**ANALYZING THE IMPACT OF CAR FEATURES ON PRICE AND PROFITABILITY**

**CONTENTS**

* TECH STACK
* DATASET OVERVIEW
* DATA PRE-PROCESSING
* ANALYSIS
* DASHBOARD
* CONCLUSION

**TECH STACK**

* Python
* Microsoft Excel
* Tableau

**DATASET OVERVIEW**

**Dataset Description:**

The dataset contains information on various car models and their specifications, and is titled "Car Features and MSRP".

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
* **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
* **Driven\_Wheels:** the type of wheels driven by the car
* **Number of Doors:** the number of doors the car has
* **Market Category:** the market category the car belongs to
* **Vehicle Size:** the size of the car
* **Vehicle Style:** the style of the car
* **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
* **MSRP:** the manufacturer's suggested retail price of the car

**DATA PRE-PROCESSING**

**Handling Duplicate Value:**

Found duplicate rows on analysis. Except the first instance, dropped all other duplicate rows.

**Handling null values:**

* For null values in Engine HP column. We searched the values of Engine HP by searching for rows with the same Make, Model and Year. For rest of the rows which were unaffected by the process, we found that there were all electric cars. So searched for the car model’s Engine HP in the website evcompare.io and replaced null values with its correct values.
* For null values in Engine Cylinder column, where Engine Fuel Type is electric, we replace them with 0 as on analysis, we found that electric cars have 0 Engine Cylinders, which is logical.
* For null values in Engine Cylinder column where Engine Fuel Type is other than electric, we found that there were only two such car models. So searched for information online and replaced them with the correct values.
* For null values in Number of Doors column, we found that there were only two such car models. So searched for the information online and replaced them with the correct values.

**Handling errors**

* Column Transmission Type had some rows with column values UNKNOWN. We searched for the information online and replaced the UNKNOWN with the correct ones.

**Handling outliers**

* For the outliers in Engine HP column, we Checked Market Category column for all rows with high values than four quartile mark of Engine HP. All cars were either Exotic or High Performance or Luxury vehicles, so didn’t change anything.

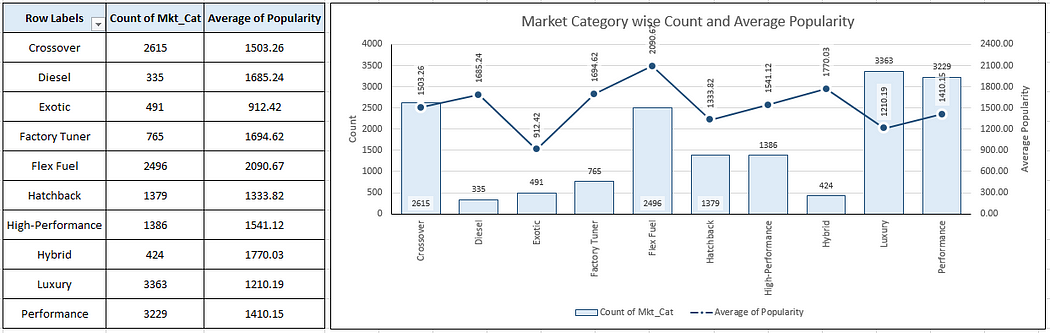
**ANALYSIS**

**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:**

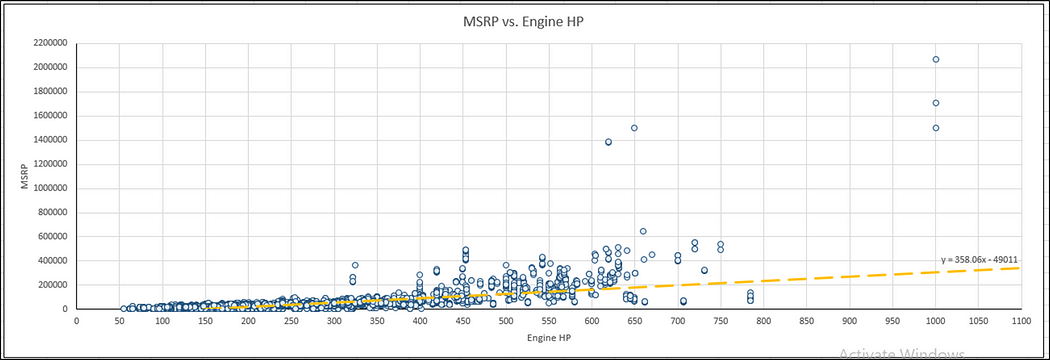


* We can observe that the average popularity of the cars based on their market Category mainly ranges from 1200 to 1800 with the exception of Exotic cars being the lowest popular and the Flex Fuel cars being the most popular.
* The Dataset has comparatively higher number of Performance and Luxury cars, followed by Crossover and Flex Fuel cars.

