The Battle of Neighborhoods-Finding the Best NYC Apartment Rentals

Muriel Kosaka

Intro|Business Problem

Background

According the latest *NYC&Company* release New York City welcomed about 65.2 million tourists in 2018 year - 51.6 million domestic and 13.5 million international visitors. And these numbers are continuously increasing from year to year.

New York City has the largest selection of lodging choices in the country – from the hostels to the luxury hotels. The prices vary from 100$ till several thousand dollars with average price 292 USD per night.

Intro|Business Problem Continued

Problem Description and Project Goal

In New York City there are almost 300 hotels with over 75,000 hotel rooms and Airbnb has more than 50,000 apartment listings in New York City in 2018 year - it can be hard to find the right fit or know how much you will get with your money.

In this project we will try to find the most optimal neighborhoods on Manhattan where a tourist can rent an accommodation via Airbnb service and have a pleasant stay in NYC and a possibility to attend the most visited attractions like Central Park, Times Square and so on.

Target Audience

This investigation would interest New York City's visitors who prefers short stays (from 1 night) and wants to select the best neighborhoods on Manhattan, New York.

Data

In our investigation we will use the free and public available datasets.

We will try to evaluate available Airbnb 2019-year accommodations on Manhattan, New York and define the most reasonable apartments sets (clusters) for the visitors.

Based on definition of our problem, we suppose that factors that will help us are:

- accommodations average price per person by the neighborhood
- number of tourist attractions near the accommodation
- number of crimes nearby the accommodation.

Following data sources are needed for our project:
- New York City apartment listing from the Inside Airbnb site
- New York Neighborhoods Tabulation Areas – official NYC neighborhoods names and coordinates
- Foursquare API to extract data about venues - food places, museums, galleries, shopping centers, and so on

Data Cleaning

Airbnb New York City apartment listing

For our project records were filtered as:

- Borough - Manhattan, New York only
- Number of reviews >= 10
- Availability >= 10 days/year
- Last Scraped/Reviewed later than 2019-10-01
- Minimum nights >= 1
- Excluded Hostels and Camper/RV
- Excluded Shared rooms.

After filter was applied, we had 1,571 accommodations in our dataset.

In this project we are trying to detect Manhattan's Neighborhoods that have accommodations for rent with positive reviews, reasonable prices, low number of crimes and tourists' attractions nearby.

In the first step we have collected the following data:
- Airbnb Accommodations with their NYC Tabulation Area (official neighborhood names)
- Defined NYC Tabulation Area (official neighborhood name) for each Manhattan's crime case.

The second step in our analysis will be a calculation and exploration different neighborhoods of Manhattan. We will explore the following characteristics:
- average price per person;
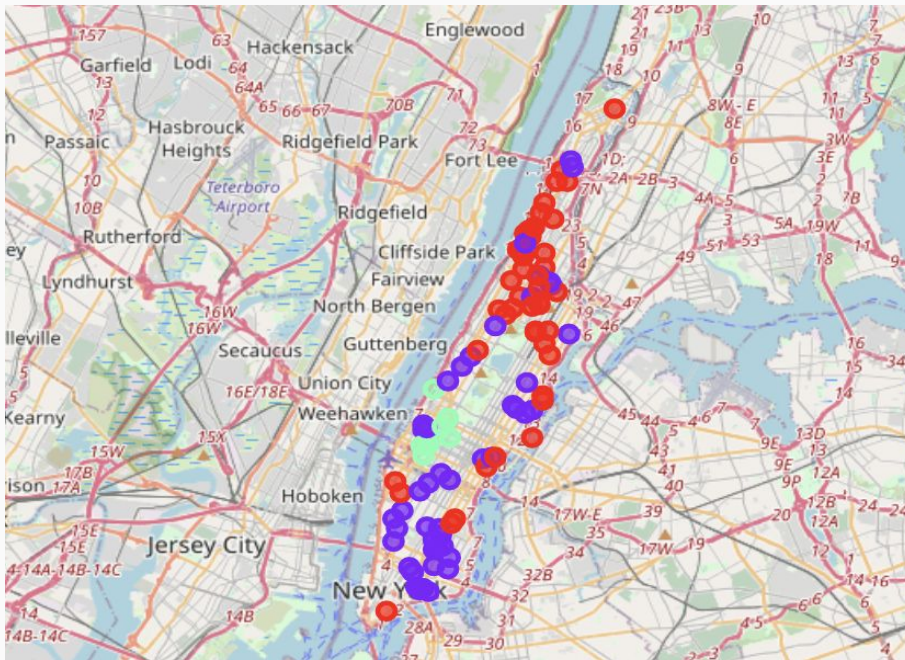- number of accommodations available.

