Analysing the Impact of Car Features on Price and Profitability

PROJECT MADE BY:

SAMIYA ALAM

samiyaalam1710@gmail.com

Project Description

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. In recent years, there has been a growing trend towards electric and hybrid vehicles and increased interest in alternative fuel sources such as hydrogen and natural gas. At the same time, traditional gasoline powered cars remain dominant in the market, with varying fuel types and grades available to consumers.

The project aims to analyze a dataset containing various features of cars, including their make, model, year, engine specifications, and pricing. The primary objective is to understand the factors that influence car prices, fuel efficiency, and popularity, and to derive actionable insights for car manufacturers to optimize pricing and product development decisions.

Business Problem or Question

The central question addressed by this project is: "How can a car manufacturer optimize pricing and product development decisions to maximize profitability while meeting consumer demand?"

Data Pre-Processing

First, I downloaded the dataset provided in CSV format and loaded it into Excel. I adjusted the column widths for better visibility and applied a data cleaning strategy using standard Excel functions. This involved counting and removing blank rows using Excel's special functions.

Initially, the dataset contained 11,915 rows and 16 columns. After cleaning, it was reduced to 11,813 rows.

I converted the dataset into a table format and highlighted the header rows for clarity. I then analysed the dataset to examine the impact of various car features, such as model, vehicle type, engine horsepower, market price, production year, fuel efficiency (MPG), and number of doors, on different car models and manufacturers.

CLEAN DATASET FILE:

https://docs.google.com/spreadsheets/d/11Blx97K6GLaE_PxCNDeU5VnzUaNkkUxU/edit?usp=sharing&ouid=118309411958556729568&rtpof=true&sd=true

Make	Model	Year	Engine Fuel Type	Engine HP	Engine Cylinders	Transmission Typ	Driven_Wheel	Number of Door	Mark
BMW	1 Series M	2011	premium unleaded (required)	335	6	MANUAL	rear wheel drive	2	Facto
BMW	1 Series	2011	premium unleaded (required)	300	6	MANUAL	rear wheel drive	2	Luxur
BMW	1 Series	2011	premium unleaded (required)	300	6	MANUAL	rear wheel drive	2	Luxur
BMW	1 Series	2011	premium unleaded (required)	230	6	MANUAL	rear wheel drive	2	Luxur
BMW	1 Series	2011	premium unleaded (required)	230	6	MANUAL	rear wheel drive	2	Luxur
BMW	1 Series	2012	premium unleaded (required)	230	6	MANUAL	rear wheel drive	2	Luxur
BMW	1 Series	2012	premium unleaded (required)	300	6	MANUAL	rear wheel drive	2	Luxur
BMW	1 Series	2012	premium unleaded (required)	300	6	MANUAL	rear wheel drive	2	Luxur
BMW	1 Series	2012	premium unleaded (required)	230	6	MANUAL	rear wheel drive	2	Luxur
BMW	1 Series	2013	premium unleaded (required)	230	6	MANUAL	rear wheel drive	2	Luxur
BMW	1 Series	2013	premium unleaded (required)	300	6	MANUAL	rear wheel drive	2	Luxur
BMW	1 Series	2013	premium unleaded (required)	230	6	MANUAL	rear wheel drive	2	Luxur
BMW	1 Series	2013	premium unleaded (required)	300	6	MANUAL	rear wheel drive	2	Luxur
BMW	1 Series	2013	premium unleaded (required)	230	6	MANUAL	rear wheel drive	2	Luxur
BMW	1 Series	2013	premium unleaded (required)	230	6	MANUAL	rear wheel drive	2	Luxur
BMW	1 Series	2013	premium unleaded (required)	320	6	MANUAL	rear wheel drive	2	Luxur
BMW	1 Series	2013	premium unleaded (required)	320	6	MANUAL	rear wheel drive	2	Luxur
Audi	100	1992	regular unleaded	172	6	MANUAL	front wheel drive	4	Luxur
Audi	100	1992	regular unleaded	172	6	MANUAL	front wheel drive	4	Luxur
Audi	100	1992	regular unleaded	172	6	AUTOMATIC	all wheel drive	4	Luxur
Audi	100	1992	regular unleaded	172	6	MANUAL	front wheel drive	4	Luxur
Audi	100	1992	regular unleaded	172	6	MANUAL	all wheel drive	4	Luxur
Audi	100	1993	regular unleaded	172	6	MANUAL	front wheel drive	4	Luxur
Audi	100	1993	regular unleaded	172	6	AUTOMATIC	all wheel drive	4	Luxur
Audi	100	1993	regular unleaded	172	6	MANUAL	front wheel drive	4	Luxur
Audi	100	1993	regular unleaded	172	6	MANUAL	front wheel drive	4	Luxur
Audi	100	1993	regular unleaded	172	6	MANUAL	all wheel drive	4	Luxur
Audi	100	1994	regular unleaded	172	6	AUTOMATIC	front wheel drive	4	Luxur
Audi	100	1994	regular unleaded	172	6	MANUAL	all wheel drive	4	Luxur
Audi	100	1994	regular unleaded	172	6	MANUAL	front wheel drive	4	Luxur
Audi	100	1994	regular unleaded	172	6	AUTOMATIC	front wheel drive	4	Luxur
Audi	100	1994	regular unleaded	172	6	AUTOMATIC	all wheel drive	4	Luxur
FIAT	124 Spider	2017	premium unleaded (recommended)	160	4	MANUAL	rear wheel drive	2	Perfo
FIAT	124 Spider	2017	premium unleaded (recommended)	160	4	MANUAL	rear wheel drive	2	Perfo
FIAT	124 Spider	2017	premium unleaded (recommended)	160	4	MANUAL	rear wheel drive	2	Perfo

Approach

This project employed a comprehensive analytical approach to extract valuable insights from the dataset. Initially, a thorough data cleaning process was conducted to ensure data quality and reliability.

Following data cleaning, Descriptive statistics were performed to understand the distribution, variability, and basic characteristics of the data. This initial exploration set the stage for more advanced analyses.

Various visualization techniques were employed to uncover relationships and trends within the data. Using Tableau, scatter plots, line charts, and packed bubble charts were created to visually represent relationships between different car features, such as horsepower, MPG, and MSRP, across various brands and models. These visualizations helped identify patterns and correlations that were not immediately apparent from the raw data.

To further investigate the factors influencing car prices and popularity, regression and correlation analyses were conducted. These statistical methods quantified the relationships between different features and identified key drivers of car prices and consumer interest. Driving and vehicle performance data were analyzed to uncover factors influencing fuel efficiency.

Overall, this project aimed to provide a comprehensive understanding of the impact of car data analysis on the automotive industry, offering insights, solutions, and guidance for industry stakeholders to harness the power of data for positive transformations.

Tech-Stack Used

- **Excel**: Used for data cleaning, preprocessing, and analysis.
- **Tableau**: Utilized for advanced visualizations and some interactive dashboards.