**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 chat to visualise the relationship between these variables

**Result:**

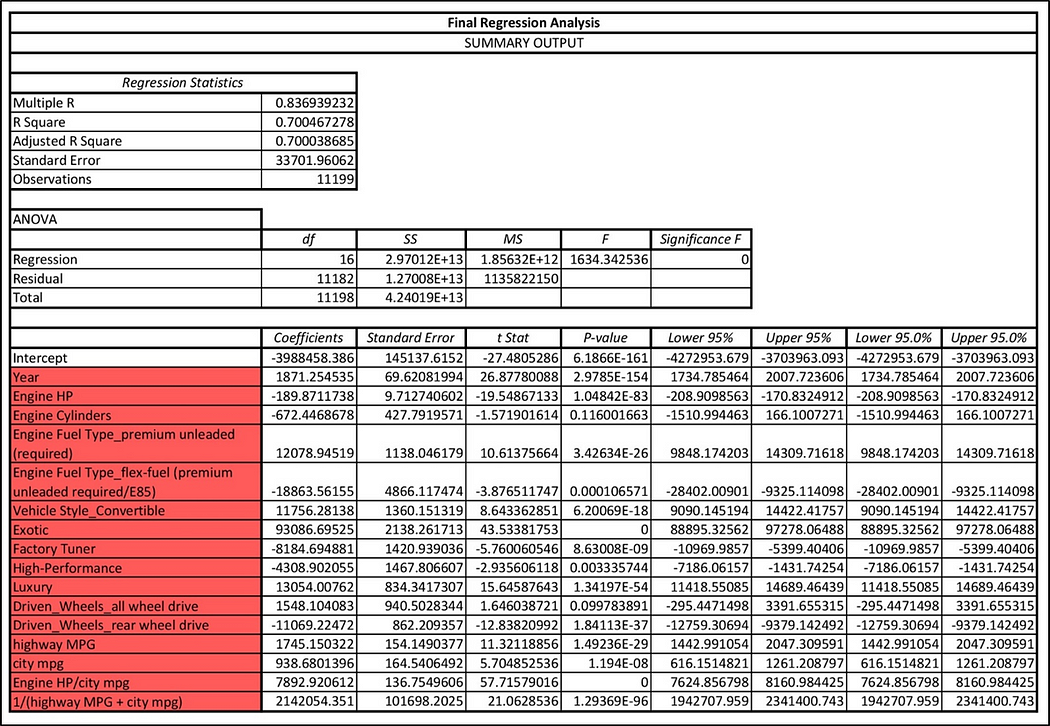


We can observe that the relationship is positive as the trend line has positive slope. This is logical as higher Engine HP requires more complex level of design and engineering and more expensive sub parts Also cars with higher Engine HP are mostly Performance cars.

