EDA ON AIRBNB BOOKING

Abstract:

Airbnb is a global online marketplace that connects travellers seeking accommodations with hosts offering unique stays or experiences. Founded in 2008, the platform operates as a peer-to-peer network, enabling individuals to rent out their spaces or provide tourism experiences. Its business model is built on fostering trust through user reviews, secure payments, and verification systems. By leveraging the principles of the sharing economy, Airbnb has disrupted traditional hospitality, providing cost-effective and personalized alternatives to hotels. The platform's innovative approach has redefined travel, though it faces challenges such as regulatory compliance, market saturation, and competition. Despite these hurdles, Airbnb continues to expand, offering innovative features and partnerships to enhance the user experience.

Problem Statement:

Since its launch in 2008, Airbnb has transformed the travel industry by offering unique and personalized lodging experiences that cater to diverse traveller needs. Central to Airbnb's success is its ability to leverage data insights drawn from millions of listings worldwide. These insights play a pivotal role in shaping the platform's operational strategies and enhancing user satisfaction. For instance, the analysis of a dataset comprising approximately 49,000 observations provides valuable information about host behaviour, customer preferences, and emerging market trends. By interpreting this data, Airbnb gains a deeper understanding of the factors that drive user engagement, enabling the company to make informed strategic decisions and remain competitive in the ever-evolving hospitality sector. This data-driven approach has not only optimized Airbnb's platform but has also contributed to redefining modern travel experiences.

Dataset Overview:

The dataset contains 48,895 records and 16 columns, covering various details about Airbnb listings in NYC.

Include column descriptions:

- **id:** Unique listing id.
- **name:** Name of the property.

- **host_id:** Unique identifier for each listed host.
- **host_name:** Name of the host.
- **neighbourhood_group:** Location
- **neighbourhood:** Area
- **latitude:** Latitude coordinates
- **longitude:** Longitude coordinates
- **room_type:** Type of room being rented (e.g., Entire home/apt, Private room).
- **price:** Price per night for renting the property.
- **minimum_nights:** Minimum number of nights required for a booking or stay
- **number_of_reviews:** Number of reviews written for the listing
- **last_review:** Date of the most recent review.
- **reviews_per_month:** Average number of reviews per month.
- **calculated_host_listings_count:** Total no of listings against the host id
- **availability_365:** Number of days when listing is available for booking.

The goal of this project is to perform Exploratory Data Analysis (EDA) on the Airbnb NYC 2019 dataset. EDA involves examining the dataset to find useful insights and patterns. By analysing this data, we aim to discover how different factors affect property prices, understand which neighbourhoods are more popular, and identify trends that can be useful for Airbnb hosts and potential guests.

Introduction:

Airbnb is a global online marketplace that connects travellers seeking accommodations with hosts offering unique stays or experiences. Founded in 2008, the platform operates as a peer-to-peer network, enabling individuals to rent out their spaces or provide tourism experiences. Its business model is built on fostering trust through user reviews, secure payments, and verification systems. By leveraging the principles of the sharing economy, Airbnb has disrupted traditional hospitality, providing cost-effective and personalized alternatives to hotels. The platform's innovative approach has redefined travel, though it faces challenges such as regulatory compliance, market saturation, and competition. Despite these hurdles, Airbnb continues to expand, offering innovative features and partnerships to enhance the user experience.

Reason for Analysis:

Analysing Airbnb's dataset serves several important purposes that contribute to the company's growth, operational efficiency, and user satisfaction. These include:

- **Understanding Host Behaviour:** Insights into host practices, such as pricing strategies, response times, and listing quality, allow Airbnb to provide better guidance to hosts and maintain high standards across the platform.
- **Identifying Customer Preferences:** Analysing data on customer bookings, preferences, and reviews helps uncover trends in traveller needs, such as preferred locations, amenities, or pricing levels. This enables Airbnb to tailor its recommendations and improve user experience.
- **Tracking Market Trends:** The data reveals patterns in demand, such as seasonality, peak travel periods, and popular destinations. This information is crucial for forecasting and adapting to changes in the market.
- **Enhancing Decision-Making:** Data-driven insights empower Airbnb to make informed strategic decisions, from optimizing algorithms for search and recommendations to launching new features and services.
- **Improving Competitiveness:** In a highly competitive hospitality industry, data analysis helps Airbnb identify opportunities to stay ahead, such as addressing underserved markets or improving customer retention strategies.

Tools and Techniques used:

Tools: - PowerBI:

- Power BI is a business intelligence tool used to transform raw data into interactive dashboards and meaningful insights.
- It allows users to connect to multiple data sources, clean data, create reports, and share them easily.
- Power BI provides a user-friendly interface and supports real-time updates for effective data analysis and decision-making.

