SUMIT SHAMLAL CHAURE

Analyzing the Impact of Car Features on Price and Profitability

Trainity Project 7 (Final Project - 3) – Dashboard & EDA Visualization

Figure 1 - https://trainity.link/data/project07

INDEX

Contents

LOOM PPT (Explanation – Watch these to get more anaalysis inisghts & for dashboard interaction)

- Introduction (Project Summary, Data cleaning, Aim & Introduction)
- Questions (Main Task Query & Actual output along-with analysis explanation)
- Dashboard (Interactive dashboards can be explained either by visting the published link and exploring or by watching the video explanation to get the insights more detailed.)

DESCRIPTION (Loom)

Problem Statement:

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.

For the given dataset, as a Data Analyst, the client has asked How can a car manufacturer optimize pricing and product development decisions to maximize profitability while meeting consumer demand?

This problem could be approached by analysing 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.

Data Cleaning:

This step involves pre-processing the data to make it suitable for analysis. It includes handling missing values, removing duplicates, converting data types if necessary, and possibly feature engineering.

Business & Data Analytics Skills:

The given tasks below based on the business problem would require advanced Excel skills and knowledge of data analysis techniques such as regression analysis, pivot tables, sensitivity analysis, optimization, and time series analysis.

However, by answering these questions and building an interactive dashboard, a data analyst could provide valuable insights to a car manufacturer and help them optimize their pricing and product development decisions to maximize profitability while meeting consumer demand.

Understanding the Dataset:

The dataset 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).

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:** 11,951
- **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

This dataset could be useful for a variety of data analysis tasks, such as:

- Exploring trends in car features and pricing over time
- Comparing the fuel efficiency of different types of cars
- Investigating the relationship between a car's features and its popularity
- Predicting the price of a car based on its features and market category

However, it's important to note that the dataset was last updated in 2017, so it may not reflect current trends or prices in the automotive industry.

Sumit S. C

A data analyst could use this dataset to gain insights into various aspects of the automotive industry, such as:

Analyzing trends in car features and pricing over time: By examining the variables in the dataset, a data analyst could identify how car features and prices have changed over time, which could help manufacturers make informed decisions about product development and pricing.

Comparing the fuel efficiency of different types of cars: By looking at the MPG variables in the dataset, a data analyst could compare the fuel efficiency of different types of cars and identify which types are the most efficient. This could help consumers make informed decisions about which car to purchase.

Investigating the relationship between a car's features and its popularity: By examining the popularity variable in the dataset, a data analyst could identify which features are most popular among consumers and how they affect a car's popularity. This could help manufacturers make informed decisions about product development and marketing.

Predicting the price of a car based on its features and market category: By using the various features and market category variables in the dataset, a data analyst could develop a model to predict the price of a car. This could help manufacturers and consumers understand how different features affect the price of a car and make informed decisions about pricing and purchasing.

Overall, this dataset could be a valuable resource for data analysts interested in exploring various aspects of the automotive industry and could provide insights that could inform decisions related to product development, marketing, and pricing.

Report and Data Story:

After your analysis, you'll create a report that tells a story with your data. This should include your initial problem, your findings, and the insights you've gained. Use visualizations to help tell your story and make your findings more understandable.

Remember, as a data analyst, your goal is not just to answer questions but to provide insights that can drive decision-making. Your analysis should aim to provide actionable insights that can help stakeholders make informed decisions.

Requirements –

1) Project Description:

The aim of the project is to find the use the knowledge of Excel Formulas, Descriptive statistics & various test involved in making better decision during data cleaning & manipulation like regression tests. The project contains a dataset pertaining to the automotive specifications which gives us a glimpse about car features that impacts the MSRP of the vehicle.

The overall dataset gives us a varied array of datapoints like the mileage of various models & market category, the brands, year of the model, transmission & engine cylinders as well as drive modes etc. that influences the customers opinions about the purchase and the popularity that overall determines the sales forecast which the dealers can analyze and decide what should be the ideal price to get more sales and earn profits over the sales.

In our analyze we think about the business part of the client and analyze the impact of various factors & features that will make variation in MSRP and bring out the desired results to either change MSRP or do marketing strategies or some other solutions to maintain or increase the sales on the models.

The dataset used in the project was collected and made available on Kaggle by Cooper Union, a private college located in New York City by the title – "Car Features and MSRP".

The data cleaning part is mentioned separately below as the steps involved are lengthy.

For certain null values I have assumed to use the mode values of columns as in HP category things are usually in certain range but for electric vehicles like Tesla the actual HP is in reality higher than the mode but imputing the values considering the values from online sources will result in partial treatment with other data so I considered imputing the statistics value instead to not make any drastic impact on outliers in data. Same with N/A value in market category (see data cleaning)

2) Approach:

I first analyzed the data and looked for null values, blanks, duplicates and treated them using basic functions like delete cells, find & replace, remove blank rows etc. For certain cell values I

changed them to suit better with other values like – (Different case for column values or improper naming) etc. After confirming that the data has little to less outliers (Only Bugatti had highest HP engine above 1k HP while the maximum data lied below 600) and saved the processed data to work on with the operations. Also, I removed the most irrelevant columns from the dataset like for 1 task we needed to do regression test to see the impact of features which needed me to treat the categorical value as numerical and remove certain columns which had no significant impacts.

The analysis based on excel functions, Descriptive analysis, pivot tables has insights at the bottom of the screengrabs to let the others understand the aim of each analysis. I used Power BI to create the visual dashboards which are interactive as I don't have Power BI pro or Work email I cannot share the dashboard on BI server to let others see them so I made gif recordings of most dashboard also in video PPT I tried to cover most parts.

3) Tech-Stack Used:

Excel – The basic data manipulation, handling and overall pivot charts and the statistics has been handled using MS Excel.

Google Sheets – Used to do the basic data manipulation and to get column stats (gif added)

Power BI – Made the interactive dashboard for the questions asked using BI tool to make them more interactive.

Word – The report is written in word/docx format using MS Word and then exported to pdf.

Loom – To record and share the Video PPT to explain the overall analysis and approach used in the project.

Google Drive – To upload all the essential files attached in the report for reference & pdf upload.

ShareX – To capture gif recording and screengrabs and host the HD image used in the projects to see the details in more depth.

4) Insights:

The summary for each query is given with the screenshot but to summarize the overall thing I concluded that among the total data set we had **11915 data rows, 16 columns & a totals of 190519** cells in the given dataset.

After removing the blanks, duplicates and adjusting the non-relevant columns we made the dataset to ***11199 rows, 16 columns & 179200 cells*** for our final calculation. The data needed to be treated for Q3 as it needed regression analysis/test which operates on numerical values for that I have explained it more in data cleaning.

For more data insight on the questions look at the respective questions for screenshot and pivot charts & the dashboards or the video PPT for the insights.

5) Result:

The Project has given me a good idea about the importance and vast variety of excel usage which helps us to look deep into plain numbers and generate a visually insightful data which can help business to gain knowledge and prepare for future as well as give out trends to focus on from the numerical data. The statistics section has helped me learn about the various concept which are useful for majority of the operation for handling and displaying basic charts and generate a meaningful dashboard as well as the use of Descriptive stats to get more deeper insights in the data. The use of test to decide for the impact of various column values on a certain feature just like correlation does helped me to understand the overall impact of features that might impact the business in these cases the MSRP of vehicle. The Interactive dashboard helped me dive deep into insights and analysis of simple dataset which can be made more interactive and helped me analyze the overall impact of features on other factors which ultimately helps the business to make decision for sales.

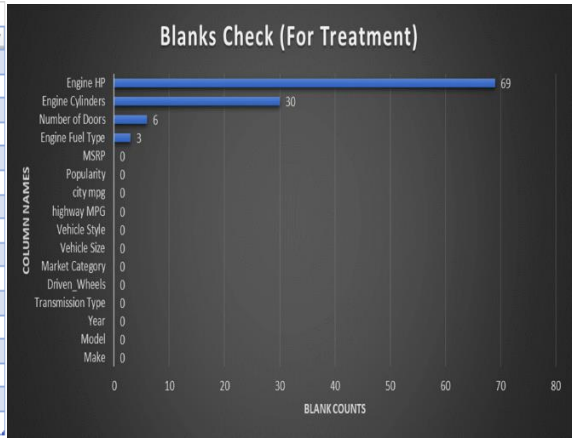
Data Cleaning Task:

Will not explain much just have added the links of gif/process. (These were the basic steps to clean and adjust the data).

- ➢ Data Count & Column Stats (Google Sheets/Docs Overview)

- The dataset had around 715 duplicate rows which we simply removed using excel function.
  - ➢ Data Import & Cleaning (Duplicate Removal)
- There were certain null values which we treated by using mode values as it was the appropriate medium and ideal as in case with no. of cylinder, HP these things are fixed so getting the average or mean does not make sense and it will give decimal value which is not case in real life.

Count	11199	For categorical data we don't get descriptive stats directly so need to use pivots.(Div/0,#N/A,#Value - since Categorcial values don't give mean mode median or division)					
Column Name	No. of Blanks	Blank %	Mean	Mode	Median	Decision	What to Fill
Make	0	0.00%	#DIV/0!	#N/A	#NUM!	No Treatment	
Model	0	0.00%	#DIV/0!	#N/A	#NUM!	No Treatment	
Year	0	0.00%	2010.714528	2016	2015	No Treatment	
Transmission Type	0	#DIV/0!	#DIV/0!	#N/A	#NUM!	#DIV/0!	Mode(See Pivot Below - UNKNOWN type)
Driven_Wheels	0	#DIV/0!	#DIV/0!	#N/A	#NUM!	#DIV/0!	
Market Category	0	#DIV/0!	#DIV/0!	#N/A	#NUM!	#DIV/0!	
Vehicle Size	0	#DIV/0!	#DIV/0!	#N/A	#NUM!	#DIV/0!	
Vehicle Style	0	#DIV/0!	#DIV/0!	#N/A	#NUM!	#DIV/0!	
highway MPG	0	#DIV/0!	26.61059023	24	25	#DIV/0!	
city mpg	0	#DIV/0!	19.73185106	17	18	#DIV/0!	
Popularity	0	#DIV/0!	1558.483347	1385	1385	#DIV/0!	
MSRP	0	#DIV/0!	41925.92714	2000	30675	#DIV/0!	
Engine Fuel Type	3	#VALUE!	#DIV/0!	#N/A	#NUM!	#VALUE!	Mode(See Pivot Below)
Number of Doors	6	#VALUE!	3.454123113	4	4	#VALUE!	Mode(4)
Engine Cylinders	30	#VALUE!	5.665950398	4	6	#VALUE!	Mode(4) (Also See Pivot for 0 zero cylinder)
Engine HP	69	#VALUE!	253.3888589	200	239	#VALUE!	Mode(200 Also note Tesla actually has higher HP but we are not treating it)



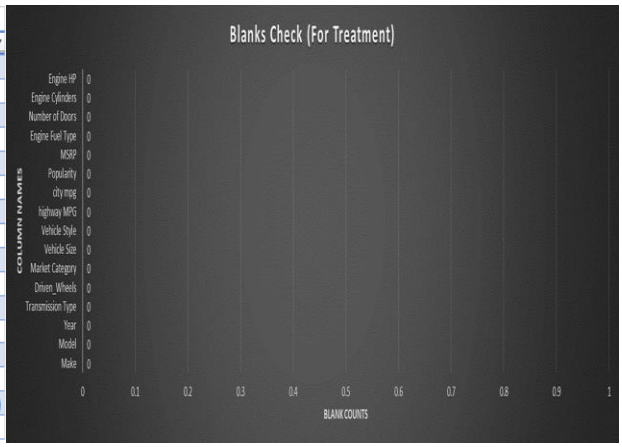
Before Treatment - Blanks Count & Stats – Calculation Gif , Pivot HD, Chart HD

- ➢ Data Imputation using Stats(calculation – mode)
- ➢ Data Imputation Manual way (filter) & Automatic (Go to special > Blanks > Replace)
- ➢ Data Imputation Numerical Columns (Pivot Stats Calculation)
- ➢ Data Imputation Categorical Columns (Mode – Pivot Table Count)

- There were certain cell values which had formatting error or data did not match with rest of columns so we used **Proper function** in formula to rectify the column values like Driven_wheels, Transmission type, Engine Fuel type etc.

- The market category column had N/A value which was the highest in the range so imputing it with mode would result in a large disparity in our analysis so I kept the value same and did analysis because imputing it with some market category would change the 2nd value to be highest which might not be the case so I kept it same & proceeded with operations.

