Name: Nikhil Zodape

Email Id.: Nikszodape@gmail.com

Project Name: Analyzing the Impact of Car Features on Price and Profitability

Using: Python, Excel and Tableau.

Problem Statement:

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 Overview:

- Th dataset provides details about the various car features like the Company, Year and Manufacture, Engine Type and Power etc.

The brief overview of the Dataset:

- **Number of observation:** 11199
- **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

Data Pre-Processing:

1. **Checking the data type of dataset.**

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 11199 entries, 0 to 11198
Data columns (total 16 columns):
 #   Column             Non-Null Count  Dtype
---  ------             --------------  -----
 0   Make               11199 non-null  object
 1   Model              11199 non-null  object
 2   Year               11199 non-null  int64
 3   Engine Fuel Type   11196 non-null  object
 4   Engine HP          11130 non-null  float64
 5   Engine Cylinders   11199 non-null  int64
 6   Transmission Type  11199 non-null  object
 7   Driven_Wheels      11199 non-null  object
 8   Number of Doors    11193 non-null  float64
 9   Market Category    7823 non-null   object
 10  Vehicle Size       11199 non-null  object
 11  Vehicle Style      11199 non-null  object
 12  highway MPG        11199 non-null  int64
 13  city mpg           11199 non-null  int64
 14  Popularity         11199 non-null  int64
 15  MSRP               11199 non-null  int64
dtypes: float64(2), int64(6), object(8)
memory usage: 1.4+ MB
```

2. **Checking for Duplicates values in the dataset:**
   - There are no duplicate values present in the dataset.

3. Checking the Null Values:

```
      💡 Click here to ask Blackbox to help you code faster
   1  round(df.isnull().sum() / len(df) * 100,2)

Make                    0.00
Model                   0.00
Year                    0.00
Engine Fuel Type        0.03
Engine HP               0.62
Engine Cylinders        0.00
Transmission Type       0.00
Driven_Wheels           0.00
Number of Doors         0.05
Market Category        30.15
Vehicle Size            0.00
Vehicle Style           0.00
highway MPG             0.00
city mpg                0.00
Popularity              0.00
MSRP                    0.00
dtype: float64
```

- There are fer nan values present in the dataset so we will fill those Null values accordingly with meaning full insight and also, we will be conducting some own research to fill those Null values so that we can't lose much meaningful data.
- Also we can see there are 30% Null values present in Market Category we will be dropping those Null values from the dataset.

4. Now we will export the data as xlsx to do the further analysis and also build the Dashboard with meaningful insights.

Task Analysis:

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.

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

Result:
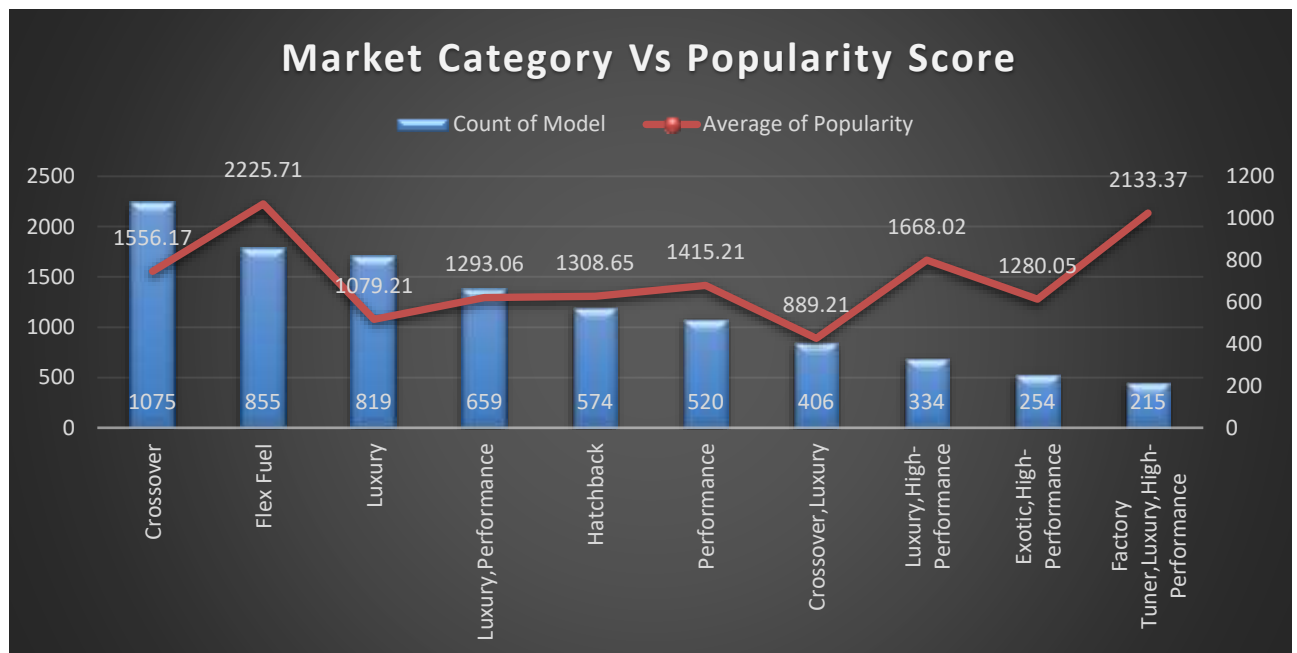
The market category "Flex Fuel" stands out as the most popular, average total of 2225.71 models. It is closely trailed by the "Crossover" market category, which has a popularity of 1556.17.