Data Cleaning:

1. **Handling Missing values:**

The dataset was loaded from excel into powerBI.

a) **name column:**
   - The **name** column contained null values, indicating that names were not provided for certain records.
   - To address this, the missing values were replaced with the placeholder value "**Unknown**".
   - This ensured consistency in the dataset and allowed for accurate analysis without losing any records due to missing data.

b) **host_name column:**
   - The **host_name** column contained null values, indicating that names were not provided for certain records.
   - To address this, the missing values were replaced with the placeholder value "**Unknown**".
   - This ensured consistency in the dataset and allowed for accurate analysis without losing any records due to missing data.

c) **reviews_per_month:**
   - The **reviews_per_month** column contained null values.
   - Upon analysis, it was observed that the corresponding values in the **number_of _reviews** column were **0** for these null entries.
   - Based on this observation, it was concluded that filling the null values in the reviews_per_month column with **0** was the most appropriate approach.
   - This ensures the data remains consistent and accurately represents the lack of reviews for those entries.

2. Removed Columns:

The following columns were removed from the dataset as they were deemed irrelevant for the analysis:

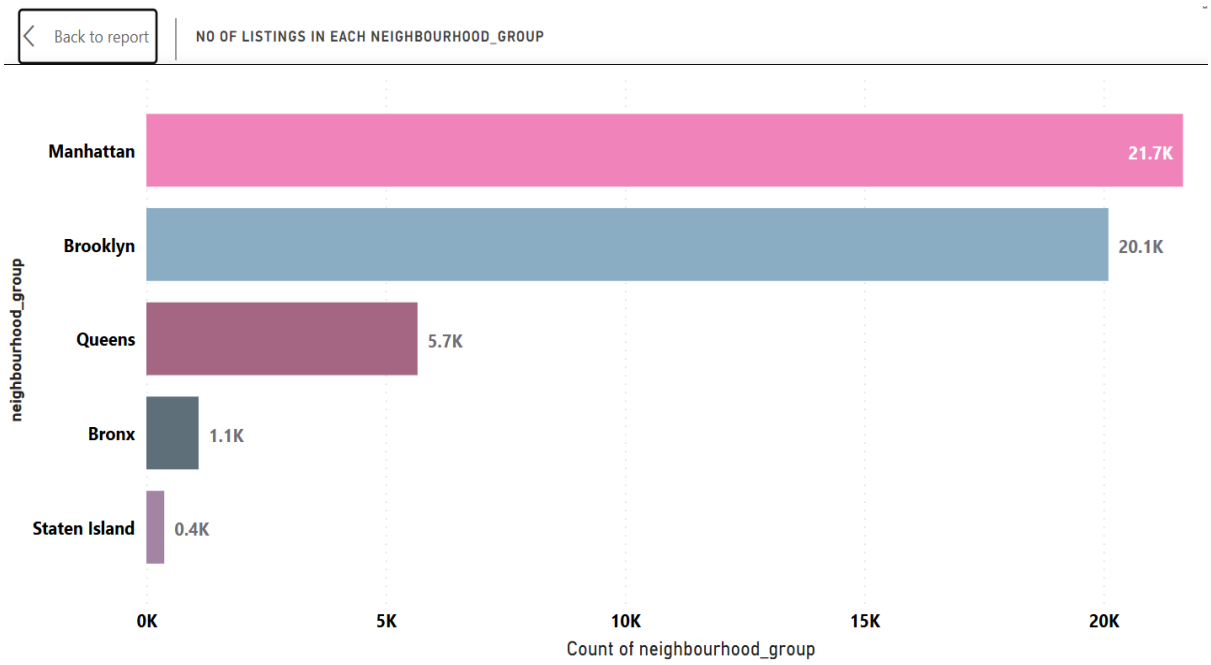
a) latitude.
b) longitude.
c) last_review.
d) availability_365.

These columns were not directly contributing to the objectives of the analysis and were excluded to simplify the dataset and focus on the relevant variables.

3. **Added Columns:**
    - A new column named **Revenue** was added to the dataset.
    - The Revenue column was calculated based on the values in the Price and Minimum Nights columns using the formula:
    - Revenue = Price × Minimum Nights.
    - This addition helps in analysing the potential earnings for each listing and provides a more comprehensive understanding of the dataset.
    - And also added another column **price_category.**
    - **Low**: Price < ₹60, indicating budget-friendly listings.
    - **Medium**: Price between ₹60 and ₹400, offering moderate pricing.
    - **High**: Price > ₹400, for premium listings.

