Where to buy a house in Scarborough, Toronto

- Gloria Li

- 15-04-2019

Content

- Background
- Business problem
- Data
- Methodology
- Results
- Conclusion
- Discussion

Background

- Scarborough is a popular destination for new immigrants in Canada to reside. As a result, it is one of the most diverse and multicultural areas in the Greater Toronto Area, being home to various religious groups and places of worship. The northeast corner of Scarborough is largely rural with some of Toronto's last remaining farms, leading to Scarborough's reputation of being greener than any other part of Toronto.
- My family are considering to move to Toronto next year. After consulting several friends, we have decided to choose Scarborough as our dreaming living location.

Business problem

- There are 37 neighbourhoods in Scarborough. **Which neighbourhood is the perfect living area?**
  - I have 2 kids, one is 6 years old, the other is 4. So If there is a park or playground, it will be perfect.
  - We also look for the location where there are some restaurants, but not too many.
  - I also prefer some grocery stores nearby for continence.
  - My husband hates population density areas, so I would rather not choose some locations with shopping malls, coffee shops, airport and bus stations.
- We know little about Scarborough. I must do a deep analysis on Scarborough's neighborhoods, compare them by using Foursquare data to find what venues are near each location, making it easier to decide where we would like to buy a dreaming house.

Data

1. By performing web scrapping techniques, I will import ***Postal Codes of Toronto*** (https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M,). I will transform the data into a pandas dataframe including:
    - PostCode, borough, and neighborhood.
2. Foursquare location data, we need to get the latitude and the longitude coordinates of each neighborhood. I will use a link to a csv file that has the ***geographical coordinates*** of each postal code: http://cocl.us/Geospatial_data.
3. To explore the neighbourhood, I will use ***Foursquare API***, where I can extract the following fields: Venue Name, Venue Category, Venue Location. It will be important to understand and visualize Scarborough's features.

Data sample

Postal Codes of Toronto

Postcode	Borough	Neighbourhood
M1A	Not assigned	Not assigned
M1B	Scarborough	Rouge
M1B	Scarborough	Malvern
M1C	Scarborough	Highland Creek

Geographical coordinates

Postal Code	Latitude	Longitude
M1B	43.8066863	-79.194353
M1C	43.7845351	-79.160497
M1E	43.7635726	-79.188712
M1G	43.7709921	-79.216917

Foursquare

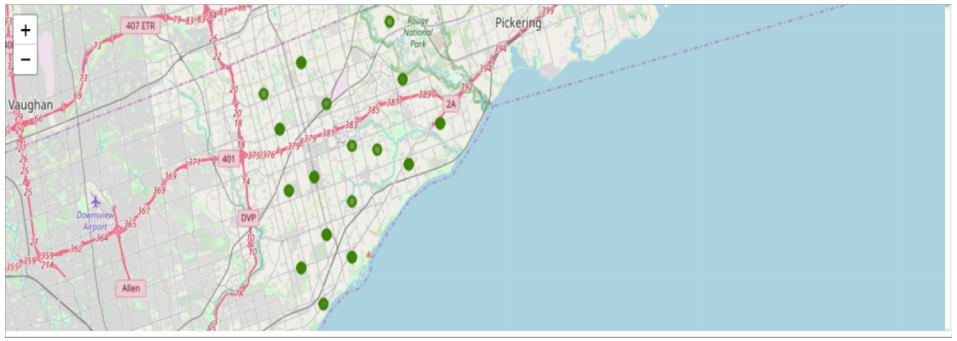
	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Rouge	43.806686	-79.194353	Wendy's	43.807448	-79.199056	Fast Food Restaurant
1	Malvern	43.806686	-79.194353	Wendy's	43.807448	-79.199056	Fast Food Restaurant
2	Highland Creek	43.784535	-79.160497	Royal Canadian Legion	43.782533	-79.163085	Bar
3	Highland Creek	43.784535	-79.160497	Scarborough Historical Society	43.788755	-79.162438	History Museum
4	Rouge Hill	43.784535	-79.160497	Royal Canadian Legion	43.782533	-79.163085	Bar

Methodology-1

1. My master data which has the main components Postcode, Borough, Neighborhood, Latitude and Longitude information of the neighborhood.

