STORYTELLING CASE STUDY

Airbnb,NYC

DEEPALI AGGARWAL

TANISHQ BHALLA

Problem Background

For the past few months, Airbnb has seen a major decline in revenue. 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.

The different leaders at Airbnb want to understand some important insights based on various attributes in the dataset so as to increase the revenue.

Objective :

To prepare for the next best steps that Airbnb needs to take as a business, you have been asked to analyse a dataset consisting of various Airbnb listings in New York as a data analyst.

Dataset :

Contains information about different Airbnb listings along with their hosts, locations, prices and other attributes. Below is the column description –

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

```
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
```

Data Analysis :

Step 1: Imported appropriate libraries and read the data into a data frame.

In [1]:
```
# Importing required libraries

import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

In [2]:
```
# Ignoring warning

import warnings

warnings.filterwarnings('ignore')
```

Reading the dataset

In [3]:
```
airbnb= pd.read_csv(r"C:\Users\Aarti\Downloads\AB_NYC_2019.csv")
```

In [4]:
```
airbnb.head()
```

Out[4]:

	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	

Step 2: Data understanding/ Normal routine check

Data understanding

In [5]:
```
# Checking the dimension

airbnb.shape
```

Out[5]:
```
(48895, 16)
```

In [6]:
```
# Checking column names

airbnb.columns
```

Out[6]:
```
Index(['id', 'name', 'host_id', 'host_name', 'neighbourhood_group',
       'neighbourhood', 'latitude', 'longitude', 'room_type', 'price',
       'minimum_nights', 'number_of_reviews', 'last_review',
       'reviews_per_month', 'calculated_host_listings_count',
       'availability_365'],
      dtype='object')
```

In [7]:
```
# Checking column information

airbnb.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 48895 entries, 0 to 48894
Data columns (total 16 columns):
 #   Column                          Non-Null Count  Dtype  
---  ------                          --------------  -----  
 0   id                              48895 non-null  int64  
 1   name                            48879 non-null  object 
 2   host_id                         48895 non-null  int64  
 3   host_name                       48874 non-null  object 
 4   neighbourhood_group             48895 non-null  object 
 5   neighbourhood                   48895 non-null  object 
 6   latitude                        48895 non-null  float64
 7   longitude                       48895 non-null  float64
 8   room_type                       48895 non-null  object 
 9   price                           48895 non-null  int64  
 10  minimum_nights                  48895 non-null  int64  
 11  number_of_reviews               48895 non-null  int64  
 12  last_review                     38843 non-null  object 
 13  reviews_per_month               38843 non-null  float64
 14  calculated_host_listings_count  48895 non-null  int64  
 15  availability_365                48895 non-null  int64  
dtypes: float64(3), int64(7), object(6)
memory usage: 6.0+ MB
```


In [8]:
```
# Checking numerical summary of the dataset

airbnb.describe()
```

Out[8]:

	id	host_id	latitude	longitude	price	minimum_nights	number_of_reviews	reviews_per_month	calculated_host_listings
count	4.889500e+04	4.889500e+04	48895.000000	48895.000000	48895.000000	48895.000000	48895.000000	38843.000000	48895.
mean	1.901714e+07	6.762001e+07	40.728949	-73.952170	152.720687	7.029962	23.274466	1.373221	7.
std	1.098311e+07	7.861097e+07	0.054530	0.046157	240.154170	20.510550	44.550582	1.680442	32.
min	2.539000e+03	2.438000e+03	40.499790	-74.244420	0.000000	1.000000	0.000000	0.010000	1.
25%	9.471945e+06	7.822033e+06	40.690100	-73.983070	69.000000	1.000000	1.000000	0.190000	1.
50%	1.967728e+07	3.079382e+07	40.723070	-73.955680	106.000000	3.000000	5.000000	0.720000	1.
75%	2.915218e+07	1.074344e+08	40.763115	-73.936275	175.000000	5.000000	24.000000	2.020000	2.
max	3.648724e+07	2.743213e+08	40.913060	-73.712990	10000.000000	1250.000000	629.000000	58.500000	327.

Observations:

- Dataset contains **48895** rows and **16** columns.
- 3 columns are of float type, 7 columns of int type and 6 columns of object type.
- Average price of rentals in NY is **$153**, average no. of nights spent is **7 days**, and average availability of an Airbnb is **112 days**.
- There are some listings which appear to be free.
- Few customer(s) are also exists who have rented for almost **3 years**.
- There are rentals which are available for all times of the year too i.e., **365 days**.
- Outliers are also present in columns: minimum_nights, number_of_reviews, reviews_per_month and calculated_host_listings_count columns.

Step 3: Data Quality Inspection

```
In [9]: # Checking Null value count

        airbnb.isnull().sum()
```

```
Out[9]: id                                0
        name                             16
        host_id                           0
        host_name                        21
        neighbourhood_group               0
        neighbourhood                     0
        latitude                          0
        longitude                         0
        room_type                         0
        price                             0
        minimum_nights                    0
        number_of_reviews                 0
        last_review                   10052
        reviews_per_month             10052
        calculated_host_listings_count    0
        availability_365                  0
        dtype: int64
```

```
In [10]: # Mssing data
         # There are around 10,000+ null values in last_review and reviews_per_month column
         # There are around 20+ null values in host_name column
         # There are 16 null values in nam column
```

```
In [11]: # Checking number of unique values in each column of the dataset

         airbnb.nunique()
```

```
Out[11]: id                                48895
         name                              47896
         host_id                           37457
         host_name                         11452
         neighbourhood_group                   5
         neighbourhood                       221
         latitude                          19048
         longitude                         14718
         room_type                             3
         price                               674
         minimum_nights                      109
         number_of_reviews                   394
         last_review                        1764
         reviews_per_month                   937
         calculated_host_listings_count       47
         availability_365                    366
         dtype: int64
```



```
In [12]: # There are 5 neighbourhood groups, 221 neighbouhoods are 3 room_types
```

