Problem Statement

Title: *Predicting House Prices Using K-Nearest Neighbors Regression*

Objective

To build a **K-Nearest Neighbors (KNN) regression model** that predicts the **price of a house** based on:

- Its **size (in square feet)**
- Its **number of bedrooms**

Background

In real estate, pricing a house accurately is crucial for both sellers and buyers. Traditional regression models assume a functional form, but KNN regression offers a **non-parametric** approach by estimating prices based on **similar historical houses**. This model predicts the price by averaging the prices of the **k most similar houses** in the dataset, based on a distance metric.

Dataset Description

Each data point includes:

- **X_1:** Size (sqft)
- **X_2:** Bedrooms
- **Y:** Price (in thousands of dollars)

Sample Dataset

House	Size (sqft)	Bedrooms	Price (K USD)
1	1400	3	245
2	1600	3	312
3	1700	4	279
4	1875	3	308
5	1100	2	199
6	1550	4	219
7	2350	4	405
8	2450	5	324
9	1425	3	319
10	1700	3	255

Goals

- Implement a **KNN regression model** to predict house prices
- Choose an appropriate value of **k** (number of neighbors)
- Use **Euclidean distance** to find nearest neighbors
- Predict price of a new house by **averaging prices of the k nearest neighbors**
- Evaluate model performance using:
  - **Mean Squared Error (MSE)**
  - **Root Mean Squared Error (RMSE)**
  - **R^2 Score**

Assumptions

- Houses with similar size and bedroom count tend to have similar prices
- The relationship between features and target does **not need to be linear**
- All features are **numerically scaled** (or normalized, if needed)

Example Task

> Predict the price of a new house with:
>
> - Size = 1800 sqft
> - Bedrooms = 3
>
> Using **k = 3**

Solution

🎯 Goal

Predict the price of a **new house**:

- Size = 1800 sqft
- Bedrooms = 3
- Using **k = 3 nearest neighbors**

📊 Step 1: Dataset Recap

House	Size (sqft)	Bedrooms	Price (K USD)
1	1400	3	245
2	1600	3	312
3	1700	4	279
4	1875	3	308

House	Size (sqft)	Bedrooms	Price (K USD)
5	1100	2	199
6	1550	4	219
7	2350	4	405
8	2450	5	324
9	1425	3	319
10	1700	3	255

📏 Step 2: Compute Euclidean Distance

Distance formula:

$ d = \sqrt{(x_1 - x_1')^2 + (x_2 - x_2')^2} $

Let the query house be (Q = (1800, 3))

House	(Size, Bed)	Distance to Q	Price
1	(1400, 3)	$\sqrt{(400^2 + 0^2)} = 400$	245
2	(1600, 3)	200	312
3	(1700, 4)	$\sqrt{(100^2 + 1^2)} \approx 100.005$	279
4	(1875, 3)	75	308
5	(1100, 2)	$\sqrt{(700^2 + 1^2)} \approx 700.0007$	199
6	(1550, 4)	$\sqrt{(250^2 + 1^2)} \approx 250.002$	219
7	(2350, 4)	$\sqrt{(550^2 + 1^2)} \approx 550.0009$	405
8	(2450, 5)	$\sqrt{(650^2 + 2^2)} \approx 650.0031$	324
9	(1425, 3)	375	319
10	(1700, 3)	100	255

☑ Step 3: Identify k = 3 Nearest Neighbors

Sort by distance:

House	Distance	Price
4	75	308
10	100	255
3	100.005	279

☑ Nearest 3 houses: House 4, House 10, House 3

🧮 Step 4: Predict Price (Average of k nearest neighbors)

$ \hat{y} = \frac{308 + 255 + 279}{3} = \frac{842}{3} \approx 280.67 $

📦 Final Prediction

> 🏠 A house with **1800 sqft** and **3 bedrooms** is predicted to cost:
>
> $ \boxed{$280,670} $

☑ Summary

Step	Result
k	3
Distance Metric	Euclidean
Prediction	$280.67K
Neighbors	Houses 4, 10, 3