***"The best Airbnb entire home/apartment in Sheepshead Bay, Brooklyn - NYC"***

Matteo Vadi – December 28, 2020

# 1. INTRODUCTION

## 1.1 BACKGROUND

All of us have to look for a place to stay for a longer or shorter time period, at least once in our lives. And, in those situations, sooner or later we have to face with the choice between two or more available alternatives, without knowing which to choose having little information about the place we are going to visit.
This work accepts the challenge and tries to give an answer to the following question: which of these apartments is better to choose? This document specifically focuses on rental entire homes / apartments in the Airbnb platform for the neighborhood of Sheepshead Bay within the borough of Brooklyn, New York city, but can be extended to other neighborhoods, boroughs, cities and/or rental properties.
Airbnb, Inc. is an American vacation rental online marketplace company based in San Francisco, California US. It allows to host and rent different properties, accessible to consumers on its website or via an App. Through the service, users can arrange lodging, primarily homestays, and tourism experiences or list their properties for rental.

## 1.2 PROBLEM

This project aims to determine which entire home / apartment in the Airbnb platform is the best one to choose when deciding for a place to stay in Sheepshead Bay, Brooklyn NYC, in terms of venues and points of interest you can find in the nearby area. In particular, I have decided to focus only on the most important types of venues for an accurate decision in my personal opinion, within a radius of 600 meters (an appropriate distance that can be covered by foot) from the place, that you can find listed below:

- **Travel & Transport**:
  - Metro station
  - Tram station
  - Bus station
  - Taxi
  - Taxi stand
  - Rental car location
- **Food**
- **Professional & Other Places**:
  - Parking
  - Medical Center:
    - Emergency room
    - Hospital

- Shop & Services:
    - Market
    - ATM
    - Laundromat
    - Dry cleaner
    - Internet Café
    - Auto Garage
    - Food & Drink Shop
    - Shopping Mall
    - Shopping Plaza
    - Pharmacy

These venues categories and sub-categories refer to the differentiation you can find in the Foursquare Developers webpage (https://developer.foursquare.com/docs/build-with-foursquare/categories/ ).
The list introduced above is considering the important venues in choosing a place to stay whatever the reason of the journey is, in case of shorter time period for tourism visits or longer ones for work transfers. Indeed, this work does not consider Arts & Entertainment, College & Universities, Events or Outdoors & Recreation venues among others, since those depend on the specific purpose of the journey.
This project deals with **Foursquare APIs** to explore the geographical location of NYC and to cluster apartments in Brooklyn borough (**k-means clustering algorithm**), in order to find the most suitable home / apartment to choose among all. In fact, the best apartment according to this project work will be the one which will present the higher number of these kinds of venues in the nearby area. This information can be then used in choosing the apartment together with other common aspects, such as availability, price and minimum overnight stays.

## 1.3 INTEREST

The interest that such a project may arouse could encompass several stakeholders. First of all, surely the final user of the Airbnb platform or, generally speaking, all the people that travel with more or less frequency, allowing them to have a different evaluation tool from the common reviews/feedbacks on websites.
Similarly, this work could also arouse interest in those individuals who are on the other side and who are thinking of buying an apartment for renting it, allowing them to find the best area of a city (a specific neighborhood of New York City in this case) where to buy the property.

# 2. DATASET

## 2.1 DATA SOURCES

The original dataset describes the listing activity and metrics in NYC for 2019. You can find it to the Kaggle website (https://www.kaggle.com/dgomonov/new-york-city-airbnb-open-data); in particular, it is part of Airbnb open data and the original source can be found on its website (http://insideairbnb.com/). The dataset includes all needed information to find out more about hosts, geographical availability, necessary metrics to make predictions and draw conclusions. For the aim of this project work, it will focus only on the division between *neighborhood* and *neighborhood_groups* (actually, the NYC's boroughs) for the Foursquare API identification.

## 2.2 FEATURES SELECTION

*Figure 1. The original dataframe with the .head() function.*

```
The df dataframe shows 48895 records (rows) and 16 attributes (columns)
```

| | id | name | host_id | host_name | neighbourhood_group | neighbourhood | latitude | longitude | room_type | price | minimum_nights | number_of_reviews | last_review | reviews_per_month | calculated_host_listings_count | availability_365 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 2539 | Clean & quiet apt home by the park | 2787 | John | Brooklyn | Kensington | 40.64749 | -73.97237 | Private room | 149 | 1 | 9 | 2018-10-19 | 0.21 | 6 | 365 |
| **1** | 2595 | Skylit Midtown Castle | 2845 | Jennifer | Manhattan | Midtown | 40.75362 | -73.98377 | Entire home/apt | 225 | 1 | 45 | 2019-05-21 | 0.38 | 2 | 355 |
| **2** | 3647 | THE VILLAGE OF HARLEM....NEW YORK ! | 4632 | Elisabeth | Manhattan | Harlem | 40.80902 | -73.94190 | Private room | 150 | 3 | 0 | NaN | NaN | 1 | 365 |
| **3** | 3831 | Cozy Entire Floor of Brownstone | 4869 | LisaRoxanne | Brooklyn | Clinton Hill | 40.68514 | -73.95976 | Entire home/apt | 89 | 1 | 270 | 2019-07-05 | 4.64 | 1 | 194 |
| **4** | 5022 | Entire Apt: Spacious Studio/Loft by central park | 7192 | Laura | Manhattan | East Harlem | 40.79851 | -73.94399 | Entire home/apt | 80 | 10 | 9 | 2018-11-19 | 0.10 | 1 | 0 |

After a first quick view, the dataset imported in the form of csv document through the correspondent pandas library (*pd.DataFrame.read_csv*), shows 48895 records (rows) and 16 attributes (columns). Among these, we select only ones that concern the rental of entire houses / apartments, not considering private rooms or shared rooms, but the type of analysis that will be carried out could also be extended to the latter.

