

The Battle of Neighborhoods

Best locations for a new sporting goods store in Chicago

By Maksim Balunov



2020

Introduction

Chicago is one of the most populated cities in the USA. Chicago is one of ten U.S. cities to have teams from the five major American professional team sports (baseball, football, basketball, hockey, soccer). Chicago has been named as the Best Sports City by Sporting News three times in 1993, 2006 and 2010.

The Task At Hand

The company that is specialized on selling sporting goods decided to expand its business to Chicago. They already have multiple shops in big cities, and found out that most profitable of them are in the neighborhood of various sportive objects, especially various gyms. As far as not all of their shops are highly profitable, they need to be more selective and intelligent in their shop positioning. Chicago is also a new market for them and they suffer from the lack of information.

During the discussion of a business problem we've formulated some criteria, that should be met in desired neighborhood. Business highly relies on gym visitors, so the area should be highly dense with gyms. But not only the number of them is significant. During the previous analysis company found out, that the popularity of gyms plays an important role. They found out, that in areas where gyms are more frequent venues (in other words most typical or common), their shops are more profitable. So we need to make qualitative and quantitative analysis. Foursquare API will be very helpful in making data-driven decision.

The goal

So what is the goal of that work? It is obvious. We need to make a data-driven decision, which locations are the best to begin with expansion to Chicago. The results should be easily understandable by stakeholders.

Additionally, this work may be interesting to various other companies, as it explains methods and ways of finding solution.

Data

Various sources of data were used not only to resolve the problem, described above, but also to get acquainted with city of Chicago and some aspects of its life. Some of that sources are not directly used in this work, but they are a part of investigation work.

1. [Chicago Open data portal](#) - The City of Chicago's open data portal lets you find city data, lets you find facts about your neighborhood, lets you create maps and graphs about the city, and lets you freely download the data for your own analysis. Many of these datasets are updated at least once a day, and many of them are updated several times a day.
2. [Wikipedia](#) - Wikipedia is a free encyclopedia, that contains a great amount of data. Several pages were used to get detailed information to solve a task. [List of neighborhoods in Chicago](#) for neighborhood exploration, and [Community areas in Chicago](#) for a community areas understanding.
3. [Foursquare API](#) - An incredible tool to get information about venues, coordinates and many other well detailed information.

