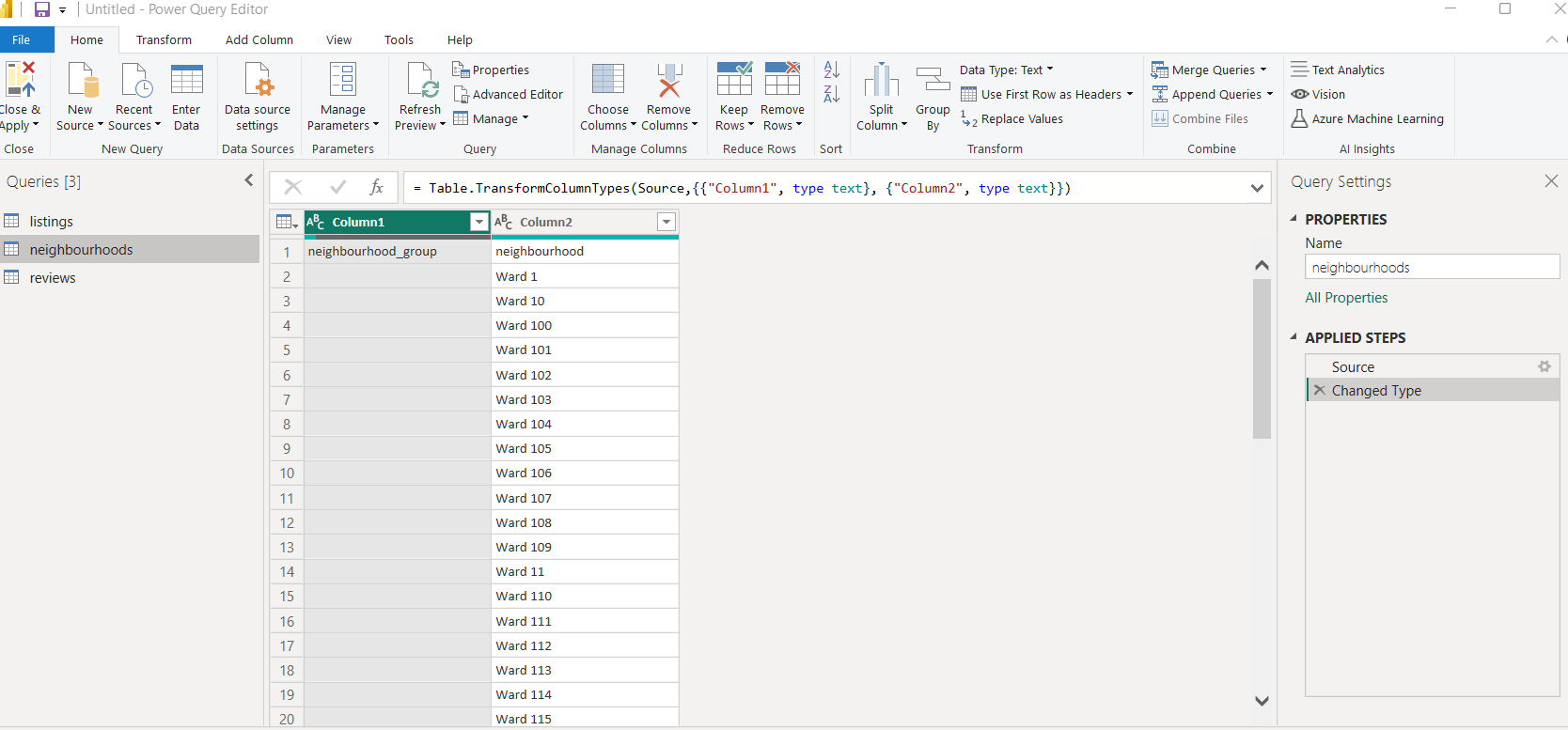
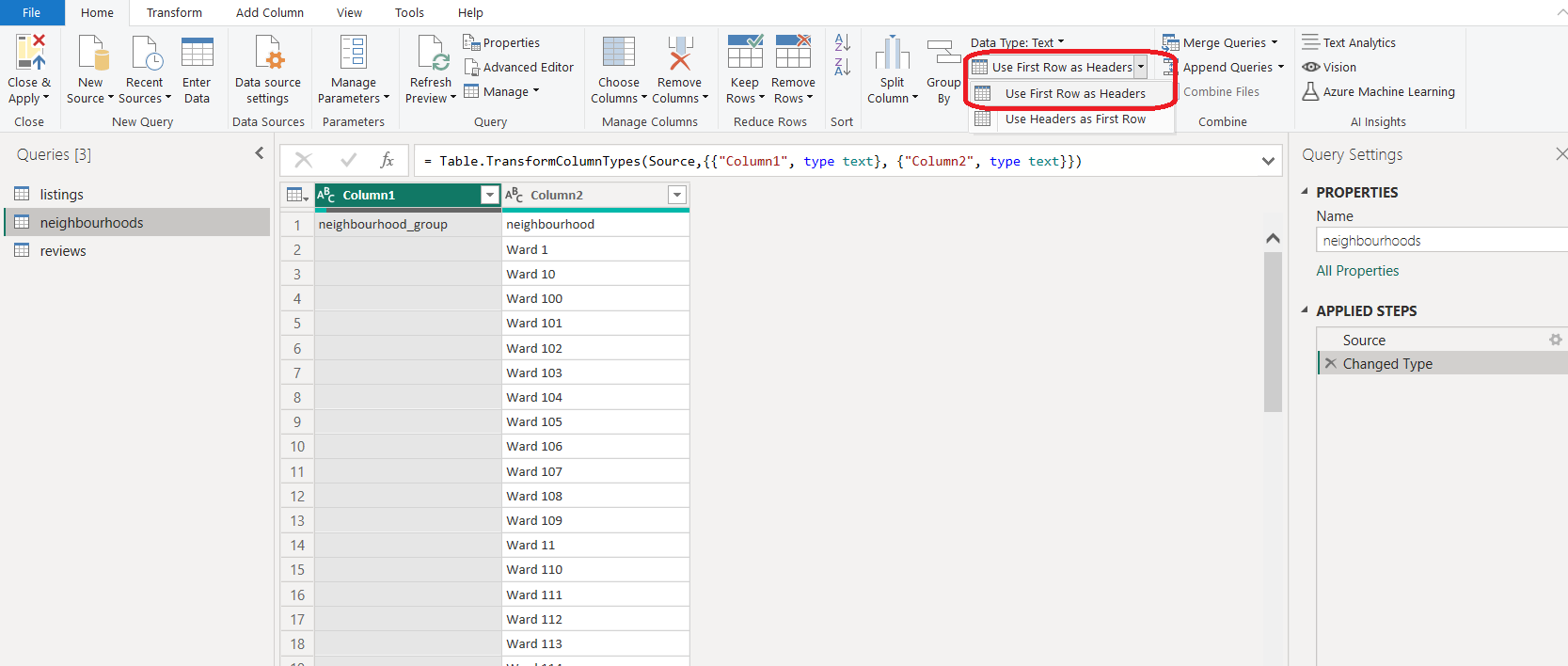
**Data Clean Project for the Cape Town Airbnb Dataset**