Data Visualization:

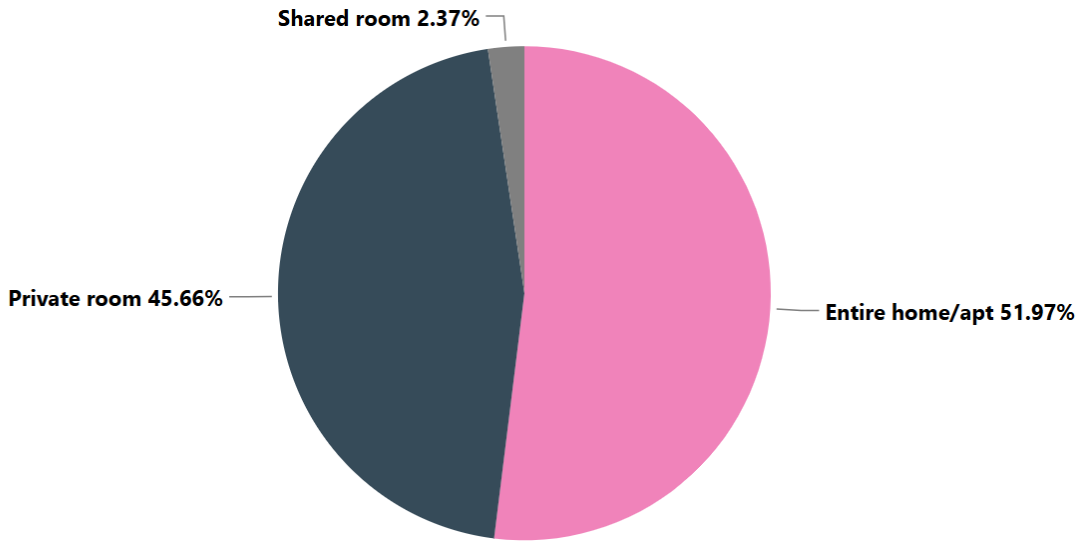
Number of Listings in Each Neighbourhood Group:



- **Manhattan and Brooklyn**: These two neighbourhoods dominate in terms of listing counts, indicating their popularity as Airbnb destinations. These areas likely offer attractions, amenities, and a higher demand for short-term rentals.

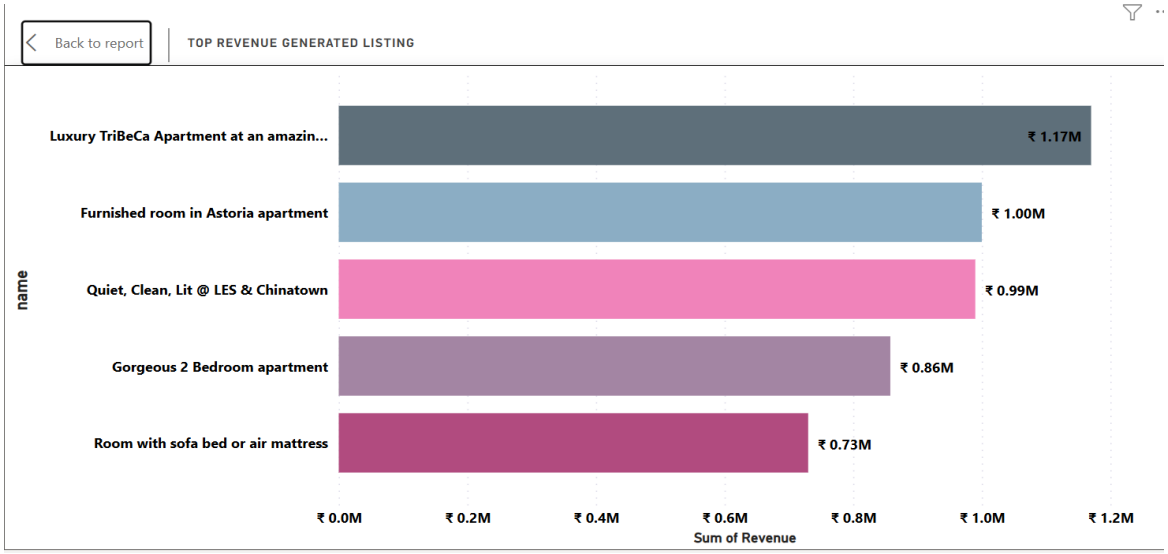
- **Staten Island**: With the least number of listings, it appears less attractive to hosts, possibly due to fewer tourist attractions or limited demand for short-term rentals.

Room Type Distribution:



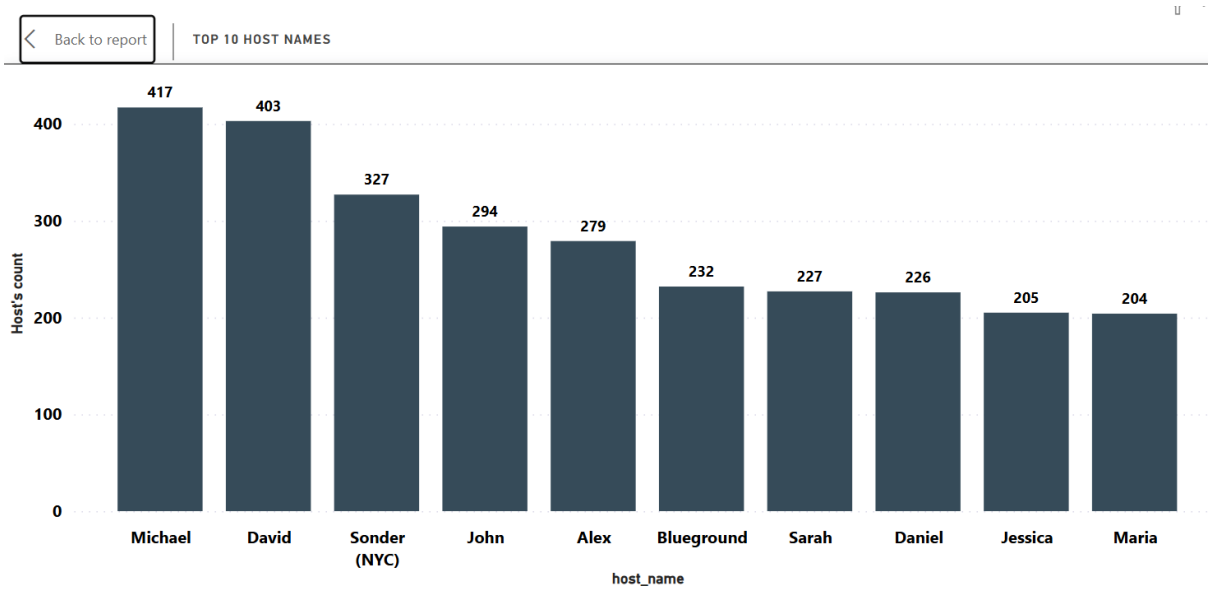
- **Entire Home/Apt: 51.97%** of listings, indicating that hosts prefer to offer entire properties as rentals. This could be due to the higher potential earnings from renting out full units.
- **Private Room: 45.66%** of listings, showing that a significant portion of hosts choose to rent private rooms within their properties.
- **Shared Room: 2.37%** of listings, a much smaller share, likely due to the reduced appeal of shared accommodations compared to private or entire units.

Top Revenue-Generating Listings:



- **Luxury properties** such as the "Luxury Triplex Penthouse" and "Furnished Private Room" generate the highest revenue, emphasizing that high-end accommodations attract a premium price. These listings contribute significantly to the total revenue and highlight the demand for upscale, unique Airbnb offerings.

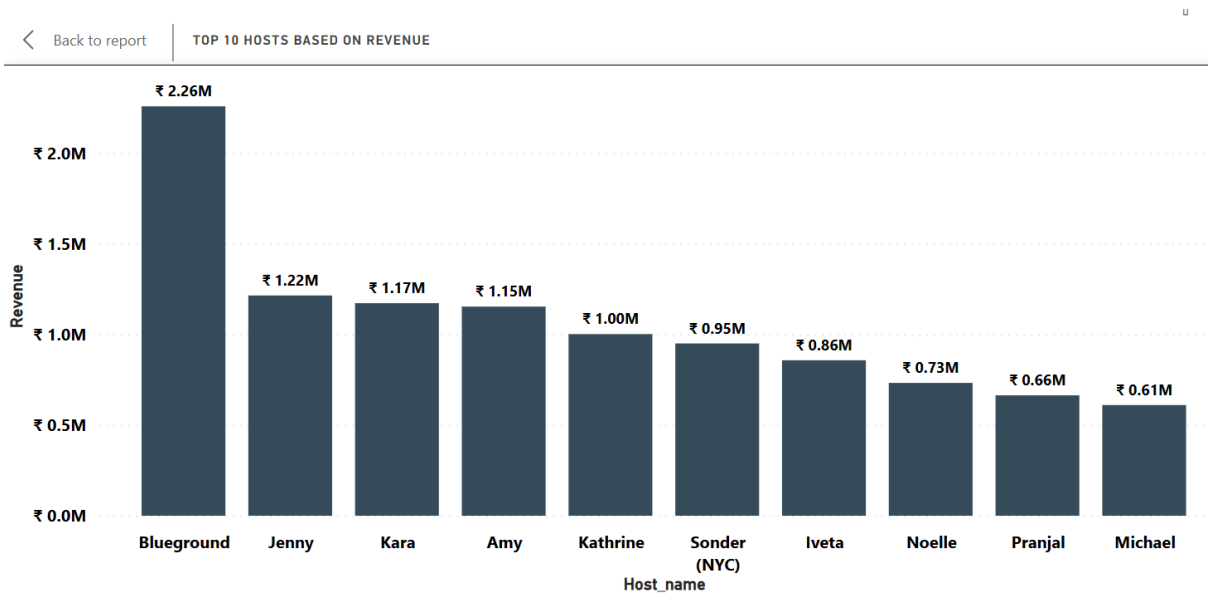
Top 10 Host Names by Count:



- **Michael** has the highest number of listings, followed by **David** and **Sonder**. This shows that these hosts have a significant presence in the NYC Airbnb market, either through

