# Access to Parks and Fitness Facilities in Manhattan

Racine Nassau
September 8, 2020

## 1. Introduction

According to a study published in Mayo Clinic Proceedings, less than 3% of Americans live a healthy lifestyle. One of the key factors characterizing a healthy lifestyle is exercise, but many Americans lack easy access to parks and fitness facilities. Among the 100 largest metro areas in the US, only 54% of people live less than half a mile from a park.

This issue is explored by investigating the access to parks and fitness facilities across different neighborhoods in Manhattan. The Foursquare API is used to collect data on the number of parks, gyms, sports facilities such as basketball and tennis courts, and fitness classes in a given area. Using this data, the neighborhoods are grouped into clusters. The demographics of each cluster are then compared to explore the relationship between factors such as income and access to green spaces and fitness facilities.

This type of analysis could be used to identify city areas with inequitable access to parks and fitness facilities, and may be used to help city planners address these gaps.

## 2. Data

In order to explore access to green spaces and fitness facilities and its relationship to socio-economic factors, the following datasets are needed:

1. Demographic data by neighborhood
2. GeoJSON file of Neighborhood Tabulation Areas
3. Longitude and latitude points representing each neighborhood
4. A list of nearby parks and sports/fitness facilities for each neighborhood

### 2.1 Dataset #1
Demographic and socio-economic data is collected by the United States Census Bureau and is organized into areas called *census tracts*, generally encompassing 2,500-8,000 people. This Kaggle dataset provides such data for NYC, but does not specify which neighborhood each census tract belongs to. Each neighborhood consists of multiple census tracts, so data to map these census tracts to their respective neighborhoods is also required. This 2010 Census Tract to 2010 Neighborhood Tabulation Area Equivalency CSV file specifies the neighborhood each census tract belongs to. Using this dataset, these census tracts are converted to their respective neighborhoods and the overall demographics of each neighborhood are calculated.

## 2.2 Dataset #2

Dataset #1 provides the population of each neighborhood, but in order to calculate the population density of each neighborhood, the area of each neighborhood is needed as well. The Neighborhood Tabulation Areas (NTAs) downloaded from the [NYC Department of City Planning website](#) provide this information. These NTAs can also be used to visualize Manhattan neighborhoods on a map, and give the boundaries of each neighborhood.

## 2.3 Dataset #3

Before collecting Foursquare location data for Dataset #4, representative latitude and longitude points of each neighborhood are required. Dataset #2 maps each neighborhood as a polygon. This is useful because the centroid of a polygon can be a good approximation for representative geographical coordinates. Calculating the centroid of each neighborhood, representative latitude and longitude points of each neighborhood are obtained.

## 2.4 Dataset #4

This dataset is built using Foursquare location data. The Foursquare API is used to find parks, gyms, sports facilities, and fitness classes within 500 meters of the geographical coordinates representing each neighborhood. Rather than comparing the number of these "venues" in each neighborhood, it seemed more meaningful to explore how many venues a resident of a particular neighborhood could walk to within 500 m, making it easier to compare across neighborhoods of different sizes.
This project combines and cleans data from different sources to get geographical, demographic, and economic data for each neighborhood in Manhattan.

Foursquare location data can be collected for different "categories" defined by the Foursquare API. There are two main categories of interest. The first one is the *Athletics & Sports* category, which refers to venues such as gyms, basketball courts, soccer fields, yoga studios, and other fitness classes. The second category is the *Park* category. Both of these categories consist of multiple sub-categories, and not all of them are relevant to this analysis. The relevant sub-categories are filtered and grouped into four new categories as follows:

- **Parks** (includes parks, gardens, trails, and playgrounds)
- **Gyms** (includes different types of gyms)
- **Sports Facilities** (includes different types of sports fields and facilities, such as basketball courts and soccer fields)
- **Fitness Classes** (includes yoga studios, martial arts schools, and Pilates studios)

The number of venues in each of these four categories is counted for each neighborhood, providing a dataset structured as in *Table 1*.

