

# Amazon Machine Learning Homework

## Data Pre-processing

Due to the size of the data, I was unable to process the whole data set so decided to select the first year of the dataset, 1995. I then subset the data to only show columns 3 5 7 12 (if the index was 1 based).

## Summary

For the models I decided to use both a multiple linear regression on the property type, lease duration and the town or city in which the property was located. As well I used a random forest ensemble regressor to compare models. I used 2 values to compare the models which were the  $R^2$   $R$ -squared evaluates the scatter of the data points around the fitted regression line and I also calculated the mean difference of the observation to the predicted value.