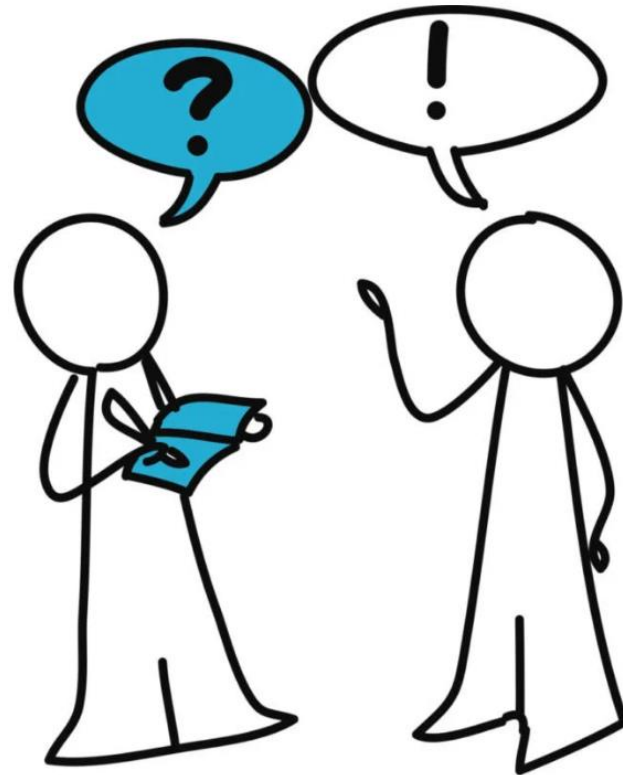# Capstone Project-1
Airbnb Booking Analysis

By – Viral Bhatu Shewale

Content :

- ❑ **Problem statement**
- ❑ **Data Summary**
- ❑ **Preparing dataset**
- ❑ **Exploring dataset**
- ❑ **Data visualization**
- ❑ **Conclusion**

What is Airbnb?

➢ Airbnb was founded in 2008 as a platform enabling locals to list their homes for short-term rental and travellers to have a lodging option alternative to hotels.

➢ Airbnb is the largest provider of rooms available to travellers. Along with homes for rent, Airbnb.com allows its users to search through "Experiences", which include pre-arranged multi-destination trips and "Restaurants,".

Problem Statement

➢ Since 2008, guests and hosts have used Airbnb to expand travelling possibilities.

➢ Today, Airbnb has become a one-of-a-kind service that is used and recognized by the whole world and generates a lot of data.

➢ Data that can be analyzed and used for security, business decisions, understanding of customers' and providers' (hosts) behaviour and performance on the platform, guiding marketing initiatives, implementation of innovative additional services and much more.

Data Summary

Shape : Columns = **16** , Rows = **48896**

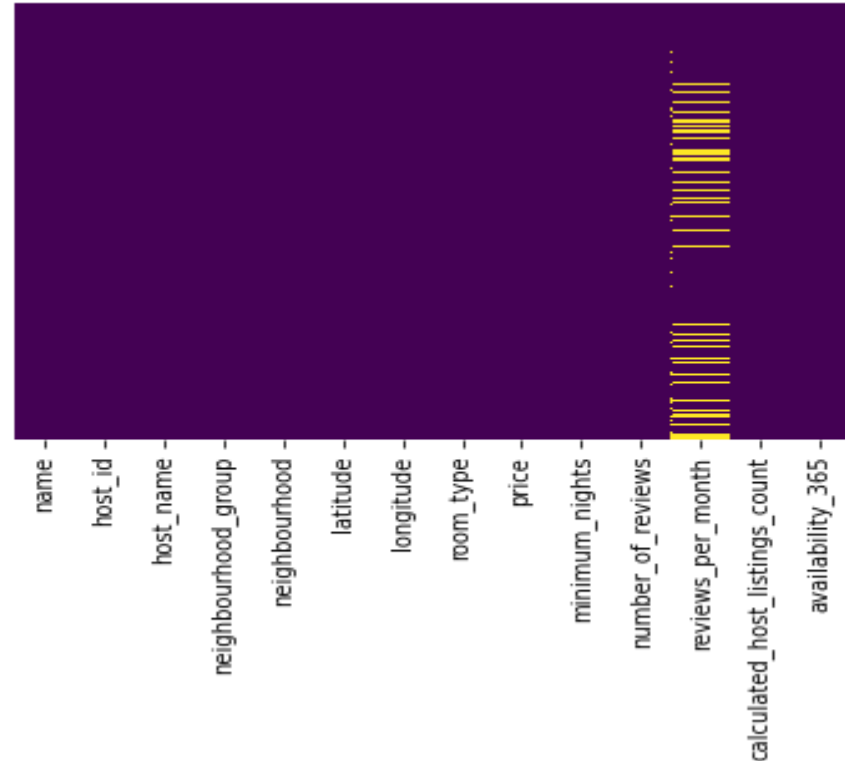|					
1. Id	int64		11. minimum_nights	int64
2. name	object		12. maximum_nights	int64
3. host_id	int64		13. last_review	object
4. host_name	object		14. reviews_per_month	float64
5. neighbourhood_group	object		15. calculated_host_listings_count	int64
6. neighbourhood	object		16. availability_365	int64
7. latitude	float64			
8. longitude	float64			
9. room_type	object			
10. price	int64			

