Case Study (Air Bnb)

Data Analysis Using Chat GPT4

Submitted by: Hamza Riaz

Group 11

Contents

Objective:

The purpose of this task is to analyze the Airbnb market in a specific city. The datasets provided include information about Airbnb listings, such as room types, pricing, reviews, and locations. The goal is to understand the current market, identify trends, and provide recommendations to small Airbnb hosts looking to optimize their listings.

Dataset:

It includes three files related to Airbnb data, each in a different format:

1. airbnb_last_review.tsv - This file is in TSV (Tab-Separated Values) format, which is similar to CSV but uses tabs as separators between the data fields.
2. airbnb_price.csv - This is a CSV (Comma-Separated Values) file. It's a common format for storing tabular data and is easily readable and writable by most spreadsheet programs.
3. airbnb_room_type.xlsx - This is an Excel file. It might contain one or more sheets with data in a table format.

Data Analysis:

Basic Information about Data:

The files contain complementary information about Airbnb listings, including details about the hosts, prices, neighborhoods, room types, and reviews. There's a common listing column across all files, which suggests that they can be merged or joined for a more comprehensive analysis.

1. Last Review Data (airbnb_last_review.tsv)

- **Number of Rows:** 25,209
- **Number of Columns:** 3
- **Columns:**
    - listing_id: Presumably the unique identifier for each Airbnb listing.
    - host_name: The name of the host.
    - last_review: The date of the last review received.
- **Sample Data:**
    - Listing ID 2595, hosted by Jennifer, last reviewed on May 21, 2019.
    - Listing ID 3831, hosted by LisaRoxanne, last reviewed on July 05, 2019.

2. Price Data (airbnb_price.csv)

- **Number of Rows:** 25,209
- **Number of Columns:** 3
- **Columns:**
  - listing_id: The unique identifier for each Airbnb listing.
  - price: The price of the listing (in dollars).
  - nbhood_full: The neighborhood and larger area where the listing is located.
- **Sample Data:**
  - Listing ID 2595, priced at 225 dollars, located in Manhattan, Midtown.
  - Listing ID 3831, priced at 89 dollars, located in Brooklyn, Clinton Hill.

3. Room Type Data (airbnb_room_type.xlsx)

- **Number of Rows:** 17,614
- **Number of Columns:** 4
- **Columns:**
  - Unnamed: 0: Appears to be an index or serial number.
  - listing_id: The unique identifier for each Airbnb listing.
  - room_type: The type of room offered (e.g., Entire home/apt, Private room).
  - number_of_reviews: The total number of reviews received for the listing.
- **Sample Data:**
  - Listing ID 2595, Entire home/apt, with 48 reviews.
  - Listing ID 3831, Entire home/apt, with 295 reviews.

Missing Values, Outliers, Anomalies:

Analyzed the data for any Missing Values, Outliers, or any other Anomalies.

1. Missing Values

- **Columns with Significant Missing Values:**
  - host_name, last_review, price, nbhood_full have similar counts of missing values (~3,940).
  - room_type, number_of_reviews have a higher count of missing values (~11,535).
- **Suggestions:**
  - If a column like host_name is not crucial for analysis, consider dropping it.
  - For price and room_type, investigate if missing values correspond to inactive or unlisted properties.
  - For last_review and number_of_reviews, missing values might indicate listings without reviews. Consider replacing them with appropriate indicators (e.g., No Reviews or 0).

2. Data Types

- Data types seem appropriate for each column.

3. Duplicate Records

- No duplicate records based on listing_id.

4. Outliers

- **Numerical Columns:**
  - price ranges from 0 to 7,500 dollars, with a mean of 141.78 dollars. The maximum value seems unusually high.
  - number_of_reviews range from 1 to 675 reviews. The maximum value is high but could be plausible for a very popular listing.
- **Suggestions:**
  - Investigate listings with extremely high prices or review counts to confirm their validity.
  - Consider using statistical methods (e.g., IQR) to identify and handle outliers.

5. Consistency in Categorical Data

- **Room Types:** The data includes 'Entire home/apt', 'Private room', 'Shared room', 'Hotel room', and some missing values.
- **Suggestions:**
  - Ensure that the categorization is consistent and correctly labeled.
  - Decide how to handle missing values in room_type based on the analysis requirements (e.g., drop, categorize as 'Other').

Overall, the dataset required some cleaning and preprocessing to address these issues.

Handling Irregularities

Missing Values:

- "host_name", "nbhood_full", and "room_type" missing values are replaced with "unknown".
- "last_review" missing values are replaced with "no reviews".
- "number_of_reviews" missing values are replaced with 0.

