Capstone Project
Airbnb Booking Analysis
By
Pooja Hoolageri
Cohort -CANBERRA

Contents

- Introduction
- Problem statement
- Objective
- Data summary
- Data analysis and visualizations
  - ❑Neighbourhood group
  - ❑Price
  - ❑Room types
  - ❑Number of reviews
  - ❑Longitude and latitude
- Conclusion

Introduction

Airbnb hosts price their properties to maximize profits what approach do they use to price their properties expensively to get higher profit margins, or should they be more competitive to generate more sales? Many Airbnb hosts struggle with this pricing problem: they don't know if they're undercutting potential profits through low pricing.

The goal of this data exploration is to understand the significance related to the Airbnb dataset, we got to know different hosts having different kinds of properties for rent.

Prices of rooms depending upon location and facilities provided and with a minimum number of nights of stay required, also check various aspects related to price, room type, availability 365, reviews per month etc.

Problem statement

Airbnb Bookings Analysis:

- EDA - Exploratory data analysis (EDA) is used to analyze and investigate data sets and summarize their main characteristics, often employing data visualization methods.(Wikipedia)

- Since 2008, guests and hosts have used Airbnb to expand on travelling possibilities and present a more unique, personalized way of experiencing the world.

- 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 decision, guiding marketing initiatives, implementation of innovative additional services and much more.

Objective

- Finding the total number of listings available in New York.

- To find the average price for available room types/neighborhoods.

- To find the hosts (Who has the most number of listings and who has more number of reviews for their listings) through out New York.

- To find how price is affecting the Maximum reviewed places.

- To see the price variation for different listings throughout New York.

Data summary

➢ The Airbnb data set contains 48895 records with 16 variables.

Important columns :

➢ 'name' - These are names of the properties given by specific hosts. Or i can say, 'name' is a column which has the descriptions provided by the respective hosts for the reference of their customers.

➢ 'host_id' and 'host_name' - Many properties are being offered by many hosts. This 'host_id' and 'host_name' holds the records of all those hosts.

➢ 'neighbourhood-group' and 'neighbourhood' - These columns holds the information about the city and areas of the properties which are offered in here.

➢Longitude' and 'Latitude' => As the name suggests it just contains the longitude and latitude of the property location ⬜ 'Room_type' => It displays the room_type of the property (either private room / entire home / shared room)

➢ 'price' => Its an important column which holds the price value of all those properties.

➢ 'number_of_reviews' and 'reviews_per_month' => It clearly tells about the number of reviews and reviews per month for those properties and hosts hospitality.

Data analysis and visualizations

➢ Rentals/properties present in Neighborhood group , Neighborhood, Room type

➢ Overall contributions of each neighborhood in the count of listings throughout NYC

➢ Total count of room types available in NYC

➢ Average price for each neighborhood group

➢ Average Price for each neighborhood areas in respective neighborhood groups

➢ Price distribution data in every neighborhood-group

➢ Overall Price distribution through NY

➢ In search of a famous Host

➢ Number of reviews for each neighborhood-group

➢ Average price for room-type throughout NY

Treating NA
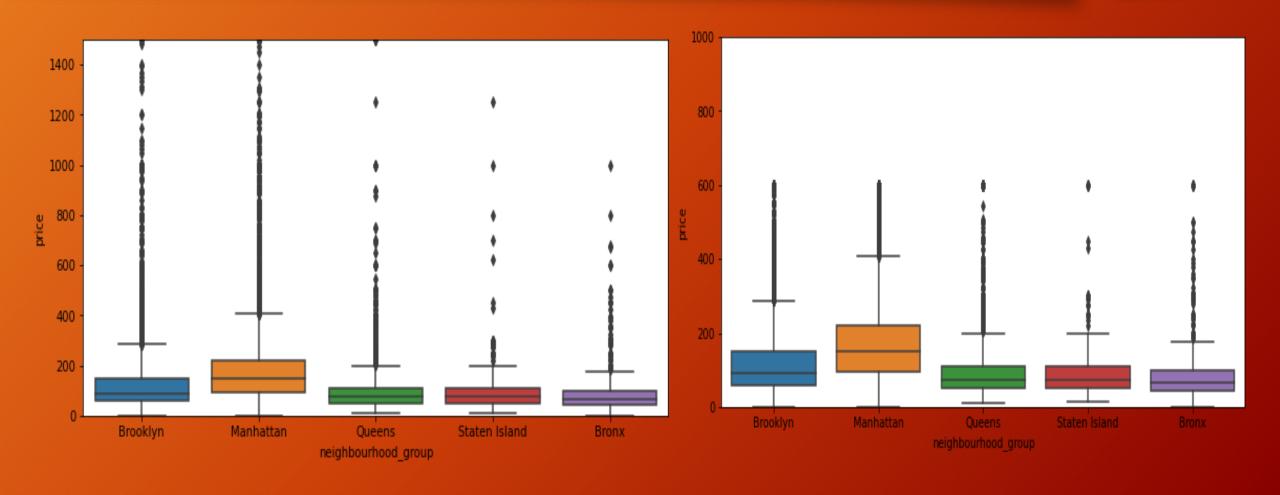
AI

```
id                               0
name                            11
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
```