Data Cleaning on the neighbourhood table

Before



After

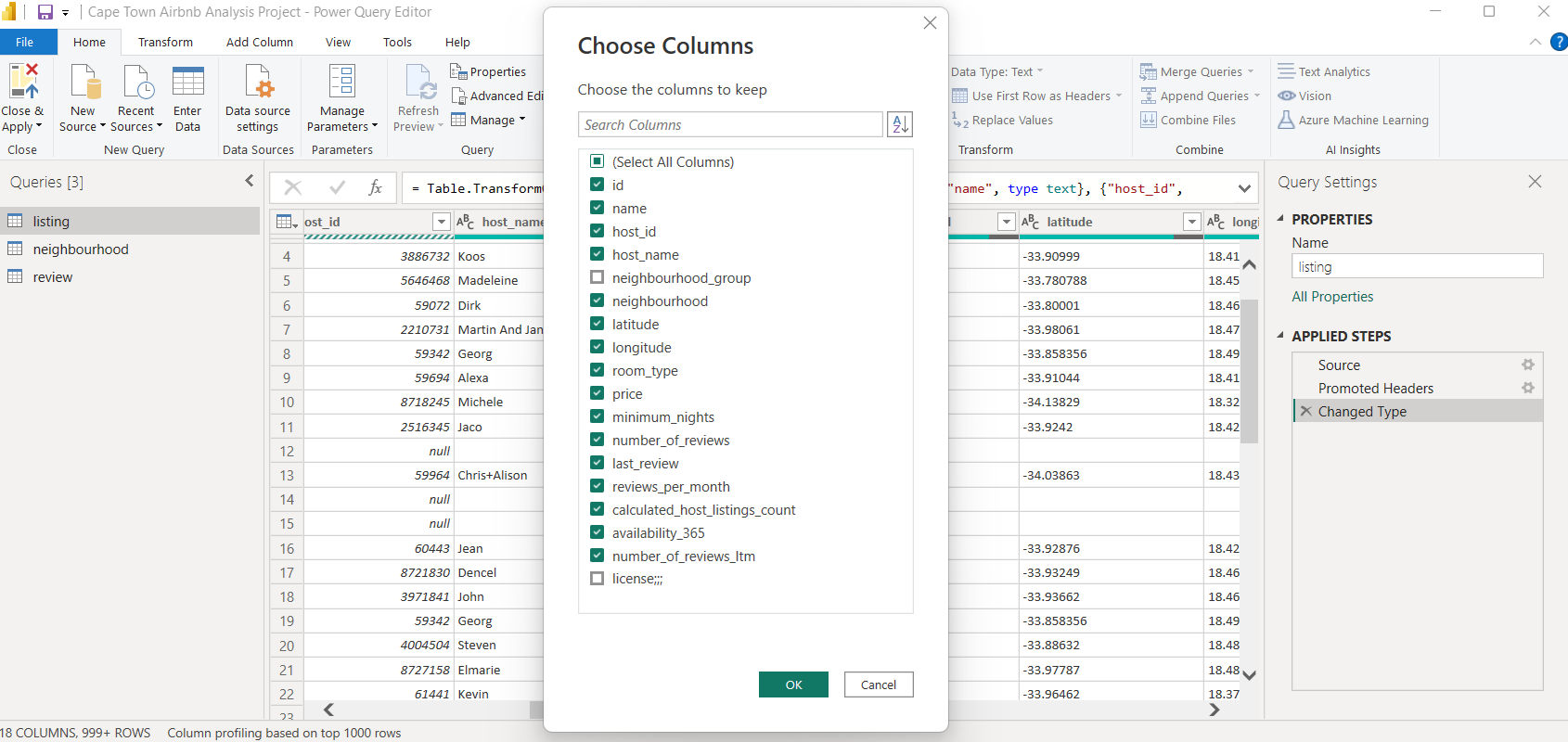


Graphical user interface, application, table, Word

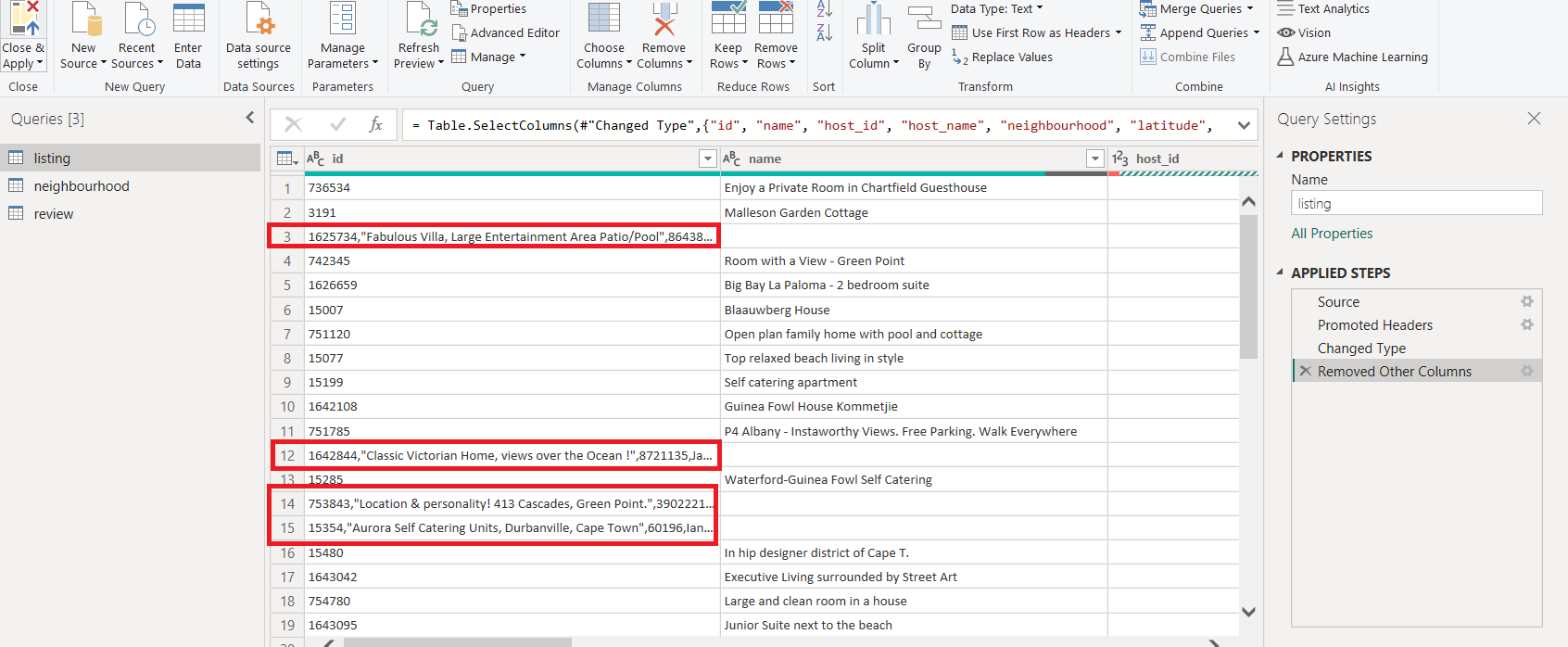
Description automatically generated

