



airbnb

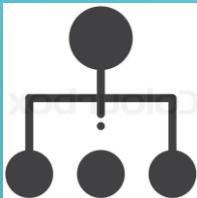
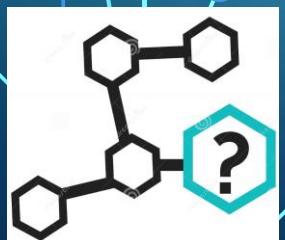
EXPLORATORY DATA ANALYSIS ON AIRBNB

*By :- Ajay Negi &
Mantresh Kumar*

Process under EDA



- Numbers
- Strings
- Boolean
- Undefined
- Null



Data Extraction

Identify Categorical and Numerical Columns

Check Data Types of Columns

Duplicate Removal

Handle Null Values

Process under EDA (cont.)



Handling Outliers

Invalid Data Fixing

Plotting Graphs

Observation & Conclusion

What Airbnb is and how does it works ?



Airbnb is an online marketplace that connects people who want to rent out their house, rooms or apartments with people who are looking for accommodation in specific locales.

The company has come a long way since its cofounders first came up with the idea, back in 2007 during recession, to invite paying guests to sleep on an air mattress in their living room. According to Airbnb's latest data, it has in excess of 5.6 million listings, covering more than 100,000 cities and towns and 220-plus countries worldwide.

Airbnb offers people an easy, relatively stress-free way to earn passive income by renting their property.

Guests often find Airbnb is cheaper, as more character, and is homier than hotels

Things To Be Discussed :

- No. of neighbourhoods in different group.
- Neighbourhood group which have most numbers of apartment.
- In which neighbourhood group people have stayed the most.
- Reviews according to each neighbourhood group.
- No. of reviews to each room types.
- Percentage of room types in every neighbourhood group.
- Comparing the prices according to each room type in every neighbourhood group.
- Room Types, people have stayed for most no. of nights.

Things To Be Discussed :

- Correlation between price, minimum nights and no. of reviews.
- Availability of rooms in different neighbourhood groups during each Year.
- Availability of different types of room during each year.
- Year, Airbnb received the most no. of reviews.
- Years, in which people stayed for most number of nights.
- Price fluctuation of each room type during each year.
- Month in which people spend most no. of nights according to each room type.
- Active hosts in different neighbourhood group.
- Conclusion

Data Extraction and Identify Columns

Airbnb Dataset
Rows 48895
Columns 16

Data is Extracted from a CSV Files

	<u>id</u>	<u>name</u>	<u>host_id</u>	<u>host_name</u>	<u>neighbourhood_group</u>	<u>neighbourhood</u>	<u>latitude</u>	<u>longitude</u>	<u>room_type</u>	<u>price</u>	<u>minimum_nights</u>	<u>number_of_reviews</u>	<u>last_review</u>	<u>reviews_per_month</u>	<u>calculated_host_listings_count</u>	<u>availability_365</u>
0	2539	Clean & quiet apt home by the park	2787	John	Brooklyn	Kensington	40.64749	-73.97237	Private room	149	1	9	2018-10-19	0.21	6	365
1	2595	Skylit Midtown Castle	2845	Jennifer	Manhattan	Midtown	40.75362	-73.98377	Entire home/apt	225	1	45	2019-05-21	0.38	2	355
2	3647	THE VILLAGE OF HARLEM,...NEW YORK !	4632	Elisabeth	Manhattan	Harlem	40.80902	-73.94190	Private room	150	3	0	Nan	Nan	1	365
3	3831	Cozy Entire Floor of Brownstone	4869	LisaRoxanne	Brooklyn	Clinton Hill	40.68514	-73.95976	Entire home/apt	89	1	270	2019-07-05	4.64	1	194
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	2018-11-19	0.10	1	0

Numerical Columns

1. Id
2. host_id
3. Latitude
4. Longitude
5. Price
6. Minimum nights
7. Number Of reviews
8. Reviews_per_month
9. Calculated_host_listings_count
10. Availability_365

Catagorical Columns

1. Name
2. Host_name
3. Neighbourhood_groups
4. Neighbourhood
5. room_Type
6. last_review

Date and Time

- ❖ last_review

Importing Folium and Adding Markers to show every neighbourhood groups

Folium is a Python Library used for visualizing geospatial data.

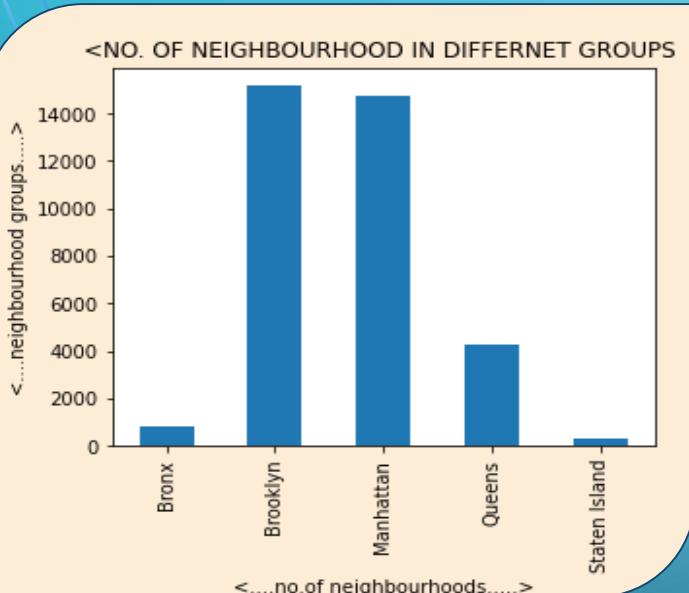


These 5 Places are also called Boroughs of New York :-

1. Manhattan
2. Brooklyn
3. Bronx
4. Queens
5. Staten Island

Together Forms The New York City.

Number Of Neighborhoods In Different Groups.

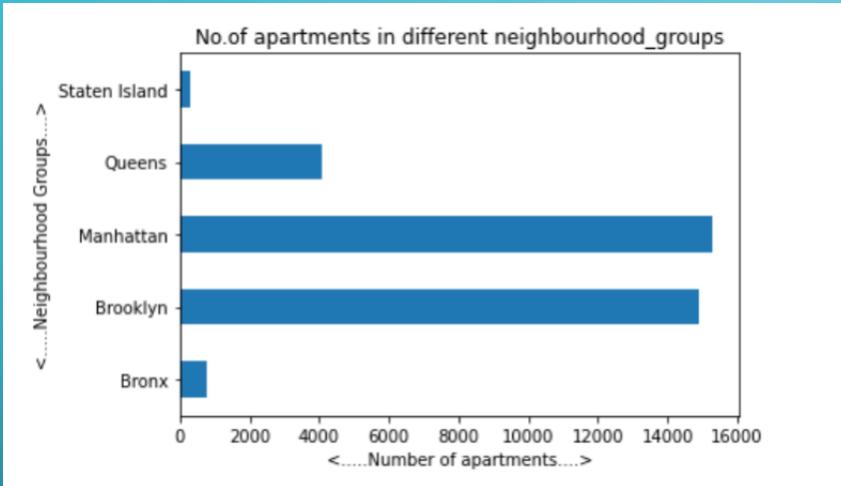


Neighbourhood Groups having the most no. of neighborhood

1. Brooklyn
2. Manhattan
3. Queens
4. Staten Island
5. Bronx

In the bar graph above represents the average price of apt. in every neighbourhood group ..so by seeing the graph ,what we get is the price of apt. in BRONX is cheapest which is below 70 ,whereas QUEENS and STATEN ISLAND have approx. same prices which falls under range of $70 \geq 80$ and the prices in BROOKLYN is up to 100 and MANHATTAN is most expensive that is up to 140.

