

# Finding the Optimal Location to Open a Martial Arts Gym in NYC

D. Berry – 06/08/2020

## 1.0 Introduction

### 1.1 Background & Interest

Martial Arts are “codified systems and traditions of combat practiced for a number of reasons such as self-defense; military and law enforcement applications; competition; physical, mental and spiritual development; and entertainment or the preservation of a nation's intangible cultural heritage.” They are key to a healthy society, self-discipline, and self-defense.

Many Martial Arts Gyms are successful, but there are never enough gyms, especially in a urban city like NYC. Moreover, there are many kinds of martial arts, and thus, many opportunities for bringing a new style to a neighborhood.

### 1.2 Problem

In this project we will try to find an optimal location for a gym. Specifically, this report will be targeted to stakeholders interested in opening a Martial Arts Gym in New York, NY.

Since there are a lot of gyms in NYC we will try to detect locations that are not already crowded with gyms. We are also particularly interested in areas with no Martial Arts Gyms in the vicinity. We would also prefer locations as close to city center as possible, assuming that first two conditions are met.

We will use data science techniques to generate a few most promising neighborhoods based on this criteria. Advantages of each area will then be clearly expressed so that best possible final location can be chosen by stakeholders.

## 2.0 Data Acquisition & Cleaning

Based on definition of our problem, factors that will influence our decision are:

- number of existing gyms in the neighborhood (any type of gym)
- number of and distance to Martial Arts Gyms in the neighborhood, if any
- distance of neighborhood from city center

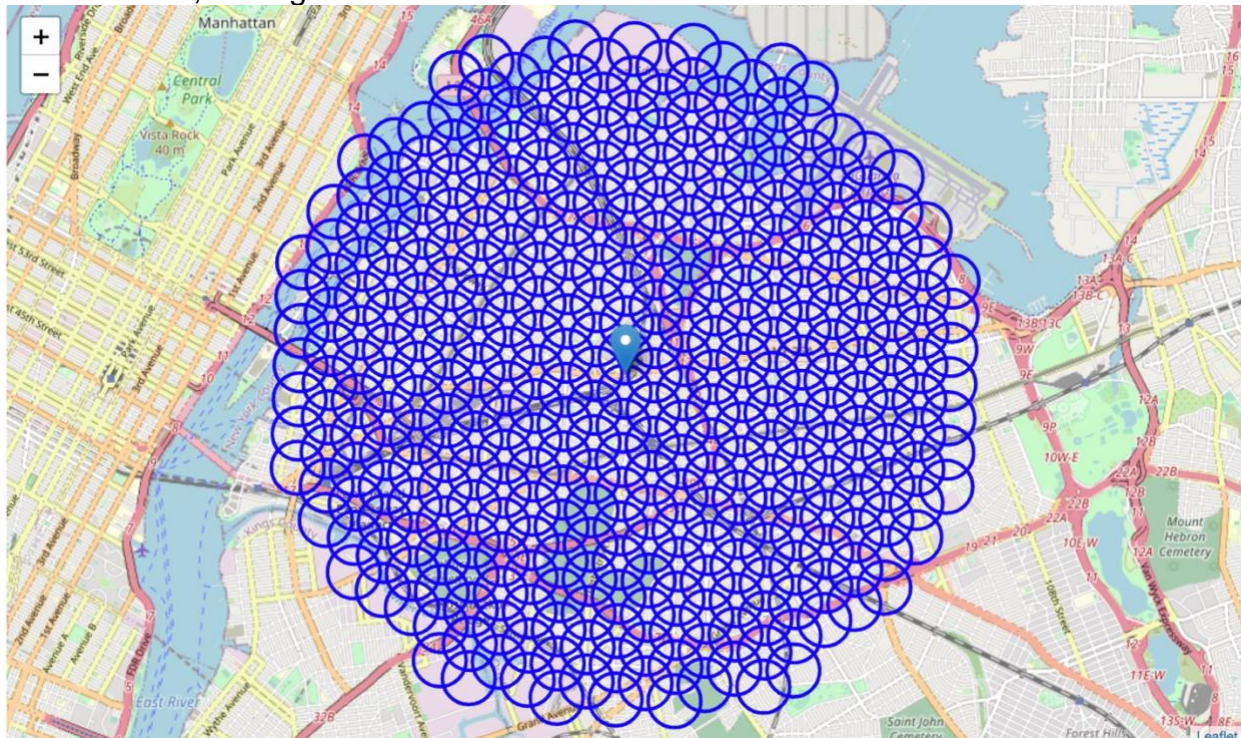
Regularly spaced grids of locations were used, centered around city center, to define our neighborhoods.

