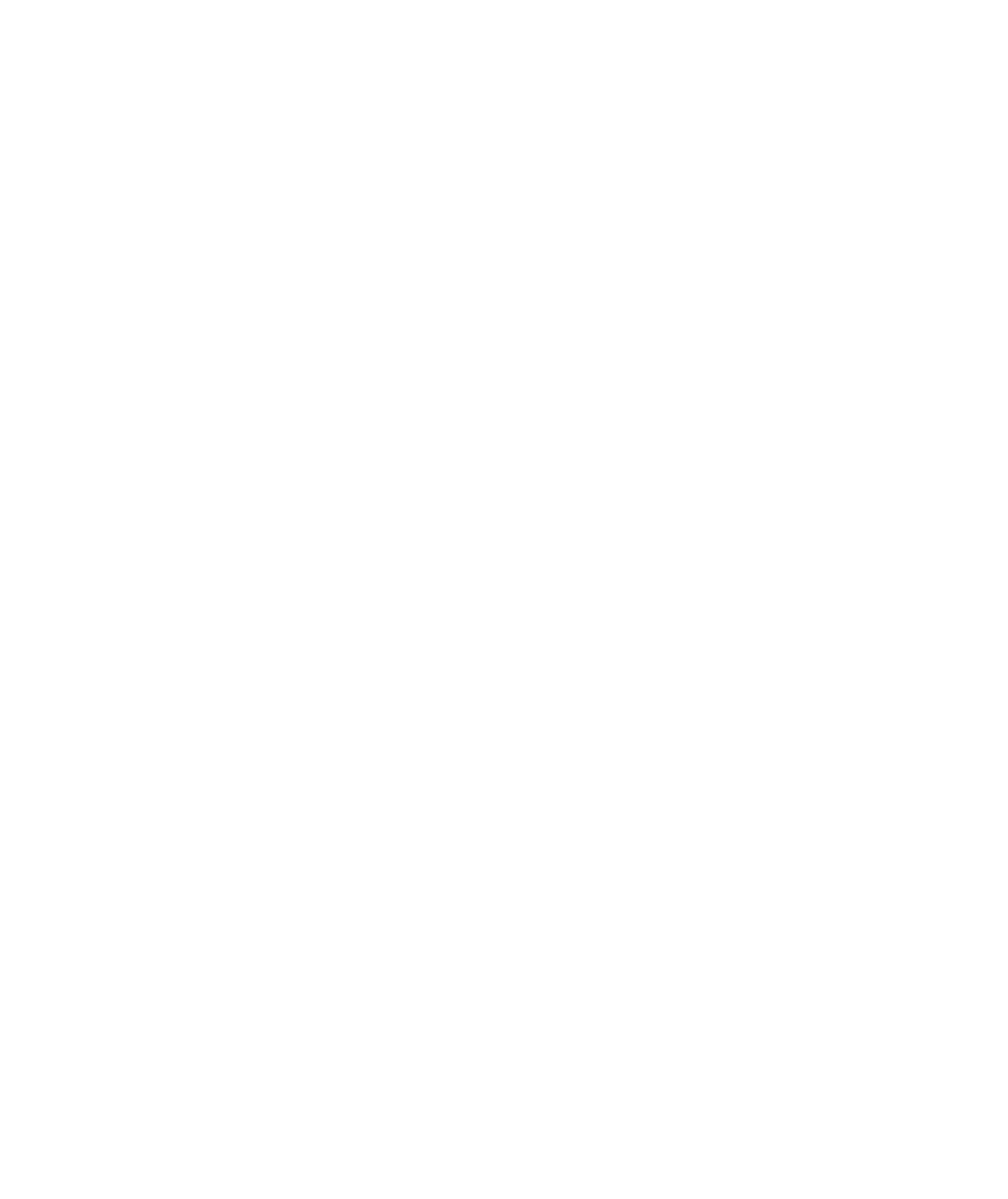
**Detailed Methodology Document**



**A**

**I**

**R**

**B**

**N**

**B**

**NYC**

**Submitted By: Ashish Arora**

**Suraj Dhamdhere**

**Abhishek Ranjan**

**Date: 11-April-2023**

**Index**

Content

1. Research Problem

2. Business Understanding

3. About Data

4. Whom are we presenting

5. Data Assumptions

6. Tools Used

7. Data Methodology

8. Export Final Data Frame in Excel File.

9. Analysis Methodology

10. Important Findings and conclusion.

**Research Problem:**

* For the past few months, Airbnb has seen a major decline in revenue due to lockdown imposed during the pandemic.
* Now that the restrictions have started lifting and people have started to travel more. Hence, Airbnb wants to make sure that it is fully prepared for this change.

**Business Understanding:**

* Airbnb is an American company based in San Francisco, California. It operates an online marketplace for lodging, primarily homestays for vacation rentals, and tourism activities.
* The platform is accessible via website and mobile app.
* Afterall, being an online marketplace for hosting personal home stays and private apartments in majority, the company had two types of customers, one who hosts their place and another who books the place.
* Airbnb earns commission from both ends and hence have to make sure both of its customers are able to generate value from their business.

**About Data:**

* Decline in the revenue could be for two major reasons, either the sites hosted on the platform are not able to provide better user experience or there could be a competitor in the market capturing the market share.
* Data which is provided to us use the information of the hosted places on the platform to see where and what can be done to improve the end consumer experience.
* The data would majorly include the location and region of the hosted places, in our case we are targeting Borough (New York City) — the Bronx, Brooklyn, Manhattan, Queens and Staten Island, followed by their hosts details, prices of the hosted sites and reviews received by the end consumer.

**Whom are we presenting?**

**Data Analysis Managers:** These people manage the data analysts directly for processes and their technical expertise is basic.

**Lead Data Analyst:** The lead data analyst looks after the entire team of data and business analysts and is technically sound.

**Head of Acquisitions and Operations, NYC**: This head looks after all the property and hosts acquisitions and operations. Acquisition of the best properties, price negotiation, and negotiating the services the properties offer falls under the purview of this role.

