

**Lab Manual**

**SUBJECT : Machine Learning**

**Submitted By : M Wahaaj Siddique**

**Submitted To : Sir Saeed**

**Registration Number : 2023-BS-AI-007**

**Table of Contents**

**Project 1: Car Price Prediction using Regression.................................................................2** Summary.....................................................................................................................................2 Objectives...................................................................................................................................3 Abstract......................................................................................................................................3Explanation of Steps..................................................................................................................3 Importing Libraries....................................................................................................................4 Data Cleaning.............................................................................................................................6 Data Transformation...................................................................................................................6 Linear Regression.......................................................................................................................7

Model Evaluation & Visualization.............................................................................................8 **Project 2: Car price Classification using ML........................................................................9** Summary...................................................................................................................................9Objectives.................................................................................................................................9Abstract....................................................................................................................................10Explanation of Steps.................................................................................................................11 Code Steps (1–22)...............................................................................................................12-14 Visualizations and Evaluation...............................................................................................,..15

Confusion Matrix………………………………………………......…………..…..………...16

Model evaluation and confusion matrix…………………………..…………….....................17

**Project Number 1:**

**1. Summary**

* Predicting car prices is a **real-world challenge** in the age of digital car marketplaces.
* The project leverages **regression techniques** to estimate the market value of cars based on multiple features.
* Uses a dataset containing attributes like **brand, model, year, mileage, engine size, fuel type, and transmission**.
* Helps buyers, sellers, and dealerships to make **data-driven pricing decisions**.
* Aims to build a model that is not only **accurate** but also **practical and scalable**.

**2. Objective**

* To analyze and uncover the **key factors influencing car prices**.
* To perform **data preprocessing** including cleaning, transformation, and feature engineering.
* To implement and compare **multiple regression models** (Linear, Random Forest, Gradient Boosting, etc.).
* To evaluate models using **performance metrics** such as MAE, RMSE, and R² Score.
* To develop a **smart predictive system** that can assist users in estimating the fair price of a used car.

**3. Abstract**

* In today’s fast-evolving automobile market, **pricing a vehicle correctly is crucial**.
* This project applies machine learning regression algorithms to predict used car prices with high precision.
* Features like **age of the car, brand reputation, mileage, and engine power** are analyzed to predict price variations.
* The dataset undergoes rigorous **cleaning, encoding, and feature scaling** to prepare it for modeling.
* Models such as **Linear Regression**, **Random Forest**, and **XGBoost** are trained and tested.
* The best-performing model serves as a **virtual car appraiser**, providing instant, reliable price estimates.
* This solution empowers users to make **confident and informed decisions** in car buying and selling.

**Explanation of Steps:**

1. Load the car price dataset and explore structure.

2. Clean the dataset and handle missing values.

3. Encode categorical features and scale numerical ones.

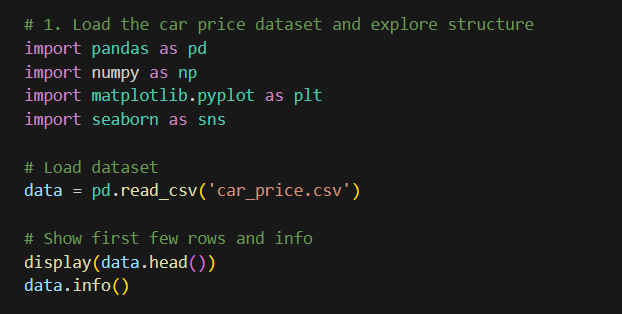
4. Split data into training and testing sets.

