

ANALYZING IMPACT OF CAR FEATURES ON PRICE AND PROFITABILITY

Project Description:

The project analyzing impact of car features on price and profitability is about finding the required insights for the client from the automotive industry which is evolving very fast . The problem statement is how can a car manufacturer optimize pricing and product development decision while meeting consumer demand. We are provided by a raw data called Car_data.csv .

A brief overview of the dataset:

- Number of observations: 11,159
- Number of variables: 16
- File type: CSV (Comma Separated Values)

The variables in the dataset are:

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

After downloading the raw data we use various cleaning techniques to clean the raw data to make the data ready for analysis . During cleaning we get rid of all duplicate values in the data set . Then we delete all the blank rows to ensure better analysis. Once the cleaning is done we look to delete the columns which are of no use in analysis but in this dataset there was no column which could be deleted.

Approach :

The main approach towards this project is to first understand the dataset provided. Then using various cleaning techniques we can clean the data and get rid of null values and duplicates to prepare data for analysis stage . We will use various data analysis techniques like using pivot tables, fuctions , regression analysis to find the insights required At the end we will display the insights extracted using various tables and charts to make an interactive dashboard and make it more easy to quickly understand the insights acquired

Tech used : The main software used during the project is Microsoft Excel

Insights :

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

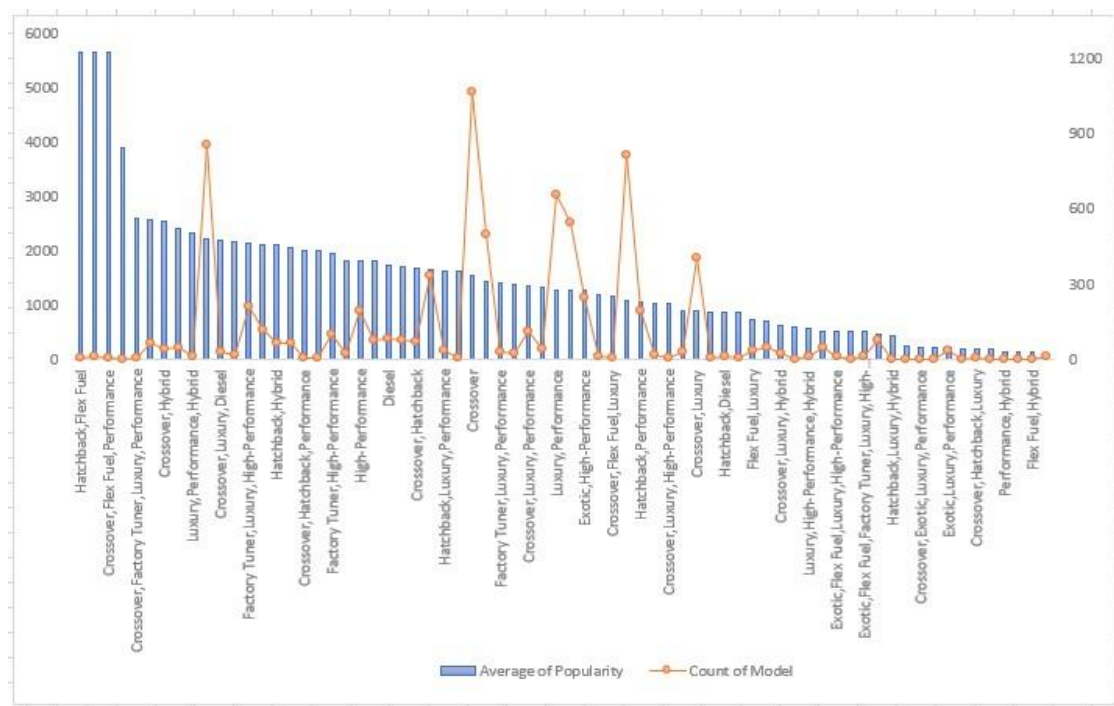
- **Task 1.A:** Create a pivot table that shows the number of car models in each market category and their corresponding popularity scores.

