

Barcelona Airbnb Insights

Daniel Herrera & Rodrigo González

2024-02-10

1. Presentation of the case



1.1. Airbnb

Airbnb, founded in 2008, has grown from a simple idea of renting out air mattresses in a San Francisco apartment to a global phenomenon in the hospitality industry. It has significantly altered the way people travel by offering unique accommodations from local hosts in over 100,000 cities worldwide. Its platform caters to a diverse range of preferences, from single rooms to entire homes, providing more personalized and often more cost-effective lodging options than traditional hotels. This peer-to-peer model has not only democratized travel accommodations but also enabled millions of hosts to generate additional income. As of 2022, Airbnb had hosted over 800 million guest arrivals, showcasing its vast growth and popularity among travelers seeking authentic experiences. This growth trajectory highlights Airbnb's impact on the travel and tourism sector, prompting a reevaluation of traditional hospitality models and regulatory frameworks globally.

1.2. Airbnb controversies

Airbnb's rise has sparked controversies, especially around its impact on local housing markets, community dynamics, and regulatory compliance. The platform has been criticized for contributing to housing shortages by converting long-term rental properties into short-term tourist accommodations, leading to increased rent and property prices. In cities like Barcelona, this has exacerbated existing tensions between residents and tourists, contributing to over-tourism and altering neighborhood characters. Additionally, Airbnb has faced legal challenges regarding compliance with local housing laws and regulations, with accusations of facilitating illegal rentals. These issues have prompted cities worldwide to implement stricter regulations on short-term rentals, balancing the need for tourism revenue with protecting residents' interests and preserving housing affordability.

Recently, Airbnb has seen a noticeable increase in rental prices, attributed to a combination of factors including heightened demand, limited supply, and the gradual recovery of the travel industry post-pandemic. This surge reflects a broader trend in the accommodation sector where prices are climbing as travelers return in large numbers, seeking unique and safe lodging options. The rise in prices has sparked discussions on affordability and access, urging both Airbnb and hosts to balance profitability with providing value to guests.

1.3. Airbnb in Barcelona

Airbnb's controversies in Barcelona have been a focal point of discussions about the impact of short-term rentals on cities globally. Barcelona, a prime tourist destination, has faced significant challenges due to the proliferation of Airbnb listings, leading to tensions between the platform, the city's residents, and local authorities.

One major controversy revolves around housing affordability and availability. Critics argue that Airbnb contributes to rising rents and the displacement of long-term residents in favor of short-term tourists. The city has seen protests from local groups and activists who claim that neighborhoods have lost their identity and cohesion due to the influx of tourists staying in Airbnb properties. These concerns have been echoed by housing advocacy groups like the Platform for People Affected by Mortgages (PAH), which has been vocal about the negative impacts of short-term rentals on the local housing market.

The Barcelona City Council, led by Mayor Ada Colau, has taken a firm stance against illegal tourist rentals. Since coming to office in 2015, Colau has introduced strict regulations aimed at curbing the growth of short-term rental platforms like Airbnb. In 2016, the city imposed a moratorium on new tourist licenses and fined Airbnb €600,000 for listing unlicensed properties, highlighting the city's commitment to enforcing its regulations.

Further actions were taken in subsequent years, including the introduction of the PEUAT (Special Urban Plan for Tourist Accommodation) in 2017, which aimed to limit the expansion of tourist accommodations and ensure a balance between tourism and local residents' needs. Despite these measures, conflicts persist, with Airbnb arguing that the city's approach is overly restrictive and harms local hosts who rely on income from short-term rentals.

1.4. Data sources

Despite the significance of Airbnb data, the private nature of the company, being publicly listed, rules out its open publication. Consequently, an alternative approach is necessary for data collection. In this report, we opt to depend on robust and reliable projects that have already proven their significance in regulating this reservation platform, as well as data sourced directly from the municipality.

We have established two distinct sources of data, the first involves the *Inside Airbnb* project, while the second is taken from the Open Data portal of the Barcelona City Council.

Inside Airbnb Adding data to the debate

1.4.1. Inside Airbnb

The Inside Airbnb (<http://insideairbnb.com/>) project is a comprehensive, independent online platform that provides data and analysis on the operations of Airbnb in cities around the world. It was started by activist and data scientist Murray Cox in 2016 as a response to the lack of transparency regarding how Airbnb's business model impacts local housing markets, neighborhoods, and communities. The project aggregates publicly available information on Airbnb listings and reviews to present a more detailed picture of the platform's presence in various cities.

Inside Airbnb has become a crucial resource for researchers, policymakers, journalists, and community groups concerned about the rapid growth of short-term rentals and their effect on cities. By offering detailed data on the number of listings, types of properties, location, pricing, and host information, the project sheds light on how Airbnb contributes to issues like housing affordability, gentrification, and regulatory compliance.

One of the most significant repercussions of the Inside Airbnb project has been its influence on public policy and regulation. Cities like New York, San Francisco, and Barcelona have used data from Inside Airbnb to understand the scale and impact of short-term rentals, leading to the development and enforcement of more stringent short-term rental regulations. For instance, New York City has implemented rules requiring hosts to register with the city, a move aimed at cracking down on illegal listings—a policy informed by insights gained from Inside Airbnb's data.

Inside Airbnb's work has not been without criticism, particularly from Airbnb and hosts who argue that the platform's data might be misinterpreted or used to unjustifiably restrict short-term rentals. Nonetheless, the project's impact in highlighting the need for balance between tourism benefits and community well-being

remains undeniable.

In our case we have used the data set of the city of Barcelona through two files, one being the facelifts and the other being the GEOJson of the locations. The quality of the data is indisputable and has been recognized by various actors. Likewise, the update is continuous and we work with data from less than two months ago, these being from mid-December.

We thank Inside Airbnb for its collaboration with consumers and the governments in making this data public, easily reachable and visible.



1.4.2. Open data BCN

The portal <https://opendata-ajuntament.barcelona.cat/> (<https://opendata-ajuntament.barcelona.cat/>) is an initiative by the Barcelona City Council aimed at promoting transparency, citizen participation, and innovation through open access to public data. This digital platform offers a wide array of datasets on various city aspects such as transportation, environment, demographics, and economy, enabling developers, researchers, and the general public to analyze, create, and share applications that enhance the understanding and management of the city.

In our case we have used the data set of "Population of Barcelona aggregated by sex according to the Municipal Register of Inhabitants on January 1 of each year". Barcelona City Council offers an easy-to-use API from which we have only indexed the fields of our interest.

We thank Barcelona City Council for its collaboration with citizens and general public for publishing this data of general interest.

https://opendata-ajuntament.barcelona.cat/data/en/dataset/pad_mdbas_sexe/resource/d0e4ec78-e274-4300-a3bc-cb85cf79014d (https://opendata-ajuntament.barcelona.cat/data/en/dataset/pad_mdbas_sexe/resource/d0e4ec78-e274-4300-a3bc-cb85cf79014d)



1.5. Municipal Consumer Information Office (OMIC)

The Municipal Consumer Information Office (OMIC) in Barcelona plays a crucial role in protecting consumer rights within the city. This institution provides essential services such as advice, mediation, and conflict resolution between consumers and businesses. Established to ensure that consumer rights are respected and promoted, OMIC offers a valuable resource for citizens facing issues with products or services.

OMIC operates in various areas, including handling complaints and claims, promoting responsible consumption practices, and educating about consumer rights. It also offers workshops and informative talks, contributing to raising awareness and educating consumers about their rights and how to effectively exercise them.

Additionally, OMIC monitors commercial practices within the city to ensure compliance with current consumer legislation. This includes inspecting commercial establishments and implementing corrective actions in cases of infringement. Its work is essential in maintaining a fair balance between consumer interests and those of businesses, promoting a transparent and fair market.

We consider that OMIC is the perfect client for a report that reflects the impact that Airbnb is having on the city of Barcelona. Likewise, and after the major controversies over the price increase that Airbnb has had lately, it is interesting to observe the profile of accommodation and large license holders that Barcelona City Council has had under its sights for years.

This study is the first phase and after adequate negotiation a second could be launched with much more detailed data.

2. Data exploration

2.1. Data structure

2.1.1 Inside Airbnb

Each entry collected by Inside Airbnb, corresponding to one geographical location or city, is structured unanimously in the following manner:

Country/City	File Name	Description
Barcelona	listings.csv.gz	Detailed Listings data
Barcelona	calendar.csv.gz	Detailed Calendar Data
Barcelona	reviews.csv.gz	Detailed Review Data
Barcelona	listings.csv	Summary information and metrics for listings in (good for visualisations).
Barcelona	reviews.csv	Summary Review data and Listing ID (to facilitate time based analytics and visualisations linked to a listing).
Barcelona	neighbourhoods.csv	Neighbourhood list for geo filter. Sourced from city or open source GIS files.
Barcelona	neighbourhoods.geojson	GeoJSON file of neighbourhoods of the city.

Given this recurring structure and the richness of the data, the initial goal of the project was to compare data from different European cities. This objective is evident in the code implementation for loading data into data frames, designed to be replicable and applicable to any city's data, provided the files are loaded into the working directory and stored in folders named after the respective cities. This approach was eventually discarded, as it became clear that focusing on Barcelona as a study case is already in itself an interesting focus.

The data collected from Barcelona corresponds to the 13 of December, 2023.