If we have a look at the number of null values throughout the data set, then we can observe that we found some of NA values in few columns.

```
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
reviews_per_month         0
availability_365          0
dtype: int64
```

- Last_review column contains 10052 records with null values. This column is not much of a help to analysis. So we drop that column.
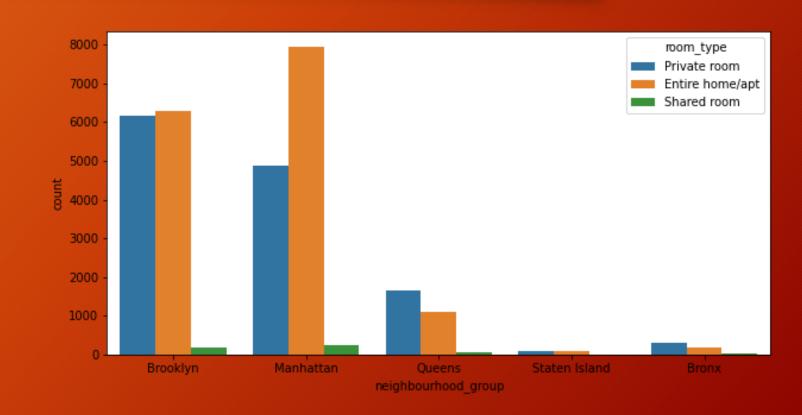- We then equated the null values of 'reviews_per_month' column as 0.
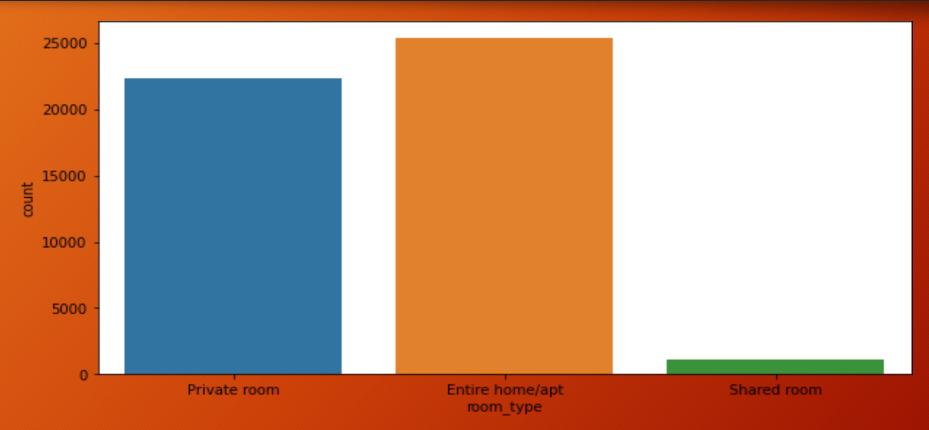
Outliers

Neighbourhood groups

Number of properties available in different neighborhood groups is more in Manhattan(Home room type)
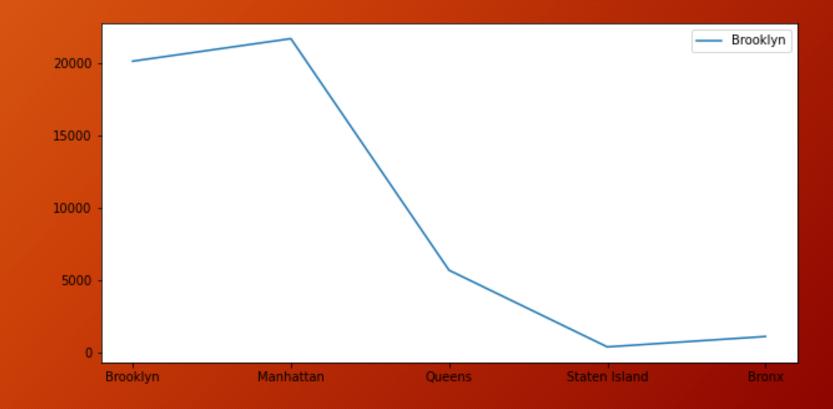
Room type throughout New York

