Assignment 3

Diamonds Price Prediction

Prepare a clean, well-documented Kaggle notebook that demonstrates full regression modeling using **Linear Regression, Ridge, Lasso, and ElasticNet techniques** to predict the price of diamonds. Your notebook should include data preprocessing, visualization, modeling, evaluation, and interpretation of results using regularization techniques.

Dataset:

Download the dataset from this Link: Dataset Link

Requirements:

1. Dataset Description and Problem Definition

- Provide a brief description of the dataset.
- Define the problem.

2. Exploratory Data Analysis (EDA)

- Descriptive statistics for numerical features.
- Visualizations to explore distributions and relationships.
- Identify issues like skewness or outliers.

3. Data Cleaning

- Check for missing values or invalid entries.
- Handle outliers if necessary.
- Remove duplicates if found.

4. Preprocessing

- Perform all necessary preprocessing steps required to prepare the dataset for regression modeling.

5. Modeling and Evaluation

- Train and evaluate the following models:
  - Linear Regression
  - Ridge Regression

    - Lasso Regression
    - ElasticNet
- Use Mean Squared Error (MSE), RMSE, R^2 and R^2-adjusted scores to compare models.
- Plot:
    - Actual vs. predicted prices.
    - Coefficient magnitudes across models.
    - Lasso/ElasticNet zeroed features (if applicable).

6. Interpretation and Conclusion

- Summarize the results.
- Justify **which model performed best and why** (based on metrics and visual analysis).

Instructions:

- All codes should be written **independently** without using AI tools.
- Submit your notebook by **uploading it to Kaggle** and sharing the public notebook link.
- **Create a GitHub repository named IEEE-ML-2025**. **Inside it, add a folder named Wine_Quality_Regression,** which must contain:
    - Your **.ipynb** notebook.
    - A **README.md** file specific to the project.
    - A **plots/ folder** to store relevant visualizations.

Expected Deliverables:

- **Your GitHub repository link**
- **Your public Kaggle notebook link** demonstrates your full implementation.

Submission Deadline: Tuesday 15/7/2025 before 11:59 PM