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

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 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.

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 (based on the number of times it has been viewed on Edmunds.com)
- **MSRP:** the manufacturer's suggested retail price of the car

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).

Approach:

The Data Analysis is made using Google Sheets and pivot tables. Different Pivot sheets are used to analyse the required insights

TECH STACK USED:

1. Google Sheets – for storing the dataset and using the formulas to draw insights within the data set.
2. Google Search Engine and YouTube for reference.

INSIGHT:

The working file is in the following link:

https://docs.google.com/spreadsheets/d/1H-kMocEfYcT0ufNAr2bJeVZTJi9OYAginy0xnM9i5rU/edit?usp=sharing

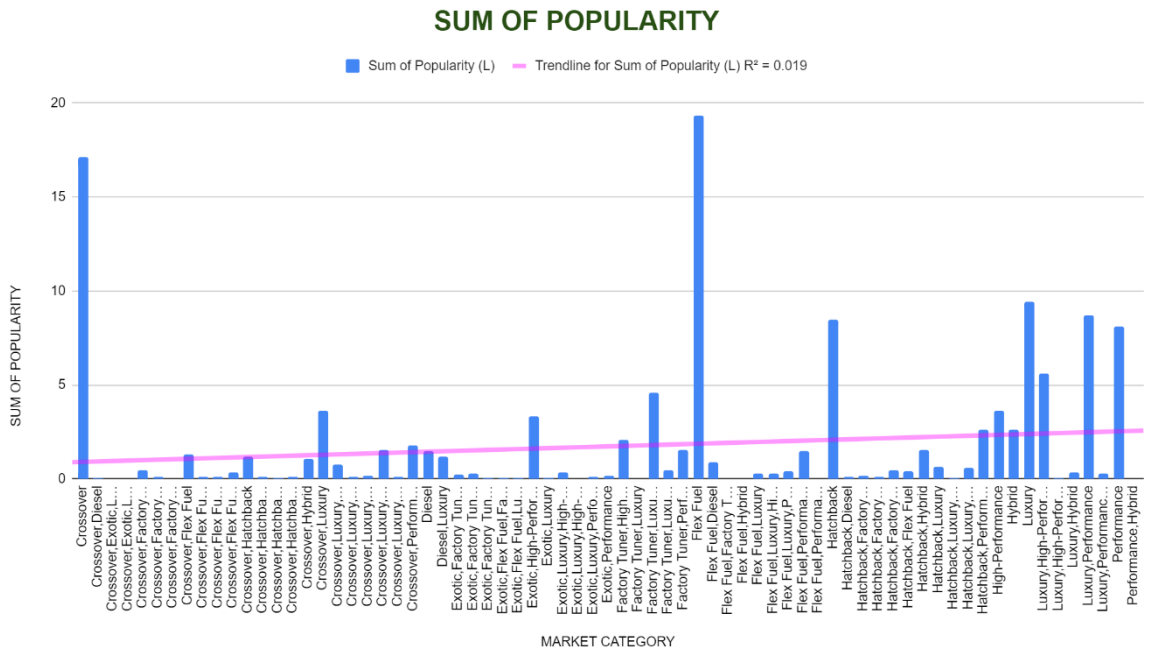
Insight Required: How does the popularity of a car model vary across different market categories?

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

A	B	C
Market Category	COUNTA of Model	SUM of Popularity
Crossover	1110	1715242
Crossover,Diesel	7	6111
Crossover,Exotic,Luxury,High-Performance	1	238
Crossover,Exotic,Luxury,Performance	1	238
Crossover,Factory Tuner,Luxury,High-Performance	26	47410
Crossover,Factory Tuner,Luxury,Performance	5	13037
Crossover,Factory Tuner,Performance	4	840
Crossover,Flex Fuel	64	132720
Crossover,Flex Fuel,Luxury	10	11732
Crossover,Flex Fuel,Luxury,Performance	6	9744
Crossover,Flex Fuel,Performance	6	33942
Crossover,Hatchback	72	120650
Crossover,Hatchback,Factory Tuner,Performance	6	12054
Crossover,Hatchback,Luxury	7	1428
Crossover,Hatchback,Performance	6	12054
Crossover,Hybrid	42	107662
Crossover,Luxury	410	362665
Crossover,Luxury,Diesel	34	73080
Crossover,Luxury,High-Performance	9	9335
Crossover,Luxury,Hybrid	24	15142
Crossover,Luxury,Performance	113	151968
Crossover,Luxury,Performance,Hybrid	2	7832
Crossover,Performance	69	178431
Diesel	84	145396
Diesel,Luxury	51	116025
otic,Factory Tuner,High-Performance	21	21974