```
In [13]: # Checking number of duplicate rows

         airbnb.duplicated().sum()
```

```
Out[13]: 0
```

```
In [14]: # There no duplicate rows in the dataset
```

Step 4: Data Cleaning

Four columns have NaN values, and the missing values must be dealt with.

In [15]:
```
# Replacing all NaN values in 'reviews_per_month' with 0

airbnb.fillna({'reviews_per_month':0},inplace=True)
```

In [16]:
```
airbnb.host_name.mode()
```

Out[16]:
```
0    Michael
dtype: object
```

In [17]:
```
airbnb.name.mode()
```

Out[17]:
```
0    Hillside Hotel
dtype: object
```

In [18]:
```
# Replacing all NaN values in 'Host_name' with mode value of the column

airbnb['host_name'].fillna('Michael',inplace=True)
```

In [19]:
```
# Replacing all NaN values in 'Name' with mode value of the column

airbnb['name'].fillna('Hillside Hotel',inplace=True)
```

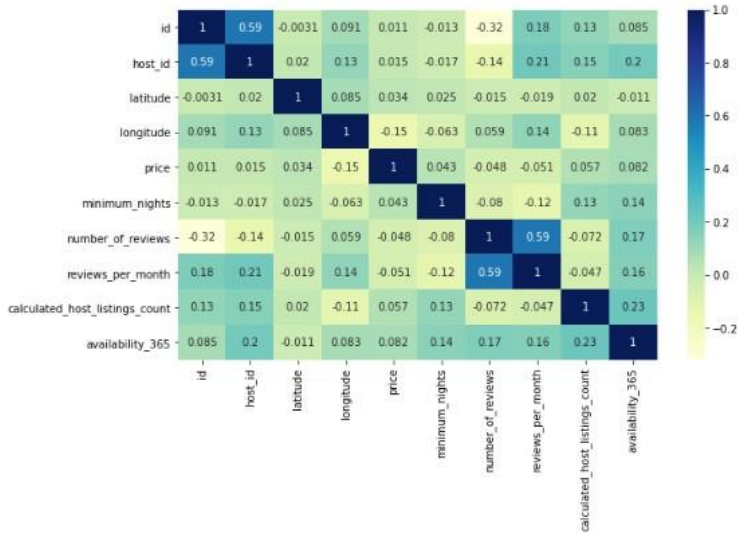

In [20]:
```
# Creating a new category for NaN values
```

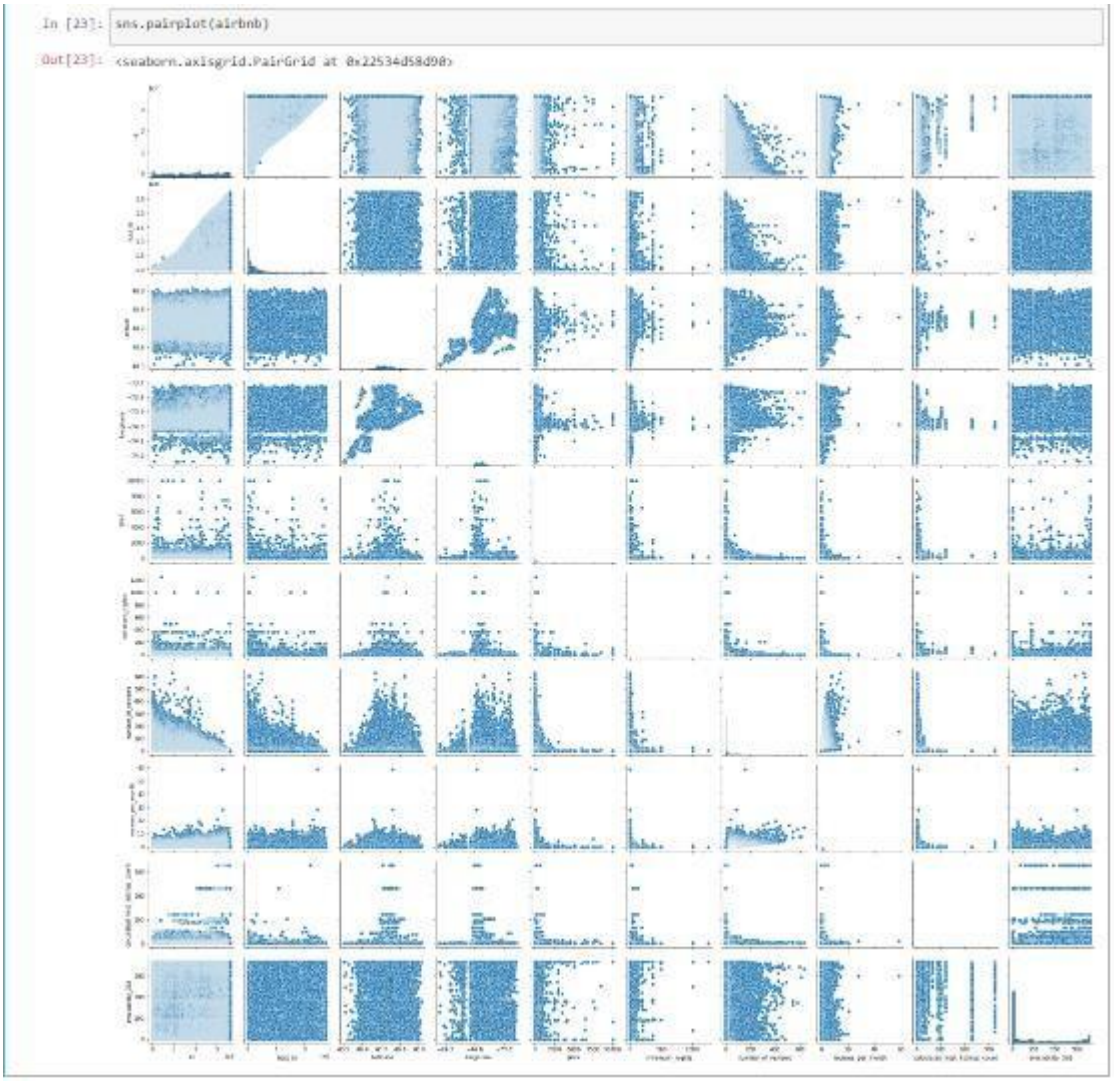
In [21]:
```
# Replacing all NaN values in 'last_review' with string 'Unavailble'

