Data Science Capstone Week 3 Assignment - PART 3

Intent: To explore, segment, and cluster the neighborhoods in the city of Toronto

Description: The Foursquare API will be used to do the following.

- The explore function will be used to get the most common venue categories in each **Toronto** neighborhood,
- Using the clustering feature, the neighborhoods will be grouped into clusters.
- The clustering algorithm employed will be the k-means clustering algorithm.
- The Folium library will be used to visualize the neighborhoods in Toronto and their emerging clusters.

In [1]:

```
import numpy as np # library to handle data in a vectorized manner

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

import json # library to handle JSON files

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

print('Libraries imported.')
```

```
Libraries imported.
```

Read the Dataframe from ".csv" file that was created in Part 2 of this project

In [77]:

```
TorontoDF = pd.read_csv('/resources/data/DataScienceCAPSTONE/Week3/completePart2DF.csv')
TorontoDF.head()
```

Out[77]:

	Unnamed: 0	Postcode	Borough	Neighbourhood	Latitude	Longitude
0	0	M3A	North York	Parkwoods	43.753259	-79.329656
1	1	M4A	North York	Victoria Village	43.725882	-79.315572
2	2	M9A	Etobicoke	Islington Avenue	43.667856	-79.532242
3	3	M3B	North York	Don Mills North	43.745906	-79.352188
4	4	M6B	North York	Glencairn	43.709577	-79.445073

Extract only the rows with "Toronto" as a substring in Borough

** Find the records that have "Toronto" in Borough name; indicate non-presence with -1 in a new column called 'Indexes'

In [78]:

```
# substring to be searched
sub ='Toronto'

# creating and passing series to new column
TorontoDF["Indexes"]= TorontoDF["Borough"].str.find(sub)
TorontoDF.head()
```

Out[78]:

	Unnamed: 0	Postcode	Borough	Neighbourhood	Latitude	Longitude	Indexes
0	0	M3A	North York	Parkwoods	43.753259	-79.329656	-1
1	1	M4A	North York	Victoria Village	43.725882	-79.315572	-1
2	2	M9A	Etobicoke	Islington Avenue	43.667856	-79.532242	-1
3	3	M3B	North York	Don Mills North	43.745906	-79.352188	-1
4	4	M6B	North York	Glencairn	43.709577	-79.445073	-1

Extract only the rows where the 'Indexes' field is NOT -1

In [79]:

```
jendf = TorontoDF[TorontoDF.Indexes != (-1)]
jendf.head()
```

Out[79]:

	Unnamed: 0	Postcode	Borough	Neighbourhood	Latitude	Longitude	Indexes
6	6	M5C	Downtown Toronto	St. James Town	43.651494	-79.375418	9
8	8	M4E	East Toronto	The Beaches	43.676357	-79.293031	5
9	9	M5E	Downtown Toronto	Berczy Park	43.644771	-79.373306	9
13	13	M5G	Downtown Toronto	Central Bay Street	43.657952	-79.387383	9
14	14	M6G	Downtown Toronto	Christie	43.669542	-79.422564	9

Extract the Five desired columns: Postcode, Borough, Neighbourhood, Latitude, and Longitude

In [80]:

```
TorontoDF = jendf[['Postcode','Borough','Neighbourhood','Latitude','Longitude']]
TorontoDF.head()
```

Out[80]:

	Postcode	Borough	Neighbourhood	Latitude	Longitude
6	M5C	Downtown Toronto	St. James Town	43.651494	-79.375418
8	M4E	East Toronto	The Beaches	43.676357	-79.293031
9	M5E	Downtown Toronto	Berczy Park	43.644771	-79.373306
13	M5G	Downtown Toronto	Central Bay Street	43.657952	-79.387383
14	M6G	Downtown Toronto	Christie	43.669542	-79.422564

Use geopy library to get the latitude and longitude values of Toronto.