The main source of data is the `listings.csv` file, which contains a collection of 18,321 listings with 18 variables:

```
dim(Barcelona)

## [1] 18321    18
```

```
str(Barcelona)
```

```
## 'data.frame': 18321 obs. of 18 variables:
## $ id : num 17475 18674 198958 23197 32711 ...
## $ name : chr "Rental unit in 08013 Barcelona · ★4.40 · 1 bedroom · 1 bed · 1 bath" "Rental unit in Barcelona · ★4.33 · 3 bedrooms · 6 beds · 2 baths" "Rental unit in Barcelona · ★4.69 · 4 bedrooms · 6 beds · 2 baths" "Rental unit in Sant Adria de Besos · ★4.77 · 3 bedrooms · 4 beds · 2 baths" ...
## $ host_id : int 65623 71615 971768 90417 135703 440825 73163 10138 55 1014050 73163 ...
## $ host_name : chr "Luca" "Mireia Maria" "Laura" "Etain (Marnie)" ...
## $ neighbourhood_group : chr "Eixample" "Eixample" "Sant Martí" "Sant Martí" ...
## $ neighbourhood : chr "la Dreta de l'Eixample" "la Sagrada Família" "Diagonal Mar i el Front Marítim del Poblenou" "el Besòs i el Maresme" ...
## $ latitude : num 41.4 41.4 41.4 41.4 41.4 ...
## $ longitude : num 2.17 2.17 2.21 2.22 2.17 ...
## $ room_type : chr "Entire home/apt" "Entire home/apt" "Entire home/apt" "Entire home/apt" ...
## $ price : int 140 121 304 200 79 48 120 120 150 226 ...
## $ minimum_nights : int 5 1 2 3 1 4 5 4 3 5 ...
## $ number_of_reviews : int 26 40 105 75 99 168 8 244 142 217 ...
## $ last_review : chr "2023-12-04" "2023-11-07" "2023-10-16" "2023-11-25" ...
## $ reviews_per_month : num 0.16 0.31 0.74 0.48 0.66 1.34 0.05 1.67 0.96 1.35 ...
## $ calculated_host_listings_count : int 1 30 9 2 3 1 3 1 1 3 ...
## $ availability_365 : int 32 39 137 300 297 18 90 129 0 228 ...
## $ number_of_reviews_ltm : int 9 7 26 11 16 29 0 44 22 27 ...
## $ license : chr "" "HUTB-002062" "HUTB-000926" "HUTB005057" ...
```

The interpretation of variables is the following:

- `id` : numerical identifier for each listing
- `name` : name of the listing
- `host_id` : numerical identifier for the host of the listing
- `host_name` : name of the host.
- `neighbourhood` : location information specifying the neighborhood the listing is in
- `neighbourhood_group` : larger district grouping of neighborhoods (simplified from 71 to 10 districts)
- `latitude` : latitude coordinate of the listing
- `longitude` : longitude coordinate of the listing
- `room_type` : categorical string describing the type of room
- `price` : daily price of the listing in the local currency (euro)
- `minimum_nights` : minimum number of nights the listing can be booked
- `number_of_reviews` : number of reviews at the time the data was captured
- `last_review` : date of the latest review for the listing
- `reviews_per_month` : average number of reviews per month for each listing
- `calculated_host_listings_count` : number of listings by the listing host, calculated directly from the data
- `availability_365` : availability of the listing in days starting from the data capture day within the next year
- `number_of_reviews_ltm` : number of reviews in the past 12 months

- `license` : compliance with City Council; includes the license ID number if available, otherwise, pending or empty

2.1.2. Open data BCN

As explained previously, the data from has been enriched with official demographic information obtained via API from the Open Data BCN project from the City Council. The topic of choice was “Population of Barcelona aggregated by sex according to the Municipal Register of Inhabitants on January 1 of each year”, found in the file `2023_pad_mdbas_sexe.csv` , from which the population grouped by neighbourhood and district was extracted and merged into our main data frame. The indexed fields are:

- `Nom_Districte` : name of the greater district
- `Nom_Barri` : name of the neighbourhood
- `Valor` : population count

The merge has been done for `neighbourhood = Nom_Barri` , resulting in a new `demographics` column in our `Barcelona_md` data frame (md for merged).

2.2. Further data enrichment

From a first assessment of the data it was observed that within it, additional numerical and categorical variables could be extracted and used to provide insightful details once plotted.

2.2.1. Star rating

The `name` field was found to include among other information the star rating from 0 to 5 preceded by a star character, a pattern easily reproducible with regular expressions. With pipes and filters it has been extracted as a numerical value and introduced as an additional column. Cases where names do not contain a star rating are listed as NA.

2.2.2. License status

String items found under `license` were observed to either contain a license ID matching the City Council's figure HUTB (Habitatge d'Ús Turístic de Barcelona), the word 'Exempt', or be missing at all. A new variable `LicenseGrouping` was established to contain three new categories: `Exempt` , `License is displayed` and `License is not displayed` .

2.2.3. Large and small tenants

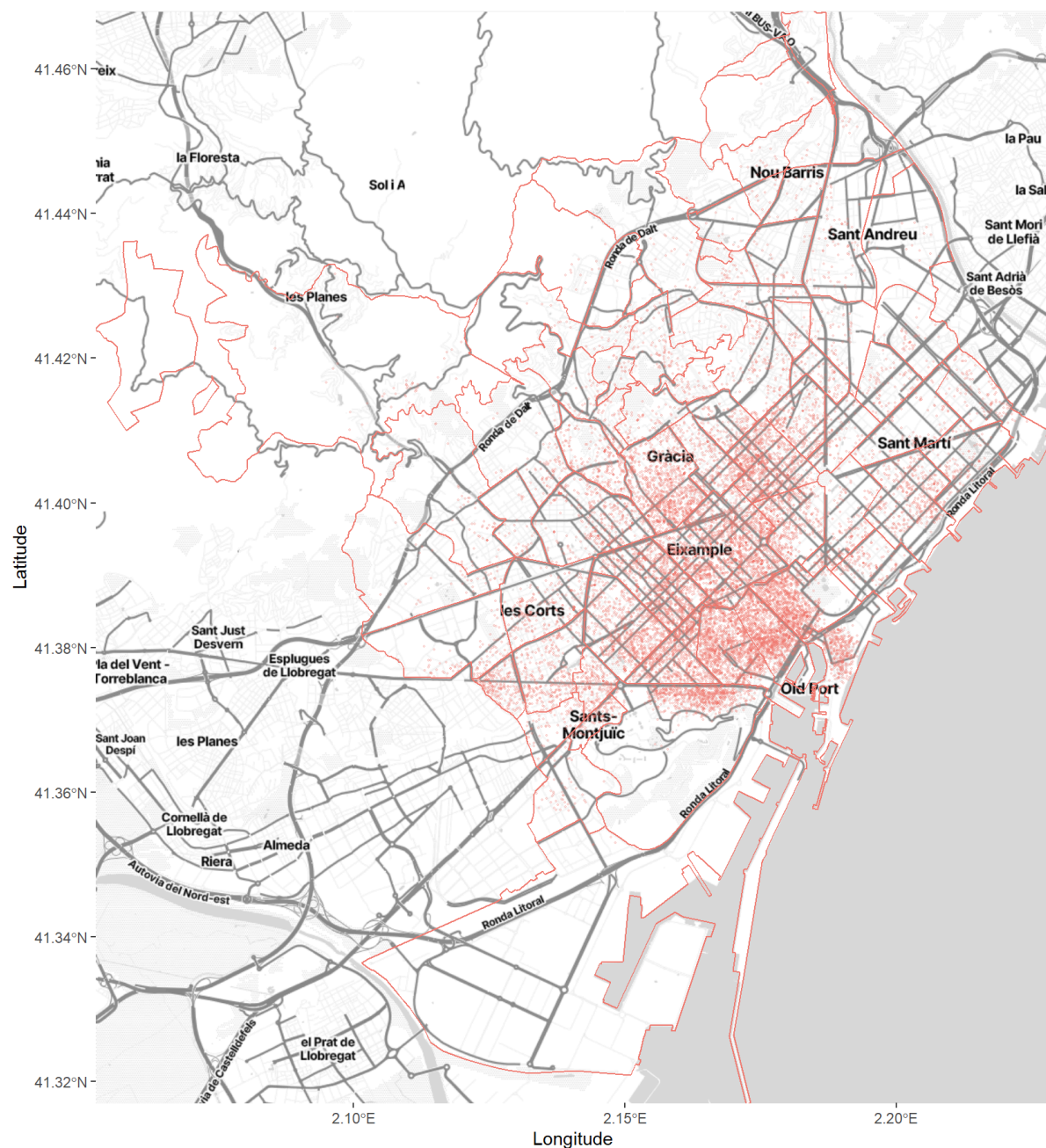
By considering the variable `calculated_host_listings_count` and referencing the Spanish Housing Law of 12/2023 (<https://www.boe.es/buscar/act.php?id=BOE-A-2023-12203>), which outlines the differentiation between small and large tenants when there are 5 or more properties involved, a new categorical variable, `TenantSizeGrouping` , was established to capture this observation.

2.3. Visualizations

2.3.1. Airbnb listings map overview

Airbnb listings overview

Barcelona

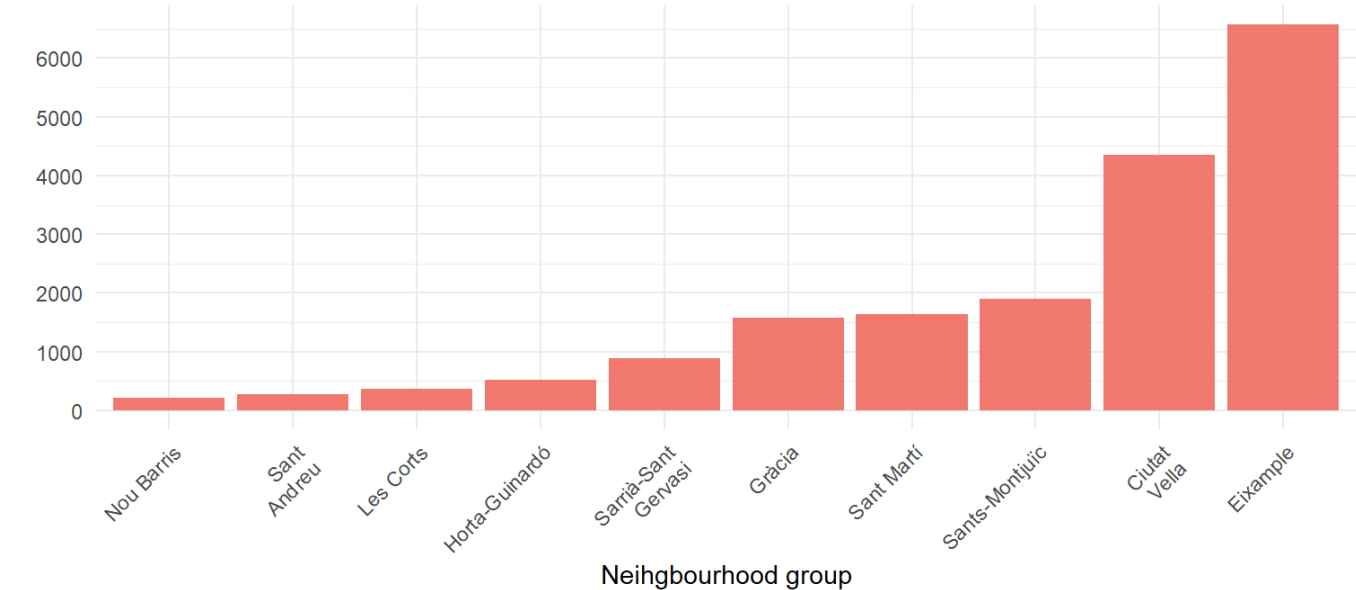


A straightforward visualization of the dataset's listings already reveals diverse concentrations, aligning with popular tourist and overnight stay destinations in Barcelona. The primary cluster is situated in the expansive neighborhood of *Eixample* and *Ciutat Vella*, along with *Gràcia*, and the vicinity encompassing the main train station of *Sants* and *Montjuïc*. The density significantly diminishes beyond these central regions. It is essential to acknowledge that the recorded data corresponds to the official district divisions of Barcelona, delineated by their boundaries. It is plausible that additional clusters of listings may exist outside these demarcations. For instance, areas like the proximity of the airport in the southwestern town of *El Prat de Llobregat* might host distinct clusters. Incorporating such regions into future studies could uncover valuable insights in this regard.

