Report: Finding locations to open a gym in Brooklyn, NY

Jibin Joseph

25/09/2019

1. Introduction / Business Problem

A fitness club group is interested in opening their gym/fitness center in Brooklyn, NY. This project report is for the director board of the fitness club to suggest potential gym/fitness center locations closer to the city center of Brooklyn, NY & away from other gym/fitness centers, Boxing clubs or Gym pools.

There are several gym & fitness centers already operating in the Brooklyn area. **Our goal is to identify locations within 5km from the Brooklyn city center and 2.5km away from an existing Gym or Fitness club. We will leverage the Foursquare Places API to find the candidate neighborhood centers for the Gym.**

The analysis would be helpful to the director board of the company in finding the best location for their endeavor. When with the conditions are met, the suggested locations will be attractive to the people living in the neighborhoods to avail the facilities closer to them.

2. Data

We have to find the dataset with neighborhoods, latitude, longitude of Brooklyn. We can explore each neighborhood to find the most common venue categories in each neighborhood using the explore function of **Foursquare Places API**. Once the venues are identified, we can filter the venue categories to find the existing gym/fitness centers in each neighborhood and use the data for further analysis.

Fortunately, we have a dataset available free in the New York University's spatial data repository **geo.nyu.edu.** The dataset is named 'New York City Neighborhood Names' in JSON format consists of the boroughs, neighborhood names, geo coordinates, etc of New York. Here are the links to the New York neighborhood dataset:

Description: https://geo.nyu.edu/catalog/nyu-2451-34572
Download GeoJSON: https://geo.nyu.edu/catalog/nyu_2451_34572

As the dataset is already available in a structured format with all available features. The dataset will save a substantial amount of time we spent on data pre-processing, formatting, Normalization, feature selection, etc (Known as data wrangling). But while exploring the venues in each neighborhood & filtering the gym/fitness center, we have to add important

features such as coordinates of the location in UTM cartesian coordinate system (X/Y coordinates in meters) and the distance from the Brooklyn city center to the venue.

We will create python functions to convert between WGS84 spherical coordinate system (latitude/longitude degrees) and UTM Cartesian coordinate system (X/Y coordinates in meters) as well as finding the distance of each Gym/Fitness center from the city center.

3. Visualization

3.1 Visualize the data we have

Let's visualize the Brooklyn neighborhood locations using the data we have using python's **folium** library for map rendering. The following figure (Figure - 1) will show the map with Brooklyn city center marked, and the candidate neighborhood centers in blue circles.

Fig – 1 Brooklyn city center location & neighborhoods

3.1 Explore neighborhoods using Foursquare

Using Foursquare Places API's 'explore' endpoint, get a list of recommended venues in each neighborhood location. We are looking for 'Gym' or 'Fitness center' category. We will filter the venue categories and shortlist the venues belongs to 'Gym' or 'Fitness center. Let's visualize all the identified gym locations using folium. (Fig - 2)

Now, we have all the Gym/Fitness centers in Brooklyn exactly in the vicinity of the neighborhood centers. We can use this data for further analysis to find optimum locations for a new gym with preconditions on the distance to the Brooklyn city center & from an existing gym, fitness center or boxing club. We will explain this in greater detail in the following section

Fig – 2 Gym/Fitness Centers in Brooklyn

4. Exploratory Data Analysis

Our main goal is to identify locations to open a new gym in Brooklyn closer to the city center with no gym or fitness clubs nearby. We are focusing our interest to 5km around the city center and we need our new locations to be at least 2.5km away from an existing gym or fitness center. And it is clear that we are looking for areas with lower density of gym/fitness centers

4.1 Calculation of the cartesian coordinates & distance from city center

We have listed the gyms in the vicinity of neighborhood centers with venue name, latitude, longitude & category as features. We need to convert the longitude & latitude (spherical) to the cartesian coordinate system (X/Y coordinates in meters) to calculate the distance from the city center & between the nearest gym or fitness center. We will use a python function

to calculate cartesian coordinates in meters & distance from city center. Now we have the gym dataset with features venue name, latitude, longitude X, Y & Distance from the city center in Brooklyn.

