

# *The Battle of Neighborhoods*

## Capstone Project: Doug Marcum

### Contents

- |   |                                      |                                       |
|---|--------------------------------------|---------------------------------------|
| 1) <a href="#"><u>Introduction/Business Problem</u></a> | 2) <a href="#"><u>Data</u></a>       | 3) <a href="#"><u>Methodology</u></a> |
| 4) <a href="#"><u>Results</u></a>                       | 5) <a href="#"><u>Discussion</u></a> | 6) <a href="#"><u>Conclusion</u></a>  |

### 1) Introduction/Business Problem

The basis of this study is to help a small group of investors planning to open their first U.S. based brewery / restaurant expansion in Toronto. Being that Toronto is the most populated city in Canada, and continually ranks as an important global city based on a high quality of living, the choice to expand into the neighbor of the north market was an easy selection for the investing group. However, with limited knowledge of the Toronto market, the group of investors have selected us to assist in the selection of which areas of Toronto will facilitate a launch of their brewery / restaurant expansion.

They are interested in building in an area that meets the following criteria:

- *A neighborhood with an average to above average median total population*
- *Above average populations of 25-40-year-old male and female professionals*
- *A high concentration of the population having secondary education*
- *Average to above average median net household incomes*

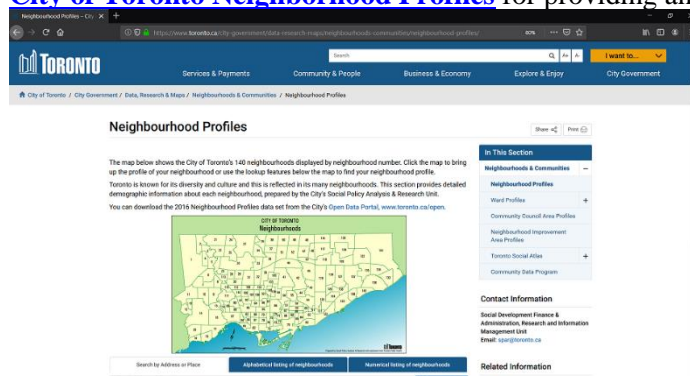
With these criteria given by the investing group, based on previous success in other markets, the objective is to locate and recommend to the investors, the target audience, which neighborhood(s) of Toronto will be the best choice to start their international growth plan. The information gained will assist in choosing the right location by providing data about the population of each neighborhood, in addition to other established venues present in these areas.

Additionally, this information could be of interest to other potential investors looking to open a new restaurant or entertainment venue in Toronto.

### 2) Data

The necessary information needed by the investing group will come from the following sources:

- [City of Toronto Neighborhood Profiles](#) for providing an overview of the neighborhoods in Toronto



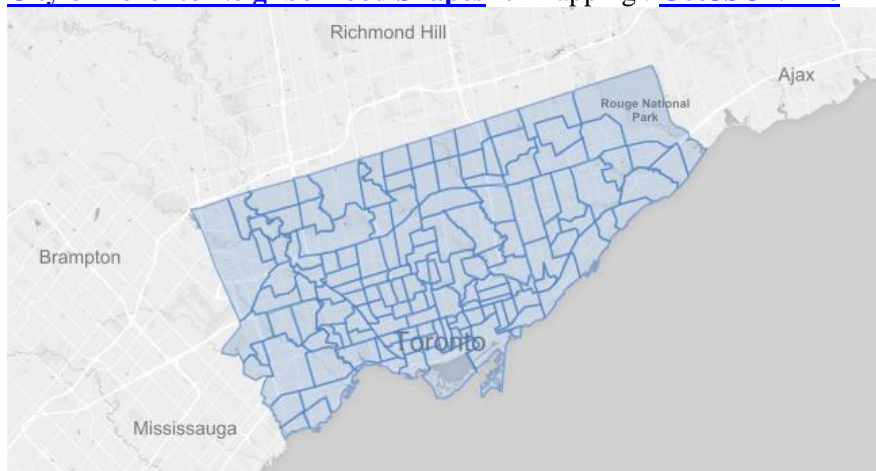
- [City of Toronto Open Data Catalogue](#) : *The Census of Population is held across Canada every five years (the last being in 2016), and collects data about age and sex, families and households, language, immigration and internal migration, ethnocultural diversity, Aboriginal peoples, housing, education, income, and labor. City of Toronto Neighborhood Profiles use this Census data to provide a portrait of the demographic, social and economic characteristics of the people and households in each City of Toronto neighborhood. The profiles present selected highlights from the data, but these accompanying data files provide the full data set assembled for each neighborhood.*

