Airbnb Booking Analysis Project Report

**Project Name**

Airbnb Booking Analysis

**Project Type**

Exploratory Data Analysis (EDA)

**Team Contribution**

* **Team Member 1:** Bhushan Khapre (08)
* **Team Member 2:** Yogesh Kuthe (63)

**Introduction**

The **Airbnb Booking Analysis** project explores Airbnb listing data for New York City to uncover trends and actionable insights. By analyzing factors like pricing, neighborhood patterns, and guest reviews, the project aims to help hosts optimize pricing, improve occupancy rates, and enhance customer satisfaction. Through visualizations and data-driven insights, it supports informed decision-making for better market alignment and growth.

**Problem Statement**

Since 2008, guests and hosts have used Airbnb to expand on traveling possibilities and present a more unique, personalized way of experiencing the world. Today, Airbnb became one of a kind service that is used and recognized by the whole world. Data analysis on millions of listings provided through Airbnb is a crucial factor for the company. These millions of listings generate a lot of data - data that can be analyzed and used for security, business decisions, understanding of customers' and providers' (hosts) behavior and performance on the platform, guiding marketing initiatives, implementation of innovative additional services and much more.

This dataset has around 49,000 observations in it with 16 columns and it is a mix between categorical and numeric values.

**Objectives**

1. Understand the dataset and clean it for analysis.
2. Explore key features like price, room types, availability, and reviews.
3. Identify patterns and trends to optimize host and guest experiences.
4. Provide actionable recommendations for pricing, marketing, and operational improvements.

**Flow Diagram**

Know Your Data

Data Wrangling

**Price Trends**

Data Visualization

**Demand for Room Type**

**Availability**

Key Findings

**Reviews**

Recommendations

**Neighborhood Insights**

Conclusion

**Steps in the Analysis**

**1. Know Your Data**

* **Dataset Variable Description:** The dataset includes columns such as

• id : Unique ID

• name : Name of the listing

• host\_id : Unique host\_id

• host\_name : Name of the host

• neighbourhood\_group : location

• neighborhood : area

• latitude : Latitude range

• longitude : Longitude range

• room\_type : Type of listing

• price : Price of listing

• minimum\_nights : Minimum nights to be paid for

• number\_of\_reviews : Number of reviews

• last\_review : Content of the last review

• reviews\_per\_month : Number of checks per month

• calculated\_host\_listing\_count : Total count

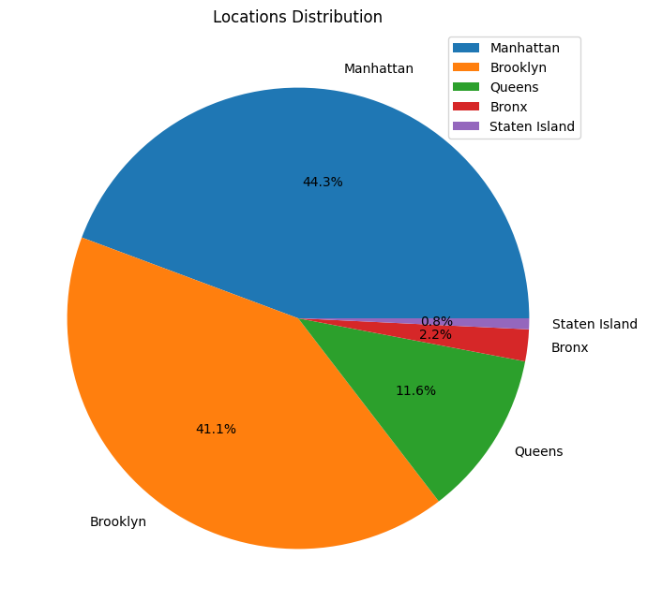
• availability\_365 : Availability around the year

**2. Data Wrangling**

* **Data Cleaning:** Removed unwanted columns like latitude and longitude. Deal with missing values using heatmaps and data wrangling techniques.
* **Unique Values Analysis:** Explored the unique categories for each feature to understand their diversity.
* **Duplicate Removal:** Identified and eliminated duplicate entries.
* **Missing Values:** Addressed null values for critical columns such as reviews\_per\_month using appropriate imputation techniques.
* **Feature Transformation:** Converted categorical variables into formats suitable for analysis.

