

# The Battle of the Neighborhoods in Toronto: Restaurants

Applied Data Science Capstone by IBM/Coursera

Sergii Tarasenko

29/07/20

---

## Table of contents

- [Introduction: Business Problem](#)
- [Data](#)
- [Methodology](#)
- [Analysis](#)
- [Results and Discussion](#)
- [Conclusion](#)

---

## Introduction: Business Problem

Toronto is the provincial capital of Ontario. It is the most populous city in Canada and the fourth most populous city in North America. Toronto has a large Italian Canadian community, with 32.2 percent of the ethnic Italians in Canada living in the Greater Toronto Area as of 2016. It is home to the fourth largest Italian population outside of Italy, behind São Paulo, Brazil, Buenos Aires, Argentina, and New York City, respectively. As of the Canada 2016 Census, there were more than 500k Italian Canadians located in the Greater Toronto Area. There are many opportunities to open a new Italian restaurant.

The aim of this project is **to help a prospective investor to find the best location for the opening of a restaurant with Italian "homemade cuisine" for locals living in Toronto, Canada.**

We are going to investigate the best location in an area that meets the following criteria:

- A neighbourhood with average to above-average total population
- Above the average population of 25-50-year-old professionals
- Average to above-average median net household incomes
- Minimal competition around

The main objective is to provide the recommendation to the prospective investor which neighbourhood of Toronto will be the best choice to start in. We will analyze data of the population and focus on the borough and explore its neighborhoods which meets the criteria above. In addition, we'll take a look at the 10 most common venues in each neighbourhood where Italian restaurants are not among the most common venue.

---

## Data

- To evaluate the neighbourhoods, we will be using the data from the 2016 Toronto Census
- Wikipedia, to matching each neighbourhood to their boroughs
- Boundaries of City of Toronto Neighbourhoods to map the neighbourhoods
- Foursquare API to obtain information on venues/competitors.

We will be using the following data sources:

- The Census of Population in Toronto [Open Data City of Toronto | NEIGHBOURHOOD PROFILES](#)

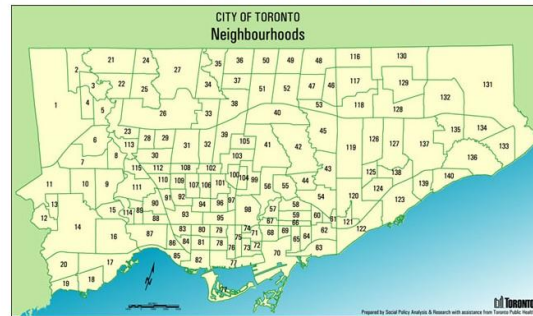
Data set: [neighbourhood-profiles-2016-csv](#)

## Neighbourhood Profiles

The map below shows the City of Toronto's 140 neighbourhoods displayed by neighbourhood number. Click the map to bring up the profile of your neighbourhood or use the lookup features below the map to find your neighbourhood profile.

Toronto is known for its diversity and culture and this is reflected in its many neighbourhoods. This section provides detailed demographic information about each neighbourhood, prepared by the City's Social Policy Analysis & Research Unit.

You can download the 2016 Neighbourhood Profiles data set from the City's [Open Data Portal](http://www.toronto.ca/open), [www.toronto.ca/open](http://www.toronto.ca/open).



The Census of Population is held across Canada every 5 years 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 Neighbourhood 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 neighbourhood. The profiles present selected highlights from the data, but these accompanying data files provide the full data set assembled for each neighbourhood.

- In these profiles, "neighbourhood" refers to the City of Toronto's 140 social planning neighbourhoods. These social planning neighbourhoods 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 neighbourhoods are consistent over time, allowing for comparison between Census years. Neighbourhood 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 neighbourhoods, as well as for the City of Toronto as a whole. The data is sourced from a number of 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.
- Each of the 140 neighbourhoods of Toronto resides within a defined borough. We will obtain this information from Wikipedia: [Wikipedia | List of city-designated neighbourhoods in Toronto](#)



[Main page](#)  
[Contents](#)  
[Current events](#)  
[Random article](#)

Article [Talk](#)



Wiki Loves Earth photo contest: Upload photos of Ukraine!

