Capstone Project-1

Air Booking Analysis

Overview of Air booking Analysis:

1. Introduction
2. Abstract
3. Problem Statement
4. Steps Involved
5. Exploratory Data Analysis
6. Statistical Tests
7. Conclusion

Introduction:

- Airbnb is an online marketplace for lodging, principally homestays for vacation rentals, and tourism activities based in the United States. Airbnb does not own any of the houses advertised; instead, it makes money from commissions on each booking.
- Airbnb is a database of accommodation bookings. This data is only available for New York City from 2011 to 2019.
- We can grasp all of the features by exploring the data. The objective is to look into the data and then evaluate it with all of the necessary information.
- We have extracted crucial insights from the data through data exploration and data analysis.
- Airbnb is an abbreviated version of AirBedandBreakfast.com, which was the company's

Abstract:

- We have extensive experience with Airbnb Booking Analysis.
- We intend to concentrate solely on New York City.
- The investigation is mostly focused on the boroughs of Brooklyn, Manhattan, Queens, Staten Island, and the Bronx.
- This data includes information about the host, lodging prices, and reviews, among other things.
- With the use of exploratory data analysis, data wrangling, visualization, and other tools, we can extract a lot of information from the analysis.

Problem Statement:

- The customer can book a room through the AirBedandBreakfast.com (airbnb website). To explore more, the customer is moving to the Queens, Bronx, Staten Island, Brooklyn, and Manhattan neighborhoods.
- The customer requires assistance in identifying a number of bookings in each borough country and guiding that customer in selecting the best place to live at an affordable cost, analyzing the room's price so that the customer can afford it, and selecting the appropriate room type based on their needs.
- The quantity of reviews, the number of reviews per month, and the availability of the rooms will all meet these criteria.
- All of these insights will be obtained through EDA and a thorough comprehension of the data.

Steps Involved:

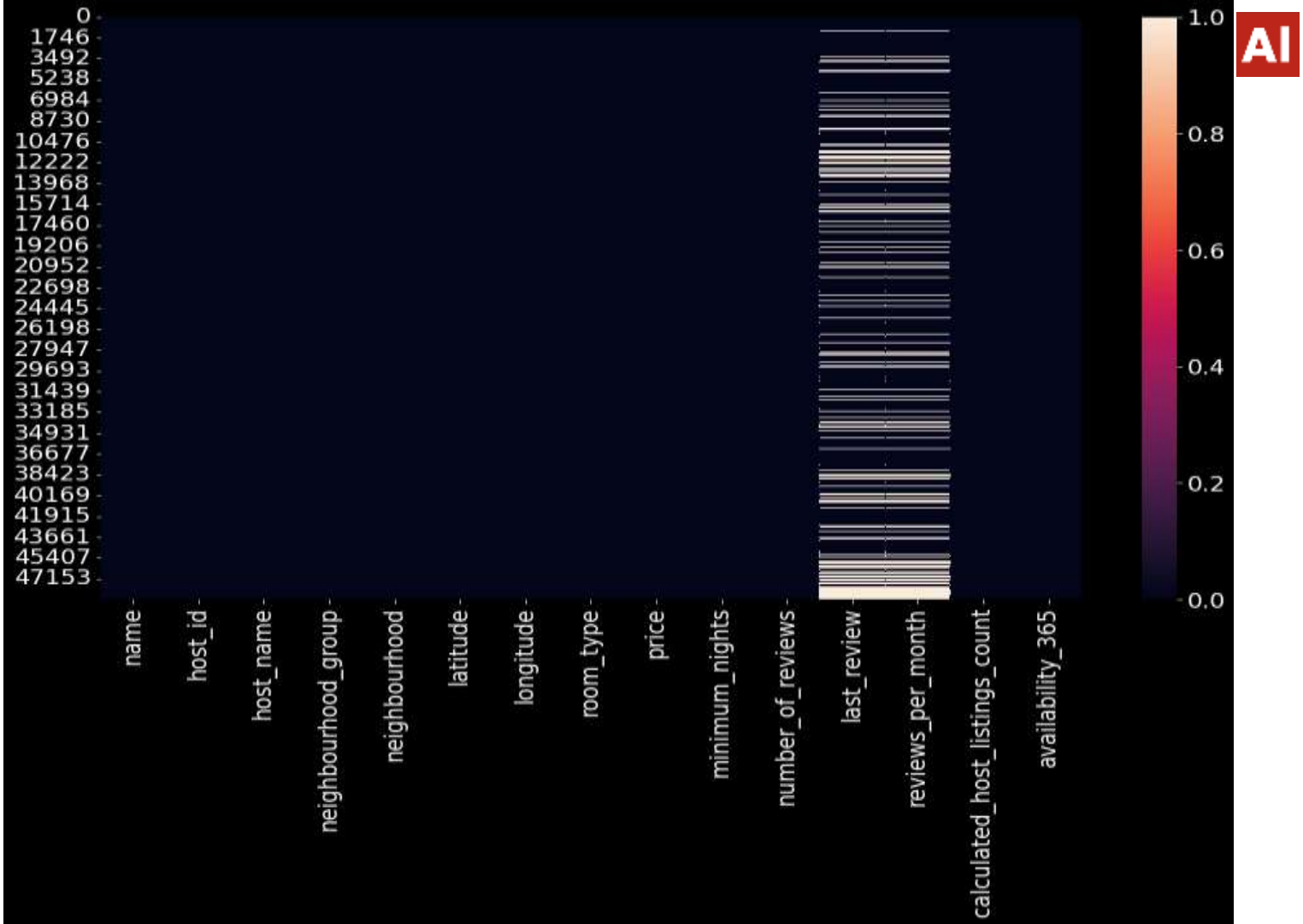
1. Data Wrangling
2. Performed Univariate and Bivariate Analysis
3. Checked correlation between each features
4. Treatment of Null values
5. Treatment of Outliers
6. Explored Visualization
7. Transformation
8. Statistical Test

Exploratory Data Analysis:

- Using the pandas package, The data is loaded. Operations performed such as casting the columns into their appropriate data types, to better comprehend the data.
- Statistical data, univariate analysis, and the accuracy of data in each column were all investigated.
- Understanding the relationship of all numerical and categorical columns. In addition, we performed Bivariate Analysis on both numerical and categorical data.
- Performed Univariate analysis for numerical columns to grasp distribution and information and see if it follows a Gaussian distribution.
- Used a variety of plots to gain a better understanding of the data and presented it in an attractive way .

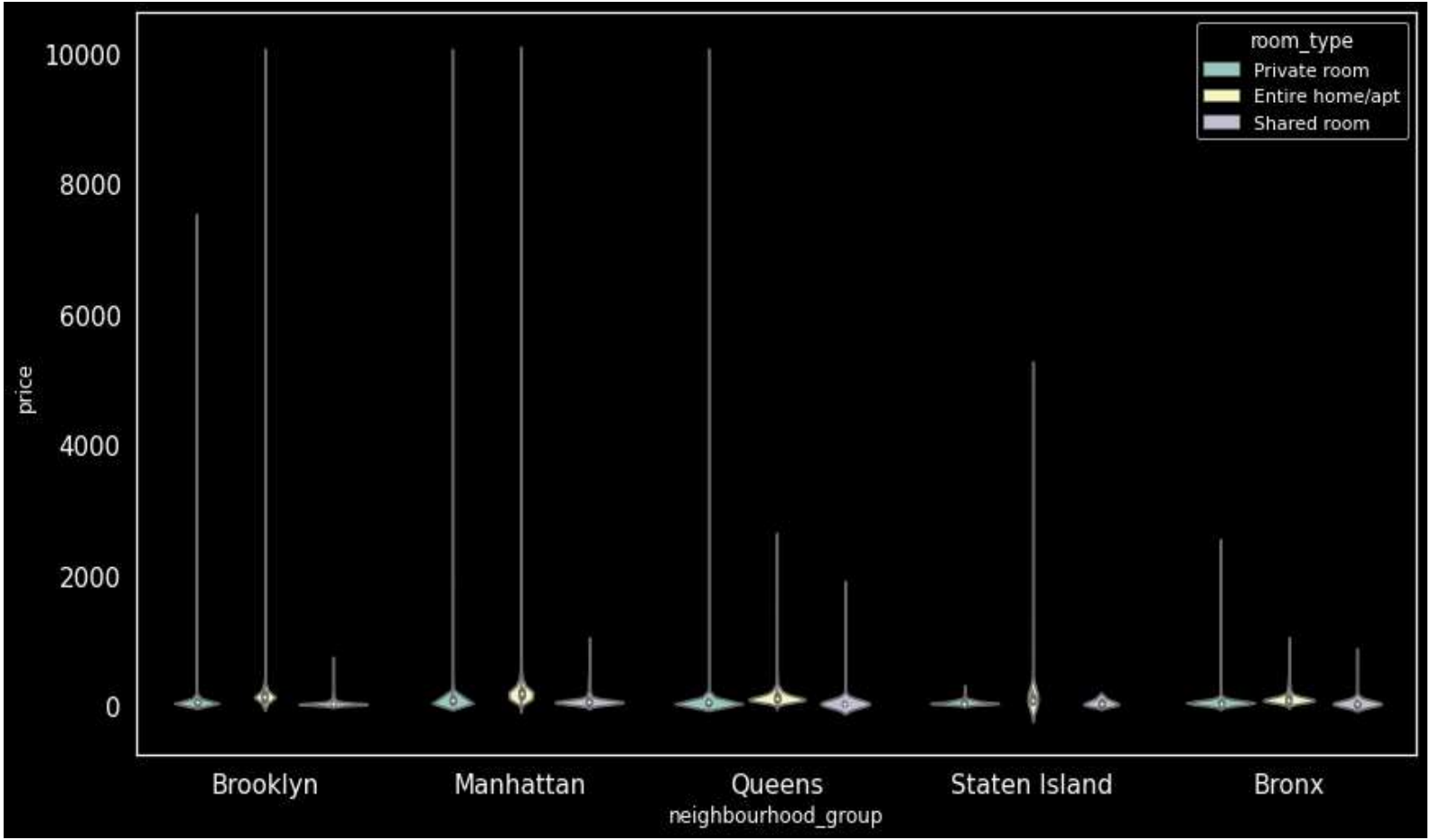
EDA contd...