*In these profiles of the City of Toronto's 140 social planning neighborhoods. These social planning neighborhoods were developed by the City of Toronto to help government and community organizations with local planning by providing socio-economic data at a meaningful geographic area. The boundaries of these social planning neighborhoods are consistent over time, allowing for comparison between Census years. Neighborhood level data from a variety of other sources are also available through the City's Wellbeing Toronto mapping application and here on the Open Data portal.*

*Each data point in this file is presented for the City's 140 neighborhoods, as well as for the City of Toronto as a whole. The data is sourced from several Census tables released by Statistics Canada. The general Census Profile is the main source table for this data, but other Census tables have also been used to provide additional information. [CSV File](#)*

Category	Topic	Data Source	Characteristic	City of Toronto	Agincourt North	Agincourt South-Malvern West	Alderwood	Annex	Banbury-Don Mills	Bathurst Manor	Bayview
0	Neighbourhood Information	City of Toronto	Neighbourhood Number	NaN	129	128	20	95	42	34	
1	Neighbourhood Information	City of Toronto	TSNS2020 Designation	NaN	No Designation	No Designation	No Designation	No Designation	No Designation	No Designation	No Designation
2	Population	Census Profile 98-316-X2016001	Population, 2016	2,731,571	29,113	23,757	12,054	30,526	27,695	15,873	2
3	Population	Census Profile 98-316-X2016001	Population, 2011	2,615,060	30,279	21,988	11,904	29,177	26,918	15,434	1
4	Population	Census Profile 98-316-X2016001	Population Change 2011-2016	4.50%	-3.90%	8.00%	1.30%	4.60%	2.90%	2.80%	35

- [City of Toronto Neighborhood Shapes](#) for mapping : [GeoJSON File](#)



- [Wikipedia for Toronto Neighborhood Borough Designation](#) : Each of the 140 social planning neighborhoods of Toronto reside within a defined borough. While the City of Toronto is a singular municipality, the 140 neighborhoods are still grouped into six distinct boroughs.

	neighborhood	borough
0	Agincourt North	Scarborough
1	Agincourt South-Malvern West	Scarborough
2	Alderwood	Etobicoke
3	Annex	Old City of Toronto
4	Banbury-Don Mills	North York

- [Foursquare API](#) to collect information on other venues/competitors in the neighborhoods of Toronto

	neighborhood	neighborhood_latitude	neighborhood_longitude	venue	venue_latitude	venue_longitude	venue_category
0	Agincourt North (129)	43.805441	-79.266712	Saravanaa Bhavan South Indian Restaurant	43.810117	-79.269275	Indian Restaurant
1	Agincourt North (129)	43.805441	-79.266712	Fahmee Bakery & Jamaican Foods	43.810170	-79.280113	Caribbean Restaurant
2	Agincourt North (129)	43.805441	-79.266712	Samosa King - Embassy Restaurant	43.810152	-79.257316	Indian Restaurant
3	Agincourt North (129)	43.805441	-79.266712	Grandeur Palace 華麗宮 (Grandeur Palace 華麗宮)	43.797885	-79.270585	Chinese Restaurant
4	Agincourt North (129)	43.805441	-79.266712	Bestco Food Market 鴻華超級市場	43.796514	-79.270790	Supermarket

### 3) Methodology

In order to establish the targeted neighborhood(s), we will explore the demographics of the neighborhoods in the city of Toronto by segmenting the data and conducting descriptive analysis using Panda. Additional data will be gleaned by web scraping and API will be used to generate data.

