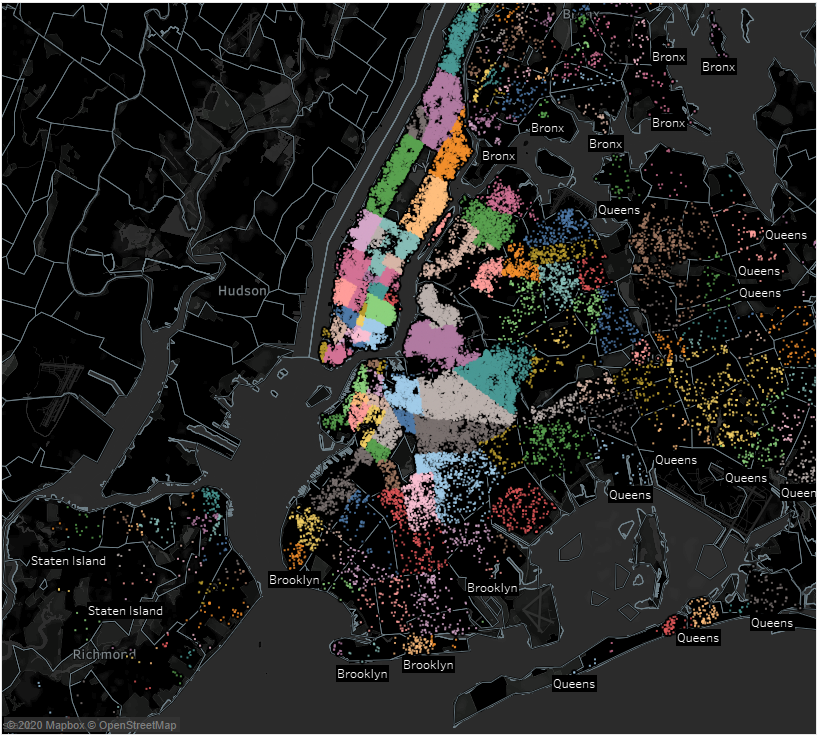
New York City Airbnb Exploratory Data Analysis

**Final Group Project**

June 22, 2020

**Introduction**

After our group has decided on our dataset, which is about the Airbnb listings in NewYork city, and we have identified our commercial enterprise question, this document intends to provide extra insights on the initial appearance of the facts. Particularly, it makes a specialty of our target variable, rate, and courting with different variables selected. This could be essential considering that it can help us to apprehend our information higher and provide instructions for besides modeling. At ultimate, we can extract the most critical insights based totally on our dashboard and our interpretations. The purpose of this project is to deliver data-driven insights for better decision making on improving their businesses in the post-epidemic era.



