Capstone Project Report

Introduction

This is coursera capstone project. In this project, we will explore, segment, and cluster the neighborhoods in the city of Toronto. We use the Foursquare API to explore neighborhoods in Toronto and use the explore function to get the most common venue categories in each neighborhood, and then use this feature to group the neighborhoods into clusters by use the k-means clustering algorithm.

Background

The problem that we try to solve in this project is "if someone is looking to open a restaurant, where would we recommend that they open it?

Methodology

To solve the problem, we use the data of neighborhoods in the city of Toronto that we can obtain from “https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M”. The Toronto dataset combined with three features: PostalCode, Brough and Neighborhood as shown in figure below.

	PostalCode	Borough	Neighborhood
0	M3A	North York	Parkwoods
1	M4A	North York	Victoria Village
2	M5A	Downtown Toronto	Regent Park, Harbourfront
3	M6A	North York	Lawrence Manor, Lawrence Heights
4	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government
...	...	...	...
98	M8X	Etobicoke	The Kingsway, Montgomery Road, Old Mill North
99	M4Y	Downtown Toronto	Church and Wellesley
100	M7Y	East Toronto	Business reply mail Processing Centre, South C...
101	M8Y	Etobicoke	Old Mill South, King's Mill Park, Sunnylea, Hu...
102	M8Z	Etobicoke	Mimico NW, The Queensway West, South of Bloor,...

103 rows × 3 columns

Fig.1 Toronto dataset from Wikipidia

In order to use the Foursquare API to get the most common venue categories, we need to merge the Toronto dataset with geospatial data of Toronto (we obtain geospatial

data from csv file in Coursera, https://cocl.us/Geospatial_data) as shown in Fig.2. Then, we can use Foursquare API to get to 100 venue categories in radius 300 m of each neighborhoods in our dataset as show in Fig.3.

	PostalCode	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	Regent Park, Harbourfront	43.654260	-79.360636
3	M6A	North York	Lawrence Manor, Lawrence Heights	43.718518	-79.464763
4	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government	43.662301	-79.389494

Fig.2 Dataset after merging with geospatial data

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Regent Park, Harbourfront	43.65426	-79.360636	Roselle Desserts	43.653447	-79.362017	Bakery
1	Regent Park, Harbourfront	43.65426	-79.360636	Tandem Coffee	43.653559	-79.361809	Coffee Shop
2	Regent Park, Harbourfront	43.65426	-79.360636	Cooper Koo Family YMCA	43.653249	-79.358008	Distribution Center
3	Regent Park, Harbourfront	43.65426	-79.360636	Body Blitz Spa East	43.654735	-79.359874	Spa
4	Regent Park, Harbourfront	43.65426	-79.360636	Corktown Common	43.655618	-79.356211	Park

Fig.3 Dataset with the most common venue categories

After that, we group dataset by neighborhood to top 10 common venue of each neighborhood as shown in Fig.4. By using the top 10 common venue data with K-means clustering algorithm, we separate dataset into 5 clusters and plot each cluster on the map as shown in Fig.5.

	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	Cocktail Bar	Seafood Restaurant	Bakery	Restaurant	Cheese Shop	Café	Beer Bar	Japanese Restaurant	Hotel
1	Brockton, Parkdale Village, Exhibition Place	Café	Performing Arts Venue	Breakfast Spot	Coffee Shop	Bakery	Stadium	Burrito Place	Restaurant	Climbing Gym	Pet Store
2	Business reply mail Processing Centre, South C...	Light Rail Station	Yoga Studio	Garden Center	Skate Park	Restaurant	Recording Studio	Pizza Place	Park	Garden	Spa
3	CN Tower, King and Spadina, Railway Lands, Har...	Airport Service	Airport Lounge	Boutique	Harbor / Marina	Plane	Coffee Shop	Boat or Ferry	Sculpture Garden	Rental Car Location	Airport Terminal
4	Central Bay Street	Coffee Shop	Italian Restaurant	Japanese Restaurant	Sandwich Place	Café	Salad Place	Dessert Shop	Middle Eastern Restaurant	Thai Restaurant	Department Store

