

**Analyzing the Impact of Car Features on Price and Profitability**

# Project Description:

- For the given dataset, the client has asked how can a car manufacturer optimize pricing and product development decisions to maximize profitability while meeting consumer demand.
- The business 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.
- The dataset used in the project 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), and is titled "Car Features and MSRP". It was collected and made publicly available on Kaggle.
- Dataset link : [car data](#)

## Approach:

### Data Preprocessing:

Cleaning the data and handling the missing information in the dataset.

### Data Analysis:

Created Visualizations using Pivot Tables and performed Regression Analysis to determine what factors influence the price of a car.

### Building Dashboard:

Created Interactive Dashboard comprising of Slicers to select across different features to further drill through the features of data.

## Tech Stack Used:

### 1. MS Excel 2019

- Used MS Excel in this project for data cleaning, data analysis and visualizations.



### 2. MS PowerPoint 2019

- Used MS PowerPoint to prepare the project report and presentation.



# Data Preprocessing:

## Steps involved in data preprocessing:

The Market Category column containing N/A row labels are dropped and added one column named average MPG based on average of highway MPG and city MPG.

The blank cells in the Engine HP column are imputed based on research on google.

For Engine Cylinder column, the blank cells are imputed as '0' based on other Electric Fuel Type cars and also imputed as '4 for some blank cells based on other Tesla Model S Engine cylinders.

The blank cells in Number of Doors column are imputed based on other same Make and Model column row labels.

The 'Unknown' row labels in Transmission Type column is replaced by 'Automatic' row labels based on Make, Model and Engine Cylinders column row labels.



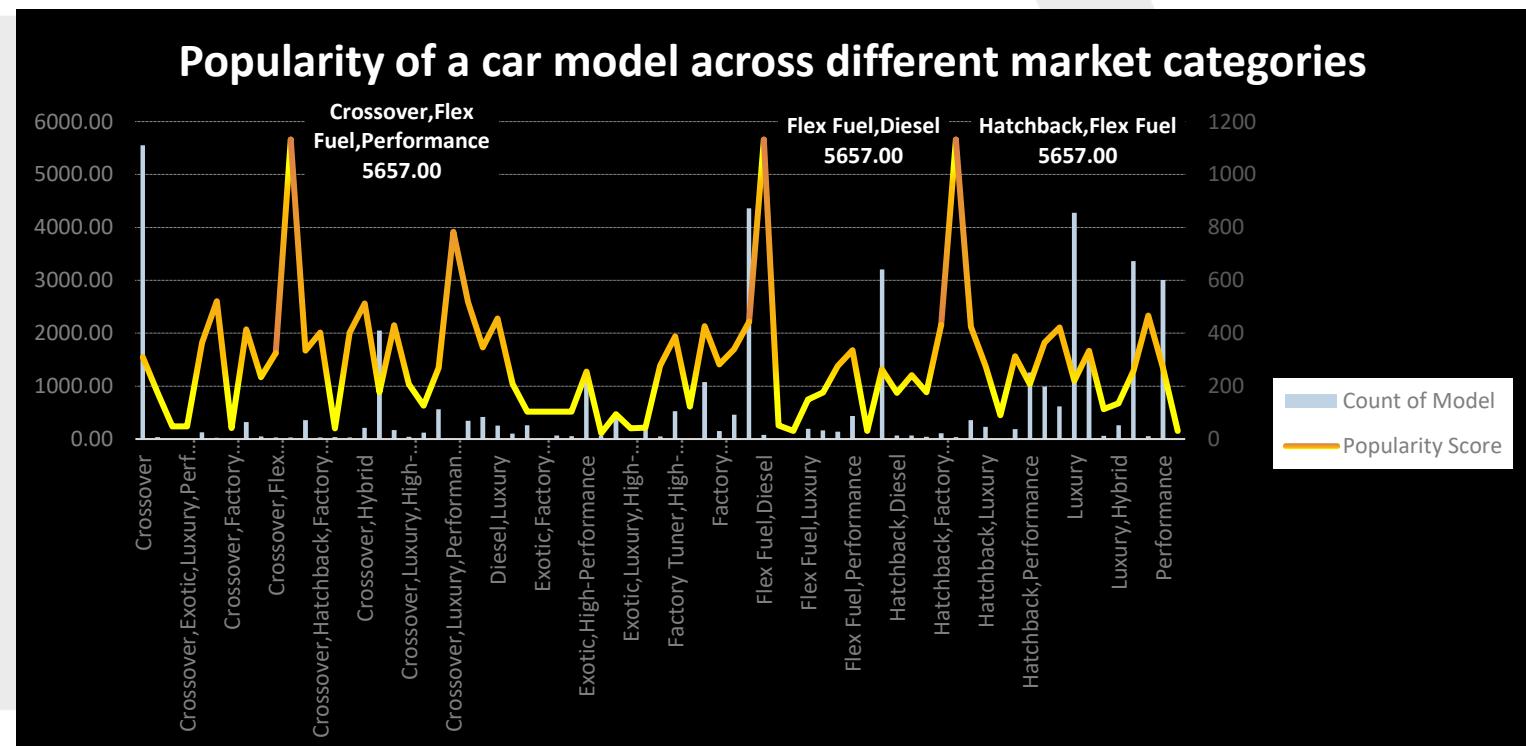
# Data Analysis tasks:

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

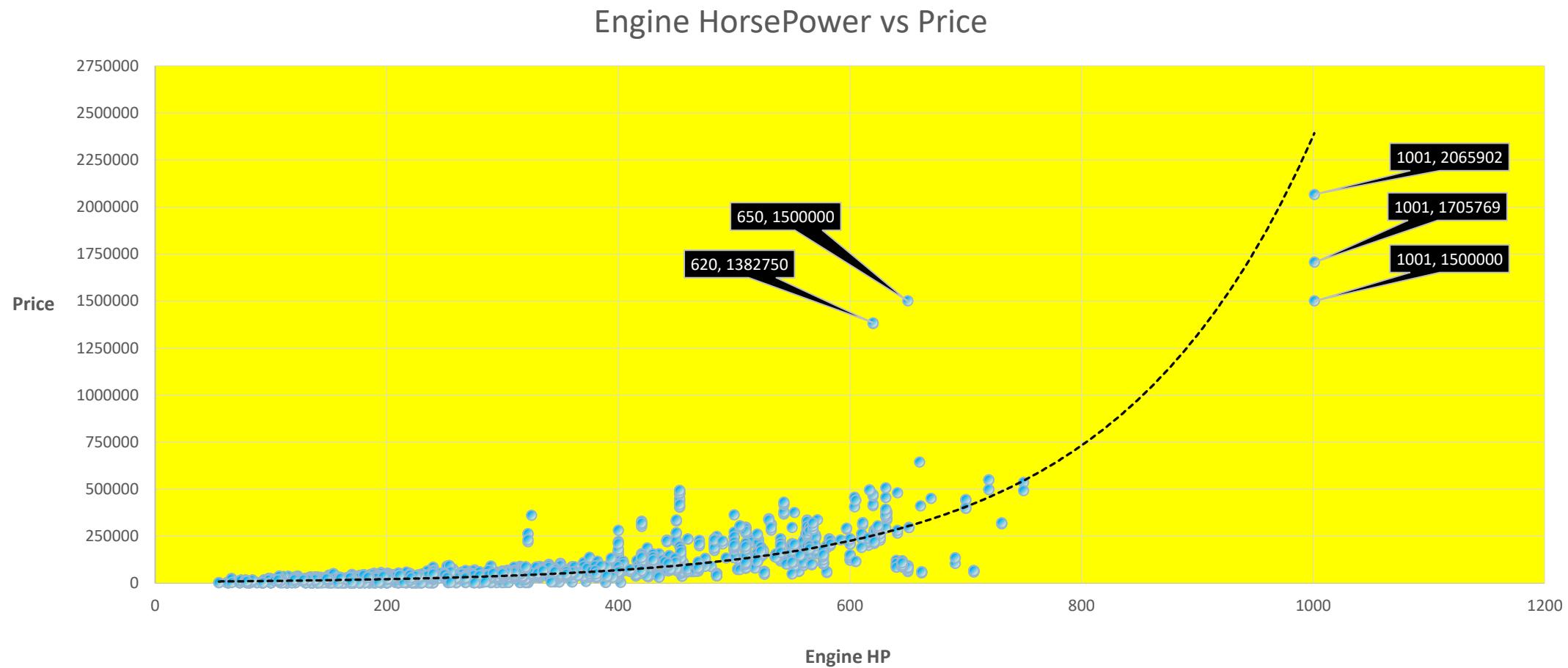
Market Category	Count of Model	Popularity Score			
Crossover	1110	1545.26	Crossover,Performance	69	2585.96
Crossover,Diesel	7	873.00	Diesel	84	1730.90
Crossover,Exotic,Luxury,High-Performance	1	238.00	Diesel,Luxury	51	2275.00
Crossover,Exotic,Luxury,Performance	1	238.00	Exotic,Factory Tuner,High-Performance	21	1046.38
Crossover,Factory Tuner,Luxury,High-Performance	26	1823.46	Exotic,Factory Tuner,Luxury,High-Performance	52	517.54
Crossover,Factory Tuner,Luxury,Performance	5	2607.40	Exotic,Factory Tuner,Luxury,Performance	3	520.00
Crossover,Factory Tuner,Performance	4	210.00	Exotic,Flex Fuel,Factory Tuner,Luxury,High-Performance	13	520.00
Crossover,Flex Fuel	64	2073.75	Exotic,Flex Fuel,Luxury,High-Performance	11	520.00
Crossover,Flex Fuel,Luxury	10	1173.20	Exotic,High-Performance	261	1271.33
Crossover,Flex Fuel,Luxury,Performance	6	1624.00	Exotic,Luxury	12	112.67
Crossover,Flex Fuel,Performance	6	5657.00	Exotic,Luxury,High-Performance	79	467.08
Crossover,Hatchback	72	1675.69	Exotic,Luxury,High-Performance,Hybrid	1	204.00
Crossover,Hatchback,Factory Tuner,Performance	6	2009.00	Exotic,Luxury,Performance	36	217.03
Crossover,Hatchback,Luxury	7	204.00	Exotic,Performance	10	1391.00
Crossover,Hatchback,Performance	6	2009.00	Factory Tuner,High-Performance	106	1941.42
Crossover,Hybrid	42	2563.38	Factory Tuner,Luxury	2	617.00
Crossover,Luxury	410	884.55	Factory Tuner,Luxury,High-Performance	215	2133.37
Crossover,Luxury,Diesel	34	2149.41	Factory Tuner,Luxury,Performance	31	1413.42
Crossover,Luxury,High-Performance	9	1037.22	Factory Tuner,Performance	92	1695.70
Crossover,Luxury,Hybrid	24	630.92	Flex Fuel	872	2217.30
Crossover,Luxury,Performance	113	1344.85	Flex Fuel,Diesel	16	5657.00
Crossover,Luxury,Performance,Hybrid	2	3916.00	Flex Fuel,Factory Tuner,Luxury,High-Performance	1	258.00
			Flex Fuel,Hybrid	2	155.00
			Flex Fuel,Luxury	39	746.54
			Flex Fuel,Luxury,High-Performance	33	878.91

Flex Fuel,Luxury,Performance	28	1380.07
Flex Fuel,Performance	87	1680.47
Flex Fuel,Performance,Hybrid	2	155.00
Hatchback	641	1318.87
Hatchback,Diesel	14	873.00
Hatchback,Factory Tuner,High-Performance	13	1205.15
Hatchback,Factory Tuner,Luxury,Performance	9	886.89
Hatchback,Factory Tuner,Performance	22	2159.05
Hatchback,Flex Fuel	7	5657.00
Hatchback,Hybrid	72	2121.25
Hatchback,Luxury	46	1379.50
Hatchback,Luxury,Hybrid	3	454.00
Hatchback,Luxury,Performance	38	1566.13
Hatchback,Performance	252	1039.65
High-Performance	199	1821.45
Hybrid	123	2105.57
Luxury	855	1102.66
Luxury,High-Performance	334	1668.02
Luxury,High-Performance,Hybrid	12	568.83
Luxury,Hybrid	52	673.63
Luxury,Performance	673	1292.62
Luxury,Performance,Hybrid	11	2333.18
Performance	601	1348.87
Performance,Hybrid	1	155.00

- From the chart below, you can see the top-3 market category is labeled with popularity scores, but the number of models for each these categories are less compared to other market categories having lower popularity scores.
- For some market categories having higher number of car models have lower popularity scores and vice versa.

