

Analyzing the Impact of Car Features on Price and Profitability

Final Project-3



SOWMIYA R | DATA ANALYTICS | 10TH AUGUST 2023

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.

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 project aims to highlight the transformative potential of car data analysis in revolutionizing the automotive industry. By demonstrating tangible benefits such as reduced maintenance costs, improved safety, and enhanced user experiences, the project seeks to encourage wider adoption of data-driven practices within the automotive ecosystem and beyond.

In today's automotive landscape, vehicles are equipped with an array of sensors, processors, and communication technologies that generate vast amounts of data. This data, if harnessed effectively, holds the potential to revolutionize multiple aspects of the automotive industry. However, to fully realize its benefits, there is a pressing need to understand and address several critical challenges.

Data sources is been used in this are the Microsoft excel and the visualizations of making visual insights using Tableau. And the cleaning technique is done by using the Excel functions.

DATA CLEANING

First of all, I have downloaded the dataset which was given in csv format , then I load the dataset in the excel and done with few auto-fit column and making it clearance . Then I used my Data cleaning strategy to clean the provided dataset with the help of normal excel functions that is I have count the number of blank rows and then removed by using the special function in excel.

And then I have analyzed that the unwanted headers are present there, so I inferred with the specified and useful headers will be left behind. So firstly I found that 11915 rows and 16 headers are present there. So after cleaning the dataset I left with a 11813 rows and 17 header rows including index created headers will be left there.

Makes it as table format and highlighted the header rows , and analyzed throughout the dataset , there inferred impact of car data features with the model of the manufacturer , vehicle type , Cylinder HP, the market price for each one of the model of the position of car, when they produced that is mentioned as year , MPG that is miles per gallons with the engine fuel, number of doors and so on.

I have attached my cleaned data excel file below:

[https://docs.google.com/spreadsheets/d/1nC5QFm0UQspbJFhFP0f2LtrX-uMF1WQf/edit?usp=drive link&ouid=103768596710140113695&rt pof=true&sd=true](https://docs.google.com/spreadsheets/d/1nC5QFm0UQspbJFhFP0f2LtrX-uMF1WQf/edit?usp=drive_link&ouid=103768596710140113695&rt=pof=true&sd=true)

F1 Engine HP										
	A	B	C	D	E	F	G	H	I	J
	Index	Make	Model	Year	Engine Fuel Type	Engine HP	Engine Cylinders	Transmission Type	Driven Wheel	Number of Doors
1	1	BMW	1 Series M	2011	premium unleaded (required)	335	6	MANUAL	rear wheel drive	2
2	2	BMW	1 Series	2011	premium unleaded (required)	300	6	MANUAL	rear wheel drive	2
3	3	BMW	1 Series	2011	premium unleaded (required)	300	6	MANUAL	rear wheel drive	2
4	4	BMW	1 Series	2011	premium unleaded (required)	230	6	MANUAL	rear wheel drive	2
5	5	BMW	1 Series	2011	premium unleaded (required)	230	6	MANUAL	rear wheel drive	2
6	6	BMW	1 Series	2012	premium unleaded (required)	230	6	MANUAL	rear wheel drive	2
7	7	BMW	1 Series	2012	premium unleaded (required)	300	6	MANUAL	rear wheel drive	2
8	8	BMW	1 Series	2012	premium unleaded (required)	300	6	MANUAL	rear wheel drive	2
9	9	BMW	1 Series	2012	premium unleaded (required)	230	6	MANUAL	rear wheel drive	2
10	10	BMW	1 Series	2013	premium unleaded (required)	230	6	MANUAL	rear wheel drive	2
11	11	BMW	1 Series	2013	premium unleaded (required)	300	6	MANUAL	rear wheel drive	2
12	12	BMW	1 Series	2013	premium unleaded (required)	230	6	MANUAL	rear wheel drive	2
13	13	BMW	1 Series	2013	premium unleaded (required)	300	6	MANUAL	rear wheel drive	2
14	14	BMW	1 Series	2013	premium unleaded (required)	230	6	MANUAL	rear wheel drive	2
15	15	BMW	1 Series	2013	premium unleaded (required)	230	6	MANUAL	rear wheel drive	2
16	16	BMW	1 Series	2013	premium unleaded (required)	320	6	MANUAL	rear wheel drive	2
17	17	BMW	1 Series	2013	premium unleaded (required)	320	6	MANUAL	rear wheel drive	2
18	18	Audi	100	1992	regular unleaded	172	6	MANUAL	front wheel drive	4
19	19	Audi	100	1992	regular unleaded	172	6	MANUAL	front wheel drive	4
20	20	Audi	100	1992	regular unleaded	172	6	AUTOMATIC	all wheel drive	4
21	21	Audi	100	1992	regular unleaded	172	6	MANUAL	front wheel drive	4
22	22	Audi	100	1992	regular unleaded	172	6	MANUAL	all wheel drive	4
23	23	Audi	100	1993	regular unleaded	172	6	MANUAL	front wheel drive	4
24	24	Audi	100	1993	regular unleaded	172	6	AUTOMATIC	all wheel drive	4
25	25	Audi	100	1993	regular unleaded	172	6	MANUAL	front wheel drive	4
26	26	Audi	100	1993	regular unleaded	172	6	MANUAL	front wheel drive	4
27	27	Audi	100	1993	regular unleaded	172	6	MANUAL	all wheel drive	4

From the above picture , can see a image of cleaned data and the excel file have been attached above the picture.

APPROACH

Perform statistical analysis to understand the distribution, variability, and basic characteristics of the collected data. Visualize data patterns, correlations, and anomalies using graphs and plots. Identify initial insights that can guide subsequent analysis. Evaluate model performance using metrics like accuracy, precision, recall, and F1-score. Showcase how predictive models can optimize maintenance schedules and reduce downtime.

Demonstrate how real-time alerts and warnings can be generated for drivers and relevant authorities. Evaluate the potential reduction in accidents and overall enhancement in road safety. Analyze driving patterns and vehicle performance data to uncover factors influencing fuel efficiency. Develop a model to predict fuel efficiency based on driving behavior, road conditions, and vehicle characteristics. Provide actionable recommendations to drivers for improving fuel economy. Quantify potential fuel savings and emissions reduction through optimized driving.

The project aims to provide a comprehensive understanding of the impact of car data analysis on the automotive industry, offering insights, solutions, and guidance for industry stakeholders to harness the power of data for positive transformations.

TECH-STACK USED

I have used Microsoft Excel and Tableau for analyzing the given tasks and gave various insights by using the Tableau Public. Excel is used in various places to make the table as pivot tables and analyzing through the conditional manner. And the Tableau is used for making bar charts , bubble charts and so on and also useful for making dashboards. And Excel is been used for doing regression analysis with the table format of making it clear with one who see this.



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: 11,159

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

These are the ones which were described for the given dataset and then explained each one of the dataset header rows provided in the form of csv file data.

This dataset could be useful for a variety of data analysis tasks, such as:

- Exploring trends in car features and pricing over time
- Comparing the fuel efficiency of different types of cars
- Investigating the relationship between a car's features and its popularity
- Predicting the price of a car based on its features and market category

TASKS: ANALYSIS

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?

BUILDING THE DASHBOARD:

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

2. Which car brands have the highest and lowest average MSRPs, and how does this vary by body style?
3. How do the different feature such as transmission type affect the MSRP, and how does this vary by body style?
4. How does the fuel efficiency of cars vary across different body styles and model years?
5. How does the car's horsepower, MPG, and price vary across different Brands?

INSIGHTS AND RESULTS:

1. **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.

The solution for the task 1 of A is that of created a pivot table with the help of showing the number of car models in each market category with respect to the popularity scores which is of creating a make or manufacturers with the popularity scores of maintaining the market category of impact of car trends.

