# Storytelling Case Study

## Airbnb, NYC

---

## Presentation I

Airbnb

# Agenda

# Background

o Airbnb is an online platform that allows individuals to rent out their unused accommodations.

o In recent months, Airbnb has experienced a significant drop in revenue due to COVID-19.

o Now that the restrictions have started lifting and people have started to travel more, Airbnb wants to make sure that it is fully prepared for this change.

o To facilitate this, an analysis has been conducted on a dataset containing various Airbnb listings in New York.
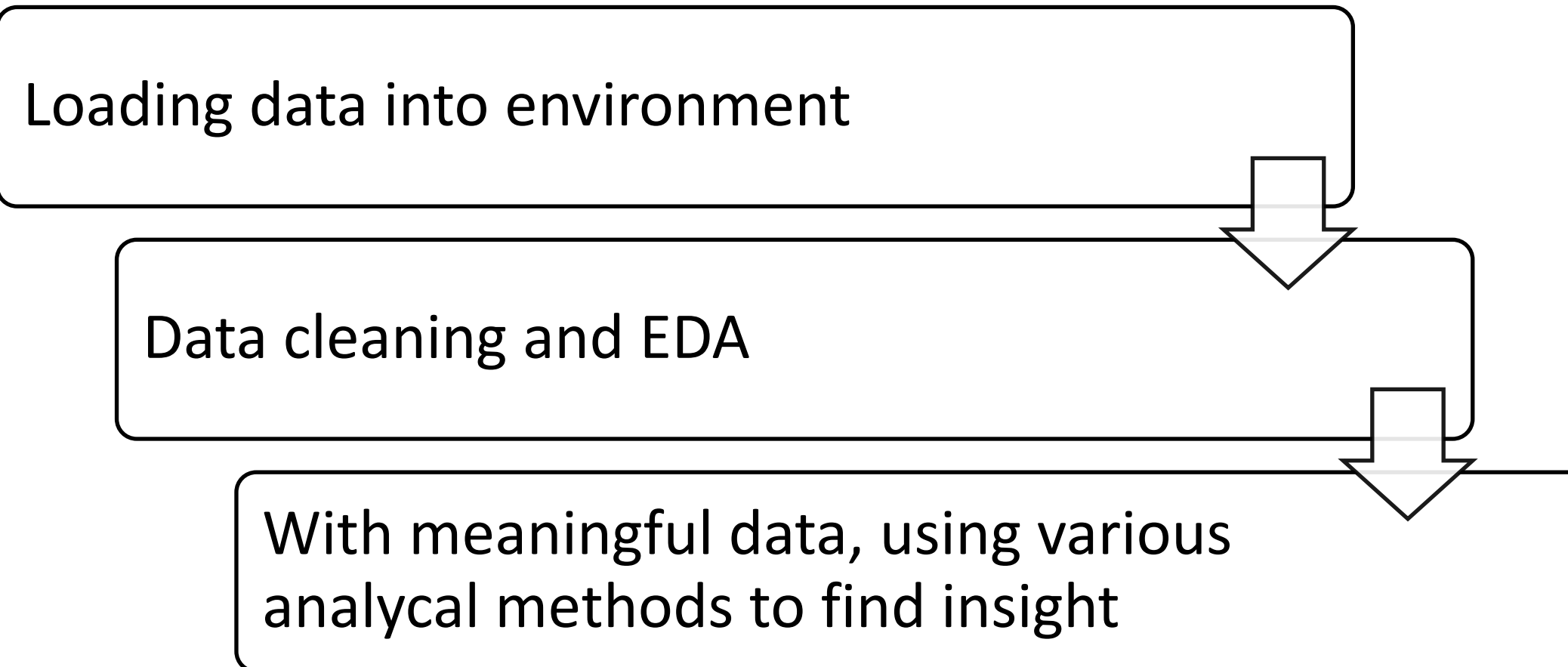
Airbnb

# Objective

- o To perform a comprehensive analysis of the New York Airbnb dataset.
- o To give some effective questions that can lead to valuable data insights
- o Process, analyze, and present findings using data visualization and statistical techniques.

Airbnb

## DATA LIFE CYCLE

Loading data into environment

Data cleaning and EDA

With meaningful data, using various analycal methods to find insight

For convenience of tracking and synchronization on the same tool, we decided to only use python in this analysis. That means we will import data into python, perform EDA, analyze and visualize in python