airbnb['last_review'].fillna('Unavailble',inplace=True)
```

Step 5: Bivariate Analysis

In [22]:
```
plt.figure(figsize=(10,6))
sns.heatmap(airbnb.corr(),cmap='YlGnBu',annot=True)
```

Out[22]:
```
<AxesSubplot:>
```



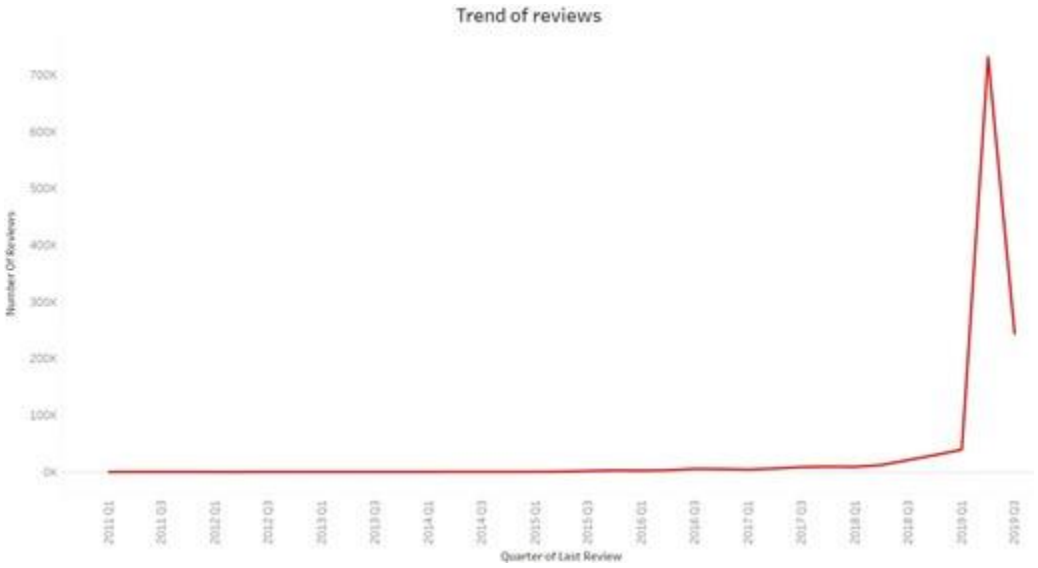


- There are no significant correlations between columns in the dataset.

Step 6: Exported cleaned data frame to a CSV file

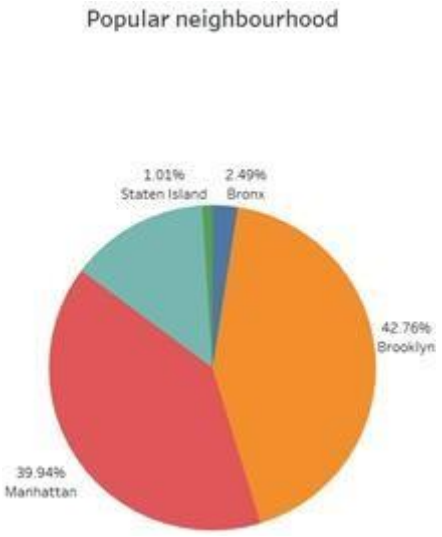
```
In [24]: airbnb.shape
Out[24]: (48895, 16)

In [25]: # Checking null values count

         airbnb.isnull().sum()
Out[25]: id                                0
         name                              0
         host_id                           0
         host_name                         0
         neighbourhood_group               0
         neighbourhood                     0
         latitude                          0
         longitude                         0
         room_type                         0
         price                             0
         minimum_nights                    0
         number_of_reviews                 0
         last_review                       0
         reviews_per_month                 0
         calculated_host_listings_count    0
         availability_365                  0
         dtype: int64

