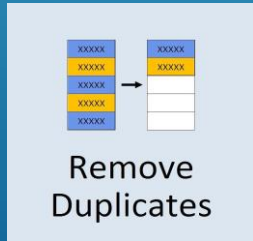
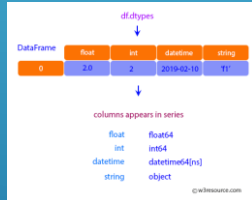
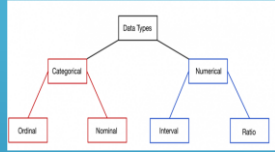


EXPLORATORY DATA ANALYSIS ON



By : NITIN KUMAR

Contents in EDA



Data Extraction

Identify Categorical & Numerical Values

Check Data Types of Column

Duplicate Removal

Handle Null Values

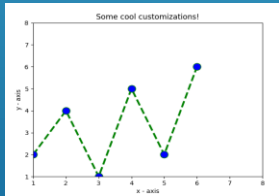
Contents in EDA



Outliers Handling



Fixing Invalid Data



Plotting Graphs

Observation & Conclusion

What is Airbnb & how does it works ?

Airbnb, Inc. is an American company that operates an online marketplace for lodging, primarily homestays for vacation rentals, and tourism activities.

The Company has come a long way since its co-founders first come up with the idea, back in 2007 back in 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 listing of 5.6 millions, covering more than 100,000 cities and towns in more than 220 countries.

Points for Discussion :

- No. of apartments in different neighbourhood_groups
- Price of apartments varying in different neighbourhood groups
- Percentage of room types in every neighbourhood groups
- Prices of room types in different neighbourhood group individually
- Comparing the prices of room types for different neighborhood groups
- Room type people prefer the most for staying
- Neighborhood Group which got the most no. of visitors
- Correlation between price, minimum nights & no. of review

Points for Discussion :

- Availability of rooms in different 'neighbourhood_groups' Yearly
- Availability of different types of room Yearly
- In which year Airbnb receives the most no. of reviews
- Which year brought the most no. of visitors
- Price fluctuation of each room type around the year
- Which room types people prefer around the different months
- Most no of active host in different neighbourhood group
- Conclusion

Data Extraction & Identifying Columns

(Data is extracted from a CSV file)

	id	name	host_id	host_name	neighbourhood_group	neighbourhood	latitude	longitude	room_type	price	minimum_nights	number_of_reviews	last_review	reviews_per_mont
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	0.2
1	2595	Skylit Midtown Castle	2845	Jennifer	Manhattan	Midtown	40.75362	-73.98377	Entire home/apt	225	1	45	2019-05-21	0.3
2	3647	THE VILLAGE OF HARLEM.....NEW YORK !	4632	Elisabeth	Manhattan	Harlem	40.80902	-73.94190	Private room	150	3	0	NaN	Na
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	4.6
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	0.1

*Airbnb dataset
has 49000 rows &
16 columns*

Importing Folium & adding markers to show every neighborhood group

(Folium is a library in python used for visualizing Geospatialdata)

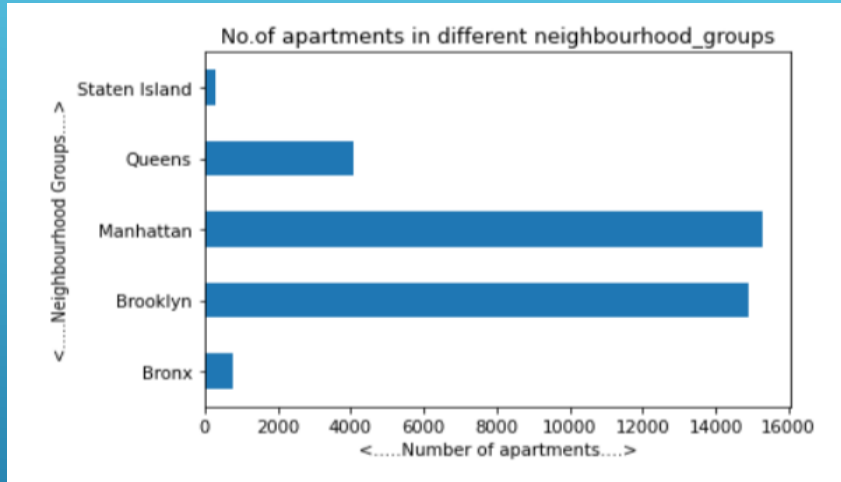


*These 5 places are also called
Boroughs of New York :-*

1. *Manhattan*
2. *Brooklyn*
3. *Bronx*
4. *Queens*
5. *Staten island*

Together forms the new York city

No. of apartments in different neighbourhood_groups



Neighbourhood groups having highest no. of apartment

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

Observation

Manhattan has the highest no of apartments followed by Brooklyn & Queens, but there is very huge gap between 2nd & 3rd neighbour. The 4th no neighbour is Bronx & the Staten Island has the least no of apartments in the whole neighbourhood

Price of apartments varying in different neighbourhood groups



Prices according to neighbourhood groups

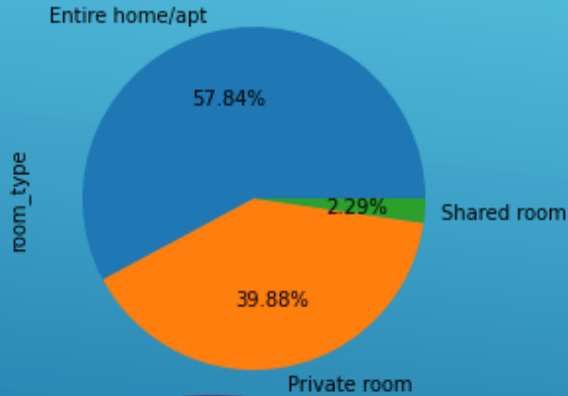
1. Manhattan
2. Brooklyn
3. Queens
4. Staten Island
5. Bronx

