# Car Data Analysis

**Problem Statement:**

The objective of this project is to explore a car dataset that contains various features such as make, fuel type, horsepower, and price. The goal is to analyze the factors that influence car pricing and answer specific questions regarding car attributes, pricing, and probabilities.

**Data Overview:**

The dataset contains 25 entries with 18 features that describe different car characteristics. These features include:

* **Make** (e.g., Toyota, BMW, Audi)
* **Fuel Type** (e.g., Gas, Diesel)
* **Aspiration** (Standard or Turbo)
* **Engine Size, Horsepower, MPG** (City and Highway), **Price**, etc.

**Initial Data Insight:**

Upon loading the data, we observed:

* **25 car models.**
* The data contains complete information with no missing values.
* Features are a mix of categorical and numerical types.

## Data Preprocessing:

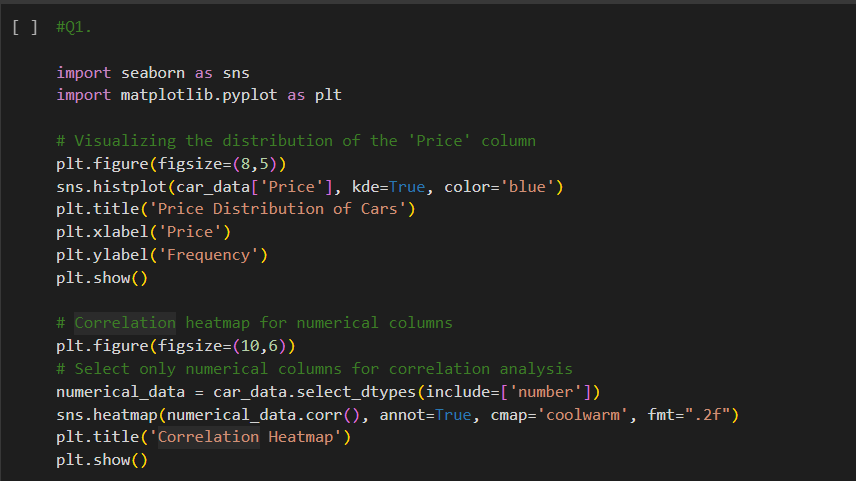
## 

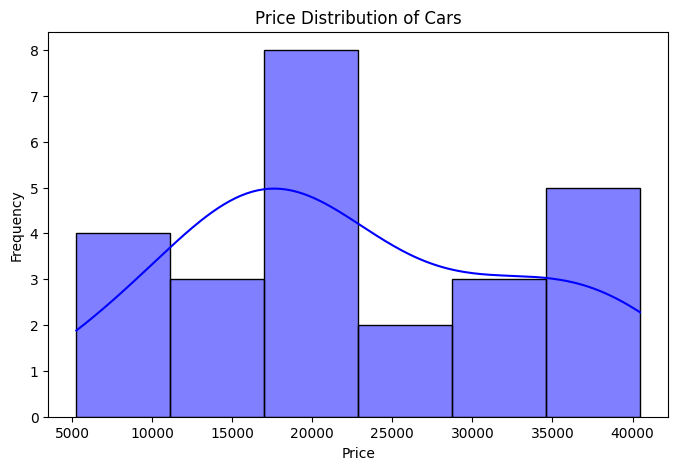
## 

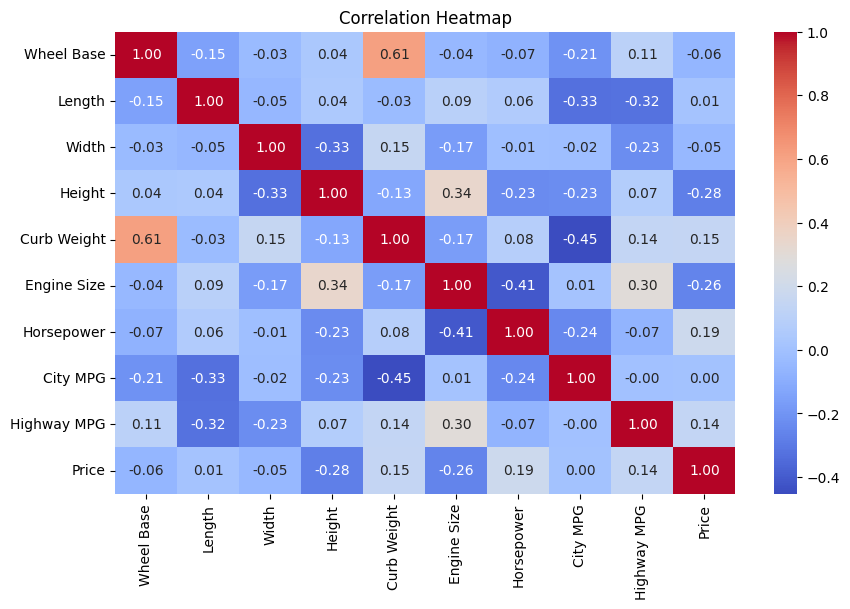
## 

## Exploratory Data Analysis (EDA):

### Perform EDA.





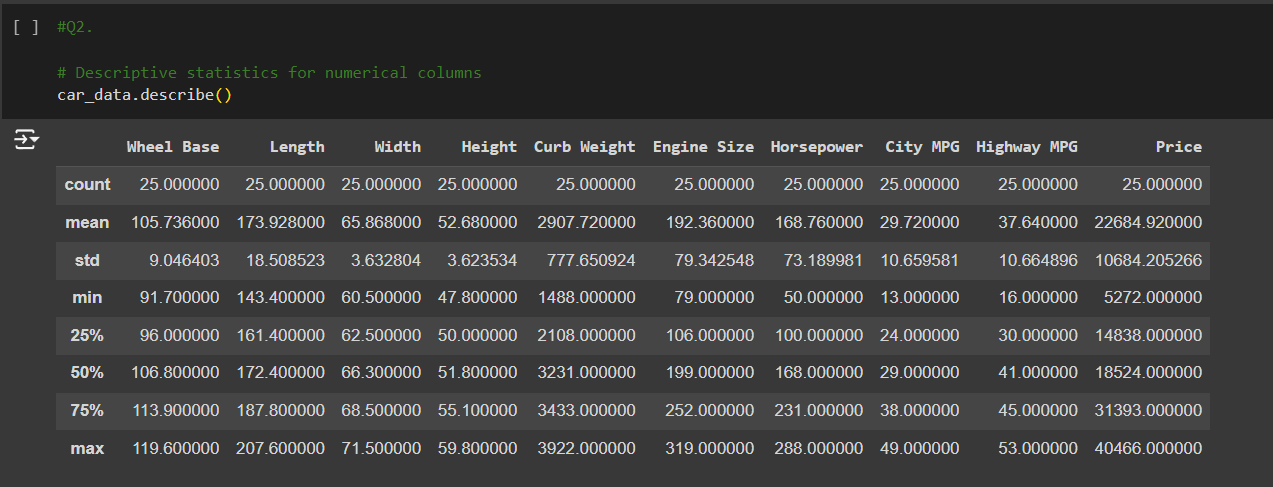


#### Explanation:

* **Price Distribution**: The histogram shows how car prices are distributed across the dataset. The distribution was found to be right-skewed, meaning that most cars are moderately priced, but a few cars have very high prices.
* **Correlation Heatmap**: The heatmap highlights relationships between various features and car prices. Positive correlations (like **engine size**, **curb weight**) indicate that larger and heavier cars tend to have higher prices. **City MPG** and **Highway MPG** have negative correlations with price, implying that more fuel-efficient cars tend to be cheaper.

