

Airbnb Data: Amsterdam 2023 Analysis Project

Katja Pönigk

Data Source: Airbnb Amsterdam Data, 2023

Data Sourcing & Summary

This data has been sourced from the [,Inside Airbnb'](#) website.

Airbnb is an online marketplace for lodging, primarily home-stays for vacation rentals, and tourism activities. It provides a platform for hosts to share accommodations with travelers around the world. There are over 6M active listings available on Airbnb platform today, 4M+ hosts provide listings on Airbnb, and over 1B+ Airbnb guest arrivals all-time (as of Sept. 28, 2021).

I have chosen this data because I am interested in and passionate about traveling and have been using the services of Airbnb myself before. I am fascinated how the company has evolved over the last years. Additionally, will I join a Travel AI startup from next month on and think insights from exploring an Airbnb data set could be beneficial.

Data Collection

The data is collected by Inside Airbnb. Inside Airbnb is a mission driven project that provides data and advocacy about Airbnb's impact on residential communities. New data is updated quarterly for the last for each region and is available for free to download on their website.

Data Contents

The Airbnb dataset "Amsterdam, North Holland, The Netherlands for 12 December, 2023" was accessed from [,Inside Airbnb'](#) on 16/01/2023.

Amsterdam, North Holland, The Netherlands

12 December, 2023 ([Explore](#))

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

Has 7 CSV files containing data about listings, reviews, neighbourhoods (incl. a geojson file) and calendar. The listings_details file contains additional variables for Airbnb in Amsterdam. The calendar has 365 records for each listing. It specifies the whether the listing is available on a

particular day (365 days ahead), and the price on that day. In addition, a reviews file is available (text mining opportunities), and a geojson file of the neighbourhoods in Amsterdam.

The column description can be found in the : [Inside Airbnb Data Dictionary](#)

Data Limitations

The dataset does not include information on external factors that could impact the Airbnb market, such as economic changes, global events, or public health crises. Another limitation is the quality of reviews. The quality of reviews can vary, and some reviews may be biased or inaccurate. Depending solely on reviews for analysis might not provide a complete and objective picture of guest satisfaction. Another limited factor might be the price. We need to be aware that the dataset may not capture all the factors influencing pricing decisions, such as short-term events, local festivals, or changes in local regulations.

Data Ethics

No private information is being used in the data. All the information here are public knowledge.

Data Cleaning and Profiling

Data Cleaning, Wrangling, Consistency Checks

Column	Type of Inconsistency/Wrangling	Comment
	no duplicates found	
	no columns renamed	
host_name, neighbourhood_group, reviews_per_month, license, host_url, host_since, host_location, host_about, host_thumbnail_url, host_verifications, neighbourhood_group_cleansed, bathrooms, bathrooms_text, beds, amenities, calendar_updated, has_availability, instant_bookable, calculated_host_listing_count, calculated_host_listing_count_entire_homes, calculated_host_listing_count_private_rooms, calculated_host_listing_count_shared_rooms, review_per_month	columns dropped	columns being dropped due being unnecessary for his analysis
price, host_response_rate, review_scores_location, host_is_superhost, last_review	missing values - 2,933 NaN values	since the missing values in these columns made up less than 10% of our data set, we deleted them.

host_response_time	column type changed	changed from mixed type to string
last_review, first_review	column type changed	changed from INT to date
host_response_rate	column type changed	changed from mixed type to float

Data Profiling

Data after cleaning & consistency check:

- 2023_Amsterdam_Airbnb_Data_csv
- Total: 5050 rows, 28 columns

#	Column	Non-Null	Count	Dtype
0	index	5050	non-null	int64
1	id	5050	non-null	int64
2	name	5050	non-null	object
3	host_id	5050	non-null	int64
4	neighbourhood	5050	non-null	object
5	latitude	5050	non-null	float64
6	longitude	5050	non-null	float64
7	room_type	5050	non-null	object
8	price	5050	non-null	float64
9	minimum_nights	5050	non-null	int64
10	number_of_reviews	5050	non-null	int64
11	last_review	5050	non-null	datetime64
12	reviews_per_month	5050	non-null	float64
13	calculated_host_listings_count	5050	non-null	int64
14	availability_365	5050	non-null	int64
15	number_of_reviews_ltm	5050	non-null	int64
16	host_response_rate	5050	non-null	float64
17	host_since	5050	non-null	object
18	host_response_time	5050	non-null	object
19	property_type	5050	non-null	object
20	accommodates	5050	non-null	int64
21	first_review	5050	non-null	datetime64
22	review_scores_location	5050	non-null	float64
23	review_scores_communication	5050	non-null	float64
24	review_scores_rating	5050	non-null	float64
25	maximum_nights	5050	non-null	int64
26	host_is_superhost	5050	non-null	float64
27	host_has_profile_pic	5050	non-null	int64

Key Questions

- Which neighborhoods are the most popular, receive the most bookings?
- What is the average list price of an Airbnb listing?
- How is the distribution of Airbnb listings across different neighborhoods in Amsterdam? (map)
- Which rental property types (e.g. private rooms, house, etc.) are available and are the most booked?

- Are there any trends or notable patterns in booking activity throughout a year?
- Are there any notable correlations between specific features (e.g., location, property type, reviews) and the overall satisfaction of guests, as reflected in the review scores?
- (Do reviews effect listing price?)