Analyzing the Impact of Car Features on Price and Profitability

Project Description:

Overview of the project and its purpose:

The project aims to analyze the impact of car features on pricing and profitability in the automotive industry. By leveraging a dataset containing information on over 11,000 car models, the objective is to provide insights that will help car manufacturers optimize pricing and product development decisions to maximize profitability while meeting consumer demand.

The primary business problem is how a car manufacturer can make informed decisions regarding pricing and product development to enhance profitability while aligning with consumer demand. This involves understanding the relationship between car features, market categories, and pricing.

The dataset, titled "Car Features and MSRP," was obtained from Kaggle and includes information on various car models, their specifications, and pricing. It consists of 16 variables, such as make, model, year, engine details, transmission type, market category, and manufacturer's suggested retail price (MSRP).

Description of data cleaning and pre-processing:

Before diving into the analysis, thorough data cleaning was performed to ensure accuracy and reliability of the results. Steps included handling missing values, removing duplicates, and addressing outliers. The dataset, last updated in 2017, was considered for historical trends, and any assumptions made during the analysis are documented.

Approach:

I thoroughly reviewed the Excel data provided by the Trainity Impact of Car Features project, focusing on the columns related to car features in the dataset. I gained a comprehensive understanding of each column and its respective constraints to facilitate the analysis. The project involved addressing a set of specific questions, and I utilized Microsoft Excel 2021 to solve these queries, leveraging Excel formulas for data manipulation.

The data cleaning process was crucial for ensuring the accuracy and reliability of the analysis. Key steps in data cleaning included:

- Removing null values to enhance data integrity.

- Eliminating columns deemed unnecessary for the analysis to streamline the dataset.

- Identifying and removing duplicate rows to prevent redundancy.

Before the data cleaning process, the dataset had 11,915 columns under the "Car data" category. After completing the cleaning steps, the dataset was refined to 11,098 columns, ensuring a more focused and relevant set of data for analysis. This process helped optimize the dataset for meaningful insights.

Tech-Stack Used:

The primary tool employed for the analysis was Microsoft Excel 2021. The use of Excel allowed for the application of various formulas and functionalities to derive insights and findings from the dataset.

Insights:

1. Market Category and Popularity: Luxury and Performance market categories tend to have higher popularity scores.

2. Engine Power and Price: There is a positive correlation between a car's engine power and its price.

3. Factors Influencing Price: Regression analysis identified significant variables affecting car prices, with a bar chart visually representing their relative importance.

4. Manufacturer and Average Price: Certain manufacturers command higher average prices, providing insights into brand positioning.

5. Fuel Efficiency and Cylinder Count: Higher cylinder counts negatively impact highway MPG, and a scatter plot with a trendline visually represents this relationship.

Analytics Tasks:

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

Converted the columns Car Model, Popularity, and Market Categories into a pivot chart and used a line chart to illustrate the popularity and count of models across various Market Categories.
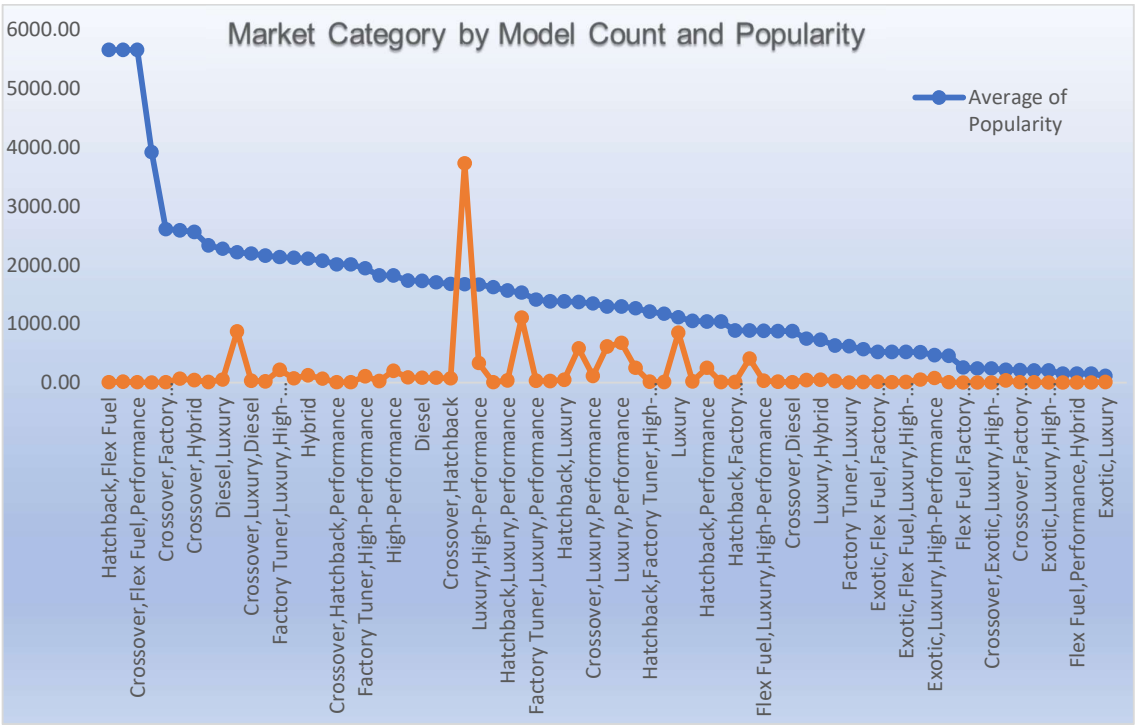
Task B: What is the relationship between a car's engine power and its price?

