Analyzing the Impact of Car Features on Price and Profitability

Excel Hyperlink

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.

For the given dataset, as a Data Analyst, the client has asked How can a car manufacturer optimize pricing and product development decisions to maximize profitability while meeting consumer demand? This problem could be approached by analyzing the relationship between a car's features, market category, and pricing, and identifying which features and categories are most popular among consumers and most profitable for the manufacturer. By using data analysis techniques such as regression analysis and market segmentation, the manufacturer could develop a pricing strategy that balances consumer demand with profitability, and identify which product features to focus on in future product development efforts. This could help the manufacturer improve its competitiveness in the market and increase its profitability over time.

Data 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.

The dataset contains information on over 11,000 car models and their specifications, including details on the car's make, model, year, fuel type, engine power, transmission, wheels, number of doors, market category, size, style, estimated miles per gallon, popularity, and manufacturer's suggested retail price (MSRP).

Overall, this dataset could be a valuable resource for data analysts interested in exploring various aspects of the automotive industry and could provide insights that could inform decisions related to product development, marketing, and pricing.

This dataset contains 11915 rows and 16 columns before cleaning the dataset.

But after handling the all null values the dataset could have 11813 rows and 16 columns are present in the new dataset.

By examining the popularity variable in the dataset, We could identify which features are most popular among consumers and how they affect a car's popularity. This could help manufacturers make informed decisions about product development and marketing. Predicting the price of a car based on its features and market category: By using the various features and market category variables in the dataset, a data could develop a model to predict the price of a car. This could help manufacturers and consumers understand how different features affect the price of a car and make informed decisions about pricing and purchasing.

APPROACH

Firstly, we downloaded the given dataset and starting from cleaning the data and handling the all null values to get accurate results. After this we are understanding the variables and relationship between them and analyzing the questions to extract the valuable insights from them. Also I applied all statistical and logical building techniques to find the relationship of variables and making pivot tables and many graphs, correlations to better visualize our insights. Also we done EDA and feature engineering and solved questions one by one and summarize the insights. We faced some problems to extracting the insights from the data but after doing thoroughly we got the required insights.

TECH-STACK USED

We used the Microsoft Excel 2019 for EDA and Feature Engineering and for whole project and Microsoft Word 2019 used to make presentation and convert into PDF.

We used MS Excel because it enables uses to format, organize and calculate data in a spreadsheet and it make easier to us to extract the valuable insights. Also it allows users to modify the fields and functions that perform computations when working with more complex data. We used the given dataset to extract the valuable insights.

INSIGHTS

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

In this question, we have to create the pivot table to shows the number of car model in each market category and their popularity scores it helps us to know which brand of car are more likely to car enthusiast and more likely to buy it will also help to know which car have more demand and which

car models are more likely to buy. It will helps to manufacturer to increases their manufacturing of particular to increase their sales and make more profit. We made pivot table of Car Models and Sum of popularity to know the each car popularity.

Here's our Top 10 most popularity car :

Accord	165150
Expedition	203652
F-150	712782
F-250	271536
Ranger	209309
Silverado 1500	216060
Tacoma	162480
Transit Connect	260222
Transit Wagon	271536
Tundra	284340

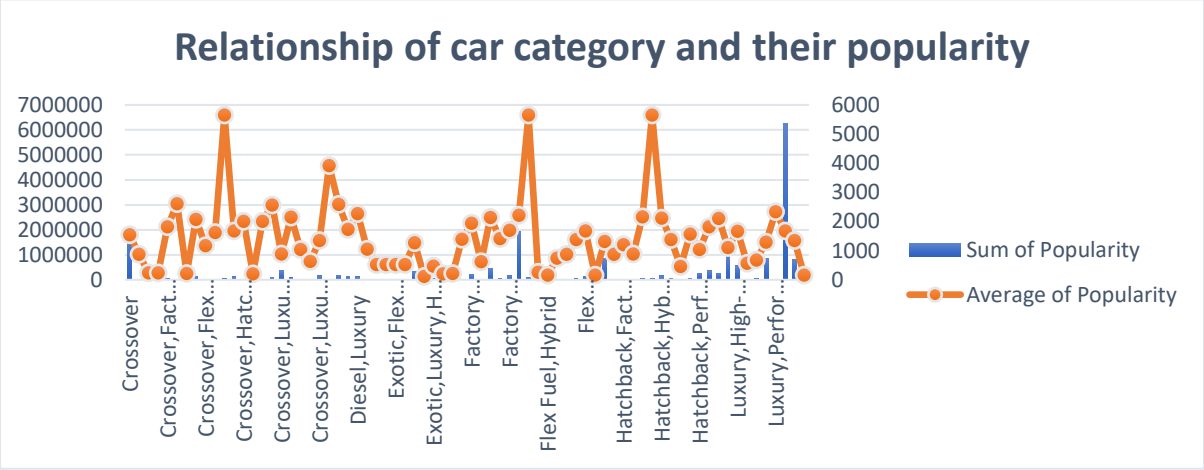
Task 1.B: Create a combo chart that visualizes the relationship between market category and popularity.

In this question, we made a combo chart to visualizes the relationship between market category and popularity. It helps us to know which kind of car are most popular and which one aren't.

It will also help to know the most favourable category and which category are more sales in the market.

