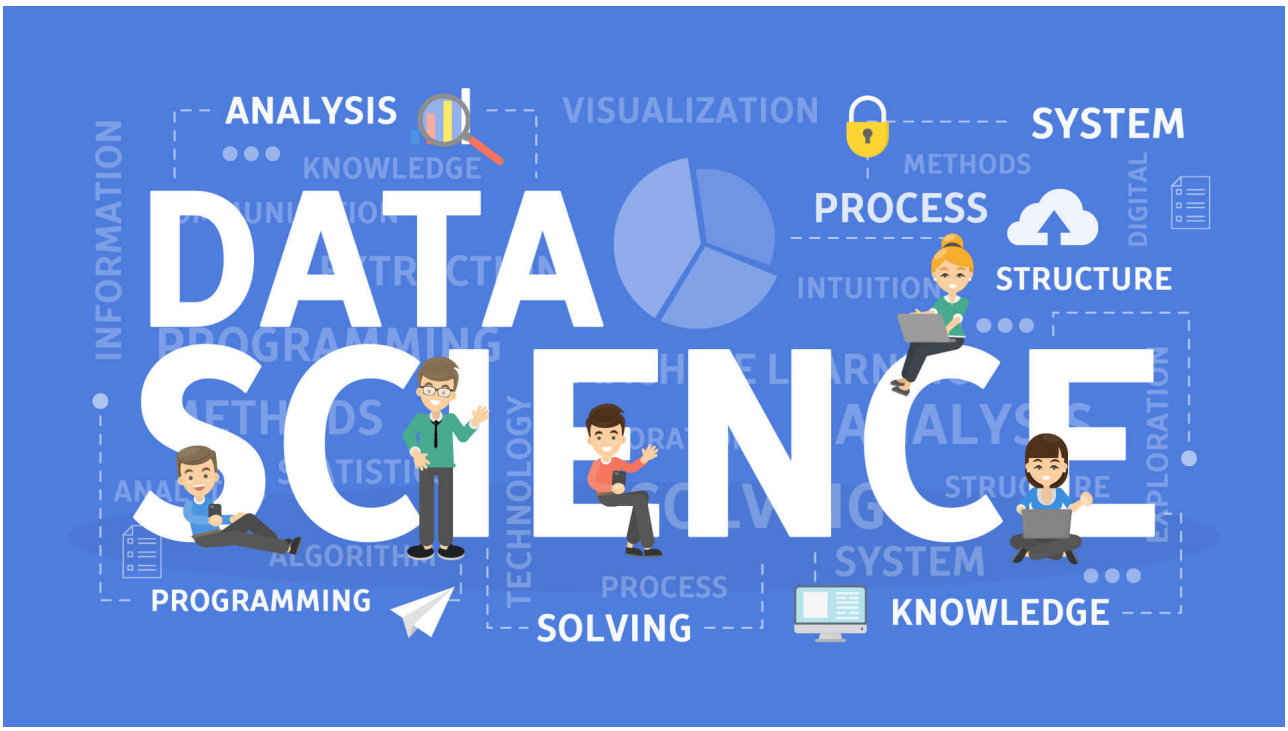


Capstone Project Report

Author : Amine Taghrout

2019 — 2020

Table des matières

1 Introduction

In this project we will try to find an optimal location for a restaurant in Etobicoke. Specifically, this report will be targeted to stakeholders interested in opening a restaurant in Etobicoke, Canada. Since there are lots of restaurants in York we will try to detect locations that are not already crowded with restaurants.
We will use our data science powers to generate a few most promising neighborhoods based on this criteria. Advantages of each area will then be clearly expressed so that best possible final location can be chosen by stakeholders.

2 Data

Based on definition of our problem, factors that will influence our decision are :

- Number of existing restaurants in the neighborhood (any type of restaurant)
- Some characteristics of neighborhoods

We decided to use regularly spaced grid of locations, centered around city center, to define our neighborhoods.
Following data sources will be needed to :

- Extract/generate the required data by scraping it from the following website : `https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M`
- Transform and treat the scraped data
- Know the number of restaurants and their type and location in every neighborhood using Foursquare API

3 Methodology

First of all, we will scrap data from `https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M` and we will transform it to a data-frame that we can use in our study. Next, we will convert addresses into their equivalent latitude and longitude values. Also, we will use the Foursquare API to explore neighborhoods in Etobicoke and we will use the explore function to get the most common venue categories in each neighborhood, and then use this feature to group the neighborhoods into clusters. You will use the k-means clustering algorithm to complete this task. Finally, you will use the Folium library to visualize the neighborhoods in Etobicoke and their emerging clusters.

