

UNIVERSITY OF CALIFORNIA, LOS ANGELES

Exploratory Data Analysis of Airbnb Listings

A report submitted in partial fulfillment of the requirements for
Statistics 20

by

Isabelle Fortaleza-Tan

University of California, Los Angeles

June 2026

1. Introduction

This report presents an exploratory data analysis of Airbnb listings. The dataset contains information about listing characteristics, host characteristics, and property information. The goal of this analysis is to explore factors associated with Airbnb listings and identify interesting patterns within the data.

2. Dataset Description

The dataset used in this analysis contains information on Airbnb listings, including listing characteristics, host attributes, pricing information, and review metrics. The original dataset contains 279,712 observations and 33 variables. Variables include information on room type, property type, number of bedrooms, accommodation capacity, nightly price, host response behavior, and guest review scores.

To focus the analysis on factors associated with listing prices, a subset of variables was selected for further exploration. Listings with missing prices or non-positive prices were removed from the analysis. Because the dataset contains several extreme price outliers, observations above the 99th percentile price (\$6,500 per night) were excluded from visualizations where appropriate to improve readability.

[1]	"listing_id"	"name"
[3]	"host_id"	"host_since"
[5]	"host_location"	"host_response_time"
[7]	"host_response_rate"	"host_acceptance_rate"
[9]	"host_is_superhost"	"host_total_listings_count"
[11]	"host_has_profile_pic"	"host_identity_verified"
[13]	"neighbourhood"	"district"
[15]	"city"	"latitude"
[17]	"longitude"	"property_type"
[19]	"room_type"	"accommodates"
[21]	"bedrooms"	"amenities"
[23]	"price"	"minimum_nights"
[25]	"maximum_nights"	"review_scores_rating"
[27]	"review_scores_accuracy"	"review_scores_cleanliness"
[29]	"review_scores_checkin"	"review_scores_communication"
[31]	"review_scores_location"	"review_scores_value"
[33]	"instant_bookable"	

Key Variables

Table 2.1: Summary of Key Variables Used in the Analysis

Variable	Description
price	Nightly listing price in USD
room_type	Type of accommodation offered
property_type	Type of property being rented
bedrooms	Number of bedrooms in the listing
accommodates	Maximum number of guests the listing can accommodate
review_scores_rating	Overall guest rating score
host_is_superhost	Indicates whether the host has Airbnb Superhost status
host_response_rate	Percentage of inquiries responded to by the host
host_acceptance_rate	Percentage of booking requests accepted by the host

Summary Statistics

Mean	Median	SD	Minimum	Maximum
609.0388	150	3442.5	1	625216

Table 2.2: Summary Statistics for Listing Prices

This table summarizes the distribution of listing prices. The large difference between the mean and median reflects the presence of extreme price outliers in the dataset.

3. Exploratory Data Analysis

3.1 Distribution of Airbnb Listing Prices

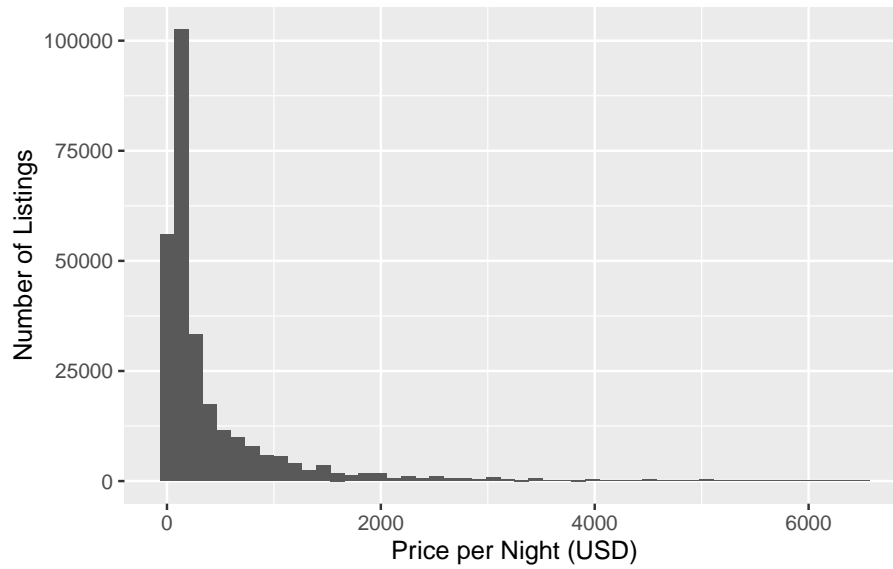


Figure 3.1: Distribution of Airbnb Listing Prices

The distribution of Airbnb listing prices is highly right-skewed. Most listings are concentrated at lower price levels, while a relatively small number of listings command substantially higher nightly rates. Even after excluding the most extreme 1% of listings, a long right tail remains visible.

