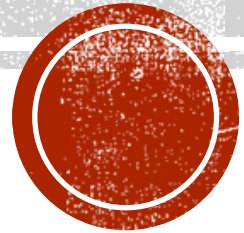
THE BATTLE OF NEIGHBORHOODS

Finding Good Housing and Good School in North York, Toronto



INTRODUCTION

- Many people are moving to various location in Canada. Before they move to a new city or state or country, they do a lot of research for good housing prices and good reputable school for children. They also want to live in good safe neighborhoods with easy access to grocery, hospital, mall, theater, cafe, park, museum, beach, public transportation, etc.

- North York is one of the best place in Canada to live in. First of all, the place is very convenient in terms of transportation. There are lot of transportation system that connects North York to the rest of Toronto and beyond. North York will give you an experience to a mouthwatering restaurants and cafe. Moreover, this place also have an abundance of shop, boutiques and malls that you can enjoy shopping. An all year round festival will give you feeling at home, where you have a chance to connect with people. The city is rich with historical landmarks. Lastly, a great education opportunities from elementary to college surrounds North York.

INTRODUCTION

- This project is to help people finding good housing and good school in North York, Toronto. It will help people make a decision to choose a good neighborhood around the surrounding neighborhoods in North York, Toronto. The features analysis used are comparison North York with its neighborhoods, inlcuding median housing price and school ratings, crime rates, road connectivity, weather conditions, and recreational facilities.

DATA

- This project will use Toronto dataset which we have scrapped from wikipedia on Week 3 project. Dataset consists of latitude and longitude, zip codes. Then we are going to get the latitude and longitude of North York, Toronto and its surrounding neighborhoods.

- Data Link: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

DATA

- **Foursquare API Data:**
- This project is using Foursquare API that can provide data location with information about venue names, locations, menus, photos and events within an area of interest. After getting the list of neighborhoods, Foursquare API is connected to gather information about venues inside each and every neighborhood. For each neighborhood, the radius is 100 meter. The data retrieved from Foursquare contained information of venues within a specified distance of the longitude and latitude of the postcodes. Here is the breakdown of the information obtained per venue:

- 1. Neighborhood
- 2. Neighborhood Latitude
- 3. Neighborhood Longitude
- 4. Venue
- 5. Name of the venue example the name of a store or restaurant
- 6. Venue Latitude
- 7. Venue Longitude
- 8. Venue Category

COORDINATES OF NORTH YORK

Using Geocoder and geopy library to find North York Coordinates

```
address = 'North York,Toronto'

geolocator = Nominatim(user_agent="smy-application")
location = geolocator.geocode(address)
latitude_x = location.latitude
longitude_y = location.longitude
print(f"The Geograpical Co-ordinate of North York, Toronto are {latitude_x}, {longitude_y}.")
```

```
The Geograpical Co-ordinate of North York, Toronto are 43.7543263, -79.44911696639593.
```

MAP OF NORTH YORK

Using Folium to visualize the data

In [51]:
```
map_North_York = folium.Map(location=[latitude_x, longitude_y], zoom_start=10)

for lat, lng, nei in zip(df_2['Latitude'], df_2['Longitude'], df_2['Neighborhood']):

    label = '{}'.format(nei)
    label = folium.Popup(label, parse_html=True)
    folium.CircleMarker(
        [lat, lng],
        radius=5,
        popup=label,
        color='blue',
        fill=True,
        fill_color='#3186cc',
        fill_opacity=0.7,
        parse_html=False).add_to(map_North_York)

map_North_York
```

Out[51]:

METHODOLOGY

- **Clustering:**
  - This project uses clustering using K-means clustering algorithm to compare the similarities of two cities, segment them, and group them into clusters to find similar neighborhoods in a big city.

In [196]:
```
neighborhoods_venues_sorted.insert(0, 'Cluster Labels', kmeans.labels_)

North_York_merged =df_2.iloc[51:75,:] # North York Borough index is 51 to 74
```

In [197]:
```
# merge toronto_grouped with toronto_data to add latitude/longitude for each neighborhood
North_York_merged = North_York_merged.join(neighborhoods_venues_sorted.set_index('Neighborhood'), on='Neighborhood')

North_York_merged
```