	Postcode	Borough	Neighborhood	Latitude	Longitude
0	M1B	Scarborough	Rouge	43.806686	-79.194353
1	M1B	Scarborough	Malvern	43.806686	-79.194353
2	M1C	Scarborough	Highland Creek	43.784535	-79.160497

2. Then I used folium package in Python to visualize geographic details of Scarborough and its neighborhoods. I created a map and used latitude and longitude values to get the visual as below



Methodology- Foursquare

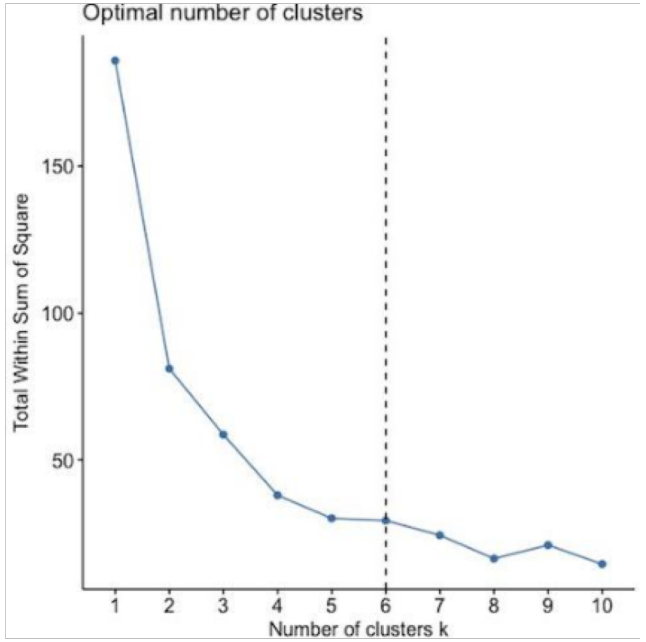
3. By using Foursquare API, I explored the neighborhoods and segmented them. I set the limit as **100 venue** and the radius **500 meter** for each neighborhoods from their given latitude and longitude information. Here is a head of the list Venues name, category, latitude and longitude information from Forsquare.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Rouge	43.806686	-79.194353	Wendy's	43.807448	-79.199056	Fast Food Restaurant
1	Malvern	43.806686	-79.194353	Wendy's	43.807448	-79.199056	Fast Food Restaurant
2	Highland Creek	43.784535	-79.160497	Royal Canadian Legion	43.782533	-79.163085	Bar
3	Highland Creek	43.784535	-79.160497	Scarborough Historical Society	43.788755	-79.162438	History Museum
4	Rouge Hill	43.784535	-79.160497	Royal Canadian Legion	43.782533	-79.163085	Bar

4. There are 52 unique categories were returned by Foursquare, then I created a table which shows list of top 5 venue category for each neighborhoods in below table.

	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	Agincourt	Sandwich Place	Breakfast Spot	Lounge	Clothing Store	Vietnamese Restaurant	Coffee Shop	Gym Pool	Grocery Store	General Entertainment	Fried Chicken Joint
1	Agincourt North	Park	Playground	Chinese Restaurant	Gym Pool	Grocery Store	General Entertainment	Fried Chicken Joint	Fast Food Restaurant	Electronics Store	Discount Store
2	Birch Cliff	General Entertainment	Skating Rink	Café	College Stadium	Vietnamese Restaurant	Clothing Store	Gym Pool	Grocery Store	Fried Chicken Joint	Fast Food Restaurant
3	Cedarbrae	Hakka Restaurant	Thai Restaurant	Athletics & Sports	Bakery	Bank	Fried Chicken Joint	Caribbean Restaurant	College Stadium	Gym Pool	Grocery Store
4	Clairlea	Bakery	Bus Line	Park	Soccer Field	Fast Food Restaurant	Bus Station	Metro Station	Coffee Shop	Grocery Store	General Entertainment

Methodology- Modeling



5. In order to identify where is the perfect location to buy a house, I need to group similar groups together. In this case, I tried clustering method: **K-means algorithm** to cluster the neighborhoods. K-Means algorithm is one of the most common cluster method of unsupervised learning.

The optimal number of clusters is 6 according to elbow method.

My merged table with clusters labels is shown as below.

	Cluster Labels	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	0	Agincourt	Sandwich Place	Breakfast Spot	Lounge	Clothing Store	Vietnamese Restaurant	Coffee Shop	Gym Pool	Grocery Store	General Entertainment	Fried Chicken Joint
1	1	Agincourt North	Park	Playground	Chinese Restaurant	Gym Pool	Grocery Store	General Entertainment	Fried Chicken Joint	Fast Food Restaurant	Electronics Store	Discount Store
2	0	Birch Cliff	General Entertainment	Skating Rink	Café	College Stadium	Vietnamese Restaurant	Clothing Store	Gym Pool	Grocery Store	Fried Chicken Joint	Fast Food Restaurant
3	0	Cedarbrae	Hakka Restaurant	Thai Restaurant	Athletics & Sports	Bakery	Bank	Fried Chicken Joint	Caribbean Restaurant	College Stadium	Gym Pool	Grocery Store
4	0	Clairlea	Bakery	Bus Line	Park	Soccer Field	Fast Food Restaurant	Bus Station	Metro Station	Coffee Shop	Grocery Store	General Entertainment