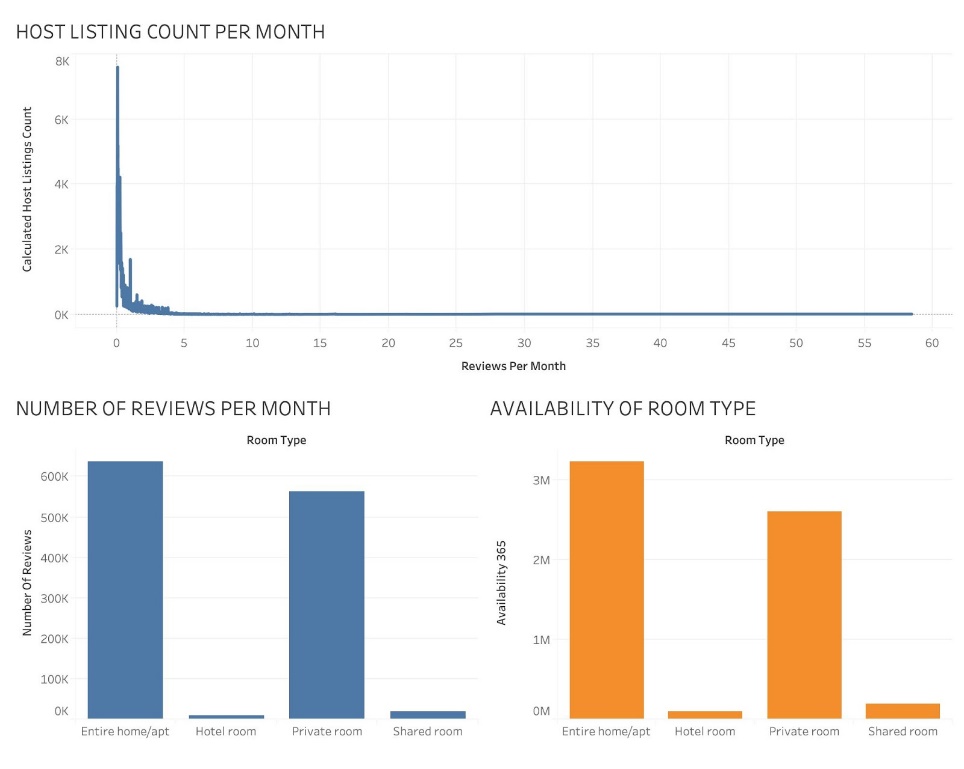
*(Figure 1. Map Views)*

**Exploratory Data Analysis**

The datasets we retrieved come from the Inside Airbnb website, which is an official database website, and Airbnb released their latest data on this platform for many years.

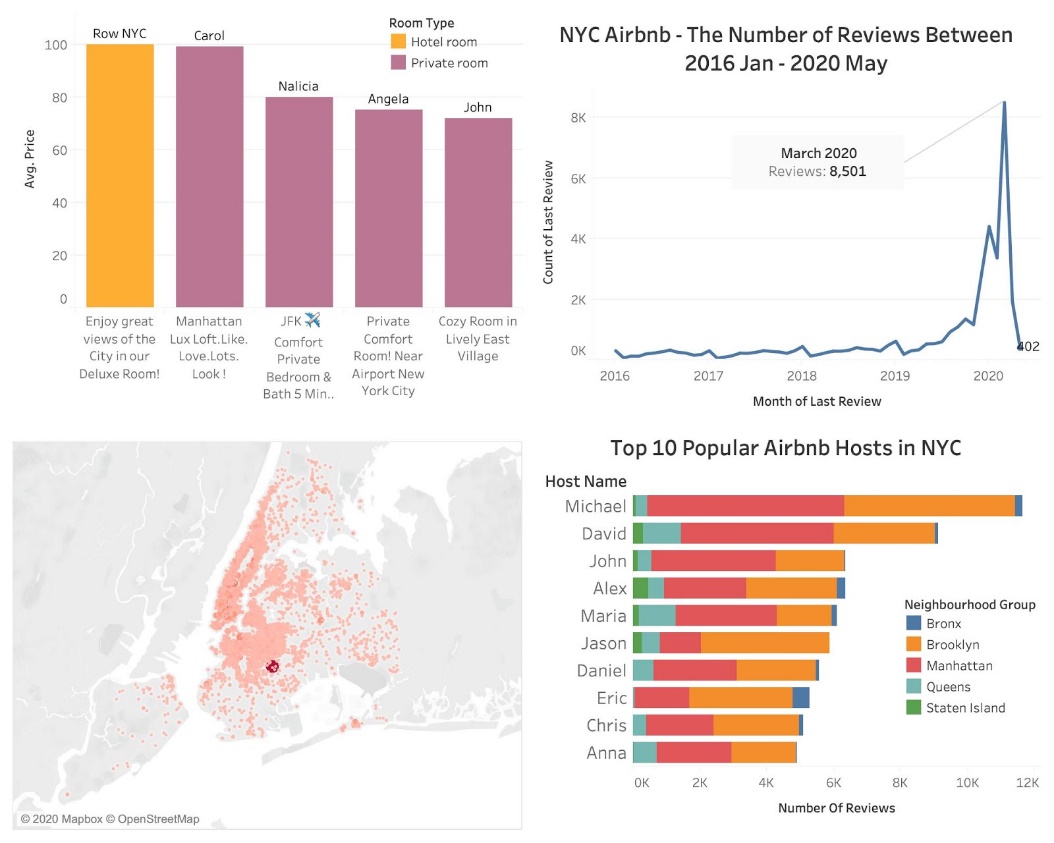
The first step we did is that we've extracted the variables we need to assist in answering. Before going into visualization, the records need to be smooth and ready. Thus we achieved a few statistics cleaning off dollar signs, commas, changing missing values with null, etc. Then, it's time for a few dashboard evaluations.

