Capstone Project - The Battle of Neighborhoods

Analysis of Toronto neighborhoods for an optimal new restaurant location

Yerzat Marat

September 5th, 2019

Table of contents

1. Introduction **3**
1.1 Business problem 3
1.2 Stakeholders 3
2. Data **4**
2.1 Data sources 4
2.2 Data description 4
3. Methodology **6**
3.1 Exploratory data analysis 6
3.2 Machine learning methods 9
4. Results 10
5. Discussion **14**
6. Conclusion **15**
7. References **16**

1. Introduction

With the proliferation of data and services in the modern world, it is sometimes hard to analyze all the information and decide on an optimal course of action. However, when used correctly, massive amounts of information can be become a powerful tool in the hands of a knowledgeable person. In this report, one will see how the use of geolocation data along with other information can help in deciding on an optimal location to open a new restaurant in Toronto.

Toronto is the provincial capital of Ontario and the most populous city in Canada, with a population of 2,731,571 in 2016. Toronto is the fastest growing city in North America, and is the anchor of an urban agglomeration. Toronto is an international center of business, finance, arts, and culture, and is recognized as one of the most multicultural and cosmopolitan cities in the world.[1]

One of the biggest factors contributing to growth of Toronto is immigration and the city is quickly establishing itself as one of the most diverse in its region. Therefore, there is ample opportunity for a quick entry into the restaurant business as the population is growing more diverse and demand for novelty increases.

1.1. Business problem

The scenario we will analyze further is the following – someone is looking to open a restaurant in Toronto and he or she cannot decide on where to open it. Through our analysis of available data, we will provide recommendations on the optimal location to open a restaurant. We will use geolocation data and open datasets provided by city of Toronto to analyze competition, potential demand and other relevant factors in different neighborhoods of the city.

1.2. Stakeholders

This report will be interesting to several groups of stakeholders:

1. Someone who is looking to open a new restaurant. This might be a newcomer to the restaurant business, someone expanding their current business or big chains looking to establish a presence.

2. This report might be interesting to banks or financial institutions when making decisions on assessing loan applications for restaurants. The likelihood of loan defaults decreases if a restaurant has more chances of success due to a convenient location. Banks might even encourage loans in a certain area to maximize chances of success.

[1] https://en.wikipedia.org/wiki/Toronto

2. Data

Opening a new business such as a restaurant requires an analysis of several factors. If done meticulously, number of factors could be well above a hundred. However, for the purpose of this report we will focus on the following – potential customers, competition and safety in several neighborhoods in Toronto. Description of factors and data sources is below.

2.1. Data sources

There are several data sources in the project:

1. Foursquare location data from a developer's account
https://foursquare.com/
2. Safety data extracted from "Toronto Police Service Public Safety Data Portal"
http://data.torontopolice.on.ca/pages/open-data
3. Population data across neighborhoods was extracted from the 2016 Census, Toronto
https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/index-eng.cfm
4. Demographic data from the Toronto Open Data Portal
https://www.toronto.ca/city-government/data-research-maps/open-data/
5. Neighborhood data and postal codes for Toronto
https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
6. Longitude and latitude coordinates of neighborhoods
CSV file "Geospatial coordinates.csv"

2.2. Data description

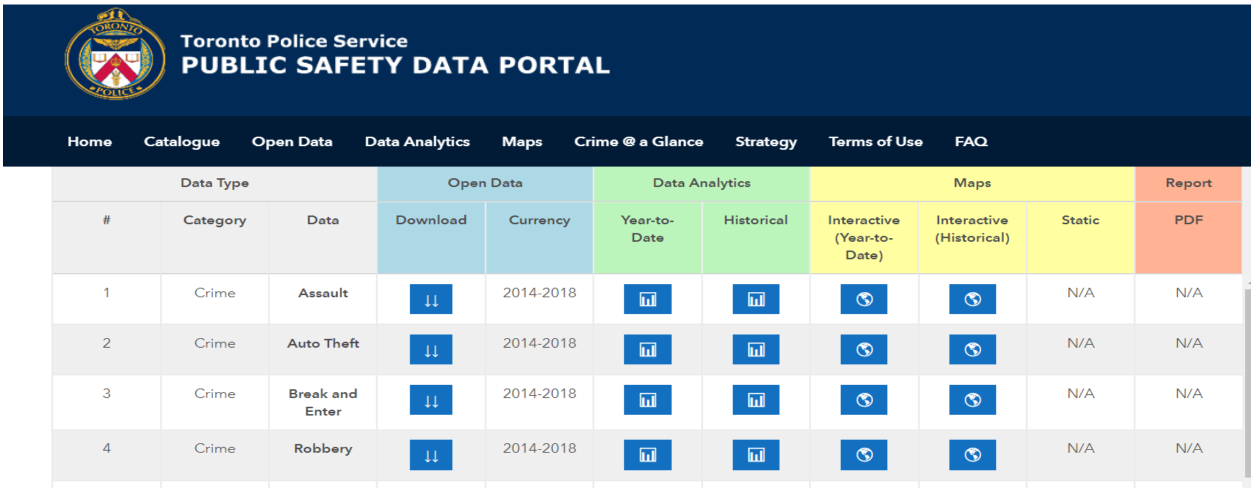
Data collected from the mentioned sources above will allow us to analyze neighborhoods and figure out which are more suitable to opening new restaurants. Using foursquare location data, we will primarily analyze competition, for example a number of restaurants operating in the neighborhood (if data allows, maybe even the types of restaurants).

