Project Report: Optimizing Pricing and Product Development in the Automotive Industry

**Project Overview:**

The project aims to assist a car manufacturer in optimizing pricing and product development decisions by analyzing a comprehensive dataset titled "Car Features and MSRP." The dataset, contains information on over 11,000 car models, including specifications like make, model, year, fuel type, horsepower, transmission type, market category, and MSRP.

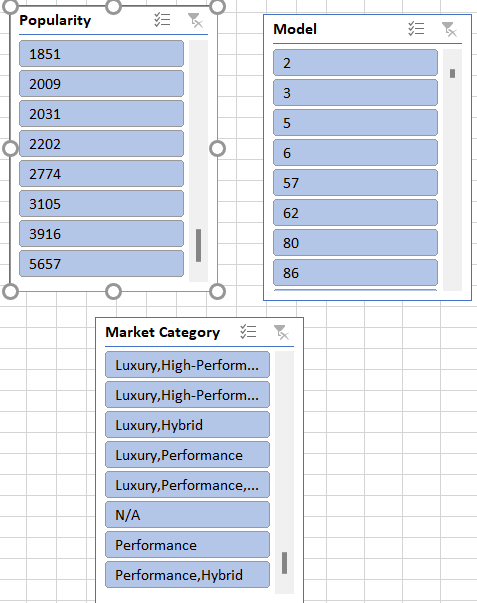
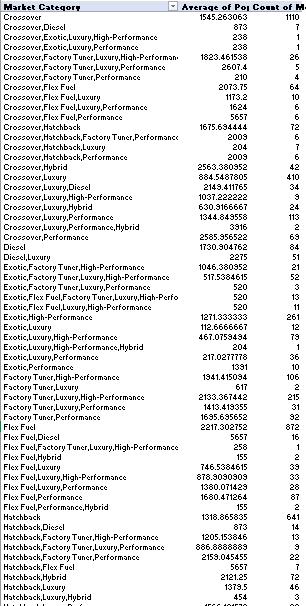
**Data Pre-processing:**

* **Handling Null Values**
* For Null values in Engine HP column, we searched the value of Engine HP by searching for rows with same Make, Model and Year. If found then replaced null value with the mode of Engine HP value of all the matching rows. For rest of the rows which were unaffected by above process, we found that they were all Electric Cars. So searched for the Car model’s Engine HP (Motor’s Power) in the website evcompare.io and replaced null values with its correct values.
* For Null values in Engine Cylinders column where Engine Fuel Type is electric, we replaced them with 0 as on analysis we found that electric Cars have 0 Engine Cylinders which is logical.
* For Null values in Engine Cylinders column where Engine Fuel Type is other than electric, we found that there were only two such Car models. So searched for the 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.

**Analysis Tasks:**

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.



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

* We can observe that the average popularity of cars based on their Market Category mainly ranges from 2000 to 4000 with the exception of Exotic cars being the lowest popular and Flex Fuel cars being the most popular.
* The dataset has comparatively higher number of Performance and Luxury cars followed by Crossovers and Flex Fuel cars.

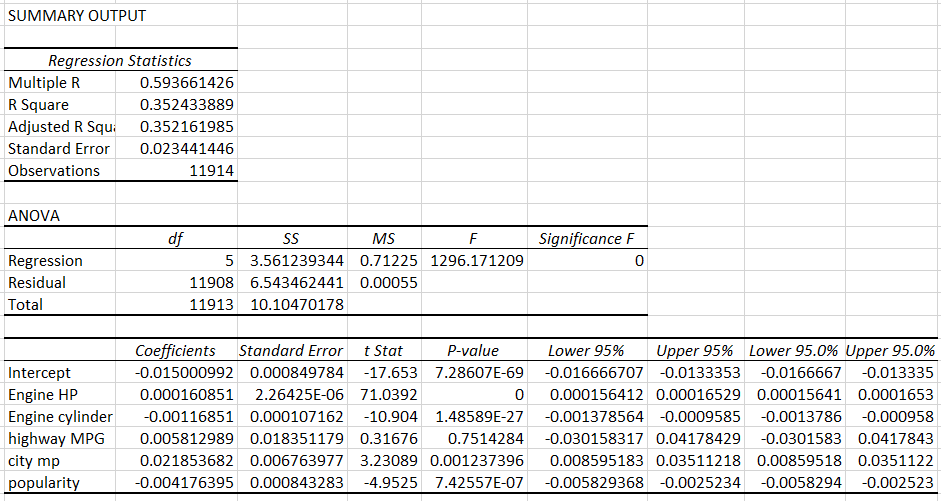
1. Insight Required: What is the relationship between a car's engine power and its price?

