A blue and white logo

Description automatically generated

**The Final Project Report**

Submitted in partial fulfillment of the requirement for project deliverable.

**PYTHON PROGRAMMING**

**SPRING SEMESTER 2024**

**Submitted by Group - 7**

**BOMMINENI ALEKHYA**

**YASHWANTH THALLA**

**MAMIDALA RUTHWIK**

**Supervised by**

**Professor Kshitij Sharma**

**PACE UNIVERSITY**

**SEIDENBERG SCHOOL OF COMPUTER SCIENCE AND INFORMATION**

**SYSTEMS**

**Abstract**

This report presents an in-depth exploratory data analysis (EDA) of the New York City Airbnb dataset, aiming to uncover insights into the short-term rental market dynamics in NYC. By leveraging data visualization and statistical analysis, the report elucidates trends in pricing, demand, and host activity across different boroughs and types of accommodations.

**Introduction**

The advent of Airbnb has revolutionized the lodging industry, offering an alternative to traditional hotel accommodations. New York City, as a bustling metropolis attracting millions of tourists annually, presents a unique market for short-term rentals. This report analyses the NYC Airbnb open dataset to understand the market's characteristics, including pricing strategies, demand fluctuations, and the impact of location and listing types.

**Dataset**

The dataset comprises Airbnb listings in New York City, detailing various aspects such as location, room type, price, minimum nights, number of reviews, and availability. With nearly 49,000 entries, it offers a comprehensive overview of the short-term rental landscape in NYC.

* **Listing ID and Name:** Unique identifiers and names of the listings.
* **Host ID and Name:** Information about the hosts.
* **Neighbourhood Group and Neighbourhood:** Geographical information providing insights into the distribution of listings.
* **Latitude and Longitude:** Precise location data for mapping and spatial analysis.
* **Room Type:** Type of listing (e.g., Entire home/apt, Private room, Shared room).
* **Price:** Nightly price, which can be analysed for pricing strategies and affordability.
* **Minimum Nights:** The minimum length of stay required, useful for understanding the target market of listings.
* **Number of Reviews:** Can indicate popularity or quality.
* **Last Review Date:** The date of the latest review, useful for temporal analyses.
* **Reviews per Month:** Activity indicator that could suggest how often a listing is booked.
* **Availability 365:** Number of days a year the listing is available, providing insight into how listings are managed.

Data Set Source: [New York City Airbnb Open Data (kaggle.com)](https://www.kaggle.com/datasets/dgomonov/new-york-city-airbnb-open-data)

**Preprocessing**

The dataset underwent several preprocessing steps, including handling missing values, removing duplicates, and ensuring data integrity. Missing values in textual fields were filled with placeholders, whereas numerical missing values were addressed appropriately to maintain analytical accuracy.

**Data Cleaning**

A screenshot of a computer program

Description automatically generated

A screen shot of a computer program

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer program

Description automatically generated

**EDA:**

**Descriptive Statistics**

A screenshot of a computer

Description automatically generated

**Univariate Analysis**

A computer screen shot of a program code

Description automatically generated

A graph with a green bar

Description automatically generated

**The histograms provide valuable insights into the distributions of key variables:**

**Prices:** The distribution of listing prices is highly skewed to the right, with the majority of listings priced under $250 per night. There are few listings with very high prices, which may represent luxury accommodations or potential outliers.

**Minimum Nights:** The distribution of minimum nights required is also right skewed, with most listings requiring fewer than 5 nights stay, indicating a market geared towards short-term rentals. Listings requiring more than 30 nights are less common and might be targeted towards long-term stays or could be outliers.

**Number of Reviews:** The number of reviews per listing is right-skewed, with a significant number of listings having fewer than 20 reviews. This suggests that while some listings are highly popular, many others have limited feedback.

These observations highlight the diversity in the New York City Airbnb market, with a wide range of prices, minimum stay requirements, and varying levels of guest engagement.

**Bivariate and Multivariate Analysis**

A screen shot of a computer program

Description automatically generated

A group of graphs with different sizes and shapes

Description automatically generated with medium confidence

**The bivariate analyses yield several interesting findings:**

**Price Distribution by Room Type:** The median price varies significantly by room type, with Entire home/apts being the most expensive, followed by Private rooms and then Shared rooms. The variation within Entire home/apts and Private rooms is substantial, indicating a wide range of options for travellers.