Observation

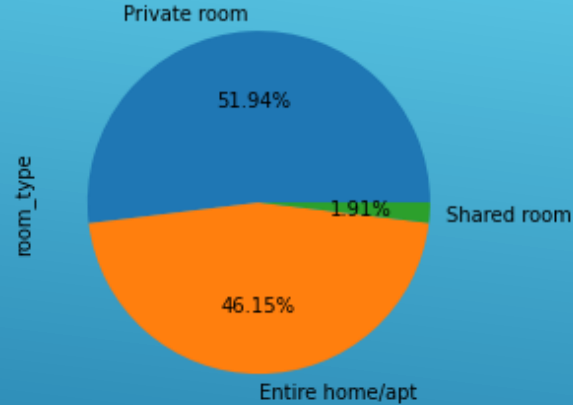
We can see that Manhattan offers the highest price to the people & Bronx gives the cheapest deal to people in night staying

Percentage of room types in every neighbourhood groups

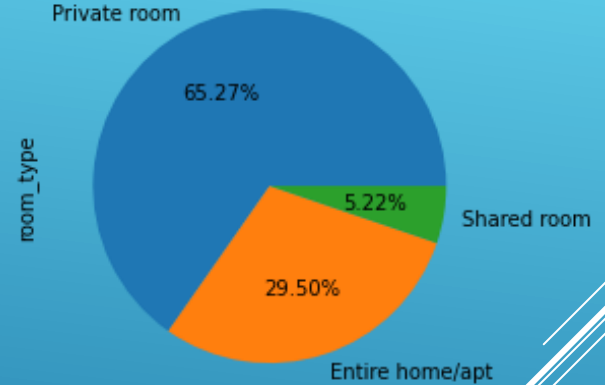
<.....% OF ROOM TYPES IN MANHATTAN.....>



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



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



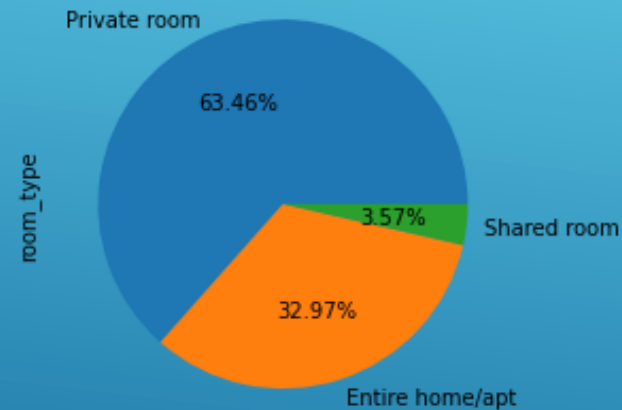
Observation

In Manhattan mostly the people rent their Entire home > Private room > Shared room

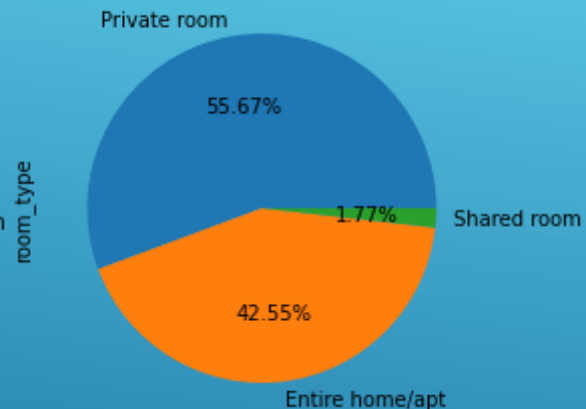
But in Brooklyn, Bronx, Queens, Staten Island mostly rent their Private room > Entire home > Shared room

Percentage of room types in every neighbourhood groups

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



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



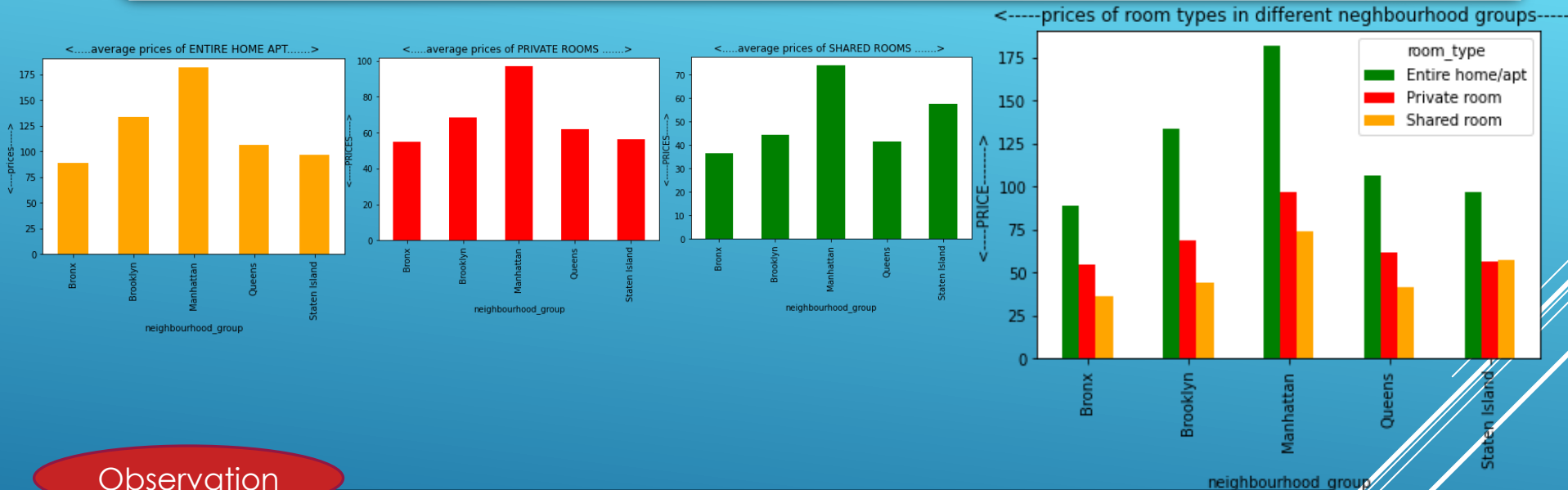
<.....% OF TYPES OF ROOMS IN ALL NEIGHBOURHOOD_GROUPS.....>



