Airbnb Data Analysis Project

Félix Raphael

October 2023

felix-raphael@outlook.com

Table of Contents

1. Introduction

1.1. Airbnb as a sharing economy and its effects on gentrification

Airbnb has become very popular for its quality, affordability and wide range of property offerings all around the world, all while allowing tourists a different experience to that found in a hotel (Barker, 2020). Also, not only users on the demand side of the platform benefit from it, but it turns out that home providers enjoy and make great use of the service provided by Airbnb as well. In other words, turning their rooms and properties into short-term rentals, hereafter referred to as STR, provides them with a substantial source of revenue. In 2021, the average Airbnb Host in the European Union made 3,225$. Furthermore, many people seem to become hosts because of the cost-of-living crisis. Thereby, over 40% of hosts would use the additional income from the platform to cover the rising cost for food and other essentials (Airbnb, 2022).

On the flipside, the rise of sharing economy platforms such as Airbnb comes with serious negative implications for various stakeholders. More specifically, the so-called “Airbnb-effect” has grown into a significant cause for concerns with regard to housing stock, prices and communities that likely exceed the benefits to travellers and property owners (Barker, 2020). In other words, the detrimental impact on housing stock consists in encouraging landlords to switch from long-term rentals and sale-market to STR, thereby lowering the housing opportunities for long-term residents (Barker, 2020). The so-called “Airbnb effect” has thereby shown to be somewhat similar to gentrification in that it slowly increases the value of an area to the detriment of the indigenous residents, many of whom are pushed out from urban areas due to financial constraints (Barker, 2020; Rabiei-Dastjerdi et al., 2022). As a consequence, residents of various cities have expressed their concerns about the platform in that it leads to entire buildings being used for the purpose of short-term renting, consequently changing the character of areas and neighbourhoods as well as having further implications on e.g., demographics, schools and local stores (Bosma & van Doorn, 2022; RFI, 2023).

A negative implication of this home-sharing economy is the professionalization of Airbnb’s hosts, which is driving more revenue to a narrower segment of hosts (Deboosere et al., 2019). In fact, Airbnb even encourages its hosts to upgrade their listings and accommodate more large-scale forms of hosting, equipping hosts that seek to close rent-gaps with professionalization programs and tools (Bosma & van Doorn, 2022). Further, recent findings (Bosma & van Doorn, 2022; Deboosere et al., 2019) suggests that a higher degree of professionalization in the context of Airbnb also leads to higher revenues for the respective hosts. The underlying rent-gap theory, which is defined as the gap between current rental income and potential rental income of a property, can be seen as a primary driver of this form

of gentrification and acts as an incentive for landlords to gain more profit out of their properties by turning their residences into STR, oftentimes in places with already dramatic housing shortages and skyrocketing rents (Bosma & van Doorn, 2022). As a result, the dramatic increase in short-term and decrease in long-term rentals has shown adverse effects on neighbourhoods and housing markets, such as increasing housing and rent prices (Rabiei-Dastjerdi et al., 2022).

1.2. Airbnb in Paris

During the past years, especially bigger, more popular cities struggled with the effects of home-sharing platforms driving over-tourism and its consequences of gentrification. As Paris has been strongly affected by those negative implications of short-term rentals, the city has been on the forefront of efforts to limit the effect of Airbnb on the rental market (Short Term Rentals, 2021). In that regard, the mayor of Paris has expressed her annoyance with Parisian residents treating home-sharing like a business (Short Term Rentals, 2021). Additionally, the deputy to the mayor Ian Brossat has made it clear that the city's objective was to preserve accommodations and the way of life of the residential areas (RFI, 2023). Moreover, alongside 21 other European cities, the city of Paris has urged the EU competition commissioner to advocate for the establishment of a comprehensive EU-wide regulatory framework for short-term rentals, replacing the current system where individual cities enforce their own regulations. This clearly illustrates the aim of French authorities to combat the long-term rental housing shortage and work on proportionate regulation that puts local families and communities first and works for all (BBC News, 2020). In terms of regulation, currently only Parisian main residences are allowed to be rented out as furnished tourist accommodation provided that they are declared to the town hall. Also, such accommodations can only be rented out for up to 120 days per year. According to Radio France Internationale, "Paris city hall has raked in 6.5 million euros in fine which have been issued by the courts against Parisian landlords who have failed to comply with its regulations on seasonal lets – mainly through the Airbnb platform" (RFI, 2023).

1.3. Gentrification and neighbourhood change through Airbnb in Paris

Considering the drastic measures that the city of Paris has taken, it is worth examining the Parisian STR market to see if notable implications with regard to gentrification can be deduced from it. In doing so, I will primarily make use of publicly available data from Airbnb for the urban area of Paris. In analyzing this data, I want to address the negative repercussions of Airbnb by comparing the density of Airbnb listings in different neighbourhoods and further expand on the degree of professionalization of Parisian hosts and explore potential differences between

individual and professional hosts regarding pricing and listing type. More specifically, I aim to touch upon the following research questions:

I) Which neighbourhoods display the highest listing density?

