

Capstone Project - The Battle of the Neighbourhoods

**Applied Data Science Capstone by
IBM/Coursera**

Bharath Reddy

Introduction

Today, we have more intellects on this planet than ever before. As the intellect becomes stronger, people look for logical solutions. More and more people are realizing physical activity is important for Fitness & Mental well-being of an individual.

There are different types of Fitness & mental well-being center's, such as GYM, Yoga, Aerobics, Boxing, Pool, recreation centers, Gymnastics and Martial Arts School etc.

Each type has its own benefits and its own style.

Toronto is the capital city of the Canadian province of Ontario. With a recorded population of 2,731,571 in 2016, it is the most populous city in Canada and the fourth most populous city in North America. Toronto is an international centre of business, finance, arts, and culture, and is recognized as one of the most multicultural and cosmopolitan cities in the world.

Starting a **Fitness & Mental well-being centre** in **Toronto, Ontario, Canada** in a right Neighbourhood can be highly rewarding.

In this project we will try to find an optimal location for starting a **Fitness & mental well-being center**.

This report will be targeted to stakeholders interested in opening a **Fitness & Mental well-being center** in **Toronto, Ontario, Canada**

We will use our data science powers to identify neighbourhoods and show which type of Fitness & Mental well-being center can be launched.

Advantages of each neighbourhoods will then be clearly expressed so that best possible final location can be chosen by stakeholders.

Data

Based on definition of our problem, factors that will influence our decision are:

- Different types of fitness centers in a neighbourhood.
- Density of fitness centers in a neighbourhood.
- Combination of fitness centers in a neighbourhood.

Following data sources will be needed to extract/generate the required information:

- List of neighbourhoods of Toronto will be gathered from wikipedia https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:M
- List of fitness centers and their location in every neighbourhood will be obtained using **Foursquare API**
- List of Toronto neighbourhood coordinates from http://cocl.us/Geospatial_data

Methodology

Brief steps of the methodology we are going to follow in this exercise.

- In the First step we will be collecting the list of neighbourhoods in Toronto and clean the data.
- As a second step for each neighbourhood we shall collect list of Fitness or wellness centers in a radius of 500 meters and clean the data.
- For Third step we will identify market share of different types of Fitness or wellness centers and as well how Fitness or Wellness centers spread across the neighbourhoods.
- In Fourth step we will segregate the neighbourhoods into clusters using K means based on the most popular venues in each neighbourhood.
- As a part of fifth and final step, we will merge clusters with count of venues in each neighbourhood, to search for optimal venue location by stakeholders.

Collected Toronto neighbourhoods by scrapping Wikipedia page
https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M.

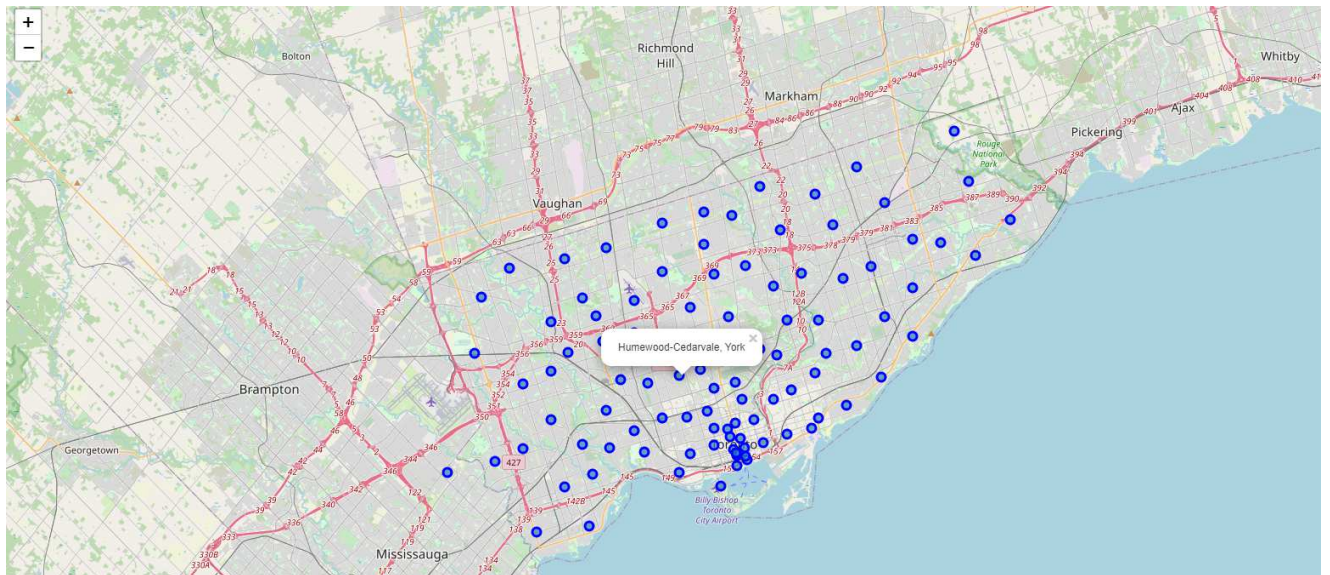
Then using http://cocl.us/Geospatial_data we got the longitude and latitude of all the neighbourhoods in Toronto region.

