Capstone Project-1
Airbnb Booking Analysis

Team members
1. Sai Venkata Ravi Teja Ande
2.Sunil Kumar Bodugu
3.Amogha K

Content:

AI

1. Introduction to Airbnb

2. General overview of Dataset

3. Data cleaning

4. Exploratory Data Analysis

5. Conclusion

1.Introduction to Airbnb

- ❖ **Airbnb stands for the Air, Bed, and Breakfast.**

- ❖ **Airbnb is an American company that operates an online marketplace for lodging, primarily homestays for vacation rentals, and tourism activities. It is based in San Francisco, California, the platform is accessible via website and mobile app.**

- ❖ **Airbnb provides a platform to rent the people's property or spare room to guest and Airbnb don't have any own property.**

- ❖ **Prices are set by the property owners and payments are collected via the Airbnb app. The majority of its revenue comes from service fees from bookings charged to both guests and hosts.**

- ❖ **Airbnb operates in more than 65000 cities and 191 countries.**

- ❖ **New York City is the most densely populated city in the United States, and one of the most popular tourism and business places globally.**

2.General overview of Dataset

➢ **Dataset (Airbnb NYC 2019.csv) for Analysis is belong to New York City.**

➢ **The analysis is focused on the boroughs of Brooklyn, Manhattan, Queens, Staten Island, and the Bronx. These Boroughs are further divided into distinctive neighborhoods.**

➢ **Data set has total 48,895 rows with 16 columns.**

Column name and details:

1. **id : Unique listing id.**

2. **name : Name of the property .**

3. **host_id : unique id for each listed host.**

4. **host_name : Name of the host.**

5. **neighbourhood_group : Location.**

6. **neighborhood : Area.**

7. **latitude : Latitude coordinates.**

8. **longitude : Longitude coordinates.**

Column name and details:

9.room_type : Listing space types.

10.price : Price in dollars($).

11.minimum_nights : Minimum nights required to stay.

12.number_of_reviews : No. of reviews recived for the listing.

13.last_review : Last reviewed date for the listing.

14.reviews_per_month : Total review per month for the listing.

15.calculated_host_listings_count : Total no of listing against the host id.

16.availability_365 : Number of days when listing is available for booking.

3.Data cleaning

AI

head - gives top 5 rows of all features

`[ ]  df.head(5)S`

	id	name	host_id	host_name	neighbourhood_group	neighbourhood	latitude	longitude	room_type	price	minimum_nights	number_of_reviews	last_review	reviews_per_month
0	2539	Clean & quiet apt home by the park	2787	John	Brooklyn	Kensington	40.64749	-73.97237	Private room	149	1	9	19-10-2018	0.21
1	2595	Skylit Midtown Castle	2845	Jennifer	Manhattan	Midtown	40.75362	-73.98377	Entire home/apt	225	1	45	21-05-2019	0.38
2	3647	THE VILLAGE OF HARLEM....NEW YORK !	4632	Elisabeth	Manhattan	Harlem	40.80902	-73.94190	Private room	150	3	0	NaN	NaN
3	3831	Cozy Entire Floor of Brownstone	4869	LisaRoxanne	Brooklyn	Clinton Hill	40.68514	-73.95976	Entire home/apt	89	1	270	05-07-2019	4.64
4	5022	Entire Apt: Spacious Studio/Loft by central park	7192	Laura	Manhattan	East Harlem	40.79851	-73.94399	Entire home/apt	80	10	9	19-11-2018	0.10

tail - gives bottom 5 rows of all features

`[ ]  df.tail(5)`

	id	name	host_id	host_name	neighbourhood_group	neighbourhood	latitude	longitude	room_type	price	minimum_nights	number_of_reviews	last_review	reviews_per_month
48890	36484665	Charming one bedroom - newly renovated rowhouse	8232441	Sabrina	Brooklyn	Bedford-Stuyvesant	40.67853	-73.94995	Private room	70	2	0	NaN	NaN
48891	36485057	Affordable room in Bushwick/East Williamsburg	6570630	Marisol	Brooklyn	Bushwick	40.70184	-73.93317	Private room	40	4	0	NaN	NaN
48892	36485431	Sunny Studio at Historical Neighborhood	23492952	Ilgar & Aysel	Manhattan	Harlem	40.81475	-73.94867	Entire home/apt	115	10	0	NaN	NaN
48893	36485609	43rd St. Time Square-cozy single bed	30985759	Taz	Manhattan	Hell's Kitchen	40.75751	-73.99112	Shared room	55	1	0	NaN	NaN
48894	36487245	Trendy duplex in the very heart of Hell's Kitchen	68119814	Christophe	Manhattan	Hell's Kitchen	40.76404	-73.98933	Private room	90	7	0	NaN	NaN

- **Finding null values in the data.**

- **Finding the percentage of null values.**

- **using head function to see the top rows of the data**

- **Using tail function to see last rows of the data.**

Finding null values in all features

```
df.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
```

