

In []:

Segmenting and Clustering Neighborhoods in Toronto

first i am importing the needed packages

```
In [1]: import requests
import numpy as np
import lxml.html as lh
import pandas as pd
```

```
In [2]: url='https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M'
#Create a handle, page, to handle the contents of the website
page = requests.get(url)
#Store the contents of the website under doc
doc = lh.fromstring(page.content)
#Parse data that are stored between <tr>..</tr> of HTML
tr_elements = doc.xpath('//tr')
```

I Used request to read the postcodes from HTML page

I used a code found in a website, and I modified it to solve the requeriments

```
In [3]: tr_elements = doc.xpath('//tr')
#Create empty list
col=[]
i=0
#For each row, store each first element (header) and an empty list
for t in tr_elements[0]:
    i+=1
    name=t.text_content()
    print (i,name)
    col.append((name,[]))
```

- 1 Postcode
- 2 Borough
- 3 Neighbourhood

```
In [4]: col
```

```
Out[4]: [('Postcode', []), ('Borough', []), ('Neighbourhood\n', [])]
```

```
In [5]: #i created Col2, because i did not like the column name with \n  
col2=[('Postcode', []), ('Borough', []), ('Neighbourhood', [])]
```

```
In [6]: col2
```

```
Out[6]: [('Postcode', []), ('Borough', []), ('Neighbourhood', [])]
```

```
In [7]: #Since out first row is the header, data is stored on the second row onwards  
for j in range(1,len(tr_elements)):  
    #T is our j'th row  
    T=tr_elements[j]  
  
    #If row is not of size 3In the , the //tr data is not from our table  
    if len(T)!=3:  
        break  
  
    #i is the index of our column  
    i=0  
  
    #Iterate through each element of the row  
    for t in T.iterchildren():  
        data=t.text_content()  
        #Check if row is empty  
        if i>0:  
            #Convert any numerical value to integers  
            try:  
                data=int(data)  
            except:  
                pass  
        #Append the data to the empty list of the i'th column  
        col2[i][1].append(data)  
        #Increment i for the next column  
        i+=1
```

```
In [8]: [len(C) for (title,C) in col2]
```

```
Out[8]: [288, 288, 288]
```

```
In [9]: Dict={title:column for (title,column) in col2}  
df=pd.DataFrame(Dict)
```

```
In [11]: #removing \n character  
df = df.replace('\n',' ', regex=True)
```

```
In [10]: df.head()
```

```
Out[10]:
```

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

removing '\n' character

```
In [12]: df.head()
```

```
Out[12]:
```

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

```
In [ ]: #taking out rows with Borough == 'Not assigned'
```

```
In [13]: #drop not assigned  
ds=df[df.Borough != 'Not assigned']
```

```
In [14]: ds.head()
```

```
Out[14]:
```

	Postcode	Borough	Neighbourhood
2	M3A	North York	Parkwoods
3	M4A	North York	Victoria Village
4	M5A	Downtown Toronto	Harbourfront
5	M5A	Downtown Toronto	Regent Park
6	M6A	North York	Lawrence Heights

```
In [15]: #reset index
ds.reset_index(drop=True, inplace=True)
ds.head()
```

```
Out[15]:
```

	Postcode	Borough	Neighbourhood
0	M3A	North York	Parkwoods
1	M4A	North York	Victoria Village
2	M5A	Downtown Toronto	Harbourfront
3	M5A	Downtown Toronto	Regent Park
4	M6A	North York	Lawrence Heights

```
In [20]: #ds['Neighbourhood'] = ds['Product Name'].str.replace('Not assigned', ds['Borough'])
#df['A'] = np.where(df.B.isin(['X', 'Y', 'Z']), 'T', df['A'])
ds['Neighbourhood'] = np.where(ds.Neighbourhood.isin(['Not assigned']), ds['Borough'], ds['Neighbourhood'])
```

/home/jupyterlab/conda/lib/python3.6/site-packages/ipykernel_launcher.py:3: SettingWithCopyWarning:

A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: <http://pandas.pydata.org/pandas-docs/stable/indexing.html#indexing-view-versus-copy>

This is separate from the ipykernel package so we can avoid doing imports until

if Neighbourhood is in Not Assigned then take Borough value

```
In [22]: #Queen's park has been asigned to Neighbourhood
ds.head(10)
```

Out[22]:

	Postcode	Borough	Neighbourhood
0	M3A	North York	Parkwoods
1	M4A	North York	Victoria Village
2	M5A	Downtown Toronto	Harbourfront
3	M5A	Downtown Toronto	Regent Park
4	M6A	North York	Lawrence Heights
5	M6A	North York	Lawrence Manor
6	M7A	Queen's Park	Queen's Park
7	M9A	Etobicoke	Islington Avenue
8	M1B	Scarborough	Rouge
9	M1B	Scarborough	Malvern