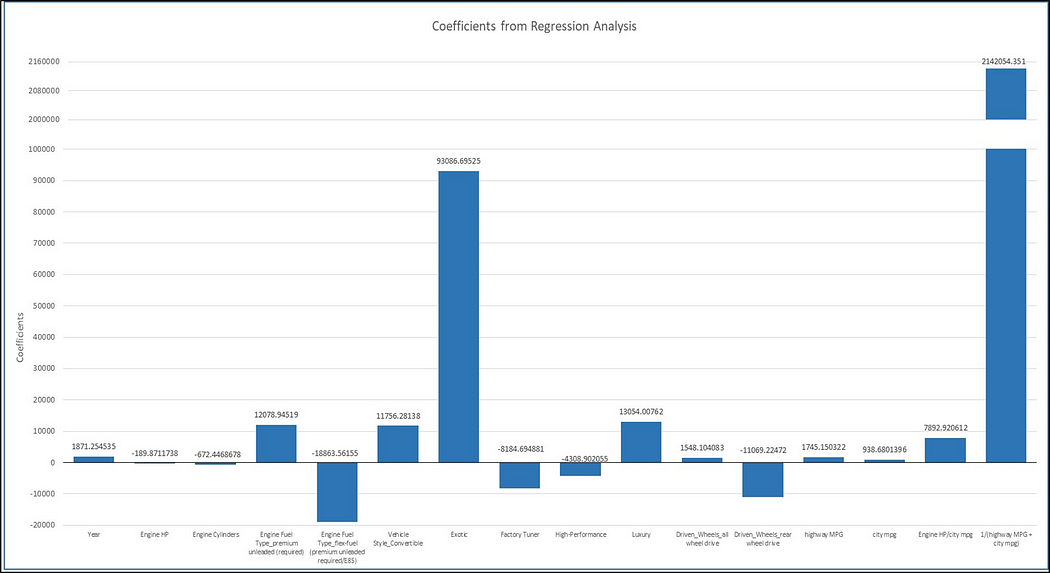
**Insight Required:** Which car feature are most important in determining a car’s price.

**Task 3:** Use regression analysis to identify the variables that have strongest relationships with car’s price. Then create a bar chart that shows the coefficient values for each variable to visualise their relative importance.

**Result :**



* Using regression analysis, we found the top columns. This also include two new columns which were Featured Engineered (Engine HP/ city mpg) and (1 /(highway MPG + city mpg)).
* We can observe that the R-Squared score is 0. 7 which can be counted as a good score.

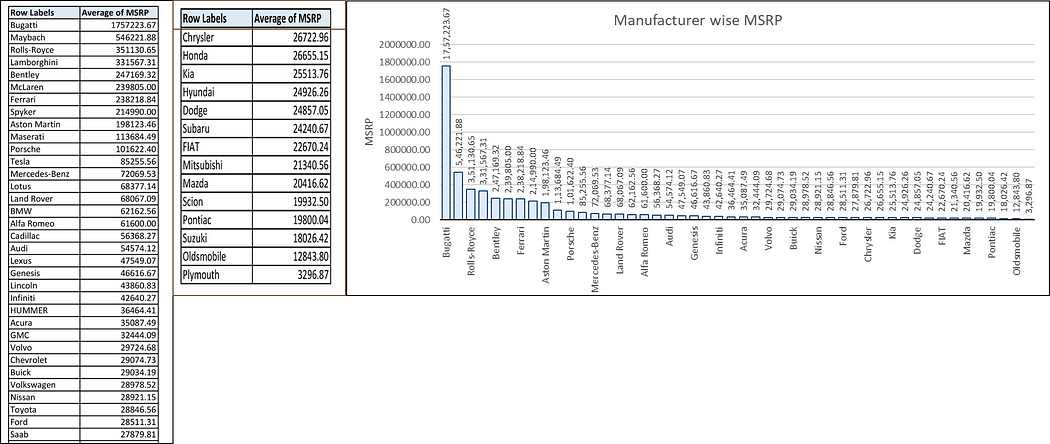


* We can observe that the highest coefficient value Bat of Engineered Feature.,1/(highway MPG + city mpg).
* This shows that the Engineered Feature is very important relationship with car’s price.

