# PROJECT PRESENTATION

# EXPLORATORY DATA ANALYSIS

# AIRBNB BOOKING

**By,**
**Shafiq Abubacker**

# CONTENT

- The Airbnb Exploratory Data Analysis project aims to delve into the vast dataset provided by Airbnb, uncovering valuable insights into the dynamics of property listings, guest preferences, host performance, and neighborhood characteristics.

- Through rigorous analysis and visualization, we seek to better understand the factors influencing the Airbnb ecosystem, providing actionable insights for hosts, potential investors, and the Airbnb platform itself.

# OVERVIEW

- Airbnb, as a global platform connecting hosts and guests since 2008, has amassed a dataset comprising around 49,000 observations with 16 columns, encompassing both categorical and numeric variables.

- The dataset includes a mix of data types, such as integers, floats, and objects (text or categorical data).

# OBJECTIVES

- **Pricing Analysis:** Understand the distribution of prices, identify factors influencing pricing, and explore optimal pricing strategies.
- **Geographical Analysis:** Visualize the distribution of listings on a map, identify popular neighborhoods, and explore trends in different boroughs.
- **Host Performance Metrics**: Analyze the relationship between host metrics (e.g., host count, listings count) and other variables to understand host performance.
- **User Experience Analysis:** Explore the number of reviews, reviews per month, and last review date to gauge user satisfaction and identify areas for improvement.
- **Seasonal Trends:** Analyze availability patterns throughout the year (availability_365) to understand seasonal demand.
- **Average Price per Neighborhood Analysis:** Exploring the variation in listing prices across different neighborhoods which can provide valuable insights into the pricing dynamics of Airbnb listings in various areas.

# DATA FEATURES

- **id:** Unique identifier for each listing.
- **name:** Name or title of the listing.
- **host_id:** Unique identifier for each host.
- **host_name:** Name of the host.
- **neighbourhood_group:** The geographical grouping of neighborhoods.
- **neighbourhood:** Specific neighborhood where the listing is located.
- **latitude and longitude:** Geographic coordinates of the listing.
- **room_type:** Type of accommodation (e.g., entire home/apartment, private room, shared room).
- **price:** Listing price per night.
- **minimum_nights:** Minimum nights required for booking.
- **number_of_reviews:** Total number of reviews for the listing.
- **last_review:** Date of the last review.
- **reviews_per_month:** Average number of reviews per month.
- **calculated_host_listings_count:** Count of listings by the host.
- **availability_365:** Number of days the listing is available in a year.

# EDA TECHNIQUES

- Descriptive Statistics
  - Comprehensive overview of the dataset's central tendencies and spread
- Data Visualization
  - Recognizing patterns, trends, and outliers, offering an intuitive understanding of the dataset's structure
- Correlation Analysis
  - Assessing the relationships between different variables
- Distribution Analysis
  - Identifying anomalies or unusual patterns in the dataset
- Categorical Data Exploration
  - Distribution of property types, neighborhood characteristics, and other categorical features
- Temporal Analysis
  - Identifying trends and seasonality

# DATA CLEANING

```
[ ]  # Import Libraries
     import pandas as pd
     import numpy as np
     import matplotlib.pyplot as plt
     import seaborn as sns
```

```
[ ]  # Load Dataset
     from google.colab import drive
     drive.mount('/content/drive')
```

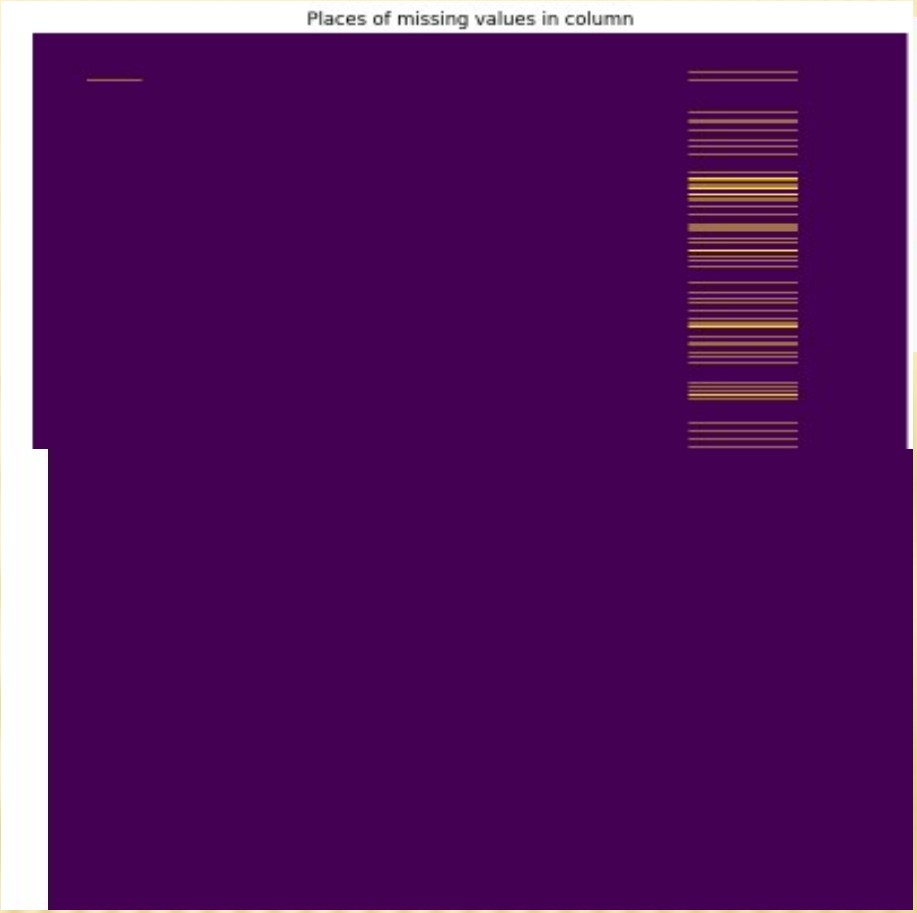
```
     # Dataset Duplicate Value Count
     df1.duplicated().value_counts()
```

```
False    48895
```

```
# Missing Values/Null Values Cour
df1.isnull().sum().sort_values(as
```

```
last_review
reviews_per_month
host_name
name
id
host_id
neighbourhood_group
neighbourhood
latitude
longitude
```

```
df = pd.read_csv("/content/drive/MyDrive/Almabetter/01 Python/EDA/Airbnb/Airbnb NYC 2019.csv")
```