Toronto neighbourhoods with Latitude, Longitude. (*Sample data)

:

	Borough	Neighbourhood	Latitude	Longitude
Postal Code				
M3A	North York	Parkwoods	43.753259	-79.329656
M4A	North York	Victoria Village	43.725882	-79.315572
M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636
M6A	North York	Lawrence Manor, Lawrence Heights	43.718518	-79.464763
M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government	43.662301	-79.389494

A visualization with all the neighbourhoods of the Toronto are shown below.



Using the Foursquare API, we acquire Fitness or Wellness around each neighbourhood within radius of 500 meters. To limit venues to only Fitness and wellness category we have used category field in the API as below.

https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&v={}&ll={,}&categoryId={}&radius={}&limit={}

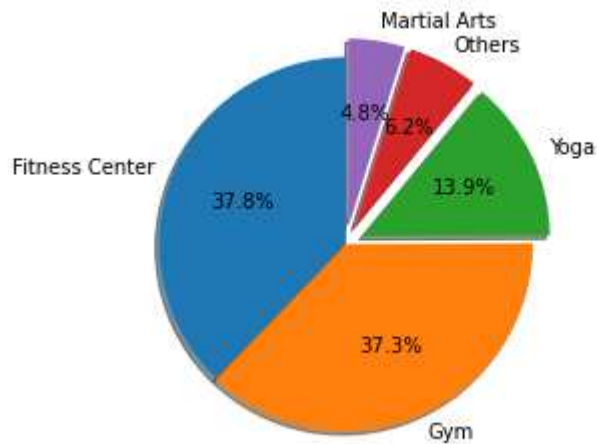
Category ID for the Fitness and Wellness centers can be found from Foursquare website <https://developer.foursquare.com/docs/build-with-foursquare/categories>

Category ID we used is “4bf58dd8d48988d175941735”.

There were 13 unique categories of Fitness and Wellness centers, but few of categories are just in few very few numbers. We had cleaned the data to have 5 different categories.

