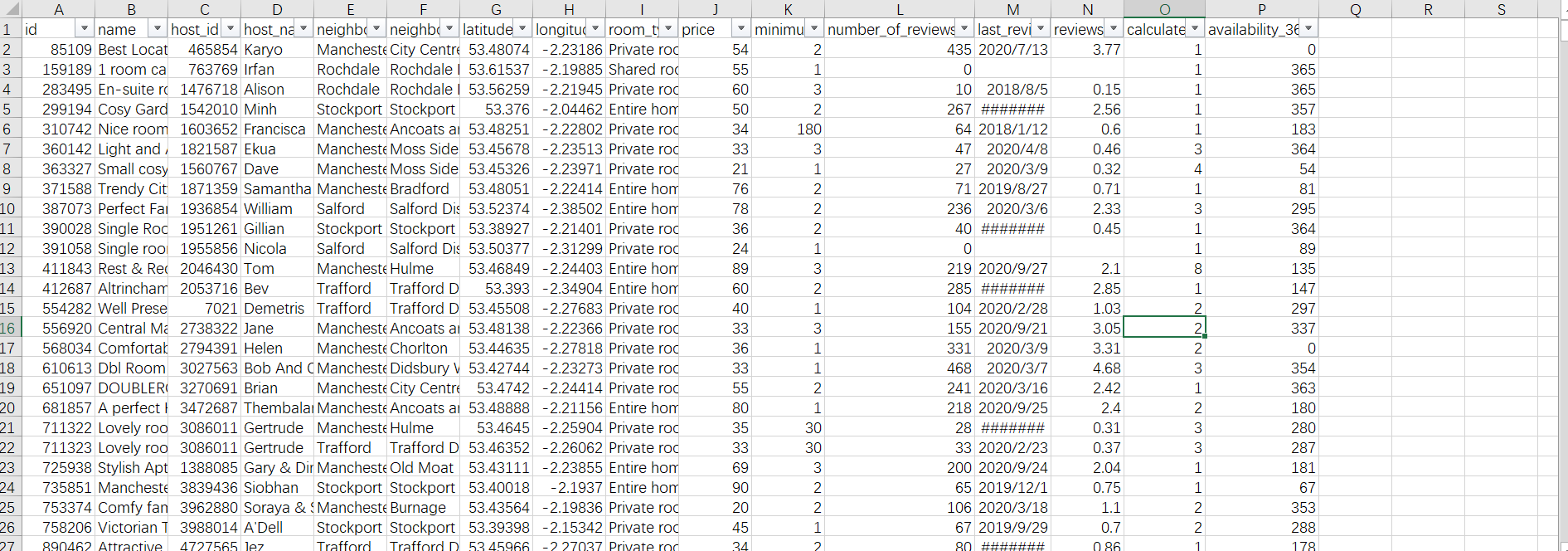
**Summary:**

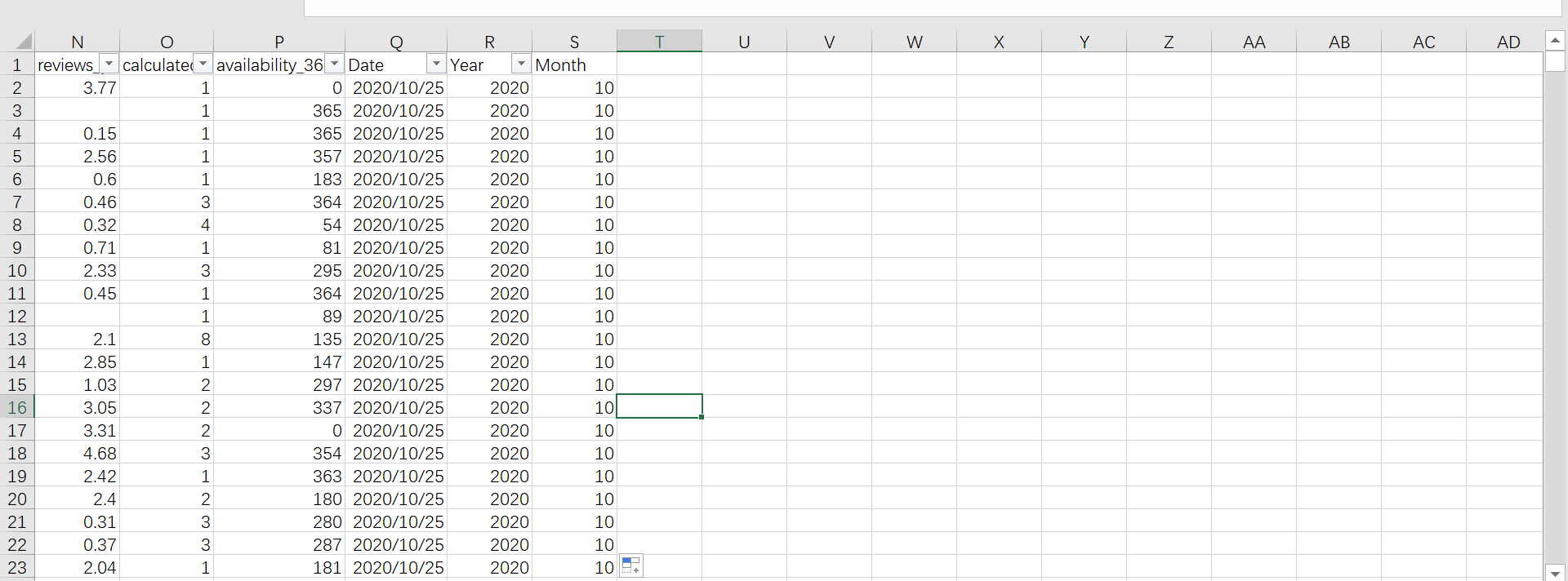
* We want to analysis the impact of Covid-19 onto the business activities of Airbnb, and we choose the data of London, England, United Kingdom and Greater Manchester, England, United Kingdom as our dataset because they are the two largest metropolitan areas in the United Kingdom.
* We also want to explore the strategies for the marketing team of Airbnb on how to shift their marketing strategies to minimize the lost by analyzing how tourist’s preference changed regarding two possible factors:
  + Type of room: we assumed that tourists prefer to rent entire apartment/house rather than shared apartment/house.
  + Geological factor: we assumed that tourists prefer to rent property outside of the metropolitan of Greater Manchester/London
* We need data to digitalize the change of tourist’s preference. However, data such as booking confirmation are not publicly released. Therefore, we decided to use ***Number of Review*** from ***listing.csv*** to digitize the change of tourist’s preference. Since the ***Number of Review*** is an aggregate data(the