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

**Result:**



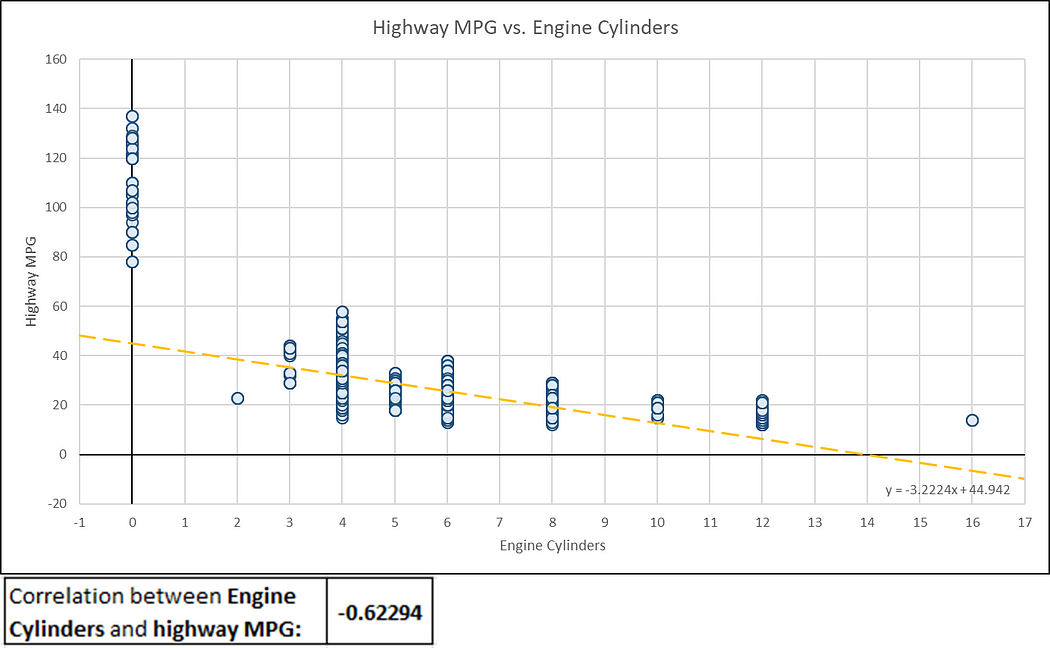
We can observe that the most expensive cars are of Bugatti brand followed by Maybach, Rolls Royce, Lamborghini etc. All these cars brands are High Performance and Luxury brands.