**Head of User Experience, NYC:** The head of user experience looks after the customer preferences and also handles the properties listed on the website and the Airbnb app. The head of user experience tries to optimize the order of property listing in certain neighbourhoods and cities in order to get every property the optimal amount of traction.

**Data Assumptions:**

1. Assumed that the data prior to the Covid-19 period was achieving the desired goals.
2. Currently, Airbnb is particularly looking for New York City — the Bronx, Brooklyn, Manhattan, Queens and Staten Island, followed by their hosts details, prices of the hosted sites and reviews received by the end consumer.
3. There are missing values in Reviews related columns like last\_review, number\_of\_reviews, and ratings\_per\_month. It is assumed that data for some listed properties is not available and there is no data discrepancy.
4. Availability\_365 tells that post pandemic, for how many days listed properties are active for hosting.
5. We only considered minimum nights and price for a listed property as major source of Revenue for properties.

**Tools Used:**

* Python, Tableau, Microsoft Power Point and Microsoft Word Document.

**Data Methodology:**

1. Data Understanding and Data Preparation is done in Python.
   1. Univariate Analysis
      1. Missing Value Treatment
      2. Outliers Treatment
      3. Feature Engineering

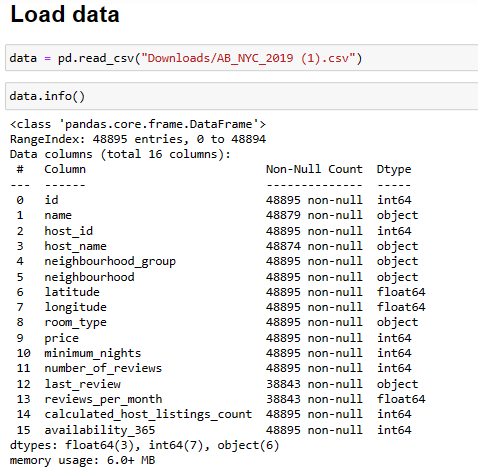
Let’s explore each step in detail.

**Data Understanding and Data Preparation:**

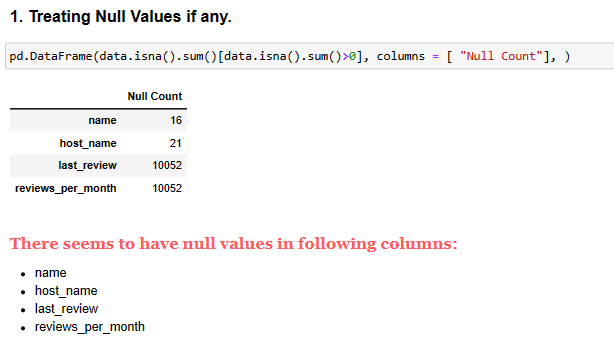
1. **Imported relevant libraries available in Python.**



1. **We started with Understanding the Data.**



1. **Missing Values Treatment**



For name and host name we filled the missing values with name not known.



For last\_review, we filled the null values with DUMMY DATE.

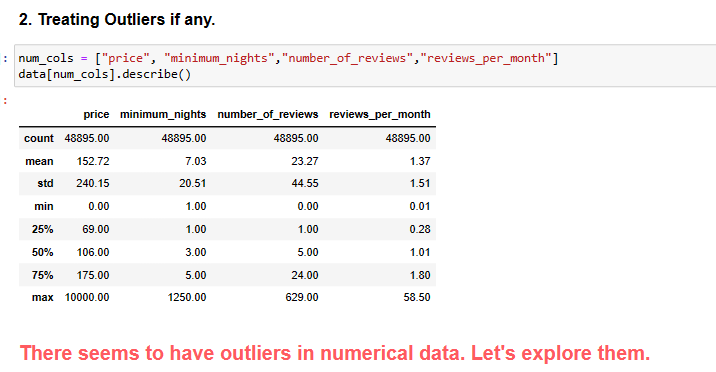


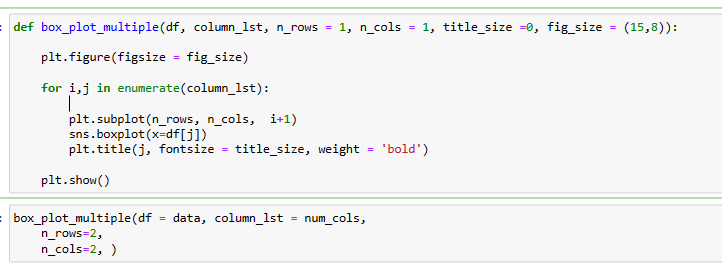
For reviews per month, we filled missing values with mean reviews in their respective neighbourhoods.

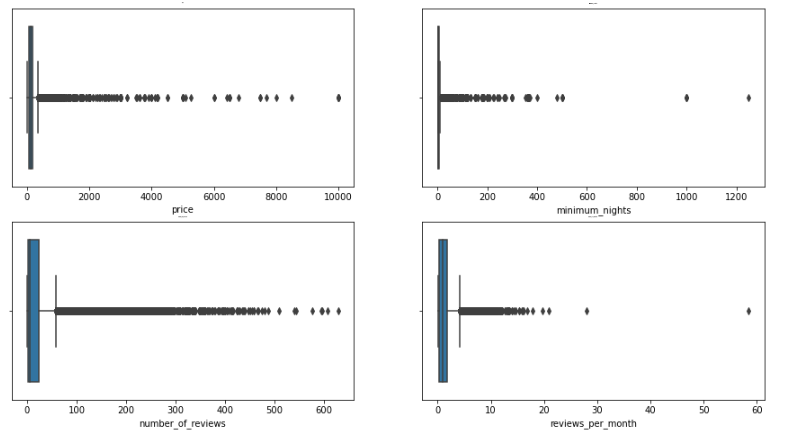


1. **Outlier Treatment:**

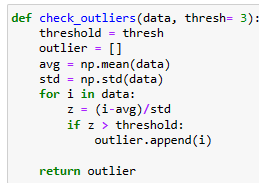
**Post analysing and treating the missing values accordingly we treated the spread in the data frame i.e. outliers. Below is the code we used to identify the spread of the outliers.**