Data Cleaning on the listings table

1. Remove unwanted columns by unchecking them.



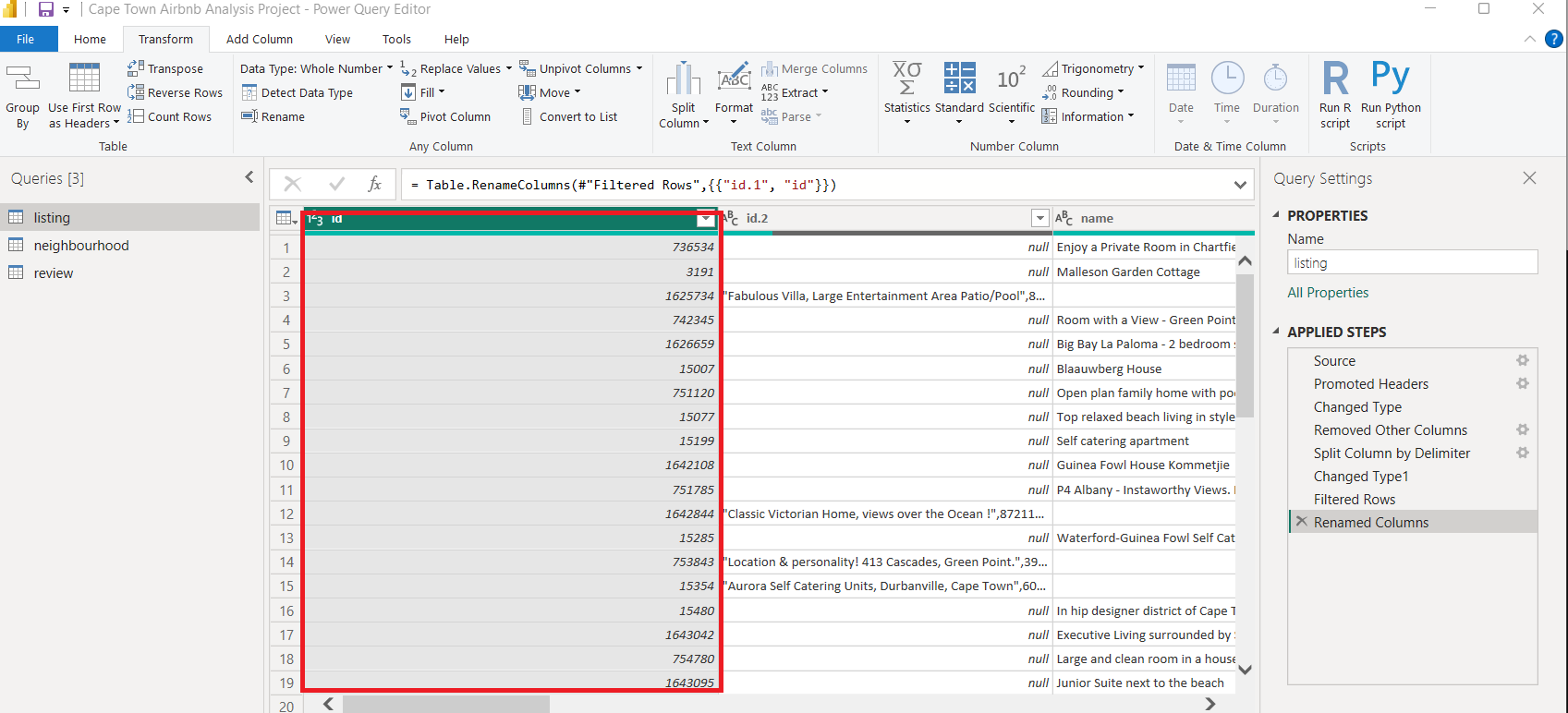
1. Remove unwanted text in host\_id field which is highlighted below