Dataset Description

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:** 11,159
- **Number of variables:** 16
- **File type:** CSV (Comma Separated Values)

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

Tasks: Analysis

Before diving into the analysis of the given dataset, it is important to perform thorough data cleaning to ensure accurate and reliable results. You need to build an interactive dashboard in Excel from the tasks given below:

Insight Required: How does the popularity of a car model vary across different market categories?
Insight Required: What is the relationship between a car's engine power and its price?
Insight Required: Which car features are most important in determining a car's price?
Insight Required: How does the average price of a car vary across different manufacturers?
Insight Required: What is the relationship between fuel efficiency and the number of cylinders in a car's engine?

Building the 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?

INSIGHTS AND RESULTS

1) 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.

I created a pivot table in Excel by loading the dataset and using the PivotTable feature. I placed "Market Category" in the rows field, "Model" in the values field (set to count), and "Popularity" in the values field (set to sum). This gave me a summary of the number of car models and their total popularity scores for each market category.

THE COMPLETE PIVOT TABLE IS ATTACHED IN THE EXCEL FILE LINK PROVIDED AT THE END OF THIS DOCUMENT

Image for reference below:

Market category	Count of Model	Average of Popularity
Crossover	1103	1529.030825
Crossover,Diesel	7	873
Crossover,Exotic,Luxury,High-Performance	1	238
Crossover,Exotic,Luxury,Performance	1	238
Crossover,Factory Tuner,Luxury,High-Performance	26	1823.461538
Crossover,Factory Tuner,Luxury,Performance	5	2607.4
Crossover,Factory Tuner,Performance	4	210
Crossover,Flex Fuel	64	2073.75
Crossover,Flex Fuel,Luxury	10	1173.2
Crossover,Flex Fuel,Luxury,Performance	6	1624
Crossover,Flex Fuel,Performance	6	5657
Crossover,Hatchback	72	1675.694444
Crossover,Hatchback,Factory Tuner,Performance	6	2009
Crossover,Hatchback,Luxury	7	204
Crossover,Hatchback,Performance	6	2009
Crossover,Hybrid	42	2563.380952
Crossover,Luxury	410	884.5487805
Crossover,Luxury,Diesel	33	2195.848485
Crossover,Luxury,High-Performance	9	1037.222222
Crossover,Luxury,Hybrid	24	630.9166667
Crossover,Luxury,Performance	113	1344.849558
Crossover,Luxury,Performance,Hybrid	2	3916
Crossover,Performance	69	2585.956522
Diesel	84	1730.904762
Diesel,Luxury	51	2275
Exotic,Factory Tuner,High-Performance	21	1046.380952
Exotic,Factory Tuner,Luxury,High-Performance	52	517.5384615
Exotic,Factory Tuner,Luxury,Performance	3	520
Exotic,Flex Fuel,Factory Tuner,Luxury,High-Performance	13	520

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

I created a combo chart in Tableau to visualize the relationship between market category and popularity. The green bar chart represents the average popularity, while the red line chart shows the count of car models. This was done using Tableau Public, allowing me to effectively analyze and visualize the impact of car features on popularity by comparing the count and average popularity of each model within different market categories.

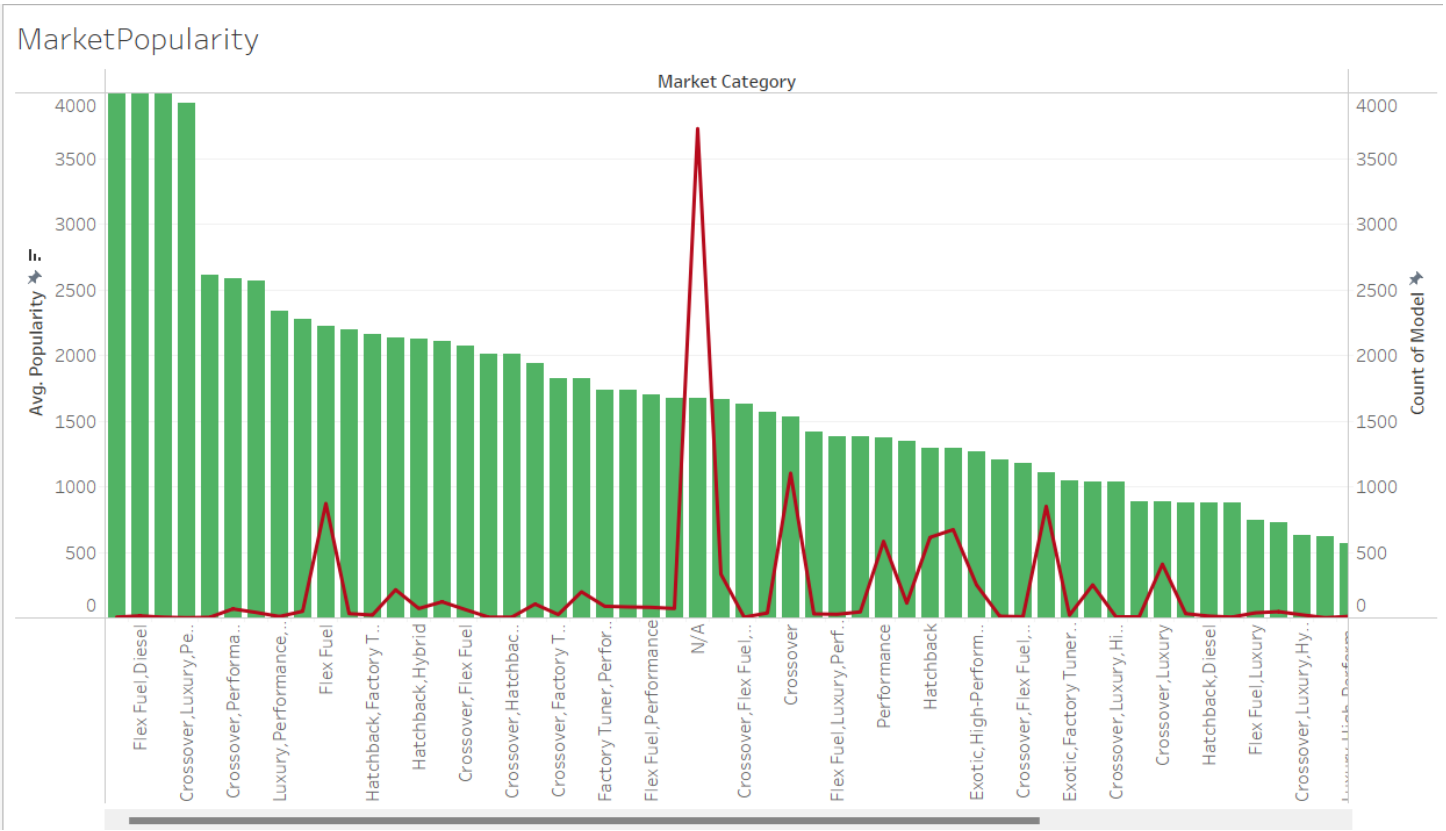
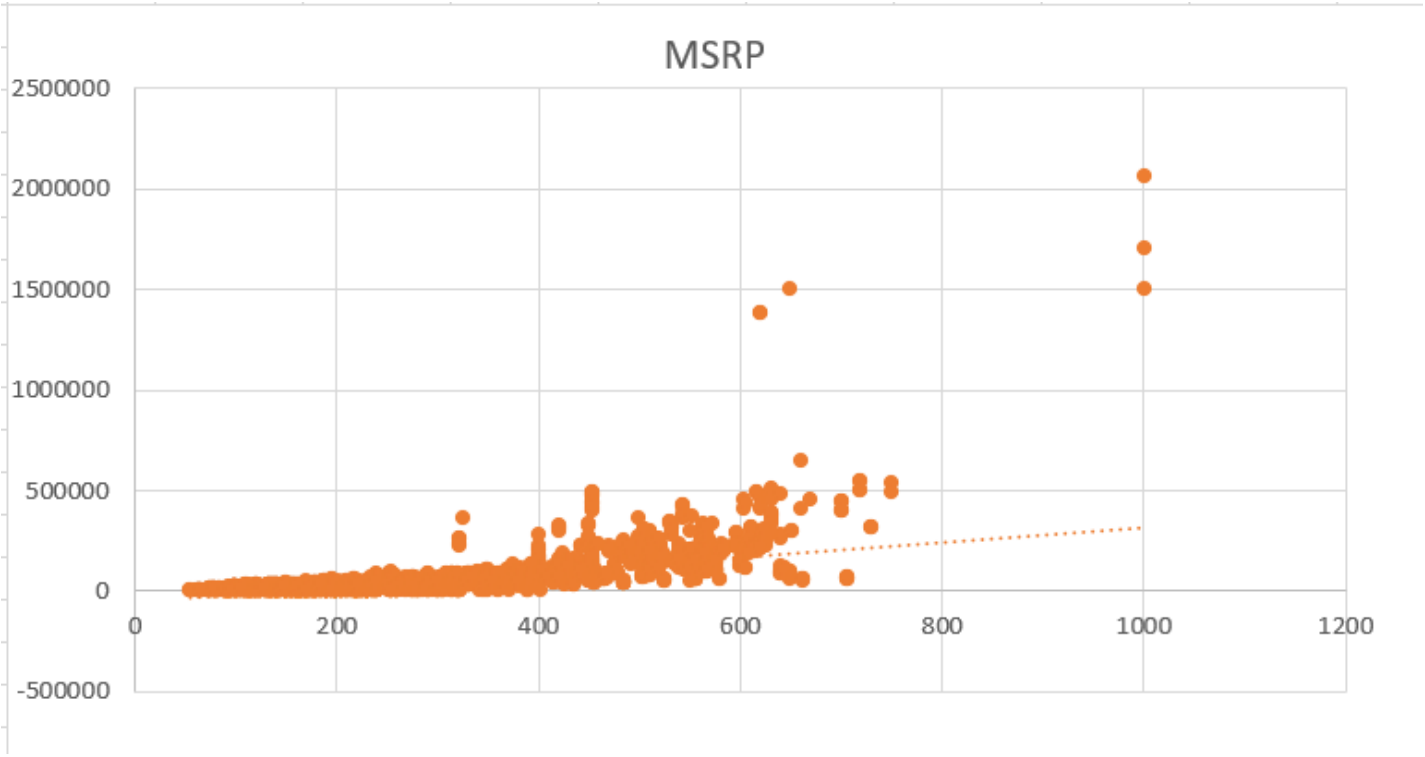