*Table 1: Number of park and fitness venues in each neighborhood by category*

| Neighborhood | Fitness Class | Gym | Park | Sports Facility |
|---|---|---|---|---|
| **Battery Park City-Lower Manhattan** | 2 | 92 | 29 | 2 |
| **Central Harlem North-Polo Grounds** | 2 | 9 | 3 | 6 |
| **Central Harlem South** | 6 | 9 | 3 | 5 |

## 3. Methodology

To verify that Dataset #3 gives reasonable geographical coordinates to represent each neighborhood, these coordinates are plotted along with their respective Neighborhood Tabulation Areas. To explore the population density across Manhattan, these Neighborhood Tabulation Areas are shaded according to population density.
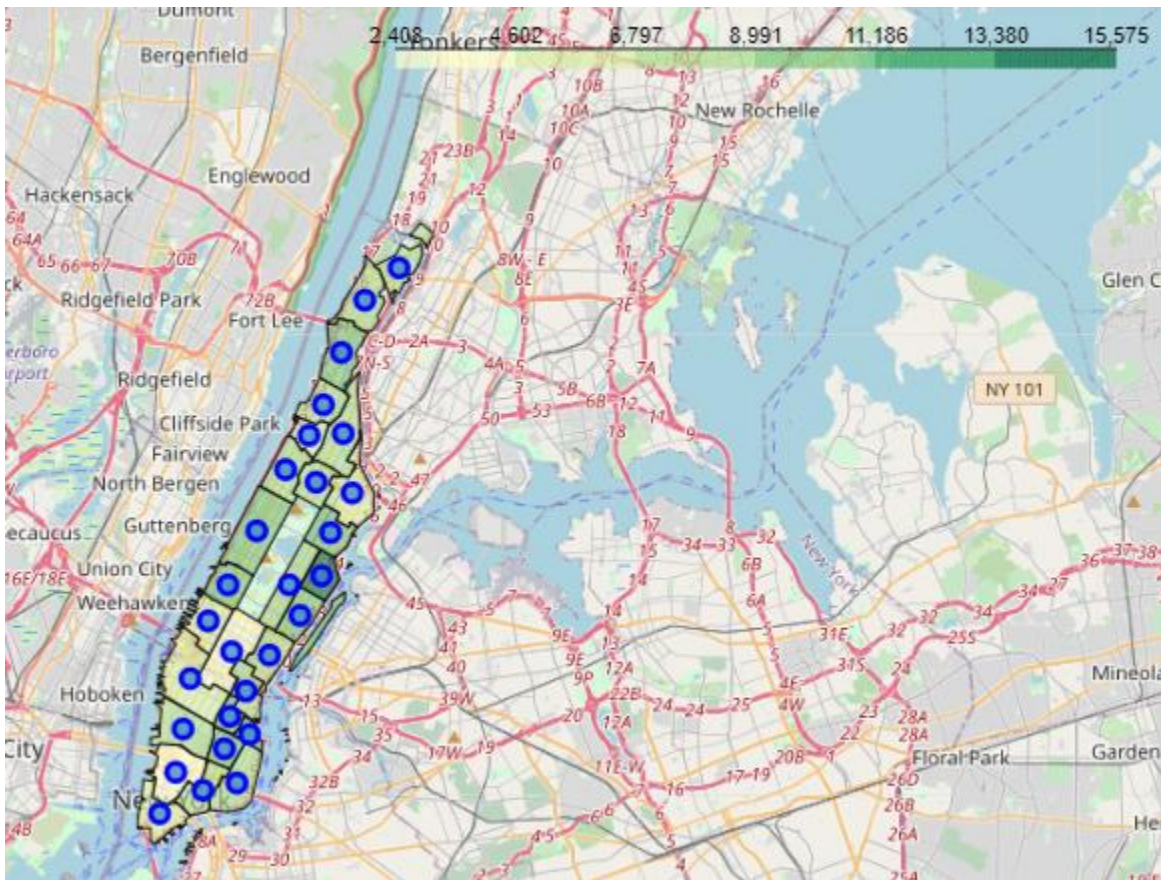


*Figure 1: Population density map plotted with representative neighborhood coordinates*

As shown in *Figure 1*, the calculated latitudes and longitudes for each neighborhood are centered reasonably to represent their respective neighborhoods.

Exploring the demographic/socio-economic data derived from the US Census data, the median income across Manhattan neighborhoods is highly variable. East Harlem North has a median income of just

$27,909, whereas Upper East Side-Carnegie Hill has a median income of $164,661. Similarly, poverty rates in Manhattan range from 5% to 37%. Even neighborhoods near each other can display huge inequalities in wealth.

