Project Description

Over the last few decades, the automobile industry has experienced significant change, with an increasing emphasis on technical innovation, environmental sustainability, and fuel efficiency. Understanding the elements that drive customer demand for vehicles is now more crucial than ever due to rising manufacturer rivalry and a changing consumer landscape.

Electric and hybrid vehicles have become more popular in recent years, and interest in alternative fuels like hydrogen and natural gas has also grown. However, traditional gasoline-powered vehicles continue to dominate the market, and buyers have access to a variety of fuel kinds and grades.

The client has requested information on the dataset for which How can a vehicle manufacturer balance satisfying consumer demand with maximising profitability in pricing and product development decisions?

This issue might be solved by determining which features and market segments are the most well-liked by consumers and the most lucrative for the manufacturer by looking at the link between a car's characteristics, market category, and pricing. The company might create a pricing strategy that combines customer demand with profitability and select which product characteristics to concentrate on in subsequent product development efforts by employing data analysis techniques like regression analysis and market segmentation. In the long run, this may help the manufacturing become more profitable and raise its marketability.

To get deeper understanding For this, I'll conduct some data analysis. These inquiries are made in an effort to address business issues:

1. How does the popularity of a car model vary across different market categories?
2. What is the relationship between a car's engine power and its price?
3. Which car features are most important in determining a car's price?
4. How does the average price of a car vary across different manufacturers?
5. What is the relationship between fuel efficiency and the number of cylinders in a car's engine?
6. How does the distribution of car prices vary by brand and body style?
7. Which car brands have the highest and lowest average MSRPs, and how does this vary by body style?
8. How do the different feature such as transmission type affect the MSRP, and how does this vary by body style?
9. How does the fuel efficiency of cars vary across different body styles and model years?
10. How does the car's horsepower, MPG, and price vary across different Brands?

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.

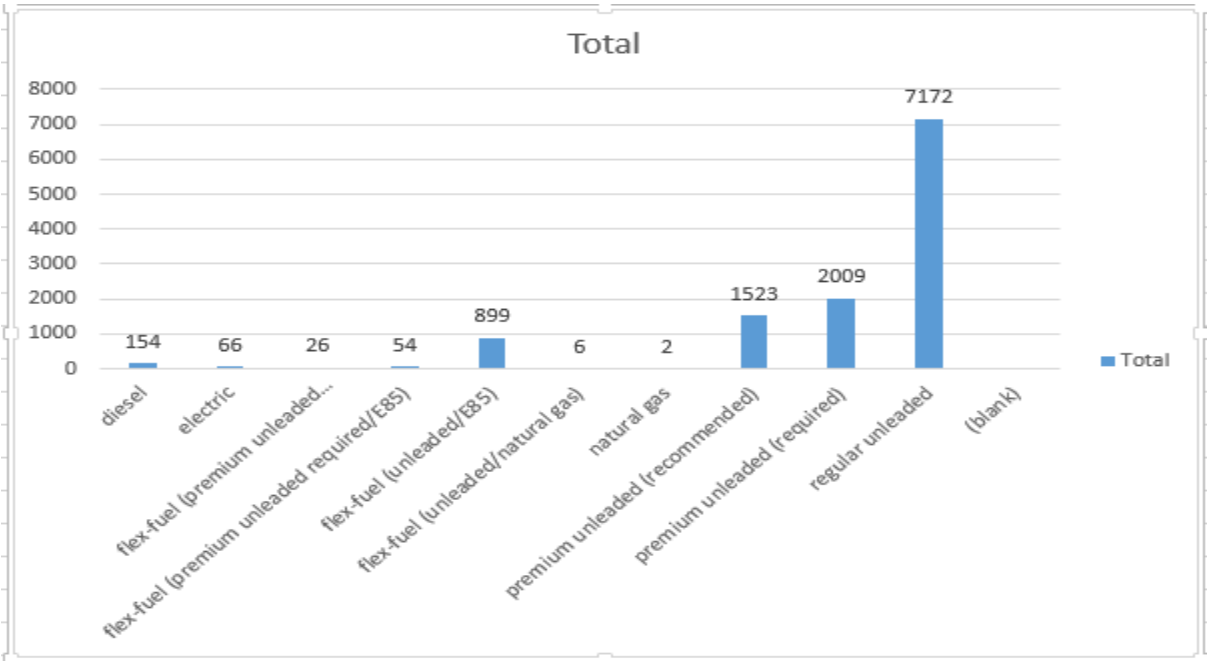
Here is a brief overview of the dataset:

- Number of observations: 11,914
- Number of variables: 16
- File type: CSV (Comma Separated Values)

The variables in the dataset are:

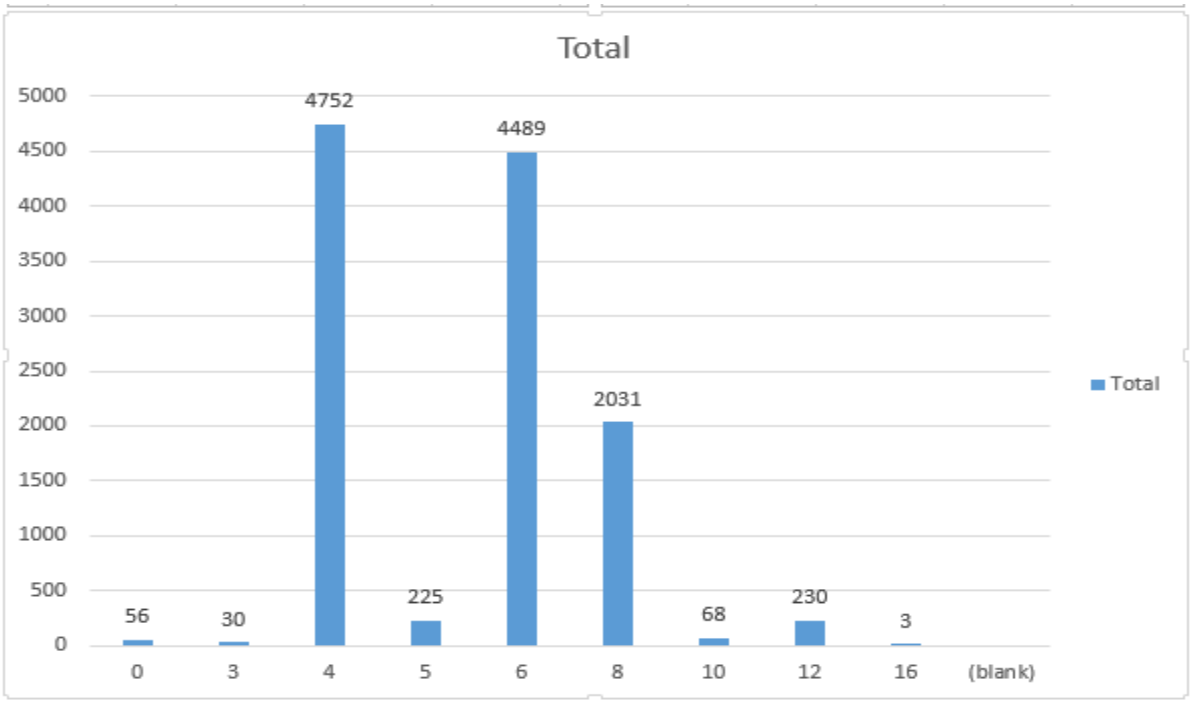
1. **Make:** the make or brand of the car
2. **Model:** the specific model of the car
3. **Year:** the year the car was released
4. **Engine Fuel Type:** the type of fuel used by the car (gasoline, diesel, etc.)
5. **Engine HP:** the horsepower of the car's engine
6. **Engine Cylinders:** the number of cylinders in the car's engine
7. **Transmission Type:** the type of transmission (automatic or manual)
8. **Driven_Wheels:** the type of wheels driven by the car (front, rear, all)
9. **Number of Doors:** the number of doors the car has
10. **Market Category:** the market category the car belongs to (Luxury, Performance, etc.)
11. **Vehicle Size:** the size of the car
12. **Vehicle Style:** the style of the car (Sedan, Coupe, etc.)
13. **Highway MPG:** the estimated miles per gallon the car gets on the highway
14. **City MPG:** the estimated miles per gallon the car gets in the city
15. **Popularity:** a ranking of the popularity of the car (based on the number of times it has been viewed on Edmunds.com)
16. **MSRP:** the manufacturer's suggested retail price of the car