**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.

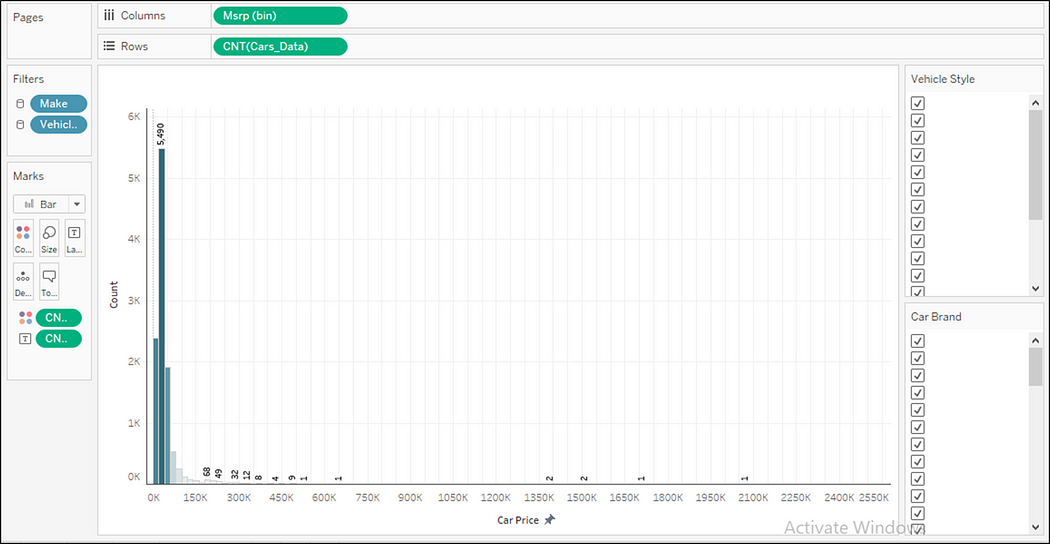
**Result:**



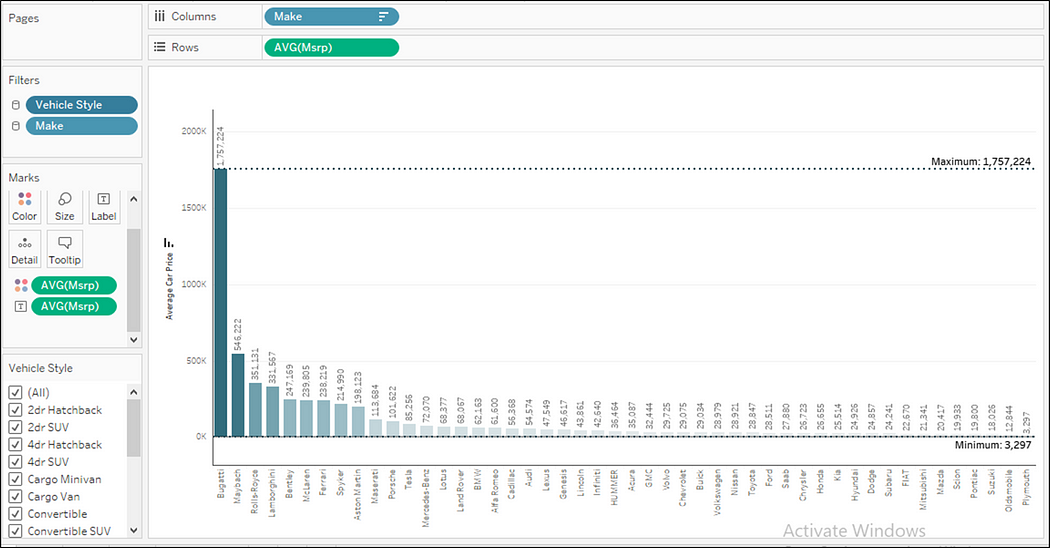
* We can observe that the plot between highway MPG and Engine Cylinders has a negative slope with value of -3.2224.
* The correlation coefficient is also negative with the value of -0.62294.
* This is logical because as number of Engine Cylinder increases the amount of fuel to be burnt also increases thus decreasing the mileage.

**DASHBOARD**

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



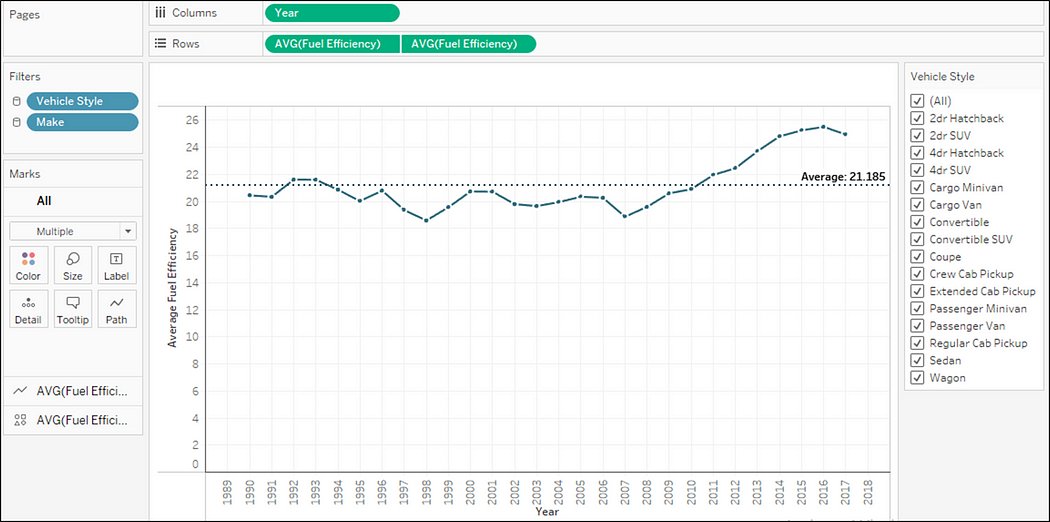
**Task 2:** Which car brands have the highest and lowest average MSRPs, and how does this vary by body style?



