Problem Statement: Predicting Housing Prices Using Regression Models

The task in this project is to investigate how machine learning regression models can be used to predict median housing values in California districts. The California Housing dataset contains socio-economic and demographic features such as average income, average number of rooms, population, and geographical information. The target variable is the median house value, expressed in units of $100,000.

The process involves: 1. Loading and exploring the California Housing dataset to understand its structure and distribution. 2. Splitting the dataset into training and testing subsets to evaluate model performance on unseen data. 3. Applying preprocessing techniques such as feature scaling using StandardScaler to normalize the input features. 4. Training regression models (e.g., Linear Regression and potentially others) to fit the training data. 5. Evaluating the models on the test set using appropriate performance metrics such as the Root Mean Squared Error (RMSE).

The key objective is to build a regression pipeline that can predict housing prices with high accuracy while maintaining generalization to unseen data. By analyzing and comparing the results, the project aims to highlight the importance of preprocessing, model selection, and evaluation strategies in real-world predictive modeling tasks.