We created a histogram which suggests the distribution of charges in Airbnb. It is apparent for us to look that the general public expenses of Airbnb are about inside the variety, with the frequency. This is an affordable price range because the general public picks out Airbnb now not just due to the neighborhood customs, however, additionally due to its decreased rate than resorts.



*(Figure 2. Tableau Dashboard 1)*

The above graphs show the host listing count per month and the number of reviews per month for a particular room type and availability of room type in a year. The highest recorded is the entire home / apt has above 600k reviews per month, and more than 3M rooms availability. EDA evaluation that intends to offer more insights on the first look of the information. Particularly, it specializes in our target variable, rate, and courting with different variables chosen. This outcome could be very critical when you consider that it can assist us in understanding our records better and offer directions for similarly modeling.

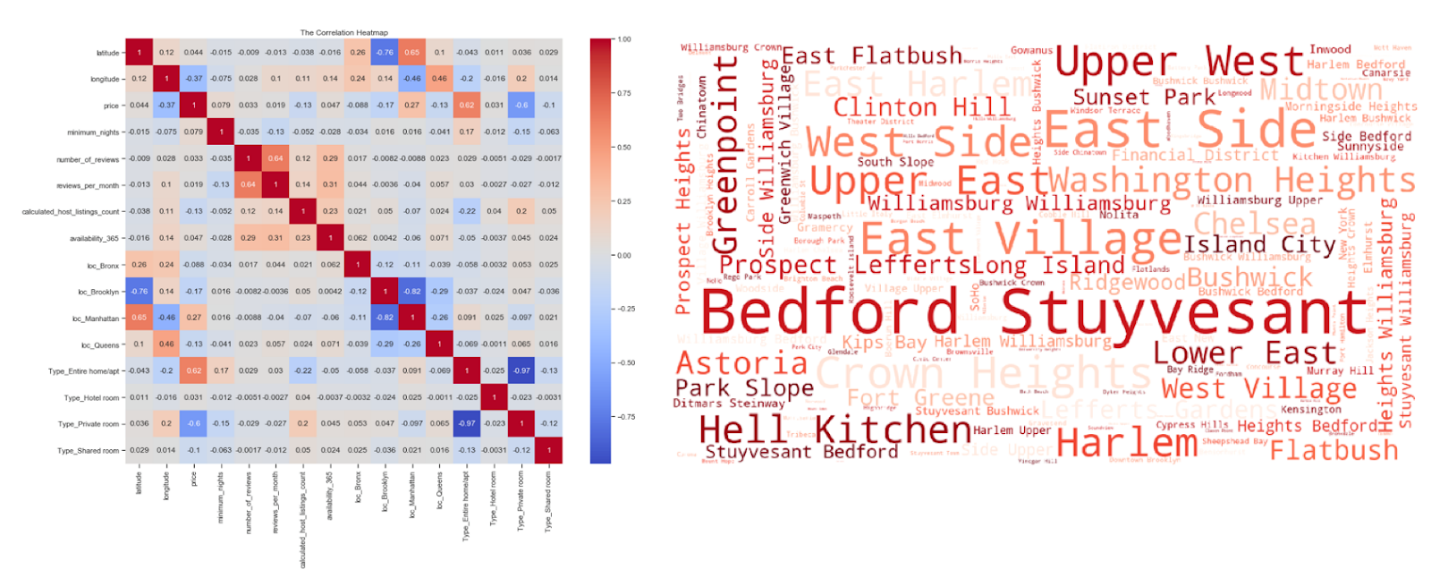


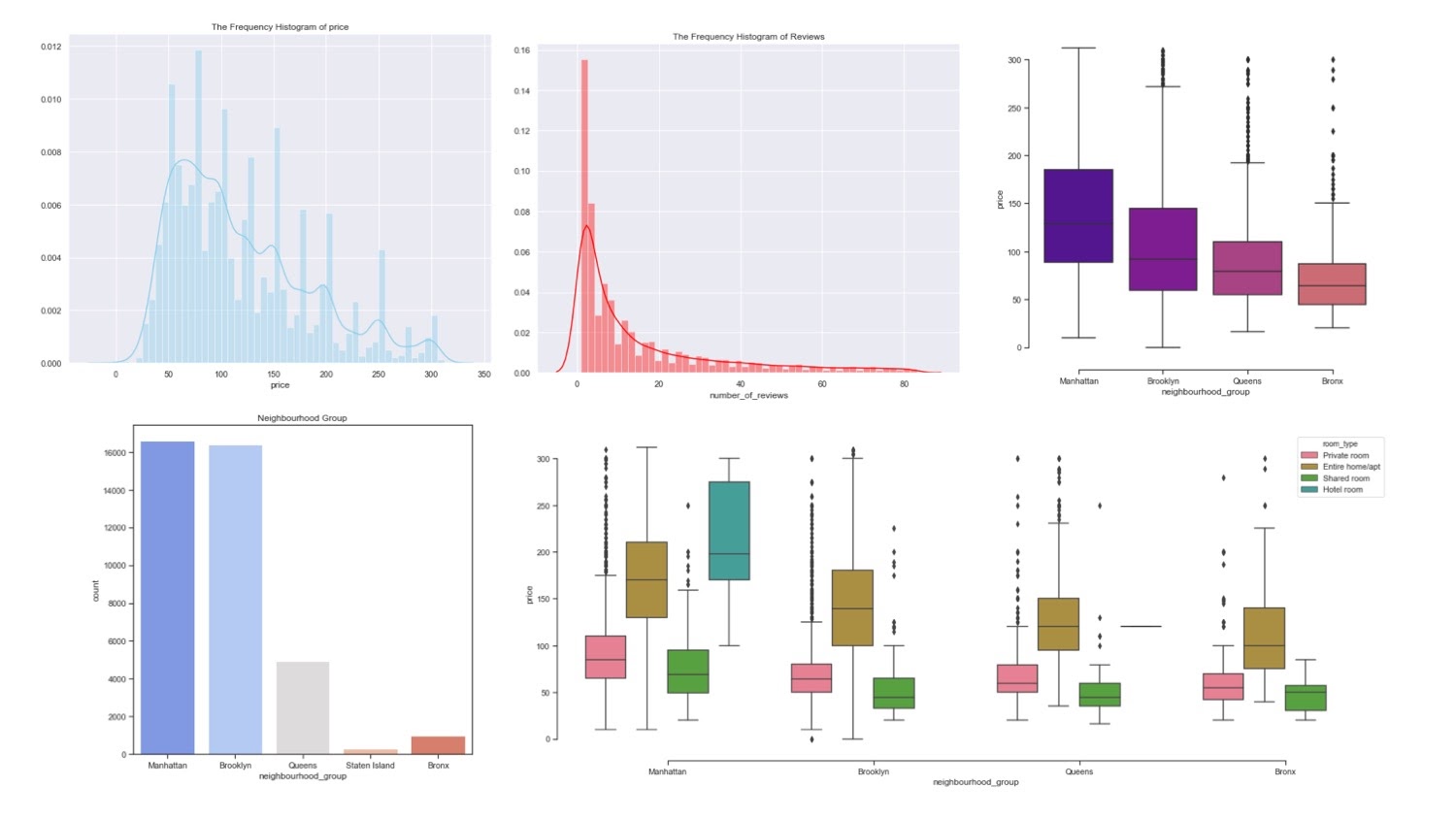
*(Figure 3. Tableau Dashboard 2)*

         As the dashboard shown above, the upper right time series plot illustrates that there is a vertiginous drop in the number of client reviews since March 2020 when the COVID-19 pandemic outbreak in the United States. New York City as one of the most popular and most populated cities in the country, has suffered a significant loss (8,501 reviews to 402). It has a huge impact on the consumer industries, especially the transportation and traveling industry. Airbnb, who has intended to run an IPO in March 2020, was forced to delay the original plan, and most of their hosts are selling their properties to cover the debts.