Firstly, we separate the required columns from the main data and made a new sheet then used pivot table and calculate the sum and average of popularity for each category of car. It will help the most popular category of car.

Here's our output :



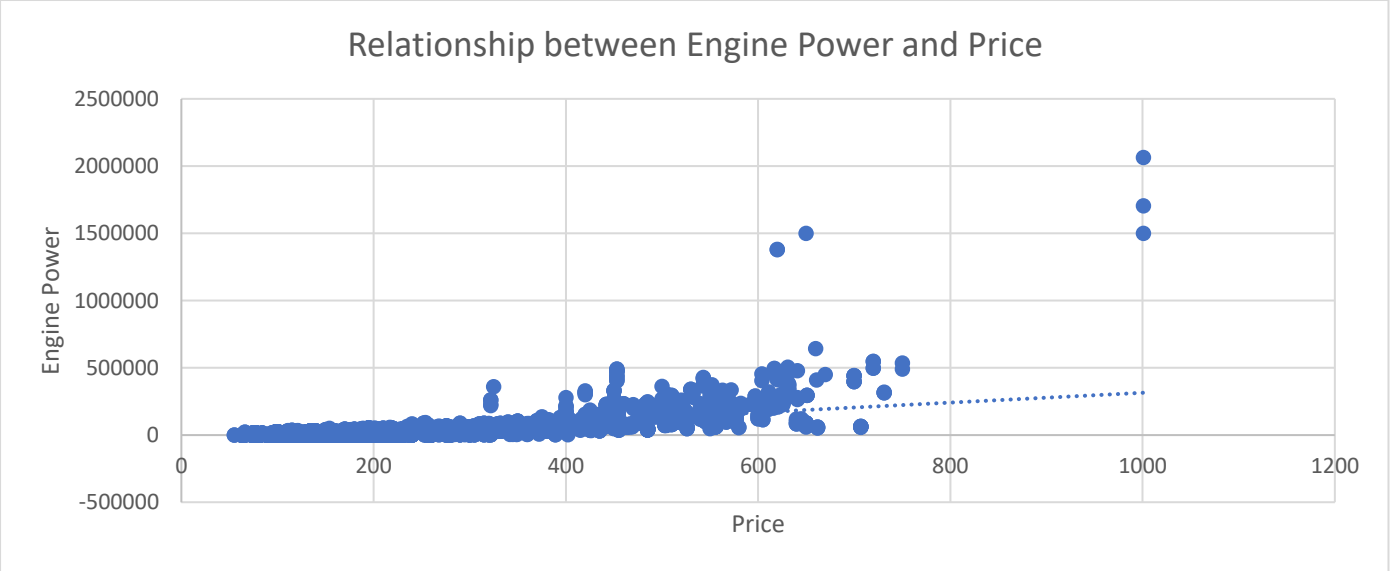
Task 2: Create a scatter chart that plots engine power on the x-axis and price on the y-axis. Add a trendline to the chart to visualize the relationship between these variables.

In this question, we have to create scatter chart of engine power and price also plot the trendline to visualize the relationship between these variables. It helps us to know the relation between the engine power and price.

We already know that Car having high engine power having big engines and these kind of car manufactured very costly and car having least power are becomes less costlier than High engine power car.

Firstly, we extract the engine power and MSRP column from main data and copy to new sheet to better understanding and by using graph function we plotter the scatter plot to know the relationship of Engine power and price of car.

Here's our output :



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.

In this question we have to regression analysis to identify the strongest relationship with a car's price and also we have to create the bar chart to shows the co efficient values for each variable to visualize their importance.

