

Milestone 1: Airbnb Visualization across European Cities

Imane Jennane, Harkeerat Singh Sawhney, Daniela Gjorgjieva

March 21, 2025

1 Dataset

As students who love traveling, we are using an Airbnb dataset covering ten European cities: Amsterdam, Athens, Barcelona, Berlin, Budapest, Lisbon, London, Paris, Rome, and Vienna, to help travelers choose their destinations based on their budget and needs. It includes details on price, room type, guest capacity, host status, cleanliness, and location attributes such as distance to city centers and metro stations. Also for each City there is a Weekend and Weekday dataset, to show the difference in prices and ratings.

The dataset is well-structured and does not require extensive preprocessing. However, an extra field was added "District" along with other fields. This was done to allow more analysis per city. Below are the fields which are in both of the datasets (Weekday & Weekend).

- **realSum:** The total price of the Airbnb listing. (Numeric)
- **room_type:** The type of room being offered (e.g., private, shared, etc.). (Categorical)
- **room_shared:** Whether the room is shared or not. (Boolean)
- **room_private:** Whether the room is private or not. (Boolean)
- **person_capacity:** The maximum number of people that can stay in the room. (Numeric)
- **host_is_superhost:** Whether the host is a superhost or not. (Boolean)
- **multi:** Whether the listing is for multiple rooms or not. (Boolean)
- **biz:** Whether the listing is for business purposes or not. (Boolean)
- **cleanliness_rating:** The cleanliness rating of the listing. (Numeric)
- **guest_satisfaction_overall:** The overall guest satisfaction rating of the listing. (Numeric)
- **bedrooms:** The number of bedrooms in the listing. (Numeric)
- **dist:** The distance from the city center. (Numeric)
- **metro_dist:** The distance from the nearest metro station. (Numeric)
- **attr_index:** Attraction index of the listing location. (Numeric)
- **attr_index_norm:** 0–100 normalized attraction index. (Numeric)
- **rest_index:** Restaurant index of the listing location. (Numeric)
- **rest_index_norm:** 0–100 normalized restaurant index. (Numeric)
- **lng:** The longitude of the listing. (Numeric)
- **lat:** The latitude of the listing. (Numeric)
- **district:** The district of the listing (Categorical)

2 Problem Statement

Our visualization provides an in-depth exploration of Airbnb listings across multiple cities, focusing on key trends and insights that shape the short-term rental market. Through interactive and comparative analysis, we aim to:

- **Compare price distributions across neighborhoods:** By examining how rental prices vary within and between different urban areas, we uncover pricing patterns, cost disparities, and potential factors driving these differences, such as local demand, amenities, and proximity to key attractions.
- **Identify dominant room types and their spatial distribution:** We analyze the prevalence of entire homes, private rooms, and shared spaces in various districts, highlighting which areas cater more to budget travelers, long-term stays, or luxury experiences.
- **Analyze guest satisfaction in relation to price and host status:** Our study investigates how ratings and reviews correlate with listing prices and whether factors like host experience (e.g., Superhost status) influence guest satisfaction.
- **Assess how location influences pricing and demand:** By mapping Airbnb listings alongside major urban features—such as tourist attractions, metro stations, business districts, and nightlife hotspots—we evaluate how accessibility and neighborhood characteristics impact both rental rates and booking popularity.

This visualization is designed for a diverse audience, including **travelers** seeking budget-friendly or premium stays, **urban planners** analyzing the impact of Airbnb on housing and tourism, and **data enthusiasts** exploring trends in the sharing economy. Through interactive filters and visual storytelling, we provide valuable insights into the dynamics of Airbnb listings, enabling better decision-making for both guests and hosts.

3 Exploratory Data Analysis

To ensure effective and meaningful visualizations, we have performed a thorough exploratory data analysis (EDA) to uncover key patterns, trends, and anomalies within the Airbnb dataset. Our approach has included the following steps:

- **Summary Statistics:** We have computed fundamental descriptive statistics, such as mean, median, and standard deviation, for key variables like listing price, guest satisfaction ratings, and distances to city centers or major attractions. This will provide an initial understanding of data distribution and variability. Figure 1

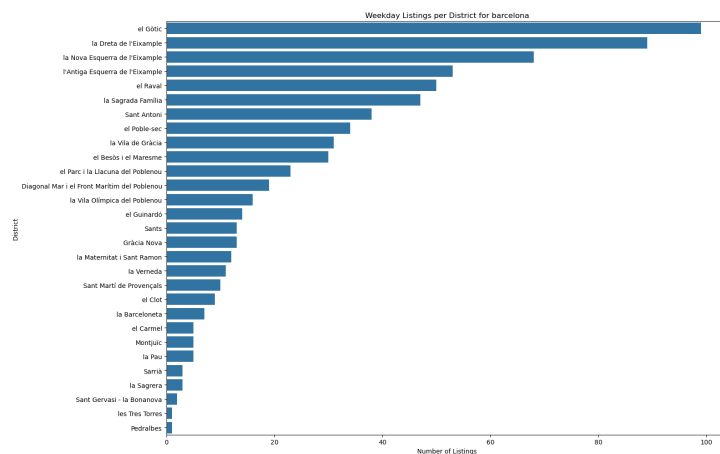


Figure 1: Visualization of Airbnb Listing Per District in Barcelona

- **Geospatial Mapping:** We have plotted Airbnb listings on interactive maps, in which we aim to visualize their spatial distribution, identifying high-density rental zones and potential gaps in coverage. This will help highlight how listings cluster around key urban features such as business districts, tourist landmarks, and transit hubs. Figure 2 show cases an interactive visualization of all the Airbnb Listing in Amsterdam and which districts they are a part of. To see the interactive version of this, it can accessed by downloading this link and viewing it in a browser.

