



Analyzing the Impact of Car Features on Price and Profitability

By
Stan Pereira

Project Description

A silver Mercedes-Benz sedan is shown from a front-three-quarter view, positioned behind the text. The car is a modern model with a prominent grille and headlights.

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.

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.

Approach



Understand the data – Understand the dataset, features and impact on other columns.

Clean the data - Identify Missing Data and Deal with it Appropriately.

Analyze Data – Analyze the data to find various relationships between features to derive conclusions.

Visualization – Visualize the data using Tables, Charts and Dashboards.

Tech Stack Used

Ability to perform calculations, data analysis, data visualization, data transformation, and data cleaning with Excel tools and functions.

Transform and clean data with features like Power Query and Flash Fill.

Microsoft Excel 2010
Version 14.0.7628.5000

Code to automate tasks and customize functions with VBA (Visual Basic for Applications).

Availability of free templates and code to customize and automate Excel.

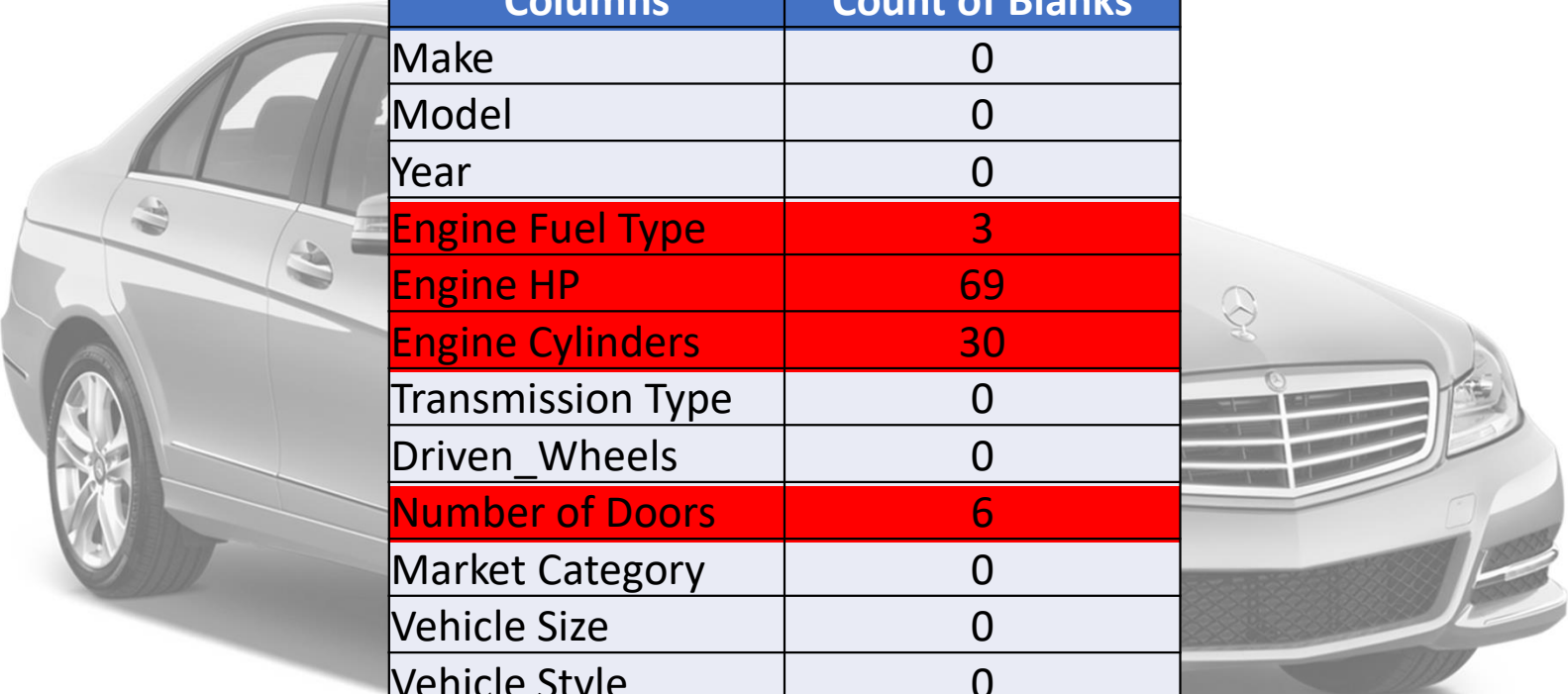
Understanding the Data



Table Information (Original)	
Total Rows	11914
Total Columns	16
Total Blanks	108
Duplicates	715

