

Site Selection Optimization Model

Based on Segmenting and Clustering

Neighborhoods in Toronto

Larry Qu

September 6, 2019

1. Introduction

There are many new but rich immigrants to Toronto these days. I have many friends from China, Hong Kong and Taiwan who want to start their new life and business in Toronto. To get the best site selection, we need to use Foursquare to understand different neighborhoods' feature.

In addition, we need to learn the condition of different owners, including their budget, age and industry.

This research can give us good example of property site selection process for business owners. This research is based on Toronto, but every city, with Foursquare data and geographic data, can follow the same process.

2. Data acquisition and cleaning

1) Data Sources

There are two main data frames. One is Toronto geographic information, which we can get from Wikipedia. The other is different business owners' data, which is from the interview between my friends and me.

2) Data Cleaning

For Toronto geographic data, firstly, we need to delete all null value tuples. And we need to combine all neighborhoods in one postcode.

Out[5]:

Unnamed: 0	Postcode	Borough	Neighborhood	Latitude	Longitude
0	0	M1B Scarborough	Rouge, Malvern	43.806686	-79.194353
1	1	M1C Scarborough	Highland Creek, Rouge Hill, Port Union	43.784535	-79.160497
2	2	M1E Scarborough	Guildwood, Morningside, West Hill	43.763573	-79.188711
3	3	M1G Scarborough	Woburn	43.770992	-79.216917
4	4	M1H Scarborough	Cedarbrae	43.773136	-79.239476

For business owners' data, we need to delete all unreasonable information, for example, unrealistic budget limit.

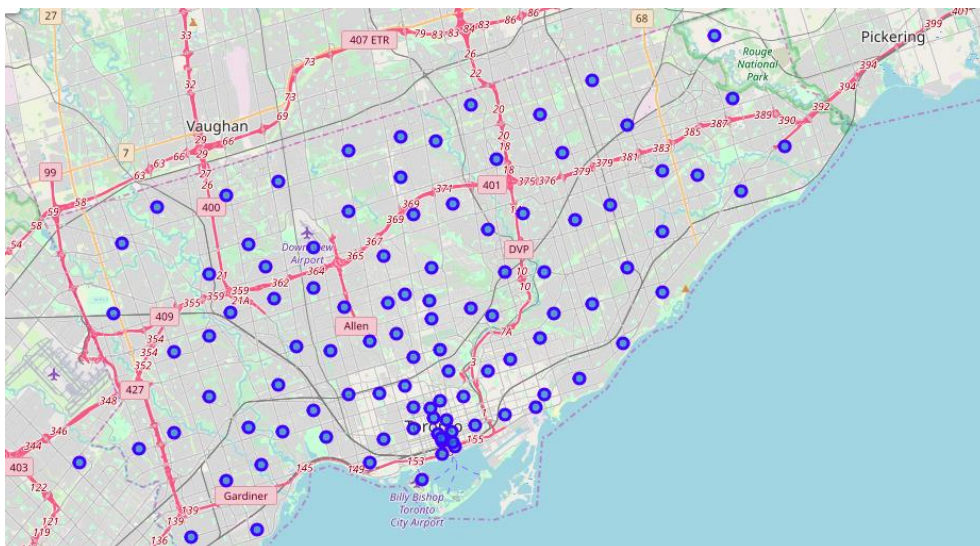
Out[6]:

	Store_name	Owner	Is_restaurant	Category	Is_chain	Age	Budget
0	ABC	Jack	0	Bar	0	0	100
1	DEF	Lisa	1	Res	0	5	500
2	GHI	Linda	1	Res	1	15	800
3	JKL	Mary	0	Gro	0	2	100
4	MNO	James	1	Res	1	7	700

3. Exploratory Data Analysis

1) Combine neighborhood information with map

Use Folium map tool to draw the map of Toronto with neighborhoods in blue spots.



2) Cluster Foursquare data with map

After combining Foursquare data with map, we should cluster different categories. With KKN method, we can cluster all neighborhoods into 5 categories.



3) Get common venues of the neighborhood

We need to know areas' common venue condition. Our Strategy is agglomeration economic but not too fierce competition.

