

# **Coursera Capstone Project**

## **Battle of the Neighborhoods**

### Table of Contents

Business Problem .....	2
Data Sources .....	2
Methodology .....	3
Step 1: Data Cleansing & Preparation .....	3
Step 2: North York Neighborhoods map visualization .....	5
Step 3: Using Four Square API to get Venues .....	6
Step 4: Cluster, Segment & Analyze Neighborhoods .....	8
Results .....	14
Cluster 1 .....	14
Cluster 2.....	15
Cluster 3.....	16
Cluster 4.....	16
Cluster 5.....	17
Cluster 6.....	17
Discussion.....	18
Recommendations .....	18
Conclusion.....	18

## Business Problem

The scope of this project is to find the most suitable location for opening a new **Cafe Equipment Cleaning & Maintenance company** in Toronto North York. Key factors will be the number of cafes in the targeted area and the possibility of combining office location with place of residence.

These factors will be decisive for our analysis because,

1. An office location with many potential customers close by will raise revenue and minimize operational & marketing costs
2. It will be beneficial for the company owner to live close by to his office, as long as one of two important personal parameters are met:
  - a. The place of residence should be near parks
  - b. There should be a variety of restaurants close by

**Target Audience:** Any individual who wishes to open a Cafe Equipment Cleaning & Maintenance company in North York Toronto and at the same time live in the same area.

Our main task would be to analyze relevant North York data in order to cluster and segment neighborhoods based on the most common venues and yet as close to parks and restaurants as possible.

## Data Sources

The necessary data and the way which will be used are listed below:

- ✚ We will acquire all Toronto neighborhoods from Wikipedia with relevant Postal Code – Borough – Neighborhood  
[https://en.wikipedia.org/wiki/List\\_of\\_postal\\_codes\\_of\\_Canada:\\_M](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M)
- ✚ We will Use geopy library to get the latitude and longitude values of Toronto Postal Codes & combine with the Geographical coordinates of the neighborhoods with the respective Postal Codes  
[https://cocl.us/Geospatial\\_data](https://cocl.us/Geospatial_data)

- ✚ We will use Four Square API to explore North York number and type of venues and cluster the neighborhoods according to our findings. we will visualize our findings by color code in relevant maps and come up to final conclusion.

## Methodology

### Step 1: Data Cleansing & Preparation

Our first task is to drop any Boroughs with a 'Not Assigned' value & group our dataframe based on Borough.

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

Based on Postal Code and the information from the Geocoder package we proceed to assign the Latitude and Longitude for every Neighborhood.

	Borough	Neighbourhood	Latitude	Longitude
0	Scarborough	Malvern, Rouge	43.806686	-79.194353
1	Scarborough	Rouge Hill, Port Union, Highland Creek	43.784535	-79.160497
2	Scarborough	Guildwood, Morningside, West Hill	43.763573	-79.188711
3	Scarborough	Woburn	43.770992	-79.216917
4	Scarborough	Cedarbrae	43.773136	-79.239476

Since we are interested only in North York Borough, we proceed to drop from our dataframe or other non-relevant information.

	Borough	Neighbourhood	Latitude	Longitude
0	North York	Hillcrest Village	43.803762	-79.363452
1	North York	Fairview, Henry Farm, Oriole	43.778517	-79.346556
2	North York	Bayview Village	43.786947	-79.385975
3	North York	York Mills, Silver Hills	43.757490	-79.374714
4	North York	Willowdale, Newtonbrook	43.789053	-79.408493
5	North York	Willowdale, Willowdale East	43.770120	-79.408493
6	North York	York Mills West	43.752758	-79.400049
7	North York	Willowdale, Willowdale West	43.782736	-79.442259
8	North York	Parkwoods	43.753259	-79.329656
9	North York	Don Mills	43.745906	-79.352188
10	North York	Don Mills	43.725900	-79.340923
11	North York	Bathurst Manor, Wilson Heights, Downsview North	43.754328	-79.442259
12	North York	Northwood Park, York University	43.767980	-79.487262
13	North York	Downsview	43.737473	-79.464763
14	North York	Downsview	43.739015	-79.506944
15	North York	Downsview	43.728496	-79.495697
16	North York	Downsview	43.761631	-79.520999
17	North York	Victoria Village	43.725882	-79.315572
18	North York	Bedford Park, Lawrence Manor East	43.733283	-79.419750
19	North York	Lawrence Manor, Lawrence Heights	43.718518	-79.464763
20	North York	Glencairn	43.709577	-79.445073
21	North York	North Park, Maple Leaf Park, Upwood Park	43.713756	-79.490074
22	North York	Humber Summit	43.756303	-79.565963
23	North York	Humberlea, Emery	43.724766	-79.532242

```
# we proceed to scrap 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 of postal codes & transform data in a new pandas dataframe
df = pd.read_html('https://en.wikipedia.org/w/index.php?title=List_of_postal_codes_of_Canada:_M')[0]

# we drop any Boroughs with value 'Not assigned'
df = df[df.Borough!='Not assigned']
df = df.groupby(['Postal Code', 'Borough'], as_index=False).agg(lambda x: ','.join(x))
df
```

```
df_geo = pd.read_csv("https://cocl.us/Geospatial_data")
df = pd.merge(df, df_geo, left_on='Postal Code', right_on='Postal Code').drop(['Postal Code'], axis=1)
df
```

```
# Segment and cluster only the neighborhoods in North York Borough
york_data = df[df['Borough']=='North York'].reset_index(drop=True)
york_data
```