# Key findings from "Room Types"



count of each room types
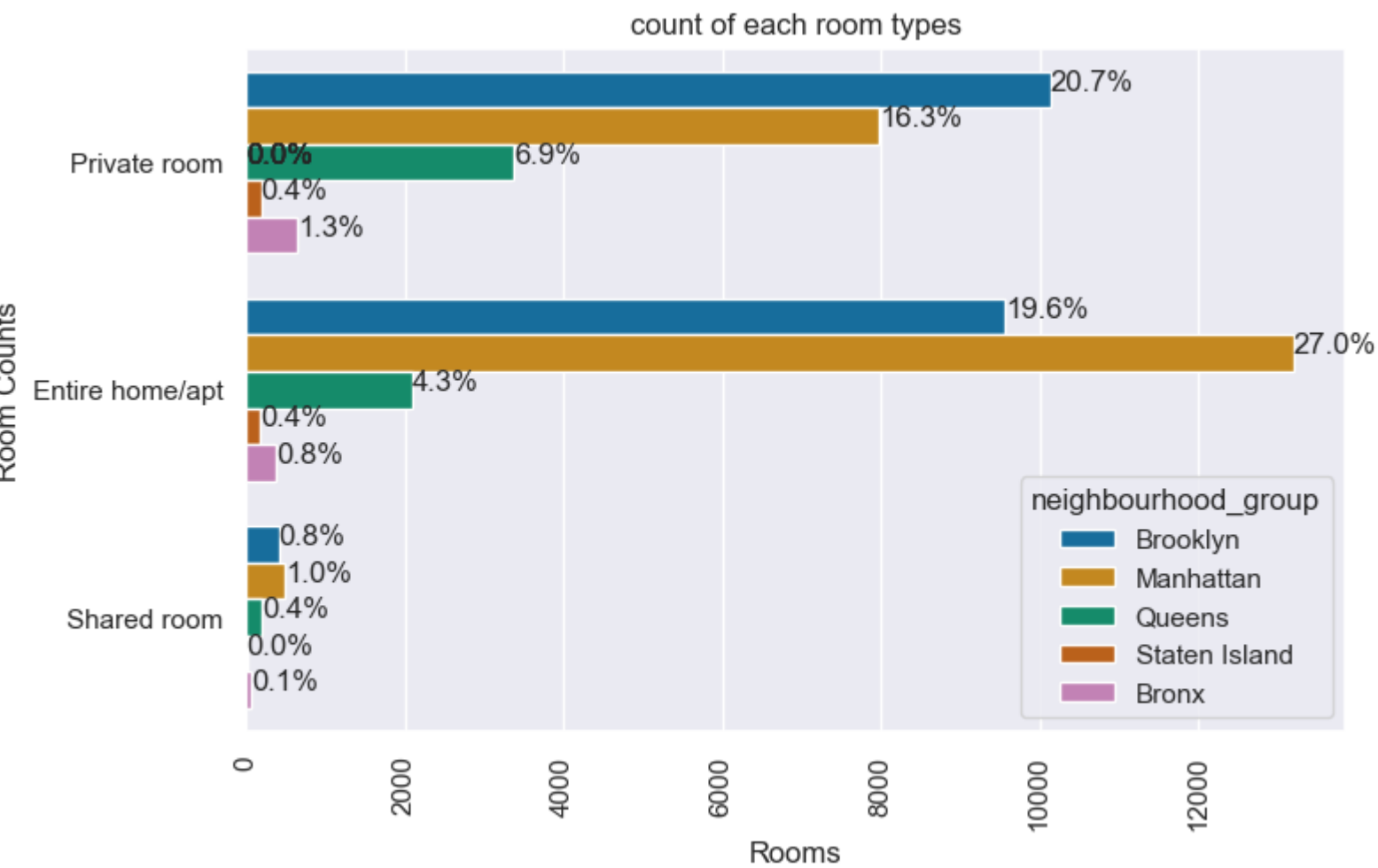
- Shared rooms represent only 2% of the total room.

- Private rooms are more prevalent in Brooklyn, followed by Manhattan

- Entire home/apt are listed 27% in Manhattan and Brooklyn comes in second with about 19.6%

⇒ *Manhattan has more properties listed than any other areas*

⇒ *The total number of shared rooms listed on Airbnb is quite limited, with very few available in Staten Island and the Bronx, where shared rooms are negligible or extremely rare.*
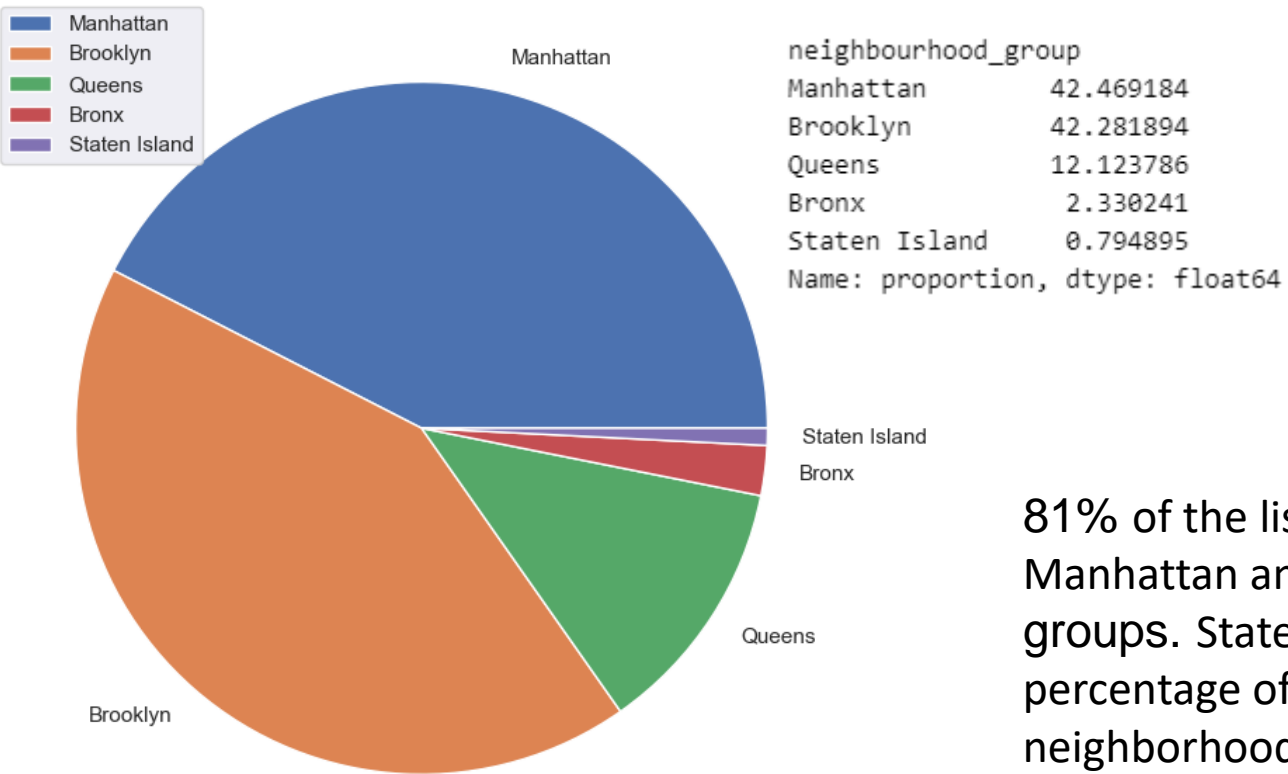
**On an average for how many nights people stayed in each room_types**

Generally people prefer staying in an Entire home or apartment, with an average duration of over 8 nights. In contrast, guests who choose shared rooms typically stay for an average of 6 to 7 nights.