In [28]: airbnb.to_csv('AB_NYC_2019_cleaned.csv')
```


Step 7: Data Visualization



- COVID-19 pandemic affected Airbnb business due to travel restrictions.
- Revenue took the largest hit in NYC in the **Q2** and **Q3** of 2019.

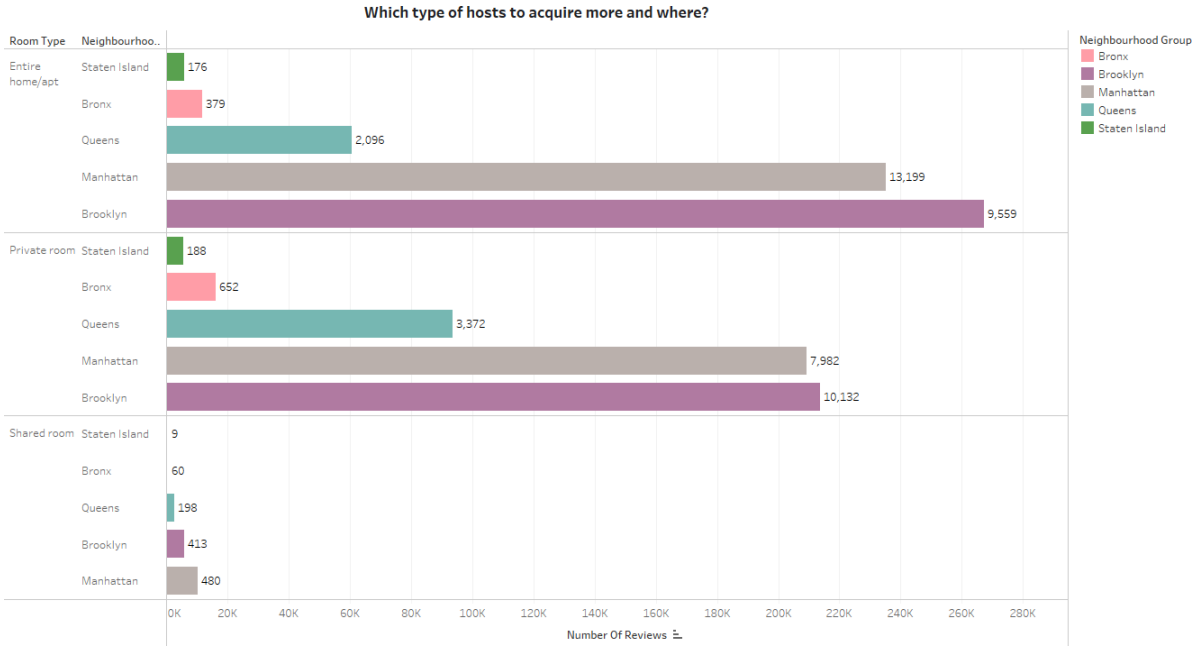
Neighbourhood with maximum listing

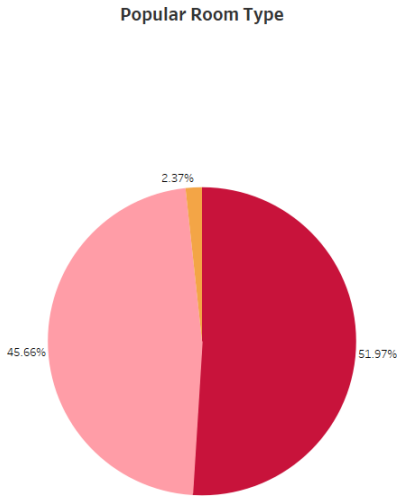


- **Manhattan** and **Brooklyn** are most popular neighborhoods with regards to total number of listings. These two neighbourhood account for **82.7%** of the listings in NYC.

SOME IMPORTANT INSIGHTS:

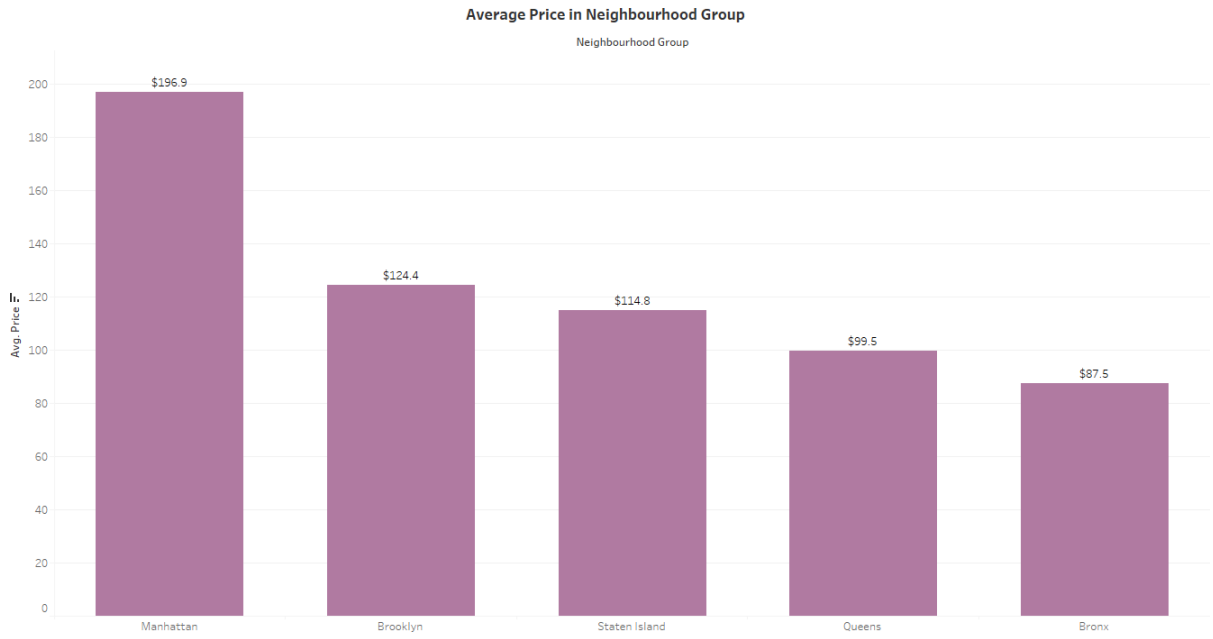
1. Which type of hosts to acquire more and where?





- Although, shared room accounts for only **2.37%** of all listings in NYC, Airbnb must acquire more no. of Entire home/apt and private room listings across all neighbourhood groups as these will be preferred by customers during the pandemic to avoid coming in contact/co-habiting with strangers.

Average price of listing per neighborhood



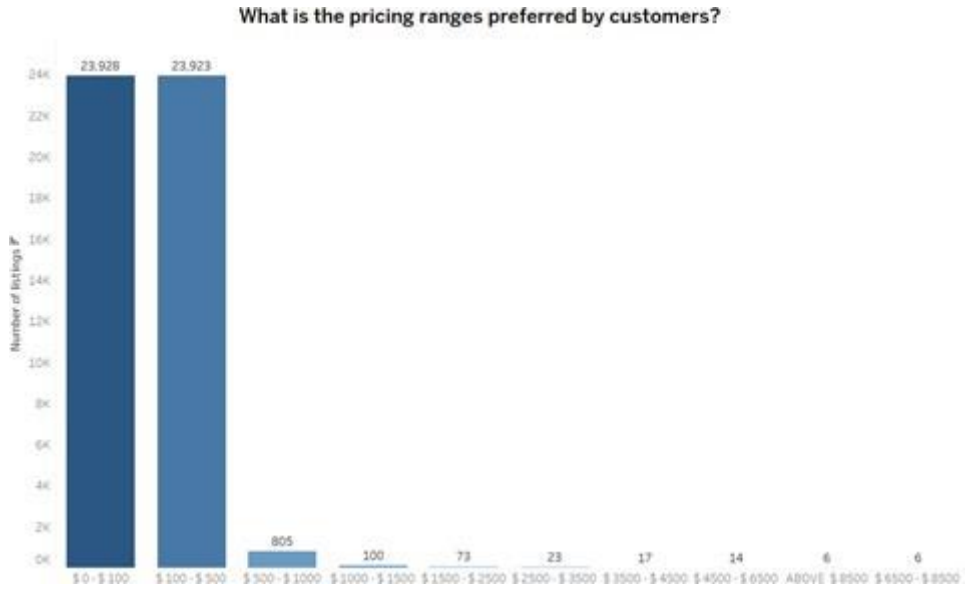
Average price of listing per neighbourhood is **$197** for **Manhattan** which is much higher than the other neighborhoods.

2. The categorization of customers based on their preferences:

i) What are the neighborhoods they need to target?

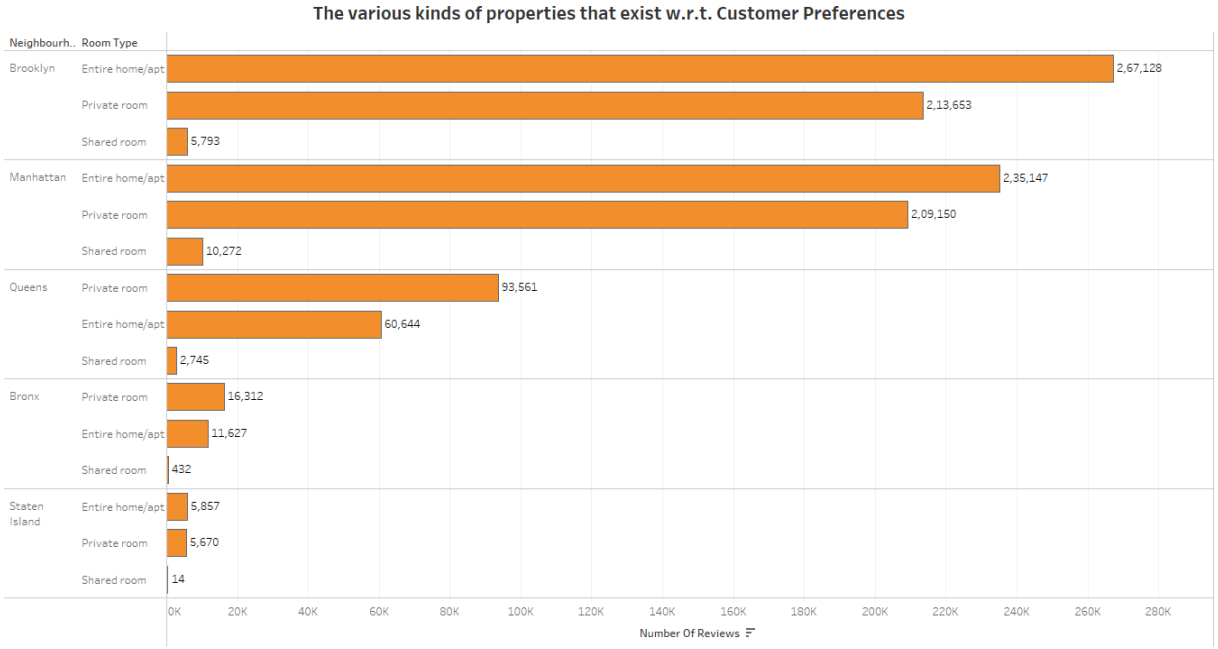
- Airbnb must acquire more customers and listings at **Bronx, Staten Island** and **Queens.**

ii) What are the pricing ranges preferred by customers?



- **$0 to $500** price range per night is preferred by customers while visiting NYC.

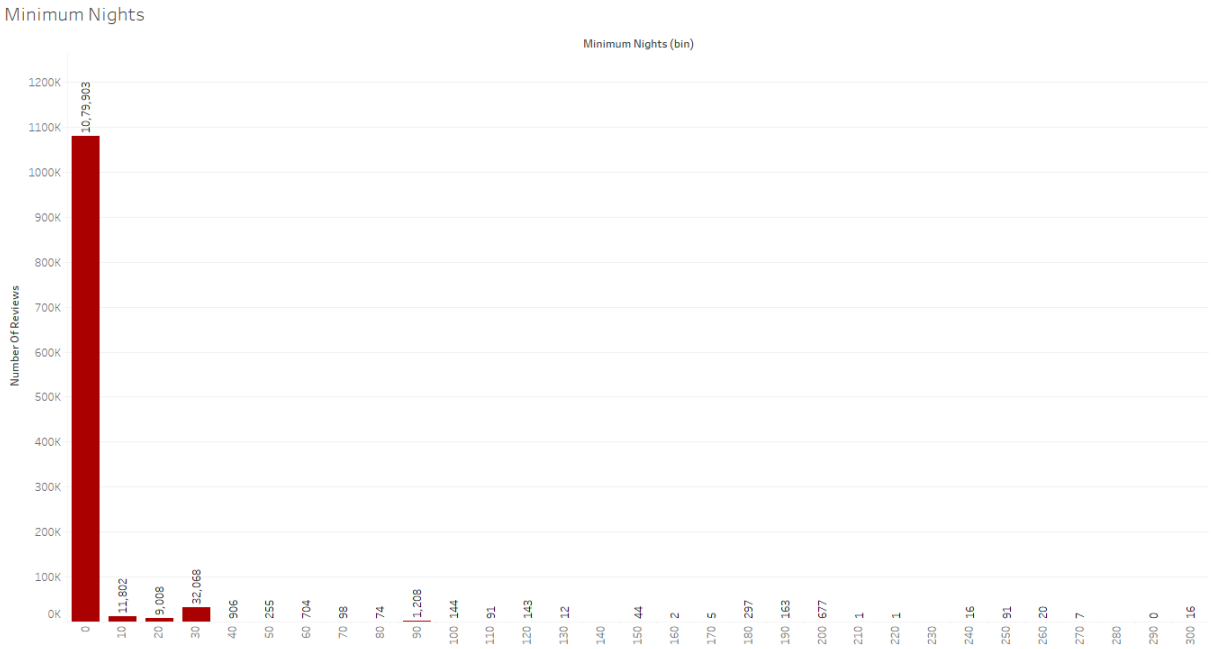
iii) The various kinds of properties that exits w.r.t. customer preferences.



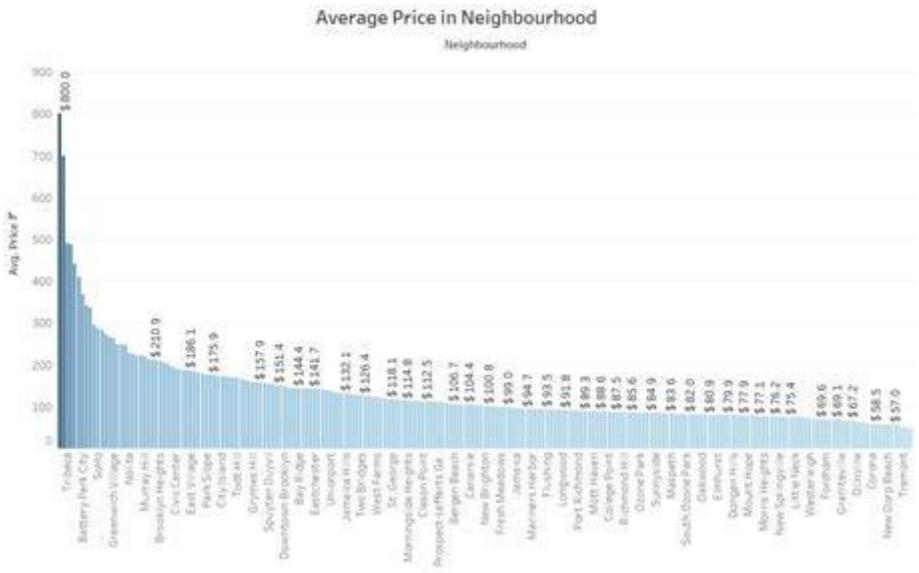
- **Manhattan and Brooklyn** prefer Entire Home/ Apartment and Private Rooms
- **Bronx and Queens** neighbourhood prefer Private rooms over Entire Home/Apartment.
- **At Staten Island,** customer doesn't have any specific preferences over Entire Home/Apartment and Private Rooms.
- **Shared rooms** are least preferred among all the neighborhoods might be because of pandemic.

iv) Adjustments in the existing properties to make it more customers oriented.

- Since, customers prefer **entire home/apt or private rooms** more. So, shared rooms can be converted to private rooms as customers' would not prefer sharing rooms with strangers due to ongoing pandemic.
- Customers prefer **$0 to $500** per night price points; hence, decrease the listings' price to attract more customers towards existing properties.
