Week 1. Peer-graded Assignment - Capstone Project Notebook

Instructions

Start a Jupyter Notebook using any platform that you are comfortable with and do the following:

1. Write some markdown to explain that this notebook will be mainly used for the capstone project.

This Jupyter Notebook is going to be used mainly for the developement of the Capstone Project.

2. Import the pandas library as pd.
3. Import the Numpy library as np.

```
import pandas as pd
import numpy as np
```

4. Print the following the statement: Hello Capstone Project Course!

```
print('Hello Capstone Project Course!')
```

```
Hello Capstone Project Course!
```

Push the Notebook to your Github repository and submit a link to the notebook on your Github repository.

Week 3. Peer-graded Assignment: Segmenting and Clustering Neighb

Instructions

For this assignment, you will be required to explore and cluster the neighborhoods in Toronto.

1. Start by creating a new Notebook for this assignment.
2. Use the Notebook to build the code to scrape the following Wikipedia page, https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M, in order to obtain the data that is in the table transform the data into a pandas dataframe like the one shown below:

```
# We import the required packages for the Peer-graded Assigment
import pandas as pd
import numpy as np
```

```
# We set the display options of pandas, so we can display the whole table when we are done.
pd.options.display.max_rows = 999
```

```
# We use the pandas' read_html function to read the wikitable
df1 = pd.read_html('http://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M', match='Postal c
```

```
# We check whether the reading of the table was successfull or not.
# df1.head()
```

3. To create the above dataframe:

- The dataframe will consist of three columns: PostalCode, Borough, and Neighborhood
- Only process the cells that have an assigned borough. Ignore cells with a borough that is Not assigned.
- More than one neighborhood can exist in one postal code area. For example, in the table on the Wikipedia is listed twice and has two neighborhoods: Harbourfront and Regent Park. These two rows will be combine neighborhoods separated with a comma as shown in row 11 in the above table.
- If a cell has a borough but a Not assigned neighborhood, then the neighborhood will be the same as the bo
- Clean your Notebook and add Markdown cells to explain your work and any assumptions you are making.
- In the last cell of your notebook, use the .shape method to print the number of rows of your dataframe.

```
# We get the index of those registers whose Boroughs are not assigned
indexNames = df1[df1.Borough=='Not assigned'].index
```

```
# We drop the rows in which the Borough is not assigned
df1.drop(indexNames,inplace=True)
```

```
# We reset the index of the DataFrame
df1.reset_index(drop=True, inplace=True)
```

```
# We replace the '/' with ',' for every Neighborhood
df1['Neighborhood'] = df1['Neighborhood'].apply(lambda x: x.replace(' /',','))
```

```
# We now check there are no Neighborhoods with 'Not assigned' values
if sum(df1['Neighborhood']=='Not assigned')==0:
    print('All Neighborhood values are Ok.')
```

```
All Neighborhood values are Ok.
```

```
# We finally display the complete table of the List_of_postal_codes_of_Canada:_M
df1
```

	Postal code	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
5	M9A	Etobicoke	Islington Avenue
6	M1B	Scarborough	Malvern, Rouge
7	M3B	North York	Don Mills
8	M4B	East York	Parkview Hill, Woodbine Gardens
9	M5B	Downtown Toronto	Garden District, Ryerson
10	M6B	North York	Glencairn
11	M9B	Etobicoke	West Deane Park, Princess Gardens, Martin Grov...
12	M1C	Scarborough	Rouge Hill, Port Union, Highland Creek
13	M3C	North York	Don Mills
14	M4C	East York	Woodbine Heights
15	M5C	Downtown Toronto	St. James Town
16	M6C	York	Humewood-Cedarvale
17	M9C	Etobicoke	Eringate, Bloordale Gardens, Old Burnhamthorpe...
18	M1E	Scarborough	Guildwood, Morningside, West Hill
19	M4E	East Toronto	The Beaches
20	M5E	Downtown Toronto	Berczy Park
21	M6E	York	Caledonia-Fairbanks
22	M1G	Scarborough	Woburn
23	M4G	East York	Leaside
24	M5G	Downtown Toronto	Central Bay Street
25	M6G	Downtown Toronto	Christie
26	M1H	Scarborough	Cedarbrae
27	M2H	North York	Hillcrest Village
28	M3H	North York	Bathurst Manor, Wilson Heights, Downsview North
29	M4H	East York	Thorncliffe Park
30	M5H	Downtown Toronto	Richmond, Adelaide, King
31	M6H	West Toronto	Dufferin, Dovercourt Village

32	M1J	Scarborough	Scarborough Village
33	M2J	North York	Fairview, Henry Farm, Oriole
34	M3J	North York	Northwood Park, York University
35	M4J	East York	East Toronto
36	M5J	Downtown Toronto	Harbourfront East, Union Station, Toronto Islands
37	M6J	West Toronto	Little Portugal, Trinity
38	M1K	Scarborough	Kennedy Park, Ionview, East Birchmount Park
39	M2K	North York	Bayview Village
40	M3K	North York	Downsview
41	M4K	East Toronto	The Danforth West, Riverdale
42	M5K	Downtown Toronto	Toronto Dominion Centre, Design Exchange
43	M6K	West Toronto	Brockton, Parkdale Village, Exhibition Place
44	M1L	Scarborough	Golden Mile, Clairlea, Oakridge
45	M2L	North York	York Mills, Silver Hills
46	M3L	North York	Downsview
47	M4L	East Toronto	India Bazaar, The Beaches West
48	M5L	Downtown Toronto	Commerce Court, Victoria Hotel
49	M6L	North York	North Park, Maple Leaf Park, Upwood Park
50	M9L	North York	Humber Summit
51	M1M	Scarborough	Cliffside, Cliffcrest, Scarborough Village West
52	M2M	North York	Willowdale, Newtonbrook
53	M3M	North York	Downsview
54	M4M	East Toronto	Studio District
55	M5M	North York	Bedford Park, Lawrence Manor East
56	M6M	York	Del Ray, Mount Dennis, Keelsdale and Silverthorn
57	M9M	North York	Humberlea, Emery
58	M1N	Scarborough	Birch Cliff, Cliffside West
59	M2N	North York	Willowdale
60	M3N	North York	Downsview
61	M4N	Central Toronto	Lawrence Park
62	M5N	Central Toronto	Roselawn
63	M6N	York	Runnymede, The Junction North

