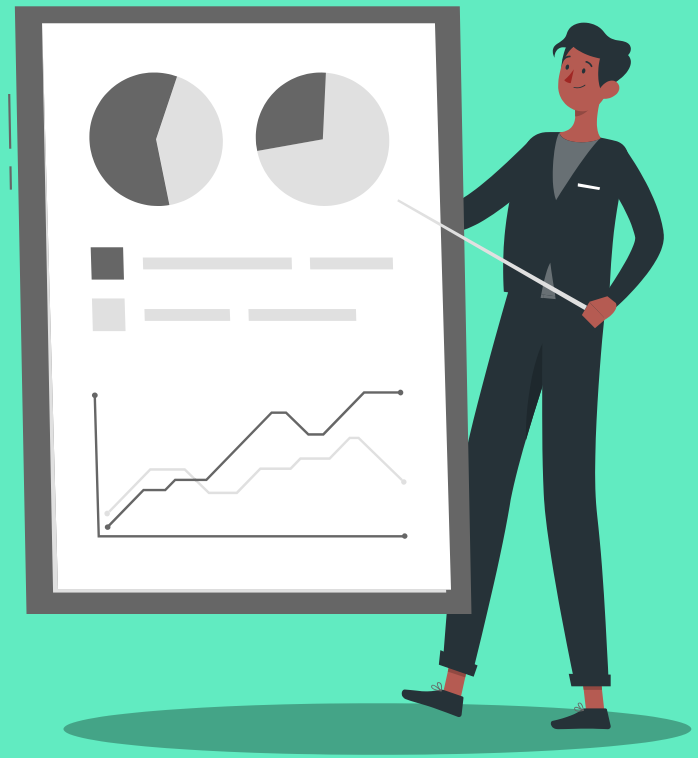


Car Feature Analysis Report

By Yash Shinde

Jupyter Notebook:- Click Here to Download

Video Presentation:- Click Here to Watch Video Presentation

Project Description!

Car companies want to make the most money while giving people the cars they want. But it's not easy with lots of competition and different kinds of cars like electric and regular ones. A Data Analyst can help by looking at data about cars and prices. They can figure out what things people like in cars and what kinds of cars make the most money. By doing this, the car company can decide how much to charge for their cars and what features to put in new cars. This helps the company do well in the market and make more money in the long run.

Table of Contents

"Unlocking the power of data can drive car manufacturers to thrive in a changing world, satisfying consumers while boosting profits."

Tech Stack Used :-

-> Jupyter Notebook
-> Tableau Public
-> PowerPoint
-> Github
-> Google Drive

Importing and Exploring Dataset

Importing and Exploring Dataset

- First, I created a new notebook and imported libraries and dataset.
- application_data = pd.read_csv("car_data.csv")
- Then I explored the dataset using the following programs.
- To see the five rows from top and bottom I used head() and tail() respectively.
- application_data.head()
- application_data.tail()
- Then by Using the column function see the names of all the columns/variables.
- application_data.columns
- By using the describe function obtained all descriptive stats for the numeric variables in the DataSet.
- application_data.describe()
- By using isnull() method of pandas data frame calculated total null values in each variable.
- car_data.isnull().sum()
- Then filled those null values with mean, median or mode. (According to variables)
- car_data['Engine HP'] = car_data['Engine HP'].fillna(car_data['Engine HP'].mean())
- car_data['Engine Cylinders'] = car_data['Engine Cylinders'].fillna(car_data['Engine Cylinders'].median())
- car_data['Engine Fuel Type'] = car_data['Engine Fuel Type'].fillna(car_data['Engine Fuel Type'].mode())

```
1 car_data.isnull().sum()
```

```
Make                    0
Model                   0
Year                    0
Engine Fuel Type        3
Engine HP              69
Engine Cylinders       30
Transmission Type       0
Driven_Wheels           0
Number of Doors         6
Market Category      3742
Vehicle Size            0
Vehicle Style           0
highway MPG             0
city mpg                0
Popularity              0
MSRP                    0
dtype: int64
```

```
car_data.duplicated()
```

```
0        False
1        False
2        False
3        False
4        False
         ...
11909    False
11910    False
11911    False
11912    False
11913    False
Length: 11914, dtype: bool
```

```
1 car_data.columns
```

```
Index(['Make', 'Model', 'Year', 'Engine Fuel Type', 'Engine HP',
       'Engine Cylinders', 'Transmission Type', 'Driven_Wheels',
       'Number of Doors', 'Market Category', 'Vehicle Size', 'Vehicle Style',
       'highway MPG', 'city mpg', 'Popularity', 'MSRP'],
      dtype='object')
```

```
1 car_data.describe()
```

	Year	Engine HP	Engine Cylinders	Number of Doors	highway MPG	city mpg	
count	11914.000000	11914.000000	11914.000000	11908.000000	11914.000000	11914.000000	119
mean	2010.384338	249.386070	5.629763	3.436093	26.637485	19.733255	15
std	7.579740	108.875192	1.778413	0.881315	8.863001	8.987798	14
min	1990.000000	55.000000	0.000000	2.000000	12.000000	7.000000	
25%	2007.000000	170.000000	4.000000	2.000000	22.000000	16.000000	5
50%	2015.000000	227.000000	6.000000	4.000000	26.000000	18.000000	13
75%	2016.000000	300.000000	6.000000	4.000000	30.000000	22.000000	20
max	2017.000000	1001.000000	16.000000	4.000000	354.000000	137.000000	56