## List of city-designated neighbourhoods in Toronto

From Wikipedia, the free encyclopedia

- Boundaries of City of Toronto Neighbourhoods. [Open Data City of Toronto | NEIGHBOURHOOD](#)



Data for mapping :

- [Neighbourhoods GeoJSON File](#)
- [Neighbourhoods CSV File](#)
- To collect information on other venues/competitors, we will use Foursquare: [Foursquare | API](#)



## Methodology

In this project, we will focus on finding neighbourhoods of Toronto to open an Italian restaurant with the "homemade cuisine" targeted on locals living nearby, that meet the initial criteria from the investor. We will limit our analysis to area ~2km around the centers of every neighbourhood.

In the first step, we will collect the required data from the Census of Population in Toronto, then scrap information about the Borough from Wikipedia and merging it with Boundaries of City of Toronto Neighbourhoods which includes the location of its centers.

To establish the targeted neighbourhoods, we will examine the demographic data of the neighborhoods by segmenting the data and performing descriptive analysis using Panda. So, the second step in our analysis will be the calculation of the scores of neighbourhoods according to initial criteria, selecting the top 20 of them, and the exploration of competitors across selected neighbourhoods using API Foursquare.

In the final step, we will focus on the most interesting areas to meet initial criteria and create clusters of locations that represent density and types of competitors.

In the end, we will present a map of the top 20 neighbourhoods indicating the competition and also create clusters (using k-means clustering) of those neighbourhoods to identify those of them with minimum competition around.

# Analysis

## Getting the raw data, cleaning

Downloaded and wrangled data from the Census of Population in Toronto, then scraped information about the Borough from Wikipedia and merged it with Boundaries of City of Toronto Neighbourhoods. Prepared the dataframe with all collected data looks like

	CDN	Borough	neighbourhood	population	target_population	after_tax_income	AREA_NAME	LONGITUDE	LATITUDE
0	129	Scarborough	Agincourt North	30279	9295	633574	Agincourt North (129)	-79.266712	43.805441
1	128	Scarborough	Agincourt South-Malvern West	21988	7940	540969	Agincourt South-Malvern West (128)	-79.265612	43.788658
2	20	Etobicoke	Alderwood	11904	4510	388460	Alderwood (20)	-79.541611	43.604937
3	95	Old City of Toronto	Annex	29177	10690	2053142	Annex (95)	-79.404001	43.671585
4	42	North York	Banbury-Don Mills	26918	9420	1158599	Banbury-Don Mills (42)	-79.349718	43.737657

We used CDN number as a key column for the all merging dataframes.

## Calculating Scores according to criteria

For the scoring system, we added a coefficient of the importance of each criterion. The actual figures were obtained as a result of consultations with several restaurateurs in Kyiv, Ukraine. This “weights” of the criteria are the follows:

- A neighbourhood with average to above-average total population = 35%
- Above the average population of 25-50-year-old professionals = 45%
- Average to above-average median net household incomes = 20%

Then the score of every category was calculated as a standardized score: the value of category divided by its median and then multiplied by a coefficient of the importance.

