# AIRBNB STORYTELLING CASE STUDY

# AGENDA

- Objective
- Background
- Data Preparation and Cleaning
- Key findings
- Recommendations
- Appendix
  - Data methodology
  - Data model Assumptions

# OBJECTIVE

- Analyze Airbnb's listing data and identify which types of hosts to acquire and target specific areas for expansion.
- Identify trends in Airbnb listing performance across NYC's boroughs
- Provide recommendations for improving revenue generation based on insights derived from data analysis.

# BACKGROUND

- Airbnb has seen a significant decline in revenue over the past few months. There is a need to better understand how listings vary in performance across different boroughs and neighborhoods of NYC.
- Leaders want to understand trends in customer preferences, host acquisitions, and property performance.
- Insights are needed to increase bookings and ensure an optimal mix of listings.

# DATA PREPARATION AND CLEANING

- Importing libraries (Pandas, NumPy, Matplotlib and Seaborn) and warnings and setting the display to maximum
- Loading the csv file
- Understanding basic structure and statistical summary of the dataset
  - Structure: Using head(), shape(), info()
  - Statistical Summary: describe()
- Ensuring correct format
- Handling the missing values
- Filtering out irrelevant or outlier data
- Checking for duplicates

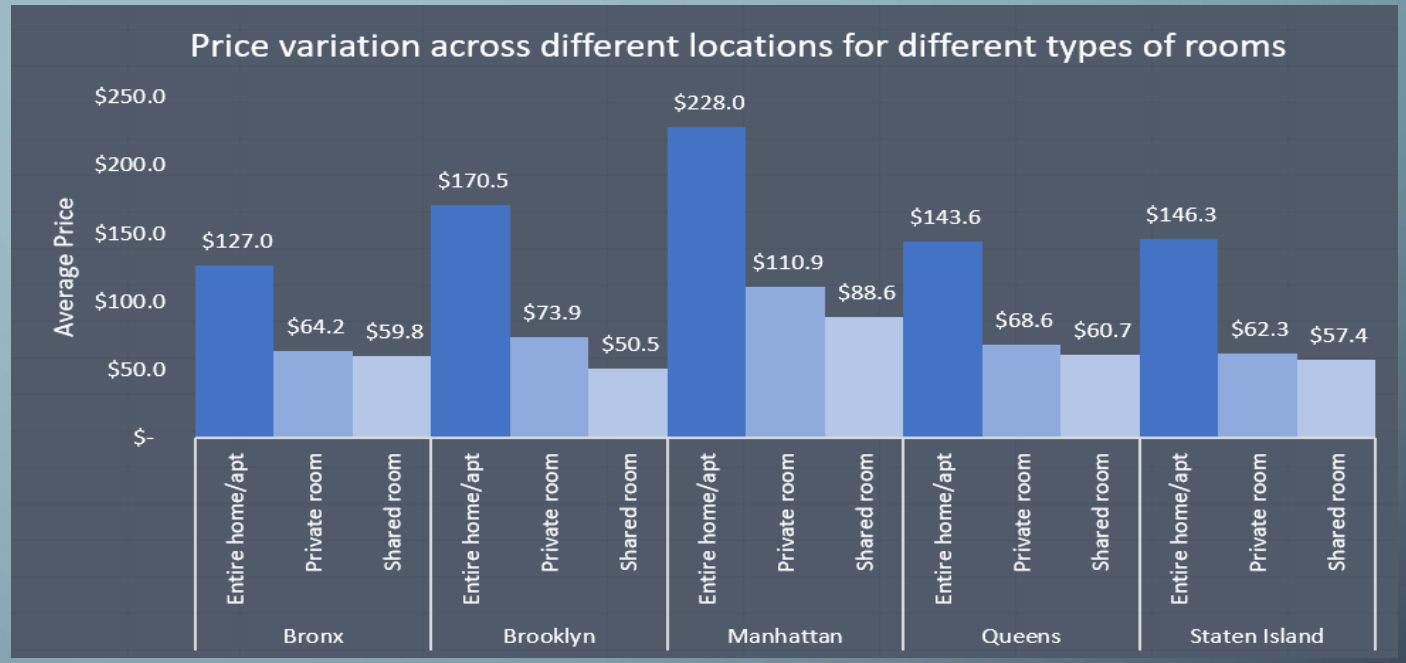
```
In [5]: airbnb.head()
```

Out[5]:

| | id | name | host_id | host_name | neighbourhood_group | neighbourhood | latitude | longitude | room_type | price | minimum_nights | number_of_revie |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 2539 | Clean & quiet apt home by the park | 2787 | John | Brooklyn | Kensington | 40.64749 | -73.97237 | Private room | 149 | 1 | |
| 1 | 2595 | Skylit Midtown Castle | 2845 | Jennifer | Manhattan | Midtown | 40.75362 | -73.98377 | Entire home/apt | 225 | 1 | |
| 2 | 3647 | THE VILLAGE OF HARLEM....NEW YORK ! | 4632 | Elisabeth | Manhattan | Harlem | 40.80902 | -73.94190 | Private room | 150 | 3 | |
| 3 | 3831 | Cozy Entire Floor of Brownstone | 4869 | LisaRoxanne | Brooklyn | Clinton Hill | 40.68514 | -73.95976 | Entire home/apt | 89 | 1 | |
| 4 | 5022 | Entire Apt: Spacious Studio/Loft by central park | 7192 | Laura | Manhattan | East Harlem | 40.79851 | -73.94399 | Entire home/apt | 80 | 10 | |

```
In [6]: airbnb.shape
Out[6]: (48895, 16)
```

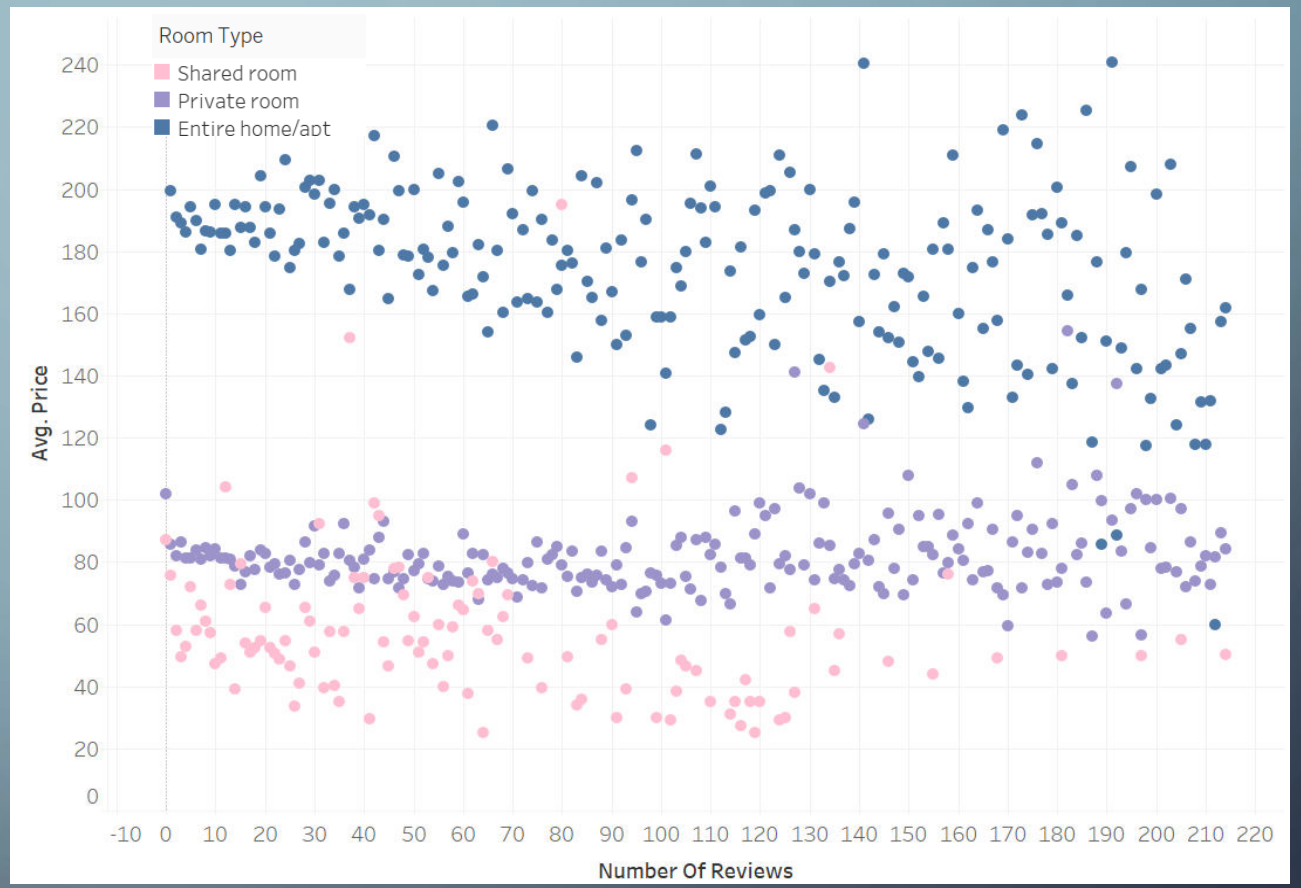
# AVERAGE PRICE PER NIGHT FOR DIFFERENT ROOM TYPES ACROSS NYC BOROUGHS




- Across all neighborhood groups, ‘Entire home/apt’ **consistently has the highest average price** compared to ‘Private room’ and ‘Shared room’ types, suggesting a customer **preference for exclusive** usage.
- Manhattan’s pricing is higher across all room types, probably due to highest listings available.
- For Bronx, Queens and Staten Island the price variation across ‘Private Room’ and Shared room is not much significant offering **budget-friendly options** for customers.

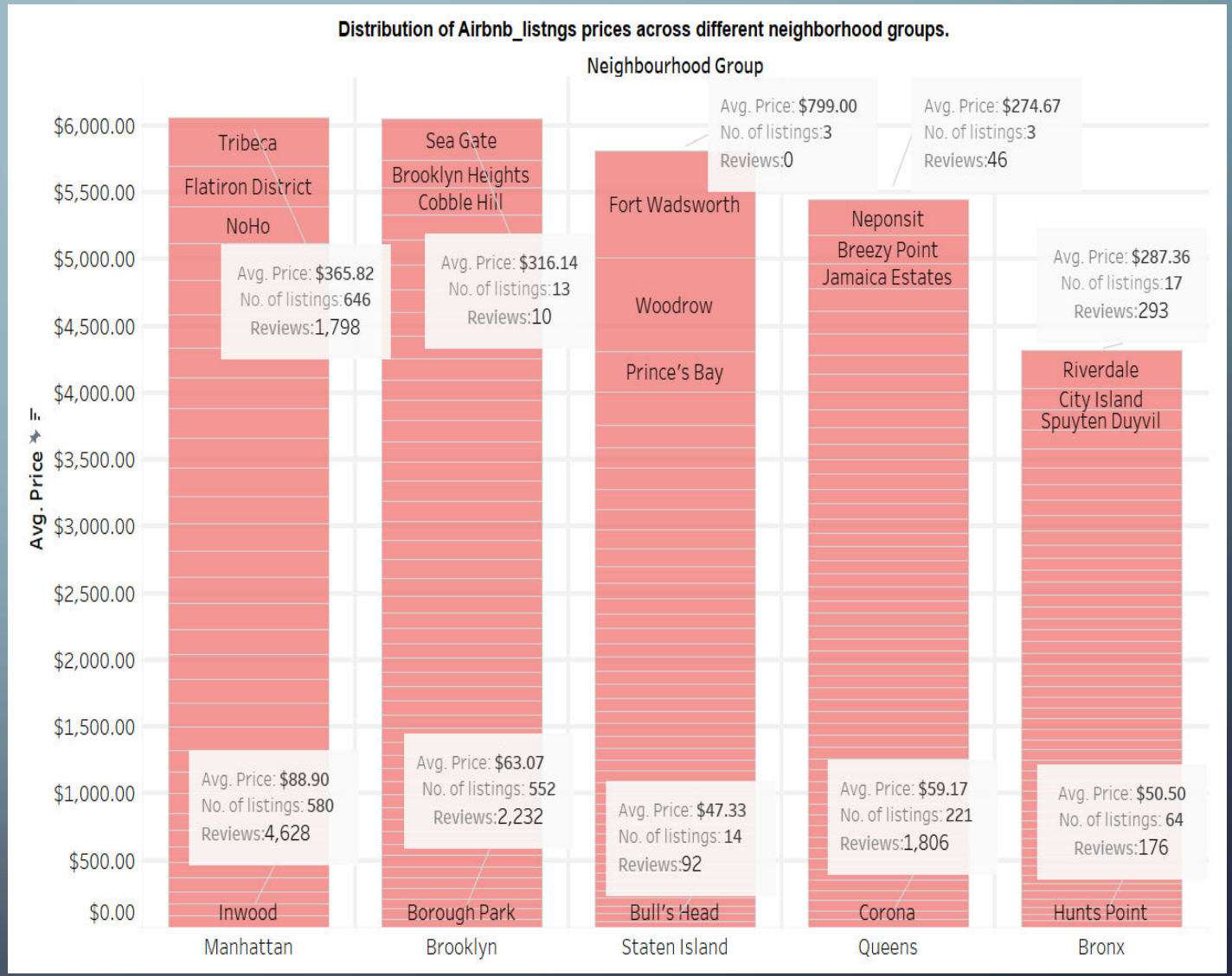
# PATTERN BETWEEN PRICING AND REVIEW W.R.T ROOM TYPE

