

## **Project 7 : By Ramana Bansal**

### **Car Features Case Study**

Worksheet Link:

<https://docs.google.com/spreadsheets/d/1oL4NCyzQplWUS3QRmcGUSUAC-H1b5K3i/edit?usp=sharing&ouid=108908340523953886832&rtpof=true&sd=true>

Video Link:

<https://drive.google.com/file/d/1peJJIP23yyeGDcVQqSTtI5K33HGgJ2Ix/view?usp=sharing>

#### **Project Description:**

The aim of the project is to aid the manufacturer in developing a pricing strategy that balances ever changing consumer demands 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.

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

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:** 11,199

**Number of variables:** 16

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

### **Approach:**

The data was formatted and cleaned as required. A set of questions that might aid in analysis were drafted. MS Excel 2010 was used to clean, analyze and visualize the data provided.

### **Tech-Stack Used:**

MS Excel 2010 was used to perform the analysis.

### **Insights:**

1. The size, number of engine cylinders, transmission type and city mpg are some of the features that impact price the most.
2. Hatchback, Flex Fuel; Flex Fuel, Diesel; Crossover, Flex Fuel, Performance were some of the most popular categories. Flex Fuel, Hybrid; Performance, Hybrid; Flex Fuel, Performance, Hybrid; Exotic, Luxury were some of the least popular categories.
3. The trendline as well as correlation statistic depict a negative relationship between Number of engine cylinders and Highway MPG.
4. The fuel efficiency has improved for most vehicles over the years, except for Cargo van and Convertible SUVs.
5. The MSRP seems to increase linearly with Engine Horsepower. The plot shows that Engine Horsepower between 400 and 700 Hp has the highest number of cars with higher MSRP. However, the highest MSRP is for cars with Engine Hp of 1001 Hp, but those models are very few in number.
6. Coupes with Automatic manual transmission types had the highest MSRPs, followed by Automatic Convertibles.
7. Some of the most expensive cars were Coupes by Ferrari and Lamborghini, Sedans by Cadillac and BMW, Convertibles by Aston Martin and Lamborghini and 4dr SUV by Land Rover and GMC.

8. Buggati has the most expensive car with highest horsepower, but with abysmal Highway mpg. Plenty of cars have high mpg and low price, but below average engine horsepower. There are a few cars that lie in the middle, with average engine horsepower, average highway mpg and average prices.
9. The lowest price of the cars tended to be \$2000, which could not be taken into account since that was the price for initial years i.e. during 1990s.

## A. Cleaning the data

The data was formatted into a table for ease of handling. The datatypes of certain values/columns were changed as required.

### 1. Removing irrelevant columns/data

The data was well formatted and there weren't any irrelevant columns.

### 2. Handling Typos and Inaccurate Values

The car prices may seem inaccurate at first glance, however, they correspond to the years the cars were made available.

### 3. Handling Duplicate Values

This step involves identifying and removing duplicate values in the data. Duplicate values can cause errors and distort the analysis results. 715 duplicates were removed. All the columns were used to find duplicates. There were 11199 unique rows.

E	F	G	H	I	J
Engine HP	Engine Cylinders	Transmission Type	Driven_Wheels	Number of Doors	Market Category
335	6	MANUAL	rear wheel drive	2	Factory Tuner,
300	6	MANUAL	rear wheel drive	2	Luxury,Perform
300	6	MANUAL	rear wheel drive	2	Luxury,High-P
230	6	MANUAL	rear wheel drive	2	Luxury,Perform
230	6	MANUAL	rear wheel drive	2	Luxury
230	6	MANUAL	rear wheel drive	2	Luxury,Perform
300				2	Luxury,Perform
300				2	Luxury,High-P
230				2	Luxury
230				2	Luxury
300				2	Luxury,High-P
230				2	Luxury,Perform
300				2	Luxury,Perform
230	6	MANUAL	rear wheel drive	2	Luxury
320	6	MANUAL	rear wheel drive	2	Luxury,High-P
320	6	MANUAL	rear wheel drive	2	Luxury,High-P
172	6	MANUAL	front wheel drive	4	Luxury
172	6	AUTOMATIC	all wheel drive	4	Luxury
172	6	MANUAL	all wheel drive	4	Luxury

#### 4. Checking for missing data

There are total 11199 rows after duplicate removal. The number of blanks, unknowns or N/A in columns are as follows:

<b>Column</b>	<b>Blanks</b>	<b>Type</b>
Engine Fuel Type	3	Categorical
Engine HP	69	Numeric
Engine Cylinders	30	Numeric
Transmission	12	Categorical
Driven_Wheels	0	Categorical
Number of Doors	6	Numeric
Market Category	3376	Categorical
Vehicle Size	0	Categorical
Vehicle Style	0	Categorical
highway MPG	0	Numeric
city mpg	0	Numeric
Popularity	0	Numeric
MSRP	0	Numeric

The blank rows of columns with blank cells < 100 were deleted using Filter and Go To. The N/A in Market Category were replaced by Blanks since the number of N/A was high and it was a significant column.

There were 11085 rows left after data cleaning.