Observation

Bronx & Queens have higher share of Private room & Shared room as compared to Brooklyn & Staten Island

Comparing the prices of room types for different neighborhood groups



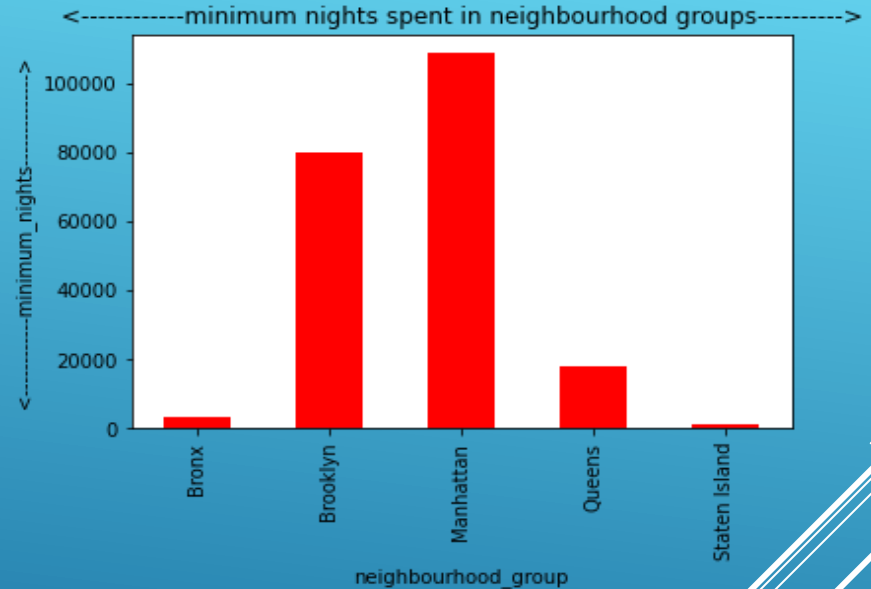
Observation

1. Manhattan Offers highest price among all the places for all room types
2. Bronx Offers the cheapest among all the places for all room types
3. Manhattan's Private rooms price is higher incompare to Bronx Entire home
4. Manhattan Shared rooms price is higher in compare to other four neighbourhood Private rooms price
5. In Staten Island, Shared rooms price is higher than it's own & Bronx's Private rooms

Neighborhood Group which got the most no. of visitors

Neighbourhood group

1. Manhattan - 108657
2. Brooklyn - 79806
3. Queens - 17758
4. Bronx - 3113
5. Staten Island - 939



Observation

1. Manhattan & Brooklyn have a great no. of visitors
2. Rest neighbours have very less visitors as compared to these two

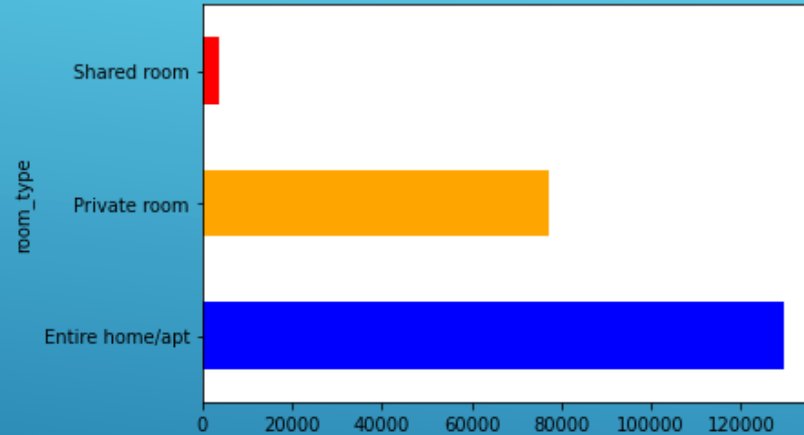
Room type people prefer the most for staying

Room type :

Entire Home/Apt - 129486

Private room - 77170

Shared room - 3617



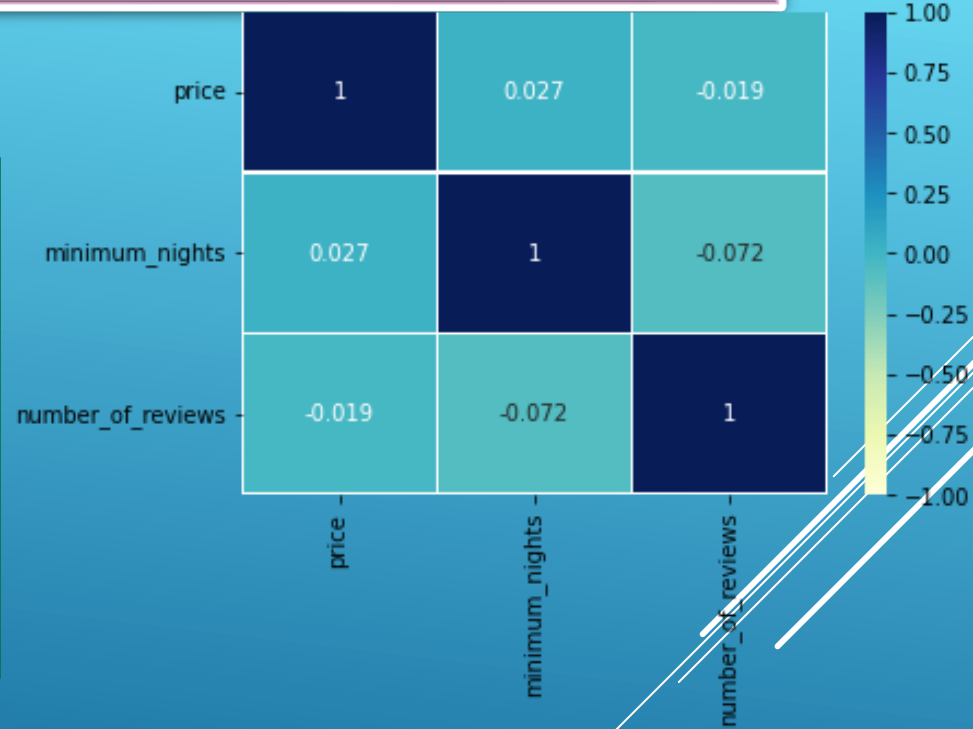
Observation

1. Most of the people prefer Entire home/apt then the other 2 options, it is because people prefer privacy & comfort instead of compromising in cheaper options
2. Private rooms & Shared rooms are very less in no. instead of being a cheaper option to people

