

Capstone Project - The Battle of Neighbourhoods

Project made for peer graded assignment for the Course Applied Data Science Capstone

By Utkarsh Sharma



1. INTRODUCTION / BUSINESS REQUIREMENT

An international Pizza Chain wants to setup their Pizza Store in Canada and they have shortlisted Toronto for their first store because Toronto is the largest city in Canada and is quite densely populated. Added advantage is to of historical monuments and frequent tourist footfall.

The Pizza Chain wants us to decide their store location.

They want us to analyse feasibility of location of store with respect to the neighbouring store in the region. They want to minimize competitors.

They want to open the store either in the main Toronto or Either in Scarborough Location(formerly called East Toronto).

2. DATA

In order to carry out this project we will be needing data from a few sources including wikipedia, Foursquare and CSV files. All the sources of data and their specifications are described below:

Wikipedia Source

We need the information of Boroughs and Neighbourhoods from the wikipedia Website. Wikipedia has a well defined table of all the details of neighbourhoods required to make the analysis of this project.

The link is here - [WIKIPEDIA LINK \(https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M\)](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M).

Foursquare API

Foursquare API will have major role in this project as this API contains all the neighbourhood details. This is a regularly updated database of the neighbourhoods. We have chosen this API because firstly it is mandated to use it, secondly it is free of cost and it provides enough calls per day to make this project feasible. The data obtained from this website is properly formatted leaving no hustle to format it. It's very intuitive.

The link to Foursquare website is [Foursquare Developer Portal \(https://foursquare.com/developers/apps\)](https://foursquare.com/developers/apps).

Geospatial Coordinates CSV File

This File has all the latitudes and longitudes stored for all the required postal codes of Canada.

The link to this file is [Geospatial_data.csv \(https://cocl.us/Geospatial_data\)](https://cocl.us/Geospatial_data)

3. Methodology

The methodology involves obtaining data from data sources, cleaning data, applying Machine Learning Algorithms and Analysis of data.

The methodology is explained below:

Importing Libraries

In [1]:

```
import pandas as pd
import numpy as np
import requests
from bs4 import BeautifulSoup
```

We have used beautiful soup to scratch data from Wikipedia Page

In [3]:

```
res = requests.get("https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M")
soup = BeautifulSoup(res.content, 'lxml')
table = soup.find_all('table')[0]
df = pd.read_html(str(table), header=0)
ff1=df[0]
ff1
ff2=ff1[ff1.Borough != 'Not assigned']
ff2=ff2.reset_index(drop=True)
ff2.loc[ff2.Neighbourhood == 'Not assigned', 'Neighbourhood'] = ff2.Borough
ff2

ff2=ff2.groupby(['Postcode', 'Borough'], as_index=False).agg(lambda x : ', '.join(x))
ff2
```

Out[3]:

	Postcode	Borough	Neighbourhood
0	M1B	Scarborough	Rouge, Malvern
1	M1C	Scarborough	Highland Creek, Rouge Hill, Port Union
2	M1E	Scarborough	Guildwood, Morningside, West Hill
3	M1G	Scarborough	Woburn
4	M1H	Scarborough	Cedarbrae
5	M1J	Scarborough	Scarborough Village
6	M1K	Scarborough	East Birchmount Park, Ionview, Kennedy Park
7	M1L	Scarborough	Clairlea, Golden Mile, Oakridge
8	M1M	Scarborough	Cliffcrest, Cliffside, Scarborough Village West
9	M1N	Scarborough	Birch Cliff, Cliffside West
10	M1P	Scarborough	Dorset Park, Scarborough Town Centre, Wexford ...
11	M1R	Scarborough	Maryvale, Wexford
12	M1S	Scarborough	Agincourt
13	M1T	Scarborough	Clarks Corners, Sullivan, Tam O'Shanter
14	M1V	Scarborough	Agincourt North, L'Amoreaux East, Milliken, St...
15	M1W	Scarborough	L'Amoreaux West, Steeles West
16	M1X	Scarborough	Upper Rouge
17	M2H	North York	Hillcrest Village
18	M2J	North York	Fairview, Henry Farm, Oriole
19	M2K	North York	Bayview Village
20	M2L	North York	Silver Hills, York Mills
21	M2M	North York	Newtonbrook, Willowdale
22	M2N	North York	Willowdale South
23	M2P	North York	York Mills West
24	M2R	North York	Willowdale West
25	M3A	North York	Parkwoods
26	M3B	North York	Don Mills North
27	M3C	North York	Flemingdon Park, Don Mills South
28	M3H	North York	Bathurst Manor, Downsview North, Wilson Heights
29	M3J	North York	Northwood Park, York University
...
73	M6C	York	Humewood-Cedarvale
74	M6E	York	Caledonia-Fairbanks