Market	Average of Popularity	Count of Model
Hatchback, Flex Fuel	5657	7
Flex Fuel, Diesel	5657	16
Crossover, Flex Fuel, Performance	5657	6
Crossover, Luxury, Performance, Hybrid	3916	2
Crossover, Factory Tuner, Luxury, Performance	2607	5
Crossover, Performance	2586	69
Crossover, Hybrid	2563	42
Diesel, Luxury	2416	47
Luxury, Performance, Hybrid	2333	11
Flex Fuel	2226	855
Crossover, Luxury, Diesel	2196	33
Hatchback, Factory Tuner, Performance	2174	21
Factory Tuner, Luxury, High-Performance	2133	215
Hybrid	2117	121
Hatchback, Hybrid	2111	64
Crossover, Flex Fuel	2074	64
Crossover, Hatchback, Performance	2009	6
Crossover, Hatchback, Factory Tuner, Performance	2009	6
Factory Tuner, High-Performance	1966	104
Crossover, Factory Tuner, Luxury, High-Performance	1823	26
High-Performance	1823	198
Factory Tuner, Performance	1818	81
Diesel	1731	84
Flex Fuel, Performance	1702	81
Crossover, Hatchback	1676	72
Luxury, High-Performance	1668	334
Hatchback, Luxury, Performance	1632	36
Crossover, Flex Fuel, Luxury, Performance	1624	6
Crossover	1539	1068
Performance	1443	503
Factory Tuner, Luxury, Performance	1413	31
Flex Fuel, Luxury, Performance	1380	28
Crossover, Luxury, Performance	1349	112
Hatchback, Luxury	1323	45
Luxury, Performance	1293	659
Hatchback	1279	547

Exotic,High-Performance	1276	246
Hatchback,Factory Tuner,High-Performance	1205	13
Crossover,Flex Fuel,Luxury	1173	10
Luxury	1084	815
Hatchback,Performance	1074	198
Exotic,Factory Tuner,High-Performance	1046	21
Crossover,Luxury,High-Performance	1037	9
Flex Fuel,Luxury,High-Performance	898	32
Crossover,Luxury	889	406
Hatchback,Factory Tuner,Luxury,Performance	887	9
Hatchback,Diesel	873	14
Crossover,Diesel	873	7
Flex Fuel,Luxury	747	39
Luxury,Hybrid	725	48
Crossover,Luxury,Hybrid	631	24
Factory Tuner,Luxury	617	2
Luxury,High-Performance,Hybrid	569	12
Exotic,Factory Tuner,Luxury,High-Performance	523	51
Exotic,Flex Fuel,Luxury,High-Performance	520	11
Exotic,Factory Tuner,Luxury,Performance	520	3
Exotic,Flex Fuel,Factory Tuner,Luxury,High-Performance	520	13
Exotic,Luxury,High-Performance	473	77
Hatchback,Luxury,Hybrid	454	3
Flex Fuel,Factory Tuner,Luxury,High-Performance	258	1
Crossover,Exotic,Luxury,Performance	238	1
Crossover,Exotic,Luxury,High-Performance	238	1
Exotic,Luxury,Performance	217	36
Crossover,Factory Tuner,Performance	210	4
Crossover,Hatchback,Luxury	204	7
Exotic,Luxury,High-Performance,Hybrid	204	1
Performance,Hybrid	1	1
Flex Fuel,Performance,Hybrid	155	2
Flex Fuel,Hybrid	155	2
Exotic,Luxury	113	12
Grand Total	1513	7736

We create the pivot table required by putting market category in rows and calculating average popularity and count of model in every market category.

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



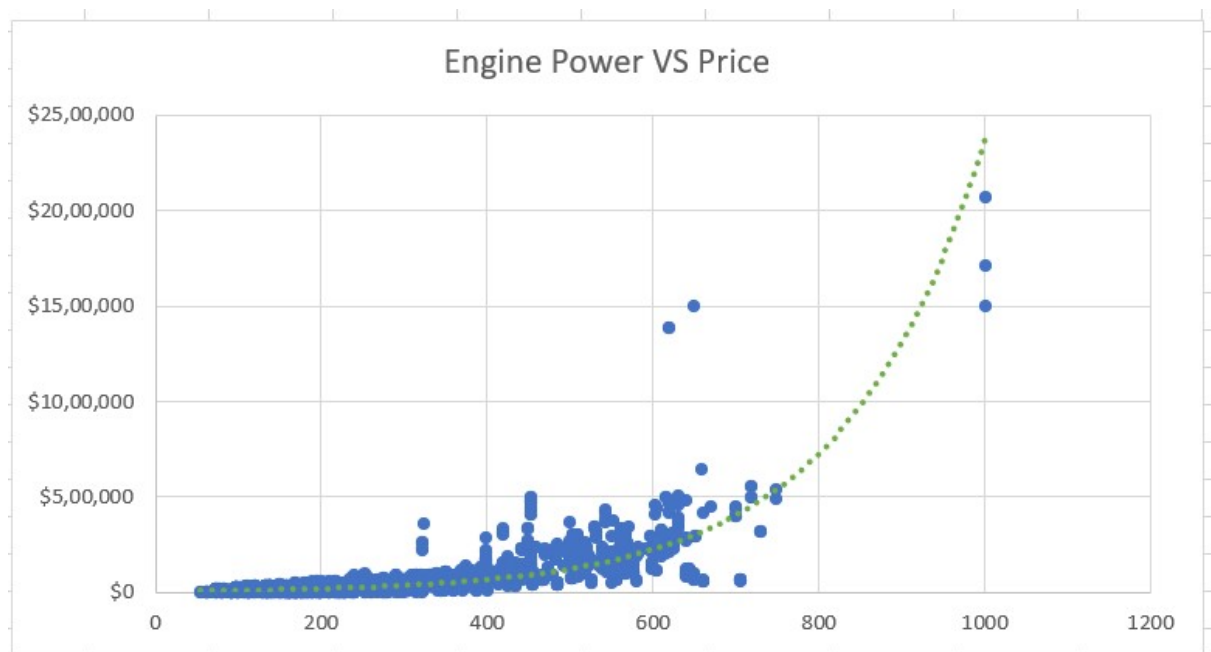
From the above graph we find out that Market Categories like