Number of room types available in New York which are of the type Private room, Entire home/Apartments, Shared room.
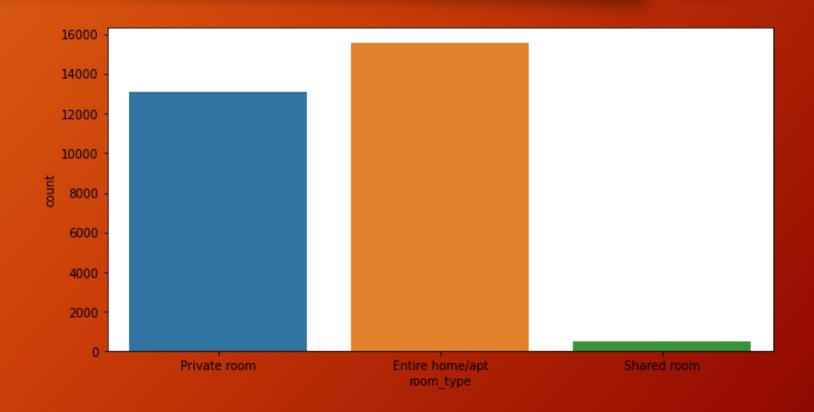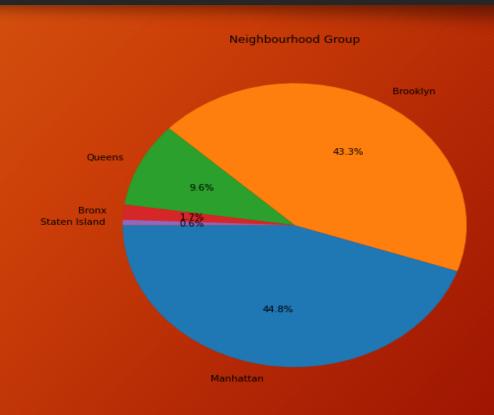
Total number of rentals/properties in each neighborhood group

This gives information about the count of properties present in those 5 cities. We can observe that Brooklyn and Manhattan has more properties when compared to all 5 cities. In particular, Manhattan is the city which is busy city with most properties to offer.

Total count of room types available in NYC

Number of room types available in New York which are of the type Private room, Entire home/Apartments, Shared room. **Overall just trying to fetch more information about the room types in general, we have more properties in home/apt. We already know that home/apt is most demanded room type throughout NY.**
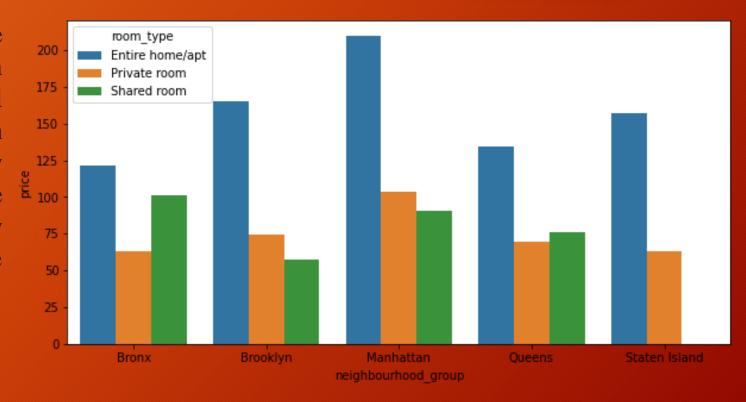
Overall contributions of each neighborhood in the count of listings throughout NYC

