

# **Capstone Project - Finding the right location to open a Gym at Toronto**

Kartick Srinivas K C

October 21, 2020

## **1. Introduction**

### **1.1 Background**

Toronto is the capital city of Ontario with a population close to 3 million. It is the most populous city in Canada. It is an international centre for business, finance and arts. We want to open a gym in Toronto. We want to find the right spot to open the gym such that it will give us maximum returns.

### **1.2 Business Problem**

In this project, we will use the power of data science to find the right spot to open a gym at Toronto city. There are a lot of gyms and fitness centres in Toronto. Our aim is to find the optimal location so that we can maximize the return on investment, which is based on the location, its popularity and nearby competitors.

We will also prefer to open the gym within the city limits. We shall analyse the advantage of each area, and the number of gym venues in each area to find the right spot

## **2. Data**

For this project, the following factors will influence our decision

- number of existing gyms and fitness centres in the neighborhood
- popularity and population of the neighborhood
- distance of neighborhood from Toronto downtown

### **2.1 Data Sources**

We will use the following data sources to extract/generate the required information:

- Neighborhood, postal codes and boroughs of Toronto city will be taken from Wikipedia ([https://en.wikipedia.org/wiki/List\\_of\\_postal\\_codes\\_of\\_Canada:\\_M](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M))
- Geographical coordinates of each postal code are taken from this [link](#).
- Number of gyms and fitness centres, their type, location and details in every neighborhood will be obtained using **Foursquare API**

## 2.2 Data Cleaning

We will use BeautifulSoup to extract the data from [Wikipedia](#). Then clean the data by removing the row that has ‘Not assigned’ value for borough. If borough has value but neighborhood has ‘Not assigned’, then neighborhood will have same value as borough. Finally, we create a Dataframe with the data.

	PostalCode	Borough	Neighborhood
0	M1B	Scarborough	Malvern, Rouge
1	M1C	Scarborough	Rouge Hill, Port Union, Highland Creek
2	M1E	Scarborough	Guildwood, Morningside, West Hill
3	M1G	Scarborough	Woburn
4	M1H	Scarborough	Cedarbrae
5	M1J	Scarborough	Scarborough Village
6	M1K	Scarborough	Kennedy Park, Ionview, East Birchmount Park
7	M1L	Scarborough	Golden Mile, Clairlea, Oakridge
8	M1M	Scarborough	Cliffside, Cliffcrest, Scarborough Village West
9	M1N	Scarborough	Birch Cliff, Cliffside West
10	M1P	Scarborough	Dorset Park, Wexford Heights, Scarborough Town...

Then using [this link](#), we extract the geographical coordinates and add as columns to the above mentioned dataframe with columns as Longitude and Latitude.

	PostalCode	Borough	Neighborhood	Latitude	Longitude
0	M1B	Scarborough	Malvern, Rouge	43.806686	-79.194353
1	M1C	Scarborough	Rouge Hill, Port Union, Highland Creek	43.784535	-79.160497
2	M1E	Scarborough	Guildwood, Morningside, West Hill	43.763573	-79.188711
3	M1G	Scarborough	Woburn	43.770992	-79.216917
4	M1H	Scarborough	Cedarbrae	43.773136	-79.239476
5	M1J	Scarborough	Scarborough Village	43.744734	-79.239476
6	M1K	Scarborough	Kennedy Park, Ionview, East Birchmount Park	43.727929	-79.262029
7	M1L	Scarborough	Golden Mile, Clairlea, Oakridge	43.711112	-79.284577
8	M1M	Scarborough	Cliffside, Cliffcrest, Scarborough Village West	43.716316	-79.239476
9	M1N	Scarborough	Birch Cliff, Cliffside West	43.692657	-79.264848
10	M1P	Scarborough	Dorset Park, Wexford Heights, Scarborough Town...	43.757410	-79.273304

We pass the above data to the Foursquare API, and use explore option to get required data of gym venues in neighborhoods of Toronto using the geographical coordinates. After getting the data from Foursquare API, we arrange the data in a data frame with columns Neighborhood, Venue, Venue Longitude, Venue Latitude, Venue Category, Venue Distance, Neighborhood latitude, Neighborhood longitude