**Price Distribution by Neighbourhood Group:** There are noticeable differences in median prices across neighbourhood groups, with Manhattan being the most expensive, followed by Brooklyn, Queens, Staten Island, and the Bronx. Manhattan and Brooklyn show a wide range of prices, reflecting their popularity and diversity of listings.

**Number of Reviews by Room Type:** Private rooms and Entire home/apts tend to have a higher number of reviews compared to Shared rooms, possibly indicating higher occupancy or guest preference. The distribution of reviews is quite wide for Private and Entire homes, suggesting variability in guest experiences or listing popularity.

**Availability by Neighbourhood Group:** Listings in Staten Island and the Bronx tend to have higher availability throughout the year, whereas Manhattan, Brooklyn, and Queens show lower availability, potentially reflecting higher demand or more active management of listings in these areas.

These insights underscore the complexity of the New York City Airbnb market, with price, demand, and availability influenced by a variety of factors including room type and geographical location.

**Temporal Trends Analysis**

Understanding how prices and availability change over time can offer insights into market seasonality, demand fluctuations, and optimal times for hosts to list their properties or for guests to book their stays.

First, we'll examine the distribution of the last review dates to understand when listings were most recently reviewed, acknowledging that our dataset stops in 2019 and may not reflect current trends.

A computer screen shot of text

Description automatically generated

A graph with numbers and a line

Description automatically generated

Peaks in the graph could correspond to popular travel seasons or specific events in New York City that attracted more guests. Notably, there's a general increase in reviews during certain periods, suggesting seasonality in booking patterns.

**Spatial Analysis with Map Visualization**

A computer screen shot of a program

Description automatically generated

A map of a city

Description automatically generated

**Correlation Analysis**

A screen shot of a computer code

Description automatically generated

A screenshot of a computer screen

Description automatically generated

**The correlation matrix provides insights into how numerical variables within the dataset are related to each other:**

**Price:** There's a very low correlation between price and other variables, indicating that factors like minimum nights, number of reviews, reviews per month, host listings count, and availability don't strongly predict the price directly. This suggests that price is influenced by other factors not captured by these variables, such as location specifics, listing quality, or amenities.

**Number of Reviews and Reviews per Month:** There's a moderate positive correlation between the number of reviews and reviews per month (0.59), which makes sense as more popular listings (with more reviews) likely receive reviews more frequently.

**Calculated Host Listings Count and Availability:** A slight positive correlation (0.23) suggests that hosts with more listings might have slightly higher availability across their listings, possibly indicating professional hosting operations.

**Room Type and Location Impact**

A screen shot of a computer program

Description automatically generated

A group of graphs with different sizes and shapes

Description automatically generated with medium confidence

**The analysis of room type and location (neighbourhood group) reveals significant insights into their impact on price and popularity:**

**Price Distribution by Room Type:** The price varies significantly with room type. Entire homes/apartments tend to have the highest median prices, followed by private rooms, and then shared rooms. The variability in prices is also much higher for entire homes/apartments, indicating a broad range of offerings, from budget to luxury.

**Price Distribution by Neighbourhood Group:** Manhattan stands out as having the highest median prices, consistent with its status as a highly sought-after location. Brooklyn and Queens follow, with Staten Island and the Bronx being the most affordable. This reflects the general market perception of these areas and their demand among tourists and visitors.

**Number of Reviews by Room Type:** Private rooms and entire homes/apartments generally have more reviews than shared rooms, suggesting they are more popular or frequently booked options. This could be due to a preference for privacy or the nature of trips (tourism, business, etc.) that NYC attracts.

**Number of Reviews by Neighbourhood Group:** The distribution of reviews across neighbourhood groups does not show as stark a contrast as price does, suggesting that while guests are willing to pay more to stay in certain areas, they are generally satisfied with their experiences across the city. Brooklyn and Manhattan, being popular tourist destinations, have a wide range of reviews, reflecting their higher traffic.

**Host Impact**

A screen shot of a computer program

Description automatically generated

A graph showing a number of listings

Description automatically generated

**The scatter plot comparing the number of listings a host manages against the reviews per month, both on logarithmic scales for better visibility, reveals some interesting patterns:**

**Variability in Reviews per Month:** There's a wide range of reviews per month across listings, regardless of how many properties a host manages. This suggests that the popularity of listings (as indicated by review frequency) is not strictly dependent on the number of listings a host has.

