3rd week assignment: Clustering Toronto

by Arturo López

In [1]:
```
import pandas as pd
```

Get data from wikipedia and generate a data frame

In [2]:
```
#get the table from wikipedia and create a dataframe
dfs=pd.read_html("https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M")
df=dfs[0]
df=df.rename(columns={'Neighbourhood': 'Neighborhood'})
df.head()
```

Out[2]:

	Postal Code	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

The table contains rows lacking data. If row has no Borough, drop it. If no neighborhood is assigned, assume its borough.

In [3]:
```
#clean the data frame
df.drop(df.loc[df['Borough']=='Not assigned'].index, inplace=True)
df.set_index(['Neighborhood']).replace('Not assigned', df.Borough)
df.head()
```

Out[3]:

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

In [4]:
```
df.shape
```
Out[4]: (103, 3)

Getting geographic data for each postal code.

In [5]:
```
locations=pd.read_csv('https://cocl.us/Geospatial_data')
```

In [6]:
```
df_loc=pd.merge(df, locations)
df_loc.head()
```

Out[6]:

	Postal Code	Borough	Neighborhood	Latitude	Longitude
0	M3A	North York	Parkwoods	43.753259	-79.329656
1	M4A	North York	Victoria Village	43.725882	-79.315572
2	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636
3	M6A	North York	Lawrence Manor, Lawrence Heights	43.718518	-79.464763
4	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government	43.662301	-79.389494

In [7]:
```
df_loc.shape
```
Out[7]: (103, 5)

Finally, segmenting and clustering neighborhoods in Toronto

In [8]:

```
#import required libraries
import numpy as np
from geopy.geocoders import Nominatim
import requests
import matplotlib.cm as cm
import matplotlib.colors as colors
from sklearn.cluster import KMeans
!pip install folium
import folium

print('libraries ready!')
```

```
Requirement already satisfied: folium in /opt/conda/envs/Python-3.7
-main/lib/python3.7/site-packages (0.12.0)
Requirement already satisfied: numpy in /opt/conda/envs/Python-3.7-
main/lib/python3.7/site-packages (from folium) (1.18.5)
Requirement already satisfied: jinja2>=2.9 in /opt/conda/envs/Pytho
n-3.7-main/lib/python3.7/site-packages (from folium) (2.11.2)
Requirement already satisfied: branca>=0.3.0 in /opt/conda/envs/Pyt
hon-3.7-main/lib/python3.7/site-packages (from folium) (0.4.2)
Requirement already satisfied: requests in /opt/conda/envs/Python-
3.7-main/lib/python3.7/site-packages (from folium) (2.24.0)
Requirement already satisfied: MarkupSafe>=0.23 in /opt/conda/envs/
Python-3.7-main/lib/python3.7/site-packages (from jinja2>=2.9->fol
ium) (1.1.1)
Requirement already satisfied: certifi>=2017.4.17 in /opt/conda/envs
/Python-3.7-main/lib/python3.7/site-packages (from requests->foliu
m) (2020.12.5)
Requirement already satisfied: idna<3,>=2.5 in /opt/conda/envs/Pyth
on-3.7-main/lib/python3.7/site-packages (from requests->folium)
(2.9)
Requirement already satisfied: chardet<4,>=3.0.2 in /opt/conda/envs
/Python-3.7-main/lib/python3.7/site-packages (from requests->foliu
m) (3.0.4)
Requirement already satisfied: urllib3!=1.25.0,!=1.25.1,<1.26,>=1.2
1.1 in /opt/conda/envs/Python-3.7-main/lib/python3.7/site-packages
(from requests->folium) (1.25.9)
libraries ready!
```

In [9]:

```
neighborhoods=df_loc.copy()
neighborhoods.head()
```

Out[9]:

	Postal Code	Borough	Neighborhood	Latitude	Longitude
0	M3A	North York	Parkwoods	43.753259	-79.329656
1	M4A	North York	Victoria Village	43.725882	-79.315572
2	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636
3	M6A	North York	Lawrence Manor, Lawrence Heights	43.718518	-79.464763
4	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government	43.662301	-79.389494

In [10]:
```
# Let's get the coordinates of Toronto
address = 'Toronto, TO'

geolocator = Nominatim(user_agent="to_explorer")
location = geolocator.geocode(address)
latitude = location.latitude
longitude = location.longitude
print('The geographical coordinate of Toronto are {}, {}.'.format(lat
itude, longitude))
```

```
The geograpical coordinate of Toronto are 43.65238435, -79.3835676
5.
```

Let's create a map of Toronto with its neighborhoods superimposed.