Although the market category "Flex Fuel" has highest popularity than "crossover" the models presents in the market for Crossover is more than Flex Fuel.

The market category "luxury", "luxury performance", and "performance" have relatively lower popularity, but they have a significant number of models present in the market.

Insight Required: What is the relationship between a car's engine power and it's 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.

Result:
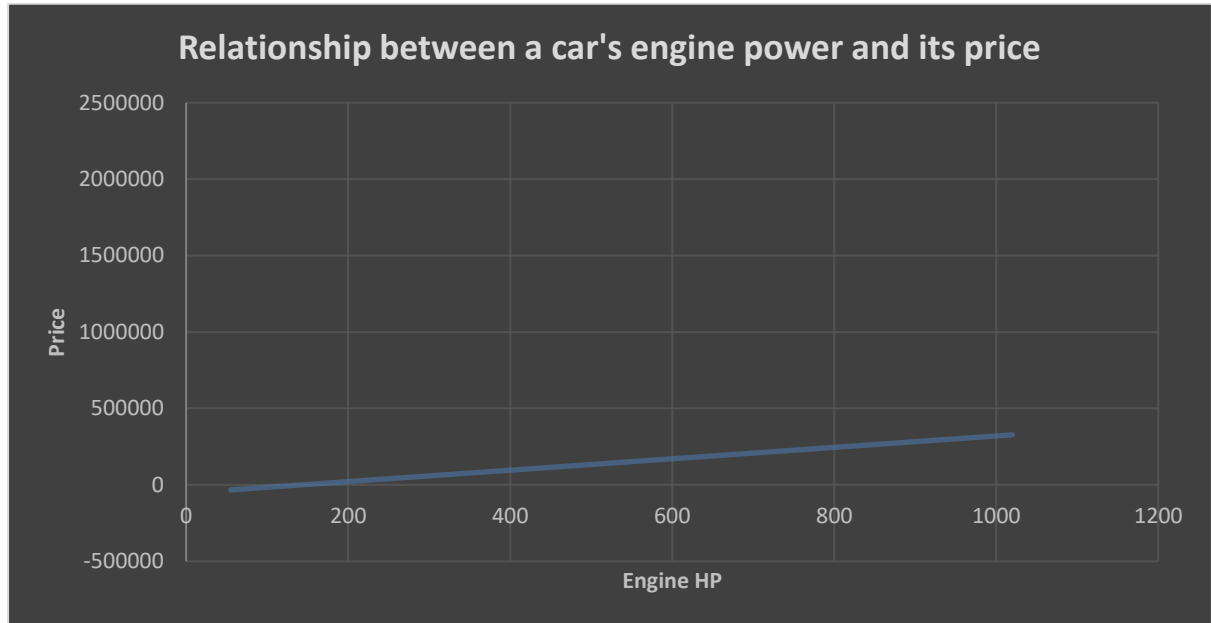
We can infer a positive relationship since the trendline exhibits a positive slope. This aligns with logical expectations, as higher Engine Horsepower (HP) necessitates a more intricate level of design and engineering, leading to increased costs for sub-parts. Additionally, vehicles with higher Engine HP are predominantly associated with performance cars.

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.