## Step 2: North York Neighborhoods map visualization

We proceed to visualize our North York Neighborhoods in a map using Folium library and color-code each Neighborhood.

```
address = 'North York'
geolocator = Nominatim(user_agent='to_explorer')
location = geolocator.geocode(address)
latitude = location.latitude
longitude = location.longitude

print("The geographical coordinates of North York are {}, {}.".format(latitude, longitude))
```

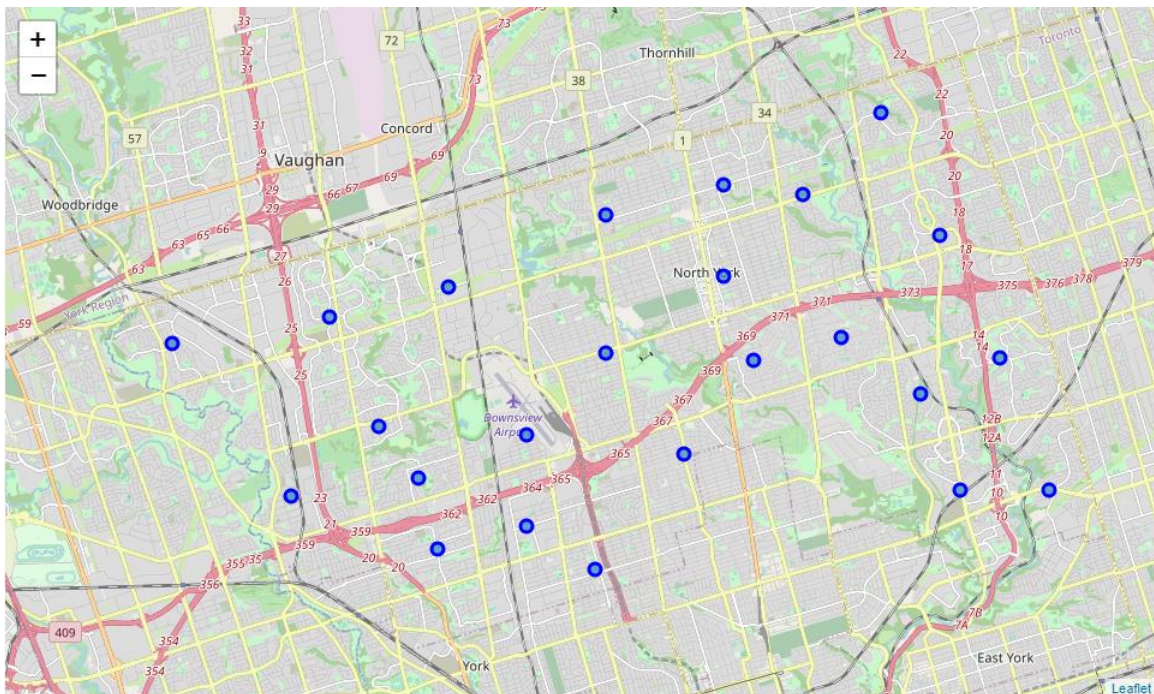
The geographical coordinates of North York are 43.7543263, -79.44911696639593.

```
# create map of North York using Latitude and Longitude values
map_north_york = folium.Map(location=[latitude, longitude], zoom_start=12)

# add markers to map
for lat, lng, borough, neighbourhood in zip(york_data['Latitude'], york_data['Longitude'], york_data['Borough'], york_data['Neighbourhood']):
    label = '{}', {}'.format(neighbourhood, 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_north_york)

map_north_york
```