The following data sources will be needed to extract/generate the required information:

- centers of candidate areas will be generated algorithmically and approximate addresses of centers of those areas will be obtained using Google Maps API reverse geocoding
- number of Gyms and their type and location in every neighborhood will be obtained using the Foursquare API
- coordinates of the NYC center will be obtained using Google Maps API geocoding

## 2.1 Gathering the Data - Constructing Location Grids

First, around the center of NYC, “Woodside, NY”, we create a grid of area candidates, equally spaced, centered around the city center and within ~12km from the center. Our neighborhoods will be defined as circular areas with a radius of 300 meters, so our neighborhood centers will be 600 meters apart. This grid area is determined **algorithmically, using geocoding**. We then back-calculate the addresses at the center of each circle, storing these in a database.



## 2.2 Collecting Venue Data using the Foursquare API

In order to gather information about the venues in our pre-defined areas of interest, we will use Foursquare, a robust and data-rich location provider. Since we're interested in venues in 'Gyms / Fitness' category, included all the subcategories of specific 'Martial Arts Gym' category, as we need info on Martial Arts Dojos in the neighborhood.

Using our developer credentials, and a pre-defined radius, we then searched for all gyms in the radius of each address in our database. The result:

- Total number of gyms: **258**

- Total number of Martial Arts Dojos: **33**
- Percentage of Martial Arts Dojos: **12.79%**
- Average number of gyms in neighborhood: **1.28**

### 3.0 Methodology

In this project we will direct our efforts on detecting areas of NYC that have low gym density, particularly those with low number of Martial Arts dojos. We will limit our analysis to area ~5km around city center.

In first step we have collected the required **data: location and type (category) of every gym within 6km from NYC center** (Woodside). We have also **identified Martial Arts Dojos** (according to Foursquare categorization).

Second step in our analysis will be calculation and exploration of '**gym density**' across different areas of NYC - we will use **heatmaps** to identify a few promising areas close to center with low number of gyms in general (*and* no Martial Arts Dojos in the vicinity) and focus our attention on those areas.

In third and final step we will focus on most promising areas and within those create **clusters of locations that meet some basic requirements** established in discussion with stakeholders: we will take into consideration locations with **no more than two gyms in radius of 200 meters**, and we want locations **without Martial Arts Dojos in radius of 400 meters**. We will present map of all such locations but also create clusters (using **k-means clustering**) of those locations to identify general zones / neighborhoods / addresses which should be a starting point for final 'street level' exploration and search for optimal venue location by stakeholders.