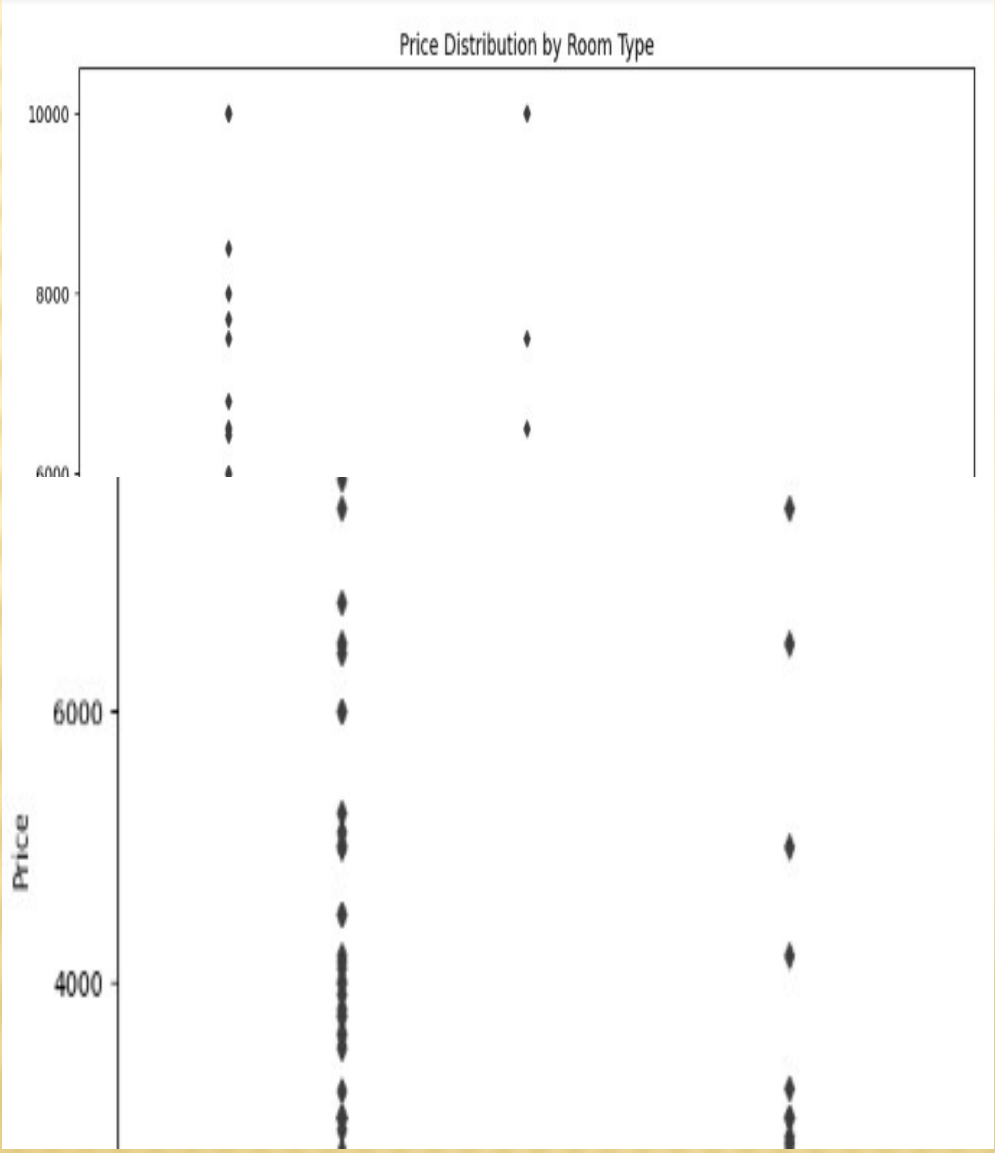

```
#Replacing Null values with 0.
null_columns=['last_review','reviews_per_month']
for col in null_columns:
  df1[col].fillna(0,inplace=True)

  # Replacing NA values with 'others'
dnull_columns=['host_name','name']
for col2 in dnull_columns:
  df1[col2].fillna("Others",inplace=True)

df1.isnull().sum().sort_values(ascending = False)
```

```
id
name
host_id
host_name
neighbourhood_group
neighbourhood
latitude
longitude
room_type
price
minimum_nights
```

Considering the missing values in 'host_name' and 'name,' they can be replaced with "Other," and for 'last_review' and 'reviews_per_month,' zeros can be used, indicating potential incompleteness or optionality during listing creation.

