

Recommendation of Promising Houses Development Area in Toronto for Real Estate Developers

1. Introduction

Toronto is the capital city of the Canadian province of Ontario. With a recorded population of 2,731,571, it is the most populous city in Canada and the fourth most populous city in North America. The city is the anchor of the Golden Horseshoe, an urban agglomeration of 9,245,438 people (as of 2016) surrounding the western end of Lake Ontario, while the Greater Toronto Area (GTA) proper had a 2016 population of 6,417,516. Toronto is an international centre of business, finance, arts, and culture, and is recognized as one of the most multicultural and cosmopolitan cities in the world.

Toronto is a prominent centre for music, theatre, motion picture production, and television production, and is home to the headquarters of Canada's major national broadcast networks and media outlets. Its varied cultural institutions, which include numerous museums and galleries, festivals and public events, entertainment districts, national historic sites, and sports activities, attract over 43 million tourists each year. Toronto is known for its many skyscrapers and high-rise buildings, in particular the tallest free-standing structure in the Western Hemisphere, the CN Tower. The city is home to the Toronto Stock Exchange, the headquarters of Canada's five largest banks, and the headquarters of many large Canadian and multinational corporations. Its economy is highly diversified with strengths in technology, design, financial services, life sciences, education, arts, fashion, aerospace, environmental innovation, food services, and tourism.

Since Toronto is the largest city in Canada and the center of economic activities, it can not be denied that the need for housing will continue to increase in the future. Peoples tend to look for a house as a place to live that is not far from their office/work in the downtown, near from public transportations and also have a comfortable and safe environment for their children. Therefore, in this assignment I will analyze the areas around Toronto that meet these criteria and it also can be used as recommendations for real estate developers to build housing in this area.

2. Data Sources

In this assignment, we will use the list of postal code of Canada from Wikipedia, we only looking for the name and postal code of the neighbours around Toronto, and apply the postal code to Google Maps Geocoding API for getting their latitude and longitude coordinates. To produce appropriate recommendation for houses development area, we will analyze the people's habits, activities and hobbies on each neighbour by exploring most visited venues on each neighbours using Foursquare API.

3. Methodology

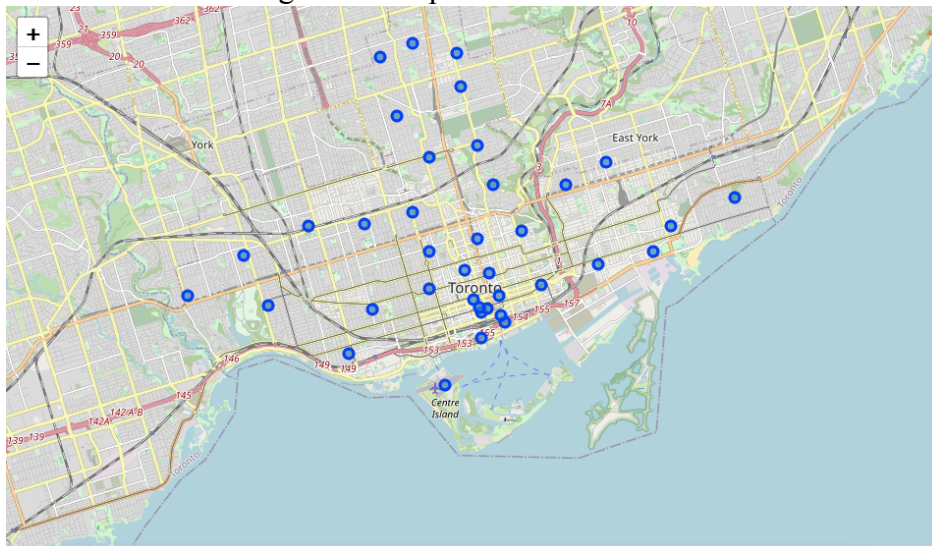
Since there is no explicit label or remarks of each neighbours regarding their peoples habit, so we will use unsupervised learning approach to cluster the neighbours based on their similar people's habit, activities and hobbies using k-means clustering.

