

# Capstone Project - 1



## Airbnb Bookings Analysis



### Team members

Amani chivilkar

Abdul Rahaman

Shivam

Vinay Kumar

# Points Of Discussion

- Introduction
- Know your Data
- Data summary
- Data Preprocessing
- Exploratory Data Analysis



# Introduction:-

- ❖ Airbnb.inc. is an American company that operates an online marketplace for lodging, primarily homestays for vacation rentals, and tourism activities. Based in San Francisco, California, the platform is accessible via website and mobile app.
- ❖ Airbnb stands for “**Air Bed and Breakfast**” Airbnb does not own any of the listed properties; instead, it profits by receiving commission from each booking. The company was founded in 2008 by Brian Chesky, Nathan Blecharczyk and Joe Gebbia. Airbnb is a shortened version of its original name, AirBedandBreakfast.com



## Columns of the DataFrame

- **Id**- Whenever a host list its property on Airbnb, a unique id is created, since every id is unique there are 48895 unique id
- **name**- It is the property description. Same name does not mean same property
- **host\_id**- Its the identity id given to individual host (For example- two different host cant have the same host\_id). So if we want to count the number of host, we count the unique host\_id.
- **neighbourhood\_group**- This are the 5 Boroughs of New York city, in which the properties are located.
- **neighbourhood**- This are the towns and villages within the 5 Boroughs
- **room\_type**- Types of properties
- **price**- Property price
- **minimum\_nights**- It is the minimum number of night you can book a property,( example-1 year rental contract)
- **number\_of\_reviews**- Total number of reviews
- **Calculated\_host\_listing\_count**- The number of times the host has listed its property.
- **availability\_365**- the number of days the property were available for booking.

# Data summary

- For this EDA project, we are using the "**Airbnb booking Analysis**" database.
- Data set contains **48895** rows and **16** columns
- This data set contains the booking details of five boroughs namely **Bronx, Brooklyn, Manhattan, Queens and Staten Island** and includes details such as host id, host name, neighbourhood, room types, price of rooms, duration of stay, number of reviews, availability and host listing counts.

**The following steps are performed on the Data set:**

**Step 1.** Overview the whole data : Look at the number of rows and columns of data set and Look at each column information that is data type, null rows

**Step 2.** Looking at missing data

**Step 3.** To establish relationship between various features of the Dataset

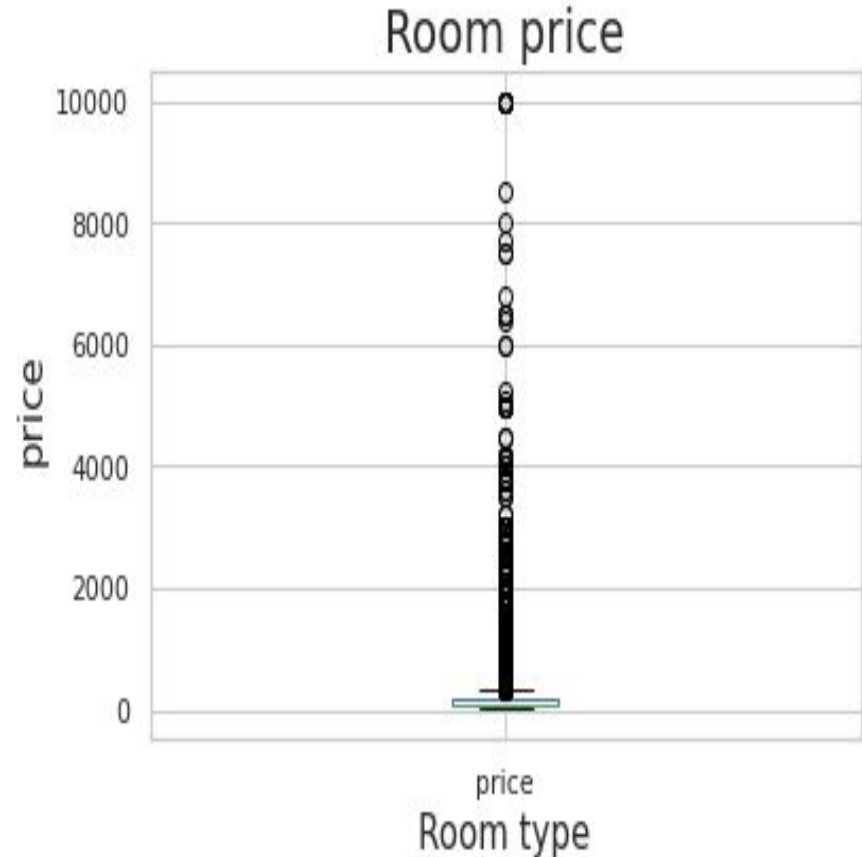
**Step 4:** Present these relationships using various Data Visualization Techniques

**Step 5:** Conclusion

## Clean the DataFrame

- Drop the column which are not required, like i'd
- Fill the NaN values in host\_name by 'no name'
- Replace NaN values in 'name' column by 'no description'
- Replace NaN values in 'reviews\_per\_month' column by 0
- If price is 0, replace it by the average price of all similar room\_type with similar neighbourhood\_group
- Convert time string to datetime
- Handling outliers

- **Handling Outliers for Price column** :By plotting boxplot,we see some outliers. We have used mean for removing outliers.Seaborn boxplot are used for visualizing the outliers in the present data.
- Plotting different features against one another:
- After doing the data cleaning we are left with **45923** rows and **16** columns.



# → Exploratory Data Analysis

## Researched questions :

- Define busiest host
- About top 5 host with highest listing counts
- Number of properties listed
- Price of properties
- Overall correlation with price
- Count the number of unique host
- Find the rate of reviews per month
- Properties location and percentage on map
- Average price of Properties
- Price at which most of the properties are available
- Top 5 host with minimum possible earnings
- What are the minimum\_nights stay set by most of the host
- Most Popular Descriptive words used by the host





## → Define busiest host ?

- I would consider the host to be busiest, if the host frequently receives bookings. Since we have not been provided with any booking data. We can consider the number of reviews received by the host as the minimum number of bookings the host could get.
- Host with maximum number of reviews is not the busiest, but the host with the highest rate of reviews (i.e max number of reviews per month) is the busiest.