**Trend Observation:** While there are hosts with many listings and varying levels of review activity, there doesn't appear to be a clear linear relationship between the number of listings managed and the average number of reviews per month. Some hosts with a high number of listings have listings with lower average reviews per month, possibly indicating that managing more listings could dilute the attention each listing receives, affecting its overall review frequency.

**Diverse Host Strategies:** The data points are spread across the plot, highlighting the diversity in hosting strategies and outcomes. Some hosts manage to maintain high levels of reviews per month across multiple listings, while others with fewer listings have both high and low review frequencies.

**EDA Summary and Insights**

**Through our in-depth EDA, we've uncovered a variety of insights into the NYC Airbnb market:**

**Pricing Dynamics:** Room type and location are significant determinants of pricing, with entire homes/apartments and listings in Manhattan commanding higher prices.

**Popularity Indicators:** The number of reviews, which can serve as a proxy for listing popularity or booking frequency, shows variability across room types and neighbourhoods, indicating a healthy demand across the board.

**Host Influence:** The analysis suggests that the relationship between the number of listings a host manages, and the popularity of those listings is complex, with no straightforward correlation. Quality, listing presentation, and guest experience may play more critical roles in driving reviews than the sheer number of listings managed.

These insights can inform both hosts and guests in making decisions within the Airbnb platform. For hosts, focusing on quality listings and guest experience, regardless of the number of properties managed, seems essential. For guests, these findings highlight the importance of considering room type and location in their booking decisions, balancing cost with their preferences and needs.

**Literature Review**

**Introduction**

The sharing economy has transformed traditional industries, with Airbnb playing a significant role in the lodging and tourism sectors. Understanding the spatial and temporal dynamics of the Airbnb market is crucial for stakeholders, including hosts, guests, policymakers, and local communities.

**Study 1: Spatial Dynamics and Market Implications**

**Reference:** “Airbnb phenomenon: a review of literature and future research directions”

**Methodology Overview:** The paper reviews academic research on Airbnb from 2010–2022, focusing on consumer behaviour, host behaviour, host–guest relationship, trust in Airbnb, dominant theories in Airbnb, Airbnb regulation, Airbnb and hotels, and macro impacts of Airbnb.

**Key Findings:**

* Most existing Airbnb research has been conducted in Europe, USA/Canada, followed by Asian countries like China, Singapore, S. Korea, and India.
* "The study provides deeper insights into the decision-making process of both guests and hosts by examining the relevant motivators and constraints." (“Airbnb phenomenon: a review of literature and future research directions”)

**Study 2: Temporal Trends and Regulatory Impacts**

**Reference:** “Progress on Airbnb: a literature review”

**Methodology Overview:** This paper reviews the extant literature on Airbnb to assess the research progress that has been accomplished to date.

**Key Findings:**

* "Consistent findings have begun to emerge on several important topics, including guests’ motivations and the geographical dispersion of listings." (“Progress on Airbnb: a literature review | Emerald Insight”)
* This research provides a concise summary of Airbnb knowledge that will assist industry practitioners as they adapt to the recent rapid emergence of Airbnb.

**Comparative Analysis**

Both studies provide valuable insights into the spatial and temporal dynamics of the Airbnb market. While the first study offers a comprehensive overview of the Airbnb phenomenon, the second study provides a concise summary of Airbnb knowledge that will assist industry practitioners.

**Synthesis and Implications**

The insights from these studies can help hosts optimize their listing performance through strategic pricing, timing, and location selection. Guests can use these insights to find value or avoid peak pricing. Policymakers can balance regulatory goals with healthy market function and community impacts. Researchers can identify gaps in current research and suggest future studies to address unresolved questions or emerging trends.

**Conclusion**

The reviewed literature enriches our understanding of the complex dynamics influencing Airbnb markets. The sharing economy continues to evolve, with significant implications for various stakeholders.

**References**

1. “Airbnb phenomenon: a review of literature and future research directions”
2. “Progress on Airbnb: a literature review”

**Architecture/Methodology**

The EDA employed a combination of statistical analyses and data visualization techniques. Key steps included descriptive statistics, correlation analysis, and bivariate and multivariate analyses. Tools such as Python, Pandas, and Seaborn facilitated the exploration, with spatial data visualized using Folium maps.

