GEOSPATIAL ANALYSIS

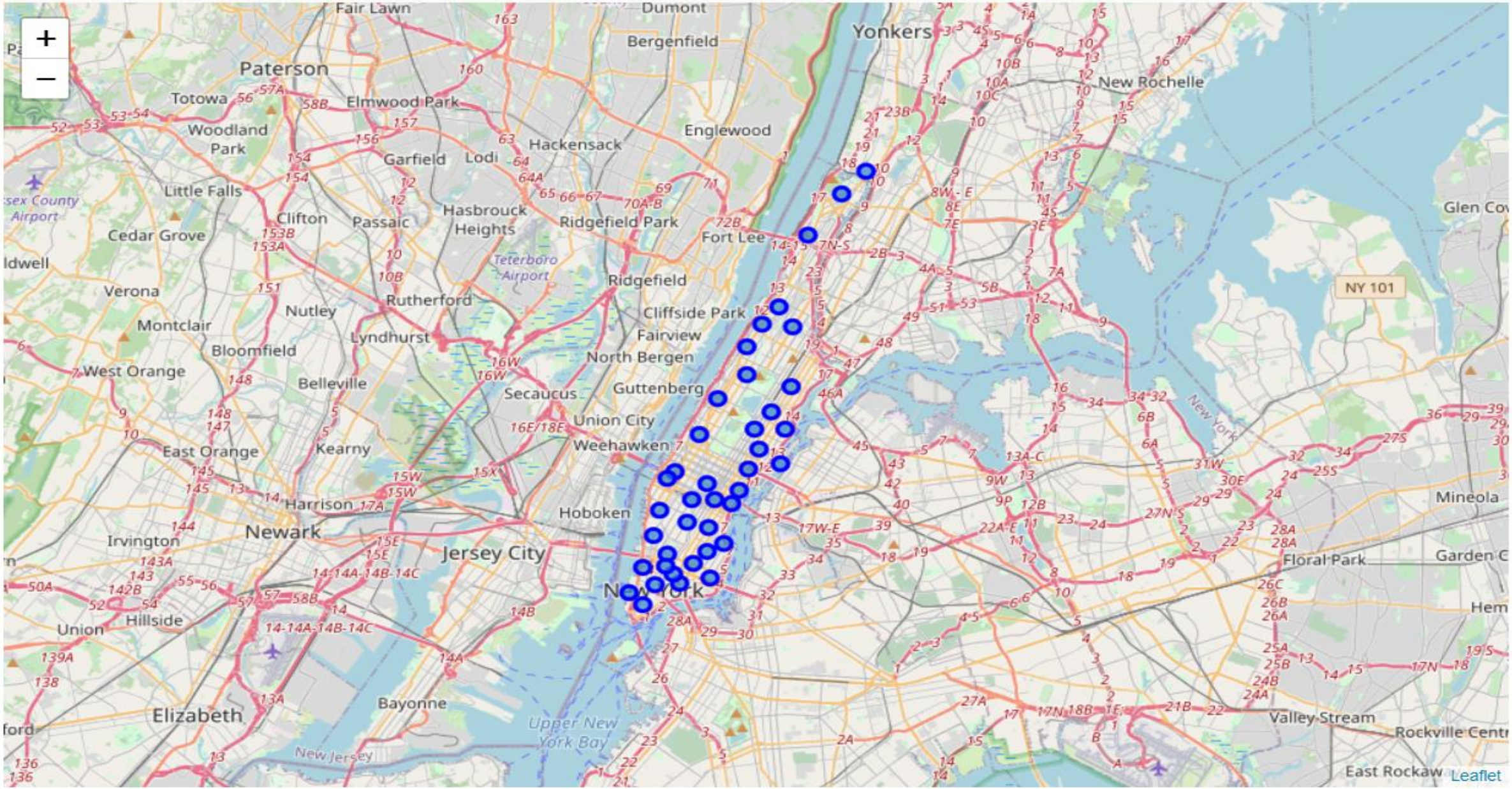
Finding the right spot to open up a franchise

Business Problem

- A business in the sports supplements industry wishes to expand.
- The task at hand is to find the right location to open a **supplement store** in Manhattan, NYC. This is crucial to the success of the supplement store franchise looking to expand.
- Many sports enthusiasts and gym goers from various ethnicities reside in New York. With a population north of 8 million, this is a prime city for business.
- The **right location** to open up a supplement store would be one where sports enthusiasts and gym goers are more likely to be found AND no supplement shops are found in close vicinity.