## North York Neighborhoods Map in Toronto



### Step 3: Using Four Square API to get Venues

We proceed to use the Foursquare API to get a list of the top 100 venues that ate in North York within a radius of 1000 meters. The resulting data will be a first indication about the most common venue types in North York neighborhoods.

```
# Get the top 100 venues that are in North York within a radius of 1000 meters
# we define necessary functions

def getNearbyVenues(names, latitudes, longitudes, radius=1000):

    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 = ['Neighbourhood',
                            'Neighbourhood Latitude',
                            'Neighbourhood Longitude',
                            'Venue',
                            'Venue Latitude',
                            'Venue Longitude',
                            'Venue Category']

    return(nearby_venues)
```

```
# run the function above to each neighborhood and create a new dataframe called north_york_venues
north_york_venues = getNearbyVenues(names = york_data['Neighbourhood'],
                                     latitudes = york_data['Latitude'],
                                     longitudes = york_data['Longitude'])

north_york_venues
```

```
Hillcrest Village
Fairview, Henry Farm, Oriole
Bayview Village
York Mills, Silver Hills
Willowdale, Newtonbrook
Willowdale, Willowdale East
York Mills West
Willowdale, Willowdale West
Parkwoods
Don Mills
Don Mills
Bathurst Manor, Wilson Heights, Downsview North
Northwood Park, York University
Downsview
Downsview
Downsview
Victoria Village
Bedford Park, Lawrence Manor East
Lawrence Manor, Lawrence Heights
Glencairn
North Park, Maple Leaf Park, Upwood Park
Humber Summit
Humberlea, Emery
```

## Out combined Neighborhood – Venue data

	Neighbourhood	Neighbourhood Latitude	Neighbourhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Hillcrest Village	43.803762	-79.363452	Tastee	43.807722	-79.356798	Bakery
1	Hillcrest Village	43.803762	-79.363452	고려삼계탕 Korean Ginseng Chicken Soup & Bibimbap	43.798391	-79.369187	Korean Restaurant
2	Hillcrest Village	43.803762	-79.363452	Cummer Park	43.799564	-79.371175	Park
3	Hillcrest Village	43.803762	-79.363452	Galati	43.797831	-79.369410	Grocery Store
4	Hillcrest Village	43.803762	-79.363452	Pizza Pizza	43.799079	-79.369449	Pizza Place

We get 628 venues in North York Area

```
# check size of resulting dataframe
north_york_venues.shape
```

```
(628, 7)
```

We check how many venues were returned for each Neighborhood

```
# Let's check how many venues were returned for each neighborhood
north_york_venues.groupby('Neighbourhood').count()
```

Neighbourhood	
Bathurst Manor, Wilson Heights, Downsview North	29
Bayview Village	14
Bedford Park, Lawrence Manor East	39
Don Mills	75
Downsview	67
Fairview, Henry Farm, Oriole	44
Glencairn	35
Hillcrest Village	21
Humber Summit	10
Humberlea, Emery	8
Lawrence Manor, Lawrence Heights	44
North Park, Maple Leaf Park, Upwood Park	12
Northwood Park, York University	24
Parkwoods	28
Victoria Village	14
Willowdale, Newtonbrook	29
Willowdale, Willowdale East	100
Willowdale, Willowdale West	11
York Mills West	20
York Mills, Silver Hills	4



Finally, we can see that there are 152 unique venue categories.

```
# Let's find out how many unique categories can be curated from all the returned venues
print('There are {} unique categories.'.format(len(north_york_venues['Venue Category'].unique())))

There are 152 unique categories.
```

## Step 4: Cluster, Segment & Analyze Neighborhoods

Our first task would be to use 'One hot encoding' in order to transform categorical data to numeric data & accordingly see how many specific venue types exist in every Neighborhood.

