Coursera Capstone

IBM Applied Data Science Capstone

The Battle of Neighbourhoods- Toronto and New York City

By: Hira Mansoor

January 2021

Introduction

Toronto and New York being the financial capital of Canada and the US respectively are one of the densely populated cities in the world. Both these cities see a lot of diversity resulting from the movement of a lot of immigrants from several parts of the world for work and settlement. These are one of the most immigrant-friendly cities, still different in so many aspects, which we are going to observe in this work. The purpose of this Capstone Project is to compare the neighborhoods of the two major cities and determine how similar or dissimilar they are. We will get to see the most common venues for both.

The purpose of this whole exercise is for submission of the final capstone project for the "IBM Data Science" course on Coursera as well as to show case my data science skills in the real-world application.

Data and Problem-solving strategy

The idea is to compare Toronto and the New York region for their similarities difference in terms of the most common venues they have. I will look into the number of cuisines, Restaurants, coffee shops and other features in both cities as well as list down the 5 most common venues in both cities Neighborhood wise. The outcome of this study will help tourists and new immigrants have an overview of the common venues in both cities and chalk out the differences between both, which might further help them in their decision of travel or immigration choice.

Methodology:

- Data frame for Toronto and New York is generated after collecting the data from https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
  And https://www.health.ny.gov/statistics/cancer/registry/appendix/neighborhoods.htm

- Both cities are then explored by using Foursquare API and K-Clustering Approach
- Folium library is used to visualize geographic overview for both of the cities.

Data Exploration and Results:

We extracted the data for both the cities from Wikipedia and health of government to get the zip codes for New York city and prepared that data set that contained Postal codes/Zip codes, boroughs, and Neighborhoods.

In [3]:
```
Canada_PC =requests.get("https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M").text
BSoup=BeautifulSoup(Canada_PC, 'html')
tab = str(BSoup.table)
dfs = pd.read_html(tab)
df=dfs[0]
# Dropping the rows where Borough is 'Not assigned'
df1 = df[df.Borough != 'Not assigned']

# Combining the neighbourhoods with same Postalcode
df2 = df1.groupby(['Postal Code','Borough'], sort= False).agg(', '.join)
df2.reset_index(inplace=True)
# Replacing the name of the neighbourhoods which are 'Not assigned' with names of Borough
df2['Neighbourhood'] = np.where(df2['Neighbourhood'] == 'Not assigned',df2['Borough'], df2['Neighbourhood'])
C_df= df2.rename(columns={"Neighbourhood":"Neighborhood"})
C_df.head(12)
```

Out[3]:

	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, Humber Valley Village
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...

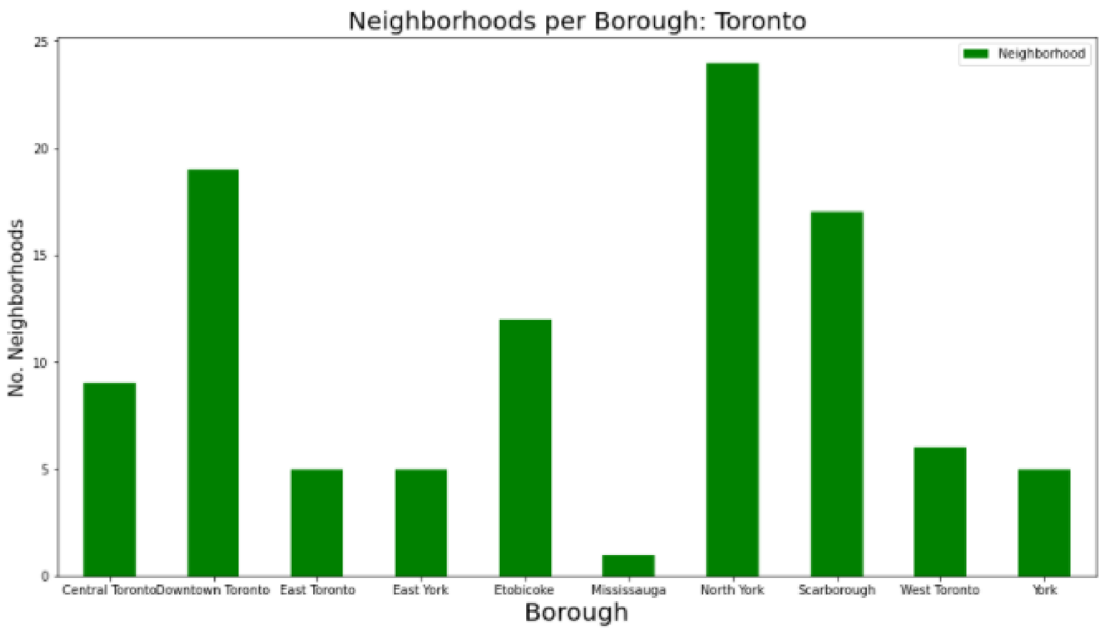
In [5]:
```
NYork_PC =requests.get("https://www.health.ny.gov/statistics/cancer/registry/appendix/neighborhoods.htm").text
BSoup=BeautifulSoup(NYork_PC, 'html')
tab_N = str(BSoup.table)
dfN = pd.read_html(tab_N)
df_N=dfN[0]
# Dropping the rows where Borough is 'Not assigned'
N_df = df_N[df_N.Borough != 'Not assigned']

# Combining the neighbourhoods with same Postalcode
NY_df = N_df.groupby(['ZIP Codes','Borough'], sort= False).agg(', '.join)
NY_df.reset_index(inplace=True)
# Replacing the name of the neighbourhoods which are 'Not assigned' with names of Borough
NY_df['Neighborhood'] = np.where(NY_df['Neighborhood'] == 'Not assigned',NY_df['Borough'], NY_df['Neighborhood'])

NY_df.head()
```