2.3.2. Airbnb listings by neighbourhood group

Airbnb listings by neighbourhood group

Barcelona



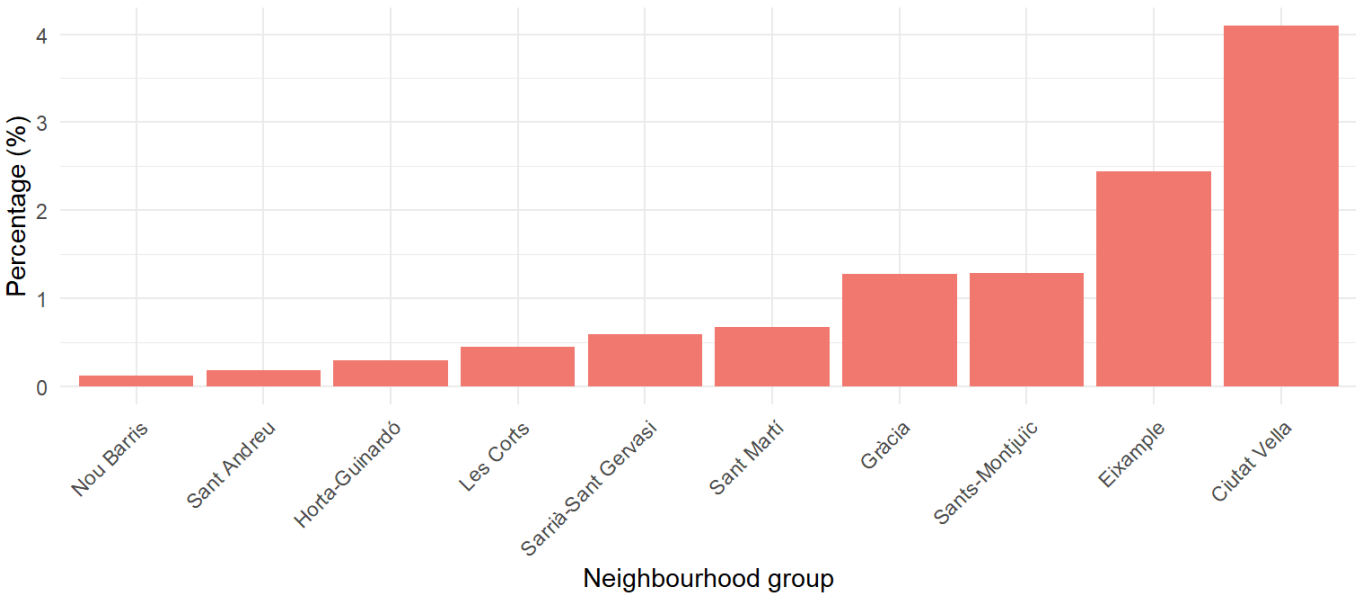
This bar chart illustrates the number of Airbnb listings across different neighbourhoods in Barcelona. The most prominent feature is the overwhelming dominance of the *Eixample* neighbourhood, with the number of listings towering over other areas. This suggests *Eixample* is a highly sought-after area for tourists or visitors using Airbnb. In contrast, neighbourhoods like *Nou Barris* and *Sant Andreu* have far fewer listings, which may indicate less tourist activity or a lower availability of rental properties on Airbnb.

The distribution indicates a potential disparity in the spread of tourism across the city, with certain areas possibly facing higher pressure from tourist accommodation. It is very interesting to note that only considering *Eixample* the number of listings is greater than the sum of the last six neighborhoods. This could have implications for local housing markets, infrastructure demand, and urban planning. The chart effectively highlights the disparities and could serve as a basis for more detailed analysis on the impact of short-term rentals on the urban landscape of Barcelona. As mentioned, there are several areas that do not reach a few hundred listings, which indicates a very low concentration of Airbnbs in relation to the people who reside there.

2.3.3. Airbnb listings per 100 residents by neighbourhood group

Airbnb listings per 100 residents by neighbourhood group

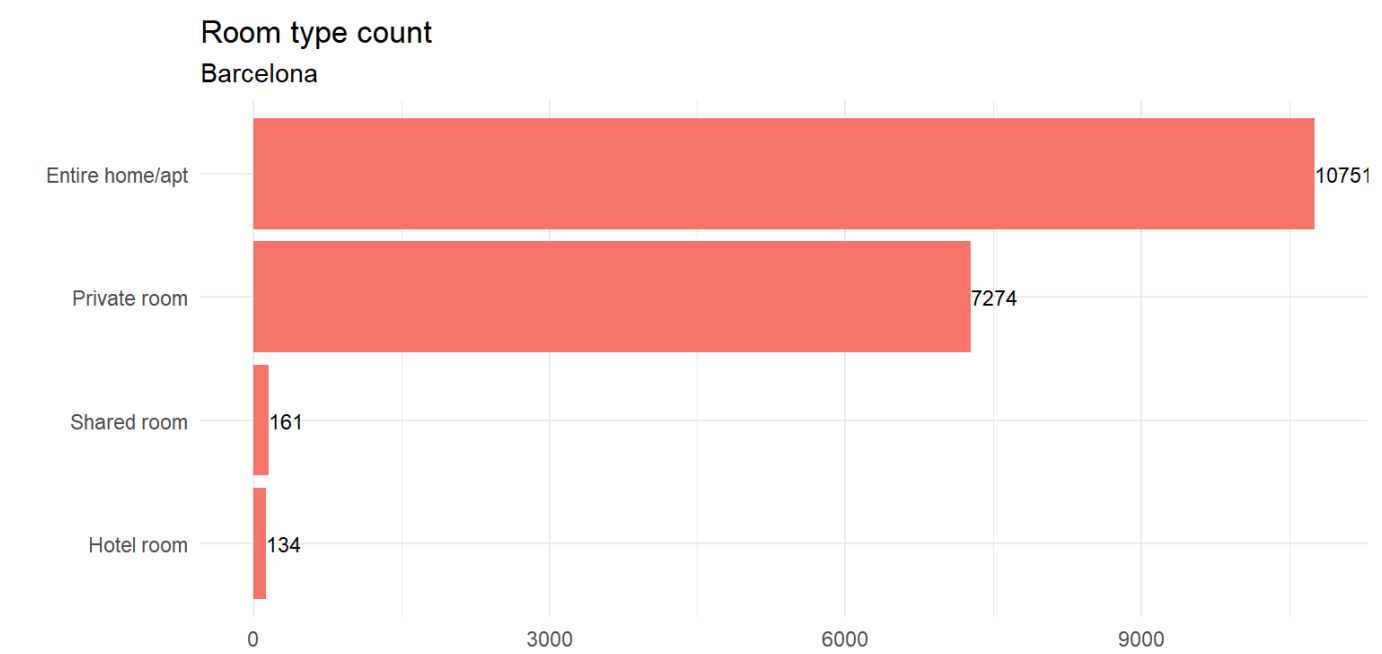
Barcelona



The data presented in this bar graph serves as an insightful complement to the previous visualization. By comparing the number of listings in each neighbourhood group with the official demographics corresponding to that district, a proportion of Airbnb listings per person can be obtained. This better reflects the different densities of listings between parts of the city. Although *Eixample* and *Ciutat Vella* still reign as the most listing-saturated districts, the higher concentration of *Ciutat Vella* compared to *Eixample* becomes obvious. This is explained by many factors, such as the higher urban density of the Old City and the resulting higher concentration of population by area, and also *Eixample* being a rather large district within Barcelona. Similar observations can be made between the other neighbourhood groups.

According to the Government of Catalonia (<https://web.gencat.cat/es/actualitat/detall/Regulacio-de-pisos-turisticos-per-garantir-laccess-a-lhabitatge>), areas with more than 5 tourist housing listings per 100 inhabitants face housing access issues. A recent Decree Law has set this limit to address the problem, with city councils having the authority to permit up to 10 through their urban planning. It is important to note that our study focuses exclusively on Airbnb listings, and other units may be legally listed as tourist accommodations on different platforms.