Popularity of a car model vary across different market categories		
Market Category	Values	
	Average of Popularity	Count of Model
Crossover	1545	1110
Crossover,Diesel	873	7
Crossover,Exotic,Luxury,High-Performance	238	1
Crossover,Exotic,Luxury,Performance	238	1
Crossover,Factory Tuner,Luxury,High-Performance	1823	26
Crossover,Factory Tuner,Luxury,Performance	2607	5
Crossover,Factory Tuner,Performance	210	4
Crossover,Flex Fuel	2074	64
Crossover,Flex Fuel,Luxury	1173	10
Crossover,Flex Fuel,Luxury,Performance	1624	6
Crossover,Flex Fuel,Performance	5657	6
Crossover,Hatchback	1676	72
Crossover,Hatchback,Factory Tuner,Performance	2009	6
Crossover,Hatchback,Luxury	204	7
Crossover,Hatchback,Performance	2009	6
Crossover,Hybrid	2563	42
Crossover,Luxury	885	410
Crossover,Luxury,Diesel	2149	34
Crossover,Luxury,High-Performance	1037	9
Crossover,Luxury,Hybrid	631	24

Crossover,Luxury,Performance	1345	113
Crossover,Luxury,Performance,Hybrid	3916	2
Crossover,Performance	2586	69
Diesel	1731	84
Diesel,Luxury	2275	51
Exotic,Factory Tuner,High-Performance	1046	21
Exotic,Factory Tuner,Luxury,High-Performance	518	52
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	1271	261
Exotic,Luxury	113	12
Exotic,Luxury,High-Performance	467	79
Exotic,Luxury,High-Performance,Hybrid	204	1
Exotic,Luxury,Performance	217	36
Exotic,Performance	1391	10
Factory Tuner,High-Performance	1941	106
Factory Tuner,Luxury	617	2
Factory Tuner,Luxury,High-Performance	2133	215
Factory Tuner,Luxury,Performance	1413	31
Factory Tuner,Performance	1696	92
Flex Fuel	2217	872
Flex Fuel,Diesel	5657	16
Flex Fuel,Factory Tuner,Luxury,High-Performance	258	1
Flex Fuel,Hybrid	155	2
Flex Fuel,Luxury	747	39
Flex Fuel,Luxury,High-Performance	879	33
Flex Fuel,Luxury,Performance	1380	28
Flex Fuel,Performance	1680	87
Flex Fuel,Performance,Hybrid	155	2
Hatchback	1319	641
Hatchback,Diesel	873	14
Hatchback,Factory Tuner,High-Performance	1205	13
Hatchback,Factory Tuner,Luxury,Performance	887	9
Hatchback,Factory Tuner,Performance	2159	22
Hatchback,Flex Fuel	5657	7
Hatchback,Hybrid	2121	72
Hatchback,Luxury	1380	46
Hatchback,Luxury,Hybrid	454	3
Hatchback,Luxury,Performance	1566	38
Hatchback,Performance	1040	252
High-Performance	1821	199
Hybrid	2106	123
Luxury	1103	855
Luxury,High-Performance	1668	334
Luxury,High-Performance,Hybrid	569	12
Luxury,Hybrid	674	52
Luxury,Performance	1293	673

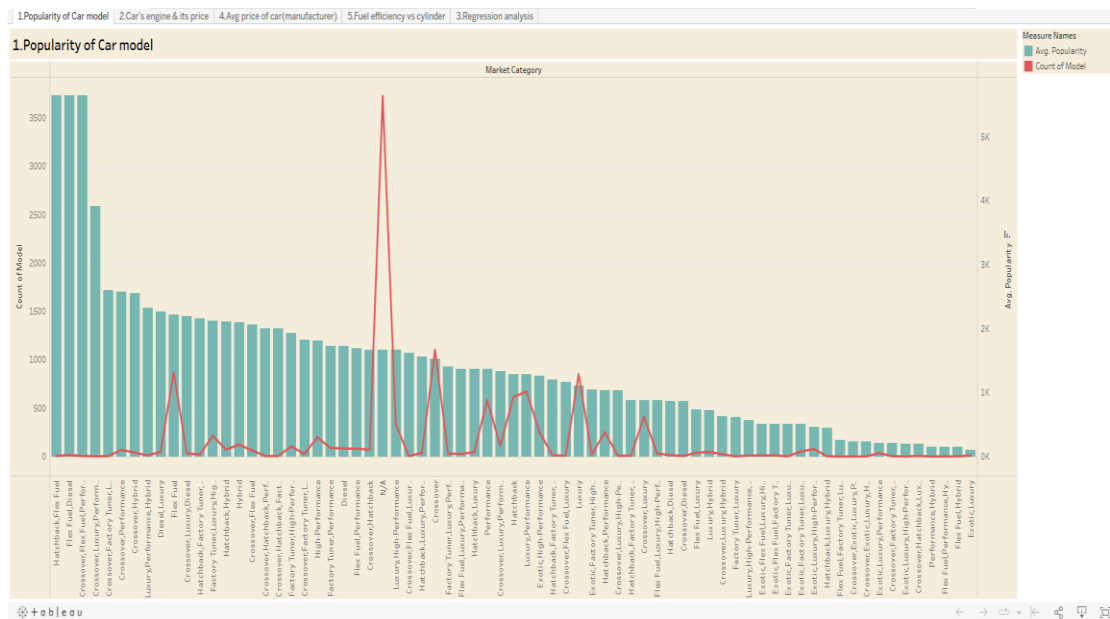
Luxury,Performance,Hybrid	2333	11
Performance	1349	601
Performance,Hybrid	155	1
Grand Total	1499	8172

Below the picture, inferred that the market category of car features with the table of creation of pivot table format and calculated with the help of excel pivot functions. Here the values are found about average of popularity scores and count of each and every models.

Here are attached a excel file for **popularity of scores.....**

[https://docs.google.com/spreadsheets/d/17UcMJqn62H03zj962rtX6BfZaAeS7nh6/edit?usp=drive link&ouid=103768596710140113695&rtpof=true&sd=true](https://docs.google.com/spreadsheets/d/17UcMJqn62H03zj962rtX6BfZaAeS7nh6/edit?usp=drive_link&ouid=103768596710140113695&rtpof=true&sd=true)

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



From the above combo chart tasks, I have created with a bar chart and combination of line chart, it varies as colors for both line and bar chart. Here I have insisted that the popularity of car is based on the average of popularity and count of each model.

So this tasks was done using the tableau public tool which is helped me to solve and analyzed by using this type of combo chart with the dataset of impact of car features that is have taken a popularity and count of each model within it.

Above that, only I showed a picture besides that I have attached a tableau public link here.....

[Tasks analysis for car data | Tableau Public](#)

2. 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 trend-line to the chart to visualize the relationship between these variables.

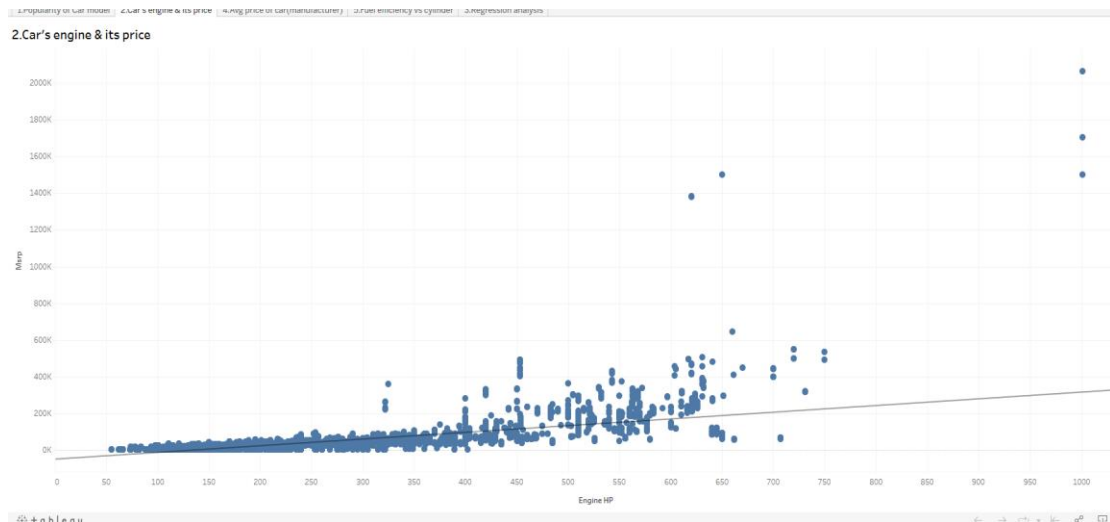
The solution for the Task 2 will be the relationship between car's engine and its price as MSRP , found out by separating from the final dataset of cleaned data which help me out to categorize the particular tasks and have been visualized with the trend line chart.