3.2 Relationship Between Room Type and Listing Price

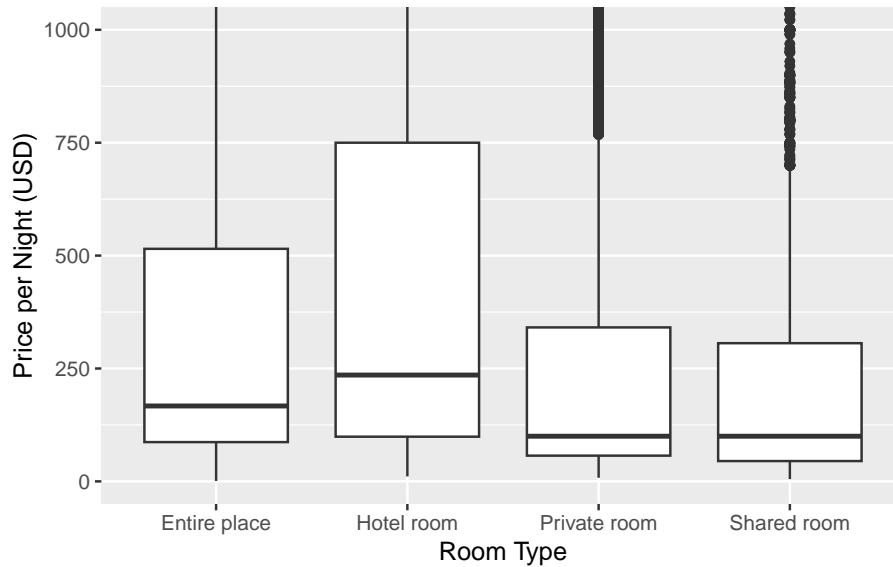


Figure 3.2: Price Distribution by Room Type

This boxplot illustrates the distribution of Airbnb listing prices across different room types. Hotel rooms have the highest median nightly price, followed by entire places. In contrast, private rooms and shared rooms generally have lower median prices, indicating that they are more affordable accommodation options. While prices vary within each room type, the overall pattern suggests that listings offering greater privacy and exclusive access tend to command higher prices. Additionally, all room types show numerous high-priced outliers, highlighting substantial variation in the Airbnb market.

3.3 Relationship Between Bedroom Number and Listing Price

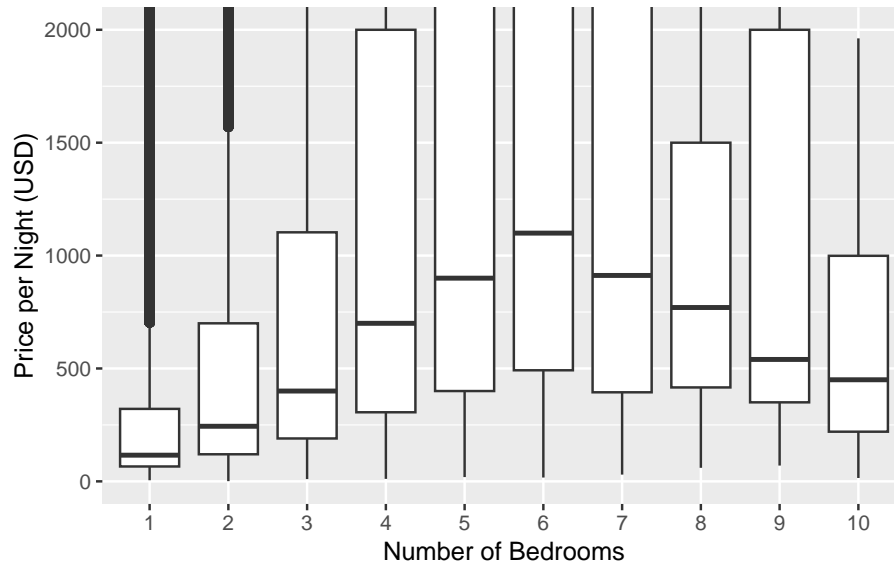


Figure 3.3: Price Distribution by Number of Bedrooms

This boxplot shows the relationship between the number of bedrooms and Airbnb listing prices. Overall, listings with more bedrooms tend to have higher nightly prices. One-bedroom listings have the lowest median price, while larger properties generally exhibit higher median prices and wider price ranges. The increasing spread of the boxplots for larger properties suggests greater variability in pricing among high-capacity listings. Although the relationship is not perfectly linear, the overall trend indicates that property size is positively associated with listing price.

3.4 Relationship between Superhost status and Listing Price

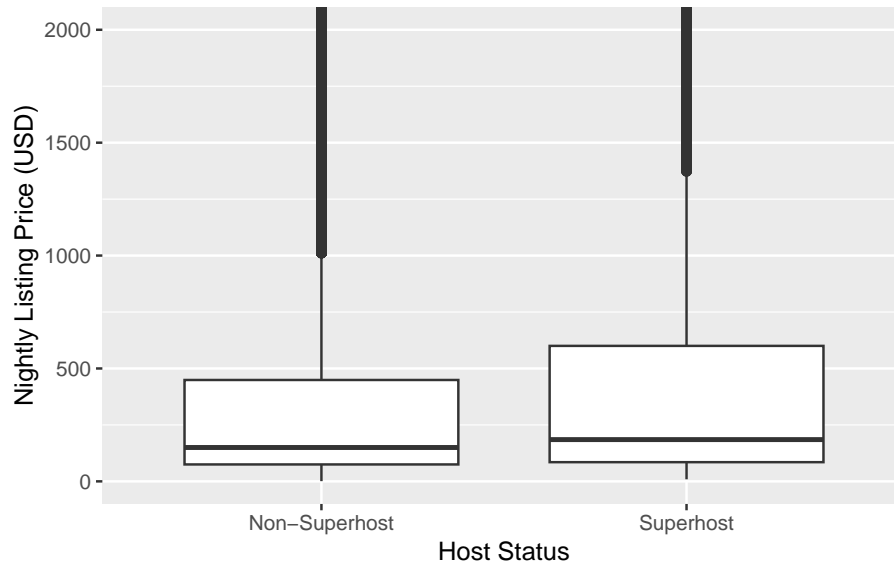


Figure 3.4: Price Distribution by Superhost status

This boxplot shows the relationship between Airbnb Superhost status and listing price. Listings hosted by Superhosts generally have a slightly higher median nightly price than listings hosted by non-Superhosts. The distribution of prices is also wider among Superhosts, indicating greater variability in pricing. While both groups contain high-priced outliers, the overall pattern suggests that Superhost status is associated with higher listing prices. This may reflect increased host experience, stronger reputations, or higher-quality accommodations offered by Superhosts.

4. Conclusion

This analysis examined factors associated with Airbnb listing prices. The results indicate that listing prices are highly right-skewed, with most listings concentrated at lower price levels and a small number of very expensive properties.

Room type, number of bedrooms, and Superhost status all appear to be associated with listing price. Hotel rooms and entire homes generally command higher prices than private or shared rooms. Listings with more bedrooms tend to have higher median prices, suggesting that larger properties can charge higher nightly rates. Superhost listings also exhibit slightly higher median prices than non-Superhost listings.

While these findings identify several interesting patterns, they do not establish causation. Future analyses could incorporate additional variables such as location, review scores, and accommodation capacity to better understand the factors that influence Airbnb pricing.

5. References

AirBnB Listings Dataset. (2023). Kaggle. Created by Ulrik Thyge Pedersen. Retrieved from <https://www.kaggle.com/datasets/ulrikthygepedersen/airbnb-listings>

Gonzalez Castro, M. (2023). An Exploratory Analysis of the Challenges of Latino Students in Two-Year Colleges and Their Perception of Challenges of Transfer to Four-Year Colleges. University of California, Los Angeles. Retrieved from <https://escholarship.org/uc/item/02f1009x>

Vo, K. P. (2025). Point Process Modeling of San Francisco Smash-and-Grabs. University of California, Los Angeles. Retrieved from <https://escholarship.org/uc/item/2xv520qb>

Zhang, Y. (2025). Evaluating Machine Learning Models for Breast Cancer Survival Prediction Using Patient Demographics and Tumor Characteristics. University of California, Los Angeles. Retrieved from <https://escholarship.org/uc/item/3wp122nf>