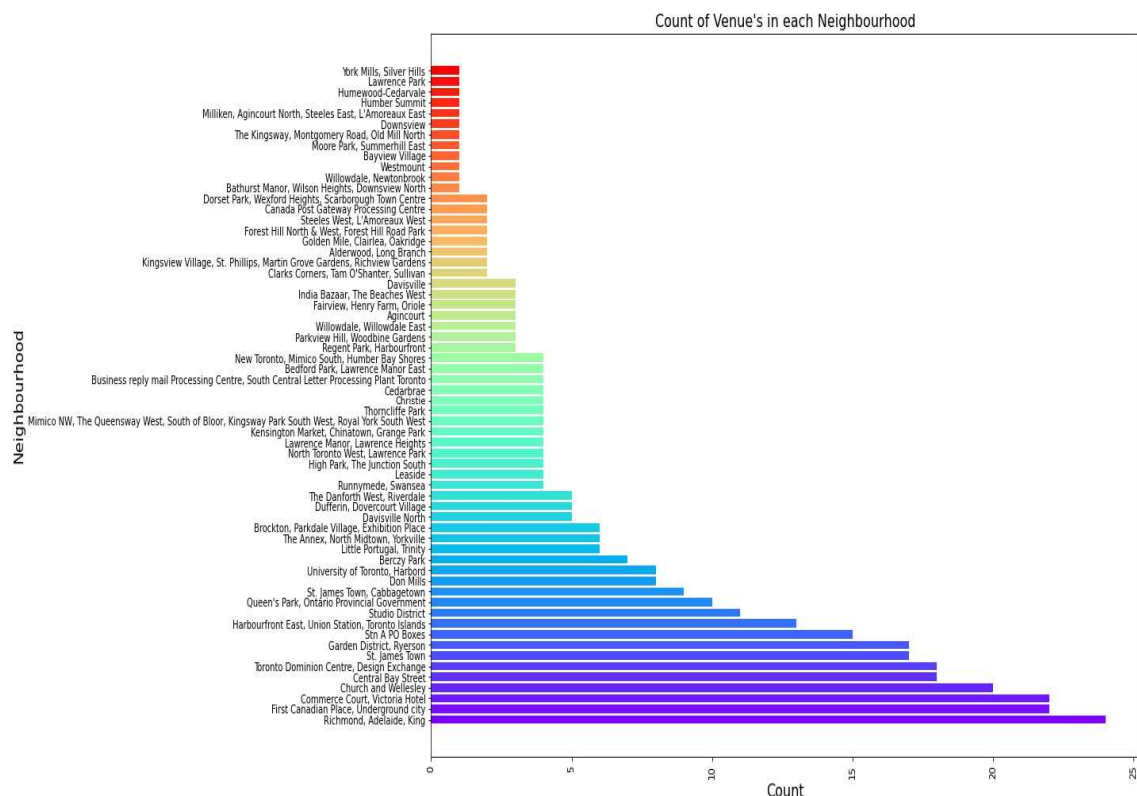
- 1) Fitness Center (Aerobics etc)
- 2) GYM
- 3) Yoga
- 4) Martial Arts
- 5) Others (boxing, climbing, cycling, pool etc).

Based on the analysis of Venues returned from the Foursquare portal, below is the current Market Share of different Fitness/Wellness Centers.



75% Market share is captured by **GYM & Fitness centers (aerobics etc)** and **Yoga**, which is currently highly popular has just **14% market share**. Martial Arts just hold 5% market share, and remaining types such as Boxing, Pools, Gymnastics, climbing studio's hold just 5% Market share

Also based on the analysis of Venues returned from the Foursquare portal, Below picture shows Neighbourhood vs number of venues.

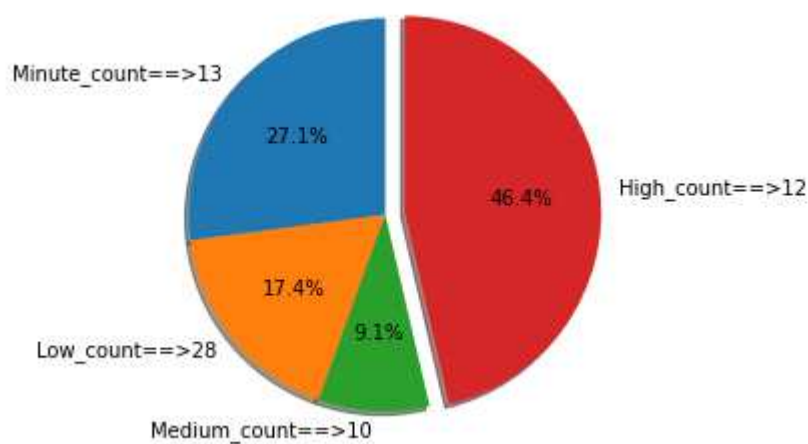


We have noticed that major number of Wellness and Fitness centers are present in only few neighbourhoods.