In order to answer this question, I will display the absolute number of listings per neighbourhood and create a density measure that I will calculate according to the following formula:

$$listing\ density_i = \frac{number\ of\ listings_i}{\frac{population_i}{1000}}$$

For the denominator, I will use population data of each neighbourhood provided by statista with ISEE as an original source. Similar to Gant (2016), who builds a measure with the number of households in a neighbourhood, I will illustrate the density of Airbnb listings in the different neighbourhoods i of Paris, with the difference that I will relate the listing number to the population size of the neighbourhood. (Short Term Rentals, 2021) suggests. I will build on the finding that more Airbnb listings accelerate gentrification through increasing rent and house prices by identifying the neighbourhoods with the highest listing density as a proxy for a high gentrification (Barron et al., 2020).

II) Professionalization

a) What is the distribution of the number of listings per host in Paris overall?

b) Which neighbourhoods are most popular among professional hosts?

After displaying the average number of listings per host for each neighbourhood, I will investigate the degree to which hosts in different neighbourhoods are professionalized. I will base my analysis on the work of Bosma & van Doorn (2022) and Abrate et al. (2022), which define professional hosts as having two or more listings. I will expand on this definition according to Deboosere et al.'s (2019) clustering approach as a measure for the degree of professionalization, grouping hosts into categories of single listing, 2-10 listings and more than 10 listings, while neglecting listings that hosts may have outside of Paris. By answering this question, I will touch upon the annoyance of Parisian city representatives towards Airbnb as a business model. Furthermore, I will once again investigate which neighbourhoods are affected the most by gentrification, since gentrification is driven by the professionalization of Airbnb hosts. In other words, having a high concentration of professionalized hosts puts a neighbourhood at risk of gentrification, since property prices on the sales market increase with

its STR revenue, which is generally shown to be higher for professional hosts (Bosma & van Doorn, 2022).

III) Host type characteristics

a) How do hosts of different types of professionalization set their prices?

b) How does professionalization relate to the type of listing?

According to Abrate et al. (2022), hosts with high professionalization set lower prices for their listings than individual hosts do (hosts with only one listing). I will examine this claim on the Parisian Airbnb dataset and further analyse how the degree of professionalization relates to the type of listing and thereby examine the claim from (Bosma & van Doorn, 2022), stating that professional hosts are characterized by listing rather entire homes than rooms or shared rooms. Answering this question serves the purpose of understanding what characterizes highly professionalized hosts and to test the claims made about their pricing strategies as well as listing types.

2. Entity Relationship Diagram

2.1. Design und ERD choices

To serve my problem statement, not many attributes from the raw Airbnb dataset are required. In that, I created an entity “location” that includes the Primary Key (PK) “neighbourhood_cleansed” (in ERD already renamed “neighbourhood”) and the attribute “population”, which refers to additional data drawn from statista. The initial source of this population data set is a report from the National Institute of Statistics and Economic Studies of France and was published in December 2020, hence containing data for 2020. The population data’s neighbourhood segmentation exactly matches that of the Airbnb dataset, which makes it an asset to this analysis.

In addition, the second table “host” entails the PK “host_id” and the column “calculated_host_listings_count”, which displays the number of listings for each host in the city of Paris. This column will later come in handy when clustering the hosts with respect to their level of professionalization (Question 3). For Question 2, in which the number of listings per host is required on a neighbourhood level, this variable will, however, not provide any help, since the counted listings are not available per neighbourhood. Therefore, this question will have to be answered by a separate counting operation on listings, performed in SQL.

Furthermore, I created a “listings” entity in which the “listing_id” attribute (initially called “id” in raw data) as a PK because it serves as a unique identifier for every variable describing an accommodation in the dataset.

Moreover, the table contains “neighbourhood” as a Foreign Key (FK), to form the relation to the “location” table and thereby enable matching the population data for each neighbourhood, in which the listings are located.
What’s more, the “listings” table encompasses the attributes “host_id” as a FK, to identify hosts and refer to their total listings in Paris represented in the table “host”, “price” and “room_type”. The variables “price” and “room_type” will provide data to answer question 3.

These tables and attributes are fully sufficient to answer all formulated questions in the aim of addressing the overarching problem statement. Therefore, further data stemming either from the listings dataset, such as host or property related data, or from the calendar or reviews dataset, are not necessary and hence will be ignored, since they don’t contribute to this analysis with regard to the questions that were presented.

In terms of data privacy, no personal data beyond the identification number for hosts were included since it does not provide additional benefit to this analysis. To address the privacy issue with “host_id”, the id’s will be replaced by randomized numbers, thereby assigning each host a new, non-traceable, identifier. This will be done in the “Cleaning”-Part of this project.

2.2. Relationship description

As for relationships of the ERD, the relationship between hosts and listings can be described by a one-to-many relationship, in the sense that one host can have one or multiple listings, whereas one or many listings can be associated with only one host. Also, for a host to be in the dataset at all, it can be assumed that a host must have at least one listing (therefore one-to-many) and that every listing can be linked to its host, since there are no missing values for the “host_id” variable.

Furthermore, the relationship between listings and location can be described by a one-to-many relationship, because one listing is assigned to one and only one location (“neighbourhood”) and one neighbourhood is expected to entail at least one listing for it to be listed in the dataset.

