Coursera Capstone

IBM Applied Data Science Capstone

Finding a new home in Toronto, Canada

By: Brian Chan

January 2020

Introduction

I am a person living in Hong Kong. I am living in the Kennedy Town region, where it is very close to the underground station. The region has a balance of the eastern and western culture. It is easily accessible to the central downtown, but far enough to enjoy a quiet and relaxed environment. Recently, I have been invited by my boss to work in Toronto. The package is a nice deal and I decided to accept it. I am very excited, and at the same time very busy at the preparation work. I am looking for an apartment in Toronto which has a similar ambience compared to my current living environment. The question is, which Neighborhood should I look for?

Business Problem

The main problem is to find a neighbourhood in Toronto that exhibits the closest characteristics compared to my current home: Kennedy town. To find out the answer, we can break it down to two questions:

1. Getting the characteristics of Kennedy Town
2. Matching the characteristics of Kennedy Town to a neighborhood (or a few neighborhoods) in Toronto for consideration.

Data

We will use the Toronto Data we have prepared in week 3, and get the characteristics of Kennedy Town from FourSquare.

Toronto Data

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

Kennedy Town Data

Let’s take a look at the Kennedy Town.

Methodology

For the Toronto data, we majorly relied on Foursquare API to retrieve all venues of each neighborhoods. After that, we calculate the number of different types of venues in each neighborhoods. From the frequency of each type of venues appearing in the neighbourhood, we can infer the characteristics of each neighborhoods.

After that, we will use the K-Means clustering algorithm to infer 10 clusters in Toronto neighborhoods. This will give us an idea of 10 different neighbourhood characteristics in Toronto.

For the Kennedy Town data, we will process the Kennedy Town data in the same way, and infer the frequencies of each type of venues in Kennedy Town. We will use the K-Means clustering algorithm to predict which cluster does Kennedy Town belongs to. The neighborhoods in that cluster will exhibit a similar characteristic compared to Kennedy Town.

Analysis

Step 1: Get the Toronto Data

We first used the foursquare API to get all venues around 500 meters of each Toronto Neighborhoods. Below shows a snapshot of the data:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Rouge, Malvern	43.806686	-79.194353	Wendy's	43.807448	-79.199056	Fast Food Restaurant
1	Rouge, Malvern	43.806686	-79.194353	Interprovincial Group	43.805630	-79.200378	Print Shop
2	Highland Creek, Rouge Hill, Port Union	43.784535	-79.160497	Royal Canadian Legion	43.782533	-79.163085	Bar
3	Guildwood, Morningside, West Hill	43.763573	-79.188711	Swiss Chalet Rotisserie & Grill	43.767697	-79.189914	Pizza Place
4	Guildwood, Morningside, West Hill	43.763573	-79.188711	G & G Electronics	43.765309	-79.191537	Electronics Store

We take a look at the Neighborhoods that return most amount of venues

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Adelaide, King, Richmond	100	100	100	100	100	100
St. James Town	100	100	100	100	100	100
Ryerson, Garden District	100	100	100	100	100	100
Harbourfront East, Toronto Islands, Union Station	100	100	100	100	100	100
First Canadian Place, Underground city	100	100	100	100	100	100
Design Exchange, Toronto Dominion Centre	100	100	100	100	100	100
Commerce Court, Victoria Hotel	100	100	100	100	100	100

Let's find out how many unique categories can be curated from all the returned venues. There are **271** unique categories in total.

We then one hot encoded the data into a dataframe like following format:

	Neighborhood	Accessories Store	Afghan Restaurant	Airport	Airport Food Court	Airport Gate	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	...	Vegetarian / Vegan Restaurant	Video Game Store	Video Store
0	Adelaide, King, Richmond	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.02	...	0.02	0.0	0.0
1	Agincourt	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.00	...	0.00	0.0	0.0
2	Agincourt North, L'Amoreaux East, Milliken, St...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.00	...	0.00	0.0	0.0
3	Albion Gardens, Beaumond Heights, Humbergate, ...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.00	...	0.00	0.0	0.0
4	Alderwood, Long Branch	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.00	...	0.00	0.0	0.0

The one hot encoding technique is a technique commonly used in machine learning to preprocess categorical data.