Neighbourhood Groups Having The No. Of Apartments



Neighbourhood Groups having the highest number of apartments



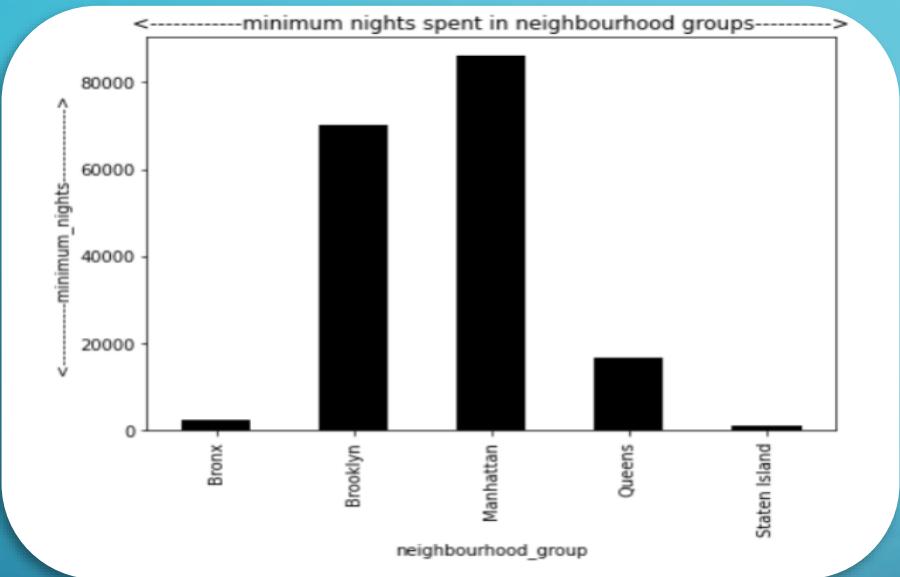
1. Manhattan
2. Brooklyn
3. Queens
4. Bronx
5. Staten Island

Reason behind highest number of apartment in Manhattan even though with much smaller area, is Density of population which is 28,872 persons/Km².

Manhattan alone contributes 71% GDP of New York City. This shows higher opportunity of growth which will attract population to settle for their survival.

The chances of growth in Queens is also much greater because the population is much higher than Manhattan. Airbnb have to work more to reach Queen's population

In which neighbourhood group people have stayed most



Neighbourhood group

Bronx 2420

Brooklyn 70085

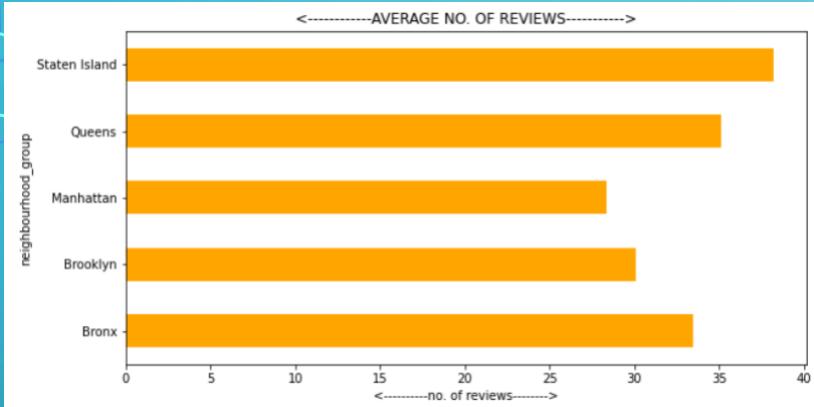
Manhattan 85923

Queens 16573

Staten Island 966

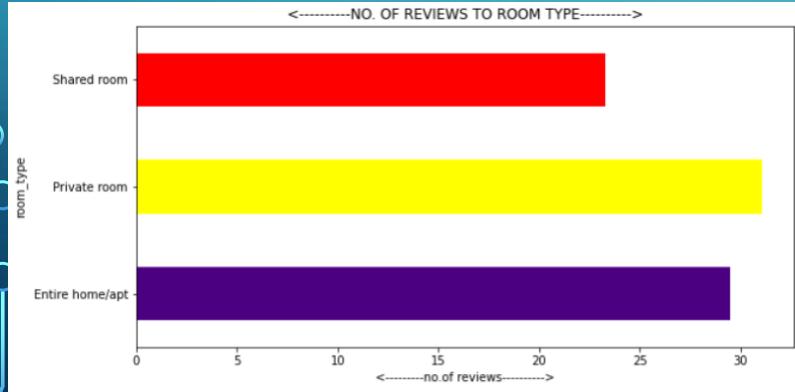
1. Most of the people have stayed in Manhattan and Brooklyn which is even more than 150,000+ minimum nights if combined.
2. Rest other neighbourhood groups are very less if compared even less than 20,000 minimum nights.

Reviews according to each Neighbourhood Group



1. Staten Island and Queens have got the highest avg. number of reviews which is $40 \geq 35$.
2. Manhattan and Brooklyn have got the lowest avg. no of reviews.
3. Whereas Bronx is somewhere between 30 to 35.