	B1		MSRP	
	A	B	C	D
1	Engine HP	MSRP		
2	335	\$ 46,135		
3	300	\$ 40,650		
4	300	\$ 36,350		
5	230	\$ 29,450		
6	230	\$ 34,500		
7	230	\$ 31,200		
8	300	\$ 44,100		
9	300	\$ 39,300		
10	230	\$ 36,900		
11	230	\$ 37,200		
12	300	\$ 39,600		
13	230	\$ 31,500		
14	300	\$ 44,400		
15	230	\$ 37,200		
16	230	\$ 31,500		
17	320	\$ 48,250		
18	320	\$ 43,550		
19	172	\$ 2,000		
20	172	\$ 2,000		
21	172	\$ 2,000		
22	172	\$ 2,000		
23	172	\$ 2,000		
24	172	\$ 2,000		

The above picture depicts that the separated data of car's **Engine HP** and **MSRP** , to find the relationship between the car's engine power and its price.

The Excel is attached in the below as a link.....

[https://docs.google.com/spreadsheets/d/1i56EwaP-cC-bma22SPcsNIthohyV024i/edit?usp=drive link&oid=103768596710140113695&rtpof=true&sd=true](https://docs.google.com/spreadsheets/d/1i56EwaP-cC-bma22SPcsNIthohyV024i/edit?usp=drive_link&oid=103768596710140113695&rtpof=true&sd=true)



From the above picture depicts that the scatter plot and combination of shown trend lines for finding the relationship between Engine HP and MSRP with respect to it.

Then I have taken for finding the visualization of scatter plot and with a trend lines is then taken Engine HP and MSRP to solve the relationship between them.

And also Engine HP will be taken as x-axis and MSRP taken as y-axis and plotting it accordingly and gone through the relation between , in the analysis it will be shown as trend lines - show trend lines and then edit the axis according to it.

Here I have attached Tableau public Link for finding the insights of this tasks....

[Tasks analysis for car data | Tableau Public](#)

3. 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.

Here I have learned how to solve the regression analysis using the excel data analysis functions, it gives the anova table , summary output with the residual and regression which inferred the regression data of car features on determining a car's price.

Then I have observed the car data with the help of determining the car features by including the header rows are below shown picture of header rows that has the car data within it.

Engine HP	Engine Cylinders	Number of Doors	city mpg	highway MPG	MSRP
-----------	------------------	-----------------	----------	-------------	------

	A	B	C	D	E	F	G
	Engine H	Engine Cylinder	Number of Door	city mpg	highway MPG	MSRP	
	335	6	2	19	26	\$ 46,135	
	300	6	2	19	28	\$ 40,650	
	300	6	2	20	28	\$ 36,350	
	230	6	2	18	28	\$ 29,450	
	230	6	2	18	28	\$ 34,500	
	230	6	2	18	28	\$ 31,200	
	300	6	2	17	26	\$ 44,100	
	300	6	2	20	28	\$ 39,300	
0	230	6	2	18	28	\$ 36,900	
1	230	6	2	18	27	\$ 37,200	
2	300	6	2	20	28	\$ 39,600	
3	230	6	2	19	28	\$ 31,500	
4	300	6	2	19	28	\$ 44,400	
5	230	6	2	19	28	\$ 37,200	
5	230	6	2	19	28	\$ 31,500	
7	320	6	2	18	25	\$ 48,250	
3	320	6	2	20	28	\$ 43,550	
9	172	6	4	17	24	\$ 2,000	
0	172	6	4	17	24	\$ 2,000	
1	172	6	4	16	20	\$ 2,000	
2	172	6	4	17	24	\$ 2,000	
3	172	6	4	16	21	\$ 2,000	
4	172	6	4	17	24	\$ 2,000	

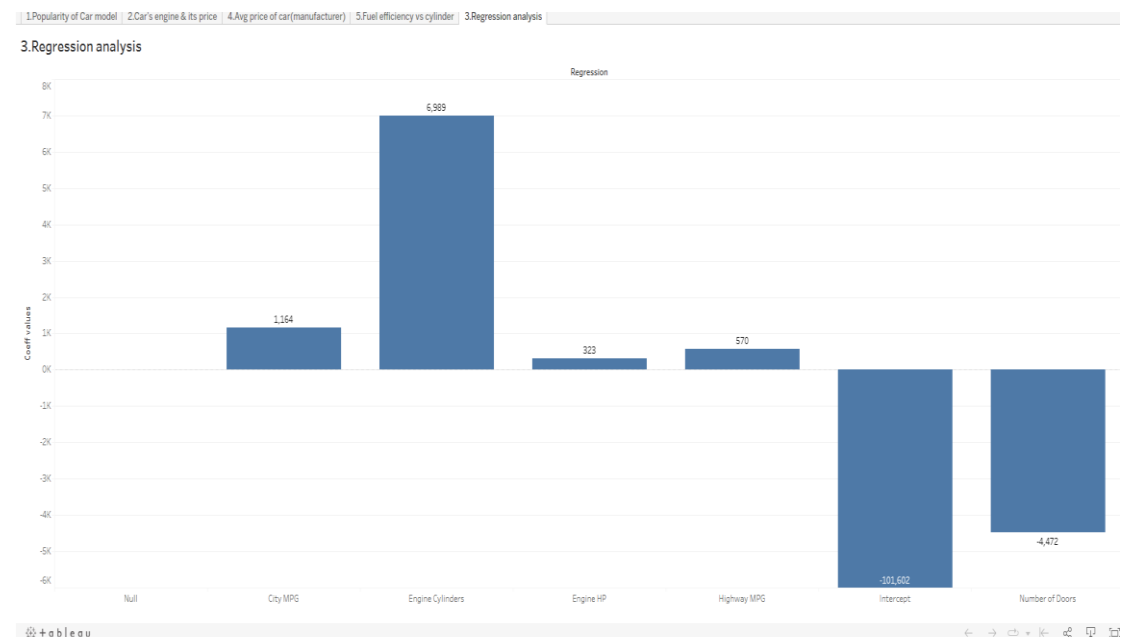
Above picture depicts that car features are most important for determining the car's price, so then analyzed with the car features are implied to the car's price which is MSRP. And determined a regression analysis in the excel function of data analysis tool.

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.680708139							
R Square	0.46336357							
Adjusted R Square	0.463136297							
Standard Error	44170.77827							
Observations	11812							
ANOVA								
	df	SS	MS	F	Significance F			
Regression	5	1.98891E+13	3.97782E+12	2038.799457	0			
Residual	11806	2.30342E+13	1951057653					
Total	11811	4.29233E+13						
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	-101601.736	3684.351697	-27.57655738	2.765E-162	-108823.673	-94379.79896	-108823.673	-94379.79896
Engine HP	322.7465574	6.01767382	53.63310924	0	310.9509241	334.5421906	310.9509241	334.5421906
Engine Cylinders	6989.177662	439.6449924	15.89732121	2.53591E-56	6127.400961	7850.954363	6127.400961	7850.954363
Number of Doors	-4472.158125	465.7180593	-9.602715711	9.35015E-22	-5385.042338	-3559.273912	-5385.042338	-3559.273912
Highway MPG	570.1808088	105.7839778	5.390048859	7.17937E-08	362.826764	777.5348535	362.826764	777.5348535
City MPG	1163.755457	121.9978136	9.539150113	1.72109E-21	924.61962	1402.891294	924.61962	1402.891294

This is the regression analysis its just a picture demonstration and I have included my excel file below this.

Here I have attached an Excel file.....

https://docs.google.com/spreadsheets/d/1SfVwWAF_zncu-QwgNlnrTQ3GsrX4MR5q/edit?usp=drive_link&oid=103768596710140113695&rtpof=true&sd=true



From the above picture about the regression analysis , that is in the tasks they have ask to make a insight of bar chart for the regression of each coefficient values with respect to it. I have done a regression analysis in the Microsoft excel above it with the help of that taken a coefficient values based on a car data header rows each of the values like engine hp, msrp and so on.

Here I have analyzed that in the bar chart , Engine cylinders have the more coefficient value that is 6989.177 and intercept and the number of rows are getting negative of the coefficient.

And also have attached tableau public link to find the visualization clearly

[Tasks analysis for car data | Tableau Public](#)

4. 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.

The solution this tasks is that average price of car vary across the different manufacturers that are depicted as make in this header row that calculates the average price car for each make.

So I have insisted on creating a pivot table with the help of manufacturers and MSRP for price of each make. And then created the pivot table with the row labels of make that is manufacturer and MSRP as Average of car price, that are included as pivot table and then insisted as the changing of

count to average of car price with the help of value field settings in the pivot table.