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.

* We can observe that the relationship is positive as the trendline 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.

1. Insight Required: Which car features are most important in determining a car's price?

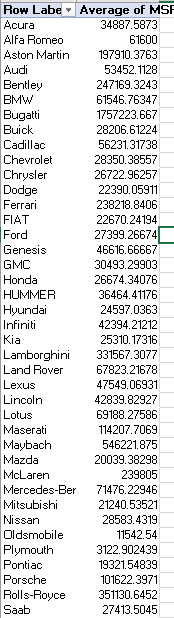
Task 3: Use regression analysis to identify the variables that have the strongest relationship with a car's price. Then create a bar chart that shows the coefficient values for each variable to visualize their relative importance.



* Using regression analysis, I found the top columns.
* We can observe that the R-Squared score is 0.35 which can be counted as a good score.
* This shows that the city mpg is very important relationship with Car’s price.

1. 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.



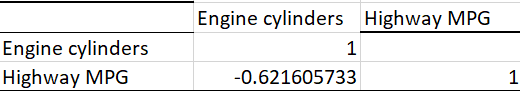
Task 4.B: Create a bar chart or a horizontal stacked bar chart that visualizes the relationship between manufacturer and average price.

We can observe that the most expensive cars are manufactured by of Buick brand followed by Maybach.

1. 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.

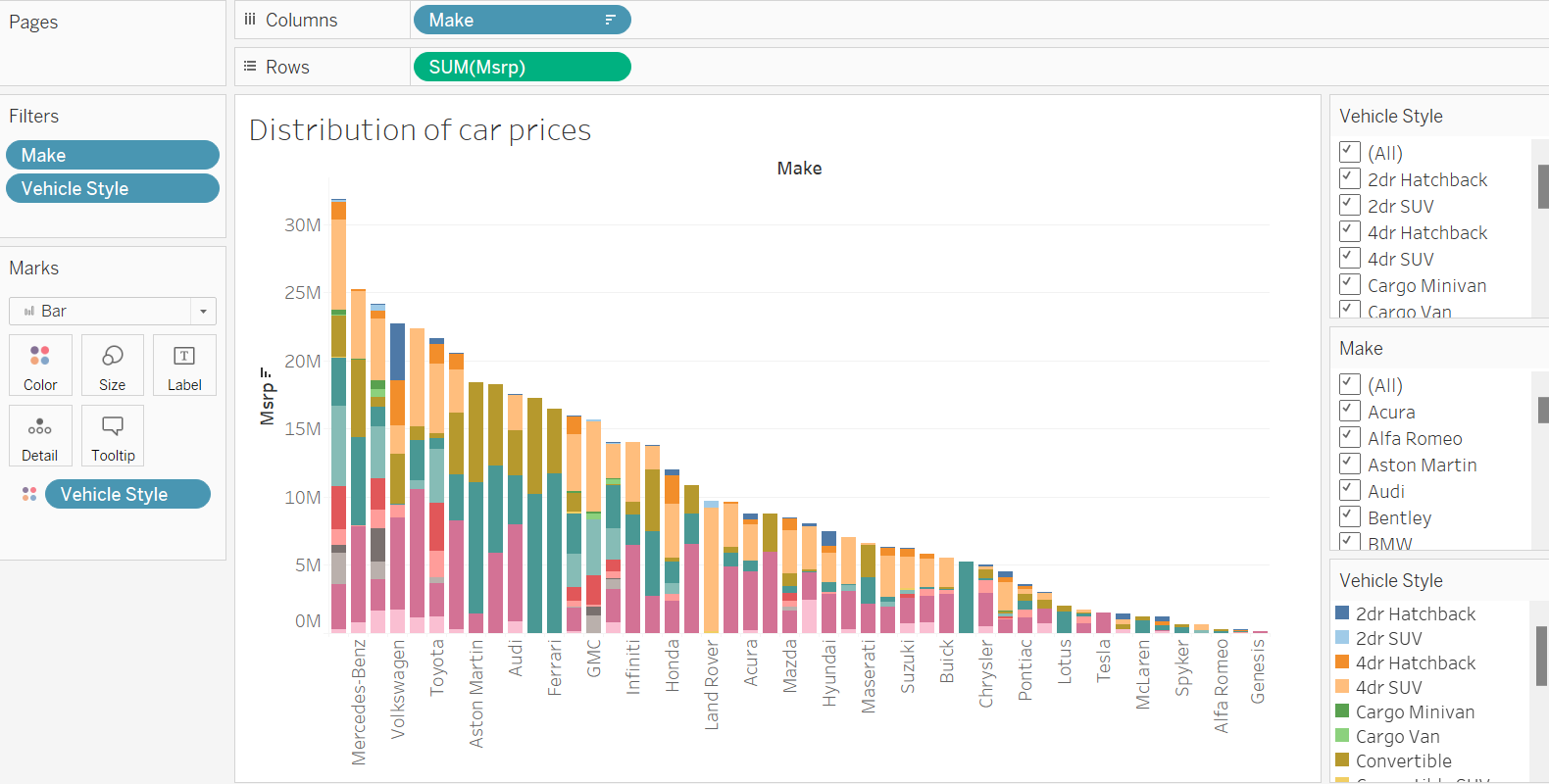
Task 5.B: Calculate the correlation coefficient between the number of cylinders and highway MPG to quantify the strength and direction of the relationship.