Table Information (After Cleaning)	
Total Rows	11199
Total Columns	16
Total Blanks	0
Duplicates	0

Data Cleaning – Distribution of Blanks



Columns	Count of Blanks
Make	0
Model	0
Year	0
Engine Fuel Type	3
Engine HP	69
Engine Cylinders	30
Transmission Type	0
Driven_Wheels	0
Number of Doors	6
Market Category	0
Vehicle Size	0
Vehicle Style	0
highway MPG	0
city mpg	0
Popularity	0
MSRP	0

Data Cleaning (Engine Fuel Type, Cylinders & Doors)

Engine Fuel Type		
Engine Fuel Type	Count	Change To
diesel	150	Diesel
electric	66	Electric
flex-fuel (premium unleaded recommended/E85)	26	Flex-Fuel Premium
flex-fuel (premium unleaded required/E85)	53	Flex-Fuel Premium
flex-fuel (unleaded/E85)	887	Flex-Fuel Regular
flex-fuel (unleaded/natural gas)	6	Flex-Fuel Regular
natural gas	2	Natural Gas
premium unleaded (recommended)	1392	Premium Unleaded
premium unleaded (required)	1956	Premium Unleaded
regular unleaded	6658	Regular Unleaded
(blank)	3	Regular Unleaded. As the records have the same model and make, and the car used regular unleaded fuel type

Blank Number of Doors				
Make	Model	Count of Make	New Value	Remarks
Ferrari	FF	1	2	Same as the model of another record with similar values
Tesla	Model S	5	4	Same as the model of another record with similar values

Engine Cylinders					
Make	Model	Engine Fuel Type	Count of Make	New Value	Remarks
Chevrolet	Bolt EV	electric	2	0	Electric Vehicles have no cylinders
Mazda	RX-7	regular unleaded	3	4	Wankel 2 engine has no cylinders but is equivalent to a 4 cylinder engine
Mazda	RX-8	premium unleaded (required)	17	4	Wankel 2 engine has no cylinders but is equivalent to a 4 cylinder engine
Mitsubishi	i-MiEV	electric	3	0	Electric Vehicles have no cylinders
Toyota	RAV4 EV	electric	1	0	Electric Vehicles have no cylinders
Volkswagen	e-Golf	electric	4	0	Electric Vehicles have no cylinders

Data Cleaning (Engine HP)

Updating the Blank Engine HP (Based on Certain Factors)
Size of the engine, Number of cylinders, Type of fuel injection system

Make	Model	Year	Engine Fuel Type	Engine Cylinders	Count of Make	New Engine HP	Remarks
Chevrolet	Impala	2015 – 2017	flex-fuel (unleaded/natural gas)	6	2	230	Online
FIAT	500e	2015 - 2017	electric	0	1	111	Online
Ford	Escape	2017	regular unleaded	4	4	168	Online
Ford	Focus	2015 - 2017	electric	0	1	143	Online
Ford	Freestar	2005	regular unleaded	6	6	201	Online
Honda	Fit EV	2013 - 2014	electric	0	1	123	Online
Kia	Soul EV	2015 - 2016	electric	0	2	109	Online
Lincoln	Continental	2017	premium unleaded (recommended)	6	4	400	Online
Lincoln	MKZ	2017	regular unleaded	4	4	245	Online
Mercedes-Benz	M-Class	2015	diesel	4	1	200	Online
Mitsubishi	i-MiEV	2014	electric	(blank)	1	66	Online
Nissan	Leaf	2014 - 2016	electric	0	3	107	Online
Tesla	Model S	2014 – 2016	electric	0	4	360	302,362 or 416
Toyota	RAV4 EV	2013 - 2014	electric	0	1	154	

Data Cleaning (Engine HP)

Updating the Blank Engine HP (Based on Certain Factors)
Size of the engine, Number of cylinders, Type of fuel injection system