- Hatchback, Flex Fuel ,
- Crossover, Flex Fuel, Performance
- Flex Fuel, Diesel

are the most popular among all.

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

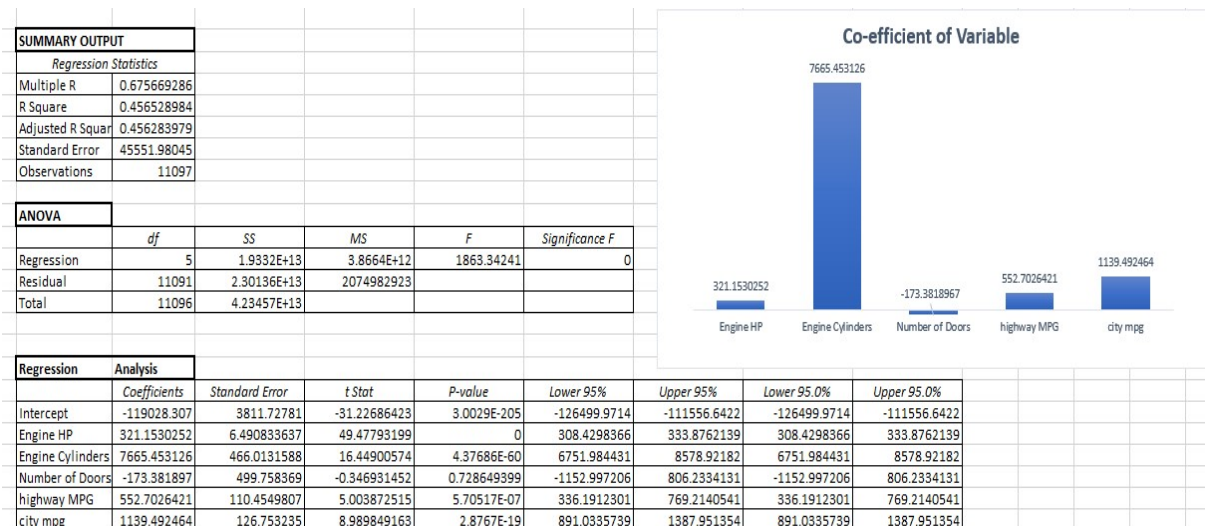
- **Task 2:** Create a scatter chart that plots engine power on the x-axis and price on the y-axis. Add a trendline to the chart to visualize the relationship between these variables.



We create this scatter plot using Engine HP and MSRP columns from the table and add a trendline to analyze the trend . From the above graph we can say that as the engine power increases the price of the car increases .

3) 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.



We already have some car features like engine hp , engine cylinders , number of doors , highway and city mpg which contribute in determining the car's price . To find the most important feature we use regression analysis on the columns mentioned above to find out the co efficient value of each variable which helps to determine the feature . The final decision is based on which variable co efficient value is highest . The variable with highest co efficient value is the most important feature . From the above graph we can say that engine cylinders is the most important feature to determine the car's price whereas number of doors is the least important feature .

4) 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.

2		
3	Manufacturer	Average of MSRP
4	Acura	\$35,087
5	Alfa Romeo	\$61,600
6	Aston Martin	\$1,98,123
7	Audi	\$54,574
8	Bentley	\$2,47,169
9	BMW	\$62,163
10	Bugatti	\$17,57,224
11	Buick	\$29,034
12	Cadillac	\$56,368
13	Chevrolet	\$29,000
14	Chrysler	\$26,723
15	Dodge	\$24,857
16	Ferrari	\$2,38,219
17	FIAT	\$22,206
18	Ford	\$28,523
19	Genesis	\$46,617
20	GMC	\$32,444
21	Honda	\$26,609
22	HUMMER	\$36,464
23	Hyundai	\$24,926
24	Infiniti	\$42,640
25	Kia	\$25,319
26	Lamborghini	\$3,31,567
27	Land Rover	\$68,067
28	Lexus	\$47,549
29	Lincoln	\$43,560
30	Lotus	\$68,377
31	Maserati	\$1,13,684
32	Maybach	\$5,46,222
33	Mazda	\$20,107

