**Used Car Data Analysis Project**

**Overview**

This project involves analyzing a dataset of used cars to derive insights and visualize trends. The project covers data cleaning, feature engineering, exploratory data analysis (EDA), and visualizations to uncover patterns related to car pricing, fuel types, brands, and other features.

**Key Features**

* **Data Cleaning**:
  + Handled missing values and checked for duplicates.
  + Removed unnecessary columns for effective analysis.
* **Feature Engineering**:
  + Created new features such as car\_age to understand the influence of a car's age on its price.
  + Split the Name column into brand and model for better analysis.
* **Exploratory Data Analysis (EDA)**:
  + Analyzed the distribution and relationships of numerical and categorical variables.
  + Conducted visualizations using **Matplotlib** and **Seaborn** to highlight trends and correlations.
* **Visualizations**:
  + Distribution plots for numerical variables.
  + Horizontal bar charts for categorical variables.
  + Violin plots, pair plots, bubble charts, and heatmaps to reveal key insights.
  + Stacked bar charts and pie charts to compare features.

**Data**

The dataset used in this project contains details of used cars, including:

* **Name**: Car brand and model.
* **Year**: Year of manufacture.
* **Fuel\_Type**: Type of fuel (e.g., Petrol, Diesel, CNG).
* **Transmission**: Gear system (e.g., Manual, Automatic).
* **Owner\_Type**: Type of ownership (e.g., First, Second).
* **Price**: Selling price of the car.
* And other related attributes.

**Key Insights**

1. **Fuel Type and Price**:
   * Violin plots reveal the impact of fuel type on the car price distribution.
   * A stacked bar chart shows the relationship between fuel type and transmission.
2. **Brand Trends**:
   * Identified the top brands and their frequency using horizontal bar charts.
   * Grouped similar brands to unify naming conventions (e.g., "Land Rover" and "Mini Cooper").
3. **Age Impact**:
   * Added a car\_age column to explore how the car's age affects its price.
4. **Correlation Analysis**:
   * Heatmaps and pair plots provided insights into numerical relationships.

**Technologies Used**

* **Python**:
  + Libraries: Pandas, NumPy, Matplotlib, Seaborn, Scipy.
* **Tools**:
  + PyCharm for code development.
  + Git and GitHub for version control.

**Project Setup**

1. Clone the repository:

<https://github.com/gobinathdatascientist/Car_sales_prediction_visualization>

1. Install required libraries:

pip install *pandas, matplotlib, seaborn, scikit-learn*

1. Run the Python script:

Cars\_sales\_prediction.py

**Visualizations**

The project includes:

* Histograms, boxplots, and KDE plots for distribution analysis.
* Stacked bar charts and heatmaps for categorical comparisons.
* Scatter plots and bubble charts for multivariate analysis.

**Future Improvements**

* Implement machine learning models to predict car prices based on features.
* Enhance interactivity with dashboards using tools like **Plotly** or **Tableau**.
* Automate feature engineering processes.

**Author**

**Gobinath**  
*Data Scientist*  
[GitHub Profile](https://github.com/yourusername)