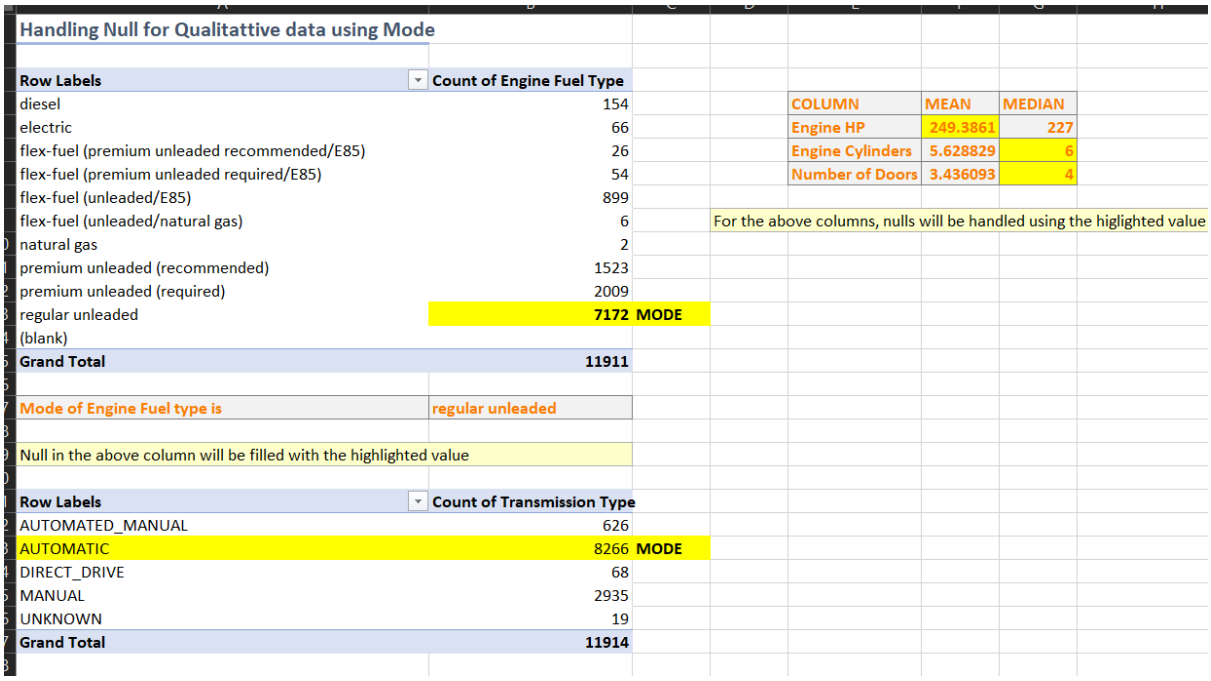
Analysing the Impact of Car Features on Price and Profitability

Hyperlink of excel file:
https://docs.google.com/file/d/1ymdhY4QJ6jhKxOfuYNRSkjAtPhDKQZgo/edit?usp=docslist_api&filetype=msexcel

Hyperlink of video presentation:
https://drive.google.com/file/d/1ycUJT2jopV8lM7vt343ypsyMnA7Ec2yJ/view?usp=drivesdk

Data Cleaning:

Below is the glimpse of data cleaning operation performed



Handling Null for Qualitattive data using Mode

Row Labels	Count of Engine Fuel Type	
diesel	154	
electric	66	
flex-fuel (premium unleaded recommended/E85)	26	
flex-fuel (premium unleaded required/E85)	54	
flex-fuel (unleaded/E85)	899	
flex-fuel (unleaded/natural gas)	6	
natural gas	2	
premium unleaded (recommended)	1523	
premium unleaded (required)	2009	
regular unleaded	7172	MODE
(blank)		
Grand Total	11911	

COLUMN	MEAN	MEDIAN
Engine HP	249.3861	227
Engine Cylinders	5.628829	6
Number of Doors	3.436093	4

For the above columns, nulls will be handled using the higlighted value

Mode of Engine Fuel type is	regular unleaded

Null in the above column will be filled with the highlighted value

Row Labels	Count of Transmission Type	
AUTOMATED_MANUAL	626	
AUTOMATIC	8266	MODE
DIRECT_DRIVE	68	
MANUAL	2935	
UNKNOWN	19	
Grand Total	11914	

