**Project Description**

**California Housing Price Analysis and Prediction**

This project involves analyzing and predicting housing prices in California using a dataset that includes various features such as location, age of the house, number of rooms, population, household, median income, and proximity to the ocean. The objective is to build predictive models to estimate the median house value based on these features.

**Walkthrough**

1. **Data Exploration and Preprocessing:**
   * Loaded the dataset and examined its structure and basic statistics.
   * Handled missing values by filling them with the median value of the respective column.
   * Created new features to better capture the information, such as the number of rooms per household and population per household.
   * Categorical feature ocean\_proximity was encoded using one-hot encoding.
2. **Feature Engineering:**
   * Added interaction terms to capture the effect of combined features.
   * Standardized the features to have zero mean and unit variance for better performance of some machine learning models.
3. **Modeling:**
   * **Linear Regression:** A simple linear approach to understand the basic relationships between the features and the target variable.
   * **Decision Tree Regressor:** A non-linear model that captures more complex relationships by partitioning the data into subsets based on feature values.
   * **Random Forest Regressor:** An ensemble method that improves the performance of decision trees by averaging the results of multiple trees to reduce overfitting and improve accuracy.
4. **Model Evaluation:**
   * Used metrics such as Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) to evaluate the performance of the models.
   * Cross-validation was employed to ensure the models generalize well to unseen data.
5. **Model Tuning:**
   * Hyperparameter tuning was performed using grid search to find the optimal parameters for each model, enhancing their performance.

**Models Used**

1. **Linear Regression:**
   * A basic yet powerful model that assumes a linear relationship between the input features and the target variable.
2. **Decision Tree Regressor:**
   * A tree-based model that splits the data into subsets based on the value of input features. It captures non-linear relationships but can be prone to overfitting.
3. **Random Forest Regressor:**
   * An ensemble method that builds multiple decision trees and merges their results to improve accuracy and control overfitting.