In [11]:
```
# 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(neighborhoods['Latitude
'], neighborhoods['Longitude'], neighborhoods['Borough'], neighborho
ods['Neighborhood']):
    label = '{}, {}'.format(neighborhood, borough)
    label = folium.Popup(label, parse_html=True)
    folium.CircleMarker(
        [lat, lng],
        radius=4,
        popup=label,
        color='purple',
        fill=True,
        fill_color='#3186cc',
        fill_opacity=0.5,
        parse_html=False).add_to(map_toronto)

map_toronto
```

Out[11]: Make this Notebook Trusted to load map: File -> Trust Notebook

The analysis will be focused on boroughs containing Toronto in its name. This will reduce the number of total neighborhoods but it will be easier to look at.

In [12]:
```
toronto_data = neighborhoods[neighborhoods['Borough'].str.contains('
Toronto')].reset_index(drop=True)
toronto_data.head()
```

Out[12]:

	Postal Code	Borough	Neighborhood	Latitude	Longitude
0	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636
1	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government	43.662301	-79.389494
2	M5B	Downtown Toronto	Garden District, Ryerson	43.657162	-79.378937
3	M5C	Downtown Toronto	St. James Town	43.651494	-79.375418
4	M4E	East Toronto	The Beaches	43.676357	-79.293031

As said, let's reduce the markers down to Toronto boroughs and create a new map. It can also help to see how many boroughs are we missing reducing the information to Toronto Boroughs.

In [13]:
```
# create map of Manhattan using latitude and longitude values
map_onlytoronto = folium.Map(location=[latitude, longitude], zoom_start=11)

# add markers to map
for lat, lng, label in zip(toronto_data['Latitude'], toronto_data['Longitude'], toronto_data['Neighborhood']):
    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_onlytoronto)

map_onlytoronto
```

Out[13]: Make this Notebook Trusted to load map: File -> Trust Notebook

Let's get Foursquare ready to use:

In [14]:
```
# @Hidden_cell
#Contains my Foursquare credentials
CLIENT_ID = 'DGDP2KPJ52CCMD51CBTR5GWF4BOBJ4TYGHRXPOY21GMX3HWU' # my Foursquare ID
CLIENT_SECRET = 'ISTM3KRMYXFT40KC3DGTVUV0NF2K2GBBFKX2M1Z5XMJK1GVZ' # my Foursquare Secret
VERSION = '20180605' # Foursquare API version
LIMIT = 100 # A default Foursquare API limit value
```

Now let's add a limit to venues results and a maximum radius where foursquare should look for each neighborhood.

In [15]:
```
LIMIT=100
radius=500
```

I will use the function define in the Lab so it looks for all the venues in the defined radius for every neighborhood.

In [16]:
```
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)
```

Now we use the function:

In [17]:
```
toronto_venues = getNearbyVenues(names=toronto_data['Neighborhood'],
                                   latitudes=toronto_data['Latitude
'],
                                   longitudes=toronto_data['Longitud
e']
                                  )
```

```
Regent Park, Harbourfront
Queen's Park, Ontario Provincial Government
Garden District, Ryerson
St. James Town
The Beaches
Berczy Park
Central Bay Street
Christie
Richmond, Adelaide, King
Dufferin, Dovercourt Village
Harbourfront East, Union Station, Toronto Islands
Little Portugal, Trinity
The Danforth West, Riverdale
Toronto Dominion Centre, Design Exchange
Brockton, Parkdale Village, Exhibition Place
India Bazaar, The Beaches West
Commerce Court, Victoria Hotel
Studio District
Lawrence Park
Roselawn
Davisville North
Forest Hill North & West, Forest Hill Road Park
High Park, The Junction South
North Toronto West, Lawrence Park
The Annex, North Midtown, Yorkville
Parkdale, Roncesvalles
Davisville
University of Toronto, Harbord
Runnymede, Swansea
Moore Park, Summerhill East
Kensington Market, Chinatown, Grange Park
Summerhill West, Rathnelly, South Hill, Forest Hill SE, Deer Park
CN Tower, King and Spadina, Railway Lands, Harbourfront West, Bath
urst Quay, South Niagara, Island airport
Rosedale
Stn A PO Boxes
St. James Town, Cabbagetown
First Canadian Place, Underground city
Church and Wellesley
Business reply mail Processing Centre, South Central Letter Proces
sing Plant Toronto
```