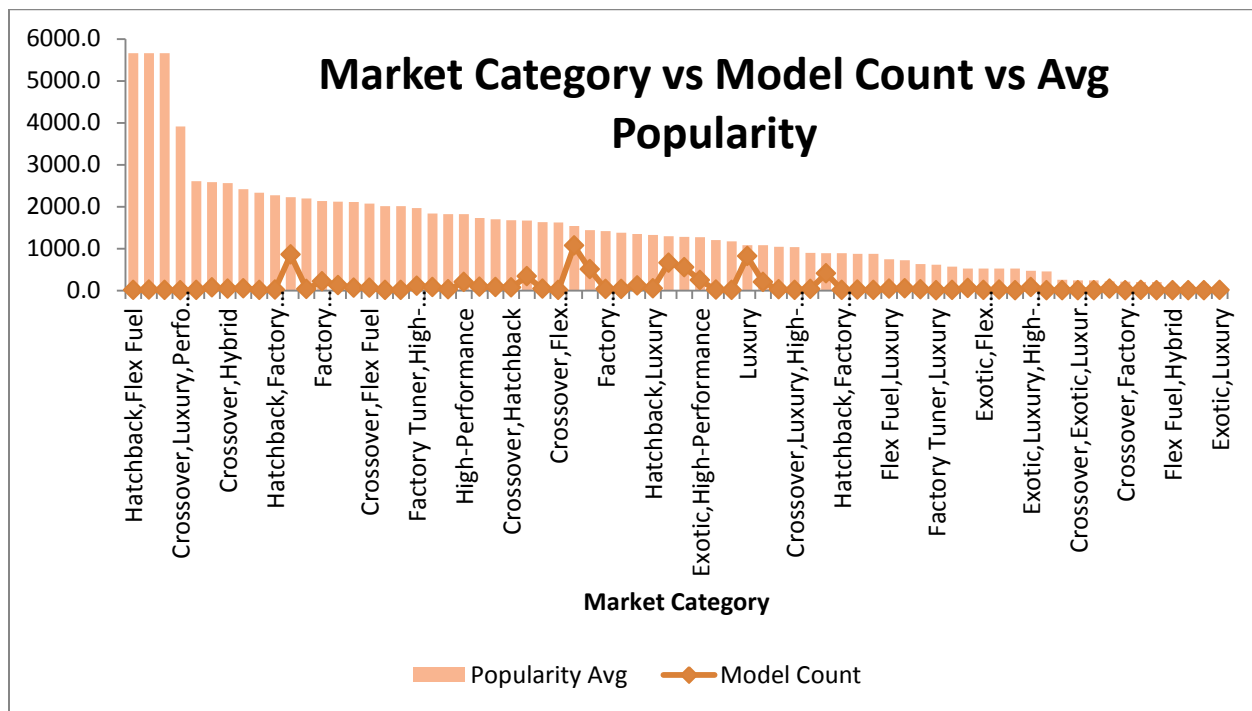
## Tasks: Analysis

**Task1:** How does the popularity of a car model vary across different market categories?

The number of N/A was high for Market Category Column. These were removed before starting with Analysis. 7732 rows were used for this particular task.

A pivot table was created with Market Category in the Row Labels and Count of Models and Average of Popularity in Values. The pivot table was sorted as per Average Popularity.

A combo chart with Columns depicting average Popularity and Line chart showing Model count for each Market Category was created using pivot table.

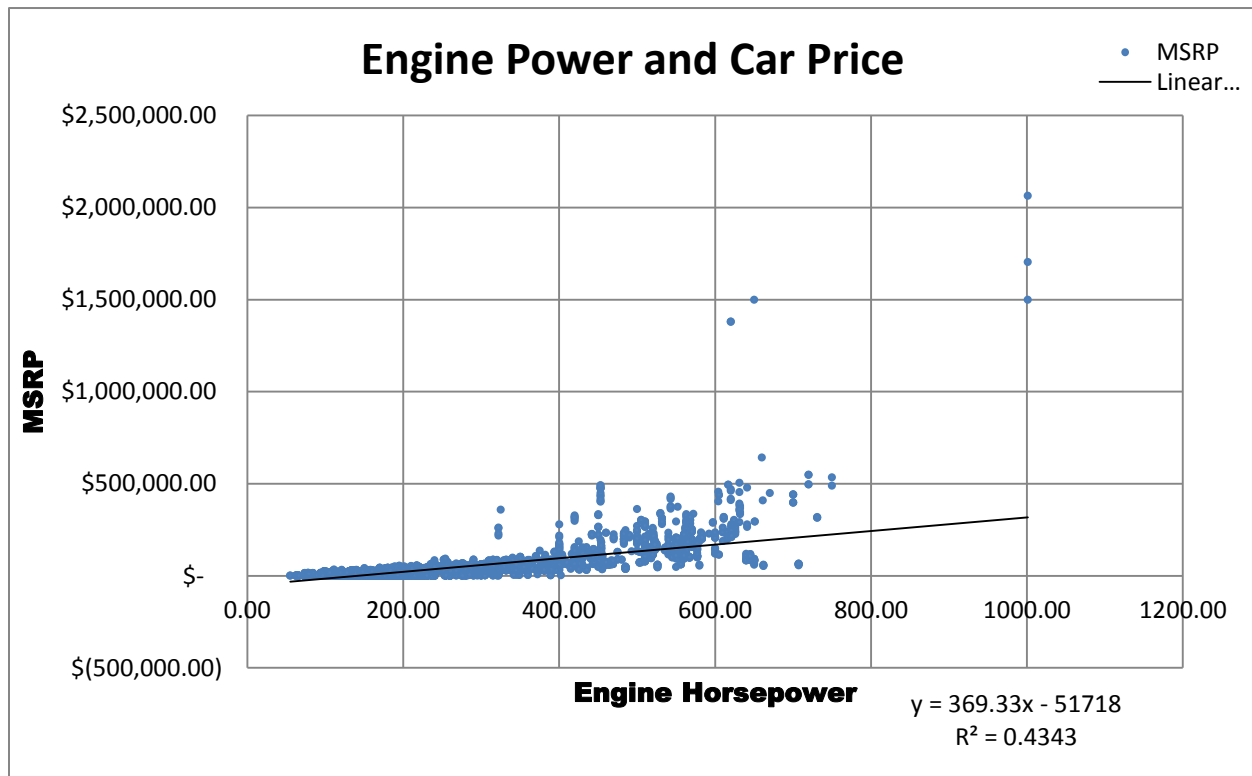


### **Insights:**

1. Popular Market Categories: Hatchback, Flex Fuel; Flex Fuel, Diesel; Crossover, Flex Fuel, Performance.
2. Least popular Market Category: Flex Fuel, Hybrid; Performance, Hybrid; Flex Fuel, Performance, Hybrid; Exotic, Luxury
3. Crossover type has the highest number of Models at 1068 and Flex Fuel has second highest at 855.

