Unveiling the Path Forward: Strategic Insights for Airbnb's Next Business Steps through New York Listing Data Analysis

Sagar Kumar Makhija

Souradeep Bose

SOUBHIK DAS

Agenda

01. Objective

02. Data Life Cycle

03. Key Findings

04. Recommendations

05. Appendix

- Data Attributes
- Data Methodologies
- Data Assumptions

Objective

Customer Preference and User Experience Understanding:

- Analyze and comprehend the evolving customer preferences in the post-COVID period.
- Identify and leverage current user experience trends specific to the Airbnb NYC business.

Strategic Recommendations for New Acquisitions:

- Provide early insights into potential areas and property types for new acquisitions.
- Develop strategies for effective acquisition processes to adapt to the changing market dynamics.

Enhancing Customer Experience:

- Evaluate current customer experience to pinpoint areas for improvement.
- Deliver actionable recommendations to enhance overall customer satisfaction and engagement.

Data Life Cycle

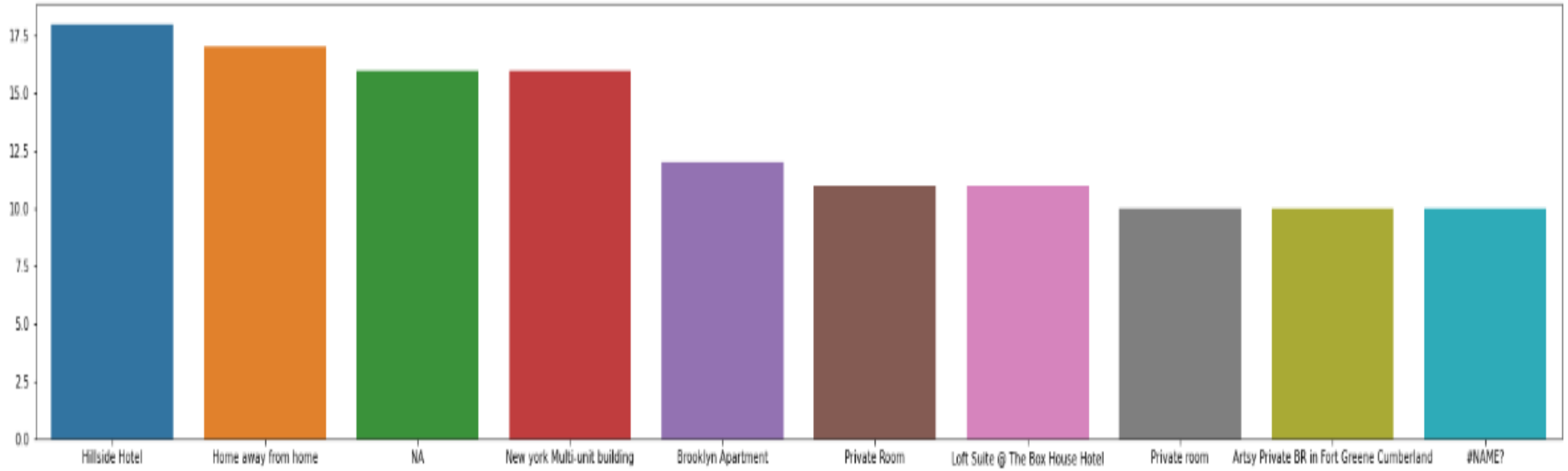
In the initial phase, data is systematically captured and subsequently loaded into diverse environments

After completing the data cleaning process, conducting Exploratory Data Analysis (EDA), and generating additional features

Significant and valuable insights are unearthed through the application of diverse analytical methods.

Key Findings

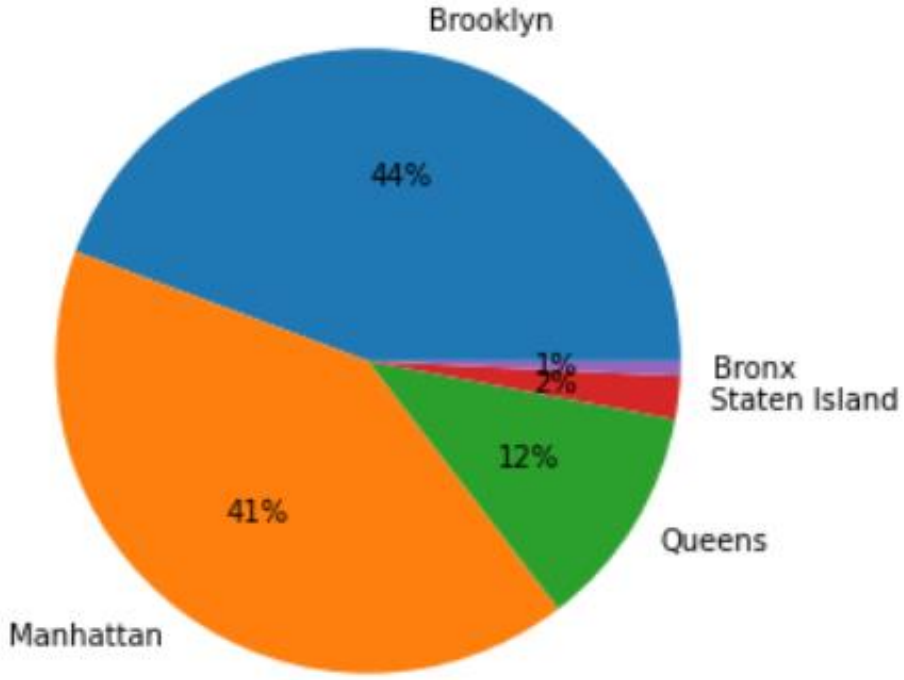
- **Top 10 Property Types booked by customers**



- **Hillside Hotel** and **Home away from home** are the top 2 property types booked.

