The Impact of Car Features on Price and Profitability

BY
Anurag
Shrivas

DATA ANALYTICS REPORT

Project Description:

a) **Overview of the project and its purpose:** 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.

b) **Business problem or question that the project aims to address: i)** Which car features are most important in determining a car's price?
ii) How does the car's horsepower, MPG, and price vary across different Brands?

c) **Description of the data sources used in the project:** 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.

d) Description of the data cleaning and preprocessing steps performed on the data:
Basically all data is cleaned but Market category column has 3742 null values so I delete those rows because approx. 31 percentage of the data is missing.

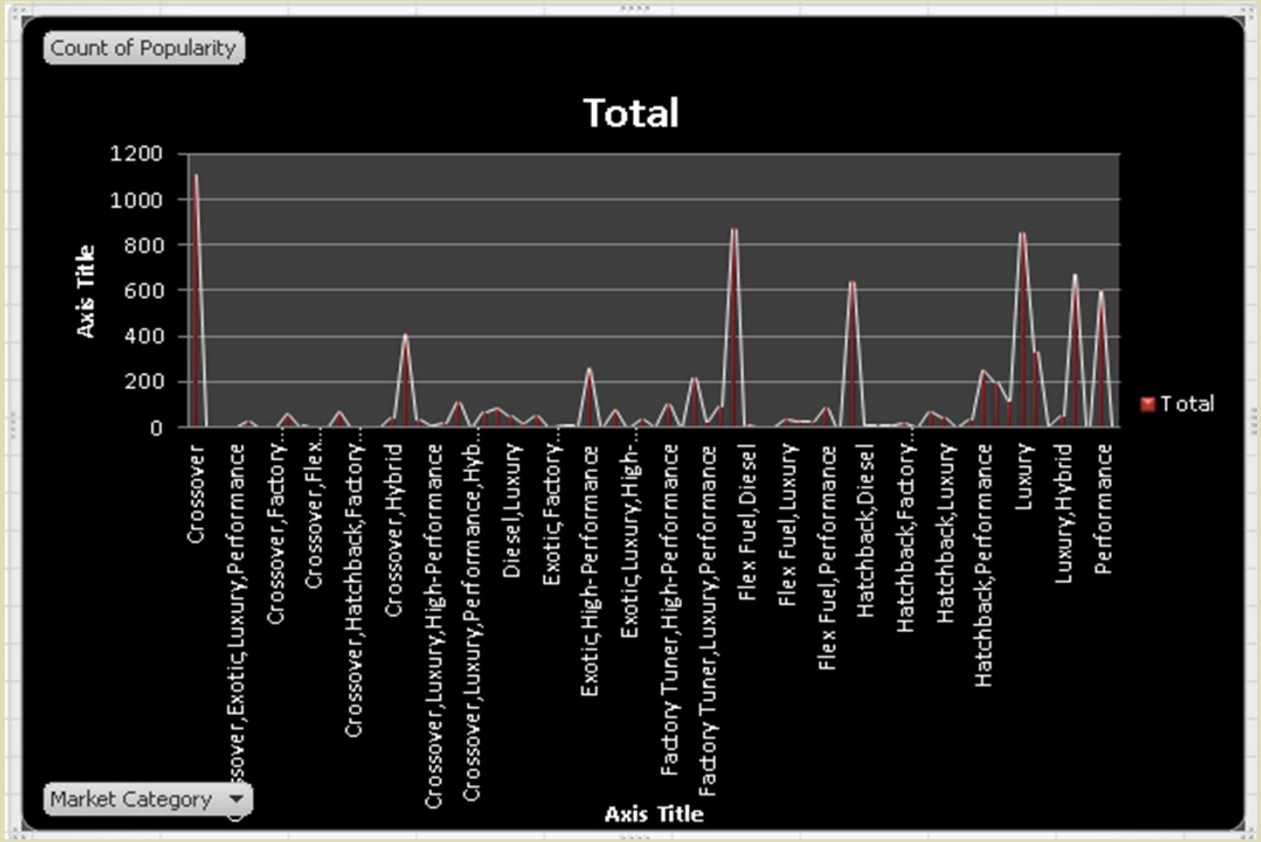
e) Any assumptions made during the project: Basically we know about car price vary across the car company & market category.

- **Insights required:** How does the popularity of a car model vary across different market categories?
- **Task1A:** Create a pivot table that shows the number of car models in each market category and their corresponding popularity scores.
- **Solution:**

Row Labels	Count of Model	Sum of Popularity
Crossover	1110	1715242
Crossover,Diesel	7	6111
Crossover,Exotic,Luxury,High-Performance	1	238
Crossover,Exotic,Luxury,Performance	1	238
Crossover,Factory Tuner,Luxury,High-Performance	26	47410
Crossover,Factory Tuner,Luxury,Performance	5	13037
Crossover,Factory Tuner,Performance	4	840
Crossover,Flex Fuel	64	132720
Crossover,Flex Fuel,Luxury	10	11732
Crossover,Flex Fuel,Luxury,Performance	6	9744
Crossover,Flex Fuel,Performance	6	33942
Crossover,Hatchback	72	120650
Crossover,Hatchback,Factory Tuner,Performance	6	12054
Crossover,Hatchback,Luxury	7	1428
Crossover,Hatchback,Performance	6	12054
Crossover,Hybrid	42	107662
Crossover,Luxury	410	362665
Crossover,Luxury,Diesel	34	73080
Crossover,Luxury,High-Performance	9	9335
Crossover,Luxury,Hybrid	24	15142

- **Insight:** Most popular and highest selling model is **CROSSOVER**

Task1.B- Create a combo chart that visualizes the relationship between market category and popularity