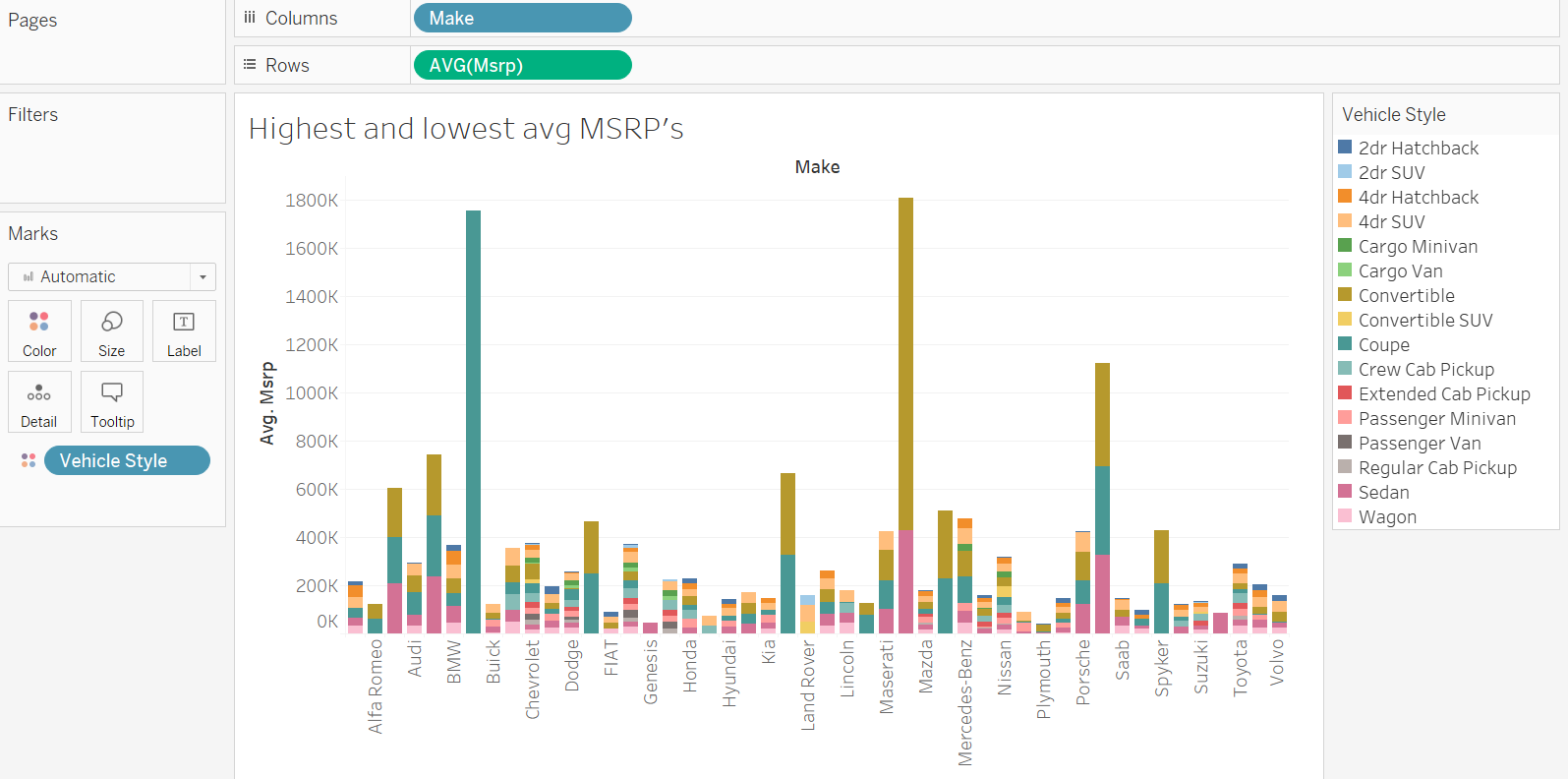
* We can observe that the plot between highway MPG and Engine Cylinders has a negative slope.
* The correlation coefficient is also Negative with a value of -0.62294.
* This is logical because as number of Engine Cylinders increases, the amount of fuel to be burnt also increasing, thus decreasing the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

