Airbnb price prediction with machine learning and deep learning

Interim Report

Björgvin Atli Júlíusson, Jón Eðvarð Viðarsson,
Tómas Kristinn L Ingólfsson

November 2022

1 Interim Report

Our plan is to begin small by implementing a simple model (linear regression), and only a subset of the features available on a single city's data. First, we attempted to guess which features would be most useful in the prediction. Before looking at the data, we assumed that the size of the property being listed and the number of bedrooms would be the most important features. However, it came to light that the data does not contain information regarding the size of the property, and upon investigation of the features available, we selected four features to start training our model with: accommodates (the number of people that can be accommodated), bathrooms (the number of bathrooms), bedrooms (the number of bedrooms), and beds (the number of beds). The next step was to select a city to train our model on. We selected Copenhagen as a starting point as it is a moderately sized city that we are familiar with, and we will evaluate the Copenhagen model on the data from Stockholm - as it is, in most ways, very similar to Copenhagen.

In the beginning of our project, all we are trying to do is familiarize ourselves as much as possible with the data, i.e. learn which features are most/ least important. That is our primary goal with the linear regression model, as opposed to obtaining stellar results. From there, we can begin to implement more complicated models. The models we have in mind at the moment are SVM regression, followed by a neural network, but these are just our initial ideas and could of course change. The goal is to produce a model that is not only accurate in predicting prices in Copenhagen and Stockholm, but all over the world.