**Tableau Progress Report 1**

**Group 12: Lingxuan Liu, Yue Zheng, Yixuan Huang, Hui Jing**

1. **Background**

Airbnb is a peer-to-peer online marketplace and homestay network. Since 2008, guests and hosts have used Airbnb to expand on traveling possibilities and present more unique, personalized way of experiencing the world.

This dataset describes the listing activity and metrics in NYC, NY for 2019. It is known that New York City is the nation’s largest short-term rental market, facilitated by hosting websites such as Airbnb. This is understandable given the high cost of hotel rooms in the city.

1. **Reason**

We choose this dataset because the following points.

Firstly, homestay is becoming more and more popular in traveling, especially for people like us, who would like to spend less money for better experience. So the analysis of the dataset could give us some useful suggestions for traveling. And Airbnb is one of the most well-known homestay network platform, which enables people to list or rent short-term lodging in residential properties, with the cost of such accommodation set by the property owner. The company receives percentage service fees from both guests and hosts in conjunction with every booking.

Secondly, this data file includes all needed information such as neighbourhood, room type, and room price to find out more about hosts, geographical availability, necessary metrics to make predictions and draw conclusions. It is suitable to be used to perform the analysis by using Tableau.

### **Inspiration**

Based on this dataset, we want to find out

* The differences among hosts and areas.
* The busiest host and the possible reason.
* Any noticeable difference of prices among different areas and the possible reason for it.
* The average price and the trend of the price Airbnb in New York.
* Making some predictions about locations, prices, reviews, etc.

**Tableau Progress Report 2**

**Group 12: Lingxuan Liu, Yue Zheng, Yixuan Huang, Hui Jing**

After watching the training video of Tableau, we can know that Tableau is not only a visual software, but also has lots of great abilities. First of all, it has considerable functions which are about data connection. It can directly connect to database like SQL, AWS, etc. Also, the data preparation and data processing can make user process more data and more quickly. Thus, we believe that Tableau is not just a better ‘Excel’, it can help us analyze data more well-rounded.

Besides the fact that tableau can serve as a tool for acquiring and loading data, we can then perform the initial understanding and do some necessary wrangle and clean the data.

After that, we then start to explore and visualize the data from different dimensions.

* The differences among hosts and the busiest host

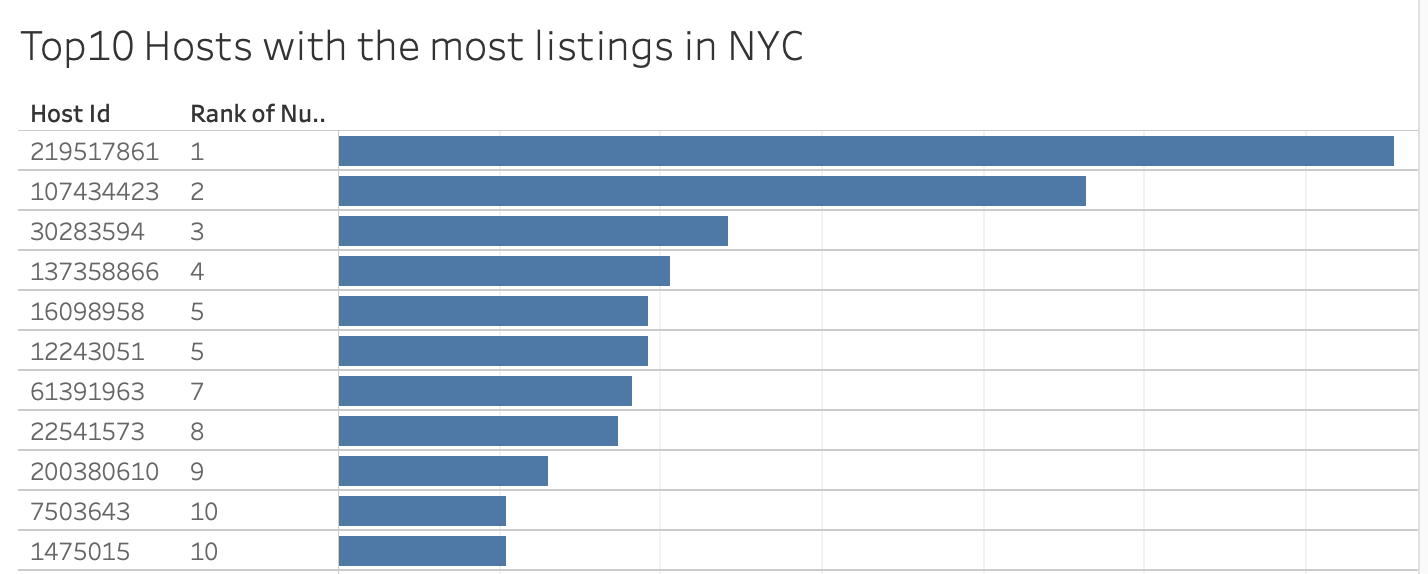
**We want to see which hosts (IDs) have the most listings on Airbnb platform.**

title: top 10 host ID with the most listings in NYC

dimension: Host Id

measure: sum(number of records)

Chart type: horizontal bar chart (showing as the following chart)