Exploratory data analysis

Point of discussion :

- ❑ **How many different types of rooms are listed on the Airbnb platform?**
- ❑ **Understanding Different Areas in which Airbnb services exist.**
- ❑ **From which neighbourhood group, most rooms are listed on the Airbnb platform.**
- ❑ **Which host is the busiest and why?**
- ❑ **The most expensive location to book on Airbnb.**
- ❑ **Most preferred room type.**
- ❑ **Highest traffic generating area.**
- ❑ **Price distribution according to different areas.**
- ❑ **Price-wise availability of rooms.**
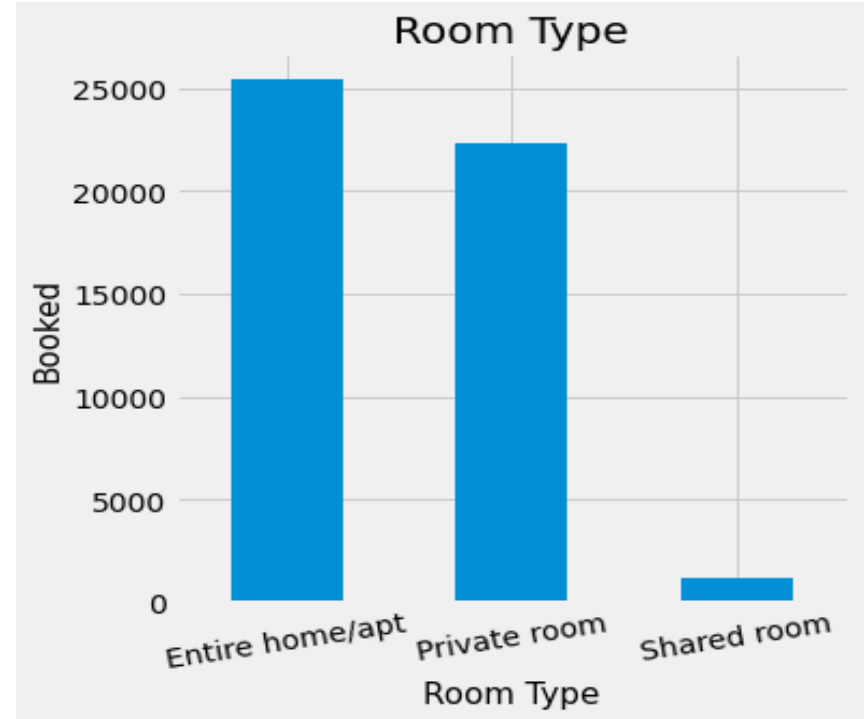- ❑ **Days-wise room availability.**

Preparing dataset:

➤ Libraries used: **Numpy, Pandas, Seaborn and Matplotlib**

➤ In the data, there were **three columns** which have **null values -** ['name', 'host_name', 'reviews_per_month']

➤ Most of the null values were present in the columns **-** 'reviews_per_month'.

➤ Handling null values:-

  * Replacing null values of column name and host_name by <u>No Name</u>.