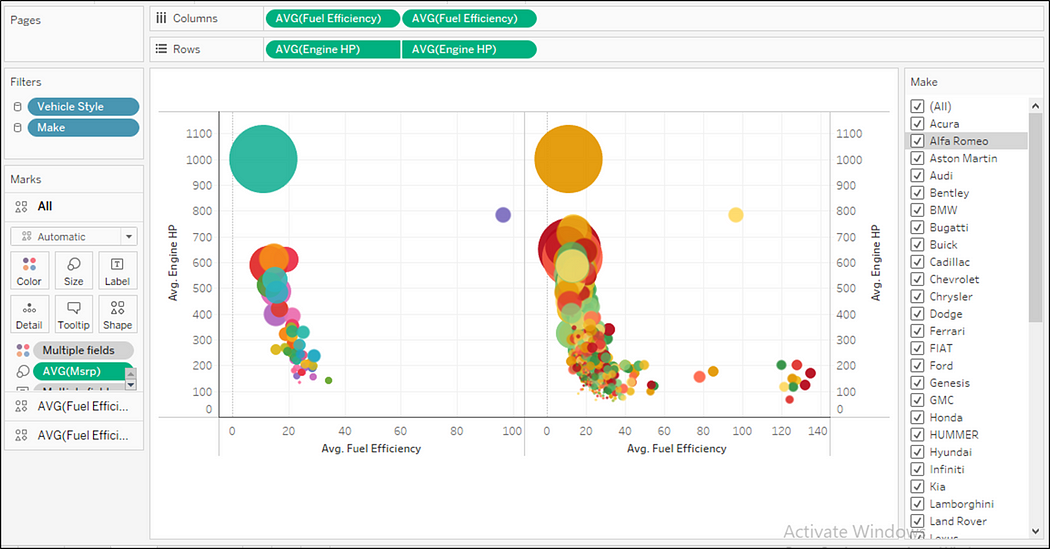
**Task 3:** How do the different feature such as transmission type affect the MSRP, and how does this vary by body style?



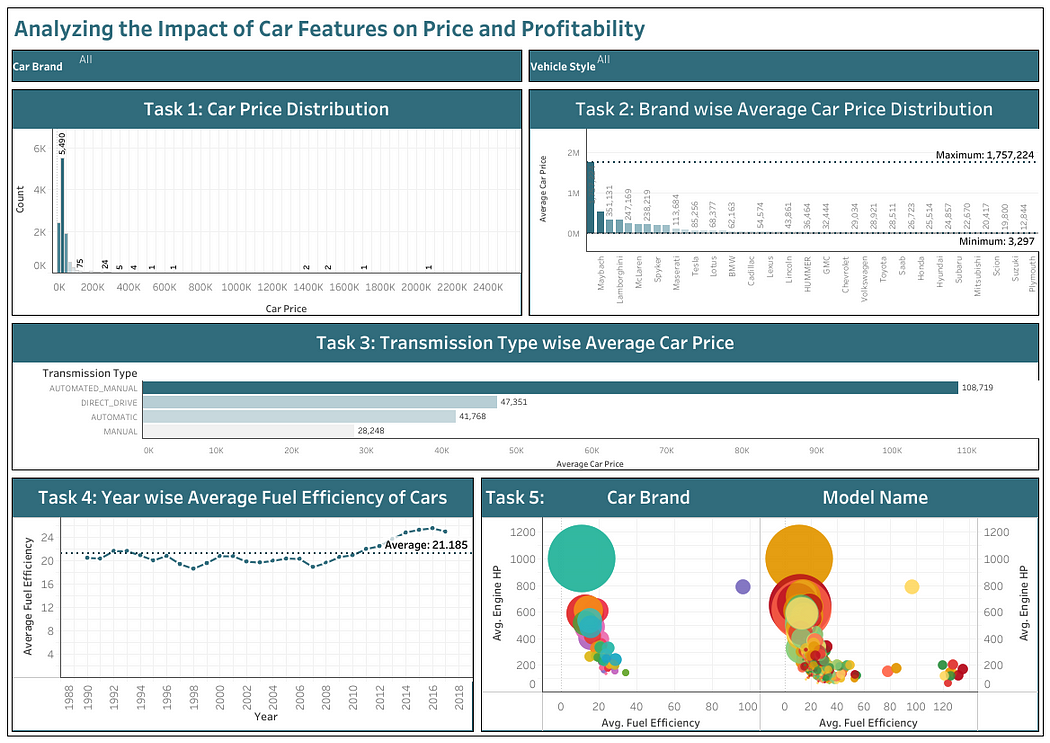
**Task 4:** How does the fuel efficiency of cars vary across different body styles and model years?



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



Result:



**CONCLUSION**

Throughout this project, I was able to understand the importance of data analytics in car feature analysis. As it provides valuable insights, which helps in making data driven decisions.

In this project, I was able to get insights like which features affects car prices, relationship and engine cylinders and its fuel efficiency etc. I also got experience in data processing, like data cleaning, handling outliers feature engineering, etc, in this project, which can be communicated to relevant stakeholders as per requirement