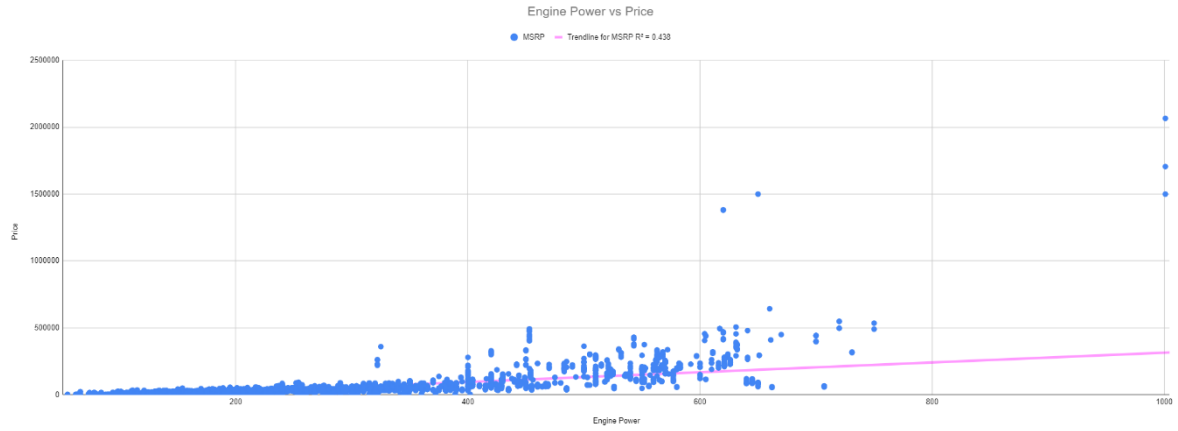
The full pivot table is in the following link: https://docs.google.com/spreadsheets/d/1H-kMocEfYcT0ufNAr2bJeVZTJi9OYAginy0xnM9i5rU/edit?usp=sharing

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



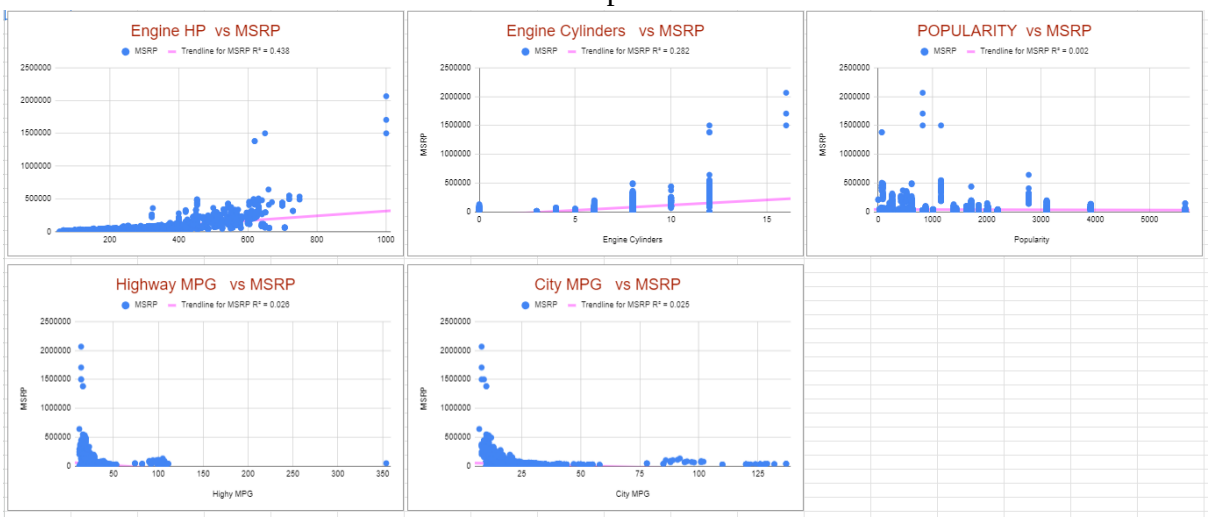
Insight Required: What is the relationship between a car's engine power and its price?