We had segregated neighbourhoods as Minute, Low, Medium and High Categories to observe the Market Share w.r.t Venues.

- 1) Minute holds for less than 2 Venues.
- 2) Low holds for less than 5 Venues.
- 3) Medium holds for less than 10 Venues.
- 4) High holds for more than 10 Venues.

Below is the pie chart for the same.



We can see that **50%** of the Total fitness/wellness centers are present in **12** Neighbourhoods.

There are **13** neighbourhoods which has just **1** fitness/wellness centers which holds for **28%** for neighbourhoods.

We created a one-hot encoding to analyse each neighbourhood. This results in a Data frame that displays the most common venue category in a neighbourhood. The result is as shown below.

(62, 6)

	Neighbourhood	Fitness Center	Gym	Martial Arts	Others	Yoga
0	Agincourt	0.000000	0.333333	0.666667	0.000000	0.000000
1	Alderwood, Long Branch	0.500000	0.500000	0.000000	0.000000	0.000000
2	Bathurst Manor, Wilson Heights, Downsview North	0.000000	0.000000	0.000000	0.000000	1.000000
3	Bayview Village	0.000000	1.000000	0.000000	0.000000	0.000000
4	Bedford Park, Lawrence Manor East	0.250000	0.250000	0.000000	0.250000	0.250000
5	Berczy Park	0.285714	0.428571	0.000000	0.142857	0.142857
6	Brockton, Parkdale Village, Exhibition Place	0.166667	0.333333	0.000000	0.166667	0.333333
7	Business reply mail Processing Centre, South Central Letter Processing Plant Toronto	0.500000	0.000000	0.250000	0.000000	0.250000
8	Canada Post Gateway Processing Centre	1.000000	0.000000	0.000000	0.000000	0.000000
9	Cedarbrae	0.500000	0.250000	0.250000	0.000000	0.000000

So, we have 62 neighbourhoods to investigate.

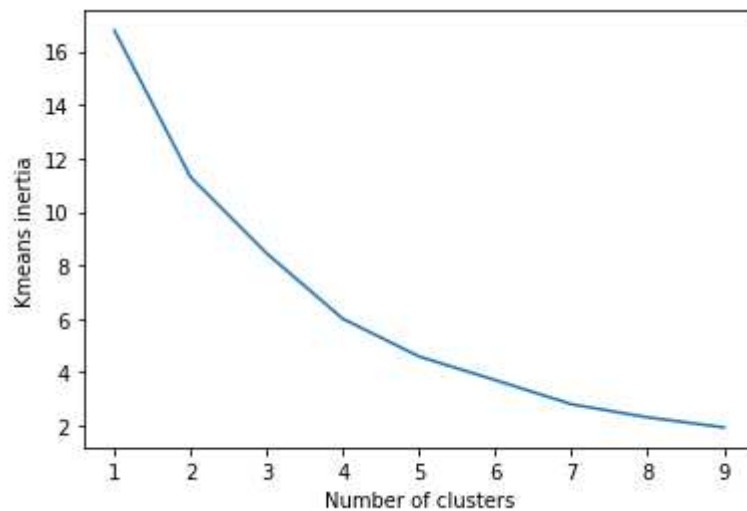
We had further optimized the neighbourhood with top 3 Venue's as below.

	Neighbourhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
0	Agincourt	Martial Arts	Gym	Fitness Center
1	Alderwood, Long Branch	Fitness Center	Gym	Martial Arts
2	Bathurst Manor, Wilson Heights, Downsview North	Yoga	Fitness Center	Gym
3	Bayview Village	Gym	Fitness Center	Martial Arts
4	Bedford Park, Lawrence Manor East	Fitness Center	Gym	Others
5	Berczy Park	Gym	Fitness Center	Others
6	Brockton, Parkdale Village, Exhibition Place	Gym	Yoga	Fitness Center
7	Business reply mail Processing Centre, South Central Letter Processing Plant Toronto	Fitness Center	Martial Arts	Yoga
8	Canada Post Gateway Processing Centre	Fitness Center	Gym	Martial Arts
9	Cedarbrae	Fitness Center	Gym	Martial Arts

We then use the K-means clustering algorithm to group the Venue's into clusters, aim was to partition 'n' observations into k clusters in which each observation belongs to the cluster.

