

# **Project: House Price Prediction and Neighborhood Analysis**

## **Objective:**

The objective of this project is to predict house prices based on various features of houses. Additionally, we'll cluster neighborhoods using K-means or K-medoids to analyze similarities and differences, then employ linear regression models to predict house prices and perform feature selection to determine the most influential factors. Also, we will train Neural Network architectures to predict house prices.

## **Dataset:**

You can use a real estate dataset such as the Boston Housing Dataset or the Ames Housing Dataset. The Ames dataset is more comprehensive and provides a wider range of features to work with.

## **Steps:**

### **1. Data Exploration and Preprocessing:**

- Load the dataset and explore its structure and features.
- Handle missing values, outliers, and perform necessary data transformations.

### **2. Neighborhood Clustering:**

- Use K-means or K-medoids clustering to group similar neighborhoods based on features like crime rate, schools, accessibility, etc. You can use hierarchical clustering to figure out cluster size.
- Visualize the clusters on a map (if geographical data is available).

### **3. Feature Selection:**

- Employ techniques like correlation analysis, forward/backward selection, or LASSO regression to select the most relevant features for house price prediction.
- Explain why certain features were chosen or discarded.

### **4. House Price Prediction:**

- Split the data into training and testing sets.
- Implement a simple linear regression model to predict house prices based on a single feature (e.g., square footage).
- Evaluate the model using metrics like Mean Squared Error (MSE) or R-squared.

## 5. Multiple Linear Regression:

- Build a multiple linear regression model using selected features to predict house prices.
- Compare the performance with the simple linear regression model.
- Discuss the impact of adding more features on the model's accuracy.

## 6. Neural Network:

- Implement a basic neural network (e.g., using TensorFlow or PyTorch) for house price prediction.
- Experiment with different architectures, activation functions, and regularization techniques.
- Compare neural network results with linear regression models.

## Deliverables:

- Jupyter Notebook or a similar document detailing:
  - Data preprocessing steps.
  - Neighborhood clustering results and visualizations.
  - Feature selection process and chosen features.
  - Simple and multiple linear regression model implementation and evaluation.
  - Neural network implementation and comparison.
- Visualizations:
  - Clusters of neighborhoods on a map (if applicable).
  - Scatter plots showing house price vs. selected features.
  - Regression plots demonstrating model predictions.

## Conclusion:

- Discuss the insights gained from neighborhood clustering.
- Analyze the impact of different features on house prices.
- Compare the performance of linear regression models and neural networks.
- Reflect on the importance of feature selection in improving model accuracy.