**Introduction of Dataset**

I have download this dataset from Airbnb. It includes the Amsterdam ’s apartments.

**About the tables’ data:**

1. Listings:

* **id**
* **name** – Name of the apartment
* **host\_id** – ID of the owner
* **host\_name** – Name of the owner
* **neighbourhood\_group** – Null Data
* **neighbourhood** – Location of the apartment
* **latitude**
* **longitude**
* **room\_type**
* **price** –/night
* **minimum\_nights**
* **number\_of\_reviews**
* **last\_review**
* **reviews\_per\_month**
* **calculated\_host\_listings\_count**
* **availability\_365** – Availability of the year

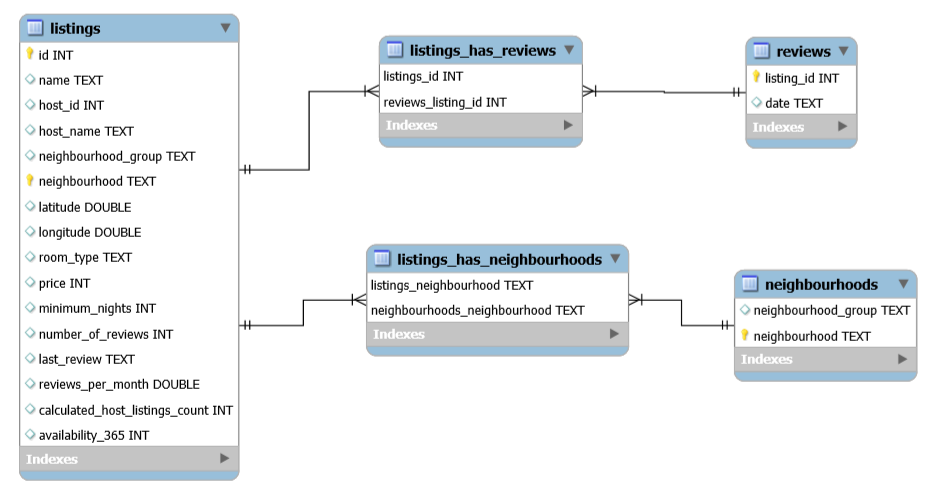
1. Reviews:

* **listing\_id**
* **date** – Date of the reviews

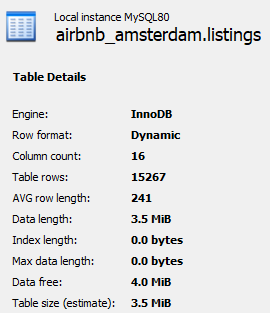
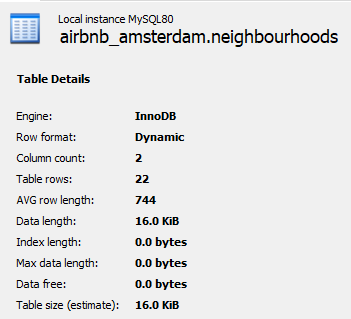
1. Neighbourhoods:

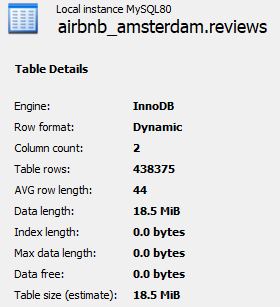
* **neighbourhood\_group**
* **neighbourhood** – Location of the apartment

