Thinking of an interesting business

Jolman Ferley Salinas

March 12, 2020

Background

Fruits and vegetables play an important role in a healthy diet. However, there could be harmful bacteria in the soil or water where these agricultural products grow that come in contact with them and contaminate them. Fresh fruits and vegetables can also become contaminated after being harvested, such as during storage or preparation, hence the importance of processing and consuming them as soon as possible.

Toronto is the capital of the province of Ontario and, with a population of 2,615,060 in habitants and is the largest city in Canada in addition to the financial center of that country.
As the economic capital of Canada, Toronto is considered a global city and one of the main financial cities in the world.

 The former municipalities that make up the current city of Toronto are East York, Etobicoke, North York, Old Toronto, Scarborough and York

Problem

An investor dedicated to the creation of emerging businesses and with great knowledge in Entrepreneurship scalability tells me that you are interested in opening a point of sale for fresh vegetables. He asks me to indicate a good place

to establish this business and a complete analysis of the place to make a good decision.

North York is one of the largest neighborhoods in Toronto, Canada with great potential to create a niche related to fresh vegetables and vegetables of all kinds. An investor is willing to offer this service guaranteeing that quality and freshness are its added value. You want to know where in North York you can place your point of sale. Consider that the best points to make this investment should be close to restaurants of all kinds since they are the ones potential. The investor wants to try to start in this neighborhood and then expand throughout Toronto, That is why it is important that recognition and quality as well as its good location allow it.

Data acquisition and cleaning

It was necessary to download the postal codes of Toronto, for this I will use the same font used in the previous week. In the same way, information is extracted from fousquare to know the commerce sites and be able to meet the objective, Also from this source and based on the postal codes the precise geographical location was extracted given the longitude and latitude.

A typical request from Foursquare will provide us with the following information: [Postal Code] [Neighborhood(s)] [Neighborhood Latitude] [Neighborhood Longitude] [Venue] [Venue Summary] [Venue Category] [Distance (meter)] [M1L] [Clairlea, Golden Mile, Oakridge] [43.711112] [-79.284577] [Tim Hortons] [This spot is popular] [Coffee Shop] [592.

Foursquare is a service based on web location applied to social networks. Geolocation allows you to locate a fixed or mobile device in a geographic location

Other important information has to do with the opinions that people give about the network since check-in is marked specific places where one is and in this way it is understood in a better way the tastes and preferences to better locate the business of selling fresh vegetables.

Additional research is carried out at the Canadian embassy to find out what people prefer when it comes to food, for this I make consultations in different sources, mainly the surveys carried out by government entities. In the same way, I checked with people who have traveled to Toronto to know their impressions about the possible investment.

2.1 Data sources

It is taken from https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M, everything related to Toronto neighborhoods and neighborhoods.

In the same way, information is consulted at https://www.canadainternational.gc.ca/colombia-colombie/index.aspx?lang=en

2.2 Data cleaning

The information obtained from Wikipedia corresponds to Postcode, Borough and Toronto Neighborhood.

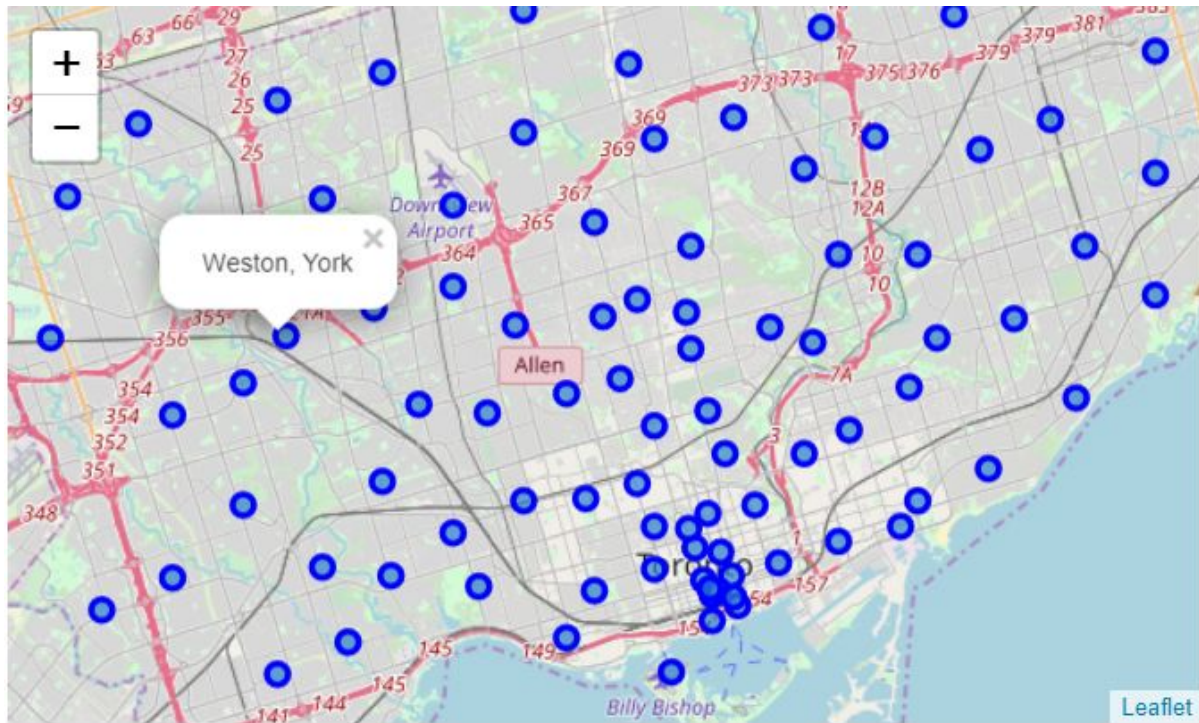|	Postcode	Borough	Neighbourhood
2	M3A	North York	Parkwoods
3	M4A	North York	Victoria Village
4	M5A	Downtown Toronto	Harbourfront
5	M6A	North York	Lawrence Heights
6	M6A	North York	Lawrence Manor

