Car Data Analysis Project

Project Overview

This project focuses on analyzing a dataset of cars to understand various factors that affect car prices. The main goals are to clean the data, organize it, and prepare it for analysis to derive meaningful insights.

Dataset Description

The dataset contains information about different cars, including:

Make : The brand of the car (e.g., Toyota, BMW).
Fuel Type : The type of fuel the car uses (e.g., gas, diesel).
Engine Size : The size of the car's engine.
Number of Doors : How many doors the car has (e.g., two, four).
Price : The selling price of the car.
- Additional specifications like `wheel_base`, `curb_weight`, `horsepower`, etc.

The dataset includes 205 rows and 26 columns.

Data Cleaning Steps

1. Handling Missing Values:
- Replaced missing values represented by `?` in columns like `normalized losses` and `num_of_doors` with the mean or mode values.

2. Data Type Conversion:

- Converted columns such as `normalized losses`, `horsepower`, `peak_rpm`, and `price` to integer type.
- Converted columns `bore` and `stroke` to float type.

3. String Replacement:
 - Replaced `-` with a space in the `make` column to improve readability.

4. Categorical to Numerical Conversion:
 - Converted text values in the `num_of_doors` column (`four`, `two`) to numeric values (`4`, `2`).

5. Handling Outliers:
  - Assessed and managed outliers in numerical columns to ensure data accuracy.

Analysis Conducted

- Descriptive statistics to summarize key numerical columns.
- Visualization of data using bar charts and pie charts to compare car prices by make and fuel type.
- Analysis of the relationship between car features (e.g., engine size, fuel type) and car prices.

Key Findings

- Larger engine sizes are generally associated with higher car prices.
- Fuel type significantly impacts car pricing, with some fuel types being more expensive than others.

Conclusion

The project provides insights into how various car features influence their market price. By cleaning and organizing the dataset, we ensured the reliability of our analysis, which can aid in decision-making for car purchases and sales.

Future Work

- Further analysis to explore the impact of other features like body style and drive wheels on car prices.
- Implementing machine learning models to predict car prices based on dataset features.

Getting Started

Prerequisites

- Python 3.x
- Pandas
- Matplotlib