Analyzing the Impact of Car Features on Price and Profitability

-ANDRI

Project Description

- With increasing competition among automobile manufacturers and a changing consumer landscape, it has become more important than ever to understand the factors that drive consumer demand for cars.
- 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.

Approach and tech stack used

- Understanding the data.
- Cleaning / pre-processing the data and handling all NULL values and outliers.
- Data analysis and visualization.
- Gaining insights/hypothesis from the analysis.

I have used MS-Office 2019 for analysis and powerpoint presentation.

Understanding the dataset

- Number of columns: 16
- Number of rows: 11915
- Number of duplicate rows: 715
- Columns with blank cells: 4

Understanding the dataset

- Number of blank cells in each column:

i. Engine HP : 69

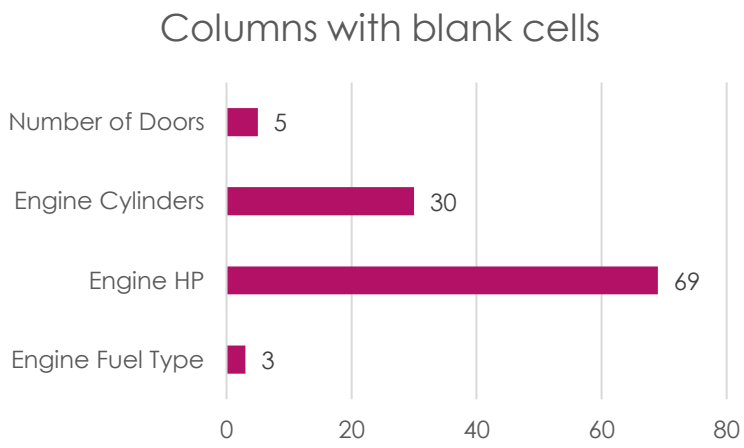
ii. Engine cylinder: 30

iii. Number of doors: 5

iv. Engine fuel type : 3

- Number of cells having unknown/NA values:

1. Market Category : 3373
2. Transmission type: 12

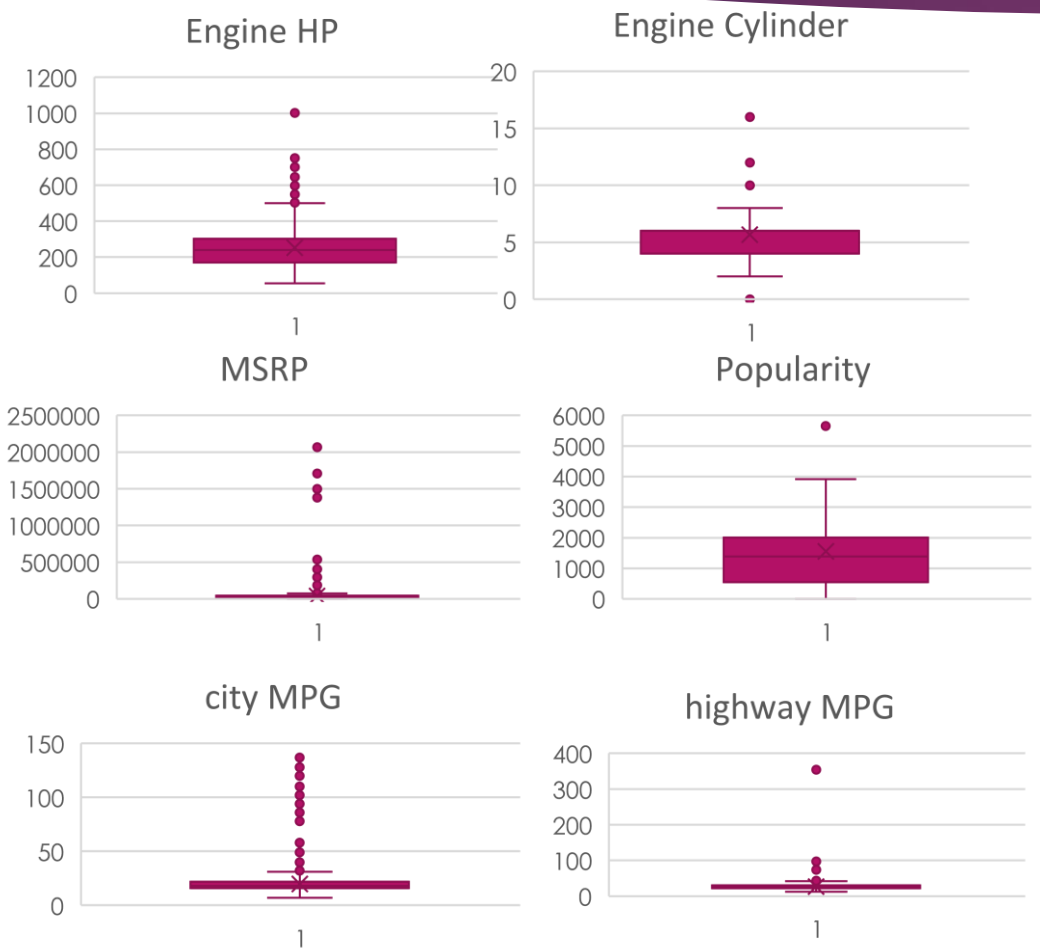


Data Cleaning: Removing and imputing NULL cells

- Removing the duplicate data from the dataset.
- Fill in the blank values in the 3 columns with the values available on the internet/online.
- Deleting the null rows from the "engine fuel type" columns as just 3 cells are blank i.e. for the Suzuki Verona model.
- To uncomplicate the dataset, changing the fuel types to a standardized value.

Fuel type	Modified Fuel type
premium unleaded (required)	Premium Unleaded
diesel	Diesel
electric	Electric
regular unleaded	Regular Unleaded
flex-fuel (unleaded/E85)	Flex-Fuel
premium unleaded (recommended)	Premium Unleaded
flex-fuel (premium unleaded recommended/E85)	Flex-Fuel
flex-fuel (unleaded/natural gas)	Flex-Fuel
natural gas	Natural Gas
flex-fuel (premium unleaded required/E85)	Flex-Fuel

Data cleaning: outliers



Outliers:

- Removing outliers from the engine cylinder i.e value 0 would remove all the electric cars from the dataset, as electric cars don't have engines, they have motors.
- Removing outliers from the MSRP column would remove almost all luxury brand cars, hence not removing the outliers.
- Removing outliers from city MPG/highway MPG would again remove the electric cars and hybrid cars hence it is kept as it is.
- Removing the outlier from the highway MPG column i.e. 354 as it was of a petrol car which is not possible.

Data cleaning: Engine Cylinders

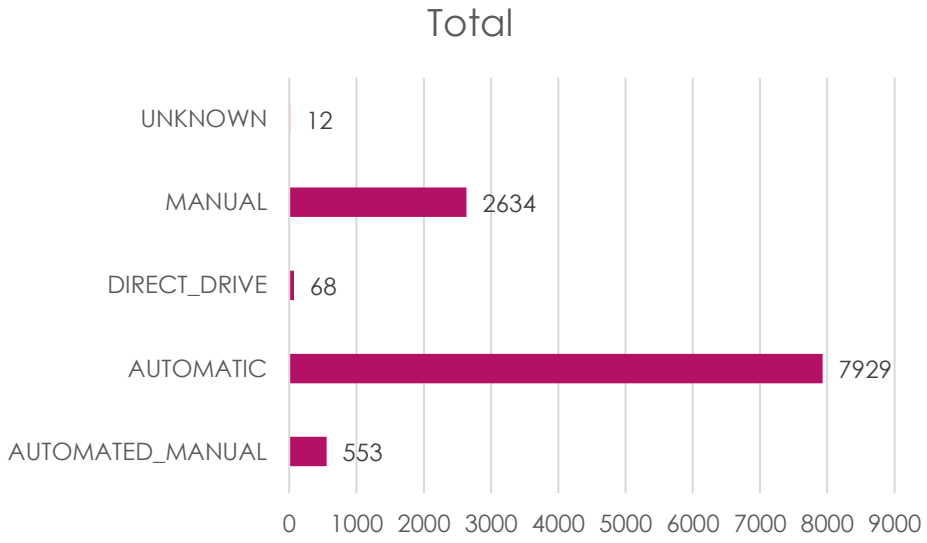
- The blank values in the "Engine Cylinders" column contains two types of cars:

1. Electric
2. Mazda

- Electric engines don't have a engine hence their number of engine cylinders would be 0.
- Mazda's cars RX-7 and RX-8 have rotary engines (Wankle engine) hence don't use engine cylinders hence the imputed value would be 0 here.

