I initially checked the whole dataset as it was using glimpse and I transformed the variable bath as a

factor. I created a histogram of the response variable and a linear regression model using all the

predictors. The histogram was strongly skewed, and it was not possible to notice any correlation

between price and all the other variables probably due to an outlier.

I removed the outlier using the methods above and I checked the model again and this time the

histogram and all the other plots were clearer.

I decided to create a model with a log transformation of the response and even though

there were still low value of p-value and R-squared = 0.8641, from the diagnostics it is possible to

notice that the error do not have a constant variance and the histogram is left skewed. So, I decided touse data without transformations. I then created single plots between price and all the other variables

in order to have a better view of trends.



The model I chosen (Model 1) has the highest

value of Adj. R-squared and the lowest p-value.

In the prediction model analysis, the correlation

between the observed model and the fit model

is 0.94017.17