5. Train regression models like Linear Regression.

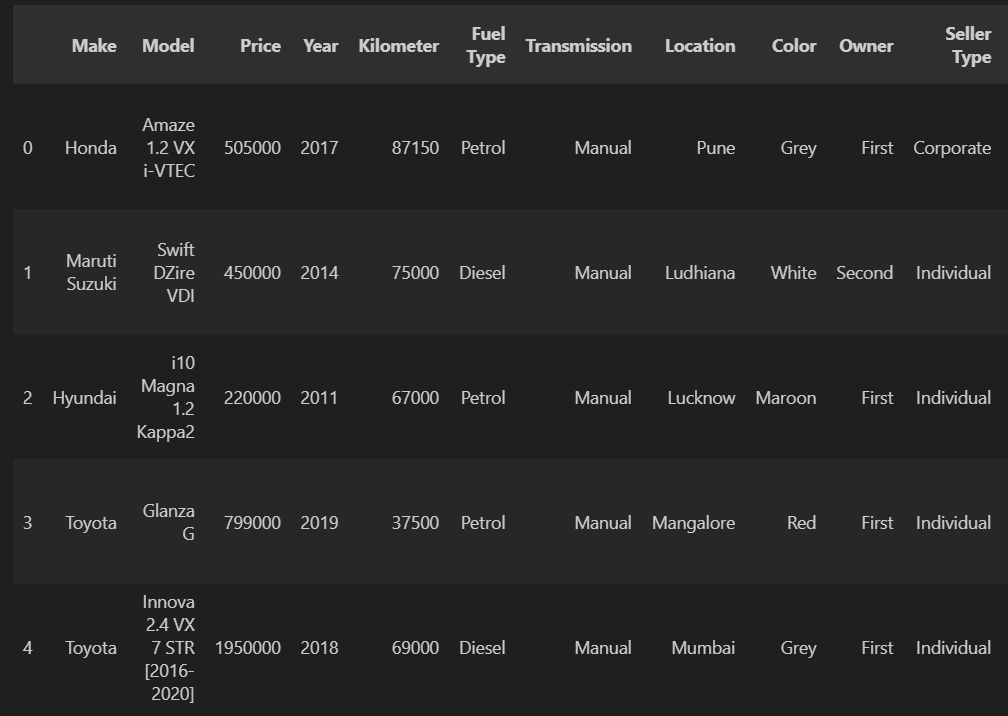
6. Evaluate the model using MSE, RMSE, and R² metrics.

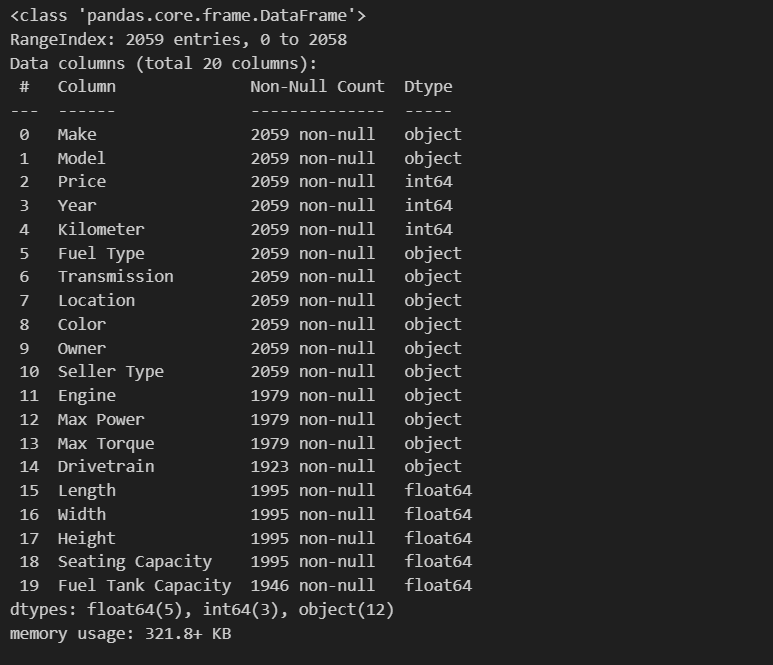
7. Visualize model predictions and data trends.