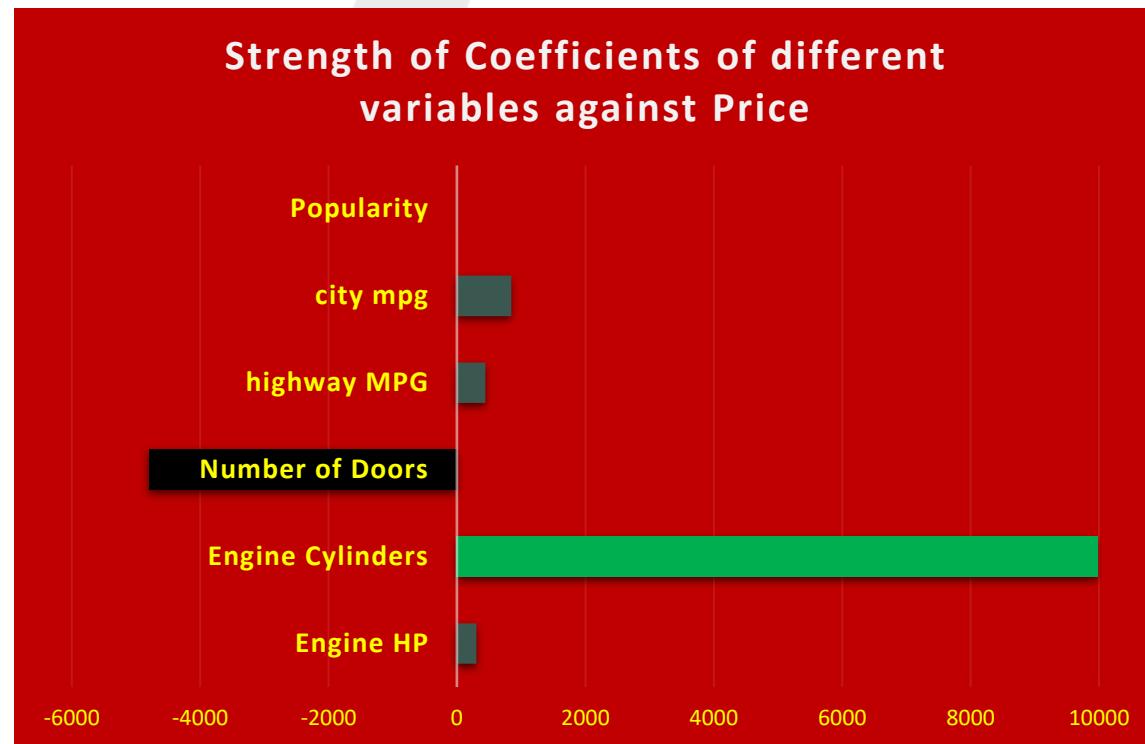


## Task 2: What is the relationship between a car's engine power and its price?



- Positive relationship is found between car's engine power and its price.
- Therefore, price can increase according to the increase of Engine Horse Power.

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



- The bar chart shows the distribution of the coefficients from regression analysis.
- Six independent variables including - Engine HP, Engine Cylinders, highway MPG, number of doors and city mpg are analyzed to determine its influence on car's price.
- Five variables influence car's price positively and one variable impact car's price negatively.
- The coefficient for number of Engine Cylinders is Highly Positive, indicating that there is a strong positive relationship between the number of cylinders in a car's engine and its price. This suggests that cars with more cylinders tend to be more expensive and vice versa.
- The coefficient for number of doors is negative, indicating that there is a strong negative relationship between the number of doors a car has and its price. This suggests that cars with fewer doors tend to be more expensive and vice versa.

## SUMMARY OUTPUT

*Regression Statistics*

Multiple R	0.681937657
R Square	0.465038968
Adjusted R Square	0.464645855
Standard Error	51226.92484
Observations	8172

## ANOVA

	df	SS	MS	F	Significance F
Regression	6	1.8626E+13	3.10434E+12	1182.965657	0
Residual	8165	2.14266E+13	2624197828		
Total	8171	4.00526E+13			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	-96810.82392	4593.218482	-21.07690376	4.5372E-96	-105814.7014	-87806.94641	-105814.7014	-87806.94641
Engine HP	302.993397	8.167039021	37.09954075	1.4631E-278	286.9839214	319.0028725	286.9839214	319.0028725
Engine Cylinders	9983.933937	564.4747481	17.68712236	9.97694E-69	8877.419734	11090.44814	8877.419734	11090.44814
Number of Doors	-4799.980364	646.8507735	-7.420537411	1.28368E-13	-6067.972548	-3531.98818	-6067.972548	-3531.98818
highway MPG	440.2851816	131.1754427	3.356460419	0.000793074	183.1479207	697.4224425	183.1479207	697.4224425
city mpg	843.1740624	121.2254327	6.95542217	3.78622E-12	605.5413541	1080.806771	605.5413541	1080.806771
Popularity	-4.029999588	0.402004197	-10.02476993	1.62205E-23	-4.818030152	-3.241969024	-4.818030152	-3.241969024

