When it comes to simple regression, we would like to find which numerical variable has the most influence on Price, and do a simple linear regression model.

A general view of correlation and coefficient between each two numerical variable are shown to do the comparison. In this graph, the deeper the color is, the stronger the correlation is. It can be found that our dependent variable Price is at the left bottom corner. The variable with the deepest color is living area. Now, the scatter plot is showed and suggests that there might be some linear relationship exist.

Thus we can say that the variable ‘Living.Area’ may be most affected by ‘Price’. Then, our dependent variable is ‘Price’ and independent variable is ‘Living area’.

Furthermore, there are four assumptions underling our linear regression model.

The first one is the relationship between Y and x is linear, and is called linearity. The blue line is reasonably well plotted like a straight line (with no obvious curve), so there is no obvious pattern.

Also, it appears the residuals are getting spread-out and the errors have constant variance.

and follow a normal distribution is approximately satisfied, since, in the QQ plot, the points are reasonably close to the diagonal line. Although there seems to be some outliers exist, the data-set is relatively large.

Thus,as all assumption met, we can concluded that a one meter square living area increase in temperature, results in a One hundred and thirteen dollars Price increase on average