In order to understand which neighborhoods are most similar in terms of access to gyms, parks, sports facilities, and fitness classes, the neighborhoods are grouped into clusters using k-means clustering. Since the appropriate number of clusters to represent this data is still unknown, the *elbow method* is used, running k-means over a range of different numbers of clusters.
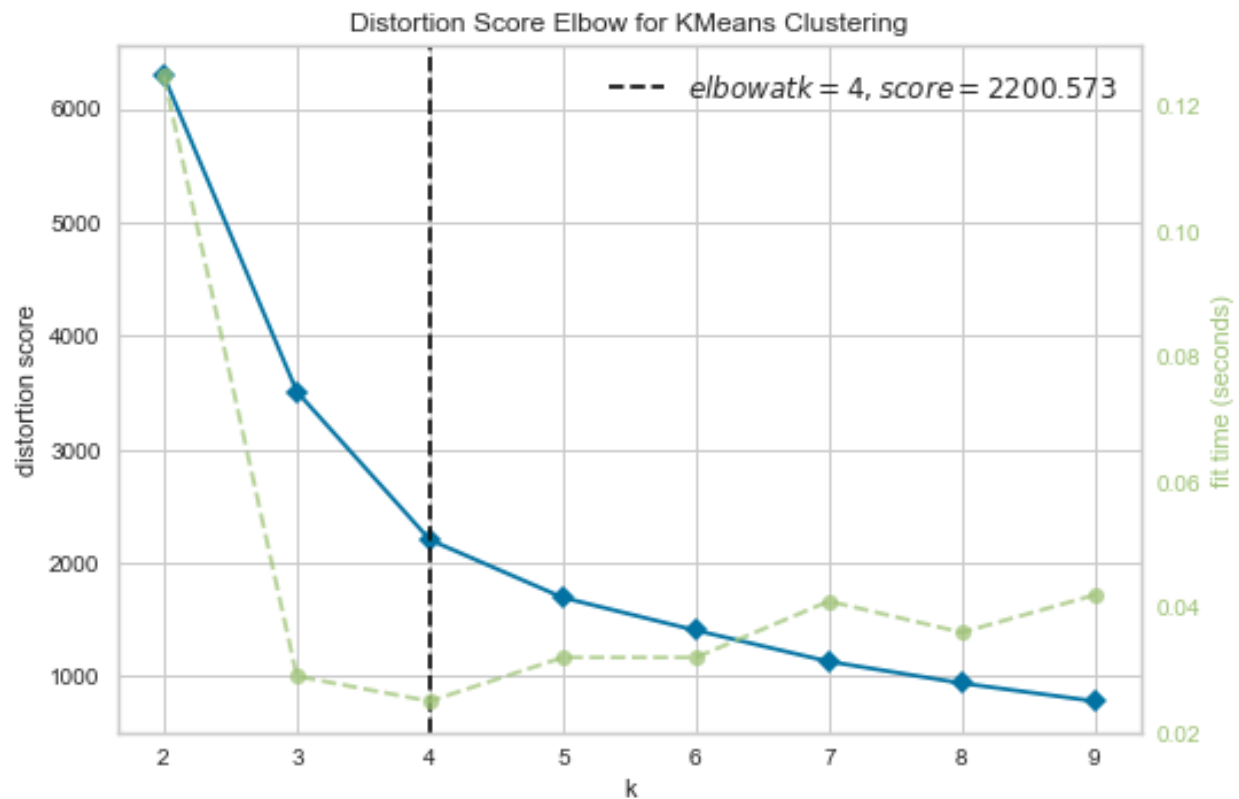


*Figure 2: Distortion score elbow for k-means clustering*

As shown in *Figure 2*, the optimal number of clusters to represent this data is four, so the neighborhoods are grouped into four clusters for further analysis.