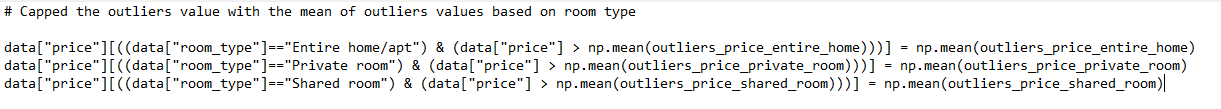




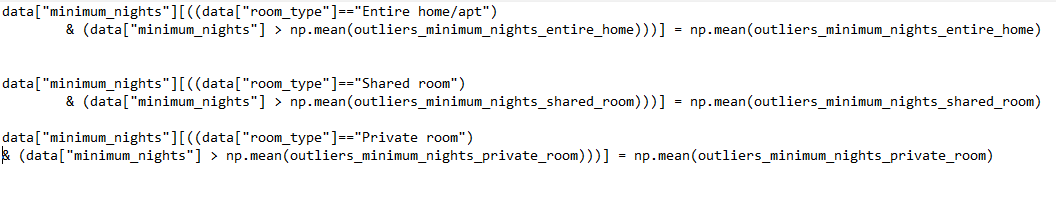
Sure there were outliers in numerical columns. In order to find the outliers value, we used z-score method with a threshold value of 3 and 4.



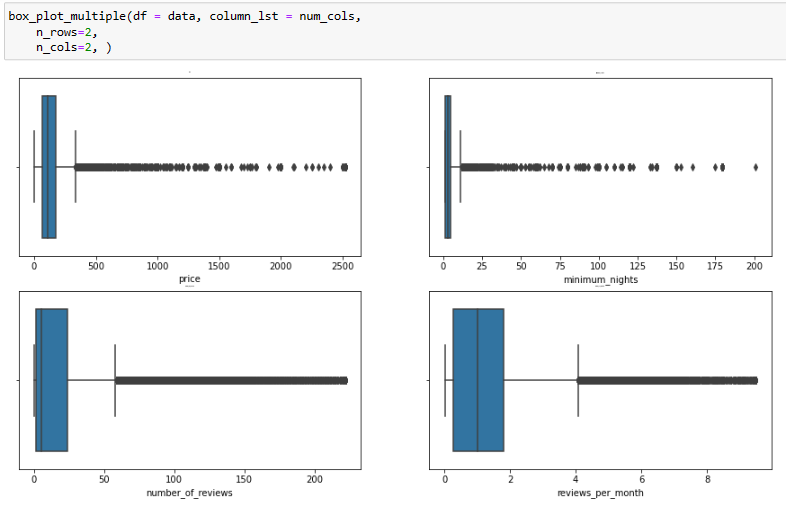
For each numerical column, we did the Univariate analysis and capped the outliers accordingly.



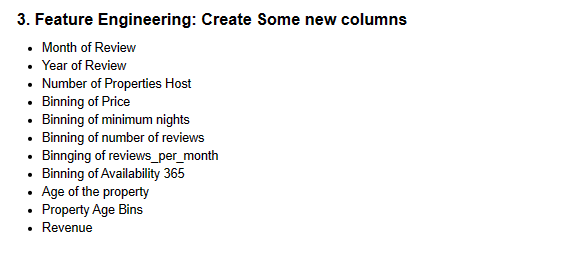
Used the same concept for capping minimum nights.



**After Treating Outliers**



3. Feature Engineering: Created Some new columns from numerical columns.



Create Month and year of last Review to find the year in which year reviews and given more and does the month follows any seasonality trend with reviews.



Binned price for finding the price range of properties and calculated host listings to find how many host hosts how many properties.



We did the same for minimum nights, number of reviews and ratings per month.

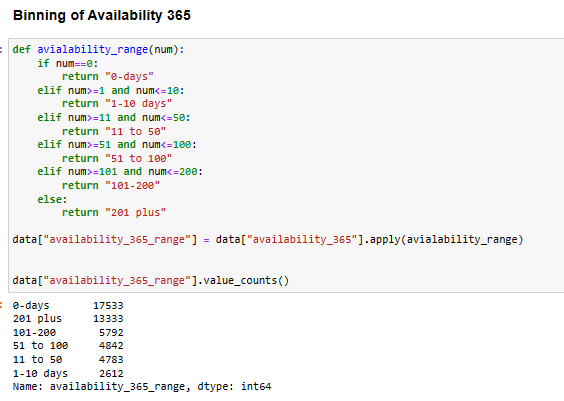
We also found the age of property by taking the help of number of reviews and reviews per month. This will help us know that a given property is being listed for how many years with the Airbnb and head of acquisition can plan out acquisitioning some of those properties.



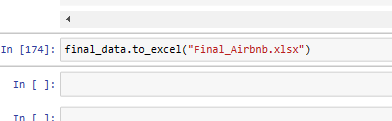
Created Revenue and Revenue Range column.



