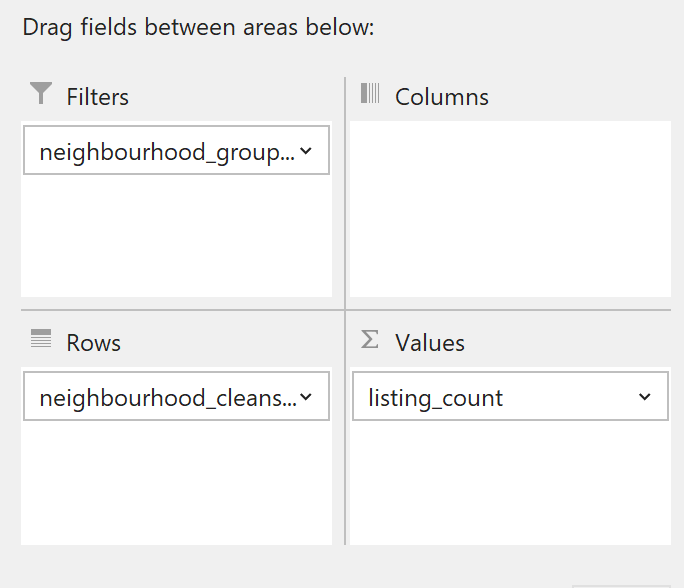
Cleaning steps:

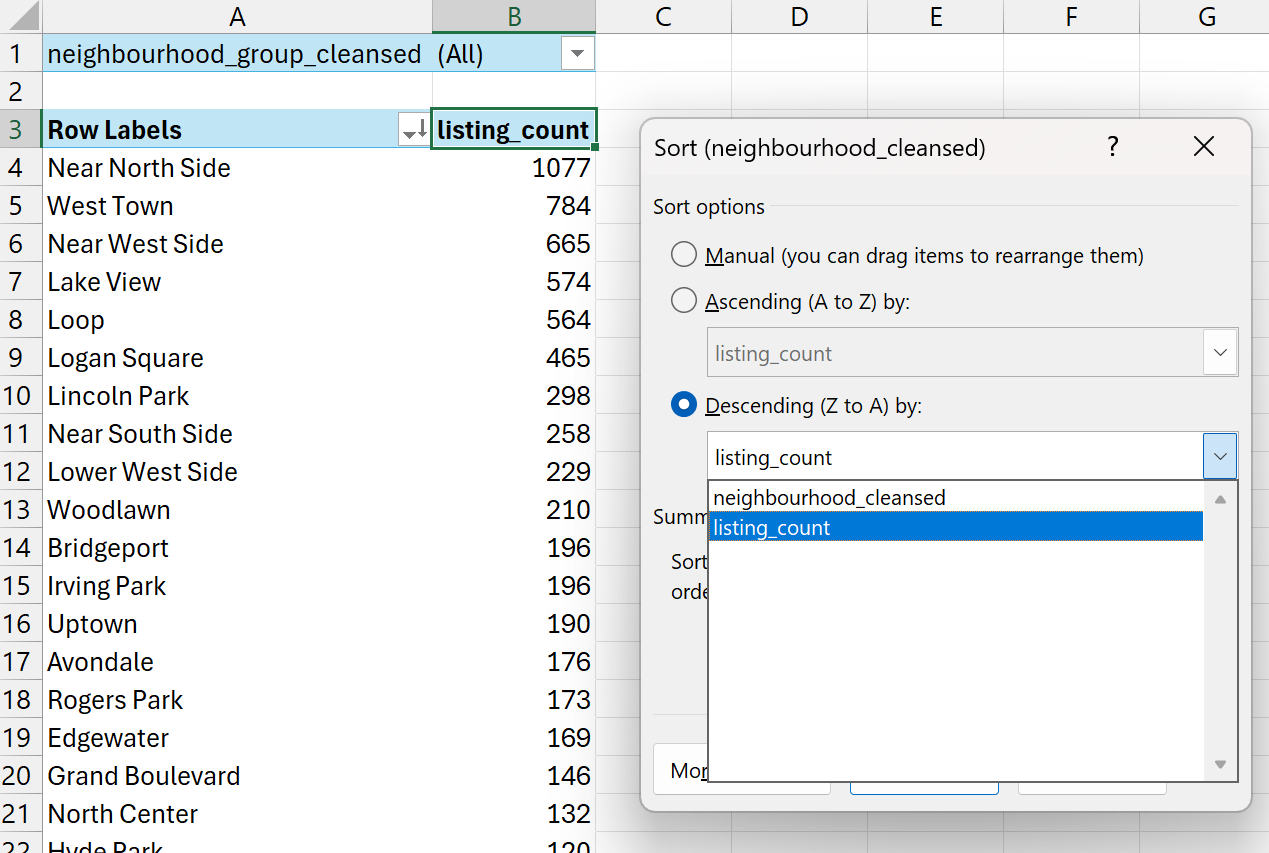
1. Open excel
2. Go to data tab
3. Get data as csv file
4. Transform data, do not import first
5. Check data over all
6. Confirm changes
7. Make a raws sheet to reference, never use it
8. Check which columns are to be used to answer question 1