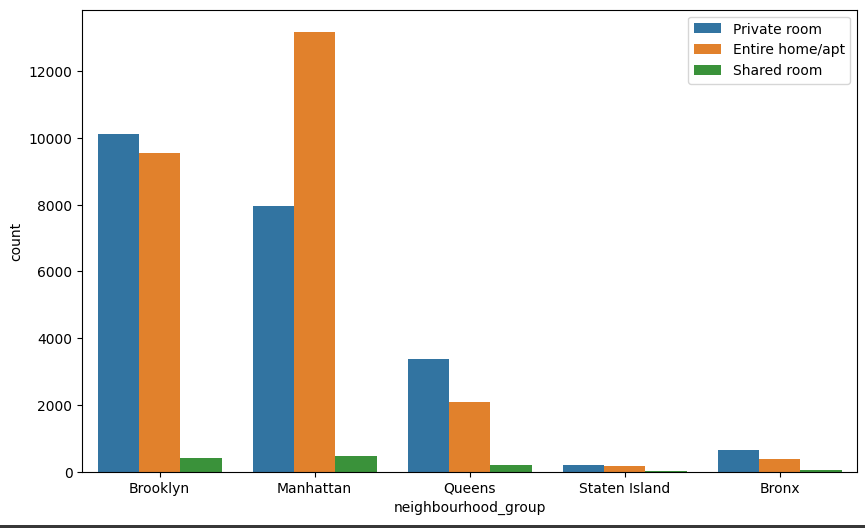
**3. Data Visualization & Insights**

**Chart 1:**Pie Chart on Dependant Variable i.e., neighbourhood\_group (Univariate)



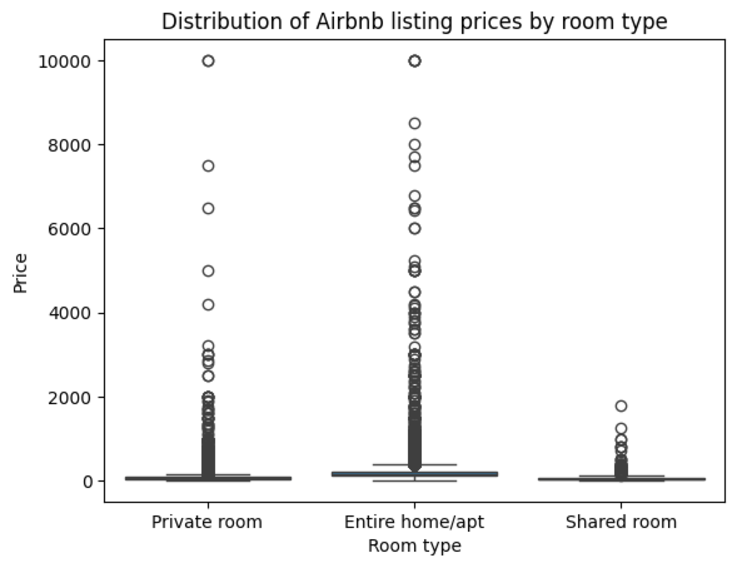
* Insight: Manhattan has the highest number of listings, reflecting its popularity among tourists.