The more number of listings are observed in Manhattan as it takes the 44.3% of the overall listings
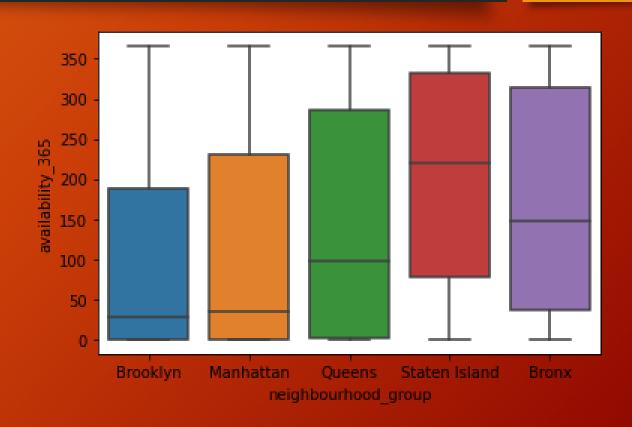
Average price for each neighborhood group

We got to know that most properties are available in Manhattan. Now we even want to know the average price of all the properties for the cities. We can observe that, Manhattan is the city which is in lead in all aspects. Its price for all room type properties are pretty high when compared to the rest of the cities. Indeed its busy and famous city.
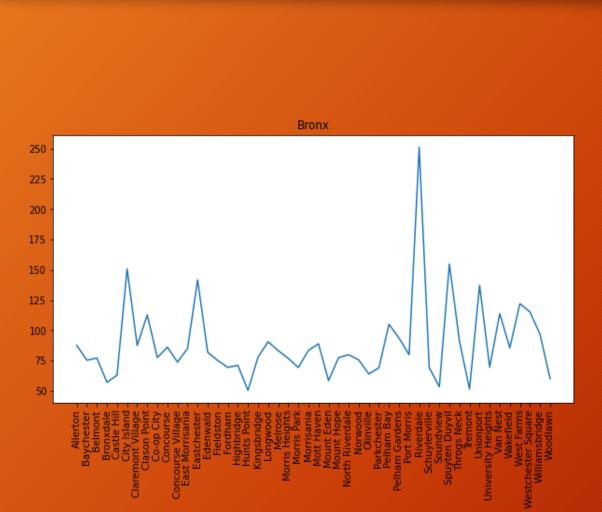
Availability of the listings and neighborhood group

We can observe that the mean of listings count in Brooklyn, Manhattan and Queens are available only for 0 to 100 days a year.
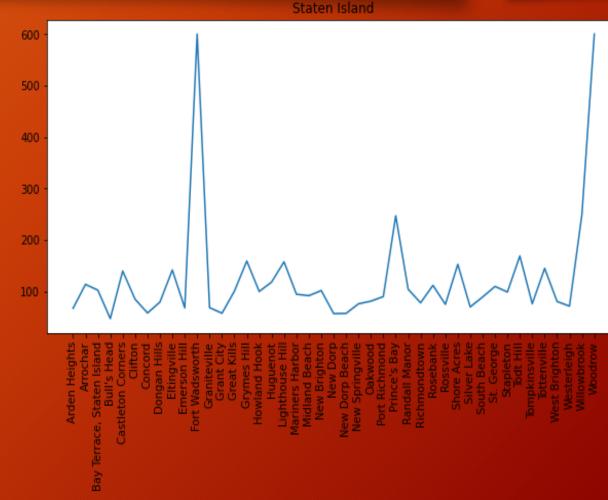
Average Price for each neighborhood areas in respective neighborhood groups
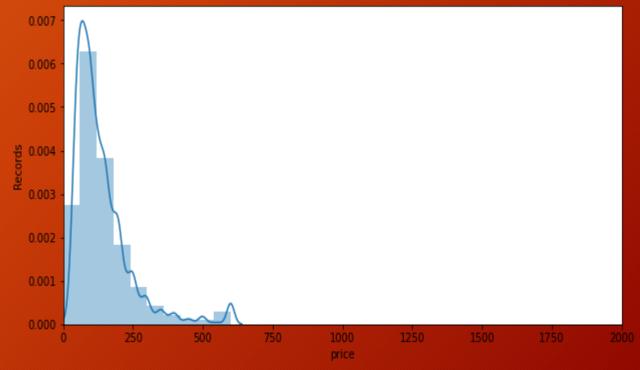
We can observe that many cities are having very fluctuating price in their neighborhood. Especially Manhattan is the city which has even very less price and even very high price in its neighborhood.
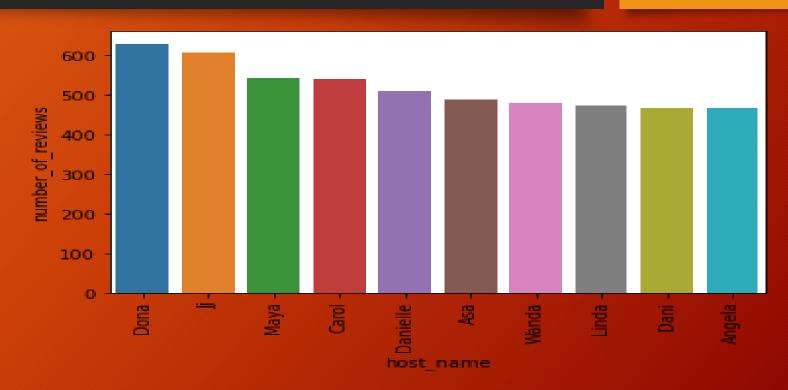
Frequency distribution of prices throughout NewYork