Count	11199		For categorical data we don't get descriptive stats directly so need to use pivots.				
Column Name	No. of Blanks	Blank %	Mean	Mode	Median	Decision	What to Fil
Make	0	0.00%	#DIV/0!	#N/A	#NUM!	No Treatment	
Model	0	0.00%	#DIV/0!	#N/A	#NUM!	No Treatment	
Year	0	0.00%	2010.714528	2016	2015	No Treatment	
Transmission Type	0	0.00%	#DIV/0!	#N/A	#NUM!	No Treatment	Mode(See Pivot Below - UNKNOWN type)
Driven_Wheels	0	0.00%	#DIV/0!	#N/A	#NUM!	No Treatment	
Market Category	0	0.00%	#DIV/0!	#N/A	#NUM!	No Treatment	
Vehicle Size	0	0.00%	#DIV/0!	#N/A	#NUM!	No Treatment	
Vehicle Style	0	0.00%	#DIV/0!	#N/A	#NUM!	No Treatment	
highway MPG	0	0.00%	26.61059023	24	25	No Treatment	
city mpg	0	0.00%	19.73185106	17	18	No Treatment	
Popularity	0	0.00%	1558.483347	1385	1385	No Treatment	
MSRP	0	0.00%	41925.92714	2000	30675	No Treatment	
Engine Fuel Type	0	0.00%	#DIV/0!	#N/A	#NUM!	No Treatment	Mode(See Pivot Below)
Number of Doors	0	0.00%	3.454415573	4	4	No Treatment	Mode(4)
Engine Cylinders	0	0.00%	5.681489419	4	6	No Treatment	Mode(4) (Also See Pivot for 0 zero cylinder)
Engine HP	0	0.00%	253.0599161	200	236	No Treatment	Mode(200)



After Blank Removal

Video PPT Loom Folder (Drive Folder)

Zip (As google drive converts excel sheets to docs and then the connections & plots are lost)

Data Analytics Tasks: (Loom)

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

Q1 - Excel Answer File

- ***Task 1.A:*** Create a pivot table that shows the number of car models in each market category and their corresponding popularity scores.

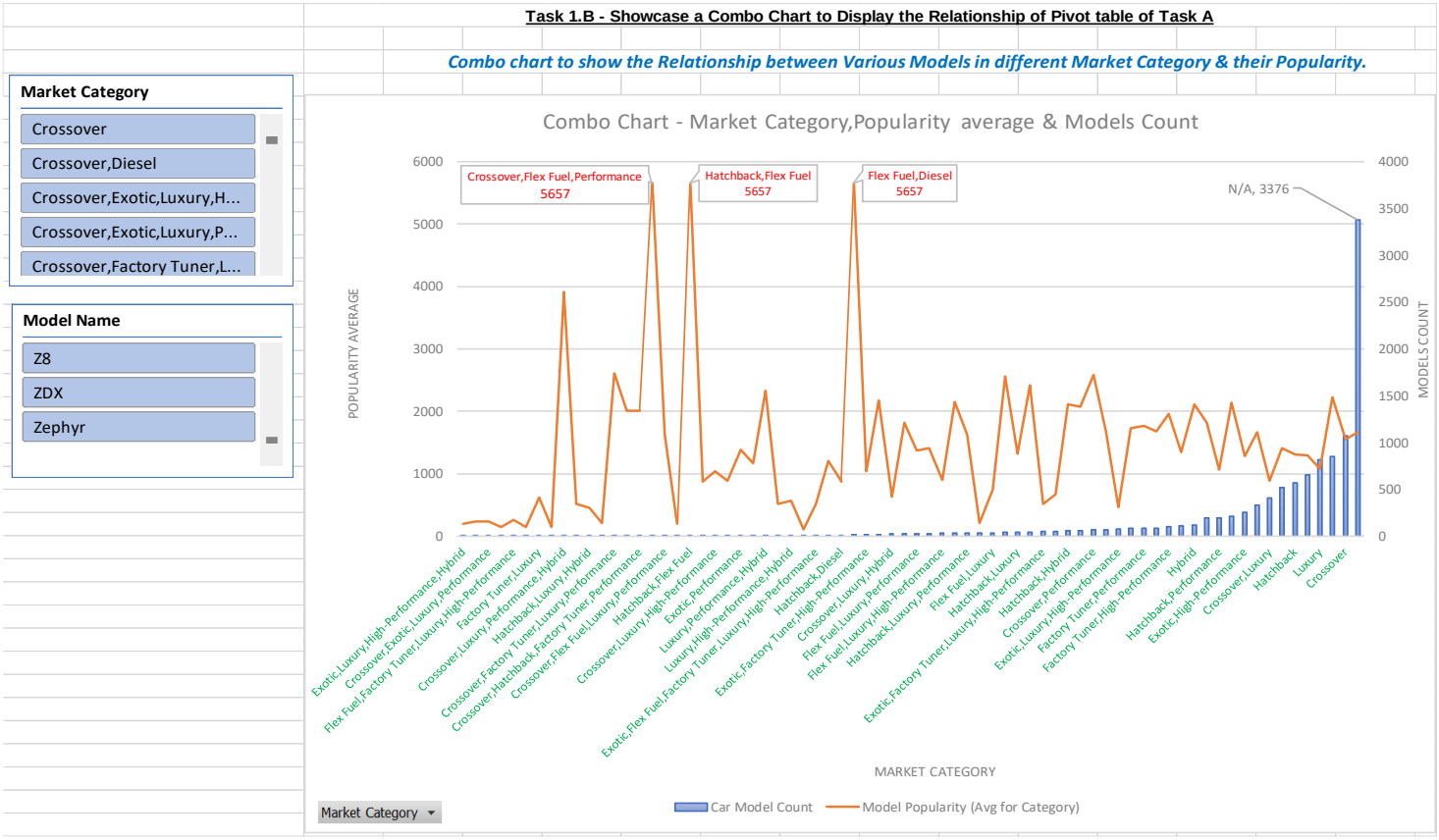
TASK 1A - Pivot Table to Show Market category distribution of car models & its popularity

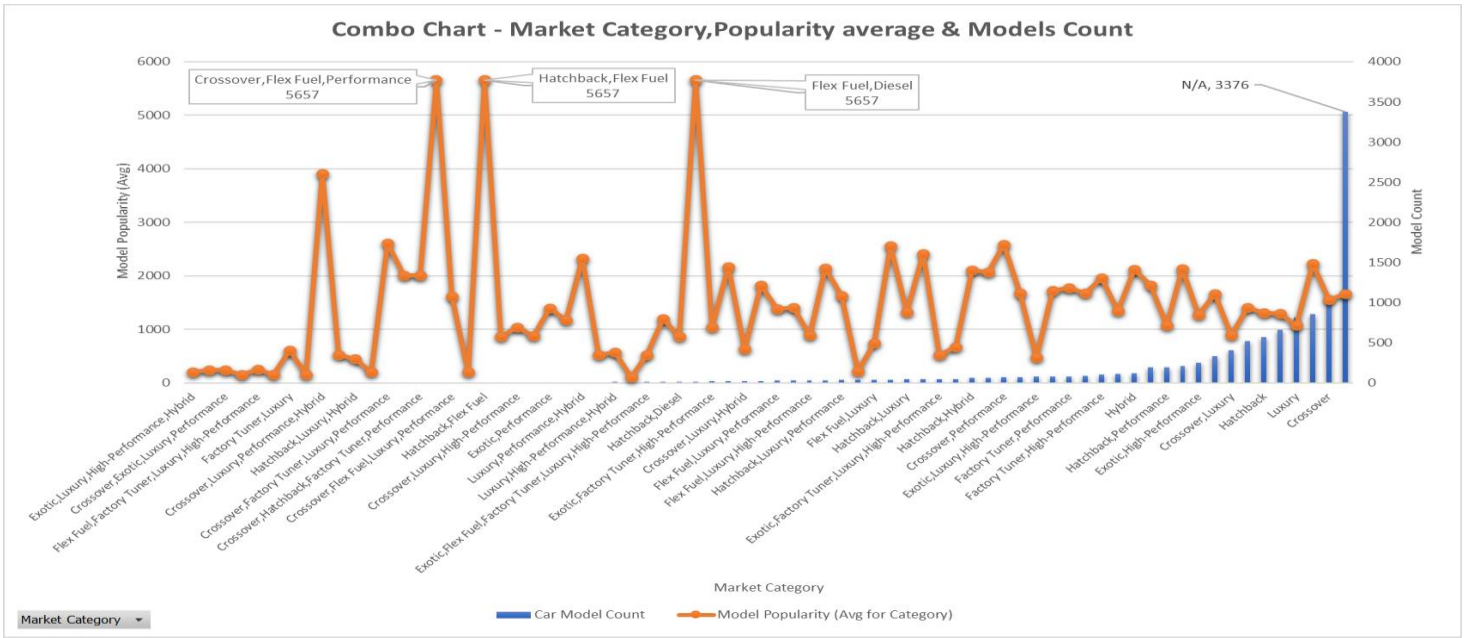
No. of Car Models(Count) in each Market Category & its Popularity Score (Average)

Market Category	Car Model Count	Model Popularity (Avg for Category)
Exotic,Luxury,High-Performance,Hybrid	1	204
Crossover,Exotic,Luxury,High-Performance	1	238
Crossover,Exotic,Luxury,Performance	1	238
Performance,Hybrid	1	155
Flex Fuel,Factory Tuner,Luxury,High-Performance	1	258
Flex Fuel,Hybrid	2	155
Factory Tuner,Luxury	2	617
Flex Fuel,Performance,Hybrid	2	155
Crossover,Luxury,Performance,Hybrid	2	3916
Exotic,Factory Tuner,Luxury,Performance	3	520
Hatchback,Luxury,Hybrid	3	454
Crossover,Factory Tuner,Performance	4	210
Crossover,Factory Tuner,Luxury,Performance	5	2607.4
Crossover,Hatchback,Performance	6	2009
Crossover,Hatchback,Factory Tuner,Performance	6	2009
Crossover,Flex Fuel,Performance	6	5657
Crossover,Flex Fuel,Luxury,Performance	6	1624
Crossover,Hatchback,Luxury	7	204
Hatchback,Flex Fuel	7	5657
Crossover,Diesel	7	873
Crossover,Luxury,High-Performance	9	1037.222222
Hatchback,Factory Tuner,Luxury,Performance	9	886.8888889
Exotic,Performance	10	1391
Crossover,Flex Fuel,Luxury	10	1173.2
Luxury,Performance,Hybrid	11	2333.181818
Exotic,Flex Fuel,Luxury,High-Performance	11	520
Luxury,High-Performance,Hybrid	12	568.8333333
Exotic,Luxury	12	112.6666667
Exotic,Flex Fuel,Factory Tuner,Luxury,High-Performance	13	520
Hatchback,Factory Tuner,High-Performance	13	1205.153846
Hatchback,Diesel	14	873
Flex Fuel,Diesel	16	5657
Exotic,Factory Tuner,High-Performance	21	1046.380952
Hatchback,Factory Tuner,Performance	21	2173.714286
Crossover,Luxury,Hybrid	24	630.9166667
Crossover,Factory Tuner,Luxury,High-Performance	26	1823.461538
Flex Fuel,Luxury,Performance	28	1380.071429
Factory Tuner,Luxury,Performance	31	1413.419355
Flex Fuel,Luxury,High-Performance	32	898.3125
Crossover,Luxury,Diesel	34	2149.411765
Hatchback,Luxury,Performance	36	1632.25
Exotic,Luxury,Performance	36	217.0277778
Flex Fuel,Luxury	39	746.5384615
Crossover,Hybrid	42	2563.380952
Hatchback,Luxury	45	1323.133333
Diesel,Luxury	47	2416.106383
Exotic,Factory Tuner,Luxury,High-Performance	51	523.0196078
Luxury,Hybrid	52	673.6346154
Hatchback,Hybrid	64	2111.15625
Crossover,Flex Fuel	64	2073.75
Crossover,Performance	69	2585.956522
Crossover,Hatchback	72	1675.694444
Exotic,Luxury,High-Performance	77	473.025974
Diesel	84	1730.904762
Factory Tuner,Performance	84	1774.047619
Flex Fuel,Performance	87	1680.471264
Factory Tuner,High-Performance	104	1966.442308
Crossover,Luxury,Performance	112	1349.089286
Hybrid	121	2116.586777
High-Performance	198	1823.378788
Hatchback,Performance	198	1073.661616
Factory Tuner,Luxury,High-Performance	215	2133.367442
Exotic,High-Performance	254	1280.047244
Luxury,High-Performance	334	1668.017964
Crossover,Luxury	406	889.2142857
Performance	520	1415.209615
Hatchback	574	1308.65331
Luxury,Performance	659	1293.062215
Luxury	819	1079.214896
Flex Fuel	855	2225.71345
Crossover	1075	1556.168372
N/A	3376	1664.832938
Grand Total	**11199**	**1558.483347**

Pivot Table 1 - Market Category Distribution of car models & its popularity (Average)

- ***Task 1.B*:** Create a combo chart that visualizes relationship b/n market category & popularity





Charts 1 – Combo-Chart to show the distribution of Popularity(avg), Market Category & Models (Count)

Insights:

1. The task 1.A helped us to get the insights about the change in popularity score according to the market category of vehicles that showed that certain vehicle category are popular among buyers may be due to price or mileage factor or maybe carrying capacity.