**1 : Importing Libraries:**

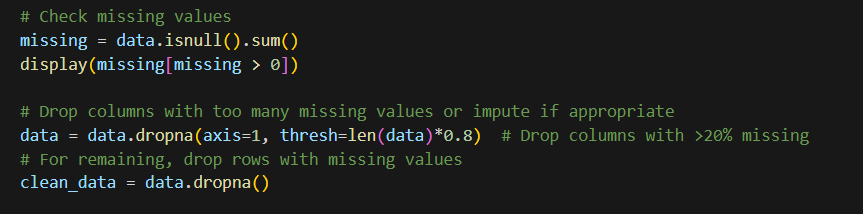


**OUTPUTS:**

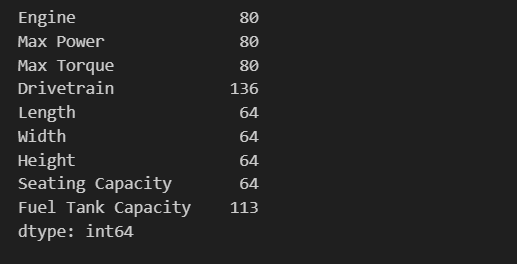




**2 : Cleaning The Dataset And Handling Missing Values:**

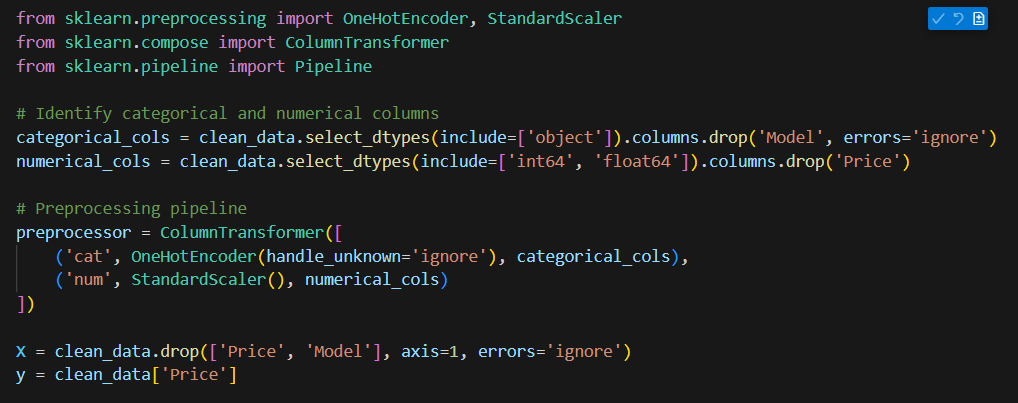


**OUTPUT:**

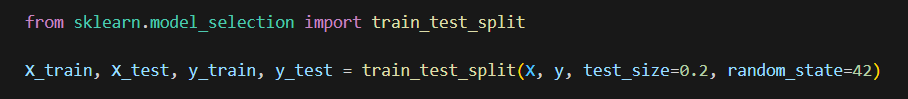


**3 : Encode categorical features and scale numerical ones:**

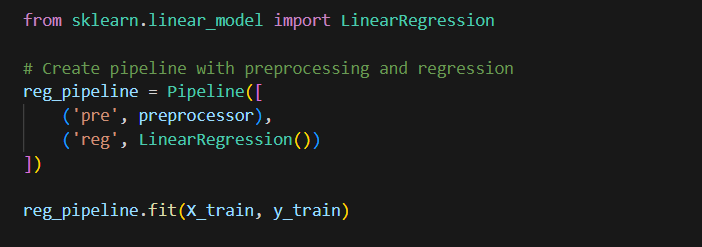
We will use one-hot encoding for categorical variables and StandardScaler for numeric features.



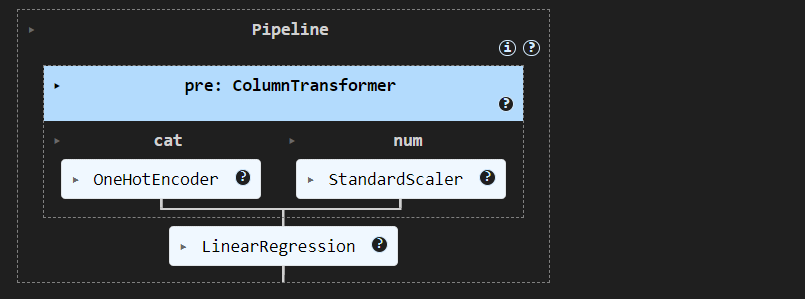
**4 : Split data into training and testing sets:**



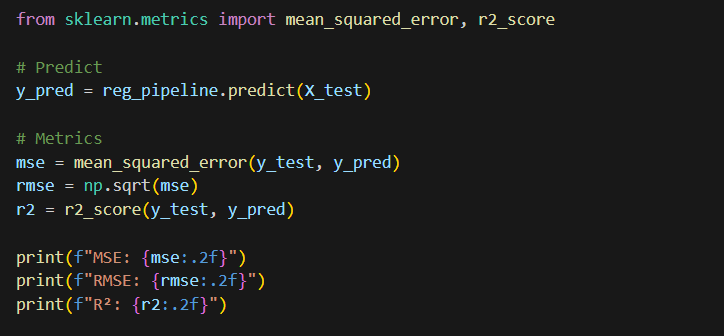
**5 : Train regression models like Linear Regression:**



**OUTPUT:**



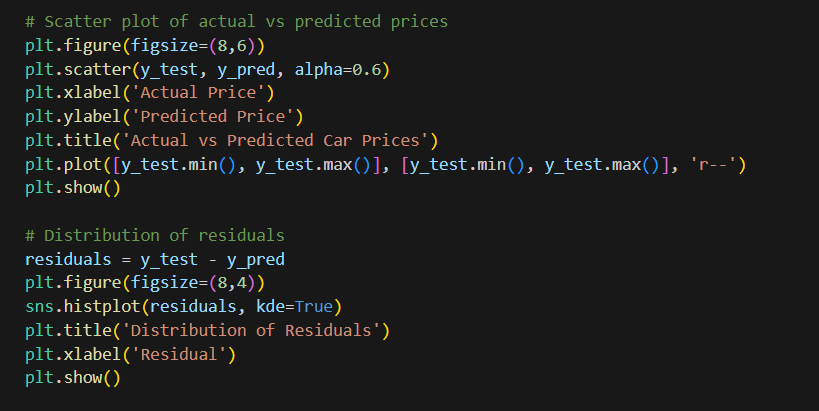
**6 : Evaluate the model using MSE, RMSE, and R² metrics:**