Reviews count per month for different property type of Top 5 host

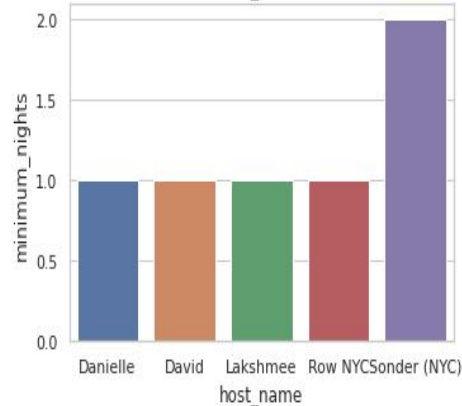
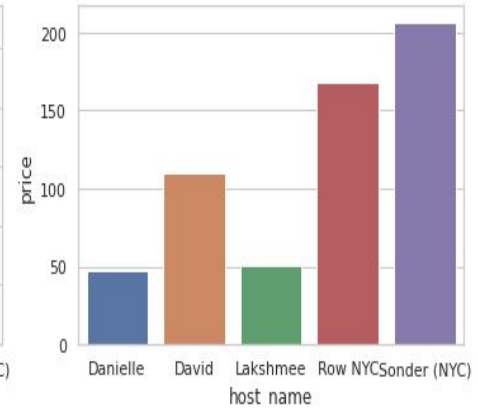
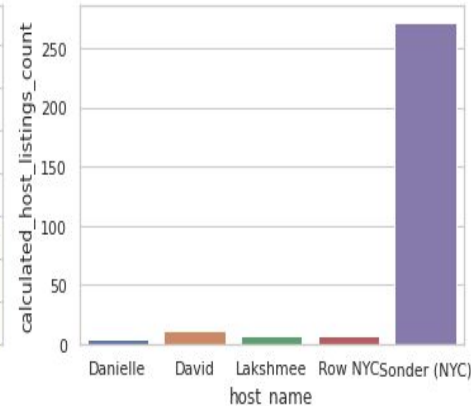
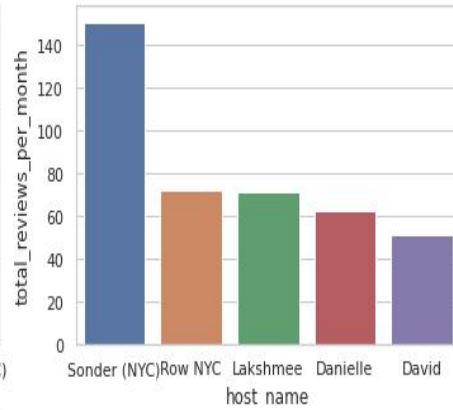
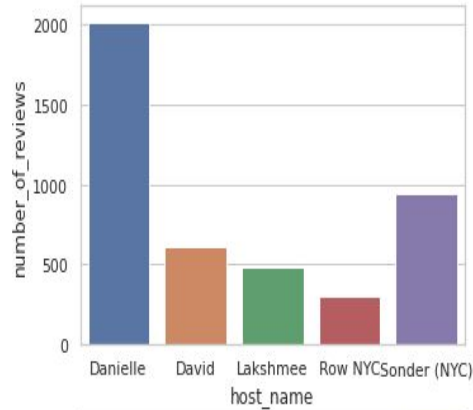
host_name	reviews_per_month
Sonder (NYC)	397.56
Row NYC	111.72
Lakshmee	80.63
Danielle	68.02
David	62.89

## → Count of unique host that has listed properties on Airbnb ?

- Since 2 different host can have the same host\_name, so we use host\_id.
- host id is unique for individual host
- This shows the count of listings of particular host id on Airbnb. But we want total number of unique host ids
- Airbnb has **35389** unique host ids or host owner

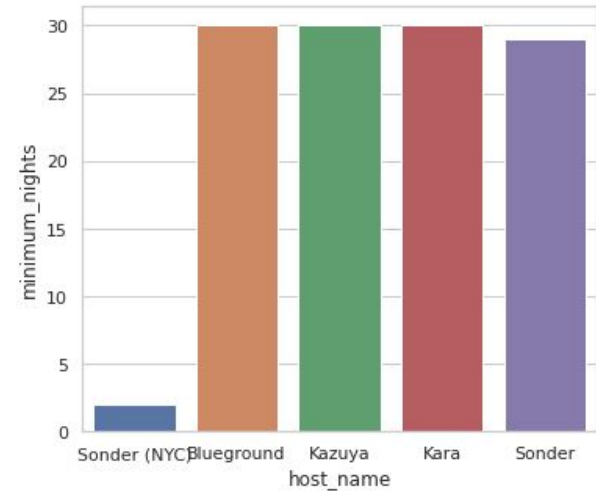
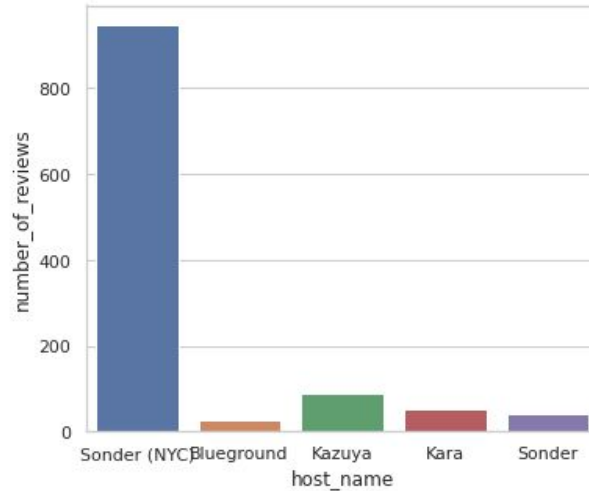
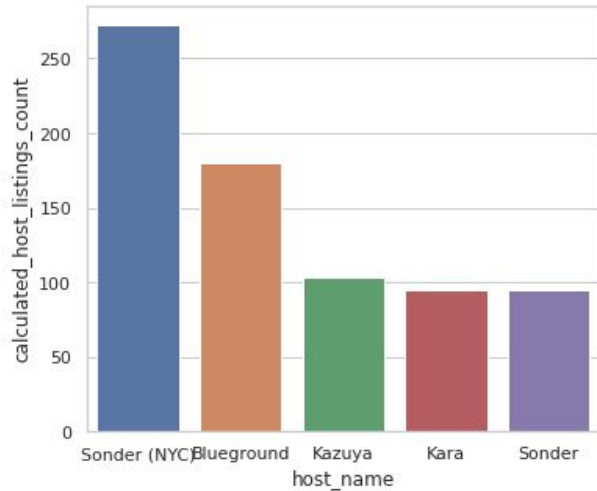
	host_id	listing_counts
0	219517861	272
1	107434423	180
2	137358866	103
3	12243051	95
4	30283594	95
...	...	...
35384	95485067	1
35385	167560332	1
35386	263742622	1
35387	205706382	1

# → All about the top 5 host with max number of reviews per month on Airbnb



- Sonder has the highest review per month, highest listing count, and most of his properties are in Manhattan which is the most densely populated among the 5 boroughs of New York city. Hence Sonder is the busiest host.
- Most of his properties are entire home/apt, with an average price of **251\$** and the minimum\_night stay is **2** nights.