2. From The popularity score we can see that *Crossover, Flex-Fuel ; Hatchback, Flex-Fuel , Flex-Fuel, Diesel* seems to be the most popular with around **5657** score each, while *Performance - Hybrid; Flex-Fuel, Hybrid* seems to be the least popular ones with **155** scores each.

3. From the count of model we can conclude that *N/A & Crossover* are the market leaders in sales (counts).

4. From the Task 1.B we get the relationship between the model's popularity and count across the market category using a combo chart to visualize the relations.

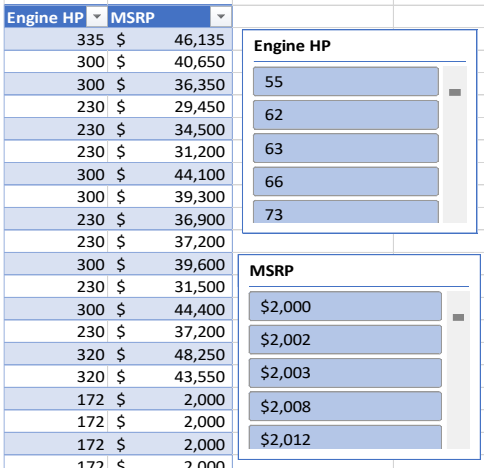
5. To get more insight I have kept the slicer and filters to drill deep on the model and category to see more information on each category impacts.

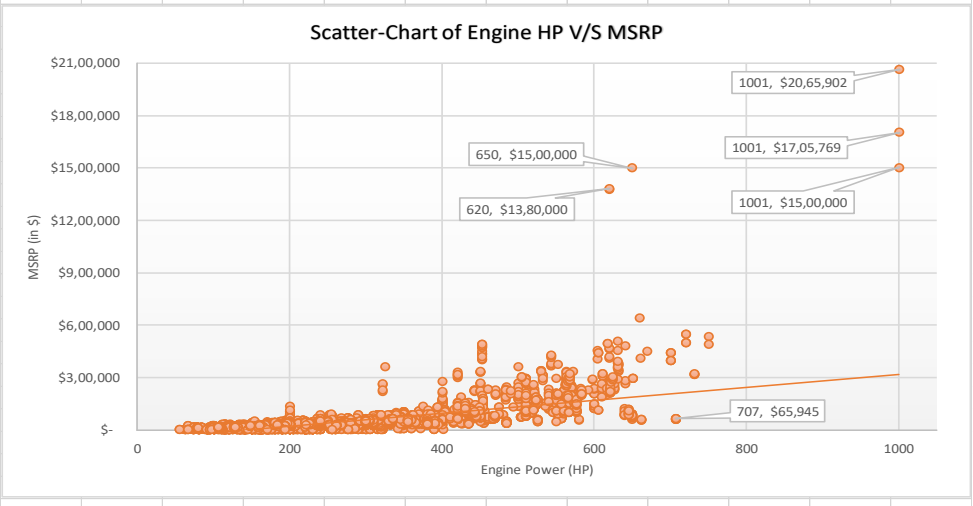
2) What is the relationship between a car's engine power and its price? Q2 - Excel_File

- ***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.

Task 2 - ScatterPlot to show the Relation between the Various Engine Power & its MSRP value for the car Model.

Engine HP	MSRP
335	$ 46,135
300	$ 40,650
300	$ 36,350
230	$ 29,450
230	$ 34,500
230	$ 31,200
300	$ 44,100
300	$ 39,300
230	$ 36,900
230	$ 37,200
300	$ 39,600
230	$ 31,500
300	$ 44,400
230	$ 37,200
320	$ 48,250
320	$ 43,550
172	$ 2,000
172	$ 2,000
172	$ 2,000
172	$ 2,000
172	$ 2,000
172	$ 2,000

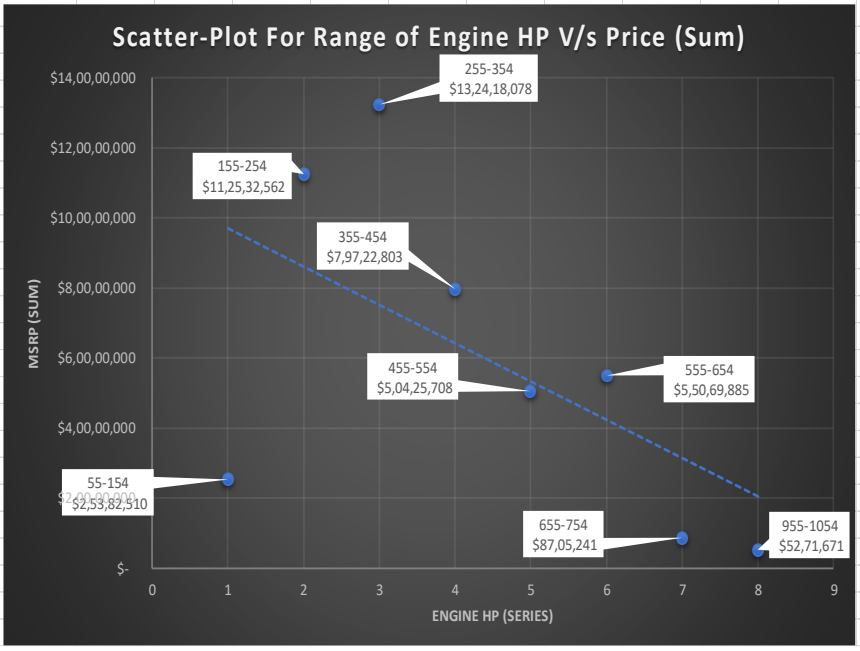




Task 2 - Using Pivot (To create a range of HP and then comparing the Price against Range of Power using a scatter.

Engine HP Range	Sum of MSRP (Currency)
55-154	$ 2,53,82,510
155-254	$ 11,25,32,562
255-354	$ 13,24,18,078
355-454	$ 7,97,22,803
455-554	$ 5,04,25,708
555-654	$ 5,50,69,885
655-754	$ 87,05,241
955-1054	$ 52,71,671
Grand Total	**$ 46,95,28,458**

Engine HP	Sum of MSRP (C
55-154	$ 2,53,82,510
155-254	$ 11,25,32,562
255-354	$ 13,24,18,078
355-454	$ 7,97,22,803
455-554	$ 5,04,25,708
555-654	$ 5,50,69,885
655-754	$ 87,05,241
955-1054	$ 52,71,671
Total	**$ 46,95,28,458**

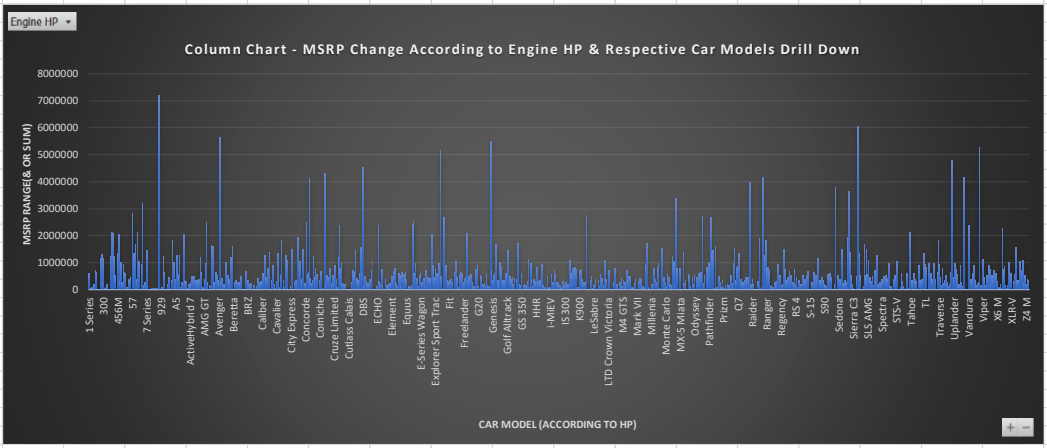


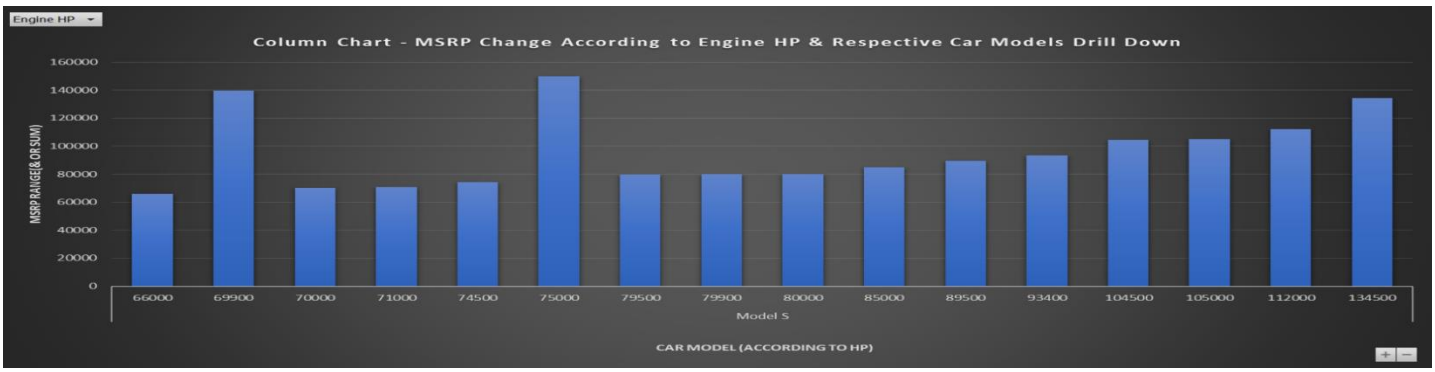
Charts 2 - Scatter plot to show the comparison of Engine HP & Change in MSRP

Task 2 (Extra Analysis) - Showcase MSRP change according to the engine HP & Car Manufacturer + Car Models [Used Pivot Table to add slicers & filters]

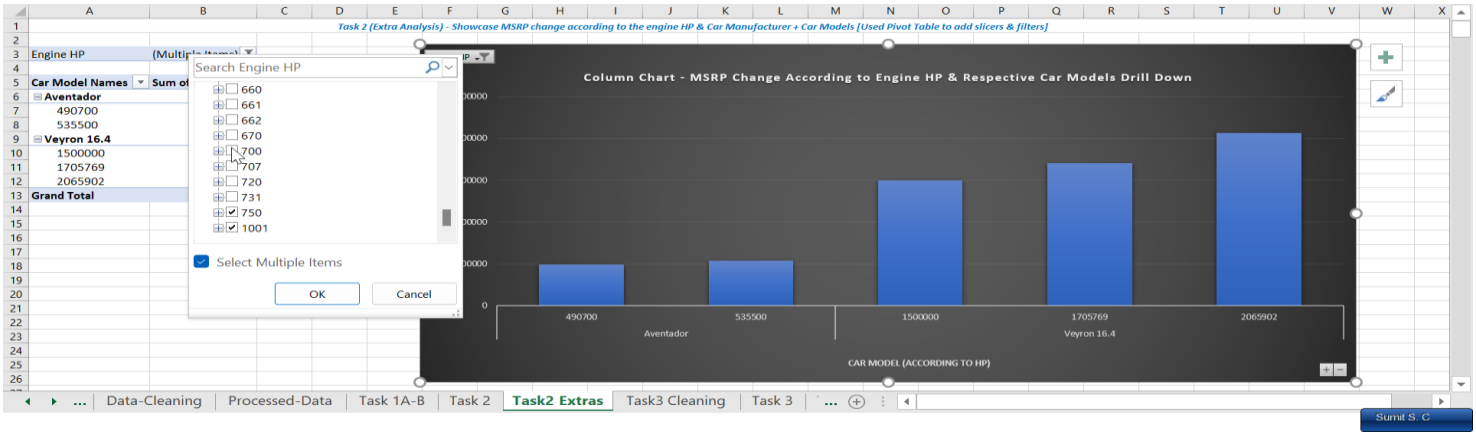
Engine HP	All

Car Model Names	Sum of MSRP
1 Series	574150
1 Series M	46135
100	22000
124 Spider	80685
190-Class	12000
2	189860
2 Series	708850
200	623390
200SX	12073
240	14000
240SX	32531
3	1176335
3 Series	1313500
3 Series Gran Turismo	274900
300	1144030
3000GT	45322
300-Class	69182
300M	183605
300ZX	35623
323	8000
350-Class	8731
350Z	1239040





Charts 3 - Drilling the data to get the model wise bars to get the value of MSRP sum



Charts 4 - Drilling same data on HP to get the models and Sum of MSRP GIF (to see the slicing)

Insights:

1. In the task 2 we analyzed the MSRP value by the engine HP using a scatterplot to check the distribution. It can be summarized that as the engine HP increases the price of car increases in linear relation.

