A REPORT

ON

Airbnb Host Listings of New York: Analysis and Recommendations

BY

Adarsha Mani

Prepared in complete fulfillment of

Capstone Project
APPLIED DATA SCIENCE CAPSTONE

COURSERA

(April 2020)

INTRODUCTION

Manhattan is considered to be one of the world's foremost commercial, cultural, and financial centers. Manhattan, thus, has the strongest tourism in all neighborhoods of New York. Tourists from all over the place have the options nowadays to book hotel rooms, hostels, and Airbnb. Since Airbnb has subsidized prices and is very practical for people who come to visit New York with their families, or in a big group of friends. Airbnb offers them a variety of choices in terms of types of accommodation: shared rooms, private rooms, hotel rooms, or entire homes/apartments. Airbnb also enables customers to customize their lifestyle and expenses on food according to their choices during their stay in the city.

OBJECTIVE 1: Airbnb is restricted to some neighborhoods in New York. The absence of Airbnb host listings in a lot of neighborhoods can hit the Airbnb business when tourism booms a few months post COVID-19. Airbnb can advertise asking citizens to host their listings in locations Airbnb hasn't expanded yet. The project will visualize the locations where Airbnb exists and where it can expand.

OBJECTIVE 2: Airbnb provides filters for potential clients to search for an appropriate host listing. The project aims to better the filter experience by listing the utilities available around the host listings, so the client can choose from the listings according to his preferences.
For example, some of the filters that can be mentioned by the client are:

- Located in Midtown Manhattan of New York
- Should be available any time of the year
- Should definitely be reviewed by some previous customers
- Listings should not have been last reviewed pre-2018
- The Airbnb will be used for a minimum of 10 days
- Visualize the unique room types available
- Filter the room type to the one that occurs at the highest frequency above
- Find the distribution of prices in the list of suggestions now

- Cost for the Airbnb should be up to 200 dollars
- List the most common utilities available around the listings

Stakeholders:

1. Potential renters (tourists)
2. Potential hosts for Airbnb
3. Airbnb media team (for advertisements)

DATA

Following sources of data were used while working on this Capstone Project:

1. Airbnb Host Listings Dataset

Name of the file: listings

Type of the file: CSV

Date Compiled: 8 April 2020

Description: The dataset contains summary information and metrics for listings in New York City.

The headers of the file are as follows:

id:	ID of the listing
name:	Name of the listing
host_id:	ID of the host
host_name:	Name of the host
neighborhood_group:	Borough of New York City
neighborhood:	Neighborhood in the Borough
latitude:	Latitude coordinate of the listing
longitude:	Longitude coordinate of the listing
room_type:	Type of room that is available for a particular listing

price:	Price of the listing
minimum_nights:	Minimum nights that listing has to be booked for
number_of_reviews:	Number of reviews the listing has
last_review:	Date it was last reviewed on
reviews_per_month:	Reviews that were given per month for a particular listing
calculated_host_listings_count:	Number of listings a particular host has
availability_365:	Number of days in a year the listing is available for Airbnb to host a customer

Source: Open data available on '**Inside Airbnb: Adding data to the debate**'

URL: http://insideairbnb.com/get-the-data.html

2. Foursquare API

Type: JSON

Date Retrieved: 3 April 2020 (In the last run of the code)

Description: The data was extracted through API calls using a separate Python code. The proximities of various amenities were determined on the basis of the client's requirements.

Data that was especially taken out of the JSON file is as follows:

Venue:	The name of the facility in the area around the listing
Venue Latitude:	Latitude coordinate of the facility
Venue Longitude:	Longitude coordinate of the facility
Venue Category:	Category of venues in which the facility lies

Source: Account on FourSquare has already been made before and the username and password were used for requesting API calls.

URL: https://foursquare.com/

METHODOLOGY

1. Dataset

1.1 New York Data

The dataset in JSON form was loaded and the features were extracted to a DataFrame. This DataFrame had the list of neighborhoods present in the Boroughs of New York along with their coordinates (latitudes and longitudes). All of these neighborhoods were then visualized on the map of New York City.

1.2 Airbnb Host Listings Dataset

The dataset in CSV form was loaded to a DataFrame. The total number of listings in New York City was calculated. The total number of listings in each borough were calculated.

2. Objective 1: Find out which Neighbourhoods in the Boroughs have no Airbnb hosts

Every specific borough was dealt with separately. All the neighborhoods that have Airbnb hosts were listed first and then, the neighborhoods that did not have hosts were found out. All this data was stored in two separate DataFrames which were then used to visualize on the map.

The percentage of neighborhoods that do not have Airbnb hosts in each borough was thus calculated.

3. Objective 2: Client Preferences

The following example contains preferences that can be used to create filters for any client. The first 9 filters were created through Basic Exploratory Analysis techniques.

