The Battle of Neighborhoods
Find the Best Place to Stay in New York City

Applied Data Science Capstone
May 18, 2021

Introduction/Business Problem

Background

- According to the latest report, New York city welcomed about **65.2 million tourists in year 2018** - 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$ per night.** The hotel **occupancy rate** is also high - in 2018 year it was **88%.** Visitors prefer short stays that are often over weekends – **averaging 2.4 nights.**

Problem Statement

- In New York city there are **almost 300 hotels with over 75.000 rooms** and Airbnb has **more than 50,000 apartment** listings in New York city in year 2018 – it can be difficult to find the right fit or know how much you will get with your money.

- In this project we aims to **find the most optimal neighborhoods on Manhattan** where a tourist can rent a room via Airbnb and have a pleasant stay, as well as the possibility to attend the most visited attractions like Central Park, Time Square, and so on.

Target Audience

- This work would interest potential visitors who prefers short stays and wants to select the best neighborhoods on Manhattan.

Data

Initial datasets

In the project we will mainly use free and publicly available datasets. Given our problem statement, we will **evaluate available Airbnb accommodations** on Manhattan, New York city and define the most reasonable apartments sets/clusters for potential visitors.

Based on defined challenge, we believe **following factors will be needed:**

- Accommodation's average price per person by the neighborhood
- Number of tourist attractions near the accommodation
- Number of crimes nearby the accommodation

Therefore, **following data sources will be needed:**

- New York city apartment listing from the Airbnb website
- New York city neighborhoods tabulation areas – official neighborhoods names and coordinates
- Foursquare API to extract data about venues – food places, museums, etc.
- New York crime data records

Data (cont')

Data Cleaning

Airbnb New York city apartment listing

- Borough – Manhattan, New York only
- Number of reviews >= 10
- Availability >= 10 days/year
- Last reviewed later than 2019-10-01
- Minimum nights >=1
- Excludes hostels and camper/RV
- Excludes shared rooms

After applying above-mentioned filtering, **2,356 accommodations are identified.**

New York Police Crime Records

- Borough – Manhattan, New York only
- Crime type – Felony and misdemeanor

In total, **101,086 crimes records are identified.**

Methodology

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

In the first step, we have collected the following data

- Airbnb accommodations with their New York city tabulation area
- Airbnb accommodations' number of crimes nearby
- Defined New York city tabulation area for each Manhattan's crime case

The second step in the analysis is a calculation and exploration different neighborhoods of Manhattan. We will explore the following characteristics

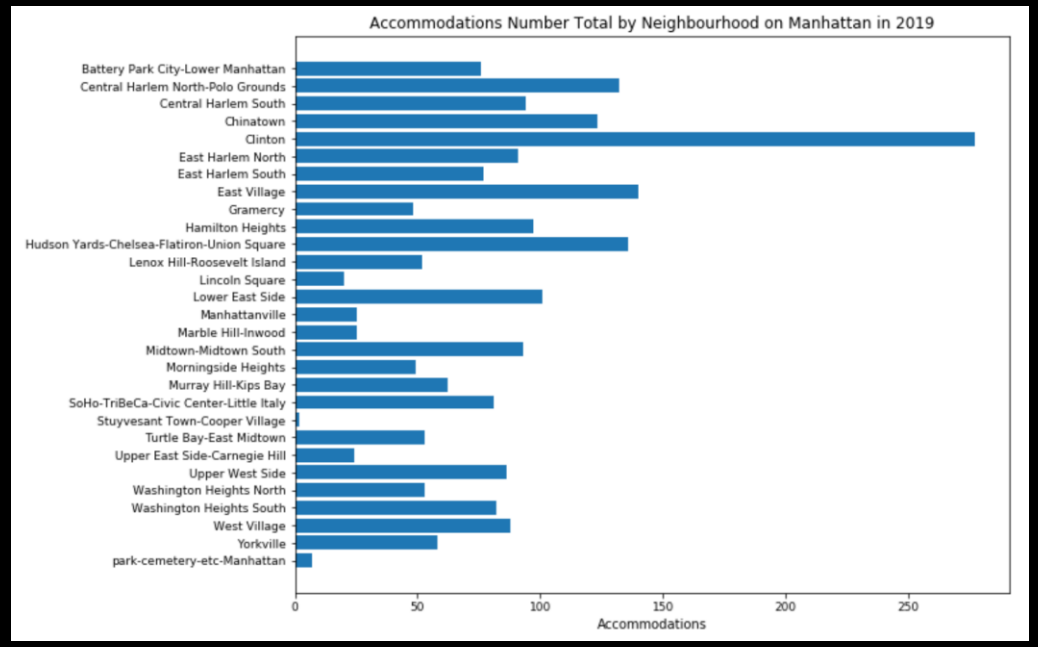
- Number of crimes in the area
- Average price per person
- Number of accommodations available