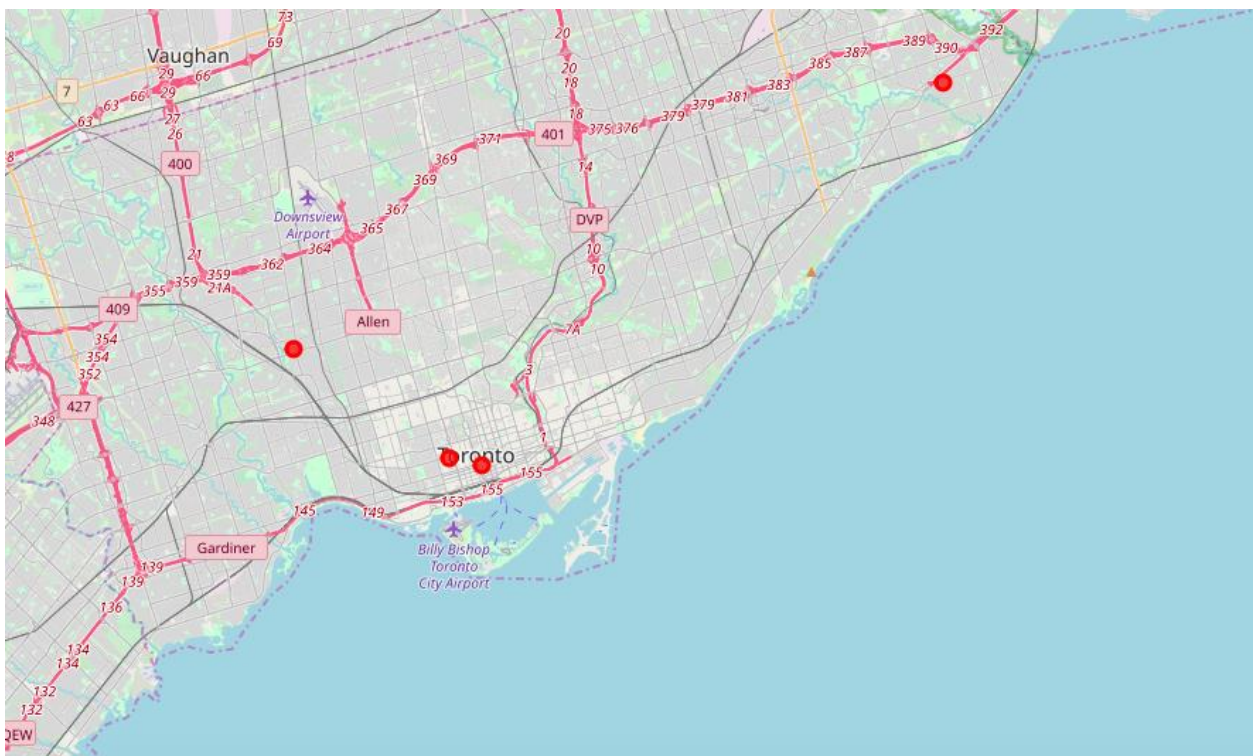
Out[92]:

	Postcode	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
1	M1C	-79.160497	0	History Museum	Bar	Women's Store	Drugstore	Diner	Discount Store	Dog Run	Doner Restaurant	Donut Shop
2	M1E	-79.188711	0	Electronics Store	Intersection	Pizza Place	Breakfast Spot	Rental Car Location	Medical Center	Mexican Restaurant	Women's Store	Dog Run
3	M1G	-79.216917	0	Coffee Shop	Korean Restaurant	Dumpling Restaurant	Diner	Discount Store	Dog Run	Doner Restaurant	Donut Shop	Drugstore
4	M1H	-79.239476	0	Bakery	Hakka Restaurant	Athletics & Sports	Bank	Thai Restaurant	Caribbean Restaurant	Fried Chicken Joint	Women's Store	Doner Restaurant
5	M1J	-79.239476	0	Playground	Construction & Landscaping	Jewelry Store	Women's Store	Drugstore	Diner	Discount Store	Dog Run	Doner Restaurant

4) Analyze each business owner's selection

We should make sure that kind of property is not 1st most common venue, but in top 4 common venues.

Take Jack as an example. When we filter the “Bar” in top 4 common venues, we find out that there are 4 candidates. After we make these four spots on the map, we can conclude that according to his budget and experience, he can only choose the M1C.



Another example is Linda's restaurant. After filtering neighborhoods, we can find there is only one candidate. The only one choice.

67	M5T	Downtown Toronto	Chinatown, Grange Park, Kensington Market	43.653206	-79.400049	0	Café	Vegetarian / Vegan Restaurant	Bar	Chinese Restaurant	Vietnamese Restaurant	Mexican Restaurant	Bakery
----	-----	------------------	---	-----------	------------	---	------	-------------------------------	-----	--------------------	-----------------------	--------------------	--------

4. Conclusion

The problem we try to solve is to optimize the site selection process. With Foursquare data and map data, by clustering different neighborhoods, and summarize common venues, we can analyze business owners request case by case.