**California Housing Price Prediction Project Report**

**Project Overview:**

The project aims to build a model for predicting median house values in California using a dataset obtained from the US Census Bureau. The dataset includes various metrics such as longitude, latitude, housing median age, total rooms, total bedrooms, population, households, median income, ocean proximity, and median house value for each block group in California.

**Steps Performed:**

1. Load the Data:

- The dataset was loaded from the "housing.csv" file.

- The first few rows of the data were printed.

- Input (X) and output (Y) data were extracted from the dataset.

2. Handle Missing Values:

- Missing values in the dataset were filled with the mean of the respective column.

Encode Categorical Data:

- Categorical columns, particularly 'ocean\_proximity', were one-hot encoded to convert them into numerical data.

4. Split the Dataset:

- The dataset was split into 80% training dataset and 20% test dataset.

5. Standardize Data:

- The training and test datasets were standardized using the StandardScaler from scikit-learn.

6. Perform Linear Regression:

- Linear Regression was performed on the training data.

- Predictions were made for the test dataset using the fitted model.

- Root Mean Squared Error (RMSE) from Linear Regression was calculated.

7. Perform Decision Tree Regression:

- Decision Tree Regression was performed on the training data.

- Predictions were made for the test dataset using the fitted model.

- RMSE from Decision Tree Regression was calculated.

8. Perform Random Forest Regression:

- Random Forest Regression was performed on the training data.

- Predictions were made for the test dataset using the fitted model.

- RMSE from Random Forest Regression was calculated.

9. Bonus Exercise - Linear Regression with One Independent Variable:

- The median\_income column was extracted from the independent variables for training and test data.

- Linear Regression with one independent variable (median\_income) was performed on the training data.

- Predictions were made for the test dataset using the fitted model.

- The fitted model was plotted for both training and test datasets.

**Results:**

- RMSE values were obtained for Linear Regression, Decision Tree Regression, and Random Forest Regression.

- The bonus exercise provided insights into the relationship between median income and median house value.