In [18]:
```
toronto_venues.groupby('Neighborhood').count()
```

Out[18]:

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Berczy Park	57	57	57	57	57	57
Brockton, Parkdale Village, Exhibition Place	24	24	24	24	24	24
Business reply mail Processing Centre, South Central Letter Processing Plant Toronto	16	16	16	16	16	16
CN Tower, King and Spadina, Railway Lands, Harbourfront West, Bathurst Quay, South Niagara, Island airport	15	15	15	15	15	15
Central Bay Street	61	61	61	61	61	61
Christie	16	16	16	16	16	16
Church and Wellesley	79	79	79	79	79	79
Commerce Court, Victoria Hotel	100	100	100	100	100	100
Davisville	33	33	33	33	33	33
Davisville North	9	9	9	9	9	9
Dufferin, Dovercourt Village	16	16	16	16	16	16
First Canadian Place, Underground city	100	100	100	100	100	100
Forest Hill North & West, Forest Hill Road Park	4	4	4	4	4	4
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	24	24	24	24	24	24
India Bazaar, The Beaches West	20	20	20	20	20	20

	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
Kensington Market, Chinatown, Grange Park	64	64	64	64	64	64
Lawrence Park	4	4	4	4	4	4
Little Portugal, Trinity	43	43	43	43	43	43
Moore Park, Summerhill East	1	1	1	1	1	1
North Toronto West, Lawrence Park	21	21	21	21	21	21
Parkdale, Roncesvalles	14	14	14	14	14	14
Queen's Park, Ontario Provincial Government	35	35	35	35	35	35
Regent Park, Harbourfront	46	46	46	46	46	46
Richmond, Adelaide, King	97	97	97	97	97	97
Rosedale	4	4	4	4	4	4
Roselawn	4	4	4	4	4	4
Runnymede, Swansea	33	33	33	33	33	33
St. James Town	80	80	80	80	80	80
St. James Town, Cabbagetown	46	46	46	46	46	46
Stn A PO Boxes	97	97	97	97	97	97
Studio District	37	37	37	37	37	37
Summerhill West, Rathnelly, South Hill, Forest Hill SE, Deer Park	15	15	15	15	15	15
The Annex, North Midtown, Yorkville	20	20	20	20	20	20
The Beaches	4	4	4	4	4	4

In [19]:
```
print('There are {} uniques categories.'.format(len(toronto_venues['
Venue Category'].unique())))
```

```
There are 234 uniques categories.
```

Neighborhood analysis.

In [20]:
```
# 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.co
lumns[:-1])
toronto_onehot = toronto_onehot[fixed_columns]

toronto_onehot.head()
```

Out[20]:

	Neighborhood	_Airport	_Airport Food Court	_Airport Lounge	_Airport Service	_Airport Terminal	_American Restaurant	_Antique Shop	_
0	Regent Park, Harbourfront	0	0	0	0	0	0	0	
1	Regent Park, Harbourfront	0	0	0	0	0	0	0	
2	Regent Park, Harbourfront	0	0	0	0	0	0	0	
3	Regent Park, Harbourfront	0	0	0	0	0	0	0	
4	Regent Park, Harbourfront	0	0	0	0	0	0	0	

5 rows × 235 columns

In [21]:
```
toronto_onehot.shape
```

Out[21]: (1615, 235)

Group rows by neighborhood and by taking the mean of the frequency of occurence of each category.

In [22]:
```
toronto_grouped = toronto_onehot.groupby('Neighborhood').mean().rese
t_index()
toronto_grouped
```

Out[22]:

	Neighborhood	_Airport	_Airport Food Court	_Airport Lounge	_Airport Service	_Airport Terminal	_American Restaurant	_Antique Shop
0	Berczy Park	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	0.000000
1	Brockton, Parkdale Village, Exhibition Place	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	0.000000
2	Business reply mail Processing Centre, South C...	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	0.000000
3	CN Tower, King and Spadina, Railway Lands, Har...	0.066667	0.066667	0.133333	0.2	0.133333	0.000000	0.000000
4	Central Bay Street	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	0.000000
5	Christie	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	0.000000
6	Church and Wellesley	0.000000	0.000000	0.000000	0.0	0.000000	0.012658	0.000000
7	Commerce Court, Victoria Hotel	0.000000	0.000000	0.000000	0.0	0.000000	0.040000	0.000000
8	Davisville	0.000000	0.000000	0.000000	0.0	0.000000	0.030303	0.000000
9	Davisville North	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	0.000000
10	Dufferin, Dovercourt Village	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	0.000000
11	First Canadian Place, Underground city	0.000000	0.000000	0.000000	0.0	0.000000	0.030000	0.000000
12	Forest Hill North & West, Forest Hill Road Park	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	0.000000
13	Garden District, Ryerson	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	0.000000
14	Harbourfront East, Union Station, Toronto Islands	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	0.000000