Checking the count of null values



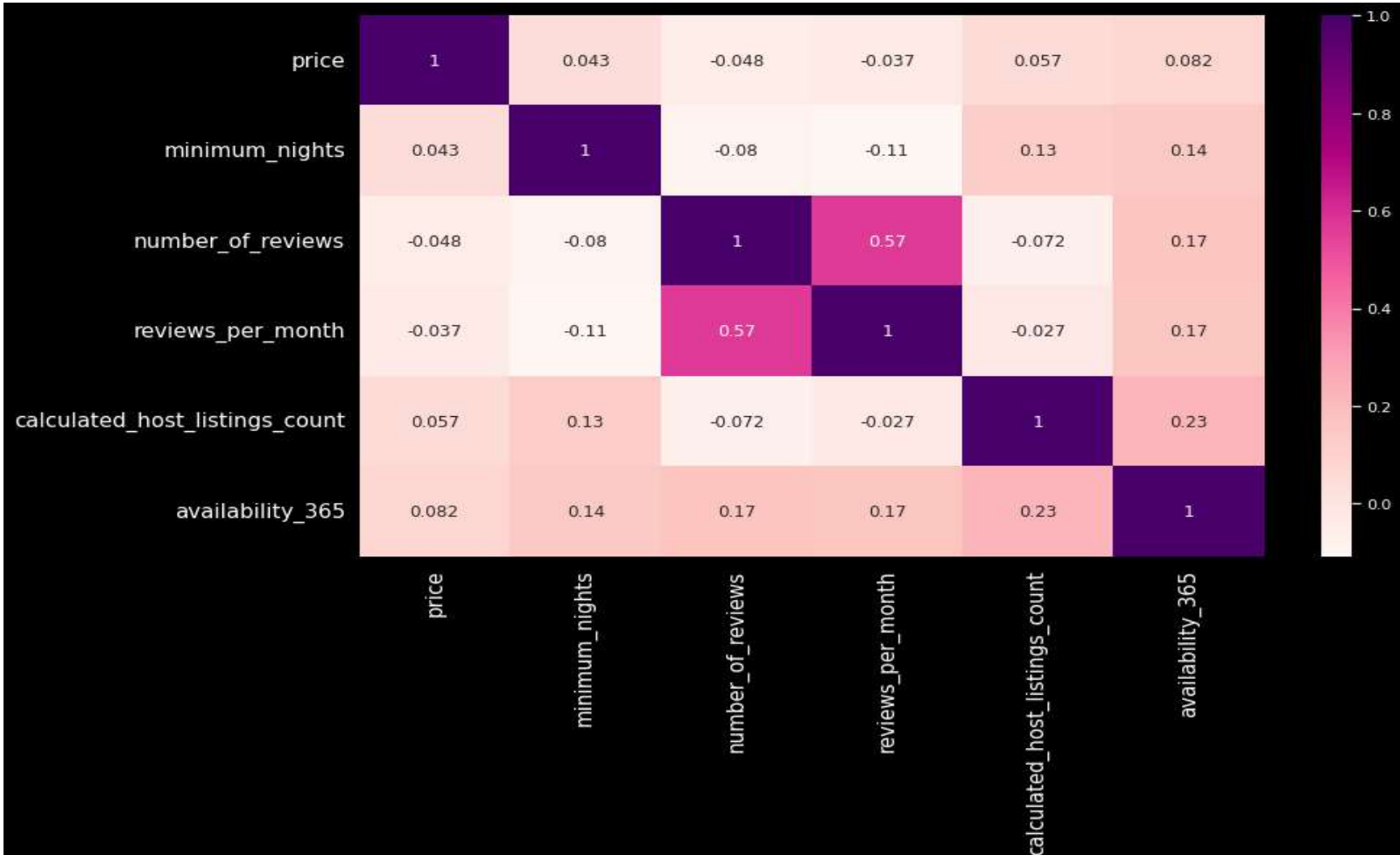
EDA contd...

Relationship between Neighborhood country vs price

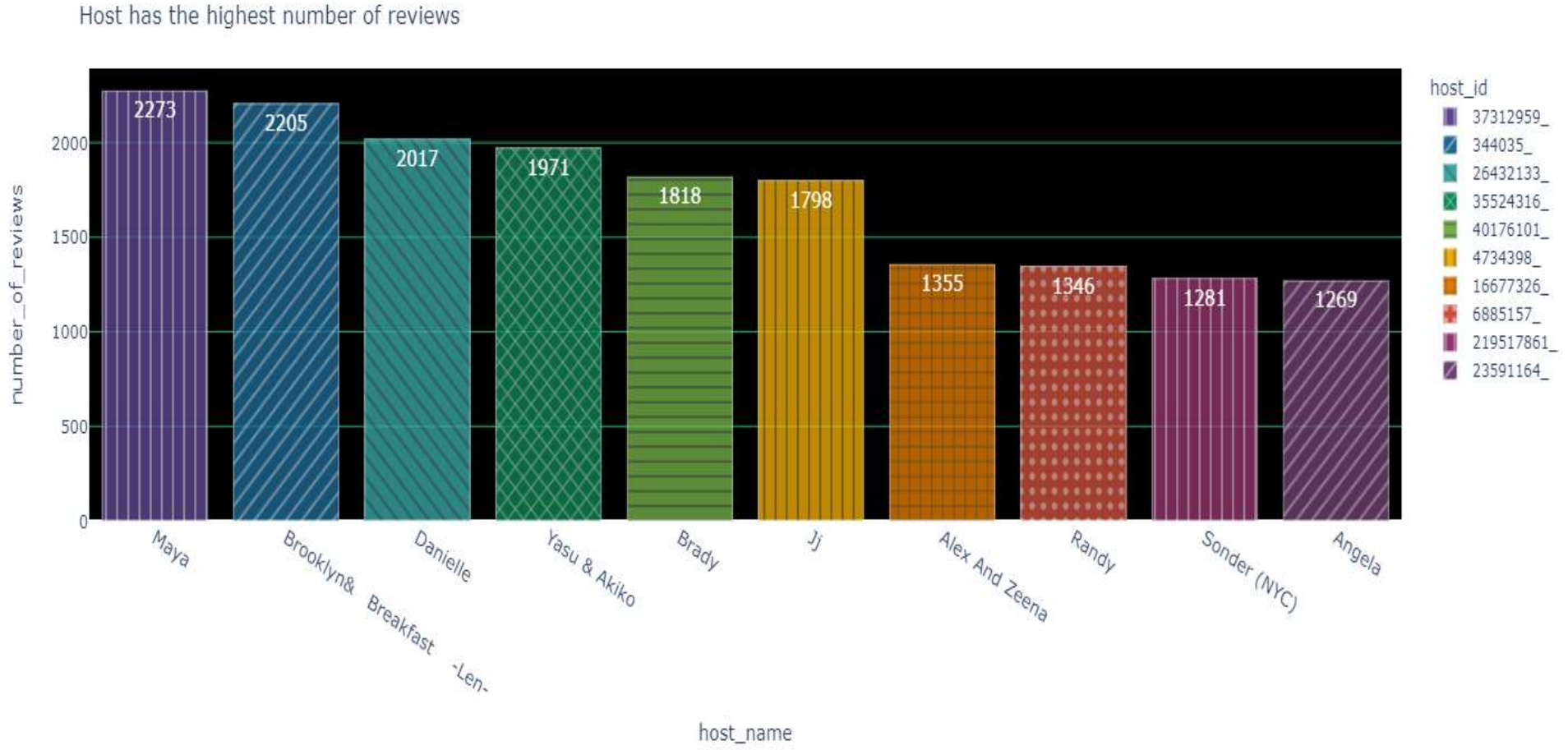


EDA contd...