Result:

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.696746931
R Square	0.485456286
Adjusted R Square	0.48473167
Standard Error	51135.20388
Observations	7823

ANOVA

	df	SS	MS	F	Significance F
Regression	11	1.92697E+13	1.75179E+12	669.9490958	0
Residual	7811	2.04243E+13	2614809076		
Total	7822	3.9694E+13			

|	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	-58733.64419	6684.662671	-8.786328806	1.87684E-18	-71837.37277	-45629.9156	-71837.37277	-45629.9156
Engine Fule Type	746.8093269	364.7850649	2.047258506	0.040666093	31.73293231	1461.885722	31.73293231	1461.885722
Transmission Type	-8572.579483	730.5164971	-11.73495673	1.55063E-31	-10004.58741	-7140.571561	-10004.58741	-7140.571561
Driven_Wheels	-2885.602914	541.2036605	-5.331824459	9.99514E-08	-3946.50699	-1824.698838	-3946.50699	-1824.698838
Vehicle Size	-19942.41293	1019.131416	-19.56804846	2.7918E-83	-21940.18337	-17944.6425	-21940.18337	-17944.6425
Vehicle Style	125.4090033	129.8068308	0.966120215	0.334013941	-129.0471395	379.8651461	-129.0471395	379.8651461
Engine HP	262.4406879	7.388637221	35.51949839	2.6394E-256	247.9569807	276.9243951	247.9569807	276.9243951
Engine Cylinders	14798.02218	550.7292063	26.86987001	3.4767E-152	13718.44549	15877.59888	13718.44549	15877.59888
Number of Doors	-2228.541744	813.9494902	-2.737936163	0.006196612	-3824.100672	-632.9828166	-3824.100672	-632.9828166
highway MPG	265.7357103	132.8805948	1.999808254	0.045555541	5.254167043	526.2172535	5.254167043	526.2172535
city mpg	725.5427246	123.2071809	5.888802253	4.05109E-09	484.0236625	967.0617866	484.0236625	967.0617866
Popularity	-3.070535328	0.409523604	-7.497822582	7.20146E-14	-3.873311238	-2.267759418	-3.873311238	-2.267759418
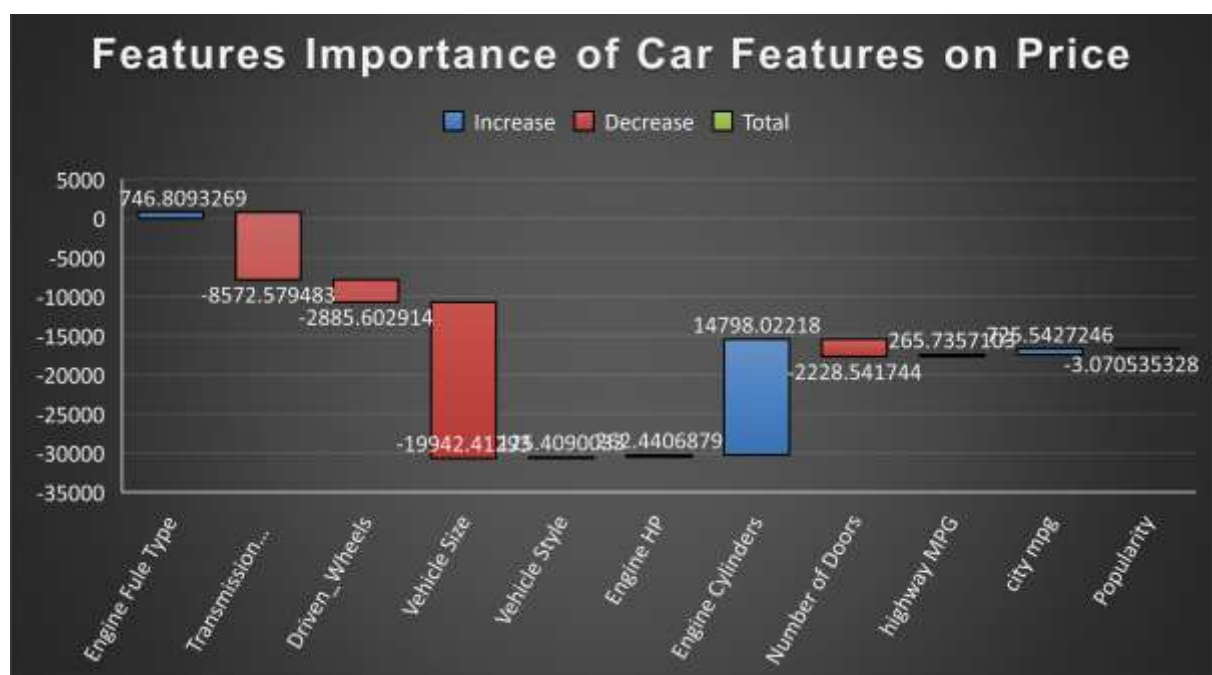
Using regression analysis, we found the top columns. The car features that has the least importance on the price of the vehicle is "Vehicle Size". And the most importance column is "Engine Cylinders".

Insight Required: How does the average price of a car vary across different manufactures?

- **Task 4.A:** Create a pivot table that shows the average price of cars for each manufacturer.
