**Project**

***Analyzing the Impact of Car Features on Price and Profitability***

[**Python Notebook**](https://colab.research.google.com/drive/15DChVlV0_-U36kVgUZITixSDxxMfT3C-?usp=sharing)

[**Excel file: Analysis of car data set**](https://docs.google.com/spreadsheets/d/1aYad2iXC4xhKAIP11wgYw7dGXuawAVdF/edit?usp=sharing&ouid=102980832699252360079&rtpof=true&sd=true)

[**Presentation Link**](https://drive.google.com/file/d/1ahmWVwGHr6aLhpAudWG2m7f4ZsJhgkvI/view?usp=sharing)

# **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.

# **Business and Data Analytics Skills:**

The given tasks below based on the business problem would require advanced Excel skills and knowledge of data analysis techniques such as regression analysis, pivot tables, sensitivity analysis, optimization, and time series analysis.

However, by answering these questions and building an interactive dashboard, a data analyst could provide valuable insights to a car manufacturer and help them optimize their pricing and product development decisions to maximize profitability while meeting consumer demand.

# **Dataset Description:**

**RangeIndex: 11914 entries, 0 to 11913**

**Data columns (total 16 columns):**

**# Column Non-Null Count Dtype**

**--- ------ -------------- -----**

**0 Make 11914 non-null object**

**1 Model 11914 non-null object**

**2 Year 11914 non-null int64**

**3 Engine Fuel Type 11911 non-null object**

**4 Engine HP 11845 non-null float64**

**5 Engine Cylinders 11884 non-null float64**

**6 Transmission Type 11914 non-null object**

**7 Driven\_Wheels 11914 non-null object**

**8 Number of Doors 11908 non-null float64**

**9 Market Category 8172 non-null object**

**10 Vehicle Size 11914 non-null object**

**11 Vehicle Style 11914 non-null object**

**12 highway MPG 11914 non-null int64**

**13 city mpg 11914 non-null int64**

**14 Popularity 11914 non-null int64**

**15 MSRP 11914 non-null int64**

**dtypes: float64(3), int64(5), object(8)**

* **Data Cleaning And EDA**

**Remove the duplicates row entries from Dataset**

* **There is 715 duplicates entries are present in dataset**
* **715 duplicates entries are remove from dataset**
* **Null Value Analysis**

| **Column Name** | **Null Count** | **Null %** |
| --- | --- | --- |
| **Make** | **0** | **0** |
| **Model** | **0** | **0** |
| **Year** | **0** | **0** |
| **Engine Fuel Type** | **3** | **0.02518** |
| **Engine HP** | **69** | **0.579151** |
| **Engine Cylinders** | **30** | **0.251805** |
| **Transmission Type** | **0** | **0** |
| **Driven\_Wheels** | **0** | **0** |
| **Number of Doors** | **6** | **0.050361** |
| **Market Category** | **3742** | **31.408427** |
| **Vehicle Size** | **0** | **0** |
| **Vehicle Style** | **0** | **0** |
| **highway MPG** | **0** | **0** |
| **city mpg** | **0** | **0** |
| **Popularity** | **0** | **0** |
| **MSRP** | **0** | **0** |

**Results :**

**Numerical columns containing null Values**

**Engine HP**

**Engine Cylinders**

**Number of Doors**

**Categorical Columns containing Null values**

**Engine Fuel Type**

**Market Category**

* **Handling missing Values**

**In Numerical columns**

**Numerical columns null values are less than 1%**

**Removed the null value entries**

**In Categorical Columns**

**Null values of Engine Fuel Type are removed**

**Null values from Market Category are replace by ‘Other’**

* **Outliers check By box Plot**

|  |  |
| --- | --- |
| **Results : Year 1993 before release**  **Car are to old Doesn't stand current market** | **Results : Engine Hp is greater than 500 HP are not prefer by majority , these are high power cars and to much costly** |
|  |  |
| **Results : most preferable number of cylinders in car are between 3 - 8 .** | **Results : More numbers of customers prefer car price less than 5 lakh** |
|  |  |
| **Results : average Highway MPG is 26.27 and good range of MPG consider between 18-40** | **Results : city MPG of car are less than city MPG average is 19.3**  **Good MPG consider in rage 16-30** |
|  | |
| **Results :Most popularity is between 549 - 2009** | |

| * **Insights :** * **From the Box Plot We Able to Find the Outliers in the dataset** * **Gain the brief Understanding of Data Present and variation in the data Visualize** | |
| --- | --- |
|

# **Tasks: Analysis**

Before diving into the analysis of the given dataset, it is important to perform thorough data cleaning to ensure accurate and reliable results. You need to build an interactive dashboard in Excel from the tasks given below:

**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.**

| **Row Labels** | **Average of Popularity** | **Count of Model** |
| --- | --- | --- |
| **Hatchback,Flex Fuel** | **5657** | **7** |
| **Flex Fuel,Diesel** | **5657** | **16** |
| **Crossover,Flex Fuel,Performance** | **5657** | **6** |
| **Crossover,Luxury,Performance,Hybrid** | **3916** | **2** |
| **Crossover,Factory Tuner,Luxury,Performance** | **2607.4** | **5** |
| **Crossover,Performance** | **2585.956522** | **69** |
| **Crossover,Hybrid** | **2563.380952** | **42** |
| **Diesel,Luxury** | **2416.106383** | **47** |
| **Luxury,Performance,Hybrid** | **2333.181818** | **11** |
| **Flex Fuel** | **2225.71345** | **855** |
| **Crossover,Luxury,Diesel** | **2195.848485** | **33** |
| **Hatchback,Factory Tuner,Performance** | **2173.714286** | **21** |
| **Factory Tuner,Luxury,High-Performance** | **2133.367442** | **215** |
| **Hybrid** | **2116.586777** | **121** |
| **Hatchback,Hybrid** | **2111.15625** | **64** |
| **Crossover,Flex Fuel** | **2073.75** | **64** |
| **Crossover,Hatchback,Factory Tuner,Performance** | **2009** | **6** |
| **Crossover,Hatchback,Performance** | **2009** | **6** |
| **Factory Tuner,High-Performance** | **1966.442308** | **104** |
| **Crossover,Factory Tuner,Luxury,High-Performance** | **1823.461538** | **26** |
| **High-Performance** | **1823.378788** | **198** |
| **Factory Tuner,Performance** | **1818.049383** | **81** |
| **Diesel** | **1730.904762** | **84** |
| **Flex Fuel,Performance** | **1702.358025** | **81** |
| **Crossover,Hatchback** | **1675.694444** | **72** |
| **Luxury,High-Performance** | **1668.017964** | **334** |
| **Other** | **1657.632689** | **3365** |
| **Hatchback,Luxury,Performance** | **1632.25** | **36** |
| **Crossover,Flex Fuel,Luxury,Performance** | **1624** | **6** |
| **Crossover** | **1539.475655** | **1068** |
| **Performance** | **1443.234592** | **503** |
| **Factory Tuner,Luxury,Performance** | **1413.419355** | **31** |
| **Flex Fuel,Luxury,Performance** | **1380.071429** | **28** |
| **Crossover,Luxury,Performance** | **1349.089286** | **112** |
| **Hatchback,Luxury** | **1323.133333** | **45** |
| **Luxury,Performance** | **1293.062215** | **659** |
| **Hatchback** | **1279.113346** | **547** |
| **Exotic,High-Performance** | **1270.326531** | **245** |
| **Hatchback,Factory Tuner,High-Performance** | **1205.153846** | **13** |
| **Crossover,Flex Fuel,Luxury** | **1173.2** | **10** |
| **Luxury** | **1084.21227** | **815** |
| **Hatchback,Performance** | **1073.661616** | **198** |
| **Exotic,Factory Tuner,High-Performance** | **1046.380952** | **21** |
| **Crossover,Luxury,High-Performance** | **1037.222222** | **9** |
| **Flex Fuel,Luxury,High-Performance** | **898.3125** | **32** |
| **Crossover,Luxury** | **889.2142857** | **406** |
| **Hatchback,Factory Tuner,Luxury,Performance** | **886.8888889** | **9** |
| **Crossover,Diesel** | **873** | **7** |
| **Hatchback,Diesel** | **873** | **14** |
| **Flex Fuel,Luxury** | **746.5384615** | **39** |
| **Luxury,Hybrid** | **724.6875** | **48** |
| **Crossover,Luxury,Hybrid** | **630.9166667** | **24** |
| **Factory Tuner,Luxury** | **617** | **2** |
| **Luxury,High-Performance,Hybrid** | **568.8333333** | **12** |
| **Exotic,Factory Tuner,Luxury,High-Performance** | **523.0196078** | **51** |
| **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,Luxury,High-Performance** | **473.025974** | **77** |
| **Hatchback,Luxury,Hybrid** | **454** | **3** |
| **Flex Fuel,Factory Tuner,Luxury,High-Performance** | **258** | **1** |
| **Crossover,Exotic,Luxury,Performance** | **238** | **1** |
| **Crossover,Exotic,Luxury,High-Performance** | **238** | **1** |
| **Exotic,Luxury,Performance** | **217.0277778** | **36** |
| **Crossover,Factory Tuner,Performance** | **210** | **4** |
| **Crossover,Hatchback,Luxury** | **204** | **7** |
| **Exotic,Luxury,High-Performance,Hybrid** | **204** | **1** |
| **Flex Fuel,Performance,Hybrid** | **155** | **2** |
| **Performance,Hybrid** | **155** | **1** |
| **Flex Fuel,Hybrid** | **155** | **2** |
| **Exotic,Luxury** | **112.6666667** | **12** |
| **Grand Total** | **1556.914685** | **11100** |