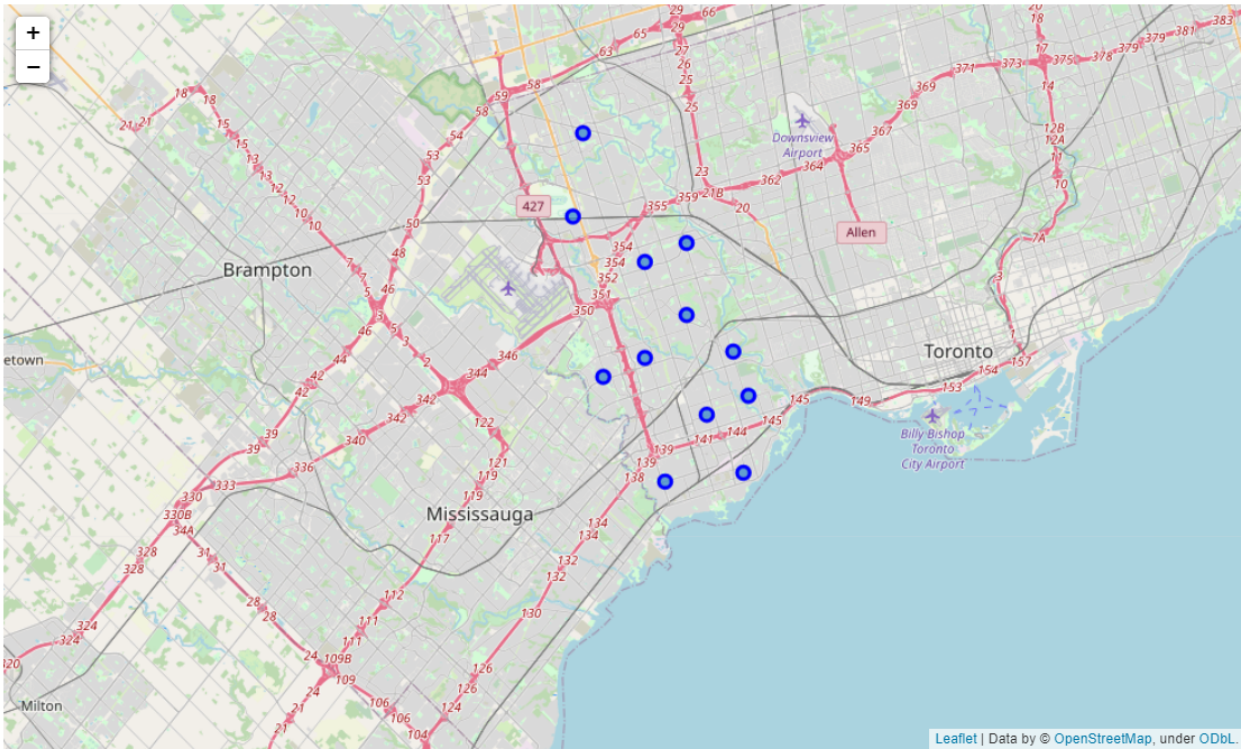
4 Analyzing

4.1 Scraping and transforming data

After scraping data from the URL which we talked about previously, we did some transformations like dropping the NA rows and the Borough that were not assigned. We also created a data-frame that contains only the neighborhood of Etobicoke as shown below.

	Postal Code	Borough	Neighborhood	Latitude	Longitude
0	M9A	Etobicoke	Islington Avenue, Humber Valley Village	43.667856	-79.532242
1	M9B	Etobicoke	West Deane Park, Princess Gardens, Martin Grov...	43.650943	-79.554724
2	M9C	Etobicoke	Eringate, Bloordale Gardens, Old Burnhamthorpe...	43.643515	-79.577201
3	M9P	Etobicoke	Westmount	43.696319	-79.532242
4	M9R	Etobicoke	Kingsview Village, St. Phillips, Martin Grove ...	43.688905	-79.554724
5	M8V	Etobicoke	New Toronto, Mimico South, Humber Bay Shores	43.605647	-79.501321
6	M9V	Etobicoke	South Steeles, Silverstone, Humbergate, Jamest...	43.739416	-79.588437
7	M8W	Etobicoke	Alderwood, Long Branch	43.602414	-79.543484
8	M9W	Etobicoke	Northwest, West Humber - Clairville	43.706748	-79.594054
9	M8X	Etobicoke	The Kingsway, Montgomery Road, Old Mill North	43.653654	-79.506944
10	M8Y	Etobicoke	Old Mill South, King's Mill Park, Sunnylea, Hu...	43.636258	-79.498509
11	M8Z	Etobicoke	Mimico NW, The Queensway West, South of Bloor,...	43.628841	-79.520999

Finally, we created a map showing Etobicoke and its different neighborhoods.



4.2 Dummies and scoring

To use clustering, we had to re-code the data by using the dummies (0-1) as shown below.