# TRENDS AND PATTERNS

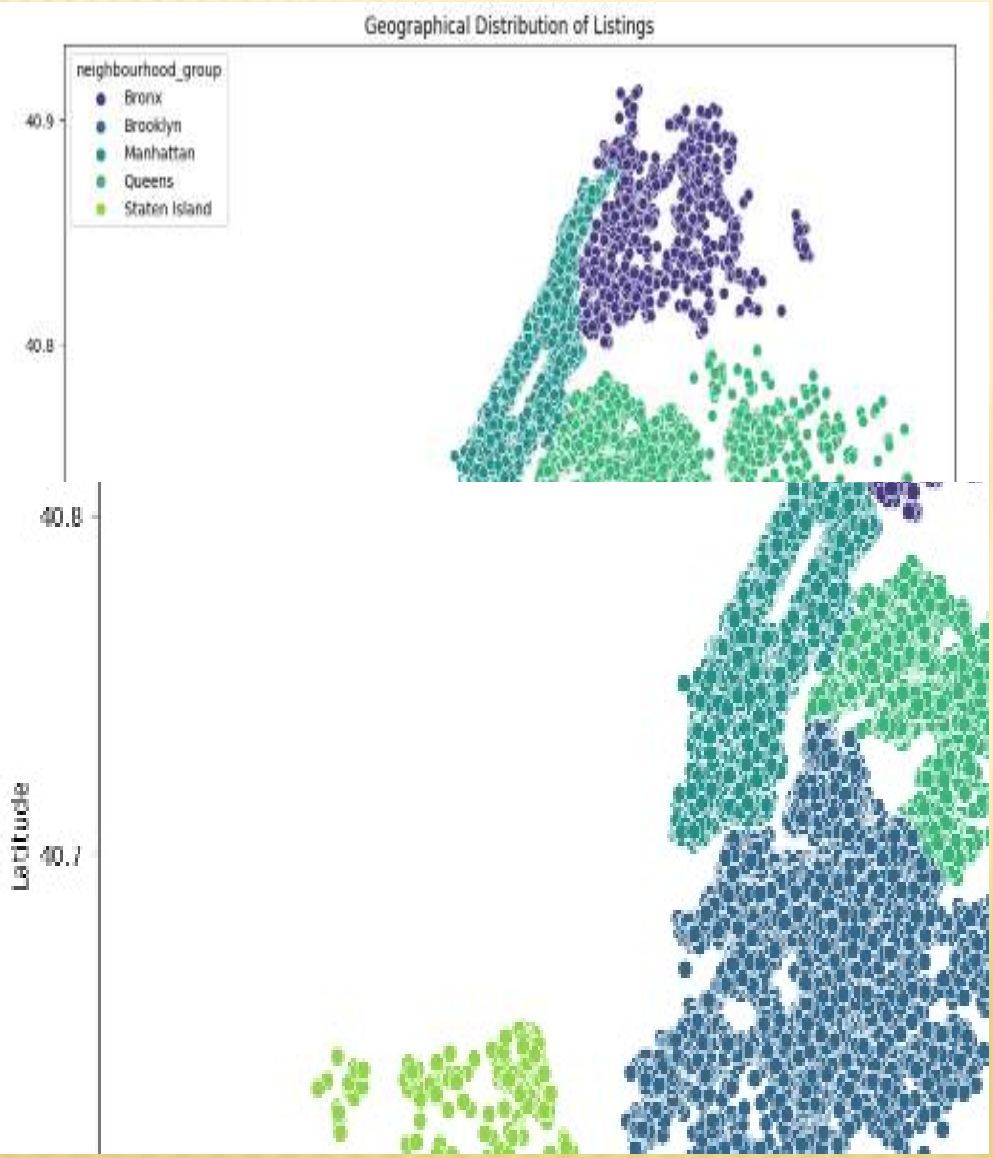
## PRICING ANALYSIS




- **Premium for Privacy and Convenience:**
  - The Price Distribution by Room Type chart indicates a clear hierarchy in pricing, with Entire home/apartment listings being the most expensive, followed by Private rooms and Shared rooms.
- **Privacy Gradation:**
  - Entire home/apartment listings offer the highest level of privacy and convenience, providing guests with exclusive use of the entire space.
  - Private rooms, while offering less privacy than entire units, still provide a more private experience compared to Shared rooms.
- **Demand Dynamics:**
  - The chart underscores that users are willing to pay a premium for increased privacy and convenience. Entire home/apartment listings, being the most expensive, are also the most in-demand, while Shared rooms, the least expensive, have lower demand, likely due to the shared nature of the space.

