

Capstone Gnana Week 1&2

October 23, 2019

0.1 Capstone Project - The Battle of the Neighborhoods (Week 1/2)

0.2 -----

0.3 Introduction to the Business Problem

This project is aimed at finding an optimal neighborhood location for a restaurant. **Specifically, this report is to advice a Thai restaurant chain that is interested in opening an outlet in Toronto.**

In Canada, especially in Toronto, being a food capital of Canada, Thai food is becoming very popular. Thai food has a multi ethnic appeal. It is basically enjoyed by South East Asian, Chinese and Indian community which is a sizeable population in Toronto. The chefs of this chain can modify the basic curry local spices and understand the need for western tastes and modified the level of spices and hotness. This chain has several outlets in the USA running successfully and wants to start operations in Canada. In Toronto, there are several successful Thai restaurants already and the question to be answered is with regard to the choice of location to beat existing competition and yet be profitable. That is to be able to find new catchment areas with similar characteristics as that of locations corresponding to the existing competitors.

We will use our data science techniques to generate a few most promising neighborhoods based on this criteria.

0.4 Data Section

Key data sets required to proceed with solution are:

1. Neighborhood details for Toronto. Web crawling to understand the Toronto neighbourhood preferences, demographics and its food industry. https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M, will be scraped using the BeautifulSoup package to get the basic neighborhood data (Canada Boroughs, Neighbourhood and Postal code) . This will be augmented by geographical coordinates (Latitude and Longitude) using the data in the csv file : http://cocl.us/Geospatial_data
2. Venue details that includes Thai and non Thai restaurants. For this Foursquare API Venues data will be used to pull 100 venues in each neighbourhood within a radius of 500m that will include the venue name, venue category, longitude and latitude details.

Data analysis: 1. Basic clean up of the datasets will be performed to make it suitable for running any data science analytics 2. Neighborhoods will be clustered and analyzed to understand the spending patterns and restaurant density in general. Geographical proximity and its association with the demographics of the physical locations will be analyzed which will provide key pointers

while recommending the location for Thai restaurants 3. Density and geospatial spread of the existing Thai restaurants will be analyzed on the map and superimposed on the neighborhood clusters.

```
[1]: import numpy as np # library to handle data in a vectorized manner

import pandas as pd # library for data analysis
pd.set_option('display.max_columns', None)
pd.set_option('display.max_rows', None)

import json # library to handle JSON files

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

!conda install -c anaconda beautifulsoup4 --yes

print('Libraries imported.')
```

Solving environment: done

```
==> WARNING: A newer version of conda exists. <==
  current version: 4.5.11
  latest version: 4.7.12
```

Please update conda by running

```
$ conda update -n base -c defaults conda
```

Package Plan

environment location: /home/jupyterlab/conda/envs/python

added / updated specs:

- geopy

The following packages will be downloaded:

package	build		
geopy-1.20.0	py_0	57 KB	conda-forge
geographiclib-1.50	py_0	34 KB	conda-forge
Total:		91 KB	

The following NEW packages will be INSTALLED:

geographiclib: 1.50-py_0 conda-forge
geopy: 1.20.0-py_0 conda-forge

Downloading and Extracting Packages

geopy-1.20.0 | 57 KB | ##### | 100%
geographiclib-1.50 | 34 KB | ##### | 100%

Preparing transaction: done

Verifying transaction: done

Executing transaction: done

Solving environment: done

=> WARNING: A newer version of conda exists. <==

current version: 4.5.11

latest version: 4.7.12

Please update conda by running

\$ conda update -n base -c defaults conda

All requested packages already installed.

Solving environment: done

```
==> WARNING: A newer version of conda exists. <==
current version: 4.5.11
latest version: 4.7.12
```

Please update conda by running

```
$ conda update -n base -c defaults conda
```

Package Plan

environment location: /home/jupyterlab/conda/envs/python

```
added / updated specs:
- beautifulsoup4
```

The following packages will be downloaded:

package	build		
soupsieve-1.9.3	py36_0	60 KB	anaconda
openssl-1.1.1	h7b6447c_0	5.0 MB	anaconda
certifi-2019.9.11	py36_0	154 KB	anaconda
beautifulsoup4-4.8.1	py36_0	153 KB	anaconda
Total:		5.4 MB	

The following NEW packages will be INSTALLED:

```
soupsieve:      1.9.3-py36_0      anaconda
```

The following packages will be UPDATED:

```
beautifulsoup4: 4.6.3-py37_0      --> 4.8.1-py36_0      anaconda
certifi:        2019.9.11-py36_0   conda-forge --> 2019.9.11-py36_0 anaconda
openssl:        1.1.1c-h516909a_0   conda-forge --> 1.1.1-h7b6447c_0 anaconda
```

Downloading and Extracting Packages

```
soupsieve-1.9.3      | 60 KB      | ##### | 100%
openssl-1.1.1        | 5.0 MB     | ##### | 100%
certifi-2019.9.11    | 154 KB     | ##### | 100%
beautifulsoup4-4.8.1 | 153 KB     | ##### | 100%
```

Preparing transaction: done

Verifying transaction: done

```
Executing transaction: done
Libraries imported.
```

0.5 1. Download and Explore Dataset

```
[2]: ##### Scraping wikipedia data to get the list of Canada neighbourhoods
```

```
[3]: from bs4 import BeautifulSoup
```

```
[4]: html = "https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M"

    ## use bs4 scraping content from url
    response = requests.get(html)

    html_soup = BeautifulSoup(response.text, 'html.parser')
    type(html_soup)
```

```
[4]: bs4.BeautifulSoup
```

```
[5]: pip install lxml
```

```
Collecting lxml
```

```
  Downloading https://files.pythonhosted.org/packages/ec/be/5ab8abdd8663c0
386ec2dd595a5bc0e23330a0549b8a91e32f38c20845b6/lxml-4.4.1-cp36-cp36m-manylinux1_
x86_64.whl (5.8MB)
```

```
      |                               | 5.8MB 120kB/s eta 0:00:01      |
| 1.0MB 1.9MB/s eta 0:00:03      |                               | 4.9MB 120kB/s
eta 0:00:07
```

```
Installing collected packages: lxml
```

```
Successfully installed lxml-4.4.1
```

```
Note: you may need to restart the kernel to use updated packages.
```

```
[6]: # get tables
    table = html_soup.find_all('table')[0]
    df = pd.read_html(str(table))[0]

    # get a record from table
    df.head()
```

```
[6]:
```

	Postcode	Borough	Neighbourhood
0	M1A	Not assigned	Not assigned
1	M2A	Not assigned	Not assigned
2	M3A	North York	Parkwoods
3	M4A	North York	Victoria Village
4	M5A	Downtown Toronto	Harbourfront

0.5.1 1. Cleaning the neighbourhood data

```
[7]: ## drop column which Borough is not assigned
df_bor = df[~df.Borough.isin(['Not assigned'])]

## combine same Postcode into one column
df_combine = df_bor.groupby(['Postcode', 'Borough'])['Neighbourhood'].apply(', '.
    ↪join).reset_index()

## assign same value of Borough to Neighbourhood which is not assigned
df_combine.loc[df_combine['Neighbourhood']=="Not assigned", 'Neighbourhood'] =
    ↪df_combine.loc[df_combine['Neighbourhood']=="Not assigned", 'Borough']
df_combine.head()
```

```
[7]:   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
```

```
[8]: df_combine.shape
```

```
[8]: (103, 3)
```

0.5.2 Assigning Latitude and Longitude to the neighbourhood data

```
[9]: a = pd.read_csv('https://cocl.us/Geospatial_data')
gdf = pd.DataFrame(a)
```

```
[10]: ## assign lat, long to postcode dataframe
df_post = pd.
    ↪merge(df_combine, gdf, how='left', left_on='Postcode', right_on='Postal Code')
df_post.drop('Postal Code', axis=1, inplace=True)
df_post.head()
```

```
[10]:   Postcode   Borough   Neighbourhood  Latitude \
0      M1B  Scarborough  Rouge, Malvern  43.806686
1      M1C  Scarborough  Highland Creek, Rouge Hill, Port Union  43.784535
2      M1E  Scarborough  Guildwood, Morningside, West Hill  43.763573
3      M1G  Scarborough  Woburn  43.770992
4      M1H  Scarborough  Cedarbrae  43.773136

      Longitude
0 -79.194353
1 -79.160497
2 -79.188711
```

```
3 -79.216917
4 -79.239476
```

```
[11]: df_post.shape
```

```
[11]: (103, 5)
```

```
[12]: address = 'Toronto, ON'

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

The geograpical coordinate of Toronto City are 43.653963, -79.387207.

