

Visualization : Improve and Extend

Michael Cubriel, Gabriel Dare, Joaquin Salas, Sandip Singh

Group Effort

Q2: What is the distribution of room types across cities?

We changed from 6 separate bar graphs to 1 condensed bar graph, added color to columns, changed title, axes, and city names. We then swapped the places of Private rooms and Hotel rooms as private rooms were always higher so it reads easier than having a gap.

Gabriel Dare

Multicollinearity

Original correlation heat matrix and table filtered out everything that had a correlation less than 0.50 which was basically everything that wasn't a comparison of a variable against itself, so the filter was removed. Duplicate values were also removed for the sake of clarity. Minor edits were made to the titles to reflect the changes.

Michael Cubriel

I added more insight to the top 10 most expensive listings by creating a bar graph showing the average price per city in descending order. I cleaned up the graph of the top 10 most expensive listings by changing the style to match the other graphs. I also added two graphs comparing the minimum number of nights vs price and the number of reviews vs price but the data did not lend well to being graphed as it showed no correlation between the two so they were deleted. If there were more cities from which data was collected perhaps there may be a correlation.

Sandip Singh

I worked on "Model 1 - Multiple Linear Regression." Initially, six scatterplots were provided with every data point in the dataset to try and show if there was any sort of linear relationship between availability and each other variable. However, the scatterplots did not do a very good job of showing if there was a linear or any other sort of relationship. This is because the data was entirely clustered to the left to the point where there was no distinction between individual data points, therefore limiting the amount of information shown. Additionally, the clustered data points were only to the left, as we moved to the right there were almost no data points whatsoever, so this was not a good model. I decided to create six different contour plots instead, which could provide better visualization when trying to find relationships. I created the contour plots using a random sample from each of the six variables. This is because when trying to create the plots with the whole data set, the program would entirely crash due to the large amount of

data. Another change I made from the original plot was to decrease the values from the x-axis to get a more precise contour plot. Since there was no obvious relationship even with the scatterplot, I tried a log transformation that could potentially make any patterns more evident. I also used contour plots for the log transformation.

Joaquin Salas

I edited two plots, the first one being “Top 10 Listings with the Highest Minimum Nights”. Before it was a table and very uninteresting to look at. I created a bar chart that easily shows the data and creates easy conclusions. On top of this, I made a new conclusion that can be seen below the graph from what I gathered by looking at the graph.

My second plot is “Top 10 Listings with the Most Reviews”. I went in the same direction as the plot before. By making the data visually appealing, I am able to better understand what is happening. This is true for this example since I was able to make a good conclusion about listings with a desirable location, price, and the amount of reviews. All of this data makes sense and can be further analyzed. All my conclusions aligned with the ones stated in the prompt.