No of reviews to each room types

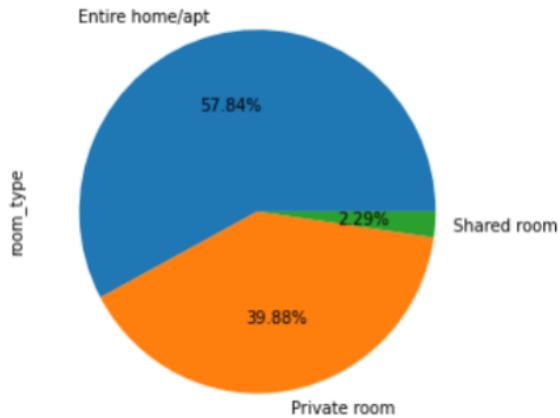


1. Private rooms have got the highest number of reviews more than 30.
2. After that Entire Home/Apt have got 2nd highest number of reviews a bit less than 30.
3. Whereas Shared rooms have got the lowest number of reviews which is betn $25 > 20$

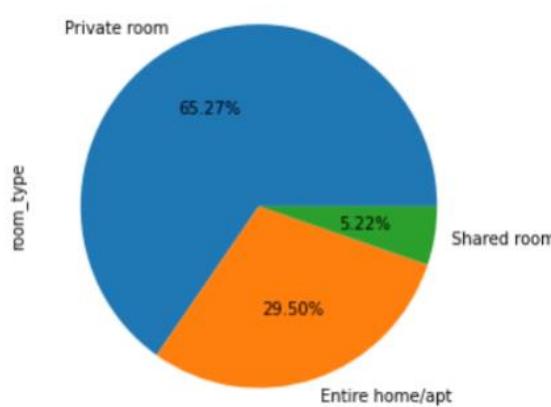
Room Types In Every Neighbourhood Groups



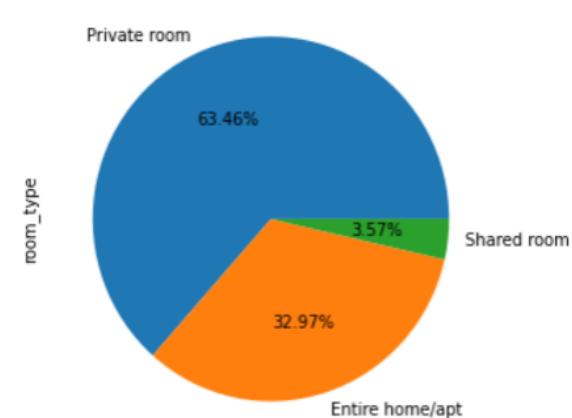
<....% OF ROOM TYPES IN MANHATTAN....>



<....% OF ROOM TYPES IN BRONX....>



<....% OF ROOM TYPES IN QUEENS....>



Observation

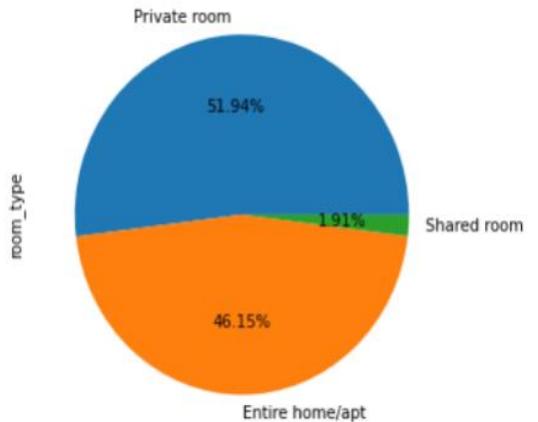
Only In Manhattan, Majority of the people rent out Entire home or apartments > Private room > Shared Rooms.

Whereas in Bronx, Brooklyn , Queens and Staten Island, Majority of the people rent out Private Rooms > Entire Home or Apartments > Shared Rooms.

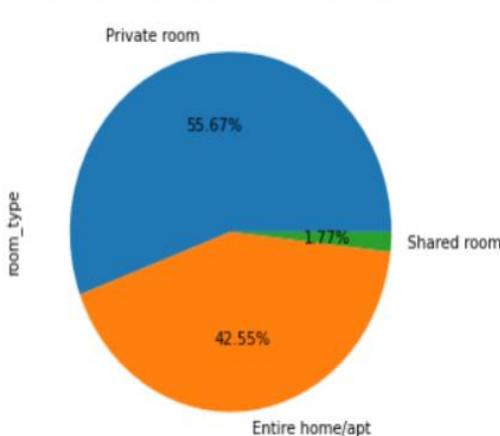
Room Types In Every Neighbourhood Groups



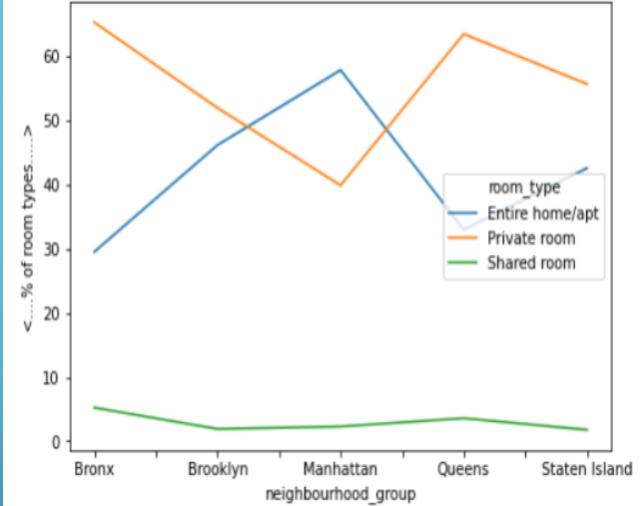
<....% OF ROOM TYPES IN BROOKLYN....>



<....% OF ROOM TYPES IN Staten Island....>



<.....% OF TYPES OF ROOMS IN ALL NEIGHBOURHOOD_GROUPS.....>

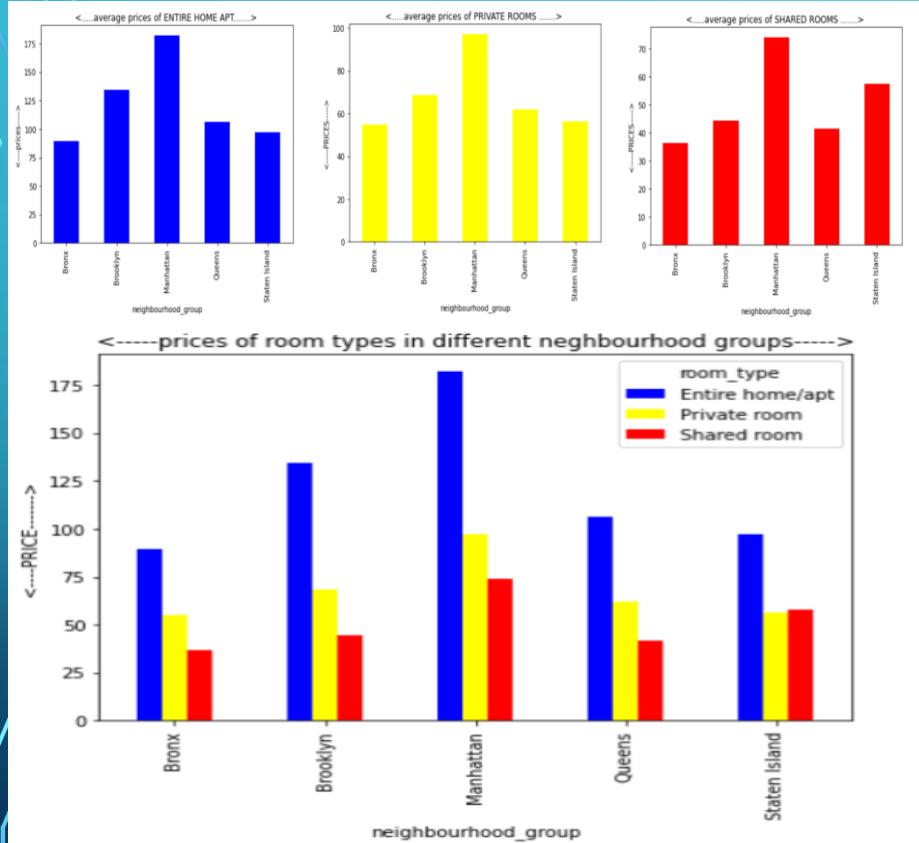


Observation

- In Brooklyn ,there are 52 % of Private room and 47 % of entire home apt and only 2 % of shared room