* Any noticeable difference of prices among different areas

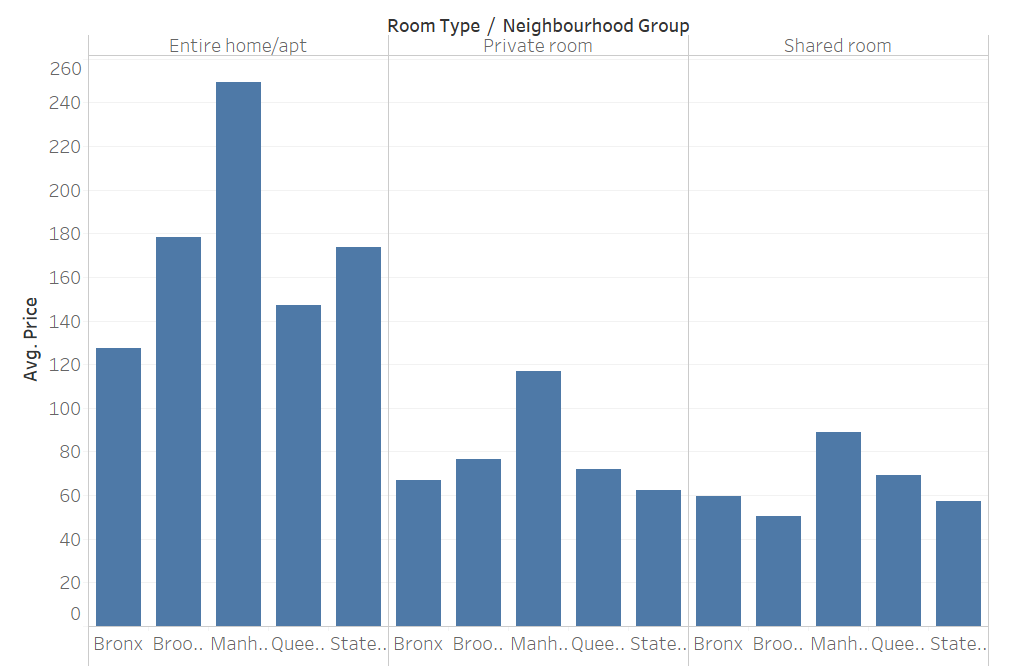
**We then want to find out more about how neighborhood groups and room types affect prices**

Title: Horizontal bar plot of the average prices for each neighberhood\_group and room type

Dimension: Room Type, Neighbourhood\_group

Measure: Price

Chart type: horizontal bar plot (showing as the following chart)



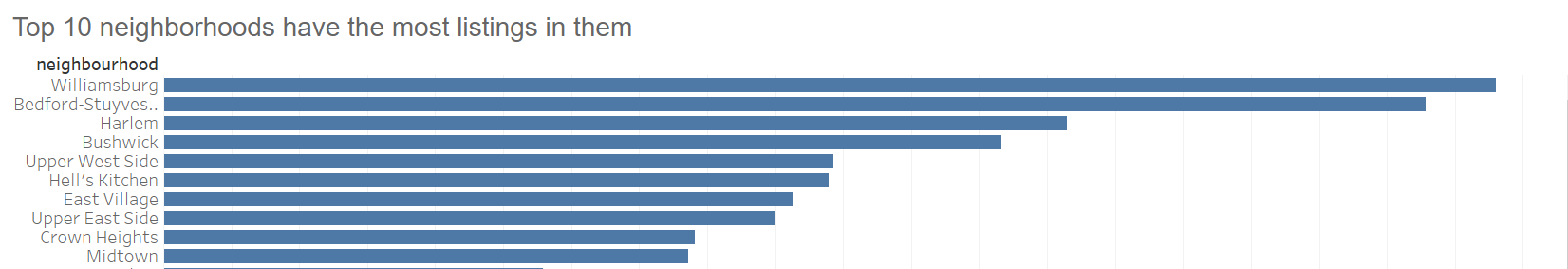
**We then want to find out more about our neighborhoods**

title: top 10 neighborhoods have the most listings in them

dimension: neighborhood

measure: sum(number of records)

Chart type: horizontal bar chart (showing as the following chart)

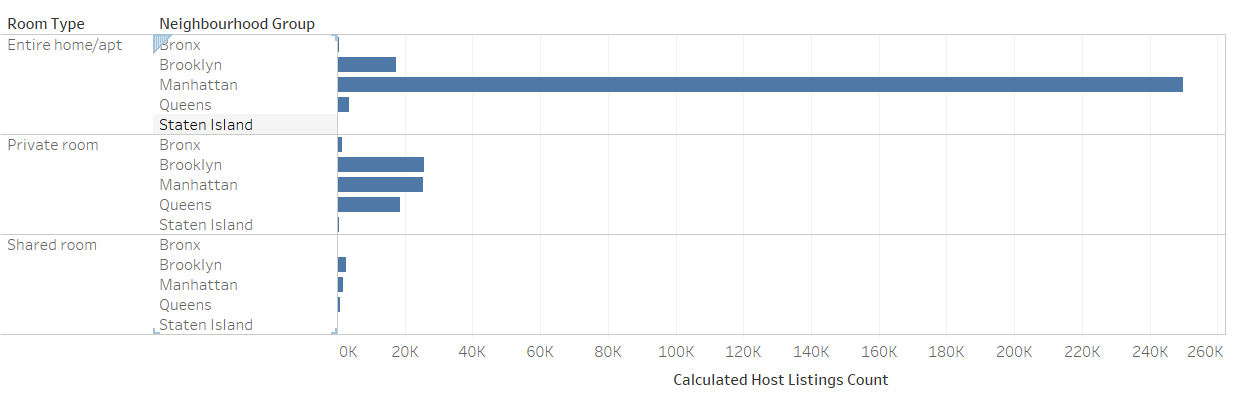


add more dimension into the analysis

dimension: oom type, neighbourhood\_group

measure: sum(number of records)