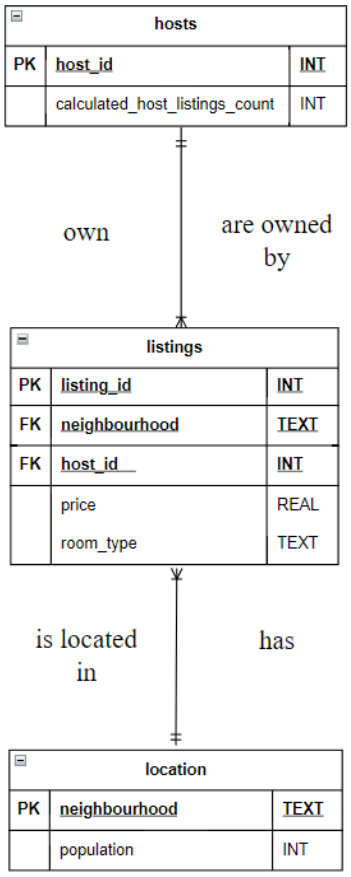

Figure 1: Physical Model of the Entity Relationship Diagram

2.3. Normalization

In terms of normalization, the suggested Entity-Relationship-Diagram (ERD) in Figure 1 fulfils the 3rd normal form in that there are no atomic values in attributes, there is no composite key and no transitive dependencies, such that attributes are only functionally dependent on key attributes of their respective entity.

Initially, the raw dataset from Inside Airbnb was not normalized at all, because the first normal form (1NF) was already violated by allowing non-atomic values, such as in column “host_location”, which could be split in “host_city” and “host_country”, or also in column amenities, which contains all the available amenities as a string in only one cell for each listing, listed as a string. By only considering the variables relevant to our analysis in one single table, without splitting the table (see Table 1), our table would only be in 2NF, because there would be transitive dependencies. In other words, in table 1, the attribute “population” is only dependent on the non-key attribute “neighbourhood” and the attribute "calculated_host_listings_count” is dependent on the non-key attribute “host_id”. In the 3rd NF,

however, all non-key attributes should be dependent on the Primary Key only, which is not the case in the table below.

	listing_id	host_id	neighbourhood	price	room_type	calculated_host_listings_count	population
0	9359	40742	Louvre	75.0	Entire home/apt	1	16149
1	167998	36737	Louvre	443.0	Entire home/apt	1	16149
2	48498	37361	Louvre	221.0	Entire home/apt	81	16149
3	2821377	37361	Louvre	198.0	Entire home/apt	81	16149
4	26241307	37361	Louvre	230.0	Entire home/apt	81	16149
...	...	...	...	...	...	...	...
60761	906974779120185085	28107	Ménilmontant	125.0	Entire home/apt	1	193044
60762	906989429815879202	20929	Ménilmontant	544.0	Entire home/apt	1	193044
60763	907008252086007569	10048	Ménilmontant	89.0	Entire home/apt	1	193044
60764	907716758564519280	41774	Ménilmontant	88.0	Entire home/apt	1	193044
60765	907924863916056172	32607	Ménilmontant	190.0	Entire home/apt	1	193044

Table 1: Relevant variables in 2NF table

The idea behind organizing the ERD in 3rd NF is that, in general, higher degree of normalization makes a database more efficient, improves query performance, increases data integrity and scalability, reduces data redundancy and provides users with a greater ease of maintenance, dealing with smaller, well-structured tables. If, for example, the entities “location” and “hosts” were not created, this would lead to high data redundancy in columns “population” and “calculated_host_listings_count”. More specifically, for each repetition of a unique “host_id” in the table, data for the variable “calculated_host_listings_count” would be redundant and for each repetition of unique neighbourhoods in the table, the column “population” would be redundant. In 3rd NF, however, the number of listings will be listed only once for each unique host in the “hosts” entity.

3. Data Cleaning

In order to further work with a clean set of data, this section will focus on the methods adopted to clean the raw data set provided by Inside Airbnb. The data cleaning was performed in Python. More specifically, the raw data was imported as a csv into Jupyter Notebooks and the resulting cleaned dataframe exported again as a csv-file for further upload into DB Browser for database implementation, following in section 4. The following subsections will serve as guidance through the different steps of the cleaning process.

3.1. Rename columns

First, three different columns were renamed. The variable “id” was changed to “listing_id” for less ambiguous wording. Next, the initial “neighbourhood” variable from the raw data set was replaced by “neighbourhood2”, since the “neighbourhood_cleansed” variable is the one used in this analysis. Subsequently, the “neighbourhood_cleansed” column was renamed “neighbourhood”. The renaming of columns can be traced back in table 2 below.

	initial_variable	new_variable
0	id	listing_id
1	neighbourhood	neighbourhood2
2	neighbourhood_cleansed	neighbourhood

Table 2: Rename columns

3.2. Clean the “price” variable

For the variable price to be useable in further data analysis, the price variable has to be cleaned such that the “$” sign is removed and the values of this variables are displayed as numerical values. More precisely, the data for the price variable was stored as a float64 data type. Other variables included in the ERD (Figure 1) already fulfilled the respective data type requirements for further processing.

3.3. Check outliers, null values and duplicates

3.3.1. Outliers

In terms of outliers, the variable “price” was investigated. Figure 3 displays the price distribution of the listings data. It shows that the prices range from 0€ to 999€ per night for an Airbnb in Paris, while the plot is strongly right skewed. To detect outliers, the Interquartile Range (IQR) method was performed. As a result, the detected outliers detected according to the IQR ranged from 387€ to 999€. Their distribution is shown in Figure 3. According to Figure 2 and Figure 3, there is no obvious reason to exclude any outliers from the analysis, since the right skewedness could simply be a result of specific, expensive neighbourhoods in Paris. Not considering prices above 387€ for the descriptive analysis context of this report would therefore highly flaw our findings.

