

2024

AIR BNB ANALYSIS

Presented by :
Sameen Khurram



Contents

Case Study	2
Objective	2
Dataset	2
Data Understanding & Cleaning	2
Price Dataset	2
Room Type Dataset	3
Anomalies	3
Merging of Data	3
Missing Values	4
Data Analysis	5
Data Analysis and Insights	5
Analysis of Room Types and Their Popularity	6
Analysis of Pricing Strategies Across Neighborhoods and Room Types	8
Relationship Between Price, Number of Reviews, and Days Since Last Review	9
Optimal Pricing Strategies	10
Room Types	10
Location Characteristics	11
Recommendations	12
Top 10 Locations	12
Bottom 10 locations	13
Best Practices for listing management	14
Distribution of Room Type	15
Price vs. Popularity	16
Review Recency and Listing Popularity	18
Pricing Trends Across Neighborhoods and Room Types	19

Case Study

Objective

This study aims to utilize GPT 4 and study the market trends so that particularly small Airbnb hosts can utilize these results and optimize our listings.

Dataset

The dataset provided was two Excel files and one TSV. The common column in all three files of “lising_id” which we used to join the data on GPT 4. Below is the information regarding the files:

1. **airbnb_last_review.tsv**

This file had the information about when the hosts received the last review.

2. **airbnb_price**

This file had information about the price of Airbnb, concerning their neighborhood.

3. **airbnb_room_type**

This file had information about the room type and the total number of reviews.

Data Understanding & Cleaning

Last Review Dataset

Columns: listing_id, host_name, last_review.

It contains information about the Airbnb listing ID, the host's name, and the date of the last review.

Price Dataset

Columns: listing_id, price, nbhood_full.

This dataset includes the listing ID, the price of the listing (which seems to be a string with the word 'dollars'), and the neighborhood information.

Room Type Dataset

Columns: Unnamed: 0, listing_id, room_type, number_of_reviews.

It includes an unidentified first column (likely an index), the listing ID, the type of room, and the number of reviews.

Anomalies

I will address inconsistencies or anomalies in the data, particularly in the *'price'* column of the **Price dataset**.

The *'price'* column in the **Price dataset** has been successfully cleaned.

The word 'dollars' was removed, and the values have been converted to numeric (float) format. This will make it easier to perform any numerical analyses or aggregations on the price data.

Merging of Data

Now, let's turn our attention to merging the datasets. To do this effectively, we need a common key across the datasets. The *'listing_id'* column appears in all three, making it a suitable key for merging.

The merged dataset includes the following columns:

listing_id: The unique identifier for each listing.

host_name: The name of the host.

last_review: The date of the last review.

price: The price of the listing (now in numeric format).

nbhood_full: The full neighborhood information.

room_type: The type of room being offered.

number_of_reviews: The number of reviews for the listing.

This merged dataset should provide a more complete view of the Airbnb listings, incorporating aspects like pricing, room type, and review information.

Missing Values

The analysis of missing values reveals the following:

Here is the percentage of missing values in each column of the merged dataset:

- listing_id: 0.00% (no missing values)
- host_name: 13.54%
- last_review: 13.52%
- price: 13.52%
- nbhood_full: 13.52%
- room_type: 39.57%
- number_of_reviews: 39.57%

Notably, the room_type and number_of_reviews columns have a significantly higher percentage of missing values compared to other columns. This may impact analyses that rely on these fields.

Imputation of Missing Values

number_of_reviews Column

Strategy

We used a median-based imputation strategy for the number_of_reviews column.

Rationale

Since number_of_reviews is a numerical column, median imputation was chosen to mitigate the impact of outliers. This approach ensures that the imputed values are representative of the typical (median) number of reviews for Airbnb listings.

Implementation

We utilized the Simple Imputer function from the sklearn.impute module, specifying the median strategy. This function automatically calculated the median number of reviews and replaced all missing values in the number_of_reviews column with this median value.

room_type Column

Strategy

For the categorical room_type column, we employed a mode-based imputation approach, conditioned on the nbhood_full (neighborhood) column.

Rationale

Imputing missing room_type values based on the most common room type in each neighborhood allows us to maintain the underlying distribution and characteristics specific to each area. This method respects the contextual relationship between the neighborhood and the type of rooms typically available there.

Implementation

We grouped the dataset by the nbhood_full column and then applied a lambda function to impute missing room_type values with the mode (most frequent value) of each group. Where a group's mode was not determinable (due to all values being equally frequent), we used the overall mode of the room_type column across the entire dataset as a fallback.

These imputation methods were chosen to preserve the integrity and distribution of your dataset while filling in missing values in a way that aligns with the existing data patterns. This approach facilitates more accurate and meaningful analyses of Airbnb listings.

Data Analysis

Data Analysis and Insights

To proceed with your requested data analysis and insights, we'll focus on the following areas:

Distribution of Room Types and Their Popularity:

We will analyze the distribution of different room types in the dataset. We'll examine the popularity of these room types based on the number of reviews they have received.

