---2---

Linear Regression Challenge

## **Predicting Energy Consumption**

### **Overview of the Task**

Energy consumption forecasting is vital for managing energy resources, optimizing usage, and implementing energy-efficient strategies. In this challenge, you are given an uncleaned dataset containing various features related to energy consumption across different regions. Your task is to clean the data, perform exploratory data analysis (EDA), and build a regression model to predict energy consumption based on the available features.

### **Problem Statement**

The dataset contains information about energy consumption but is uncleaned and may include missing values, inconsistencies, and outliers. The challenge is to preprocess this data and develop a regression model that accurately predicts energy consumption based on various features.

### **Objective**

Your main objectives are:

1. **Data Cleaning**: Address and correct issues such as missing values, inconsistencies, and outliers in the dataset.
2. **Feature Engineering**: Identify and create features that improve the performance of the regression model.
3. **Model Building**: Develop and train a regression model to predict energy consumption using the cleaned dataset.
4. **Evaluation**: Assess the performance of your model using appropriate metrics and validate its accuracy.
5. **Deployment**: Create an application or tool that allows for easy input of new data and provides consumption predictions.

### **Data Description**

**Data Source**: The dataset for this challenge is sourced from the UCI Machine Learning Repository's "Energy Efficiency Data Set" . The dataset includes:

* **Features**:
  + X1: Relative Compactness
  + X2: Surface Area
  + X3: Wall Area
  + X4: Roof Area
  + X5: Overall Height
  + X6: Orientation
  + X7: Glazing Area
  + X8: Glazing Area Distribution
* **Target Variable**:
  + Heating Load: The amount of energy required for heating the building (kWh/year).
  + Cooling Load: The amount of energy required for cooling the building (kWh/year).

### **Evaluation Criteria**

|  |  |
| --- | --- |
| **Data Collection, Connectivity, and Cleaning** | **30%** |
| **Logic and Results** | **40%** |
| **Presentation** | **5%** |
| **Code Quality** | **5%** |
| **Deployment / Running App** | **20%** |

Good luck with your energy consumption prediction model. We look forward to seeing your approach to tackling the complexities of this dataset and developing a reliable solution!