

Analyzing the Impact of Car Features on Price and Profitability

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.

This problem could be approached by analyzing the relationship between a car's features, market category, and pricing, and identifying which features and categories are most popular among consumers and most profitable for the manufacturer. By using data analysis techniques such as regression analysis and market segmentation, the manufacturer could develop a pricing strategy that balances consumer demand with profitability, and identify which product features to focus on in future product development efforts. This could help the manufacturer improve its competitiveness in the market and increase its profitability over time.

Project Overview:

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).

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.

Approach:

To analyse the given data set we need to clean the dataset and then we can perform the task given

Tasks: Analysis

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.

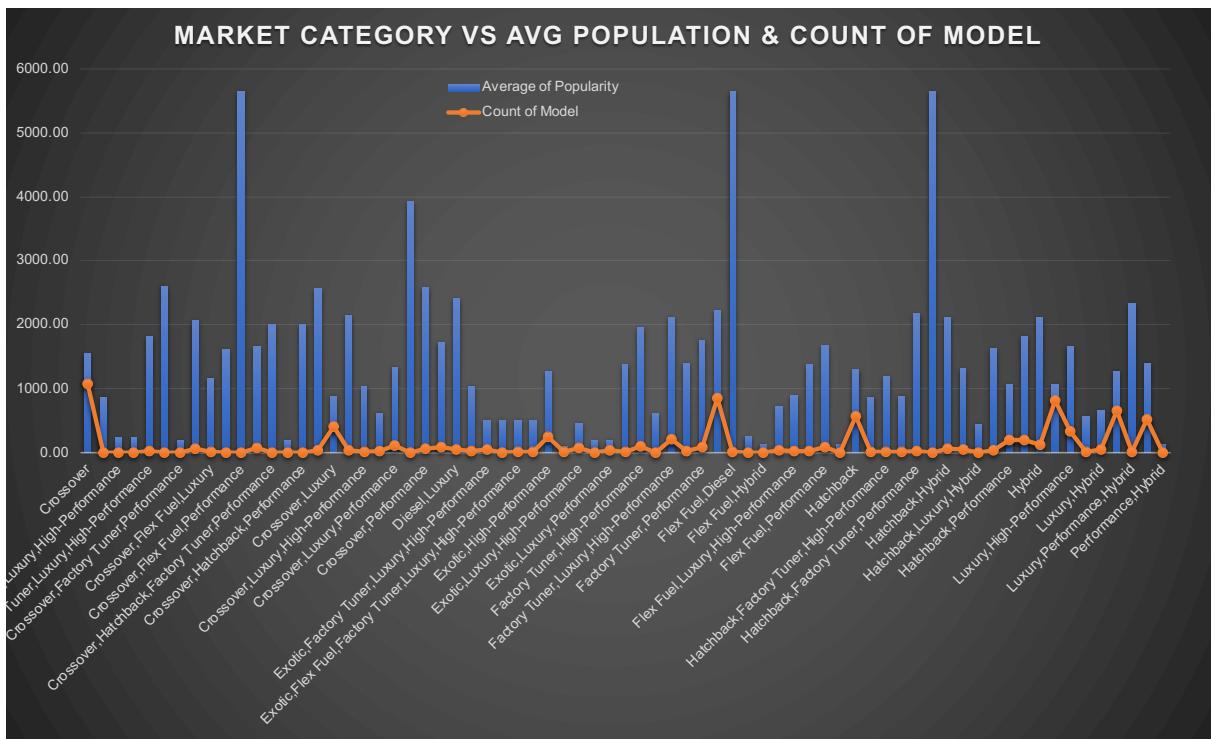
After cleaning the data we have found that there were 3736 rows that have no market category, so we have remove the rows and now we have 7823 rows having clean data.

Taking market category and population and count of models. We have created a pivot table:

Market Category	Average of Popularity	Count of Model
Crossover	1556.17	1075
Crossover,Diesel	873.00	7
Crossover,Exotic,Luxury,High-Performance	238.00	1
Crossover,Exotic,Luxury,Performance	238.00	1
Crossover,Factory Tuner,Luxury,High-Performance	1823.46	26
Crossover,Factory Tuner,Luxury,Performance	2607.40	5
Crossover,Factory Tuner,Performance	210.00	4
Crossover,Flex Fuel	2073.75	64
Crossover,Flex Fuel,Luxury	1173.20	10
Crossover,Flex Fuel,Luxury,Performance	1624.00	6
Crossover,Flex Fuel,Performance	5657.00	6
Crossover,Hatchback	1675.69	72
Crossover,Hatchback,Factory Tuner,Performance	2009.00	6
Crossover,Hatchback,Luxury	204.00	7
Crossover,Hatchback,Performance	2009.00	6
Crossover,Hybrid	2563.38	42
Crossover,Luxury	889.21	406
Crossover,Luxury,Diesel	2149.41	34
Crossover,Luxury,High-Performance	1037.22	9
Crossover,Luxury,Hybrid	630.92	24
Crossover,Luxury,Performance	1349.09	112
Crossover,Luxury,Performance,Hybrid	3916.00	2
Crossover,Performance	2585.96	69
Diesel	1730.90	84
Diesel,Luxury	2416.11	47
Exotic,Factory Tuner,High-Performance	1046.38	21
Exotic,Factory Tuner,Luxury,High-Performance	523.02	51
Exotic,Factory Tuner,Luxury,Performance	520.00	3
Exotic,Flex Fuel,Factory Tuner,Luxury,High-Performance	520.00	13
Exotic,Flex Fuel,Luxury,High-Performance	520.00	11
Exotic,High-Performance	1280.05	254
Exotic,Luxury	112.67	12
Exotic,Luxury,High-Performance	473.03	77
Exotic,Luxury,High-Performance,Hybrid	204.00	1
Exotic,Luxury,Performance	217.03	36
Exotic,Performance	1391.00	10
Factory Tuner,High-Performance	1966.44	104
Factory Tuner,Luxury	617.00	2
Factory Tuner,Luxury,High-Performance	2133.37	215
Factory Tuner,Luxury,Performance	1413.42	31
Factory Tuner,Performance	1774.05	84
Flex Fuel	2225.71	855
Flex Fuel,Diesel	5657.00	16
Flex Fuel,Factory Tuner,Luxury,High-Performance	258.00	1
Flex Fuel,Hybrid	155.00	2