For the purpose of our problem, we do not take into account the attributes about the *id*, *host_name*, *host_id*, *price*, *minimum_nights*, *number_of_reviews*, *last_reviews*, *reviews_per_month*, *calculated_host_listings_count* and *availability_365*. This types of attributes could be used in a supervised Data Mining problem (for example, if the goal was the prediction of a rental place's price), while for the aim of our clustering approach they would only be redundant. For this reason we consider the following attributes:

- **name**: the name of the entire home / apartment;
- **neighbourhood_group:** actually, the borough's name the entire home/apartment belongs to;
- **neighbourhood**: the neighborhood's name the entire home/apartment belongs to;
- **latitude**: the latitude coordinate for the entire home / apartment;
- **longitude**: the longitude coordinate for the entire home / apartment.

To have a better comprehension, the ***df*** dataset has been renamed in ***airbnb***, changing also 2 columns names. The dataset, after these changes appears as in Figure 2.

*Figure 2. The Airbnb dataset after features selection with the .head() function.*

```
The airbnb dataframe now shows 25409 records (rows) and 5 attributes (columns)
```

| | name | borough | neighborhood | latitude | longitude |
|---|---|---|---|---|---|
| **1** | Skylit Midtown Castle | Manhattan | Midtown | 40.75362 | -73.98377 |
| **3** | Cozy Entire Floor of Brownstone | Brooklyn | Clinton Hill | 40.68514 | -73.95976 |
| **4** | Entire Apt: Spacious Studio/Loft by central park | Manhattan | East Harlem | 40.79851 | -73.94399 |
| **5** | Large Cozy 1 BR Apartment In Midtown East | Manhattan | Murray Hill | 40.74767 | -73.97500 |
| **9** | Cute & Cozy Lower East Side 1 bdrm | Manhattan | Chinatown | 40.71344 | -73.99037 |

## 2.3 DATA CLEANING

After the feature selection, the *airbnb* dataset shows 25409 rows and 5 columns.
This section is partially aimed to deal with missing values in the dataset. After a quick view (Figure 3), it is possible to see that there are very few of them and all connected to the *name* attribute.

*Figure 3. Missing values in the Airbnb dataset.*

```
Out[7]: name            7
        borough         0
        neighborhood    0
        latitude        0
        longitude       0
        dtype: int64
```

Hence, we can easily remove them, by dropping the entire records with missing values. In fact, since all of them belong to the '*name*' attribute, without the name is impossible to consider them as good candidates to be the best rental place in our analysis.

*Figure 4. The airbnb dataset after having removed missing values .*

```
The airbnb dataframe shows 25402 records (rows) and 5 attributes (columns)
```

| | name | borough | neighborhood | latitude | longitude |
|---|---|---|---|---|---|
| **0** | Skylit Midtown Castle | Manhattan | Midtown | 40.75362 | -73.98377 |
| **1** | Cozy Entire Floor of Brownstone | Brooklyn | Clinton Hill | 40.68514 | -73.95976 |
| **2** | Entire Apt: Spacious Studio/Loft by central park | Manhattan | East Harlem | 40.79851 | -73.94399 |
| **3** | Large Cozy 1 BR Apartment In Midtown East | Manhattan | Murray Hill | 40.74767 | -73.97500 |
| **4** | Cute & Cozy Lower East Side 1 bdrm | Manhattan | Chinatown | 40.71344 | -73.99037 |

Then, for the goals of the following sections, let’s build other 2 different datasets starting from the *airbnb* one. One of them, named *sheepshead_bay* dataset, will be the main for the further operations.
At the end of this section, we have 3 different cleaned dataset that will be used for different aims:
- ***airbnb***, for visualizing apartments location spread in NYC;
- ***brooklyn***, with apartments in the borough of Brooklyn (Figure 5);
- ***sheepshead_bay***, the mainly used dataset about a single neighborhood in the Brooklyn borough (Figure 6).

*Figure 5. The brooklyn dataset with the .head() function.*

```
The brooklyn dataframe shows 9558 records (rows) and 5 attributes (columns)
```

| | name | borough | neighborhood | latitude | longitude |
|---|---|---|---|---|---|
| 0 | Cozy Entire Floor of Brownstone | Brooklyn | Clinton Hill | 40.68514 | -73.95976 |
| 1 | Only 2 stops to Manhattan studio | Brooklyn | Williamsburg | 40.70837 | -73.95352 |
| 2 | Perfect for Your Parents + Garden | Brooklyn | Fort Greene | 40.69169 | -73.97185 |
| 3 | Hip Historic Brownstone Apartment with Backyard | Brooklyn | Crown Heights | 40.67592 | -73.94694 |
| 4 | Sweet and Spacious Brooklyn Loft | Brooklyn | Williamsburg | 40.71842 | -73.95718 |

*Figure 6. The sheepshead_bay dataset with the .head() function.*

```
The Sheepshead Bay dataframe shows 59 records (rows) and 5 attributes (columns)
```

