Battle of the Neighborhoods

Table of Contents

1. Introduction / Business Problem:

From time to time, we see headlines like those above appearing on most media outlets. It's both depressing and sad to accept the fact that Canadians are actually considered some of the least active people in the world. Technological advancement which creates so much convenience nowadays is also one of the main culprits. More time spending on mobile gadgets, watching Netflix, sitting all

day in front of computers, etc., are preventing people from spending more time outdoors and be active. So the lack of fitness facilities or public areas to exercise becomes a good excuse for these people. In order to tackle this issue, I plan to explore which neighborhoods in Toronto which are lack of fitness facilities or public areas to exercise.

2. Target Audience

This project is aimed towards potential fitness center operators, yoga/pilates instructors who wanted to launch their business in areas that have less competition. Toronto City Council can also plan to create more green space or pocket parks in these neighborhoods.

3. Data Overview

These are the data that I used for this project:

1. Toronto's list of neighborhood from wikipedia
https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
2. geographical coordinates csv file from http://cocl.us/Geospatial_data
3. list of venues from Foursquare

Top 10 venues data obtained from Foursquare will be use to analyze each neighborhoods which are catergorized into clusters based on similar venue types. From these data, I will remove neighborhoods with Gym, Yoga Studio, Dance Studio, Park, Pool , Soccer Field, Golf Course in their top 10 venues. The neighborhoods left will be the ones most suitable to open a fitness center.

4. Methodology

4.1 — Data Cleansing and Preparation
- remove rows with boroughs that are “Not assigned”
- merging the geographical coordinates dataframe with the original dataframe based on Postal Code

	Postal Code	Borough	Neighbourhood	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
...	...	...	...	...	...
98	M9N	York	Weston	43.706876	-79.518188
99	M9P	Etobicoke	Westmount	43.696319	-79.532242
100	M9R	Etobicoke	Kingsview Village, St. Phillips, Martin Grove ...	43.688905	-79.554724
101	M9V	Etobicoke	South Steeles, Silverstone, Humbergate, Jamest...	43.739416	-79.588437
102	M9W	Etobicoke	Northwest, West Humber - Clairville	43.706748	-79.594054

103 rows × 5 columns

- venues data are pulled from the Foursquare API, providing all the local venues within a 500-meter radius in all neighborhoods. The data are then grouped together by the same venue category, as shown below

	Neighbourhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Malvern, Rouge	43.806686	-79.194353	Wendy's	43.807448	-79.199056	Fast Food Restaurant
1	Rouge Hill, Port Union, Highland Creek	43.784535	-79.160497	Great Shine Window Cleaning	43.783145	-79.157431	Home Service
2	Rouge Hill, Port Union, Highland Creek	43.784535	-79.160497	Royal Canadian Legion	43.782533	-79.163085	Bar
3	Guildwood, Morningside, West Hill	43.763573	-79.188711	RBC Royal Bank	43.766790	-79.191151	Bank
4	Guildwood, Morningside, West Hill	43.763573	-79.188711	G & G Electronics	43.765309	-79.191537	Electronics Store

```
toronto_venues.groupby('Neighbourhood').count()
```