- Customers usually prefer **1-10 days** of minimum nights per stay. But due to ongoing pandemic, customers may prefer to book properties for 30, 60 or 90 days to avoid frequent travelling. Hence, the listings must have high availability.

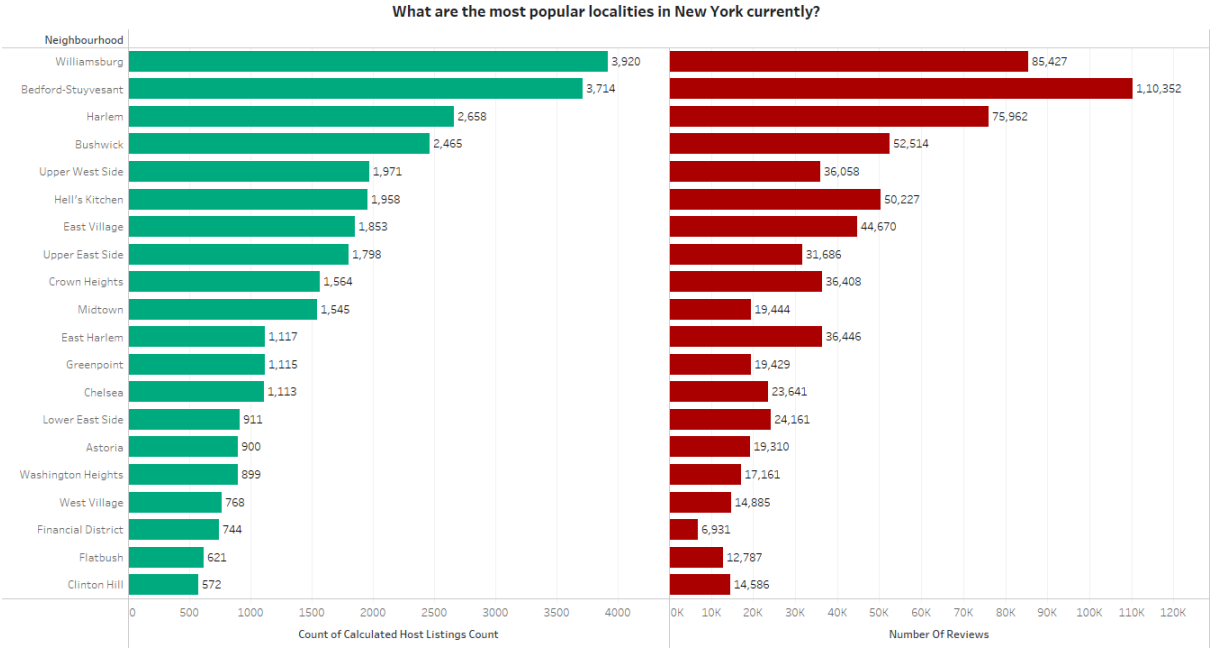


Average price per neighborhood

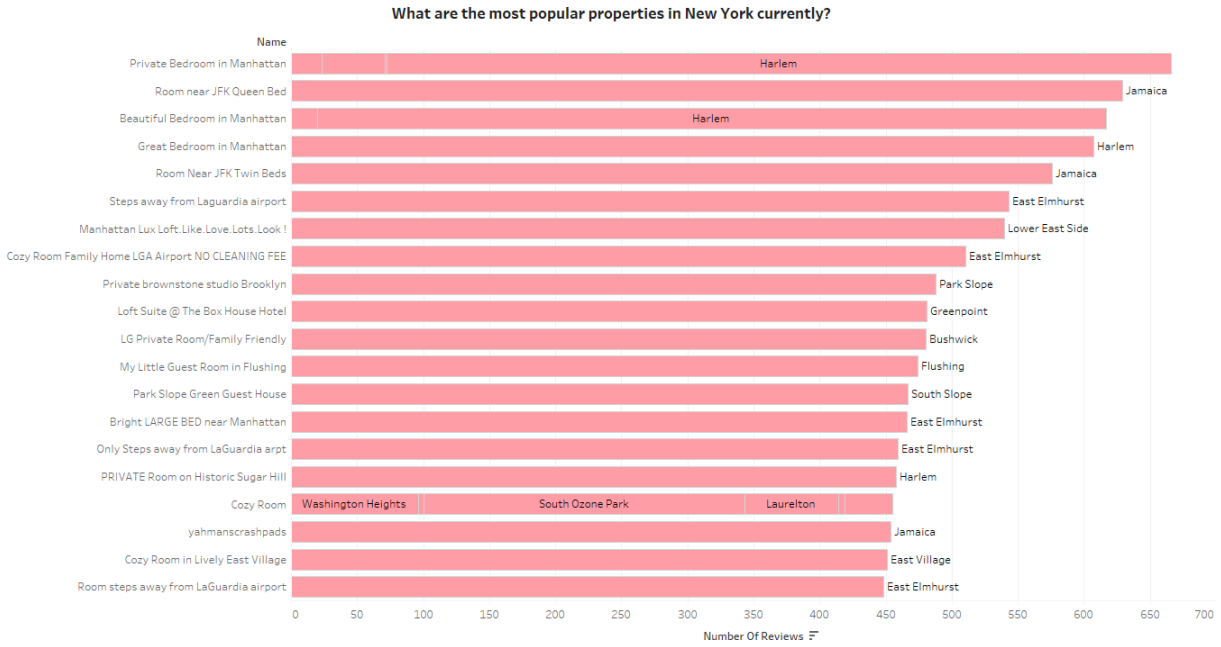


- Two neighborhoods in **Staten Island** have high average price according to above graph. This is due to availability of only one property in these neighborhoods.
- The cheapest locality at **Manhattan** is **Inwood($88.9)** and that in **Brooklyn** is **Borough Park($63.1).**

3. What are the most popular localities and properties in New York currently?



- **Williamsburg, Bedford- Stuyvesant,** and **Harlem** have the highest number of reviews and listings in NY neighborhoods.
- Although, **Williamsburg** has the highest number of listings, but its number of reviews are comparatively less.



- There are few properties with the same name across neighborhoods.
- Although, the number of listings in **Queens** is less, properties feature a lot in the top N results w.r.t the highest number of ratings. This may be attributed also to

the fact that fewer listings are available in Queens and hence, listings have higher number of ratings available.

4. How to get unpopular properties more traction ?

Assumption: Unpopular properties are the ones that have not been visited even once.

- Customers prefer **$0 to $500** price points per day. Hence, decrease the listings' price to attract more customers towards existing properties.
- Customers prefer **1-10** days of minimum nights per stay. Due to ongoing pandemic, customers may prefer to book properties for 30, 60 or 90 days to avoid frequent travelling. Properties with higher days of minimum nights per stay should also be acquired as they could have market traction.
- Listings must be highly available to target higher number of bookings.

How to get unpopular properties more traction?

Neighbourh..	Neighbourh..	Name	Last Review	Room Type	Availability 3..	Min expense per stay	Minimum Nights	Number Of Reviews	Price
Manhattan	Chelsea	New york Multi-unit building	Unavailable	Entire home/apt	0	1,750	7	0	250
		Maison 130	Unavailable	Shared room	0	250	1	0	250
		Luxury building Brand new Apartment in Ch..	Unavailable	Entire home/apt	60	1,250	5	0	250
		Gorgeous and Renovated Chelsea Apartme..	Unavailable	Entire home/apt	0	1,000	4	0	250
		Extremely large 1 bed in West Chelsea- Hig..	Unavailable	Entire home/apt	0	750	3	0	250