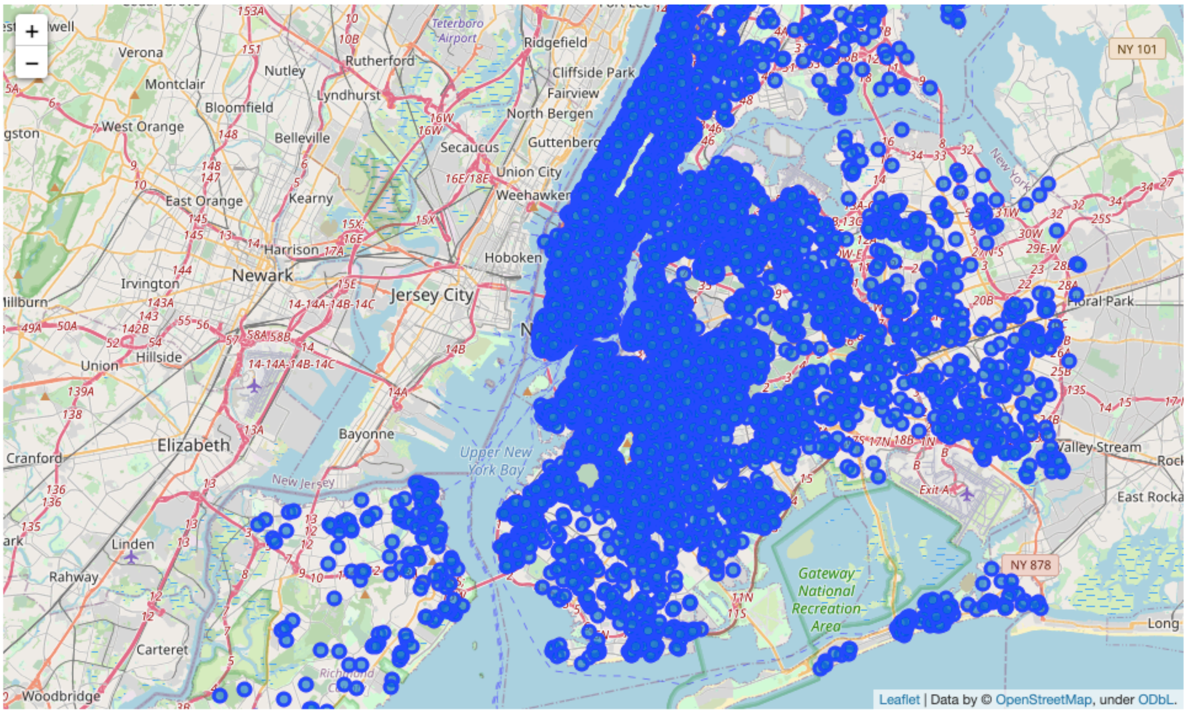
| | name | borough | neighborhood | latitude | longitude |
|---|---|---|---|---|---|
| 0 | Bright Modern Charming Housebarge | Brooklyn | Sheepshead Bay | 40.58422 | -73.94079 |
| 1 | Luxury L-Shape Studio + 3 cats | Brooklyn | Sheepshead Bay | 40.59721 | -73.95149 |
| 2 | Charming Housebarge w/ outside deck | Brooklyn | Sheepshead Bay | 40.58408 | -73.94122 |
| 3 | Great studio with 2 rooms and kitchen | Brooklyn | Sheepshead Bay | 40.58426 | -73.95949 |
| 4 | A Dream! Luxury 3 Bedroom Apt+Pking | Brooklyn | Sheepshead Bay | 40.58527 | -73.93534 |

# 3. METHODOLOGY

## 3.1 EXPLORING ALL THE APARTMENTS IN THE CITY

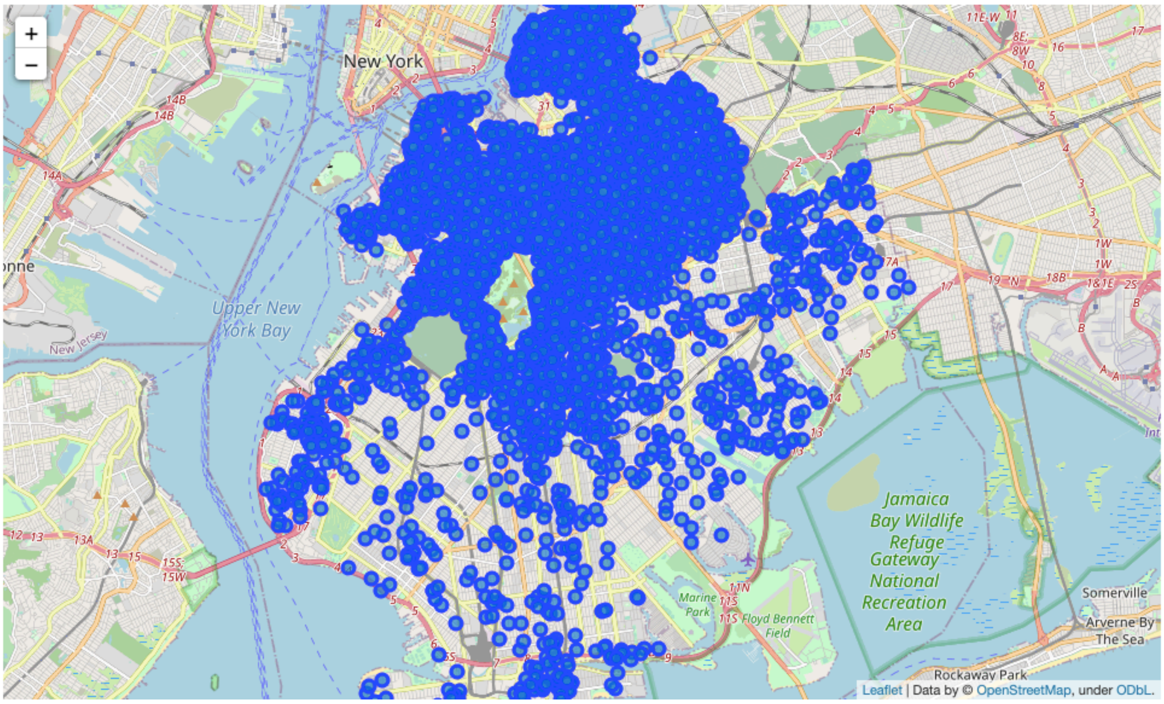
First we will explore all the apartments in NYC, just to be sure to have no insights directly from just a picture.
For this aim we will use the *GeoPy* and the *Folium* libraries. With *GeoPy* is possible to extract the geo-coordinates (latitude and longitude) of a place in the world; in particular it is defined as “a Python client for several popular geocoding web services” (from the definition on the official website: https://geopy.readthedocs.io/en/stable/ ).
With the *Folium* library instead you can build great interactive *Leaflet maps* in Python with markers superimposed on it or choropleth ones, too (see more at: https://pypi.org/project/folium/ ).
The first step has been the creation of a *folium* map of NYC with all the 25402 different apartments in the *airbnb* dataset, in order to have a first look at the location spread of the apartments in all the city. The result is in Figure 7.

