

# **Capstone Project**

## **Airbnb 2019 Data Analysis**

**- Rajat Srivastava**



- It's an online platform which lets people, rent out their properties to travellers and in same way travellers can book a property online hastelfree.
- Airbnb has 5.6 million active listings worldwide.
- There are at least 100,000 cities across the world with active Airbnb listings.

# Importance of Data Analysis

- Millions of listings are generated worldwide every year, and nearly half a million guests checked in per day.
- Analyzing these data helps in business decision making, understanding guest and host behaviour, platform's performance, implementing new marketing strategies and many more.

# Brief about the Dataset

- Dataset is of Airbnb listings of New York city for the year 2019.
- The City is divided into 5 Groups namely Manhattan, Brooklyn, Queens, Bronx, and Staten Island.
- These Groups are further divided into sub-group neighborhoods.
- There are 48,895 rows and 16 columns.
- Out of 16 columns, there are -
  - 10 numeric columns
  - 6 categorical columns

# Let's know more about each column now

- **id:** Unique identification for each property in airbnb listings.
- **name:** Name of each property in airbnb listings.
- **host\_id:** Unique identification for each host who has listed their properties on airbnb.
- **host\_name:** Name of Host who have listed their properties on airbnb. Total unique 11452 host\_names in data set.
- **neighbourhood\_group:** Names of all areas in NYC city.
- **neighborhood:** Name of all sub-areas in NYC city.
- **latitude, longitude:** Coordinates for each property in airbnb listings.
- **room\_type:** Each type of room available in properties . In all areas Private rooms, Entire home/apt, Shared room types are available.
- **price:** Price of each property in airbnb listing.
- **minimum\_night:** Mandatory number of nights room needs to be booked
- **number\_of\_review:** Number of reviews for each listed property.
- **last\_review:** Last date the property was reviewed.
- **review\_per\_month:** Number of reviews received per month by the property.
- **Calculated\_host\_listing\_counts:** Number of properties each host owns.
- **Availability\_365:** Number of days in the year the property is available for booking.

# Data Analysis Process

- Data Collection and Understanding it.
- Data Cleaning and Preparation.
- Data Analyzing and Visualizing.
- Data Interpretation and Application.

# Missing Values

```
# Checking total number of missing values in each column index  
airbnb_data.isnull().sum()
```

```
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
```

## Observations -

1. Less than 1% data is missing in name and host\_name columns.
2. Around 20% data is missing in last\_review and reviews\_per\_month columns.

# Descriptive Statistics

- Minimum price can not be 0, as no property is free of cost on airbnb listing.
- Average cost of properties are 152 but some properties can cost as high as 10000.
- Though price of such properties can be relevant, but mean cost will be skewed and we will not be able to perform accurate analysis.

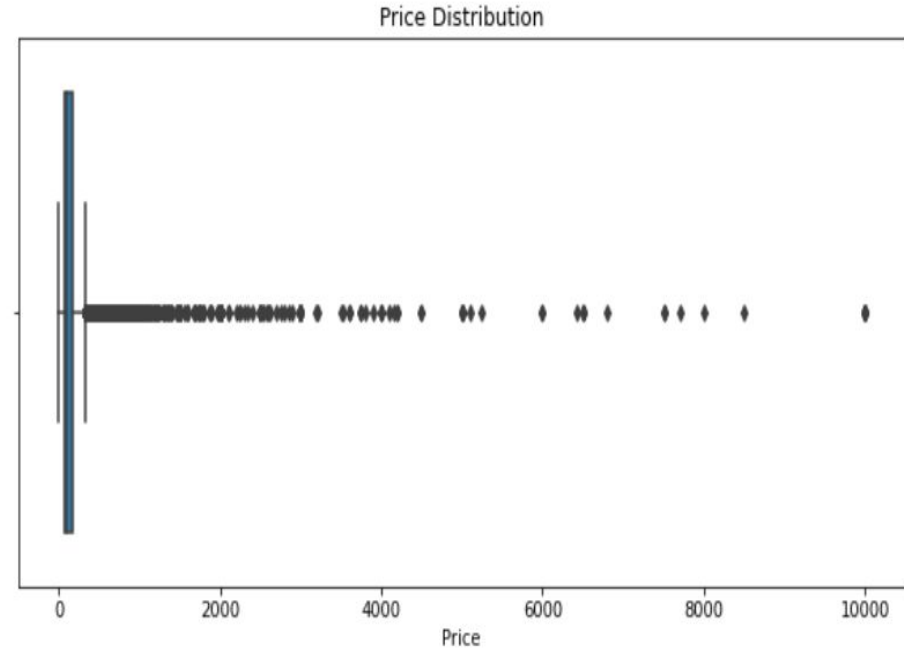
price	
mean	152.720687
median	106.000000
min	0.000000
max	10000.000000
count	48895.000000

---



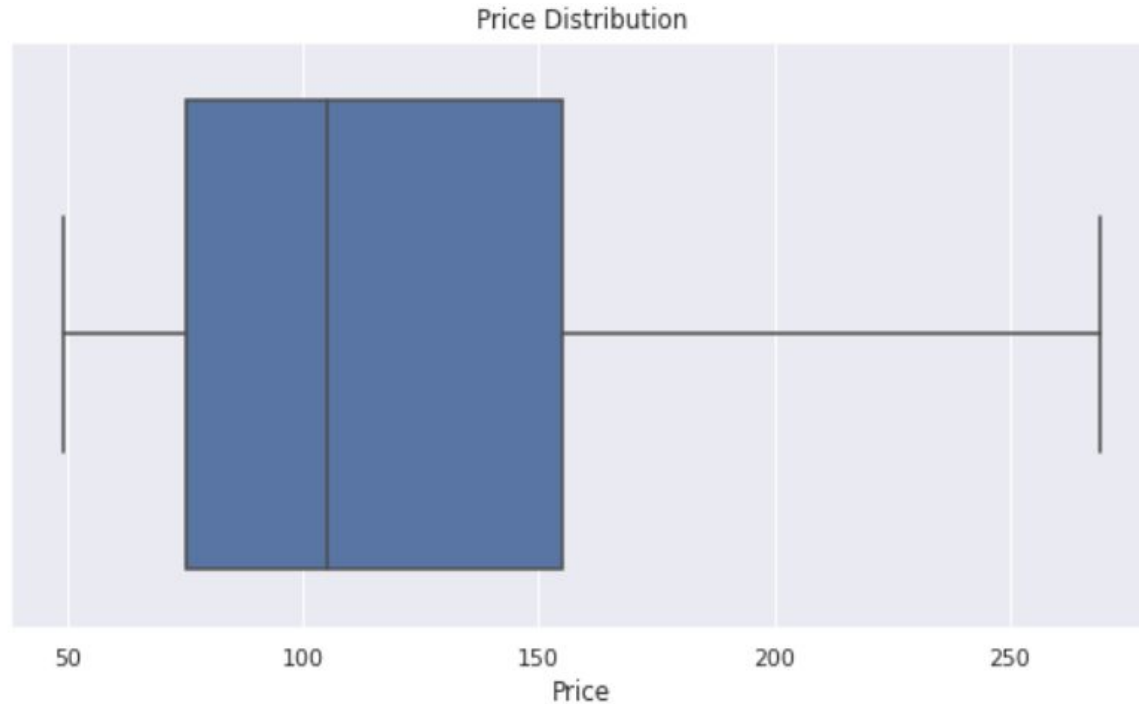
# Box plot showing outliers in data

- Values outside upper whisker are clearly visible.
- We need to remove such outliers by dropping 10th and 90th percentile of the price column from the dataset.