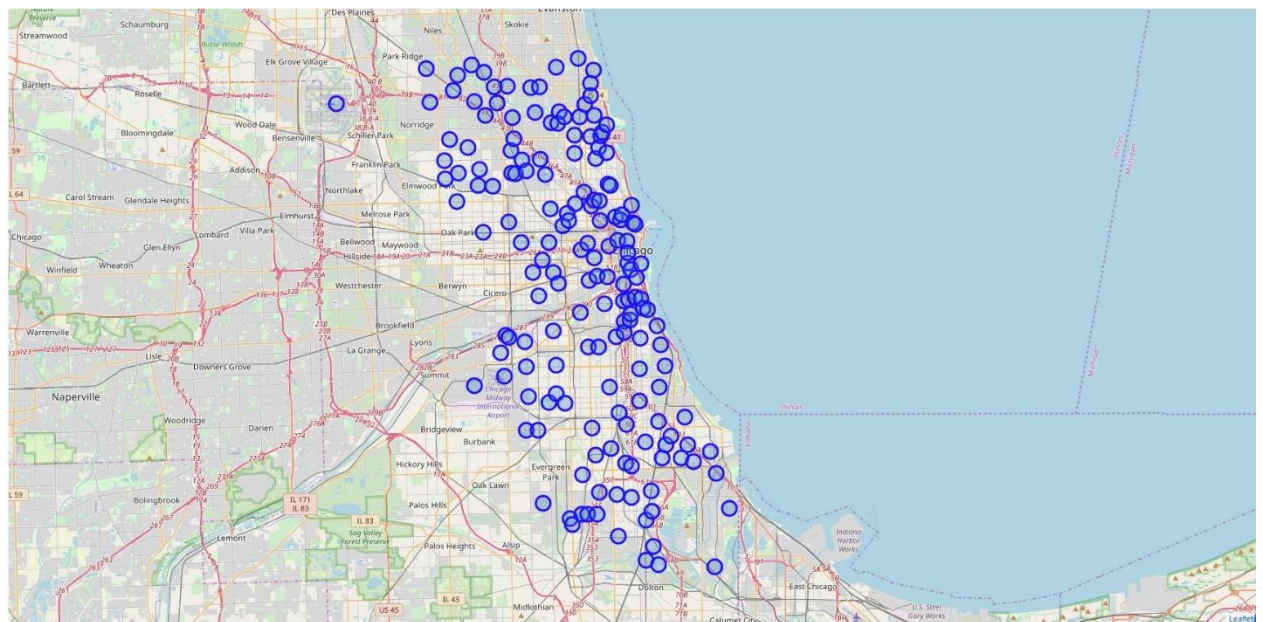
Methodology

The first step in our methodology is to get the initial data, which contains geospatial data on Chicago neighborhoods and community areas. Various sources were used, for example Chicago open data portal and number of Wikipedia pages. The most detailed data was found on [List of neighborhoods in Chicago page](#) from Wikipedia. I used BeautifulSoup package to scrap the data.

The second step is to get neighborhood's coordinates. Most of the coordinates were found lovely, but for some of them I had to change them manually using other resources like Google or Yandex. In the end we have a desired pandas dataframe:

	Neighborhood	Community area	Latitude	Longitude
0	Albany Park	Albany Park	41.971937	-87.716174
1	Altgeld Gardens	Riverdale	41.654864	-87.600446
2	Andersonville	Edgewater	41.977139	-87.669273
3	Archer Heights	Archer Heights	41.811422	-87.726165
4	Armour Square	Armour Square	41.840033	-87.633107
5	Ashburn	Ashburn	41.749352	-87.713514
6	Auburn Gresham	Auburn Gresham	41.750757	-87.662948
7	Avalon Park	Avalon Park	41.745035	-87.588658
8	Avondale	Avondale	41.938921	-87.711168
9	Back of the Yards	New City	41.807533	-87.666163
10	Belmont Central	Belmont Cragin	41.931698	-87.768670

For the better understanding of data distribution, I placed all the neighborhoods on the map:



Visualization methods are good in terms of data understanding. We can see, that neighborhoods from our dataframe are distributed across all the area of Chicago with a slightly greater density closer to the center.

The third step of our analysis involves the usage of a Foursquare API. After providing credentials let's take a look at venues list of the first neighborhood (O'Hare):

	name	categories	lat	lng
0	Garrett Popcorn Shops	Snack Place	41.976047	-87.902024
1	Argo Tea	Tea Room	41.975817	-87.902590
2	Delta Air Lines Ticket Counter	Airport Service	41.976418	-87.904570
3	Garrett Popcorn Shops	Snack Place	41.978830	-87.906890
4	O'Hare Yoga Room	Yoga Studio	41.975487	-87.903245
5	Publican Tavern	Gastropub	41.976573	-87.900908
6	Argo Tea	Tea Room	41.977470	-87.899133
7	Tortas Frontera by Rick Bayless	Mexican Restaurant	41.975421	-87.900147
8	Summer House Santa Monica	American Restaurant	41.976611	-87.905583
9	Tortas Frontera by Rick Bayless	Mexican Restaurant	41.980000	-87.906803

As we can see various categories are presented in that neighborhood. If we are able to get list of venues for a particular neighborhood, we can repeatably get all the venues iterating through all the neighborhoods.

(4309, 7)

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	O'Hare	41.977921	-87.903141	Garrett Popcorn Shops	41.976047	-87.902024	Snack Place
1	O'Hare	41.977921	-87.903141	Argo Tea	41.975817	-87.902590	Tea Room
2	O'Hare	41.977921	-87.903141	Delta Air Lines Ticket Counter	41.976418	-87.904570	Airport Service
3	O'Hare	41.977921	-87.903141	Garrett Popcorn Shops	41.978830	-87.906890	Snack Place
4	O'Hare	41.977921	-87.903141	O'Hare Yoga Room	41.975487	-87.903245	Yoga Studio

On the picture we can see, that there are roughly 4300 venues stored in Foursquare for a particular date among all neighborhoods of Chicago. Five top of them are listed above.