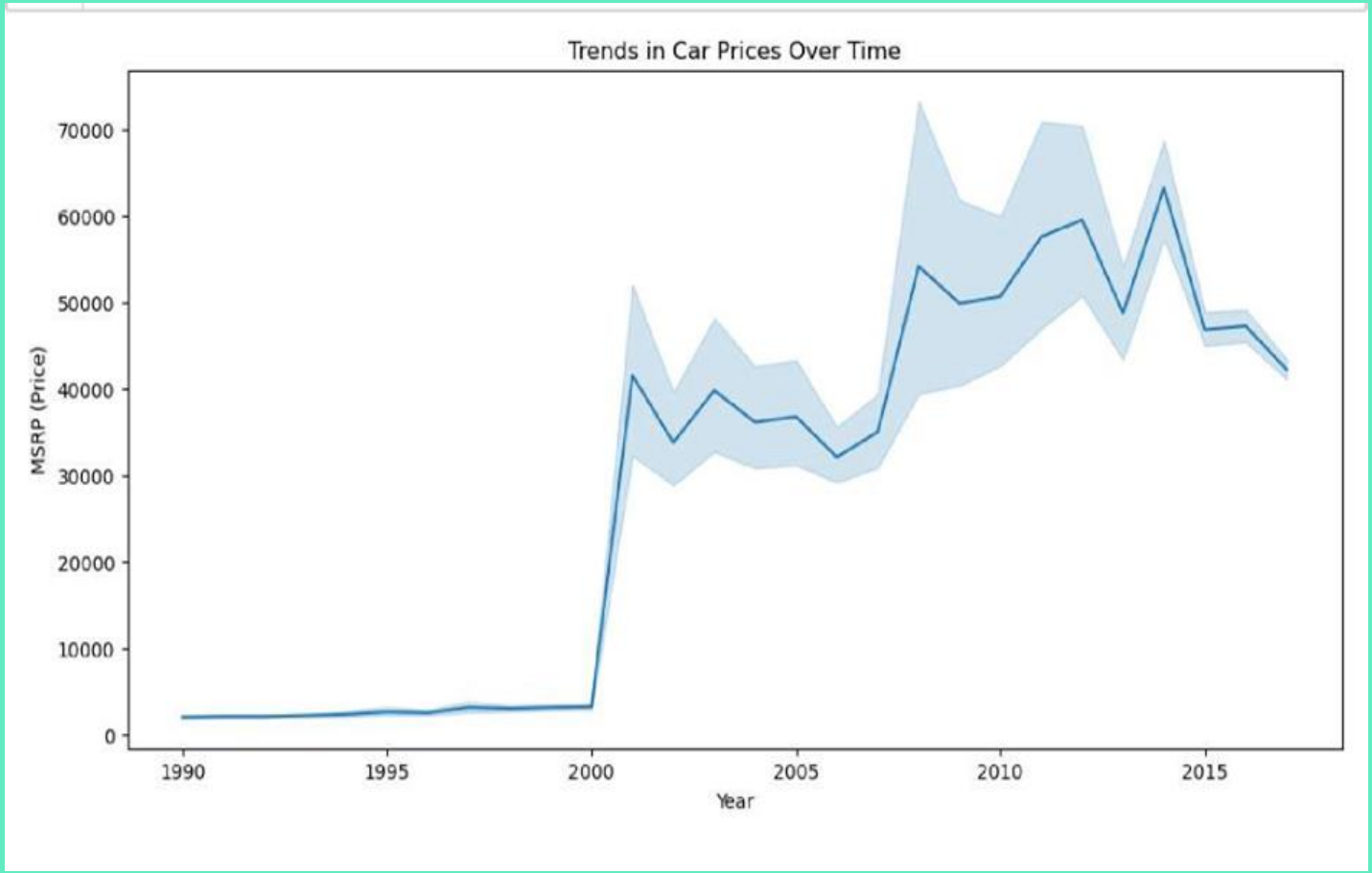
A Understanding the Dataset

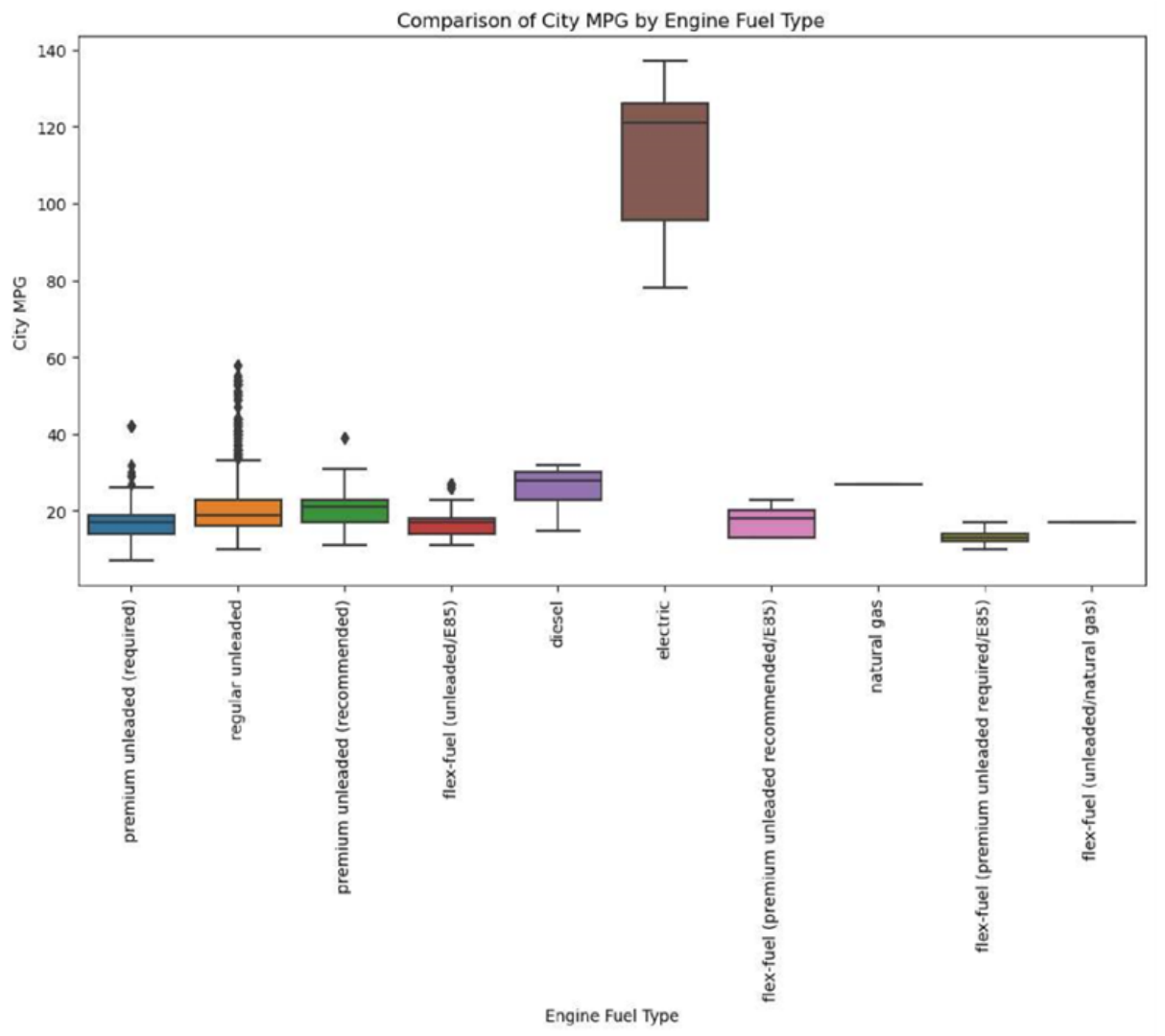
Gain insights into various aspects of the automotive industry.

Understanding the Dataset

- Analyzing trends in the car features and pricing over time.
- plt.figure(figsize=(12, 6))
- sns.lineplot(x='Year', y='MSRP', data=car_data)
- plt.title('Trends in Car Prices Over Time')
- plt.xlabel('Year')
- plt.ylabel('MSRP (Price)')
- plt.show()

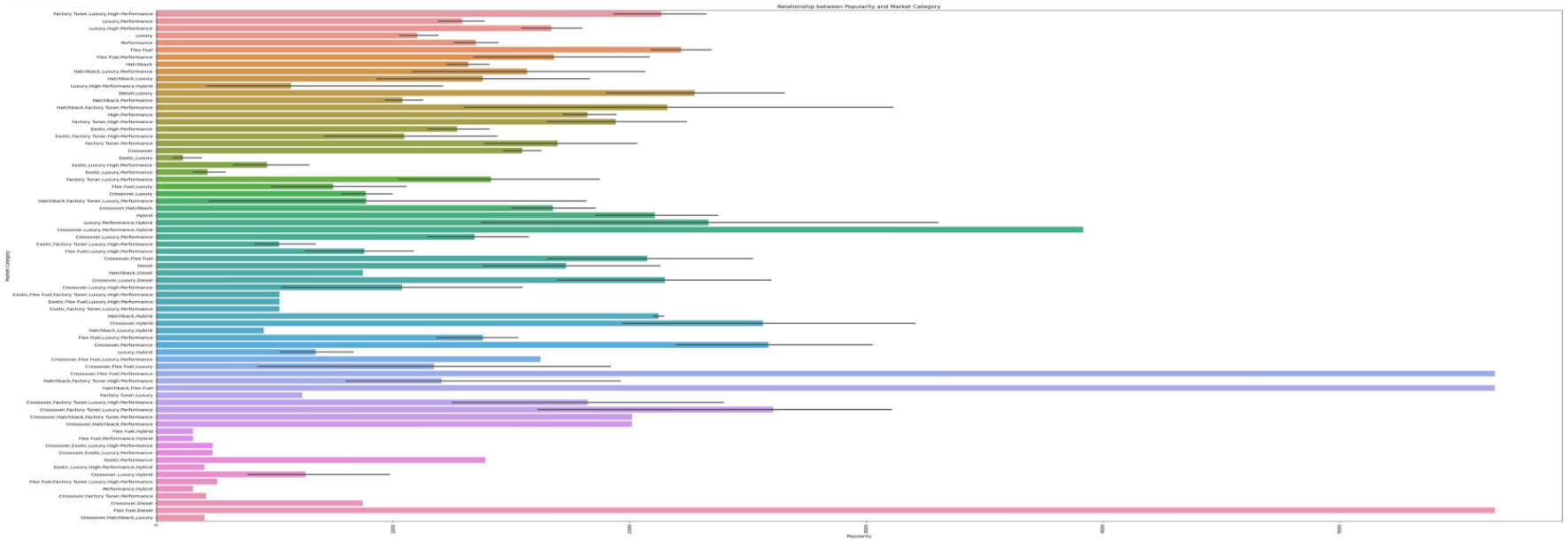
- Compare fuel efficiency of different types of cars
- plt.figure(figsize=(12, 6))
- sns.boxplot(x='Engine Fuel Type', y='city mpg', data=car_data)
- plt.title('Comparison of City MPG by Engine Fuel Type')
- plt.xlabel('Engine Fuel Type')
- plt.ylabel('City MPG')
- plt.xticks(rotation=90)
- plt.show()


Trends in Car Prices Over Time
MSRP (Price)
70000
60000
50000
40000
30000
20000
10000
0
1990
1995
2000
2005
2010
2015
Year


Comparison of City MPG by Engine Fuel Type
City MPG
140
120
100
80
60
40
20
premium unleaded (required)
regular unleaded
premium unleaded (recommended)
flex-fuel (unleaded/E85)
diesel
electric
flex-fuel (premium unleaded recommended/E85)
natural gas
flex-fuel (premium unleaded required/E85)
flex-fuel (unleaded/natural gas)
Engine Fuel Type