Finding percentage of null values

```
[ ]  missing_values_per =df.isnull().sum()*100/len(df)
```

```
[ ]  missing_values_per
```

```
id                              0.000000
name                            0.032723
host_id                         0.000000
host_name                       0.042949
neighbourhood_group             0.000000
neighbourhood                   0.000000
latitude                        0.000000
longitude                       0.000000
room_type                       0.000000
price                           0.000000
minimum_nights                  0.000000
number_of_reviews               0.000000
last_review                    20.558339
reviews_per_month              20.558339
calculated_host_listings_count  0.000000
availability_365                0.000000
dtype: float64
```

```python
mean_1 = df['reviews_per_month'].mean()
```

```python
df['reviews_per_month'].fillna(mean_1, inplace=True)
```

```python
df.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                   0
calculated_host_listings_count      0
availability_365                    0
dtype: int64
```

```python
df['reviews_per_month']
```

```
0          0.210000
1          0.380000
2          1.373221
3          4.640000
4          0.100000
             ...
48890      1.373221
48891      1.373221
48892      1.373221
48893      1.373221
48894      1.373221
Name: reviews_per_month, Length: 48895, dtype: float64
```

```python
df['reviews_per_month'].isnull().sum()
```

```
0
```

Missing_values_df

col_name	% of_null _values
id	id	0.000000
name	name	0.032723
host_id	host_id	0.000000
host_name	host_name	0.042949
neighbourhood_group	neighbourhood_group	0.000000
neighbourhood	neighbourhood	0.000000
latitude	latitude	0.000000
longitude	longitude	0.000000
room_type	room_type	0.000000
price	price	0.000000
minimum_nights	minimum_nights	0.000000
number_of_reviews	number_of_reviews	0.000000
last_review	last_review	20.558339
reviews_per_month	reviews_per_month	20.558339
calculated_host_listings_count	calculated_host_listings_count	0.000000
availability_365	availability_365	0.000000

4.Exploratory Data Analysis

1. TOP 5 property name with maximum number of Reviews.

- Room near JFK Queen Bed has highest review from customers.
- Manhattan neighborhood group have more good reviews.
- Top reviews can considered as the good service provider for the customers.

hotel_name	max_reviews
Room near JFK Queen Bed	629
Great Bedroom in Manhattan	607
Beautiful Bedroom in Manhattan	597
Private Bedroom in Manhattan	594
Room Near JFK Twin Beds	576

2. Which group has good reviews on each neighbourhood.

neighbourhood_group	number_of_reviews
Queens	629
Manhattan	607
Brooklyn	488
Staten Island	333
Bronx	321

Good reviews on each neighbourhood groups
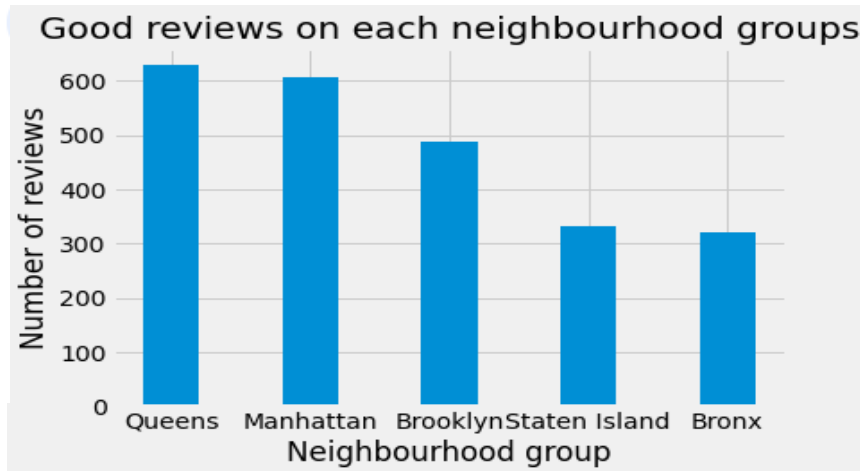
- Queen neighbourhood group has 629 reviews and it is highest reviews among other neighbourhood groups.

3. what will be the maximum and average price of all room type?

Maximum price				Average Price	
	price				price
room_type				room_type	
Entire home/apt	10000			Entire home/apt	211.79
Private room	10000			Private room	89.78
Shared room	1800			Shared room	70.13

➤ Entire home/apt has highest both in average and maximum i.e. $211.79 and $10,000

➤ Shared room has low price both in average and maximum but most of the customer will not prefer for shared room.

4.Which hosts are the busiest and why?

❑ These hosts are busy because they are charging minimum price , having good reviews and providing good services to the customers. So customer will prefer for these hosts.