In the third and final step we will

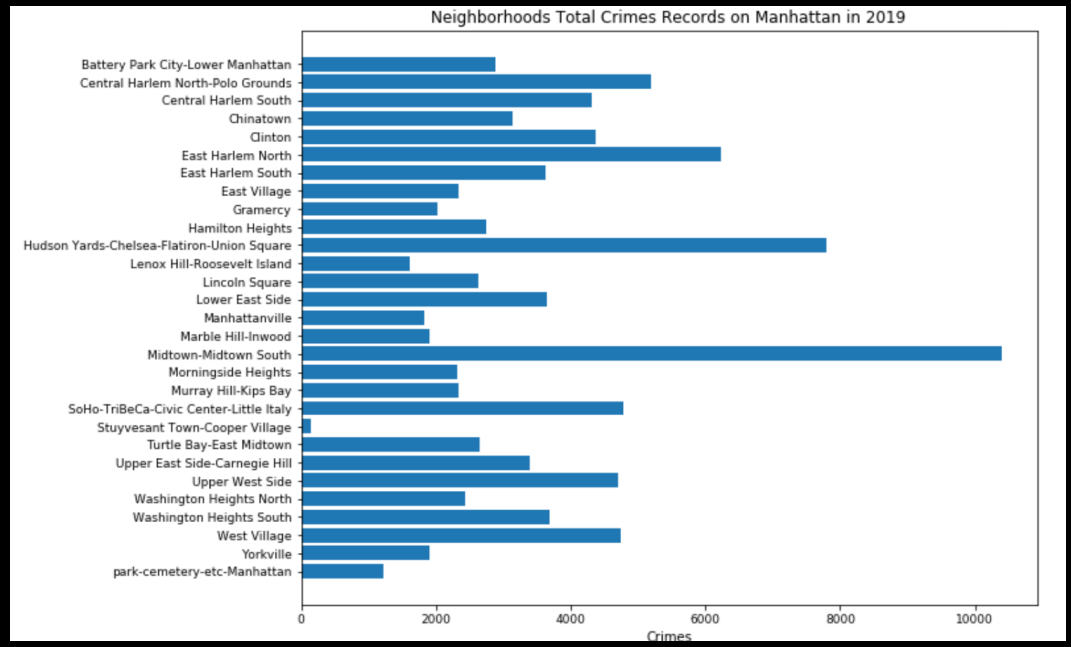
- Select top-100 Airbnb accommodations based on summary rating, number of crimes and price per person, and
- Invoke Foursquare API to find top accommodations' nearby venues
- Create and investigate clusters (k-means) for accommodations to make recommendations to potential tourists

Analysis

Apartments Total by Neighborhood Chart

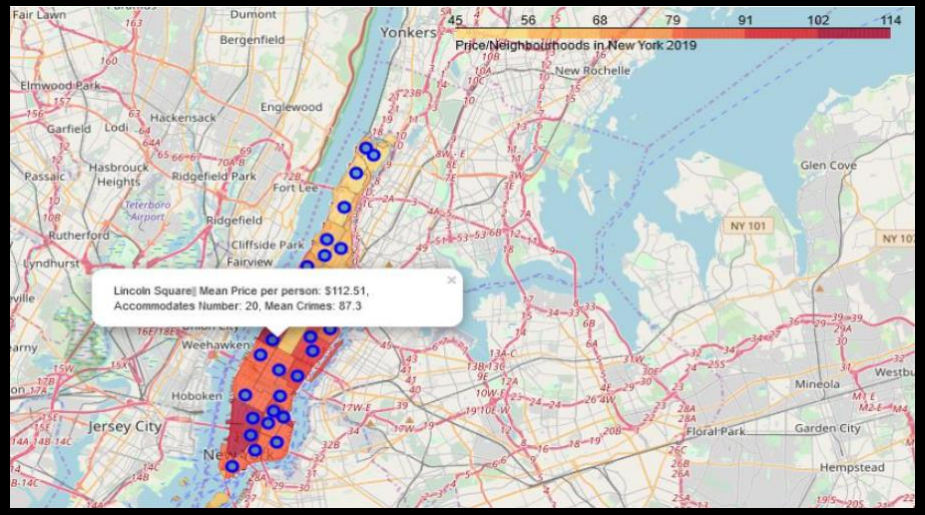


Neighborhoods Crimes Records Chart



Analysis (cont')

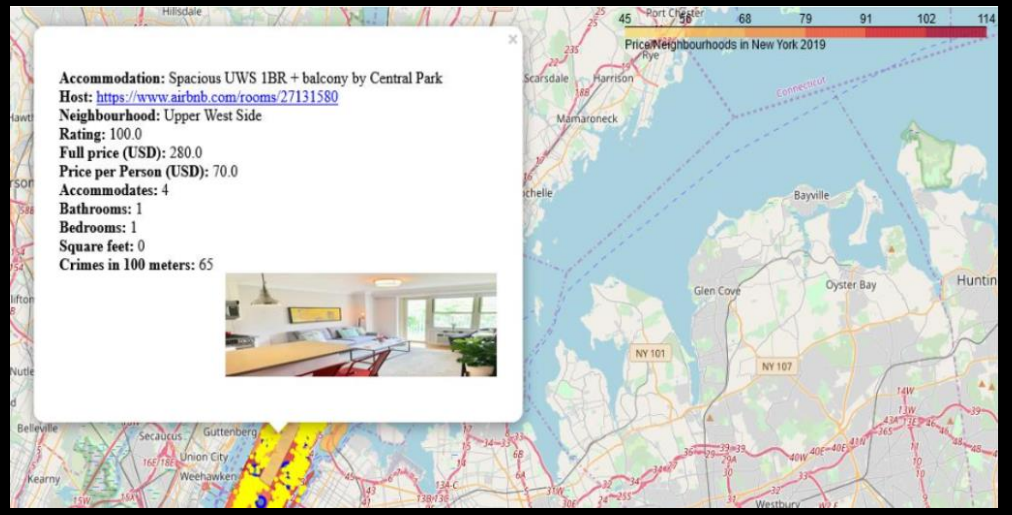
NYC Tabulation Area Neighborhoods
Average Prices per Person



Top 3 neighborhoods with **highest average price per person**

- West Village
- Lincoln Square
- Stuyvesant Town-Cooper Village

Accommodations Detailed Info Map

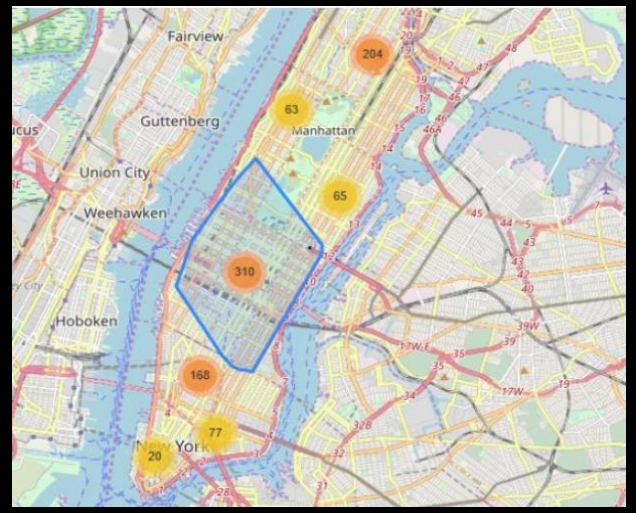


Top 3 neighborhoods with **lowest average price per person**

- Marble Hill-Inwood
- Washington Heights South
- Washington Heights North

Analysis (cont')

