In [108]:
```
import numpy as np
import pandas as pd
pd.set_option("display.max_columns", None)
pd.set_option("display.max_rows", None)
import matplotlib.pyplot as plt
import requests
import json
import folium
!conda install -c conda-forge folium=0.5.0 --yes
import xml
import matplotlib.cm as cm
import matplotlib.colors as colors
!conda install -c conda-forge geopy --yes
from geopy.geocoders import Nominatim
from bs4 import BeautifulSoup
from pandas.io.json import json_normalize
from sklearn.cluster import KMeans
```

```
Collecting package metadata (current_repodata.json): done
Solving environment: done

# All requested packages already installed.

Collecting package metadata (current_repodata.json): done
Solving environment: done

# All requested packages already installed.
```

In [109]:
```
url='https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M'
result=requests.get(url)
soup=BeautifulSoup(result.content, 'html.parser')
print(url)
print(result.status_code)
print(result.headers)
```

```
https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
200
{'Date': 'Sun, 21 Jun 2020 05:57:57 GMT', 'Vary': 'Accept-Encoding,Cookie,Authorization', 'Server': 'AT
S/8.0.7', 'X-Content-Type-Options': 'nosniff', 'P3p': 'CP="See https://en.wikipedia.org/wiki/Special:Ce
ntralAutoLogin/P3P for more info."', 'Content-Language': 'en', 'Last-Modified': 'Thu, 18 Jun 2020 21:29
:18 GMT', 'Content-Type': 'text/html; charset=UTF-8', 'Content-Encoding': 'gzip', 'Age': '43265', 'X-Ca
che': 'cp3064 hit, cp3050 hit/54', 'X-Cache-Status': 'hit-front', 'Server-Timing': 'cache;desc="hit-fro
nt"', 'Strict-Transport-Security': 'max-age=106384710; includeSubDomains; preload', 'Set-Cookie': 'WMF-
Last-Access=21-Jun-2020;Path=/;HttpOnly;secure;Expires=Thu, 23 Jul 2020 12:00:00 GMT, WMF-Last-Access-G
lobal=21-Jun-2020;Path=/;Domain=.wikipedia.org;HttpOnly;secure;Expires=Thu, 23 Jul 2020 12:00:00 GMT, G
eoIP=GB:ENG:Walthamstow:51.59:-0.02:v4; Path=/; secure; Domain=.wikipedia.org', 'X-Client-IP': '82.15.1
78.91', 'Cache-Control': 'private, s-maxage=0, max-age=0, must-revalidate', 'Accept-Ranges': 'bytes', '
Content-Length': '13207', 'Connection': 'keep-alive'}
```

In [110]:
```
post_data=soup.find('table')
fields=post_data.find_all('td')

Postcode=[]
borough =[]
neighbourhood =[]
```

In [111]:
```
for i in range(0, len(fields), 3):
    Postcode.append(fields[i].text.strip())
    borough.append(fields[i+1].text.strip())
    neighbourhood.append(fields[i+2].text.strip())

df_post=pd.DataFrame(data=[Postcode,borough,neighbourhood]).transpose()
df_post.columns=['Postcode','Borough','Neighborhood']
df_post.head()
```

Out[111]:

	Postcode	Borough	Neighborhood
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	Regent Park, Harbourfront

In [112]:
```
df_postcode=df_post[df_post.Borough !='Not assigned'].reset_index(drop=True)
df_postcode.head()
```

Out[112]:

	Postcode	Borough	Neighborhood
0	M3A	North York	Parkwoods
1	M4A	North York	Victoria Village
2	M5A	Downtown Toronto	Regent Park, Harbourfront
3	M6A	North York	Lawrence Manor, Lawrence Heights
4	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government

In [113]:
```
df_post_group=df_postcode.groupby(['Postcode','Borough'],as_index=False).agg(lambda x:','.join(x))
df_post_group.head()
```

Out[113]:

	Postcode	Borough	Neighborhood
0	M1B	Scarborough	Malvern, Rouge
1	M1C	Scarborough	Rouge Hill, Port Union, Highland Creek
2	M1E	Scarborough	Guildwood, Morningside, West Hill
3	M1G	Scarborough	Woburn
4	M1H	Scarborough	Cedarbrae

In [114]:
```
for inex,row in df_post_group.iterrows():
    if row['Neighborhood']=='Not assigned':
        row['Neighborhood']=row['Borough']

df_post_group.head()
```

Out[114]:

	Postcode	Borough	Neighborhood
0	M1B	Scarborough	Malvern, Rouge
1	M1C	Scarborough	Rouge Hill, Port Union, Highland Creek
2	M1E	Scarborough	Guildwood, Morningside, West Hill
3	M1G	Scarborough	Woburn
4	M1H	Scarborough	Cedarbrae

In [115]:
```
df_post_group.shape
```

Out[115]: (103, 3)

In [116]:
```
df_post_group.to_csv('Toronto_Postcodes.csv')
```

In [117]:
```
url_1 = 'http://cocl.us/Geospatial_data'
df_coordinates = pd.read_csv(url_1)
df_coordinates.head()
```

Out[117]:

	Postal Code	Latitude	Longitude
0	M1B	43.806686	-79.194353
1	M1C	43.784535	-79.160497
2	M1E	43.763573	-79.188711
3	M1G	43.770992	-79.216917
4	M1H	43.773136	-79.239476

In [118]:
```
df_coordinates.rename(columns={'Postal Code':'Postcode'},inplace=True)
df_coordinates.head()
```

Out[118]:

	Postcode	Latitude	Longitude
0	M1B	43.806686	-79.194353
1	M1C	43.784535	-79.160497
2	M1E	43.763573	-79.188711
3	M1G	43.770992	-79.216917
4	M1H	43.773136	-79.239476

In [119]:
```
neighborhood=pd.read_csv('Toronto_Postcodes.csv',index_col=[0])
neighborhood.head()
```

Out[119]:

	Postcode	Borough	Neighborhood
0	M1B	Scarborough	Malvern, Rouge
1	M1C	Scarborough	Rouge Hill, Port Union, Highland Creek
2	M1E	Scarborough	Guildwood, Morningside, West Hill
3	M1G	Scarborough	Woburn
4	M1H	Scarborough	Cedarbrae

In [120]:
```
n_coordinates=pd.merge(neighborhood,df_coordinates,on='Postcode')
n_coordinates.head()
```

Out[120]:

	Postcode	Borough	Neighborhood	Latitude	Longitude
0	M1B	Scarborough	Malvern, Rouge	43.806686	-79.194353
1	M1C	Scarborough	Rouge Hill, Port Union, Highland Creek	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 [121]:
```
n_coordinates.to_csv('n_coordinates.csv')
```

In [122]:
```
df=pd.read_csv('n_coordinates.csv',index_col=0)
df.head()
```

Out[122]:

	Postcode	Borough	Neighborhood	Latitude	Longitude
0	M1B	Scarborough	Malvern, Rouge	43.806686	-79.194353
1	M1C	Scarborough	Rouge Hill, Port Union, Highland Creek	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 [123]:
```
print('The data has{} boroughs and{} neighborhoods'.format(len(df['Borough'].unique()),df.shape[0]))
```

```
The data has10 boroughs and103 neighborhoods
```

In [124]:
```
df.groupby('Borough').count()['Neighborhood']
```

Out[124]:
```
Borough
Central Toronto      9
Downtown Toronto    19
East Toronto         5
East York            5
Etobicoke           12
Mississauga          1
North York          24
Scarborough         17
West Toronto         6
York                 5
Name: Neighborhood, dtype: int64
```

In [125]:
```
toronto=df[df['Borough'].str.contains('Toronto')]
toronto.reset_index(inplace=True)
toronto.head()
```