2. In our data Bugatti had outlier with 1001HP value and also the highest price at around 1.75M avg while the rest of data was below the 600HP line and below 0.6M MSRP line.

3. The 2nd scatterplot helps to analyze the distribution of price of car on the bins of engine HP range to get a better view of the trend. The trendline shows us that as the engine HP increases the count of models drops thus the sale valuation of particular HP range drops as the price is greater than most average purchases.

4. I did extra calculation of the question and drilled the dataset according to the model of the brand to get more deeper insights of Engine HP and price of cars according to brands.

3) Which car features are most important in determining a car's price?

[Q3 - Excel Answer File]

- ***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.

Table for Vlookup calculation

Make	Column1	Model S	Column2	Year	Column3	Engine Fuel Type	Column4	Transmission Type	Column5	Driven Wheels	Column6	Market Category	Column7	Vehicle Size	Column8	Vehicle Style	Colum9
Acura	1	1 Series	1	1990	1	Diesel	1	Automated_Manual	1	All Wheel Drive	1	Crossover	1	Compact	1	2dr Hatchback	1
Alfa Rome	2	1 Series M	2	1991	2	Electric	2	Automatic	2	Four Wheel Drive	2	Crossover,Diesel	2	Large	2	2dr SUV	2
Aston Mar	3	100	3	1992	3	Flex-Fuel (Premium	3	Direct_Drive	3	Front Wheel Drive	3	Crossover,Exotic,Lux	3	Midsize	3	4dr Hatchback	3
Audi	4	124 Spider	4	1993	4	Flex-Fuel (Premium	4	Manual	4	Rear Wheel Drive	4	Crossover,Exotic,Lux	4			4dr SUV	4
Bentley	5	190-Class	5	1994	5	Flex-Fuel (Unleaded	5					Crossover,Factory T	5			Cargo Minivan	5
BMW	6	2	6	1995	6	Flex-Fuel (Unleaded	6					Crossover,Factory T	6			Cargo Van	6
Bugatti	7	2 Series	7	1996	7	Natural Gas	7					Crossover,Factory T	7			Convertible	7
Buick	8	200	8	1997	8	Premium Unleaded	8					Crossover,Flex Fuel	8			Convertible SUV	8
Cadillac	9	200SX	9	1998	9	Premium Unleaded	9					Crossover,Flex Fuel,	9			Coupe	9
Chevrolet	10	240	10	1999	10	Regular Unleaded	10					Crossover,Flex Fuel,	10			Crew Cab Picku	10
Chrysler	11	240SX	11	2000	11							Crossover,Flex Fuel,	11			Extended Cab P	11
Dodge	12	3	12	2001	12							Crossover,Hatchbac	12			Passenger Minir	12
Ferrari	13	3 Series	13	2002	13							Crossover,Hatchbac	13			Passenger Van	13
FIAT	14	3 Series Gr	14	2003	14							Crossover,Hatchbac	14			Regular Cab Pic	14
Ford	15	300	15	2004	15							Crossover,Hatchbac	15			Sedan	15
Genesis	16	3000GT	16	2005	16							Crossover,Hybrid	16			Wagon	16
GMC	17	300-Class	17	2006	17							Crossover,Luxury	17				
Honda	18	300M	18	2007	18							Crossover,Luxury,Di	18				
HUMMER	19	300ZX	19	2008	19							Crossover,Luxury,Hi	19				
Hyundai	20	323	20	2009	20							Crossover,Luxury,Hy	20				
Infiniti	21	350-Class	21	2010	21							Crossover,Luxury,Pe	21				
Kia	22	350Z	22	2011	22							Crossover,Luxury,Pe	22				
Lamborgh	23	360	23	2012	23							Crossover,Performa	23				
Land Rove	24	370Z	24	2013	24							Diesel	24				
Lexus	25	4 Series	25	2014	25							Diesel,Luxury	25				
Lincoln	26	4 Series Gr	26	2015	26							Exotic,Factory Tunes	26				
Lotus	27	400-Class	27	2016	27							Exotic,Factory Tunes	27				
Maserati	28	420-Class	28	2017	28							Exotic,Factory Tunes	28				
Maybach	29	456M	29									Exotic,Flex Fuel,Fact	29				
Mazda	30	458 Italia	30									Exotic,Flex Fuel,Luxu	30				
McLaren	31	4C	31									Exotic,High-Perform	31				
Mercedes	32	4Runner	32									Exotic,Luxury	32				
Mitsubish	33	5	33									Exotic,Luxury,High-F	33				
Nissan	34	5 Series	34									Exotic,Luxury,High-F	34				
Oldsmobil	35	5 Series Gr	35									Exotic,Luxury,Perfor	35				
Plymouth	36	500	36									Exotic,Performance	36				
Pontiac	37	500-Class	37									Factory Tuner,High-	37				
Porsche	38	500e	38									Factory Tuner,Luxur	38				
Rolls-Royc	39	500L	39									Factory Tuner,Luxur	39				
Saab	40	500X	40									Factory Tuner,Luxur	40				
Scion	41	550	41									Factory Tuner,Perfo	41				
Spyker	42	560-Class	42									Flex Fuel	42				
Subaru	43	57	43									Flex Fuel,Diesel	43				
Suzuki	44	570S	44									Flex Fuel,Factory Tu	44				
Tesla	45	575M	45									Flex Fuel,Hybrid	45				
Toyota	46	599	46									Flex Fuel,Luxury	46				
Volkswage	47	6	47									Flex Fuel,Luxury,Hig	47				
Volvo	48	6 Series	48									Flex Fuel,Luxury,Per	48				
		6 Series Gr	49									Flex Fuel,Performan	49				
		6000	50									Flex Fuel,Performan	50				
		600-Class	51									Hatchback	51				
		612 Scaglie	52									Hatchback,Diesel	52				
		62	53									Hatchback,Factory T	53				
		626	54									Hatchback,Factory T	54				

Data Cleaning 1 For Regression Test (Vlookup calculation) [Vlookup calculation GIF]

Categorical Columns Tranformed to Numerical using Vlookup/IF function to use them for regression analysis

Sr. No.	Make	Make2	Model	Model2	Year	Year2	Engine Fuel Type	Engine Fuel Type2	Engine HP	Engine Cylinders	Transmission Type	Transmission Type2	Driven_Wheels	Driven_Wheels2	Number of Doors	Market Category	Market Category2	Vehicle Size	Vehicle Size2	Vehicle Style	Vehicle Style2	Highway MPG	City MPG	Popularity	MSRP
1	BMW	6	1 Series M	2	2011	22	Premium Unleaded	9	335	6	Manual	4	Rear Wheel Drive	4	2	Factory Tuner,Luxur	39	Compact	1	Coupe	9	26	19	3916	46135
2	BMW	6	1 Series	1	2011	22	Premium Unleaded	9	300	6	Manual	4	Rear Wheel Drive	4	2	Luxury,Performance	68	Compact	1	Convertible	7	28	19	3916	40650
3	BMW	6	1 Series	1	2011	22	Premium Unleaded	9	300	6	Manual	4	Rear Wheel Drive	4	2	Luxury,High-Perform	65	Compact	1	Coupe	9	28	20	3916	36350
4	BMW	6	1 Series	1	2011	22	Premium Unleaded	9	230	6	Manual	4	Rear Wheel Drive	4	2	Luxury,Performance	68	Compact	1	Coupe	9	28	18	3916	29450
5	BMW	6	1 Series	1	2011	22	Premium Unleaded	9	230	6	Manual	4	Rear Wheel Drive	4	2	Luxury	64	Compact	1	Convertible	7	28	18	3916	34500
6	BMW	6	1 Series	1	2012	23	Premium Unleaded	9	230	6	Manual	4	Rear Wheel Drive	4	2	Luxury,Performance	68	Compact	1	Coupe	9	28	18	3916	31200
7	BMW	6	1 Series	1	2012	23	Premium Unleaded	9	300	6	Manual	4	Rear Wheel Drive	4	2	Luxury,Performance	68	Compact	1	Convertible	7	26	17	3916	44100
8	BMW	6	1 Series	1	2012	23	Premium Unleaded	9	300	6	Manual	4	Rear Wheel Drive	4	2	Luxury,High-Perform	65	Compact	1	Coupe	9	28	20	3916	39300
9	BMW	6	1 Series	1	2012	23	Premium Unleaded	9	230	6	Manual	4	Rear Wheel Drive	4	2	Luxury	64	Compact	1	Convertible	7	28	18	3916	36900
10	BMW	6	1 Series	1	2013	24	Premium Unleaded	9	230	6	Manual	4	Rear Wheel Drive	4	2	Luxury	64	Compact	1	Convertible	7	27	18	3916	37200
11	BMW	6	1 Series	1	2013	24	Premium Unleaded	9	300	6	Manual	4	Rear Wheel Drive	4	2	Luxury,High-Perform	65	Compact	1	Coupe	9	28	20	3916	39600

Data Cleaning 2 - After Vlookup & Numerical value insertion table data

Cleaned data with numerical value for categorical values - for regression analysis

Make2	Model2	Year2	Engine Fuel Type2	Engine HP	Engine Cylinders	Transmission Type2	Driven_Wheels2	Number of Doors	Market Category2	Vehicle Size2	Vehicle Style2	Highway MPG	City MPG	Popularity	MSRP
6	2	22	9	335	6	4	4	2	39	1	9	26	19	3916	46135
6	1	22	9	300	6	4	4	2	68	1	7	28	19	3916	40650
6	1	22	9	300	6	4	4	2	65	1	9	28	20	3916	36350
6	1	22	9	230	6	4	4	2	68	1	9	28	18	3916	29450
6	1	22	9	230	6	4	4	2	64	1	7	28	18	3916	34500
6	1	23	9	230	6	4	4	2	68	1	9	28	18	3916	31200
6	1	23	9	300	6	4	4	2	68	1	7	26	17	3916	44100
6	1	23	9	300	6	4	4	2	65	1	9	28	20	3916	39300
6	1	23	9	230	6	4	4	2	64	1	7	28	18	3916	36900
6	1	24	9	230	6	4	4	2	64	1	7	27	18	3916	37200
6	1	24	9	300	6	4	4	2	65	1	9	28	20	3916	39600
6	1	24	9	230	6	4	4	2	68	1	9	28	19	3916	31500

Data Cleaning 3 - New table with numerical columns for regression calculations

Task 3 - Impact of Various Features on MSRP Values (Using Regression Analysis, Intercept-Coefficients & BarPlots to show the Effects