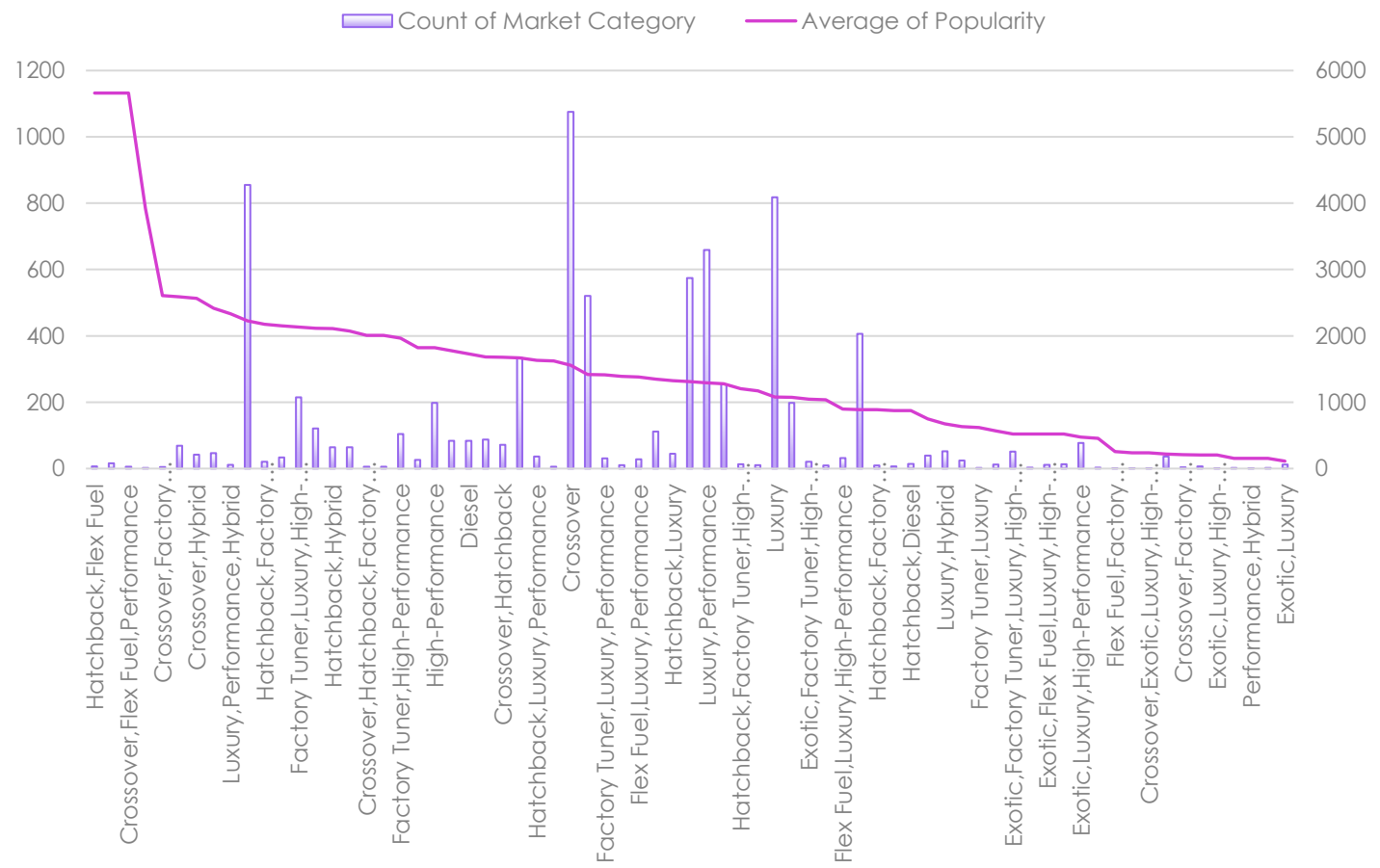
Data Cleaning: Transmission type

- Imputing the Transmission type blank cells with the mode value as there is no information present.
- Imputing “Unknown” with “Automatic”