cumulative reviews from tourists since the first day it was listed on Airbnb), we need to calculate the monthly/yearly difference from month-to-month or year-to-year. For example, if we want the number of reviews for a specific listing in January 2020, we need to subtract its ***Number of Review*** in January 2020 by ***Number of Review*** in December 2019.
* Since we need ***Number of Review*** of 2019 and 2020 for both London and Greater Manchester, we downloaded the monthly dataset of in 2019, 2020 of London and Greater Manchester from insideairbnb
  + Dataset Downloaded:
    - London: December 2018; January-December 2019; January-October 2020
    - Greater Manchester: December 2018; January-December 2019; January-July, October 2020
      * The data of Greater Manchester for August, September 2020 is missing

**Data Cleaning procedures:**

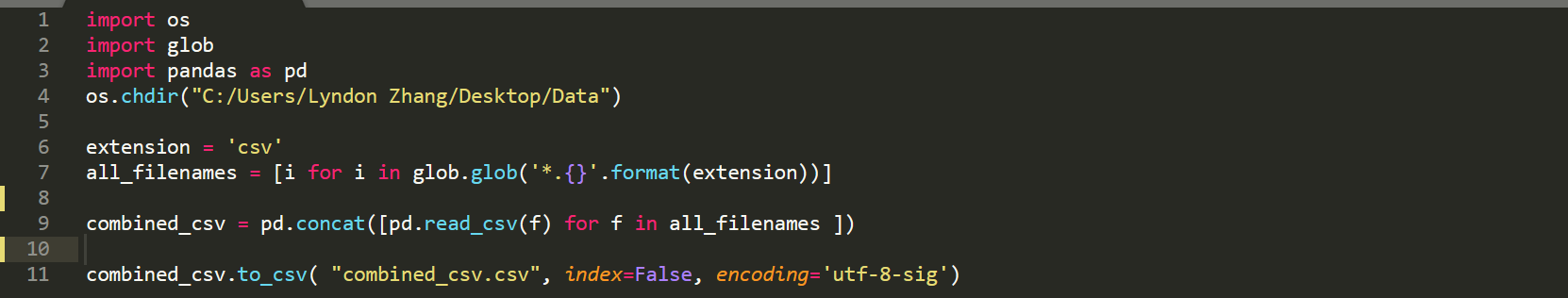
* The raw data(listings.csv, 25 October, 2020, Greater Manchester) has following attributes:

****

* After we downloaded the raw data, we need to add additional attributes - ***Date***, ***Year***, ***Month*** - to indicate the year and month of this listing.