	American Restaurant	Bakery	Bar	Baseball Field	Beer Store	Burger Joint	Bus Line	Café	Chinese Restaurant	Coffee Shop	...	Pub	Rental Car Location	Restaurant	River	Sandwich Place	Social Club	Supplement Shop
0	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0
2	0	0	0	0	1	0	0	0	0	0	...	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	1	...	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0

5 rows × 40 columns

After that, we use some metrics like the mean one to obtain scores as shown below

	Neighborhood	American Restaurant	Bakery	Bar	Baseball Field	Beer Store	Burger Joint	Bus Line	Café	Chinese Restaurant	...	Pizza Place	Pub	Rental Car Location	Restaurant
0	Alderwood, Long Branch	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	0.00	0.000000	0.000000	...	0.285714	0.142857	0.000000	0.000000
1	Eringate, Bloordale Gardens, Old Burnhamthorpe...	0.000000	0.000000	0.000000	0.0	0.142857	0.000000	0.00	0.142857	0.000000	...	0.142857	0.000000	0.000000	0.000000
2	Kingsview Village, St. Phillips, Martin Grove ...	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	0.25	0.000000	0.000000	...	0.250000	0.000000	0.000000	0.000000
3	Mimico NW, The Queensway West, South of Bloor,...	0.000000	0.066667	0.000000	0.0	0.000000	0.066667	0.00	0.000000	0.000000	...	0.000000	0.000000	0.000000	0.000000
4	New Toronto, Mimico South, Humber Bay Shores	0.071429	0.071429	0.000000	0.0	0.000000	0.000000	0.00	0.142857	0.000000	...	0.071429	0.000000	0.000000	0.071429
5	Northwest, West Humber - Clairville	0.000000	0.000000	0.333333	0.0	0.000000	0.000000	0.00	0.000000	0.000000	...	0.000000	0.000000	0.333333	0.000000
6	Old Mill South, King's Mill Park, Sunnylea, Hu...	0.000000	0.000000	0.000000	0.5	0.000000	0.000000	0.00	0.000000	0.000000	...	0.000000	0.000000	0.000000	0.000000
7	South Steeles, Silverstone, Humbergate, Jamest...	0.000000	0.000000	0.000000	0.0	0.111111	0.000000	0.00	0.000000	0.000000	...	0.111111	0.000000	0.000000	0.000000
8	The Kingsway, Montgomery Road, Old Mill North	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	0.00	0.000000	0.000000	...	0.000000	0.000000	0.000000	0.000000
9	West Deane Park, Princess Gardens, Martin Grov...	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	0.00	0.000000	0.000000	...	0.000000	0.000000	0.000000	0.000000

Finally, we ranked the neighborhoods using the 5 best scores obtained previously.

```
----Alderwood, Long Branch----
            venue  freq
0     Pizza Place  0.29
1             Gym  0.14
2  Sandwich Place  0.14
3        Pharmacy  0.14
4     Coffee Shop  0.14


----Eringate, Bloordale Gardens, Old Burnhamthorpe, Markland Wood----
          venue  freq
0  Liquor Store  0.14
1    Beer Store  0.14
2          Café  0.14
3     Pet Store  0.14
4   Coffee Shop  0.14


----Kingsview Village, St. Phillips, Martin Grove Gardens, Richview Gardens----
                venue  freq
0         Pizza Place  0.25
1      Sandwich Place  0.25
2            Bus Line  0.25
3   Mobile Phone Shop  0.25
4  American Restaurant  0.00


----Mimico NW, The Queensway West, South of Bloor, Kingsway Park South West, Royal York South West----
             venue  freq
0              Gym  0.07
1   Discount Store  0.07
2    Tanning Salon  0.07
3  Supplement Shop  0.07
4      Social Club  0.07
```

4.3 Clustering

As you can see from the table below which shows an example of clusters, we clustered the neighborhoods and we did the project.

	Postal Code	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
0	M9A	Etobicoke	Islington Avenue, Humber Valley Village	43.667856	-79.532242	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1	M9B	Etobicoke	West Deane Park, Princess Gardens, Martin Grov...	43.650943	-79.554724	1.0	Filipino Restaurant	Wings Joint	Coffee Shop	Flower Shop	Fast Food Restaurant	Drugstore	Discount Store
2	M9C	Etobicoke	Eringate, Bloordale Gardens, Old Burnhamthorpe...	43.643515	-79.577201	2.0	Pharmacy	Coffee Shop	Beer Store	Liquor Store	Café	Pizza Place	Pet Store
3	M9P	Etobicoke	Westmount	43.696319	-79.532242	2.0	Coffee Shop	Sandwich Place	Intersection	Discount Store	Pizza Place	Middle Eastern Restaurant	Chinese Restaurant
4	M9R	Etobicoke	Kingsview Village, St. Phillips, Martin Grove ...	43.688905	-79.554724	2.0	Mobile Phone Shop	Sandwich Place	Bus Line	Pizza Place	Wings Joint	Filipino Restaurant	Fast Food Restaurant

5 Conclusion

This project was really helpful because it helped us to practice what we learned from IBM courses.