Creating the last Data Set; grouping by Postcode & Borough

```
In [26]: nds=ds.groupby(['Postcode', 'Borough'])['Neighbourhood'].apply(', '.join).reset_index()  
nds
```

Out[26]:

	Postcode	Borough	Neighbourhood
0	M1B	Scarborough	Rouge,Malvern
1	M1C	Scarborough	Highland Creek,Rouge Hill,Port Union
2	M1E	Scarborough	Guildwood,Morningside,West Hill
3	M1G	Scarborough	Woburn
4	M1H	Scarborough	Cedarbrae
5	M1J	Scarborough	Scarborough Village
6	M1K	Scarborough	East Birchmount Park,Ionview,Kennedy Park
7	M1L	Scarborough	Clairlea,Golden Mile,Oakridge
8	M1M	Scarborough	Cliffcrest,Cliffside,Scarborough Village West
9	M1N	Scarborough	Birch Cliff,Cliffside West
10	M1P	Scarborough	Dorset Park,Scarborough Town Centre,Wexford He...
11	M1R	Scarborough	Maryvale,Wexford
12	M1S	Scarborough	Agincourt
13	M1T	Scarborough	Clarks Corners,Sullivan,Tam O'Shanter
14	M1V	Scarborough	Agincourt North,L'Amoreaux East,Milliken,Steel...
15	M1W	Scarborough	L'Amoreaux West
16	M1X	Scarborough	Upper Rouge
17	M2H	North York	Hillcrest Village
18	M2J	North York	Fairview,Henry Farm,Oriole
19	M2K	North York	Bayview Village
20	M2L	North York	Silver Hills,York Mills
21	M2M	North York	Newtonbrook,Willowdale
22	M2N	North York	Willowdale South
23	M2P	North York	York Mills West
24	M2R	North York	Willowdale West
25	M3A	North York	Parkwoods
26	M3B	North York	Don Mills North
27	M3C	North York	Flemingdon Park,Don Mills South
28	M3H	North York	Bathurst Manor,Downsview North,Wilson Heights
29	M3J	North York	Northwood Park,York University
...
73	M6C	York	Humewood-Cedarvale
74	M6E	York	Caledonia-Fairbanks
75	M6G	Downtown Toronto	Christie
76	M6H	West Toronto	Dovercourt Village,Dufferin

	Postcode	Borough	Neighbourhood
77	M6J	West Toronto	Little Portugal, Trinity
78	M6K	West Toronto	Brockton, Exhibition Place, Parkdale Village
79	M6L	North York	Downsview, North Park, Upwood Park
80	M6M	York	Del Ray, Keelesdale, Mount Dennis, Silverthorn
81	M6N	York	The Junction North, Runnymede
82	M6P	West Toronto	High Park, The Junction South
83	M6R	West Toronto	Parkdale, Roncesvalles
84	M6S	West Toronto	Runnymede, Swansea
85	M7A	Queen's Park	Queen's Park
86	M7R	Mississauga	Canada Post Gateway Processing Centre
87	M7Y	East Toronto	Business Reply Mail Processing Centre 969 Eastern
88	M8V	Etobicoke	Humber Bay Shores, Mimico South, New Toronto
89	M8W	Etobicoke	Alderwood, Long Branch
90	M8X	Etobicoke	The Kingsway, Montgomery Road, Old Mill North
91	M8Y	Etobicoke	Humber Bay, King's Mill Park, Kingsway Park Sout...
92	M8Z	Etobicoke	Kingsway Park South West, Mimico NW, The Queensw...
93	M9A	Etobicoke	Islington Avenue
94	M9B	Etobicoke	Cloverdale, Islington, Martin Grove, Princess Gar...
95	M9C	Etobicoke	Bloordale Gardens, Eringate, Markland Wood, Old B...
96	M9L	North York	Humber Summit
97	M9M	North York	Emery, Humberlea
98	M9N	York	Weston
99	M9P	Etobicoke	Westmount
100	M9R	Etobicoke	Kingsview Village, Martin Grove Gardens, Richvie...
101	M9V	Etobicoke	Albion Gardens, Beaumont Heights, Humbergate, Jam...
102	M9W	Etobicoke	Northwest

103 rows × 3 columns

In []:

In [25]:

```
#printing the Dtaset shape.
nds.shape
```

Out[25]: (103, 3)

In []:

In [32]:

```
filename = "https://coc1.us/Geospatial_data"
```



```
In [33]: coord = pd.read_csv(filename)
```

```
In [34]: coord.head()
```

```
Out[34]:
```