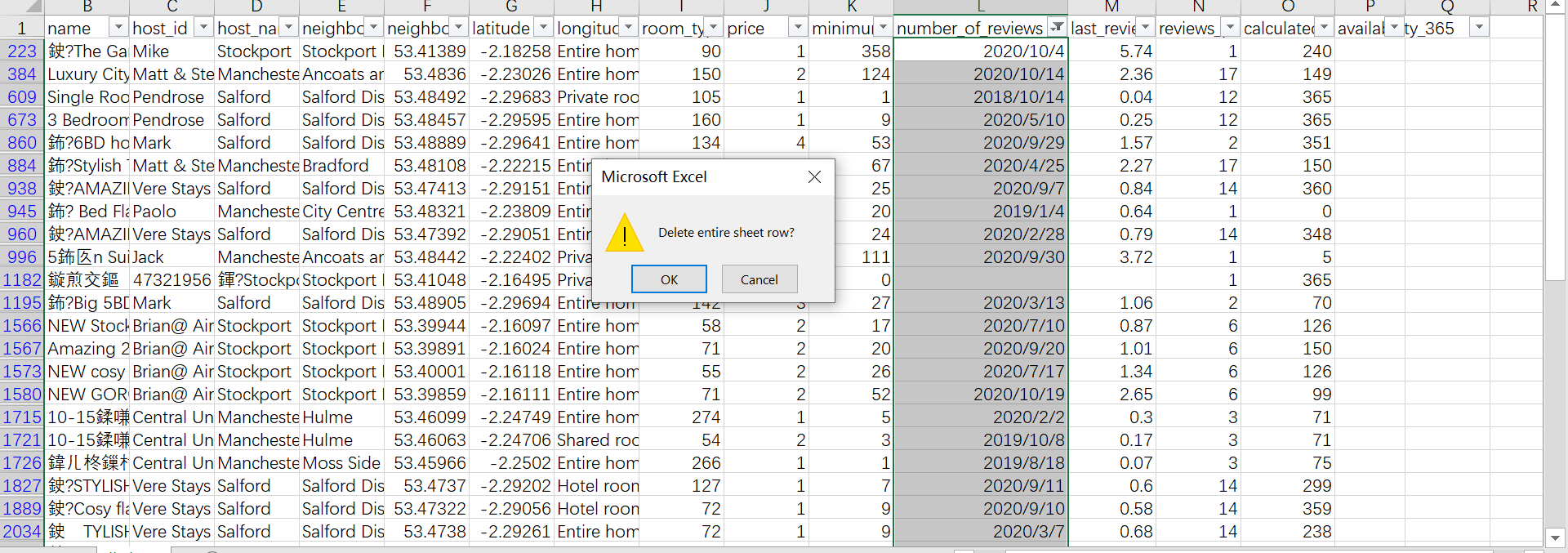


* Most data cleaning is done by using Excel. However, the comprehensive dataset is too large to open in Excel. Therefore, we use python to merge all monthly data to three separate datasets (we added the data of December in the year before, because we wanted to calculate the new monthly review of January in the year. For example, for the dataset in 2019, we add the data of December 2018, so we could calculate the new monthly view of January 2019).
  + London\_combined\_2019.csv: all monthly data of London in 2019, plus the data of December 2018 of London
  + London\_combined\_2020.csv: all monthly data of London in 2020, plus the data of December 2019 of London
  + Greater\_Manchaster\_combined\_2019-2020.csv: all monthly data of Greater Manchester in 2019 and 2020, plus the data of December 2019 of Greater Manchester