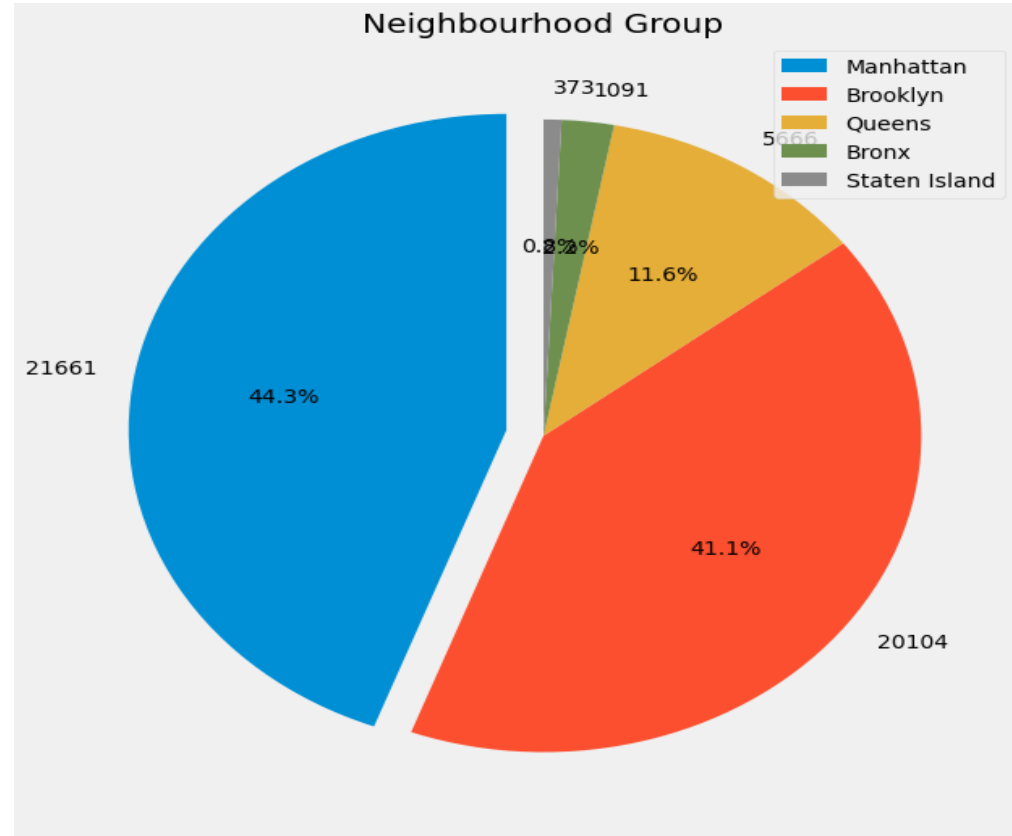 * reviews per month by its <u>mean</u>.

which is the most preferred room type by the customers for their stay?

➢ **51.97%** of customers preferred an Entire home/apartment for their stay.

➢ **45.67%** of customers preferred a Private room for their stay.

➢ A few customers prefer a shared room.

Room Type

which neighborhood group generates the highest traffic?

➢ We can see that **Manhattan** and **Brooklyn** have the most listings, covering more than **80%** of the total listings**.**

➢ **Manhattan** and **Brooklyn** are the *dominating* areas.

➢ **Staten Island** has the least number of listings (**0.76%**).

