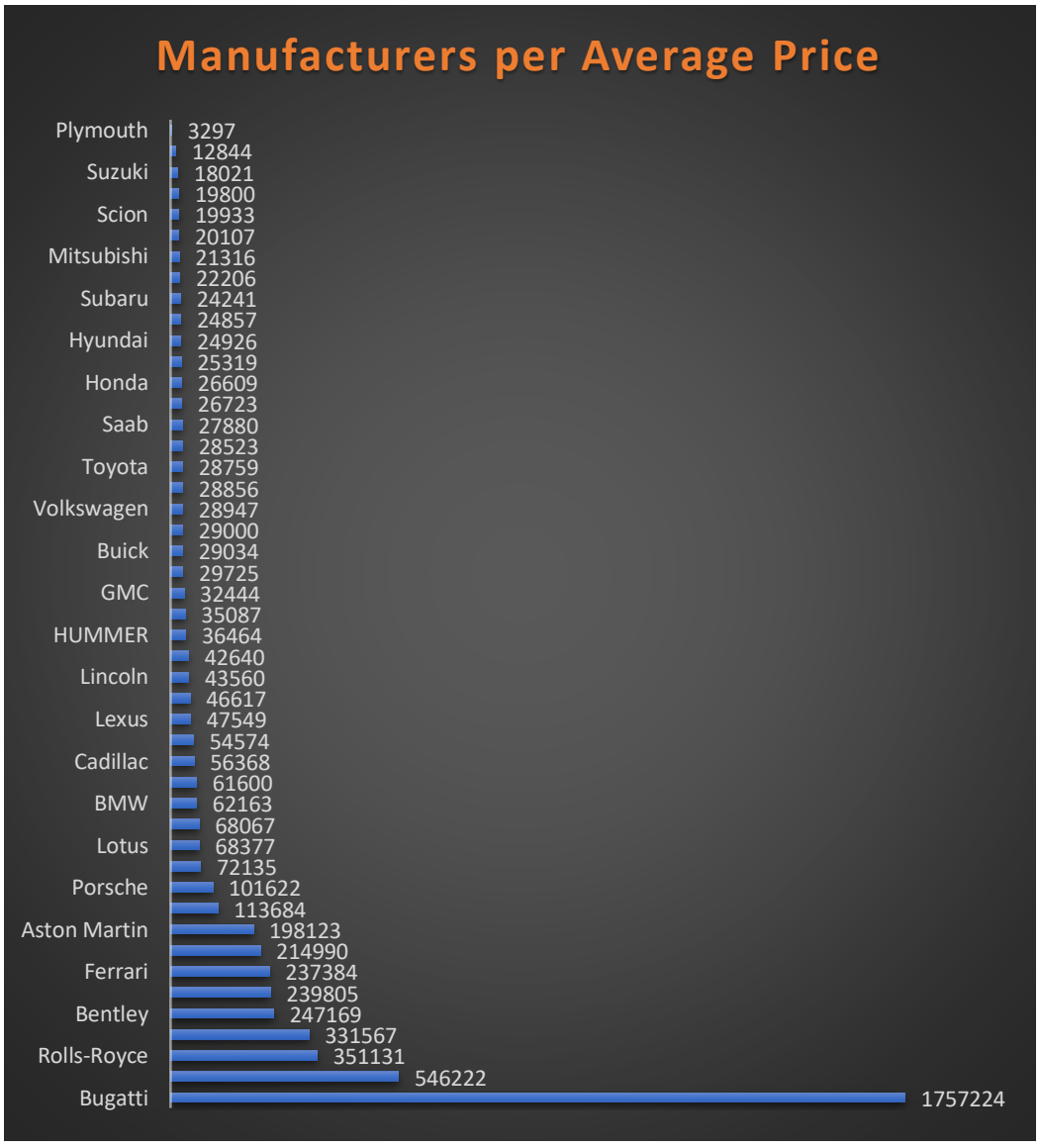
Use regression analysis to identify the variables that have the strongest relationship with a car's price. Then create a bar chart that shows the coefficient values for each variable to visualize their relative importance.