Apart from these, there are missing values in the "price" column.

The proportion of missing values in the "price" column is approximately 13.52% of the total dataset. This is a significant portion, but not overwhelmingly large. Given this proportion, the following method is used to address this issue.

Imputation with Mean/Median:

- If the data distribution of the "price" column is not heavily skewed, you can impute the missing values with the mean or median price. The median is typically more robust to outliers.

Before proceeding with this, it's helpful to understand the distribution of the 'price' data. Therefore, a basic statistical summary and visualization of the distribution of the 'price' column to inform our choice of method is made.

Distribution of Prices
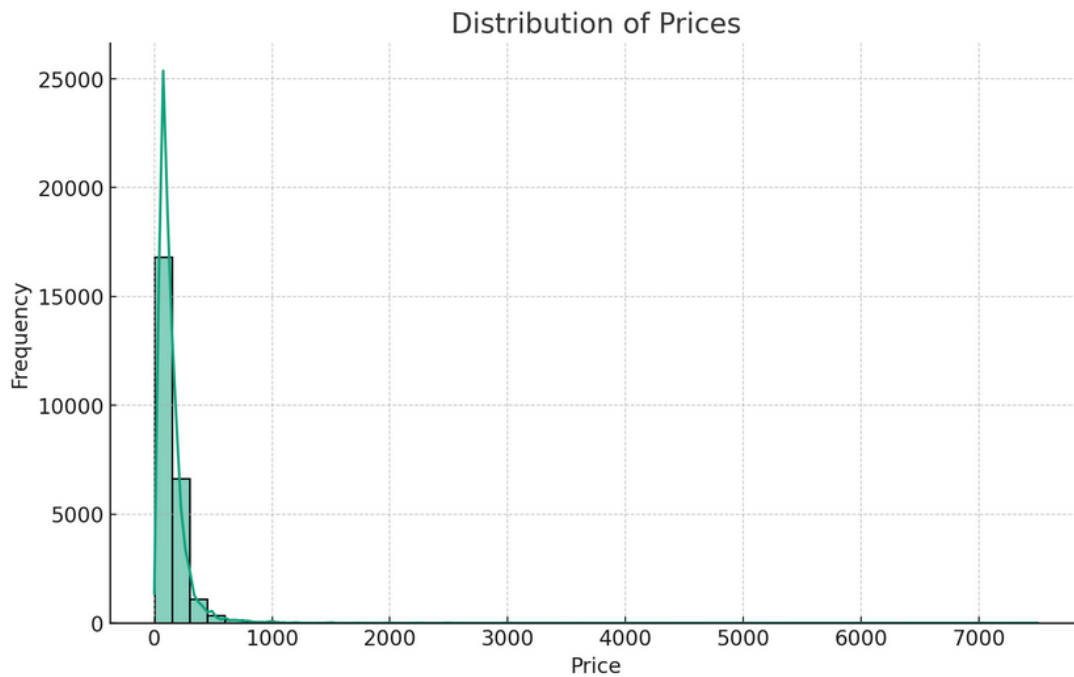
The statistical summary and the distribution plot provide insights into the 'price' data:

- **Mean Price**: $141.78
- **Median Price**: $105.00
- **Standard Deviation**: $147.35
- **Min/Max Price**: $0.00 - $7500.00

Observations:

1. **Skewed Distribution**: The distribution of prices is right skewed. There are a few listings with very high prices compared to the majority, which pulls the mean higher than the median.

2. **Presence of Extreme Values**: The maximum price is $7500, which is much higher than the mean or median, indicating the presence of outliers.

Given these observations, The following method for handling missing values in the price column.

- **Median Imputation**: Given the skewed nature of the data, imputing missing values with the median ($105.00) is recommended over the mean.

EDA (Exploratory Data Analysis)

Room Types

The analysis of room types in the dataset and their popularity, based on the number of reviews, reveals the following insights:

1. **Room Types**:
   - **Entire home/apt**: The most common type of listing, with 9,405 listings.
   - **Private room**: The second most common, with 7,752 listings.
   - **Shared room**: Significantly fewer listings, with only 357.
   - **Hotel room**: Least common among the specified types, with 100 listings.
   - **Unknown**: There are 11,535 listings with an unspecified or unknown room type.
2. **Total Reviews**:
   - **Entire home/apt** listings have accumulated the most reviews (507,530).
   - **Private rooms** are close behind with 462,604 reviews.
   - **Shared rooms** and **Hotel rooms** have far fewer reviews, with 15,276 and 3,246 respectively.
   - Listings categorized as **unknown** have no reviews.
3. **Average Reviews per Listing**:
   - **Private rooms** have the highest average number of reviews per listing (approximately 59.68), indicating high popularity or longer availability on the platform.
   - **Entire home/apt** listings have an average of about 53.96 reviews per listing.
   - **Shared rooms** have an average of approximately 42.79 reviews per listing.
   - **Hotel rooms** have the lowest average at about 32.46 reviews per listing.
   - The **unknown** category has an average of 0 reviews per listing, which might indicate data issues or new listings.