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

```
In [35]: nds.head()
```

```
Out[35]:
```

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

```
In [36]: nds.rename(columns={'Postcode':'Postal Code'}, inplace=True)
```

```
In [38]: nds.dtypes
```

```
Out[38]: Postal Code    object
Borough                object
Neighbourhood          object
dtype: object
```

joining two data sets

```
In [39]: Tor=pd.merge(nds, coord, on='Postal Code', how='left')
```

In [41]: `Tor.head(12)`

Out[41]:

	Postal Code	Borough	Neighbourhood	Latitude	Longitude
0	M1B	Scarborough	Rouge,Malvern	43.806686	-79.194353
1	M1C	Scarborough	Highland Creek,Rouge Hill,Port Union	43.784535	-79.160497
2	M1E	Scarborough	Guildwood,Morningside,West Hill	43.763573	-79.188711
3	M1G	Scarborough	Woburn	43.770992	-79.216917
4	M1H	Scarborough	Cedarbrae	43.773136	-79.239476
5	M1J	Scarborough	Scarborough Village	43.744734	-79.239476
6	M1K	Scarborough	East Birchmount Park,Ionview,Kennedy Park	43.727929	-79.262029
7	M1L	Scarborough	Clairlea,Golden Mile,Oakridge	43.711112	-79.284577
8	M1M	Scarborough	Cliffcrest,Cliffside,Scarborough Village West	43.716316	-79.239476
9	M1N	Scarborough	Birch Cliff,Cliffside West	43.692657	-79.264848
10	M1P	Scarborough	Dorset Park,Scarborough Town Centre,Wexford He...	43.757410	-79.273304
11	M1R	Scarborough	Maryvale,Wexford	43.750072	-79.295849

Subsetting the Data Set to select Borough with words==Toronto

In [44]: `Toronto = Tor[Tor['Borough'].str.contains('Toronto')].reset_index(drop=True)`
`Toronto .head()`

Out[44]:

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

Import libraries

```
In [45]: import json # library to handle JSON files

#!conda install -c conda-forge geopy --yes # uncomment this line if you have
n't completed the Foursquare API lab
from geopy.geocoders import Nominatim # convert an address into latitude and longitude values

import requests # library to handle requests
from pandas.io.json import json_normalize # transform 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
```

```

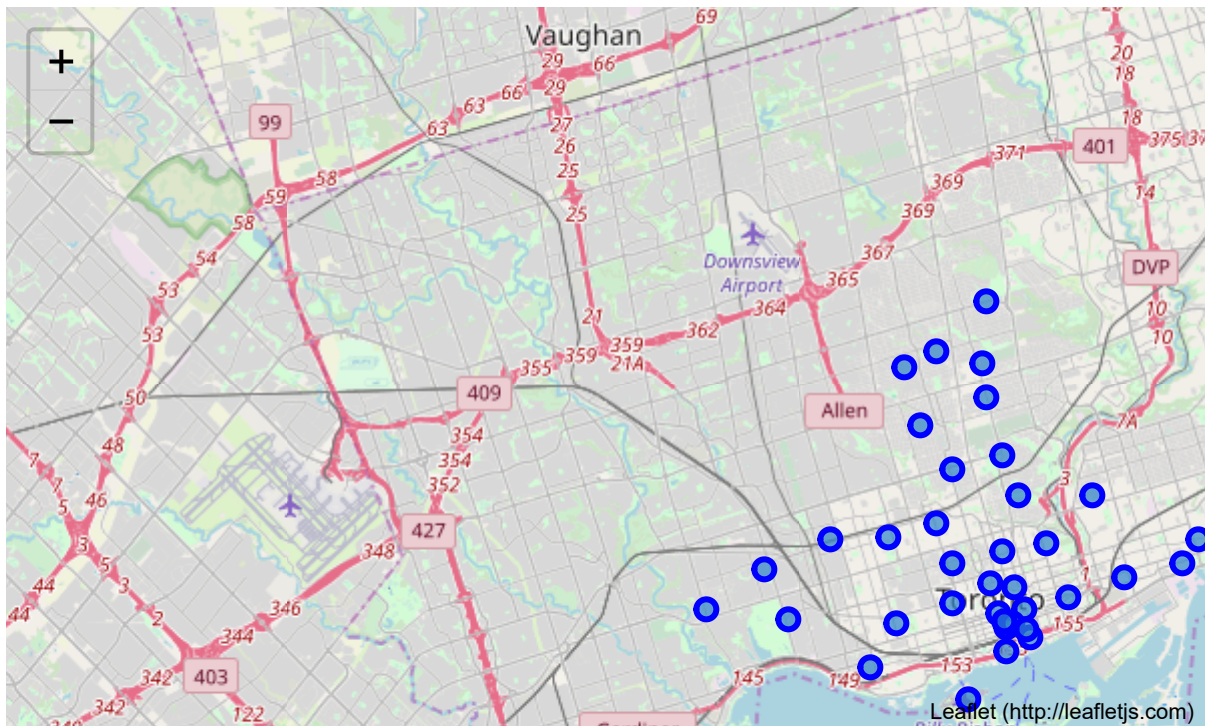
In [47]: # create map of Toronto using Latitude and Longitude values
map_Tor = folium.Map(location=[latitude, longitude], zoom_start=11)

# add markers to map
for lat, lng, label in zip(Toronto['Latitude'], Toronto['Longitude'], Toronto[
    'Neighbourhood']):
    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_Tor)

map_Tor