Chart type: present horizontal bar chart for each room type



**Tableau Progress Report 3**

**Group 12: Lingxuan Liu, Yue Zheng, Yixuan Huang, Hui Jing**

Tableau is truly useful for data analysis and visualization. By watching the Tableau training videos on website, we learned many advanced functions which include how to group and filter data, how to create dashboard, how to plot geographic data and so on. These functions could help us better analyze the Airbnb data in NYC based on the work we had done before. For example, by applying the mapping function, we could visualize the data geographically to better present the price differences among different neighbourhoods. And by creating dashboard, we could tell the story more clearly and better help the audiences understand our findings.

One of the challenges we have met would be the lack of knowledge for the environments of different neighbourhoods in NYC. Because we don’t live in NYC, we can’t do enough practical research for the NYC neighbourhoods and the conditions of Airbnb rooms in NYC. But we could do more research on Internet to learn more information to better understand the background and underlying reasons of the dataset we analyze.

Accomplishments:

* **We want to see the rough income information of each neighbourhood.**

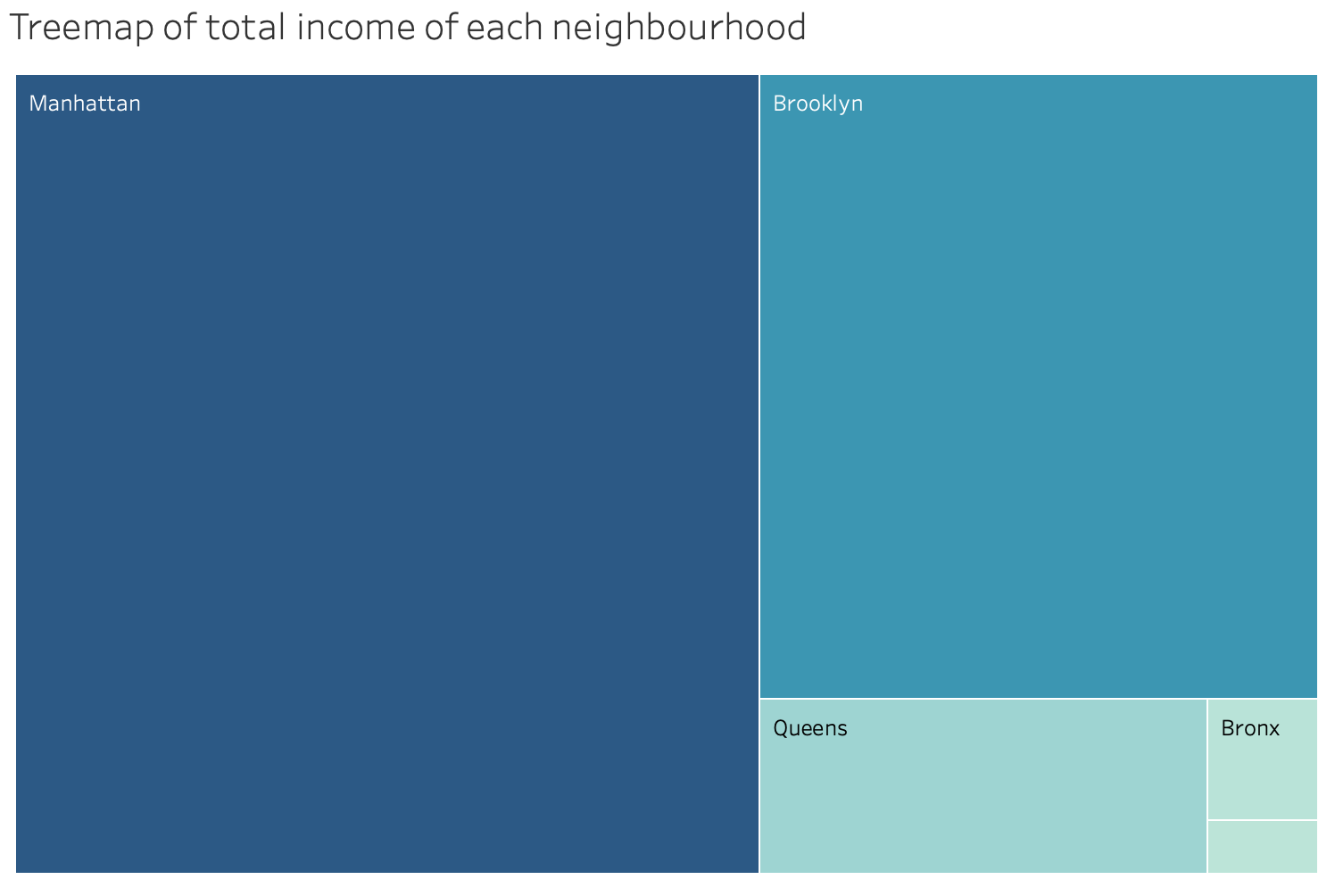
We define the income = average price \* number of record