These insights suggest that while entire homes/apartments are the most common type of listing, private rooms tend to have a slightly higher average number of reviews, indicating potentially higher turnover or longer presence on the platform. Shared rooms and hotel rooms are less common and have lower averages in terms of both total and per listing reviews. The unknown category is significant in number but lacks review data, which might require further investigation to understand its nature.
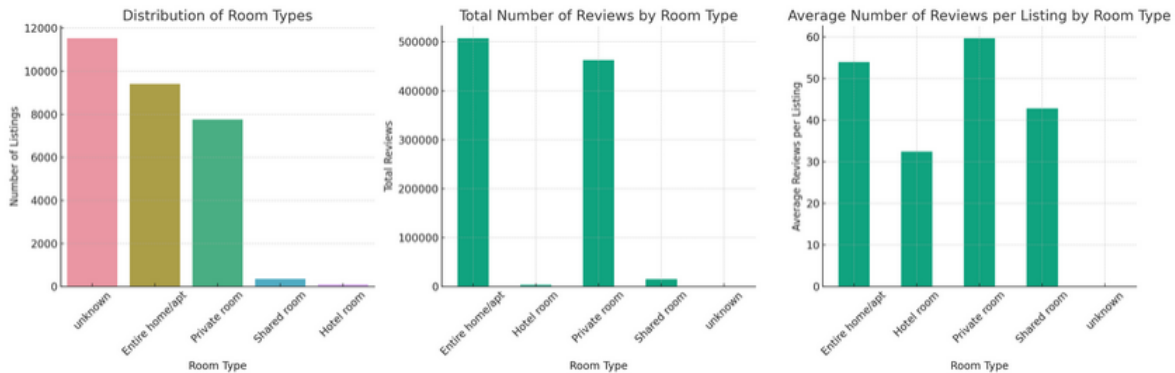
The visualizations created, supports, and explains the insights drawn from the analysis of room types and their popularity:

1. **Distribution of Room Types**:
   o The first chart shows the number of listings for each room type. Entire homes/apartments are the most common, followed by private rooms. Shared and hotel rooms are much less common, and there is a significant number of listings with an unknown room type.
2. **Total Number of Reviews by Room Type**:
   o The second chart illustrates the total number of reviews for each room type. It highlights that entire homes/apartments and private rooms have amassed a significantly higher number of reviews compared to shared and hotel rooms. The unknown category has no reviews.
3. **Average Number of Reviews per Listing by Room Type**:
   o The third chart depicts the average number of reviews per listing for each room type. Private rooms lead in terms of average reviews per listing, followed closely by entire homes/apartments. Shared rooms and hotel rooms have lower averages, and the unknown category shows zero reviews per listing.

These visualizations provide a clear picture of the popularity and distribution of different room types in the dataset, highlighting the dominance of entire homes/apartments and private rooms both in terms of their presence on the platform and the number of reviews they have received.
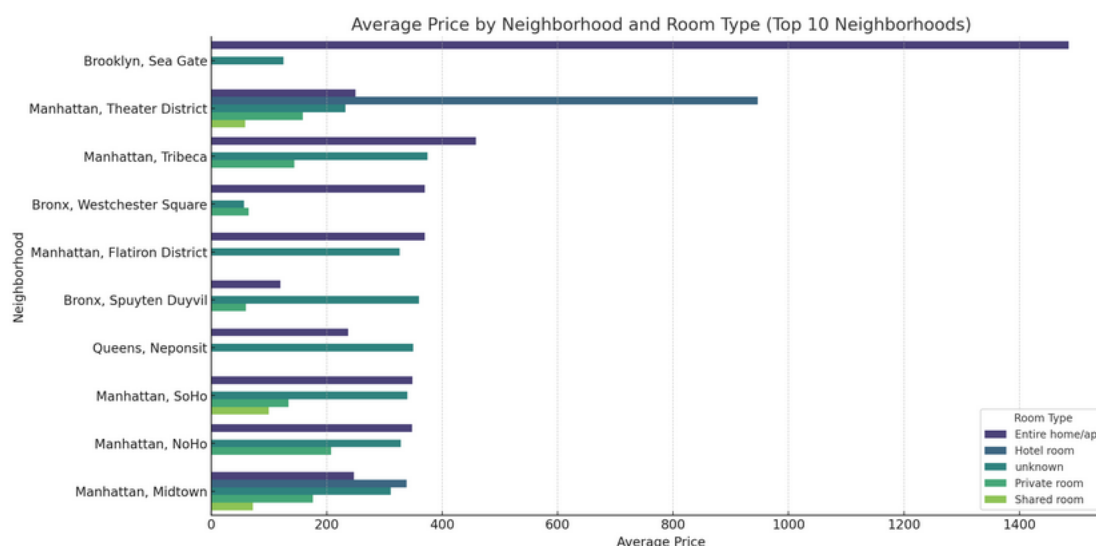
Pricing

