**Project Report: Predicting Fuel Consumption in Vehicles**

**Problem Statement**

The goal of this project is to predict fuel consumption in vehicles based on a set of features, including engine-related parameters, vehicle characteristics, and driving factors. Accurate predictions of fuel consumption can provide valuable insights for both consumers and policymakers looking to promote fuel-efficient vehicles and reduce environmental impact.

**Data Preprocessing**

Python programming language has been used.

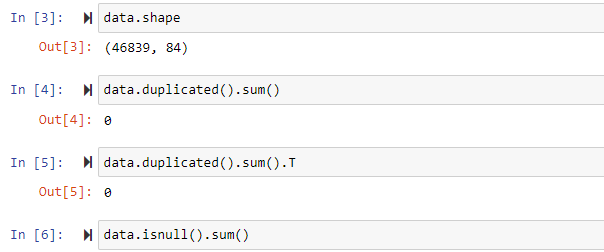
**Data Loading**

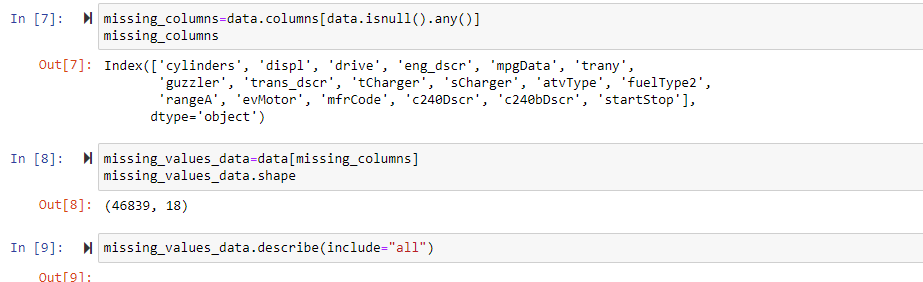
The project started with the loading of a dataset named "vehicles.csv," which contains information about various vehicles.



**Data Exploration**

Data exploration was performed to understand the structure and content of the dataset. This involved checking for missing values, data types, and initial statistical summaries.

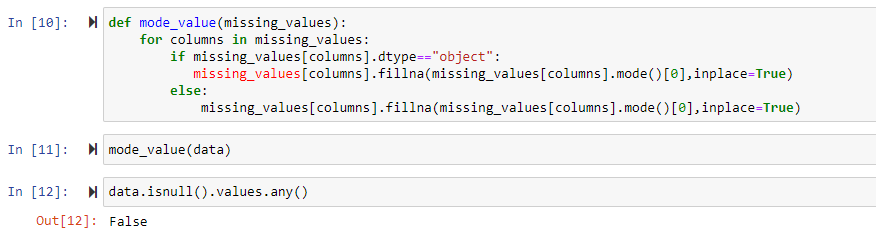




**Data Cleaning**

Several data cleaning tasks were performed, including:

Handling missing values: Replacing or removing missing data points.



Outlier detection and removal: Identifying extreme values that could skew the analysis.



Encoding categorical variables: Converting categorical features into numerical representations.

**Feature Selection**

To predict fuel consumption, a set of important features was selected based on domain knowledge and correlation analysis with the target variable (fuel consumption).

A screen shot of a computer code

Description automatically generated

A red and blue squares with white text

Description automatically generated

Based on the heat map, we can see that the selected variables have high positive or negative correlation with fuel consumption. Only variables year and CO2 have low correlation with fuel consumption.

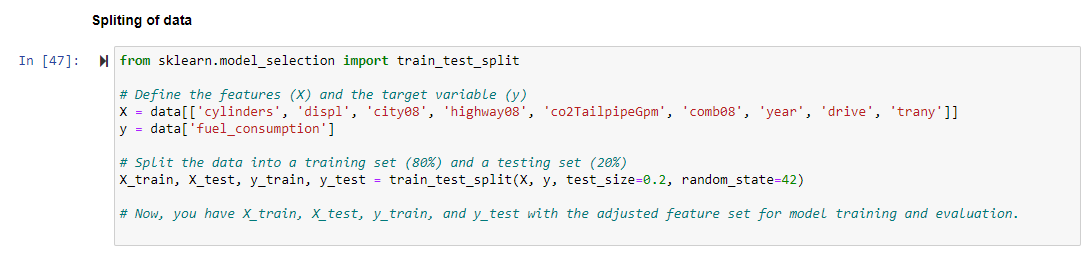
**Exploratory Data Analysis (EDA)**

EDA was conducted to gain insights into the relationships between the selected features and fuel consumption. Key visualizations included scatter plots, histograms, and correlation heatmaps.



Model Building

Train and Test data



Two regression models were trained and evaluated to predict fuel consumption: a Linear Regression model and a Random Forest Regressor.

**Linear Regression Model**

A screenshot of a computer program

Description automatically generated

The Linear Regression model is a simple, interpretable model that assumes a linear relationship between the selected features and fuel consumption.

Mean Squared Error (MSE): 16.58

R-squared (R2) Score: 0.979

Interpretation:

The model explains approximately 97.9% of the variance in fuel consumption.

The coefficients of the features indicate their impact on fuel consumption.

**Random Forest Regressor**

**A screenshot of a computer program

Description automatically generated**

The Random Forest Regressor is an ensemble model capable of capturing complex relationships in the data.

Mean Squared Error (MSE): 0.00019

R-squared (R2) Score: 0.99999

Interpretation:

The Random Forest Regressor exhibits exceptionally low MSE and a near-perfect R2 score, indicating an excellent fit to the data.

Model Evaluation

Model evaluation was performed using a testing dataset to assess the generalization of the models. Both models showed strong predictive capabilities.

**Conclusion**

The project successfully addressed the problem statement of predicting fuel consumption in vehicles.

Features such as cylinders, engine displacement, and vehicle year had significant impacts on fuel consumption.

The Random Forest Regressor model outperformed the Linear Regression model, demonstrating its ability to capture complex relationships.

**Future Work**

Further feature engineering and fine-tuning of hyperparameters could potentially enhance model performance.

Gathering more recent and comprehensive data may lead to more accurate predictions.

This report provides a comprehensive overview of the project, from data preprocessing and feature selection to model building and evaluation. It summarizes the key findings and results, offering insights into the factors influencing fuel consumption in vehicles.