**Chart 2:** Distribution of Airbnb listings in different neighborhoods as room types (Bivariate)



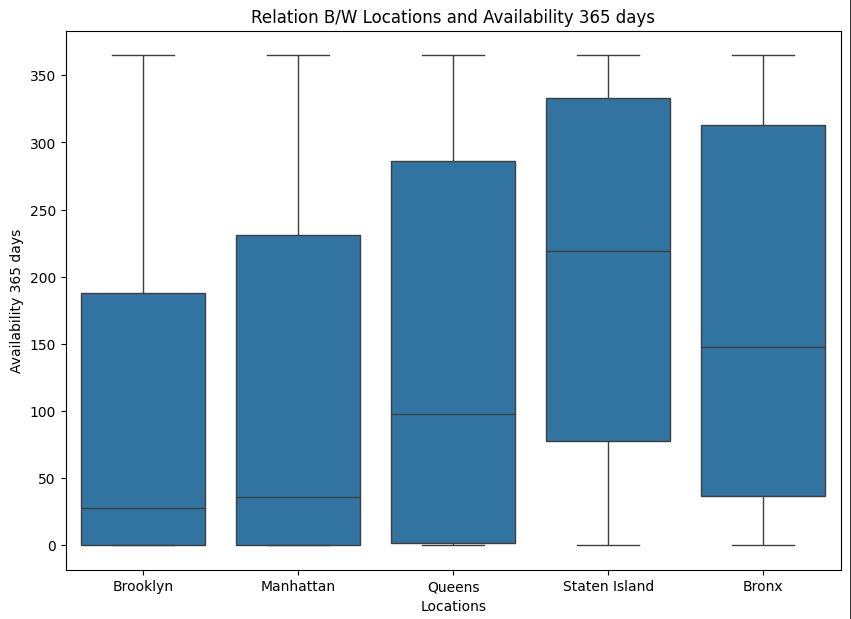
* Insight: Entire homes/apartments dominate Manhattan, while private rooms are more common in Brooklyn.

**Chart 3:** Distribution of Airbnb listing prices by room type (Bivariate)



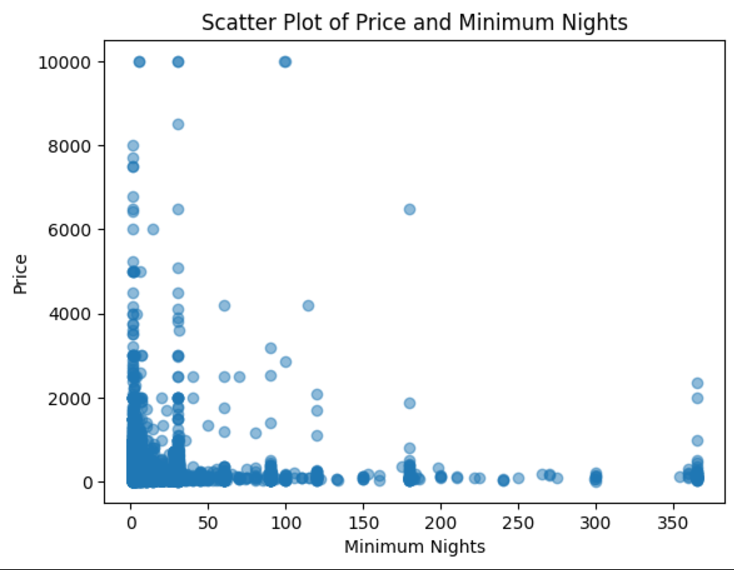
* Insight: Entire homes/apartments are the most expensive, whereas shared rooms are the cheapest.