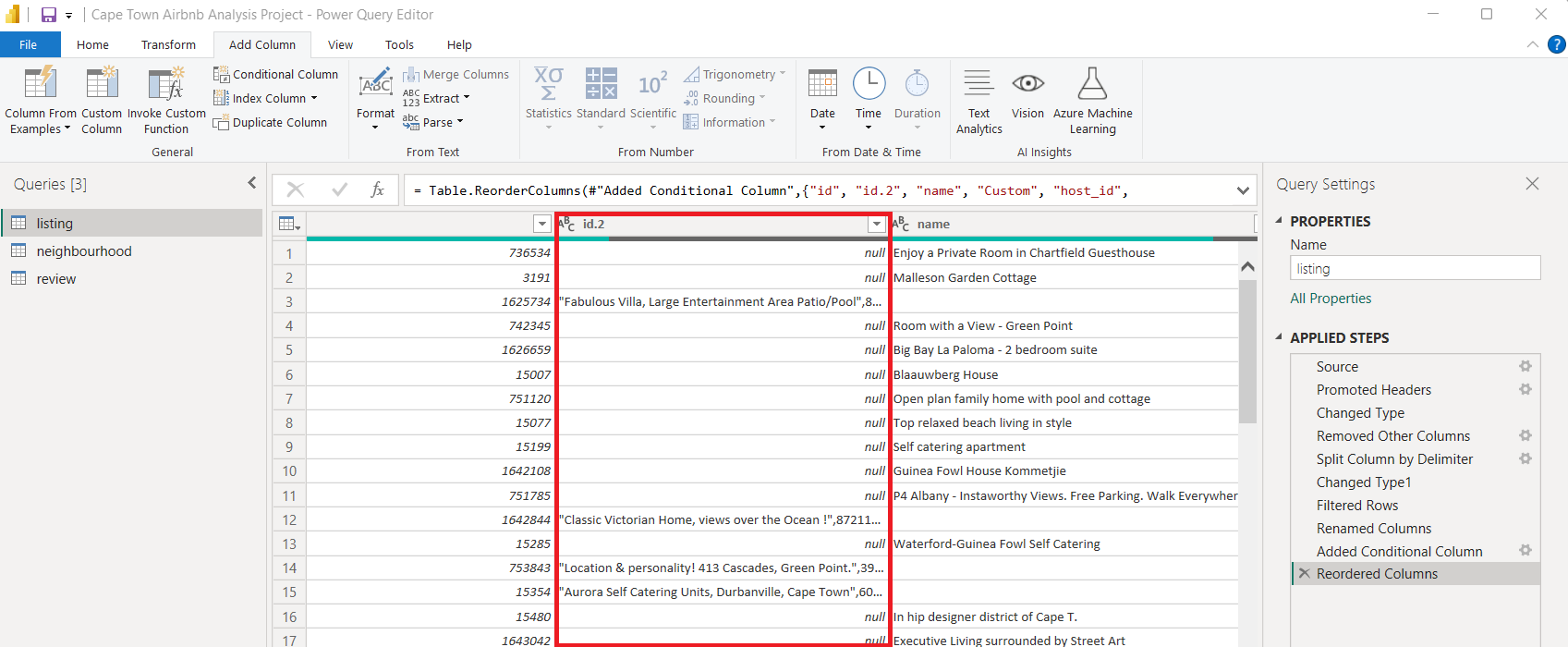
Graphical user interface, application

Description automatically generated

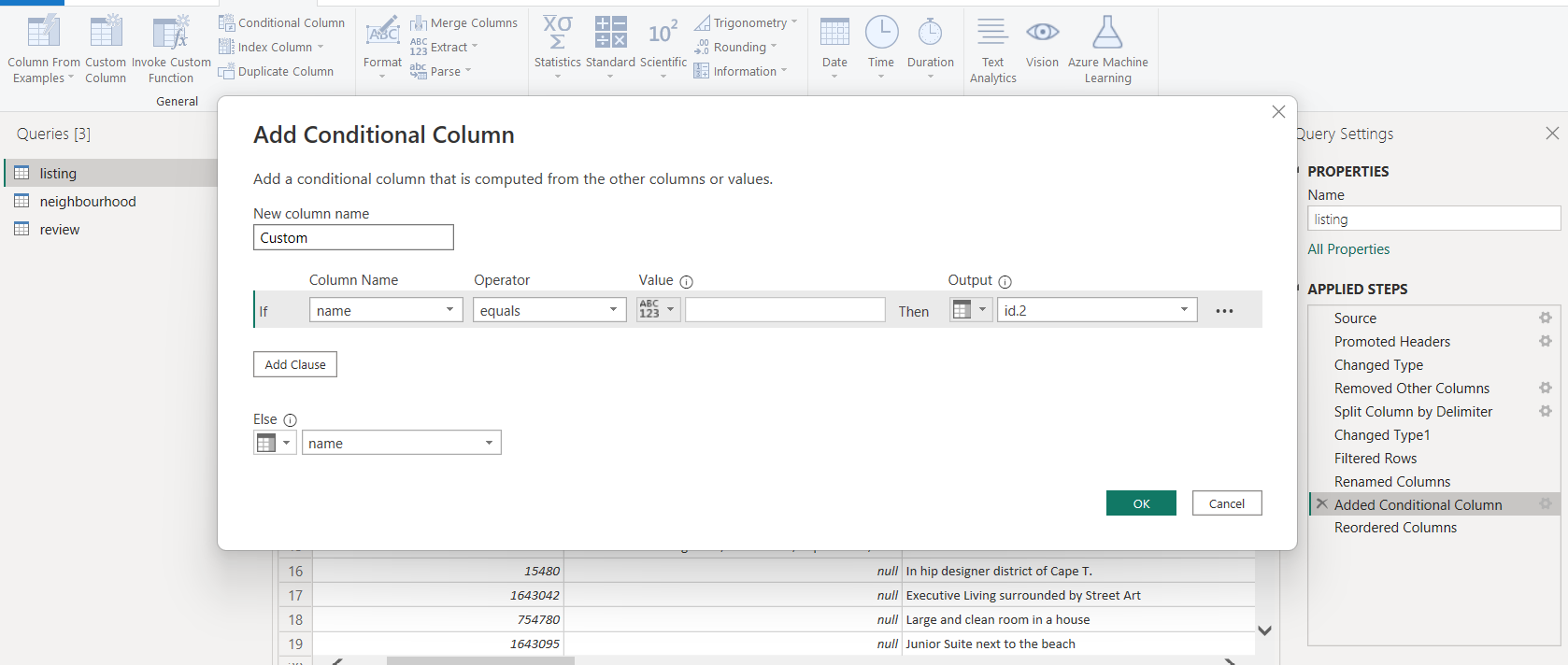
The unwanted data is removed from the id column as per below.