Most Contributing Neighbourhoods

Neighbourhood Groups

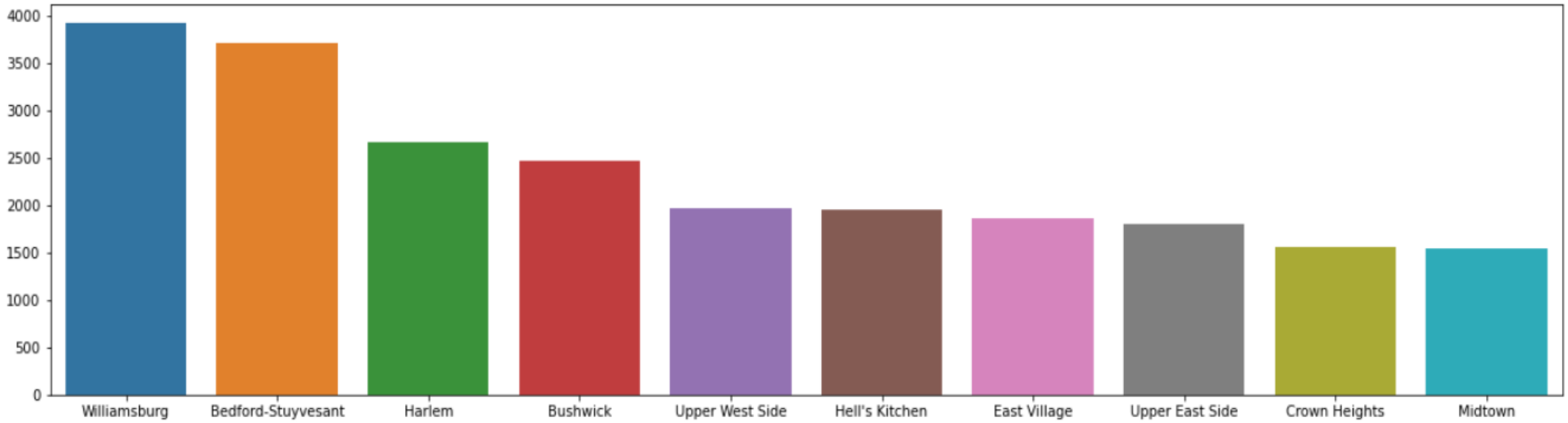


- **85% of listings** are in **Manhattan and Brooklyn** making them the most popular destinations
- Brooklyn has the maximum contribution closely followed by Manhattan.
- **Staten Island** has the lowest contribution.

Neighbourhood Group	Count of neighbourhood_group
Manhattan	44.30%
Brooklyn	41.12%
Queens	11.59%
Bronx	2.23%
Staten Island	0.76%
Grand Total	**100.00%**

Top 10 Neighbourhood Groups

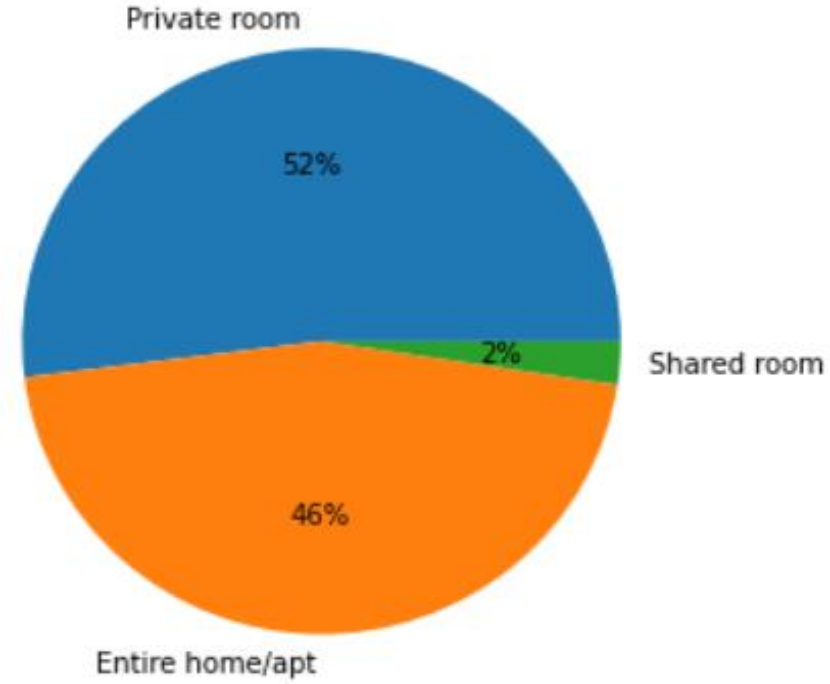
- **Neighbourhood Groups from high to low bookings**