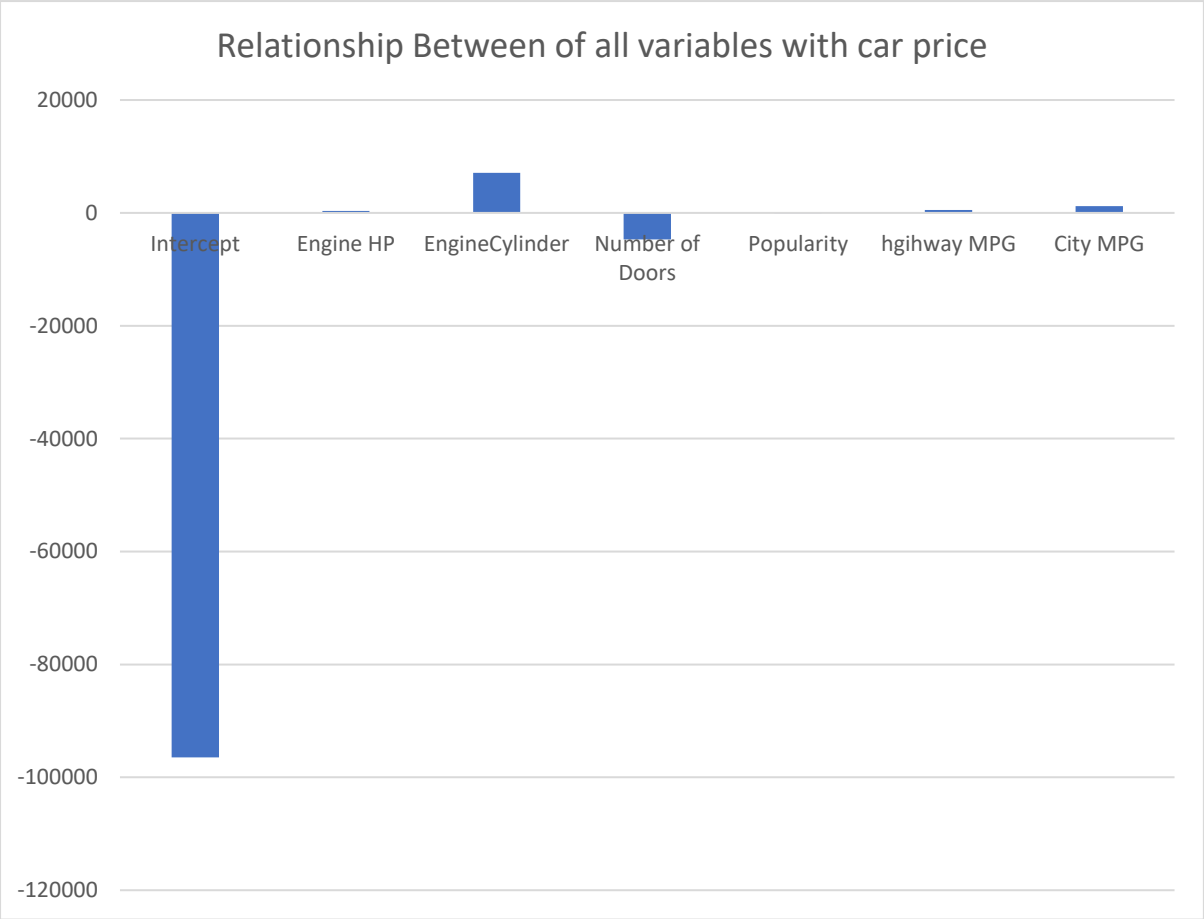
Firstly, we extract the numerical columns from main data and copy to new sheet to better understanding after this we used the data analysis function to do regression analysis after this we used the regression function and we got our output.

The regression helps us to analyze the relationships between one or two variables also Regression allows to predict or explain the variation in one variable based on another variable.

We done regression with many columns like Engine HP, Engine Cylinder, No of doors, Popularity, MPG and their car prices. Also we extracted the regression statistics to know our data properly. After this We also made the correlation table to know which columns are highly correlated.

We made the bar graph to visualize the highly correlated values in our dataset.

Here's our output :



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

In this question, we have to create the pivot table to shows the average price of cars for each manufacturer.

Firstly, we extract the required columns from main data and copy to new sheet to better understanding and we use the pivot table function to create the pivot table also we calculated the average price of cars.

This pivot table helps us to know the manufacturer average car prices and also helps us to buyers which cars they have to buy according to their budget and which one is suitable for them.

Manufacturer	Average of Price
Acura	29285.13095
Alfa Romeo	61600
Aston Martin	79054.04301
Audi	37949.78963
Bentley	166380.5
BMW	51407.43413
Bugatti	23583.33333
Buick	62172.53571
Cadillac	51295.98741

Chevrolet	40130.38924
Chrysler	32197.18182
Dodge	28326.40096
Ferrari	130217.5735
FIAT	22683.64407
Ford	39065.82373
Genesis	15845
GMC	38481.86214
Honda	42211.64206
HUMMER	28306.76471
Hyundai	45528.80858
Infiniti	38846.30909
Kia	26263.98673
Lamborghini	105661.25
Land Rover	28834.77622
Lexus	40433.55941
Lincoln	38316.60897
Lotus	23290.13793
Maserati	45490.86207
Maybach	263759.5625
Mazda	39938.39454
McLaren	225104
Mercedes-Benz	52271.08807
Mitsubishi	35761.3619
Nissan	31162.99635
Oldsmobile	23180.51333
Plymouth	23139.40244
Pontiac	31321.19355
Porsche	76125.38971
Rolls-Royce	27032.80645
Saab	27854.6036
Scion	28104.41667
Spyker	16783.33333
Subaru	58455.14844
Suzuki	32197.73276
Toyota	31186.88156
Volkswagen	33822.29317
Volvo	33791.27758
Grand Total	**40530.36065**

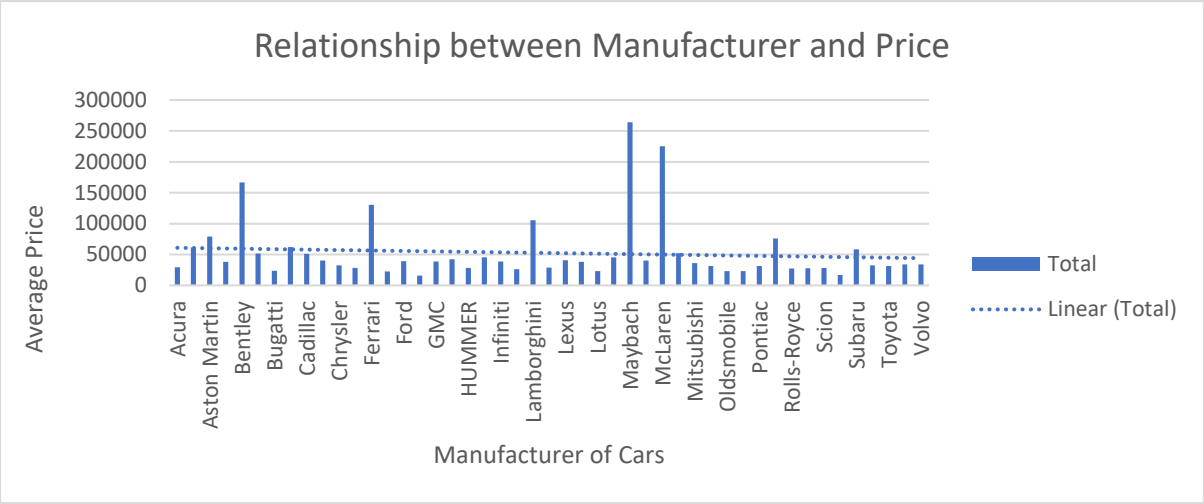
Task 4.B: Create a bar chart or a horizontal stacked bar chart that visualizes the relationship between manufacturer and average price.

