Exploring the Airbnb Listings in New York City: An Exploratory Data Analysis

Project 1 - MAT214

Joel Joseph Jaison

2023-03-21

Description:

The Airbnb dataset is a collection of information on Airbnb listings and reviews from various cities around in New York.The dataset is commonly used for data analysis and visualization to gain insights into the pricing, popularity, and performance of Airbnb listings in different areas.

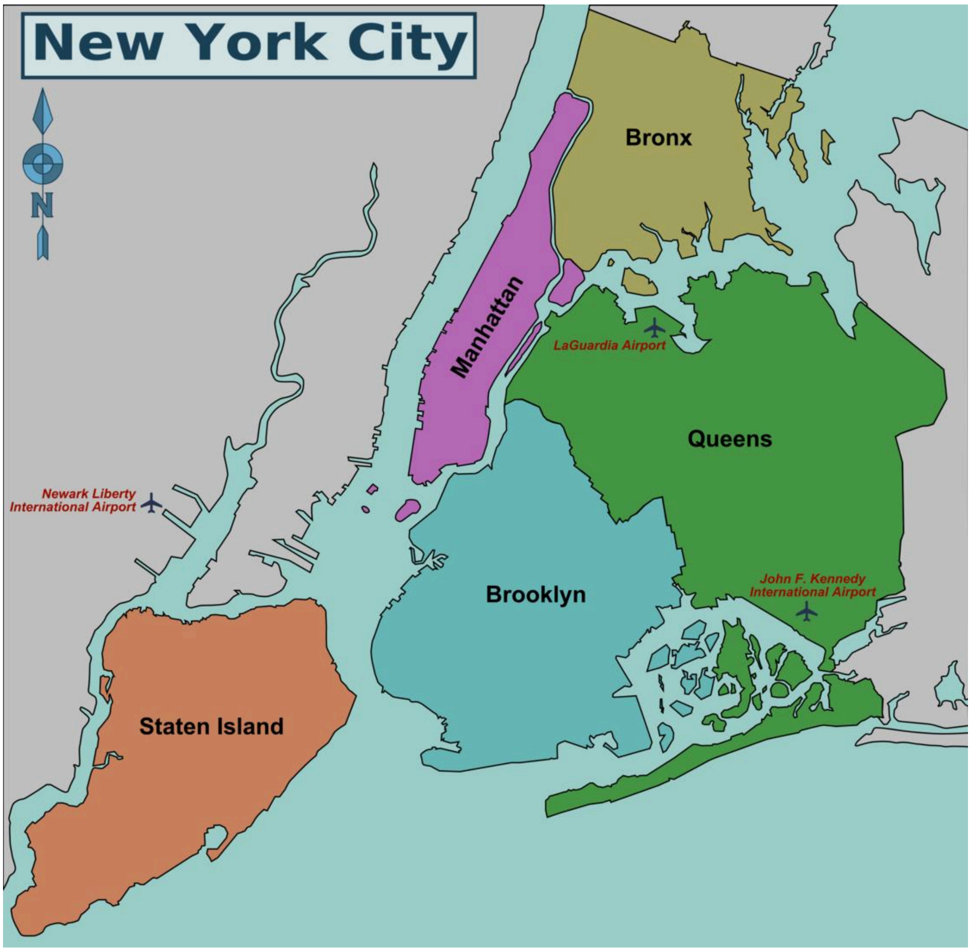

New York City
N
Bronx
Manhattan
LaGuardia Airport
Queens
Newark Liberty International Airport
Brooklyn
John F. Kennedy International Airport
Staten Island

Figure 1: Map of New York City with the five boroughs (1)

Reflective of the Project:

The objective of this project is to conduct an exploratory data analysis (EDA) of a chosen dataset that satisfies specific data requirements. The EDA will involve loading and parsing the data into R, visualizing and exploring the variation in the variables, identifying typical and unusual values, and investigating covariation between the variables. The project will also involve generating questions about the dataset to guide further analysis.

Format:

Variables	Description
id	ID number of the listing
name	Name of the listing
host_id	ID number of the host
host_name	Name of the host
neighbourhood_group	Boroughs in which the listing is located
neighbourhood	Neighborhood in which the listing is located
latitude	Latitude of the listing
longitude	Longitude of the listing
room_type	Type of room (e.g. Entire home/apt, Private room)
price	Price per night
minimum_nights	Minimum number of nights for a stay
number_of_reviews	Number of reviews for the listing
last_review	Date of the latest review
reviews_per_month	Number of reviews per month
calculated_host_listings_count	Number of listings by the same host
availability_365	Number of days in the year that the listing is available

Source:

https://www.kaggle.com/datasets/dgomonov/new-york-city-airbnb-open-data (https://www.kaggle.com/datasets/dgomonov/new-york-city-airbnb-open-data)

Task 1:

1.Parsing Data

```
airbnb <- read.csv("C:/Users/joelj/OneDrive/Desktop/AB_NYC_2019.csv")
```

This line of code loads the data stored in the file “AB_NYC_2019.csv” into R as a dataframe and assigns it to the variable name “airbnb”.This line of code assumes that our data set is in .csv format. The filepath of the file is specified as

“C:/Users/joelj/OneDrive/Desktop/” and may need to be adjusted depending on where the file is located on the user’s computer. Once loaded, the data can be manipulated and analyzed using various R functions and packages.

2. Requirements

```
str(airbnb)
```

```
## 'data.frame':    48895 obs. of  16 variables:
##  $ id                            : int  2539 2595 3647 3831 5022 5099 5121 5178 5203 5238
...
##  $ name                          : chr  "Clean & quiet apt home by the park" "Skylit Midto
wn Castle" "THE VILLAGE OF HARLEM....NEW YORK !" "Cozy Entire Floor of Brownstone" ...
##  $ host_id                       : int  2787 2845 4632 4869 7192 7322 7356 8967 7490 7549
...
##  $ host_name                     : chr  "John" "Jennifer" "Elisabeth" "LisaRoxanne" ...
##  $ neighbourhood_group           : chr  "Brooklyn" "Manhattan" "Manhattan" "Brooklyn" ...
##  $ neighbourhood                 : chr  "Kensington" "Midtown" "Harlem" "Clinton Hill" ...
##  $ latitude                      : num  40.6 40.8 40.8 40.7 40.8 ...
##  $ longitude                     : num  -74 -74 -73.9 -74 -73.9 ...
##  $ room_type                     : chr  "Private room" "Entire home/apt" "Private room" "E
ntire home/apt" ...
##  $ price                         : int  149 225 150 89 80 200 60 79 79 150 ...
##  $ minimum_nights                : int  1 1 3 1 10 3 45 2 2 1 ...
##  $ number_of_reviews             : int  9 45 0 270 9 74 49 430 118 160 ...
##  $ last_review                   : chr  "2018-10-19" "2019-05-21" "" "2019-07-05" ...
##  $ reviews_per_month             : num  0.21 0.38 NA 4.64 0.1 0.59 0.4 3.47 0.99 1.33 ...
##  $ calculated_host_listings_count: int  6 2 1 1 1 1 1 1 1 4 ...
##  $ availability_365              : int  365 355 365 194 0 129 0 220 0 188 ...
```

```
table(airbnb$neighbourhood_group)
```

```
## 
##         Bronx      Brooklyn     Manhattan        Queens Staten Island 
##          1091         20104         21661          5666           373
```

Using the str() function , we can see the structure of the dataset. Here we can see that:

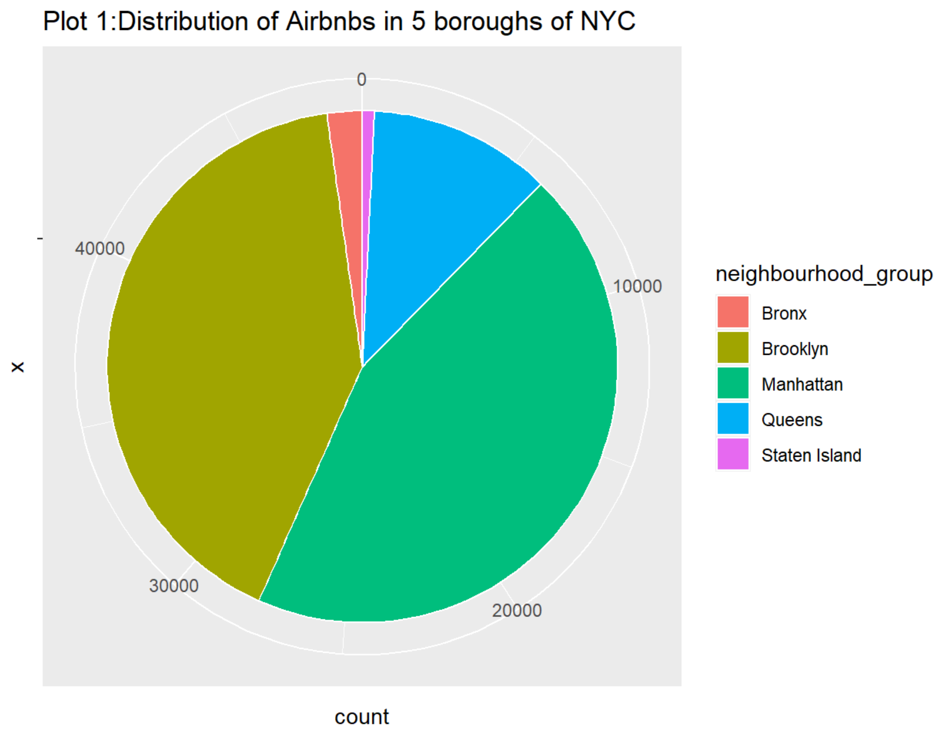
1. There is 48895 observations.
2. There are 3 numerical variables which are latitude, longitude and reviews_per_month
3. There are 2 categorical variables which are neighbourhood_group and room_type
4. There are 16 variables in total

Additionally , We see that the variable last_review consists of dates

Therefore this dataset, satisfies all our requirements

Task 2:

1. Plots of Variation of Variables


Plot 1:Distribution of Airbnbs in 5 boroughs of NYC

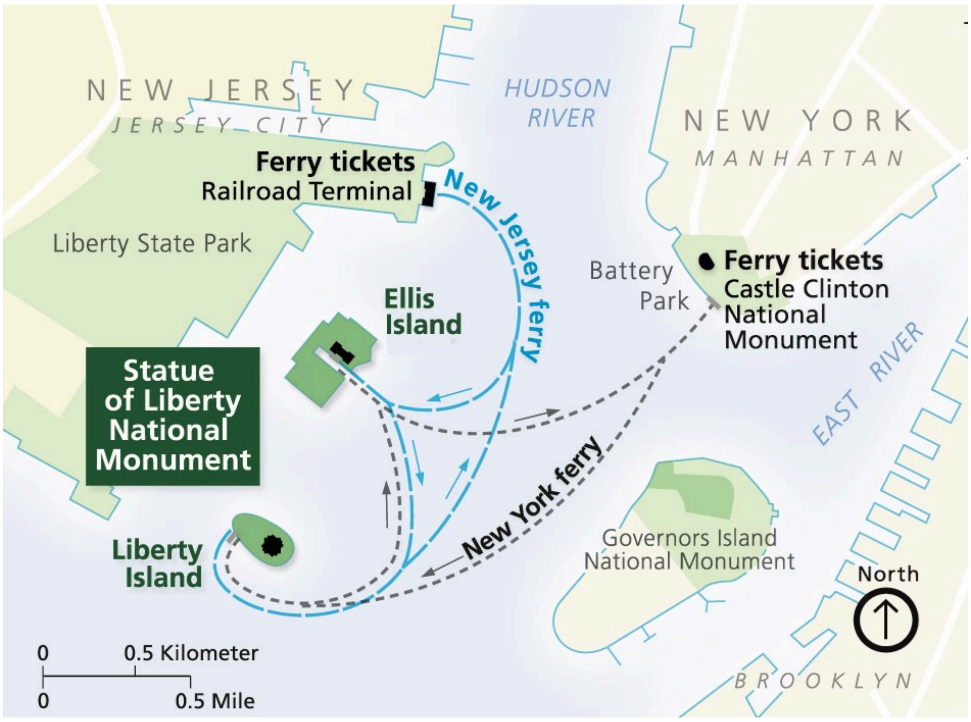
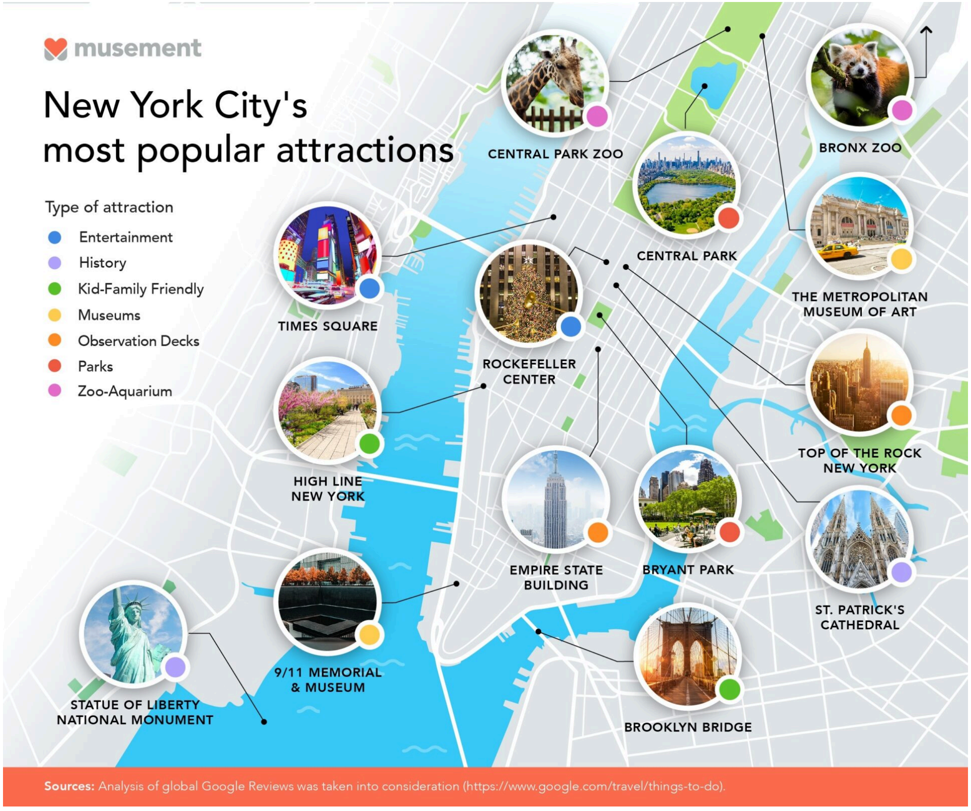