Make	Model	Year	Engine Fuel Type	Engine Cylinders	Count of Make	New Engine HP	Remarks
Chevrolet	Impala	2015 – 2017	flex-fuel (unleaded/natural gas)	6	2	230	Online
FIAT	500e	2015 - 2017	electric	0	1	111	Online
Ford	Escape	2017	regular unleaded	4	4	168	Online
Ford	Focus	2015 - 2017	electric	0	1	143	Online
Ford	Freestar	2005	regular unleaded	6	6	201	Online
Honda	Fit EV	2013 - 2014	electric	0	1	123	Online
Kia	Soul EV	2015 - 2016	electric	0	2	109	Online
Lincoln	Continental	2017	premium unleaded (recommended)	6	4	400	Online
Lincoln	MKZ	2017	regular unleaded	4	4	245	Online
Mercedes-Benz	M-Class	2015	diesel	4	1	200	Online
Mitsubishi	i-MiEV	2014	electric	(blank)	1	66	Online
Nissan	Leaf	2014 - 2016	electric	0	3	107	Online
Tesla	Model S	2014 – 2016	electric	0	4	360	302,362 or 416
Toyota	RAV4 EV	2013 - 2014	electric	0	1	154	

Data Cleaning (Other Changes)

Rename Columns

Old Name	New Name
Driven Wheels	Drive Wheels
highway MPG	Highway MPG
city mpg	City MPG

The #N/A in Market Category Column was left as is.

