DataDesc

Patrick Tawil

March 6, 2018

I'll be making a summary about the variables at hand.

Intuitively, one of the main aspect of a house that would increase its cost is the amount of square feet. Sadly with the data given, out 20,545 observations we only had 496 recorded values for square feet. This made it a very poor predictor to use.

Table 1: Summary of the Square Feet column

Mininum	1st Quartile	Median	Mean	3rd Quartile	Maximum	Number of NAs
0	97	700	673.5	1023	2799	20049

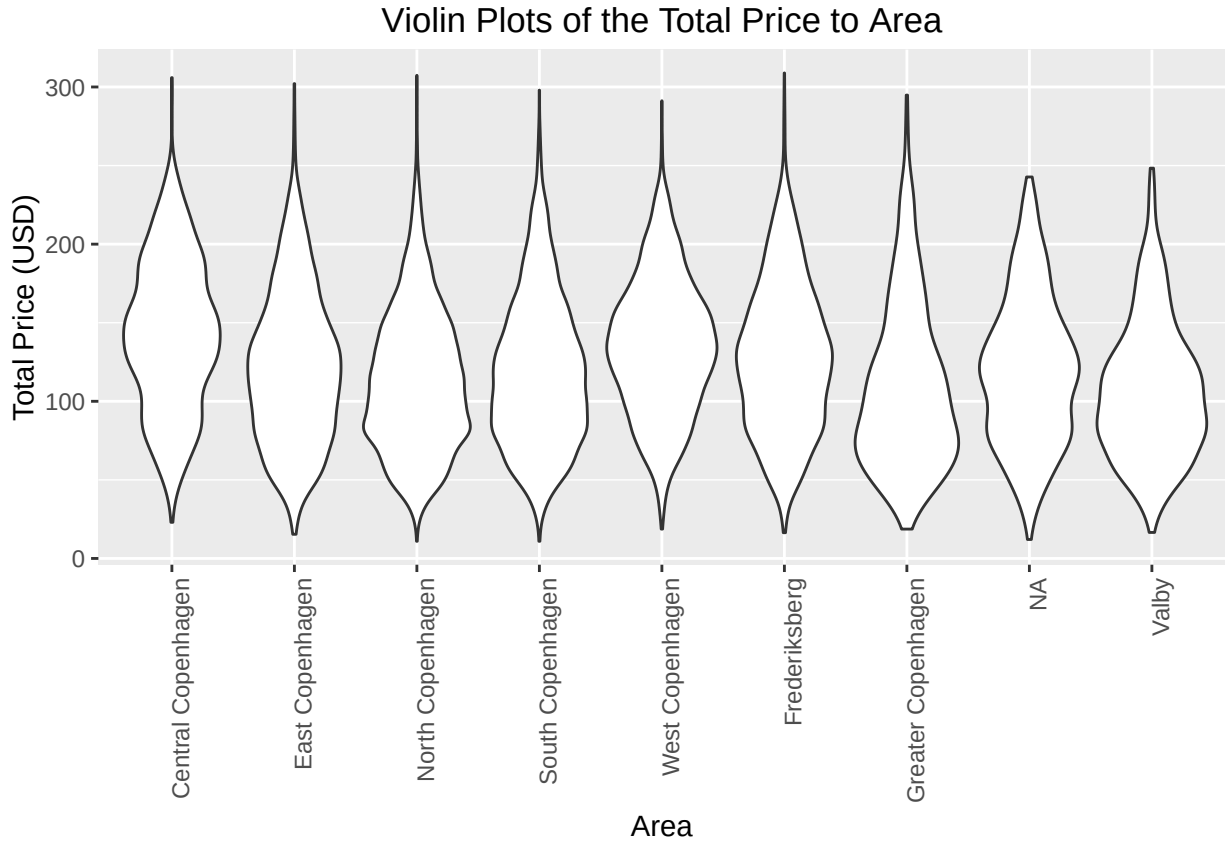
This graph shows a clear correlation for a larger square feet leading to a higher price. Unfortunately, further showing how unreliable the data on square feet we have is, we have over 100 observations with 0 square feet. Furthermore we have quite a few more with less than 70, making it really questionable if someone can really live in such a small room.

