

# # California Housing Price Prediction

*This project predicts the median housing prices in California using \*\*Machine Learning\*\*. It includes \*\*data preprocessing\*\*, \*\*visualization\*\*, \*\*model training\*\*, and \*\*evaluation\*\* using multiple regression algorithms.*

---

## ## Dataset

*The dataset used is the \*\*California Housing Dataset\*\* available in `sklearn.datasets`.*

*It contains information about:*

- *Median income*
- *Housing age*
- *Average rooms per household*
- *Average bedrooms per household*
- *Population*
- *Households*
- *Latitude and longitude*
- *Median house value (target)*

---

## ## Features

- *Data loading and exploration*
- *Data cleaning and preprocessing*

- Feature scaling
- Data visualization using Matplotlib & Seaborn
- Model training using:
  - Linear Regression
  - Decision Tree Regressor
  - Random Forest Regressor
  - Gradient Boosting Regressor
- Model evaluation ( $R^2$  Score, RMSE)

---

## ## Tech Stack

- \*\*Language:\*\* Python
- \*\*Libraries:\*\*
  - NumPy
  - Pandas
  - Scikit-learn
  - Matplotlib
  - Seaborn
  - Jupyter Notebook

---

## ## Model Evaluation

Each model is evaluated using:

- \*\*Mean Squared Error (MSE)\*\*
- \*\*Root Mean Squared Error (RMSE)\*\*

- **$R^2$  Score**

---

## ## 🧠 Results

*After testing multiple models, the **Random Forest Regressor** achieved the best accuracy with the lowest RMSE, making it suitable for predicting California housing prices.*

---

## ## 📁 File Structure