# TRENDS AND PATTERNS

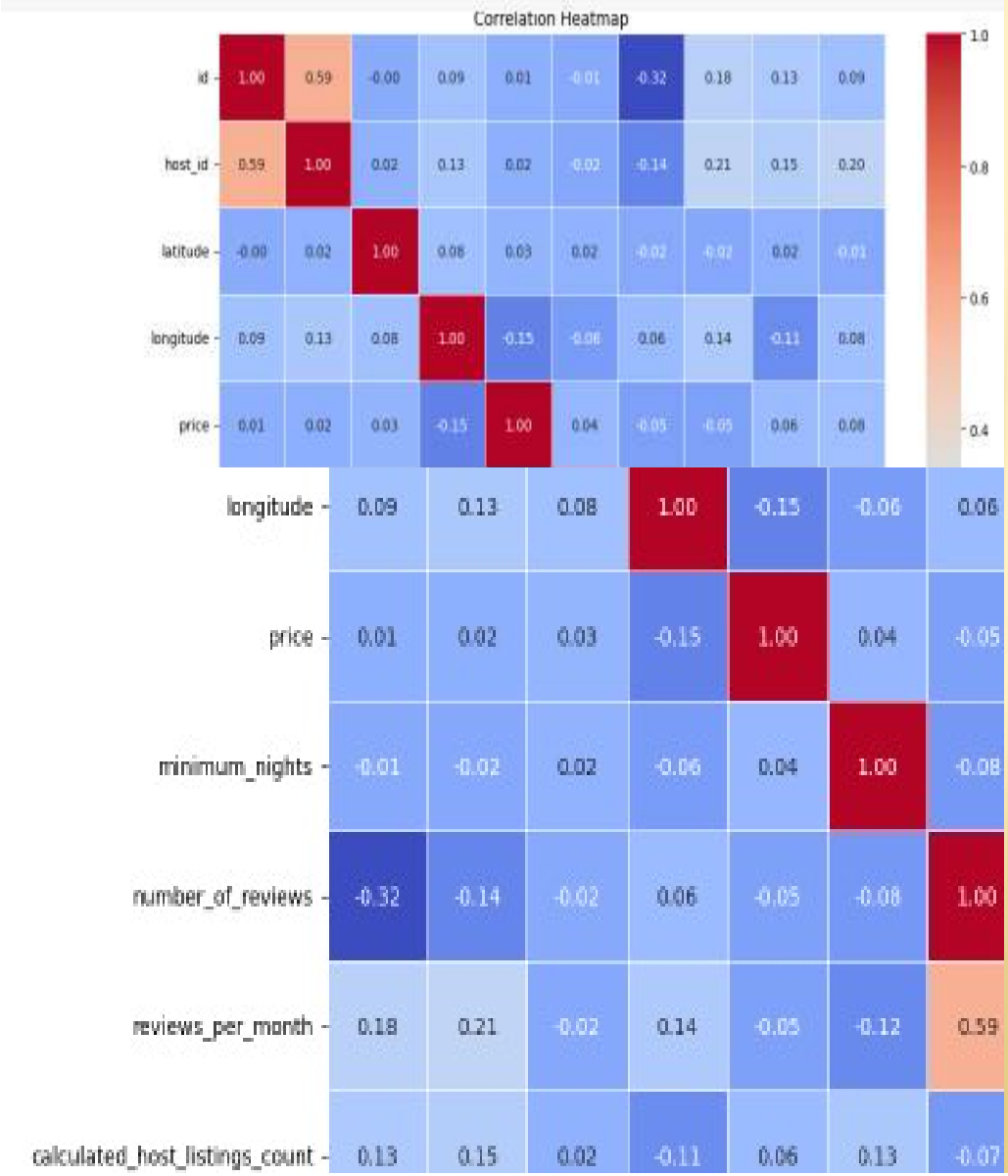
## GEOGRAPHICAL ANALYSIS




- **Listing Concentration by Borough:**
  - Manhattan boasts the highest concentration of Airbnb listings, indicating its status as the most sought-after tourist destination in New York City.
- **Rising Popularity of Brooklyn and Queens:**
  - Brooklyn and Queens are gaining popularity among travelers, offering a more affordable and authentic NYC experience compared to Manhattan. Their proximity to the city center enhances their appeal for Airbnb guests.
- **Staten Island's Lower Demand:**
  - Staten Island exhibits the lowest popularity among tourists and business travelers, resulting in a lower demand for Airbnb listings. Hosts in this borough may face pricing challenges due to decreased demand compared to other boroughs.