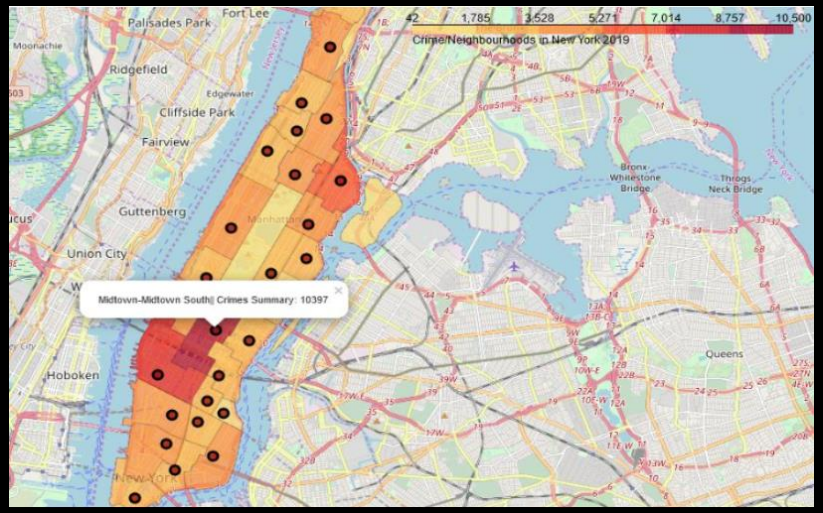
Crimes Cluster Map



Top 3 neighborhoods with **highest crime level**

- Midtown-Midtown South
- Hudson Yards-Chelsea-Flatiron-Union Square
- East Harlem North

Summary Crimes by Neighborhoods Map



Top 3 neighborhoods with **lowest crime level**

- Stuyvesant Town-Cooper Village
- Lenox Hill-Roosevelt Island
- Manhattanville

Analysis (cont')

Foursquare API Neighborhood Analysis

Because the Foursquare API has a limitation for free usage, we only analyze here the top 100 accommodations from the Airbnb dataset. So we first define **top 3 venue categories** for each accommodation **in radius of 1000 meter.** Then we define the **3 clusters** for these accommodations.

Select top 100 Airbnb accommodation by

- **review_scores_rating** – overall accommodation 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 then define our custom top-level categories for venues (see screenshot below).

```
fine_art_cat = ['Art','Arts','Museum', 'Library','Exhibit','Gallery']
eat_place_cat = ['Restaurant','Steakhouse']
shopping_cat = ['Shopping Mall','Market','Boutique']
outdoor_cat = ['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_cat = ['Nightclub','Circus','Club', 'Stadium', 'Karaoke Bar', 'Pub','Theater','Opera', 'Concert', 'Zoo']

#Join all categories' values in one
tourists_categories = fine_art_cat + eat_place_cat + shopping_cat + outdoor_cat +entertainment_cat
```

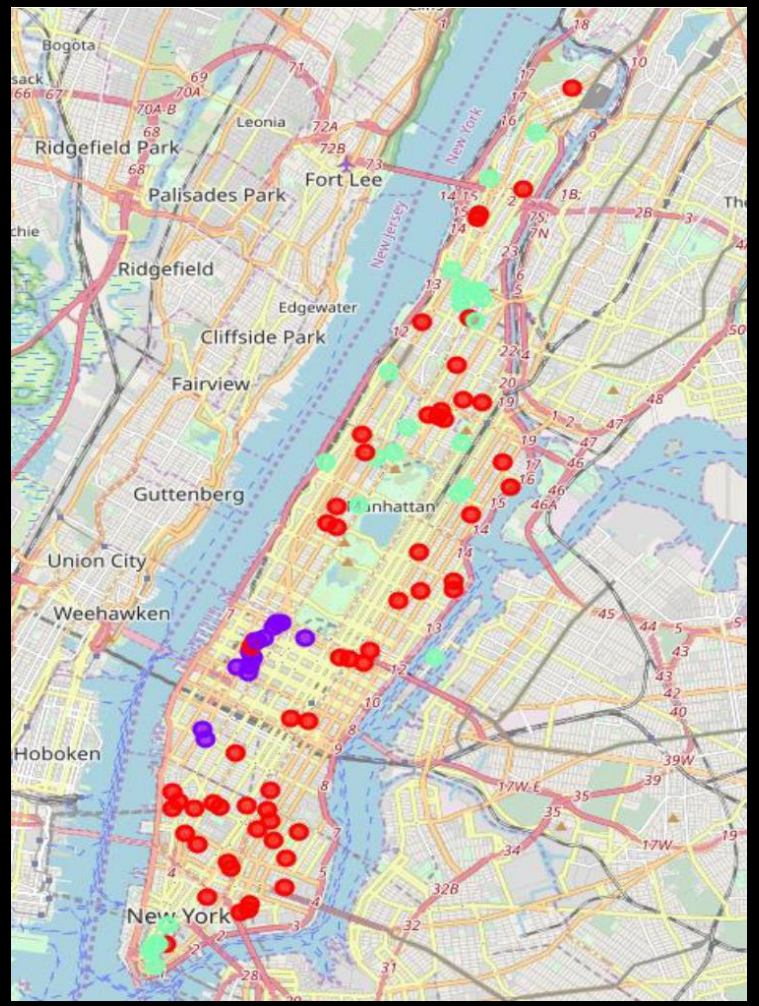
Analysis (cont')