- **Williamsburg, Beedford-Stuyvesant,Harlem & Bushwick** are the most popular neighboorhood groups with more than **5% bookings.**
- Top 10 Neighbourhood Groups only make up to **48% of the overall contribution.**

Private Rooms is preferred Room Type

Types of Room



THE PROBLEMS WITH SHARED ROOMS

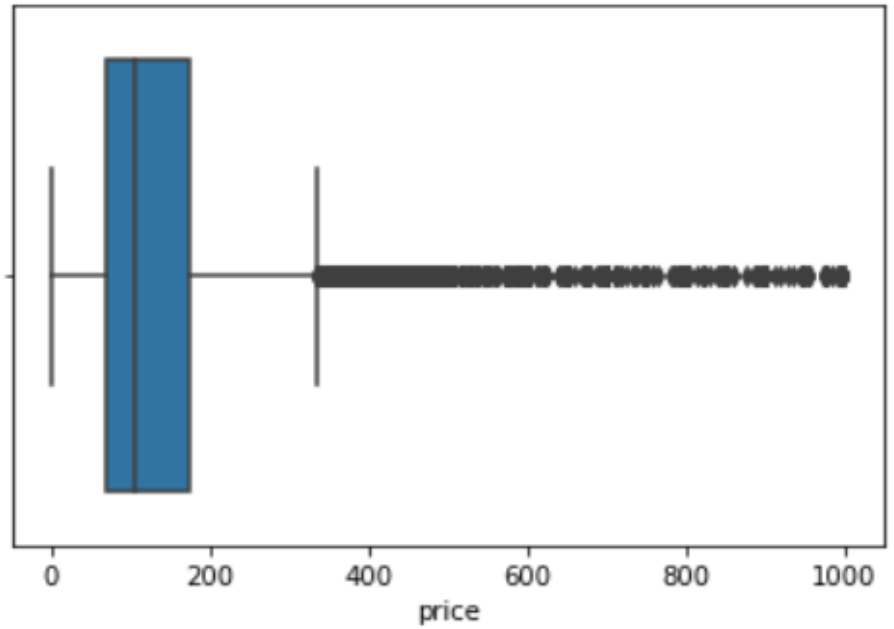
Room Types	Count of room_type
Entire home/apt	51.97%
Private room	45.66%
Shared room	2.37%
Grand Total	**100.00%**

- **98%** of guests prefer a **private room or entire home/apartment** making shared rooms less preferred choice for property type
- **Shared rooms** only account for **2% of the total types** of rooms. They are less likely to be reviewed.

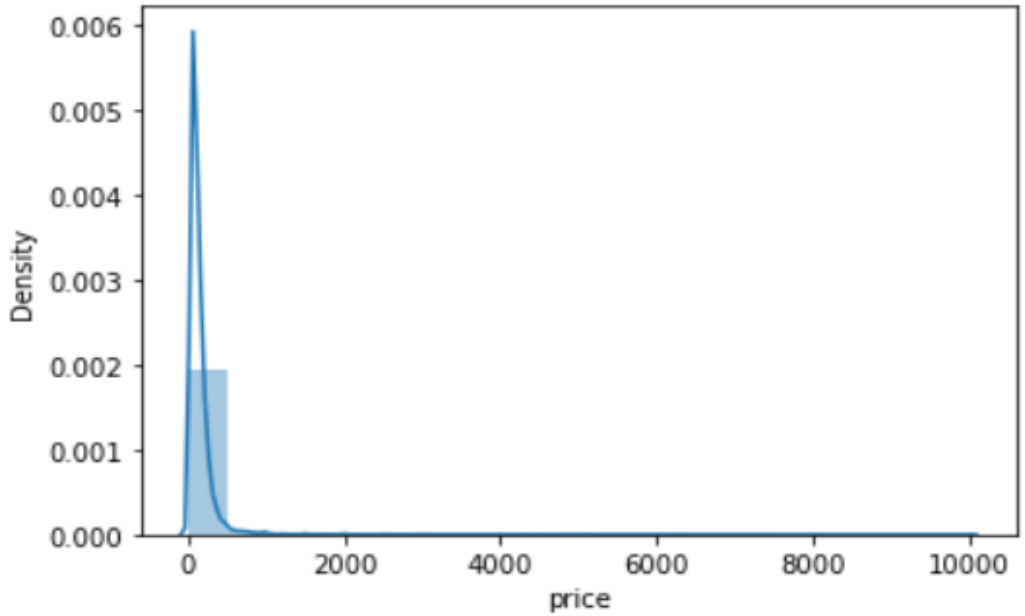
Price Distribution

- Perform univariate analysis for continuous variable **Price.**
- **Box Plot & Histogram** are being used to derive **Price analysis.**