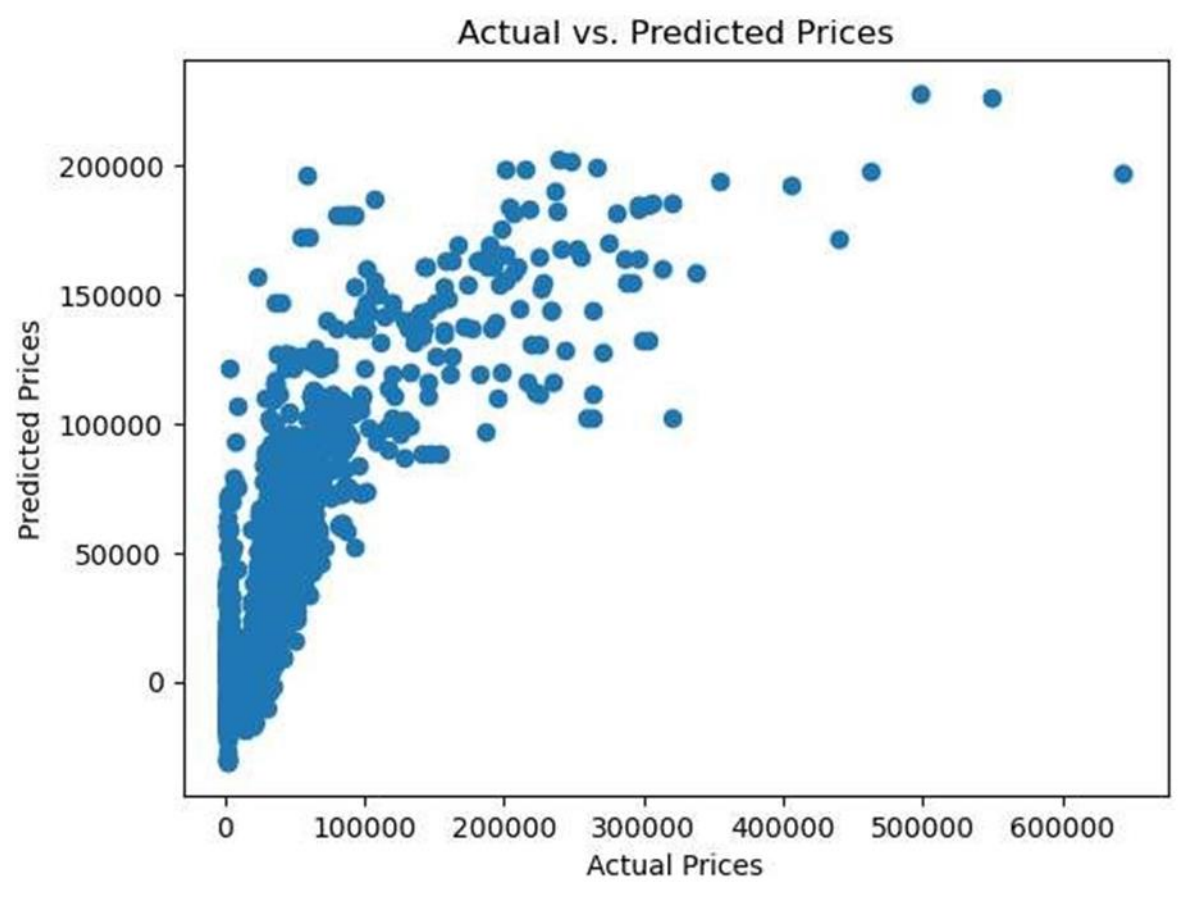
Investigating the relationship between a car's features and its popularity

```
plt.figure(figsize=(50, 25))
sns.barplot(x='Popularity', y='Market Category', data=car_data)
plt.title('Relationship between Popularity and Market Category')
plt.xlabel('Popularity')
plt.ylabel('Market Category')
plt.xticks(rotation=90)
plt.show()
```



Understanding the Dataset

- 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, I have developed 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.



B

Analysis of Dataset

Finding valuable insights using data analysis techniques such as regression analysis, pivot tables, sensitivity analysis, optimization, and time series analysis.

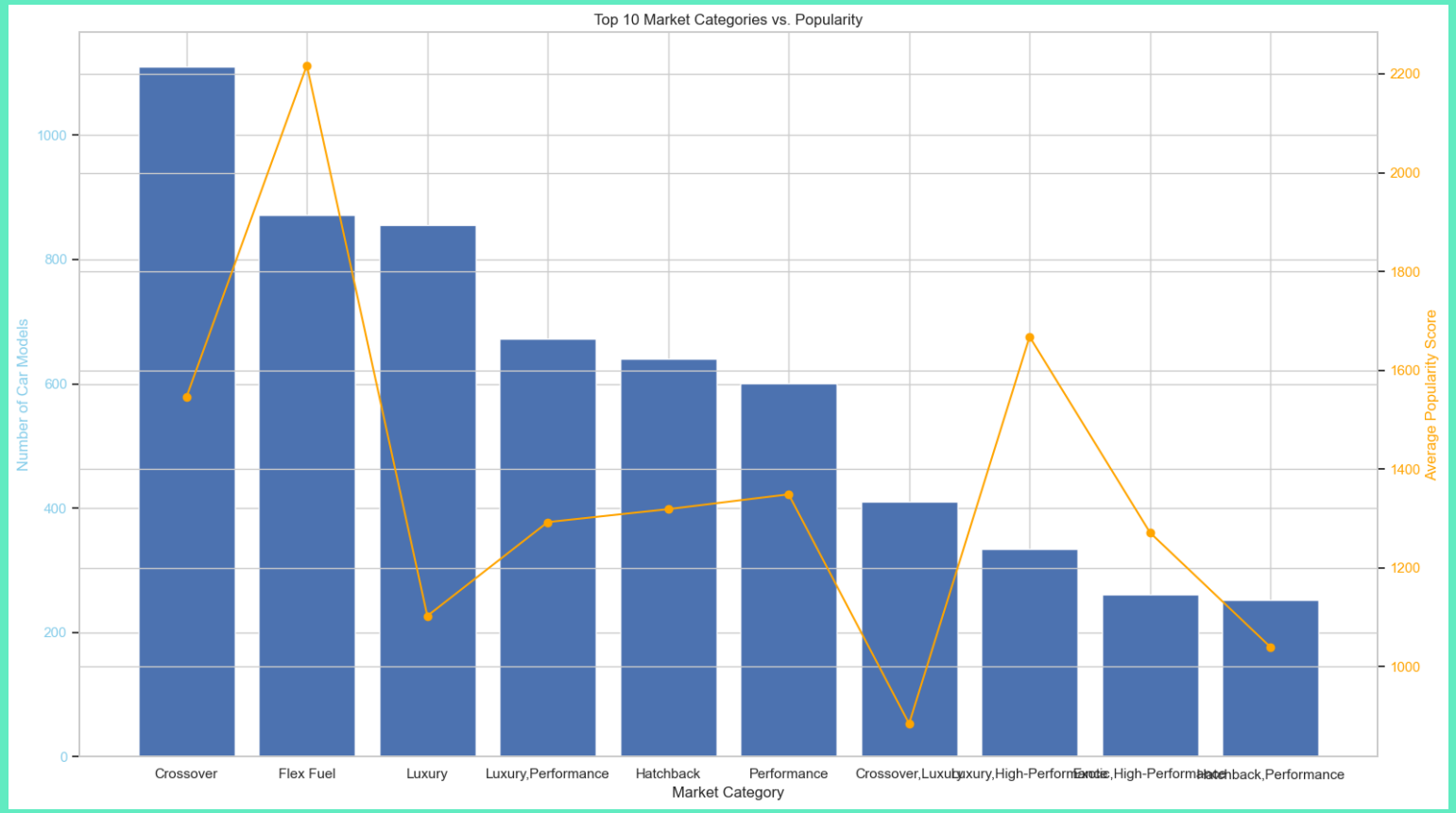
How does the popularity of a car model vary across different market categories?

