Cars Data Exploration

1. Basic Descriptive & Distribution Analysis

A. What is the distribution of car prices (show both histogram and boxplot)?

B. What is the distribution of horsepower, torque, top speed and acceleration?

C. How many cars belong to each fuel type (countplot)?

D. What is the frequency of companies in the dataset (i.e., how many models does each company have)?

E. How many electric vs hybrid vs petrol cars are there?

2. Comparison Analysis

A. What is the average price of cars by fuel type?

B. Compare horsepower, torque, and acceleration across fuel types.

C. Which fuel type offers the most torque for the least battery capacity? (Use scatterplot of torque vs. battery capacity for each fuel type.)

D. How do electric vs hybrid vs ICE (internal combustion engine) vehicles compare in acceleration?

3. Trend Analysis (specify bins)

A. Is there a trend between horsepower and top speed?

Hint: Create horsepower bins <150, 150–300, 300–500, 500–1000, 1000+ and show median max speed in each bin.

B. How does battery capacity affect acceleration?

Hint: Bin battery capacity into <50, 50–60, 60–90, 90+ and plot median 0–100 km/h time for each bin.

C. Does price increase with engine capacity or battery capacity?

Hint: For engine capacity, bin as <1500, 1500–2000, 2000–3000, 3000–4000, 4000+; for battery capacity use the same <50, 50–60, 60–90, 90+ bins.

D. How does torque correlate with acceleration?

Hint: Bin torque into <300, 300–500, 500–800, 800+ and show median for each bin.

4. Ranking Analysis

A. Which companies have the most expensive cars?

B. What are the top 10 fastest accelerating cars (lowest 0–100 km/h time)?

C. Rank the fuel types based on average car price.

D. Which cars offer the most horsepower per dollar? (Compute Horsepower/Price.)

5. Contribution / Variance Analysis

A. What percentage of total car models are contributed by each company?

B. Which fuel type contributes most to the total horsepower in the dataset?

C. What is the price contribution of electric cars compared to others?

D. How much does the average price of each fuel type deviate from the overall mean car price?

6. Correlation Analysis

A. Are horsepower and torque strongly correlated?

B. Does battery capacity correlate with car price? (Focus on Electric & Hybrid only.)

C. Is acceleration more influenced by torque or horsepower?

D. Is there any multicollinearity among features? (Use correlation heatmap, threshold = 0.7.)

7. Frequency Analysis

A. Count of car models by company and fuel type (stacked or grouped barplot).

B. What is the most common engine size?

Hint: Use the same engine displacement bins as above <1500, 1500–2000, 2000–3000, 3000–4000, 4000+ cc.

C. How frequent are cars with battery capacity above a threshold (say ≥70 kWh)?

8. Pareto Analysis (80/20 Rule)

A. Which ~20 % of companies account for 80 % of high-performance cars?

Hint: Define high performance as 0–100 km/h time ≤10 sec and compute cumulative counts.

B. Which few car models contribute most to the total price value of the dataset (80 % of total price)?

C. Which top companies dominate the electric car space (80 % of electric car total price)?

9. Extra

A. How does the average car price vary by fuel type within each company?

B. Which companies offer the most powerful electric vehicles?

C. Compare acceleration across companies for each fuel type (use median 0–100 km/h times).

D. How does battery capacity distribution differ between electric and hybrid cars? (Use KDE and boxplot.)