```
[13]: # create map of Toronto using latitude and longitude values
map_toronto = folium.Map(location=[latitude, longitude], zoom_start=10)

# add markers to map
for lat, lng, borough, neighborhood in zip(df_post['Latitude'],
    ↪df_post['Longitude'], df_post['Borough'], df_post['Neighbourhood']):
    label = '{} {}'.format(neighborhood, borough)
    label = folium.Popup(label, parse_html=True)
    folium.CircleMarker(
        [lat, lng],
        radius=5,
        popup=label,
        color='blue',
        fill=True,
        fill_color='#3186cc',
        fill_opacity=0.7,
        parse_html=False).add_to(map_toronto)

map_toronto
```

```
[13]: <folium.folium.Map at 0x7f29575e4d68>
```

```
[14]: #We are going to work with only the boroughs that contain the word "Toronto".

df_t = df_post[df_post['Borough'].str.contains("Toronto")].
    ↪reset_index(drop=True)
df_t.head()
```

```
[14]:
```

	Postcode	Borough	Neighbourhood	Latitude	\
0	M4E	East Toronto	The Beaches	43.676357	
1	M4K	East Toronto	The Danforth West, Riverdale	43.679557	
2	M4L	East Toronto	The Beaches West, India Bazaar	43.668999	
3	M4M	East Toronto	Studio District	43.659526	
4	M4N	Central Toronto	Lawrence Park	43.728020	


```

Longitude
0 -79.293031
1 -79.352188
2 -79.315572
3 -79.340923
4 -79.388790

```

Define Foursquare Credentials and Version

```
[15]: CLIENT_ID = 'YSH1ABHBNWGMJW4T5LP2ETHXS40D5TUBWJ1JFE4MLET2SITO' # your
      ↪ Foursquare ID
CLIENT_SECRET = 'GHCI4N3ONG5XOMAG3ETKE4DCOOKWNOYHAJOPI1XI1BM2BJAN' # your
      ↪ Foursquare Secret
VERSION = '20180605' # Foursquare API version

print('Your credentails:')
print('CLIENT_ID: ' + CLIENT_ID)
print('CLIENT_SECRET: ' + CLIENT_SECRET)
```

Your credentails:

CLIENT_ID: YSH1ABHBNWGMJW4T5LP2ETHXS40D5TUBWJ1JFE4MLET2SITO

CLIENT_SECRET: GHCI4N3ONG5XOMAG3ETKE4DCOOKWNOYHAJOPI1XI1BM2BJAN

Let's explore the first neighborhood in our dataframe

```
[16]: df_t.loc[0, 'Neighbourhood']
```

```
[16]: 'The Beaches'
```

Get the neighborhood's latitude and longitude values.

```
[17]: neighborhood_latitude = df_t.loc[0, 'Latitude'] # neighborhood latitude value
      neighborhood_longitude = df_t.loc[0, 'Longitude'] # neighborhood longitude value

      neighborhood_name = df_t.loc[0, 'Neighbourhood'] # neighborhood name

print('Latitude and longitude values of {} are {}, {}.'.
      ↪ format(neighborhood_name,
              ↪ neighborhood_latitude,
```



```
↪neighborhood_longitude))
```

Latitude and longitude values of The Beaches are 43.67635739999999, -79.2930312.

Now, let's get the top 100 venues that are in The Beaches within a radius of 500 meters.

```
[18]: LIMIT = 100 # limit of number of venues returned by Foursquare API

radius = 500 # define radius

# create URL
url = 'https://api.foursquare.com/v2/venues/explore?
↪&client_id={} &client_secret={} &v={} &ll={}, {} &radius={} &limit={}'.format(
    CLIENT_ID,
    CLIENT_SECRET,
    VERSION,
    neighborhood_latitude,
    neighborhood_longitude,
    radius,
    LIMIT)
url # display URL
```

```
[18]: 'https://api.foursquare.com/v2/venues/explore?&client_id=YSH1ABHBNWGMJW4T5LP2ETH
XS40D5TUBWJ1JFE4MLET2SIT0&client_secret=GHCI4N3ONG5XOMAG3ETKE4DCOOKWNOYHAJOPI1XI
1BM2BJAN&v=20180605&ll=43.67635739999999,-79.2930312&radius=500&limit=100'
```

```
[19]: results = requests.get(url).json()
```

```
[20]: # function that extracts the category of the venue
def get_category_type(row):
    try:
        categories_list = row['categories']
    except:
        categories_list = row['venue.categories']

    if len(categories_list) == 0:
        return None
    else:
        return categories_list[0]['name']

venues = results['response']['groups'][0]['items']

nearby_venues = json_normalize(venues) # flatten JSON

# filter columns
```

```

filtered_columns = ['venue.name', 'venue.categories', 'venue.location.lat',
    ↳ 'venue.location.lng']
nearby_venues = nearby_venues.loc[:, filtered_columns]

# filter the category for each row
nearby_venues['venue.categories'] = nearby_venues.apply(get_category_type,
    ↳ axis=1)

# clean columns
nearby_venues.columns = [col.split(".")[1] for col in nearby_venues.columns]

nearby_venues.head()

```

```

[20]:

```

	name	categories	lat	lng
0	Glen Manor Ravine	Trail	43.676821	-79.293942
1	The Big Carrot Natural Food Market	Health Food Store	43.678879	-79.297734
2	Grover Pub and Grub	Pub	43.679181	-79.297215
3	Upper Beaches	Neighborhood	43.680563	-79.292869

```

[21]: print('{} venues were returned by Foursquare.'.format(nearby_venues.shape[0]))

```

4 venues were returned by Foursquare.

0.6 2. Exploring Neighborhoods in Toronto

```

[22]: #function to repeat the same process to all the neighborhoods in Toronto
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)

```

```

[23]: # Using the above function on each neighborhood to create a new dataframe_
    ↪called toronto_venues
toronto_venues = getNearbyVenues(names=df_t['Neighbourhood'],
                                latitudes=df_t['Latitude'],
                                longitudes=df_t['Longitude']
                                )

print(toronto_venues.shape)
toronto_venues.head()

```

```

The Beaches
The Danforth West,Riverdale
The Beaches West,India Bazaar
Studio District
Lawrence Park
Davisville North
North Toronto West
Davisville
Moore Park,Summerhill East
Deer Park,Forest Hill SE,Rathnelly,South Hill,Summerhill West
Rosedale
Cabbagetown,St. James Town
Church and Wellesley
Harbourfront,Regent Park
Ryerson,Garden District
St. James Town
Berczy Park

```

Central Bay Street
 Adelaide,King,Richmond
 Harbourfront East,Toronto Islands,Union Station
 Design Exchange,Toronto Dominion Centre
 Commerce Court,Victoria Hotel
 Roselawn
 Forest Hill North,Forest Hill West
 The Annex,North Midtown,Yorkville
 Harbord,University of Toronto
 Chinatown,Grange Park,Kensington Market
 CN Tower,Bathurst Quay,Island airport,Harbourfront West,King and Spadina,Railway
 Lands,South Niagara
 Stn A PO Boxes 25 The Esplanade
 First Canadian Place,Underground city
 Christie
 Dovercourt Village,Dufferin
 Little Portugal,Trinity
 Brockton,Exhibition Place,Parkdale Village
 High Park,The Junction South
 Parkdale,Roncesvalles
 Runnymede,Swansea
 Business Reply Mail Processing Centre 969 Eastern
 (1715, 7)

```
[23]:
```

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude \
0	The Beaches	43.676357	-79.293031
1	The Beaches	43.676357	-79.293031
2	The Beaches	43.676357	-79.293031
3	The Beaches	43.676357	-79.293031
4	The Danforth West,Riverdale	43.679557	-79.352188

	Venue	Venue Latitude	Venue Longitude \
0	Glen Manor Ravine	43.676821	-79.293942
1	The Big Carrot Natural Food Market	43.678879	-79.297734
2	Grover Pub and Grub	43.679181	-79.297215
3	Upper Beaches	43.680563	-79.292869
4	Pantheon	43.677621	-79.351434

	Venue Category
0	Trail
1	Health Food Store
2	Pub
3	Neighborhood
4	Greek Restaurant

```
[24]: #Let's check the size of the resulting dataframe
print(toronto_venues.shape)
```

```
toronto_venues.head()
```

```
(1715, 7)
```

```
[24]:
```

	Neighborhood	Neighborhood	Latitude	Neighborhood	Longitude	\
0		The Beaches	43.676357		-79.293031	
1		The Beaches	43.676357		-79.293031	
2		The Beaches	43.676357		-79.293031	
3		The Beaches	43.676357		-79.293031	
4	The Danforth West,	Riverdale	43.679557		-79.352188	

		Venue	Venue	Latitude	Venue	Longitude	\
0		Glen Manor Ravine		43.676821		-79.293942	
1	The Big Carrot	Natural Food Market		43.678879		-79.297734	
2		Grover Pub and Grub		43.679181		-79.297215	
3		Upper Beaches		43.680563		-79.292869	
4		Pantheon		43.677621		-79.351434	

	Venue Category
0	Trail
1	Health Food Store
2	Pub
3	Neighborhood
4	Greek Restaurant