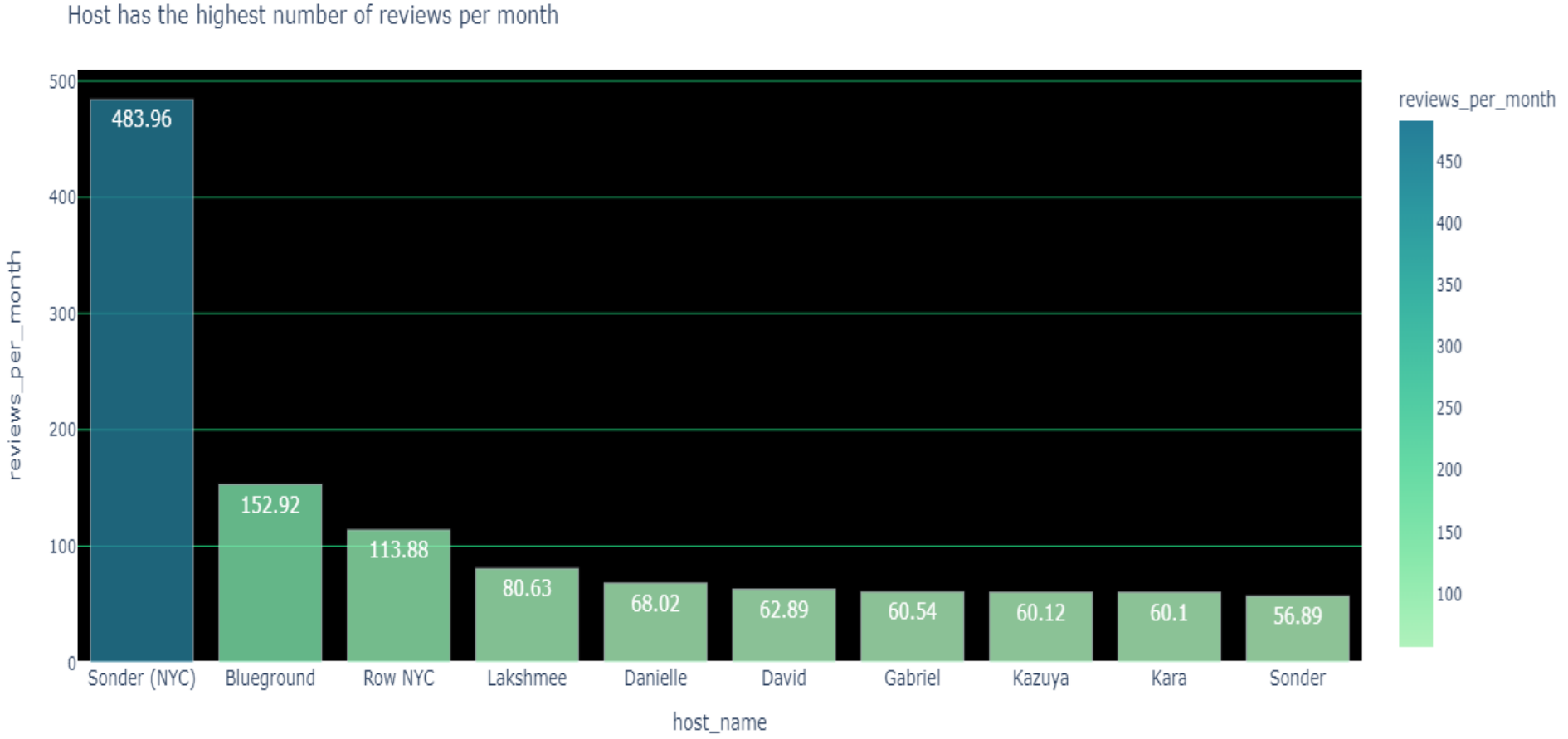
Relationship between all features



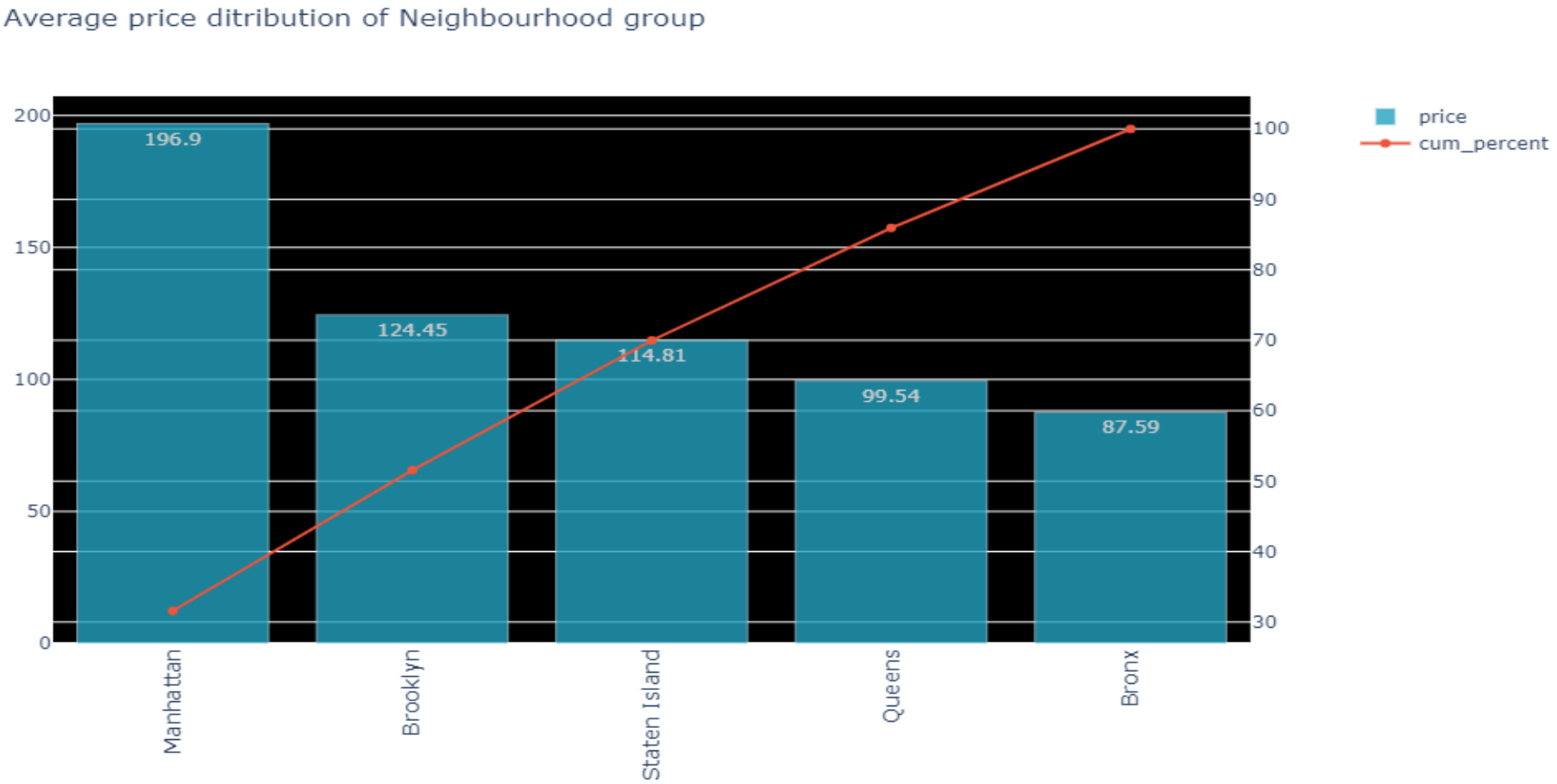
EDA contd...


Host has the highest number of reviews
number_of_reviews
0
500
1000
1500
2000
2273
2205
2017
1971
1818
1798
1355
1346
1281
1269
Maya
Brooklyn& Breakfast -Len-
Danielle
Yasu & Akiko
Brady
Jj
Alex And Zeena
Randy
Sonder (NYC)
Angela
host_name
host_id
37312959_
344035_
26432133_
35524316_
40176101_
4734398_
16677326_
6885157_
219517861_
23591164_

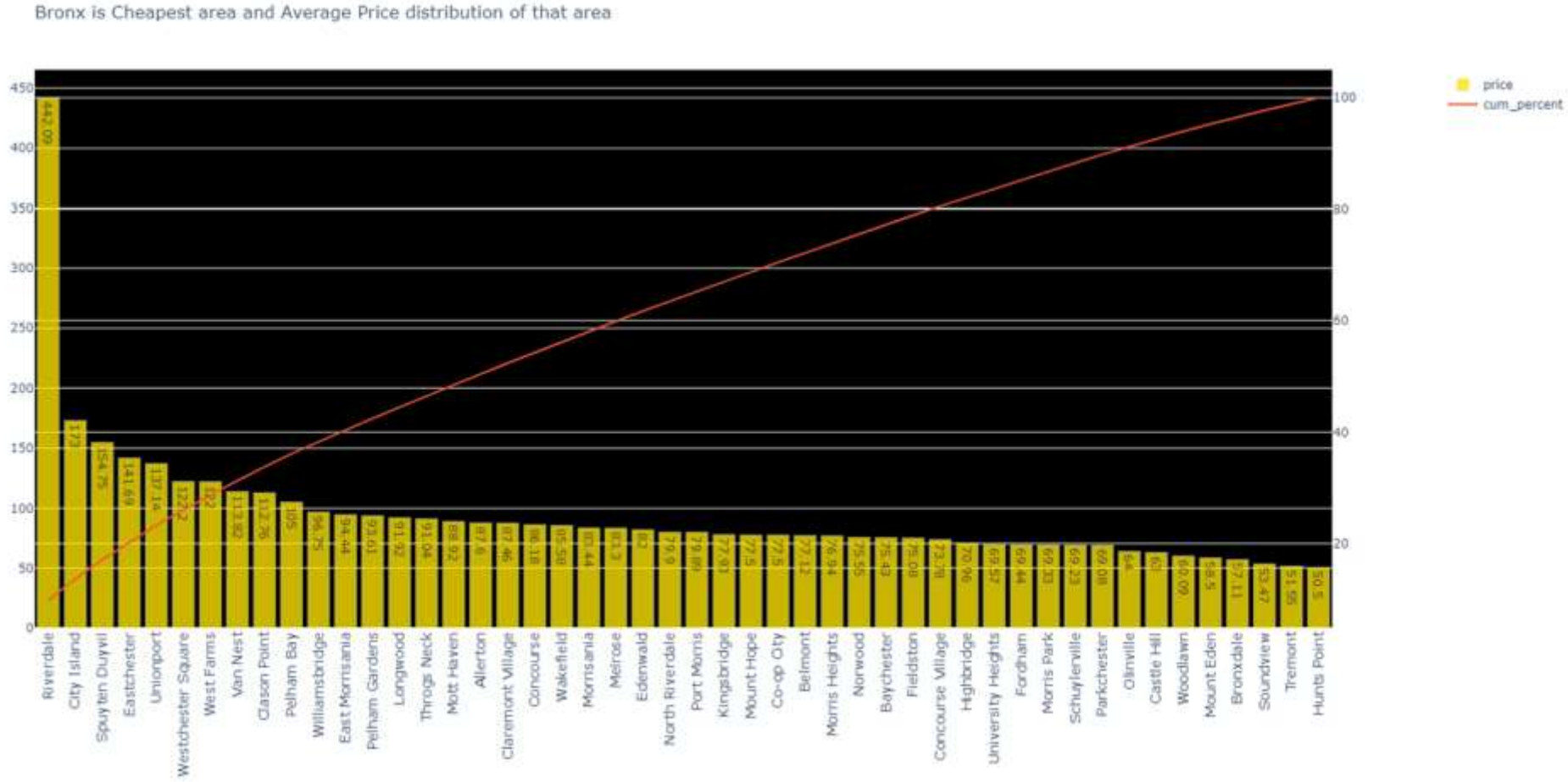
EDA contd...


Host has the highest number of reviews per month
reviews_per_month
500
400
300
200
100
0
483.96
152.92
113.88
80.63
68.02
62.89
60.54
60.12
60.1
56.89
Sonder (NYC)
Blueground
Row NYC
Lakshmee
Danielle
David
Gabriel
Kazuya
Kara
Sonder
host_name
reviews_per_month
450
400
350
300
250
200
150
100