Correlation between price, minimum nights & no. of review

Observation

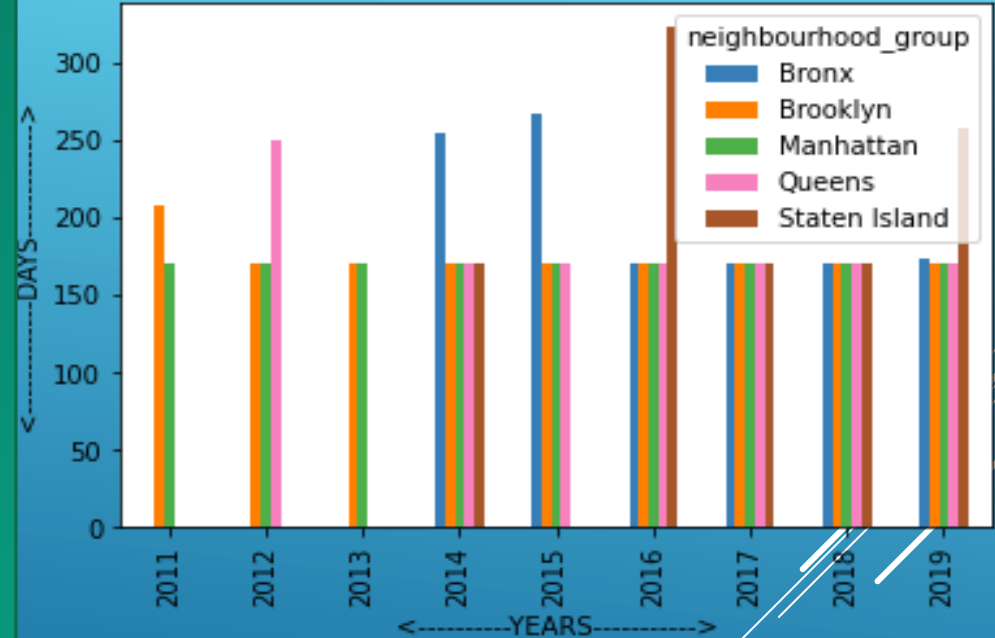
1. We are getting Negative correlation between price & number of reviews which is -0.019
2. Also we get a Negative correlation of minimum nights & number of reviews which is -0.072



Availability of rooms in different 'neighbourhood_groups' Yearly

- **In 2011:** Rooms are available only in Manhattan & Brooklyn, Brooklyn had higher availability of rooms in 2011
- **In 2012:** Queens comes in the competition With the other two but was higher in availability, whereas Manhattan & Brooklyn had same number of availability
- **In 2013:** Again the same Rooms are available only in Manhattan & Brooklyn
- **In 2014 & 2015:** Rooms are available in all the neighbourhood group but Bronx has highest availability
- **In 2016:** Rooms are available in all neighbourhood group but has higher availability in Staten Island
- **In 2017 & 2018:** Rooms are available in all neighbourhood group & the availability also remains same for all
- **In 2019:** Rooms are available in all neighbourhood group & availability is highest in Staten Island & Bronx goes slightly up from rest of 3 in availability

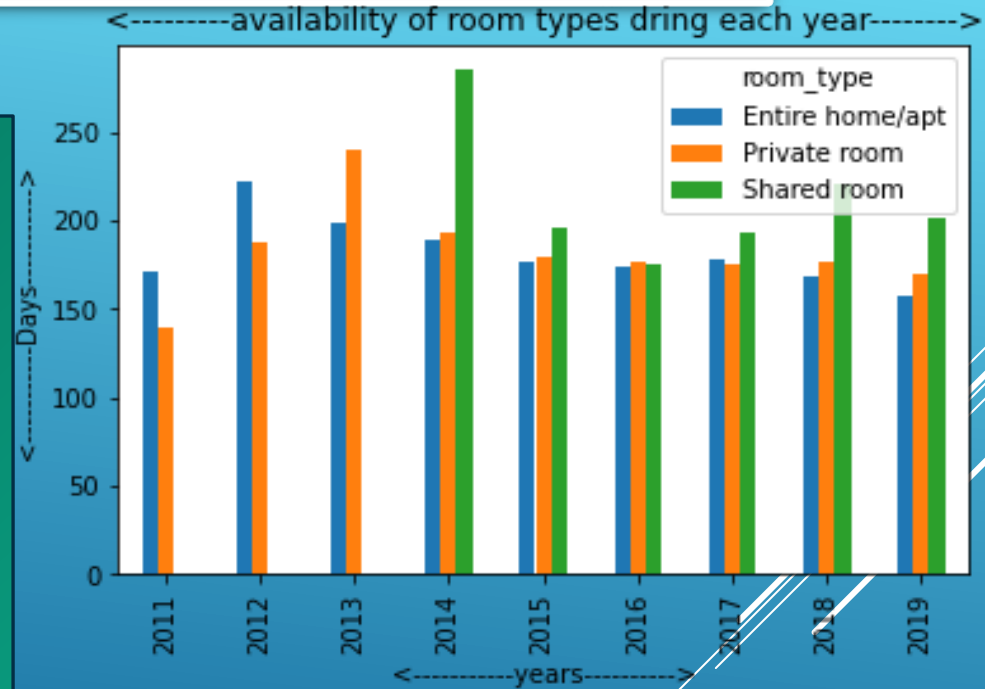
<-----availability of rooms in each neighbourhood group during all years----->



