Lee Ferris

For the final project, I split my code up based on the main function and then the rest of the functions. The first function I have reads over the csv file, organizing it based on how long each line is and if it is missing any information that is necessary. The next function normalized the data to have a mean of 0 and a standard deviation of 1. The linear regression function calculates the slope and the intercept of the data set. It does this by finding the gradients of the slope and intercept to determine how they minimize the error. The coefficient of determination calculates R^2 by subtracting the quotient of the residual sum of squares over the total sum of squares from 1, resulting in R^2. The mean squared error and mean absolute error are relatively similar, but one squares the difference between the actual and predicted while one takes the absolute value.

With the results of running the code, we can see that the relationship is not linear because the coefficient of determination is practically 0. While you would think that as the number of bedrooms a house has increases, the value of the house should too. However, this data set includes samples from every state. So, smaller houses in more expensive states may be more expensive than larger houses in cheaper states, which explains why R^2 may be low. Also, the MAE and MSE are both very large, showing that the linear regression performed poorly. This also shows that the linear regression is not a good fit for the data set.