        The most popular Airbnb host in New York City is Michael (with approximately 12k reviews in total). And over 90% of his listings are located in Manhattan and Brooklyn. So as the rest of the top 9 hosts, there are a small number of reviews of their listings in Staten Island due to its location and inconvenience. Therefore, Airbnb should notify those hosts with a debt issue and business problem due to the COVID-19 pandemic to sell their properties in Bronx, or Queens to cover their loss at first and keep their popular properties in Manhattan if it is financially possible.

According to the correlation heatmap below, we can see there is a relatively strong positive correlation between Type\_Entire apt (0.62), loc\_Manhattan(0.27), and price. Based on the neighborhood word cloud, Upper East, East Village, Upper West, Hell Kitchen, etc. in Manhattan are more popular. Bedford - Stuyvesant, and Williamsburg in Brooklyn were the hot areas. Hosts will be able to make strategic plans accordingly as long as the pandemic outbreak fades.

*(Figure 4. Correlation Heatmap & Neighborhood Word Cloud )*

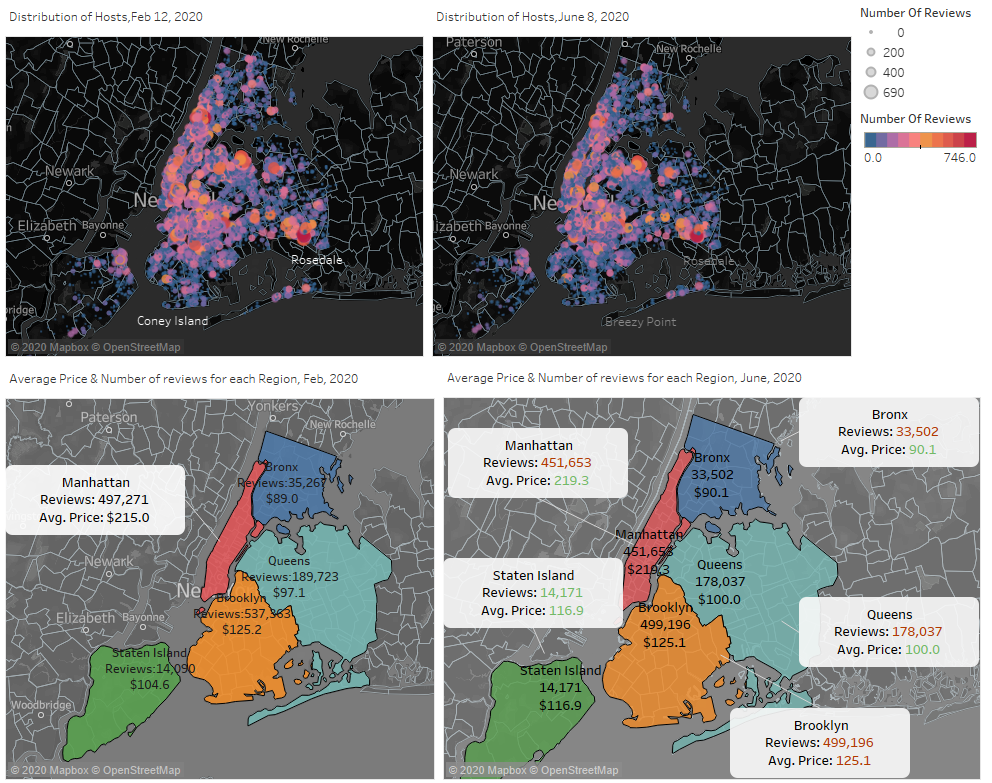


*(Figure 5. Python Data Visualization Dashboard )*

The upper left histogram above illustrates the distribution of the price for each listing in NYC. The box plot in the upper right corner shows listings in Manhattan have the highest average price at 150 dollars, and properties in the Bronx have the average price under 100 dollars. Hotel rooms have the highest price range (between 200 and 300) in Manhattan.