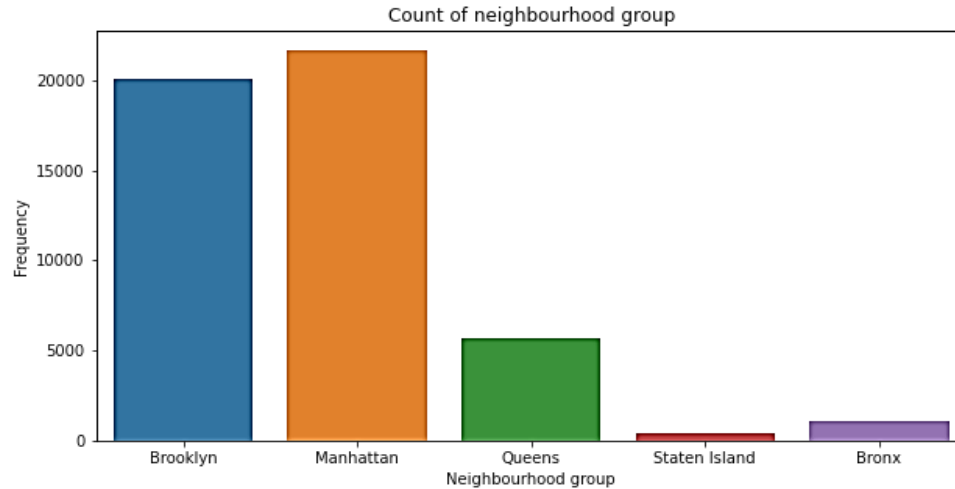❑ Most of the room type in busiest list are private room type and customers will like to stay.
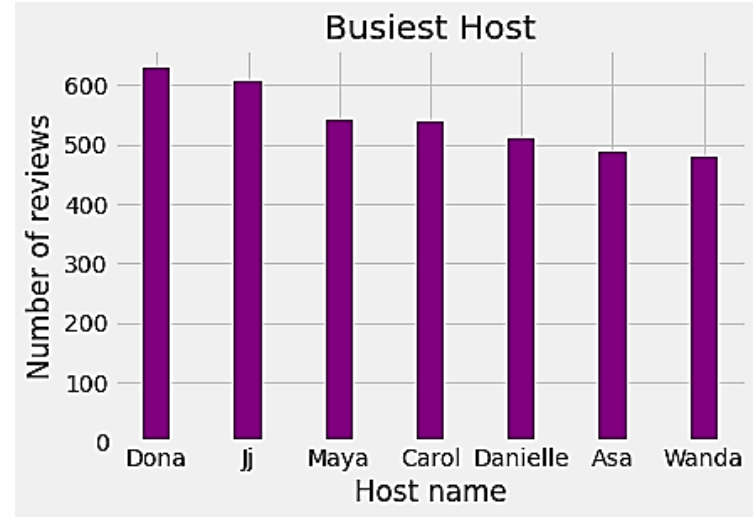
**5.Scatter Plot to show the neighbourhood group based on Latitude and Longitude.**
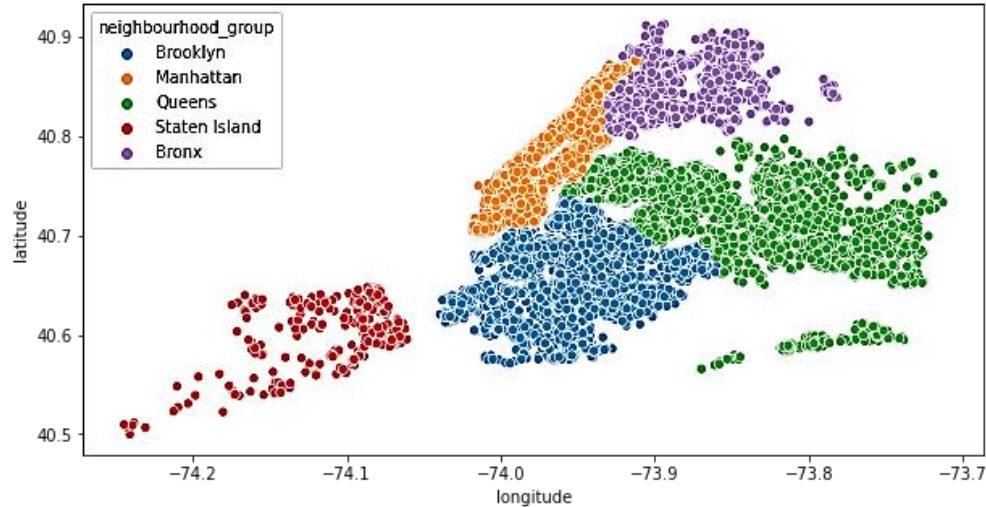
❑ Using latitude and longitude points were able to view all NYC listings in Airbnb.

Bookings of Neighbourhood group.

Count of neighbourhood group

- ➢ Manhattan has the highest bookings with 21661.
- ➢ Brooklyn has the second-highest with 20104 bookings, Queens has the third highest bookings with 5666.
- ➢ The Bronx and Staten Island trail the pack with 1091 and 373 bookings.

6.Average booking prices in neighbourhood group.