The analysis of pricing across different neighborhoods and room types reveals some interesting patterns:

1. **High-End Neighborhoods and Luxury Listings**: Some neighborhoods, particularly in Manhattan (like Tribeca, Flatiron District, SoHo, NoHo), have very high average prices for entire home/apartment listings. This suggests a luxury market in these areas.
2. **Unique or Exceptional Listings**: The highest average price is observed in Brooklyn's Sea Gate neighborhood for entire home/apartments, indicating the presence of premium, possibly unique or exceptional listings.
3. **Hotel Rooms**: In some neighborhoods, like the Theater District in Manhattan, hotel rooms have a high average price, which is typical for hotel accommodations in prime locations.
4. **Unknown Room Types**: There are significant average prices for listings with unknown room types in some areas, indicating potential data categorization issues or unique property types.

Based on these insights, here are some pricing strategies to consider:

1. **Dynamic Pricing in High-End Neighborhoods**: For neighborhoods with high average prices, consider dynamic pricing strategies that adjust rates based on demand, seasonality, and local events. This can help maximize revenue for luxury and high-end listings.
2. **Premium Pricing for Unique Properties**: For unique or exceptional listings (like those in Brooklyn's Sea Gate), setting a premium price may be appropriate due to their uniqueness or luxury features.
3. **Competitive Pricing in Popular Neighborhoods**: In popular neighborhoods with a mix of room types, pricing competitively (slightly below the average) could increase bookings, especially for private rooms and shared spaces.
4. **Special Offers for Off-Peak Times**: During off-peak seasons or times, offering discounts or special packages can attract more guests, increasing occupancy rates and overall revenue.
5. **Review and Adjust Unknown Category**: Investigate and categorize the 'unknown' room types properly. Once categorized, apply neighborhood and room-type-specific pricing strategies.
6. **Segmented Approach**: Differentiate pricing strategies for different segments - luxury, mid-range, and budget accommodations should each have tailored pricing approaches.
7. **Monitor Competitors**: Regularly monitor competitor pricing in each neighborhood and room type to stay competitive and attractive to potential guests.

It's important to balance maximizing revenue with maintaining competitive pricing to attract guests. Regular market analysis and adjustments to pricing strategies are key to staying relevant in the dynamic short-term rental market.

Average Price by Neighborhood and Room Type (Top 10 Neighborhoods)

The visualization provides a clear representation of the average pricing across different neighborhoods and room types, particularly for the top 10 neighborhoods with the highest average prices:

1. **Variation in Prices by Neighborhoods**: The chart shows significant variation in average prices across different neighborhoods. This reflects the diverse range of property types and market demands in each area.
2. **High Prices in Certain Neighborhoods**: Neighborhoods like Brooklyn's Sea Gate and Manhattan's Theater District, Tribeca, and Flatiron District exhibit higher average prices, especially for entire homes/apartments and hotel rooms. This aligns with the understanding of these areas being more upscale or having unique properties.
3. **Room Type Impact on Pricing**: The influence of room type on pricing is evident. Entire homes/apartments generally command higher prices compared to private rooms or shared rooms. Hotel rooms in certain neighborhoods also show high pricing, consistent with the luxury or boutique hotels likely present in those areas.
4. **Pricing Anomalies**: The presence of high prices in the 'unknown' room type category in some neighborhoods suggests either premium properties not categorized under standard room types or a need for better data categorization.

This visualization supports the insight that neighborhood and room type are significant factors in pricing strategy. It highlights the need for differentiated pricing strategies depending on the location and type of accommodation, catering to the unique characteristics and demand in each market segment.

**Analyzing the relationship between price, number of reviews.**

The visualizations offer insights into the relationship between price, the number of reviews.

Price vs Number of Reviews:

- The scatter plot shows the relationship between the number of reviews and the price of listings. There isn't a clear linear relationship between the two variables. However, it appears that listings with a very high number of reviews tend to have moderate prices. This could indicate that more affordably priced listings are more frequently booked and reviewed.
- High-priced listings don't necessarily have a high number of reviews. This could be due to less frequent bookings or a shorter presence on the platform.