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

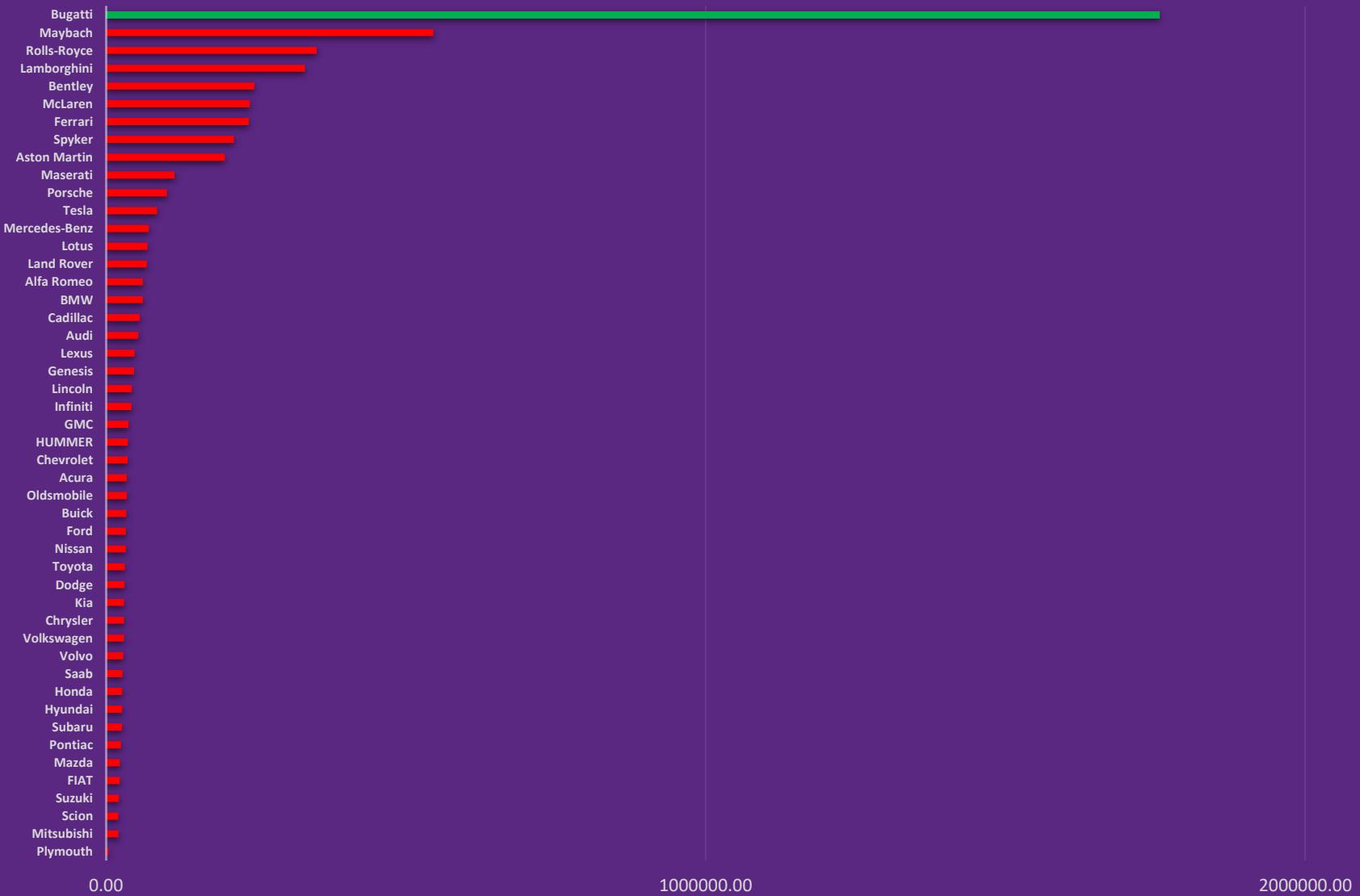
### Manufacturer | Average Price

Acura	34887.59
Alfa Romeo	61600.00
Aston Martin	197910.38
Audi	53452.11
Bentley	247169.32
BMW	61546.76
Bugatti	1757223.67
Buick	33770.40
Cadillac	56231.32
Chevrolet	35875.11
Chrysler	29978.87
Dodge	30995.38
Ferrari	238218.84
FIAT	23060.61
Ford	33178.53
Genesis	46616.67
GMC	37385.75
Honda	27024.98
HUMMER	36464.41
Hyundai	26985.96
Infiniti	42394.21
Kia	30149.31
Lamborghini	331567.31
Land Rover	67823.22

### Manufacturer | Average Price

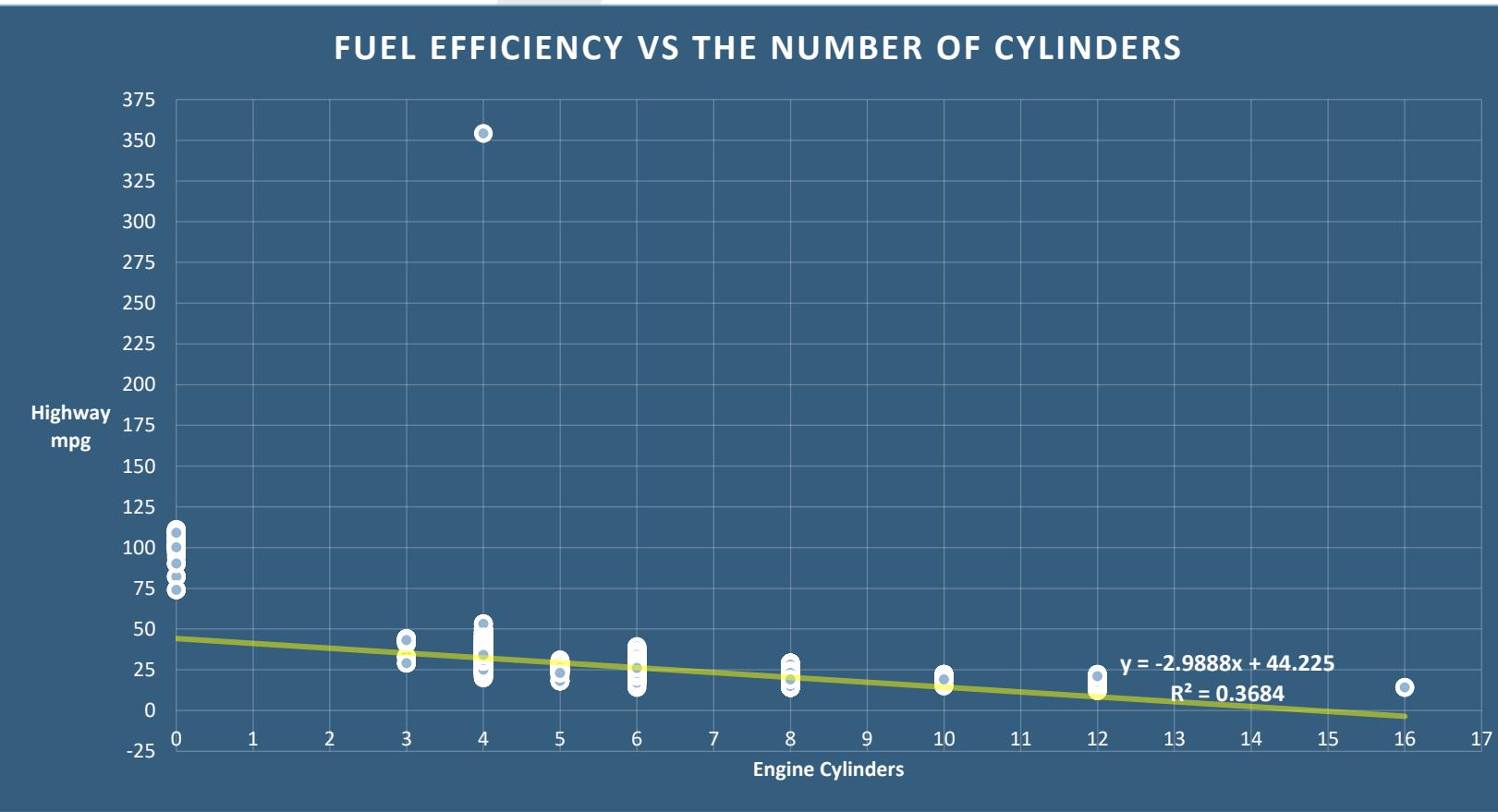
Lexus	47549.07
Lincoln	42839.83
Lotus	69188.28
Maserati	114207.71
Maybach	546221.88
Mazda	23254.44
McLaren	239805.00
Mercedes-Benz	71476.23
Mitsubishi	20286.88
Nissan	32893.15
Oldsmobile	34868.00
Plymouth	4076.82
Pontiac	24728.13
Porsche	101622.40
Rolls-Royce	351130.65
Saab	27413.50
Scion	20395.94
Spyker	213323.33
Subaru	26407.22
Suzuki	21153.05
Tesla	85255.56
Toyota	31229.92
Volkswagen	29932.35
Volvo	28541.16