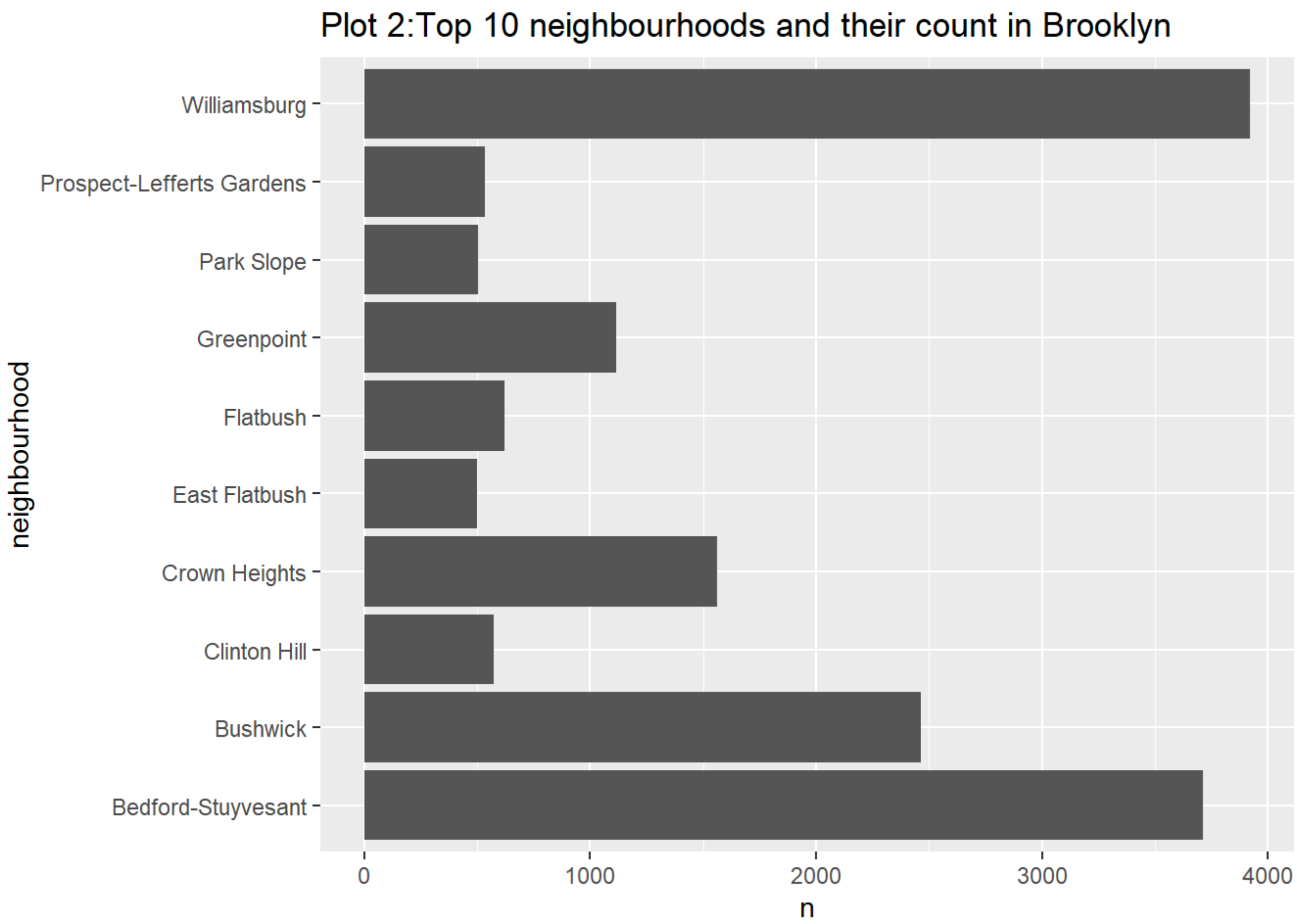
Figure 2: Ferry Map of New York City (2)



Sources: Analysis of global Google Reviews was taken into consideration (https://www.google.com/travel/things-to-do).

Figure 3: Map with the most famous attractions in New York City(3)

This is a polar bar chart which provides a visualization of the distribution of airbnb listings in New York City. Manhattan has the largest number of airbnb listing among all the neighbourhoods. It was found from Figure 2 that Manhattan is one of the closest borough in New York which has easy access to the Statue of Liberty which is one of the most famous attractions in The United States. And from Figure 3 we can see that Manhattan has other important attractions like the Empire State Building, Rockfeller Center and even includes Times Square where about 330,000 people pass through it on a daily basis(4). We also see that distribution in Brooklyn is almost equal to that of Manhattan, this is because Brooklyn has some famous attractions like the Brooklyn Bridge (Figure 3). But surely other factors like Low Seasonality, Lenient Airbnb Laws and Regulations,Access to Public Transportation and Amenities and many others could also be a factor to the distribution of airbnbs in NYC but in this project we focus on just the proximity to famous attractions(5)



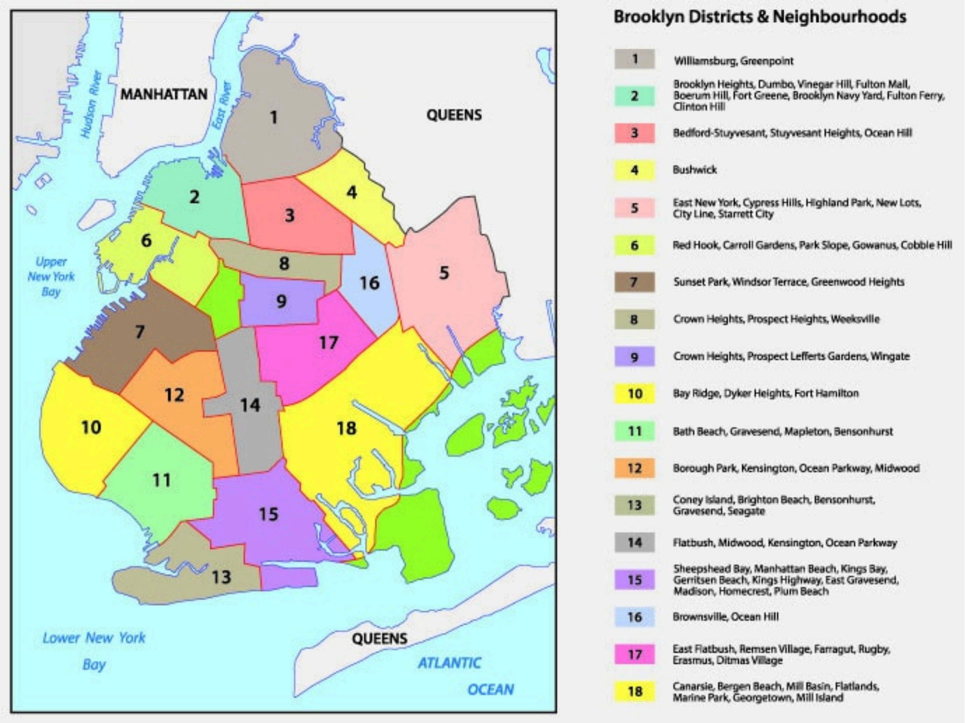

MANHATTAN
Hudson River
East River
QUEENS
Upper New York Bay
Lower New York Bay
QUEENS
ATLANTIC OCEAN
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
Brooklyn Districts & Neighbourhoods
1 Williamsburg, Greenpoint
2 Brooklyn Heights, Dumbo, Vinegar Hill, Fulton Mall, Boerum Hill, Fort Greene, Brooklyn Navy Yard, Fulton Ferry, Clinton Hill
3 Bedford-Stuyvesant, Stuyvesant Heights, Ocean Hill
4 Bushwick
5 East New York, Cypress Hills, Highland Park, New Lots, City Line, Starrett City
6 Red Hook, Carroll Gardens, Park Slope, Gowanus, Cobble Hill
7 Sunset Park, Windsor Terrace, Greenwood Heights
8 Crown Heights, Prospect Heights, Weeksville
9 Crown Heights, Prospect Lefferts Gardens, Wingate
10 Bay Ridge, Dyker Heights, Fort Hamilton
11 Bath Beach, Gravesend, Mapleton, Bensonhurst
12 Borough Park, Kensington, Ocean Parkway, Midwood
13 Coney Island, Brighton Beach, Bensonhurst, Gravesend, Seagate
14 Flatbush, Midwood, Kensington, Ocean Parkway
15 Sheepshead Bay, Manhattan Beach, Kings Bay, Gerritsen Beach, Kings Highway, East Gravesend, Madison, Homecrest, Plum Beach
16 Brownsville, Ocean Hill
17 East Flatbush, Remsen Village, Farragut, Rugby, Erasmus, Ditmas Village
18 Canarsie, Bergen Beach, Mill Basin, Flatlands, Marine Park, Georgetown, Mill Island

Figure 4: Map with the neighbourhoods in Brooklyn(6)