Fig.4 Top 10 common venue of each neighborhood

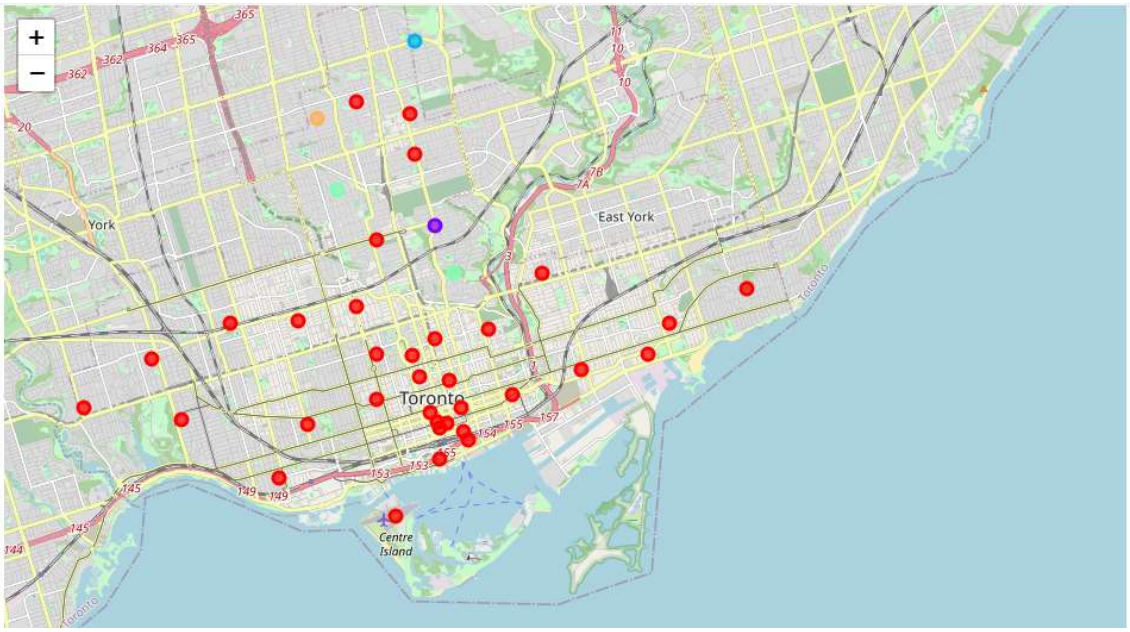

Fig.5 Map with 5 clusters of each neighborhood in different color

Result

The result from clustering shows that the most common venue category in city of Toronto is cluster 0 as shown in the map as red marker. The result in cluster 0 shows that the most common venues in the cluster are coffee shop and cafe, Next categories of common venue in the cluster are many types of restaurant and food store as shown in Fig.6.

We can suggest that if someone is looking to open a restaurant in city of Toronto, they should open their restaurant in cluster 1,2,3 or 4 neighborhoods because in cluster 0 there are too many restaurants in neighborhood.

	Borough	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
0	Downtown Toronto	Regent Park, Harbourfront	0	Coffee Shop	Pub	Bakery	Park	Breakfast Spot	Café	Theater
1	Downtown Toronto	Queen's Park, Ontario Provincial Government	0	Coffee Shop	Sushi Restaurant	Bank	Bar	Beer Bar	Smoothie Shop	Sandwich Place
2	Downtown Toronto	Garden District, Ryerson	0	Clothing Store	Coffee Shop	Cosmetics Shop	Bubble Tea Shop	Middle Eastern Restaurant	Café	Italian Restaurant
3	Downtown Toronto	St. James Town	0	Café	Coffee Shop	Cocktail Bar	American Restaurant	Gastropub	Creperie	Italian Restaurant
4	East Toronto	The Beaches	0	Trail	Health Food Store	Pub	Doner Restaurant	Dim Sum Restaurant	Diner	Discount Store
5	Downtown Toronto	Berczy Park	0	Coffee Shop	Cocktail Bar	Seafood Restaurant	Bakery	Restaurant	Cheese Shop	Café
6	Downtown	Central Bay	0	Coffee	Italian	Japanese	Sandwich	Café	Salad Place	Dessert

Fig.6 The result of clustering in cluster 0