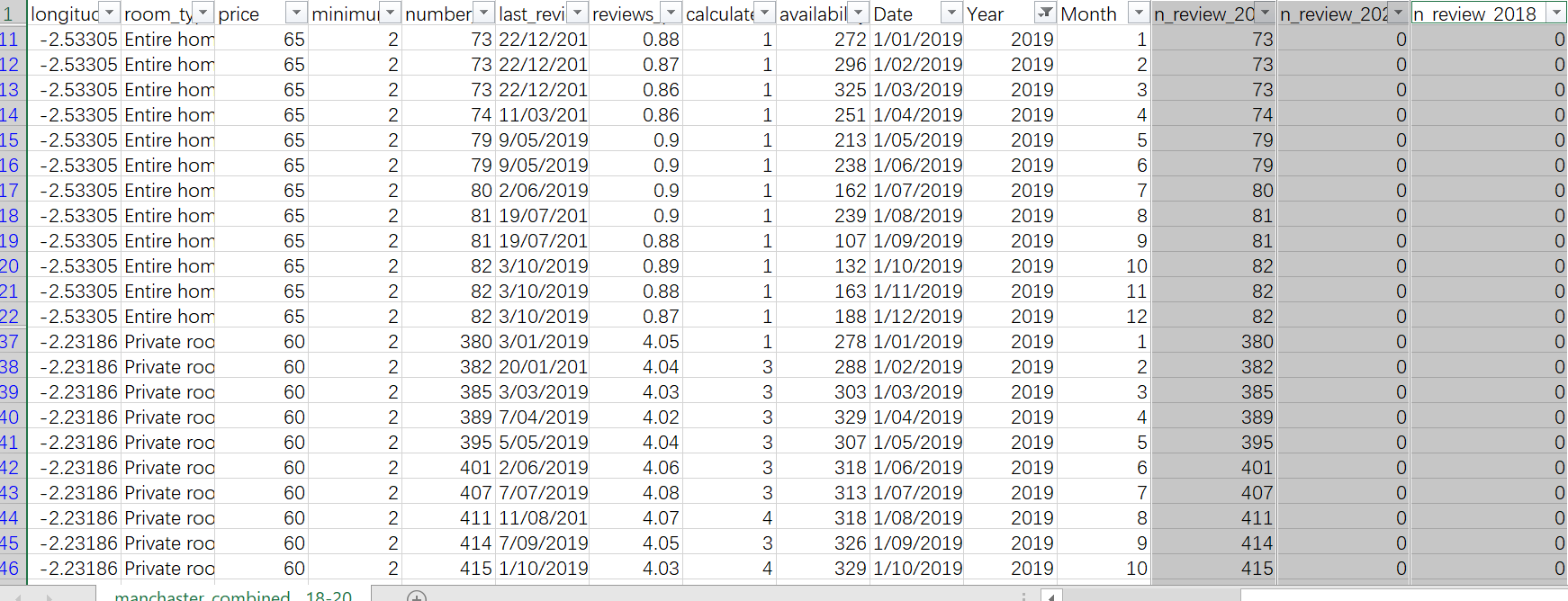




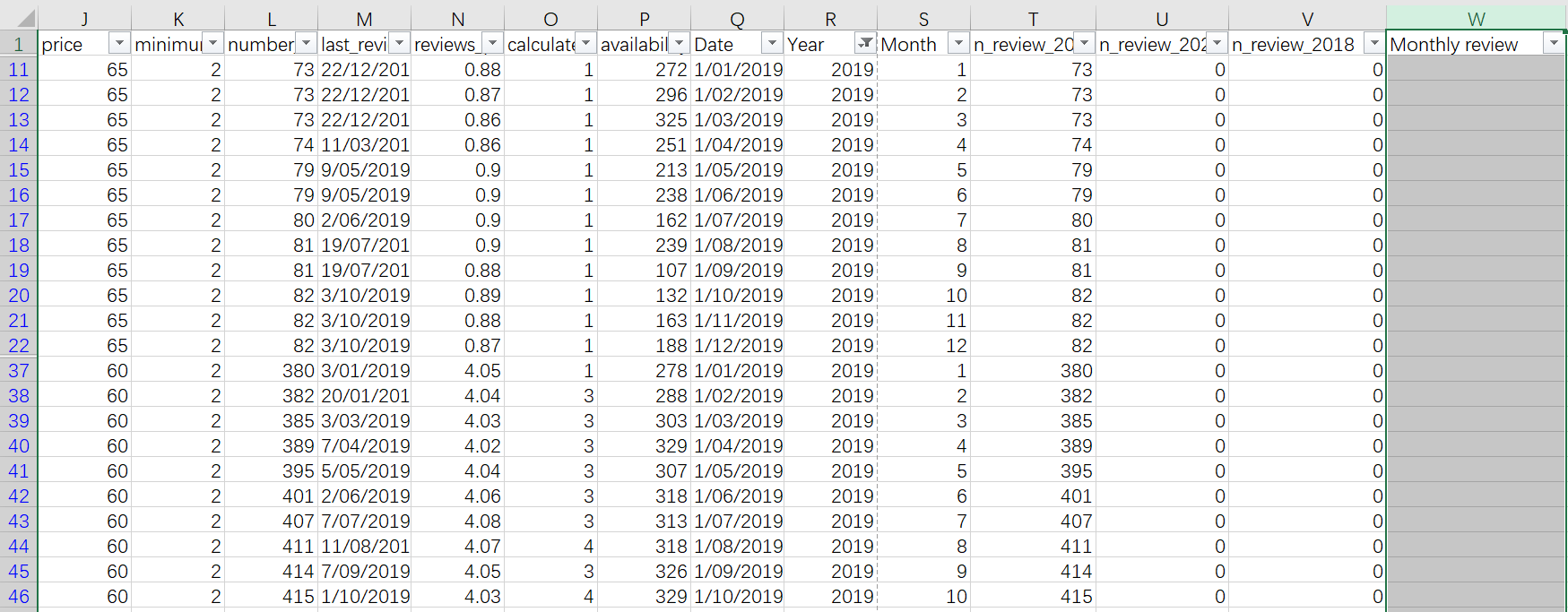
* For the three separate datasets, we repeat the following procedures:
  + Cleaning Empty Data and Data Misalignment: since we need to use function to calculate the difference of ***Number of Review*** in different year and months, we first ensure that all listings with empty or misaligned data in ***Number of Review*** are deleted by using filter:



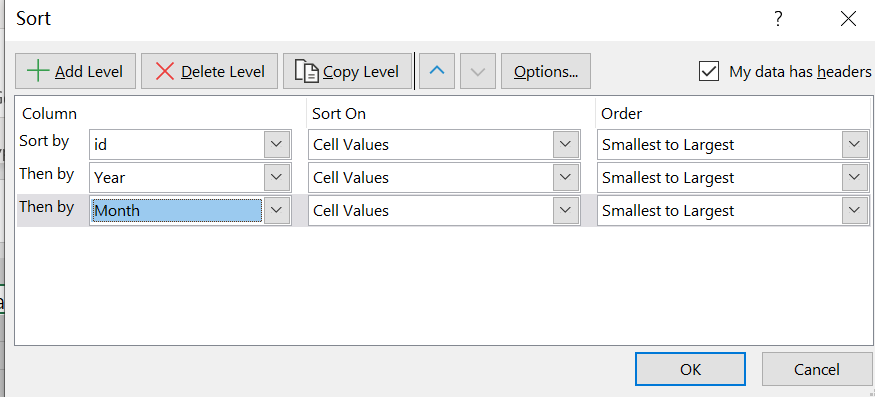
* + We created auxiliary attributes named ***number\_review\_2018***, ***number\_review\_2019***, ***number\_review\_2020***, whose value are directed copied from ***number\_of\_review***. For example, for a listing whose ***Year*** is 2019, its value in ***number\_review\_2019*** is directly copied from ***number\_of\_reivew***, while its value in ***number\_review\_2018*** and ***number\_review\_2020*** are 0. We created these auxiliary attributes, since while we merge all datasets and import the comprehensive one into Power Bi for visualization, we could avoid of filtering the value of ***number\_of\_view*** by year if, for example, we want to calculate the total number of views in a specific year.