Min Stays in different room types listed on Airbnb

# Data Insights
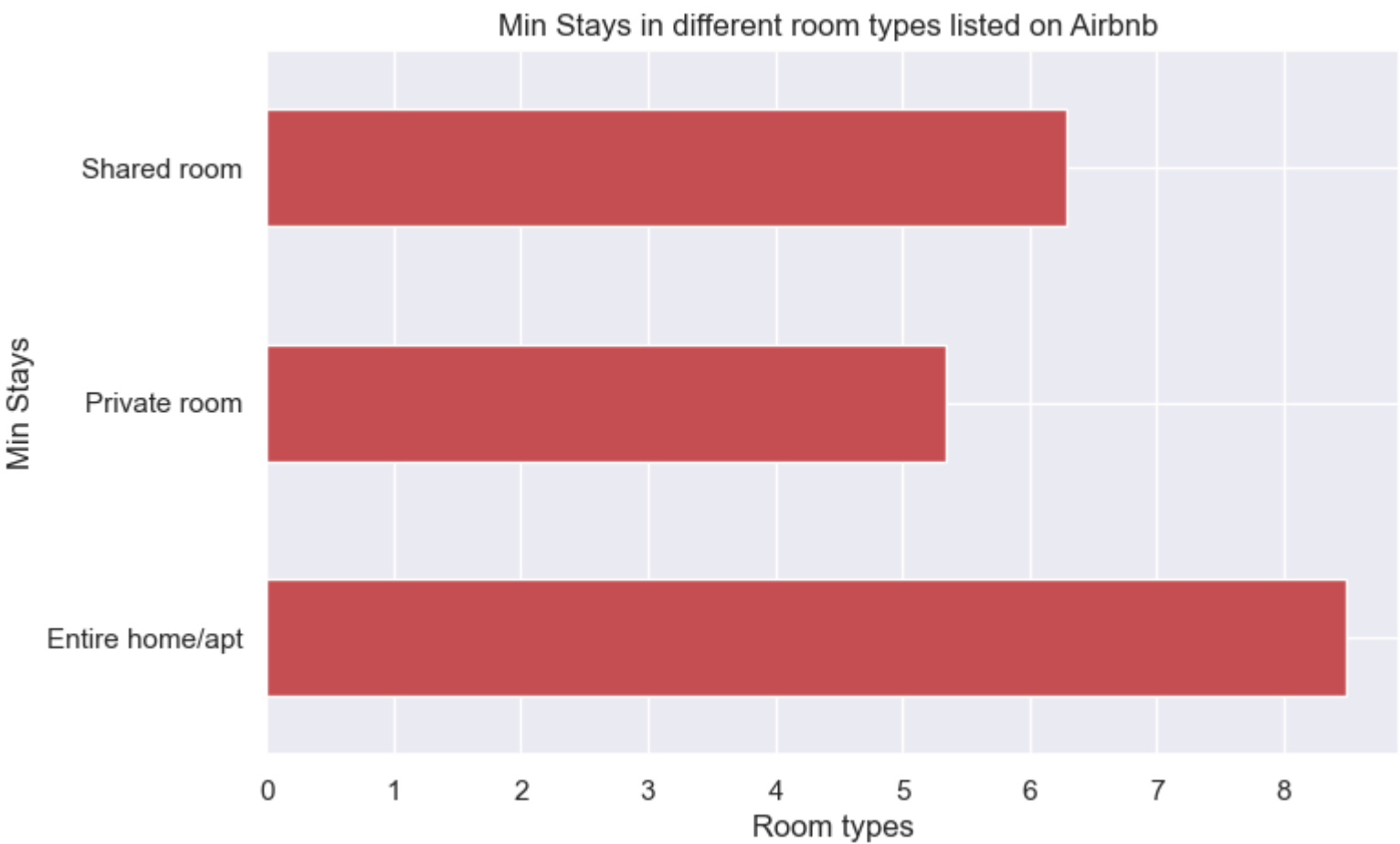
## Key findings from "Neighborhoods"

o Staten Island is higher availability throughout the year, often exceeding 300 days. On average, these listings are available for about 210 days per year.

o Following Staten Island, the Bronx exhibits an average availability of around 150 days for its listings each year.



