

Part 3: Advanced Analytics

7. Statistical Analysis

Problem Statement: Perform a statistical analysis on a given dataset to identify significant trends and correlations. Provide a summary of your findings.

Summary of Statistical Analysis

Dataset Overview: The dataset contains information about various car attributes including cylinders, displacement, horsepower, weight, acceleration, year, origin, and mpg.

1. Data Exploration and Correlation Analysis:

- **Correlation Matrix:** Calculated and visualized the correlation between continuous variables such as cylinders, displacement, horsepower, weight, acceleration, and mpg.
- **Heatmap:** Used a heatmap to display correlations, indicating how strongly different variables are related.

2. Distribution of Continuous Variables:

- **Histograms:**
 - **MPG:** Displayed the distribution of mpg with a histogram and KDE plot.
 - **Horsepower:** Analysed the distribution of horsepower similarly.

3. Scatter Plots for Relationship Analysis:

- **MPG vs. Weight:** Visualized the relationship between mpg and weight.
- **MPG vs. Displacement:** Visualized the relationship between mpg and displacement.

4. Analysis by Categorical Variables:

- **Frequency Counts:**
 - **Cylinders:** Showed the frequency of different numbers of cylinders.
 - **Origin:** Showed the frequency of car origins.
- **Average MPG:**
 - **By Cylinders:** Calculated and visualized the average mpg by the number of cylinders.
 - **By Origin:** Calculated and visualized the average mpg by car origin.

5. Hypothesis Testing:

- **T-Test:** Compared the mean mpg between cars with 4 cylinders and 6 cylinders, finding whether the difference is statistically significant.
- **ANOVA:** Tested for significant differences in mpg across different numbers of cylinders, assessing if at least one group mean is significantly different.

Conclusion: The analysis provides insights into the relationships between various car attributes and mpg, and evaluates statistical significance in differences observed across different groups of cylinders. The visualizations and statistical tests help in understanding patterns and trends within the dataset.