**Creating tables DB Schema / ER diagram:**



**Shape of tables:**

****

****

**Tasks:**

**What type of ’minimum nights’ were prefered based on the number of the reviews?**

**Code:**

SELECT

l.minimum\_nights, COUNT(r.date) AS number\_of\_dates

FROM

listings AS l

JOIN

reviews AS r ON r.listing\_id = l.id

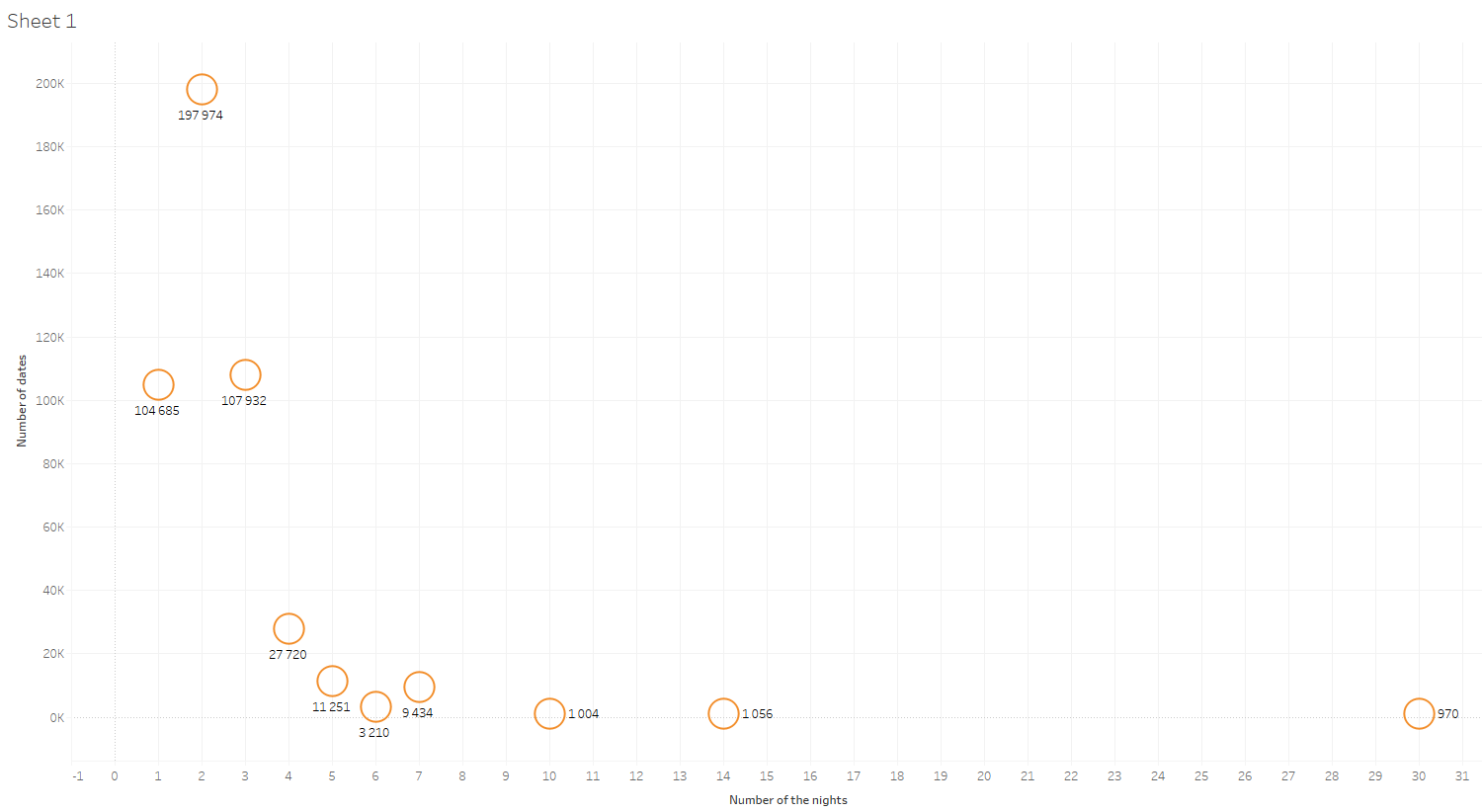
GROUP BY l.minimum\_nights

ORDER BY COUNT(r.date) DESC

LIMIT 10;