In this question, we have to create the chart to visualize the relationship between the manufacturer and their average price.

We already extracted the required column to the new sheet and also created the pivot table, Now we created the bar graph to visualize the relationship of manufacturer and their average price.

Also, this graph helps to visualize to the customers very easily and they can better understand through the graph.

Here's our output :



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.

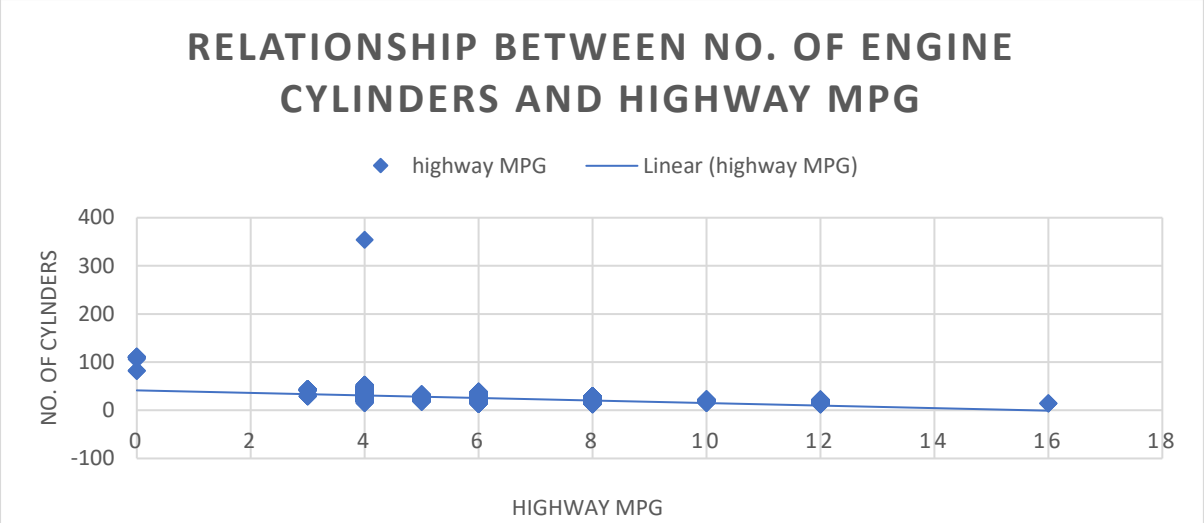
In this question, we have to create the scatter plot of No of cylinders and highway MPG.

Firstly, we extract the No of cylinders and Highway MPG column from the main data and copied to the new sheet to better understanding after this we use the graph function and made the scatter plot to know their relationships.

Also, this scatter plot helps to know their correlations between them.

We know that, the higher no of cylinders having car produce more power and it burns more fuel so higher no of cylinders car having low MPG and least no of cylinders gave more MPG. This scatter plot helps to customers which car giving more or less MPG and which one have to buy according to their requirements.

Here's our output :



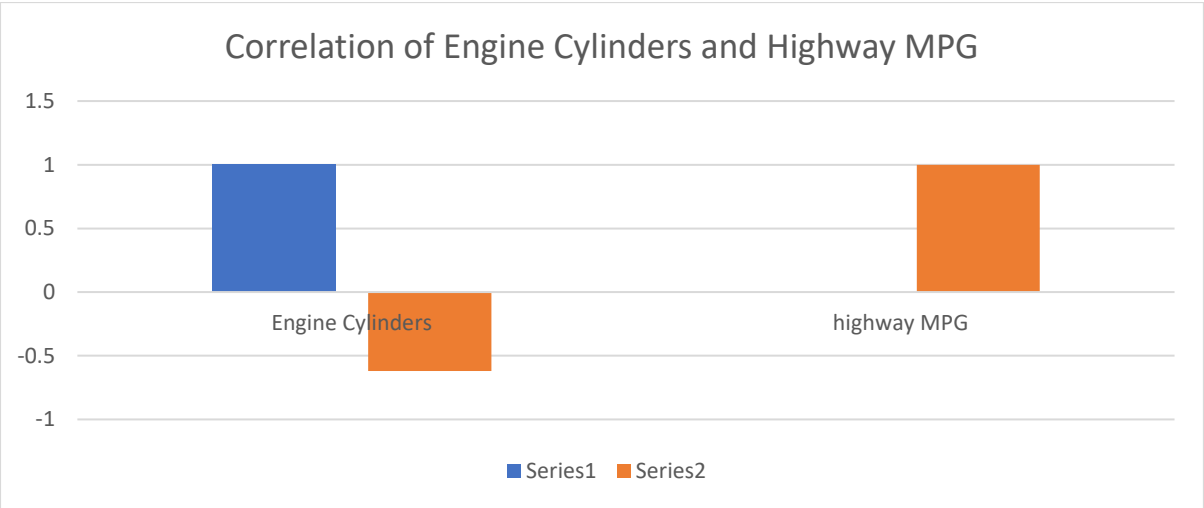
Task 5.B: Calculate the correlation coefficient between the number of cylinders and highway MPG to quantify the strength and direction of the relationship.