Data Cleaning and preprocessing

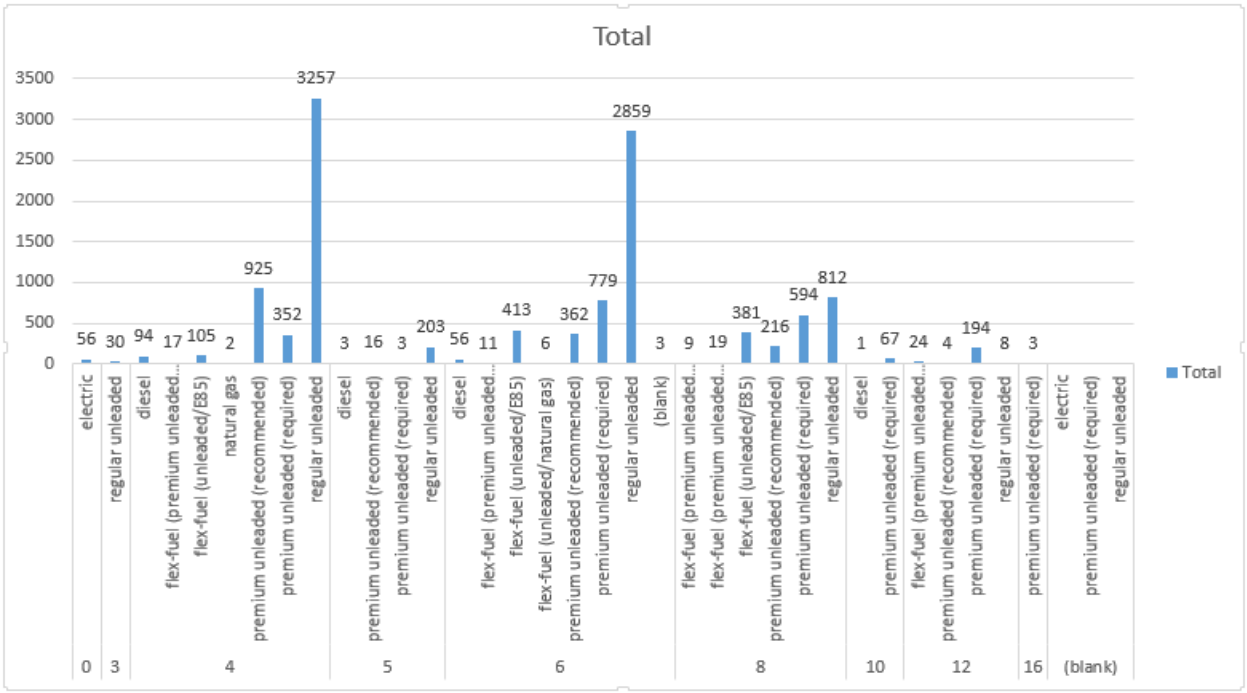


Engine Fuel Type: Regular unleaded engine fuel types has the largest counts on this graph relative to the other fuel types, thus we must fill all of the blank spaces with regular unleaded.

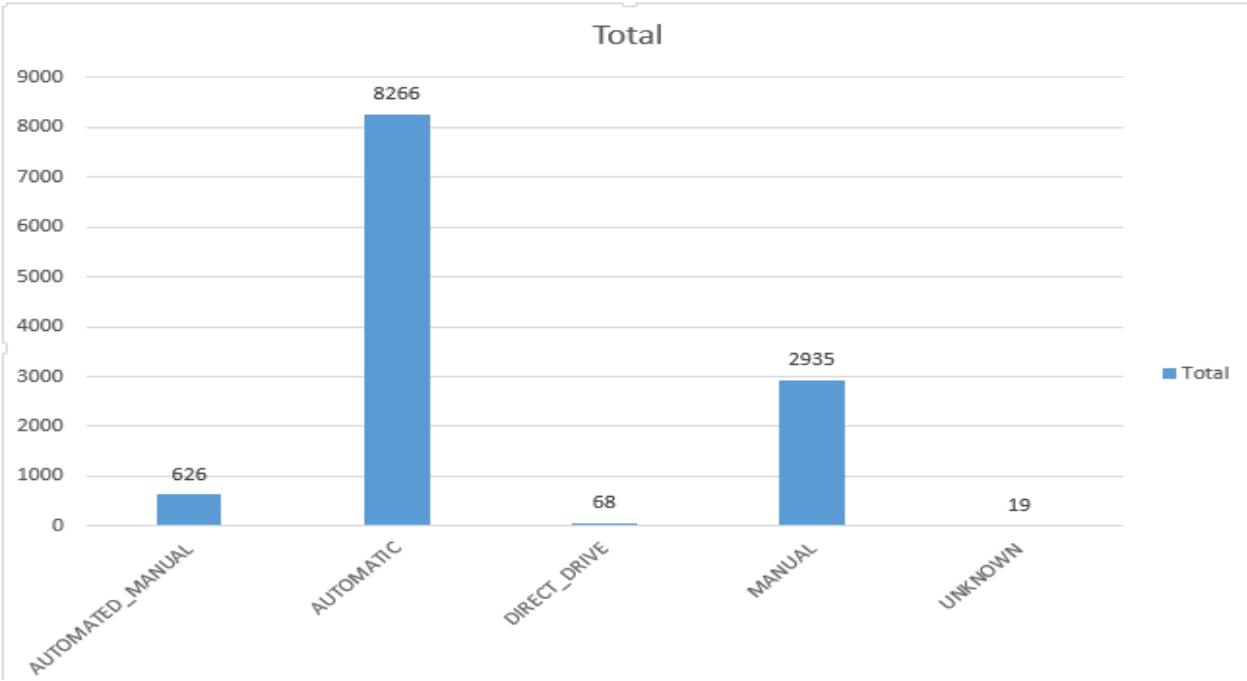
Engine HP: Because there are no outliers, we may fill in blank values in the engine HP column with its mean.249.3919 is the mean value.