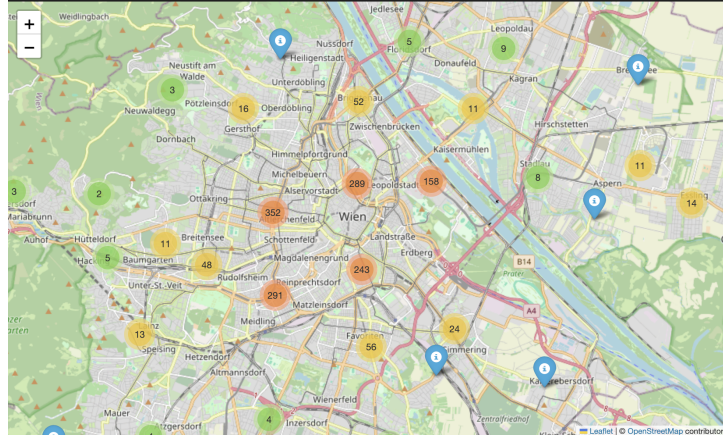


Figure 2: Map Cluster Visualization of Listings Throughout Amsterdam)

- **Correlation Analysis:** We have investigate relationships between different variables, such as price versus location, satisfaction ratings versus host experience, and proximity to amenities versus demand. Correlation matrices and scatter plots will help us determine how various factors interact and influence one another. Figure 3 shows the differences in a price of a Airbnb Listing as compare to weekdays and weekends.

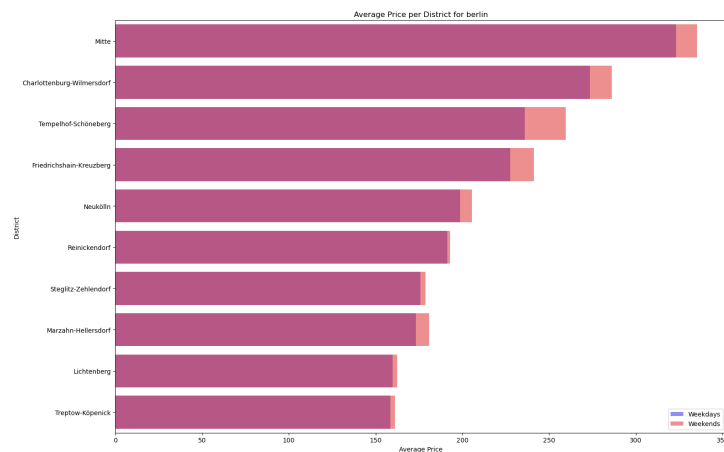


Figure 3: Difference In Average Price Per Listing in District (Weekend vs Weekday)

- **Consistency Amongst Dataset:** We have ensure that for each city, the weekend and weekday dataset are consistent, meaning they are of equal shape and contains Airbnb Accommodations which are in both Datasets.

All of this analysis can be found on the Jupiter Book called as "exploratory-data-analysis.ipynb" under the Dataset Directory. You can as well access it in the link Notebook as well.

4 What others have already done with the data?

Several existing projects have analyzed Airbnb listings, focusing on various aspects such as pricing trends, spatial distribution, and economic impact:

- **Inside Airbnb:** An open-source platform that visualizes Airbnb data to analyze housing market trends, host activity, and affordability concerns.
- **Kaggle Projects:** Numerous data science competitions and community projects have explored pricing prediction, neighborhood comparisons, and guest satisfaction analysis.
- **Urban Studies Research:** Academic papers have examined Airbnb's influence on urban housing markets, gentrification, and tourism-related economic shifts.

5 Why is your approach original?

While previous studies have primarily concentrated on single-city analyses or specific Airbnb trends, our approach introduces a broader and more comprehensive perspective. We aim to bridge gaps in existing research by leveraging a combination of multi-city comparisons, interactive geospatial mapping, and an integrated approach to understanding price, satisfaction, and location influence.

Our approach is unique because it:

- **Emphasizes multi-city comparisons:** Unlike studies that focus on a single urban area, we analyze Airbnb listings across ten European cities. This allows us to uncover regional differences, compare market dynamics across diverse locations, and highlight city-specific trends. For example, we can examine how Airbnb pricing in high-tourism cities like Paris and Rome compares to more budget-friendly options like Budapest or Athens.
- **Integrates interactive geospatial visualizations:** By mapping Airbnb listings with spatial data overlays, we move beyond static charts and tables. Users can explore how rental prices, room types, and guest satisfaction vary across neighborhoods. These visual tools help reveal spatial clustering patterns, high-demand zones, and rental hot spots influenced by urban infrastructure such as metro stations, tourist landmarks, and business districts.
- **Combines pricing, guest satisfaction, and location insights into a unified analysis:** Many studies focus solely on pricing or guest reviews in isolation. Our project interconnects these variables, allowing for a richer understanding of how they influence one another.
- **Comparing Districts within a City with Each Other** We have extended our dataset by including which District the Airbnb Accommodations are located in. With that we are able to compare amongst other Districts which one offers better value in terms price and other variables.

By combining these elements, our project offers a distinctive perspective on Airbnb listings across Europe, filling a gap in existing research and providing valuable insights for multiple user groups.

6 Our source of Inspiration

We want to make an interface in which we can represent our data in a very interactive way by using the actual geographical location. One of the examples from which we have gotten our inspiration is French Champagne Brand called Chartogne-Taillet. As soon as we enter the site, we are greeted with an interactive map that looks as if it is hand-drawn on the paper. It showcases vineyards, houses and trees from which we can choose the specific vineyard to get more information. Even-though, we do not have much knowledge about Champagne we still were able to not only have a enjoyable experience but also learn from the story which it shared with us.