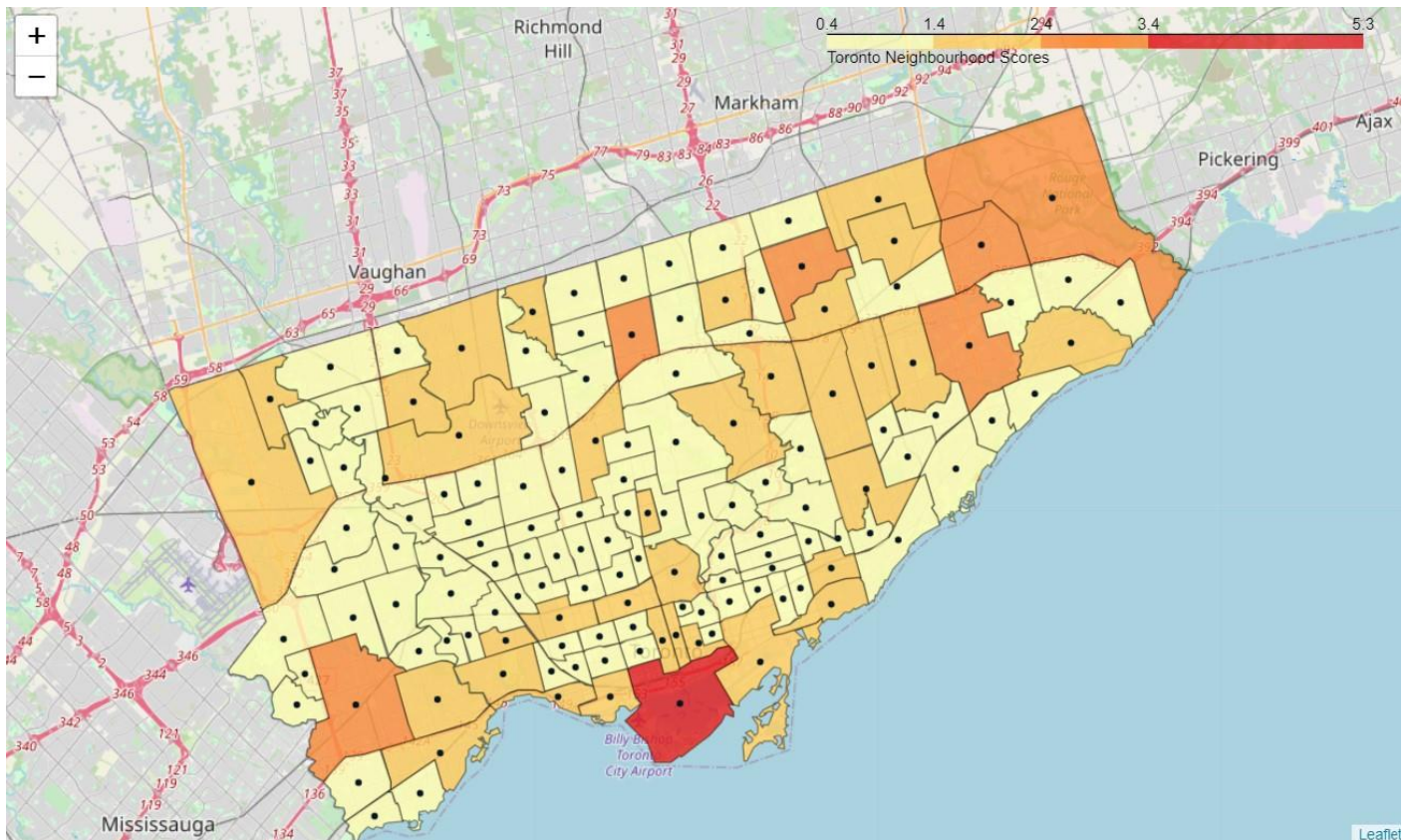
	CDN_score	p_score	tp_score	income_score	total_score
0	129	0.65	0.69	0.24	1.58
1	128	0.47	0.59	0.20	1.26
2	20	0.25	0.34	0.15	0.74
3	95	0.62	0.79	0.77	2.19
4	42	0.58	0.70	0.44	1.71

As a result, we got the following merged dataframe:

	cdn	borough	neighbourhood	population	target_population	after_tax_income	area_name	longitude	latitude	total_score
0	129	Scarborough	Agincourt North	30279	9295	633574	Agincourt North (129)	-79.266712	43.805441	1.58
1	128	Scarborough	Agincourt South-Malvern West	21988	7940	540969	Agincourt South-Malvern West (128)	-79.265612	43.788658	1.26
2	20	Etobicoke	Alderwood	11904	4510	388460	Alderwood (20)	-79.541611	43.604937	0.74
3	95	Old City of Toronto	Annex	29177	10690	2053142	Annex (95)	-79.404001	43.671585	2.19
4	42	North York	Banbury-Don Mills	26918	9420	1158599	Banbury-Don Mills (42)	-79.349718	43.737657	1.71