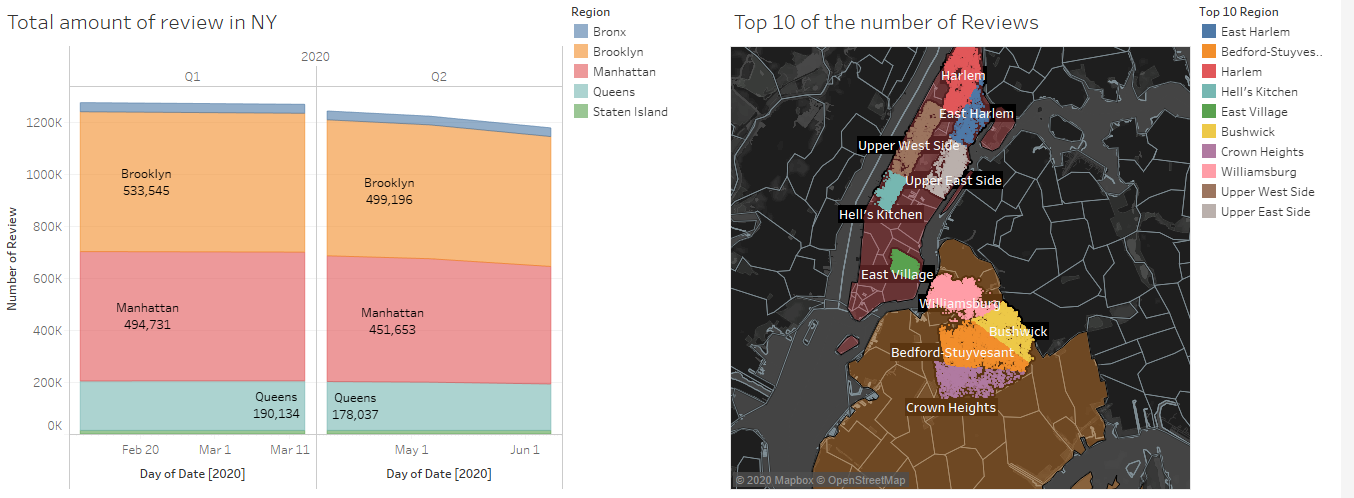
**Regional Analysis**

Since customers only can leave reviews for their hosts after they finished the period of rent,    the number of reviews is considered a critical measurement of counting the number of orders successfully completed by Airbnb. Comparing the number of reviews of each of the five New York City boroughs before and after the outbreak, we found that the overall amounts of the reviews tend to decrease while the average price of each borough slightly up but also beginning to decline.



*(Figure 6. Map Dashboard 3)*

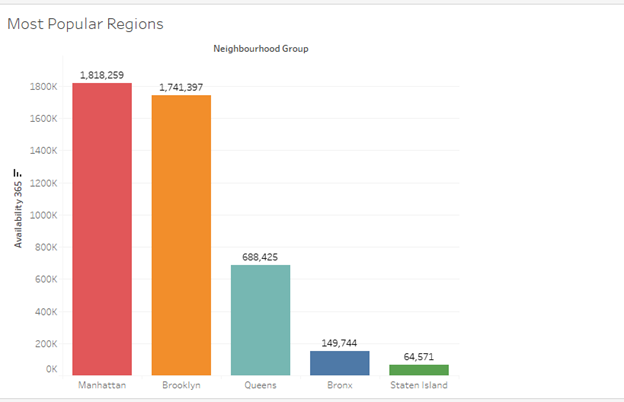
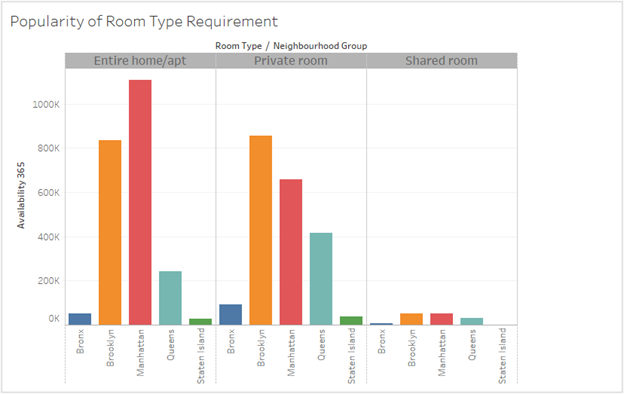
According to the first and second heatmap in the dashboard above, affected by the pandemic, the number of reviews in upper Manhattan and middle Brooklyn was significantly reduced. However, both the number of reviews and the average price in Staten Island abnormal growth, which indicated that people expected to live in areas with low population density to reduce their chances of getting infected. Therefore, because of the unique conditions during the epidemic, Airbnb should consider increasing the development of the Staten Island borough for temporary to meet customers' needs.



