Capstone Project

Finding the best place for a new Spiritual Development Center in Toronto

Introduction/Business Problem

Discussion of the problem: "Which Toronto neighborhood is the most ideal to start a new Spiritual Development Center?"

My Canadian client devoted to the problem of spiritual growth and personal development is interested to open the Spiritual Development Center in Toronto. She believes that the Center should be situated in the area popular within active people, involved in sport and a healthy lifestyle. She also thinks that restaurants and coffee shops might negatively affect the spiritual growth of her future students, so the ideal neighborhood for the Center should contain less restaurants than venues related to the active lifestyle.

My client asks wants me to consult her on which locality or neighborhood of Toronto is best suited to start the Spiritual Development Center.

Target Audience

Similarly to my client, other businesses can be interested to know which Toronto neighbourhoods attract active people. These information can be used for new venue openings and for advertisement targeting of such venues as sport goods shops, healthy food shops, nutritionists, supplement shops, alternative medicine centers and others.

Data Description:

Description:

This project relies mainly on the data available from Foursquare and Wikipedia.

Dataset 1: List of postal codes of Toronto along with the boroughs and neighborhoods

Since we focus on Toronto in this project, we will be looking to procure all the demographic information related to city, including all the boroughs and neighborhoods of Toronto along with their associated zip code.

To do this, we rely on the publicly available Wikipedia page for the same, titled List of postal codes of Canada:_M. As can be seen from the description of the page: "Postal codes beginning with M are located within the city of Toronto in the province of Ontario."

We will be scraping this data from the Wikipedia page with the help of Python's pandas and Wikipedia packages.

Dataset 2: Geographical coordinates of the neighborhoods

In order to plot the neighborhoods on the map, we will also be using the geographical co-ordinates of the neighborhoods of Toronto. Although this data can be obtained using the Google Maps Geocoding API, given the unreliability of the package we use data from the following source: http://cocl.us/Geospatial_data

In this project, we will be leveraging the Foursquare API to obtain the geographical location data of various neighborhoods of Toronto. This data will be used to explore the venues in each of the neighborhoods, which will help us to identify the best possible location for our client's Center.

We will redefine venues categories to satisfy the client's requirement as "Food provider", "Health provider" and "Indifferent" venues.

Methodology

Neighbourhoods data pre-processing

After scraping the initial data from Wikipedia, there were some improvements required before it could be used for analysis. Some modifications that were made to it were:

- Dropping all the rows from the derived table where boroughs were not assigned
- Combining different neighborhoods with the same postcode
- For neighborhoods with no name, assigning a borough name to it for the purposes of simplicity
- Concatenating the geospatial coordinates obtained from Dataset 2 to this dataframe.

Redefining venues categories

Accordingly to the client's requirements, we need to redefine venues categories as "Food provider", "Health provider" and "Indifferent". To do that, we formed three lists of the Foursquare venue categories using corresponding keywords for "Food provider" for "Health provider". Then we replace the Foursquare venue categories by there new categories.

- *The list of keywords for the selecting "Food provider" category*: restaurant, coffee, spot, caf (excl: gaming), snack, bistro, pizza, place, joint, noodle, cupcake, candy, bakery, food court, ice cream, diner, , steakhouse, gastropub, creperie, tea, beer, brewery, liquor, deli, cheese, dessert, fish & chips, pub, speakeasy, market (excl: flea), bar, food (excl: health), grocery (excl: organic).
- *The list of keywords for the selecting "Health provider" category:* health, gym, field, park, stadium, sport (excl: bar), court (excl: food), studio (excl: recording), supplement, smoothie, organic, skat, pool, swim, bike, trail, arena.
- Foursquare venue categories not selected as "Food provider" and "Health provider" are defined as "Indifferent" venues.

Data preparation for the clustering of neighborhoods

To find group of Toronto neighborhoods with satisfactory venue environment, we used k-means clustering. Each Postal Code area (which could contain a few neighbourhoods) was characterized by the proportion of "Food provider", "Health provider" and "Indifferent" venues. These proportions were employed as features for k-mean clustering.