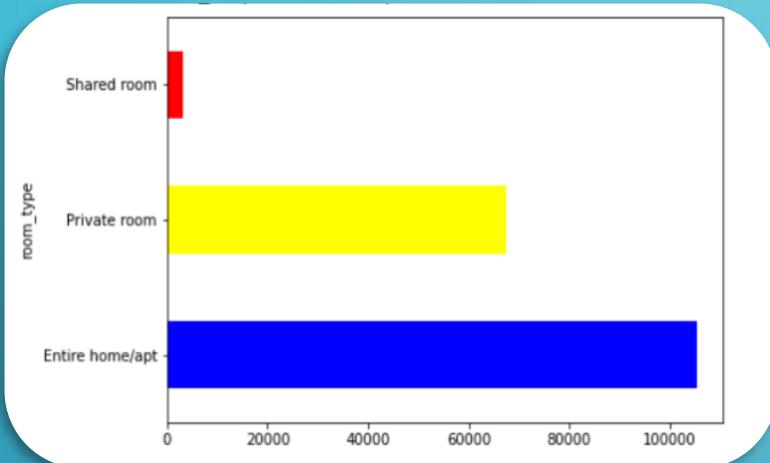
- In Staten Island there is 56% of Private rooms and 42% of entire home/apt and very less of shared room i.e. 1.70%.

\$ Comparing The Prices on the basis of Room Types in Every Neighbourhood Group \$



1. Even the shared rooms price in Manhattan is more than the price of Private Rooms in other four neighbourhood groups.
2. Bronx is the cheapest from all in terms of every room type.
3. Manhattan's Price is the highest among all in terms of all room types.
4. In Staten Island, Shared rooms price is higher than Staten Island and Bronx's Private rooms.
5. Manhattan's Private rooms price is even higher than Entire Home/Apartment's price in Bronx.

Room Types in which people have stayed for most number of nights.



Room type

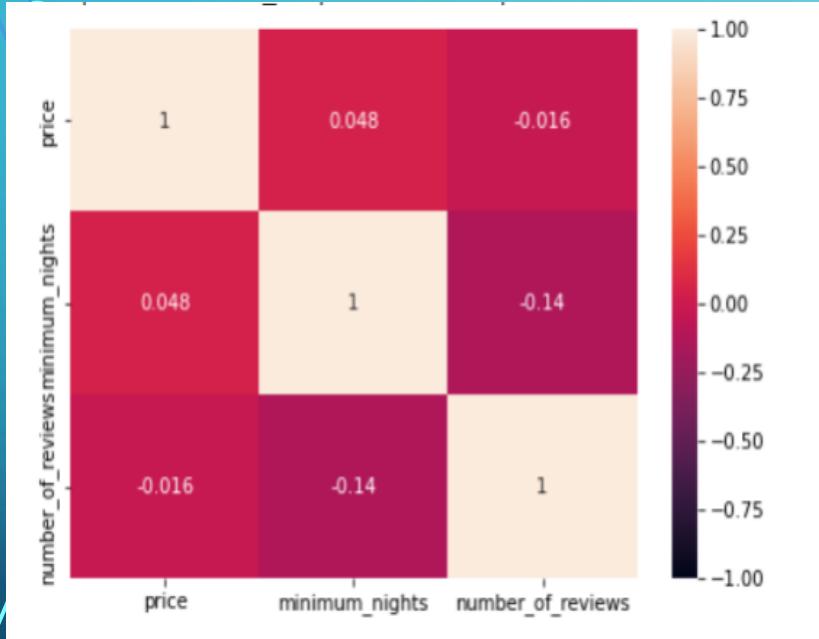
Entire home/apt 105390

Private room 67357

Shared room 3220

1. Most of the people have stayed in the Entire Home/Apt which is higher than 1000.
2. Secondly, Most people have preferred Private rooms to stay which is between 8000 to 6000.
3. Whereas in Shared rooms very less people have preferred to stay which is even less than 1000.

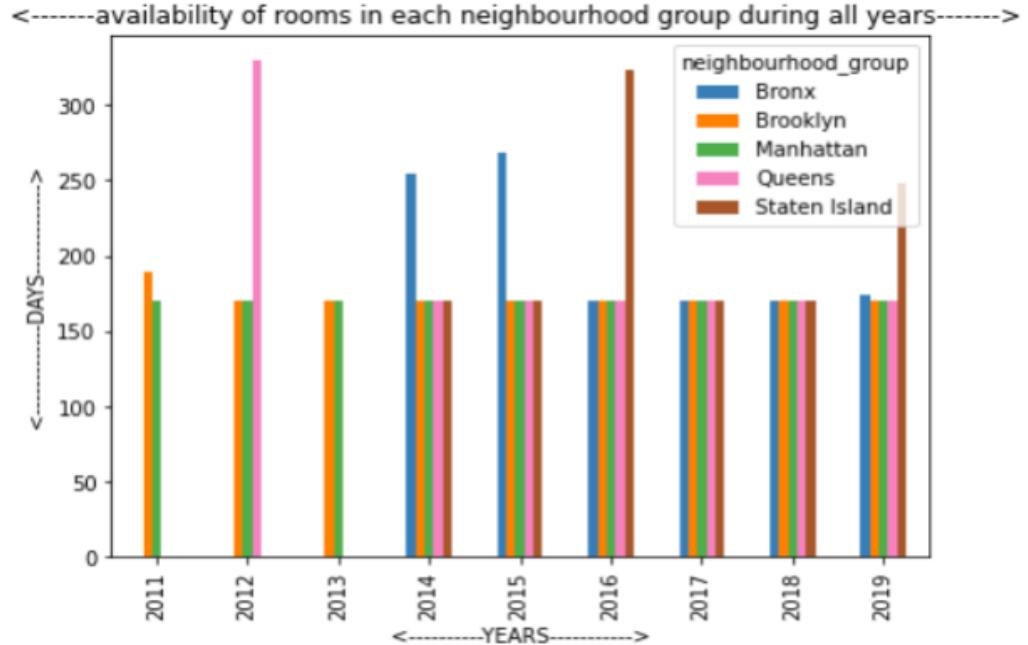
Let's see the correlation between price ,minimum nights and number of reviews



There is negative correlation between number of reviews and price which is -0.016.