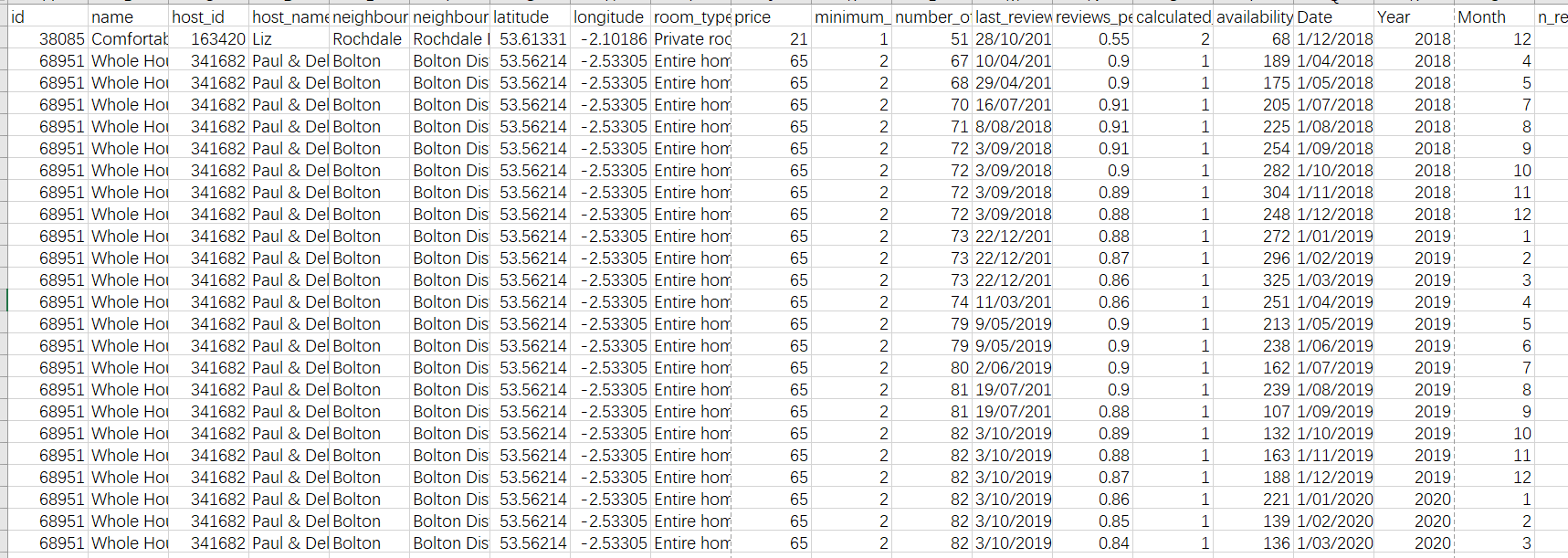


* + To show the monthly total review of any specific listings, we add the auxiliary attribute named as ***Monthly review***