64	M9N	York	Weston
65	M1P	Scarborough	Dorset Park, Wexford Heights, Scarborough Town...
66	M2P	North York	York Mills West
67	M4P	Central Toronto	Davisville North
68	M5P	Central Toronto	Forest Hill North & West
69	M6P	West Toronto	High Park, The Junction South
70	M9P	Etobicoke	Westmount
71	M1R	Scarborough	Wexford, Maryvale
72	M2R	North York	Willowdale
73	M4R	Central Toronto	North Toronto West
74	M5R	Central Toronto	The Annex, North Midtown, Yorkville
75	M6R	West Toronto	Parkdale, Roncesvalles
76	M7R	Mississauga	Canada Post Gateway Processing Centre
77	M9R	Etobicoke	Kingsview Village, St. Phillips, Martin Grove ...
78	M1S	Scarborough	Agincourt
79	M4S	Central Toronto	Davisville
80	M5S	Downtown Toronto	University of Toronto, Harbord
81	M6S	West Toronto	Runnymede, Swansea
82	M1T	Scarborough	Clarks Corners, Tam O'Shanter, Sullivan
83	M4T	Central Toronto	Moore Park, Summerhill East
84	M5T	Downtown Toronto	Kensington Market, Chinatown, Grange Park
85	M1V	Scarborough	Milliken, Agincourt North, Steeles East, L'Amo...
86	M4V	Central Toronto	Summerhill West, Rathnelly, South Hill, Forest...
87	M5V	Downtown Toronto	CN Tower, King and Spadina, Railway Lands, Har...
88	M8V	Etobicoke	New Toronto, Mimico South, Humber Bay Shores
89	M9V	Etobicoke	South Steeles, Silverstone, Humbergate, Jamest...
90	M1W	Scarborough	Steeles West, L'Amoreaux West
91	M4W	Downtown Toronto	Rosedale
92	M5W	Downtown Toronto	Stn A PO Boxes
93	M8W	Etobicoke	Alderwood, Long Branch
94	M9W	Etobicoke	Northwest
95	M1X	Scarborough	Upper Rouge
96	M4X	Downtown Toronto	St. James Town, Cabbagetown

97	M5X	Downtown Toronto	First Canadian Place, Underground city
98	M8X	Etobicoke	The Kingsway, Montgomery Road, Old Mill North
99	M4Y	Downtown Toronto	Church and Wellesley
100	M7Y	East Toronto	Business reply mail Processing CentrE
101	M8Y	Etobicoke	Old Mill South, King's Mill Park, Sunnylea, Hu...
102	M8Z	Etobicoke	Mimico NW, The Queensway West, South of Bloor,...

4. Submit a link to your Notebook on your Github repository. (10 marks)
Note: There are different website scraping libraries and packages in Python. For scraping the above table, you ca
the table into a pandas dataframe.

Here is a link to a csv file that has the geographical coordinates of each postal code:
http://cocl.us/Geospatial_data/Geospatial_Coordinates.csv . Use the csv file to create the following dataframe:

```
df2 = pd.read_csv(r'http://cocl.us/Geospatial_data/Geospatial_Coordinates.csv')
# df2.head()
```

```
# We now merge both DataFrames
df = df1.merge(df2[['Latitude','Longitude']], how = 'left', left_on=df1['Postal code'],right_on=df2[
df
```

	key_0	Postal code	Borough	Neighborhood	Latitude
0	M3A	M3A	North York	Parkwoods	43.753259
1	M4A	M4A	North York	Victoria Village	43.725882
2	M5A	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260
3	M6A	M6A	North York	Lawrence Manor, Lawrence Heights	43.718518
4	M7A	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government	43.662301
5	M9A	M9A	Etobicoke	Islington Avenue	43.667856
6	M1B	M1B	Scarborough	Malvern, Rouge	43.806686
7	M3B	M3B	North York	Don Mills	43.745906
8	M4B	M4B	East York	Parkview Hill, Woodbine Gardens	43.706397
9	M5B	M5B	Downtown Toronto	Garden District, Ryerson	43.657162
10	M6B	M6B	North York	Glencairn	43.709577
11	M9B	M9B	Etobicoke	West Deane Park, Princess Gardens, Martin Grov...	43.650943
12	M1C	M1C	Scarborough	Rouge Hill, Port Union, Highland Creek	43.784535
13	M3C	M3C	North York	Don Mills	43.725900
14	M4C	M4C	East York	Woodbine Heights	43.695344
15	M5C	M5C	Downtown Toronto	St. James Town	43.651494
16	M6C	M6C	York	Humewood-Cedarvale	43.693781
17	M9C	M9C	Etobicoke	Eringate, Bloordale Gardens, Old Burnhamthorpe...	43.643515
18	M1E	M1E	Scarborough	Guildwood, Morningside, West Hill	43.763573
19	M4E	M4E	East Toronto	The Beaches	43.676357
20	M5E	M5E	Downtown Toronto	Berczy Park	43.644771
21	M6E	M6E	York	Caledonia-Fairbanks	43.689026
22	M1G	M1G	Scarborough	Woburn	43.770992
23	M4G	M4G	East York	Leaside	43.709060
24	M5G	M5G	Downtown Toronto	Central Bay Street	43.657952
25	M6G	M6G	Downtown Toronto	Christie	43.669542
26	M1H	M1H	Scarborough	Cedarbrae	43.773136
27	M2H	M2H	North York	Hillcrest Village	43.803762
28	M3H	M3H	North York	Bathurst Manor, Wilson Heights, Downsview North	43.754328
29	M4H	M4H	East York	Thorncliffe Park	43.705369
30	M5H	M5H	Downtown Toronto	Richmond, Adelaide, King	43.650571
31	M6H	M6H	West Toronto	Dufferin, Dovercourt Village	43.669005