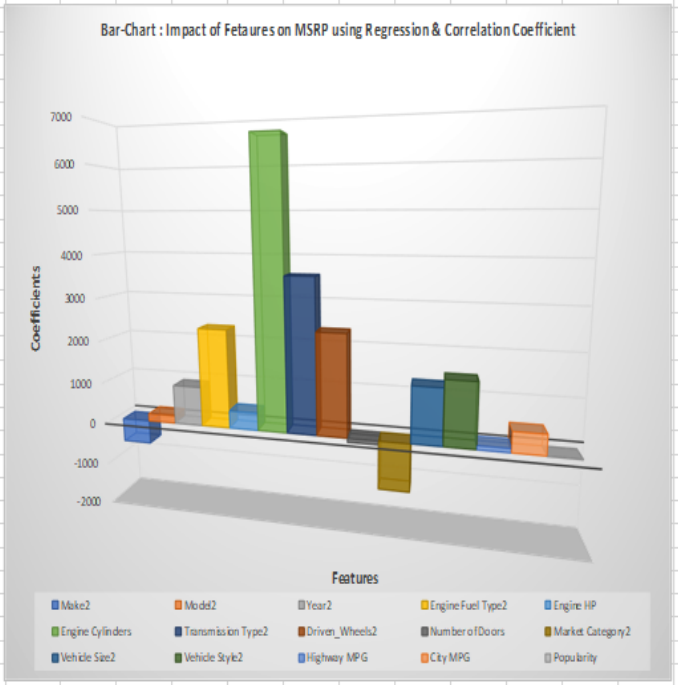
SUMMARY OUTPUT

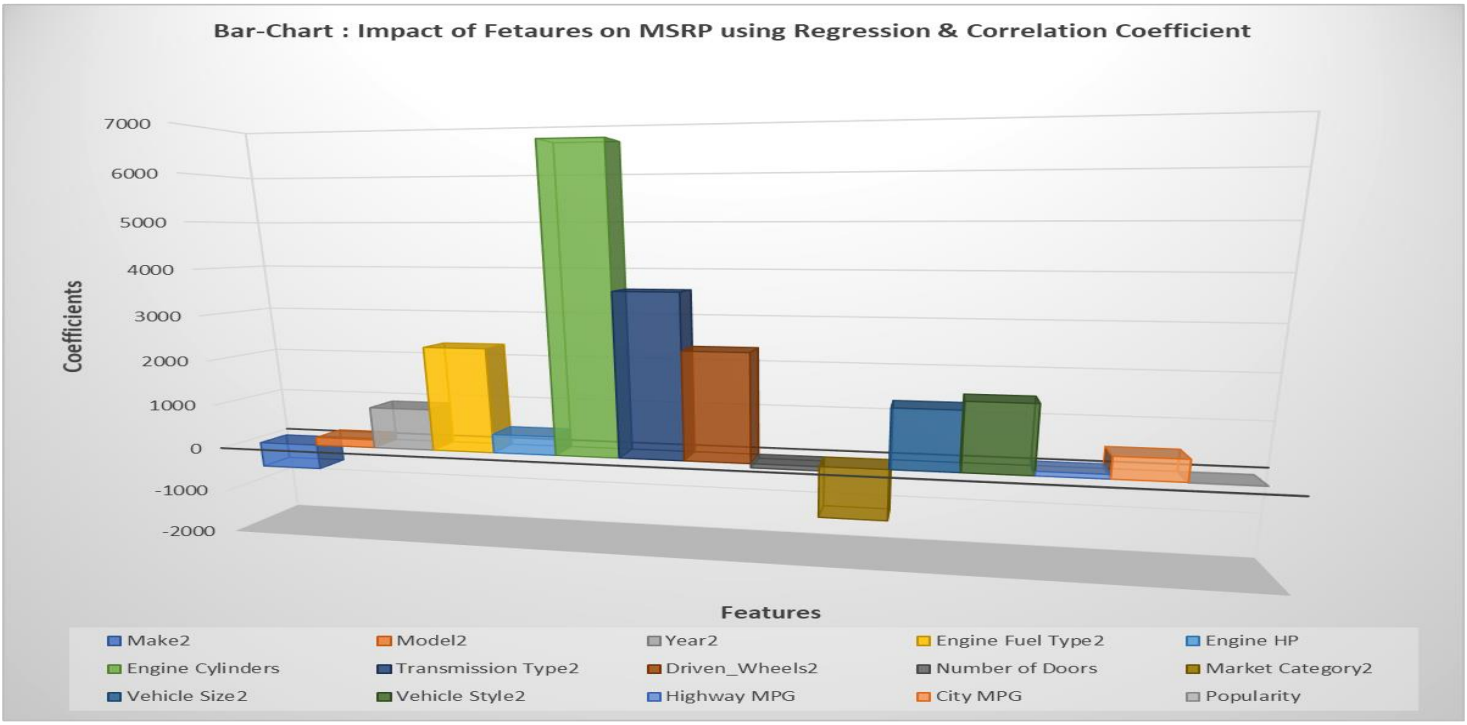
Regression Statistics	
Multiple R	0.860201093
R Square	0.739945921
Adjusted R Square	0.735602033
Standard Error	36582.47693
Observations	914

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	15	3.41947E+12	2.27965E+11	170.3418613	3.2101E-250
Residual	898	1.20177E+12	1338277619		
Total	913	4.62124E+12			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-90274.43033	18714.54663	-4.823757269	1.65449E-06	-127003.772	-53545.08865	-127003.772	-53545.08865
Make2	-564.889301	147.7635595	-3.822926997	0.000140976	-854.8914247	-274.8871773	-854.8914247	-274.8871773
Model2	179.928662	65.18926338	2.760096566	0.005896116	51.98761292	307.8697111	51.98761292	307.8697111
Year2	927.8523953	249.0318299	3.725838564	0.000206828	439.100231	1416.60456	439.100231	1416.60456
Engine Fuel Type2	2325.365834	921.4384161	2.52362588	0.011786668	516.9423066	4133.789361	516.9423066	4133.789361
Engine HP	416.2166312	28.41007022	14.6503204	9.76707E-44	360.4587656	471.9744968	360.4587656	471.9744968
Engine Cylinders	6616.684742	1505.023983	4.396398209	1.23227E-05	3662.910804	9570.45868	3662.910804	9570.45868
Transmission Type2	3582.906435	1447.859826	2.474622453	0.013521435	741.3233968	6424.489473	741.3233968	6424.489473
Driven_Wheels2	2369.160011	1229.89528	1.926310354	0.054380989	-44.64379877	4782.963821	-44.64379877	4782.963821
Number of Doors	-91.32479006	1949.499625	-0.046845246	0.962646983	-3917.430719	3734.781139	-3917.430719	3734.781139
Market Category2	-1120.143878	111.5351572	-10.04296678	1.4526E-22	-1339.043805	-901.2439507	-1339.043805	-901.2439507
Vehicle Size2	1304.506579	1781.098704	0.732416781	0.464105335	-2191.094151	4800.10731	-2191.094151	4800.10731
Vehicle Style2	1485.818507	379.1747246	3.918558941	9.5845E-05	741.6466985	2229.990316	741.6466985	2229.990316
Highway MPG	96.34531861	708.707399	0.135945129	0.891895127	-1294.570354	1487.260991	-1294.570354	1487.260991
City MPG	470.5983864	588.2495298	0.799997896	0.423923589	-683.9055618	1625.102335	-683.9055618	1625.102335
Popularity	-10.49628806	1.497949295	-7.007105043	4.77382E-12	-13.43617715	-7.556398968	-13.43617715	-7.556398968





Charts 5 Impacts of features on MSRP

Insights:

1. The data cleaning part need to be done in question as for the regression test we need numerical value by doing basic *VLOOKUP* and treatment of data we get the new refined dataset on which we do data analysis function to perform regression analysis which gave us the coefficient on which we plotted the required graph.

2. From the new data and regression analysis we get the relationship between various features that indirectly impacts the price of cars.

3.The basic bar-chart shows us that few values lie below the zero indicating that they might not be impacting much on the MSRP.

4. For more details watch the video ppt for explanation & insights or open the excel sheet to see them.

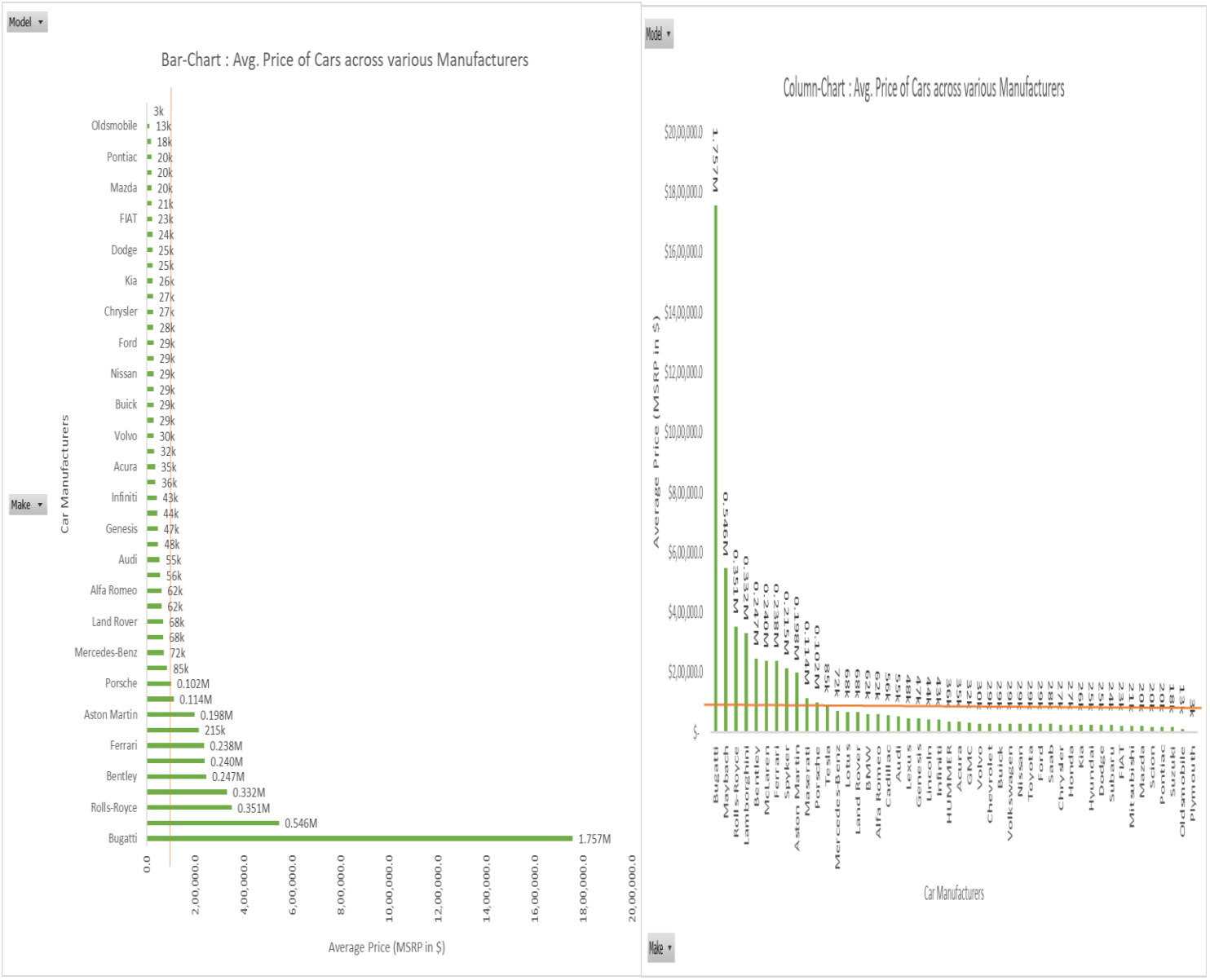
4) How does the average price of a car vary across different manufacturers?

Q4 - Excel_Answer_File

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

Model	(All)	
Row Labels	**Average of MSRP**	
Bugatti	$	17,57,223.7
Maybach	$	5,46,221.9
Rolls-Royce	$	3,51,130.6
Lamborghini	$	3,31,567.3
Bentley	$	2,47,169.3
McLaren	$	2,39,805.0
Ferrari	$	2,38,218.8
Spyker	$	2,14,990.0
Aston Martin	$	1,98,123.5
Maserati	$	1,13,684.5
Porsche	$	1,01,622.4
Tesla	$	85,255.6
Mercedes-Benz	$	72,069.5
Lotus	$	68,377.1
Land Rover	$	68,067.1
BMW	$	62,162.6
Alfa Romeo	$	61,600.0
Cadillac	$	56,368.3
Audi	$	54,574.1
Lexus	$	47,549.1
Genesis	$	46,616.7
Lincoln	$	43,860.8
Infiniti	$	42,640.3
HUMMER	$	36,464.4
Acura	$	35,087.5
GMC	$	32,444.1
Volvo	$	29,724.7
Chevrolet	$	29,074.7
Buick	$	29,034.2
Volkswagen	$	28,978.5
Nissan	$	28,921.2
Toyota	$	28,846.6
Ford	$	28,511.3
Saab	$	27,879.8
Chrysler	$	26,723.0
Honda	$	26,655.1
Kia	$	25,513.8
Hyundai	$	24,926.3
Dodge	$	24,857.0
Subaru	$	24,240.7
FIAT	$	22,670.2
Mitsubishi	$	21,340.6
Mazda	$	20,416.6
Scion	$	19,932.5
Pontiac	$	19,800.0
Suzuki	$	18,026.4
Oldsmobile	$	12,843.8
Plymouth	$	3,296.9
Grand Total	**$**	**41,925.9**

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



Charts 6 Bar chart - avg price of car models across various manufacturers

Insights:

1. The pivot chart shows us the Task 4.A part to showcase the average MSRP for each manufacturer.

2. The bar and column charts in Task 4.B gives us a clear picture about the MSRP for various brands and we can even get the top brands and cheap brands directly from the categorization.