	Postcode	Borough	Neighbourhood
75	M6G	Downtown Toronto	Christie
76	M6H	West Toronto	Dovercourt Village, Dufferin
77	M6J	West Toronto	Little Portugal, Trinity
78	M6K	West Toronto	Brockton, Exhibition Place, Parkdale Village
79	M6L	North York	Maple Leaf Park, North Park, Upwood Park
80	M6M	York	Del Ray, Keelesdale, Mount Dennis, Silverthorn
81	M6N	York	The Junction North, Runnymede
82	M6P	West Toronto	High Park, The Junction South
83	M6R	West Toronto	Parkdale, Roncesvalles
84	M6S	West Toronto	Runnymede, Swansea
85	M7A	Queen's Park	Queen's Park
86	M7R	Mississauga	Canada Post Gateway Processing Centre
87	M7Y	East Toronto	Business reply mail Processing Centre969 Eastern
88	M8V	Etobicoke	Humber Bay Shores, Mimico South, New Toronto
89	M8W	Etobicoke	Alderwood, Long Branch
90	M8X	Etobicoke	The Kingsway, Montgomery Road, Old Mill North
91	M8Y	Etobicoke	Humber Bay, King's Mill Park, Kingsway Park So...
92	M8Z	Etobicoke	Kingsway Park South West, Mimico NW, The Queen...
93	M9A	Etobicoke	Islington Avenue
94	M9B	Etobicoke	Cloverdale, Islington, Martin Grove, Princess ...
95	M9C	Etobicoke	Bloordale Gardens, Eringate, Markland Wood, Ol...
96	M9L	North York	Humber Summit
97	M9M	North York	Emery, Humberlea
98	M9N	York	Weston
99	M9P	Etobicoke	Westmount
100	M9R	Etobicoke	Kingsview Village, Martin Grove Gardens, Richv...
101	M9V	Etobicoke	Albion Gardens, Beaumond Heights, Humbergate, ...
102	M9W	Etobicoke	Northwest

103 rows × 3 columns

Converted data to readable format above by making changes to the original table

Reading CSV file and then merging it with the table we obtained above so that we get all the latitudes and longitudes

In [4]:

```
latdata = pd.read_csv("https://coc1.us/Geospatial_data")

ff2=pd.merge(ff2, latdata.rename(columns={'Postal Code':'Postcode'}), on='Postcode', how='left')
ff2
```

Out[4]:

	Postcode	Borough	Neighbourhood	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
5	M1J	Scarborough	Scarborough Village	43.744734	-79.239476
6	M1K	Scarborough	East Birchmount Park, Ionview, Kennedy Park	43.727929	-79.262029
7	M1L	Scarborough	Clairlea, Golden Mile, Oakridge	43.711112	-79.284577
8	M1M	Scarborough	Cliffcrest, Cliffside, Scarborough Village West	43.716316	-79.239476
9	M1N	Scarborough	Birch Cliff, Cliffside West	43.692657	-79.264848
10	M1P	Scarborough	Dorset Park, Scarborough Town Centre, Wexford ...	43.757410	-79.273304
11	M1R	Scarborough	Maryvale, Wexford	43.750072	-79.295849
12	M1S	Scarborough	Agincourt	43.794200	-79.262029
13	M1T	Scarborough	Clarks Corners, Sullivan, Tam O'Shanter	43.781638	-79.304302
14	M1V	Scarborough	Agincourt North, L'Amoreaux East, Milliken, St...	43.815252	-79.284577
15	M1W	Scarborough	L'Amoreaux West, Steeles West	43.799525	-79.318389
16	M1X	Scarborough	Upper Rouge	43.836125	-79.205636
17	M2H	North York	Hillcrest Village	43.803762	-79.363452
18	M2J	North York	Fairview, Henry Farm, Oriole	43.778517	-79.346556
19	M2K	North York	Bayview Village	43.786947	-79.385975
20	M2L	North York	Silver Hills, York Mills	43.757490	-79.374714
21	M2M	North York	Newtonbrook, Willowdale	43.789053	-79.408493
22	M2N	North York	Willowdale South	43.770120	-79.408493
23	M2P	North York	York Mills West	43.752758	-79.400049
24	M2R	North York	Willowdale West	43.782736	-79.442259
25	M3A	North York	Parkwoods	43.753259	-79.329656
26	M3B	North York	Don Mills North	43.745906	-79.352188
27	M3C	North York	Flemingdon Park, Don Mills South	43.725900	-79.340923

