

In [1]:

```
import pandas as pd
import wikipedia as wp
from bs4 import BeautifulSoup
%matplotlib inline
```

We get the data and load it into a pandas's DataFrame

In [2]:

```
html = wp.page("List of postal codes of Canada: M").html().encode("UTF-8")
```

In [3]:

```
df = pd.read_html(html, header = 0)[0]
```

We ignore cells with a borough that is Not assigned.

In [4]:

```
df = df[df.Borough != 'Not assigned']
```

More than one neighbourhood can exist in one postal code area. We combine neighbourhoods in a Neighbourhood column, separating values with a comma.

In [5]:

```
df = df.groupby(['Postcode', 'Borough'])['Neighbourhood'].apply(list).apply(lambda x:','.join(x)).to_frame()
```

We reset the index of the DataFrame

In [6]:

```
df = df.reset_index()
```

If a cell has a borough but a Not assigned neighbourhood, then the Neighbourhood will be set to name of the borough.

In [7]:

```
for index, row in df.iterrows():
    if row['Neighbourhood'] == 'Not assigned':
        row['Neighbourhood'] = row['Borough']
```

Now that we have built a dataframe of the postal code of each neighborhood along with the borough name and neighborhood name, in order to utilize the Foursquare location data, we need to get the latitude and the longitude coordinates of each neighborhood.

In [8]:

```
import requests
import io
```

In [9]:

```
url = "http://cocl.us/Geospatial_data"
alpha = requests.get(url).content
df2 = pd.read_csv(io.StringIO(alpha.decode('utf-8')))
```

We rename the first column to allow merging the two DataFrames on Postcode

In [10]:

```
df2.columns = ['Postcode', 'Latitude', 'Longitude']
df3 = pd.merge(df2, df, on='Postcode')
```

We then reorder column names and show the dataframe

In [11]:

```
df3 = df3[['Postcode', 'Borough', 'Neighbourhood', 'Latitude', 'Longitude']]
df3.head()
```

Out[11]:

	Postcode	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

Visualize Map of Toronto

In [12]:

```
import folium
```

We want to create map of Toronto using latitude and longitude values

In [13]:

```
toronto_latitude = 43.6532; toronto_longitude = -79.3832
map_toronto = folium.Map(location = [toronto_latitude, toronto_longitude], zoom_start = 10)
```

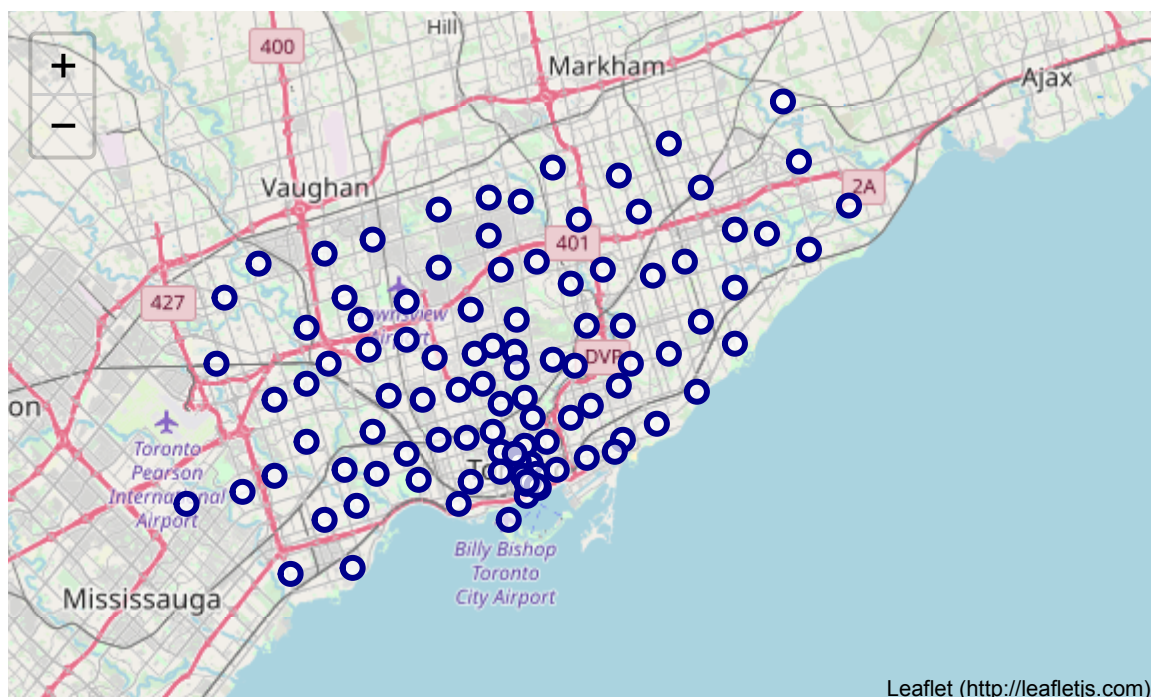
We then add markers to the map and display the map