2.3.4. Room type count



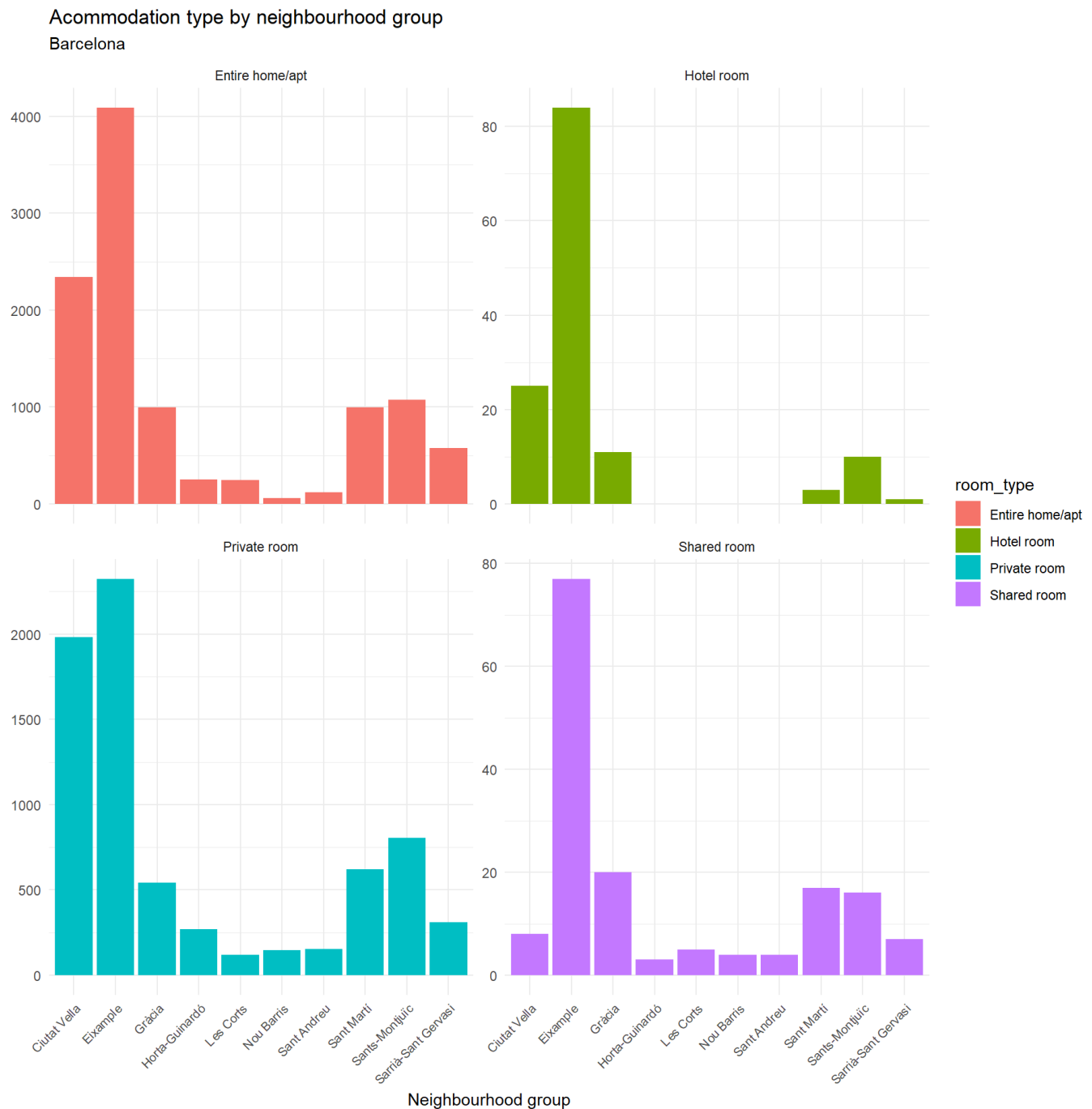
There are four categories of room types displayed: Entire home/apt , Private room , Shared room , and Hotel room .

The Entire home/apt category has the highest count, surpassing the 10,000 mark, suggesting that entire apartments or homes are the most common type of property listed in Barcelona. The Private room category follows, with roughly half as many listings as entire homes, indicating a significant presence in the market but less than the full property rentals.

Shared rooms and Hotel rooms have significantly fewer listings compared to the other two categories. Shared rooms barely register on the scale, suggesting they are a less popular option among the listings. Hotel rooms have the smallest count, indicating that traditional hotel stays are much less commonly listed on platforms likely compared to short-term rental options.

The visual emphasizes the prevalence of whole property rentals in Barcelona’s accommodation offerings, with a substantial secondary market for private rooms. The minimal presence of shared and hotel room listings could reflect market demand or possibly restrictions and regulations within the city.

2.3.5. Accommodation type by neighbourhood group



This chart provides a detailed breakdown of Airbnb listings in Barcelona by room type and neighbourhood group, with four distinct categories of accommodation: entire home/apt , hotel room , private room , and shared room . The y-axis represents the number of listings, which is not consistent across the categories, indicating the use of a free scale within the facets to better display the range of data.

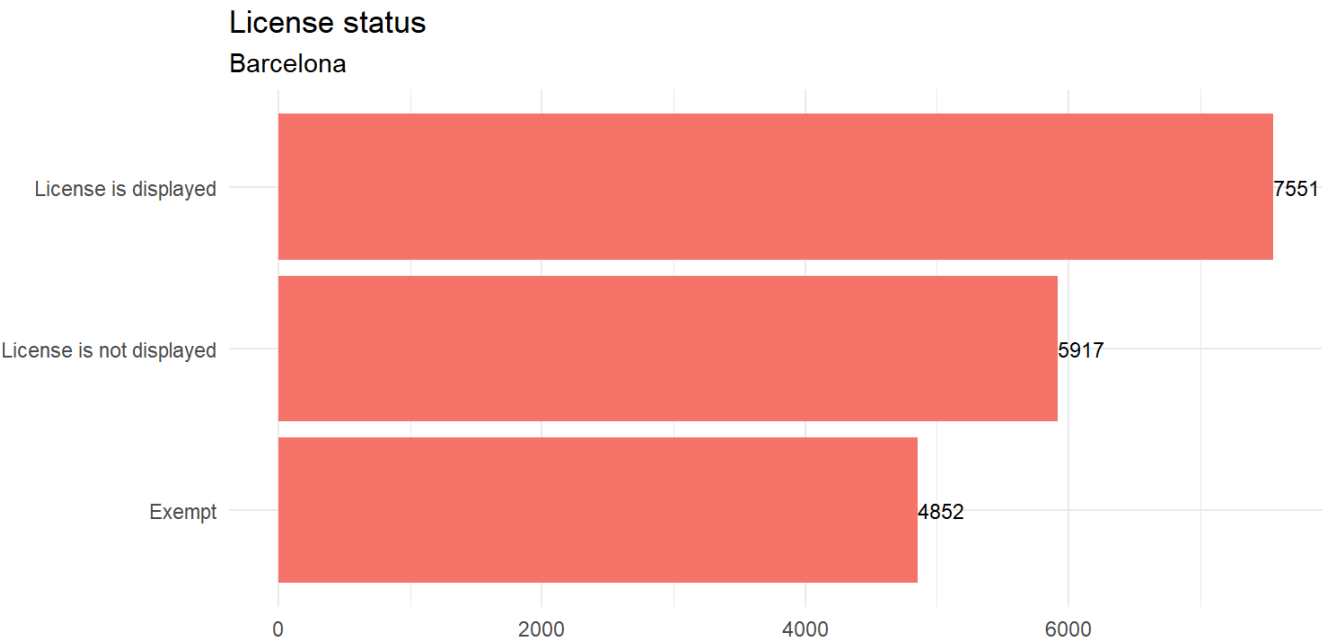
- Focusing first on the Entire home/apt category, we observe a pronounced peak in *Eixample*, with over 4,000 listings, which significantly overshadows the counts in other areas, where the second-highest listing *Ciutat Vella* has just above 1,000. This indicates a heavy concentration of full-property rentals in that particular part of the city, which shows that *Eixample* is a popular central area known for tourist attractions.
- The Hotel room category exhibits a very different scale, peaking at around 80 listings in the most prominent neighbourhood, being again *Eixample* for this category. This suggests that hotel rooms are a minor part of the Airbnb market in Barcelona or that hotels prefer to use other channels for renting out rooms. There are several neighborhoods that do not even present a single hotel listing.

- For Private rooms , the distribution seems more even, yet *Eixample* stands out with nearly 2,000 listings, which is about triple the number of listings in the second next populous neighborhood in this category, *Sants-Montjuïc*. The shape of the graph is surprisingly similar to the Entire home/apt .
- The Shared room type displays the least number of listings across neighborhoods, with the highest being under 80. The low count could indicate that shared rooms are not a preferred choice for visitors to Barcelona, or such listings are rare.

The vast discrepancy between the number of Entire home/apt listings and other types suggests that visitors to Barcelona may prefer the privacy and space of an entire apartment. The data could also imply a potential regulatory environment that either supports whole-home rentals or one that has yet to address this preference in the sharing economy.

The chart is an excellent tool for stakeholders to assess market saturation in various neighborhoods and room types. For investors and property managers, areas with lower counts could represent potential growth opportunities. Conversely, neighborhoods with high listings might face more competition, affecting pricing strategies. For policymakers such as OMIC and the Barcelona City hall, such data can be crucial in understanding how short-term rentals are distributed across the city and may guide decisions on tourism management, zoning, and housing policies to balance the needs of residents and visitors.

2.3.6. License status

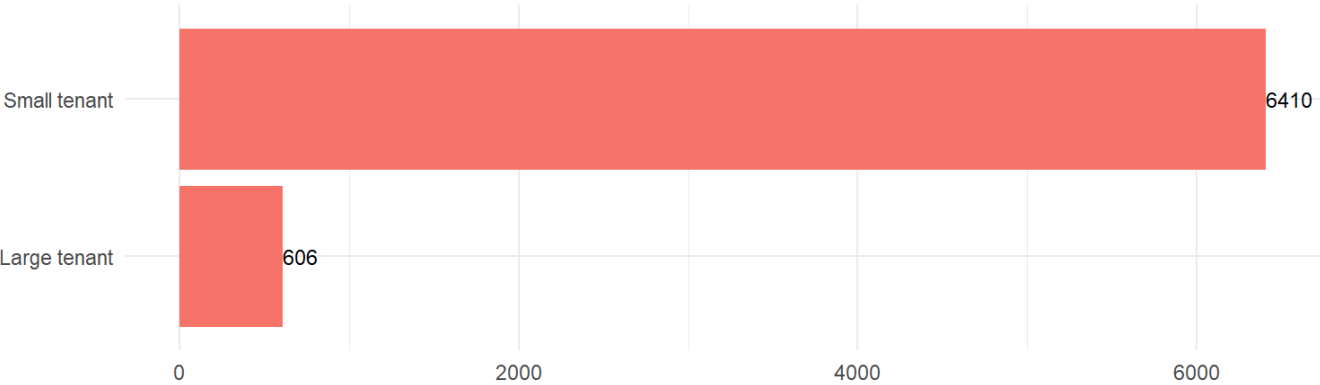


This visualization provides a rough insight into the licensing status of listings. According to the law, full-home listings require a registration and a license number. Instances under the License not displayed group may either be listed illegally or have their license approval pending. Exempt could mean a listing only includes part of a housing unit or a room, and thus does not have to be legally registered as full touristic housing. It would be valuable to observe the evolution of the License is displayed group over time, to evaluate whether the city’s efforts to regulate touristic housing have any effectiveness.