[34]:

	name	categories	address	cc	city	country	crossStreet	distance	formattedAddress	labeledLatLngs	lat	lng	neighborhood	postalCode	state
0	Korin	Furniture / Home Store	57 Warren St	US	New York	United States	Church St	73	[57 Warren St (Church St), New York, NY 10007, United States]	[{'label': 'display', 'lat': 40.71482437714839, 'lng': -74.00940425461492}]	40.714824	-74.009404	Tribeca	10007	NY
1	Juice Press	Vegetarian / Vegan Restaurant	83 Murray St	US	New York	United States	btwn Greenwich St & W Broadway	202	[83 Murray St (btwn Greenwich St & W Broadway), New York, NY 10007, United States]	[{'label': 'display', 'lat': 40.71478769908051, 'lng': -74.0111317502157}]	40.714788	-74.011132	NaN	10007	NY
2	Chambers Street Wines	Wine Shop	148 Chambers St	US	New York	United States	btwn West Broadway & Hudson St	88	[148 Chambers St (btwn West Broadway & Hudson St), New York, NY 10007, United States]	[{'label': 'display', 'lat': 40.715773063928374, 'lng': -74.00971823312332}]	40.715773	-74.009718	NaN	10007	NY
3	Takahachi Bakery	Bakery	25 Murray St	US	New York	United States	at Church St	187	[25 Murray St (at Church St), New York, NY 10007, United States]	[{'label': 'display', 'lat': 40.713652845301894, 'lng': -74.0088038953017}]	40.713653	-74.008804	NaN	10007	NY
4	Takahachi	Sushi Restaurant	145 Duane St	US	New York	United States	btwn W Broadway & Church St	146	[145 Duane St (btwn W Broadway & Church St), New York, NY 10013, United States]	[{'label': 'display', 'lat': 40.71652647412374, 'lng': -74.00810108466207}]	40.716526	-74.008101	NaN	10013	NY

Image 1. An example of venue information obtained by a request to a Foursquare API

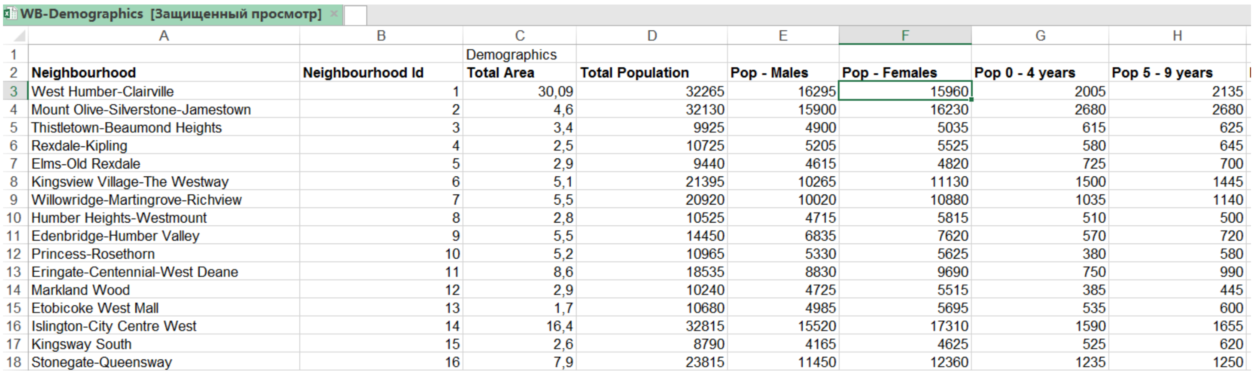
The safety data from Toronto Police will give us data on the crime rates of certain areas as we don’t want to open a restaurant in an area with high levels of crime. It is available in CSV format and will be transformed into Pandas data frames for cleaning.



Data Type			Open Data		Data Analytics		Maps			Report
#	Category	Data	Download	Currency	Year-to-Date	Historical	Interactive (Year-to-Date)	Interactive (Historical)	Static	PDF
1	Crime	Assault		2014-2018					N/A	N/A
2	Crime	Auto Theft		2014-2018					N/A	N/A
3	Crime	Break and Enter		2014-2018					N/A	N/A
4	Crime	Robbery		2014-2018					N/A	N/A

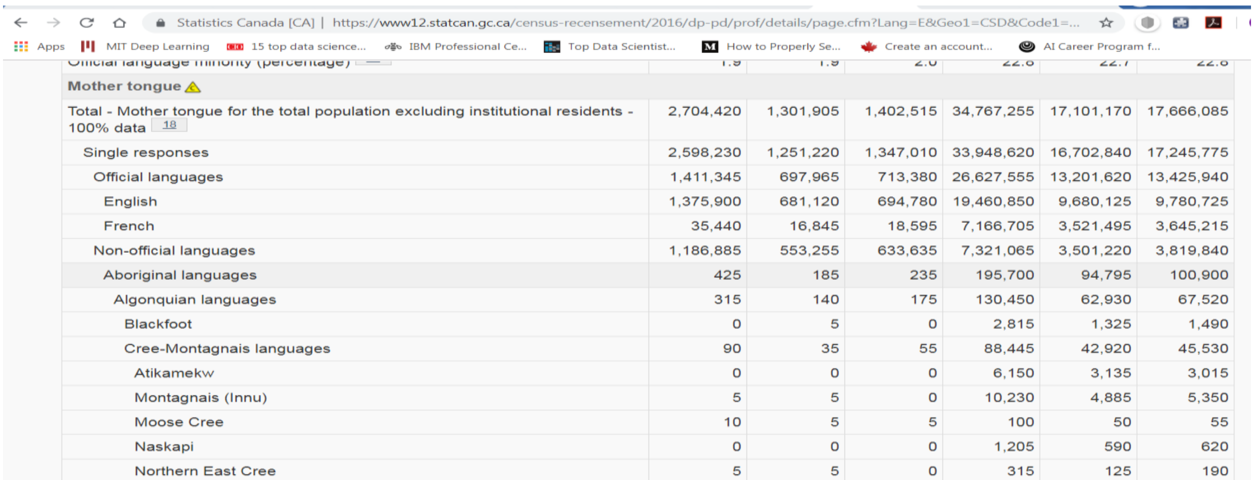
Image 2. Data on Toronto crime from the Police Service.