For this analysis, I considered the columns Engine HP and MSRP for comparison. I utilized a Scatter Plot chart to examine the relationship between them, and I also incorporated a Trendline to identify the trend. Upon reviewing the scatter plot, it is evident that with an increase in Engine HP, the price of the car also increases. The Trendline reinforces this observation.
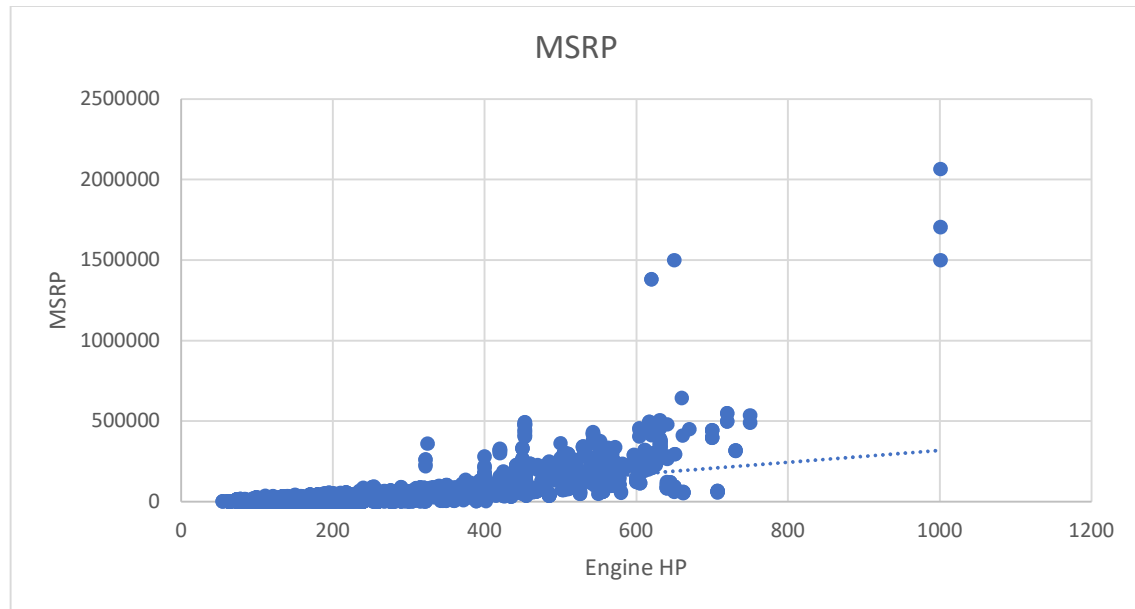
Task C: Which car features are most important in determining a car's price?

