

# Manhattan's Safest Neighborhood

Sean Brown

November 15, 2019

## 1. Introduction

### 1.1 Background

New York City is composed of five boroughs, Manhattan, Brooklyn, the Bronx, Queens, and Staten Island, each composed of several neighborhoods with diverse cultures and lifestyles. Of the five boroughs, Manhattan is arguably the most recognizable and alluring location for those aspiring to find their career in the “administrative, business, and financial center” of New York and, possibly, the United States<sup>1</sup>. According to Neighborhood Scout, Manhattan sees approximately 35,200 violent and property crimes annually and has a Crime Index of 30, meaning it is safer than 30% of U.S. cities<sup>22</sup>. As with any major decision involving the establishment of residence, personal and property safety is an important factor to consider when choosing a neighborhood for oneself or one’s family. Therefore, the ability to decide a neighborhood based on available crime data would be beneficial for a prospective resident of Manhattan.

### 1.2 Problem

A new President of Finance has been hired by an investment bank in Manhattan and must decide which neighborhood he should relocate his family to from his current home in San Francisco. Given his hefty corporate salary and attractive bonus structure, he is less worried about cost and more worried about safety for his family. Being an avid martial arts practitioner, he also wants to make sure that his new home is in close proximity to a dojo where he can continue to polish his karate skills. As part of his contract, the investment bank has hired this Data Scientist to analyze potential neighborhoods that would fulfill their new President of Finance’s criteria. The ultimate recommendation will be the neighborhood that minimizes proximity to crime and maximizes proximity to martial arts dojos.

---

<sup>1</sup> <https://www.britannica.com/place/New-York-City/The-boroughs>

<sup>2</sup> <https://www.neighborhoodscout.com/ny/new-york/crime>

## 1.3 Interest

In addition to the investment banker being interested in his new potential living situation, any resident or potential resident of Manhattan may be interested in having an informed view of neighborhood safety within the borough.

# 2. Data Acquisition and Cleaning

## 2.1 Data Sources

The borough and neighborhood data for Manhattan has been provided by the [NYU Spatial Data Repository](#), which includes the centroids for each of Manhattan's five boroughs. [Kaggle](#), owned by Google and home to the largest and most diverse data community in the world, was used to gather New York City crime data from 2014-2015 by borough with location data. Finally, the [Foursquare Places API](#) was used to gather venue data for Manhattan, including longitude and latitude and types of venues.

## 2.2 Data Cleaning

### 2.2.1 Neighborhood Dataset

Using Python's json library, the geojson file from NYU was loaded, parsed, and appended to a pandas DataFrame. The DataFrame was then filtered to include only those neighborhoods located within the borough of Manhattan.

	Borough	Neighborhood	Latitude	Longitude
0	Manhattan	Marble Hill	40.876551	-73.910660
1	Manhattan	Chinatown	40.715618	-73.994279
2	Manhattan	Washington Heights	40.851903	-73.936900
3	Manhattan	Inwood	40.867684	-73.921210
4	Manhattan	Hamilton Heights	40.823604	-73.949688

Table 1 - Manhattan Neighborhood (First 5)

The data was mapped with Folium to determine its validity.