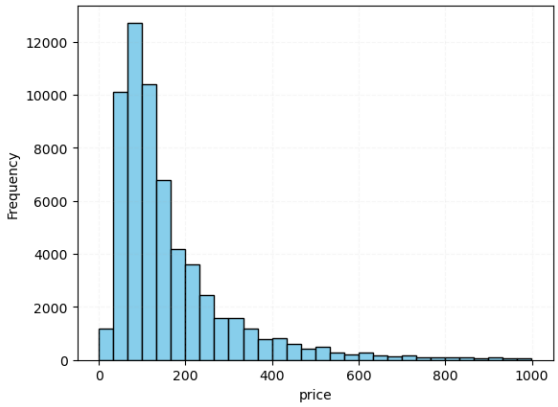

Figure 2: Price Distribution

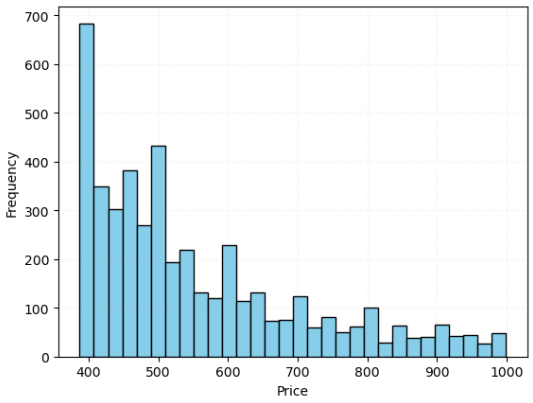

Figure 3: Distribution of Price Outliers

3.3.2. Null values and 0€ prices

The “price” variable is further source of missing values. All the other variables display complete data. The chosen approach to deal with the 914 missing values in the “price” column is conservative. Similarly, there are 26 price values of 0€, which will be treated the same way, since reporting a price of 0€ for a listing can be considered as having no value displayed at all. This is why, in the interest of providing representative results to the formulated research questions, rows in which the price is either missing or equal to 0, will be deleted from the sample. By doing so, one must take into account that, especially in answering questions unrelated to prices, deleting these rows weakens the reliability of the analysis. However, since there are only 940 rows affected by this, this appears like a reasonable approach to choose. The alternatives such as imputation of mean or average values may be just as, or even more flawed, than the conservative approach of excluding these observations with regard to the quality and reliability of results. With the 0€ prices now excluded from the clean data set, the prices range from 8€ to 999€.

3.3.3. Duplicates

Since the variables “host_id”, “neighbourhood”, “price”, “room_type” and “calculated_host_listings_count” allow for duplicate values, the column “listing_id” is consequently the only variable that needs to be checked for duplicate values. As expected and required, the “listing_id” column does not contain duplicate values.

3.4. Categorical variables

To check the data quality of categorical variables, the columns “neighbourhood” and “room_type” were examined. The objective of this procedure is to find out if string values of categorical values may entail typing errors or other mistakes that must be cleaned prior to data analysis. For that, for both categorical variables the unique values were observed. In both

cases, there were no typing errors. Furthermore, it is important to check the right spelling of the “neighbourhood”, since the population table will later be joined on the “neighbourhood” column.

3.5. Data Privacy

In terms of data privacy, the data used should comply with certain privacy requirements. In that, since the chosen variables for the analysis also entail “host_id”, which is considered personal data according to European Regulations (GPDR), this variable has to be transformed, such that it is not possible to traced back the identification number of the specific host. More particularly, under the chosen methodology in doing so, a new dataframe was created in Python, in which the unique “host_id” values from the dataset are stored. Then, a second column was created to store the range of numbers from 1 to the number of unique hosts (45,710). At the same time, this range of numbers was being stored in a random order. This created the new “host_id”. As an additional check that the pseudonymization worked, I checked that the old id is never equal to the new id. Finally, the initial “host_id” column was replaced by the newly created, pseudonymized, “host_id” column inside the dataframe.

3.6. Cleaning the population data

As a last step to the cleaning process, the population data had to be inspected. For that, the columns “district” and “Population” were renamed to “neighbourhood” and “population”. The renaming of district is thereby very important with respect to the join operation that will follow later during the database implementation. Therefore, the column on which the table “population_data” is joined on, has to match the exact name of the column in the “listings” table.

Furthermore, the values of the population column are not in numerical form, which is why transformation is required at the point.

After cleaning the population data, exporting the resulting tables, being “data” (raw listings data set cleaned), “population_data” (population data set cleaned) and “matching_table” (with the old and new host id’s), as csv files, concludes this section of data cleaning.

4. Database Implementation

4.1. Create tables

With the both the listings data (now named “data”) and the population data (“population_data”) cleaned, the database can now be set up. For that, both data and “population_data” were uploaded in DB Browser as csv files. By doing so, both data sets were created as tables in DB Browser, in which the data types had to be checked for correctness, since DB Browser may have misinterpreted the data by default. In addition to that, “listing_id” and “neighbourhood” were chosen as PK for the tables “data” and “population_data” respectively. Further adjustments were not necessary in these tables, because they only serve as a basis to extract the cleaned data to then populate the entity tables from the ERD (Figure 1). Next, those three new tables according to the established ERD were created. In that, the tables “location” with attributes “neighbourhood” and “population” were created with the specification that “neighbourhood” as the primary key, may not be null and should be unique. Similarly, for creating the table “hosts”, the primary key “host_id” shall be specified as not null and unique as well. This is because PKs by their logic are not allowed to entail null values, as those are the attributes that are supposed to be described by other attributes in the respective tables. Finally, the table “listings” was created with its attributes listed in Figure 1, again specifying the PK “listing_id” as not null and unique, but also defining the FKs “host_id”, referencing the PK in table “host”, and “neighbourhood”, referencing the PK in table “location”, both also specified as not null but not unique, since their values are allowed to be repeating in this table. As a last step, the matching table was uploaded to DB Browser to retrace which pseudonymized host id was assigned to which initial host id.