```
[25]: #Let's check how many venues were returned for each neighborhood
```

```
toronto_venues.groupby('Neighborhood').count()
```

```
[25]:
```

	Neighborhood	Latitude	\
	Neighborhood		
	Adelaide,King,Richmond		100
	Berczy Park		56
	Brockton,Exhibition Place,Parkdale Village		23
	Business Reply Mail Processing Centre 969 Eastern		18
	CN Tower,Bathurst Quay,Island airport,Harbourfr...		16
	Cabbagetown,St. James Town		42
	Central Bay Street		87
	Chinatown,Grange Park,Kensington Market		100
	Christie		17
	Church and Wellesley		87
	Commerce Court,Victoria Hotel		100
	Davisville		36
	Davisville North		8
	Deer Park,Forest Hill SE,Rathnelly,South Hill,S...		15
	Design Exchange,Toronto Dominion Centre		100
	Dovercourt Village,Dufferin		15

First Canadian Place,Underground city	100
Forest Hill North,Forest Hill West	4
Harbord,University of Toronto	36
Harbourfront East,Toronto Islands,Union Station	100
Harbourfront,Regent Park	52
High Park,The Junction South	24
Lawrence Park	4
Little Portugal,Trinity	64
Moore Park,Summerhill East	4
North Toronto West	25
Parkdale,Roncesvalles	15
Rosedale	4
Roselawn	1
Runnymede,Swansea	37
Ryerson,Garden District	100
St. James Town	100
Stn A PO Boxes 25 The Esplanade	97
Studio District	39
The Annex,North Midtown,Yorkville	22
The Beaches	4
The Beaches West,India Bazaar	21
The Danforth West,Riverdale	42

	Neighborhood Longitude \
Neighborhood	
Adelaide,King,Richmond	100
Berczy Park	56
Brockton,Exhibition Place,Parkdale Village	23
Business Reply Mail Processing Centre 969 Eastern	18
CN Tower,Bathurst Quay,Island airport,Harbourfr...	16
Cabbagetown,St. James Town	42
Central Bay Street	87
Chinatown,Grange Park,Kensington Market	100
Christie	17
Church and Wellesley	87
Commerce Court,Victoria Hotel	100
Davisville	36
Davisville North	8
Deer Park,Forest Hill SE,Rathnelly,South Hill,S...	15
Design Exchange,Toronto Dominion Centre	100
Dovercourt Village,Dufferin	15
First Canadian Place,Underground city	100
Forest Hill North,Forest Hill West	4
Harbord,University of Toronto	36
Harbourfront East,Toronto Islands,Union Station	100
Harbourfront,Regent Park	52
High Park,The Junction South	24

Lawrence Park	4
Little Portugal,Trinity	64
Moore Park,Summerhill East	4
North Toronto West	25
Parkdale,Roncesvalles	15
Rosedale	4
Roselawn	1
Runnymede,Swansea	37
Ryerson,Garden District	100
St. James Town	100
Stn A PO Boxes 25 The Esplanade	97
Studio District	39
The Annex,North Midtown,Yorkville	22
The Beaches	4
The Beaches West,India Bazaar	21
The Danforth West,Riverdale	42

	Venue	Venue Latitude \
Neighborhood		
Adelaide,King,Richmond	100	100
Berczy Park	56	56
Brockton,Exhibition Place,Parkdale Village	23	23
Business Reply Mail Processing Centre 969 Eastern	18	18
CN Tower,Bathurst Quay,Island airport,Harbourfr...	16	16
Cabbagetown,St. James Town	42	42
Central Bay Street	87	87
Chinatown,Grange Park,Kensington Market	100	100
Christie	17	17
Church and Wellesley	87	87
Commerce Court,Victoria Hotel	100	100
Davisville	36	36
Davisville North	8	8
Deer Park,Forest Hill SE,Rathnelly,South Hill,S...	15	15
Design Exchange,Toronto Dominion Centre	100	100
Dovercourt Village,Dufferin	15	15
First Canadian Place,Underground city	100	100
Forest Hill North,Forest Hill West	4	4
Harbord,University of Toronto	36	36
Harbourfront East,Toronto Islands,Union Station	100	100
Harbourfront,Regent Park	52	52
High Park,The Junction South	24	24
Lawrence Park	4	4
Little Portugal,Trinity	64	64
Moore Park,Summerhill East	4	4
North Toronto West	25	25
Parkdale,Roncesvalles	15	15
Rosedale	4	4

Roselawn	1	1
Runnymede, Swansea	37	37
Ryerson, Garden District	100	100
St. James Town	100	100
Stn A PO Boxes 25 The Esplanade	97	97
Studio District	39	39
The Annex, North Midtown, Yorkville	22	22
The Beaches	4	4
The Beaches West, India Bazaar	21	21
The Danforth West, Riverdale	42	42

	Venue Longitude	\
Neighborhood		
Adelaide, King, Richmond	100	
Berczy Park	56	
Brockton, Exhibition Place, Parkdale Village	23	
Business Reply Mail Processing Centre 969 Eastern	18	
CN Tower, Bathurst Quay, Island airport, Harbourfr...	16	
Cabbagetown, St. James Town	42	
Central Bay Street	87	
Chinatown, Grange Park, Kensington Market	100	
Christie	17	
Church and Wellesley	87	
Commerce Court, Victoria Hotel	100	
Davisville	36	
Davisville North	8	
Deer Park, Forest Hill SE, Rathnelly, South Hill, S...	15	
Design Exchange, Toronto Dominion Centre	100	
Dovercourt Village, Dufferin	15	
First Canadian Place, Underground city	100	
Forest Hill North, Forest Hill West	4	
Harbord, University of Toronto	36	
Harbourfront East, Toronto Islands, Union Station	100	
Harbourfront, Regent Park	52	
High Park, The Junction South	24	
Lawrence Park	4	
Little Portugal, Trinity	64	
Moore Park, Summerhill East	4	
North Toronto West	25	
Parkdale, Roncesvalles	15	
Rosedale	4	
Roselawn	1	
Runnymede, Swansea	37	
Ryerson, Garden District	100	
St. James Town	100	
Stn A PO Boxes 25 The Esplanade	97	
Studio District	39	

The Annex,North Midtown,Yorkville	22
The Beaches	4
The Beaches West,India Bazaar	21
The Danforth West,Riverdale	42

	Venue Category
Neighborhood	
Adelaide,King,Richmond	100
Berczy Park	56
Brockton,Exhibition Place,Parkdale Village	23
Business Reply Mail Processing Centre 969 Eastern	18
CN Tower,Bathurst Quay,Island airport,Harbourfr...	16
Cabbagetown,St. James Town	42
Central Bay Street	87
Chinatown,Grange Park,Kensington Market	100
Christie	17
Church and Wellesley	87
Commerce Court,Victoria Hotel	100
Davisville	36
Davisville North	8
Deer Park,Forest Hill SE,Rathnelly,South Hill,S...	15
Design Exchange,Toronto Dominion Centre	100
Dovercourt Village,Dufferin	15
First Canadian Place,Underground city	100
Forest Hill North,Forest Hill West	4
Harbord,University of Toronto	36
Harbourfront East,Toronto Islands,Union Station	100
Harbourfront,Regent Park	52
High Park,The Junction South	24
Lawrence Park	4
Little Portugal,Trinity	64
Moore Park,Summerhill East	4
North Toronto West	25
Parkdale,Roncesvalles	15
Rosedale	4
Roselawn	1
Runnymede,Swansea	37
Ryerson,Garden District	100
St. James Town	100
Stn A PO Boxes 25 The Esplanade	97
Studio District	39
The Annex,North Midtown,Yorkville	22
The Beaches	4
The Beaches West,India Bazaar	21
The Danforth West,Riverdale	42

```
[26]: #Unique venue categories in Toronto

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

There are 237 uniques categories.