K-mean clustering of neighborhoods

The reasonable number of clusters (k=3) was selected after manual inspection the results of clustering with different k. As a selection criterion we used the number of neighbourhoods within the same borough which have been distributed to the same cluster by their proportion of "Food provider", "Health provider" and "Indifferent" venues. These method gives us relatively large areas; each contains few adjoining neighbourhoods.

Results

Neighbourhoods data pre-processing

The initial data from Wikipedia (Dataset 1) consisted of 288 rows, contained postal code, borough and neighbourhood names. Dropping all the rows from the derived table where boroughs were not assigned left us 211 rows. Combining different neighborhoods with the same postal code left us 103 rows. After these modifications, Dataset 1 was concatenated with the geospatial coordinates (Dataset 2) into one dataframe. The resulting dataframe contained 103 rows and consisted of postal code, borough and neighbourhood names, latitude and longitude of the centre of each area with the same postal code.

Redefining venues categories

Using Foursquare API we obtained the information about Toronto venues located within 500 metres from the center of each postal code area limiting the number of venues by 300 (Dataset 3). The dataset contained 2233 venues belonged to 274 categories. As described in "Methodology" section, we formed three lists of the Foursquare venue categories using corresponding keywords:

- 113 Foursquare venue categories were redefined as "Food Providers",
- 31 were redefined as "Health Providers",
- others were redefined as "Indifferent" venues (Figure 1).

Figure 1. Words clouds representing the Foursquare venue categories redefined to as A: "Food Providers", B: "Indifferent" venues and C: "Health Providers"

Data preparation for the clustering of neighborhoods

Each postal code area contained one or few neighbourhoods was characterized by the frequency of "Food provider", "Health provider" and "Indifferent" venues (Figure 2). Due to the wide difference in total venue numbers between the areas, the data was converted to the proportions of "Food provider", "Health provider" and "Indifferent" venues in each area.
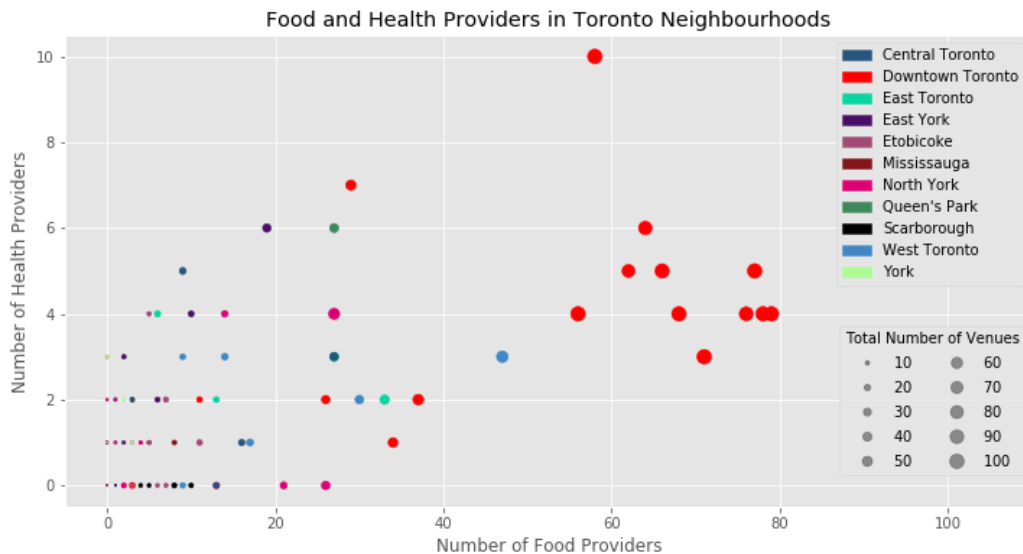
Figure 2. The frequency of "Food provider", "Health provider" and "Indifferent" venues in the Toronto Postal code areas.