In [7]:

```
!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
```

```
Solving environment: done


==> WARNING: A newer version of conda exists. <==
  current version: 4.5.11
  latest version: 4.7.11

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-1.49         |             py_0          32 KB  conda
-forge
    ------------------------------------------------------------
                                           Total:          90 KB

The following NEW packages will be INSTALLED:

    geographiclib: 1.49-py_0   conda-forge
    geopy:         1.20.0-py_0 conda-forge


Downloading and Extracting Packages
geopy-1.20.0         | 57 KB     | ####################################
# | 100%
geographiclib-1.49   | 32 KB     | ####################################
# | 100%
Preparing transaction: done
Verifying transaction: done
Executing transaction: done
```

In [81]:

```
address = 'Toronto, Ontario'
geolocator = Nominatim(user_agent="toronto_explorer")
location = geolocator.geocode(address)
latitude = location.latitude
longitude = location.longitude
print('The geograpical coordinates of Toronto are {}, {}.'.format(latitude, longitu
de))
```

```
The geograpical coordinates of Toronto are 43.653963, -79.387207.
```

Create a map of Toronto with neighborhoods superimposed on top.

In [82]:

```
# 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, pc, borough, neighborhood in zip(TorontoDF['Latitude'], TorontoDF['Longitude'], TorontoDF['Postcode'],
                                               TorontoDF['Borough'], TorontoDF['Neighbourhood']):
    label = '{}-- {}, {}'.format(pc, 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
```

Out[82]:

1. Use Foursquare API to explore Toronto

Define Foursquare Credentials and Version

In [83]:

```
CLIENT_ID = '3PLDGNKB1ZAH12N1UDC3XOZIYLGYUADHUO2V5ZWRLMK0WCHL' # your Foursquare ID
CLIENT_SECRET = 'NLCEFCCXOGBVPLEJOFT34TOESGRGBSJDCPY5MO3IKTMTYDKT' # your Foursquar
e Secret
VERSION = '20190815' # Foursquare API version
print('... got your credentials!')
#print('Your credentails:')
#print('CLIENT_ID: ' + CLIENT_ID)
#print('CLIENT_SECRET:' + CLIENT_SECRET)
```

```
... got your credentials!
```

Explore the first neighborhood in our dataframe "TorontoDF"

Get the neighborhood's name.

In [84]:

```
neighborhood_name = TorontoDF.iloc[0,2]
neighborhood_name
```

Out[84]:

```
'St. James Town'
```

Get the neighborhood's latitude and longitude values.

In [85]:

```
neighborhood_latitude = TorontoDF.iloc[0,3] # neighborhood latitude value
neighborhood_longitude = TorontoDF.iloc[0,4] # neighborhood longitude value

#neighborhood_name = TorontoDF.loc[0,'Neighborhood'] # neighborhood name

print('Latitude and longitude values of {} are {}, {}.'.format(neighborhood_name,
                                                               neighborhood_latitud
e,
                                                               neighborhood_longitu
de))
```

```
Latitude and longitude values of St. James Town are 43.6514939, -79.375
4179.
```

EXPLORE ... get the top 100 venues that are in St.James Town within a radius of 500 meters.

In [86]:

```
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
```

Out[86]:

```
'https://api.foursquare.com/v2/venues/explore?&client_id=3PLDGNKB1ZAH12
N1UDC3XOZIYLGYUADHUO2V5ZWRLMK0WCHL&client_secret=NLCEFCCXOGBVPLEJOFT34T
OESGRGBSJDCPY5MO3IKTMTYDKT&v=20190815&ll=43.6514939,-79.3754179&radius=
500&limit=100'
```

Send the GET request and examine the results

In [87]:

```
results = requests.get(url).json() # VERY LARGE json file... comment display after
 viewing
#results
```

From the Foursquare lab in the previous module, we know that all the information is in the *items* key. Before we proceed, let's borrow the **get_category_type** function from the Foursquare lab.