0.7 3. Analyze Each Neighborhood

```
[27]: # one hot encoding
toronto_onehot = pd.get_dummies(toronto_venues[['Venue Category']], prefix="",
↪prefix_sep="")

# add neighborhood column back to dataframe
toronto_onehot['Neighbourhood'] = 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()
```

```
[27]:
```

	Neighbourhood	Afghan Restaurant	Airport	\
0	The Beaches	0	0	
1	The Beaches	0	0	
2	The Beaches	0	0	
3	The Beaches	0	0	
4	The Danforth West,Riverdale	0	0	

	Airport Food Court	Airport Gate	Airport Lounge	Airport Service	\
0	0	0	0	0	
1	0	0	0	0	
2	0	0	0	0	
3	0	0	0	0	
4	0	0	0	0	

	Airport Terminal	American Restaurant	Antique Shop	Aquarium	Art Gallery	\
0	0	0	0	0	0	
1	0	0	0	0	0	
2	0	0	0	0	0	
3	0	0	0	0	0	
4	0	0	0	0	0	

	Art Museum	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	\
0	0	0	0	0	
1	0	0	0	0	
2	0	0	0	0	

3	0	0	0	0
4	0	0	0	0

	Auto Workshop	BBQ Joint	Baby Store	Bagel Shop	Bakery	Bank	Bar	\
0	0	0	0	0	0	0	0	
1	0	0	0	0	0	0	0	
2	0	0	0	0	0	0	0	
3	0	0	0	0	0	0	0	
4	0	0	0	0	0	0	0	

	Baseball Stadium	Basketball Stadium	Beach	Bed & Breakfast	Beer Bar	\
0	0	0	0	0	0	
1	0	0	0	0	0	
2	0	0	0	0	0	
3	0	0	0	0	0	
4	0	0	0	0	0	

	Beer Store	Belgian Restaurant	Bistro	Board Shop	Boat or Ferry	\
0	0	0	0	0	0	
1	0	0	0	0	0	
2	0	0	0	0	0	
3	0	0	0	0	0	
4	0	0	0	0	0	

	Bookstore	Boutique	Brazilian Restaurant	Breakfast Spot	Brewery	\
0	0	0	0	0	0	
1	0	0	0	0	0	
2	0	0	0	0	0	
3	0	0	0	0	0	
4	0	0	0	0	0	

	Bubble Tea Shop	Building	Burger Joint	Burrito Place	Bus Line	Butcher	\
0	0	0	0	0	0	0	
1	0	0	0	0	0	0	
2	0	0	0	0	0	0	
3	0	0	0	0	0	0	
4	0	0	0	0	0	0	

	Café	Cajun / Creole Restaurant	Camera Store	Candy Store	\
0	0	0	0	0	
1	0	0	0	0	
2	0	0	0	0	
3	0	0	0	0	
4	0	0	0	0	

	Caribbean Restaurant	Cheese Shop	Chinese Restaurant	Chocolate Shop	\
0	0	0	0	0	

1	0	0	0	0
2	0	0	0	0
3	0	0	0	0
4	0	0	0	0

	Church	Climbing Gym	Clothing Store	Cocktail Bar	Coffee Shop \
0	0	0	0	0	0
1	0	0	0	0	0
2	0	0	0	0	0
3	0	0	0	0	0
4	0	0	0	0	0

	College Arts Building	College Gym	College Rec Center \
0	0	0	0
1	0	0	0
2	0	0	0
3	0	0	0
4	0	0	0

	Colombian Restaurant	Comfort Food Restaurant	Comic Shop	Concert Hall \
0	0	0	0	0
1	0	0	0	0
2	0	0	0	0
3	0	0	0	0
4	0	0	0	0

	Convenience Store	Cosmetics Shop	Coworking Space	Creperie \
0	0	0	0	0
1	0	0	0	0
2	0	0	0	0
3	0	0	0	0
4	0	0	0	0

	Cuban Restaurant	Cupcake Shop	Dance Studio	Deli / Bodega \
0	0	0	0	0
1	0	0	0	0
2	0	0	0	0
3	0	0	0	0
4	0	0	0	0

	Department Store	Dessert Shop	Dim Sum Restaurant	Diner	Discount Store \
0	0	0	0	0	0
1	0	0	0	0	0
2	0	0	0	0	0
3	0	0	0	0	0
4	0	0	0	0	0

	Dog Run	Doner Restaurant	Donut Shop	Dumpling Restaurant	\
0	0	0	0	0	
1	0	0	0	0	
2	0	0	0	0	
3	0	0	0	0	
4	0	0	0	0	

	Eastern European Restaurant	Electronics Store	Ethiopian Restaurant	\
0	0	0	0	
1	0	0	0	
2	0	0	0	
3	0	0	0	
4	0	0	0	

	Event Space	Falafel Restaurant	Farmers Market	Fast Food Restaurant	\
0	0	0	0	0	
1	0	0	0	0	
2	0	0	0	0	
3	0	0	0	0	
4	0	0	0	0	

	Filipino Restaurant	Fish & Chips Shop	Fish Market	Flea Market	Food	\
0	0	0	0	0	0	
1	0	0	0	0	0	
2	0	0	0	0	0	
3	0	0	0	0	0	
4	0	0	0	0	0	

	Food & Drink Shop	Food Court	Food Truck	Fountain	French Restaurant	\
0	0	0	0	0	0	
1	0	0	0	0	0	
2	0	0	0	0	0	
3	0	0	0	0	0	
4	0	0	0	0	0	

	Fried Chicken Joint	Fruit & Vegetable Store	Furniture / Home Store	\
0	0	0	0	
1	0	0	0	
2	0	0	0	
3	0	0	0	
4	0	0	0	

	Gaming Cafe	Garden	Garden Center	Gastropub	Gay Bar	\
0	0	0	0	0	0	
1	0	0	0	0	0	
2	0	0	0	0	0	
3	0	0	0	0	0	

4 0 0 0 0 0

	General Entertainment	General Travel	German Restaurant	Gift Shop	\
0	0	0	0	0	
1	0	0	0	0	
2	0	0	0	0	
3	0	0	0	0	
4	0	0	0	0	

	Gluten-free Restaurant	Gourmet Shop	Greek Restaurant	Grocery Store	Gym	\
0	0	0	0	0	0	
1	0	0	0	0	0	
2	0	0	0	0	0	
3	0	0	0	0	0	
4	0	0	1	0	0	

	Gym / Fitness Center	Harbor / Marina	Health & Beauty Service	\
0	0	0	0	
1	0	0	0	
2	0	0	0	
3	0	0	0	
4	0	0	0	

	Health Food Store	Historic Site	History Museum	Hobby Shop	Hookah Bar	\
0	0	0	0	0	0	
1	1	0	0	0	0	
2	0	0	0	0	0	
3	0	0	0	0	0	
4	0	0	0	0	0	

	Hospital	Hostel	Hotel	Hotel Bar	Hotpot Restaurant	Ice Cream Shop	\
0	0	0	0	0	0	0	
1	0	0	0	0	0	0	
2	0	0	0	0	0	0	
3	0	0	0	0	0	0	
4	0	0	0	0	0	0	

	Indian Restaurant	Indie Movie Theater	Indoor Play Area	Intersection	\
0	0	0	0	0	
1	0	0	0	0	
2	0	0	0	0	
3	0	0	0	0	
4	0	0	0	0	

	Italian Restaurant	Japanese Restaurant	Jazz Club	Jewelry Store	\
0	0	0	0	0	
1	0	0	0	0	

2	0	0	0	0
3	0	0	0	0
4	0	0	0	0

	Jewish Restaurant	Juice Bar	Korean Restaurant	Lake	\
0	0	0		0	0
1	0	0		0	0
2	0	0		0	0
3	0	0		0	0
4	0	0		0	0

	Latin American Restaurant	Light Rail Station	Lingerie Store	\
0		0	0	0
1		0	0	0
2		0	0	0
3		0	0	0
4		0	0	0

	Liquor Store	Lounge	Mac & Cheese Joint	Malay Restaurant	Market	\
0	0	0	0		0	0
1	0	0	0		0	0
2	0	0	0		0	0
3	0	0	0		0	0
4	0	0	0		0	0

	Martial Arts Dojo	Mediterranean Restaurant	Men's Store	\
0	0		0	0
1	0		0	0
2	0		0	0
3	0		0	0
4	0		0	0

	Mexican Restaurant	Middle Eastern Restaurant	Miscellaneous Shop	\
0	0		0	0
1	0		0	0
2	0		0	0
3	0		0	0
4	0		0	0

	Modern European Restaurant	Molecular Gastronomy Restaurant	\
0		0	0
1		0	0
2		0	0
3		0	0
4		0	0

Monument / Landmark	Movie Theater	Moving Target	Museum	Music Venue	\
---------------------	---------------	---------------	--------	-------------	---

0	0	0	0	0	0
1	0	0	0	0	0
2	0	0	0	0	0
3	0	0	0	0	0
4	0	0	0	0	0

	Neighborhood	New American Restaurant	Nightclub	Noodle House	Office \
0	0	0	0	0	0
1	0	0	0	0	0
2	0	0	0	0	0
3	1	0	0	0	0
4	0	0	0	0	0

	Opera House	Optical Shop	Organic Grocery	Other Great Outdoors \
0	0	0	0	0
1	0	0	0	0
2	0	0	0	0
3	0	0	0	0
4	0	0	0	0

	Outdoor Sculpture	Park	Performing Arts Venue	Pet Store	Pharmacy \
0	0	0	0	0	0
1	0	0	0	0	0
2	0	0	0	0	0
3	0	0	0	0	0
4	0	0	0	0	0

	Pizza Place	Playground	Plaza	Poke Place	Portuguese Restaurant \
0	0	0	0	0	0
1	0	0	0	0	0
2	0	0	0	0	0
3	0	0	0	0	0
4	0	0	0	0	0