4.2. Populate tables

Once the ERD tables were created, they have to be populated. For that, the order in which the tables are populated is crucial because the tables contain foreign keys. Therefore, first the tables “location” and “hosts” were populated, so that the third table “listings” can reference those tables when populated with regard to its foreign keys.

5. Findings

5.1. Query description

To answer the first question of displaying the listing density, I formulated two different options on how to approach this with SQL. Recall that listing density was defined as the number of listings in a neighbourhood divided by the population of that neighbourhood per 1000 residents. In the first approach (Q1 – Option 1), I achieved this by performing a SQL query that involved joining the "population_data" and "listings" tables using a LEFT JOIN on the "neighbourhood" column. The result was grouped by neighbourhood, and I used ROUND to ensure precision. Finally, I calculated the density by dividing the count of listings by population per 1000 and ordered the results in descending order of density. The second approach (Q1 – Option 2) involves creating a Common Table Expression (CTE) named "DensityRankedNeighbourhoods" to calculate the population, population per 1000 residents, and density for each neighbourhood. Then, in the main query, I selected the neighbourhood, population, density, and used DENSE_RANK() to rank the neighbourhoods based on density. This approach makes the ranking clearer and more concise compared to the previous option. Moreover, the second approach is better suited because it avoids redundant aggregation (i.e. through aggregation functions “ROUND” and “COUNT”), already displays a rank column and provides more flexibility, as one can easily change the ranking criteria for example.

To answer Question 2a, I aimed to analyze the distribution of the number of listings per host in Paris. I performed a SQL query that counted the number of hosts for each value of "calculated_host_listings_count" and ordered the results in ascending order of this count. This query provided a fragmented distribution of hosts, which was hard to interpret. To improve on readability, I clustered hosts into three categories as mentioned in section 1: individual hosts with 1 listing, semi-professional hosts with 2 to 10 listings, and professional hosts with more than 10 listings. I achieved this by running separate SQL queries for each category, counting the hosts that fall into these groups.

For Question 2b, I focused on finding the neighbourhoods where professional hosts, those with more than 10 listings, have their listings. To do this, I first created a view called "professional_hosts" to subset hosts with more than 10 listings. Then, I queried this view to count the number of professional listings in each neighbourhood, allowing me to identify which neighbourhoods are most attractive to professional hosts.

To address Question 3a, I calculated the average price for listings based on the professionalization level of hosts. I ran separate SQL queries to calculate the average price for listings belonging to individual hosts, semi-professional hosts and professional hosts. This

query was rather low in complexity. A more sophisticated or complex query, however, was not necessary since the question does not entail any further complexity.

For Question 3b, I examined the distribution of room types among hosts with different professionalization levels. I performed separate SQL queries for individual hosts, semi-professional hosts, and professional hosts. These queries counted the frequency of each room type within each host category, providing insights into the types of accommodations offered by hosts at different professionalization levels.

5.2. Discussion

5.2.1. Airbnb Density throughout neighbourhoods

As our assumption in section 1 built on the listing density of a neighbourhood as a proxy for gentrification, I identified four neighbourhoods, in which the listing density is especially high, namely “Bourse”, “Louvre”, “Temple” and “Hôtel-de-Ville”. In figure 5, the listing densities of all neighbourhoods are represented by the size of their respective bubble. As can be observed on this map, these high-density neighbourhoods, circled in red, represent the first four arrondissements of Paris, which are concentrated in the centre of the city. It appears that neighbourhoods are systematically less dense in listings as they move away from the city centre. This would, based on our assumption, be an indicator that the neighbourhoods that gather around the city centre are more at risk or more touched by gentrification through Airbnb. Please find an overview of the top 4 neighbourhoods by density in table 3 and a density comparison of all 20 Parisian neighbourhoods in figure 5.

rank	neighbourhood	population	density
1	Bourse	21277	102.27
2	Louvre	16149	85.08
3	Temple	33651	82.73
4	Hôtel-de-Ville	29326	69.56