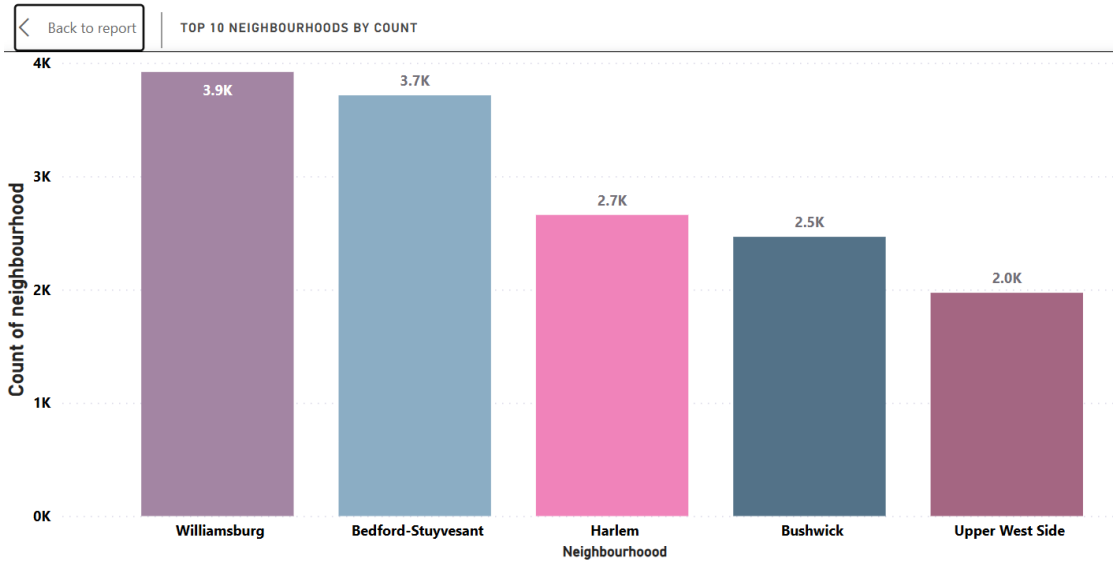
owning multiple properties or managing numerous listings. Their dominance suggests a well-established hosting operation and could reflect their market strategy.

Top 10 Hosts Based on Revenue:



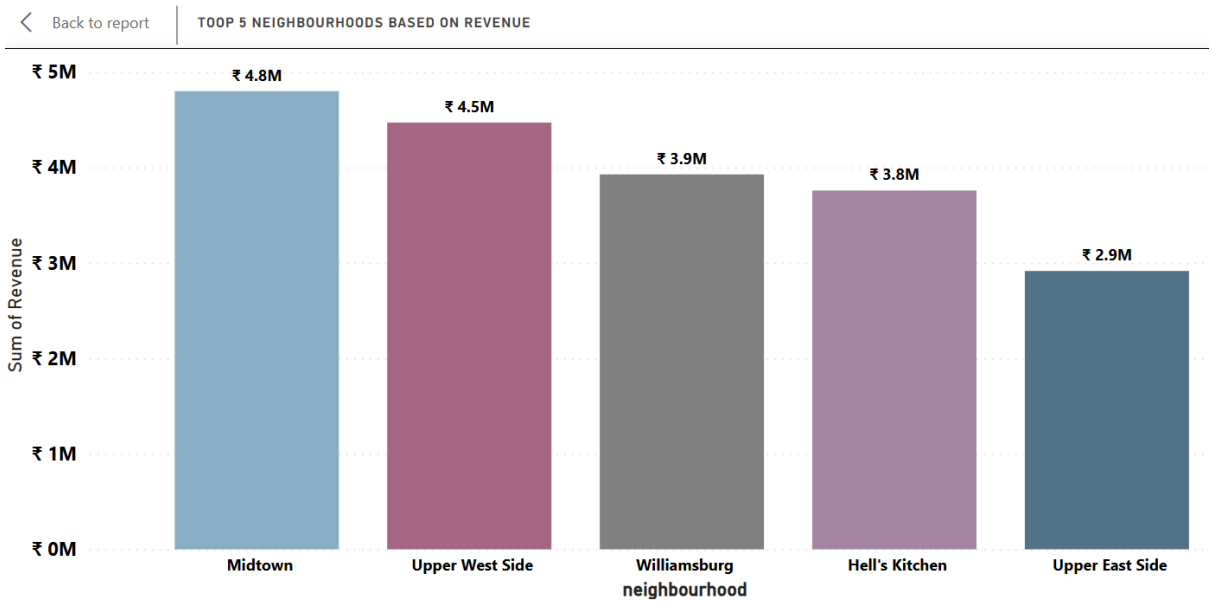
- **Blueground -** is the top revenue-generating host with ₹2.3M, followed by **Jenny** and **Kara**. The revenue concentration on professional hosts like Blueground highlights the importance of managing high-value properties, which yield higher earnings, possibly due to premium pricing or higher occupancy rates. This insight suggests that hosts who focus on quality and pricing strategy can see significant returns.