In this question, we have to calculate the correlation between No of cylinders and highway MPG and know their relationships.

Firstly, we extract the required columns from main data and copied to new sheet to better understanding. After this we used the correlation function to find the correlation of No of cylinders and highway MPG and also we used the graph function to plot the bar graph and to visualize our output.

This correlation helps us to determine the relationship of no of engine cylinders and Highway MPG also it help us to assess the strength and direction of the linear relationship between pairs of variables.

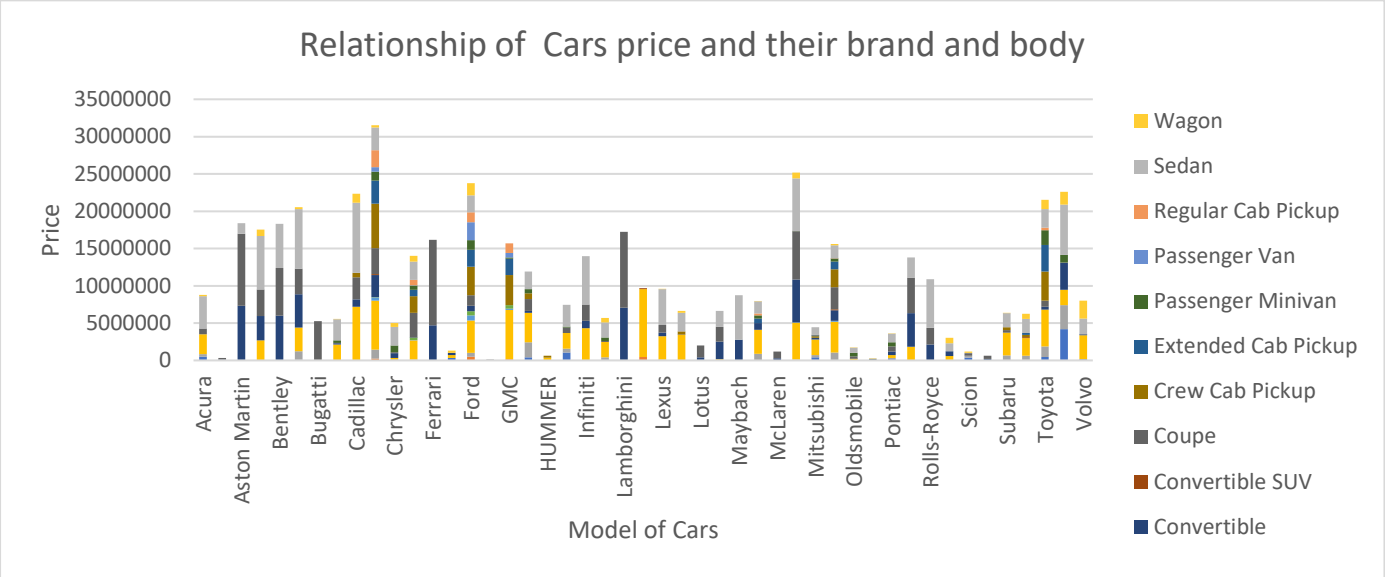
Correlation of Engine Cylinders and Highway MPG		
	Engine Cylinders	*highway MPG*
Engine Cylinders	1	
highway MPG	-0.620312551	1



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

In this question we have to prepare the distribution of car prices vary by brand and their body style. So firstly, we extract the Make, Vehicle Style, MSRP columns from the main data and copied in new sheet to better understanding and make easy to do feature engineering.

After this we used the pivot table of those columns and calculated the sum of MSRP according to their brand and Body style of vehicle to know their relationship then we plot the graph to better visualization our output and also made the slicers of those columns to check one by one according to their requirements.