Out[125]:

	index	Postcode	Borough	Neighborhood	Latitude	Longitude
0	37	M4E	East Toronto	The Beaches	43.676357	-79.293031
1	41	M4K	East Toronto	The Danforth West, Riverdale	43.679557	-79.352188
2	42	M4L	East Toronto	India Bazaar, The Beaches West	43.668999	-79.315572
3	43	M4M	East Toronto	Studio District	43.659526	-79.340923
4	44	M4N	Central Toronto	Lawrence Park	43.728020	-79.388790

In [126]:
```
toronto.drop('index',axis=1,inplace=True)
toronto.head()
```

```
/Users/arielchien/opt/anaconda3/lib/python3.7/site-packages/pandas/core/frame.py:4102: SettingWithCopyW
arning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.h
tml#returning-a-view-versus-a-copy
  errors=errors,
```

Out[126]:

	Postcode	Borough	Neighborhood	Latitude	Longitude
0	M4E	East Toronto	The Beaches	43.676357	-79.293031
1	M4K	East Toronto	The Danforth West, Riverdale	43.679557	-79.352188
2	M4L	East Toronto	India Bazaar, The Beaches West	43.668999	-79.315572
3	M4M	East Toronto	Studio District	43.659526	-79.340923
4	M4N	Central Toronto	Lawrence Park	43.728020	-79.388790

In [127]:
```
toronto.groupby('Borough').count()['Neighborhood']
```

Out[127]:
```
Borough
Central Toronto      9
Downtown Toronto    19
East Toronto         5
West Toronto         6
Name: Neighborhood, dtype: int64
```

In [128]:
```
latitude= toronto['Latitude'].mean()
longitude= toronto['Longitude'].mean()
print('The geographical coordinates of Toronto are {}, {}'.format(lat_toronto, lon_toronto))
```

```
The geographical coordinates of Toronto are 43.66713498717948, -79.38987324871795
```

In [129]:
```
boroughs = toronto['Borough'].unique().tolist()
borough_color = {}
for borough in boroughs:
    borough_color[borough]= '#%02X%02X%02X' % tuple(np.random.choice(range(256), size=3))
```

In [132]:
```
map_toronto=folium.Map(location=[latitude,longitude], zoom_start=14)
for lat,lng,borough,neighborhood in zip(toronto['Latitude'], toronto['Longitude'],toronto['Borough'],tor
onto['Neighborhood']):
    label='{},{}'.format(neighborhood,borough)
    label = folium.Popup(label,parse_html=True)
    folium.CircleMarker(
        [lat,lng],
        radius=5,
        popup=label,
        color=borough_color[borough],
        fill_color=borough_color[borough],
        fill_opacity=0.7,parse_html=False).add_to(map_toronto)

map_toronto
```

Out[132]: Make this Notebook Trusted to load map: File -> Trust Notebook

In [134]:
```
CLIENT_ID = '2OHAAQBXAPTXCBIVYRH2HVS51P5AORXJIR430HI1T4J1AHSS'
CLIENT_SECRET = 'QSPNHIGXUS5J04IJZ4A1AMUBHGHX1AOBOGFT1SYUVZZFPRXW'
VERSION = '20180605'
LIMIT = 100
radius = 500
```

In [135]:
```
def getNearbyVenues(names, latitudes, longitudes, radius=500):

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

        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)

        results = requests.get(url).json()["response"]['groups'][0]['items']

        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 [137]:
```
toronto_venues=getNearbyVenues(names=toronto['Neighborhood'],
                               latitudes=toronto['Latitude'],
                               longitudes=toronto['Longitude'])
```

```
The Beaches
The Danforth West, Riverdale
India Bazaar, The Beaches West
Studio District
Lawrence Park
Davisville North
North Toronto West,  Lawrence Park
Davisville
Moore Park, Summerhill East
Summerhill West, Rathnelly, South Hill, Forest Hill SE, Deer Park
Rosedale
St. James Town, Cabbagetown
Church and Wellesley
Regent Park, Harbourfront
Garden District, Ryerson
St. James Town
Berczy Park
Central Bay Street
Richmond, Adelaide, King
Harbourfront East, Union Station, Toronto Islands
Toronto Dominion Centre, Design Exchange
```