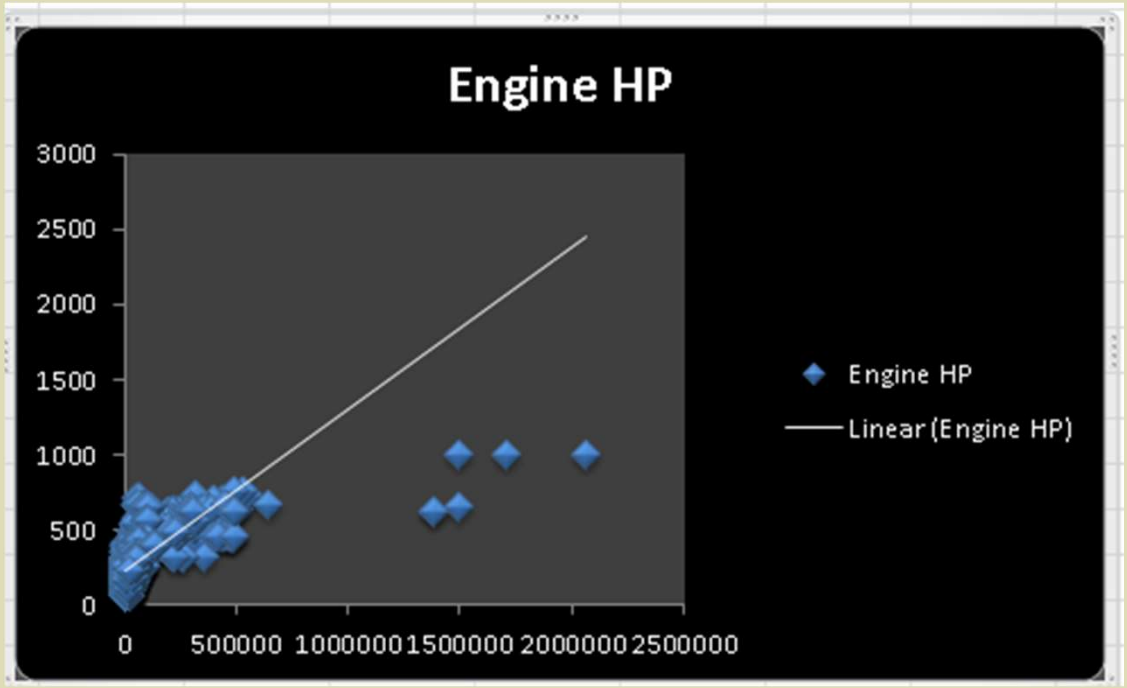


Insight: Most of the market covered by Crossover

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.

Solution:



Insight: Most of car engine are 0-800hp which car price are between 0-600000. Trendline shows engine hp and price shows linear trend so we can say that as well as Engine Hp increases as well as price be 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.

Solution: Here I provide the Jupyter notebook link where I show regression analysis.

Notebook Link:

https://drive.google.com/file/d/1zyi5UTceqIqQhxwTrc37T5LRlITREoV9/view?usp=sharing

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.

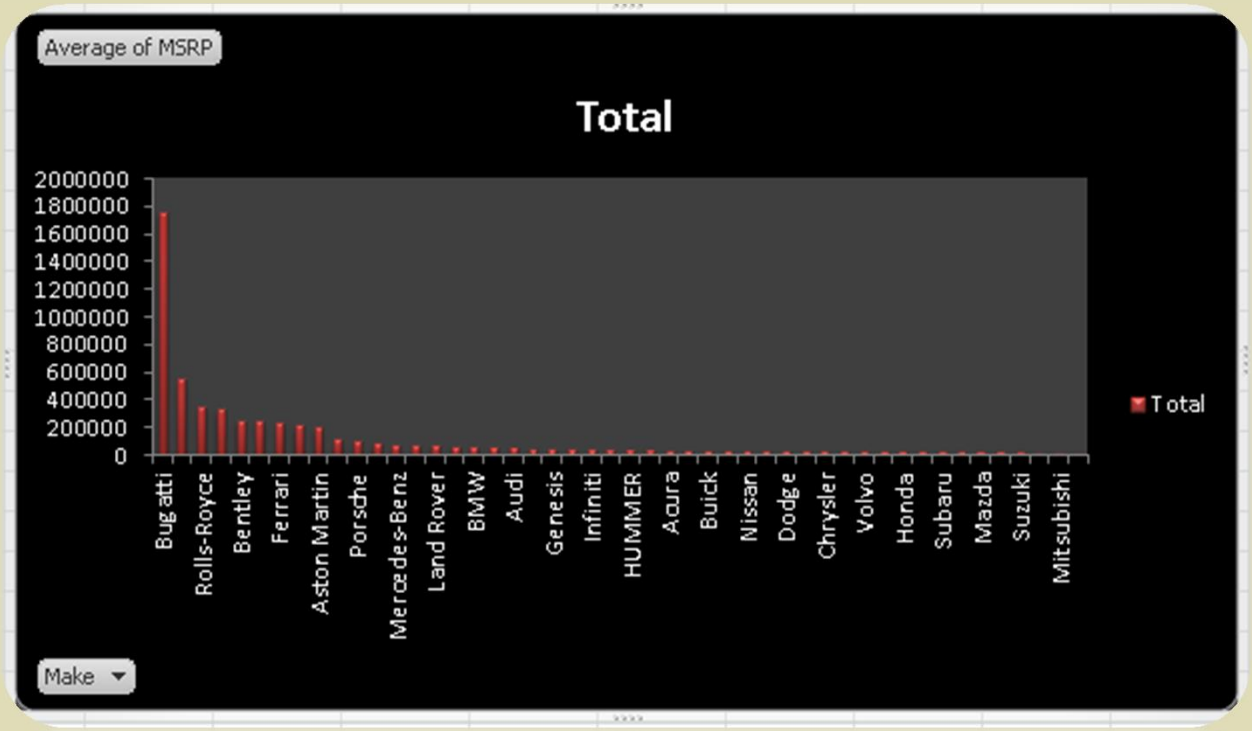
Solution:

Row Labels	Average of MSRP
Acura	34887.5873
Alfa Romeo	61600
Aston Martin	197910.3763
Audi	53452.1128
Bentley	247169.3243
BMW	61546.76347
Bugatti	1757223.667
Buick	28206.61224
Cadillac	56231.31738
Chevrolet	28350.38557
Chrysler	26722.96257
Dodge	22390.05911
Ferrari	238218.8406
FIAT	22670.24194
Ford	27399.26674
Genesis	46616.66667
GMC	30493.29903
Honda	26674.34076
HUMMER	36464.41176
Hyundai	24597.0363
Infiniti	42394.21212
Kia	25310.17316
Lamborghini	331567.3077
Land Rover	67823.21678
Lexus	47549.06931
Lincoln	42839.82927

Insights: This can help to understand which manufacturer is highest average price and who is selling most expensive car.

