UNIVERSITÀ DI PISA

Visual Analytics Report

AirBnb posting analysis
Exploring the dynamics of Airbnb postings over the years

Professor:
Salvatore Rinzivillo

Student:
Alla Usova - 660973

Academic Year 2024/2025

Table of contents

1 Introduction

Airbnb has transformed the hospitality industry by offering a peer-to-peer lodging platform that is widely used around the world. As the volume of listings continues to grow, understanding patterns in host behavior, room types, pricing, and geographic distribution becomes increasingly important for researchers, policymakers, and travelers. This project aims to provide an interactive visualization tool that allows users to explore Airbnb data across multiple dimensions, uncover trends, and support analytical tasks through intuitive and linked visual displays.

2 Research questions

This project explores the Airbnb dataset with the aim of understanding how host registrations have evolved over time across different countries. In particular, the focus is on examining geographic patterns and trends related to the types of properties listed. The central research questions that guided the design of the visualizations are:

- How has the number of Airbnb hosts changed over the years, and how does this growth vary by country?
- What are the dominant types of properties listed in different regions, and how have these preferences shifted over time?
- Can we detect signs of professionalization in certain areas (e.g., the rise of entire apartments offered)?

These questions shaped the selection of dimensions, filters, and interactive visualizations, aiming to support both exploratory analysis and targeted investigations.

3 Dataset Overview

3.1 Data description

The dataset includes over 1 500 000 Airbnb listings worldwide from the early years of the platform (2008–2015), and contains metadata such as:

- **Location (latitude, longitude, country)**
- **Room type**
- **Host ID**
- **Host since year**
- etc

For this analysis I decided to use the data from the following European countries: Italy, Germany, France, Spain, Switzerland, Austria, Netherlands and Denmark.

3.2 Key Dimensions for Analysis

The following dimensions were selected due to their analytical importance and their ability to reveal meaningful patterns in the data:

- **Country, Region, City:** to understand geographical trends
- **Property type:** to study accommodation preferences
- **Host since year:** to analyze platform growth over time

These dimensions allow for both macro-level and detailed exploration of the Airbnb ecosystem and are sufficient to address the research questions defined in this project.

3.3 Data preparation

Before visualization, the data were subjected to several pre-processing steps:

- Removal of duplicate or irrelevant listings
- Handling of missing or malformed values for the region and city by finding the dataset with zipcodes and joining it to our Kaggle dataset
- Conversion of date formats and extraction of relevant time features (year, month)

At first, I worked with a smaller CSV file to make things easier—it had about 5 entries per combination of Country, City, Host Since Year, and Property Type. This helped me quickly build and test my visualizations.

But later, I realized I needed the full dataset for a complete analysis. The problem was, the full CSV was way too big and slow to work with. So I decided to pre-aggregate the data into a JSON structure. The aggregations were done across the main dimensions: Country, City, Year, and Property Type.

This reduced the data volume and made the application load much faster. Instead of computing aggregates on the fly in the browser, most of the heavy work was done once in Python and then stored efficiently.

4 State-of-art

I referenced and improved upon:

InsideAirbnb.com - Offers dashboards with tabular insights as well as a map, but the data they are analyzing is just for the last year and focusing mostly on static data not visualizing the dynamics.

Airbnb dashboards in Tableau - Mostly bar/pie charts without geographic interactivity or with geographical interactivity, but focusing mostly on prices

Our project aims to improve upon these limitations by introducing dynamic filtering, linked brushing, and original geographic visualization using D3.js. and offers an integrated and interactive exploration experience focusing on the number of hosts across the years.

5 Design Choices and Implementation

As I am primary working with geografical data, the map seems necessary step to visualize the property postings. Then for the distributions of data across other dimension I decided to use barchart as it is one of the best ways to visualize different types of data.

As the dataset has strong spatial and temporal dimensions, it was important to design visualizations that support both a high-level overview and detailed inspection. The main focus was placed on clarity, responsiveness, and intuitive interaction. I used Vue.js as the front-end framework and D3.js/Plotly for rendering charts and maps.

5.1 Map

At the center of the interface is a custom-built interactive map created with D3.js. Each country is shown as a circle, where the radius is proportional to the number of Airbnb listings, and the color represents the most common property type in that country. The color scheme is consistent across all views, allowing users to mentally connect visual elements.

A zoom mode is available for more detailed exploration. When activated, the user can view grouped listings as individual points positioned according to their geographic coordinates. These points represent aggregated data across the selected dimensions (such as city, property type, and year). Hovering over a point displays a tooltip with contextual information, including the city, property type, and the year when the listings were recorded.