```

Out[47]:



```
In [51]: #definig Foursquare Credentials
CLIENT_ID = 'YA3G4TBQ4M5XTM1CPAAP32RYCGNQW4F0MAXXRWZU2R3N1KWQ' # your Foursqua
re ID
CLIENT_SECRET = 'DTIPAYCVDLIMU53HUCN4WRH40RVL4BOZEF3M404FIJJHHQVW' # your Four
square Secret
VERSION = '20180605' # Foursquare API version

print('Your credentails:')
print('CLIENT_ID: ' + CLIENT_ID)
print('CLIENT_SECRET:' + CLIENT_SECRET)
radius=500
LIMIT=100
```

Your credentails:

CLIENT_ID: YA3G4TBQ4M5XTM1CPAAP32RYCGNQW4F0MAXXRWZU2R3N1KWQ

CLIENT_SECRET:DTIPAYCVDLIMU53HUCN4WRH40RVL4BOZEF3M404FIJJHHQVW

Function to extract venues from Foursquare API

```

In [52]: 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)

```

In [58]: *# type your answer here*

```
Tor_venues = getNearbyVenues(names=Toronto['Neighbourhood'],  
                             latitudes=Toronto['Latitude'],  
                             longitudes=Toronto['Longitude']  
                             )
```

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

```
In [59]: print(Tor_venues.shape)
Tor_venues.head()
```

(1695, 7)

Out[59]:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	The Beaches	43.676357	-79.293031	The Big Carrot Natural Food Market	43.678879	-79.297734	Health Food Store
1	The Beaches	43.676357	-79.293031	Grover Pub and Grub	43.679181	-79.297215	Pub
2	The Beaches	43.676357	-79.293031	Glen Manor Ravine	43.676821	-79.293942	Trail
3	The Beaches	43.676357	-79.293031	Upper Beaches	43.680563	-79.292869	Neighborhood
4	The Danforth West,Riverdale	43.679557	-79.352188	Pantheon	43.677621	-79.351434	Greek Restaurant

```
In [57]: Tor_venues.dtypes
```

```
Out[57]: Neighborhood      object
Neighborhood Latitude      float64
Neighborhood Longitude      float64
Venue                      object
Venue Latitude              float64
Venue Longitude              float64
Venue Category              object
dtype: object
```



```
In [60]: Tor_venues.groupby('Neighborhood').count()
```

Out[60]:

	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
Adelaide,King,Richmond	100	100	100	100	100	100
Berczy Park	55	55	55	55	55	55
Brockton,Exhibition Place,Parkdale Village	21	21	21	21	21	21
Business Reply Mail Processing Centre 969 Eastern	19	19	19	19	19	19
CN Tower,Bathurst Quay,Island airport,Harbourfront West,King and Spadina,Railway Lands,South Niagara	14	14	14	14	14	14
Cabbagetown,St. James Town	46	46	46	46	46	46
Central Bay Street	88	88	88	88	88	88
Chinatown,Grange Park,Kensington Market	100	100	100	100	100	100
Christie	15	15	15	15	15	15
Church and Wellesley	87	87	87	87	87	87
Commerce Court,Victoria Hotel	100	100	100	100	100	100
Davisville	38	38	38	38	38	38
Davisville North	8	8	8	8	8	8
Deer Park,Forest Hill SE,Rathnelly,South Hill,Summerhill West	14	14	14	14	14	14
Design Exchange,Toronto Dominion Centre	100	100	100	100	100	100
Dovercourt Village,Dufferin	20	20	20	20	20	20
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	34	34	34	34	34	34
Harbourfront East,Toronto Islands,Union Station	100	100	100	100	100	100
Harbourfront,Regent Park	48	48	48	48	48	48
High Park,The Junction South	23	23	23	23	23	23
Lawrence Park	3	3	3	3	3	3
Little Portugal,Trinity	62	62	62	62	62	62

