**Project Overview:**

The California Housing Price Prediction project aims to build a model that predicts the median house value in California using the provided dataset. The dataset includes 10 metrics for each block group in California, such as the population, median income, and median housing price. The model should be able to learn from the data and predict the median housing price in any district, given all the other metrics.

**Analysis Tasks:**

1. **Load the data:**

The first step was to read the "housing.csv" file into the program and print the first few rows of the data. Next, we extracted input (X) and output (Y) data from the dataset.

1. **Handle missing values:**

We filled the missing values in the dataset with the mean of the respective column to avoid errors in the analysis.

1. **Encode categorical data:**

To enable analysis of categorical data, we converted the categorical column in the dataset to numerical data.

1. **Split the dataset:**

The data was split into an 80% training dataset and a 20% test dataset.

1. **Standardize data:**

Both the training and test datasets were standardized.

1. **Perform Linear Regression:**

We performed Linear Regression on the training data, predicted the output for the test dataset using the fitted model, and printed the root mean squared error (RMSE) from Linear Regression using the mean\_squared\_error function from sklearn.metrics.

1. **Perform Decision Tree Regression:**

Next, we performed Decision Tree Regression on the training data, predicted the output for the test dataset using the fitted model, and printed the RMSE from Decision Tree Regression.

1. **Perform Random Forest Regression:**

We then performed Random Forest Regression on the training data, predicted the output for the test dataset using the fitted model, and printed the RMSE from Random Forest Regression.

1. **Bonus exercise: Perform Linear Regression with one independent variable:**

We extracted just the median\_income column from the independent variables (from X\_train and X\_test), performed Linear Regression to predict housing values based on median\_income, predicted the output for the test dataset using the fitted model, and plotted the fitted model for training data as well as for test data to check if the fitted model satisfies the test data.

**Challenges:**

There were no major challenges encountered during this project.

**Insights:**

The Random Forest Regression model had the lowest RMSE of 49,269.16 compared to Linear Regression and Decision Tree Regression, which had RMSE values of 71,821.63 and 69,414.67, respectively. This suggests that the Random Forest Regression model performed better in predicting the median housing price in California.

**Conclusion:**

The California Housing Price Prediction project aimed to build a model that predicts the median house value in California using the provided dataset. We successfully built and tested multiple regression models on the dataset, and found that the Random Forest Regression model performed best in predicting median housing prices. This model could be useful for real estate professionals, investors, and policymakers in making informed decisions about the housing market in California.