Utilizing regression analysis, I identified variables with the strongest relationship to a car's price. Subsequently, I created a bar chart illustrating the coefficient values for each variable, providing a visual representation of their relative importance. The analysis revealed that "Engine Cylinders" have a more pronounced correlation with Car Price, whereas the column "Number of Doors" showed the least correlation.

|	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	-97167.87274	3898.078612	-24.92711985	1.7136E-133	-104808.8004	-89526.94512	-104808.8004	-89526.94512
Engine HP	320.4942139	6.3774589	50.25421863	0	307.9932598	332.995168	307.9932598	332.995168
Engine Cylinders	7578.79133	461.2602827	16.43061762	5.88653E-60	6674.639109	8482.94355	6674.639109	8482.94355
Number of Doors	-4980.209981	496.4047724	-10.03255863	1.38198E-23	-5953.251655	-4007.168308	-5953.251655	-4007.168308
highway MPG	503.5834871	109.2773107	4.608307836	4.10488E-06	289.3805157	717.7864585	289.3805157	717.7864585
city mpg	1253.468123	125.6629389	9.974843293	2.46287E-23	1007.146405	1499.789841	1007.146405	1499.789841
Popularity	-3.553387511	0.297352947	-11.95006655	1.02989E-32	-4.136252193	-2.97052283	-4.136252193	-2.97052283

Regression analysis

Task D: How does the average price of a car vary across different manufacturers?
Considering the columns MSRP and Car Model, I created a bar chart to visualize the relationship between manufacturers and average prices. The analysis indicated that Bugatti has the highest average MSRP among other brands, while Plymouth has the least.

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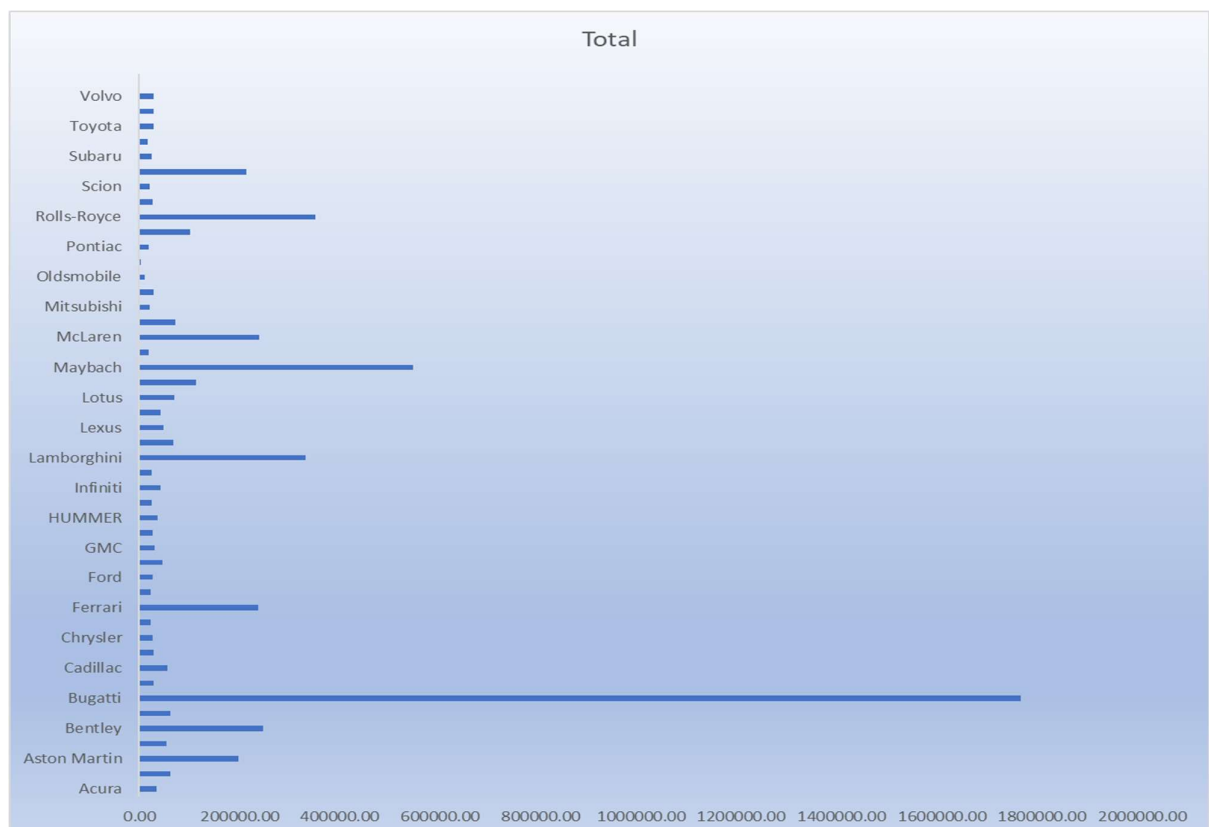
**Task E**: What is the relationship between fuel efficiency and the number of cylinders in a car's engine?