Let's group them by neighborhoods to take a look on distribution:

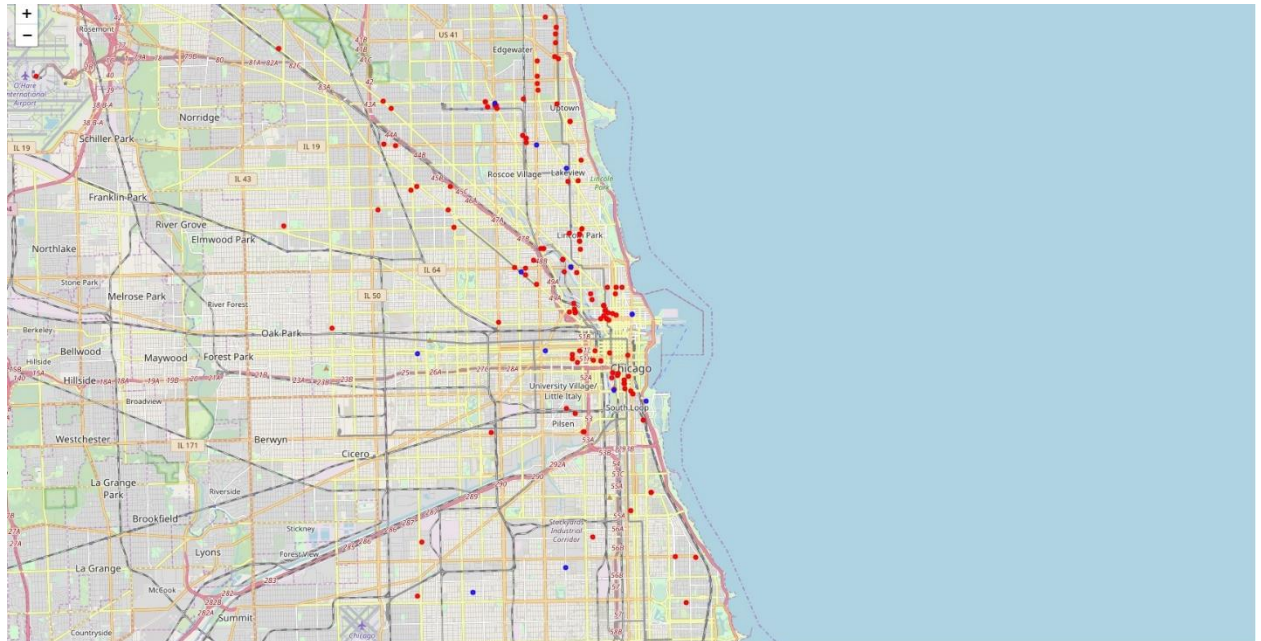
	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
Albany Park	15	15	15	15	15	15
Altgeld Gardens	3	3	3	3	3	3
Andersonville	77	77	77	77	77	77
Archer Heights	18	18	18	18	18	18
Armour Square	11	11	11	11	11	11
Ashburn	6	6	6	6	6	6
Auburn Gresham	7	7	7	7	7	7
Avalon Park	12	12	12	12	12	12
Avondale	27	27	27	27	27	27
Back of the Yards	31	31	31	31	31	31

We can see that venues are distributed very uneven among neighborhoods. Some of them have extremely low number of venues. As far as one of the goals of that work is to define the popularity of gyms in neighborhood, that can seriously impact on that work. So we will not take into account neighborhoods with extremely low number of venues inside.

Some preparations were done. And now we are ready to list all gyms of Chicago:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
69	O'Hare	41.977921	-87.903141	Hilton O'Hare Fitness Center And Spa.	41.977547	-87.904083	Gym
161	Norwood Park East	41.987245	-87.793098	Dog House Crossfit	41.987218	-87.790157	Gym
222	Brickyard	41.929444	-87.788131	XSport Fitness	41.925177	-87.787765	Gym
331	The Island	41.887876	-87.764851	YMCA	41.889375	-87.764858	Gym
498	Kelvyn Park	41.929318	-87.737968	Anytime Fitness	41.930930	-87.743310	Gym / Fitness Center

For better understanding let's put them on the map. Gyms are red, Sporting goods shops are blue.



Foursquare has different categories for different kinds of gyms: Boxing Gym, Climbing Gym, Gym, Gym / Fitness Center. For our work, all of them are just gyms. So we group them together and list top 20 neighborhoods.