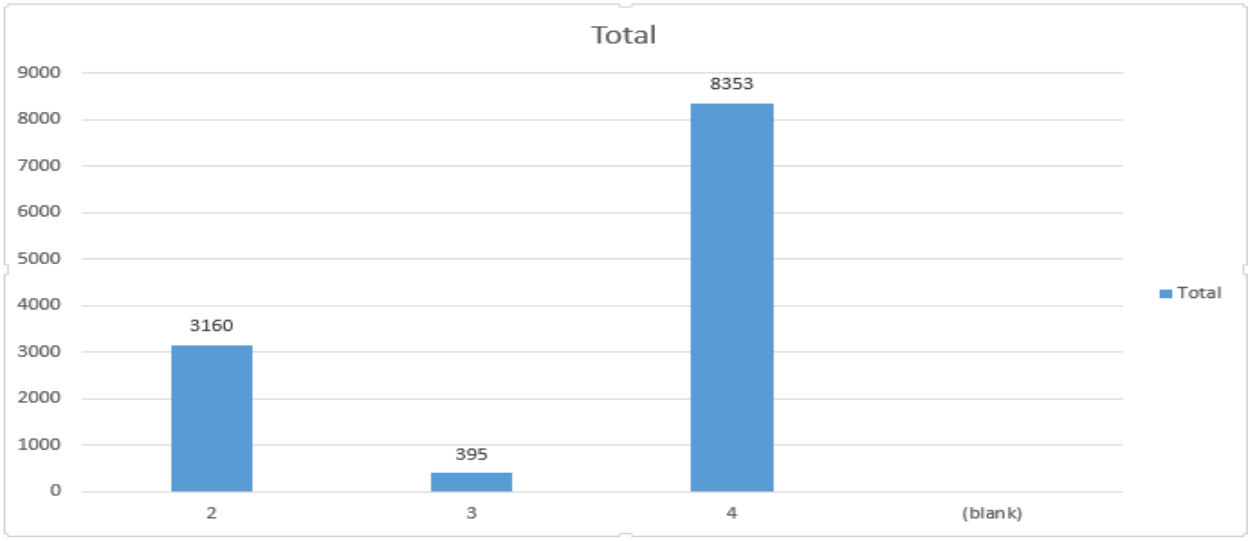
Engine Cylinders: According to the graph, the maximum number of engine cylinders is 4, hence all null values must be replaced with 4 except for those for vehicles using electric.



Engine Cylinders: This graph indicates that the fuel type is electric, thus we must change all null values in the engine cylinder column to 0.



Transmission Type: Since Automatic has the highest count among the unknown values in the transmission type column, I will replace all of them with it.



Number of Doors: I'm going to fill in all of the blanks in the columns for the number of doors in this graph with 4, as its count is larger than the other numbers.

Market Category: Here, we may assume that all N/A values are null, which equals 3742; thus, we must replace all N/A with the second maximum Crossover, which is 1110.

Assumptions

1. Here, I'm assuming that every N/A value in the market category column is null, necessitating replacement.
2. Since it is not possible to make a scatter and bubble graph in a pivot table for tasks 3 and 5 of a dashboard, I established a pivot table, copied the numbers, and then made the graphs because it takes time to calculate each pair individually.

Approach

I must first comprehend the dataset, including the number of rows and columns as well as the number of numerical and categorical variables, before I can begin to clean the data.I'm going to fill in any null values throughout the data cleaning process and look for outliers.Then, in order to uncover insightful information, I'll perform regression analysis along with analysis using a pivot table and graph.

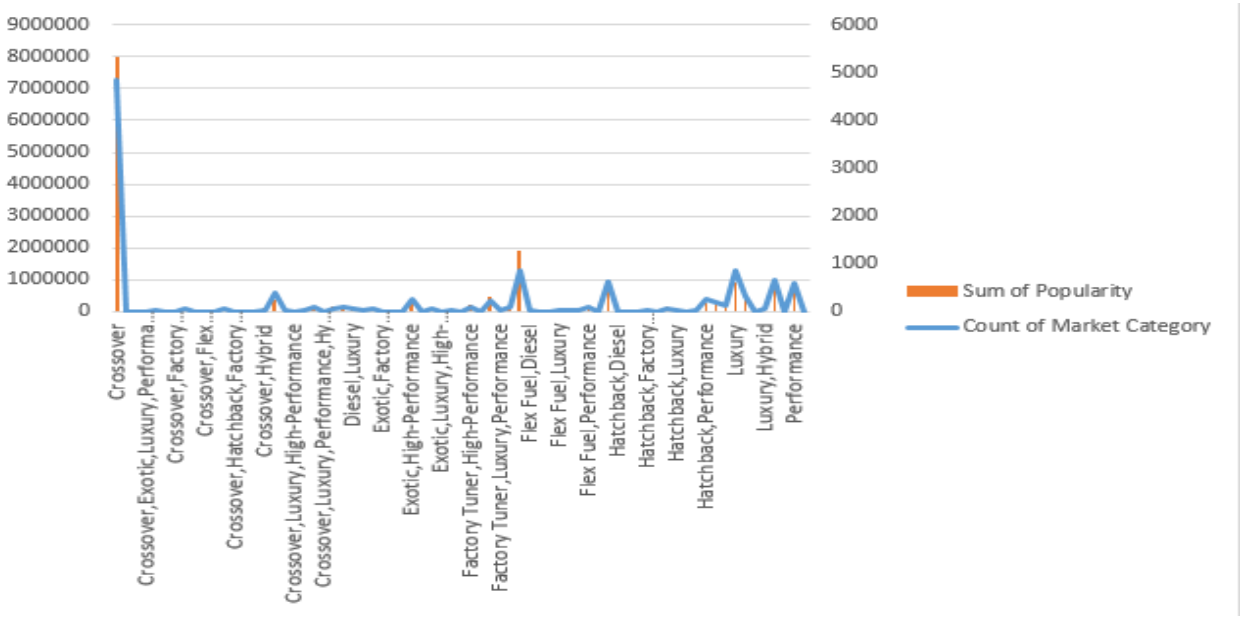
Regression allows me to determine other factors' coefficents with regard to MSRP and detect correlation, which is helpful for figuring out how two variables are related to one another. Graphs can help you comprehend the distribution of variables better.

Tech-Stack Used

Using Microsoft Excel 2013, which I will be using in this instance, I can clean the data and create a pivot table, which is helpful for data analysis. Excel graphs may be used to visualise data as well. Additionally, I use the Excel data analysis tool for regression analysis.