	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
Moore Park,Summerhill East	2	2	2	2	2	2
North Toronto West	16	16	16	16	16	16
Parkdale,Roncesvalles	15	15	15	15	15	15
Rosedale	4	4	4	4	4	4
Roselawn	2	2	2	2	2	2
Runnymede,Swansea	38	38	38	38	38	38
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	94	94	94	94	94	94
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	18	18	18	18	18	18
The Danforth West,Riverdale	42	42	42	42	42	42

Analyze Each Neighborhood

```
In [61]: # one hot encoding
Tor_onehot = pd.get_dummies(Tor_venues[['Venue Category']], prefix="", prefix_sep="")

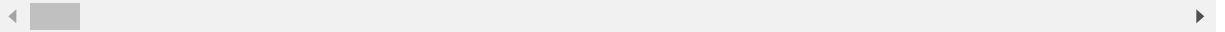
# add neighborhood column back to dataframe
Tor_onehot['Neighborhood'] = Tor_venues['Neighborhood']

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

Tor_onehot.head()
```

Out[61]:

	Yoga Studio	Adult Boutique	Afghan Restaurant	Airport	Airport Food Court	Airport Gate	Airport Lounge	Airport Service	Airport Terminal	American Restaurant
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 [62]: Tor_onehot.shape
```

Out[62]: (1695, 237)

```
In [63]: Tor_grouped = Tor_onehot.groupby('Neighborhood').mean().reset_index()  
Tor_grouped
```

Out[63]:

	Neighborhood	Yoga Studio	Adult Boutique	Afghan Restaurant	Airport	Airport Food Court	Airport Gate	Airport Lounge
0	Adelaide,King,Richmond	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0000
1	Berczy Park	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0000
2	Brockton,Exhibition Place,Parkdale Village	0.047619	0.000000	0.000000	0.000000	0.000000	0.000000	0.0000
3	Business Reply Mail Processing Centre 969 Eastern	0.052632	0.000000	0.000000	0.000000	0.000000	0.000000	0.0000
4	CN Tower,Bathurst Quay,Island airport,Harbourf...	0.000000	0.000000	0.000000	0.071429	0.071429	0.071429	0.1428
5	Cabbagetown,St. James Town	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0000
6	Central Bay Street	0.011364	0.000000	0.000000	0.000000	0.000000	0.000000	0.0000
7	Chinatown,Grange Park,Kensington Market	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0000
8	Christie	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0000
9	Church and Wellesley	0.011494	0.011494	0.011494	0.000000	0.000000	0.000000	0.0000
10	Commerce Court,Victoria Hotel	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0000
11	Davisville	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0000
12	Davisville North	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0000
13	Deer Park,Forest Hill SE,Rathnelly,South Hill,...	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0000
14	Design Exchange,Toronto Dominion Centre	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0000
15	Dovercourt Village,Dufferin	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0000
16	First Canadian Place,Underground city	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0000
17	Forest Hill North,Forest Hill West	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0000
18	Harbord,University of Toronto	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0000
19	Harbourfront East,Toronto Islands,Union Station	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0000
20	Harbourfront,Regent Park	0.020833	0.000000	0.000000	0.000000	0.000000	0.000000	0.0000
21	High Park,The Junction South	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0000
22	Lawrence Park	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0000
23	Little Portugal,Trinity	0.016129	0.000000	0.000000	0.000000	0.000000	0.000000	0.0000

	Neighborhood	Yoga Studio	Adult Boutique	Afghan Restaurant	Airport	Airport Food Court	Airport Gate	Airport Lounge
24	Moore Park,Summerhill East	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0000
25	North Toronto West	0.062500	0.000000	0.000000	0.000000	0.000000	0.000000	0.0000
26	Parkdale,Roncesvalles	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0000
27	Rosedale	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0000
28	Roselawn	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0000
29	Runnymede,Swansea	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0000
30	Ryerson,Garden District	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0000
31	St. James Town	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0000
32	Stn A PO Boxes 25 The Esplanade	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0000
33	Studio District	0.026316	0.000000	0.000000	0.000000	0.000000	0.000000	0.0000
34	The Annex,North Midtown,Yorkville	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0000
35	The Beaches	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0000
36	The Beaches West,India Bazaar	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0000
37	The Danforth West,Riverdale	0.023810	0.000000	0.000000	0.000000	0.000000	0.000000	0.0000

In [64]: `Tor_grouped.shape`

Out[64]: (38, 237)

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

```
In [65]: num_top_venues = 5

for hood in Tor_grouped['Neighborhood']:
    print("-----"+hood+"-----")
    temp = Tor_grouped[Tor_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	0.06
1	Café	0.05
2	Steakhouse	0.04
3	Bar	0.04
4	American Restaurant	0.04