*(Figure 7. Reviews Graphs )*

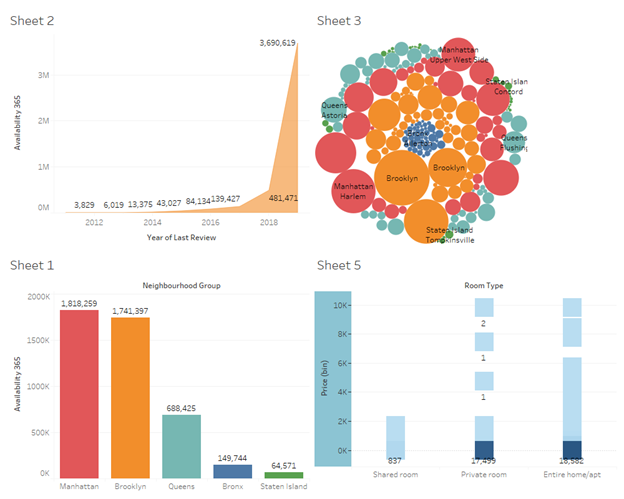
Moreover, based on the figures above, the difference in the number of reviews between Q1 and Q2 is significant that the share of Brooklyn and Manhattan starts to decrease smoothly, and the gap will continue to widen since the pandemic getting worse in the United States. Besides, by the time of cleaning raw data, many accounts labeled zero rent in the price column of the latest dataset, which is abnormal, and we deduced two possibilities for this situation. The first one might be that many hosts suspended their cooperation with Airbnb temporarily, thereby, Airbnb labeled those hosts with a zero rental price. Secondly, Airbnb may have been involved in falsifying data to cope with the consequences of a rapid rise in prices, which is not an expected result that the shareholders want to see.

**Room Type Analysis**



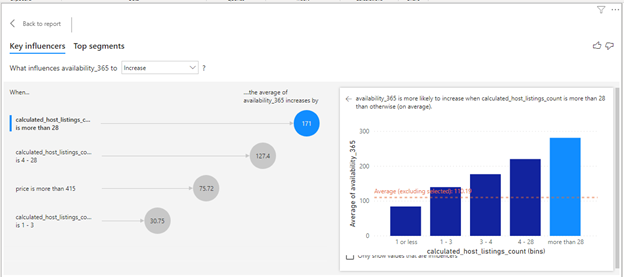
*(Figure 8. Room Type Bar Plot )*

Airbnb clients rate their stay on the premise of area, cleanliness, and a host of other parameters. The area scores got to be a firm pointer of the request of the neighborhood. Profoundly appraised communities will tend to have a superior network (tram stations), will tend to be closer to the city hotspots (Times Square, Real Estate, Divider Road). Manhattan gets the most noteworthy area scores for the downtown region (esp underneath Central Stop). In Staten Island, the ranges near to the State Stop have the most remarkable area scores. Brooklyn neighborhoods near Manhattan tend to have higher area ratings. From the above bar graphs, we can see that these are the popular regions in NYC with customer's requirement, our analysis showed us that demand for the entire apartment is more than any of the other conditions whereas shared Room scored the lowest score. Both graphs verify our report that Manhattan, Brooklyn neighborhoods have the maximum popularity and demand due to their geographical locations.