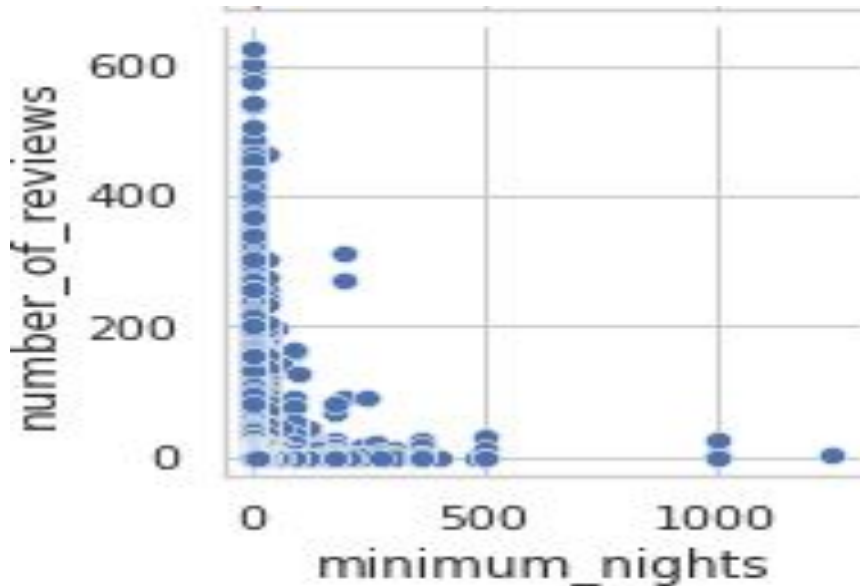
## → Find top 5 host with highest listing counts



- Sonder has the most number of listing with (**272**) followed by Blueground (**170**)
- In spite of Blueground having the second highest listing count, he has got very less reviews among all, that is because the minimum night required to book Blueground's property is 30 nights. where as for Sonder is 2 nights.
- Scatterplot will shows the relation between minimum\_nights and total\_reviews of all the host

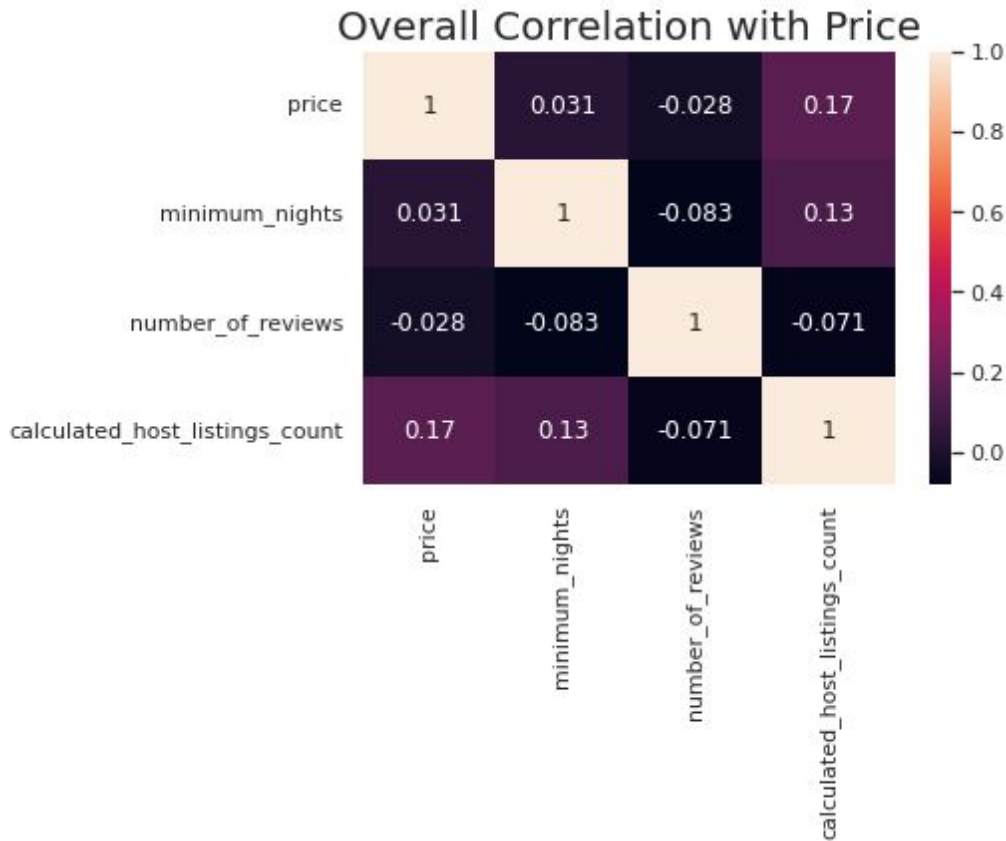
## → Relation between minimum\_nights and total\_reviews