The choropleth folium map, visualizing neighbourhoods by scores



## Exploring venues using API Foursquare

Using the geographical coordinates of each neighborhood in Toronto, were made calls to the Foursquare API and returned the top venues within a radius of 2 km from the centers of neighbourhoods. We got 11153 venues and 343 unique venue types.

	neighbourhood	neighbourhood_latitude	neighbourhood_longitude	venue	venue_latitude	venue_longitude	venue_category
0	Agincourt North	43.805441	-79.266712	Menchie's	43.808338	-79.268288	Frozen Yogurt Shop
1	Agincourt North	43.805441	-79.266712	Fahmee Bakery & Jamaican Foods	43.810170	-79.280113	Caribbean Restaurant
2	Agincourt North	43.805441	-79.266712	Saravanaa Bhavan South Indian Restaurant	43.810117	-79.269275	Indian Restaurant
3	Agincourt North	43.805441	-79.266712	Samosa King - Embassy Restaurant	43.810152	-79.257316	Indian Restaurant
4	Agincourt North	43.805441	-79.266712	Bestco Food Market 鴻華超級市場	43.796514	-79.270790	Supermarket

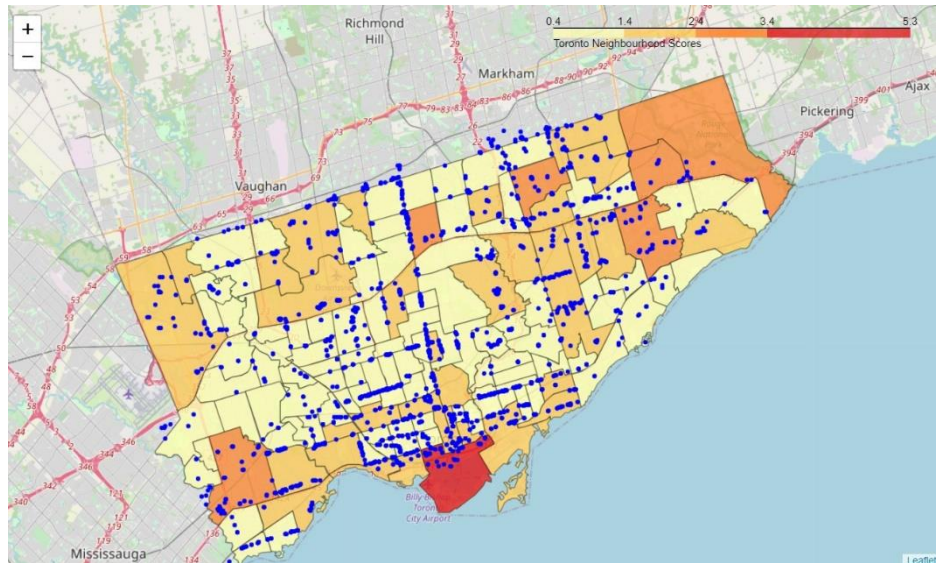
## Exploring competitors

As we focused on particular subject, we described only the venues, concerned the restaurants.

A map visualizing of all competitors

```
[132] df_comp.groupby(['venue_category']).size()
```

```
Churrascaria      1
Comfort Food Restaurant  6
Cuban Restaurant  16
Dim Sum Restaurant 18
Dumpling Restaurant  6
Eastern European Restaurant 19
Egyptian Restaurant  5
Ethiopian Restaurant 15
Falafel Restaurant  20
Fast Food Restaurant 217
Filipino Restaurant  15
French Restaurant  51
Gastropub         104
German Restaurant  2
Greek Restaurant  91
Hawaiian Restaurant  3
Hong Kong Restaurant  3
Hotpot Restaurant  3
Hungarian Restaurant  8
Indian Chinese Restaurant  2
Indian Restaurant  133
Indonesian Restaurant  9
Irish Pub          2
Italian Restaurant 309
Japanese Restaurant 167
```



As can be seen from the map, the concentration of restaurants is observed mainly on the central streets. As we are going to open a restaurant for locals with "Italian home cuisine", we will focus on an area that meets our initial criteria.