We can see that the "**Engine Cylinders**" are more related to the Car Price and the least related column is Number of doors.

## Task D – How does the average price of a car vary across different manufacturers?

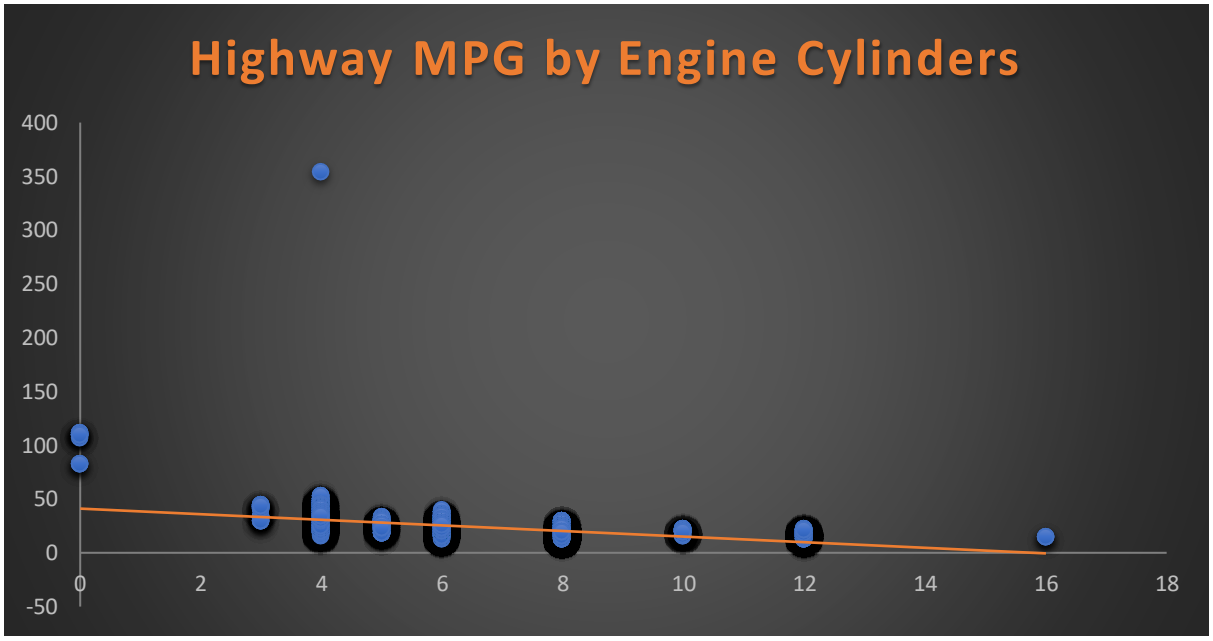
Considered the columns MSRP and car model for the analysis. Created a bar chart that visualizes the relationship between manufacturer and average price.




From the above Bar chart, we could see that **Bugatti** is having a **greater average** of MSRP among other brands. And **Plymouth** is having the **least average** of MSRP.

## Task E – What is the relationship between fuel efficiency and the number of cylinders in a car's engine?

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.




We could see that the **Trendline is coming down** when there is an **increase in the Engine Cylinders**.

Calculated the correlation coefficient between the number of cylinders and highway MPG to quantify the strength and direction of the relationship with an Excel formula:

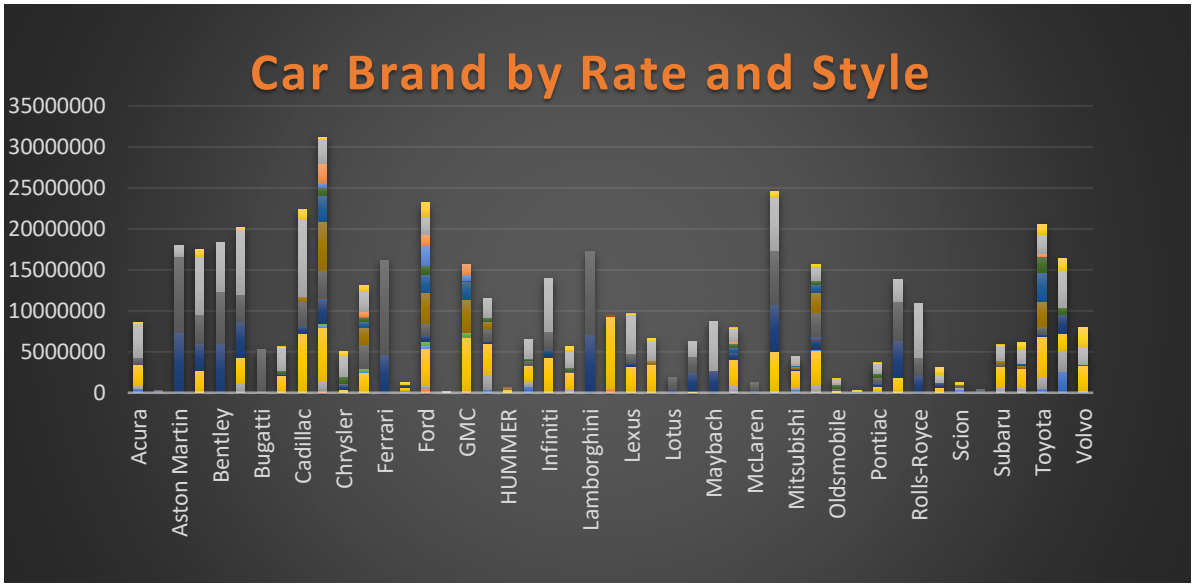
=CORREL(A1:A11098,B1:B11098)

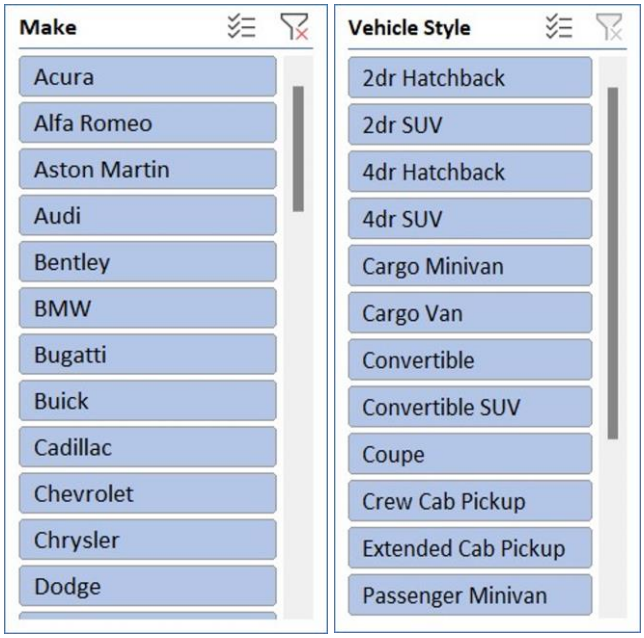
And found the correlation is negative (-0.6147).

## Building a Dashboard:

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

Used Stacked column chart to show the distribution of car prices by brand and body style. Used filters and slicers to make the chart interactive. Calculated the total MSRP for each brand and body style using Pivot Tables.






Used slicer to filter the values in the Stacked Column chart to do the analysis.