#### Data Group 1

##### Stage A – Census Data

1. Data was pulled into from the City of Toronto Neighborhoods Profile Census CSV File to create a dataframe.
2. This dataframe contains all the census data (2016) of the neighborhoods of Toronto that will be filtered.
3. Data is filtered into columns based on neighborhood population, male and female age groups, education level, and after-tax income.

	neighborhood	population	male	female	higher_education	after_tax_income
0	Agincourt North	29113	2530	2715	4240	26955
1	Agincourt South-Malvern West	23757	2490	2525	4615	27928
2	Alderwood	12054	1220	1230	1980	39159
3	Annex	30526	4745	4980	12640	80138
4	Banbury-Don Mills	27695	2115	2500	8060	51874

##### Stage B – Web scraping to align neighborhoods with boroughs

1. [Wikipedia page](#) for Toronto Neighborhood Borough Designations is scraped using BeautifulSoup.
2. Scraped data is transformed to dataframe.
3. Merge this dataframe with Census Data dataframe.

```
In [9]: df_xy = pd.merge(df_pc, toronto_hoods, left_index=True, right_index=True, how='inner')
df_xy = df_xy.drop(['neighborhood_y'], axis=1)
df_xy = df_xy.rename(columns={'neighborhood_x': 'neighborhood'})
df_xy.head()
```

Out[9]:

	neighborhood	borough	population	male	female	higher_education	after_tax_income
0	Agincourt North	Scarborough	29113	2530	2715	4240	26955
1	Agincourt South-Malvern West	Scarborough	23757	2490	2525	4615	27928
2	Alderwood	Etobicoke	12054	1220	1230	1980	39159
3	Annex	Old City of Toronto	30526	4745	4980	12640	80138
4	Banbury-Don Mills	North York	27695	2115	2500	8060	51874

##### Stage C – Pull Toronto shape file

1. Get the [shape file](#).
2. Remove unnecessary data and merge to previous dataframe.

	borough	neighborhood	longitude	latitude	population	male	female	higher_education	after_tax_income
0	Scarborough	Agincourt North (129)	-79.266712	43.805441	29113	2530	2715	4240	26955
1	Scarborough	Agincourt South-Malvern West (128)	-79.265612	43.788658	23757	2490	2525	4615	27928
2	Etobicoke	Alderwood (20)	-79.541611	43.604937	12054	1220	1230	1980	39159
3	Old City of Toronto	Annex (95)	-79.404001	43.671585	30526	4745	4980	12640	80138
4	North York	Banbury-Don Mills (42)	-79.349718	43.737657	27695	2115	2500	8060	51874

## Data Group 2

### Stage A – Establish Medians and scoring system

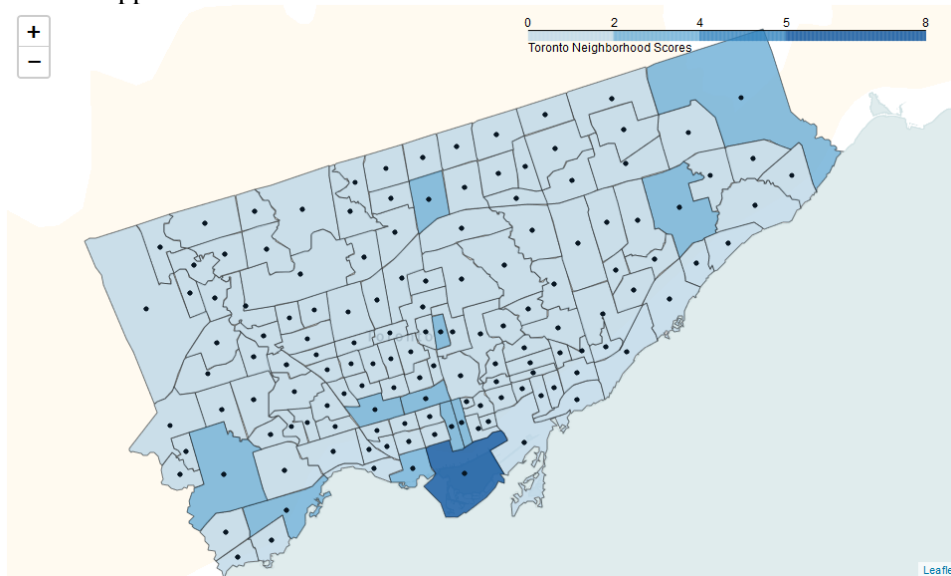
- Calculate medians of the demographic columns across the 140 neighborhoods.