**Task2:** What is the relationship between a car's engine power and its price?

A scatter chart was plotted with engine power on the x-axis and price on the y-axis. A trendline was added to the chart to visualize the relationship between these variables.



**Insights:**

The MSRP seems to increase linearly with Engine Horsepower.

The plot shows that Engine Horsepower between 400 and 700 Hp has the highest number of cars with higher MSRP.

The highest MSRP is for cars with Engine Hp of 1001 Hp, but those models are very few in number.

**Task3:** Which car features are most important in determining a car's price?

Regression analysis was used to identify the variables that have the strongest relationship with a car's price. We can find Regression between Numerical values only, therefore categorical values were assigned Numerical values for Analysis.

No. of observations: 11085

Dependent variable: MSRP

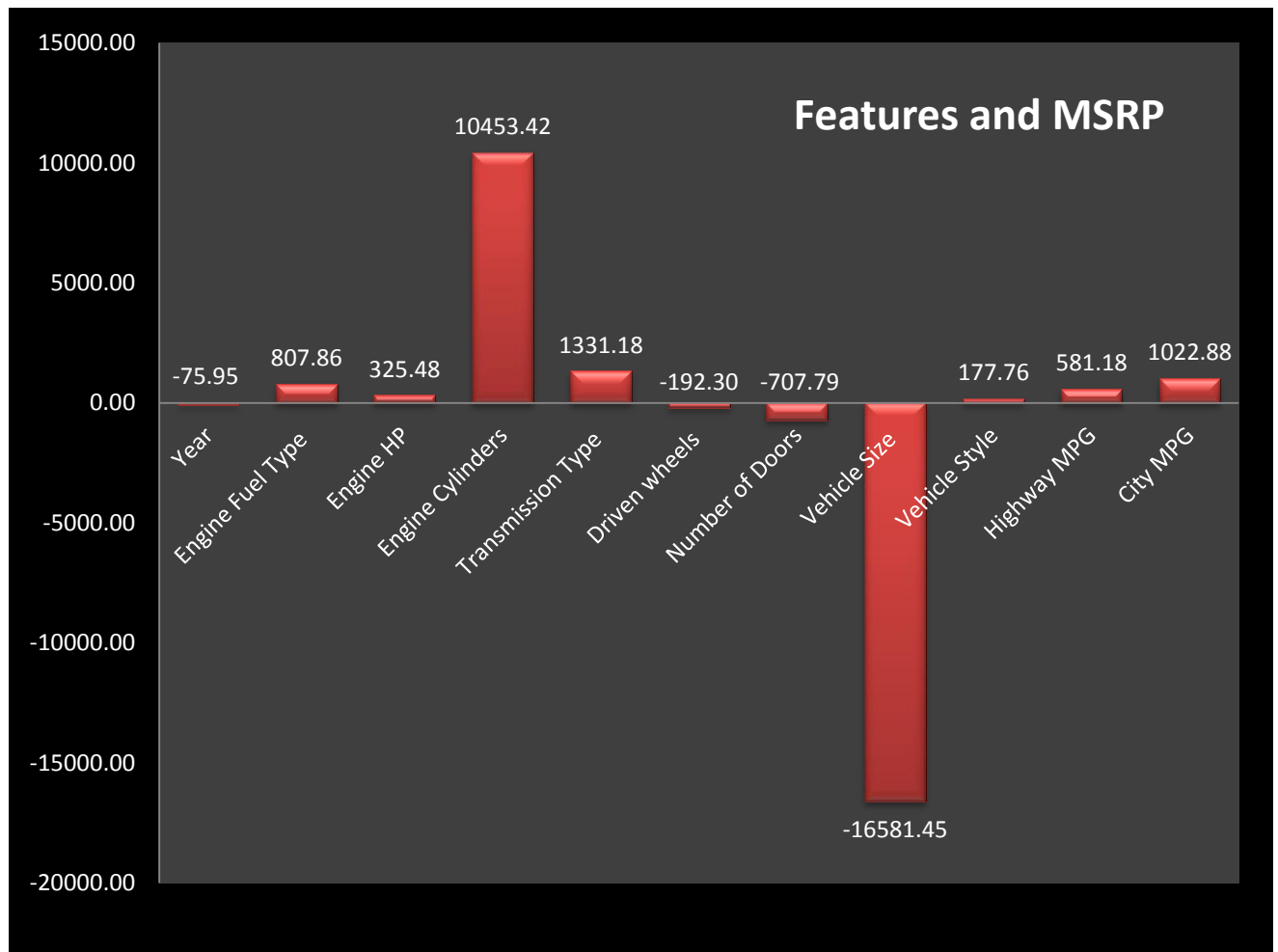
Independent variables: Year, Engine Fuel Type, Engine HP, Engine Cylinders, Transmission Type, Driven wheels, Number of Doors, Vehicle Size, Vehicle Style, highway MPG, city mpg.