Table 3: Top 4 neighbourhoods by listing density

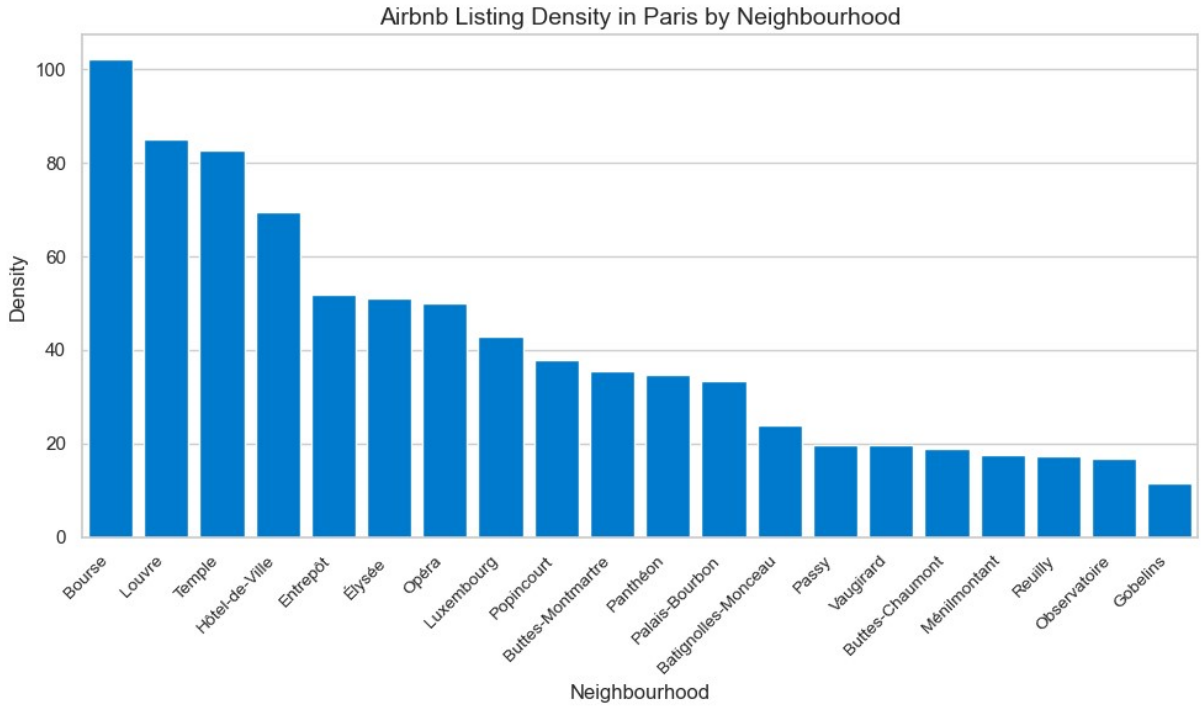

Figure 4: Comparison of Airbnb density by neighbourhood

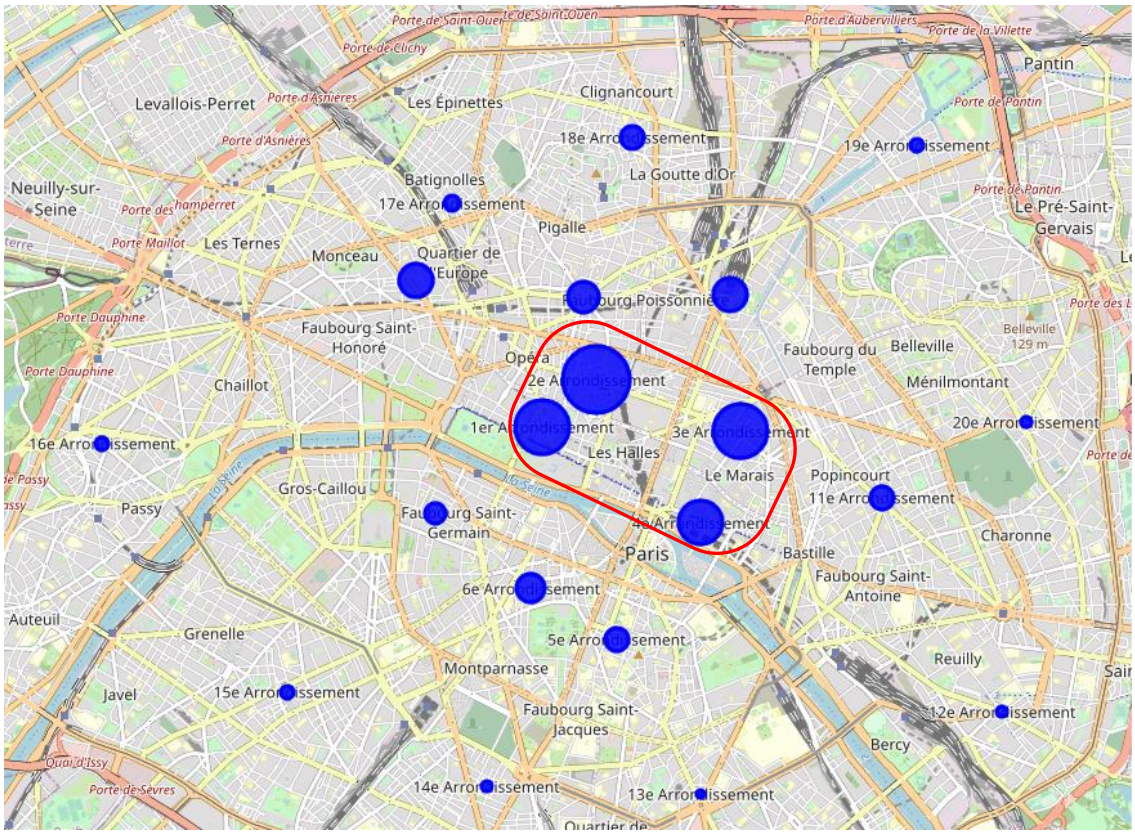

Figure 5: Airbnb density by neighbourhood

5.2.2. Professionalization distribution & popular neighbourhoods for Airbnb as a business

For question 2, the aim was to display the relative distribution of the number of listings per host (*part a)*) and to then analyze which neighbourhoods are the most popular among professional hosts. This could consequently shed light on neighbourhoods that could possibly be more affected by gentrification (*part b)*), since business-oriented hosts are considered to be a driver of gentrification (Bosma & van Doorn, 2022).