Availability of different types of room Yearly

Observation

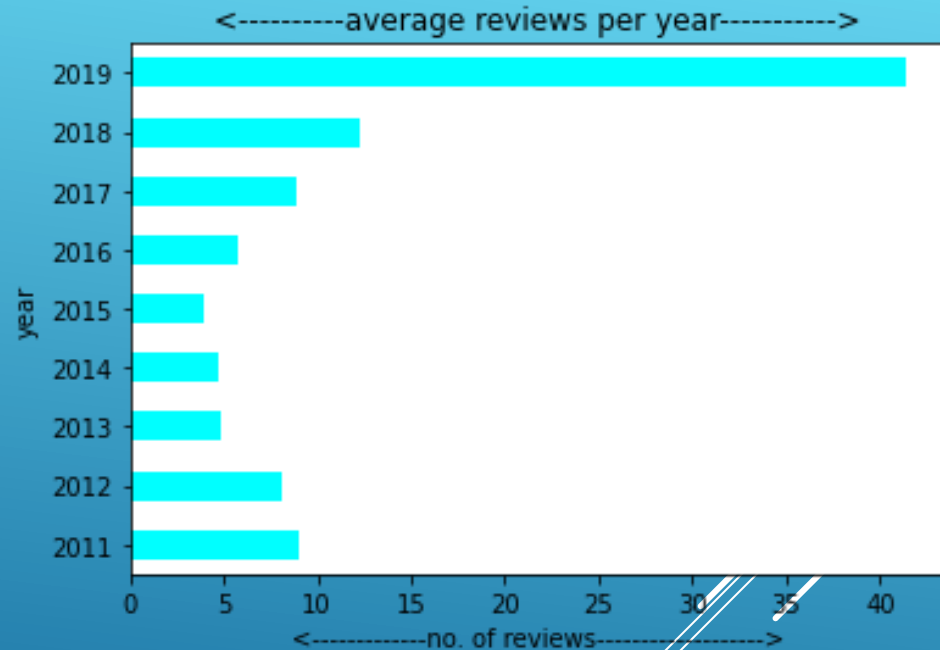
- **In 2011:** Only the Entire home/apt & Private rooms options were available to people.
- **In 2012:** Availability was increased in both options but was higher in Entire home/apt
- **In 2013:** Availability of Private rooms increased & Entire home/apt decreased as of previous year
- **In 2014:** Shared rooms came in as an option, availability series (high to low) Shared room > Private room > Entire home/apt
- **In 2015:** Availability series was same as of 2014 but the shared room availability came down
- **In 2016:** Availability series (high to low) private room > shared room > entire home/apt
- **In 2017:** Availability series (high to low) Shared room > entire home/apt > private room
- **In 2018 & 2019:** Availability series (high to low) Shared room > Private room > Entire home/apt.



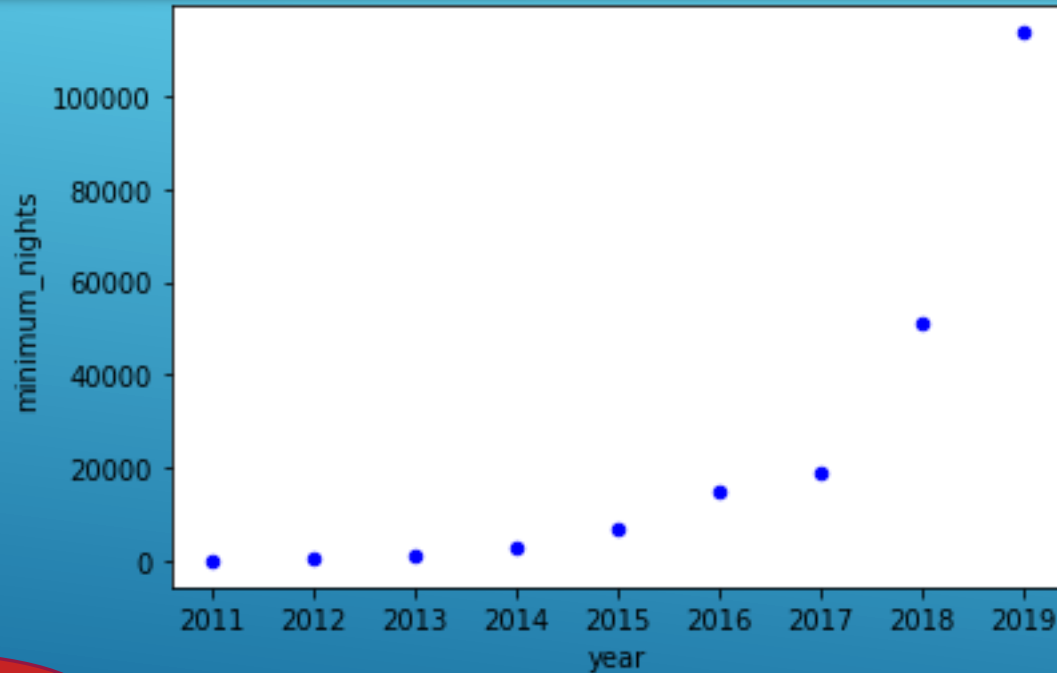
In which year Airbnb receives the most no. of reviews

Observation

- In 2015 Airbnb receives the least no of reviews
- In 2019 Airbnb receives the highest no of reviews