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

Market Category	Number of Car Models	Average Popularity Score
Crossover	1110	1545.263063
Flex Fuel	872	2217.302752
Luxury	855	1102.657310
Luxury,Performance	673	1292.615156
Hatchback	641	1318.865835
...	...	...
Exotic,Luxury,High-Performance,Hybrid	1	204.000000
Flex Fuel,Factory Tuner,Luxury,High-Performance	1	258.000000
Crossover,Exotic,Luxury,Performance	1	238.000000
Crossover,Exotic,Luxury,High-Performance	1	238.000000
Performance,Hybrid	1	155.000000

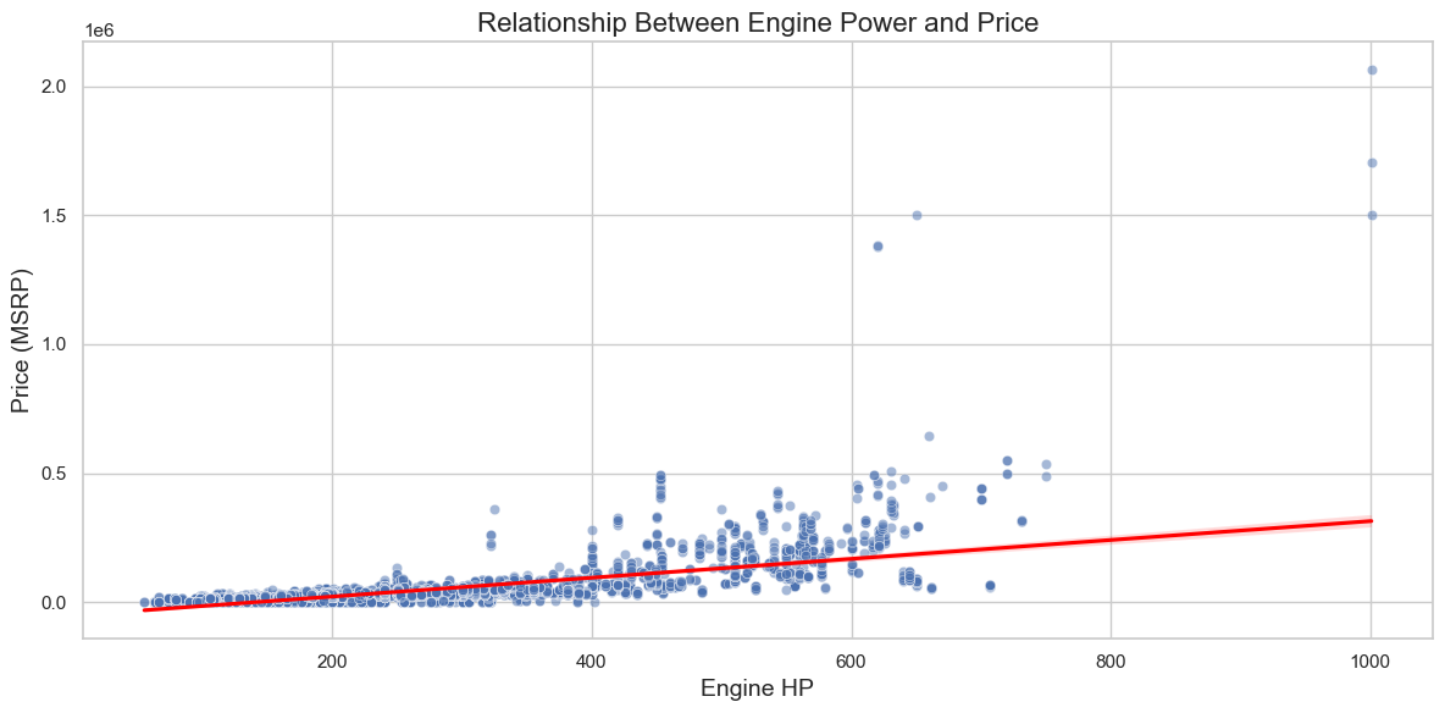
71 rows × 2 columns

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



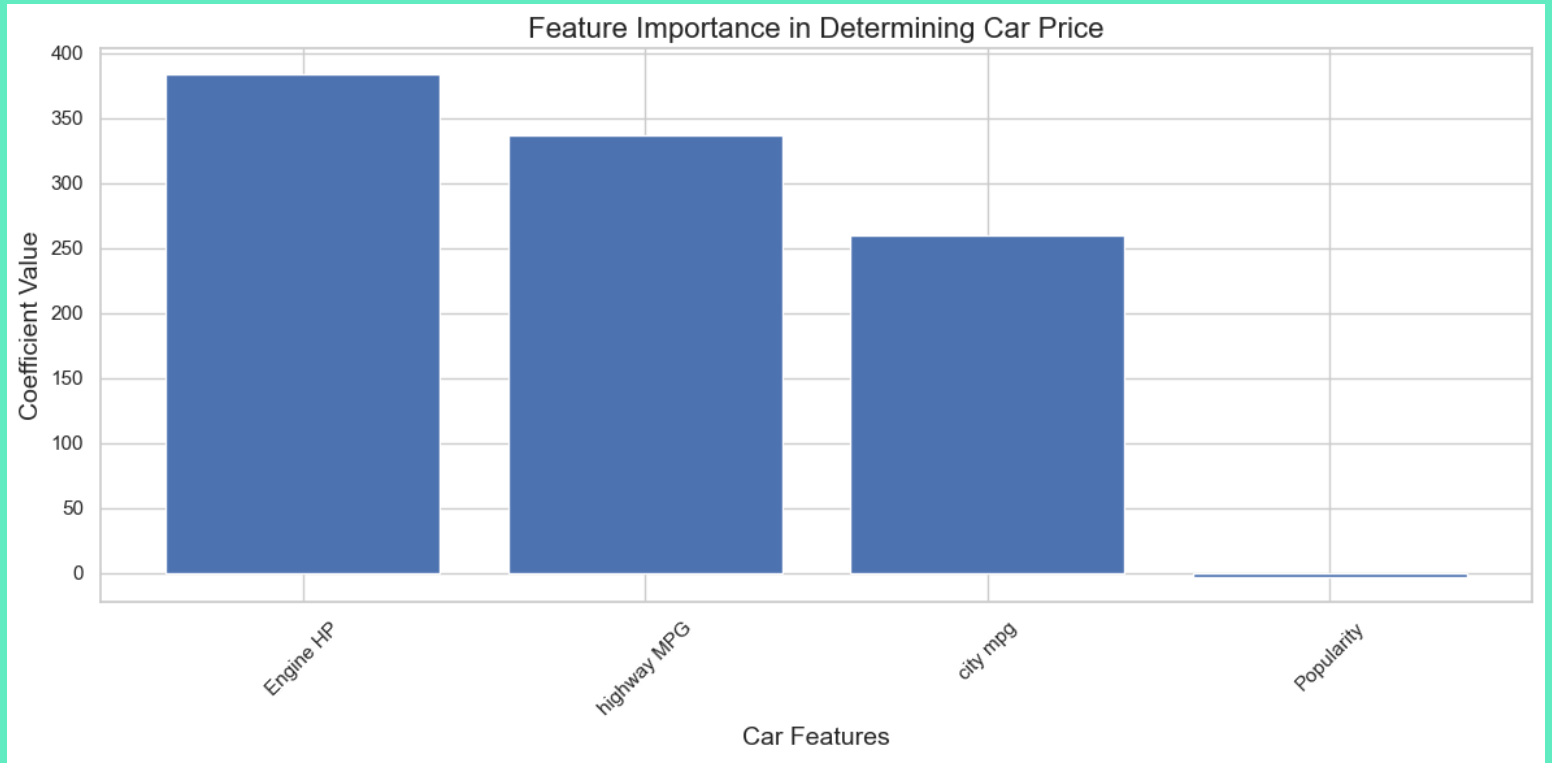
What is the relationship between a car's engine power and its price?

- Task 2: Creating 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.



Which car features are most important in determining a car's price?

- Task 3: Using regression analysis to identify the variables that have the strongest relationship with a car's price.