K-mean clustering of neighborhoods

After manual inspection the results of clustering with different k, k=3 was selected as providing the best results. All postal code areas were separated to three clusters. Each cluster contained areas characterized by the same most common venue category: "Food provider", "Health provider" and "Indifferent" (Figure 3).

Cluster exploration

"Food Providers"

The cluster characterized by "Food Providers" as the most common venue category by k-mean algorithm is the largest one, contains 61 postal code areas from 11 Toronto boroughs. The highest proportion of "Health Providers" (10%) was found in the area with postal codes started with M5J, which contains Harbourfront East, Toronto Islands, Union Station neighbourhoods (Table 1). The average "Health/food ratio" in this cluster was equal to 0.15 with highest ration of 0.8 demonstrated by postal code area: "M8W" corresponding to Alderwood and Long Branch neighbourhoods.

The cluster characterized by "Indifferent" as the most common venue category by k-mean algorithm contains 26 postal code areas from 7 Toronto boroughs. The highest proportion of "Health Providers" (23%) was found in the area with postal code M4R corresponds to North Toronto West neighbourhood (Table 2). The highest "Health/food ratio" in this cluster was 1.5 demonstrated by postal code area: "M4C" corresponding to Woodbine Heights. There are only 10 venues in the neighbourhood with only two "Food Providers" and three "Health Providers".

Lastly, the cluster characterized by "Health Providers" as the most common venue category contains 16 postal code areas from 8 Toronto boroughs. The highest proportion of "Health Providers" (75%) was found in the areas with postal code M4W and M6C linked to Rosedale neighbourhood in Downtown Toronto and Humewood-Cedarvale in York, correspondingly (Table 3).
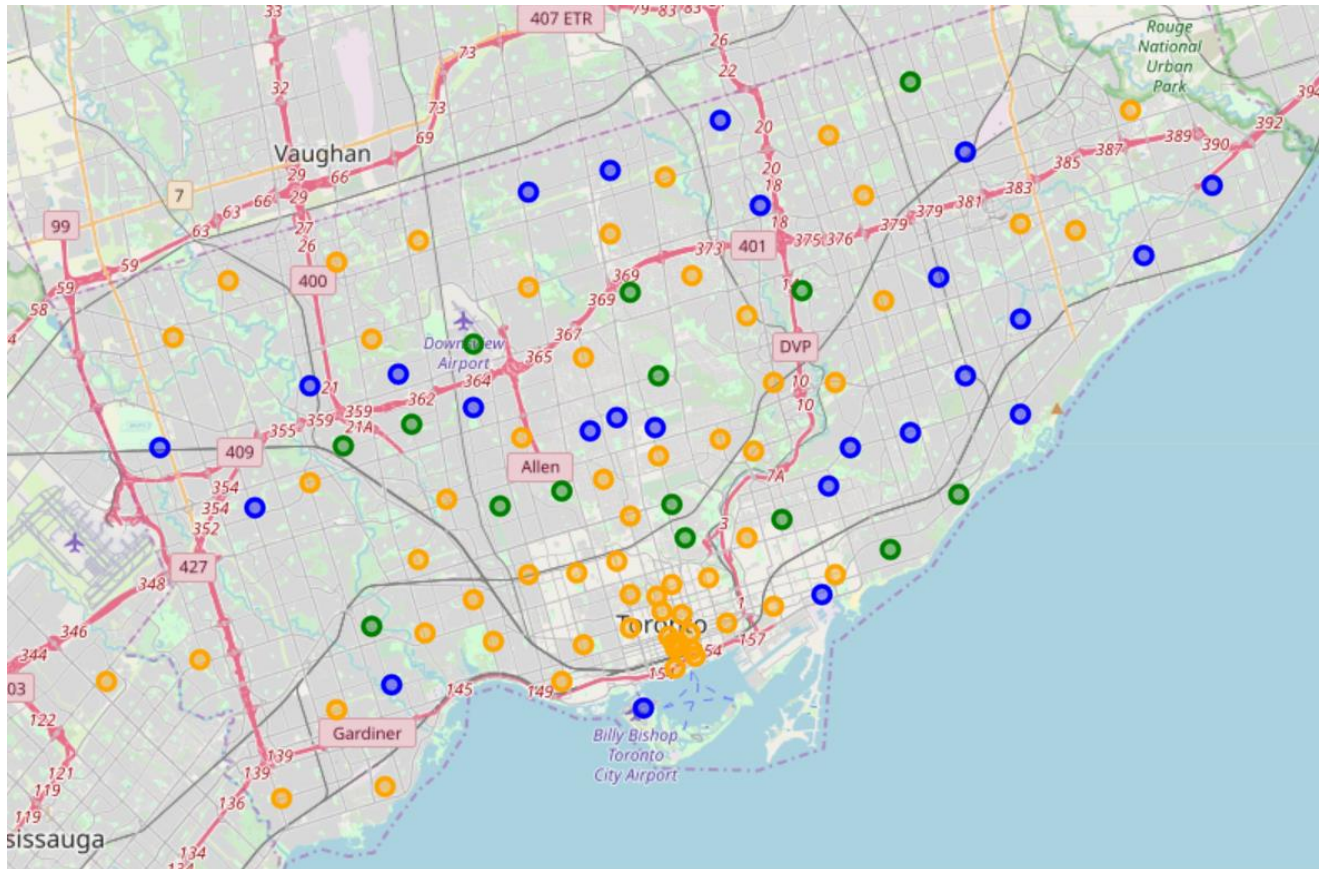