Outlier Analysis



Relationship between price and density

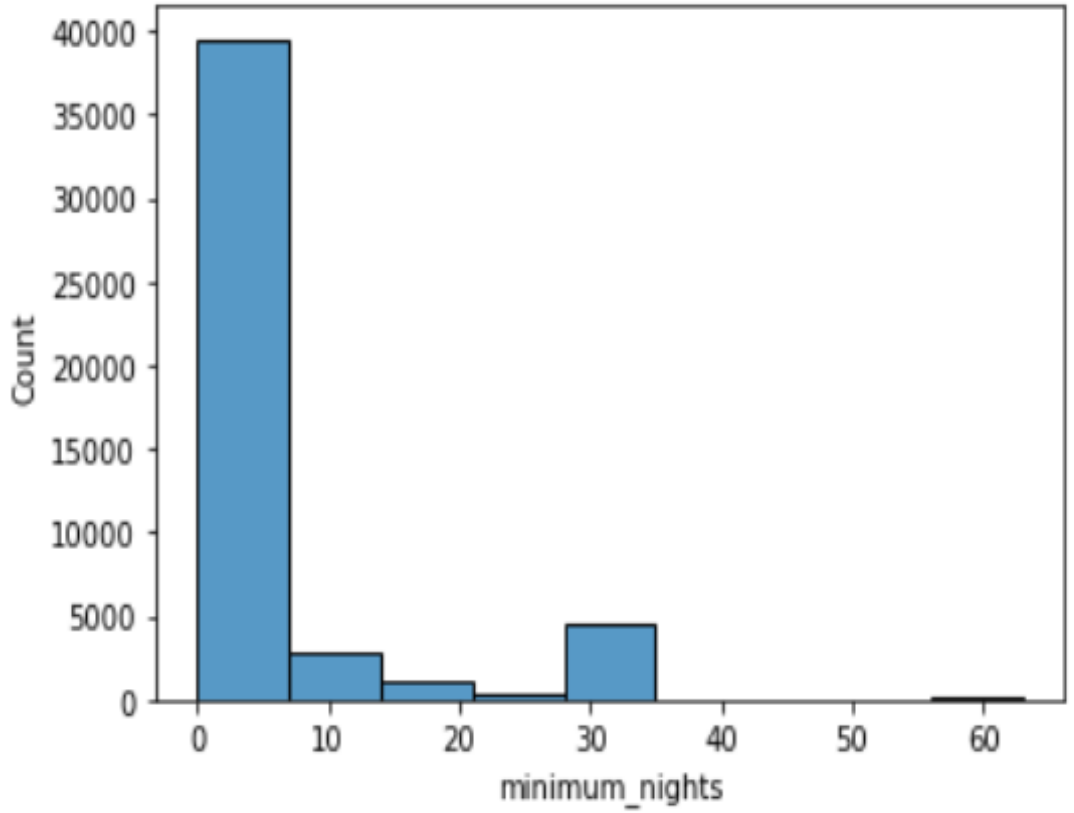


- The above histograms and box plots show that majority of listed properties are relatively **low priced (<$400)** which shows that tenants **prefer cheaper properties** on Airbnb

Minimum Nights

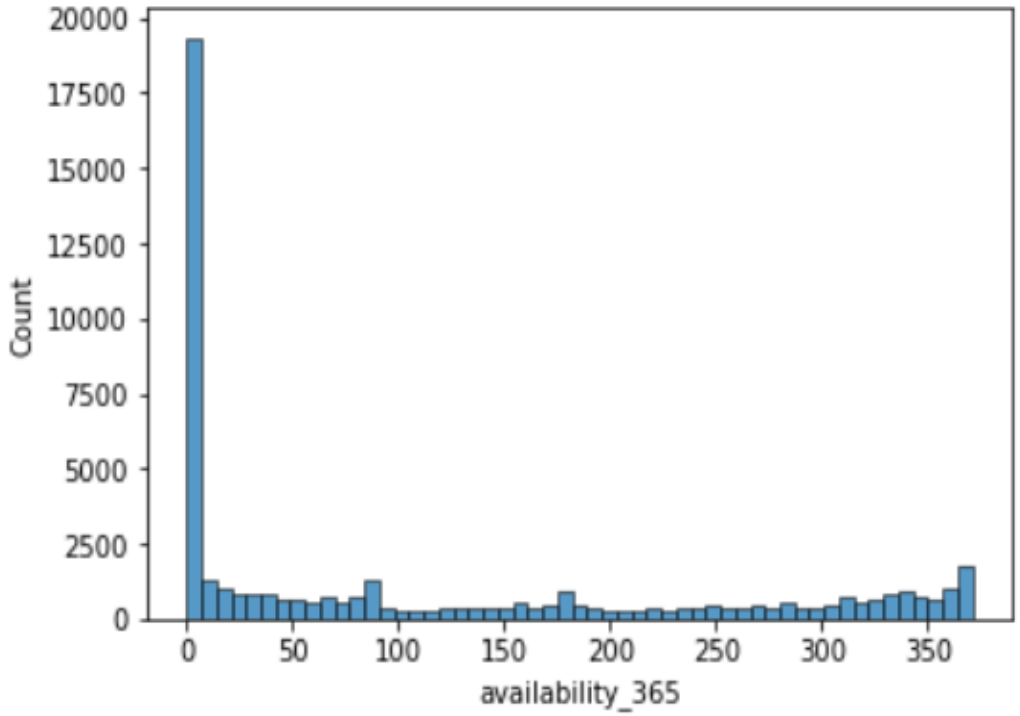
Minimum Nights Booking Preference

- Majority of tenants prefer to book establishments with **minimum limit of 7 days**, with establishments with **30 days stay policy** being the second most popular one.
- **85% of tenants** prefer to book with minimum limit 7 days followed by **7.69% with 30 days** stay policy