Title: Treemap of the total income of each neighbourhood

Dimension: Neighbourhood Group

Measure: income = average price \* number of record

Chart type: treemap



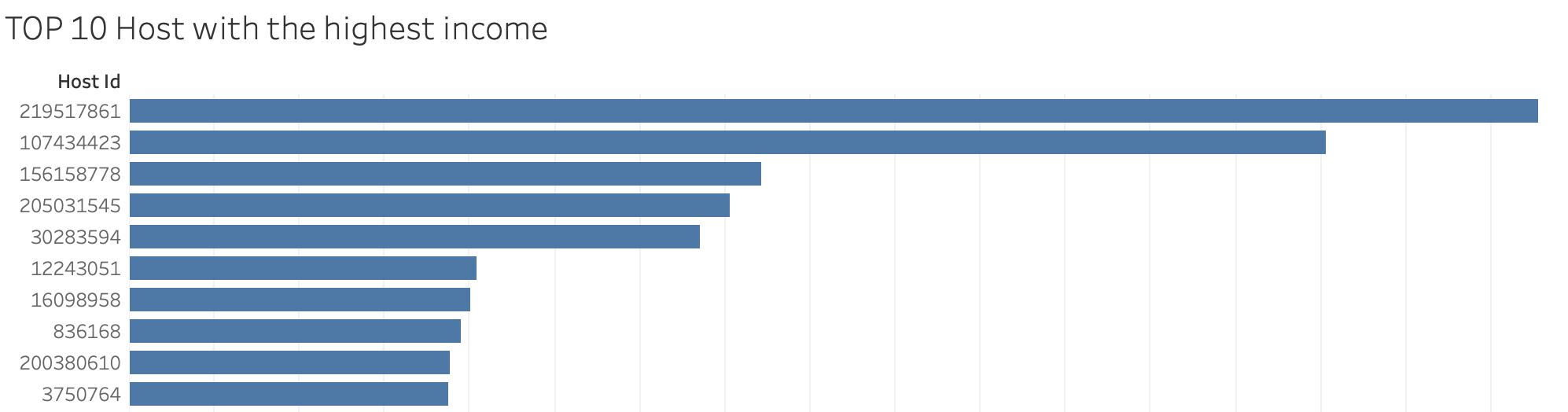
* **We want to see which hosts (IDs) earns the highest in NYC, and we can compare it with the hosts (IDs) have the most listings on Airbnb platform.**

Title: Top 10 host with the highest income in NYC

Dimension: Host Id

Measure: income = average price \* number of record

Chart type: horizontal bar chart (showing as the following chart)



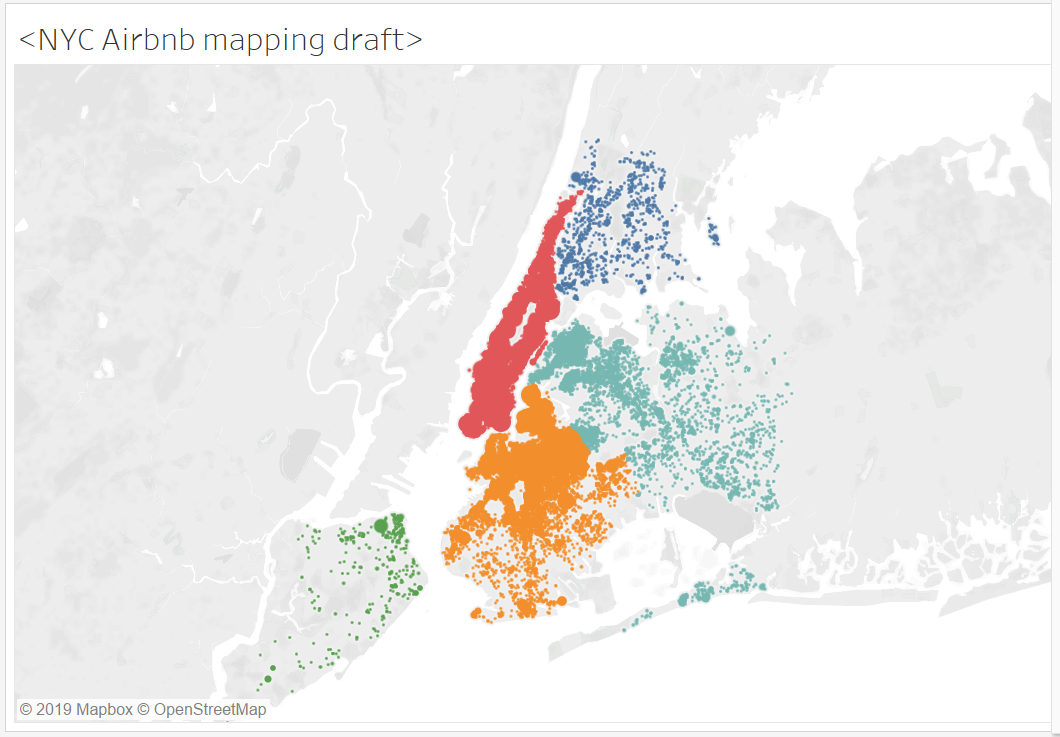
* **We want to see the mapping of Airbnb data in NYC to know about the differences about prices and neighbourhoods**