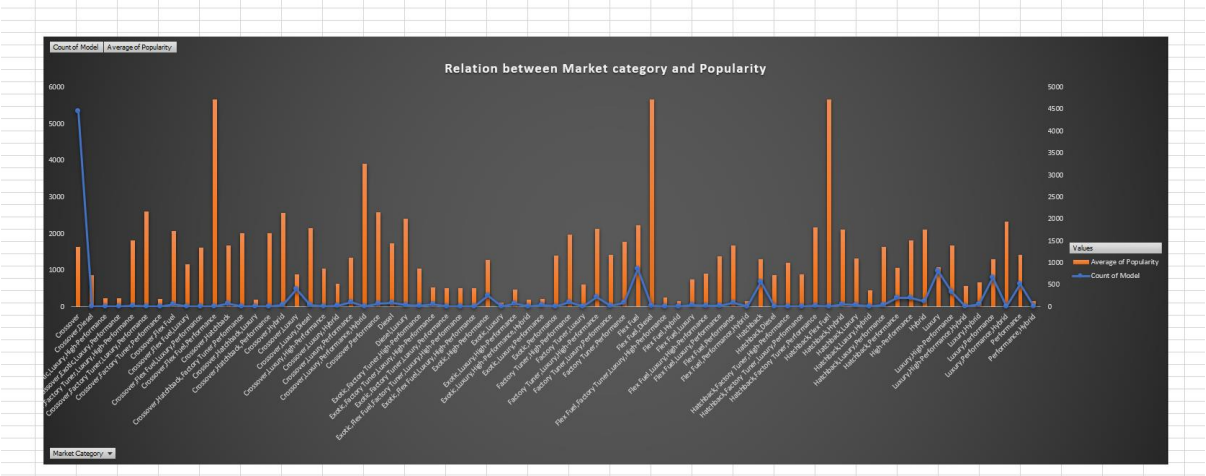
In the given dataset, there were 3 quantitative and 3 qualitative data to be handled and was handled qualitative data using mode and quantitative using mean and mode.

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

Market Category	Count of Model	Average of Popularity
Crossover	4451	1638.588407
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	406	889.2142857
Crossover,Luxury,Diesel	34	2149.411765
Crossover,Luxury,High-Performance	9	1037.222222
Crossover,Luxury,Hybrid	24	630.9166667
Crossover,Luxury,Performance	112	1349.089286
Crossover,Luxury,Performance,Hybrid	2	3916
Crossover,Performance	69	2585.956522
Diesel	84	1730.904762
Diesel,Luxury	47	2416.106383
Exotic,Factory Tuner,High-Performance	21	1046.380952
Exotic,Factory Tuner,Luxury,High-Performance	51	523.0196078
Exotic,Factory Tuner,Luxury,Performance	3	520
Exotic,Flex Fuel,Factory Tuner,Luxury,High-Performanc	13	520
Exotic,Flex Fuel,Luxury,High-Performance	11	520
Exotic,High-Performance	254	1280.047244
Exotic,Luxury	12	112.6666667
Exotic,Luxury,High-Performance	77	473.025974
Exotic,Luxury,High-Performance,Hybrid	1	204
Exotic,Luxury,Performance	36	217.0277778
Exotic,Performance	10	1391
Factory Tuner,High-Performance	104	1966.442308

To arrive at the above result, I created a pivot table using the final data and added market category in the row field and count of model and average of popularity in value field.

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

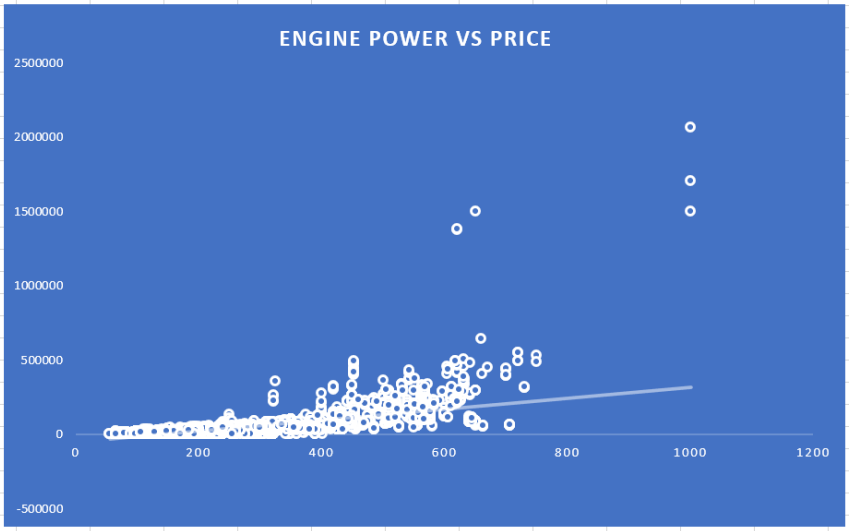


Above combo chart was created using the result of the pivot table , column chart represents the average of popularity and line represents the count of model

INSIGHT: The popularity of a car model is significantly influenced by its market category, with performance, luxury, and exotic segments showing higher demand. At the same time, a larger number of models in a market category doesn't necessarily equate to higher popularity.