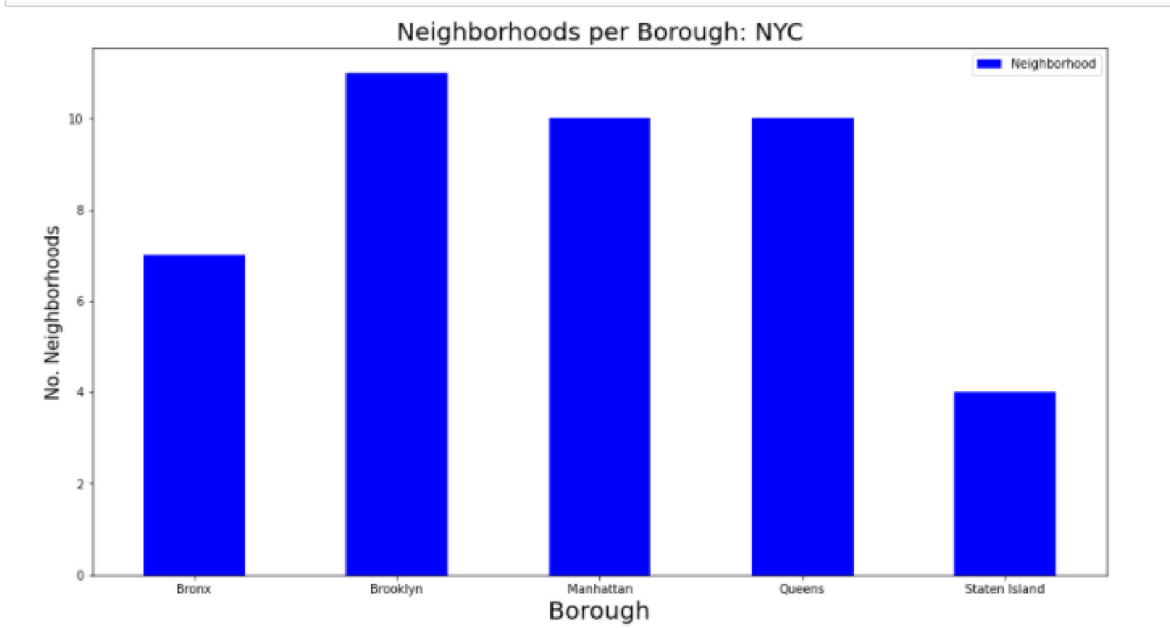
Out[5]:

	ZIP Codes	Borough	Neighborhood
0	10453, 10457, 10460	Bronx	Central Bronx
1	10458, 10467, 10468	Bronx	Bronx Park and Fordham
2	10451, 10452, 10456	Bronx	High Bridge and Morrisania
3	10454, 10455, 10459, 10474	Bronx	Hunts Point and Mott Haven
4	10463, 10471	Bronx	Kingsbridge and Riverdale

Data frame was then visualized in bar plot to see higher number of neighborhoods in each borough. After the analysis, it occurred that North York, Toronto and Brooklyn, NY has the highest number of Neighborhoods.

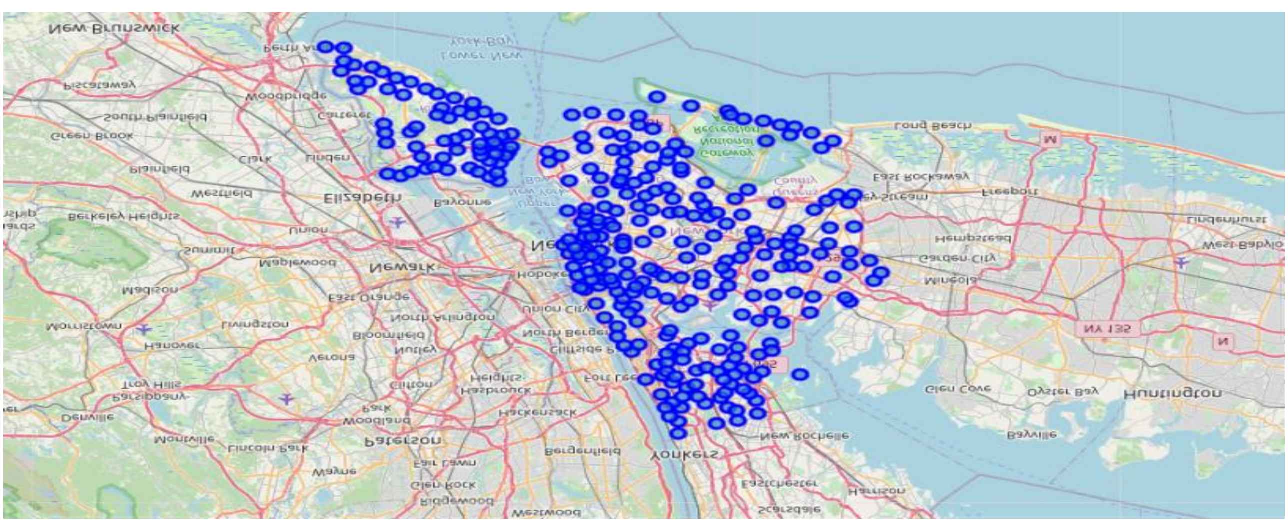
In [176]:
```
# Lets plot a graph for neighbor per Borough
clr = "green"
C_df.groupby('Borough')['Neighborhood'].count().plot.bar(figsize=(15,8), color=clr)
plt.title('Neighborhoods per Borough: Toronto', fontsize = 20)
plt.xlabel('Borough', fontsize = 20)
plt.ylabel('No. Neighborhoods',fontsize = 15)
plt.xticks(rotation = 'horizontal')
plt.legend()
plt.show()
```