Manufacturer	Average of MSRP
Acura	₹ 34,887.59
Alfa Romeo	₹ 61,600.00
Aston Martin	₹ 1,97,910.38
Audi	₹ 53,452.11
Bentley	₹ 2,47,169.32
BMW	₹ 61,546.76
Bugatti	₹ 17,57,223.67
Buick	₹ 28,206.61
Cadillac	₹ 56,231.32
Chevrolet	₹ 28,273.36
Chrysler	₹ 26,722.96
Dodge	₹ 22,390.06
Ferrari	₹ 2,37,383.82
FIAT	₹ 22,206.02
Ford	₹ 27,393.42
Genesis	₹ 46,616.67
GMC	₹ 30,493.30
Honda	₹ 26,629.82
HUMMER	₹ 36,464.41
Hyundai	₹ 24,597.04
Infiniti	₹ 42,394.21
Kia	₹ 25,112.39
Lamborghini	₹ 3,31,567.31
Land Rover	₹ 67,823.22
Lexus	₹ 47,549.07
Lincoln	₹ 42,494.37
Lotus	₹ 69,188.28
Maserati	₹ 1,14,207.71
Maybach	₹ 5,46,221.88
Mazda	₹ 19,719.06
McLaren	₹ 2,39,805.00
Mercedes-Benz	₹ 71,537.81
Mitsubishi	₹ 21,215.47
Nissan	₹ 28,513.37
Oldsmobile	₹ 11,542.54
Plymouth	₹ 3,122.90
Pontiac	₹ 19,321.55
Porsche	₹ 1,01,622.40
Rolls-Royce	₹ 3,51,130.65
Saab	₹ 27,413.50
Scion	₹ 19,932.50
Spyker	₹ 2,13,323.33
Subaru	₹ 24,827.50

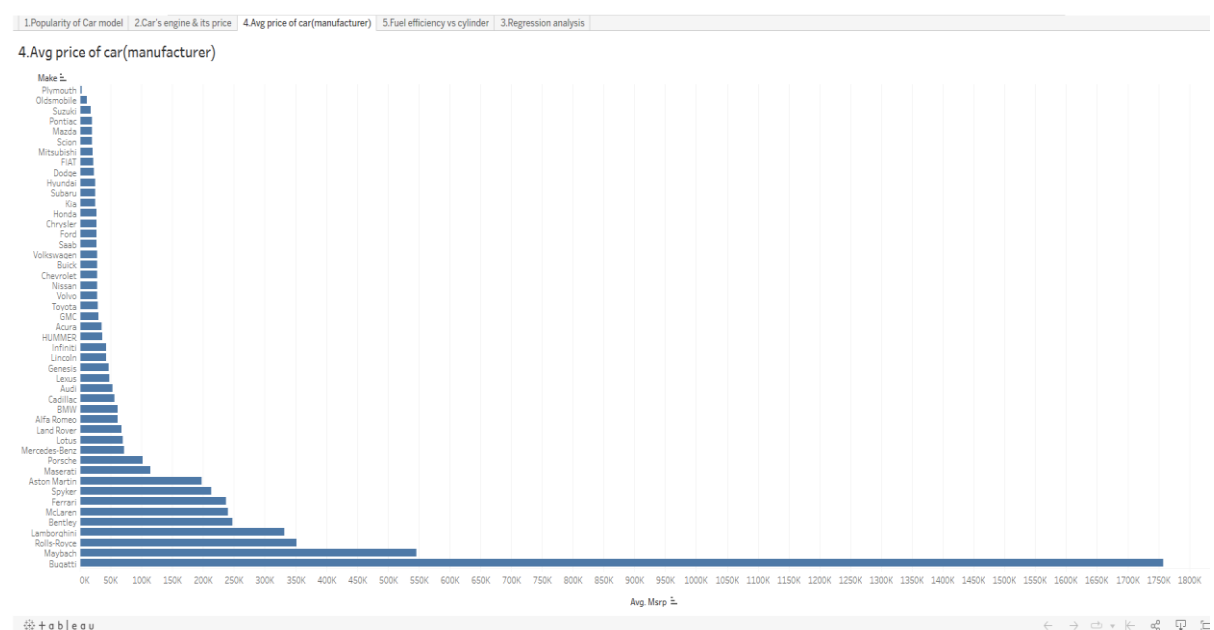
Suzuki	₹	17,900.96
Toyota	₹	28,946.15
Volkswagen	₹	28,076.20
Volvo	₹	28,541.16
Grand Total	₹	40,559.94

Above the picture depicts that the pivot table of calculated average of car' price as the make as the manufacturer and grand total of total amount of car is been determined.

The excel file for determining the average of car' price tasks is attached here....

https://docs.google.com/spreadsheets/d/1K28K6_t8dE48yUZLqsMBdg-BPzYDGOoU/edit?usp=drive_link&oid=103768596710140113695&rtpof=true&sd=true

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 pasted picture represents that the horizontal a stacked bar chart for insights of average price of car with respect to the manufacturer. Taking Average of price that is in the header row of MSRP is been taken as x-axis and Manufacturer will be taken as y-axis .

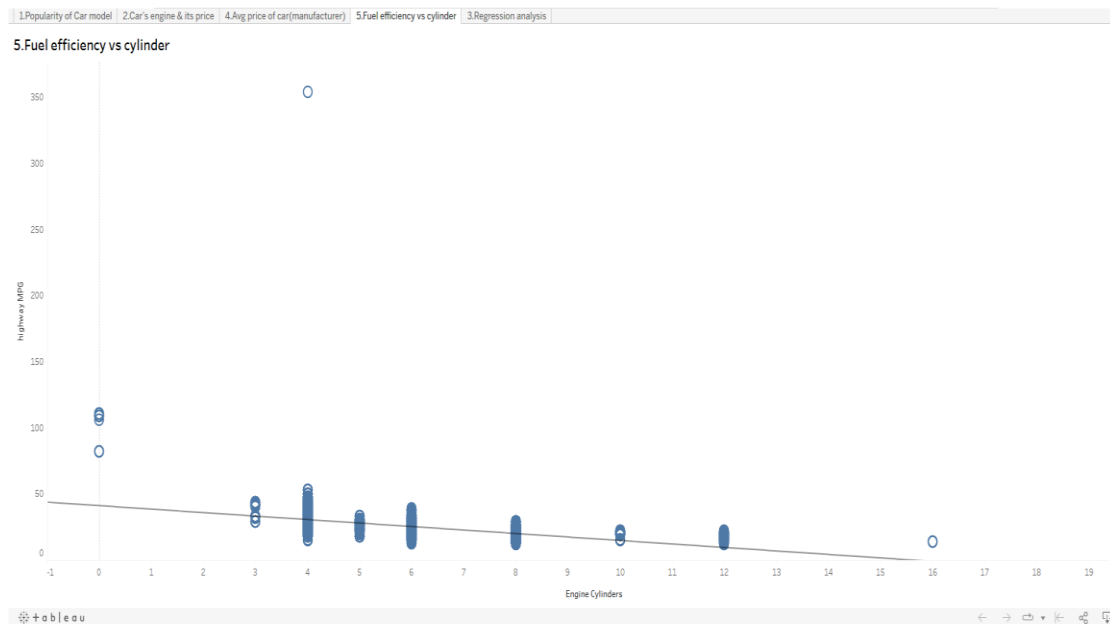
Here my tasks is to find the average of car price with each of the manufacturer in that **Bugatti** has the highest car price.

Also then I have attached a Tableau public link that has the charts which have the tasks is been as follows....

[Tasks analysis for car data | Tableau Public](#)

5. 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.



The above picture represents the scatter plot of finding the fuel efficiency versus number of cylinder and with a trend lines for the reference of insisting the correlation of fuel efficiency with each of the cylinders.

It represents that the main cause of using the cylinders for each of the Highway MPG is that shown as the scatter plots with the trend lines about this tasks. Engine cylinder 4 has the highest scattered plots of its MPG.

Also that I have attached a Tableau public Link to view the scattered plot as the clear view.....

[Tasks analysis for car data | Tableau Public](#)

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

In this tasks , I found the solution that the correlation of coefficient between number of cylinders and the highway MPG played a role in this tasks. It quantifies the strength of each of the cylinders numbered and a direction of relationship between them.