In [14]:

```
for lat, lng, borough, neighborhood in zip(df3['Latitude'], df3['Longitude'], df3['Borough'], df3['Neighbourhood']):
    label = '{} , {}'.format(neighborhood, borough)
    label = folium.Popup(label, parse_html=True)
    folium.CircleMarker(
        [lat, lng],
        radius=5,
        popup=label,
        color='darkblue',
        fill=True,
        fill_color='white',
        fill_opacity=0.7).add_to(map_toronto)
```

map_toronto

Out[14]:



We create a new dataframe with neighborhoods in Scarborough

In [15]:

```
Scarborough_data = df3[df3['Borough'] == 'Scarborough'].reset_index(drop = True)
Scarborough_data.head()
```

Out[15]:

	Postcode	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

In [16]:

```
Scarborough_latitude = 43.77223; Scarborough_longitude = -79.25666
map_Scarborough = folium.Map(location = [Scarborough_latitude, Scarborough_longitude],
                              zoom_start = 11)
for lat,lng,label in zip(Scarborough_data['Latitude'],Scarborough_data['Longitude'],Scarborough_data['Neighbourhood']):
    label = folium.Popup(label, parse_html=True)
    folium.CircleMarker(
        [lat, lng],
        radius=5,
        popup=label,
        color='darkblue',
        fill=True,
        fill_color='white',
        fill_opacity=0.7).add_to(map_Scarborough)
map_Scarborough
```

Out[16]:



In [17]:

```
CLIENT_ID = '*****'  
CLIENT_SECRET = '*****'  
VERSION = '*****'
```

Explore Neighbourhood in Scarborough

In [18]:

```
LIMIT = 100  
radius = 500  
url = 'https://api.foursquare.com/v2/venues/explore?client_id={}&client_secret={}&ll={}  
,{}&v={}&radius={}&limit={}'.format(CLIENT_ID, CLIENT_SECRET, Scarborough_latitude, Sca  
rborough_longitude, VERSION, radius, LIMIT)
```

In [19]:

```
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)
```

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

In [20]:

```
Scarborough_venues = getNearbyVenues(names=Scarborough_data['Neighbourhood'],
                                     latitudes=Scarborough_data['Latitude'],
                                     longitudes=Scarborough_data['Longitude']
                                     )
```

Rouge, Malvern
Highland Creek, Rouge Hill, Port Union
Guildwood, Morningside, West Hill
Woburn
Cedarbrae
Scarborough Village
East Birchmount Park, Ionview, Kennedy Park
Clairlea, Golden Mile, Oakridge
Cliffcrest, Cliffside, Scarborough Village West
Birch Cliff, Cliffside West
Dorset Park, Scarborough Town Centre, Wexford Heights
Maryvale, Wexford
Agincourt
Clarks Corners, Sullivan, Tam O'Shanter
Agincourt North, L'Amoreaux East, Milliken, Steeles East
L'Amoreaux West, Steeles West
Upper Rouge

In [21]:

```
Scarborough_venues.head()
```

Out[21]:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Rouge, Malvern	43.806686	-79.194353	Wendy's	43.807448	-79.199056	Fast Food Restaurant
1	Highland Creek, Rouge Hill, Port Union	43.784535	-79.160497	Royal Canadian Legion	43.782533	-79.163085	Event Space
2	Highland Creek, Rouge Hill, Port Union	43.784535	-79.160497	Affordable Toronto Movers	43.787919	-79.162977	Moving Target
3	Guildwood, Morningside, West Hill	43.763573	-79.188711	Swiss Chalet Rotisserie & Grill	43.767697	-79.189914	Pizzeria
4	Guildwood, Morningside, West Hill	43.763573	-79.188711	G & G Electronics	43.765309	-79.191537	Electronics Store