```
Commerce Court, Victoria Hotel
Roselawn
Forest Hill North & West, Forest Hill Road Park
The Annex, North Midtown, Yorkville
University of Toronto, Harbord
Kensington Market, Chinatown, Grange Park
CN Tower, King and Spadina, Railway Lands, Harbourfront West, Bathurst Quay, South Niagara, Island airp
ort
Stn A PO Boxes
First Canadian Place, Underground city
Christie
Dufferin, Dovercourt Village
Little Portugal, Trinity
Brockton, Parkdale Village, Exhibition Place
High Park, The Junction South
Parkdale, Roncesvalles
Runnymede, Swansea
Queen's Park, Ontario Provincial Government
Business reply mail Processing Centre, South Central Letter Processing Plant Toronto
```

In [138]:
```
print(toronto_venues.shape)
toronto_venues.head()
```

```
(1623, 7)
```

Out[138]:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	The Beaches	43.676357	-79.293031	Glen Manor Ravine	43.676821	-79.293942	Trail
1	The Beaches	43.676357	-79.293031	The Big Carrot Natural Food Market	43.678879	-79.297734	Health Food Store
2	The Beaches	43.676357	-79.293031	Grover Pub and Grub	43.679181	-79.297215	Pub
3	The Beaches	43.676357	-79.293031	Upper Beaches	43.680563	-79.292869	Neighborhood
4	The Danforth West, Riverdale	43.679557	-79.352188	MenEssentials	43.677820	-79.351265	Cosmetics Shop

In [139]:
```
toronto_venues.groupby('Neighborhood').count()
```

Out[139]:

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Berczy Park	58	58	58	58	58	58
Brockton, Parkdale Village, Exhibition Place	22	22	22	22	22	22
Business reply mail Processing Centre, South Central Letter Processing Plant Toronto	18	18	18	18	18	18
CN Tower, King and Spadina, Railway Lands, Harbourfront West, Bathurst Quay, South Niagara, Island airport	14	14	14	14	14	14
Central Bay Street	65	65	65	65	65	65

Christie	17	17	17	17	17	17
Church and Wellesley	76	76	76	76	76	76
Commerce Court, Victoria Hotel	100	100	100	100	100	100
Davisville	35	35	35	35	35	35
Davisville North	8	8	8	8	8	8
Dufferin, Dovercourt Village	17	17	17	17	17	17
First Canadian Place, Underground city	100	100	100	100	100	100
Forest Hill North & West, Forest Hill Road Park	6	6	6	6	6	6
Garden District, Ryerson	100	100	100	100	100	100
Harbourfront East, Union Station, Toronto Islands	100	100	100	100	100	100
High Park, The Junction South	25	25	25	25	25	25
India Bazaar, The Beaches West	20	20	20	20	20	20
Kensington Market, Chinatown, Grange Park	63	63	63	63	63	63
Lawrence Park	3	3	3	3	3	3
Little Portugal, Trinity	44	44	44	44	44	44
Moore Park, Summerhill East	2	2	2	2	2	2
North Toronto West, Lawrence Park	21	21	21	21	21	21
Parkdale, Roncesvalles	15	15	15	15	15	15
Queen's Park, Ontario Provincial Government	32	32	32	32	32	32
Regent Park, Harbourfront	45	45	45	45	45	45
Richmond, Adelaide, King	94	94	94	94	94	94
Rosedale	4	4	4	4	4	4
Roselawn	3	3	3	3	3	3

Runnymede, Swansea	34	34	34	34	34	34
St. James Town	79	79	79	79	79	79
St. James Town, Cabbagetown	48	48	48	48	48	48
Stn A PO Boxes	97	97	97	97	97	97
Studio District	40	40	40	40	40	40
Summerhill West, Rathnelly, South Hill, Forest Hill SE, Deer Park	17	17	17	17	17	17
The Annex, North Midtown, Yorkville	20	20	20	20	20	20
The Beaches	4	4	4	4	4	4
The Danforth West, Riverdale	43	43	43	43	43	43
Toronto Dominion Centre, Design Exchange	100	100	100	100	100	100
University of Toronto, Harbord	34	34	34	34	34	34

In [140]:
```
print('There are {} uniques categories.'.format(len(toronto_venues['Venue Category'].unique())))
```