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

Solution:

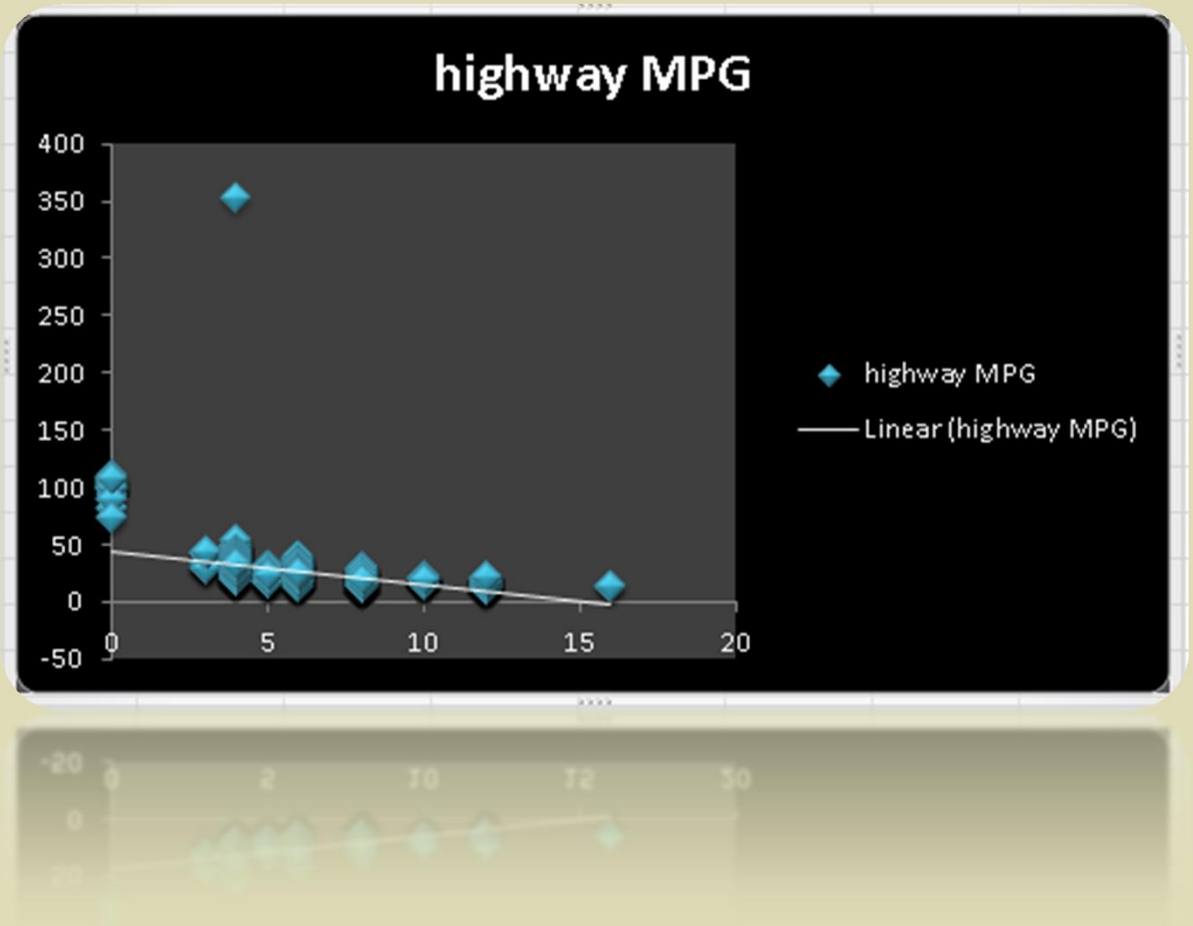


Insight of 4.A and 4.B: Average price of all manufacturer is 50042. Highest average price is 17,57,223 (**BUGATTI)** and the lowest average price is 20,395 (**SCION**)

Insight Required: What is the relationship between fuel efficiency and the number of cylinders in a car's engine?

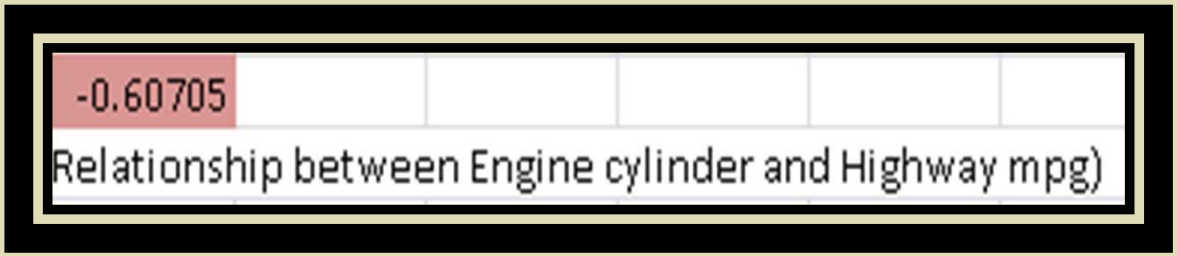
Task 5.A: Create a scatter plot with the number of cylinders on the x-axis and highway MPG on the y-axis. Then create a trendline on the scatter plot to visually estimate the slope of the relationship and assess its significance.