5.2 Bar charts

The bar charts are implemented using Plotly and offer smooth transitions, tooltips, and dynamic updates. Each chart shows a different aspect of the dataset:

- **Number of hosts over time**: a horizontal bar chart grouped by year
- **Top countries or cities by host count**: depending on the filter, it shows either countries or top 10 cities in the selected country
- **Distribution of property types**: stacked bar or grouped by country
- **Year-on-year growth of listings**: derived metric visualized as bar height

By default, the bar charts are sorted in ascending order based on the key. However, if the sortDesc prop is passed, the data is instead sorted in descending order by value. This behavior helps users interpret the data more quickly and effectively. To improve readability and avoid visual clutter, the city-level chart displays only the top 10 cities based on the number of hosts and the same is valid for Property Type distribution chart.

I also accounted for cases where long bars could cause overlapping or hidden text labels. A maximum bar width was introduced so that value labels remain legible and always appear outside the bars. All charts update in response to interactions with the filters.

I also applied consistent color scales across charts, using a palette that complements Airbnb's brand red ('FF5A5F'). For example, '5AC8FF' (soft cyan) and 'FFA07A' (peach) were selected to differentiate categories while preserving accessibility and contrast.

5.2.1 Increase rate calculation

One of the key metrics computed in this project is the increase rate of host registrations over time. This metric helps to quantify the growth trend and identify moments of acceleration or stagnation in Airbnb's expansion.

The increase rate was computed globally as well as at every level that could be chosen with enabling the filters, allowing for comparisons across different markets.

This metric was visualized using a bar chart, which clearly showed:

Rapid growth in the early years of the platform, particularly between 2008-2012 A slowdown or plateau in some regions after 2013 and even decrease after 2015

Overall, the increase rate is a powerful and intuitive indicator for understanding temporal dynamics, and it plays a central role in this project's analytical framework.

5.3 Filters interaction logic

A key feature of the interface is the coordinated interaction between filters and visualizations.

Dropdown menus filters allow users to dynamically refine the dataset based on selected criteria such as country, year, and room type. These filters are implemented using Vue's 'v-model' two-way data binding, which ensures that any change in user input is immediately reflected across all relevant components.

To manage and coordinate the filtering logic, I used the Crossfilter.js library. Crossfilter provides fast multidimensional filtering and grouping capabilities, which allowed us to define dimensions for Country, City, Year, and Property Type. Once a user interacts with a filter, Crossfilter recalculates the subset of the data in real-time. This updated data is then passed to each visualization component, such as the map or bar charts, using Vue's reactive props.

This architecture made it straightforward to maintain consistency across visualizations without manually managing state or triggering updates. For example, selecting a specific country will not only update the bar chart to show top cities within that country, but also zoom the map and adjust the room type distribution accordingly.

In addition, I applied conditional logic to the layout. When "All" countries are selected, the system displays a country-level bar chart; when a specific country is selected, the interface automatically switches to a chart of top cities in that country. This makes the experience more tailored and reduces cognitive load for the user.

6 Analytical use case and results

This project aimed to explore patterns in Airbnb host activity using interactive visualizations. Below, I revisit each research question and summarize the insights obtained from the analysis:

- **How has the number of Airbnb hosts changed over the years, and how does this growth vary by country?** The line chart revealed a clear upward trend in host registrations globally, with particularly rapid growth observed in the first 5 years after the platform launch. Countries like Italy and France showed a sharp increase, indicating a growing interest in short-term rentals in these regions.

- **Which countries and cities have the highest concentration of hosts?** The geographic map and country-level bar chart showed that France have the highest number of hosts reflecting high tourism demand or local hosting culture.

- **What are the dominant types of properties listed in different regions, and how have these preferences shifted over time?**

  The property type distribution chart showed that "Apartment" option is the most common listing type in all countries. However, "Bed and Breakfast" is still popular in Italy, and the "House" option is popular in Spain and Switzerland. Over time, there has been a visible shift towards more "entire apartment" listings, especially in urban areas.

- **Can we detect signs of professionalization in certain areas (e.g., the rise of entire apartments offered)?**

  Yes, there are signs of professionalization in certain areas, especially urban regions. In many major cities, entire apartments have consistently been the dominant property type over the years. While this in itself is not a new trend, what stands out is the continuous growth in the number of newly registered hosts offering this type of accommodation. This suggests that the platform is increasingly attracting users who may treat Airbnb as a professional or semi-professional activity, rather than occasional peer-to-peer hosting.

Overall, the interactive visualizations made it easier to identify temporal and spatial trends in Airbnb host activity. The design supports both a high-level overview and more detailed analysis through filtering and zooming.

Thanks to the coordinated views, the user can switch between countries, track yearly changes, and immediately see how the distribution of property types varies — all in one interface. This makes the tool powerful not just for static reporting, but also for exploratory analysis.

While the current implementation allows users to explore host dynamics and property type distributions across countries and time, it does not include economic indicators such as price trends, occupancy rates, or revenue per listing. Including these aspects could enrich the analysis and support more business-oriented or policy-related decisions. In the future, this project could be extended to incorporate temporal pricing data, user review scores, or even seasonal patterns. Additionally, adding clustering or anomaly detection techniques might help highlight unusual hosting behaviors or market disruptions in specific regions.