Tableau Public LINK:
https://public.tableau.com/views/cardataanalysis_17213254958860/MarketPopularity?:language=en-US&publish=yes&:sid=&:redirect=auth&:display_count=n&:origin=viz_share_link

Insight Required: 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.

In excel I created a scatter chart by plotting engine power (HP) on the x-axis and price (MSRP) on the y-axis. I then added a trendline to the chart to visualize and better understand the relationship between these two variables.



Complete task can be seen in the excel file I attached in the end.

Insight Required: 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.

I conducted a regression analysis in Excel to identify which car features have the strongest relationship with a car's price. I included variables such as *Engine HP, Engine Cylinders, Number of Doors, city mpg, highway MPG, MSRP* in the regression model. After running the analysis, I created a bar chart to display the coefficient values for each variable. This chart visually represents the relative importance of each feature in determining the car's price.

Results below:

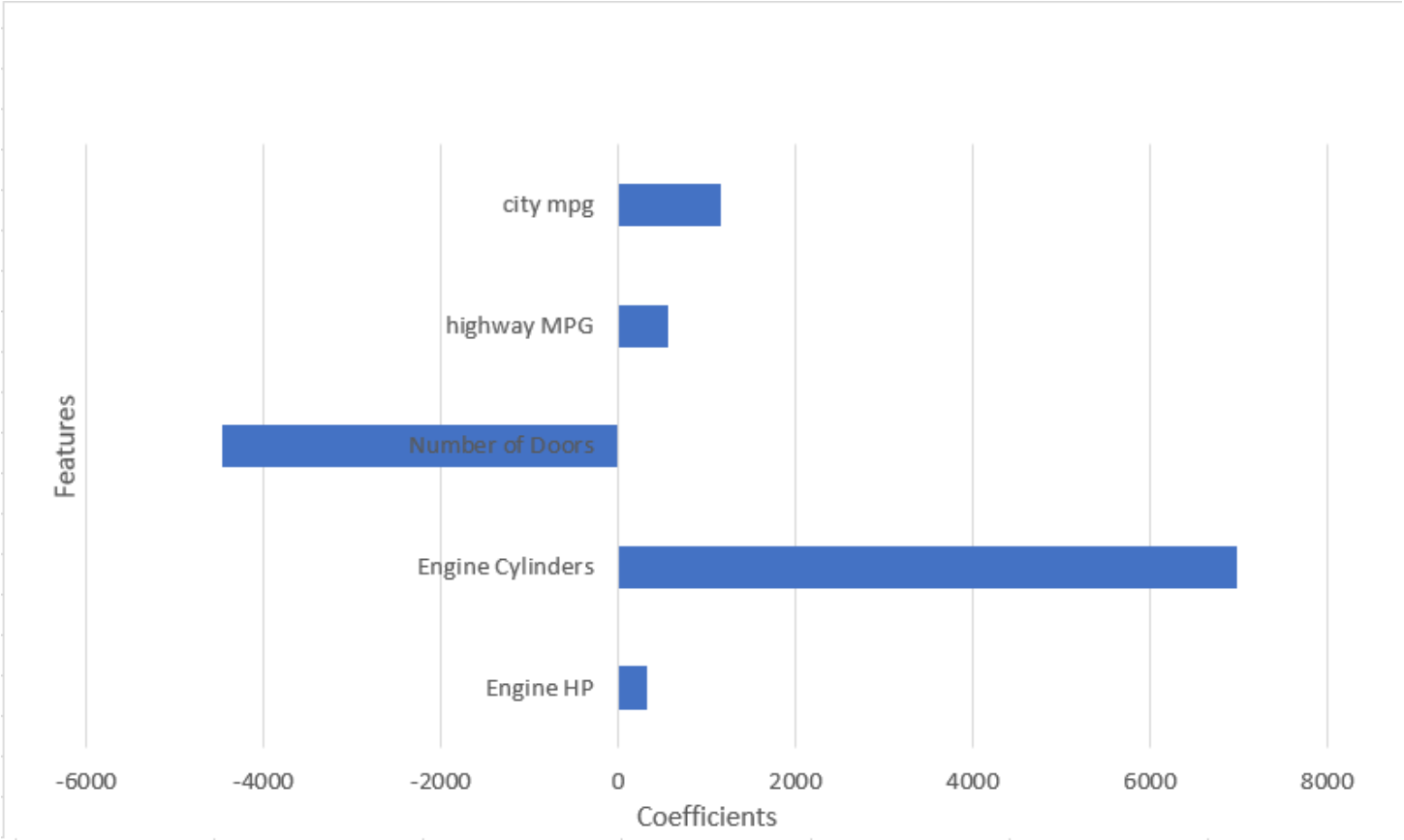
SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.680708139
R Square	0.46336357
Adjusted R Square	0.463136297
Standard Error	44170.77827
Observations	11812