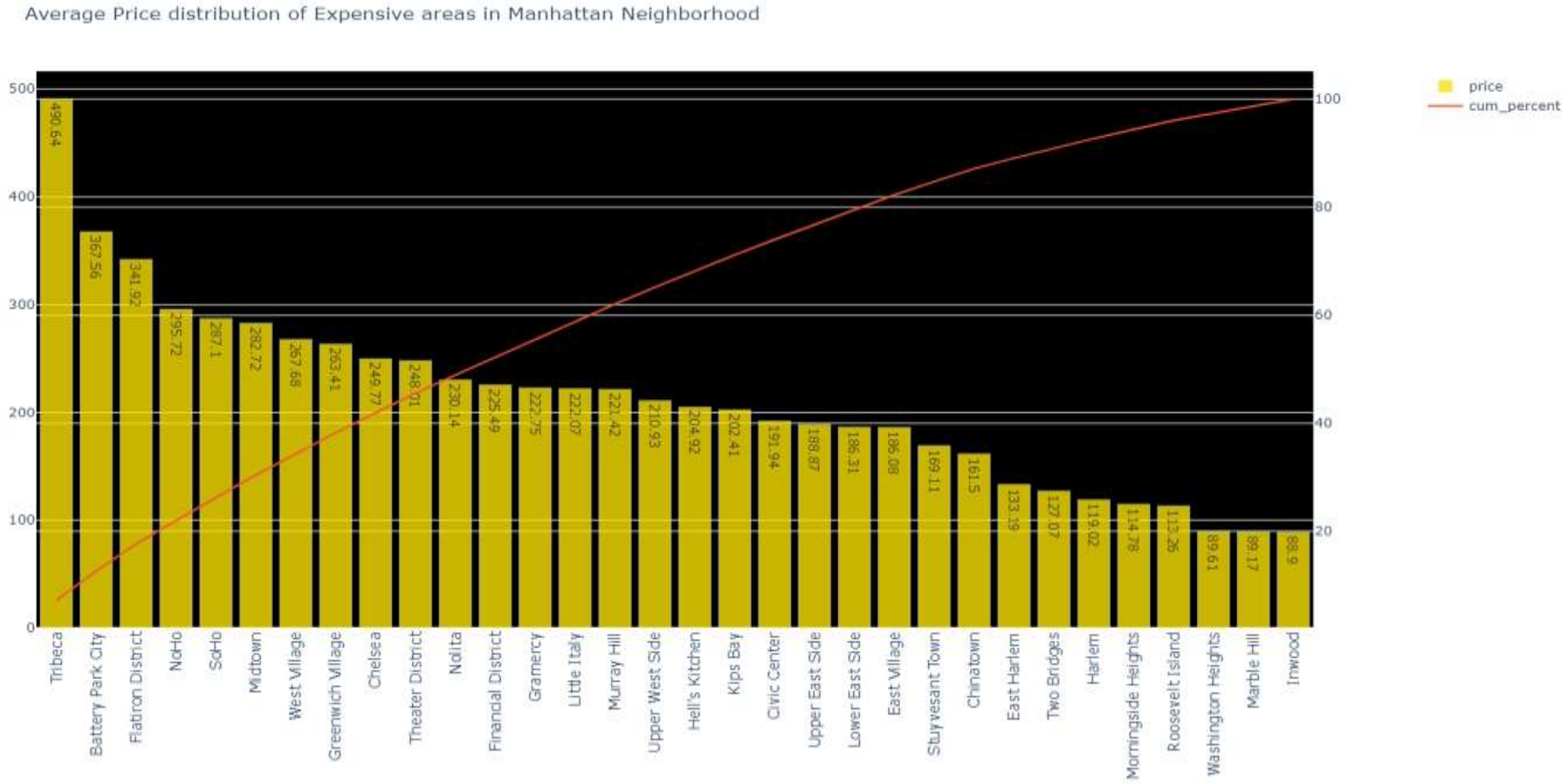
EDA contd...


Average price ditribution of Neighbourhood group
price
cum_percent
196.9
124.45
114.81
99.54
87.59
Manhattan
Brooklyn
Staten Island
Queens
Bronx
0
50
100
150
200
30
40
50
60
70
80
90
100

EDA contd...


Bronx is Cheapest area and Average Price distribution of that area
price
cum_percent
Riverdale
City Island
Spuyten Duyvil
Eastchester
Unionport
Westchester Square
West Farms
Van Nest
Olason Point
Pelham Bay
Williamsbridge
East Morrisania
Pelham Gardens
Longwood
Throgs Neck
Mott Haven
Allerton
Claremont Village
Concourse
Wakefield
Morrisania
Melrose
Elderwald
North Riverdale
Port Morris
Kingsbridge
Mount Hope
Co-op City
Belmont
Morris Heights
Norwood
Baychester
Fieldston
Concourse Village
Highbridge
University Heights
Fordham
Morris Park
Schuylerville
Parkchester
Olinville
Castle Hill
Woodlawn
Mount Eden
Bronxdale
Soundview
Tremont
Hunts Point

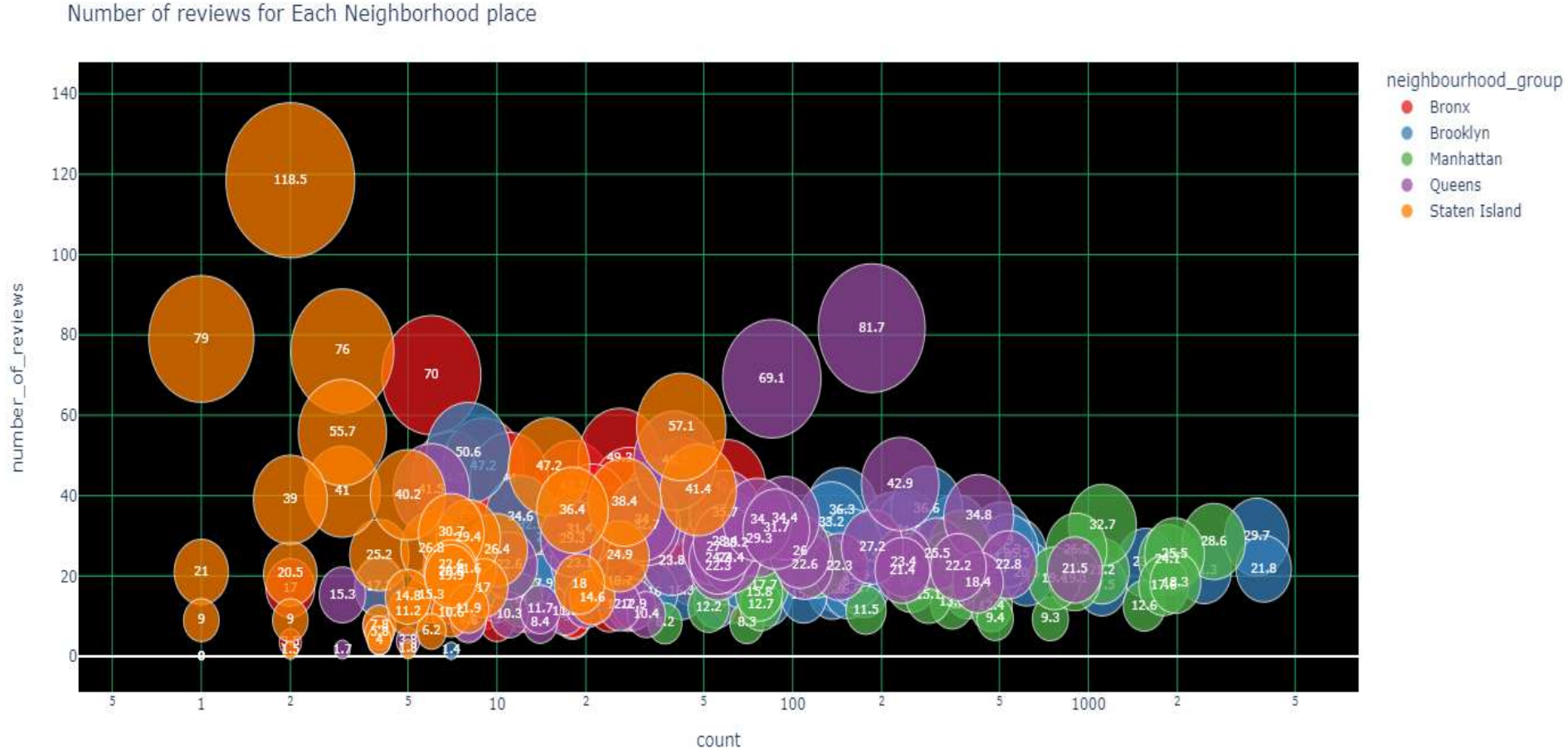
EDA contd...


Average Price distribution of Expensive areas in Manhattan Neighborhood
price
cum_percent
500
400
300
200
100
0
100
80
60
40
20
490.64
367.56
341.92
295.72
287.1
282.72
267.68
263.41
249.77
248.01
230.14
225.49
222.75
222.07
221.42
210.93
204.92
202.41
191.94
188.87
186.31
186.08
169.11
161.5
133.19
127.07
119.02
114.78
113.36
89.61
88.17
88.9
Tribeca
Battery Park City
Flatiron District
NoHo
SoHo
Midtown
West Village
Greenwich Village
Chelsea
Theater District
Nolita
Financial District
Gramercy
Little Italy
Murray Hill
Upper West Side
Hell's Kitchen
Kips Bay
Civic Center
Upper East Side
Lower East Side
East Village
Stuyvesant Town
Chinatown
East Harlem
Two Bridges
Harlem
Morningside Heights
Roosevelt Island
Washington Heights
Marble Hill
Inwood

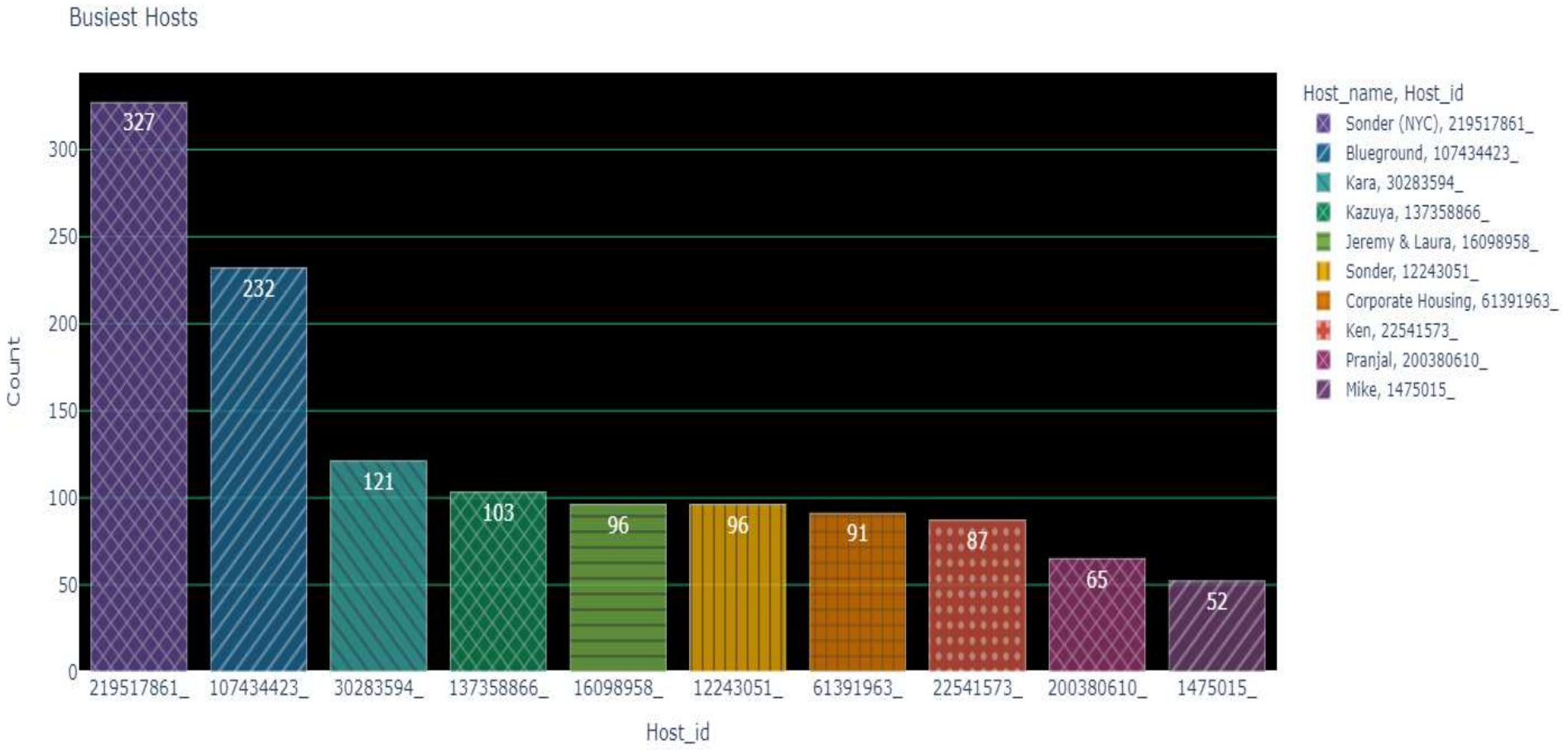
EDA contd...