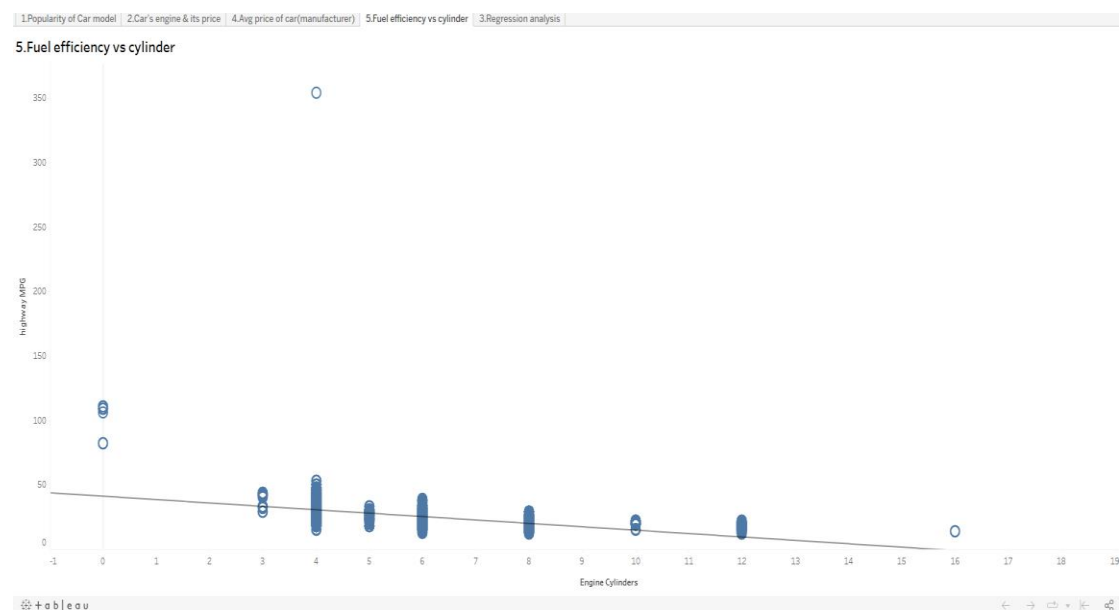
Clipboard		Font		Alignment			
E6		fx					
	A	B	C	D	E	F	
1	Engine Cylinders	highway MPG					
2	6	26			Correlation, r=	-0.62031	
3	6	28					
4	6	28					
5	6	28					
6	6	28					
7	6	28					
8	6	26					
9	6	28					
10	6	28					
11	6	27					
12	6	28					
13	6	28					
14	6	28					
15	6	28					
16	6	28					
17	6	25					
18	6	28					
19	6	24					
20	6	24					
21	6	20					
22	6	24					
23	6	21					
24	6	24					

From the above picture is my analysis of finding the fuel efficiency with the number of cylinders and found out that coefficient values as **$r = -0.620312551$** .

This was achieved by the excel correl function for finding the fuel efficiency of correlation value. With this r value I can make use of analyzing the empowerment of number of cylinders with the Highway MPG .

So here I have attached a excel file....

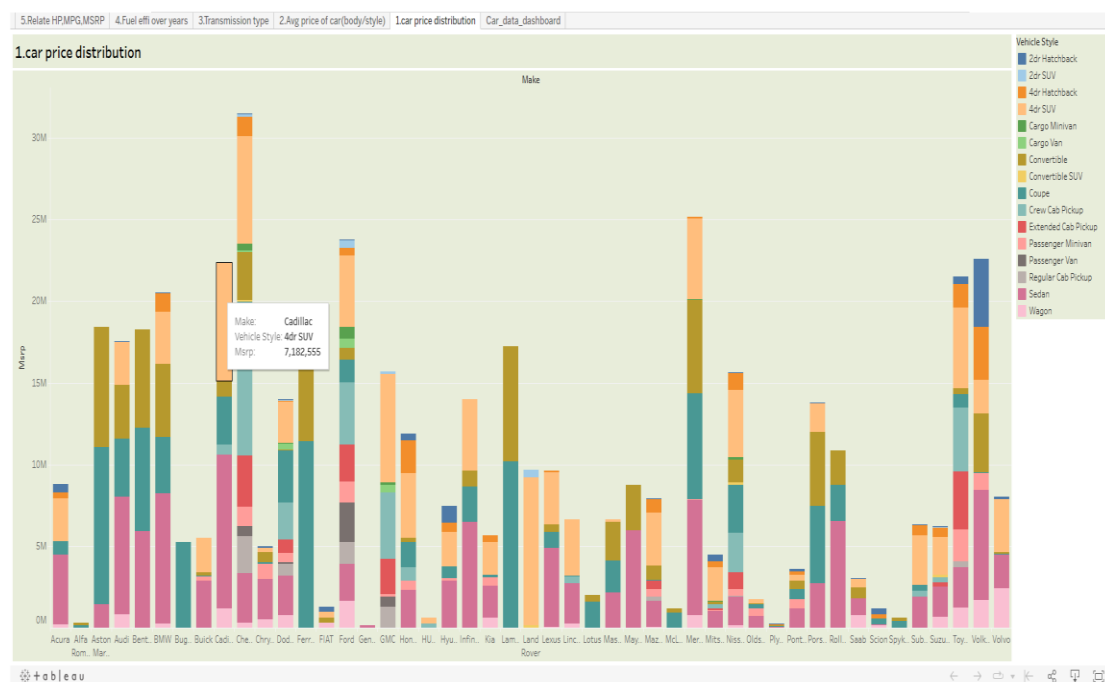
https://docs.google.com/spreadsheets/d/17B5Ee5Y9I2HClCbClQd2MPq2aEUW4Abs/edit?usp=drive_link&ouid=103768596710140113695&rtpof=true&sd=true



BUILDING THE 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.



In this dashboard analysis, I have inferred that the distribution of car price with respect to the vehicle style and manufacturer for each of the car. Now here what is the tasks is that to make a stacked column chart for making the

visualizations of finding the distribution of car price based on branch and vehicle style.

The tableau public link is attached here....

[dashboardall | Tableau Public](#)

DISTRIBUTION OF CAR PRICE BY BRAND AND BODY STYLE												
Brands	2dr Hatchback	2dr SUV	4dr Hatchback	4dr SUV	Cargo Minivan	Cargo Van	Convertible	Convertible SUV	Coupe	Crew Cab Pickup	Extended Cab Pickup	Passenger Minivan
Acura	\$ 4,80,917		\$ 3,57,440	\$ 26,63,505					\$ 7,93,748			
Alfa Romeo							\$ 1,29,800		\$ 1,78,200			
Aston Martin							\$ 73,21,655		\$ 96,35,275			
Audi	\$ 4,000			\$ 26,74,900			\$ 32,91,405		\$ 35,56,290			
Bentley							\$ 60,12,870		\$ 63,56,760			
BMW	\$ 80,097		\$ 11,44,950	\$ 31,60,950			\$ 45,02,671		\$ 34,19,051			
Bugatti									\$ 52,71,671			
Buick				\$ 21,41,770			\$ 1,79,325		\$ 18,534			\$ 3,30,065
Cadillac				\$ 71,82,555			\$ 9,85,607		\$ 29,53,574	\$ 5,99,150		
Chevrolet	\$ 8,000	\$ 2,13,310	\$ 12,09,735	\$ 65,69,568	\$ 4,20,150	\$ 78,688	\$ 29,53,245	\$ 1,06,300	\$ 35,04,525	\$ 59,27,617	\$ 31,17,951	\$ 11,78,515
Chrysler	\$ 98,805			\$ 2,50,545			\$ 6,30,105		\$ 1,14,510			\$ 9,22,295
Dodge	\$ 48,000	\$ 44,000	\$ 18,000	\$ 25,72,405	\$ 60,520	\$ 3,000	\$ 2,000		\$ 32,64,627	\$ 22,35,775	\$ 8,64,172	\$ 5,57,425
Ferrari							\$ 3,811		\$ 1,14,18,289			
FIAT	\$ 3,25,315			\$ 3,69,305			\$ 7,965					
Ford	\$ 36,000	\$ 4,79,873	\$ 4,80,155	\$ 43,70,871	\$ 6,80,770	\$ 5,000	\$ 0,007		\$ 13,98,144	\$ 38,12,353	\$ 22,85,584	\$ 12,71,330
Genesis												
GMC		\$ 1,44,319		\$ 66,41,919	\$ 1,42,750	\$ 4,68,085				\$ 40,62,482	\$ 21,83,866	\$ 1,50,630
Honda	\$ 4,13,200		\$ 20,15,270	\$ 39,53,209			\$ 2,52,135		\$ 15,88,705	\$ 7,87,720		\$ 5,53,185
HUMMER				\$ 3,77,490						\$ 2,42,405		
Hyundai	\$ 10,38,050		\$ 5,28,880	\$ 21,28,890					\$ 7,24,070			\$ 1,33,075
Infiniti				\$ 43,40,200			\$ 9,80,050		\$ 21,75,750			
Kia			\$ 4,06,960	\$ 20,49,645					\$ 1,42,630			\$ 4,94,650