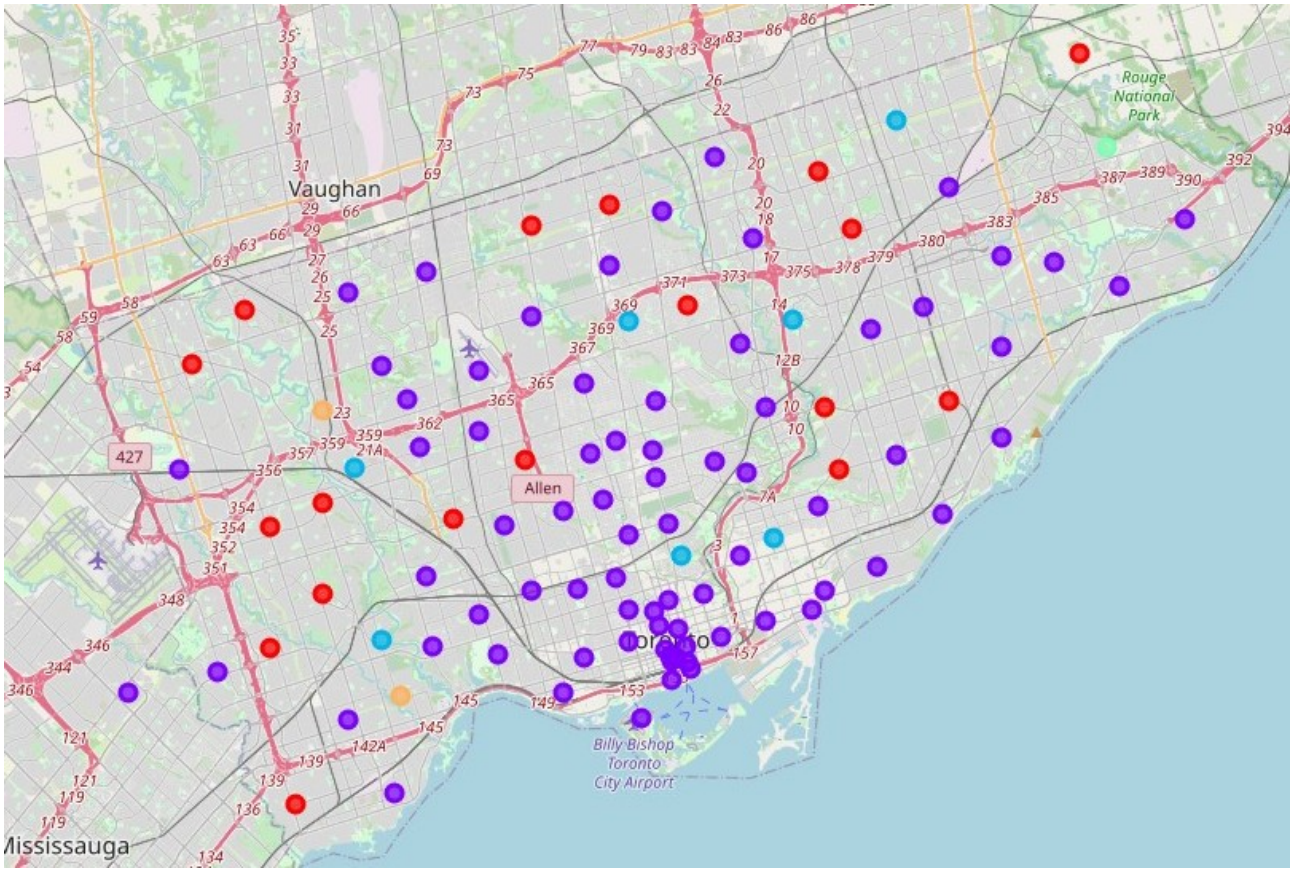
Neighbourhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Agincourt	5	5	5	5	5	5
Alderwood, Long Branch	8	8	8	8	8	8
Bathurst Manor, Wilson Heights, Downsview North	20	20	20	20	20	20
Bayview Village	4	4	4	4	4	4
Bedford Park, Lawrence Manor East	25	25	25	25	25	25
...	...	...	...	...	...	...
Willowdale, Willowdale East	30	30	30	30	30	30
Willowdale, Willowdale West	6	6	6	6	6	6
Woburn	4	4	4	4	4	4
Woodbine Heights	7	7	7	7	7	7
York Mills West	4	4	4	4	4	4

4.2 — Data Exploration/Machine Learning

After cleansing the data, I proceed to analyze it. First, I use one hot encoding to group rows by neighborhood and by taking the mean of the frequency of occurrence of each category. I create the new dataframe and display the top 10 venues for each neighborhood.

	Neighbourhood	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	Agincourt	Latin American Restaurant	Skating Rink	Clothing Store	Lounge	Breakfast Spot	Yoga Studio	Dance Studio	Dog Run	Distribution Center	Discount Store
1	Alderwood, Long Branch	Pizza Place	Coffee Shop	Sandwich Place	Skating Rink	Gym	Pharmacy	Pub	Curling Ice	Dance Studio	Deli / Bodega
2	Bathurst Manor, Wilson Heights, Downsview North	Coffee Shop	Bank	Park	Middle Eastern Restaurant	Diner	Sandwich Place	Deli / Bodega	Bridal Shop	Restaurant	Ice Cream Shop
3	Bayview Village	Chinese Restaurant	Bank	Japanese Restaurant	Café	Yoga Studio	Dance Studio	Drugstore	Donut Shop	Dog Run	Distribution Center
4	Bedford Park, Lawrence Manor East	Coffee Shop	Italian Restaurant	Sandwich Place	Restaurant	Greek Restaurant	Pub	Indian Restaurant	Liquor Store	Locksmith	Butcher

Then, I use K-Means clustering method to group neighborhoods of similar venue categories together, into 5 clusters. And proceed to add the cluster labels into the Dataframe. We can view the map of Toronto which 5 different clusters. Using Folium.