Figure 3. The map of Toronto with postal code areas marked correspondingly to the most common venue category: orange marks, "Food provider"; green marks, "Health provider"; blue marks, "Indifferent" venues. Marks are placed in the geographical centres of each postal code area.

Table 1. Postal codes and neighborhoods directed to the cluster characterized by "Food Providers" as the most common venue category by k-mean algorithm, with number of venues, number and proportion of food and health providers and health/food ratio.

Borough	Neighborhood	Postal Code	Venues	1st Most Common Venue	Health Provider (proportion)	Food Provider (proportion)	Health / Food Ratio
Central Toronto	Davisville	M4S	34	Food Provider	3(0.09)	27(0.79)	0.111111
	Deer Park, Forest Hill SE, Rathnelly, South Hill, Summerhill West	M4V	16	Food Provider	0(0.00)	13(0.81)	0
	Forest Hill North, Forest Hill West	M5P	4	Food Provider	1(0.25)	2(0.50)	0.5
	The Annex, North Midtown, Yorkville	M5R	20	Food Provider	1(0.05)	16(0.80)	0.0625
Downtown Toronto	Adelaide, King, Richmond	M5H	100	Food Provider	3(0.03)	71(0.71)	0.042254
	Berczy Park	M5E	55	Food Provider	2(0.04)	37(0.67)	0.054054
	Cabbagetown, St. James Town	M4X	45	Food Provider	1(0.02)	34(0.76)	0.029412
	Central Bay Street	M5G	80	Food Provider	5(0.06)	62(0.78)	0.080645
	Chinatown, Grange Park, Kensington Market	M5T	92	Food Provider	4(0.04)	76(0.83)	0.052632
	Christie	M6G	16	Food Provider	2(0.13)	11(0.69)	0.181818
	Church and Wellesley	M4Y	87	Food Provider	6(0.07)	64(0.74)	0.09375
	Commerce Court, Victoria Hotel	M5L	100	Food Provider	5(0.05)	77(0.77)	0.064935
	Design Exchange, Toronto Dominion Centre	M5K	100	Food Provider	4(0.04)	79(0.79)	0.050633
	First Canadian Place, Underground city	M5X	100	Food Provider	4(0.04)	78(0.78)	0.051282
	Harbord, University of Toronto	M5S	34	Food Provider	2(0.06)	26(0.76)	0.076923
	Harbourfront East, Toronto Islands, Union Station	M5J	100	Food Provider	10(0.10)	58(0.58)	0.172414
	Harbourfront, Regent Park	M5A	50	Food Provider	7(0.14)	29(0.58)	0.241379
	Ryerson, Garden District	M5B	100	Food Provider	4(0.04)	56(0.56)	0.071429
	St. James Town	M5C	100	Food Provider	4(0.04)	68(0.68)	0.058824
	Stn A PO Boxes 25 The Esplanade	M5W	97	Food Provider	5(0.05)	66(0.68)	0.075758
East Toronto	Studio District	M4M	38	Food Provider	3(0.08)	27(0.71)	0.111111
	The Beaches West, India Bazaar	M4L	17	Food Provider	2(0.12)	13(0.76)	0.153846
	The Danforth West, Riverdale	M4K	42	Food Provider	2(0.05)	33(0.79)	0.060606
East York	Leaside	M4G	33	Food Provider	6(0.18)	19(0.58)	0.315789
	Thorncliffe Park	M4H	17	Food Provider	4(0.24)	10(0.59)	0.4