neighbourhood_group
Manhattan        42.469184
Brooklyn         42.281894
Queens           12.123786
Bronx             2.330241
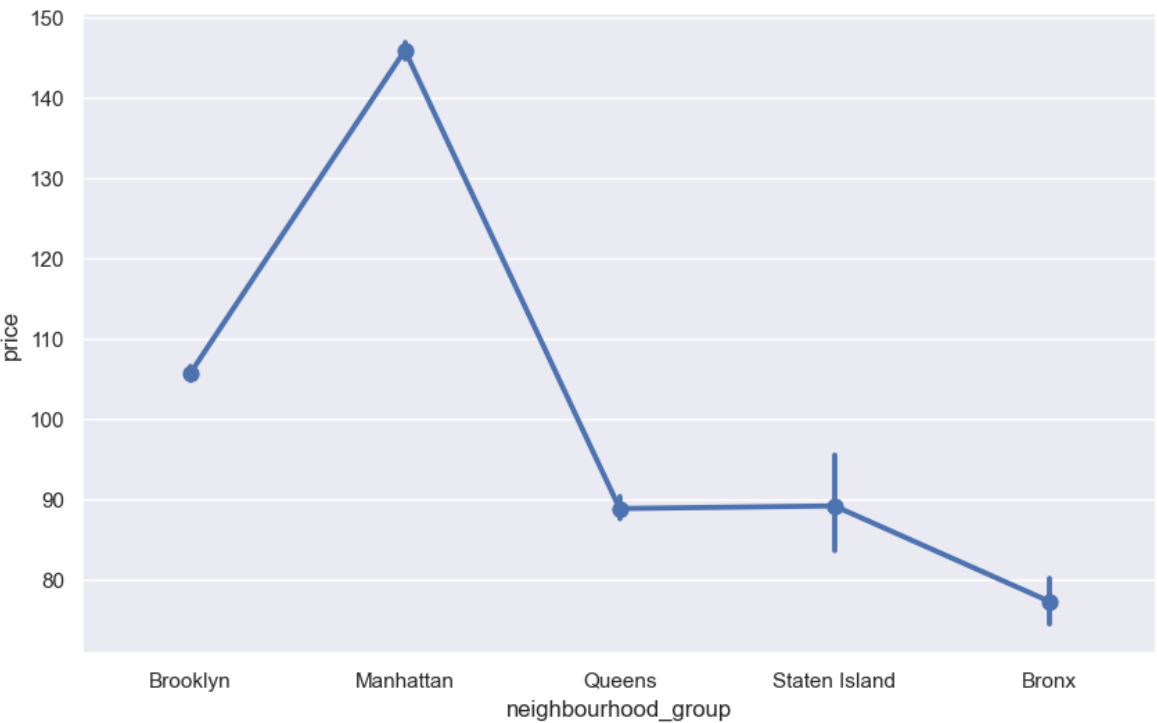Staten Island     0.794895
Name: proportion, dtype: float64

81% of the listings are from the Manhattan and Brooklyn neighborhood groups. Staten Island has the lowest percentage of contributions among the neighborhood groups
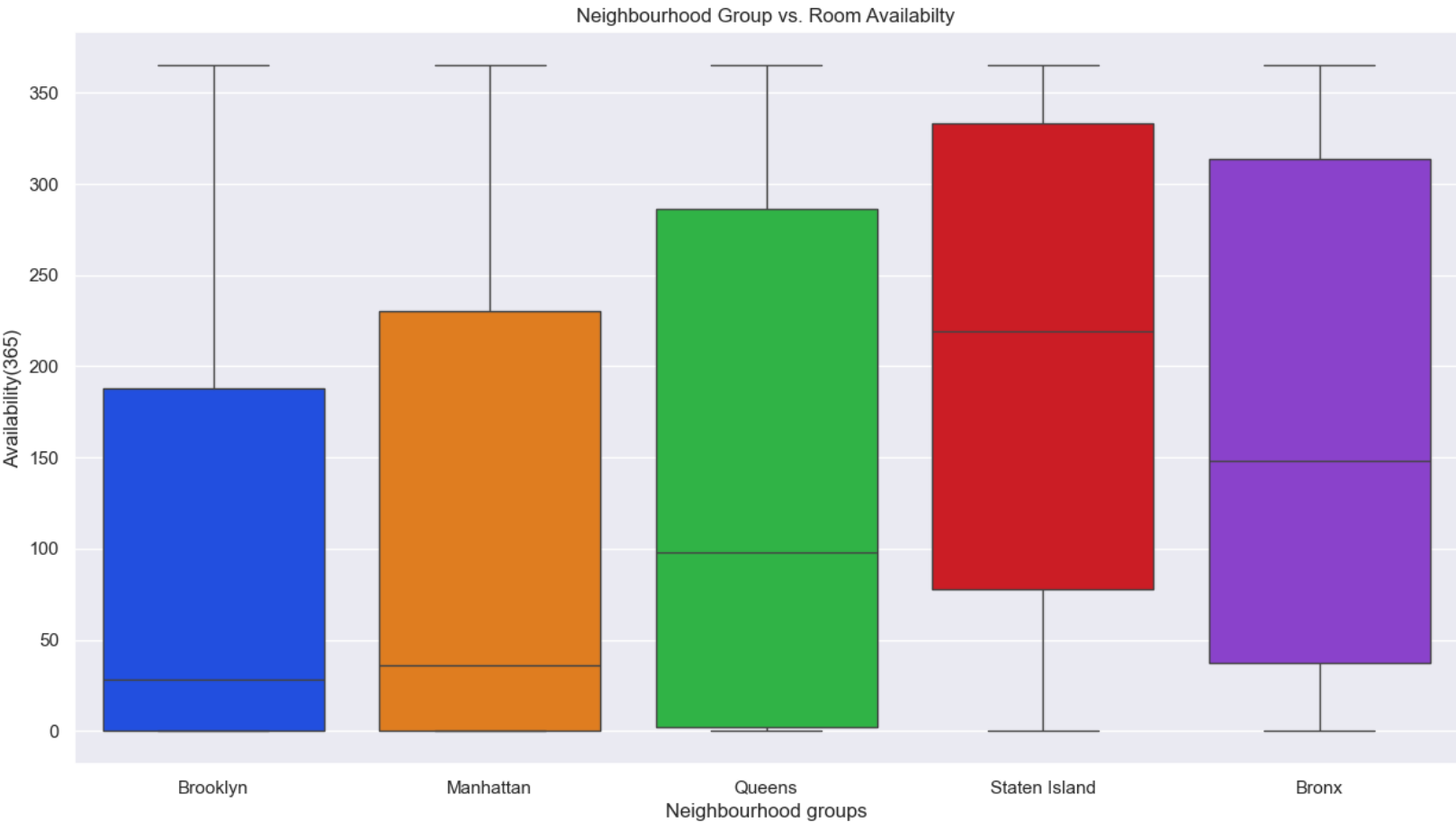
### The average price each neighborhood groups





o Manhattan remains the most expensive area, with average listing prices exceeding $140. Brooklyn follows, with an average price of around $80.