# Box plot showing no outliers in data

- All price values are under the range.
- Price column has no outlier and ready for further analysis



# Data Cleaning and Preparation

```
▶ # Checking for null values again before moving along with further analysis  
  
airbnb_data_new.isnull().sum()
```

```
↳ 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  
   calculated_host_listings_count 0  
   availability_365   0  
   dtype: int64
```

Our dataset is all ready for data analysis and visualization.

# 1A. Host with most listed properties

host_name	neighbourhood_group	calculated_host_listings_count
-----------	---------------------	--------------------------------

Sonder (NYC)	Manhattan	327
--------------	-----------	-----

Blueground	Manhattan	232
------------	-----------	-----

Kara	Manhattan	121
------	-----------	-----

Kazuya	Manhattan	103
--------	-----------	-----

Kazuya	Queens	103
--------	--------	-----

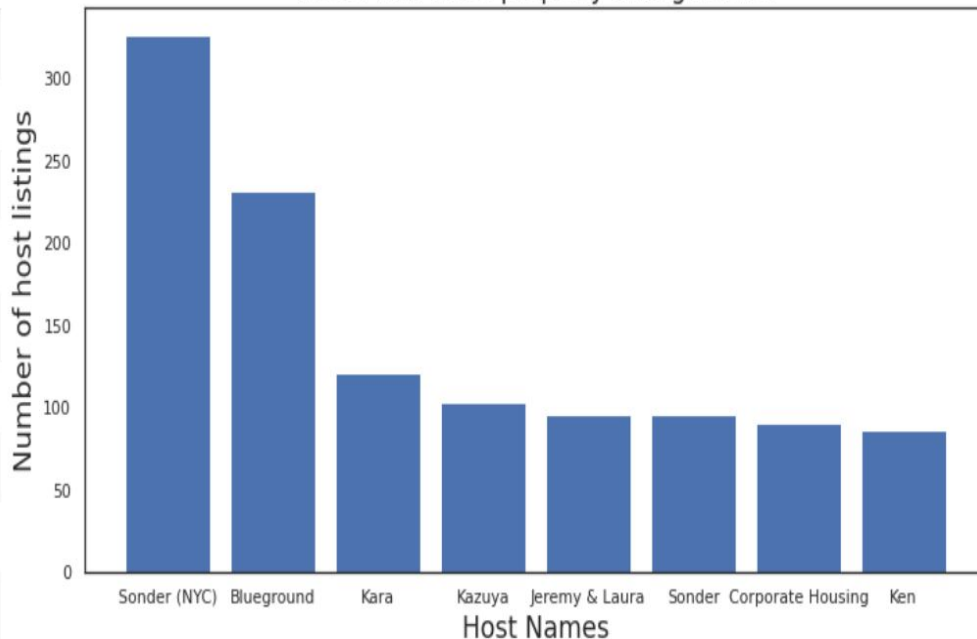
Jeremy & Laura	Manhattan	96
----------------	-----------	----

Sonder	Manhattan	96
--------	-----------	----

Corporate Housing	Manhattan	91
-------------------	-----------	----

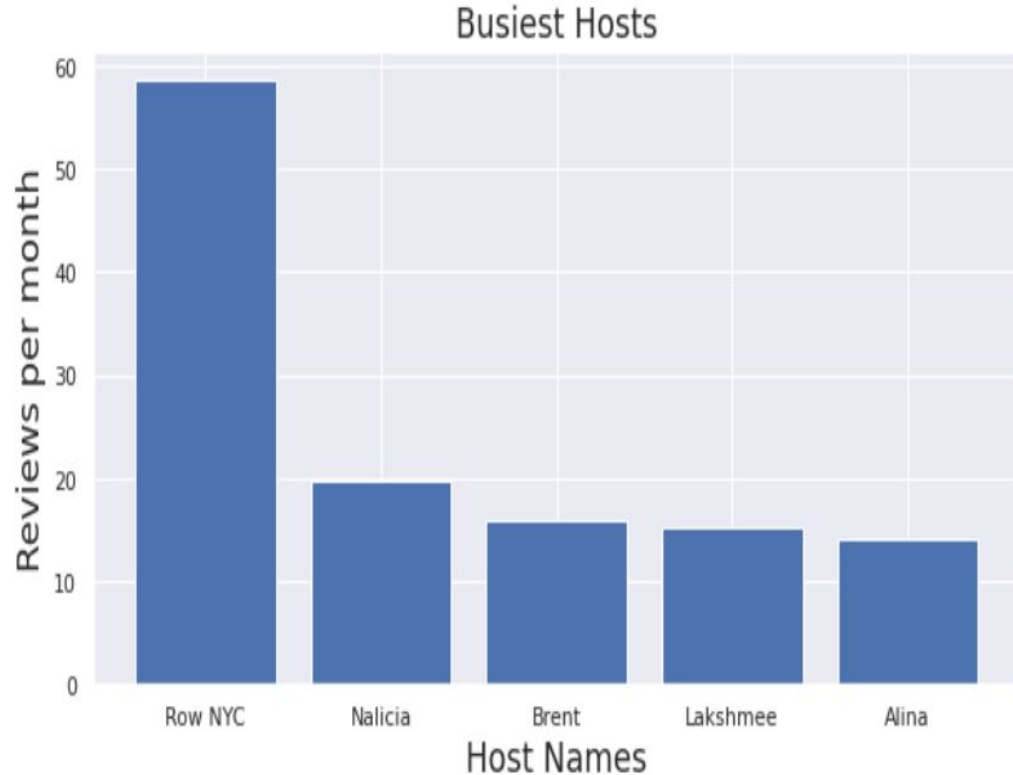
Ken	Brooklyn	87
-----	----------	----

Hosts with most property listing in NYC



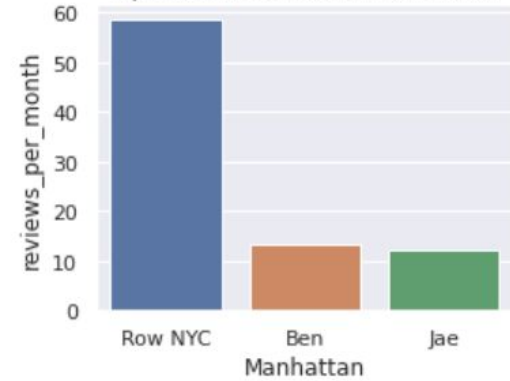
## 1B. Busiest and most popular hosts in NYC

host_name	host_id	neighbourhood_group	neighbourhood	reviews_per_month
Row NYC	244361589	Manhattan	Theater District	58.50
Nalicia	156684502	Queens	Springfield Gardens	19.75
Brent	217379941	Queens	Springfield Gardens	15.78
Lakshmee	232251881	Queens	Jamaica	15.32
Alina	259630588	Brooklyn	Williamsburg	14.00