Table 4 depicts the absolute and relative number of hosts for different types of professionalization. As mentioned in section 1, I clustered the hosts in different categories by their number of listings, also summarized in table 4. The results show that 92% of the hosts are individual hosts, which only have one listing associated to them. 8% of the hosts are categorized as semi-professionals, having between two and ten listings, while only 1% of the hosts count as professionalized, having more than ten listings. These results also imply that 18,839 listings from our database (~ 30%) are owned by hosts with more than one listing.

These results suggest that individual or private hosts are in the vast majority in both their number, as well as in the number of listings they own, but that the share of listings owned by (semi-)professional hosts is still substantial, considering the sharing economy context of the Airbnb platform. In other words, almost one in three Airbnb listings in Paris is owned by a (semi-)professional host, which can be considered a large share, especially since Airbnb's target audience was initially the private, i.e., individual, host type. As professionalization is positively associated with gentrification, this appears to be at the root of the city's STR issues.

	Host category	Number of Listings	Number of Hosts	%-share of hosts
0	individual	1	41927	92%
1	semi-professionalized	2-10	3545	8%
2	professionalized	>10	238	1%

Table 4: Host type distribution

As for part b) of the second question, table 5 exhibits the top 5 neighbourhoods by their popularity among professional hosts, i.e., hosts having more than ten listings. The results show that the neighbourhoods "Temple", "Bourse", "Passy", "Vaurigard", and "Élysée" are most attractive to professional, business-oriented host, judging by the number of listings these hosts have in the respective neighbourhood. When comparing these results to

our findings regarding the listing density of neighbourhoods, it can be observed that three of the top 5 neighbourhoods among professionals are also three of the top 5 neighbourhoods judging by density from question 1, namely “Temple”, “Bourse” and “Élysée”, thereby reinforcing the significance of findings from question 1. As a consequence, especially these three neighbourhoods appear to be strongly affected by the rise of Airbnb and therefore may run high risk of gentrification effects.

5.2.3. Pricing and type of listing by degree of professionalization

In question 3, the main goal was to show characteristics of listings by degrees of professionalization. In other words, we want to understand what type of listings highly professional hosts own in comparison to individual hosts’ listing types (*part b)*) and what differences there might be in setting prices between host categories (*part a)*). The underlying claims to examine based on research findings discussed in section 1 were that professional hosts set higher prices for their listings compared to individual hosts (Abrate et al., 2022) and that professional hosts are characterized by listing entire homes and apartments rather than rooms or shared rooms (Bosma et al., 2022).

Regarding the price setting of different host categories (*part a)*), the analysis shows that the average listing price for individual hosts is approximately 135€, for semi-professional hosts 214€ and for professional hosts even 251€. This clearly provides evidence for Abrate et al,’s (2022) claim and further depicts the problem with professionalization with regard to gentrification. Business-oriented hosts setting high prices exerts pressure on the housing market of respective cities and neighbourhoods, which drives up general housing prices. The negative consequences of this Airbnb effect were already discussed in length in section 1, thereby casting this finding in a negative light.

For part b) of this question, the results displayed in figure 6 and table 5 depict that the finding of Bosma et al. (2022) is only partially applicable to our use case. While it holds for the professional hosts, it does not hold for the semi-professional hosts, when compared to the individual hosts. However, professional hosts almost exclusively provide entire homes as listing types, accounting for more than 93% of their listings provided. Consequently, this finding illustrates how apartments, that could be rented out long-term, e.g., by families, are provided for STR instead and thereby drive housing shortage and price increases, affecting potential permanent residents.

