

# SQL Data Cleaning Portfolio Project notes

The purpose of this project is to showcase my data cleaning abilities within SQL.

*Data I'm using for this project can be found here:*

[https://github.com/NyeemahE/PortfolioProjects/blob/main/Nashville%20Housing%20Data%20for%20Data%20Cleaning%20\(unaltered\).xlsx](https://github.com/NyeemahE/PortfolioProjects/blob/main/Nashville%20Housing%20Data%20for%20Data%20Cleaning%20(unaltered).xlsx)

## Viewing and altering the data in Excel

- I opened my data in excel to get a look at what exactly I was dealing with. Typically I would do this in the SQL program of my choice however the one I am using (MySQL Workbench) does not import data well if certain parameters are unmet.
- MySQL workbench will omit rows of data that have blank cells if the datatype does not allow for empty cells. To get around this and save myself from data loss I use the F5 key to highlight all blank cells and replace them with "NULL".
- Number data types are also particular in MySQL, if a special character is in a cell under certain numeric data types the row will be omitted and result in data loss.
- To avoid this I checked the import logs for which columns were having errors in MySQL, found that it was the salesDate.
- I reformatted the salesDate column to "numbers" in Excel so that it would properly import into my database.
- I saved the file as a csv and uploaded it to MySQL.

## SQL data cleaning

- The specific queries used can be found in the " " file on my github
- Standardized date format using the str\_to\_date function of the SalesDate column
- Populated null property addresses based on the Parcel ID
- Separated property address column into two columns address and city..
- Separate owner address column into three columns address, city and state.

Nyeemah Elliott

- Altered “SoldAsVacant” field to have only “Yes” or “No” output instead of N, Y, Yes and No.
- Deleted columns that are not necessary for analysis.