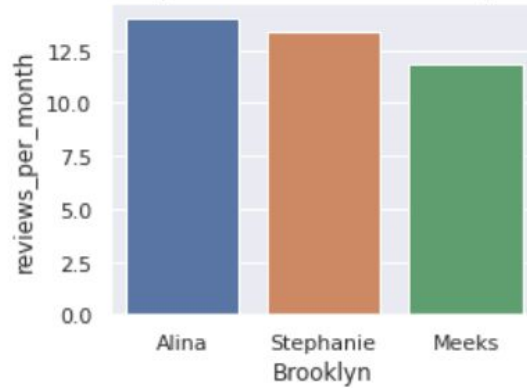


# 1C. Popular Hosts in neighbourhood groups

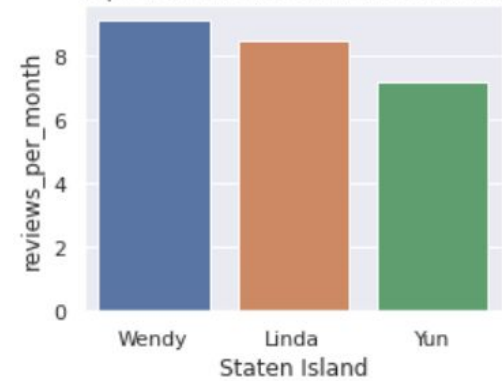
Top 3 favourite hosts in Manhattan



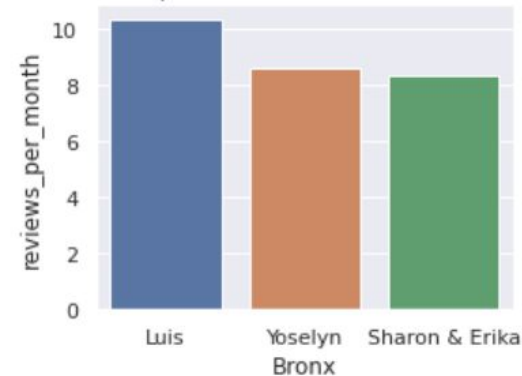
Top 3 favourite hosts in Brooklyn



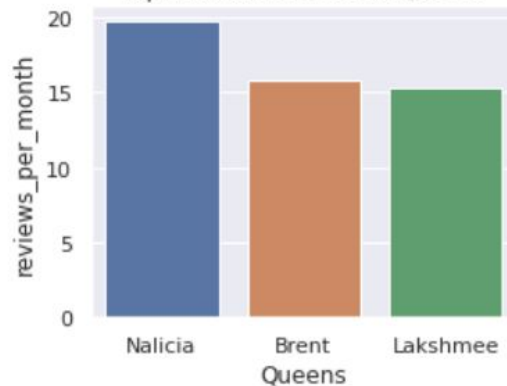
Top 3 favourite hosts in Staten Island



Top 3 favourite hosts in Bronx



Top 3 favourite hosts in Queens



## Top hosts in different neighbourhood groups

1. Manhattan - Row Nyc,Ben,Jae
2. Bronx - Luis,Yoselyn,Sharon & Erika
3. Brooklyn - Alina, Stephanie, Meeks
4. Queens - Nalicia,Brent,Lakshmee
5. Staten Island - Wendy,Linda,Yun