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

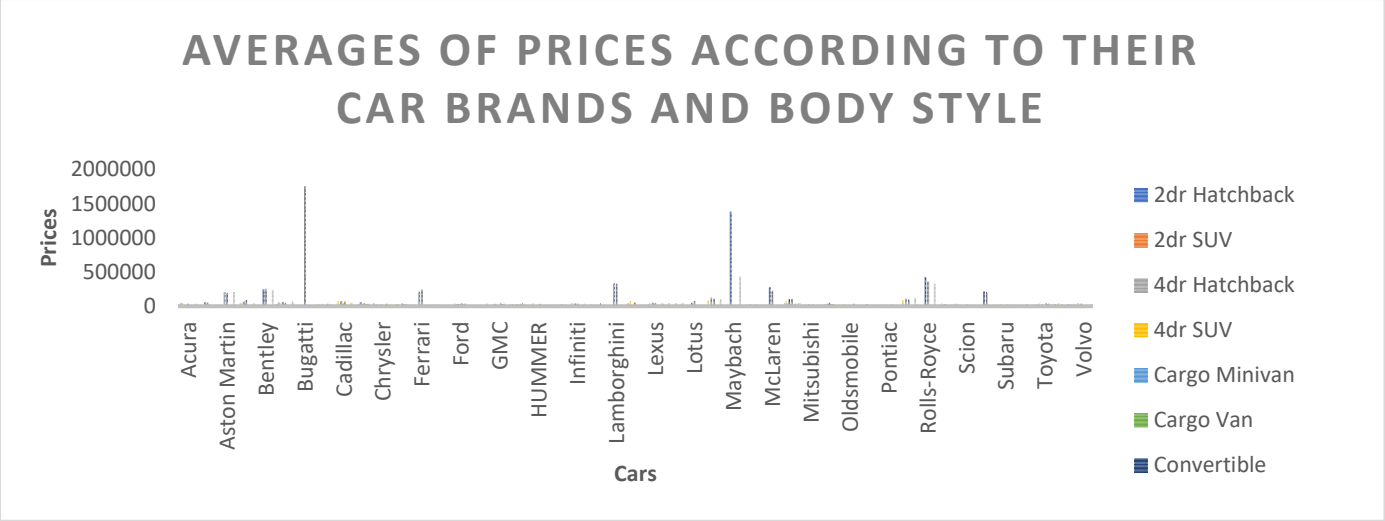
In this question, we have to analyze the car brands their highest and lowest average MSRP and how does they vary by their body style.

Firstly, we extract the Make, Vehicle Style, MSRP from the main data and copied to the new sheet to better understanding our variables which are required then we made a pivot table by using the pivot table function and also calculate the average of MSRP to know the values of car brands according to their body style.

Some car brands have their more values so those car could be more MSRP because they have lots of features which attract the customers and they could buy according to their requirements.

Also, we plotted the graph to better visualization and better understanding to the customers.

Here's our output :

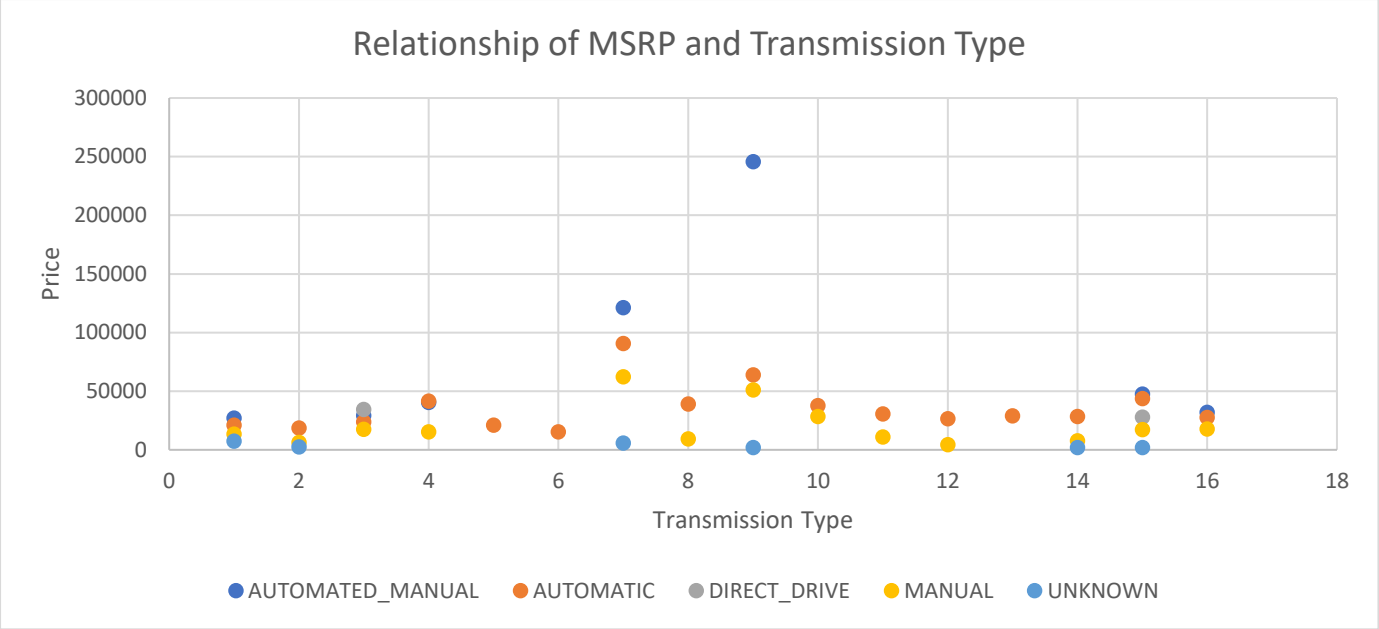


Task 3: How do the different feature such as transmission type affect the MSRP, and how does this vary by body style?

In this question, we have to analyze the different features of cars which affect the MSRP and how they vary by body style.

Firstly, we extract the required columns from the main data and copied to new sheet to better understanding after this we made the pivot table by using the pivot table function and we also calculated the average of MSRP according to their transmission type of cars. Also we plotted the scatter plot to know their relation between them and visualize our variables.