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

# add neighbourhood column back to dataframe
york_onehot['Neighbourhood'] = north_york_venues['Neighbourhood']

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

york_onehot.head()

# Let's group rows by neighborhood and by taking the mean of the frequency of occurrence of each category
york_grouped = york_onehot.groupby('Neighbourhood').mean().reset_index()
york_grouped.head()
```

	Neighbourhood	Accessories Store	Airport	American Restaurant	Art Gallery	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	Auto Workshop	Automotive Shop	BBQ Joint	Baby Store	Bagel Shop	Bakery
0	Bathurst Manor, Wilson Heights, Downsview North	0.0	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	0.0	0.0	0.0	0.000000	0.000000	0.000000
1	Bayview Village	0.0	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	0.0	0.0	0.0	0.000000	0.000000	0.000000
2	Bedford Park, Lawrence Manor East	0.0	0.000000	0.025641	0.000000	0.0	0.000000	0.000000	0.0	0.0	0.0	0.025641	0.025641	0.025641
3	Don Mills	0.0	0.000000	0.013333	0.013333	0.0	0.026667	0.000000	0.0	0.0	0.0	0.000000	0.000000	0.000000
4	Downsview	0.0	0.014925	0.014925	0.000000	0.0	0.000000	0.029851	0.0	0.0	0.0	0.000000	0.000000	0.014925

```
# check dataframe size
york_grouped.shape
```

```
(20, 153)
```

We proceed to print each Neighborhood along with top 5 most common venues

```
# Let's print each neighborhood along with the top 5 most common venues
num_top_venues = 5

for hood in york_grouped['Neighbourhood']:
    print("----"+hood+"----")
    temp = york_grouped[york_grouped['Neighbourhood'] == 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')
```



## ----Bathurst Manor, Wilson Heights, Downsview North----

	venue	freq
0	Coffee Shop	0.07
1	Bank	0.07
2	Mobile Phone Shop	0.03
3	Ski Chalet	0.03
4	Ski Area	0.03

## ----Bayview Village----

	venue	freq
0	Japanese Restaurant	0.14
1	Gas Station	0.14
2	Bank	0.14
3	Grocery Store	0.14
4	Trail	0.07

## ----Bedford Park, Lawrence Manor East----

	venue	freq
0	Italian Restaurant	0.08
1	Coffee Shop	0.08
2	Bank	0.05
3	Sandwich Place	0.05
4	Butcher	0.03

## ----Don Mills----

	venue	freq
0	Coffee Shop	0.08
1	Restaurant	0.08
2	Japanese Restaurant	0.07
3	Gym	0.05
4	Supermarket	0.04

## ----Downsview----

	venue	freq
0	Vietnamese Restaurant	0.07
1	Coffee Shop	0.07
2	Grocery Store	0.06
3	Hotel	0.06
4	Pizza Place	0.04

## ----Fairview, Henry Farm, Oriole----

	venue	freq
0	Coffee Shop	0.11
1	Clothing Store	0.11
2	Restaurant	0.05
3	Juice Bar	0.05
4	Japanese Restaurant	0.05

## ----Glencairn----

	venue	freq
0	Grocery Store	0.11
1	Fast Food Restaurant	0.09
2	Italian Restaurant	0.06
3	Pizza Place	0.06
4	Gas Station	0.06

## ----Hillcrest Village----

	venue	freq
0	Park	0.10
1	Pharmacy	0.10
2	Coffee Shop	0.10
3	Ice Cream Shop	0.05
4	Shopping Mall	0.05

## ----Humber Summit----

	venue	freq
0	Electronics Store	0.2
1	Italian Restaurant	0.1
2	Pharmacy	0.1
3	Medical Center	0.1
4	Bank	0.1

## ----Humberlea, Emery----

	venue	freq
0	Convenience Store	0.12
1	Gas Station	0.12
2	Discount Store	0.12
3	Auto Workshop	0.12
4	Storage Facility	0.12

## ----Lawrence Manor, Lawrence Heights----

	venue	freq
0	Clothing Store	0.11
1	Fast Food Restaurant	0.07
2	Restaurant	0.07
3	Coffee Shop	0.07
4	Vietnamese Restaurant	0.05

## ----North Park, Maple Leaf Park, Upwood Park----

	venue	freq
0	Coffee Shop	0.17
1	Pizza Place	0.08
2	Intersection	0.08
3	Dim Sum Restaurant	0.08
4	Convenience Store	0.08

## ----Northwood Park, York University----

	venue	freq
0	Coffee Shop	0.12
1	Furniture / Home Store	0.12
2	Restaurant	0.08
3	Pizza Place	0.08
4	Bar	0.04

## ----Parkwoods----

	venue	freq
0	Park	0.11
1	Convenience Store	0.07
2	Bus Stop	0.07
3	Shopping Mall	0.07
4	Pharmacy	0.07

## ----Victoria Village----

	venue	freq
0	Coffee Shop	0.14
1	Boxing Gym	0.07
2	Intersection	0.07
3	French Restaurant	0.07
4	Playground	0.07

## ----Willowdale, Newtonbrook----

	venue	freq
0	Korean Restaurant	0.14
1	Café	0.10
2	Middle Eastern Restaurant	0.07
3	Diner	0.07
4	Pizza Place	0.07

## ----Willowdale, Willowdale East----

	venue	freq
0	Coffee Shop	0.07
1	Korean Restaurant	0.06
2	Bubble Tea Shop	0.06
3	Ramen Restaurant	0.06
4	Japanese Restaurant	0.05

## ----Willowdale, Willowdale West----

	venue	freq
0	Pharmacy	0.18
1	Bus Line	0.09
2	Convenience Store	0.09
3	Park	0.09
4	Coffee Shop	0.09

## ----York Mills West----

	venue	freq
0	Park	0.15
1	Restaurant	0.10
2	Coffee Shop	0.10
3	Pet Store	0.05
4	Bubble Tea Shop	0.05

## ----York Mills, Silver Hills----

	venue	freq
0	Park	0.75
1	Pool	0.25
2	Accessories Store	0.00
3	Miscellaneous Shop	0.00
4	Mobile Phone Shop	0.00

