Multiple Linear Regression

**Step 1: Load and Explore the Dataset**

**EDA**:

Inspect the structure and types of variables.

Generate summary statistics for each variable.

Visualize relationships : Scatter plots (e.g., Price vs. Age, Price vs. KM).

Box plots for categorical variables (e.g., FuelType, Automatic).

Correlation matrix to identify multicollinearity.

Check for missing values and handle them (e.g., imputation or removal).

**Preprocessing**:

Encode categorical variables (FuelType).

Normalize or scale numerical variables if necessary.

Handle outliers (e.g., by capping or transformation).

**Step 2: Split the Dataset**

Split the data into training (80%) and testing (20%) sets using a random state for reproducibility.

**Step 3: Build and Interpret Multiple Linear Regression Models**

**Model 1**: Include all variables as predictors.

**Model 2**: Use stepwise regression or select variables based on feature importance (using correlation, p-values, etc.).

**Model 3**: Use polynomial features or interaction terms to capture non-linear effects.

Interpret coefficients:

Assess the direction and magnitude of influence for each predictor.

**Step 4: Evaluate Model Performance**

Evaluation metrics:

Coefficient of determination.

Adjusts for the number of predictors.

Mean Squared Error (MSE) or Root Mean Squared Error (RMSE).

Mean Absolute Error (MAE).

Residual analysis:

Plot residuals vs. predicted values to check assumptions.

**Step 5: Apply Lasso and Ridge Regression**

Perform Lasso and Ridge regression using the training data.

Use cross-validation to tune hyperparameters (α\alphaα).

Compare coefficients and performance metrics with the linear regression model.

Visualize how coefficients shrink in Lasso and Ridge models.