- **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.



The above charts show the linear regression Analysis for different features. And also have the R^2 to show the correlation among those features. We observe that the price is directly correlated to Engine HP and Engine Cylinders.

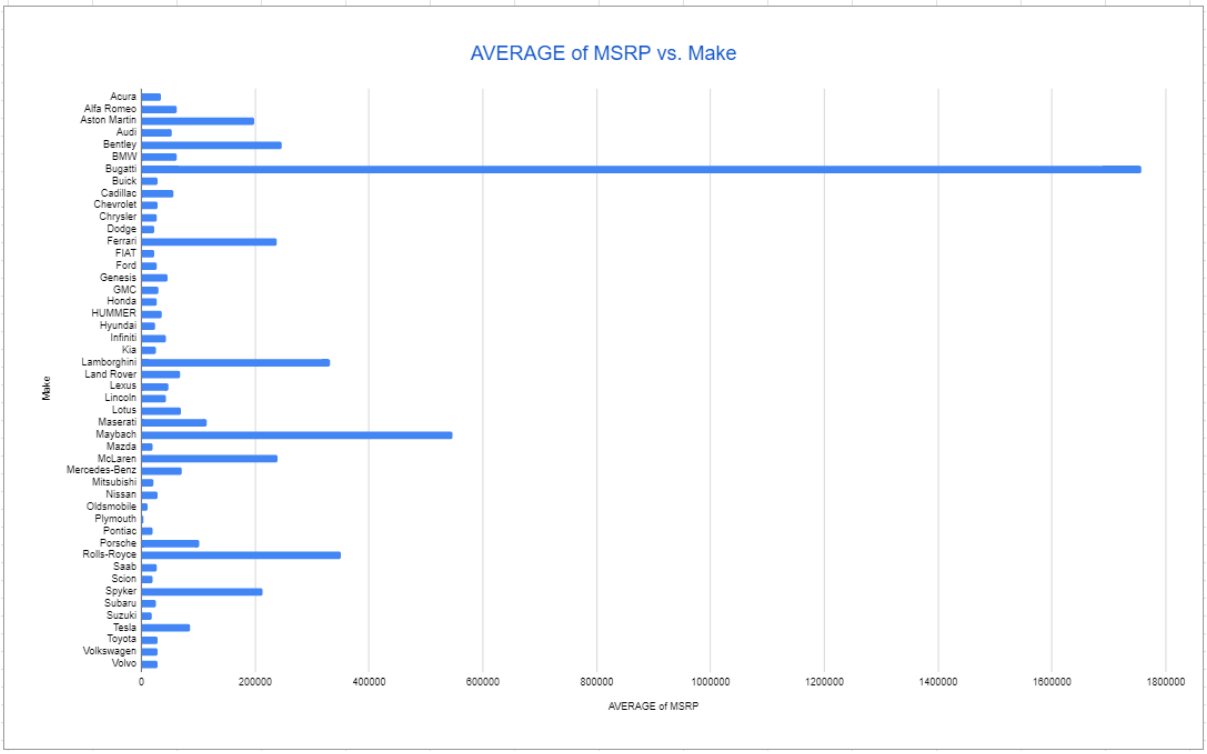
Insight Required: How does the average price of a car vary across different manufacturers?

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

A	B
Make	AVERAGE of MSRP
Acura	34887.5873
Alfa Romeo	61600
Aston Martin	197910.3763
Audi	53452.1128
Bentley	247169.3243
BMW	61546.76347
Bugatti	1757223.667
Buick	28206.61224
Cadillac	56231.31738
Chevrolet	28350.38557
Chrysler	26722.96257
Dodge	22390.05911
Ferrari	238218.8406
FIAT	22670.24194
Ford	27399.26674
Genesis	46616.66667
GMC	30493.29903
Honda	26674.34076
HUMMER	36464.41176
Hyundai	24597.0363
Infiniti	42394.21212
Kia	25310.17316
Lamborghini	331567.3077
Land Rover	67823.21678
Lexus	47549.06931
coln	42839.82927