1. The regression model explains about 48.7% of data i.e. about 48.7% of MSRP values are explained by selected Independent variables. We use Adjusted R square value for estimating the same in case of Multiple Regression.
2. The regression model obtained is reliable at alpha levels 0.05 and 0.01. We use Significance F to estimate the significance.
3. We use p-values of t stats for each variable/feature to understand if it's a good feature to estimate our MSRP. If p-value is greater than 0.05, then the feature/ variable is not a good coefficient for determining our MSRP.
4. Based on point 3, the coefficients of Year, Engine Fuel Type and Driven wheels are not a good fit at  $\alpha = 0.05$ . However, the coefficients of Engine Fuel Type and Driven wheels could be a good fit at  $\alpha = 0.1$ . Since the rest of the features have p-values  $< 0.05$ , their coefficients are considered for Analysis.
5. The coefficients of Year, Engine Fuel Type, Driven wheels, Number of doors and Vehicle Style are not a good fit at  $\alpha = 0.05$ . However, the coefficients of Engine Fuel Type and Vehicle Style could be a good fit at  $\alpha = 0.1$ .

Feature	Coefficient	Relation with MSRP
Year	-75.95	Negative Relation with MSRP
Engine Fuel Type	807.86	Positive Relation with MSRP
Engine HP	325.48	Positive Relation with MSRP
Engine Cylinders	10453.42	Positive Relation with MSRP
Transmission Type	1331.18	Positive Relation with MSRP
Driven wheels	-192.30	Negative Relation with MSRP
Number of Doors	-707.79	Negative Relation with MSRP
Vehicle Size	-16581.45	Negative Relation with MSRP
Vehicle Style	177.76	Positive Relation with MSRP
Highway MPG	581.18	Positive Relation with MSRP
City MPG	1022.88	Positive Relation with MSRP



A bar chart was created that shows the coefficient values for each variable to visualize their relative importance.



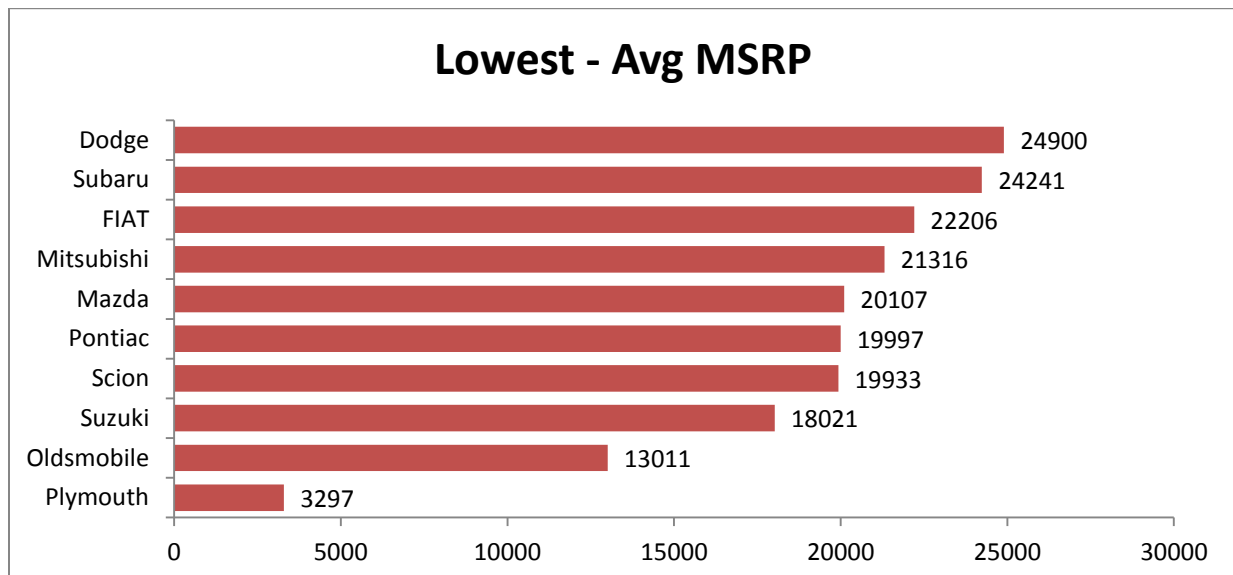
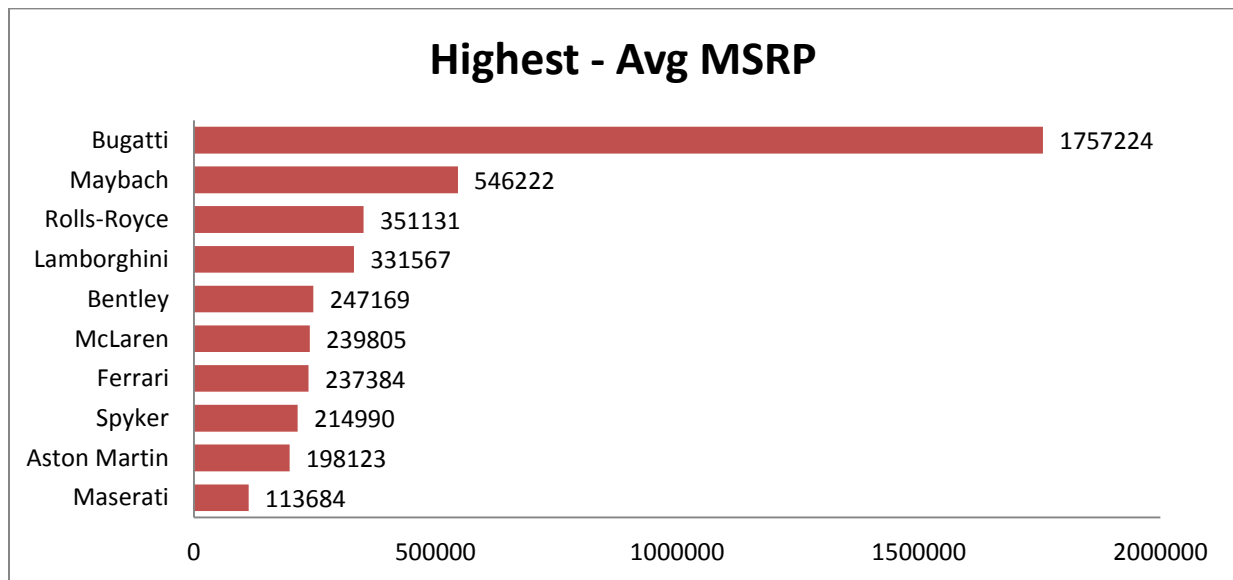
### Insights:

The size, number of engine cylinders, transmission type and city mpg are some of the features that impact price the most.