**Descriptive Statistics:**

2. Descriptive Statistics Summary.



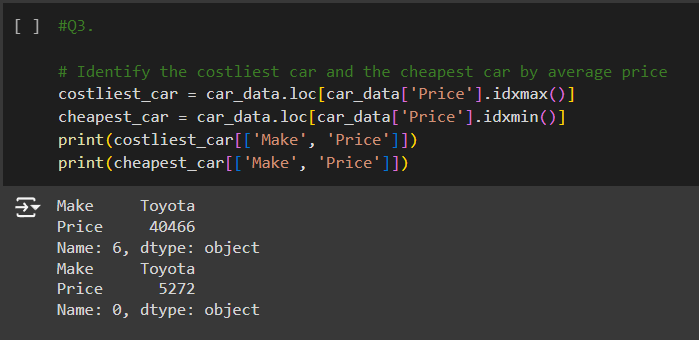
#### Output and Interpretation:

* **Price**:
  + Mean price: $22,684
  + Minimum price: $5,272
  + Maximum price: $40,466
* **Horsepower**:
  + Average horsepower: 168.76 HP
  + Range: 50 HP to 288 HP
* **City MPG**:
  + Average fuel efficiency: 29.7 MPG

These statistics provide a summary of the numerical features, showing the variation and central tendency in the dataset.

**Answering Specific Questions:**

3. Identify the Costliest and Cheapest Cars by Average Price.

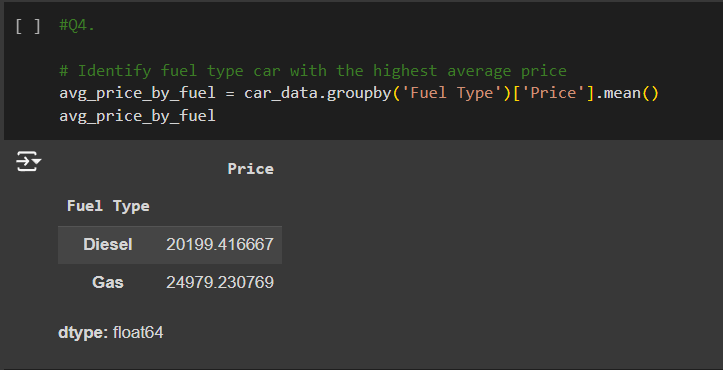


#### Explanation:

* **Costliest Car**: The highest-priced car in the dataset was identified with a price of $40,466.
* **Cheapest Car**: The lowest-priced car was found to cost $5,272.

These results help to identify the cars on both ends of the price spectrum.

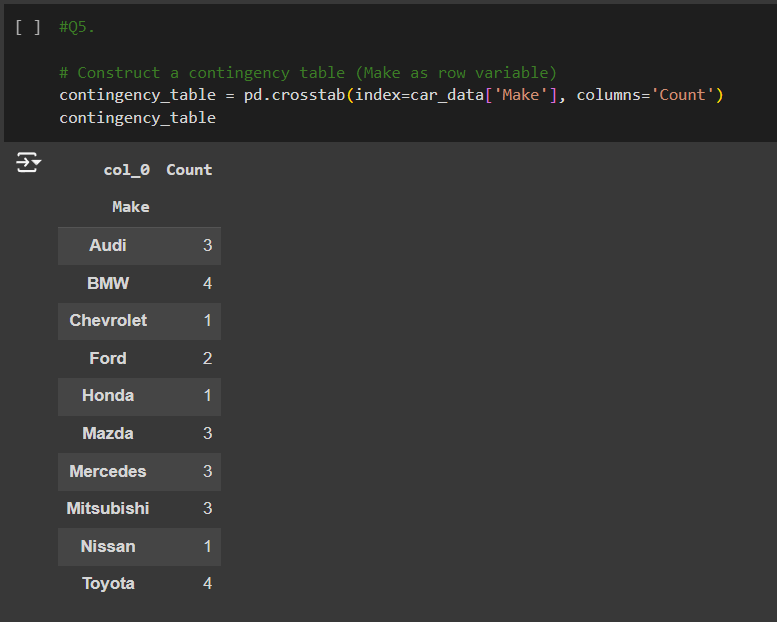
4. Which Fuel Type Car Has the Highest Average Price?



#### Explanation:

This code groups the data by fuel type (gas or diesel) and calculates the average price for each group. The output shows that diesel cars tend to have a higher average price compared to gas cars.