33	Mazda	\$20,107
34	McLaren	\$2,39,805
35	Mercedes-Benz	\$72,135
36	Mitsubishi	\$21,316
37	Nissan	\$28,856
38	Oldsmobile	\$12,844
39	Plymouth	\$3,297
40	Pontiac	\$19,899
41	Porsche	\$1,01,622
42	Rolls-Royce	\$3,51,131
43	Saab	\$27,880
44	Scion	\$19,933
45	Spyker	\$2,14,990
46	Subaru	\$24,241
47	Suzuki	\$18,021
48	Toyota	\$28,759
49	Volkswagen	\$28,947
50	Volvo	\$29,725
51	Grand Total	\$41,928

We create the pivot by keeping the Make(brand) of car in rows and then calculate average msrp for each brand in pivot table .

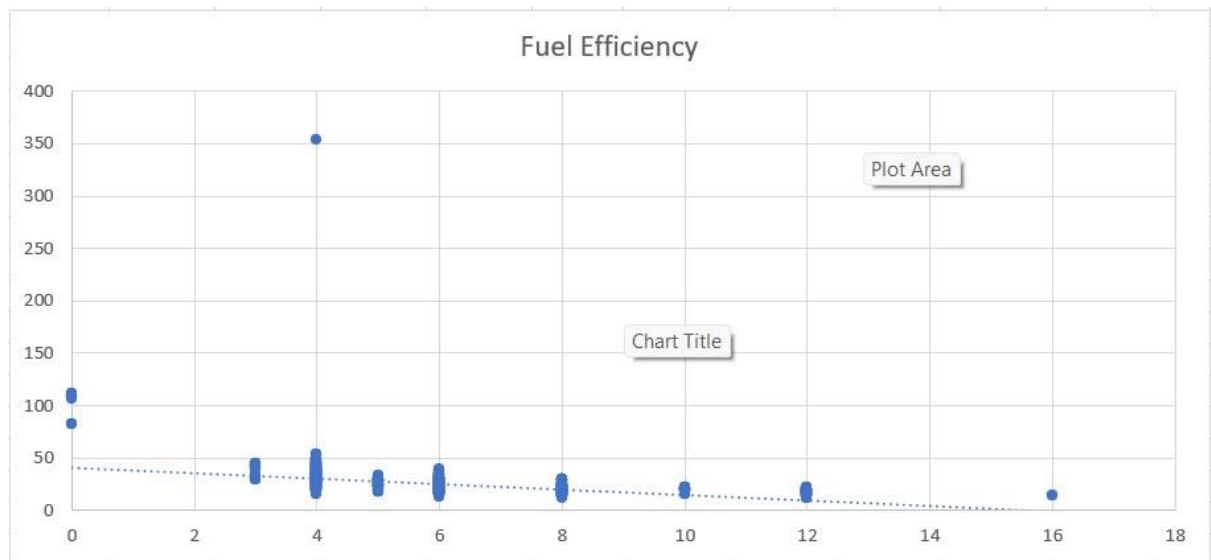
- **Task 4.B:** Create a bar chart or a horizontal stacked bar chart that visualizes the relationship between manufacturer and average price.



From the above tasks performed we can analyze that Buggati Brand of car have the highest average msrp where Plymouth brand has lowest average msrp of the car .

5) 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.

Co-relation coefficient	-0.61474
--------------------------------	-----------------

From the above task performed we can say that along the line as the number of cylinders in a car increases it has a negative effect on the highway mpg i.e highway mpg also decreases . If we want to keep highway mpg as high as possible there should only be 4 cylinders in the car .

We also found out the co-relation coefficient between cylinders and highway mpg using CORREL() function in excel . It gave a negative value which indicates that the relationship between these two variable is not strong . We also added a trend line to analyze the trend which suggests that it is going in negative direction which is not good for the industry .