3. The chart shows us that Bugatti brand is an ultra-luxury brand whose MSRP is almost double of the 2nd top brand, it clearly establishes the fact that owning a Bugatti means a person has many cars and maybe a Yatch and they need to be a millionaire to buy which is seen in the exorbitant price of $1.75M while the Plymouth is your daily average household secondhand cars range under $3k dollars.

4. We can get many insights from the price distribution, I have pulled a line above $100k to show that below these lines many people buy cars or can be considered as the mainstream ones brands while above it I changed the denotation to $M calculation showing these are luxury and premium brand cars which sales limited quantity and has userbase from elite class.

---

5) What is the relationship between fuel efficiency and the number of cylinders in a car's engine? Q5 - Excel_Answer_File

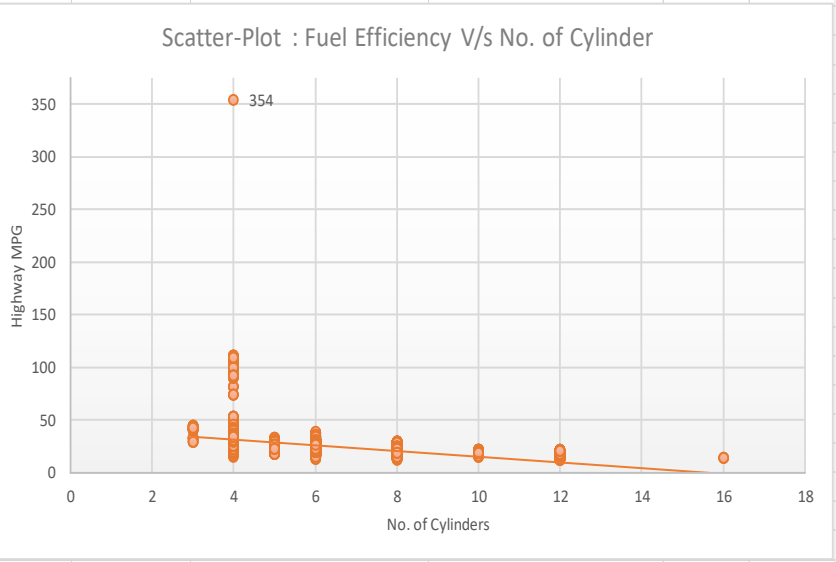
- **Task 5.A:** Create a scatter plot with the number of cylinders on the x-axis and highway MPG on the y-axis. Then create a trendline on the scatter plot to visually estimate the slope of the relationship and assess its significance.

Scatter chart to showcase the effect of cylinders & Fuel Efficiency (on Highways) [Decreasing trending]

Engine Cylinders	Highway MPG
6	26
6	28
6	28
6	28
6	28
6	28
6	26
6	28
6	28
6	27
6	28
6	28
6	28
6	28
6	25
6	28
6	24
6	20
6	21
6	24
6	20
6	21

Correleation between Cylinder & Efficiency = **-0.524935279**

The negative value indicated that as the no. of cylinders increases the efficiency decreases & overall the trend is going down.

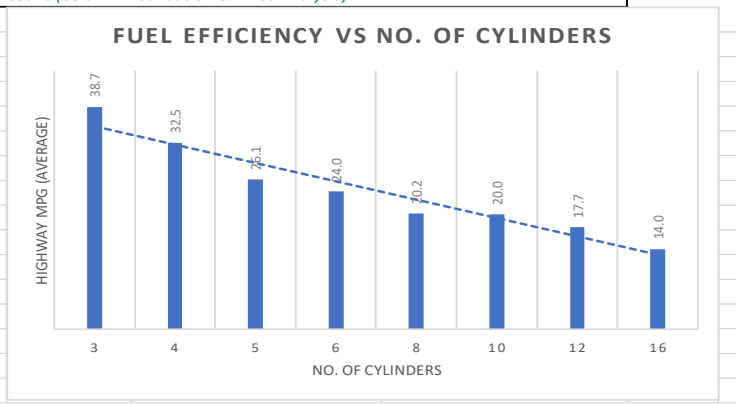


Charts 7 - Scatterplot for fuel efficiency vs No. of cylinder

Task 5 - Relationship Between Fuel Efficeny & No. of Cylinders

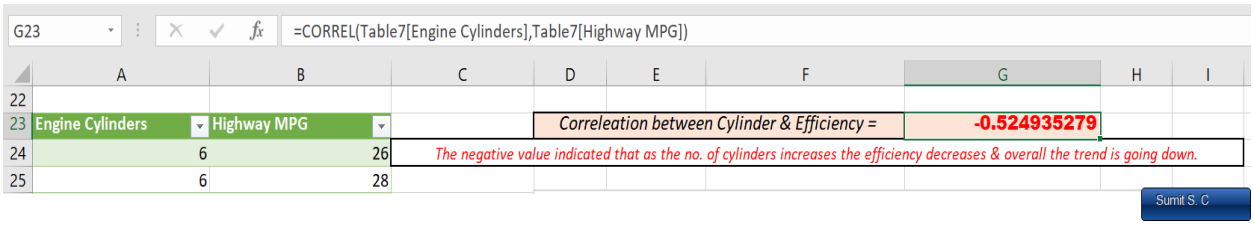
Fuel Efficiency over Cylinders count (Column Distribution & Pivot Analysis)

Engine Cylinders (No.)	Average of Highway MPG	Eng. Cylinder Count
3	38.7	30
4	32.5	4448
5	26.1	169
6	24.0	4291
8	20.2	1965
10	20.0	65
12	17.7	228
16	14.0	3
Grand Total	**26.6**	**11199**



Charts 8 (Extra) - Bar chart for same analysis but on Bin data instead & average MPG to see the relationship & trendline

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



Pivot Table 2 - Using CORREL func to get the coefficient for no. of cylinder & highway MPG

Insights:

1. The task 5.A helped us analyze the relation between the no. of cylinder & fuel efficiency using the scatterplot.

2. From the graph it is evident that as the number of cylinder increase it impacts the mileage of the cars which is true more cylinders gives more power & speed on the compensation of fuel burning while lower cylinders means less power torque but better mileage.

3. The task 5.B we did the **CORREL** of the two column values to get the coefficient of correlation which indicates us that there is negative trend meaning that our scatterplot is true the efficiency decreases with increase in cylinders count.

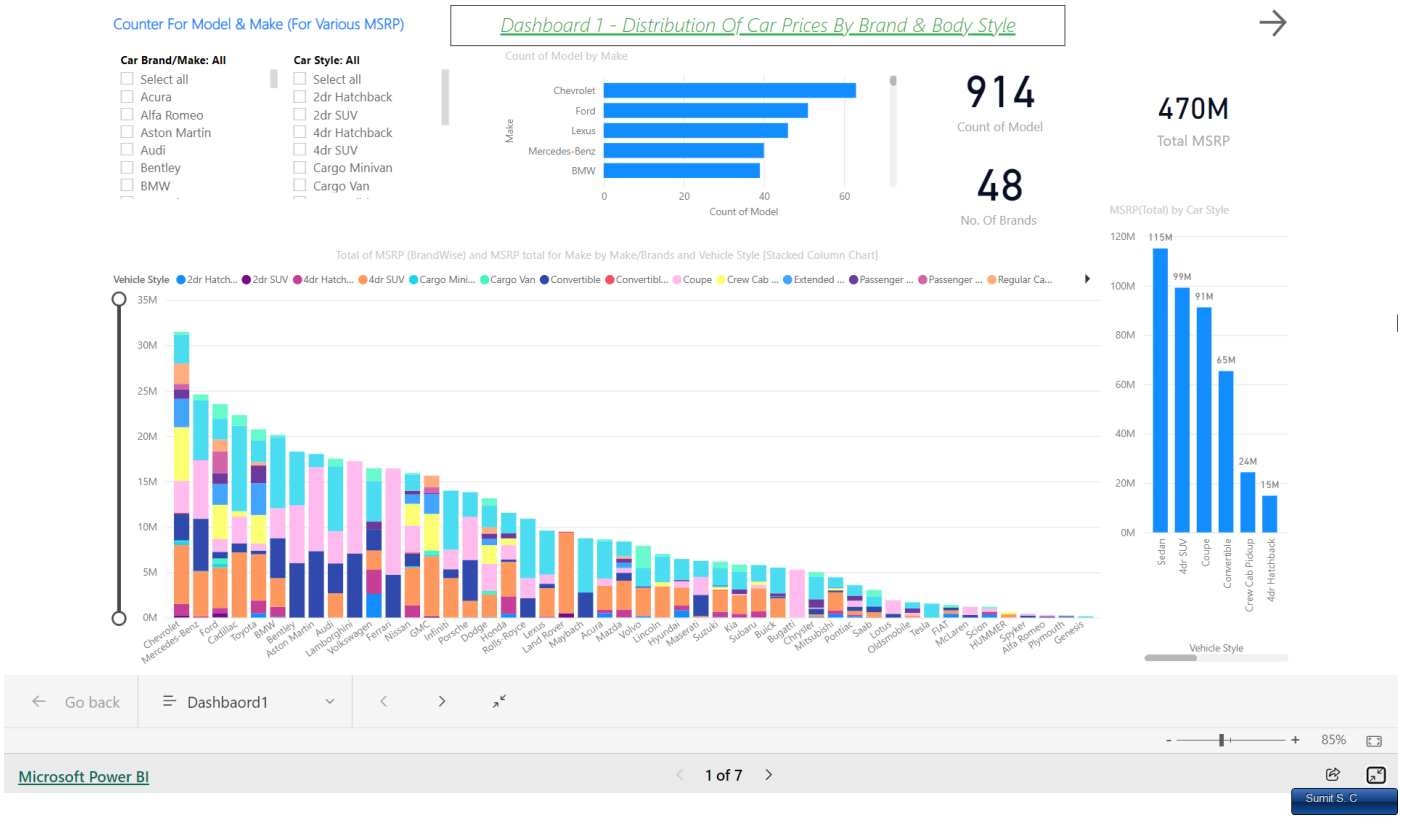
4. I did a simple bar chart on the binned data of cylinder and its count over the average of highway MPG and we can see the trend more vividly. (Extra Analysis)

DASHBOARD (PowerBi Web link, Loom)

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

Dashboard 1

- ***Hints*:** Stacked column chart to show the distribution of car prices by brand and body style. Use filters and slicers to make the chart interactive. Calculate the total MSRP for each brand and body style using SUMIF or Pivot Tables.