*Figure 7. Airbnb rental homes/apartments’ location spread in NYC.*




From this very rough visualization is impossible to say anything; even if we take only entire home / apartments, without considering private or shared rooms, there are too many of them for having a first insight about their location spread.
This is why our analysis will focus only on a smaller dataset connected to a smaller region, the neighborhood of Sheepshead Bay within Brooklyn. So the second step has been a *folium* map with all rental entire homes/apartments in the borough of Brooklyn (Figure 8, starting from *brooklyn* dataset and its 9558 records).

*Figure 8. Airbnb rental homes/apartments' location spread in Brooklyn, NYC.*



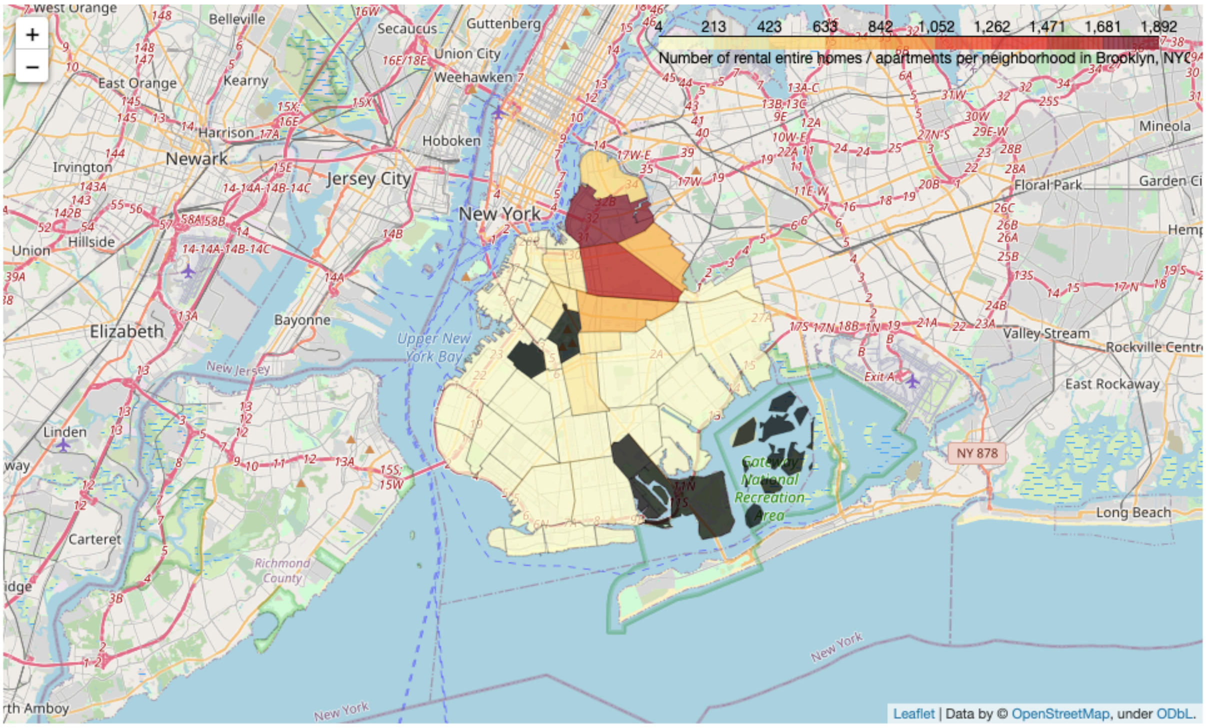

This seems to be more interesting. The higher number of rental entire homes/apartments in Airbnb platform within the borough of Brooklyn NYC, are located on the Northern part. This is probably due to the proximity with the city center and with the fact that actually Brooklyn is one of the most populated borough in New York City, with an high number of ethnic minorities (for example, consider that in the Southern part of Brooklyn there is one of the biggest Chinatown in the world, based on Sunset Park), that can prevent the presence of tourist rental places.

To have a deeper understanding of this fact, I have built a ***choropleth map*** visualizing the frequency of rental entire homes / apartments with respect to each neighborhood in Brooklyn borough. For such a type of map, the *Folium* library requires a *geoJSON* file explaining the boundaries and limit of the area under analysis; in our case, we will need a *geoJSON* file with the boundaries of all the neighborhoods in the borough of Brooklyn (playing a little bit with the map displayed, you can obtain it here: [https://geojson.io/#map=0/0/0](https://geojson.io/#map=0/0/0) ). You can find the output map in Figure 9.

The *choropleth map* confirmed what I was supposing, that is the higher frequency of rental homes/apartments is in the Northern part. The dark zones in the map are actually neighborhoods in which there are no places to stay, since there is a park or something different, as the Green-Wood Cemetery.