In the third and final step we will
select Top-100 Airbnb accommodations based on summary rating, number of crimes and price per person, and
- invoke Foursquare API to find Top accommodations' nearby venues
- create and investigate clusters (using k-means clustering) for our accommodations to make some recommendations to our tourists.
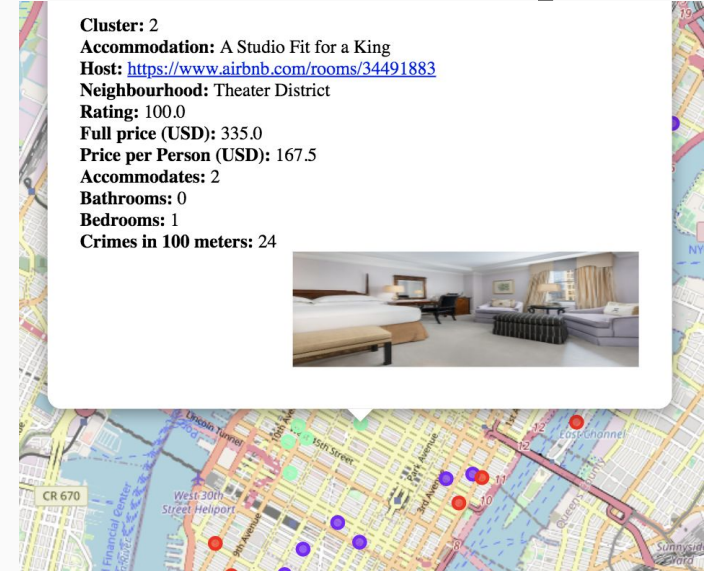
Methodology

Cluster: 2
Accommodation: A Studio Fit for a King
Host: https://www.airbnb.com/rooms/34491883
Neighbourhood: Theater District
Rating: 100.0
Full price (USD): 335.0
Price per Person (USD): 167.5
Accommodates: 2
Bathrooms: 0
Bedrooms: 1
Crimes in 100 meters: 24

Top five neighborhoods with HIGHEST price per person

- Hudson Yards-Chelsea-Flatiron-Union Square : 76.20USD : 78 accommodations
- Stuy Town-Cooper Village : 75 USD : 1 accommodation
- West Village : 72.79 USD : 45 accommodations
- Lincoln Square : 72.38 USD : 16 accommodations
- Battery Park - Lower Manhattan : 67.78 USD : 15 accommodations

Top five neighborhoods with LOWEST price per person

- Washington Heights North: 30.54 USD : 16 accommodations
- Marble Hill - Inwood: 35.91 USD : 10 accommodations
- Washington Heights South: 36 USD : 37 accommodations
- Hamilton Heights: 37.99 USD : 63 accommodations
- Manhattanville: 38.08 USD : 26 accommodations

Analysis
Foursquare API Neighborhoods Analysis

Because of the Foursquare API limitations for free usage lets analyze Top-100 Accommodations from the Airbnb data set.
We define Top-3 Venue Categories for each accommodation in radius of 1000 meters.
Then we will try to define the 3 clusters for these accommodations.
Select Top-100 Airbnb accommodations by

- **review_scores_rating** - overall accommodations rating - from maximum 100 to lower values;
- **full_price** - from lower price to higher;
- **price_per_person** - from lower price to higher;
- **crimes** - from lower number to higher;

We define our custom Top-Level categories for Venues

```python
# Define categories from Foursquares
fine_art = ['Art','Arts','Museum', 'Library','Exhibit','Gallery']
eateries = ['Restaurant','Steakhouse']
shopping = ['Shopping Mall','Market','Boutique']
outdoor = ['Sculpture Garden','Scenic Lookout','Roof Deck','Outdoor Sculpture','Monument / Landmark',
           'Memorial Site','Lighthouse','Historic Site','Harbor / Marina','Fountain','Event Space','Bridge',
           'Waterfront','Church','Building','Garden','Historic Site','Lake','Park',
           'Pier','Rest Area','River','Synagogue','Field']
entertainment = ['Nightclub','Circus','Club', 'Stadium', 'Karaoke Bar', 'Pub','Theater','Opera', 'Concert', 'Zoo']

#Join all categories' values in one
tourists_categories = fine_art + eateries + shopping + outdoor +entertainment
```

Analysis
Foursquare API Neighborhoods Analysis

Calculate the top 3 venue categories for each accommodations.

Then run K-Means to cluster the neighborhoods into 3 clusters

Cluster Labels	name	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	1st Most Common Venue Share	2nd Most Common Venue Share	3rd Most Common Venue Share
0	1 Bedroom Apartment — Amazing Location in Midtown!	Food Place	Entertainment	Fine Art	0.89	0.07	0.04
0	2-story Apt in the heart of The Upper West Side!	Food Place	Sightseeing	Shopping	0.81	0.08	0.04
1	A Studio Fit for a King	Entertainment	Food Place	Fine Art	0.53	0.33	0.1
2	Aesthetically Designed Modern Private Room	Food Place	Sightseeing	Entertainment	0.68	0.26	0.05
0	Art filled peaceful paradise EV Union Square	Food Place	Sightseeing	Shopping	0.77	0.08	0.08