Results

Cluster 0: Restaurant and transportation centre

	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
5	Guildwood	0	Rental Car Location	Intersection	Medical Center	Pizza Place	Breakfast Spot	Electronics Store	Mexican Restaurant	Spa	Fast Food Restaurant	Clothing Store
6	Morningside	0	Rental Car Location	Intersection	Medical Center	Pizza Place	Breakfast Spot	Electronics Store	Mexican Restaurant	Spa	Fast Food Restaurant	Clothing Store
7	West Hill	0	Rental Car Location	Intersection	Medical Center	Pizza Place	Breakfast Spot	Electronics Store	Mexican Restaurant	Spa	Fast Food Restaurant	Clothing Store
9	Cedarbrae	0	Hakka Restaurant	Thai Restaurant	Athletics & Sports	Bakery	Bank	Fried Chicken Joint	Caribbean Restaurant	College Stadium	Gym Pool	Grocery Store
14	Clairlea	0	Bakery	Bus Line	Park	Soccer Field	Fast Food Restaurant	Bus Station	Metro Station	Coffee Shop	Grocery Store	General Entertainment
15	Golden Mile	0	Bakery	Bus Line	Park	Soccer Field	Fast Food Restaurant	Bus Station	Metro Station	Coffee Shop	Grocery Store	General Entertainment
16	Oakridge	0	Bakery	Bus Line	Park	Soccer Field	Fast Food Restaurant	Bus Station	Metro Station	Coffee Shop	Grocery Store	General Entertainment
20	Birch Cliff	0	General Entertainment	Skating Rink	Café	College Stadium	Vietnamese Restaurant	Clothing Store	Gym Pool	Grocery Store	Fried Chicken Joint	Fast Food Restaurant
21	Cliffside West	0	General Entertainment	Skating Rink	Café	College Stadium	Vietnamese Restaurant	Clothing Store	Gym Pool	Grocery Store	Fried Chicken Joint	Fast Food Restaurant
22	Dorset Park	0	Indian Restaurant	Vietnamese Restaurant	Latin American Restaurant	Light Rail Station	Pet Store	Chinese Restaurant	Coffee Shop	Grocery Store	General Entertainment	Fried Chicken Joint
23	Scarborough Town Centre	0	Indian Restaurant	Vietnamese Restaurant	Latin American Restaurant	Light Rail Station	Pet Store	Chinese Restaurant	Coffee Shop	Grocery Store	General Entertainment	Fried Chicken Joint
24	Wexford Heights	0	Indian Restaurant	Vietnamese Restaurant	Latin American Restaurant	Light Rail Station	Pet Store	Chinese Restaurant	Coffee Shop	Grocery Store	General Entertainment	Fried Chicken Joint
25	Maryvale	0	Auto Garage	Bakery	Smoke Shop	Sandwich Place	Breakfast Spot	Vietnamese Restaurant	Coffee Shop	Gym Pool	Grocery Store	General Entertainment
26	Wexford	0	Auto Garage	Bakery	Smoke Shop	Sandwich Place	Breakfast Spot	Vietnamese Restaurant	Coffee Shop	Gym Pool	Grocery Store	General Entertainment
27	Agincourt	0	Sandwich Place	Breakfast Spot	Lounge	Clothing Store	Vietnamese Restaurant	Coffee Shop	Gym Pool	Grocery Store	General Entertainment	Fried Chicken Joint

Results

Cluster 1: Living zone

	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
10	Scarborough Village	1	Playground	Vietnamese Restaurant	Clothing Store	Gym Pool	Grocery Store	General Entertainment	Fried Chicken Joint	Fast Food Restaurant	Electronics Store	Discount Store
31	Agincourt North	1	Park	Playground	Chinese Restaurant	Gym Pool	Grocery Store	General Entertainment	Fried Chicken Joint	Fast Food Restaurant	Electronics Store	Discount Store
32	L'Amoreaux East	1	Park	Playground	Chinese Restaurant	Gym Pool	Grocery Store	General Entertainment	Fried Chicken Joint	Fast Food Restaurant	Electronics Store	Discount Store
33	Milliken	1	Park	Playground	Chinese Restaurant	Gym Pool	Grocery Store	General Entertainment	Fried Chicken Joint	Fast Food Restaurant	Electronics Store	Discount Store
34	Steeles East	1	Park	Playground	Chinese Restaurant	Gym Pool	Grocery Store	General Entertainment	Fried Chicken Joint	Fast Food Restaurant	Electronics Store	Discount Store