The visualization suggests that while the number of reviews does not directly correlate with higher prices, more affordable listings might receive more reviews due to higher booking frequency. It's important to consider that high prices might limit the number of bookings and reviews but cater to a different market segment compared to more frequently booked, moderately priced listings.

Price Distribution by Last Review across different months of 2019.
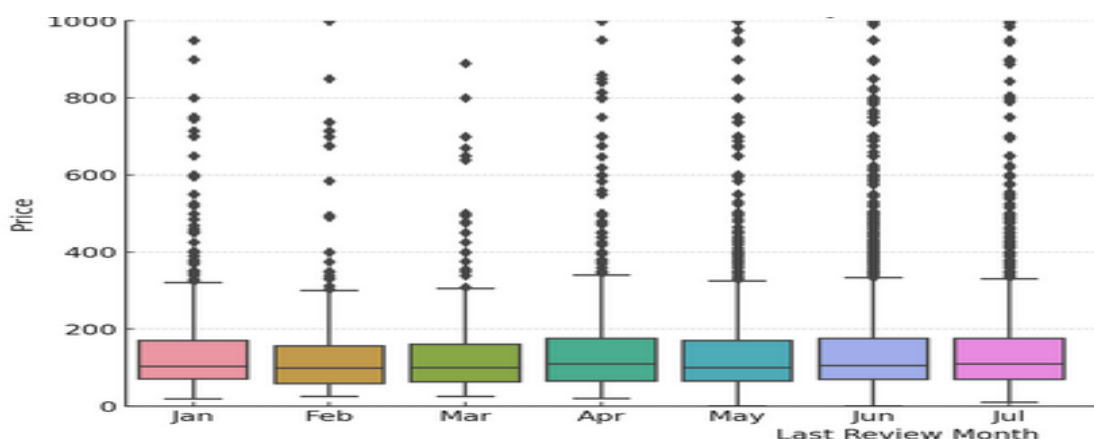
The visualization illustrates the price distribution by the last review month for the year 2019:

1. **Consistent Median Prices**: The median prices across different months remain relatively consistent. This suggests a stable pricing strategy throughout the year for the listings in the dataset.
2. **Monthly Variability**: There is noticeable variability in the upper price range each month, indicated by the length of the boxplot whiskers and the presence of outliers. This variability could be due to a mix of standard and luxury listings, seasonal pricing changes, or special events that lead to higher prices in certain months.
3. **Outliers**: Each month shows outliers above the upper whiskers, indicating that there are listings with significantly higher prices than the average in every month. These could be luxury properties or listings with unique features commanding higher prices.
4. **Seasonal Trends**: While the median price remains stable, there might be slight increases in the upper price ranges during peak travel months (like June, July, and August). However, the presence of outliers throughout the year suggests that high-priced listings are available regardless of the season.

This visualization indicates that while the general pricing strategy remains stable across months, there is a significant range in the prices, potentially accommodating various market segments from budget to luxury travelers. The consistent presence of high-priced outliers throughout the year also suggests a segment of the market consistently targeting higher-end customers.

Actionable Insights based on these visualizations:

Based on the "Price Distribution by Last Review Month (2019)" visualization, here are some actionable insights for hosts or property managers in the Airbnb market:

1. **Stable Pricing Strategy**: Given the relatively consistent median prices across months, maintaining a stable base price throughout the year seems effective. However, it's

important to regularly review and adjust prices based on market trends and competitor pricing.

2. **Seasonal Adjustments**: While the median price remains stable, the variability in the upper price range suggests opportunities for seasonal pricing adjustments. During peak travel months (like summer or holiday seasons), consider slightly increasing prices for properties, especially in high-demand areas, to capitalize on increased demand.

3. **Targeting Different Market Segments**: The presence of high-priced outliers in each month indicates a segment of the market with luxury or unique properties. Tailoring listings and marketing strategies to different segments - budget, mid-range, and luxury - can help maximize occupancy and revenue. For luxury properties, focus on highlighting unique features and experiences that justify higher prices.

4. **Promotions During Off-Peak Months**: To attract guests during traditionally slower months, consider offering special promotions or discounts. This can help increase occupancy rates when demand is lower.

5. **Monitoring and Dynamic Pricing**: Implement dynamic pricing tools that automatically adjust prices based on real-time market data, demand patterns, and local events. This approach can help optimize revenue, especially during periods of fluctuating demand.

6. **Quality and Services**: For listings that are priced higher (either consistently or during peak times), ensure that the quality and services offered match the pricing. High prices should correlate with high guest satisfaction to maintain good reviews and repeat business.