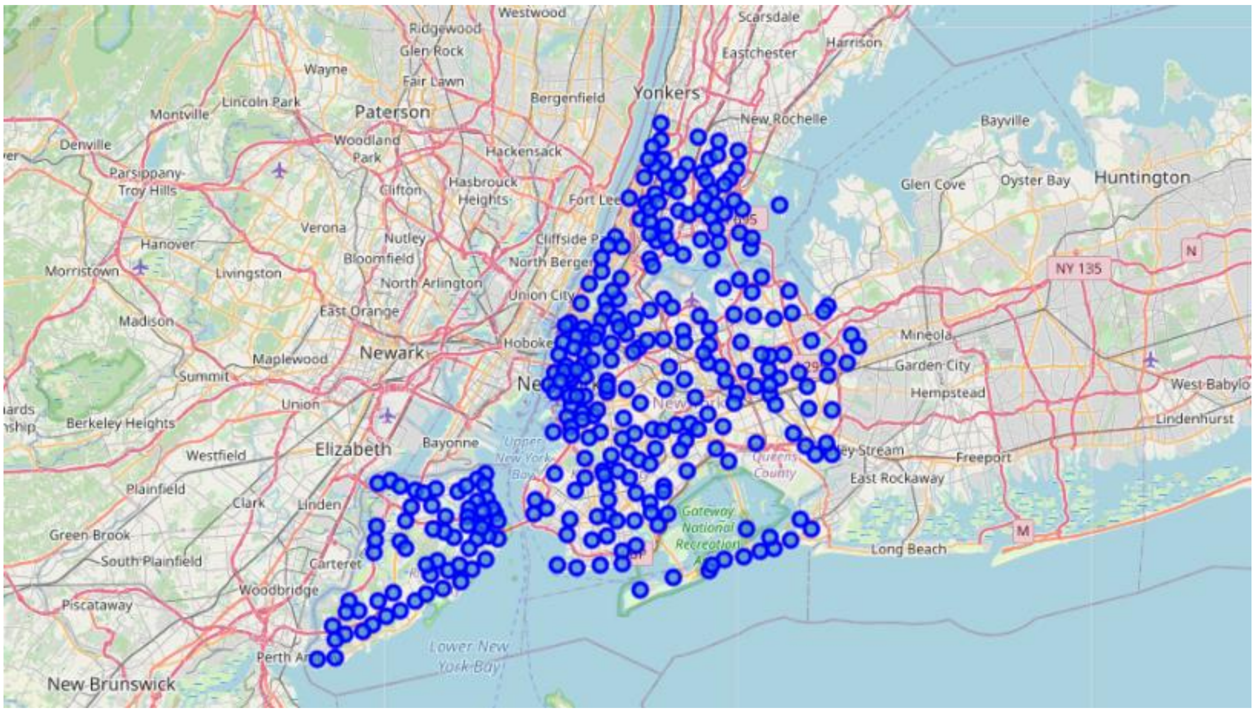


We also created maps to visualize the neighborhoods in both the cities.

```
In [17]: # create map of New York using latitude and longitude values
         map_newyork = folium.Map(location=[latitude_NYork, longitude_NYork], zoom_start=10)

         # add markers to map
         for lat, lng, borough, neighborhood in zip(neighborhoods['Latitude'], neighborhoods['Longitude'], neighborhoods['Borough'], neighborhoods['Neighborhood']):
             label = '{}, {}'.format(neighborhood, borough)
             label = folium.Popup(label, parse_html=True)
             folium.CircleMarker(
                 [lat, lng],
                 radius=5,
                 popup=label,
                 color='blue',
                 fill=True,
                 fill_color='#3186cc',
                 fill_opacity=0.7,
                 parse_html=False).add_to(map_newyork)

         map_newyork
```



```
In [18]: # create map of Toronto using latitude and longitude values
map_toronto = folium.Map(location=[latitude_toronto, longitude_toronto], zoom_start=10)

# add markers to map
for lat, lng, borough, neighborhood in zip(Cdf['Latitude'], Cdf['Longitude'], Cdf['Borough'], Cdf['Neighborhood']):
    label = '{}, {}'.format(neighborhood, borough)
    label = folium.Popup(label, parse_html=True)
    folium.CircleMarker(
        [lat, lng],
        radius=2,
        popup=label,
        color='blue',
        fill=True,
        fill_color='#3186cc',
        fill_opacity=0.75,
        parse_html=False).add_to(map_toronto)

map_toronto
```



We have used API Foursquare to extract the 5 most common Venues in each Neighborhood in both the cities.

```
In [30]: def return_most_common_venues(row, T_top_venues):
    row_categories = row.iloc[1:]
    row_categories_sorted = row_categories.sort_values(ascending=False)

    return row_categories_sorted.index.values[0:T_top_venues]
```

```
In [31]: T_top_venues = 5

indicators = ['st', 'nd', 'rd']

# create columns according to number of top venues
columns = ['Neighborhood']
for ind in np.arange(T_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
T_neighborhoods_venues_sorted = pd.DataFrame(columns=columns)
T_neighborhoods_venues_sorted['Neighborhood'] = toronto_grouped['Neighborhood']

for ind in np.arange(toronto_grouped.shape[0]):
    T_neighborhoods_venues_sorted.iloc[ind, 1:] = return_most_common_venues(toronto_grouped.iloc[ind, :], T_top_venues)

T_neighborhoods_venues_sorted.head()
```

Out[35]:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Bathurst Manor, Wilson Heights, Downsview North	Coffee Shop	Fried Chicken Joint	Deli / Bodega	Yoga Studio	Dessert Shop
1	Bedford Park, Lawrence Manor East	Sushi Restaurant	Coffee Shop	Italian Restaurant	Clothing Store	College Gym
2	CN Tower, King and Spadina, Railway Lands, Har...	Performing Arts Venue	Yoga Studio	Gastropub	Coffee Shop	College Gym
3	Central Bay Street	Pharmacy	Sandwich Place	Coffee Shop	Yoga Studio	Clothing Store
4	Commerce Court, Victoria Hotel	American Restaurant	Soup Place	Fast Food Restaurant	Deli / Bodega	Gym

