

## **Impact of Car Features on Price and Profitability**

**Project Description:** In this project, we shall work with the car features dataset and attempt to derive insights about the different car features that affect the manufacturer's suggested retail price (MSRP).

**Approach:** Download the dataset into our local repository. Clean the dataset by removing missing values and unwanted columns. Perform some exploratory data analysis (EDA) before attempting the tasks put forth. Finally, bring all the visualizations together to be presented in a dashboard.

**Tech-Stack Used:** Microsoft Excel, Tableau

### **Excel Workbook:**

<https://docs.google.com/spreadsheets/d/1XqRS79roTqLEVw31iqE87LXYSm48lBvd/edit?usp=sharing&ouid=105088911601746010535&rtpof=true&sd=true>

### **Tableau Dashboard:**

[https://public.tableau.com/views/ImpactofCarFeaturesonPriceandProfitability\\_16920053297150/CarPricesAnalysis?:language=en-GB&:display\\_count=n&:origin=viz\\_share\\_link](https://public.tableau.com/views/ImpactofCarFeaturesonPriceandProfitability_16920053297150/CarPricesAnalysis?:language=en-GB&:display_count=n&:origin=viz_share_link)

### **Problem Statement:**

How can a car manufacturer optimize pricing and product development decisions to maximize profitability while meeting consumer demand?

This problem could be approached by analyzing 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:**

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.

Link:

[https://drive.google.com/file/d/1HgHNOQx\\_wKnvApgBQ5DuKS1KlxmkDWPc/view?usp=share\\_link](https://drive.google.com/file/d/1HgHNOQx_wKnvApgBQ5DuKS1KlxmkDWPc/view?usp=share_link)

Before diving into the analysis of the given dataset, it is important to perform thorough data cleaning to ensure accurate and reliable results.

The dataset contains 11915 rows and 16 variables that describe the features of a car and its list price. On cleaning we have 11813 rows.

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

## Analysis

**Task 1:** How does the popularity of a car model vary across different market categories?

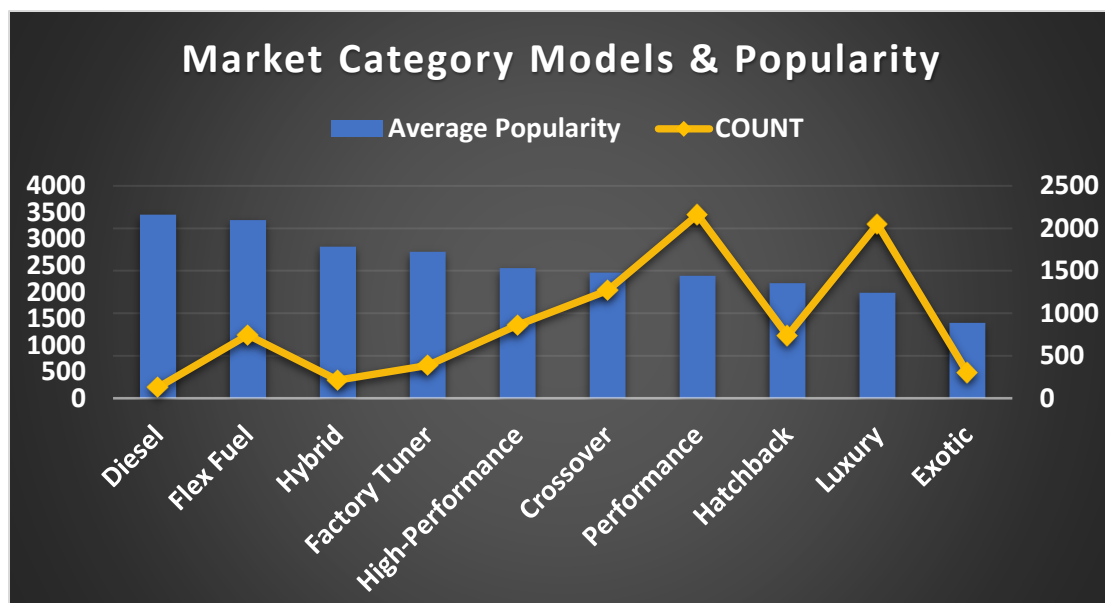
For this task, we will create a pivot table that shows the number of car models in each market category and their corresponding popularity scores.

We shall also create a combo chart that visualizes the relationship between market category and popularity.

