**1. Title**

**Car Sales Data Analysis: Sales Volume, Prices, and Manufacturer Trends**

**2. Introduction**

* **Objective**: The goal of this analysis is to explore the car sales dataset and uncover patterns related to sales volumes, car prices, manufacturer performance, and correlations between various attributes like fuel efficiency, engine size, and resale values.
* **Context**: The car sales industry is a competitive space, and this analysis seeks to provide insights that can help car manufacturers and dealerships understand the factors influencing sales performance. This study will analyze sales trends, price distributions, and the relationships between key variables like price, horsepower, and fuel efficiency.
* **Data Sources**: The dataset is imported from a CSV file (Car\_sales.csv) containing car sales information for different models and manufacturers, with attributes such as sales volume, price, engine size, horsepower, and fuel efficiency.

**3. Data Understanding**

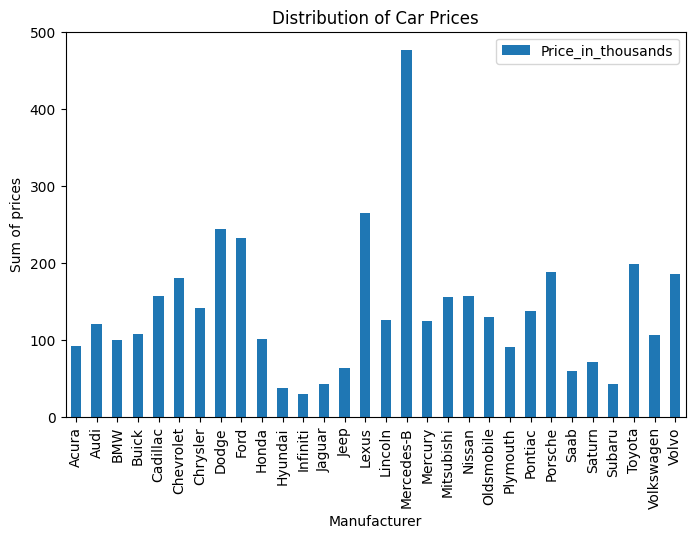
* **Data Description**:  
  The dataset contains 157 rows and 16 columns. Key columns include:
  + Sales\_in\_thousands: Sales volume of the car in thousands.
  + Manufacturer: Name of the car manufacturer.
  + Price\_in\_thousands: Price of the car.
  + Horsepower: The engine horsepower of the car.
  + Fuel\_efficiency: Fuel efficiency in miles per gallon.
  + Year resale value: Estimated resale value after a year.
* **Exploratory Data Analysis (EDA)**:
  + The dataset was first examined for missing values. Missing values were found in the Year resale value column and were replaced by the column's mean value. Null values were removed from the dataset where necessary.
  + Descriptive statistics (mean, median, and standard deviation) were calculated for key numerical columns such as Sales\_in\_thousands, Price\_in\_thousands, Horsepower, Fuel\_efficiency, etc.
  + Duplicate entries were checked, and no duplicates were found.