Another predictor that is usually expected to raise the price of a house is the region it is located in. Unfortunately the data itself did not provide common neighborhoods, therefore we had to create a few out of the lattitude and longitude given.

Table 2: Mean and Median of Price Depending on the Region

Area	Mean	Median
Center Copenhagen	142.3402	141.1364
West Copenhagen	134.5285	132.8634
Frederiksberg	127.8715	124.0941
East Copenhagen	121.8465	116.1521
NA	119.5094	116.1521
South Copenhagen	118.936	115.3248
North Copenhagen	113.2218	107.7137
Greater Copenhagen	109.3131	98.94437
Valby	107.2231	99.77167

Now we can check the distribution of each Area. Most of them are alike but we can clearly see that houses in the Valby region are generally cheaper than their counterparts.

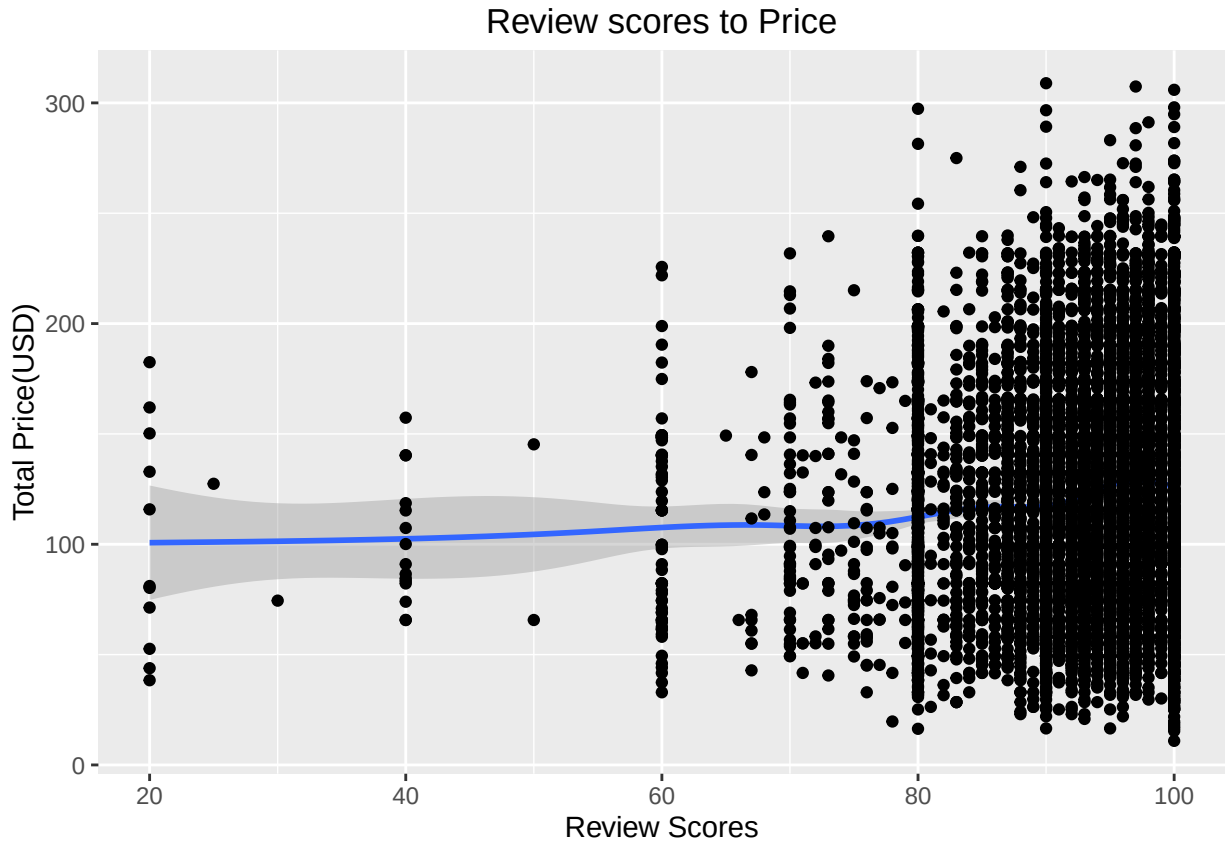


You can also tell that places like Greater Coppenhagen have some extreme values as the median is smaller than the mean. This violin plot confirms it,most observations are concentrted at the under 100 range but the extreme values raise the average.

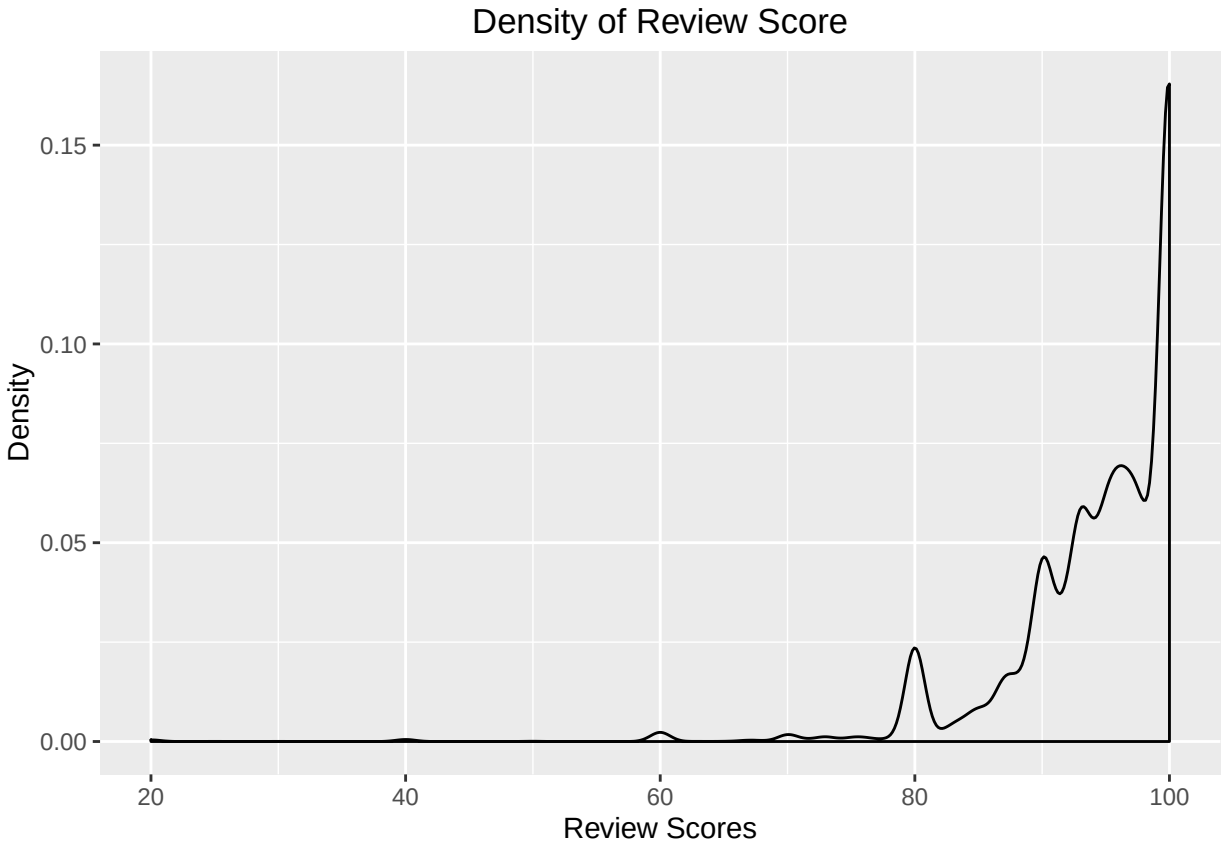
It is important to note that some areas have much mroe observations than others. For example Valby only has 772 listings compared to N. Copenhagen with 4886. This can either be due to people living there not being able to afford having someone else live in their house, for lack of space. Considering Valby is larger than N. Copenhagen it can also indicate that the density of people is greater in that region.

Table 3: Frequency Table of Listings Depending on Area

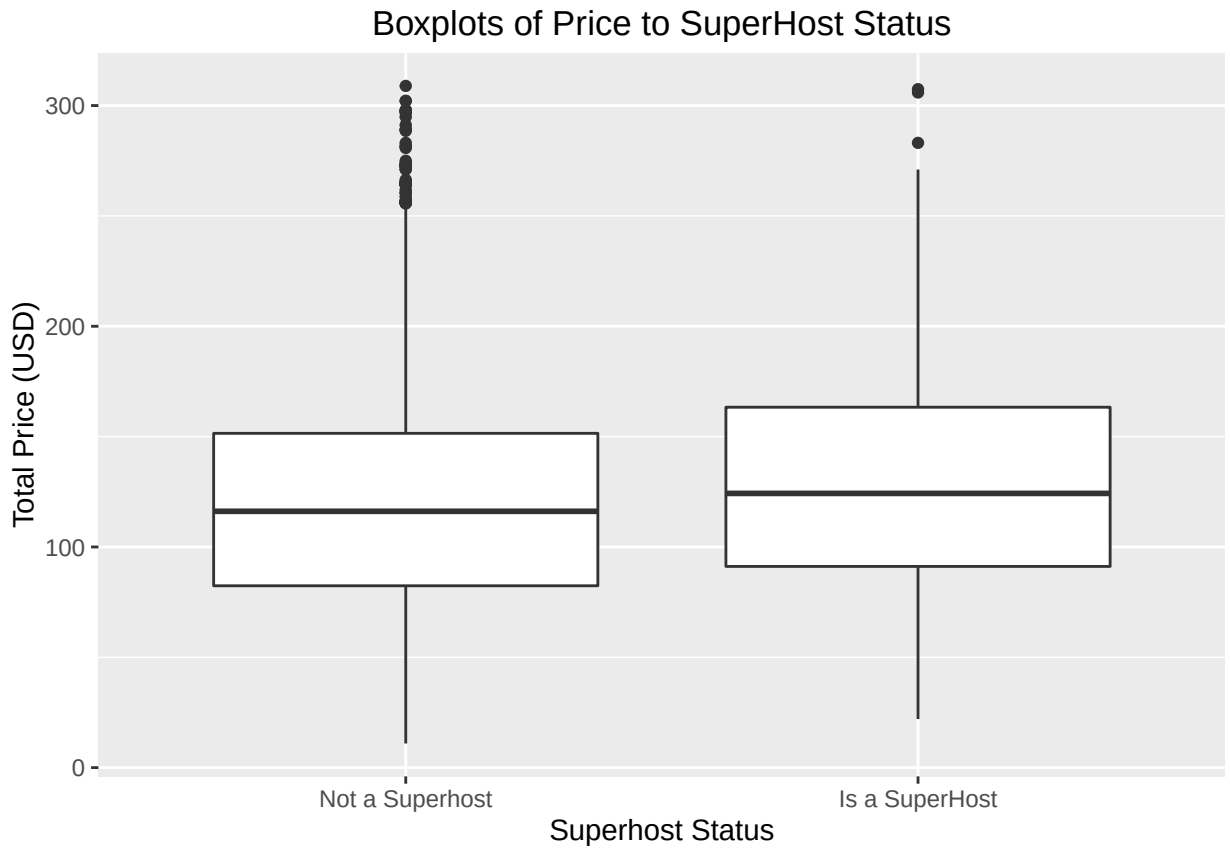
Area	Frequency
Center Copenhagen	2292
West Copenhagen	2549
Frederiksberg	2408
East Copenhagen	2379
NA	653
South Copenhagen	3798
North Copenhagen	4886
Greater Copenhagen	808
Valby	772



Here we can see rating does not seem to really influence anything. Though it is hard to say for any value thats under 60, the standard error becomes really small as we get pas the 80 range. There is a small increase past the 90 margin, which might just be enough to be considered statistically significant. But this also shows that low reviews are pretty rare as can be shown here



Most values are near the 90/100+ range. Odds are most people just click top score if they had no issues.


Boxplots of Price to SuperHost Status
Total Price (USD)
300
200
100
0
Not a Superhost
Is a SuperHost
Superhost Status