	Postcode	Borough	Neighbourhood	Latitude	Longitude
28	M3H	North York	Bathurst Manor, Downsview North, Wilson Heights	43.754328	-79.442259
29	M3J	North York	Northwood Park, York University	43.767980	-79.487262
...
73	M6C	York	Humewood-Cedarvale	43.693781	-79.428191
74	M6E	York	Caledonia-Fairbanks	43.689026	-79.453512
75	M6G	Downtown Toronto	Christie	43.669542	-79.422564
76	M6H	West Toronto	Dovercourt Village, Dufferin	43.669005	-79.442259
77	M6J	West Toronto	Little Portugal, Trinity	43.647927	-79.419750
78	M6K	West Toronto	Brockton, Exhibition Place, Parkdale Village	43.636847	-79.428191
79	M6L	North York	Maple Leaf Park, North Park, Upwood Park	43.713756	-79.490074
80	M6M	York	Del Ray, Keelesdale, Mount Dennis, Silverthorn	43.691116	-79.476013
81	M6N	York	The Junction North, Runnymede	43.673185	-79.487262
82	M6P	West Toronto	High Park, The Junction South	43.661608	-79.464763
83	M6R	West Toronto	Parkdale, Roncesvalles	43.648960	-79.456325
84	M6S	West Toronto	Runnymede, Swansea	43.651571	-79.484450
85	M7A	Queen's Park	Queen's Park	43.662301	-79.389494
86	M7R	Mississauga	Canada Post Gateway Processing Centre	43.636966	-79.615819
87	M7Y	East Toronto	Business reply mail Processing Centre969 Eastern	43.662744	-79.321558
88	M8V	Etobicoke	Humber Bay Shores, Mimico South, New Toronto	43.605647	-79.501321
89	M8W	Etobicoke	Alderwood, Long Branch	43.602414	-79.543484
90	M8X	Etobicoke	The Kingsway, Montgomery Road, Old Mill North	43.653654	-79.506944
91	M8Y	Etobicoke	Humber Bay, King's Mill Park, Kingsway Park So...	43.636258	-79.498509
92	M8Z	Etobicoke	Kingsway Park South West, Mimico NW, The Queen...	43.628841	-79.520999
93	M9A	Etobicoke	Islington Avenue	43.667856	-79.532242
94	M9B	Etobicoke	Cloverdale, Islington, Martin Grove, Princess ...	43.650943	-79.554724

	Postcode	Borough	Neighbourhood	Latitude	Longitude
95	M9C	Etobicoke	Bloordale Gardens, Eringate, Markland Wood, Ol...	43.643515	-79.577201
96	M9L	North York	Humber Summit	43.756303	-79.565963
97	M9M	North York	Emery, Humberlea	43.724766	-79.532242
98	M9N	York	Weston	43.706876	-79.518188
99	M9P	Etobicoke	Westmount	43.696319	-79.532242
100	M9R	Etobicoke	Kingsview Village, Martin Grove Gardens, Richv...	43.688905	-79.554724
101	M9V	Etobicoke	Albion Gardens, Beaumont Heights, Humbergate, ...	43.739416	-79.588437
102	M9W	Etobicoke	Northwest	43.706748	-79.594054

103 rows × 5 columns

In [6]:

```
!conda install -c conda-forge geopy --yes # uncomment this line if you haven't complete
d the Foursquare API lab
from geopy.geocoders import Nominatim # convert an address into latitude and longitude
values

import requests # library to handle requests
from pandas.io.json import json_normalize # tranform JSON file into a pandas dataframe

# Matplotlib and associated plotting modules
import matplotlib.cm as cm
import matplotlib.colors as colors

# import k-means from clustering stage
from sklearn.cluster import KMeans

!conda install -c conda-forge folium=0.5.0 --yes # uncomment this line if you haven't c
ompleted the Foursquare API lab
import folium # map rendering library

print('Libraries imported.')
```

```
Fetching package metadata .....
Solving package specifications: .
```

Package plan for installation in environment /opt/conda/envs/DSX-Python35:

The following NEW packages will be INSTALLED:

```
geographiclib: 1.49-py_0    conda-forge
geopy:         1.18.1-py_0  conda-forge
```

```
geographiclib- 100% |#####| Time: 0:00:00 18.4
4 MB/s
```

```
geopy-1.18.1-p 100% |#####| Time: 0:00:00 29.2
7 MB/s
```

```
Fetching package metadata .....
Solving package specifications: .
```

Package plan for installation in environment /opt/conda/envs/DSX-Python35:

The following NEW packages will be INSTALLED:

```
altair: 2.2.2-py35_1 conda-forge
branca: 0.3.1-py_0    conda-forge
folium: 0.5.0-py_0    conda-forge
vincent: 0.4.4-py_1   conda-forge
```

```
altair-2.2.2-p 100% |#####| Time: 0:00:00 35.7
1 MB/s
```

```
branca-0.3.1-p 100% |#####| Time: 0:00:00 25.6
8 MB/s
```

```
vincent-0.4.4- 100% |#####| Time: 0:00:00 27.3
5 MB/s
```

```
folium-0.5.0-p 100% |#####| Time: 0:00:00 35.5
6 MB/s
```

```
Libraries imported.
```

In [7]:

```
address = 'Canada, CA'

geolocator = Nominatim()
location = geolocator.geocode(address)
latitude = location.latitude
longitude = location.longitude
print('The geograpical coordinate of Canada are {}, {}'.format(latitude, longitude))
```

The geographical coordinate of Canada are 61.0666922, -107.9917071.

Extracting Data Of Toronto and Scarborough(Formerly East Toronto)

In [19]:

```
toronto_data = ff2[ff2['Borough'].str.contains('Toronto')].reset_index(drop=True)
toronto_data.head()
```

Out[19]:

	Postcode	Borough	Neighbourhood	Latitude	Longitude
0	M6H	West Toronto	Dovercourt Village, Dufferin	43.669005	-79.442259
1	M6J	West Toronto	Little Portugal, Trinity	43.647927	-79.419750
2	M6K	West Toronto	Brockton, Exhibition Place, Parkdale Village	43.636847	-79.428191
3	M6P	West Toronto	High Park, The Junction South	43.661608	-79.464763
4	M6R	West Toronto	Parkdale, Roncesvalles	43.648960	-79.456325

In [32]:

```
scarborough_data = ff2[ff2['Borough'].str.contains('Scarborough')].reset_index(drop=True)
scarborough_data.head()
```

Out[32]:

	Postcode	Borough	Neighbourhood	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

In [22]:

```
saddress = 'Scarborough, CA'

geolocator = Nominatim()
location = geolocator.geocode(saddress)
slatitude = location.latitude
slongitude = location.longitude
print('The geograpical coordinate of Scarborough are {}, {}'.format(slatitude, slongitude))
```

The geograpical coordinate of Scarborough are 43.773077, -79.257774.