```

# Let's put that into a pandas dataframe
# we create 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]

# we 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 = ['Neighbourhood']
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
neighbourhoods_venues_sorted = pd.DataFrame(columns=columns)
neighbourhoods_venues_sorted['Neighbourhood'] = york_grouped['Neighbourhood']

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

neighbourhoods_venues_sorted

```

Our final data would be a sorted list of Neighborhoods with the 10 most common venues as seen below

	Neighbourhood	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	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Bathurst Manor, Wilson Heights, Downsview North	Bank	Coffee Shop	Restaurant	Community Center	Shopping Mall	Dog Run	Sandwich Place	Bridal Shop	Diner	Chinese Restaurant
1	Bayview Village	Gas Station	Bank	Grocery Store	Japanese Restaurant	Trail	Café	Restaurant	Chinese Restaurant	Skating Rink	Park
2	Bedford Park, Lawrence Manor East	Coffee Shop	Italian Restaurant	Sandwich Place	Bank	Bridal Shop	Restaurant	Butcher	Pub	Café	Pharmacy
3	Don Mills	Coffee Shop	Restaurant	Japanese Restaurant	Gym	Burger Joint	Supermarket	Mobile Phone Shop	Beer Store	Asian Restaurant	Pizza Place
4	Downsview	Coffee Shop	Vietnamese Restaurant	Hotel	Grocery Store	Gas Station	Pharmacy	Pizza Place	Intersection	Turkish Restaurant	Fast Food Restaurant
5	Fairview, Henry Farm, Oriole	Coffee Shop	Clothing Store	Sandwich Place	Japanese Restaurant	Bank	Bakery	Restaurant	Juice Bar	Grocery Store	Movie Theater
6	Glencairn	Grocery Store	Fast Food Restaurant	Coffee Shop	Gas Station	Pizza Place	Italian Restaurant	Convenience Store	Pub	Photography Lab	Pharmacy
7	Hillcrest Village	Pharmacy	Park	Coffee Shop	Convenience Store	Chinese Restaurant	Sandwich Place	Restaurant	Residential Building (Apartment / Condo)	Korean Restaurant	Recreation Center
8	Humber Summit	Electronics Store	Bank	Pizza Place	Pharmacy	Park	Italian Restaurant	Shopping Mall	Medical Center	Bakery	Yoga Studio
9	Humberlea, Emery	Discount Store	Convenience Store	Park	Auto Workshop	Golf Course	Storage Facility	Gas Station	Bakery	Yoga Studio	Event Space

Now that we have the 10 most common venue types we will cluster & visualize our Neighborhoods using the K-Means Machine Learning Algorithm. After some testing we can see that the best number of clusters is 6.