	Neighborhood	_Airport	_Airport Food Court	_Airport Lounge	_Airport Service	_Airport Terminal	_American Restaurant	_Antique Shop
15	High Park, The Junction South	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	0.041667
16	India Bazaar, The Beaches West	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	0.000000
17	Kensington Market, Chinatown, Grange Park	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	0.000000
18	Lawrence Park	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	0.000000
19	Little Portugal, Trinity	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	0.000000
20	Moore Park, Summerhill East	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	0.000000
21	North Toronto West, Lawrence Park	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	0.000000
22	Parkdale, Roncesvalles	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	0.000000
23	Queen's Park, Ontario Provincial Government	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	0.000000
24	Regent Park, Harbourfront	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	0.021739
25	Richmond, Adelaide, King	0.000000	0.000000	0.000000	0.0	0.000000	0.020619	0.000000
26	Rosedale	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	0.000000
27	Roselawn	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	0.000000
28	Runnymede, Swansea	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	0.000000
29	St. James Town	0.000000	0.000000	0.000000	0.0	0.000000	0.037500	0.000000
30	St. James Town, Cabbagetown	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	0.000000
31	Stn A PO Boxes	0.000000	0.000000	0.000000	0.0	0.000000	0.010309	0.010309
32	Studio District	0.000000	0.000000	0.000000	0.0	0.000000	0.054054	0.000000
33	Summerhill West, Rathnelly, South Hill, Forest...	0.000000	0.000000	0.000000	0.0	0.000000	0.066667	0.000000
34	The Annex, North Midtown, Yorkville	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	0.000000

	Neighborhood	_Airport	_Airport Food Court	_Airport Lounge	_Airport Service	_Airport Terminal	_American Restaurant	_Antique Shop
35	The Beaches	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	0.000000

In [23]:
```
toronto_grouped.shape
```

Out[23]: (39, 235)

Which are the 5 most common venues for each neighborhood?