# TRENDS AND PATTERNS

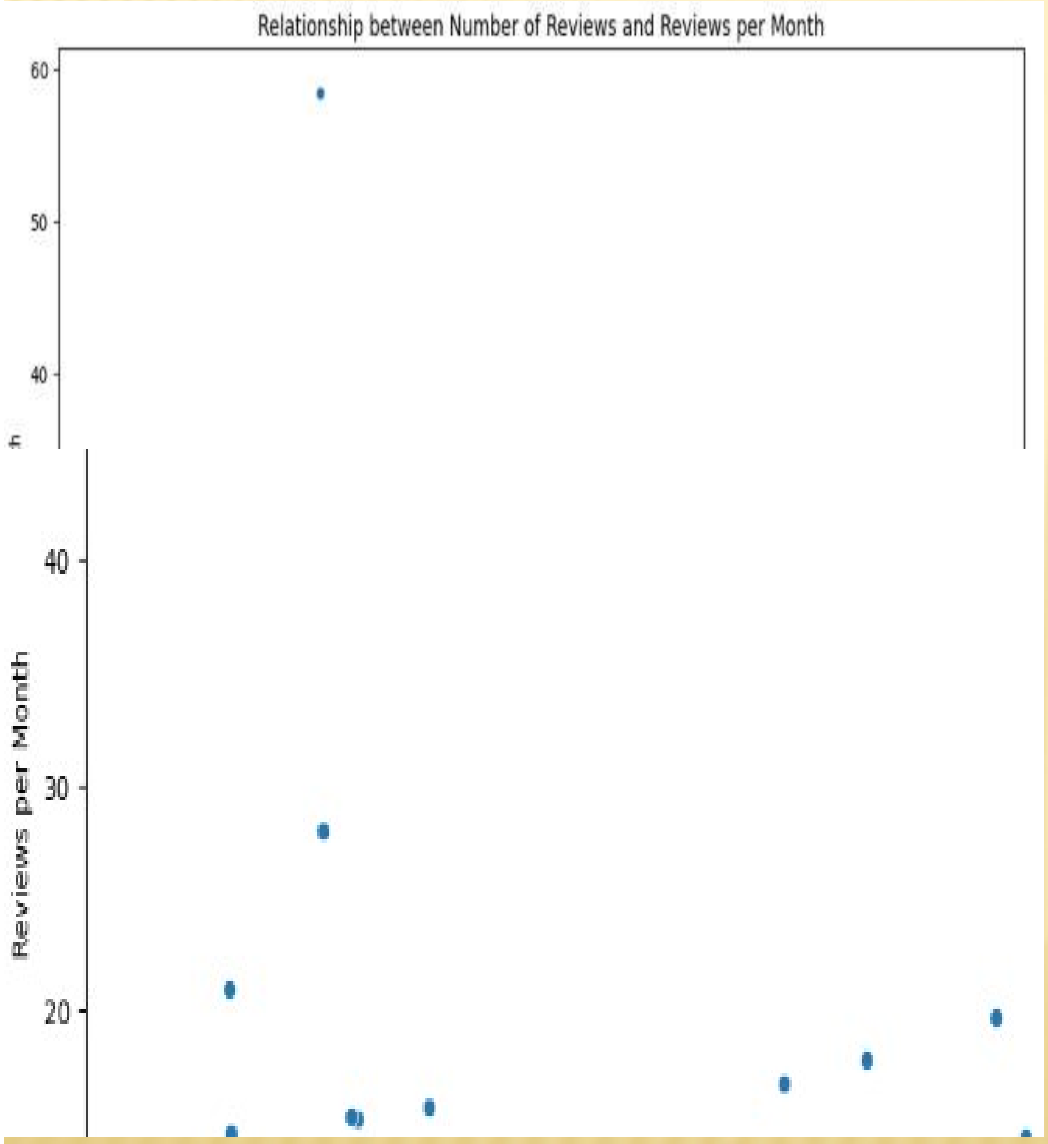
## HOST PERFORMANCE METRICS




- **Positive Correlation Insights:**
  - A positive correlation exists between the number of listings, number of reviews, and availability, indicating that hosts excelling in one metric are likely to excel in others as well.
- **Performance Metrics Alignment:**
  - Hosts with a greater number of listings demonstrate higher review counts and response rates. This correlation suggests that hosts deeply committed to their Airbnb business are more inclined to deliver a positive guest experience.
- **Investment and Guest Satisfaction:**
  - The connection between higher listing numbers and positive performance metrics implies that hosts heavily invested in their Airbnb endeavors are more apt to provide a satisfactory guest experience, as reflected in review counts and responsiveness.