Also binned the availability 365 columns in following labels:



**8. Export final Data Frame to Excel file**

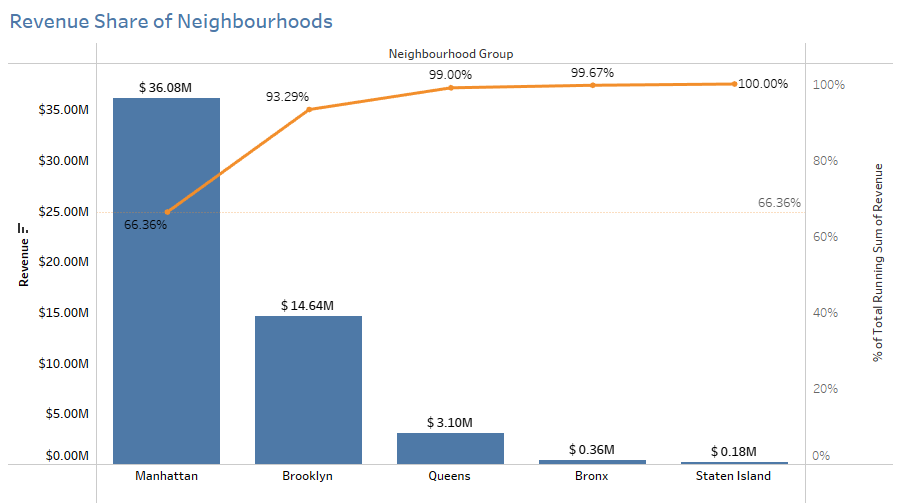
****

**9. Analysis Methodology**

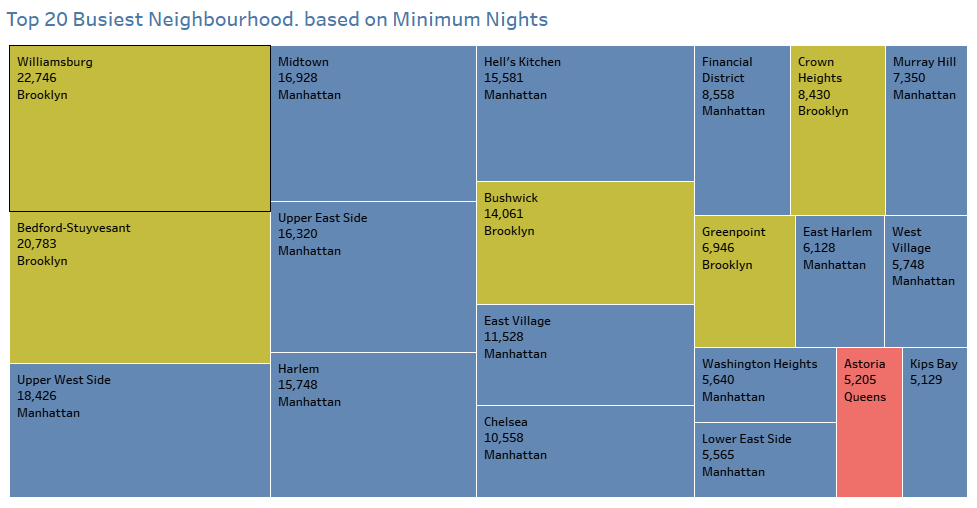
Our Analysis is based on the assumptions that leaders at Airbnb want to understand some important insights based on various attributes so as to increase the revenue such as -

1. Which type of hosts to acquire more and where?
2. The categorization of customers based on their preferences.
3. What are the neighborhoods they need to target?
4. What are the pricing ranges preferred by customers?
5. What are the various kinds of properties existed w.r.t. customer preferences?
6. What can be the possible adjustments to make in the existing properties to make them more customer-oriented?
7. What are the most popular localities and properties in New York currently?
8. How to get unpopular properties more traction?

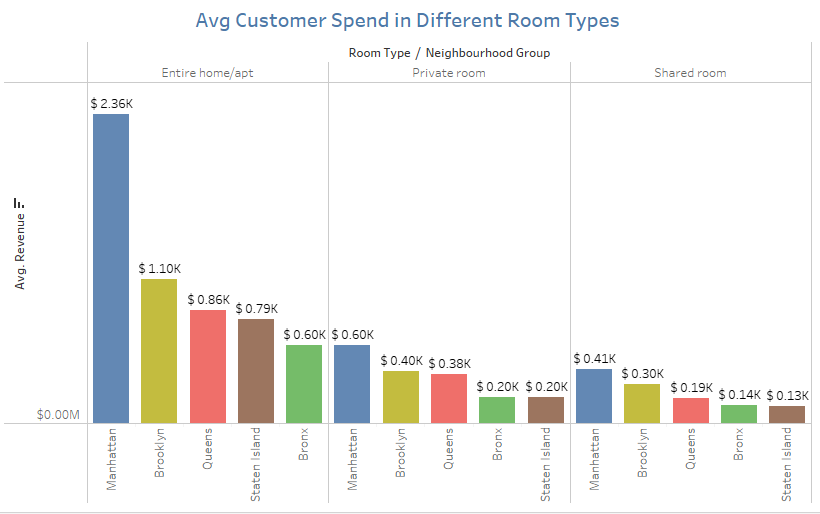
These are some of the analysis which I have conducted on Tableau.



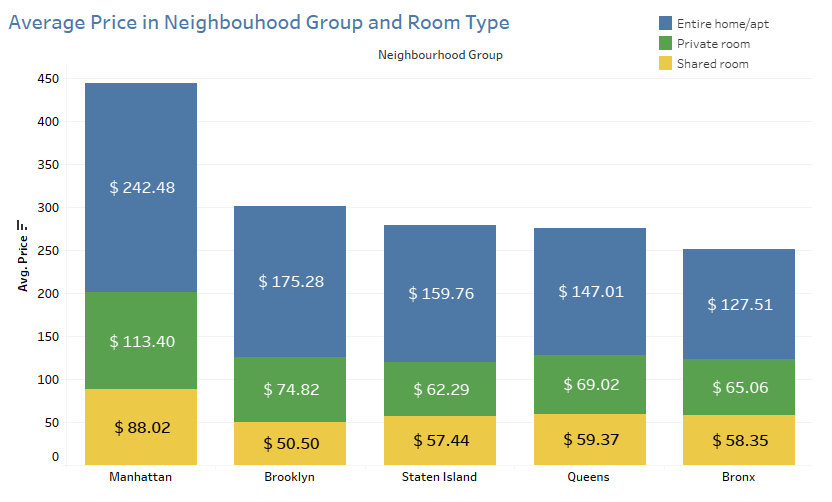
Manhattan is most prominent neighbourhood group in New York. • Around 66% of revenue is coming from Manhattan itself followed by Brooklyn which has 23% share in overall revenue.



