**Step 1: Problem Definition:** The problem at hand would be defined as attempting to predict the price of a home in California based upon several parameters which include Longitude, Latitude, Total Rooms, Total Bedrooms, Population, Households and Median Income. To help with the problem, I used Regression given the continuous nature of the data. I have also used Random Forest to attempt to predict the value as well.

**Step 2: Data Splitting:** To help with addressing the problem and analyzing the data frame as whole, the data was split into two separate data frames between df\_features and df\_response. Df\_features is the data frame with the features that we want to use in our prediction such as Total Rooms and is used at the X value. Df\_response is the column of housing values that we will use as our Y axis and is the value that we are trying to predict.

**Step 3: Model Selection, Model Fitting, and Model Evaluation:** As stated previously, Regression as well as Random Forest were used as models to predict the housing value. The data was fitted using the training data and then testing with test data to get see how well each model worked. The Regression model has an R squared of .6387253266157136 which is not very accurate overall and tells us that this model is not doing a very good job at prediction. We are reconfirmed of this with our standard error of 51549.4225 for the price which is a very large amount of error. The Random Forest model returned an accuracy rating of 0.8146263790906706 which is much better than the regression model, however, could use some improvement if we were actually trying to rely on this model for prediction. Out of the two, the Random Forest certainly had the best performance.

**Step 4: Finals Remarks:** Overall there are some very interesting analytics to be taken away from the models. One of the biggest surprises we can see from the heat map where Median Income was actually correlated with Median House Value which could be something worth investigating more. Our scatter plot also gives us a very good visual of how the closer to the ocean our houses appear to be, the higher the value seems to rise which of course makes sense too. In general, our regression model did not give us very convincing results with a R squared value near 64% which would indicate that the model should not be very trusted overall. The random forest giving us an accuracy of 81.46% means that it definitely beats the regression model, however I would suggest that perhaps breaking the data set up further and maybe attempting to do some deeper analysis would lead to more accurate results.