|  |  |  |
| --- | --- | --- |
| Purpose | Best-fit column | Why? |
| Neighbourhood flag | neighbourhood\_cleansed | Already “scrubbed” by Airbnb—spelling is standardized, so counts won’t split between “Juffair”, “Juffair ”, “Al-Juffair” etc. |
| Alt/ sanity check | neighbourhood & host\_neighbourhood | Raw text the host typed. Good to eyeball for blanks or odd spellings, but don’t trust for final tally. |
| Higher-level grouping | neighbourhood\_group  \_cleansed | If you later want to roll a borough-level view (e.g., “Capital Governorate vs Muharraq”), use this. |
| Unique listing key | id (or listing\_url) | Lets you COUNT distinct listings per neighborhood and be |

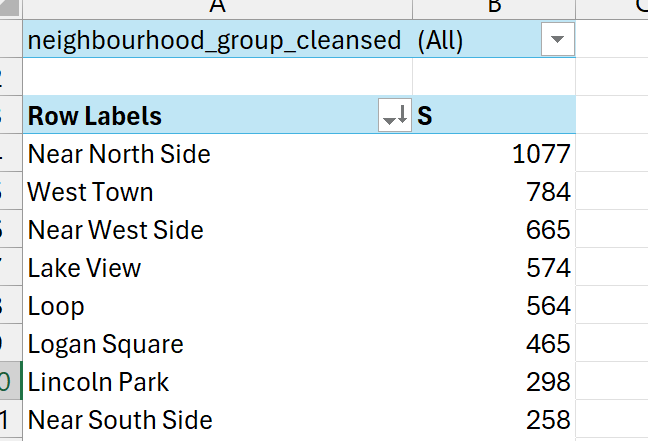
1. Run pivot with these only



1. “Count of id” was renamed “listing\_count”
2. Sort largest to smallest by count



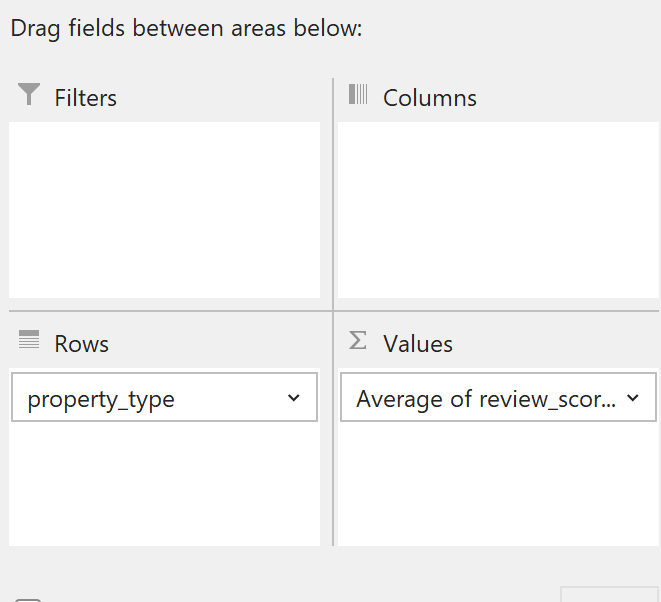
1. Final answer for question 1 is here, 1077, that’s west town



1. Check which columns are to be used to answer question 2

|  |  |  |
| --- | --- | --- |
| Purpose | Best-fit column | Why? |
| Filter for blanks → they should be rare. | property\_type | Tells us whether it’s an *Apartment*, *Villa*, *Tiny house*, etc. |
| Make sure it’s numeric; min should be 1, max 5. - Check how many rows are NA. | review\_scores\_rating | Overall rating (1 – 5). This is the “positivity” signal. |
| Look for outliers (e.g., 0 or 1000+). | *(optional weight)* number\_of\_reviews | Lets you weight ratings so a place with 300 reviews counts more than one with 2. |

1. Why are we taking the review score rating only, without considering the remaining scores
2. Its a roll-up metric
3. The most complete one, the remaining ratings are missing some data more than this one
4. Makes my life easier with the pivot table
5. Doing this will remove the outliers issue we had and will keep the highest score as 5 only, no more than 5, like what we had
6. Now we filter the column for blanks, and remove them. We wont use maiden or mean here, cuz it ratings matter only when they are made by users
7. Make a new pivot table, based on the newly cleaned data