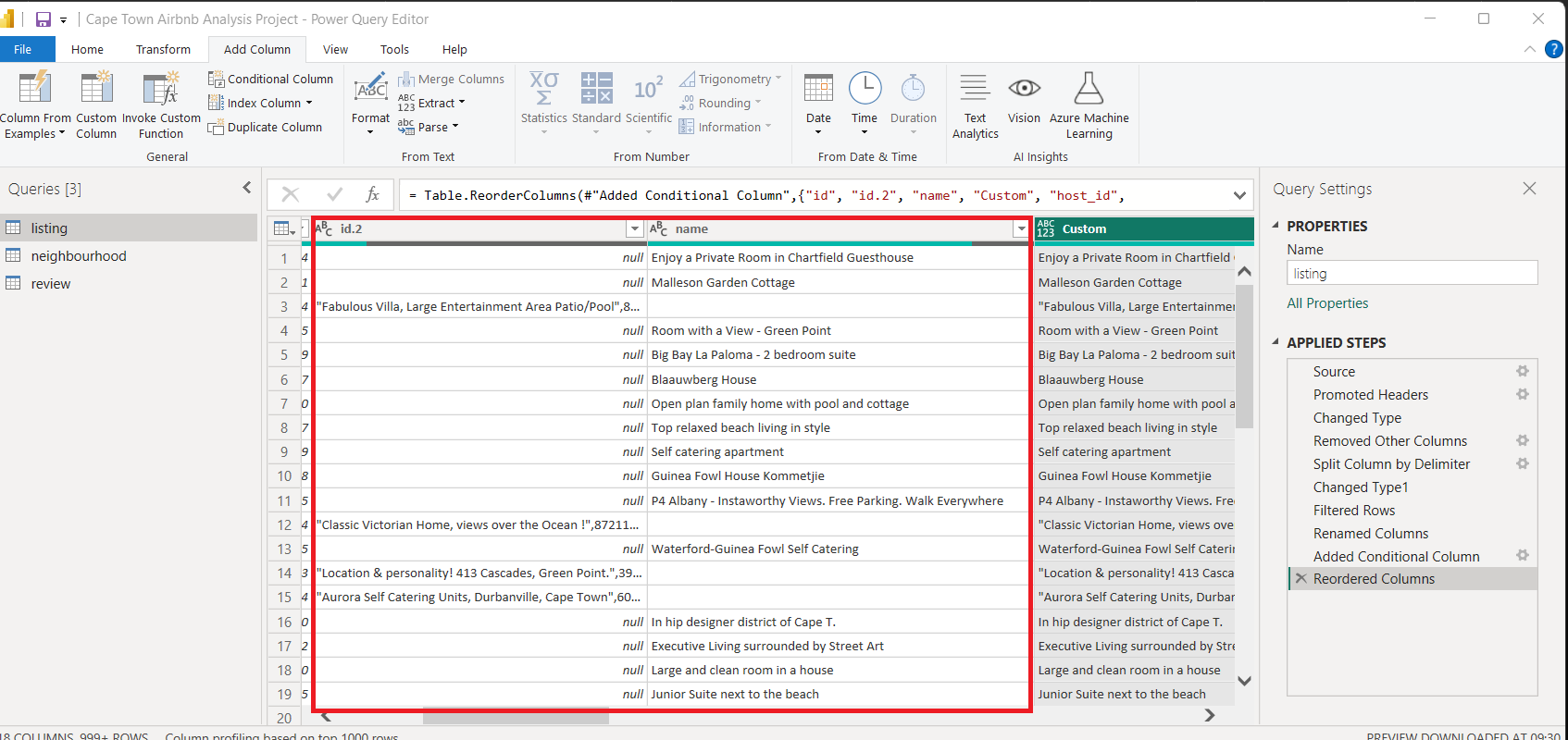
1. After splitting the id column, it created an new column with only the data we didn’t want in step 2. Now we need to fill in the blank values in the name column with the values from the new column which was created when splitting the id column.

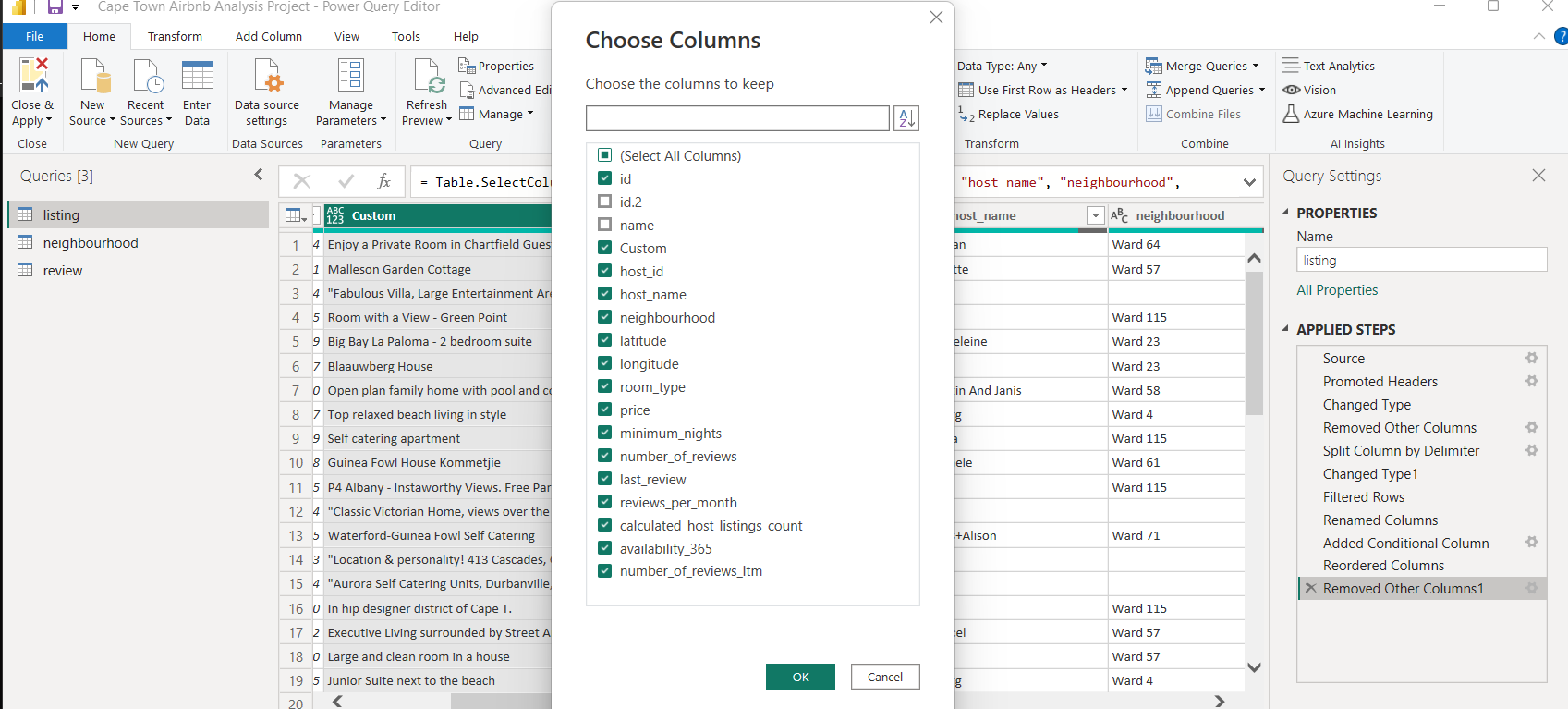


I created a conditional format which fills the blank values in name column with the values from the new split column which was created from the id column.