In [23]:

```
address = 'City of Toronto, CA'

geolocator = Nominatim()
location = geolocator.geocode(address)
latitude = location.latitude
longitude = location.longitude
print('The geograpical coordinate of Toronto are {}, {}'.format(latitude, longitude))
```

The geograpical coordinate of Toronto are 43.7170226, -79.4197830350134.

Using the Foursquare API to get neighbourhood details of the city using latitudes and longitudes

In [25]:

```
# @hidden_cell
CLIENT_ID = 'YKWGPQS4TQPABUAXESR20Y5M0NXJB0M5MFHDYETPL32SMWDA' # your Foursquare ID
CLIENT_SECRET = 'EN3MDBC2XJ12ST3BSM0C4Q3DEIEZ2F4RS3QPLADR044IYXCD' # your Foursquare Secret
VERSION = '20180605' # Foursquare API version
```

In [130]:

```
LIMIT = 500 # limit of number of venues returned by Foursquare API
radius = 500 # define radius

def getNearbyVenues(names, latitudes, longitudes, radius=500):

    venues_list=[]
    for name, lat, lng in zip(names, latitudes, longitudes):
        print(name)

        # create the API request URL
        url = 'https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret
=*&v={}&ll={},{&radius={}&limit={}'.format(
            CLIENT_ID,
            CLIENT_SECRET,
            VERSION,
            lat,
            lng,
            radius,
            LIMIT)

        # make the GET request
        results = requests.get(url).json()["response"]['groups'][0]['items']

        # return only relevant information for each nearby venue
        venues_list.append([(
            name,
            lat,
            lng,
            v['venue']['name'],
            v['venue']['location']['lat'],
            v['venue']['location']['lng'],
            v['venue']['categories'][0]['name']) for v in results])

    nearby_venues = pd.DataFrame([item for venue_list in venues_list for item in venue_
list])
    nearby_venues.columns = ['Neighborhood',
                            'Neighborhood Latitude',
                            'Neighborhood Longitude',
                            'Venue',
                            'Venue Latitude',
                            'Venue Longitude',
                            'Venue Category']

    return(nearby_venues)
```

In [131]:

```
toronto_venues = getNearbyVenues(names=toronto_data['Neighbourhood'],  
                                  latitudes=toronto_data['Latitude'],  
                                  longitudes=toronto_data['Longitude']  
                                )
```

Dovercourt Village, Dufferin
Little Portugal, Trinity
Brockton, Exhibition Place, Parkdale Village
High Park, The Junction South
Parkdale, Roncesvalles
Runnymede, Swansea

In [132]:

```
scarborough_venues = getNearbyVenues(names=scarborough_data['Neighbourhood'],  
                                       latitudes=scarborough_data['Latitude'],  
                                       longitudes=scarborough_data['Longitude']  
                                      )
```

Rouge, Malvern
Highland Creek, Rouge Hill, Port Union
Guildwood, Morningside, West Hill
Woburn
Cedarbrae
Scarborough Village
East Birchmount Park, Ionview, Kennedy Park
Clairlea, Golden Mile, Oakridge
Cliffcrest, Cliffside, Scarborough Village West
Birch Cliff, Cliffside West
Dorset Park, Scarborough Town Centre, Wexford Heights
Maryvale, Wexford
Agincourt
Clarks Corners, Sullivan, Tam O'Shanter
Agincourt North, L'Amoreaux East, Milliken, Steeles East
L'Amoreaux West, Steeles West
Upper Rouge

Venue Details

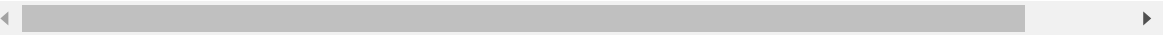
In [133]:

```
print(toronto_venues.shape)
toronto_venues.head()
```

(179, 7)

Out[133]:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	C
0	Dovercourt Village, Dufferin	43.669005	-79.442259	The Greater Good Bar	43.669409	-79.439267	Bar
1	Dovercourt Village, Dufferin	43.669005	-79.442259	Parallel	43.669516	-79.438728	Mide East Res
2	Dovercourt Village, Dufferin	43.669005	-79.442259	Planet Fitness Toronto Galleria	43.667588	-79.442574	Gym Fitne Cen
3	Dovercourt Village, Dufferin	43.669005	-79.442259	Happy Bakery & Pastries	43.667050	-79.441791	Bake
4	Dovercourt Village, Dufferin	43.669005	-79.442259	FreshCo	43.667918	-79.440754	Sup