```
# set number of clusters
kclusters = 6

york_grouped_clustering = york_grouped.drop('Neighbourhood', 1)

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

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

```
array([1, 0, 2, 2, 1, 2, 1, 1, 5, 4])
```

```
# Let's create a new dataframe that includes the cluster as well as the top 10 venues for each neighborhood.
# add clustering labels
neighbourhoods_venues_sorted.insert(0, 'Cluster Labels', kmeans.labels_)

york_merged = df

# merge toronto_grouped with df to add Latitude/Longitude for each neighbourhood
york_merged = york_merged.join(neighbourhoods_venues_sorted.set_index('Neighbourhood'), on='Neighbourhood')
york_merged = york_merged.dropna()
york_merged = york_merged.astype({"Cluster Labels": int})
york_merged.head()
```

	Borough	Neighbourhood	Latitude	Longitude	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	8th Most Common Venue
17	North York	Hillcrest Village	43.803762	-79.363452	1	Pharmacy	Park	Coffee Shop	Convenience Store	Chinese Restaurant	Sandwich Place	Restaurant	Residential Building (Apartment / Condo)
18	North York	Fairview, Henry Farm, Oriole	43.778517	-79.346556	2	Coffee Shop	Clothing Store	Sandwich Place	Japanese Restaurant	Bank	Bakery	Restaurant	Juice Bar
19	North York	Bayview Village	43.786947	-79.385975	0	Gas Station	Bank	Grocery Store	Japanese Restaurant	Trail	Café	Restaurant	Chinese Restaurant
20	North York	York Mills, Silver Hills	43.757490	-79.374714	3	Park	Pool	Diner	Falafel Restaurant	Event Space	Electronics Store	Eastern European Restaurant	Dumpling Restaurant
21	North York	Willowdale, Newtonbrook	43.789053	-79.408493	1	Korean Restaurant	Café	Pizza Place	Middle Eastern Restaurant	Park	Diner	Coffee Shop	Supermarket

```
# Finally, Let's visualize the resulting clusters
# create map
import matplotlib.cm as cm
import matplotlib.colors as colors

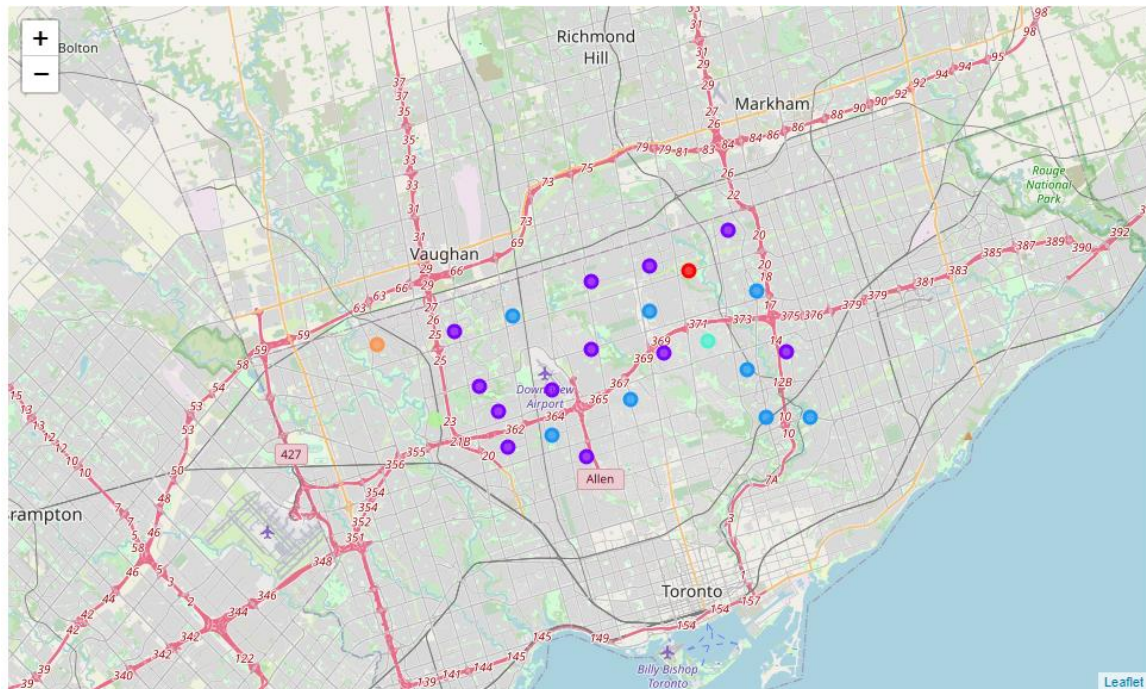
map_clusters = folium.Map(location=[latitude, longitude], zoom_start=11)

# 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(york_merged['Latitude'], york_merged['Longitude'], york_merged['Neighbourhood'], york_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
```

We visualize our clustered results in the following map



## Results

### Cluster 1

```
# Cluster 1
york_merged.loc[york_merged['Cluster Labels'] == 0]
```

	Borough	Neighbourhood	Latitude	Longitude	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	8th Most Common Venue	9th Most Common Venue
19	North York	Bayview Village	43.786947	-79.385975	0	Gas Station	Bank	Grocery Store	Japanese Restaurant	Trail	Café	Restaurant	Chinese Restaurant	Skate Ri

Cluster 1 contains only Bayview Village neighborhood. There are a lot of Gas stations, banks & grocery stores in the area. Cafes are the 6<sup>th</sup> most common venue. Based on our desired parameters, Bayview Village is not a desired area to open a new office for our company or as a place of residence.