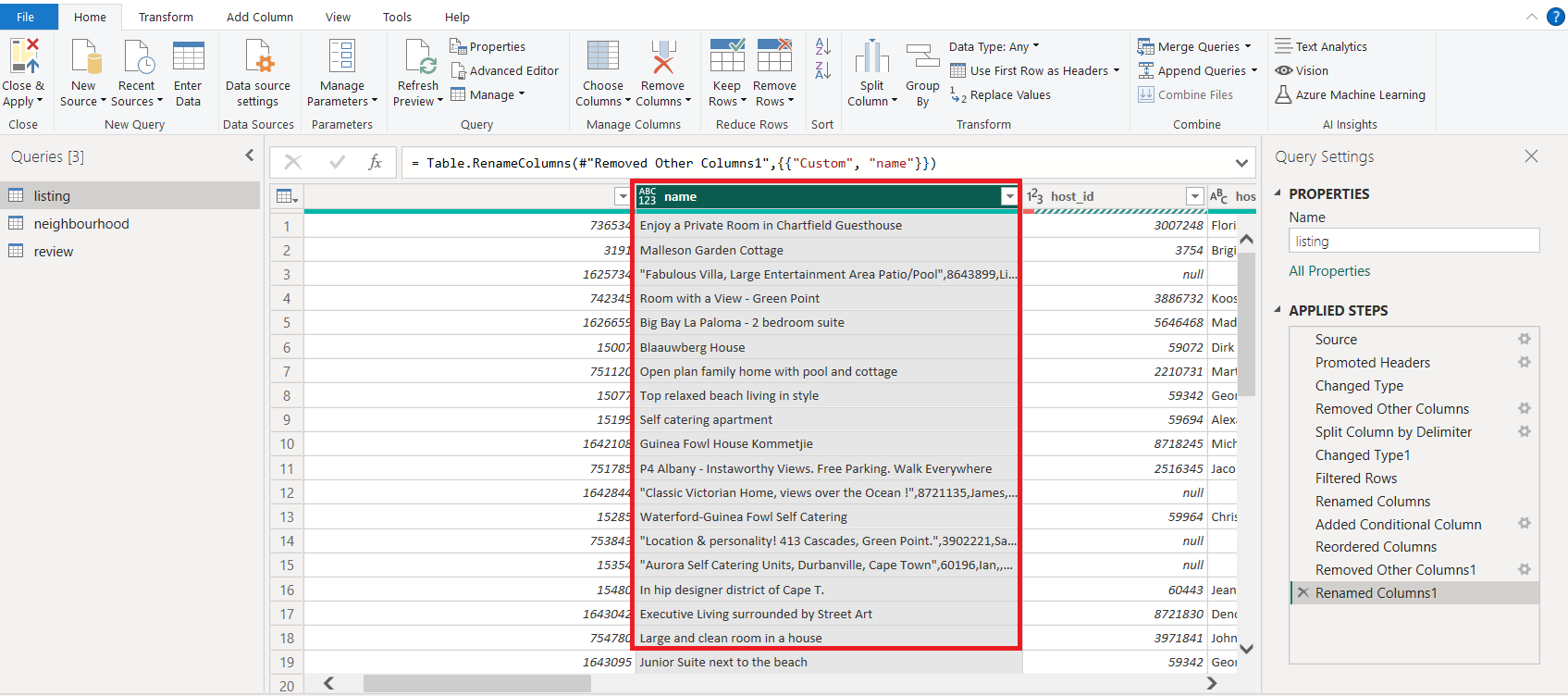


We are now going to remove the 2 columns highlighted in red below as they are no longer needed because we created a new column which populated the blank columns in the old name column.