Out[197]:

stalcode	Neighborhood	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	
M2H	Hillcrest Village	43.80225	-79.35558	2	Park	Pharmacy	Residential Building (Apartment / Condo)	Chinese Restaurant	Bakery	Fast Food Restaurant	Elementary School	Donut Shop	I R
M2J	Fairview, Henry Farm, Oriole	43.78097	-79.34781	0	Clothing Store	Coffee Shop	Fast Food Restaurant	Bakery	Juice Bar	Women's Store	Restaurant	Bank	
M2K	Bayview Village	43.78112	-79.38060	0	Gas Station	Flower Shop	Park	Trail	Asian Restaurant	Donut Shop	Eastern European Restaurant	Electronics Store	El
M2L	York Mills, Silver Hills	43.75698	-79.38060	2	Concert Hall	Park	Falafel Restaurant	Eastern European Restaurant	Electronics Store	Elementary School	Escape Room	Ethiopian Restaurant	
M2M	Willowdale, Newtonbrook	43.79182	-79.41372	0	Korean Restaurant	Middle Eastern Restaurant	Coffee Shop	Café	Pizza Place	Diner	Japanese Restaurant	Sandwich Place	R
M2N	Willowdale, Willowdale East	43.76774	-79.40728	0	Coffee Shop	Japanese Restaurant	Pizza Place	Ramen Restaurant	Fast Food Restaurant	Café	Sandwich Place	Restaurant	R
M2P	York Mills West	43.74778	-79.40033	0	Coffee Shop	Park	Restaurant	Convenience Store	Intersection	Sandwich Place	Burrito Place	Thai Restaurant	
M2R	Willowdale, Willowdale West	43.77989	-79.44678	0	Coffee Shop	Pharmacy	Park	Grocery Store	Pizza Place	Butcher	Dumpling Restaurant	Eastern European Restaurant	El
M3A	Parkwoods	43.75245	-79.32991	2	Park	Food & Drink Shop	Burger Joint	Pet Store	Yoga Studio	Falafel Restaurant	Electronics Store	Elementary School	

In [198]:
```
kclusters = 10
```

METHODOLOGY

- **Most Common venues near Neighborhood :**

In [193]:
```
import numpy as np
num_top_venues = 10

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

columns = ['Neighborhood']
for ind in np.arange(num_top_venues):
    try:
        columns.append(f"{ind+1}{indicators[ind]} Most Common Venue")
    except:
        columns.append(f"{ind+1}th Most Common Venue")

neighborhoods_venues_sorted = pd.DataFrame(columns=columns)
neighborhoods_venues_sorted['Neighborhood'] = North_York_grouped['Neighborhood']

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

neighborhoods_venues_sorted.head()
```

Out[193]:

	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	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Agincourt	Shopping Mall	Pool	Bakery	Sushi Restaurant	Latin American Restaurant	Supermarket	Breakfast Spot	Bubble Tea Shop	Skating Rink	Shanghai Restaurant
1	Alderwood, Long Branch	Convenience Store	Pizza Place	Sandwich Place	Gym	Pharmacy	Coffee Shop	Print Shop	Pub	Gas Station	Event Space
2	Bathurst Manor, Wilson Heights, Downsview North	Coffee Shop	Park	Mobile Phone Shop	Mediterranean Restaurant	Pizza Place	Middle Eastern Restaurant	Deli / Bodega	Restaurant	Sandwich Place	Fried Chicken Joint
3	Bayview Village	Gas Station	Flower Shop	Park	Trail	Asian Restaurant	Donut Shop	Eastern European Restaurant	Electronics Store	Elementary School	Escape Room
4	Bedford Park, Lawrence Manor East	Pizza Place	Sandwich Place	Italian Restaurant	Coffee Shop	Comfort Food Restaurant	Intersection	Butcher	Café	Sports Club	Liquor Store