Flex Fuel,Luxury	746.54	39
Flex Fuel,Luxury,High-Performance	898.31	32
Flex Fuel,Luxury,Performance	1380.07	28
Flex Fuel,Performance	1680.47	87
Flex Fuel,Performance,Hybrid	155.00	2
Hatchback	1308.65	574
Hatchback,Diesel	873.00	14
Hatchback,Factory Tuner,High-Performance	1205.15	13
Hatchback,Factory Tuner,Luxury,Performance	886.89	9
Hatchback,Factory Tuner,Performance	2173.71	21
Hatchback,Flex Fuel	5657.00	7
Hatchback,Hybrid	2111.16	64
Hatchback,Luxury	1323.13	45
Hatchback,Luxury,Hybrid	454.00	3
Hatchback,Luxury,Performance	1632.25	36
Hatchback,Performance	1073.66	198
High-Performance	1823.38	198
Hybrid	2116.59	121
Luxury	1079.21	819
Luxury,High-Performance	1668.02	334
Luxury,High-Performance,Hybrid	568.83	12
Luxury,Hybrid	673.63	52
Luxury,Performance	1293.06	659
Luxury,Performance,Hybrid	2333.18	11
Performance	1415.21	520
Performance,Hybrid	155.00	1
Grand Total	1512.59	7823

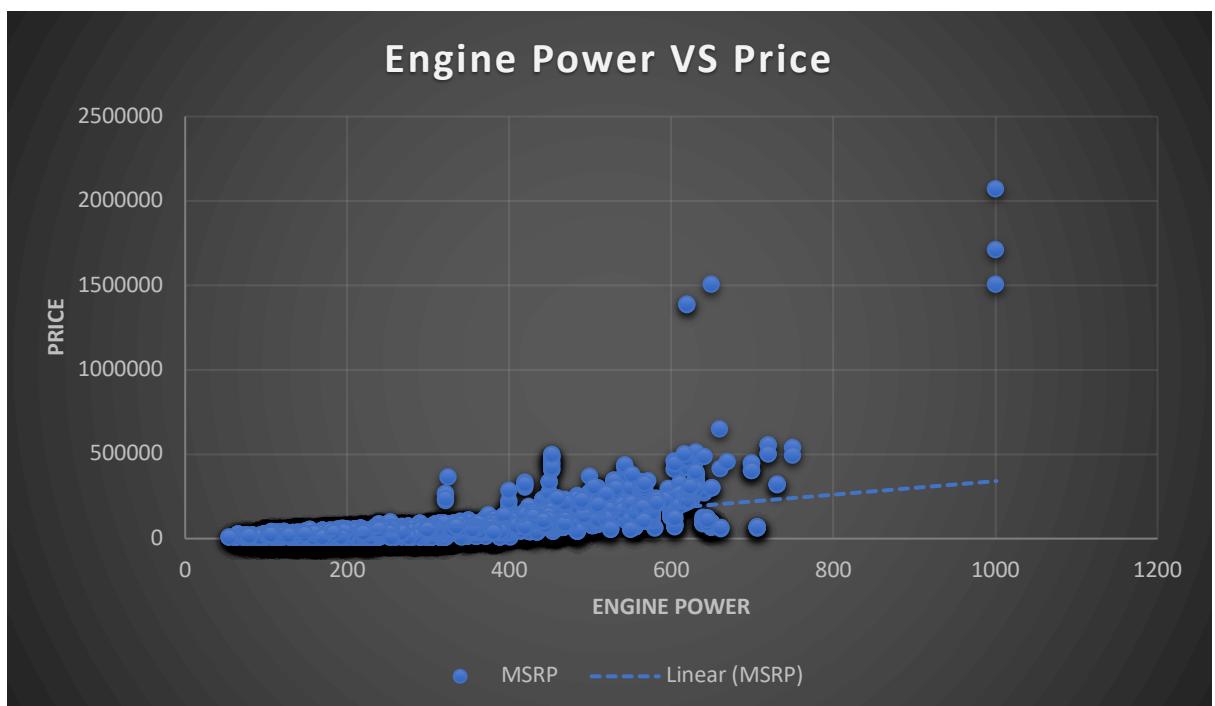
- **Task 1.B:** Create a combo chart that visualizes the relationship between market category and popularity.
So taking the pivot table in consideration we have created the chart that shows the average popularity and count model of each market category.



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.

Ans: So from the data we will take engine power and price for the plot.



Here you can see that as the engine power increases the price also increases.

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.

Ans: To use the regression analysis we will separate the factors that affects the price from the data set.

We will use data analysis tool of excel to perform regression analysis.