In [24]:
```
num_top_venues = 5

for hood in toronto_grouped['Neighborhood']:
    print("----"+hood+"----")
    temp = toronto_grouped[toronto_grouped['Neighborhood'] == hood].
T.reset_index()
    temp.columns = ['venue','freq']
    temp = temp.iloc[1:]
    temp['freq'] = temp['freq'].astype(float)
    temp = temp.round({'freq': 2})
    print(temp.sort_values('freq', ascending=False).reset_index(drop
=True).head(num_top_venues))
    print('\n')
```

```
----Berczy Park----
                  venue  freq
0          _Coffee Shop  0.09
1         _Cocktail Bar  0.05
2             _Beer Bar  0.04
3  _Seafood Restaurant  0.04
4          _Cheese Shop  0.04


----Brockton, Parkdale Village, Exhibition Place----
                  venue  freq
0                 _Café  0.12
1       _Breakfast Spot  0.08
2            _Nightclub  0.08
3          _Coffee Shop  0.08
4  _Italian Restaurant  0.04


----Business reply mail Processing Centre, South Central Letter Pr
ocessing Plant Toronto----
                    venue  freq
0     _Light Rail Station  0.12
1            _Pizza Place  0.06
2          _Auto Workshop  0.06
3  _Fast Food Restaurant  0.06
4          _Burrito Place  0.06


----CN Tower, King and Spadina, Railway Lands, Harbourfront West,
Bathurst Quay, South Niagara, Island airport----
                venue  freq
0    _Airport Service  0.20
1     _Airport Lounge  0.13
2  _Airport Terminal  0.13
3            _Airport  0.07
4       _Boat or Ferry  0.07


----Central Bay Street----
                  venue  freq
0          _Coffee Shop  0.20
1       _Sandwich Place  0.05
2                 _Café  0.05
3  _Italian Restaurant  0.05
4     _Thai Restaurant  0.03


----Christie----
             venue  freq
0  _Grocery Store  0.25
1            _Café  0.19
2            _Park  0.12
3    _Coffee Shop  0.06
4      _Nightclub  0.06


----Church and Wellesley----
                  venue  freq
0          _Coffee Shop  0.08
```

```
1  _Japanese Restaurant  0.06
2     _Sushi Restaurant  0.06
3              _Gay Bar  0.04
4           _Restaurant  0.04


----Commerce Court, Victoria Hotel----
                  venue  freq
0          _Coffee Shop  0.12
1           _Restaurant  0.07
2                _Hotel  0.06
3                 _Café  0.06
4  _American Restaurant  0.04


----Davisville----
             venue  freq
0    _Dessert Shop  0.09
1  _Sandwich Place  0.09
2            _Café  0.06
3     _Coffee Shop  0.06
4     _Pizza Place  0.06


----Davisville North----
               venue  freq
0       _Pizza Place  0.11
1              _Park  0.11
2    _Sandwich Place  0.11
3    _Breakfast Spot  0.11
4  _Department Store  0.11


----Dufferin, Dovercourt Village----
          venue  freq
0     _Pharmacy  0.12
1       _Bakery  0.12
2         _Bank  0.06
3  _Music Venue  0.06
4          _Bar  0.06


----First Canadian Place, Underground city----
                  venue  freq
0          _Coffee Shop  0.10
1                 _Café  0.07
2                _Hotel  0.05
3           _Restaurant  0.04
4  _Japanese Restaurant  0.04


----Forest Hill North & West, Forest Hill Road Park----
                 venue  freq
0  _Mexican Restaurant  0.25
1               _Trail  0.25
2        _Jewelry Store  0.25
3     _Sushi Restaurant  0.25
4             _Airport  0.00
```

```
----Garden District, Ryerson----
              venue  freq
0   _Clothing Store  0.09
1      _Coffee Shop  0.09
2   _Cosmetics Shop  0.03
3            _Hotel  0.03
4  _Bubble Tea Shop  0.03


----Harbourfront East, Union Station, Toronto Islands----
                 venue  freq
0         _Coffee Shop  0.13
1            _Aquarium  0.05
2                _Café  0.04
3               _Hotel  0.04
4  _Italian Restaurant  0.03


----High Park, The Junction South----
                 venue  freq
0  _Mexican Restaurant  0.08
1     _Thai Restaurant  0.08
2                _Café  0.08
3         _Flea Market  0.04
4           _Speakeasy  0.04


----India Bazaar, The Beaches West----
                   venue  freq
0  _Fast Food Restaurant  0.10
1        _Sandwich Place  0.05
2          _Liquor Store  0.05
3         _Movie Theater  0.05
4      _Sushi Restaurant  0.05


----Kensington Market, Chinatown, Grange Park----
                            venue  freq
0                           _Café  0.08
1                    _Coffee Shop  0.06
2             _Mexican Restaurant  0.05