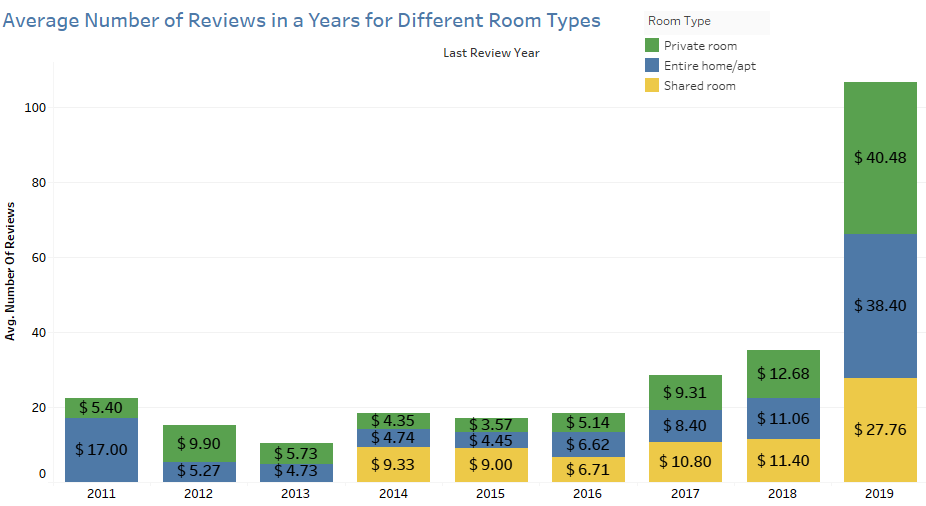
These 20 neighbourhoods hosted most number of nights. • 5 of them are from Brooklyn, 14 are from Manhattan and 1 from Queens.



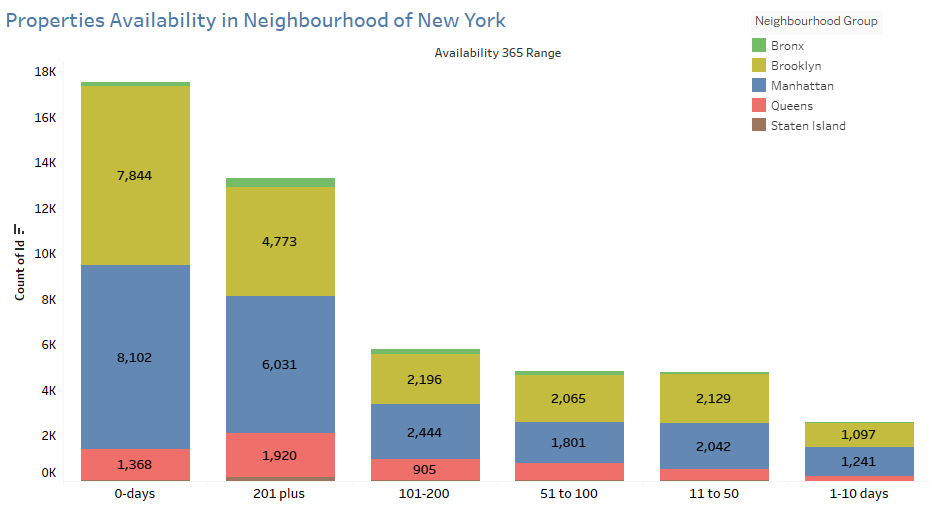
Average customer spends in different room types. Manhattan has the highest customer spend of 2.36 thousand dollars



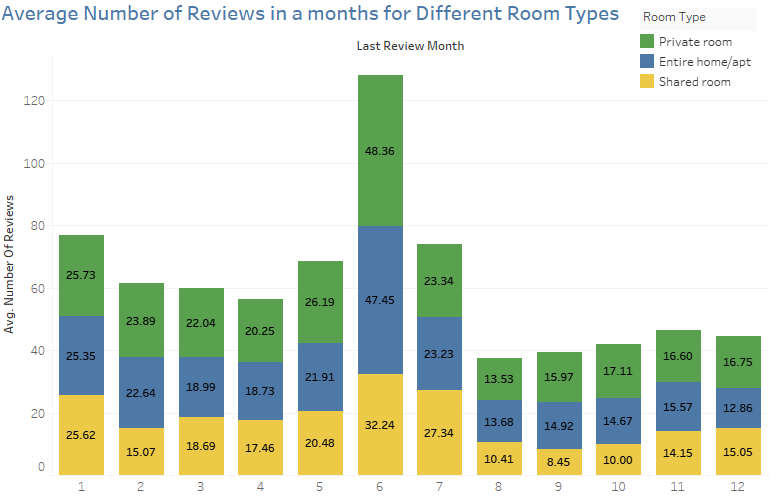
An average price range preferred by a customer is: o Entire Home: 150$ to 120$ o Private Room: 70$ to 60$ o Shared Room: 50$ to 60$ • whereas the most traction generating neighbourhoods i.e. Manhattan and Brooklyn offers a higher price range which might demotivate the customer to book