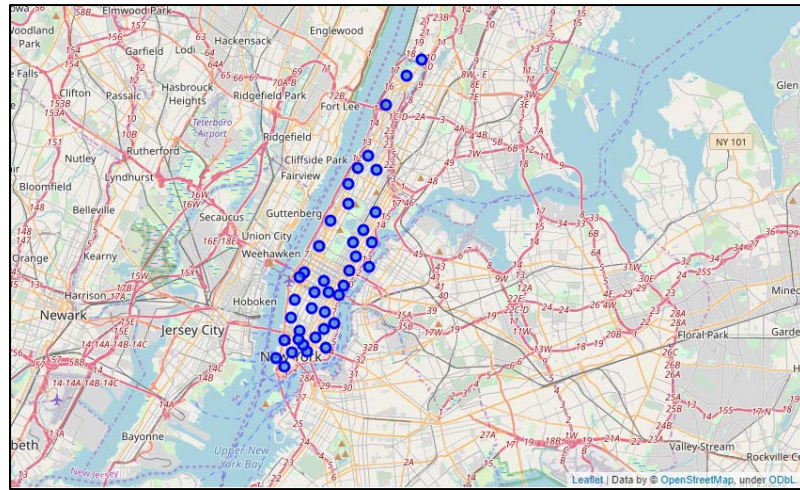


Figure 1 - Neighborhoods of Manhattan

## 2.2.2 Crime Dataset

The New York City crime dataset was loaded into a pandas DataFrame and filtered by borough to include only crimes in Manhattan that were reported in 2015. The dataset included 24 columns, but only the report date, law category (felony or misdemeanor), and latitude and longitude coordinates were required for this analysis. Once filtered to the relevant columns, the data set was split into Felony and Misdemeanor crimes and mapped for visualization.

	RPT_DT	LAW_CAT_CD	Latitude	Longitude
2	12/31/2015	FELONY	40.802607	-73.945052
4	12/31/2015	MISDEMEANOR	40.738002	-73.987891
6	12/31/2015	MISDEMEANOR	40.720200	-73.988735
9	12/31/2015	MISDEMEANOR	40.765618	-73.963623
11	12/31/2015	MISDEMEANOR	40.732828	-73.986063

Table 2 – Crime Dataset (First 5)

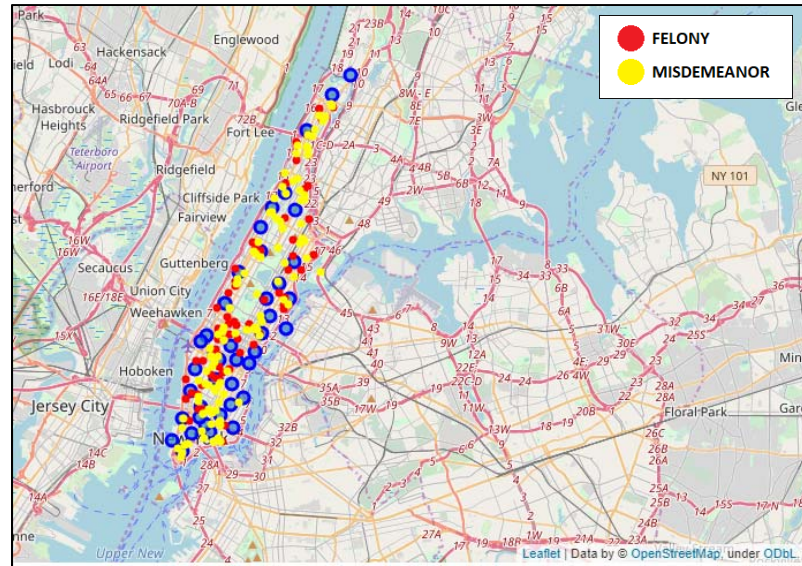


Figure 2 – Crimes of Manhattan

### 2.2.2 Martial Arts Dojo Dataset

The Places API from Foursquare was used to find all venues in Manhattan within each neighborhood and load them into a DataFrame. The DataFrame was then filtered to include only Martial Arts dojos and mapped for visualization.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
1237	Murray Hill	40.748303	-73.978332	Coban's Muay Thai Camp	40.750355	-73.981554	Martial Arts Dojo
1764	Little Italy	40.719324	-73.997305	Five Points Academy	40.719566	-73.999819	Martial Arts Dojo
1924	Soho	40.722184	-74.000657	Five Points Academy	40.719566	-73.999819	Martial Arts Dojo
2500	Carnegie Hill	40.782683	-73.953256	Modern Martial Arts NYC Eastside	40.778487	-73.953136	Martial Arts Dojo
2638	Civic Center	40.715229	-74.005415	Modern Martial Arts NYC Tribeca	40.715431	-74.007362	Martial Arts Dojo

