Logo

Description automatically generated

**Module 4: Final Project**

**By**

**Smit Pareshbhai Ranpariya, Neha Ajinkya Nagurkar & Vishvesh Gaurangbhai Vyas**

**Project: Airbnb**

**College of Professional Studies: Northeastern University**

**Prof. Hootan Kamran**

**Subject: ALY6110: Big Data and Data Management**

**Date: June 8, 2024**

#### Project: **Basic Analysis and Dashboard**

### **Introduction**

This comprehensive Airbnb dataset provides a detailed overview of the platform's listings in New York City, spanning various neighborhoods, room types, and pricing strategies. The data includes 48,895 entries with key attributes such as listing ID, host information, geographic coordinates, room type, price, and review metrics. Visualizations such as bar graphs, heatmaps, treemaps, and maps offer insights into trends like annual review counts, average monthly reviews, price variations, booking patterns, and host performance. This rich dataset facilitates a deep understanding of Airbnb's market dynamics, revealing how different areas and types of accommodations perform, seasonal booking trends, and host successes. These insights are invaluable for market analysis, helping stakeholders—hosts, guests, and analysts—make informed decisions and strategies to optimize their participation in the Airbnb ecosystem.

**Why Airbnb?**

Airbnb revolutionizes the travel and accommodation industry by offering diverse, personalized lodging experiences that go beyond traditional hotel stays. It caters to a wide spectrum of travelers—from budget-conscious backpackers to luxury-seeking tourists—by providing access to unique accommodations such as private rooms, entire homes, castles, treehouses, and even yachts. This variety allows guests to choose stays that suit their preferences and budgets, fostering a more tailored travel experience.

One of Airbnb's core strengths is its community-based approach. Both hosts and guests benefit from a robust review system that builds trust and ensures quality. Positive reviews can enhance a host's reputation, leading to higher occupancy rates and better earnings, while guests rely on these reviews to make informed decisions about where to stay.

Airbnb also contributes significantly to the local economy. Hosts can monetize their unused spaces, providing them with an additional income stream. This is especially beneficial in cities where the cost of living is high. Guests, on the other hand, often find Airbnb rentals to be more affordable than hotels, especially for longer stays or when traveling in groups.

Furthermore, Airbnb enables guests to immerse themselves in local culture. Staying in residential neighborhoods rather than tourist hotspots allows travelers to experience destinations more authentically, interacting with locals and discovering hidden gems.

The platform’s user-friendly interface and widespread availability make it accessible and convenient for users worldwide. It integrates seamlessly with travel plans, offering features like trip planning, experiences, and local recommendations. In essence, Airbnb redefines travel by merging comfort, affordability, and cultural immersion, making it a preferred choice for millions of travelers globally.

**Dataset Overview**

The Airbnb dataset under review encompasses 48,895 entries, each detailing various aspects of individual listings. Key attributes include the listing ID, name, host ID, host name, neighborhood group, neighborhood, latitude, longitude, room type, price, minimum nights required, number of reviews, last review date, reviews per month, calculated host listings count, and availability over the year.

Geographically, the listings cover multiple neighborhoods across New York City, providing latitude and longitude coordinates for precise spatial analysis. This allows for detailed studies on the distribution of Airbnb properties and their concentration in different urban areas.