Till now we discussed the price range only for neighborhood / neighbourhood_group / room type. Now lets focus on how well the price is varying throughout NY. We can observe that most of the price are in between the range 0 to 250.(We can say those properties are not that high until you want to opt for a fancy property with high price.

Hosts who got more number of reviews

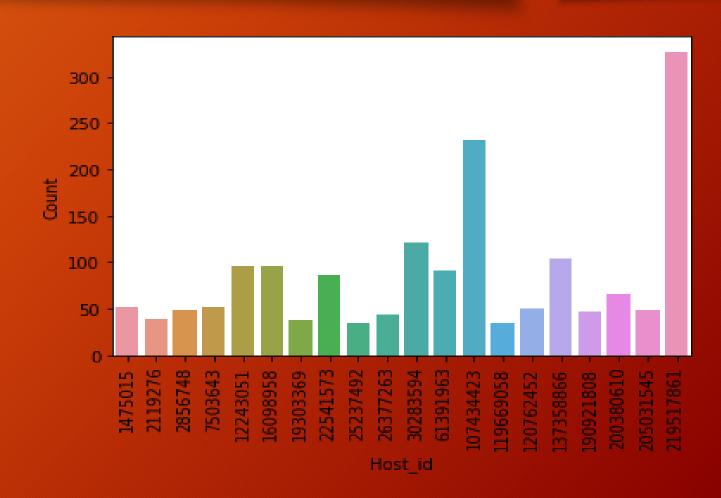Dona is the host who got more number of reviews throughout New York.

	id	name	host_id	host_name	neighbourhood_group	neighbourhood	latitude	longitude	room_type	price	minimum_nights	number_of_reviews	reviews_per_month	availability_365
11759	9145202	Room near JFK Queen Bed	47621202	Dona	Queens	Jamaica	40.66730	-73.76831	Private room	47	1	629	14.58	333
2031	903972	Great Bedroom in Manhattan	4734398	Jj	Manhattan	Harlem	40.82065	-73.94025	Private room	49	1	607	7.75	293
2030	903947	Beautiful Bedroom in Manhattan	4734398	Jj	Manhattan	Harlem	40.82124	-73.93838	Private room	49	1	597	7.72	342

Famous host / host with more properties

Here I tried searching the famous host, i felt that whoever has the most number of properties(doesn't matter if its a private room/home/shared room) is the famous host. He has been offering those properties for the customers. And if the property count is more that means he has more to offer. So obviously he is the one famous host.

*Above is the visualization which describes top 10 hosts whoever has the most number of properties to offer.

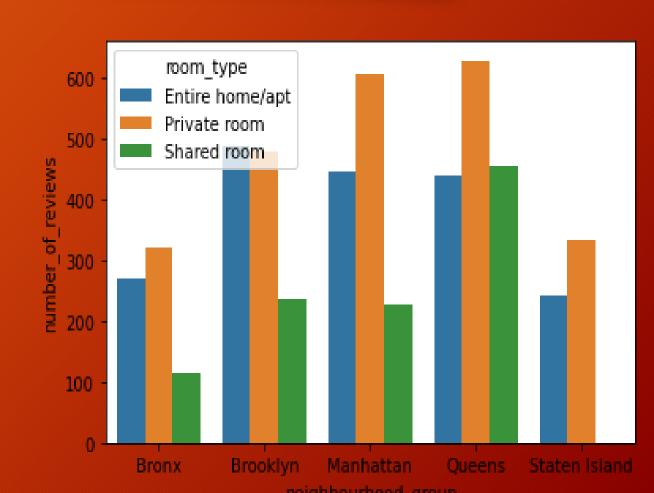We can see that the top1 host who is famous(according to me) is Sonder. And he is from Manhattan.