Total Bookings of Borough Country vs Average Price
NYC Airbnb
Manhattan
Harlem
East Village
Upper East Side
Midtown
Upper West Side
East Harlem
Lower East Side
Financial District
Murray Hill
Kips Bay
Washington Heights
Gramercy
Hell's Kitchen
Chelsea
Chinatown
Inwood
West Village
SoHo
Nolita
NoHo
Brooklyn
Williamsburg
Bushwick
Crown Heights
Bedford-Stuyvesant
Greenpoint
Park Slope
East Flatbush
Fort Greene
South Slope
Gowanus
Flatbush
Sunset Park
Clinton Hill
Kensington
Queens
Astoria
Long Island City
Flushing
Ridgewood
Sunnyside
Jamaica
Elmhurst
Woodside
Bronx
Mean_price
800
700
600
500
400
300
200
100

EDA contd...

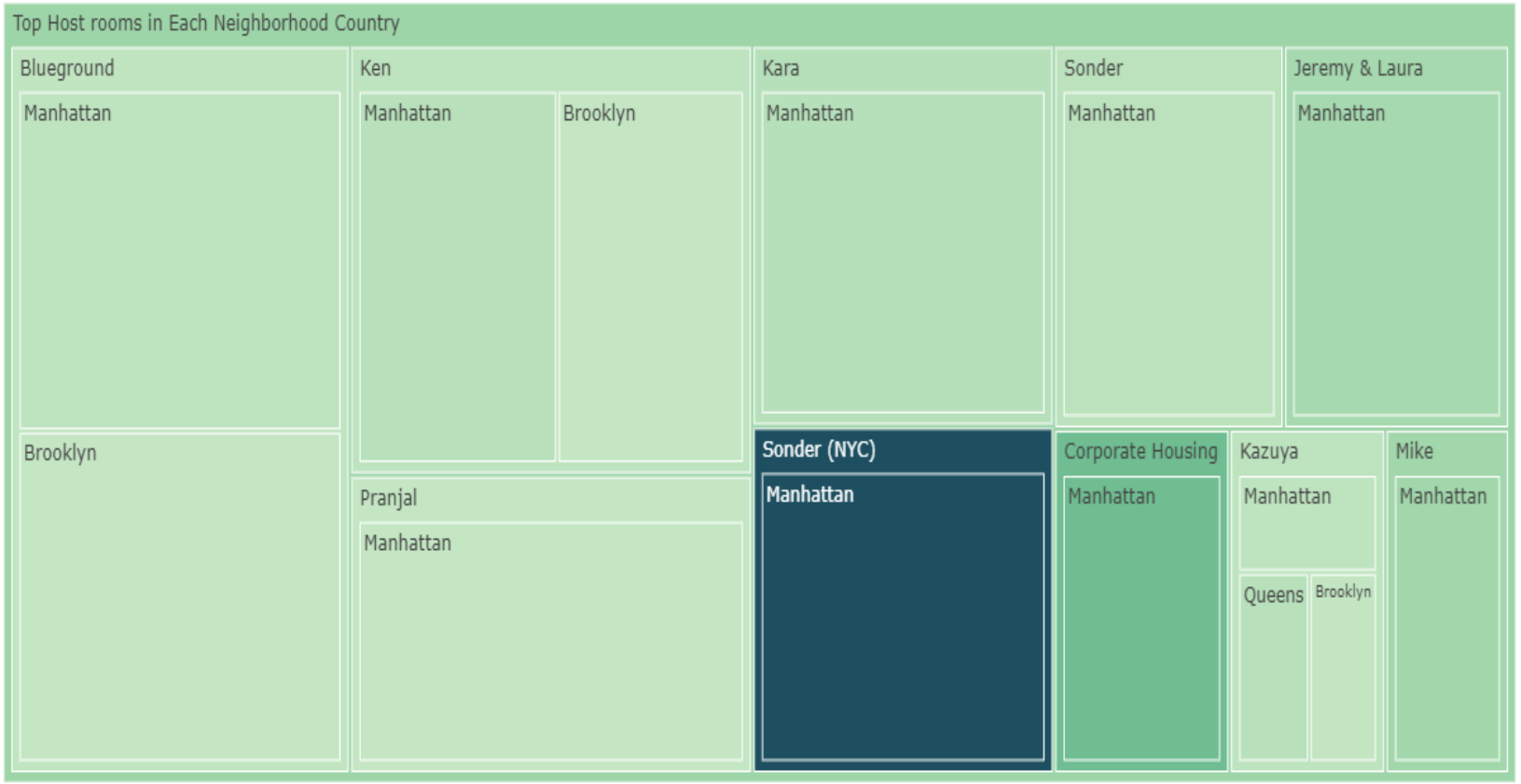

Number of reviews for Each Neighborhood place
number_of_reviews
count
neighbourhood_group
Bronx
Brooklyn
Manhattan
Queens
Staten Island

EDA contd...


Busiest Hosts
Count
0
50
100
150
200
250
300
327
232
121
103
96
96
91
87
65
52
219517861_
107434423_
30283594_
137358866_
16098958_
12243051_
61391963_
22541573_
200380610_
1475015_
Host_id
Host_name, Host_id
Sonder (NYC), 219517861_
Blueground, 107434423_
Kara, 30283594_
Kazuya, 137358866_
Jeremy & Laura, 16098958_
Sonder, 12243051_
Corporate Housing, 61391963_
Ken, 22541573_
Pranjal, 200380610_
Mike, 1475015_

EDA contd...

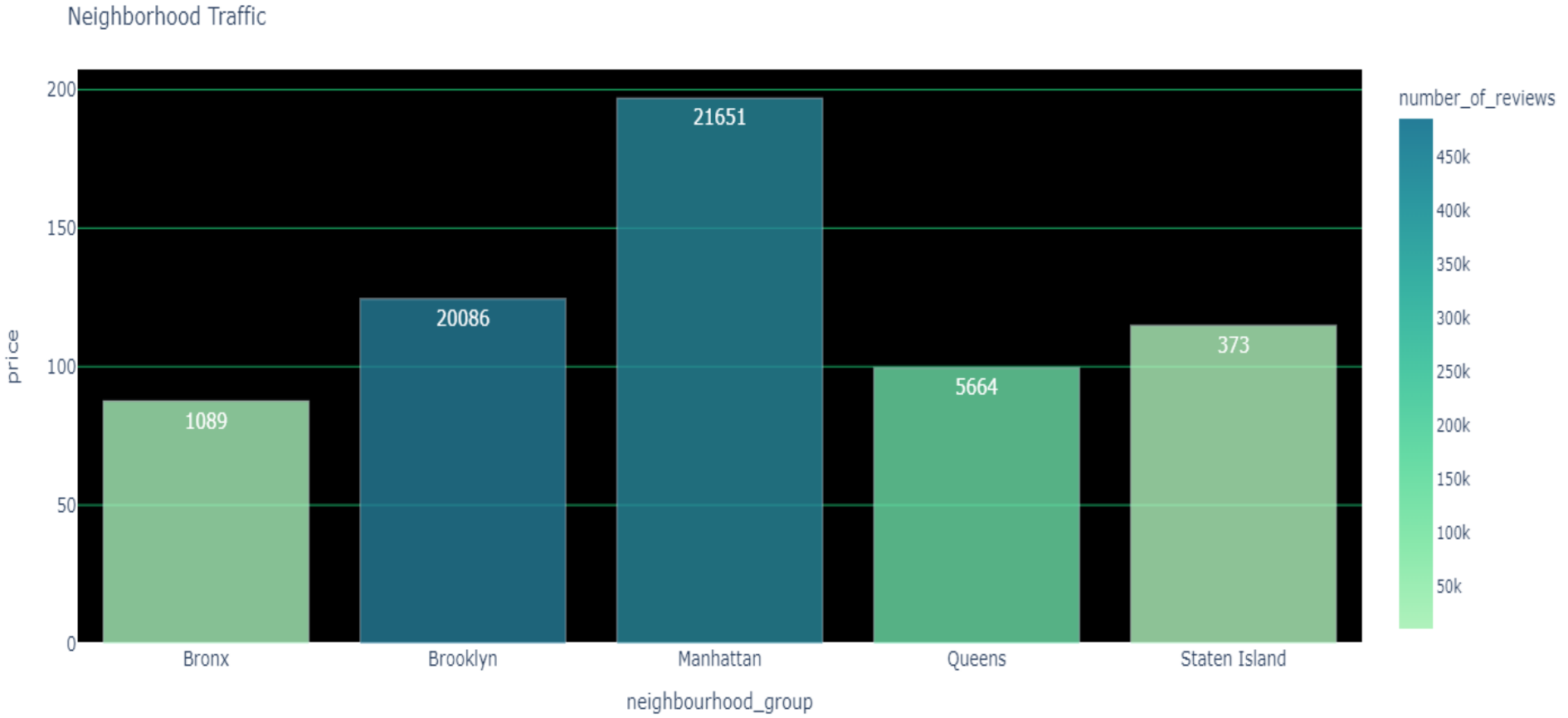
Busiest Hosts Rental place with bookings count and number of reviews