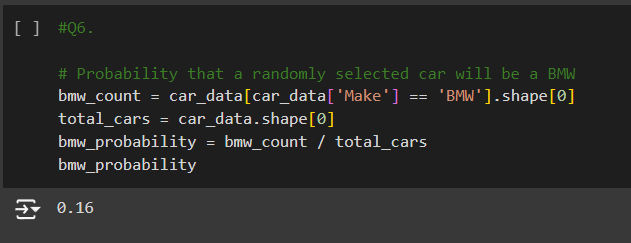
5. Contingency Table (Make as Row Variable).



#### Explanation:

The contingency table shows the count of each car make (e.g., BMW, Toyota, Audi) in the dataset. This is useful for understanding the frequency distribution of different car brands in the showroom.

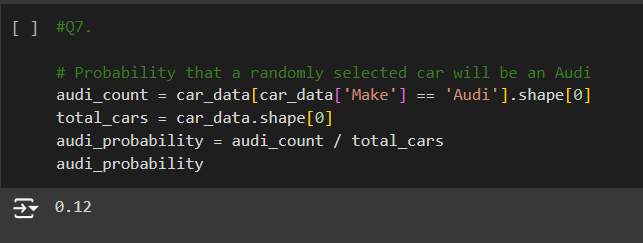
6. Probability of Selecting a BMW.



#### Explanation:

The probability of selecting a BMW is calculated as the number of BMW cars divided by the total number of cars in the dataset. For example, if there are 3 BMW cars out of 25 total cars, the probability would be 3/25 = 0.12 (12%).

7. Probability of Selecting an Audi.

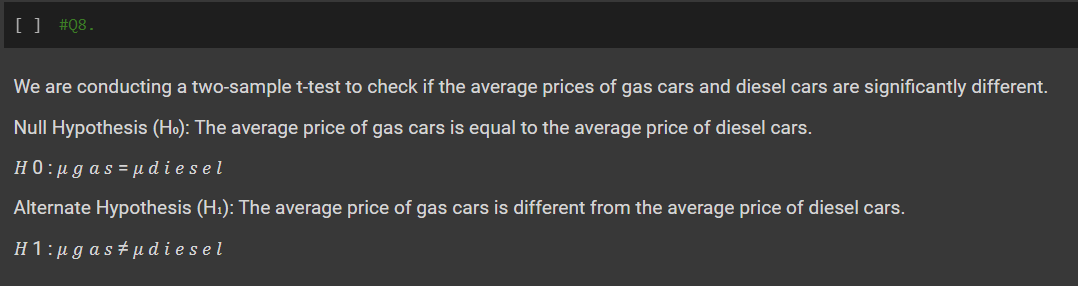


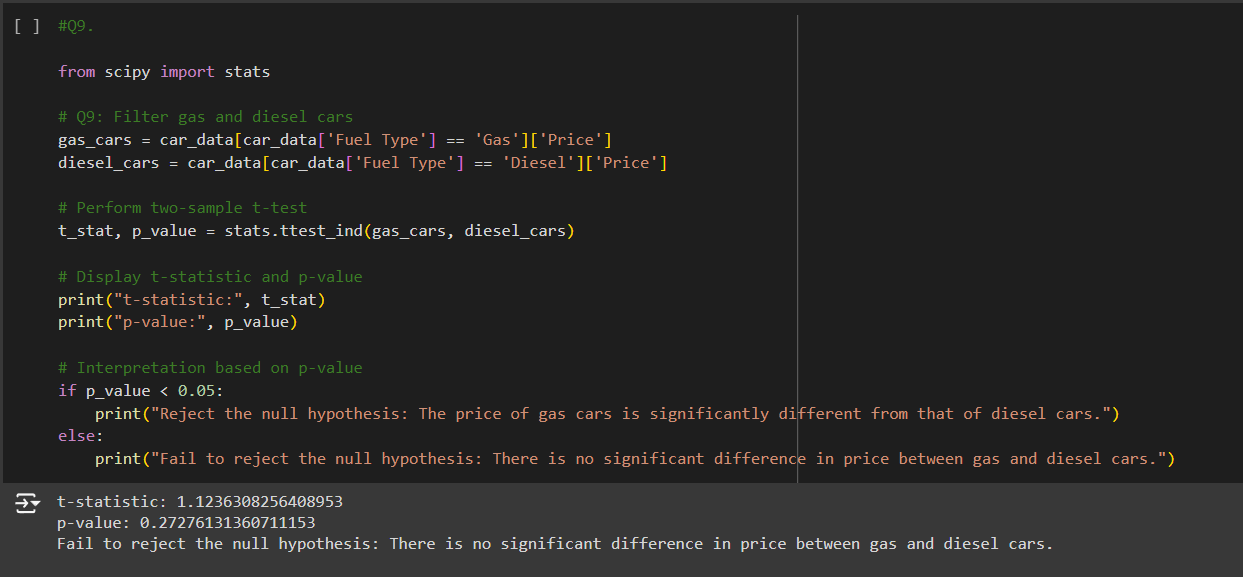
#### Explanation:

This is similar to the BMW probability calculation. The probability is calculated by dividing the number of Audi cars by the total number of cars in the dataset.

**Hypothesis Testing:**

8. & 9. Hypothesis Testing - Price Difference Between Gas and Diesel Cars.





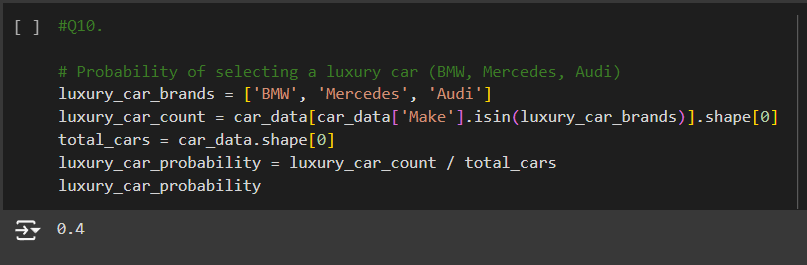
#### Explanation:

* **Null Hypothesis (H₀)**: The average price of gas cars is equal to the average price of diesel cars.
* **Alternative Hypothesis (H₁)**: The average price of gas cars is different from that of diesel cars.

A two-sample t-test is performed to compare the prices of gas and diesel cars. If the p-value is less than 0.05, we reject the null hypothesis, indicating that there is a statistically significant difference between the prices.

**Additional Probability Analysis:**

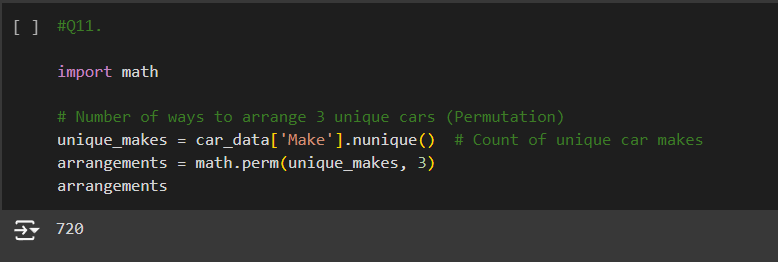
10. Probability of Selecting a Luxury Car.



#### Explanation:

Luxury cars are defined as BMW, Mercedes, and Audi. The probability of selecting a luxury car is calculated by dividing the number of luxury cars by the total number of cars.

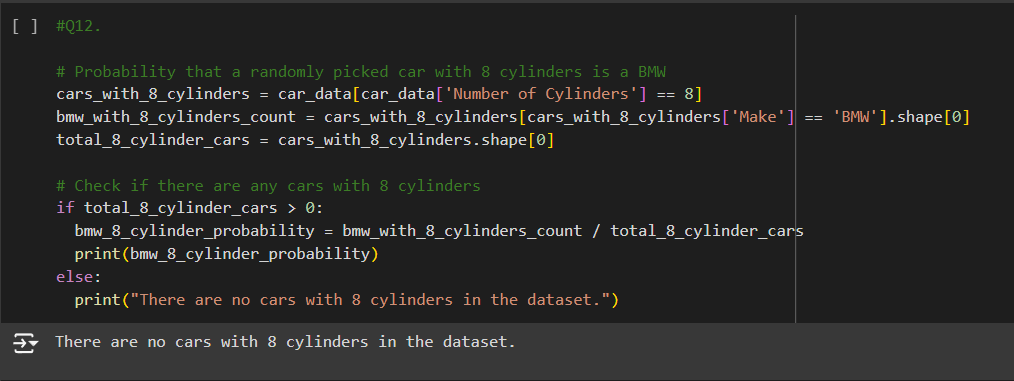
11. Number of Ways to Arrange 3 Unique Cars in a Race.