Average price vs Neighbourhood group
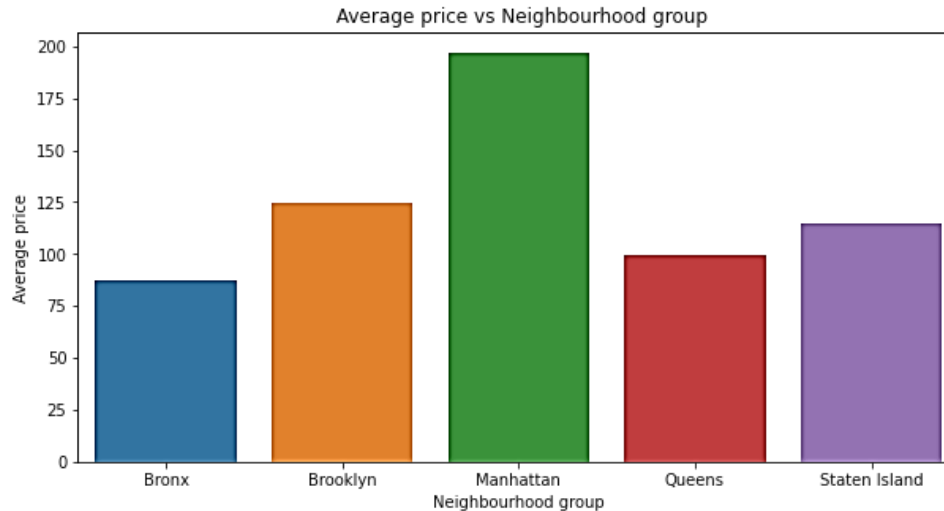
neighbourhood_group	average price
Bronx	87.50
Brooklyn	124.38
Manhattan	196.88
Queens	99.52
Staten Island	114.81

✓ Manhattan is having highest booking price and most expensive borough ,followed by Brooklyn.

✓ Bronx is having lowest average booking price and affordable for customers.

Count of Neighbourhood

Count of Neighbourhood
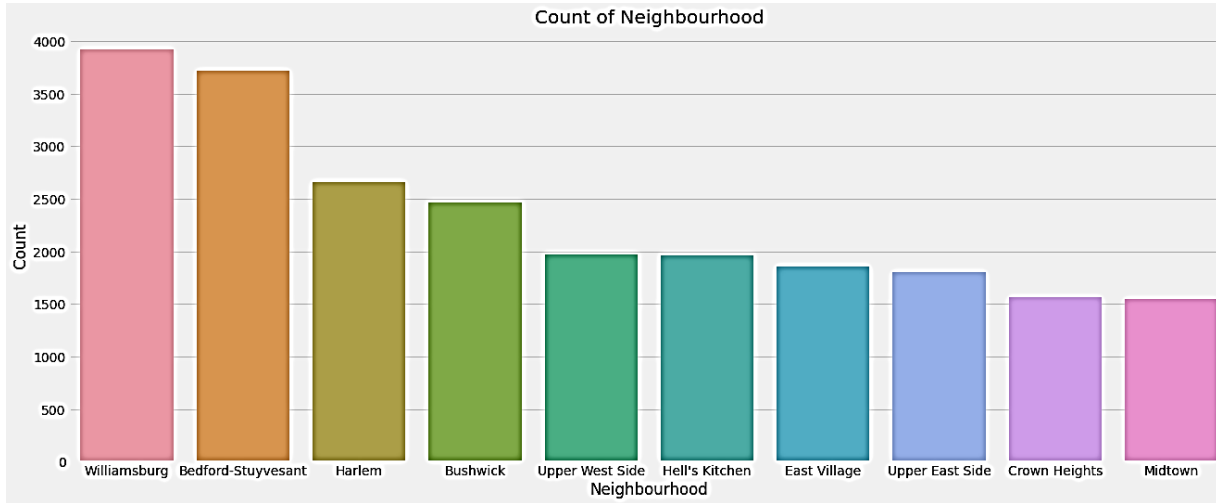
| neighbourhood	counts
Williamsburg	3920
Bedford-Stuyvesant	3714
Harlem	2658
Bushwick	2465
Upper West Side	1971
Hell's Kitchen	1958
East Village	1853
Upper East Side	1798
Crown Heights	1564
Midtown	1545

✓ There are 221 unique neighbourhoods.

✓ Williamsburg is the top neighborhood with 3920 bookings.

✓ The above plot shows us the top 10 neighbourhood towns where customers were interested to stay.

Room/property type.

| Room type	count of no of rooms
Entire home/apt	25409
Private room	22326
Shared room	1160
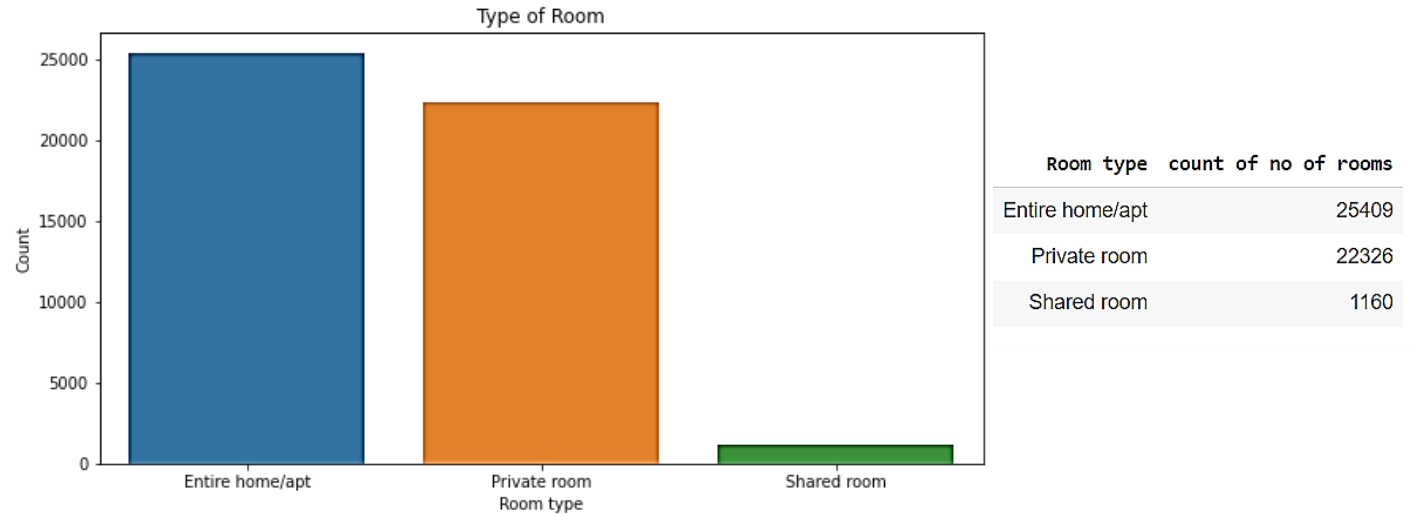
- o Majority of customers were interested in booking entire home/apt and Private room
- o As expected very few customers were interested in sharing room type.