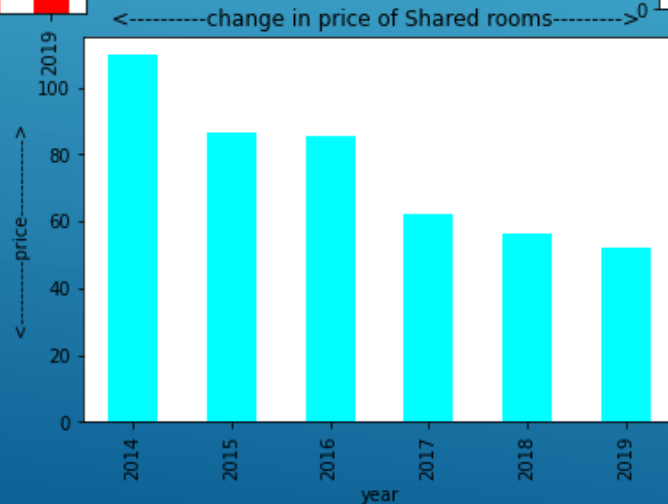
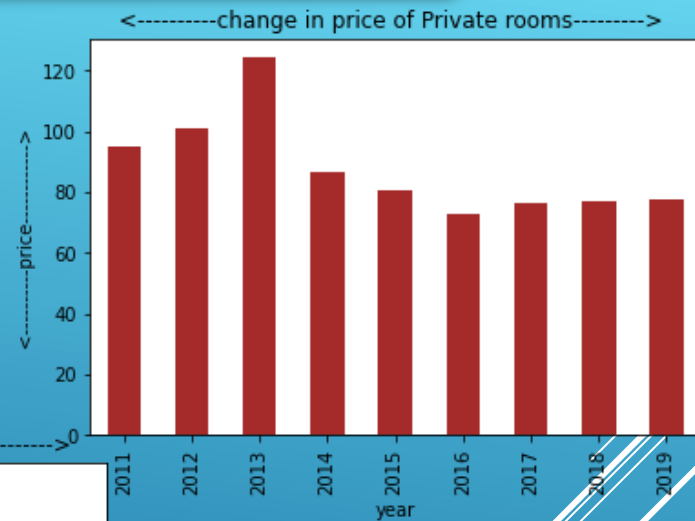
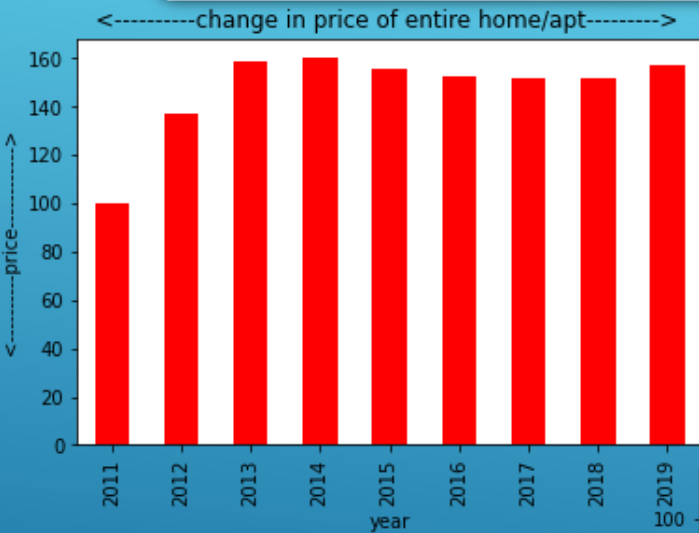
Which year brought the most no. of visitors



Observation

We can see a positive correlation between year & minimum nights in this scatter plot, as the years increase from 2011-2019 the minimum nights also increases

Price fluctuation of each room type around the year



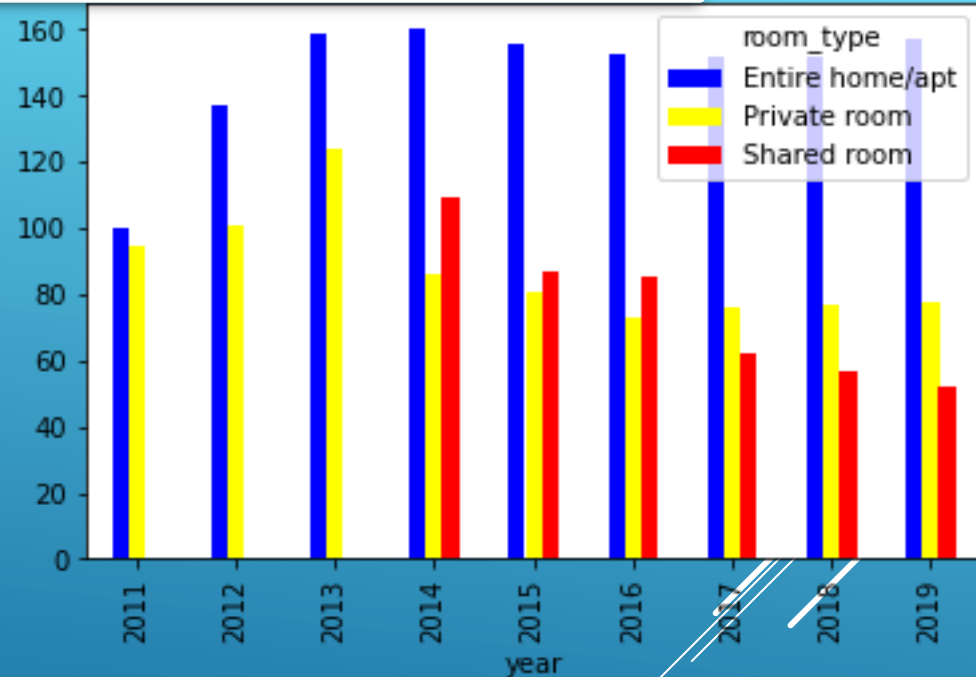
Price fluctuation of each room type around the year