**Building the Dashboard:**

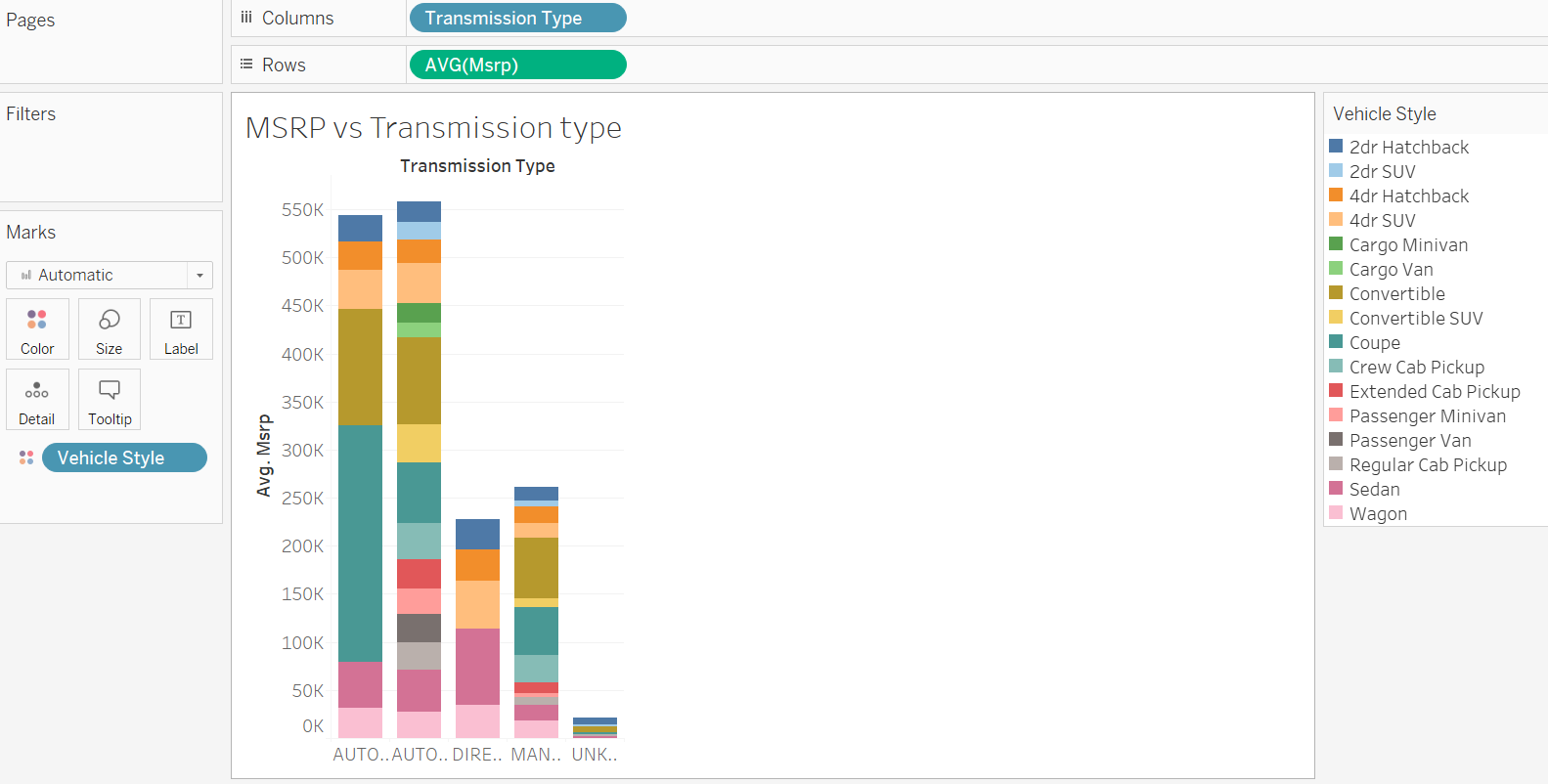
Task 1: How does the distribution of car prices vary by brand and body style?