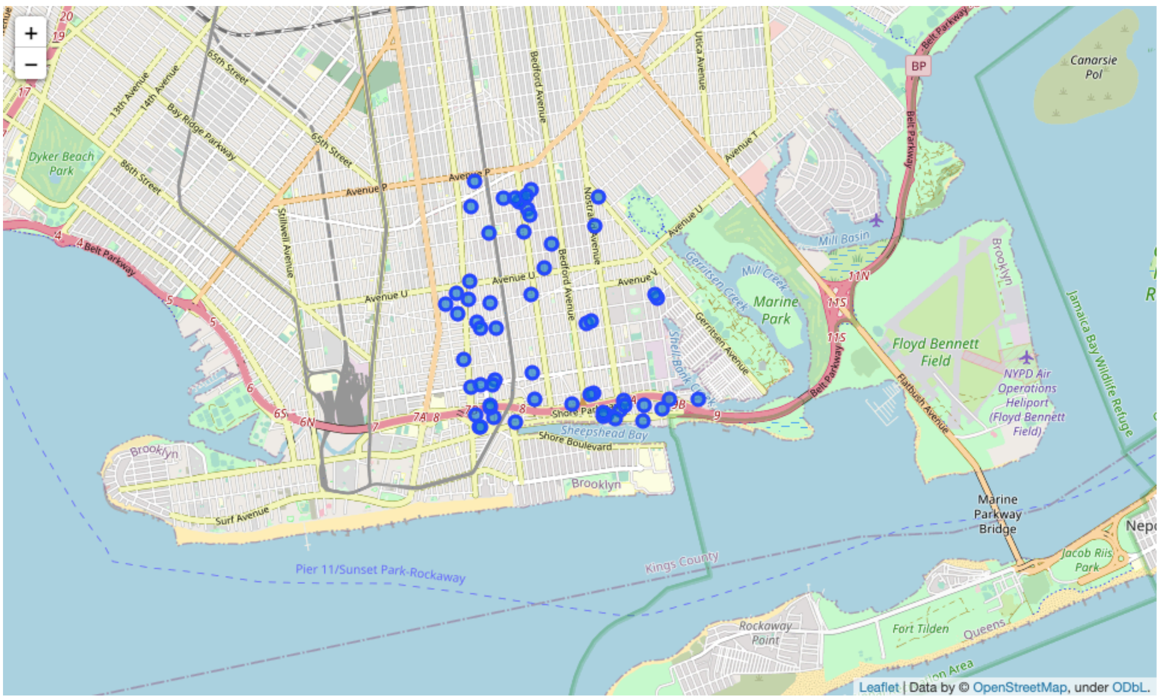
This is why we will deal with a neighborhood in the Southern part of the Brooklyn borough, as it is less difficult to find the best rental entire home / apartment in this area.

*Figure 9. The choropleth map displaying the frequency of the number of rental entire homes/apartments in the borough of Brooklyn, NYC.*




I decided to focus on the **Sheepshead Bay** as the objective neighborhood and the last step in this section has been the realization of a *Folium* map with all the rental entire homes/apartments (59 records of *sheepshead_bay* dataset) superimposed on it (Figure 10).

*Figure 10. Airbnb rental homes/apartments' location spread in Sheepshead Bay, Brooklyn NYC.*



## 3.2 EXPLORING VENUES IN THE APARTMENTS' NEARBY AREA

In this section, we will deal with some kind of exploratory analysis, in order to understand a little bit more the nearby area of each apartment within the neighborhood of Sheepshead Bay in Brooklyn. To go more in-depth, I decided to proceed exploring a single Airbnb home/apartment in Sheepshead Bay and then repeating the procedure for all of them. It is important to notice that for both approaches we will explore a radius of 600 meters from the location of the rental place, since this can be a good distance to be covered by foot with no problem.

### 3.2.1 Exploring a single Airbnb apartment in Sheepshead Bay

For this first part, we will deal with the very first apartment you can find in the *sheepshead_bay* dataset, called ***Bright Modern Charming Housebarge***. With a defined function, named ***get_venues***, is possible to retrieve all the venues connected to one of the 4 specific "root" categories (*Food, Travel & Transport, Shop & Services, Professional & Other Places*) defined at the beginning of the project; repeating the function for all the 4 "root" categories and filtering those which require it (all of them except for the *Food* one), we can build a dataset with all the venues of interest for such a place under analysis. The dataset for this apartment is shown in Figure 11.

*Figure 11. The bright_df showing all the venues of interest for the very first apartment in the sheepshead_bay dataset.*

| | name | lat | lng | categories | sub_category |
|---|---|---|---|---|---|
| 0 | Roll N Roaster | 40.584143 | -73.939555 | 4bf58dd8d48988d1c5941735 | Sandwich Place |
| 1 | Liman Restaurant | 40.583368 | -73.941053 | 4f04af1f2fb6e1c99f3db0bb | Turkish Restaurant |
| 2 | Opera Cafe Lounge | 40.583842 | -73.944481 | 4f04af1f2fb6e1c99f3db0bb | Turkish Restaurant |
| 3 | Il Fornetto | 40.583378 | -73.939487 | 4bf58dd8d48988d110941735 | Italian Restaurant |
| 4 | Rocca Cafe & Lounge | 40.583371 | -73.940841 | 4f04af1f2fb6e1c99f3db0bb | Turkish Restaurant |
| 5 | Cats on the Bay | 40.583824 | -73.946849 | 4bf58dd8d48988d1c4941735 | Restaurant |
| 6 | Randazzo's Clam Bar | 40.583832 | -73.947583 | 4bf58dd8d48988d1ce941735 | Seafood Restaurant |
| 7 | Applebee's Grill + Bar | 40.583888 | -73.943235 | 4bf58dd8d48988d14e941735 | American Restaurant |
| 8 | Maria's Ristorante Italiano | 40.584205 | -73.936825 | 4bf58dd8d48988d110941735 | Italian Restaurant |
| 9 | Signature Restaurant | 40.583817 | -73.947776 | 4bf58dd8d48988d1c4941735 | Restaurant |
| 10 | Dunkin' | 40.586017 | -73.936737 | 4bf58dd8d48988d148941735 | Donut Shop |
| 11 | Passage | 40.583739 | -73.947066 | 5293a7563cf9994f4e043a44 | Russian Restaurant |
| 12 | Seaport Buffet | 40.583819 | -73.947188 | 52e81612bcbc57f1066b79f4 | Buffet |
| 13 | Emmons Bagel | 40.583854 | -73.944670 | 4bf58dd8d48988d179941735 | Bagel Shop |
| 14 | Maria's | 40.584263 | -73.936714 | 4bf58dd8d48988d1ce941735 | Seafood Restaurant |
| 15 | Maria Ristorante | 40.584273 | -73.936425 | 4bf58dd8d48988d110941735 | Italian Restaurant |
| 16 | Marmaris Restaurant | 40.584214 | -73.936272 | 4bf58dd8d48988d1c0941735 | Mediterranean Restaurant |
| 17 | Jumpin Bean | 40.584124 | -73.936226 | 4bf58dd8d48988d1c1941735 | Mexican Restaurant |
| 18 | XO Creperie | 40.583790 | -73.947231 | 52e81612bcbc57f1066b79f2 | Creperie |
| 19 | Fil's Seafood Grill | 40.584290 | -73.935760 | 4bf58dd8d48988d1ce941735 | Seafood Restaurant |
| 20 | Chinatown Restaurant | 40.586529 | -73.945533 | 4bf58dd8d48988d145941735 | Chinese Restaurant |
| 21 | Bay Side Mini Mart | 40.584327 | -73.934831 | 4bf58dd8d48988d146941735 | Deli / Bodega |
| 22 | MTA NYCTA Bus at Nostand Ave & Shore Pkwy / Em... | 40.584838 | -73.938890 | 4bf58dd8d48988d1fe931735 | Bus Station |
| 23 | MTA B4, B44, BM3 (Emmons Ave/Nostrand Ave) | 40.584027 | -73.938333 | 4bf58dd8d48988d1fe931735 | Bus Station |
| 24 | B44 - Nostrand Ave/Shore Pkwy | 40.585827 | -73.938947 | 4bf58dd8d48988d1fe931735 | Bus Station |
| 25 | Greater New York Endoscpy | 40.584148 | -73.945717 | 4bf58dd8d48988d196941735 | Hospital |