1. Location: Midtown Manhattan of New York
2. Availability: Any time of the year
3. Reviews: >=1
4. Reviews: Not reviewed pre-2018
5. Minimum Nights: 10 days
6. Visualization: Unique room types available

7. Room Type: Room that occurs at the highest frequency
8. Prices: Visualize the distribution of prices
9. Prices: <= $200
10. Venues/Utilities: Most common utilities around the listings

The last filter required the usage of Foursquare API. Of all the venues returned by Foursquare API, hotels were removed as the client already has a residence in the Airbnb host listings.
The listings, thus, found out were clustered using K means Clustering. Silhouette Coefficient is used to find out the most optimal number of clusters.

$$\text{Silhouette Coefficient} = (x-y)/ \max(x,y)$$

Where,
y is the mean intracluster distance: mean distance to the other instances in the same cluster.
x depicts mean nearest cluster distance i.e. mean distance to the instances of the next closest cluster.
The coefficient varies between -1 and 1. A value close to 1 implies that the instance is close to its cluster is a part of the right cluster. Whereas, a value close to -1 means that the value is assigned to the wrong cluster.

RESULTS

1. New York has 5 boroughs and 306 neighborhoods.
2. The count of neighborhoods in each borough of New York:

Borough	Neighborhood	Latitude	Longitude
Bronx	52	52	52
Brooklyn	70	70	70
Manhattan	40	40	40
Queens	81	81	81
Staten Island	63	63	63

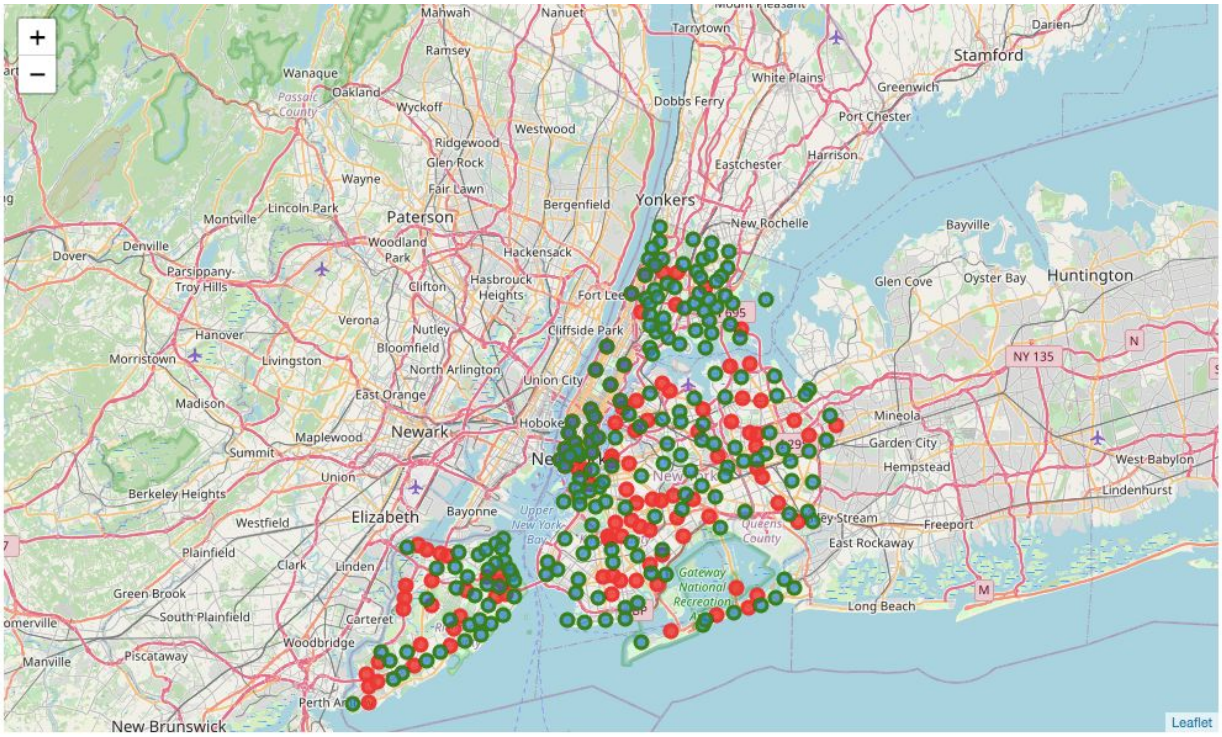
3. New York City has 50796 Airbnb listings spread all over 5 boroughs.