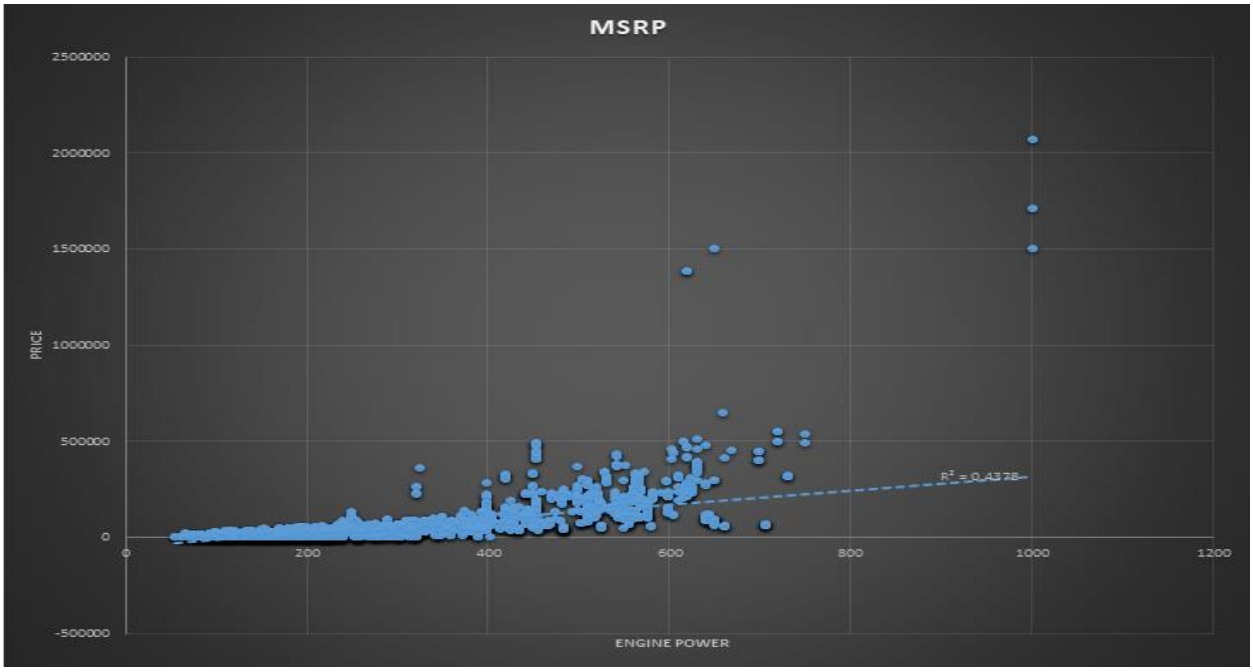
Insight

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



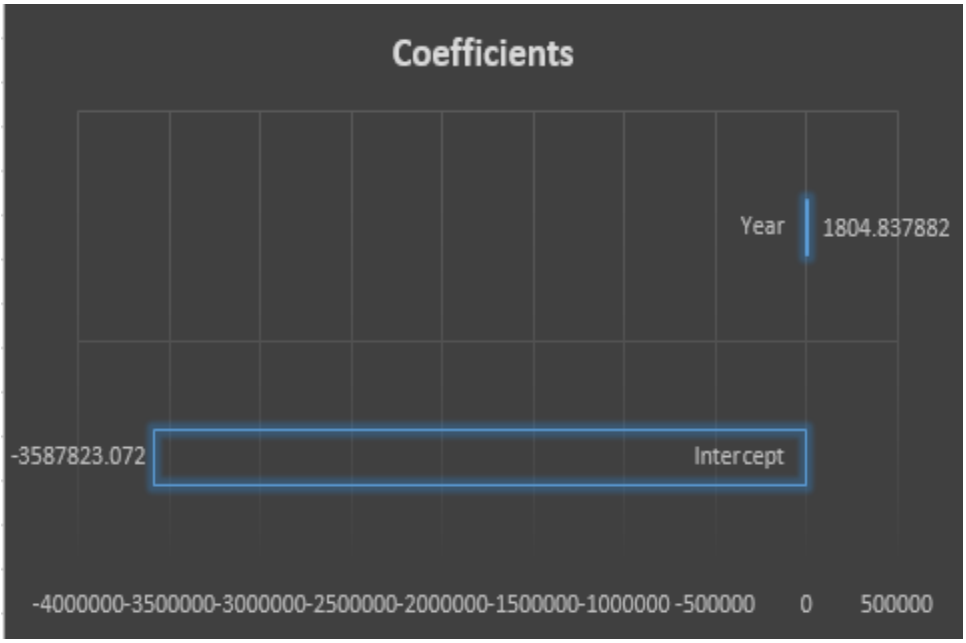
Observation : In this graph we can see that count of market category is proportional to sum of popularity.so we can say that popularity depends on market category means crossover market is more popular as compare to others.

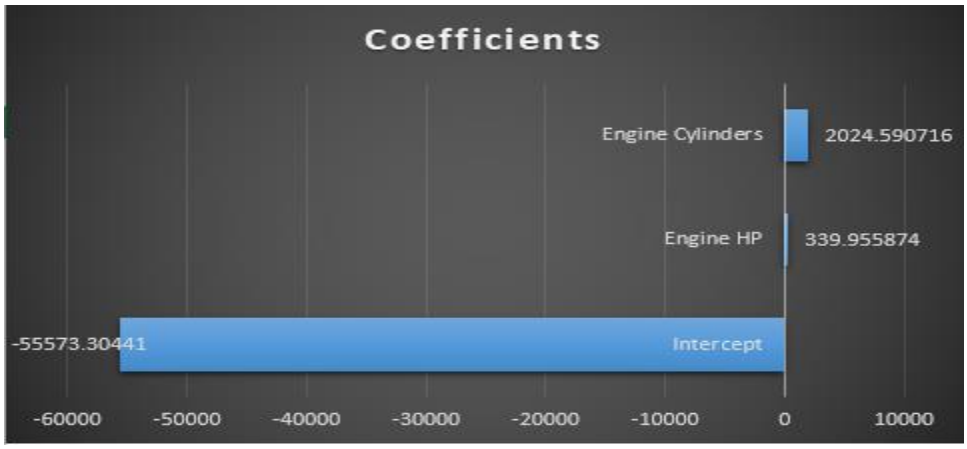
2. What is the relationship between a car's engine power and its price?

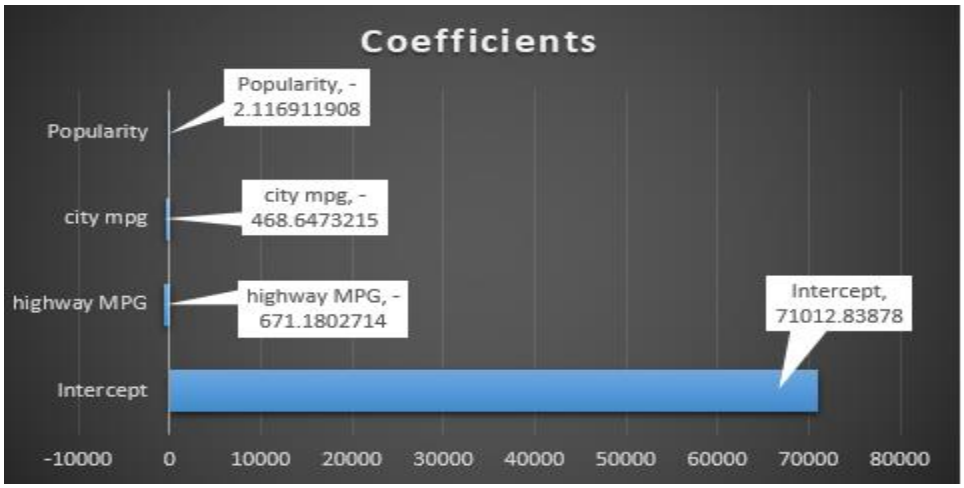


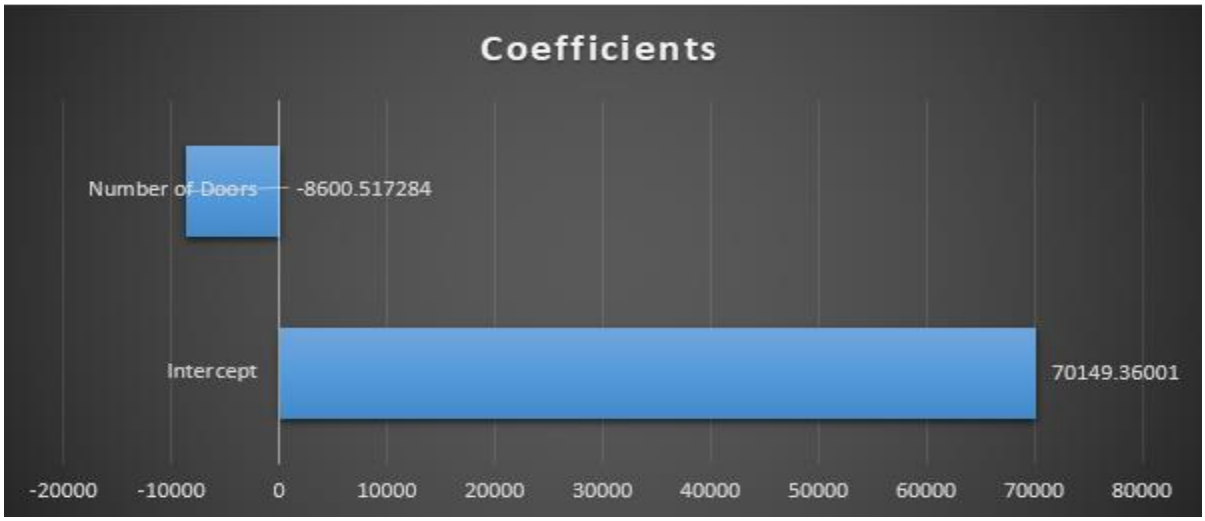
Observation: In this graph we can see that car engine power is linearly proportional to MSRP also we can see that some values in MSRP behave like a outliers but we know that some cars are expensive like Bugatti.so we can say that when MSRP increases when power of car increases.