*(Figure 9. Dashboard 4 )*

In the below chart, we have exclusively used PowerBi tool for analyzing the Key Influencers and their impact on data, we have selected the Availability\_365 as our influencing factor. How it keeps getting changes if we manipulate the calculated\_host\_listing count and price to improve our model, it clearly shows us with numbers that if we increase the calculated\_host\_listing count till 28, then we can increase our availability up to 171 resulting in expansion as well as more option for customers for selection. Our current range varies from 4 to 28 resulting in 127 counts of availability and breakdown of the same if we reduce the count of host listing. The outcome gave us a brief idea about how the parameters can affect the results of price and availability. Similarly, we can use the other settings as critical influencers and perform analysis on the same to find out the best possible outcome for the same.



*(Figure 9. Influencers Visualization )*

**Conclusion & Recommendation**

According to Airbnb's performance in the epidemic situation, we will give corresponding recommendations based on the above series of analyses.

Strategically, since the pandemic is getting worse in the United States, Airbnb should temporarily increase its efforts to explore more potential cooperation partners in Staten Island, but also notify those hosts with a debt issue and business problem due to the COVID-19 pandemic to sell their properties in Bronx, or Queens to cover their loss at first and keep their popular properties in Manhattan if it is financially possible.

Tactically, Airbnb should incentivize more hosts to increase their timely reaction price because bedrooms, accommodates are challenging to manipulate from Airbnb's point of view. Moreover, Airbnb can do an extra evaluation to analyze similarity, which includes what determines the excessive average price for that region, how they charge adjustments in District will impact the average charge changes as a whole.

**References**

Cheng, Mingming, and Xin Jin. "What Do Airbnb Users Care About? An Analysis of Online Review Comments." International Journal of Hospitality Management 76 (2019): 58-70. Web.

Chicago Board Options Exchange, CBOE Volatility Index: VIX [VIXCLS], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/VIXCLS, May 31, 2020.

Cox, M. (2014, October). Inside Airbnb. Adding data to the debate. Retrieved June 24, 2020, from <http://insideairbnb.com/about.html>

Fitzgerald, M. (2020, March 14). Coronavirus and market volatility shut down the IPO market for potential listings like Airbnb. CNBC. Retrieved from<https://www.cnbc.com/2020/03/14/coronavirus-and-market-volatility-shuts-down-the-ipo-market-for-potential-listings-like-airbnb.html>

Gupta, S. (2019, January 05). Airbnb Rental Listings Dataset Mining. Retrieved June 24, 2020, from <https://towardsdatascience.com/airbnb-rental-listings-dataset-mining-f972ed08ddec>

Ju, Yongwook, Ki-Joon Back, Youngjoon Choi, and Jin-Soo Lee. "Exploring Airbnb Service Quality Attributes and Their Asymmetric Effects on Customer Satisfaction." International Journal of Hospitality Management 77 (2019): 342-52. Web.

     New York City. Adding data to the debate. (n.d.). Retrieved May 31, 2020, from<http://insideairbnb.com/newyorkcity/?neighbourhood=&filterEntireHomes=false&filterHighlyAvailable=false&filterRecentReviews=false&filterMultiListings=false>

Renaissance Capital. (2020, March 10). How the coronavirus has reshaped the 2020 IPO timeline.<https://www.renaissancecapital.com/IPO-Center/News/67674/How-the-coronavirus-has-reshaped-the-2020-IPO-timeline>

Warren, K. (2020, May 7). Airbnb has laid off 25% of its staff. Meet CEO Brian Chesky, who cofounded the company in 2008 to help pay his San Francisco apartment's rent and is now worth $4.1 billion. Business Insider. Retrieved from<https://www.businessinsider.com/brian-chesky-airbnb-life-career-net-worth-relationship-philanthropy>