Solution:



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

Solution:



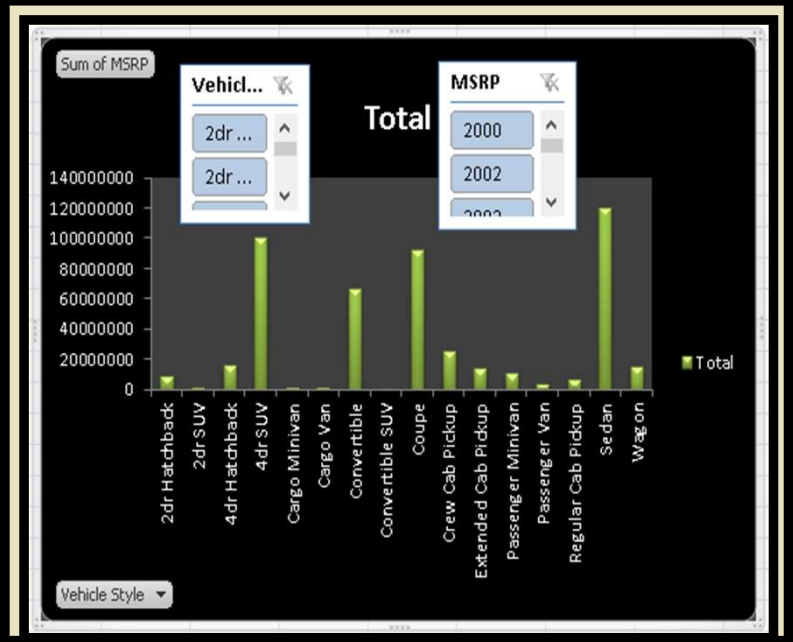
Insight of 5.A and 5.B: Highway MPG and number of cylinders are negatively correlated . As we saw in the chart number of Cylinder is increases at the same time highway MPG is going down.

BUILDING DASHBOARD

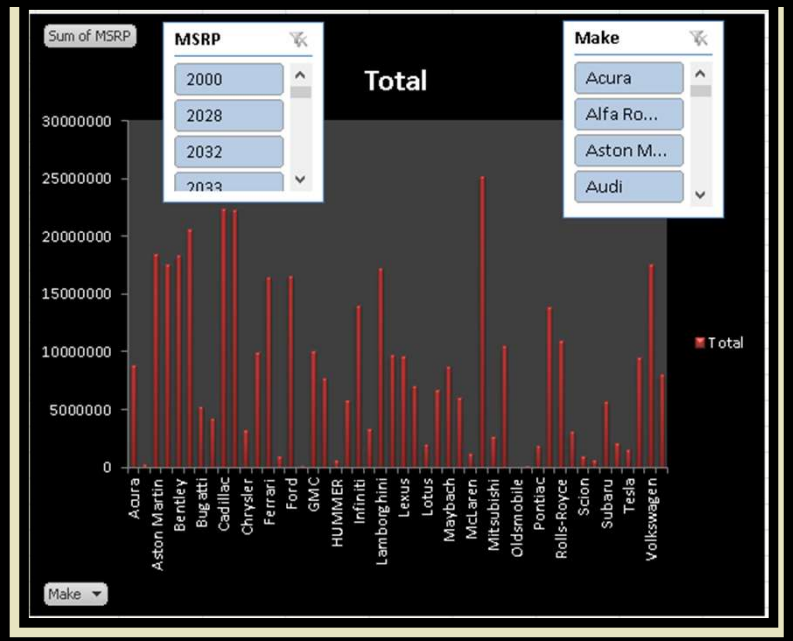
Task 1: How does the distribution of car prices vary by brand and body style?

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.

Solution:



Row Labels	Sum of MSRP
2dr Hatchback	8535063
2dr SUV	489179
4dr Hatchback	15739448
4dr SUV	89097172
Cargo Minivan	114820
Cargo Van	1044640
Convertible	63961758
Convertible SUV	276806
Coupe	89180376
Crew Cab Pickup	13427203
Extended Cab Pickup	5895574
Passenger Minivan	1718406
Passenger Van	3596315
Regular Cab Pickup	4783639
Sedan	99848014
Wagon	11241736
Grand Total	**408950149**

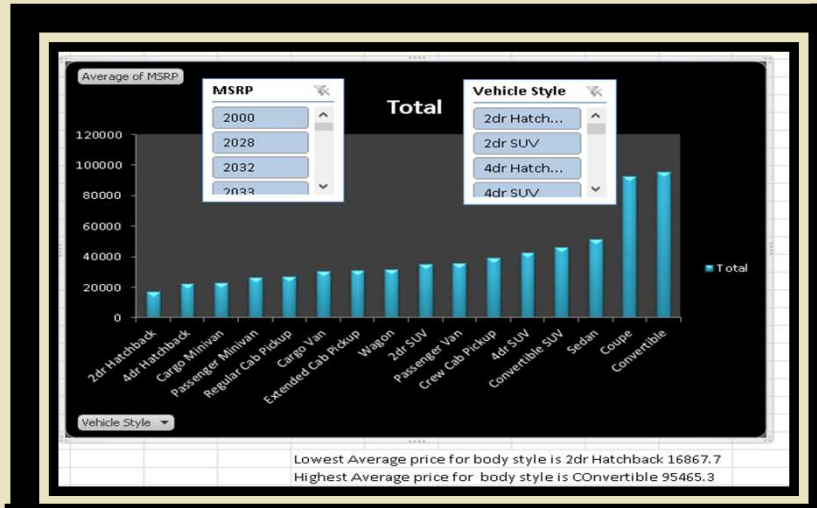


Row Labels	Sum of MSRP
Acura	8791672
Alfa Romeo	308000
Aston Martin	18405665
Audi	17532293
Bentley	18290530
BMW	20556619
Bugatti	5271671
Buick	4221300
Cadillac	22323833
Chevrolet	22314321
Chrysler	3237718
Dodge	9918521
Ferrari	16437100
FIAT	945485
Ford	16556084
Genesis	139850
GMC	10094152
Honda	7675093
HUMMER	619895
Hyundai	5774995
Infiniti	13990090
Kia	3286275