**Chart 4:** Relation B/W Locations and Availability 365 days (Bivariate)



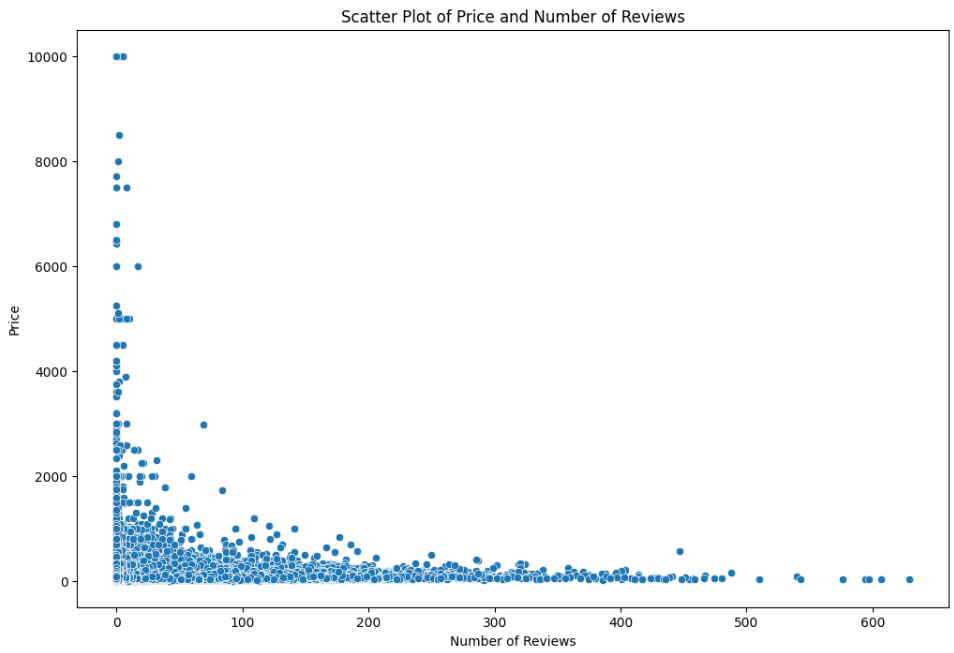
* Insight: Listings in Manhattan and Brooklyn show higher year-round availability.

**Chart 5:** Scatter Plot of Price and Minimum Nights (Bivariate)



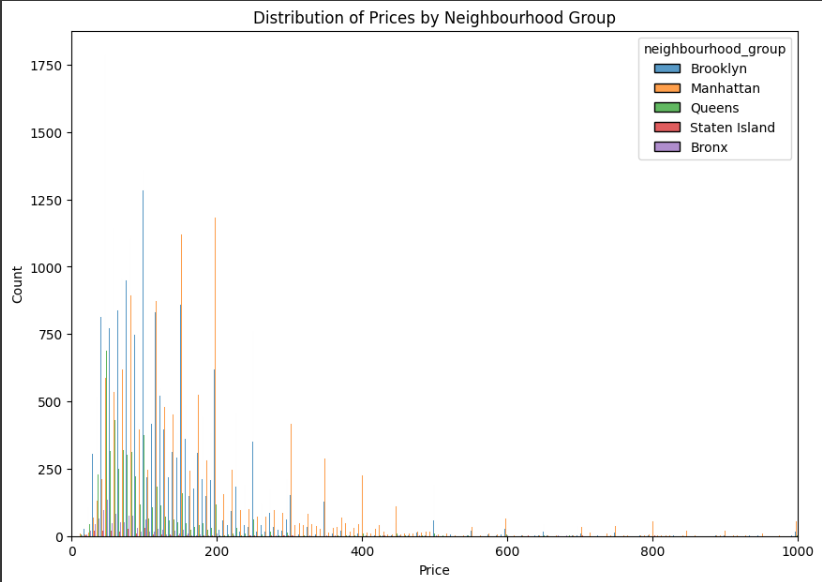
* Insight: Outliers exist with very high prices or extended minimum night stays, which could skew average pricing data.