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



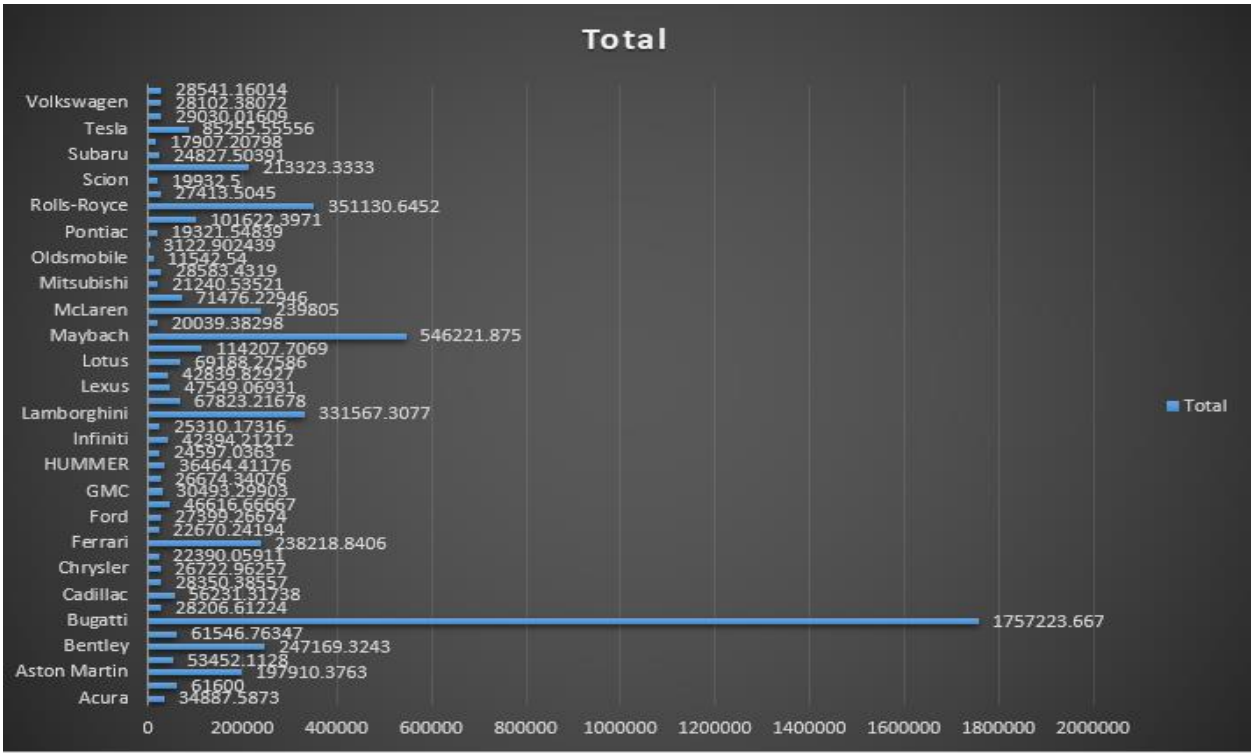






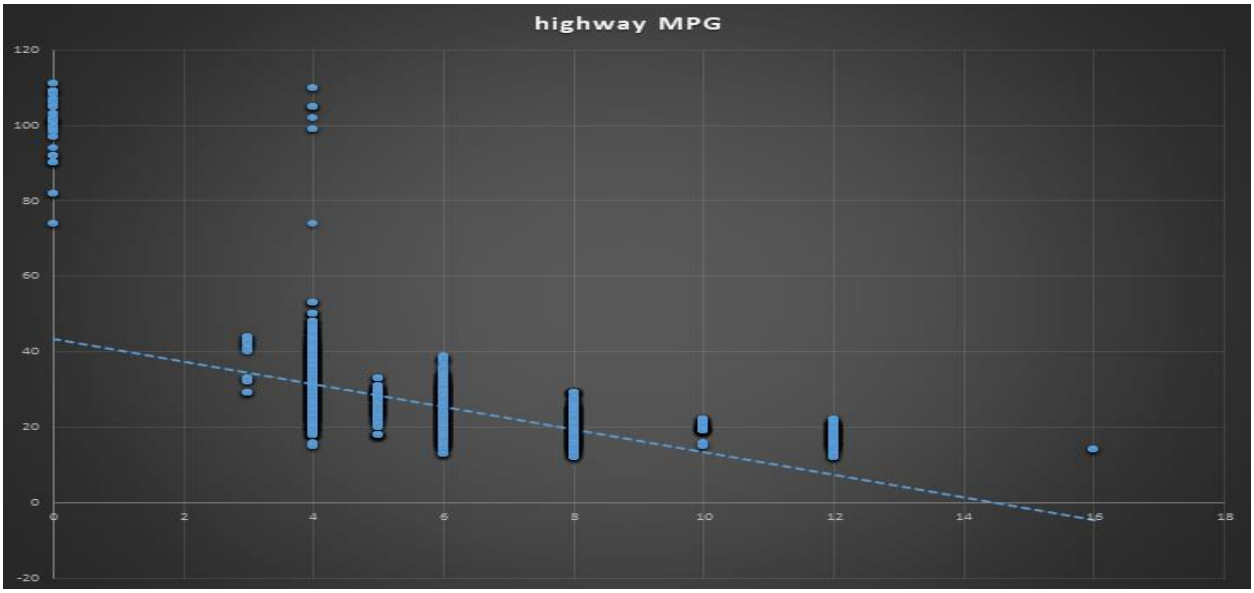
Observation: Here, we can see that the engine cylinders coefficient is larger than the year, city mpg, and high mpg for automobile features; thus, we can conclude that this is a crucial characteristic of any manufacturer as it is inversely related to the engine horsepower.

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



Observation: Here we can see that Bugatti has a highest average price and second highest is Maybach.Also avg car price is depend on every brand some are expensive because they are dealing with luxury segment and high HP.