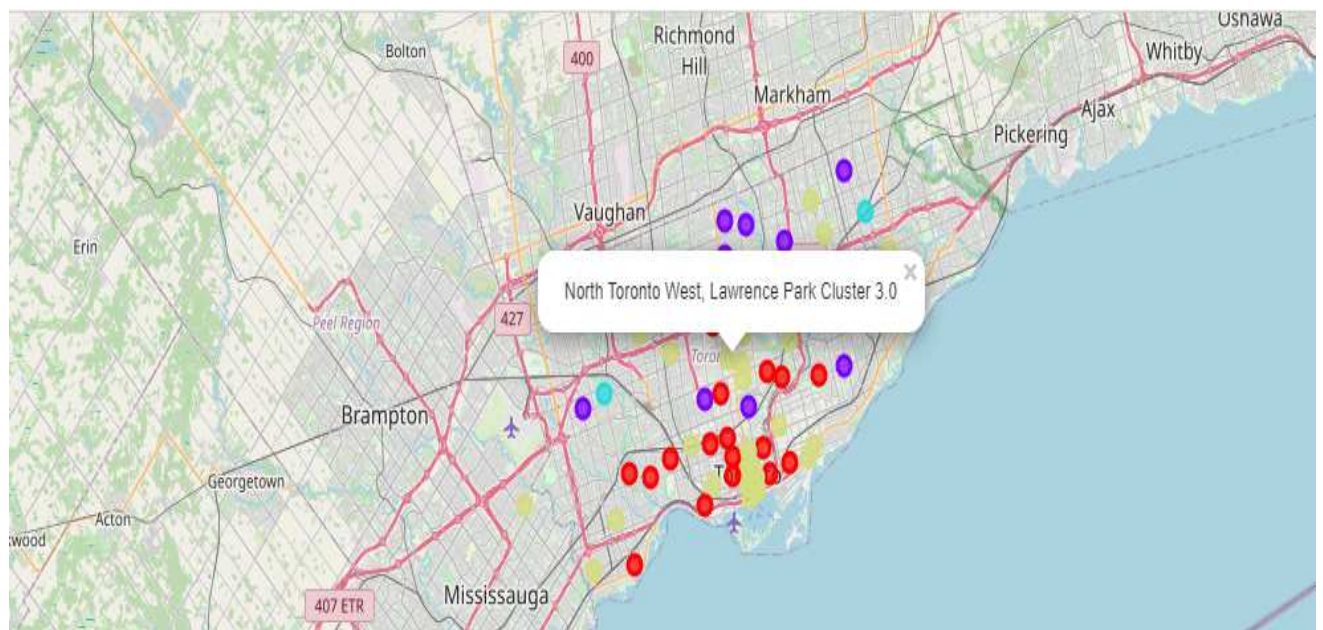
We used Elbow method to identify the K value.

Below is the Graph.



With Elbow method we have concluded that optimal value for number of clusters can be either 3 or 4, we have chosen to stick with 4 clusters after verifying the generated clusters.

Clusters are plotted on the Toronto map as below.