- The higher the minimum nights , the lesser the reviews



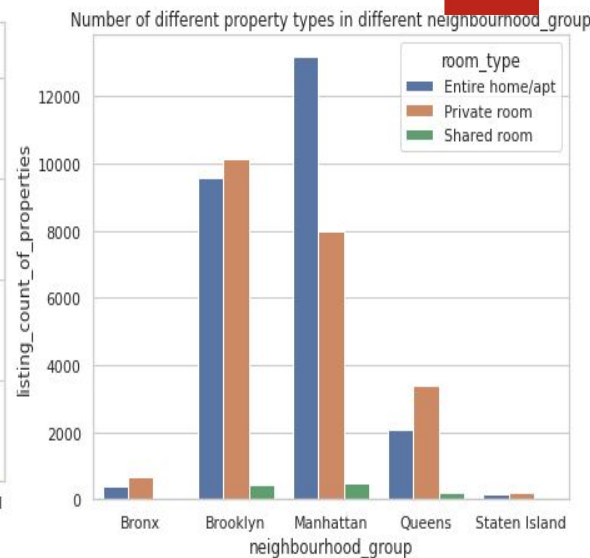
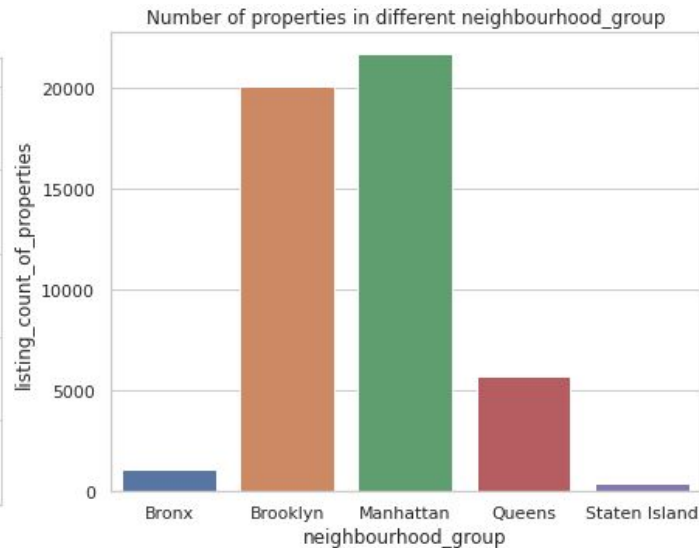
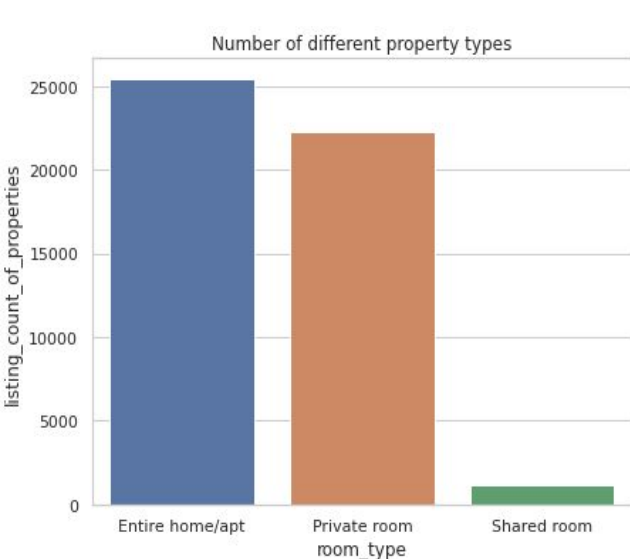
- minimum\_nights and total\_reviews are negatively correlated

## → Overall Correlation with Price



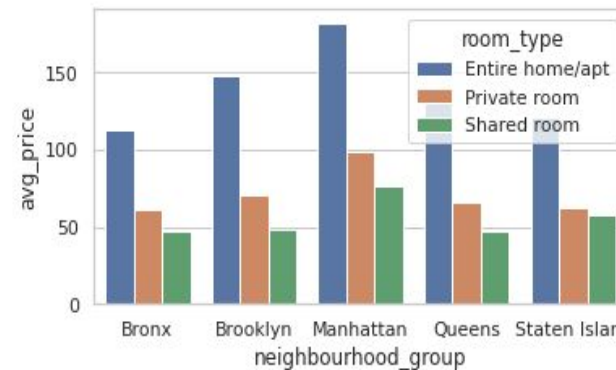
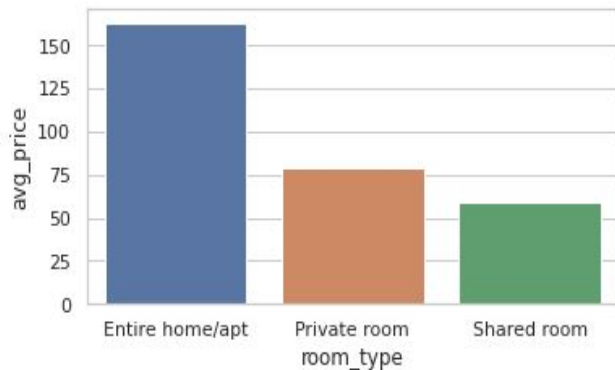
- Most of the properties with cheaper price ,often gets more number of reviews as compare to the properties with higher price. That is maybe because people prefer renting cheaper property, which leads to more reviews.
- Less minimum nights, higher is the number of review . This is because if a host gets rental for his properties, for 60 days, then he will not get booking for atleast 60 days, he gets atmost 1 review within 2-3 months. And the host who has set minimum\_nights as 1 or 2 will get atleast 1 booking within 2-3 months, hence he will get atleast 1 review with in 2-3 months.
- number\_of\_reviews is negatively correlated with host listing count, this doesn't make sense

## → The number of properties listed on Airbnb



- **49.6%** is Entire home/apt
- **47.97%** private room
- **2.5%** is shared room
- Manhattan is the smallest neighborhood\_group, but has about **42.5%** of the properties listed on Airbnb, that's because it is densely populated city of New York
- Out of all the properties situated in Staten Island, only **0.8%** of the properties are Shared room and Out of all the properties situated in Bronx has only **5.2%** of the properties are Shared room