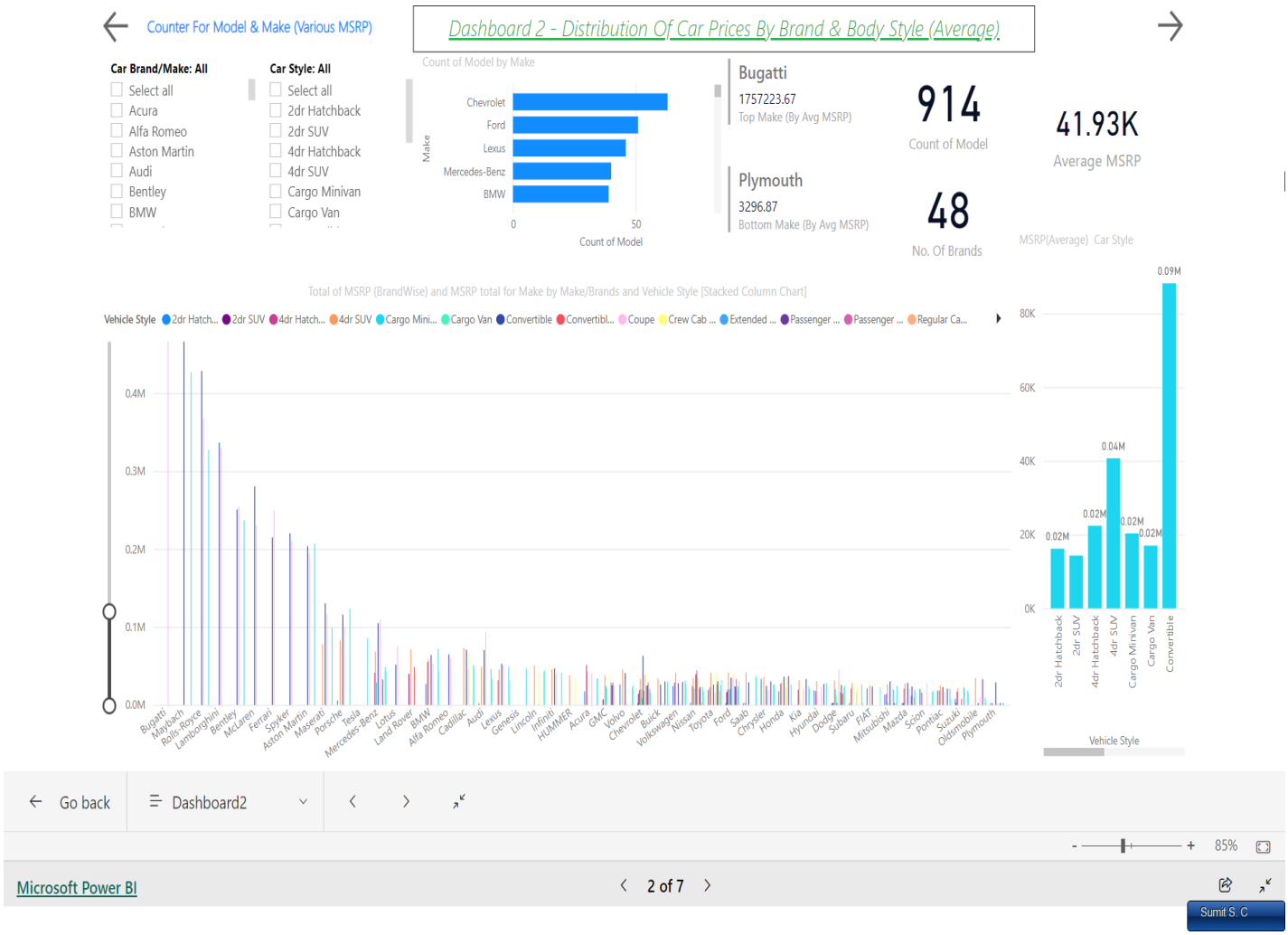
Dashboard 1 - Distribution of car prices by brand & body style(Sum of MSRP - Total). (Gif of dashboard)

Insights:

1. The dashboard can be interacted to get the distribution of prices across brand & vehicle style.
2. We can see the impact of brands on MSRP showing loyalty as well as the preference for style when buying as a trend.
3. The stacked column takes into the various brands under style so we can see the dist. Neatly.

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

- ***Hints*:** Clustered column chart to compare the average MSRPs across different car brands and body styles. Calculate the average MSRP for each brand and body style using AVERAGEIF or Pivot Tables.



Dashboard 2 - Distribution of car price by brand & style (average MSRP) (Gif Dashboard 2)

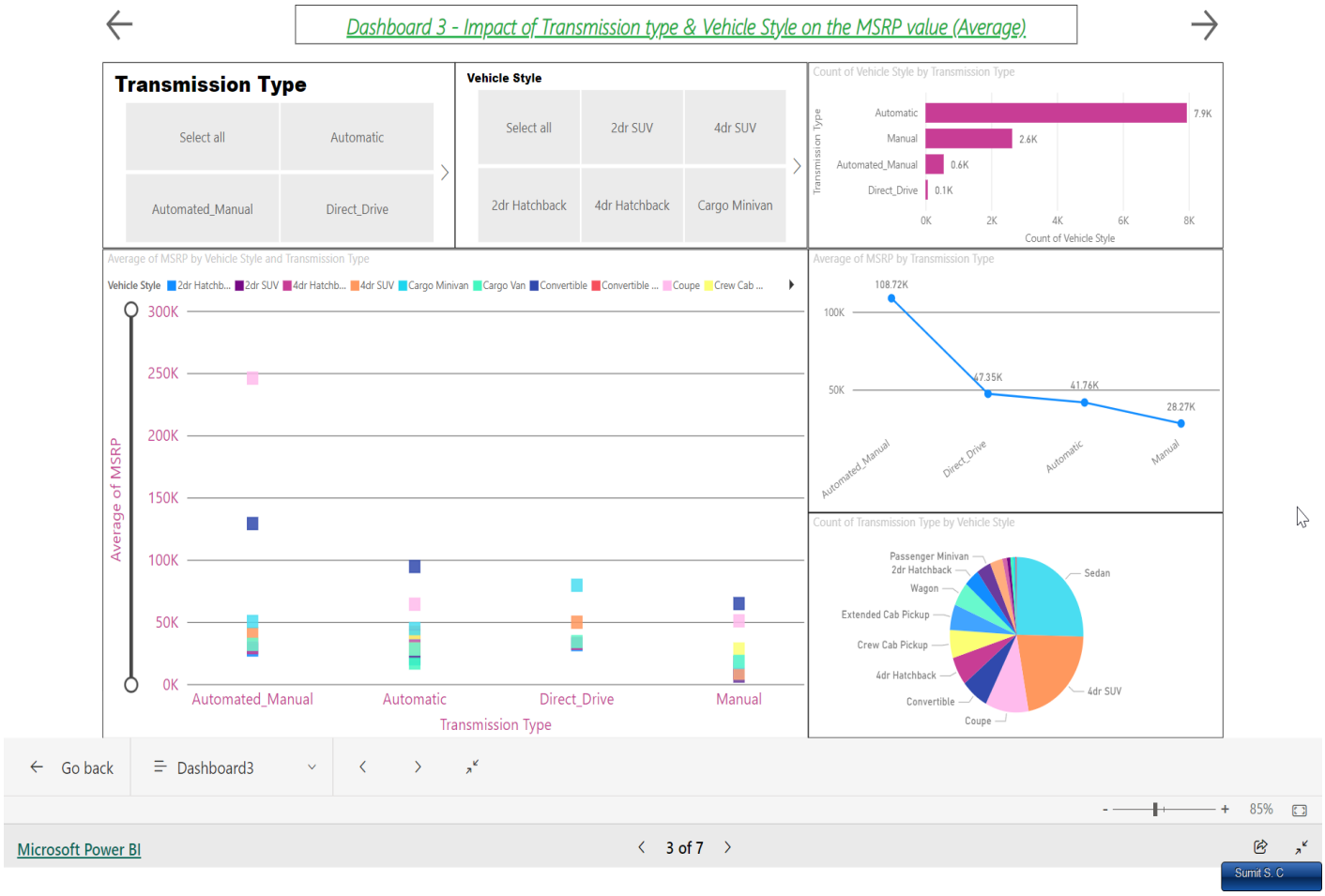
Insights:

1. The question helped us analyze the same analysis like previous but these time the difference is that we did the average MSRP calculation and used a bar chart which drills the data more and we get in depth knowledge.

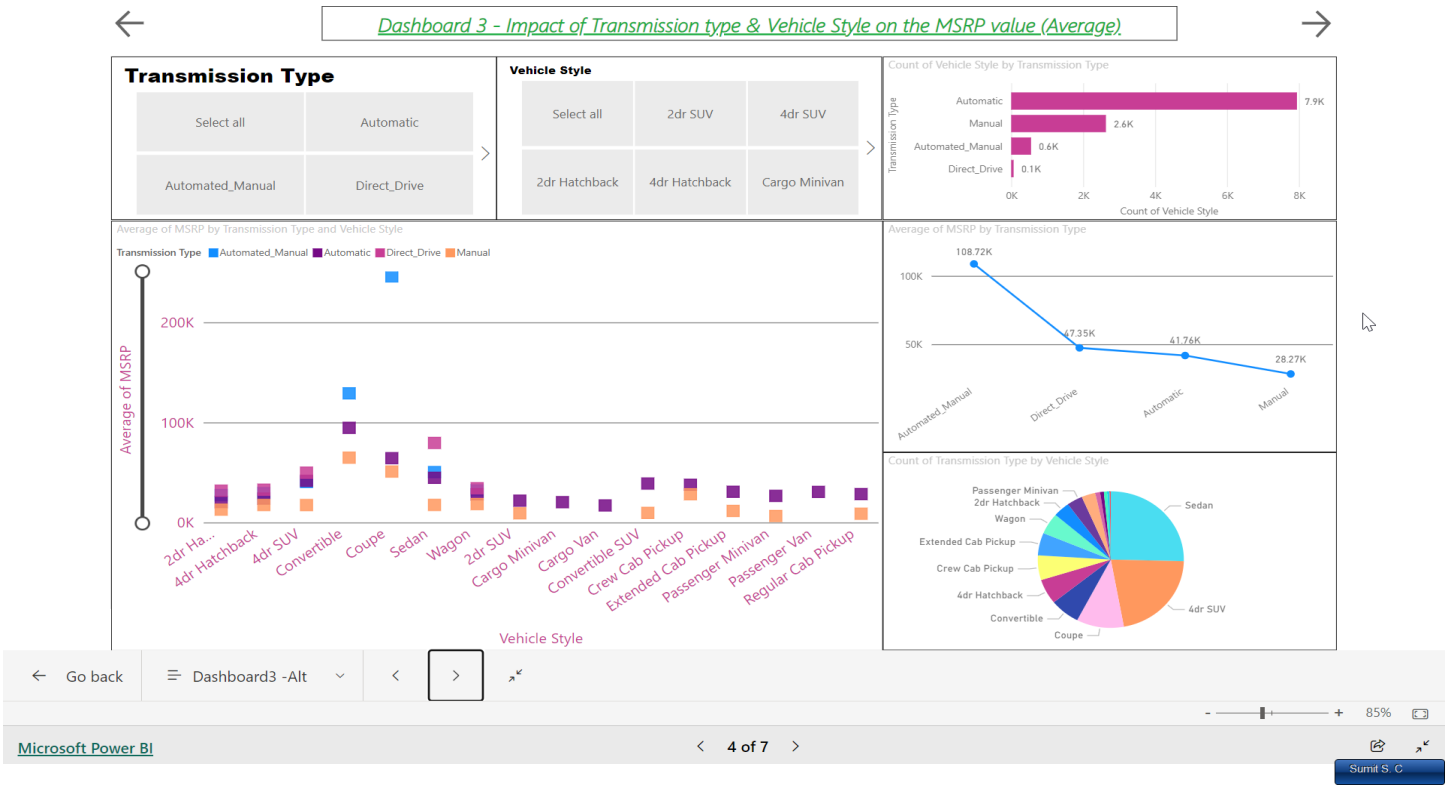
2. Like previous insights we can toggle the filters & slicers to see the trends.

3) How do the different feature such as transmission type affect the MSRP, and how does this vary by body style? Dashboard 3 Dashboard 3(Alt)

- **Hints:** Scatter plot chart to visualize the relationship between MSRP and transmission type, with different symbols for each body style. Calculate the average MSRP for each combination of transmission type and body style using AVERAGEIFS or Pivot Tables.