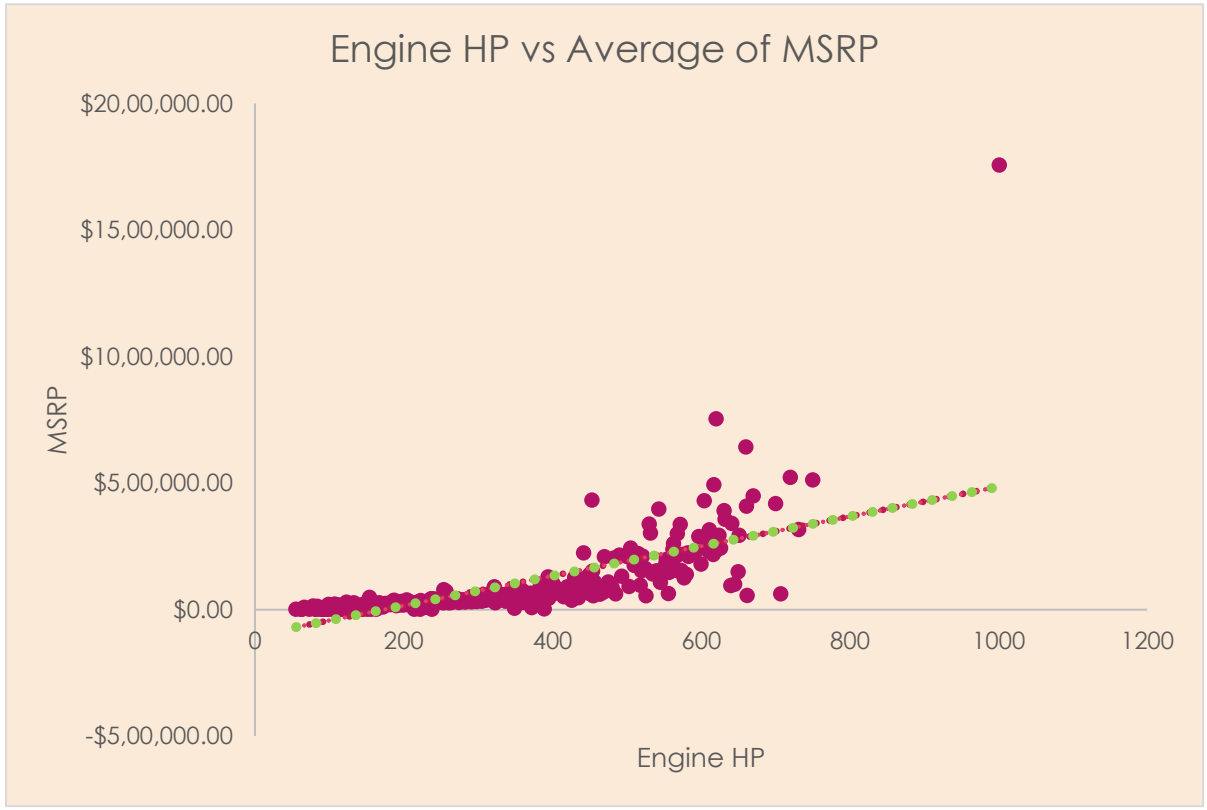
Analysis: Popularity of car model across different market categories

- Hatchback, Flex-Fuel has the highest popularity while the Exotic luxury cars have the lowest popularity.
- Most selling cars in the market are from Crossover category.