4. The count of Airbnb listings in each borough:

Borough	ID
Bronx	1249
Brooklyn	20455
Manhattan	22448
Queens	6266
Staten Island	378

(ID here is actually the count of listings in each borough)

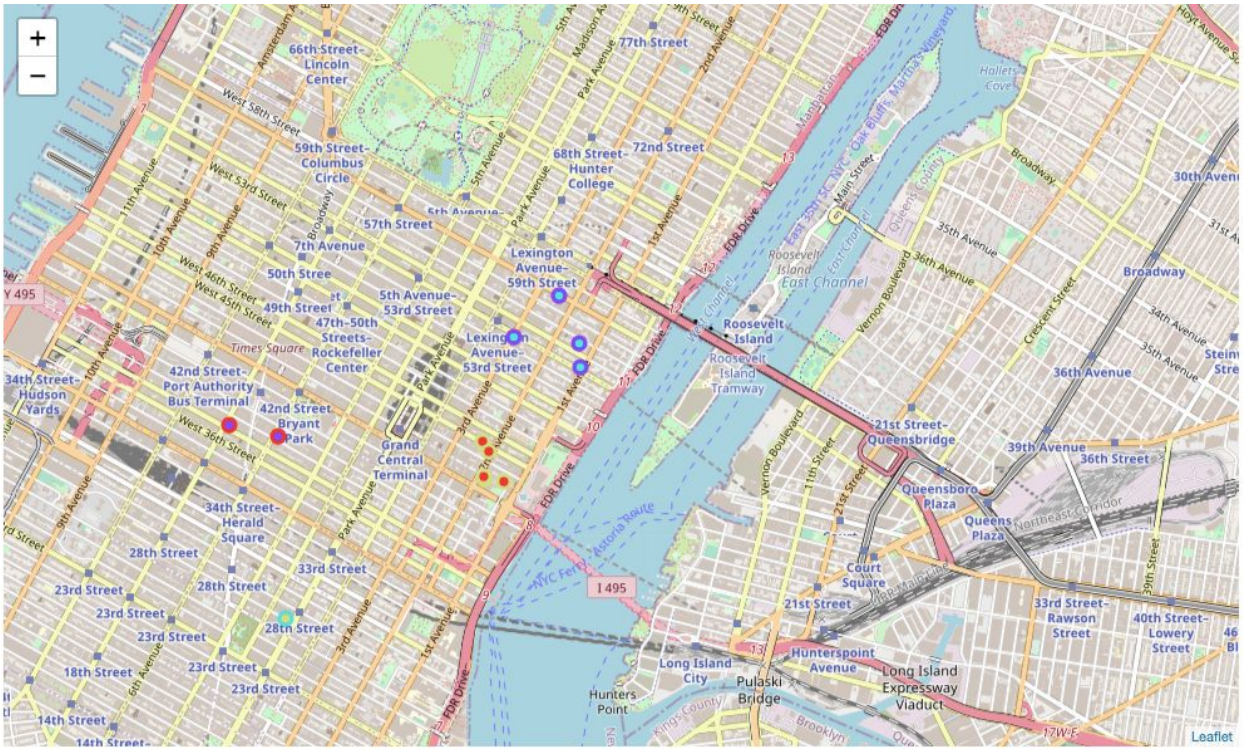
5. 108 neighborhoods out of 306 neighborhoods in New York City do not have Airbnb hosts.



The red markers are the neighborhoods that do not have Airbnb hosts and the green markers are the neighborhoods that have Airbnb hosts.

Borough	Neighborhood With No Airbnb	Neighborhood With Airbnb	Total Neighborhoods	Percentage of Neighborhoods with no Airbnb
Bronx	6	46	52	11.538462
Brooklyn	29	41	70	41.428571
Manhattan	23	23	46	50.000000
Queens	31	50	81	38.271605
Staten Island	25	38	63	39.682540

6. Host listings available for rent after all the applied filters: 12



7. The host listings were clustered into 4 clusters that have been described as follows:

a. **Cluster 1:** The most common venues available around the host listings include Coffee Shops, Japanese Restaurants, and American Restaurants.

b. **Cluster 2:** The most common venues available around the host listings include Theaters, Korean Restaurants, and Burger Joints.

c. **Cluster 3:** The most common venues available around the host listings include Coffee Shops, Gym/Fitness Centres, and American Restaurants.

d. **Cluster 4:** The most common venues available around the host listings include Korean Restaurants, Gym/Fitness Centres, and Indian Restaurants.

DISCUSSION

There are 108 neighborhoods of New York City that do not have Airbnb hosts. Given that the tourism of this city has been booming, it will be beneficial for the company if it can attract hosts to rent their houses/apartments in the other neighborhoods as well.

Filtering System for Airbnb has been bettered by adding another facility to it by presenting a list of venues that are available around each of the listings. This can add to customer experience as the customer can make a better pick out of his available options.

CONCLUSION

Further work on the expansion of the Airbnb network will include finding out the locations in New York that see the most traction from the tourists. If these locations do not yet have Airbnb hosts, the company will have to take a targetted approach to get new hosts.