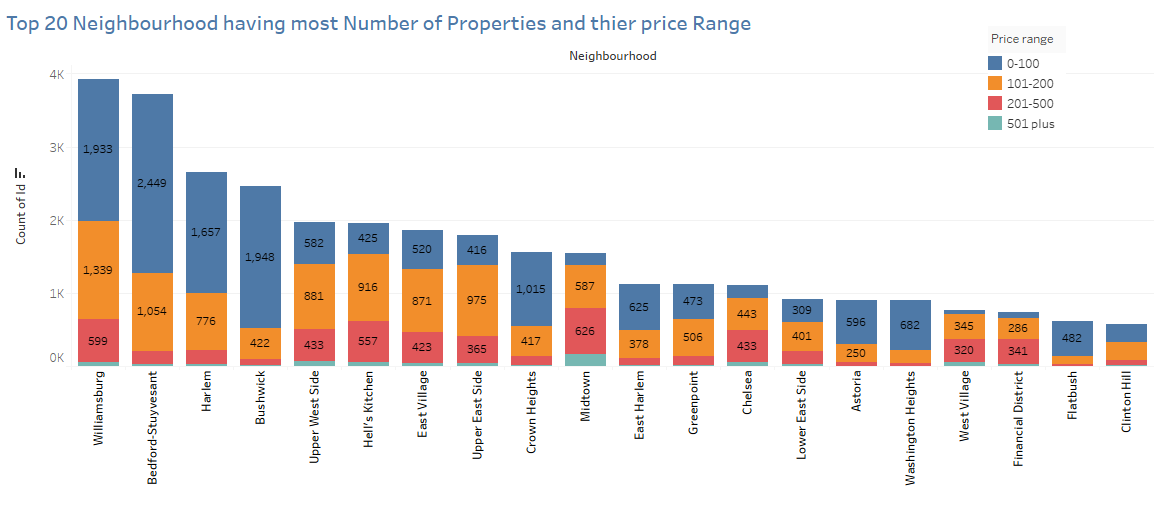
From 2014, reviews starting gaining up.



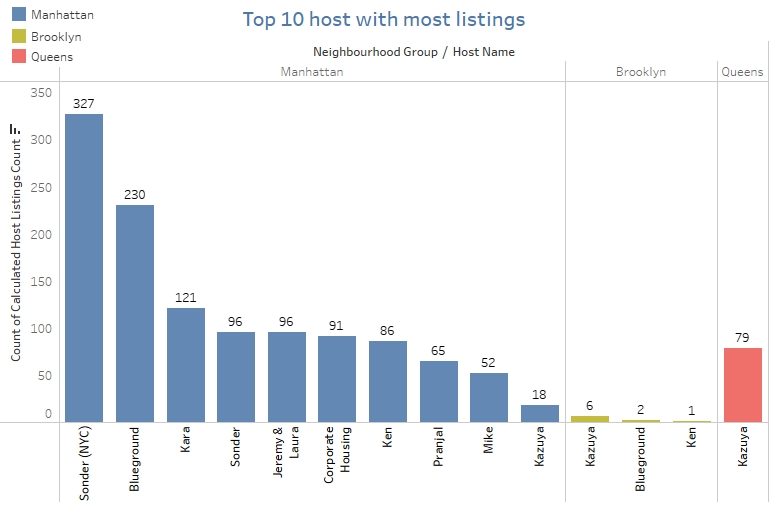
Many properties in Manhattan and Brooklyn is not available now. We can consult these hosts again and can provide monetary assistance in renovation as acquiring new hosts may cost more.



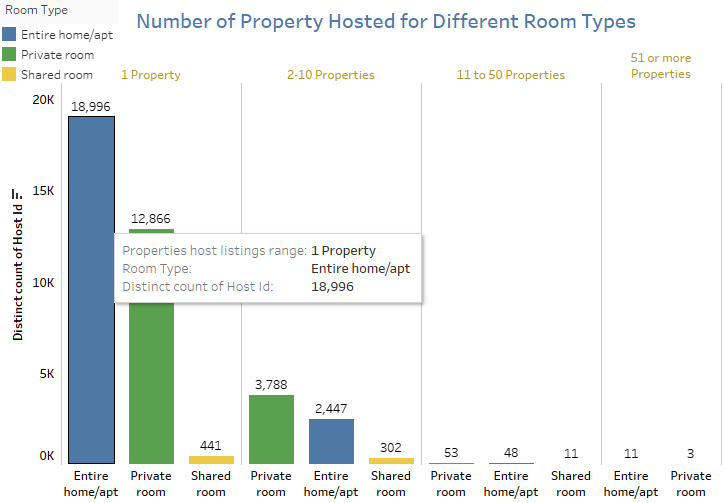
On Average, all months share almost same average, except the month of June. • At the start of year, average numbers of reviews are more and as the year precedes these average reviews are going down. • Average Number of reviews for: • Private Room – 16-25 • Entire Home – 12-25 • Shared Room – 8-25



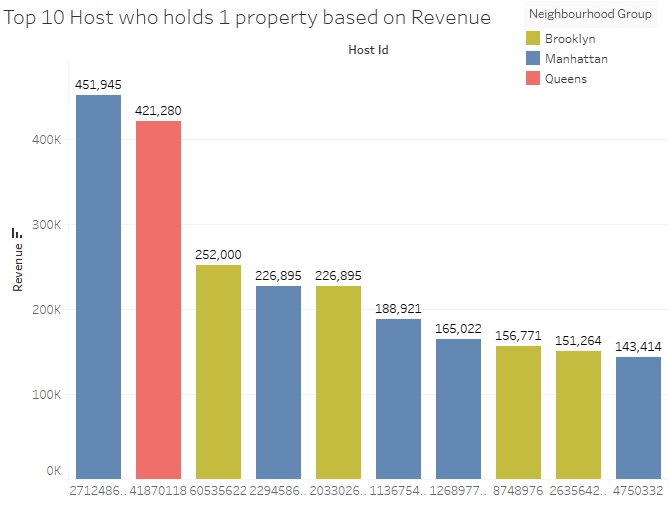
Neighbourhood which has most number of properties prefers a price range of 0-100 and then followed by 101-200. • Williamsburg, Belford-Stuyesant and Harlem has the most number of properties in their neighbourhood