- **Task 2:** Below is a scatter chart that plots engine power on the x-axis and price on the y-axis. The trendline to the chart to visualize the relationship between these variables.

Engine HP	MSRP
335	46135
300	40650
300	36350
230	29450
230	34500
230	31200
300	44100
300	39300
230	36900
230	37200
300	39600
230	31500
300	44400
230	37200
320	48250
320	43550
172	2000
172	2000
172	2000
172	2000
172	2000
172	2000
172	2000
172	2000
172	2000
172	2000



Above chart was created by separately taking two columns Engine HP and MSRP in a separate sheet and a scatter chart was plotted using the result with a trendline added later on.

INSIGHT: The trendline indicates a **positive correlation** between engine power and price, meaning that as engine horsepower increases, car prices tend to rise. However, the slope is relatively modest, suggesting other factors also influence price beyond just engine power.

- **Task 3:** Below is a regression analysis to identify the variables that have the strongest relationship with a car's price. And a bar chart that shows the coefficient values for each variable to visualize their relative importance.

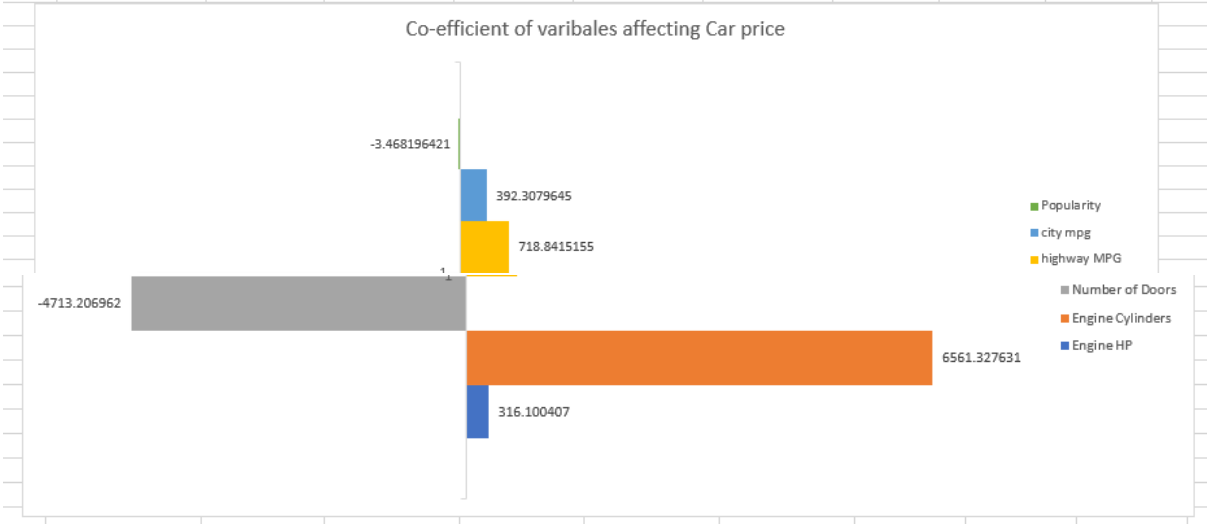
SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.679184837
R Square	0.461292042
Adjusted R Square	0.461003242
Standard Error	45176.81596
Observations	11199

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	6	1.95597E+13	3.25994E+12	1597.27253	0
Residual	11192	2.28423E+13	2040944700		
Total	11198	4.24019E+13			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-80528.23831	3554.297338	-22.65658459	3.7159E-111	-87495.28654	-73561.1901	-87495.2865	-73561.1901
Engine HP	316.100407	6.24836011	50.58933888	0	303.8525217	328.3482924	303.8525217	328.3482924
Engine Cylinders	6561.327631	445.2155862	14.73741674	1.05916E-48	5688.626738	7434.028524	5688.626738	7434.028524
Number of Doors	-4713.206962	496.1650682	-9.499272045	2.54292E-21	-5685.777805	-3740.63612	-5685.77781	-3740.63612
highway MPG	718.8415155	106.9993693	6.718184603	1.92929E-11	509.1039231	928.5791079	509.1039231	928.5791079
city mpg	392.3079645	100.9970609	3.884350307	0.000103195	194.3359529	590.2799762	194.3359529	590.2799762
Popularity	-3.468196421	0.296181285	-11.70970821	1.73458E-31	-4.048763858	-2.88762898	-4.04876386	-2.88762898



Above result was calculated by using the excel feature known as data analysis tool kit from which I regression analysis was used. While doing so it asked for independent column which is MSRP and then for dependent columns which are Engine HP, Engine Cylinder, Number of doors, Highway MPG, City MPG and Popularity.

