

# **CAR PRICE ANALYSIS**

Exploring factors influencing car prices

By Asish Jose

# INTRODUCTION

Analysis of the car price dataset to explore whether specific features can be used to predict car prices. The dataset includes various attributes such as engine type, fuel system, dimensions, horsepower, and fuel efficiency.

# PROJECT GOALS

- Identify Key Price Determinants: Determine the most influential factors that significantly affect car prices.
- Analyze Feature Correlations with Price: Explore the relationship between car price and various features using correlation analysis to identify key drivers of price variation

# DATA CLEANING

- Missing Values Handling: For columns like 'num-of-doors' (which might have missing values), you can either impute with the mode (most frequent value) or remove rows if imputation isn't feasible.
- Outliers Treatment: Identify outliers in numerical features such as 'curb-weight', 'engine-size', and 'horsepower' using techniques like Z-score or IQR. You could remove or cap extreme values to avoid their influence on the model.
- Normalization: Normalize continuous variables like 'engine-size', 'curb-weight', and 'horsepower' to bring them onto a similar scale for better model performance. This can be done using Min-Max Scaling or Standardization.
- Categorical to Numerical: Convert categorical variables such as 'make', 'num-of-doors', 'body-style', 'engine-type', and 'fuel-system' to numerical values using encoding methods like one-hot encoding or label encoding, depending on the model you're using.

# EXPLORATORY DATA ANALYSIS (EDA)

- Summary Statistics: Calculate mean, median, and standard deviation for numerical features like 'engine-size', 'horsepower', 'curb-weight', and 'price'. This will give a general idea of central tendency and spread.
- Visualizations:
  - Use histograms for understanding the distribution of 'price', 'horsepower', and 'engine-size'.
  - Box plots to identify outliers in 'curb-weight', 'engine-size', and 'price'.
  - Scatter plots to visualize relationships, for example, between 'horsepower' and 'price' or 'engine-size' and 'price'.
- Feature Correlations: Show correlations between features like 'engine-size', 'horsepower', and 'price'. A heatmap can help visualize these relationships.

# STATISTICAL ANALYSIS

- Hypothesis Testing: Performed ANOVA to determine if there are significant differences in 'price' based on categorical features like 'aspiration-std' vs 'aspiration-turbo' or 'fuel-system' (diesel vs gas).
- Covariance and Correlation Matrices: Analyze how variables like 'engine-size', 'horsepower', 'curb-weight', and 'price' are related. A high correlation between 'horsepower' and 'price' may suggest a strong relationship for predictive modeling.
- Inferential Statistics: Use regression analysis or p-values to validate which features significantly influence 'price'. Features like 'engine-size', 'curb-weight', and 'horsepower' may show high significance.

# KEY FINDINGS & CONCLUSION

- Most Influential Factors: Based on the analysis, highlight that 'engine-size', 'horsepower', and 'drive-wheels' are likely the most influential features affecting car price. For example, cars with a larger 'engine-size' tend to have higher prices.
- Strategic Insights: These findings can help in setting more accurate pricing strategies for cars based on key features like 'engine-size' and 'horsepower'. They can also help predict prices for future car models by focusing on influential features.



**THANK YOU**