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	5	1.98891E+13	3.97782E+12	2038.799457	0
Residual	11806	2.30342E+13	1951057653		
Total	11811	4.29233E+13			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-101601.736	3684.351697	-27.57655738	2.765E-162	-108823.673	-94379.79896	-108823.673	-94379.79896
Engine HP	322.7465574	6.01767382	53.63310924	0	310.9509241	334.5421906	310.9509241	334.5421906
Engine Cylinders	6989.177662	439.6449924	15.89732121	2.53591E-56	6127.400961	7850.954363	6127.400961	7850.954363
Number of Doors	-4472.158125	465.7180593	-9.602715711	9.35015E-22	-5385.042338	-3559.273912	-5385.042338	-3559.273912
highway MPG	570.1808088	105.7839778	5.390048859	7.17937E-08	362.826764	777.5348535	362.826764	777.5348535
city mpg	1163.755457	121.9978136	9.539150113	1.72109E-21	924.61962	1402.891294	924.61962	1402.891294



Insight Required: How does the average price of a car vary across different manufacturers?

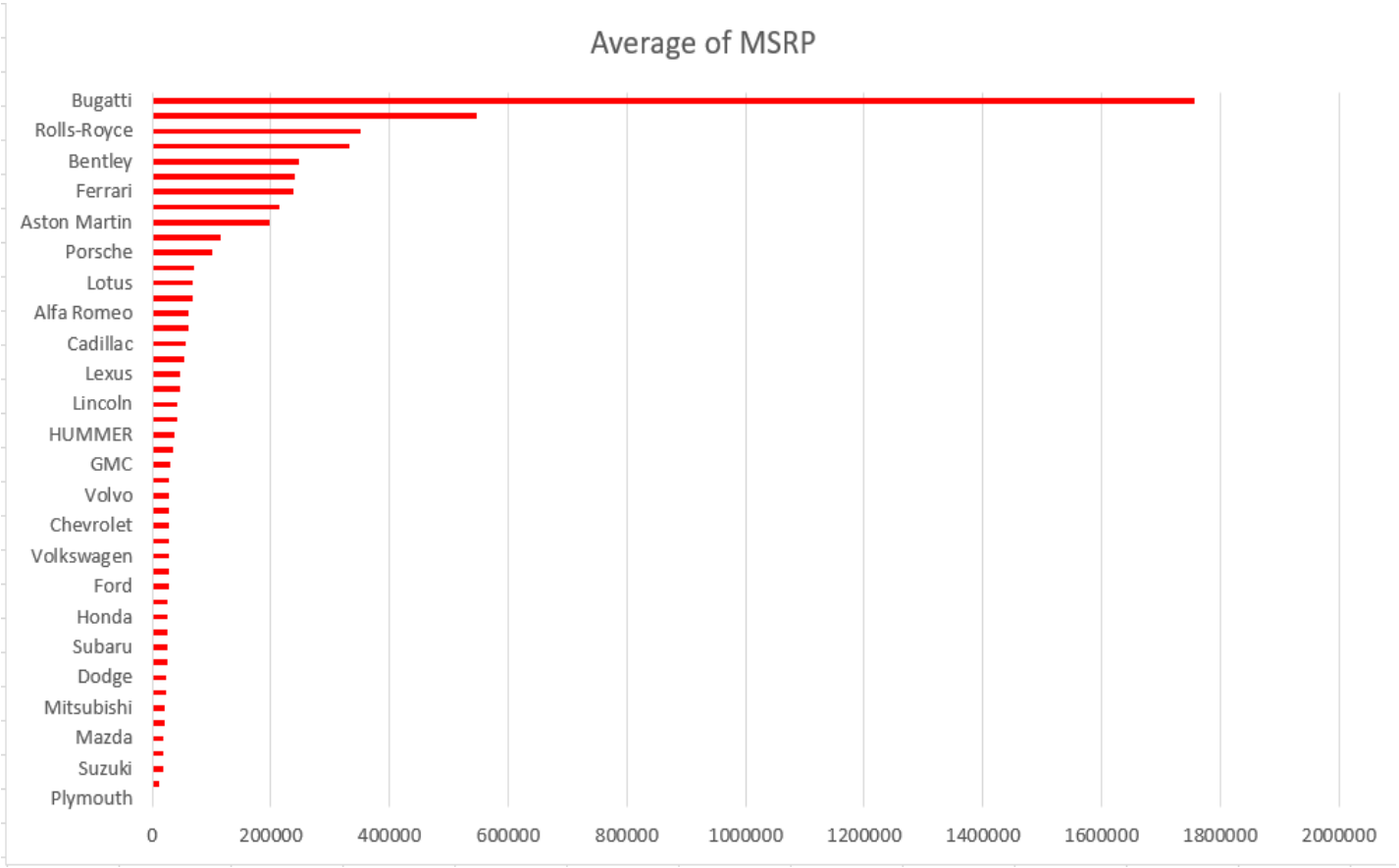
Task 4.A: Create a pivot table that shows the average price of cars for each manufacturer.

I created a pivot table in Excel to show the average price of cars for each manufacturer. I loaded the dataset, selected the data range, and inserted a PivotTable.