Dashboard 3 - Impact of transmission type & Vehicle style on MSRP(Value) (Gif of Dashboard 3)



Insights:

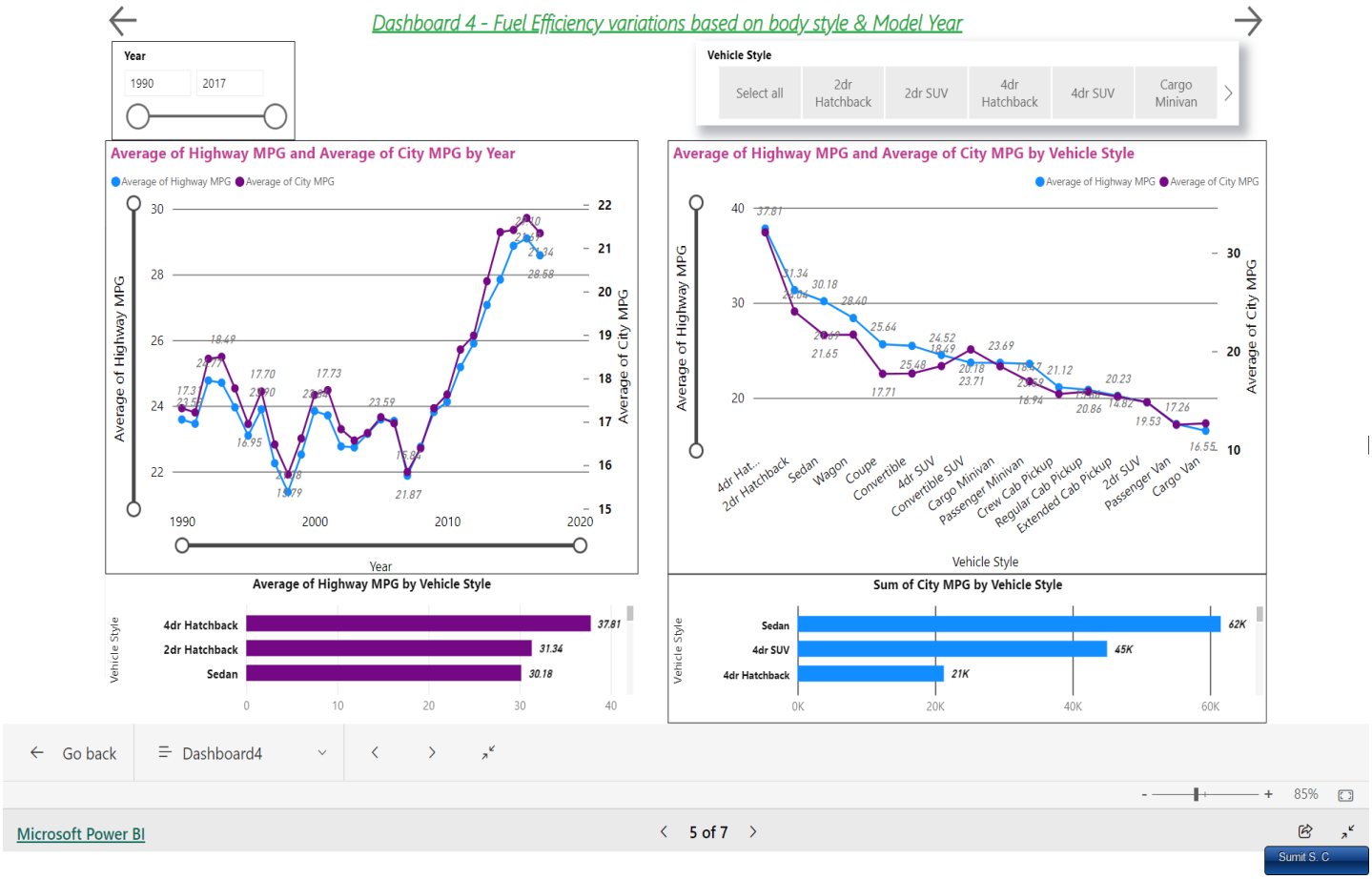
1. The dashboard focus on seeing the distribution of MSRP based on vehicle type & transmission types.

2. It is interesting to see the distribution of cars based once using the transmission over scatterplot and once using the vehicle type the scattering changes shows us the variation in detail.

3. The other charts and counters bars helps us to get more refined insights.

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

- **Hints:** Line chart to show the trend of fuel efficiency (MPG) over time for each body style. Calculate the average MPG for each combination of body style and model year using AVERAGEIFS or Pivot Tables.



Dashboard 4 - Fuel Efficiency Variation based on body style & model year (Gif of Dashboard)

Insights:

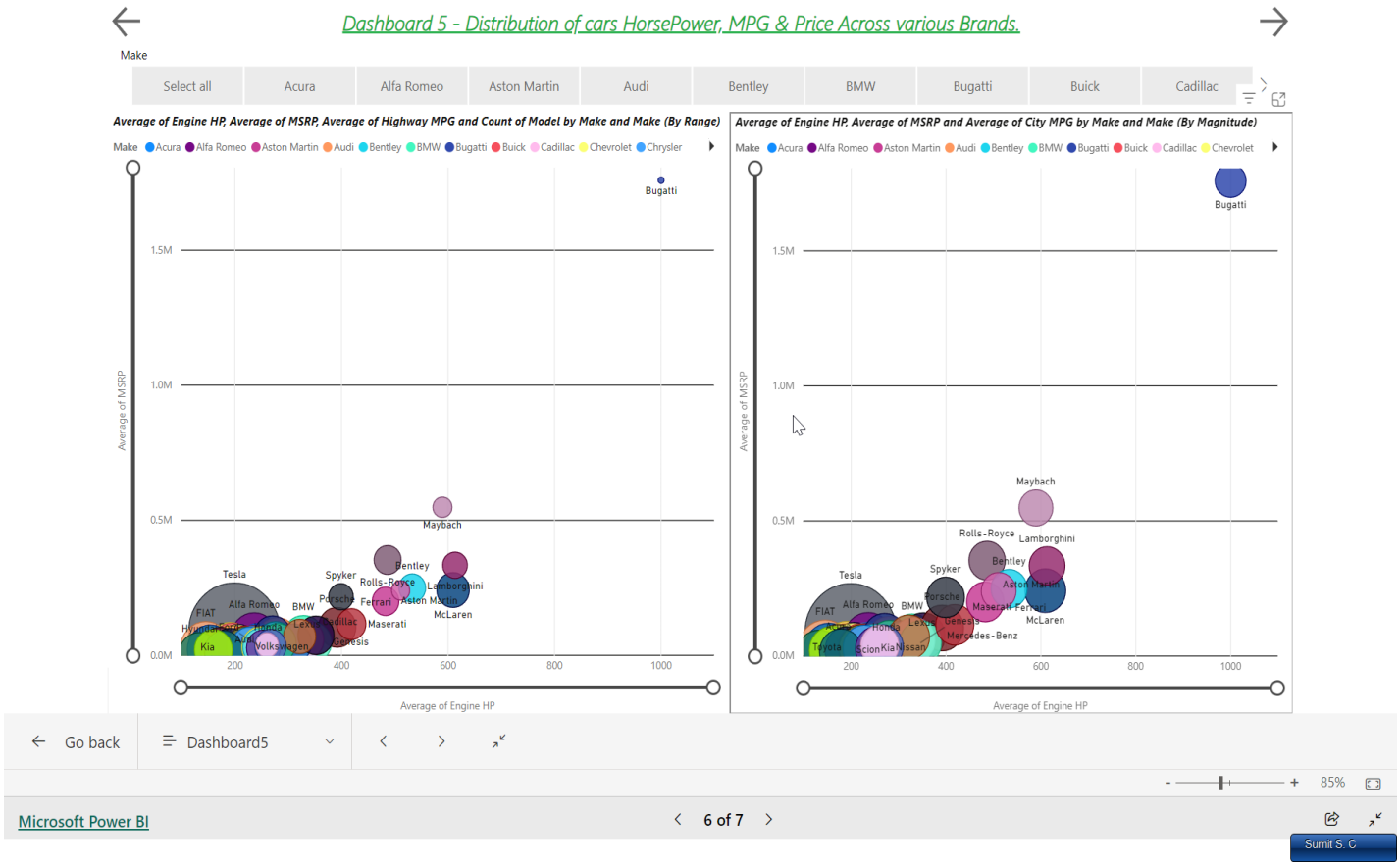
1. The task helped us to make a dashboard to see the variation of fuel efficiency over the body style & year of the model of car.

2. The year slicer helps us to get to know how has the technology improved the mileage over the years and it does show significant changes.

3. The body style shows us the variation of mileage to showing us the impacts of different type of vehicles.

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

- **Hints:** Bubble chart to visualize the relationship between horsepower, MPG, and price across different car brands. Assign different colors to each brand and label the bubbles with the car model name. Calculate the average horsepower, MPG, and MSRP for each car brand using AVERAGEIFS or Pivot Tables.



Insights:

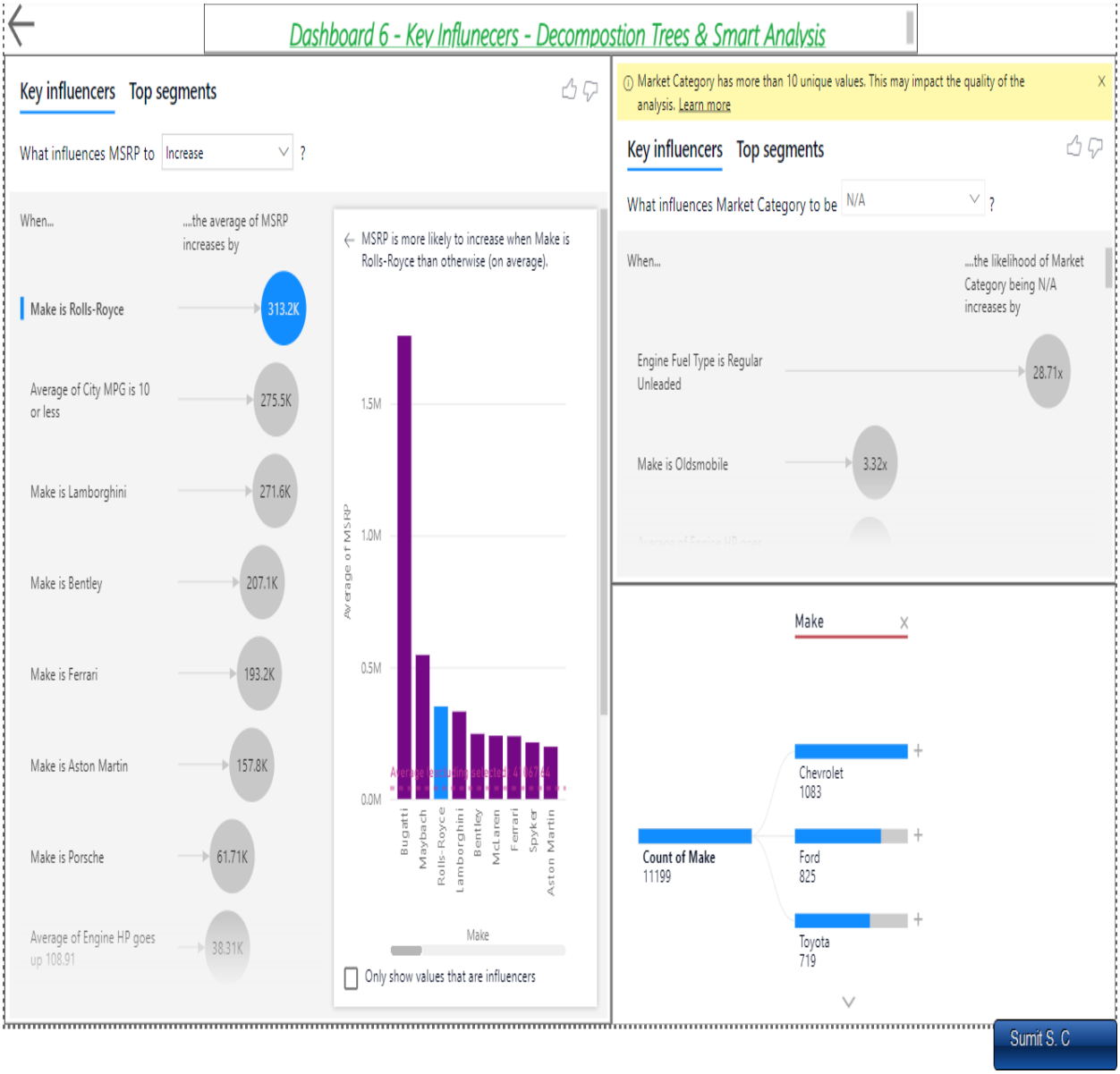
1. The question helped us to understand the engine HP, Efficiency & MSRP of brands distribution.

2. In the 1st graph I used the highway mpg and choose the bubble marker to be in range param so the bubbles changed according to the overall factors present across each calculation.

3. While in 2nd graph I used the city MPG and choose the bubble over magnitude so as the value increases for factors the size changes in size too but its same as previous chart in working.

Extra – Used the Power BI KPI & decomposition function to get automated analysis. (Just like our main analysis and title of project is – To see the impact of different factors on car price we used the BI intelligent features to drill the data to get more insights.)

Dashboard 6 (Extra – KPI, Decomposition tree to find the insights of data)[Due to extensive data drilling the web view might not show all KPI but the .pbix file will run the analysis)



Insights :

1. For the required analysis category, we drilled it over other columns which were turned into parameters for our main question and then the AI tool calculated various parameters and segments and gave us key insights it got from our data.
2. In Decomposition tree we get the same analysis but we do it manually we go on drilling the main criteria on various sub criteria and the results changes every time with change in the parameter selection.
3. This advanced function in Power BI helps us to understand the data more beautifully and can help us to learn about the insights and factors impacting the overall dataset.

Important Links:

Drive Folder Link (Has all the files & readme for navigation for loom ppt , files, reports & dashboard)

Individual Excel Sheets (Individual Questions File)

Final Excel sheet (including cleaned dataset)

Word File Link & Pdf File (Will Be in the drive folder – can't add before I upload the file)

Video Presentation

Loom Folder

Power BI Dashboard Links (Web, Deployed, Drive-File)

Screengrabs & Gif of dashboard & process (Drive Folder)

Thank You