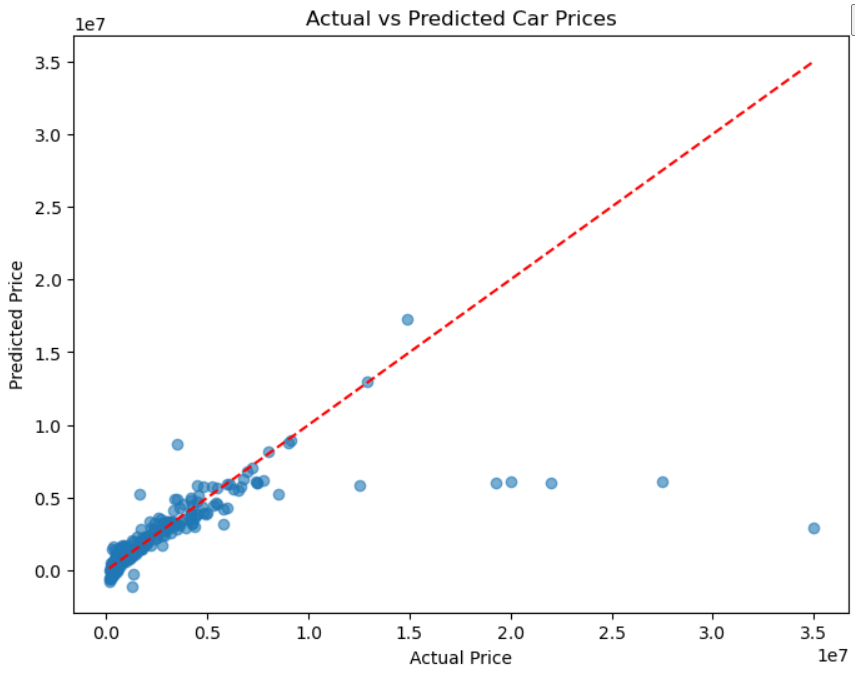
**OUTPUT:**

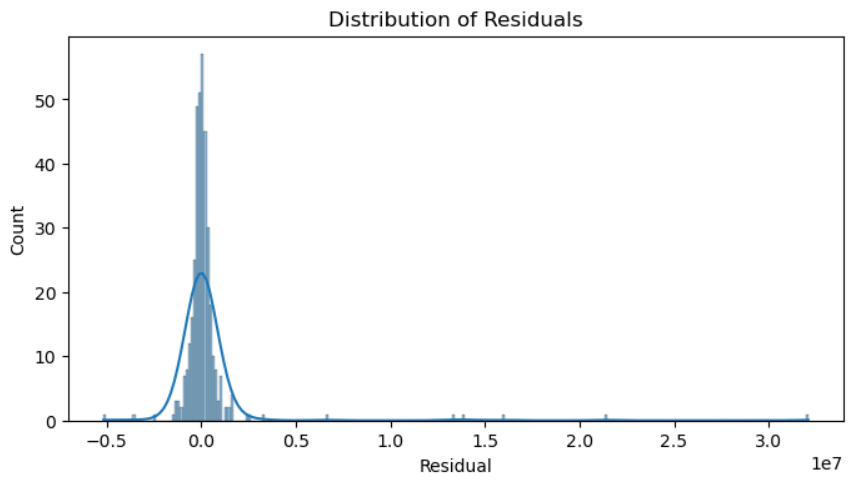


**7 : Visualize model predictions and data trends:**



**OUTPUT:**





**Project Number 2 : Diabetes Classification using Machine Learning:**

**1. Summary**

* Diabetes is a chronic disease affecting millions worldwide, and **early detection** can be life-saving.
* This project uses **machine learning classification algorithms** to identify whether a person is diabetic based on medical attributes.
* The dataset includes features like **glucose level, BMI, age, insulin levels, blood pressure, and more**.
* The model classifies individuals as **diabetic or non-diabetic** using data-driven insights.
* Aims to build a system that can **support medical professionals** and **raise early health alerts** for at-risk individuals.