Title: NYC Airbnb mapping draft

Color: Neighbourhood

Size: Price

Detail: Name



**Tableau Final Report**

**Group 12: Lingxuan Liu, Yue Zheng, Yixuan Huang, Hui Jing**

**Introduction**

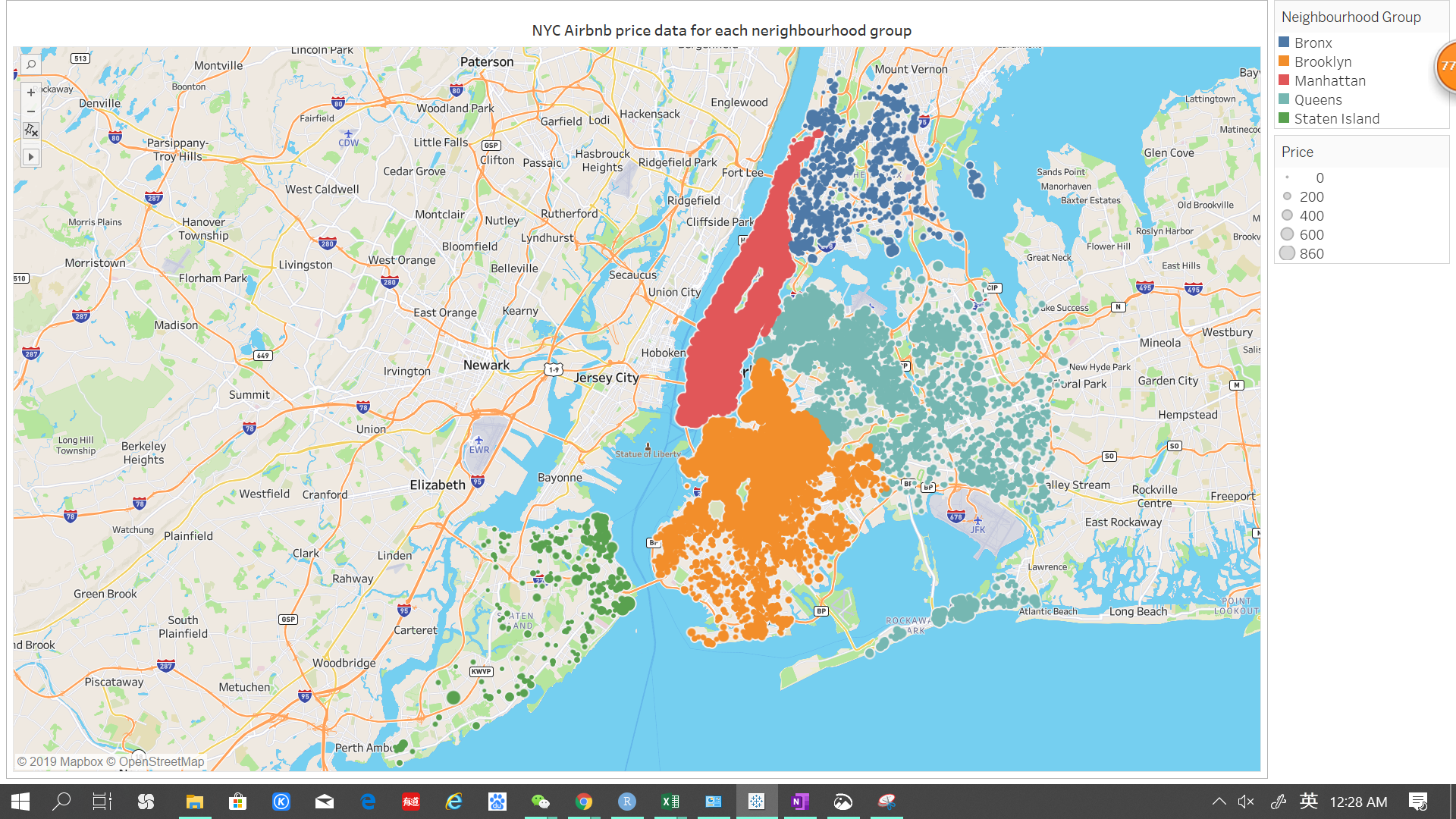
Airbnb is a peer-to-peer online marketplace and homestay network. Since 2008, guests and hosts have used Airbnb to expand on traveling possibilities and present more unique, personalized way of experiencing the world.

Our goal is to find out the most busiest hosts, the average price in each area, the most popular room and room type and so on.