All the Duplicate values were removed

One Outlier 'Audi A6', had an highway MPG of 354, which was corrected to 34

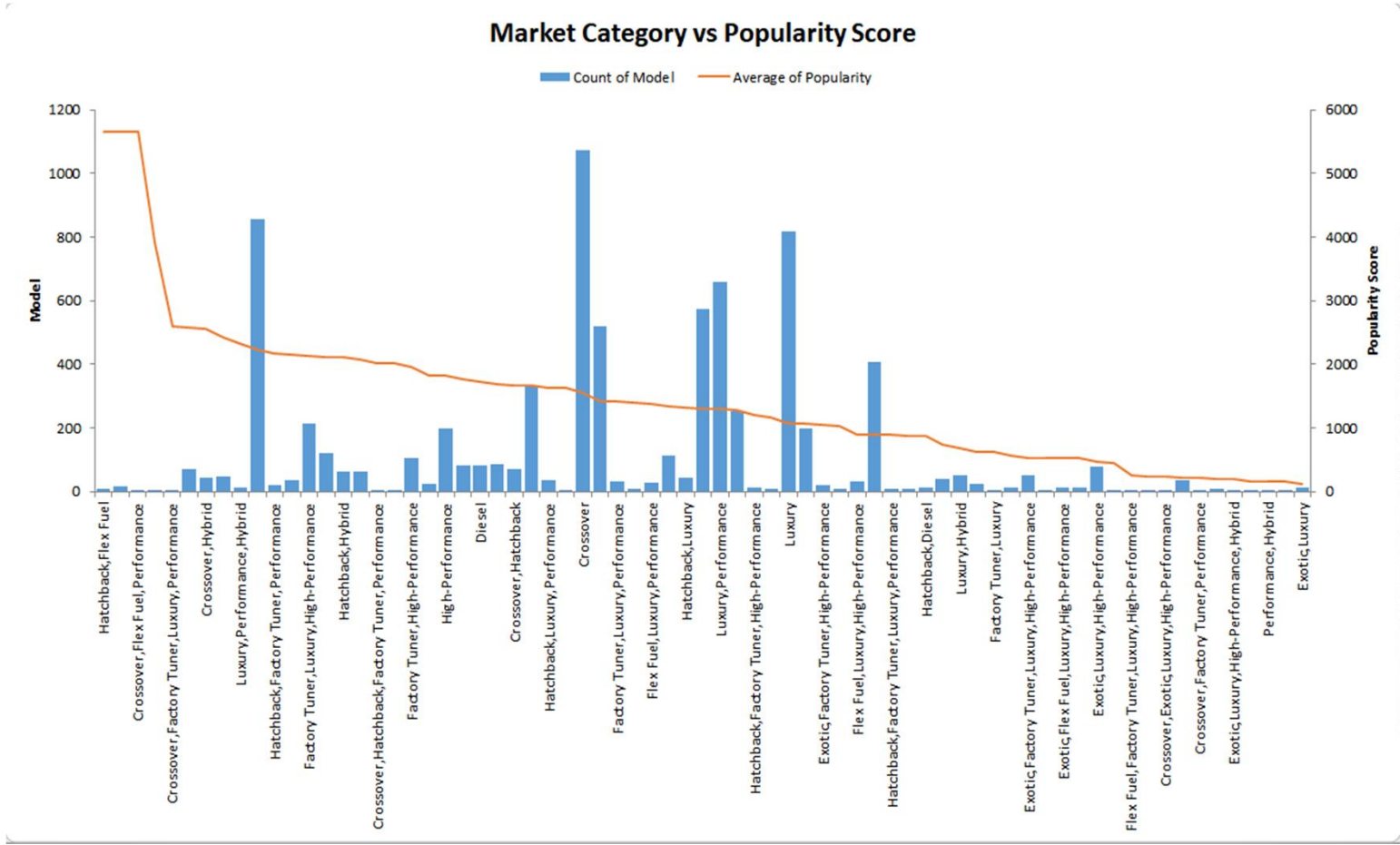
Data Analysis



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.

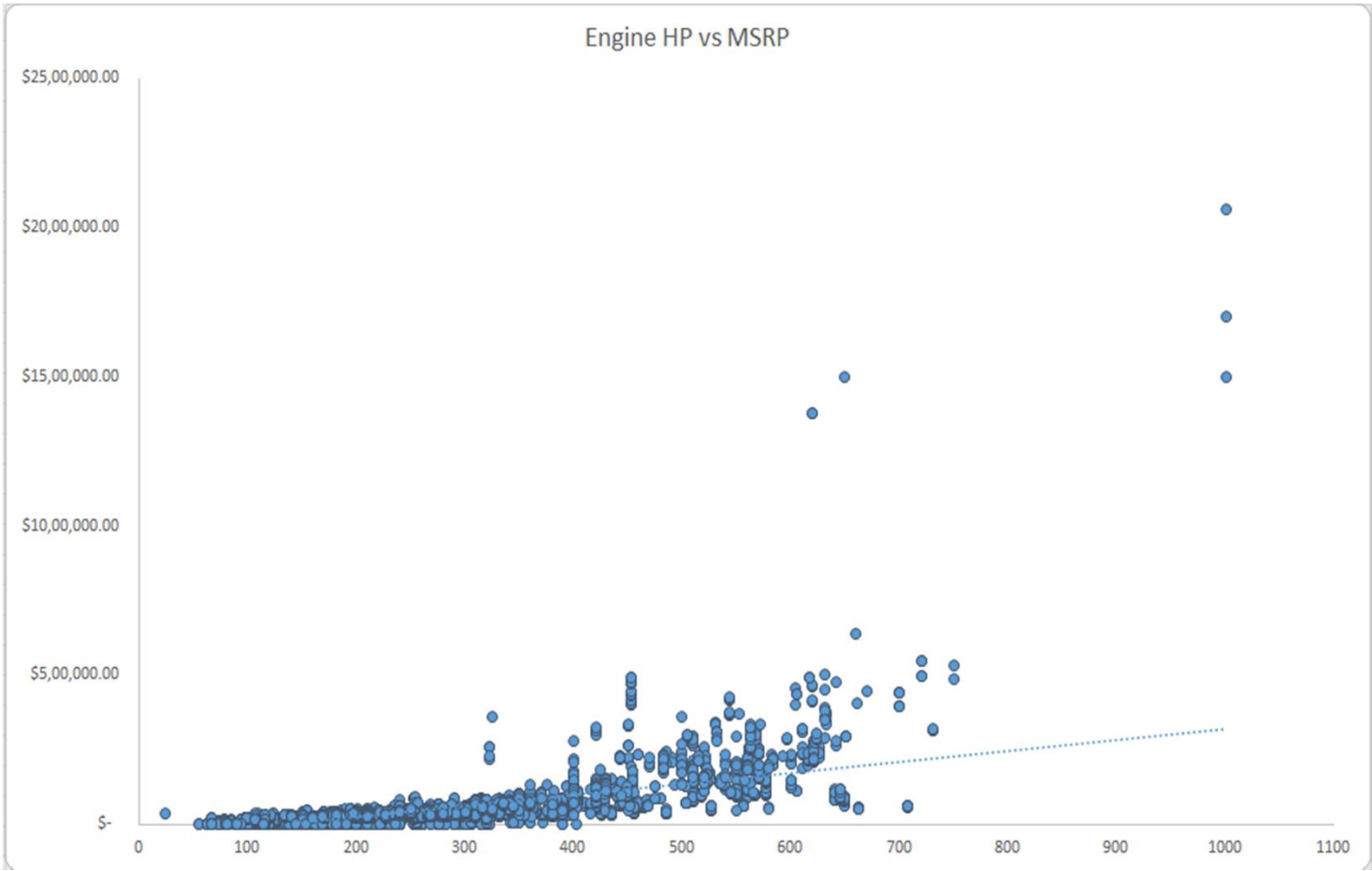
Task 1.B. Create a combo chart that visualizes the relationship between market category and popularity.