Results

Cluster 2 Restaurant

	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	2	Fast Food Restaurant	Vietnamese Restaurant	Thrift / Vintage Store	Hakka Restaurant	Gym Pool	Grocery Store	General Entertainment	Fried Chicken Joint	Electronics Store	Discount Store
1	Malvern	2	Fast Food Restaurant	Vietnamese Restaurant	Thrift / Vintage Store	Hakka Restaurant	Gym Pool	Grocery Store	General Entertainment	Fried Chicken Joint	Electronics Store	Discount Store

Cluster 3 Museum

	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	Highland Creek	3	History Museum	Bar	Coffee Shop	Hakka Restaurant	Gym Pool	Grocery Store	General Entertainment	Fried Chicken Joint	Fast Food Restaurant	Electronics Store
3	Rouge Hill	3	History Museum	Bar	Coffee Shop	Hakka Restaurant	Gym Pool	Grocery Store	General Entertainment	Fried Chicken Joint	Fast Food Restaurant	Electronics Store
4	Port Union	3	History Museum	Bar	Coffee Shop	Hakka Restaurant	Gym Pool	Grocery Store	General Entertainment	Fried Chicken Joint	Fast Food Restaurant	Electronics Store

Results

Cluster 4 Shopping centre

	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
8	Woburn	4	Coffee Shop	Korean Restaurant	Vietnamese Restaurant	History Museum	Gym Pool	Grocery Store	General Entertainment	Fried Chicken Joint	Fast Food Restaurant	Electronics Store
11	East Birchmount Park	4	Discount Store	Hobby Shop	Coffee Shop	Bus Station	Department Store	Chinese Restaurant	Hakka Restaurant	Gym Pool	Grocery Store	General Entertainment
12	Ionview	4	Discount Store	Hobby Shop	Coffee Shop	Bus Station	Department Store	Chinese Restaurant	Hakka Restaurant	Gym Pool	Grocery Store	General Entertainment
13	Kennedy Park	4	Discount Store	Hobby Shop	Coffee Shop	Bus Station	Department Store	Chinese Restaurant	Hakka Restaurant	Gym Pool	Grocery Store	General Entertainment

Cluster 5 Motel/Museum

	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
17	Cliffcrest	5	American Restaurant	Motel	History Museum	Gym Pool	Grocery Store	General Entertainment	Fried Chicken Joint	Fast Food Restaurant	Electronics Store	Discount Store
18	Cliffside	5	American Restaurant	Motel	History Museum	Gym Pool	Grocery Store	General Entertainment	Fried Chicken Joint	Fast Food Restaurant	Electronics Store	Discount Store
19	Scarborough Village West	5	American Restaurant	Motel	History Museum	Gym Pool	Grocery Store	General Entertainment	Fried Chicken Joint	Fast Food Restaurant	Electronics Store	Discount Store

Conclusion

- Cluster 1 combines parks, playground, grocery store, gym and restaurant, which address my selection requirements.
- Cluster 1 includes neighborhoods such as Scarborough Village, Agincourt North, L'Amoreaux East, Milliken and Steeles East. I can look for properties within these neighborhoods with confidence.

	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
10	Scarborough Village	1	Playground	Vietnamese Restaurant	Clothing Store	Gym Pool	Grocery Store	General Entertainment	Fried Chicken Joint	Fast Food Restaurant	Electronics Store	Discount Store
31	Agincourt North	1	Park	Playground	Chinese Restaurant	Gym Pool	Grocery Store	General Entertainment	Fried Chicken Joint	Fast Food Restaurant	Electronics Store	Discount Store
32	L'Amoreaux East	1	Park	Playground	Chinese Restaurant	Gym Pool	Grocery Store	General Entertainment	Fried Chicken Joint	Fast Food Restaurant	Electronics Store	Discount Store
33	Milliken	1	Park	Playground	Chinese Restaurant	Gym Pool	Grocery Store	General Entertainment	Fried Chicken Joint	Fast Food Restaurant	Electronics Store	Discount Store
34	Steeles East	1	Park	Playground	Chinese Restaurant	Gym Pool	Grocery Store	General Entertainment	Fried Chicken Joint	Fast Food Restaurant	Electronics Store	Discount Store

Discussion

- For this assignment project, I only used the location data to get idea of neighborhood characteristics, which include neighbourhoods, coordinates of the location, venues and venues categories. To purchase a dreaming house, we need **more comprehensive data**, for example, economics index, school evaluation, population, house price, house types etc.
- Moreover, I only used k-means analysis without any comparison with other techniques. In the future, I will practice more complex data to do **more advanced analysis** such as hierarchy clustering.