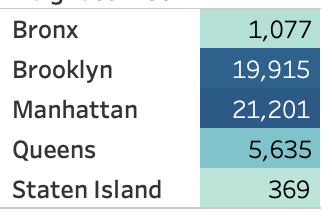
**Data Description**

* The Airbnb listing activity and metrics in NYC, NY for 2019
* Information covered: host information, location, price, room type, reviews, and other basic information
* Size: 5,888 KB
* One table with 48896 records and 16 columns
* Each record represents a Airbnb home information
* Source: Kaggle datasets <https://www.kaggle.com/dgomonov/new-york-city-airbnb-open-data>
* Clean data on Tableau Prep to filter the price as $0 to $850 to remove the outlier

**Analysis and Findings**

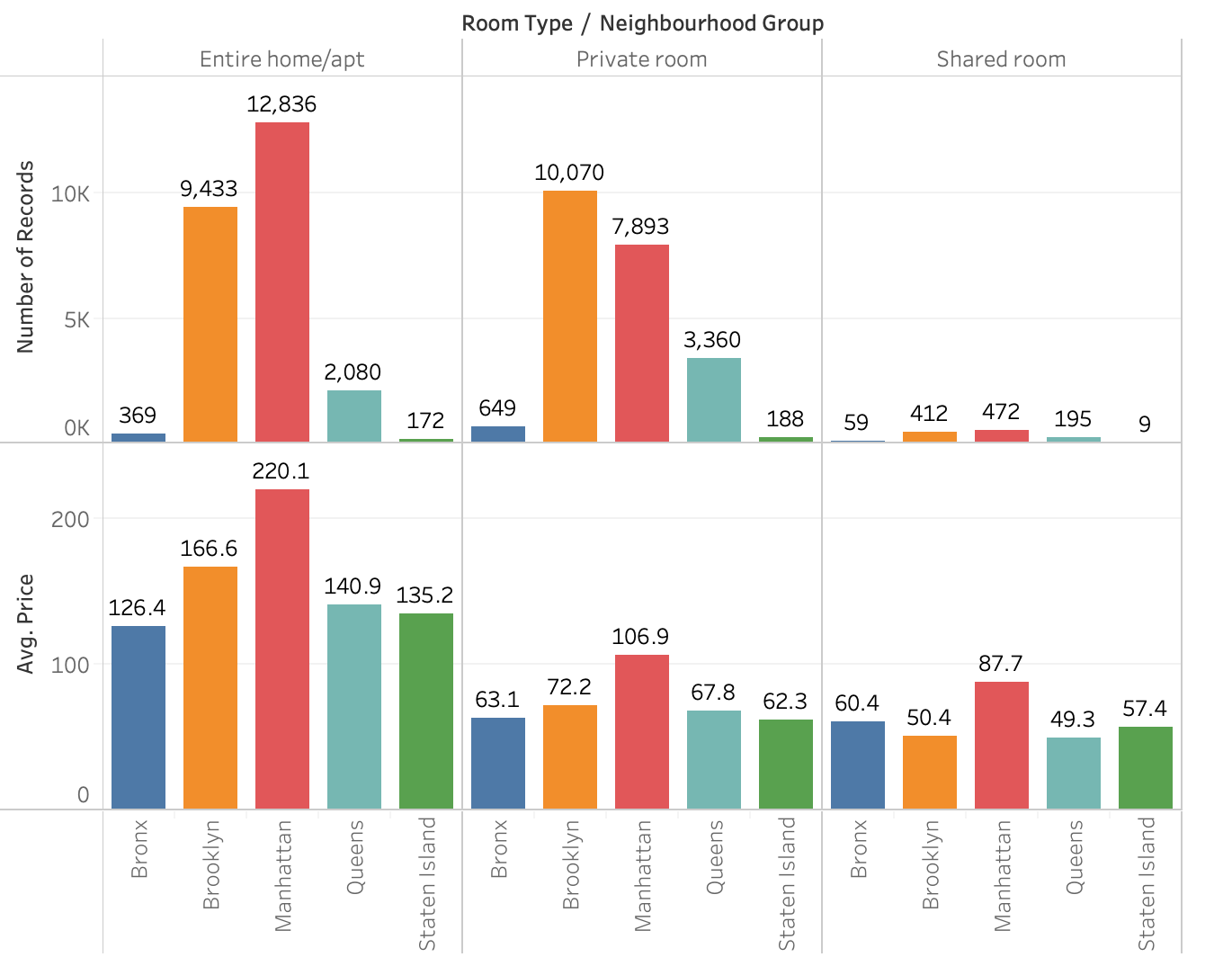
****

We created the mapping based on the latitude and longitude of each record from the dataset. Each spot represents one Airbnb record in the dataset. The neighbourhood group is classified by different colors, and the price per night is represented by the size of the spots. As we can see from the graph, the spots are mostly aggregate in Manhattan, the south of Queens, and the north of Brooklyn area. So we can tell that Airbnb is very popular in these areas, and we suppose that because these areas are busiest in NYC. The Staten Island seems to have the lowest amount of Airbnb homes. Generally, the spots are mostly concentrated in the central area of NYC, and are dispersed in the boulder area.