You can actually see that for the "**Bright Modern Charming Housebarge**" apartment there are **26** important venues in the nearby area, within a radius of 600 meters.

Just some consideration about the *get_venues* function; it requires as inputs your Foursquare developer credentials, together with the Foursquare "root" category ID (*category_id*) and the geo-coordinates of the apartment you want to analyze (so it is important that you have the geo-coordinates of each apartment as in all the datasets of this project work; if you refer to the Airbnb open data platform, probably all of them have such a similar structure). With these inputs the function uses the Foursquare API calls to have all the venues connected to the defined *category_id*; playing a little bit with the flatten JSON file given back from Foursquare, the function gives as ouput a dataset with the all the venues for that "root" category (then, outside the function, you have to filter only ones that are of interest for the project work, by looking at the sub-category ID given for each record). Repeating the function for all the 4 "root" categories for the very first apartment in the *sheepshead_bay* dataset, and merging all the output you have the dataset above in Figure 11 (the *categories* attribute in this picture is actually a column with all the sub-categories ID; the column *sub_category* display instead the name-type of that sub-category).

### 3.2.2 Exploring all the Airbnb apartments in Sheepshead Bay

In this section, we will use the ***getNearbyVenues*** function to extract the venues of interest for all the rental entire home / apartments in the Sheepshead Bay (hence, repeating what's done before for all the records in the *sheepshead_bay* dataframe).
Some considerations about the function introduced before; this function utilizes the other *get_venues* function to repeat the process in a for loop for all the apartments in the *sheepshead_bay* dataset, extracting the venues of interest for all of them and storing them in a global dataset, called *dataframe,* and finally assigning it to a new global varible called *sheepshead_data*. Hence, in this variable you can find each venues of interest in the nearby area (600 meters) for each rental entire home/apartment in the Sheepshead Bay neighborhood in the borough of Brooklyn.

# 4. MODEL DEVELOPMENT

In this section we will deal with the Machine Learning model utilized for clustering apartments depending on the important venues within 600 meters from each rental place. I decided to use a ***k-means algorithm*** for doing so, splitting all the apartments in the *sheepshead_data* into 7 clusters (labelled from 0 up to 7) in a new global variable, called *sheepshead_grouped*. I opted for **7 clusters** on the basis of the number of important venues and according to the dimension of the dataset, so to have a clear differentiation between better and worse ones. The *k-means algorithm* has been chosen for its simplicity, with centroids that are found in an iterative way running the specific function within the scikit-learn library in Pyhton.

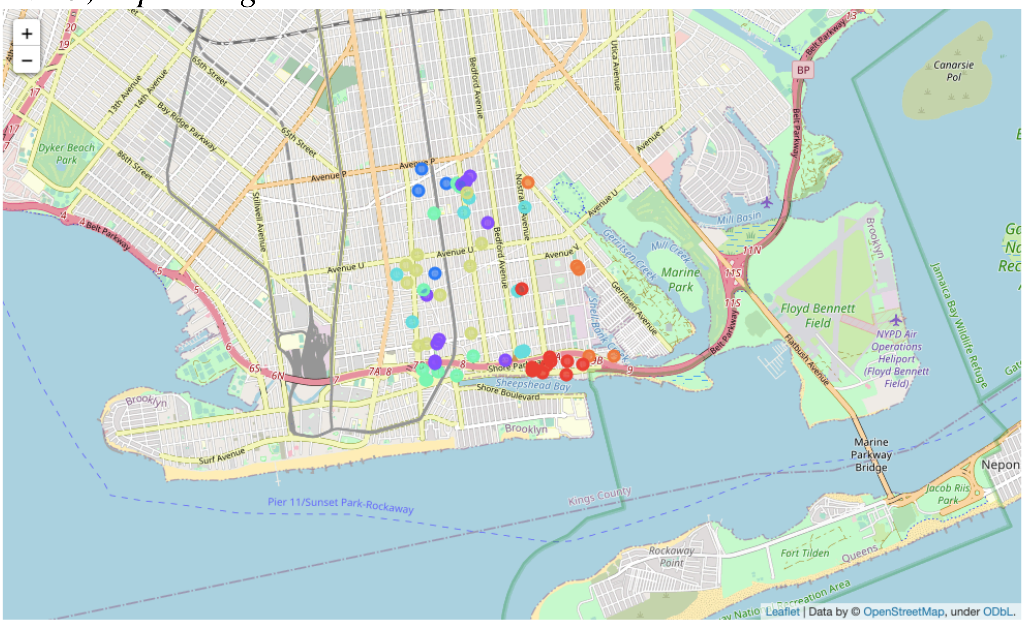
The results after the *k-means clustering* are then stored in a new object called *cluster_df,* which shows, after some manipulation, each apartment in the original *sheepshead_bay* dataset together with its geo-coordinates, the cluster it belongs to and the number of important venues in the nearby area (Figure 12).

