

# Finding the Ideal Location to set up a Coffee Shop in Toronto

Eric Tay

July 15 2020

## 1. Introduction:

A business owner is looking to open up a coffee shop (mid-high range) in Toronto, and is looking for advice as to where to open up at. The project will be focused on generating the data required to convince the business owner to open the business at a specific neighbourhood/location, based on factors such as customer profiles and proximity of competing businesses.

Ideally, we would want to target a location with a large population, large percentage of working population, high income and with few competing businesses to avoid cannibalizing on market share. This is to ensure that the business can draw in customers effectively.

## 2. Data:

To execute this project, we need the following data:

### 1. FourSquare Api

- to get number of coffee shops/cafes per neighborhood
- a sample query looks like this: [https://api.foursquare.com/v2/venues/explore?client\\_id=](https://api.foursquare.com/v2/venues/explore?client_id=)

### 2. Wikipedia (Mapping Neighborhoods to Bouroughs)

- to get Borough names from Neighborhoods.
- Source: [https://en.wikipedia.org/wiki/List\\_of\\_postal\\_codes\\_of\\_Canada:\\_M](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M)

### 3. Neighborhood Profiles

- to understand neighborhood profiles for Toronto
- Main source from Canadian Census, but incorporates other data sources as well
- Source: [https://ckan0.cf.opendata.inter.prod-toronto.ca/download\\_resource/ef0239b1-832b-4d0b-a1f3-4153e53b189e?format=csv](https://ckan0.cf.opendata.inter.prod-toronto.ca/download_resource/ef0239b1-832b-4d0b-a1f3-4153e53b189e?format=csv)

### 4. Toronto Geospatial Data

- As geopy is a little unreliable, I would be falling back on the geospatial data provided by the city of Toronto for each region.
- Source: "[https://ckan0.cf.opendata.inter.prod-toronto.ca/download\\_resource/a083c865-6d60-4d1d-b6c6-b0c8a85f9c15?format=csv&projection=4326](https://ckan0.cf.opendata.inter.prod-toronto.ca/download_resource/a083c865-6d60-4d1d-b6c6-b0c8a85f9c15?format=csv&projection=4326)"

## 3. Methodology:

Firstly, I combined sources 2 and 4 to create the geospatial profile of all the neighborhoods in Toronto. Secondly, I added on the neighborhood profile from source 3 to complete the picture (Population, Working Population and average income after taxes)

From the population, % working population and average income data, I scored each region based how each of the features compares with the median ( $\frac{x}{x_{median}}$ ), and applied an equal weightage to all 3 factors.

I summed the scores across all the 3 features (in the variable total\_score) and narrowed down the top 40 neighbourhoods based on this initial scoring.

Following that, I extracted coffee shops/cafes/tea room data tagged to the 40 neighborhoods from the Foursquare API data. The number of such shops, combined with the total\_score variable, were the 2 features used in k-means clustering of the data on a geomap. A k of 4 was used in the clustering.

## 4. Results:

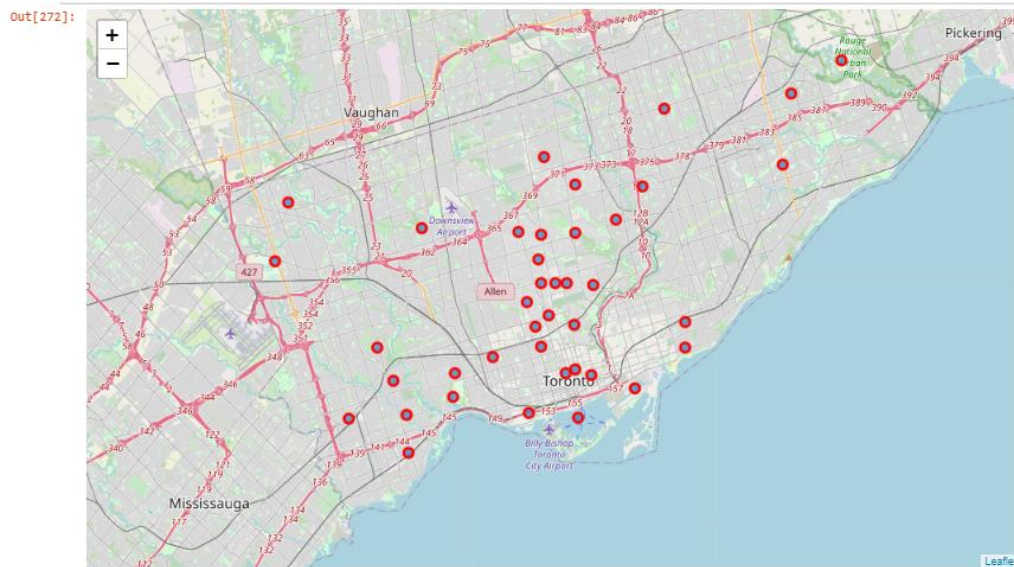
The merging of all the relevant datasets (Sources 2, 3 and 4) yielded this table below:

Merging Census Data with Geodata								
The below will show the joining of Census data with Geospatial data								
<pre>In [114]: df_total=pd.merge(toronto_census,df_pos,on='CDN') df_total=df_total.sort_values('CDN').reset_index().drop(['index','WorkingPop'], axis=1) df_total</pre>								
<pre>Out[114]:</pre>								
	neighborhood	CDN	population	after_tax_income	% Working Pop	borough	LONGITUDE	LATITUDE
0	West Humber-Clairville	1	33312	28066	0.415616	Etobicoke	-79.596356	43.716180
1	Mount Olive-Silverstone-Jamestown	2	32954	24122	0.413152	Etobicoke	-79.587259	43.746868
2	Thistletown-Beaumont Heights	3	10360	28842	0.401544	Etobicoke	-79.563491	43.737988
3	Rexdale-Kipling	4	10529	30201	0.408396	Etobicoke	-79.566228	43.723725
4	Elms-Old Rexdale	5	9456	28355	0.391286	Etobicoke	-79.548983	43.721519
5	Kingsview Village-The Westway	6	22000	31447	0.392500	Etobicoke	-79.547863	43.698993
6	Willowridge-Martingrove-Richview	7	22156	36713	0.367395	Etobicoke	-79.554221	43.683645
7	Humber Heights-Westmount	8	10948	38150	0.346182	Etobicoke	-79.522416	43.692233
8	Edenbridge-Humber Valley	9	15535	72156	0.382362	Etobicoke	-79.522458	43.670886
9	Princess-Rosethorn	10	11051	71025	0.346123	Etobicoke	-79.544559	43.666051
10	Eringate-Centennial-West Deane	11	18588	39203	0.371745	Etobicoke	-79.580445	43.658017

The scoring table is as follows (Only the first 5 rows are shown)

	neighborhood	CDN	borough	LONGITUDE	LATITUDE	pop_score	income_score	work_score	total_score
79	Waterfront Communities-The Island	77	Old City of Toronto	-79.377202	43.633880	1.312427	0.497174	0.533259	2.34
43	Bridle Path-Sunnybrook-York Mills	41	North York	-79.378904	43.731013	0.184500	1.771537	0.259867	2.22
100	Rosedale-Moore Park	98	Old City of Toronto	-79.379669	43.682820	0.416608	1.235014	0.295162	1.95
103	Forest Hill South	101	Old City of Toronto	-79.414318	43.694526	0.213690	1.306094	0.307509	1.83
53	Willowdale East	51	North York	-79.401484	43.770602	1.004217	0.336443	0.399413	1.74

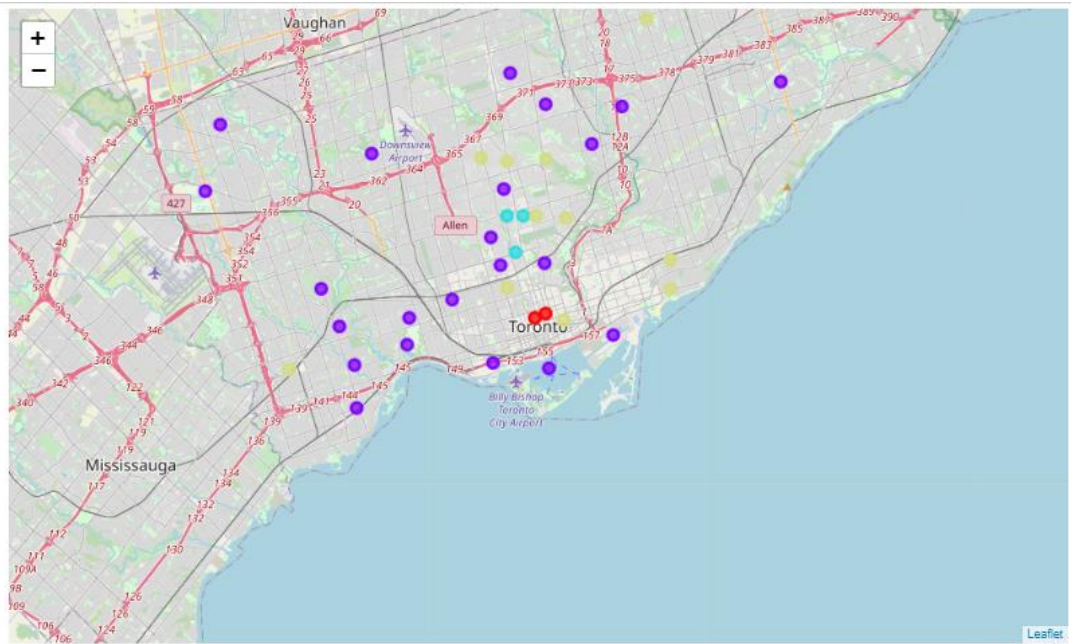
From this scoring, only the top 40 neighborhoods were chosen, and the map below shows the location of these neighborhoods:



From the Foursquare API data, I was able to pull out the number of coffee shops per district, as shown in the table below:

neighborhood	No. of Coffee Shops
Bay Street Corridor	30
Church-Yonge Corridor	30
Mount Pleasant West	18
Yonge-St.Clair	15
Yonge-Eglinton	11
The Beaches	8
Mount Pleasant East	6
Annex	6
Moss Park	4
Lawrence Park North	4
Bedford Park-Nortown	4
Bridle Path-Sunnybrook-York Mills	3
Leaside-Bennington	3
East End-Danforth	3
L'Amoreaux	3
Islington-City Centre West	3
Banbury-Don Mills	2
Malvern	2
Casa Loma	2
Niagara	2
Rouge	2
Mimico (includes Humber Bay Shores)	1
High Park North	1
Mount Olive-Silverstone-Jamestown	1
Dovercourt-Wallace Emerson-Junction	1
West Humber-Clairville	1
High Park-Swansea	1
Lawrence Park South	1

The clustering with K means and the resultant map is shown below:

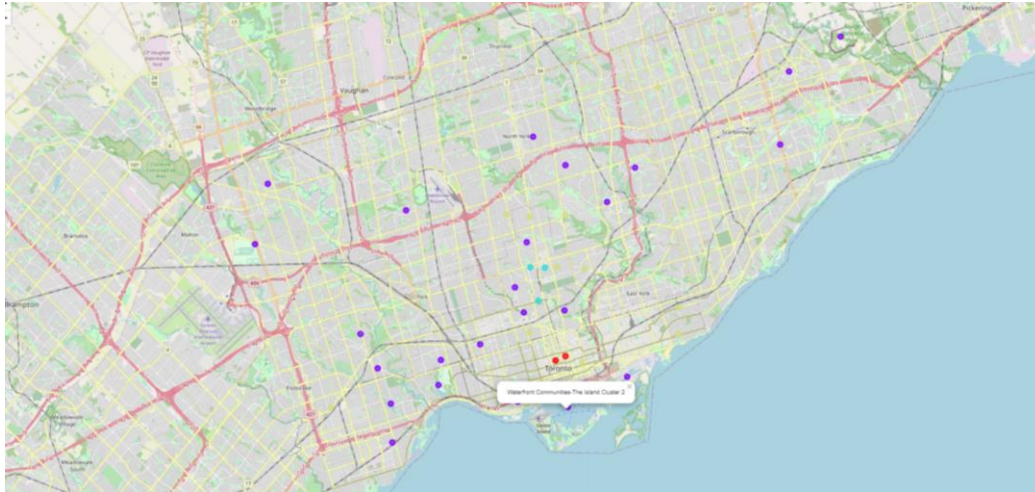


The table below shows the finalized dataset:

	Clusters	L	neighborhood	CDN	borough	LONGITUDE	LATITUDE	pop_score	Income_score	work_score	total_score	No. of Coffee Shops
39	1		Waterfront Communities-The Island	77	Old City of Toronto	-79.377202	43.633880	1.312427	0.497174	0.533259	2.34	0.0
28	1		Rosedale-Moore Park	98	Old City of Toronto	-79.379669	43.682820	0.416608	1.235014	0.295162	1.95	0.0
30	1		Edenbridge-Humber Valley	9	Etobicoke	-79.522458	43.670886	0.309325	0.660762	0.297961	1.27	0.0
31	1		Forest Hill South	101	Old City of Toronto	-79.414318	43.694526	0.213690	1.306094	0.307509	1.83	0.0
32	1		Woburn	137	Scarborough	-79.228586	43.766740	1.064967	0.250373	0.319734	1.64	0.0
29	1		Stonegate-Queensway	16	Etobicoke	-79.501128	43.635518	0.498803	0.448731	0.331602	1.28	0.0
34	1		StAndrew-Windfields	40	North York	-79.379037	43.756248	0.354664	0.649123	0.297496	1.30	0.0
35	1		Downsview-Roding-CFB	26	North York	-79.490497	43.733292	0.697938	0.272653	0.334475	1.31	0.0
36	1		Parkwoods-Donalda	45	North York	-79.330180	43.755033	0.693020	0.324255	0.330916	1.35	0.0
37	1		Kingsway South	15	Etobicoke	-79.510577	43.653520	0.184600	0.895924	0.272755	1.35	0.0
33	1		Willowdale East	51	North York	-79.401484	43.770602	1.004217	0.336443	0.399413	1.74	0.0
38	1		South Riverdale	70	Old City of Toronto	-79.335651	43.649292	0.555053	0.396488	0.420019	1.37	0.0
27	1		High Park North	88	Old City of Toronto	-79.466302	43.657565	0.441279	0.420261	0.409991	1.27	1.0
26	1		Mount Olive-Silverstone-Jamieson	2	Etobicoke	-79.587259	43.746868	0.656164	0.220895	0.321954	1.20	1.0
25	1		High Park-Swansea	87	Old City of Toronto	-79.467872	43.645065	0.476383	0.502485	0.376848	1.36	1.0
24	1		Lawrence Park South	103	Old City of Toronto	-79.406039	43.717212	0.302237	1.021839	0.301356	1.63	1.0
23	1		Mimico (includes Humber Bay Shores)	17	Etobicoke	-79.500137	43.615924	0.678274	0.402660	0.405991	1.48	1.0
22	1		West Humber-Clairville	1	Etobicoke	-79.596356	43.716180	0.663292	0.257012	0.323875	1.24	1.0
21	1		Dovercourt-Wallace Emerson-Junction	93	Old City of Toronto	-79.438541	43.665677	0.729259	0.309310	0.421069	1.46	1.0
20	1		Bambury-Don Mills	42	North York	-79.349718	43.737657	0.551449	0.475031	0.304165	1.33	2.0
19	1		Niagara	82	Old City of Toronto	-79.412420	43.636681	0.620841	0.501853	0.582824	1.71	2.0
17	1		Malvern	132	Scarborough	-79.222517	43.803658	0.872004	0.242717	0.317887	1.43	2.0
16	1		Rouge	131	Scarborough	-79.186343	43.821201	0.925805	0.308934	0.310224	1.54	2.0
18	1		Cassa Loma	96	Old City of Toronto	-79.408007	43.681852	0.218389	1.053404	0.323628	1.60	2.0
15	3		Islington-City Centre West	14	Etobicoke	-79.543317	43.633463	0.875409	0.389648	0.365837	1.63	3.0
14	3		East End-Danforth	62	Old City of Toronto	-79.299359	43.684174	0.425728	0.398961	0.364101	1.19	3.0
13	3		L'Amoreaux	117	Scarborough	-79.314084	43.795716	0.875967	0.255895	0.304847	1.44	3.0
12	3		Leaside-Bennington	56	East York	-79.366072	43.703797	0.335071	0.782922	0.298916	1.42	3.0
11	3		Bridle Path-Sunnybrook-York Mills	41	North York	-79.378904	43.731013	0.184500	1.771537	0.259867	2.22	3.0
8	3		Moss Park	73	Old City of Toronto	-79.367297	43.656518	0.408305	0.420114	0.476162	1.30	4.0
10	3		Bedford Park-Nortown	39	North York	-79.420227	43.731486	0.462664	0.784588	0.282046	1.53	4.0
9	3		Lawrence Park North	105	Old City of Toronto	-79.403978	43.730060	0.290847	0.724351	0.319025	1.33	4.0
7	3		Annex	95	Old City of Toronto	-79.404001	43.671585	0.607819	0.733856	0.383939	1.73	6.0
6	3		Mount Pleasant East	99	Old City of Toronto	-79.384924	43.704852	0.334015	0.584810	0.353979	1.27	6.0
5	3		The Beaches	63	Old City of Toronto	-79.299601	43.671050	0.429431	0.618033	0.346508	1.39	8.0
4	2		Yonge-Eglinton	100	Old City of Toronto	-79.403590	43.704689	0.235294	0.598492	0.388434	1.22	11.0
3	2		Yonge-St.Clair	97	Old City of Toronto	-79.397871	43.687859	0.249451	0.737675	0.370723	1.36	15.0
2	2		Mount Pleasant West	104	Old City of Toronto	-79.393360	43.704435	0.590535	0.416122	0.449303	1.46	18.0
1	0		Church-Yonge Corridor	75	Old City of Toronto	-79.379017	43.659649	0.624027	0.394126	0.466962	1.49	30.0
0	0		Bay Street Corridor	76	Old City of Toronto	-79.385721	43.657511	0.513657	0.397679	0.394682	1.31	30.0

## 5. Discussion:

The K-means clustering seems to be effective as a first cut to isolate the areas where it may not be feasible to set up. Areas with the “ClusterL” label of 0, 2, and 3 seemed to have a lot of coffee shops in the area and have relatively low total\_scores. Hence, we can focus our attention on the 24 venues labelled Cluster 1.



From which, 1 particular venue stood out: Waterfront Communities – The Island. There are no coffee shops set up in the region, and it has a high population and % of working population.

## 6. Conclusion:

Hence, based on the data above, the business owner should set up a coffee shop at Waterfront Communities – The Island.