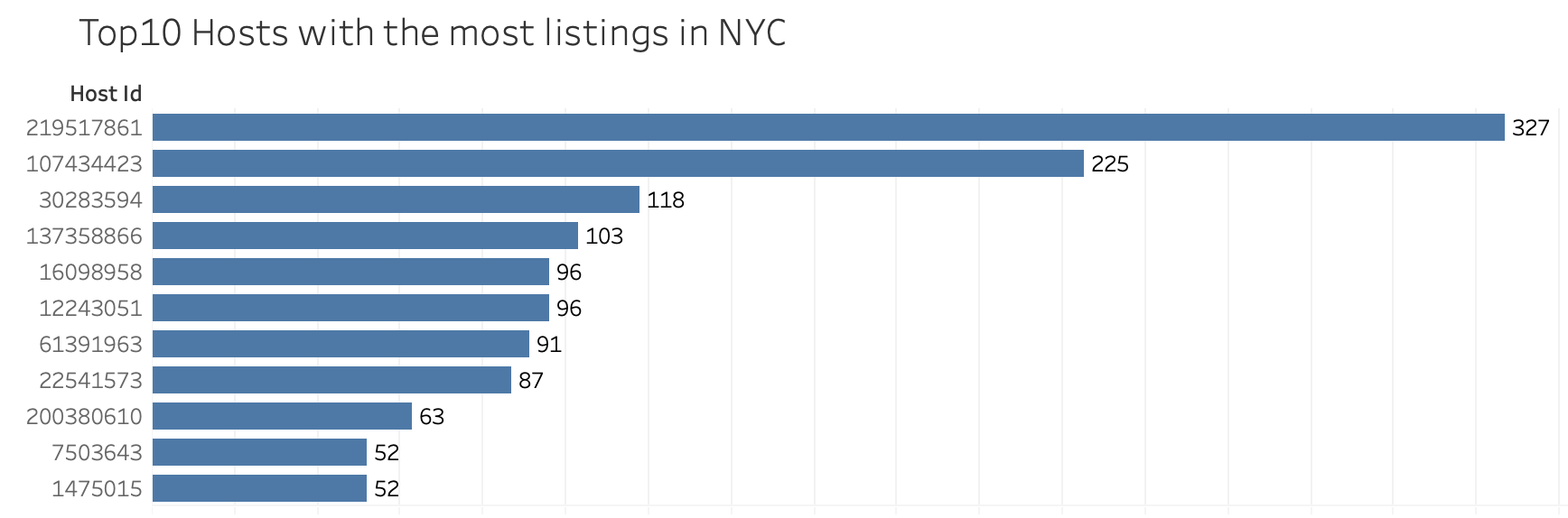


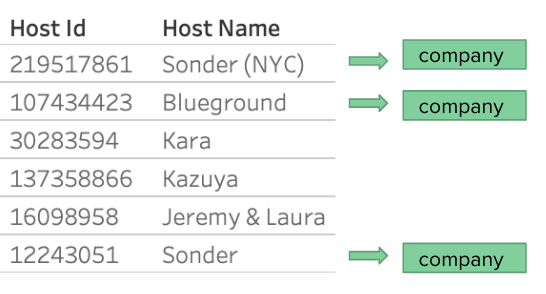
To compare the number of listings in different neighbourhood groups, we created a treemap. The darkness and the area of the squares represent the number of listings. Obviously, Manhattan and Brooklyn have the most listings with around 20 thousand homes respectively.

The bar chart at the bottom is about the average price in each neighbourhood group. Manhattan has the largest average price than the others with 196.9 dollars, and Bronx has the lowest price with 87.5 dollars. If we compare the treemap and the bar chart, even though Manhattan and Brooklyn have every similar amount of listings, but the average price in Manhattan is much larger than that of Brooklyn. So we suppose that the people in Manhattan should be richer than the people in Brooklyn in average.