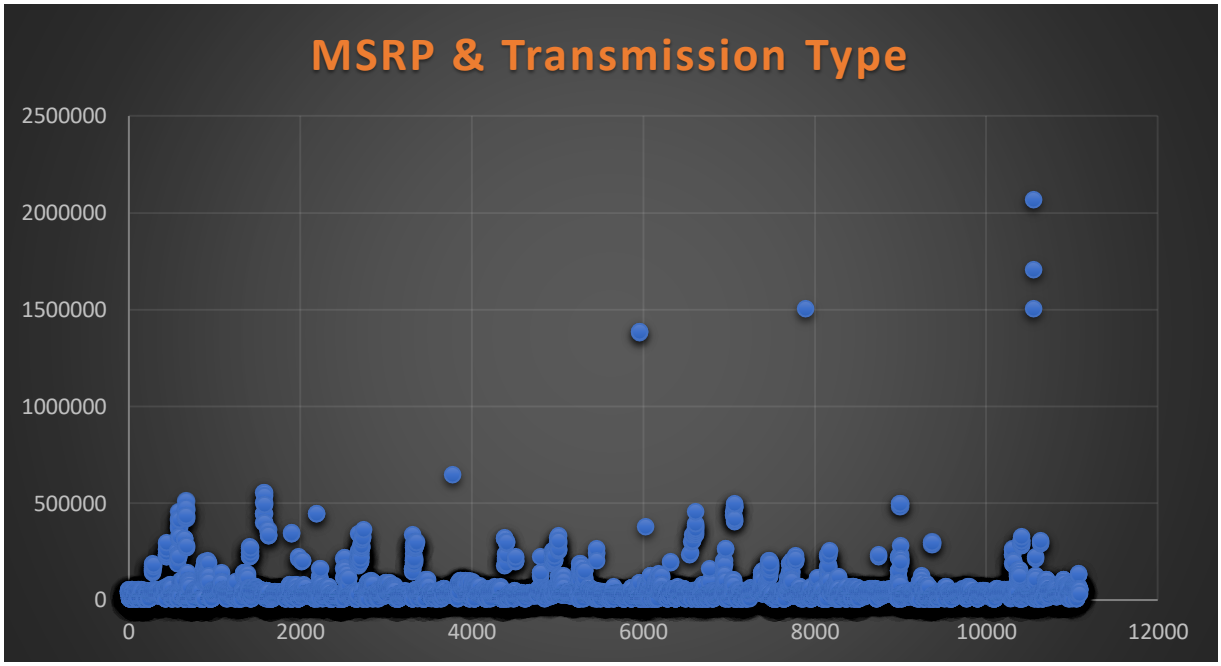
## Task 2: Which car brands have the highest and lowest average MSRPs, and how does this vary by body style?

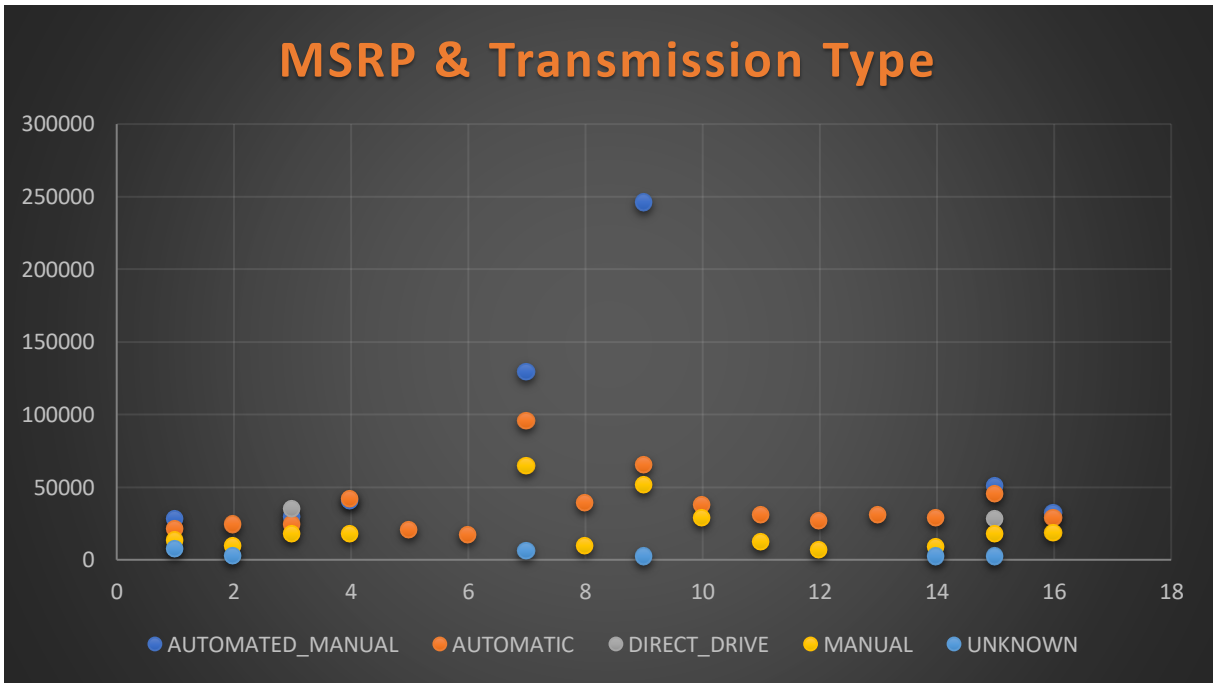
Used Clustered column chart to compare the average MSRPs across different car brands and body styles. Calculated the average MSRP for each brand and body style using Pivot Tables.