o Queens, Staten Island and Bronx have similar average prices for their listings, indicating a more affordable market compared to Manhattan and Brooklyn.

**Deep dive into cost of living**

o Top 5 most expensive listings belongs to Manhattan (being the most expensive place of living in NYC)
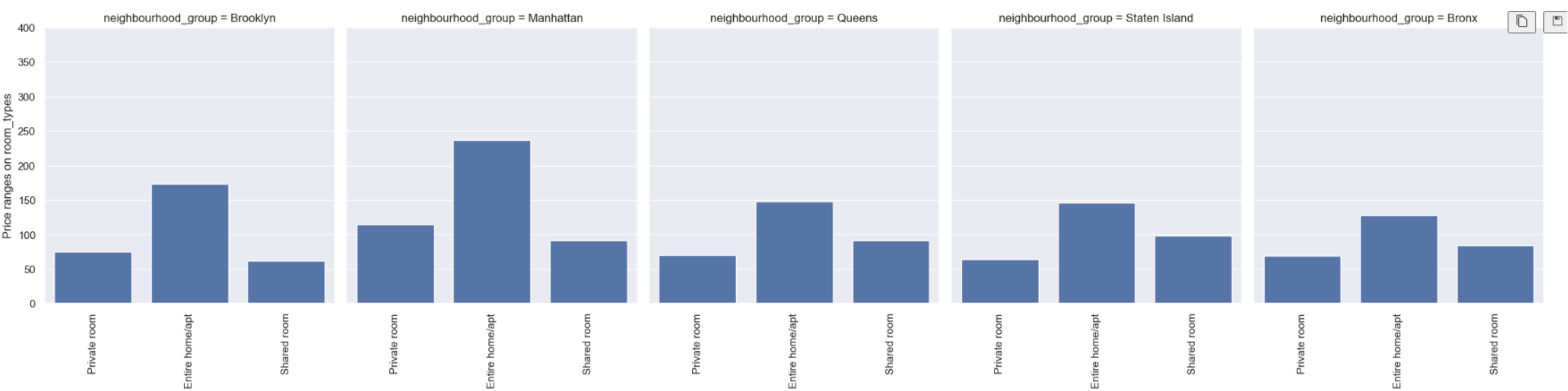o Top 5 cheapest listings mostly comes from Brooklyn, Queens and all are private rooms

*<Top 5 most expensive>*

| | name | neighbourhood_group | neighbourhood | host_name | room_type |
|---|---|---|---|---|---|
| 38498 | LUXURIOUS 5 bedroom, 4.5 bath home | Manhattan | Upper West Side | Lisa | Entire home/apt |
| 48304 | Next to Times Square/Javits/MSG! Amazing 1BR! | Manhattan | Hell's Kitchen | Rogelio | Entire home/apt |
| 46533 | Amazing Chelsea 4BR Loft! | Manhattan | Chelsea | Viberlyn | Entire home/apt |
| 30824 | Designer's Beautiful 2BR Apartment in NOLITA/SOHO | Manhattan | Nolita | Ilo And Richard | Entire home/apt |
| 22992 | Modern Townhouse for Photo, Film & Daytime Events | Manhattan | Upper West Side | Lanie | Entire home/apt |

*<Top 5 most cheapest>*

| | name | neighbourhood_group | neighbourhood | host_name | room_type |
|---|---|---|---|---|---|
| 12516 | cute and cozy room in brooklyn | Brooklyn | Bedford-Stuyvesant | Ornella | Private room |
| 7864 | Comfortable and Large Room | Brooklyn | Flatbush | Kay | Private room |
| 29967 | Large bed room share bathroom | Queens | Elmhurst | Cha | Private room |
| 39100 | 15 minutes From Times Square!! | Manhattan | Washington Heights | Ari | Private room |
| 28700 | Cozy room in Loft Apartment - Brooklyn | Queens | Ridgewood | Estefani | Private room |

*<Top 5 neighborhoods with respect to average price/day>*

| | neighbourhood_group | neighbourhood | price |
|---|---|---|---|
| 188 | Staten Island | Fort Wadsworth | 800.000000 |
| 88 | Brooklyn | Sea Gate | 487.857143 |
| 34 | Bronx | Riverdale | 442.090909 |
| 121 | Manhattan | Tribeca | 410.531429 |
| 163 | Queens | Neponsit | 274.666667 |

*<Room_types vs price on different neighborhood groups>*

o Room Type Popularity: Entire homes/apartments are the most sought-after, indicating a preference for privacy and space
o Manhattan has the highest price for room types with Entire home/apt ranging to nearly 240 USD/night, followed by Private room with 110 USD/night. And it's obvious being the most expensive place to live

# Insights

## On an average for how many nights people stayed in each room_types

People generally prefer staying in an entire home or apartment, with an average duration of over 8 nights. In contrast, guests who choose shared rooms typically stay for an average of 6 to 7 nights.