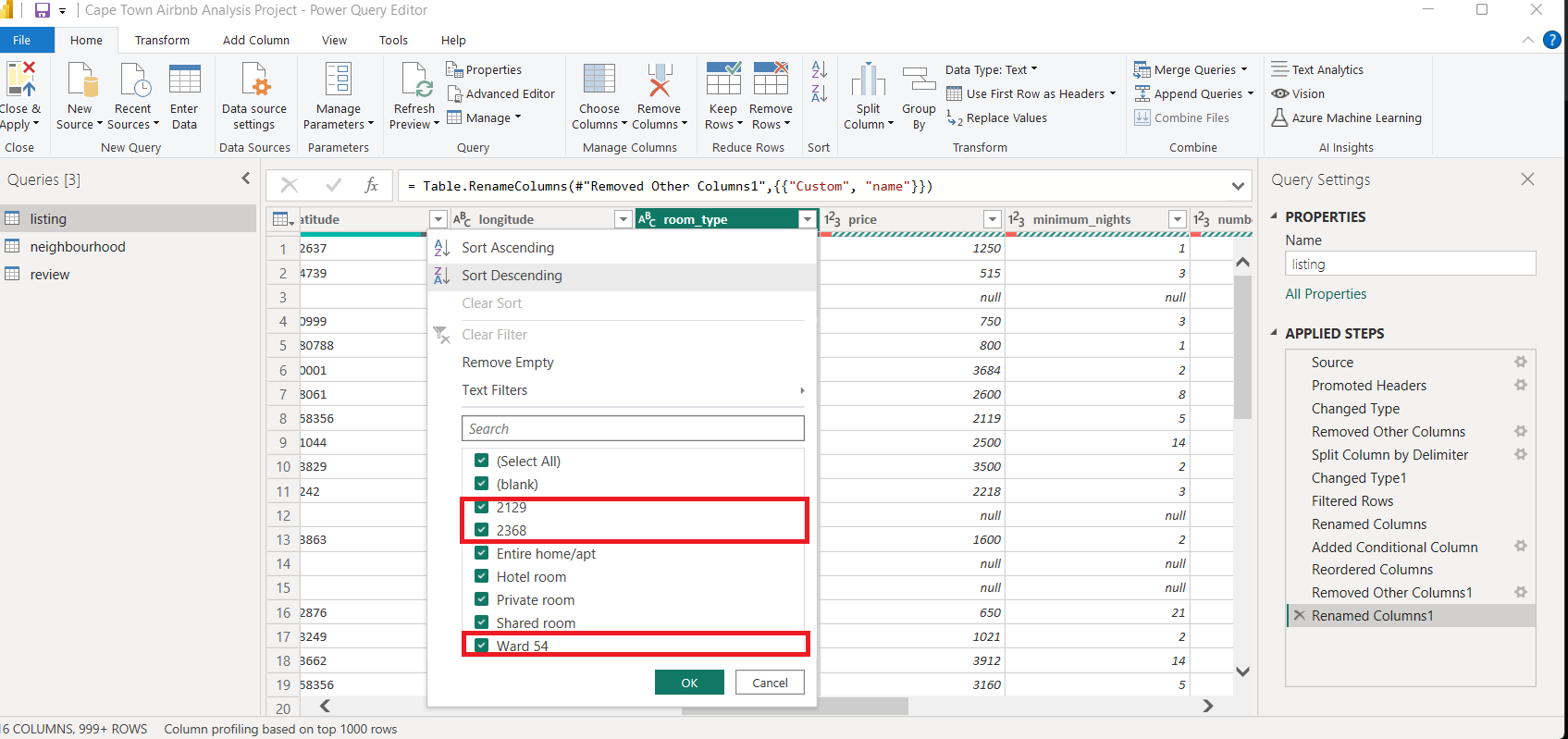




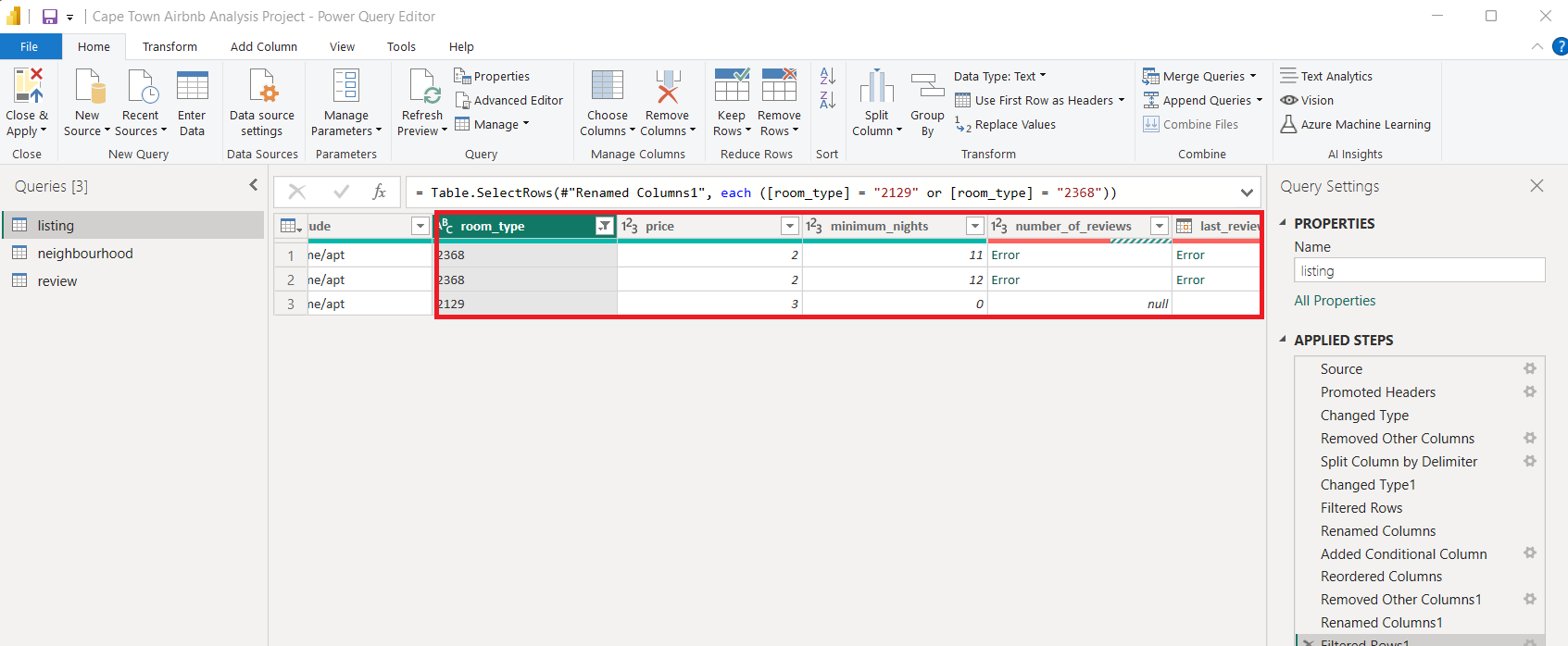
We have renamed the new column as name as highlighted in red below.