Room types occupied by a neighbourhood group
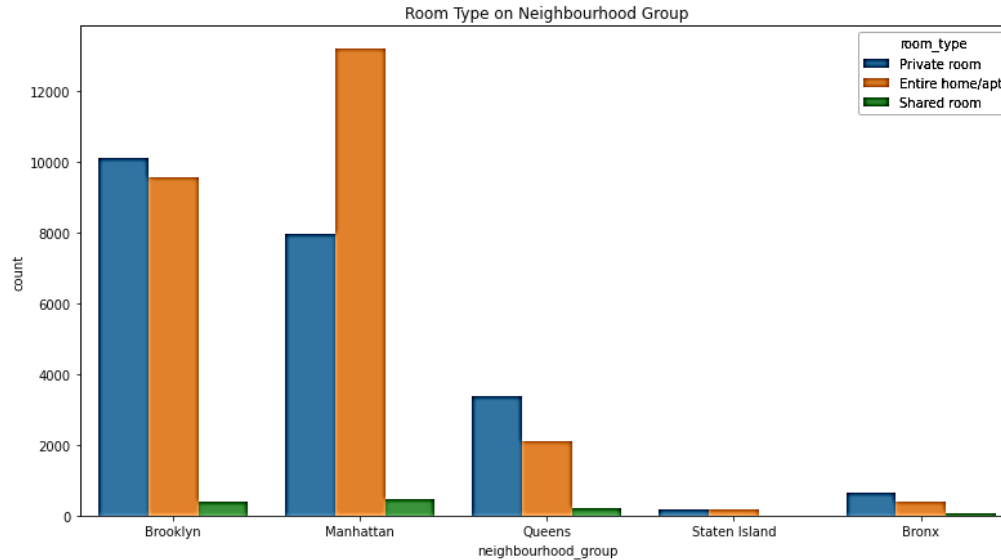
❑ In all 5 NYC boroughs except Manhattan, majority of people are occupied of Private room type.

❑ In Manhattan, Entire home/apt type property is highest occupied by people.

Price variations of the room types.

Price vs Room Type
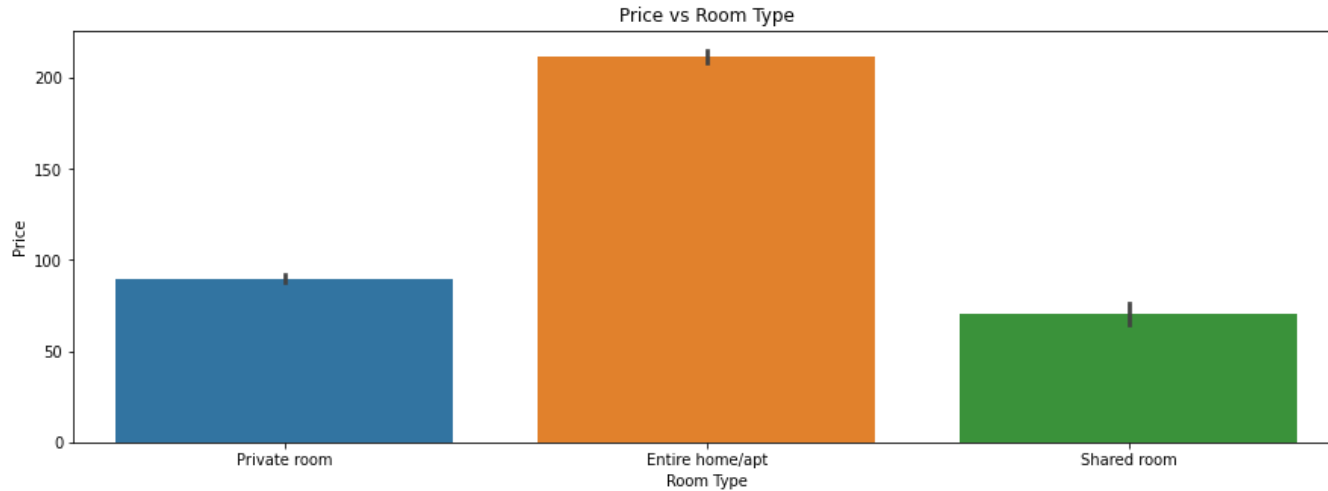
- Despite being costlier than other room types 'Entire home/apt', majority of customers were interested to choose this.