32	M1J	M1J	Scarborough	Scarborough Village	43.744734	-
33	M2J	M2J	North York	Fairview, Henry Farm, Oriole	43.778517	-
34	M3J	M3J	North York	Northwood Park, York University	43.767980	-
35	M4J	M4J	East York	East Toronto	43.685347	-
36	M5J	M5J	Downtown Toronto	Harbourfront East, Union Station, Toronto Islands	43.640816	-
37	M6J	M6J	West Toronto	Little Portugal, Trinity	43.647927	-
38	M1K	M1K	Scarborough	Kennedy Park, Ionview, East Birchmount Park	43.727929	-
39	M2K	M2K	North York	Bayview Village	43.786947	-
40	M3K	M3K	North York	Downsview	43.737473	-
41	M4K	M4K	East Toronto	The Danforth West, Riverdale	43.679557	-
42	M5K	M5K	Downtown Toronto	Toronto Dominion Centre, Design Exchange	43.647177	-
43	M6K	M6K	West Toronto	Brockton, Parkdale Village, Exhibition Place	43.636847	-
44	M1L	M1L	Scarborough	Golden Mile, Clairlea, Oakridge	43.711112	-
45	M2L	M2L	North York	York Mills, Silver Hills	43.757490	-
46	M3L	M3L	North York	Downsview	43.739015	-
47	M4L	M4L	East Toronto	India Bazaar, The Beaches West	43.668999	-
48	M5L	M5L	Downtown Toronto	Commerce Court, Victoria Hotel	43.648198	-
49	M6L	M6L	North York	North Park, Maple Leaf Park, Upwood Park	43.713756	-
50	M9L	M9L	North York	Humber Summit	43.756303	-
51	M1M	M1M	Scarborough	Cliffside, Cliffcrest, Scarborough Village West	43.716316	-
52	M2M	M2M	North York	Willowdale, Newtonbrook	43.789053	-
53	M3M	M3M	North York	Downsview	43.728496	-
54	M4M	M4M	East Toronto	Studio District	43.659526	-
55	M5M	M5M	North York	Bedford Park, Lawrence Manor East	43.733283	-
56	M6M	M6M	York	Del Ray, Mount Dennis, Keelsdale and Silverthorn	43.691116	-
57	M9M	M9M	North York	Humberlea, Emery	43.724766	-
58	M1N	M1N	Scarborough	Birch Cliff, Cliffside West	43.692657	-
59	M2N	M2N	North York	Willowdale	43.770120	-
60	M3N	M3N	North York	Downsview	43.761631	-
61	M4N	M4N	Central Toronto	Lawrence Park	43.728020	-
62	M5N	M5N	Central Toronto	Roselawn	43.711695	-
63	M6N	M6N	York	Runnymede, The Junction North	43.673185	-

64	M9N	M9N	York	Weston	43.706876	-
65	M1P	M1P	Scarborough	Dorset Park, Wexford Heights, Scarborough Town...	43.757410	-
66	M2P	M2P	North York	York Mills West	43.752758	-
67	M4P	M4P	Central Toronto	Davisville North	43.712751	-
68	M5P	M5P	Central Toronto	Forest Hill North & West	43.696948	-
69	M6P	M6P	West Toronto	High Park, The Junction South	43.661608	-
70	M9P	M9P	Etobicoke	Westmount	43.696319	-
71	M1R	M1R	Scarborough	Wexford, Maryvale	43.750072	-
72	M2R	M2R	North York	Willowdale	43.782736	-
73	M4R	M4R	Central Toronto	North Toronto West	43.715383	-
74	M5R	M5R	Central Toronto	The Annex, North Midtown, Yorkville	43.672710	-
75	M6R	M6R	West Toronto	Parkdale, Roncesvalles	43.648960	-
76	M7R	M7R	Mississauga	Canada Post Gateway Processing Centre	43.636966	-
77	M9R	M9R	Etobicoke	Kingsview Village, St. Phillips, Martin Grove ...	43.688905	-
78	M1S	M1S	Scarborough	Agincourt	43.794200	-
79	M4S	M4S	Central Toronto	Davisville	43.704324	-
80	M5S	M5S	Downtown Toronto	University of Toronto, Harbord	43.662696	-
81	M6S	M6S	West Toronto	Runnymede, Swansea	43.651571	-
82	M1T	M1T	Scarborough	Clarks Corners, Tam O'Shanter, Sullivan	43.781638	-
83	M4T	M4T	Central Toronto	Moore Park, Summerhill East	43.689574	-
84	M5T	M5T	Downtown Toronto	Kensington Market, Chinatown, Grange Park	43.653206	-
85	M1V	M1V	Scarborough	Milliken, Agincourt North, Steeles East, L'Amo...	43.815252	-
86	M4V	M4V	Central Toronto	Summerhill West, Rathnelly, South Hill, Forest...	43.686412	-
87	M5V	M5V	Downtown Toronto	CN Tower, King and Spadina, Railway Lands, Har...	43.628947	-
88	M8V	M8V	Etobicoke	New Toronto, Mimico South, Humber Bay Shores	43.605647	-
89	M9V	M9V	Etobicoke	South Steeles, Silverstone, Humbergate, Jamest...	43.739416	-
90	M1W	M1W	Scarborough	Steeles West, L'Amoreaux West	43.799525	-
91	M4W	M4W	Downtown Toronto	Rosedale	43.679563	-
92	M5W	M5W	Downtown Toronto	Stn A PO Boxes	43.646435	-
93	M8W	M8W	Etobicoke	Alderwood, Long Branch	43.602414	-
94	M9W	M9W	Etobicoke	Northwest	43.706748	-
95	M1X	M1X	Scarborough	Upper Rouge	43.836125	-
96	M4X	M4X	Downtown Toronto	St. James Town, Cabbagetown	43.667967	-