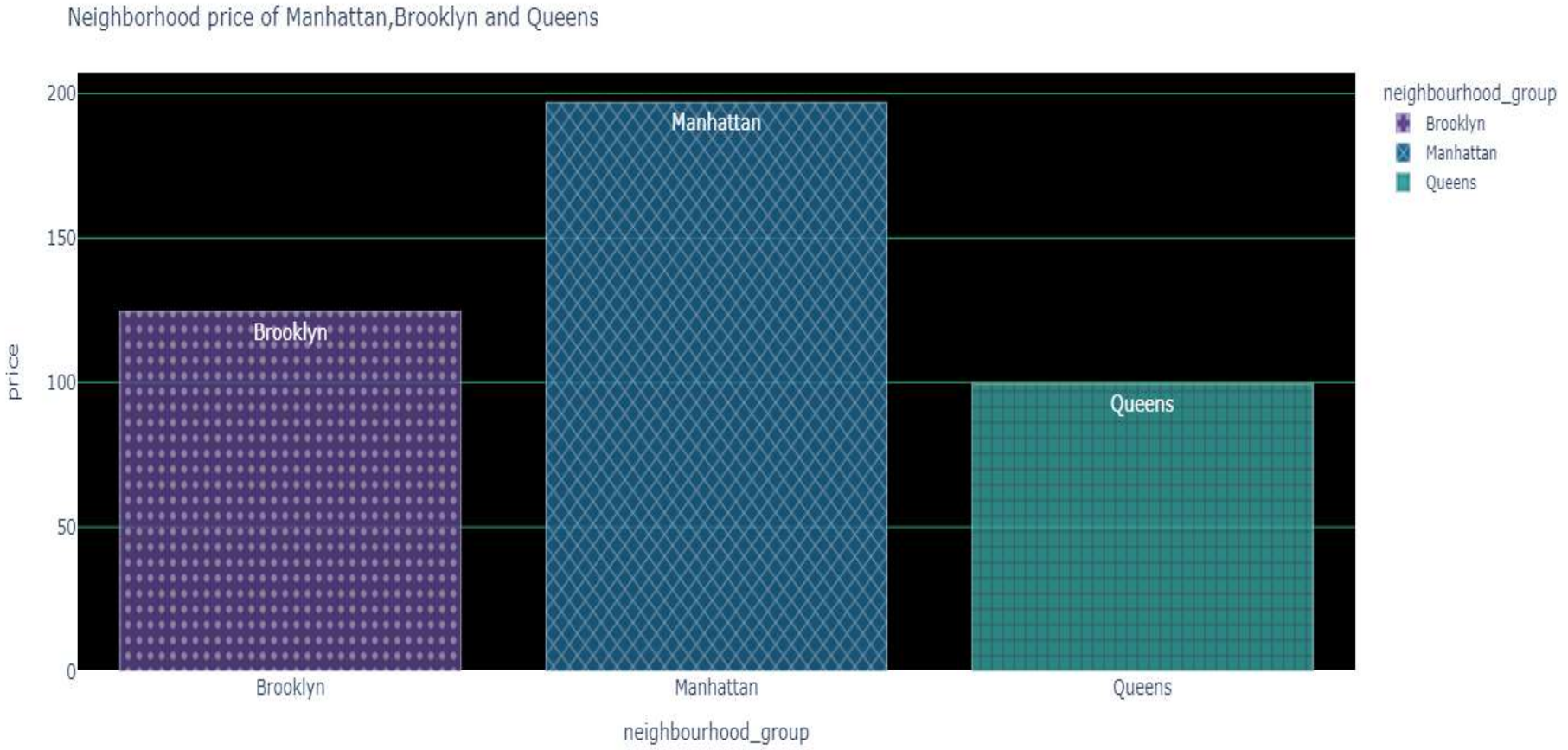
EDA contd...


Bronx
Brooklyn
Manhattan
Queens
Staten Island
Bronx
Manhattan
Queens
Brooklyn
Staten Island

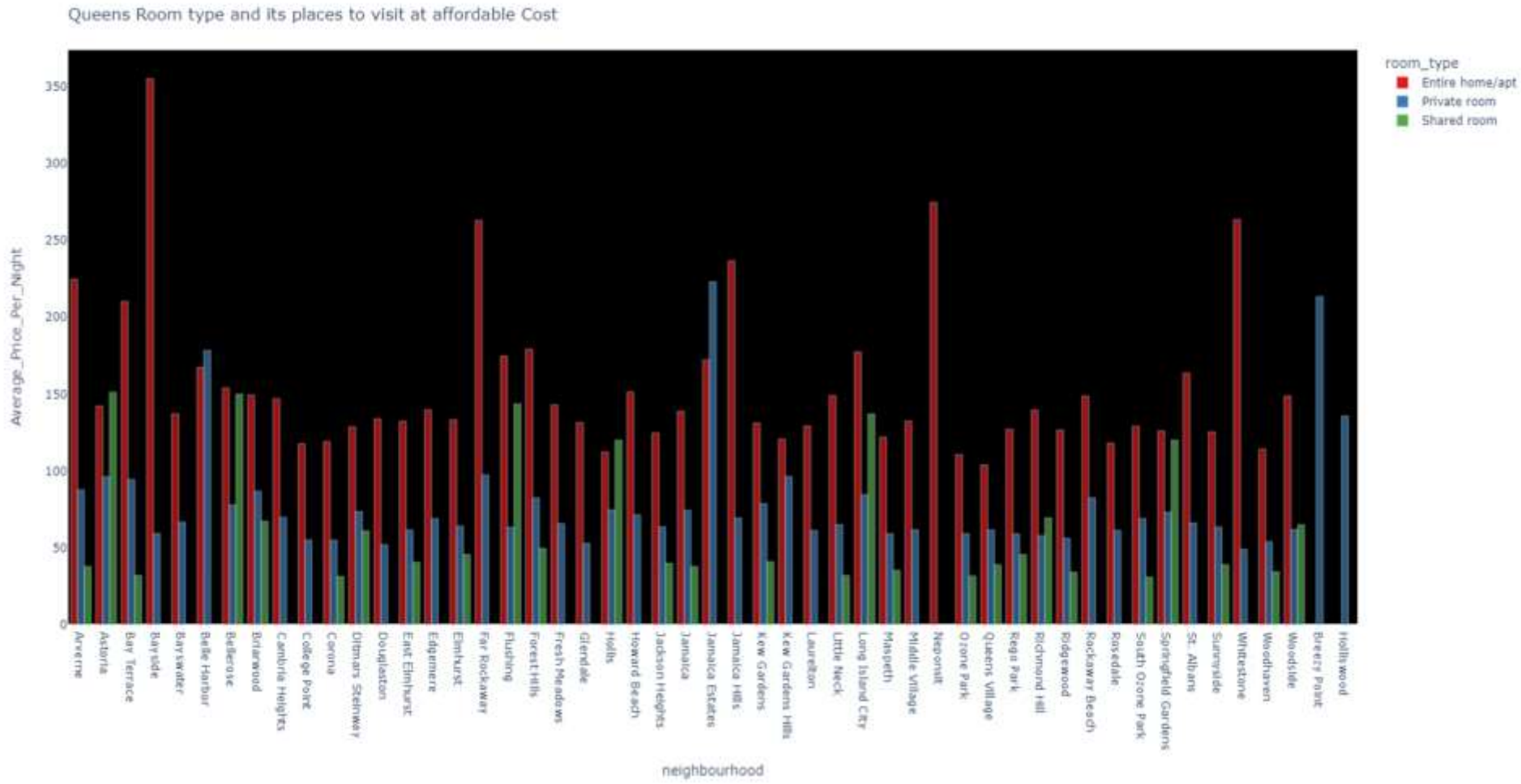
EDA contd...


Neighborhood Traffic
200
150
100
50
0
price
21651
20086
1089
5664
373
Bronx
Brooklyn
Manhattan
Queens
Staten Island
neighbourhood_group
number_of_reviews
450k
400k
350k
300k
250k
200k
150k
100k
50k

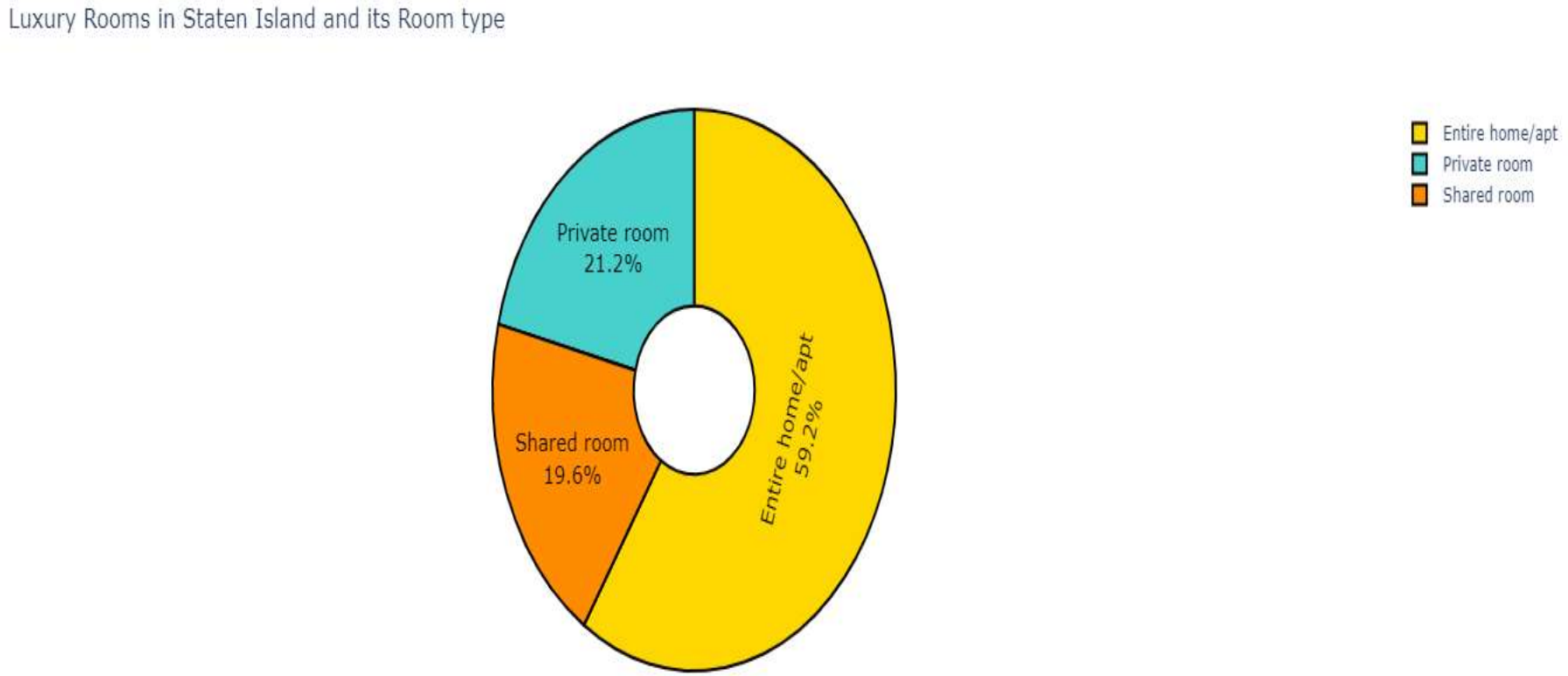
EDA contd...