There is negative correlation between number of reviews and minimum nights which is -0.14.

Availability of rooms in different neighbourhood groups during each year



Year 2019 : Rooms are available in all neighbourhood and availability is the highest in **Staten Island(200+ days)** and in **Bronx** slightly higher than rest excluding **Staten Island**.

Year 2017 & 2018 : Rooms are available in all neighbourhood and availability of all are also same.

Year 2011 : Rooms are available only in **Manhattan and Brooklyn** in which availability was higher in Brooklyn than Manhattan

Year 2012 : Rooms are available only in **Manhattan, Brooklyn & Queens** in which availability was much higher in Queens (300+ days) whereas in Brooklyn and Manhattan remains same (200>150).

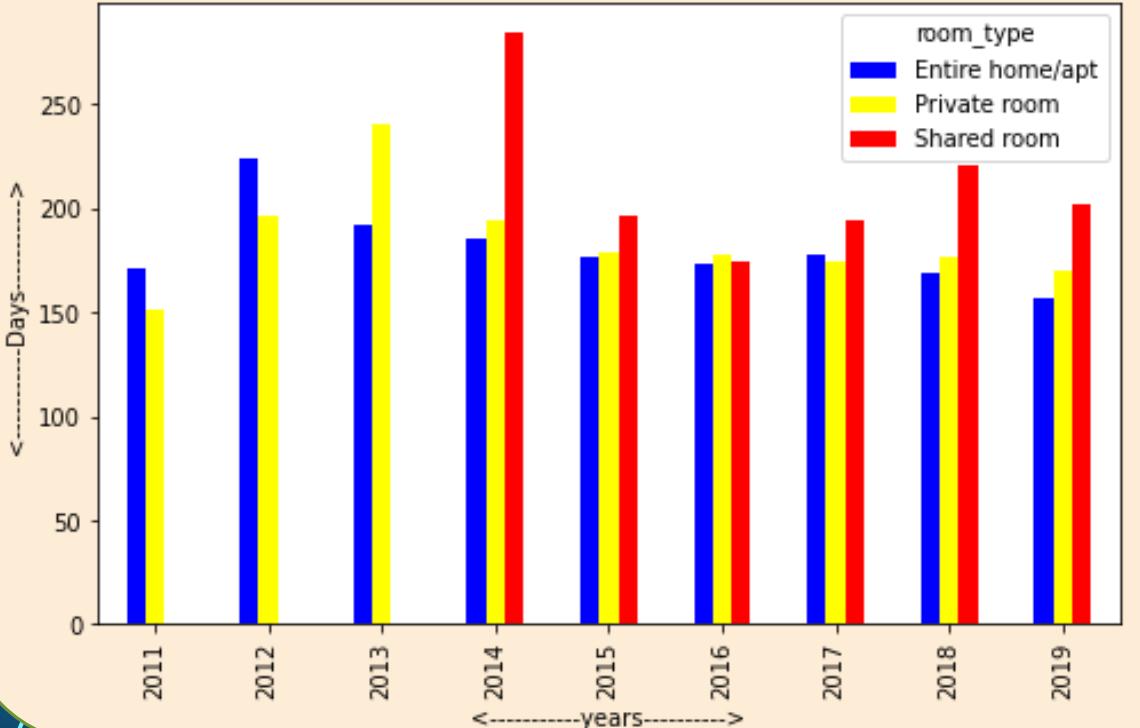
Year 2013 : Rooms are available only in **Manhattan & Brooklyn** and remains same like previous year.

Year 2014 & 2015 : Rooms available in all neighbourhood but **Bronx** has highest availability.

Year 2016 : Rooms are available in all neighbourhood but **Staten Island** has highest availability.

Availability of different types of room during each year

<-----availability of room types during each year----->



Year 2011: Only **Entire home/apt** (150+Days) and **Private rooms** (150 days) are available. No shared rooms were available.

Year 2012: Only **Entire home/apt** (200+Days) and **Private rooms** (200 > 150days) are available. No shared rooms were available.

Year 2013: Only **Entire home/apt** (200 > 150days) and **Private rooms** (200+days) are available. No shared rooms were available.

Year 2014: Availability Series (High to Low)
Shared Room > Private room > Entire Room

Year 2015: Availability Series (High to Low)
Shared Room > Private room > Entire Room

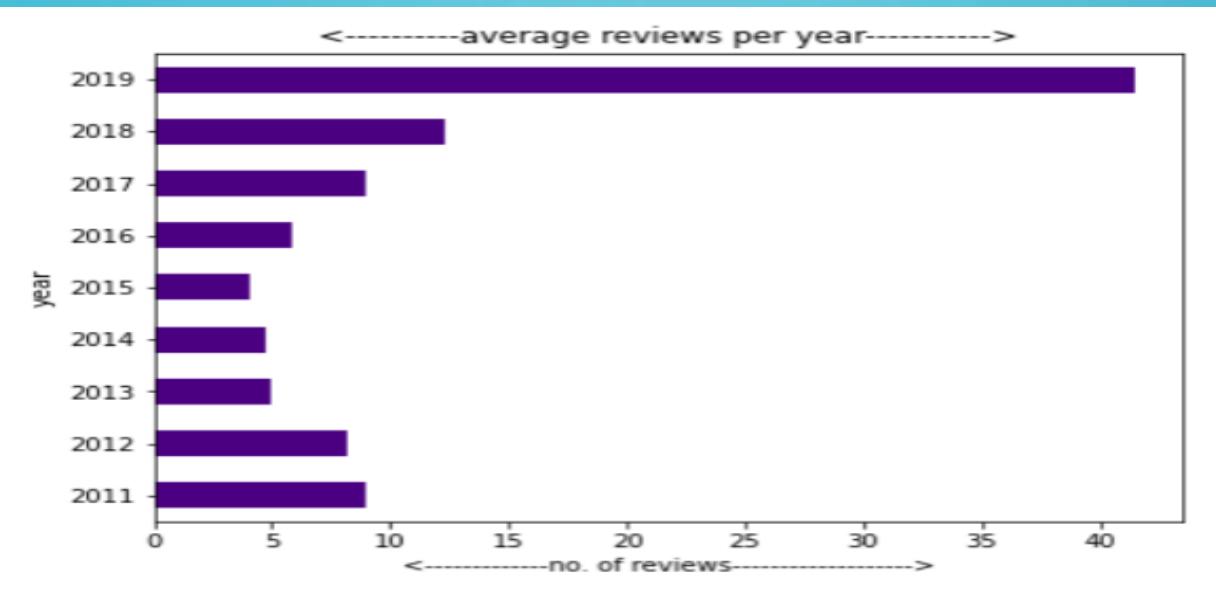
Year 2018 & 2019:

Availability Series (High to Low)
Shared Room > Private room > Entire Room

Year 2017: Availability Series (High to Low)
Shared Room > Private room > Entire Room