97	M5X	M5X	Downtown Toronto	First Canadian Place, Underground city	43.648429	-
98	M8X	M8X	Etobicoke	The Kingsway, Montgomery Road, Old Mill North	43.653654	-
99	M4Y	M4Y	Downtown Toronto	Church and Wellesley	43.665860	-
100	M7Y	M7Y	East Toronto	Business reply mail Processing CentrE	43.662744	-
101	M8Y	M8Y	Etobicoke	Old Mill South, King's Mill Park, Sunnylea, Hu...	43.636258	-
102	M8Z	M8Z	Etobicoke	Mimico NW, The Queensway West, South of Bloor,...	43.628841	-

5. Explore and cluster the neighborhoods in Toronto. You can decide to work with only boroughs that contain the replicate the same analysis we did to the New York City data. It is up to you.

Just make sure:

1. to add enough Markdown cells to explain what you decided to do and to report any observations you make
2. to generate maps to visualize your neighborhoods and how they cluster together.

Once you are happy with your analysis, submit a link to the new Notebook on your Github repository. (3 marks)

1. Download and Explore Dataset

```
# To simplify the analysis, we have decided to work only with Boroughs that contain the word Toronto
downtown_data = df[df['Borough'] == 'Downtown Toronto'].reset_index(drop=True)
downtown_data.head()
```

	key_0	Postal code	Borough	Neighborhood	Latitude	Longitud
0	M5A	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.36063
1	M7A	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government	43.662301	-79.38949
2	M5B	M5B	Downtown Toronto	Garden District, Ryerson	43.657162	-79.37893
3	M5C	M5C	Downtown Toronto	St. James Town	43.651494	-79.37541
4	M5E	M5E	Downtown Toronto	Berczy Park	43.644771	-79.37330

```
from geopy.geocoders import Nominatim # convert an address into latitude and longitude values

# Let's get the geographic coordinates of Toronto's Downtown

address = 'Downtown Toronto'

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

```
The geograpical coordinate of Manhattan are 43.6541737, -79.38081164513409.
```

Let's visualize Toronto and the neighborhoods in it.

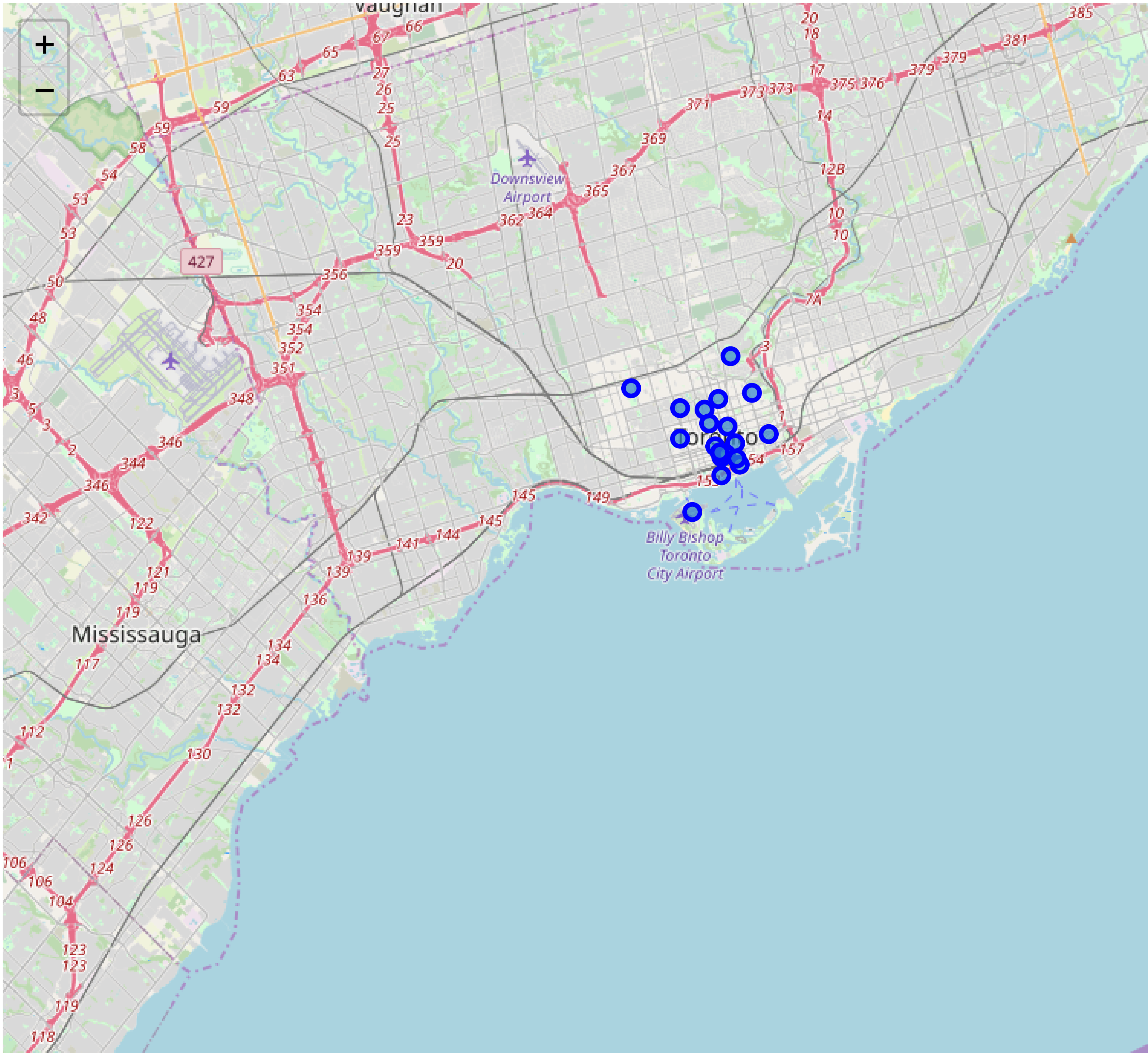
```
# Create map of Downtown Toronto using latitude and longitude values