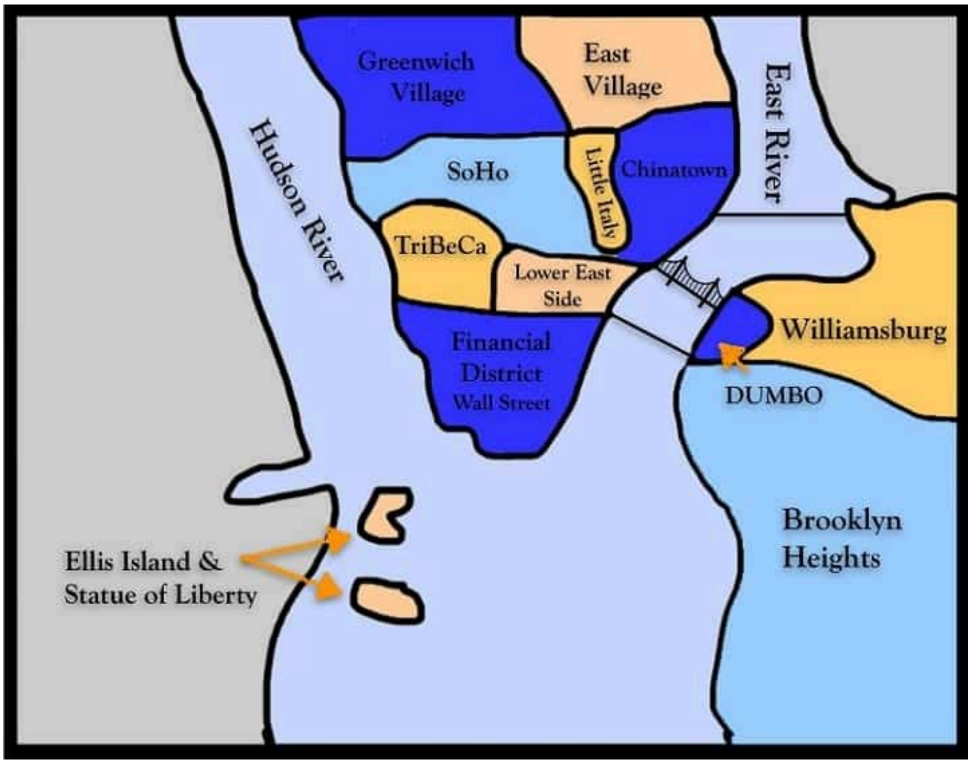
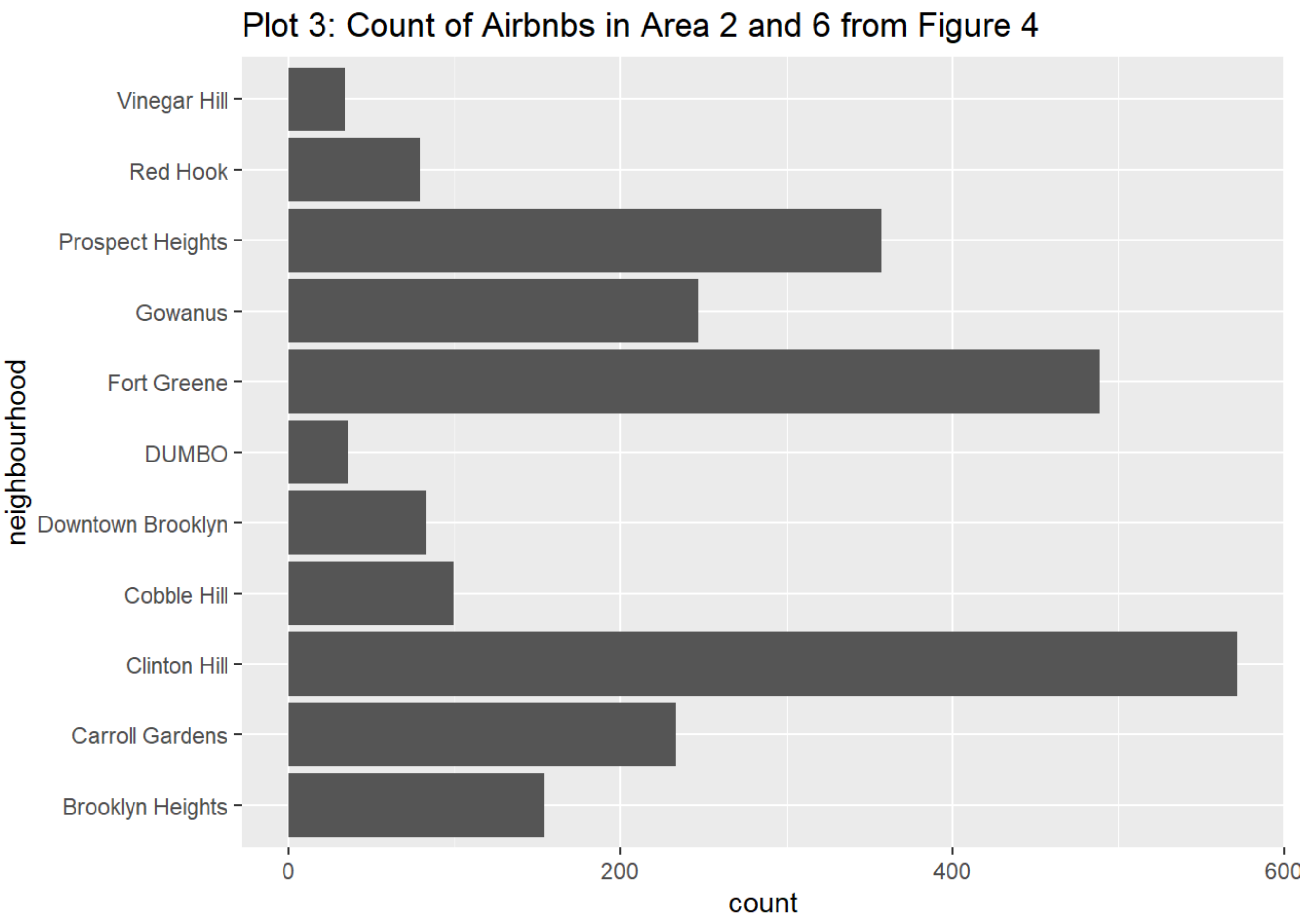
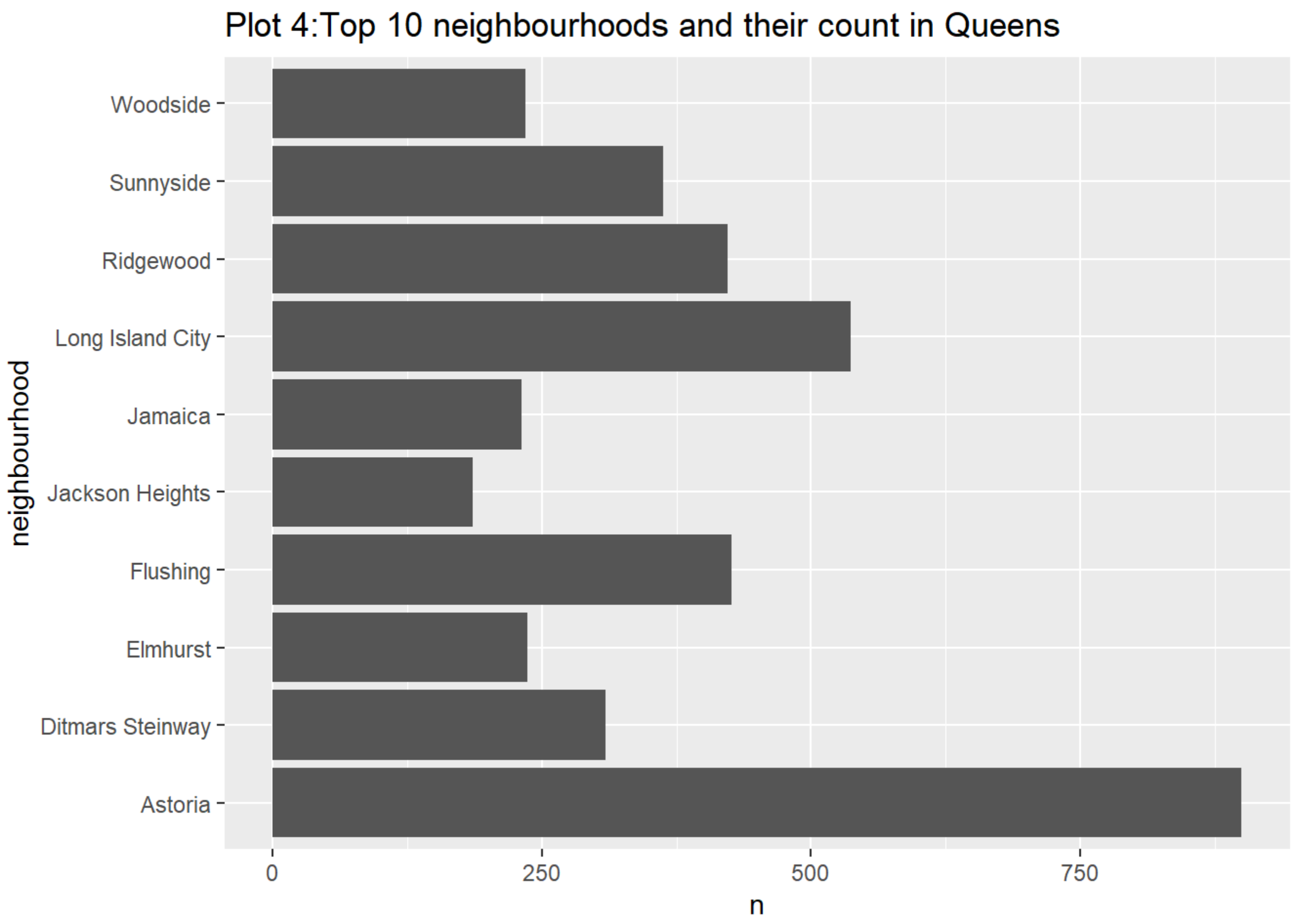