Min Stays in different room types listed on Airbnb
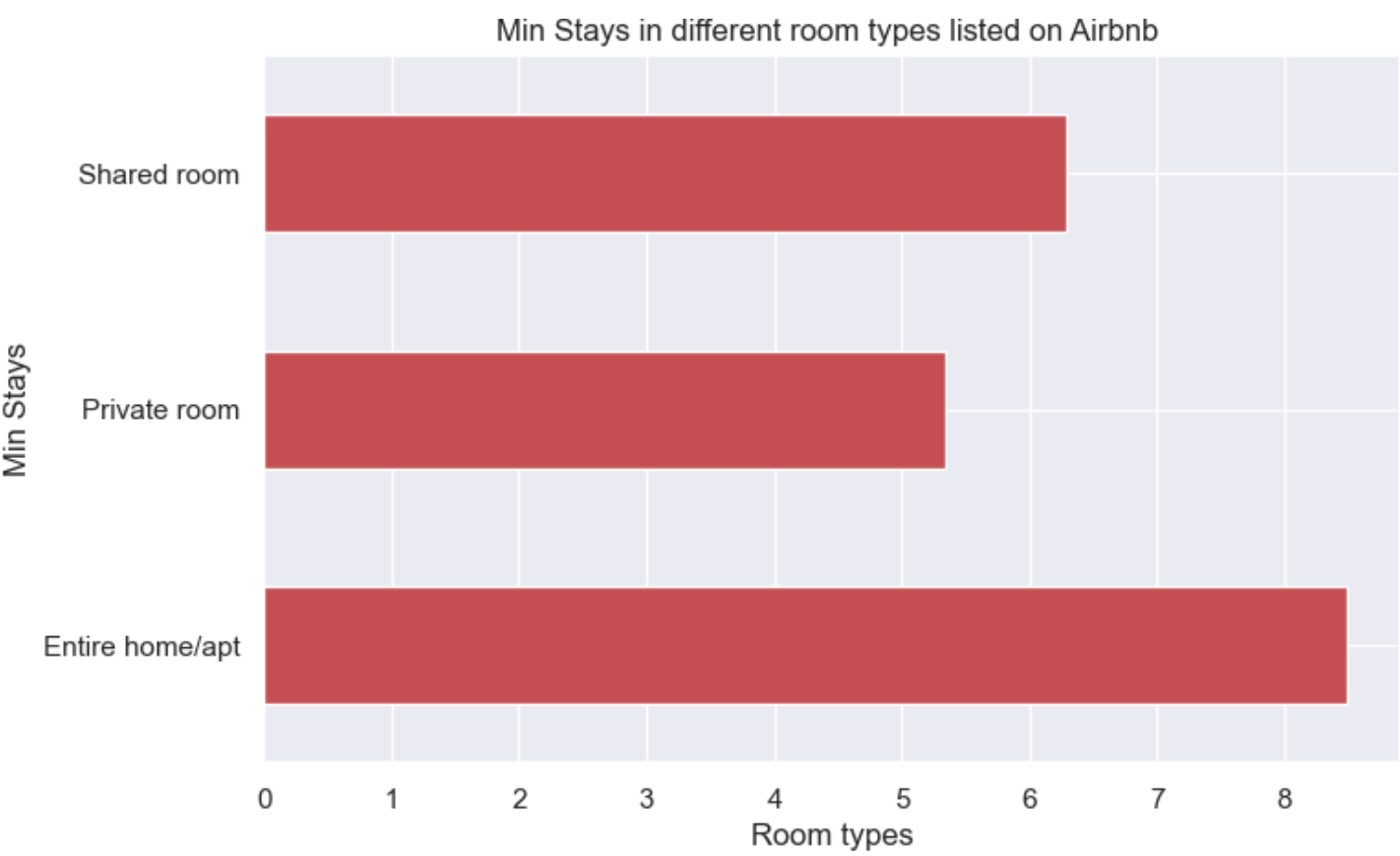
## Examine how monthly reviews differ by room types across each neighborhood group

Private rooms received the highest number of reviews per month, with Manhattan leading at over 50 reviews monthly, followed closely by other areas. For the Entire home/apt category, both Manhattan and Queens received the most reviews. In contrast, shared rooms garnered significantly fewer reviews, primarily from Staten Island, followed by the Bronx.