Median Population: 16749.5      Median Higher Education: 4122.5  
 Median Male: 1800.0      Median After Tax Income: 36538.5  
 Median Female: 1952.5

- From the criteria delivered by the investor group, each category was given a standardized score based the category being divided by its median score and then multiplied by a factor of given importance. The columns are the summed to create a total score for each neighborhood. The dataframe is then merged to create a dataframe with all necessary data.

	borough	neighborhood	longitude	latitude	pop_score	male_score	female_score	edu_score	income_score	total_score
0	Scarborough	Agincourt North (129)	-79.266712	43.805441	0.260721	0.351389	0.347631	0.154275	0.147543	1.26
1	Scarborough	Agincourt South-Malvern West (128)	-79.265612	43.788658	0.212756	0.345833	0.323303	0.167920	0.152869	1.20
2	Etobicoke	Aldenwood (20)	-79.541611	43.604937	0.107949	0.169444	0.157490	0.072044	0.214344	0.72
3	Old City of Toronto	Annex (95)	-79.404001	43.671585	0.273375	0.659028	0.637644	0.459915	0.438650	2.47
4	North York	Banbury-Don Mills (42)	-79.349718	43.737657	0.248022	0.293750	0.320102	0.293269	0.283942	1.44

- From here, utilizing a choropleth folium map, a clearer picture of the neighborhoods of Toronto becomes apparent.