It is evident from the chart that **Bugatti** is the car brand with **Highest Average** and **Plymouth** is the car brand with **Lowest Average**.

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

Used Scatter plot chart to visualize the relationship between MSRP and transmission type, with different symbols for each body style. Calculated the average MSRP for each combination of transmission type and body style using Pivot Tables.



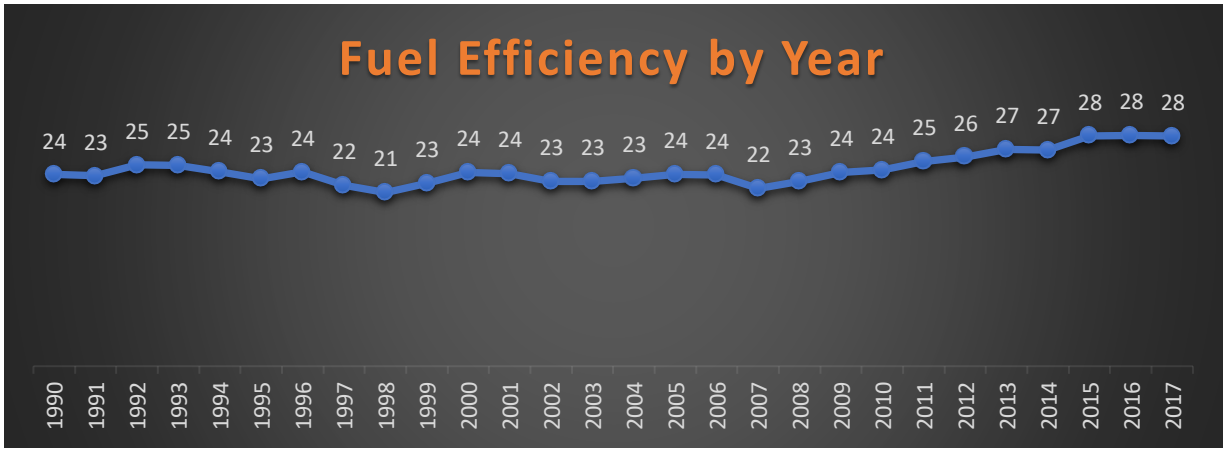




The price is more for the transmission type **Automated Manual**.

The price is less for the transmission type **Manual**.