Top 10 Neighbourhoods by Count:



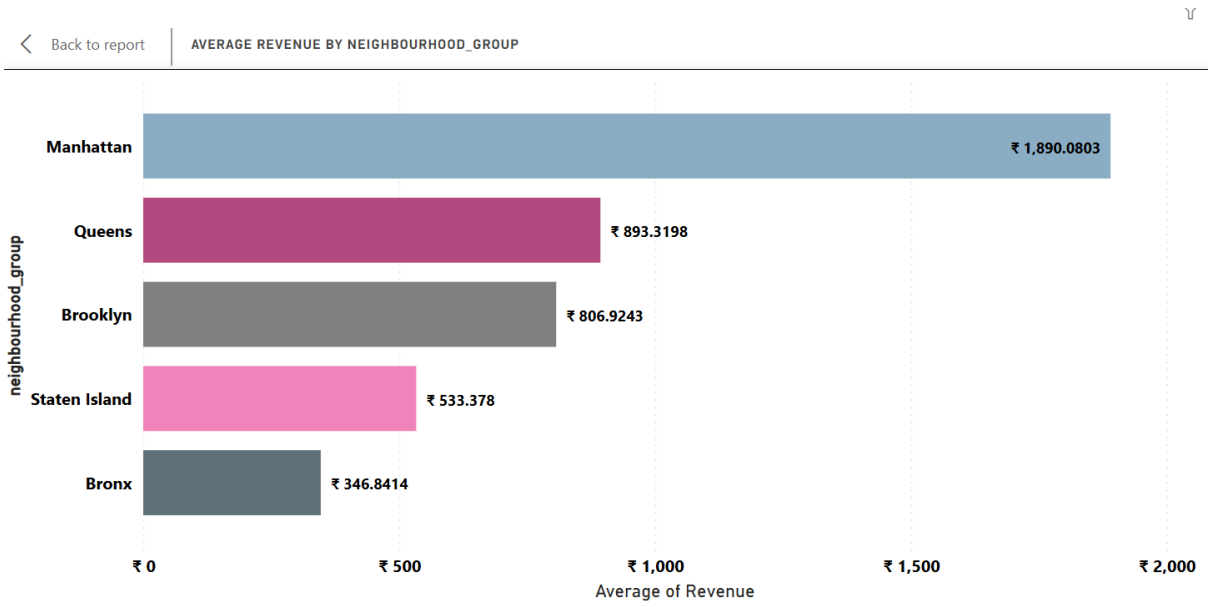
- **Williamsburg** and **Bedford-Stuyvesant** lead in listings, indicating their popularity as neighbourhood choices for Airbnb hosts. These neighbourhoods likely have a high demand for short-term rentals, possibly due to their proximity to popular tourist spots or vibrant local culture.
- **Neighbourhood popularity** suggests that hosts in these areas benefit from a steady stream of visitors.

Top 5 Neighbourhoods Based on Revenue:



- **Midtown** generates the highest revenue, followed by **Williamsburg** and **Harlem**. This indicates that properties in Midtown command higher prices and occupancy rates, making it a lucrative location for hosts. Premium pricing and the area's central location likely contribute to the high revenue generation.
- Hosts in these neighbourhoods should focus on maintaining quality listings to maximize earnings.

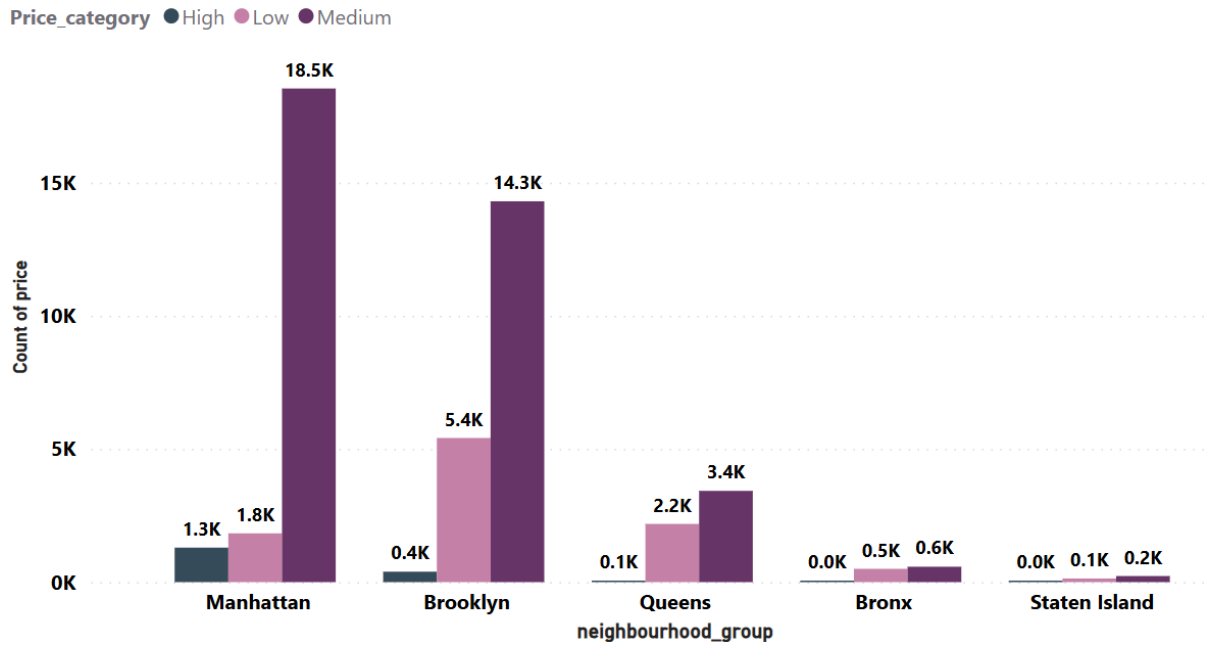
Average Revenue by Neighbourhood Group:



- **Manhattan** listings yield the highest average revenue per listing, showing that this area is the most lucrative for hosts. The area's demand for high-end listings and tourism likely contributes to this trend.

- **Staten Island**, with the lowest average revenue, suggests that it may not attract the same level of demand or premium pricing.

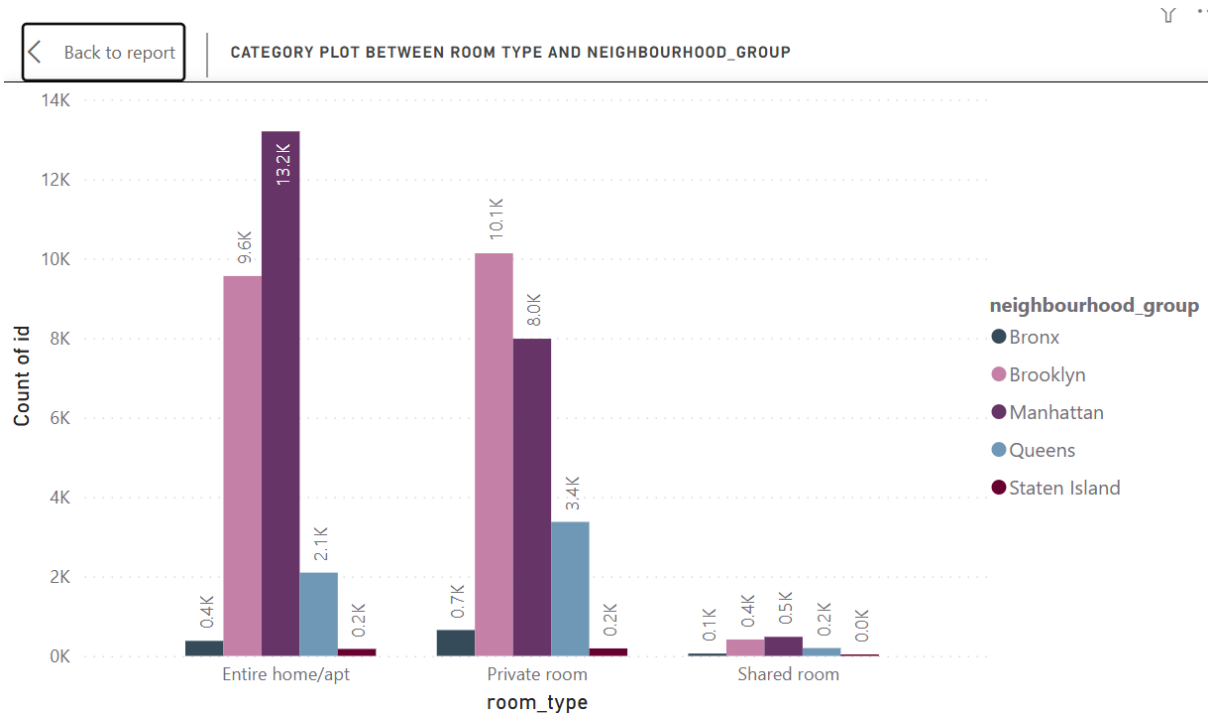
Count of Price by Neighbourhood Group and Price Category:



- **Manhattan and Brooklyn** dominate in all price categories (Low, Medium, High), with most listings falling under the Medium price category. This shows that while hosts offer a variety of price points, the bulk of listings are positioned in the mid-range, possibly catering to a wide demographic of tourists.

- This trend can be useful for understanding pricing strategies in different neighbourhoods.