----Berczy Park----

	venue	freq
0	Coffee Shop	0.09
1	Cocktail Bar	0.05
2	Bakery	0.04
3	Italian Restaurant	0.04
4	Farmers Market	0.04

----Brockton,Exhibition Place,Parkdale Village----

	venue	freq
0	Breakfast Spot	0.10
1	Café	0.10
2	Coffee Shop	0.10
3	Yoga Studio	0.05
4	Gym	0.05

----Business Reply Mail Processing Centre 969 Eastern----

	venue	freq
0	Yoga Studio	0.05
1	Auto Workshop	0.05
2	Park	0.05
3	Comic Shop	0.05
4	Recording Studio	0.05

----CN Tower,Bathurst Quay,Island airport,Harbourfront West,King and Spadina,
Railway Lands,South Niagara----

	venue	freq
0	Airport Lounge	0.14
1	Airport Service	0.14
2	Airport Terminal	0.14
3	Boutique	0.07
4	Sculpture Garden	0.07

----Cabbagetown,St. James Town----

	venue	freq
0	Coffee Shop	0.09
1	Restaurant	0.07
2	Pub	0.04
3	Café	0.04
4	Park	0.04

----Central Bay Street----

	venue	freq
--	-------	------

0	Coffee Shop	0.16
1	Italian Restaurant	0.05
2	Café	0.05
3	Burger Joint	0.03
4	Middle Eastern Restaurant	0.03

----Chinatown,Grange Park,Kensington Market----

	venue	freq
0	Café	0.08
1	Vegetarian / Vegan Restaurant	0.06
2	Mexican Restaurant	0.04
3	Dumpling Restaurant	0.04
4	Bar	0.04

----Christie----

	venue	freq
0	Grocery Store	0.20
1	Café	0.20
2	Park	0.13
3	Convenience Store	0.07
4	Baby Store	0.07

----Church and Wellesley----

	venue	freq
0	Japanese Restaurant	0.07
1	Coffee Shop	0.07
2	Sushi Restaurant	0.06
3	Restaurant	0.03
4	Gay Bar	0.03

----Commerce Court,Victoria Hotel----

	venue	freq
0	Coffee Shop	0.11
1	Café	0.07
2	Hotel	0.06
3	Restaurant	0.05
4	American Restaurant	0.04

----Davisville----

	venue	freq
0	Pizza Place	0.11
1	Dessert Shop	0.08
2	Sandwich Place	0.08
3	Restaurant	0.05
4	Sushi Restaurant	0.05

----Davisville North----

	venue	freq
0	Grocery Store	0.12
1	Gym	0.12
2	Clothing Store	0.12

3 Park 0.12
4 Sandwich Place 0.12

----Deer Park,Forest Hill SE,Rathnelly,South Hill,Summerhill West----

	venue	freq
0	Pub	0.14
1	Coffee Shop	0.14
2	Light Rail Station	0.07
3	Pizza Place	0.07
4	Bagel Shop	0.07

----Design Exchange,Toronto Dominion Centre----

	venue	freq
0	Coffee Shop	0.12
1	Café	0.07
2	Hotel	0.06
3	Restaurant	0.05
4	Deli / Bodega	0.03

----Dovercourt Village,Dufferin----

	venue	freq
0	Pharmacy	0.10
1	Supermarket	0.10
2	Bakery	0.10
3	Middle Eastern Restaurant	0.05
4	Liquor Store	0.05

----First Canadian Place,Underground city----

	venue	freq
0	Café	0.09
1	Coffee Shop	0.08
2	Restaurant	0.04
3	Hotel	0.04
4	Bakery	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	Miscellaneous Shop	0.00

----Harbord,University of Toronto----

	venue	freq
0	Café	0.12
1	Italian Restaurant	0.06
2	Japanese Restaurant	0.06
3	Bar	0.06
4	Bakery	0.06

----Harbourfront East,Toronto Islands,Union Station----

	venue	freq
0	Coffee Shop	0.12
1	Hotel	0.05
2	Aquarium	0.05
3	Italian Restaurant	0.04
4	Café	0.04

----Harbourfront,Regent Park----

	venue	freq
0	Coffee Shop	0.17
1	Park	0.06
2	Bakery	0.06
3	Pub	0.06
4	Mexican Restaurant	0.04

----High Park,The Junction South----

	venue	freq
0	Café	0.09
1	Mexican Restaurant	0.09
2	Bar	0.09
3	Furniture / Home Store	0.04
4	Park	0.04

----Lawrence Park----

	venue	freq
0	Swim School	0.33
1	Park	0.33
2	Bus Line	0.33
3	Yoga Studio	0.00
4	Nightclub	0.00