In [63]:

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

    return row_categories_sorted.index.values[0:NY_top_venues]
```

In [64]:

```
NY_top_venues = 5

indicators = ['st', 'nd', 'rd']

# create columns according to number of top venues
columns = ['Neighborhood']
for ind in np.arange(NY_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
NY_neighborhoods_venues_sorted = pd.DataFrame(columns=columns)
NY_neighborhoods_venues_sorted['Neighborhood'] = NYork_grouped['Neighborhood']

for ind in np.arange(NYork_grouped.shape[0]):
    NY_neighborhoods_venues_sorted.iloc[ind, 1:] = return_most_common_venues(NYork_grouped.iloc[ind, :], NY_top_venues)

NY_neighborhoods_venues_sorted.head()
```

Out[64]:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Allerton	Deli / Bodega	Gas Station	Donut Shop	Women's Store	Dog Run
1	Arlington	Construction & Landscaping	Bus Stop	Food	Filipino Restaurant	Fast Food Restaurant
2	Astoria Heights	Laundromat	Women's Store	Doctor's Office	Filipino Restaurant	Fast Food Restaurant
3	Battery Park City	Park	Boat or Ferry	Playground	Cooking School	Historic Site
4	Bay Ridge	Hookah Bar	Market	Pizza Place	Pool Hall	Chinese Restaurant

Next approach is **Clustering** which is one of the most common exploratory data analysis technique used to get an intuition about the structure of the data. We have specified number clusters to 5 for both the cities and have clustered the neighborhoods based on the venue categories. The code is given below:

```
In [65]: # set number of clusters
T_kclusters = 5

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

# run k-means clustering
T_kmeans = KMeans(n_clusters=T_kclusters, random_state=0).fit(toronto_grouped_clustering)

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

Out[65]: array([4, 4, 1, 1, 1, 1, 4, 4, 1, 1], dtype=int32)
```

```
In [66]: # add clustering labels
T_neighborhoods_venues_sorted.insert(0, 'CA Cluster Labels', T_kmeans.labels_)

toronto_merged = Cdf

# merge toronto_grouped with Ndf to add latitude/longitude for each neighborhood
toronto_merged = toronto_merged.join(T_neighborhoods_venues_sorted.set_index('Neighborhood'), on='Neighborhood')
toronto_merged = toronto_merged.dropna()
toronto_merged.head()
```

Out[66]:

	PostalCode	Borough	Neighborhood	Latitude	Longitude	CA Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
2	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636	1.0	Park	Breakfast Spot	Spa	Yoga Studio	Cupcake Shop
3	M6A	North York	Lawrence Manor, Lawrence Heights	43.718518	-79.464763	1.0	Clothing Store	Boutique	Yoga Studio	Coffee Shop	College Gym
4	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government	43.662301	-79.389494	1.0	Sculpture Garden	Yoga Studio	Clothing Store	Coffee Shop	College Gym
7	M3B	North York	Don Mills	43.745906	-79.352188	4.0	Restaurant	Baseball Field	Yoga Studio	Dessert Shop	Coffee Shop
9	M5B	Downtown Toronto	Garden District, Ryerson	43.657162	-79.378937	4.0	Art Gallery	Coffee Shop	Yoga Studio	Dessert Shop	College Gym

```
In [67]: # set number of clusters
NY_kclusters = 5

NYork_grouped_clustering = NYork_grouped.drop('Neighborhood', 1)

# run k-means clustering
NY_kmeans = KMeans(n_clusters=NY_kclusters, random_state=0).fit(NYork_grouped_clustering)

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

Out[67]: array([2, 1, 0, 0, 0, 0, 0, 0, 0, 0], dtype=int32)
```

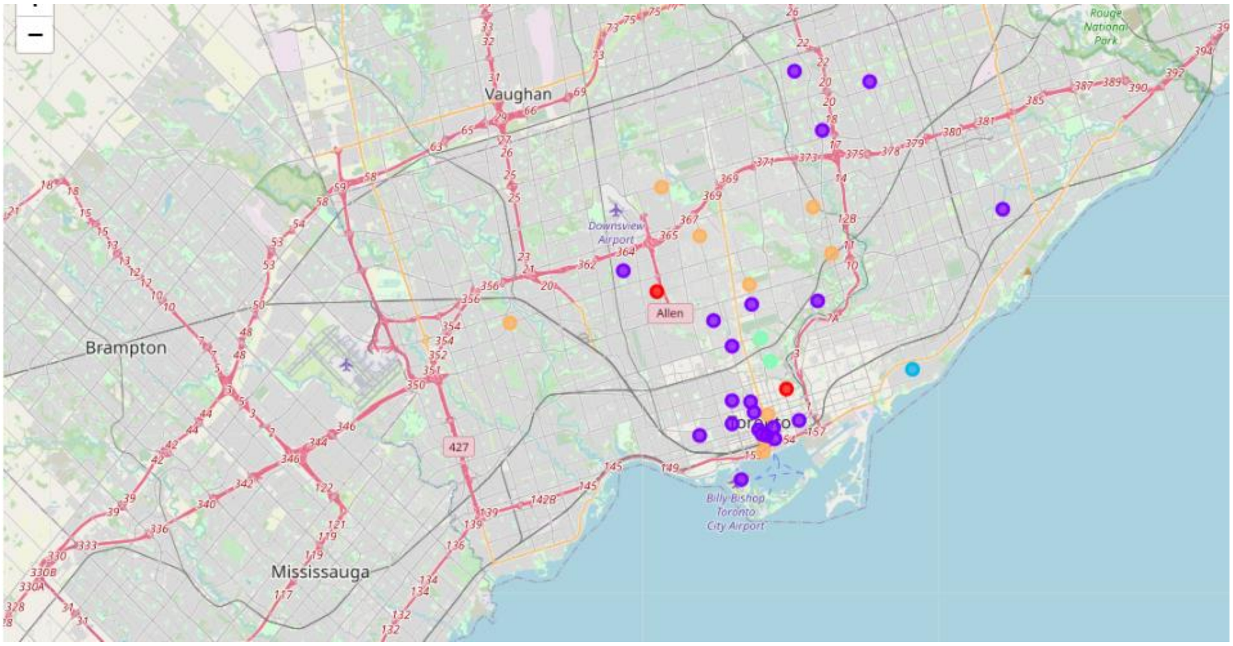
```
In [68]: # add clustering labels
NY_neighborhoods_venues_sorted.insert(0, 'NY Cluster Labels', NY_kmeans.labels_)