Because the postal codes are repeated for different neighborhoods, they are grouped by these so as not to fall into redundancies of information.

	Postcode	Borough	Neighbourhood
0	M1B	Scarborough	Rouge, Malvern
1	M1C	Scarborough	Highland Creek, Rouge Hill, Port Union
2	M1E	Scarborough	Guildwood, Morningside, West Hill
3	M1G	Scarborough	Woburn
4	M1H	Scarborough	Cedarbrae
5	M1J	Scarborough	Scarborough Village
6	M1K	Scarborough	East Birchmount Park, Ionview, Kennedy Park
7	M1L	Scarborough	Clairlea, Golden Mile, Oakridge
8	M1M	Scarborough	Cliffcrest, Cliffside, Scarborough Village West
9	M1N	Scarborough	Birch Cliff, Cliffside West

Geographical coordinates are added to the grouped dataframe to have everything necessary for the next steps.
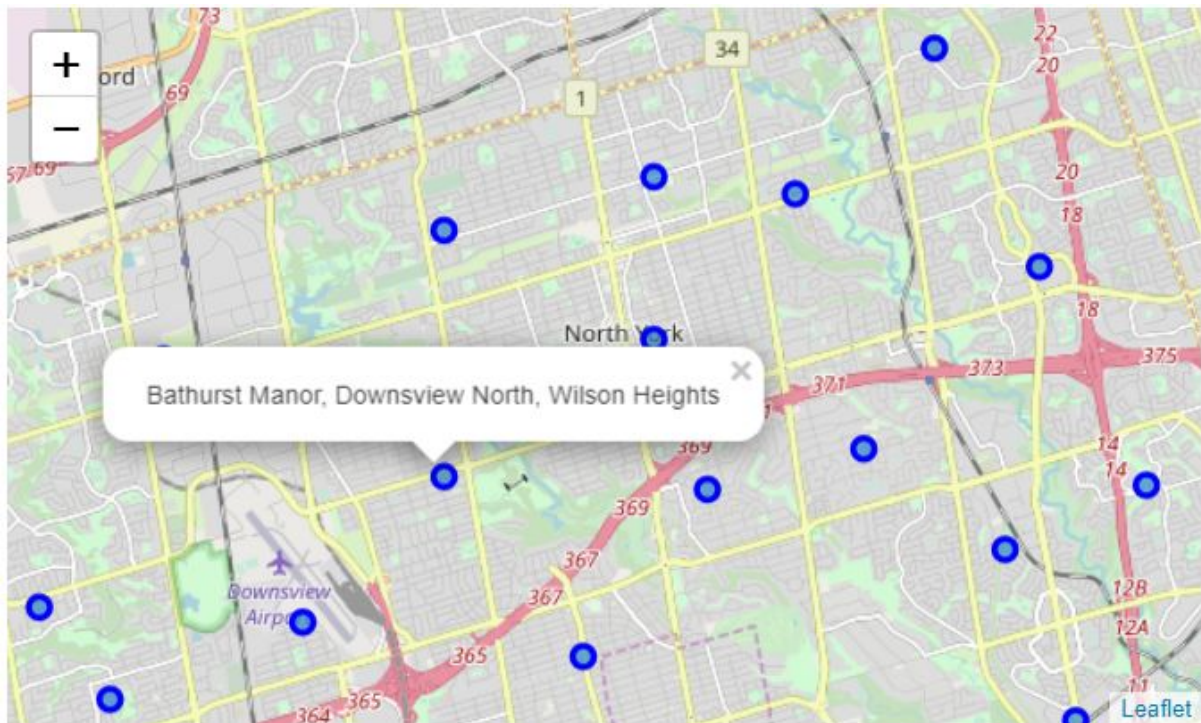
2.3 Selection

After conducting a review of all the neighborhoods that exist in Toronto and making a comparison with the information found in web searches, the Canadian Embassy and visual exploration (maps), North York was selected to search for neighborhoods and venues.