Some customers wants comfort and some wants power in their car so it will vary to customers and their choice which type of cars they have to own. Automatic Transmission cars can easy to drive so it is more prefreble to them. Manual Type cars are less likely to buy because its quiet uncomfort while driving in traffic but automatic cars are easy to handle. So more customers will buy the automatic and their price is more compare to automatic cars.



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

In this question, we have to analyze the fuel efficiency of cars according to their different body styles and their model years.

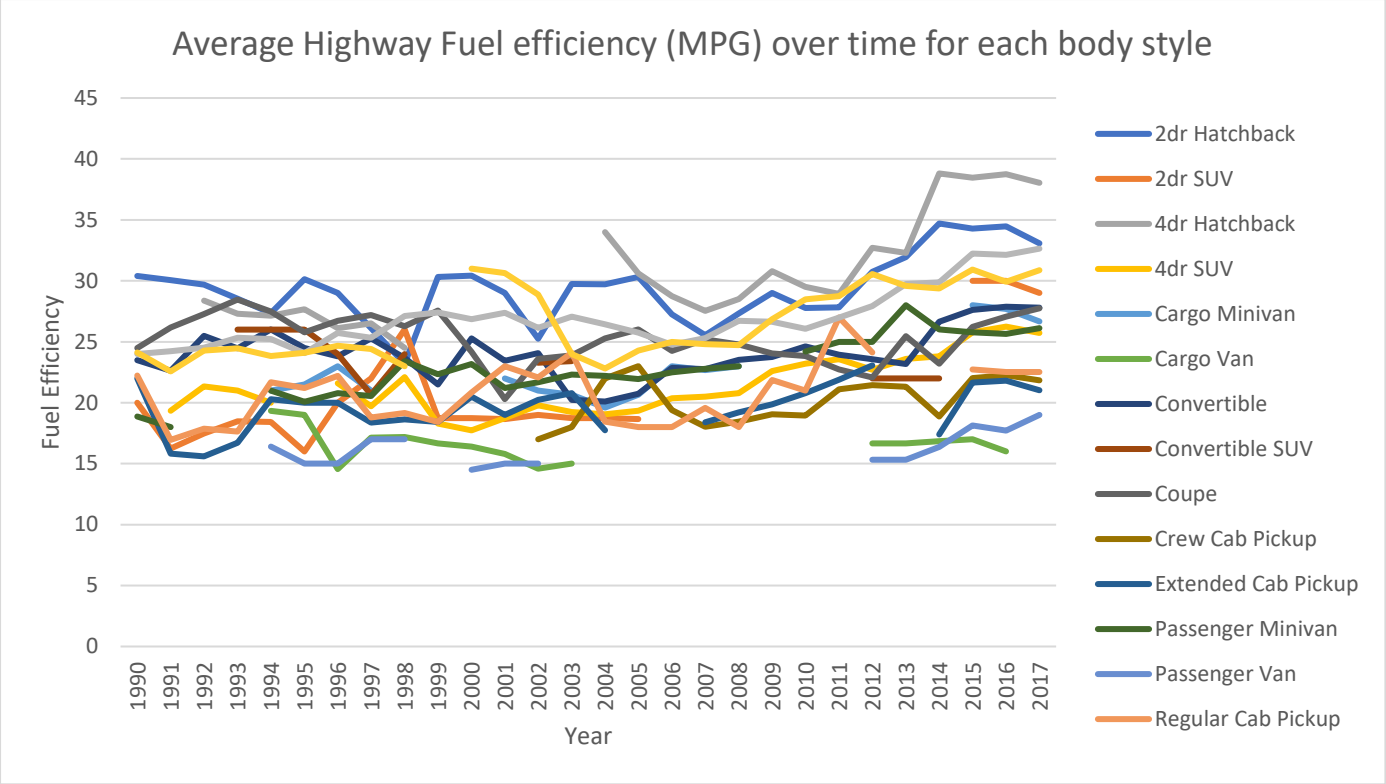
Firstly, we extract the required columns from the main data and copied to new sheet then we made the pivot table with using the pivot table function and also find the average of highway MPG according to their years and body style.

Also we done the same process for city MPG and both are different because in city the cars consume more gasoline due to drive in low and high speed, braking, traffic and we are not maintaining the constant speed and highway MPG are quiet high because we maintaining the constant speed and in this condition the gasoline will less burns. We analyze both separately.

The High engine powers cars consume more fuel and they gave less MPG compare to Low engine powers cars also the size of cars could be factor in this scenarios the SUV's Cars consume more gasoline due to their sizes and their weights and they are more comfortable. Old cars also consume more gasoline due to their ages.

Customers buys cars according to their prefrences and their choice which type of car they want according to their purposes. This graph helps us to better visualizations and to better understanding

Here's our output :



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

In this question we have to analyze the relationship between the car's horsepower, MPG and their price across different brands.