## Cluster 2

```
# Cluster 2
york_merged.loc[york_merged['Cluster Labels'] == 1]
```

	Borough	Neighbourhood	Latitude	Longitude	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	8th Most Common Venue
17	North York	Hillcrest Village	43.803762	-79.363452	1	Pharmacy	Park	Coffee Shop	Convenience Store	Chinese Restaurant	Sandwich Place	Restaurant	Restaurant (Apartment)
21	North York	Willowdale, Newtonbrook	43.789053	-79.408493	1	Korean Restaurant	Café	Pizza Place	Middle Eastern Restaurant	Park	Diner	Coffee Shop	Supermarket
23	North York	York Mills West	43.752758	-79.400049	1	Park	Restaurant	Coffee Shop	Playground	Gas Station	Dog Run	Business Service	
24	North York	Willowdale, Willowdale West	43.782736	-79.442259	1	Pharmacy	Butcher	Convenience Store	Pizza Place	Eastern European Restaurant	Coffee Shop	Grocery Store	
25	North York	Parkwoods	43.753259	-79.329656	1	Park	Convenience Store	Pharmacy	Shopping Mall	Bus Stop	Discount Store	Shop & Service	Fast Food Restaurant
28	North York	Bathurst Manor, Wilson Heights, Downsview North	43.754328	-79.442259	1	Bank	Coffee Shop	Restaurant	Community Center	Shopping Mall	Dog Run	Sandwich Place	Bridge
30	North York	Downsview	43.737473	-79.464763	1	Coffee Shop	Vietnamese Restaurant	Hotel	Grocery Store	Gas Station	Pharmacy	Pizza Place	Interpretation Center
31	North York	Downsview	43.739015	-79.506944	1	Coffee Shop	Vietnamese Restaurant	Hotel	Grocery Store	Gas Station	Pharmacy	Pizza Place	Interpretation Center
32	North York	Downsview	43.728496	-79.495697	1	Coffee Shop	Vietnamese Restaurant	Hotel	Grocery Store	Gas Station	Pharmacy	Pizza Place	Interpretation Center
33	North York	Downsview	43.761631	-79.520999	1	Coffee Shop	Vietnamese Restaurant	Hotel	Grocery Store	Gas Station	Pharmacy	Pizza Place	Interpretation Center
72	North York	Glencairn	43.709577	-79.445073	1	Grocery Store	Fast Food Restaurant	Coffee Shop	Gas Station	Pizza Place	Italian Restaurant	Convenience Store	
79	North York	North Park, Maple Leaf Park, Upwood Park	43.713756	-79.490074	1	Coffee Shop	Convenience Store	Athletics & Sports	Bakery	Gas Station	Chinese Restaurant	Park	Dog Run

Cluster 2 contains a bigger number of neighborhoods. We can see that we have either parks or cafes in top common venue types and there are a lot of restaurants. Based on our desired parameters, these neighborhoods are worth consideration as a desired area to open a new office for our company and as a place of residence.