However, the one hot encoded dataframe is quite hard for human to interpret. To make it easier to draw insights, we retrieve the top 10 most common venues for each neighbourhood.

	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	Adelaide, King, Richmond	Coffee Shop	Bar	Café	Cosmetics Shop	Steakhouse	Asian Restaurant	Thai Restaurant	Burger Joint	Restaurant	Hotel
1	Agincourt	Lounge	Latin American Restaurant	Breakfast Spot	Skating Rink	Donut Shop	Dim Sum Restaurant	Diner	Discount Store	Dog Run	Doner Restaurant
2	Agincourt North, L'Amoreaux East, Milliken, St...	Playground	Park	Yoga Studio	Donut Shop	Dessert Shop	Dim Sum Restaurant	Diner	Discount Store	Dog Run	Doner Restaurant
3	Albion Gardens, Beaumond Heights, Humbergate, ...	Grocery Store	Beer Store	Pharmacy	Pizza Place	Liquor Store	Fried Chicken Joint	Fast Food Restaurant	Sandwich Place	Japanese Restaurant	Discount Store
4	Alderwood, Long Branch	Pizza Place	Pharmacy	Skating Rink	Coffee Shop	Pool	Pub	Sandwich Place	Gym	Airport Terminal	Falafel Restaurant

Step 2: Get the Kennedy Town Data

We will use the foursquare API to get the venues around Kennedy Town. More specifically, we will get the top 100 venues within a radius of 500m around Kennedy Town.

The below table shows the first 5 records returned by the FourSquare API.

	name	categories	lat	lng
0	Winstons Coffee	Coffee Shop	22.281374	114.127172
1	Comptoir	French Restaurant	22.281209	114.126975
2	Little Creatures	Brewery	22.283950	114.128264
3	Sun Hing Restaurant (新興食家)	Dim Sum Restaurant	22.283036	114.128209
4	Catch.	Breakfast Spot	22.283152	114.126988

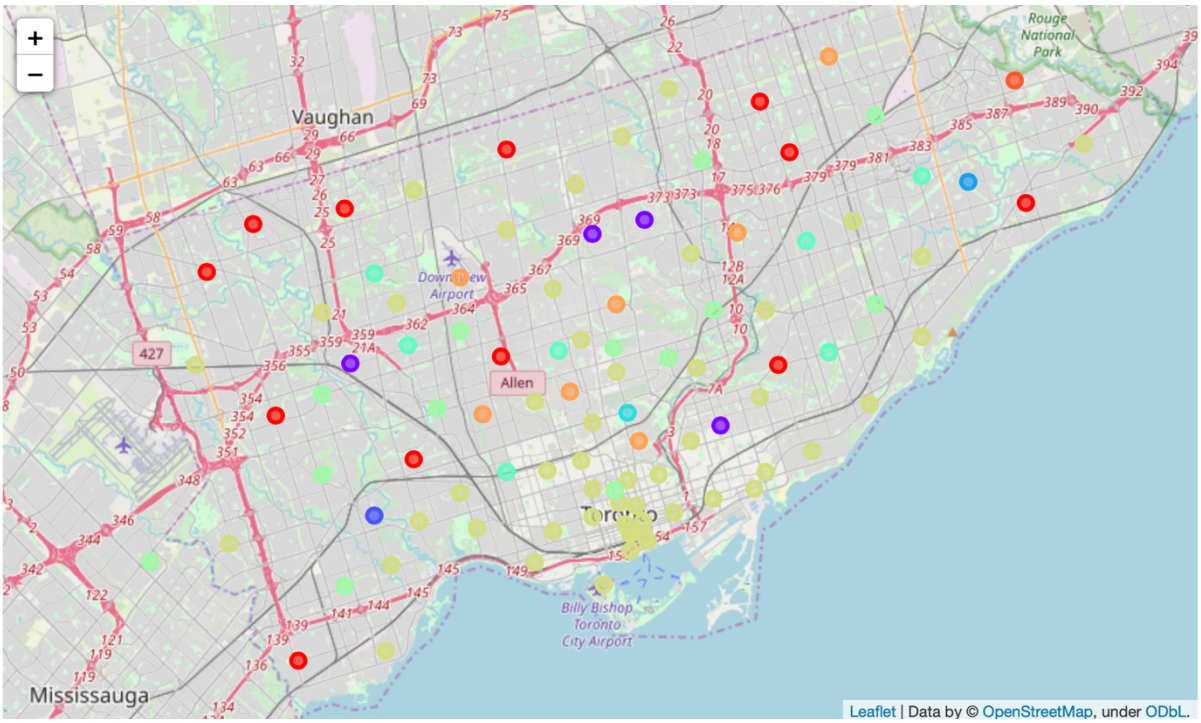
In total, 78 venues are returned by foursquare. Let's take a look at the most common venues too.