	Neighborhood	Gyms	Longitude	Latitude	Community area
0	River North	9	-87.635719	41.896471	Near North Side
1	Printer's Row	7	-87.628900	41.873787	The Loop
2	Dearborn Park	6	-87.628954	41.866553	Near South Side
3	Greektown	5	-87.646932	41.878445	Near West Side
4	Park West	5	-87.645661	41.920892	Lincoln Park
5	East Village	5	-87.655358	41.896198	West Town
6	Lincoln Square	5	-87.688965	41.968682	Lincoln Square
7	Bucktown	4	-87.670123	41.916211	Logan Square
8	Cabrini-Green	4	-87.640937	41.898805	Near North Side
9	Near North Side	4	-87.634497	41.900033	Near North Side
10	Heart of Chicago	3	-87.648108	41.856977	Lower West Side
11	Lincoln Park	3	-87.647832	41.921699	Lincoln Park
12	Andersonville	3	-87.669273	41.977139	Edgewater
13	Goose Island	3	-87.655765	41.910048	Near North Side
14	Nortown	3	-87.660611	41.910530	West Ridge
15	West Loop	3	-87.638578	41.882332	Near West Side
16	North Center	3	-87.679160	41.956107	North Center
17	Wicker Park	3	-87.678155	41.907802	West Town
18	Edgewater	3	-87.663952	41.983369	Edgewater
19	Edgewater Glen	3	-87.664046	41.992391	Edgewater

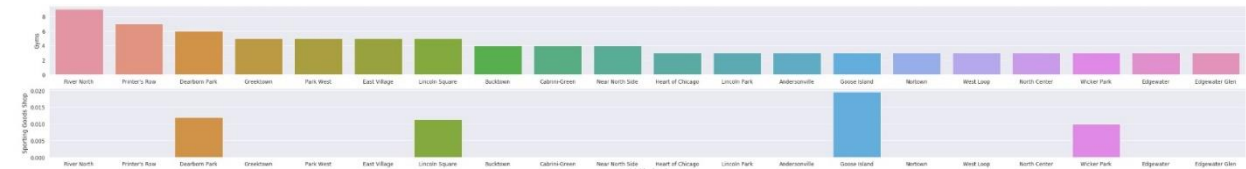
That dataframe contains an answer for one of the works questions: What neighborhoods contain the most number of gyms. The next step is to define top neighborhoods by popularity of gyms. This is made through investigation through all of neighborhoods and finding the relation to all other existing venues. Top 20 of them are listed.

	Neighborhood	Gyms	Longitude	Latitude	Community area
0	Kennedy Park	0.285714	-87.683482	41.687226	Morgan Park
1	Nortown	0.166667	-87.660611	41.910530	West Ridge
2	Bucktown	0.166667	-87.670123	41.916211	Logan Square
3	East Village	0.147059	-87.655358	41.896198	West Town
4	Near North Side	0.133333	-87.634497	41.900033	Near North Side
5	Norwood Park East	0.125000	-87.793098	41.987245	Norwood Park
6	River North	0.113924	-87.635719	41.896471	Near North Side
7	Cabrini-Green	0.102564	-87.640937	41.898805	Near North Side
8	Mayfair	0.100000	-87.737837	41.968087	Albany Park
9	Canaryville	0.100000	-87.640329	41.814756	New City
10	Greektown	0.098039	-87.646932	41.878445	Near West Side
11	Irving Park	0.095238	-87.736447	41.953365	Irving Park
12	Kelvyn Park	0.090909	-87.737968	41.929318	Hermosa
13	Kenwood	0.086957	-87.597991	41.809144	Kenwood
14	The Island	0.083333	-87.764851	41.887876	Austin
15	Groveland Park	0.083333	-87.610256	41.833644	Douglas
16	Edgewater Glen	0.081081	-87.664046	41.992391	Edgewater
17	Dearborn Park	0.072289	-87.628954	41.866553	Near South Side
18	Printer's Row	0.070000	-87.628900	41.873787	The Loop
19	Park West	0.067568	-87.645661	41.920892	Lincoln Park

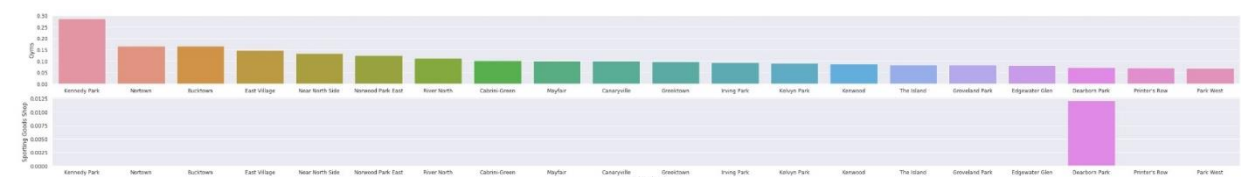
The dataframe of the most common venues for all neighborhoods was also created:

Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0 Albany Park	Sandwich Place	Grocery Store	Mobile Phone Shop	Hookah Bar	Fried Chicken Joint	Bus Station	Bakery	Chinese Restaurant	Korean Restaurant	Cocktail Bar
1 Andersonville	Gyms	Sandwich Place	Coffee Shop	Italian Restaurant	Lounge	Bookstore	Breakfast Spot	Beer Bar	Burger Joint	Salon / Barbershop
2 Archer Heights	Mobile Phone Shop	Mexican Restaurant	Gas Station	Grocery Store	Gyms	Big Box Store	Bank	Bar	Park	Sandwich Place
3 Armour Square	Chinese Restaurant	Sandwich Place	Cosmetics Shop	Gas Station	Italian Restaurant	Sports Bar	Hot Dog Joint	American Restaurant	Asian Restaurant	Exhibit
4 Ashburn	Light Rail Station	Clothing Store	Locksmith	Park	Mexican Restaurant	Fried Chicken Joint	Electronics Store	English Restaurant	Entertainment Service	Ethiopian Restaurant
5 Auburn Gresham	Cosmetics Shop	Pharmacy	Fast Food Restaurant	Lounge	Discount Store	Dim Sum Restaurant	Greek Restaurant	Dance Studio	Currency Exchange	English Restaurant
6 Avalon Park	Boutique	Fast Food Restaurant	Burger Joint	Diner	Cajun / Creole Restaurant	Sandwich Place	Grocery Store	ATM	Pizza Place	Deli / Bodega
7 Avondale	Bus Station	Chinese Restaurant	Grocery Store	Food Truck	Gyms	Diner	Electronics Store	Sandwich Place	Donut Shop	Supermarket
8 Back of the Yards	Mexican Restaurant	Pizza Place	Grocery Store	Bank	Intersection	Shoe Store	Bus Station	Chinese Restaurant	Fried Chicken Joint	Clothing Store
9 Belmont Central	Restaurant	Grocery Store	Mexican Restaurant	Thrift / Vintage Store	Discount Store	Chinese Restaurant	Nightclub	American Restaurant	Bakery	Middle Eastern Restaurant

Top 20 neighborhoods were found by both of the metrics. The next question to answer if there are existing sporting goods shops in selected neighborhoods. That can give us an understanding of other companies’ strategy. First barplot our first metric:



We can see that there is not much intersection among the number of gyms and shops, but still other companies tend to place their shops in such neighborhoods. What is about popularity?



Regarding to this plot most of our competitors don't consider that metric.

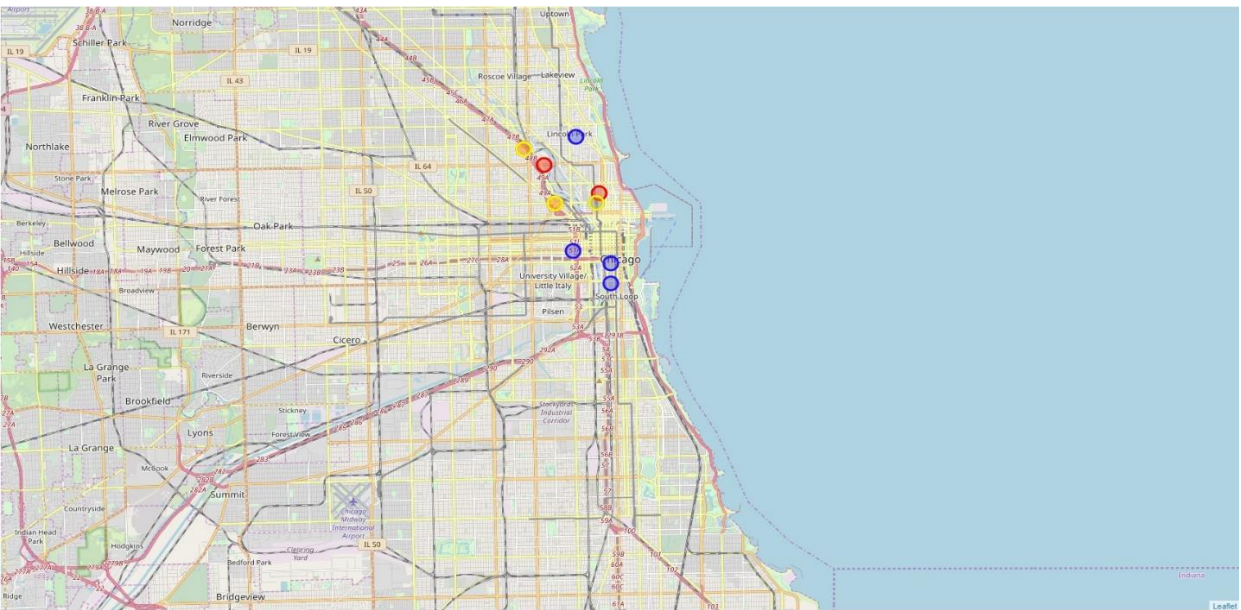
As far as our own company's strategy relies on the both of metrics we will create a combined one. First let's find an intersection if any. Good news, that nearly a half of neighborhoods are in both top datasets. As far as for the first case we count numbers and for the second relationship it will be great to normalize columns of both metrics. After that the final value of combination is derived, and the dataframe is sorted.