Pricing Strategies Across Different Neighborhoods and Room Types:

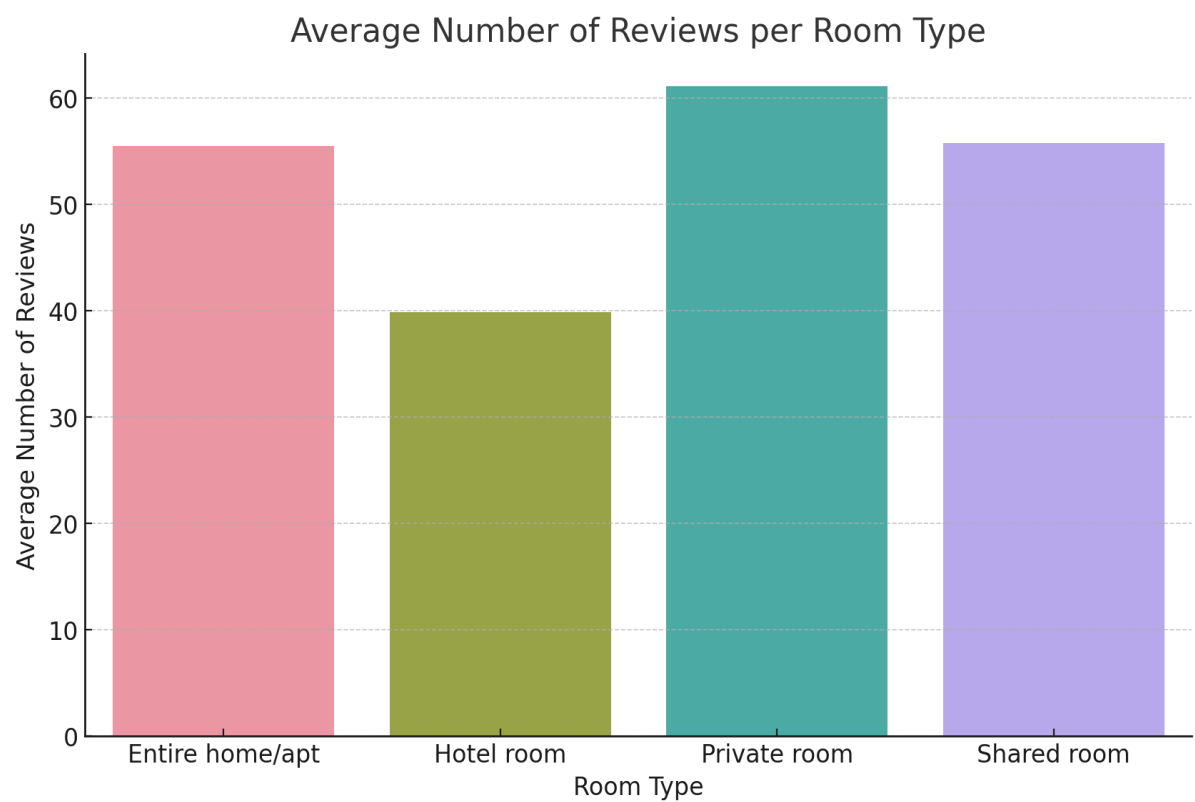
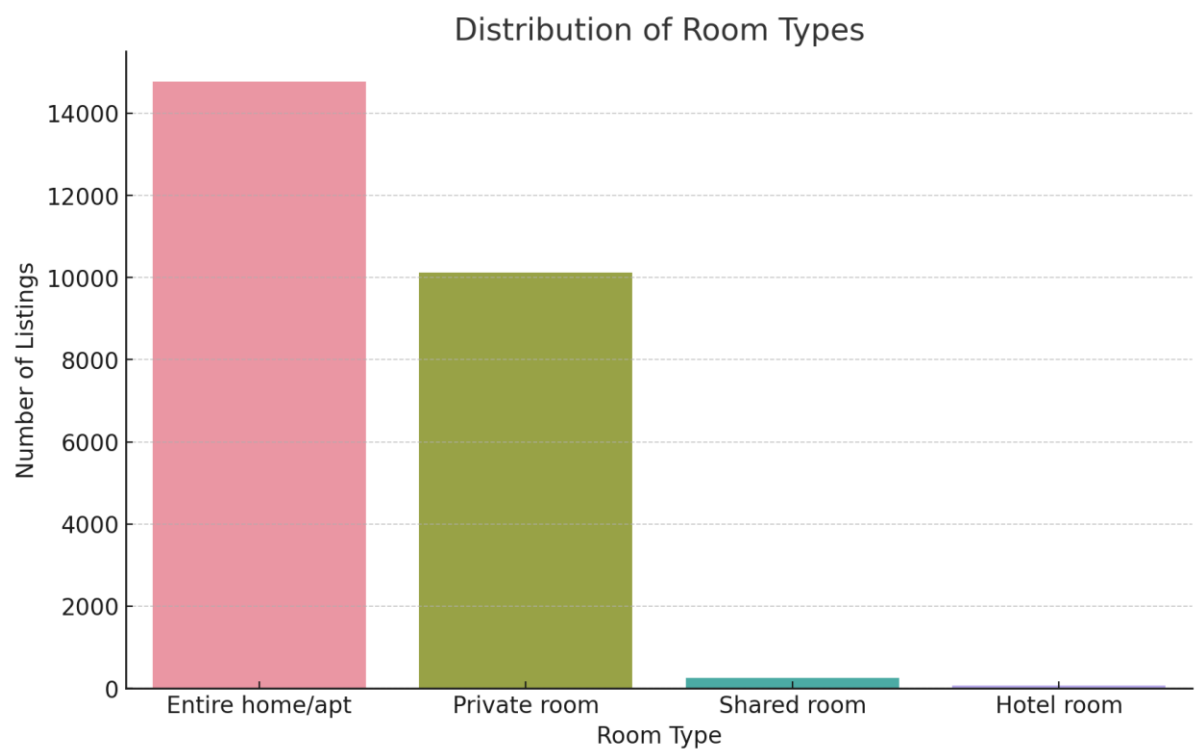
We will investigate how pricing varies across different neighborhoods. We will also examine how room types influence pricing.

Relationship Between Price, Number of Reviews, and Last Review Date:

We will analyze if there's a correlation between price and the number of reviews. We will investigate how the date of the last review relates to price and number of reviews.

Starting with the first analysis: the distribution of room types and their popularity. We'll create a summary of the different room types and calculate the average number of reviews they receive to gauge popularity.