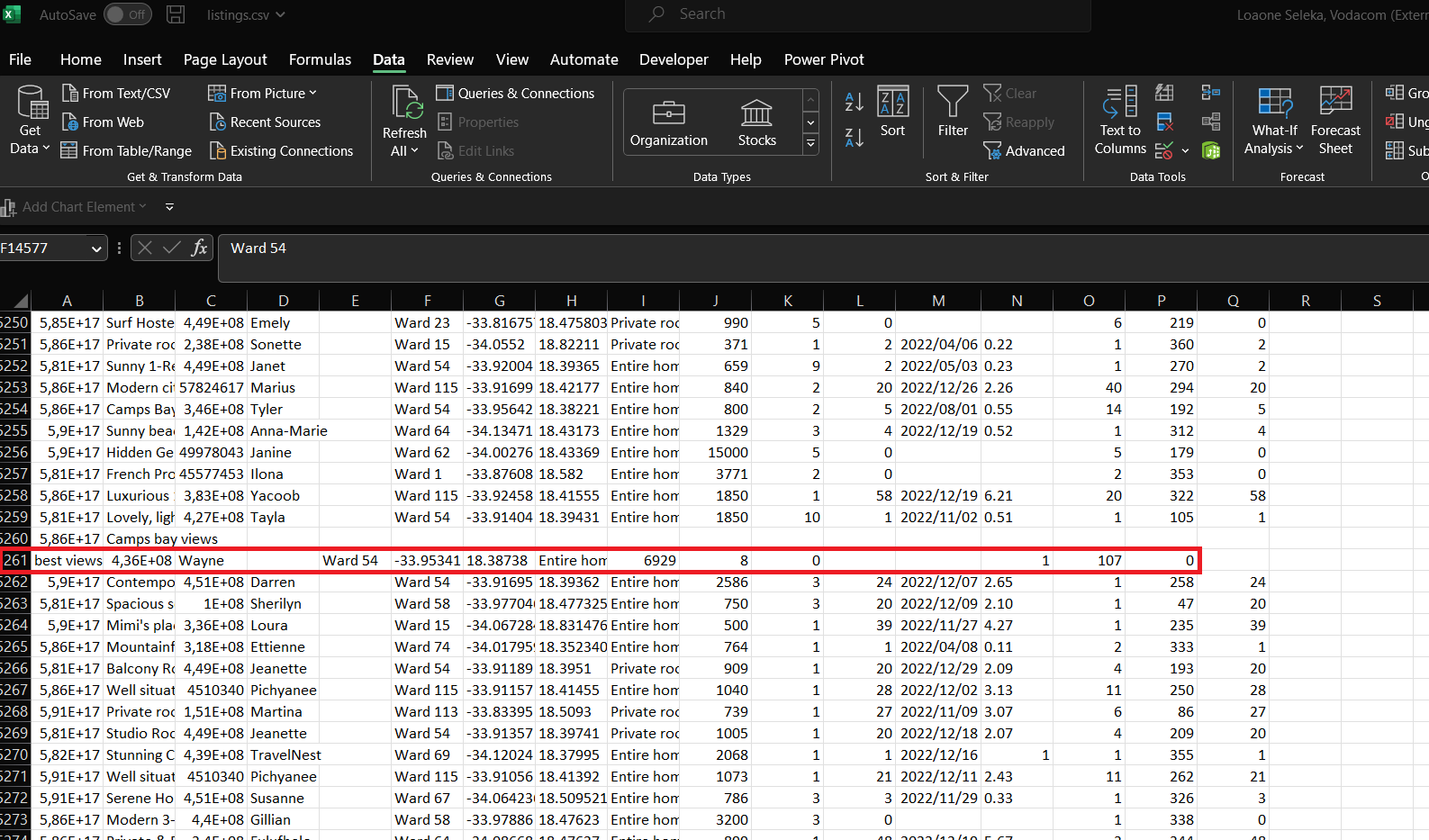
1. We are seeing some inconsistent values in the room type column and we need to remove those values.



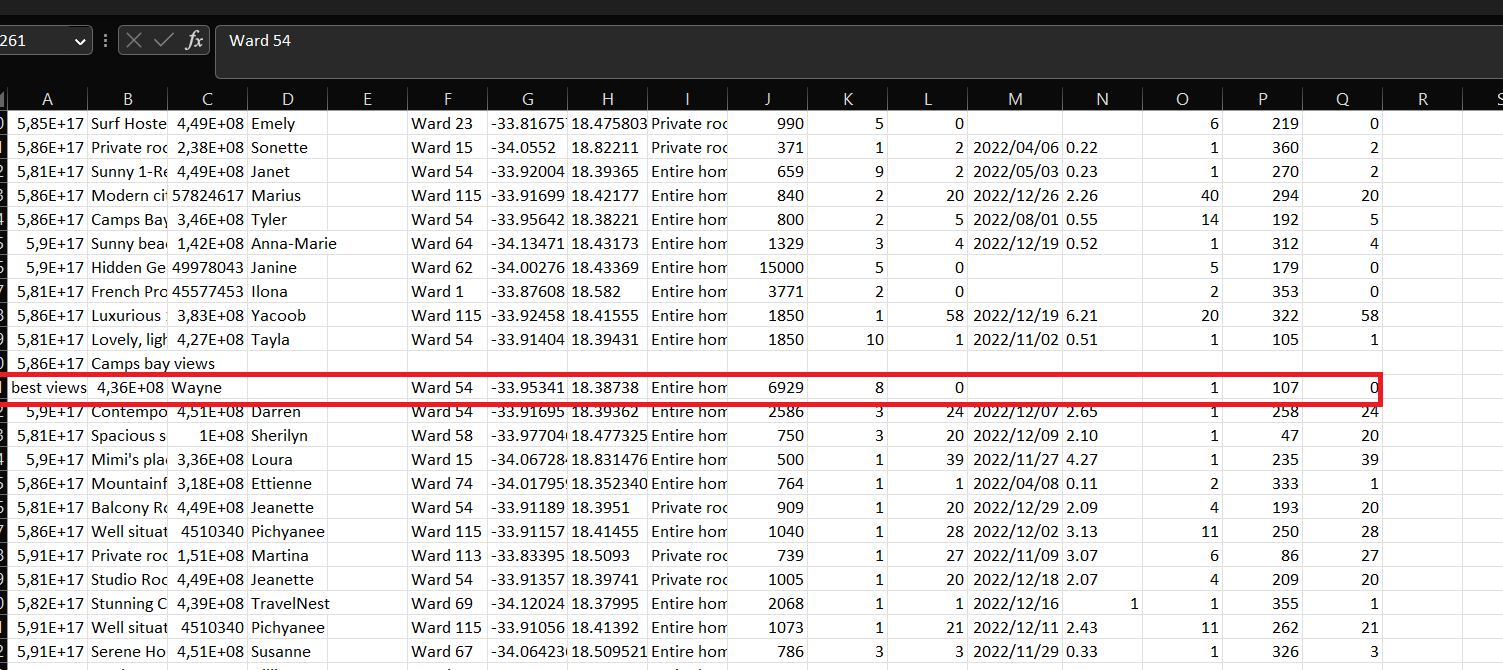
Firstly we filtered using the numeric values and we see that those values are in the incorrect columns so they need to be moved to the correct columns.



After investigating, looks like the data in those filtered rows shifted to the left so I need to fix it from the data sources which is an Excel file



I have shifted the data to the correct columns.



After fixing the data on the data source, the room\_type column is showing the correct data.