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

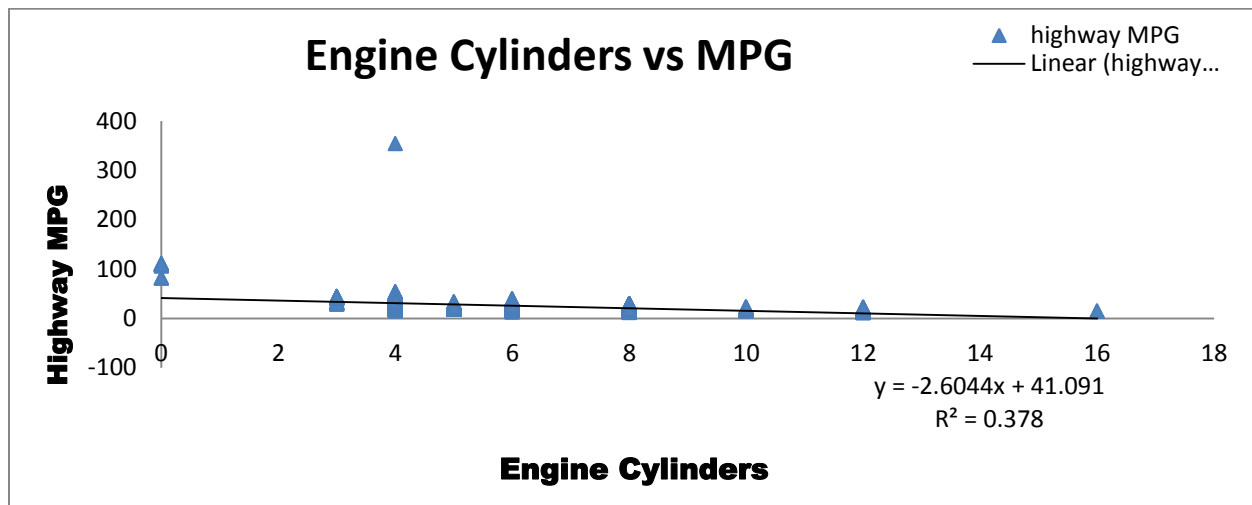
A pivot table was created with Make in the Row Labels and Average of MSRP in Values. The pivot table was sorted as per Average MSRP.

Two bar charts were created to visualize the relationships between manufacturer and Highest MSRPs and Lowest MSRPs.



**Task5:** What is the relationship between fuel efficiency and the number of cylinders in a car's engine?

A scatter plot was created with the number of cylinders on the x-axis and highway MPG on the y-axis. A trendline was added on the scatter plot to visually estimate the slope of the relationship and assess its significance.



The correlation coefficient between the number of cylinders and highway MPG was calculated to quantify the strength and direction of the relationship.

**Correlation: -0.614855443**

There is a negative correlation between the number of Engine Cylinders and highway MPG, signifying an inverse relationship.

### Insights:

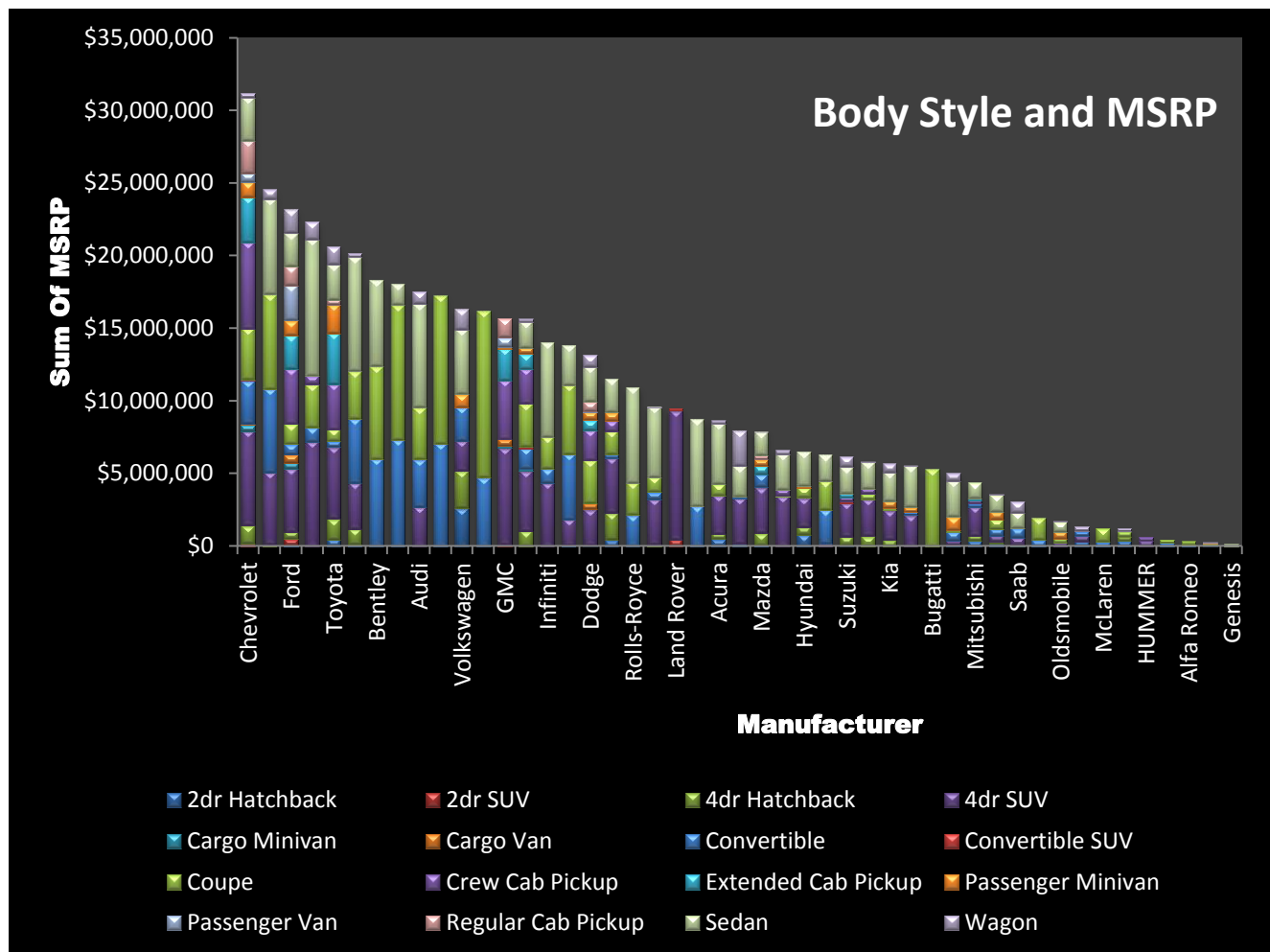
The trendline as well as correlation statistic depict a negative relationship between Number of engine cylinders and Highway MPG.