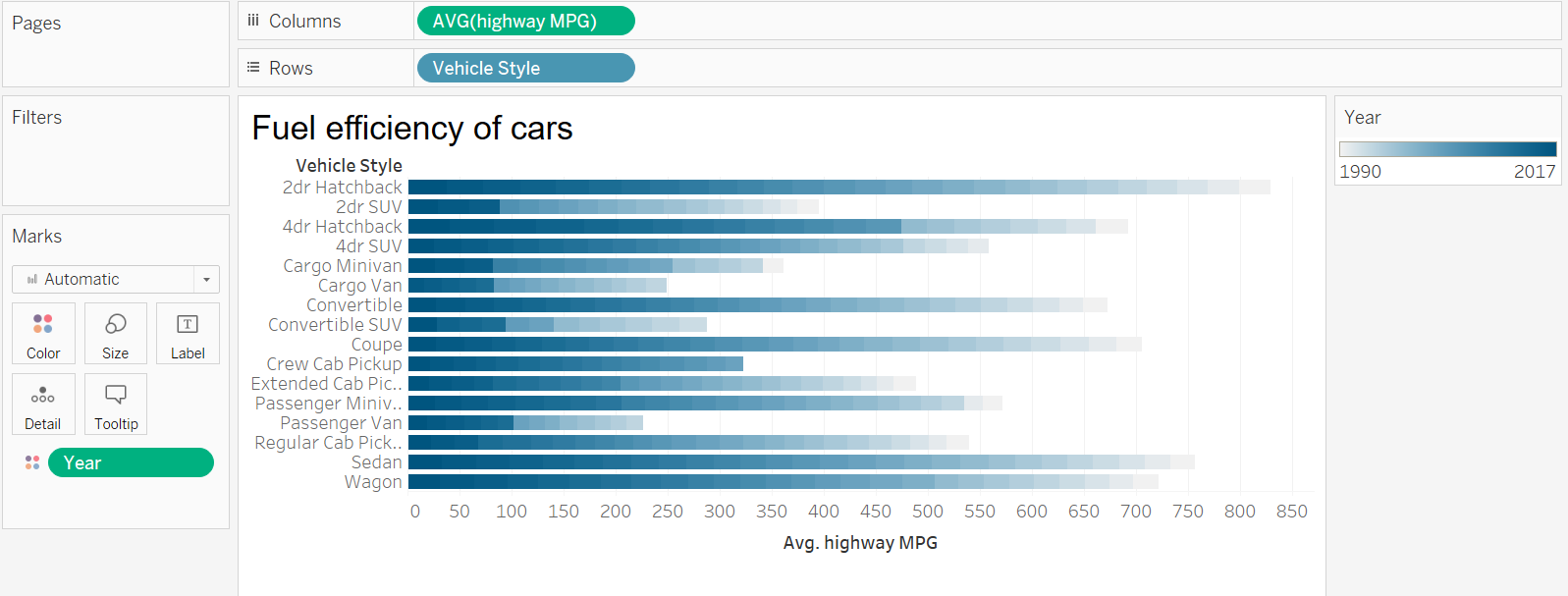
Task 2: Which car brands have the highest and lowest average MSRPs, and how does this vary by body style?