## → The Average price of properties



Average price of

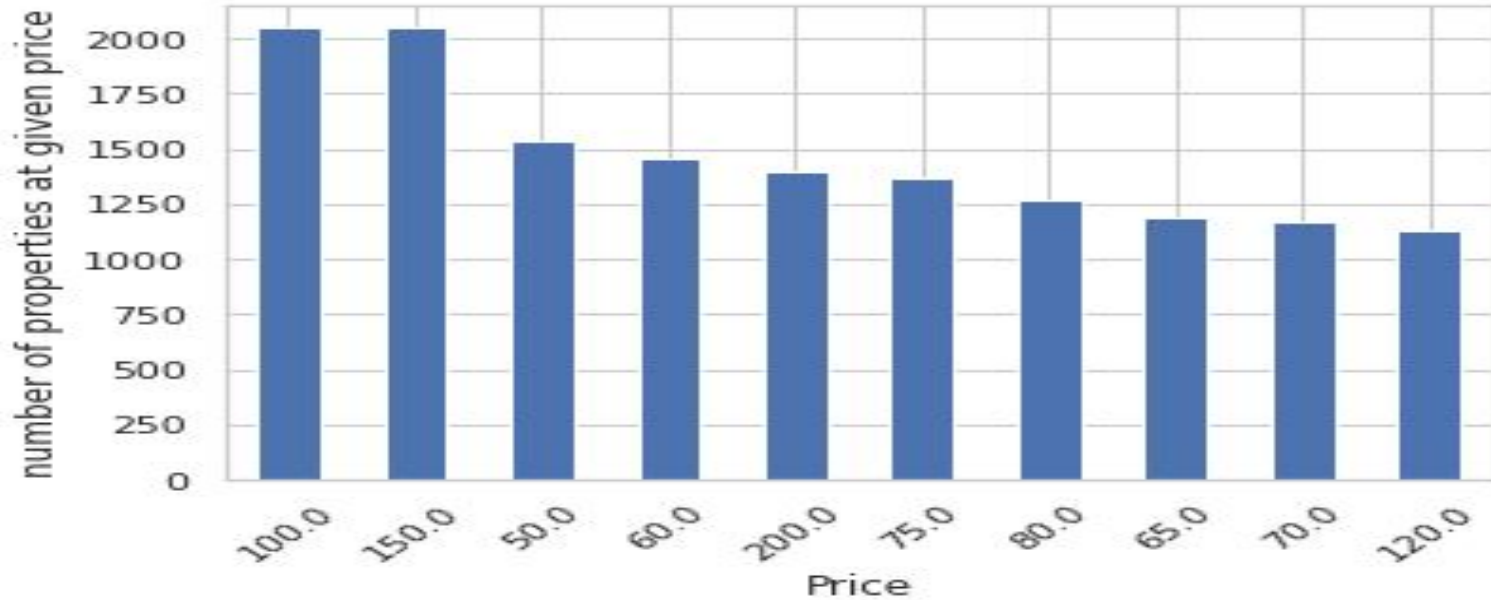
- Entire home/apt is 162\$
- Private room is 75\$
- Shared\_room is 60\$
- Since there is very less difference between the avg price of private room and shared room, people will prefer renting private room

Properties in Manhattan are more expensive(145\$), because its is densely populated neighborhood\_group of New York

- The avg price of every property type in Manhattan are higher than the other
- The average price for private room and shared room are almost same in Bronx, Brooklyn, Queens and Staten Island.

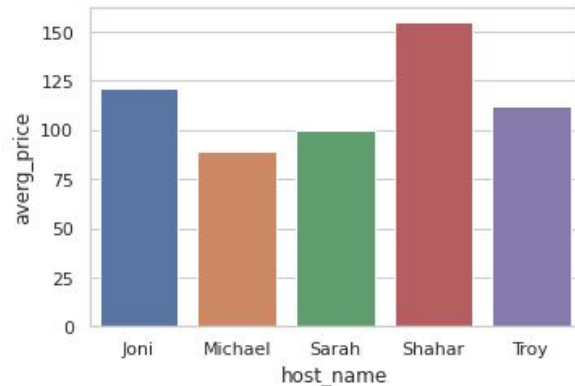
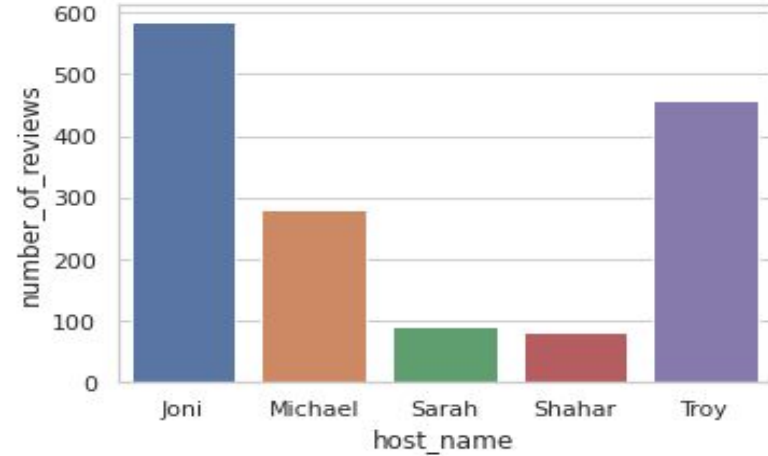
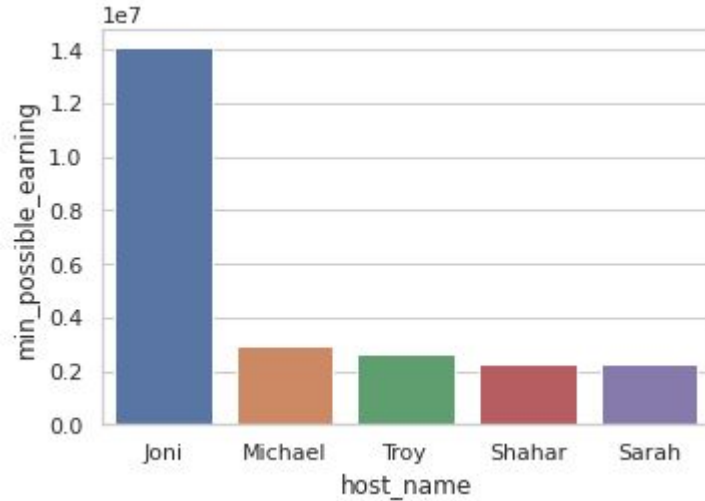