Analyze each Neighbourhood

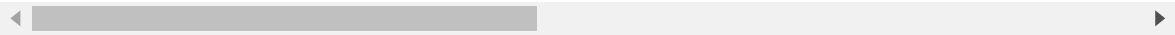
In [22]:

```
Scarborough_onehot = pd.get_dummies(Scarborough_venues[['Venue Category']], prefix="",
prefix_sep="")
Scarborough_onehot['Neighborhood'] = Scarborough_venues['Neighborhood']
fixed_columns = [Scarborough_onehot.columns[-1]] + list(Scarborough_onehot.columns[:-1
])
Scarborough_onehot = Scarborough_onehot[fixed_columns]
Scarborough_onehot.head()
```

Out[22]:

	Neighborhood	Accessories Store	American Restaurant	Athletics & Sports	Auto Garage	Bakery	Bank	Bar	Breakfast Spot
0	Rouge, Malvern	0	0	0	0	0	0	0	0
1	Highland Creek, Rouge Hill, Port Union	0	0	0	0	0	0	1	0
2	Highland Creek, Rouge Hill, Port Union	0	0	0	0	0	0	0	0
3	Guildwood, Morningside, West Hill	0	0	0	0	0	0	0	0
4	Guildwood, Morningside, West Hill	0	0	0	0	0	0	0	0

5 rows × 55 columns



In [23]:

```
Scarborough_grouped = Scarborough_onehot.groupby('Neighborhood').mean().reset_index()  
Scarborough_grouped
```

Out[23]:

	Neighborhood	Accessories Store	American Restaurant	Athletics & Sports	Auto Garage	Bakery	Bank	Bar	Br
0	Agincourt	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	0
1	Agincourt North, L'Amoreaux East, Milliken, St...	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	0
2	Birch Cliff, Cliffside West	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	0
3	Cedarbrae	0.000000	0.000000	0.142857	0.000000	0.142857	0.142857	0.0	0
4	Clairlea, Golden Mile, Oakridge	0.000000	0.000000	0.000000	0.000000	0.200000	0.000000	0.0	0
5	Clarks Corners, Sullivan, Tam O'Shanter	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	0
6	Cliffcrest, Cliffside, Scarborough Village West	0.000000	0.333333	0.000000	0.000000	0.000000	0.000000	0.0	0
7	Dorset Park, Scarborough Town Centre, Wexford ...	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	0
8	East Birchmount Park, Ionview, Kennedy Park	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	0
9	Guildwood, Morningside, West Hill	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	0
10	Highland Creek, Rouge Hill, Port Union	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.5	0
11	L'Amoreaux West, Steeles West	0.000000	0.076923	0.000000	0.000000	0.000000	0.000000	0.0	0
12	Maryvale, Wexford	0.166667	0.000000	0.000000	0.166667	0.166667	0.000000	0.0	0
13	Rouge, Malvern	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	0
14	Scarborough Village	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	0
15	Woburn	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	0

16 rows × 55 columns

In [24]:

```
import numpy as np
```

In [25]:

```
#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]

#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'] = Scarborough_grouped['Neighborhood']

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

neighborhoods_venues_sorted
```

Out[25]:

	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
0	Agincourt	Skating Rink	Breakfast Spot	Lounge	Clothing Store	Vietnamese Restaurant	Chinese Restaurant
1	Agincourt North, L'Amoreaux East, Milliken, St...	Park	Playground	Vietnamese Restaurant	Caribbean Restaurant	General Entertainment	Furniture Home Store
2	Birch Cliff, Cliffside West	Café	General Entertainment	Skating Rink	College Stadium	Caribbean Restaurant	Furniture Home Store
3	Cedarbrae	Thai Restaurant	Athletics & Sports	Bakery	Bank	Fried Chicken Joint	Chinese Restaurant
4	Clairlea, Golden Mile, Oakridge	Bus Line	Bakery	Metro Station	Soccer Field	Intersection	Fast Food Restaurant
5	Clarks Corners, Sullivan, Tam O'Shanter	Pizza Place	Noodle House	Thai Restaurant	Fried Chicken Joint	Fast Food Restaurant	Italian Restaurant
6	Cliffcrest, Cliffside, Scarborough Village West	American Restaurant	Skating Rink	Motel	Vietnamese Restaurant	Chinese Restaurant	General Entertainment
7	Dorset Park, Scarborough Town Centre, Wexford ...	Indian Restaurant	Pet Store	Chinese Restaurant	Furniture / Home Store	Latin American Restaurant	Vietnamese Restaurant
8	East Birchmount Park, Ionview, Kennedy Park	Discount Store	Department Store	Bus Station	Coffee Shop	Train Station	Chinese Restaurant
9	Guildwood, Morningside, West Hill	Electronics Store	Rental Car Location	Breakfast Spot	Pizza Place	Medical Center	Mexican Restaurant
10	Highland Creek, Rouge Hill, Port Union	Bar	Moving Target	Vietnamese Restaurant	Chinese Restaurant	General Entertainment	Furniture Home Store
11	L'Amoreaux West, Steeles West	Chinese Restaurant	Fast Food Restaurant	Japanese Restaurant	Grocery Store	Burger Joint	Pharmacy
12	Maryvale, Wexford	Accessories Store	Auto Garage	Bakery	Shopping Mall	Sandwich Place	Breakfast Spot
13	Rouge, Malvern	Fast Food Restaurant	Vietnamese Restaurant	Chinese Restaurant	Grocery Store	General Entertainment	Furniture Home Store
14	Scarborough Village	Spa	Playground	Vietnamese Restaurant	Caribbean Restaurant	General Entertainment	Furniture Home Store
15	Woburn	Coffee Shop	Insurance Office	Korean Restaurant	Vietnamese Restaurant	Hakka Restaurant	General Entertainment