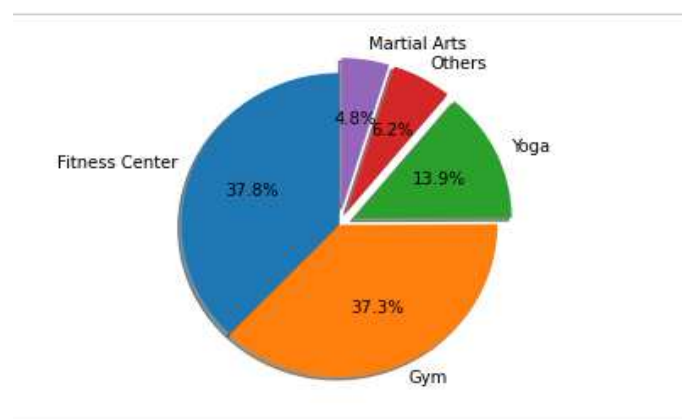
The colours Red, Purple, Blue and Yellow represents Cluster 0, 1, 2 and 3 Respectively.

Results and Discussion

Our analysis shows that although there is a good number of Fitness/Wellness centers in Toronto, but we have first identified how different types of Fitness and wellness centers trends in the neighbourhoods. **75%** Market share is captured by **GYM & Fitness centers (aerobics etc)** and **Yoga**, which is currently highly popular has just **14% market share**.

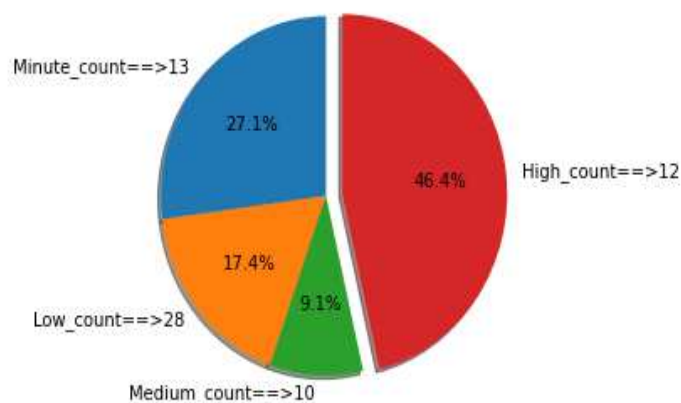
Martial Arts just hold 5% market share, and remaining types such as Boxing, Pools, Gymnastics, climbing studio's hold just 5% Market share.

There is still potential of the market share capturing for Yoga, Martial arts, and Climbing studios etc.



On analysis of how fitness/wellness centers are spread across the neighbourhood, we can see that **50%** of the Total fitness/wellness centers are present in **12** Neighbourhoods. There are **13** neighbourhoods which has just **1** fitness/wellness centers which holds for **28%** for neighbourhoods.

These less density neighbourhood's potential for new fitness/wellness centers.



Cluster 0: Cluster 0 contains neighbourhoods which has GYM, Fitness centers and Yoga as preference.

Postal Code	Neighbourhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	Total No of Venues
M4M	Studio District	Gym	Yoga	Others	11
M4X	St. James Town, Cabbagetown	Gym	Fitness Center	Martial Arts	9
M5S	University of Toronto, Harbord	Gym	Yoga	Others	8
M6K	Brockton, Parkdale Village, Exhibition Place	Gym	Yoga	Fitness Center	6
M5R	The Annex, North Midtown, Yorkville	Gym	Yoga	Fitness Center	6
M8V	New Toronto, Mimico South, Humber Bay Shores	Gym	Fitness Center	Yoga	4
M4G	Leaside	Gym	Fitness Center	Yoga	4
M6G	Christie	Fitness Center	Yoga	Gym	4
M4H	Thornccliffe Park	Yoga	Fitness Center	Gym	4
M5M	Bedford Park, Lawrence Manor East	Fitness Center	Gym	Others	4
M6P	Hinx Park, The Junction South	Yoga	Gym	Martial Arts	4

Cluster 1: Cluster 1 contains neighbourhoods which has GYM as Major preference.