Foursquare API Neighborhood Analysis (cont')

First of all, we calculate the top 3 venues categories for each accommodation.
Then we run k-means to cluster the neighborhood 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
1	**Stylish, Quiet, Centrally Located (9th & 52nd)	Food Place	Entertainment	Fine Art	0.62	0.35	0.04
2	157-C	Food Place	Sightseeing	Fine Art	0.52	0.29	0.14
2	A neat bedroom in a cozy 3-bedroom apartment	Food Place	Sightseeing	Shopping	0.48	0.41	0.04
0	Art filled peaceful paradise EV Union Square	Food Place	Sightseeing	Shopping	0.71	0.07	0.07
0	Artsy Parisian Apt in Greenwich Village	Food Place	Entertainment	Sightseeing	0.65	0.26	0.09

Analysis (cont')



Foursquare API Neighborhoods Analysis

Now, we can examine each cluster and determine our custom venue categories that distinguish each cluster.

Cluster 0 – Mix (red dots) characteristics:

- Average price_per_person
- Average crimes rate
- Contains 58% from all top accommodations

Cluster 1 – Entertainment (blue dots) characteristics

- Highest average price_per_person among all clusters
- Highest average crimes rate among all clusters
- Contains 15% from all top accommodations

Cluster 2 – Sightseeing (light-green dots) characteristics

- Lowest average price_per_person
- Lowest crimes rate among all clusters
- Contains 27% from all top accommodations

Results and Discussion

During the analysis, 3 clusters have been defined.
All clusters have a **Food Place** category as the first common venues. This is what we have in common among clusters.

But they are distinguished by the other characteristics as

- Average **price_per-person**
- Average **crimes rate**
- The second common veues
- Number of available Airbnb accommodations
- Neighborhoods location

Cluster 0 - Mix is the most generic cluster

- Average price_per_person 100$
- Average crimes rate 67
- Mix of all venue categories: Fine Arts, Shopping, Entertainment
- Contains 58% from all accommodations selected from analysis (top 100 Airbnb accommodations)

Results and Discussion

Cluster 1 – Entertainment is the smallest cluster with the following qualities (nightclubs, stadiums, pubs, etc.)

- Highest average price_per_person among all clusters 111$
- Highest average crimes rate among all clusters 102
- Entertainment is 1st and the 2nd top common venue categories
- Contains 15% from all top accommodations

Cluster 2 – Sightseeing is the cheapest one with many sightseeing attractions nearby (Monument/Landmark, Memorial Lake, etc.)

- Lowest average price_per_person 59$
- Lowest crimes rate among all cluster 65
- Sightseeing is the second top common venue category
- Contains 27% from all top accommodations

We identified three clusters from which a visitor could choose an appropriate accommodation based on individual preferences.

Limitation

- We limited our investigation by Manhattan Borough only
- Foursquare free account has a limitation of 950 calls/day

Conclusion

To conclude, the basic data analysis was performed to **identify Manhattan's neighborhoods clusters** for a short stay visit. During the analysis, we cleansed the investigated Manhattan's neighborhoods' datasets, found statistical characteristics and visualize them.

The aim of the this project is to help potential Manhattan visitors select the most suitable Airbnb accommodations to stay based on the most **common venues, price policy, and safety characteristics.**

- If a person is interested in entertainment, e.g., nightlife, pubs, concerts, and movies, we recommend paying attention for accommodations located in **Cluster 1 – Entertainment:** Chelsea, Hell's Kitchen, and Midtown Airbnb's offerings. But the person should take into the consideration of the **relatively high price and crime rate** for these neighborhood

- If a person is looking for an area with lower prices and nice views nearby, we recommend looking at **Cluster 2 – Sightseeing**

- If a person doesn't have any preferences – investigate proposals from **Cluster 0 – Mix.** It has an **average price** and spreads over almost all Manhattan's neighborhood

Area of Improvement

- **Include other New York city boroughs** – Bronx, Brooklyn, Queens, and Staten Island in the future
- **Utilize other services** such as Google API to find nearby venues in the future
- **Include hotels** in the future research (in this project, we only included accommodations offered via Airbnb)