## Average Price of a car across different manufacturers



- Bugatti is a top manufacturer in terms of average price.
- Plymouth is a brand that is associated with more affordable or budget-friendly price.
- Car prices can be influenced due to variety of factors such as differences in manufacturing costs, local market trend, features and technology used.

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

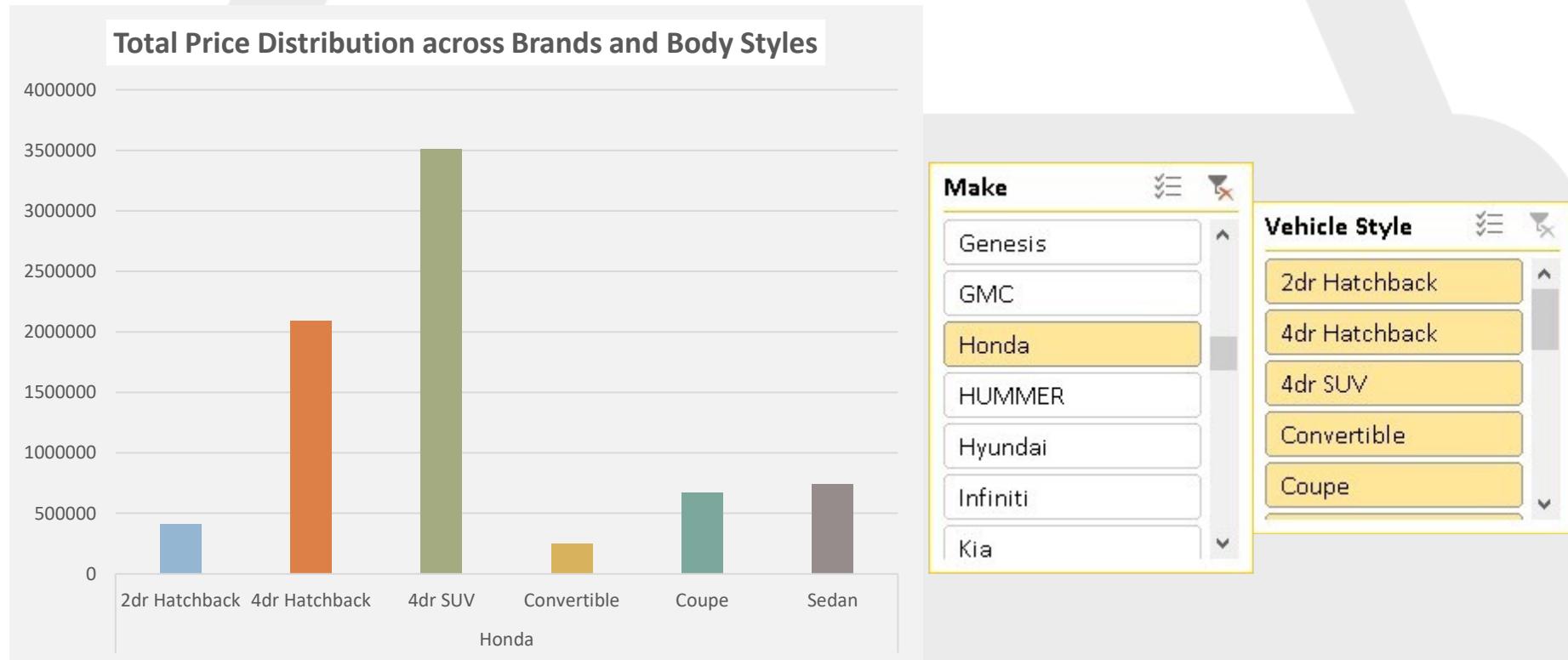


Correlation Coefficient	
	Engine Cylinders
Highway MPG	-0.61

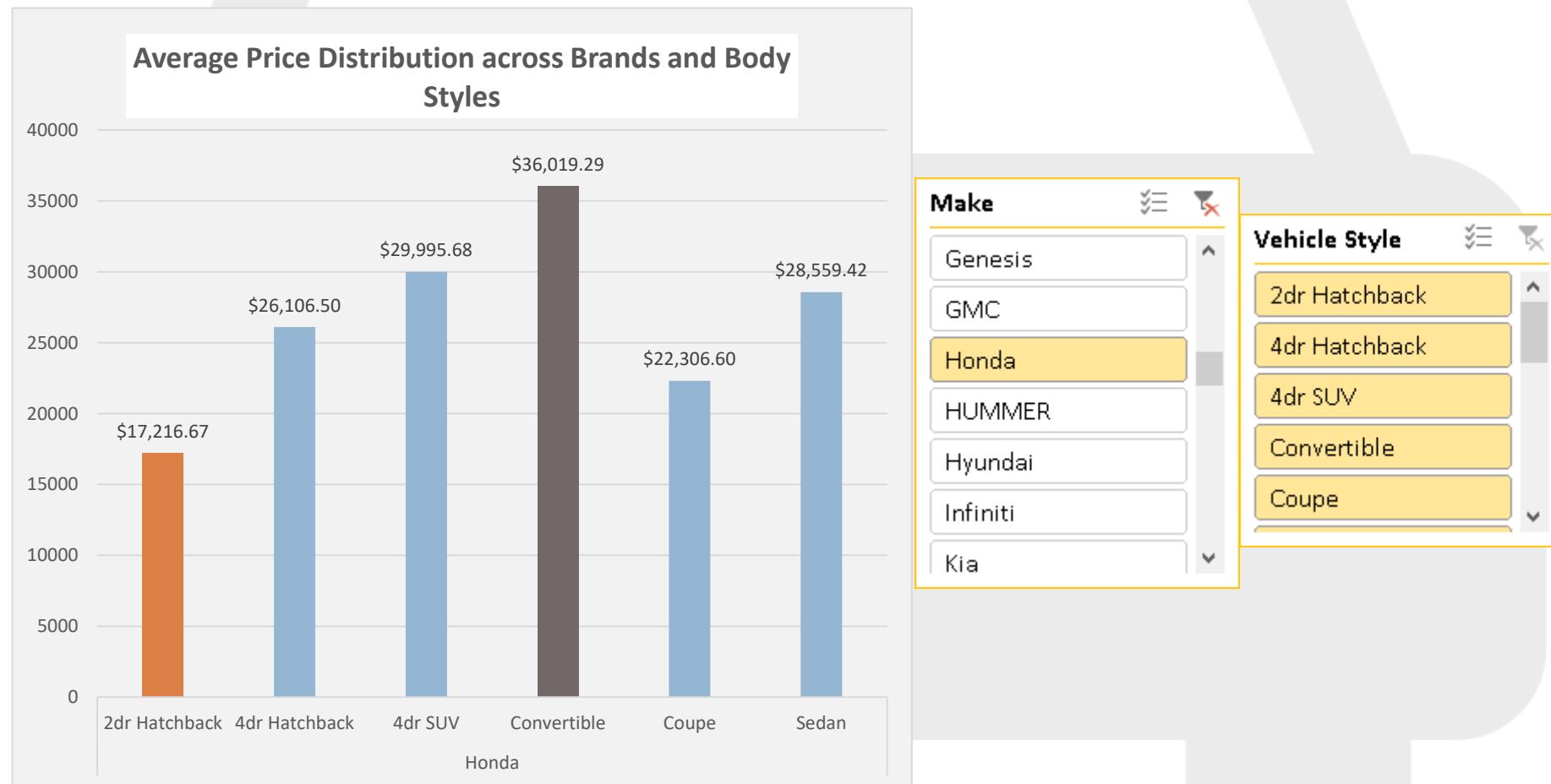