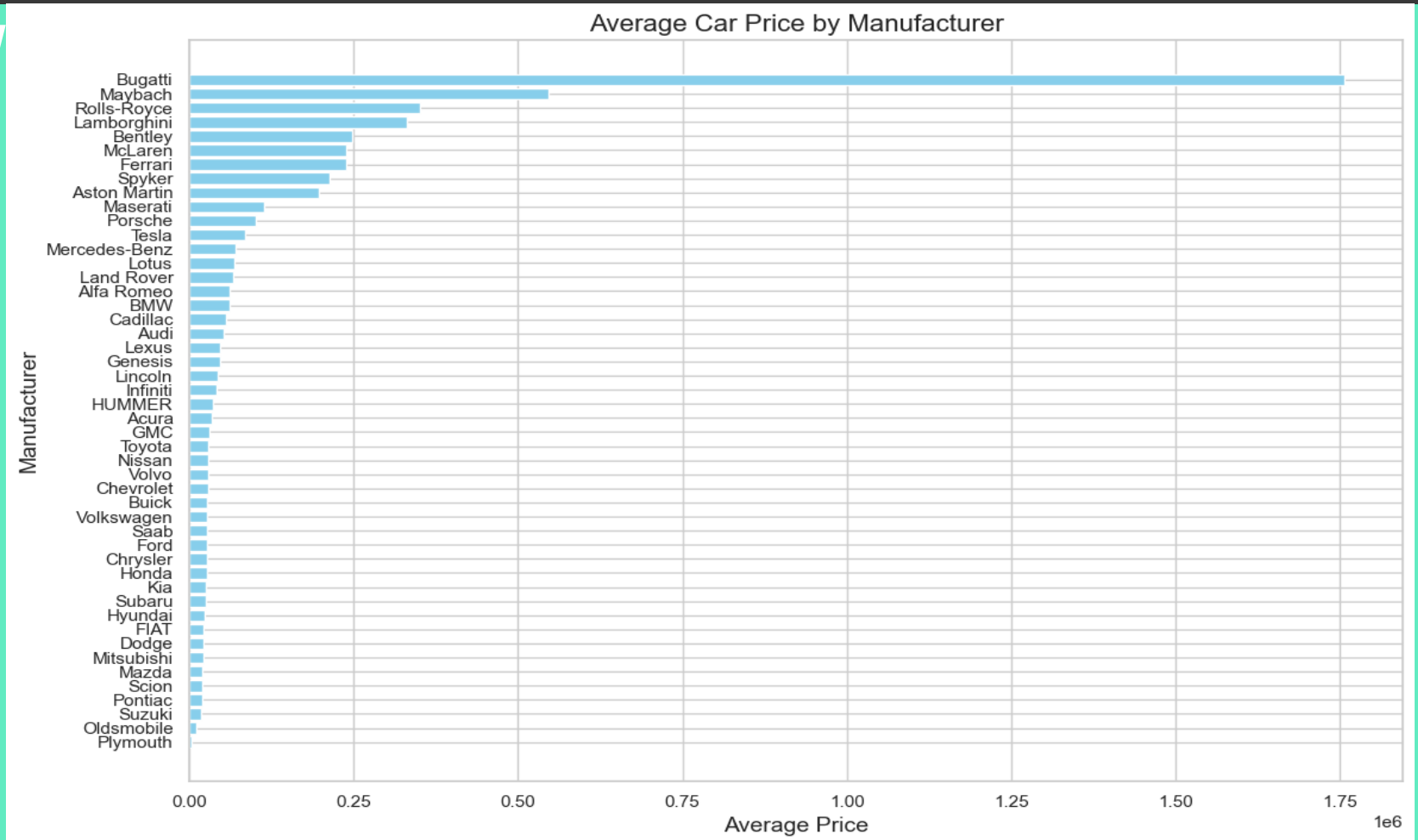


How does the average price of a car vary across different manufacturers?

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

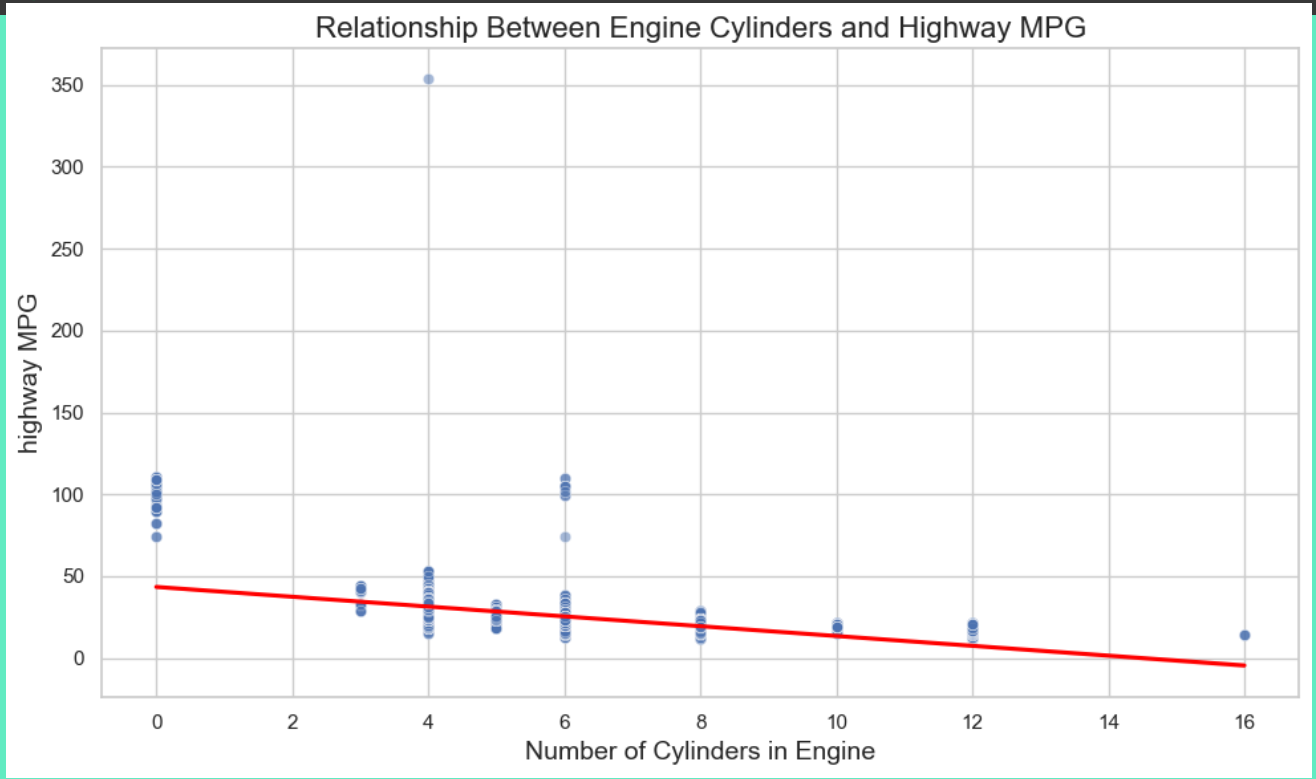
Make	MSRP
Plymouth	3.122902e+03
Oldsmobile	1.154254e+04
Suzuki	1.790721e+04
Pontiac	1.932155e+04
Scion	1.993250e+04
Mazda	2.003938e+04
Mitsubishi	2.124054e+04
Dodge	2.239006e+04
FIAT	2.267024e+04
Hyundai	2.459704e+04
Subaru	2.482750e+04
Kia	2.531017e+04
Honda	2.667434e+04
Chrysler	2.672296e+04
Ford	2.739927e+04
Saab	2.741350e+04
Volkswagen	2.810238e+04
Buick	2.820661e+04
Chevrolet	2.835039e+04
Volvo	2.854116e+04
Nissan	2.858343e+04
Toyota	2.903002e+04
GMC	3.049330e+04
Acura	3.488759e+04
HUMMER	3.646441e+04
Infiniti	4.239421e+04
Lincoln	4.283983e+04
Genesis	4.661667e+04
Lexus	4.754907e+04
Audi	5.345211e+04
Cadillac	5.623132e+04
BMW	6.154676e+04
Alfa Romeo	6.160000e+04
Land Rover	6.782322e+04
Lotus	6.918828e+04
Mercedes-Benz	7.147623e+04
Tesla	8.525556e+04
Porsche	1.016224e+05
Maserati	1.142077e+05
Aston Martin	1.979104e+05
Spyker	2.133233e+05
Ferrari	2.382188e+05
McLaren	2.398050e+05
Bentley	2.471693e+05
Lamborghini	3.315673e+05
Rolls-Royce	3.511306e+05
Maybach	5.462219e+05
Bugatti	1.757224e+06

➢ Task 4.B: Creating a bar chart to visualize the relationship between manufacturer & average price.



What is the relationship between fuel efficiency and the number of cylinders in a car's engine?

- Task 5. A: Creating a scatter plot and trend line with the number of cylinders on the x-axis and highway MPG on the y-axis.



What is the relationship between fuel efficiency and the number of cylinders in a car's engine?

- Task 5.B: Calculate the correlation coefficient between the number of cylinders and highway MPG

```
# Calculate the correlation coefficient
correlation = car_data['Engine Cylinders'].corr(car_data['highway MPG'])

# Print the correlation coefficient
print(f"Correlation coefficient between Engine Cylinders and Highway MPG:
```

```
Correlation coefficient between Engine Cylinders and Highway MPG: -0.60
```