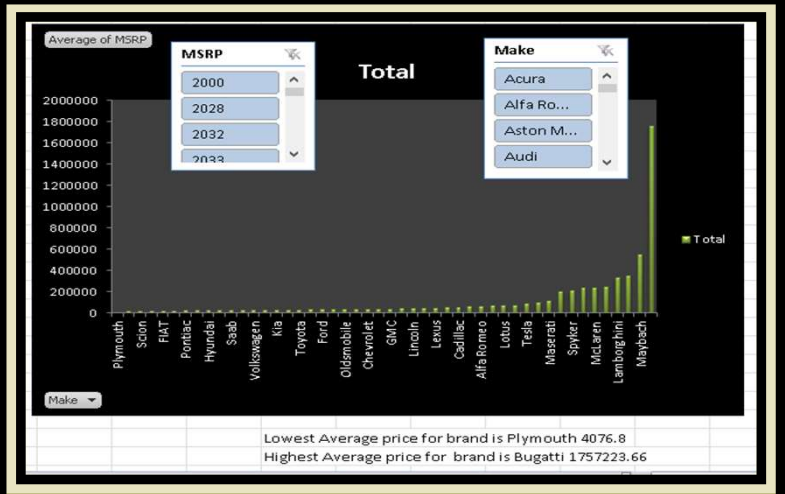
Task 2: Which car brands have the highest and lowest average MSRPs, and how does this vary by body style?

Hints: Clustered column chart to compare the average MSRPs across different car brands and body styles. Calculate the average MSRP for each brand and body style using AVERAGEIF or Pivot Tables.

Solution:



Row Labels	Average of MSRP
2dr Hatchback	16867.71344
4dr Hatchback	22420.8661
Cargo Minivan	22964
Passenger Minivan	26437.01538
Regular Cab Pickup	27179.76705
Cargo Van	30724.70588
Extended Cab Pickup	30866.87958
Wagon	31489.45658
2dr SUV	34941.35714
Passenger Van	35963.15
Crew Cab Pickup	39032.56686
4dr SUV	42407.03094
Convertible SUV	46134.33333
Sedan	51441.5322
Coupe	92896.225
Convertible	95465.31045
Grand Total	**50042.84741**

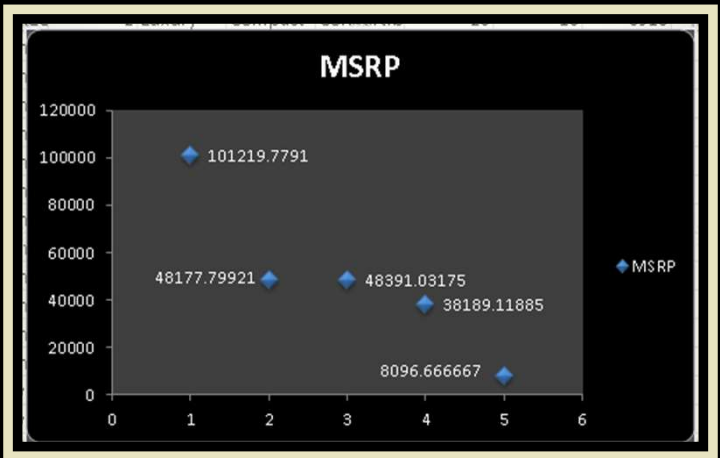


Row Labels	Average of MSRP
Acura	34887.5873
Alfa Romeo	61600
Aston Martin	197910.3763
Audi	53452.1128
Bentley	247169.3243
BMW	61546.76347
Bugatti	1757223.667
Buick	28206.61224
Cadillac	56231.31738
Chevrolet	28350.38557
Chrysler	26722.96257
Dodge	22390.05911
Ferrari	238218.8406
FIAT	22670.24194
Ford	27399.26674

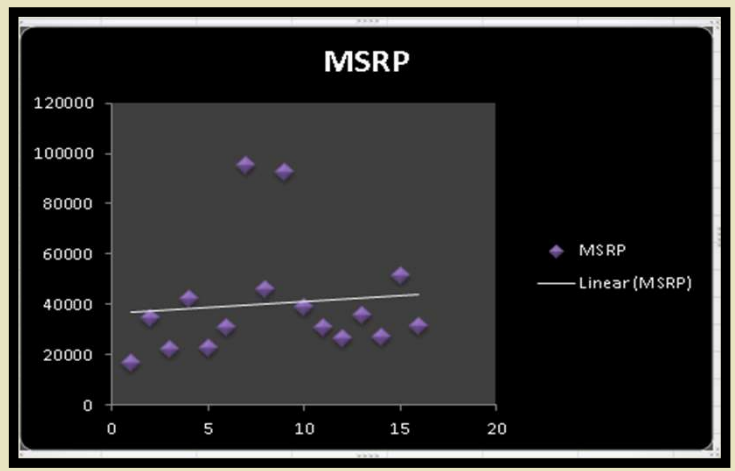
Task 3: How do the different feature such as transmission type affect the MSRP, and how does this vary by body style?

Hints: Scatter plot chart to visualize the relationship between MSRP and transmission type, with different symbols for each body style. Calculate the average MSRP for each combination of transmission type and body style using AVERAGEIFS or Pivot Tables.

Solution:



Row Labels	Average of MSRP
AUTOMATED_MANUAL	101219.7791
AUTOMATIC	48177.79921
DIRECT_DRIVE	48391.03175
MANUAL	38189.11885
UNKNOWN	8096.666667
Grand Total	50042.84741