	Neighborhood	Gyms_x	Longitude_x	Latitude_x	Community area_x	Sporting Goods Shop_x	Gyms_y	Longitude_y	Latitude_y	Community area_y	Sporting Goods Shop_y	Gyms
0	River North	1.000000	-87.635719	41.896471	Near North Side	0.000000	0.467779	-87.635719	41.896471	Near North Side	0.000000	1.467779
6	Bucktown	0.166667	-87.670123	41.916211	Logan Square	0.000000	1.000000	-87.670123	41.916211	Logan Square	0.000000	1.166667
5	East Village	0.333333	-87.655358	41.896198	West Town	0.000000	0.802139	-87.655358	41.896198	West Town	0.000000	1.135472
9	Northtown	0.000000	-87.660611	41.910530	West Ridge	0.000000	1.000000	-87.660611	41.910530	West Ridge	0.000000	1.000000
8	Near North Side	0.166667	-87.634497	41.900033	Near North Side	0.000000	0.663636	-87.634497	41.900033	Near North Side	0.000000	0.830303
1	Printer's Row	0.666667	-87.628900	41.873787	The Loop	0.000000	0.024545	-87.628900	41.873787	The Loop	0.000000	0.691212
3	Greektown	0.333333	-87.646932	41.878445	Near West Side	0.000000	0.307487	-87.646932	41.878445	Near West Side	0.000000	0.640820
2	Dearborn Park	0.500000	-87.628954	41.866553	Near South Side	0.012048	0.047645	-87.628954	41.866553	Near South Side	0.012048	0.547645
7	Cabrini-Green	0.166667	-87.640937	41.898805	Near North Side	0.000000	0.353147	-87.640937	41.898805	Near North Side	0.000000	0.519814
4	Park West	0.333333	-87.645661	41.920892	Lincoln Park	0.000000	0.000000	-87.645661	41.920892	Lincoln Park	0.000000	0.333333
10	Edgewater Glen	0.000000	-87.664046	41.992391	Edgewater	0.000000	0.136364	-87.664046	41.992391	Edgewater	0.000000	0.136364

The next plot shows the comparison of initial metrics, derived one and existing sporting goods shops:



The final map of the best locations is also made:



Results and discussion

Three metrics were used in this work. The first of them was based on the number of gyms (red circles), and the other on their popularity (blue circles). The third one is a combination of first two (yellow circles). It was found out that both of initial metrics shows that desired neighborhoods are located near the center of the city. The most desirable neighborhoods are:

1. River North
2. Bucktown
3. East Village

Bar plots in addition show, that there are not many sporting goods shops in neighborhoods selected by popularity metrics and combined one. Our company strategy relies on the combined metrics, so its results show the low level of the competition.

Conclusions

In conclusion, we can say that different methodologies, data sources and metrics can be used to resolve that question. This work highly relies on Foursquare API. The industry is changing and the location of gyms and sporting goods shops changes. First, I tended to use cauterization methods to resolve that issue, but instead a custom model was created. As for my point of view it is easier to understand and gives a clear answer. The way I've chosen starts with finding a large amount of locations, and then, step by step, narrowing the number of them. The answer to the initial question is data-driven and rather easy for stakeholders to understand. Moreover, it is reusable and may help in further investigations.

It could be very time-consuming to find a location for the store without used methods. This work shows the easy way to make data-driven recommendations.