	Neighborhood	Venue	Venue Latitude	Venue Longitude	Venue Category	Venue Distance	Neighborhood Latitude	Neighborhood Longitude
0	Cedarbrae	Fitness Distinction	43.774717	-79.239998	Gym / Fitness Center	180	43.773136	-79.239476
1	Cedarbrae	Xplosion Fitness Resolutions	43.775060	-79.239952	Gym	217	43.773136	-79.239476
2	Cedarbrae	Supreme Fitness	43.776590	-79.237579	Gym / Fitness Center	413	43.773136	-79.239476
3	Cedarbrae	Olympian Martial Arts Studio	43.774686	-79.240908	Martial Arts School	207	43.773136	-79.239476
4	Golden Mile, Clairlea, Oakridge	Leveled Fitness	43.714290	-79.281266	Gym	442	43.711112	-79.284577
5	Golden Mile, Clairlea, Oakridge	Warden Hilltop C.C. Weights Room	43.714257	-79.280644	Gym	471	43.711112	-79.284577
6	Dorset Park, Wexford Heights, Scarborough Town...	Tempus Performance	43.759552	-79.277403	Gym	406	43.757410	-79.273304
7	Dorset Park, Wexford Heights, Scarborough Town...	United Martial Arts Canada	43.759020	-79.268078	Martial Arts School	456	43.757410	-79.273304

### 3. Methodology

We shall concentrate on detecting areas of Toronto that have low density of gyms or other fitness centres.

In first step we have collected the required data: Toronto neighborhood, location and type (category) of every gym and fitness centres in Toronto

Second step in our analysis will be calculation and exploration of gym / fitness centres density across different areas of Toronto - we will use heatmaps to identify a few promising areas close to center with low number of gyms / fitness centres in general and focus our attention on those areas.

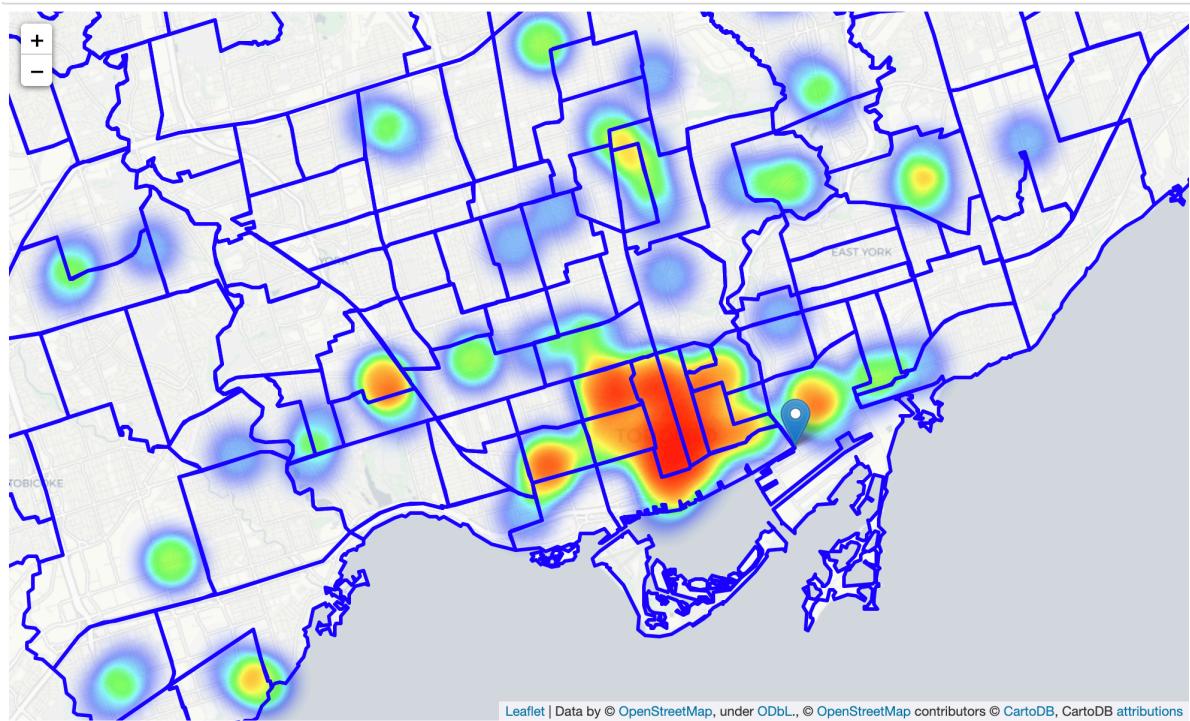
In third and final step we will focus on most promising areas and within those create clusters of locations that have gyms / fitness centres (using k-means clustering)

We will then analyse each cluster, neighborhood, and the number of gyms/fitness centres across each area and identify the right spot.