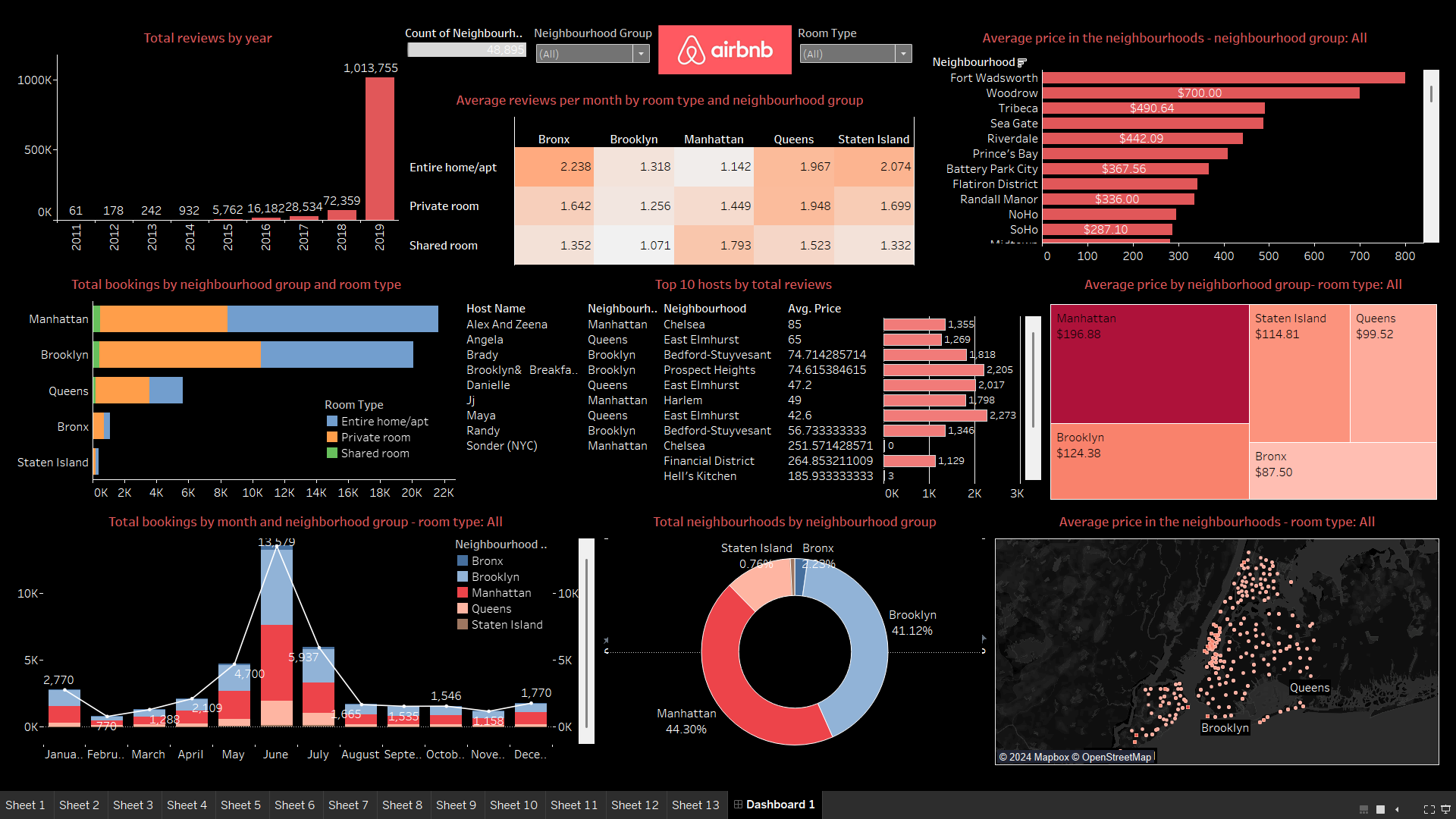
The dataset categorizes accommodations into different room types, including private rooms, entire homes/apartments, and shared spaces, highlighting the variety of options available to guests. Pricing information shows significant variability, reflecting the range of economic levels and the diversity of offerings—from budget-friendly stays to luxurious accommodations.

Review data, including the total number of reviews and average reviews per month, serves as a proxy for guest satisfaction and listing popularity. The last review date helps in assessing the recency of guest feedback, which is crucial for evaluating current host performance and property conditions.

The availability attribute indicates how often a listing is available throughout the year, which can be used to gauge occupancy rates and potential earnings. The calculated host listings count reveals whether hosts manage multiple properties, offering insights into the prevalence of professional hosting versus casual, single-property listings.

Overall, this dataset is a rich resource for understanding the dynamics of Airbnb's market in New York City. It supports diverse analyses, from market trends and pricing strategies to the impact of short-term rentals on local housing markets and tourism patterns.

**Airbnb Dashboard**



The Airbnb dashboard provides a comprehensive analysis of key metrics related to listings in New York City. It highlights the growth in reviews from 2011 to 2019, demonstrating increasing user engagement. The dashboard showcases the total listings by neighborhood, average monthly reviews by room type and neighborhood, and average listing prices across different areas. It also details total bookings by neighborhood and room type, top hosts by total reviews, and average prices by neighborhood groups. Seasonal booking trends are illustrated, showing peak demand in summer. Additionally, the geographic distribution of listings and their prices are mapped, indicating dense clusters and price variations across the city. This dashboard aids in understanding market trends, guest preferences, and host performance for strategic decision-making.