**2. Objectives**

* To analyze key health indicators that contribute to the development of diabetes.
* To perform **data preprocessing**, handling missing values, outliers, and feature scaling.
* To implement and compare **classification algorithms** such as:
  + Logistic Regression
  + Decision Trees
  + Random Forest
  + Support Vector Machines (SVM)
  + K-Nearest Neighbors (KNN)
* To evaluate model performance using metrics like **accuracy, precision, recall, F1-score, and ROC-AUC**.
* To develop a **predictive tool** that can assist in early diagnosis and preventive care.

**3. Abstract**

* Diabetes is a serious metabolic disorder that, if undiagnosed, can lead to severe health complications.
* In this project, we develop a machine learning model to **classify individuals as diabetic or non-diabetic**.
* Using a labeled medical dataset (e.g., PIMA Indian Diabetes Dataset), the model is trained on various health indicators.
* The data is carefully preprocessed: **handling null values, normalizing features**, and balancing class distribution if needed.
* Several classification models are trained and compared to determine the most accurate and reliable approach.
* The final output is a **fast, accurate, and scalable prediction model** that can assist clinics, health apps, and patients in detecting diabetes risk.
* This project demonstrates how **technology can revolutionize early disease detection** and empower people with preventive healthcare insights.

**Explanation of Steps:**

1. Import necessary libraries and load the dataset.

2. Clean the dataset and handle missing/null values.

3. Perform exploratory data analysis using visualization tools.

4. Encode categorical variables and scale numerical features.

5. Split the dataset into training and testing sets.

6. Train classification models like logistic regression.

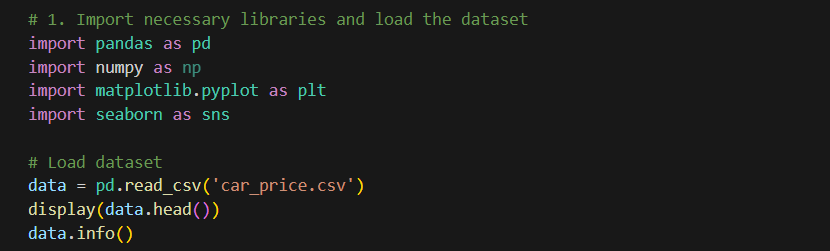
7. Evaluate model performance using metrics like confusion matrix and classification report.

8. Interpret results and visualize outcomes.

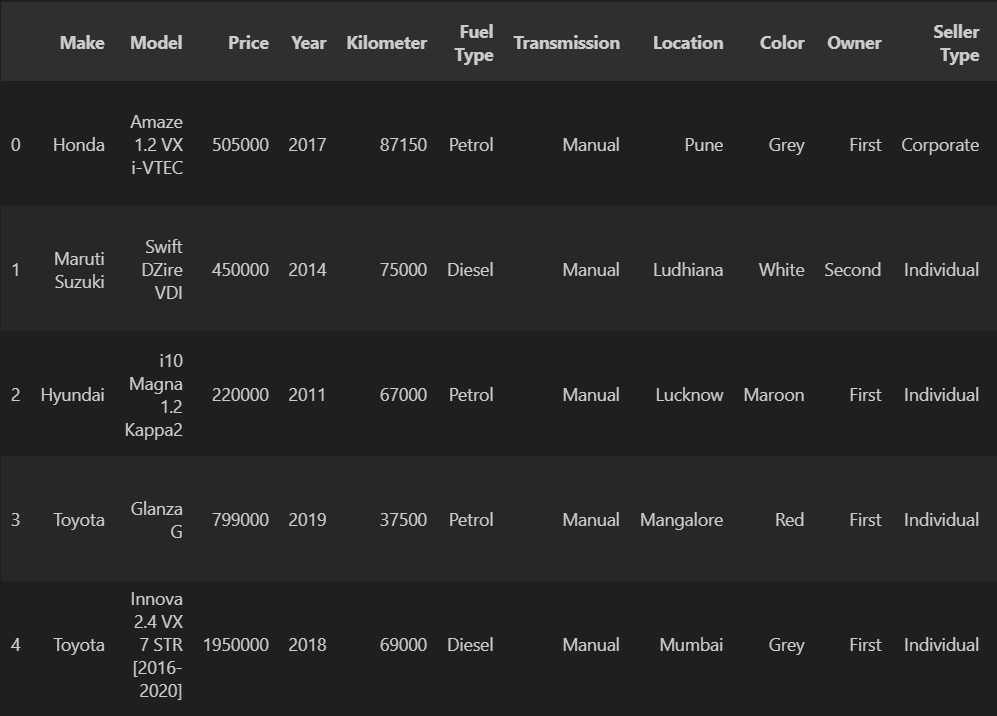
**STEPS:**

**1 : Car Price Classification using Logistic Regression:**

This project aims to classify car prices (e.g., high vs low) using machine learning classification techniques. We will use a dataset of used cars and follow a complete ML workflow.