4.2 Heatmap

One of the best ways to identify the current density of gym/fitness centers is to use a **heatmap**. The heatmap will help us to identify a few promising areas with a low number of gyms in general. And of course, we are only focusing on an area of 5km around the Brooklyn city center. (Fig - 3)

Fig – 3 Heatmap: Gym/Fitness Centers in Brooklyn

Here we can see the pockets from the heatmap where low density of gym/fitness centers in our focus area.

4.3 New location candidates

As we already know, 5km around the city center is our target area. Let's create denser grid of location candidates in our area of interest. Let's make the location candidates 200m apart from each other.

We will generate the location candidates using our python function and save the latitude and longitude of the location candidates in the candidate location data frame. We will also

calculate the cartesian coordinates in meter & distance to the nearest gym using python functions and add in to the candidate location dataframe.

4.4 Filter the candidate location data

As our goal states, we are only interested in locations which doesn't have a gym/fitness center closer. We will filter the candidate location dataframe and list the only locations with at least 2.5km to the nearest gym/fitness center within our interest area (5km from the city center).

Fig – 4 Location candidates

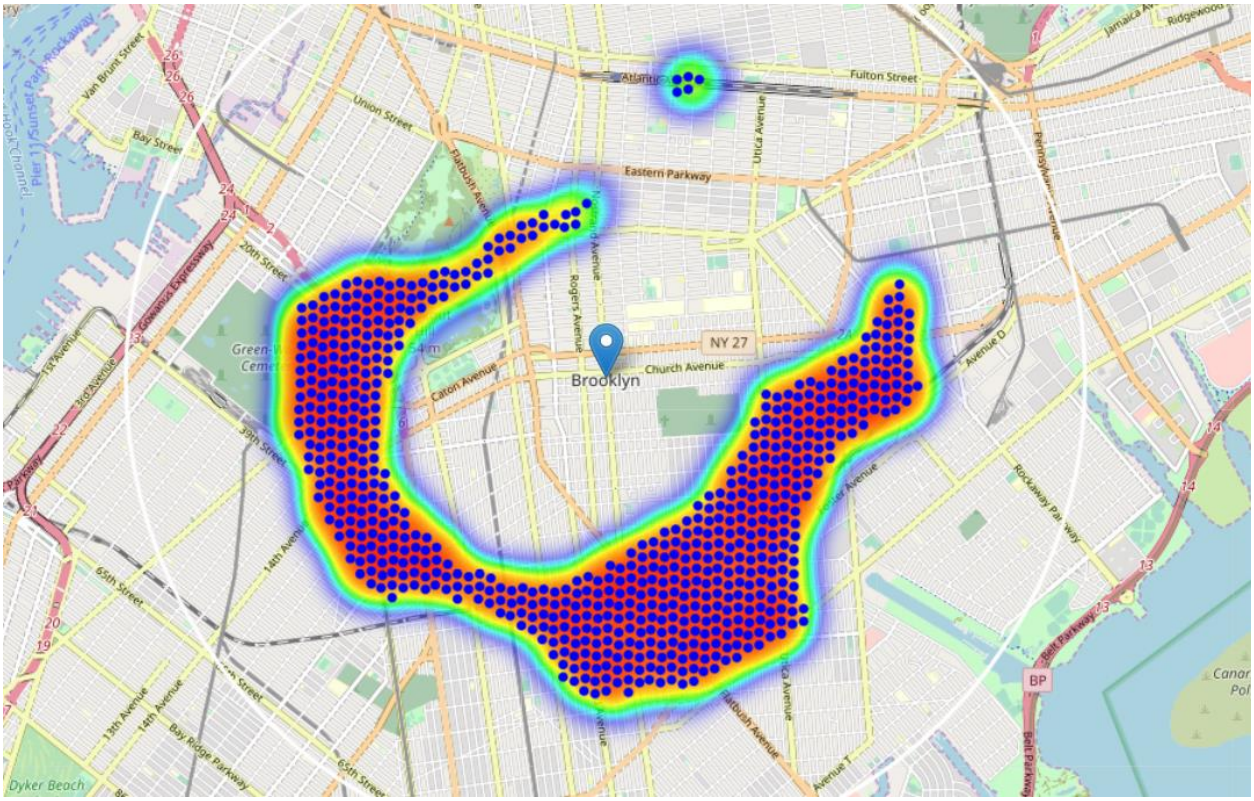

Fig – 5 Heatmap: Location candidates

Look at the generated location candidates superimposed in the map using python's folium library & the candidate location dataset we already generated. (Fig - 4 & Fig - 5).

5. Modeling

We have the clear indication of zones with low number of gym/fitness centers in vicinity. Let's apply our machine learning model on those locations to create centers of zones containing good locations. Those zone centers will be the result of our analysis

We have already identified this as an **unsupervised learning problem**. Having **unlabeled data**, we should apply a clustering model to segment the locations. Here we are using **K-Means clustering** model. It's one of the simplest clustering models and it is vastly used for clustering in many data science applications, especially useful if you need to quickly discover insights from unlabeled data.

Let's apply the K-Means clustering model on the candidate locations data. The figure (Fig - 6) shows the resulted grouping of location candidates. The centers (Fig - 7) are placed in the middle of the zone's dense with candidate locations.