```
There are 236 uniques categories.
```

In [141]:
```
toronto_venues['Venue Category'].unique()[:100]
```

Out[141]:
```
array(['Trail', 'Health Food Store', 'Pub', 'Neighborhood',
       'Cosmetics Shop', 'Greek Restaurant', 'Ice Cream Shop',
       'Italian Restaurant', 'Brewery', 'Fruit & Vegetable Store',
       'Yoga Studio', 'Restaurant', 'Pizza Place', 'Juice Bar',
       'Bookstore', 'Bubble Tea Shop', 'Dessert Shop',
       'Furniture / Home Store', 'Spa', 'Grocery Store', 'Coffee Shop',
       'Bakery', 'Caribbean Restaurant', 'Japanese Restaurant',
       'Indian Restaurant', 'Café', 'Lounge', 'Frozen Yogurt Shop',
       'Liquor Store', 'American Restaurant', 'Gym', 'Fish & Chips Shop',
       'Fast Food Restaurant', 'Sushi Restaurant', 'Park', 'Pet Store',
       'Steakhouse', 'Burrito Place', 'Movie Theater', 'Sandwich Place',
       'Board Shop', 'Fish Market', 'Gay Bar', 'Thai Restaurant',
       'Seafood Restaurant', 'Cheese Shop', 'Comfort Food Restaurant',
       'Middle Eastern Restaurant', 'Stationery Store', 'Coworking Space',
       'Wine Bar', 'Latin American Restaurant', 'Gastropub',
       'Gym / Fitness Center', 'Bar', 'Convenience Store', 'Bank',
       'Diner', 'Clothing Store', 'Swim School', 'Bus Line',
       'Food & Drink Shop', 'Breakfast Spot', 'Department Store', 'Hotel',
       'Chinese Restaurant', 'Salon / Barbershop', 'Mexican Restaurant',
       'Sporting Goods Shop', 'Shoe Store', 'Gift Shop',
       'Rental Car Location', 'Toy / Game Store', 'Gas Station',
       'Farmers Market', 'Gourmet Shop', 'Pharmacy', 'Deli / Bodega',
       'Tennis Court', 'Playground', 'Supermarket', 'Sports Bar',
       'Fried Chicken Joint', 'Athletics & Sports',
       'Vietnamese Restaurant', 'Light Rail Station', 'Bagel Shop',
       'Jewelry Store', 'General Entertainment', 'Butcher',
       'Taiwanese Restaurant', 'Market', 'Beer Store', 'Snack Place',
       'Dance Studio', 'Theme Restaurant', 'Beer Bar', 'Ramen Restaurant',
       'Burger Joint', 'Creperie'], dtype=object)
```

In [142]:
```
"Thai Restaurant" in toronto_venues['Venue Category'].unique()
```

Out[142]: True

In [145]:
```
toronto_onehot = pd.get_dummies(toronto_venues[['Venue Category']], prefix="", prefix_sep="")

toronto_onehot['Neighborhood'] = toronto_venues['Neighborhood']

fixed_columns = [toronto_onehot.columns[-1]] + list(toronto_onehot.columns[:-1])
toronto_onehot = toronto_onehot[fixed_columns]