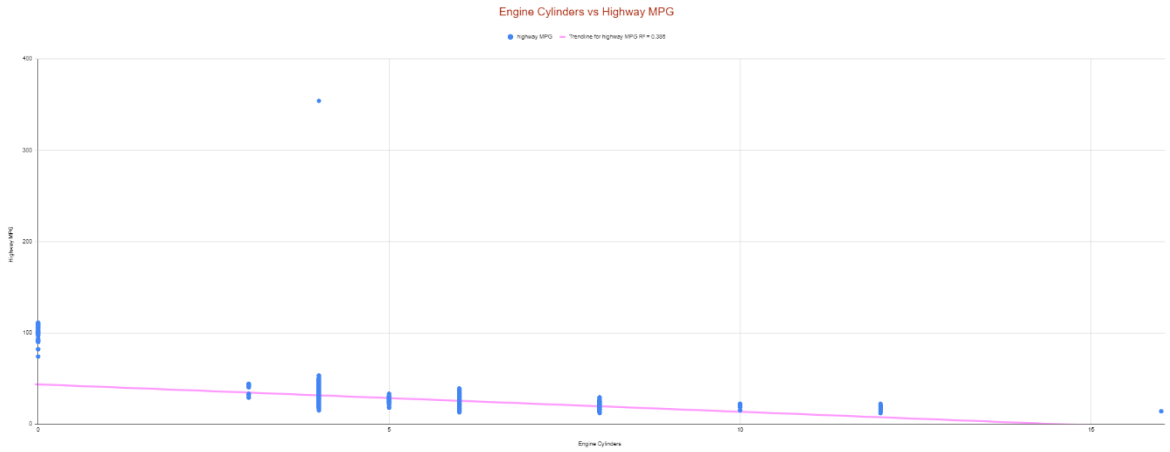
If we sort the average price in descending order we find that Bugatti sells at highest prices as compared to others and Plymouth sells at lowest prices in the given data.

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



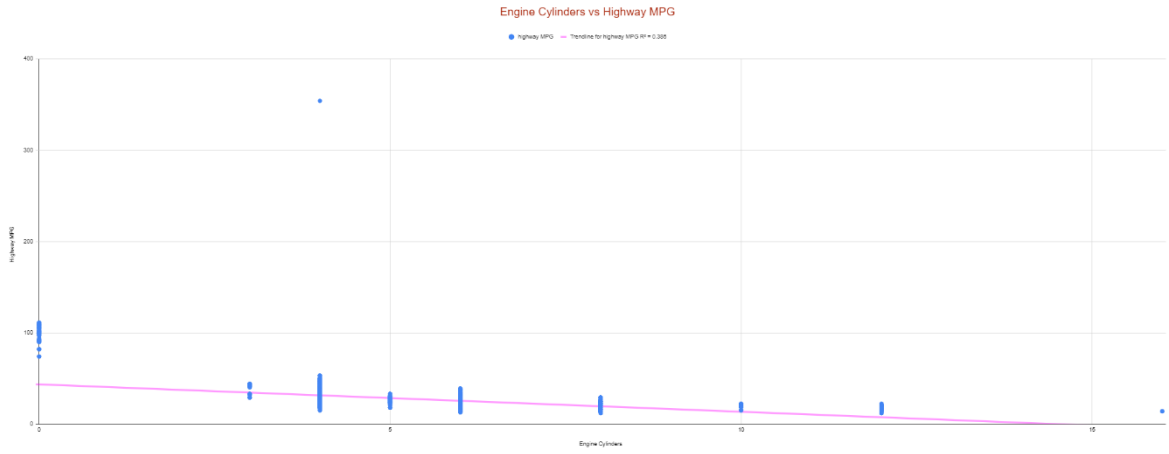
Insight Required: What is the relationship between fuel efficiency and the number of cylinders in a car's engine?

- **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.



The Engine Cylinders has a high correlation with the Highway MPG where R^2 is 0.386. However, since the trendline is negative we can decide that as the cylinders increase the Highway MPG value decreases.

- **Task 5.B:** Calculate the correlation coefficient between the number of cylinders and highway MPG to quantify the strength and direction of the relationship.