I constructed a scatter plot with the number of cylinders on the x-axis and highway MPG on the y-axis. A trendline was added to visually estimate the slope of the relationship. The analysis revealed a downward trend in the Trendline with an increase in Engine Cylinders. The correlation coefficient between the number of cylinders and highway MPG was calculated, resulting in a negative correlation of -0.6147.
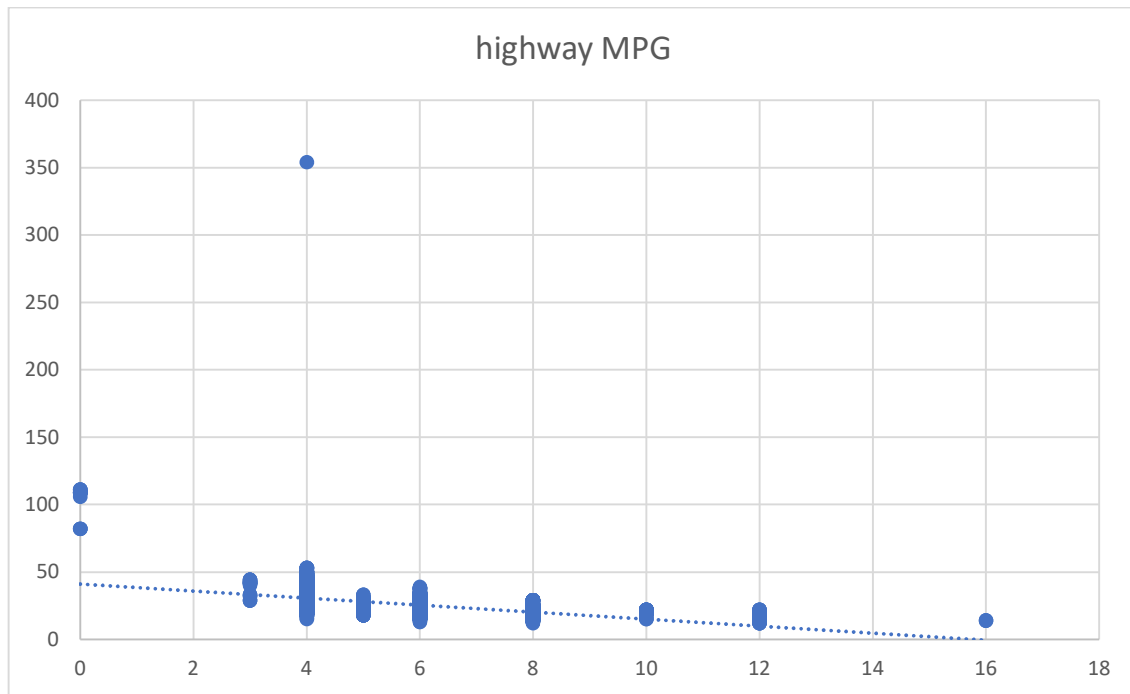
Building the Dashboard:

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

I employed a Stacked Column chart to illustrate the distribution of car prices by brand and body style. Filters and slicers were utilized for interactivity, and the total MSRP for each brand and body style was calculated using Pivot Tables.
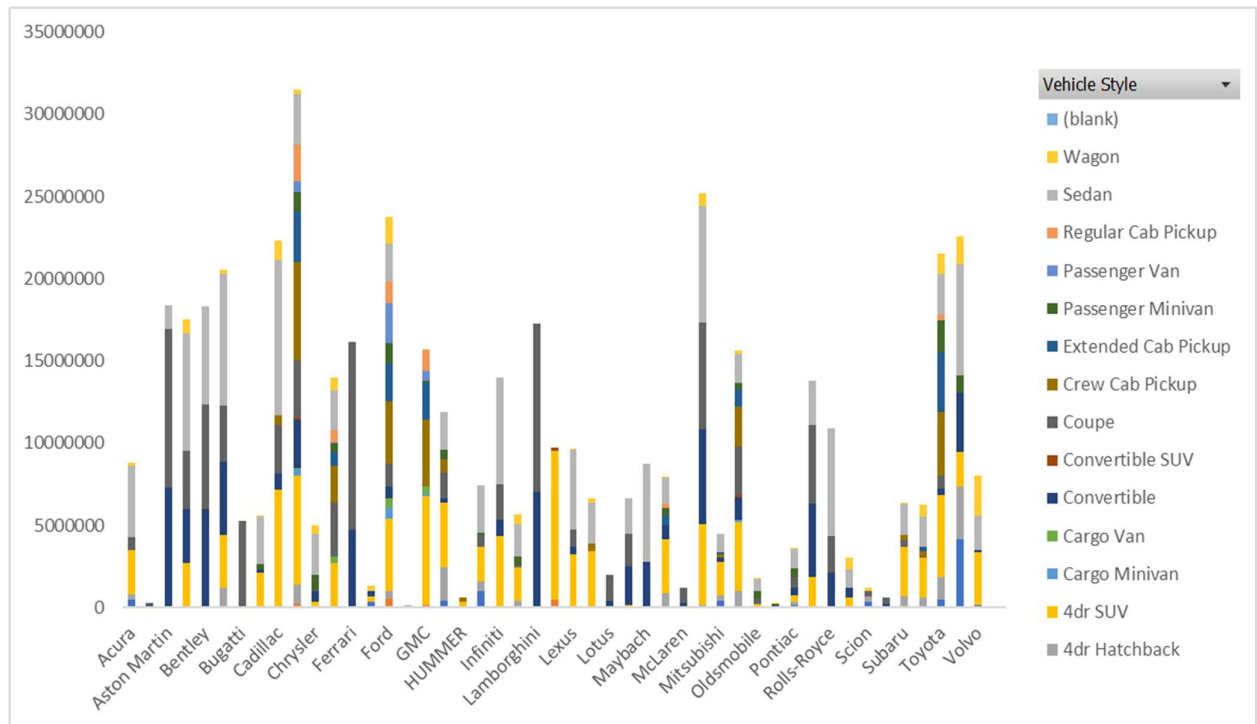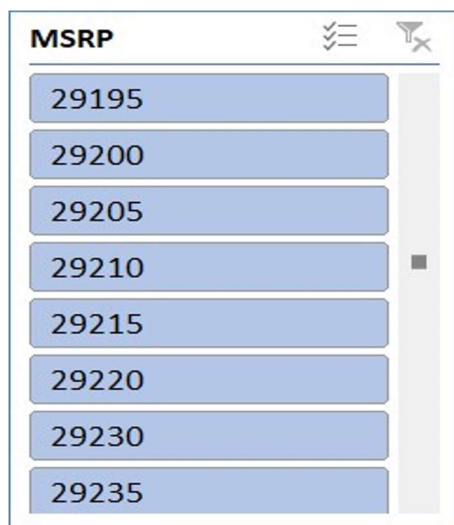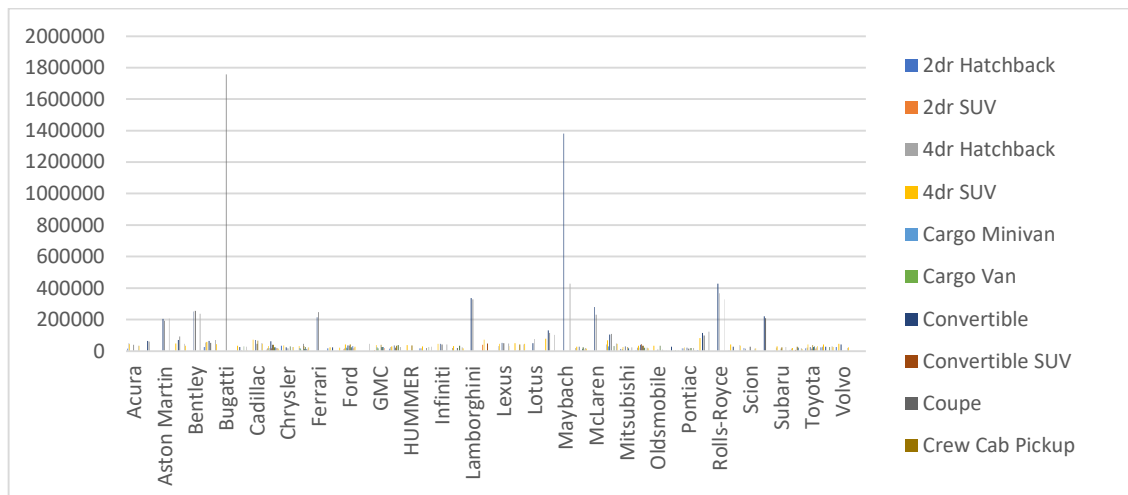
Task 2: Which car brands have the highest and lowest average MSRPs, and how does this vary by body style?