Analysis of Room Types and Their Popularity



Distribution of Room Types:

- The most common room type is 'Entire home/apt' with 14,766 listings.
- 'Private room' is the second most common with 10,118 listings.
- 'Shared room' and 'Hotel room' are significantly less common, with 247 and 78 listings respectively.

Popularity Based on Reviews:

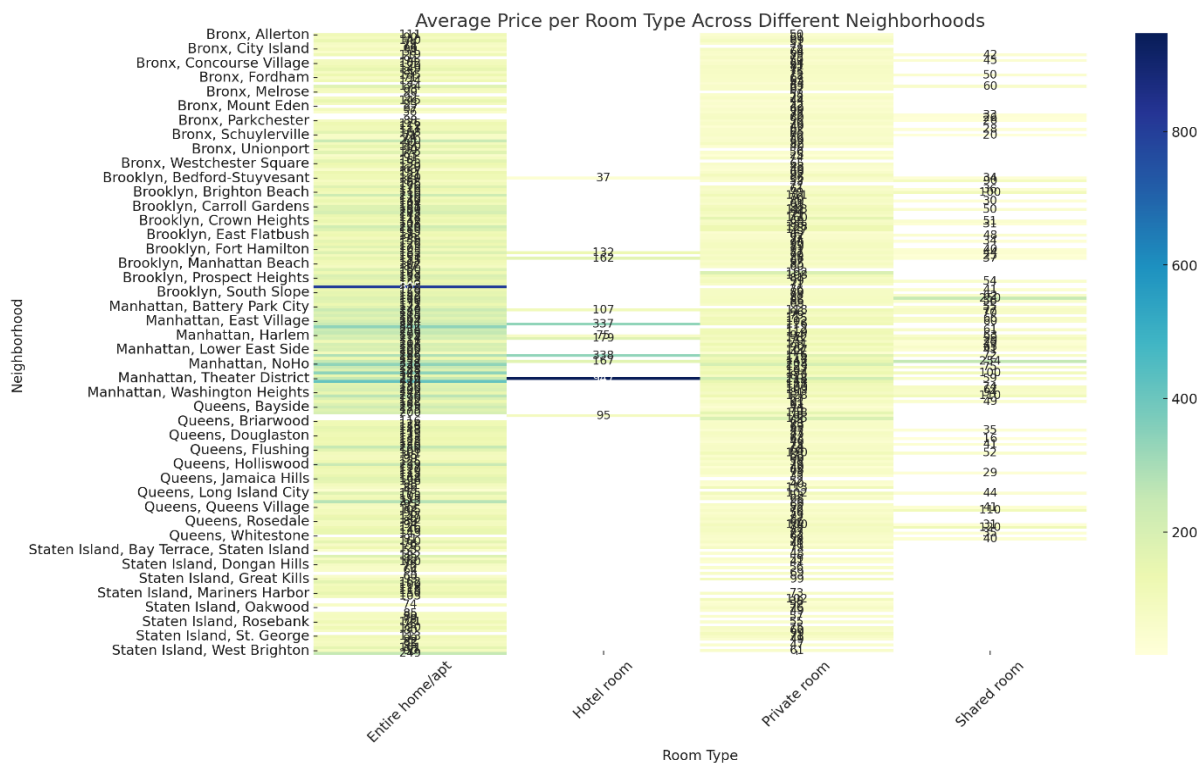
- 'Private rooms' have the highest average number of reviews (about 61.11), suggesting they are quite popular among guests.
- 'Entire home/apt' and 'Shared room' types also have similar average review counts, around 55.48 and 55.79 respectively.
- 'Hotel rooms' have the lowest average number of reviews, approximately 39.86.

Observations and Insights:

- 'Entire home/apt' and 'Private room' are the most prevalent types of listings, indicating a higher demand or supply in these categories.
- Despite being less common, 'Private rooms' seem to be more frequently reviewed than 'Entire home/apt', which might indicate higher guest turnover or satisfaction.

Next, let's investigate pricing strategies across different neighborhoods and room types. This will involve analyzing the average price for each room type across various neighborhoods. We'll create a visualization to better understand these pricing dynamics.

Analysis of Pricing Strategies Across Neighborhoods and Room Types



The heatmap displays the average price of different room types across various neighborhoods. Here are some observations:

Variation in Prices:

Prices vary significantly across neighborhoods and room types. In some neighborhoods, certain room types (like 'Hotel room') may not be available, as indicated by the blank spaces in the heatmap.

Neighborhood Influence:

Neighborhoods play a crucial role in determining the pricing of listings. Some neighborhoods have higher average prices for 'Entire home/apt', which might be due to factors like location, demand, and amenities.

Room Type Influence:

'Entire home/apt' types tend to have higher average prices compared to other room types in the same neighborhood.

'Private room' and 'Shared room' types are typically more affordable.

Observations and Insights

The pricing strategy for Airbnb listings is highly influenced by both the room type and the neighborhood.

This analysis can help hosts in setting competitive prices and guests in finding accommodations that fit their budget and preferences.

Finally, let's understand the relationship between price, the number of reviews, and the last review date. We'll look for correlations and patterns that might exist between these variables. This will involve statistical analysis and possibly some visualizations for clearer insights.