3          _Vietnamese Restaurant  0.05
4  _Vegetarian / Vegan Restaurant  0.05


----Lawrence Park----
               venue  freq
0              _Park  0.25
1          _Bus Line  0.25
2  _Business Service  0.25
3       _Swim School  0.25
4           _Airport  0.00


----Little Portugal, Trinity----
                            venue  freq
0                            _Bar  0.09
1                    _Coffee Shop  0.07
```

```
2               _Asian Restaurant  0.05
3          _Vietnamese Restaurant  0.05
4  _Vegetarian / Vegan Restaurant  0.05


----Moore Park, Summerhill East----
                 venue  freq
0          _Summer Camp   1.0
1              _Airport   0.0
2     _Malay Restaurant   0.0
3  _Martial Arts School   0.0
4       _Massage Studio   0.0


----North Toronto West, Lawrence Park----
                 venue  freq
0       _Clothing Store  0.14
1          _Coffee Shop  0.10
2          _Yoga Studio  0.05
3  _Sporting Goods Shop  0.05
4                 _Park  0.05


----Parkdale, Roncesvalles----
            venue  freq
0  _Breakfast Spot  0.14
1       _Gift Shop  0.14
2     _Coffee Shop  0.07
3            _Bar  0.07
4        _Dog Run  0.07


----Queen's Park, Ontario Provincial Government----
               venue  freq
0        _Coffee Shop  0.20
1   _Sushi Restaurant  0.06
2        _Yoga Studio  0.03
3  _College Cafeteria  0.03
4      _Burrito Place  0.03


----Regent Park, Harbourfront----
         venue  freq
0  _Coffee Shop  0.15
1       _Bakery  0.07
2         _Café  0.07
3         _Park  0.07
4          _Pub  0.04


----Richmond, Adelaide, King----
              venue  freq
0      _Coffee Shop  0.08
1             _Café  0.05
2       _Restaurant  0.04
3  _Thai Restaurant  0.03
4              _Gym  0.03
```

```
----Rosedale----
            venue  freq
0           _Park  0.50
1     _Playground  0.25
2          _Trail  0.25
3        _Airport  0.00
4  _Movie Theater  0.00


----Roselawn----
                      venue  freq
0  _Health & Beauty Service  0.25
1             _Home Service  0.25
2           _Ice Cream Shop  0.25
3                   _Garden  0.25
4    _Performing Arts Venue  0.00


----Runnymede, Swansea----
                venue  freq
0               _Café  0.09
1        _Coffee Shop  0.09
2  _Italian Restaurant  0.06
3    _Sushi Restaurant  0.06
4                _Pub  0.06


----St. James Town----
                 venue  freq
0         _Coffee Shop  0.06
1                _Café  0.05
2        _Cocktail Bar  0.04
3  _American Restaurant  0.04
4           _Gastropub  0.04


----St. James Town, Cabbagetown----
         venue  freq
0  _Restaurant  0.07
1 _Pizza Place  0.07
2 _Coffee Shop  0.07
3        _Park  0.04
4   _Pet Store  0.04


----Stn A PO Boxes----
                venue  freq
0        _Coffee Shop  0.10
1         _Restaurant  0.03
2           _Beer Bar  0.03
3  _Italian Restaurant  0.03
4              _Hotel  0.03


----Studio District----
                 venue  freq
0         _Coffee Shop  0.08
1                _Café  0.05
2  _American Restaurant  0.05
```

```
3             _Gastropub  0.05
4               _Brewery  0.05


----Summerhill West, Rathnelly, South Hill, Forest Hill SE, Deer P
ark----
           venue  freq
0   _Coffee Shop  0.13
1          _Bank  0.07
2  _Liquor Store  0.07
3   _Supermarket  0.07
4    _Restaurant  0.07


----The Annex, North Midtown, Yorkville----
             venue  freq
0  _Sandwich Place  0.15
1            _Café  0.15
2     _Coffee Shop  0.10
3  _History Museum  0.05
4       _BBQ Joint  0.05


----The Beaches----
                venue  freq
0       _Neighborhood  0.25
1  _Health Food Store  0.25
2              _Trail  0.25
3                _Pub  0.25
4             _Museum  0.00


----The Danforth West, Riverdale----
                     venue  freq
0        _Greek Restaurant  0.19
1             _Coffee Shop  0.07
2      _Italian Restaurant  0.07
3  _Furniture / Home Store  0.05
4           _Ice Cream Shop  0.05


----Toronto Dominion Centre, Design Exchange----
                  venue  freq
0          _Coffee Shop  0.10
1                _Hotel  0.08
2                 _Café  0.06
3  _American Restaurant  0.03
4   _Italian Restaurant  0.03


----University of Toronto, Harbord----
                  venue  freq
0                 _Café  0.15
1                  _Bar  0.06
2   _Italian Restaurant  0.06
```