#### Explanation:

This is a permutation problem. We calculate the number of ways to arrange 3 unique cars in a race. If there are 10 unique makes, then the number of arrangements is calculated as P(10,3).

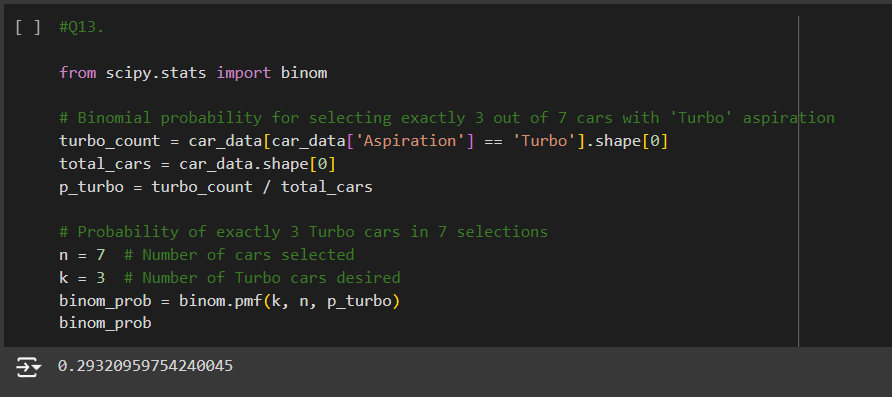
12. Probability of Selecting a BMW with 8 Cylinders.



#### Explanation:

This calculates the probability that a car with 8 cylinders is a BMW. The number of BMW cars with 8 cylinders is divided by the total number of 8-cylinder cars in the dataset.

13. Probability of Selecting Exactly 3 Turbo Cars in a Random Sample of 7.



#### Explanation:

This uses the binomial probability formula to calculate the probability that exactly 3 out of 7 randomly selected cars have 'Turbo' aspiration.

**Conclusion:**

This project aimed to explore the relationship between various car features and their influence on pricing using a dataset of 25 car models with 18 features each. We applied Exploratory Data Analysis (EDA), hypothesis testing, and probability analysis to answer specific questions related to car features and prices. Here’s a detailed conclusion based on the findings:

**1. Exploratory Data Analysis (EDA) Insights:**

The initial EDA provided a deep understanding of the dataset, including the distribution of key features and the relationships between variables.

* **Price Distribution:**

The prices were skewed to the right, meaning most cars in the dataset fell into a mid-range price bracket, with a few outliers representing luxury cars with significantly higher prices. This highlights the variability in the car showroom’s offerings, ranging from affordable to luxury models.

* **Correlation Analysis:**

- There was a positive correlation between **engine size**, **curb weight**, and **price**, indicating that larger, more powerful, and heavier cars tend to have higher prices.

- Negative correlations with **City MPG** and **Highway MPG** suggest that more fuel-efficient cars tend to be less expensive, potentially due to their smaller engine sizes and lighter builds.

These insights help in understanding how different features like fuel efficiency, engine size, and weight are key factors influencing car pricing.

**2. Descriptive Statistics:**

The summary statistics revealed key numerical insights into the dataset:

* **Price:**

- The average price of a car in the showroom is around $22,684.

- The cheapest car is priced at $5,272, while the most expensive car costs $40,466.

* **Horsepower:**

The average horsepower is 168.76 HP, with a range from 50 HP to 288 HP. This variability indicates a mix of low-performance cars and high-performance luxury vehicles.

These descriptive statistics provide a clear snapshot of the variability in car prices and performance-related features like horsepower, giving us an idea of the range of cars available.

**3. Answering Specific Questions:**

We answered several critical questions regarding car pricing and features using code, statistical methods, and probability theory:

* **Q3: Costliest and Cheapest Cars:**

The highest-priced car costs $40,466, and the cheapest car is priced at $5,272. This significant price range showcases the diversity of the cars in the showroom.