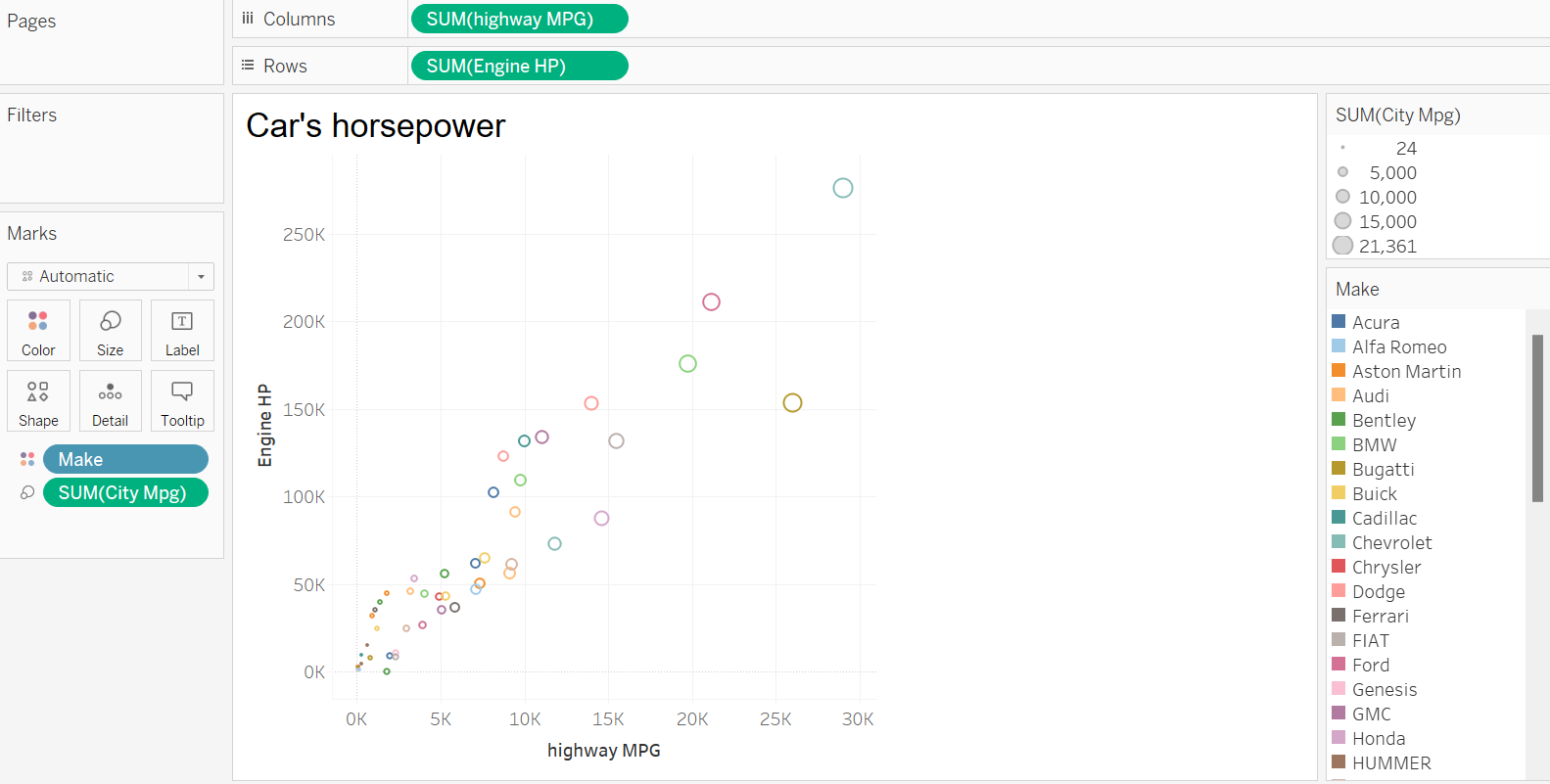
Task 3: How do the different feature such as transmission type affect the MSRP, and how does this vary by body style?



