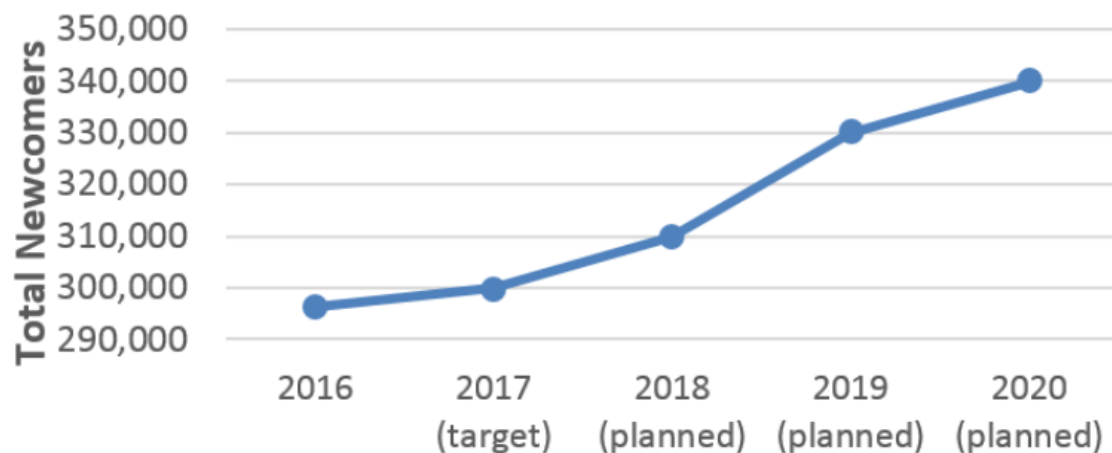


Applied Data Science Capstone Project

Cluster and Segment Neighborhoods in Toronto to Open a Restaurant

1. Introduction/Business Problem

- Immigrants to Canada has increased in the past year
- The total number of migrants is expected to increase by 340,000 by year 2020
- Locating a restaurant has always been a concern for newcomers coming with different taste and preference



- The goal of this exercise would be to use a clustering & segmentation approach to select a borough in Toronto, being a capital city to open a restaurant

2. Source and Pre-processing the Data

To solve this problem datasets will be used in combination with the Foursquare location data. Data will be used to cluster and segment neighborhoods in Toronto, Canada

- The first step in data collection is to extract the list of Toronto for which the Toronto neighborhoods dataset can be scraped online from Wikipedia which includes the Postcode, Borough, and Neighborhood name.
- Next, the Geocoder library can be used to fetch latitude and longitude coordinates for each of the neighborhoods.

- Finally, the Foursquare location API will be used to extract the list of venues surrounding each of the neighborhoods and this list, which contains venues like restaurants/gym/coffee shops/ parks, will be used to cluster and segment neighborhoods in Toronto Area. The data will be merged, and further analysis will be performed to clean and prepare it for modeling.

3. Methodology

- The first step is to import the dataframe of Canada Postal Code, Borough, Neighborhood from Wikipedia using the library beautiful soup to parse the data from Wikipedia. This dataframe would look as below and would include all the borough of Toronto

	PostalCode	Borough	Neighborhood
0	M2A	Not assigned	Not assigned
1	M3A	North York	Parkwoods
2	M4A	North York	Victoria Village
3	M5A	Downtown Toronto	Harbourfront
4	M6A	North York	Lawrence Heights

- We would then need to obtain the latitude and the longitude data for the above codes, using the geospatial dataset as 'toronto_data.csv' https://cocl.us/Geospatial_data. Once we obtain the latitude and longitude information, we would merge it with the above data frame as outlined above.

	PostalCode	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

- We then use foursquare API to obtain all the venues in the Toronto area by providing the latitude and Longitude information of Toronto. We give a radius limit of 600 m and the number of venues to be restricted to 100.

```
: # type your answer here
LIMIT = 100 # limit of number of venues returned by Foursquare API
radius = 600 # define radius
toronto_latitude = toronto_new_df['Latitude'][0]
toronto_longitude = toronto_new_df['Longitude'][0]
url = "https://api.foursquare.com/v2/venues/explore?&client_id=B2U1CZARFPRTJSOPL00IKR2T4UJQNVKPL3TVIOVMOTJQ5BKF&client_secret=KXMDGKFPHLCMM1D3RCS0QNSDN3R4UYXRJS15YAWLMAHP3XQF&v=20190501&ll= 43.653963,-79.387207&radius=200&limit=100".format(
    CLIENT_ID,
    CLIENT_SECRET,
    VERSION,
    toronto_latitude,
    toronto_longitude,
    radius,
    LIMIT)

: url

: 'https://api.foursquare.com/v2/venues/explore?&client_id=B2U1CZARFPRTJSOPL00IKR2T4UJQNVKPL3TVIOVMOTJQ5BKF&client_secret=KXMDGKFPHLCMM1D3RCS0QNSDN3R4UYXRJS15YAWLMAHP3XQF&v=20190501&ll= 43.653963,-79.387207&radius=200&limit=100'

: results = requests.get(url).json()
```