Relationship Between Price, Number of Reviews, and Days Since Last Review

The correlation matrix provides insights into the relationships between the following variables:

Price and Number of Reviews:

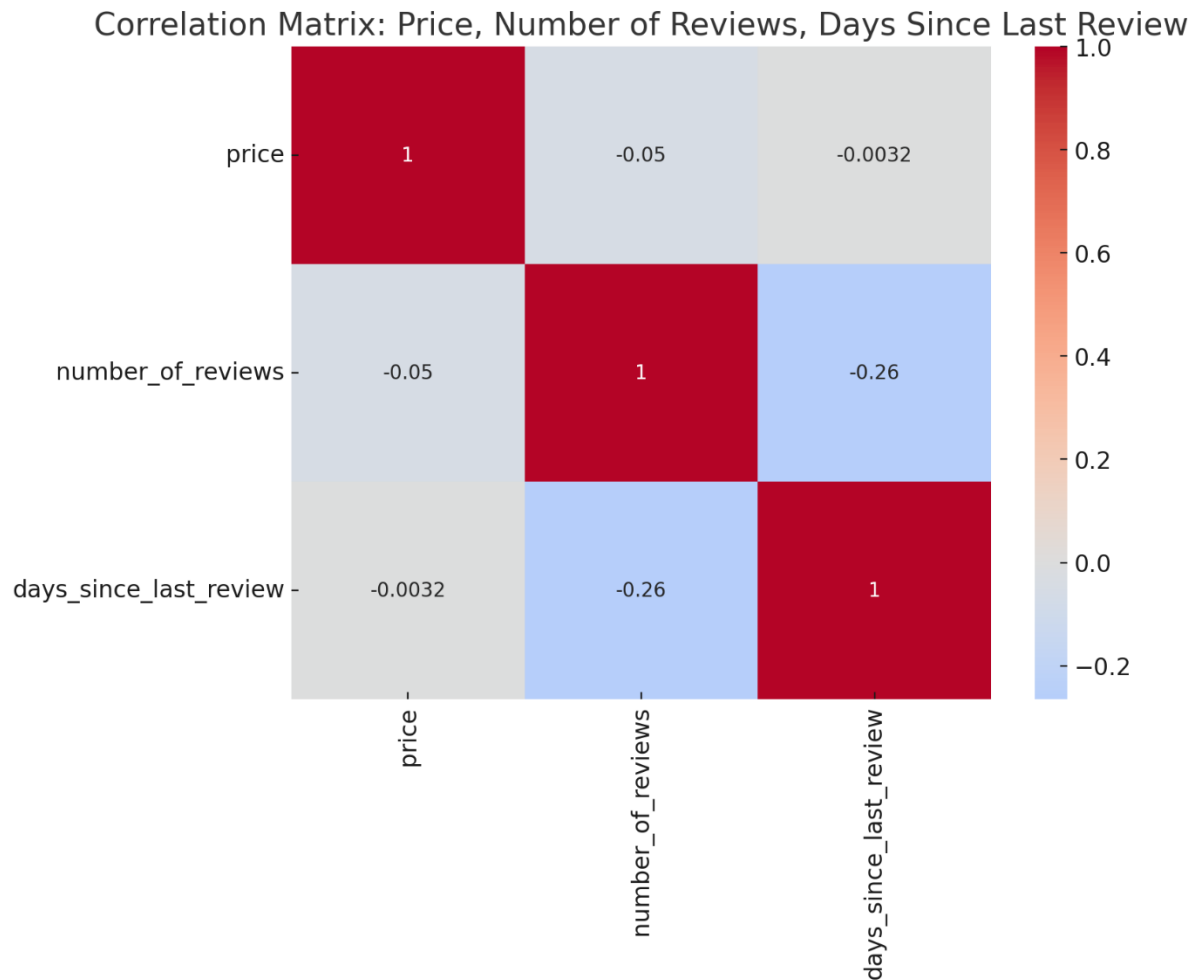
There is a weak negative correlation (-0.0499) between price and the number of reviews. This suggests that higher-priced listings might have slightly fewer reviews, but the correlation is too weak to draw a definitive conclusion.

Price and Days Since Last Review:

The correlation between price and days since the last review is negligible (-0.0032). This indicates that there's no significant relationship between how recently a property was reviewed and its price.

Number of Reviews and Days Since Last Review:

There is a weak negative correlation (-0.2648) between the number of reviews and days since the last review. This implies that listings reviewed more recently tend to have more reviews, which could be due to newer listings gaining popularity or actively managed listings receiving more frequent feedback.



Observations and Insights

Price does not strongly influence the frequency or recency of reviews.

Recent reviews might be an indicator of popularity or active management of a listing, but they do not significantly correlate with higher or lower prices.

This analysis can assist in understanding the dynamics of guest interactions and pricing strategies but suggests that factors other than price and recency of reviews might play a more significant role in the popularity of a listing.

Optimal Pricing Strategies

To develop optimal pricing strategies for different room types and locations in the context of Airbnb listings, it's important to consider various factors that influence demand and guest preferences. Here's a detailed guide based on room types and location characteristics:

Room Types

Entire Home/Apartment

- **High Demand Areas (e.g., city centers, tourist hotspots):** Price competitively but higher than other room types, considering the privacy and space offered.
- **Residential or Outskirts:** Offer lower rates than central areas but maintain a premium over private and shared rooms. Emphasize features like space and privacy.

Private Room

- **Urban and Well-connected Areas:** Price competitively, slightly lower than entire homes. Focus on cleanliness, comfort, and added amenities like Wi-Fi or breakfast.
- **Suburban or Less Central Areas:** Offer budget-friendly rates. Highlight local experiences and accessibility to public transport.

Shared Room

- **Popular Among Budget Travelers:** Generally, the lowest-priced option. Best suited for hostels or budget accommodations, especially in areas popular among backpackers or young travelers.

Location Characteristics

- **Tourist Hotspots**
- **High Demand:** Can command higher prices, especially during peak tourist seasons.
- **Strategy:** Use dynamic pricing to adjust rates based on demand fluctuations and local events.

Business Districts

- **Target Audience:** Business travelers.
- **Strategy:** Offer competitive rates on weekdays with possible premium pricing. Focus on amenities like fast Wi-Fi and a workspace.