# TRENDS AND PATTERNS

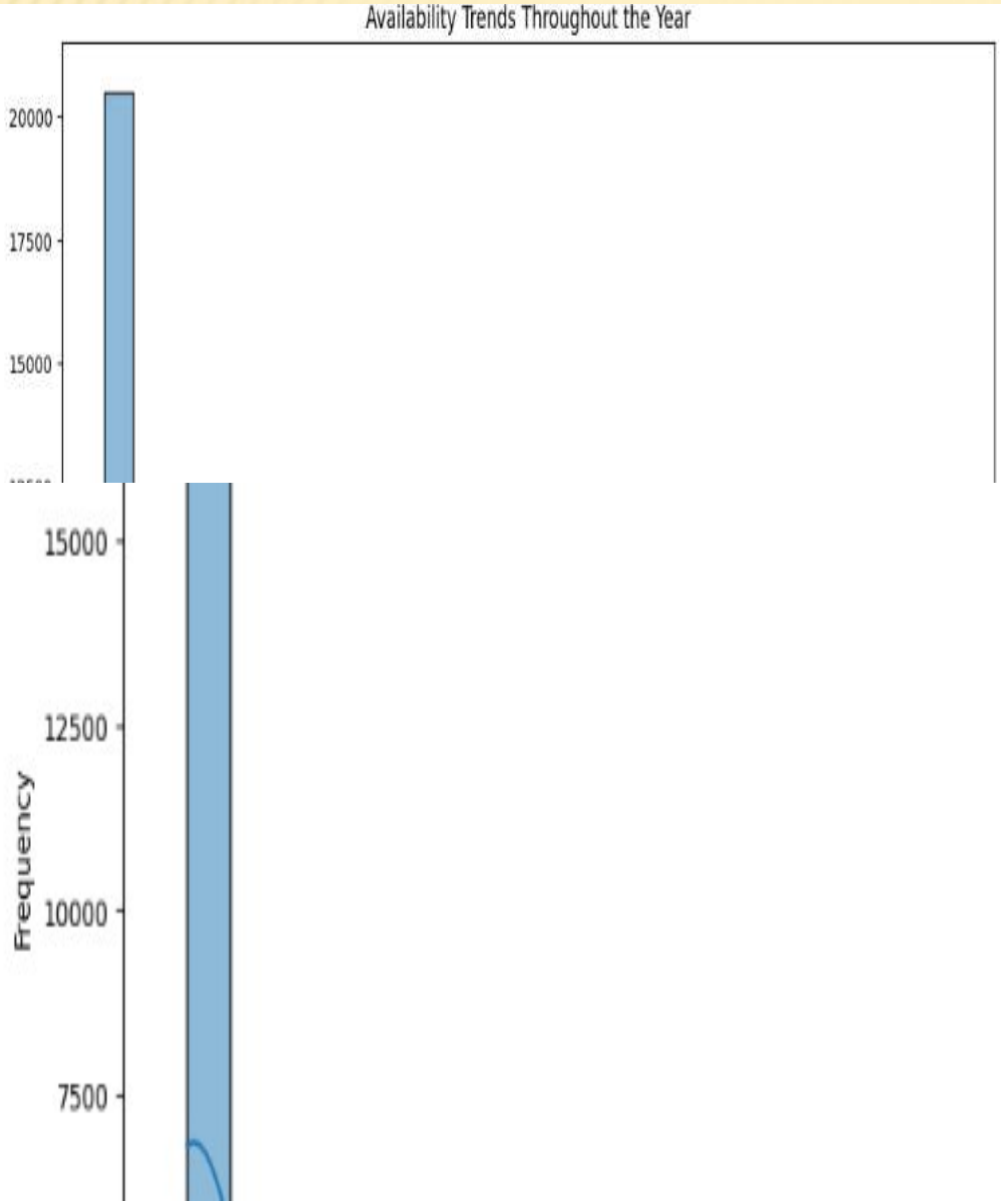
## USER EXPERIENCE ANALYSIS




- **Positive Correlation Highlight:**
  - The chart illustrates a positive correlation between the total number of reviews and reviews per month, emphasizing a symbiotic relationship where listings with a higher overall review count also tend to receive more reviews on a monthly basis.
- **Focus on Guest Experience:**
  - The insight from the plot suggests that hosts should prioritize providing an exceptional guest experience. A positive experience contributes to a higher total review count, which, in turn, correlates with increased monthly reviews.
- **Enhanced Visibility and Guest Attraction:**
  - Listings accumulating more reviews are likely to gain higher visibility in search results, attracting a larger number of guests. This underlines the importance of consistently delivering positive guest experiences to optimize search placement and overall listing performance.

# TRENDS AND PATTERNS

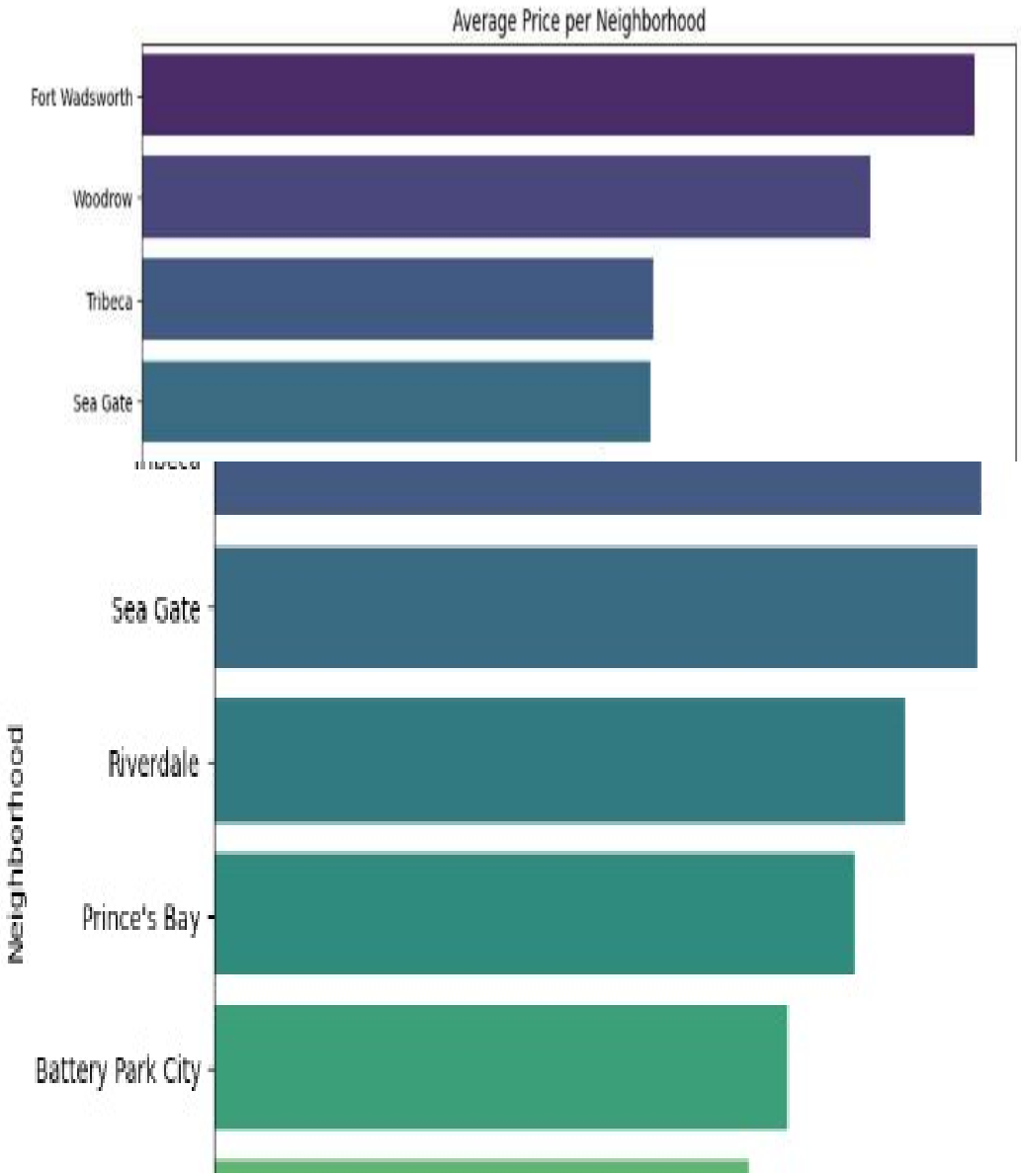
## SEASONAL TRENDS




- **Prevalent Trend for Short Stays:**
  - The chart analysis points to a predominant trend, with the majority of guests favoring shorter stays, typically lasting less than 10 days.
- **Substantial Segment Preferring Extended Stays:**
  - A significant finding is the presence of around one-fifth of guests opting for extended stays exceeding 350 days, indicating a notable preference for prolonged accommodations.
- **Limited Presence in Intermediate Stay Range:**
  - The distribution underscores a smaller proportion of guests choosing stays ranging from 10 days to 350 days. This suggests a tendency for either shorter durations or significantly longer stays.
- **Divergent Guest Stay Preferences:**
  - The overall pattern reflects divergent guest preferences, with a substantial majority favoring either brief or extensive stays, while fewer guests fall within the intermediate stay duration range.