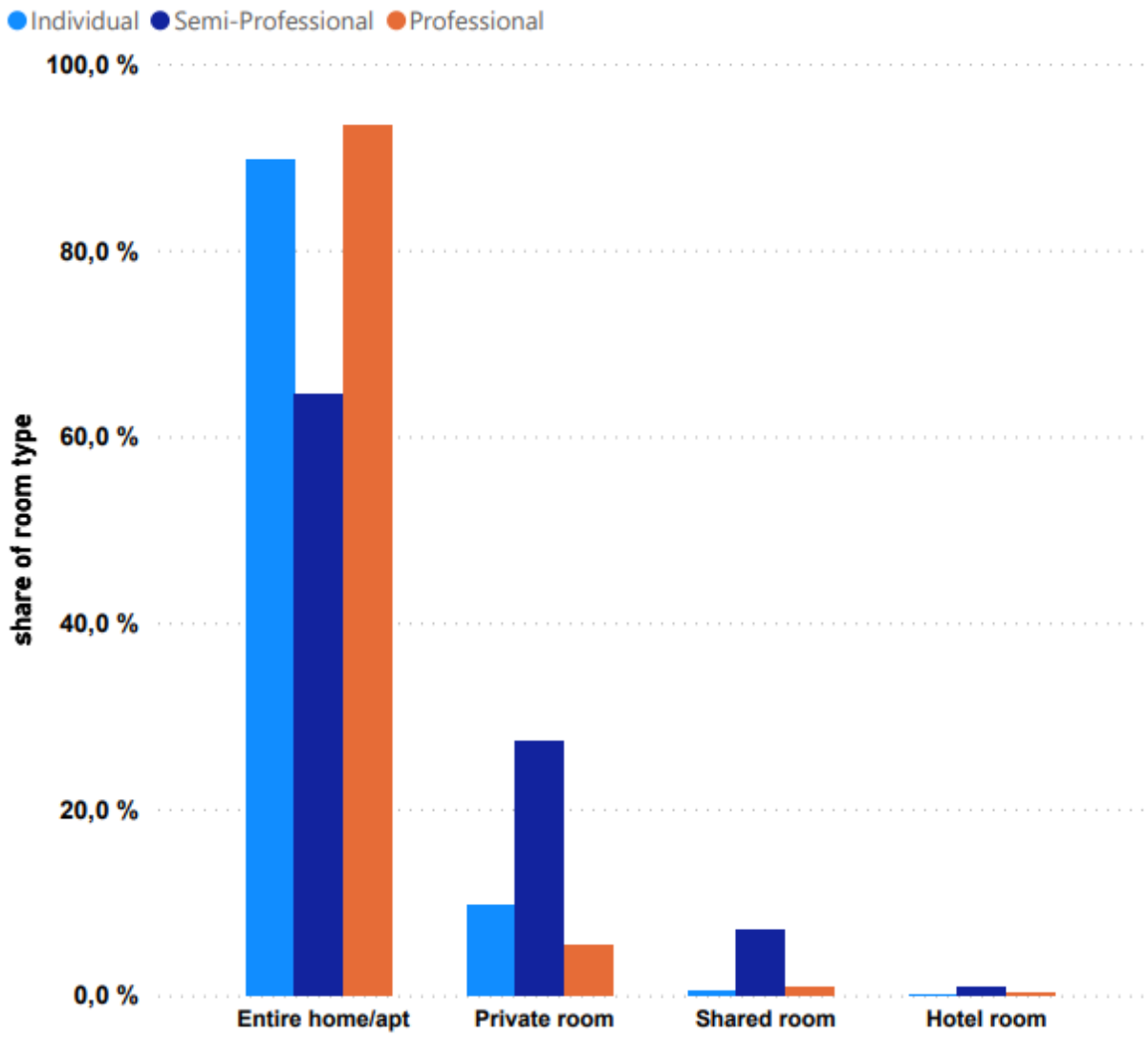

Figure 6: Room type by host category

Room type	Individual	Semi-Professional	Professional
Entire home/apt	89,72 %	64,50 %	93,36 %
Private room	9,79 %	27,41 %	5,34 %
Shared room	0,45 %	7,07 %	0,94 %
Hotel room	0,04 %	1,02 %	0,36 %
Total	**100,00 %**	**100,00 %**	**100,00 %**

Table 5: Room type by host category

6. References

Abrate, G., Sainaghi, R., & Mauri, A. G. (2022). Dynamic pricing in Airbnb: Individual versus professional hosts. Journal of Business Research, 141, 191–199. https://doi.org/10.1016/j.jbusres.2021.12.012

Airbnb. (2022, May 31). New survey: EU Hosts use Airbnb income to afford rising living costs. Airbnb Newsroom. https://news.airbnb.com/new-survey-eu-hosts-use-airbnb-income-to-afford-rising-living-costs/

Barker, G. (2020). The Airbnb Effect On Housing And Rent. Forbes. https://www.forbes.com/sites/garybarker/2020/02/21/the-airbnb-effect-on-housing-and-rent/

Barron, K., Kung, E., & Proserpio, D. (2020). The Effect of Home-Sharing on House Prices and Rents: Evidence from Airbnb. Marketing Science, 40(1). https://doi.org/10.1287/mksc.2020.1227

Bosma, J. R., & van Doorn, N. (2022). The Gentrification of Airbnb: Closing Rent Gaps Through the Professionalization of Hosting. Space and Culture, 120633122210906. https://doi.org/10.1177/12063312221090606

Deboosere, R., Kerrigan, D. J., Wachsmuth, D., & El-Geneidy, A. (2019). Location, location and professionalization: a multilevel hedonic analysis of Airbnb listing prices and revenue. Regional Studies, Regional Science, 6(1), 143–156. https://doi.org/10.1080/21681376.2019.1592699

France Airbnb: Paris hails victory over short-stay rents. (2020, September 22). *BBC News*. https://www.bbc.com/news/world-europe-54246005

France: population of Paris by arrondissement. (n.d.). Statista. https://www.statista.com/statistics/1046193/population-by-district-arrondissements-paris-france/

Gant, A. C. (2016). Holiday Rentals: The New Gentrification Battlefront. Sociological Research Online, 21(3), 1–9. https://doi.org/10.5153/sro.4071

Inside Airbnb. Adding data to the debate. (2019). Inside Airbnb. http://insideairbnb.com/

Paris authorities rule to restrict short-term rental operations. (2021, February 22). Short Term Rentals. https://shorttermrentalz.com/news/paris-restrictions/

Paris hails success of tough rules for short-term lets on Airbnb. (2023, August 12). RFI. https://www.rfi.fr/en/france/20230812-paris-city-hall-hails-success-of-tough-rules-for-short-term-lets-on-airbnb

Rabiei-Dastjerdi, H., McArdle, G., & Hynes, W. (2022). Which came first, the gentrification or the Airbnb? Identifying spatial patterns of neighbourhood change using Airbnb data. Habitat International, 125, 102582. https://doi.org/10.1016/j.habitatint.2022.102582