The longitude and latitude of all North York neighborhoods is located.

	Postcode	Borough	Neighbourhood	Latitude	Longitude
0	M2H	North York	Hillcrest Village	43.803762	-79.363452
1	M2J	North York	Fairview, Henry Farm, Oriole	43.778517	-79.346556
2	M2K	North York	Bayview Village	43.786947	-79.385975
3	M2L	North York	Silver Hills, York Mills	43.757490	-79.374714
4	M2M	North York	Newtonbrook, Willowdale	43.789053	-79.408493

2.4 Using Foursquare

Foursquare is used to extract all venues from North York

```
LIMIT = 100
radius = 500
url = 'https://api.foursquare.com/v2/venues/explore?client_id={}&client_secret={
```

53 venues were returned by Foursquare.

```
Hillcrest Village
Fairview, Henry Farm, Oriole
Bayview Village
Silver Hills, York Mills
Newtonbrook, Willowdale
Willowdale South
York Mills West
Willowdale West
Parkwoods
Don Mills North
Flemingdon Park, Don Mills South
Bathurst Manor, Downsview North, Wilson Heights
Northwood Park, York University
CFB Toronto, Downsview East
Downsview West
Downsview Central
Downsview Northwest
Victoria Village
Bedford Park, Lawrence Manor East
Lawrence Heights, Lawrence Manor
Glencairn
Downsview, North Park, Upwood Park
Humber Summit
Emery, Humberlea
```

Additionally, a dataframe is created with the names, the location of the venue and the neighborhood, and the type of venue it is.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Hillcrest Village	43.803762	-79.363452	Eagle's Nest Golf Club	43.805455	-79.364186	Golf Course
1	Hillcrest Village	43.803762	-79.363452	New York Fries	43.803664	-79.363905	Fast Food Restaurant

This dataframe is grouped by Neighborhood

```
North_York_venues.groupby('Neighborhood').count()
```

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Bathurst Manor, Downsview North, Wilson Heights	22	22	22	22	22	22
Bayview Village	4	4	4	4	4	4
Bedford Park, Lawrence Manor East	24	24	24	24	24	24
CFB Toronto, Downsview East	3	3	3	3	3	3
Don Mills North	6	6	6	6	6	6
Downsview Central	3	3	3	3	3	3

Finally one hot encoding is done to prepare the dataframe for the application of Machine Learning Technique.

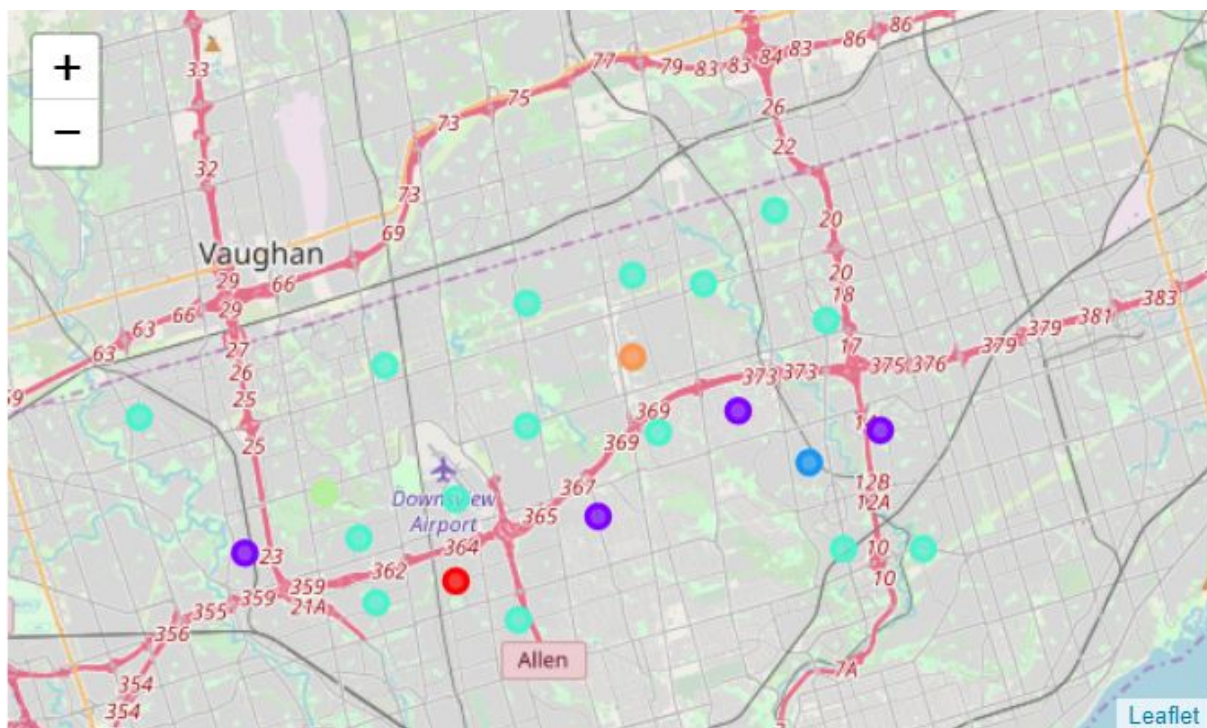|	Neighborhood	Accessories Store	Airport	American Restaurant	& Crafts Store	Asian Restaurant	Bakery	Bank
0	Bathurst Manor, Downsview North, Wilson Heights	0.0	0.000000	0.000000	0.0	0.0	0.0	0.045455
1	Bayview Village	0.0	0.000000	0.000000	0.0	0.0	0.0	0.250000
2	Bedford Park, Lawrence Manor East	0.0	0.000000	0.041667	0.0	0.0	0.0	0.000000
	CFB Toronto							