Observation

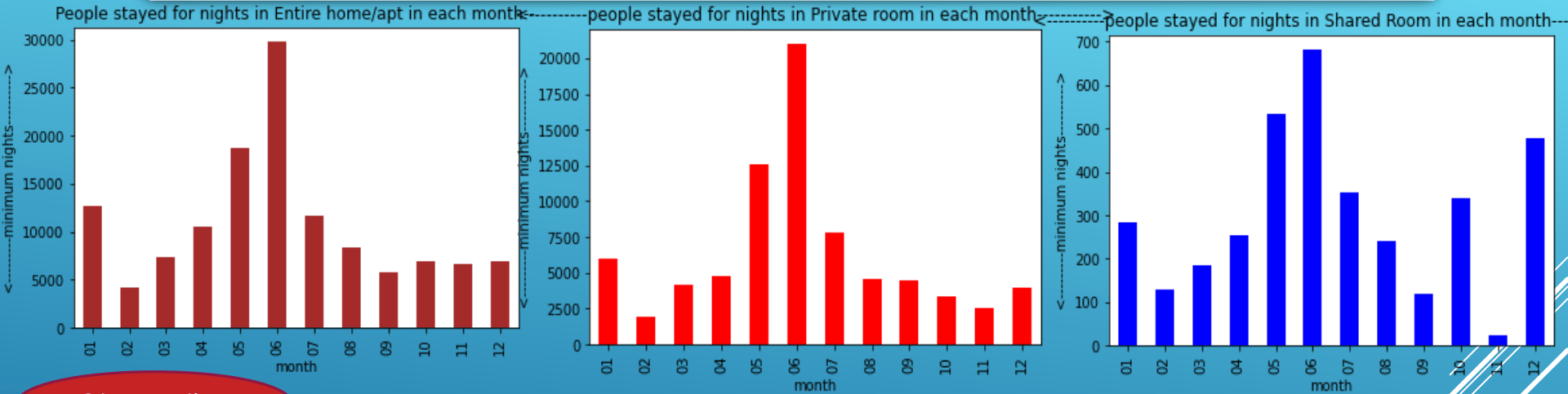
Entire home/apt: The prices of Entire home/apt started increasing from 2011 and stopped at 2013, after that the price became stable between 140-160

Private rooms: At beginning the price of private rooms increased & was highest in 2013, after that it started downfall & became stable at 60-80

Shared rooms: Shared rooms market started from 2014 & it started receiving downfall & in 2019 it is between 40-60



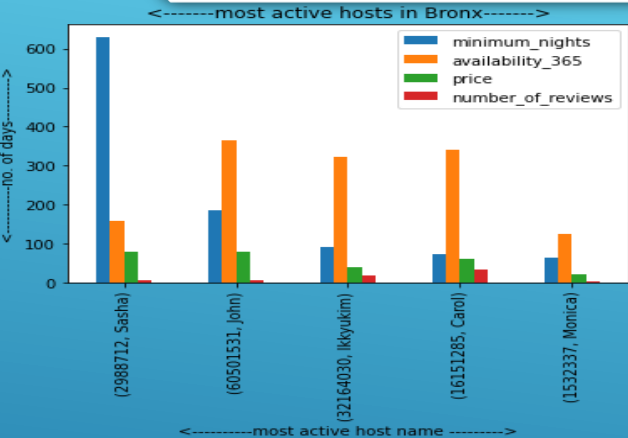
Which room types people prefer around the different months



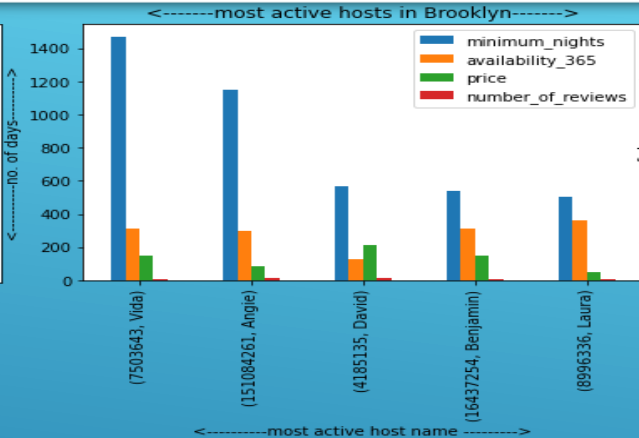
Observation

- **Entire home/apt**: May, June & January brings the most number of customers for Entire home/apt
- **Private rooms**: May, June & July brings the most number of customers for Private rooms
- **Shared rooms**: May, June & December brings the most number of customers for Shared rooms

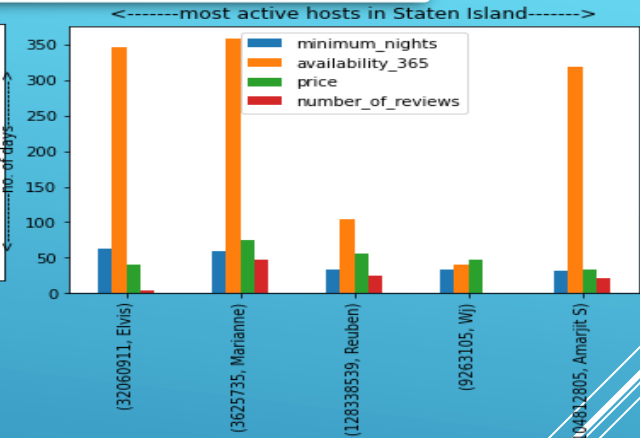
Most no of active host in different neighbourhood group



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



In Brooklyn these are top 5 active hosts, Whose room availability was higher during each year & the people have stayed for most no. of nights in their apartment