Table 1 (cont-ed).

Borough	Neighborhood	Postal Code	Venues	1st Most Common Venue	Health Provider (proportion)	Food Provider (proportion)	Health / Food Ratio
Etobicoke	Albion Gardens, Beaumond Heights, Humbergate, Jamestown, Mount Olive, Silverstone, South Steeles, Thistletown	M9V	9	Food Provider	0(0.00)	7(0.78)	0
	Alderwood, Long Branch	M8W	10	Food Provider	4(0.40)	5(0.50)	0.8
	Bloordale Gardens, Eringate, Markland Wood, Old Burnhamthorpe	M9C	9	Food Provider	1(0.11)	5(0.56)	0.2
	Humber Bay Shores, Mimico South, New Toronto	M8V	16	Food Provider	1(0.06)	11(0.69)	0.090909
	Kingsway Park South West, Mimico NW, The Queensway West, Royal York South West, South of Bloor	M8Z	14	Food Provider	2(0.14)	7(0.50)	0.285714
	Westmount	M9P	8	Food Provider	0(0.00)	6(0.75)	0
Mississauga	Canada Post Gateway Processing Centre	M7R	11	Food Provider	1(0.09)	8(0.73)	0.125
North York	Bathurst Manor, Downsview North, Wilson Heights	M3H	20	Food Provider	0(0.00)	13(0.65)	0
	Bayview Village	M2K	4	Food Provider	0(0.00)	3(0.75)	0
	Bedford Park, Lawrence Manor East	M5M	24	Food Provider	0(0.00)	21(0.88)	0
	Don Mills North	M3B	5	Food Provider	2(0.40)	3(0.60)	0.666667
	Downsview Northwest	M3N	4	Food Provider	1(0.25)	2(0.50)	0.5
	Downsview West	M3L	4	Food Provider	1(0.25)	2(0.50)	0.5
	Flemingdon Park, Don Mills South	M3C	21	Food Provider	4(0.19)	14(0.67)	0.285714
	Glencairn	M6B	5	Food Provider	1(0.20)	4(0.80)	0.25
	Humber Summit	M9L	3	Food Provider	0(0.00)	3(1.00)	0
	Northwood Park, York University	M3J	5	Food Provider	1(0.20)	4(0.80)	0.25
	Silver Hills, York Mills	M2L	1	Food Provider	0(0.00)	1(1.00)	0
	Victoria Village	M4A	5	Food Provider	1(0.20)	3(0.60)	0.333333
	Willowdale South	M2N	35	Food Provider	0(0.00)	26(0.74)	0
Queen's Park	Queen's Park	M7A	39	Food Provider	6(0.15)	27(0.69)	0.222222

Table 1 (cont-ed).