In [88]:

```
# 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']
```

Clean the json file and put into a pandas dataframe

In [89]:

```
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()
```

Out[89]:

	name	categories	lat	lng
0	Crepe TO	Creperie	43.650063	-79.374587
1	Gyu-Kaku Japanese BBQ	Japanese Restaurant	43.651422	-79.375047
2	GoodLife Fitness Toronto 137 Yonge Street	Gym	43.651242	-79.378068
3	GEORGE Restaurant	Restaurant	43.653346	-79.374445
4	Terroni	Italian Restaurant	43.650927	-79.375602

In [90]:

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

```
100 venues were returned by Foursquare.
```

2. Explore Many Neighborhoods in Toronto

Create a function to repeat the same process to all the neighborhoods in Toronto

In [91]:

```
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_se
cret={}&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 ve
nue_list])
    nearby_venues.columns = ['Neighborhood',
                  'Neighborhood Latitude',
                  'Neighborhood Longitude',
                  'Venue',
                  'Venue Latitude',
                  'Venue Longitude',
                  'Venue Category']

    return(nearby_venues)
```

Run the above function on each neighborhood and create a new dataframe called *toronto_venues*.

In [92]:

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

```
St. James Town
The Beaches
Berczy Park
Central Bay Street
Christie
Studio District
Lawrence Park
Roselawn
Davisville North
North Toronto West
Davisville
Rosedale
Stn A PO Boxes 25 The Esplanade
Church and Wellesley
Business Reply Mail Processing Centre 969 Eastern
CN Tower, Bathurst Quay, Island airport, Harbourfront West, King and Sp
adina, Railway Lands, South Niagara
Deer Park, Forest Hill SE, Rathnelly, South Hill, Summerhill West
The Annex, North Midtown, Yorkville
Brockton, Exhibition Place, Parkdale Village
Harbourfront East, Toronto Islands, Union Station
Chinatown, Grange Park, Kensington Market
Adelaide, King, Richmond
Cabbagetown, St. James Town
Forest Hill North, Forest Hill West
Design Exchange, Toronto Dominion Centre
Harbourfront, Regent Park
The Beaches West, India Bazaar
First Canadian Place, Underground city
Moore Park, Summerhill East
Parkdale, Roncesvalles
Harbord, University of Toronto
Ryerson, Garden District
Little Portugal, Trinity
The Danforth West, Riverdale
High Park, The Junction South
Runnymede, Swansea
Dovercourt Village, Dufferin
Commerce Court, Victoria Hotel
```

In [93]:

```
toronto_venues.tail()
```

Out[93]:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
1704	Commerce Court, Victoria Hotel	43.648198	-79.379817	Tim Hortons	43.646256	-79.379573	Coffee Shop
1705	Commerce Court, Victoria Hotel	43.648198	-79.379817	Sam James Coffee Bar (SJCB)	43.650319	-79.376217	Coffee Shop
1706	Commerce Court, Victoria Hotel	43.648198	-79.379817	Szechuan Express	43.646973	-79.379549	Chinese Restaurant
1707	Commerce Court, Victoria Hotel	43.648198	-79.379817	Carisma	43.649617	-79.375434	Italian Restaurant
1708	Commerce Court, Victoria Hotel	43.648198	-79.379817	D.W. Alexander	43.648333	-79.373826	Cocktail Bar

In [94]:

```
toronto_venues.shape
```

Out[94]:

```
(1709, 7)
```

Display number of venues found per Neighborhood

In [95]:

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

Out[95]:

	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
Adelaide, King, Richmond	100	100	100	100	100	100
Berczy Park	56	56	56	56	56	56
Brockton, Exhibition Place, Parkdale Village	23	23	23	23	23	23
Business Reply Mail Processing Centre 969 Eastern	18	18	18	18	18	18
CN Tower, Bathurst Quay, Island airport, Harbourfront West, King and Spadina, Railway Lands, South Niagara	17	17	17	17	17	17
Cabbagetown, St. James Town	42	42	42	42	42	42
Central Bay Street	86	86	86	86	86	86
Chinatown, Grange Park, Kensington Market	100	100	100	100	100	100
Christie	16	16	16	16	16	16
Church and Wellesley	86	86	86	86	86	86
Commerce Court, Victoria Hotel	100	100	100	100	100	100
Davisville	37	37	37	37	37	37
Davisville North	9	9	9	9	9	9
Deer Park, Forest Hill SE, Rathnelly, South Hill, Summerhill West	15	15	15	15	15	15
Design Exchange, Toronto Dominion Centre	100	100	100	100	100	100
Dovercourt Village, Dufferin	15	15	15	15	15	15
First Canadian Place, Underground city	100	100	100	100	100	100
Forest Hill North, Forest Hill West	4	4	4	4	4	4
Harbord, University of Toronto	35	35	35	35	35	35
Harbourfront East, Toronto Islands, Union Station	100	100	100	100	100	100
Harbourfront, Regent Park	52	52	52	52	52	52
High Park, The Junction South	22	22	22	22	22	22
Lawrence Park	5	5	5	5	5	5