- The slope of the linear equation is negative, which indicates an inverse relationship between the x and y variables, i.e., as the value of y increases, the value of x decreases, and vice versa.
- A correlation coefficient value of -0.61 indicates a moderate negative correlation between the number of engine cylinders and highway MPG.
- This means that as the number of cylinders in a car's engine increases, the highway MPG generally decreases, and vice versa.
- Therefore, we can conclude that there is a significant relationship between the number of cylinders and highway MPG.

# Dashboard Tasks:

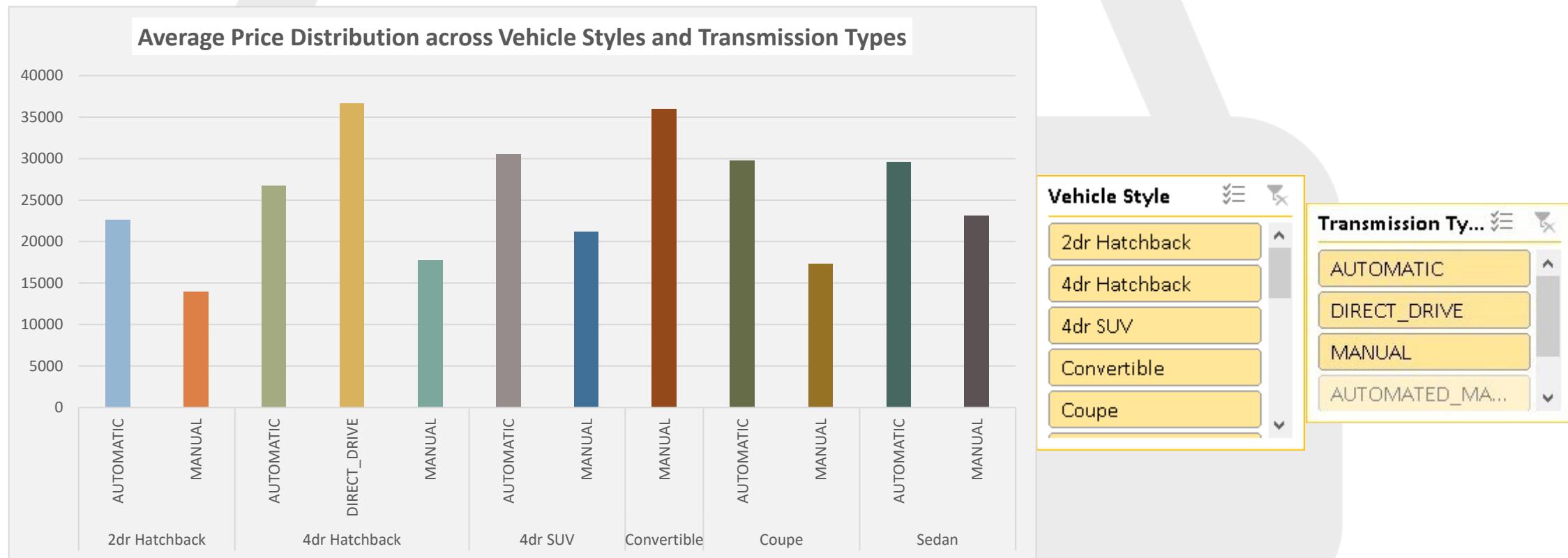
Task 1: How does the distribution of car prices vary by brand and body style?