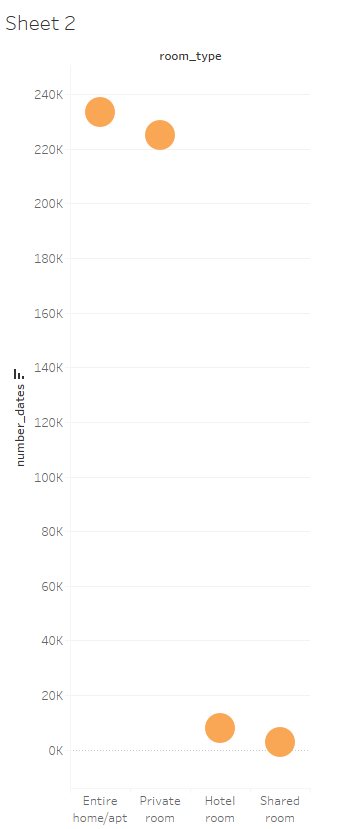
**Result:**

|  |  |  |
| --- | --- | --- |
| **Row number** | **minimum\_nights** | **number\_of\_dates** |
| 1 | 2 | 197974 |
| 2 | 3 | 107932 |
| 3 | 1 | 104685 |
| 4 | 4 | 27720 |
| 5 | 5 | 11251 |
| 6 | 7 | 9434 |
| 7 | 6 | 3210 |
| 8 | 14 | 1056 |
| 9 | 10 | 1004 |
| 10 | 30 | 970 |



The most prefered minimum nights is the 2 nights section.

**What kind of ’room type’ was the most visited one?**

**Code:**

SELECT

l.room\_type, COUNT(r.date) AS number\_of\_dates

FROM

listings AS l

JOIN

reviews AS r ON r.listing\_id = l.id

GROUP BY l.room\_type

ORDER BY COUNT(r.date) DESC

LIMIT 10;

**Result:**

|  |  |  |
| --- | --- | --- |
| **Row number** | **room\_type** | **number\_of\_dates** |
| 1 | Entire home/apt | 233631 |
| 2 | Private room | 224881 |
| 3 | Hotel room | 7846 |
| 4 | Shared room | 2923 |

The most visited room type was the „Entire home/apt”.