INSIGHT: The most important car features in determining a car's price are:

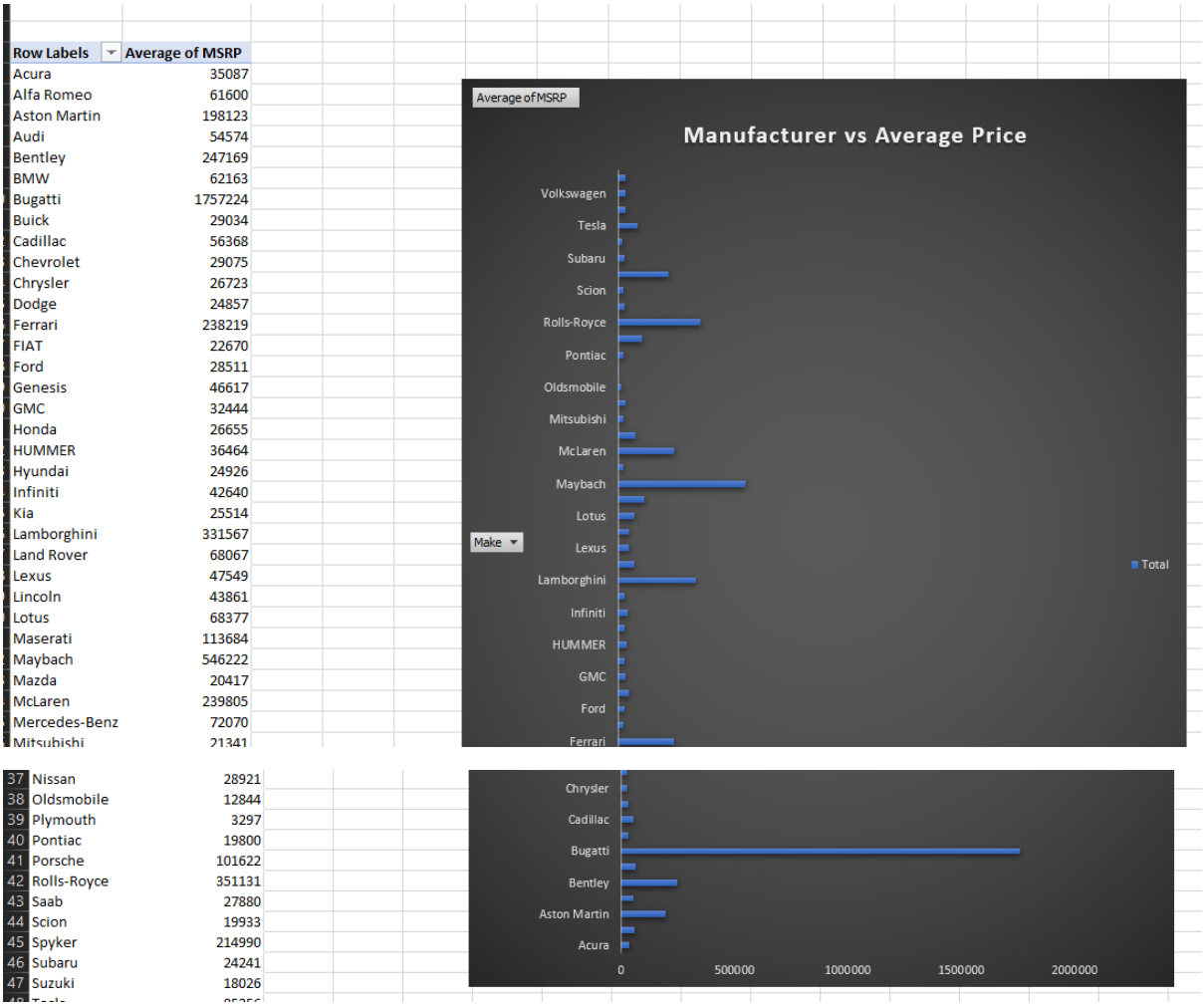
1. **Engine HP** – It has a significant positive impact on price, with a high coefficient (316.10) and strong statistical significance (P-value = 0).
2. **Engine Cylinders** – Also positively influences price with a high coefficient (6561.33) and strong significance (P-value = 1.06E-48).
3. **Highway MPG** and **City MPG** – Both positively affect price, but with smaller coefficients compared to engine features.
4. **Popularity** – Negatively impacts price with a small but significant negative coefficient (-3.47).