Most Reviewed room_types in each Neighbourhood Groups
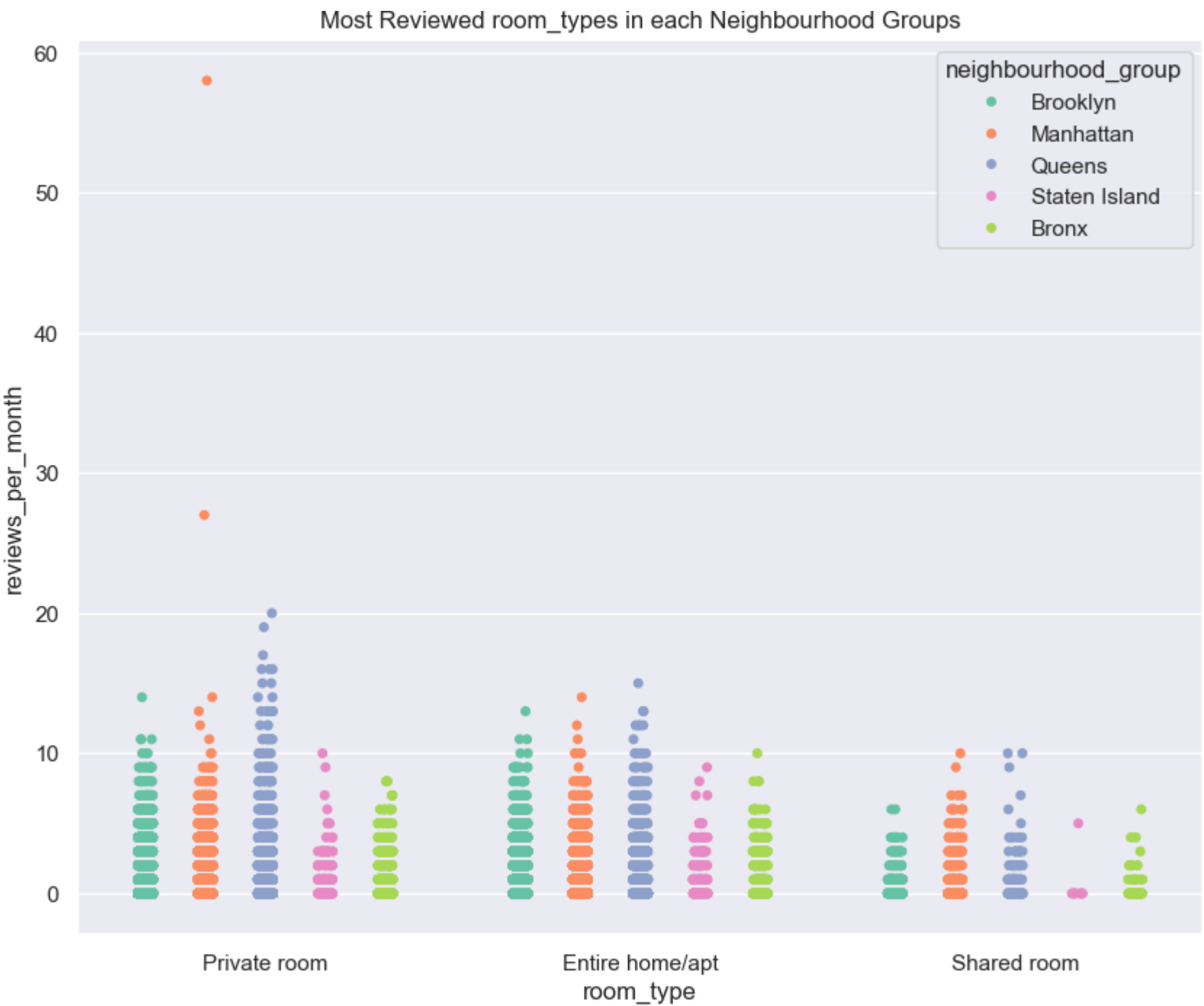
Distribution of type of rooms by longitude and latitude


Distribution of type of rooms across NYC
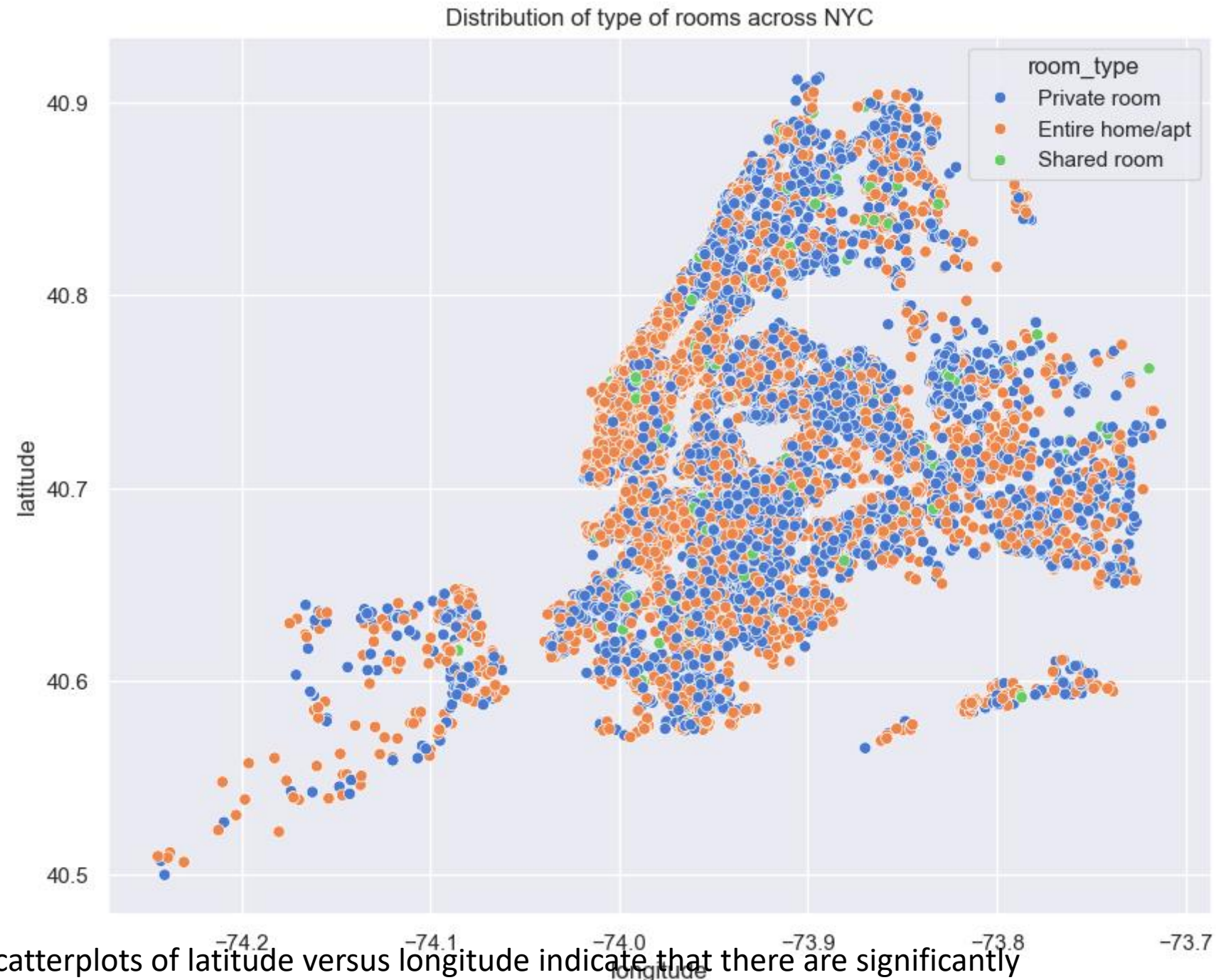