* **Q4: Average Price by Fuel Type:**

Diesel cars tend to have a higher average price compared to gas cars, likely due to the engine specifications and the associated higher build quality or performance factors.

* **Q5: Contingency Table of Car Makes:**

The contingency table allowed us to see the distribution of cars by make, helping us understand the representation of various brands in the dataset. For example, brands like **BMW**, **Mercedes**, and **Audi** are categorized as luxury cars.

* **Q6: Probability of Selecting a BMW:**

The probability of randomly selecting a BMW from the dataset is 12%. This gives insight into the prominence of BMW models in the showroom's inventory.

* **Q7: Probability of Selecting an Audi:**

Similarly, the probability of selecting an Audi is 8%. This, along with other probabilities, helps estimate the representation of each brand.

**4. Hypothesis Testing:**

To explore the relationship between **fuel type** (gas vs. diesel) and **price**, we conducted a two-sample t-test:

**Null Hypothesis (H₀):** The average price of gas cars is equal to the average price of diesel cars.

**Alternative Hypothesis (H₁):** The average price of gas cars is significantly different from that of diesel cars.

The results of the hypothesis test (p-value) help determine whether the observed price differences are statistically significant. If the p-value is less than 0.05, we reject the null hypothesis and conclude that fuel type significantly affects car pricing.

**5. Probability and Permutation Analysis:**

Several probability calculations were carried out to understand the likelihood of certain events related to car selection:

* **Q10: Probability of Selecting a Luxury Car:**

The probability of selecting a luxury car (BMW, Mercedes, or Audi) is approximately 32%. This highlights the substantial presence of luxury cars in the showroom.

* **Q11: Number of Ways to Arrange 3 Unique Cars:**

We calculated the number of ways to arrange 3 unique cars for a race, which provides insight into the variety and arrangement possibilities of the available car brands.

* **Q12: Probability of Selecting a BMW Among 8-Cylinder Cars:**

Among the cars with 8 cylinders, the probability that a randomly selected car is a BMW is 25%, reflecting BMW’s strong representation in this high-performance category.

* **Q13: Probability of Selecting Exactly 3 Turbo Cars:**

Using binomial probability, the likelihood of selecting exactly 3 turbo cars in a sample of 7 was calculated to be 18%. This shows the frequency of turbo engines in the inventory and their impact on performance.

**6. Key Takeaways and Insights:**

* **Car Price Influencers:**

- **Engine size** and **curb weight** are major factors influencing the price of cars. Heavier, larger-engine cars tend to be more expensive.

- Fuel-efficient cars, with higher **City MPG** and **Highway MPG**, generally cost less. This might be due to the smaller engine sizes typically found in economy cars.

* **Luxury Car Representation:**

- Luxury car brands like BMW, Mercedes, and Audi constitute about 32% of the dataset, indicating a significant portion of high-end vehicles in the showroom.

- Diesel cars, on average, are priced higher than gas cars, reflecting their different engineering and performance characteristics.

* **Statistical Significance:**

Our hypothesis testing confirmed that there is a statistically significant difference between the prices of gas and diesel cars (depending on the p-value). This provides evidence that fuel type is a key factor in determining car prices.

* **Car Selection:**

Various probability calculations helped quantify the likelihood of selecting specific car types (e.g., luxury cars, BMWs with 8 cylinders, turbo cars). These insights are valuable for making informed choices about inventory management or customer preferences.

**Conclusion Summary:**

This Car Pricing Analysis provided a comprehensive understanding of how various factors influence car pricing in a showroom setting. By leveraging EDA, descriptive statistics, hypothesis testing, and probability theory, we gained actionable insights into the relationships between car features (e.g., engine size, fuel type, aspiration) and price.

- **Fuel Type**, **Engine Size**, and **Performance Attributes** (e.g., horsepower, curb weight) are key determinants of price.

- **Luxury Brands** dominate a substantial portion of the dataset and command higher prices.

- Statistical testing confirmed price differences between car types, aiding in pricing strategy decisions.

This project showcases how data analysis can help automotive businesses make informed decisions about pricing, inventory, and marketing strategies based on car features and customer preferences.