The demographic and population data will let us analyze potential demand and maybe give us a clue in the type of restaurant that should be opened (for example, the predominance of a certain language in the area might mean that type of cuisine will be popular). It is available in CSV format and will be read into a Pandas data frame for cleaning and analysis.



	A	B	C	D	E	F	G	H
1			Demographics					
2	Neighbourhood	Neighbourhood Id	Total Area	Total Population	Pop - Males	Pop - Females	Pop 0 - 4 years	Pop 5 - 9 years
3	West Humber-Clairville	1	30,09	32265	16295	15960	2005	2135
4	Mount Olive-Silverstone-Jamestown	2	4,6	32130	15900	16230	2680	2680
5	Thistletown-Beaumond Heights	3	3,4	9925	4900	5035	615	625
6	Rexdale-Kipling	4	2,5	10725	5205	5525	580	645
7	Elms-Old Rexdale	5	2,9	9440	4615	4820	725	700
8	Kingsview Village-The Westway	6	5,1	21395	10265	11130	1500	1445
9	Willowridge-Martingrove-Richview	7	5,5	20920	10020	10880	1035	1140
10	Humber Heights-Westmount	8	2,8	10525	4715	5815	510	500
11	Edenbridge-Humber Valley	9	5,5	14450	6835	7620	570	720
12	Princess-Rosethorn	10	5,2	10965	5330	5625	380	580
13	Eringate-Centennial-West Deane	11	8,6	18535	8830	9690	750	990
14	Markland Wood	12	2,9	10240	4725	5515	385	445
15	Etobicoke West Mall	13	1,7	10680	4985	5695	535	600
16	Islington-City Centre West	14	16,4	32815	15520	17310	1590	1655
17	Kingsway South	15	2,6	8790	4165	4625	525	620
18	Stonegate-Queensway	16	7,9	23815	11450	12360	1235	1250

Image 3. A view of the demographic data for Toronto neighborhoods in xls format.



Mother tongue						
Total - Mother tongue for the total population excluding institutional residents - 100% data 18	2,704,420	1,301,905	1,402,515	34,767,255	17,101,170	17,666,085
Single responses	2,598,230	1,251,220	1,347,010	33,948,620	16,702,840	17,245,775
Official languages	1,411,345	697,965	713,380	26,627,555	13,201,620	13,425,940
English	1,375,900	681,120	694,780	19,460,850	9,680,125	9,780,725
French	35,440	16,845	18,595	7,166,705	3,521,495	3,645,215
Non-official languages	1,186,885	553,255	633,635	7,321,065	3,501,220	3,819,840
Aboriginal languages	425	185	235	195,700	94,795	100,900
Algonquian languages	315	140	175	130,450	62,930	67,520
Blackfoot	0	5	0	2,815	1,325	1,490
Cree-Montagnais languages	90	35	55	88,445	42,920	45,530
Atikamekw	0	0	0	6,150	3,135	3,015
Montagnais (Innu)	5	5	0	10,230	4,885	5,350
Moose Cree	10	5	5	100	50	55
Naskapi	0	0	0	1,205	590	620
Northern East Cree	5	5	0	315	125	190

Image 4. Data on Toronto languages

The neighborhood data and postal codes with longitude and latitude are needed for tools to make the visualizations and make the analysis easier to understand.

	Postcode	Borough	Neighborhood	Latitude	Longitude
0	M1B	Scarborough	Rouge,Malvern	43.806686	-79.194353
1	M1C	Scarborough	Highland Creek,Rouge Hill,Port Union	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

Image 5. Transformed data from postal code and neighborhood data with geographical coordinates

3. Methodology

3.1. Exploratory Data Analysis

First, I collected and transformed the data into Pandas dataframes. I scraped the Wikipedia page for a list of postal codes with longitude and latitude, after that I created a map of Toronto with the neighborhood postcodes superimposed on top of it using Folium.