**Results :**

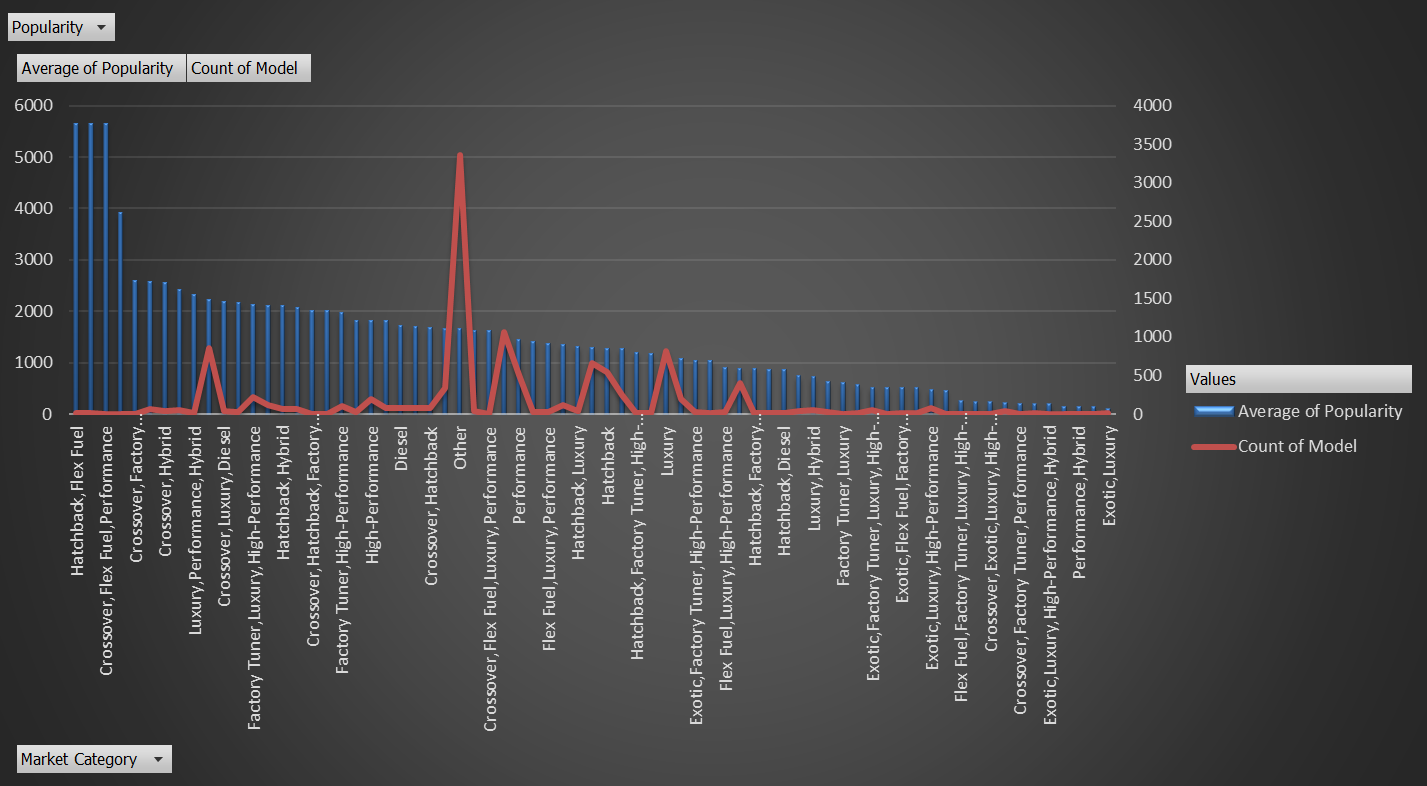
* 1. **car models in each market category and their corresponding popularity scores show in the pivot table.**
  2. **Top 5 market category and their corresponding popularity scores**

| **Row Labels** | **Average of Popularity** | **Count of Model** |
| --- | --- | --- |
| **Hatchback,Flex Fuel** | **5657** | **7** |
| **Flex Fuel,Diesel** | **5657** | **16** |
| **Crossover,Flex Fuel,Performance** | **5657** | **6** |
| **Crossover,Luxury,Performance,Hybrid** | **3916** | **2** |
| **Crossover,Factory Tuner,Luxury,Performance** | **2607.4** | **5** |

**Task 1.B:**

**Create a combo chart that visualizes the relationship between market category and popularity**.

* **Visualization market category and popularity**.**:**



**Results :**

* **combo chart that visualizes the relationship between market category and popularity.**
* **Clear understanding of market categories with popularity score by Visualization of combo chart**

**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.**

* **Visualization**

|  |
| --- |
| **Results : Graph shows as the Engine HP increases MSRP is Increased** |
|  |
| **Results : Graph shows MSRP and The HP variation trendline between Engine HP 100-800 hp** |

* **Insights :**
  + **As the Engine Horsepower increase Price of Car Increase**

**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.

**High coefficients** (either positive or negative) as they indicate large changes in price for changes in the feature.

**Low p-values** (typically < 0.05) indicating strong statistical significance.

**High t-statistics** suggesting the feature is important.

| **Regression Analysis** | **Coefficients** | **t Stat** | **P-value** |
| --- | --- | --- | --- |
| **Year** | **1681.250447** | **49.07272713** | **0** |
| **Engine HP** | **221.4567535** | **151.963588** | **0** |
| **Engine Cylinders** | **9075.136122** | **64.85147414** | **0** |
| **Number of Doors** | **-3509.155523** | **-12.37778723** | **5.94975E-35** |
| **highway MPG** | **-1269.655719** | **-31.65455522** | **2.2619E-210** |
| **city mpg** | **-2037.735128** | **-41.39125312** | **0** |
| **Popularity** | **-0.317007109** | **-1.853992229** | **0.063767542** |

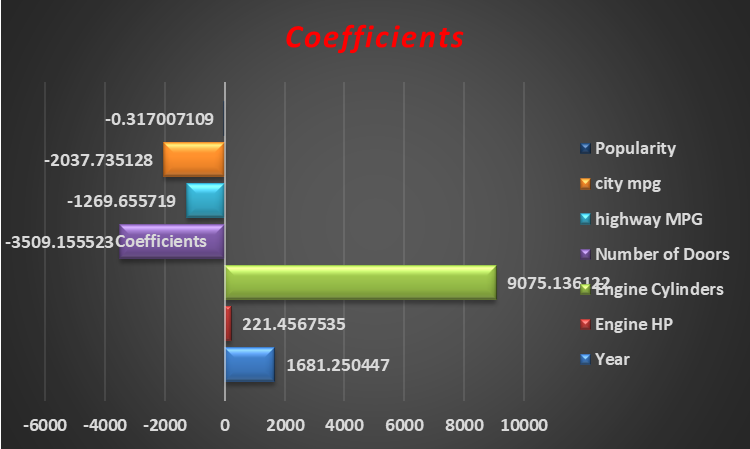
**Results:**

**Based on the Regression Analysis Engine Cylinders, Number of Doors,Year , and Engine HP are more important features**

**Regression Analysis by Coefficients**

| **Regression Analysis** | **Coefficients** |
| --- | --- |
| **Year** | **1681.250447** |
| **Engine HP** | **221.4567535** |
| **Engine Cylinders** | **9075.136122** |
| **Number of Doors** | **-3509.155523** |
| **highway MPG** | **-1269.655719** |
| **city mpg** | **-2037.735128** |
| **Popularity** | **-0.317007109** |