NYork_merged = neighborhoods

# merge toronto_grouped with Ndf to add latitude/longitude for each neighborhood
NYork_merged = NYork_merged.join(NY_neighborhoods_venues_sorted.set_index('Neighborhood'), on='Neighborhood')
NYork_merged = NYork_merged.dropna()
NYork_merged.head()
```

Out[68]:

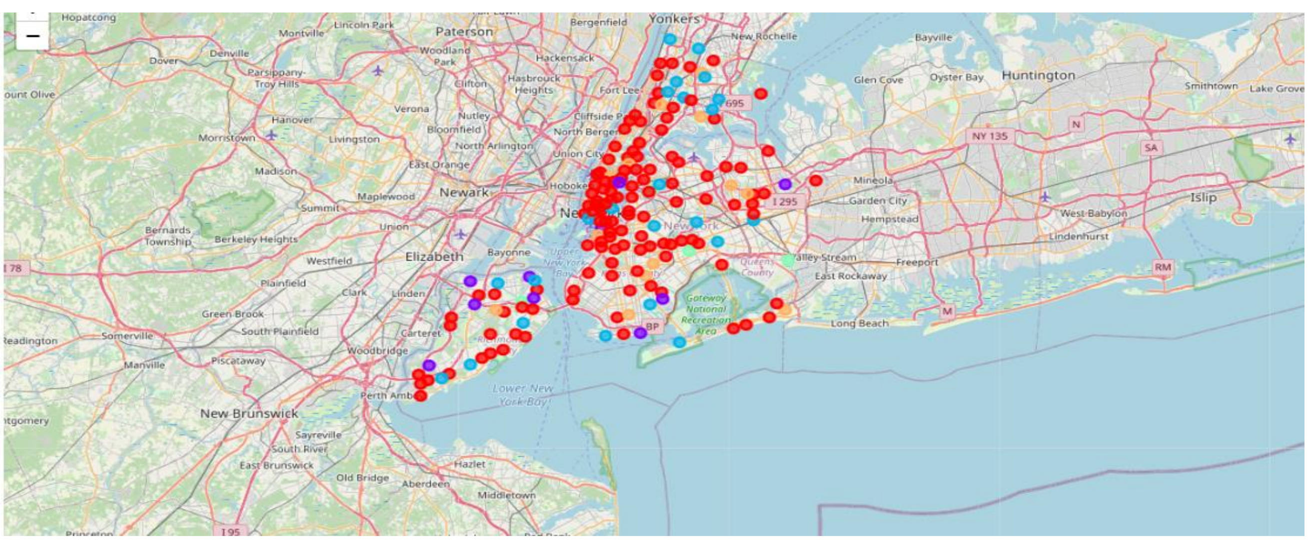
	Borough	Neighborhood	Latitude	Longitude	NY Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
5	Bronx	Kingsbridge	40.881687	-73.902818	0.0	Mattress Store	Supermarket	Thrift / Vintage Store	Gourmet Shop	Women's Store
7	Bronx	Woodlawn	40.898273	-73.867315	2.0	Pub	Deli / Bodega	Pizza Place	Bar	Pharmacy
8	Bronx	Norwood	40.877224	-73.879391	0.0	History Museum	Women's Store	Doctor's Office	Filipino Restaurant	Fast Food Restaurant
12	Bronx	City Island	40.847247	-73.786488	0.0	Pizza Place	Jewelry Store	French Restaurant	Frozen Yogurt Shop	Diner
15	Bronx	Morris Heights	40.847898	-73.919672	0.0	Convenience Store	Women's Store	Fish & Chips Shop	Filipino Restaurant	Fast Food Restaurant

Below is the code and map to visualize each cluster that belong to a different color with different traits. Detailed code is provided below:

```
In [124]: map_clusters_T = folium.Map(location=[latitude_toronto, longitude_toronto], zoom_start=11)

# set color scheme for the clusters
x_T = np.arange(T_kclusters)
ys_T = [i + x_T + (i*x_T)**2 for i in range(T_kclusters)]
T_colors_array = cm.rainbow(np.linspace(0, 1, len(ys_T)))
T_rainbow = [colors.rgb2hex(i) for i in T_colors_array]

# add markers to the map
markers_colors = []
for lat, lon, poi, cluster in zip(toronto_merged['Latitude'], toronto_merged['Longitude'], toronto_merged['Neighborhood'],toronto_merged['CA Cluster Labels']):
    label = folium.Popup(str(poi) + ' Cluster ' + str(cluster), parse_html=True)
    folium.CircleMarker(
        [lat, lon],
        radius=5,
        popup=label,
        color = T_rainbow[int(cluster)-1],
        fill=True,
        fill_color=T_rainbow[int(cluster)-1],
        fill_opacity=0.7).add_to(map_clusters_T)

map_clusters_T
```



In [70]:
```
latitudes_NYork = 40.7127281
longitudes_NYork = -74.0060152
map_clusters_NY = folium.Map(location=[latitudes_NYork, longitudes_NYork], zoom_start=11)