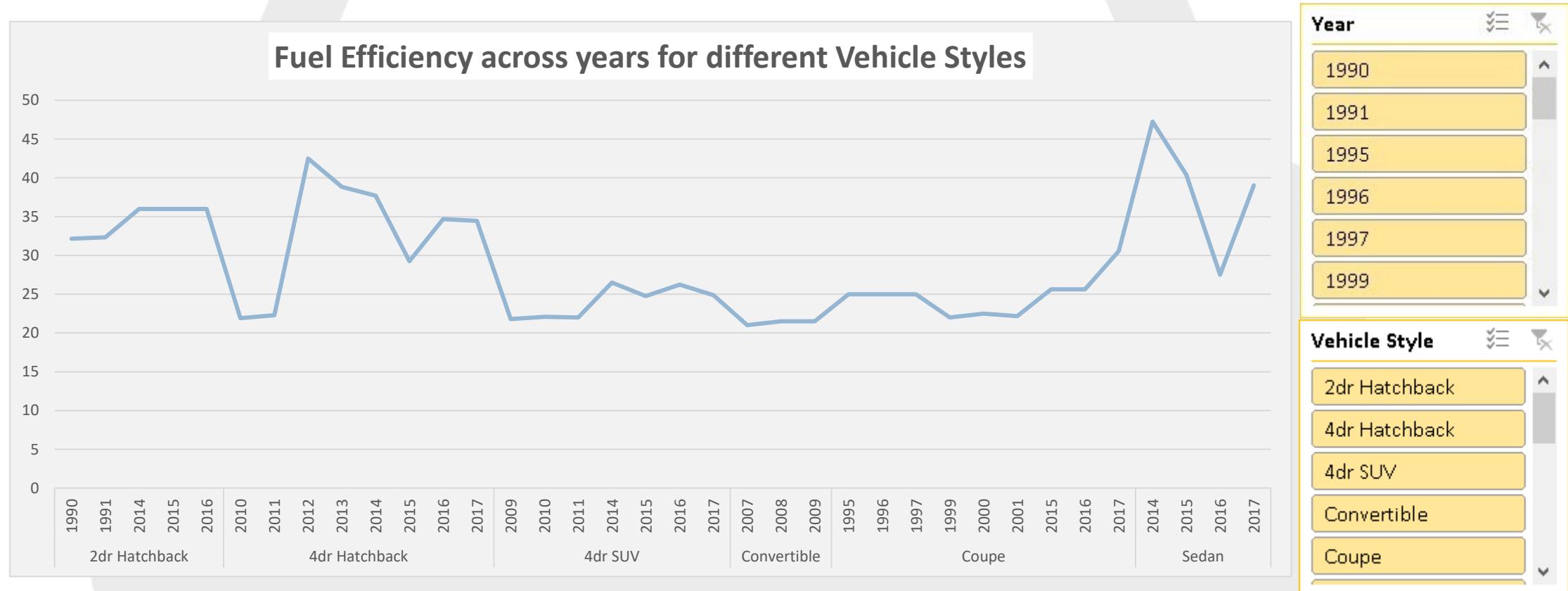
Task 2: Which car brands have the highest and lowest average MSRPs, and how does this vary by body style?



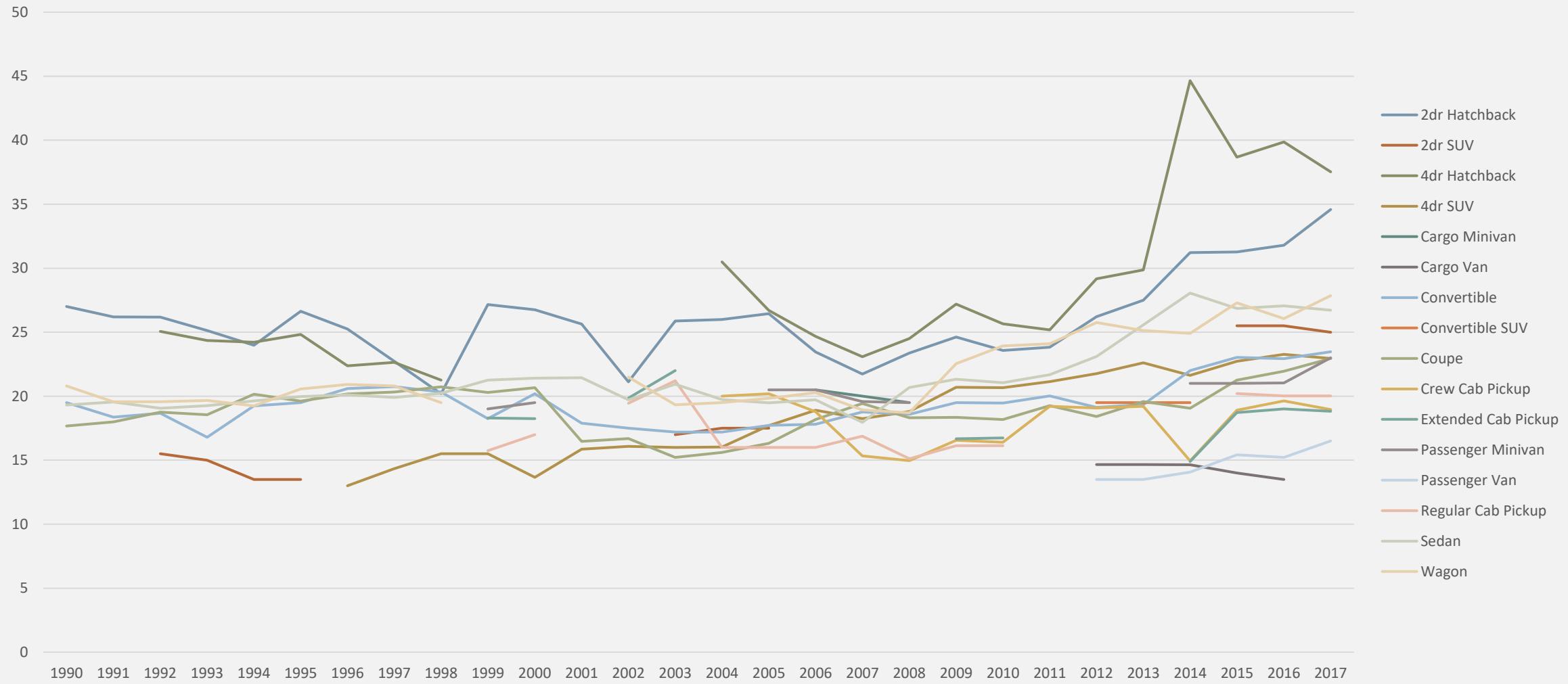
Task 3: How do the different feature such as transmission type affect the MSRP, and how does this vary by body style?