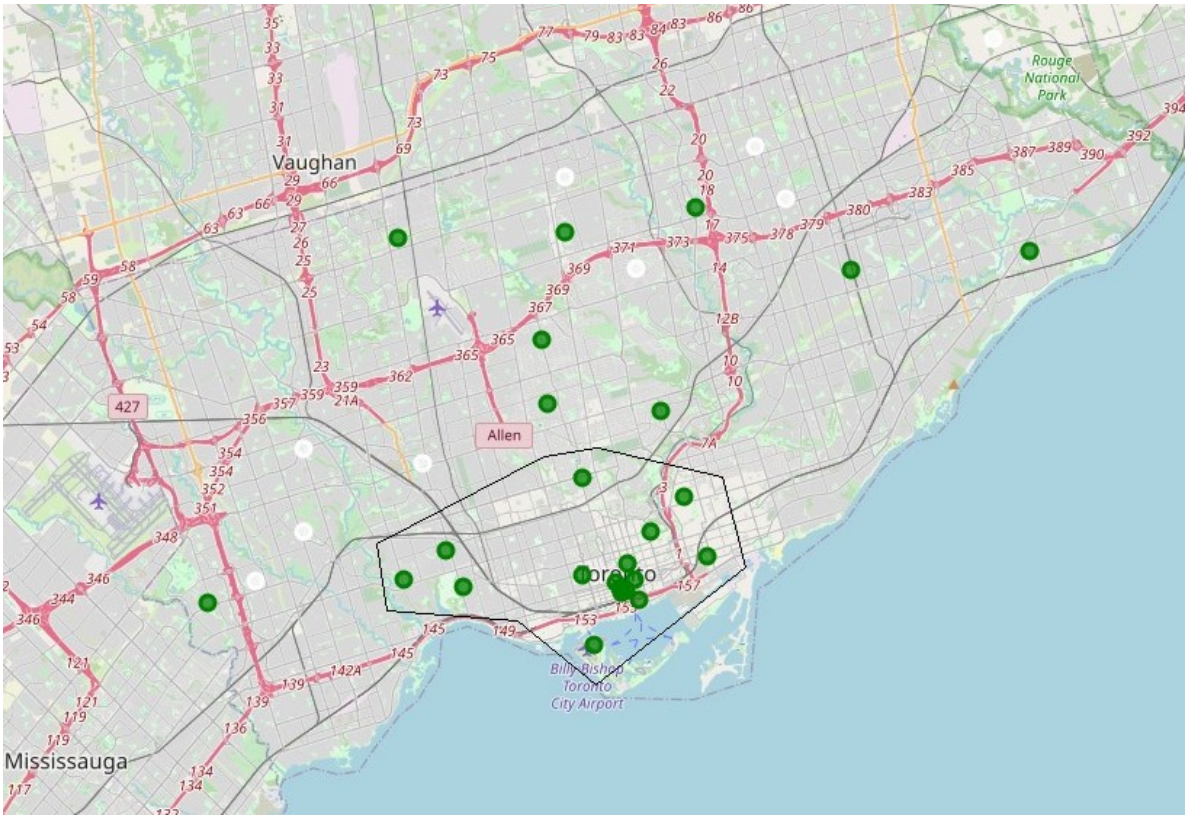


From the final table showing neighborhoods with their respective top 10 venues, I will remove neighborhoods that have these venues - Gym, Yoga Studio, Dance Studio, Park, Pool, Soccer Field, Golf Course, Stadium, Çollege Stadium, Hockey Arena. What's left will be the results.

	Neighbourhood	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	Agincourt	Latin American Restaurant	Skating Rink	Clothing Store	Lounge	Breakfast Spot	Yoga Studio	Dance Studio	Dog Run	Distribution Center	Discount Store
1	Alderwood, Long Branch	Pizza Place	Coffee Shop	Sandwich Place	Skating Rink	Gym	Pharmacy	Pub	Curling Ice	Dance Studio	Deli / Bodega
2	Bathurst Manor, Wilson Heights, Downsview North	Coffee Shop	Bank	Park	Middle Eastern Restaurant	Diner	Sandwich Place	Deli / Bodega	Bridal Shop	Restaurant	Ice Cream Shop
3	Bayview Village	Chinese Restaurant	Bank	Japanese Restaurant	Café	Yoga Studio	Dance Studio	Drugstore	Donut Shop	Dog Run	Distribution Center
4	Bedford Park, Lawrence Manor East	Coffee Shop	Italian Restaurant	Sandwich Place	Restaurant	Greek Restaurant	Pub	Indian Restaurant	Liquor Store	Locksmith	Butcher

5. Results and Discussion:

After removing neighborhoods with sports facilities and parks as their top 10 venues, I use folium to show the neighborhoods that are lack of these facilities. There are a total of 34 neighborhoods and all of them belong to Cluster 0 or Cluster 1. As you can see from the map, many of them are concentrated in Downdown or Central Toronto area, while others scatter around the fringe of the city. This is logical as central area is usually expensive and lack of big space.



For these reasons, a viable business plan is to open fitness centers that target busy working adults in the central area (the area where I draw the boundary) as there are a lot of offices in this area. They can exercise during breaks, before work or after work.

6. Conclusion

This project has shown me the practical application of Data Analysis to resolve a real world situation. I like the K means clustering method and some other powerful libraries available in Python. Data visualization tool like Folium also makes me understand the impact of visualisation and its role in helping to make a better decision.