print(toronto_onehot.shape)
toronto_onehot.head()
```

```
(1623, 236)
```

Out[145]:

	Yoga Studio	Afghan Restaurant	Airport	Airport Food Court	Airport Gate	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	Antique Shop	Aquarium	Art Gallery	Art Museum	Arts & Crafts Store	A Restau
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	

In [146]:
```
toronto_grouped = toronto_onehot.groupby('Neighborhood').mean().reset_index()
toronto_grouped
```

Out[146]:

	Neighborhood	Yoga Studio	Afghan Restaurant	Airport	Airport Food Court	Airport Gate	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	Antique Shop	Aquarium	Gall

0	Berczy Park	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.017
1	Brockton, Parkdale Village, Exhibition Place	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.000
2	Business reply mail Processing Centre, South C...	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.000
3	CN Tower, King and Spadina, Railway Lands, Har...	0.000000	0.000000	0.071429	0.071429	0.071429	0.142857	0.142857	0.071429	0.000000	0.000000	0.00	0.000
4	Central Bay Street	0.015385	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.000
5	Christie	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.000
6	Church and Wellesley	0.026316	0.013158	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.013158	0.000000	0.00	0.000
7	Commerce Court, Victoria Hotel	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.040000	0.000000	0.00	0.010
8	Davisville	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.028571	0.000000	0.00	0.000
9	Davisville North	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.000
10	Dufferin, Dovercourt Village	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.000
11	First Canadian Place, Underground city	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.030000	0.000000	0.00	0.010

12	Forest Hill North & West, Forest Hill Road Park	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.000(
13	Garden District, Ryerson	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.010(
14	Harbourfront East, Union Station, Toronto Islands	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.05	0.010(
15	High Park, The Junction South	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.040000	0.00	0.000(
16	India Bazaar, The Beaches West	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.000(
17	Kensington Market, Chinatown, Grange Park	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.000(
18	Lawrence Park	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.000(
19	Little Portugal, Trinity	0.022727	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.022(
20	Moore Park, Summerhill East	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.000(
21	North Toronto West, Lawrence Park	0.047619	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.000(
22	Parkdale, Roncesvalles	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.000(
	Queen's Park,												

23	Ontario Provincial Government	0.031250	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.0000
24	Regent Park, Harbourfront	0.022222	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.022222	0.00	0.0222
25	Richmond, Adelaide, King	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.021277	0.000000	0.00	0.0106
26	Rosedale	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.0000
27	Roselawn	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.0000
28	Runnymede, Swansea	0.029412	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.0000
29	St. James Town	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.037975	0.000000	0.00	0.0126
30	St. James Town, Cabbagetown	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.0000
31	Stn A PO Boxes	0.010309	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.010309	0.010309	0.00	0.0206
32	Studio District	0.025000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.050000	0.000000	0.00	0.0000
33	Summerhill West, Rathnelly, South Hill, Forest...	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.058824	0.000000	0.00	0.0000
34	The Annex, North Midtown, Yorkville	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.0000
35	The Beaches	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.0000
36	The Danforth West, Riverdale	0.023256	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.023256	0.000000	0.00	0.0000

37	Toronto Dominion Centre, Design Exchange	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.030000	0.000000	0.00	0.010(
38	University of Toronto, Harbord	0.029412	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.000(

In [147]:
```
toronto_grouped.shape
```

Out[147]:
```
(39, 236)
```

In [151]:
```
len(toronto_grouped[toronto_grouped['Thai Restaurant'] > 0])
```

Out[151]:
```
12
```

In [153]:
```
toronto_thai = toronto_grouped[['Neighborhood','Thai Restaurant']]
toronto_thai.head()
```

Out[153]:

	Neighborhood	Thai Restaurant
0	Berczy Park	0.017241
1	Brockton, Parkdale Village, Exhibition Place	0.000000
2	Business reply mail Processing Centre, South C...	0.000000
3	CN Tower, King and Spadina, Railway Lands, Har...	0.000000
4	Central Bay Street	0.030769

In [155]:
```
kclusters = 5