Maximum and minimum revenue generated by each room type in the neighbors.

```
The maximum revenue generated by Entire home/apt is
 room_type       Entire home/apt  Private room  Shared room
neighbourhood
Williamsburg            389724.0      171265.0       2718.0
-----------------------------------------------------------------
The minimum revenue generated by Entire home/apt is
 room_type       Entire home/apt  Private room  Shared room
neighbourhood
New Dorp                   57.0           NaN          NaN

******************************************************************

The maximum revenue generated by private room is
 room_type       Entire home/apt  Private room  Shared room
neighbourhood
Williamsburg            389724.0      171265.0       2718.0
-----------------------------------------------------------------
The minimum revenue generated by private room  is
 room_type       Entire home/apt  Private room  Shared room
neighbourhood
Graniteville              186.0          20.0          NaN

******************************************************************

The maximum revenue generated by Shared room is
 room_type       Entire home/apt  Private room  Shared room
neighbourhood
Hell's Kitchen          301786.0       89713.0       9488.0
-----------------------------------------------------------------
The minimum revenue generated by Shared room  is
 room_type       Entire home/apt  Private room  Shared room
neighbourhood
Randall Manor            5859.0         512.0         13.0
```
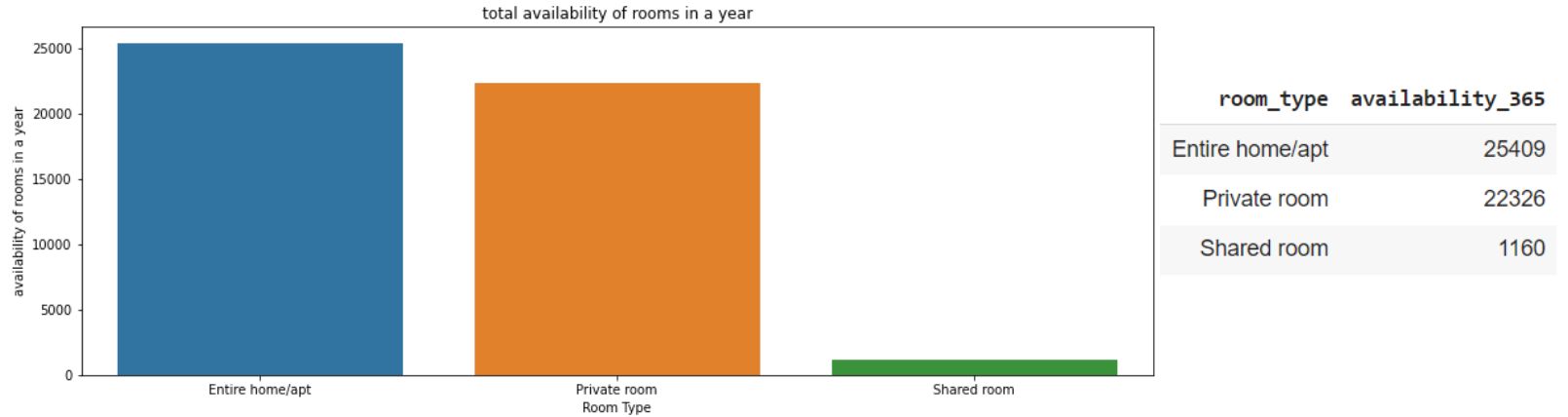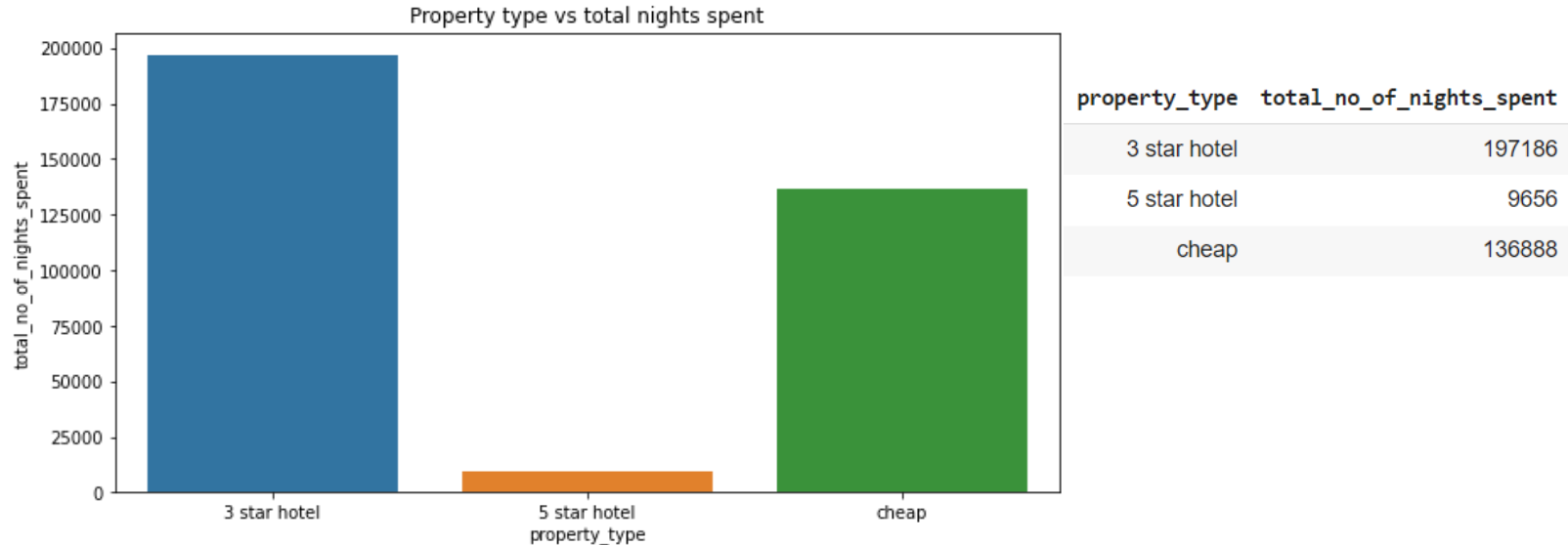
So from outputs, the neighbourhood 'Williamsburg' has generated highest revenue for Entire home/apt and Private room types. The reason behind this may be the presence of tourist places near this neighbourhood.