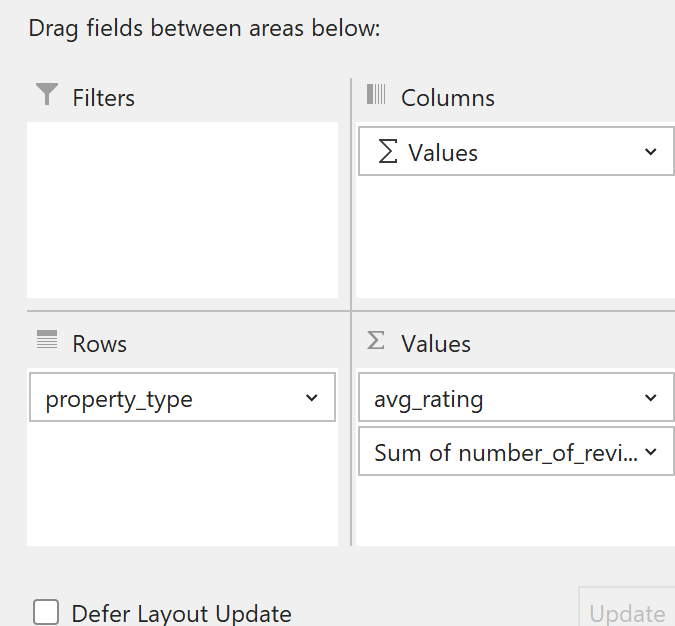


1. Renamed the following column

A screenshot of a computer

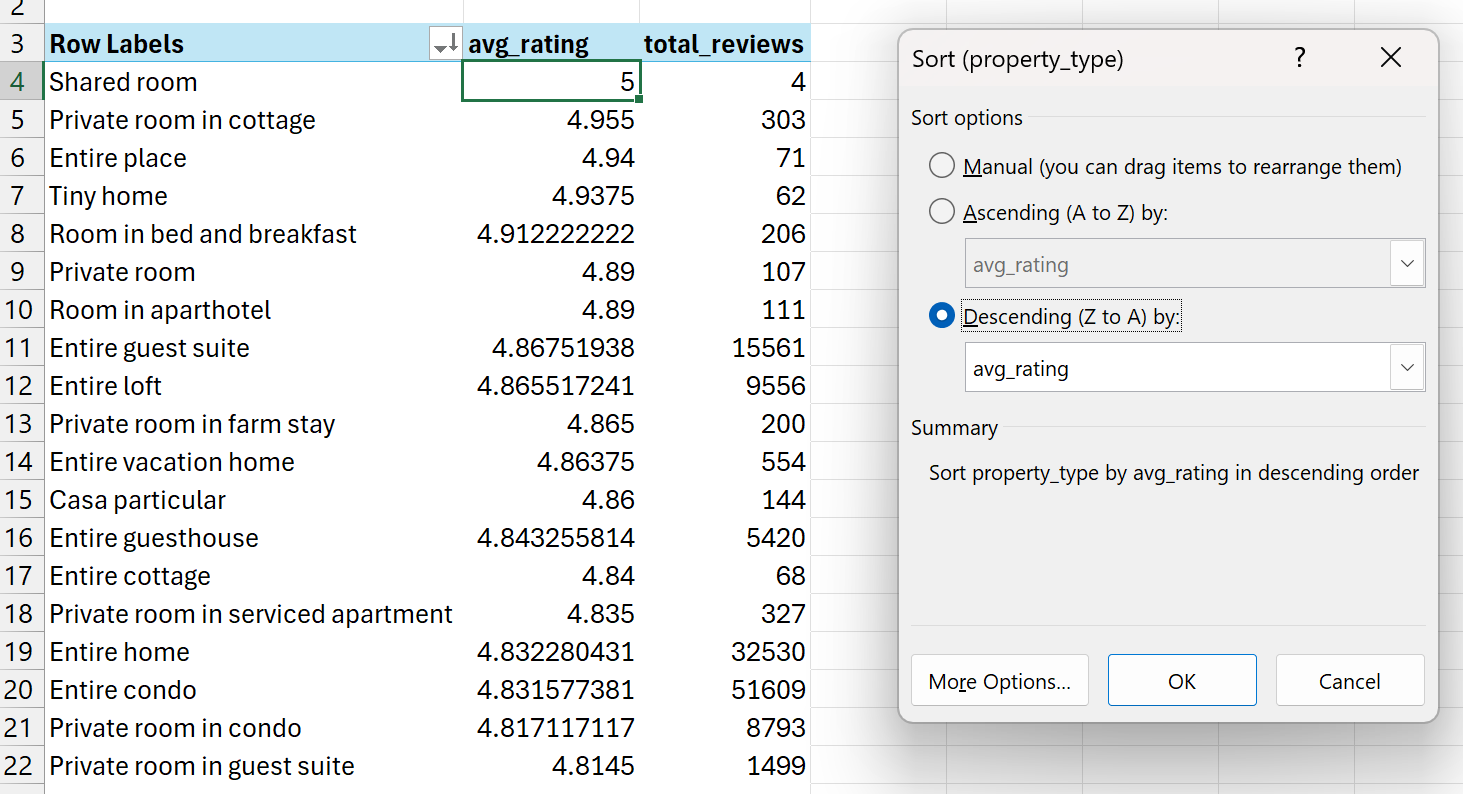
AI-generated content may be incorrect.