# TRENDS AND PATTERNS

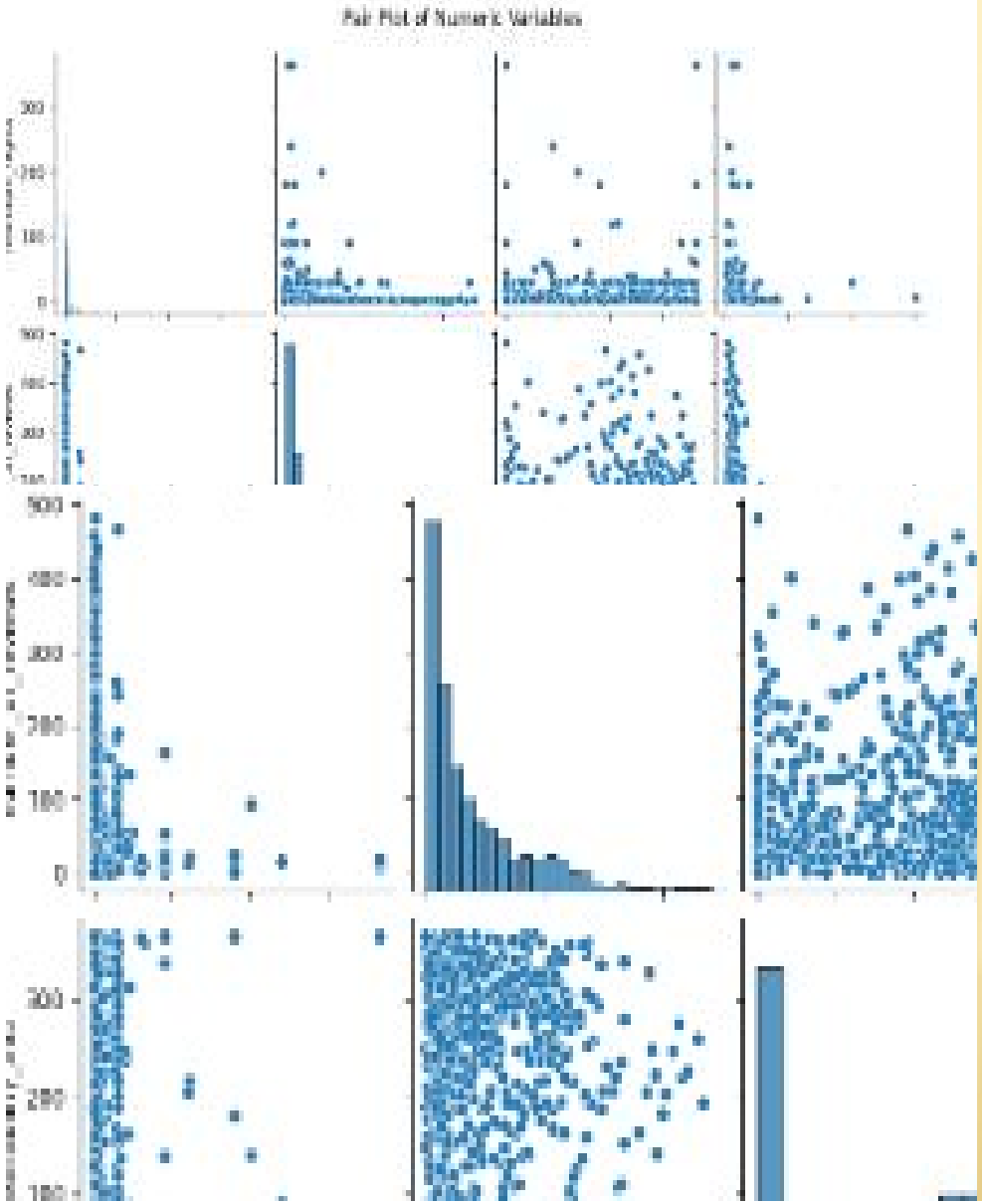
## AVERAGE PRICE PER NEIGHBORHOOD ANALYSIS




- **Exclusive Neighborhoods with High Prices:**
  - Fort Wadsworth and Woodrow in Staten Island stand out with notably high average prices of $800 and $700, respectively. This suggests a potential catering to a more high-end or exclusive market in these areas.
- **Tribeca's Upscale Reputation:**
  - Tribeca in Manhattan boasts an average price of $500, aligning with its reputation as an upscale neighborhood known for luxury accommodations. This reflects the trendy and affluent nature of Tribeca.
- **Diverse Pricing in Brooklyn:**
  - Sea Gate in Brooklyn, with an average price slightly below $500, indicates that Brooklyn features a diverse range of neighborhoods with varying price points, catering to different segments of the market.
- **Moderately Priced Neighborhood in the Bronx:**
  - Riverdale in the Bronx and Staten Island falls within the average price range of $400 to $450, positioning it as a neighborhood with moderately high pricing compared to other areas in the Bronx.
- **Manhattan's Diverse Range of Pricing:**
  - NoHo in Manhattan stands out with an average price of $295, showcasing a diverse pricing range within Manhattan itself. This diversity caters to various budgets and preferences among potential guests.