7. **Customized Experience Offerings**: For months with higher priced outliers, consider offering customized experiences or additional services (like guided tours, local experiences, etc.) that align with guests looking for a premium stay.

8. **Data-Driven Decision Making**: Regularly analyze market data, competitor pricing, and customer feedback to inform pricing strategies. Stay informed about local events and holidays that might affect travel patterns and adjust pricing and availability accordingly.

By applying these insights, hosts and property managers can more effectively strategize their pricing, catering to different customer needs and maximizing revenue across the year.

## Recommendations:

Suggestions for Hosts:

To increase the attractiveness of their Airbnb listings and consequently enhance guest experience and bookings, hosts can follow these best practices:

1. **High-Quality Photos**: Use professional, high-resolution photos that accurately represent the space. Include well-lit photos of all areas the guests can access. Consider seasonal updates to photos to reflect changes in decor or surroundings.

2. **Detailed and Accurate Descriptions**: Provide clear, detailed, and honest descriptions of the space. Include specifics about the amenities, local attractions, and any unique features of the property.
3. **Regular Updates and Maintenance**: Keep the listing information up to date. Regularly check and update the amenities, rules, and any changes in the local area that might be of interest to guests.
4. **Responsive Communication**: Respond promptly to inquiries and booking requests. Engaging quickly with potential guests increases the likelihood of securing bookings.
5. **Guest Reviews**: Actively encourage guests to leave reviews after their stay. Respond to reviews, both positive and negative, in a professional and courteous manner. Address any issues raised in negative reviews and highlight improvements made.
6. **Offer Competitive and Dynamic Pricing**: Use dynamic pricing strategies to stay competitive and maximize earnings, especially during peak seasons or special events in your area.
7. **Local Information and Recommendations**: Provide guests with a guidebook or information about local attractions, restaurants, and tips. Personal recommendations can enhance the guest experience.
8. **Amenities and Comfort**: Invest in quality bedding, towels, and essential amenities. Offer additional perks like free Wi-Fi, coffee/tea, or small toiletries.
9. **Cleanliness and Safety**: Maintain impeccable cleanliness. During times when health concerns are paramount, ensure to follow enhanced cleaning protocols. Safety features like smoke detectors, first aid kits, and clear emergency instructions should be in place.
10. **Check-In/Check-Out Process**: Simplify the check-in and check-out process. Consider keyless entry or secure lockbox systems for convenience.
11. **Unique Experiences**: Offer something unique that sets your listing apart, like local artwork, unique interior design, or experiences like a cooking class or bike tours.
12. **Leverage social media and Online Marketing**: Use social media platforms to promote your listing. Share updates, special offers, or interesting aspects of your property or location.
13. **Solicit Feedback for Continuous Improvement**: Ask guests for feedback on what could be improved and act on this feedback to continually enhance the guest experience.
14. **Sustainability Practices**: Implement eco-friendly practices like recycling, using energy-efficient appliances, and offering sustainable amenities. This can be a selling point for environmentally conscious travelers.
15. **Personal Touches and Hospitality**: Small personal touches like a welcome note, a small gift, or offering assistance with luggage can make a big difference in a guest's experience.

By implementing these practices, hosts can significantly increase the attractiveness of their listings, leading to higher occupancy rates, better reviews, and potentially higher revenue.

Impact of Recency of Review (Last review date) on perceived quality or popularity of a listing.