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

Row Labels	Average of MSRP
Acura	35087
Alfa Romeo	61600
Aston Martin	198123
Audi	54574
Bentley	247169
BMW	62163
Bugatti	1757224
Buick	29034
Cadillac	56368
Chevrolet	29075
Chrysler	26723
Dodge	24857
Ferrari	238219
FIAT	22670
Ford	28511
Genesis	46617
GMC	32444
Honda	26655
HUMMER	36464
Hyundai	24926
Infiniti	42640
Kia	25514
Lamborghini	331567
Land Rover	68067
Lexus	47549
Lincoln	43861
Lotus	68377
Maserati	113684
Maybach	546222
Mazda	20417
McLaren	239805
Mercedes-Benz	72070
Mitsubishi	21341

  To arrive at the above result, I initially created pivot table from the final data and took make in the rows section and average of MSRP in the column section.

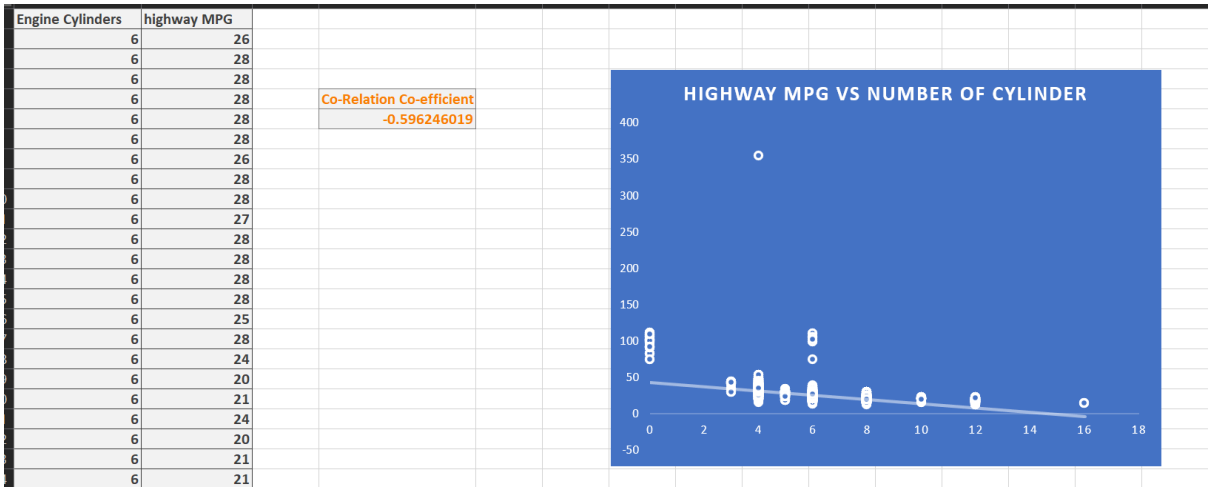
- **Task 4.B: Below is a horizontal stacked bar chart that visualizes the relationship between manufacturer and average price.**



Above chart was created by using the result of pivot table created earlier.

INSIGHT: The average price of cars varies significantly across different manufacturers. Luxury brands such as Bugatti, Bentley, and Ferrari show considerably higher average prices. On the other hand, manufacturers like Chevrolet, Ford, and Hyundai are positioned at the lower end with more affordable average prices.

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



Below chart was created by separately taking two columns engine cylinders and highway MPG on new sheet and creating a chart using the result

- **Task 5.B: Below is the correlation coefficient between the number of cylinders and highway MPG.**

Above result was calculated using the excel function COREL which inputted two columns Number of Cylinder and highway MPG.

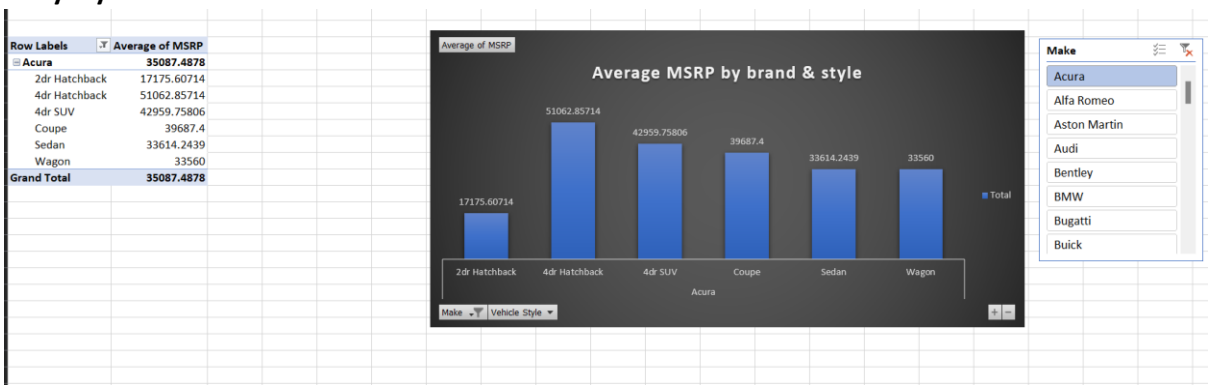
INSIGHT: The correlation coefficient of -0.596 indicates a moderate negative relationship between fuel efficiency and the number of cylinders in a car's engine. This means that as the number of cylinders increases, fuel efficiency tends to decrease.

