

Project Brief

Dulce Valcarcel

1. Objective

tbd at the end

2. Content

tbd at the end

3. Source

The Airbnb report for Berlin in 2024 provides valuable insights into the city's short-term rental market, utilizing data compiled by Inside Airbnb from the official Airbnb website. This comprehensive report is a result of thorough analysis and exploration of the available data, shedding light on various aspects of the Airbnb landscape in Berlin. The information presented covers details such as property listings, availability, host characteristics, and other relevant metrics. The dataset, sourced from <http://insideairbnb.com/explore>, serves as a reliable and up-to-date resource for understanding the dynamics of the short-term rental market in Berlin as of January 15, 2024.

4. Collection

The Airbnb report for Berlin in 2024 draws upon publicly available information meticulously compiled from the Airbnb website, specifically from the platform <http://insideairbnb.com/explore>. The primary objective of harnessing this data is to undertake a comprehensive quantification of the impact that short-term

rentals have on the local housing landscape and residential communities within the city. By delving into the wealth of information made publicly accessible by Inside Airbnb, the report aims to unravel the intricate dynamics between short-term rental listings and the broader housing ecosystem in Berlin. This involves an in-depth analysis of property listings, availability patterns, host characteristics, and various other metrics that collectively contribute to a nuanced understanding of how the proliferation of short-term rentals influences the local housing market and residential neighborhoods. The report, underpinned by its commitment to transparency and reliance on factual data, serves as an invaluable resource for policymakers, researchers, and the general public seeking to grasp the multifaceted impact of short-term rentals on the city's housing landscape as of January 15, 2024.

5. Content

The dataset utilized in the Airbnb report for Berlin in 2024 encompasses a rich compilation of information that spans the breadth of the short-term rental landscape. It provides an exhaustive listing of properties available for rent in Berlin, offering a detailed glimpse into the diverse array of accommodations that the city has to offer. This encompasses a wide spectrum of property types, ranging from apartments and houses to unique and specialized lodging options. The dataset further extends its scope by incorporating an availability calendar, projecting the occupancy status for the upcoming 365 days. This forward-looking feature allows for a comprehensive analysis of the temporal patterns in property availability, shedding light on the ebb and flow of demand throughout the year.

Additionally, the dataset includes a wealth of user reviews, providing valuable insights into the guest experience for each listing. This aspect not only aids prospective renters in making informed decisions but also contributes to a holistic understanding of the quality and desirability of different properties in Berlin. By encapsulating these key elements—property listings, availability calendars, and user reviews—the dataset forms a robust foundation for exploring the dynamics of the short-term rental market in Berlin for the entirety of 2024.

a) Datasets

- **Listing long** → id, host_response_time, host_response_rate, host_is_superhost, property_type, accommodates, maximum_nights, first_review, review_scores_rating, review_scores_accuracy, review_scores_cleanliness, review_scores_checkin, review_scores_communication, review_scores_location, review_scores_value, instant_bookable
- **Listing short** → id, name, host_id, host_name, neighbourhood_group, neighborhood, latitude, longitude, room_type, price, minimum_nights, number_of_reviews, last_review, reviews_per_month, calculated_host_listings_count, availability_365
- **Calendar** → Id, date, month, available, adjusted_price, season
- **Reviews** → Id, date, reviewer_id, reviewer_name, comments

6. Limitations and Ethics

The dataset utilized in this report is a snapshot capturing the state of Airbnb listings as of December 18, 2023. It is important to note that the nature of this dataset being a snapshot means that the information related to future listings is subject to change, as it may be updated or modified by Airbnb after the designated date. This temporal dimension underscores the dynamic nature of the short-term rental market, where property availability, pricing, and other listing details can evolve over time.

It is crucial to emphasize that the dataset is curated with the utmost respect for user privacy. No private information is leveraged in this analysis, ensuring compliance with ethical data handling practices. Location information associated with the listings has been anonymized to safeguard the identity and privacy of both hosts and guests. This anonymization process is undertaken to prioritize data security and uphold the privacy standards of the individuals involved in the Airbnb ecosystem.

While every effort is made to compile accurate and reliable information, it is important to acknowledge that the accuracy of the data sourced from the Airbnb site is not the sole responsibility of Inside Airbnb. As with any large-scale

platform, discrepancies or inaccuracies may arise, and users should be mindful of potential variations in the data. Moreover, it is essential to consider that some hosts might not consistently update their calendars, leading to potential discrepancies in the availability information provided. This acknowledgment serves as a reminder that, despite rigorous data curation, external factors such as user behavior and platform updates can influence the precision of the information presented in this report.