	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
Little Portugal, Trinity	66	66	66	66	66	66
Moore Park, Summerhill East	4	4	4	4	4	4
North Toronto West	20	20	20	20	20	20
Parkdale, Roncesvalles	15	15	15	15	15	15
Rosedale	5	5	5	5	5	5
Roselawn	1	1	1	1	1	1
Runnymede, Swansea	35	35	35	35	35	35
Ryerson, Garden District	100	100	100	100	100	100
St. James Town	100	100	100	100	100	100
Stn A PO Boxes 25 The Esplanade	98	98	98	98	98	98
Studio District	38	38	38	38	38	38
The Annex, North Midtown, Yorkville	23	23	23	23	23	23
The Beaches	4	4	4	4	4	4
The Beaches West, India Bazaar	19	19	19	19	19	19
The Danforth West, Riverdale	43	43	43	43	43	43

Display how many unique categories can be curated from all the returned venues

In [96]:

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

```
There are 235 uniques categories.
```

3. Analyze Each Neighborhood

In [97]:

```
# one hot encoding
toronto_onehot = pd.get_dummies(toronto_venues[['Venue Category']], prefix="", pref
ix_sep="")

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

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

toronto_onehot.head()
```

Out[97]:

	Yoga Studio	Afghan Restaurant	Airport	Airport Food Court	Airport Gate	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	Antique Shop
0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0

In [98]:

```
toronto_onehot.shape
```

Out[98]:

```
(1709, 235)
```

Next, group rows by neighborhood and by taking the mean of the frequency of occurrence of each category

In [99]:

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

Out[99]:

	Neighborhood	Yoga Studio	Afghan Restaurant	Airport	Airport Food Court	Airport Gate	Airport Lounge	Airport Service	Airp Termi
0	Adelaide, King, Richmond	0.000000	0.0	0.000000	0.000000	0.000000	0.000000	0.000000	0.0000
1	Berczy Park	0.000000	0.0	0.000000	0.000000	0.000000	0.000000	0.000000	0.0000
2	Brockton, Exhibition Place, Parkdale Village	0.000000	0.0	0.000000	0.000000	0.000000	0.000000	0.000000	0.0000
3	Business Reply Mail Processing Centre 969 Eastern	0.055556	0.0	0.000000	0.000000	0.000000	0.000000	0.000000	0.0000
4	CN Tower, Bathurst Quay, Island airport, Harbo...	0.000000	0.0	0.058824	0.058824	0.058824	0.117647	0.176471	0.1176