BUILDING THE DASHBOARD

- **Task 1:** Below is stacked column chart showing the distribution of car prices vary by brand and body style

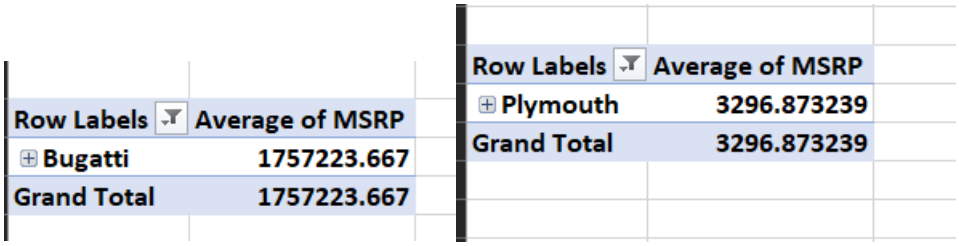
Above is interactive chat with a slicer for make showing the total price by style. And it was created using a pivot table

- **Task 2: Clustered column chart showing the average MSRPs across different car brands and body styles**



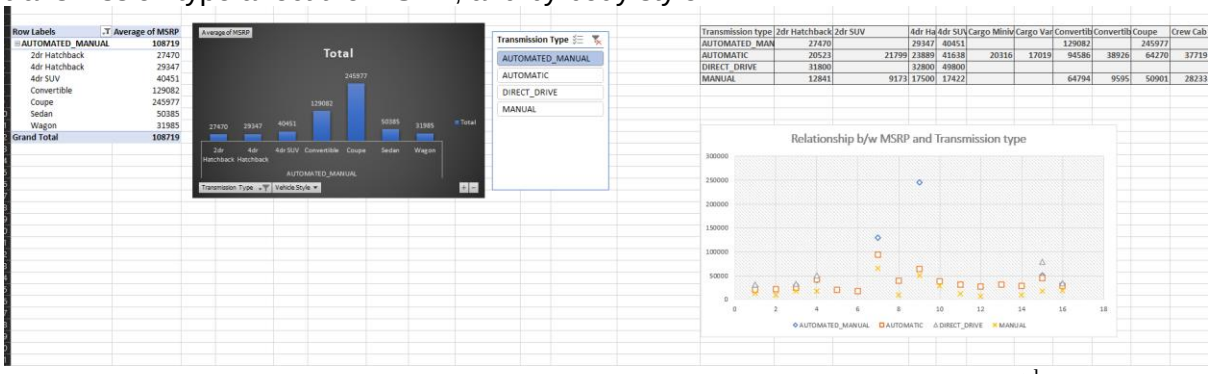
Again, this an interactive chart with slicer to Make/Brand showing average price of the car across different body style of the car. The above result was created using a pivot table.

Car brands having the highest and lowest average MSRPs is given below



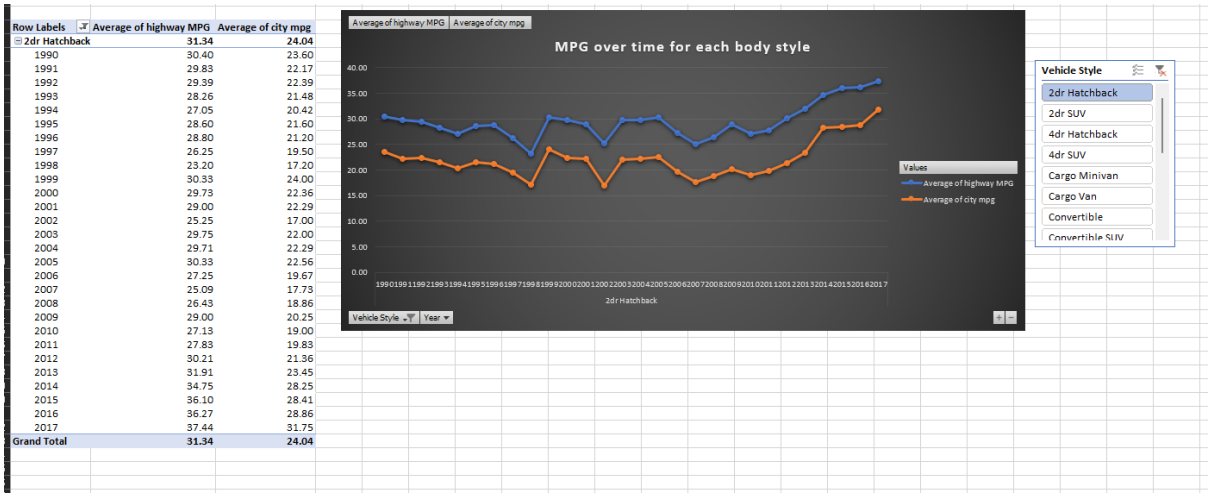
Bugatti has the highest average MSRP while Plymouth has the lowest average MSRP.