**Chart 6:** Scatter Plot of Price and Number of Reviews (Bivariate)



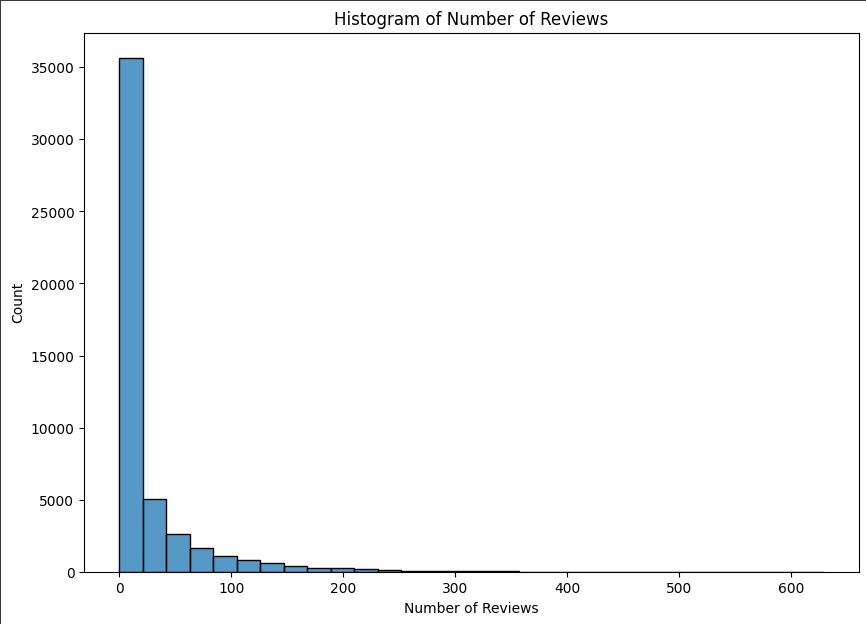
Insight: Listings with moderate pricing tend to receive more reviews, indicating a sweet spot for affordability and value.

**Chart 7:** Distribution of Prices by Neighbourhood Group (Bivariate)



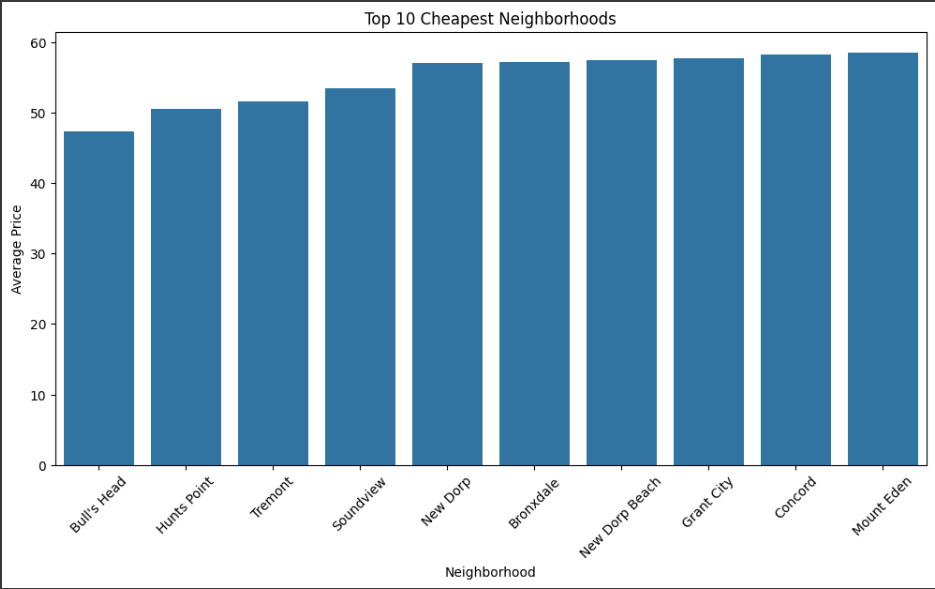
* Insight: Manhattan listings have the highest average prices, while Staten Island is the most affordable.