**Total Reviews by Year**

This bar graph tracks the total number of Airbnb reviews received each year from 2011 to 2019. Starting with a modest 61 reviews in 2011, there is a noticeable and consistent increase over the years. The reviews surge dramatically in 2018 and peak in 2019 with 1,013,755 reviews. This exponential growth indicates the rising popularity and increased user engagement with Airbnb over the years. The sharp increase in reviews, especially in the last two years, may reflect both an increase in the number of listings and more frequent usage by guests. This trend is significant for understanding Airbnb's market expansion and its growing acceptance among travelers.

**Count of Neighbourhood Group**

This count, totaling 48,895 listings, provides an overview of Airbnb's penetration across various neighborhood groups. The graph helps in identifying the distribution of listings across different areas. With filters for neighborhood groups and room types, users can drill down to see more specific data, such as the number of listings in a particular neighborhood or the prevalence of certain room types. This count is essential for understanding Airbnb's market spread and for conducting detailed spatial and market analyses.

**Average Reviews per Month by Room Type and Neighbourhood Group**

This heatmap compares the average monthly reviews received by different room types across various neighborhood groups. For instance, entire homes/apartments in the Bronx receive an average of 2.238 reviews per month, the highest among all combinations. This visualization highlights which room types and neighborhoods are most popular with guests, providing insights into user preferences and market demand. By examining these averages, hosts and analysts can gauge the performance and attractiveness of different listings.

**Average Price in the Neighbourhoods - Neighbourhood Group: All**

This bar chart showcases the average price of Airbnb listings across different neighborhoods. Fort Wadsworth leads with an average price of $700, followed by Woodrow at $490.64. This data highlights significant price disparities between neighborhoods, indicating varied market segments. High average prices in certain neighborhoods suggest a concentration of luxury accommodations or high-demand areas. Conversely, lower prices in other neighborhoods may attract budget-conscious travelers. Understanding these price variations helps in market positioning and pricing strategies for hosts.

**Total Bookings by Neighbourhood Group and Room Type**

This stacked bar chart displays the total number of bookings in each neighborhood group, segmented by room type. Manhattan and Brooklyn dominate in total bookings, with a high preference for entire homes/apartments. This visualization reveals the most popular accommodation types and areas, providing valuable insights for potential hosts and market analysts. By understanding which combinations of room types and locations attract the most bookings, stakeholders can tailor their offerings to meet market demand and maximize occupancy rates.

**Top 10 Hosts by Total Reviews**

This bar chart lists the top 10 Airbnb hosts based on the total number of reviews received, highlighting 'Alex And Zeena' from Manhattan with 85 reviews as the top host. The chart also includes average price data for these hosts, providing a comparative look at how review counts correlate with listing prices. This information is crucial for understanding what makes top hosts successful, whether it’s their service quality, pricing strategy, or location. New hosts can use these insights to improve their own listings and customer service to attract more reviews.

**Average Price by Neighborhood Group - Room Type: All**

This treemap displays the average price of Airbnb listings by neighborhood group, with Manhattan ($196.88) being the highest and Bronx ($87.50) the lowest. The varying sizes and color intensities of the segments highlight price disparities across neighborhoods. This visualization helps in quickly assessing which areas are more expensive and may cater to higher-end markets versus more affordable neighborhoods. It provides a clear comparative view of pricing, assisting hosts in setting competitive rates and helping guests find accommodations that fit their budgets.

**Total Bookings by Month and Neighborhood Group - Room Type: All**

This line chart captures the monthly booking trends across different neighborhood groups, showing a clear peak in June with 5,937 bookings. The seasonal pattern suggests higher demand during summer months, with noticeable lows in January and February. Such seasonal trends are essential for hosts to understand booking behaviors and to plan their pricing and availability accordingly. It also helps potential guests to anticipate peak travel periods, which can impact their travel planning and budget.