Manufacturer	Average of MSRP
Bugatti	1757223.667
Maybach	546221.875
Rolls-Royce	351130.6452
Lamborghini	331567.3077
Bentley	247169.3243
McLaren	239805
Ferrari	237383.8235
Spyker	213323.3333
Aston Martin	197910.3763
Maserati	114207.7069
Porsche	101622.3971
Mercedes-Benz	71537.80966
Lotus	69188.27586
Land Rover	67823.21678
Alfa Romeo	61600
BMW	61546.76347
Cadillac	56231.31738
Audi	53452.1128
Lexus	47549.06931
Genesis	46616.66667
Lincoln	42494.37179
Infiniti	42394.21212
HUMMER	36464.41176
Acura	34887.5873
GMC	30493.29903
Toyota	28946.15343
Volvo	28541.16014
Nissan	28513.36679
Chevrolet	28273.35695
Buick	28206.61224
Volkswagen	28076.2
Saab	27413.5045
Ford	27393.42051
Chrysler	26722.96257
Honda	26629.81879
Kia	25112.38938
Subaru	24827.50391
Hyundai	24597.0363
Dodge	22390.05911
FIAT	22206.01695
Mitsubishi	21215.47143
Scion	19932.5
Mazda	19719.05707
Pontiac	19321.54839
Suzuki	17900.9569
Oldsmobile	11542.54
Plymouth	3122.902439

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

I created a bar chart to visualize the relationship between manufacturer and average price. Using the pivot table data, I inserted a bar chart and formatted it to clearly show the average prices across different manufacturers.



Insight Required: 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.

I used Tableau to analyze the relationship between fuel efficiency and the number of cylinders in a car's engine. I added a trendline to the scatter plot to visually estimate the slope of the relationship and assess its significance.

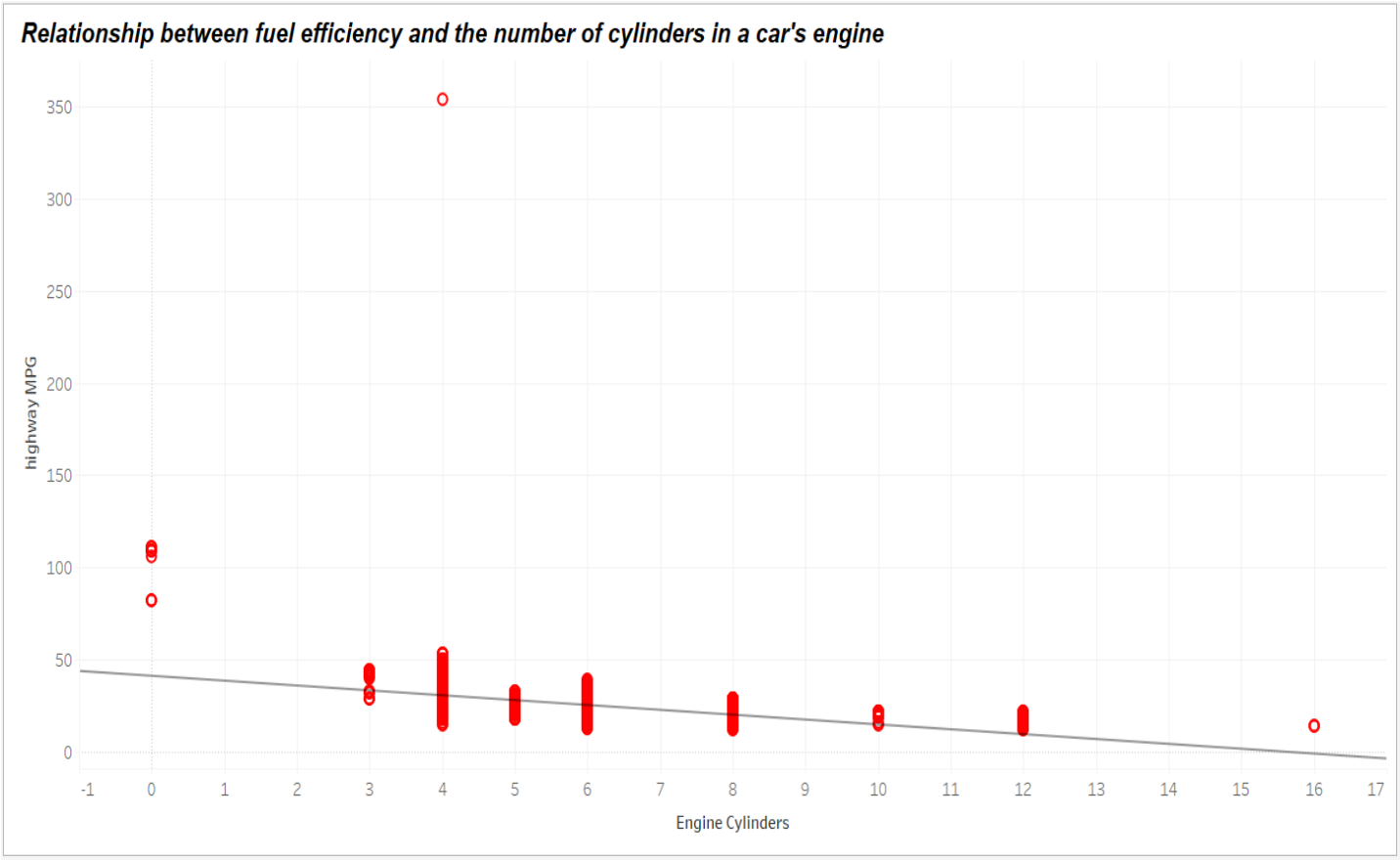


Tableau public LINK:
https://public.tableau.com/views/cardataanalysis_17213254958860/EfficeiencyVScylinders?:language=en-US&publish=yes&:sid=&:redirect=auth&:display_count=n&:origin=viz_share_link

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

I used Tableau's built-in analytics tools to calculate the correlation coefficient between the number of cylinders and highway MPG.

Interpret Result The correlation coefficient provided a quantitative measure of the strength and direction of the relationship, indicating how fuel efficiency varies with the number of engine cylinders.

Building the Dashboard:

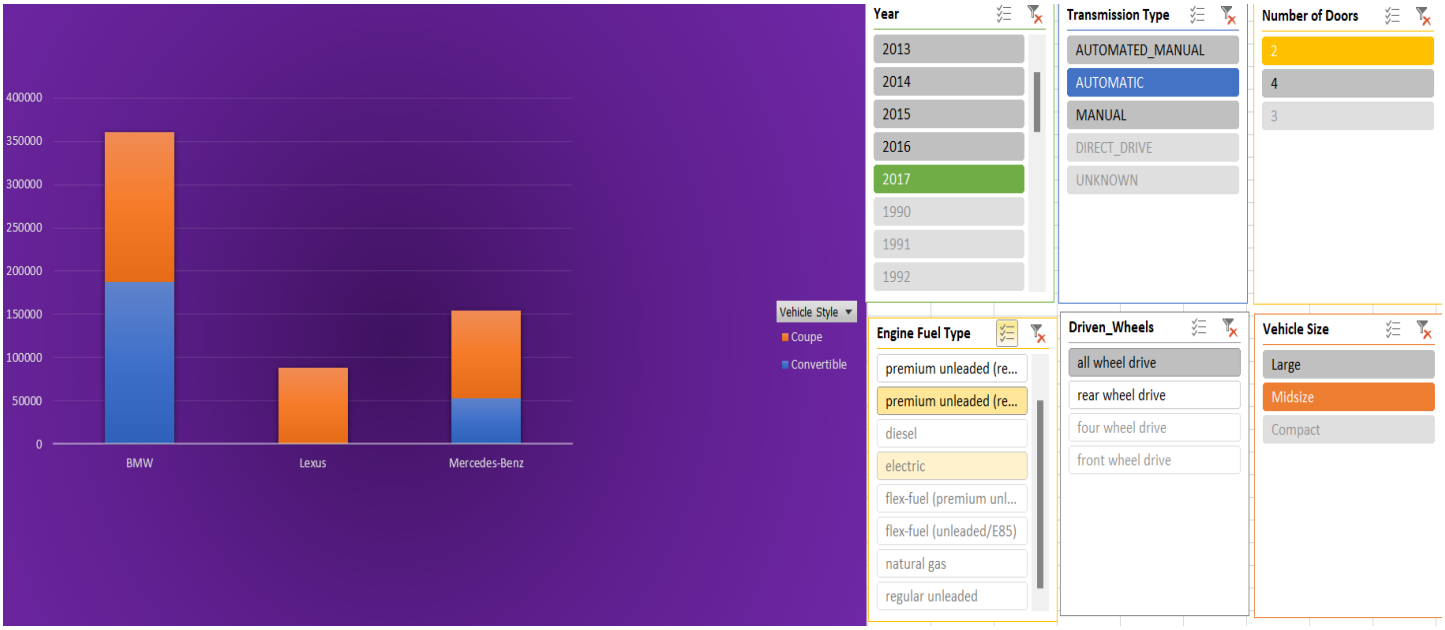
Now for the Next portion of the Project, I needed to create the Interactive Dashboards.

I Used 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?

I analyzed how the distribution of car prices varies by brand and body style using Excel. I used the PivotTable data to insert a stacked column chart, showing the distribution of car prices by brand and body style. I added filters and slicers to make the chart interactive, allowing for easy exploration of the data by brand and body style.

Applying few filters:

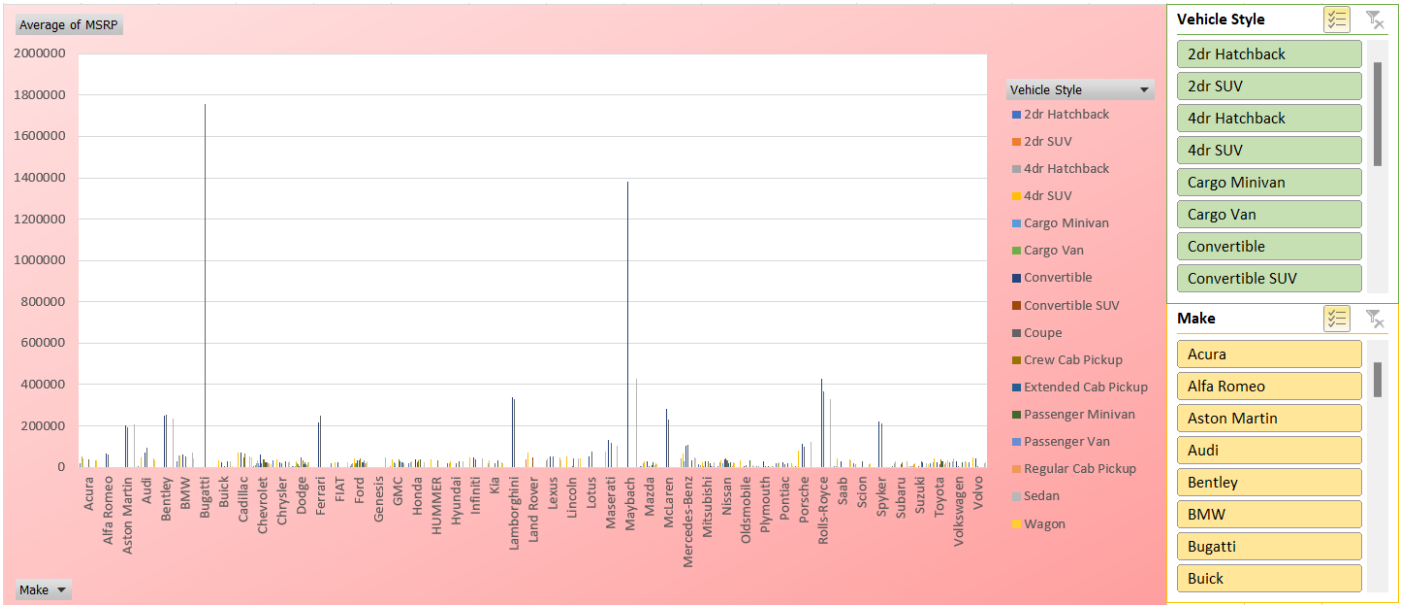


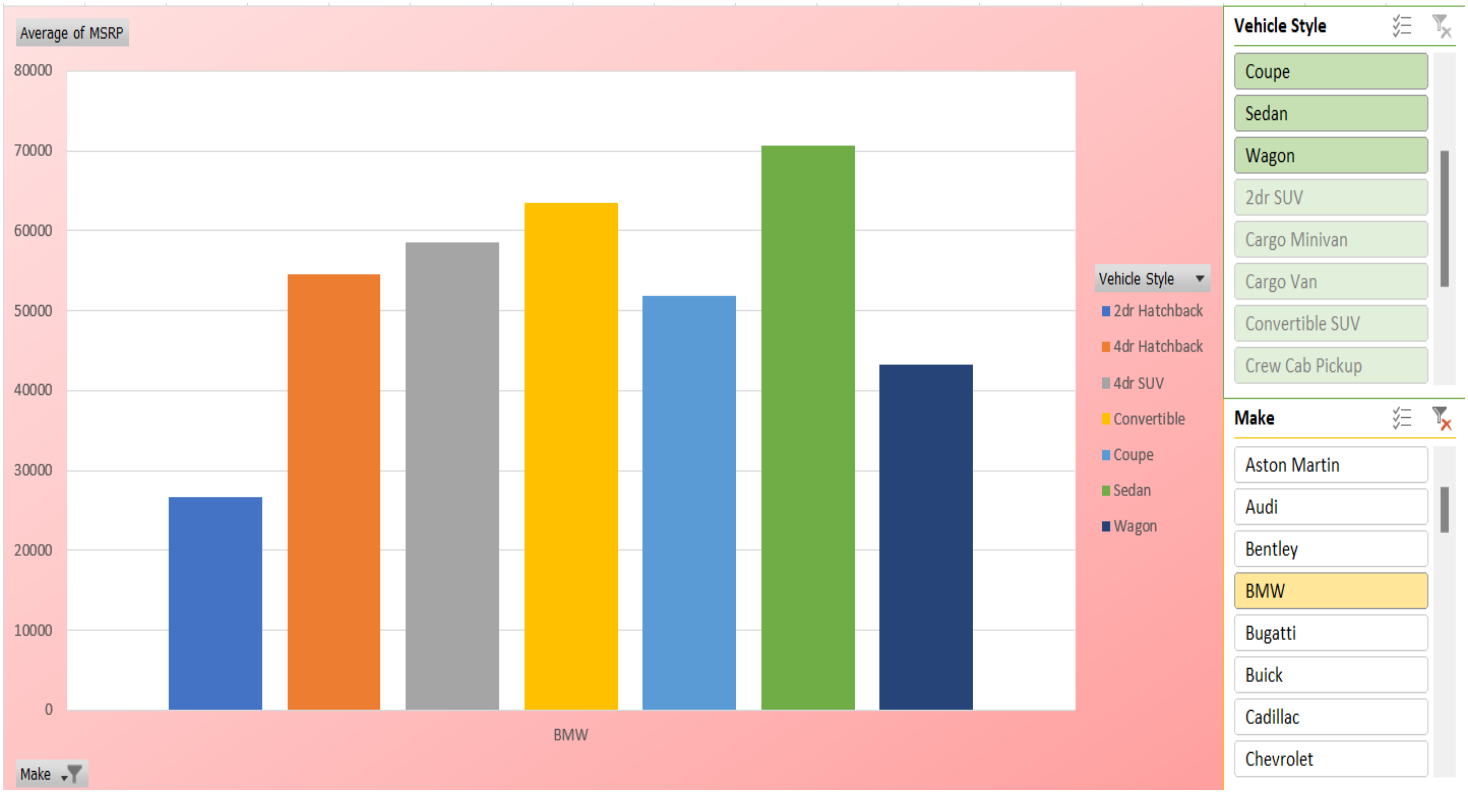
Task 2: Which car brands have the highest and lowest average MSRPs, and how does this vary by body style?