Data Requirements

- Details about the neighborhoods in Manhattan can be found on this → https://geo.nyu.edu/catalog/nyu_2451_34572 json file found on the web.
- All the neighborhoods can be found on the json file and placed onto a Dataframe.
- We'll analyse the venues in the neighborhood using Foursquare's API.
- The API returns all the venues along with the venue category
- Venues in each neighborhood that are frequented by sports enthusiasts and gym goers will be taken into account
- Existing supplement shops will also be taken into account


Fair Lawn
Dumont
Bergenfield
Yonkers
Paterson
Totowa
Elmwood Park
New Rochelle
Woodland Park
Englewood
Hackensack
Garfield
Lodi
Little Falls
Clifton
Passaic
Hasbrouck Heights
Ridgefield Park
Fort Lee
Cedar Grove
Teterboro Airport
Ridgefield
Verona
Nutley
Rutherford
Cliffside Park
Montclair
Lyndhurst
Fairview
Bloomfield
North Bergen
West Orange
Belleville
Secaucus
Guttenberg
Union City
East Orange
Kearny
Weehawken
Harrison
Hoboken
Irvington
Newark
Jersey City
New York
Hillside
Union
Elizabeth
Bayonne
Upper New York Bay
New Jersey
NY 101
Mineola
Floral Park
Valley Stream
Rockville Centre
East Rockaway
Leaflet

Selecting Venue Categories

We'll be working with the following venue categories:

- Athletics & Sports Venues
- Baseball Fields
- Basketball Courts
- Boxing Gyms,
- Climbing Gym
- Cycle Studios
- Gyms, Gym/Fitness Centres, Gym Pools
- Soccer Fields
- Sports clubs, Sporting Goods Shops
- Tennis courts
- Supplement Shops

	Neighborhood	Supplement Shop	Athletics & Sports	Baseball Field	Basketball Court	Boxing Gym	Gym	Cycle Studio	Gym / Fitness Center	Gym Pool	Soccer Field	Sports Club	Sporting Goods Shop	Tennis Court
0	Battery Park City	0.0	0.015873	0.0	0.0	0.0	0.047619	0.000000	0.000000	0.0	0.0	0.00	0.0	0.0
1	Carnegie Hill	0.0	0.000000	0.0	0.0	0.0	0.034091	0.000000	0.034091	0.0	0.0	0.00	0.0	0.0
2	Central Harlem	0.0	0.000000	0.0	0.0	0.0	0.021739	0.021739	0.043478	0.0	0.0	0.00	0.0	0.0
3	Chelsea	0.0	0.000000	0.0	0.0	0.0	0.010000	0.020000	0.000000	0.0	0.0	0.00	0.0	0.0
4	Chinatown	0.0	0.000000	0.0	0.0	0.0	0.000000	0.000000	0.000000	0.0	0.0	0.01	0.0	0.0

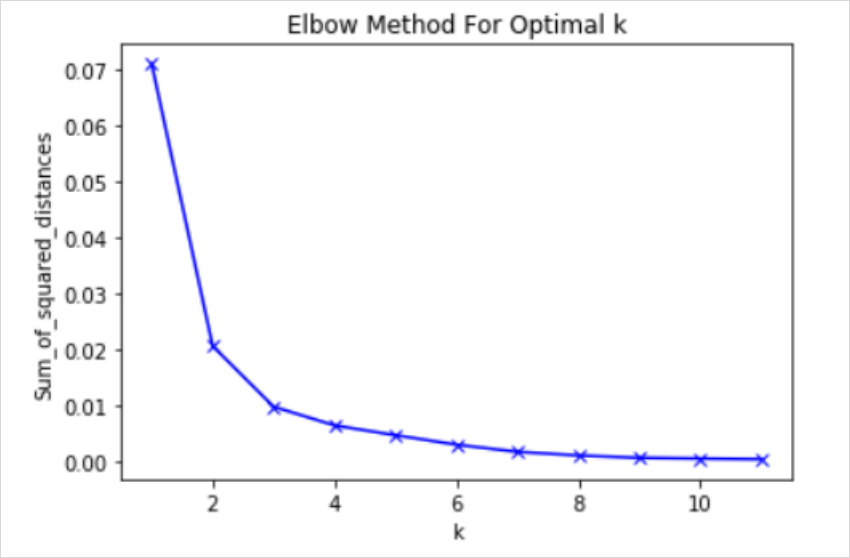
Methodology

- We'll analyse each neighborhood and take a **mean** of all sports facilities to quantify the **concentration** of each sports facility in each neighborhood.
- We have ignored all stadiums and other facilities which are used infrequently throughout the year and are reserved for tournaments.
- Further, we'll sum all column mean values to get a value on the **total** sports facility concentration in **each** neighborhood. This step will exclude supplement shops.