Year 2016: Availability Series (High to Low)
Private room > Shared Room > Entire Room

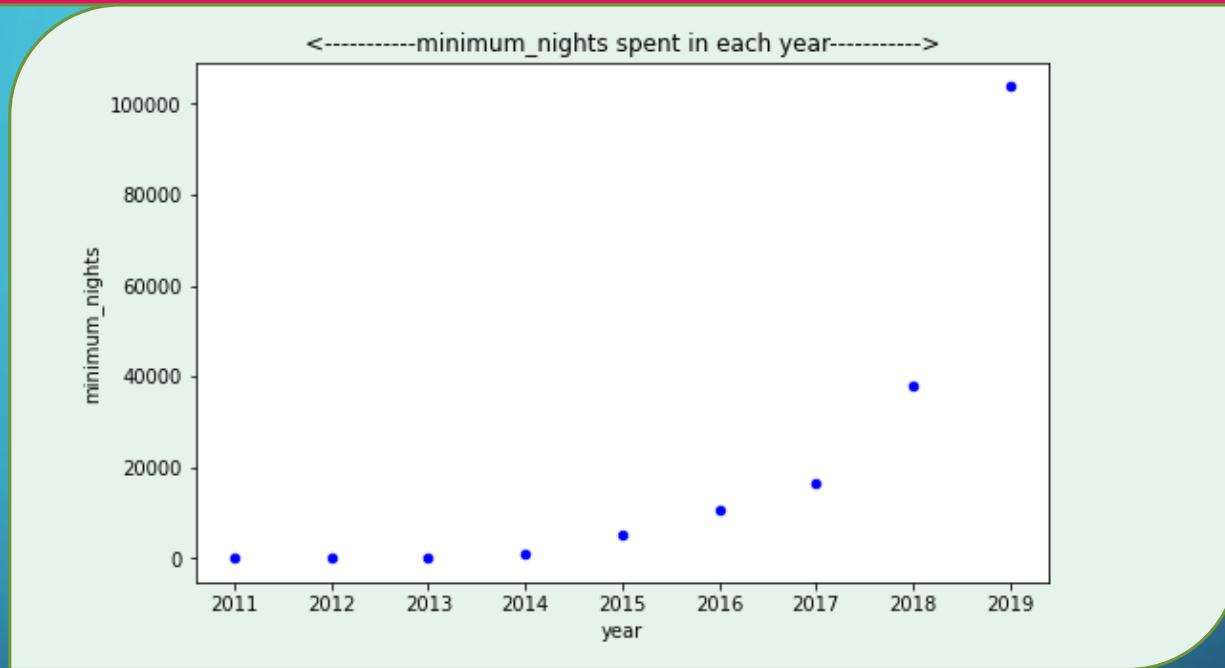
Year in which Airbnb receives the most no. of reviews



In the Year 2015 Airbnb receives the lowest number of reviews which is even less than 5.

In the Year 2019 Airbnb receives the highest number of reviews which is more than 40.

Years, in which people stayed for most numbers of nights 😊

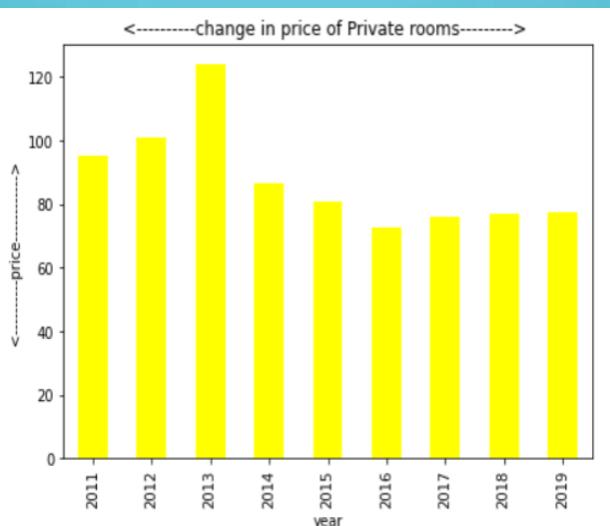


In this scatter plot, there is a positive correlation between the years and the minimum nights. As the years increase from 2011 to 2019, the number of minimum nights also increases. 📈

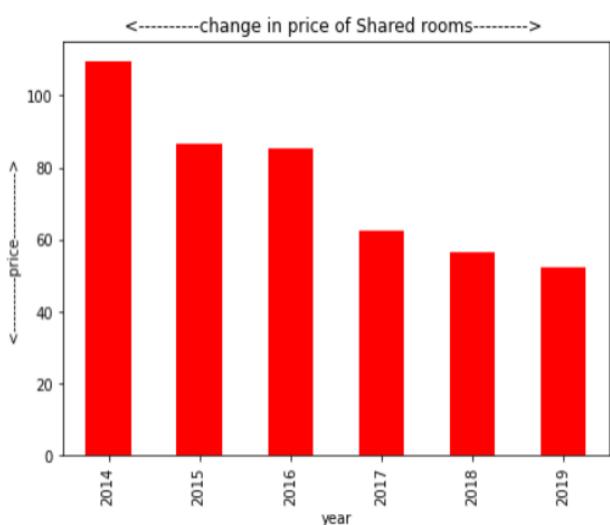
Price fluctuation of each room type during each year



Entire Home/Apt.

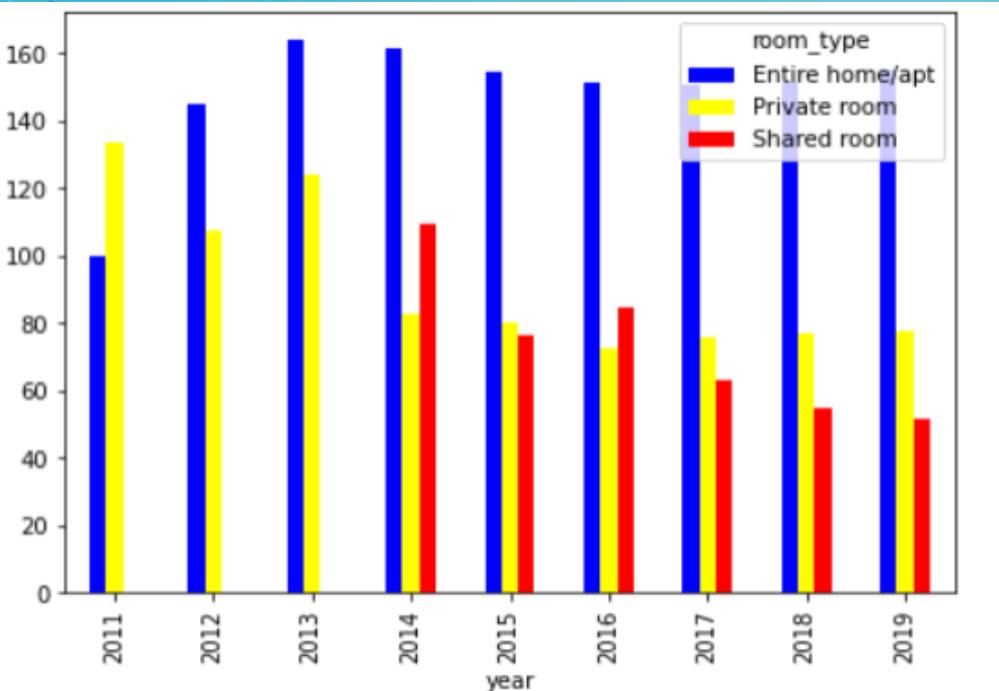


Private Rooms



Shared Rooms

Price fluctuation of each room type during each year



Entire Home/Apt.