METHODOLOGY

- **Work Flow**

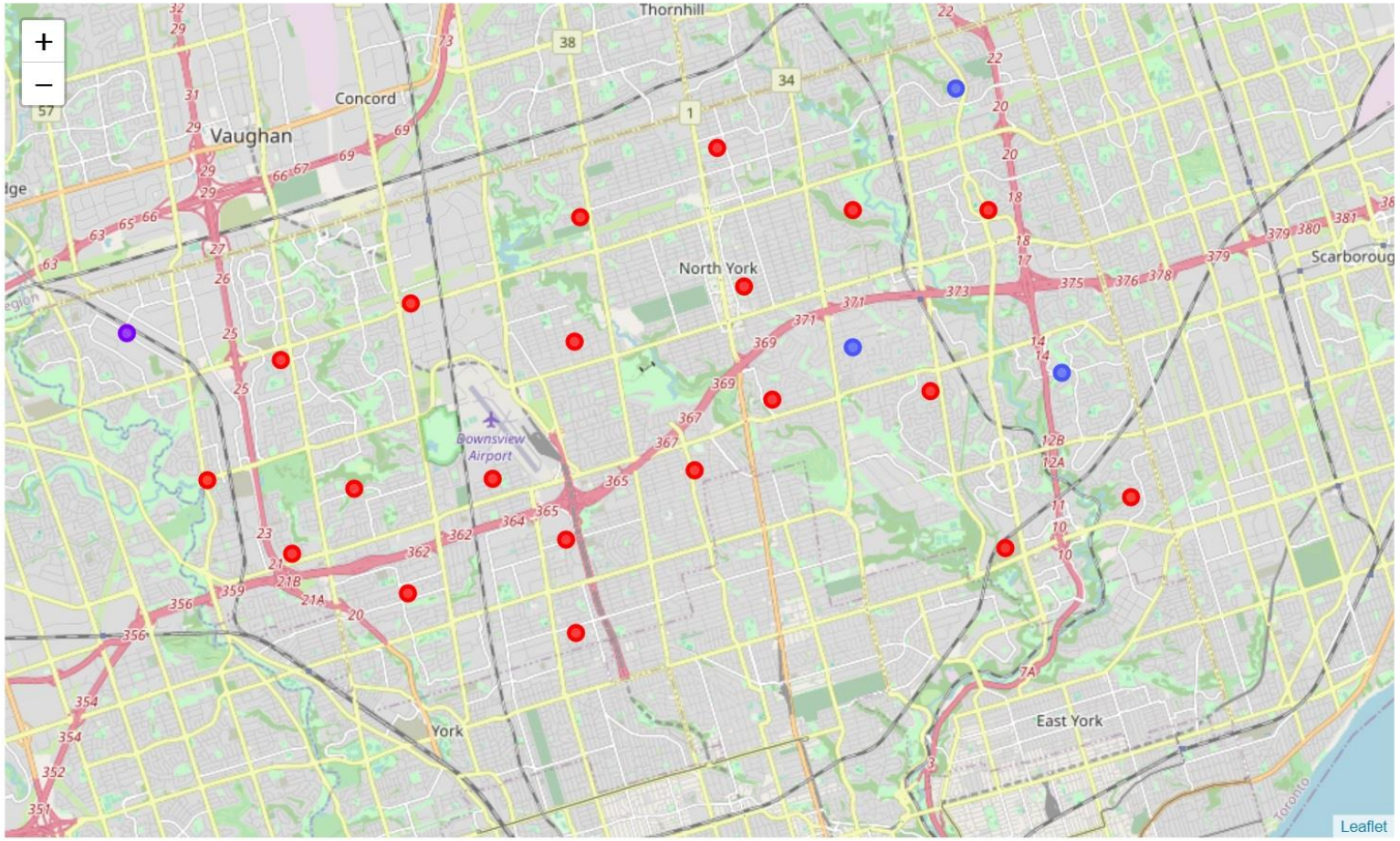
After extracting and reading the data, it translates into a pandas dataframe then we are preprocessing and cleaning the data to make sure there is no null/NaN. We get the North York coordinates and visualize it with Folium. These data elements are needed to call Foursquare web service to get the venues available in that neighborhood. Foursquare API is used as data gathering source because it has database of millions of places, especially their places API which provides the ability to perform location search, location sharing and details about a business. Using credentials of Foursquare API features of nearby places with the number of places per neighborhood parameter be set to 100 and the radius parameter would be set to 500.

- **Libraries**
  - Pandas: For creating and manipulating dataframes.
  - Beautiful Soup and Requests: To scrap and library to handle http requests.
  - Matplotlib: Python Plotting Module.
  - JSON: Library to handle JSON files.
  - XML: To separate data from presentation and XML stores data in plain text format.
  - Folium: Python visualization library would be used to visualize the neighborhoods cluster distribution of using interactive leaflet map.
  - Geocoder: To retrieve Location Data.
  - Scikit Learn: For importing k-means clustering.

RESULTS

- Here are the results of cluster map in North York and the graph based on average housing price and school ratings cluster map

Out[199]:



Average_Housing_Price
600000
500000
400000
300000
200000
100000
0
Lawrence Park
Davisville North
North Toronto West, Lawrence Park
Davisville
Moore Park, Summerhill East
Summerhill West, Rathnelly, South Hill, Forest Hill SE, Deer Park
Roselawn
Forest Hill North & West, Forest Hill Road Park
The Annex, North Midtown, Yorkville
Rosedale
St. James Town, Cabbagetown
Church and Wellesley
Regent Park, Harbourfront
Garden District, Ryerson
St. James Town
Bay Street
Central Bay Street
Richmond, Adelaide, King
Harbourfront East, Union Station, Toronto Islands
Toronto Dominion Centre, Design Exchange
Commerce Court, Victoria Hotel
Stn A PO Boxes
Kensington Market, Chinatown, Grange Park
CN Tower, King and Spadina, Railway Lands, Harbourfront West, Bathurst Quay, South Niagara, Island airport
Stn A PO Boxes
First Canadian Place, Underground city
Christie
Queen's Park, Ontario Provincial Government
The Beaches
The Danforth West, Riverdale
India Bazaar, The Beaches West
Studio District
Business reply mail Processing Centre, South Central Letter Processing Plant Toronto
Parkview Hill, Woodbine Gardens
Woodbine Heights
Leaside
Thorncliffe Park
East Toronto, Broadview North (Old East York)
New Toronto, Mimico South, Humber Bay Shores
Alderwood, Long Branch
The Kingsway, Montgomery Road, Old Mill North
Islington Avenue, Humber Valley Village
West Deane Park, Princess Gardens, Martin Grove, Islington, Cloverdale
Etringate, Bloordale Gardens, Old Burnhamthorpe, Markland Wood
Westmount
Kingsview Village, St. Phillips, Martin Grove Gardens, Richview Gardens
South Steeles, Silverstone, Humbergate, Jamestown, Mount Olive, Beaumond Heights, Thistletown, Albion Gardens
Northwest, West Humber - Clairville
Canada Post Gateway Processing Centre
Rexdale, Kipling
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
Downsview
Victoria Village
Hillcrest Village
Bedford Park, Lawrence Manor East
Lawrence Manor, Lawrence Heights
Glencairn
North Park, Maple Leaf Park, Upwood Park
Humber Summit
Humberlea, Emery
Humber Heights, Westmount
Rouge Hill, Port Union, Highland Creek
Guildwood, Morningside, West Hill
Woburn
Cedarbrae
Scarborough Village
Kennedy Park, Ionview, East Birchmount Park
Golden Mile, Clairlea, Oakridge
Cliffside, Cliffcrest, Scarborough Village West
Birch Cliff, Cliffside West
Dorset Park, Wexford Heights, Scarborough Town Centre
Wexford, Maryvale
Agincourt
Clarks Corners, Tam O'Shanter, Sullivan
Milliken, Agincourt North, Steeles East, L'Amoreaux East
Steeles West, L'Amoreaux West
Upper Rouge
Dufferin, Dovercourt Village
Little Portugal, Trinity
Brockton, Parkdale Village, Exhibition Place
High Park, The Junction South
Parkdale, Roncesvalles
Runnymede, Swansea
Runnymede-Cedarvale
Oakwood-Vaughan
Del Ray, Mount Dennis, Keelsdale and Silverthorn
Runnymede, The Junction North
Weston