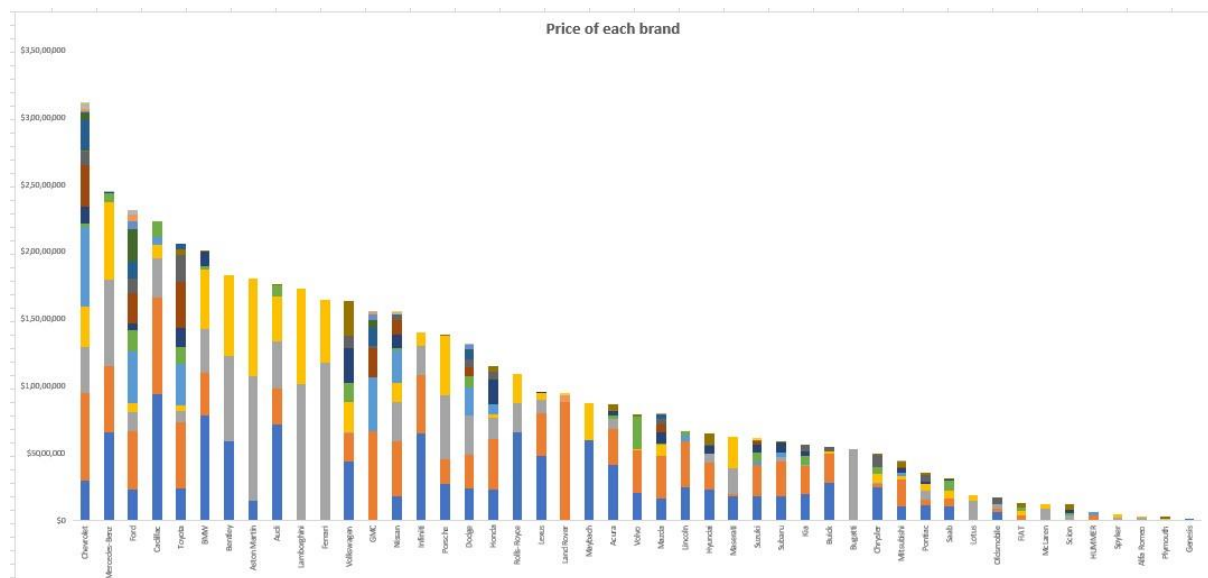
Building a dashboard :

The next part of the project is to create an interactive dashboard in excel to help the client visualize the data easily and make decisions based on the dashboard result . We make use of various charts, slicers and filters to help make the dashboard very interactive .

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

[illegible]

To find out the distribution of car prices for different brands and body we first create a pivot by assigning the brand to rows and body style to columns and then calculating msrp for each brand and body style with the help of pivot table . To understand the pivot table better we created a stacked column chart and come to the conclusion that Chevrolet brand is producing car of every style so their total msrp is highest , whereas Genesis brand has lowest total msrp .



For this task we again create pivot table same like we did in the previous task the only difference is that instead of using sum of msrp in values section from above task this time we give average of msrp to pivot table to be calculated for each body type of every brand .

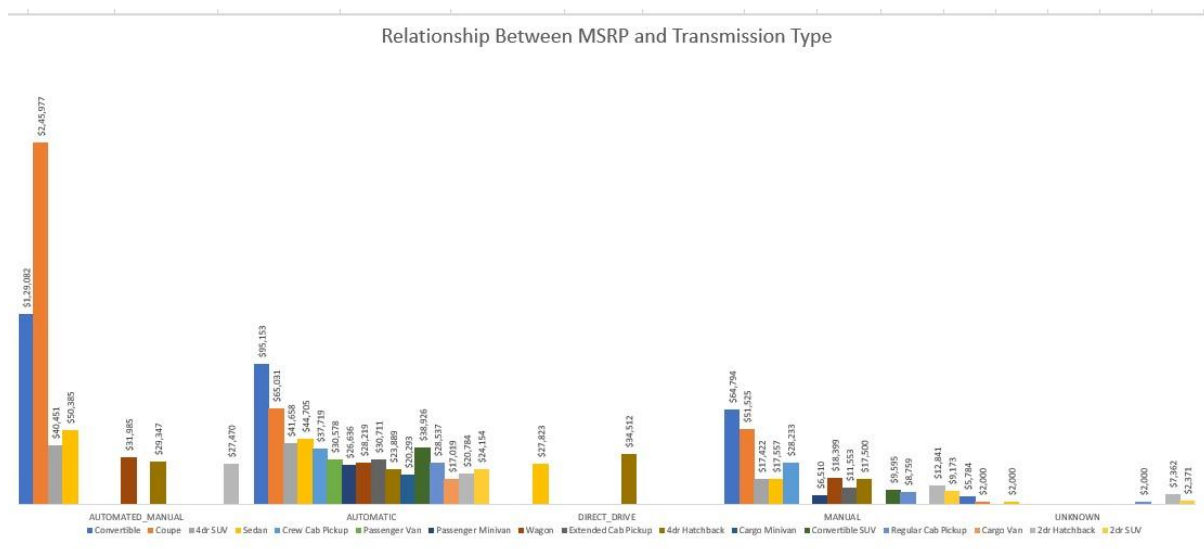
[illegible]

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

For this we first create a pivot table using vehicle style for rows, transmission type for columns and with the help of pivot table we calculate average msrp .