In [100]:

```
toronto_grouped.shape
```

Out[100]:

```
(38, 235)
```

Display each neighborhood along with the top 5 most common venues

In [101]:

```
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')
```

```
----Adelaide, King, Richmond----
         venue  freq
0  Coffee Shop  0.08
1         Café  0.05
2   Steakhouse  0.04
3          Bar  0.04
4        Hotel  0.03


----Berczy Park----
          venue  freq
0   Coffee Shop  0.09
1  Cocktail Bar  0.05
2          Café  0.04
3   Cheese Shop  0.04
4      Beer Bar  0.04


----Brockton, Exhibition Place, Parkdale Village----
            venue  freq
0     Coffee Shop  0.09
1  Breakfast Spot  0.09
2          Bakery  0.09
3            Café  0.09
4             Gym  0.04


----Business Reply Mail Processing Centre 969 Eastern----
                venue  freq
0  Light Rail Station  0.11
1         Yoga Studio  0.06
2       Auto Workshop  0.06
3       Garden Center  0.06
4              Garden  0.06


----CN Tower, Bathurst Quay, Island airport, Harbourfront West, King an
d Spadina, Railway Lands, South Niagara----
              venue  freq
0   Airport Service  0.18
1    Airport Lounge  0.12
2  Airport Terminal  0.12
3          Boutique  0.06
4     Boat or Ferry  0.06


----Cabbagetown, St. James Town----
                venue  freq
0         Coffee Shop  0.07
1         Pizza Place  0.05
2  Italian Restaurant  0.05
3                Café  0.05
4  Chinese Restaurant  0.05


----Central Bay Street----
                       venue  freq
```

```
0                Coffee Shop  0.15
1         Italian Restaurant  0.05
2                       Café  0.05
3             Ice Cream Shop  0.05
4  Middle Eastern Restaurant  0.03


----Chinatown, Grange Park, Kensington Market----
                           venue  freq
0                           Café  0.08
1  Vegetarian / Vegan Restaurant  0.06
2                            Bar  0.05
3             Mexican Restaurant  0.04
4          Vietnamese Restaurant  0.04


----Christie----
               venue  freq
0      Grocery Store  0.19
1               Café  0.19
2               Park  0.12
3  Convenience Store  0.06
4          Nightclub  0.06


----Church and Wellesley----
                 venue  freq
0          Coffee Shop  0.07
1  Japanese Restaurant  0.06
2     Sushi Restaurant  0.05
3              Gay Bar  0.03
4           Restaurant  0.03


----Commerce Court, Victoria Hotel----
         venue  freq
0  Coffee Shop  0.11
1        Hotel  0.06
2         Café  0.06
3   Restaurant  0.05
4   Steakhouse  0.03


----Davisville----
            venue  freq
0     Pizza Place  0.08
1  Sandwich Place  0.08
2    Dessert Shop  0.08
3        Pharmacy  0.05
4            Café  0.05


----Davisville North----
            venue  freq
0           Hotel  0.22
1    Dance Studio  0.11
2             Gym  0.11
```

```
3  Clothing Store  0.11
4            Park  0.11


----Deer Park, Forest Hill SE, Rathnelly, South Hill, Summerhill West--
--
                 venue  freq
0          Coffee Shop  0.13
1                  Pub  0.13
2          Pizza Place  0.07
3  Fried Chicken Joint  0.07
4           Bagel Shop  0.07


----Design Exchange, Toronto Dominion Centre----
           venue  freq
0    Coffee Shop  0.12
1           Café  0.08
2          Hotel  0.06
3     Restaurant  0.05
4  Deli / Bodega  0.03


----Dovercourt Village, Dufferin----
                    venue  freq
0                  Bakery  0.13
1                Pharmacy  0.13
2             Supermarket  0.13
3                 Brewery  0.07
4  Furniture / Home Store  0.07


----First Canadian Place, Underground city----
         venue  freq
0  Coffee Shop  0.10
1         Café  0.08
2        Hotel  0.05
3   Restaurant  0.04
4    Gastropub  0.03


----Forest Hill North, Forest Hill West----
              venue  freq
0              Park  0.25
1             Trail  0.25
2  Sushi Restaurant  0.25
3     Jewelry Store  0.25
4       Yoga Studio  0.00


----Harbord, University of Toronto----
                 venue  freq
0                 Café  0.14