- **Task 4.B:** Create a bar chart or a horizontal stacked bar chart that visualizes the relationship between manufacturer and average price.

Result:
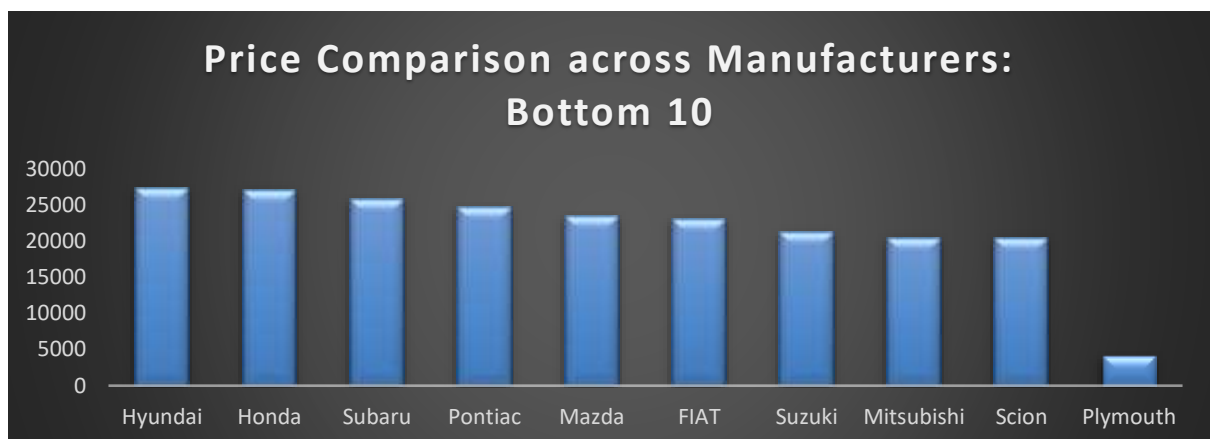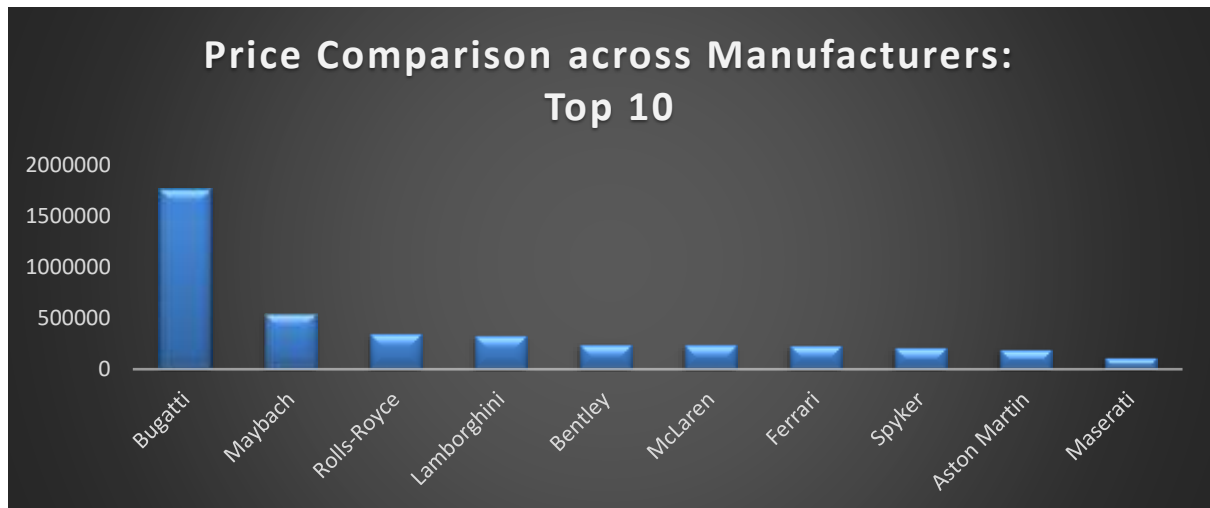
The manufacturers "Bugatti," "Maybach," and "Rolls Royce" stand out for having the highest average prices for cars. Among them, "Bugatti" boasts the highest average price range, while "Plymouth" has the lowest average price range among all manufacturers.

Upon observation, it becomes evident that Bugatti cars are the most expensive, followed by Maybach, Rolls-Royce, Lamborghini, and others. Notably, these brands are renowned for producing high-performance and luxury vehicles.