Impact of Review Recency on Number of Reviews

Impact of Review Recency on Price
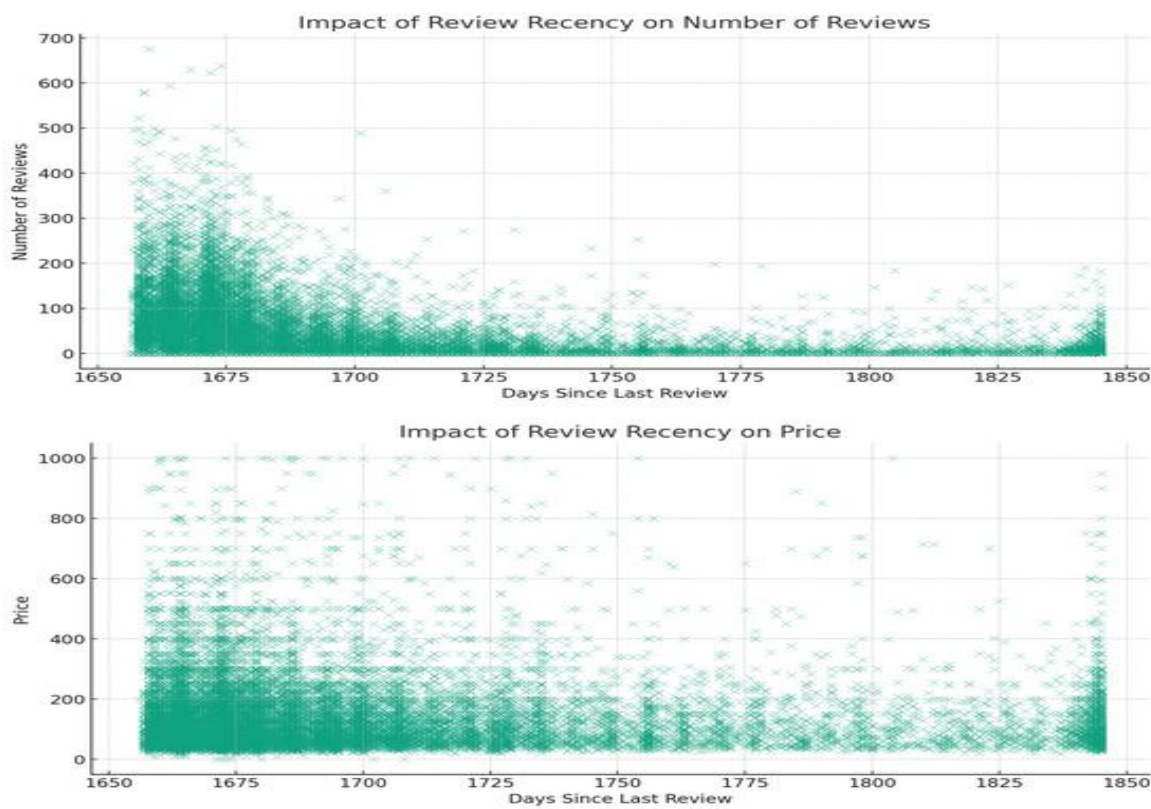
The visualizations provide insights into the impact of review recency on the number of reviews and the price of Airbnb listings:

1. **Impact of Review Recency on Number of Reviews**:
   o The first chart shows that listings with more recent reviews tend to have a higher number of reviews. This suggests that active listings, which are regularly booked and reviewed, accumulate more reviews over time.
   o Listings with a long time since the last review tend to have fewer reviews, possibly indicating less frequent bookings or newer listings.
2. **Impact of Review Recency on Price**:
   o The second chart illustrates the relationship between the recency of reviews and the price of listings. There isn't a strong visible correlation between the days since the last review and the price. This suggests that the recency of reviews may not directly influence the pricing strategy of a listing.
   o Both high and low-priced listings can have recent or older reviews, indicating diverse pricing strategies regardless of review recency.

Actionable Insights:

- **Encourage Recent Reviews**: Hosts should encourage guests to leave reviews promptly after their stay to ensure the listing appears active and popular.

- **Regularly Update Listings**: Keeping the listing updated and responding to reviews can help maintain guest interest and lead to more frequent bookings.
- **Price Strategy**: While review recency doesn't directly correlate with pricing, understanding market trends and competitor pricing is crucial for effective pricing strategies.
- **Focus on Guest Experience**: Ensuring a high-quality guest experience can lead to more positive and recent reviews, potentially increasing the listing's attractiveness and booking frequency.