1	Average of MSRP	Transmission Type					
2	Vehicle Style	AUTOMATED_MANUAL	AUTOMATIC	DIRECT_DRIVE	MANUAL	UNKNOWN	Grand Total
3	Convertible	\$1,29,082	\$95,153		\$64,794	\$5,784	\$88,217
4	Coupe	\$2,45,977	\$65,031		\$51,525	\$2,000	\$78,480
5	4dr SUV	\$40,451	\$41,658		\$17,422		\$40,736
6	Sedan	\$50,385	\$44,705	\$27,823	\$17,557	\$2,000	\$40,175
7	Crew Cab Pickup		\$37,719		\$28,233		\$37,183
8	Passenger Van		\$30,578				\$30,578
9	Passenger Minivan		\$26,636		\$6,510		\$26,215
10	Wagon	\$31,985	\$28,219		\$18,399		\$26,084
11	Extended Cab Pickup		\$30,711		\$11,553		\$23,042
12	4dr Hatchback	\$29,347	\$23,889	\$34,512	\$17,500		\$22,061
13	Cargo Minivan		\$20,293				\$20,293
14	Convertible SUV		\$38,926		\$9,595		\$17,975
15	Regular Cab Pickup		\$28,537		\$8,759	\$2,000	\$17,855
16	Cargo Van		\$17,019				\$17,019
17	2dr Hatchback	\$27,470	\$20,784		\$12,841	\$7,362	\$16,063
18	2dr SUV		\$24,154		\$9,173	\$2,371	\$14,307
19	Grand Total	\$1,08,719	\$41,852	\$33,620	\$28,285	\$3,648	\$41,928

To understand this pivot table more easily we created column chart .



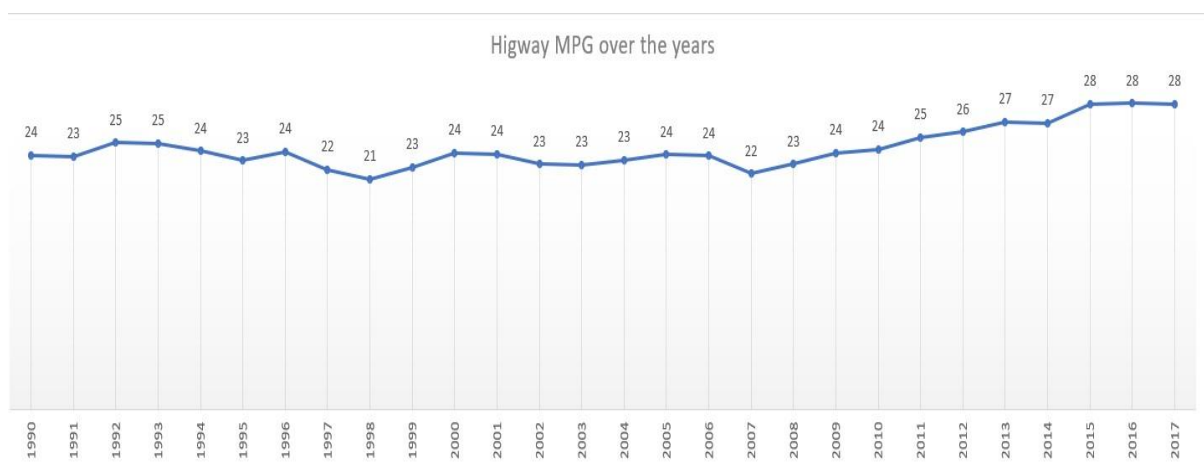
From above chart we can understand that Coupe and Convertible body style vehicles have high average msrp , whereas unknown transmission type vehicles have the lowest average msrp.

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

To find the efficiency of cars across different body styles and model years we use pivot table . We find out the average fuel efficiency (Highway MPG) of cars of different body style by assigning year and body style to rows and then asking pivot table to calculate average fuel efficiency(Highway MPG) over the years . Then we subgroup the year column to get fuel efficiency(Highway MPG) only for different years .

1	Years	Average of highway MPG
2	+ 1990	24
3	+ 1991	23
4	+ 1992	25
5	+ 1993	25
6	+ 1994	24
7	+ 1995	23
8	+ 1996	24
9	+ 1997	22
10	+ 1998	21
11	+ 1999	23
12	+ 2000	24
13	+ 2001	24
14	+ 2002	23
15	+ 2003	23
16	+ 2004	23
17	+ 2005	24
18	+ 2006	24
19	+ 2007	22
20	+ 2008	23
21	+ 2009	24
22	+ 2010	24
23	+ 2011	25
24	+ 2012	26
25	+ 2013	27
26	+ 2014	27

