# 🏡 House Price Prediction Practice

In this practice exercise, you will work on predicting house prices using gradient descent. You will also clean messy real-world data before building your model.

## 📋 Steps for Gradient Descent with Real Data

### 1. Import and Load Data

Load the .csv file using pandas.

### 2. Data Cleaning

Fix missing values (like NaN, n/a, or unknown).  
Convert wrong formats (e.g., '1,800' ➔ 1800).  
Change data types to numeric if needed.

### 3. Feature Selection

Choose columns you will use to predict price. Example features: bedrooms, bathrooms, size\_sqft, location\_score.

### 4. Set Up Prediction Model

Prediction function:  
f(x) = w1 \* x1 + w2 \* x2 + w3 \* x3 + w4 \* x4 + b

### 5. Initialize Parameters

Set all weights (w's) and bias (b) to 0 (or small random numbers).

### 6. Calculate Predictions

For each data point, predict:  
f(x) = w1 \* x1 + w2 \* x2 + w3 \* x3 + w4 \* x4 + b

### 7. Compute Cost Function (MSE)

Cost function:  
J(w, b) = (1/m) Σ (f(x(i)) - y(i))^2

### 8. Find Gradients

For each weight:  
(1/m) Σ (f(x(i)) - y(i)) \* xj(i)  
  
For bias:  
(1/m) Σ (f(x(i)) - y(i))

### 9. Update Parameters

Update rule:  
wj = wj - α × derivative for wj  
b = b - α × derivative for b

### 10. Repeat

Repeat steps 6–9 for many epochs until the cost gets very small.

## 📂 Dataset Information

You will be working with a dataset containing 200 rows of house data. Some cleaning is necessary.

**File:** house\_prices.csv

**Rows:** 200

**Columns:** bedrooms, bathrooms, size\_sqft, location\_score, year\_built, price

**Messy parts to clean:** - Missing values: NaN, n/a, unknown  
- Wrong formats: '1,800', '2,000' in size\_sqft  
- Non-numeric values: 'N/A', 'None' in price

## 🏆 Your Challenge Checklist

Complete the following steps to practice gradient descent on real data:

* Load the data into pandas.
* Clean the data (fix missing or wrong values).
* Select good features to predict price.
* Implement gradient descent manually.
* Calculate cost and gradients.
* Update weights and bias using learning rate α.
* Repeat and monitor if the cost is decreasing.