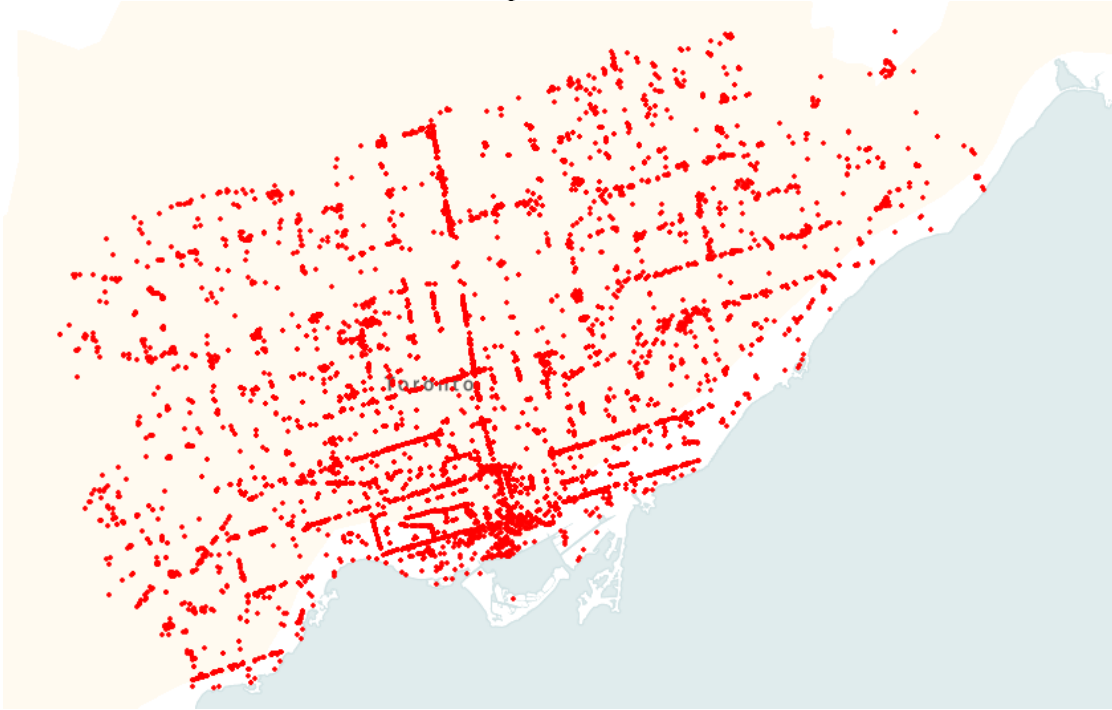
## Data Group 3

### Stage A – Toronto geographical data is utilized as the input into the Foursquare API, that returned venues from all of Toronto

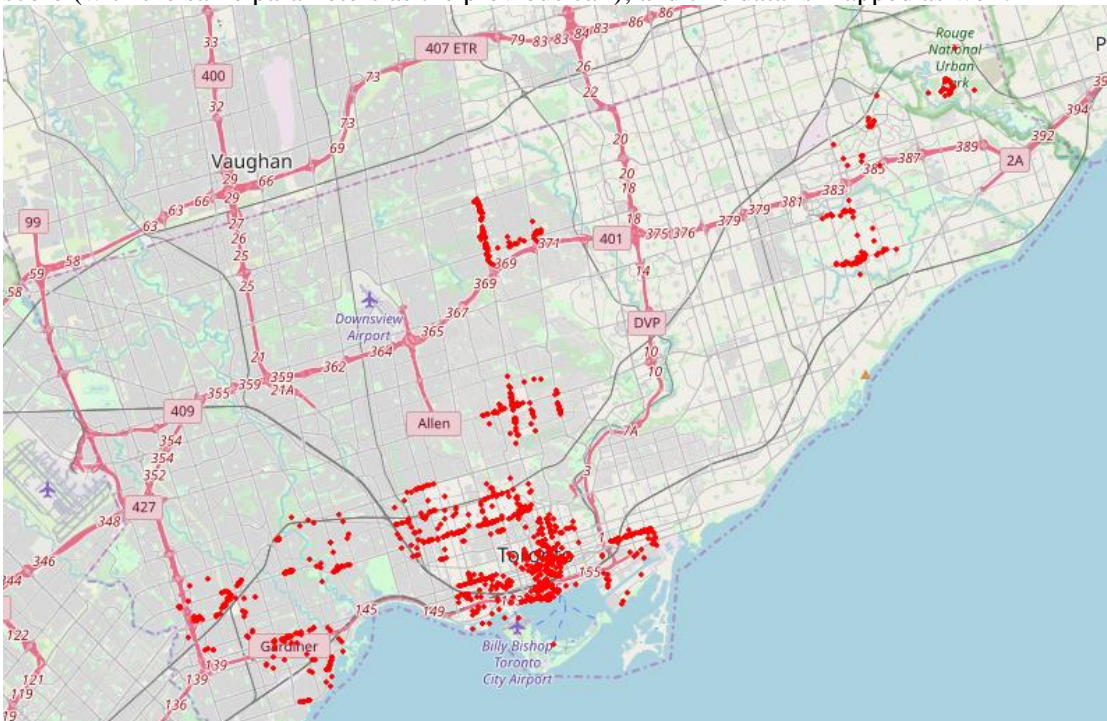
- Using the geographical coordinates of each neighborhood in Toronto, calls are made to the Foursquare API to return the top 100 venues in a radius of 1610 meters, approximately a one-mile radius.

