**ANALYSIS:**

* + - The boston dataset is imported from the sklearn.datasets and a brief overview of the attributes, features and objective of the data are done.
    - The dataset has its own inbuilt functions and data members.
    - A scatterplot of the target feature, which is the prices, is plotted using the matplotlob.pyplot module.
    - Then, in order to use the seaborn library, the whole of the Boston dataset is converted to a dataframe, including the target feature.
    - An lmpot is done with the number of rooms per dwelling as x axis and the cost as y axis. the lmplot automatically fits a linear regression model for the given scatterplot.
    - The best fit line equation for the given scatter plot is found out by finding the parameters of the line equation for which the sum of the squares of the distance between the y coordinate of the line and the actual y value for the same x value is least. This method is called the sum of least squares.
    - Seaborn library automatically does this. But we can explicitly find out the best fit using the provisions of the numpy library:
      * The series is first converted to a 2 dimensional array using the ‘vstack()’ in numpy.
      * For every row in the array X, a new column with value 1 is created.
      * Now , using the ‘np.linalg.lstsq()’ we find the values of m,b by supplying the matrix X and Y(Y is the array containing all the prices, the target feature).
      * now we can plot the best fit line whose equation is y=m\*x + b
      * we plot this on top of the scatter plot using the plot function in plt.
      * The total error is found out as the value in index 1 of the list object returned by the ‘np.linalg.lstsq()’ method.
      * Through this we can find the root mean square error
    - The sklearn module is imported, LinearRegression is imported from sklearn.linear\_model
    - A LinearRegression object is created.
    - The training data consists of the training values (the X dataframe excluding the target feature) and the target data consists of the target values( the Y data frame)
    - Using the ‘.fit()’ method of the LinearRegression object with X,Y mentioned before as the parameters, the linear fit is modelled.
    - The coefficients of the fitted line can be obtained using the ‘.coef\_’ attribute of the LinearRegression object
    - Similarly the intercept coefficient can be obtained using the ‘.intercept\_’ attribute of the LinearRegression object.
    - Predicting prices:
      * In order to predict the prices, we need to split our available dataset into a training set and a testing set. This is done using the ‘sklearn.model\_selection.train\_test\_split()’ function that takes in the X(data used for prediction) and the Y(the target variable data), also optionally the fraction of the given data to be split into training/testing set. This function gives 4 different indexable objects as output that are stored as X\_train,X\_test,Y\_train and Y\_test respectively.
      * X\_train andY\_train are used in the ‘.fit()’ method to model the fitted line
      * ‘.predict()’ method predicts the target set by returning a predicted set and takes in the dataset as input.
      * The error is the difference between this predicted set and the target set. Summing up squares of all these errors gives us the root mean square error. The coefficient of determination is also measure of the accuracy of the model. It basically is the square of the correlation coefficient between the predicted and the target random variables.
      * Once that is done, a residual plot is plotted to check for the residual error pattern. A residual data point is the difference between the predicted data point to the corresponding target data point.
      * The residual plot has the residuals in the y axis and the predicted points as the x values.
      * A proper model should have the scatter plot centred on y=0 with no detectable patterns. If the scatter plot is completely stochastic, it implies that the errors are due to complete randomness of the real world and there are no deterministic variables that have been missed while making the predictive model.