Neighborhood price of Manhattan,Brooklyn and Queens
200
150
100
50
0
price
Brooklyn
Manhattan
Queens
Brooklyn
Manhattan
Queens
neighbourhood_group
neighbourhood_group
Brooklyn
Manhattan
Queens

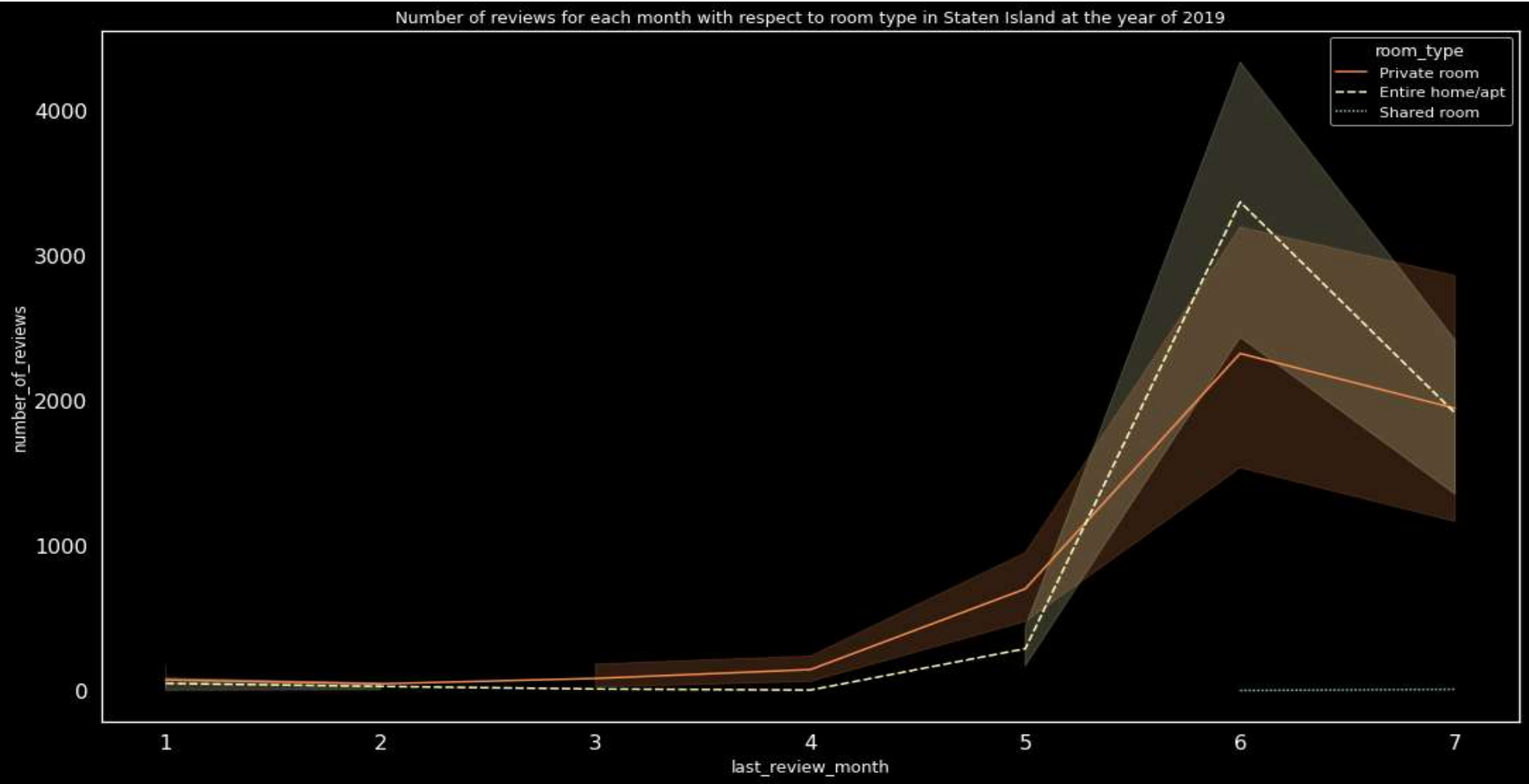
EDA contd...


Queens Room type and its places to visit at affordable Cost
Average_Price_Per_Night
0
50
100
150
200
250
300
350
neighbourhood
Arverne
Astoria
Bay Terrace
Bayside
Baysawater
Belle Harbor
Bellerose
Briarwood
Cambria Heights
College Point
Corona
Ditmars Steinway
Douglaston
East Elmhurst
Edgemere
Elmhurst
Far Rockaway
Flushing
Forest Hills
Fresh Meadows
Glendale
Holls
Howard Beach
Jackson Heights
Jamaica
Jamaica Estates
Jamaica Hills
Kew Gardens
Kew Gardens Hills
Laurelton
Little Neck
Long Island City
Maspeth
Middle Village
Neponsit
Ozone Park
Queens Village
Rego Park
Richmond Hill
Ridgewood
Rockaway Beach
Rosedale
South Ozone Park
Springfield Gardens
St. Albans
Sunnyside
Whitestone
Woodhaven
Woodside
Breezy Point
Hollis Wood
room_type
Entire home/apt
Private room
Shared room

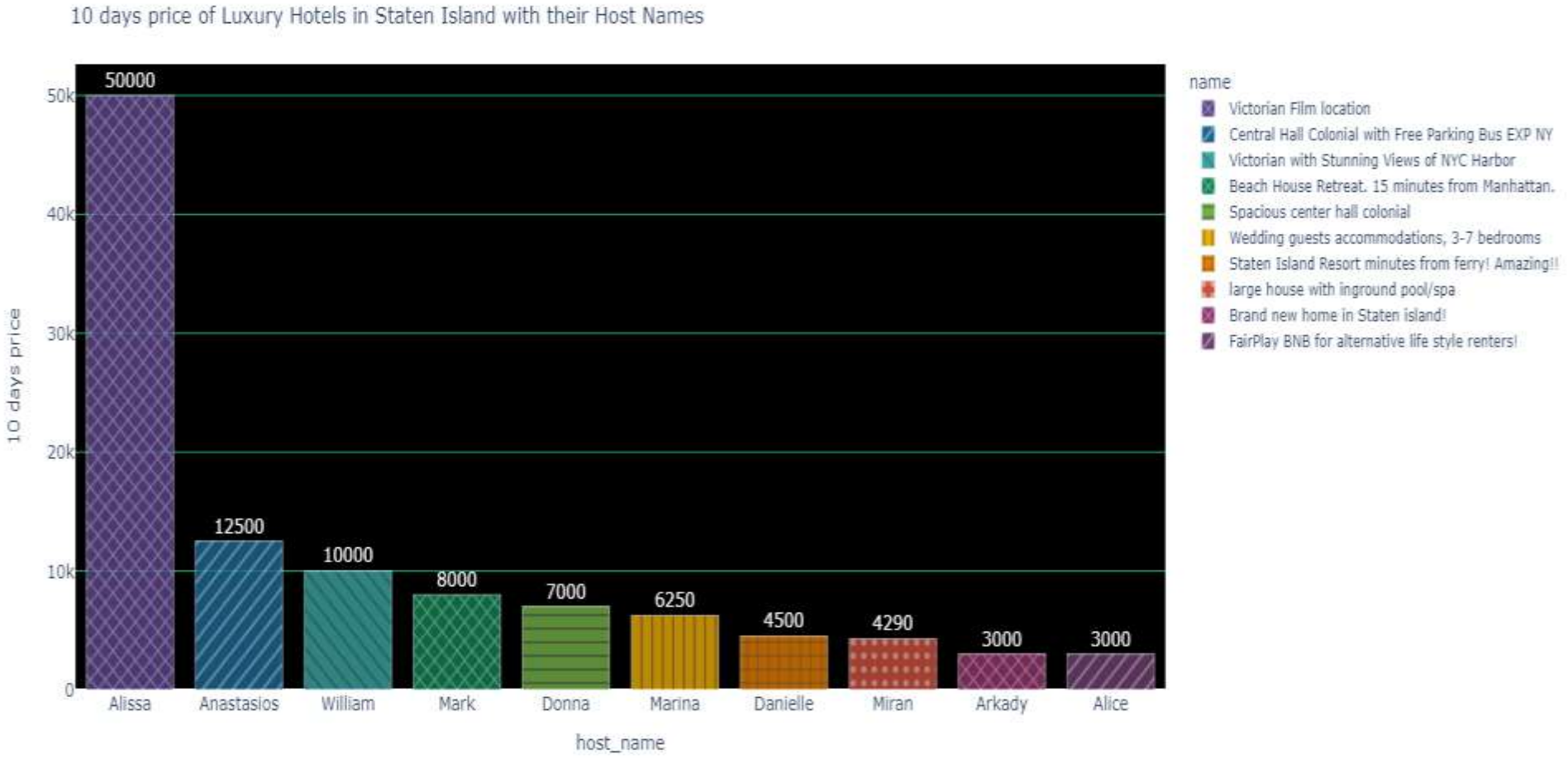
EDA contd...