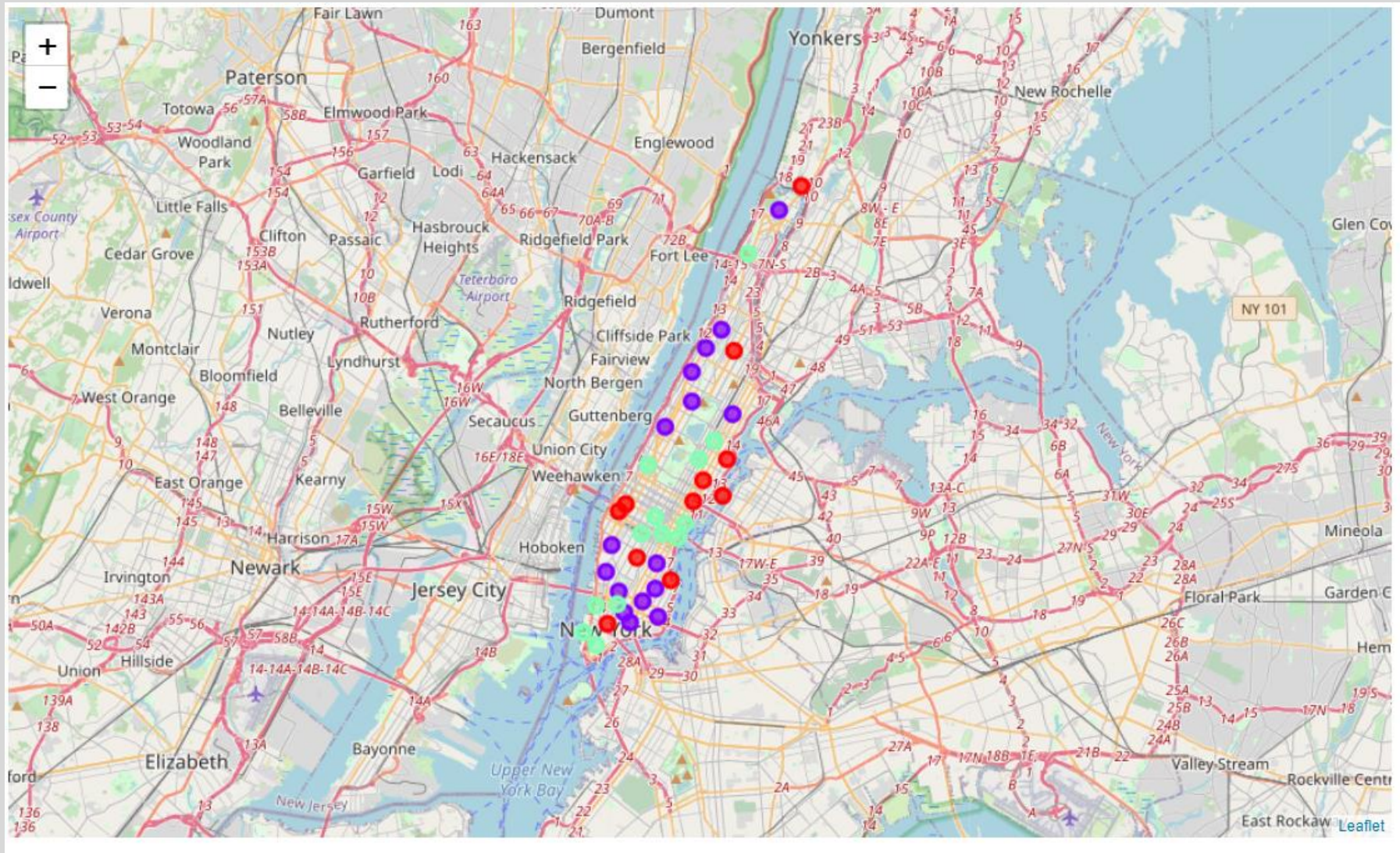
Neighborhood	Supplement Shop	Athletics & Sports	Baseball Field	Basketball Court	Boxing Gym	Gym	Cycle Studio	Gym / Fitness Center	Gym Pool	Soccer Field	Sports Club	Sporting Goods Shop	Tennis Court	Sum
Flatiron	0.000000	0.000000	0.000000	0.000000	0.000000	0.040000	0.020000	0.070000	0.000000	0.000000	0.01	0.020000	0.000000	0.160000
Hudson Yards	0.000000	0.017544	0.000000	0.000000	0.000000	0.035088	0.017544	0.052632	0.000000	0.000000	0.00	0.000000	0.000000	0.122807
Marble Hill	0.040000	0.000000	0.000000	0.000000	0.000000	0.080000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.120000
Roosevelt Island	0.000000	0.000000	0.000000	0.000000	0.000000	0.038462	0.000000	0.038462	0.000000	0.038462	0.00	0.000000	0.000000	0.115385
Stuyvesant Town	0.000000	0.000000	0.055556	0.000000	0.000000	0.000000	0.000000	0.055556	0.000000	0.000000	0.00	0.000000	0.000000	0.111111

KMeans Clustering

- We'll use the **KMeans** Machine Learning alogorithm to cluster neighborhoods based on the sports facilites and supplement shops present. The **elbow method** will be deployed to calculate the optimal k value.