1            Bookstore  0.06
2                  Bar  0.06
3  Japanese Restaurant  0.06
4           Restaurant  0.06
```

```
----Harbourfront East, Toronto Islands, Union Station----
            venue  freq
0     Coffee Shop  0.13
1        Aquarium  0.05
2           Hotel  0.05
3            Café  0.04
4  Scenic Lookout  0.03


----Harbourfront, Regent Park----
         venue  freq
0  Coffee Shop  0.15
1       Bakery  0.06
2         Park  0.06
3         Café  0.06
4          Pub  0.06


----High Park, The Junction South----
                 venue  freq
0   Mexican Restaurant  0.09
1                 Café  0.09
2            Bookstore  0.05
3  Arts & Crafts Store  0.05
4          Music Venue  0.05


----Lawrence Park----
                venue  freq
0  Photography Studio   0.2
1                Park   0.2
2              Lawyer   0.2
3         Swim School   0.2
4            Bus Line   0.2


----Little Portugal, Trinity----
                   venue  freq
0                    Bar  0.12
1            Coffee Shop  0.06
2       Asian Restaurant  0.05
3           Cocktail Bar  0.03
4  Vietnamese Restaurant  0.03


----Moore Park, Summerhill East----
                        venue  freq
0                  Playground  0.25
1                Tennis Court  0.25
2                         Gym  0.25
3                       Trail  0.25
4  Modern European Restaurant  0.00


----North Toronto West----
```

```
                 venue  freq
0       Clothing Store  0.15
1          Coffee Shop  0.10
2  Sporting Goods Shop  0.10
3          Yoga Studio  0.05
4                  Spa  0.05


----Parkdale, Roncesvalles----
            venue  freq
0  Breakfast Spot  0.13
1       Gift Shop  0.13
2     Coffee Shop  0.07
3   Movie Theater  0.07
4             Bar  0.07


----Rosedale----
        venue  freq
0        Park   0.4
1  Playground   0.2
2       Trail   0.2
3    Building   0.2
4   Nightclub   0.0


----Roselawn----
                      venue  freq
0                    Garden   1.0
1               Yoga Studio   0.0
2   New American Restaurant   0.0
3  Mediterranean Restaurant   0.0
4               Men's Store   0.0


----Runnymede, Swansea----
                venue  freq
0         Coffee Shop  0.09
1                Café  0.09
2    Sushi Restaurant  0.06
3         Pizza Place  0.06
4  Italian Restaurant  0.06


----Ryerson, Garden District----
                       venue  freq
0                Coffee Shop  0.09
1             Clothing Store  0.07
2                       Café  0.03
3             Cosmetics Shop  0.03
4  Middle Eastern Restaurant  0.03


----St. James Town----
                venue  freq
0         Coffee Shop  0.07
1          Restaurant  0.06
```

```
2              Café  0.06
3             Hotel  0.05
4  Italian Restaurant  0.04


----Stn A PO Boxes 25 The Esplanade----
                venue  freq
0         Coffee Shop  0.11
1          Restaurant  0.04
2  Seafood Restaurant  0.04
3                Café  0.04
4               Hotel  0.03


----Studio District----
                 venue  freq
0                 Café  0.11
1          Coffee Shop  0.08
2               Bakery  0.05
3   Italian Restaurant  0.05
4  American Restaurant  0.05


----The Annex, North Midtown, Yorkville----
            venue  freq
0     Coffee Shop  0.13
1  Sandwich Place  0.13
2            Café  0.13
3     Pizza Place  0.09
4  History Museum  0.04


----The Beaches----
                     venue  freq
0        Health Food Store  0.25
1                    Trail  0.25
2                      Pub  0.25
3              Yoga Studio  0.00
4  New American Restaurant  0.00


----The Beaches West, India Bazaar----
               venue  freq
0               Park  0.11
1  Fish & Chips Shop  0.05
2         Steakhouse  0.05
3   Sushi Restaurant  0.05
4       Liquor Store  0.05


----The Danforth West, Riverdale----
                    venue  freq
0        Greek Restaurant  0.21
1             Coffee Shop  0.09
2      Italian Restaurant  0.07
3  Furniture / Home Store  0.05
4          Ice Cream Shop  0.05
```