# set color scheme for the clusters
x_NY = np.arange(NY_kclusters)
ys_NY = [i + x_NY + (i*x_NY)**2 for i in range(NY_kclusters)]
NY_colors_array = cm.rainbow(np.linspace(0, 1, len(ys_NY)))
NY_rainbow = [colors.rgb2hex(i) for i in NY_colors_array]

# add markers to the map
markers_colors = []
for lat, lon, poi, cluster in zip(NYork_merged['Latitude'], NYork_merged['Longitude'], NYork_merged['Neighborhood'],NYork_merged['NY Cluster Labels']):
    label = folium.Popup(str(poi) + ' Cluster ' + str(cluster), parse_html=True)
    folium.CircleMarker(
        [lat, lon],
        radius=5,
        popup=label,
        color = NY_rainbow[int(cluster)-1],
        fill=True,
        fill_color=NY_rainbow[int(cluster)-1],
        fill_opacity=0.7).add_to(map_clusters_NY)

map_clusters_NY
```



```
In [138]: toronto_merged.loc[toronto_merged['CA Cluster Labels'] == 0, toronto_merged.columns[[1] + list(range(5, toronto_merged.shape[1]))]]
```

Out[138]:

	Borough	CA Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
10	North York	0.0	Italian Restaurant	Yoga Studio	Gastropub	Coffee Shop	College Gym
96	Downtown Toronto	0.0	Italian Restaurant	Yoga Studio	Gastropub	Coffee Shop	College Gym

```
In [140]: toronto_merged.loc[toronto_merged['CA Cluster Labels'] == 1, toronto_merged.columns[[1] + list(range(5, toronto_merged.shape[1]))]]
```

Out[140]:

	Borough	CA Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
2	Downtown Toronto	1.0	Park	Breakfast Spot	Spa	Yoga Studio	Cupcake Shop
3	North York	1.0	Clothing Store	Boutique	Yoga Studio	Coffee Shop	College Gym
4	Downtown Toronto	1.0	Sculpture Garden	Yoga Studio	Clothing Store	Coffee Shop	College Gym
15	Downtown Toronto	1.0	Japanese Restaurant	Gay Bar	Diner	Breakfast Spot	Italian Restaurant
24	Downtown Toronto	1.0	Pharmacy	Sandwich Place	Coffee Shop	Yoga Studio	Clothing Store
27	North York	1.0	Fast Food Restaurant	Yoga Studio	Dessert Shop	Coffee Shop	College Gym
29	East York	1.0	Housing Development	Indian Restaurant	Coffee Shop	Sandwich Place	College Gym
30	Downtown Toronto	1.0	Garden	Coffee Shop	Vegetarian / Vegan Restaurant	Food Court	Japanese Restaurant
32	Scarborough	1.0	Playground	Yoga Studio	Gastropub	Coffee Shop	College Gym
33	North York	1.0	Theater	Bank	Yoga Studio	Dessert Shop	Coffee Shop
37	West Toronto	1.0	Yoga Studio	Bar	Japanese Restaurant	Coffee Shop	Brewery
48	Downtown Toronto	1.0	American Restaurant	Soup Place	Fast Food Restaurant	Deli / Bodega	Gym
68	Central Toronto	1.0	Furniture / Home Store	Yoga Studio	Dessert Shop	Coffee Shop	College Gym
79	Central Toronto	1.0	Café	Dessert Shop	Toy / Game Store	Coffee Shop	Italian Restaurant
80	Downtown Toronto	1.0	College Gym	Yoga Studio	Dessert Shop	Coffee Shop	Concert Hall
84	Downtown Toronto	1.0	Cocktail Bar	Café	Thrift / Vintage Store	Farmers Market	Liquor Store
86	Central Toronto	1.0	Liquor Store	Supermarket	Yoga Studio	Clothing Store	Coffee Shop

```
In [144]: NYork_merged.loc[NYork_merged['NY Cluster Labels'] == 0, NYork_merged.columns[[1] + list(range(5, NYork_merged.shape[1]))]]
```

Out[144]:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
5	Kingsbridge	Mattress Store	Supermarket	Thrift / Vintage Store	Gourmet Shop	Women's Store
8	Norwood	History Museum	Women's Store	Doctor's Office	Filipino Restaurant	Fast Food Restaurant
12	City Island	Pizza Place	Jewelry Store	French Restaurant	Frozen Yogurt Shop	Diner
15	Morris Heights	Convenience Store	Women's Store	Fish & Chips Shop	Filipino Restaurant	Fast Food Restaurant
18	West Farms	Diner	Women's Store	Dog Run	Fish & Chips Shop	Filipino Restaurant
...	...	...	...	...	...	...
295	Highland Park	Gym / Fitness Center	Park	Women's Store	Discount Store	Fast Food Restaurant
300	Erasmus	Bar	Women's Store	Dog Run	Fish & Chips Shop	Filipino Restaurant
301	Hudson Yards	Pet Store	Women's Store	Doctor's Office	Filipino Restaurant	Fast Food Restaurant
302	Hammels	Beach	Women's Store	Dog Run	Fish & Chips Shop	Filipino Restaurant
303	Bayswater	Construction & Landscaping	Convenience Store	Fish & Chips Shop	Filipino Restaurant	Fast Food Restaurant

122 rows × 6 columns