There are two terms in regression analysis:

Dependent variables: which are dependent on others.

Independent variable: Independent of other variables.

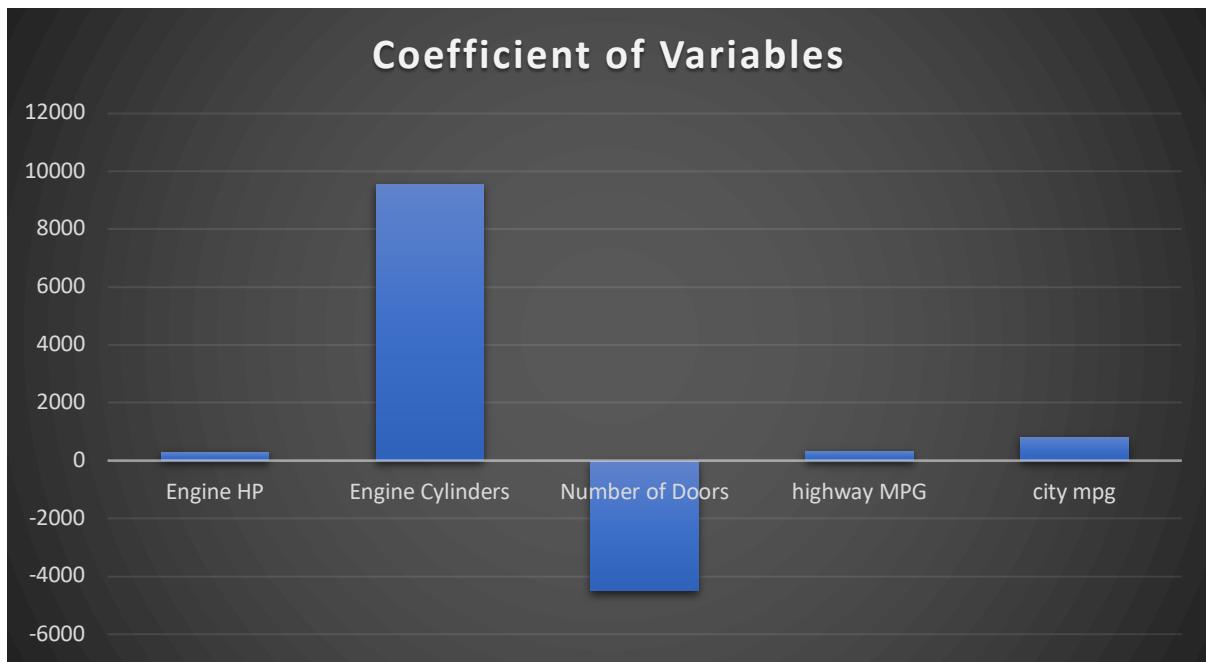
Keeping price as dependent variable and others as independent variable, we will perform regression analysis.

SUMMARY OUTPUT	
<i>Regression Statistics</i>	
Multiple R	0.674362061
R Square	0.454764189
Adjusted R Square	0.454415171
Standard Error	52592.742
Observations	7817

ANOVA					
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression	5	1.80202E+13	3.60404E+12	1302.982311	0
Residual	781 1	2.16052E+13	2765996511		
Total	781 6	3.96254E+13			

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>	<i>Lower 95.0%</i>	<i>Upper 95.0%</i>
Intercept	-100359.2592	4731.036189	-21.21295529	3.83317E-97	-109633.3568	-91085.16155	-109633.3568	-91085.16155
Engine HP	307.4500058	8.405758945	36.57611499	1.6929E-270	290.9724677	323.9275439	290.9724677	323.9275439
Engine Cylinders	9554.811755	567.9670627	16.82282721	2.06565E-62	8441.444244	10668.17926	8441.444244	10668.17926
Number of Doors	-4497.564278	683.1603885	-6.583467592	4.89312E-11	-5836.741549	-3158.387007	-5836.741549	-3158.387007
highway MPG	354.9214507	135.3370704	2.622499879	0.008745674	89.62455767	620.2183437	89.62455767	620.2183437
city mpg	836.5851653	125.1264491	6.685917894	2.45271E-11	591.3038236	1081.866507	591.3038236	1081.866507

	<i>Coefficients</i>
Intercept	-100359.2592
Engine HP	307.4500058
Engine Cylinders	9554.811755
Number of Doors	-4497.564278
highway MPG	354.9214507
city mpg	836.5851653



From the coefficient of variable relationship we can see that engine cylinder has highest relationship with the price.