The 5 most popular venues of each neighborhood fitted into a dataframe

In [25]:
```
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]
```

In [26]:
```
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.head()
```

Out[26]:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Cor \
0	Berczy Park	_Coffee Shop	_Cocktail Bar	_Beer Bar	_Bakery	_Restaurant	_Cheese Shop	_Fa M
1	Brockton, Parkdale Village, Exhibition Place	_Café	_Breakfast Spot	_Nightclub	_Coffee Shop	_Performing Arts Venue	_Furniture / Home Store	_E
2	Business reply mail Processing Centre, South C...	_Light Rail Station	_Gym / Fitness Center	_Garden	_Brewery	_Spa	_Farmers Market	Resta
3	CN Tower, King and Spadina, Railway Lands, Har...	_Airport Service	_Airport Lounge	_Airport Terminal	_Airport	_Boat or Ferry	_Plane	_I Lo
4	Central Bay Street	_Coffee Shop	_Sandwich Place	_Café	_Italian Restaurant	_Burger Joint	_Bubble Tea Shop	_

Let's cluster the neighborhoods

For that K-means will be the choice

In [27]:
```
# set number of clusters
kclusters = 6

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

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

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

Out[27]:
```
array([3, 3, 3, 3, 0, 3, 3, 3, 3, 3], dtype=int32)
```

Create a dataframe with the 10 most common venues and its cluster

In [28]:
```
# add clustering labels
neighborhoods_venues_sorted.insert(0, 'Cluster Labels', kmeans.label
s_)

toronto_merged = toronto_data

# merge manhattan_grouped with manhattan_data to add latitude/longit
ude for each neighborhood
toronto_merged = toronto_merged.join(neighborhoods_venues_sorted.set
_index('Neighborhood'), on='Neighborhood')

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

Out[28]:

	Postal Code	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd M Comn Ver
0	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636	0	_Coffee Shop	_C
1	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government	43.662301	-79.389494	0	_Coffee Shop	_Su Restau
2	M5B	Downtown Toronto	Garden District, Ryerson	43.657162	-79.378937	3	_Coffee Shop	_Cloth St
3	M5C	Downtown Toronto	St. James Town	43.651494	-79.375418	3	_Coffee Shop	_C
4	M4E	East Toronto	The Beaches	43.676357	-79.293031	4	_Neighborhood	_He F St