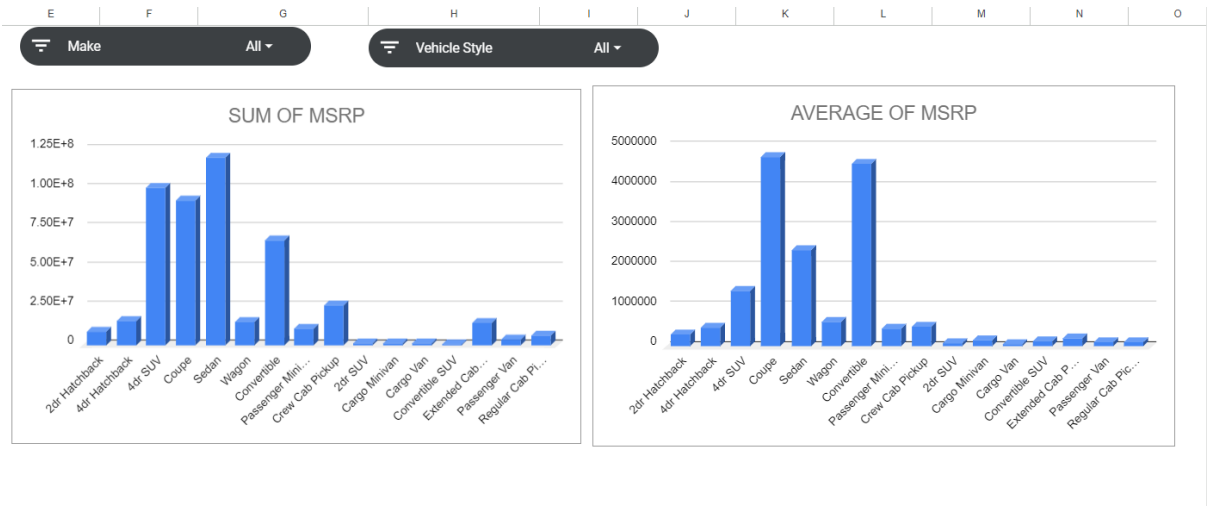


The correlation coefficient is 0.38 and has a negative correlation.

DASHBOARD

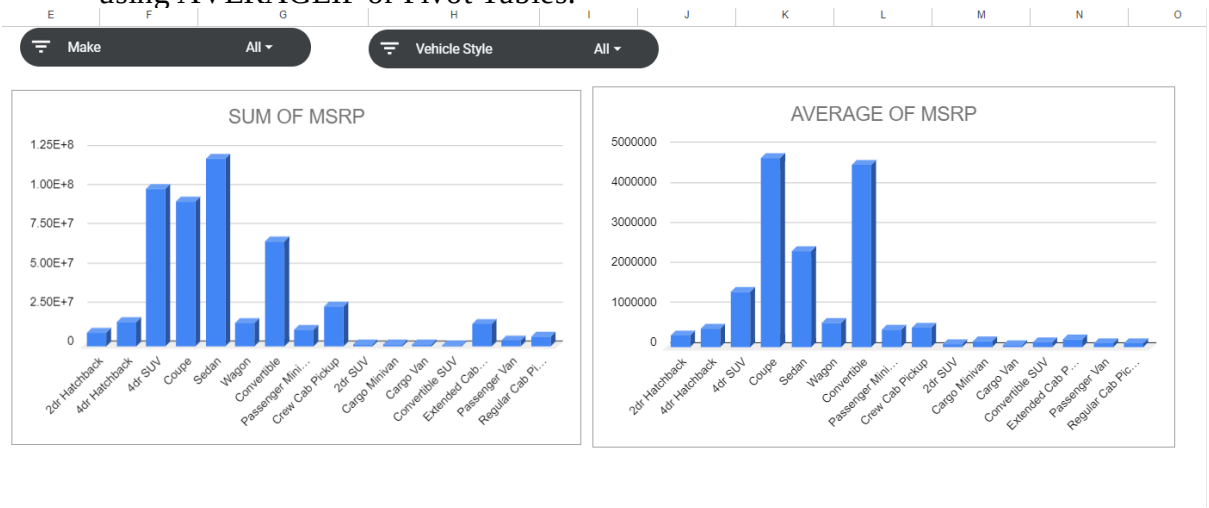
Task 1: How does the distribution of car prices vary by brand and body style?

- **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.



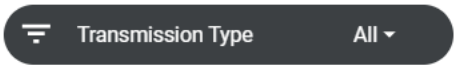
Task 2: Which car brands have the highest and lowest average MSRPs, and how does this vary by body style?