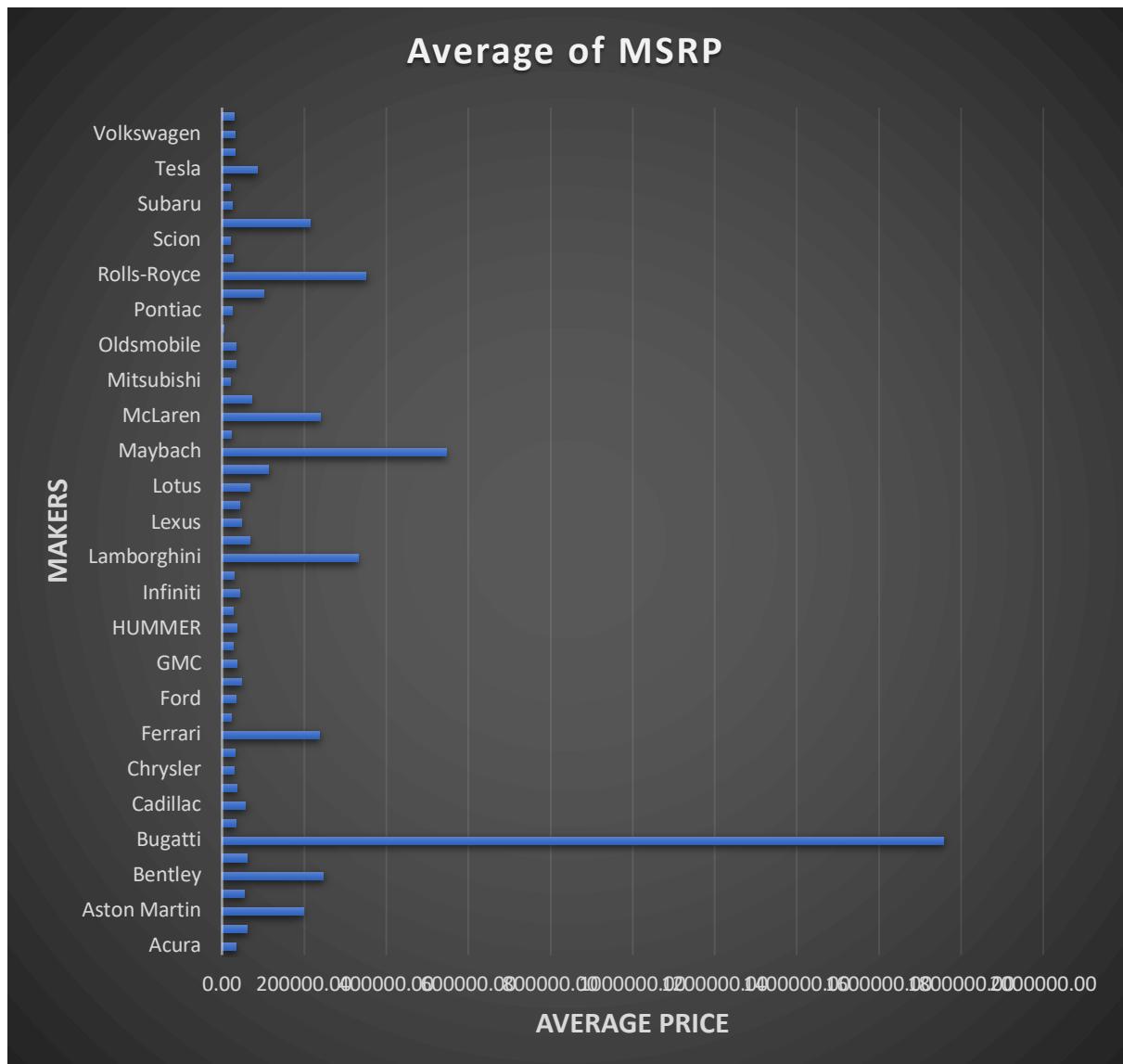
Insight Required: How does the average price of a car vary across different manufacturers?

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

Makers	Average of MSRP
Acura	35087.49
Alfa Romeo	61600.00
Aston Martin	198123.46
Audi	54574.12
Bentley	247169.32
BMW	62162.56
Bugatti	1757223.67
Buick	33770.40
Cadillac	56368.27
Chevrolet	36011.28
Chrysler	29978.87
Dodge	31186.80
Ferrari	238218.84
FIAT	23060.61
Ford	33565.55
Genesis	46616.67
GMC	37385.75
Honda	27042.11
HUMMER	36464.41
Hyundai	27274.32
Infiniti	42640.27
Kia	30149.31
Lamborghini	331567.31

Land Rover	68067.09
Lexus	47549.07
Lincoln	43860.83
Lotus	68377.14
Maserati	113684.49
Maybach	546221.88
Mazda	23503.50
McLaren	239805.00
Mercedes-Benz	72069.53
Mitsubishi	20417.26
Nissan	32893.15
Oldsmobile	34868.00
Plymouth	4189.08
Pontiac	24728.13
Porsche	101622.40
Rolls-Royce	351130.65
Saab	27879.81
Scion	20395.94
Spyker	214990.00
Subaru	25831.60
Suzuki	21203.17
Tesla	85255.56
Toyota	30944.87
Volkswagen	30922.08
Volvo	29724.68