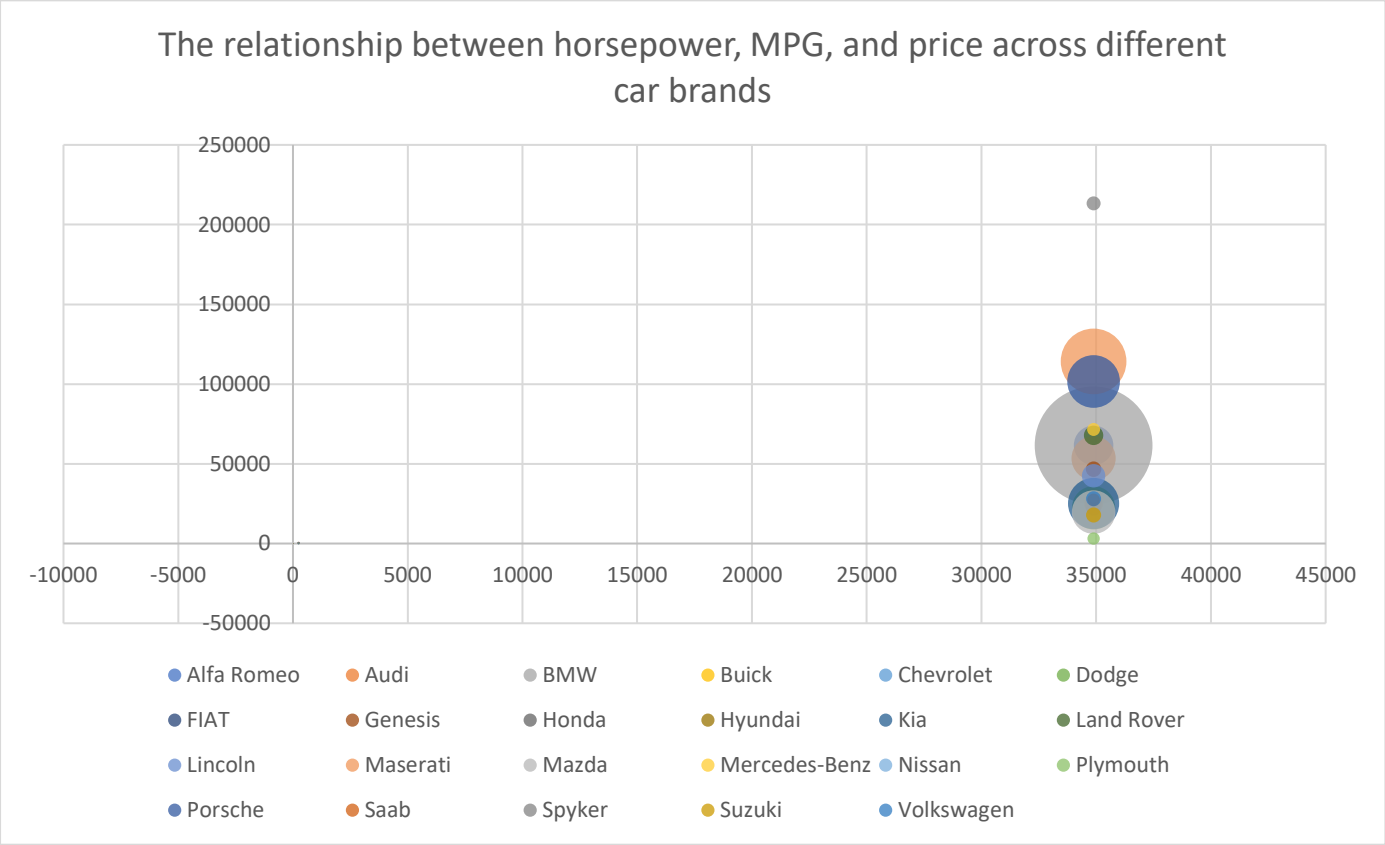
Firstly, we extract the required columns from the main data to the new sheet to better understanding after this we made the pivot table by using the pivot table functions and also calculate the average of Engine HP, Highway MPG and MSRP.

We plotted the Bubble Chart to better visualizations and to better understanding to the customers.

We know that the high horse power car produces more power also their MPG are too low because high burn of gasoline and these type of cars having high price.

There are many brands and they are making the high horsepower car to those types of customers for sales and earn profits but some customers are focusing only decent powers and better MPG with their good prices they can afford this type of car.

Here's our output :



RESULT

- From this project, we learned and focused on various functions and applying the data analysis ways and techniques to solve the questions and extract the valuable insights. Also I

learned the regression analysis to understand the relationships between variables and to extract the required output.

- In this project, I dive into new features and learned how to use slicers regression and many more from this project.
- Also I learned how the Data Analyst derived the valuable insights and conclusions from the data analysis process and Regressions analysis helps us to understand the strength of relationship between variables.
- I learned how data analysts interested in exploring various aspects of the automotive industry and could provide insights that could inform decisions related to product development, marketing, and pricing.
- I learned how a data analyst could identify how car features and prices have changed over time, which could help manufacturers make informed decisions about product development and pricing.
- Also I learned how to deal with the automotive industry has been rapidly evolving over the past few decades, with a growing focus on fuel efficiency, environmental sustainability, and technological innovation.
- I learnt how to deal with null or missing values and how to clean the data it can helps to enhance the better feature engineering.
- Also, I gained skills how to use pivot table and how to visualize by using graph, chart, scatter plot to the given output to better understanding the data.

Linkedin Profile : https://www.linkedin.com/in/pawan-kashyap-832515230

Instagram : https://www.instagram.com/pawankkkashyap

THANK YOU