**Chart 8:** Histogram of Number of Reviews (Univariate)



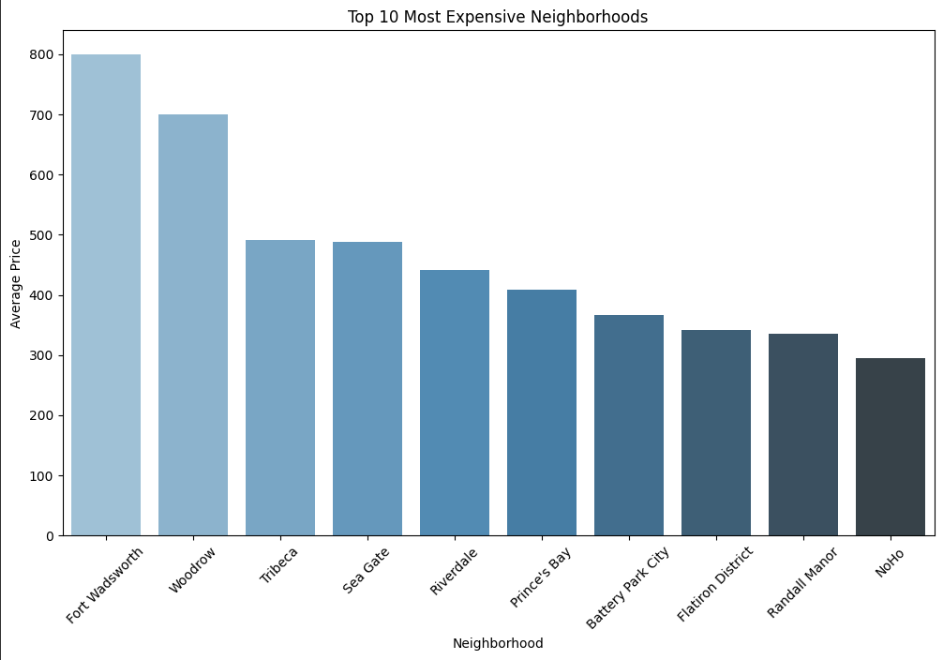
* Insight: A majority of listings have fewer than 50 reviews, highlighting the need for hosts to actively encourage guest feedback.

**Chart 9:** Top 10 Cheapest Neighborhoods (Bivariate)



* Insight: Neighborhoods like Bay Ridge and Flatlands in Brooklyn offer affordable options.

**Chart 10:** Top 10 Most Expensive Neighborhoods'(Bivariate)



* Insight: Tribeca and SoHo in Manhattan lead with the highest average prices.

**Key Findings**

**1. Price Trends:** While hosts in Manhattan might charge high prices, other areas are more popular due to affordability.   
**2. Demand for Room Type:** While individual rooms serve tourists on a tight budget, entire residences or apartments predominate in upscale districts.   
**3. Availability:** Consistently available properties draw more reviews and prospective visitors.   
**4. Reviews:** Listings with fair prices typically receive more reviews, which increases their appeal.   
**5. Neighborhood Insights:** By utilizing low pricing and positive evaluations, hosts in less well-known neighborhoods can increase their visibility.

**Recommendations**

1. Optimize pricing strategy by adjusting prices to match market demand and seasonality trends. Hosts can also take into consideration the pricing distribution by neighborhood group and adjust their prices accordingly.
2. Target marketing efforts on the top 10 cheapest and most expensive neighborhoods to increase visibility and attract potential renters. Hosts and rental agencies can also use the density plot of price distribution by neighborhood group to identify pricing trends in different areas.
3. Address areas of improvement raised in reviews and ratings to improve customer satisfaction and increase positive word-of-mouth referrals. Hosts can also leverage the distribution of the number of reviews and ratings to assess their performance and identify areas for improvement.
4. Incentivize experienced hosts to continue using the platform through special offers and promotions.
5. Analyze the distribution of the number of reviews and ratings to identify popular listings and areas for growth.

**Conclusion**

1. Pricing strategy is an important factor for hosts and rental agencies to consider. By analyzing pricing trends by neighborhood group and over time, hosts can optimize their pricing to match market demand and seasonality.
2. Marketing efforts should be targeted towards the top 10 cheapest and most expensive neighborhoods to increase visibility and attract potential renters.
3. Customer satisfaction is key to driving positive word-of-mouth referrals and repeat business. Hosts should address areas of improvement raised in reviews and ratings to improve customer satisfaction.
4. Experienced hosts can be incentivized to continue using the platform through special offers and promotions.
5. The number of reviews and ratings is a good indicator of listing popularity and areas for growth. Airbnb can use this data to identify popular listings and areas for expansion.

Overall, leveraging data insights from Airbnb's platform can help hosts and rental agencies improve their business strategies and drive growth.