**Total Neighbourhoods by Neighbourhood Group**

This donut chart presents the distribution of Airbnb listings across different neighborhood groups, showing Manhattan (44.30%) and Brooklyn (41.12%) as the dominant areas. This distribution highlights the primary focus areas for Airbnb activity, indicating where the majority of listings and likely the demand is concentrated. Understanding this distribution is crucial for market analysis, helping to identify potential growth areas and the overall market landscape of Airbnb in New York City.

**Average Price in the Neighbourhoods - Room Type: All**

This map visualizes the geographic distribution of Airbnb listings and their average prices. Each dot represents a listing, with denser clusters in areas like Manhattan and Brooklyn, indicating high listing concentrations. The color coding by price provides a visual understanding of the spatial price distribution, helping to identify expensive versus more affordable areas. This spatial representation is valuable for both hosts and guests in identifying prime locations and understanding how location influences pricing.

**Model Implemented in R**

**Summary of the Dataset**

The summary statistics give us an overview of the data:

* **Price**: The price of listings ranges significantly, with a maximum of 10,000 and a median of 101. This suggests a wide variability in listing prices, potentially due to differences in location, room type, or amenities offered.
* **Minimum Nights**: Most listings seem to have a minimum stay requirement of just a few nights, which is typical for short-term rentals.
* **Number of Reviews**: Listings have up to 629 reviews, but the median is 9, indicating that while some listings are very popular, most have a moderate amount of feedback.
* **Review Frequency**: The reviews\_per\_month have a wide range as well, with a maximum of 58.5, suggesting some listings are reviewed very frequently, possibly indicating high turnover rates.

**K-Means Clustering**

The k-means clustering output shows centroids of five clusters, which are likely to correspond to different neighborhoods or areas within a city:

* **Cluster Centers**: The latitude and longitude cluster centers are fairly close to each other, which could indicate that the listings are concentrated in a relatively small geographical area or that the city itself is not very large in terms of Airbnb market spread.

**Correlation Matrix**

The correlation matrix shows relationships between different variables:

* **Price and Availability**: There's a positive correlation (0.078) between price and availability\_365, suggesting that higher-priced listings might be available more days in the year.
* **Reviews and Review Frequency**: There's a strong positive correlation (0.549) between number\_of\_reviews and reviews\_per\_month, which implies that listings with more reviews tend to have a higher frequency of reviews. This could indicate higher guest turnover and/or higher popularity.
* **Availability and Review Frequency**: There's also a positive correlation (0.186) between availability\_365 and reviews\_per\_month, which might suggest that listings available for more days get reviewed more often, perhaps due to higher booking rates.

**Key Insights**

1. **Variable Listing Prices**: The wide range of prices indicates a diverse market, with options ranging from budget to luxury. The median price provides a better indication of the typical cost for a traveler, as the mean could be skewed by very high-priced listings.
2. **Geographic Market Segmentation**: The clustering results show distinct geographical groupings, which can be useful for identifying popular areas or regions with higher or lower pricing dynamics.
3. **Popularity and Engagement**: The strong correlation between the number of reviews and reviews per month suggests that more reviews may lead to more visibility and subsequent bookings, emphasizing the importance of managing customer satisfaction and feedback.

These insights can be used by Airbnb hosts to optimize their listings and pricing strategies, and by the platform to understand market dynamics and inform strategic decisions.

**Code**

**A computer screen shot of a computer program

Description automatically generated**

**Output**

**A screen shot of a computer screen

Description automatically generated**

### **Conclusion**

The Airbnb dataset for New York City reveals significant insights into the platform's operations, market trends, and user preferences. Visual analyses indicate a sharp increase in reviews and popularity, particularly peaking in 2019, and highlight substantial price variations across different neighborhoods. Manhattan and Brooklyn emerge as dominant areas for listings and bookings, with entire homes/apartments being the most popular room type. Seasonal booking trends show a peak in summer months, providing strategic timing information for hosts. High-performing hosts are identified by their review counts and pricing strategies, offering benchmarks for others. Overall, this dataset serves as a valuable resource for understanding Airbnb's market behavior, guiding hosts in optimizing their listings, and helping guests find suitable accommodations. The insights derived from this data can drive informed decision-making and strategic planning within the Airbnb community.

**Reference:**

[**https://www.kaggle.com/datasets/dgomonov/new-york-city-airbnb-open-data**](https://www.kaggle.com/datasets/dgomonov/new-york-city-airbnb-open-data)