5. What is the relationship between fuel efficiency and the number of cylinders in a car's engine?



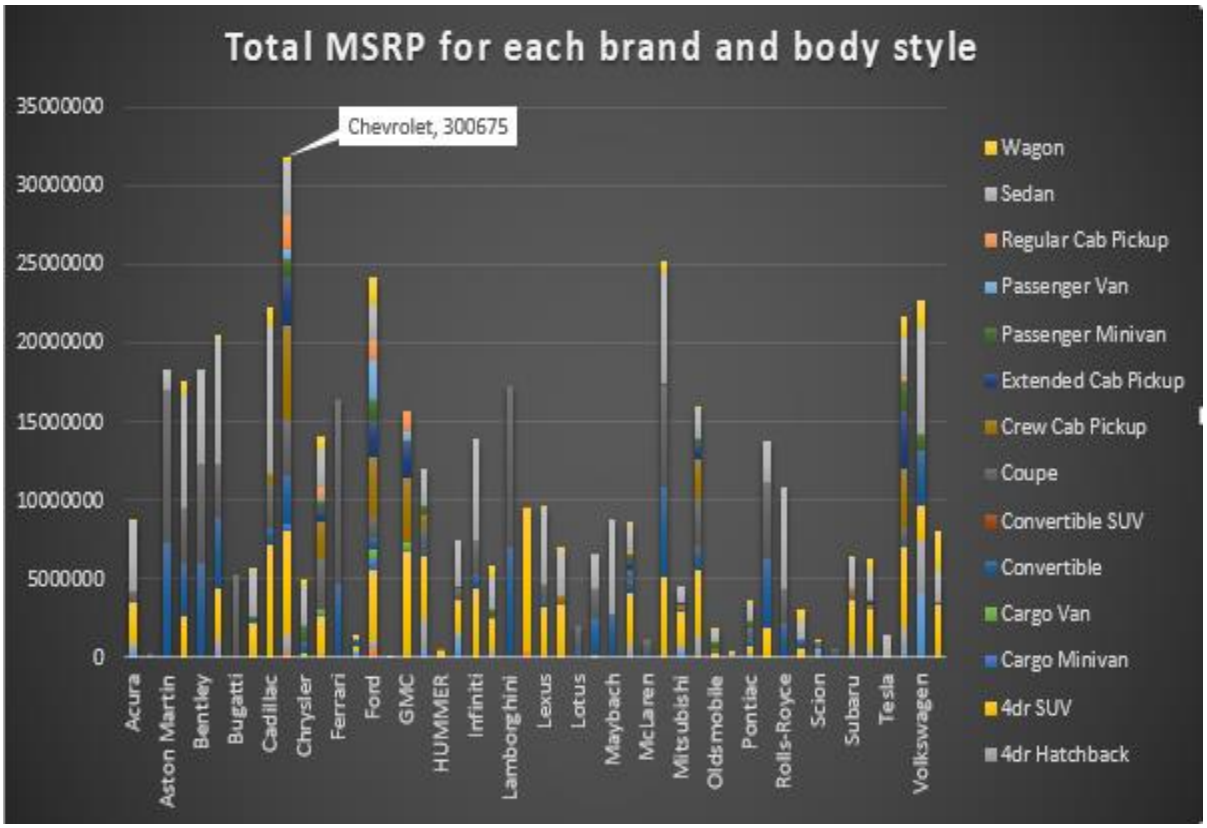
Observation : We are comparing the fuel economy in this scatter plot with regard to the number of cylinders, and since we can see that efficiency decreases as the number of cylinders increases, we may conclude that the relationship between the two is inverse or we can say that number of cylinders increases then power also increases and fuel efficiency decreases.

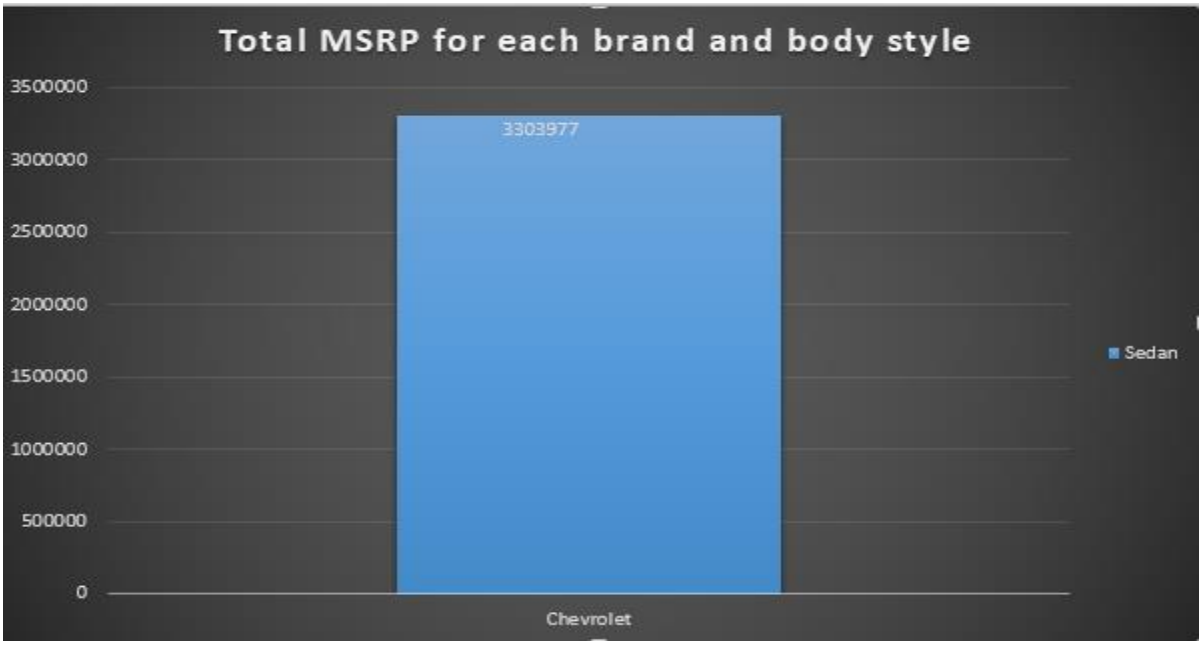
Correlation

	Engine Cylinders	*highway MPG*
Engine Cylinders	1	
highway MPG	-0.6074225	1

Observation: Here we can see that engine cylinders and highway MPG are high negative correlated to each other so I can say that when engine cylinders are increases then MPG decreases.

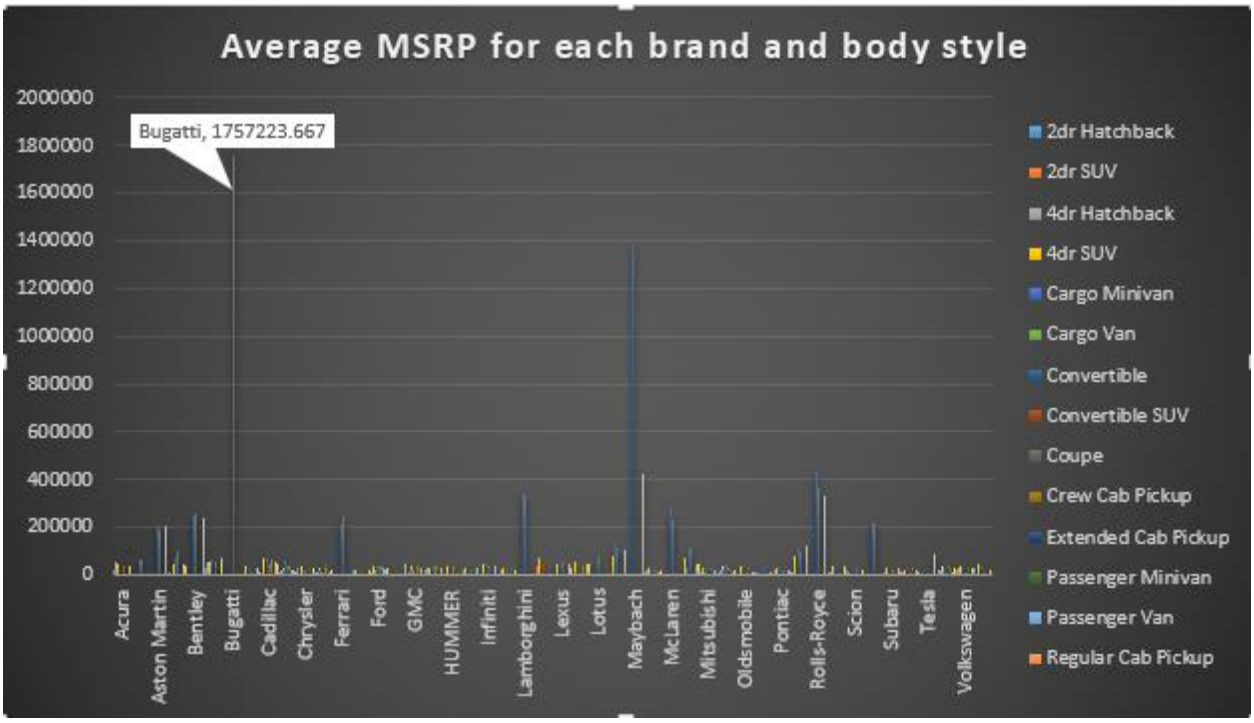
6. How does the distribution of car prices vary by brand and body style?





