# Project Submission: Predicting House Prices Using Multiple Regression

## 1.Introduction

**Problem Statement:** The objective of this project is to build a regression model that can predict house prices based on various features (e.g., size, location, number of bedrooms).

**Business Relevance:** Predicting house prices is a common task in the real estate industry. Accurate predictions can help real estate agents, buyers, and sellers make informed decisions.

## 2. Dataset Description

**Source:** Simulated dataset for regression analysis.

**Features:**  
- Size of the House (in square feet)  
- Number of Bedrooms  
- Location (City/Region)  
- Year Built

**Target Variable:** House Price (dependent variable)

**Data Preprocessing:**- Data cleaned and handled for missing values.  
- Features normalized/scaled if needed.  
- Data split into training and testing sets.

## 3. Model Evaluation

**Model Used:** Multiple Linear Regression

**Why Chosen:** Multiple Linear Regression was chosen to predict house prices because it allows us to model the relationship between the target variable (house price) and multiple independent variables.

**Evaluation Metrics:**- R² Score: 0.973  
- Mean Absolute Error (MAE): $40,169.09  
- Mean Squared Error (MSE): 2,317,970,141.92  
- Root Mean Squared Error (RMSE): $48,145.30

**Interpretation of Results:**- The R² score indicates that the model explains 97.3% of the variance in house prices.  
- The MAE, MSE, and RMSE highlight the average prediction error, with an MAE of $40,169 indicating that the model's predictions are, on average, this much off from actual values.

## 4. Model Improvements (Optional)

**Potential improvements for future work include:  
- Feature Engineering:** Adding more features, such as the age of the house, neighborhood characteristics, or market trends.

**- Model Tuning:** Trying different algorithms like Random Forest or Gradient Boosting to see if they perform better.

**- Hyperparameter Tuning:** Performing hyperparameter optimization using Grid Search or Randomized Search.

## 5. Conclusion

The Multiple Linear Regression model has provided a strong starting point for predicting house prices, with a high R² score and a reasonable prediction error (MAE). The next steps involve testing more complex models and tuning the existing model for even better performance.

**6. Github Link**

https://github.com/RedDaredevils/Data-Science-Project/blob/main/House\_Price\_Prediction.ipynb