To understand this pivot table we plot line chart



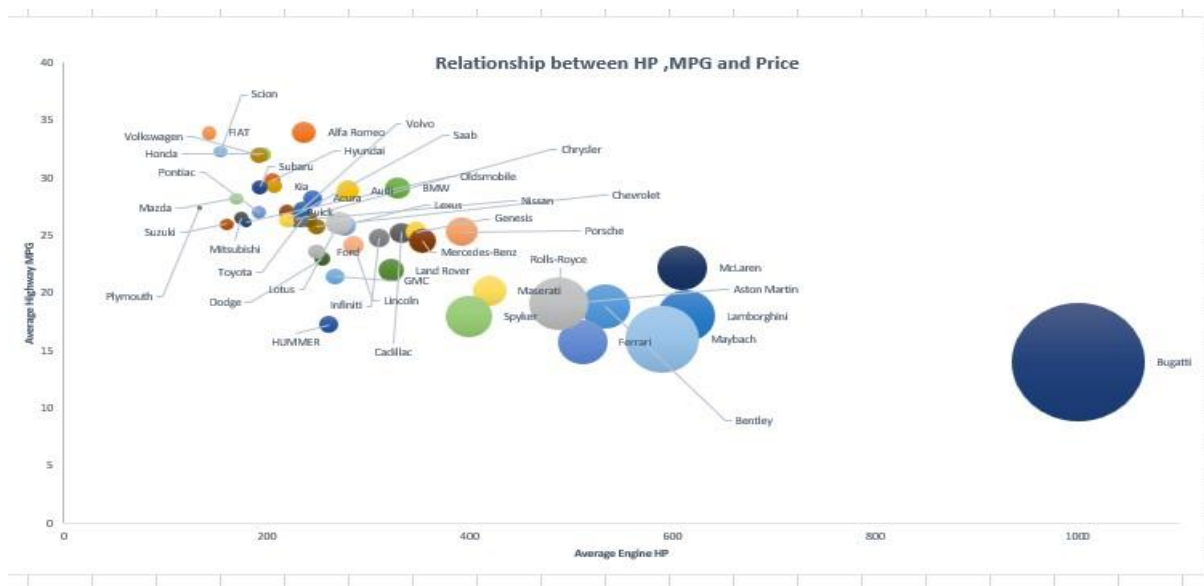
From the chart we understand that Fuel efficiency i.e. Highway MPG was increasing and decreasing year by year it was always between 24-22 Highway MPG from 1990-2007 but after 2007 fuel efficiency kept increasing .

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

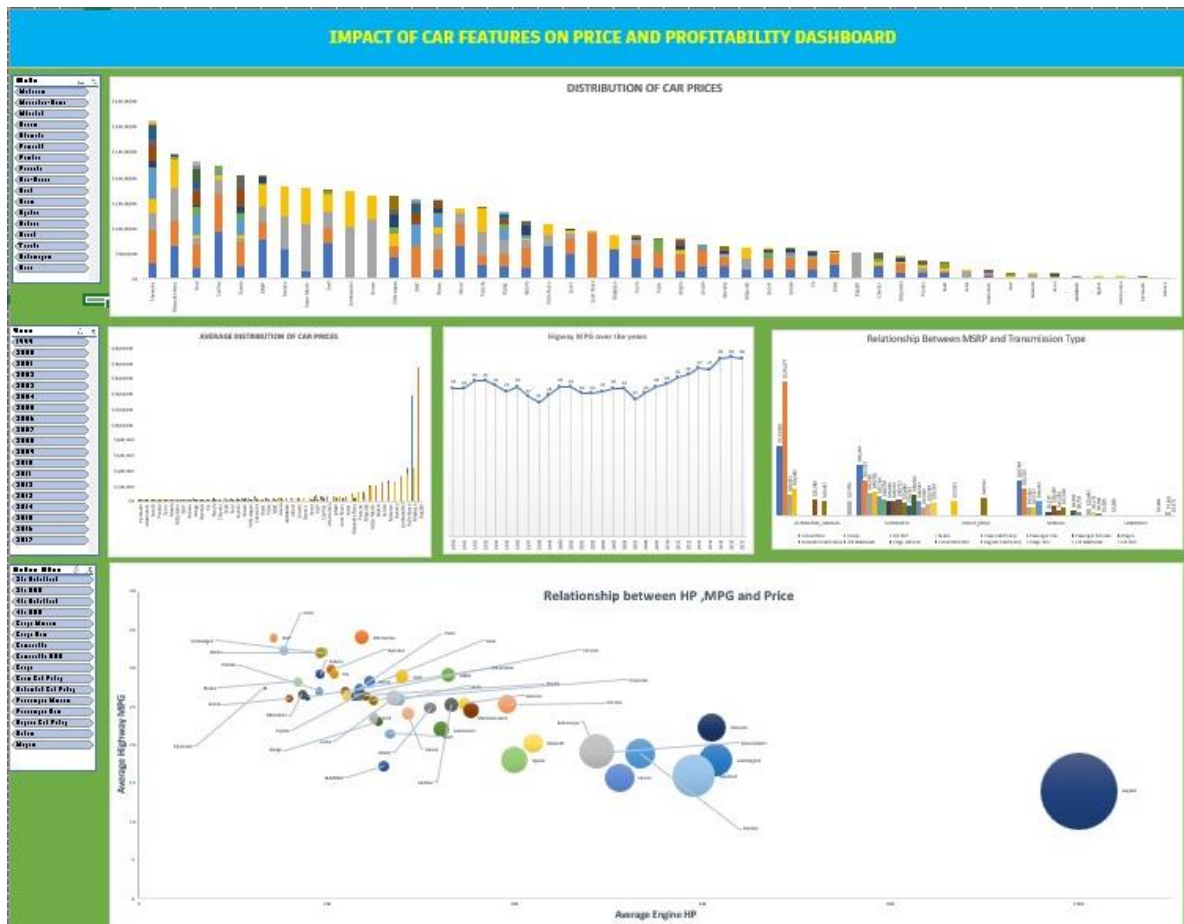
For this we create pivot table for Average Engine HP , MPG and MSRP for different brands