Top School Rating
0
2
4
6
8
10
Malvern, Rouge
Rouge Hill, Port Union, Highland Creek
Guildwood, Morningside, West Hill
Woburn
Cedarbrae
Scarborough Village
Kennedy Park, Ionview, East Birchmount Park
Golden Mile, Clairlea, Oakridge
Cliffside, Cliffcrest, Scarborough Village West
Birch Cliff, Cliffside West
Dorset Park, Wexford Heights, Scarborough Town Centre
Wexford, Maryvale
Agincourt
Clarks Corners, Tam O'Shanter, Sullivan
Milliken, Agincourt North, Steeles East, L'Amoreaux East
Steeles West, L'Amoreaux West
Upper Rouge
Hillcrest Village
Fairview, Henry Farm, Oriole
Bayview Village
York Mills, Silver Hills
Willowdale, Newtonbrook
Willowdale, Willowdale East
Willowdale, Newtonbrook West
Willowdale, Willowdale West
Parkwoods
Don Mills
Flemingdon Park
Bathurst Manor, Wilson Heights, Downsview North
Northwood Park, York University
Downsview
Downsview
Downsview
Victoria Village
Parkview Hill, Woodbine Gardens
Woodbine Heights
The Beaches
Leaside
Thorncliffe Park
East Toronto, Broadview North (Old East York)
The Danforth West, Riverdale
India Bazaar, The Beaches West
Studio District
Lawrence Park
Davisville North
North Toronto West, Lawrence Park
Davisville
Moore Park, Summerhill East
Rosedale
Summerhill West, Rathnelly, South Hill, Forest Hill SE, Deer Park
St. James Town, Cabbagetown
Church and Wellesley
Regent Park, Harbourfront
Garden District, Ryerson
St. James Town
Berczy Park
Central Bay Street
Richmond, Adelaide, King
Harbourfront East, Union Station, Toronto Islands
Toronto Dominion Centre, Design Exchange
Commerce Court, Victoria Hotel
Bedford Park, Lawrence Manor East
Roselawn
Forest Hill North & West, Forest Hill Road Park
The Annex, North Midtown, Yorkville
Kensington Market, Chinatown, Grange Park
CN Tower, King and Spadina, Railway Lands, Harbourfront West, Bathurst Quay, South Niagara, Island Airport
Stn A PO Boxes
First Canadian Place, Underground city
Lawrence Manor, Lawrence Heights
Glencairn
Humewood-Cedarvale
Caledonia-Fairbanks
Christie
Dufferin, Dovercourt Village
Little Portugal, Trinity
Brockton, Parkdale Village, Exhibition Place
North Park, Maple Leaf Park, Upwood Park
Del Ray, Mount Dennis, Keelesdale and Silverthorn
Runnymede, The Junction South
High Park, The Junction North
Parkdale, Roncesvalles
Runnymede, Swansea
Queen's Park, Ontario Provincial Government
Business reply mail Processing Centre, South Central Letter Processing Plant Toronto
New Toronto, Mimico South, Humber Bay Shores
Alderwood, Long Branch
The Kingsway, Montgomery Road, Old Mill North
K, Sunnylea, Humber Bay, Mimico NE, The Queensway East, Royal York South West
Mimico NW, The Queensway West, South of Bloor, Kingsway Park South West, Royal York South West
Islington Avenue, Humber Valley Village
West Deane Park, Princess Gardens, Martin Grove, Islington, Cloverdale
Erringate, Bloordale Gardens, Old Burnhamthorpe, Markland Wood
Humberlea, Emery
Weston
Westmount
Kingsview Village, St. Phillips, Martin Grove Gardens, Richview Gardens
Steeles, Silverstone, Humbergate, Jamestown, Mount Olive, Beaumond Heights, Thistletown
Northwest, West Humber - Clairville

RESULTS

- **The Location:**

North York is an incredibly multicultural neighborhood, it has a superb collection of restaurants and there are plenty of great festivals throughout the year that celebrate the diversity there. These festival events are a great opportunity to listen to diverse music, taste cuisine from around the world and learn about another culture. A lot of history has taken place in North York, which means that there are plenty of interesting landmark sites to explore. North York is connected to the city and the rest of the neighborhoods and Toronto by some very convenient public transit links and transportation system. It's home to two of the five major shopping malls in Toronto. North York is a good destination for new immigrants in Canada to stay.

- **Foursquare API:**

This project use Foursquare API as data gathering source because it has database of millions of places, especially their places API which provides the ability to perform location search, location sharing and details about a business. It provides data location with information about venue names, locations, menus, photos and events within an area of interest. After getting the list of neighborhoods, Foursquare API is connected to gather information about venues inside each and every neighborhood. For each neighborhood, the radius is 100 meter. Using credentials of Foursquare API features of nearby places with the number of places per neighborhood parameter be set to 100 and the radius parameter would be set to 500.The data retrieved from Foursquare contained information of venues within a specified distance of the longitude and latitude of the postcodes.

DISCUSSION

- We get a lot of information about the neighborhood in North York. There are 307 unique venues categories in North York and its nearby neighborhood such as shopping mall, bakery, pool, grocery store, pharmacy, restaurant, coffee shop, diner, bus line, gym center, gas station, etc.

- In each neighborhoods, we showed the top 5 venues and and 10 most common venues. For example at Bayview village neighborhood, the top 5 most common venues are gas station, flower shop, park, trail and asian restaurant. At Willowdale, Newtonbrook neighborhood, the top 5 most common venues are korean restaurant. middle east restaurant, coffee shop, cafe and pizza place.

- Looking at the data, we can analyze that there are so much diversity in the North York neighborhood. There are lot of school with great rating for children. The talk about North York is a good recommendation place to move can be supported with these data!

- The results can be checked from the map and the graph and they can be used for people who wants to move to North York to find good housing and good school.

CONCLUSION

- Using k-means cluster algorithm to separate the neighborhood into 10 different clusters and for 103 different lattitude and logitude from dataset, which have very-similar neighborhoods around them.
- The charts results showed particular neighborhood based on average house prices and school rating.
- Folium mapping is a very powerful technique to visualize and to consolidate information and make efficient good analysis and decision.
- This Data Science project showed a practical application to resolve a real life situation that can bring personal and financial impact.
- This project can be improved by making it more accurate to find best house in North York to live with effective living cost.