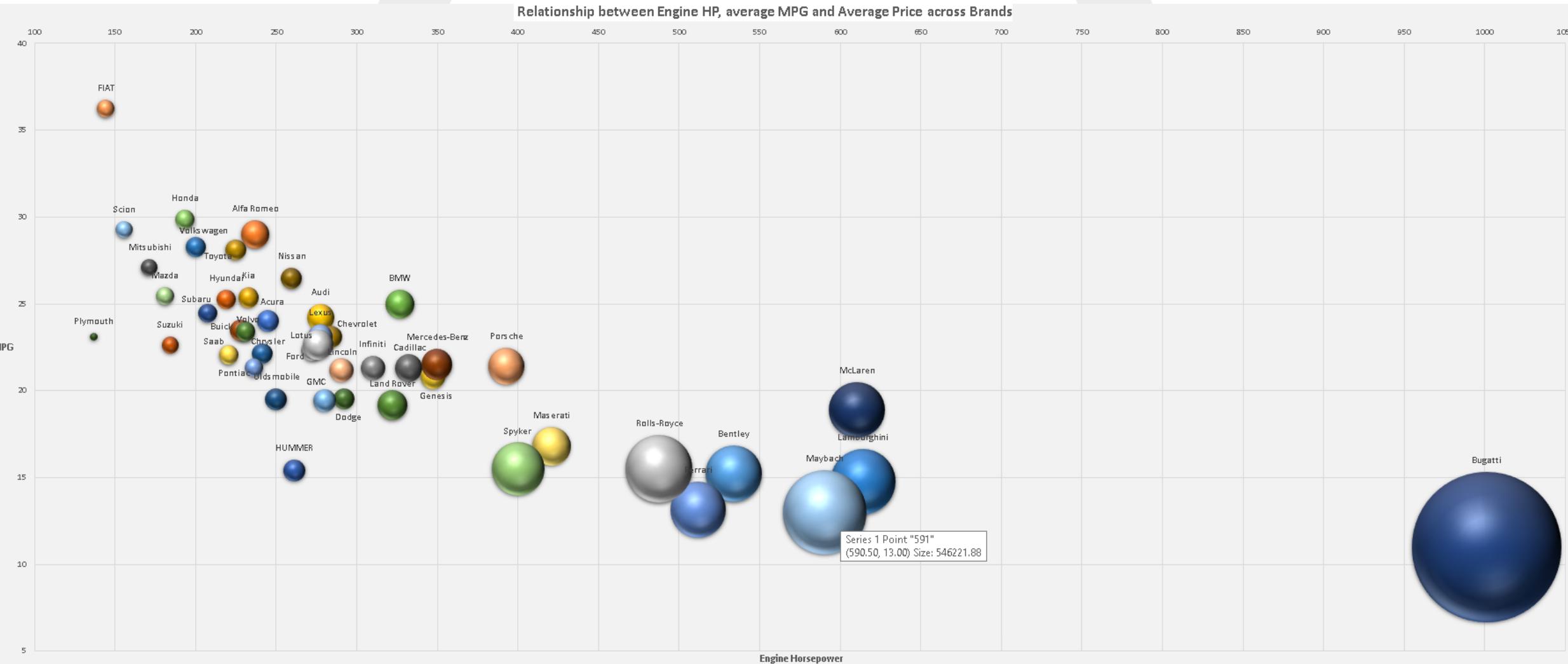
Task 4: How does the fuel efficiency of cars vary across different body styles and model years?



## Fuel Efficiency across years for different Vehicle Styles



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



# Result and Conclusion:

- Engine HP, Engine Cylinders and Transmission type are some top car features that impact on price significantly.
- Fuel efficiency has increased over the years.
- The average price of a car across majority of manufacturers is almost same or neck-to-neck.
- Number of Cylinders can impact on fuel efficiency negatively.
- Cars with higher price segment provides less fuel efficiency.
- In terms of Profitability, Fuel Efficiency and budget-friendly price, there are untapped opportunities and there is lots of scope in productivity of Market Segments like ‘Crossover, Hatchback, Factory Tuner, Performance’, ‘Crossover, Hatchback, Performance’, ‘Flex Fuel, Hybrid’, ‘Hatchback, Flex Fuel’, ‘Hatchback, Luxury, Hybrid’, ‘Hatchback, Diesel’ for car manufacturers to capitalize and capture the market.
- This project has helped me a lot to get a good understanding of the Automotive Industry and what are the top features which can impact car prices.
- Strengthened my data analysis skills by applying statistical knowledge such as performing regression analysis and correlation test.
- Project link → [Car\\_data.xlsx](#) .  
It is recommended to download and open the project link in MS EXCEL only since some of the visuals might differ from MS Excel in Google Spreadsheets.
- The drawback is that the dataset was last updated in 2017, so it will not reflect the current trends or prices in the automotive industry.