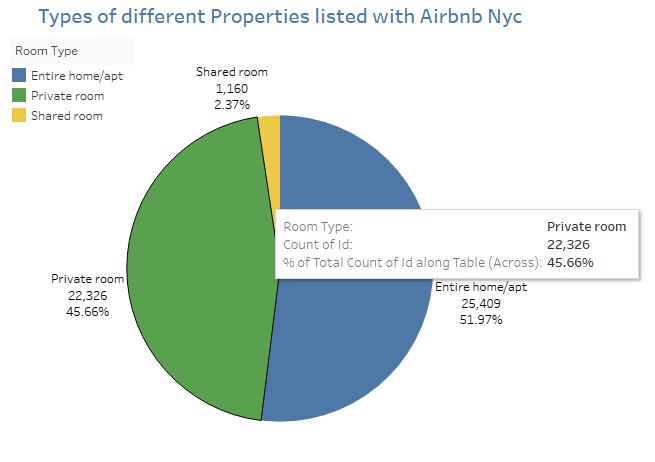
Sonder hosts around 327 properties with us, followed by Blue ground which hosts 230 properties and so on. • Kazuya, from Queens, also hosts 79 properties. We should interact with such hosts and introduce them in marketing ads which may help in acquiring more hosts in neighbourhood.



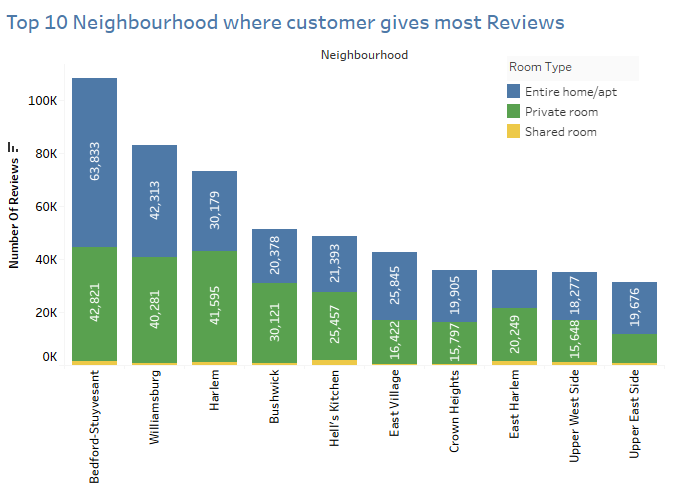
Host which host only 1 property mostly has Entire Home followed by private rooms. • Same trend is seen by the host who holds 2-10 properties



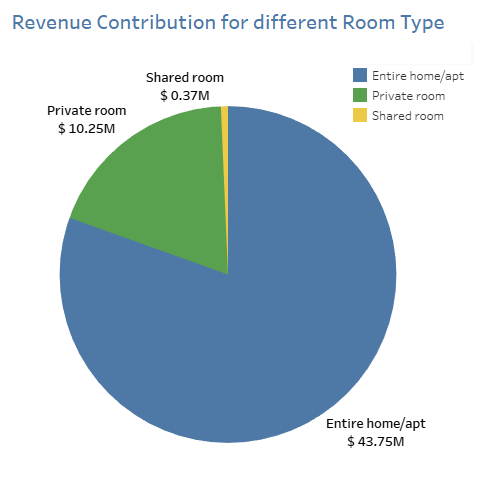
Most hosts have 1 property. • Considering that, 5 hosts are from Manhattan and 1 from Queens and 4 from Brooklyn.



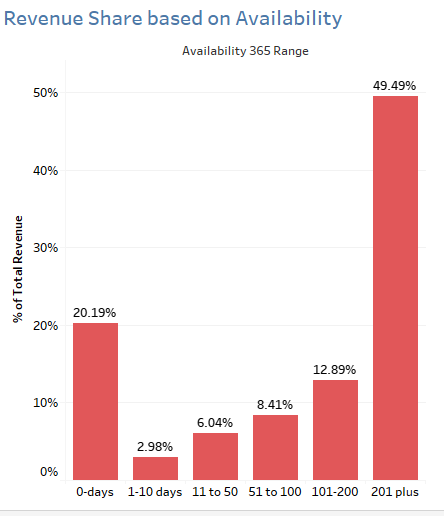
Out of all listed properties, Entire home properties count for 25,409. • Followed by that, 47% of Listed properties host Private rooms • only 2.37% Shared rooms are there.



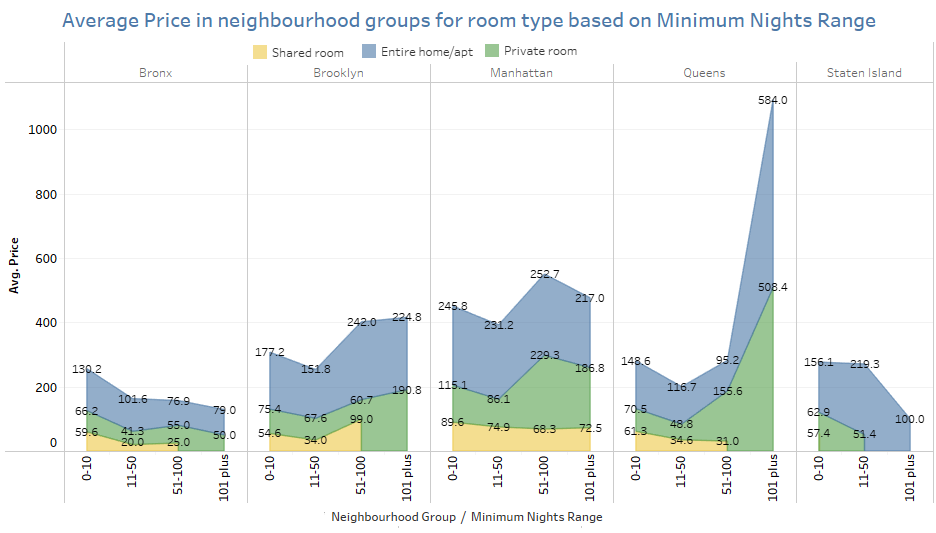
Customers in these 10 neighbourhoods shared most number of reviews. • This could be either because customers are getting problem in these neighbourhood or they are extra active in these neighbourhoods. • Thus we should find a conclusive evidence for them