Analysis
Foursquare API Neighborhoods Analysis
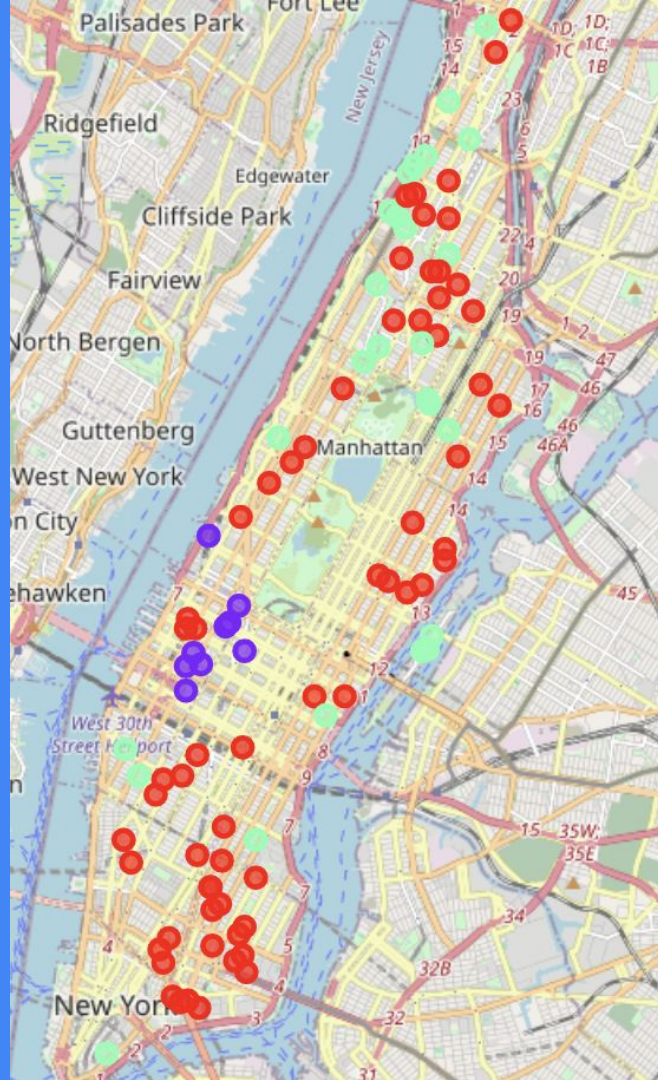
Cluster 0 (red dots)- Mix characteristics

- lowest average price_per_person - $49;
- mix of all Venue Categories (Fine Arts, Shopping, Entertainment);
- contains 58% from all accommodations selected from analysis (Top-100 Airbnb accommodations);
- spreads almost on all Manhattan's areas.

Cluster 1 (purple dots)- Entertainment

- average price_per_person among all clusters - $62;
- highest average crimes rate among all clusters - 102;
- Entertainment is 1st and the 2nd Top Common Venue Categories;
- contains 15% from all top accommodations (Top-100 Airbnb accommodations);
- spreads on Chelsea, Hell's Kitchen, and Midtown Airbnb's Neighborhoods.

Cluster 2 (light blue dots)- Sightseeing

- highest average price_per_person - $75;
- Sightseeing is the second top Common Venue Category;
- contains 27% from all top accommodations (Top-100 Airbnb accommodations);
spreads on East Harlem, Financial District, Harlem, Inwood, Morningside Heights, Roosevelt Island, Upper West Side, Washington Heights, West Village.

Results and Discussion

During the analysis, three clusters were defined.
All clusters have a 'Food Place' category as the First Common Venues. This is what we have in common among our clusters.

But they are distinguished by the other characteristics as
- average **Price per person;**
- the second Common Venues;
- number of available Airbnb accommodations;
- neighborhoods location.

Cluster 0 – Mix is the most generic cluster - it has a
- lowest average price_per_person - $49;
- contains 58% from all accommodations selected from analysis (Top-100 Airbnb accommodations);
- spreads almost on all Manhattan's areas.

Results and Discussion

Cluster 1 - Entertainment is the smallest cluster with the following qualities (Nightclub, Stadium, Pub, Theater, Concert and so on):
- average *price_per_person* among all clusters - $62;
- *Entertainment* is 1st and the 2nd Top Common Venue Categories;
- contains 15% from all top accommodations (Top-100 Airbnb accommodations);
- spreads on *Chelsea, Hell's Kitchen, and Midtown* Airbnb's Neighborhoods.
Cluster 2 - Sightseeing is the cheapest one with many Sightseeing attractions nearby (Monument/Landmark, Memorial Site, Historic Site, Lake, Park, Pier, and so on)
- highest average price_per_person - $75;
- *Sightseeing* is the second top Common Venue Category;
- contains 27% from all top accommodations (Top-100 Airbnb accommodations);
- spreads on *East Harlem, Financial District, Harlem, Inwood, Morningside Heights, Roosevelt Island, Upper West Side, Washington Heights, West Village.*

We identified three clusters from which a visitor could choose an appropriate accommodation based on his/her preferences or needs.

Limitations
-limited our investigation by Manhattan Borough only;
- Foursquare free account has a limitation of 950 calls/day so maybe it's worth to upgrade our free account to analyze Top-1000