Borough	Neighborhood	Postal Code	Venues	1st Most Common Venue	Health Provider (proportion)	Food Provider (proportion)	Health / Food Ratio
Scarborough	Cedarbrae	M1H	7	Food Provider	1(0.14)	5(0.71)	0.2
	Clarks Corners, Sullivan, Tam O'Shanter	M1T	13	Food Provider	0(0.00)	8(0.62)	0
	L'Amoreaux West	M1W	11	Food Provider	0(0.00)	10(0.91)	0
	Maryvale, Wexford	M1R	8	Food Provider	0(0.00)	5(0.63)	0
	Rouge, Malvern	M1B	1	Food Provider	0(0.00)	1(1.00)	0
	Woburn	M1G	3	Food Provider	0(0.00)	3(1.00)	0
West Toronto	Brockton, Exhibition Place, Parkdale Village	M6K	22	Food Provider	3(0.14)	14(0.64)	0.214286
	Dovercourt Village, Dufferin	M6H	16	Food Provider	3(0.19)	9(0.56)	0.333333
	High Park, The Junction South	M6P	24	Food Provider	1(0.04)	17(0.71)	0.058824
	Little Portugal, Trinity	M6J	63	Food Provider	3(0.05)	47(0.75)	0.06383
	Parkdale, Roncesvalles	M6R	15	Food Provider	0(0.00)	9(0.60)	0
	Runnymede, Swansea	M6S	36	Food Provider	2(0.06)	30(0.83)	0.066667
York	Del Ray, Keelesdale, Mount Dennis, Silverthorn	M6M	4	Food Provider	1(0.25)	3(0.75)	0.333333
	The Junction North, Runnymede	M6N	4	Food Provider	0(0.00)	3(0.75)	0

Table 2. Postal codes and neighborhoods directed to the cluster characterized by "Indifferent" as the most common venue category by k-mean algorithm, with number of venues, number and proportion of food and health providers and health/food ratio

Borough	Neighborhood	Postal Code	Venues	1st Most Common Venue	Health Provider (proportion)	Food Provider (proportion)	Health / Food Ratio
Central Toronto	Davisville North	M4P	8	Indifferent	2(0.25)	3(0.38)	0.666667
	North Toronto West	M4R	22	Food Provider	5(0.23)	9(0.41)	0.555556
	Roselawn	M5N	2	Indifferent	0(0.00)	1(0.50)	0
Downtown Toronto	CN Tower, Bathurst Quay, Island airport, Harbourfront West, King and Spadina, Railway Lands, South Niagara	M5V	17	Indifferent	0(0.00)	3(0.18)	0
East Toronto	Business Reply Mail Processing Centre 969 Eastern	M7Y	19	Indifferent	4(0.21)	6(0.32)	0.666667

Table 2 (cont-ed).