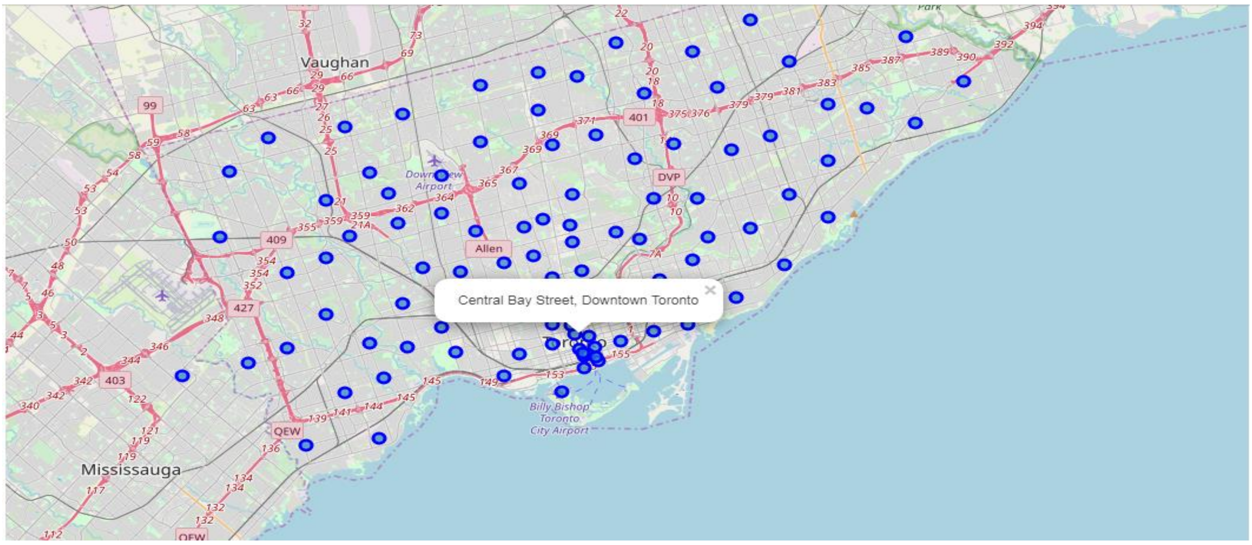


Image 6. Map of Toronto with neighborhoods superimposed as blue dots

Then I used the Foursquare API to explore the neighborhoods where I passed the coordinates of each neighborhood to the Foursquare API, which returned a list of venues within a given radius of 500 meters and up to 100 venues max. The code requested and returned venues for all 140 neighborhoods in Toronto. The size of the resulting dataframe was 2249 rows by 7 columns, meaning we have 2249 venues in our dataset with neighborhood name, latitude, longitude, venuename and geocoordinates and its category. The dataframe snapshot is below:

[16]:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Rouge,Malvern	43.806686	-79.194353	Wendy's	43.807448	-79.199056	Fast Food Restaurant
1	Highland Creek,Rouge Hill,Port Union	43.784535	-79.160497	Chris Effects Painting	43.784343	-79.163742	Construction & Landscaping
2	Highland Creek,Rouge Hill,Port Union	43.784535	-79.160497	Royal Canadian Legion	43.782533	-79.163085	Bar
3	Guildwood,Morningside,West Hill	43.763573	-79.188711	Swiss Chalet Rotisserie & Grill	43.767697	-79.189914	Pizza Place
4	Guildwood,Morningside,West Hill	43.763573	-79.188711	G & G Electronics	43.765309	-79.191537	Electronics Store

Image 7. Dataframe of all venues in Toronto.

Then I calculated the number of venues returned by each neighborhood and unique categories in the dataset. Some neighborhoods maxed out at 100 venues and some only returned 3, there were 277 unique categories of venues.

	Neighborhood Latitude
Neighborhood	
Adelaide,King,Richmond	100
Agincourt	4
Agincourt North,L'Amoreaux East,Milliken,Steeles East	4
Albion Gardens,Beaumond Heights,Humbergate,Jamestown,Mount Olive,Silverstone,South Steeles,Thistletown	11
Alderwood,Long Branch	9
...	...
Willowdale West	4
Woburn	3
Woodbine Gardens,Parkview Hill	12
Woodbine Heights	8
York Mills West	3

Image 8. View of the number of venues per neighborhood

After that, I analyzed each neighborhood. To do that, I performed one hot encoding, grouped neighborhoods, and calculated the frequency of occurrence of venues for each category in neighborhoods.

	Neighborhood	Yoga Studio	Accessories Store	Afghan Restaurant	Airport	Airport Food Court	Airport Gate	Airport Lounge	Airport Service	Airport Terminal	...	Trail	Train Station	Vegetarian / Vegan Restaurant
0	Adelaide,King,Richmond	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.02
1	Agincourt	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.00
2	Agincourt North,L'Amoreaux East,Milliken,Steel...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.00
3	Albion Gardens,Beaumond Heights,Humbergate,Jam...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.00
4	Alderwood,Long Branch	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.00
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
96	Willowdale West	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.00
97	Woburn	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.00
98	Woodbine Gardens,Parkview Hill	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.00
99	Woodbine Heights	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.00
100	York Mills West	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.00

101 rows × 277 columns

Image 9. Results of the frequency of occurrence of venues per neighborhood

Then, I printed each neighborhood and ranked the venues by frequency of occurrence, listing the top 5 per each. This is what it looks like for Adelaide, King and Richmond neighborhood:

```
----Adelaide,King,Richmond----
                venue  freq
0         Coffee Shop  0.08
1                Café  0.05
2                 Bar  0.04
3          Steakhouse  0.04
4  American Restaurant  0.03
```

Image 10. Results of the frequency of occurrence of venues for Adelaide, King and Richmond

Then I put the above in a pandas dataframe and displayed the top 10 venues for each neighborhood. This is the resulting dataframe:

	Neighborhood	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	Adelaide,King,Richmond	Coffee Shop	Café	Bar	Steakhouse	Thai Restaurant	Cosmetics Shop	Hotel	Burger Joint	American Restaurant	Restaurant
1	Agincourt	Lounge	Skating Rink	Breakfast Spot	Sandwich Place	Drugstore	Discount Store	Dog Run	Doner Restaurant	Donut Shop	Eastern European Restaurant
2	Agincourt North,L'Amoreaux East,Milliken,Steel...	Park	Playground	Gym	Coffee Shop	Women's Store	Drugstore	Diner	Discount Store	Dog Run	Doner Restaurant
3	Albion Gardens,Beaumond Heights,Humbergate,Jam...	Grocery Store	Fast Food Restaurant	Coffee Shop	Japanese Restaurant	Discount Store	Sandwich Place	Beer Store	Fried Chicken Joint	Pizza Place	Pharmacy
4	Alderwood,Long Branch	Pizza Place	Coffee Shop	Gym	Skating Rink	Pharmacy	Sandwich Place	Pub	Pool	Dog Run	Dessert Shop

Image 11. Top 10 venue categories for each neighborhood

Next, I analyzed the Census data. Below is a processed slice as a dataframe from the 2016 Census, sorting the popularity of different languages. It has several categories, irrelevant for our analysis, such as “Official languages”, “Single responses” etc.

