

# Capstone Project 1

## Airbnb Bookings Analysis

By:  
Ghanal Kaushik

# Summary

1. Problem Statement
2. Data Overview and preprocessing
3. Exploratory Data Analysis
4. Conclusion



# Problem Statement

Airbnb is an online platform that allows property owners to list their place as holiday accommodation and allows travelers to find a place to stay while they are away from home.

Data analysis on millions of listings provided through Airbnb is a crucial factor for the company. These millions of listings generate a lot of data - data that can be analyzed and used for security, business decisions, understanding of customers' and providers' (hosts) behavior and performance on the platform, guiding marketing initiatives, implementation of innovative additional services and much more.

# Data Overview

**Id**: Unique Id of each property

**name**: Name of properties, which also describes the property

**host\_id**: id of host, names maybe similar for two hosts

**neighbourhood\_group**: have 5 different regions of New York

**neighbourhood**: these are subregions of neighbourhood\_group

**latitude**: contains coordinates of latitude for that location

# Data Overview cont.

**longitude**: contains coordinates of longitude for that location

**room\_type**: there are three types of property available

**price**: price of stay (one night basis)

**minimum\_nights**: min number of nights you have to book host's place

**number\_of\_reviews**: the number of reviews a property receives

**last\_review**: date of last review that is posted for that property

# Data Overview cont.

reviews\_per\_month: number of reviews for each property per month

calculated\_host\_listings\_count: number of listings for each property

availability\_365: number of days each property is available out of 365 days of year

# Preprocessing

We decided to remove some less important columns like: 'name', 'host\_name', 'last\_review'

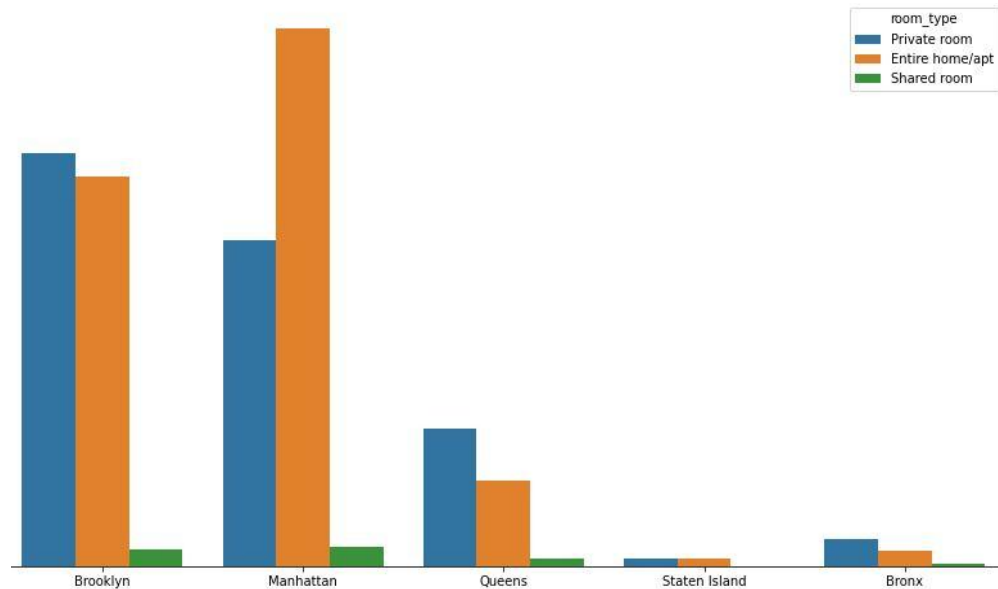
Also there were some null values, so we decided to replace those values with 0.

```
id                0
host_id           0
neighbourhood_group  0
neighbourhood     0
latitude          0
longitude         0
room_type         0
price            0
minimum_nights    0
number_of_reviews  0
reviews_per_month 10052
calculated_host_listings_count  0
availability_365   0
dtype: int64
```

# Insights from EDA

## Count of hotels in each neighbourhood\_group

- Private room and entire home are the mostly in demand.
- Shared room is the least popular in every region.