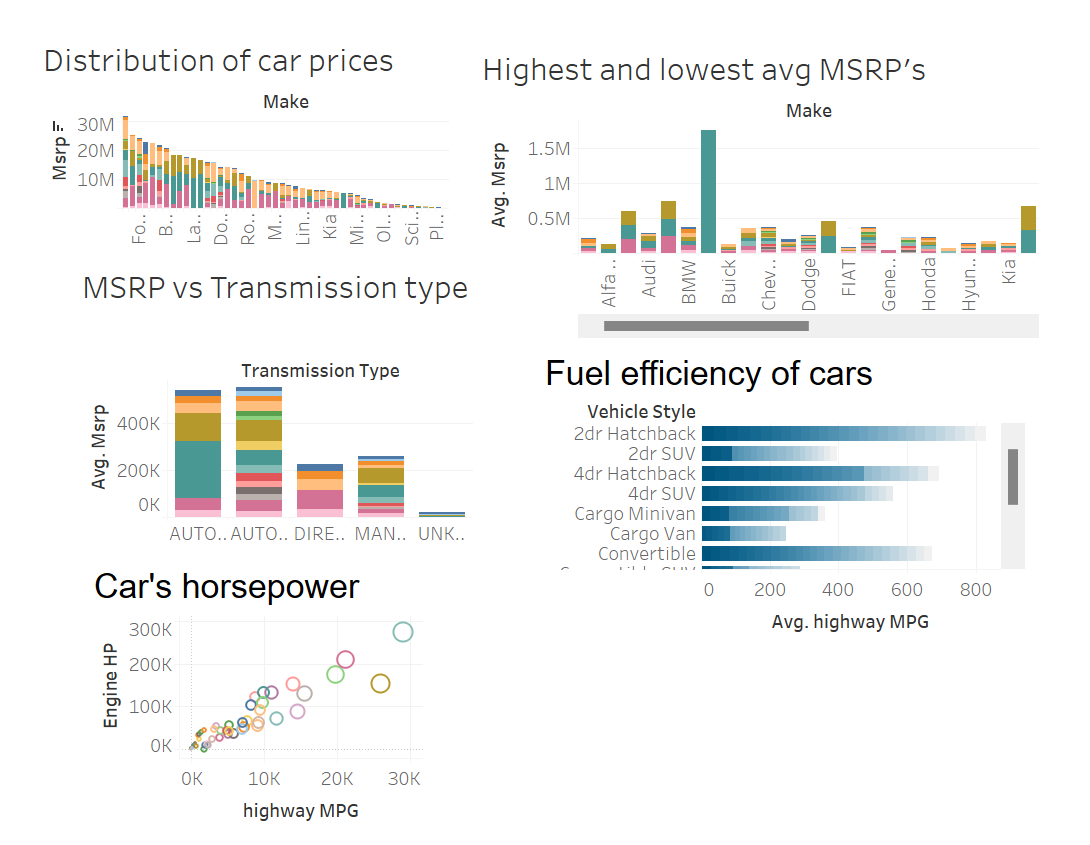
Task 4: How does the fuel efficiency of cars vary across different body styles and model years?



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



**Dashboard:**



**Tech Stack Used:**

* Microsoft Excel 2019 — A spreadsheet editor software used mainly by professionals to enter data in table format, perform computations, plot graphs etc.
* Tableau Public 2024.1 — A visualization tool to represent data in graphs and plots. Mainly used to create Dashboard

**Conclusion:**

Through 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 effects Car Price, relationship between Engine Cylinders and it’s fuel efficiency etc. I also got experience in Data Preprocessing like Data Cleaning, handling Outliers.