**Prediction of price and sqft\_area using Raw house data**

This project addresses the needs of **real estate investors** who aim to maximize returns by investing in properties with optimal **price per square foot (PPS)**, low taxes, and high rental potential. The model is designed to assist investors in estimating either the **price per square foot at a given location** or the **square footage they can afford** based on their budget and preferences.

The project involves a two-phase modeling approach:

1. **K-Nearest Neighbors (KNN) Classification**:
   * KNN is trained using latitude and longitude to classify properties into **7 price bins** (e.g., $50–150, $150–250, ..., $650–750 PPS).
   * The model achieved a **training accuracy of 81.38%** and a **test accuracy of 70.51%**.
   * The output bin provides a **coarse range** for PPS at any given coordinate.
2. **Regression Models for Exact PPS**:
   * A **Simple Linear Regression** and a **Multivariate Linear Regression (MVLR)** model were implemented to predict **exact PPS**.
   * MVLR used key features: bedrooms, bathrooms, garage, fireplaces, HOA, and zipcode.
   * Surprisingly, **Simple Linear Regression outperformed MVLR**, possibly due to multicollinearity or overfitting in the MVLR model.

Additional steps included:

* **Data Cleaning & Feature Engineering**: Added pps and bin columns, normalized inputs, and removed noisy records.
* **Train-Test Split**: 80% for training, 20% for testing.
* **Model Use Cases**:
  + **Budget Scenario**: Predict maximum square footage an investor can buy based on their budget and preferences.
  + **Location Scenario**: Predict house price per square foot at a specific location using known features.

Overall, the system provides **location-based pricing insights** and personalized investment guidance, enabling data-driven decisions in real estate investment.