Category Plot Between Room Type and Neighbourhood Group:



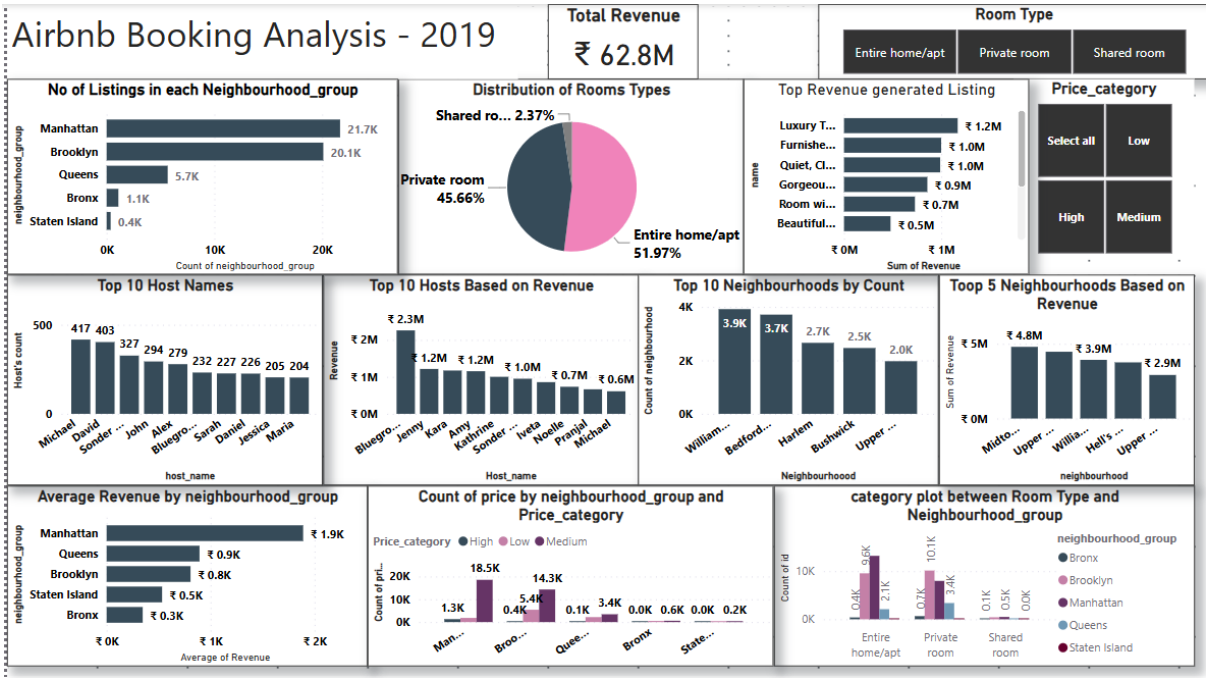
Entire Home/Apt dominates across all neighbourhoods, especially in **Manhattan and Brooklyn**, where the demand for full-property rentals is high. This suggests that hosts in these neighbourhoods focus on providing more premium, independent accommodations rather than shared spaces, which aligns with trends in higher-end travel and tourism.

Total Revenue: ₹62.8M

- **Insight**: The total revenue of ₹62.8M signifies the overall earnings generated by Airbnb hosts in NYC. This figure acts as a key performance indicator (KPI) that reflects the

health of the Airbnb market. By tracking this value, stakeholders can monitor growth, set targets, and evaluate the effectiveness of pricing strategies.

Dashboard:



Key Insights from the Dashboard:

1. **Neighbourhood Insights**:

    - **Manhattan** and **Brooklyn** are the most active neighborhoods in terms of both listings and revenue, making them prime targets for hosts looking to maximize earnings.
    - **Staten Island** has the least activity, suggesting potential for growth if hosts can find ways to increase demand or offer attractive listings in the area.

2. **Room Type Trends**:

- **Entire Home/Apt** is the preferred room type, with hosts focusing on providing independent properties. This trend is particularly strong in **Manhattan and Brooklyn**.

- **Private Rooms** also hold a significant portion of the market, but **Shared Rooms** remain a niche offering.

3. **Revenue Insights**:

    - High revenue is concentrated in premium neighborhoods like **Midtown** and **Williamsburg**, reinforcing the idea that premium pricing and high occupancy are key to maximizing Airbnb earnings.
    - Hosts such as **Blueground** dominate revenue generation, suggesting that professional hosts focusing on high-value properties are more likely to succeed.

4. **Host Insights**:

    - **Michael**, **David**, and **Blueground** are key players, demonstrating that hosts with a significant number of listings or revenue generation dominate the Airbnb market.
    - Professional hosts like **Blueground** focus on high-value properties, leading to higher earnings and sustained success.

5. **Price and Revenue Correlation**:

    - **Medium price category** listings dominate in all neighborhoods, but **High price category** listings contribute significantly to total revenue, even though they represent fewer listings. This suggests that premium listings yield high returns.

Conclusion:

- The analysis highlights key opportunities for Airbnb hosts, including focusing on **premium properties** in **Manhattan and Brooklyn** and leveraging the high demand for **Entire Home/Apt** rentals. Hosts who offer higher-priced, well-managed listings can significantly boost their revenue.
- The dashboard serves as an essential tool for understanding market trends, guiding decisions related to **pricing**, **room types**, and **location strategies**. By analyzing these insights, stakeholders can make informed, data-driven decisions to improve performance and profitability in the Airbnb market.