- 'Entire homes/apt' have the **highest average prices**, ranging from around **$100 to above $240**, while private room are in the mid-range around $60 to $120, and 'Shared room' have the **lowest prices,** typically **below $80**.
- **Private rooms** show a consistent review count across different price levels, indicating **steady occupancy.**
- There is no strong correlation between prices and reviews, such as 'Entire homes/apt' have higher average prices but do not necessarily receive more reviews compared to 'Private rooms'.



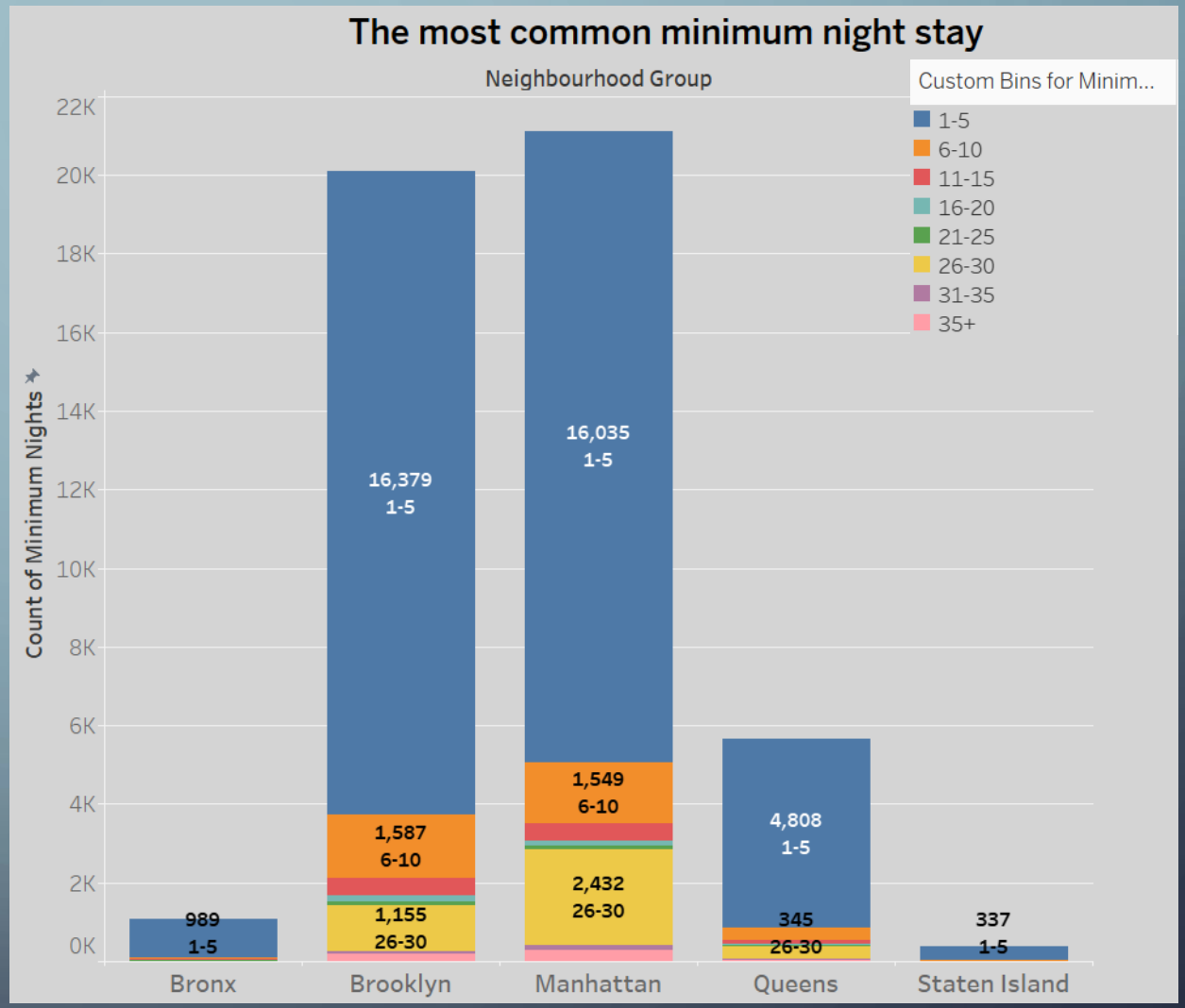
# MIN TO MAX AVERAGE PRICE DISTRIBUTION ACROSS NEIGHBORHOODS