Row Labels	Average of Popularity		Market Category	COUNT	Average Popularity
Crossover	1529		Diesel	205	2160
Crossover,Diesel	873		Flex Fuel	1191	2099
Crossover,Exotic,Luxury,High-Performance	238		Hybrid	343	1784
Crossover,Exotic,Luxury,Performance	238		Factory Tuner	618	1724
Crossover,Factory Tuner,Luxury,High-Performance	1823		High-Performance	1378	1531
Crossover,Factory Tuner,Luxury,Performance	2607		Crossover	2026	1478
Crossover,Factory Tuner,Performance	210		Performance	3456	1443
Crossover,Flex Fuel	2074		Hatchback	1181	1357
Crossover,Flex Fuel,Luxury	1173		Luxury	3279	1243
Crossover,Flex Fuel,Luxury,Performance	1624		Exotic	482	887
Crossover,Flex Fuel,Performance	5657				
Crossover,Hatchback	1676				
Crossover,Hatchback,Factory Tuner,Performance	2009				
Crossover,Hatchback,Luxury	204				
Crossover,Hatchback,Performance	2009				
Crossover,Hybrid	2563				
Crossover,Luxury	885				

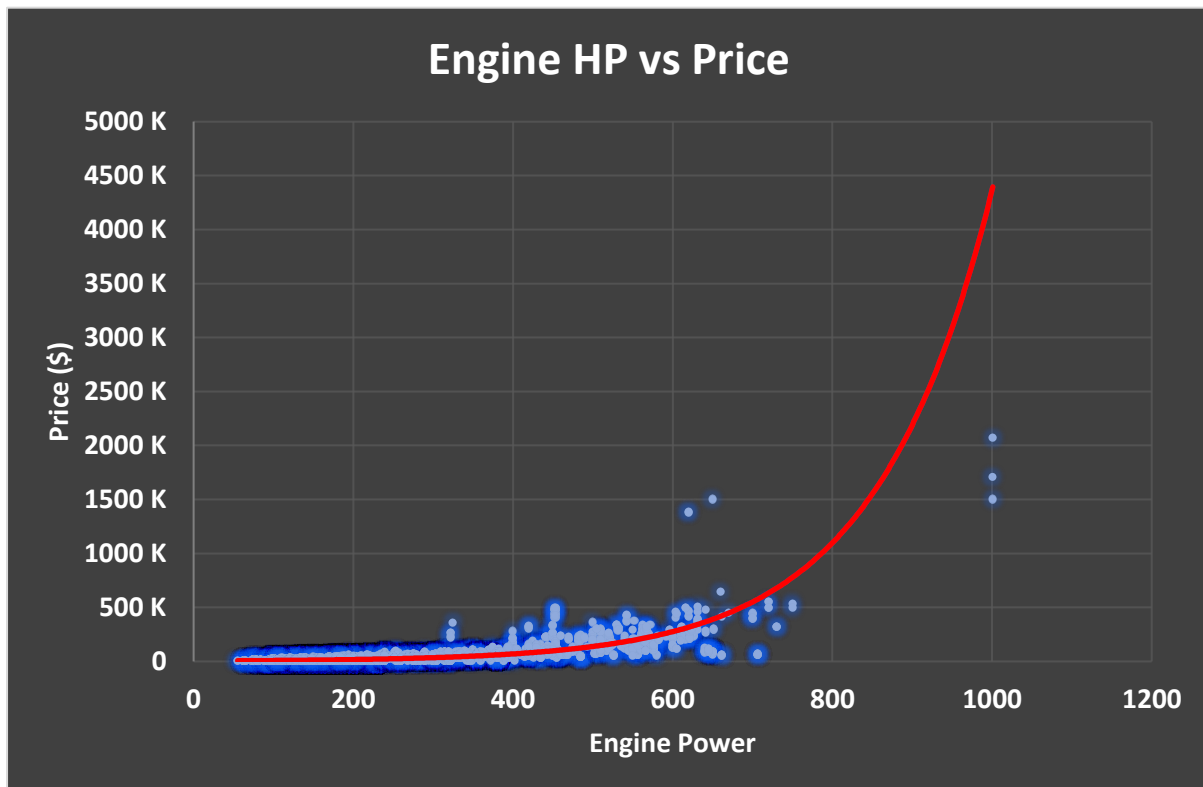
Used Excel functions such as COUNTIF and AVERAGEIF to split the joint market categories into individual ones and find their popularity.

**Insights:** As you can observe, Diesel has the highest popularity with 2160 average views.



**Task 2:** What is the relationship between a car's engine power and its price?

We shall create a scatter chart that plots engine power on the x-axis and price on the y-axis and add a trendline to the chart to visualize the relationship between these variables.