7. Hypothesis

a) **Pricing, Accommodation Types, and Neighborhood Dynamics:** Delving into the Berlin Airbnb dataset, this analysis focuses on the interplay between pricing, varied accommodation types, and neighborhood influences. The hypothesis suggests a distinct connection between short-term rental prices, the diversity of accommodation offerings, and the specific characteristics of different neighborhoods. By unraveling patterns in pricing linked to accommodation types and pinpointing how neighborhoods contribute to cost variations, this exploration aims to provide concise insights into the intricate dynamics of short-term rentals in Berlin.

- Are certain types of listings especially popular? type and accommodation
- How is the listing distribution per neighborhood?
- How is the price distribution per neighborhood?
- Which listings are more profitable? And per neighborhood?

b) **Impact of Short-Term Rentals on Housing:** The dataset allows for the exploration of the impact of short-term rentals on the local housing landscape and residential communities in Berlin. The hypothesis is that there is a discernible influence of short-term rental listings on the availability, pricing, and overall dynamics of the housing market.

- How do short-term rentals impact the local housing landscape in Berlin?

- Are there noticeable changes in housing availability? (deficit)
- What is the correlation between short-term rental density and residential community dynamics? Population
- How rent-prices have changed in Berlin in the last 5 years?

c) **Host Characteristics and Listing Quality:** An assumption can be made that certain host characteristics, such as being a superhost or having a high response rate, may correlate with the quality of listings. The hypothesis is that hosts with positive characteristics are likely to provide better-maintained and more desirable properties, contributing to higher ratings and increased demand.

- What role do host characteristics play in the quality of listings and guest satisfaction?
- Is there a correlation between host superhost status or high response rates and positive guest reviews?
- Do certain host characteristics contribute to higher demand for listings?
- How do reviews and ratings contribute to the perceived quality of listings?

d) **Temporal Patterns in Property Availability:** The dataset includes an availability calendar, prompting the hypothesis that there are temporal patterns in property availability. Seasonal variations and trends throughout the year may influence the demand for short-term rentals, affecting pricing and occupancy rates.

- Are there temporal patterns in property availability, and how do they influence pricing and occupancy rates?
- What are the peak periods of demand for short-term rentals in Berlin?
- How do pricing trends align with the availability calendar throughout the year?

8. Key Questions

- a) How do short-term rentals impact the local housing landscape in Berlin?
 - Are there noticeable changes in housing availability?
 - What is the correlation between short-term rental density and residential community dynamics?

- b) What role do host characteristics play in the quality of listings and guest satisfaction?
 - Is there a correlation between host superhost status or high response rates and positive guest reviews?
 - Do certain host characteristics contribute to higher demand for listings?

- c) Are there temporal patterns in property availability, and how do they influence pricing and occupancy rates?
 - What are the peak periods of demand for short-term rentals in Berlin?
 - How do pricing trends align with the availability calendar throughout the year?

9. Data Profile

Dataset: listing_short

Summary information and metrics for listings in Berlin

VARIABLES INFORMATION

Variables	Description	Data type	Comments
id	Airbnb's unique identifier for the listing	Qualitative, nominal	
name	Name of the listing item	Qualitative, nominal	
host_id	Airbnb's unique identifier for the host/user	Qualitative, nominal	
host_name	Personal name of the host	Qualitative, nominal	
neighbourhood_group	The neighbourhood group as geocoded using the latitude and longitude against neighborhoods as defined by open or public digital shapefiles.	Qualitative, nominal	German characters were transform to normal ones for easy readability (ä > ae, ö > oe, ü > ue, ß > ss)

neighbourhood	The neighbourhood as geocoded using the latitude and longitude against neighborhoods as defined by open or public digital shapefiles.	Qualitative, nominal	German characters were transform to normal ones for easy readability (ä > ae, ö > oe, ü > ue, ß > ss)
latitude	Uses the World Geodetic System (WGS84) projection for latitude	Qualitative, ordinal	
longitude	Uses the World Geodetic System (WGS84) projection for longitude	Qualitative, ordinal	
room_type	<p>All homes are grouped into the following three room types:</p> <ul style="list-style-type: none"> - Entire place - Private room - Shared room 	Qualitative, nominal	
price	daily price in local currency (€)	Quantitative, continuous	
minimum_nights	Minimum number of night stay for the listing	Qualitative, ordinal	
number_of_reviews	The number of reviews the listing has	Qualitative, ordinal	
last_review	The date of the last/newest review	Date (d/m/yyyy)	

reviews_per_month	The number of reviews the listing has over the lifetime of the listing	Qualitative, ordinal	
calculated_host_listings_count	The number of listings the host has in the current in the neighbourhood	Qualitative, ordinal	
availability_365	Number of days this listing is available in the future. Note a listing may be available because it has been booked by a guest or blocked by the host.	Qualitative, ordinal	
number_of_reviews_ltm	The number of reviews the listing has (in the last 12 months)	Qualitative, ordinal	
license	The licence/permit/registration number according with Berlin law	Qualitative, nominal	