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

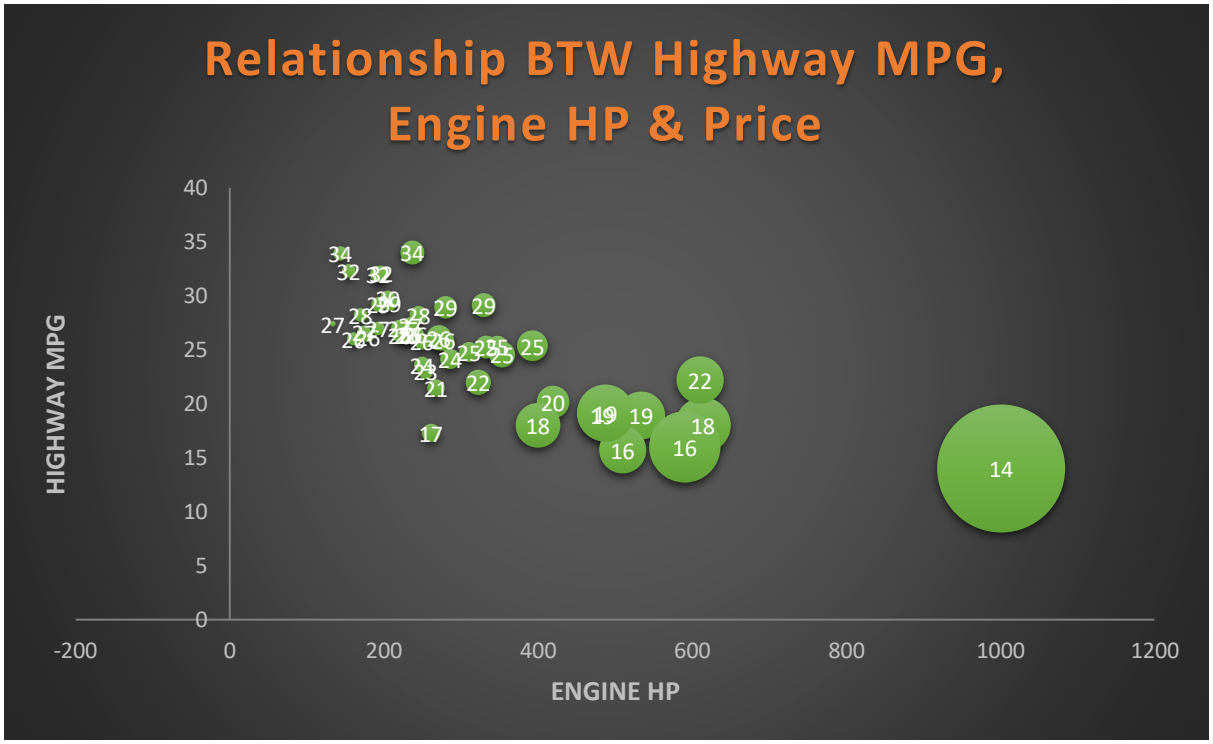
Used Line chart to show the trend of fuel efficiency (MPG) over time for each body style. Calculated the average MPG for each combination of body style and model year using Pivot Tables.




Although there was a dip in the year 2007, the **Fuel efficiency is slowly increasing year by year**.

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

Used Bubble chart to visualize the relationship between horsepower, MPG, and price across different car brands. Assigned different colours to each brand and label the bubbles with the car model name. Calculated the average horsepower, MPG, and MSRP for each car brand using Pivot Tables.




As a result, we could see that when there is an **increase in the Engine HP**, the **MRSP also increases** and there is a **decrease in the Highway MPG.**

## Result:

Through this project I was able to understand the Pivot Tables being used in the Excel which can be used to find the Correlation and various charts on how to use them. I got used to the Excel Pivot tables and how to convert the Raw Data into meaningful insights. And the steps which I used are – cleansing the data and using the formulas and pivot table to create the chart for the analysis and also learnt how to convert the data into a visualized chart so that the insights can be drawn within seconds by seeing the graphs instead of searching the whole data.

As a result, we could summarize as:

1. **The "Crossover" has a greater number of counts and with the popularity it is: "Hatchback, Flex Fuel, Crossover, Performance".**
2. **With the increase in the Engine HP, the Price of the car also increases.**
3. **The "Engine Cylinders" are more related to the Car Price.**
4. **Bugatti is having a greater average of MSRP and Plymouth is having the least average.**
5. **The transmission type Automated Manual is expensive.**
6. **The Fuel efficiency is slowly increasing year by year.**
7. **When there is an increase in the Engine HP, the MRSP also increases and there is a decrease in the Highway MPG.**

I have achieved the end result and I think I have contributed my full support into the Analysis. I hope this project helps the Analysis and it achieves what it was tend to achieve.

### Hyperlink for the Excel sheet:

**Trainity Impact of Car Features Excel File**