- The optimal value for k is 3


Fair Lawn
Dumont
Bergenfield
Yonkers
Paterson
Totowa
Elmwood Park
Woodland Park
Englewood
New Rochelle
Hackensack
Garfield
Lodi
Little Falls
Clifton
Passaic
Hasbrouck Heights
Ridgefield Park
Fort Lee
Cedar Grove
Teterboro Airport
Ridgefield
Verona
Nutley
Rutherford
Cliffside Park
Montclair
Lyndhurst
Fairview
Bloomfield
North Bergen
West Orange
Belleville
Secaucus
Guttenberg
Union City
East Orange
Kearny
Weehawken
Harrison
Hoboken
Irvington
Newark
Jersey City
Union
Hillside
Elizabeth
Bayonne
Upper New York Bay
New Jersey
NY 101
Mineola
Floral Park
Valley Stream
Rockville Centr
East Rockaw
Leaflet

Results

Cluster 0

	Neighborhood	Sum	Supplement Shop
0	Marble Hill	0.120000	0.04
6	Central Harlem	0.086957	0.00
9	Yorkville	0.090000	0.00
10	Lenox Hill	0.100000	0.00
11	Roosevelt Island	0.115385	0.00
14	Clinton	0.100000	0.00
32	Civic Center	0.103093	0.00
34	Sutton Place	0.100000	0.00
37	Stuyvesant Town	0.111111	0.00
38	Flatiron	0.160000	0.00
39	Hudson Yards	0.122807	0.00

Cluster 1

	Neighborhood	Sum	Supplement Shop
1	Chinatown	0.010000	0.0
3	Inwood	0.000000	0.0
4	Hamilton Heights	0.000000	0.0
5	Manhattanville	0.000000	0.0
7	East Harlem	0.024390	0.0
12	Upper West Side	0.022727	0.0
17	Chelsea	0.030000	0.0
18	Greenwich Village	0.020000	0.0
19	East Village	0.000000	0.0
20	Lower East Side	0.020000	0.0
22	Little Italy	0.000000	0.0
24	West Village	0.020000	0.0
25	Manhattan Valley	0.023256	0.0
26	Morningside Heights	0.024390	0.0
27	Gramercy	0.000000	0.0
31	Noho	0.020000	0.0

Cluster 2

	Neighborhood	Sum	Supplement Shop
2	Washington Heights	0.058140	0.011628
8	Upper East Side	0.057471	0.000000
13	Lincoln Square	0.084211	0.000000
15	Midtown	0.080000	0.000000
16	Murray Hill	0.063158	0.000000
21	Tribeca	0.064103	0.000000
23	Soho	0.040000	0.000000
28	Battery Park City	0.063492	0.000000
29	Financial District	0.050000	0.000000
30	Carnegie Hill	0.068182	0.000000
33	Midtown South	0.050000	0.000000
35	Turtle Bay	0.040000	0.000000
36	Tudor City	0.066667	0.000000

Conclusion

- Analysing each cluster reveals that clusters 0 and 2 have a significant portion of sports facilities and closer inspection reveals that cluster 0 has a siginificant chunk of sports facilities.
- Meanwhile, cluster 1 has to be avoided like the plague as it has very few sports facilities; therefore, it's in the best interest of stakeholders to avoid cluster 1 and focus on cluster 0.

	Neighborhood	Sum	Supplement Shop
0	Marble Hill	0.120000	0.04
6	Central Harlem	0.086957	0.00
9	Yorkville	0.090000	0.00
10	Lenox Hill	0.100000	0.00
11	Roosevelt Island	0.115385	0.00
14	Clinton	0.100000	0.00
32	Civic Center	0.103093	0.00
34	Sutton Place	0.100000	0.00
37	Stuyvesant Town	0.111111	0.00
38	Flatiron	0.160000	0.00
39	Hudson Yards	0.122807	0.00

- Final decision has to be taken by stakeholders to determine which of the locations described above are ideal depending on the vicinity of the surroundings itself.