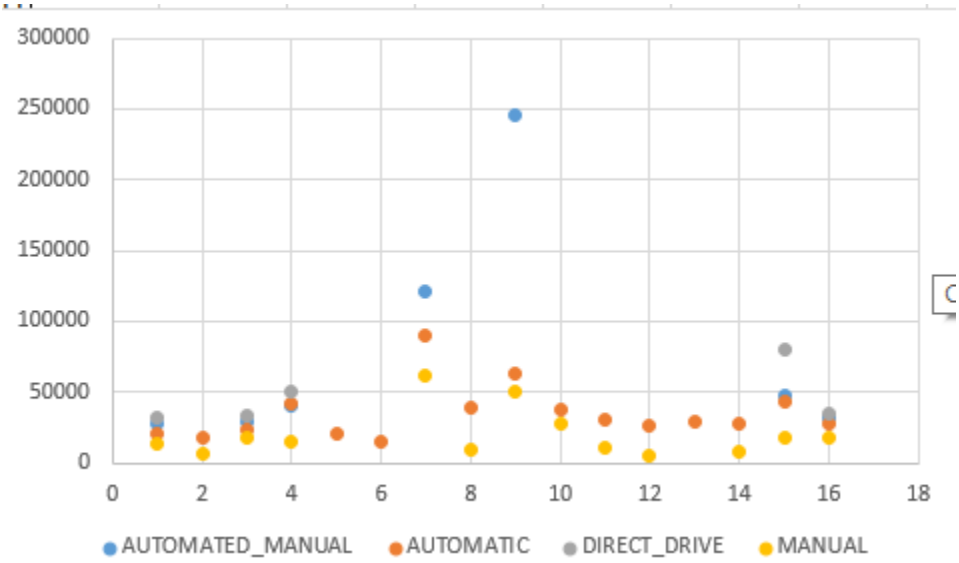
Observation: From both the graphs we can see that in Chevrolet Brand with sedan style have the highest MSRP.

7. Which car brands have the highest and lowest average MSRPs, and how does this vary by body style?



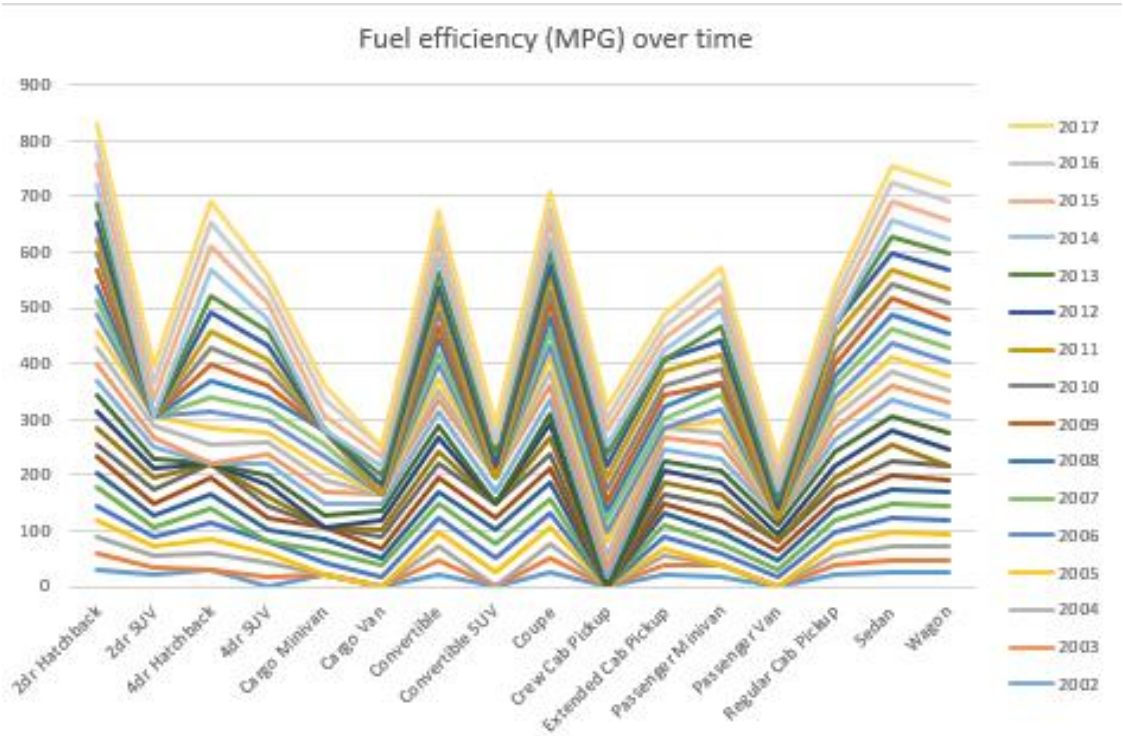
Observation: Highest Average MSRP is Bugatti of coupe body style and the lowest Average MSRP is Chrysler of coupe body style.

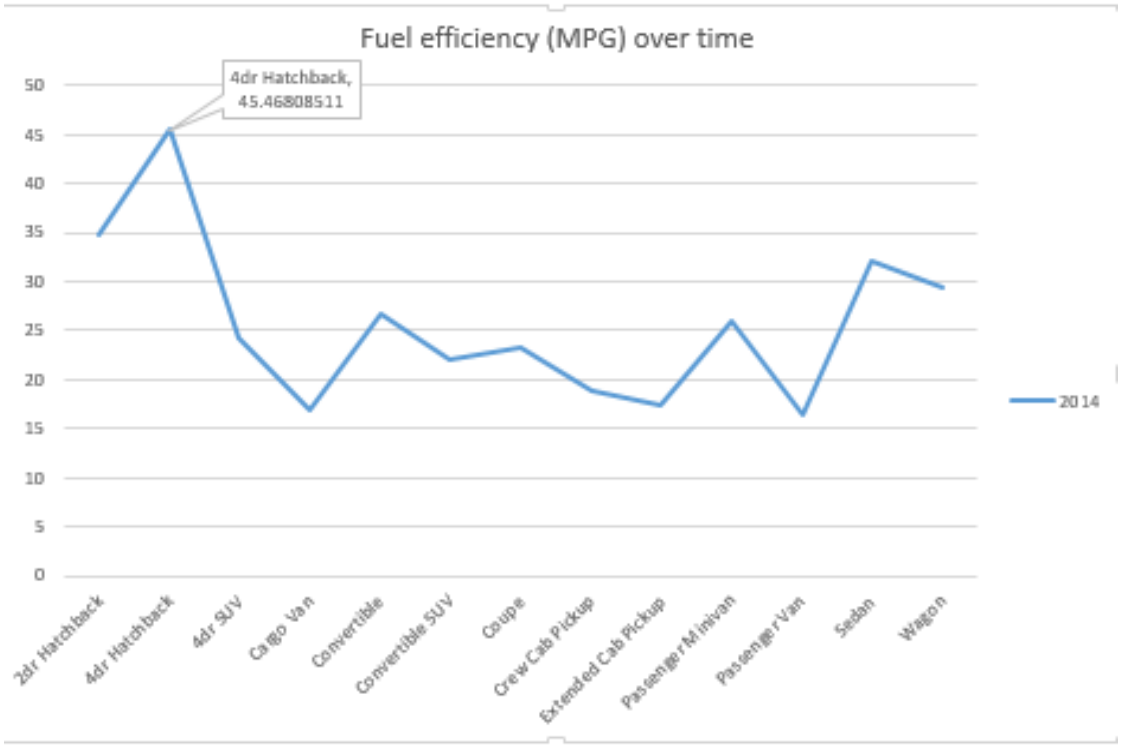
8. How do the different feature such as transmission type affect the MSRP, and how does this vary by body style?