Residential or Suburban Areas

- **Lower Demand:** Prices should be more budget-friendly.
- **Strategy:** Emphasize tranquility, space, and local lifestyle. Offer discounts for longer stays.

Near Special Attractions (e.g., beaches, mountains)

- **Unique Selling Point:** Natural attractions or unique local experiences.
- **Strategy:** Price can be higher if the location offers a unique experience. Highlight proximity to attractions and views.

Conclusion

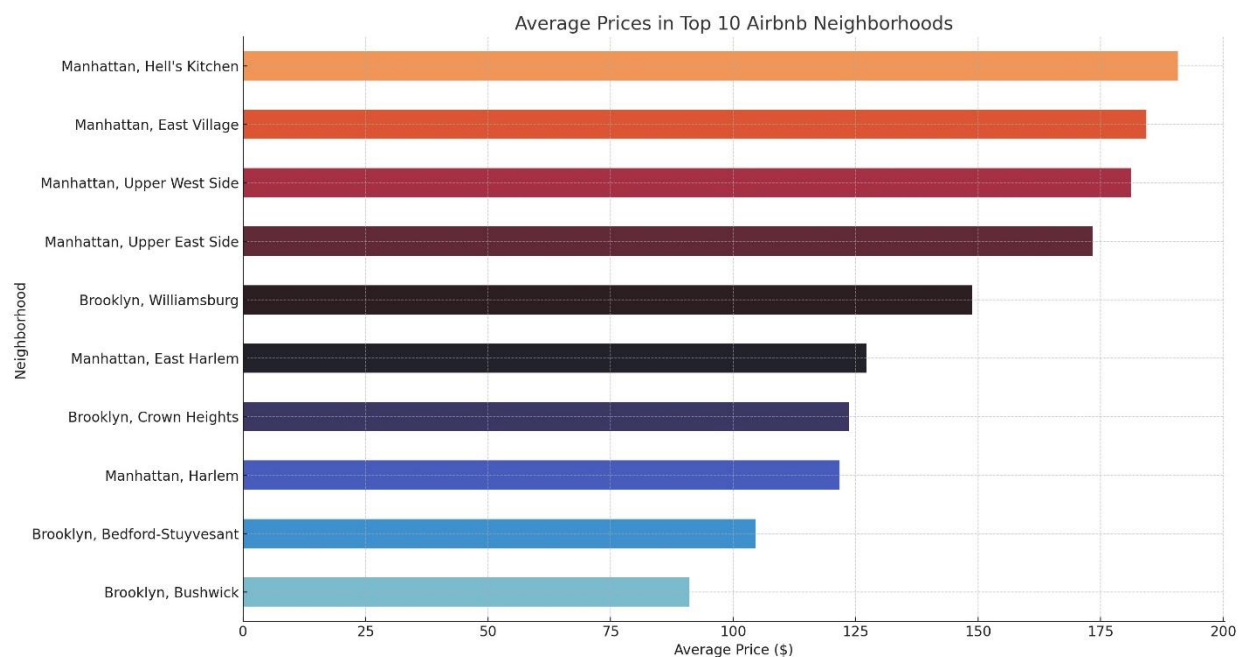
Optimal pricing strategies should be dynamic and tailored to the type of room, location, and changing market conditions. Understanding guest profiles, local demand cycles, and competitive landscape is key to setting effective pricing. Regular adjustments and a focus on guest experience can help maximize both occupancy rates and revenue.

Recommendations

- **Start with Competitive Pricing:** Initially, hosts can set rates slightly below average for similar listings in your area to attract guests and garner reviews.
- **Prioritize Guest Reviews:** Hosts should focus on obtaining positive early reviews and respond professionally to all feedback.
- **High-Quality Listing Details:** They should ensure their listing features clear, professional photos and a detailed, accurate description.
- **Responsive Communication:** They need to be quick to respond to inquiries and bookings and should provide comprehensive guest information.
- **Update Regularly for Search Optimization:** Regular updates and responsiveness can improve their Airbnb search ranking.
- **Tailor to Room Type and Location:** They should customize their space and services to suit the specific room type and location, emphasizing local experiences or conveniences.
- **Essential Amenities and Extras:** They can offer in-demand amenities and thoughtful extras, such as Wi-Fi, workspaces, or local treats.
- **Adhere to Safety and Cleanliness Standards:** Maintain high cleanliness standards and ensure all safety measures are in place, in line with Airbnb's guidelines.

By focusing on these key areas, hosts can enhance the appeal and success of your Airbnb listing in a competitive market.