- **Task 3:** Below is the scatter plot chart showing the different feature such as transmission type affect the MSRP, and by body style



And this was calculated using pivot table with slicer to transmission type and 2nd image shows the relation between MSRP and Transmission type.

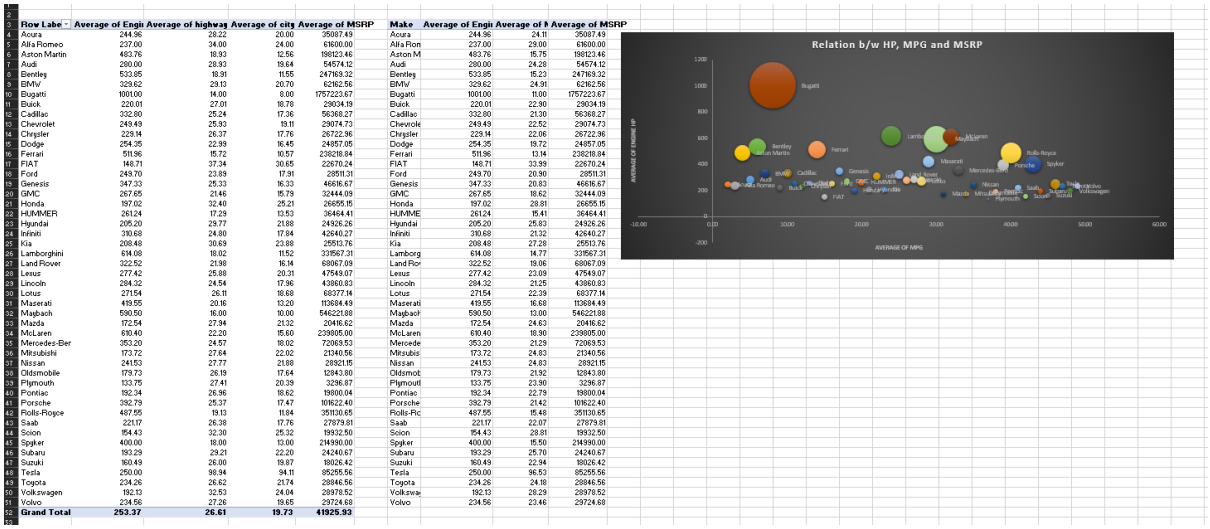
- **Task 4:** Below is the line chart showing the trend of fuel efficiency (MPG) over time for each body style and the average MPG for each combination of body style.



Row Labels	Average of highway MPG	Average of city mpg
2dr Hatchback	31.34	24.04
1990	30.40	23.60
1991	29.83	22.17
1992	29.39	22.39
1993	28.26	21.48
1994	27.05	20.42
1995	28.60	21.60
1996	28.80	21.20
1997	26.25	19.50
1998	23.20	17.20
1999	30.33	24.00
2000	29.73	22.36
2001	29.00	22.29
2002	25.25	17.00
2003	29.75	22.00
2004	29.71	22.29
2005	30.33	22.56
2006	27.25	19.67
2007	25.09	17.73
2008	26.43	18.86
2009	29.00	20.25
2010	27.13	19.00
2011	27.83	19.83
2012	30.21	21.36
2013	31.91	23.45
2014	34.75	28.25
2015	36.10	28.41
2016	36.27	28.86
2017	37.44	31.75
Grand Total	31.34	24.04

Again, a pivot table was used with car style as a row and average of city and highway MPG as values. A slicer was used on vehicle style to extract MPG over years to a specific style.

- **Task 5: Below is bubble chart showing the car's horsepower, MPG, and price across different Brands**

