Addressing Issues from Iteration 1

Instructors Comment:

Part 1

d) Two plots for distribution, spread discussed?

You should fix your plots, using a single function don't allow you to properly analyze your distribution and spread. For example, price contains extremely values for probably luxury houses/apartments. If you manipulate your ticks properly or even remove them from the data (your boxplot shows you that) you will see the variation properly in your plot. Please redo this part for a proper assessment. I believe most your variables are skewed and you couldn't show that with your plots.

Resolution:

We replotted both the histograms and boxplots. We increased the number of bins (75 bins) for the histogram to see the distribution better. And for the boxplot we used a log axis for the y-axis so the plot isn't too squeezed together and it is better to visualize the distribution.

Instructors Comment:

Part 2

g) Is the technique to find outliers proper? Was it justified?

Your technique seems to work but it was risky. With a positively skewed distribution (this is an insight from other values) it would be more indicated to use around the 95 percentiles as a threshold to remove 'extreme values'. Fix that for the next iteration.

Resolution:

Instead of using the pervious technique of removing outliers that are 3 standard deviations of the mean with higher or lower limit, we used 95^{th} percentile and replaced them with the median as the distributions were skewed.

Check **Iteration 2, Part 1 – Addressing instructor's comments from Iteration 1** (in the notebook) to see the changes that were implemented.