Borough	Neighborhood	Postal Code	Venues	1st Most Common Venue	Health Provider (proportion)	Food Provider (proportion)	Health / Food Ratio
East York	Woodbine Gardens, Parkview Hill	M4B	13	Food Provider	2(0.15)	6(0.46)	0.333333
	Woodbine Heights	M4C	10	Indifferent	3(0.30)	2(0.20)	1.5
Etobicoke	Humber Bay, King's Mill Park, Kingsway Park South East, Mimico NE, Old Mill South, The Queensway East, Royal York South East, Sunnylea	M8Y	3	Indifferent	1(0.33)	0(0.00)	NA
	Kingsview Village, Martin Grove Gardens, Richview Gardens, St. Phillips	M9R	4	Indifferent	1(0.25)	1(0.25)	1
	Northwest	M9W	2	Indifferent	0(0.00)	0(0.00)	NA
North York	Downsview Central	M3M	4	Indifferent	1(0.25)	1(0.25)	1
	Emery, Humberlea	M9M	3	Indifferent	1(0.33)	0(0.00)	NA
	Fairview, Henry Farm, Oriole	M2J	59	Indifferent	4(0.07)	27(0.46)	0.148148
	Hillcrest Village	M2H	5	Indifferent	1(0.20)	2(0.40)	0.5
	Lawrence Heights, Lawrence Manor	M6A	13	Indifferent	0(0.00)	2(0.15)	0
	Newtonbrook, Willowdale	M2M	1	Indifferent	0(0.00)	0(0.00)	NA
	Willowdale West	M2R	7	Indifferent	0(0.00)	3(0.43)	0
Scarborough	Agincourt	M1S	5	Indifferent	1(0.20)	2(0.40)	0.5
	Clairlea, Golden Mile, Oakridge	M1L	9	Indifferent	2(0.22)	3(0.33)	0.666667
	Cliffcrest, Cliffside, Scarborough Village West	M1M	2	Indifferent	0(0.00)	1(0.50)	0
	Dorset Park, Scarborough Town Centre, Wexford Heights	M1P	8	Indifferent	0(0.00)	4(0.50)	0
	East Birchmount Park, Ionview, Kennedy Park	M1K	5	Indifferent	0(0.00)	2(0.40)	0
	Guildwood, Morningside, West Hill	M1E	8	Indifferent	0(0.00)	3(0.38)	0
	Highland Creek, Rouge Hill, Port Union	M1C	2	Indifferent	0(0.00)	1(0.50)	0
	Scarborough Village	M1J	1	Indifferent	0(0.00)	0(0.00)	NA

Table 3. Postal codes and neighborhoods directed to the cluster characterized by "Health Providers" as the most common venue category by k-mean algorithm, with number of venues, number and proportion of food and health providers and health/food ratio

Borough	Neighborhood	Postal Code	Venues	1st Most Common Venue	Health Provider (proportion)	Food Provider (proportion)	Health / Food Ratio
Central Toronto	Lawrence Park	M4N	4	Health Provider	2(0.50)	1(0.25)	2
	Moore Park, Summerhill East	M4T	2	Indifferent	1(0.50)	0(0.00)	NA
Downtown Toronto	Rosedale	M4W	4	Health Provider	3(0.75)	0(0.00)	NA
East Toronto	The Beaches	M4E	4	Health Provider	2(0.50)	1(0.25)	2
East York	East Toronto	M4J	2	Indifferent	1(0.50)	0(0.00)	NA
Etobicoke	The Kingsway, Montgomery Road, Old Mill North	M8X	2	Indifferent	1(0.50)	0(0.00)	NA
North York	CFB Toronto, Downsview East	M3K	2	Indifferent	1(0.50)	0(0.00)	NA
	Downsview, North Park, Upwood Park	M6L	4	Health Provider	2(0.50)	1(0.25)	2
	Parkwoods	M3A	2	Health Provider	1(0.50)	1(0.50)	1
	York Mills West	M2P	4	Indifferent	2(0.50)	0(0.00)	NA
Scarborough	Agincourt North, L'Amoreaux East, Milliken, Steeles East	M1V	3	Indifferent	1(0.33)	1(0.33)	1
	Birch Cliff, Cliffside West	M1N	4	Health Provider	2(0.50)	1(0.25)	2
York	Caledonia-Fairbanks	M6E	5	Health Provider	2(0.40)	2(0.40)	1
	Humewood-Cedarvale	M6C	4	Health Provider	3(0.75)	0(0.00)	NA
	Weston	M9N	1	Health Provider	1(1.00)	0(0.00)	NA

Discussion

The observed clusters allow us to choose the neighbourhoods for the opening of new Spiritual Development Center. It can be situated in York (Caledonia-Fairbanks, Humewood-Cedarvale, Weston) or North York (CFB Toronto, Downsview East, Downsview, North Park, Upwood Park, Parkwoods, York Mills West). However, I would recommend my client to conduct deeper research which include population data in different Toronto neighbourhoods and marketing research data. The modification of methodology by rethinking of influence of the restaurants and food markets on the spiritual development needs might be useful as well.

Conclusion

Despite the clear and univocal results provided by our research, the decision cannot be made using the limitations of the client's data and requirements.