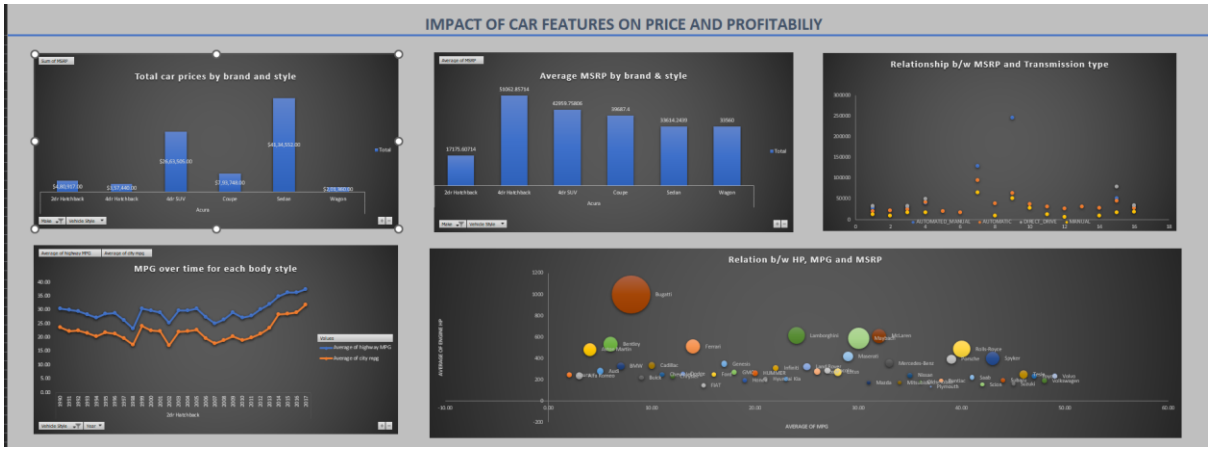


Row Labe	Average of Engi	Average of highway	Average of city	Average of MSRP
Acura	244.96	28.22	20.00	35087.49
Alfa Romeo	237.00	34.00	24.00	61600.00
Aston Martin	483.76	16.93	12.56	198123.46
Audi	280.00	28.93	19.64	54574.12
Bentley	533.85	18.91	11.55	247169.32
BMW	329.62	29.13	20.70	62162.56
Bugatti	1001.00	14.00	8.00	1757223.67
Buick	220.01	27.01	18.78	29034.19
Cadillac	332.80	25.24	17.36	56368.27
Chevrolet	249.49	25.93	19.11	29074.73
Chrysler	229.14	26.37	17.76	26722.96
Dodge	254.35	22.99	16.45	24857.05
Ferrari	511.96	15.72	10.57	238218.84
FIAT	148.71	37.34	30.65	22670.24
Ford	249.70	23.89	17.91	28511.31
Genesis	347.33	25.33	16.33	46616.67
GMC	267.65	21.46	15.79	32444.09
Honda	197.02	32.49	25.21	26695.15
HUMMER	261.24	17.29	13.53	36464.41
Hyundai	205.20	29.77	21.88	24926.26
Infiniti	330.68	24.89	17.84	42640.27
Kia	208.48	30.69	23.88	25610.76
Lamborghini	614.08	18.02	11.52	331567.31
Land Rover	322.52	21.98	16.14	69067.09
Lexus	277.42	25.88	20.31	47549.07
Lincoln	284.32	24.54	17.96	43860.83
Lotus	271.54	26.11	18.68	69377.14
Maserati	419.55	20.16	13.20	112694.49
Maybach	590.50	16.00	10.00	546221.88
Mazda	172.54	27.94	21.32	20416.62
McLaren	610.40	22.20	15.60	239805.00
Mercedes-Ber	353.20	24.57	18.02	72069.53
Mitsubishi	173.72	27.64	22.02	21340.56
Nissan	241.53	27.77	21.88	28921.15
Oldsmobile	179.73	26.19	17.64	12943.80
Plymouth	133.75	27.41	20.39	3296.87
Pontiac	192.34	26.96	18.62	19800.04
Porsche	392.79	25.37	17.47	101622.40
Rolls-Royce	497.55	19.13	11.84	351130.65
Saab	221.17	26.38	17.76	27879.81
Scion	154.43	32.30	25.32	19932.50
Spyker	400.00	18.00	13.00	214990.00
Subaru	193.29	29.21	22.20	24240.67
Suzuki	160.49	26.00	19.87	18026.42
Tesla	250.00	98.94	94.11	85255.56
Toyota	234.26	25.62	21.74	28946.56
Volkswagen	192.13	32.53	24.04	28978.52
Volvo	234.56	27.26	19.65	29724.68
Grand Total	253.37	26.61	19.73	41925.93

To arrive at the above output, I used pivot table with make as row and average of engine HP, average of highway and city MPG and average of MSRP. The separately took all these columns and merged city and highway MPG to arrive at 4 final columns and used this result to draw a bubble chart as shown above. Each car brand with different bubble colour.

Final Dashboard:

Below is the final interactive dashboard using the above results.



Project Description: The Project is based on the impact of car features on price and profitability and its purpose is to test and improve my data analytics skill. The business problems are the car features affecting the price and profitability. Data used is the data of cars over the years with vital features of different brands. Data cleaning process is explained above

Approach: The excel and statistical methods and all the relevant approaches based on this was used

Tech-Stack Used: Microsoft Excel 2019

Insight: Provided above for the required fields

Result: Provided and explained in detail above**.**