----Little Portugal,Trinity----

	venue	freq
0	Bar	0.13
1	Asian Restaurant	0.05
2	Men's Store	0.05
3	Coffee Shop	0.05
4	New American Restaurant	0.03

----Moore Park,Summerhill East----

	venue	freq
0	Playground	0.5
1	Tennis Court	0.5
2	Plane	0.0
3	Pizza Place	0.0
4	Men's Store	0.0

----North Toronto West----

	venue	freq
--	-------	------

0	Coffee Shop	0.12
1	Yoga Studio	0.06
2	Fast Food Restaurant	0.06
3	Spa	0.06
4	Chinese Restaurant	0.06

----Parkdale,Roncesvalles----

	venue	freq
0	Gift Shop	0.13
1	Breakfast Spot	0.13
2	Bookstore	0.07
3	Dessert Shop	0.07
4	Dog Run	0.07

----Rosedale----

	venue	freq
0	Park	0.50
1	Playground	0.25
2	Trail	0.25
3	Nightclub	0.00
4	Men's Store	0.00

----Roselawn----

	venue	freq
0	Ice Cream Shop	0.5
1	Garden	0.5
2	Yoga Studio	0.0
3	Noodle House	0.0
4	Middle Eastern Restaurant	0.0

----Runnymede,Swansea----

	venue	freq
0	Café	0.08
1	Pizza Place	0.08
2	Coffee Shop	0.08
3	Italian Restaurant	0.05
4	Diner	0.05

----Ryerson,Garden District----

	venue	freq
0	Coffee Shop	0.10
1	Clothing Store	0.06
2	Café	0.04
3	Cosmetics Shop	0.04
4	Middle Eastern Restaurant	0.03

----St. James Town----

	venue	freq
0	Coffee Shop	0.07
1	Café	0.06
2	Hotel	0.05

3	Restaurant	0.05
4	Cosmetics Shop	0.04

----Stn A PO Boxes 25 The Esplanade----

	venue	freq
0	Coffee Shop	0.11
1	Café	0.04
2	Restaurant	0.04
3	Cocktail Bar	0.03
4	Seafood Restaurant	0.03

----Studio District----

	venue	freq
0	Café	0.11
1	Coffee Shop	0.08
2	Italian Restaurant	0.05
3	Gastropub	0.05
4	Bakery	0.05

----The Annex,North Midtown,Yorkville----

	venue	freq
0	Café	0.13
1	Sandwich Place	0.13
2	Coffee Shop	0.13
3	Pizza Place	0.09
4	History Museum	0.04

----The Beaches----

	venue	freq
0	Health Food Store	0.25
1	Pub	0.25
2	Trail	0.25
3	Yoga Studio	0.00
4	Monument / Landmark	0.00

----The Beaches West,India Bazaar----

	venue	freq
0	Park	0.11
1	Gym	0.06
2	Sushi Restaurant	0.06
3	Sandwich Place	0.06
4	Steakhouse	0.06

----The Danforth West,Riverdale----

	venue	freq
0	Greek Restaurant	0.21
1	Coffee Shop	0.10
2	Italian Restaurant	0.07
3	Ice Cream Shop	0.07
4	Furniture / Home Store	0.05

Let's put that into a pandas dataframe

First, let's write a function to sort the venues in descending order.

```
In [66]: 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.

```

In [72]: 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'] = Tor_grouped['Neighborhood']

for ind in np.arange(Tor_grouped.shape[0]):
    neighborhoods_venues_sorted.iloc[ind, 1:] = return_most_common_venues(Tor_
grouped.iloc[ind, :], num_top_venues)

neighborhoods_venues_sorted.head()

```

Out[72]:

	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	American Restaurant	Thai Restaurant	Steakhouse	Bar	
1	Berczy Park	Coffee Shop	Cocktail Bar	Seafood Restaurant	Café	Cheese Shop	Farmers Market	Be
2	Brockton,Exhibition Place,Parkdale Village	Coffee Shop	Café	Breakfast Spot	Yoga Studio	Italian Restaurant	Pet Store	Cl
3	Business Reply Mail Processing Centre 969 Eastern	Yoga Studio	Auto Workshop	Pizza Place	Gym / Fitness Center	Recording Studio	Restaurant	E
4	CN Tower,Bathurst Quay,Island airport,Harbourf...	Airport Lounge	Airport Service	Airport Terminal	Boat or Ferry	Harbor / Marina	Boutique	

Cluster Neighborhoods


```
In [68]: # set number of clusters
kclusters = 5

Tor_grouped_clustering = Tor_grouped.drop('Neighborhood', 1)