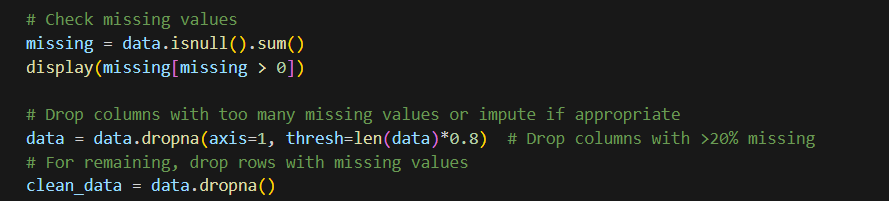
**OUTPUT:**





**2 : Clean the dataset and handle missing/null values:**

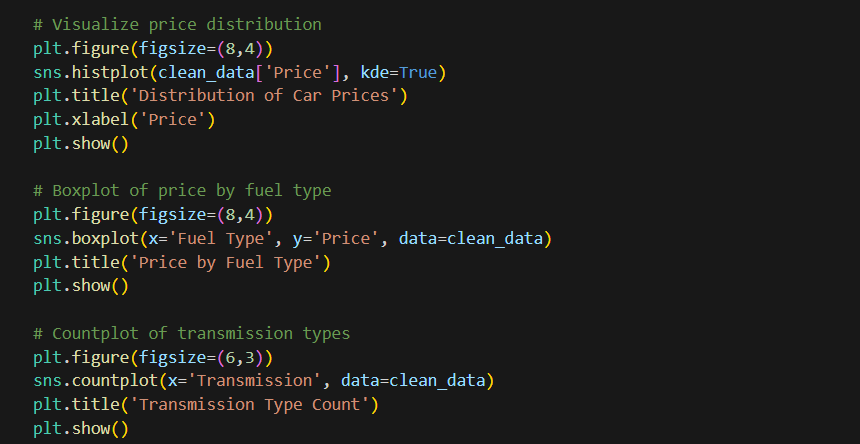
**We will check for missing values and drop or impute as needed.**