* **Visualization of Coefficients**

****

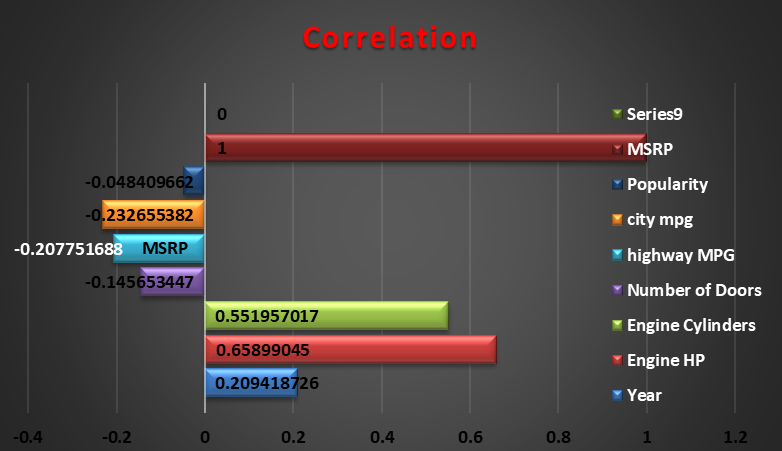
**Results :**

**More coefficient variation based on Engine Cylinders as the Number of Engine Cylinder Increase More Significant change in Cars Prices**

**Correlation with MSRP**

| **Correlation** | **MSRP** |
| --- | --- |
| **Year** | **0.209419** |
| **Engine HP** | **0.65899** |
| **Engine Cylinders** | **0.551957** |
| **Number of Doors** | **-0.14565** |
| **highway MPG** | **-0.20775** |
| **city mpg** | **-0.23266** |
| **Popularity** | **-0.04841** |
| **MSRP** | **1** |

* **Visualization of Correlation of MSRP**



**Results : Engine Hp , Engine Cylinders and Year are highly correlated**

**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.**

| **Manufacturer** | **Average of MSRP** |
| --- | --- |
| Bugatti | 1757223.667 |
| Maybach | 546221.875 |
| Rolls-Royce | 351130.6452 |
| Lamborghini | 331567.3077 |
| Bentley | 247169.3243 |
| McLaren | 239805 |
| Ferrari | 237383.8235 |
| Spyker | 214990 |
| Aston Martin | 198123.4615 |
| Maserati | 113684.4909 |
| Porsche | 101622.3971 |
| Mercedes-Benz | 72135.02647 |
| Lotus | 68377.14286 |
| Land Rover | 68067.08633 |
| BMW | 62162.55864 |
| Alfa Romeo | 61600 |
| Cadillac | 56368.26515 |
| Audi | 54574.1215 |
| Lexus | 47549.06931 |
| Genesis | 46616.66667 |
| Lincoln | 43560.01316 |
| Infiniti | 42640.27134 |
| HUMMER | 36464.41176 |
| Acura | 35087.4878 |
| GMC | 32444.08506 |
| Volvo | 29724.68421 |
| Buick | 29034.18947 |
| Chevrolet | 29000.2214 |
| Volkswagen | 28947.36879 |
| Nissan | 28856.42329 |
| Toyota | 28758.76676 |
| Ford | 28522.86207 |
| Saab | 27879.80734 |
| Chrysler | 26722.96257 |
| Honda | 26608.88399 |
| Kia | 25318.75 |
| Hyundai | 24926.26255 |
| Dodge | 24857.04537 |
| Subaru | 24240.67364 |
| FIAT | 22206.01695 |
| Mitsubishi | 21316.35122 |
| Mazda | 20106.55612 |
| Scion | 19932.5 |
| Pontiac | 19800.0442 |
| Suzuki | 18026.4152 |
| Oldsmobile | 12843.79545 |
| Plymouth | 3296.873239 |
| **Grand Total** | **41894.83009** |

**Results :**

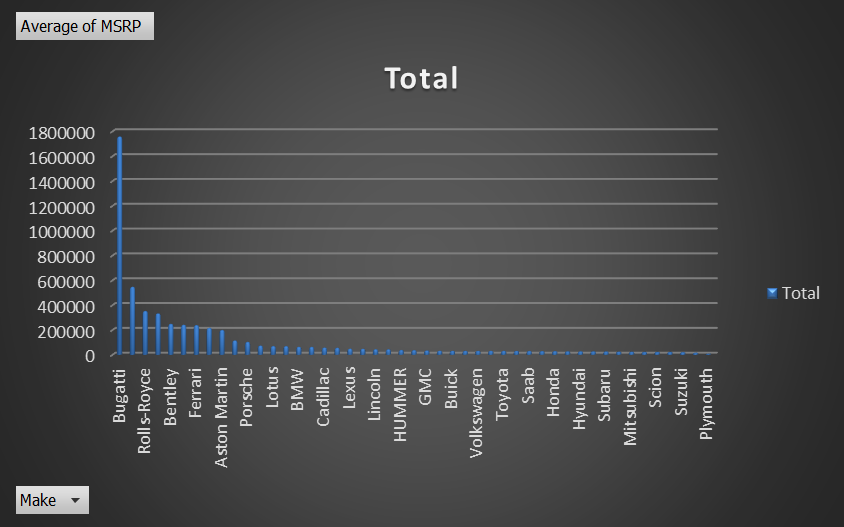
* **Pivot table describe the decreasing average price of manufacturer**
* **Top 5 manufacturer are**

| **Manufacturer** | **Average of MSRP** |
| --- | --- |
| Bugatti | 1757223.667 |
| Maybach | 546221.875 |
| Rolls-Royce | 351130.6452 |
| Lamborghini | 331567.3077 |
| Bentley | 247169.3243 |

**Task 4.B:**

**Create a bar chart or a horizontal stacked bar chart that visualizes the relationship between manufacturer and average price**.