2.3.7. Large vs. small tenants

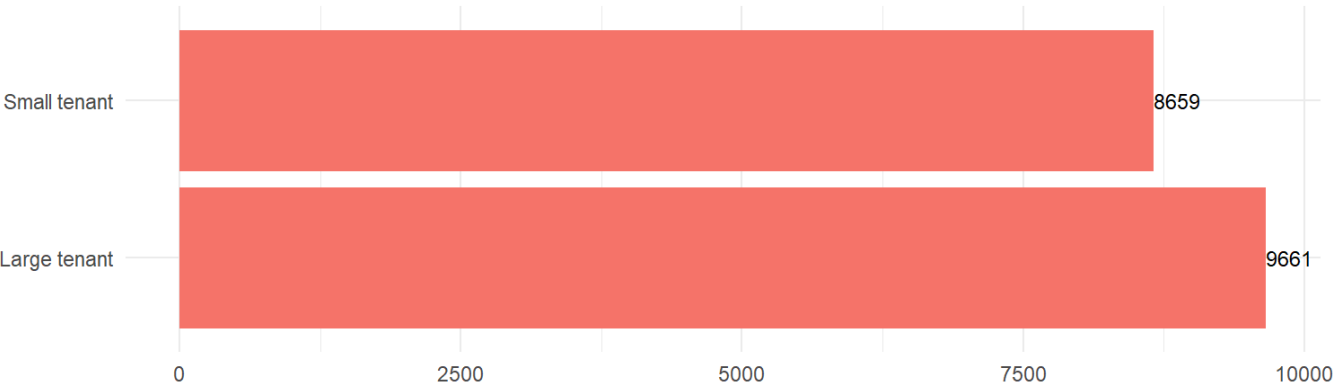
Large vs. small tenants - Host count

Barcelona



Large vs. small tenants - Listing count

Barcelona



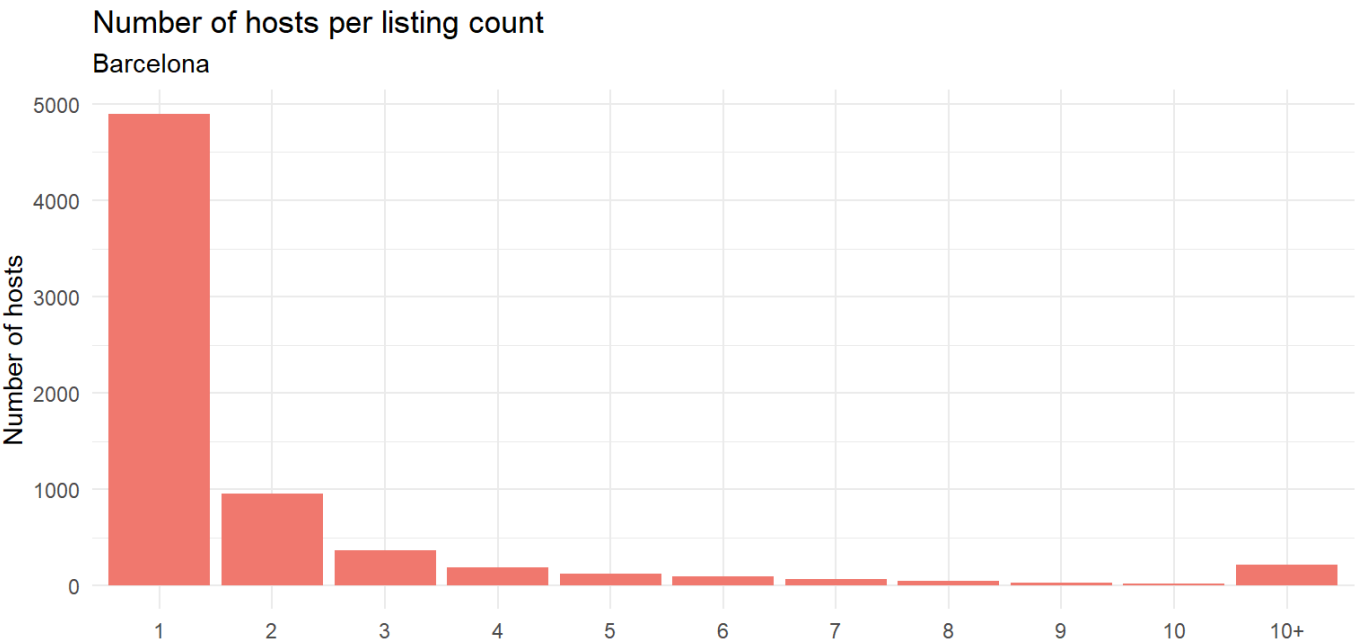
This set of charts looks at two host categories: Small tenant and Large tenant , based on the number of listings they possess. Given the context of rental tension zones, a recent legislation (<https://www.boe.es/buscar/act.php?id=BOE-A-2023-12203>) changed the definition of a large property owner from 10 to 5 properties. This meant that several former small tenants now qualify as large tenants, potentially altering the market dynamics and affecting the application of policies within stressed areas.

In the first plot, the count of Small tenant vastly outnumbers Large tenant , indicating a much higher proportion of landlords with fewer properties. This however is reflected otherwise when looking at the overall Listing count in the second plot, which shows a very close match between number of listings that are registered by Small tenant and Large tenant .

From this observation we can conclude that, although much fewer in number, large tenants represent the majority of the overall market. It is also worth noting that the presence of more small tenants could reflect a diverse range of property offerings, from single rooms to entire apartments, which may cater to different segments of the population. This might have a stabilizing effect on rental prices, as a diverse supply can meet varied demand. However, if small tenants begin to consolidate or if their listings push housing costs above the 30% income threshold, it could lead to increased regulatory scrutiny.

Regarding the law’s stipulations, if the large tenants’ listings are concentrated in areas where housing costs exceed 30% of household income or where rental prices have outpaced CPI by more than 3 percentage points, these areas could be designated as rental tension zones. This would trigger regulatory measures that could include rent capping or other controls to protect tenants.

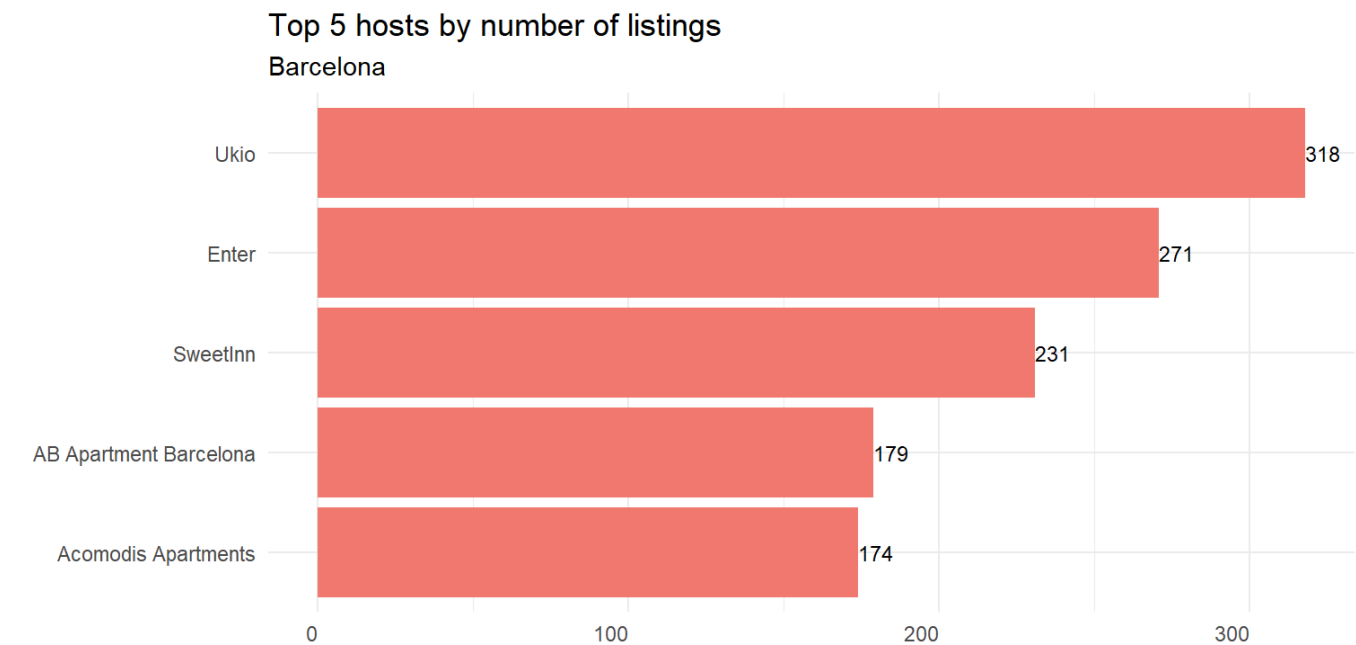
2.3.8. Number of hosts per listing count



The above plot illustrates another defining observation, by sorting all Airbnb hosts according to the number of listings they have registered on the platform. Represented in the variable `Number of hosts`, out of a total of 7,015 hosts, an overwhelming majority is responsible for only one listing, with a quick decrease in number as the `Listing count` increases