o scatterplots of latitude versus longitude indicate that there are significantly fewer Shared rooms available throughout NYC compared to Private rooms and Entire homes/apartments.

o 90 percent of the listings on Airbnb consist of either private rooms or entire homes/apartments, with very few guests choosing shared rooms.
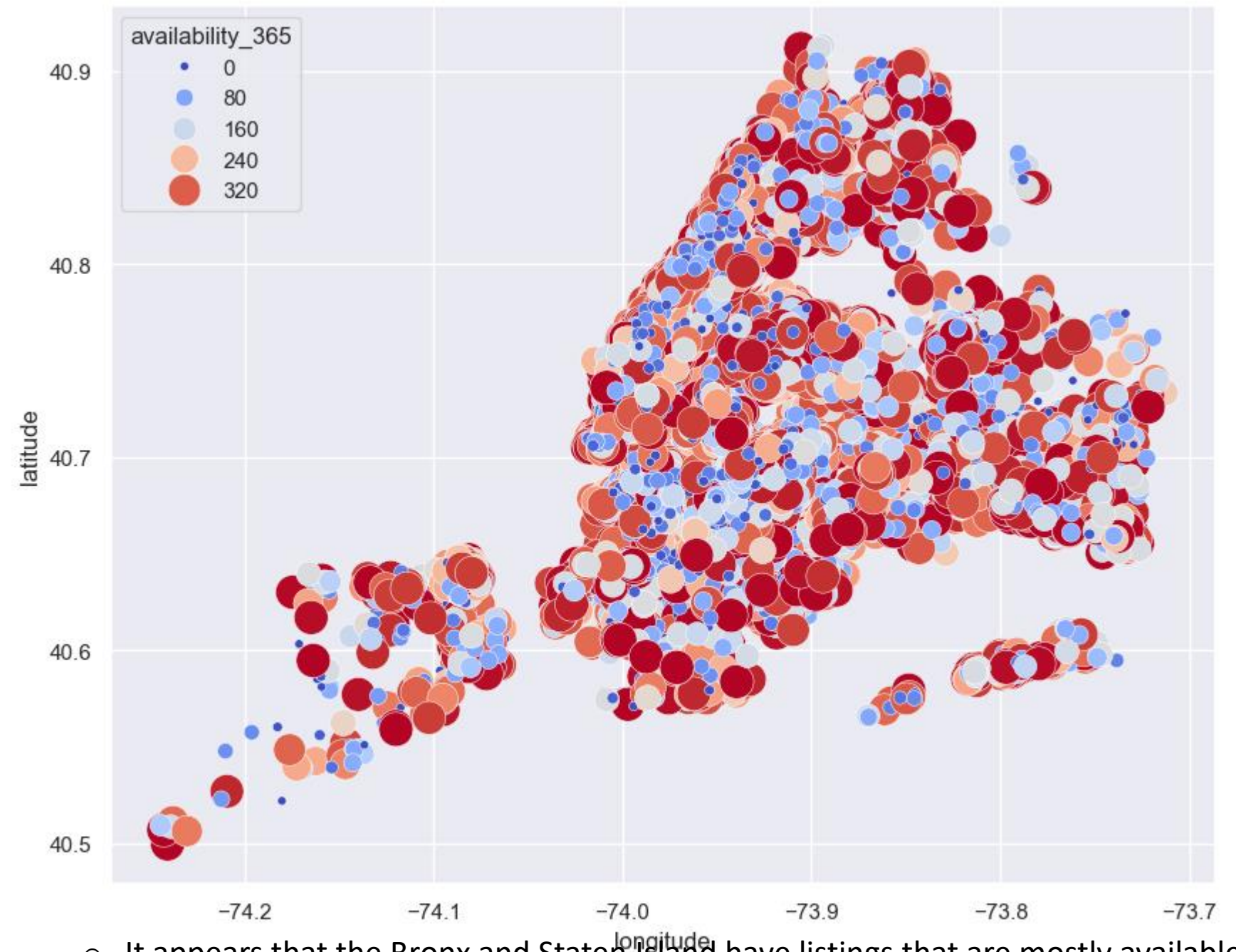
o It appears that the Bronx and Staten Island have listings that are mostly available throughout the year, likely because their prices are lower compared to other boroughs like Manhattan, Brooklyn, and Queens.

# Optimization

o The data collection team should gather information on review scores to enhance subsequent analyses.

o Should collect more details about whether the review is positive or negative

o Use diverse visuals to enhance stakeholder experience

o A clustering machine learning model can be developed to identify groups of similar objects in datasets with two or more quantitative variables.

# Appendix

Data Source

| 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

# Appendix

## Data Methodology

o Conducted a thorough analysis of New York Airbnb Dataset.

o Data Cleaning

o Used group aggregation, pivot table and other statistical methods.

o Created charts and visualizations

o Tool: Python

## 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

o The company does not expand into new market in NYC