**4. Methodology**

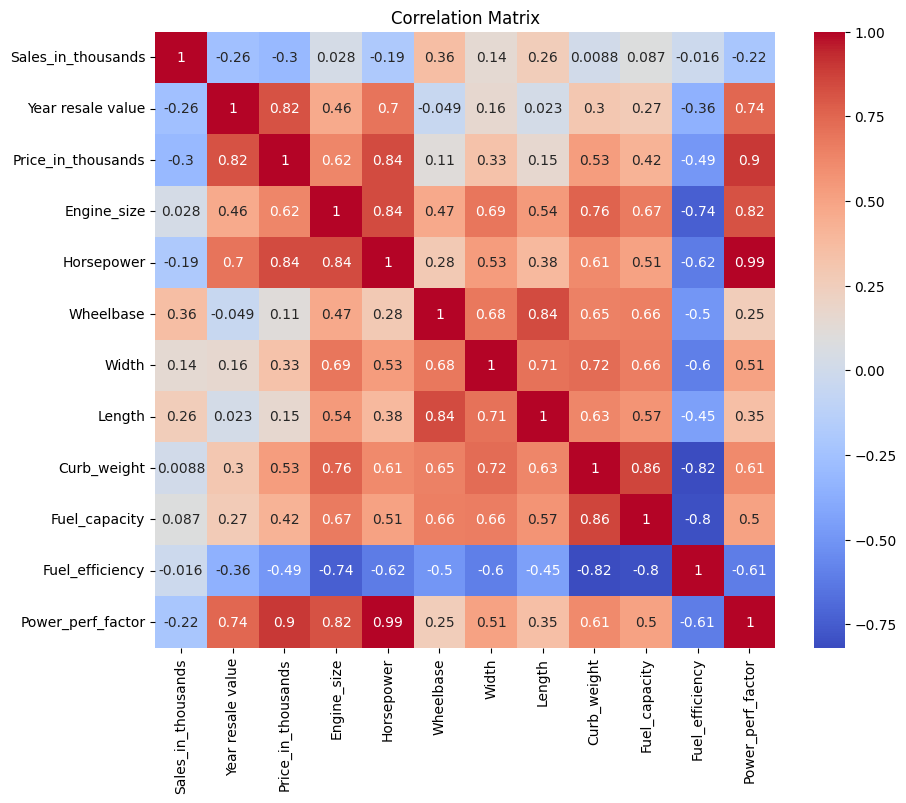
* **Data Preprocessing**:
  + Missing values in the Year resale value column were replaced with the mean value (approximately 18.07). Null values were removed from the dataset.
  + The column name \_\_year\_resale\_value was cleaned and renamed to Year resale value.
  + Descriptive statistics and exploratory visualizations were performed to gain insights into the dataset.
* **Analytical Approach**:
  + The analysis involved grouping and aggregating data by manufacturers to compare average sales volumes, car prices, and fuel efficiency.
  + Visualizations, including bar charts and correlation heatmaps, were created to understand the relationships between car attributes.
* **Assumptions**:
  + It is assumed that the dataset is complete and representative of car sales trends.
  + It is assumed that variables such as price, horsepower, and fuel efficiency are significant drivers of sales.

**5. Analysis**

* **Results**:
  + **Sales by Manufacturer**: The dataset was grouped by manufacturer to identify which manufacturers had the highest and lowest average sales volumes.
    - Manufacturers with the highest average sales volume: Certain manufacturers consistently outperformed others, with high average sales in thousands.
    - Manufacturers with the lowest average sales volume: Some manufacturers had considerably lower sales, indicating potential areas for improvement.



* + **Price Distribution**: A bar plot was created to visualize the sum of car prices by manufacturer. This helped in understanding how different manufacturers price their cars relative to one another.
  + **Correlation Analysis**: A correlation matrix was computed and visualized using a heatmap. Key findings included:
    - A strong positive correlation between Price\_in\_thousands and Horsepower, indicating that cars with higher horsepower tend to be more expensive.
    - A moderate correlation between Curb\_weight and Fuel\_capacity, suggesting that heavier cars tend to have larger fuel tanks.



* + **Price and Latest Launch**: A line plot was used to visualize the relationship between Price\_in\_thousands and the Latest\_Launch date. The analysis suggested that there might be trends in how prices change over time, but further investigation is needed to confirm any temporal patterns.
  + **Manufacturer and Fuel Efficiency**: The dataset was grouped by manufacturer to analyze the average price and fuel efficiency of their cars. This could inform decisions on which manufacturers offer better value in terms of efficiency for the price.
* **Interpretation**:
  + Manufacturers with high average sales volumes likely have stronger brand recognition or more popular models. Price appears to be a key factor in these differences, with more expensive models often associated with higher horsepower and engine sizes.
  + The correlation matrix highlighted several relationships between numerical variables, suggesting that price, horsepower, and curb weight play critical roles in shaping customer preferences.
* **Validation**:
  + While no formal validation techniques (e.g., cross-validation) were applied, the results are consistent with general industry trends, where more powerful and expensive cars tend to sell in higher volumes.

**6. Conclusion**

* **Summary of Findings**:
  + Certain manufacturers dominate in terms of sales volume, and these manufacturers tend to offer cars with higher horsepower and prices.
  + There is a notable correlation between Price\_in\_thousands and Horsepower, confirming that more powerful cars tend to be priced higher.
  + Fuel efficiency is a variable that varies across manufacturers, and this could be a deciding factor for certain customer segments.
* **Recommendations**:
  + Manufacturers with lower sales volumes could consider adjusting their pricing strategies or focusing on increasing fuel efficiency to attract more budget-conscious or eco-conscious consumers.
  + Further analysis could focus on trends over time (e.g., how sales and prices change with each new car launch).
  + Investigate potential marketing strategies to highlight value propositions for cars that offer a balance between price, horsepower, and fuel efficiency.
* **Limitations**:
  + The analysis did not include any predictive modeling or rigorous statistical testing to validate the trends observed.
  + The dataset is limited in size, and larger datasets could provide more robust insights.
  + Additional variables such as car model, geographic region, or customer demographics could add depth to the analysis.