Figure 5: Map of Brooklyn and its proximity to Liberty Island(7)

To further this analysis, I looked into a bar graph of the Top 10 neighbourhoods based on the largest number of airbnbs in Brooklyn. From this plot I was able to understand that Bedford-Stuyvesant, Willamsburg and Bushwick have very high distribution compared to the other neighbourhoods and on further analysis I was able to understand that all these three neighbourhoods are very close to Manhattan(Figure 4).From this I came to the conclusion that people tend to stay in Brooklyn and then travel all the way to Manhattan instead of staying directly at Manhattan. But why would people tend to stay in Brooklyn rather than staying in Manhattan directly? This will be discussed later in this project. Another observation that can be found from Figure 5 is that neighbourhoods from areas 2 and 6 are very close to the Liberty Island, then surely people would want to travel from Brooklyn to Liberty Island. But why aren't these neighbourhoods in the top 10 ?



I created a plot to analyze why these two specific areas were not included in the top 10 list. The plot revealed that the count of certain attributes was significantly lower in these areas, which explains why they did not receive a higher ranking. But again why isn’t it more in number especially because of the proximity to Liberty Island. Further examination, specifically Figure 2, indicated that the transportation options available to Liberty Island were only accessible from New Jersey or Manhattan, not Brooklyn. This lack of transportation connection to Liberty Island may explain why there is no incentive for individuals to reside in Brooklyn if their main interest is visiting Liberty Island.



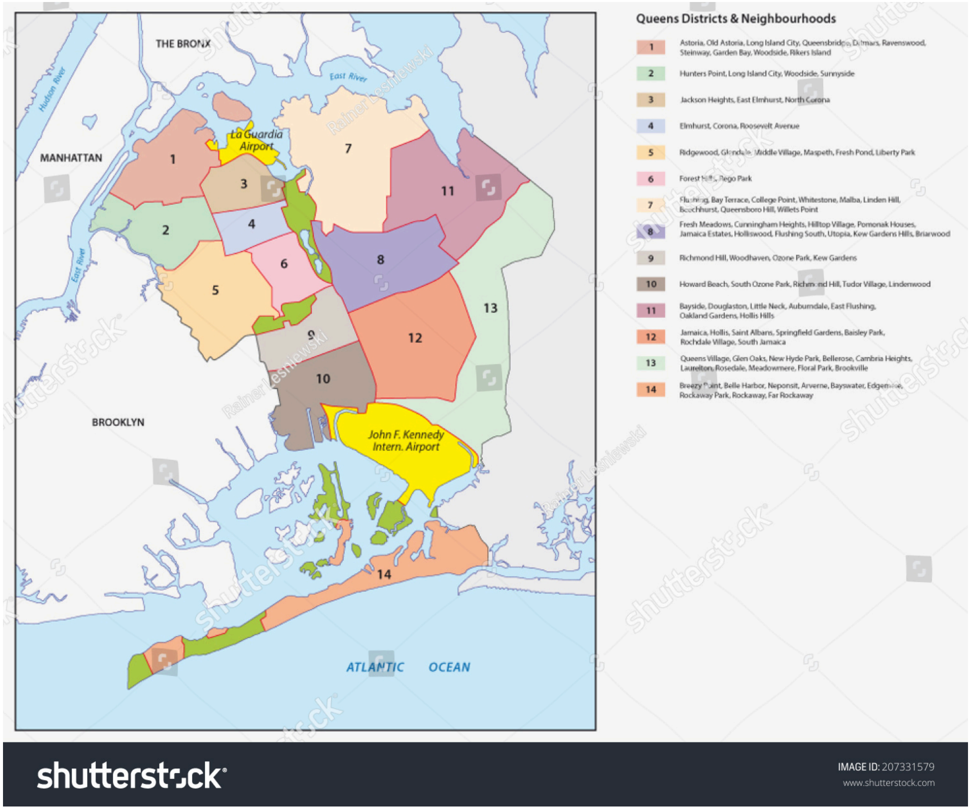
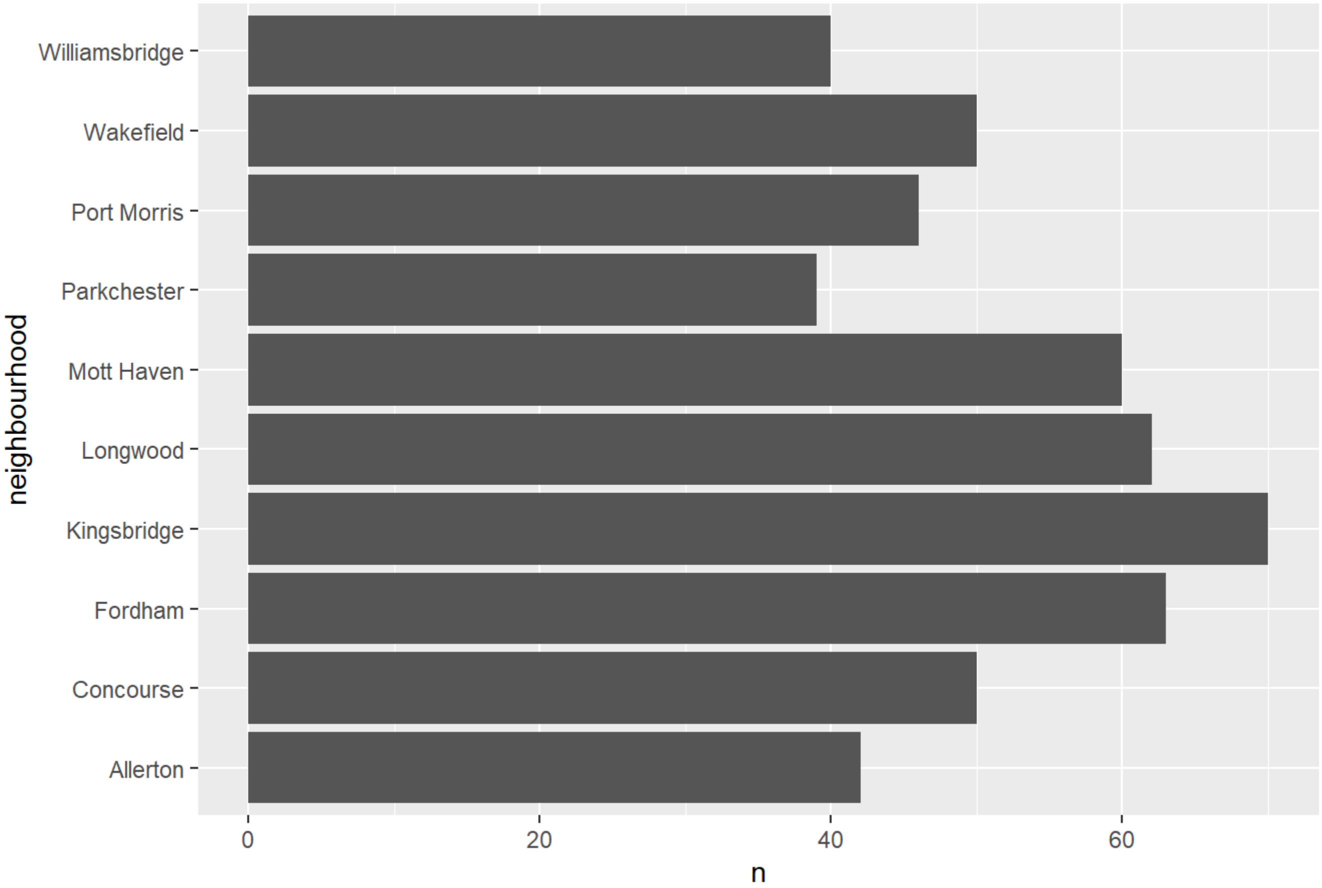