So, let's take a look at the top 20 rated neighbourhoods by scores

	cdn	borough	neighbourhood	population	target_population	after_tax_income	area_name	longitude	latitude	total_score
0	77	Old City of Toronto	Waterfront Communities-The Island	43361	29340	3291752	Waterfront Communities-The Island (77)	-79.377202	43.633880	4.35
1	51	North York	Willowdale East	45041	20050	1527084	Willowdale East (51)	-79.401484	43.770602	3.03
2	137	Scarborough	Woburn	53350	17845	1111814	Woburn (137)	-79.228586	43.766740	2.89
3	14	Etobicoke	Islington-City Centre West	38084	16655	1536690	Islington-City Centre West (14)	-79.543317	43.633463	2.63
4	131	Scarborough	Rouge	45912	15340	1209928	Rouge (131)	-79.186343	43.821201	2.58

neighbourhood	index
Annex	26
Banbury-Don Mills	36
Church-Yonge Corridor	28
Dovercourt-Wallace Emerson-Junction	31
Downsview-Roding-CFB	17
Islington-City Centre West	9
L'Amoreaux	25
Malvern	19
Mimico (includes Humber Bay Shores)	37
Mount Olive-Silverstone-Jamestown	15
Mount Pleasant West	34
Niagara	32
Parkwoods-Donalda	26
Rosedale-Moore Park	31
Rouge	2
South Riverdale	19
Waterfront Communities-The Island	18
West Humber-Clairville	10
Willowdale East	13
Woburn	19

Exploring competitors in the top selected 20 neighbourhoods we answered the following question: How many competitors located in each of the top selected neighbourhood?

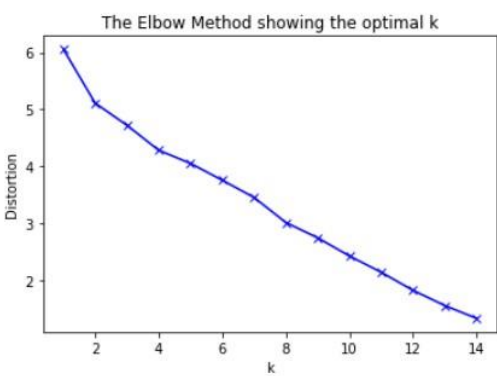
"Rouge" is a neighbourhood in the northeastern area of Toronto, Ontario, within the former city of Scarborough. It is Toronto's largest neighbourhood by surface area; however, unlike other neighbourhoods, most of its area remains undeveloped, as the neighbourhood is adjacent to Rouge National Urban Park." (c) Wikipedia

So, we decided to drop it as it's not interesting for us.

## Clustering neighbourhoods and examine clusters

With the prepared data for clustering, we run k-means to cluster neighborhoods into three (3) clusters. The cluster number was established by Elbow Method.

It is not so obvious, but we set k=3.