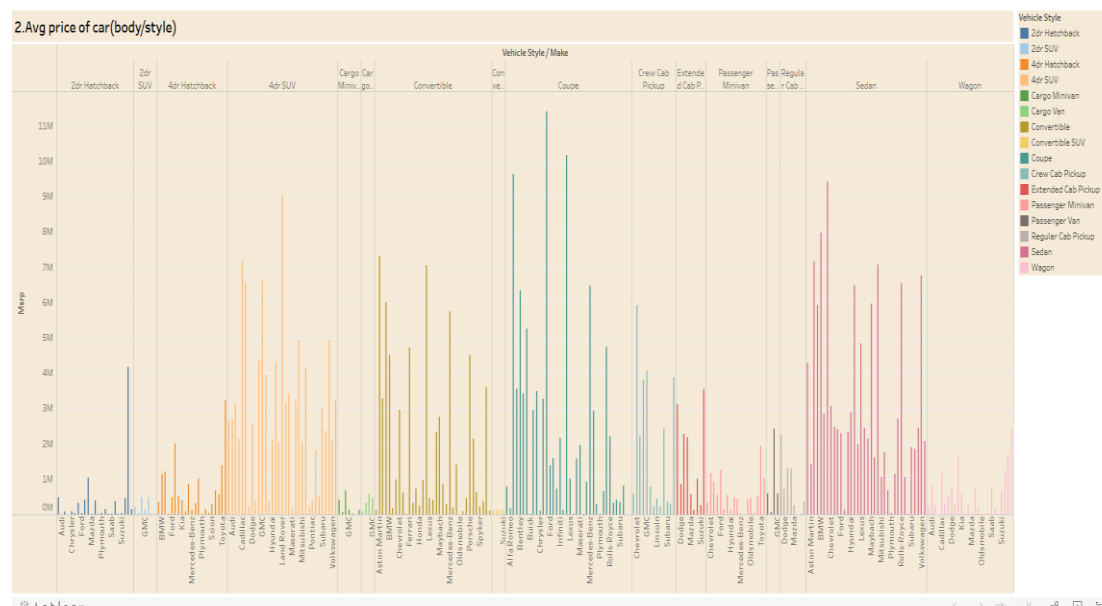
With the help of making pivot tables in the Excel sheet can analyzed with the column stacked chart. Here I have attached dashboard excel file contains pivot tables for calculating the dashboard visualization....

[https://docs.google.com/spreadsheets/d/1b1wxdIAMs2myBUJ_JBz9xwXdu4xoHt3t/edit?usp=drive link&ouid=103768596710140113695&rtpof=true&sd=true](https://docs.google.com/spreadsheets/d/1b1wxdIAMs2myBUJ_JBz9xwXdu4xoHt3t/edit?usp=drive_link&ouid=103768596710140113695&rtpof=true&sd=true)

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.



For this dashboard analysis , I have insisted on finding the average price car based on both body and style position. So I have created a column chart to visualize the values here I have done both the body and style together with each of the brand with its price that is MSRP.

Here I have attached a Tableau public link for find this column chart....[dashboardall | Tableau Public](#)

AVERAGE MSRP ACROSS DIFFERENT CARS BRANDS AND BODY STYLE																	
Body Styles	2dr Hatchback	2dr SUV	4dr Hatchback	4dr SUV	Cargo Minivan	Cargo Van	Convertible	Convertible SUV	Coupe	Crew Cab Pickup	Extended Cab Pickup	Passenger Minivan	Passenger Van	Regular Cab Pickup	Sedan	Wagon	Grand Total
	\$ 17.175		\$ 51.063	\$ 42.960			\$ 64.900		\$ 35.887						\$ 33.252	\$ 33.550	\$ 34.888
							\$ 2.03.379		\$ 59.400						\$ 2.06.962		\$ 61.600
\$ 2.000				\$ 48.635			\$ 70.030		\$ 93.597						\$ 44.462	\$ 33.894	\$ 1.97.910
							\$ 2.50.536		\$ 2.54.270						\$ 2.36.636		\$ 2.47.859
\$ 26.689			\$ 54.521	\$ 58.536			\$ 63.418		\$ 51.084						\$ 70.782	\$ 43.267	\$ 61.547
									\$ 17.57.224								\$ 17.57.224
				\$ 33.956			\$ 25.618		\$ 2.093			\$ 30.006			\$ 27.947	\$ 2.053	\$ 28.207
				\$ 72.551			\$ 70.401		\$ 45.440	\$ 66.572					\$ 50.953	\$ 47.364	\$ 56.231
\$ 2.000	\$ 8.888		\$ 18.329	\$ 32.047	\$ 20.807	\$ 7.753	\$ 62.875	\$ 17.717	\$ 38.939	\$ 39.256				24.307	\$ 18.625	\$ 19.799	\$ 18.625
\$ 32.955				\$ 35.792			\$ 24.235		\$ 16.095						\$ 26.184	\$ 26.372	\$ 26.723
\$ 2.000	\$ 2.000	\$ 2.000	\$ 30.993	\$ 20.173	\$ 12.537	\$ 2.000	\$ 45.981		\$ 31.052					14.142	\$ 9.343	\$ 21.780	\$ 22.390
							\$ 2.14.719		\$ 2.48.224								\$ 2.37.384
\$ 19.136				\$ 24.620			\$ 23.426										\$ 22.206
\$ 2.000	\$ 13.711		\$ 18.468	\$ 42.028	\$ 21.274	\$ 17.698	\$ 34.762		\$ 34.101	\$ 41.439	\$ 23.808	\$ 23.115	\$ 32.425	\$ 17.798	\$ 21.280	\$ 27.259	\$ 27.293
																	\$ 46.617
	\$ 5.551		\$ 36.696	\$ 23.792	\$ 16.723					\$ 29.062	\$ 36.629	\$ 25.105	\$ 26.247	\$ 21.070			\$ 30.493
\$ 17.217		\$ 25.637	\$ 28.896				\$ 36.019		\$ 21.763	\$ 34.249		\$ 36.879			\$ 36.001		\$ 26.630
			\$ 37.749							\$ 34.629							\$ 36.464
\$ 18.537			\$ 17.629	\$ 30.410					\$ 20.688			\$ 26.616			\$ 27.102		\$ 24.597
			\$ 45.696				\$ 46.689		\$ 40.292						\$ 40.588		\$ 42.394
			\$ 19.379	\$ 31.533					\$ 20.376			\$ 32.977			\$ 23.298	\$ 18.217	\$ 25.112
							\$ 3.36.402		\$ 3.28.292								\$ 3.31.567
\$ 38.700			\$ 70.911				\$ 48.577										\$ 67.623
		\$ 31.567	\$ 45.942				\$ 52.482		\$ 50.824						\$ 48.685	\$ 31.105	\$ 47.545
			\$ 50.332					\$ 2.112		\$ 41.205					\$ 41.665	\$ 44.951	\$ 42.454
							\$ 51.659		\$ 75.967								\$ 89.988
			\$ 77.500				\$ 1.30.165		\$ 1.16.017						\$ 1.02.562		\$ 1.14.208
							\$ 13.81.375								\$ 4.26.594		\$ 5.46.222
\$ 2.000	\$ 2.000	\$ 20.888	\$ 27.880				\$ 20.081		\$ 2.000		\$ 16.801	\$ 23.323		\$ 9.155	\$ 19.739	\$ 16.675	\$ 19.719
							\$ 2.80.225		\$ 2.29.700								\$ 2.38.805
							\$ 1.10.430		\$ 1.106.734								\$ 1.106.734
												\$ 22.500			\$ 49.353	\$ 44.956	\$ 71.636