THE BRONX
MANHATTAN
BROOKLYN
Hudson River
East River
La Guardia Airport
John F. Kennedy Intern. Airport
ATLANTIC OCEAN
1
2
3
4
5
6
7
8
9
10
11
12
13
14
Queens Districts & Neighbourhoods
1 Astoria, Old Astoria, Long Island City, Queensbridge, Ditmars, Ravenswood, Steinway, Garden Bay, Woodside, Rikers Island
2 Hunters Point, Long Island City, Woodside, Sunnyside
3 Jackson Heights, East Elmhurst, North Corona
4 Elmhurst, Corona, Roosevelt Avenue
5 Ridgewood, Glendale, Middle Village, Maspeth, Fresh Pond, Liberty Park
6 Forest Hills, Rego Park
7 Flushing, Bay Terrace, College Point, Whitestone, Malba, Linden Hill, Beechhurst, Queensboro Hill, Willets Point
8 Fresh Meadows, Cunningham Heights, Hilltop Village, Pomonak Houses, Jamaica Estates, Holliswood, Flushing South, Utopia, Kew Gardens Hills, Briarwood
9 Richmond Hill, Woodhaven, Ozone Park, Kew Gardens
10 Howard Beach, South Ozone Park, Richmond Hill, Tudor Village, Lindenwood
11 Bayside, Douglaston, Little Neck, Auburndale, East Flushing, Oakland Gardens, Hollis Hills
12 Jamaica, Hollis, Saint Albans, Springfield Gardens, Baisley Park, Rochdale Village, South Jamaica
13 Queens Village, Glen Oaks, New Hyde Park, Bellerose, Cambria Heights, Laurelton, Rosedale, Meadowmere, Floral Park, Brookville
14 Breezy Point, Belle Harbor, Neponsit, Arverne, Bayswater, Edgemere, Rockaway Park, Rockaway, Far Rockaway

Figure 6: Map with the neighbourhoods in Queens(8)

We were able to find some good relations between neighbourhoods of Brooklyn with Manhattan , thereby I tried to find such variations in other neighbourhood groups. When we analyze on the neighbourhoods in Queens, we find that the same kind of observations are found here, like the fact that Astoria and Long Island City has highest distribution among the top neighbourhoods in Queens. And these two neighbourhoods are close to Manhattan (Figure 6). Again the same factor of proximity to Manhattan makes this place more attractive than the other neighbourhoods.

Plot 5:Top 10 neighbourhoods and their count in Bronx



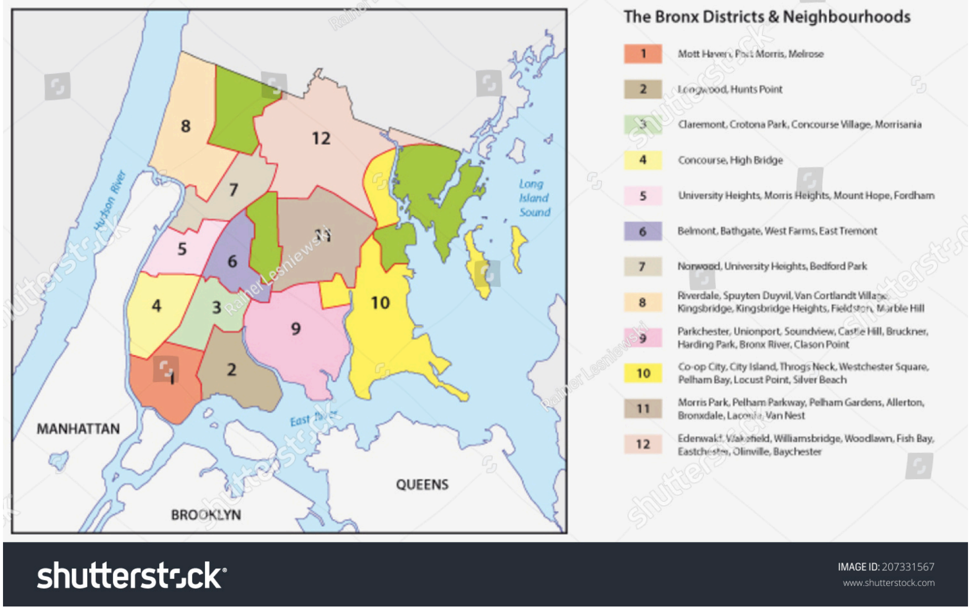

8
12
7
5
6
11
4
3
9
10
1
2
Hudson River
Long Island Sound
East River
MANHATTAN
BROOKLYN
QUEENS
The Bronx Districts & Neighbourhoods
1 Mott Haven, Port Morris, Melrose
2 Longwood, Hunts Point
3 Claremont, Crotona Park, Concourse Village, Morrisania
4 Concourse, High Bridge
5 University Heights, Morris Heights, Mount Hope, Fordham
6 Belmont, Bathgate, West Farms, East Tremont
7 Norwood, University Heights, Bedford Park
8 Riverdale, Spuyten Duyvil, Van Cortlandt Village, Kingsbridge, Kingsbridge Heights, Fieldston, Marble Hill
9 Parkchester, Unionport, Soundview, Castle Hill, Bruckner, Harding Park, Bronx River, Clason Point
10 Co-op City, City Island, Throgs Neck, Westchester Square, Pelham Bay, Locust Point, Silver Beach
11 Morris Park, Pelham Parkway, Pelham Gardens, Allerton, Bronxdale, Laconia, Van Nest
12 Edenwald, Wakefield, Williamsbridge, Woodlawn, Fish Bay, Eastchester, Olinville, Baychester

Figure 7: Map with the neighbourhoods in Bronx(9)