Type of rooms are mostly available through out the year.

total availability of rooms in a year

| room_type	availability_365
Entire home/apt	25409
Private room	22326
Shared room	1160

❖ By this analysis , Entire home/apt are more available through out the year and it is the most preferable room type by the customers in Manhattan.

❖ Shared room are less available through out the year because of less booking.

Classified the properties listed on Airbnb using price column.

Property type vs total nights spent

property_type	total_no_of_nights_spent
3 star hotel	197186
5 star hotel	9656
cheap	136888

➢ 3 star hotels has most demand in the NYC area when compared to other two room types.

➢ 3 star rooms are having average price compare to other room type.

➢ People uses the cheap room who will stay only for 1-2 days.

Top 5 Busiest Host.
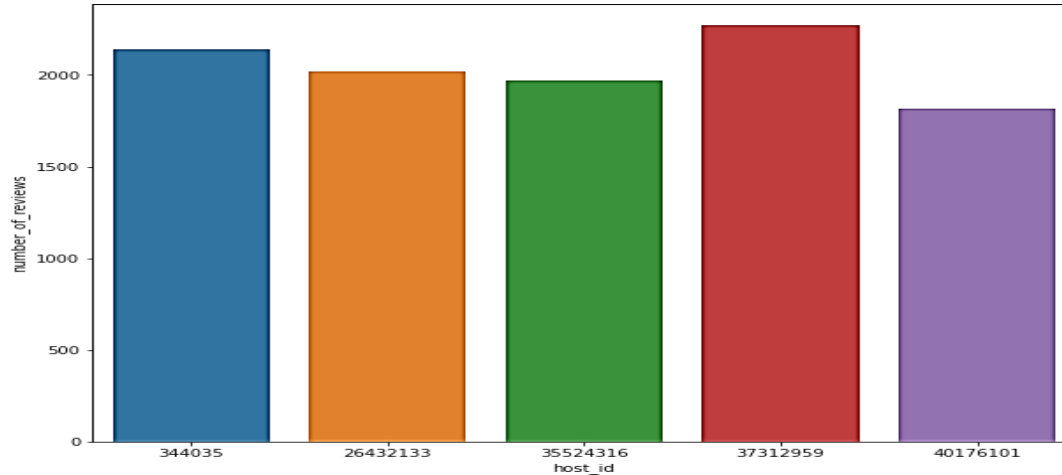
| host_id	number_of_reviews
37312959	2273
344035	2141
26432133	2017
35524316	1971
40176101	1818

Review is the one of the important criteria with online activity these days. This gives a lot of insights to a particular place for tourist and they can swing mood when it comes to online booking. So far the host with host id has 37312959 has more number reviews.