## 2A.Total properties in each neighbourhood group

neighbourhood_group	count
---------------------	-------

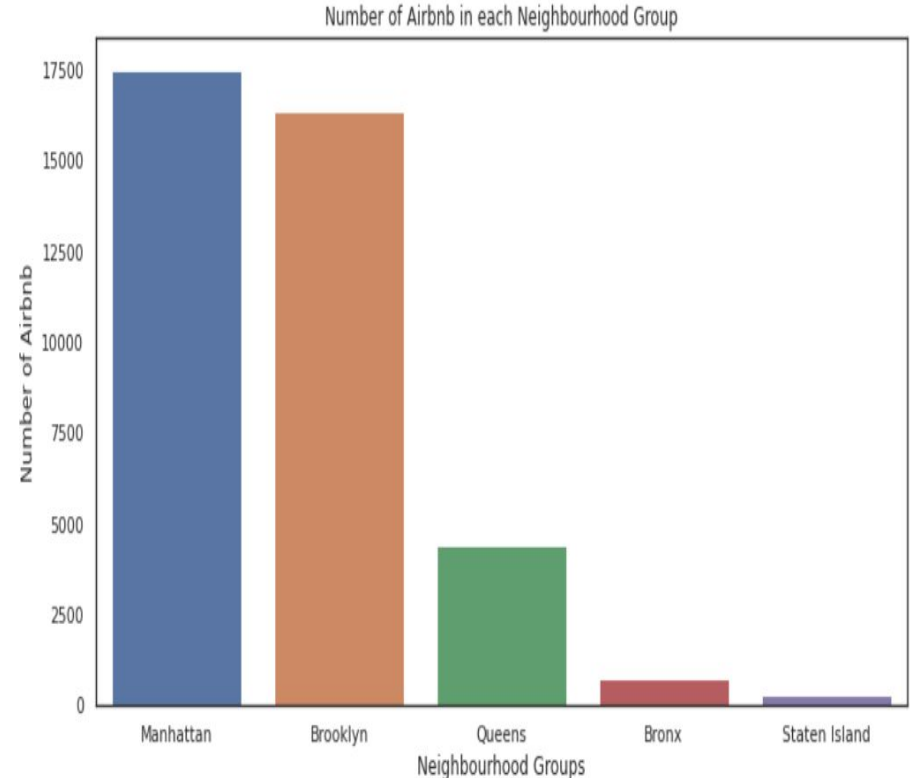
Manhattan	17513
-----------	-------

Brooklyn	16373
----------	-------

Queens	4413
--------	------

Bronx	758
-------	-----

Staten Island	278
---------------	-----



## 2B. Popularity of neighbourhood groups

neighbourhood_group	number_of_reviews
---------------------	-------------------

Manhattan	607
-----------	-----

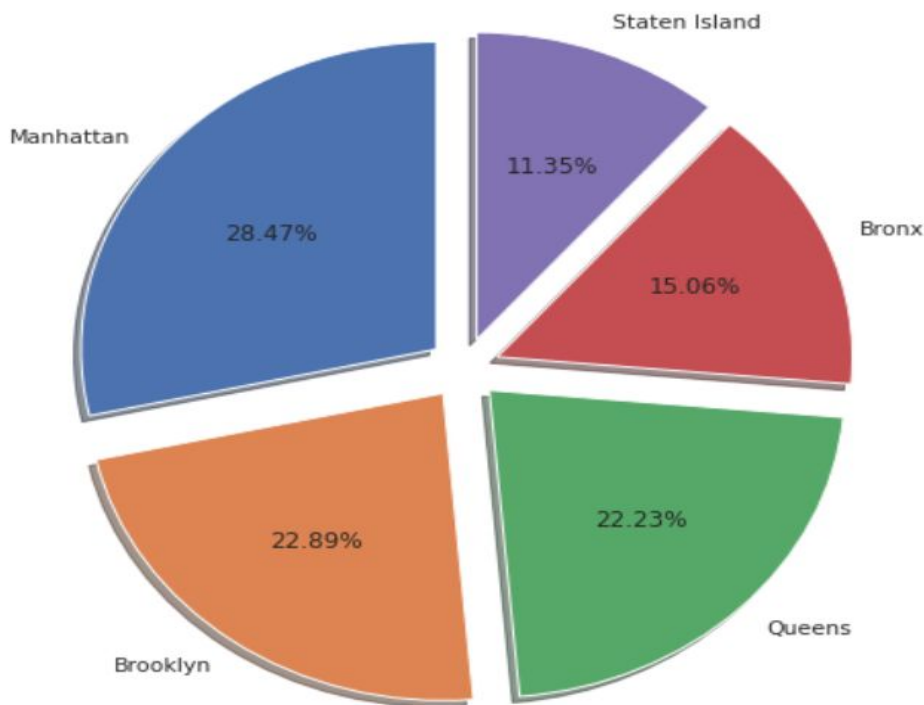
Brooklyn	488
----------	-----

Queens	474
--------	-----

Bronx	321
-------	-----

Staten Island	242
---------------	-----

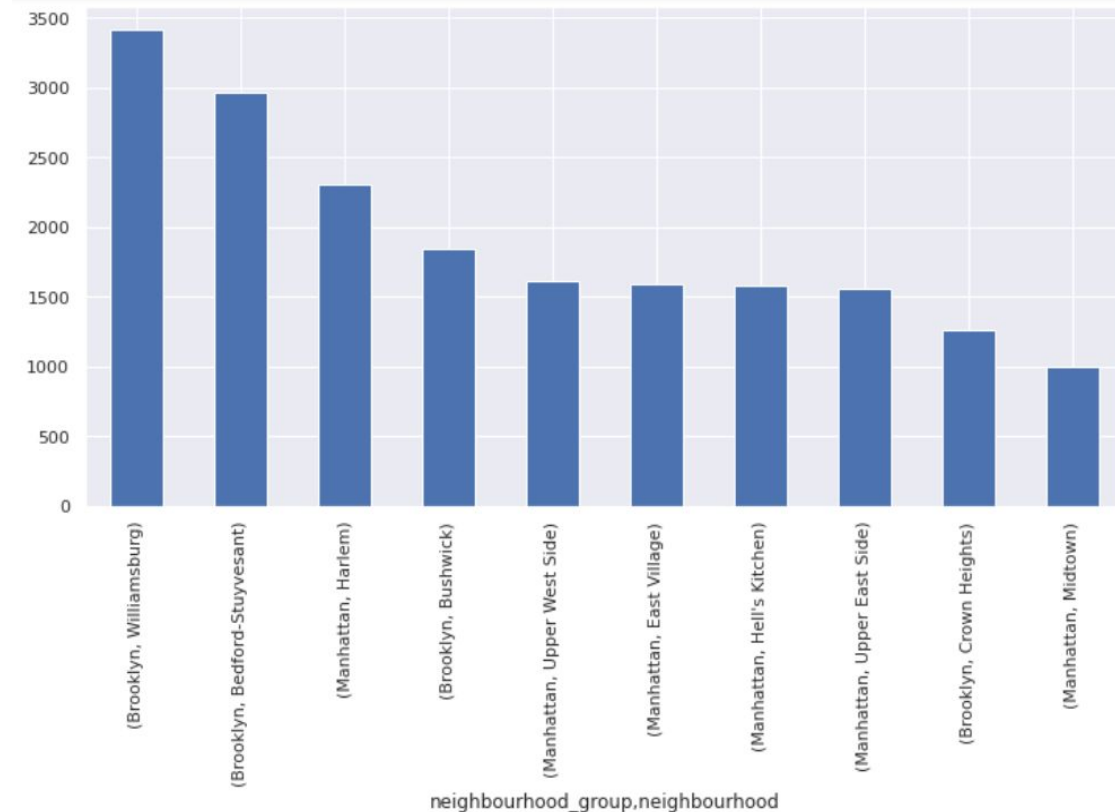
Number of reviews in each neighbourhood group



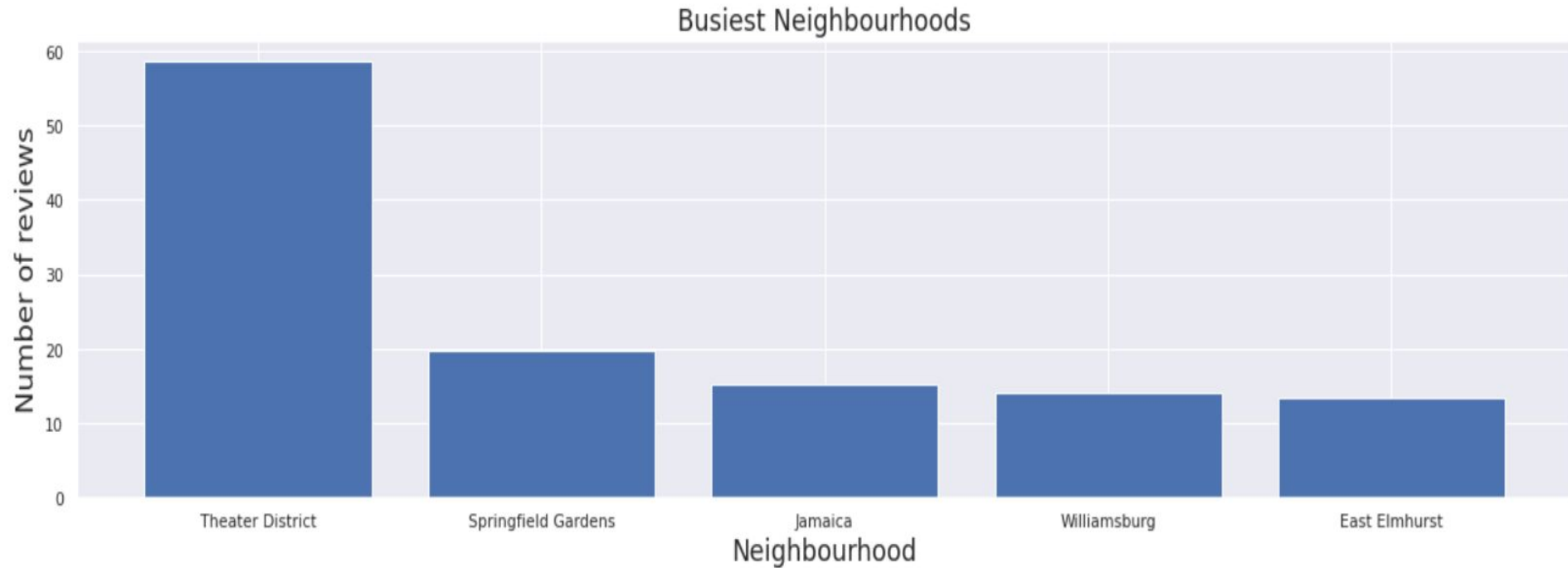


## 3A. Neighbourhoods with most listings.

neighbourhood_group	neighbourhood	
Brooklyn	Williamsburg	3415
	Bedford-Stuyvesant	2963
Manhattan	Harlem	2311
Brooklyn	Bushwick	1846
Manhattan	Upper West Side	1619
	East Village	1590
	Hell's Kitchen	1576
	Upper East Side	1561
Brooklyn	Crown Heights	1257
Manhattan	Midtown	1002