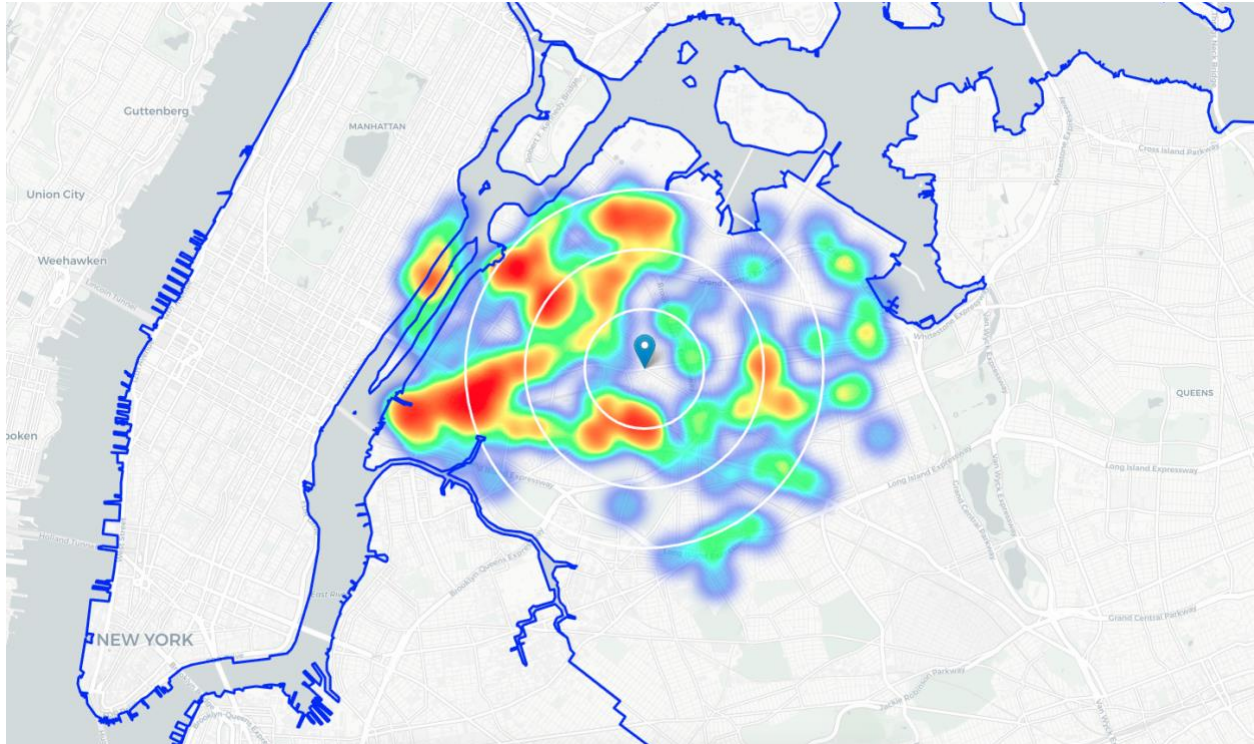
### 4.0 Data Analysis

With our list of gyms in and around our pre-defined addresses around the city center, to analyze the data, we will calculate and explore of '**gym density**' across different areas of NYC - we will use **heatmaps** to identify a few promising areas close to center with low number of gyms in general (*and* no Martial Arts Dojos in the vicinity) and focus our attention on those areas.