Luxury Rooms in Staten Island and its Room type
Private room
21.2%
Shared room
19.6%
Entire home/apt
59.2%
Entire home/apt
Private room
Shared room

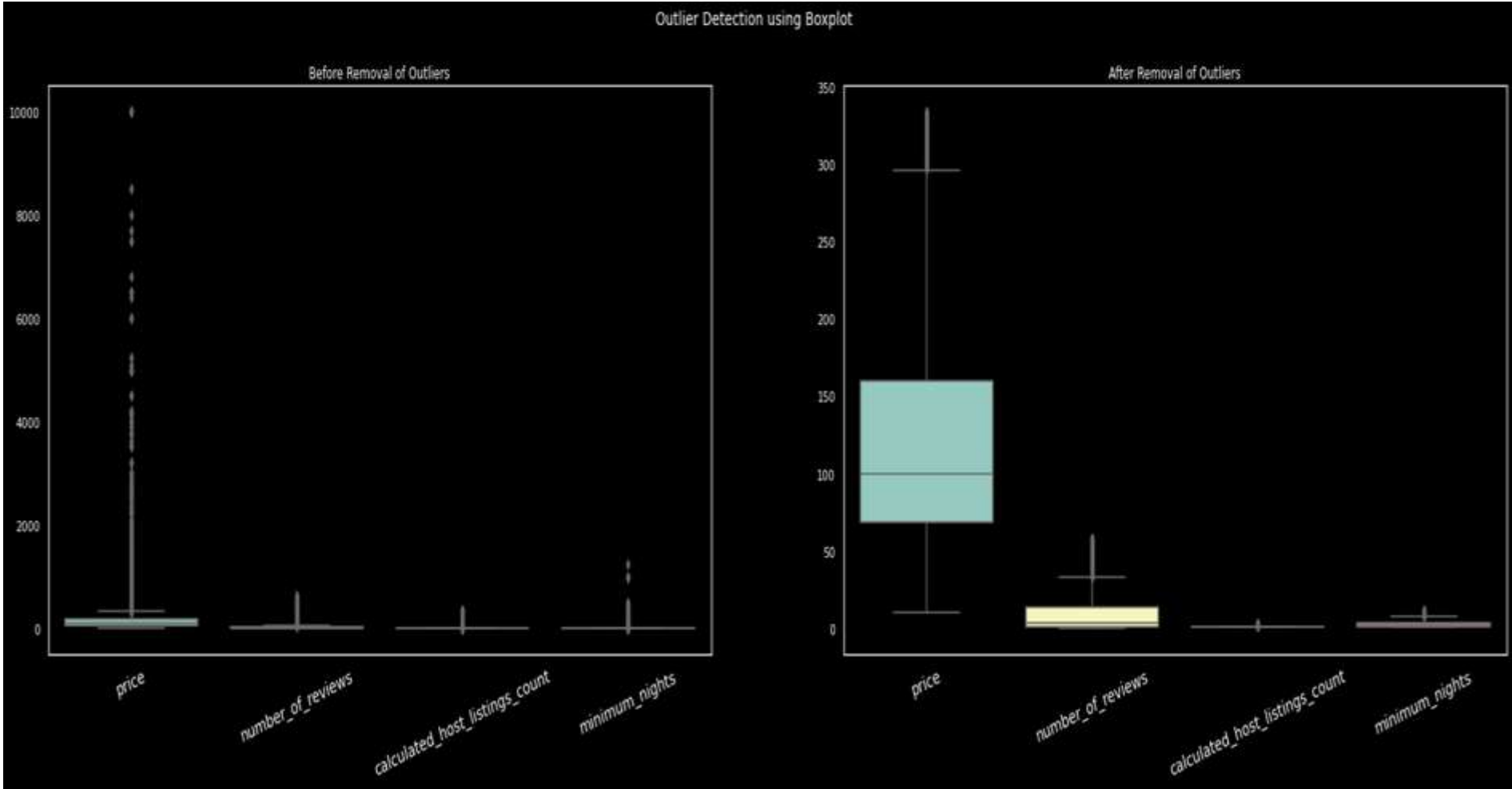
EDA contd...


Number of reviews for each month with respect to room type in Staten Island at the year of 2019
room_type
Private room
Entire home/apt
Shared room
number_of_reviews
0
1000
2000
3000
4000
1
2
3
4
5
6
7
last_review_month

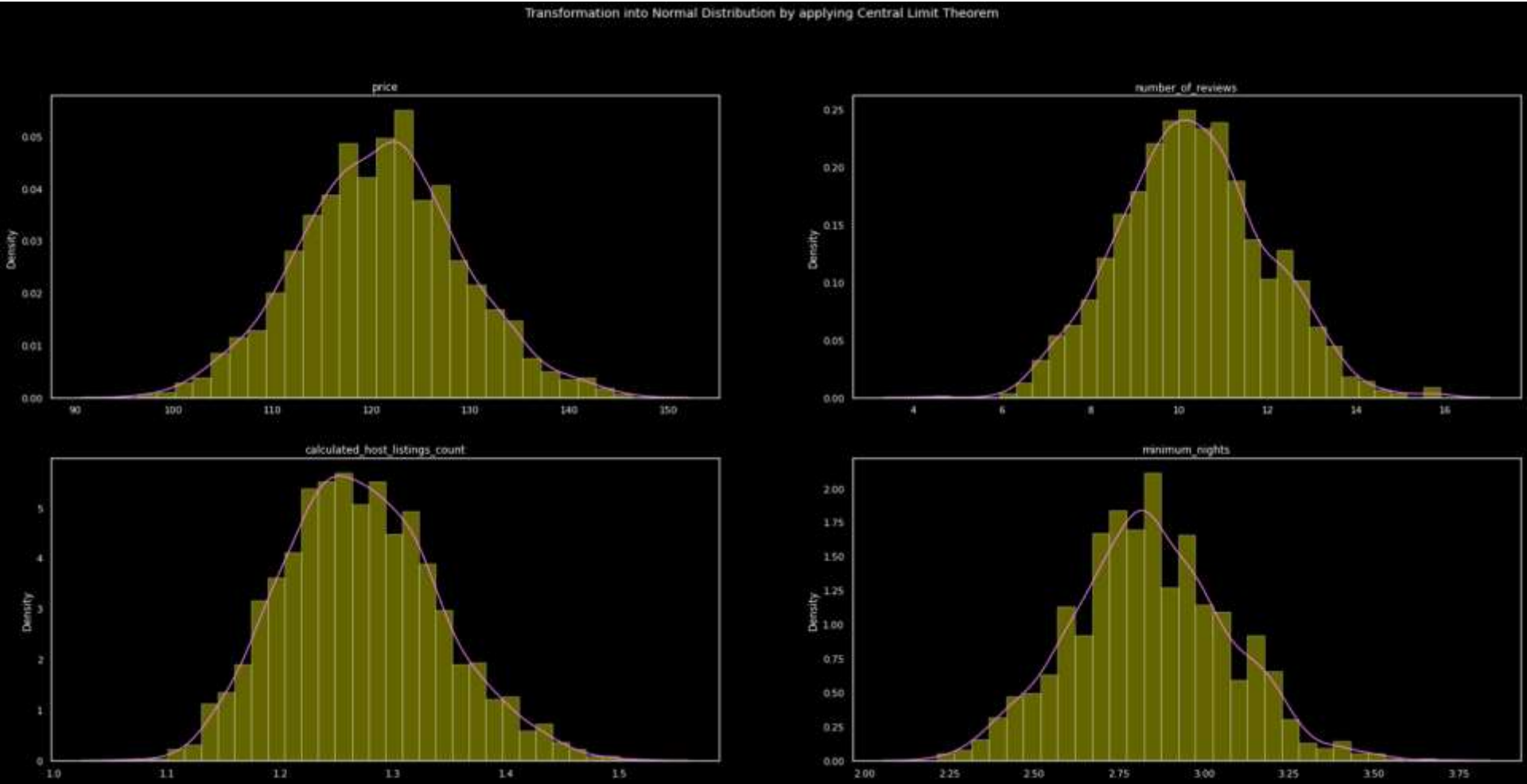
EDA contd...


10 days price of Luxury Hotels in Staten Island with their Host Names
50000
12500
10000
8000
7000
6250
4500
4290
3000
3000
50k
40k
30k
20k
10k
0
10 days price
Alissa
Anastasios
William
Mark
Donna
Marina
Danielle
Miran
Arkady
Alice
host_name
name
Victorian Film location
Central Hall Colonial with Free Parking Bus EXP NY
Victorian with Stunning Views of NYC Harbor
Beach House Retreat. 15 minutes from Manhattan.
Spacious center hall colonial
Wedding guests accommodations, 3-7 bedrooms
Staten Island Resort minutes from ferry! Amazing!!
large house with inground pool/spa
Brand new home in Staten island!
FairPlay BNB for alternative life style renters!

EDA contd...


Outlier Detection using Boxplot
Before Removal of Outliers
10000
8000
6000
4000
2000
0
price
number_of_reviews
calculated_host_listings_count
minimum_nights
After Removal of Outliers
350
300
250
200
150
100
50
0
price
number_of_reviews
calculated_host_listings_count
minimum_nights

EDA contd...


Transformation into Normal Distribution by applying Central Limit Theorem
price
number_of_reviews
calculated_host_listings_count
minimum_nights
Density

Statistical Tests:

- A statistical test is a tool that allows you to make quantitative conclusions about a process or set of activities
- Our goal is to do a statistical test to see if the prices of rooms in different neighbourhoods are similar or not.
- To do so, we built a group of neighbourhoods and took the mean of the prices, then chose two neighbourhood groups with prices that are closer to each other, then we used Hypothesis Testing to get what we wanted. When the prices of rooms in neighbouring groups are equal, the Null Hypothesis is accepted, and when they are not, the Alternate Hypothesis is accepted.

- **Result of the Hypothesis were:**
  - · price of rooms in Bronx and Queens are not equal
  - · price of rooms in Brooklyn and Staten Island are not equal
  - · price of rooms in Queens and Staten island are equal.
  - · price of rooms in Manhattan are not equal to any other.
- Another Statistical test we have done to find out whether the room type are similar to neighbourhood groups so to find out we have used Chi Square test.
  - · HO: Not Related (no association)
    - alpha > 0.05
  - · HA: Related (association)
    - alpha < 0.05
- ***Chi-square = ∑(fe−fo)2fe***
- Our goal was to find out the value of alpha and alpha value was less than 0.05 **so we can say neighbourhood groups and room types are related.**

Conclusion:

- Exploratory data analysis was used to aid. Customers can use the data to study and comprehend additional information.
- EDA assigns a host name based on the number of reviews and the amount of reviews every month to identify the best and most customer-friendly hosts.It also shows the host who has availability of 365 days.
- Customers can select the busiest host and learn why they are so busy. EDA provides each host with their own renting space
- It also provided useful information about Staten Island for his upcoming trip by analysing the number of reviews and luxury lodging options.
- With the help of these EDA, a greater amount of valuable information is provided to customers and make better decisions.