I analyzed which car brands have the highest and lowest average MSRPs and how this varies by body style using Excel. I used the PivotTable data to insert a clustered column chart, comparing the average MSRPs across different car brands and body styles.

I formatted the chart to clearly show the variations in average MSRPs, making it easier to identify which brands have the highest and lowest average prices for each body style.

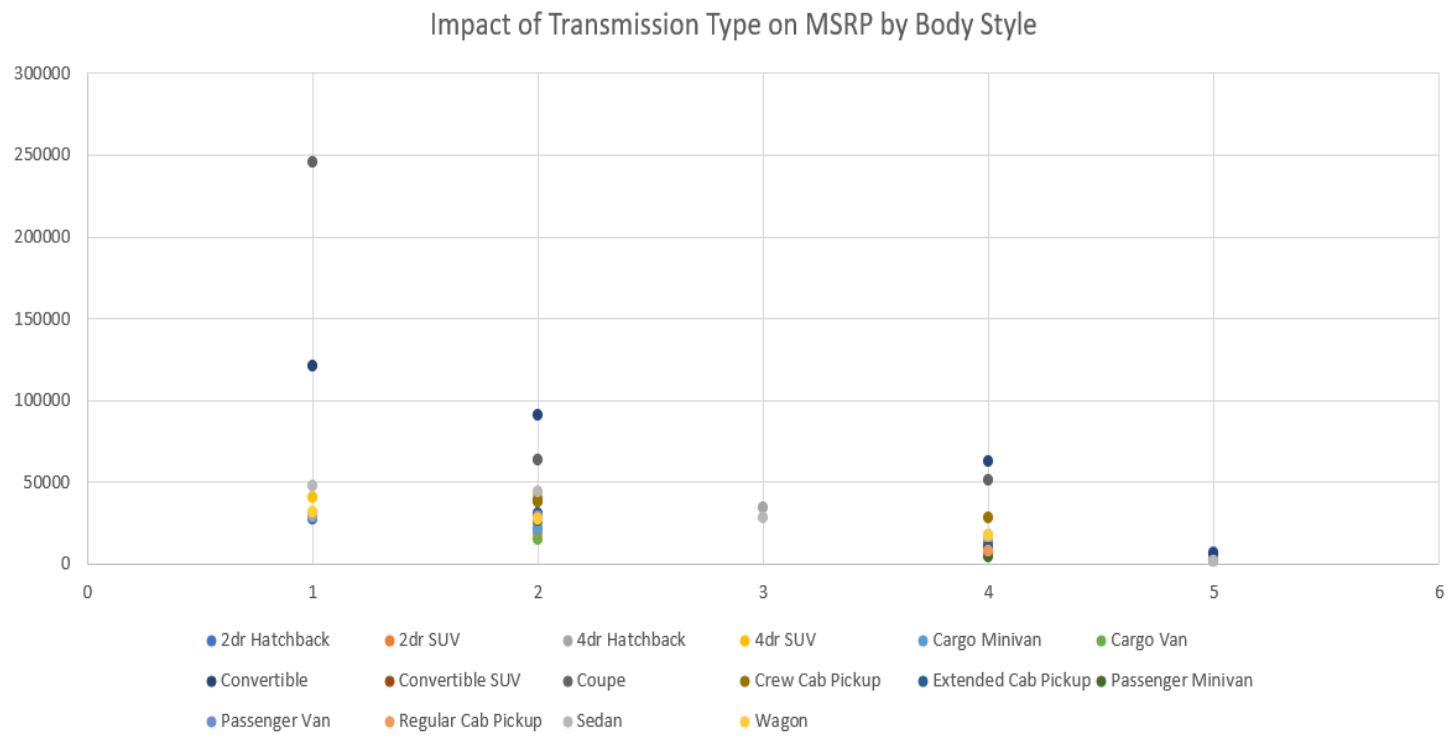
Pictures below shows chart without any filters and with filters for comparisons for different brands.