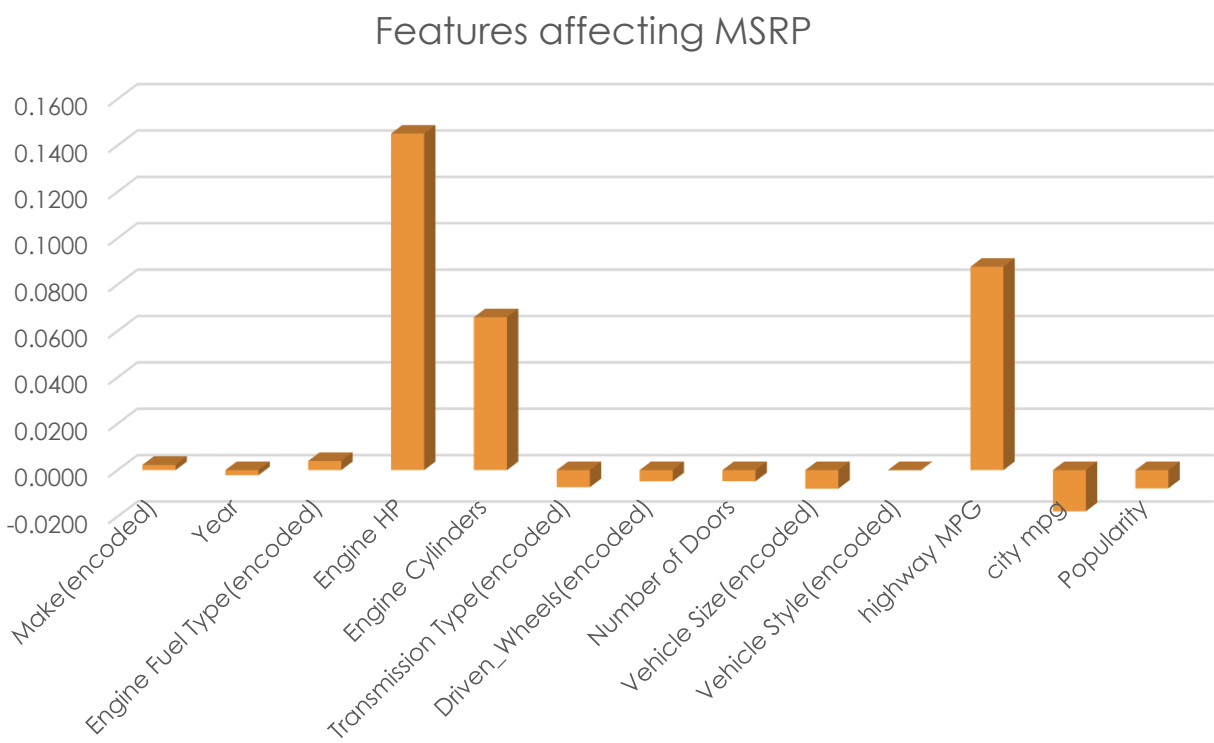
Analysis: Engine power vs Price

- Engine HP and MSRP have the positive linear relation.
- The outlier (1001 HP) here is of the Bugatti Veyron 16.4 model having 16 engine cylinders which is a premium quality car.