‘Crossover’ has the most number of models, but the ‘Hatchback, Flex Fuel’ is the most popular.

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.



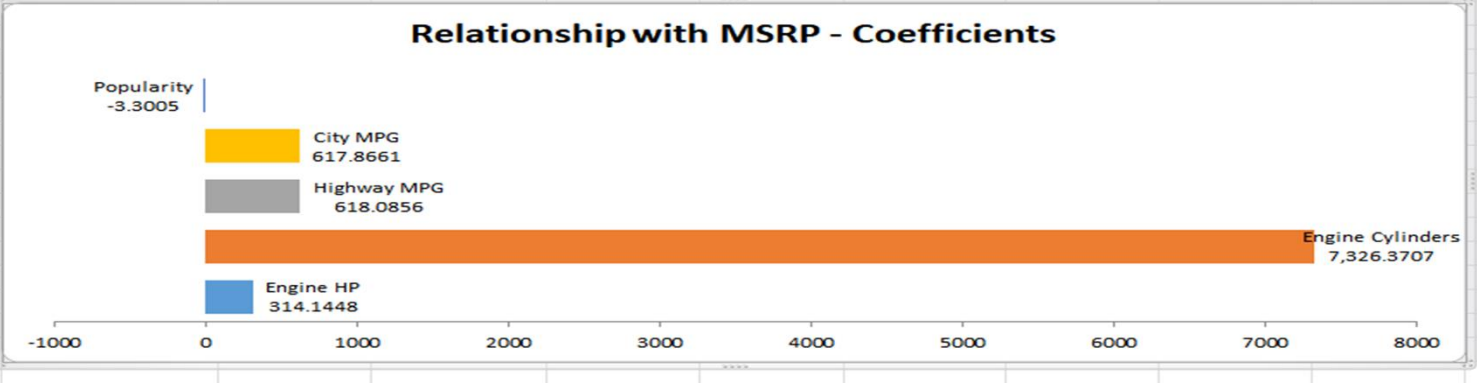
'Engine HP' and 'MSRP' have a positive linear relation.

Note the Dodge Challenger which is high on HP, but low on MSRP. This could be due to other features such as low city and highway MPG, or fuel type.

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.

SUMMARY OUTPUT									
Regression Statistics									
Multiple R	0.677769875								
R Square	0.459372003								
Adjusted R Square	0.4591305								
Standard Error	45255.23123								
Observations	11199								
ANOVA									
	df	SS	MS	F	Significance F				
Regression	5	1.94783E+13	3.89565E+12	1902.140051	0				
Residual	11193	2.29237E+13	2048035954						
Total	11198	4.24019E+13							
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%	
Intercept	-102560.8823	3115.699988	-32.91744478	6.5455E-227	-108668.2025	-96453.56213	-108668.2025	-96453.56213	
Engine HP	314.1448474	6.28046852	50.01933318	0	301.8340241	326.4556708	301.8340241	326.4556708	
Engine Cylinders	7326.370725	454.4774819	16.12042624	8.19089E-58	6435.514895	8217.226555	6435.514895	8217.226555	
Highway MPG	618.0856489	107.1773191	5.766944482	8.2858E-09	407.9992457	828.1720521	407.9992457	828.1720521	
City MPG	617.8660941	101.4618006	6.089642511	1.16876E-09	418.9831128	816.7490754	418.9831128	816.7490754	
Popularity	-3.300531052	0.296297403	-11.13925069	1.14296E-28	-3.881326095	-2.719736009	-3.881326095	-2.719736009	