#### 3.1 Explanatory Data Analysis

#### Displaying the gym locations as a heat map in Toronto :

The following map shows the location of the gym / fitness venues as a heat map in Toronto map. We can infer that most of the venues are located at downtown.



### **Analysing Each Neighborhood with the count of gym venues:**

The following table shows the count of venues in each neighborhood. This table helps somewhat to reject the neighborhood having high number of venues.

Neighborhood	Venue
Agincourt	3
Alderwood, Long Branch	2
Bathurst Manor, Wilson Heights, Downsview North	1
Bayview Village	1
Bedford Park, Lawrence Manor East	4
Berczy Park	7
Brockton, Parkdale Village, Exhibition Place	6
Business reply mail Processing Centre, South Central Letter Processing Plant Toronto	4
Canada Post Gateway Processing Centre	2
Cedarbrae	4
Central Bay Street	18
Christie	4
Church and Wellesley	21
Clarks Corners, Tam O'Shanter, Sullivan	2
Commerce Court, Victoria Hotel	22
Davisville	3
Davisville North	5

## Creating a dataframe that shows the type and number of Gyms/fitness centres venues in each neighborhood:

The following table shows the type of gym/fitness centre venues in each neighborhood.

	Neighborhood	Boxing Gym	Climbing Gym	College Gym	College Rec Center	Cycle Studio	Gym	Gym / Fitness Center	Gym Pool	Gymnastics Gym	Hotel Pool	Martial Arts School	Pilates Studio	Yoga Studio
0	Agincourt	0	0	0	0	0	1	0	0	0	0	2	0	0
1	Alderwood, Long Branch	0	0	0	0	0	1	1	0	0	0	0	0	0
2	Bathurst Manor, Wilson Heights, Downsview North	0	0	0	0	0	0	0	0	0	0	0	0	1
3	Bayview Village	0	0	0	0	0	1	0	0	0	0	0	0	0
4	Bedford Park, Lawrence Manor East	0	0	0	0	0	1	1	0	1	0	0	0	1
5	Berczy Park	0	0	0	0	0	3	2	1	0	0	0	0	1
6	Brockton, Parkdale Village, Exhibition Place	0	1	0	0	0	2	1	0	0	0	0	0	2
7	Business reply mail Processing Centre, South C...	0	0	0	0	0	0	2	0	0	0	1	0	1

## Clustering the locations:

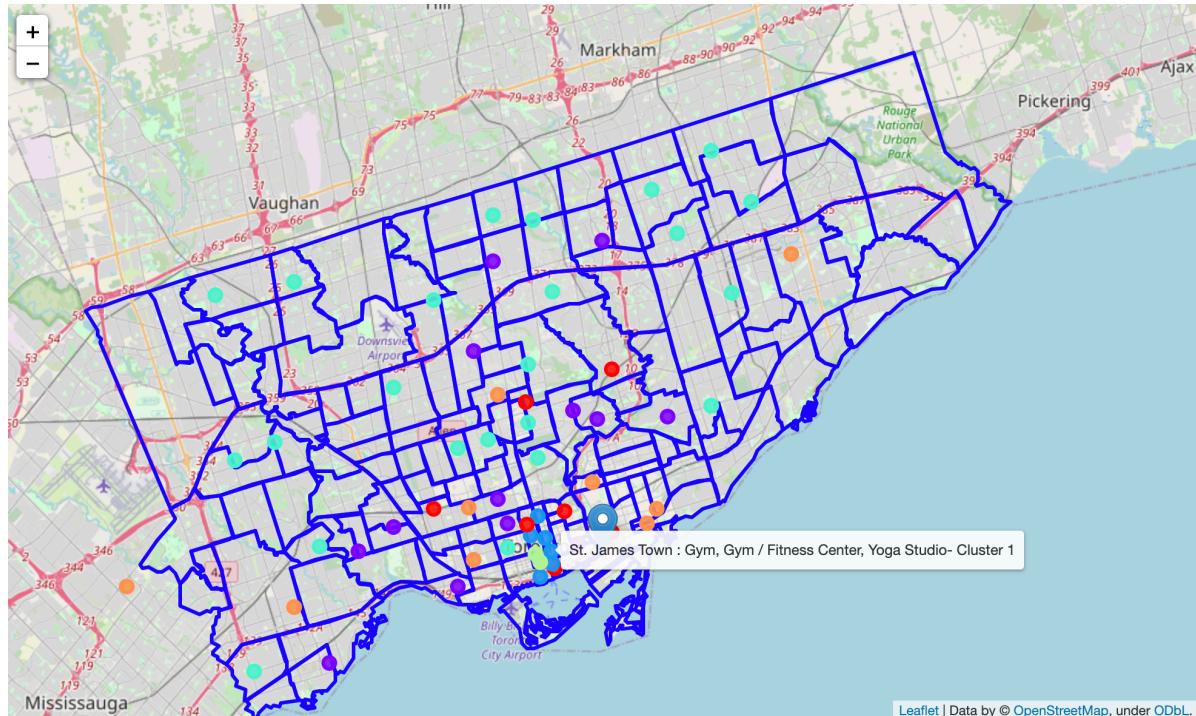
We clustered the data with k=6 using k-means clustering technique. The following table shows the data with Cluster label.