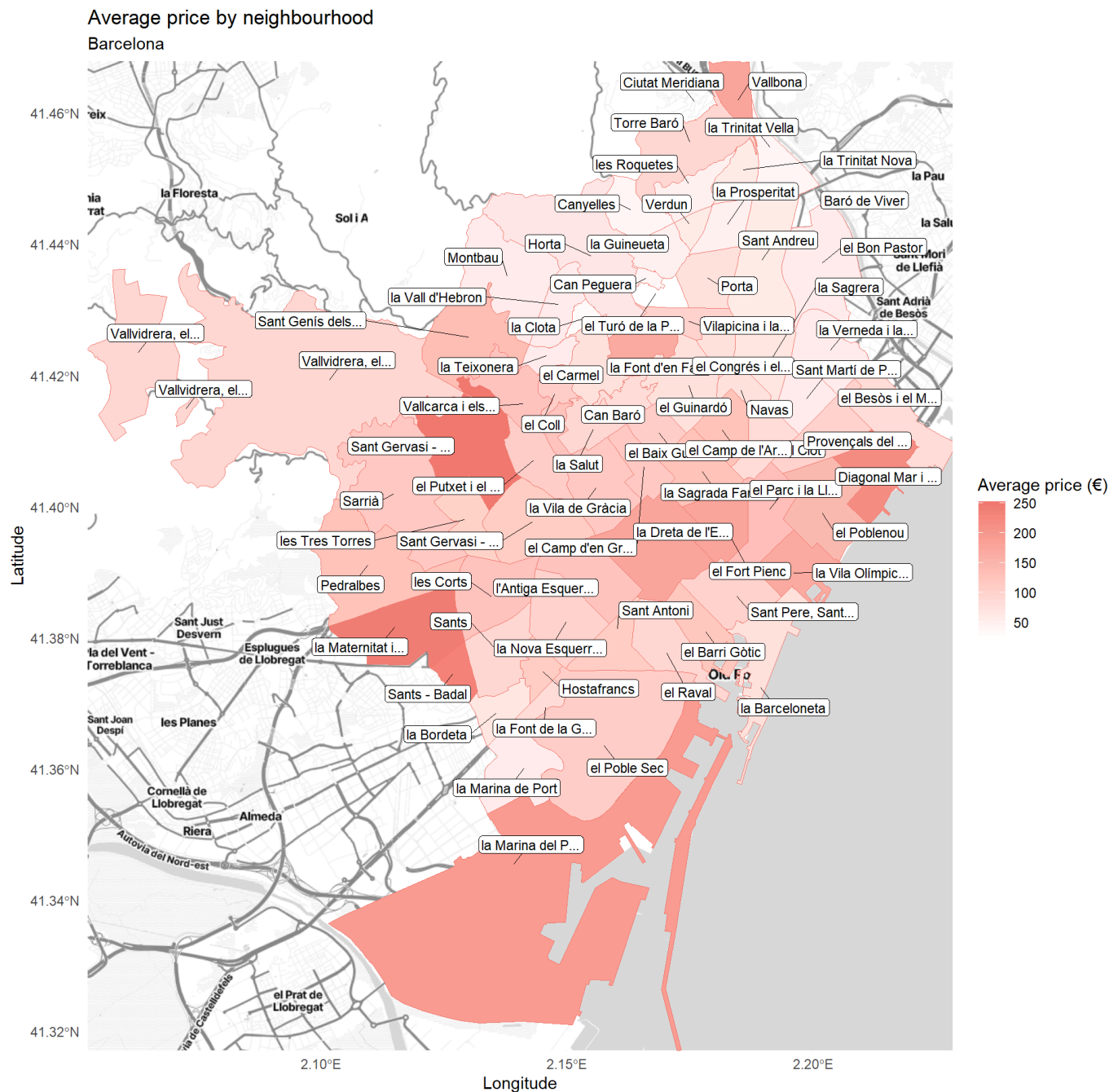
It becomes obvious that the market is overwhelmingly populated by 1-property hosts, suggesting a high level of competitiveness and relatively low monopoly of the market

2.3.9. Top 5 hosts by number of listings



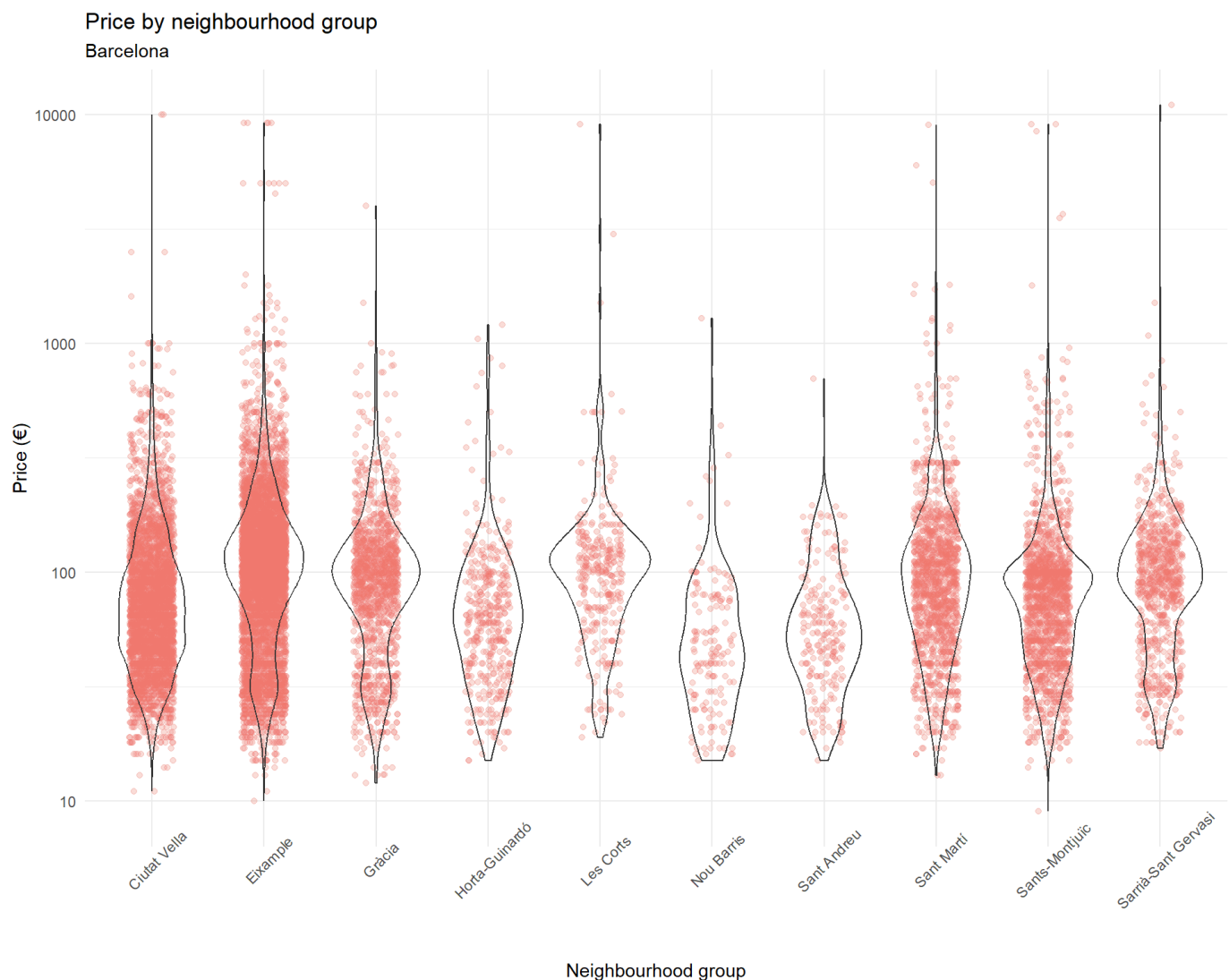
To provide further insight into the host number and listing count disparity, by querying through unique items of `host_id`, and sorting based on `calculated_host_listings_count` we can obtain the 5 top hosts by number of listings. A brief look at the graph reveals that each of the five entities is accountable for a significantly greater number of properties than what is deemed as characteristic of large tenants according to the law. Also, the assumption could be made that the names reflect that such hosts are not actual individual human users of the platform, but rather companies or trusts, which operate large touristic housing rental businesses across Barcelona.

2.3.10. Average price by neighbourhood



The map above illustrates the districts of Barcelona (the 71 divisions designated by the City Council), with colors representing a gradient based on the average price of Airbnb listings in each area. The average prices range from approximately €50 to €250, revealing a general trend of decreasing prices in neighborhoods farther from the center and popular districts, with a subtle decline from south to north. However, this pattern is disrupted by stark outliers, notably in the neighborhoods of *Sant Gervasi - la Bonanova* and *la Maternitat i Sant Ramon*. These anomalies can be attributed to a few exclusive listings in the vicinity of €10,000 per night. The reason for these extremes may be the unique nature of these listings as highly exclusive overnight units, such as villas or large fully furnished houses, or also potential data quality issues. As this report focuses solely on observations, a more in-depth analysis of these factors would be suitable for a future iteration of the study.

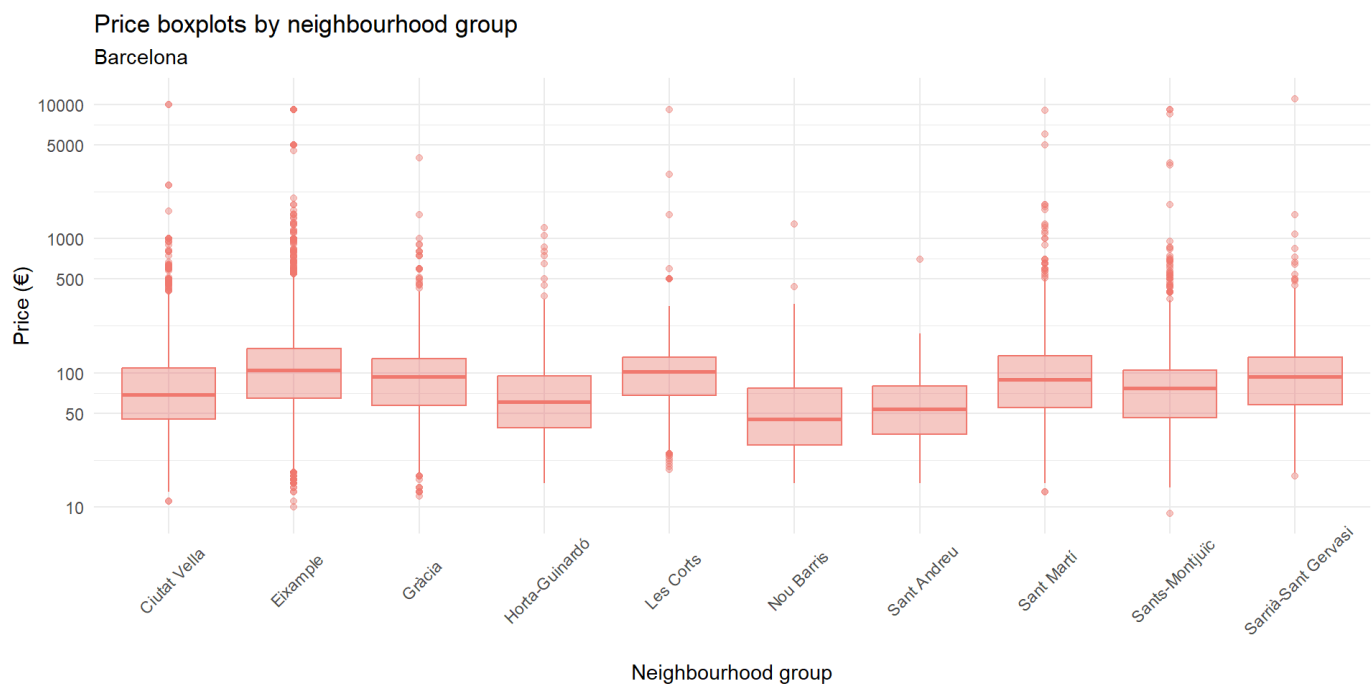
2.3.11. Price by neighbourhood group



This plot provides some further insight into the distribution of the `price` variable that is not conveyed in the previous map plot. The `price` of listings is represented as points on a scatter plot, and clustered by `neighbourhood_group` in order to reflect the actual variation of values, and locate the *center* of the distributions. Because most listings' `price` values are located around the €100 range (median is €87), the items with much higher pricing significantly skew the plot. Thus, in order to improve the readability of the *median region*, the scale has been adjusted to be logarithmic. The additional violin plot lines provide more information about the distribution and location of medians.