	neighborhood	neighborhood_latitude	neighborhood_longitude	venue	venue_latitude	venue_longitude	venue_category
0	Agincourt North (129)	43.805441	-79.266712	Saravanaa Bhavan South Indian Restaurant	43.810117	-79.269275	Indian Restaurant
1	Agincourt North (129)	43.805441	-79.266712	Fahmee Bakery & Jamaican Foods	43.810170	-79.280113	Caribbean Restaurant
2	Agincourt North (129)	43.805441	-79.266712	Samosa King - Embassy Restaurant	43.810152	-79.257316	Indian Restaurant
3	Agincourt North (129)	43.805441	-79.266712	Grandeur Palace 華麗宮 (Grandeur Palace 華麗宮)	43.797885	-79.270585	Chinese Restaurant
4	Agincourt North (129)	43.805441	-79.266712	Bestco Food Market 鴻華超級市場	43.796514	-79.270790	Supermarket

- 
2. The data is then visualized via a folium map.



- 
- 
3. Foursquare is called on again to narrow the list to the neighborhoods within the top fifteen (15) total score (with the same parameters as the previous call), and this data is mapped as well.





## Data Group 4

Stage A – Analyze the Top 15 neighborhoods in Toronto

- Utilize one hot encoding to transform our list of established venues in the top 15 neighborhoods to return a shape of (1198, 208).

	neighborhood	Airport	American Restaurant	Antique Shop	Aquarium	Arcade	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	BBQ Joint	Baby Store	Bagel Shop	Bakery	Bank	Bar
0	Waterfront Communities-The Island (77)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	Waterfront Communities-The Island (77)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	Waterfront Communities-The Island (77)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	Waterfront Communities-The Island (77)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	Waterfront Communities-The Island (77)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

- Group by neighborhoods. Filter out venues related to retail and personal care categories to focus on activity centered venues (i.e. bars, nightclubs, restaurants, attractions, etc...). Create a small dataframe to display totals for each neighborhood.

	neighborhood	total
0	Annex (95)	57
1	Bay Street Corridor (76)	59
2	Church-Yonge Corridor (75)	62
3	Dovercourt-Wallace Emerson-Junction (93)	59
4	Islington-City Centre West (14)	36
5	Lambton Baby Point (114)	34
6	Malvern (132)	10
7	Mimico (includes Humber Bay Shores) (17)	51
8	Mount Pleasant West (104)	53
9	Niagara (82)	56
10	Rouge (131)	2
11	South Riverdale (70)	58
12	Waterfront Communities-The Island (77)	59
13	Willowdale East (51)	56
14	Woburn (137)	30

- Drop neighborhood Rouge due to extreme limitations on venues after final filter.