1. Added one extra field which is not needed, you can skip this, i also renamed it

A screenshot of a computer

AI-generated content may be incorrect.

1. Now sort by average rating as follow

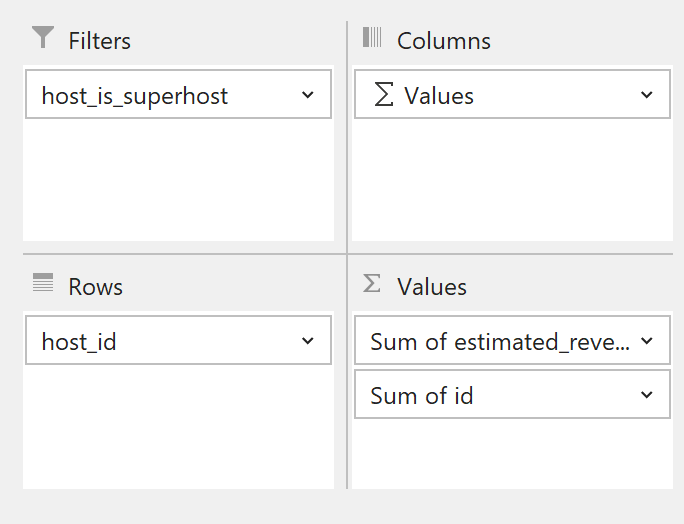


As you can see, there is no outlier issue, similar to the one we had last time

1. Renamed the sheet to rating\_leaders
2. This wraps up question 2
3. Starting question three with these columns

|  |  |
| --- | --- |
| column | why |
| Host\_id | Lets us roll-up all the listings that belong to the same host. |
| host\_is\_superhost | Easy flag for “successful” (Airbnb already did the badge work). |
| price | Nightly rate. Must be numeric—no dollar signs lurking. |
| availability | Days the listing is still open over the next year.So (365 − availability\_365) ≈ nights booked. |

1. First we will check these four columns for blanks or text formatted
2. Host\_id checks out, host\_is\_superhost does not, price does not (blanks), availability\_365 is the one im usin, there are no blanks
3. I will delete price and availability blanks first, filter then delete the empty rows on these cells, keeping the empty cells will break the math, there is no filling here, just delete
4. created a new column named estimated\_revenue as a currency, the formula is =price \* (365 - availability\_365)
5. we create a new pivot with the new data, add the following data



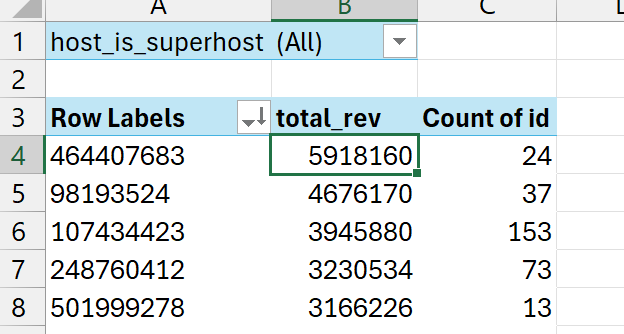
1. next im just renaming columns

A screenshot of a computer

AI-generated content may be incorrect. A screenshot of a computer

AI-generated content may be incorrect.

1. then sort highest-earning superhost by projected annual revenue



1. deleted any column that is not be used
2. next step, only if needed is to check the remaining data for any corruption, I did not do this
3. added a neighbourhood filter, west town, to both sheets rating\_leaders and money\_makers
4. xlookup was used to retrieve the minimum nights allowed to stay from the row data
5. this value was used to calculate a new total revenue, mr sayed did not like the old one, a new one was made to satisfy him
6. original formula was =[@price]\*(365-[@[availability\_365]]), mr sayed formula is =[@price]\*[@[number\_of\_reviews]]\*[@[stayed night]]
7. generated chart for all three pivot tables on separate sheets