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

Result:
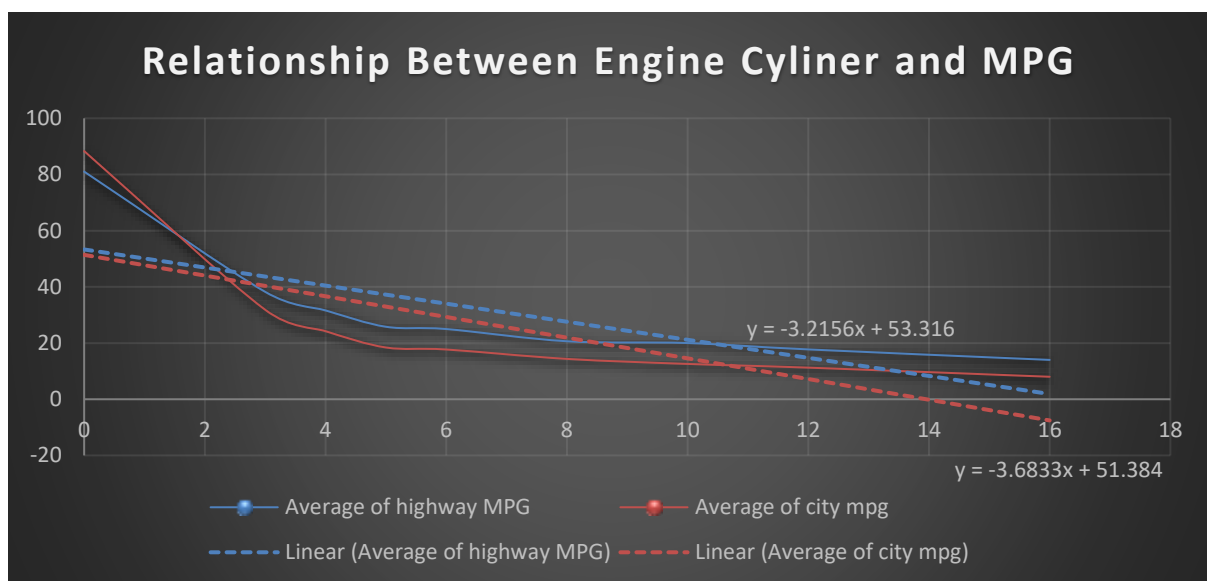
	CORRELATION	
	Engine Cylinders	*Average of highway MPG*
Engine Cylinders	1	
Average of highway MPG	-0.780403825	1

There is an inverse relationship between the number of engine cylinders and the average highway MPG.

Cars with fewer engine cylinders tend to have higher average highway MPG, while cars with more engine cylinders have lower average highway MPG.