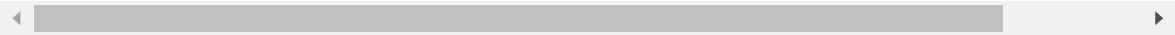
In [134]:

```
print(scarborough_venues.shape)
scarborough_venues.head()
```

(86, 7)

Out[134]:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude
0	Rouge, Malvern	43.806686	-79.194353	Wendy's	43.807448	-79.199056
1	Rouge, Malvern	43.806686	-79.194353	Interprovincial Group	43.805630	-79.200378
2	Highland Creek, Rouge Hill, Port Union	43.784535	-79.160497	RIGHT WAY TO GOLF	43.785177	-79.161108
3	Highland Creek, Rouge Hill, Port Union	43.784535	-79.160497	Royal Canadian Legion	43.782533	-79.163085
4	Guildwood, Morningside, West Hill	43.763573	-79.188711	Swiss Chalet Rotisserie & Grill	43.767697	-79.189914



In [135]:

```
print('There are {} uniques categories.'.format(len(toronto_venues['Venue Category'].unique())))
# one hot encoding
toronto_onehot = pd.get_dummies(toronto_venues[['Venue Category']], prefix="", prefix_sep="")

# add neighborhood column back to dataframe
toronto_onehot['Neighborhood'] = toronto_venues['Neighborhood']

# move neighborhood column to the first column
fixed_columns = [toronto_onehot.columns[-1]] + list(toronto_onehot.columns[:-1])
toronto_onehot = toronto_onehot[fixed_columns]

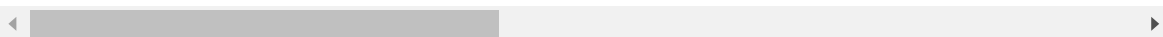
toronto_onehot.head()
```

There are 89 uniques categories.

Out[135]:

	Neighborhood	American Restaurant	Antique Shop	Art Gallery	Arts & Crafts Store	Asian Restaurant	Bakery	Bank	Ba
0	Dovercourt Village, Dufferin	0	0	0	0	0	0	0	1
1	Dovercourt Village, Dufferin	0	0	0	0	0	0	0	0
2	Dovercourt Village, Dufferin	0	0	0	0	0	0	0	0
3	Dovercourt Village, Dufferin	0	0	0	0	0	1	0	0
4	Dovercourt Village, Dufferin	0	0	0	0	0	0	0	0

5 rows × 90 columns



USING ONE HOT ENCODING

In [136]:

```
print('There are {} uniques categories.'.format(len(scarborough_venues['Venue Category']
].unique()))
# one hot encoding
scarborough_onehot = pd.get_dummies(scarborough_venues[['Venue Category']], prefix="",
prefix_sep="")

# add neighborhood column back to dataframe
scarborough_onehot['Neighborhood'] = scarborough_venues['Neighborhood']

# move neighborhood column to the first column
fixed_columns = [scarborough_onehot.columns[-1]] + list(scarborough_onehot.columns[:-1
])
scarborough_onehot = scarborough_onehot[fixed_columns]

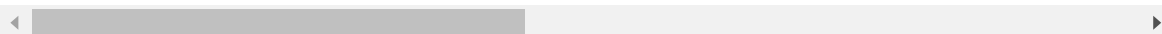
scarborough_onehot.head()
```

There are 54 uniques categories.

Out[136]:

	Neighborhood	American Restaurant	Athletics & Sports	Bakery	Bank	Bar	Breakfast Spot	Burger Joint	Bus Line	
0	Rouge, Malvern	0	0	0	0	0	0	0	0	(
1	Rouge, Malvern	0	0	0	0	0	0	0	0	(
2	Highland Creek, Rouge Hill, Port Union	0	0	0	0	0	0	0	0	(
3	Highland Creek, Rouge Hill, Port Union	0	0	0	0	1	0	0	0	(
4	Guildwood, Morningside, West Hill	0	0	0	0	0	0	0	0	(

5 rows × 55 columns



In [137]:

```
toronto_grouped = toronto_onehot.groupby('Neighborhood').mean().reset_index()
scarborough_grouped = scarborough_onehot.groupby('Neighborhood').mean().reset_index()
```

In [138]:

```
def return_most_common_venues(row, num_top_venues):  
    row_categories = row.iloc[1:]  
    row_categories_sorted = row_categories.sort_values(ascending=False)  
  
    return row_categories_sorted.index.values[0:num_top_venues]
```

Checking Most Common Venues for Toronto

In [139]:

```
num_top_venues = 10

indicators = ['st', 'nd', 'rd']

# create columns according to number of top venues
columns = ['Neighborhood']
for ind in np.arange(num_top_venues):
    try:
        columns.append('{}{} Most Common Venue'.format(ind+1, indicators[ind]))
    except:
        columns.append('{}th Most Common Venue'.format(ind+1))

# create a new dataframe
neighborhoods_venues_sorted = pd.DataFrame(columns=columns)
neighborhoods_venues_sorted['Neighborhood'] = toronto_grouped['Neighborhood']

for ind in np.arange(toronto_grouped.shape[0]):
    neighborhoods_venues_sorted.iloc[ind, 1:] = return_most_common_venues(toronto_grouped.iloc[ind, :], num_top_venues)

neighborhoods_venues_sorted
```

Out[139]:

	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
0	Brockton, Exhibition Place, Parkdale Village	Breakfast Spot	Coffee Shop	Café	Convenience Store	Italian Restaurant	Furniture Home Store
1	Dovercourt Village, Dufferin	Pharmacy	Bakery	Supermarket	Pizza Place	Fast Food Restaurant	Middle Eastern Restaurant
2	High Park, The Junction South	Mexican Restaurant	Café	Grocery Store	Speakeasy	Diner	Fast Food Restaurant
3	Little Portugal, Trinity	Bar	Men's Store	Asian Restaurant	Restaurant	Café	Coffee Shop
4	Parkdale, Roncesvalles	Breakfast Spot	Gift Shop	Dessert Shop	Coffee Shop	Restaurant	Burger Joint
5	Runnymede, Swansea	Coffee Shop	Café	Pizza Place	Sushi Restaurant	Italian Restaurant	Gym



Checking Most Common Venues for Scarborough

In [140]:

```
num_top_venues = 10

indicators = ['st', 'nd', 'rd']

# create columns according to number of top venues
columns = ['Neighborhood']
for ind in np.arange(num_top_venues):
    try:
        columns.append('{}{} Most Common Venue'.format(ind+1, indicators[ind]))
    except:
        columns.append('{}th Most Common Venue'.format(ind+1))

# create a new dataframe
neighborhoods_venues_sorted2 = pd.DataFrame(columns=columns)
neighborhoods_venues_sorted2['Neighborhood'] = scarborough_grouped['Neighborhood']

for ind in np.arange(scarborough_grouped.shape[0]):
    neighborhoods_venues_sorted2.iloc[ind, 1:] = return_most_common_venues(scarborough_
grouped.iloc[ind, :], num_top_venues)

neighborhoods_venues_sorted2
```

Out[140]:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	
0	Agincourt	Skating Rink	Sandwich Place	Breakfast Spot	Lounge	Vietnamese Restaurant	C
1	Agincourt North, L'Amoreaux East, Milliken, St...	Park	Bakery	Playground	Chinese Restaurant	Hakka Restaurant	C
2	Birch Cliff, Cliffside West	Skating Rink	General Entertainment	Café	College Stadium	Vietnamese Restaurant	C
3	Cedarbrae	Caribbean Restaurant	Thai Restaurant	Athletics & Sports	Bakery	Bank	F
4	Clairlea, Golden Mile, Oakridge	Bus Line	Bakery	Intersection	Fast Food Restaurant	Metro Station	F
5	Clarks Corners, Sullivan, Tam O'Shanter	Pizza Place	Noodle House	Shopping Mall	Pharmacy	Fast Food Restaurant	F
6	Cliffcrest, Cliffside, Scarborough Village West	American Restaurant	Motel	Coffee Shop	Hakka Restaurant	Grocery Store	C
7	Dorset Park, Scarborough Town Centre, Wexford ...	Indian Restaurant	Vietnamese Restaurant	Pet Store	Latin American Restaurant	Chinese Restaurant	E
8	East Birchmount Park, Ionview, Kennedy Park	Discount Store	Hobby Shop	Bus Station	Department Store	Train Station	C
9	Guildwood, Morningside, West Hill	Breakfast Spot	Rental Car Location	Electronics Store	Medical Center	Pizza Place	M
10	Highland Creek, Rouge Hill, Port Union	Bar	Golf Course	Vietnamese Restaurant	Coffee Shop	Hobby Shop	F
11	L'Amoreaux West, Steeles West	Fast Food Restaurant	Chinese Restaurant	Noodle House	Pizza Place	Japanese Restaurant	C

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	
12	Maryvale, Wexford	Bakery	Smoke Shop	Breakfast Spot	Middle Eastern Restaurant	Vietnamese Restaurant	(C S
13	Rouge, Malvern	Fast Food Restaurant	Print Shop	Vietnamese Restaurant	Chinese Restaurant	Hakka Restaurant	(C S
14	Scarborough Village	Playground	Vietnamese Restaurant	Chinese Restaurant	Hakka Restaurant	Grocery Store	(C (
15	Woburn	Coffee Shop	Korean Restaurant	Vietnamese Restaurant	Hakka Restaurant	Grocery Store	(C (▼