```
In [169]: NYork_merged.loc[NYork_merged['NY Cluster Labels'] == 1, NYork_merged.columns[[1] + list(range(5, NYork_merged.shape[1]))]]
```

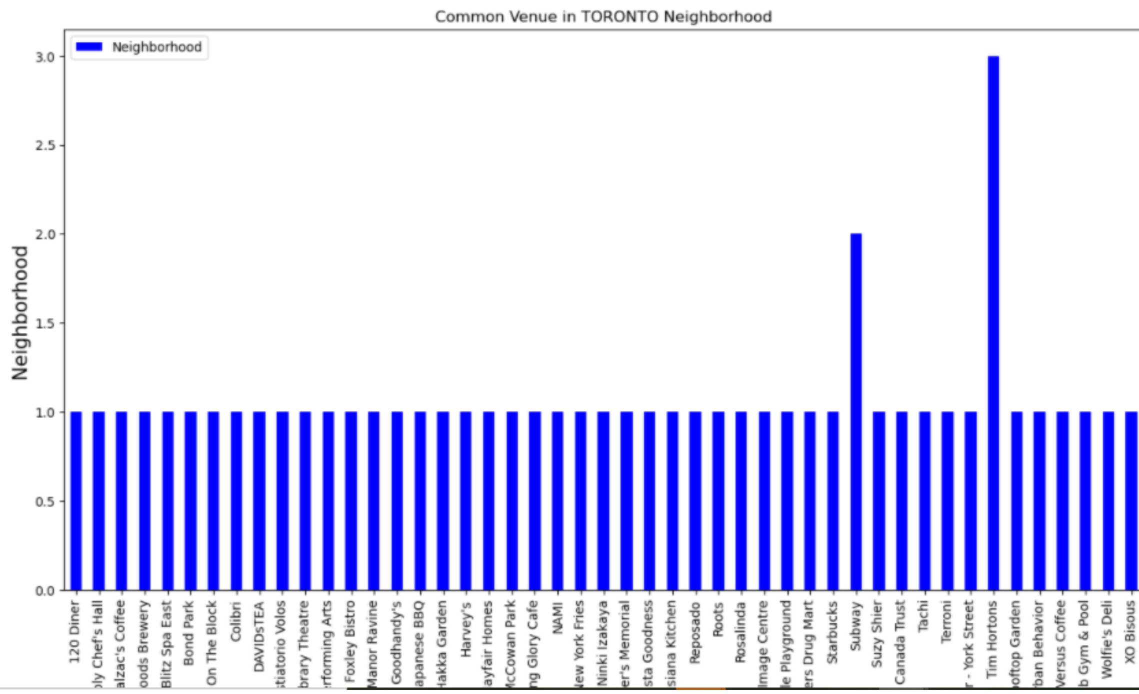
Out[169]:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
64	Brooklyn Heights	Pharmacy	Women's Store	Doctor's Office	Filipino Restaurant	Fast Food Restaurant
77	Manhattan Beach	Ice Cream Shop	Bus Stop	Women's Store	Doctor's Office	Filipino Restaurant
161	Oakland Gardens	Wine Shop	Bus Stop	Food	Filipino Restaurant	Fast Food Restaurant
198	New Brighton	Bus Stop	Women's Store	Food	Filipino Restaurant	Fast Food Restaurant
227	Arlington	Construction & Landscaping	Bus Stop	Food	Filipino Restaurant	Fast Food Restaurant
246	Bulls Head	Pharmacy	Bus Stop	Women's Store	Doctor's Office	Fast Food Restaurant
262	Mill Basin	Pharmacy	Women's Store	Doctor's Office	Filipino Restaurant	Fast Food Restaurant

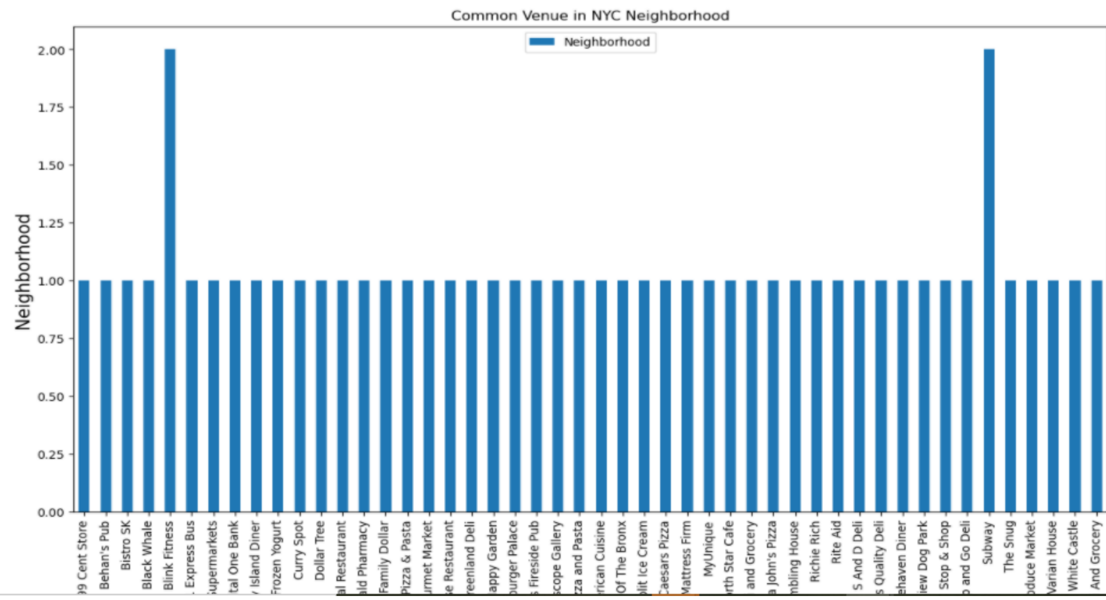
Upon analyzing clustering method, we explored that the most common venues in both the cities are Italian and fast-food restaurants, Deli/ Bodega, Coffee shops and Gym.

To see explicitly, below is the code to see the most common venues in Toronto and New York city. We discovered that Tim Horton and Subway are the 2 most common venues in Toronto. On the other side, New York's 2 most common venues are Subway and Blink Fitness.

In [135]:
```
Tor = toronto_venues.head(50)[['Venue','Neighborhood']]
plt.figure(figsize=(9,5), dpi = 100)
plt.title('Common Venue in TORONTO Neighborhood')
plt.xlabel('Venue', fontsize = 15)
plt.ylabel('Neighborhood', fontsize=15)
Tor.groupby('Venue')['Neighborhood'].count().plot.bar(figsize=(15,8), color=clr)
plt.legend()
plt.show()
```



In [178]:
```
NY = NewYork_venues.head(50)[['Neighborhood','Venue']]
plt.figure(figsize=(10,5), dpi = 100)
plt.title('Common Venue in NYC Neighborhood')
plt.xlabel('Venue', fontsize = 10)
plt.ylabel('Neighborhood', fontsize=15)
NY.groupby('Venue')['Neighborhood'].count().plot.bar(figsize=(15,8))
plt.legend()
plt.show()
```

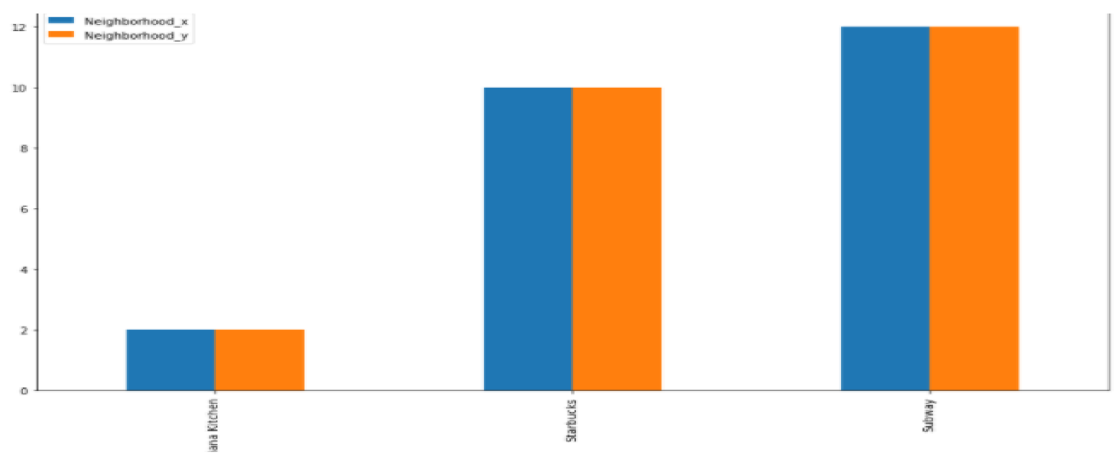


After we combine both the data frames from above, we see that Subways, Starbucks and Popeyes are the most popular venues in most of the neighborhoods in both the cities.

