Battle of The Neighborhoods

Madhava Narendran

1. Introduction

1.1 Business Problem:

As an Asian who is planning to move to Toronto from another city, how can I find the most Asian-friendly neighborhoods to settle down in Toronto?

1.2 Interest:

Asians who are planning to move to Toronto but are not familiar with Toronto.

2. Data acquisition

2.1 Postcode, Borough and Neighborhood information of Toronto is retrieved from Wikipedia webpage https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M. BeautifulSoup is used to do scraping and some data clean is done to get only the Postcodes, Borough and Neighborhood information.

	Postcode	Borough	Neighborhood
0	M3A	North York	Parkwoods
1	M4A	North York	Victoria Village
2	M5A	Downtown Toronto	Harbourfront , Regent Park
3	M6A	North York	Lawrence Manor, Lawrence Heights
4	M7A	Queen's Park	Queen's Park
5	M9A	Etobicoke	Islington Avenue
6	M1B	Scarborough	Malvern, Rouge
7	M3B	North York	Don Mills North
8	M4B	East York	Woodbine Gardens, Parkview Hill
9	M5B	Downtown Toronto	Ryerson, Garden District

2.2 Geo-spatial data is retrieved from http://cocl.us/Geospatial_data to get Latitude and Longitude for all the Neighborhoods in Toronto

	Postcode	Borough	Neighborhood	Latitude	Longitude
0	M3A	North York	Parkwoods	43.753259	-79.329656
1	M4A	North York	Victoria Village	43.725882	-79.315572
2	M5A	Downtown Toronto	Harbourfront , Regent Park	43.654260	-79.360636
3	M6A	North York	Lawrence Manor, Lawrence Heights	43.718518	-79.464763
4	M7A	Queen's Park	Queen's Park	43.662301	-79.389494
5	M9A	Etobicoke	Islington Avenue	43.667856	-79.532242
6	M1B	Scarborough	Malvern, Rouge	43.806686	-79.194353
7	M3B	North York	Don Mills North	43.745906	-79.352188
8	M4B	East York	Woodbine Gardens, Parkview Hill	43.706397	-79.309937
9	M5B	Downtown Toronto	Ryerson, Garden District	43.657162	-79.378937

2.3 Venue information is retrieved from FourSquare

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Parkwoods	43.753259	-79.329656	Brookbanks Park	43.751976	-79.332140	Park
1	Parkwoods	43.753259	-79.329656	PetSmart	43.748639	-79.333488	Pet Store
2	Parkwoods	43.753259	-79.329656	KFC	43.754387	-79.333021	Fast Food Restaurant
3	Parkwoods	43.753259	-79.329656	Variety Store	43.751974	-79.333114	Food & Drink Shop
4	Parkwoods	43.753259	-79.329656	Ranchdale Park	43.751388	-79.322138	Park
5	Parkwoods	43.753259	-79.329656	Joey	43.753441	-79.321640	Burger Joint
6	Parkwoods	43.753259	-79.329656	Three Valleys Park	43.751195	-79.337356	Park
7	Parkwoods	43.753259	-79.329656	Splash pad @ Brookbanks park	43.759167	-79.331294	Playground
8	Victoria Village	43.725882	-79.315572	Victoria Village Arena	43.723481	-79.315635	Hockey Arena
9	Victoria Village	43.725882	-79.315572	Tim Hortons	43.725517	-79.313103	Coffee Shop

2.4 Ratio of Asian Restaurant among all venue categories is calculated for all the neighborhoods since it is used as a factor in k-means.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Ratio
0	Adelaide, Richmond, King	43.650571	-79.384568	0.090000
1	Agincourt	43.794200	-79.262029	0.100000
2	Agincourt North, Steeles East, L'Amoreaux East...	43.815252	-79.284577	0.117647
3	Alderwood, Long Branch	43.602414	-79.543484	0.000000
4	Bayview Village	43.786947	-79.385975	0.250000

2.5 The other factor for the k-means algorithm is the Top 5 venue categories for each neighborhood.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Adelaide, Richmond, King	Asian Restaurant	Coffee Shop	Café	Bar	Steakhouse
1	Agincourt	Asian Restaurant	Skating Rink	Sandwich Place	Restaurant	Badminton Court
2	Agincourt North, Steeles East, L'Amoreaux East...	Pizza Place	Asian Restaurant	Fast Food Restaurant	Noodle House	BBQ Joint
3	Alderwood, Long Branch	Pizza Place	Convenience Store	Pharmacy	Gas Station	Skating Rink
4	Bayview Village	Asian Restaurant	Bank	Grocery Store	Café	Skating Rink

2.6 The two data frames from 2.4 and 2.5 are merged to get the final data frame for the k-means algorithm. And since k-means can't be applied to process float and string data at the same time, the categories are mapped into corresponding numbers,

Neighborhood	Ratio	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
Adelaide, Richmond, King	0.090000	120	7	25	42	106
Agincourt	0.100000	120	95	79	14	273
Agincourt North, Steeles East, L'Amoreaux East, Milliken, Ontario	0.117647	9	120	2	189	131
Alderwood, Long Branch	0.000000	9	124	46	260	95
Bayview Village	0.250000	120	45	57	25	95

3. Methodology

Use k-means algorithm to cluster the neighborhoods based on 1) ratio of Asian Restaurant among all venue categories, and 2) top 5 venue categories in each neighborhood.

4. Result

The neighborhoods in Toronto are clustered into 10 clusters. And Cluster #7 (the yellow circles shown below) has the highest ratio of Asian restaurant and the most numbers in top 5 venue categories. So neighborhoods in cluster 7 are recommended.