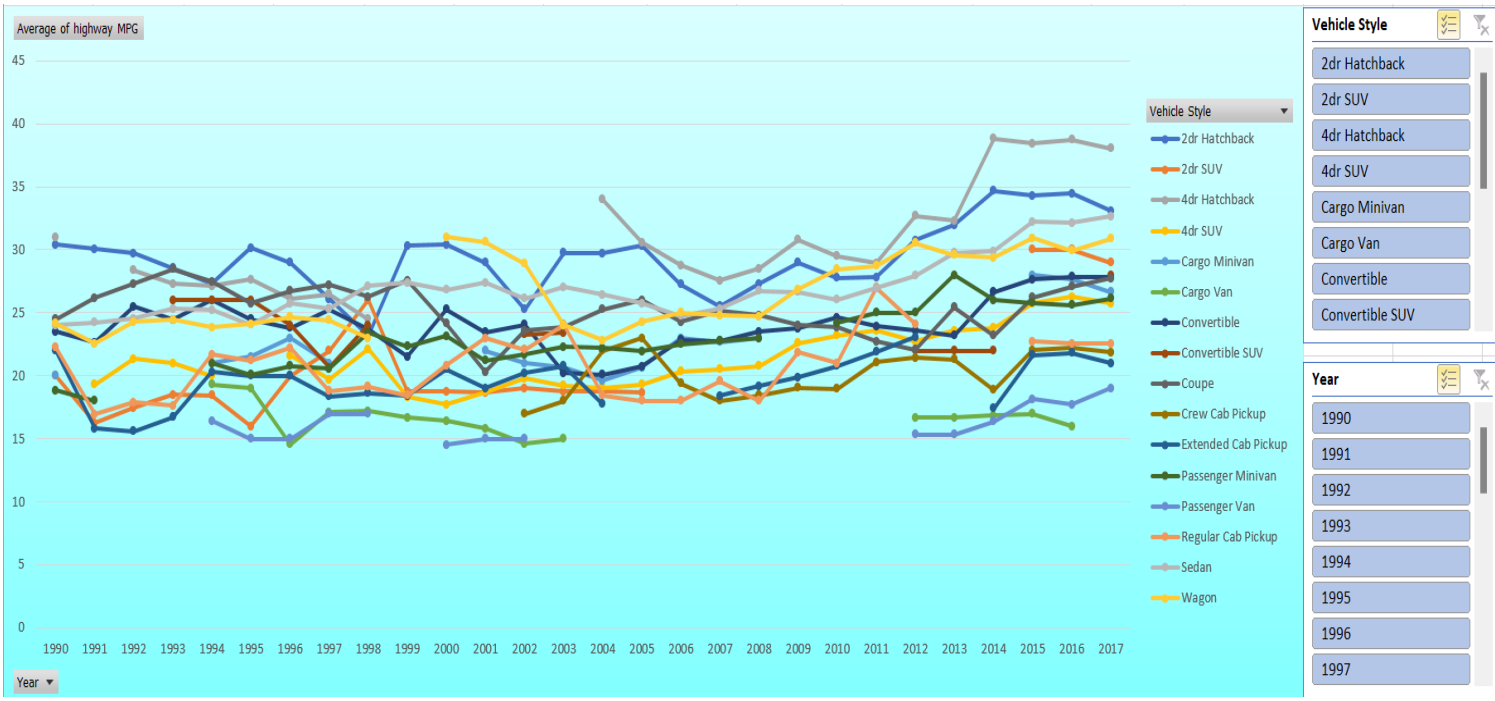
Task 3: How do the different feature such as transmission type affect the MSRP, and how does this vary by body style?

I used the PivotTable data to insert a scatter plot chart. I differentiated the data points by assigning different symbols for each body style. I formatted the scatter plot to clearly visualize the relationship between MSRP and transmission type, showing how this relationship varies by body style.



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

I used the PivotTable data to insert a line chart. I formatted the line chart to show the trend of fuel efficiency over time for each body style, clearly illustrating how fuel efficiency varies by body style and model year.



Above line graph shows fuel efficiency of couple, passenger van and wagon body style between the years 2012 to 2017.

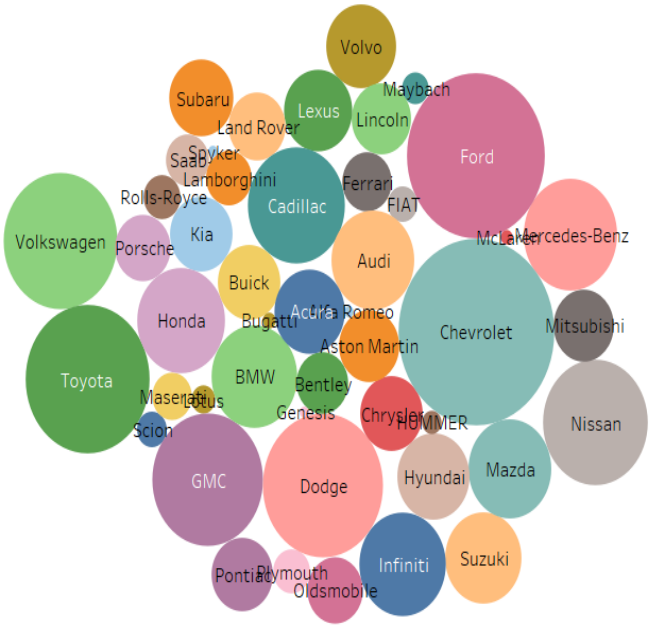
Task 5: How does the car's horsepower, MPG, and price vary across different Brands?

I used Tableau to analyze how car's horsepower, MPG, and price vary across different brands. I calculated the average horsepower, MPG, and MSRP for each car brand using Tableau's calculation features.

I assigned different colors to each brand for clear differentiation. I labeled the bubbles with the car model names to provide more detailed information.

I formatted the bubble chart to clearly show the variations in horsepower, MPG, and price across different brands.

Comparative Analysis of Car Brands: Horsepower, Fuel Efficiency, and MSRP



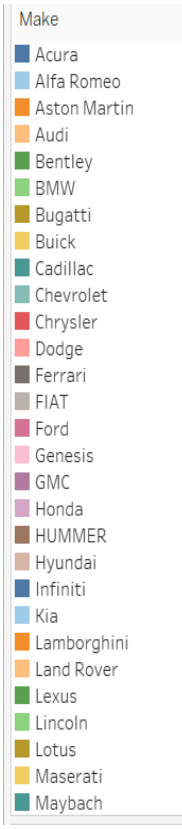


Tableau Public LINK:

https://public.tableau.com/views/cardataanalysis_17213254958860/dashboard5?:language=en-US&publish=yes&:sid=&:redirect=auth&:display_count=n&:origin=viz_share_link

****COMPLETE EXCEL FILE WITH TASKS****

LINK: https://docs.google.com/spreadsheets/d/1pR1acg7GVC3R-UgrncbirgT_tH2cxjSk/edit?usp=sharing&ouid=118309411958556729568&rtpof=true&sd=true

RESULT

The analysis yielded several key insights. Features like engine power, number of cylinders, and vehicle size significantly impact car prices. Higher horsepower and more cylinders correlate with higher MSRPs, guiding premium pricing

strategies for these models. In terms of popularity by market category, luxury and performance categories have higher popularity scores, indicating a consumer preference that manufacturers can capitalize on by enhancing features and marketing efforts in these segments. Additionally, a correlation exists between engine power and car price, helping manufacturers understand pricing dynamics and set competitive prices for high-performance vehicles. Lastly, fuel efficiency trends vary significantly across body styles and model years, with certain body styles showing improved MPG over time, aiding manufacturers in targeting fuel-efficient designs.

Limitations and Uncertainties

- **Data Quality:** The accuracy of results is contingent on the quality and completeness of the dataset.
- **External Factors:** Economic conditions, fuel prices, and technological advancements can impact car features and prices, which were not accounted for in this analysis.
- **Model Generalizability:** The models may not generalize well to all market conditions or geographic locations.

Future Directions

- **Include More Variables:** Incorporate additional variables like safety features, brand reputation, and consumer reviews for a more comprehensive analysis.
- **Advanced Modeling:** Use machine learning models to predict car prices and popularity with higher accuracy.

****COMPLETE EXCEL FILE WITH TASKS****

LINK: https://docs.google.com/spreadsheets/d/1pR1acg7GVC3R-UgrncbirgT_tH2cxjSk/edit?usp=sharing&ouid=118309411958556729568&rtpof=true&sd=true

*******THANK YOU*******

SAMIYA ALAM