Postal Code	Neighbourhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	Total No of Venues
M2J	Fairview, Henry Farm, Oriole	Gym	Yoga	Fitness Center	3
M2N	Willowdale, Willowdale East	Gym	Fitness Center	Martial Arts	3
M1L	Golden Mile, Clairlea, Oakridge	Gym	Fitness Center	Martial Arts	2
M9R	Kingsview Village, St. Phillips, Martin Grove Gardens, Richview Gardens	Gym	Fitness Center	Martial Arts	2
M6C	Humewood-Cedarvale	Gym	Fitness Center	Martial Arts	1
M2K	Bayview Village	Gym	Fitness Center	Martial Arts	1
M9L	Humber Summit	Gym	Fitness Center	Martial Arts	1
M2M	Willowdale, Newtonbrook	Gym	Fitness Center	Martial Arts	1
M4T	Moore Park, Summerhill East	Gym	Fitness Center	Martial Arts	1
M1V	Milliken, Agincourt North, Steeles East, L'Amoreaux East	Gym	Fitness Center	Martial Arts	1

Cluster 2: Cluster 2 contains neighbourhoods which has Martial Art schools as preference.

Postal Code	Neighbourhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	Total No of Venues
M1S	Agincourt	Martial Arts	Gym	Fitness Center	3
M1P	Dorset Park, Wexford Heights, Scarborough Town Centre	Gym	Martial Arts	Fitness Center	2
M2L	York Mills, Silver Hills	Martial Arts	Fitness Center	Gym	1
M9P	Westmount	Martial Arts	Fitness Center	Gym	1

Cluster 3: Cluster 4 has GYM, Fitness centers and Yoga as favourites And Cluster 4 also holds high density fitness centers neighbourhoods.

Postal Code	Neighbourhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	Total No of Venues
M5H	Richmond, Adelaide, King	Gym	Fitness Center	Others	24
M5X	First Canadian Place, Underground city	Fitness Center	Gym	Martial Arts	22
M5L	Commerce Court, Victoria Hotel	Fitness Center	Gym	Yoga	22
M4Y	Church and Wellesley	Gym	Fitness Center	Yoga	20
M5K	Toronto Dominion Centre, Design Exchange	Fitness Center	Gym	Martial Arts	18
M5G	Central Bay Street	Fitness Center	Gym	Yoga	18
M5B	Garden District, Ryerson	Fitness Center	Gym	Others	17
M5C	St. James Town	Fitness Center	Gym	Yoga	17
M5W	Stn A PO Boxes	Fitness Center	Gym	Yoga	15
M5J	Harbourfront East, Union Station, Toronto Islands	Fitness Center	Gym	Others	13
M7A	Queen's Park, Ontario Provincial Government	Fitness Center	Gym	Yoga	10

Purpose of this analysis was to show how the distribution of types of Fitness/Wellness centers across neighbourhoods and as well as share of Fitness/Wellness centers across the neighbourhoods.

Clustered data clubbed with Market share analysis shall be used as starting points for final exploration by stakeholders.

Conclusion

Purpose of this project was to identify optimal neighbourhood and type of fitness or wellness center in Toronto neighbourhoods. From Foursquare data we have first identified how different types of Fitness and wellness centers trends in the neighbourhoods. 75% Market share is captured by GYM & Fitness centers (aerobics etc) and Yoga, which is currently highly popular has just 14% market share. Martial Arts just hold 5% market share, and remaining types such as Boxing, Pools, Gymnastics, climbing studio's hold just 5% Market share.

On analysis of how fitness/wellness centers are spread across the neighbourhood, we can see that 50% of the Total fitness/wellness centers are present in 12 Neighbourhoods. There are 13 neighbourhoods which has just 1 fitness/wellness centers which holds for 28% for neighbourhoods.

Neighbourhoods clustering was then performed, with K-means clustering, by exploring the frequency of the type of Fitness/Wellness center's available in each neighbourhood. Further Identified Clusters are clubbed with density of Fitness/Wellness center's in each neighbourhood, which to be used as starting points for final exploration by stakeholders.

Final decision on optimal location and type of fitness or wellness center will be made by stakeholders based on specific characteristics of neighbourhoods and locations in every Cluster, Such as attractiveness of each location, Population in the neighbourhood, Connectivity, real estate availability, prices, social and economic dynamics of every neighbourhood etc.