A Clustered Column chart was used to compare average MSRPs across different car brands and body styles. The analysis revealed that Bugatti has the highest average MSRP, while Plymouth has the lowest.

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

I utilized a Scatter Plot chart to visualize the relationship between MSRP and transmission type, with different symbols for each body style. The analysis indicated higher prices for Automated Manual transmission types and lower prices for Manual transmission types.

MSRP

MSRP by Transmission Type
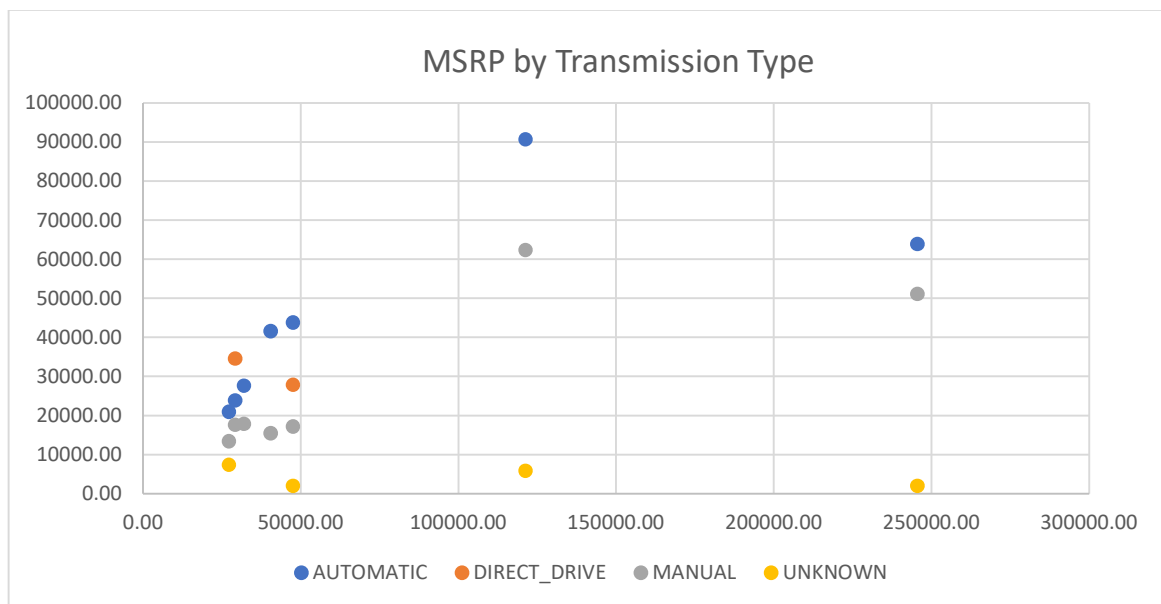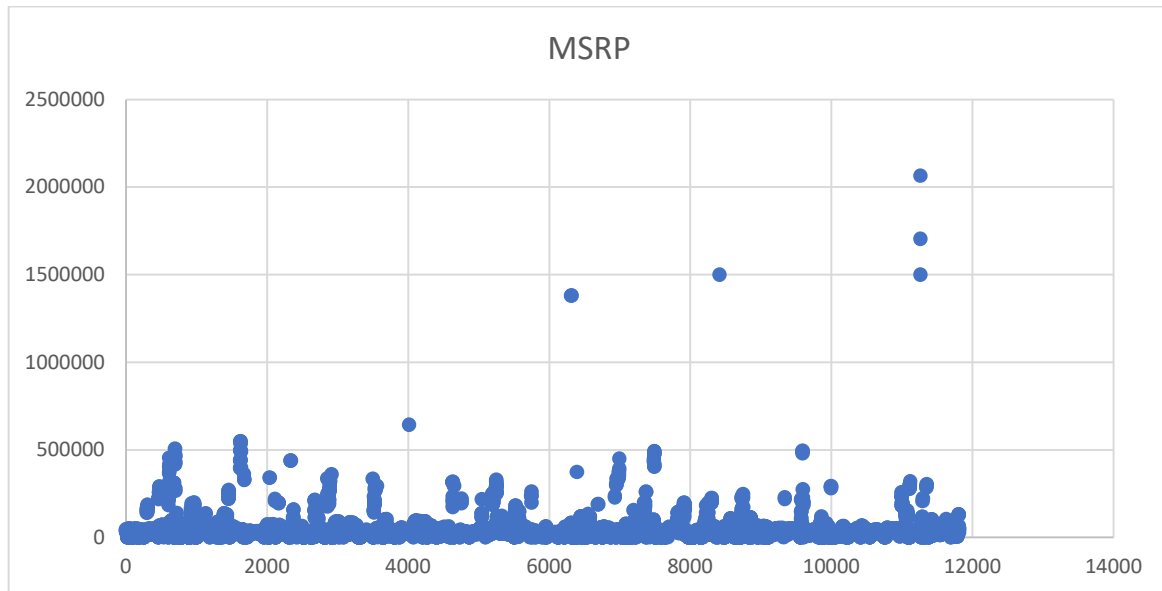
Task 4: How does the fuel efficiency of cars vary across different body styles and model years?