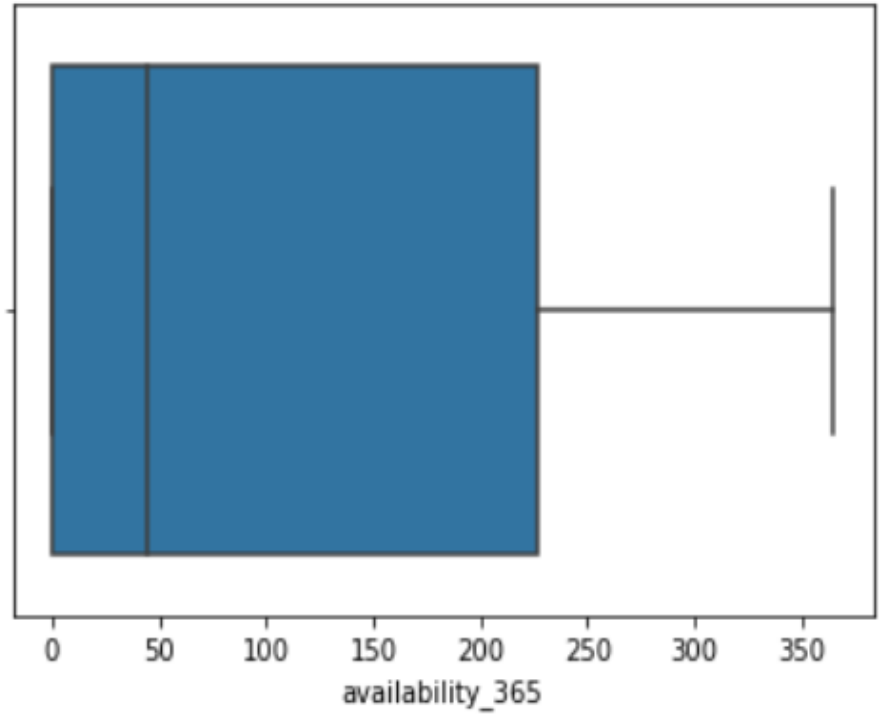


Univariate Analysis

Availability_365



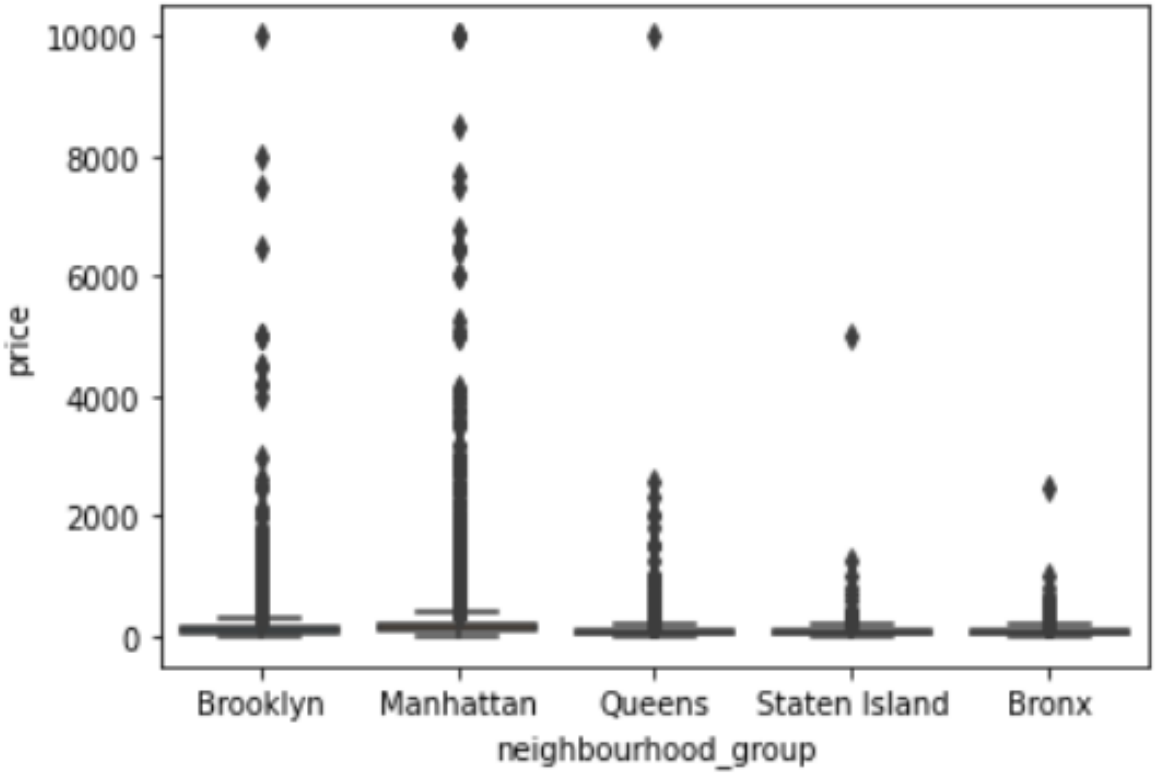
No of Days when listing is available for booking