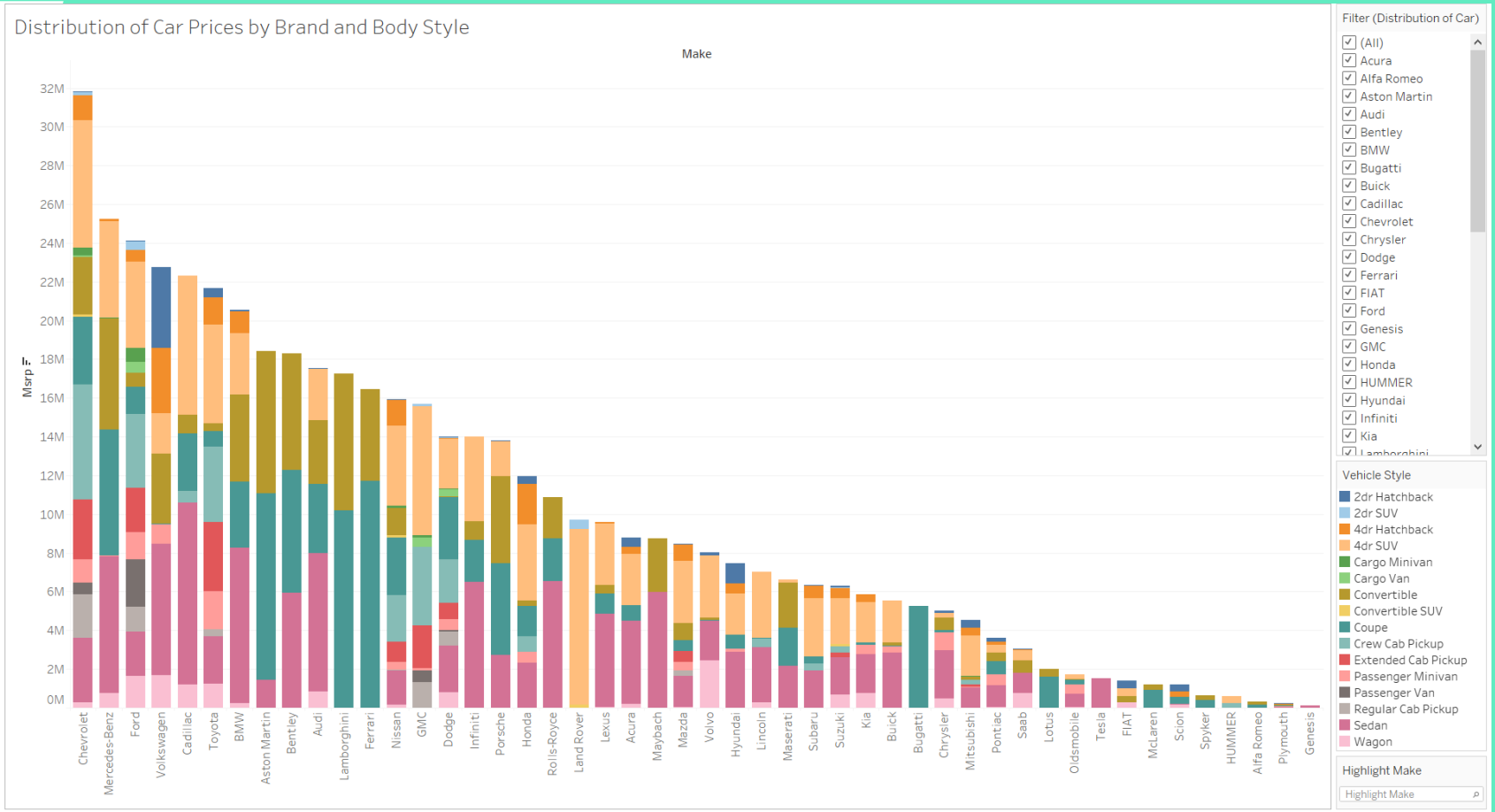
- Inference :- A negative value indicates a negative correlation. i.e., as number of cylinders increases, highway MPG decreases

C

Building the Dashboard

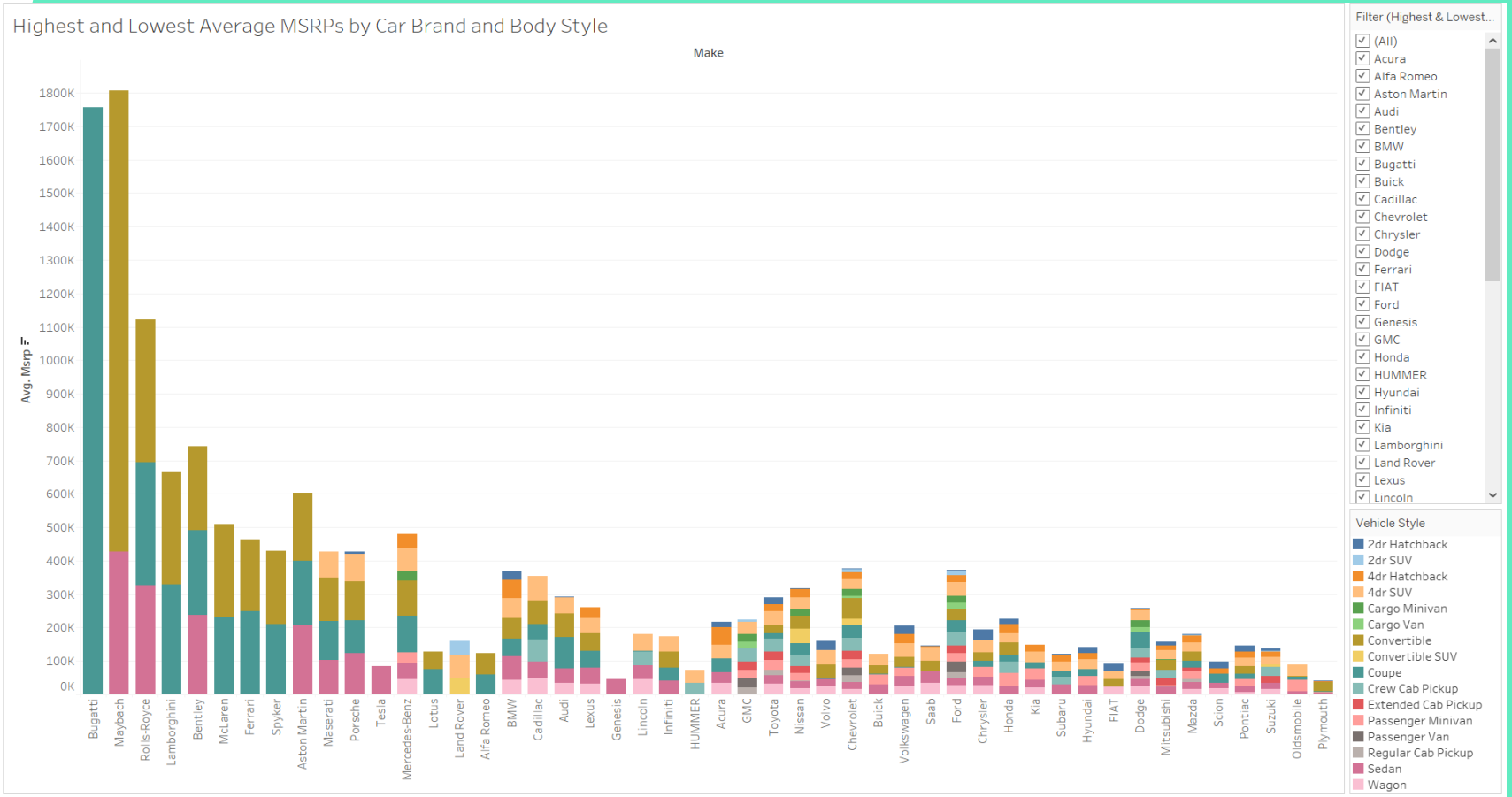
Creating the Interactive Dashboard using filters and slicers.