import folium # map rendering library
# Matplotlib and associated plotting modules
import matplotlib.cm as cm
import matplotlib.colors as colors

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

# add markers to map
for lat, lng, label in zip(downtown_data['Latitude'], downtown_data['Longitude'], downtown_data['Nei
    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_downtown)

map_downtown
```



Next, we are going to start utilizing the Foursquare API to explore the neighborhoods and segment them.

Define Foursquare Credentials and Version

```
CLIENT_ID = 'xxx' # your Foursquare ID
CLIENT_SECRET = 'xxx' # your Foursquare Secret
VERSION = '20180605' # Foursquare API version

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

```
print('CLIENT_SECRET:' + CLIENT_SECRET)
```

```
Your credentails:
CLIENT_ID: xxx
CLIENT_SECRET:xxx
```

2. Explore Neighborhoods in Manhattan

```
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
```

```
# Let's create a function to get the top 100 venues that are in a radius of 500
# meters for all the neighborhoods in Downtown Toronto

LIMIT = 100 # limit of number of venues returned by Foursquare API

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={
            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)
```

```
downtown_venues = getNearbyVenues(names=downtown_data['Neighborhood'],
                                  latitudes=downtown_data['Latitude'],
                                  longitudes=downtown_data['Longitude']
                                 )
```

```
Regent Park, Harbourfront
Queen's Park, Ontario Provincial Government
Garden District, Ryerson
St. James Town
Berczy Park
Central Bay Street
Christie
Richmond, Adelaide, King
Harbourfront East, Union Station, Toronto Islands
Toronto Dominion Centre, Design Exchange
Commerce Court, Victoria Hotel
University of Toronto, Harbord
Kensington Market, Chinatown, Grange Park
CN Tower, King and Spadina, Railway Lands, Harbourfront West, Bathurst  Quay, South Niagara, Islan
Rosedale
Stn A PO Boxes
St. James Town, Cabbagetown
First Canadian Place, Underground city
Church and Wellesley
```

```
# Let's check the size of the resulting dataframe
print(downtown_venues.shape)
downtown_venues.head()
```

(1259, 7)

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	V
0	Regent Park, Harbourfront	43.65426	-79.360636	Roselle Desserts	
1	Regent Park, Harbourfront	43.65426	-79.360636	Tandem Coffee	
2	Regent Park, Harbourfront	43.65426	-79.360636	Cooper Koo Family YMCA	
3	Regent Park, Harbourfront	43.65426	-79.360636	Body Blitz Spa East	
4	Regent Park, Harbourfront	43.65426	-79.360636	Morning Glory Cafe	

Let's check how many venues were returned for each neighborhood

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

Neighborhood	Neighborhood Latitude
Berczy Park	56
CN Tower, King and Spadina, Railway Lands, Harbourfront West, Bathurst Quay, South Niagara, Island airport	16
Central Bay Street	73
Christie	18
Church and Wellesley	77
Commerce Court, Victoria Hotel	100
First Canadian Place, Underground city	100
Garden District, Ryerson	100
Harbourfront East, Union Station, Toronto Islands	100
Kensington Market, Chinatown, Grange Park	66
Queen's Park, Ontario Provincial Government	34
Regent Park, Harbourfront	45
Richmond, Adelaide, King	100
Rosedale	4
St. James Town	91
St. James Town, Cabbagetown	46
Stn A PO Boxes	97
Toronto Dominion Centre, Design Exchange	100
University of Toronto, Harbord	36

Let's find out how many unique categories can be curated from all the returned venues

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

```
There are 207 uniques categories.
```

3. Analyze Each Neighborhood

```
# one hot encoding
downtown_onehot = pd.get_dummies(downtown_venues[['Venue Category']], prefix="", prefix_sep="")

# add neighborhood column back to dataframe
```

```
downtown_onehot['Neighborhood'] = downtown_venues['Neighborhood']

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

downtown_onehot.head()
```

	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

5 rows × 207 columns

And let's examine the new dataframe size.

```
downtown_onehot.shape
```

```
(1259, 207)
```

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

```
downtown_grouped = downtown_onehot.groupby('Neighborhood').mean().reset_index()
downtown_grouped
```