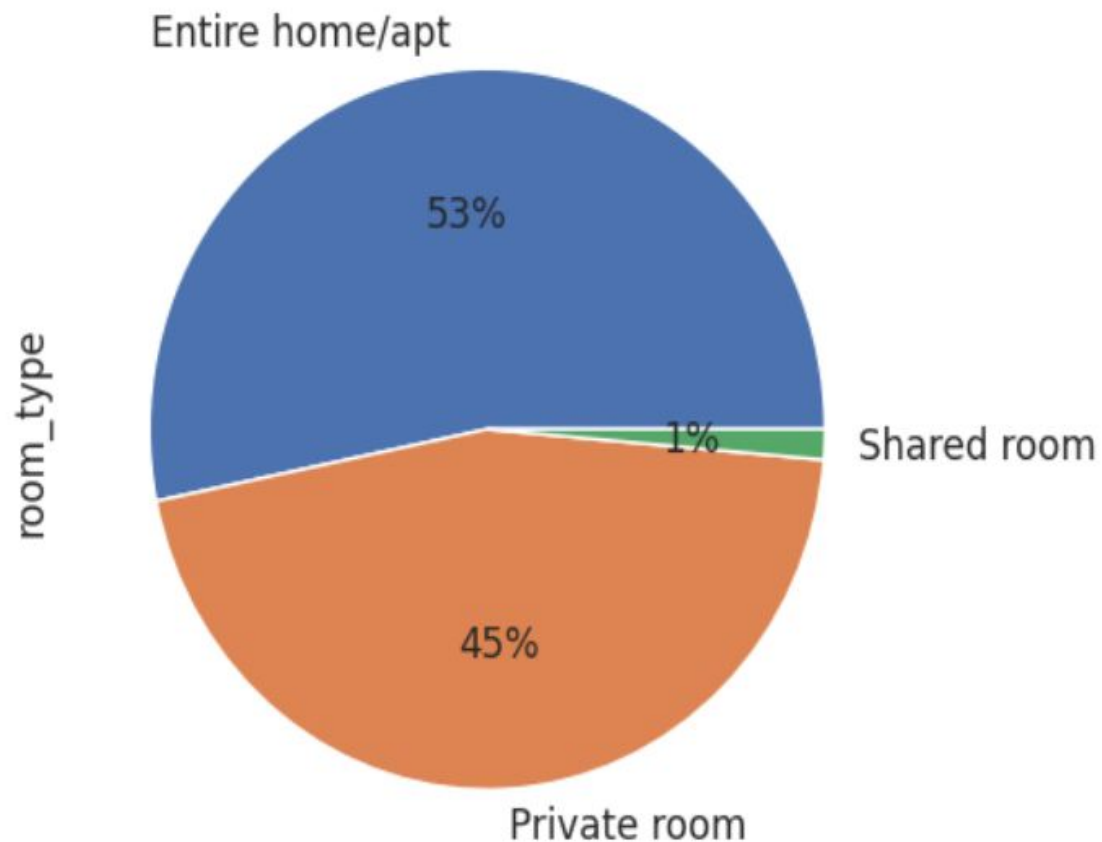


## 3B. Busiest and most popular neighbourhoods



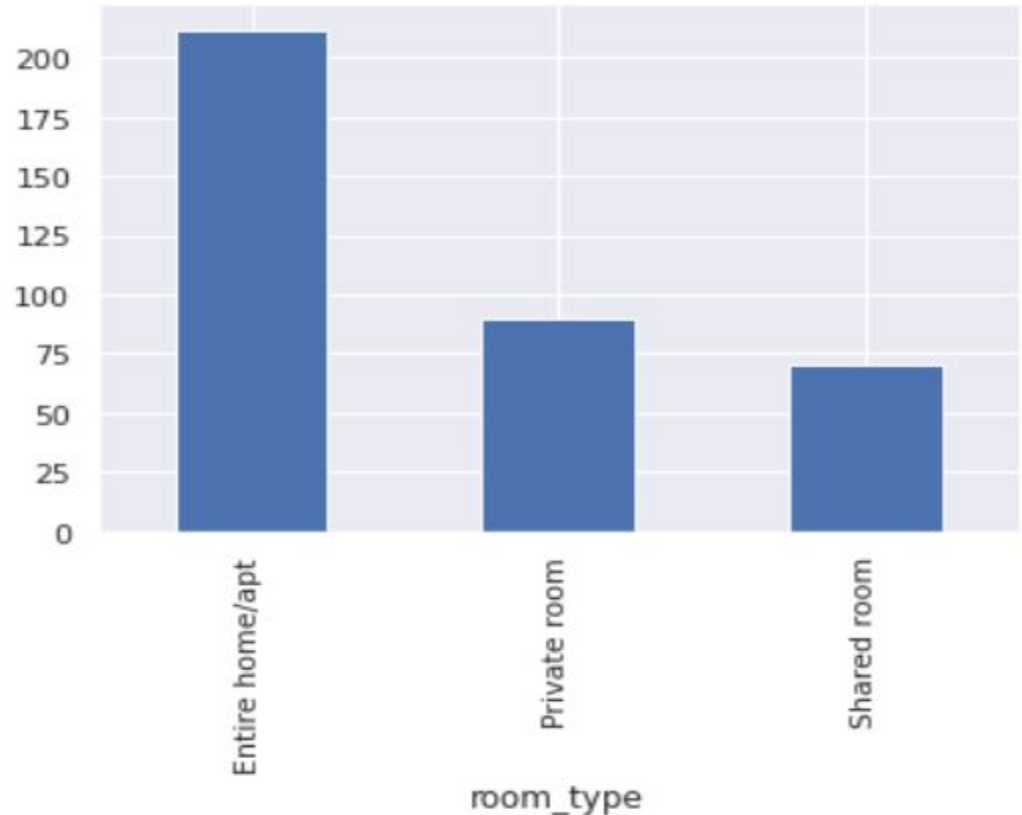
## 4A. Room types availability across NYC.

Entire home/apt	20926
Private room	17877
Shared room	532



## 4B. Average cost comparison of room types

room_type	
Entire home/apt	211.794246
Private room	89.780973
Shared room	70.127586