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

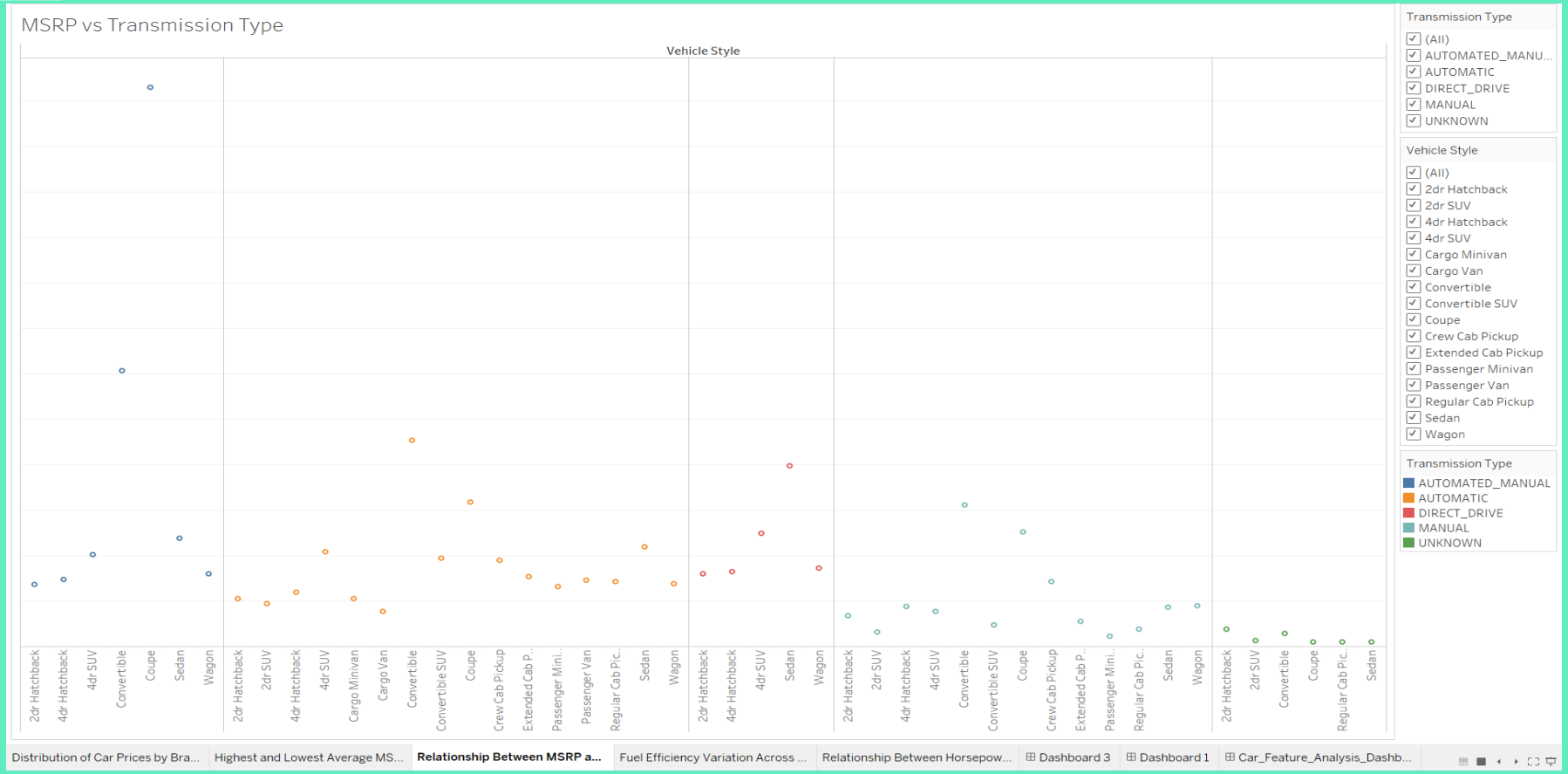


Distribution of Car Prices by Br... | Highest and Lowest Average MS... | Relationship Between MSRP and... | Fuel Efficiency Variation Across ... | Relationship Between Horsepow... | Dashboard 3 | Dashboard 1 | Car_Feature_Analysis_Dashb...

Task 2: Highest and Lowest Average MSRPs by Car Brand and Body Style

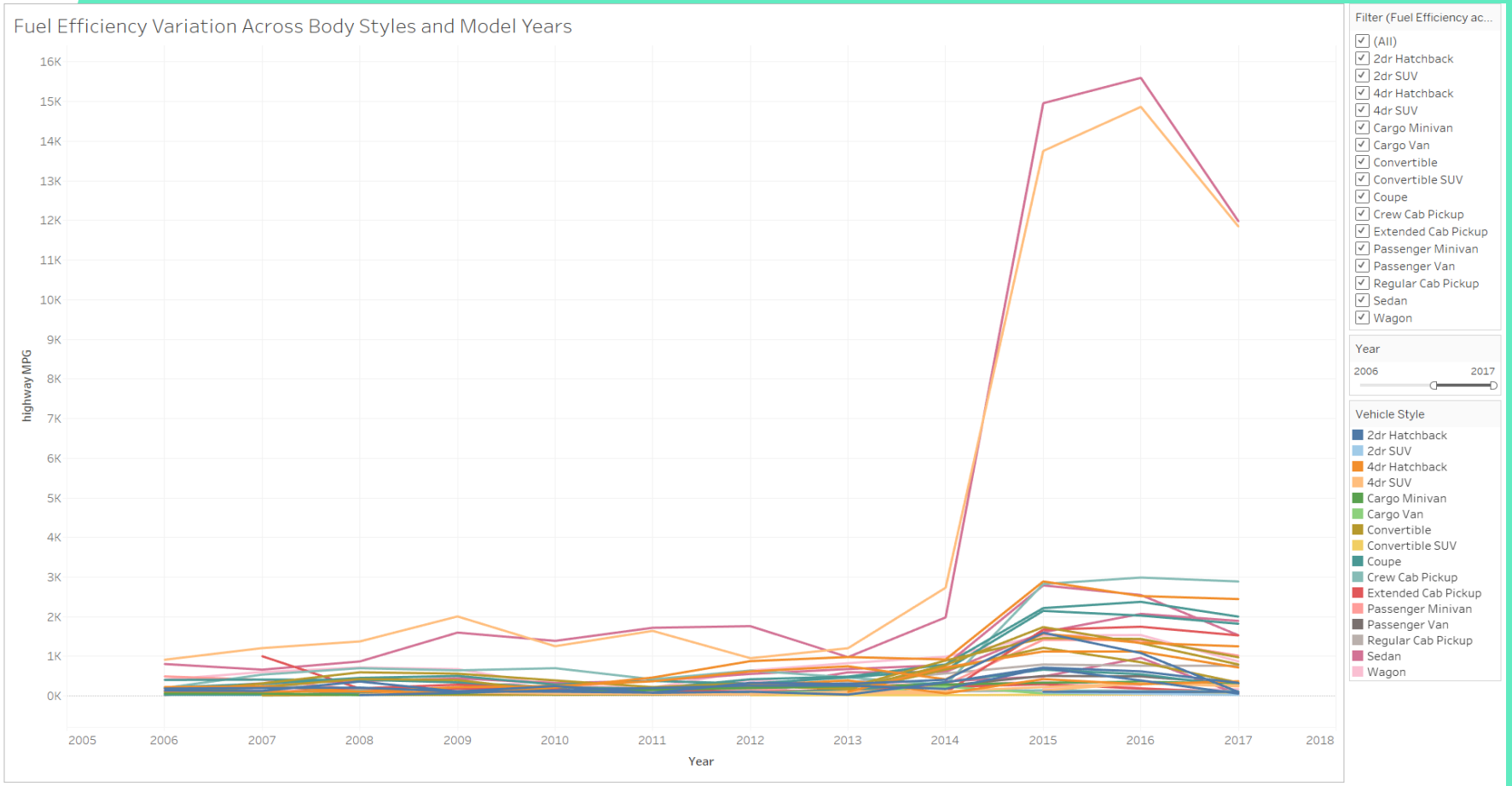


Task 3: How do the different features such as transmission type affect the MSRP, and how does this vary by body style?