Corroborating observations from previous graphs, the majority of listings is seen to concentrate in the *Ciutat Vella* and *Eixample* districts, with the latter containing the highest median, as well as highest listing count, of all districts. On the other hand, the furthest-located districts of *Nou Barris* and *Sant Andreu* are seen to contain the least and overall lowest-priced listings of the city.

2.3.12. Price boxplots by neighbourhood group



The previous plot could be summarized in the above boxplot, which locates the quartiles and medians for the price variable in each neighbourhood group. As an example, the following are the medians from the neighbourhood groups mentioned in the previous plot, which represent the lowest and highest extremes in all 10:

```
median(Barcelona_md$price[Barcelona_md$neighbourhood_group == "Ciutat Vella"], na.rm = TRUE)
```

```
## [1] 69
```

```
median(Barcelona_md$price[Barcelona_md$neighbourhood_group == "Eixample"], na.rm = TRUE)
```

```
## [1] 105
```

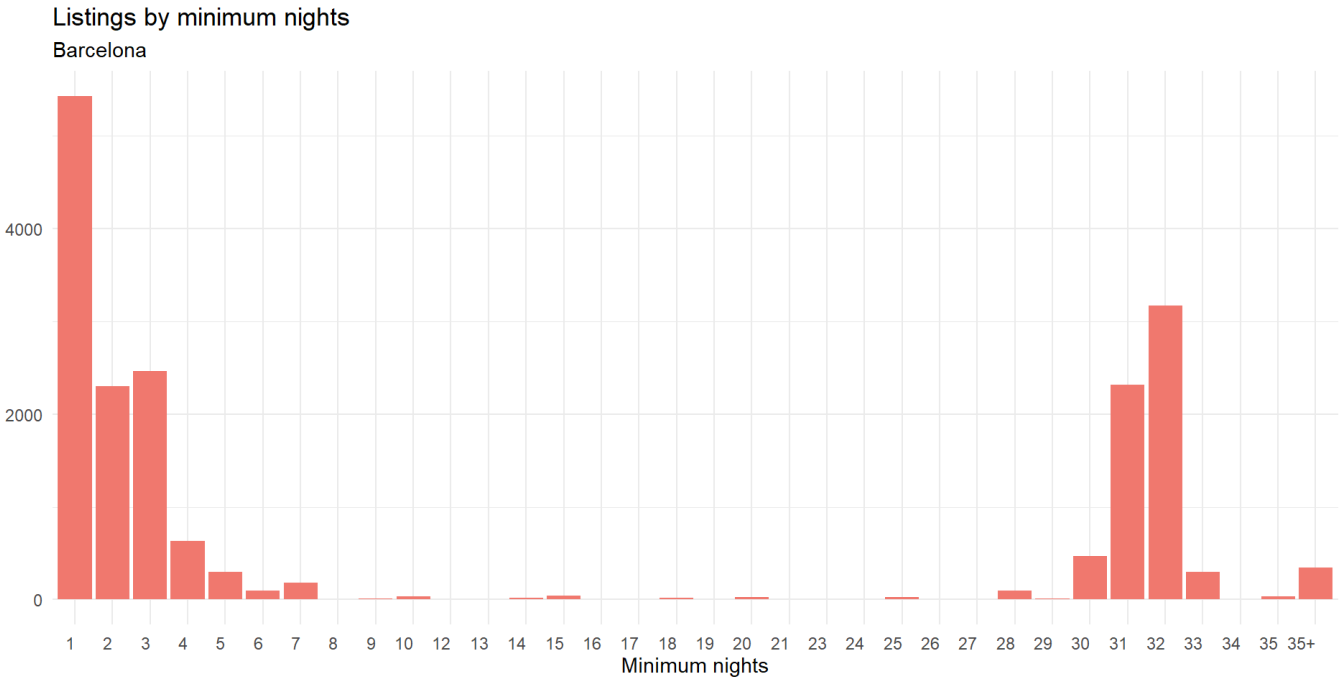
```
median(Barcelona_md$price[Barcelona_md$neighbourhood_group == "Nou Barris"], na.rm = TRUE)
```

```
## [1] 45
```

```
median(Barcelona_md$price[Barcelona_md$neighbourhood_group == "Sant Andreu"], na.rm = TRUE)
```

```
## [1] 54
```

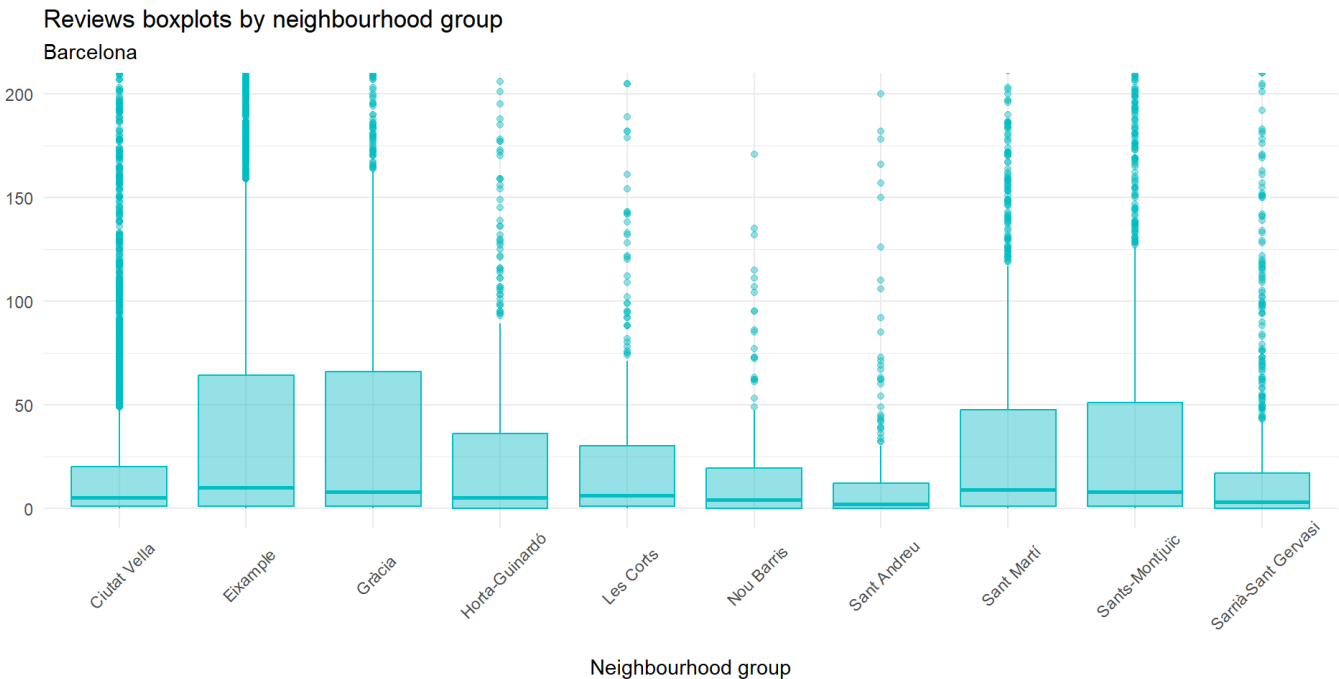
2.3.13. Listings by minimum nights



This chart examines the distribution of listing counts based on the minimum nights required, aiming to illustrate this variable throughout the entire dataset. An initial observation can be made that the majority of listings require a minimum reservation of one night, followed by a rapid decline in frequency until reaching a full week. Further on, a significant upsurge in listings happens around the 30-day mark, indicating that a substantial cluster of listings entail a minimum stay of around a month.

Such findings lend support to the argument that Airbnb serves as a portal for an increasingly diverse range of rental types, extending beyond the original notion of short-term single-room leases, which were the original hallmark of the platform.

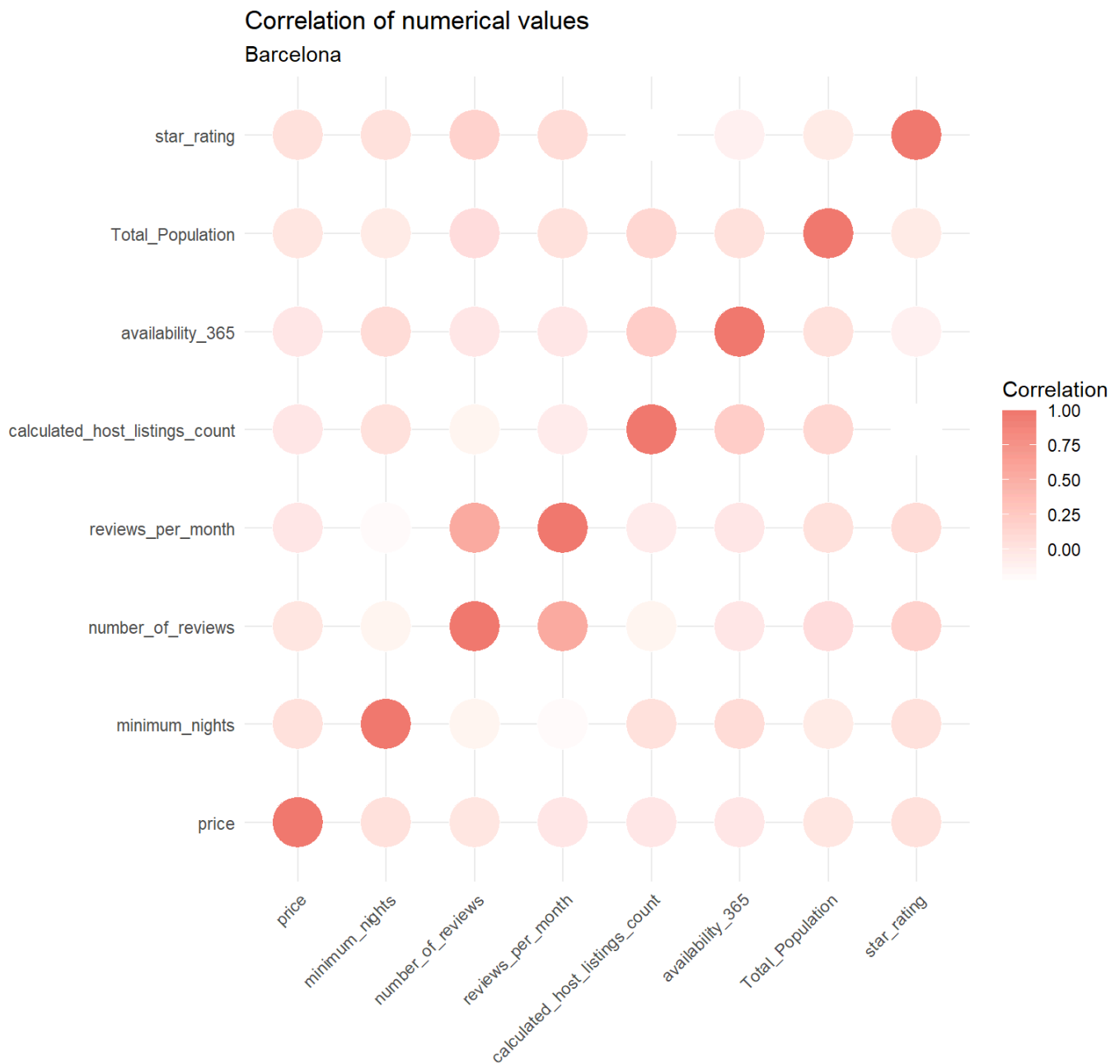
2.3.14. Reviews boxplots by neighbourhood group