	Poutine Place	Pub	Ramen Restaurant	Record Shop	Rental Car Location \
0	0	0	0	0	0
1	0	0	0	0	0
2	0	1	0	0	0
3	0	0	0	0	0
4	0	0	0	0	0

	Restaurant	Roof Deck	Sake Bar	Salad Place	Salon / Barbershop \
0	0	0	0	0	0
1	0	0	0	0	0
2	0	0	0	0	0
3	0	0	0	0	0
4	0	0	0	0	0

	Sandwich Place	Scenic Lookout	Sculpture Garden	Seafood Restaurant	\
0	0	0	0	0	
1	0	0	0	0	
2	0	0	0	0	
3	0	0	0	0	
4	0	0	0	0	

	Shoe Store	Shopping Mall	Skate Park	Skating Rink	Smoke Shop	\
0	0	0	0	0	0	
1	0	0	0	0	0	
2	0	0	0	0	0	
3	0	0	0	0	0	
4	0	0	0	0	0	

	Smoothie Shop	Snack Place	Soup Place	Southern / Soul Food Restaurant	\
0	0	0	0		0
1	0	0	0		0
2	0	0	0		0
3	0	0	0		0
4	0	0	0		0

	Spa	Speakeasy	Sporting Goods Shop	Sports Bar	Stadium	Stationery Store	\
0	0	0	0	0	0		0
1	0	0	0	0	0		0
2	0	0	0	0	0		0
3	0	0	0	0	0		0
4	0	0	0	0	0		0

	Steakhouse	Strip Club	Summer Camp	Supermarket	Supplement Shop	\
0	0	0	0	0	0	
1	0	0	0	0	0	
2	0	0	0	0	0	
3	0	0	0	0	0	
4	0	0	0	0	0	

	Sushi Restaurant	Swim School	Taco Place	Tailor Shop	\
0	0	0	0	0	
1	0	0	0	0	
2	0	0	0	0	
3	0	0	0	0	
4	0	0	0	0	

	Taiwanese Restaurant	Tanning Salon	Tapas Restaurant	Tea Room	\
0	0	0	0	0	
1	0	0	0	0	
2	0	0	0	0	

3	0	0	0	0
4	0	0	0	0

	Tennis Court	Thai Restaurant	Theater	Theme Restaurant	\
0	0	0	0	0	
1	0	0	0	0	
2	0	0	0	0	
3	0	0	0	0	
4	0	0	0	0	

	Thrift / Vintage Store	Toy / Game Store	Trail	Train Station	\
0	0	0	1	0	
1	0	0	0	0	
2	0	0	0	0	
3	0	0	0	0	
4	0	0	0	0	

	Vegetarian / Vegan Restaurant	Video Game Store	Vietnamese Restaurant	\
0	0	0	0	
1	0	0	0	
2	0	0	0	
3	0	0	0	
4	0	0	0	

	Wine Bar	Wings Joint	Women's Store	Yoga Studio
0	0	0	0	0
1	0	0	0	0
2	0	0	0	0
3	0	0	0	0
4	0	0	0	0

```
[28]: toronto_onehot.shape
```

```
[28]: (1715, 238)
```

```
[29]: #Grouping rows by neighborhood and by taking the mean of the frequency of
      ↳ occurrence of each category
toronto_grouped = toronto_onehot.groupby('Neighbourhood').mean().reset_index()
```

```
[30]: toronto_grouped.shape
```

```
[30]: (38, 238)
```

```
[31]: #function to sort the venues in descending order
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]
```

[32]: #Dataframe for the top 10 venues for each neighborhood.

```
num_top_venues = 10

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

# create columns according to number of top venues
columns = ['Neighbourhood']
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['Neighbourhood'] = toronto_grouped['Neighbourhood']

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.head()
```

[32]:

	Neighbourhood	1st Most Common Venue	\
0	Adelaide,King,Richmond	Coffee Shop	
1	Berczy Park	Coffee Shop	
2	Brockton,Exhibition Place,Parkdale Village	Coffee Shop	
3	Business Reply Mail Processing Centre 969 Eastern	Light Rail Station	
4	CN Tower,Bathurst Quay,Island airport,Harbourf...	Airport Service	

	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	\
0	Café	Bar	Thai Restaurant	
1	Cocktail Bar	Café	Farmers Market	
2	Breakfast Spot	Café	Climbing Gym	
3	Yoga Studio	Auto Workshop	Garden Center	
4	Airport Lounge	Airport Terminal	Boat or Ferry	

	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	\
0	Hotel	American Restaurant	Breakfast Spot	
1	Steakhouse	Beer Bar	Bakery	
2	Stadium	Burrito Place	Sandwich Place	
3	Garden	Fast Food Restaurant	Farmers Market	
4	Harbor / Marina	Sculpture Garden	Coffee Shop	

	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Restaurant	Steakhouse	Asian Restaurant
1	Cheese Shop	Italian Restaurant	Seafood Restaurant
2	Caribbean Restaurant	Restaurant	Yoga Studio
3	Comic Shop	Park	Gym / Fitness Center
4	Boutique	Bar	Airport Gate

0.8 4. Cluster Neighborhoods

```
[33]: #Let's create a new dataframe that includes the cluster as well as the top 10
      ↪venues for each neighborhood.
```

```
[34]: kclusters = 5

toronto_grouped_clustering = toronto_grouped.drop('Neighbourhood', 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]
```

```
[34]: array([0, 0, 0, 0, 0, 0, 0, 0, 0, 0], dtype=int32)
```

```
[35]: #Let's create a new dataframe that includes the cluster as well as the top 10
      ↪venues for each neighborhood.

# add clustering labels
neighborhoods_venues_sorted.insert(0, 'Cluster Labels', kmeans.labels_)
```

```
[36]: neighborhoods_venues_sorted.head()
```

	Cluster Labels	Neighbourhood \
0	0	Adelaide,King,Richmond
1	0	Berczy Park
2	0	Brockton,Exhibition Place,Parkdale Village
3	0	Business Reply Mail Processing Centre 969 Eastern
4	0	CN Tower,Bathurst Quay,Island airport,Harbourf...

	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue \
0	Coffee Shop	Café	Bar
1	Coffee Shop	Cocktail Bar	Café
2	Coffee Shop	Breakfast Spot	Café
3	Light Rail Station	Yoga Studio	Auto Workshop

4	Airport Service	Airport Lounge	Airport Terminal
---	-----------------	----------------	------------------

	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	\
0	Thai Restaurant	Hotel	American Restaurant	
1	Farmers Market	Steakhouse	Beer Bar	
2	Climbing Gym	Stadium	Burrito Place	
3	Garden Center	Garden	Fast Food Restaurant	
4	Boat or Ferry	Harbor / Marina	Sculpture Garden	

	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	\
0	Breakfast Spot	Restaurant	Steakhouse	
1	Bakery	Cheese Shop	Italian Restaurant	
2	Sandwich Place	Caribbean Restaurant	Restaurant	
3	Farmers Market	Comic Shop	Park	
4	Coffee Shop	Boutique	Bar	

	10th Most Common Venue
0	Asian Restaurant
1	Seafood Restaurant
2	Yoga Studio
3	Gym / Fitness Center
4	Airport Gate

```
[37]: toronto_merged = df_t

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

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

```
[37]: Postcode      Borough      Neighbourhood  Latitude \
0      M4E      East Toronto      The Beaches  43.676357
1      M4K      East Toronto      The Danforth West,Riverdale  43.679557
2      M4L      East Toronto      The Beaches West,India Bazaar  43.668999
3      M4M      East Toronto      Studio District  43.659526
4      M4N      Central Toronto      Lawrence Park  43.728020
```


	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	\
0	-79.293031	4	Health Food Store	Pub	
1	-79.352188	0	Greek Restaurant	Coffee Shop	
2	-79.315572	0	Park	Pizza Place	
3	-79.340923	0	Café	Coffee Shop	
4	-79.388790	2	Dim Sum Restaurant	Park	

	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	\
--	-----------------------	-----------------------	-----------------------	---

0	Trail	Neighborhood	Event Space
1	Italian Restaurant	Ice Cream Shop	Furniture / Home Store
2	Fast Food Restaurant	Sushi Restaurant	Movie Theater
3	Bakery	Italian Restaurant	American Restaurant
4	Swim School	Bus Line	Diner

	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue \
0	Ethiopian Restaurant	Electronics Store	Falafel Restaurant
1	Bubble Tea Shop	Sports Bar	Juice Bar
2	Ice Cream Shop	Fish & Chips Shop	Pub
3	Sandwich Place	Bar	Stationery Store
4	Falafel Restaurant	Event Space	Ethiopian Restaurant

	9th Most Common Venue	10th Most Common Venue
0	Eastern European Restaurant	Dessert Shop
1	Spa	Liquor Store
2	Intersection	Italian Restaurant
3	Fish Market	Coworking Space
4	Electronics Store	Eastern European Restaurant