Finally, a representation in a *Folium* map has been realized, showing the previous picture in Figure 10, with different colors depending on the clusters (Figure 13).

*Figure 12. The cluster_df dataset with the head method.*

| | name | latitude | longitude | Cluster Labels | number of important venues |
|---|---|---|---|---|---|
| 35 | 1 bd Apartment next to Subway | 40.58591 | -73.95099 | 4 | 66 |
| 4 | A Dream! Luxury 3 Bedroom Apt+Pking | 40.58527 | -73.93534 | 0 | 27 |
| 48 | A beautiful huge quiet Brooklyn apartment. | 40.60010 | -73.94974 | 5 | 52 |
| 23 | Amazing apartment in NYC | 40.58721 | -73.96018 | 5 | 47 |
| 57 | Beautiful Sheepshead Bay experience. | 40.58387 | -73.93957 | 0 | 25 |
| 47 | Beautiful apartment near the city and the beach!! | 40.59870 | -73.96031 | 5 | 47 |
| 51 | Beauty of Brooklyn and short ride to Manhattan. | 40.58546 | -73.94570 | 1 | 55 |
| 10 | Boutique apartment by the Ocean | 40.58294 | -73.95897 | 4 | 63 |
| 36 | Brigham Place | 40.59689 | -73.93353 | 6 | 18 |
| 22 | Bright Cozy Home in Brooklyn | 40.60776 | -73.95545 | 2 | 74 |

*Figure 13. Airbnb rental homes/apartments' location spread in Sheepshead Bay, Brooklyn NYC, depending on the clusters.*



# 5. RESULTS & DISCUSSION

First of all, it is really important to analyze in-depth each cluster identified. For this reason, in Figure 15 and 16 you can find the best and the worst clusters in terms of number of venues of interest in the nearby area, in which the 59 rental entire homes / apartments of the original *sheepshead_bay* are divided into.

*Figure 15. The worst cluster in terms of important venues is the one labelled as # 6.*

| | name | latitude | longitude | Cluster Labels | number of important venues |
|---|---|---|---|---|---|
| | | | | | |
| **36** | Brigham Place | 40.59689 | -73.93353 | 6 | 18 |
| **58** | Endless Elegance in Brooklyn | 40.59731 | -73.93381 | 6 | 21 |
| **50** | Family Friendly 2 Bedroom Apartment with Parking. | 40.58594 | -73.93182 | 6 | 22 |
| **20** | Marine park studio | 40.60778 | -73.94194 | 6 | 17 |
| **13** | Rain or Shine Studios: Dedicated to Production | 40.58599 | -73.92762 | 6 | 14 |

The worst cluster is labelled as **Cluster #6** (in orange within Figure 14) and is the one involving five apartments that are located in the Eastern part of the Sheepshead Bay (two of them South-East, two Central-East and one North-East). This is probably due to their high proximity to the Marine Park in the Eastern part of the neighborhood under analysis, that prevents the presence of any venues of interest such as restaurants, shopping or services venues and so on.

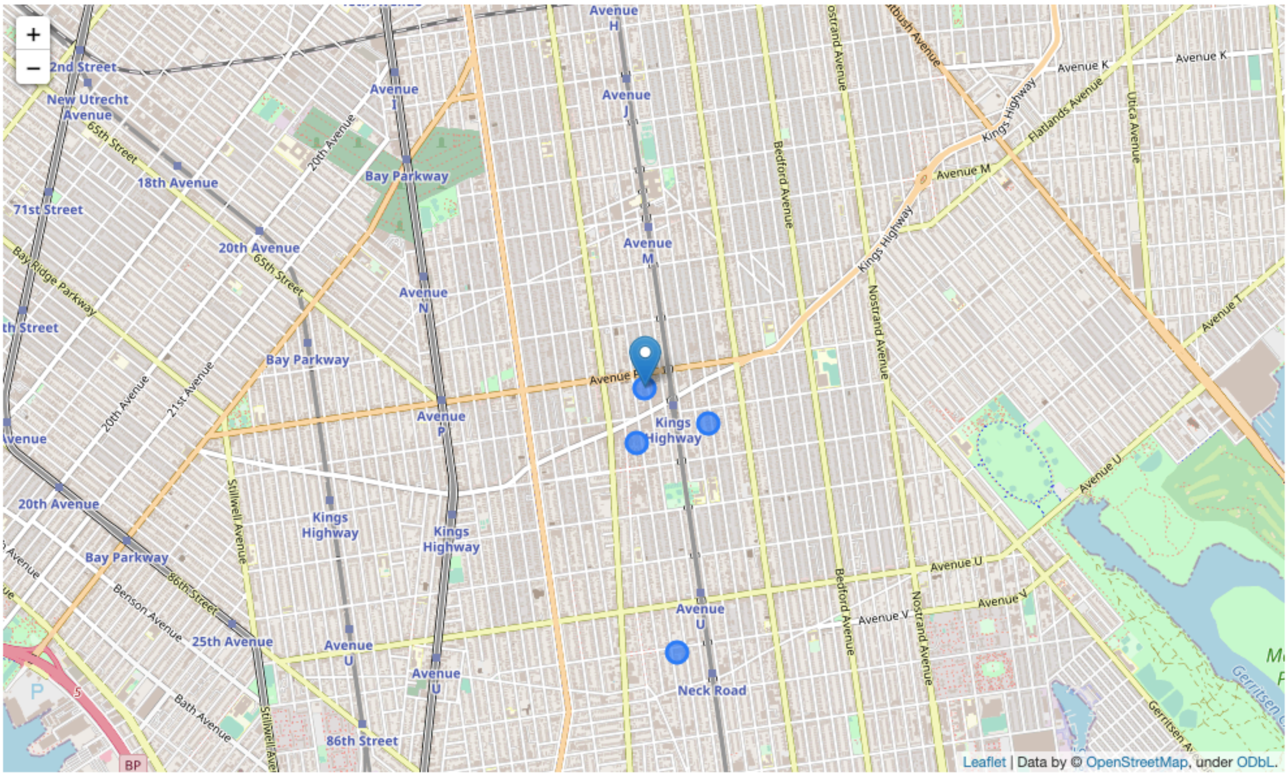
*Figure 16. The best cluster in terms of important venues is the one labelled as # 2.*