This set of boxplots looks at the distribution of number of reviews listings receive, grouped by neighbourhood groups. Although not as rich in information as other plots from the report, it does reflect certain trends which have been observed in other visualizations. Assuming that the number of reviews is a reflection of the level of activity it has received, once again the chart is dominated by the most popular neighbourhoods of *Eixample*, *Gràcia*, *Sant Martí* and *Sants-Montjuïc*. An exception to this observation is *Ciutat Vella*, which is shown to have

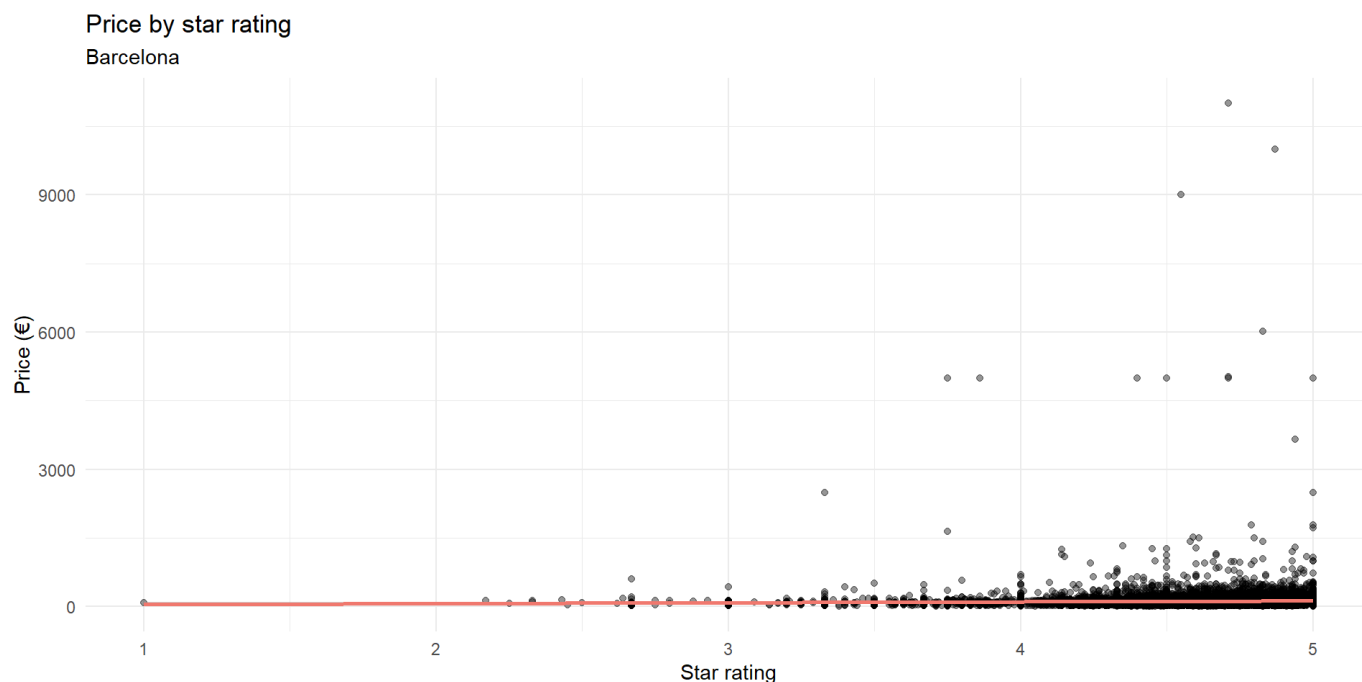
a narrow boxplot, reflecting a smaller number of reviews. This may be do to the listings having a shorter lifespan, and not having been able to accumulate as many reviews, perhaps due to a more dynamic Airbnb market in that concentrated area. This statement would however have to be proven through other analyses.

2.3.15. Correlation of numerical values



The above plot reflects the correlation of all numerical variables in the data set. A positive correlation is an indicator of a 'more is more' relationship between them, evaluated from 0 to 1. The interesting observation though is the fact that there are relatively low correlations between most variables. The single exception is the straightforward correlation between `reviews_per_month` and `number_of_reviews`, which is somewhat higher than 0.50. There are also variables shown to have no correlation at all, such as `star_rating` and `calculated_host_listings_count`, which seems also to be quite a reasonable conclusion.

2.3.16. Price by star rating



The last plot examines the behaviour of the `price` of all listings, plotted as scattered points, in relation to the `star_rating`, to explore possible correlations between the two. A general trend can be observed, that most listings tend to have high ratings, ranging from 2 to 5, forming a large cluster between the 4 and 5 marks. Also, the higher-priced listings are seen to be associated with higher star reviews. Outliers with nightly rates above €5000 are seen, skewing the perception of the overall trends of the main cluster.

The red line represents a linear regression model based on the two observations, which shows a positive linear increase in the price that is explained by the star rating. However, this causality cannot be proven only through this model. The linear model shows an intercept of 29.13, and a positive slope of 20.13 for the `star_rating` variable, indicating that for each increase star rating (1 star), the price would theoretically increase €20.13.

```
lm(price ~ star_rating, data = Barcelona_md)
```

```
##
## Call:
## lm(formula = price ~ star_rating, data = Barcelona_md)
##
## Coefficients:
## (Intercept)  star_rating
##      29.13      20.13
```

3. Additional chapter

An additional segment of the project worthy of mention are the alternative workflows and R packages that we explored, discarded or eventually implemented into our work.

One notable case was our strategy for accessing the Open Data BCN demographics data set, which relied on the Barcelona City Council's API. We opted for this path with the sole purpose of exploring alternative data loading workflows, despite the straightforward alternative of downloading and loading the `2023_pad_mdbsas_sexe.csv` file from the website being more easily available.

Among other libraries that weren't covered during the Bootcamp, we incorporated the `gridExtra` library to manipulate the `ggplot2` layout, allowing for efficient division and arrangement of layouts to optimize visualization.

For the 'Average price by neighbourhood' plot (number 10), the package `ggrepel` was tested and loaded in order to position the labels for the neighbourhood polygons in such a way that they could be easily readable.

We also experimented with various libraries, which can be seen loaded yet unused in the R code, such as `plotly`, `RColorBrewer`, `jsonlite`, `reshape2`, `leaflet`, and `tidyterra`. However, practical challenges prompted the exclusion of certain libraries from the final RMarkdown report. A notable case was the implementation of `plotly` for interactive map creation, which proved problematic when confronted with large-sized map tiles, resulting in performance issues and unwieldy HTML file sizes. Consequently, a decision was made to shift to conventional PNG images to mitigate these challenges and ensure a more and resource-efficient workflow. This iterative process underscores the importance of a cautious and technically sound approach to library selection and implementation in data visualization workflows.

As a secondary additional chapter of the project we opted for a GitHub-based workflow, which was a new experience for both of us. This allowed us to collaborate simultaneously from different devices, and ensure the reproducibility of the code, whilst keeping track of changes and version updates.

4. Conclusion

The present study has been an great opportunity to get to know a project aimed at enhancing transparency within a company recognized for its limited data openness, as well as to explore data sets and workflows from official Open Data sources. Airbnb is in itself an interesting topic and controversial subject matter, as it has had a notable impact on the short-term rental and holiday accommodation sector across the world. The choice of Barcelona as the focal point is rooted in its reputation as a distinctive case study showcasing the influence of Airbnb on heavily frequented tourist destinations. Exploring this context is essential to achieving an understanding of the challenges posed by the platform. Furthermore, aligning the findings and statistical indicators with local laws and regulations has helped add a qualitative dimension to the assessment.

Our concise analysis reveals immediate concerns regarding the compliance of listings with legal requirements and highlights the emergence of a monopolistic market in the city. It also reflects economic and demographic discrepancies within the city, which could help lend support to studies in other knowledge areas. The observed levels of habitation stress in Barcelona, attributed in part to the phenomenon of 'overtourism' and a global housing crisis, are discernible within the Airbnb data. The dataset chosen for this project not only serves as a practical application of R code but also proves to be a source of rich and insightful information extending beyond a mere coding exercise.

5. Appendix

5.1. Experiences using generative AI tools

Understanding that generative AI tools are here to stay as an invaluable companion of present-day and future programmers, we implemented them as a support and proof-checking tool, easily accelerating our productivity several times over. We observed that while quick in generating syntactically correct R code, it still requires the user to have a correct grasp of the packages being used and general R workflows. Without accurate prompts or careful interpretation of the outputs, AI remains unable to comprehend the full context of the project and address every single question adequately. Fine tuning results and providing additional context of the structure of the data sets turned out to be an essential step in our AI-powered workflow.

We employed ChatGPT versions 3.5 and 4, with the latter incorporating a Data Analyst functionality. This feature was capable amongst other things of interpreting screenshots of sample plot types and returning the corresponding `ggplot2` code, or processing the source data in csv format for a better understanding of its structure, its variable names, data types, etc. One noteworthy application was in constructing correct regex

patterns, which we would easily be able to express in natural language, but would require more knowledge and time to be written and tested manually. It is worth noting that the AI-generated code suggestions consistently aligned with our knowledge of R packages as well as the scope of the course. They not only facilitated the exploration of solutions but also prompted consideration of additional settings for certain functions that we might have overlooked initially.