Dataset: listing_long			
Detailed information and metrics for listings in Berlin			
VARIABLES INFORMATION			
Variables	Description	Data type	Comments
id	Airbnb's unique identifier for the listing	Qualitative, nominal	It will be deleted
listing_url	URL to the Airbnb listing	Qualitative, nominal	

scrape_id	Inside Airbnb "Scrape" this was part of	Qualitative, nominal	It will be deleted
last_scraped	UTC. The date and time this listing was "scraped".	Date (d/m/yyyy)	It will be deleted
name	Name of the listing item	Qualitative, nominal	It will be deleted
description	Detailed description of the listing	Qualitative, nominal	It will be deleted
neighborhood_overview	Host's description of the neighbourhood	Qualitative, nominal	It will be deleted
picture_url	URL to the Airbnb hosted regular sized image for the listing	Qualitative, nominal	It will be deleted
host_id	Airbnb's unique identifier for the host/user	Qualitative, nominal	It will be deleted
host_url	The Airbnb page for the host	Qualitative, nominal	It will be deleted
host_name	Personal name of the host	Qualitative, nominal	It will be deleted
host_since	The date the host/user was created/ registered.	Date (d/m/yyyy)	It will be deleted
host_location	The host's self reported location	Qualitative, nominal	It will be deleted
host_about	Description about the host	Qualitative, nominal	
host_response_time	Average time of response the host take to answer questions	Quantitative, continuous	
host_response_rate	Average value of answers provided by the host	Quantitative, continuous	

host_acceptance_rate	That rate at which a host accepts booking requests	Quantitative, continuous	It will be deleted
host_is_superhost	Whether a host is highly-rated, experienced, reliable, and responsive (t/f)	Boolean	
host_thumbnail_url	Link to the small image representation of a larger image of the host	Qualitative, nominal	It will be deleted
host_picture_url	Link to the larger image of the host	Qualitative, nominal	It will be deleted
host_neighbourhood	Host's description of the neighbourhood	Qualitative, nominal	It will be deleted
host_listings_count	The number of listings the host has (per Airbnb calculations)	Qualitative, ordinal	It will be deleted
host_total_listings_count	The number of listings the host has (per Airbnb calculations)	Qualitative, ordinal	It will be deleted
host_verifications	Verifications provided by the host	Qualitative, nominal	It will be deleted
host_has_profile_pic	Host's profile picture	Qualitative, nominal	It will be deleted
host_identity_verified	Verifications provided by Airbnb	Qualitative, nominal	It will be deleted
neighbourhood	The neighbourhood group as geocoded using the latitude and longitude against neighborhoods as defined by open or public digital shapefiles.	Qualitative, nominal	It will be deleted

neighbourhood_cleanse d	The neighbourhood as geocoded using the latitude and longitude against neighborhoods as defined by open or public digital shapefiles.	Qualitative, nominal	It will be deleted
neighbourhood_group_ cleanse	The neighbourhood group as geocoded using the latitude and longitude against neighborhoods as defined by open or public digital shapefiles.	Qualitative, nominal	It will be deleted
latitude	Uses the World Geodetic System (WGS84) projection for latitude and longitude.	Qualitative, ordinal	It will be deleted
longitude	Uses the World Geodetic System (WGS84) projection for latitude and longitude.	Qualitative, ordinal	It will be deleted
property_type	Self selected property type. Not normalised by Airbnb.	Qualitative, nominal	
room_type	<p>All homes are grouped into the following three room types:</p> <ul style="list-style-type: none"> - Entire place - Private room - Shared room 	Qualitative, nominal	It will be deleted
accommodates	The maximum capacity of the listing. Maximum number of person.	Qualitative, ordinal	
bathrooms	The number of bathrooms in the listing	Qualitative, ordinal	It will be deleted

bathrooms_text	On the Airbnb web-site, the bathrooms field has evolved from a number to a textual description. For older scrapes, bathrooms is used.	Qualitative, nominal	It will be deleted
bedrooms	The number of bedrooms	Qualitative, ordinal	It will be deleted
beds	The number of bed(s)	Qualitative, ordinal	It will be deleted
amenities	Useful feature or facility of a building or place	Qualitative, nominal	It will be deleted
price	daily price in local currency (€)	Quantitative, continuous	It will be deleted
minimum_nights	Minimum number of night stay for the listing	Qualitative, ordinal	It will be deleted
maximum_nights	Maximum number of night stay for the listing	Qualitative, ordinal	
minimum_minimum_nights	the smallest minimum_night value from the calender 2024	Qualitative, ordinal	It will be deleted
maximum_minimum_nights	the largest minimum_night value from the calender 2024	Qualitative, ordinal	It will be deleted
minimum_maximum_nights	the smallest maximum_night value from the calender 2024	Qualitative, ordinal	It will be deleted
maximum_maximum_nights	the largest maximum_night value from the calender 2024	Qualitative, ordinal	It will be deleted
minimum_nights_avg_ntm	the average minimum_night value from the calender 2024	Qualitative, ordinal	It will be deleted

maximum_nights_avg_n tm	the average maximum_night value from the calendar 2024	Qualitative, ordinal	It will be deleted
calendar_updated	Last date when the calendar was updated by the host	Date (d/m/yyyy)	It will be deleted
has_availability	Whether the listing is available [t=true; f=false]	Boolean	It will be deleted
availability_30	Number of days this listing is available in a period of 30 days. Note a listing may be available because it has been booked by a guest or blocked by the host.	Qualitative, ordinal	It will be deleted
availability_60	Number of days this listing is available in a period of 60 days. Note a listing may be available because it has been booked by a guest or blocked by the host.	Qualitative, ordinal	It will be deleted
availability_90	Number of days this listing is available in a period of 90 days. Note a listing may be available because it has been booked by a guest or blocked by the host.	Qualitative, ordinal	It will be deleted
availability_365	Number of days this listing is available in the next 365 days. Note a listing may be available because it has been booked by a guest or blocked by the host.	Qualitative, ordinal	It will be deleted
calendar_last_scraped	Last date when the calendar was scrapped from Airbnb site	Date (d/m/yyyy)	It will be deleted
number_of_reviews	The number of reviews the listing has	Qualitative, ordinal	It will be deleted

number_of_reviews_ltm	The number of reviews the listing has (in the last 12 months)	Qualitative, ordinal	It will be deleted
number_of_reviews_l30d	The number of reviews the listing has (in the last 30 days)	Qualitative, ordinal	It will be deleted
first_review	The date of the first/oldest review	Date (d/m/yyyy)	
last_review	The date of the last/newest review	Date (d/m/yyyy)	It will be deleted
review_scores_rating	Average of all ratings published on Airbnb for each listing	Quantitative, continuous	
review_scores_accuracy	Average of all ratings related to accuracy published on Airbnb for each listing	Quantitative, continuous	
review_scores_cleanliness	Average of all ratings related to the cleaning parameters published on Airbnb for each listing	Quantitative, continuous	
review_scores_checkin	Average of all ratings related to the checking published on Airbnb for each listing	Quantitative, continuous	
review_scores_communication	Average of all ratings related to the communication with the host published on Airbnb for each listing	Quantitative, continuous	
review_scores_location	Average of all ratings related to the location of the listing published on Airbnb for each listing	Quantitative, continuous	
review_scores_value	Average of all ratings related to the value published on Airbnb for each listing	Quantitative, continuous	

license	The licence/permit/registration number according with Berlin law	Qualitative, nominal	It will be deleted
instant_bookable	Whether the guest can automatically book the listing without the host requiring to accept their booking request. [t=true; f=false]	Boolean	An indicator of a commercial listing
calculated_host_listings_count	The number of listings the host has in the current scrape, in the city/region geography.	Qualitative, ordinal	It will be deleted
calculated_host_listings_count_entire_homes	The number of Entire home/apt listings the host has in the current scrape, in the city/region geography	Qualitative, ordinal	It will be deleted
calculated_host_listings_count_private_rooms	The number of Private room listings the host has in the current scrape, in the city/region geography	Qualitative, ordinal	It will be deleted
calculated_host_listings_count_shared_rooms	The number of Shared room listings the host has in the current scrape, in the city/region geography	Qualitative, ordinal	It will be deleted
reviews_per_month	The number of reviews the listing has over the lifetime of the listing	Qualitative, ordinal	It will be deleted

Dataset: calendar

Detailed calendar data shows the bookings per listing object in 2024 in Berlin

VARIABLES INFORMATION			
Variables	Description	Data type	Comments
id	Airbnb's unique identifier for the listing	Qualitative, nominal	
date	The booked date in the listing's calendar	Date (d/m/yyyy)	
month	Month of the booking	Qualitative, nominal	Derivated column
available	Whether the date is available for a booking or not (t/f)	Boolean	
adjusted_price			It will be deleted

Dataset: reviews			
Detailed review data submitted by the bookers for each listing			
VARIABLES INFORMATION			
Variables	Description	Data type	Comments
listing_id	Airbnb's unique identifier for the listing	Qualitative, nominal	

id	Airbnb's unique identifier for the review	Qualitative, nominal	
date	Submission date of the review	Date (d/m/yyyy)	
reviewer_id	Airbnb's unique identifier for the person who left the review	Qualitative, nominal	
reviewer_name	Name of the person who left the review	Qualitative, nominal	
comments	Text containing the review submitted	Qualitative, nominal	