Predicting housing prices : using machine learning

# Introduction

 **Objective:** Develop a machine learning model to predict housing prices based on various features.

 **Dataset:** Used the [name of dataset, e.g., Boston Housing dataset] with the following column names: ['CRIM', 'ZN', 'INDUS', 'CHAS', 'NOX', 'RM', 'AGE', 'DIS', 'RAD', 'TAX', 'PTRATIO', 'B', 'LSTAT', 'MEDV'].

# Dataset and Features

* **Description:** Provide a brief description of the dataset and the features.
  + CRIM: Per capita crime rate by town.
  + ZN: Proportion of residential land zoned for lots over 25,000 sq. ft.
  + INDUS: Proportion of non-retail business acres per town.
  + CHAS: Charles River dummy variable (1 if tract bounds river; 0 otherwise).
  + NOX: Nitric oxides concentration (parts per 10 million).
  + RM: Average number of rooms per dwelling.
  + AGE: Proportion of owner-occupied units built prior to 1940.
  + DIS: Weighted distances to five Boston employment centers.
  + RAD: Index of accessibility to radial highways.
  + TAX: Full-value property tax rate per $10,000.
  + PTRATIO: Pupil-teacher ratio by town.
  + B: 1000(Bk - 0.63)^2 where Bk is the proportion of Black residents by town.
  + LSTAT: Percentage of lower status of the population.
  + MEDV: Median value of owner-occupied homes in $1000s.

# Data Preprocessing

## Loading and Inspecting the Data

* **Steps:**
  + Load the dataset using pandas.
  + Display the first few rows of the dataset.
  + Check for missing values.

## Handling Missing Values

* **Steps:**
  + Identify missing values.
  + Choose an appropriate method to handle missing values (e.g., mean imputation).

## Feature Scaling and Normalisation

* **Steps:**
  + Apply feature scaling (e.g., StandardScaler or MinMaxScaler) to ensure all features contribute equally to the model.

## Splitting the Data

* **Steps:**
  + Split the data into training and testing sets (e.g., 80-20 split).

# Model Development

## 4.1 Model Selection

* **Model:** Choose a simple regression model (e.g., Linear Regression).

## 4.2 Training the Model

* **Steps:**
  + Train the model on the training dataset.
  + Display the model coefficients.

## 4.3 Evaluating the Model

* **Metrics:**
  + Mean Absolute Error (MAE)
  + Mean Squared Error (MSE)
* **Steps:**
  + Evaluate the model on the testing dataset.
  + Display the evaluation metrics.

# Model Evaluation and Interpretation

## Visualisation

* **Plots:**
  + Actual vs. Predicted prices.
  + Residual plot.

## Interpretation

* **Coefficients:**
  + Interpret the model coefficients to understand the impact of different features on housing prices.
* **Findings:**
  + Summarise key insights derived from the model.

# Documentation

* **Code Documentation:** Include comments in the code to explain each step.
* **Challenges:** Discuss any challenges faced during the project.

# Conclusion

* **Summary:** Summarise the model's performance, key findings, and any potential improvements for future work.

# Appendix

* **Source Code:** Provide a link or include the complete code used for data preprocessing, model training, and evaluation.
* **Additional Visualisations:** Include any additional plots or graphs that provide further insights.