## → Price at which most of the properties are available



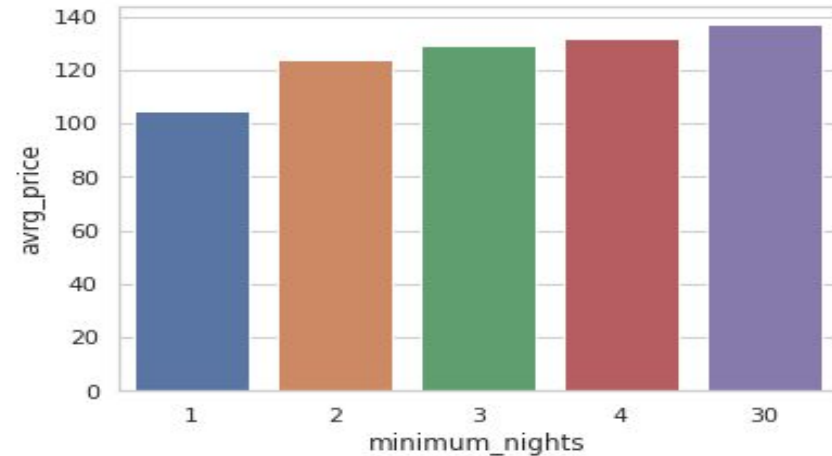
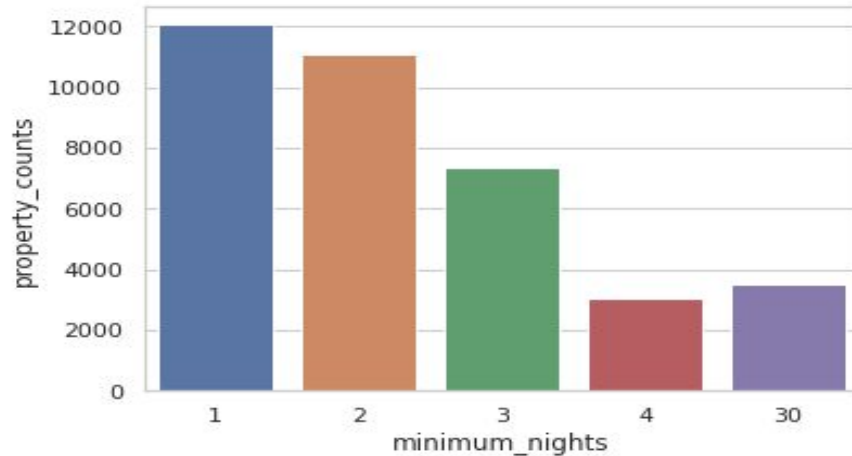
- Most of the properties listed on Airbnb costs between **100** to **200** dollar per day

## → Top 5 host with minimum possible earnings

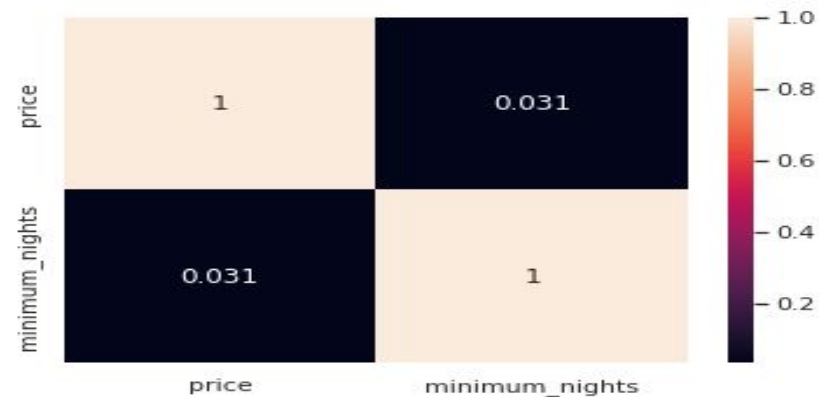


- Joni's earning is maximum (1.4 Billion \$) compared to other hosts
- His average room price is around 120\$ with minimum\_night stay of 200 nights
- He has listed only 2 times on Airbnb and has 580+ reviews
- He owns 2 Entire home/apt in Brooklyn

→ What are the minimum\_nights stay set by most of the host and the average price of property with particular minimum nights



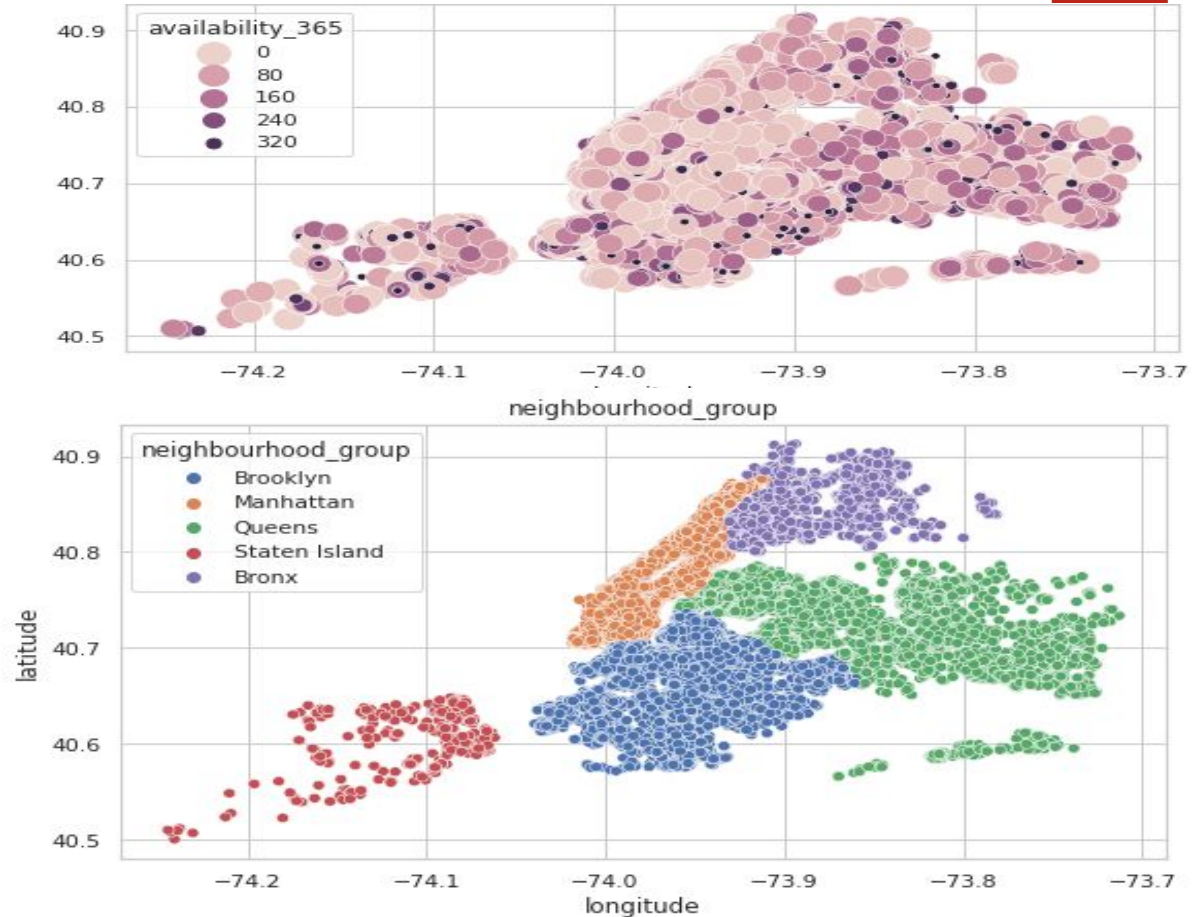
- Maximum number of properties listed on Airbnb has minimum nights stay of 1, 2 and 3 nights
- Lesser the minimum nights cheaper is the price this implies that they are positive correlated
- minimum nights and price have small positive correlation, our assumption is false



