HOUSE PRICE PREDICTION

Data 200 – Applied Statistical Analysis

PROBLEM STATEMENT

In the real estate industry, accurately estimating the price of a house is crucial for buyers, sellers, and investors.
This project aims to build a predictive model using historical housing data to estimate house prices based on features such as:
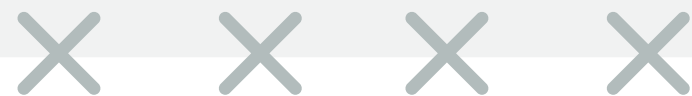
- Area (square footage)
- Number of bedrooms and bathrooms
- Location
- Year built and other variables
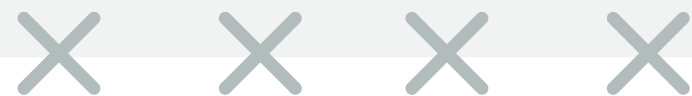
TYPE OF PROBLEM & STATISTICAL MODEL TO BE USED

Linear Regression

BRIEF DESCRIPTION OF THE CHOSEN MODEL

Linear Regression Model

Linear Regression estimates the relationship between multiple independent variables (e.g., area, rooms) and a dependent variable (house price).
It uses a best-fit line by minimizing the difference between predicted and actual values.
This helps both in:
- Making price predictions
- Understanding which features impact the price the most

FORMULA

$$\text{Price} = \beta_0 + \beta_1 \times \text{Area} + \beta_2 \times \text{Bedrooms} + \ldots + \varepsilon$$

Where:
- β_0 is the intercept — the predicted price when all features are 0.
- β_1, β_2, ... are coefficients — they tell us how much each feature (like area or bedrooms) affects the price.
- ε (epsilon) is the error — the difference between the actual and predicted prices.