		Beautiful 1BR in Chelsea	Unavailable	Entire home/apt	0	250	1	0	250
		Awesome Studio in Chelsea	Unavailable	Entire home/apt	0	250	1	0	250
		Sonder | 21 Chelsea | Serene 1BR + Rooftop	Unavailable	Entire home/apt	225	7,221	29	0	249
		Modern sunlit luxury studio in Chelsea/Flat..	Unavailable	Entire home/apt	45	996	4	0	249
		Open + Cozy Chelsea Studio w/ Gym, Lovely..	Unavailable	Entire home/apt	345	7,440	30	0	248
		Sonder | 21 Chelsea | Bright 1BR + Rooftop	Unavailable	Entire home/apt	330	7,105	29	0	245
		Adorable Hudson Yards Studio w/ Gym + Do..	Unavailable	Entire home/apt	0	7,290	30	0	243
		Swanky Central Chelsea Studio w/ Balcony..	Unavailable	Entire home/apt	180	7,260	30	0	242
		Great deal-private NON shared apt in Chels..	Unavailable	Entire home/apt	0	480	2	0	240
		Cozy Chelsea Carriage House Apt	Unavailable	Entire home/apt	0	240	1	0	240
		West 33rd street, Lux 1bd Serviced Apartm..	Unavailable	Entire home/apt	360	7,170	30	0	239
		Sonder | 21 Chelsea | Artsy 1BR + Rooftop	Unavailable	Entire home/apt	330	6,931	29	0	239
		Best Location! 3BR / 1 BA Chelsea! Steps to ..	Unavailable	Entire home/apt	15	468	2	0	234
		Beautiful, Bright Apartment in Heart of Ch..	Unavailable	Entire home/apt	0	904	4	0	226
		West Chelsea 1 bedroom 2 bathroom outdo..	Unavailable	Entire home/apt	0	3,150	14	0	225
		Urban Oasis in Chelsea	Unavailable	Entire home/apt	15	450	2	0	225
		Sixth Ave Chelsea, Studio Serviced Apt*	Unavailable	Entire home/apt	225	6,750	30	0	225
		QUIET OASIS in Prime Chelsea LRGE APT PR..	Unavailable	Entire home/apt	270	6,750	30	0	225
		Quaint 1BR in West Village	Unavailable	Entire home/apt	0	225	1	0	225
		PRIME Chelsea Modern Luxury Home *****	Unavailable	Entire home/apt	225	225	1	0	225
		New Studio Great Location in NYC	Unavailable	Entire home/apt	75	675	3	0	225
		Luxury large studio in Chelsea, at the Highli..	Unavailable	Entire home/apt	15	1,575	7	0	225
		Flatiron Oasis	Unavailable	Entire home/apt	0	225	1	0	225
		Chelsea Apartment (2 bedroom)	Unavailable	Entire home/apt	0	225	1	0	225
		Bright Chelsea Entire 2 Bedroom Apartment	Unavailable	Entire home/apt	75	675	3	0	225
		2 Floor Minimalist Loft located in Lower Ch..	Unavailable	Entire home/apt	0	450	2	0	225
		Sunny, quiet 2 BR apt in the heart of Manha..	Unavailable	Entire home/apt	165	669	3	0	223
		West Village cozy 2 bedroom NYC	Unavailable	Entire home/apt	0	880	4	0	220
		Stylish New York Apartment	Unavailable	Entire home/apt	0	3,080	14	0	220
		Midtown West Hotel - Bright Queen Room	Unavailable	Private room	360	219	1	0	219