**Avarage price based on the neighborhoods (with number of the minimum nights).**

**Code:**

SELECT

neighbourhood,

AVG(minimum\_nights) AS avarage\_minimum\_nights,

AVG(price) AS avg\_price,

AVG(price) \* AVG(minimum\_nights) AS price\_w\_nights

FROM

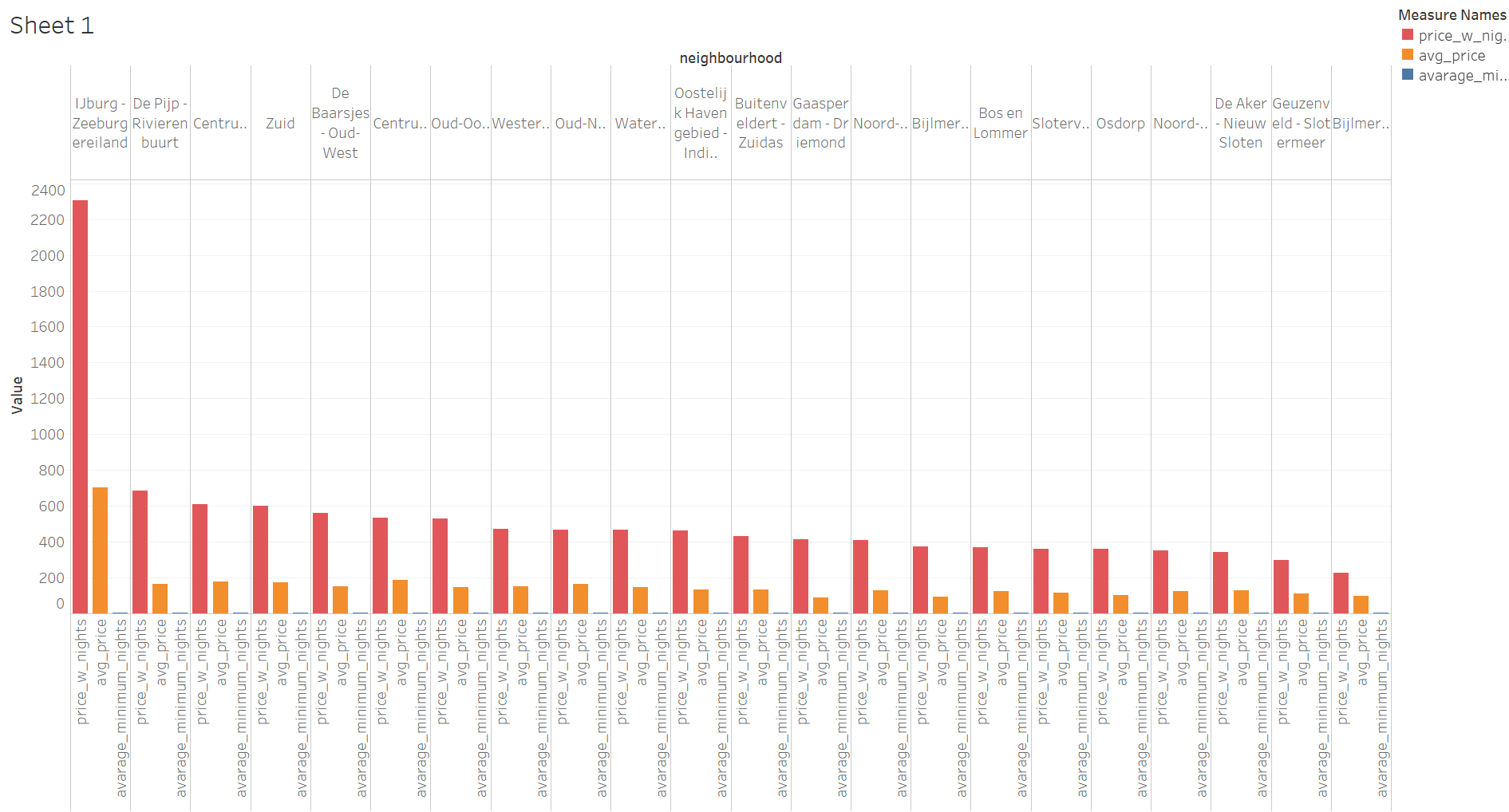
listings

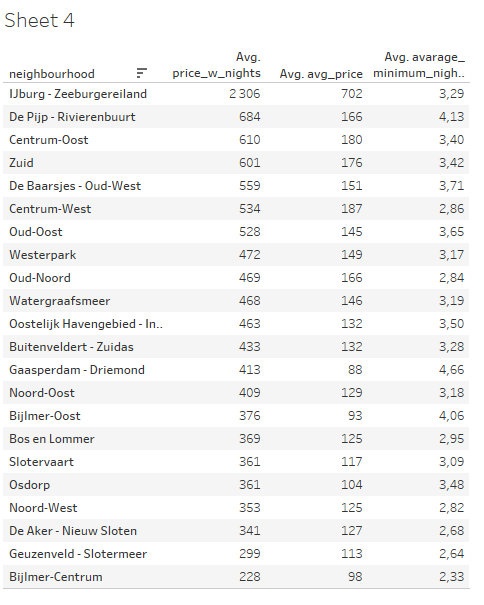
GROUP BY neighbourhood

ORDER BY AVG(price) \* AVG(minimum\_nights) DESC

LIMIT 100;

**Result:**





Crystal-clean visible that, the „minimum of the nights” affect the final output, so we had better examine the avarage of the min. nights based on the neighborhood!

As we can see, the number of min. nights can affect the final price based on the neighborhood.

**What is the most visited neighborhoods?**

**Code:**

SELECT

l.neighbourhood, COUNT(l.neighbourhood)