	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	Kennedy_Town	Hotpot Restaurant	Dim Sum Restaurant	Taco Place	Taiwanese Restaurant	Bus Stop	Fish & Chips Shop	Bus Station	Market	Metro Station	Noodle House

It seems that Kennedy Town is filled with a lot of restaurants!

Step 3 – Group Toronto Neighborhoods into 10 clusters.

We apply the k-means clustering algorithm to cluster the Toronto neighborhoods into 10 clusters. The map below shows the clustering results.



We see that a lot of the neighborhoods are in the yellow cluster (cluster 7).

The following table shows a snapshot of the top neighborhoods in cluster 7.

	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
1	Highland Creek, Rouge Hill, Port Union	7.0	Golf Course	Bar	Yoga Studio	Drugstore	Diner	Discount Store	Dog Run	Doner Restaurant
5	Scarborough Village	7.0	Convenience Store	Playground	Drugstore	Dim Sum Restaurant	Diner	Discount Store	Dog Run	Doner Restaurant
8	Cliffcrest, Cliffside, Scarborough Village West	7.0	American Restaurant	Motel	Skating Rink	Donut Shop	Dim Sum Restaurant	Diner	Discount Store	Dog Run
9	Birch Cliff, Cliffside West	7.0	College Stadium	Skating Rink	Café	General Entertainment	Donut Shop	Dim Sum Restaurant	Diner	Discount Store
10	Dorset Park, Scarborough Town Centre, Wexford ...	7.0	Indian Restaurant	Pet Store	Vietnamese Restaurant	Chinese Restaurant	Doner Restaurant	Dim Sum Restaurant	Diner	Discount Store
17	Hillcrest Village	7.0	Golf Course	Athletics & Sports	Pool	Mediterranean Restaurant	Dog Run	Yoga Studio	Doner Restaurant	Dim Sum Restaurant
19	Bayview Village	7.0	Japanese Restaurant	Chinese Restaurant	Bank	Café	Empanada Restaurant	Electronics Store	Eastern European Restaurant	Dumpling Restaurant
22	Willowdale South	7.0	Ramen Restaurant	Sushi Restaurant	Pizza Place	Café	Restaurant	Sandwich Place	Coffee Shop	Steakhouse
26	Don Mills North	7.0	Gym / Fitness Center	Japanese Restaurant	Caribbean Restaurant	Café	Diner	Discount Store	Dog Run	Doner Restaurant
28	Bathurst Manor, Downsview North, Wilson Heights	7.0	Coffee Shop	Pet Store	Diner	Middle Eastern Restaurant	Bank	Restaurant	Deli / Bodega	Fast Food Restaurant

Upon closer examination of these cluster, we can see a lot of restaurants. Indeed, the environment resembles what Kennedy Town is!

Results

We found that Kennedy Town belongs to Cluster 7 of the Toronto Neighborhoods. Some of the Cluster 7 Neighborhoods are:

- Highland Creek
- Rouge Hill
- Port Union
- Scarborough Village
- Cliffcrest, Cliffside,
- Scarborough Village West
- Birch Cliff
- Cliffside West

Discussion

Taking a look at the most common places in Cluster 7 neighborhoods, we see that most of them really contains a lot of restaurants: e.g. Bakery, Hakka Restaurant, Pizza Place, Caribbean Restaurant, Thai Restaurant etc.
Therefore, from this perspective, the algorithm really works as expected.

Conclusion

Most of the restaurants are concentrated in the Cluster 3 Neighborhoods in Toronto. I will start searching my new home in those Cluster 7 neighborhoods.