		Studio in Full Service Chelsea Bldg	Unavailable	Entire home/apt	0	215	1	0	215
		CHELSEA BEAUTY Prime location	Unavailable	Entire home/apt	120	215	1	0	215
		Beautiful Apt - Flatiron/ Chelsea	Unavailable	Entire home/apt	0	215	1	0	215
		Homey Hudson Yards w/ Gym + Doorman, n..	Unavailable	Entire home/apt	300	6,390	30	0	213
		CLASSIC 1BR IN WEST 15TH STREET-CHELS..	Unavailable	Entire home/apt	360	6,300	30	0	210
		AMAZING CHELSEA Apartment HUGE space	Unavailable	Entire home/apt	330	210	1	0	210
		Sonder | 21 Chelsea | Vibrant Studio + Roof..	Unavailable	Entire home/apt	360	6,061	29	0	209
		Sonder | 21 Chelsea | Chic 1BR + Rooftop	Unavailable	Entire home/apt	330	6,061	29	0	209
		Sonder | 21 Chelsea | Airy Studio + Rooftop	Unavailable	Entire home/apt	315	6,061	29	0	209
		Sixth Ave Chelsea, Studio Serviced Apartm..	Unavailable	Entire home/apt	345	6,150	30	0	205

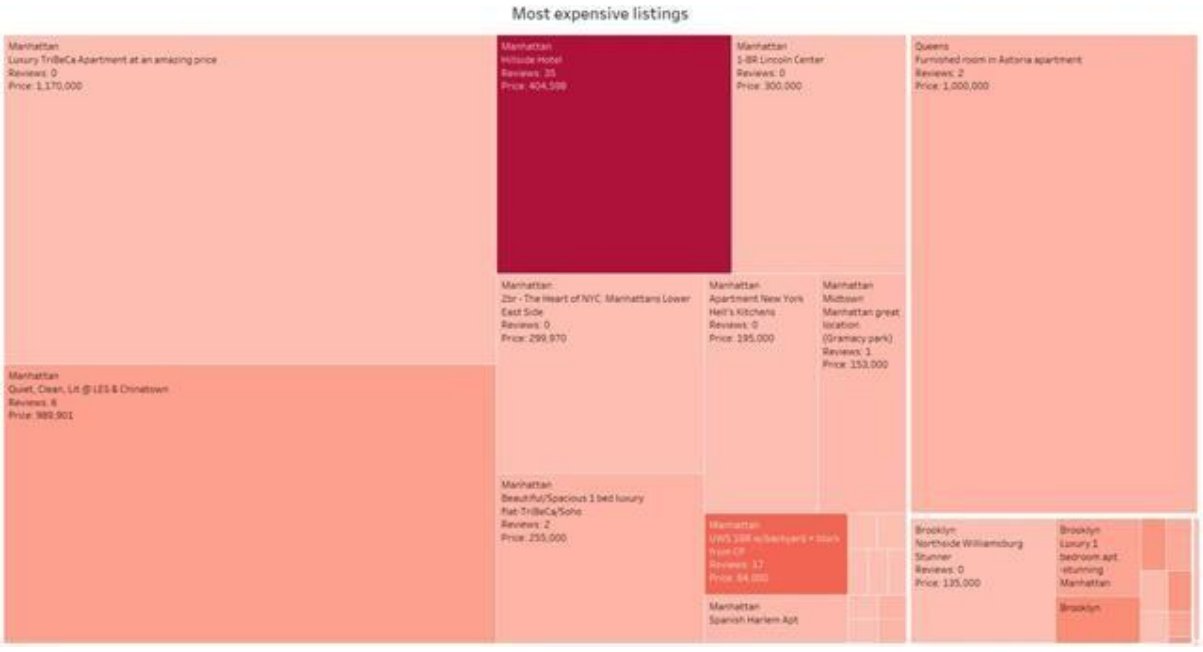
5. Most expensive listings by price per night

Most expensive listings

Name	Neighbourhood	Neighbourhood Group	
Luxury 1 bedroom apt. -stunning Manhattan views	Greenpoint	Brooklyn	10,000
Furnished room in Astoria apartment	Astoria	Queens	10,000
1-BR Lincoln Center	Upper West Side	Manhattan	10,000
Spanish Harlem Apt	East Harlem	Manhattan	9,999
Quiet, Clean, Lit @ LES & Chinatown	Lower East Side	Manhattan	9,999
2br - The Heart of NYC: Manhattans Lower East Side	Lower East Side	Manhattan	9,999
Beautiful/Spacious 1 bed luxury flat-TriBeCa/Soho	Tribeca	Manhattan	8,500
Film Location	Clinton Hill	Brooklyn	8,000
East 72nd Townhouse by (Hidden by Airbnb)	Upper East Side	Manhattan	7,703
Gem of east Flatbush	East Flatbush	Brooklyn	7,500
70' Luxury MotorYacht on the Hudson	Battery Park City	Manhattan	7,500
3000 sq ft daylight photo studio	Chelsea	Manhattan	6,800
SUPER BOWL Brooklyn Duplex Apt!!	Clinton Hill	Brooklyn	6,500
Luxury TriBeCa Apartment at an amazing price	Tribeca	Manhattan	6,500
Apartment New York ..	Upper West Side	Manhattan	6,500
Park Avenue Mansion by (Hidden by Airbnb)	Upper East Side	Manhattan	6,419
UWS 1BR w/backyard + block from CP	Upper West Side	Manhattan	6,000
Luxury townhouse Greenwich Village	Greenwich Village	Manhattan	6,000
SuperBowl Penthouse Loft 3,000 sqft	Little Italy	Manhattan	5,250
Chelsea Gallery Space for events and exhibitions	Chelsea	Manhattan	5,220
Midtown Manhattan great location (Gramacy park)	Midtown	Manhattan	5,100
Victorian Film location	Randall Manor	Staten Island	5,000
NearWilliamsburg bridge 11211 BK	Bedford-Stuyvesant	Brooklyn	5,000
Fulton 2	Cypress Hills	Brooklyn	5,000