### Cluster 3

```
# Cluster 3
york_merged.loc[york_merged['Cluster Labels'] == 2]
```

	Borough	Neighbourhood	Latitude	Longitude	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	8th Most Common Venue
18	North York	Fairview, Henry Farm, Oriole	43.778517	-79.346556	2	Coffee Shop	Clothing Store	Sandwich Place	Japanese Restaurant	Bank	Bakery	Restaurant	Juice Bar
22	North York	Willowdale, Willowdale East	43.770120	-79.408493	2	Coffee Shop	Ramen Restaurant	Korean Restaurant	Bubble Tea Shop	Fast Food Restaurant	Pizza Place	Japanese Restaurant	Sandwich Place
26	North York	Don Mills	43.745906	-79.352188	2	Coffee Shop	Restaurant	Japanese Restaurant	Gym	Burger Joint	Supermarket	Mobile Phone Shop	Beer Store
27	North York	Don Mills	43.725900	-79.340923	2	Coffee Shop	Restaurant	Japanese Restaurant	Gym	Burger Joint	Supermarket	Mobile Phone Shop	Beer Store
29	North York	Northwood Park, York University	43.767980	-79.487262	2	Coffee Shop	Furniture / Home Store	Pizza Place	Restaurant	Bank	Sandwich Place	Metro Station	Fast Food Restaurant
34	North York	Victoria Village	43.725882	-79.315572	2	Coffee Shop	Portuguese Restaurant	French Restaurant	Hockey Arena	Men's Store	Boxing Gym	Playground	Sporting Goods Shop
62	North York	Bedford Park, Lawrence Manor East	43.733283	-79.419750	2	Coffee Shop	Italian Restaurant	Sandwich Place	Bank	Bridal Shop	Restaurant	Butcher	Pub
71	North York	Lawrence Manor, Lawrence Heights	43.718518	-79.464763	2	Clothing Store	Restaurant	Coffee Shop	Fast Food Restaurant	Dessert Shop	Vietnamese Restaurant	Furniture / Home Store	Fried Chicken Joint

Cluster 3 contains also a big number of neighborhoods. We can see that Coffee shops are the most common venue by far. There are a lot of restaurants but not many parks. Based on our desired parameters, these neighborhoods are worth consideration as a desired area to open a new office for our company and as a place of residence.

### Cluster 4

```
# Cluster 4
york_merged.loc[york_merged['Cluster Labels'] == 3]
```

	Borough	Neighbourhood	Latitude	Longitude	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	8th Most Common Venue	9th Most Common Venue
20	North York	York Mills, Silver Hills	43.75749	-79.374714	3	Park	Pool	Diner	Falafel Restaurant	Event Space	Electronics Store	Eastern European Restaurant	Dumpling Restaurant	Dog R

Cluster 4 contains only 2 neighborhoods (York Mills & Silver Hills). We can see that Parks are the most common venue. There are some restaurants nearby but not Coffee shops. Based on our desired parameters, we cannot consider these neighborhoods for opening a new office, but we can consider them as a place of residence.

## Cluster 5

```
# Cluster 5
york_merged.loc[york_merged['Cluster Labels'] == 4]
```

	Borough	Neighbourhood	Latitude	Longitude	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	8th Most Common Venue	9th Most Common Venue
97	North York	Humberlea, Emery	43.724766	-79.532242	4	Discount Store	Convenience Store	Park	Auto Workshop	Golf Course	Storage Facility	Gas Station	Bakery	Yo Stuc

Cluster 5 contains only Humberlea, Emery neighborhood. We can see that Discount Stores and Convenience Stores are the most common venue. There are Parks nearby but not Coffee shops. Based on our desired parameters, we cannot consider these neighborhoods for opening a new office, but we can consider them as a place of residence.

## Cluster 6

```
# Cluster 6
york_merged.loc[york_merged['Cluster Labels'] == 5]
```

	Borough	Neighbourhood	Latitude	Longitude	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	8th Most Common Venue	9th Most Common Venue
96	North York	Humber Summit	43.756303	-79.565963	5	Electronics Store	Bank	Pizza Place	Pharmacy	Park	Italian Restaurant	Shopping Mall	Medical Center	Bake

Cluster 6 contains only Humber Summit neighborhood. We can see that Electronic Stores and Banks are the most common venue. There are Parks nearby and some Restaurants but not Coffee shops. Based on our desired parameters, we cannot consider these neighborhoods for opening a new office, but we can consider them as a place of residence.

## Discussion

Based on our overall results we will drop Clusters 1, 4, 5 & 6 with Neighborhoods as potential candidates for a new office location or a place of residence. Cluster 2 and Cluster 3 Neighborhoods seem more promising.

By examining more closely our data and after more investigation we decide not to consider Downsview as a candidate especially as a residential area because of the close proximity to the airport.

It seems that our best bets are North Park, Maple Leaf Park & Upwood Park from Cluster 2 and all Neighborhoods of Cluster 3.

## Recommendations

Our efforts would be more precise if we could take under consideration:

- ✚ General competition in the area
- ✚ Coffee shop types in order to determine the possible new equipment or maintenance budget
- ✚ Import other parameters for our residential candidates such as crime rates, median income etc

## Conclusion

By using Data Analysis and Machine Learning techniques we were able to cluster our areas in a concise and meaningful way and be presented with clear results which help us in order to solve our business problem.