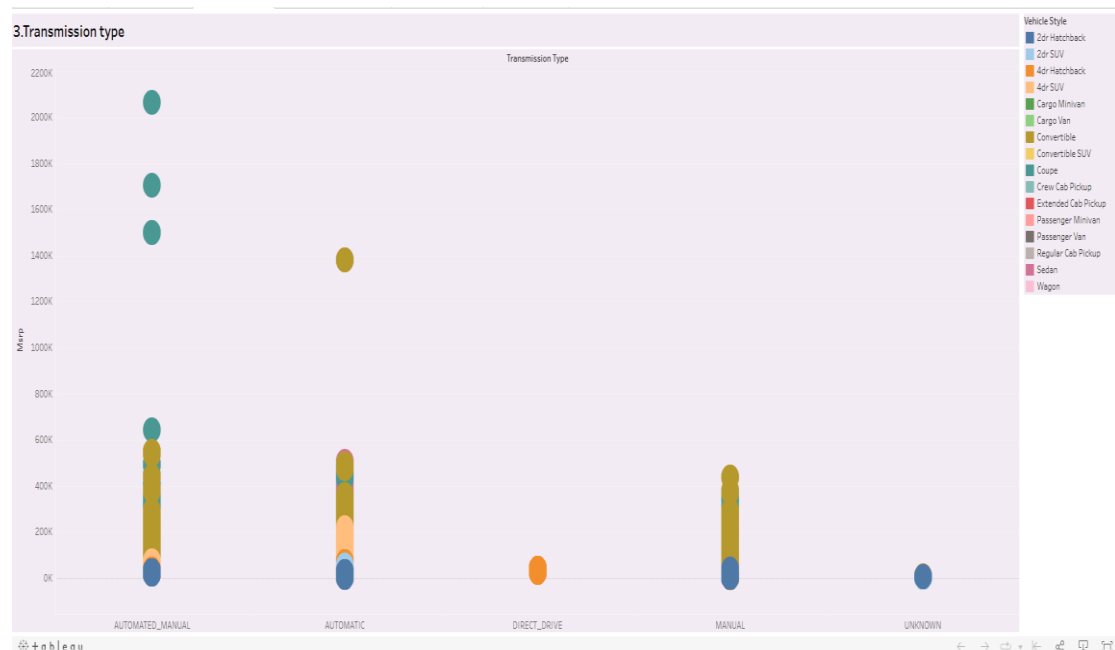
With the making the suggested pivot tables, I have done here that body style and the brand of the price I have insisted. With this pivot table I can easily make a column chart to compare body styles with the brands.

Here I have attached a excel file link to get full view of including total of each of the brands in pivot table....
https://docs.google.com/spreadsheets/d/1M7zOZuTGB9XstFaLoxuRe7o-JbT4bwJj/edit?usp=drive_link&oid=103768596710140113695&rtpof=true&sd=true

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

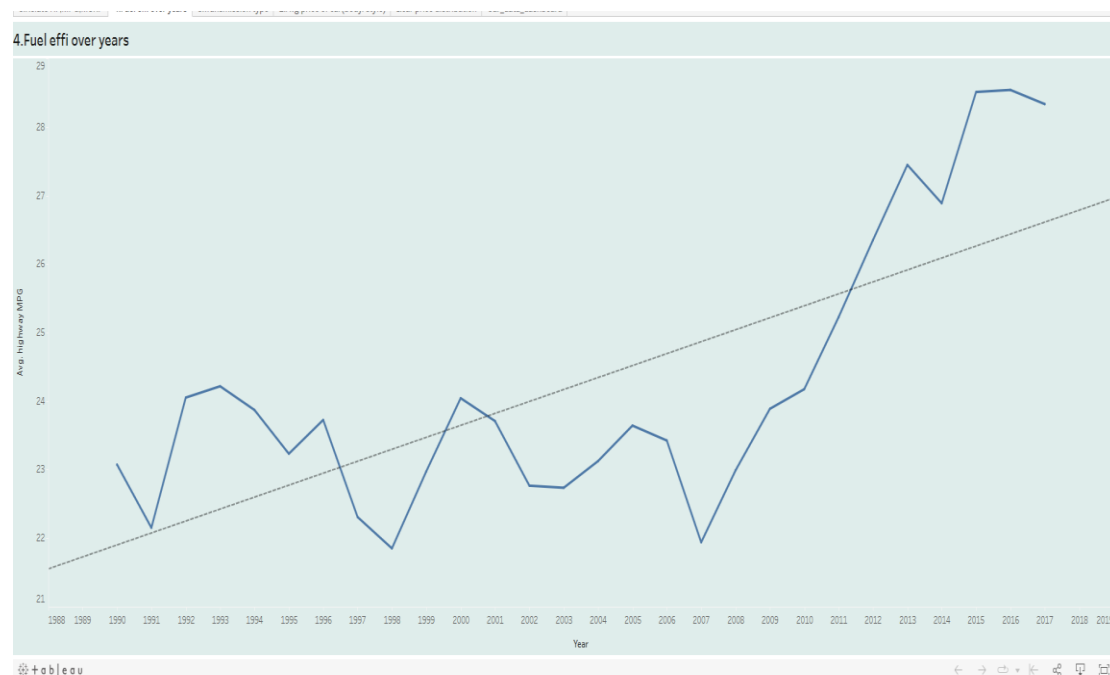
Hints: Scatter plot chart to visualize the relationship between MSRP and transmission type, with different

symbols for each body style. Calculate the average MSRP for each combination of transmission type and body style using AVERAGEIFS or Pivot Tables.



For this dashboard analysis, I have inferred that the transmission type of vehicle styles as a feature that affects the MSRP of car to shown below the pivot tables. Here I have taken vehicle style as x-axis and msrp as y-axis to get the designed scatter plot to visualize the relationship between the msrp and vehicle style, and also legends shown as for the reference of each of the color given to vehicle styles.

Here I have attached a tableau public link to find this dashboard... [dashboardall | Tableau Public](#)



For this dashboard analysis, it is make a Line chart to show the trend of fuel efficiency (MPG) over time for each body style. It represents that the fuel efficiency of finding the average number of highway MPG with respect to the year.

And calculated the average MPG for each combination of body style and model year with the visualization of line chart with the trend lines.

Here I have attached a tableau public link....

[dashboardall | Tableau Public](#)

		AVERAGE MSRP FOR EACH COMBINATION OF BODY STYLE AND MODEL YEAR																
Average of highway MPG	vehicle styles																	
Year	2dr Hatchback	2dr SUV	4dr Hatchback	4dr SUV	Cargo Minivan	Cargo Van	Convertible	Convertible SUV	Coupe	Crew Cab Pickup	Extended Cab Pickup	Passenger Minivan	Passenger Van	Regular Cab Pickup	Sedan	Wagon	Grand Total	
1990		30	20	31		20		24	25			22	19		22	24	24	23
1991		30	16		19			23	26			16	18		17	24	23	22
1992		30	17	28	21			26	27			16			18	25	24	24
1993		29	18	27	21			24	26	28		17			18	25	24	24
1994		27	18	27	20	21	19	26	26	27		20	21	15	22	25	24	24
1995		30	16	28		22	19	25	26	26		20	20	15	21	24	24	23
1996		29	20	26	22	23	15	24	24	27		20	21	15	22	26	25	24
1997		26	22	27	20	21	17	25	21	27		18	21	17	19	25	24	23
1998		23	26	25	22		17	24	24	26		16	22	17	16	27	23	23
1999		30	19		18		17	22		28		18	22		18	27		23
2000		30	19		18		16	25		24		21	23	15	21	27	31	24
2001		29	19		19	22	16	23		20		19	21	15	23	27	31	24
2002		25	19		20	21	15	24	23	24	17	20	22	15	22	26	29	23
2003		30	19		19	21	15	20	23	24	18	21	22		24	27	24	23
2004		30	19	34	19	20		20		25	22	18	22		18	26	23	23
2005		30	19	31	19	21		21		26	23	22			18	26	24	24
2006		27		29	20	23		23	24	19		23			18	25	25	23
2007		26		28	21	23		23	25	18	18	23	23		20	25	25	22
2008		27		29	21	23		24	25	18	19	23	23		18	27	25	23
2009		29		31	23			24	24	19	20				22	27	27	24
2010		28		30	23			25	24	19	21	24			21	26	28	24
2011		28		29	24			24	23	21	22	25			27	27	29	25
2012		31		33	23		17	24	22	22	21	23	25	15	24	28	31	26
2013		32		32	24		17	23	22	25	21	28	15		30	30		27
2014		35		35	24		17	27	22	23	19	17	26	16		30	29	27
2015		34	30	36	26	28	17	28		26	22	22	26	18	23	32	31	29
2016		34	30	36	26	28	16	28		27	22	22	26	18	23	32	30	29
2017		33	29	36	26	27		28	28	28	22	21	26	19	23	33	31	28
Grand Total		31	19	35	24	25	17	26	24	26	21	24	17		21	30	28	26

With the help of this pivot table in excel I can make a line chart for finding the average msrp with the body style of each year. I have calculated a average MPG for each combination of body style and model year.

