Excel Data Cleaning & Analysis Project – Nashville Airbnb Data from June 22nd, 2024

# **Purpose**

This project was performed to exercise my skills on data cleaning withing Excel and taking what I have learned from Alex the Analyst on YouTube to perform some analysis and loosely base this project on his Full Project in Excel in his Data Analyst Bootcamp playlist (<https://www.youtube.com/watch?v=opJgMj1IUrc&list=PLUaB-1hjhk8FE_XZ87vPPSfHqb6OcM0cF&index=27>).

# **Data Cleaning**

# Import and Backup Data

I imported my data from <https://insideairbnb.com/get-the-data/>.

The datasets I am working with are Airbnb data from Nashville, TN from June 22nd, 2024.

I was hoping for data from Atlanta, but this website, which scraps the official Airbnb website for listings, did not have data on the Atlanta metro.

I opened the file and immediately created a new sheet for a working dataset and left the original data in a sheet titled ‘listings\_raw\_data.’

In the new sheet, I named it ‘listings\_working\_data’ and copied over the raw data.

This is done to maintain an original copy of the raw data so that if I, inevitably, make a mistake, I can retrace my steps with the unaltered data if need be.

I saved the worksheet in 2 forms in a secure location on my local device: one as my working file and one as a backup.

# Initial Data Inspection

I performed my first look of the data and noticed there were some empty cells spotted through the sheet and the column titled ‘neighbourhood\_group’ was completely blank.

Because there was not any useful data in this column, I went ahead and deleted it.

I applied a filter to each of the columns and looked at the entries for each one and I did not see any strange or unexpected characters nor did I see text strings where I expected numbers or dates, which was nice.

# Remove Duplicates

Next, I used the ‘Remove Duplicates’ feature for the data in Excel.

The function found no duplicates in the data.

# Handle Missing Values

I went through all the columns I wanted to keep, and replaced the missing values with ‘N/A’ as a value on the listings\_working\_data sheet.

This was in the following columns:

Price, minimum\_nights, last\_review, reviews\_per\_month, and license.

One exception was I added the value of ‘0’ to the column number\_of\_reviews in the empty cells.

I also went ahead and created a new sheet for the analysis phase that removed those rows that did not have price as that would throw off my calculations. In this new sheet, I also renamed the neighbourhood column to neighbourhood\_district so that the values could be sorted ascendingly.

# Data Formatting

I then looked at the data to determine if I needed to reformat any of the data. I do not believe that I need to as it is in a very workable state as is and it seems to have conformity in its data types and capitalizations.

# Remove Unwanted Characters and Spaces

I used the CLEAN function on the ‘name column’ as it looked like there might be some weird characters included in that column.

I did not find a reason to use the TRIM function as there seemed to not be any issues in that regard.

# Data Parsing and Transformation

There was no need to split or concatenate data from the cells that I could determine.

# Delete Unnecessary Formatting

Again, I did not find any reason for this to be done as I didn’t perceive any of the formatting as unnecessary.

# Save and Export

Saved and ready for analysis!

# **Data Analysis**

I completed my analysis by creating pivot tables that I then copied into a new sheet with slicers to create a dashboard and a graphic showing the different locations of each of the neighborhoods.

The dashboard I created allows you to see and modify 2 graphs; one that shows you the average price per listing per neighborhood district and the other shows the average availability per 365 days per neighborhood district. This dashboard has slicers that allows the user to filter out the data so that they can see listing information for each accommodation type as well as per neighborhood district.

I also added a graphic I created in Excel with the 3D Map tool that shows you the different districts of the Airbnb listings by average longitude and latitude.