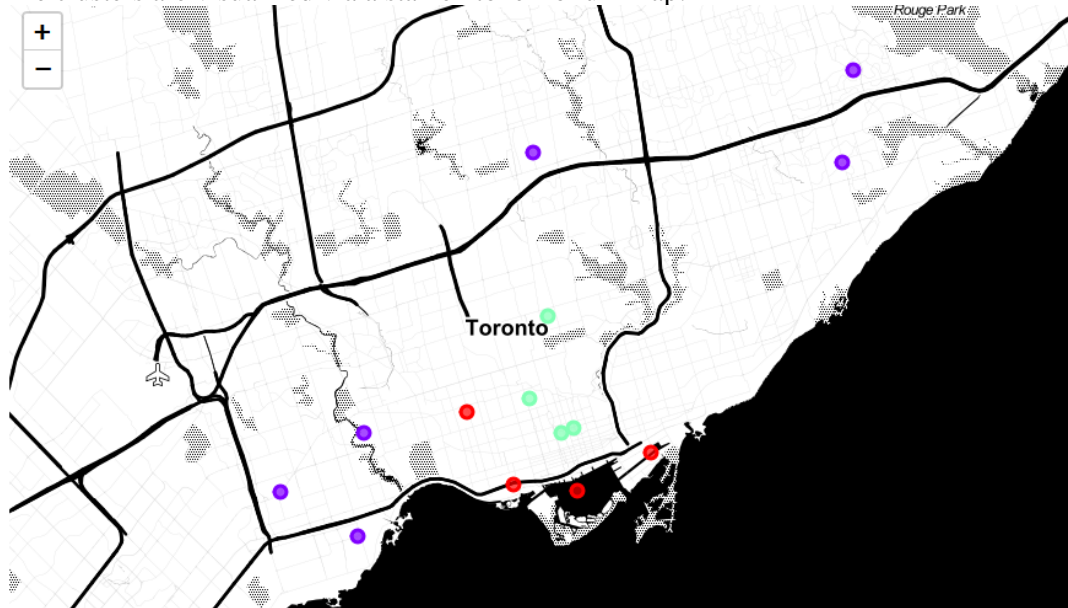
## 4) Results

With the data now ready, we run k-means to cluster the neighborhoods into three (3) clusters. The cluster number was established after multiple samplings and iterations. With our clusters established, this dataframe is merged with the total scores data to provide us with our final pieces of criteria in selecting the appropriate neighborhood(s).

1. The final dataframe:

	borough	neighborhood	longitude	latitude	total_score	cluster_labels
0	Old City of Toronto	Waterfront Communities-The Island (77)	-79.377202	43.633880	6.86	0
1	North York	Niagara (82)	-79.412420	43.636681	3.53	0
2	North York	Willowdale East (51)	-79.401484	43.770602	3.46	1
3	Old City of Toronto	Church-Yonge Corridor (75)	-79.379017	43.659649	2.73	2
4	Scarborough	Islington-City Centre West (14)	-79.543317	43.633463	2.63	1
5	Scarborough	Dovercourt-Wallace Emerson-Junction (93)	-79.438541	43.665677	2.48	0
6	Old City of Toronto	Mount Pleasant West (104)	-79.393360	43.704435	2.48	2
7	Old City of Toronto	Annex (95)	-79.404001	43.671585	2.47	2
8	Scarborough	Woburn (137)	-79.228586	43.766740	2.45	1
9	Scarborough	Mimico (includes Humber Bay Shores) (17)	-79.500137	43.615924	2.27	1
10	Old City of Toronto	Bay Street Corridor (76)	-79.385721	43.657511	2.19	2
11	Etobicoke	Malvern (132)	-79.222517	43.803658	1.92	1
12	Scarborough	South Riverdale (70)	-79.335651	43.649292	1.89	0
13	Scarborough	Lambton Baby Point (114)	-79.496045	43.657420	1.86	1

2. The clusters are visualized via a stamen toner folium map:



## 5) Discussion

From the results discovered and presented, the following observations and recommendations can be made:

1. Based on the criteria given by the investor group and the cluster data, the main neighborhood recommendation would be for the neighborhood of Waterfront Communities – The Island. With a total score of 6.86 and being in the cluster 0 (a mean of 58 venues/neighborhood), this area is a prime location for a flagship brewery / restaurant.
2. A secondary recommendation is made for the neighborhood of Islington-City Centre West. This neighborhood a ranking in the top five total score assessment, as well as having less competition (only 36 venues within a one-mile radius).

3. Additionally, from the clustering, it becomes clear that city of Toronto radiates from that point in a circular pattern outward in terms of entertaining attractions. While it was not included in the research and analysis, commercial square footage rental rates tend to follow this trend as well.

## **6) Conclusion**

In conclusion, the scope of this of the analysis is somewhat limited. The hospitality industry is ever changing, and the information afforded us may be dated due to relying on user information via Foursquare. Overall though, the model created can easily be replicated again and again with monitored data via the Foursquare API and the data from the forthcoming census in 2021. With the data analyzed and scoring system established by the investor group, we stand by the recommendations made.