4. Result

First we use beautiful soup for scrapping the Toronto's neighbour and its postal code information from Wikipedia, then apply it to Google Maps Geocoding API. In this step we can get the following data :

	Postal_Code	Borough	Neighborhood	Latitude	Longitude
0	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636
1	M5B	Downtown Toronto	Garden District, Ryerson	43.657162	-79.378937
2	M5C	Downtown Toronto	St. James Town	43.651494	-79.375418
3	M4E	East Toronto	The Beaches	43.676357	-79.293031
4	M5E	Downtown Toronto	Berczy Park	43.644771	-79.373306
5	M5G	Downtown Toronto	Central Bay Street	43.657952	-79.387383
6	M6G	Downtown Toronto	Christie	43.669542	-79.422564
7	M5H	Downtown Toronto	Richmond, Adelaide, King	43.650571	-79.384568
8	M6H	West Toronto	Dufferin, Dovercourt Village	43.669005	-79.442259
9	M4J	East YorkEast Toronto	The Danforth East	43.685347	-79.338106
10	M5J	Downtown Toronto	Harbourfront East, Union Station, Toronto Islands	43.640816	-79.381752
11	M6J	West Toronto	Little Portugal, Trinity	43.647927	-79.419750
12	M4K	East Toronto	The Danforth West, Riverdale	43.679557	-79.352188
13	M5K	Downtown Toronto	Toronto Dominion Centre, Design Exchange	43.647177	-79.381576
14	M6K	West Toronto	Brockton, Parkdale Village, Exhibition Place	43.636847	-79.428191
15	M4L	East Toronto	India Bazaar, The Beaches West	43.668999	-79.315572
16	M5L	Downtown Toronto	Commerce Court, Victoria Hotel	43.648198	-79.379817
17	M4M	East Toronto	Studio District	43.659526	-79.340923
18	M4N	Central Toronto	Lawrence Park	43.728020	-79.388790
19	M5N	Central Toronto	Roselawn	43.711695	-79.416936

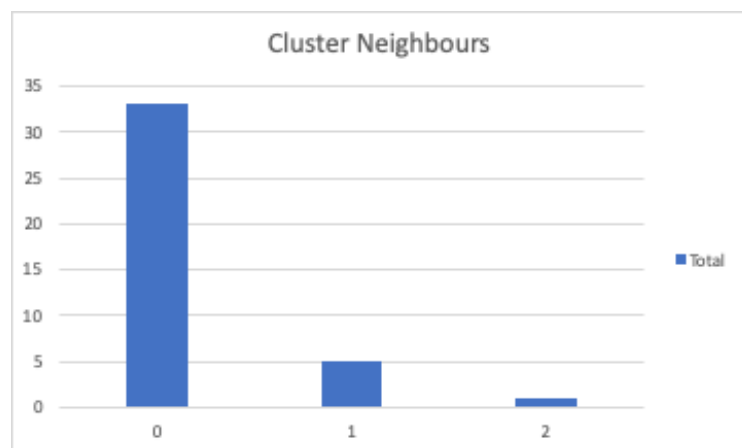
Here is Toronto's neighbours map from above data :



Then explore all venue list on each neighbours using Foursquare API and calculate the visit frequency of each venues. From this step we can recognize the most until the less visited venues.

	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	Berczy Park	Coffee Shop	Bakery	Cocktail Bar	Seafood Restaurant	Beer Bar	Farmers Market	Pharmacy	Restaurant	Cheese Shop	Jazz Club
1	Brockton, Parkdale Village, Exhibition Place	Café	Breakfast Spot	Nightclub	Bakery	Coffee Shop	Pet Store	Climbing Gym	Stadium	Burrito Place	Restaurant
2	CN Tower, King and Spadina, Railway Lands, Har...	Airport Lounge	Airport Service	Airport Terminal	Harbor / Marina	Bar	Rental Car Location	Coffee Shop	Plane	Boat or Ferry	Boutique
3	Central Bay Street	Coffee Shop	Sandwich Place	Italian Restaurant	Café	Japanese Restaurant	Department Store	Salad Place	Burger Joint	Bubble Tea Shop	Thai Restaurant
4	Christie	Grocery Store	Café	Coffee Shop	Park	Bank	Candy Store	Italian Restaurant	Baby Store	Restaurant	Nightclub