* **Visualization: Manufacturer and Average Price**



**Results :**

**Visualization shows decreasing Average car price**

**Top 5 Brands are**

1. **Bugatti**
2. **MaybachMaybach**
3. **Roll-Royce**
4. **Lamborghini**
5. **Bentley**

**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.**

* **Visualization**

|  |
| --- |
| **Results: As the Engine Cylinders number increase the Highway MPG decrease** |
|  |
| **Results : Analysing Engine cylinder trend in range 4-8 with highway MPG , trendline is in decreasing order** |

**Task 5.B:**

**Calculate the correlation coefficient between the number of cylinders and highway MPG to quantify the strength and direction of the relationship.**

**Correlation coefficient between the number of cylinders and highway MPG**

| **Correlation** | **highway MPG** |
| --- | --- |
| **Engine Cylinders** | **-0.614706275** |

**Results : Engine Cylinders and Highway MPG are negatively correlated as number of Engine Cylinders increases there is decrease in highway MPG**

## **Building the Dashboard:**

Now for the Next portion of the Project, you need to create the Interactive Dashboard.

Use filters and slicers to make the chart interactive. The client has requested these questions given below:

**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.
* **Dashboard Visualization**



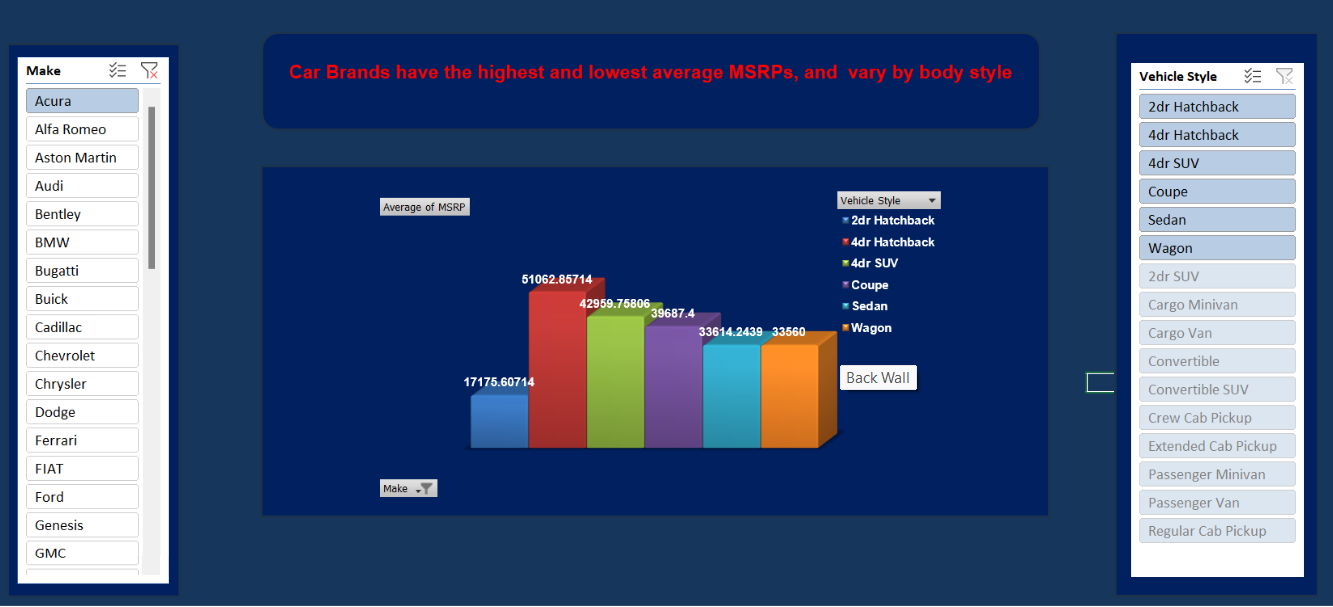
**Results :**

**Interactive dashboard to visulize the change in sum of car price varies with Brand and Body Style**

**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.
* **Dashboard Visualization**



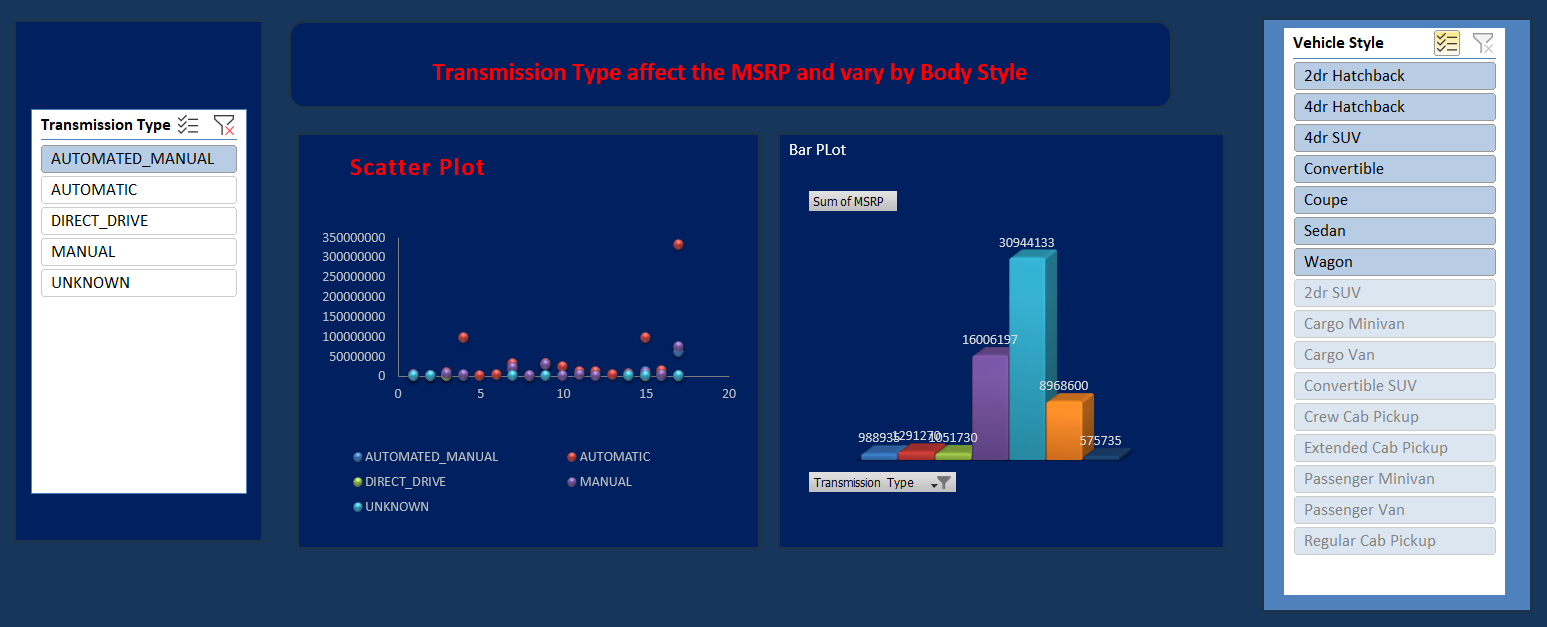