Clustering for Toronto

In [141]:

```
# set number of clusters
kclusters = 5

toronto_grouped_clustering = toronto_grouped.drop('Neighborhood', 1)

# run k-means clustering
kmeans = KMeans(n_clusters=kclusters, random_state=0).fit(toronto_grouped_clustering)

# check cluster labels generated for each row in the dataframe
#kmeans.labels_[0:10]

toronto_merged = toronto_data

# add clustering labels
toronto_merged['Cluster Labels'] = kmeans.labels_

# merge toronto_grouped with toronto_data to add latitude/longitude for each neighborhood
toronto_merged = toronto_merged.join(neighborhoods_venues_sorted.set_index('Neighborhood'), on='Neighbourhood')

#toronto_merged.head() # check the last columns!

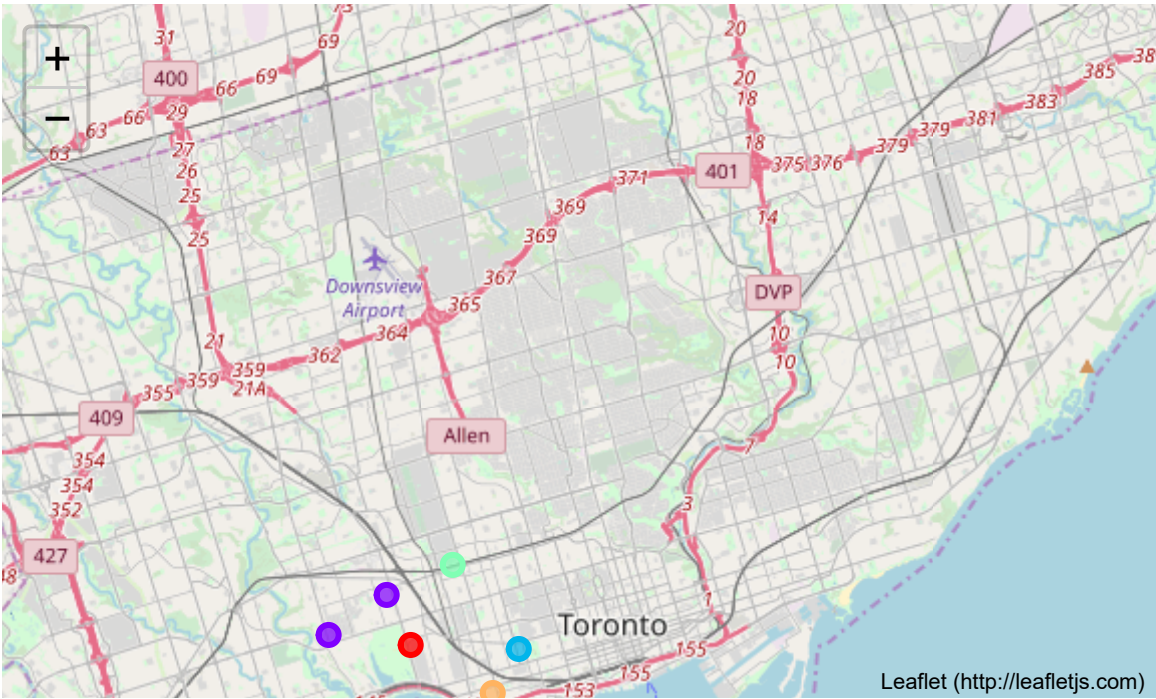
map_clusters = folium.Map(location=[latitude, longitude], zoom_start=11)

# set color scheme for the clusters
x = np.arange(kclusters)
ys = [i+x+(i*x)**2 for i in range(kclusters)]
colors_array = cm.rainbow(np.linspace(0, 1, len(ys)))
rainbow = [colors.rgb2hex(i) for i in colors_array]

# add markers to the map
markers_colors = []
for lat, lon, poi, cluster in zip(toronto_merged['Latitude'], toronto_merged['Longitude'], toronto_merged['Neighbourhood'], toronto_merged['Cluster Labels']):
    label = folium.Popup(str(poi) + ' Cluster ' + str(cluster), parse_html=True)
    folium.CircleMarker(
        [lat, lon],
        radius=5,
        popup=label,
        color=rainbow[cluster-1],
        fill=True,
        fill_color=rainbow[cluster-1],
        fill_opacity=0.7).add_to(map_clusters)

map_clusters
```


Out[141]:



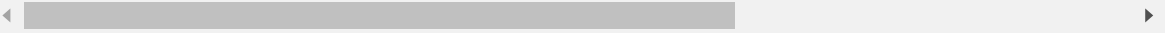
SORTED DATA FOR TORONTO

In [142]:

neighborhoods_venues_sorted

Out[142]:

	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
0	Brockton, Exhibition Place, Parkdale Village	Breakfast Spot	Coffee Shop	Café	Convenience Store	Italian Restaurant	Furniture Home Store
1	Dovercourt Village, Dufferin	Pharmacy	Bakery	Supermarket	Pizza Place	Fast Food Restaurant	Middle Eastern Restaurant
2	High Park, The Junction South	Mexican Restaurant	Café	Grocery Store	Speakeasy	Diner	Fast Food Restaurant
3	Little Portugal, Trinity	Bar	Men's Store	Asian Restaurant	Restaurant	Café	Coffee Shop
4	Parkdale, Roncesvalles	Breakfast Spot	Gift Shop	Dessert Shop	Coffee Shop	Restaurant	Burger Joint
5	Runnymede, Swansea	Coffee Shop	Café	Pizza Place	Sushi Restaurant	Italian Restaurant	Gym