#### 4.1 Gym Locations Around the City Center

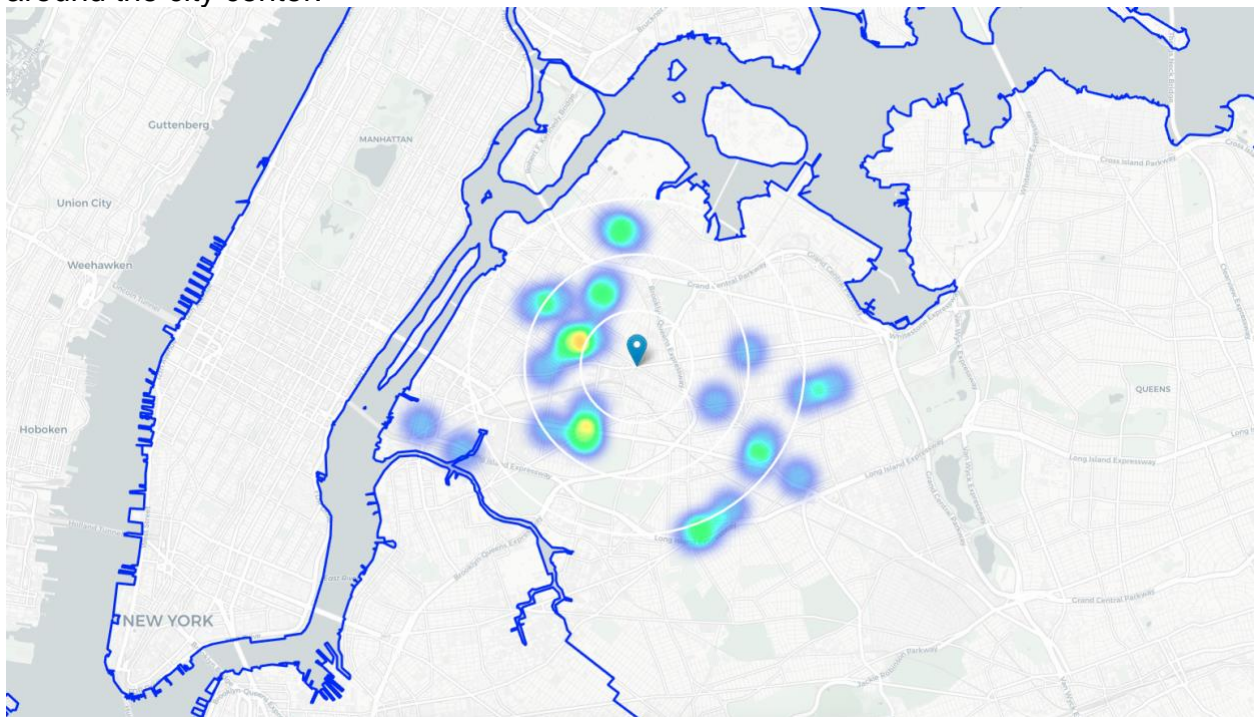
The heatmap below was constructed by calculating the density of gyms around the city center.





#### 4.2 Martial Arts Gym Locations Around the City Center

The heatmap below was constructed by calculating the density of martial arts gyms around the city center.