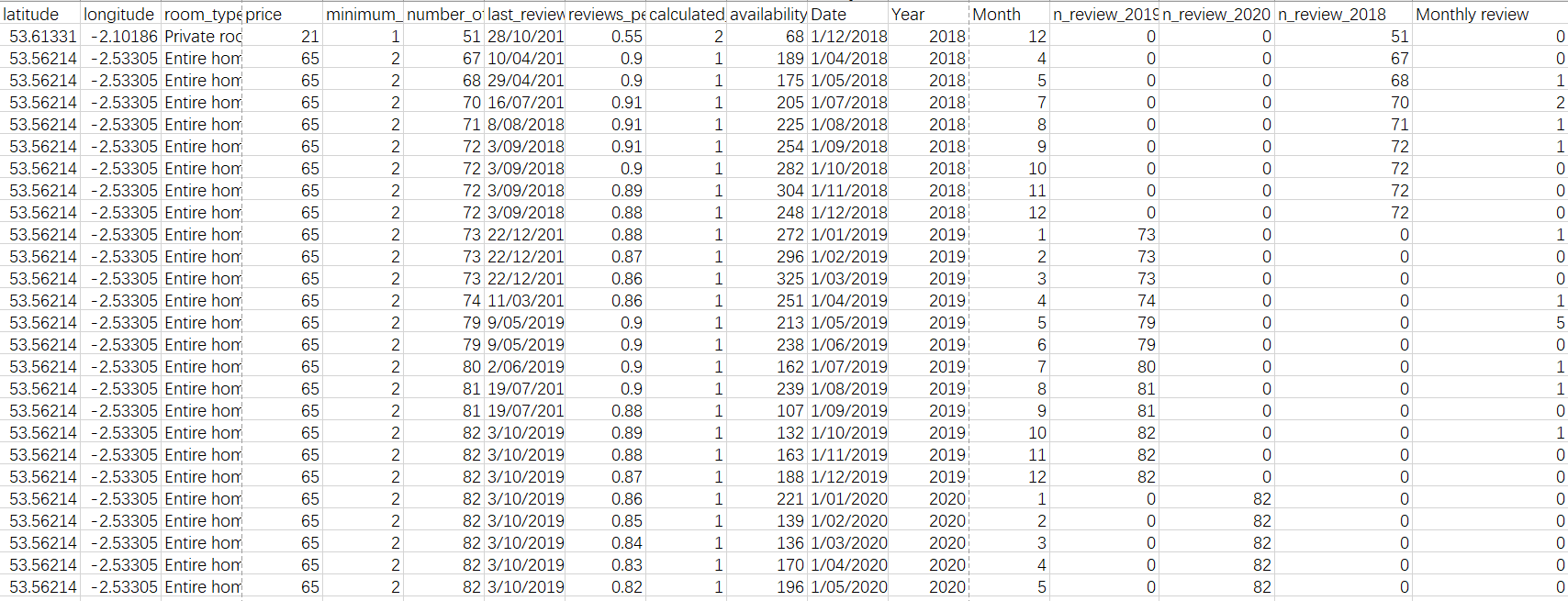


* + To calculate the value for attribute ***Monthly review***, we need to subtract the value of ***number of reviews*** by the same attribute in last year by following procedures:
    - Using sort function of Excel, so all listings is grouped by its unique id in the order of Year and Month

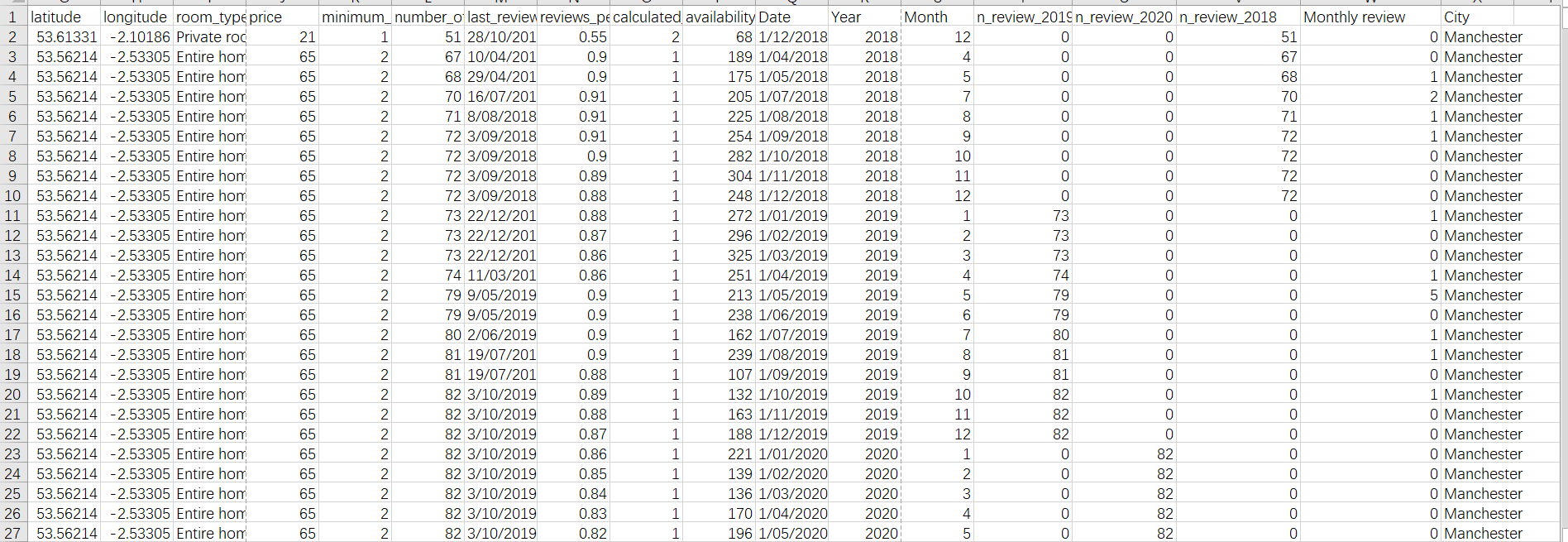




* + - Then, using function “=IF(A3=A2, L3-L2,0)” to calculate the monthly new review data for every listings (if the ***id*** of listing matches, set the value of ***Monthly review*** to be the difference of the value of new reviews between this month and the latest month; otherwise, set it to be 0)
      * We also need to set the ***Monthly review***ofthe first month to be 0 manually



* + Eventually, we created another auxiliary attribute to indicate that whether the listing is in Greater Manchester or London



* The data cleaning procedures in excel is finished. We then used the same python program to merge the three datasets into a comprehensive one and import it into Power Bi. In power Bi, we used the built-in filter to clean all blank cells.