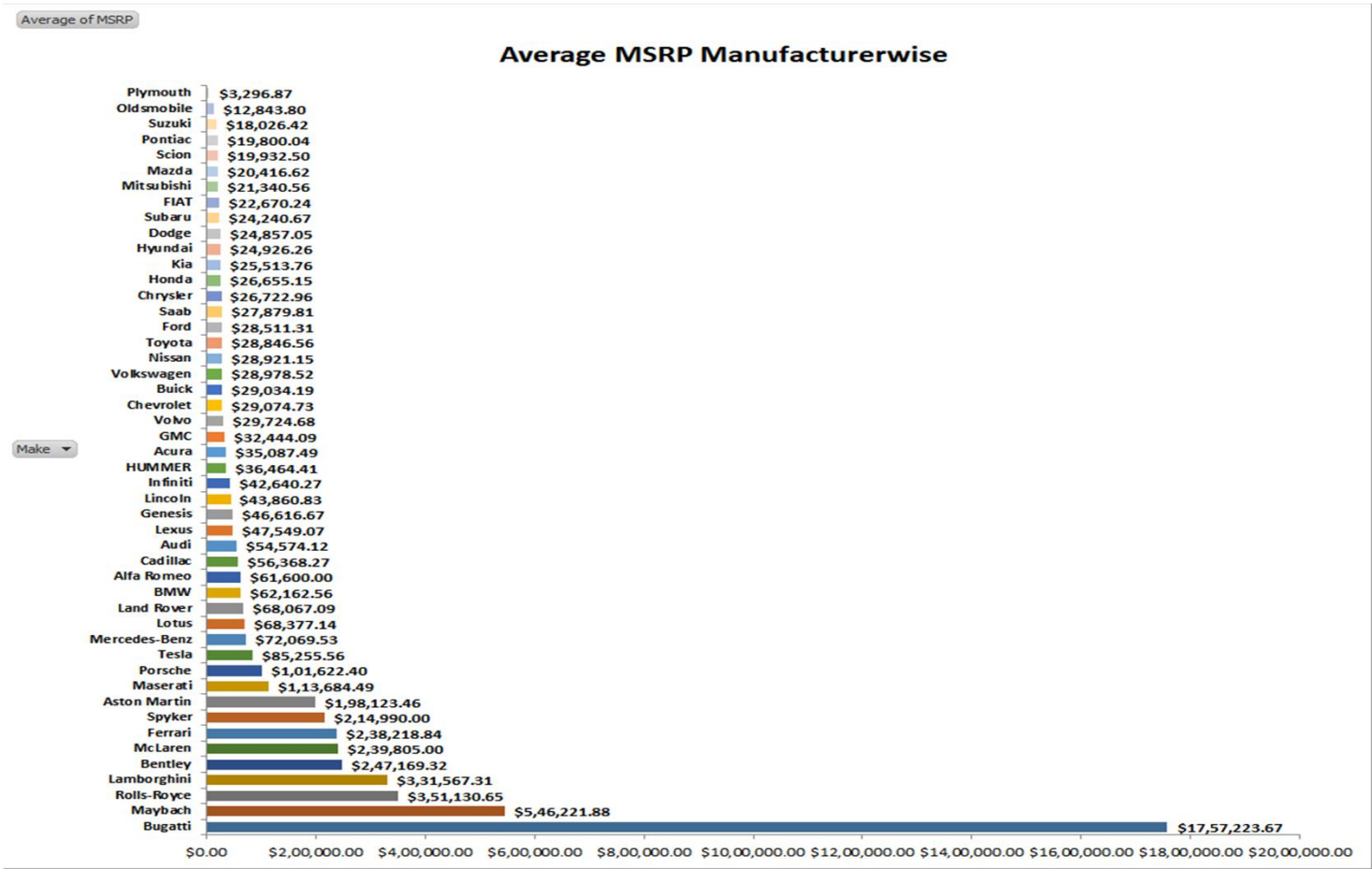
‘Engine Cylinder’ and ‘MSRP’ have the highest positive coefficient.

Whereas, ‘Popularity’ has a negative coefficient.

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.