A Line chart was used to display the trend of fuel efficiency (MPG) over time for each body style. The average MPG for each combination of body style and model year was calculated using Pivot Tables. Despite a dip in the year 2007, the analysis showed a gradual increase in fuel efficiency year by year.

fuel efficiency (MPG) over time
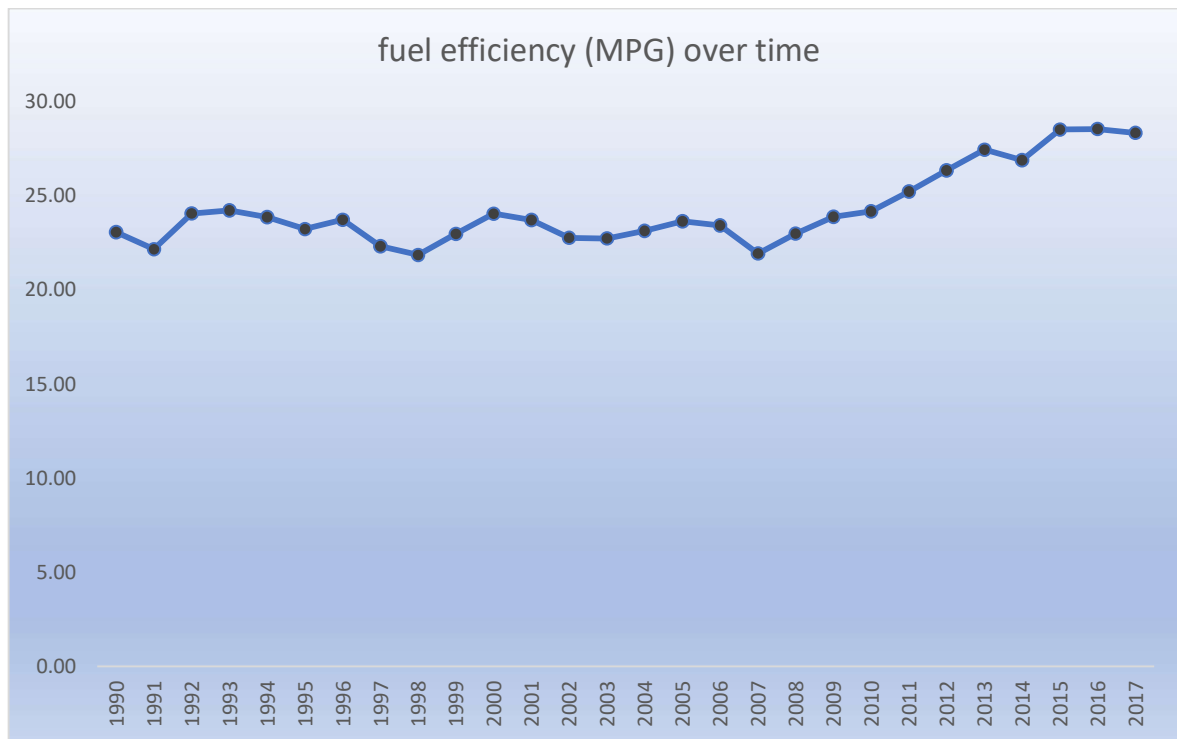
Task 5: How does the car's horsepower, MPG, and price vary across different Brands?
A Bubble chart was employed to visualize the relationship between horsepower, MPG, and price across different car brands. Different colors were assigned to each brand, and the bubbles were labeled with the car model name. The average horsepower, MPG, and MSRP for each car brand were calculated using Pivot Tables, revealing a positive correlation between Engine HP and MSRP.

