---3---

Linear Regression Challenge

## **Predicting Real Estate Prices**

## **Overview of the Task**

Predicting real estate prices is essential for various stakeholders in the real estate market, including buyers, sellers, and investors. In this challenge, you are provided with an unstructured dataset containing various features related to real estate properties and their sale prices. Your goal is to clean and structure the data, perform exploratory data analysis (EDA), and develop a regression model to predict property prices.

### **Problem Statement**

The dataset provided is sourced from Kaggle and contains a mixture of structured and unstructured data. The challenge is to preprocess the unstructured data, handle missing values, and address inconsistencies to build a robust regression model for predicting real estate prices.

### **Objective**

Your main objectives are:

1. **Data Cleaning and Structuring**: Handle unstructured and missing data, resolve inconsistencies, and prepare the dataset for analysis.
2. **Feature Engineering**: Extract meaningful features from unstructured data and select relevant features for the regression model.
3. **Model Building**: Develop and train a regression model to predict property prices based on the cleaned and structured dataset.
4. **Evaluation**: Assess the performance of the model using appropriate metrics and validate its accuracy on unseen data.
5. **Deployment**: Create an application or tool that allows users to input property data and receive price predictions.

### **Data Description**

**Data Source**: The dataset for this challenge is sourced from the Kaggle dataset. The dataset includes:

* **Features**:
  + X1 transaction date: The date when the transaction occurred.
  + X2 house age: The age of the house at the time of the transaction, measured in years.
  + X3 distance to the nearest MRT station: The distance from the house to the nearest MRT (Mass Rapid Transit) station, measured in meters.
  + X4 number of convenience stores: The number of convenience stores within walking distance from the house.
  + X5 latitude: The latitude coordinate of the house.
  + X6 longitude: The longitude coordinate of the house.
  + Y house price of unit area: The price of the house per unit area (target variable), measured in currency units per square meter.

### **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 real estate price prediction model. We look forward to seeing how you develop an effective solution for predicting property prices!