These insights highlight the importance of active management and guest engagement in maintaining an attractive and competitive Airbnb listing.
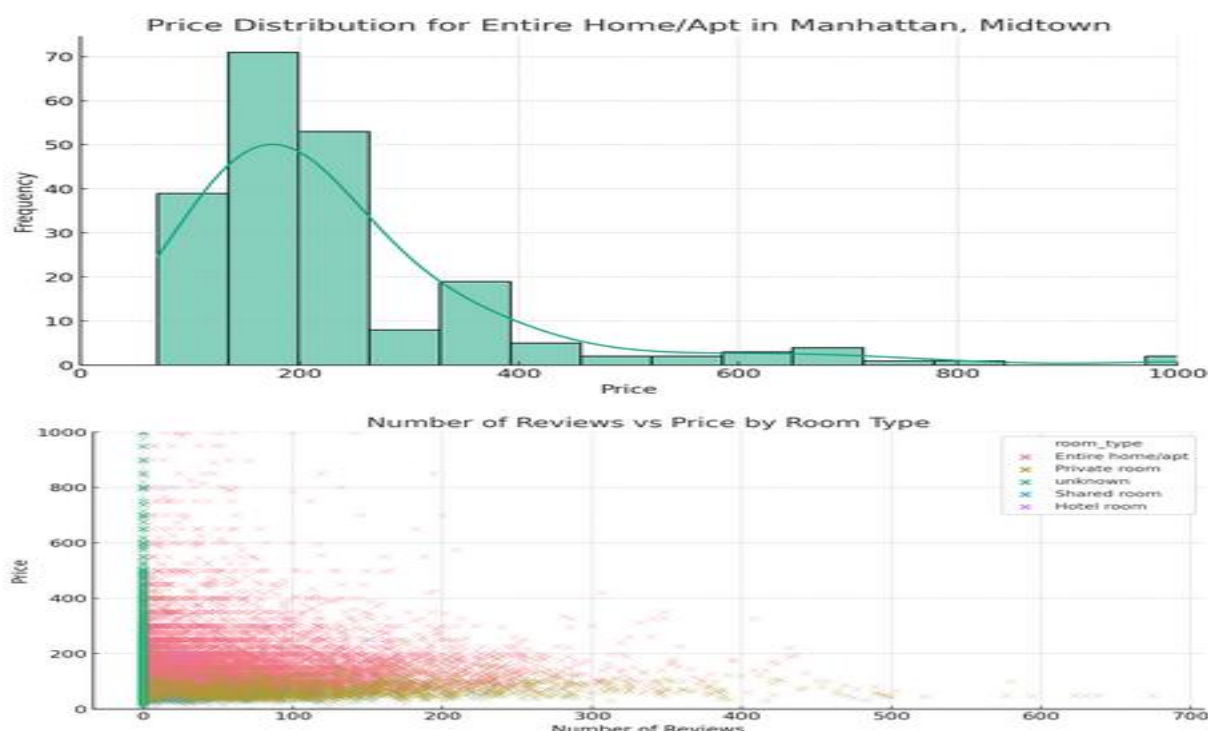
Recommendations

Based on the analysis, here are some key recommendations for a new Airbnb host, supported by visualizations to illustrate these points:

1. **Price Competitively**:
    - Start with a competitive pricing strategy, especially if you're entering a market with a variety of options. Use the median price in your neighborhood and room type as a benchmark.
2. **Focus on Getting Reviews**:
    - Early reviews are crucial. Consider initially pricing slightly lower to attract bookings and accumulate reviews. Positive reviews can significantly impact your listing's attractiveness.
3. **Maintain an Active and Updated Listing**:
    - Regularly update your listing information and respond promptly to guest inquiries and reviews. This demonstrates attentiveness and can influence potential guests.
4. **Quality of Listing**:
    - Invest in high-quality photos and detailed, accurate descriptions. Highlight unique features of your space and neighborhood.
5. **Guest Experience**:
    - Provide excellent hospitality to ensure positive guest experiences, which leads to good reviews. Small touches can make a big difference.

Here are a few visualizations to support these recommendations:

- **Visualization 1**: Showing the distribution of prices in a particular neighborhood and room type to guide competitive pricing.
- **Visualization 2**: Illustrating the importance of reviews on booking frequency.

    Let's proceed with the visualizations:

Price Distribution for Entire Home/Apt in Manhattan, Midtown

Number of Reviews vs Price by Room Type

The visualizations support the recommendations for a new Airbnb host as follows:

1. **Price Distribution for Entire Home/Apt in Manhattan, Midtown**:
   - The histogram shows the distribution of prices for "Entire home/apt" listings in Manhattan, Midtown. This visualization can guide a new host in setting a competitive price. For instance, positioning the price around the peak of the distribution (which represents the most common price range) could be a good starting point.
2. **Number of Reviews vs Price by Room Type**:
   - The scatter plot illustrates the relationship between the number of reviews and the price, segmented by room type. It's noticeable that listings with a higher number of reviews tend to have moderate prices, suggesting that competitively priced listings may receive more bookings and, consequently, more reviews. For a new host, this emphasizes the importance of accumulating reviews, which can be facilitated by initially pricing the listing attractively.

Recommendations for a New Host:

- **Competitive Pricing**: Start with a price that's competitive within your specific neighborhood and room type. Use market data to inform your pricing strategy.

- **Prioritize Reviews**: Aim to gather positive reviews quickly, as they significantly impact the attractiveness of your listing. Consider pricing strategies or promotions that encourage initial bookings.
- **Quality and Accuracy**: Ensure your listing stands out with high-quality photos and detailed descriptions. Highlight unique aspects of your space.
- **Responsive and Active Hosting**: Be responsive to guest inquiries and feedback. Regularly update your listing to keep it active and engaging.

By adhering to these recommendations and continuously evaluating market trends and guest feedback, a new Airbnb host can effectively position their listing in the market and enhance their chances of success.