Cluster Neighbourhoods

In [26]:

```
from sklearn.cluster import KMeans
```

In [27]:

```
# set number of clusters
kclusters = 4
Scarborough_grouped_clustering = Scarborough_grouped.drop('Neighborhood', 1)

# run k-means clustering
kmeans = KMeans(n_clusters=kclusters, random_state=0).fit(Scarborough_grouped_clustering)
kmeans.labels_ = np.append(kmeans.labels_,1)
kmeans.labels_
```

Out[27]:

array([3, 0, 2, 3, 3, 3, 2, 3, 3, 3, 3, 3, 3, 1, 0, 3, 1])

In [28]:

```
Scarborough_merged = Scarborough_data
Scarborough_merged['Cluster Labels'] = kmeans.labels_
Scarborough_merged = Scarborough_merged.join(neighborhoods_venues_sorted.set_index('Neighborhood'), on='Neighbourhood')
Scarborough_merged.head()
```

Out[28]:

	Postcode	Borough	Neighbourhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue
0	M1B	Scarborough	Rouge, Malvern	43.806686	-79.194353	3	Fast Food Restaurant	Vietnamese Restaurant
1	M1C	Scarborough	Highland Creek, Rouge Hill, Port Union	43.784535	-79.160497	0	Bar	
2	M1E	Scarborough	Guildwood, Morningside, West Hill	43.763573	-79.188711	2	Electronics Store	Recreation Center
3	M1G	Scarborough	Woburn	43.770992	-79.216917	3	Coffee Shop	Insurance Agency
4	M1H	Scarborough	Cedarbrae	43.773136	-79.239476	3	Thai Restaurant	Athletic Center

Visualize clusters on the map

In [29]:

```
import matplotlib.cm as cm
import matplotlib.colors as colors
```

In [30]:

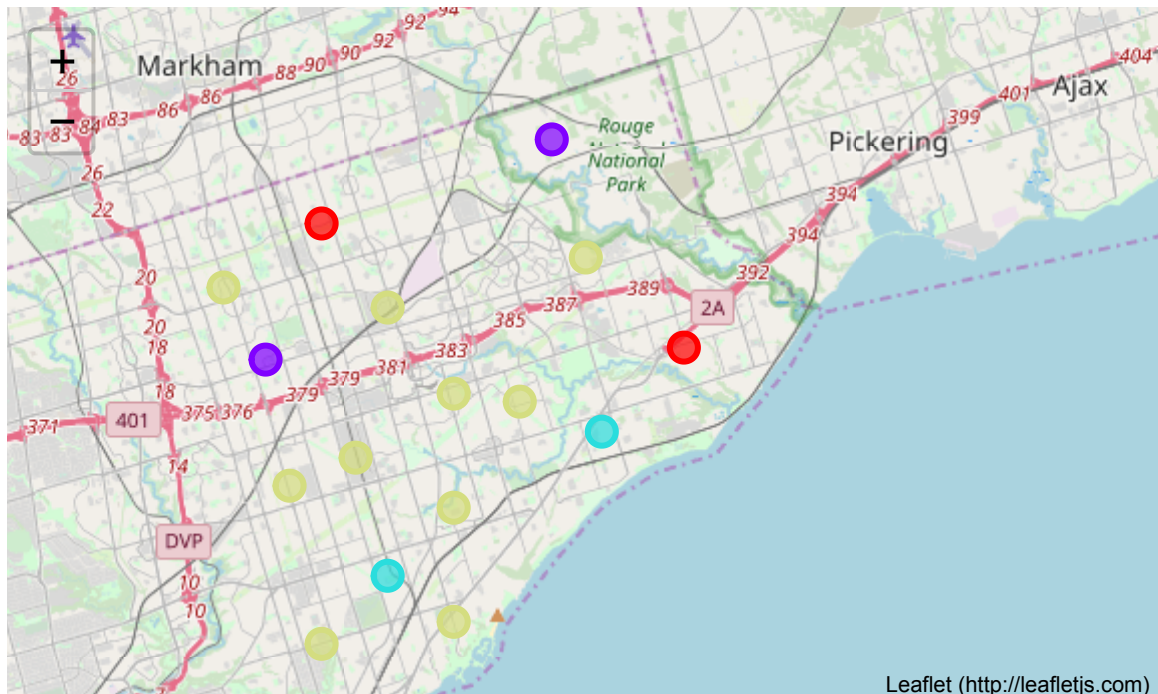
```
map_clusters = folium.Map(location=[Scarborough_latitude, Scarborough_longitude], zoom_
start=11)

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]

markers_colors = []
for lat, lon, poi, cluster in zip(Scarborough_merged['Latitude'], Scarborough_merged['L
ongitude'], Scarborough_merged['Neighbourhood'], Scarborough_merged['Cluster Labels']):
    label = folium.Popup(str(poi) + ' Cluster ' + str(cluster), parse_html=True)
    folium.CircleMarker(
        [lat, lon],
        radius=7,
        popup=label,
        color=rainbow[cluster-1],
        fill=True,
        fill_color=rainbow[cluster-1],
        fill_opacity=0.7).add_to(map_clusters)

map_clusters
```

Out[30]:



Examine the Clusters

Cluster 1

In [31]:

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

Out[31]:

	Borough	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue
1	Scarborough	0	Bar	Moving Target	Vietnamese Restaurant	Chinese Restaurant	General Entertainment	Furniture / Home Storage
14	Scarborough	0	Park	Playground	Vietnamese Restaurant	Caribbean Restaurant	General Entertainment	Furniture / Home Storage

Cluster 2

In [32]:

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

Out[32]:

	Borough	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue
13	Scarborough	1	Pizza Place	Noodle House	Thai Restaurant	Fried Chicken Joint	Fast Food Restaurant	Italian Restaurant
16	Scarborough	1	NaN	NaN	NaN	NaN	NaN	NaN

Cluster 3

In [33]:

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

Out[33]:

	Borough	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	
2	Scarborough	2	Electronics Store	Rental Car Location	Breakfast Spot	Pizza Place	Medical Center	Mexican Restaurant	Vic R
6	Scarborough	2	Discount Store	Department Store	Bus Station	Coffee Shop	Train Station	Chinese Restaurant	Vic R



Cluster 4

In [34]:

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

Out[34]:

	Borough	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	
0	Scarborough	3	Fast Food Restaurant	Vietnamese Restaurant	Chinese Restaurant	Grocery Store	General Entertainment	
3	Scarborough	3	Coffee Shop	Insurance Office	Korean Restaurant	Vietnamese Restaurant	Hakka Restaurant	Er
4	Scarborough	3	Thai Restaurant	Athletics & Sports	Bakery	Bank	Fried Chicken Joint	
5	Scarborough	3	Spa	Playground	Vietnamese Restaurant	Caribbean Restaurant	General Entertainment	
7	Scarborough	3	Bus Line	Bakery	Metro Station	Soccer Field	Intersection	
8	Scarborough	3	American Restaurant	Skating Rink	Motel	Vietnamese Restaurant	Chinese Restaurant	Er
9	Scarborough	3	Café	General Entertainment	Skating Rink	College Stadium	Caribbean Restaurant	
10	Scarborough	3	Indian Restaurant	Pet Store	Chinese Restaurant	Furniture / Home Store	Latin American Restaurant	
11	Scarborough	3	Accessories Store	Auto Garage	Bakery	Shopping Mall	Sandwich Place	
12	Scarborough	3	Skating Rink	Breakfast Spot	Lounge	Clothing Store	Vietnamese Restaurant	
15	Scarborough	3	Chinese Restaurant	Fast Food Restaurant	Japanese Restaurant	Grocery Store	Burger Joint	

In []: