# Defining the Question (The Problem Statement).

The data for Nashville housing is in Excel format and needs to be cleaned. In this project, I am going to import the data in SQL and clean it using SQL queries.

# Collecting the Data.

Data was collected from: <https://github.com/AlexTheAnalyst/PortfolioProjects/blob/2dbf63f2f2e8f7c3ff458abc8dc90ddd555f3e38/Nashville%20Housing%20Data%20for%20Data%20Cleaning%20(reuploaded).xlsx>

Snapshot of the data in Excel:

A screenshot of a computer

Description automatically generated

# Cleaning the Data.

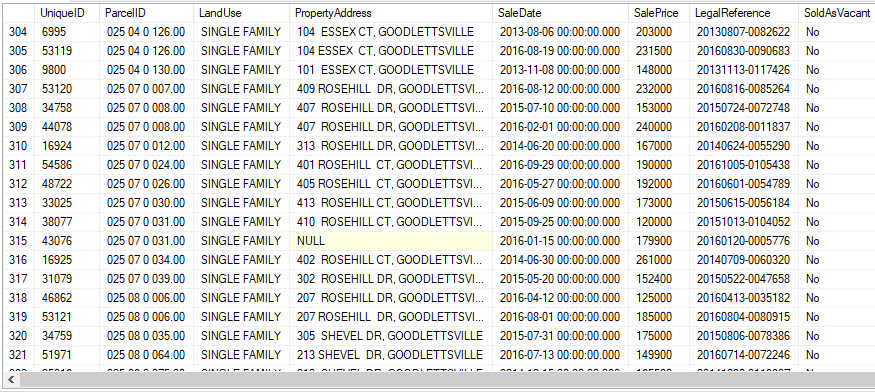
We are going to accomplish the following data-cleaning tasks:

1. Standardize the Date Format in SaleDate Field to YYYYMMDD format and remove the time information (which looks redundant)

A screenshot of a computer

Description automatically generated

1. Populate Property Address data where NULL.
2. Breaking out Address into Individual Columns (Address, City, State) from PropertyAddress field.
3. Change Y and N to Yes and No in the "Sold as Vacant" field.



1. Remove Duplicates, and
2. Delete Unused Columns.