Brand	Average of Engine HP	Average of highway MPG	Average of MSRP
Acura	245	28	\$35,087
Alfa Romeo	237	34	\$61,600
Aston Martin	484	19	\$198,123
Audi	280	29	\$54,574
Bentley	534	19	\$247,169
BMW	330	29	\$62,163
Bugatti	1001	14	\$17,57,224
Buick	220	27	\$29,034
Cadillac	333	25	\$56,368
Chevrolet	250	26	\$29,000
Chrysler	229	26	\$26,723
Dodge	254	23	\$24,867
Ferrari	512	16	\$2,38,219
FIAT	144	34	\$22,206
Ford	250	24	\$28,523
Genesis	347	25	\$46,617
GMC	268	21	\$32,444
Honda	197	32	\$26,609
HUMMER	261	17	\$36,464
Hyundai	205	30	\$24,926
Infiniti	311	25	\$42,640
Kia	208	29	\$25,319
Lamborghini	614	18	\$3,31,567
Land Rover	323	22	\$68,067
Lexus	277	26	\$47,549
Lincoln	286	24	\$43,560
Lotus	272	26	\$68,377
Maserati	420	20	\$1,13,684
Maybach	591	16	\$5,46,222
Mazda	170	28	\$20,107
McLaren	610	22	\$2,39,805
Mercedes-Benz	354	25	\$72,135
Mitsubishi	174	27	\$21,316
Nissan	241	26	\$28,956
Oldsmobile	180	26	\$12,844
Plymouth	134	27	\$3,297
Pontiac	192	27	\$19,899
Porsche	393	25	\$1,01,622
Rolls-Royce	488	19	\$3,51,131
Saab	221	26	\$27,880
Scion	154	32	\$19,333
Spyker	400	18	\$2,14,990
Subaru	193	29	\$24,241
Suzuki	161	26	\$18,021
Toyota	234	26	\$28,759
Volkswagen	193	32	\$28,947
Volvo	235	27	\$29,725
Grand Total	254	26	\$41,928

To visualize this more efficiently we create a bubble chart where X-axis represents Avg Engine HP and Y-axis represents Avg Highway MPG for every brand and bubble size depends upon the average msrp of that brand .



From the above graph we understand that as the engine hp increases MPG decreases and price also increases . For eg Bugatti car brand has the highest engine hp but has the lowest MPG and highest avg msrp .



Dataset attached of data analysis :

https://drive.google.com/file/d/1Dlqi-suqGO4FWJVLb_Uh7Hd6vPx5XTCG/view?usp=sharing

Dataset attached of Dashboard :

<https://drive.google.com/file/d/1XbFz55KqDZfq8YOQx83hkbx95XfmBlp2/view?usp=sharing>

Links for loom video

<https://www.loom.com/share/7c974d04d630435b84513467c3faa9f> (Part 1)

<https://www.loom.com/share/636d105fa8364e3a8225141bb1901c88> (Part 2)

<https://www.loom.com/share/1d2bf1f964fa467592d355467fe54929> (Part 3)

<https://www.loom.com/share/c986c173f02d42e2bb1d1dfc265c5d5e> (Part 4)

Results :

The most prominent results from the above analysis are as follows:

- Hatchback, Flex Fuel , Crossover, Flex Fuel, Performance , Flex Fuel, Diesel are most popular market categories for car .
- There are more models of Crossover category in the market
- More the HP more is the price of the car
- Engine cylinders is most important features in determining the price of the car
- As the no of engine cylinders in a car increases its highway mpg decreases .

During this project I got experience to handle vast amounts of data and perform analysis and use various charts for visualization and derive results . I also got experience to design a basic interactive dashboard in excel which will help me to become a better data analyst in future .