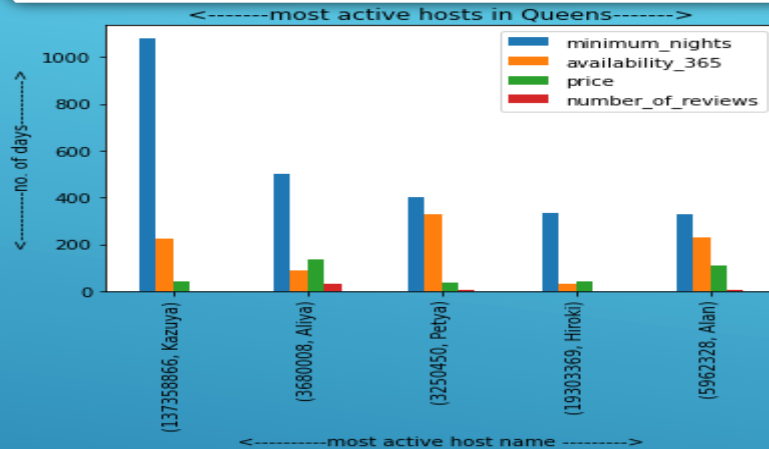
In Staten Island no of night spend by people is less than the other neighbourhood groups because the no of apartment is less as compared to others

	minimum_nights	availability_365	price	number_of_reviews
host_id host_name				
2988712 Sasha	630	159.285714	77.857143	6.428571
60501531 John	186	365.000000	79.000000	7.000000
32164030 Ikkyukim	90	321.000000	40.000000	18.333333
16151285 Carol	73	339.250000	59.500000	31.750000
1532337 Monica	64	124.250000	21.250000	3.000000

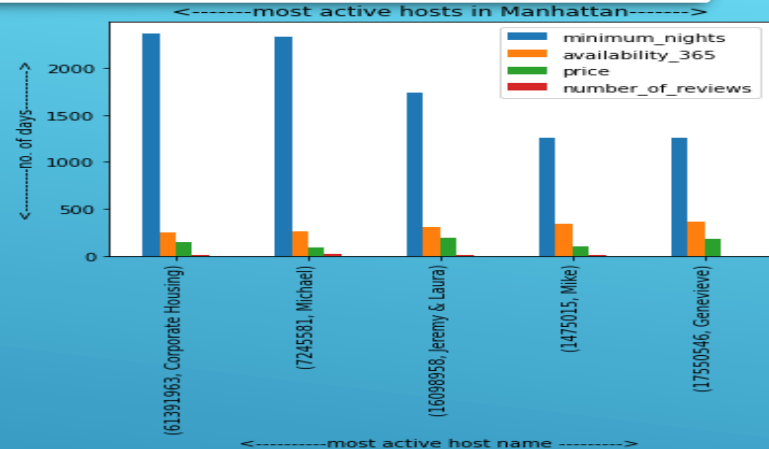
	minimum_nights	availability_365	price	number_of_reviews
host_id host_name				
7503643 Vida	1470	308.942066	149.408163	4.938776
151084261 Angie	1149	295.666667	83.666667	14.000000
4185135 David	565	130.290312	212.000000	16.500000
16437254 Benjamin	540	308.611111	149.555556	4.888889
8996336 Laura	500	365.000000	50.000000	10.000000

	minimum_nights	availability_365	price	number_of_reviews
host_id host_name				
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

Most no of active host in different neighbourhood group



In Queens these are the top 5 hosts in graph, but in this we can see the big gap between no 1 & other 4 hosts in minimum nights number



In the Manhattan graph we can see the top 5 active hosts, the minimum nights in this is higher than other & the reason can be the availability of their apartment are higher than other hosts

host_id	host_name	minimum_nights	availability_365	price	number_of_reviews
137358866	Kazuya	1080	225.786736	42.888889	1.500000
3680008	Aliya	500	90.000000	134.000000	30.000000
3250450	Petya	403	327.500000	38.571429	5.714286
19303369	Hiroki	334	30.870966	40.636364	1.818182
5962328	Alan	330	230.090909	110.727273	3.454545

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

Conclusion

- Manhattan gives lead to every neighbourhood in almost every field in terms of number of apartments, highest number of visitors. It has one of the lowest availability days. In Manhattan people mostly prefer Entire home/apt, so its visitors mostly prefer Entire home/apt.
- Manhattan's Private rooms is almost equal to Queens & Staten Island Entire home/apt & greater than price of Bronx Entire home/apt. Even its Shared rooms is higher priced than the Private rooms of all the neighbours.
- Brooklyn has 2nd most apartments in the neighbourhood & 2nd most costly after Manhattan. It has 2nd highest share in Entire home/apt. Brooklyn gets highest visitors after Manhattan. It had highest availability days in 2011.
- Queens has the 3rd most apartments in the neighbourhood, but has price almost equal to Bronx & Staten Island apartments although it has huge gap in the number of apartments with them. It had highest availability days in 2012. The number of visitors is decent according to its price. It has 2nd highest share in Private rooms & 2nd lowest share in Entire home apartment.
- Bronx is the most cheapest option in all the neighbourhood group but is 2nd last in visitor numbers & 2nd last in the number of apartments. It has highest share in its Private rooms, Shared rooms & least share of Entire home/apt. It was introduced by Airbnb from 2014 with highest availability of room & apartment.

Conclusion

- Staten Island has less number of apartments as compared to Bronx still charges more price than Bronx as a result gets less number of visitors than Bronx, Even though having highest availability days of rooms in the neighbourhood.
- Staten Island has almost same price of Private & Shared rooms. It has least share for it's Shared rooms & 2nd highest in Entire home/apt in the neighbourhood. It also came in market from 2014.
- Most of the people prefer Entire home/apt instead of Private & Shared rooms which is a cheaper option, so we can conclude that people prefer on the comfort & privacy instead on compromising them. So there is a huge gap between Private & shared rooms customers with Entire home/apt customers.
- In 2011 the availability was least for Private rooms & Entire home/apt. After 2013 there was a downfall of availability in all three options as the shared rooms was introduced by Airbnb in 2014. In 2019, shared rooms has highest availability & Entire home/apt has least availability.
- There was a downfall in the reviews of Airbnb from 2011-2015. After 2015 it started increasing with a high rate & has the maximum number of review till now in 2019.
- In 2011, the price of Entire home/apt was least & after 2013 it became stable at 140-160. Private rooms had a price hike in beginning but started downfall after 2013 & now it is stable at 60-80. The Shared rooms was highest in its 1st year 2014 but it came down due to being overpriced & stable at 40-60.