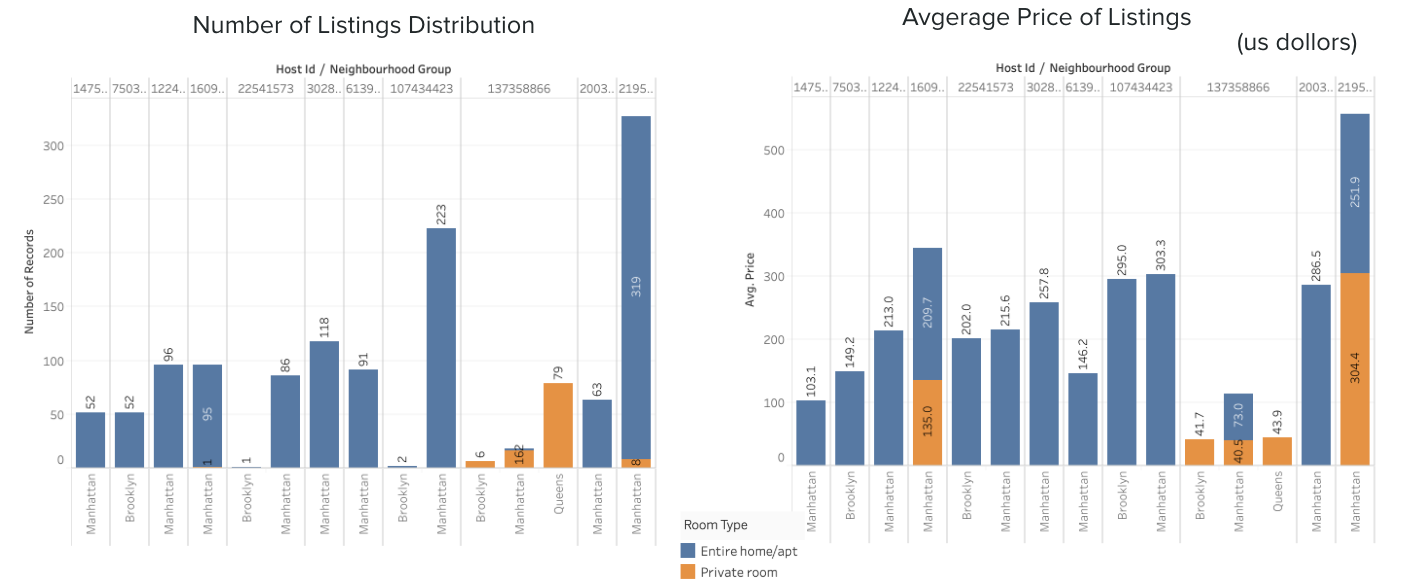


The bar chart above describes the total listing distribution in room type and neighbourhood group. The single bar chart can deliver a lot of informations.The color of the bars represent the different neighbourhood groups, and the size of the bars represent the average price per night for the Airbnb homes. The bar chart is grouped by room types, which are entire home, private room, and shared room. As we can see from the graph, generally, the shared rooms have the lowest price compared to the other room types, and the entire home or apt have the highest price. The amount of shared room is also the lowest compared to the other room types.

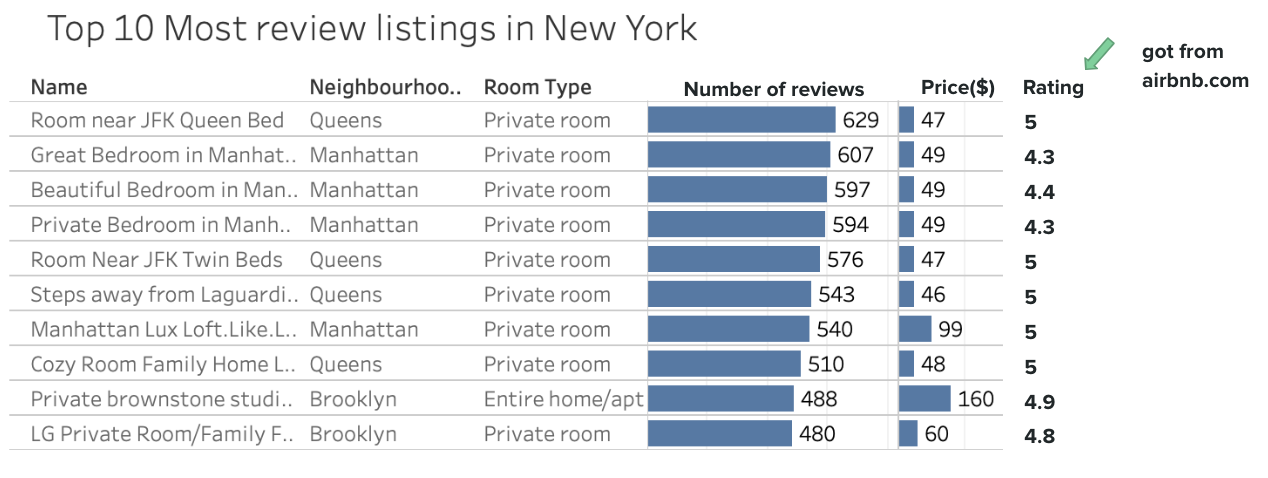




As we all known, house owners of airbnb usually are individuals who have spare room in their home and are willing to rent for the travelers, however, When we finished our investigation of the top 10 hosts with the most listings in NYC, we found there are several big company there are several big companies holding a large number of listings in the Airbnb. Eg. Sonder, the busiest host in Airbnb hold 327 listings, which means that some big corporations who have a hotel management style have already entered the Airbnb.



Further dive into the analysis of the top 10 hosts, we found that there is no shared room in the listings, as you can see, the majority of the room type of the Top 10 is entire room/apt. There is a host named kazuya, whose room type covering the private room offering the room with the lowest price among the top 10 hosts listings.



Whether we will choose a listing in the Airbnb, largely depends on its number of reviews and ratings, after filtering out the 10 listings with the most number of reviews, together with the rating information we found in Airbnb.com, we found that the top 10 listings who has the most number of reviews all have a relatively high rating score and a lower price.

**Summary**

In our analysis process, as you seen, by applying the mapping function, we could visualize the data geographically to better present the price differences among different neighbourhoods. And by creating dashboard, we could tell the story more clearly and better help the audiences understand our findings.

Our strength is that this data file includes all needed information such as neighbourhood, room type, and room price to find out more about hosts, geographical availability, necessary metrics to make predictions and draw conclusions. Like the longitude and latitude can allow us to get the mapping graph, which clearly shows the distribution of price in each neighbourhood. Also, because our dataset is sufficient, it contains 48896 records, so our conclusion like top 10 host with most listing and top 10 review listings are very persuasive.

The shortage we have had would be the lack of knowledge for the environments of different neighbourhoods in NYC. Because we don’t live in NYC, we can’t do enough practical research for the NYC neighbourhoods and the conditions of Airbnb rooms in NYC. So some of records might not be precise and real. Although we do some cleaning, we cannot make sure our data is perfect because of lack of knowledge for house pricing.

In a word, homestay is becoming more and more popular in traveling, especially for people like us, who would like to spend less money for better experience. So the analysis of the dataset could give us some useful suggestions for traveling. We hope we could find more comfort and highly cost effective room in Airbnb, not only in nyc but also in other cities.