## **Building the Dashboard:**

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

A pivot table was created with Make in the Row Labels, Vehicle style in Column Labels, and Sum of MSRP in Values. The pivot table was sorted as per Sum MSRP.

A stacked column chart was created to show the distribution of car prices by brand and body style. Filters and slicers were added to make the chart interactive.



### **Insights:**

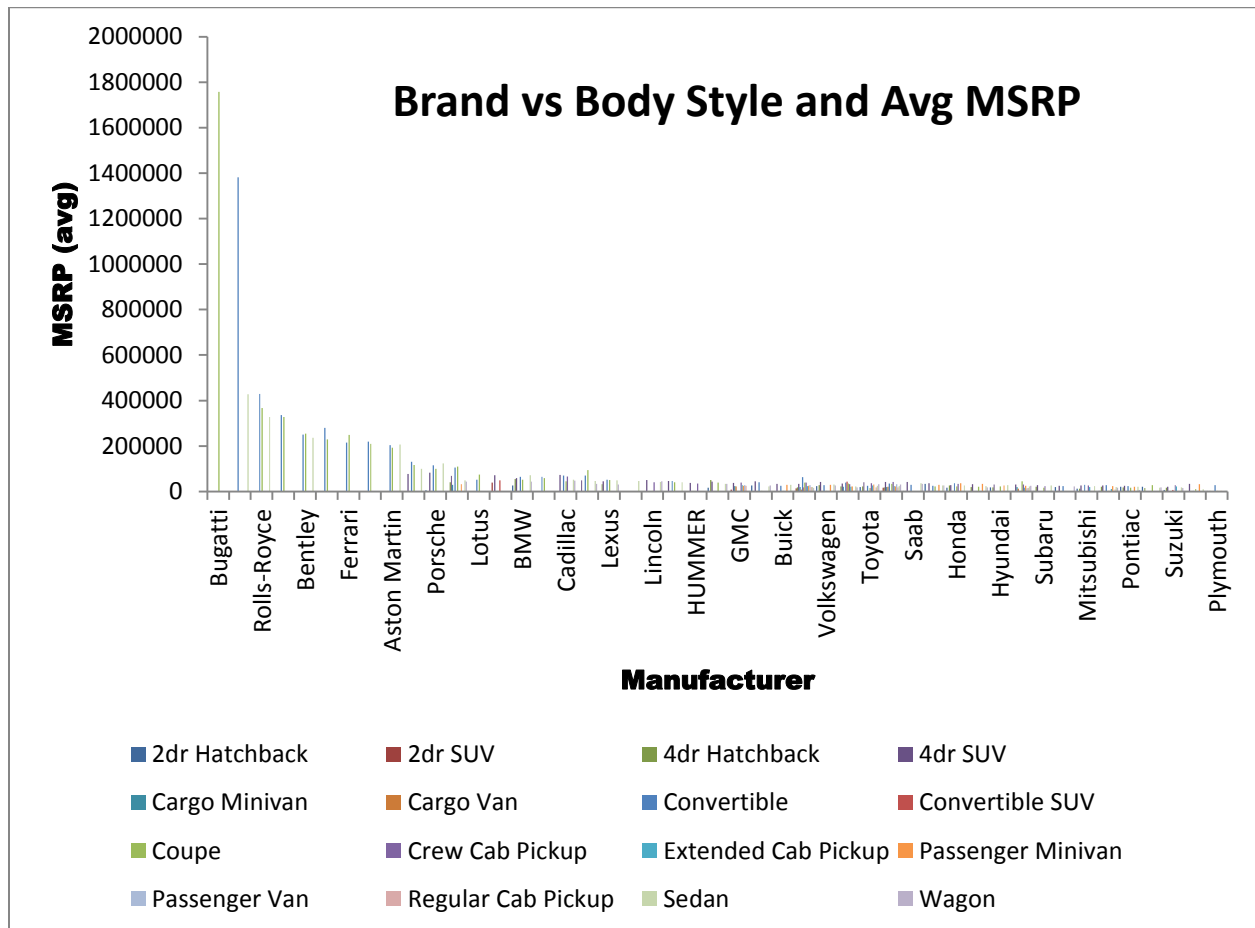
#### **Most expensive cars:**

1. Coupes by Ferrari and Lamborghini
2. Sedans by Cadillac and BMW
3. Convertibles by Aston Martin and Lamborghini
4. 4dr SUV by Land Rover and GMC

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

A pivot table was created with Make in the Row Labels, Vehicle style in Column Labels, and Average of MSRP in Values. The pivot table was sorted as per Average MSRP.