	Neighborhood	Yoga Studio	Afghan Restaurant	Airport	Airport Food Court	Airport Gate	Airport Lounge	Airport Service	Airport Terminal	Am Rest
0	Berczy Park	0.000000	0.000000	0.0000	0.0000	0.0000	0.000	0.000	0.0000	0.
1	CN Tower, King and Spadina, Railway Lands, Har...	0.000000	0.000000	0.0625	0.0625	0.0625	0.125	0.125	0.0625	0.
2	Central Bay Street	0.013699	0.000000	0.0000	0.0000	0.0000	0.000	0.000	0.0000	0.
3	Christie	0.000000	0.000000	0.0000	0.0000	0.0000	0.000	0.000	0.0000	0.
4	Church and Wellesley	0.025974	0.012987	0.0000	0.0000	0.0000	0.000	0.000	0.0000	0.
5	Commerce Court, Victoria Hotel	0.000000	0.000000	0.0000	0.0000	0.0000	0.000	0.000	0.0000	0.
6	First Canadian Place, Underground city	0.000000	0.000000	0.0000	0.0000	0.0000	0.000	0.000	0.0000	0.
7	Garden District, Ryerson	0.000000	0.000000	0.0000	0.0000	0.0000	0.000	0.000	0.0000	0.
8	Harbourfront East, Union Station, Toronto Islands	0.000000	0.000000	0.0000	0.0000	0.0000	0.000	0.000	0.0000	0.
9	Kensington Market, Chinatown, Grange Park	0.000000	0.000000	0.0000	0.0000	0.0000	0.000	0.000	0.0000	0.
10	Queen's Park, Ontario Provincial Government	0.029412	0.000000	0.0000	0.0000	0.0000	0.000	0.000	0.0000	0.
11	Regent Park, Harbourfront	0.022222	0.000000	0.0000	0.0000	0.0000	0.000	0.000	0.0000	0.
12	Richmond, Adelaide, King	0.000000	0.000000	0.0000	0.0000	0.0000	0.000	0.000	0.0000	0.
13	Rosedale	0.000000	0.000000	0.0000	0.0000	0.0000	0.000	0.000	0.0000	0.
14	St. James Town	0.000000	0.000000	0.0000	0.0000	0.0000	0.000	0.000	0.0000	0.
15	St. James Town,	0.000000	0.000000	0.0000	0.0000	0.0000	0.000	0.000	0.0000	0.

	Cabbagetown									
16	Stn A PO Boxes	0.000000	0.000000	0.0000	0.0000	0.0000	0.000	0.000	0.0000	0.
17	Toronto Dominion Centre, Design Exchange	0.000000	0.000000	0.0000	0.0000	0.0000	0.000	0.000	0.0000	0.
18	University of Toronto, Harbord	0.027778	0.000000	0.0000	0.0000	0.0000	0.000	0.000	0.0000	0.

19 rows × 207 columns

Let's confirm the new size

Let's print each neighborhood along with the top 5 most common venues

```
num_top_venues = 5

for hood in downtown_grouped['Neighborhood']:
    print("----"+hood+"----")
    temp = downtown_grouped[downtown_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  0.05
1          Restaurant  0.04
2  Italian Restaurant  0.04
3                Café  0.04
4  Seafood Restaurant  0.04


----CN Tower, King and Spadina, Railway Lands, Harbourfront West, Bathurst  Quay, South Niagara, I
                venue  freq
0      Airport Lounge  0.12
1     Airport Service  0.12
2         Coffee Shop  0.06
3             Airport  0.06
4   Airport Food Court  0.06


----Central Bay Street----
                 venue  freq
0          Coffee Shop  0.16
1   Italian Restaurant  0.05
2                 Café  0.05
3       Sandwich Place  0.04
4  Japanese Restaurant  0.04


----Christie----
           venue  freq
0  Grocery Store  0.22
1           Café  0.17
2           Park  0.11
3      Nightclub  0.06
4    Gas Station  0.06


----Church and Wellesley----
                 venue  freq
0          Coffee Shop  0.06
1              Gay Bar  0.05
2  Japanese Restaurant  0.05
3    Sushi Restaurant  0.04
4           Restaurant  0.04


----Commerce Court, Victoria Hotel----
         venue  freq
0  Coffee Shop  0.10
1   Restaurant  0.07
2         Café  0.07
3        Hotel  0.06
4          Gym  0.04


----First Canadian Place, Underground city----
         venue  freq
0  Coffee Shop  0.12
1         Café  0.07
2   Restaurant  0.06
3          Gym  0.03
```

```
3          Gym  0.03
4        Hotel  0.03


----Garden District, Ryerson----
                venue  freq
0       Clothing Store  0.09
1          Coffee Shop  0.09
2      Bubble Tea Shop  0.03
3   Italian Restaurant  0.03
4  Japanese Restaurant  0.03


----Harbourfront East, Union Station, Toronto Islands----
               venue  freq
0        Coffee Shop  0.12
1           Aquarium  0.05
2              Hotel  0.04
3         Restaurant  0.04
4  Italian Restaurant  0.04


----Kensington Market, Chinatown, Grange Park----
                   venue  freq
0            Coffee Shop  0.06
1                   Café  0.06
2  Vietnamese Restaurant  0.06
3     Mexican Restaurant  0.05
4                    Bar  0.05


----Queen's Park, Ontario Provincial Government----
                venue  freq
0         Coffee Shop  0.24
1               Diner  0.06
2         Yoga Studio  0.03
3  Distribution Center  0.03
4        Burger Joint  0.03


----Regent Park, Harbourfront----
            venue  freq
0     Coffee Shop  0.16
1            Park  0.07
2          Bakery  0.07
3             Pub  0.07
4  Breakfast Spot  0.04


----Richmond, Adelaide, King----
         venue  freq
0  Coffee Shop  0.09
1         Café  0.05
2          Gym  0.04
3   Restaurant  0.04
4       Bakery  0.03


----Rosedale----
                venue  freq
0                Park  0.50
```

```
0               Park  0.50
1         Playground  0.25
2              Trail  0.25
3  Moroccan Restaurant  0.00
4             Market  0.00


----St. James Town----
          venue  freq
0          Café  0.05
1   Coffee Shop  0.05
2  Cocktail Bar  0.04
3      Beer Bar  0.03
4    Restaurant  0.03


----St. James Town, Cabbagetown----
         venue  freq
0  Coffee Shop  0.09
1   Restaurant  0.07
2         Park  0.07
3       Bakery  0.04
4         Café  0.04


----Stn A PO Boxes----
                venue  freq
0         Coffee Shop  0.11
1  Italian Restaurant  0.04
2  Seafood Restaurant  0.04
3                Café  0.04
4          Restaurant  0.04


----Toronto Dominion Centre, Design Exchange----
                venue  freq
0         Coffee Shop  0.13
1               Hotel  0.08
2                Café  0.07
3          Restaurant  0.05
4  Seafood Restaurant  0.03


----University of Toronto, Harbord----
                venue  freq
0                Café  0.14
1           Bookstore  0.08
2              Bakery  0.06
3          Restaurant  0.06
4  Italian Restaurant  0.06
```