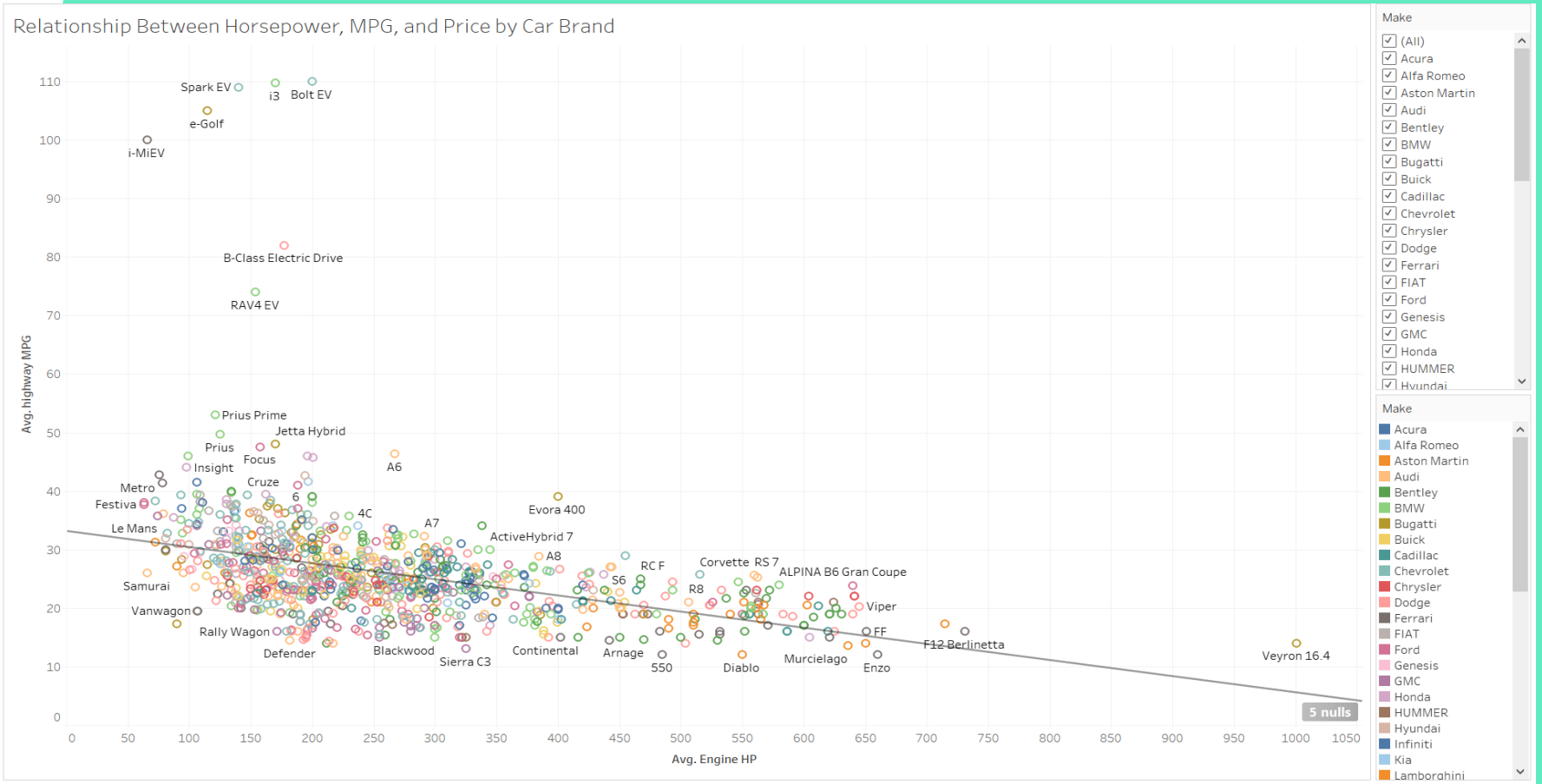


Distribution of Car Prices by Bra... | Highest and Lowest Average MS... | **Relationship Between MSRP a...** | Fuel Efficiency Variation Across ... | Relationship Between Horsepow... | Dashboard 3 | Dashboard 1 | Car_Feature_Analysis_Dashb...

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



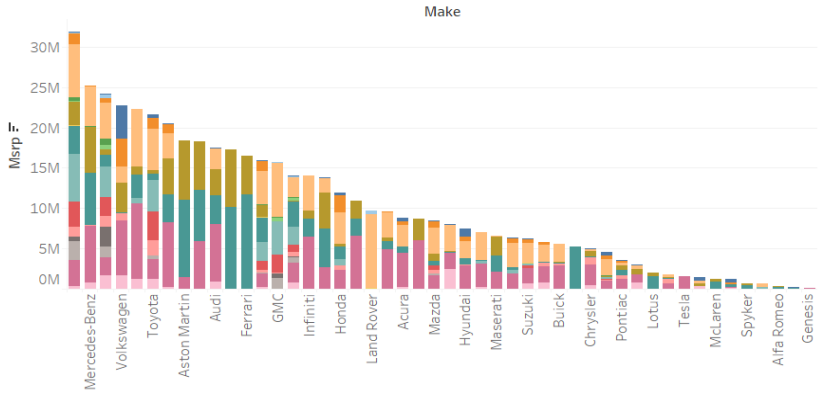
Task 5: How does the car's horsepower, MPG, and price vary across different Brands?



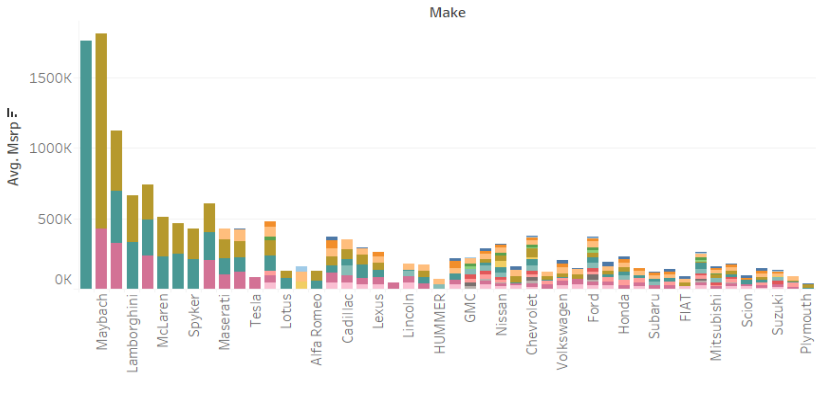
Distribution of Car Prices by Bra... | Highest and Lowest Average MS... | Relationship Between MSRP and... | Fuel Efficiency Variation Across ... | **Relationship Between Horsepo...** | Dashboard 3 | Dashboard 1 | Car_Feature_Analysis_Dashb...

Car_Feature_Analysis_Dashboard

Distribution of Car Prices by Brand and Body Style



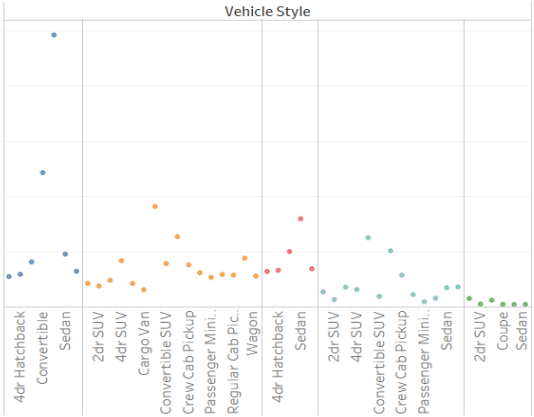
Highest and Lowest Average MSRPs by Car Brand and Body Style



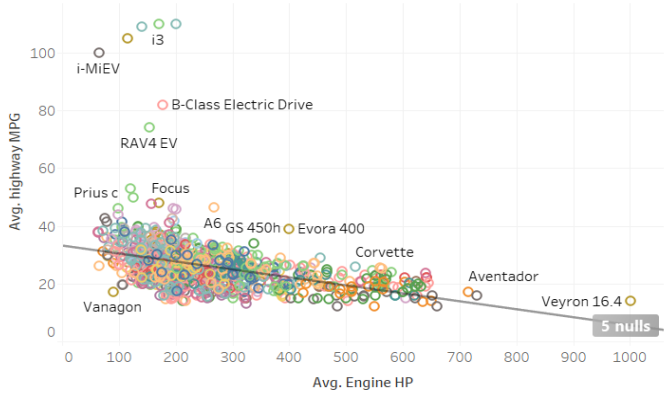
Filter (Distribution ..
(All)
Acura
Alfa Romeo
Aston Martin
Audi
Bentley
BMW
Bugatti
Buick
Cadillac
Chevrolet
Chrysler
Dodge

Filter (Highest & Lo..
(All)
Acura
Alfa Romeo
Aston Martin
Audi
Bentley
BMW
Bugatti
Buick
Cadillac
Chevrolet
Chrysler
Dodge

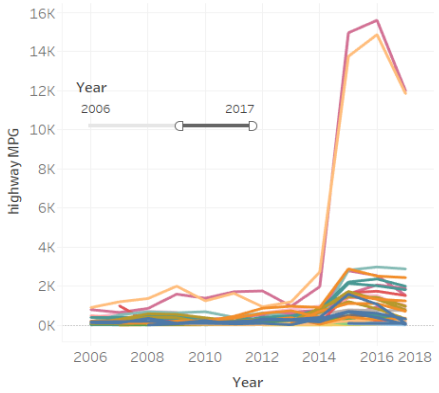
MSRP vs Transmission Type



Relationship Between Horsepower, MPG, and Price by Car Brand



Fuel Efficiency Variation Across Body Styles and Model Years



Filter (Fuel Efficienc..
(All)
2dr Hatchb...
2dr SUV
4dr Hatchb...
4dr SUV
Cargo Miniv...
Cargo Van
Convertible
Convertible ...
Coupe
Crew Cab Pi...
Extended C...
Passenger ...

Distribution of Car Prices by Bra... | Highest and Lowest Average MS... | Relationship Between MSRP and... | Fuel Efficiency Variation Across ... | Relationship Between Horsepow... | Dashboard 3 | Dashboard 1 | Car_Feature_Analysis_Dash...

Conclusion

- Thus, I have completed a Car Feature Analysis.
- Given key findings and all meaningful trends or patterns I have discovered.
- I have learned to use Python Libraries (pandas, matplotlib, numpy and seaborn) to analyze the dataset.
- I have learned how to create a dashboard using Tableau Public.
- I have learned to use Jupyter Notebook for Data Analysis.
- GitHub Repository and drive links are given as follows.

GitHub Repository:- https://github.com/ShindeYash/Impact_of_Car_Features_Analysis

Jupyter Notebook:- https://drive.google.com/file/d/1oebLWE6TFEC1Yix7CrDRP-FvPtNPxEVj/view?usp=sharing

Drive Link:- https://drive.google.com/drive/folders/1UE-rOaYAqR_Qnp4i-erhj4NEAJKw8vyw?usp=sharing

Video Presentation:- https://www.loom.com/share/fd5d07602fa04530a0209a6bb0d357e2?sid=c97d33e5-4b25-412d-bb78-68d6d9d48c61

Thanks!

Do you have any questions?

yashpradeepshinde@gmail.com
Yash Shinde