## 4. Results

Now that the neighborhoods are grouped into four clusters, the clusters can be visualized along with socio-economic data for each neighborhood. Three socio-economic factors (population density, income, and poverty rates) are chosen as features of interest to explore further. These features along with the neighborhood clusters are mapped in *Figure 3*.
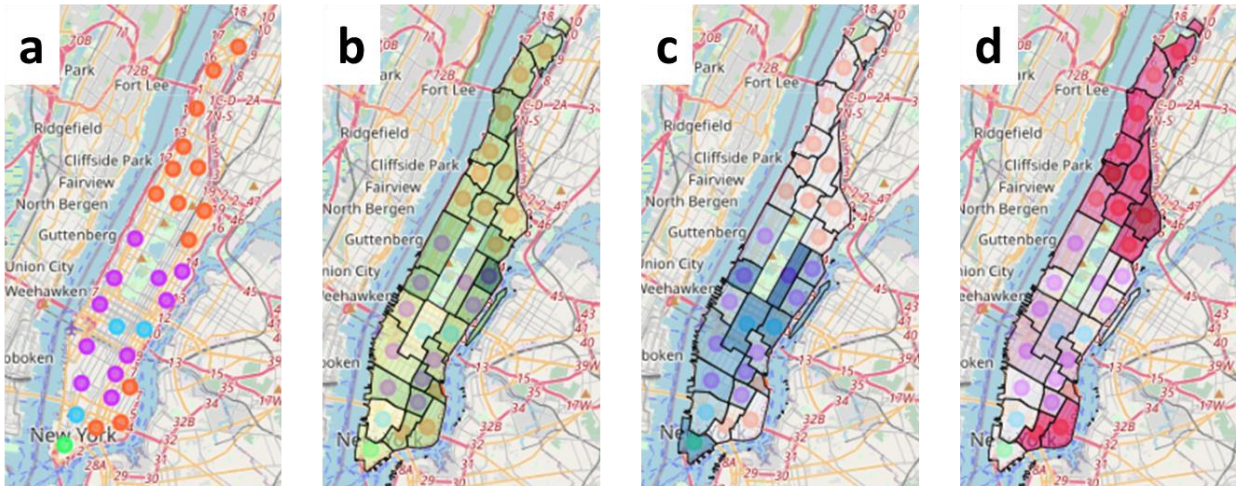
*Figure 3: Map of neighborhood clusters (a), population density (b), income (c), and poverty rates (d)*

At first glance of *Figure 3*, it seems that the orange cluster falls into the neighborhoods with the highest poverty rates and lowest income. To examine these clusters further, these three socio-economic factors for each neighborhood are averaged over the clusters they belong to. These results are shown in *Table 2*.

*Table 2: Cluster breakdown by park/fitness access and socio-economic factors*

| Cluster Label | Fitness Class | Gym | Park | Sports Facility | Population Density (people/ sq mi) | Income ($) | Poverty Rate |
|---|---|---|---|---|---|---|---|
| **Orange** | 2.6 | 5.9 | 6.6 | 6.0 | 8202 | 45,241 | 27.3% |
| **Purple** | 8.6 | 26.4 | 9.5 | 7.6 | 9050 | 106,227 | 9.9% |
| **Blue** | 18.0 | 57.0 | 15.0 | 10.0 | 4728 | 107,442 | 9.7% |
| **Green** | 2.0 | 92.0 | 29.0 | 2.0 | 6050 | 135,073 | 7.4% |

## 5. Discussion

The information in *Table 2* provides insight into what defines each cluster. Each cluster will be analyzed and discussed in this section.

### 5.1 Orange Cluster

The neighborhoods in the orange cluster are characterized by low numbers of parks, gyms, sports facilities such as basketball courts, and fitness classes. The people living in these neighborhoods have low access to both public and private sports/fitness facilities, despite living in areas of relatively high population density. The orange cluster is characterized by far lower income than the other clusters, along with significantly higher poverty rates. This data suggests that Manhattan's poorest residents not only lack access to paid private fitness facilities such as gyms, but also to free public fitness/recreation areas such as parks and sports fields.

Geographically, the orange cluster falls completely into only two areas of Manhattan. Every neighborhood in northern Manhattan falls into the orange cluster. The remaining three neighborhoods of the orange cluster are centered around the Lower East Side. It could be recommended that city planners focus on these two geographical areas for future investments into parks and recreation.

### 5.2 Purple Cluster

The purple cluster has more sports/fitness facilities in every category than the orange cluster. However, the difference is especially great in the fitness class and gym categories. With income levels more than twice as high and poverty rates roughly one third those of the orange cluster, the residents of the purple cluster may have more discretionary income to spend on fitness related services such as gyms and yoga classes.

### 5.3 Blue Cluster

The neighborhoods in the blue cluster tend to be high income and have very high numbers of parks, gyms, sports facilities, and fitness classes. Although the population density of this cluster is significantly lower than the others, the residents of these neighborhoods enjoy high access to a variety of public and private sports/fitness facilities.

### 5.4 Green Cluster

The green cluster consists of only one neighborhood, Battery Park City-Lower Manhattan. What sets this neighborhood apart from the others is that it has significantly more gyms and parks than any other neighborhood, despite relatively low numbers of fitness classes and sports fields. This neighborhood is characterized by very high income compared to the other neighborhoods.

## 6. Conclusion

The neighborhoods with low income and high poverty rates have relatively low access to both public and private fitness/recreation areas. The data analyzed suggests that Manhattan's parks and public sports facilities are not distributed according to population density, but rather are more heavily concentrated in areas of higher income. If Manhattan has future resources to invest in new public parks and sports facilities, it may be more equitable to focus investments in the neighborhoods of the orange cluster.