**Results**

Our analysis revealed:

* **Price Dynamics:** Significant variation in pricing based on location and room type, with Manhattan and Entire home/apts commanding higher prices.
* **Demand Indicators:** Number of reviews served as a proxy for demand, showing popularity across various room types and neighbourhoods.
* **Host Activity:** No clear correlation was found between the number of listings a host manages and the popularity of those listings, indicating the multifaceted nature of hosting success.

**Recommendations or Potential Avenues for Further Analysis**

Based on the exploratory data analysis (EDA) of the NYC Airbnb dataset, several recommendations for hosts, guests, and policymakers can be derived. Additionally, there are multiple avenues for further analysis that could provide deeper insights into the Airbnb market dynamics. Here are some tailored recommendations and potential areas for further research:

**Recommendations**

For Hosts:

1. **Dynamic Pricing:** Consider adopting dynamic pricing strategies to maximize revenue, especially for listings in high-demand neighbourhoods or during peak seasons.
2. **Optimize Listing Quality:** Enhance listing descriptions, quality of photos, and amenities offered to improve attractiveness and competitive edge, potentially increasing both pricing and occupancy rates.
3. **Focus on Reviews:** Encourage guests to leave reviews and actively manage listings to address any feedback, as listings with higher numbers of positive reviews tend to attract more guests.

For Guests:

1. **Booking Timing:** Plan and book stay during off-peak periods to find better deals, especially in high-demand areas like Manhattan.
2. **Explore Emerging Neighbourhoods:** Consider staying in up-and-coming neighbourhoods outside central areas for better prices without significantly compromising on accessibility to major attractions.

For Policymakers:

1. **Regulation Impact Studies:** Conduct studies on the impact of short-term rental regulations to balance community needs, housing affordability, and the economic benefits of tourism.
2. **Data Transparency:** Encourage transparency and data sharing between platforms like Airbnb and city governments to better understand market dynamics and inform policy decisions.

**Potential Avenues for Further Analysis**

1. **Longitudinal Impact of Regulations:** Analyse how specific regulatory changes over time affect listing prices, availability, and distribution to assess the long-term impacts of policy decisions on the Airbnb market.
2. **Impact of External Factors:** Explore how external factors like major events, economic shifts, or pandemics (e.g., COVID-19) impact the short-term rental market in terms of pricing, demand, and host strategies.
3. **Comparative Analysis Between Cities:** Conduct a comparative study between NYC and other major cities to identify unique market dynamics and regulatory impacts, providing a broader perspective on Airbnb's role in urban environments.
4. **Neighbourhood-Level Analysis:** Dive deeper into neighbourhood-level trends to uncover granular insights into local market conditions, identifying areas of high growth potential or saturation.
5. **Predictive Modelling:** Develop predictive models to forecast future trends in prices, demand, and listing availability, helping stakeholders make informed decisions based on anticipated market developments.
6. **Socioeconomic Impacts:** Examine the socioeconomic impacts of Airbnb on local communities, including effects on rental prices, housing availability, and neighbourhood character, to inform balanced and equitable policy frameworks.
7. **Host Economic Analysis:** Analyse the economic impact on hosts, particularly those relying on Airbnb income, to understand the platform's role in supporting livelihoods and identifying support mechanisms for hosts during market downturns.

These recommendations and avenues for further analysis aim to leverage the insights gained from the EDA to benefit various stakeholders and guide future research towards areas that could significantly impact understanding and decision-making in the Airbnb marketplace.

**Conclusion**

The NYC Airbnb market is influenced by a complex interplay of factors, including location, listing type, and host strategy. While pricing and demand vary widely, the data suggests opportunities for both hosts and guests to navigate the market effectively. Future work could delve deeper into predictive modelling to forecast market trends.

**References**

* Source 1: [Data Exploration on NYC Airbnb (kaggle.com)](https://www.kaggle.com/code/dgomonov/data-exploration-on-nyc-airbnb)
* Source 2: [New York City Airbnb Open Data (kaggle.com)](https://www.kaggle.com/datasets/dgomonov/new-york-city-airbnb-open-data/code)

**GitHub Repository for Code**

All code used for the data analysis, preprocessing, and visualization is available at a GitHub repository:

<https://github.com/Yash0831/CS661-MIDTERM-PROJECT>