# TRENDS AND PATTERNS

## PAIR PLOT



- **Pricing Variability with Stay Duration:**
  - The pricing pattern shows variability concerning shorter minimum-night stays, with prices ranging from lower to higher values. Conversely, for longer minimum-night stays, prices consistently remain below $500.
- **Year-Round Accessibility for Short Stays:**
  - For minimum-night stays of less than 100, a diverse array of accommodations is available throughout the year. This suggests flexibility and accessibility for potential guests looking for shorter stay durations.
- **Consistent Availability Below $500:**
  - Accommodations priced below $500 are consistently available year-round, ensuring a continuous accessibility for guests with varying budget considerations.
- **Guest Engagement and Satisfaction for Short Stays:**
  - Notably, accommodations with a minimum-night requirement of less than 50 consistently receive a higher number of reviews. This trend indicates increased guest engagement and satisfaction, particularly for shorter stay durations.
- **Positive Guest Experience Trend:**
  - The consistent upward trend in the number of reviews for shorter minimum-night stays implies a sustained positive guest experience. This aligns with the availability of diverse options, creating a favorable environment for guests looking for shorter-term accommodations.

# KEY INSIGHTS

- 1. Diversify Pricing Strategies:
  - Tailor pricing strategies to the diverse range of neighborhoods and accommodation types identified in the analysis. Consider implementing dynamic pricing models to optimize rates based on demand and seasonality.
- 2. Enhance Listing Features:
  - Encourage hosts to highlight unique features and amenities in their listings, especially in neighborhoods with higher average prices. This can contribute to a competitive advantage and justify premium pricing.
- 3. Optimize Minimum-Night Policies:
  - Evaluate the impact of minimum-night requirements on booking frequency. Consider adjusting these policies to align with market trends and guest preferences, potentially attracting a wider audience.
- 4. Marketing Campaigns:
  - Leverage insights on high-value neighborhoods for targeted marketing campaigns. Highlight premium offerings and unique selling points to attract guests seeking luxury accommodations.
- 5. Host Support and Training:
  - Provide support and training programs for hosts, especially those in neighborhoods with budget-friendly accommodations. This can help hosts optimize their listings, improve guest experiences, and enhance overall satisfaction

6. Competitive Analysis:
Regularly monitor competitive landscapes in different neighborhoods. Stay informed about market trends, competitor pricing strategies, and guest preferences to adapt and stay competitive.

7. User Experience Enhancement:
Focus on improving the overall user experience for both hosts and guests on the Airbnb platform. Implement features that facilitate communication, streamline bookings, and enhance the booking process.

8. Incentivize High-Performing Hosts:
Introduce incentive programs for high-performing hosts, such as discounts on fees or priority placement in search results. Recognizing and rewarding top hosts can contribute to a positive host community and improved guest satisfaction.

9. Data-Driven Decision Making:
Encourage hosts to leverage data insights for strategic decision-making. This includes adjusting pricing based on seasonality, optimizing minimum-night requirements, and adapting to changing market dynamics.

10. Continuous Monitoring and Adaptation:
Emphasize the importance of ongoing monitoring and adaptation to market changes. Regularly revisit and adjust strategies based on updated data to ensure alignment with current trends and guest preferences.

# SUMMARY

- The analysis highlights Manhattan and Brooklyn as prime locations for Airbnb rentals, driven by their high demand and diverse offerings. Manhattan's global appeal, coupled with its tourist attractions, contributes to both demand and higher prices. Brooklyn, with its strategic proximity to Manhattan, presents a cost-effective alternative for guests.
- Key neighborhoods like Williamsburg, Bedford-Stuyvesant, Harlem, Bushwick, and the Upper West Side exhibit strong demand. While Manhattan offers lucrative returns, it also faces intense competition, indicating a need for strategic investment.
- The data underscores the preference for short-term stays, with limited demand for stays exceeding 30 nights. Hosts are advised to tailor property investments to accommodate shorter stays for optimal occupancy.
- Entire homes or private rooms dominate listings, providing a range of choices for travelers. However, competition is fierce, urging hosts to explore neighborhoods with fewer listings for differentiation.
- Variations in availability across neighborhoods suggest opportunities for hosts to identify less saturated markets. The dominance of a few hosts in the market underscores the need for differentiation and strategic property investment.
- Queens neighborhoods near the airport stand out for potentially higher reviews due to their convenience for travelers. While Manhattan and Brooklyn offer a diverse range of options, competition is high, making strategic investments essential for success.

# FUTURE WORK

- Dynamic Pricing Strategies:
  - Implementing dynamic pricing strategies based on seasonal demand, local events, and other factors to optimize revenue and attract more guests.
- Host Training and Support Programs:
  - Developing training and support programs for hosts, focusing on enhancing guest experiences, communication skills, and effective marketing to improve overall performance.
- User Experience Enhancement:
  - Investigating and implementing features to enhance the user experience on the Airbnb platform, potentially leading to increased engagement and satisfaction.
- Market Expansion Opportunities:
  - Exploring opportunities for market expansion in neighborhoods or regions with untapped potential, informed by the analysis of pricing trends and demand dynamics.
- Integration of External Data Sources:
  - Integrating external data sources such as local events, public transportation accessibility, and safety metrics to provide more comprehensive insights for both hosts and guests.

# THANK YOU