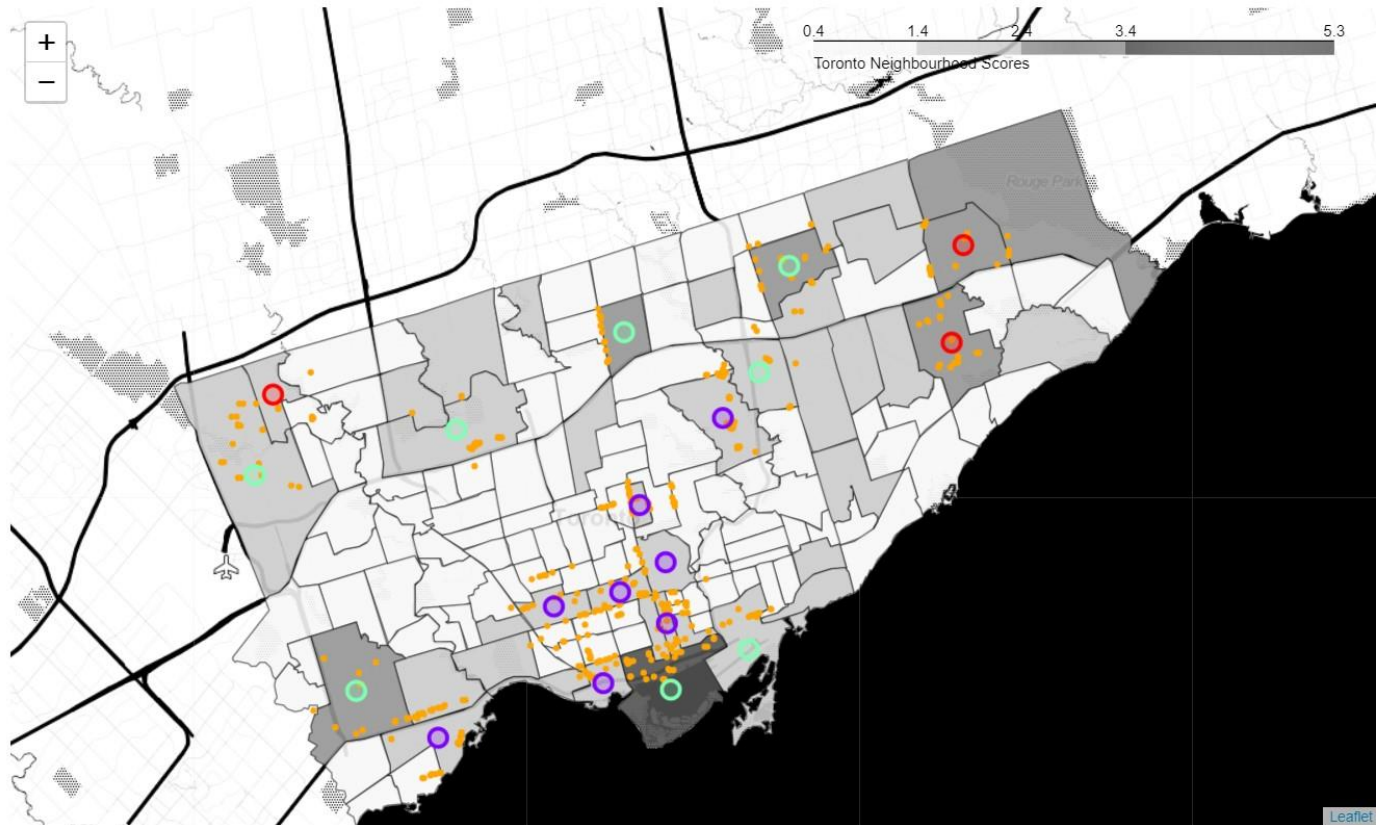
the

After our clusters established, we got the final dataframe of top selected neighbourhoods distributing by clusters, added the percentage of the target population. From now we are completely ready for examining to obtain results.