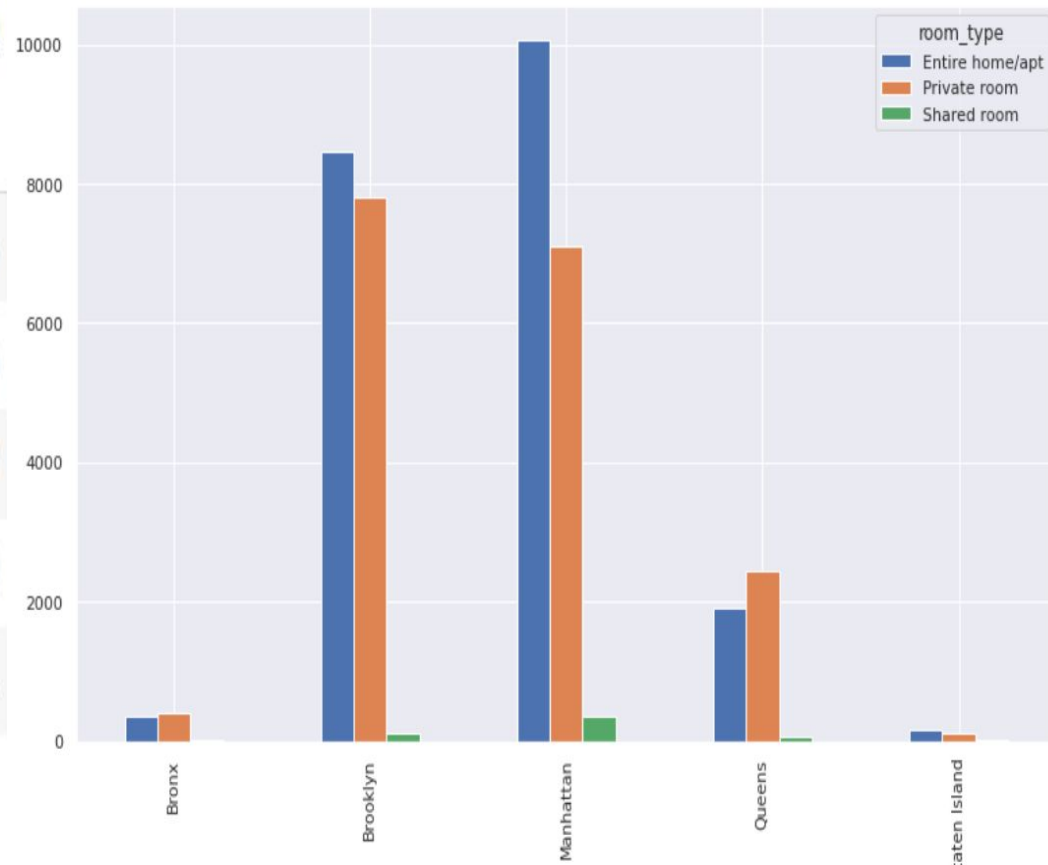


## 4C. Room type availability

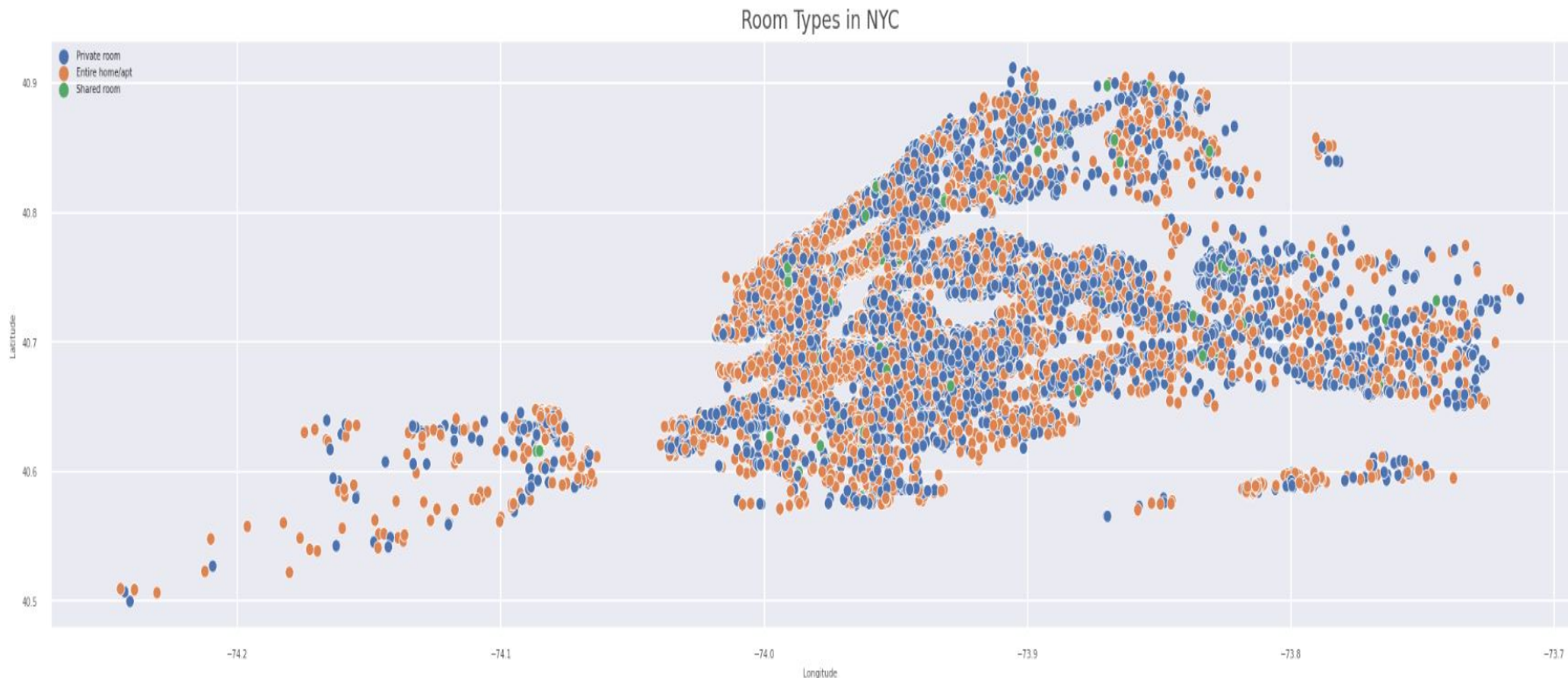
room\_type Entire home/apt Private room Shared room

neighbourhood\_group

Bronx	342	393	23
Brooklyn	8456	7814	103
Manhattan	10054	7109	350
Queens	1917	2444	52
Staten Island	157	117	4



## 4D. Location wise room types representation



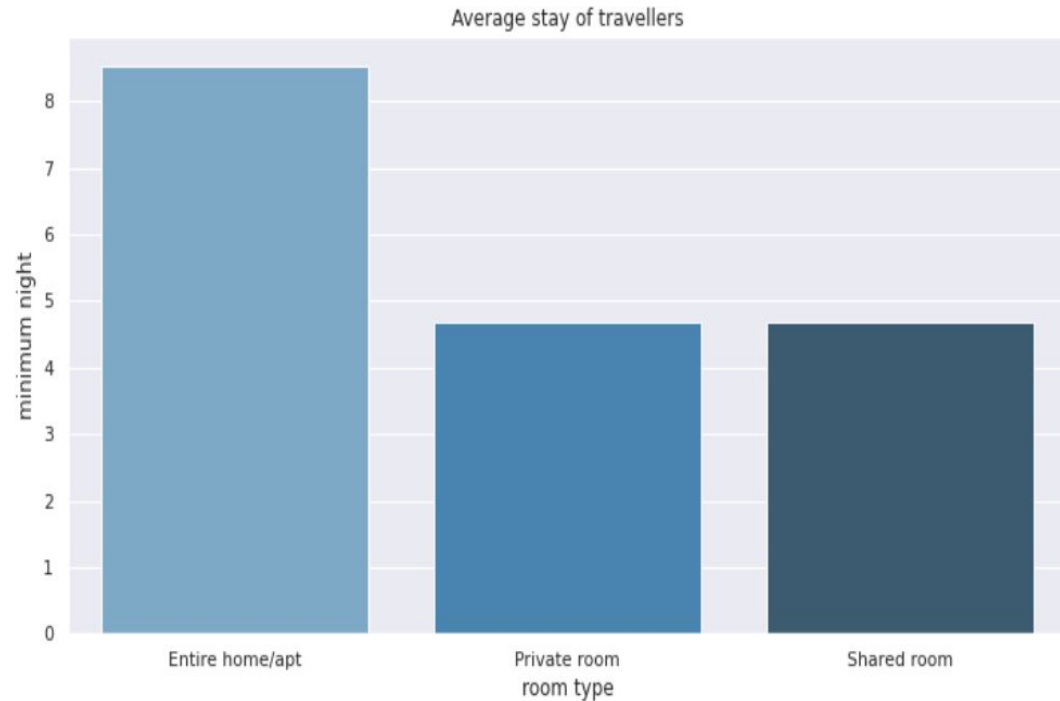
## 4E. Average occupancy of different room types