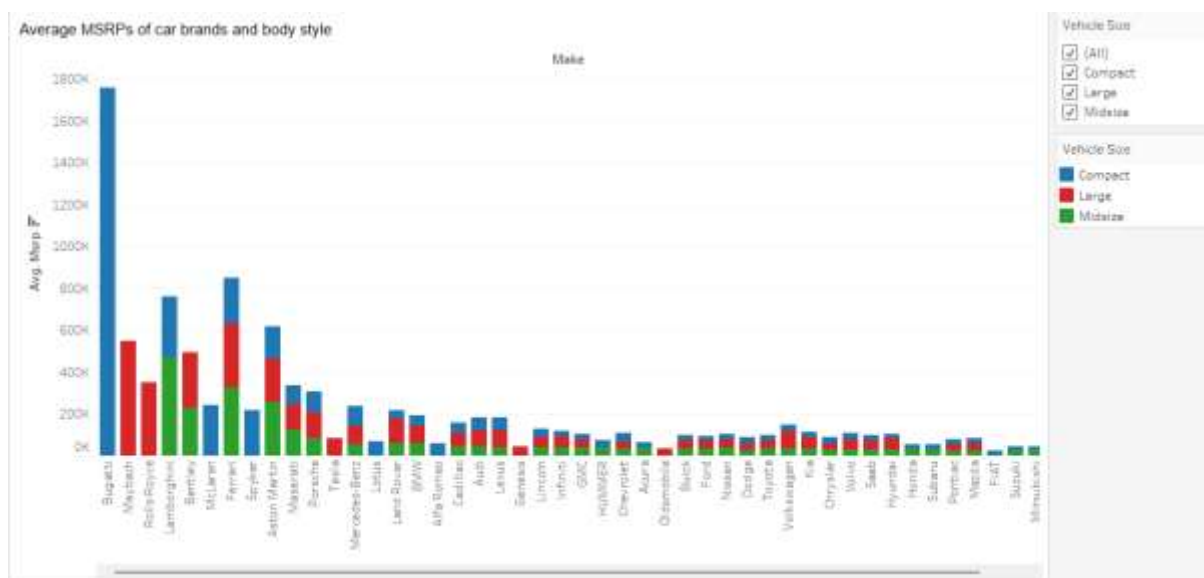
Task Dashboard:

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

Result:


Distribution of car prices by brand and body style
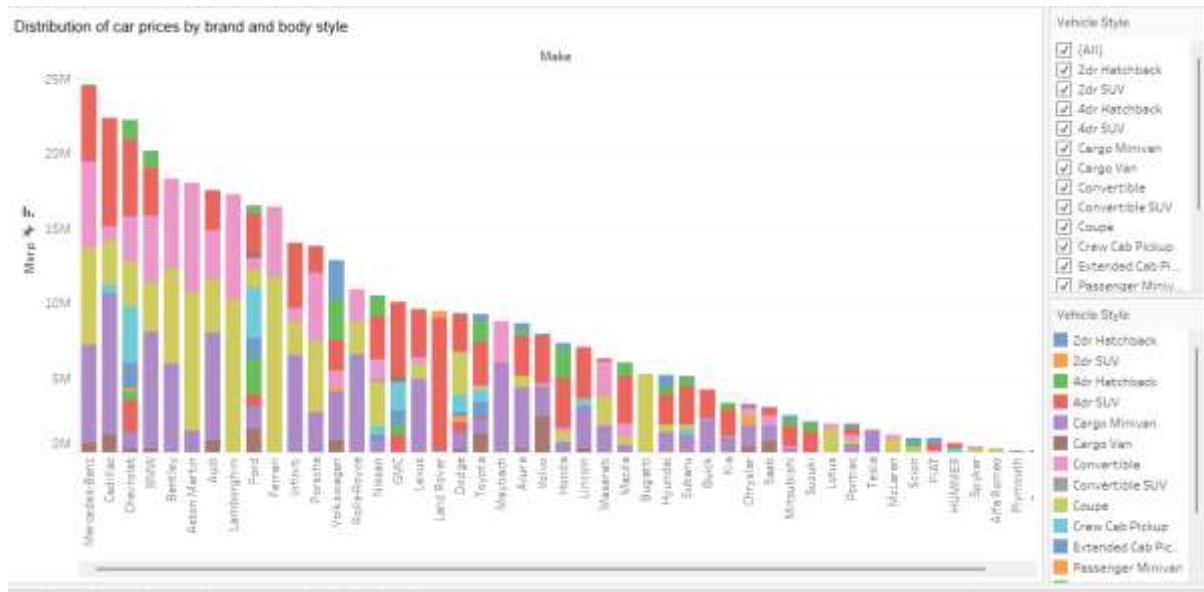
2. Which car brands have the highest and lowest average MSRPs, and how does this vary by body style?

Result:

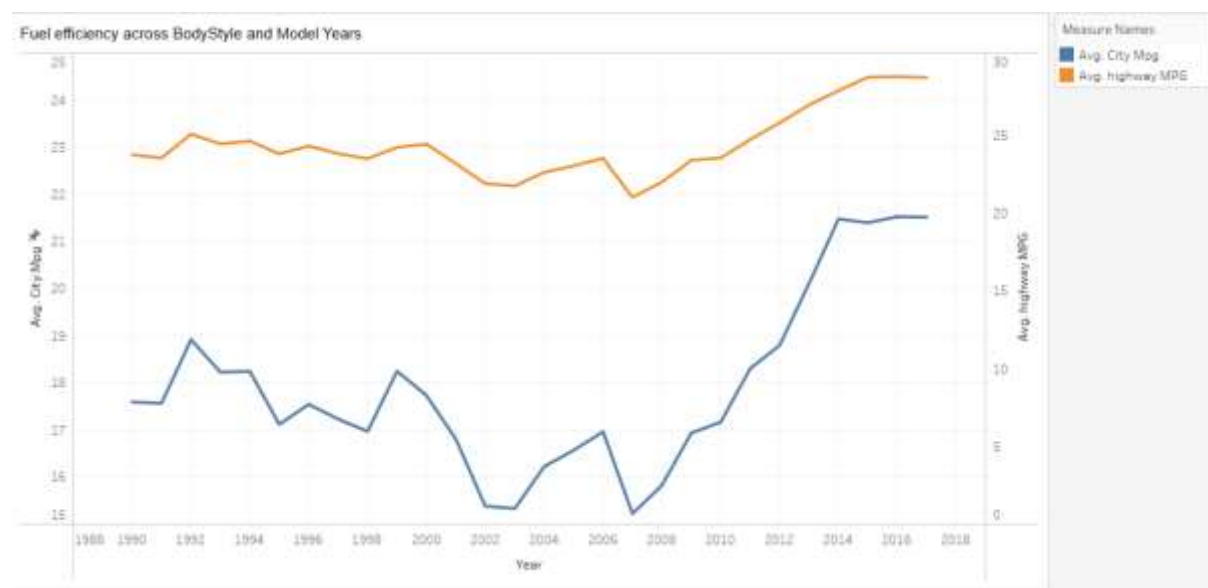Average MSRPs of car brands and body style