- **Low-priced areas** like Inwood (Manhattan), Borough Park (Brooklyn), and Corona (Queens) have a high number of listings and reviews with **prices ranging from $88 to $50**.
- **Bronx** shows balanced occupancy across both high-price (Riverdale) and low-price (Hunts Point) neighborhoods, indicating **diverse demand**.
- Areas with fewer listings, such as Staten Island's high-priced neighborhoods, show no reviews, suggesting they may require more visibility or pricing adjustments.

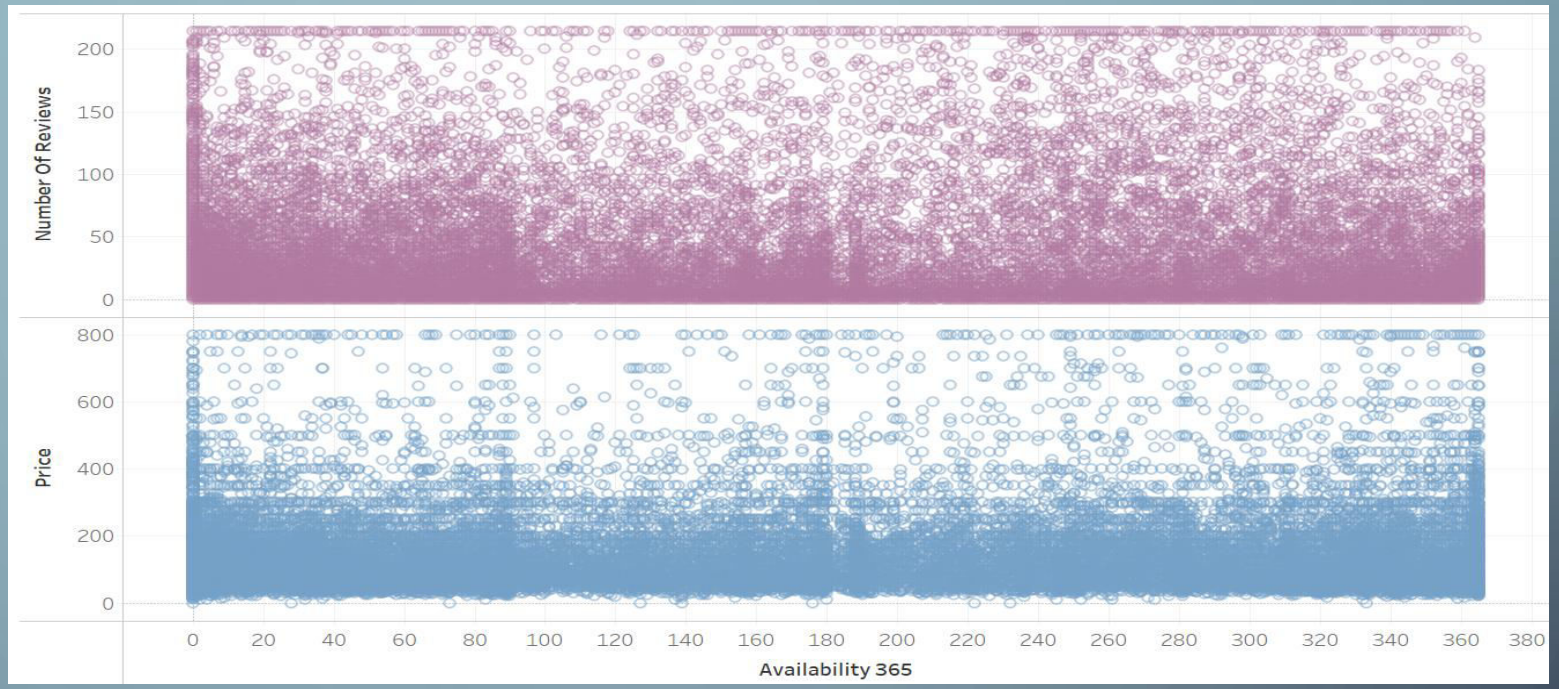


# MINIMUM NIGHTS STAY PREFERENCES OF CUSTOMERS

- The majority of listings across all neighborhood groups require 1-5 minimum nights for booking, indicating a trend toward shorter stays.
- Manhattan and Brooklyn shows more diversity in minimum night requirements compared to Bronx and Staten Island, which are more concentrated in the 1-5 nights category.
- Long- term stay options i.e. 26+ nights booking availability is in Manhattan, Brooklyn and smaller presence in Queens.