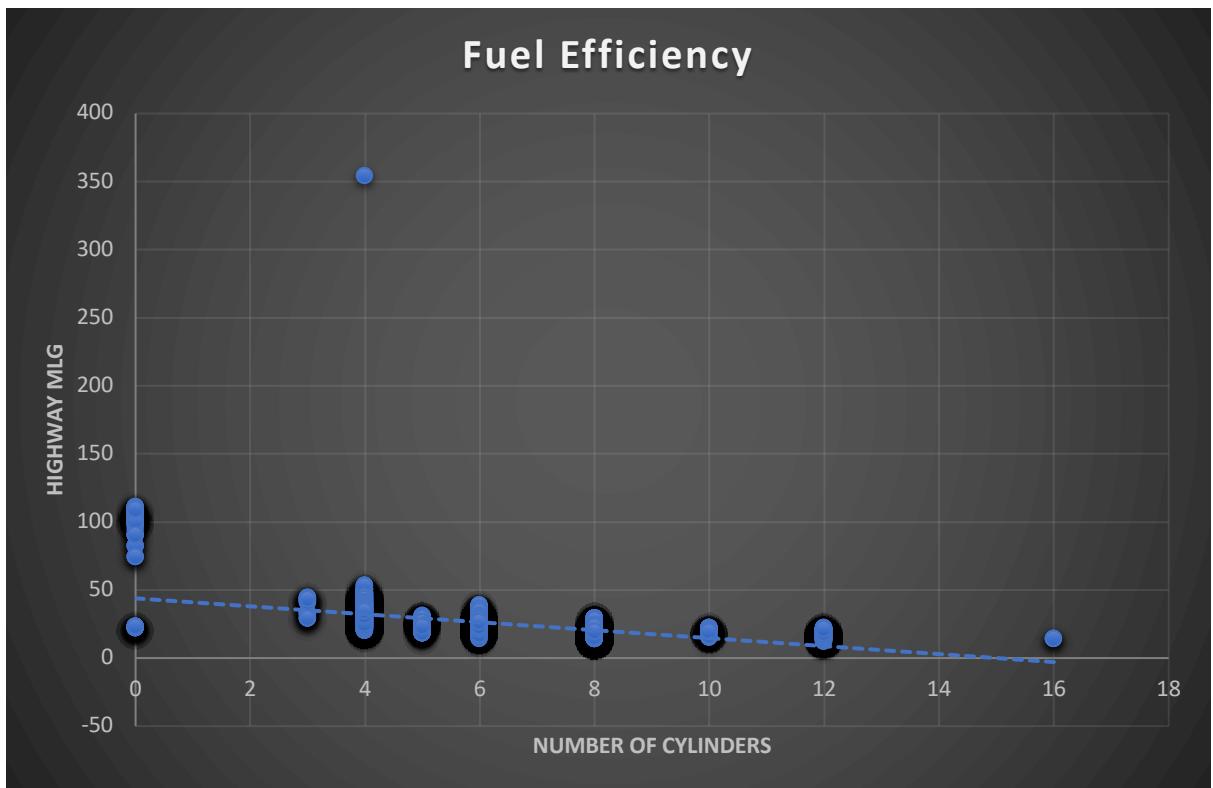
- **Task 4.B:** Create a bar chart or a horizontal stacked bar chart that visualizes the relationship between manufacturer and average price.



From the above chart we can see that Bugatti has got the highest average price followed by Maybach, Rolls-Royce and so on.

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



We can see that as number of cylinders increases the mileage decreases . Hence it has negative relationship with each other.

Correlation coefficient	-0.5956186
-------------------------	------------

Building the Dashboard:

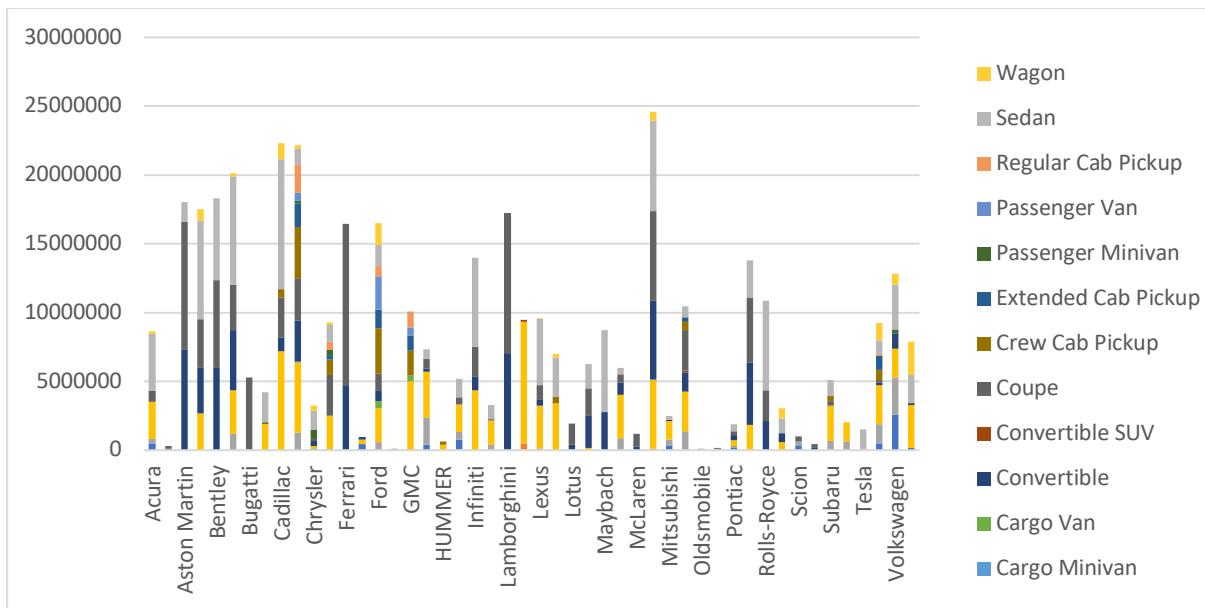
Now for the Next portion of the Project, you need to create the Interactive Dashboard. Use filters and slicers to make the chart interactive. The client has requested these questions given below:

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

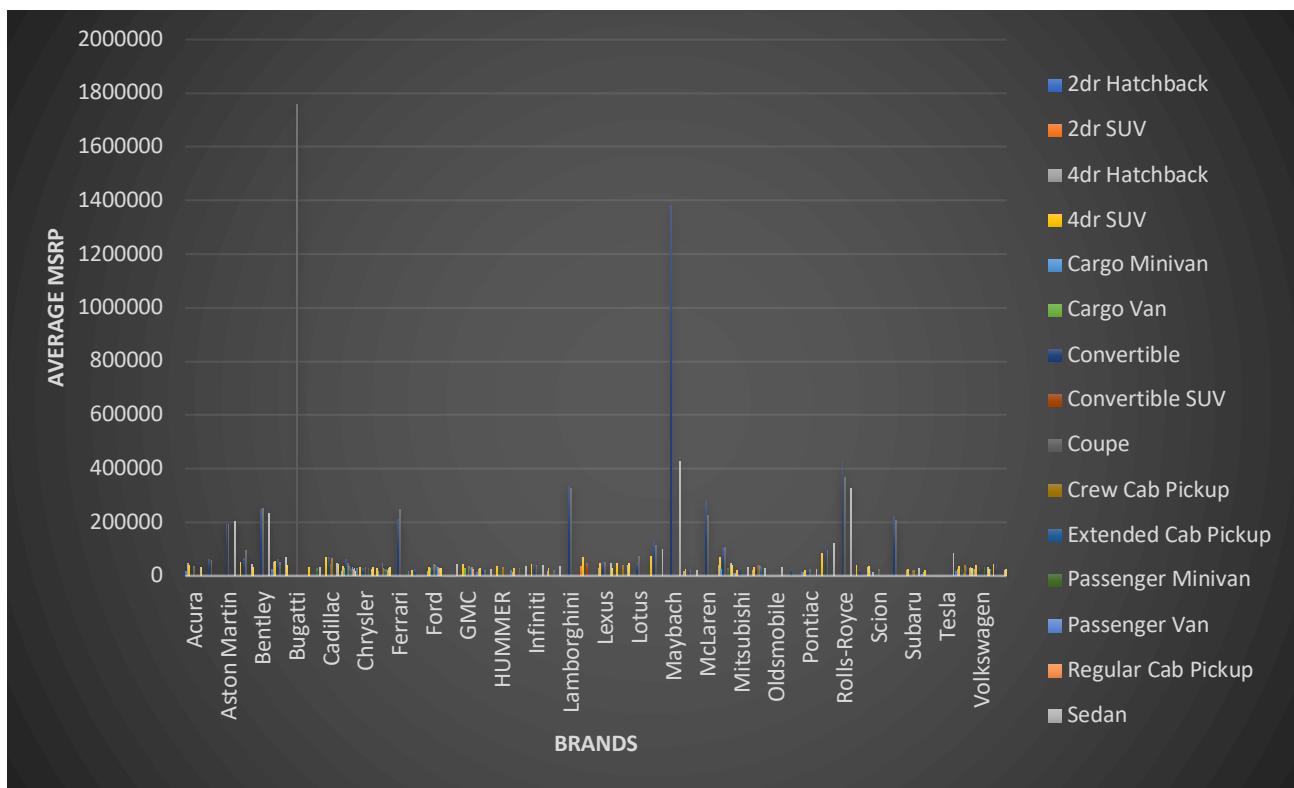
- **Hints:** Stacked column chart to show the distribution of car prices by brand and body style. Use filters and slicers to make the chart interactive. Calculate the total MSRP for each brand and body style using SUMIF or Pivot Tables.

Answer:

To get the desired output we need to create pivot table to create a stacked bar chart having slicers.

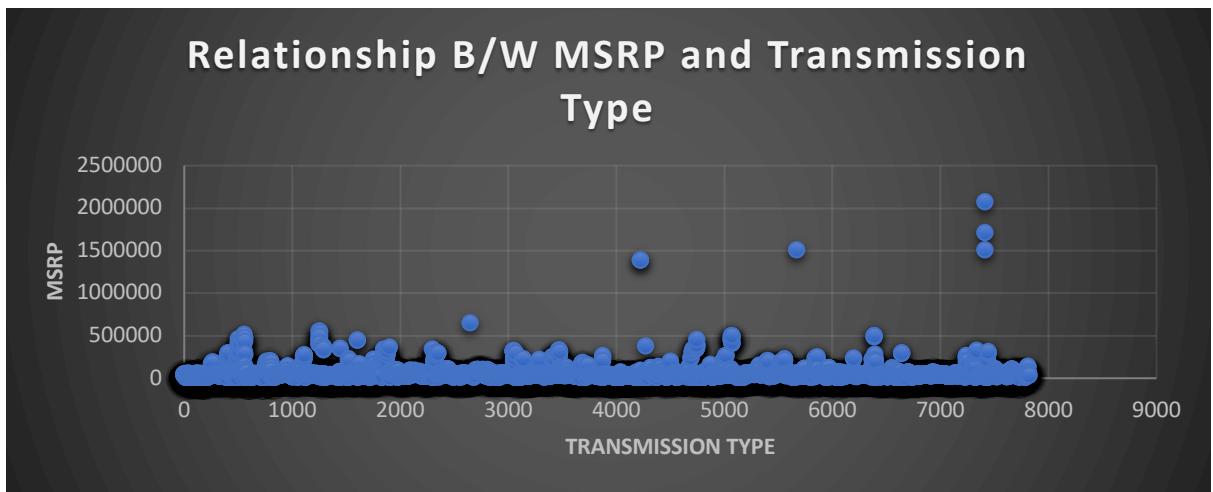


- **Task 2:** Which car brands have the highest and lowest average MSRPs, and how does this vary by body style?



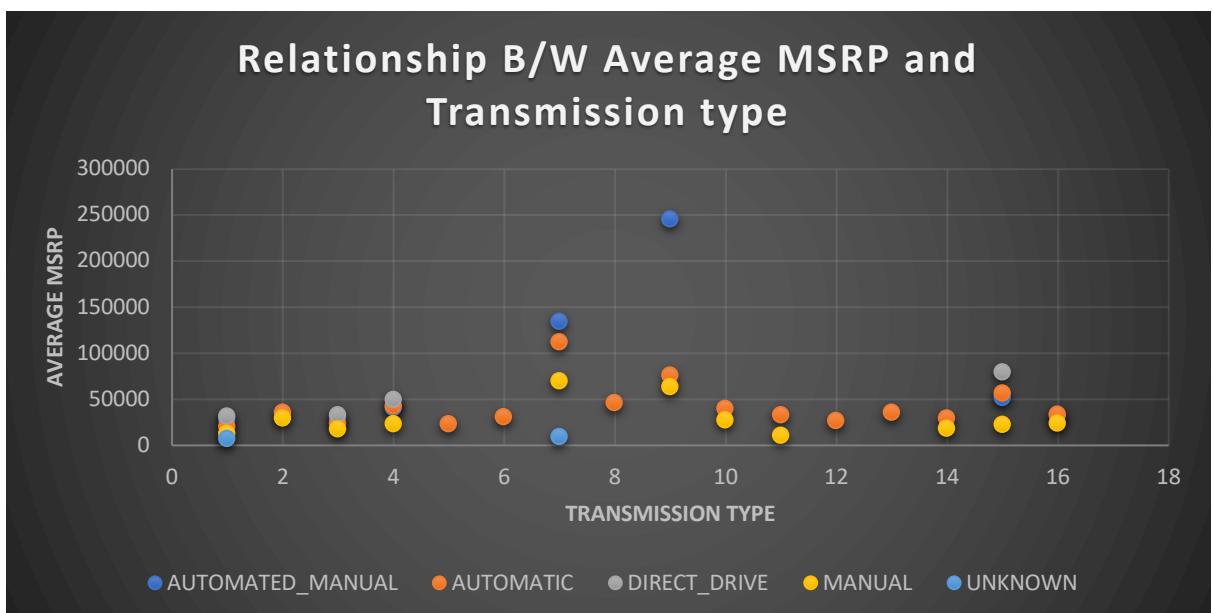
From the above chart we can conclude that Bugatti has got the highest average price in the market.

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

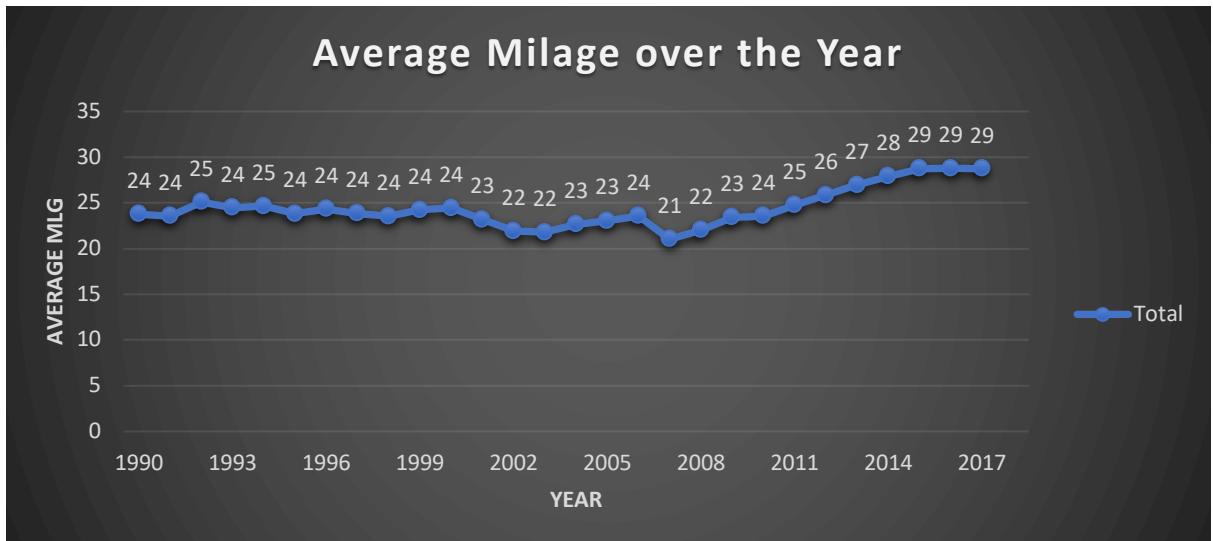


We can see that Automate_Manual(Coupe) has highest MSRP.

Average MSRP of Each combination of Transmission type and Body Style						
Body Type	AUTOMATED_MANUAL	AUTOMATIC	DIRECT_DRIVE	MANUAL	UNKNOWN	
2dr Hatchback	27470	20784	31800	12841	7362	
2dr SUV		35895		29223		
4dr Hatchback	29347	23889	32800	17500		
4dr SUV	40451	42945	49800	23131		
Cargo Minivan		22964				
Cargo Van		30725				
Convertible	134528	112168		69717	9567	
Convertible SUV		46134				
Coupe	245977	76217		63567		
Crew Cab Pickup		39607		27361		
Extended Cab Pickup		33140		10651		
Passenger Minivan		26707				
Passenger Van		35963				
Regular Cab Pickup		29210		18045		
Sedan	51531	56695	79512	22541		
Wagon	31985	33377		23704		



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



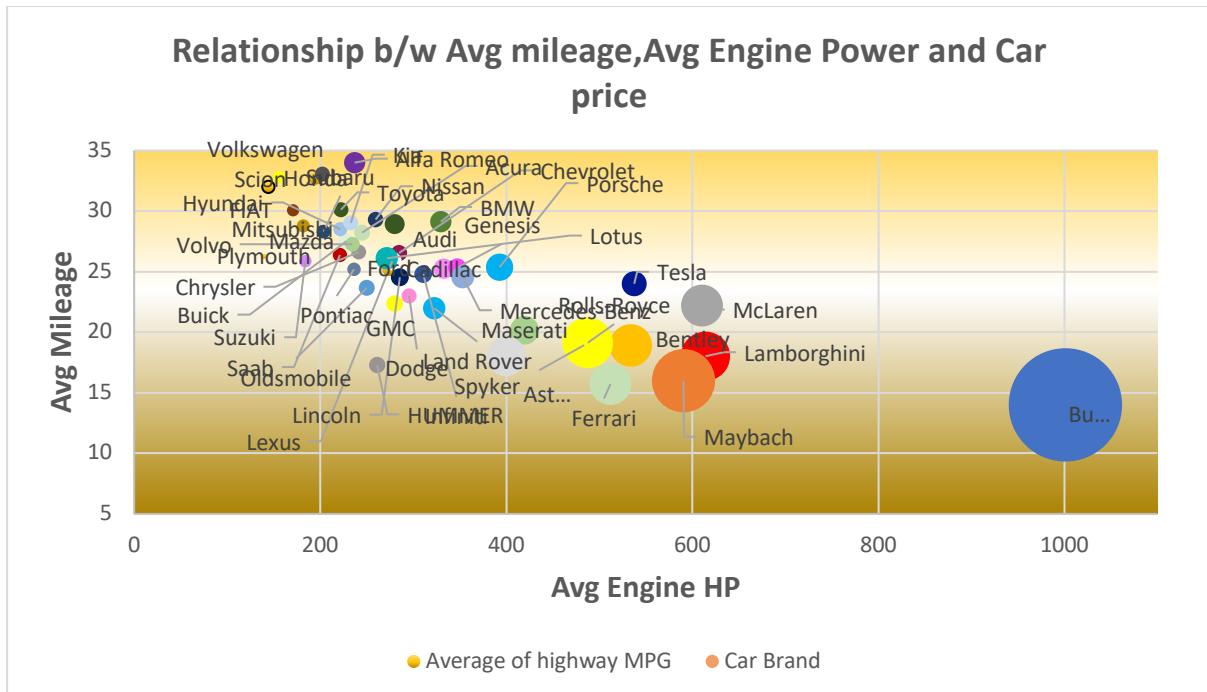
From the above chart we can see that after 2007 the mileage of the vehicles are slowly increasing.

Average of Each combination of Body Style and Year																			
Average of highway MPG	Body Style	2dr Hatchback	2dr SUV	4dr Hatchback	4dr SUV	Cargo Minivan	Cargo Van	Convertible	Convertible SUV	Coupe	Crew Cab Pickup	Extended Cab Pickup	Pasenger Minivan	Pasenger Van	Regular Cab Pickup	Sedan	Wagon	(blank)	Grand Total
Year																			
1990	30		31					24		20					22	24		24	
1991	30							22		21					16	22	22	24	
1992	29	17	28					21		22					21	22		25	
1993	28	17	28					20		22					22	22		24	
1994	27	15	27					23		24					22	22		25	
1995	29	15	28					23		23					23	24		24	
1996	29		26	14				24		23					23	25		24	
1997	26	14	27	16				24	14	24					23	24		24	
1998	23		25	18				24		24					24	23		24	
1999	30			18				22		23		21	22		18	25		24	
2000	30			16				24		24		21	23		19	25		24	
2001	29			18				21		20					25			23	
2002	25			18				21		20	15	23	23		22	23	25	22	
2003	30	19		18				20		19		25			24	25	23	22	
2004	30	19	34	18				20		19	23				18	24	23	23	
2005	30	19	31	21				21		19	23		24		18	23	23	23	
2006	27		29	21	24			22		22	21		24		18	23	23	24	
2007	25		27	21	24			22		23	18	18	23		19	21	22	21	
2008	26		28	21	23			22		22	17		23		17	24	22	22	
2009	29		31	23				23		22	19	19			19	25	27	23	
2010	27		30	23				23		22	19	19			19	25	28	24	
2011	28		29	24				24		23	21				25	28		25	
2012	30		32	24				17	23	22	21	20	25	15		27	30	26	
2013	32		33	25				17	23	22	23	21			15		29	29	
2014	35		45	25				17	26	22	23	17	17	25	16		32	29	
2015	36	30	42	26				17	27		25	22	21	25	18	23	31	29	
2016	36	30	42	27	24	16	27			26	23	22	25	18	23	31	29	29	
2017	37	29	40	26				28	28	27	22	22	28	19	23	31	31	29	
(blank)																			
Grand	31	22	38	25	24	17	25	23	24	21	21	24	18	22	29	27		27	

Average of Each Combination of Body Type and Year

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

Ans:



This bubble chart shows that how mileage changes as we increase the Engine power .

Results: The Coupe body style has highest MSRP in market.

To enable bulk of car company must make fuel efficient cars.

A car with automated and manual will be more of benefit than others.

Drive Link: <https://drive.google.com/drive/folders/1HiJHvadyDtwWhdxUNrHbq9kU4GJ2-qcP?usp=sharing>

*****THE END*****