Top 10 Locations



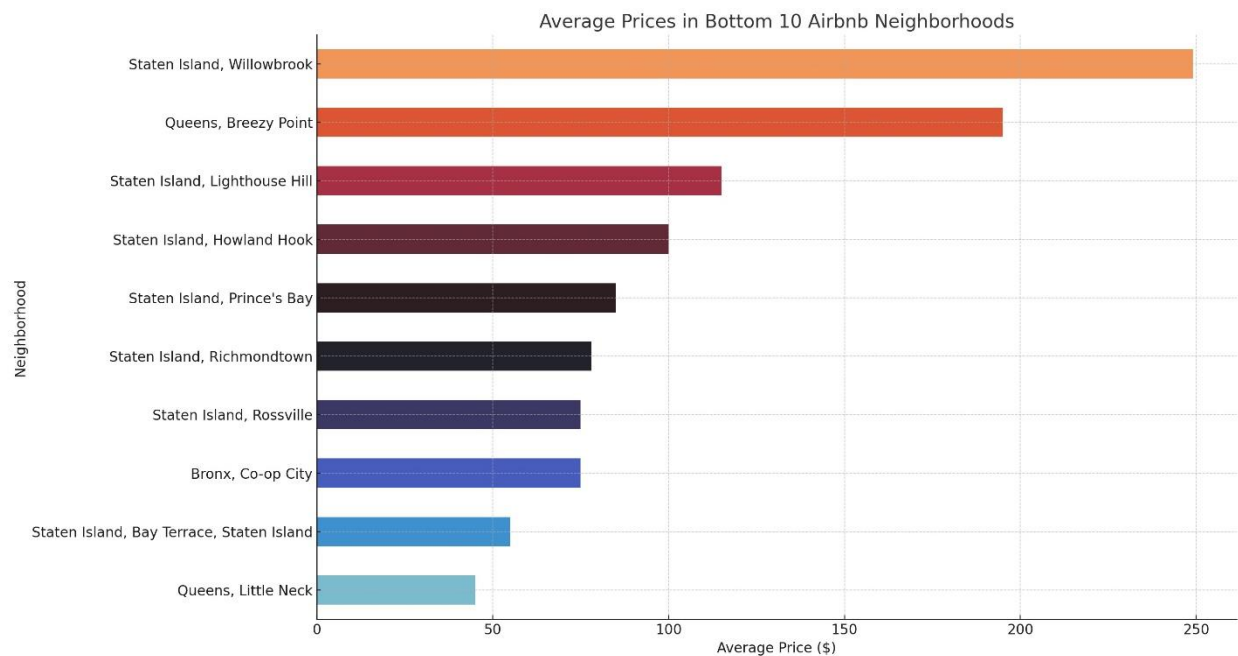
In these neighborhoods, hosts should focus on maximizing the value of their listings by emphasizing unique features that justify the higher price points. For neighborhoods like Tribeca,

SoHo, and West Village, which combine popularity with higher prices, hosts should highlight luxury amenities, proximity to key attractions, or unique aspects of their property.

In areas like Sea Gate and Neponsit, where the review count is lower, hosts might focus on creating an exceptional guest experience to boost their review count. This can include offering personalized services, exceptional hospitality, or unique local experiences.

For neighborhoods like Midtown and Murray Hill, where there's a balance between price and popularity, hosts should maintain competitive pricing but also focus on differentiating their listings through quality, service, and responsiveness.

Bottom 10 locations



In the dataset, the bottom 10 neighborhoods, characterized by fewer listings, present unique opportunities for hosts. These neighborhoods include Co-op City in the Bronx, Breezy Point and Little Neck in Queens, and several areas in Staten Island like Bay Terrace, Howland Hook, Lighthouse Hill, Prince's Bay, Richmondtown, Rossville, and Willowbrook.

The pricing in these areas varies significantly reflecting the diversity in these less-saturated markets. The number of reviews per listing also varies, indicating different levels of guest traffic and popularity. For instance, Richmondtown has an average of 95 reviews per listing, suggesting a high guest turnover, whereas Breezy Point shows an average of only 5 reviews, indicating lower booking frequency.

For hosts in these areas, focusing on niche marketing strategies that highlight the unique local attributes of these less-traveled neighborhoods could be beneficial.

Best Practices for listing management

The following report outlines best practices for Airbnb hosts to enhance the attractiveness of their listings. A well-managed listing can significantly increase booking rates and overall guest satisfaction.

Frequent Updates to the Listing:

- **Regular Refresh:** Update your listing regularly (at least once a month) to reflect any new amenities, improvements, or changes in house rules.
- **Seasonal Adjustments:** Modify your description and photos to align with different seasons or local events.
- **High-Quality, Accurate Photos:** Professional Photography: Use high-resolution, well-lit photos. Consider hiring a professional photographer.
- **Variety and Detail:** Include multiple photos covering all areas of your space.

Detailed and Clear Descriptions:

- **Highlight Unique Features:** Mention any unique aspects of your space or location.
- **Clarity and Accuracy:** Ensure your description accurately reflects your space to set the right expectations.

Responsive and Engaging Host:

- **Prompt Responses:** Respond quickly to inquiries and booking requests.
- **Proactive Communication:** Reach out to guests with all necessary information before their stay.

Management of Reviews:

- **Encourage Reviews:** Politely ask guests to leave a review after their stay.
- **Respond to Reviews:** Address both positive and negative reviews professionally. Show appreciation for feedback and address any concerns raised.
- **Competitive and Dynamic Pricing:** Regularly review and adjust your pricing based on local demand, seasonality, and competition.

Amenities and Comfort:

- **Essential Amenities:** Provide basic amenities like Wi-Fi, clean linens, and toiletries.
- **Additional Comforts:** Consider extras like a coffee machine, a workspace, or local guides.

Cleanliness and Safety:

- **Strict Cleaning Protocols:** Maintain high cleanliness standards. This is particularly crucial in the context of COVID-19.
- **Safety Features:** Equip your space with safety features like smoke detectors, first aid kits, and fire extinguishers.

Local Experience and Personal Touch:

- **Local Recommendations:** Provide guests with recommendations for local dining, shopping, and attractions.
- **Personal Welcome:** Small gestures like a welcome note or a local snack can greatly enhance guest experience.

Leverage Technology:

Utilize online tools for property management, automated messaging, and market analysis.

Regulatory Compliance:

Ensure adherence to local laws and Airbnb policies regarding short-term rentals.

Conclusion

Implementing these best practices can significantly enhance the attractiveness of Airbnb listings. A well-managed, clearly described, and competitively priced listing, combined with a responsive and thoughtful host, can lead to higher booking rates and positive reviews, ultimately contributing to the success of the Airbnb host.

Distribution of Room Type



Visualizations

The pie charts above illustrate the distribution of room types in each neighborhood.