**Results :**

**Interactive dashboard to visulize the change in Average of car price varies with Brand and Body Style**

**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.
* **Dashboard Visualization**

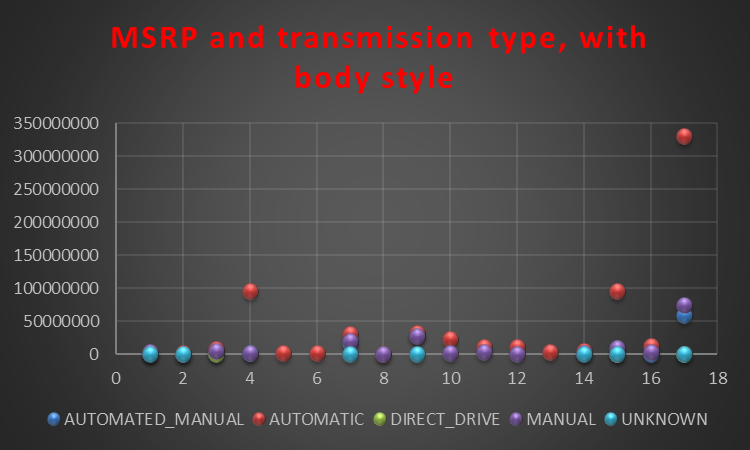


**Results :**

**Interactive dashboard to visulize the change Transmission Type and Body Style How of car price varies MSRP .**

**Scatter plot to analysis How Transmission Type Affect the MSRP by Vary Body Style**

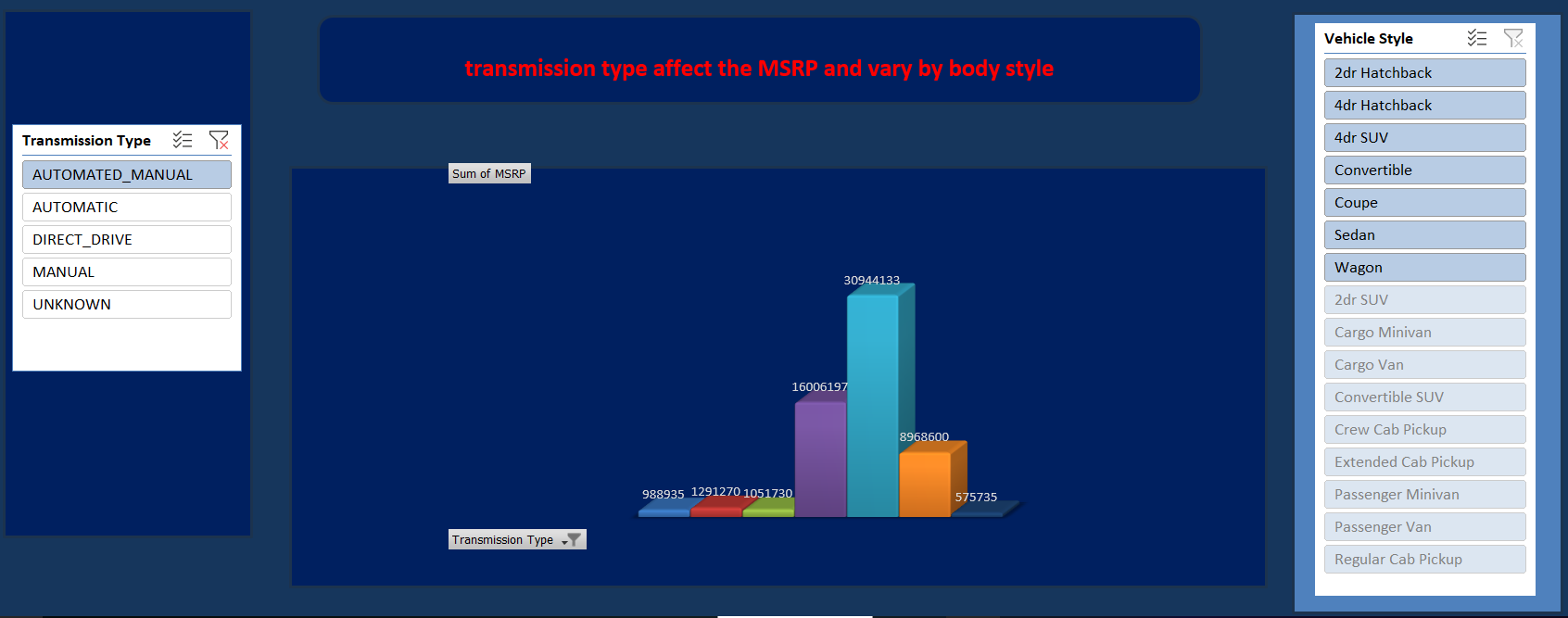
* **Visualization by Scatter plot**



**Results :**

**Scatter plot to analysis How Transmission Type Affect the MSRP by Vary Body Style**

* **Visualization**

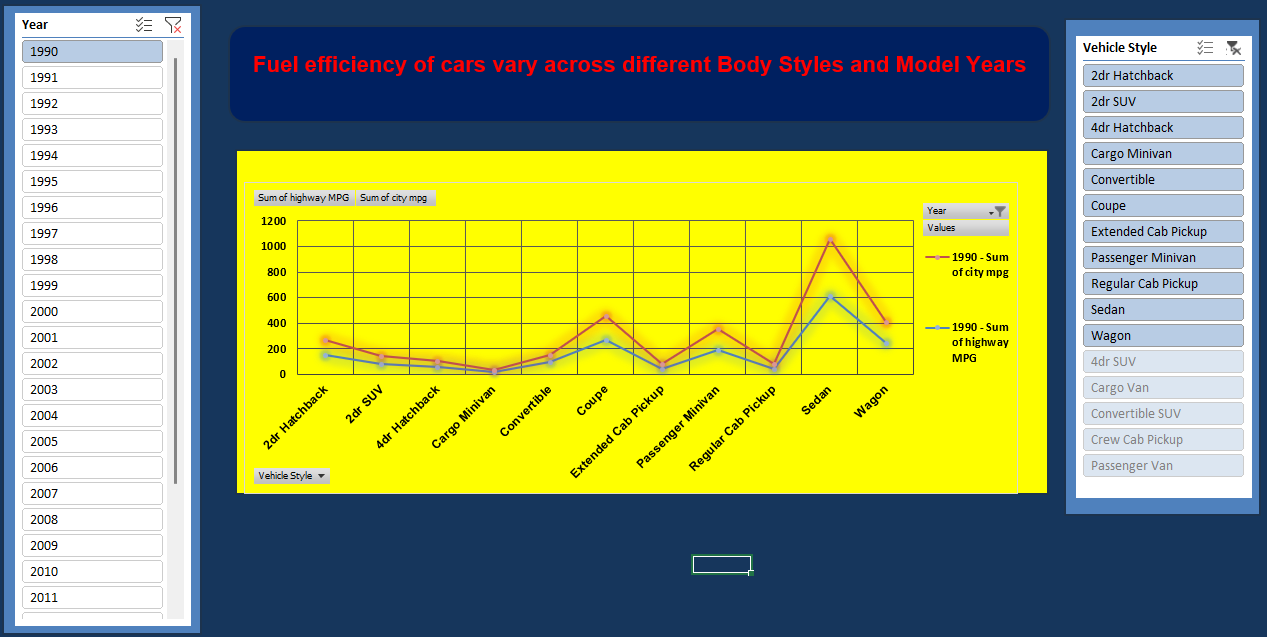


**Results :**

**Interactive dashboard to visulize the change Transmission Type and Body Style How of car price varies MSRP**

**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.
* **Dashboard Visualization**