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

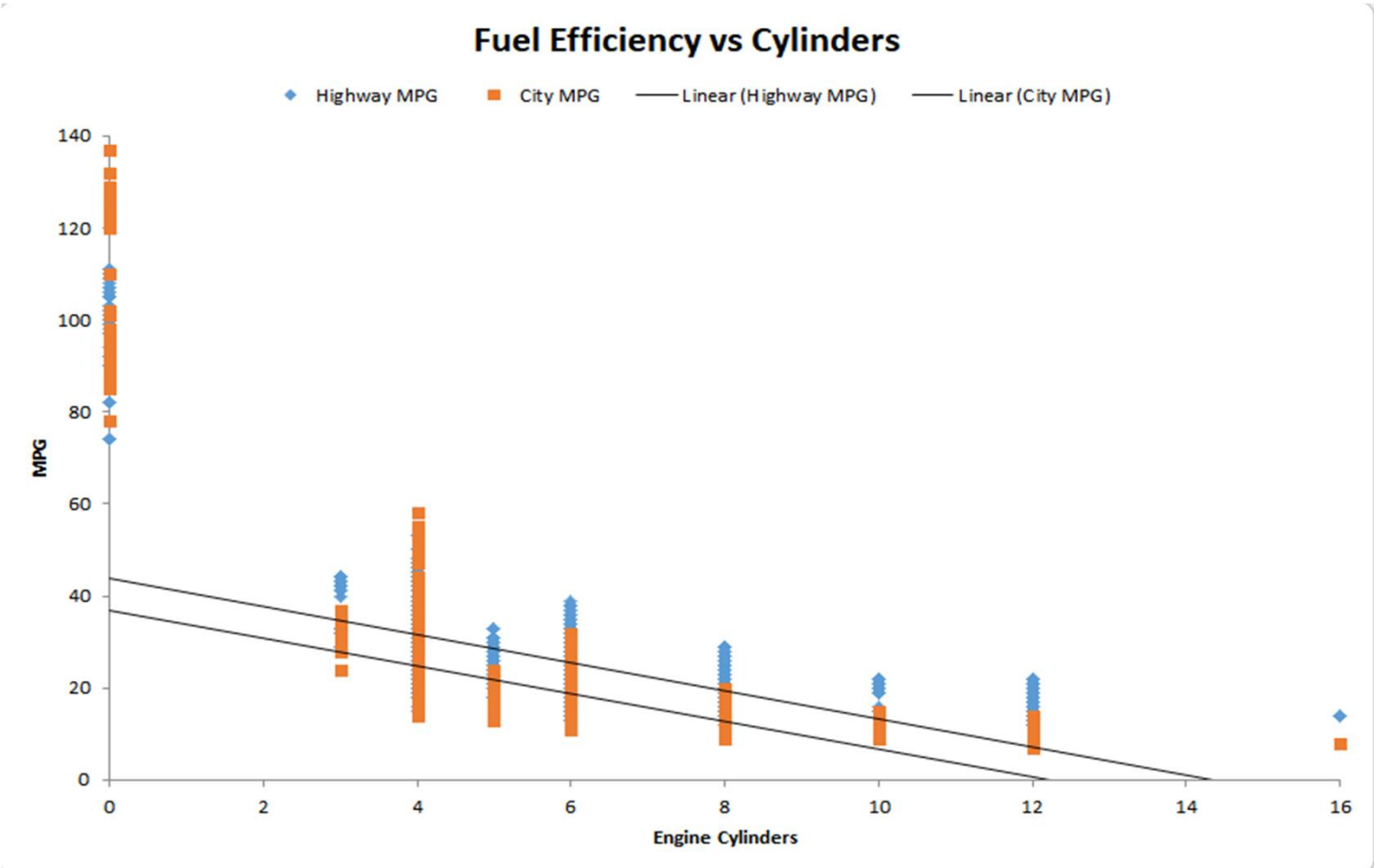


Luxury Brands like Bugatti, Maybach and Rolls-Royce were the most expensive.

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.

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



For both Highway and City MPG, as the cylinders increase, the fuel efficiency decreases.