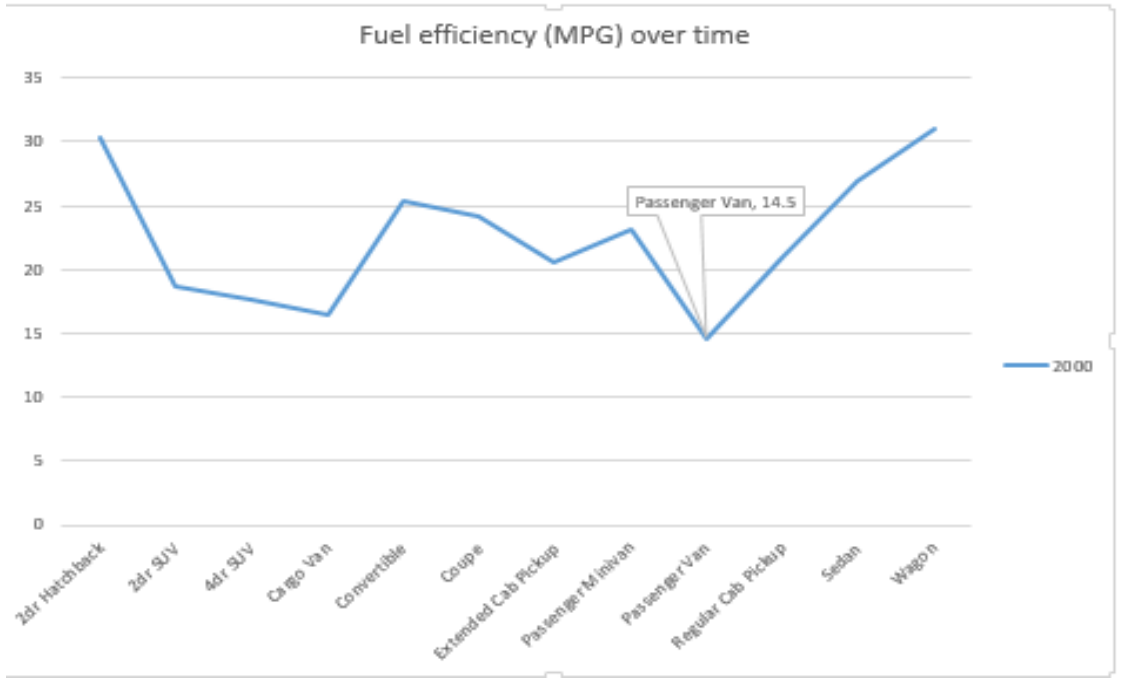


Observation: According to this graph we can see that MSRP is higher in Automatic as compare to other transmission type.

9.How does the fuel efficiency of cars vary across different body styles and model years?







Observation: We can quickly comprehend the distribution of fuel economy over time from the three graphs above. We can observe that the passenger van in 2000 had the lowest efficiency, 14.5, while the 4dr hatchback in 2014 had the best efficiency, 45.46.

10. How does the car's horsepower, MPG, and price vary across different Brands?

Row Labels	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	277.7737003	27.82874618	53457.77676
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	246.985175	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	148.6802546	37.33870968	22670.24194
Ford	243.1908003	24.00681044	27399.26674
Genesis	347.3333333	25.33333333	46616.66667
GMC	259.8446602	21.4038835	30493.29903
Honda	195.9883828	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
Kia	207.748743	30.65367965	25310.17316
Lamborghini	614.0769231	18.01923077	331567.3077
Land Rover	322.0979021	22.12587413	67823.21678
Lexus	277.4158416	25.87623762	47549.06931
Lincoln	283.177655	24.48780488	42839.82927
Lotus	275.9655172	26.55172414	69188.27586
Maserati	420.7931034	20.29310345	114207.7069
Maybach	590.5	16	546221.875
Mazda	171.9929078	27.85106383	20039.38298
McLaren	610.4	22.2	239805
Mercedes-Benz	349.8962944	24.83002833	71476.22946
Mitsubishi	173.7858776	27.54460094	21240.53521
Nissan	240.0912532	27.79928315	28583.4319
Oldsmobile	177.4666667	26.23333333	11542.54
Plymouth	131.5609756	27.96341463	3122.902439
Pontiac	190.2956989	27.06989247	19321.54839
Porsche	392.7941176	25.36764706	101622.3971
Rolls-Royce	487.5483871	19.12903226	351130.6452
Saab	220.5225225	26.35135135	27413.5045
Scion	154.4333333	32.3	19932.5
Spyker	400	18	213323.3333
Subaru	197.3085938	28.68359375	24827.50391
Suzuki	160.2877493	26.03418803	17907.20798
Tesla	249.3919284	98.94444444	85255.55556
Toyota	236.1833564	26.45308311	29030.01609
Volkswagen	189.7577256	32.12855377	28102.38072
Volvo	230.9715302	27.20284698	28541.16014

Observation : Here we can see that all three variables are correlated to each other when HP engine is increases then MPG is decreases and MSRP is increases so we can say that fuel efficiency is dependent on HP when it increases fuel consumption increases.

Result

- This graph demonstrates how the number of market categories is related to the total level of popularity.
- In this graph, we can see that the MSRP and engine power of automobiles are linearly related. We can also observe that some MSRP values behave strangely, but we know that some cars are pricey, like the Bugatti.
- Here, we can see that the engine cylinders coefficient is more than the year, city mpg, and high mpg for vehicle characteristics; thus, we can draw the conclusion that this is an essential quality of any manufacturer as it is inversely connected to the engine horsepower.
- Here, it is clear that Bugatti has the highest average price, with Maybach coming in second.
- The fuel economy and the number of cylinders are being compared in this scatter plot, and since we can see that efficiency falls as the number of cylinders rises, we may infer that the connection between the two is inverse.
- As you can see, engine cylinders and highway MPG have a strong negative correlation. Therefore, as engine cylinders grow, MPG falls.
- The two graphs show that the Chevrolet Brand's MSRP for sedans is the highest.
- The Bugatti coupe body type has the highest average MSRP, and the Chrysler coupe body style has the lowest average MSRP.
- This graph shows that the MSRP for an automatic gearbox is more than for other gearbox types.
- The three graphs above allow us to rapidly understand the distribution of fuel economy over time. We can see that in 2000, the passenger van had the lowest efficiency (14.5), while the greatest efficiency (4dr hatchback) in 2014 was 45.46.
- Here, we can see that all three variables are associated with one another; as engine horsepower grows, MPG declines and MSRP rises, indicating that as engine horsepower rises, so does fuel consumption.

Google Drive Link : PDF/PPT and Video Presentation,Excel Files in the Analysis Files