	room_type	minimum_nights
--	-----------	----------------

0	Entire home/apt	8.536462
---	-----------------	----------

1	Private room	4.680595
---	--------------	----------

2	Shared room	4.678571
---	-------------	----------



## 4F. Average room charges

neighbourhood_group	price
---------------------	-------

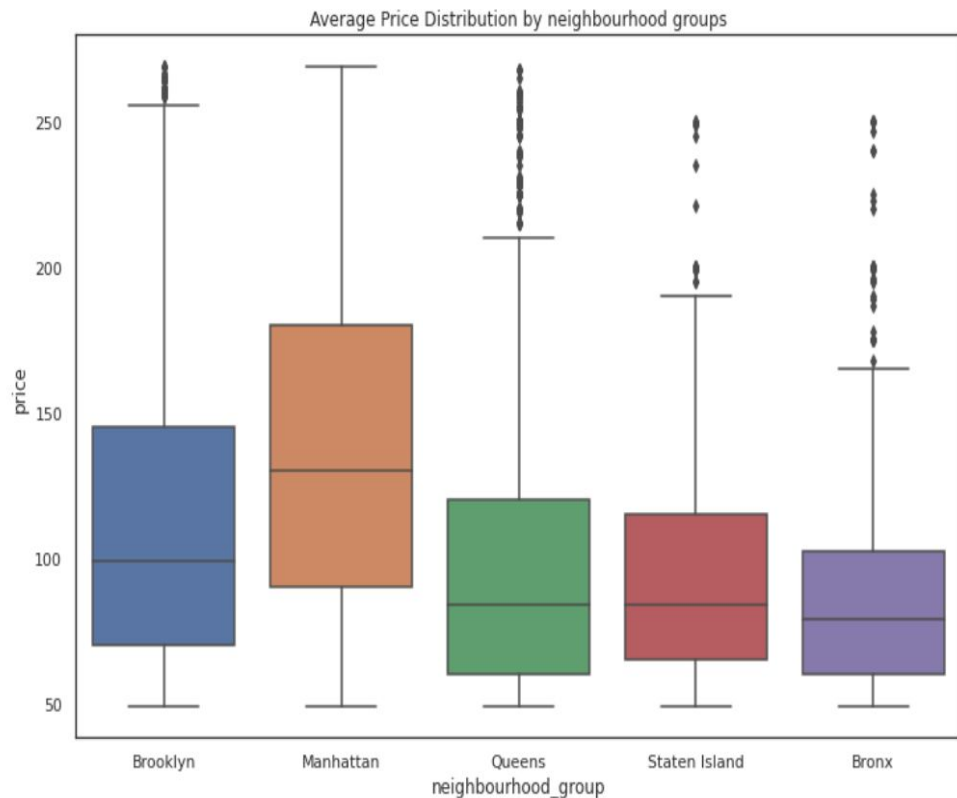
Manhattan	138.378804
-----------	------------

Brooklyn	110.887620
----------	------------

Queens	97.330614
--------	-----------

Staten Island	96.798561
---------------	-----------

Bronx	89.456464
-------	-----------

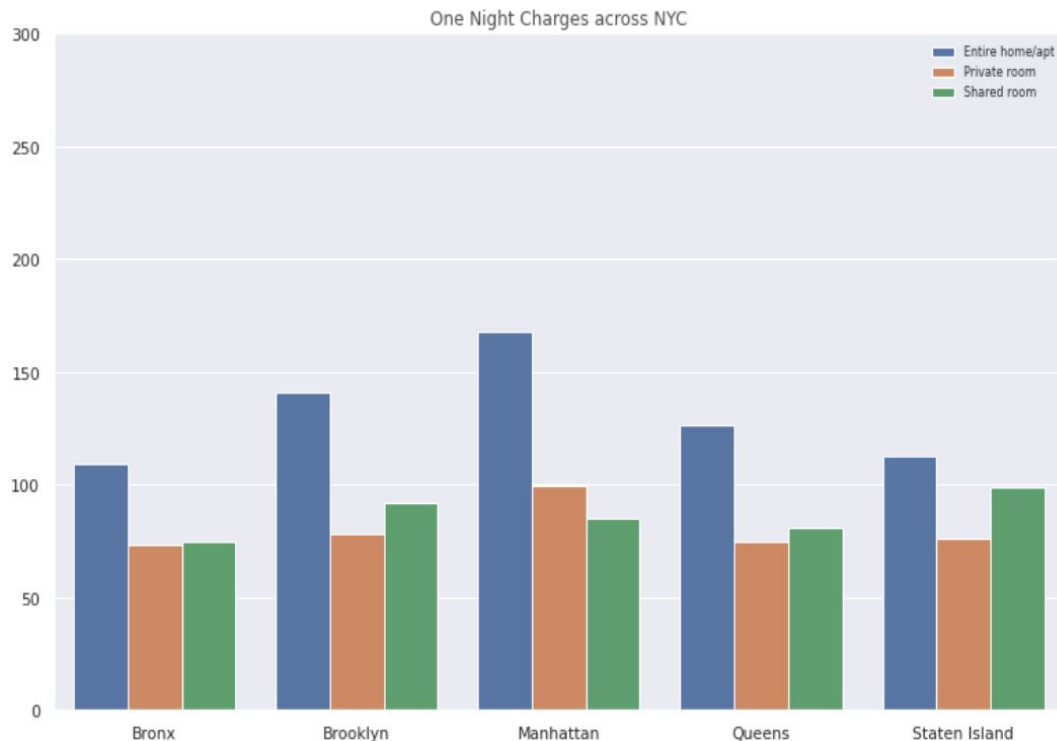




## 4G. Room charges of different room types

```
room_type Entire home/apt Private room Shared room
neighbourhood_group
```

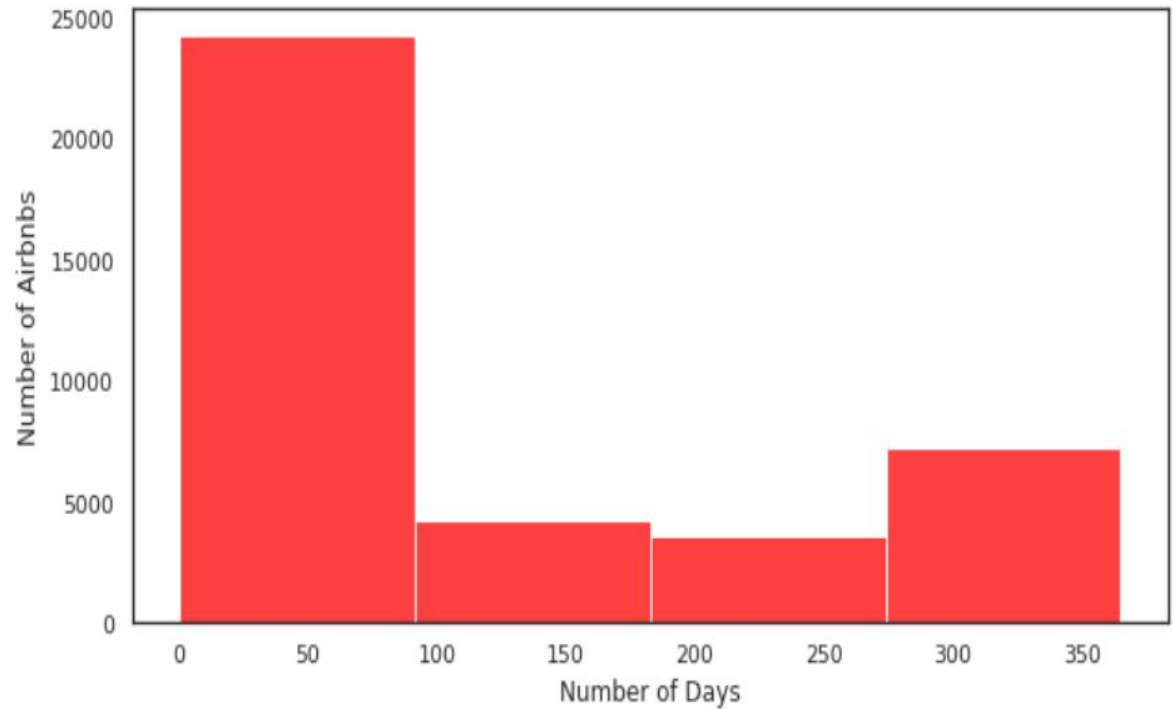
Bronx	109.040936	73.282443	74.608696
Brooklyn	141.084910	78.462759	91.660194
Manhattan	167.890292	99.263047	85.137143
Queens	126.232134	75.010229	80.923077
Staten Island	112.331210	75.888889	98.750000



## 5A. Properties availability- days/year.

Defaulter Property Count  
**24598**

Around 60% of airbnbs are open for less than 100 days across the year. This is alarming to companies reputation and companies growth.