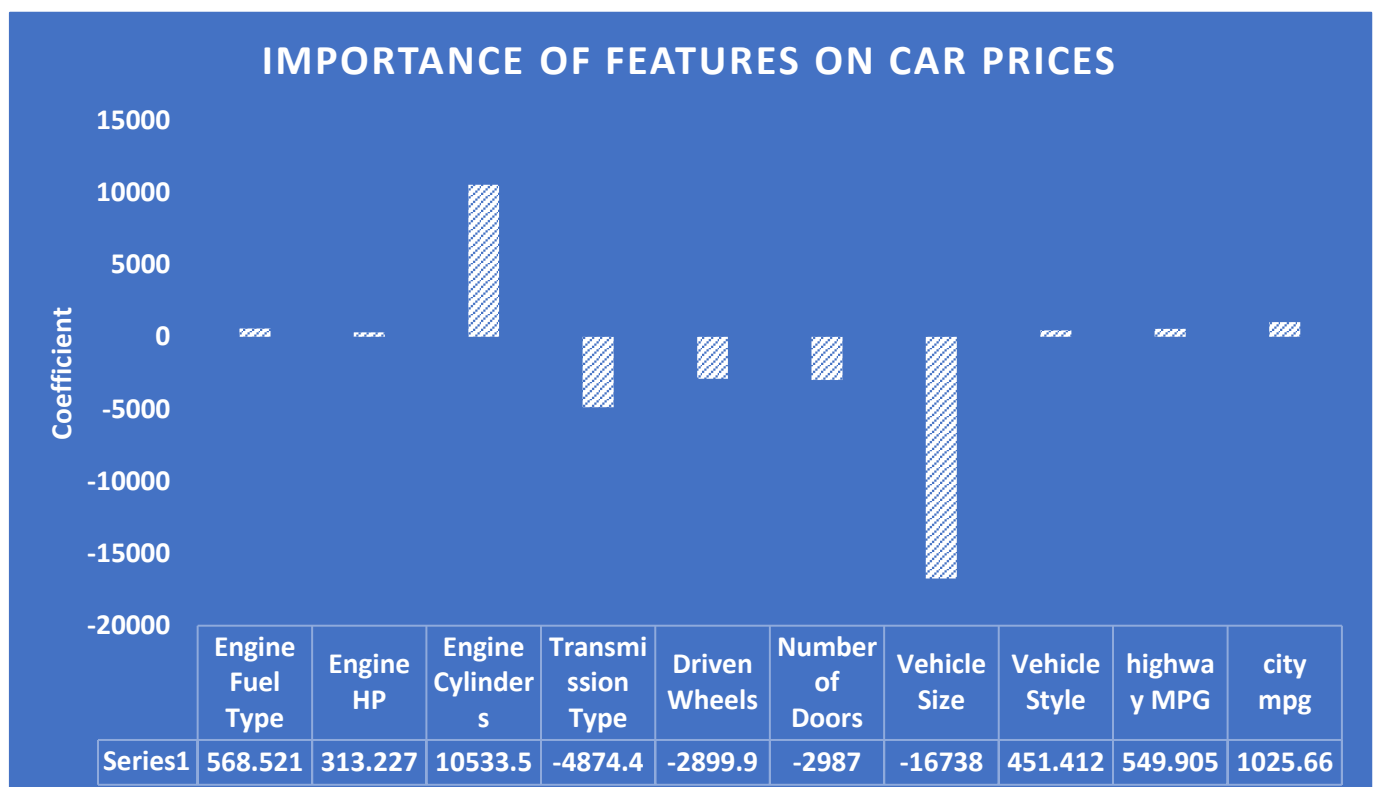


**Insights:** Notice clearly that there is a direct relationship between the Engine Horsepower with price.

### Task 3: Which car features are most important in determining a car's price?

For this task, we shall use **Regression Analysis** to identify the variables that have the strongest relationship with a car's price. Then we'll create a bar chart that shows the coefficient values for each variable to visualize their relative importance.