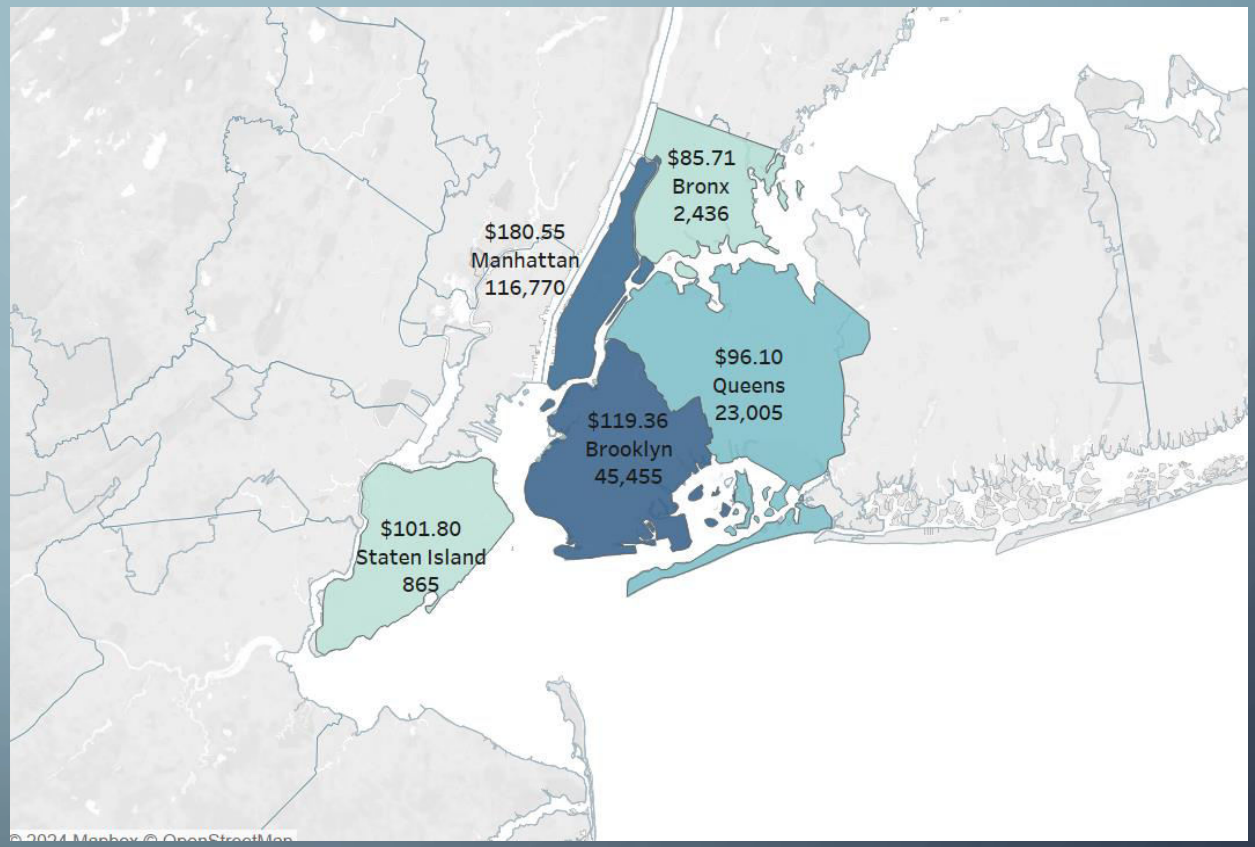
# RELATIONSHIP BETWEEN LISTING AVAILABILITY, PRICE, AND NUMBER OF REVIEWS




- The scatterplot suggests no clear relationship between price and availability thus prices do not seem to be heavily depended on the availability of the listing.
- Similarly, number of reviews does not seem to have a clear or strong relationship with how many days a listing is available.
- Listings available between 100 and 200 days per year received fewer reviews compared to those available for nearly the entire year