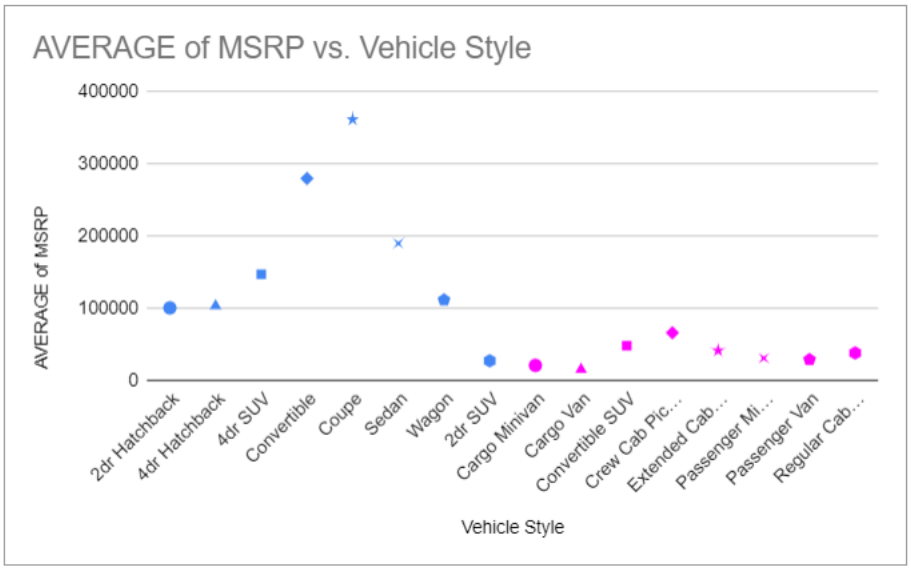
- **Hints:** Clustered column chart to compare the average MSRPs across different car brands and body styles. Calculate the average MSRP for each brand and body style using AVERAGEIF or Pivot Tables.



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

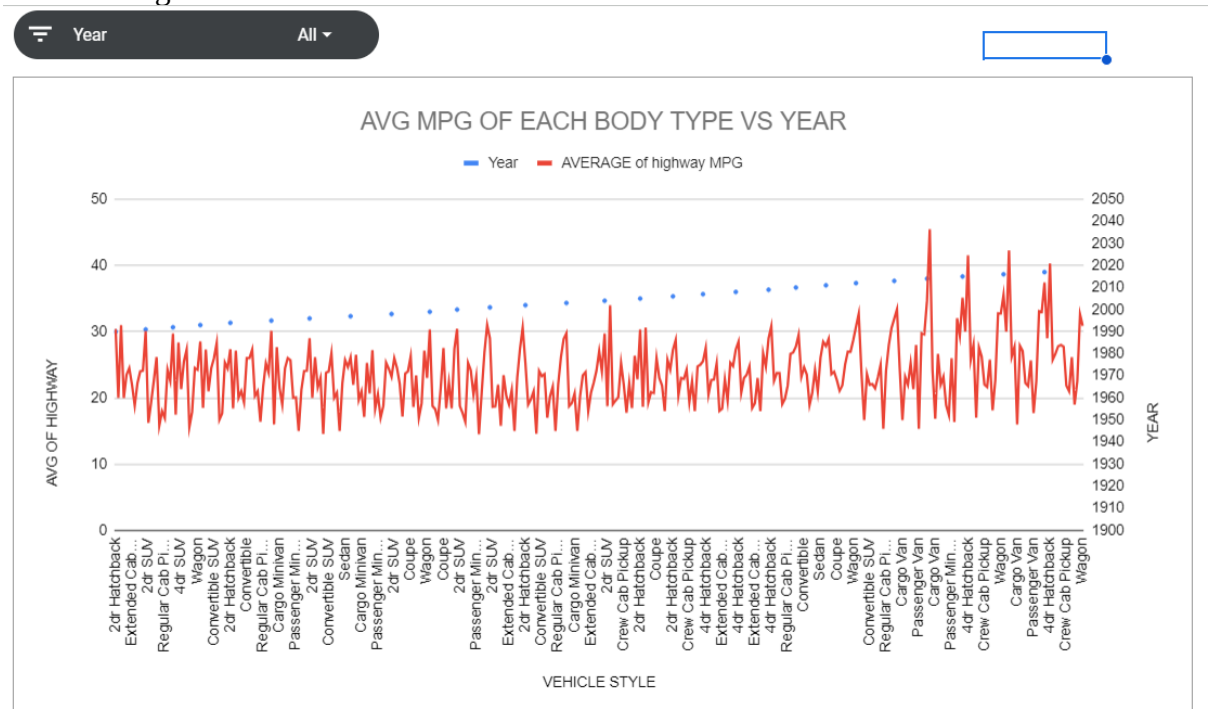
- **Hints:** Scatter plot chart to visualize the relationship between MSRP and transmission type, with different symbols for each body style. Calculate the average MSRP for each combination of transmission type and body style using AVERAGEIFS or Pivot Tables.