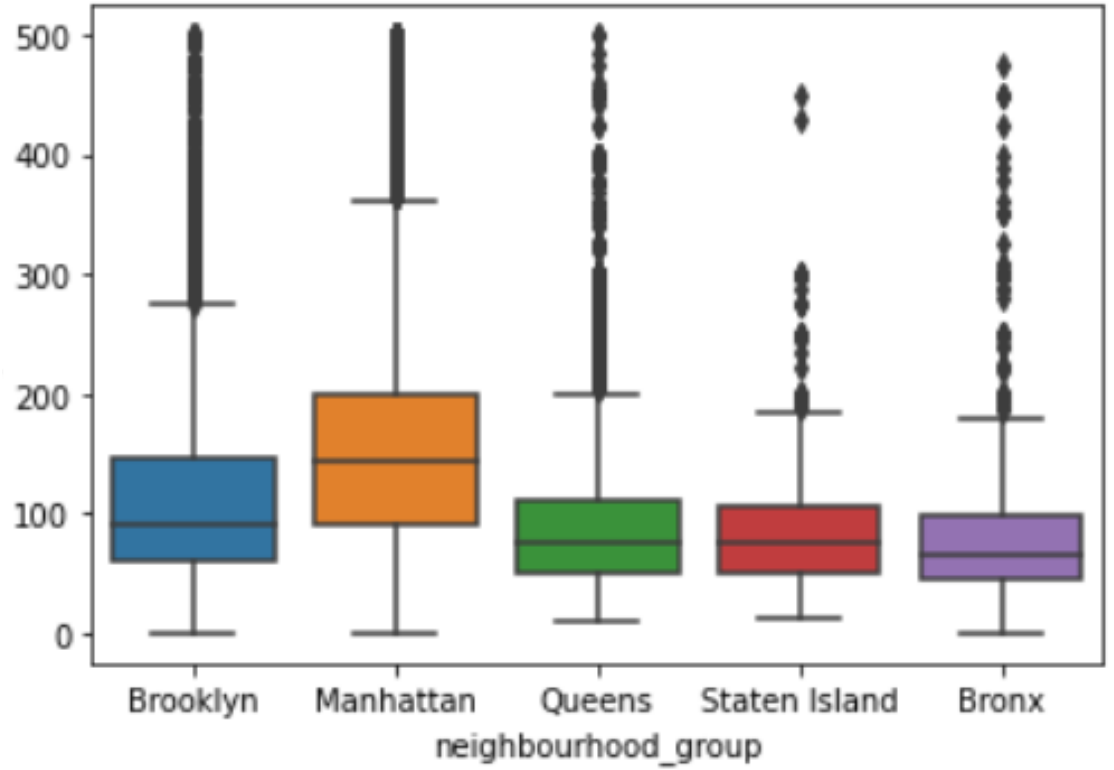


- Majority listings are booked for the entire year

Bivariate Analysis

- Analysis of **Price** vs **Neighbourhood_group**