Dashboard

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

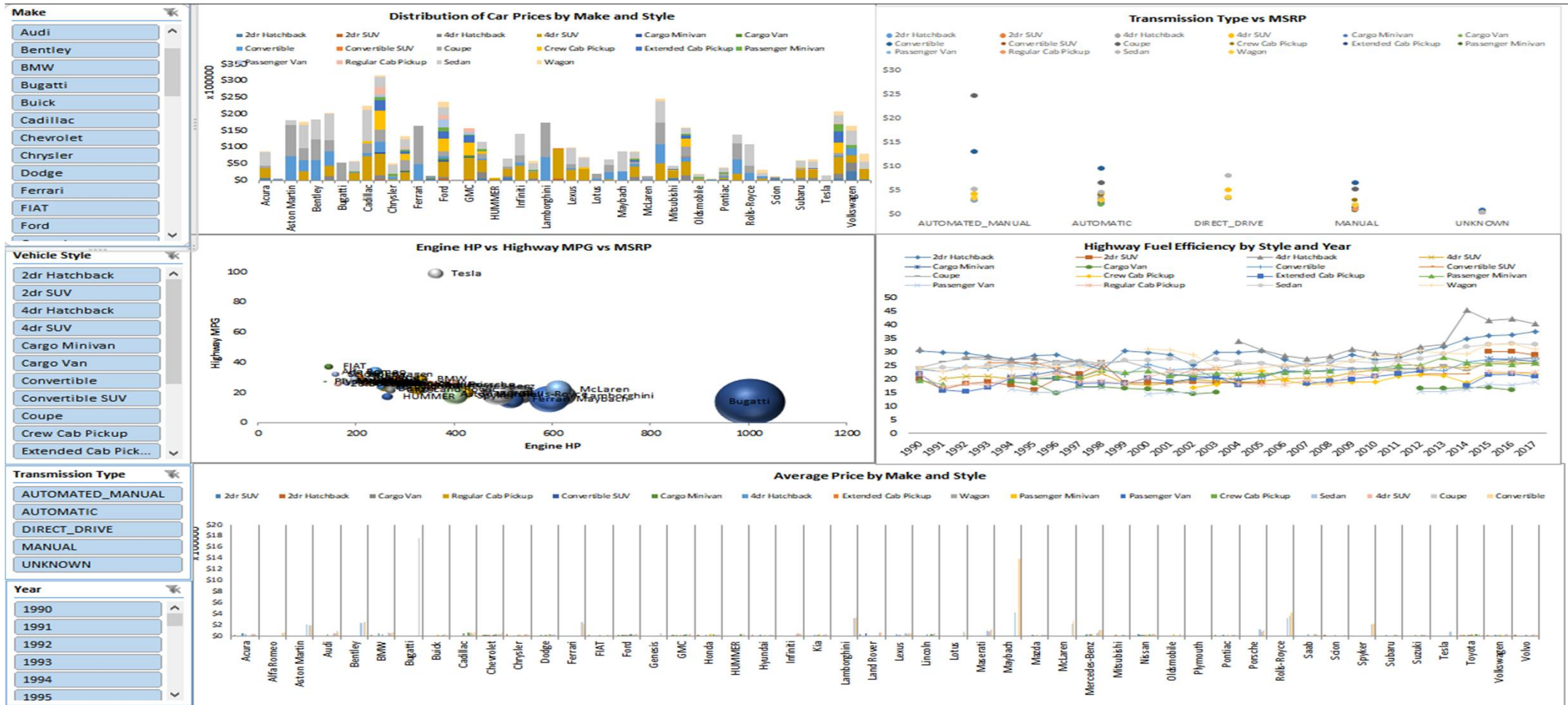
Task 2: Which car brands have the highest and lowest average MSRPs, and how does this vary by body style?

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?

Dashboard



Summary

This particular project involved an extensive usage of Excel and Statistics.

Most of the analysis was done by using Pivot Tables and Charts, for which a lot of research was done.

Creating a dashboard was a new and interesting challenge, which was difficult, and could have been easily done using PowerBI or Tableau.

In Statistics, achieving the Regression Analysis for many columns together was difficult, but after researching it, was able to understand and complete the task.

The major challenge faced was in understanding the data. Updating the NULL/Missing values proved a challenge as most of the data required internet research. There were some data irregularities, which were dealt with.

Overall, this was a very challenging but creative and informative project.

Link to Excel File

[Car Data.xlsx](#)



THANK YOU