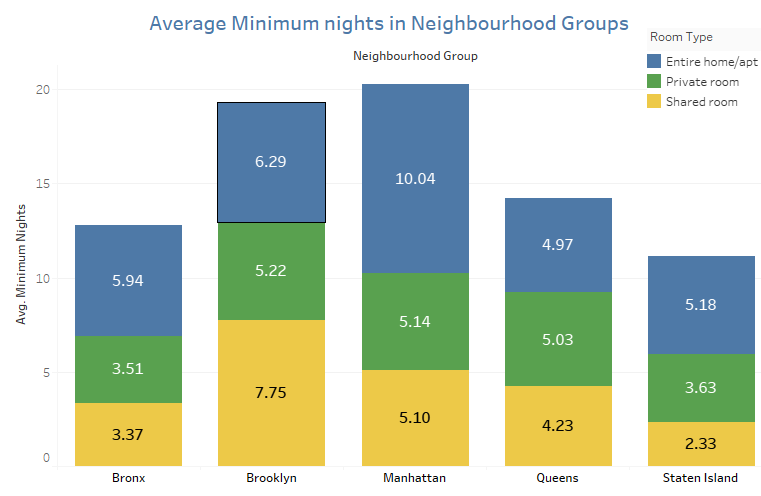
Entire home as so far earned us functional revenue of 44 million dollars followed by private rooms with 10 million revenue. • Post Covid-19 it is also evident that people would rarely go for shared rooms. • Better if we keep our focus on acquiring host with Entire home and private rooms.

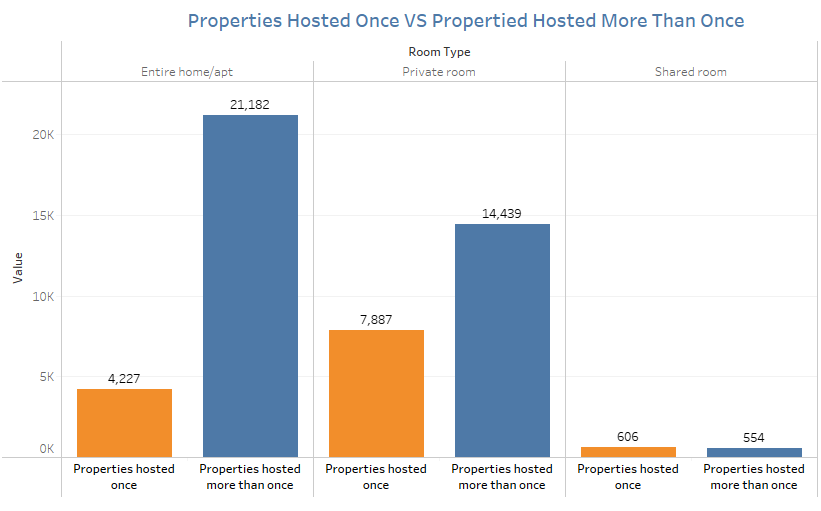


Properties which are available for 201 days or more share almost 50% of Revenue. • Properties which are available for 0 days post Covid-19 contribute 20% to overall revenue. • These properties should be contacted again. • We should make a marketing policy such that it push host to keep its property available for most number of days.



As the night increases, the average price of listed properties is going down. Except if minimum nights are in range 51-100. • Average price is very high at Queens if nights are hosted for more than 100 days.





Entire homes are hosted twice or more are 5 times more than it is hosted once which shows they are more likely to get booked again. • Likelihood of Private room getting booked is twice of hosting once.



**10. Important Findings and Conclusion**

1. There seems to be no positive or any type of correlation between the numerical variables. 2. Manhattan is the only Neighborhood in the Borough that lies in offering the Highest Price range properties on the platform followed by others with a Medium Price range on average.
2. Prices offered above 120$ on average are considered to be a High Price, between 80$ to 120$, Medium Price range and less than 80$ to be considered Low Price range property.
3. Having a high price range, Entire home/apt types rooms are available for less than 100 days on average followed by Private rooms on an average of 105 days and Shared rooms around 155 days on average being the lowest in price.
4. Manhattan has the highest number of places listed around more than 10 by a single host with an average price of 230 $ followed by Brooklyn with an average price of 108$. On the other hand, all the hosts have less than 2 properties listed in either of the Borough on an average price range between 80 $ to 170 $.
5. Brooklyn has received the highest number of reviews based on the availability to stay open for more than 200 days in a year. This is followed by Staten Island and then the Bronx. On the other hand there are some sites in Staten Island which are not open for a single day at all and hence could be the reason they have received very low reviews from the end consumer. We need to check which these places are and what issues are they facing?
6. Majority of the customers prefer a price range of 120$ to 130$ on average for a stay. As most of them have provided a good number of reviews between this price ranges.
7. Michael, David, Alex, John and Daniel are the Top 5 hosts that seem to have received the highest number of reviews for their listed sites and have also sites listed with High price range.
8. Staten Island - Silver Lake, Staten Island - Richmondtown, Staten Island - Eltingville, Staten Island - Huguenot and Brooklyn - Manhattan Beach are the Top 5 locations with Low Price range that have received the highest number of reviews on average being the lowest in Price range.
9. On the contrary, Queens - Neposit, Manhattan - NoHo, Manhattan - Tribeca, Staten Island - Willowbrook and Manhattan - Flatiron District being highest in Price range have received low number of reviews.
10. The average number of reviews started increasing exponentially after 2015-2016. And majority of the customers provide higher number reviews either between the months of May till July or in the starting of the year which shows the higher booking window in a year.
11. There are 5766 properties that are open for more than 300 days in a year. Around 2286 of them are from Brooklyn followed by Manhattan of around 1947 properties. And on the other hand, the properties that stay open for less than 50 days a year belong to Queens or Staten Island.
12. We can confirm that the greatest parameter for any customer to prefer a property and provide a review is having a maximum or minimum night stay window booking and their probability of being open for more days in a year to some extent.