FROM

listings AS l

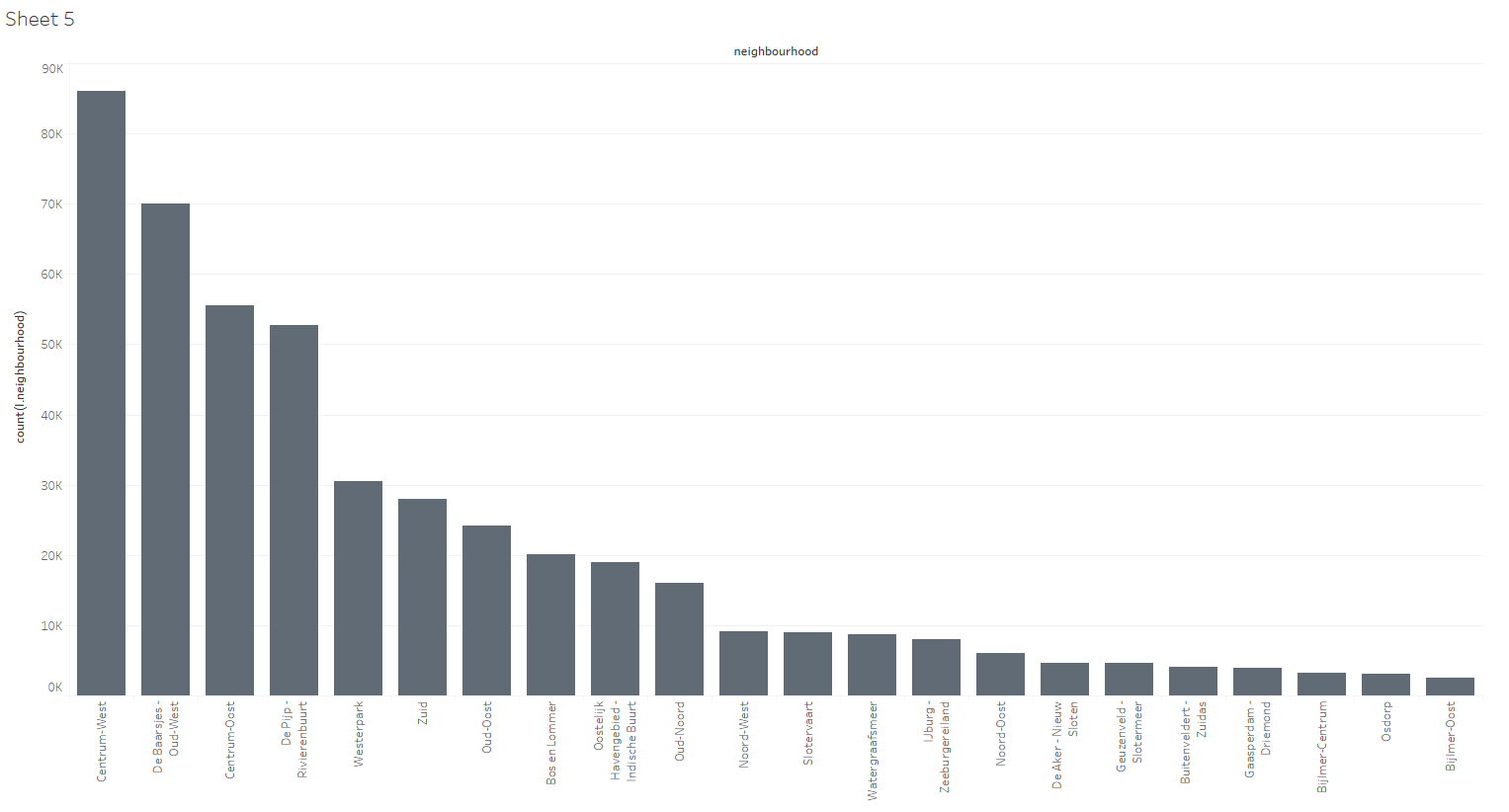
JOIN

reviews AS r ON l.id = r.listing\_id

GROUP BY l.neighbourhood

ORDER BY COUNT(l.neighbourhood) DESC;

**Result:**



**Who does have more than 20 apartments? People name and location of apartments.**

**Code:**

SELECT

host\_id, host\_name, COUNT(host\_id) as number\_of\_apartments

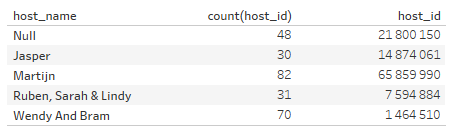
FROM

listings

GROUP BY host\_id

HAVING number\_of\_apartments > 20

ORDER BY number\_of\_apartments DESC;



**Code:**

SELECT

host\_id,

host\_name,

latitude,

longitude,

room\_type,

price,

minimum\_nights

FROM

listings

WHERE

host\_id IN (SELECT

host\_id

FROM

listings

GROUP BY host\_id

HAVING COUNT(host\_id) > 20);