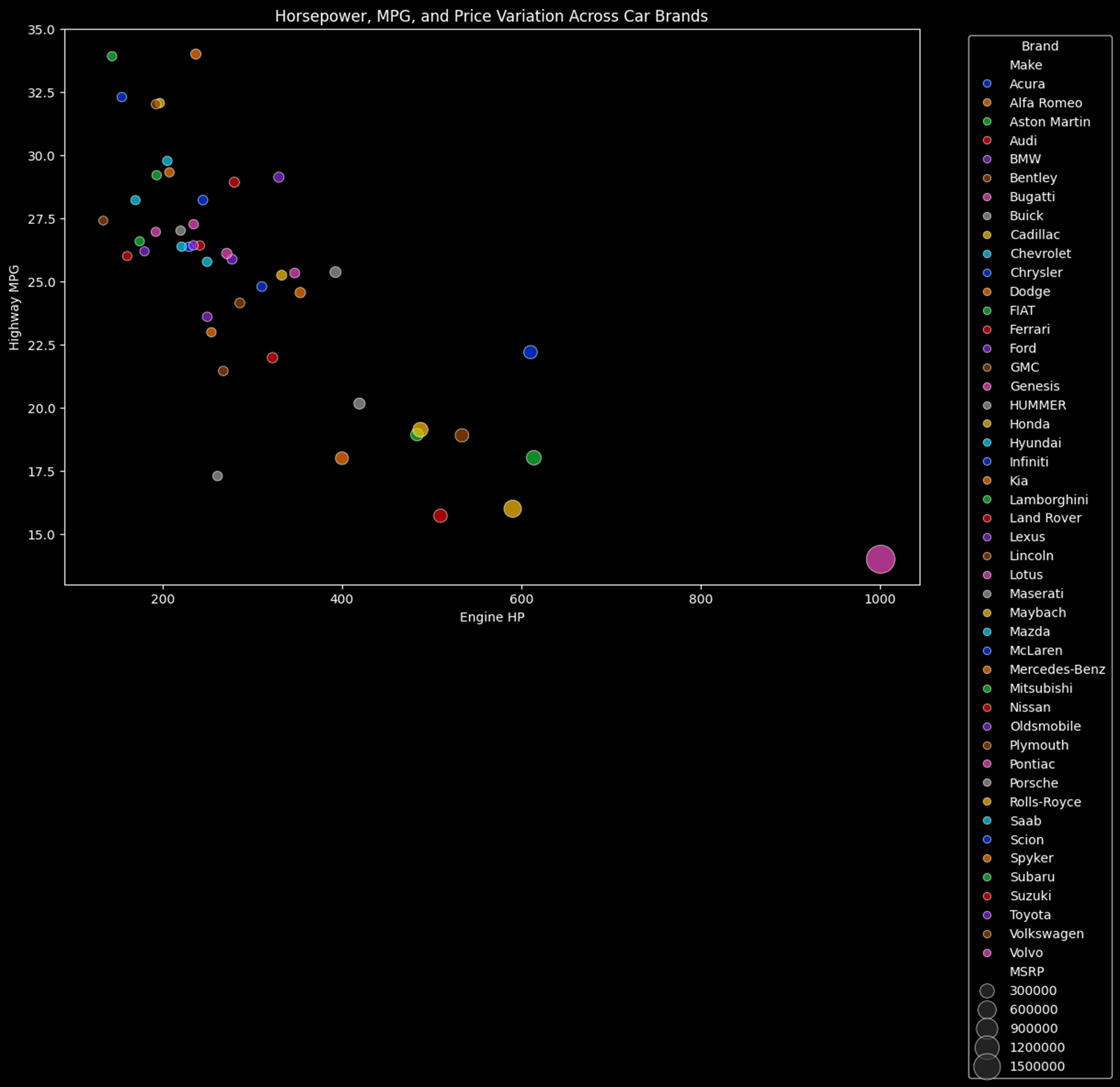
**Results :**

**Interactive dashboard to visualize the change in Fuel Efficiency of car across Body Style and Model Year .**

**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.
* **Dashboard Visualization**

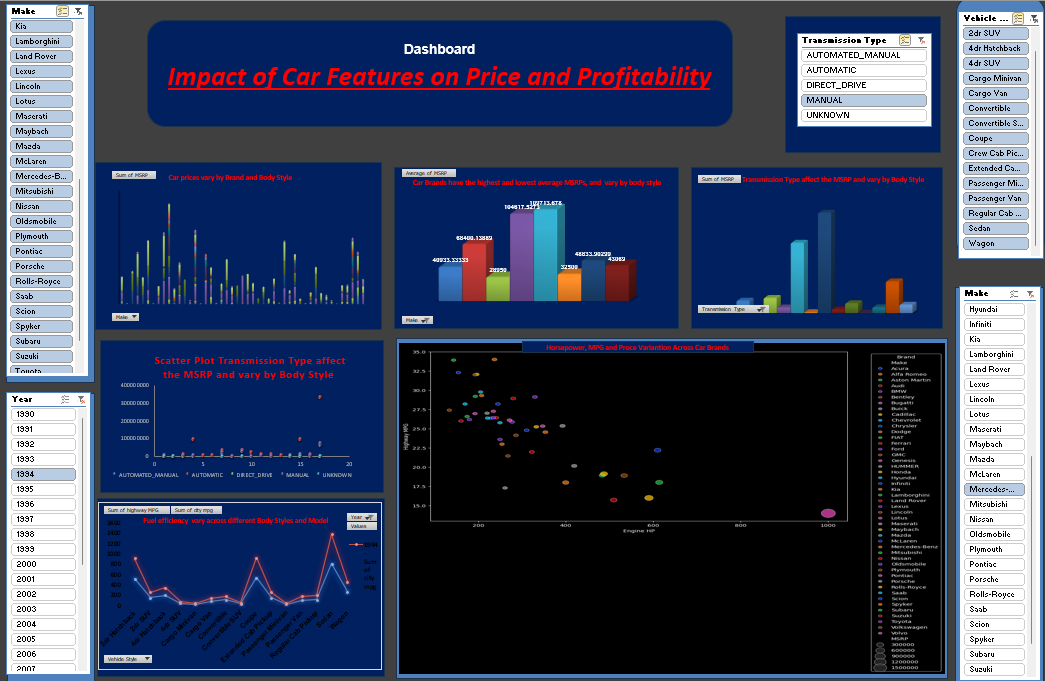


**Results :**

**Visualize the change in Fuel Efficiency of cars across Body Style and Model Year .**

**Final Dashboard**

* **Dashboard Visualization**



**Results :**

**Interactive Dashboard to analyse the Impact of car features on Price and Profitability .**

[**Excel file: Analysis of car data set**](https://docs.google.com/spreadsheets/d/1aYad2iXC4xhKAIP11wgYw7dGXuawAVdF/edit?usp=sharing&ouid=102980832699252360079&rtpof=true&sd=true)

[**Presentation Link**](https://drive.google.com/file/d/1ahmWVwGHr6aLhpAudWG2m7f4ZsJhgkvI/view?usp=sharing)