Number of reviews for each neighbourhood_group

We can observe that Queens has the most reviewed place in NY. Through this visualization we will get good comparison of the other neighborhood as well.
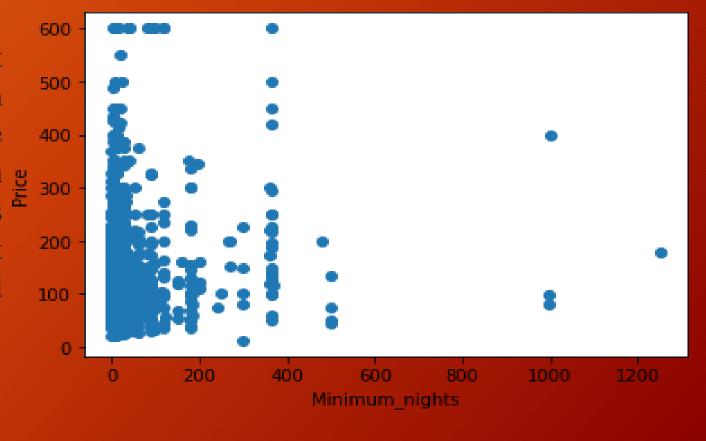
Overall,

(1)So from the above result, we have the average price per night as 65.4.

(2)And we can observe that host id 47621202 has most number of reviews for his property which is in Queens. We can even consider him as the famous host for some reason as he has most reviewed place in NY. Again its all up to our conception.

(3)the most reviewed place has very less price. That means we can conclude that the price which ever is very low, people tend to go to that home/private room. Hence it got most number of reviews.

(4)From observed data (above 4 result), we have the total number of properties in Queens is 5666 and out of which Now we can see that 629 reviews were on single property. Its pretty good. So we can come to the conclusion that Queens is the place/city which got most reviewed place. It may be because its comparably cheap when it compared to other cities like Manhattan/Brooklyn.

(5) The top reviewed place was available for almost 333days out of 365 days of a year. Which gives us good picture that people tend to choose the property with less price and which is mostly available
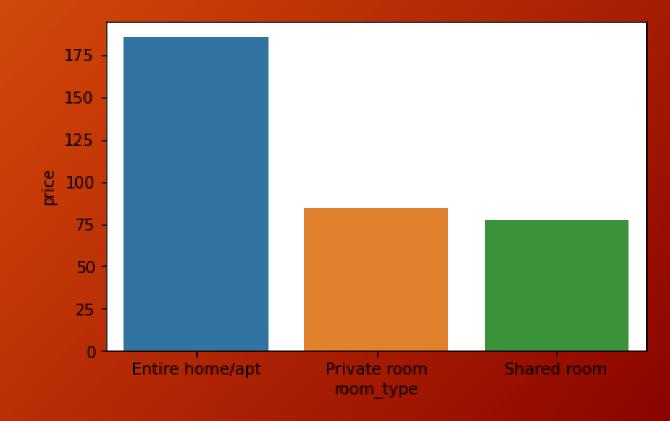
Average price in descending order based on minimum nights of stay

We can clearly observe that we can't get much interpterion from minimum number of nights and price visualization. Even there are places with less price which has the minimum nights as 1000 and vice versa. Because it mainly depends on the neighborhood group/city where you are staying.
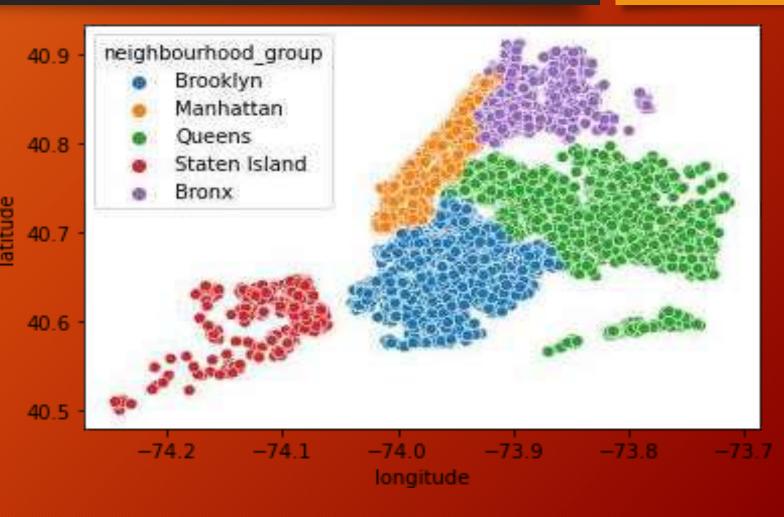
Average price for room type throughout NY