Cluster Labels	Neighborhood	Boxing Gym	Climbing Gym	College Gym	College Rec Center	Cycle Studio	Gym	Gym / Fitness Center	Gym Pool	Gymnastics Gym	Hotel Pool	Martial Arts School	Pilates Studio	Yoga Studio
0 2	Agincourt	0	0	0	0	0	1	0	0	0	0	2	0	0
1 2	Alderwood, Long Branch	0	0	0	0	0	1	1	0	0	0	0	0	0
2 2	Bathurst Manor, Wilson Heights, Downsview North	0	0	0	0	0	0	0	0	0	0	0	0	1
3 2	Bayview Village	0	0	0	0	0	1	0	0	0	0	0	0	0
4 0	Bedford Park, Lawrence Manor East	0	0	0	0	0	1	1	0	1	0	0	0	1
5 5	Berczy Park	0	0	0	0	0	3	2	1	0	0	0	0	1
6 0	Brockton, Parkdale Village, Exhibition Place	0	1	0	0	0	2	1	0	0	0	0	0	2
7 4	Business reply mail Processing Centre, South C...	0	0	0	0	0	0	2	0	0	0	1	0	1

The following table shows creating a dataframe with the venues and count included along with cluster labels.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Gym Venues Count	Gym Venues	Cluster Labels
0	Agincourt	43.794200	-79.262029	3	Gym, Martial Arts School	2
1	Alderwood, Long Branch	43.602414	-79.543484	2	Gym, Gym / Fitness Center	2
2	Bathurst Manor, Wilson Heights, Downsview North	43.754328	-79.442259	1	Yoga Studio	2
3	Bayview Village	43.786947	-79.385975	1	Gym	2
4	Bedford Park, Lawrence Manor East	43.733283	-79.419750	4	Gym, Gym / Fitness Center, Gymnastics Gym, Yog...	0
5	Berczy Park	43.644771	-79.373306	7	Gym, Gym / Fitness Center, Gym Pool, Yoga Studio	5
6	Brockton, Parkdale Village, Exhibition Place	43.636847	-79.428191	6	Climbing Gym, Gym, Gym / Fitness Center, Yoga ...	0
7	Business reply mail Processing Centre, South C...	43.662744	-79.321558	4	Gym / Fitness Center, Martial Arts School, Yog...	4

We create a map to see gym situated in different neighborhoods. Each cluster is differentiated with different color. We can also hover over each circle to see the neighborhood, gym venues and the Cluster label



## 4. Results

We used foursquare to analyse Toronto thoroughly about all the gym locations. We found that most of the gym venues are crowded inside the Toronto Downtown, and there are many neighborhoods in Toronto that do not have a gym.

Even though there are many opportunities to open a gym outside the downtown of Toronto, opening a gym in downtown will have maximum Return on Investment(ROI). There are neighborhoods such as CN Tower, King and Spadina, Railway Lands, Harbourfront West, Bathurst Quay, South Niagara and Island airport inside the downtown that have no gym or any of the fitness centres.

## 5. Discussion

There are pockets of areas in Toronto downtown that don't have gym/fitness centre. These areas are CN Tower, King and Spadina, Railway Lands, Harbourfront West, Bathurst Quay, South Niagara and Island airport. Opening a gym in any of these areas will give maximum return on investment since these neighborhoods are busy all the time with many offices and houses.

My recommendation is to open a gym/fitness center in one of the areas mentioned above.

## 6. Conclusion

The purpose of this project is to find the right location for opening a Gym at Toronto City. After thorough analysis, we found that opening a gym at postal code 'M5V' with neighborhoods such as CN Tower, King and Spadina, Railway Lands, Harbourfront West, Bathurst Quay, South Niagara and Island airport. Even though these neighborhoods are in downtown Toronto, there are no gyms in these areas. Opening a gym in one of these area will give maximum return on investment.