	Characteristic	City of Toronto
139	Mother tongue for the total population excludi...	2,704,415
140	Single responses	2,598,230
141	Official languages	1,411,345
142	English	1,375,905
144	Non-official languages	1,186,885
212	Non-Aboriginal languages	1,186,465
261	Indo-European languages	589,415
356	Sino-Tibetan languages	250,960
357	Chinese languages	245,285
319	Italic (Romance) languages	207,440
300	Indo-Iranian languages	194,765
301	Indo-Aryan languages	138,625
264	Balto-Slavic languages	120,445
268	Slavic languages	116,955
358	Cantonese	114,670
360	Mandarin	111,405
387	Multiple responses	106,190
237	Austronesian languages	99,755

Image 12. Top 10 venue categories for each neighborhood

We see from the table below that the total population surveyed was 2.7 million people. English is listed as a mother tongue for 1.375 million people or roughly half the population, which is another proof point for the diversity of Toronto.

№	Characteristic	City of Toronto
1	Mother tongue for the total population	2,704,415
2	Single responses	2,598,230
3	Official languages	1,411,345
4	English	1,375,905
5	Non-official languages	1,186,885
6	Non-Aboriginal languages	1,186,465
7	Indo-European languages	589,415
8	Sino-Tibetan languages	250,960
9	Chinese languages	245,285
10	Italic (Romance) languages	207,440

Table 1. Processed data on the popularity of languages in Toronto from the 2016 Census.

3.2. Machine Learning Methods

In this project, I used a classification machine learning method for clustering neighborhoods, namely k-means. After tweaking the algorithm for different k values I have settled on k=6, meaning there would be 7 clusters as the algorithm starts from 0. Running values lower gave less venues in each cluster and running values higher than four did not improve the results very much. Then I put that into a new dataframe with cluster values in it.

	Postcode	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	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	M1B	Scarborough	Rouge,Malvern	43.806686	-79.194353	3.0	Fast Food Restaurant	Dumpling Restaurant	Diner	Discount Store	Dog Run	Doner Restaurant	Donut Shop	Drugstore	Eastern European Restaurant	Festival
1	M1C	Scarborough	Highland Creek,Rouge Hill,Port Union	43.784535	-79.160497	4.0	Construction & Landscaping	Bar	Women's Store	Eastern European Restaurant	Dog Run	Doner Restaurant	Donut Shop	Drugstore	Dumpling Restaurant	Electronics Store
2	M1E	Scarborough	Guildwood,Morningside,West Hill	43.763573	-79.188711	4.0	Medical Center	Electronics Store	Rental Car Location	Mexican Restaurant	Breakfast Spot	Intersection	Pizza Place	Empanada Restaurant	Eastern European Restaurant	Dumpling Restaurant
3	M1G	Scarborough	Woburn	43.770992	-79.216917	4.0	Coffee Shop	Korean Restaurant	Eastern European Restaurant	Discount Store	Dog Run	Doner Restaurant	Donut Shop	Drugstore	Dumpling Restaurant	Women's Store
4	M1H	Scarborough	Cedarbrae	43.773136	-79.239476	4.0	Hakka Restaurant	Athletics & Sports	Fried Chicken Joint	Bakery	Caribbean Restaurant	Thai Restaurant	Bank	Discount Store	Doner Restaurant	Donut Shop

Image 13. Dataframe with cluster numbers in it.

4. Results

The resulting map with clusters superimposed on top of it looked like this:

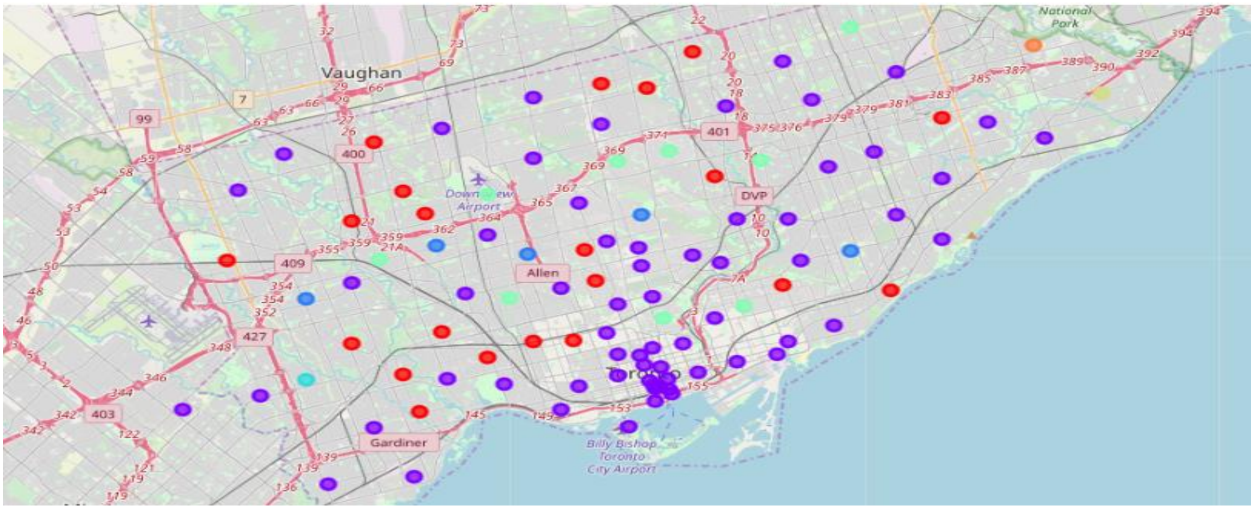


Image 14. Map of Toronto with clusters superimposed

- Cluster 0
- Cluster 1
- Cluster 2
- Cluster 3
- Cluster 4
- Cluster 5
- Cluster 6

Next, I examined each cluster and determined the discriminating venue categories that distinguish each cluster. Based on the defining categories, I then assigned a name to each cluster.

Cluster 0: Recreational

	Neighborhood	Cluster Labels	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
4	Cedarbrae	0	Hakka Restaurant	Bank	Athletics & Sports	Bakery	Thai Restaurant	Caribbean Restaurant	Fried Chicken Joint	Eastern European Restaurant	Dumpling Restaurant	Drugstore
9	Birch Cliff,Cliffside West	0	Café	General Entertainment	Skating Rink	College Stadium	Concert Hall	Dim Sum Restaurant	Event Space	Ethiopian Restaurant	Empanada Restaurant	Electronics Store
16	Upper Rouge	0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
17	Hillcrest Village	0	Mediterranean Restaurant	Fast Food Restaurant	Golf Course	Dog Run	Pool	Donut Shop	Dim Sum Restaurant	Diner	Discount Store	Doner Restaurant
19	Bayview Village	0	Café	Japanese Restaurant	Bank	Chinese Restaurant	Dim Sum Restaurant	Discount Store	Dog Run	Doner Restaurant	Donut Shop	Drugstore
21	Newtonbrook,Willowdale	0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
26	Don Mills North	0	Baseball Field	Gym / Fitness Center	Caribbean Restaurant	Japanese Restaurant	Café	Women's Store	Diner	Discount Store	Dog Run	Doner Restaurant
31	Downsview West	0	Shopping Mall	Park	Grocery Store	Bank	Hotel	Donut Shop	Diner	Discount Store	Dog Run	Doner Restaurant
32	Downsview Central	0	Baseball Field	Food Truck	Women's Store	Discount Store	Dog Run	Doner Restaurant	Donut Shop	Drugstore	Dumpling Restaurant	Dim Sum Restaurant
33	Downsview Northwest	0	Liquor Store	Grocery Store	Discount Store	Athletics & Sports	Comic Shop	Dim Sum Restaurant	Falafel Restaurant	Event Space	Ethiopian Restaurant	Empanada Restaurant
36	Woodbine Heights	0	Park	Skating Rink	Curling Ice	Cosmetics Shop	Beer Store	Pharmacy	Ethiopian Restaurant	Empanada Restaurant	Electronics Store	Eastern European Restaurant
63	Roselawn	0	Garden	Pool	Women's Store	Donut Shop	Dim Sum Restaurant	Diner	Discount Store	Dog Run	Doner Restaurant	Drugstore
64	Forest Hill North,Forest Hill West	0	Jewelry Store	Trail	Park	Sushi Restaurant	Dim Sum Restaurant	Diner	Discount Store	Dog Run	Doner Restaurant	Donut Shop

Image 15. List of neighborhoods in Cluster 0

There seemed to be a mix of entertainment venues (parks, malls, restaurants, baseball field) so I decided to name this cluster "Recreational".

Cluster 1: Entertainment

	Neighborhood	Cluster Labels	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
2	Guildwood,Morningside,West Hill	1	Medical Center	Electronics Store	Pizza Place	Intersection	Breakfast Spot	Rental Car Location	Mexican Restaurant	Doner Restaurant	Diner	Discount Store
3	Woburn	1	Coffee Shop	Korean Restaurant	Women's Store	Dumpling Restaurant	Diner	Discount Store	Dog Run	Doner Restaurant	Donut Shop	Drugstore
5	Scarborough Village	1	Playground	Construction & Landscaping	Women's Store	Drugstore	Dim Sum Restaurant	Diner	Discount Store	Dog Run	Doner Restaurant	Donut Shop
6	East Birchmount Park,Ionview,Kennedy Park	1	Department Store	Discount Store	Coffee Shop	Hobby Shop	Convenience Store	Dumpling Restaurant	Diner	Dog Run	Doner Restaurant	Donut Shop
8	Cliffcrest,Cliffside,Scarborough Village West	1	Motel	American Restaurant	Department Store	Dim Sum Restaurant	Diner	Discount Store	Dog Run	Doner Restaurant	Donut Shop	Drugstore
...	...	...	...	...	...	...	...	...	...	...	...	...
92	Kingsway Park South West,Mimico NW,The Queensw...	1	Hardware Store	Tanning Salon	Wings Joint	Fast Food Restaurant	Discount Store	Convenience Store	Gym	Burrito Place	Burger Joint	Sandwich Place
95	Bloordale Gardens,Eringate,Markland Wood,Old B...	1	Liquor Store	Beer Store	Convenience Store	Coffee Shop	Café	Pizza Place	Pharmacy	Ethiopian Restaurant	Event Space	Empanada Restaurant
96	Humber Summit	1	Pizza Place	Empanada Restaurant	Women's Store	Donut Shop	Dim Sum Restaurant	Diner	Discount Store	Dog Run	Doner Restaurant	Dumpling Restaurant
99	Westmount	1	Intersection	Sandwich Place	Pizza Place	Coffee Shop	Middle Eastern Restaurant	Discount Store	Chinese Restaurant	Dog Run	Dim Sum Restaurant	Diner
101	Albion Gardens,Beaumond Heights,Humbergate,Jam...	1	Grocery Store	Fried Chicken Joint	Pharmacy	Pizza Place	Fast Food Restaurant	Coffee Shop	Beer Store	Sandwich Place	Women's Store	Dog Run

Image 16. List of neighborhoods in Cluster 1

This cluster has 64 rows with mostly places to eat and entertainment, therefore it falls into the entertainment category.

Cluster 2: Residential

	Neighborhood	Cluster Labels	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
7	Clairlea,Golden Mile,Oakridge	2	Bakery	Park	Intersection	Fast Food Restaurant	Metro Station	Bus Line	Bus Station	Soccer Field	Costume Shop	Construction & Landscaping
44	Lawrence Park	2	Park	Bus Line	Swim School	Women's Store	Donut Shop	Diner	Discount Store	Dog Run	Doner Restaurant	Drugstore
72	Glencairn	2	Pub	Filipino Restaurant	Bakery	Japanese Restaurant	Women's Store	Discount Store	Dog Run	Doner Restaurant	Donut Shop	Drugstore
79	Downsview,North Park,Upwood Park	2	Basketball Court	Park	Construction & Landscaping	Bakery	Eastern European Restaurant	Dog Run	Doner Restaurant	Donut Shop	Drugstore	Dumpling Restaurant
100	Kingsview Village,Martin Grove Gardens,Richvie...	2	Mobile Phone Shop	Park	Pizza Place	Bus Line	Donut Shop	Dim Sum Restaurant	Diner	Discount Store	Dog Run	Doner Restaurant

Image 17. List of neighborhoods in Cluster 2

This cluster lists parks, bakeries and transit stops as most ranked venues, implying this is a residential area.

Cluster 3: Business area

	Neighborhood	Cluster Labels	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
94	Cloverdale,Islington,Martin Grove,Princess Gar...	3	Bank	Women's Store	Drugstore	Diner	Discount Store	Dog Run	Doner Restaurant	Donut Shop	Dumpling Restaurant	Fast Food Restaurant

Image 18. List of neighborhoods in Cluster 3

Although this cluster has one row, it comprises several neighborhoods where the top ranked venue is banks and stores, implying this is a business heavy area.

Cluster 4: Parks

	Neighborhood	Cluster Labels	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
14	Agincourt North,L'Amoreaux East,Milliken,Steel...	4	Playground	Park	Women's Store	Donut Shop	Dessert Shop	Dim Sum Restaurant	Diner	Discount Store	Dog Run	Doner Restaurant
20	Silver Hills,York Mills	4	Park	Women's Store	Dumpling Restaurant	Diner	Discount Store	Dog Run	Doner Restaurant	Donut Shop	Drugstore	Eastern European Restaurant
23	York Mills West	4	Park	Bank	Convenience Store	Women's Store	Dumpling Restaurant	Discount Store	Dog Run	Doner Restaurant	Donut Shop	Drugstore
25	Parkwoods	4	Park	Food & Drink Shop	Women's Store	Drugstore	Diner	Discount Store	Dog Run	Doner Restaurant	Donut Shop	Dumpling Restaurant
30	CFB Toronto,Downsview East	4	Park	Airport	Women's Store	Dumpling Restaurant	Diner	Discount Store	Dog Run	Doner Restaurant	Donut Shop	Drugstore
40	East Toronto	4	Park	Intersection	Coffee Shop	Convenience Store	Drugstore	Dim Sum Restaurant	Diner	Discount Store	Dog Run	Doner Restaurant
50	Rosedale	4	Park	Trail	Playground	Building	Dim Sum Restaurant	Diner	Discount Store	Dog Run	Doner Restaurant	Donut Shop
74	Caledonia-Fairbanks	4	Park	Women's Store	Market	Fast Food Restaurant	Comic Shop	Concert Hall	Event Space	Ethiopian Restaurant	Comfort Food Restaurant	Empanada Restaurant
98	Weston	4	Park	Convenience Store	Women's Store	Dumpling Restaurant	Diner	Discount Store	Dog Run	Doner Restaurant	Donut Shop	Drugstore

Image 19. List of neighborhoods in Cluster 4

This cluster has neighborhoods with the top ranked venue being parks, so this area is all about parks, playgrounds and trails.

Cluster 5 and 6: Cheap eats

	Neighborhood	Cluster Labels	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
1	Highland Creek,Rouge Hill,Port Union	5	Bar	Women's Store	Diner	Discount Store	Dog Run	Doner Restaurant	Donut Shop	Drugstore	Dumpling Restaurant	Fast Food Restaurant

Image 20. List of neighborhoods in Cluster 5

	Neighborhood	Cluster Labels	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	Rouge,Malvern	6	Fast Food Restaurant	Drugstore	Dim Sum Restaurant	Diner	Discount Store	Dog Run	Doner Restaurant	Donut Shop	Dumpling Restaurant	Harbor / Marina

Image 21. List of neighborhoods in Cluster 6

Clusters 5 and 6 have mostly fast food restaurants as venues, so we call these cheap eats.

From the above results, it seems to be logical to concentrate on Cluster 1 for our analysis. However, before going further, I wanted to see which neighborhoods from the clusters had at least 35 venues returned from Foursquare.

	Neighborhood	Neighborhood Latitude
0	Adelaide,King,Richmond	100
20	Chinatown,Grange Park,Kensington Market	100
82	St. James Town	100
79	Ryerson,Garden District	100
44	First Canadian Place,Underground city	100
32	Design Exchange,Toronto Dominion Centre	100
27	Commerce Court,Victoria Hotel	100
50	Harbourfront East,Toronto Islands,Union Station	100
83	Stn A PO Boxes 25 The Esplanade	98
19	Central Bay Street	88
22	Church and Wellesley	86
43	Fairview,Henry Farm,Oriole	68
65	Little Portugal,Trinity	64
8	Berczy Park	57
51	Harbourfront,Regent Park	50
15	Cabbagetown,St. James Town	45
88	The Danforth West,Riverdale	41
84	Studio District	40
74	Queen's Park	39
78	Runnymede,Swansea	36
95	Willowdale South	35
49	Harbord,University of Toronto	35

Image 22. List of neighborhoods in Cluster 2

The above is a list of neighborhoods from postal codes on Wikipedia. I then chose Chinese restaurants as a target restaurant to be opened and decided to check the distribution of Chinese speaking people across Toronto from the Census data. One of the things that was challenging was the fact that the neighborhoods in the Census dataset had different names in some rows and since we have grouped the neighborhoods before it was difficult to do a straight comparison. Below is a list of neighborhoods ranked by the prevalence of Chinese language from the Census dataset.

	chinese
City of Toronto	245,285
Milliken	16,510
Steeles	15,655
Agincourt North	14,670
Willowdale East	14,455
L'Amoreaux	14,450
Agincourt South-Malvern West	9,935
Tam O'Shanter-Sullivan	7,585
Hillcrest Village	7,255
Waterfront Communities-The Island	6,715
Bay Street Corridor	6,440
Don Valley Village	5,900
Kensington-Chinatown	5,170
Pleasant View	4,960
Bayview Village	4,895
South Riverdale	4,775
Newtonbrook East	4,015

Image 22. List of neighborhoods ranked by popularity of Chinese languages.

Based on the above list I then analyzed which neighborhoods had a few Chinese restaurants. I did that by looking at the most popular venues. However, I had to be careful not to take into account neighborhoods that had less than 35 venues. The table below summarizes the findings.

№	Neighborhood	Size of population that speaks Chinese	Number of venues returned from Foursquare	Is a Chinese restaurant in the top 5 venues? (ranking)	Worth pursuing?
1	Milliken	16,510	2*	No	Need further investigating
2	Steeles	15,655	9*	No	Need further investigating
3	Agincourt North	14,670	2*	NA	NA
4	Willowdale East	14,455	none	NA	NA
5	L'Amoreaux	14,450	13*	Yes (1)	No
6	Agincourt South-Malvern West	9,935	5*	Yes (3)	Need further investigating
7	Tam O'Shanter-Sullivan	7,585	11*	No	Need further investigating
8	Hillcrest Village	7,255	5*	No	Need further investigating
9	Waterfront Communities-The Island	6,715	15*	No	Need further investigating
10	Bay Street Corridor	6,440	86	No	Yes
11	Don Valley Village	5,900	No data	NA	NA
12	Kensington-Chinatown	5,170	100	Yes (3^{rd})	No
13	Pleasant View	4,960	No data	NA	NA
14	Bayview Village	4,895	4*	Yes ($4,5^{th}$)	No
15	South Riverdale	4,775	44	No	Yes
16	Newtonbrook East	4,015	No data	NA	NA

Table 2. Analysis of neighborhoods for suitability of opening a Chinese restaurant

* - means that there were not enough venues returned by Foursquare to be able to make a decision

5. Discussion

As an example, I have analyzed a suitability of opening a Chinese restaurant in different neighborhoods. Based on the above analysis the first thought is to analyze cluster 0 and 1 as they are the clusters with the most entertainment and recreation in them. However, a further analysis of this data coupled with Census data on Chinese speaking population across indicates that certain neighborhoods are already filled with a number of Chinese restaurants indicating that the supply is already there. After the analysis, there are several categories of neighborhoods: definitely worth pursuing, needs further analysis, not worth pursuing, not enough data to make a decision. Below are the neighborhoods in each category.

Definitely worth pursuing

Neighborhoods of Bay Street Coridor and South Riverdale are good candidates for further analysis since they are well represented in Foursquare with 100 and 86 venues returned respectively, they have a large population of Chinese speaking people and do not have Chinese restaurants in the top 10 most occurring venues.

Needs further analysis

The following neighborhoods are in this category - Milliken, Steeles, Agincourt South-Malvern West, Tam O'Shanter-Sullivan, Hillcrest Village, Waterfront Communities-The Island. Some of them look promising, others don't but the main reason they need further analysis is that there were not enough venues returned by Foursquare to make a final decision. Miliken and Steeles are of most interest because of the large number of Chinese speaking people in them.

Not worth pursuing

L'Amoreaux, Kensington-Chinatown, Bayview Village would be definitely hard to make an influence because of the predominance of Chinese restaurants there. Chinese restaurants rank in the top 5 most occurring venues in these neighborhoods, with L'Amoreaux having the most frequent venue being a Chinese restaurant.

Not enough data to make a decision

The following neighborhoods did not have data returned by Foursquare - Agincourt North, Willowdale East, Don Valley Village, Pleasant View, Newtonbrook East. We need to look at other sources of data for these neighborhoods.

After the above analysis, I was planning to analyze the safety data and demographic but due to time constraints on the project that will have to be done some other time.

6. Conclusion

According to my analysis, a good place to look to open a Chinese restaurant would be neighborhoods of Bay Street Coridor and South Riverdale. They have a substantial Chinese speaking population and do not have many Chinese restaurants; the analysis was done on a significant number of venues from Foursquare. I have analyzed the demand and supply side in this project.

Next steps I recommend would be:

1. Look at safety data from the Police portal to see how safe these neighborhoods are and whether it would be worth it opening a restaurant there
2. I would also look at demographic data as research shows that certain age groups are more willing to spend on eating out than other. Also, I would analyze these neighborhoods for predominance of families or single people as that would help in designing the restaurant.
3. There were not enough data in some cases and I was not able to definitively make a recommendation. I would look at other sources of data for the missing pieces such as Google geolocation data, Tripadvisor and such.

References

1. Coursera Capstone: Segmenting and Clustering Neighborhoods in Toronto
https://github.com/yerzatm/week-5-of-data-science-capstone

2. Population data across neighborhoods from the 2016 Census, Toronto
https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/index-eng.cfm

3. Demographic data from the Toronto Open Data Portal
https://www.toronto.ca/city-government/data-research-maps/open-data/

4. Neighborhood data and postal codes for Toronto
https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

5. Foursquare API
https://developer.foursquare.com/docs