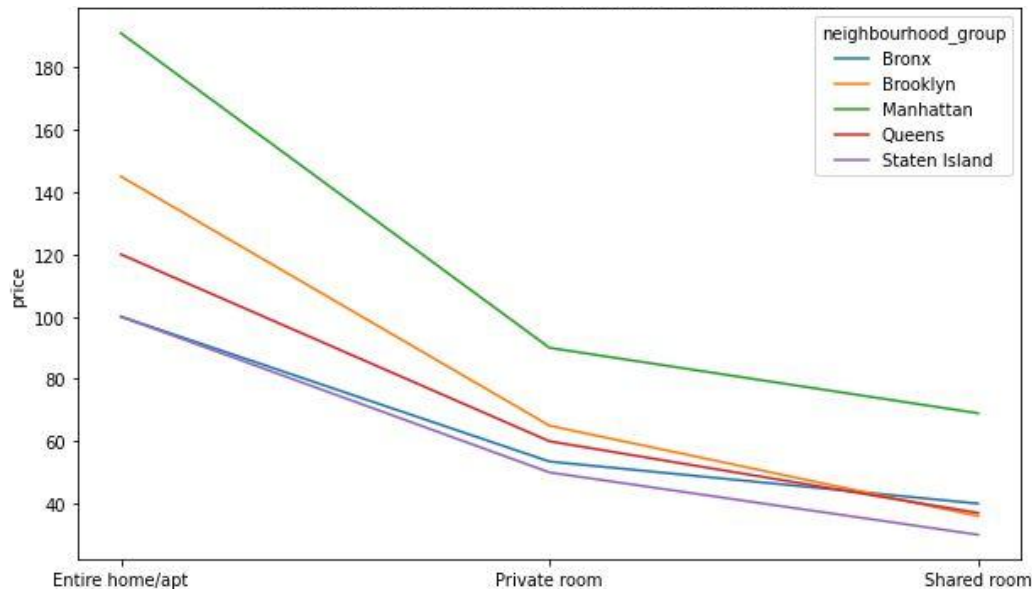




# Insights from EDA

## Median price of hotels in each neighbourhood\_group

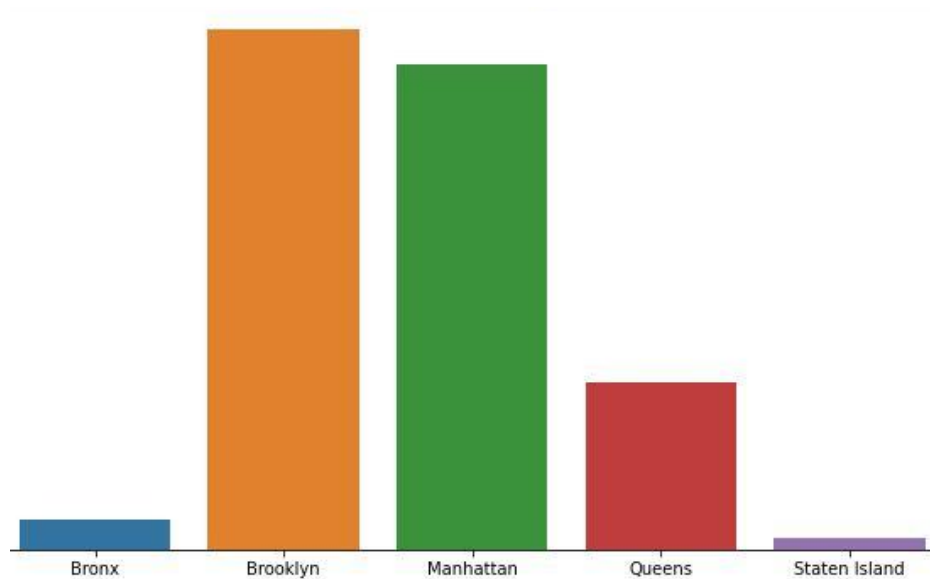
- Manhattan is the most expensive
- General trend: Manhattan > Brooklyn > Queens > Bronx > Staten Island



# Insights from EDA

Number of reviews each neighbourhood\_group receives

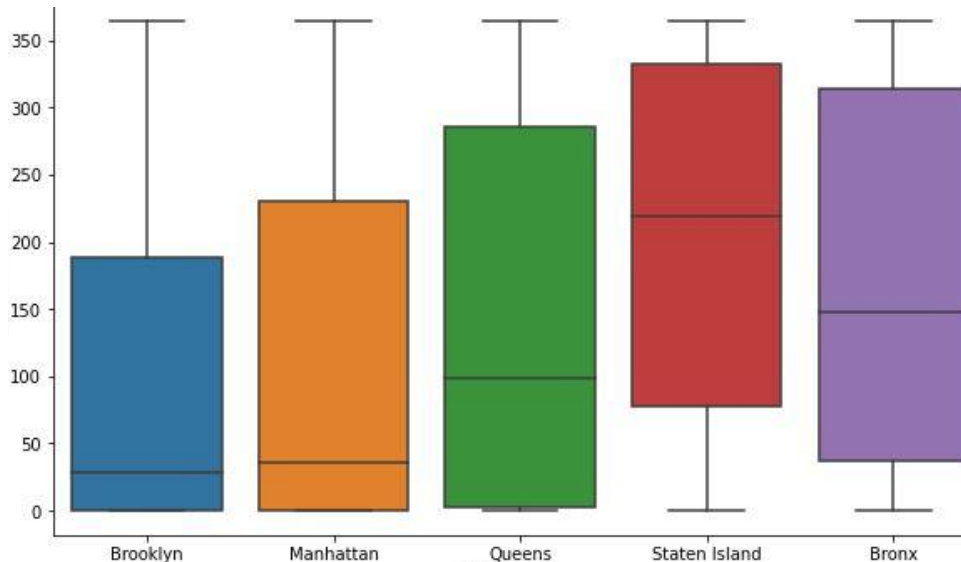
- Brooklyn and Manhattan receive most reviews
- General trend is roughly the same as previous slide



# Insights from EDA

## Availability of hotels

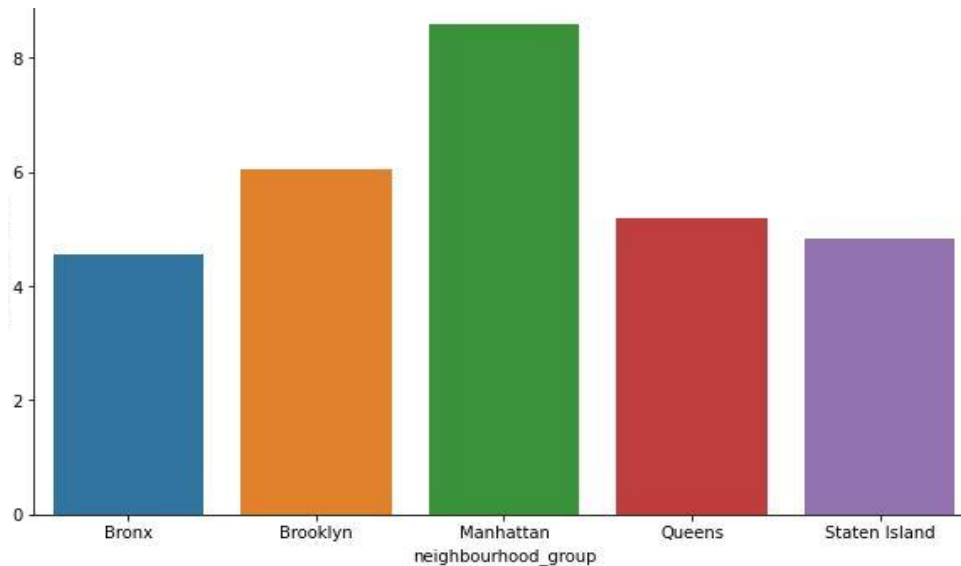
- **Brooklyn and Manhattan are the least available. Most of the hotels are booked year around**
- **General trend is roughly the opposite as price of hotels**



# Insights from EDA

## Minimum nights a person spends at a hotel

- People have to book hotels for more days in Manhattan and Brooklyn. They also stay much longer there.
- General trend is roughly same as price of hotels



# Insights from EDA

	host_id	neighbourhood_group	calculated_host_listings_count
<b>34757</b>	219517861	Manhattan	106929
<b>29495</b>	107434423	Manhattan	53360
<b>19636</b>	30283594	Manhattan	14641
<b>12846</b>	12243051	Manhattan	9216
<b>14482</b>	16098958	Manhattan	9216
<b>25743</b>	61391963	Manhattan	8281
<b>31174</b>	137358866	Queens	8137
<b>17149</b>	22541573	Manhattan	7482
<b>33975</b>	200380610	Manhattan	4225
<b>3055</b>	1475015	Manhattan	2704

Hosts with most number of listings  
are from Manhattan

	host_id	room_type	number_of_reviews
<b>24497</b>	47621202	Private room	629
<b>7710</b>	4734398	Private room	607
<b>22225</b>	37312959	Private room	543
<b>4593</b>	2369681	Private room	540
<b>19453</b>	26432133	Private room	510
<b>13714</b>	12949460	Entire home/apt	488
<b>1976</b>	792159	Private room	480
<b>5059</b>	2680820	Private room	474
<b>161</b>	42273	Entire home/apt	467
<b>18297</b>	23591164	Private room	466

Busiest hosts are have private room  
or entire home/apt

# Conclusion

- **Most hotels:** Manhattan > Brooklyn > Queens > Bronx > Staten Island
- **Most expensive neighbourhood\_group:** Manhattan > Brooklyn > Queens > Bronx > Staten Island
- **Entire home/apt and private rooms** are the most popular hotel types
- **Reason behind the trend of regions being expensive** can be: The total reviews received by each group follows the same trend, people's preference to spend more days in each group follows the same trend and also because Manhattan and Brooklyn are famous tourist spots and financial centers in NYC which attract a large number of tourists and rich personalities

# Conclusion

- As a result, Manhattan and Brooklyn hotels are the least available for the whole year as they are mostly booked. All neighbourhood\_groups follow the opposite trend here.
- Again as a result, people have to stay longer in Manhattan and Brooklyn hotels as compared to other regions. It follows the same trend as hotel prices
- The Busiest hosts offer either private rooms or entire home/apt for rent
- Converting shared room type Airbnbs in Manhattan and Brooklyn to entire home type or private room type will improve revenue generation

**Thank you**