SORTED DATA FOR SCARBOROUGH

In [143]:

```
neighborhoods_venues_sorted2
```

Out[143]:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	
0	Agincourt	Skating Rink	Sandwich Place	Breakfast Spot	Lounge	Vietnamese Restaurant	C
1	Agincourt North, L'Amoreaux East, Milliken, St...	Park	Bakery	Playground	Chinese Restaurant	Hakka Restaurant	C
2	Birch Cliff, Cliffside West	Skating Rink	General Entertainment	Café	College Stadium	Vietnamese Restaurant	C
3	Cedarbrae	Caribbean Restaurant	Thai Restaurant	Athletics & Sports	Bakery	Bank	F
4	Clairlea, Golden Mile, Oakridge	Bus Line	Bakery	Intersection	Fast Food Restaurant	Metro Station	F
5	Clarks Corners, Sullivan, Tam O'Shanter	Pizza Place	Noodle House	Shopping Mall	Pharmacy	Fast Food Restaurant	F
6	Cliffcrest, Cliffside, Scarborough Village West	American Restaurant	Motel	Coffee Shop	Hakka Restaurant	Grocery Store	C
7	Dorset Park, Scarborough Town Centre, Wexford ...	Indian Restaurant	Vietnamese Restaurant	Pet Store	Latin American Restaurant	Chinese Restaurant	E
8	East Birchmount Park, Ionview, Kennedy Park	Discount Store	Hobby Shop	Bus Station	Department Store	Train Station	C
9	Guildwood, Morningside, West Hill	Breakfast Spot	Rental Car Location	Electronics Store	Medical Center	Pizza Place	M
10	Highland Creek, Rouge Hill, Port Union	Bar	Golf Course	Vietnamese Restaurant	Coffee Shop	Hobby Shop	F
11	L'Amoreaux West, Steeles West	Fast Food Restaurant	Chinese Restaurant	Noodle House	Pizza Place	Japanese Restaurant	C

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	
12	Maryvale, Wexford	Bakery	Smoke Shop	Breakfast Spot	Middle Eastern Restaurant	Vietnamese Restaurant	(C S
13	Rouge, Malvern	Fast Food Restaurant	Print Shop	Vietnamese Restaurant	Chinese Restaurant	Hakka Restaurant	(C S
14	Scarborough Village	Playground	Vietnamese Restaurant	Chinese Restaurant	Hakka Restaurant	Grocery Store	(C (C
		Coffee	Korean	Vietnamese	Hakka	Grocery	(C

In [144]:

```
toronto_venues.head()
```

Out[144]:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	
0	Dovercourt Village, Dufferin	43.669005	-79.442259	The Greater Good Bar	43.669409	-79.439267	B
1	Dovercourt Village, Dufferin	43.669005	-79.442259	Parallel	43.669516	-79.438728	M E R
2	Dovercourt Village, Dufferin	43.669005	-79.442259	Planet Fitness Toronto Galleria	43.667588	-79.442574	G F C
3	Dovercourt Village, Dufferin	43.669005	-79.442259	Happy Bakery & Pastries	43.667050	-79.441791	B
4	Dovercourt Village, Dufferin	43.669005	-79.442259	FreshCo	43.667918	-79.440754	S

checking counts of pizza places in toronto

In [145]:

```
newtv=toronto_venues.groupby(["Venue Category"], as_index=False).count()
newtv2=pd.DataFrame(newtv,columns=['Venue Category','Venue'])
newtv3 = newtv2.sort_values(['Venue'], ascending=[0])
newtv3.head(10)
```

Out[145]:

	Venue Category	Venue
7	Bar	13
15	Café	11
20	Coffee Shop	10
46	Italian Restaurant	6
64	Pizza Place	6
5	Bakery	5
53	Men's Store	4
11	Breakfast Spot	4
69	Restaurant	4
63	Pharmacy	3

checking counts of pizza places in Scarborough

In [146]:

```
newsv=scarborough_venues.groupby(["Venue Category"], as_index=False).count()
news2=pd.DataFrame(newsv,columns=['Venue Category','Venue'])
news3 = news2.sort_values(['Venue'], ascending=[0])
news3.head(10)
```

Out[146]:

	Venue Category	Venue
2	Bakery	5
18	Fast Food Restaurant	5
13	Coffee Shop	4
42	Pizza Place	4
5	Breakfast Spot	4
12	Chinese Restaurant	4
46	Sandwich Place	2
41	Pharmacy	2
19	Fried Chicken Joint	2
43	Playground	2

This was the whole methodology used for the project. Final data has been made and now we will draw inferences on the findings.

4. RESULTS

- The 1st Common venue for Toronto are Breakfast, Pharmacy, Mexican Restaurant, Bars and Coffee Shops.
- The 1st Common venue for Scarborough are Skating Rink, Park, Caribbean Restaurant, Bus line, pizza place, american restaurant indian restaurant etc.
- Pizza Place count is 6 in Toronto.
- Pizza place count is 4 in Scarborough.
- Highest venue count is 13 in Toronto.
- Highest venue count is 5 in Scarborough.

5. DISCUSSION

We can clearly see that Highest venue count in Toronto is 13 whereas in Scarborough is 5. Therefore, on the basis of 1st common venue algorithm it would be unfair to give clear advantage to Toronto for Pizza place because any pizza place will not occur in 1st common venue due to presence of other stores in higher count. So our discussion now moves straightaway to the Pizza Place counts. The decision should be taken on the pizza place counts in individual regions. We can see that total Pizza place in Toronto is 6 whereas in Scarborough it is 4. It would be an advantage to set Pizza place in Scarborough clearly.

6. CONCLUSION

The best solution is to open the Pizza Place in Scarborough because there are less competitors and Pizza places are 1st most common visiting places in Scarborough.