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

# run k-means clustering
kmeans = KMeans(n_clusters=kclusters, random_state=1)
kmeans.fit_transform(toronto_grouped_clustering)

kmeans.labels_[0:15]
```

Out[155]:
```
array([3, 1, 1, 1, 0, 1, 1, 3, 0, 1, 1, 3, 1, 4, 1], dtype=int32)
```

In [156]:
```
toronto_thai_merged = toronto_thai.copy()
toronto_thai_merged["Cluster Labels"] = kmeans.labels_
```

In [157]:
```
toronto_thai_merged.head(5)
```

Out[157]:

	Neighborhood	Thai Restaurant	Cluster Labels
0	Berczy Park	0.017241	3
1	Brockton, Parkdale Village, Exhibition Place	0.000000	1
2	Business reply mail Processing Centre, South C...	0.000000	1
3	CN Tower, King and Spadina, Railway Lands, Har...	0.000000	1
4	Central Bay Street	0.030769	0

In [158]:
```
toronto_thai_merged=toronto_thai_merged.join(toronto_venues.set_index("Neighborhood"), on="Neighborhood"
)
print(toronto_thai_merged.shape)
toronto_thai_merged.head()
```

```
(1623, 9)
```

Out[158]:

	Neighborhood	Thai Restaurant	Cluster Labels	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Berczy Park	0.017241	3	43.644771	-79.373306	LCBO	43.642944	-79.372440	Liquor Store
0	Berczy Park	0.017241	3	43.644771	-79.373306	The Keg Steakhouse + Bar - Esplanade	43.646712	-79.374768	Restaurant
0	Berczy Park	0.017241	3	43.644771	-79.373306	Meridian Hall	43.646292	-79.376022	Concert Hall
0	Berczy Park	0.017241	3	43.644771	-79.373306	Fresh On Front	43.647815	-79.374453	Vegetarian / Vegan Restaurant
0	Berczy Park	0.017241	3	43.644771	-79.373306	Hockey Hall Of Fame (Hockey Hall of Fame)	43.646974	-79.377323	Museum

In [159]:
```
toronto_thai_merged.sort_values(["Cluster Labels"], inplace=True)
toronto_thai_merged.head()
```

Out[159]:

	Neighborhood	Thai Restaurant	Cluster Labels	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
25	Richmond, Adelaide, King	0.031915	0	43.650571	-79.384568	Forno Cultura	43.648533	-79.382535	Bakery
25	Richmond, Adelaide, King	0.031915	0	43.650571	-79.384568	Starbucks	43.646891	-79.381871	Coffee Shop
25	Richmond, Adelaide, King	0.031915	0	43.650571	-79.384568	McEwan Foods	43.647160	-79.381044	Deli / Bodega
25	Richmond, Adelaide, King	0.031915	0	43.650571	-79.384568	Druxy's	43.648015	-79.379907	Deli / Bodega
25	Richmond, Adelaide, King	0.031915	0	43.650571	-79.384568	Bosk at Shangri-La	43.649023	-79.385826	Asian Restaurant

In [171]:
```
map_thai= folium.Map(location=[latitude,longitude],zoom_start=11)

markers_colors={}
markers_colors[0] = 'blue'
markers_colors[1] = 'green'
markers_colors[2] = 'red'
markers_colors[3] = 'yellow'
markers_colors[4] = 'cyan'
for lat, lng, cluster in zip(toronto_thai_merged['Neighborhood Latitude'],toronto_thai_merged['Neighborh
ood Longitude'],toronto_thai_merged['Cluster Labels']):
    folium.features.CircleMarker(
        [lat, lng],
        radius=5,
        color =markers_colors[cluster],
        fill_color=markers_colors[cluster],
        fill_opacity=0.7).add_to(map_thai)