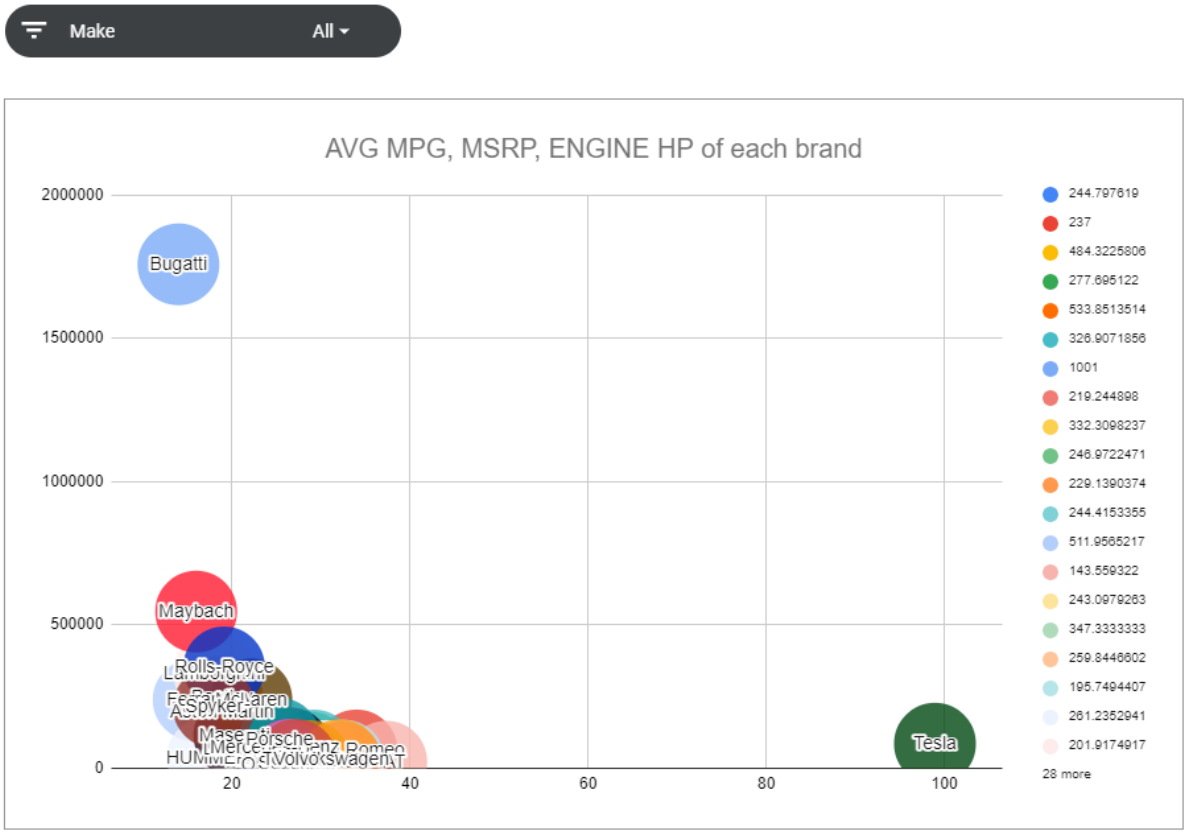
Task 4: How does the fuel efficiency of cars vary across different body styles and model years?

- **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 colours 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.



RESULT:

1. The least popular market category is Performance, Hybrid and the most popular is Flex Fuel
2. The Engine power is positively correlated to Price with a coefficient value of 0.438
3. Engine HP and Engine Cylinders show a positive correlation with the price.
4. Bugatti sells cars at highest price whereas Plymouth sells its cars at the least price
5. The number of engine cylinders are negatively correlated with highway MPG.
6. The vehicle type with highest average price across all manufacturers is Coupe whereas with highest total price across all manufacturers is Sedan.
7. 2drHatchback, 4drHatchback, Wagon are the vehicles with maximum average highway MPG.
8. Bugatti is the manufacturing company with largest average price, with low highway MPG.