Let's put that into a pandas dataframe¶

```
# First, let's write a function to sort the venues in descending order.
def return_most_common_venues(row, num_top_venues):
```

```
    row_categories = row.iloc[1:]
    row_categories_sorted = row_categories.sort_values(ascending=False)

    return row_categories_sorted.index.values[0:num_top_venues]

# Now let's create the new dataframe and display the top 10 venues for each neighborhood.
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'] = downtown_grouped['Neighborhood']

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

neighborhoods_venues_sorted.head()
```

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th C
0	Berczy Park	Coffee Shop	Seafood Restaurant	Cocktail Bar	Café	Cheese Shop	Res
1	CN Tower, King and Spadina, Railway Lands, Har...	Airport Lounge	Airport Service	Harbor / Marina	Boat or Ferry	Plane	Ren L
2	Central Bay Street	Coffee Shop	Café	Italian Restaurant	Sandwich Place	Middle Eastern Restaurant	Jap Res
3	Christie	Grocery Store	Café	Park	Baby Store	Nightclub	Coffe
4	Church and Wellesley	Coffee Shop	Gay Bar	Japanese Restaurant	Sushi Restaurant	Restaurant	Pizza

4. Cluster Neighborhoods

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

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

```
# set number of clusters
kclusters = 5

downtown_grouped_clustering = downtown_grouped.drop('Neighborhood', 1)

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

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

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

Let's create a new dataframe that includes the cluster as well as the top 10 venues for each neighborhood.

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

downtown_merged = downtown_data

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

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

	key_0	Postal code	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd M Com Ve
0	M5A	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636	0	Coffee Shop	Pub	F
1	M7A	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government	43.662301	-79.389494	0	Coffee Shop	Diner	Y St
2	M5B	M5B	Downtown Toronto	Garden District, Ryerson	43.657162	-79.378937	2	Coffee Shop	Clothing Store	Mi Eas Restau
3	M5C	M5C	Downtown Toronto	St. James Town	43.651494	-79.375418	2	Café	Coffee Shop	Coc
4	M5E	M5E	Downtown Toronto	Berczy Park	43.644771	-79.373306	2	Coffee Shop	Seafood Restaurant	Coc

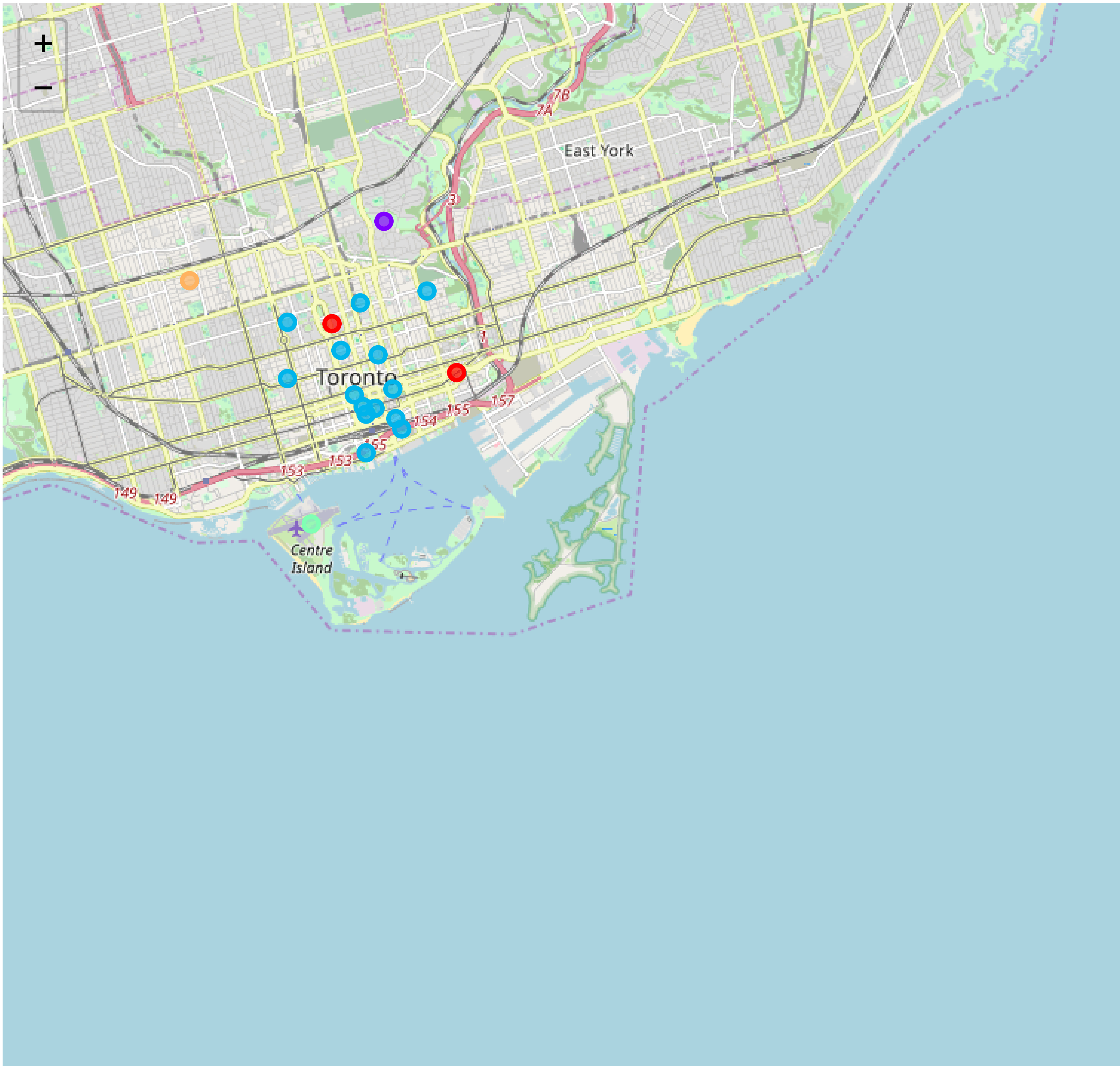
Finally, let's visualize the resulting clusters