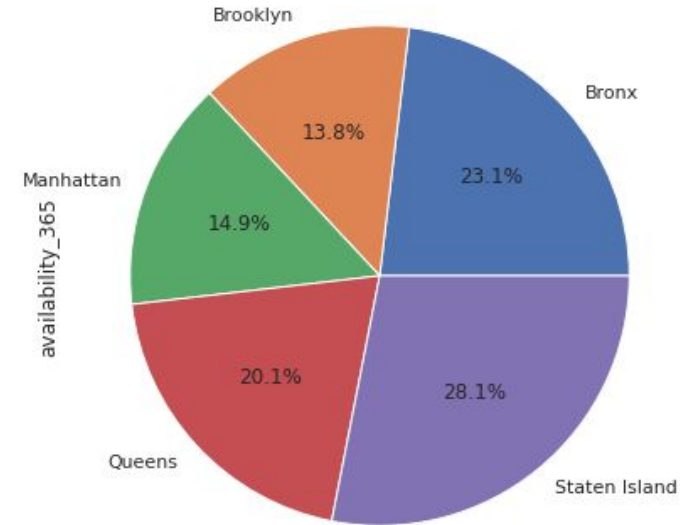
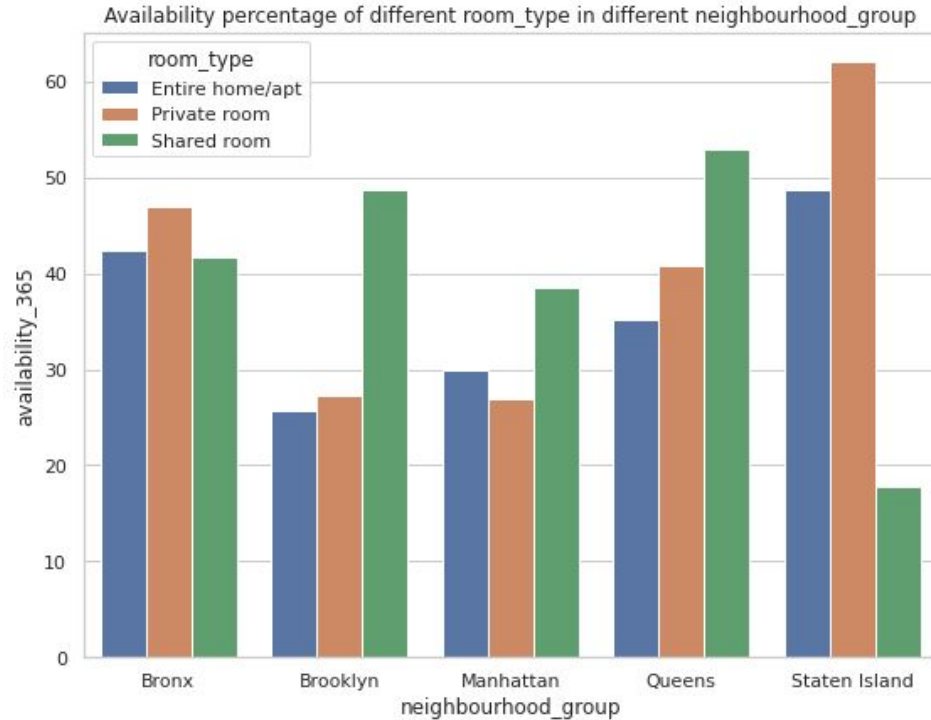
## → Availability of properties in all neighbourhood\_group

AI

- One of the factor on which the availability of property depends is booking days.
- Availability of properties in Manhattan and in the north part of Brooklyn is less as compared to other neighbourhood\_group
- Availability of properties for booking in southern region of Brooklyn, Bronx and Queen is more  
This shows that properties in this region gets booked for few day
- People prefer renting apartment in Manhattan and in the northern part of Brooklyn

