**Data Science Project**

Linlong Wu

1. Data Handling

Before setting up the model, I trimmed the data set to get rid of the outliers. To investigate the distribution of variables, I plotted the pair-plots. I observed zero values for x, y, z which isn’t logical. After filtering out the zero values, I set bands for each numerical variable to remove the outliers.

1. Ordinal Encoding for Categorical Variables

For each categorical variable, I applied ordinal encoding to assign numerical values.

1. Modelling

For prediction, I used random forest model on the train set, and applied the model to predict prices based on test set variables.