Distribution Analysis

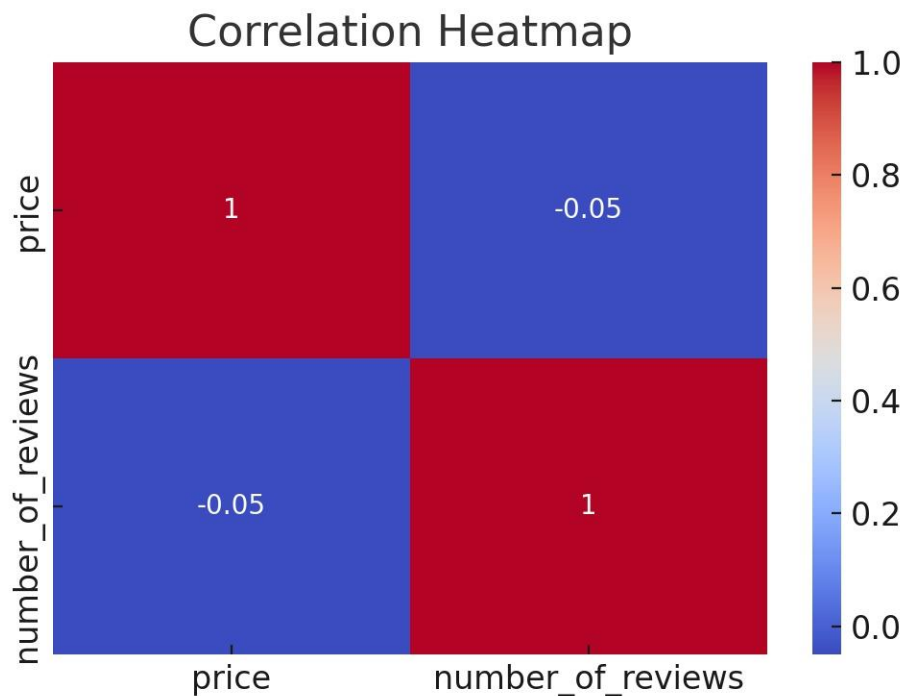
The availability of room types varies significantly across different neighborhoods.

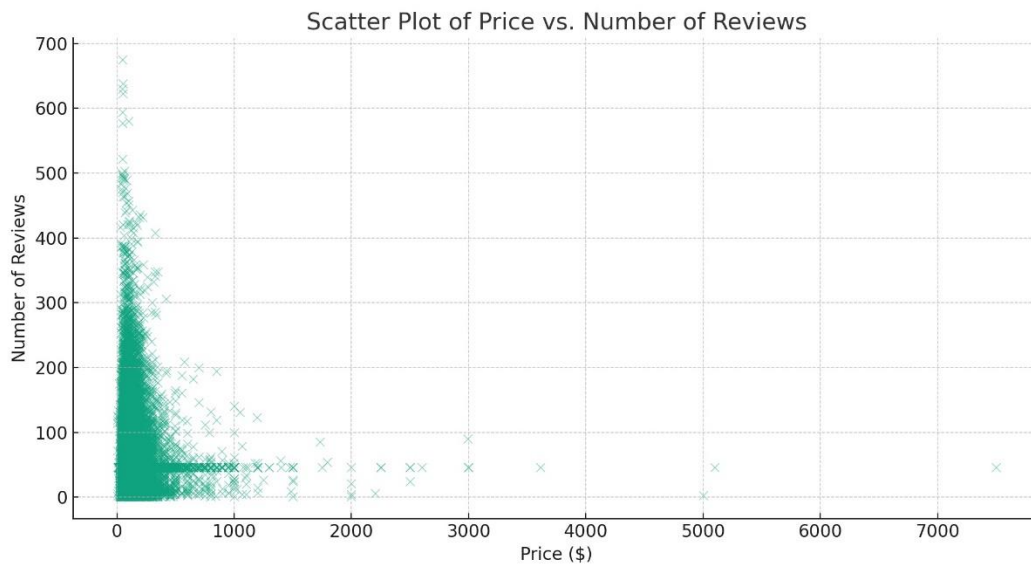
Key observations

- **Entire Home/Apt** is predominant in most neighborhoods, highlighting its popularity or availability.
- **Private Rooms** also represent a significant portion of listings, particularly in certain neighborhoods.
- **Shared Rooms and Hotel Rooms** are less common overall.
- Neighborhoods like Manhattan and Brooklyn have a higher proportion of "Entire home/apt" and "Private room" types, which might be due to the higher tourist footfall and residential nature.
- Neighborhoods with a high concentration of "Private rooms" might indicate areas with more shared housing or budget-friendly options.

The diversity in room types across neighborhoods suggests varying traveler preferences and residential patterns.

Price vs. Popularity





Visualization

The scatter plot shows the relationship between price and the number of reviews. It appears that while there are listings with a high number of reviews across various price points, very high-priced listings tend to have fewer reviews.

The correlation heatmap confirms the weak negative correlation between price and the number of reviews.

Correlation Analysis

The Pearson correlation coefficient between price and the number of reviews is approximately -0.0499, suggesting a very weak negative correlation.

Regression Analysis

Given the weak correlation, a linear regression model might not provide meaningful insights. The relationship does not appear to be strongly linear.