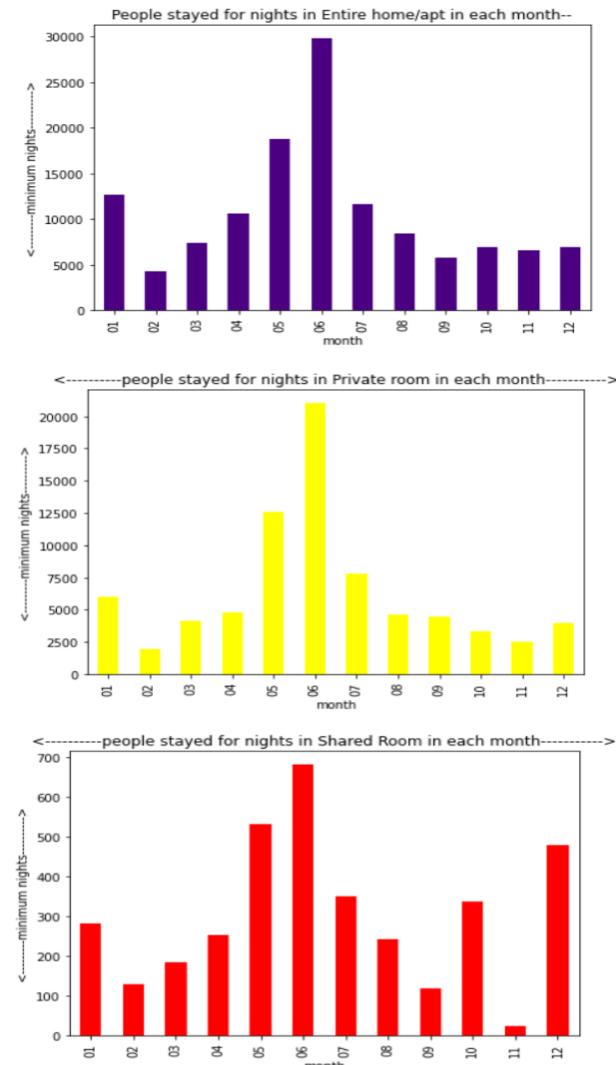
The prices of the entire home apt in 2011 were minimum and it increased and reached its max till the year 2013 but after 2013, the prices were nearly the same i.e. between 140 to 160.

Private Rooms

In 2011 the prices of private rooms were highest i.e. up to 140 and from 2014 the prices were almost the same all the years i.e. between 60 to 80.

Shared Rooms

As we can see in the above bar graph it is showing the data from the year 2014 because the shared rooms were introduced by Airbnb in 2014.