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.704849919							
R Square	0.496813408							
Adjusted R Square	0.496387015							
Standard Error	42781.05498							
Observations	11812							
ANOVA								
	df	SS	MS	F	Significance F			
Regression	10	2.13249E+13	2.13249E+12	1165.153269	0			
Residual	11801	2.15984E+13	1830218665					
Total	11811	4.29233E+13						
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	-79326.43616	5055.152454	-15.69219462	6.1349E-55	-89235.36921	-69417.50311	-89235.36921	-69417.50311
Engine Fuel Type	568.521242	279.7755719	2.032061763	0.042169756	20.11495023	1116.927534	20.11495023	1116.927534
Engine HP	313.2271598	6.102345055	51.32898205	0	301.2655565	325.1887632	301.2655565	325.1887632
Engine Cylinders	10533.52668	450.3784287	23.38816873	2.7023E-118	9650.710638	11416.34273	9650.710638	11416.34273
Transmission Type	-4874.422577	482.5787998	-10.1007806	6.85323E-24	-5820.356663	-3928.48849	-5820.356663	-3928.48849
Driven Wheels	-2899.911917	395.7892664	-7.326908945	2.50867E-13	-3675.724195	-2124.099638	-3675.724195	-2124.099638
Number of Doors	-2987.038008	530.3099997	-5.632626219	1.81545E-08	-4026.533124	-1947.542893	-4026.533124	-1947.542893
Vehicle Size	-16737.65054	652.7358405	-25.64230351	3.6277E-141	-18017.1205	-15458.18057	-18017.1205	-15458.18057
Vehicle Style	451.4124127	86.13560644	5.240717879	1.62712E-07	282.5724093	620.2524161	282.5724093	620.2524161
highway MPG	549.9052892	104.1667787	5.279085101	1.32134E-07	345.7212126	754.0893658	345.7212126	754.0893658
city mpg	1025.663211	119.5048074	8.582610466	1.04314E-17	791.4140667	1259.912355	791.4140667	1259.912355



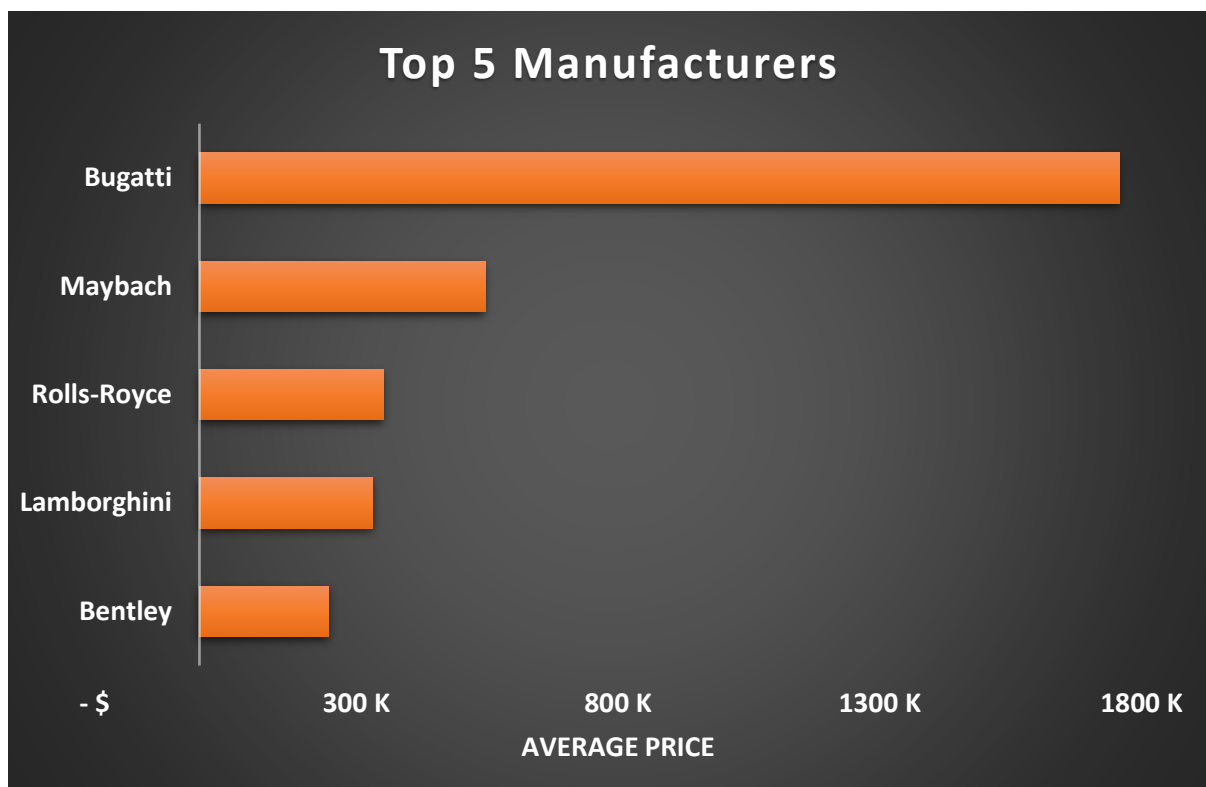
While vehicle size and transmission type have the most negative coefficient, number of engine cylinders has the highest positive coefficient with MSRP.