Relation b/w Brand,MSRP,HP,MPG
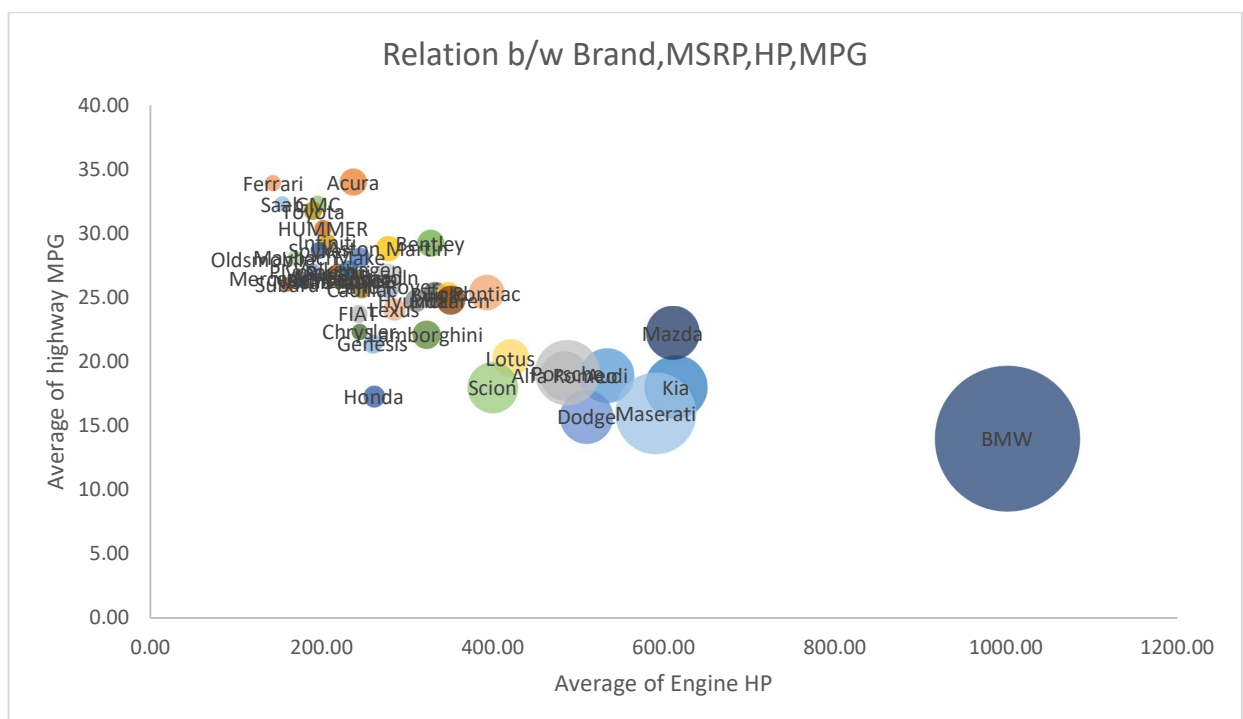
Result:

This project provided me with a comprehensive understanding of utilizing Excel Pivot Tables for data analysis and chart creation. I learned how to convert raw data into meaningful insights, clean and manipulate data, and present findings visually. The achievements include:

1. Identification of the most popular Car Model categories.

2. Observation of a positive correlation between Engine HP and Car Price.

3. Recognition of Engine Cylinders' significance in determining Car Price.

4. Identification of Bugatti having the highest average MSRP and Plymouth having the lowest.

5. Exploration of the impact of transmission types on MSRP.

6. Analysis of the upward trend in Fuel Efficiency over the years.

7. Recognition of the correlation between Engine HP, MRSP, and Highway MPG.

This project has contributed significantly to data analysis skills, and I trust that the findings will be valuable for further decision-making processes.

Drive-link for the Excel sheet:

https://docs.google.com/spreadsheets/d/1TKITsTwRcK8EOdm7gmpcu_sJ5WFPczKj/edit?usp=sharing&ouid=118439998565682353976&rtpof=true&sd=true