```
In [137]: Tor = toronto_venues[['Venue','Neighborhood']]
          NY = NewYork_venues[['Neighborhood','Venue']]
          Tor_NY.set_index('Venue')
          Tor_NY.groupby('Venue').count().plot.bar(figsize=(15,8))
          Tor_NY=pd.merge(Tor,NY, on = 'Venue')
          Tor_NY
```

Out[137]:

	Venue	Neighborhood_x	Neighborhood_y
0	Subway	Central Bay Street	Fordham
1	Subway	Central Bay Street	Westchester Square
2	Subway	Central Bay Street	Kensington
3	Subway	Central Bay Street	Inwood
4	Subway	Thorncliffe Park	Fordham
5	Subway	Thorncliffe Park	Westchester Square
6	Subway	Thorncliffe Park	Kensington
7	Subway	Thorncliffe Park	Inwood
8	Subway	Steeles West, L'Amoreaux West	Fordham
9	Subway	Steeles West, L'Amoreaux West	Westchester Square
10	Subway	Steeles West, L'Amoreaux West	Kensington
11	Subway	Steeles West, L'Amoreaux West	Inwood
12	Popeyes Louisiana Kitchen	Bathurst Manor, Wilson Heights, Downsview North	Sunset Park
13	Popeyes Louisiana Kitchen	Bathurst Manor, Wilson Heights, Downsview North	Central Harlem
14	Starbucks	Harbourfront East, Union Station, Toronto Islands	Park Slope
15	Starbucks	Harbourfront East, Union Station, Toronto Islands	Long Island City
16	Starbucks	Toronto Dominion Centre, Design Exchange	Park Slope
17	Starbucks	Toronto Dominion Centre, Design Exchange	Long Island City
18	Starbucks	Toronto Dominion Centre, Design Exchange	Park Slope
19	Starbucks	Toronto Dominion Centre, Design Exchange	Long Island City
20	Starbucks	Bedford Park, Lawrence Manor East	Park Slope
21	Starbucks	Bedford Park, Lawrence Manor East	Long Island City
22	Starbucks	Westmount	Park Slope
23	Starbucks	Westmount	Long Island City



Discussion and Recommendation:

Based on the result of our analysis, we found that Largest neighborhoods in Toronto is North York and Brooklyn in New York. After clustering the data of the respective neighborhoods, both cities (Boroughs) have venues which can be explored and attract the Tourists. The neighborhoods are much similar in features like Theaters, opera houses, food places, clubs, museums, parks etc. As far as concern to dissimilarity, it differs in terms of some unique places like historical places and monuments.

As far as recommendation is concerned, we recommend Toronto Dominion and long Island see would be having similar characteristics on a larger scale.

Conclusion

After K Clustering, it appeared that Italian Food, Deli/Bodega, fast-food Restaurant, and many other places were in common in both Toronto and New York City neighborhoods. However, the most common venues are Popeyes, Starbucks and Subway. As we know that every place is unique in its own way, so that’s argument is present in both neighborhoods. The dissimilarity exists in terms of some different venues and facilities but not on a larger extent.