A Clustered column chart was created to compare the average MSRPs across different car brands and body styles.

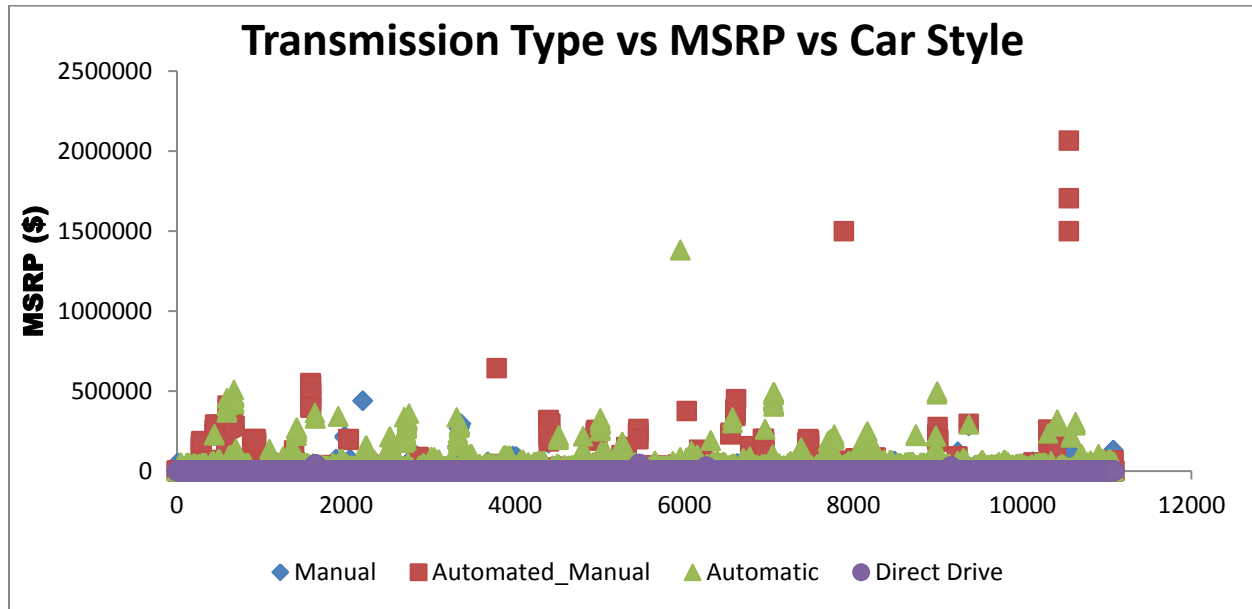


### Insights:

The most expensive cars are Coupes by Bugatti, Rolls-Royce and Lamborghini, and Convertibles by Maybach, Rolls-Royce and Lamborghini.

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

A Scatter plot chart was created to visualize the relationship between MSRP and transmission type, with different symbols for transmission type. 4 separate columns were created to record MSRP (y-value) for each Transmission Type. Then, each column was selected as a separate series for separate color.



A pivot table was also created with Vehicle Style in the Row Labels, Transmission type in Column Labels, and Average of MSRP in Values.

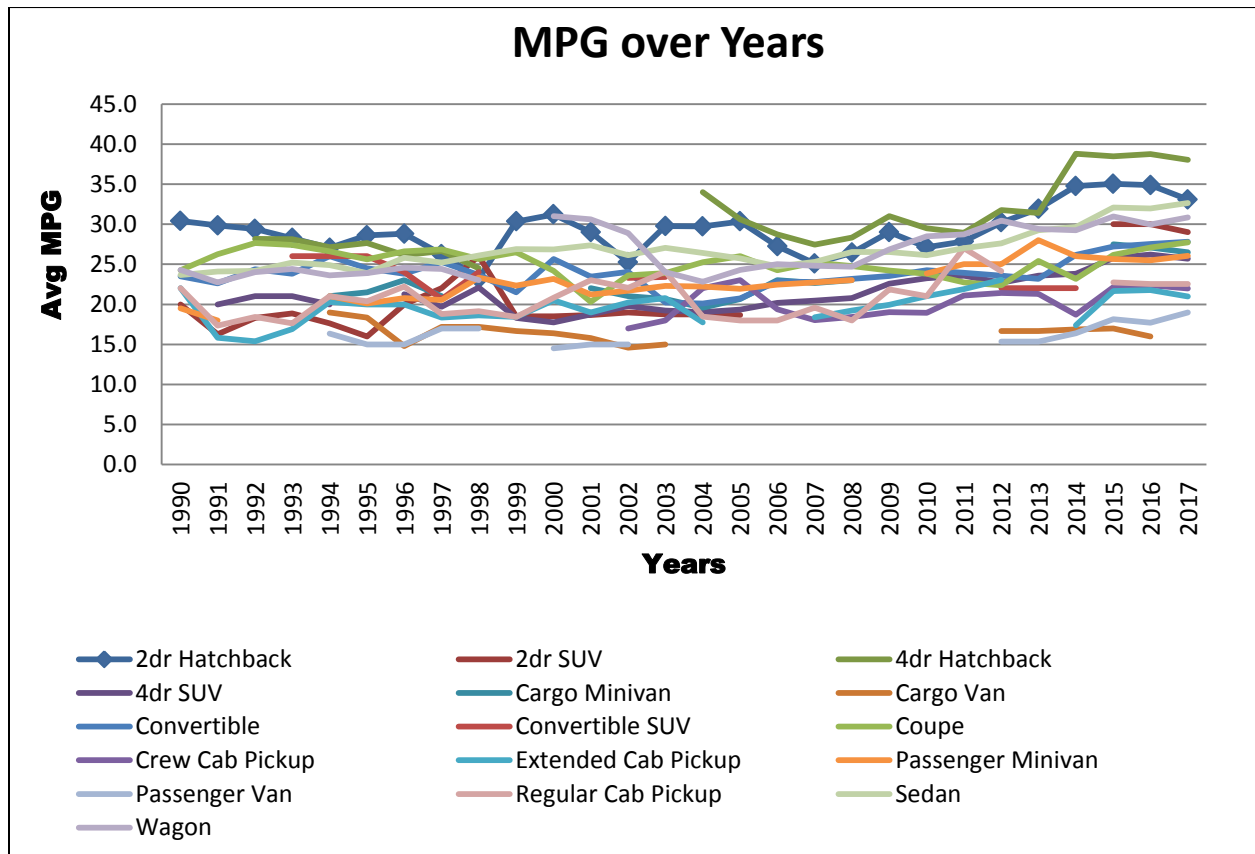
### Insights:

Coupes with Automatic manual transmission types had the highest MSRPs, followed by Automatic Convertibles.

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

A pivot table was also created with Years in Row labels, Vehicle Style in the Column Labels and Average Highway mpg in Values.

Line chart with Filters for Vehicle style was created to show the trend of fuel efficiency (MPG) over time for each body style.



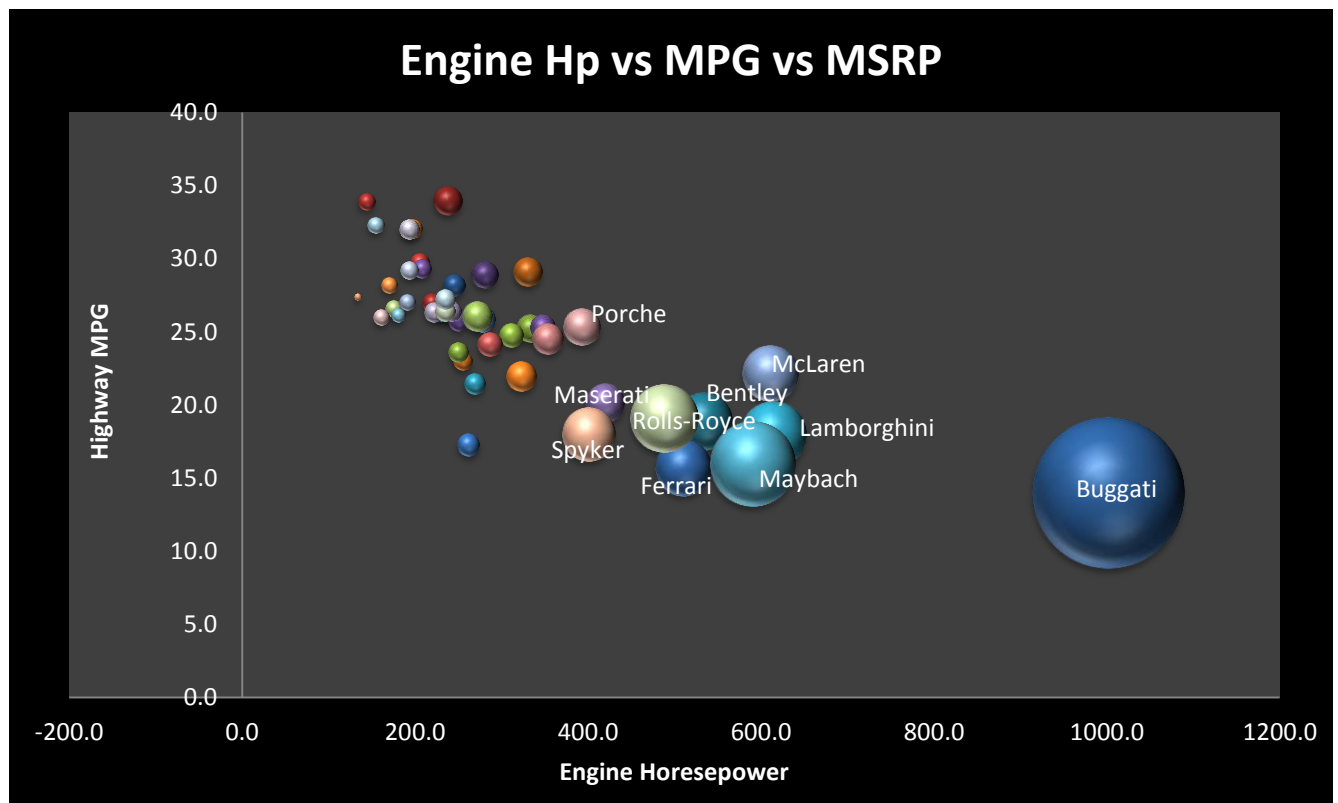
### Insights:

The fuel efficiency has improved for most vehicles over the years, except for Cargo van and Convertible SUVs.

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

A pivot table was created with Make (Manufacturer) in Row labels and Engine Horsepower (Average), Highway mpg (Average) and MSRP (Average) in Values.

A Bubble chart was created to visualize the relationship between horsepower, MPG, and price across different car brands. The chart had Engine horsepower on x-axis and Highway mpg on y-axis. The size of bubbles indicated the price of car while the color indicated different brands.



### Insights:

Bugatti has the most expensive car with highest horsepower, but with abysmal Highway mpg. Plenty of cars have high mpg and low price, but below average engine horsepower. There are a few cars that lie in the middle, with average engine horsepower, average highway mpg and average prices.



A **dashboard** was created with charts from dashboard tasks 1 to 5. Slicers were added for easy filtration of data to be viewed.