	neighbourhood	cluster_labels	cdn	borough	population	target_population	after_tax_income	area_name	longitude	latitude	total_score	competitors	p_target_population
0	Annex	1	95	Old City of Toronto	29177	10690	2053142	Annex (95)	-79.404001	43.671585	2.19	26	36.64%
1	Banbury-Don Mills	1	42	North York	26918	9420	1158599	Banbury-Don Mills (42)	-79.349718	43.737657	1.71	36	35.00%
2	Church-Yonge Corridor	1	75	Old City of Toronto	28349	13240	1221015	Church-Yonge Corridor (75)	-79.379017	43.659649	2.05	28	46.70%
3	Dovercourt-Wallace Emerson-Junction	1	93	Old City of Toronto	34631	15360	1037114	Dovercourt-Wallace Emerson-Junction (93)	-79.438541	43.665677	2.27	31	44.35%
4	Downsview-Roding-CFB	2	26	North York	34659	12825	824881	Downsview-Roding-CFB (26)	-79.490497	43.733292	2.01	17	37.00%
5	Islington-City Centre West	2	14	Etobicoke	38084	16655	1536690	Islington-City Centre West (14)	-79.543317	43.633463	2.63	9	43.73%
6	L'Amoreaux	2	117	Scarborough	44919	14135	989212	L'Amoreaux (117)	-79.314084	43.795716	2.38	25	31.47%
7	Malvern	0	132	Scarborough	45086	14655	888593	Malvern (132)	-79.222517	43.803658	2.39	19	32.50%
8	Mimico (includes Humber Bay Shores)	1	17	Etobicoke	26541	14145	1275825	Mimico (includes Humber Bay Shores) (17)	-79.500137	43.615924	2.10	37	53.29%
9	Mount Olive-Silverstone-Jamestown	0	2	Etobicoke	32788	11225	575316	Mount Olive-Silverstone-Jamestown (2)	-79.587259	43.746868	1.75	15	34.24%
10	Mount Pleasant West	1	104	Old City of Toronto	28593	13100	1163508	Mount Pleasant West (104)	-79.393360	43.704435	2.02	34	45.82%
11	Niagara	1	82	Old City of Toronto	21274	16460	1531187	Niagara (82)	-79.412420	43.636681	2.25	32	77.37%
12	Parkwoods-Donalda	2	45	North York	34617	12440	956565	Parkwoods-Donalda (45)	-79.330180	43.755033	2.03	26	35.94%
13	Rosedale-Moore Park	1	98	Old City of Toronto	20631	6630	2332496	Rosedale-Moore Park (98)	-79.379669	43.682820	1.81	31	32.14%
14	South Riverdale	2	70	Old City of Toronto	25642	12790	971149	South Riverdale (70)	-79.335651	43.649292	1.86	19	49.88%
15	Waterfront Communities-The Island	2	77	Old City of Toronto	43361	29340	3291752	Waterfront Communities-The Island (77)	-79.377202	43.633880	4.35	18	67.66%
16	West Humber-Clairville	2	1	Etobicoke	34100	11025	737298	West Humber-Clairville (1)	-79.596356	43.716180	1.83	10	32.33%
17	Willowdale East	2	51	North York	45041	20050	1527084	Willowdale East (51)	-79.401484	43.770602	3.03	13	44.51%
18	Woburn	0	137	Scarborough	53350	17845	1111814	Woburn (137)	-79.228586	43.766740	2.89	19	33.45%



A map, visualizing the resulting clusters with all competitors in selected neighbourhoods



This concludes our analysis. We have found the top 19 neighbourhoods rated by scores, calculated according to the initial criteria, and examined the competitive environment.

## Results and Discussion

The first look at the map shows us that there are significant numbers of restaurants different types and cuisines in the southern part of Toronto, close to central streets, and to the shore of Lake Ontario.

Based on the found and presented results, the following recommendations can be made:

- the main recommendation for the prospective investor would be to focus on neighbourhoods from cluster 2, as this area has good average scores and low density of competitors.
- the main neighborhood recommendation would be for **"Islington-City Centre West"**. With a score of 2.63, this neighbourhood is a commercial and residential area in Etobicoke, Toronto, Ontario, Canada. One of four central business districts outside Downtown Toronto with good perspectives to catch numerous of white-collar in lunchtime and a good number of locals for supper.
- the second and the third places go to **"Woburn"** 2.89 and **"Malvern"** 2.39 from cluster 0, both are with an average density of competition and with no Italian cuisine in the top common venues across the entire neighbourhood.

## Conclusion

The final decision on optimal restaurant location will be made by the future investor based on the recommendations given in this analysis and also on specific characteristics of neighborhoods in every recommended area, taking into consideration additional factors like real estate availability, prices, etc.

Also, before starting spending money on renting and renovating the facility, don't forget about the outdated data used in this analysis, such as information about the venues from API Foursquare and the 2016 Census Data. However, the methodology and main calculations will allow you to make the right decision when applying the updated initial data.