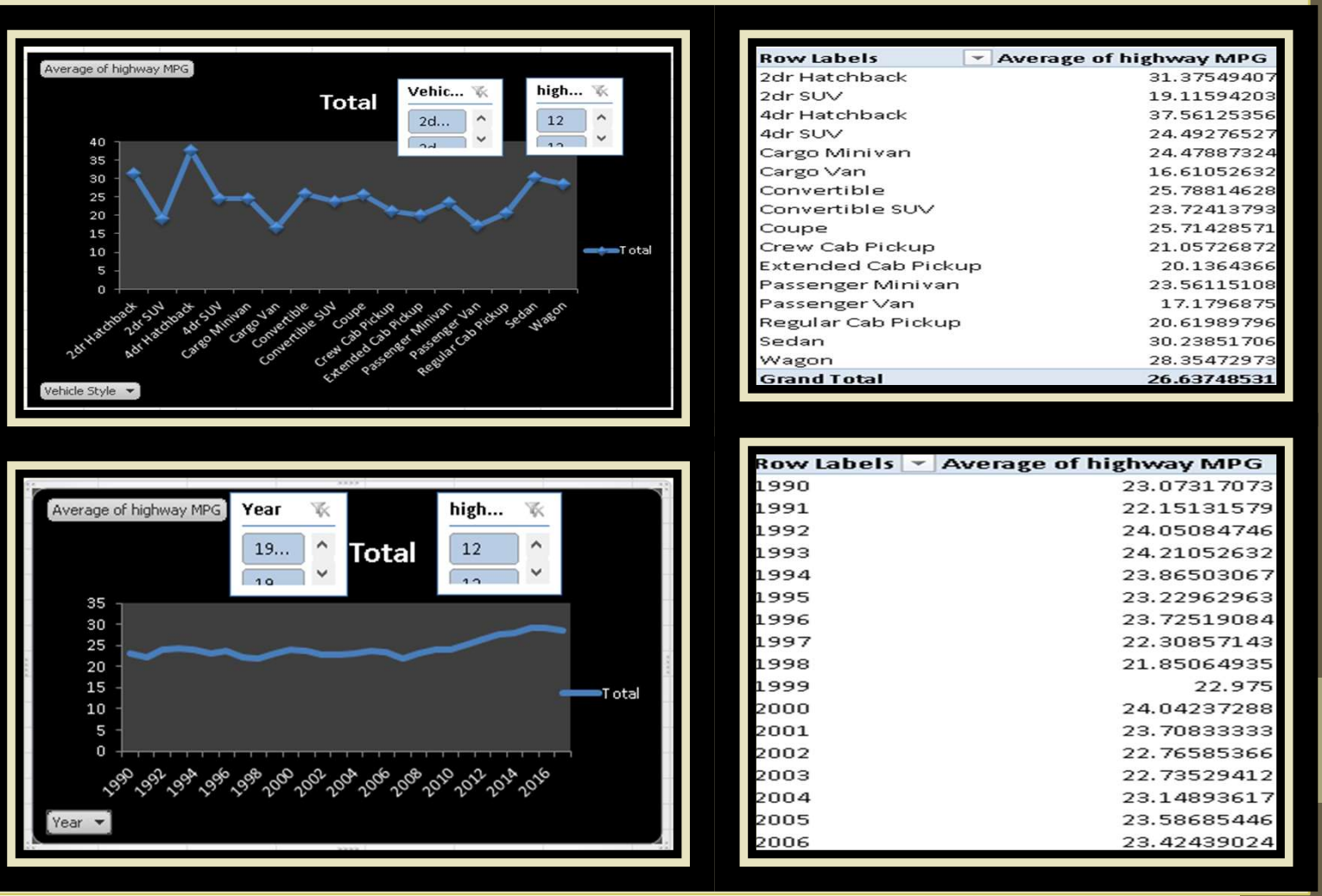


Row Labels	Average of MSRP
2dr Hatchback	16867.71344
2dr SUV	34941.35714
4dr Hatchback	22420.8661
4dr SUV	42407.03094
Cargo Minivan	22964
Cargo Van	30724.70588
Convertible	95465.31045
Convertible SUV	46134.33333
Coupe	92896.225
Crew Cab Pickup	39032.56686
Extended Cab Pickup	30866.87958
Passenger Minivan	26437.01538
Passenger Van	35963.15
Regular Cab Pickup	27179.76705
Sedan	51441.5322
Wagon	31489.45658
Grand Total	50042.84741

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

Hints: Line chart to show the trend of fuel efficiency (MPG) over time for each body style. Calculate the average MPG for each combination of body style and model year using AVERAGEIFS or Pivot Tables.

Solution:



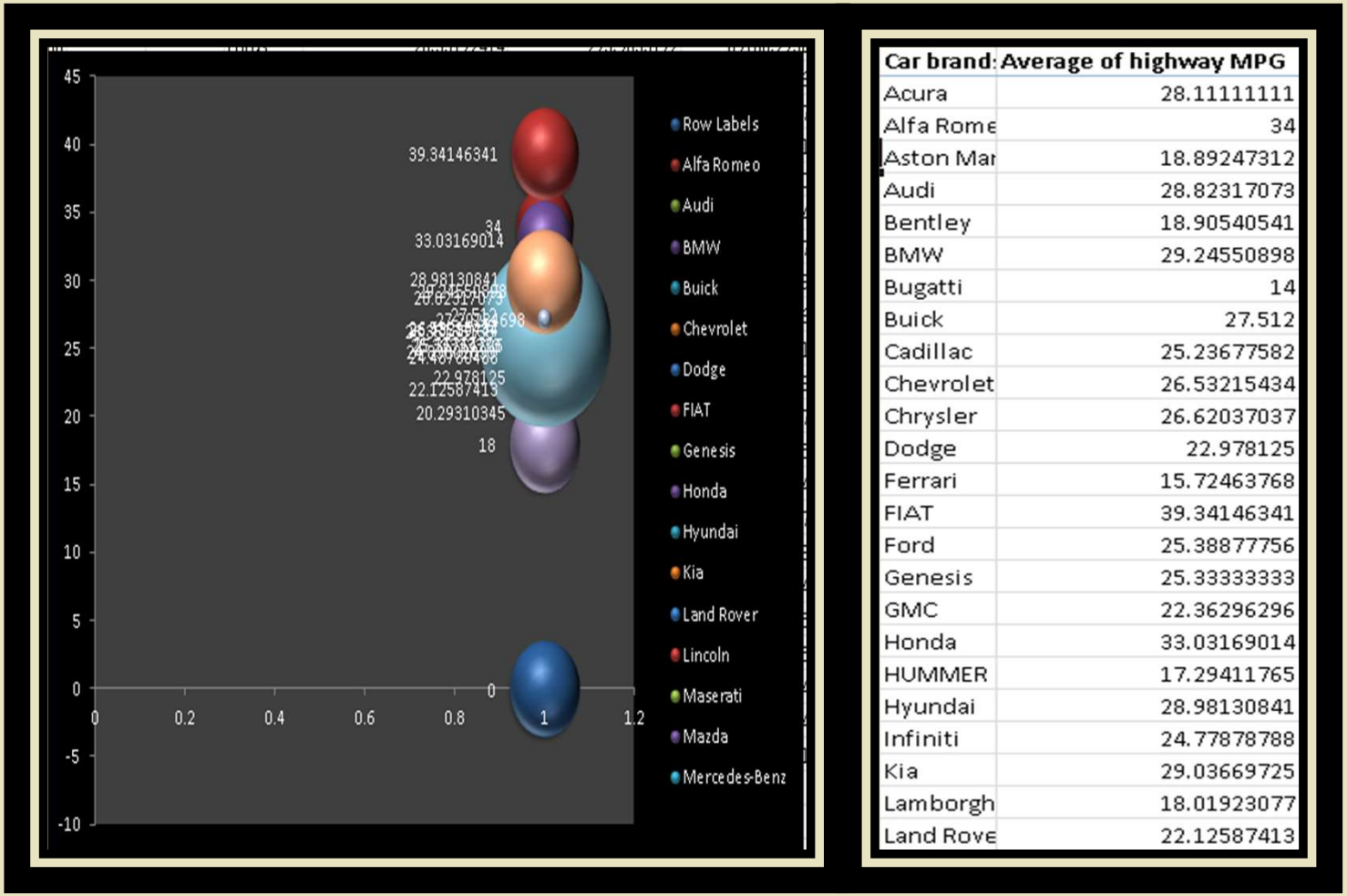
Row Labels	Average of highway MPG
2dr Hatchback	31.37549407
2dr SUV	19.11594203
4dr Hatchback	37.56125356
4dr SUV	24.49276527
Cargo Minivan	24.47887324
Cargo Van	16.61052632
Convertible	25.78814628
Convertible SUV	23.72413793
Coupe	25.71428571
Crew Cab Pickup	21.05726872
Extended Cab Pickup	20.1364366
Passenger Minivan	23.56115108
Passenger Van	17.1796875
Regular Cab Pickup	20.61989796
Sedan	30.23851706
Wagon	28.35472973
Grand Total	**26.63748531**

Row Labels	Average of highway MPG
1990	23.07317073
1991	22.15131579
1992	24.05084746
1993	24.21052632
1994	23.86503067
1995	23.22962963
1996	23.72519084
1997	22.30857143
1998	21.85064935
1999	22.975
2000	24.04237288
2001	23.70833333
2002	22.76585366
2003	22.73529412
2004	23.14893617
2005	23.58685446
2006	23.42439024

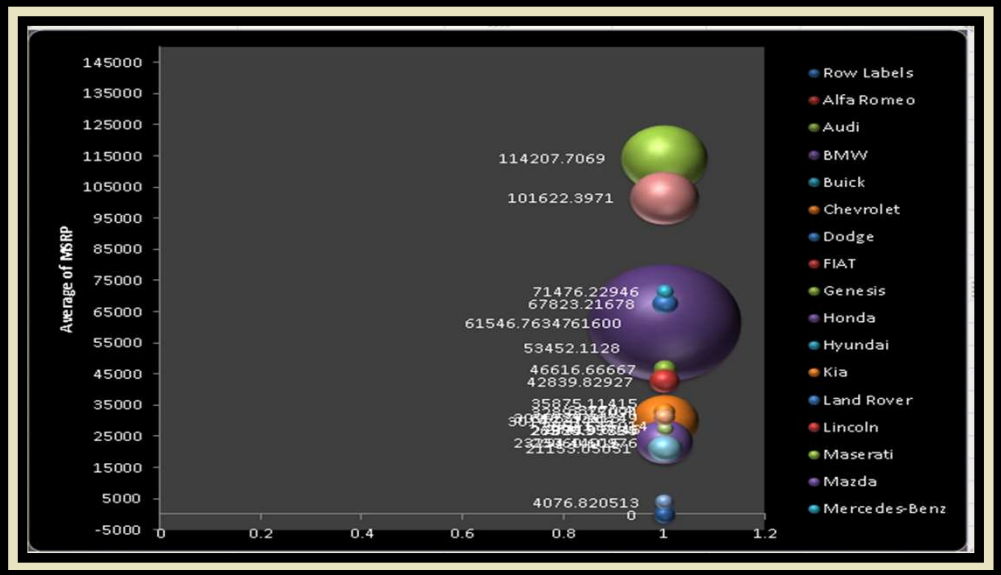
Task 5: How does the car's horsepower, MPG, and price vary across different Brands?

Hints: Bubble chart to visualize the relationship between horsepower, MPG, and price across different car brands. Assign different colors to each brand and label the bubbles with the car model name. Calculate the average horsepower, MPG, and MSRP for each car brand using AVERAGEIFS or Pivot Tables.