Insights and Observations

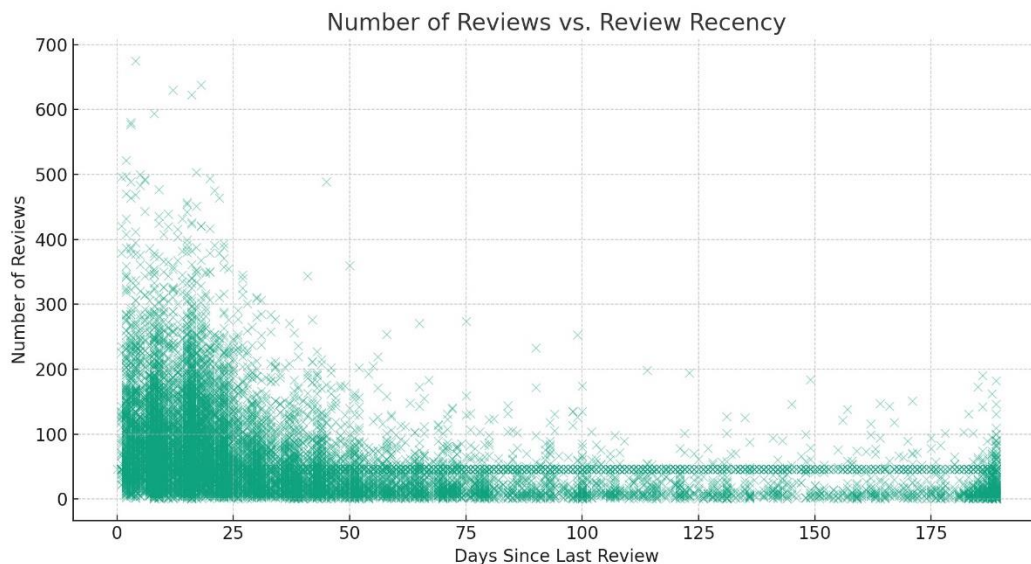
The data suggests that the price of a listing does not strongly influence its popularity, as indicated by the number of reviews. It is observed that extremely high-priced listings have fewer reviews, which might be due to less frequent rentals or a smaller customer base. Listings in the moderate price range tend to have a higher number of reviews, possibly indicating a balance between affordability and attractiveness to a broader range of guests.

Summary Report

While there is a weak negative correlation between price and number of reviews, it is not strong enough to draw definitive conclusions about the price's impact on a listing's popularity. Other factors likely play a more significant role in influencing how often guests choose and review an Airbnb listing.

Review Recency and Listing Popularity

The number of reviews a listing has received were analyzed.



Visualization

A scatter plot has been created to visualize the relationship between the number of reviews and the recency of reviews.

Insights and Observations

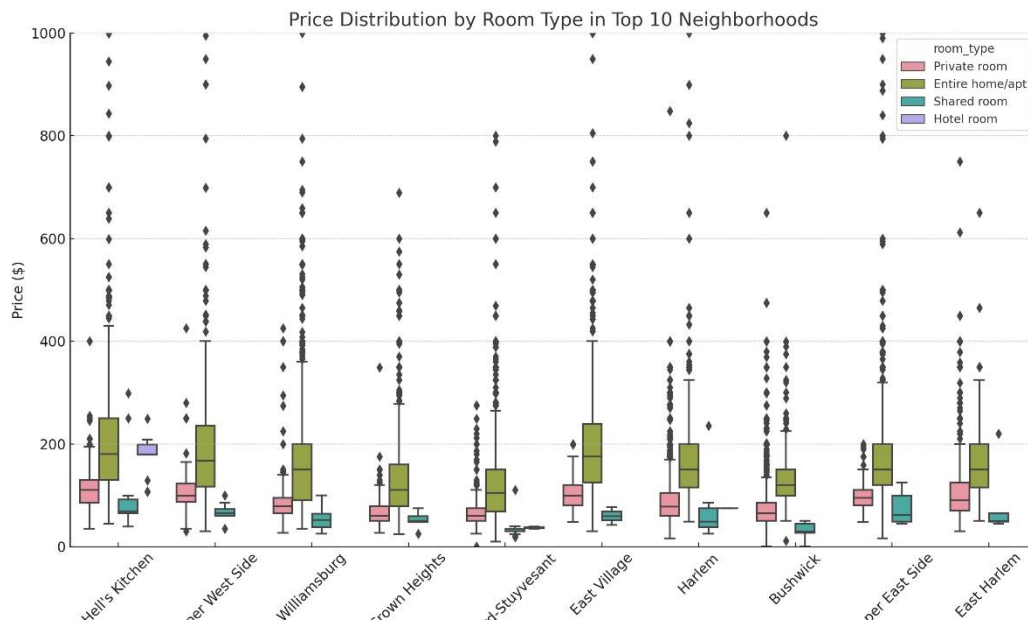
The scatter plot suggests a varied distribution. While there are listings with a high number of reviews across different levels of recency, a concentration of listings with fewer reviews seems to have a wider range of recency. It appears that listings with a very **high number of reviews** tend to have **more recent reviews**, which could indicate ongoing popularity.

However, the data does not distinctly show that recent reviews always correlate with higher numbers of reviews.

Summary Report

While there is some indication that listings with a higher number of reviews tend to have more recent reviews, the relationship is not straightforward. Recency may be one of several factors influencing a listing's popularity.

Pricing Trends Across Neighborhoods and Room Types



Visualization

A box plot has been created to visualize the price distribution for different room types across the top 10 neighborhoods (based on the number of listings).

Insights and Observations

- The box plot reveals significant variations in price distribution across different room types and neighborhoods.
- Generally, 'Entire home/apt' tends to be priced higher than 'Private room' and 'Shared room' across neighborhoods.
- Some neighborhoods exhibit a wider range of prices, indicating a diverse range of accommodation options.
- The presence of outliers, especially in high-demand neighborhoods, suggests that some listings are priced much higher than the average.

Summary Report

The analysis demonstrates clear pricing trends across different neighborhoods and room types. The choice of neighborhood and the type of room significantly influence the pricing of Airbnb listings. This information can be valuable for both hosts in setting competitive prices and guests in finding accommodations that fit their budget and preferences.