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

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

```
Out[68]: array([0, 0, 0, 0, 0, 0, 0, 0, 0, 0], dtype=int32)
```

```
In [73]: # add clustering labels
neighborhoods_venues_sorted.insert(0, 'Cluster Labels', kmeans.labels_)

Tor_merged = Toronto.rename(columns={'Neighbourhood': 'Neighborhood'})

# merge toronto_grouped with toronto_data to add Latitude/Longitude for each neighborhood
Tor_merged = Tor_merged.join(neighborhoods_venues_sorted.set_index('Neighborhood'), on='Neighborhood')

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

```
Out[73]:
```

	Postal Code	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
0	M4E	East Toronto	The Beaches	43.676357	-79.293031	0	Trail	Pub	Head FC Str
1	M4K	East Toronto	The Danforth West, Riverdale	43.679557	-79.352188	0	Greek Restaurant	Coffee Shop	Ice Cream Sh
2	M4L	East Toronto	The Beaches West, India Bazaar	43.668999	-79.315572	0	Park	Pizza Place	Su Restaur
3	M4M	East Toronto	Studio District	43.659526	-79.340923	0	Café	Coffee Shop	Gastrop
4	M4N	Central Toronto	Lawrence Park	43.728020	-79.388790	4	Park	Swim School	Bus L

```
In [ ]:
```

```

In [75]: # create map
map_clusters = folium.Map(location=[latitude, longitude], zoom_start=11)

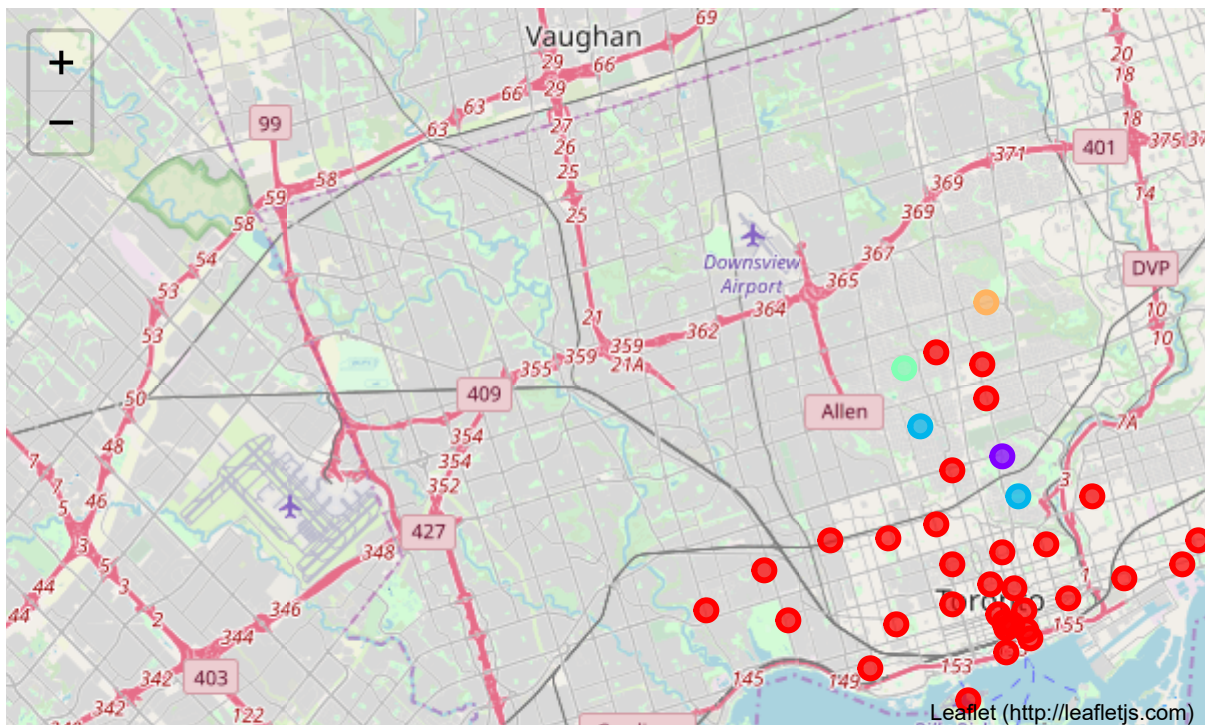
# set color scheme for the clusters
x = np.arange(kclusters)
ys = [i + x + (i*x)**2 for i in range(kclusters)]
colors_array = cm.rainbow(np.linspace(0, 1, len(ys)))
rainbow = [colors.rgb2hex(i) for i in colors_array]

# add markers to the map
markers_colors = []
for lat, lon, poi, cluster in zip(Tor_merged['Latitude'], Tor_merged['Longitude'], Tor_merged['Neighborhood'], Tor_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[75]:



In []:

In []:

In []:

In []: