

Analyzing the Impact of Car Features on Price and Profitability

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.

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.

Approach : To begin with I imported the given dataset into Microsoft Excel using necessary commands. I saw the task to done for given questions, then I saw table in the dataset and based on the task I did mind mapping like which columns and Excel commands to use for required task and later the same was used in Excel.

Tech Stack Used : Microsoft Excel 2016.

Here first we removed all duplicate values and removed all blank cells containing rows. Then I checked for outliers and replaced outlier with mode or median imputation.

Row Labels	Count of Transmission	
	Type	percentage
AUTOMATED_MANUAL	553	4.94%
AUTOMATIC	7932	70.83%
DIRECT_DRIVE	68	0.61%
MANUAL	2634	23.52%
UNKNOWN	12	0.11%
Grand Total	11199	100.00%

Here I replaced all UNKNOWN in transmission type column with AUTOMATIC because I have used mode imputation here i.e Automatic has high % compared to other transmission types.

Tasks: Analysis

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

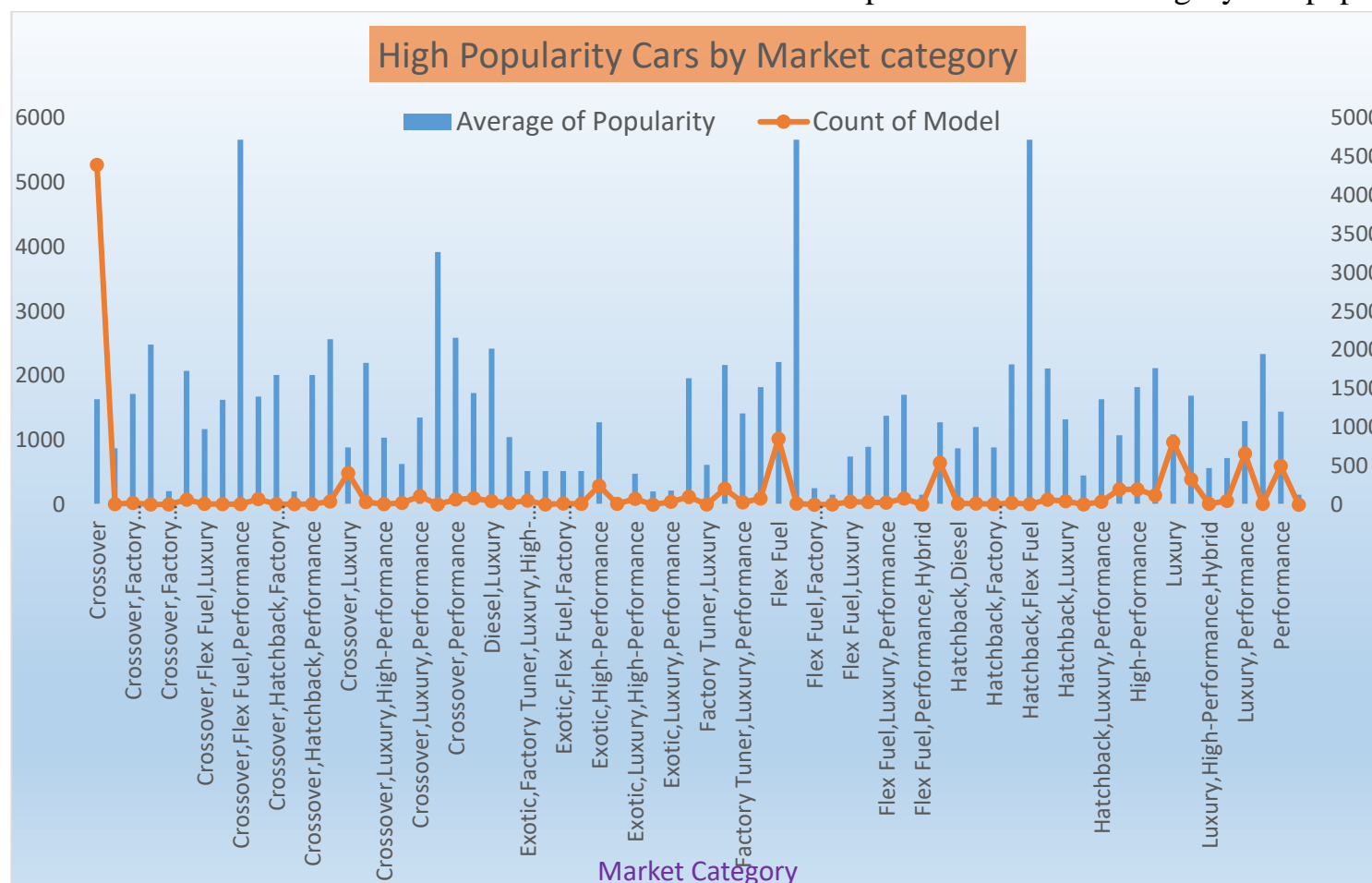
- **Task 1.A:** Create a pivot table that shows the number of car models in each market category and their corresponding popularity scores.

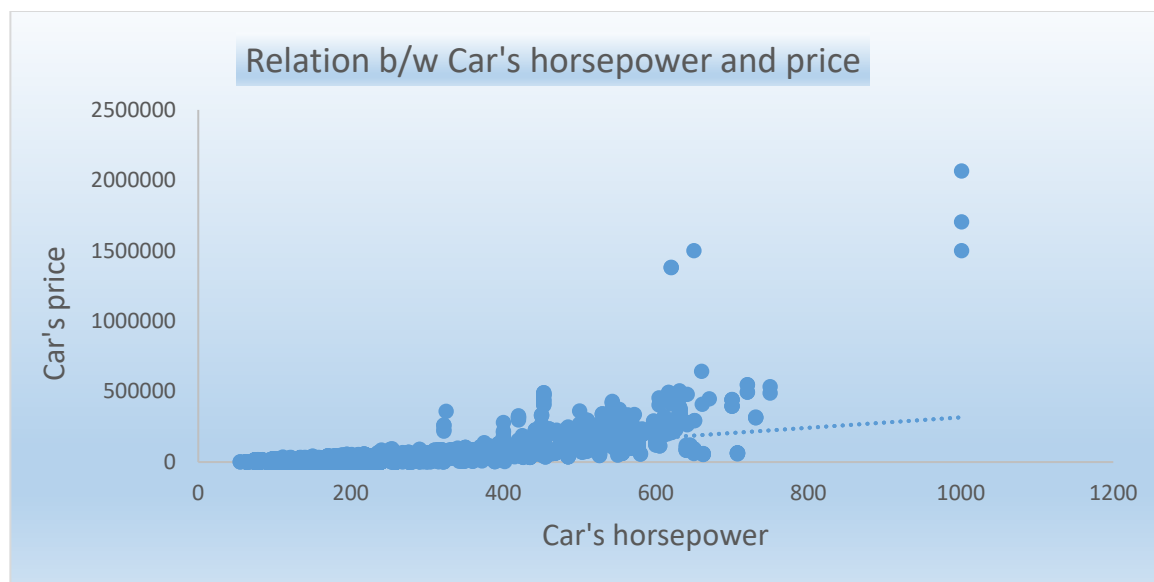
Car Model	Average of Popularity	Count of Model
Crossover	1632.159106	4387
Crossover,Diesel	873	7
Crossover,Factory Tuner,Luxury,High-Performance	1716.666667	24
Crossover,Factory Tuner,Luxury,Performance	2483	4
Crossover,Factory Tuner,Performance	210	4
Crossover,Flex Fuel	2073.75	64
Crossover,Flex Fuel,Luxury	1173.2	10
Crossover,Flex Fuel,Luxury,Performance	1624	6
Crossover,Flex Fuel,Performance	5657	6

Car Model	Average of Popularity	Count of Model
Crossover,Hatchback	1675.694444	72
Crossover,Hatchback,Factory Tuner,Performance	2009	6
Crossover,Hatchback,Luxury	204	7
Crossover,Hatchback,Performance	2009	6
Crossover,Hybrid	2563.380952	42
Crossover,Luxury	889.2142857	406
Crossover,Luxury,Diesel	2195.848485	33
Crossover,Luxury,High-Performance	1037.222222	9
Crossover,Luxury,Hybrid	630.9166667	24
Crossover,Luxury,Performance	1349.089286	112
Crossover,Luxury,Performance,Hybrid	3916	2
Crossover,Performance	2585.956522	69
Diesel	1730.904762	84
Diesel,Luxury	2416.106383	47
Exotic,Factory Tuner,High-Performance	1046.380952	21
Exotic,Factory Tuner,Luxury,High-Performance	523.0196078	51
Exotic,Factory Tuner,Luxury,Performance	520	3
Exotic,Flex Fuel,Factory Tuner,Luxury,High-Performance	520	13
Exotic,Flex Fuel,Luxury,High-Performance	520	11
Exotic,High-Performance	1277.358025	243
Exotic,Luxury	112.6666667	12
Exotic,Luxury,High-Performance	479.2933333	75
Exotic,Luxury,High-Performance,Hybrid	204	1
Exotic,Luxury,Performance	217.0277778	36
Factory Tuner,High-Performance	1961.823529	102
Factory Tuner,Luxury	617	2
Factory Tuner,Luxury,High-Performance	2164.014493	207
Factory Tuner,Luxury,Performance	1413.419355	31
Factory Tuner,Performance	1821.012821	78
Flex Fuel	2211.555294	850
Flex Fuel,Diesel	5657	16
Flex Fuel,Factory Tuner,Luxury,High-Performance	258	1
Flex Fuel,Hybrid	155	2
Flex Fuel,Luxury	746.5384615	39
Flex Fuel,Luxury,High-Performance	898.3125	32
Flex Fuel,Luxury,Performance	1380.071429	28
Flex Fuel,Performance	1702.358025	81
Flex Fuel,Performance,Hybrid	155	2
Hatchback	1278.915129	542
Hatchback,Diesel	873	14
Hatchback,Factory Tuner,High-Performance	1205.153846	13
Hatchback,Factory Tuner,Luxury,Performance	886.8888889	9
Hatchback,Factory Tuner,Performance	2173.714286	21
Hatchback,Flex Fuel	5657	7
Hatchback,Hybrid	2111.15625	64
Hatchback,Luxury	1323.133333	45
Hatchback,Luxury,Hybrid	454	3
Hatchback,Luxury,Performance	1632.25	36
Hatchback,Performance	1073.661616	198
High-Performance	1823.378788	198
Hybrid	2116.586777	121
Luxury	1090.861557	809
Luxury,High-Performance	1687.243902	328
Luxury,High-Performance,Hybrid	568.8333333	12
Luxury,Hybrid	724.6875	48
Luxury,Performance	1294.717325	658

Car Model	Average of Popularity	Count of Model
Luxury,Performance,Hybrid	2333.181818	11
Performance	1442.753012	498
Performance,Hybrid	155	1
Grand Total	1558.191385	11004

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





It is evident from the above chart as car's horse power increases, the car's price also increases linearly but when horse power goes 1000hp its price increases drastically.

Insight Required: Which car features are most important in determining a car's price?

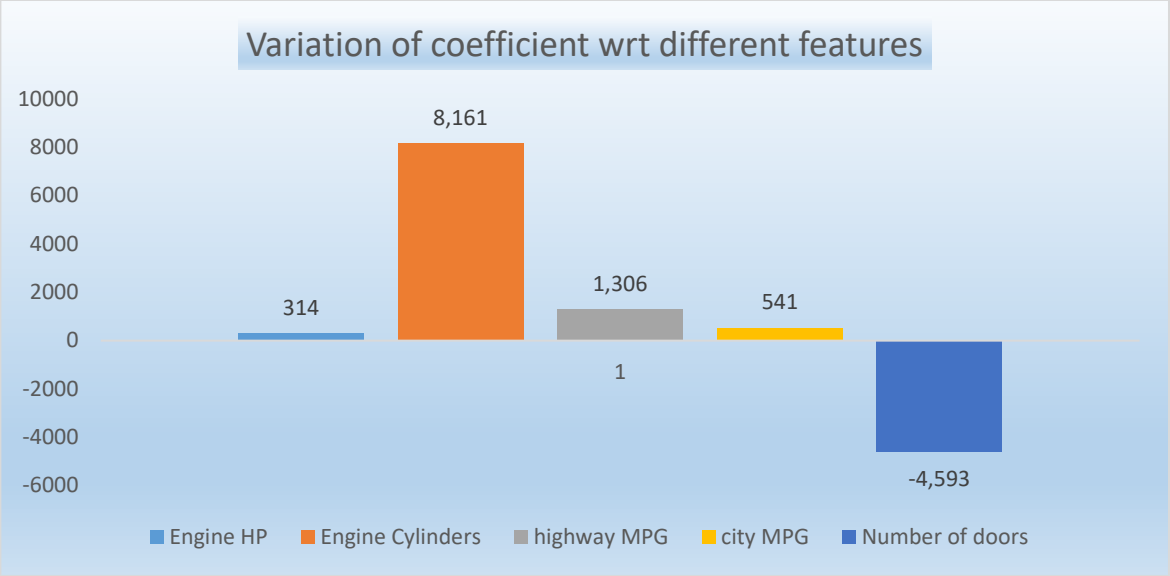
- **Task 3:** Use regression analysis to identify the variables that have the strongest relationship with a car's price. Then create a bar chart that shows the coefficient values for each variable to visualize their relative importance.

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.676822802
R Square	0.458089106
Adjusted R Square	0.457842738
Standard Error	45469.71797
Observations	11004

ANOVA					
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression	5	1.92212E+13	3.8442E+12	1859.369885	0
Residual	10998	2.27383E+13	2067495252		
Total	11003	4.19595E+13			

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>
Intercept	-113075.3242	4341.365013	-26.04603	3.5609E-145	-121585.1798	-104565
Engine HP	314.1403207	6.542880648	48.0125403	0	301.3150988	326.9655
Engine Cylinders	8160.6947	483.8428347	16.8664164	4.91093E-63	7212.275793	9109.114
highway MPG	1306.219167	176.9611069	7.38139125	1.67953E-13	959.3435962	1653.095
city MPG	540.6973765	173.0375661	3.12473984	0.001784252	201.5126508	879.8821
Number of doors	-4592.70232	502.0349188	-9.1481731	6.80994E-20	-5576.78098	-3608.62



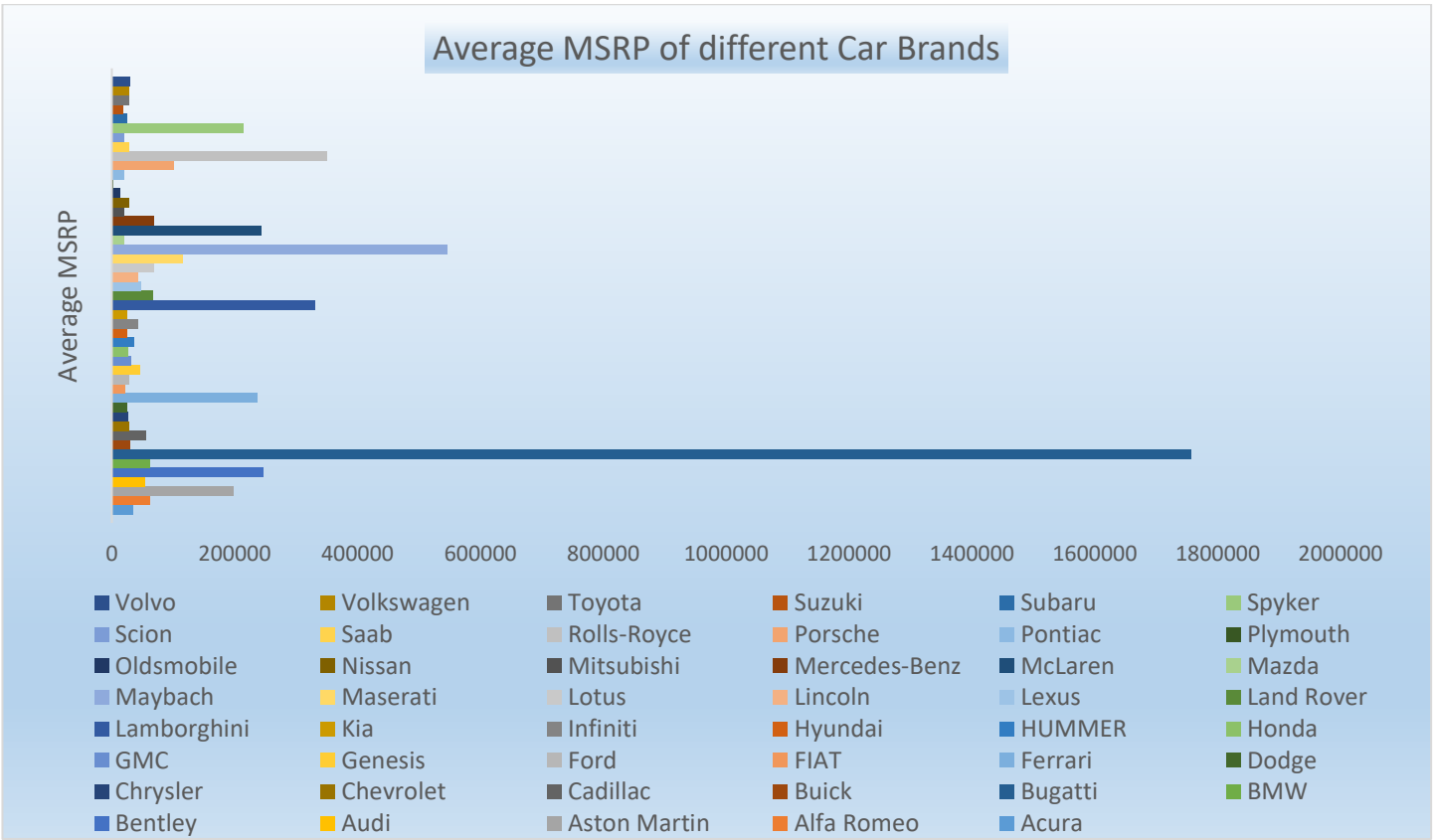
Insight Required: How does the average price of a car vary across different manufacturers?

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

Car Make	Average of MSRP
Acura	35413
Alfa Romeo	61600
Aston Martin	198123
Audi	54609
Bentley	247169
BMW	61984
Bugatti	1757224
Buick	29482
Cadillac	56368
Chevrolet	29000
Chrysler	26723
Dodge	24857
Ferrari	237384
FIAT	22191
Ford	28674
Genesis	46617
GMC	32444
Honda	26489
HUMMER	36464
Hyundai	24926
Infiniti	42640
Kia	25319
Lamborghini	331567
Land Rover	68067
Lexus	47549
Lincoln	43560
Lotus	68377
Maserati	116058
Maybach	546222

Car Make	Average of MSRP
Mazda	20007
McLaren	243542
Mercedes-Benz	68830
Mitsubishi	20747
Nissan	28856
Oldsmobile	12844
Plymouth	3330
Pontiac	19679
Porsche	101622
Rolls-Royce	351131
Saab	27880
Scion	19933
Spyker	214990
Subaru	24317
Suzuki	18042
Toyota	28759
Volkswagen	28947
Volvo	29725
Grand Total	41826

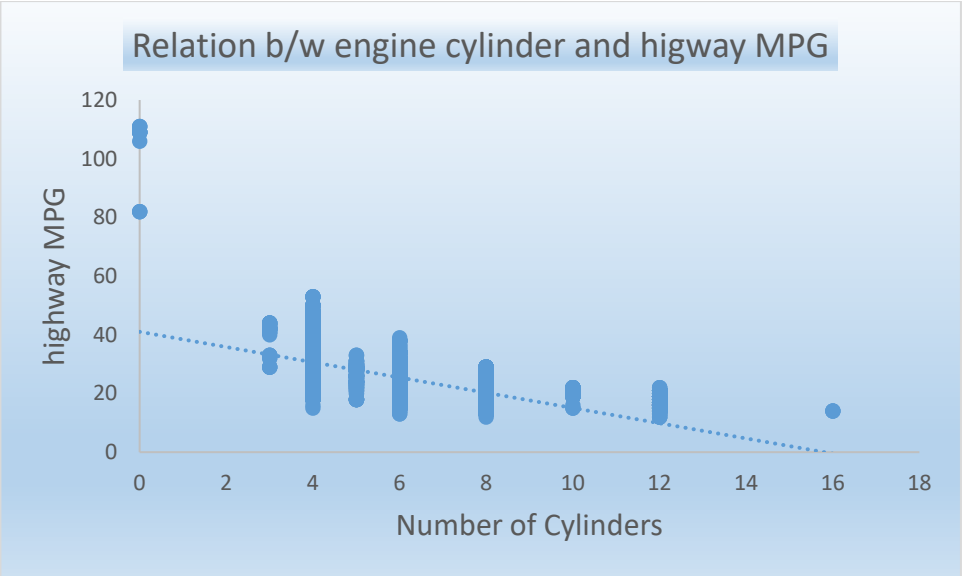
- **Task 4.B:** Create a bar chart or a horizontal stacked bar chart that visualizes the relationship between manufacturer and average price.



Top 3 highest average price car brands dominated in the above chart are

1. Bugatti
2. Maybach
3. Rolls-Royce

- 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.



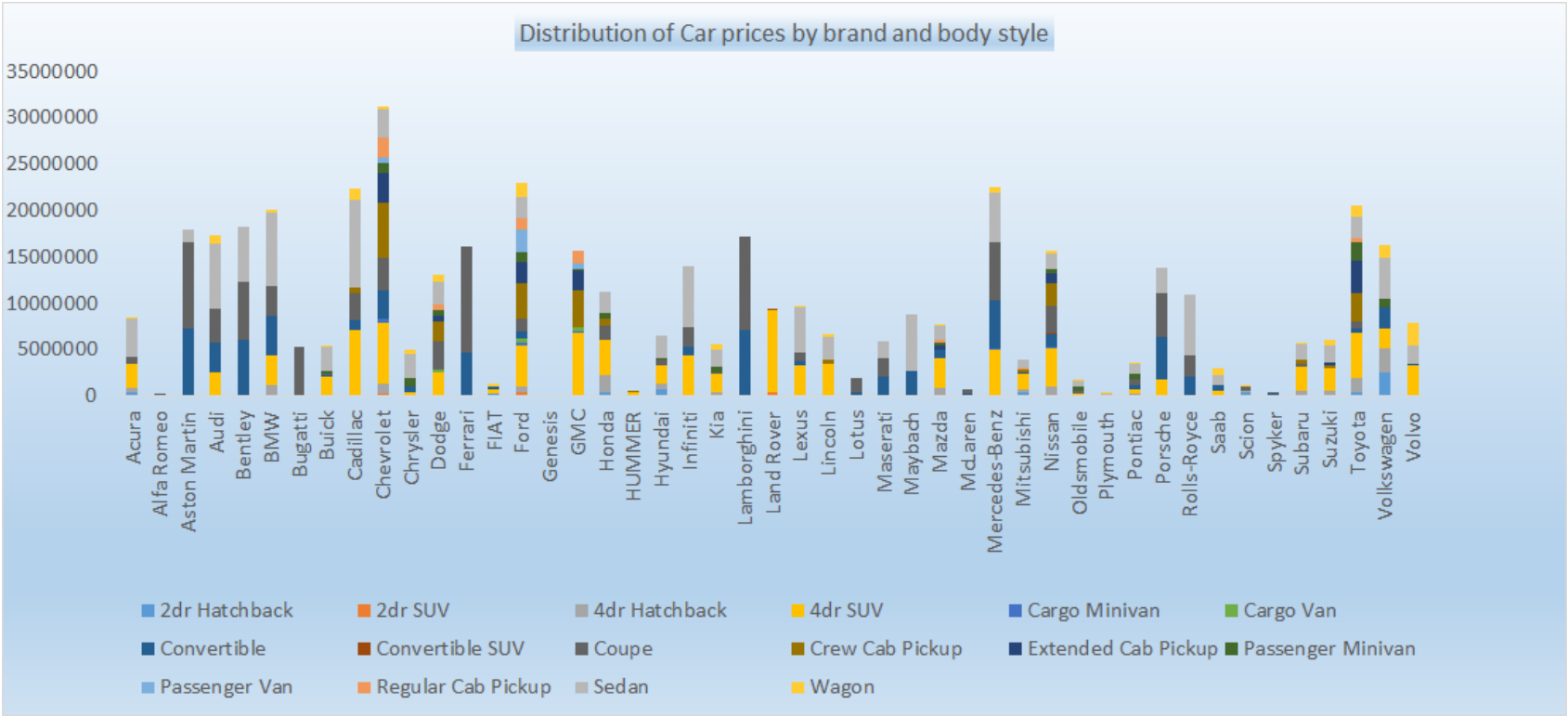
It is clear from the chart as the cylinder count in increases,the highway mileage decreases. 4 cylinder is ideal cylinder count to get better mileage.

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

The correlation coefficient between the number of cylinders and highway MPG is **-0.67277** which is a negative correlation that means as cylinder count increases, the highway mileage decreases.

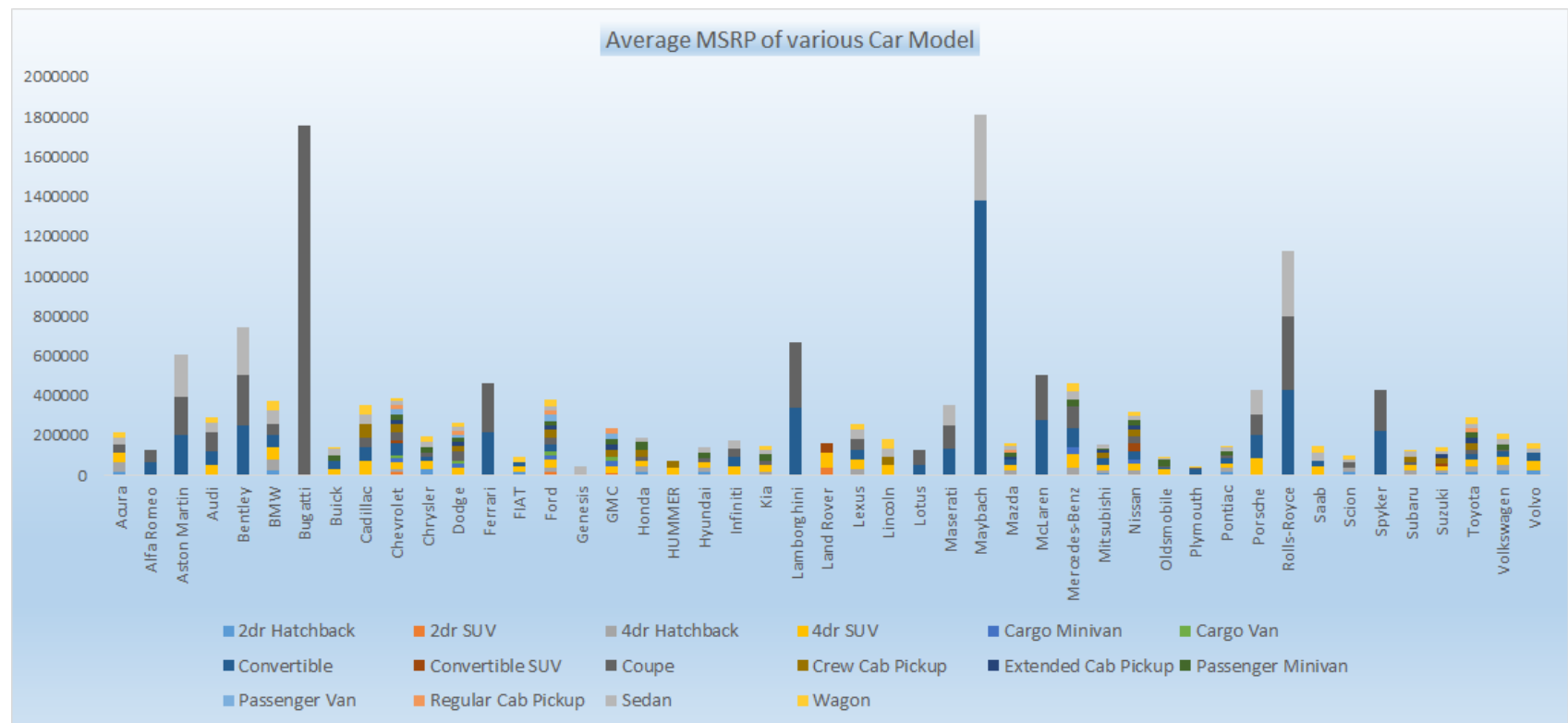
Building the Dashboard:

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



Clearly from the above chart Chevrolet high price distribution .

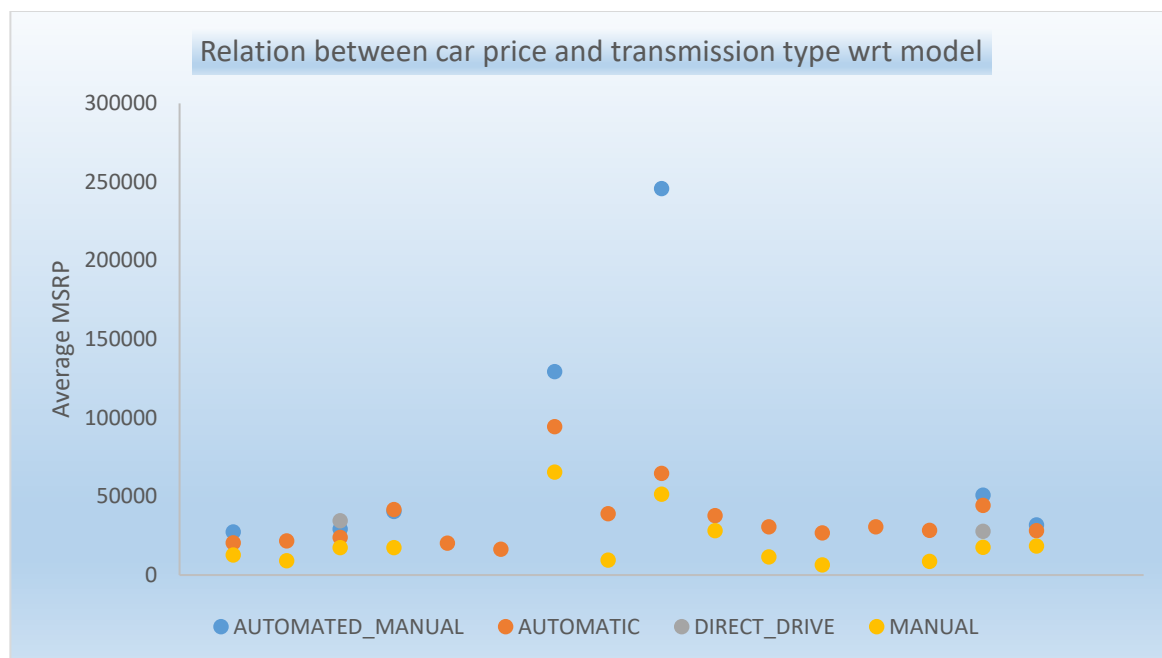
Task 2: Which car brands have the highest and lowest average MSRPs, and how does this vary by body style?



Clearly the Bugatti Coupe model car highest average price and Plymouth car lowest average price.

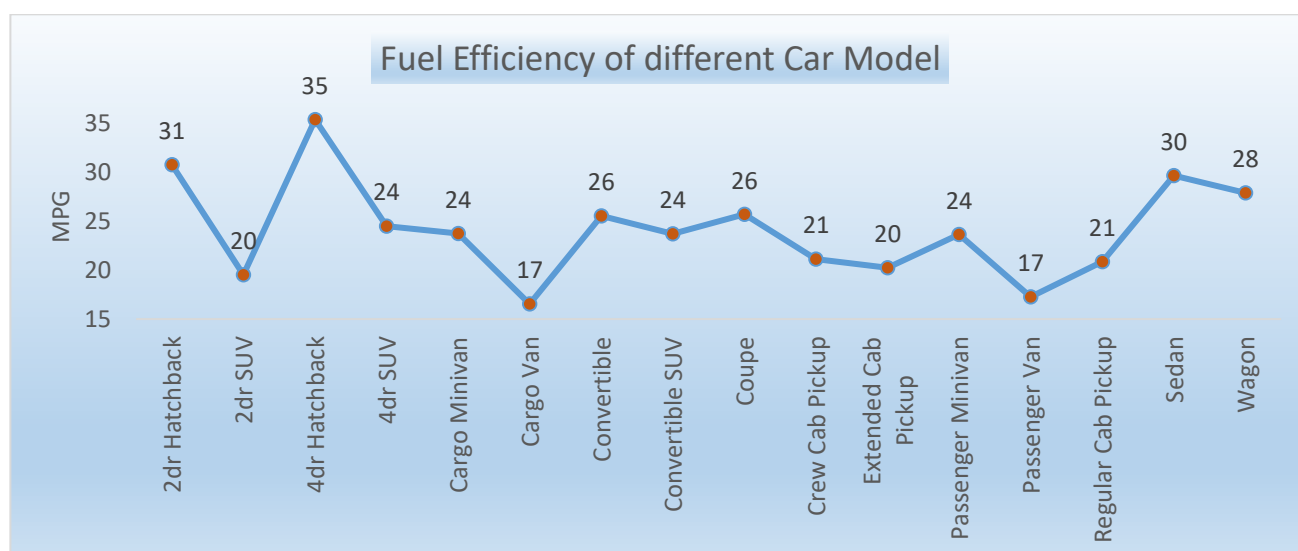
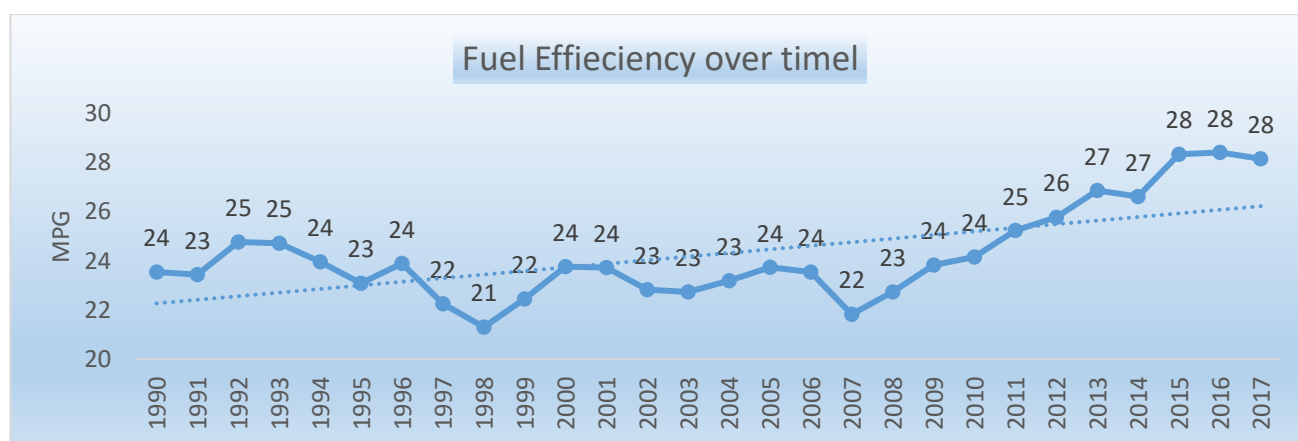
Task 3: How do the different feature such as transmission type affect the MSRP, and how does this vary by body style?

Car Model	AUTOMATED_MANUAL	AUTOMATIC	DIRECT_DRIVE	MANUAL
2dr Hatchback	27470	20523		12819
2dr SUV		21799		9173
4dr Hatchback	29347	23930	34512	17486
4dr SUV	40451	41644		17422
Cargo Minivan		20294		
Cargo Van		16372		
Convertible	129468	94306		65536
Convertible SUV		38926		9595
Coupe	245772	64698		51525
Crew Cab Pickup		37719		28233
Extended Cab Pickup		30711		11553
Passenger Minivan		26740		6510
Passenger Van		30578		
Regular Cab Pickup		28371		8759
Sedan	50775	44365	27823	17731
Wagon	31985	28219		18419

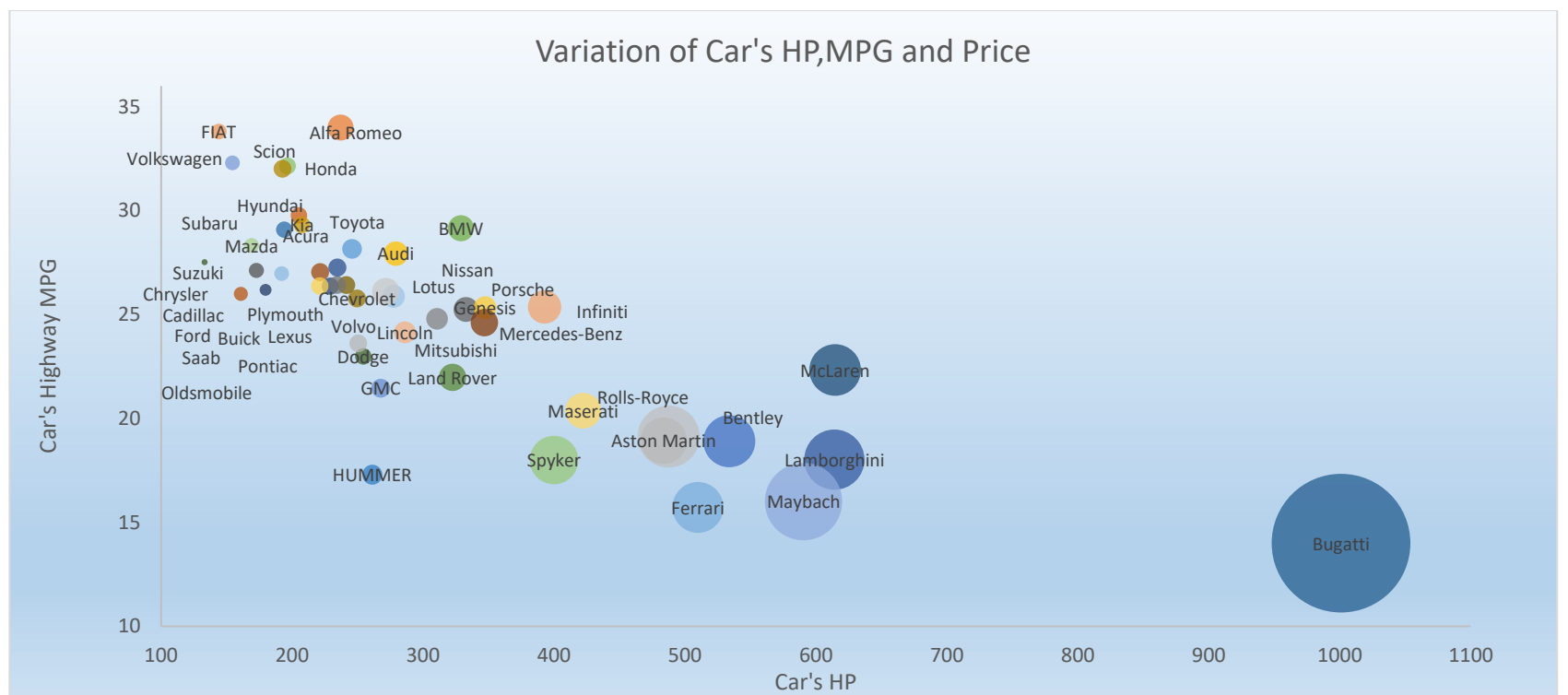


It is evident from chart that the price of automated_manual cars are highest then followed by automatic then direct drive and at last manual. Here in the above chart Automated_Manual with coupe model has the highest average price.

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



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



Clearly we can see that Bugatti has the large area of bubble which means high price of car. It also shows Bugatti has highest horsepower and lowest mileage. As power increases mileage decreases.

Excel worksheet analysis link:

https://docs.google.com/spreadsheets/d/1tkjB6Jxi3pZcmOL6kSk_JI-7F7-zTP-r/edit?usp=sharing&oid=113123883243133768755&rtpof=true&sd=true