Broadway 1	Bedford-Stuyvesant	Brooklyn	5,000
Beautiful 1 Bedroom in Nolita/Soho	Nolita	Manhattan	5,000
4-Floor Unique Event Space 50P Cap. - #10299B	Harlem	Manhattan	5,000
Photography Location	Clinton Hill	Brooklyn	4,500
Northside Williamsburg Stunner	Williamsburg	Brooklyn	4,500
bay ridge & sunset park furnished apartment	Bay Ridge	Brooklyn	4,200

- Listing at **Queens** is available for **$10,000** per night and that in **Staten Island** is available for **$5,000**. These are premium listings for Queens and Staten Island respectively.
- There are multiple premium listings for **Brooklyn** and **Manhattan**.

6. Most expensive listings by minimum expense per stay



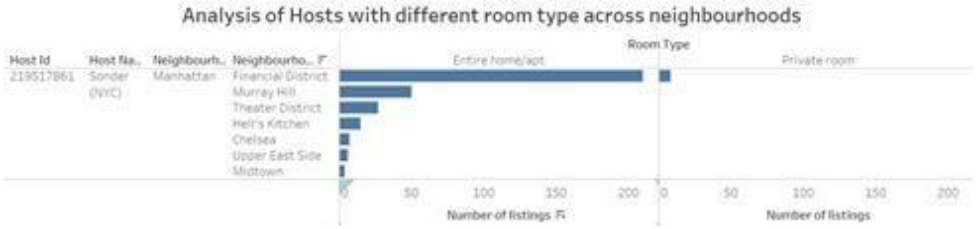
- Most expensive listing costs **$1,170,00** per stay but no reviews available for this listing.
- **Hillside Hotel** and **UMS 1BR w/backyard+block** from **CP** at Manhattan are most visited **premium** listings.

7. Hosts with maximum listings

Hosts with maximum listings

Host Id	Host Name	Distinct count of Id	Number Of Reviews
219517861	Sonder (NYC)	327	1,281
107434423	Blueground	232	29
30283594	Kara	121	65
137358866	Kazuya	103	87
16098958	Jeremy & Laura	96	138
12243051	Sonder	96	43
61391963	Corporate Housing	91	417
22541573	Ken	87	55
200380610	Pranjal	65	1
7503643	Vida	52	242
1475015	Mike	52	162
120762452	Stanley	50	84
205031545	Red Awning	49	127
2856748	Ruchi	49	10
190921808	John	47	281
26377263	Stat	43	52
2119276	Host	39	335
19303369	Hiroki	37	22
119669058	Melissa	34	498
25237492	Juliana	34	322
213781715	Anting	33	32
113805886	Yaacov	33	123
76104209	Rated	33	3
238321374	Eyal	32	11
50760546	CRNY Monthly Rentals	31	72

- There are many hosts with more than 10 listings at different neighbourhoods.
- Only **Sonder(NYC)** has a large number of reviews, hence many customers have stayed at his listings.
- All other listings have very less reviews, either customers haven't reviewed the listings as much or due to poor reviews, these listings have not been visited frequently. The quality of these listings must be reviewed by Airbnb.
- **Sonder(NYC)** has highest number of listings across NY and he primarily owns Entire room/apt.



8. Price distribution of listings across neighborhoods

- **Premium properties** in **Bronx** and **Queens** must be targeted as they cost an upwards of **$2500**.
- **Non premium properties** in **Manhattan** and **Brooklyn** must be targeted. Premium properties are exorbitant and may not attract a lot of bookings. Cheaper listings are usually preferred in these localities.