Apply k-means clustering for above data to cluster the neighbours based on similarity of most visited venues. In this assignment we will cluster the neighbours into 3 clusters (k=3). Here is the summary of the result clusters :



We get 3 cluster of neighbours, then analyze each cluster and name it.

a. Cluster 0

	Borough	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	Downtown Toronto	0	Coffee Shop	Bakery	Pub	Park	Breakfast Spot	Restaurant	Café	Theater	Yoga Studio	Dessert Shop
1	Downtown Toronto	0	Clothing Store	Coffee Shop	Bubble Tea Shop	Middle Eastern Restaurant	Café	Italian Restaurant	Cosmetics Shop	Lingerie Store	Pizza Place	Diner
2	Downtown Toronto	0	Café	Coffee Shop	Gastropub	Cosmetics Shop	Cocktail Bar	Department Store	Moroccan Restaurant	Lingerie Store	Italian Restaurant	Farmers Market
4	Downtown Toronto	0	Coffee Shop	Bakery	Cocktail Bar	Seafood Restaurant	Beer Bar	Farmers Market	Pharmacy	Restaurant	Cheese Shop	Jazz Club
5	Downtown Toronto	0	Coffee Shop	Sandwich Place	Italian Restaurant	Café	Japanese Restaurant	Department Store	Salad Place	Burger Joint	Bubble Tea Shop	Thai Restaurant
6	Downtown Toronto	0	Grocery Store	Café	Coffee Shop	Park	Bank	Candy Store	Italian Restaurant	Baby Store	Restaurant	Nightclub
7	Downtown Toronto	0	Coffee Shop	Café	Restaurant	Deli / Bodega	Clothing Store	Thai Restaurant	Hotel	Gym	Salad Place	Sushi Restaurant
8	West Toronto	0	Pharmacy	Bakery	Music Venue	Brewery	Café	Bar	Bank	Supermarket	Middle Eastern Restaurant	Brazilian Restaurant
10	Downtown Toronto	0	Coffee Shop	Aquarium	Hotel	Café	Scenic Lookout	Italian Restaurant	Fried Chicken Joint	Restaurant	Brewery	Sporting Goods Shop

From the data above, we can conclude that Cluster 0 is cluster for lifestyle neighbour because it contains neighbours with most visited venues who are related to lifestyle interest, such as coffee-shop, café, store, restaurants, etc.

b. Cluster 1

	Borough	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
3	East Toronto	1	Park	Health Food Store	Trail	Pub	Wine Bar	Dumpling Restaurant	Distribution Center	Dog Run	Doner Restaurant	Donut Shop
9	East York/East Toronto	1	Park	Convenience Store	Metro Station	Wine Bar	Distribution Center	Falafel Restaurant	Event Space	Ethiopian Restaurant	Escape Room	Electronics Store
18	Central Toronto	1	Park	Bus Line	Swim School	Wine Bar	Distribution Center	Falafel Restaurant	Event Space	Ethiopian Restaurant	Escape Room	Electronics Store
21	Central Toronto	1	Park	Trail	Jewelry Store	Sushi Restaurant	Wine Bar	Discount Store	Event Space	Ethiopian Restaurant	Escape Room	Electronics Store
33	Downtown Toronto	1	Park	Playground	Trail	Wine Bar	Diner	Event Space	Ethiopian Restaurant	Escape Room	Electronics Store	Eastern European Restaurant

From the data above, we can conclude that Cluster 1 is cluster for Family neighbour because it contains neighbours with most visited venues who are related to family interest, such as park, playground, school, public transportation, restaurants, etc.

c. Cluster 2

	Borough	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
19	Central Toronto	2	Garden	Home Service	Wine Bar	Farmers Market	Falafel Restaurant	Event Space	Ethiopian Restaurant	Escape Room	Electronics Store	Eastern European Restaurant

5. Conclusion

From the results above, we can conclude that cluster 1 is the cluster which most meet our requirements for developing houses area. So, real estate developers can focus on below area for developing new houses (purple dots) :

- Rosedale
- Forest Hill North & West
- Lawrence Park
- The Beaches
- The Danforth East