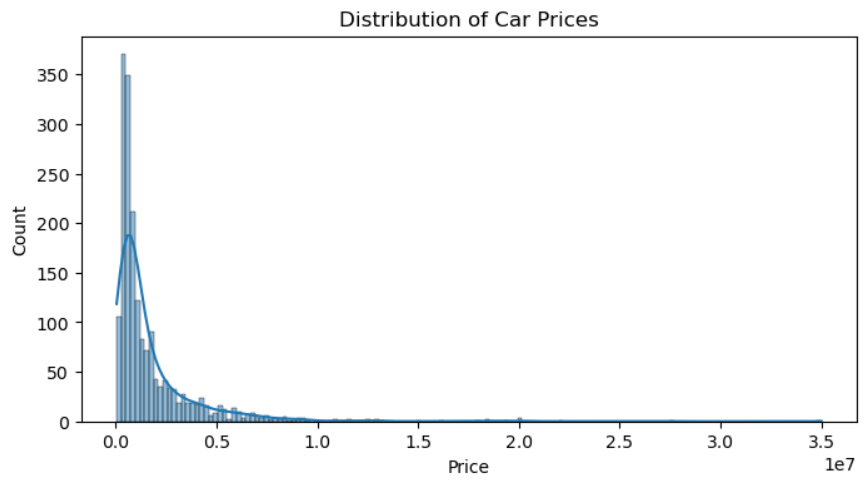
**OUTPUT:**

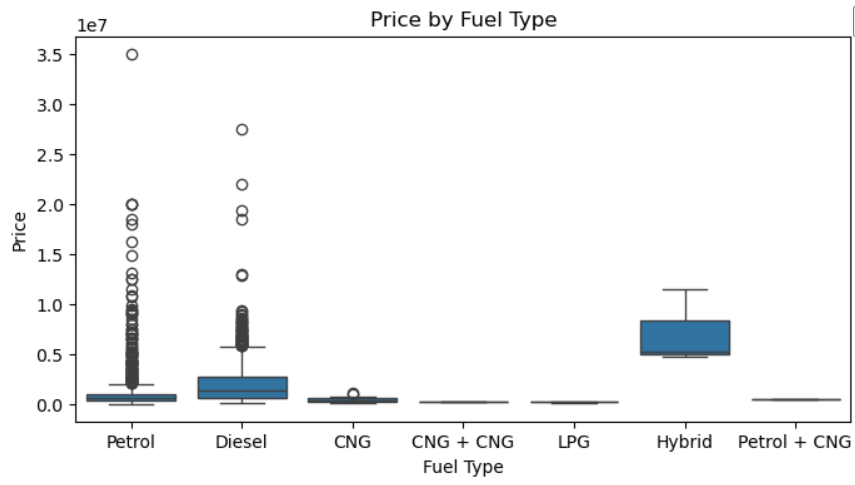


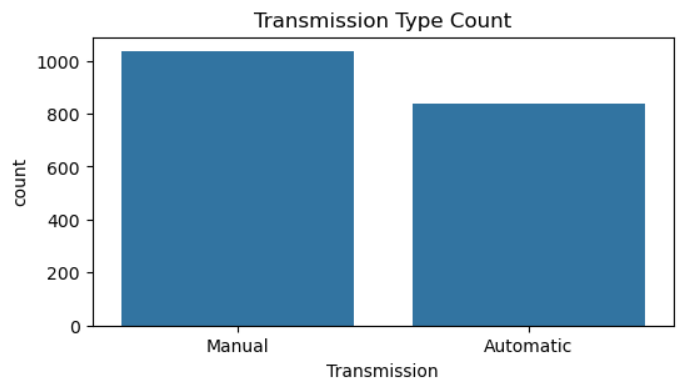
**3 : Perform exploratory data analysis using visualization tools:**



**OUTPUT:**

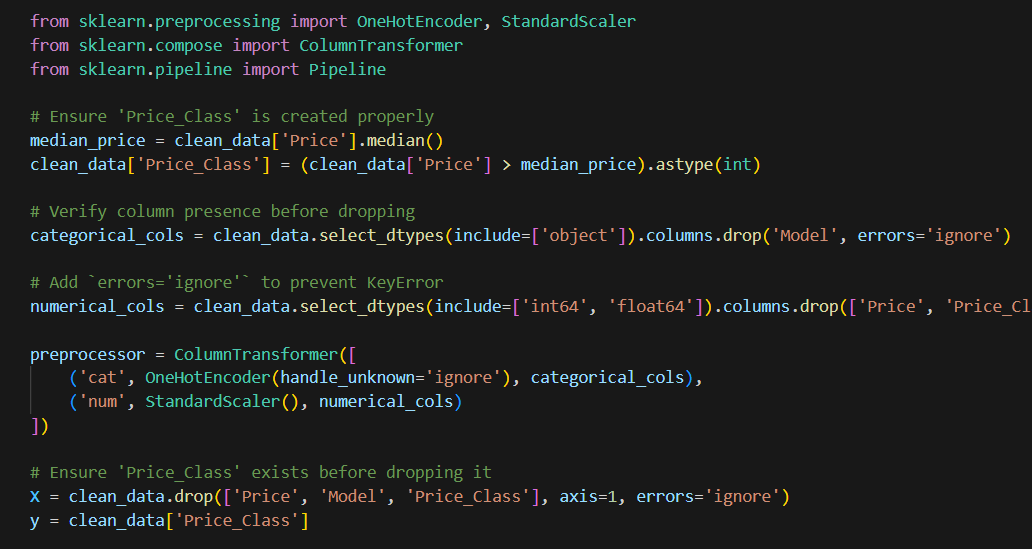




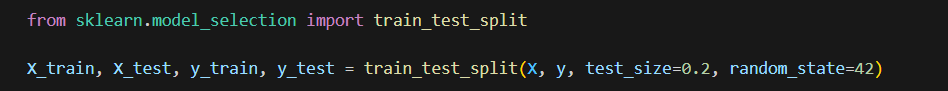


**4 : Encode categorical variables and scale numerical features:**

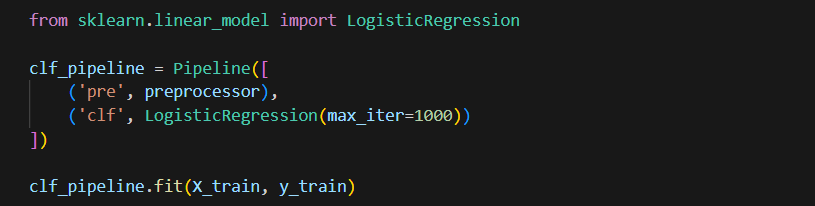
We will use one-hot encoding for categorical variables and StandardScaler for numeric features.