```
[38]: # Create map to visualize the clusters
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
```

```
[38]: <folium.folium.Map at 0x7f2955ae27f0>
```

0.9 5. Examine Clusters

Cluster0

```
[39]: toronto_merged.loc[toronto_merged['Cluster Labels'] == 0, toronto_merged.
      ↪columns[[2] + list(range(5, toronto_merged.shape[1]))]]
```

```
[39]:
```

	Neighbourhood	Cluster Labels	\
1	The Danforth West,Riverdale	0	
2	The Beaches West,India Bazaar	0	
3	Studio District	0	
5	Davisville North	0	
6	North Toronto West	0	
7	Davisville	0	
9	Deer Park,Forest Hill SE,Rathnelly,South Hill,...	0	
11	Cabbagetown,St. James Town	0	
12	Church and Wellesley	0	
13	Harbourfront,Regent Park	0	
14	Ryerson,Garden District	0	
15	St. James Town	0	
16	Berczy Park	0	
17	Central Bay Street	0	
18	Adelaide,King,Richmond	0	
19	Harbourfront East,Toronto Islands,Union Station	0	
20	Design Exchange,Toronto Dominion Centre	0	
21	Commerce Court,Victoria Hotel	0	
23	Forest Hill North,Forest Hill West	0	
24	The Annex,North Midtown,Yorkville	0	
25	Harbord,University of Toronto	0	
26	Chinatown,Grange Park,Kensington Market	0	
27	CN Tower,Bathurst Quay,Island airport,Harbourf...	0	
28	Stn A PO Boxes 25 The Esplanade	0	
29	First Canadian Place,Underground city	0	
30	Christie	0	
31	Dovercourt Village,Dufferin	0	
32	Little Portugal,Trinity	0	
33	Brockton,Exhibition Place,Parkdale Village	0	
34	High Park,The Junction South	0	
35	Parkdale,Roncesvalles	0	
36	Runnymede,Swansea	0	
37	Business Reply Mail Processing Centre 969 Eastern	0	

	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	\
1	Greek Restaurant	Coffee Shop	Italian Restaurant	
2	Park	Pizza Place	Fast Food Restaurant	
3	Café	Coffee Shop	Bakery	
5	Clothing Store	Gym	Park	
6	Clothing Store	Coffee Shop	Sporting Goods Shop	
7	Sandwich Place	Dessert Shop	Pizza Place	

9	Pub	Coffee Shop	Pizza Place
11	Coffee Shop	Restaurant	Pizza Place
12	Coffee Shop	Japanese Restaurant	Sushi Restaurant
13	Coffee Shop	Park	Pub
14	Clothing Store	Coffee Shop	Cosmetics Shop
15	Coffee Shop	Hotel	Café
16	Coffee Shop	Cocktail Bar	Café
17	Coffee Shop	Café	Ice Cream Shop
18	Coffee Shop	Café	Bar
19	Coffee Shop	Hotel	Aquarium
20	Coffee Shop	Café	Hotel
21	Coffee Shop	Hotel	Café
23	Mexican Restaurant	Trail	Sushi Restaurant
24	Sandwich Place	Café	Coffee Shop
25	Café	Bar	Italian Restaurant
26	Bar	Chinese Restaurant	Vegetarian / Vegan Restaurant
27	Airport Service	Airport Lounge	Airport Terminal
28	Coffee Shop	Café	Restaurant
29	Coffee Shop	Café	Hotel
30	Grocery Store	Café	Park
31	Pharmacy	Bakery	Supermarket
32	Bar	Coffee Shop	Asian Restaurant
33	Coffee Shop	Breakfast Spot	Café
34	Mexican Restaurant	Café	Bar
35	Breakfast Spot	Gift Shop	Bookstore
36	Sushi Restaurant	Café	Coffee Shop
37	Light Rail Station	Yoga Studio	Auto Workshop

	4th Most Common Venue	5th Most Common Venue \
1	Ice Cream Shop	Furniture / Home Store
2	Sushi Restaurant	Movie Theater
3	Italian Restaurant	American Restaurant
5	Breakfast Spot	Hotel
6	Gift Shop	Ice Cream Shop
7	Coffee Shop	Gym
9	Light Rail Station	Sports Bar
11	Italian Restaurant	Pub
12	Gay Bar	Burger Joint
13	Bakery	Café
14	Middle Eastern Restaurant	Café
15	Restaurant	Breakfast Spot
16	Farmers Market	Steakhouse
17	Italian Restaurant	Sandwich Place
18	Thai Restaurant	Hotel
19	Café	Fried Chicken Joint
20	Restaurant	American Restaurant
21	Restaurant	American Restaurant

23	Jewelry Store	Yoga Studio
24	Pizza Place	Indian Restaurant
25	Japanese Restaurant	Bookstore
26	Café	Bakery
27	Boat or Ferry	Harbor / Marina
28	Italian Restaurant	Seafood Restaurant
29	Restaurant	Steakhouse
30	Candy Store	Convenience Store
31	Bank	Pet Store
32	Men's Store	New American Restaurant
33	Climbing Gym	Stadium
34	Thai Restaurant	Speakeasy
35	Bank	Dog Run
36	Italian Restaurant	Restaurant
37	Garden Center	Garden

	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue \
1	Bubble Tea Shop	Sports Bar	Juice Bar
2	Ice Cream Shop	Fish & Chips Shop	Pub
3	Sandwich Place	Bar	Stationery Store
5	Sandwich Place	Asian Restaurant	Food & Drink Shop
6	Mexican Restaurant	Diner	Dessert Shop
7	Café	Italian Restaurant	Sushi Restaurant
9	Supermarket	Sushi Restaurant	Bagel Shop
11	Bakery	Café	Breakfast Spot
12	Restaurant	Ramen Restaurant	Gym
13	Mexican Restaurant	Theater	Breakfast Spot
14	Sporting Goods Shop	Pizza Place	Fast Food Restaurant
15	Clothing Store	Cosmetics Shop	Italian Restaurant
16	Beer Bar	Bakery	Cheese Shop
17	Burger Joint	Chinese Restaurant	Japanese Restaurant
18	American Restaurant	Breakfast Spot	Restaurant
19	Italian Restaurant	Brewery	Restaurant
20	Gastropub	Bar	Seafood Restaurant
21	Seafood Restaurant	Steakhouse	Italian Restaurant
23	Diner	Falafel Restaurant	Event Space
24	Pharmacy	Convenience Store	Cosmetics Shop
25	Restaurant	Bakery	Poutine Place
26	Coffee Shop	Mexican Restaurant	Vietnamese Restaurant
27	Sculpture Garden	Coffee Shop	Boutique
28	Cocktail Bar	Hotel	Beer Bar
29	Gastropub	Bar	American Restaurant
30	Baby Store	Coffee Shop	Italian Restaurant
31	Bar	Gym / Fitness Center	Park
32	Bakery	Cocktail Bar	Pizza Place
33	Burrito Place	Sandwich Place	Caribbean Restaurant
34	Furniture / Home Store	Bookstore	Fast Food Restaurant

35	Bar	Italian Restaurant	Dessert Shop
36	Pizza Place	Bar	Diner
37	Fast Food Restaurant	Farmers Market	Comic Shop

9th Most Common Venue		10th Most Common Venue	
1	Spa		Liquor Store
2	Intersection		Italian Restaurant
3	Fish Market		Coworking Space
5	Donut Shop		Discount Store
6	Cosmetics Shop		Bagel Shop
7	Pharmacy		Greek Restaurant
9	Restaurant		Fried Chicken Joint
11	Market		Snack Place
12	Gastropub		Hotel
13	Gym / Fitness Center		Health Food Store
14	Restaurant		Ramen Restaurant
15	Bakery		Gastropub
16	Italian Restaurant		Seafood Restaurant
17	Salad Place		Bubble Tea Shop
18	Steakhouse		Asian Restaurant
19	Scenic Lookout		Sports Bar
20	Italian Restaurant		Deli / Bodega
21	Gym		Gastropub
23	Ethiopian Restaurant		Electronics Store
24	Pub		Burger Joint
25	Beer Bar		Beer Store
26	Dumpling Restaurant		Gaming Cafe
27	Bar		Airport Gate
28	Bakery		Cosmetics Shop
29	Deli / Bodega		Gym
30	Diner		Athletics & Sports
31	Brewery		Café
32	Café		French Restaurant
33	Restaurant		Yoga Studio
34	Park		Flea Market
35	Restaurant	Eastern European Restaurant	
36	Food	Fish & Chips Shop	
37	Park	Gym / Fitness Center	

Cluster0 seems to have venues that are popular in the busy areas as is probably marks the business district of Toronto with lots of cafes, pubs and light weight entertainment centers to suit the busy lifestyle of the working class. This cluster also encompasses the bulk of Toronto neighbourhoods indicating the lifestyle or key characteristic of the city itself.

Cluster1