Booking price range for each neighbourhood group.
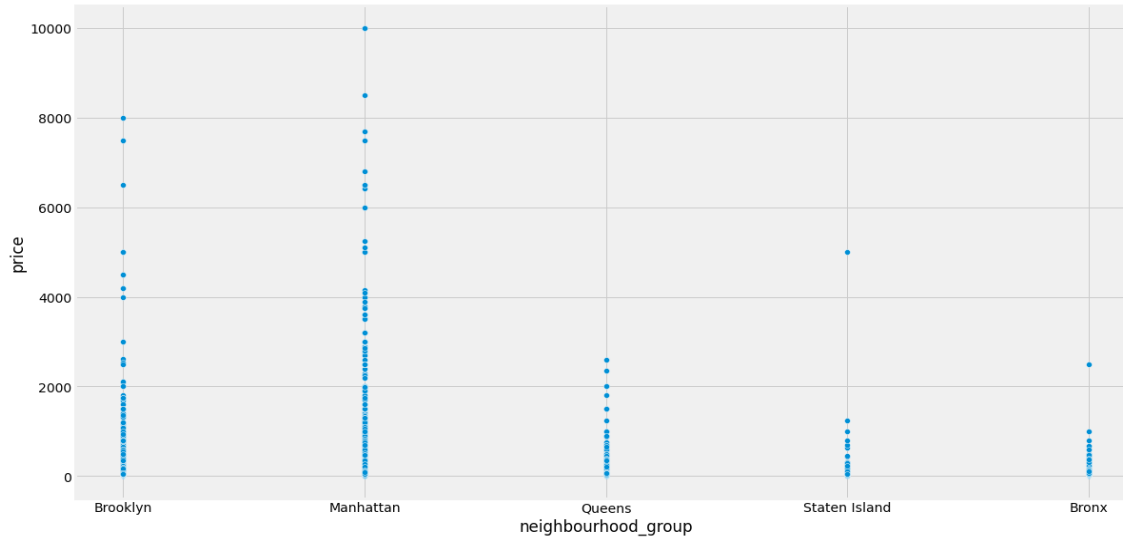
- ❖ 'Manhattan' neighbourhood has maximum price so it is expensive from all other groups,
- ❖ 'Bronx' neighbourhood is the affordable property with less amount.

The average price set by the hosts who have more than 50 properties.

compare the average prices set by the hosts who have more than 50 properties with rest of them.
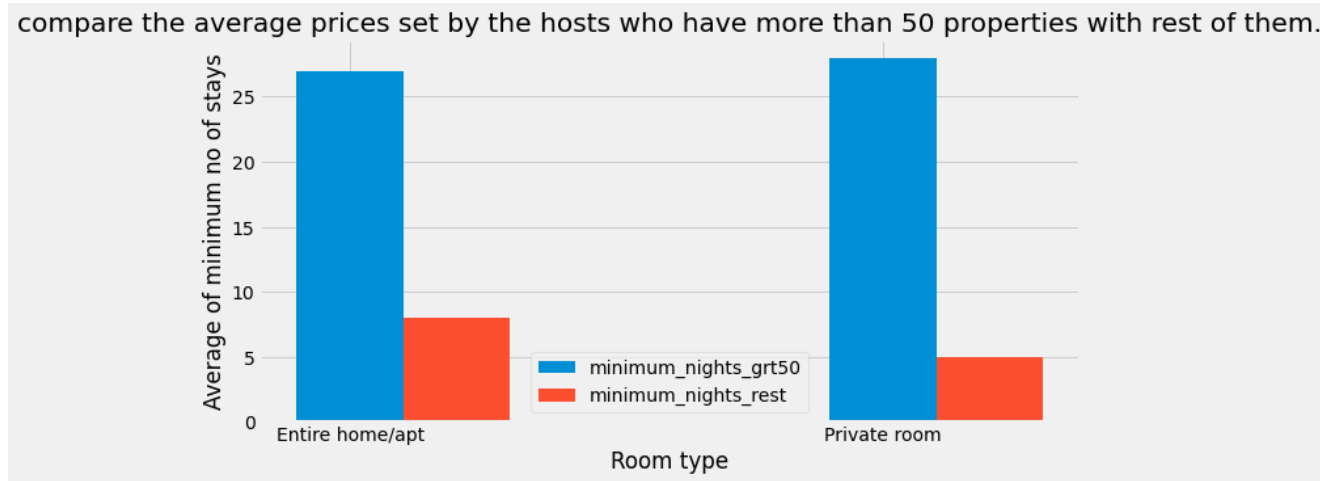
- By this analysis , one of the factor might be the minimum no of nights the customers are staying at their properties is way larger in comparison to rest.
- They might be providing their customers with large variety of complimentary facilities which will eventually forces the customer to stay there.

7.Oldest property.

propert_name	age_of_Property
Sweet and Spacious Brooklyn Loft	11.0
2 bedroom - Upper East Side-great for kids	10.0
BLUE TRIM GUEST HOUSE	10.0
Modern 1 BR / NYC / EAST VILLAGE	10.0
2 BR / 2 Bath Duplex Apt with patio! East Village	10.0

- Sweet and Spacious Brooklyn Loft is the oldest property in the hosting list.

5.Conclusion

❑ **Manhattan neighborhood has most listed Airbnb as compared to others.**

❑ **Most Airbnb users are either choose "Entire home/apt" or "Private room" type for stay.**

❑ **Revenue generated by "Entire home/apt" is highest irrespective of neighbourhood group.**

❑ **"Entire home/apt" type Airbnb in Manhattan is the most expensive Airbnb; while "Shared room" type Airbnb in Staten Island is the most affordable.**

❑ **Peoples doesn't like to stay which has room type as shared room.**

❑ **If people are not staying more than 1-2 days are might be travellers.**

❑ **Queens and Manhattan has most number of Visitors'**

❑ **Many Airbnb users require less than 5 nights of minimum stay. So, short term stays are more popular.**

AI

Thank You!