## 5B. Properties operating less than 100 days.

neighbourhood_group	count
---------------------	-------

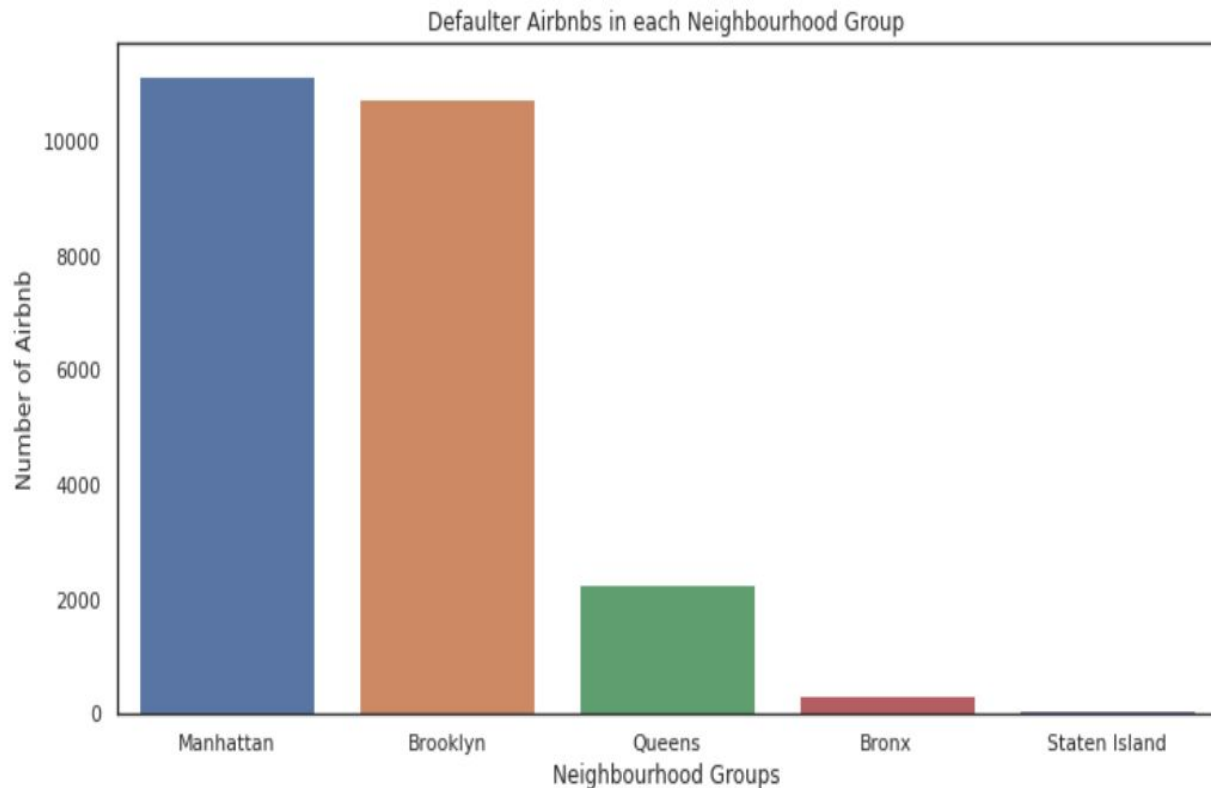
Manhattan	11161
-----------	-------

Brooklyn	10753
----------	-------

Queens	2257
--------	------

Bronx	333
-------	-----

Staten Island	94
---------------	----



# Key Observations for decision making

- 1** - Most number of Airbnb are present in Manhattan followed by Brooklyn, 86-87% of Airbnb revenue is generated from these two neighbourhood groups. Least number of Airbnb are present in Queens, Bronx and Staten Island, which are not even able to generate 15% of the company revenue.
- 2** - Manhattan followed by Brooklyn and Queens are the costliest neighbourhood groups because of their popularity. Whereas Bronx followed by Staten Island are the cheapest of the neighbourhood groups.
- 3** - As a result of high popularity and demand entire home room type is the costliest followed by private and shared rooms. But availability of shared rooms are just 1% of entire listings in NYC.
- 4** - Around 60% of total airbnbs are open for less than 100 days across the year. This is alarming to companies reputation and companies growth.
- 5** - Though most of the listings of airbnbs and revenue generated from Manhattan and Brooklyn, but percentage of defaulters( operating less than 100 days/ year) is very high which is alarming to companies growth and reputation. Actions needs to be taken to cope up with this scenario.

# Conclusions -



## 1. Hosts eligible for airbnb awards for the year 2019.

### A. Host with highest number of listings on Airbnb

Sonder(NYC) with 327 properties.

### B. Top 3 most popular host in entire NYC.

Row Nyc, Nalicia, Brent

### C. Top 3 most popular host in Manhattan.

Row Nyc, Ben, Jae

### D. Top 3 most popular host in Brooklyn.

Alina, Stephanie, Meeks

### E. Top 3 most popular host in Bronx.

Luis, Yoselyn, Sharon & Erika

### F. Top 3 most popular host in Queens.

Nalicia, Brent, Lakshmee

### G. Top 3 most popular host in Staten Island.

Wendy, Linda, Yun

## 2.Popularity on various parameters

### A.Popularity of Neighbourhood groups

Manhattan > Brooklyn > Queens > Bronx > Staten Island

### B.Popularity of Neighbourhoods

Theater District > Springfield Gardens > Jamaica > Williamsburg > East Elmhurst

### C.Room Type popularity based on average stay

Entire home/apt>Private room>Shared room

## 3.Availability on various parameters

### A.Neighbourhood groups with most listings

Manhattan > Brooklyn > Queens > Bronx > Staten Island

### B.Neighbourhoods with most listings

Williamsburg(Brooklyn)>Bedford-Stuyvesant(Brooklyn)>Harlem(Manhattan)>Bushwick(Brooklyn)>Upper West Side(Manhattan)

### C.Room types with most listings

Entire home/apt>Private room>Shared room

## 4.Price comparison on various parameters

### A.Price comparison of rooms types

Entire home/apt>Private room>Shared room

### B.Price comparison of rooms across neighbourhood groups

Manhattan > Brooklyn > Queens > Staten Island > Bronx

## 5.Defaulters(available less than 100 days/year.)

### A.Total defaulters comparison across neighbourhood groups

Manhattan > Brooklyn > Queens > Bronx > Staten Island

# Q&A