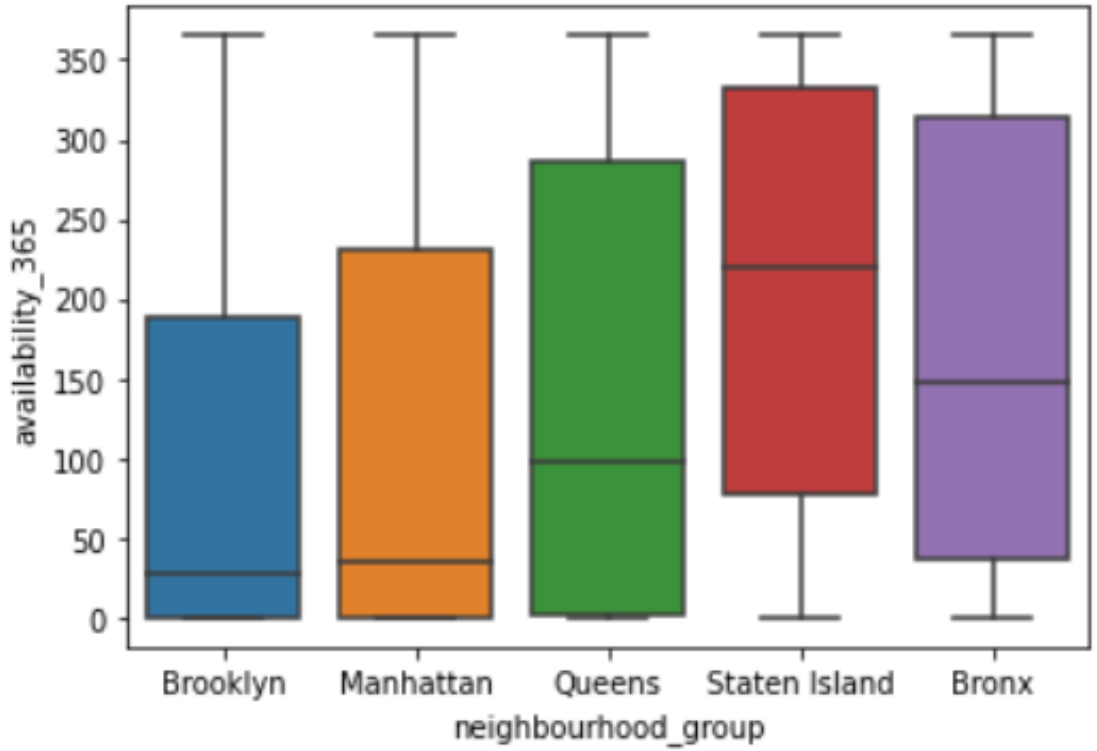




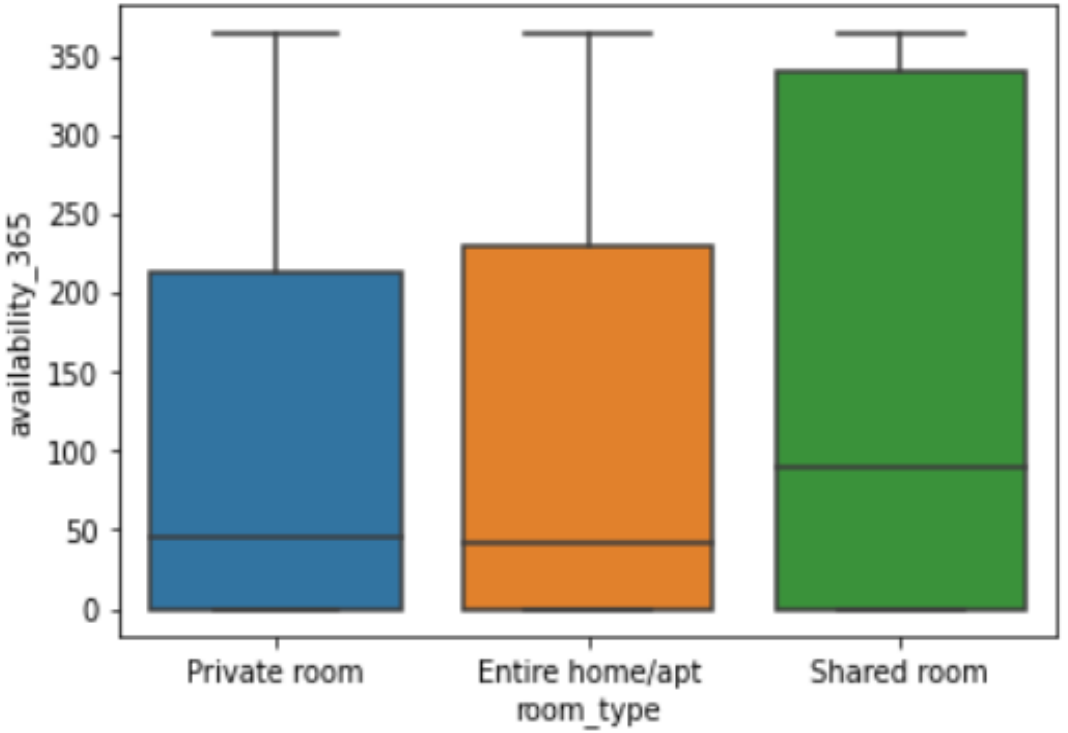
- Prices are **higher in Manhattan** making it one of the more attractive area