- We then obtain all the categories of the nearby venues along with the Latitude and Longitude information

	name	categories	lat	lng
0	Downtown Toronto	Neighborhood	43.653232	-79.385296
1	Cafe Plenty	Café	43.654571	-79.389450
2	Rolltation	Japanese Restaurant	43.654918	-79.387424
3	Sansotei Ramen 三草亭	Ramen Restaurant	43.655157	-79.386501
4	Fugo Desserts	Ice Cream Shop	43.654923	-79.387382

	toronto Latitude	toronto Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Places						
Adelaide, King, Richmond	100	100	100	100	100	100
Berczy Park	56	56	56	56	56	56
Brockton, Exhibition Place, Parkdale Village	22	22	22	22	22	22
Business Reply Mail Processing Centre 969 Eastern	17	17	17	17	17	17
CN Tower, Bathurst Quay, Island airport, Harbourfront West, King and Spadina, Railway Lands, South Niagara	18	18	18	18	18	18

```
print('There are {} uniques categories.'.format(len(toronto_venues['Venue Category'].unique())))
```

- There are 240 uniques categories.

To the places we add the venue and the venue category

	Places	toronto Latitude	toronto Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	The Beaches	43.676357	-79.293031	Glen Manor Ravine	43.676821	-79.293942	Trail
1	The Beaches	43.676357	-79.293031	The Big Carrot Natural Food Market	43.678879	-79.297734	Health Food Store
2	The Beaches	43.676357	-79.293031	Grover Pub and Grub	43.679181	-79.297215	Pub
3	The Beaches	43.676357	-79.293031	Upper Beaches	43.680563	-79.292869	Neighborhood
4	The Beaches	43.676357	-79.293031	Dip 'n Sip	43.678897	-79.297745	Coffee Shop

- We then identify the unique venue categories on the basis of which this clustering is performed.

	toronto Latitude	toronto Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Places						
Adelaide, King, Richmond	100	100	100	100	100	100
Berczy Park	56	56	56	56	56	56
Brockton, Exhibition Place, Parkdale Village	22	22	22	22	22	22
Business Reply Mail Processing Centre 969 Eastern	17	17	17	17	17	17
CN Tower, Bathurst Quay, Island airport, Harbourfront West, King and Spadina, Railway Lands, South Niagara	18	18	18	18	18	18

```
print('There are {} uniques categories.'.format(len(toronto_venues['Venue Category'].unique())))
```

There are 240 uniques categories.

- We then use the one hot encoding approach, where we can group places which have similar characteristics of venue categories

	Places	Afghan Restaurant	Airport	Airport Food Court	Airport Gate	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	Antique Shop	...	Thrift / Vintage Store	Toy / Game Store	Trail
0	The Beaches	0	0	0	0	0	0	0	0	0	...	0	0	1
1	The Beaches	0	0	0	0	0	0	0	0	0	...	0	0	0
2	The Beaches	0	0	0	0	0	0	0	0	0	...	0	0	0
3	The Beaches	0	0	0	0	0	0	0	0	0	...	0	0	0
4	The Beaches	0	0	0	0	0	0	0	0	0	...	0	0	0

- We then identify the frequency of venue categories for each place in toronoto

----Adelaide, King, Richmond----

	venue	freq
0	Coffee Shop	0.07
1	Café	0.05
2	Steakhouse	0.04
3	Thai Restaurant	0.04
4	Asian Restaurant	0.03

----Berczy Park----

	venue	freq
0	Coffee Shop	0.09
1	Cheese Shop	0.04
2	Bakery	0.04
3	Steakhouse	0.04
4	Farmers Market	0.04

- We then rank the 10 most common categories for each place based on the above frequencies

	Places	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	Café	Thai Restaurant	Steakhouse	Salad Place	Asian Restaurant	Burger Joint	Restaurant	Bar	Bake
1	Berczy Park	Coffee Shop	Seafood Restaurant	Beer Bar	Bakery	Steakhouse	Cocktail Bar	Cheese Shop	Farmers Market	Café	Japa Rest

- We then use K means clustering algorithm to form cluster of places and neighborhoods that have similar characteristics

Cluster Labels	Places	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
0	Adelaide, King, Richmond	Coffee Shop	Café	Thai Restaurant	Steakhouse	Salad Place	Asian Restaurant	Burger Joint	Restaurar
0	Berczy Park	Coffee Shop	Seafood Restaurant	Beer Bar	Bakery	Steakhouse	Cocktail Bar	Cheese Shop	Farmers Market

- We then use the above information and merge it to the dataframe which has the postal code and borough details

	PostalCode	Borough	Latitude	Longitude	Places	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue
0	M4E	East Toronto	43.676357	-79.293031	The Beaches	4	Neighborhood	Pub	Coffee Shop	Health Food Store
1	M4K	East Toronto	43.679557	-79.352188	The Danforth West, Riverdale	0	Greek Restaurant	Coffee Shop	Italian Restaurant	Ice Cream Shop
2	M4L	East Toronto	43.668999	-79.315572	The Beaches West, India Bazaar	0	Pet Store	Ice Cream Shop	Pizza Place	Movie Theater

- We then ascertain which cluster has the maximum count

```
0      33
1       2
4       1
3       1
2       1
```

- **Conclusion : The maximum number of clusters are in the East Tornado borough largely because of its similar venue charaterictics. Hence it would prudent to open a resturant in East Toronto bourgh in Canada.**