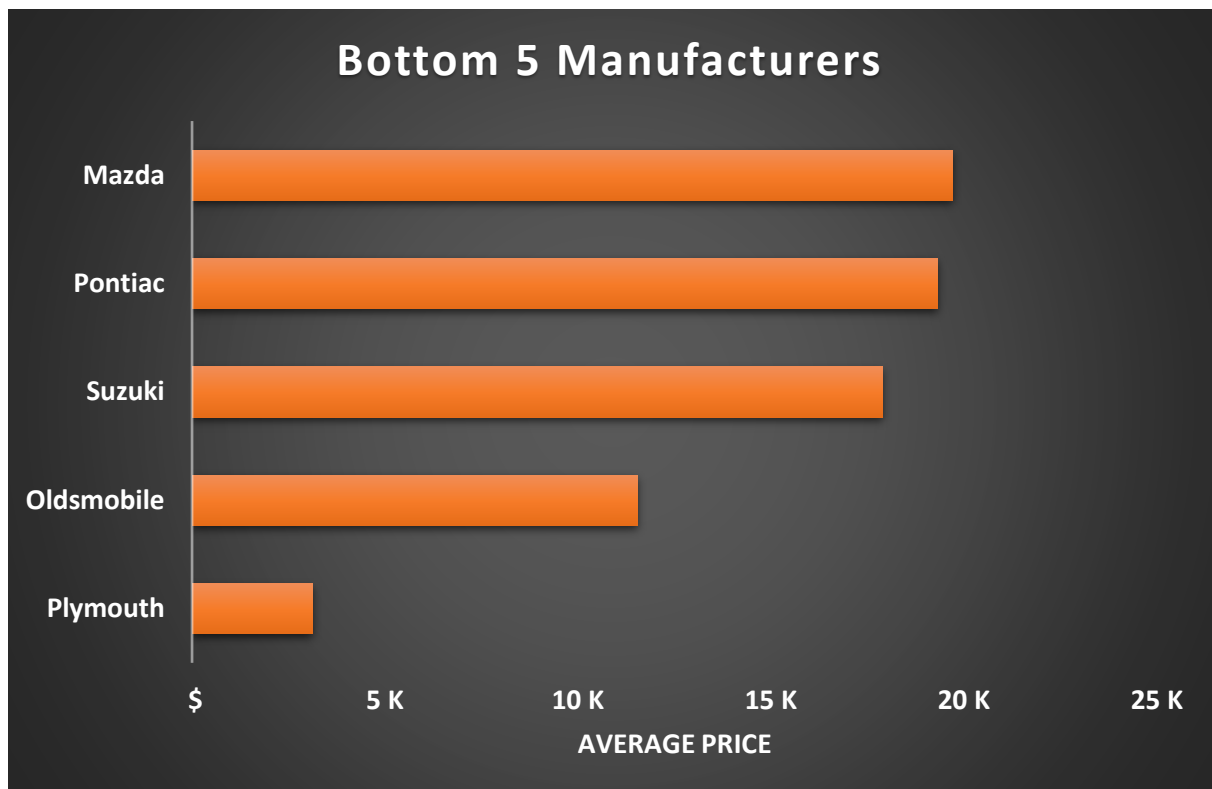
**Task 4:** How does the average price of a car vary across different manufacturers?

We'll first create a general pivot table that shows the average price of cars for each manufacturer. Then, we'll filter out the top 5 and bottom 5 manufacturer before plotting a bar chart that visualizes the relationship between manufacturer and average price.

Top 5	
Manufacturer	Average Price
Bentley	247 K
Lamborghini	332 K
Rolls-Royce	351 K
Maybach	546 K
Bugatti	1757 K
Grand Total	343 K

Bottom 5	
Manufacturer	Average Price
Plymouth	3 K
Oldsmobile	12 K
Suzuki	18 K
Pontiac	19 K
Mazda	20 K
Grand Total	17 K





From the first chart or pivot table, it is clear that Bugatti is the most expensive car by a good amount, having an average price of \$1.8M.