| | name | latitude | longitude | Cluster Labels | number of important venues |
|---|---|---|---|---|---|
| | | | | | |
| **22** | Bright Cozy Home in Brooklyn | 40.60776 | -73.95545 | 2 | 74 |
| **32** | Brooklyn in the house Private Apt Midwood Q su... | 40.60954 | -73.95966 | 2 | 84 |
| **12** | Spacious 1 Bedroom Available | 40.60676 | -73.96012 | 2 | 75 |
| **16** | ღ Spacious and chill studio 웃♥유 | 40.59631 | -73.95743 | 2 | 73 |

The best cluster instead is the one labelled as **Cluster #2** (in blue within Figure 14) and involves four apartments that are mostly located in the Northern part of the Sheepshead Bay neighborhood. In particular, three of them are really close to the intersection between Kings Highway and Avenue P (Figure 17).
This is justified by the fact that this area of the Sheepshead Bay neighborhood is one of the most attractive for its proximity to the rest of the city (especially with the other neighborhoods in Brooklyn to the North) and for the establishment of a Business Improvement District (BID) along part of the road.
In fact, a BID is a defined area within which businesses are required to pay an additional tax (or levy) in order to fund projects within the district's boundaries. BIDs may go by other names, such as business improvement area (BIA), business revitalization zone (BRZ), community improvement district (CID), special services

area (SSA), or special improvement district (SID). These districts typically fund services which are perceived by some businesses as being inadequately performed by government with its existing tax revenues, such as cleaning streets, providing security, making capital improvements, construction of pedestrian and streetscape enhancements, and marketing the area. The other clusters are in the middle ranges.

*Figure 17. Location spread of the Airbnb's rental entire homes/apartments within the best cluster (Cluster #2).*




Starting from these clusters is possible to find the best apartment according to our analysis. In fact, first-hand it should belong to Cluster #2; in order to find this apartment we can proceed just grouping the results found in the *sheepshead_data* dataframe and taking the maximum value. After grouping, the best rental entire home / apartment in the Sheepshead Bay seems to be "**Brooklyn in the house Private Apt Midwood Q subway**" with a score of **84** (number of venues of interest according to our analysis) and it is represented in the Figure 17 with a pointer-marker on it.
In its nearby area you can find the following venues:

- Turkish Restaurant
- Pizza Place
- Burrito Place
- Russian Restaurant
- Sushi Restaurant
- Middle Eastern Restaurant
- Mediterranean Restaurant
- Halal Restaurant
- Donut Shop

- Café
- Bakery
- Bagel Shop
- Steakhouse
- Caucasian Restaurant
- Eastern European Restaurant
- Food Truck
- Restaurant
- Deli / Bodega
- Diner
- Burger Joint
- Fast Food Restaurant
- Salad Place
- Fried Chicken Joint
- Chinese Restaurant
- Israeli Restaurant
- Mexican Restaurant
- Japanese Restaurant
- Asian Restaurant
- Indian Restaurant
- Italian Restaurant
- Metro Station
- Bus Station
- Rental Car Location
- Pharmacy
- Shopping Mall
- ATM
- Hospital
- Parking

It seems to be very good actually, since there are a lot of different restaurants in the nearby area (from Mediterranean cousine to Fast Foods, from Turkish restaurant to Russian ones) together with a bus station, a metro station, a shopping mall, an ATM, a parking and a rental car location. At the same time, hoping to never need them, it is close to the hospital and to a pharmacy, too. So, whatever the reason of the journey is, it seems to be really the best considering the venues in a radius of 600 meters.

# 6. CONCLUSION & FUTURE DEVELOPMENTS

The analysis made in this project work can be extended to:

- **other cities**;
- **other boroughs in the city** (even more than one, keeping in mind possible problems related to the Foursquare API calls' limit in case of developer Sandbox Tier Account or free Personal Tier Account);
- **other neighborhoods in the city** (even more than one, keeping in mind possible problems related to the Foursquare API calls' limit in case of developer Sandbox Tier Account or free Personal Tier Account);
- **other rental properties** (focusing on shared or private rooms for example);
- **other kinds of venues of interest** (focusing on other venues to call from Foursquare, depending on the reason of the journey).

At the same time, keep in mind that the approach followed until there has to be considered as an additional evaluating tool when deciding which apartment to choose among different available alternatives. In this direction, a possible future development would be **integrating the analysis just made with the common tools for a correct decision, such as feedbacks or comments in different trusted websites**. An idea can be extracting only positive feedbacks from 3 or 4 trusted websites for each apartments and using them as an additional score for the most appropriate decision.

Moreover, perhaps the most interesting future development would be **using each apartment's score** found before (in terms of venues of interest) **as an additional feature (a kind of extra-column) when predicting the price of each rental properties in the original dataset** (this would be a supervised ML approach, out-of-scope for this project work).