Analyzing the Impact of Car Features on Price and Profitability

VICTOR SHAH

Dataset link : https://docs.google.com/spreadsheets/d/1-UzV0aXl0rl3WalFDZteQvRa6xe1atgG/edit?usp=drive_link&ouid=111667124296828215097&rtpof=true&sd=true

Video link: https://drive.google.com/file/d/1G-aZ9bvKMX0BE7jhyfjr_1e2hStIKGSm/view?usp=drive_link

Tableau link: https://public.tableau.com/app/profile/victor.shah/viz/AnalyzingtheImpactofCarFeaturesonPriceandProfitability_17050770668790/Dashboard1#2

Project Description

As a data analyst my primary responsibility is how can a car manufacturer optimize pricing and product development decisions to maximize profitability while meeting consumer demand.

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 automotive industry has been rapidly evolving over the past few decades, with a growing focus on fuel efficiency, environmental sustainability, and technological innovation.

Approach

- Downloading the dataset: The first step is downloading the excel file (.csv) into the local device. Make sure the downloaded file is having the extension (.xlsx).
- Understanding the worksheet: The next step is to examine the structure of the table holding the data in the Excel Sheet. (Project 7 Dataset)
- Identifying the key tables: Identification of the primary key from the dataset of excel files.
- Data Cleaning: This is the preprocessing step that makes the data suitable for analysis. It includes handling missing values, removing duplicates.
- Data Visualization: To use EDA to understand how car features affects the market prices for each brand.
- Using Tableau: To further enhance the visualization, the dashboard is made to represent the findings to the stakeholders or other employees.

DATA ANALYTICS TASKS

Insight Required: How does the popularity of a car model vary across different market categories?

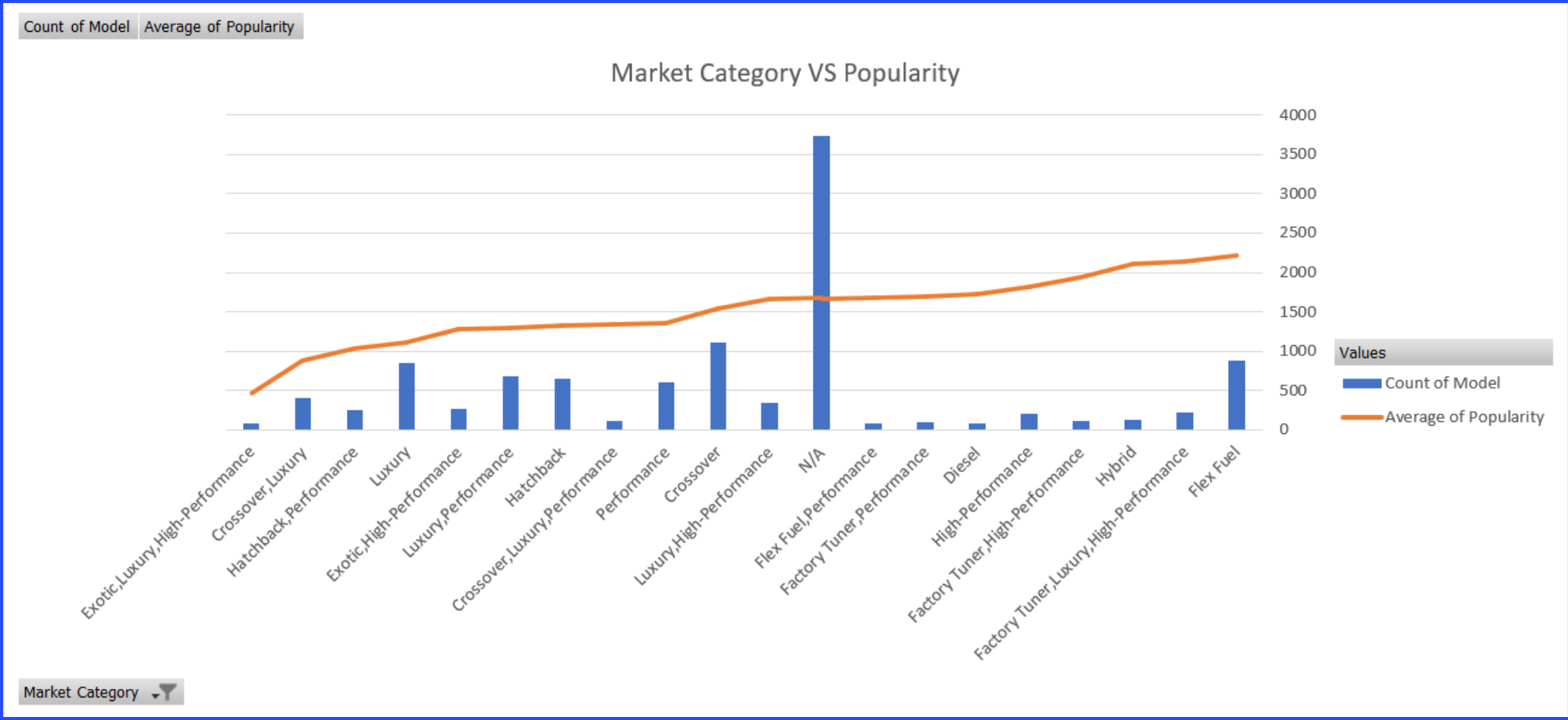
Task 1.A: Create a pivot table that shows the number of car models in each market category and their corresponding popularity scores.

The table beside shows the top 20 Market Category's corresponding to their popularity scores.

Market Category	Count of Model	Average of Popularity
Flex Fuel	872	2217.302752
Factory Tuner,Luxury,High-Performance	215	2133.367442
Hybrid	123	2105.569106
Factory Tuner,High-Performance	106	1941.415094
High-Performance	199	1821.447236
Diesel	84	1730.904762
Factory Tuner,Performance	92	1695.695652
Flex Fuel,Performance	87	1680.471264
N/A	3742	1676.889364
Luxury,High-Performance	334	1668.017964
Crossover	1110	1545.263063
Performance	601	1348.873544
Crossover,Luxury,Performance	113	1344.849558
Hatchback	641	1318.865835
Luxury,Performance	673	1292.615156
Exotic,High-Performance	261	1271.333333
Luxury	855	1102.65731
Hatchback,Performance	252	1039.646825
Crossover,Luxury	410	884.5487805
Exotic,Luxury,High-Performance	79	467.0759494
Grand Total	**10849**	**1551.14711**

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

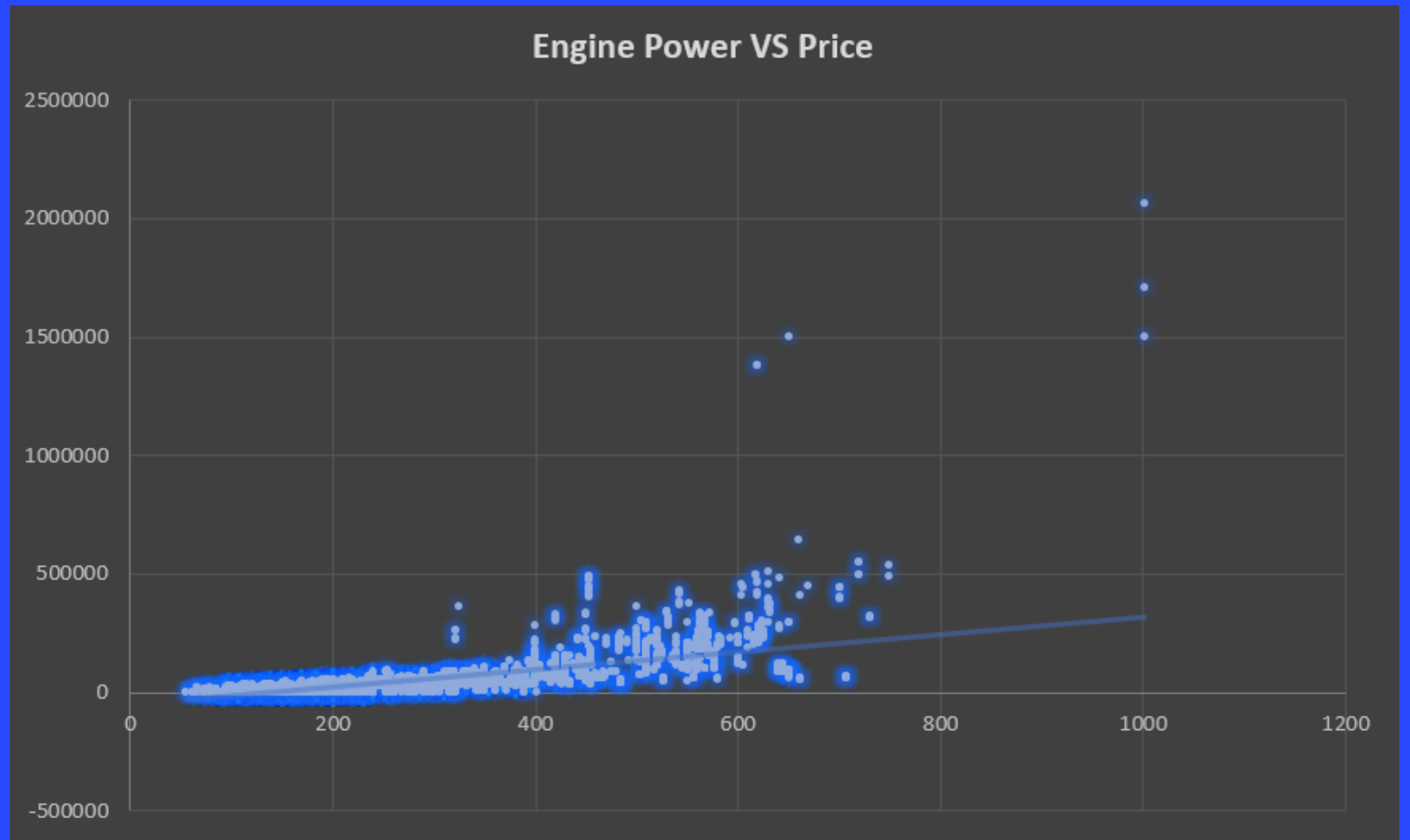
The chart below is a combo chart that shows the top 20 market categories based on their popularity scores.



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.

The chart below is a scatter chart having a rising trendline. This means the price increases with the engine power.



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.

The table below is the regression analysis to identify the relationship with the car's price.

SUMMARY OUTPUT

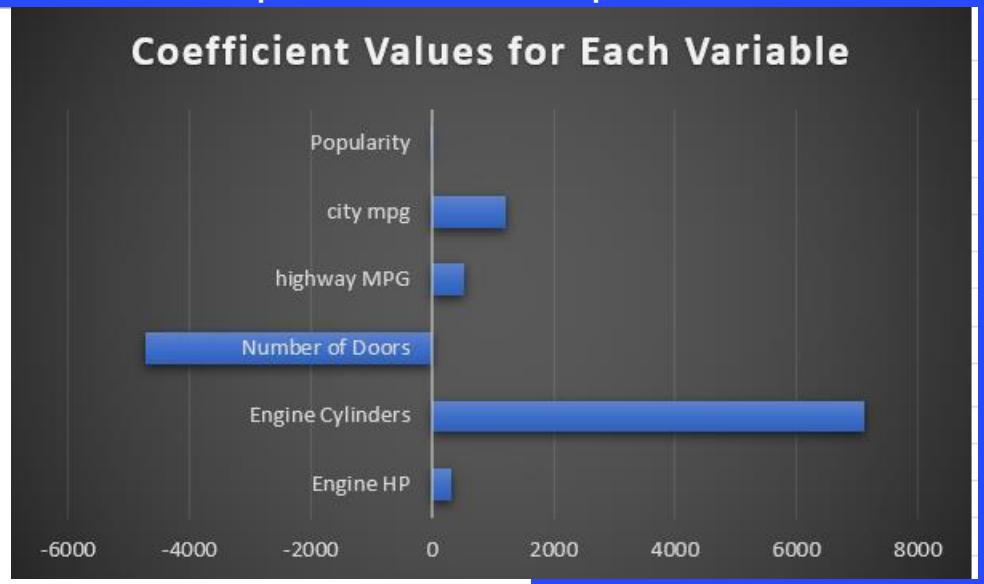
Regression Statistics	
Multiple R	0.685306311
R Square	0.46964474
Adjusted R Square	0.46937525
Standard Error	43908.53542
Observations	11815

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	6	2.01594E+13	3.35989E+12	1742.720244	0
Residual	11808	2.27653E+13	1927959483		
Total	11814	4.29247E+13			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-96512.01109	3687.749498	-26.17097803	8.3335E-147	-103740.6082	-89283.41393	-103740.6082	-89283.41393
Engine HP	323.0463366	5.979918896	54.02185919	0	311.3247094	334.7679638	311.3247094	334.7679638
Engine Cylinders	7118.342131	437.0790222	16.28616742	5.4287E-59	6261.59517	7975.089093	6261.59517	7975.089093
Number of Doors	-4721.527663	463.4140498	-10.18857254	2.81017E-24	-5629.895621	-3813.159704	-5629.895621	-3813.159704
highway MPG	534.0108739	105.1984966	5.076221534	3.90877E-07	327.8044724	740.2172754	327.8044724	740.2172754
city mpg	1219.711775	121.3635325	10.05006817	1.14281E-23	981.819237	1457.604312	981.819237	1457.604312
Popularity	-3.324711821	0.281317034	-11.81838077	4.75336E-32	-3.876139599	-2.773284044	-3.876139599	-2.773284044

The Bar Chart besides shows the coefficient values for each variable



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

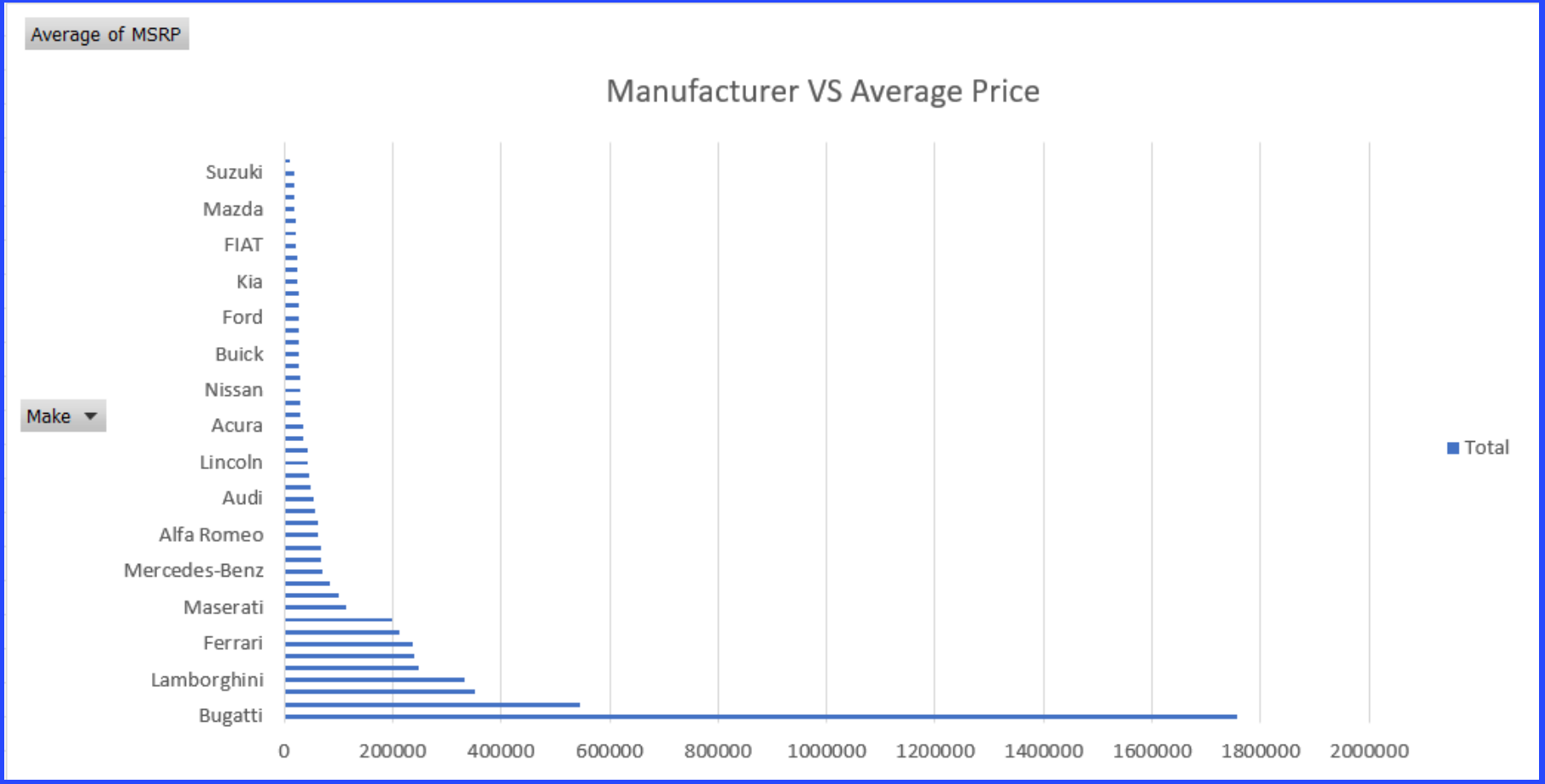
Task 4.A: Create a pivot table that shows the average price of cars for each manufacturer.

The table beside shows the average price for each car brand/manufacturer.

Manufacturer	Average of MSRP
Bugatti	1757223.667
Maybach	546221.875
Rolls-Royce	351130.6452
Lamborghini	331567.3077
Bentley	247169.3243
McLaren	239805
Ferrari	238218.8406
Spyker	213323.3333
Aston Martin	197910.3763
Maserati	114207.7069
Porsche	101622.3971
Tesla	85255.55556
Mercedes-Benz	71476.22946
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	42839.82927
Infiniti	42394.21212
HUMMER	36464.41176
Acura	34887.5873
GMC	30493.29903
Toyota	29030.01609
Nissan	28583.4319
Volvo	28541.16014
Chevrolet	28350.38557
Buick	28206.61224
Volkswagen	28102.38072
Saab	27413.5045
Ford	27399.26674
Chrysler	26722.96257
Honda	26674.34076
Kia	25310.17316
Subaru	24827.50391
Hyundai	24597.0363
FIAT	22670.24194
Dodge	22390.05911
Mitsubishi	21240.53521
Mazda	20039.38298
Scion	19932.5
Pontiac	19321.54839
Suzuki	17907.20798
Oldsmobile	11542.54
Plymouth	3122.902439
Grand Total	**40594.73703**

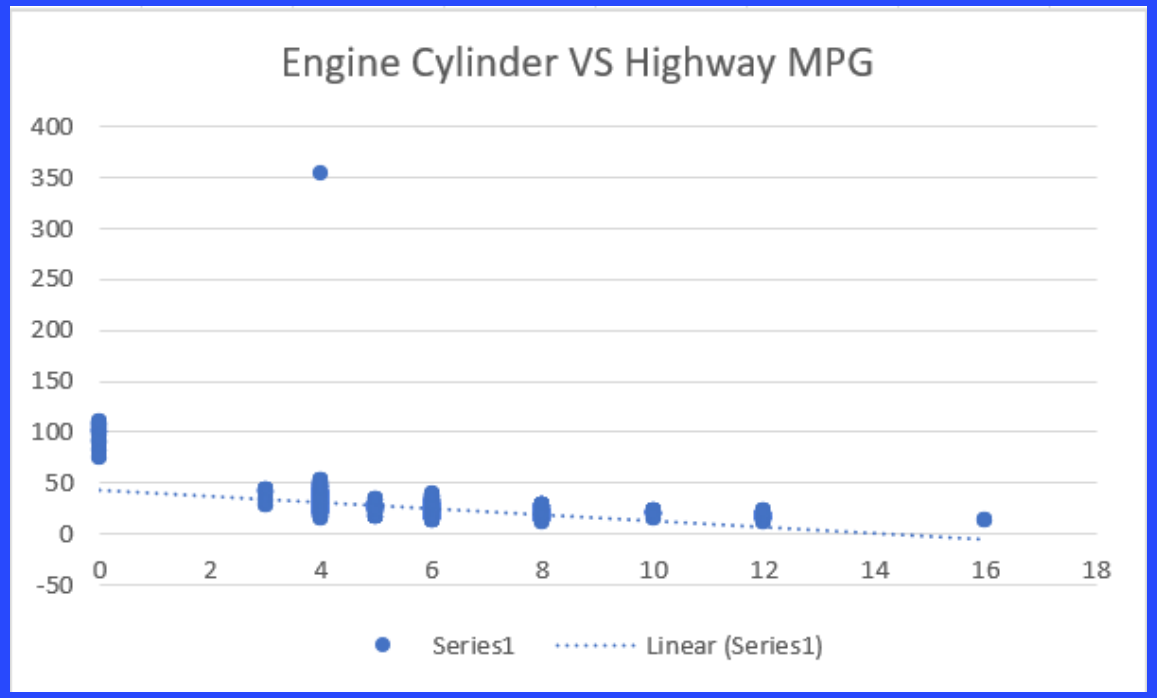
Task 4.B: Create a bar chart or a horizontal stacked bar chart that visualizes the relationship between manufacturer and average price.

The chart below is a horizontal bar chart that visualizes the relationship between the manufacturer/brand and the average price.



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.



The chart beside shows the scatter plot with the number of cylinders on the x-axis and highway MPG on the y-axis.

The negative trendline shows that the increase in the engine cylinder leads towards the decrease in the Hight MPG

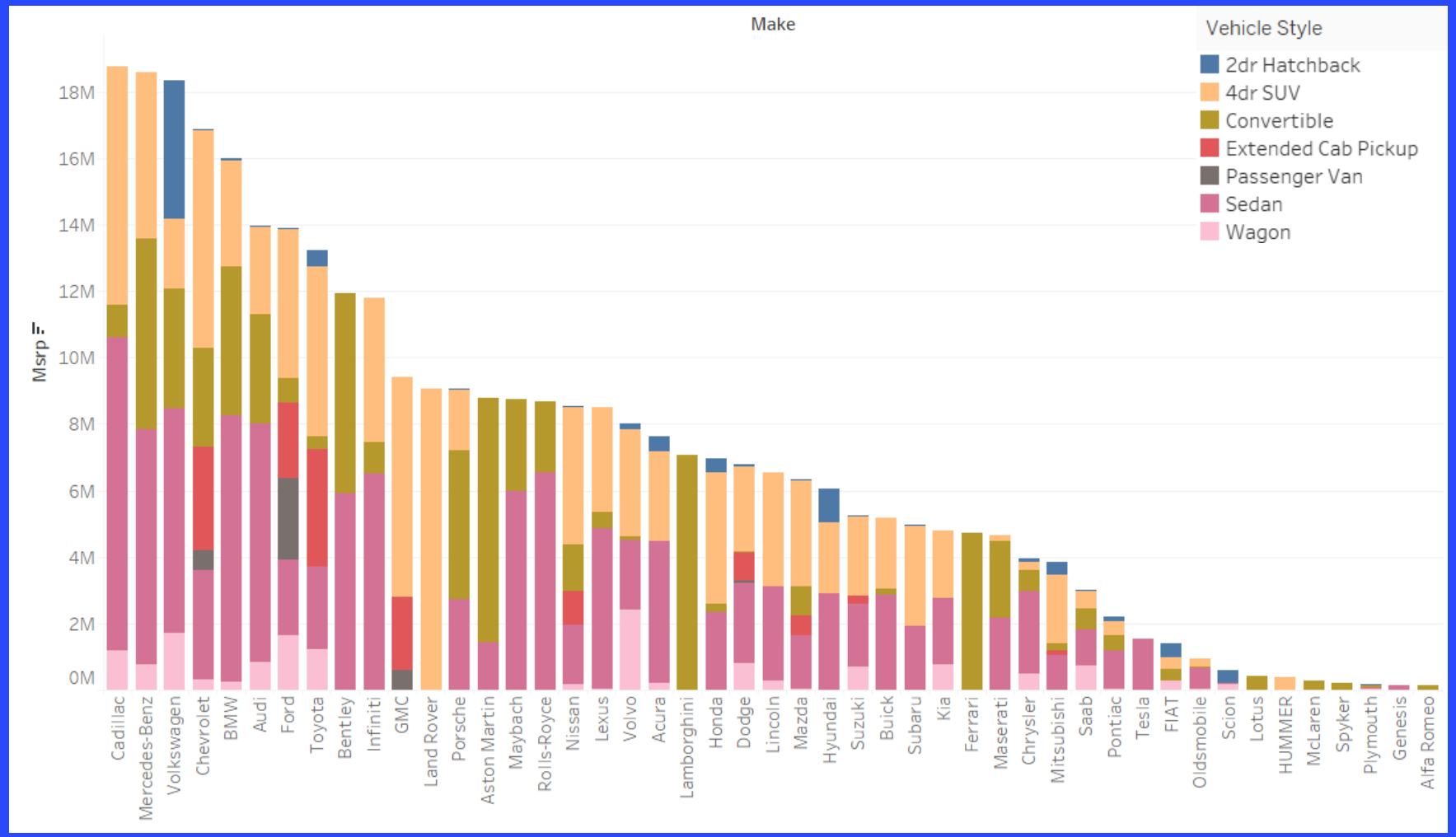
Task 5.B: Calculate the correlation coefficient between the number of cylinders and highway MPG to quantify the strength and direction of the relationship.

Correlation Coefficient between Engine Cylinder and Highway MPG is	-0.621606
The negative value means that the trendline will move in the downward direction	

BUILDING THE DASHBOARD

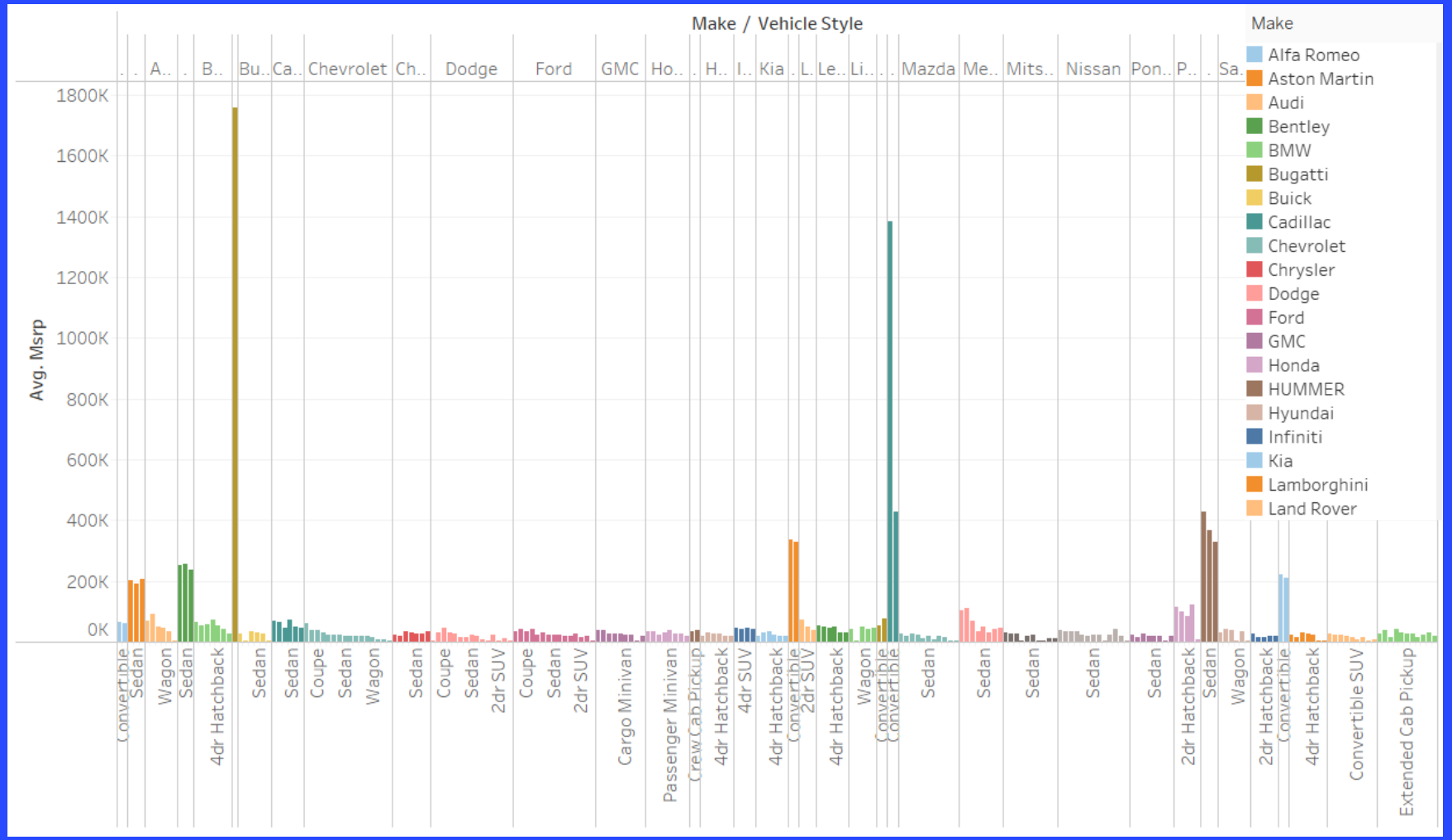
Task 1: How does the distribution of car prices vary by brand and body style?

Brand	Sum of MSRP
Acura	**8791672**
2dr Hatchback	480917
4dr Hatchback	357440
4dr SUV	2663505
Coupe	793748
Sedan	4294702
Wagon	201360
Alfa Romeo	**308000**
Convertible	129800
Coupe	178200
Aston Martin	**18405665**
Convertible	7321655
Coupe	9635275
Sedan	1448735
Audi	**17532293**
2dr Hatchback	4000
4dr SUV	2674900
Convertible	3291405
Coupe	3556290
Sedan	7158348
Wagon	847350
Bentley	**18290530**



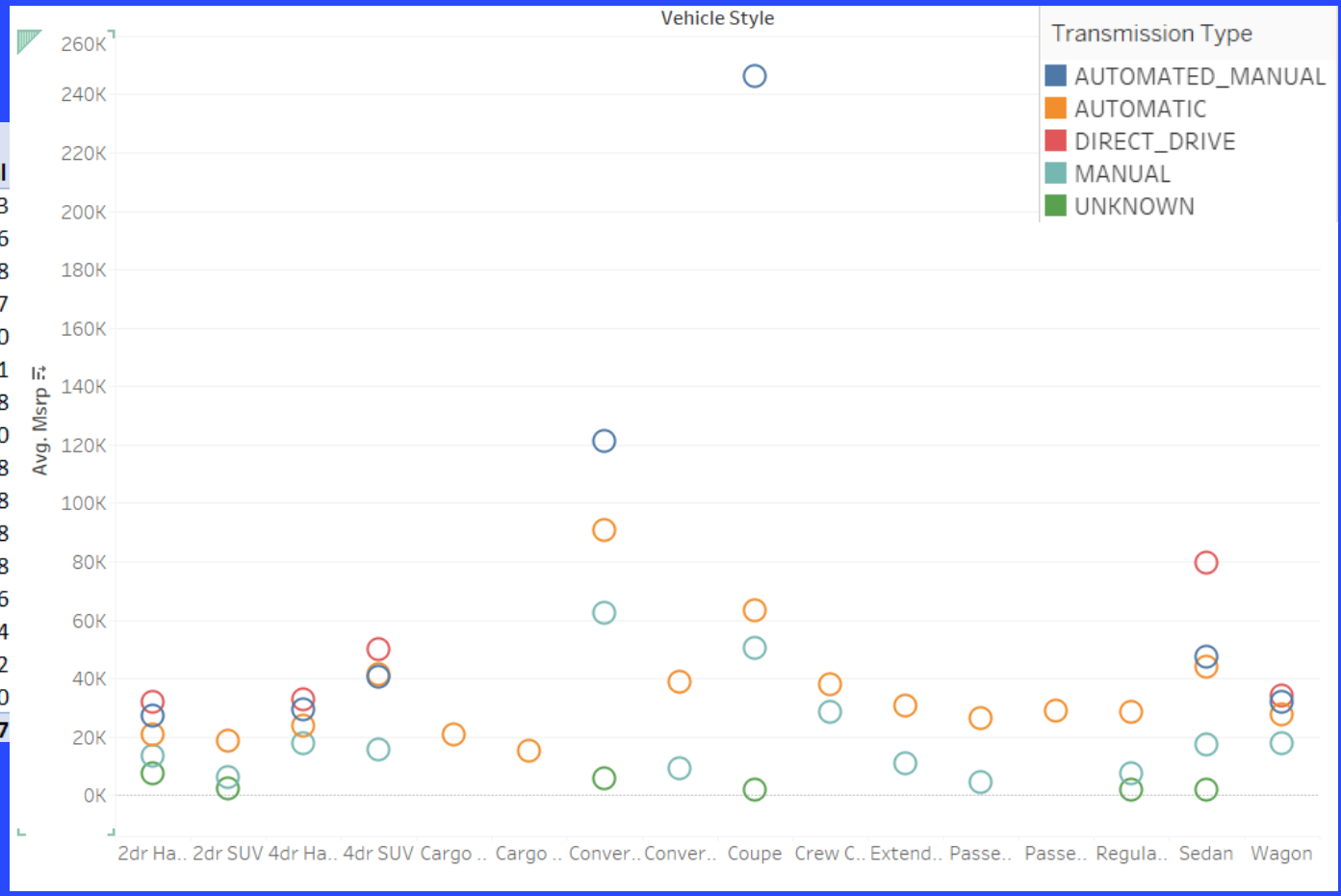
Task 2: Which car brands have the highest and lowest average MSRPs, and how does this vary by body style?

Brand	Average of MSRP
Acura	**34887.5873**
2dr Hatchback	17175.60714
4dr Hatchback	51062.85714
4dr SUV	42959.75806
Coupe	39687.4
Sedan	33292.26357
Wagon	33560
Alfa Romeo	**61600**
Convertible	64900
Coupe	59400
Aston Martin	**197910.3763**
Convertible	203379.3056
Coupe	192705.5
Sedan	206962.1429
Audi	**53452.1128**
2dr Hatchback	2000
4dr SUV	48634.54545
Convertible	70029.89362
Coupe	93586.57895
Sedan	44461.78882
Wagon	33894



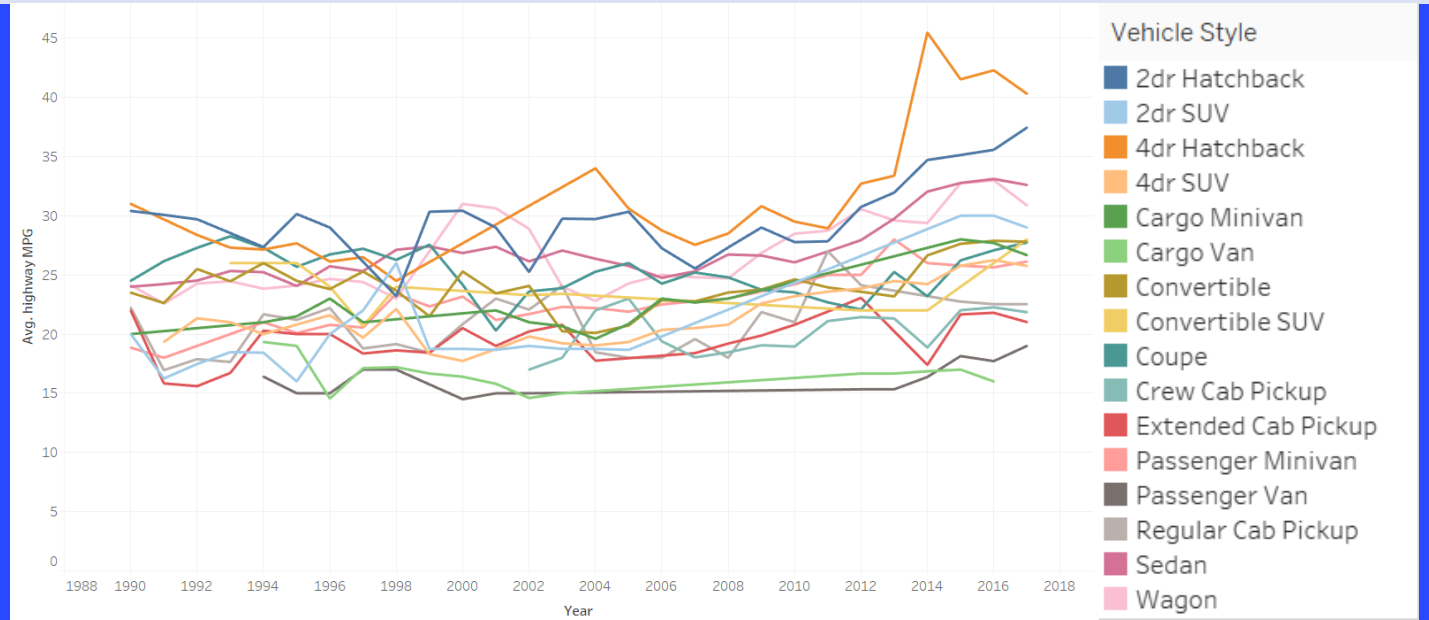
Task 3: How do the different feature such as transmission type affect the MSRP, and how does this vary by body style?

Sum of MSRP	Column Labels					
Vehicle Style	AUTOMATED_MANUAL	AUTOMATIC	DIRECT_DRIVE	MANUAL	UNKNOWN	Grand Total
2dr Hatchback	1549315	2615808	95400	4259817	14723	8535063
2dr SUV		819069		567343	9484	1395896
4dr Hatchback	1579450	8389455	1213590	4556953		15739448
4dr SUV	1051730	97733282	149400	1635205		100569617
Cargo Minivan		1485390				1485390
Cargo Van		1451621				1451621
Convertible	16369647	30454162		19954482	11567	66789858
Convertible SUV		311404		193896		505300
Coupe	31239133	31812649		29280936	4000	92336718
Crew Cab Pickup		24269438		1077700		25347138
Extended Cab Pickup		11213270		2797238		14010508
Passenger Minivan		10617898		66080		10683978
Passenger Van		3713946				3713946
Regular Cab Pickup		4537355		1700499	16000	6253854
Sedan	9927230	98436743	1590245	9740844	2000	119697062
Wagon	575735	11956512	171250	2426803		15130300
Grand Total	**62292240**	**339818002**	**3219885**	**78257796**	**57774**	**483645697**



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

Average of highway MPG	Column Labels										
Body Style	**1990**	**1991**	**1992**	**1993**	**1994**	**1995**	**1996**	**1997**	**1998**	**1999**	**2000**
2dr Hatchback	30.4	30.06666667	29.6969697	28.53333333	27.35	30.14285714	29	26.11111111	23.2	30.33333333	30.41666667
2dr SUV	20	16.25	17.47058824	18.47368421	18.42857143	16	20	22	26	18.75	18.75
4dr Hatchback	31		28.375	27.3	27.14285714	27.66666667	26.125	26.5	24.5		
4dr SUV		19.33333333	21.33333333	21	20		21.6	19.7	22.11111111	18.3	17.73333333
Cargo Minivan	20				21	21.5	23	21			
Cargo Van					19.33333333	19	14.55555556	17.125	17.2	16.66666667	16.4
Convertible	23.5	22.625	25.5	24.46153846	26	24.5	23.8	25.28571429	23.66666667	21.5	25.28571429
Convertible SUV				26	26	26	24	20.66666667	24		
Coupe	24.5	26.15789474	27.28571429	28.25925926	27.29166667	25.67741935	26.72727273	27.20689655	26.26666667	27.55555556	24.16666667
Crew Cab Pickup											
Extended Cab Pickup	22	15.83333333	15.6	16.71428571	20.28571429	20	20	18.35714286	18.625	18.42307692	20.5
Passenger Minivan	18.85714286	18			21	20.08333333	20.77777778	20.55555556	23.4	22.33333333	23.16666667
Passenger Van					16.4	15	15	17	17		14.5
Regular Cab Pickup	22.23076923	16.95238095	17.88235294	17.64705882	21.66666667	21.2	22.2	18.78571429	19.15151515	18.42857143	20.83333333
Sedan	24	24.2195122	24.52083333	25.32758621	25.22727273	24.06451613	25.72727273	25.31818182	27.12	27.40540541	26.84444444
Wagon	24.13333333	22.57142857	24.26666667	24.46153846	23.83333333	24.1	24.66666667	24.4	23		31
Grand Total	**23.07317073**	**22.15131579**	**24.05084746**	**24.21052632**	**23.86503067**	**23.22962963**	**23.72519084**	**22.30857143**	**21.85064935**	**22.975**	**24.04237288**



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

Brand	Average of Engine HP	Average of highway MPG	Average of MSRP
Acura	244.797619	28.11111111	34887.5873
Alfa Romeo	237	34	61600
Aston Martin	484.3225806	18.89247312	197910.3763
Audi	277.695122	28.82317073	53452.1128
Bentley	533.8513514	18.90540541	247169.3243
BMW	326.9071856	29.24550898	61546.76347
Bugatti	1001	14	1757223.667
Buick	219.244898	26.94897959	28206.61224
Cadillac	332.3098237	25.23677582	56231.31738
Chevrolet	246.9722471	25.81567231	28350.38557
Chrysler	229.1390374	26.36898396	26722.96257
Dodge	244.4153355	22.34504792	22390.05911
Ferrari	511.9565217	15.72463768	238218.8406
FIAT	143.559322	37.33870968	22670.24194
Ford	243.0979263	24.00681044	27399.26674
Genesis	347.3333333	25.33333333	46616.66667
GMC	259.8446602	21.4038835	30493.29903
Honda	195.7494407	32.57461024	26674.34076
HUMMER	261.2352941	17.29411765	36464.41176
Hyundai	201.9174917	30.39273927	24597.0363
Infiniti	310.0666667	24.77878788	42394.21212

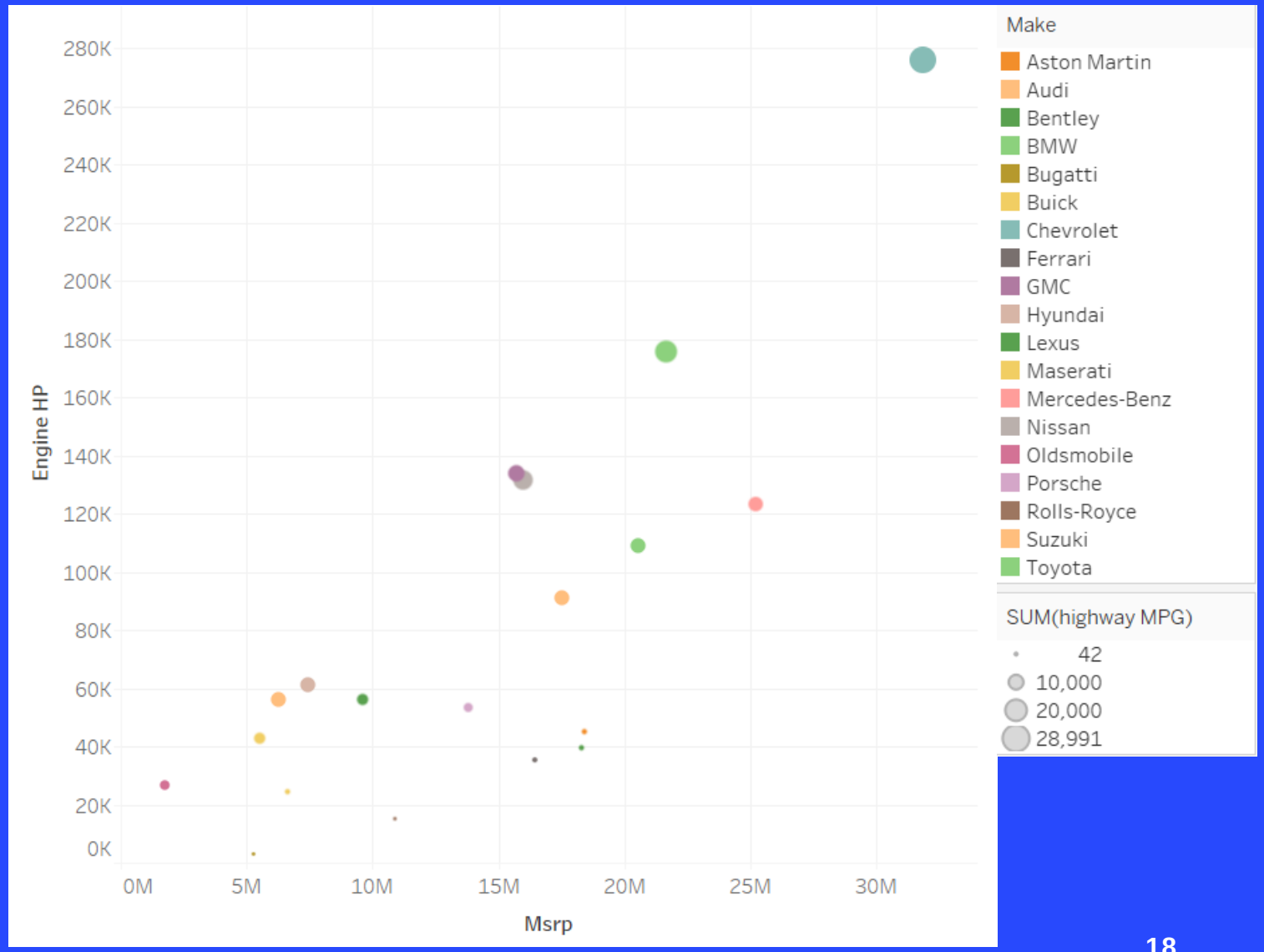
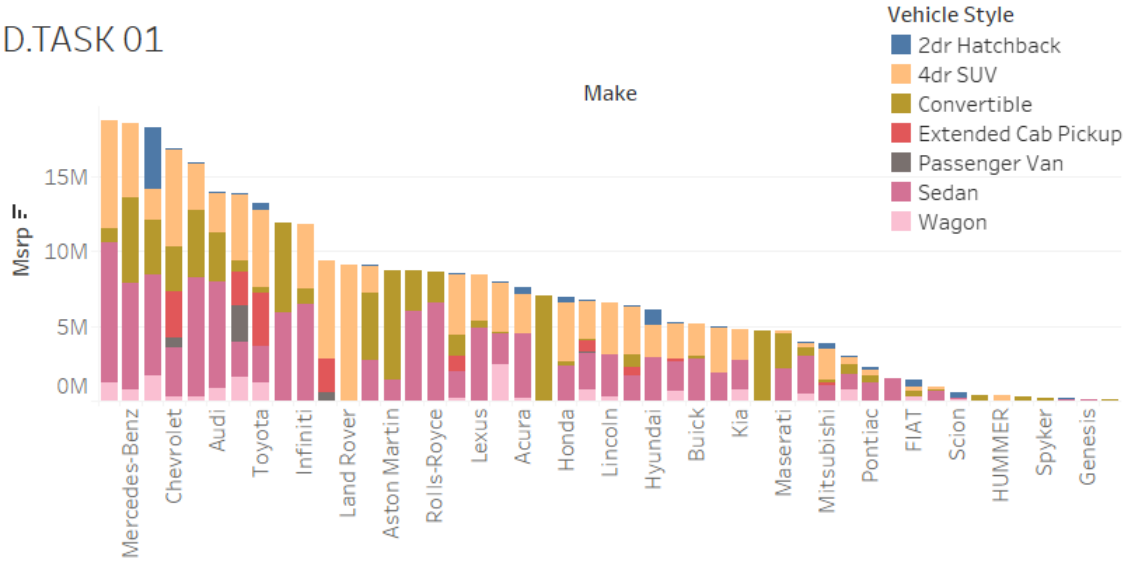
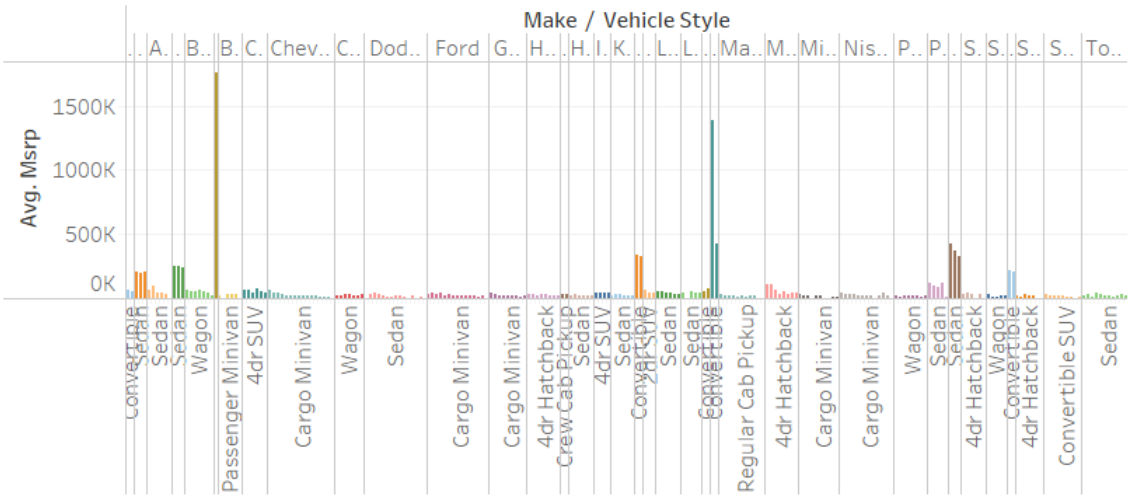


Tableau Dashboard

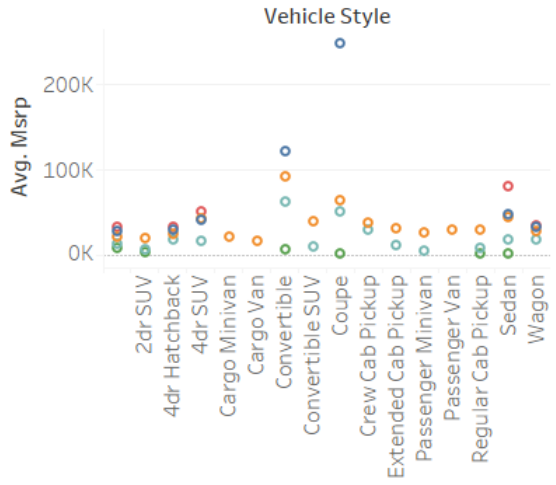
D.TASK 01



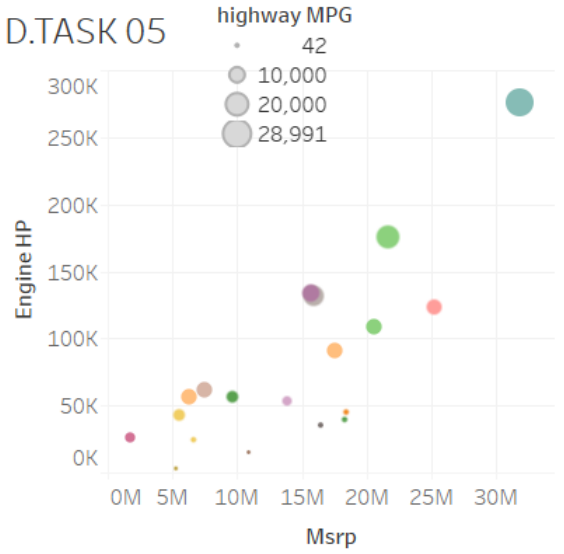
D.TASK 02



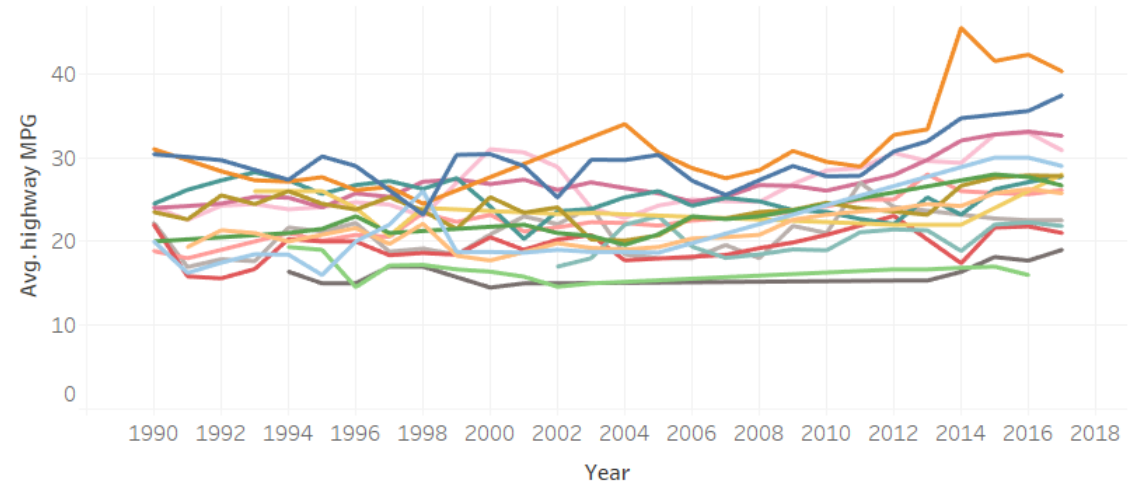
D.TASK 03



D.TASK 05



D.TASK 04



Tech-Stack Used

- Microsoft Excel: It is a spreadsheet program from Microsoft and a component of its Office product for business application. This enables users to format, calculate and organize data in a spreadsheet.
- MS Excel Functions: They are predefined formulas that perform calculations by using specific values, called arguments, in a particular order or structure. Some of the functions are:
- 1. Text function: clean(), substitute(), replace(), concatenate(), trim(), etc.
- 2. Mathematical and Statistical functions: sum(), sumif(), count(), max(), average(), median(), mode(), stdev(), correl(), etc.
- Data Visualization in Excel: Bar, Column, Scatter, Stacked Chart, horizontal chart bubble chart.
- Tableau: It is a software company that offers collaborative data visualization software for organizations working with business information analytics.

Insights

- We were able to find out the popularity of a car model varying across different market categories.
- To find out the relationship between the car’s engine power and its price.
- Which car features are most important in determining a car’s price?
- How the average price of a car vary across different manufacturers?
- What is the relationship between fuel efficiency and the number of cylinders in a car’s engine?
- Which car brands have the highest and lowest average MSRP, and how does this vary by body style?
- How does the distribution of car prices vary by brand and body style?

Results

- Remembering to adapt excel functions on specific dataset.
- These learned insights helped me understand specific business questions which were addressed by MS Excel.
- Learning about Excel Text and Statistical functions. The importance of average(), median(), mode(), text(), functions.
- We were able to build different charts for visualization for answering the business questions. Some of the charts used were bar graph, stacked Chart and heatmap.
- Achieving the ability to learn and write MS Excel functions to execute different business questions.
- Solving Company related problems using different visualization charts offered by Tableau

Thank you