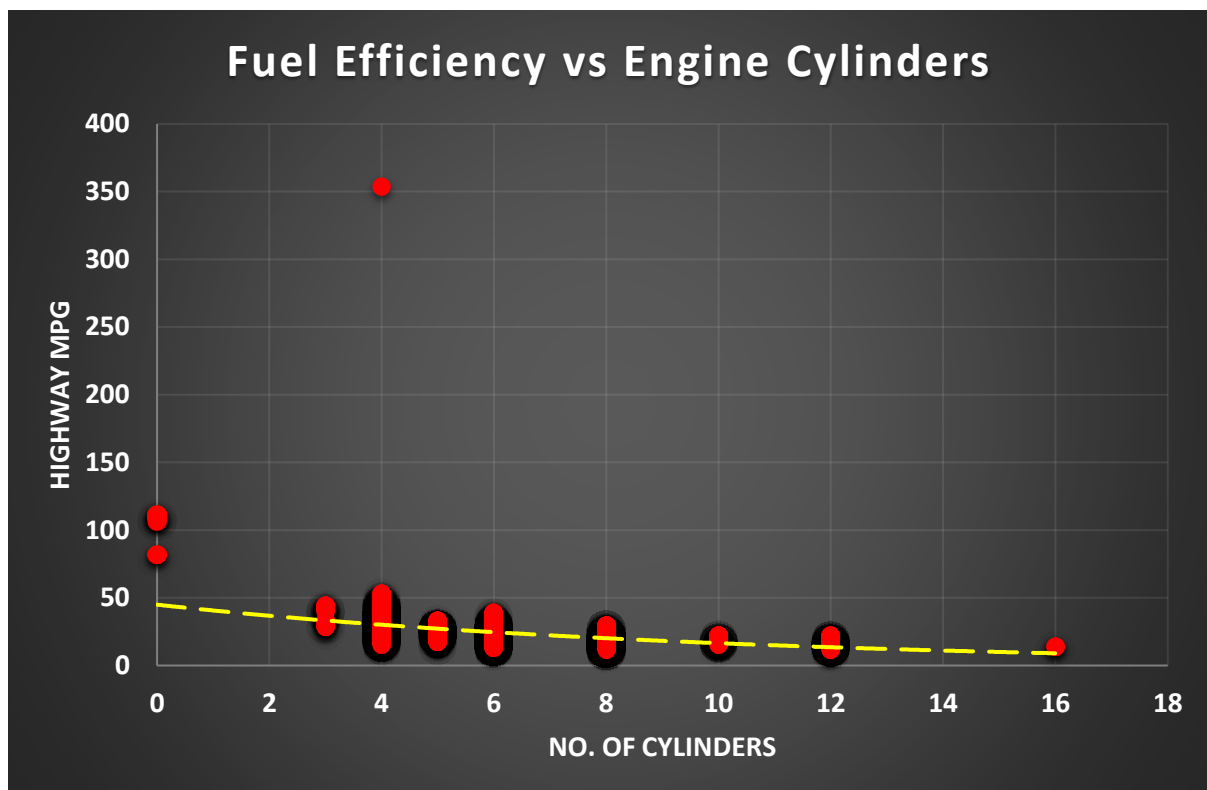
Plymouth, on the other hand, is the cheapest of them all (\$4K)

**Task 5:** What is the relationship between fuel efficiency and the number of cylinders in a car's engine?

Fuel efficiency is measured through MPG.

We'll 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.

Finally, calculate the correlation coefficient between the number of cylinders and highway MPG to quantify the strength and direction of the relationship.



Observe that there are some outliers (e.g. [4 cylinders, 350 MPG], [0 cylinders, 100 MPG]).

We can clearly notice an inverse relationship between highway MPG and the number of cylinders in a car.

We calculated the correlation coefficient between the two variables and it turned out to be negative just as the graph suggested

Correlation Coefficient
-0.620312551



## **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?

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

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

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

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

All these questions have been answered through visualizations made on Tableau.

## **Closing Note:**

In this project, I undertook a comprehensive data analysis approach to determine the relationship between various features of a car and its price as well as the natural relationship among the independent variables. By performing regression analysis, I found that engine horsepower and model year of a car have a strong correlation with its MSRP while size and fuel transmission type had inverse relationships. I also determined the most popular market categories that car brands should focus on.

By leveraging Excel pivot tables and Tableau visualizations, I derived valuable insights about the car features dataset that will undoubtedly benefit a car manufacturer to optimize its pricing and product development decisions to maximize profitability, while simultaneously meeting consumer demand.