Neighbourhood Group

3731091
5666
21661
44.3%
0.8% 2%
11.6%
41.1%
20104

Manhattan
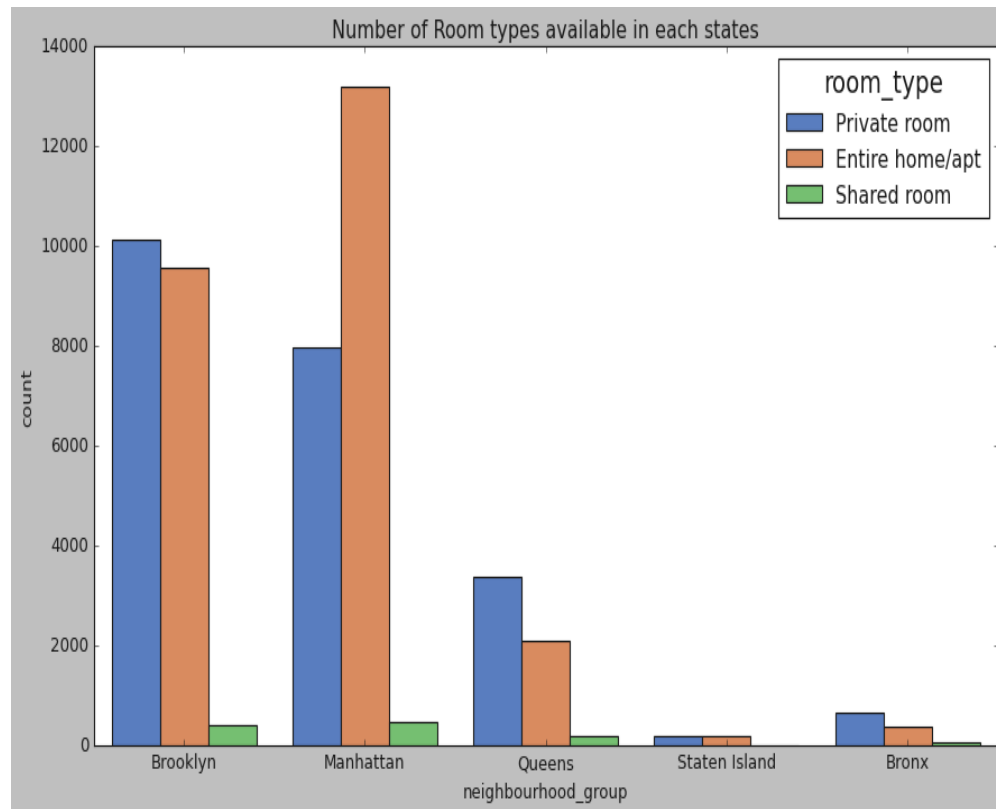Brooklyn
Queens
Bronx
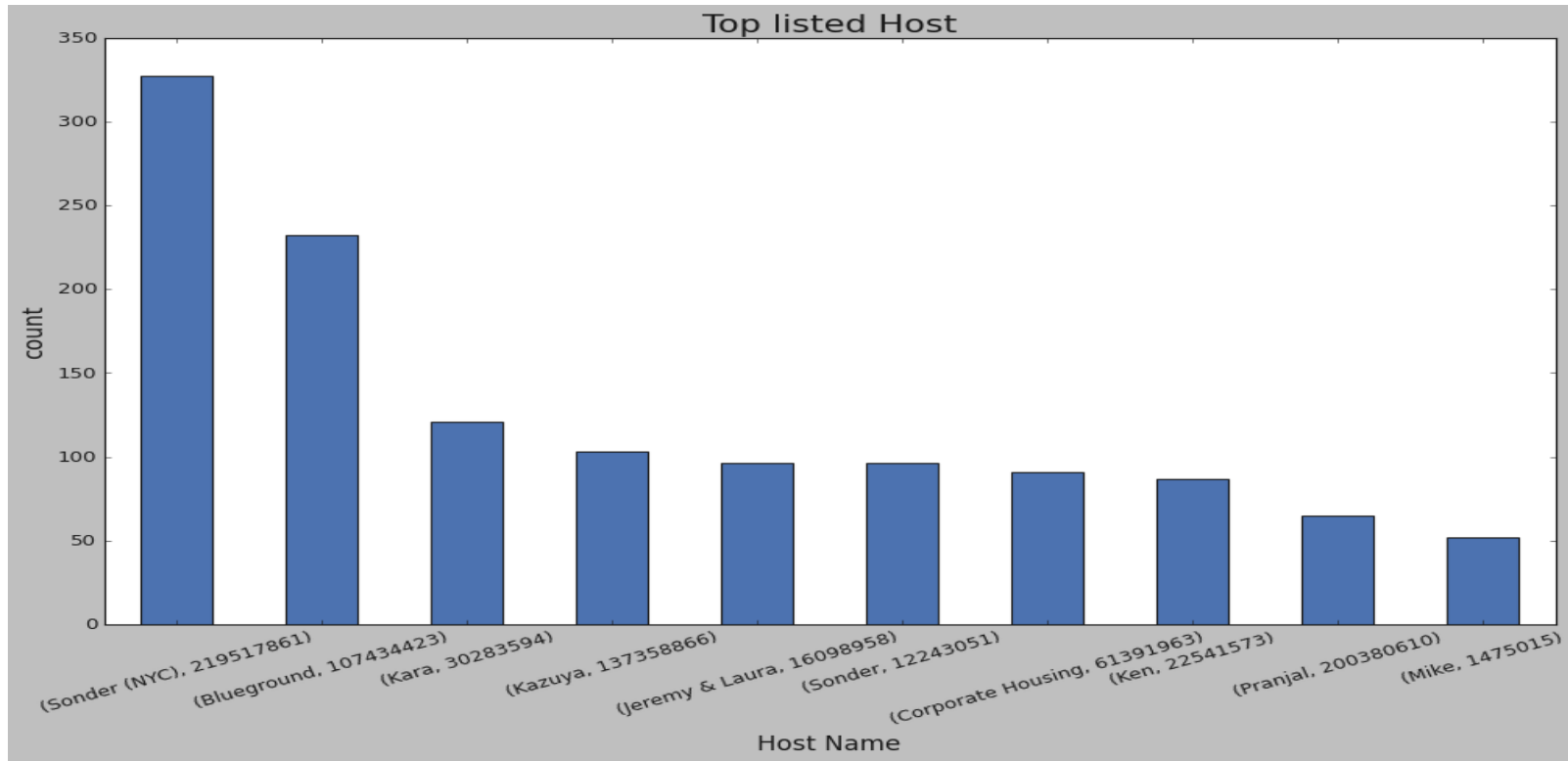Staten Island

Rooms are listed according to a different locations.

➤ **Manhattan** has the most number of **Entire home/apt** and **Brooklyn** has the most number of **private rooms.**

room_type neighbourhood_group	Entire home/apt	Private room	Shared room
Bronx	379	652	60
Brooklyn	9559	10132	413
Manhattan	13199	7982	480
Queens	2096	3372	198
Staten Island	176	188	9

Number of Room types available in each states

which host is the most listed on the Airbnb platform?

➢ We got the top 10 listed hosts.
➢ Host Sonder (NYC) (id **219517861)** has +**300** listings on the Airbnb platform.

Which host is busiest and why?

Factors that we are going to assume to check for the busiest host.

1. The number of properties listed on the Airbnb platform.
2. The number of days the property was available.
3. Based on the average monthly review of that property.

Host Name	Listing Rank	Avr_Availability	Avr_monthly review
Sonder (NYC)	1	301	1.72
Blueground	2	253	1.25
Kara	3	274	0.92

➢ Host **Sonder (NYC)** have a total of **327** listed rooms out of that **319** are **Entire home/apt** (which is the most preferred type of room by people) and only 9 private rooms.
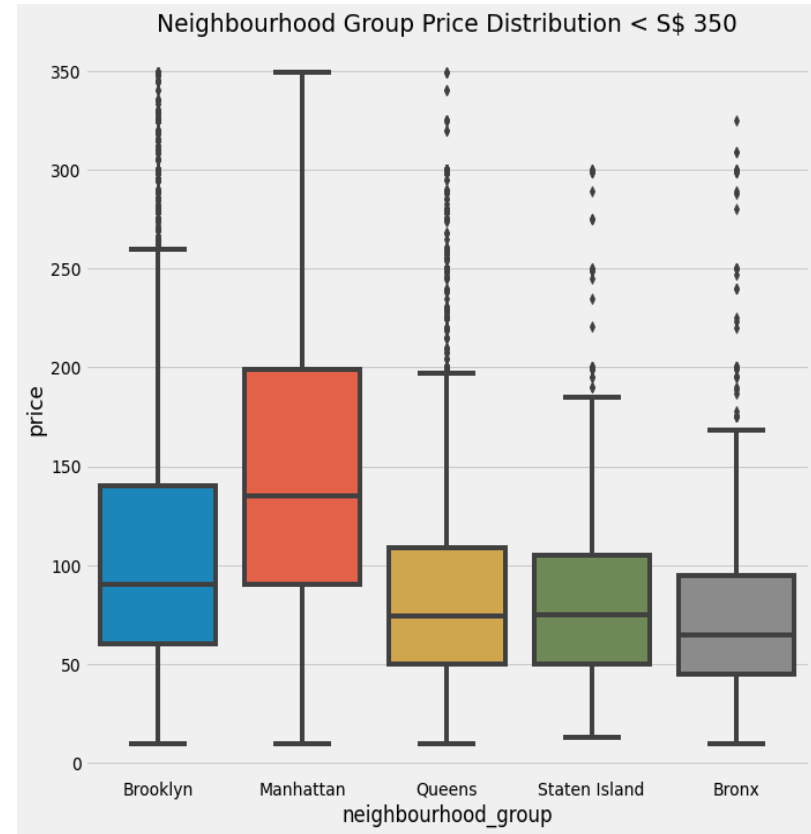
neighbourhood_group	Bronx	Brooklyn	Manhattan	Queens	Staten Island
host_name					
other host	1091.0	20104.0	21334.0	5666.0	373.0
Sonder	NaN	NaN	327.0	NaN	NaN

➢ All the rooms are listed in **Manhattan** (which is the most traffic-generating neighbourhood group).

➢ Availability of property listed by the host Sonder was +**300** days.

Hence, we can say that **Sonder is the busiest host**.
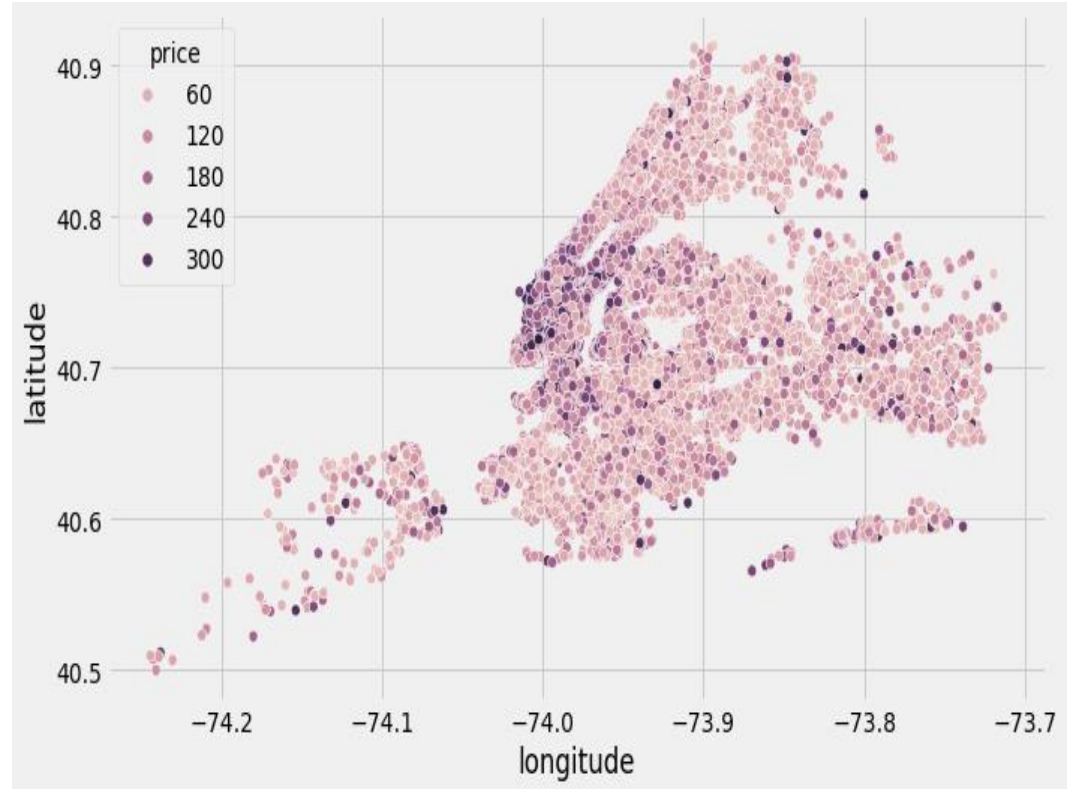
Price distribution :

➢ Price of accommodation in **Manhattan** is pretty higher and affordable for high class people**.** But those who can not afford to stay in a prime location can visit the place which is close to the prime location like **Brooklyn** and **Queens**.

➢ **Bronx** is the <u>cheapest location</u> in **New York**.

Neighbourhood Group Price Distribution < S$ 350
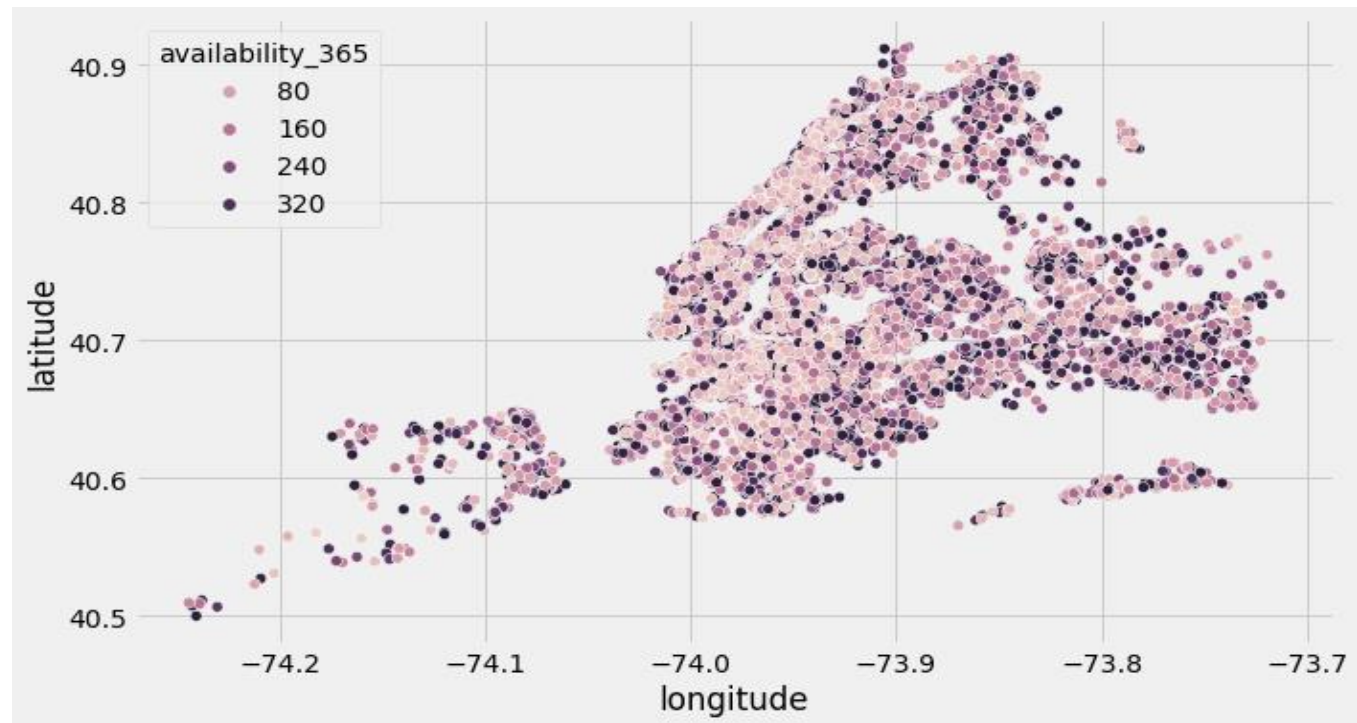
Availability of rooms at the various price range:

➢ We can see most of the rooms available at a price range from **60** to **180**. Less number of rooms available at a price range <u>greater than</u> **500**.

	Price	Total room available
0	price<60	8594
1	60<price<120	17829
2	120<price<180	10510
3	180<price<240	5325
4	240<price<300	2719
5	300<price<500	2683
6	price>500	1044

Days wise room availability:

	availability 365	available rooms
0	>=80	9955
1	>=160	5255
2	>=240	4501
3	>=320	5039
4	<320	6541

Conclusion:

➢ Simply by performing EDA on the dataset, we have identified various new insights into how the Airbnb listings are distributed, how the price range varies according to location, and we know where are the listings located.

➢ There are three types of rooms onto which the majority of people preferred the <u>Private room</u> and the <u>Entire home/apartment</u>.

➢ we found out that <u>Manhattan</u> and <u>Brooklyn</u> are dominating (Prime locations) locations.

➢ The price of accommodation in <u>Manhattan</u> is pretty higher and affordable for high−class people.

➢ <u>Sonder (NYC)</u> (Host_ID = 219517861) is <u>the busiest host</u> on the Airbnb platform.

➢ <u>Manhattan</u> and <u>Brooklyn</u> had the <u>most listings</u>, covering more than **80%** of the total listings.

➢ Most of the traffic is generated in <u>Manhattan</u> and <u>Brooklyn</u>.