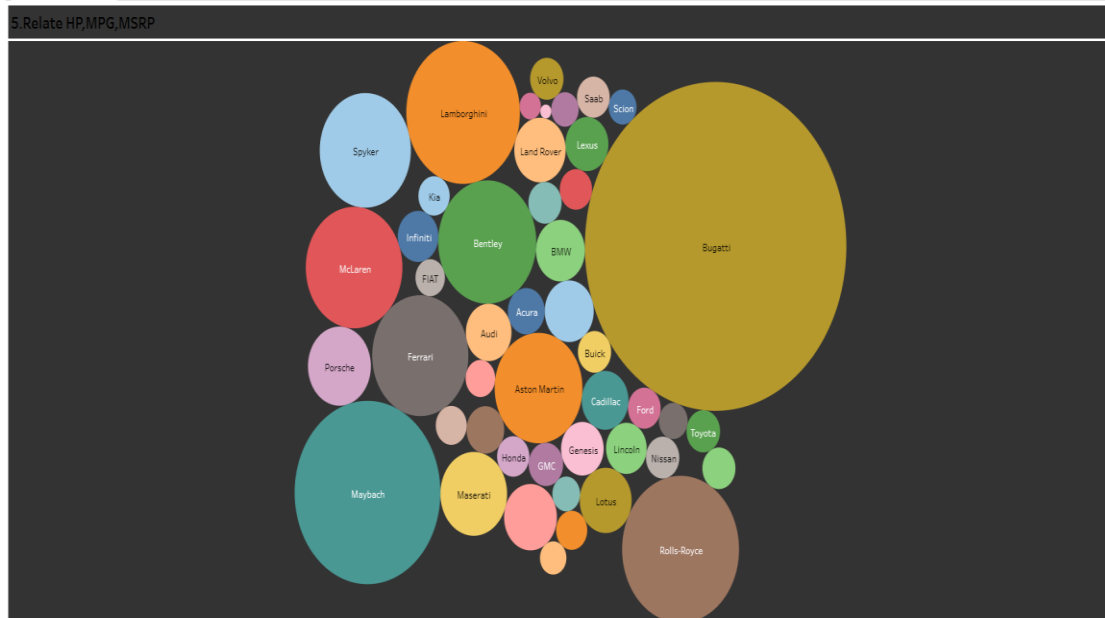
Here I have attached a excel link....

https://docs.google.com/spreadsheets/d/1FiZn8qYeWuLOvuhWMKEsA6gzyoNZy4cp/edit?usp=drive_link&ouid=103768596710140113695&rtpof=true&sd=true

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.



For this dashboard analysis, I have insisted on making a bubble chart with the help of 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. And have calculated the average horsepower, MPG, and MSRP for each car brand.

Here I have attached a tableau public link to find the fullscreen view of my dashboard analysis....

[dashboardall | Tableau Public](#)

AVERAGE OF ENGINE HP, MPG & MSRP FOR EACH CAR BRAND			
	values		
Brands	Average of Engine HP	Average of highway MPG	Average of MSRP
Acura	245	28	\$ 34,888
Alfa Romeo	237	34	\$ 61,600
Aston Martin	484	19	\$ 1,97,910
Audi	278	29	\$ 53,452
Bentley	534	19	\$ 2,47,169
BMW	327	29	\$ 61,547
Bugatti	1001	14	\$ 17,57,224
Buick	219	27	\$ 28,207
Cadillac	332	25	\$ 56,231
Chevrolet	247	26	\$ 28,273
Chrysler	229	26	\$ 26,723
Dodge	244	22	\$ 22,390
Ferrari	510	16	\$ 2,37,384
FIAT	144	34	\$ 22,206
Ford	243	24	\$ 27,393
Genesis	347	25	\$ 46,617
GMC	260	21	\$ 30,493
Honda	196	32	\$ 26,630
HUMMER	261	17	\$ 36,464
Hyundai	202	30	\$ 24,597
Infiniti	310	25	\$ 42,394
Kia	207	29	\$ 25,112
Lamborghini	614	18	\$ 3,31,567
Land Rover	322	22	\$ 67,823
Lexus	277	26	\$ 47,549
Lincoln	285	24	\$ 42,494
Lotus	276	27	\$ 69,188
Maserati	421	20	\$ 1,14,208
Maybach	591	16	\$ 5,46,222
Mazda	169	28	\$ 19,719
McLaren	610	22	\$ 2,39,805
Mercedes-Benz	350	25	\$ 71,538
Mitsubishi	174	27	\$ 21,215
Nissan	240	26	\$ 28,513
Oldsmobile	177	26	\$ 11,543
Plymouth	132	28	\$ 3,123
Pontiac	190	27	\$ 19,322
Porsche	393	25	\$ 1,01,622
Rolls-Royce	488	19	\$ 3,51,131
Saab	221	26	\$ 27,414
Scion	154	32	\$ 19,933

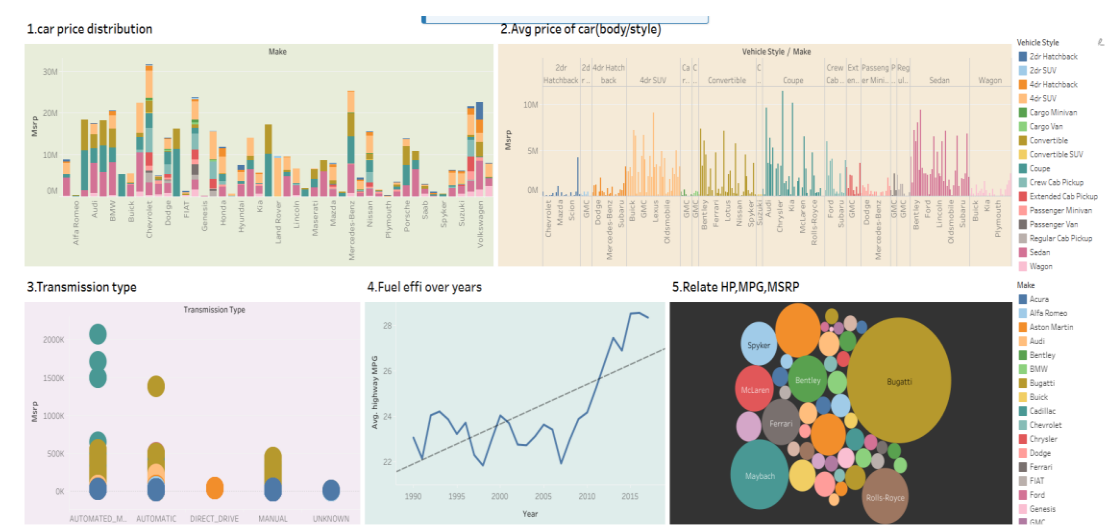
Spyker	400	18	\$	2,13,323
Subaru	197	29	\$	24,828
Suzuki	160	26	\$	17,901
Toyota	236	26	\$	28,946
Volkswagen	190	32	\$	28,076
Volvo	231	27	\$	28,541
Grand Total	250	26	\$	40,560

With the help of suggested pivot table, I had created a bubble for engine hp, highway mpg and msrp with respect to the each of the brands. And have calculated the average horsepower, MPG, and MSRP for each car brand using excel pivot table functions.

Here I have attached a excel file link....

https://docs.google.com/spreadsheets/d/1Byc9Qo5B-w2EbPvsBZyZJZBT7GC29777/edit?usp=drive_link&oid=103768596710140113695&rtpof=true&sd=true

DASHBOARD(overall)



Link: [dashboards/1](https://public.tableau.com/v/1/dashboards/1) | Tableau Public

RESULT

In this project , I have learnt how to make a analysis of each of the dashboards, I was very thrilled to making a beautiful insights as a data analyst. I have used tableau public and excel to find the tasks and dashboard analysis.

Future directions continue to be improvements in software and algorithms. Increasing the level of parallelism is desired for dealing with large scale computational and memory requirements.

Limitations:

Many advanced car features can significantly increase the cost of the vehicle. This can make the car less affordable for some consumers. Advanced features often require specialized maintenance and repairs, which can be more expensive and may require specific expertise not readily available at all repair shops.

DRIVE LINK

https://drive.google.com/drive/folders/1teT0DcMzWfXiVx_33PFkRoUtyh_v9gJK?usp=drive_link