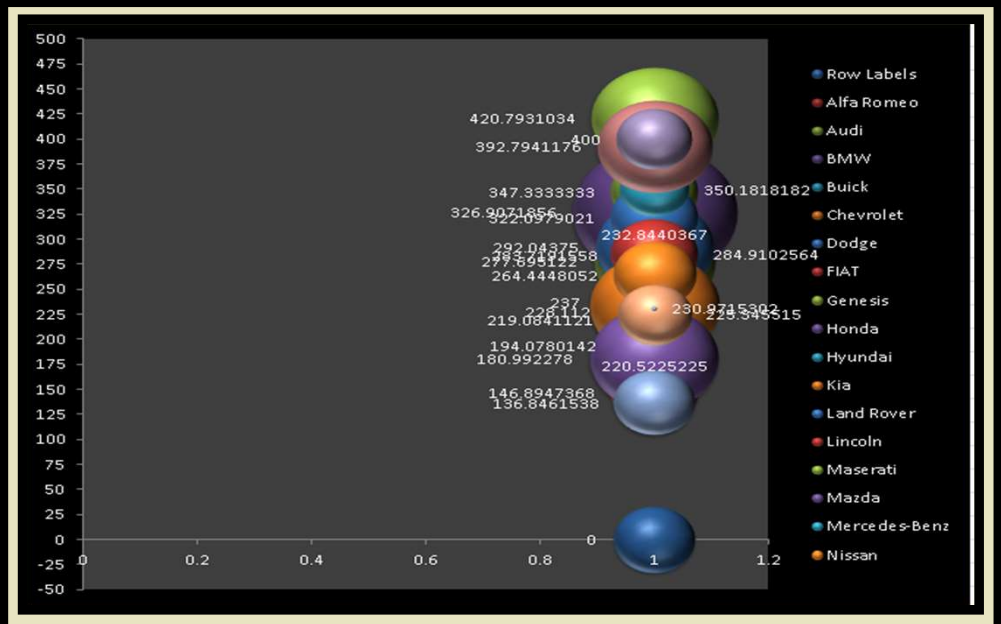
Solution:



Car brand	Average of highway MPG
Acura	28.11111111
Alfa Rome	34
Aston Mar	18.89247312
Audi	28.82317073
Bentley	18.90540541
BMW	29.24550898
Bugatti	14
Buick	27.512
Cadillac	25.23677582
Chevrolet	26.53215434
Chrysler	26.62037037
Dodge	22.978125
Ferrari	15.72463768
FIAT	39.34146341
Ford	25.38877756
Genesis	25.33333333
GMC	22.36296296
Honda	33.03169014
HUMMER	17.29411765
Hyundai	28.98130841
Infiniti	24.77878788
Kia	29.03669725
Lamborgh	18.01923077
Land Rove	22.12587413



Row Label	Average of MSRP
Acura	34887.5873
Alfa Rome	61600
Aston Mar	197910.3763
Audi	53452.1128
Bentley	247169.3243
BMW	61546.76347
Bugatti	1757223.667
Buick	33770.4
Cadillac	56231.31738
Chevrolet	35875.11415
Chrysler	29978.87037
Dodge	30995.37813
Ferrari	238218.8406
FIAT	23060.60976
Ford	33178.52505
Genesis	46616.66667
GMC	37385.74815
Honda	27024.97535
HUMMER	36464.41176
Hyundai	26985.95794
Infiniti	42394.21212
Kia	30149.31193
Lamborgh	331567.3077
Land Rove	67823.21678



Car brand	Average of Engine HP
Acura	244.797619
Alfa Rome	237
Aston Mar	484.3225806
Audi	277.695122
Bentley	533.8513514
BMW	326.9071856
Bugatti	1001
Buick	228.112
Cadillac	332.3098237
Chevrolet	283.7191558
Chrysler	241.3888889
Dodge	292.04375
Ferrari	511.9565217
FIAT	146.8947368
Ford	273.6341463
Genesis	347.3333333
GMC	279.9851852
Honda	194.0780142
HUMMER	261.2352941
Hyundai	219.0841121
Infiniti	310.0666667
Kia	232.8440367
Lamborgh	614.0769231
Land Rove	322.0979021

APPROACH & TECH-STACK USED

a) **Description of the analytical methods used in the project, such as descriptive statistics, visualization, machine learning, or optimization:** First download the dataset and find median use the formula AVERAGE, I used pivot table to make a chart like bar chart, column chart, Line chart, scatter chart. I used linear regression for find the those feature that affect price.

b) **Explanation of the reasoning behind the choice of analytical methods:** I used both of the feature numerical and categorical . I used pivot table for more of the analysis because the use of pivot table we find easily average ,sum, count values in fraction of second and easily create a chart .

c) **Description of the modeling techniques used, if applicable**: I used Linear regression for creating a model .

d) **Any challenges or limitations encountered during the project:** When I create a bubble chart that time I face some of the problems like column has much more value in it if data has less values that time we create a beautiful bubble chart .

Tech-Stack Used:

I used Microsoft excel for all of the analysis because it has so much helpful things to easily analyse the data. Some of tools that are so much helpful when we analyse the data like: Pivot table, chart, Sort & filter, table.

Insight

I. If we find which brand and model are good in market this can help to understand which type of car people like .

II. If we find average price of the car for each brand this can help to understand which price is good for highest selling.

III. If we find which car brand and model are popular this can help to understand which car is more popular in the market.

IV. If we find relationship between engine hp and average price this help to understand how much affect horse power to the price.

V. If we find which type of vehicle like : Cube, Convertible ,Sedan , Wagon are selling good.

VI. If we find average price of year wise that can help to understand how price vary across the year.

VII. If we find relationship between market category and popularity that help to understand which market category is more popular .

VIII. If we find which brand and body style is good for price . So this can help to company to easily tag the price to any car.

VIII) If we understand the trend between Highway mpg and number of cylinder that can help to understand how much average when the car riding on highway.

IX) If we find which brand and body style is good for price . So this can help to company to easily tag the price to any car.

X) If we find which transmission type affect the price this can help to understand which type of transmission is good for car . Car company easily find the trend and use this as advantage of the company.

XI) If we see highway mpg changes according the year that help to understand which year is highway mpg is good.

XII) Some of the features are important for decide the price of the car like: Make, vehicle style, transmission type, Engine hp, Highway mpg.

Result:

a) Average Engine hp for all of the brand is 274.411.

b) Average Highway mpg is 27.02 for all of the brand.

c) **Bugatti** has the highest average price of all of the car brand.

d) **Sedan** is highest sum of price car all over the **BODY STYLE.**

e) **CONVERTIBLE** is a highest average price.

f) **Automated _Manual** is the highest average price in all transmission type.

g) **2010-2017** the average mpg is increasing.

Link of analysis file:-
https://drive.google.com/file/d/1HnR4IFdWtAFOlV5mU9coAaKUi9NKoysm/view?usp=sharing