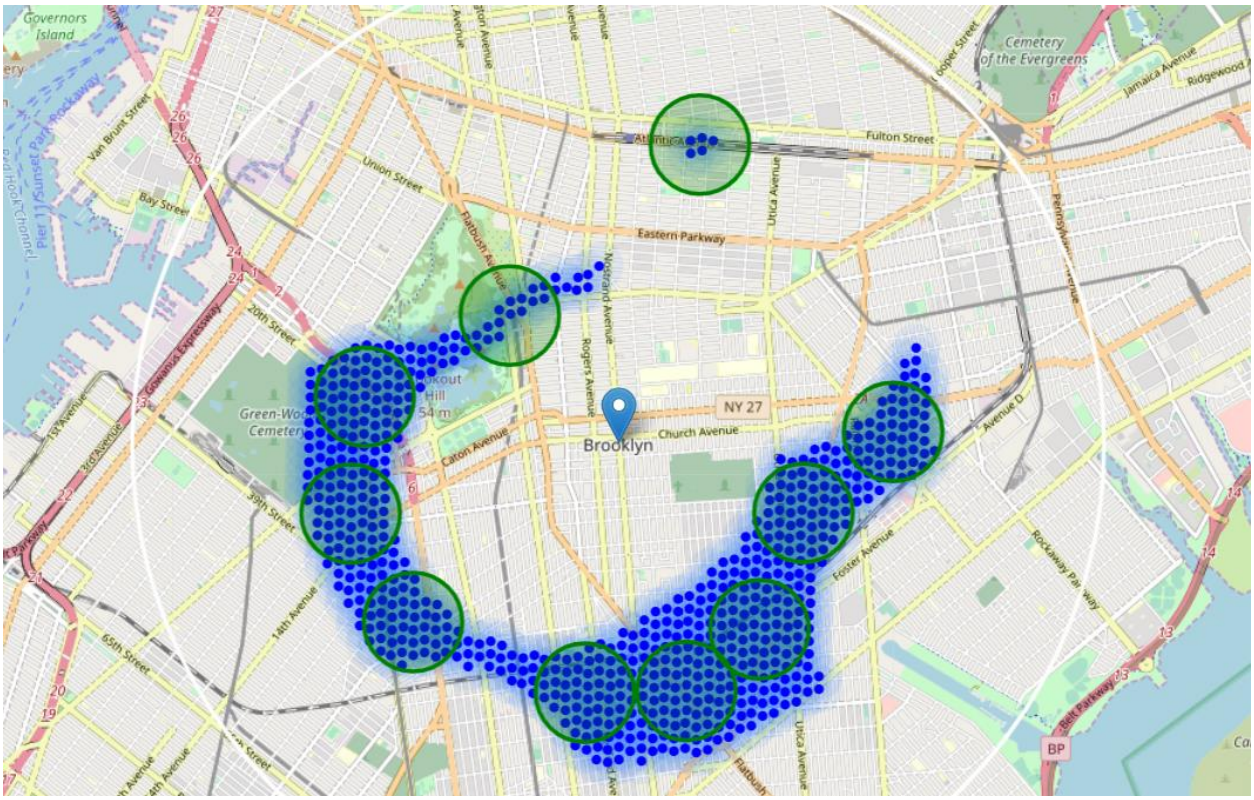

Fig – 6 Clusters of Location candidates

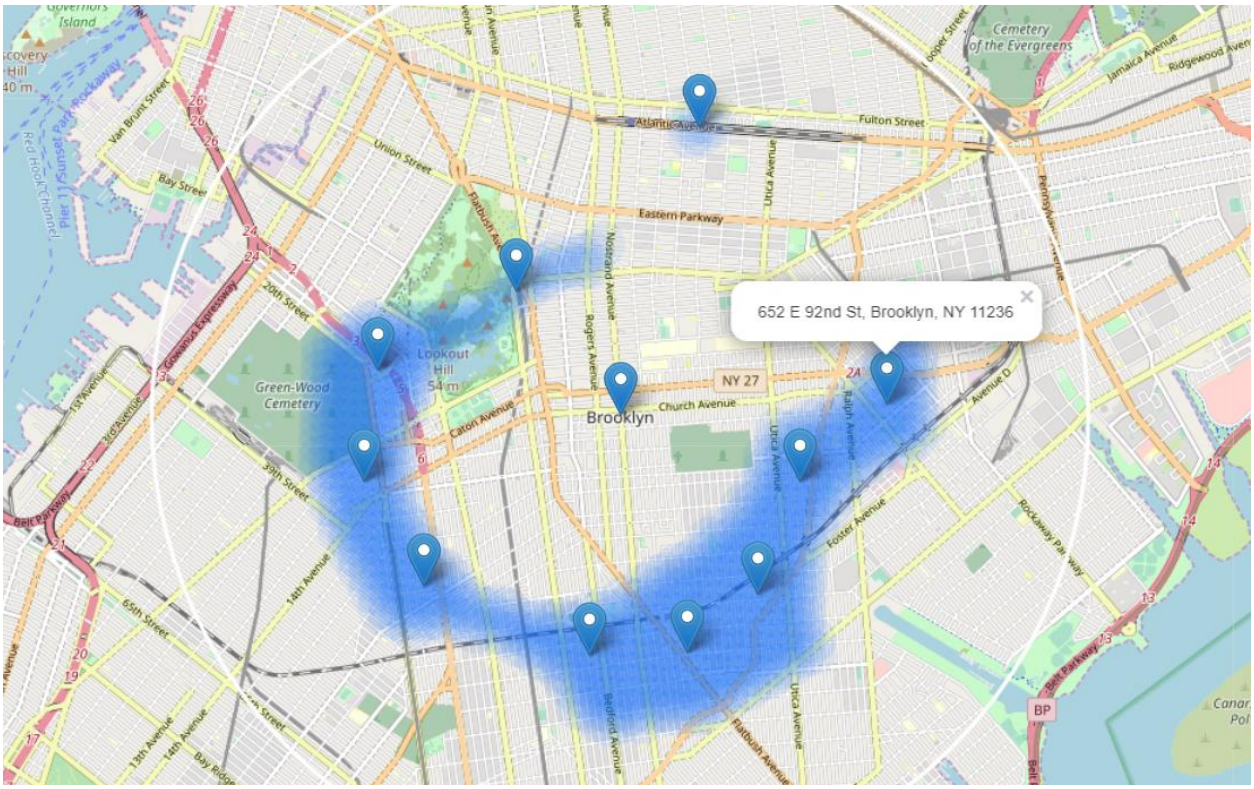

Fig – 7 Locations for new gym/fitness center

Let's find the addresses of the locations using **Google Map API reverse geocoding**. These are the best possible locations suggestion we are making (Fig - 8) for the new gym/fitness center. This will help the management to find the best possible location based on the neighborhood specifics.

```
============================================================
Addresses of centers of areas recommended for further analysis
============================================================

39 Ocean Ave, Brooklyn, NY 11225                  => 2.6km from Brooklyn center
3602 Avenue J, Brooklyn, NY 11210                 => 4.0km from Brooklyn center
638 E 53rd St, Brooklyn, NY 11203                 => 3.1km from Brooklyn center
401 Avenue F, Brooklyn, NY 11218                  => 4.3km from Brooklyn center
48 Clara St, Brooklyn, NY 11218                   => 4.4km from Brooklyn center
4524 Glendale Ct, Brooklyn, NY 11234              => 3.7km from Brooklyn center
652 E 92nd St, Brooklyn, NY 11236                 => 4.3km from Brooklyn center
1562 Atlantic Ave, Brooklyn, NY 11213             => 4.9km from Brooklyn center
593 20th St, Brooklyn, NY 11218                   => 4.1km from Brooklyn center
951 E 23rd St, Brooklyn, NY 11210                 => 3.9km from Brooklyn center
```