Now we are focusing on the column room type alone. We trying to find the average price for all room types throughout NY. We can see that Entire home/Apt is in demand and it has the price high when compared to private room/shared room.
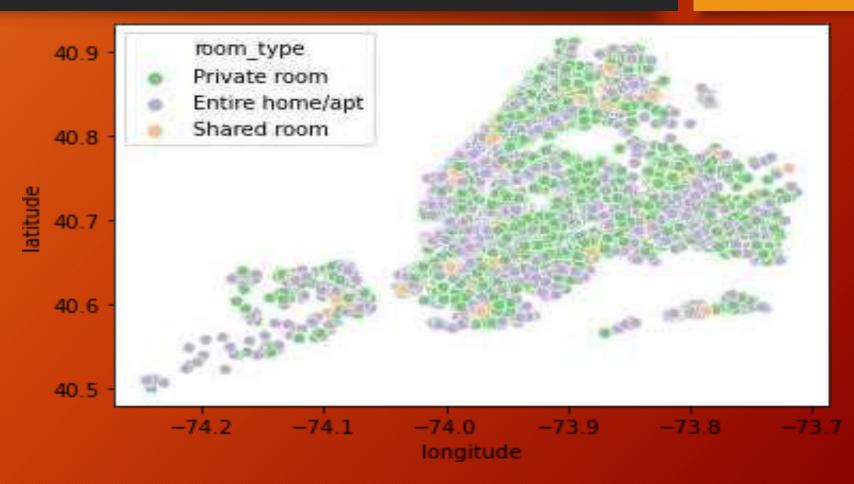
Longitude and latitude

We can see the geo locations of all neighborhood groups throughout
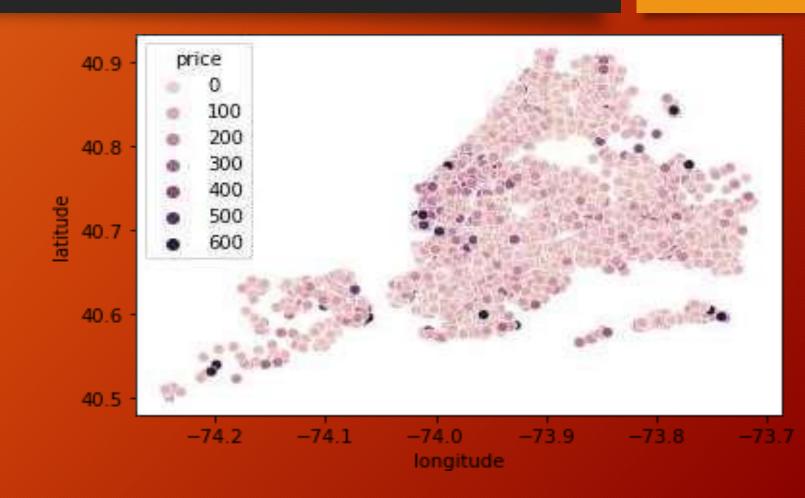New York.

Room types available in respective location

We can see the different room types available throughout New York

Price variations for different location

We can clearly see the places which is offering listings with maximum price throughout New York

Conclusion:

(1) Price depends on Neighbourhood_group. Its high in Manhattan.

(2) Within neighborhood groups, price fluctuates between the range. But Manhattan is the place where we see a lot ups and downs in price, giving the conclusion that Manhattan is the city which contains highest price and as well as lowest price.

(3) We can get the famous host (Sonder from Manhattan) from the number of properties he is offering (OR) We can find the famous host(Jordan from Queens) with respect to most reviewed place as well.

(4) People reviewed at most in properties of Queens which has the lowest prices for their properties as well.

(5) Throughout NY, Entire home/apt is the room type which is mostly in demand.

(6) Manhattan is the place which is famous and can be a good option for the companies to invest on properties of Entire home/Apt.