几个月中人们在不同类型的房间中度过最多的时间

对于整个家庭/公寓

- ✓ 根据数据，人们在整套房屋/公寓中度过最多的时间是在六月。 ✓

对于私人房间

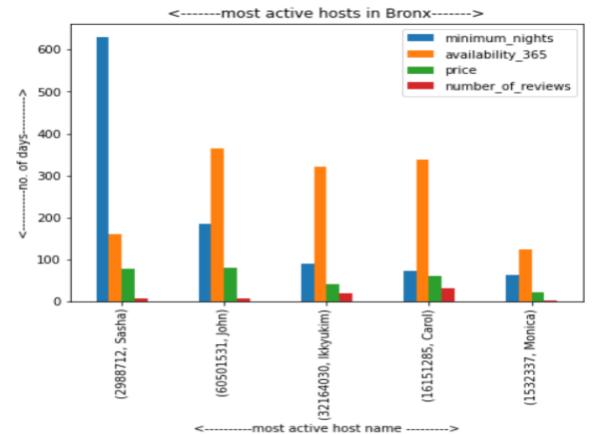
- 🏠 人们在私人房间里度过的时间最长，六月达到约21,000晚。

对于合租房间

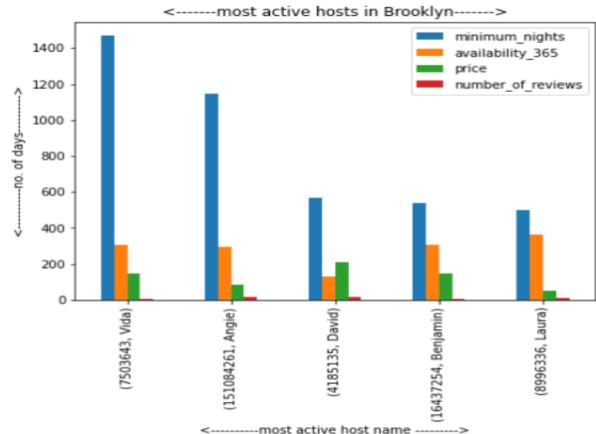
- ✓ 在上述图表中，我们可以看到人们在六月度过了最多的夜晚。 ✓

Most active host in different neighbourhood group

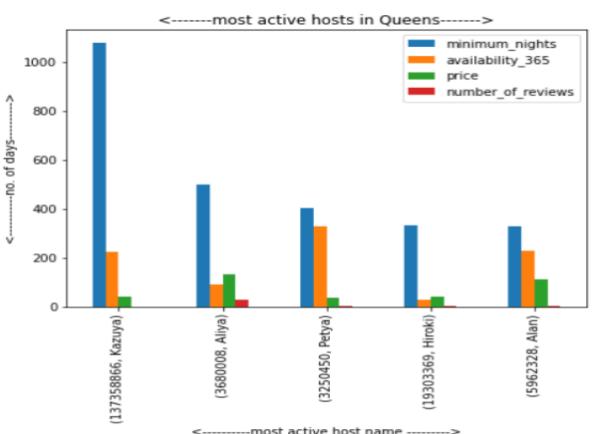
Bronx



Brooklyn



Queens



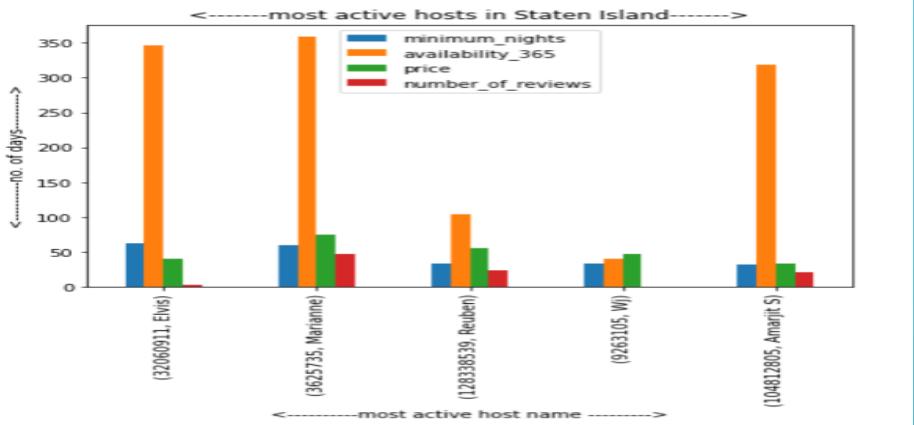
The above graph states the relationship of minimum nights with other factors such as availability of rooms, price of rooms, and the number of reviews.

In Brooklyn, these are the top 5 active hosts, whose room's availability was higher during each year and the people have stayed for most no. of nights in their apartments.

In Queens, these are the most active hosts, but if we compare all these 5 hosts, Kazuya has been the most active host in this neighborhood, there is a huge difference between Kazuya and the other 4 hosts.

Most active host in different neighbourhood group

Staten Island

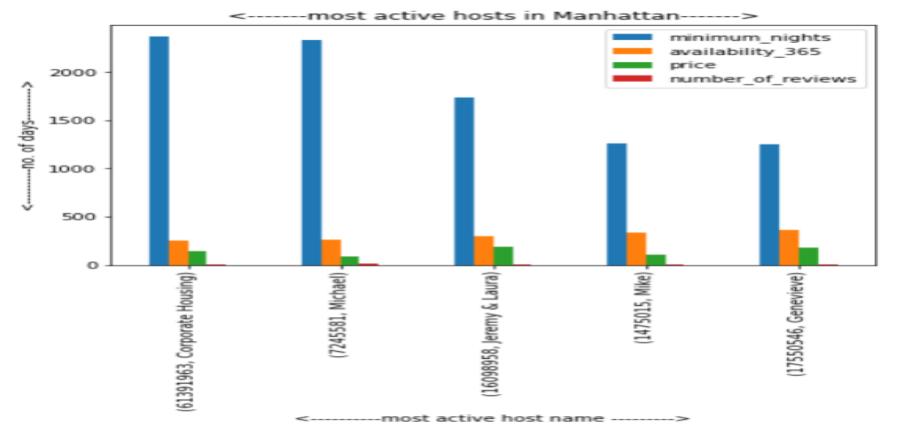


minimum_nights availability_365 price number_of_reviews

host_id	host_name	minimum_nights	availability_365	price	number_of_reviews
32060911	Elvis	63	345.500000	40.000000	3.500000
3625735	Marianne	60	358.000000	75.000000	47.500000
128338539	Reuben	33	104.500000	55.500000	24.500000
9263105	Wj	33	40.500000	47.500000	1.000000
104812805	Amarjit S	32	318.375000	33.375000	21.000000

In Staten Island, the night spent by people is low as compared to other neighbourhood groups because the no. of the apartment is less here.

Manhattan



minimum_nights availability_365 price number_of_reviews

host_id	host_name	minimum_nights	availability_365	price	number_of_reviews
61391963	Corporate Housing	2370	250.042057	144.620253	5.278481
7245581	Michael	2328	260.263158	88.789474	14.842105
16098958	Jeremy & Laura	1740	302.820356	188.448276	2.327586
1475015	Mike	1260	335.000000	102.785714	3.857143
17550546	Genevieve	1250	365.000000	180.000000	2.000000

In the above graph of the most active host in Manhattan, the night spent by people are higher and the reason behind this could be the availability of their apartments are higher as compared to other hosts and the negative correlation between price and minimum night spend.

Conclusions

- **Manhattan alone contributes 71% GDP of New York City.** This shows a **higher opportunity** for growth, that is why people mostly prefer to stay in Manhattan the reasons could be anything like for their work, etc.
- **Prices are higher in Manhattan,** the reason is that people want to stay in Manhattan because people who live far from their homes for their work. Airbnb provides the feeling of home rather than a feeling of staying in hotels.
- People have stayed mostly in Entire Home/apartments that could be the reason that private rooms have more number of reviews because if people stay in an entire home apt. for a long period, it gives them the feeling of home which is a purpose of Airbnb that's why **reviews are given to hotels not homes.**
- People have stayed for **most no. of nights in Manhattan** but the maximum no. of reviews are given to Staten Island, as we stated the reason in the above point that people with permanent residency, don't give reviews and in Staten Island people have stayed for least no. of nights.

Conclusions

- In Bronx, Manhattan, Queens, and Staten Island there is a **negative correlation between the no. of entire home apt and private rooms**. In *Bronx, Queens, and Staten Island*, in future, **the no. of entire home apt. could go up than the no. of private rooms** and in *Manhattan*, **the no. of private rooms could go up than the no. of entire home apt.**
- As we saw the graph, people have stayed for most no. of nights in an entire home apt and private room and for least no. of nights in a shared room, which shows that **people want to maintain their privacy**, they mostly avoid sharing room with another person, that is why people have stayed for least no. of nights in shared rooms.
- In all the neighborhood groups, **Manhattan is the most expensive** neighborhood group in New York according to the prices of all room types and the cheapest is the Bronx.
- Even after being a most expensive group in New York, people have mostly loved to stay in Manhattan.
- There was a negative correlation between no. of reviews and minimum nights and as we had explained in the earlier point, reviews are given to hotels, not to homes, therefore more no. of the night spend, lesser the reviews.

Conclusions

- The **shared rooms** were not introduced till the year 2013, in 2014 the shared rooms were introduced.
- In terms of **availability**, only *Manhattan* and *Brooklyn* are the neighborhood groups, that were consistent during all the years.
- Airbnb got the most no. of reviews in 2019 because the people stayed for most no. of nights in the year of 2019.
- The prices of entire home apt were least in the year 2011 and it kept increasing till the year 2013, but after 2013 till the year 2019, the prices of entire home apt are nearly same, therefore we could expect that in upcoming years the prices of entire home will be same i.e. between 140 to 160.
- The prices of private rooms were highest in 2013, after 2014, the prices of the private room were nearly the same, so we could expect that the prices of private rooms in the upcoming year will be the same as last year.
- When the shared room was introduced in 2014, the prices were highest in that year but due to low demand, the prices of the shared room kept going down, therefore we can assume that in next year, the shared rooms will be cheaper.
- The price difference between the shared room and private, entire home apt is huge, yet people chose to stay in private rooms and entire home apt, which shows that **the per capita income** of people in New York is enough that they can **afford private rooms and entire home apt**.