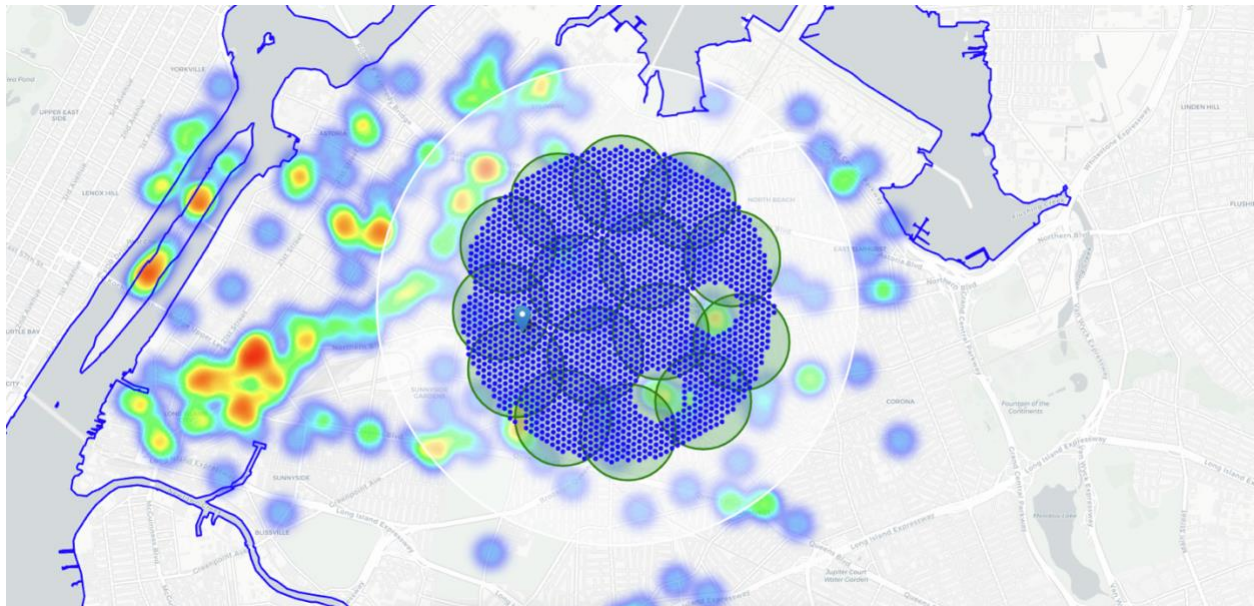


As observed above, there are several pockets of low-density Martial Arts Gyms areas to the east, north, and south of the center.

### 4.3 Determining Optimal New Location Areas - K-Means Cluster Analysis

Based on the data collected, we now have locations of potential candidates for a new Martial Arts Gym, at least based on nearby competition. What we have now is a clear indication of zones with low number of gyms in vicinity, and *no* Martial Arts gyms at all nearby.

Let us now **cluster** those locations (using K-Means clustering) to create **centers of zones containing good locations**. Those zones, their centers and addresses will be the final result of our analysis.



Our clusters represent groupings of most of the candidate locations and cluster centers are placed nicely in the middle of the zones 'rich' with location candidates.

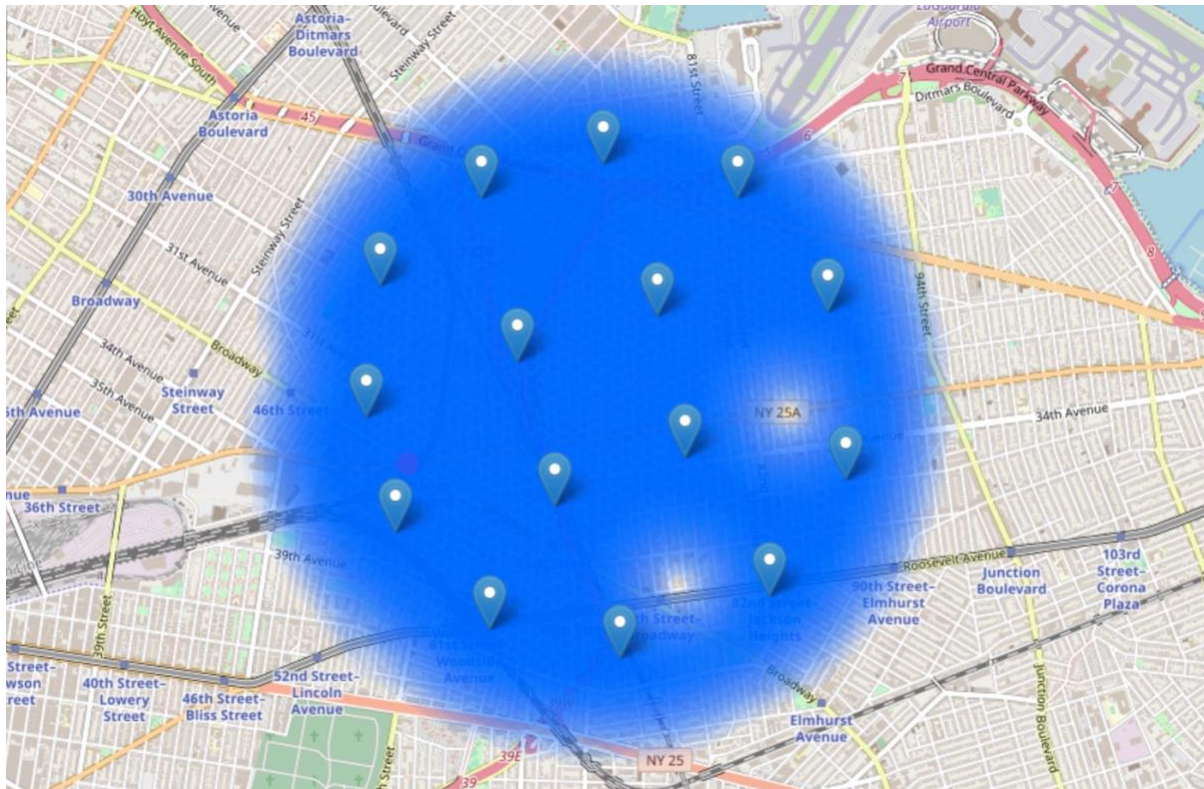
Addresses of those cluster centers will be a good starting point for exploring the neighborhoods to find the best possible location based on neighborhood specifics.