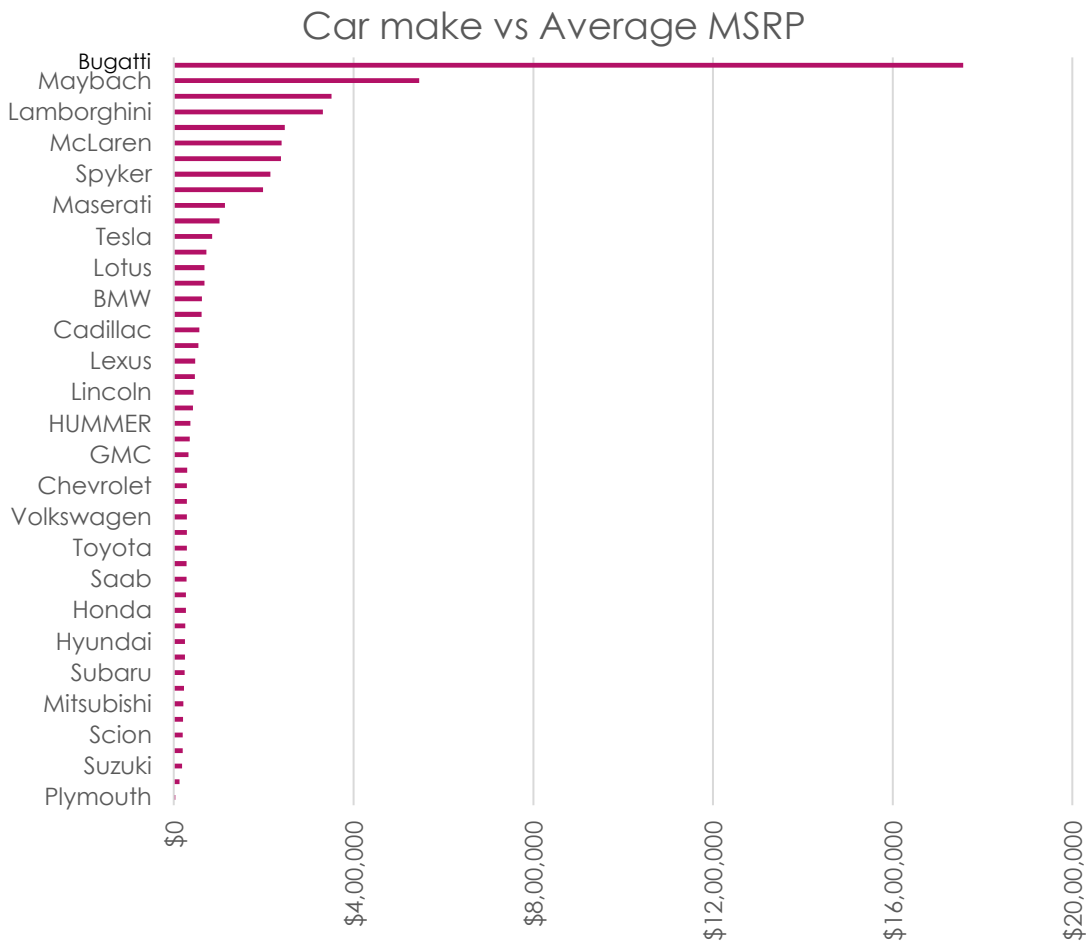
Analysis: Features affecting car price

- Engine HP and cylinders have the highest effect on determining a car's price as they have the highest positive coefficient.



Analysis: Car manufacturer vs MSRP

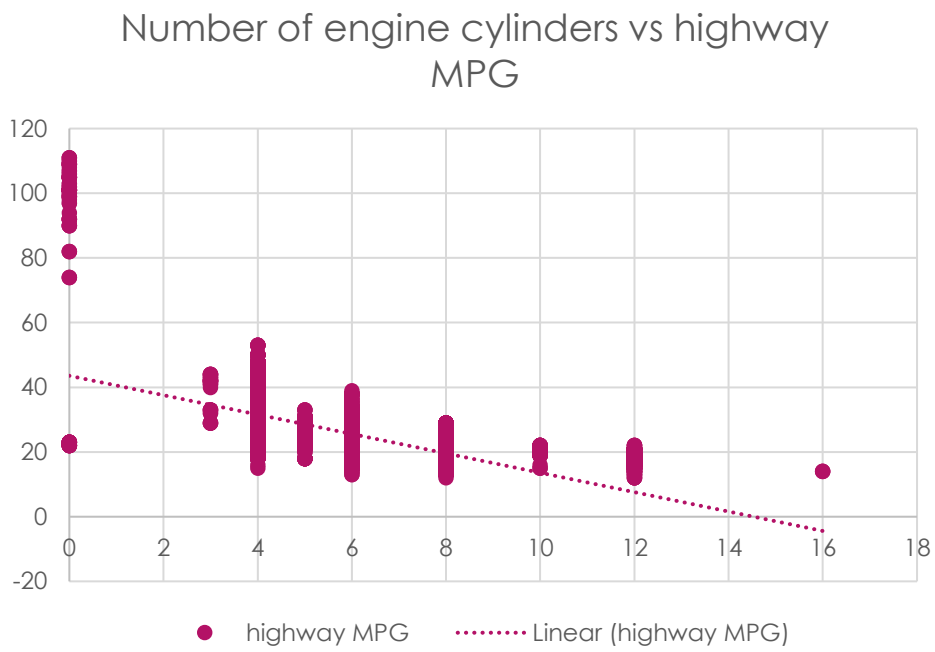
- Luxury cars are the expensive ones with Bugatti being the most expensive while the daily use cars for masses are cheaper.



Analysis: Impact of no. of cylinders on MPG

- Number of cylinders has negative correlation with highway MPG.
- As the number of cylinders in a car increases, the MPG decreases.
- This is also shown in the correlation analysis where correlation coeff is -0.77.

	Number of cylinders	*Average of highway MPG*	*Average of city mpg*
Number of cylinders	1		
Average of highway MPG	-0.776841117	1	
Average of city mpg	-0.729796378	0.996467599	1



Key Insights:

- From the analysis, it is evident that the feature affecting car sale price the most is the Engine HP.
- As the number of cylinders increases, the horsepower increases, and the cost also increases but the MPG of the car becomes less, this is seen in super luxury cars that people buy for luxury purposes and not for daily use.

Drive Links:

- Analysis excel sheet: https://docs.google.com/spreadsheets/d/1tsGavkaPh-0LvMNKRveqBBKnZ_Q18Lh-/edit?usp=sharing&ouid=115109770037321084146&rtpof=true&sd=true

ThankYou