Table 3 – Dojo Dataset (First 5)

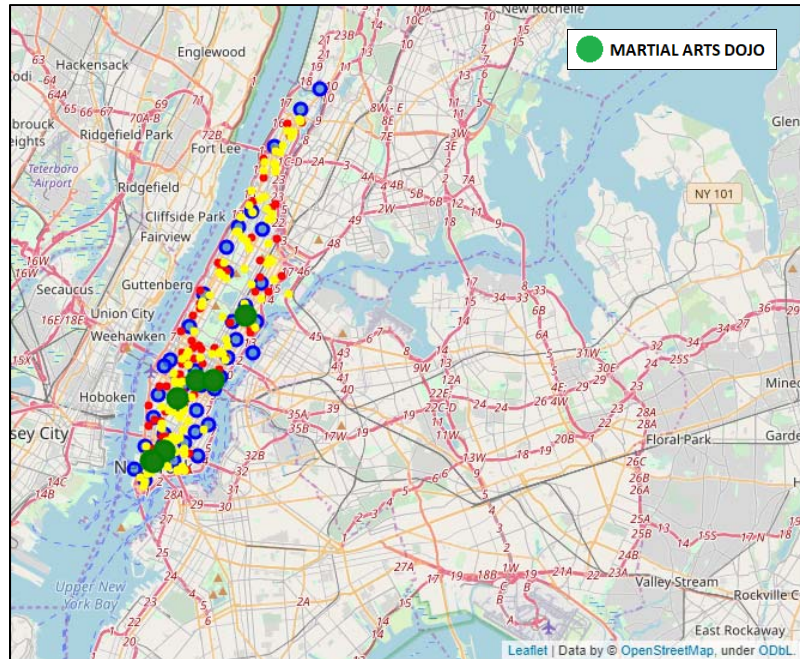


Figure 3 – Manhattan Martial Arts Dojos

## 2.3 Final Datasets

After cleaning the initial data, the final datasets are as follows:

Dataset	Entries
Manhattan Neighborhoods	40
Manhattan 2015 Felony Crimes	68
Manhattan 2015 Misdemeanor Crimes	109
Manhattan Martial Arts Dojos	9

Table 4 – Final Datasets

## 3. Exploratory Data Analysis

### 3.1 Target Variable

In order to determine the optimal location for the President of Finance (PoF) to purchase a new home, a scoring system needed to be created to rank each neighborhood and determine the best option. In the analysis, the important dependent variables were the distances to the centroids of Felony

and Misdemeanor crimes, each weighted according to their importance, and the distance to the centroid of Martial Arts dojos (to give the PoF several venue options to enjoy his pastime). This analysis will seek to maximize the distance to the crime centroids and minimize the distance to the dojo centroid, providing a neighborhood with the highest score according to the following equation:

$$\text{Score} = \frac{(0.8 \times C_F + 0.2 \times C_M)}{C_D}$$

Where  $C_F$  is the distance to the centroid of Felony Crimes,  $C_M$  is the distance to the centroid of Misdemeanor crimes, and  $C_D$  is the distance to the centroid of dojos calculated as follows:

$$D = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

The distance to felony crimes was arbitrarily weighted more heavily than the distance to misdemeanor crimes.

### 3.2 Crime and Dojo Centroids

The centroids of the crime and dojo point clusters can be found using the following formula:

Centroid Formula: The centroid of the points  $(x_1, y_1), (x_2, y_2), (x_3, y_3), \dots, (x_n, y_n)$  is

$$\left( \frac{x_1 + x_2 + x_3 + \dots + x_n}{n}, \frac{y_1 + y_2 + y_3 + \dots + y_n}{n} \right)$$

Using the formula, the centroid for each cluster is as follows:

Dataset	Centroid (Latitude, Longitude)
Manhattan 2015 Felony Crimes	(40.77285, -73.97293)
Manhattan 2015 Misdemeanor Crimes	(40.77304, -73.97014)
Manhattan Martial Arts Dojos	(40.73799, -73.98698)

The centroids are mapped below:



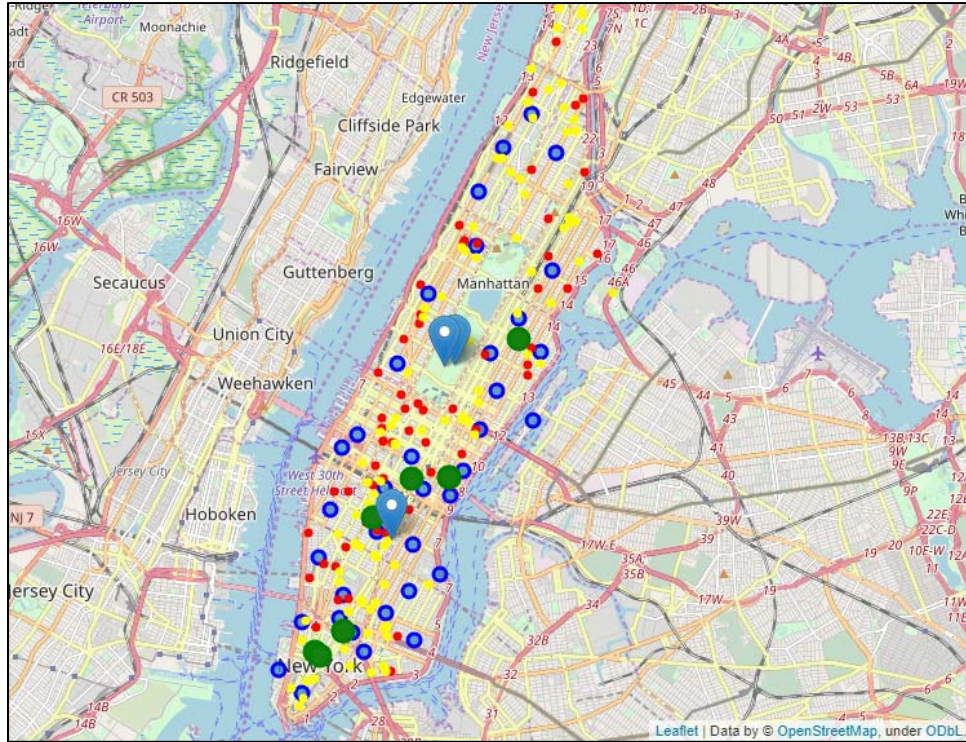


Figure 4 – Dataset Centroids

## 4. Data Analysis

### 4.1 Target Variable

Using the Score equation from 3.1, the location scores for each of the Manhattan neighborhoods were calculated and plotted.

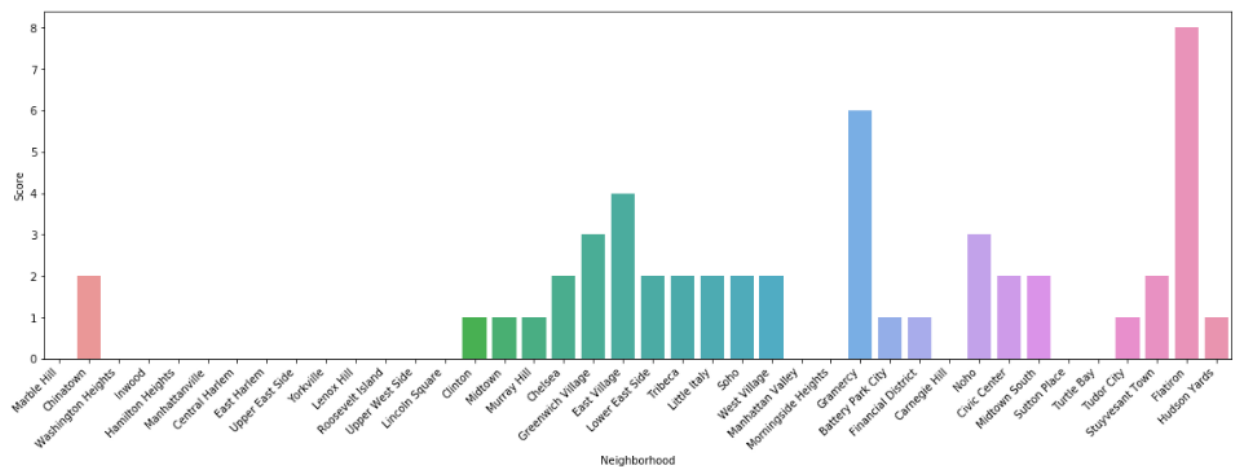


Figure 5 – Neighborhood Scores Plot

## 4. Results

The highest score was determined to be the **Flatiron** neighborhood, which a location score of 8, with the second highest being **Gramercy** at 6.

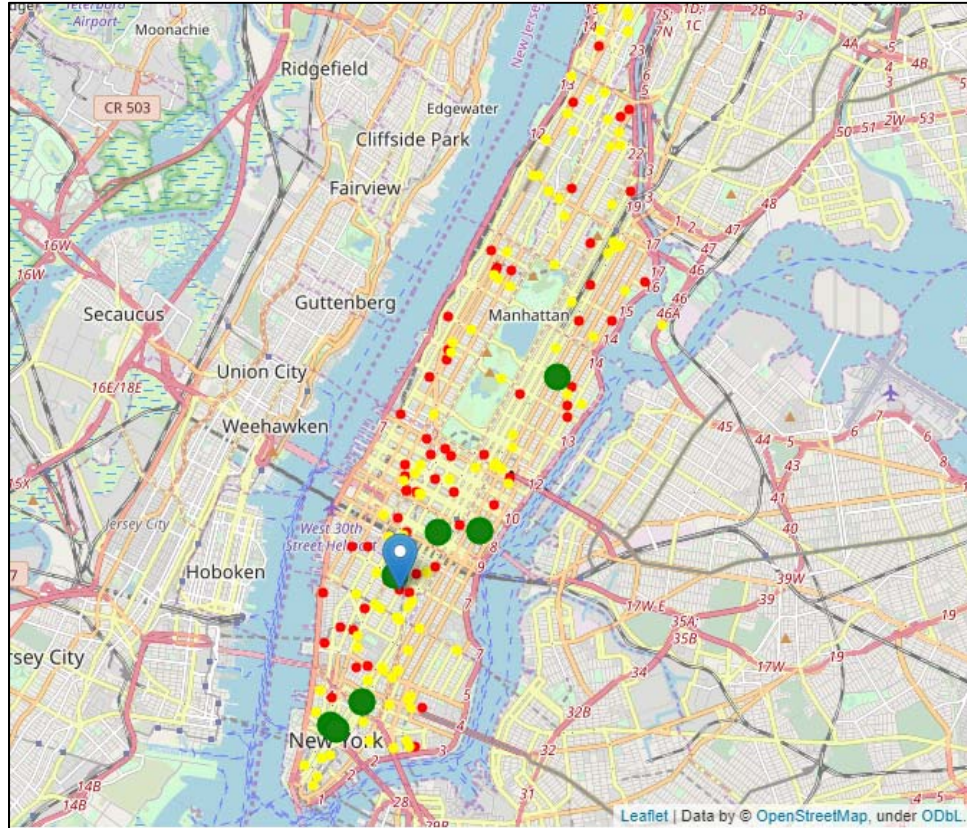


Figure 6 – Final Manhattan Map

## 5. Conclusion

The purpose of this project was to identify a Manhattan neighborhood for an investment bank's new President of Finance that would be optimally located based on available crime data and proximity to martial arts dojos, the President's favorite pastime.

Using data analytics and available Python libraries, it was determined that the best neighborhood for the President and his family would be **Flatiron**.