3. Modeling

3.1 Applying one of Machine Learning Techniques (K-Means Clustering)

Here we cluster neighborhoods via K-means clustering method. We think that 6 clusters is enough and can cover the complexity of our problem.
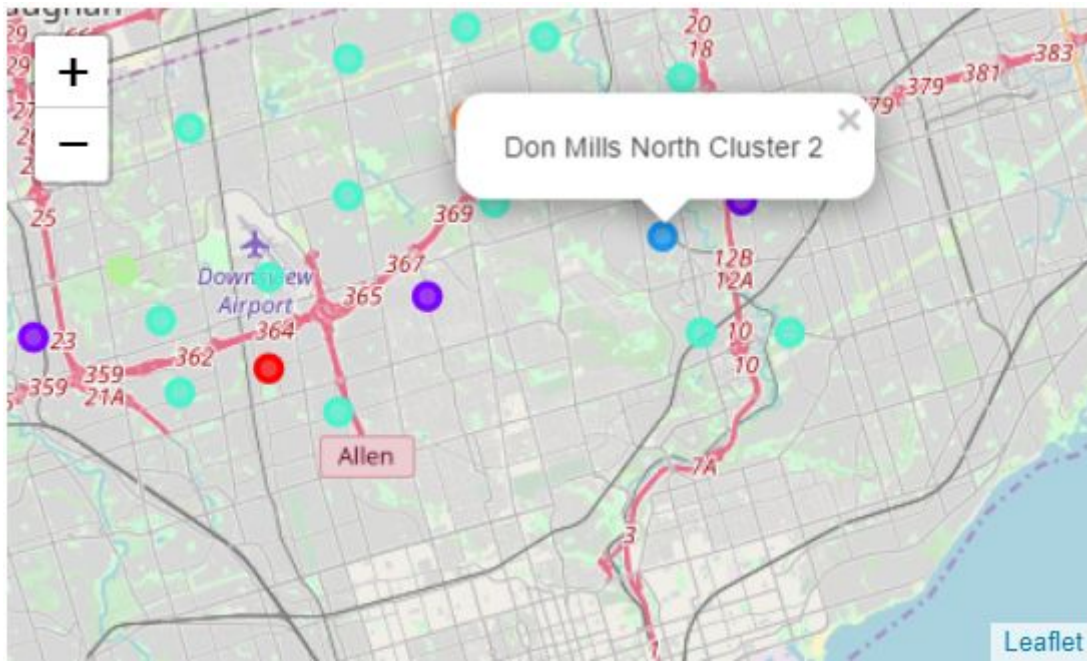
3.2 North York merged

Merge toronto grouped with toronto data to add latitude/longitude for each neighborhood and plot the definitive clustering.

4. Conclusion

It is recommended that the investor search for and establish his fresh vegetable business in the Don Mills North neighborhood since the most common venue is Caribbean Restaurant and the others are related to healthy living, which is precisely what fresh vegetables give. The others are not recommended because they are mostly for sale or entertainment, which would not be a good place.

Don Mills North Cluster 2

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	Caribbean Restaurant	Gym / Fitness Center	Café	Baseball Field	Basketball Court	Japanese Restaurant	Women's Store	Discount Store	Construction & Landscaping	Convenience Store