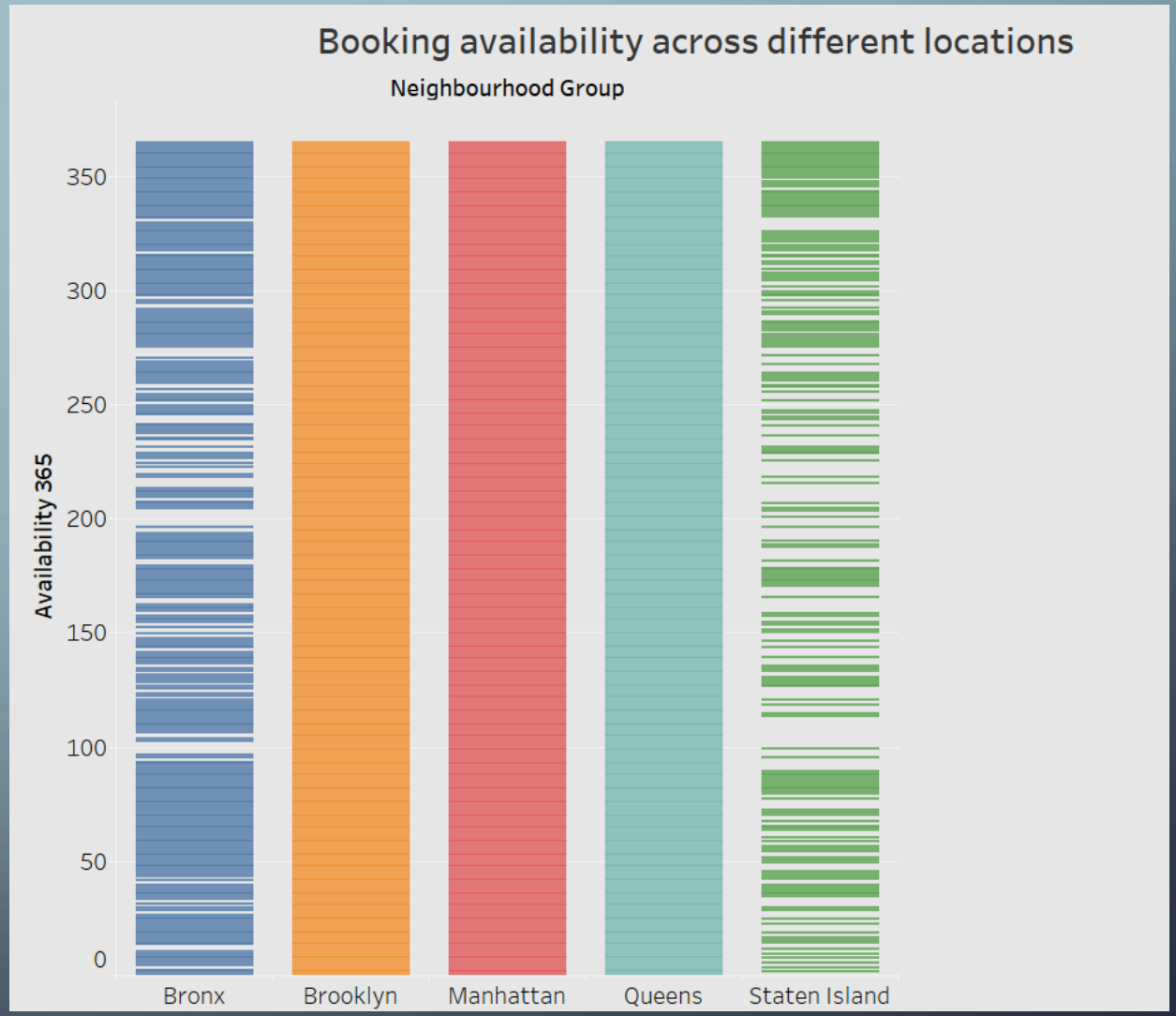
# AVERAGE PRICES AND TOTAL LISTINGS ACROSS NEIGHBORHOODS GROUPS

- Manhattan has the highest average price per night at $180.55 and also the largest number of listings (116,770).
- The Bronx shows the lowest average price at $85.71 and has fewer listings (2,436) compared to other neighborhood groups.
- Queens has a considerable number of listings (23,005) and an average price of $96.10, offering a more affordable alternative.



# BOOKING AVAILABILITY ACROSS DIFFERENT LOCATIONS

- Listings in Manhattan, Brooklyn and Queens show consistently high availability throughout the year suggesting **higher competition**.
- The Bronx and Staten Island show inconsistent availability, potentially due to fewer listings that are either available for very few days or almost the entire year.
- Despite having fewer listings, the majority listings of Staten island are keeping their booking availability for over 300 days.



# APPENDIX: DATA METHODOLOGY

- Data description:
- Data wrangling:
  - Checking duplicate entries

```
# check for duplicates
airbnb.duplicated().sum()
```

```
0
```

  - Handling the missing values by imputing and removing rows with less than 5% of missing values.

```
# filling the missing values for reviews_per_month with 0
airbnb['reviews_per_month'].fillna(0, inplace = True)
```

```
# drop the 'last_review' column
airbnb.drop('last_review', axis=1, inplace=True)
```

```
# drop rows with missing values in the 'name' column
airbnb = airbnb[~airbnb['name'].isnull()]
```

```
# check missing values
round((airbnb.isnull().sum()/len(airbnb)*100).sort_values(ascending = False), 2)
```

```
name                              0.0
host_id                           0.0
neighbourhood_group               0.0
neighbourhood                     0.0
latitude                          0.0