```
[40]: toronto_merged.loc[toronto_merged['Cluster Labels'] == 1, toronto_merged.
      ↪columns[[2] + list(range(5, toronto_merged.shape[1]))]]
```

```
[40]:   Neighbourhood  Cluster Labels 1st Most Common Venue 2nd Most Common Venue \
22      Roselawn              1              Garden              Yoga Studio

      3rd Most Common Venue 4th Most Common Venue 5th Most Common Venue \
22  Filipino Restaurant      Farmers Market      Falafel Restaurant

      6th Most Common Venue 7th Most Common Venue 8th Most Common Venue \
22          Event Space  Ethiopian Restaurant      Electronics Store

      9th Most Common Venue 10th Most Common Venue
22  Eastern European Restaurant      Dumpling Restaurant
```

Cluster1 seems like a residential zone with the basic and essential venues particularly useful for the vulnerable ie children and aged.¶ There is also geographical proximity of these neighbourhoods most stretching away from downtown to the north

Cluster2

```
[41]: toronto_merged.loc[toronto_merged['Cluster Labels'] == 2, toronto_merged.
      ↪columns[[2] + list(range(5, toronto_merged.shape[1]))]]
```

```
[41]:   Neighbourhood  Cluster Labels 1st Most Common Venue 2nd Most Common Venue \
4  Lawrence Park              2      Dim Sum Restaurant              Park

      3rd Most Common Venue 4th Most Common Venue 5th Most Common Venue \
4          Swim School              Bus Line              Diner

      6th Most Common Venue 7th Most Common Venue 8th Most Common Venue \
4  Falafel Restaurant      Event Space  Ethiopian Restaurant

      9th Most Common Venue      10th Most Common Venue
4  Electronics Store  Eastern European Restaurant
```

Cluster2 seems to have venues characteristic of people who want to enjoy peaceful and leisure living. Is a residential locality surrounded by posh neighbourhoods. Is geographically toward the north, away from the key business district.

Cluster3

```
[42]: toronto_merged.loc[toronto_merged['Cluster Labels'] == 3, toronto_merged.
      ↪columns[[2] + list(range(5, toronto_merged.shape[1]))]]
```

```
[42]:   Neighbourhood  Cluster Labels 1st Most Common Venue \
8  Moore Park,Summerhill East              3              Playground
10      Rosedale              3              Park
```

	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	\
8	Park	Summer Camp	Tennis Court	
10	Playground	Trail	Dessert Shop	

	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	\
8	Dessert Shop	Event Space	Ethiopian Restaurant	
10	Falafel Restaurant	Event Space	Ethiopian Restaurant	

	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
8	Electronics Store	Eastern European Restaurant	Dumpling Restaurant
10	Electronics Store	Eastern European Restaurant	Dumpling Restaurant

Cluster3 seems like a zone inhabited with health conscious poeple. Wiki notes on Moore park - "Moore Park is one of Toronto's most affluent neighbourhoods."

Cluster4

```
[43]: toronto_merged.loc[toronto_merged['Cluster Labels'] == 4, toronto_merged.
      ↪columns[[2] + list(range(5, toronto_merged.shape[1]))]]
```

	Neighbourhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	\
0	The Beaches	4	Health Food Store	Pub	

	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	\
0	Trail	Neighborhood	Event Space	

	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	\
0	Ethiopian Restaurant	Electronics Store	Falafel Restaurant	

	9th Most Common Venue	10th Most Common Venue
0	Eastern European Restaurant	Dessert Shop

Cluster4 seems to have venues particularly suited for a peaceful community and large greenspace in central Toronto and seems prosperous. It is aligned to wikipedia content on Rosedale: "It is located north of Downtown Toronto and is one of its oldest suburbs. It is also one of the wealthiest and most highly priced neighbourhoods in Canada.[2] Rosedale has been ranked the best neighbourhood in Toronto to live in by Toronto Life.[3] It is known as the area where the city's 'old money' lives,[4] and is home to some of Canada's richest and most famous citizens including Gerry Schwartz, founder of Onex Corporation, and Ken Thomson of Thomson Corporation, the latter of whom was the richest man in Canada at the time of his death in 2006"

0.9.1 Finding top 10 common restaurant categories in Toronto

```
[44]: t_venue_cat_common=toronto_venues.groupby(['Venue Category']).size().
      ↪reset_index(name='count').sort_values('count', ascending=False)
t_res=t_venue_cat_common[t_venue_cat_common['Venue Category'].str.
      ↪contains("Restaurant")].reset_index(drop=True)
t_res.head(10).plot('Venue Category', 'count', kind='bar', figsize=(8,4),
      ↪width=.25,colormap='Paired')
```

```
[44]: <matplotlib.axes._subplots.AxesSubplot at 0x7f294fdfdcf8>
```

0.9.2 Finding neighborhoods with high restaurant density in Toronto

```
[45]: toronto_res=toronto_venues[toronto_venues['Venue Category'].str.
      ↪contains("Restaurant")].reset_index(drop=True)
#toronto_thai
toronto_res_nei=toronto_res.groupby(['Neighborhood']).size().
      ↪reset_index(name='count').sort_values('count', ascending=False)
toronto_res_nei
```

```
[45]:
```

	Neighborhood	count
6	Chinatown,Grange Park,Kensington Market	33
0	Adelaide,King,Richmond	28
8	Church and Wellesley	28
9	Commerce Court,Victoria Hotel	26
15	First Canadian Place,Underground city	25
13	Design Exchange,Toronto Dominion Centre	24
27	St. James Town	24
26	Ryerson,Garden District	23
5	Central Bay Street	23
28	Stn A PO Boxes 25 The Esplanade	21
22	Little Portugal,Trinity	21
32	The Danforth West,Riverdale	15
18	Harbourfront East,Toronto Islands,Union Station	14
1	Berczy Park	11
25	Runnymede,Swansea	11
10	Davisville	10
17	Harbord,University of Toronto	10
29	Studio District	10
4	Cabbagetown,St. James Town	10
20	High Park,The Junction South	7
19	Harbourfront,Regent Park	6
12	Deer Park,Forest Hill SE,Rathnelly,South Hill,...	4
30	The Annex,North Midtown,Yorkville	4
24	Parkdale,Roncesvalles	4
2	Brockton,Exhibition Place,Parkdale Village	3
31	The Beaches West,India Bazaar	3

23	North Toronto West	3
7	Christie	2
3	Business Reply Mail Processing Centre 969 Eastern	2
16	Forest Hill North,Forest Hill West	2
11	Davisville North	1
21	Lawrence Park	1
14	Dovercourt Village,Dufferin	1

0.9.3 Finding neighborhoods with Thai restaurants in Toronto

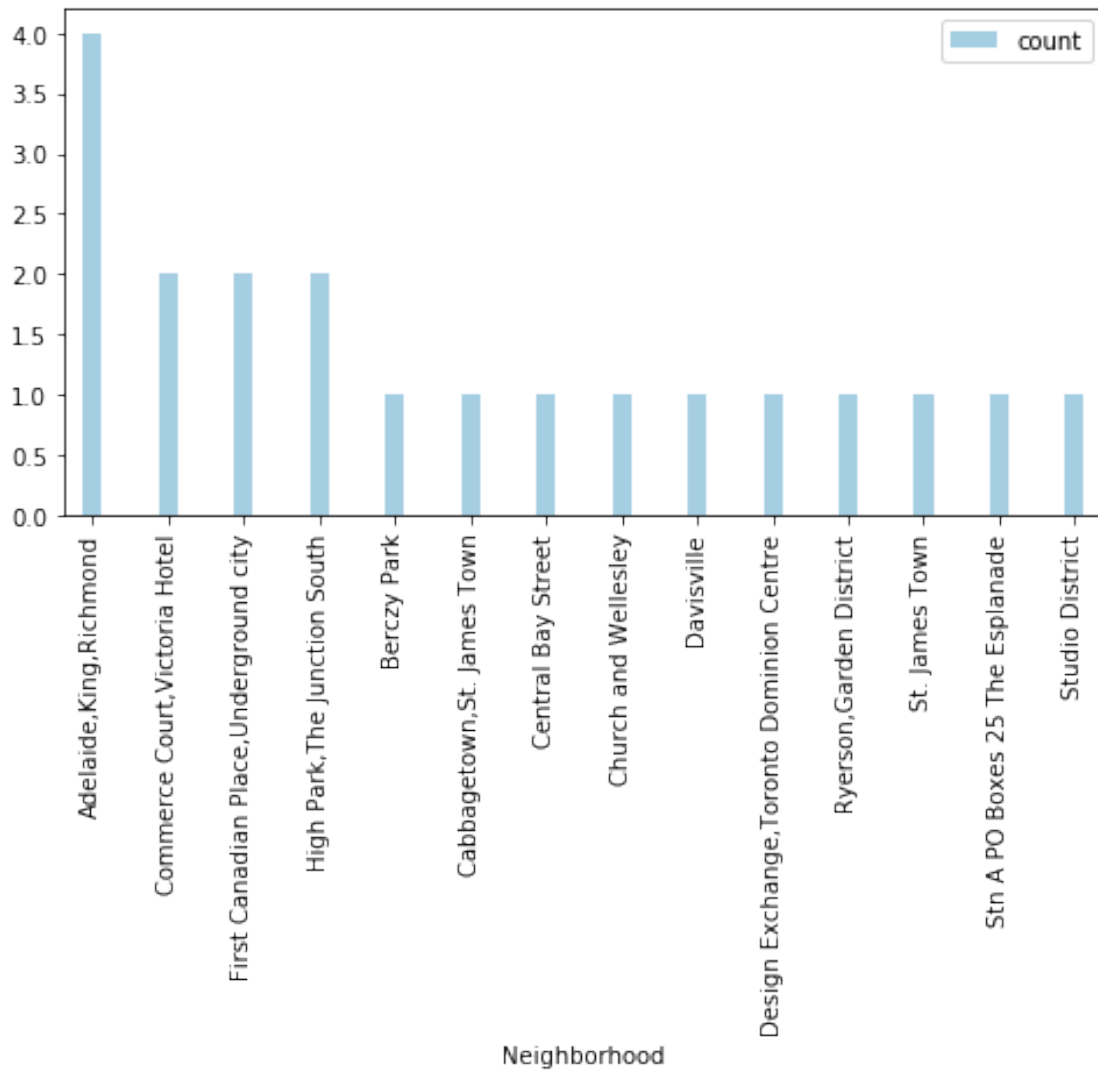
```
[46]: toronto_thai=toronto_venues[toronto_venues['Venue Category'].str.contains("Thai_
↳Restaurant")].reset_index(drop=True)
#toronto_thai
toronto_thai_nei=toronto_thai.groupby(['Neighborhood']).size().
↳reset_index(name='count').sort_values('count', ascending=False)
toronto_thai_nei
```