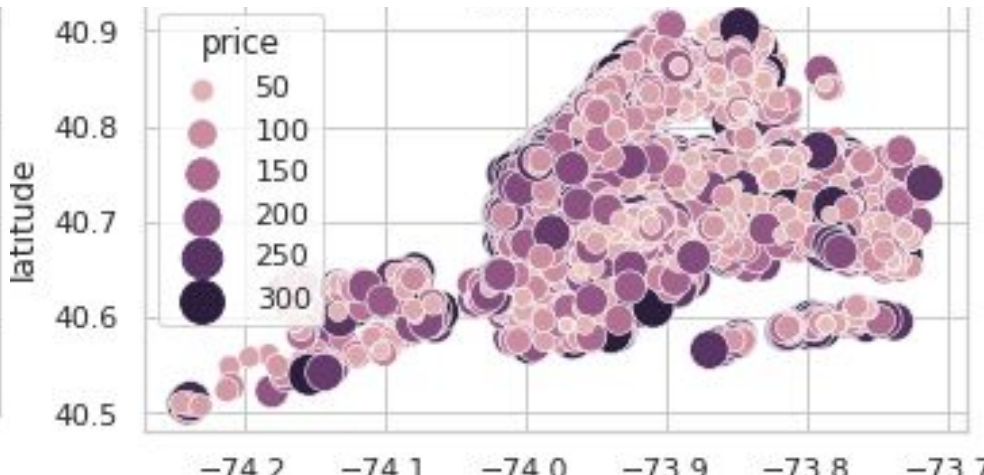
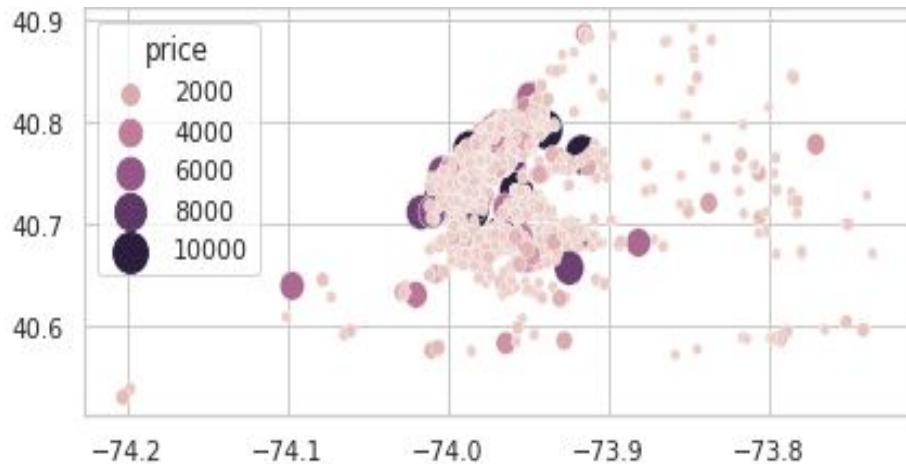
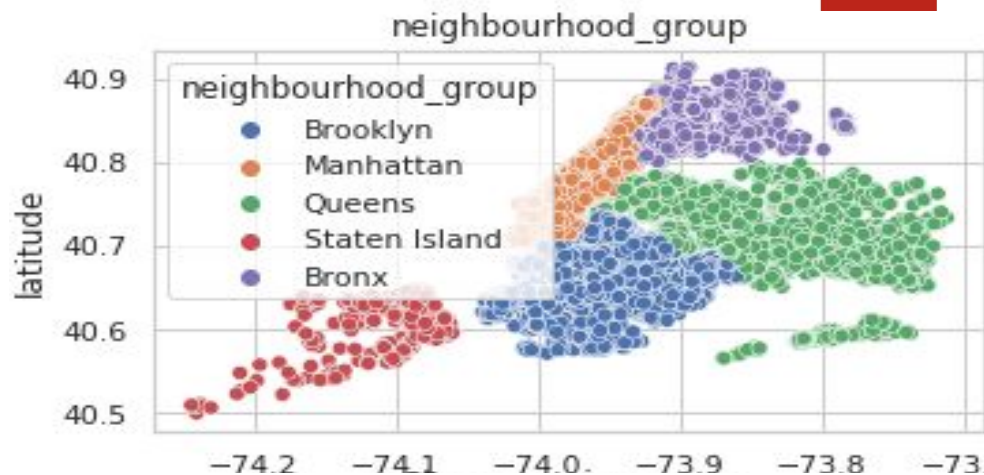


## → Availability percentage of properties in different neighbourhood\_group

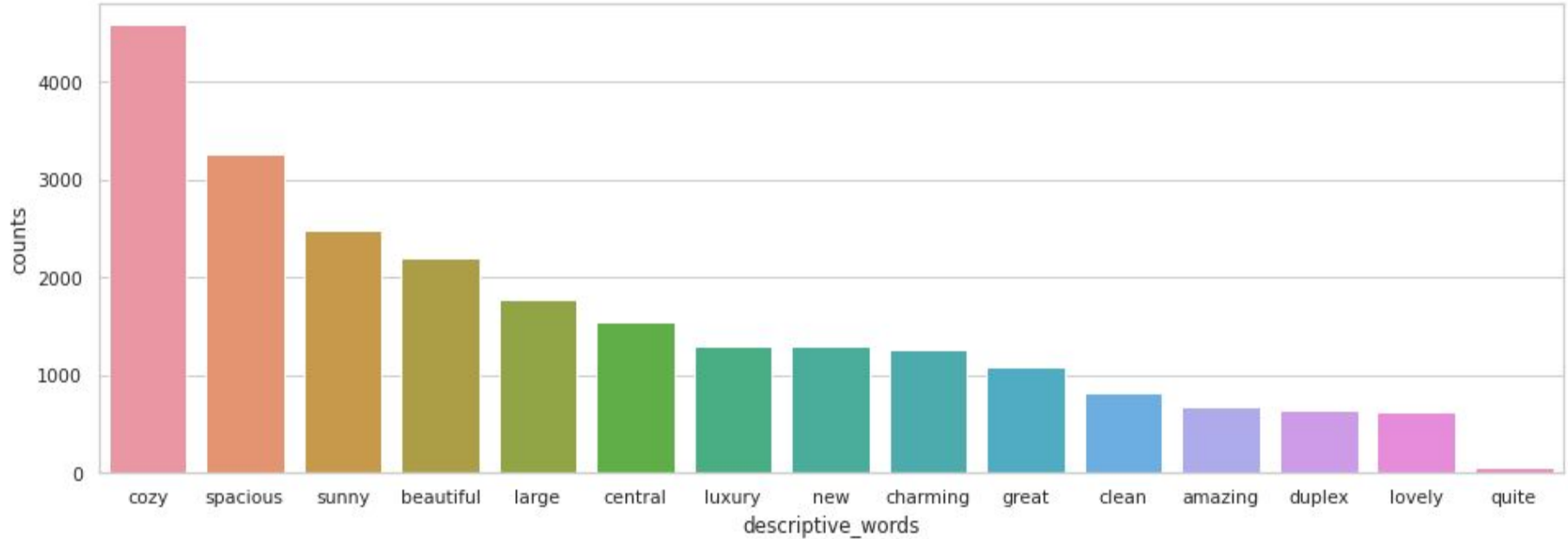


## → Scatterplot of longitude and latitude showing price range

- In every neighbourhood\_group there is always cheaper option
- Most of the expensive properties are situated in Manhattan and in northern region of Brooklyn
- Very few expensive properties are situated in Bronx and queen
- Staten Island has hardly 10-8 expensive properties



## ➔ Most Popular Descriptive words used by the host to describe their



- People use these words i.e **Spacious,cozy, beautiful ,large ,luxury** words more number of times in searching or reviewing any particular property type



# Conclusion:-

- Entire room/apt is the highest property type listed on airbnb
- Most of the properties listed on Airbnb are situated in Manhattan
- Brooklyn has got the highest number of private room.
- Manhattan has the highest number of entire room home or apartment and shared room
- Most of the property types have one or two minimum night stay





## → Conclusion cont..

- Apartment in Manhattan are more expensive whereas apartment in Bronx and Queen are cheaper
- The avg price of private room and shared room are almost same, so people will prefer renting private room over shared room
- Most of the properties listed on Airbnb cost between 100 to 200 dollars per day.
- Availability of property for booking is comparatively less in Manhattan
- Sonder is the busiest host among all, with 397 reviews per month, highest listing count of 272 listings on Airbnb
- minimum\_nights and total \_reviews are negatively correlated



Thank You