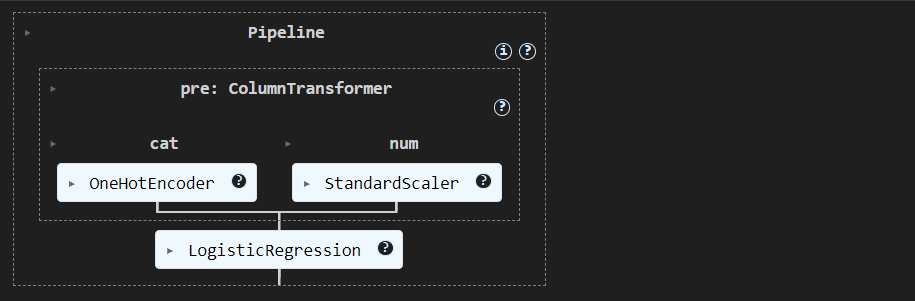
**5 : Split the dataset into training and testing sets:**



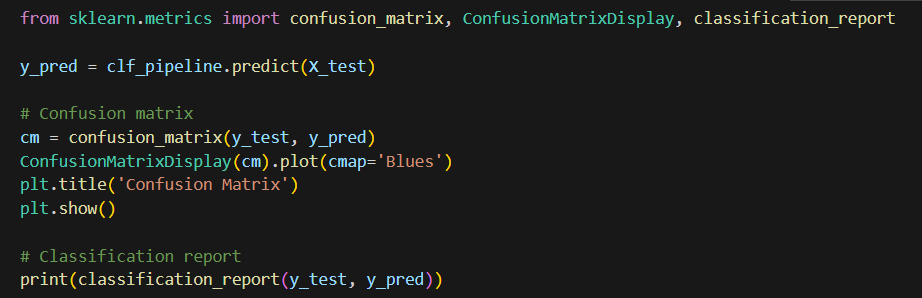
**6 : Train classification models like logistic regression:**



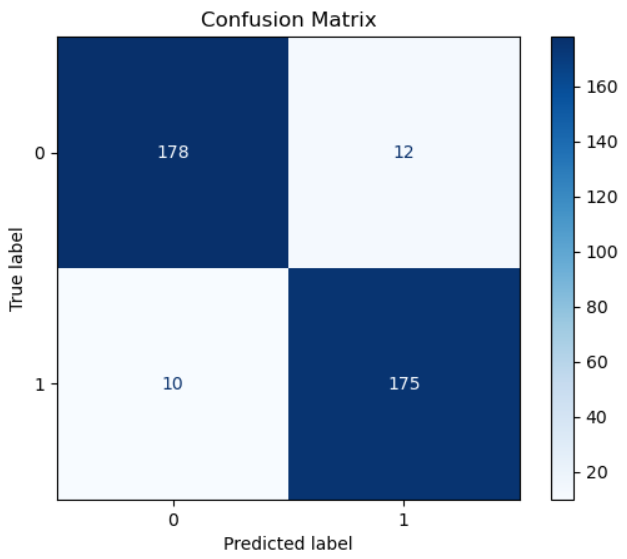
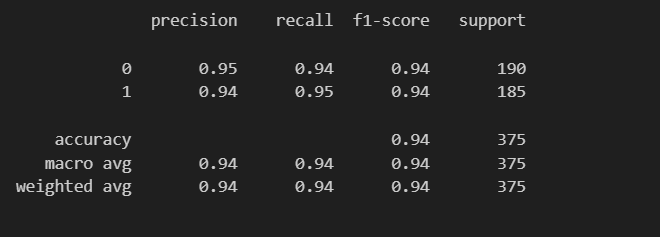
**OUTPUT:**



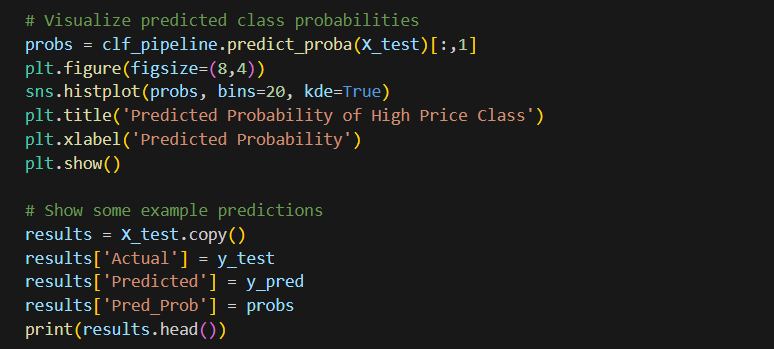
**7 : Evaluate model performance using metrics like confusion matrix and classification report:**



**OUTPUT:**

**8 : Interpret results and visualize outcomes:**



**OUTPUT**:

