

## The Battle of the Neighborhood

### 1. Introduction & Business Problem

#### 1.1 Problem Background

The city of Toronto is the most populated city in Canada, it is the capital of the province of Ontario and home to more than 2.7 million people, making it the fourth most populated city in America. North. In the 20th century, the city experienced significant industrial development and subsequently became the English-speaking banking, financial and commercial heart of Canada and was named in the 21st century as one of the most important financial centers in the world thanks to his medical research, his film production, technology and computer science, the arts, education and many more.

With its diverse cultural institutions, including many museums and art galleries, festivals and public events, and many more, Toronto becomes America's most visited city.

These important tourist visits generate a very high competitive on the market. As it is a very developed city, so is the business case. That's why any new investment or company that wants to move to Toronto needs to use market-based insights that will help them understand the business environment, allowing for a strategy to reduce risk. And increase the return on investment.

#### 1.2 Problem Description

A restaurant is a business establishment that serves prepared meals and beverages on the premises in exchange for a payment.

The food is usually prepared by a chef. The term covers a multiplicity of places and a great diversity of types of cuisine, both local and foreign. The restaurants are sometimes the device reserved for serving meals within a larger entity (hotel, university, airport ...), we speak then of collective restaurant as opposed to the kitchen site. They can also be associated with a catering or grocery business. The restaurant offers more or less comfort conditions, and the restaurant is called "fast" when the customer can order and eat in minutes or tens of minutes, possibly standing.

While searching on the internet I found that Toronto has several types of restaurants:

- Specific theme restaurants such as: Greek restaurants, French restaurants, Halal restaurants, Jewish restaurants and many more.

- 2-there are restaurants in hotels like the Drake Hotel.
- There are pizzerias also like Pizza Nova or Pizzayolo.
- 4-bistros like Bistro 990
- 5- Global restaurants like Big smoke Burger or fan's restaurant.
- 6- Restoring bars like the Rivoli

All this leads to a competitive market that requires distinguishing which are the criteria of location to choose to optimize income:

- Presence of supplier nearby?
- The demography of Toronto.
- Number of competitors nearby
- The demography of the neighborhood.
- The contribution of the people of Toronto.
- The list can continue ...

This analysis allows an investor to be able to choose an optimal location for his business and can use this analysis to move or open an extension of his business.

### 1.3 Target Audience

To find the right location a company named me as responsible Data Scientist in the objectives to analyze and find the neighborhood that will be the best choice to open a restaurant and present the arguments that go with it .Any entity considering opening a restaurant would be interested in this work.

### 1.4 Success Criteria

The criteria of success of this project is to find a location of this restoration and to justify the reasons for, the generated income satisfies the customer.

## 2. Description of Data

### 2.1 Data 1 Toronto neighborhood demography

This dataset contains figures on several demographics of the Toronto quarries in 2016, among them:

- Number of inhabitants of different ethnicities.

- Average monthly income.
- Number of women and men.
- Dominant languages
- Number of individuals per age group.
- Family member.

And many other more or less useful features.

The links is available via the following link:

<https://www.toronto.ca/city-government/data-research-maps/neighbourhoods-communities/neighbourhood-profiles/>

	Neighbourhood Number Neighbourhood Information	T SNS2020 Designation Neighbourhood Information	Population, 2016 Population and dwellings	Population, 2011 Population and dwellings	Population Change 2011-2016 Population and dwellings	Total private dwellings Population and dwellings	Private dwellings occupied by usual residents Population and dwellings	Population density per square kilometre Population and dwellings	Land area in square kilometres Population and dwellings	Children (0-14 years) Age characteristics	Youth (15-24 years) Age characteristics
	Agincourt North	129.0 No Designation	29113.0	30279.0	-3.9	9371.0	9120.0	3929.0	7.41	3840.0	3705.0
	Agincourt South-Malvern West	128.0 No Designation	23757.0	21988.0	8.0	8535.0	8136.0	3034.0	7.83	3075.0	3360.0
	Alderwood	20.0 No Designation	12054.0	11904.0	1.3	4732.0	4616.0	2435.0	4.95	1760.0	1235.0
	Annex	95.0 No Designation	30526.0	29177.0	4.6	18109.0	15934.0	10863.0	2.81	2360.0	3750.0
	Banbury-Den Mills	42.0 No Designation	27695.0	26918.0	2.9	12473.0	12124.0	2775.0	9.98	3605.0	2730.0
	Bathurst Manor	34.0 No Designation	15873.0	15434.0	2.8	6418.0	6089.0	3377.0	4.70	2325.0	1940.0
	Bay Street Corridor	76.0 No Designation	25797.0	19348.0	33.3	18436.0	15074.0	14097.0	1.83	1695.0	6860.0
	Bayview Village	52.0 No Designation	21396.0	17671.0	21.1	10111.0	9532.0	4195.0	5.10	2415.0	2505.0
	Bayview Woods-Steeles	49.0 No Designation	13154.0	13530.0	-2.8	4895.0	4698.0	3240.0	4.06	1515.0	1635.0
	Bedford Park-Nortown	39.0 No Designation	23236.0	23185.0	0.2	9052.0	8607.0	4209.0	5.52	4555.0	3210.0
10 rows x 2185 columns											

## 2.2 Data 2 foursquare

I used the Foursquare API to get all the sites within 1 mile (1.6 KM) and the center is at the coordinates of the Toronto neighborhoods.

This provides information on the longitude and latitude, the name of the establishment, the neighborhood it belongs to and other information.

In short, I took all the sites that are located in the neighborhoods of Toronto.

Neighborhoods	address	categories	cc	city	country	crossStreet	distance	formattedAddress	id	labeledLatL
0 Agincourt North	1571 Sandhurst Circle	Fast Food Restaurant	CA	Scarborough	Canada	at Finch Ave E	86	[1571 Sandhurst Circle (at Finch Ave E), Scarb...	4b79cec5f964a52065132fe3	{{'label': 'display', 43.808331321150
1 Agincourt North	115 Commander Blvd	Badminton Court	CA	Scarborough	Canada	NaN	1636	[115 Commander Blvd, Scarborough ON M1S 3M7, C...	537682a2498e21d3bd145911	{{'label': 'display', 43.793546085577
2 Agincourt North	McCowan Rd	Bus Line	CA	Toronto	Canada	Sheppard	2086	[McCowan Rd (Sheppard), Toronto ON, Canada]	4edd36982c5b50aba1f0ea86	{{'label': 'display', 43.790052037088
3 Agincourt North	5039 Finch Ave E	Residential Building (Apartment / Condo)	CA	Toronto	Canada	NaN	54	[5039 Finch Ave E, Toronto ON M1S 5L6, Canada]	4f64cd28e4b03be1036bcc7f	{{'label': 'display', 43.807633636322
4 Agincourt North	1571 Sandhurst Circle, Woodside Square	Discount Store	CA	Scarborough	Canada	in Woodside Square	290	[1571 Sandhurst Circle, Woodside Square (in Woo...	4bf43bc3cad2c92805af9b99	{{'label': 'display', 43.808894, 'lr

### 2.3 Data 3 Wikipedia List of city-designated neighbourhoods in Toronto.

This dataset contains all Toronto neighborhoods with their boroughs. It serves to assign a district to each neighborhood in Dataset 1.

[https://en.wikipedia.org/wiki/List\\_of\\_city-designated\\_neighbourhoods\\_in\\_Toronto](https://en.wikipedia.org/wiki/List_of_city-designated_neighbourhoods_in_Toronto)

	CDN number	City-designated area	Former city/borough	Neighbourhoods covered	Map
0	129	Agincourt North	Scarborough	Agincourt and Brimwood	NaN
1	128	Agincourt South-Malvern West	Scarborough	Agincourt and Malvern	NaN
2	20	Alderwood	Etobicoke	Alderwood	NaN
3	95	Annex	Old City of Toronto	The Annex and Seaton Village	NaN
4	42	Banbury-Don Mills	North York	Don Mills	NaN

## 3. Methodology

In the first place I extracted the data that I mentioned before then I analyzed its data to know their content, their size, the type of information they contain, missing information, etc.

### 3.1 Data 1 Toronto neighborhood demography

This CSV file is quite large and complete, it contains 2383 demographic information like the population of the kind by neighborhood, the average income of the inhabitants of each neighborhood and even number of dwellings by immigration, language, ethnicity or age group for each neighborhood.

I consider this file quite complete, nevertheless it contains features with null values that I deleted.

### 3.2 Data 2 foursquare

With the API of foursquare I was able to obtain information on each site (13658 sites) of each neighborhood, such as the coordinates of each site, its name and the neighborhood where it belongs.

However, there was a lack of information on the neighborhood or other things that I had to deal with later.

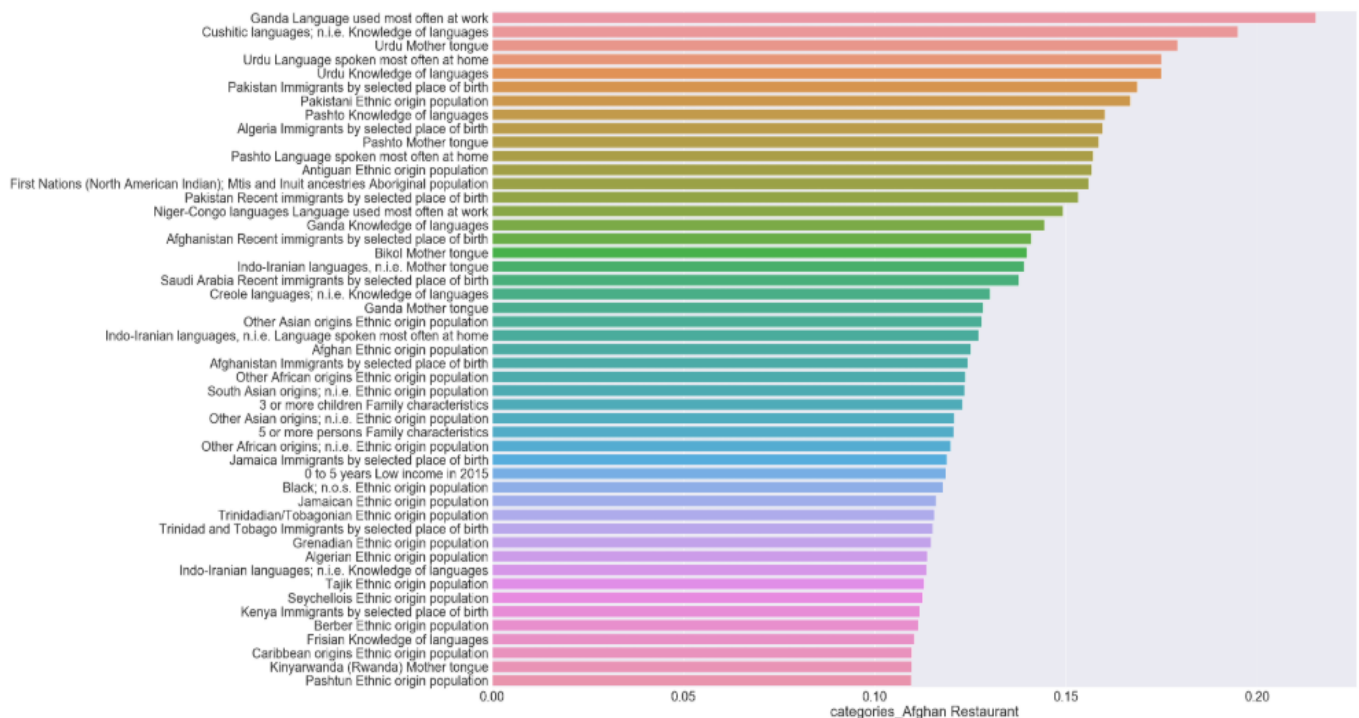
### 3.3 Data 3 Wikipedia List of city-designated neighborhoods in Toronto.

This is a simple enough file found on Wikipedia that served to correspond each neighborhood to its Brough.

### 3.4 Analysis and visualization of the data

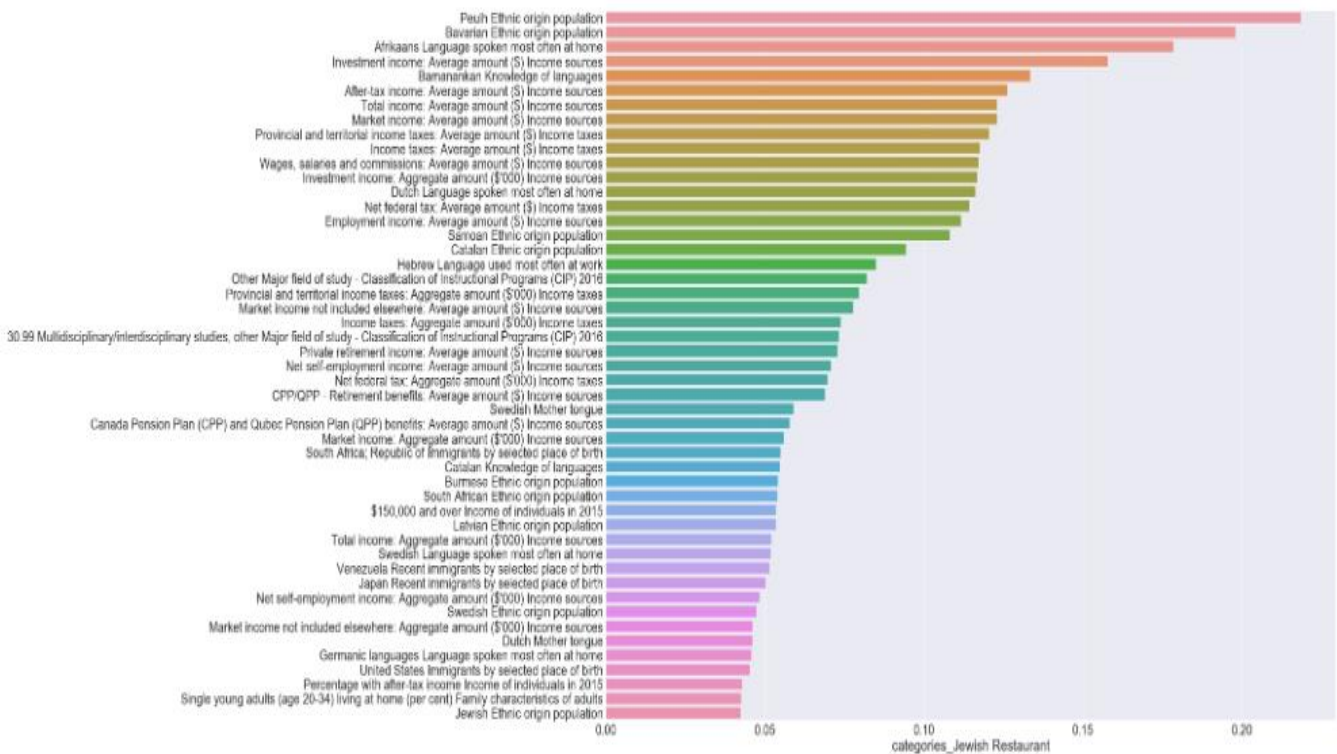
I asked myself the following question: under which criterion traders open their restaurants, by asking me this question I noticed that the restaurant category differs and I told myself that each category of restaurant targets specific people so there are some place selection criterion to open a restaurant depending on its category. Afterwards, I decided to study the correlation between information on demography and the positioning of the restaurants. However, at least I will have some idea of the location selection criteria.

It gave the following graphs:



This graph represents the correlation between Afghan restaurants and demographic information.

As we can see Ganda, gush tic and Urdu are the languages most present at the location of the Afghan restaurants, it is logical because these languages are spoken in the east of Africa and south east of Asia.

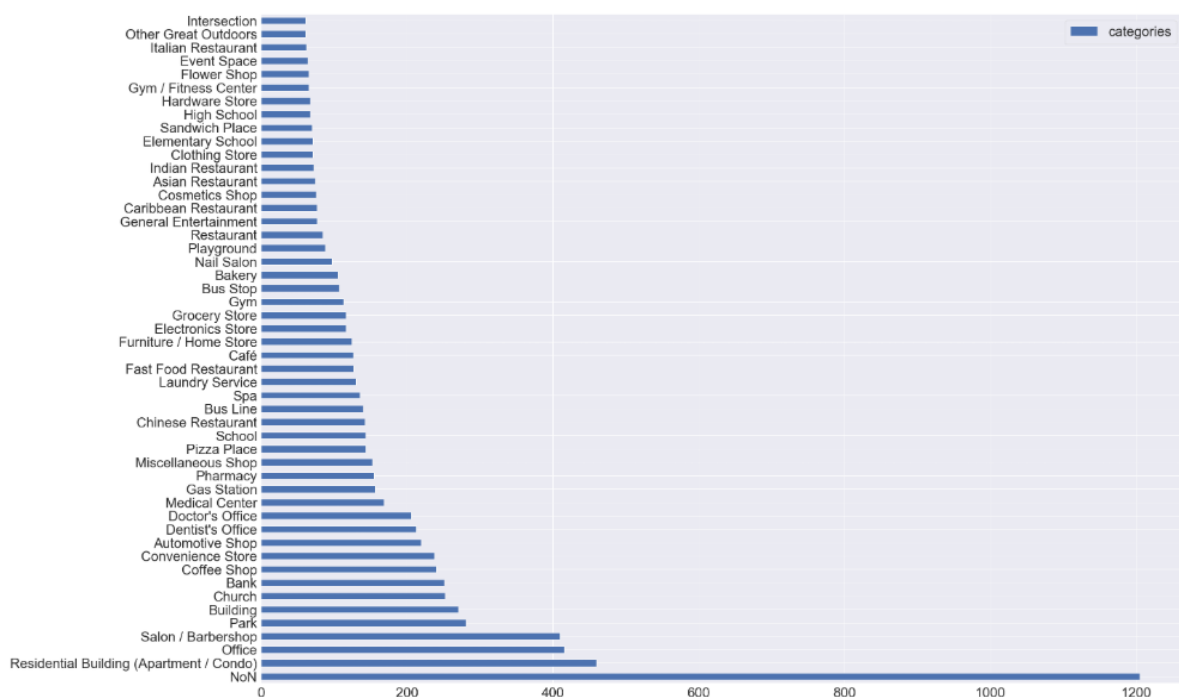


This graph represents the correlation between jewish restaurants and demographic information.

Peul and Bavarian ethnic are the most present in the neighborhoods with the most Jewish restaurant and the average income and also high.

Through the previous examples it can be deduced that the choice of location of a restaurant can change in relation to the type of restaurant, we speak on this point in the last part of the report.

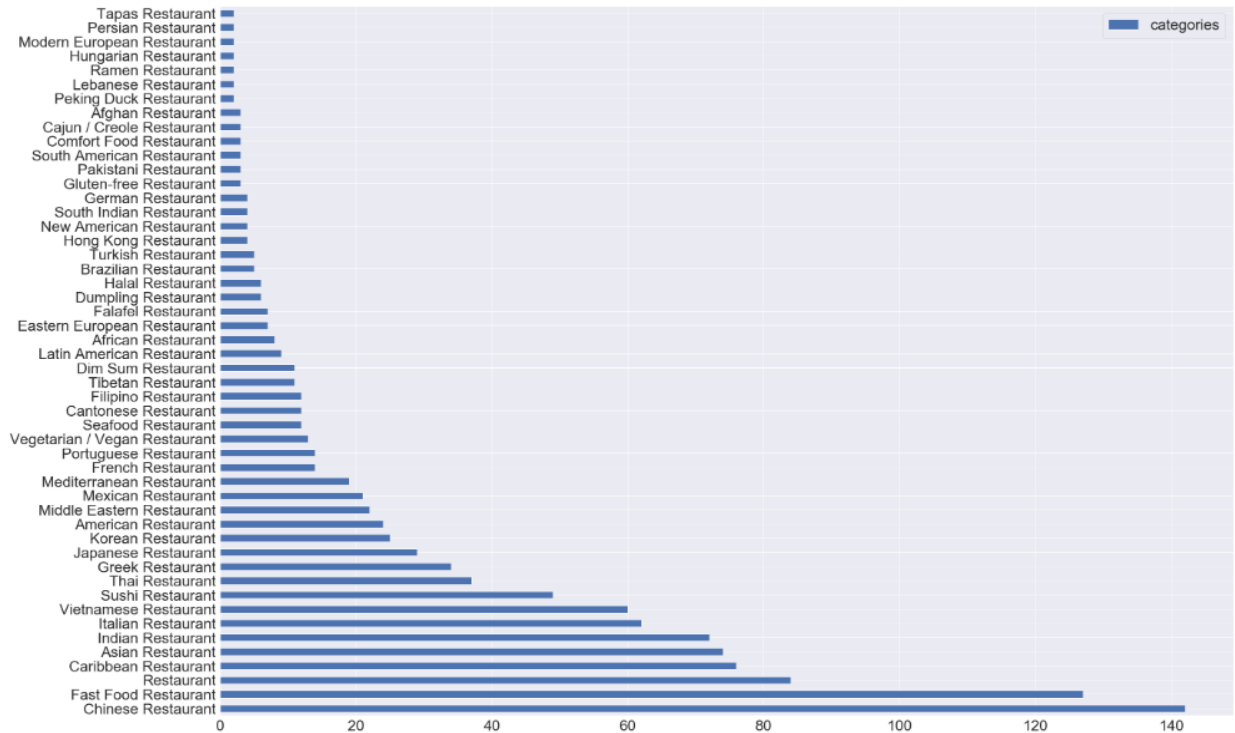
As the categories matter I wanted to know which the most dominant categories are.





The value NoN indicates no category for the site found by foursquare and that is generally present. However, Salon / Barbershop and Park are most present in Toronto.

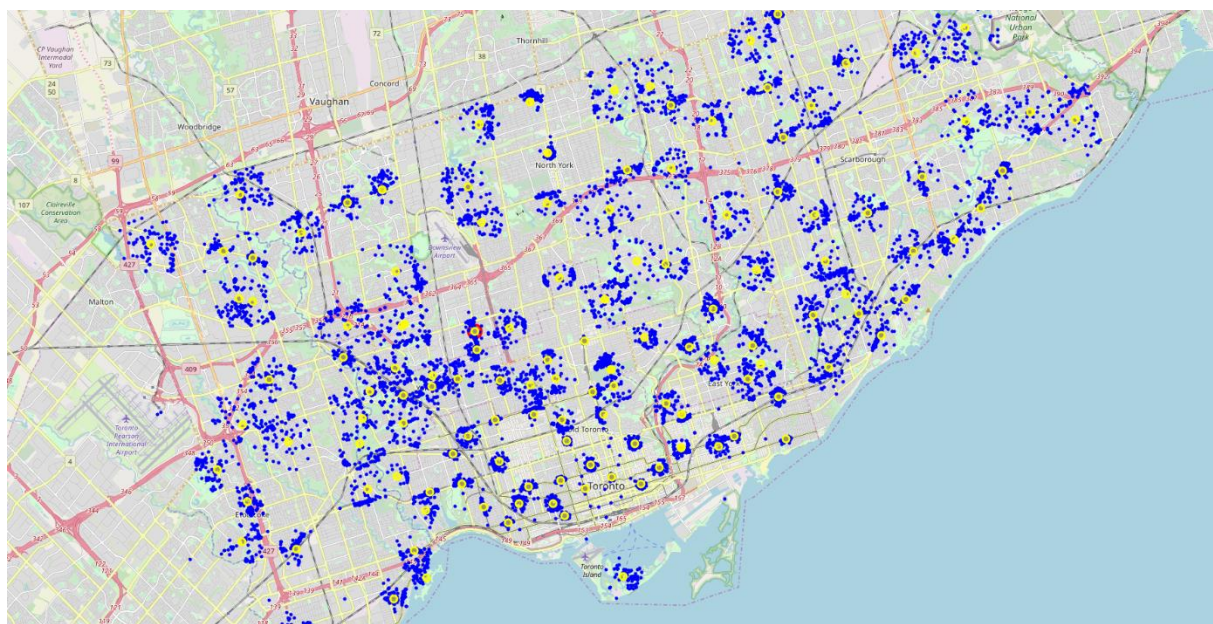
I redid the same analysis by choosing only the category of restaurant.



We can notice that the Chinese restaurant, the fast food restaurant and the restaurants without specify are the categories of restaurants most spread to Toronto.

Therefore it would be wise to choose a category not widely available because the competition is less harsh. I will mention this point in the last point of the report.

Before moving to the modeling step, I would like to mount the layout of the sites compared to neighborhood a Toronto.



The photo shows a significant density of sites in the west of Toronto.

### 3.6 Machine Learning

I chose to do a clustering on the data to know which neighborhoods groups are likely to be better to open a restaurant.

But before starting machine learning, I must choose the features to consider and I must prepare the data for what is consumable by Kmeans.

For the category of the sites I chose "Restaurant" because it is the subject of study, for the demographic information I have sorted the most likely information of favored the opening of a restaurant which are the following ones:

- Population
- Population density
- Persons living alone (per cent)
- Total income: Average
- Non-permanent residents Immigran
- Youth (15-24 years)
- Working Age (25-54 years)
- Females Population
- Males Population
- After-tax income
- langitude (to save the location)
- latitude (to save the location)

Then I normalized the previous features, and I summed the lines to get a score of each line. This score is an indicator if a neighborhood is favorable for restaurant opening or not.

Then I took the top 50 neighborhoods by sorting the score from the largest to the smallest.

Finally I added to each one of the 50 neighborhoods the number of restaurants in his area.

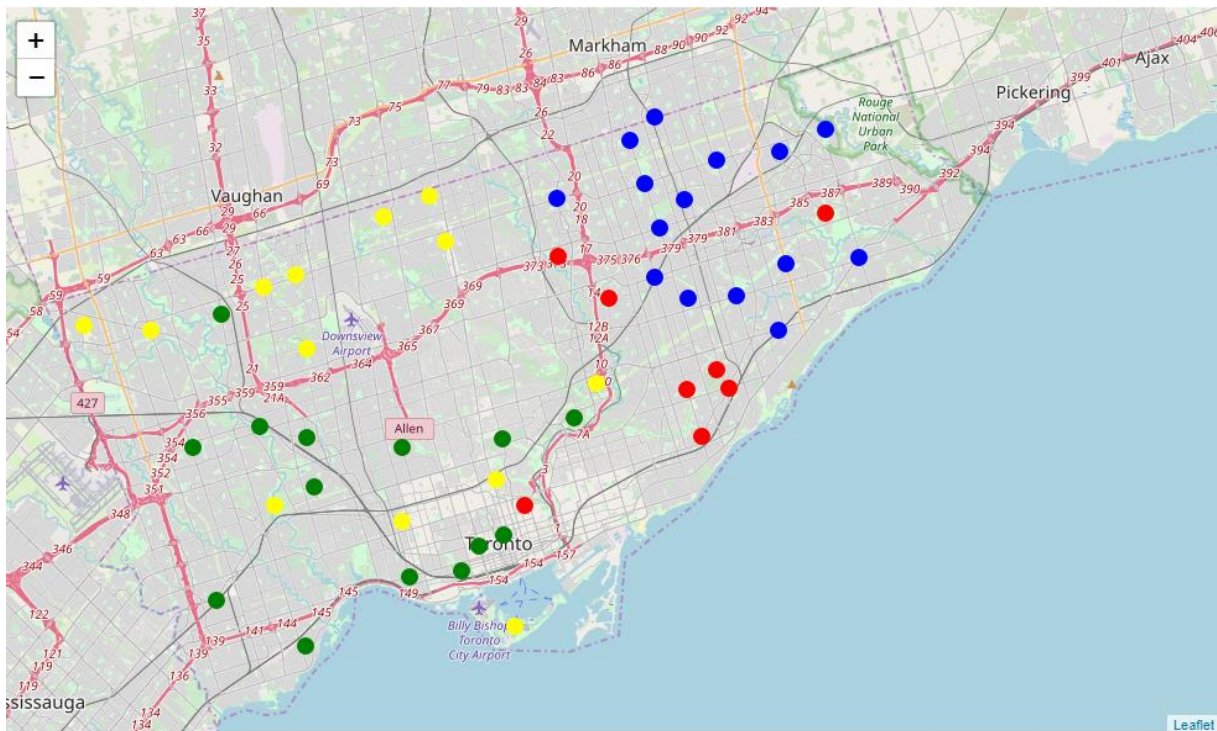
In short, I add 2 features that are: the score of each neighborhood in relation to his demography and the number of restaurants it contains.

At the end the DataFrame that will be consumed by Kmeans contains the following features:



- Longitude.
- Latitude.
- Score.
- Number of restaurant.

I formed the K Means clustering algorithm with these data, by choosing 4 clusters (this choice was made after several tests). And I displayed the result on the map by assigning each cluster of neighborhoods a unique color, and the result is:



#### 4. Discuss the Results

Kmeans returned 4 clusters and each cluster has its own characteristics:

- **Cluster 1 in red:** groups together neighborhoods with a relatively low score and with a relatively low restaurant count in its perimeter  
average Number\_of\_Restaurant cluster 1 : 12.0  
average score cluster 1 : 2.27
- **Cluster 2 in blue:** groups neighborhoods with a relatively high score and a relatively high number of restaurants also in its perimeter  
average Number\_of\_Restaurant cluster 2 : 18.8  
average score cluster 2 : 2.39

- **Cluster 3 in green** : groups neighborhoods with a relatively low score and a relatively high number of restaurants also in its perimeter  
average Number\_of\_Restaurant cluster 3 : 19.64  
average score cluster 3 : 2.30
- **Cluster 4 in yellow** : groups neighborhoods with a relatively high score and a relatively low number of restaurants also in its perimeter  
average Number\_of\_Restaurant cluster 4 : 13.692307692307692  
average score cluster 4 : 2.4528401540266525

## 5. Recommendations

The recommendations I bring for a restaurant opening after all this project:

- The criteria for opening a restaurant depend partly on the category of restaurant, if you want to open a Chinese restaurant, you target neighborhoods with a Chinese population quite present, as well as for an Italian restaurant or other.
- The Chinese restaurants and the fast foods are the most present by conceiving the competition is rough so it is preferable to choose another category.
- I recommend the East Willowdale, Mount Olive-Silverstone-Jamestown, Waterfront Communities-The Island and Dovercourt-Wallace Emerson-Junction neighborhoods to open a restaurant because they are part of a cluster that has the fewest restaurants so less competition and they have a high demographic score so it is favorable to a high turnover.

## 6. Conclusion

we can rely on the results quoted before even if it remains imprecise and that because of the lack of data provided by the foursquare API, a premium account will give us the possibility of seeing the note of the people and it will facilitate us the work better, by using a far-off regression we will get the right selection of a neighborhood and the result will be more accurate.

Also later we will consider more data to reinforce our choice like, the trade surrounding the restaurants, the crime scene of neighborhoods and others. Thanx =)