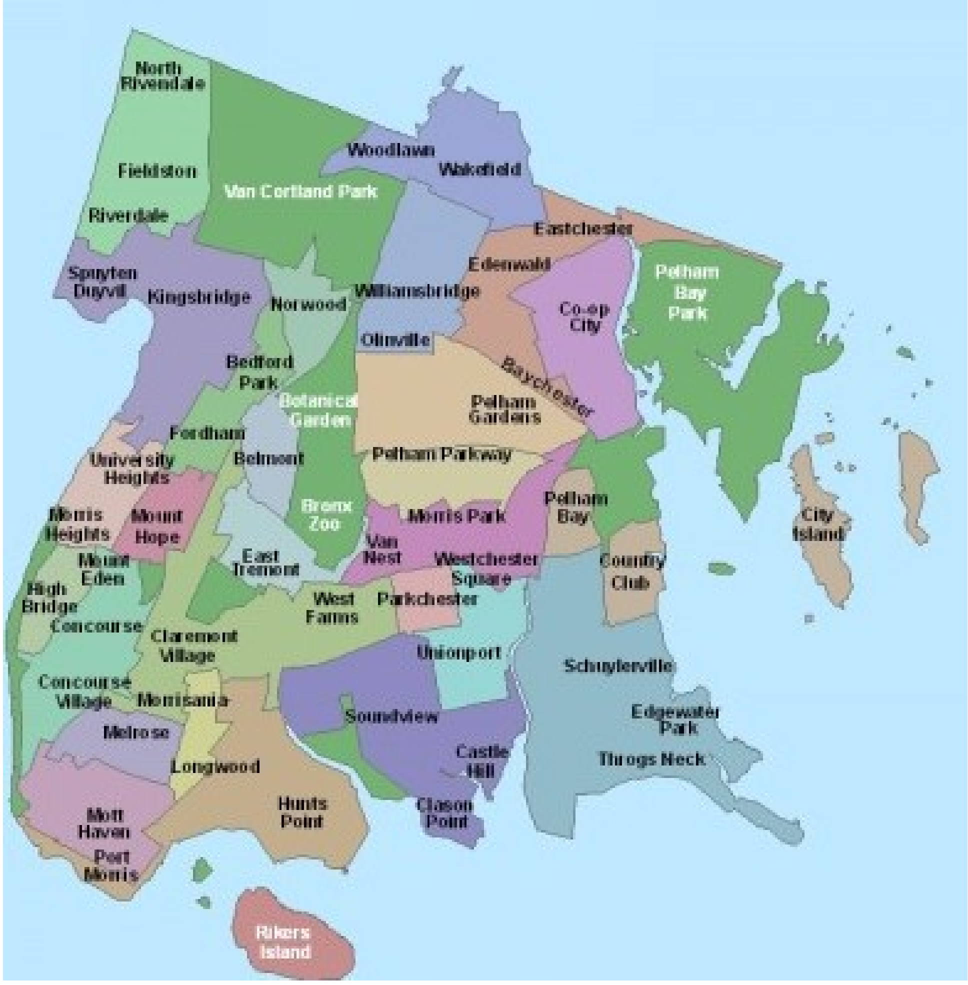


Figure 8: Map with the neighbourhoods and Bronx zoo in Bronx(10)

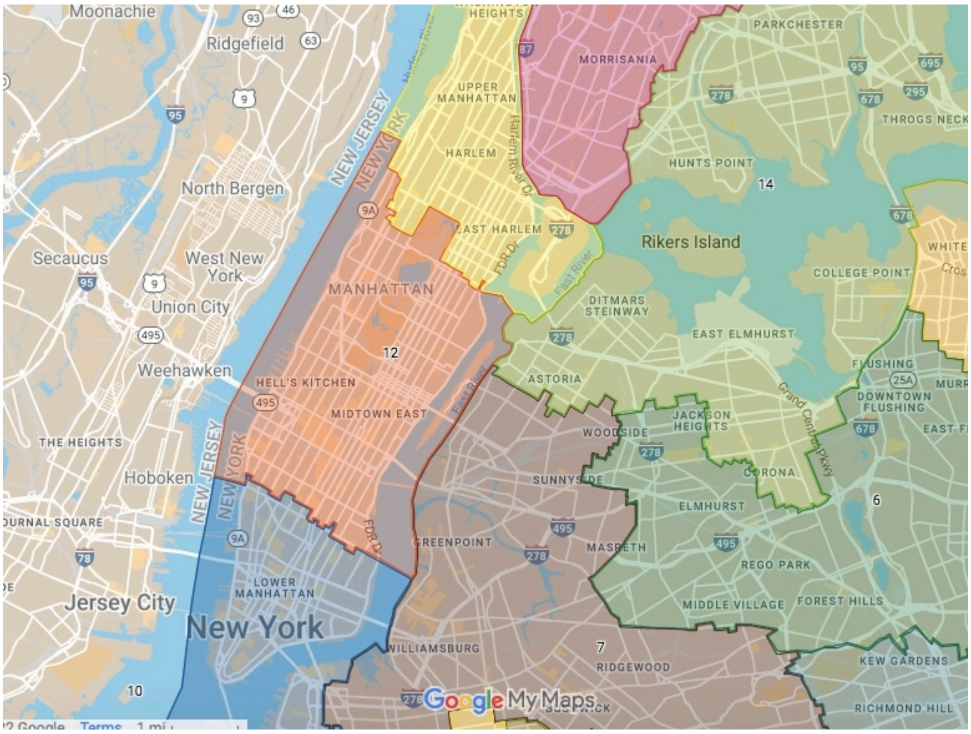
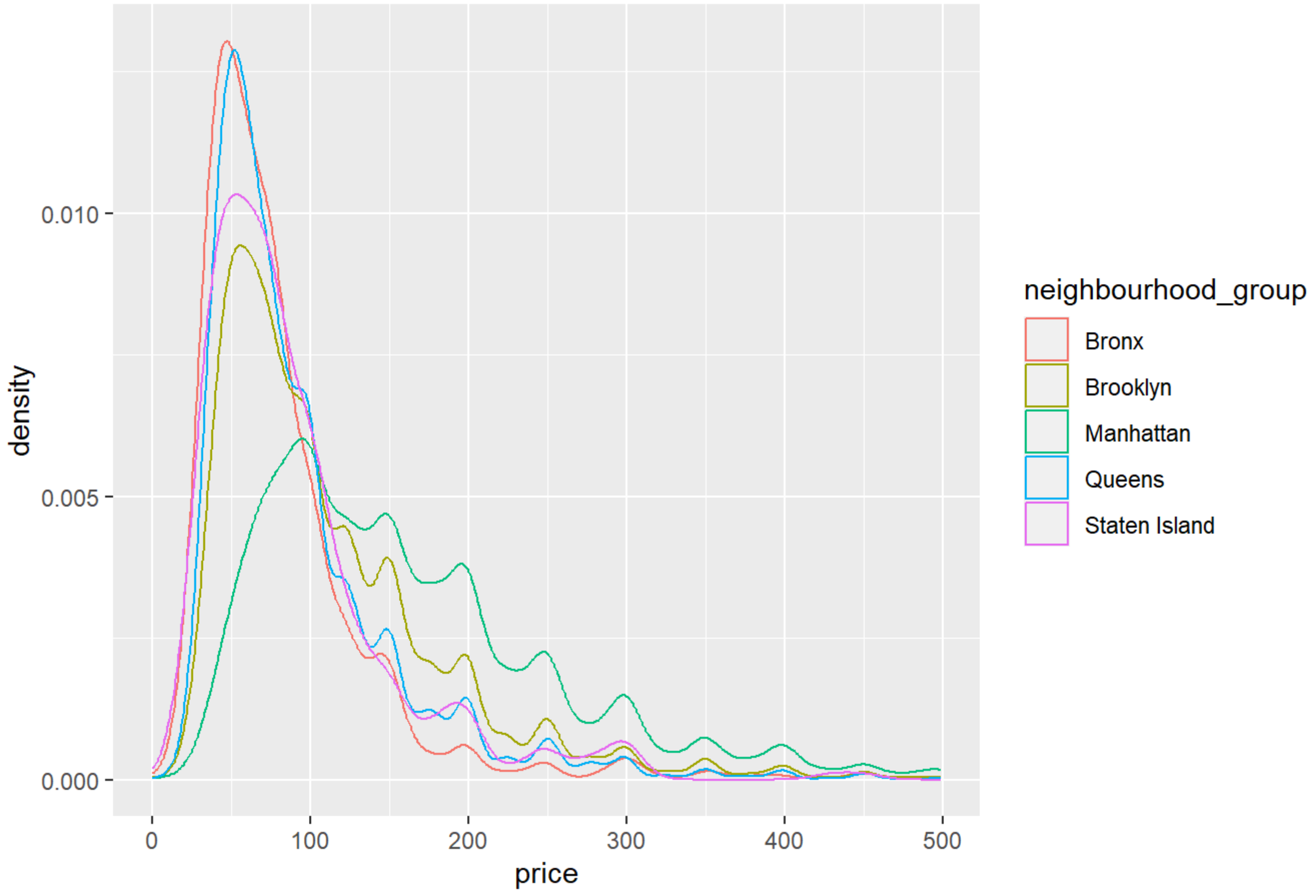


Figure :9 Map Showing the three different divisions of Manhattan. (11)

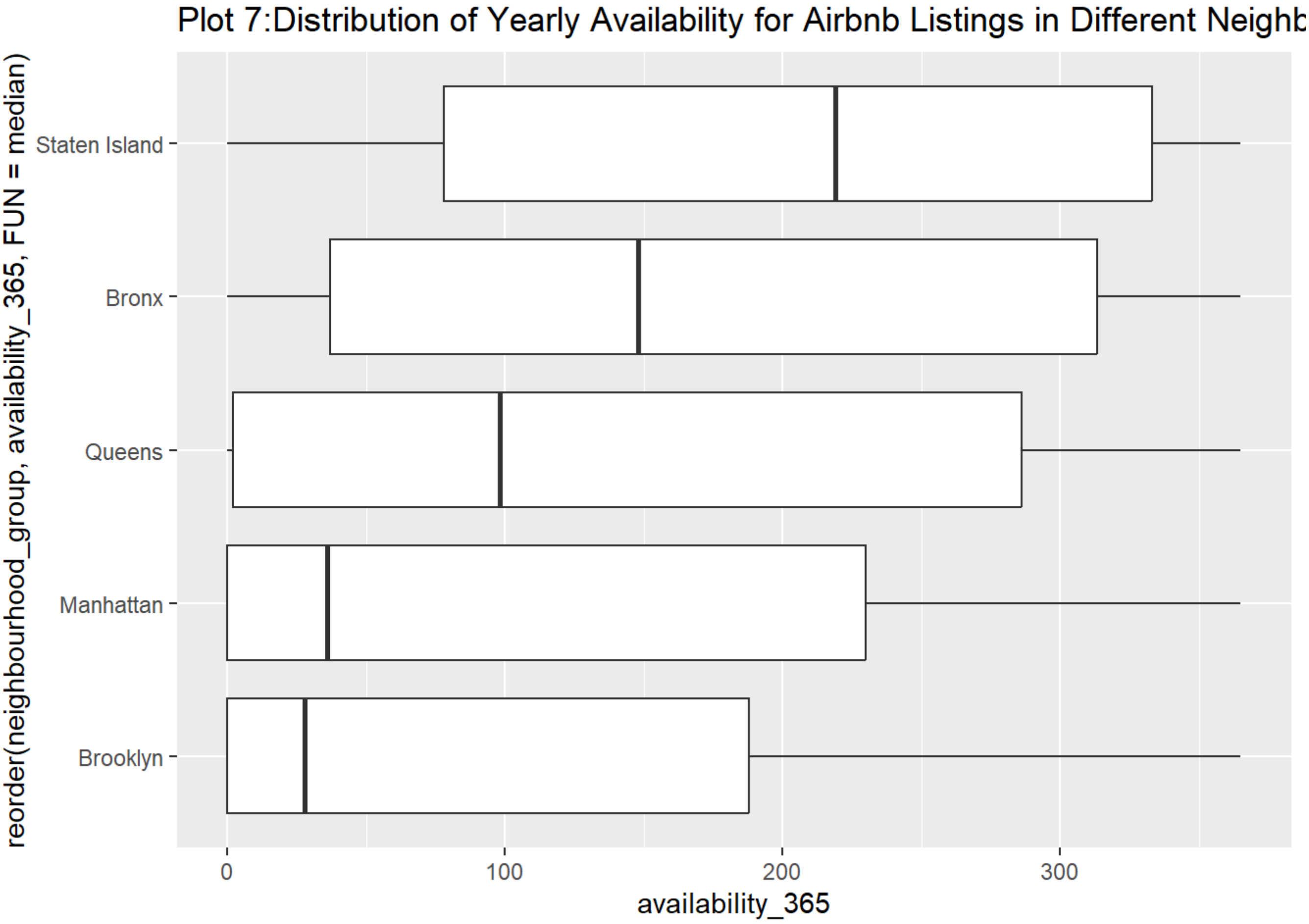
I was curious if the same would happen in Bronx just like in Brooklyn and Queens . But what I had found is that the number of airbnbs in places like Mott Haven and Port Morris which are close to Manhattan (Figure 7) are significant but places like Fordham and Kingsbridge are a bit more higher . To see why this is happening , we look into Figure 3, where we see that one of the famous attractions in NYC is the Bronx Zoo and if we look into Figure 8 we see that Fordham and Kingsbridge are very close to Bronx Zoo and thereby people tend to stay in these neighbourhoods. Also another reason why people would not want to stay in neighbourhoods in Bronx which are near to the Manhattan is that they are located more towards the upper Manhattan where there are very less number of attractions as compared to that in lower or middle Manhattan (Figure 3 and Figure 9) .But again why would people tend to stay in Brooklyn or Queens to go to Manhattan rather than staying right at Manhattan ?

2. Plots of Covariation of Variables

Plot 6:Price Density by Neighborhood Group for Airbnb Listings under $500



To further understand why people tend to stay in Queens or Brooklyn rather than staying in Manhattan , I looked into a density plot with price less than 500 against each neighbourhood group . From this I was able to understand that Manhattan has a higher density of prices between 100 and 200 which means that the price is relatively more expensive compared to the other boroughs. Queens and Brooklyn have similar density curves, which means that they have a similar distribution but there is a slight peak density for Queens indicating that listings in Queens tend to be slightly more affordable than Brooklyn. And this solves our dilemma about why people tend to stay in Brooklyn and Queens rather than directly staying in Manhattan because staying in Brooklyn and Queens is less expensive and probably traveling to Manhattan will be very cheap. But again if we look at the count of the number of airbnbs in Brooklyn and Queens , we see that the number of Airbnbs in Willamsburg is a bit less than 4000 units and in Astoria is only a bit more than 750 (Plot 2 and Plot 4)). This is because Astoria is connected to Midtown Manhattan and if we look at Figure 3 we see that there are less number of famous attractions in Midtown Manhattan and on the other hand Brooklyn is connected to lower Manhattan which contains most of the famous attractions (Figure 9).Bronx and Staten Island have relatively fewer listings compared to other neighborhood groups, and their density curves are lower than others. But if we look back again at Figure 3 we see that there is barely any attractions in Staten Island and there is one famous attraction in Bronx which is the Bronx Zoo, then why is the price in Staten Island almost similar to that of Bronx?



To answer this question we look at a box plot of the availability of Airbnb listings for each neighbourhood group, where the x-axis represents the neighbourhood group and the y-axis represents the availability in days. Here we see that Staten Island has a very high availability throughout the year compared to that of the other boroughs. This solves our problem as we can identify from Plot 1 that the distribution in Staten Island is very low, this in turn means that the supply of Airbnbs in Staten Island is less and when there is a very low supply , there will surely be a high demand and thereby prices will increase. From this we are able to understand why prices in Staten Island is quite similar to that in Bronx.

Questions:

1. How many Airbnb hosts in New York City have multiple listings?

2. How has the number of Airbnb listings in New York City changed over time?
3. What is the average number of reviews for Airbnb properties in New York City?
4. Are there any seasonal patterns in Airbnb booking patterns in New York City?
5. What is the average length of stay for Airbnb guests in New York City?

References:

(1)https://nycmap360.com/nyc-boroughs-map (https://nycmap360.com/nyc-boroughs-map)

(2)https://www.nps.gov/stli/planyourvisit/directions.htm (https://www.nps.gov/stli/planyourvisit/directions.htm)

(3)https://attractionsmagazine.com/top-attractions-new-york-city-featured-new-map/ (https://attractionsmagazine.com/top-attractions-new-york-city-featured-new-map/)

(4)https://artsandculture.google.com/entity/times-square/m07qdr?hl=en (https://artsandculture.google.com/entity/times-square/m07qdr?hl=en)

(5)https://www.mashvisor.com/blog/how-to-find-the-best-area-for-airbnb-investment/ (https://www.mashvisor.com/blog/how-to-find-the-best-area-for-airbnb-investment/)

(6)https://bklyndesigns.com/map-of-brooklyn-neighborhoods/ (https://bklyndesigns.com/map-of-brooklyn-neighborhoods/)

(7)https://freetoursbyfoot.com/new-york-city-neighborhoods/ (https://freetoursbyfoot.com/new-york-city-neighborhoods/)

(8)https://www.shutterstock.com/image-vector/new-york-districts-queens-207331579 (https://www.shutterstock.com/image-vector/new-york-districts-queens-207331579)

(9)https://www.shutterstock.com/image-vector/new-york-districts-bronx-map-207331567 (https://www.shutterstock.com/image-vector/new-york-districts-bronx-map-207331567)

(10)https://www.pinterest.com/pin/26880929000419030/ (https://www.pinterest.com/pin/26880929000419030/)

(11)https://patch.com/new-york/upper-west-side-nyc/uws-merged-ues-new-congressional-map-see-it (https://patch.com/new-york/upper-west-side-nyc/uws-merged-ues-new-congressional-map-see-it)