map_thai
```

Out[171]: Make this Notebook Trusted to load map: File -> Trust Notebook

In [163]:
```
#Cluster 0
toronto_thai_merged.loc[(toronto_thai_merged['Cluster Labels'] ==0) & (toronto_thai_merged['Venue Category'] == 'Thai Restaurant') ]
```

Out[163]:

	Neighborhood	Thai Restaurant	Cluster Labels	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
8	Davisville	0.028571	0	43.704324	-79.388790	Thai Spicy House	43.701962	-79.387513	Thai Restaurant
32	Studio District	0.025000	0	43.659526	-79.340923	EAT BKK Thai Kitchen	43.660450	-79.343113	Thai Restaurant
25	Richmond, Adelaide, King	0.031915	0	43.650571	-79.384568	Pai	43.647923	-79.388579	Thai Restaurant
25	Richmond, Adelaide, King	0.031915	0	43.650571	-79.384568	Thai Island	43.649146	-79.383798	Thai Restaurant
25	Richmond, Adelaide, King	0.031915	0	43.650571	-79.384568	Ruby Thai (First Canadian Place)	43.649091	-79.381600	Thai Restaurant
4	Central Bay Street	0.030769	0	43.657952	-79.387383	Salad King	43.657601	-79.381620	Thai Restaurant
4	Central Bay Street	0.030769	0	43.657952	-79.387383	Thai Express	43.661630	-79.387340	Thai Restaurant

In [166]:
```
#Cluster 1
toronto_thai_merged.loc[(toronto_thai_merged['Cluster Labels'] ==1) & (toronto_thai_merged['Venue Category'] == 'Thai Restaurant')]
```

Out[166]:

Neighborhood	Thai Restaurant	Cluster Labels	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category

In [167]:
```
#Cluster 2
toronto_thai_merged.loc[(toronto_thai_merged['Cluster Labels'] ==2) & (toronto_thai_merged['Venue Category'] == 'Thai Restaurant')]
```

Out[167]:

	Neighborhood	Thai Restaurant	Cluster Labels	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
15	High Park, The Junction South	0.08	2	43.661608	-79.464763	Isaan Der	43.665311	-79.468078	Thai Restaurant
15	High Park, The Junction South	0.08	2	43.661608	-79.464763	Silk	43.665291	-79.466238	Thai Restaurant

In [168]:
```
#Cluster 3
toronto_thai_merged.loc[(toronto_thai_merged['Cluster Labels'] ==3) & (toronto_thai_merged['Venue Category'] == 'Thai Restaurant')]
```

Out[168]:

	Neighborhood	Thai Restaurant	Cluster Labels	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
30	St. James Town, Cabbagetown	0.020833	3	43.667967	-79.367675	Thai Room - Carlton	43.664159	-79.368189	Thai Restaurant
0	Berczy Park	0.017241	3	43.644771	-79.373306	Sukhothai	43.648487	-79.374547	Thai Restaurant
7	Commerce Court, Victoria Hotel	0.020000	3	43.648198	-79.379817	Sukhothai	43.648487	-79.374547	Thai Restaurant
7	Commerce Court, Victoria Hotel	0.020000	3	43.648198	-79.379817	Ruby Thai (First Canadian Place)	43.649091	-79.381600	Thai Restaurant
11	First Canadian Place, Underground city	0.020000	3	43.648429	-79.382280	Ruby Thai (First Canadian Place)	43.649091	-79.381600	Thai Restaurant
11	First Canadian Place, Underground city	0.020000	3	43.648429	-79.382280	Thai Island	43.649146	-79.383798	Thai Restaurant

In [169]:
```
#Cluster 4
toronto_thai_merged.loc[(toronto_thai_merged['Cluster Labels'] ==4) & (toronto_thai_merged['Venue Category'] == 'Thai Restaurant')]
```

Out[169]:

	Neighborhood	Thai Restaurant	Cluster Labels	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
13	Garden District, Ryerson	0.010000	4	43.657162	-79.378937	Salad King	43.657601	-79.381620	Thai Restaurant
29	St. James Town	0.012658	4	43.651494	-79.375418	Sukhothai	43.648487	-79.374547	Thai Restaurant
31	Stn A PO Boxes	0.010309	4	43.646435	-79.374846	Sukhothai	43.648487	-79.374547	Thai Restaurant

Summary

There's many Thai restaurants are located nearby Commerce Court (cluster 3)and Richmond, Adelaide, King(cluster 0) However, there's only few Thai restaurants nearby High Park(ckuster 2). It's a good opportunity to open a Thai restaurant nearby. Meanwhile, there's no any Thai restaurant nearby East York(cluster 1). Will recommand to open Thai restaurant nearby East York as well.