3. How does the difference features such as transmission type affect the MSRP, and how does this vary by body style?
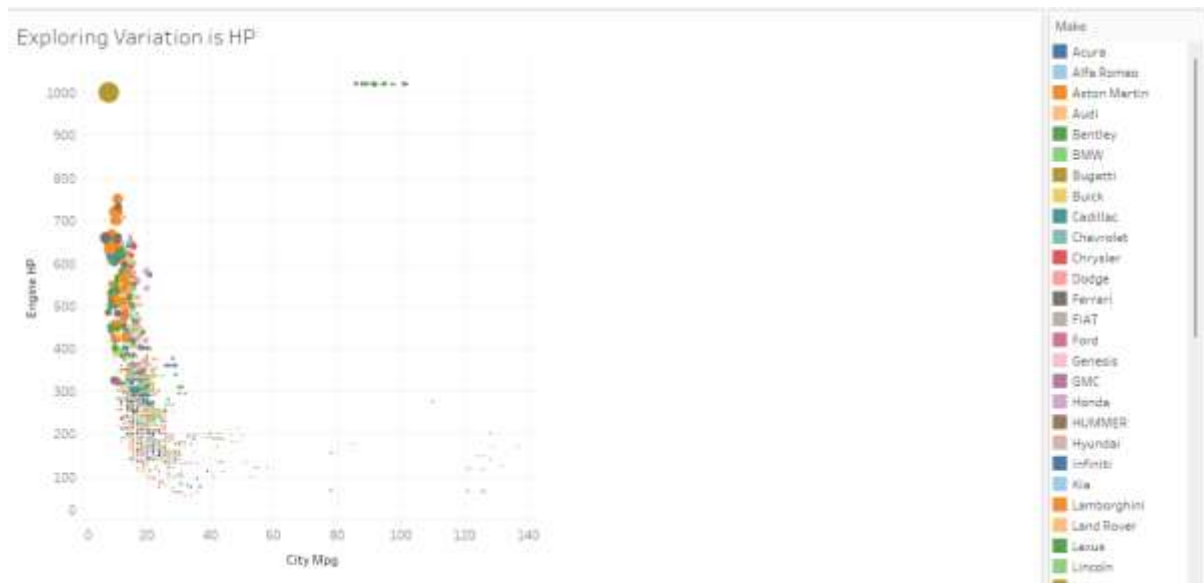
Result:

Impact of Transmission Type on MSRP by Body Style
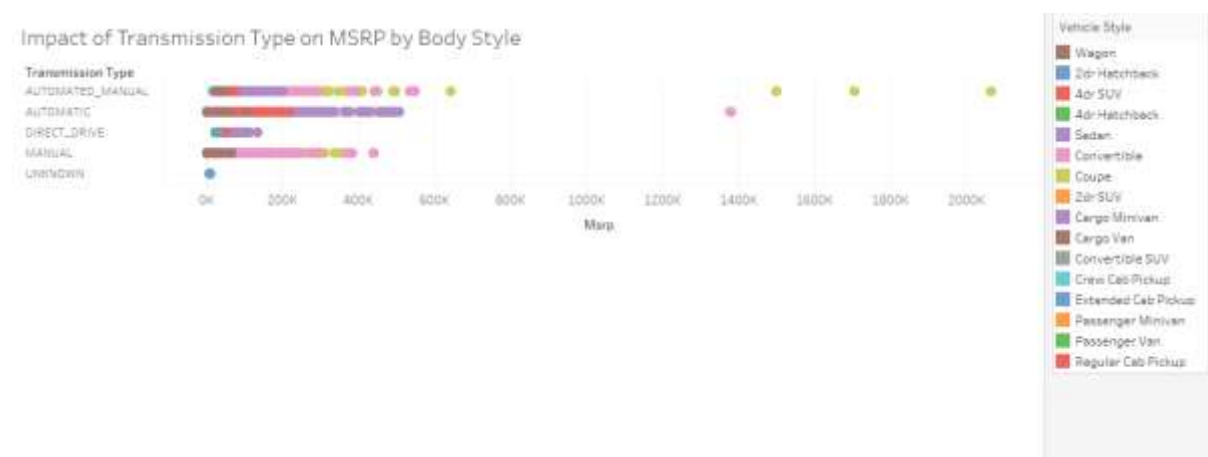
4. How does the fuel efficiency of cars vary across different body styles and model years?
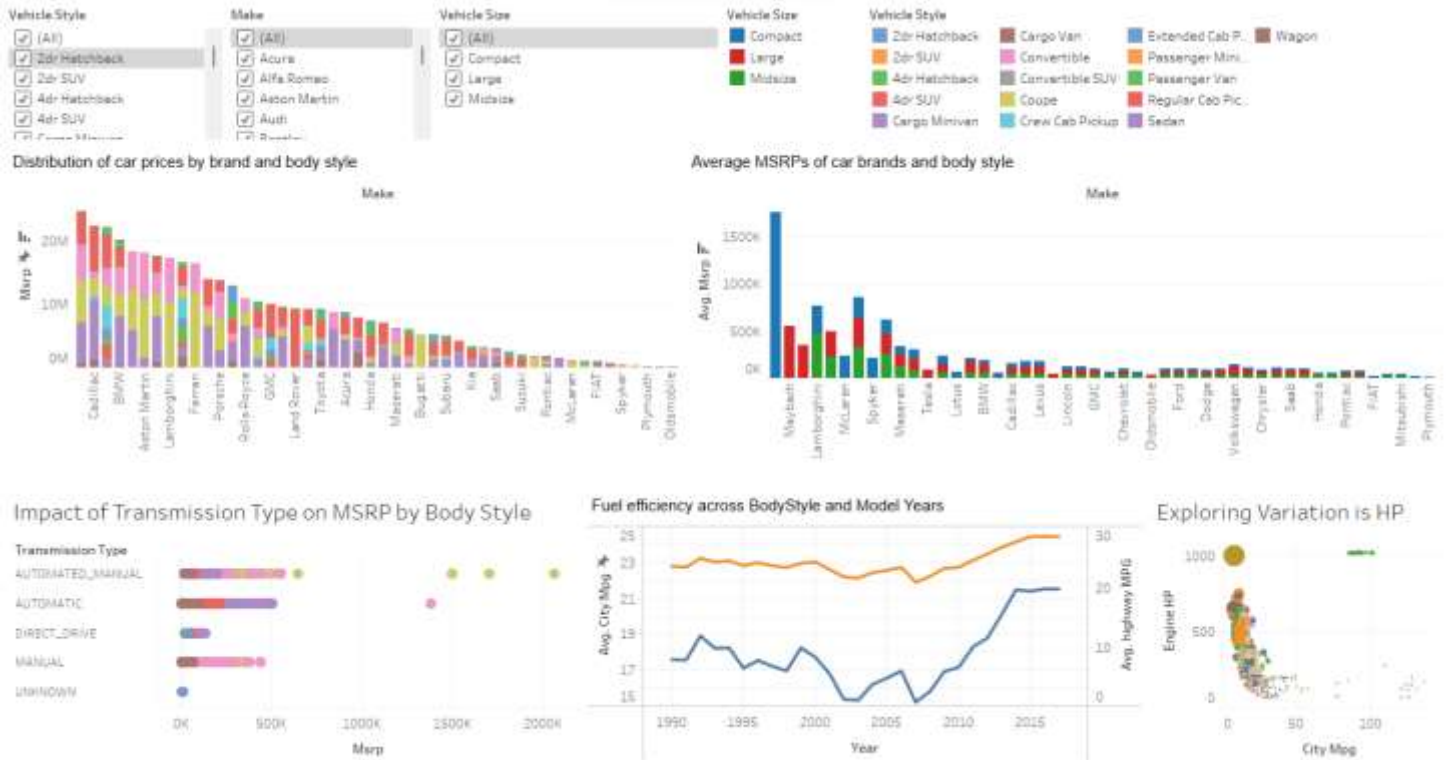
Result:

Fuel efficiency across BodyStyle and Model Years

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

Result:

DASHBOARD

Link of complete Python workbook:
https://drive.google.com/file/d/1Hq_DIdWjMeXngvYToJuTmUjg8vvHILfy/view?usp=drive_link

Link of complete Excel workbook:
https://docs.google.com/spreadsheets/d/1ozXDaXEvh_xZR1chQb59ImfosW3Xuudr/edit?usp=drive_link&ouid=107933073293773244015&rtpof=true&sd=true

Link of complete Tableau Dashboard:
https://public.tableau.com/views/ImpactofCarFeaturesonPriceandProfitability_17060832707050/Dashboard1?:language=en-US&publish=yes&:display_count=n&:origin=viz_share_link

Video Link: https://drive.google.com/file/d/1BQzkMrl-b-XEvwa7C5EGmE6Q1r42-lNj/view?usp=drive_link