Next, we find the addresses of the center of these clusters using reverse geocoding. These addresses are overlaid on our cluster plot to show the candidate areas on a map.



# Addresses of centers of areas recommended for further analysis

28-45 47th St, Queens, NY 11103, USA	=> 1.4km from NYC Center
87-7 35th Ave, Jackson Heights, NY 11372, USA	=> 3.5km from NYC Center
39-36 63rd St, Woodside, NY 11377, USA	=> 1.5km from NYC Center
23-58 83rd St, Flushing, NY 11370, USA	=> 3.4km from NYC Center
75-25 31st Ave, East Elmhurst, NY 11370, USA	=> 2.3km from NYC Center
41-31 71st St, Flushing, NY 11377, USA	=> 2.3km from NYC Center
62-5 30th Ave, Woodside, NY 11377, USA	=> 1.2km from NYC Center
22-37 74th St, Flushing, NY 11370, USA	=> 2.8km from NYC Center
50-33 Broadway, Woodside, NY 11377, USA	=> 0.5km from NYC Center
40-15 81st St, Flushing, NY 11373, USA	=> 3.1km from NYC Center
30-50 88th St, Flushing, NY 11369, USA	=> 3.6km from NYC Center
23-25 49th St, Long Island City, NY 11105, USA	=> 2.2km from NYC Center
34-40 65th St, Flushing, NY 11377, USA	=> 1.2km from NYC Center
76-9 34th Ave, Jackson Heights, NY 11372, USA	=> 2.2km from NYC Center
37-32 55th St, Flushing, NY 11377, USA	=> 0.6km from NYC Center



## 5.0 Results & Discussion

Our analysis shows that although there are a decent number of gyms in NYC. There are pockets of low martial arts gym density fairly close to city center. The highest concentration of gyms was detected west of the city center, so we focused our attention to areas north, east and south.

The result of our analysis is 15 zones containing largest number of potential new gym locations based on number of and distance to existing venues - both gyms in general and Martial Arts Gyms particularly. This, of course, does not imply that those zones are actually optimal locations for a new gym. The purpose of this analysis was to only provide info on areas close to the NYC center but not crowded with existing gyms. Recommended zones should therefore be considered only as a starting point for more detailed analysis which could eventually result in location which has not only no nearby competition but also other factors taken into account and all other relevant conditions met.

## **6.0 Conclusion**

The purpose of this project was to identify NYC areas close to the city center with a low number of gyms, and in particular, a low number of Martial Arts gyms, in order to aid stakeholders in narrowing down the search for optimal location for a new Martial Arts Dojo in NYC. By calculating gym density distribution from Foursquare data we have first identified general boroughs that justify further analysis (south, north, and east of the city center in Woodside, NYC), and then generated extensive collection of locations which satisfy some basic requirements regarding existing nearby gyms. Clustering of those locations was then performed in order to create major zones of interest (containing greatest number of potential locations) and addresses of those zone centers were created to be used as starting points for final exploration by stakeholders.

The final decision on optimal gym location will be made by stakeholders based on specific characteristics of neighborhoods and locations in every recommended zone, taking into consideration additional factors like attractiveness of each location (proximity to park or water), levels of noise / proximity to major roads, real estate availability, prices, social and economic dynamics of every neighborhood etc.