```
[46]:
```

	Neighborhood	count
0	Adelaide,King,Richmond	4
5	Commerce Court,Victoria Hotel	2
8	First Canadian Place,Underground city	2
9	High Park,The Junction South	2
1	Berczy Park	1
2	Cabbagetown,St. James Town	1
3	Central Bay Street	1
4	Church and Wellesley	1
6	Davisville	1
7	Design Exchange,Toronto Dominion Centre	1
10	Ryerson,Garden District	1
11	St. James Town	1
12	Stn A PO Boxes 25 The Esplanade	1
13	Studio District	1

```
[47]: toronto_thai_nei.plot('Neighborhood', 'count', kind='bar', figsize=(8,4),
↳width=.25,colormap='Paired')
```

```
[47]: <matplotlib.axes._subplots.AxesSubplot at 0x7f294fc91710>
```



```
[48]: # merge toronto_grouped with toronto_data to add latitude/longitude for each
      ↪neighborhood
toronto_thai_merged = toronto_merged.join(toronto_thai_nei.
      ↪set_index('Neighborhood'), on='Neighbourhood').sort_values('count',
      ↪ascending=False)
toronto_thai_merged.dropna(axis=0, subset=('count', ), inplace=True)
toronto_thai_merged
```

```
[48]: Postcode      Borough      Neighbourhood \
18      M5H      Downtown Toronto      Adelaide,King,Richmond
21      M5L      Downtown Toronto      Commerce Court,Victoria Hotel
29      M5X      Downtown Toronto      First Canadian Place,Underground city
34      M6P      West Toronto      High Park,The Junction South
3       M4M      East Toronto      Studio District
```

7	M4S	Central Toronto		Davisville
11	M4X	Downtown Toronto		Cabbagetown, St. James Town
12	M4Y	Downtown Toronto		Church and Wellesley
14	M5B	Downtown Toronto		Ryerson, Garden District
15	M5C	Downtown Toronto		St. James Town
16	M5E	Downtown Toronto		Berczy Park
17	M5G	Downtown Toronto		Central Bay Street
20	M5K	Downtown Toronto	Design Exchange, Toronto Dominion Centre	
28	M5W	Downtown Toronto	Stn A PO Boxes 25 The Esplanade	

	Latitude	Longitude	Cluster Labels	1st Most Common Venue \
18	43.650571	-79.384568	0	Coffee Shop
21	43.648198	-79.379817	0	Coffee Shop
29	43.648429	-79.382280	0	Coffee Shop
34	43.661608	-79.464763	0	Mexican Restaurant
3	43.659526	-79.340923	0	Café
7	43.704324	-79.388790	0	Sandwich Place
11	43.667967	-79.367675	0	Coffee Shop
12	43.665860	-79.383160	0	Coffee Shop
14	43.657162	-79.378937	0	Clothing Store
15	43.651494	-79.375418	0	Coffee Shop
16	43.644771	-79.373306	0	Coffee Shop
17	43.657952	-79.387383	0	Coffee Shop
20	43.647177	-79.381576	0	Coffee Shop
28	43.646435	-79.374846	0	Coffee Shop

	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue \
18	Café	Bar	Thai Restaurant
21	Hotel	Café	Restaurant
29	Café	Hotel	Restaurant
34	Café	Bar	Thai Restaurant
3	Coffee Shop	Bakery	Italian Restaurant
7	Dessert Shop	Pizza Place	Coffee Shop
11	Restaurant	Pizza Place	Italian Restaurant
12	Japanese Restaurant	Sushi Restaurant	Gay Bar
14	Coffee Shop	Cosmetics Shop	Middle Eastern Restaurant
15	Hotel	Café	Restaurant
16	Cocktail Bar	Café	Farmers Market
17	Café	Ice Cream Shop	Italian Restaurant
20	Café	Hotel	Restaurant
28	Café	Restaurant	Italian Restaurant

	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue \
18	Hotel	American Restaurant	Breakfast Spot
21	American Restaurant	Seafood Restaurant	Steakhouse
29	Steakhouse	Gastropub	Bar
34	Speakeasy	Furniture / Home Store	Bookstore

3	American Restaurant	Sandwich Place	Bar
7	Gym	Café	Italian Restaurant
11	Pub	Bakery	Café
12	Burger Joint	Restaurant	Ramen Restaurant
14	Café	Sporting Goods Shop	Pizza Place
15	Breakfast Spot	Clothing Store	Cosmetics Shop
16	Steakhouse	Beer Bar	Bakery
17	Sandwich Place	Burger Joint	Chinese Restaurant
20	American Restaurant	Gastropub	Bar
28	Seafood Restaurant	Cocktail Bar	Hotel

	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue	count
18	Restaurant	Steakhouse	Asian Restaurant	4.0
21	Italian Restaurant	Gym	Gastropub	2.0
29	American Restaurant	Deli / Bodega	Gym	2.0
34	Fast Food Restaurant	Park	Flea Market	2.0
3	Stationery Store	Fish Market	Coworking Space	1.0
7	Sushi Restaurant	Pharmacy	Greek Restaurant	1.0
11	Breakfast Spot	Market	Snack Place	1.0
12	Gym	Gastropub	Hotel	1.0
14	Fast Food Restaurant	Restaurant	Ramen Restaurant	1.0
15	Italian Restaurant	Bakery	Gastropub	1.0
16	Cheese Shop	Italian Restaurant	Seafood Restaurant	1.0
17	Japanese Restaurant	Salad Place	Bubble Tea Shop	1.0
20	Seafood Restaurant	Italian Restaurant	Deli / Bodega	1.0
28	Beer Bar	Bakery	Cosmetics Shop	1.0

It is clear that Cluster3 which houses all the existing Thai restaurants and is probably better suited for a new one as well. Let us find out which is the best neighbourhood in Cluster3 that is suited for the new restaurant.

```
[49]: # Create map to superimpose Italian restaurants on the cluster map
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)

# add the Thai restaurants as yellow circle markers
for lat, lng, label in zip(toronto_thai['Neighborhood Latitude'],
    ↳toronto_thai['Neighborhood Longitude'], toronto_thai['Venue']):
    folium.features.CircleMarker(
        [lat, lng],
        radius=2,
        color='black',
        popup=label,
        fill = True,
        fill_color='yellow',
        fill_opacity=0.6
    ).add_to(map_clusters)

map_clusters

```

[49]: <folium.folium.Map at 0x7f294fbe0438>

On the basis of the visual inspection of the existing Thai restaurants in Cluster0, any new restaurant can be positioned in the western Cluster0 neighborhoods ie in the area between High Park and Salad King which shares the characteristics of neighborhoods in which the restaurant business is thriving. Top picks are the neighborhoods of Dovercourt Village and Christie given the low restaurant density, competition is less in that area and hence can prove to be quite profitable.