Put that into a *pandas* dataframe

Sort the venues in descending order.

In [102]:

```
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]
```

Create the new dataframe and display the top 10 venues for each neighborhood.

In [103]:

```
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_g
rouped.iloc[ind, :], num_top_venues)

neighborhoods_venues_sorted.head()
```

Out[103]:

	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 Most Common Venue
0	Adelaide, King, Richmond	Coffee Shop	Café	Bar	Steakhouse	Hotel	Burger Joint	Cosmetics Shop
1	Berczy Park	Coffee Shop	Cocktail Bar	Bakery	Steakhouse	Cheese Shop	Farmers Market	Café
2	Brockton, Exhibition Place, Parkdale Village	Coffee Shop	Café	Breakfast Spot	Bakery	Gym	Intersection	Performing Arts Venue
3	Business Reply Mail Processing Centre 969 Eastern	Light Rail Station	Yoga Studio	Auto Workshop	Comic Shop	Pizza Place	Recording Studio	Restaurant
4	CN Tower, Bathurst Quay, Island airport, Harbo...	Airport Service	Airport Terminal	Airport Lounge	Plane	Bar	Boutique	Sculpture Garden

4. Cluster Neighborhoods

Run *k*-means to cluster the neighborhood into 5 clusters.

In [104]:

```
# set number of clusters
kclusters = 5

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

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

# check cluster labels generated for each row in the dataframe
kmeans.labels_[0:30]
```

Out[104]:

```
array([0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 3, 0, 0, 0,
0,
       4, 0, 2, 0, 0, 3, 1, 0], dtype=int32)
```

Create a new dataframe that includes the cluster as well as the top 10 venues for each neighborhood.

In [105]:

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

#df.rename(columns={'two':'new_name'}, inplace=True)
TorontoDF.rename(columns={'Neighbourhood':'Neighborhood'}, inplace=True)
toronto_merged = TorontoDF

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

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

```
/home/jupyterlab/conda/envs/python/lib/python3.6/site-packages/pandas/c
ore/frame.py:4223: SettingWithCopyWarning:
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-d
ocs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  return super().rename(**kwargs)
```

Out[105]:

	Postcode	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3r Co
6	M5C	Downtown Toronto	St. James Town	43.651494	-79.375418	0	Coffee Shop	Restaurant	
8	M4E	East Toronto	The Beaches	43.676357	-79.293031	0	Health Food Store	Trail	
9	M5E	Downtown Toronto	Berczy Park	43.644771	-79.373306	0	Coffee Shop	Cocktail Bar	
13	M5G	Downtown Toronto	Central Bay Street	43.657952	-79.387383	0	Coffee Shop	Ice Cream Shop	
14	M6G	Downtown Toronto	Christie	43.669542	-79.422564	0	Grocery Store	Café	