```
# 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(downtown_merged['Latitude'], downtown_merged['Longitude'], downtow
    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
```



5. Examine Clusters

Now, you can examine each cluster and determine the discriminating venue categories that distinguish each clu
categories, you can then assign a name to each cluster.

Cluster 1

```
1 downtown_merged.loc[downtown_merged['Cluster Labels'] == 0, downtown_merged.columns[[1] + list(range
```

```
downtown_merged.loc[downtown_merged['Cluster Labels'] == 0, downtown_merged.columns[[1] + list(range
```

	Postal code	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6t
0	M5A	-79.360636	0	Coffee Shop	Pub	Park	Bakery	Café	B
1	M7A	-79.389494	0	Coffee Shop	Diner	Yoga Studio	Beer Bar	Distribution Center	I

Cluster 2

```
downtown_merged.loc[downtown_merged['Cluster Labels'] == 1, downtown_merged.columns[[1] + list(range
```

	Postal code	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6
14	M4W	-79.377529	1	Park	Playground	Trail	Dance Studio	Dumpling Restaurant	Do

Cluster 3

```
downtown_merged.loc[downtown_merged['Cluster Labels'] == 2, downtown_merged.columns[[1] + list(range
```

	Postal code	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
2	M5B	-79.378937	2	Coffee Shop	Clothing Store	Middle Eastern Restaurant	Bubble Tea Shop	Café
3	M5C	-79.375418	2	Café	Coffee Shop	Cocktail Bar	Italian Restaurant	Beer Bar
4	M5E	-79.373306	2	Coffee Shop	Seafood Restaurant	Cocktail Bar	Café	Cheese Shop
5	M5G	-79.387383	2	Coffee Shop	Café	Italian Restaurant	Sandwich Place	Middle Eastern Restaurant
7	M5H	-79.384568	2	Coffee Shop	Café	Restaurant	Gym	Thai Restaurant
8	M5J	-79.381752	2	Coffee Shop	Aquarium	Restaurant	Café	Hotel
9	M5K	-79.381576	2	Coffee Shop	Hotel	Café	Restaurant	Bar
10	M5L	-79.379817	2	Coffee Shop	Café	Restaurant	Hotel	Gym
11	M5S	-79.400049	2	Café	Bookstore	Japanese Restaurant	Bar	Italian Restaurant
12	M5T	-79.400049	2	Café	Vietnamese Restaurant	Coffee Shop	Bar	Vegetarian / Vegan Restaurant
15	M5W	-79.374846	2	Coffee Shop	Restaurant	Seafood Restaurant	Café	Italian Restaurant
16	M4X	-79.367675	2	Coffee Shop	Restaurant	Park	Pub	Italian Restaurant
17	M5X	-79.382280	2	Coffee Shop	Café	Restaurant	Gym	Seafood Restaurant
18	M4Y	-79.383160	2	Coffee Shop	Gay Bar	Japanese Restaurant	Sushi Restaurant	Restaurant

Cluster 4

```
downtown_merged.loc[downtown_merged['Cluster Labels'] == 3, downtown_merged.columns[[1] + list(range
```

	Postal code	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6
13	M5V	-79.39442	3	Airport Lounge	Airport Service	Harbor / Marina	Boat or Ferry	Plane	R

Cluster 5

```
1 downtown_merged.loc[downtown_merged['Cluster Labels'] == 4, downtown_merged.columns[[1] + list(range
```

	Postal code	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6
6	M6G	-79.422564	4	Grocery Store	Café	Park	Baby Store	Nightclub	Coff

```
1
```

/