Bivariate Analysis

- Analysis of **Availability** vs **Neighbourhood_group**



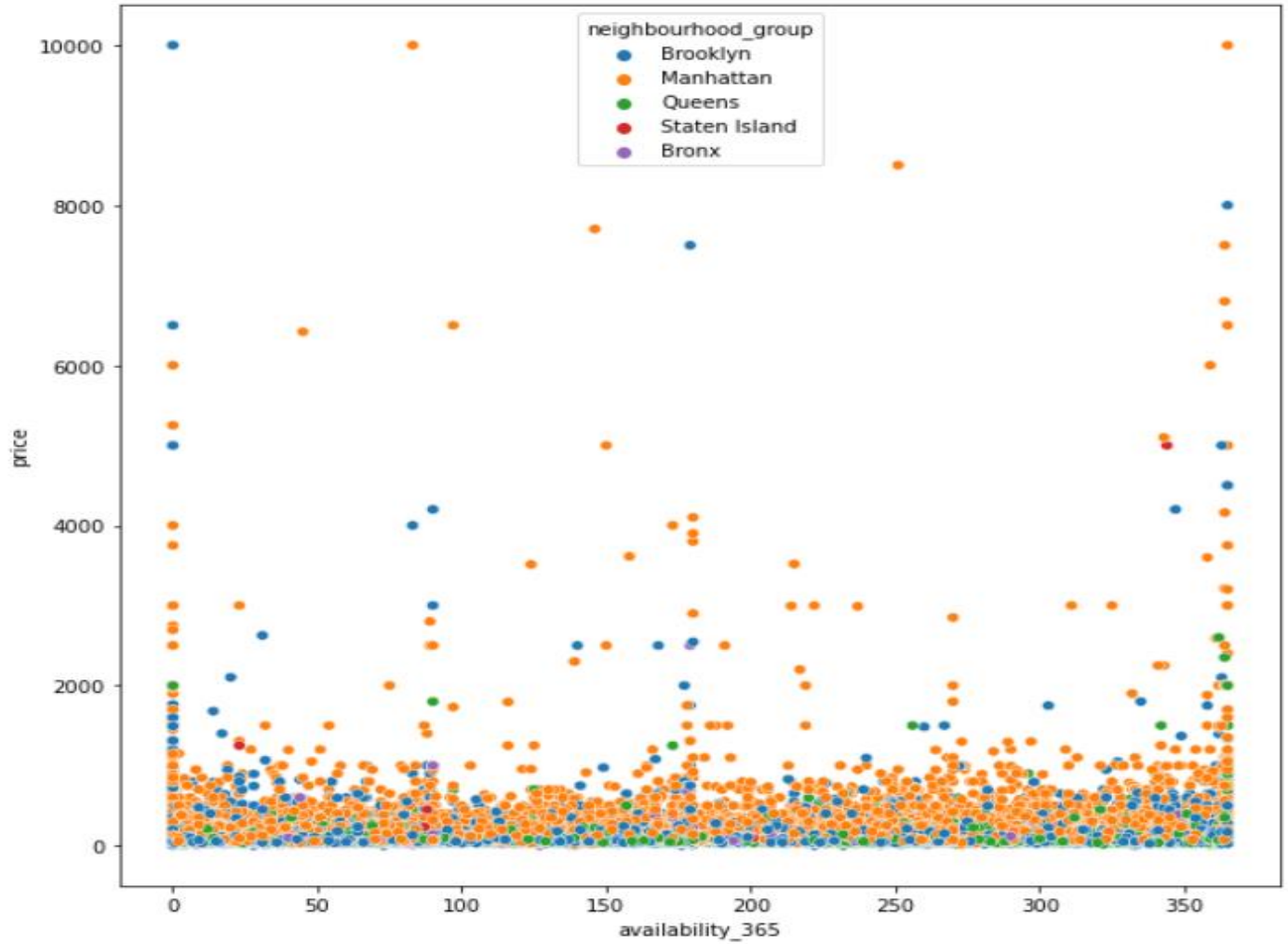
- Analysis of **Availability** vs **Property_type**



- Rooms in **Staten Island are least booked** even though prices are **lower while Brooklyn properties** are rented out for more time during the year

- **Entire home/apartments are priced higher** and yet booked throughout the year making them attractive options for promoting

Price vs Minimum Nights



Conclusion

Results & Observations

- Robust and impactful insights are uncovered by analyzing diverse attributes within the dataset.
- Abundant and diverse visuals can be incorporated into the presentations for stakeholders.
- The data collection team is advised to gather information on review scores to enhance subsequent analyses.

Properties with Least Min Nights to Stay offer Maximum Bookings

- As the number of nights to stay increases, the price increases and the bookings naturally go down.
- It has been observed, that the hosts offering min nights to stay have received the highest booking in the past.

Customer Preferences for Entire Apt/Pvt. Room should remain high post-COVID

- Shared rooms account for only 2.4% & thus needs to be reviewed.
- Private Room or Entire apartment is preferred choice & needs to be explored in other neighbourhoods.

Appendix –Data Attributes

Column	Description
id	listing ID
name	name of the listing
host_id	host ID
host_name	name of the host
neighbourhood_group	location
neighbourhood	area
latitude	latitude coordinates
longitude	longitude coordinates
room_type	listing space type
price	
minimum_nights	amount of nights minimum
number_of_reviews	number of reviews
last_review	latest review
reviews_per_month	number of reviews per month
calculated_host_listings_count	amount of listing per host
availability_365	number of days when listing is available for booking

Dataset Description

- New York Airbnb dataset contains information about different Airbnb listings along with their hosts, locations, prices and other attributes.
- The columns in the dataset are self-explanatory. You can refer to the diagram given below to get a better idea of what each column signifies.

Appendix – Data Methodology

Performed an in-depth examination of the dataset containing New York Airbnb information:

- Employed Python to clean the dataset, addressing missing values and outliers.
- Utilized exploratory data analysis techniques to discern customer preferences.
- Applied group aggregation, pivot tables, and various statistical methods for in-depth analysis.
- Generated charts and visualizations using Matplotlib and Seaborn libraries in Python.

Appendix – Data Assumptions

```
Categorical Variables:
    - room_type
    - neighbourhood_group
    - neighbourhood

Continous Variables(Numerical):
    - Price
    - minimum_nights
    - number_of_reviews
    - reviews_per_month
    - calculated_host_listings_count
    - availability_365
- Continous Variables could be binned in to groups too

Location Varibles:
    - latitude
    - longitude

Time Varibale:
    - last_review
```

Variable Categories

- Assuming a classification of variables into distinct types, including categorical, numeric, location, and time. The significance lies in discerning the appropriate plotting method for each variable.
- Assuming the data from the pre-COVID-19 period reflects the intended revenue outcomes.
- Assuming the company's strategies are formulated under the expectation that travel will increase in the post-COVID period.

Thank you