longitude                         0.0
room_type                         0.0
price                             0.0
minimum_nights                    0.0
number_of_reviews                 0.0
reviews_per_month                 0.0
calculated_host_listings_count    0.0
availability_365                  0.0
dtype: float64
```

| 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 |

# APPENDIX: DATA METHODOLOGY

- Identifying and handing outlier: Capping values above the 99th percentile

```python
# capping the ouliers at 99th percentile
num_cols = ['price', 'minimum_nights', 'number_of_reviews', 'reviews_per_month', 'calculated_host_listings_count',
            'availability_365']

for col in num_cols:
    percentile_99 = airbnb[col].quantile(0.99)
    airbnb[col] = np.where(airbnb[col] > percentile_99, percentile_99, airbnb[col])
```

- Creating Calculated Fields:
  - To calculate average price across the neighbourhood groups to see aggregate

avg_price_by_NG

```
{ FIXED [Neighbourhood Group] : AVG([Price]) }
```

  - Binning minimum nights

Custom Bins for Mini

```
IF [Minimum Nights] <= 5 THEN "1-5"
ELSEIF [Minimum Nights] <= 10 THEN "6-10"
ELSEIF [Minimum Nights] <= 15 THEN "11-15"
ELSEIF [Minimum Nights] <= 20 THEN "16-20"
ELSEIF [Minimum Nights] <= 25 THEN "21-25"
ELSEIF [Minimum Nights] <= 30 THEN "26-30"
ELSEIF [Minimum Nights] <= 35 THEN "31-35"
ELSE "35+"
END
```

- Supporting Tools:
  - Python
  - Excel
  - Tableau

# APPENDIX: DATA ASSUMPTIONS

- We assumed that a higher number of reviews reflects the popularity and demand for a listing, suggesting higher occupancy rates.
- We assumed that hosts' pricing and availability strategies are consistent over time.
- We assumed that higher pricing generally correlates with higher quality.
- We assumed that historical data on reviews and occupancy would be indicative of future trends.