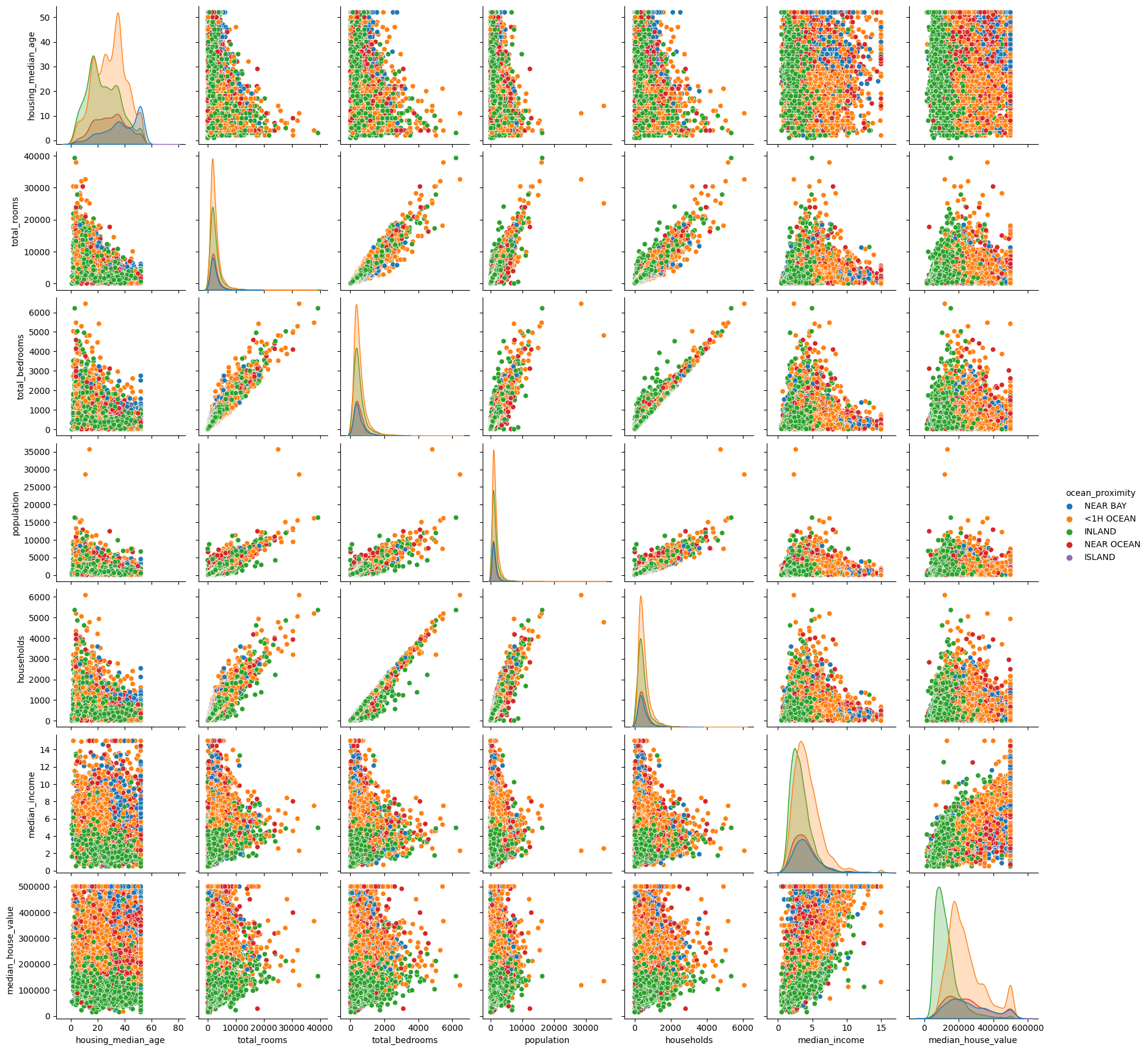
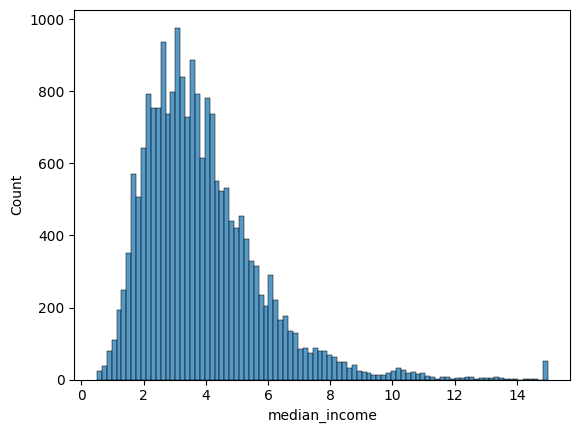
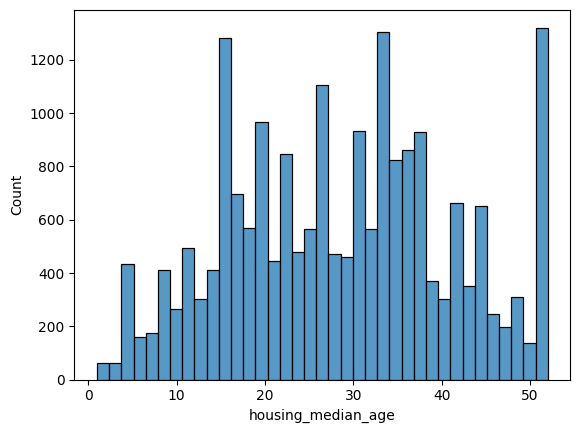
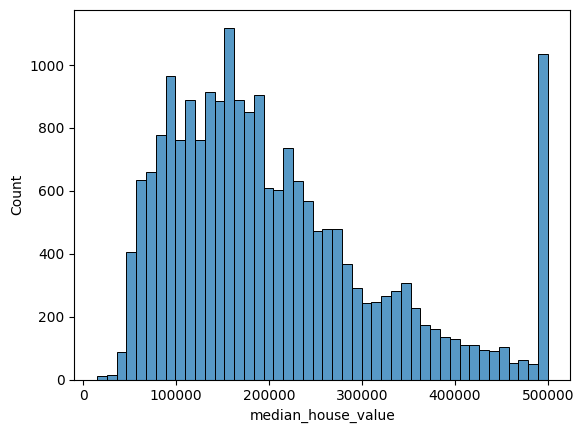
## Dataset

California Housing Prices data is selected which contains the median house prices for California districts that have been derived from the 1990 census. The dataset contains information about 20640 houses from the California district each having these features longitude, latitude, housing\_median\_age, total\_rooms, total\_bedrooms, population, households, median\_income, median\_house\_value and ocean\_proximity. I have selected housing\_median\_age and median\_income as the predictors to determine the median\_house\_value as the response variable.



Housing median age looks normally distributed, however, there are only a few outliers in median income, hence these two features are most promising in determining the median\_house\_value.





The data is scaled to normal before fitting the model. The linear regression results are:

Linear Regression R2 score: 0.508740578668139

Linear Regression RMSE score: 80299

The SGD Regression algorithm parameters are tuned before training. It performed similarly the same as the linear regression, its results are:

SGD Regression R2 score: 0.5086611493522423

SGD Regression RMSE score: 80306

The coefficient of determination (R2) is slightly better for linear regression and also its Root Mean Squared Error (RMSE) is slightly less than the SGD model.