Fig – 8 The resulted 10 location addresses

We have identified 10 addresses representing centers of zones. The zones are already in the with low number of gyms nearby and closer (around 5km) to the Brooklyn city center. This concludes our analysis and I strongly believe the identified locations will be a good starting point for exploring area neighborhoods in search of the best gym/fitness center location in Brooklyn, NY.

6. Results & Discussion

Our analysis clearly identifies areas with lower density of gym/fitness centers in Brooklyn. We focused on an area of 5km around Brooklyn as we need the gym/fitness center location closer to the Brooklyn city center. We have created a dense grid of location candidates spaced 200m apart and applied filtering on the dataset to list the only location candidates with no other gym or fitness center present within 2.5km.

We have used machine learning model **K-Means Clustering** cluster the location candidates into create zones of interest which contain dense location candidates. We used **Google Map API reverse geocoding** to generate the addresses of the 10 zone centers

The result of our analysis in the 10 location (Fig - 8) which are the centers of the location candidate zones. The result is based on target area – around 5km from Brooklyn city center - and 2.5km to a nearest gym or fitness center. Our goal is to provide this locations to the director board of the fitness club. And based on the analysis I recommend this 10 locations in your pursuit to identify the best location for a gym/fitness center. But it is entirely possible that there is a very good reason for no gym/fitness centers in those areas.

Regardless of the competition, the reasons can make those areas unsuitable for a gym/fitness center. These recommended locations will be a good starting point for further analysis and finding the best location for the new gym/fitness center.

7. Conclusion

We have identified 10 locations for a new gym/fitness center with low density of gym/fitness centers based on our goal to assist the director board of a fitness club. By finding the existing gym/fitness centers using **Foursquare API** we identified areas with low density of gym/fitness centers. We generated a number of locations in these areas based on the criteria and listed the location candidates for a new gym/fitness center.

We performed clustering of those locations and grouped the location candidates into zones based on the density of location candidates. The address of those zone centers are the 10 locations we have identified as potential locations for a new gym/fitness center.

The final decision on the new location will be made by the director board of the fitness club. They should consider specific characteristics of the neighborhoods of each recommended location. Additional factors like residential areas, traffic, roads, social & economic dynamics of the neighborhoods should be taken into consideration.