Now let's check the resulting clusters

In [29]:

```
# create 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['Neighborhood'], 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[29]: Make this Notebook Trusted to load map: File -> Trust Notebook

Startint with 5 clusters shows that most of the neighborhoods fall into the same cluster. Maybe KMeans needs a higher numbers of clusters to improve the fitting. So the second round I used 6 clusters.

We look into the clusters and see their common venues.

Cluster 1

In [30]:
```
toronto_merged.loc[toronto_merged['Cluster Labels'] == 0, toronto_me
rged.columns[[1] + list(range(5, toronto_merged.shape[1]))]]
```

Out[30]:

	Borough	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue
0	Downtown Toronto	0	_Coffee Shop	_Café	_Park	_Bakery	_Breakfast Spot	_Pub
1	Downtown Toronto	0	_Coffee Shop	_Sushi Restaurant	_Yoga Studio	_Fried Chicken Joint	_Beer Bar	_Japanese Restaurant
6	Downtown Toronto	0	_Coffee Shop	_Sandwich Place	_Café	_Italian Restaurant	_Burger Joint	_Bubble Tea Shop
24	Central Toronto	0	_Café	_Sandwich Place	_Coffee Shop	_Liquor Store	_Indian Restaurant	_Pub
28	West Toronto	0	_Café	_Coffee Shop	_Sushi Restaurant	_Pub	_Italian Restaurant	_Yoga Studio
31	Central Toronto	0	_Coffee Shop	_Bagel Shop	_Fried Chicken Joint	_Liquor Store	_Sandwich Place	_Restaurant

Cluster 2

In [31]:
```
toronto_merged.loc[toronto_merged['Cluster Labels'] == 1, toronto_me
rged.columns[[1] + list(range(5, toronto_merged.shape[1]))]]
```

Out[31]:

	Borough	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7t Cc
29	Central Toronto	1	_Summer Camp	_Yoga Studio	_Farmers Market	_Event Space	_Ethiopian Restaurant	_Escape Room	_Ele

Cluster 3

In [32]:
```
toronto_merged.loc[toronto_merged['Cluster Labels'] == 2, toronto_merged.columns[[1] + list(range(5, toronto_merged.shape[1]))]]
```

Out[32]:

	Borough	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue
18	Central Toronto	2	_Park	_Bus Line	_Business Service	_Swim School	_Falafel Restaurant	_Ethiopian Restaurant
33	Downtown Toronto	2	_Park	_Playground	_Trail	_Yoga Studio	_Deli / Bodega	_Escape Room

Cluster 4

In [33]:
```
toronto_merged.loc[toronto_merged['Cluster Labels'] == 3, toronto_me
rged.columns[[1] + list(range(5, toronto_merged.shape[1]))]]
```

Out[33]:

	Borough	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6 C
2	Downtown Toronto	3	_Coffee Shop	_Clothing Store	_Japanese Restaurant	_Café	_Cosmetics Shop	Re
3	Downtown Toronto	3	_Coffee Shop	_Café	_Gastropub	_American Restaurant	_Cocktail Bar	
5	Downtown Toronto	3	_Coffee Shop	_Cocktail Bar	_Beer Bar	_Bakery	_Restaurant	-
7	Downtown Toronto	3	_Grocery Store	_Café	_Park	_Nightclub	_Italian Restaurant	_Re
8	Downtown Toronto	3	_Coffee Shop	_Café	_Restaurant	_Gym	_Clothing Store	
9	West Toronto	3	_Pharmacy	_Bakery	_Grocery Store	_Pool	_Café	Re
10	Downtown Toronto	3	_Coffee Shop	_Aquarium	_Hotel	_Café	_Restaurant	
11	West Toronto	3	_Bar	_Coffee Shop	_Asian Restaurant	_Vietnamese Restaurant	_Restaurant	_Ve Re
12	East Toronto	3	_Greek Restaurant	_Coffee Shop	_Italian Restaurant	_Bookstore	_Ice Cream Shop	_Fu Hor
13	Downtown Toronto	3	_Coffee Shop	_Hotel	_Café	_Salad Place	_American Restaurant	_J. Re
14	West Toronto	3	_Café	_Breakfast Spot	_Nightclub	_Coffee Shop	_Performing Arts Venue	_Fu Hor
15	East Toronto	3	_Fast Food Restaurant	_Sushi Restaurant	_Pub	_Sandwich Place	_Burrito Place	
16	Downtown Toronto	3	_Coffee Shop	_Restaurant	_Café	_Hotel	_Italian Restaurant	_A Re
17	East Toronto	3	_Coffee Shop	_American Restaurant	_Bakery	_Brewery	_Café	_Ga
19	Central Toronto	3	_Garden	_Health & Beauty Service	_Ice Cream Shop	_Home Service	_Escape Room	_Ele
20	Central Toronto	3	_Gym / Fitness Center	_Breakfast Spot	_Hotel	_Food & Drink Shop	_Department Store	
22	West Toronto	3	_Café	_Mexican Restaurant	_Thai Restaurant	_Grocery Store	_Furniture / Home Store	_Fa Re
23	Central Toronto	3	_Clothing Store	_Coffee Shop	_Yoga Studio	_Sporting Goods Shop	_Grocery Store	_Ic
25	West Toronto	3	_Breakfast Spot	_Gift Shop	_Restaurant	_Cuban Restaurant	_Eastern European Restaurant	_I

	Borough	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6 C
26	Central Toronto	3	_Sandwich Place	_Dessert Shop	_Café	_Italian Restaurant	_Gym	Re
27	Downtown Toronto	3	_Café	_Bar	_Italian Restaurant	_Japanese Restaurant	_Bookstore	
30	Downtown Toronto	3	_Café	_Coffee Shop	_Mexican Restaurant	_Vietnamese Restaurant	_Vegetarian / Vegan Restaurant	_Ca Re
32	Downtown Toronto	3	_Airport Service	_Airport Lounge	_Airport Terminal	_Airport	_Boat or Ferry	
34	Downtown Toronto	3	_Coffee Shop	_Restaurant	_Japanese Restaurant	_Café	_Beer Bar	_
35	Downtown Toronto	3	_Coffee Shop	_Pizza Place	_Restaurant	_Italian Restaurant	_Park	_F
36	Downtown Toronto	3	_Coffee Shop	_Café	_Hotel	_Restaurant	_Gym	_J Re

Cluster 5

In [34]:
```
toronto_merged.loc[toronto_merged['Cluster Labels'] == 4, toronto_merged.columns[[1] + list(range(5, toronto_merged.shape[1]))]]
```

Out[34]:

	Borough	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7 C
4	East Toronto	4	_Neighborhood	_Health Food Store	_Pub	_Trail	_Yoga Studio	_Dog Run	

Cluster 6

In [35]:
```
toronto_merged.loc[toronto_merged['Cluster Labels'] == 5, toronto_merged.columns[[1] + list(range(5, toronto_merged.shape[1]))]]
```

Out[35]:

	Borough	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7t Co
21	Central Toronto	5	_Jewelry Store	_Trail	_Mexican Restaurant	_Sushi Restaurant	_Yoga Studio	_Dessert Shop	_Et Res

So here we can see that teo of the clusters take the majority of neighborhoods while the other 4 clusters are very unpopulated. There is not a gib variety of neighborhoods in Toronto from this analysis.