Visualize the resulting clusters with Folium

In [108]:

```
# 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[108]:

5. Examine Clusters

Cluster 0

In [109]:

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

Out[109]:

	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
6	Downtown Toronto	0	Coffee Shop	Restaurant	Café	Hotel	Italian Restaurant	Clothing Store
8	East Toronto	0	Health Food Store	Trail	Pub	Dessert Shop	Farmers Market	Falafel Restaurant
9	Downtown Toronto	0	Coffee Shop	Cocktail Bar	Bakery	Steakhouse	Cheese Shop	Farmers Market
13	Downtown Toronto	0	Coffee Shop	Ice Cream Shop	Café	Italian Restaurant	Middle Eastern Restaurant	Burger Joint
14	Downtown Toronto	0	Grocery Store	Café	Park	Athletics & Sports	Nightclub	Baby Store
24	East Toronto	0	Café	Coffee Shop	Bakery	Italian Restaurant	American Restaurant	Yoga Studio
31	Central Toronto	0	Hotel	Gym	Sandwich Place	Park	Breakfast Spot	Clothing Store
34	Central Toronto	0	Clothing Store	Coffee Shop	Sporting Goods Shop	Mexican Restaurant	Diner	Dessert Shop
37	Central Toronto	0	Dessert Shop	Sandwich Place	Pizza Place	Café	Pharmacy	Italian Restaurant
40	Downtown Toronto	0	Coffee Shop	Café	Seafood Restaurant	Restaurant	Hotel	Beer Bar
43	Downtown Toronto	0	Coffee Shop	Japanese Restaurant	Sushi Restaurant	Gay Bar	Restaurant	Men's Store
44	East Toronto	0	Light Rail Station	Yoga Studio	Auto Workshop	Comic Shop	Pizza Place	Recording Studio
47	Downtown Toronto	0	Airport Service	Airport Terminal	Airport Lounge	Plane	Bar	Boutique
50	Central Toronto	0	Coffee Shop	Pub	Light Rail Station	Restaurant	Vietnamese Restaurant	Supermarket
55	Central Toronto	0	Coffee Shop	Café	Sandwich Place	Pizza Place	Pharmacy	Indian Restaurant
57	West Toronto	0	Coffee Shop	Café	Breakfast Spot	Bakery	Gym	Intersection
61	Downtown Toronto	0	Coffee Shop	Hotel	Aquarium	Café	Brewery	Scenic Lookout
62	Downtown Toronto	0	Café	Vegetarian / Vegan Restaurant	Bar	Mexican Restaurant	Dumpling Restaurant	Chinese Restaurant
71	Downtown Toronto	0	Coffee Shop	Café	Bar	Steakhouse	Hotel	Burger Joint

	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
75	Downtown Toronto	0	Coffee Shop	Restaurant	Bakery	Pizza Place	Chinese Restaurant	Italian Restaurant
77	Downtown Toronto	0	Coffee Shop	Café	Hotel	Restaurant	Bar	American Restaurant
78	Downtown Toronto	0	Coffee Shop	Pub	Bakery	Park	Café	Theater
79	East Toronto	0	Park	Gym	Italian Restaurant	Sushi Restaurant	Pub	Burrito Place
81	Downtown Toronto	0	Coffee Shop	Café	Hotel	Restaurant	American Restaurant	Seafood Restaurant
84	West Toronto	0	Breakfast Spot	Gift Shop	Italian Restaurant	Eastern European Restaurant	Restaurant	Movie Theater
86	Downtown Toronto	0	Café	Bar	Japanese Restaurant	Bookstore	Restaurant	Bakery
88	Downtown Toronto	0	Coffee Shop	Clothing Store	Cosmetics Shop	Café	Middle Eastern Restaurant	Restaurant
89	West Toronto	0	Bar	Coffee Shop	Asian Restaurant	Café	Bakery	New American Restaurant
92	East Toronto	0	Greek Restaurant	Coffee Shop	Italian Restaurant	Furniture / Home Store	Ice Cream Shop	Caribbean Restaurant
94	West Toronto	0	Mexican Restaurant	Café	Park	Arts & Crafts Store	Fast Food Restaurant	Bookstore
96	West Toronto	0	Coffee Shop	Café	Sushi Restaurant	Pizza Place	Italian Restaurant	Restaurant
97	West Toronto	0	Supermarket	Pharmacy	Bakery	Café	Furniture / Home Store	Bank
101	Downtown Toronto	0	Coffee Shop	Café	Hotel	Restaurant	Italian Restaurant	Steakhouse

Cluster 1

In [110]:

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

Out[110]:

	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	7th Most Common Venue
28	Central Toronto	1	Garden	Wings Joint	Discount Store	Fast Food Restaurant	Farmers Market	Falafel Restaurant	Event Space

Cluster 2

In [111]:

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

Out[111]:

	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	7th Most Common Venue
83	Central Toronto	2	Gym	Playground	Trail	Tennis Court	Wings Joint	Dumpling Restaurant	Discount Store

Cluster 3

In [112]:

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

Out[112]:

	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	7th Mo Comm Venu
39	Downtown Toronto	3	Park	Playground	Trail	Building	Wings Joint	Diner	Fala Restaura
76	Central Toronto	3	Park	Trail	Jewelry Store	Sushi Restaurant	Wings Joint	Discount Store	Farme Mark

Cluster 4

In [114]:

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

Out[114]:

	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	7th Mos Commo Venu
27	Central Toronto	4	Photography Studio	Lawyer	Park	Bus Line	Swim School	Diner	Falafe Restaurar