Final Project Report | Coursera Capstone Project

Finding a Better Place in North York, Toronto

1. Introduction:

The purpose of this Project is to help people in exploring better facilities around their neighborhood. It will help people making smart and efficient decision on selecting great neighborhood out of numbers of other neighborhoods in **North York, Toronto**.

Lots of people are migrating to various states of Canada and this project is for those people who are looking for better neighborhoods. For ease of accessing to Cafe, School, Super market, medical shops, grocery shops, mall, theatre, hospital, like-minded people, etc.

2. Data Section:

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

Will use North York, Toronto dataset which we scrapped from Wikipedia on Week 3. Dataset consisting of latitude and longitude, zip codes.

Foursquare API Data:
We will need data about different venues in different neighbourhoods of that specific borough. In order to gain that information, we will use "Foursquare" locational information. Foursquare is a location data provider with information about all manner of venues and events within an area of interest. Such information includes venue names, locations, menus and even photos. As such, the foursquare location platform will be used as the sole data source since all the stated required information can be obtained through the API.
After finding the list of neighbourhoods, we then connect to the Foursquare API to gather information about venues inside each and every neighbourhood. For each neighbourhood, we have chosen the radius to be 100 meters.
The data retrieved from Foursquare contained information of venues within a specified distance of the longitude and latitude of the postcodes.

The information obtained per venue as follows:
1. Neighborhood
2. Neighborhood Latitude
3. Neighborhood Longitude
4. Venue
5. Name of the venue e.g. the name of a store or restaurant
6. Venue Latitude
7. Venue Longitude
8. Venue Category

IBM Data Science

Made by: Anirudh Jilla

Final Project Report | Coursera Capstone Project

Map of North York:

3. Methodology Section

Clustering Approach: To compare the similarities of two cities, we decided to explore neighborhoods, segment them, and group them into clusters to find similar neighborhoods in a big city like New York and Toronto. To be able to do that, we need to cluster data which is a form of unsupervised machine learning: k-means clustering algorithm.

Using K-Means Clustering Approach:

Building K-Means clustering model

```
In [33]:  # Using K-Means to cluster neighborhood into 3 clusters
          NY_grouped_clustering = NY_grouped.drop('Neighbourhood', 1)
          kmeans = KMeans(n_clusters=3, random_state=0).fit(NY_grouped_clustering)
          kmeans.labels_

Out[33]:  array([0, 0, 0, 0, 0, 0, 2, 0, 0, 2, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
                 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0,
                 0, 0, 0, 0, 0, 0, 0, 0, 2, 0, 1, 0, 0, 0, 0, 0, 2, 0, 0,
                 2, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
                 0, 0, 0, 0, 2, 2, 0, 2])
```

```
In [34]:  neighborhoods_venues_sorted.insert(0, 'Cluster Labels', kmeans.labels_)

          NY_merged = df[df['Borough']=='North York']

          # merge toronto_grouped with toronto_data to add latitude/longitude for each neighborhood
          NY_merged = NY_merged.join(neighborhoods_venues_sorted.set_index('Neighbourhood'), on='Neighbourhood')

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

Out[34]:

	Postal_Code	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	8 C
17	M2H	North York	Hillcrest Village	43.803762	-79.363452	0.0	Fast Food Restaurant	Pharmacy	Tennis Court	Bakery	Sandwich Place	Chinese Restaurant	Restaurant	Re
18	M2J	North York	Fairview, Henry Farm, Oriole	43.778517	-79.346556	0.0	Clothing Store	Coffee Shop	Fast Food Restaurant	Juice Bar	Shoe Store	Baseball Field	Cosmetics Shop	Ja Re
19	M2K	North York	Bayview Village	43.786947	-79.385975	0.0	Bank	Japanese Restaurant	Café	Playground	Chinese Restaurant	Grocery Store	Intersection	
20	M2L	North York	York Mills, Silver Hills	43.757490	-79.374714	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
22	M2N	North York	Willowdale	43.770120	-79.408493	0.0	Coffee Shop	Pizza Place	Ramen Restaurant	Korean Restaurant	Japanese Restaurant	Hotel	Middle Eastern Restaurant	S

IBM Data Science

Made by: Anirudh Jilla

Final Project Report | Coursera Capstone Project

Most Common venues near Neighborhood:

Most common venues nearby

```
In [30]: num_top_venues = 10

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

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))

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

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

neighborhoods_venues_sorted.head()
```

Out[30]:

	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	Pool Hall	Coffee Shop	Breakfast Spot	Fabric Shop	Skating Rink	Lounge	Badminton Court	Latin American Restaurant	Clothing Store	Motorcycle Shop
1	Alderwood, Long Branch	Convenience Store	Pizza Place	Pub	Coffee Shop	Athletics & Sports	Sandwich Place	Gym	Gas Station	Event Space	Fabric Shop
2	Bathurst Manor, Wilson Heights, Downsview North	Park	Coffee Shop	Bank	Pizza Place	Restaurant	Shopping Mall	Diner	Gas Station	Sandwich Place	Sushi Restaurant
3	Bayview Village	Bank	Japanese Restaurant	Café	Playground	Chinese Restaurant	Grocery Store	Intersection	Escape Room	Drugstore	Dry Cleaner
4	Bedford Park, Lawrence Manor East	Italian Restaurant	Coffee Shop	Cosmetics Shop	Sushi Restaurant	Sandwich Place	Pharmacy	Pizza Place	Liquor Store	Bank	Bakery

Work Flow: Using credentials of Foursquare API features of near-by places of the neighborhoods would be mined. Due to http request limitations the number of places per neighborhood parameter would reasonably be set to 100 and the radius parameter would be set to 500.

```
: # Nearby Venues
NY_Venues = getNearbyVenues(names=df['Neighbourhood'],
                            latitudes=df['Latitude'],
                            longitudes=df['Longitude']
                            )
```

```
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
York Mills West
Willowdale
Parkwoods
Don Mills
Don Mills
Bathurst Manor, Wilson Heights, Downsview North
Northwood Park, York University
Downsview
Downsview
```

4. Results Section

Map of Clusters in North York:
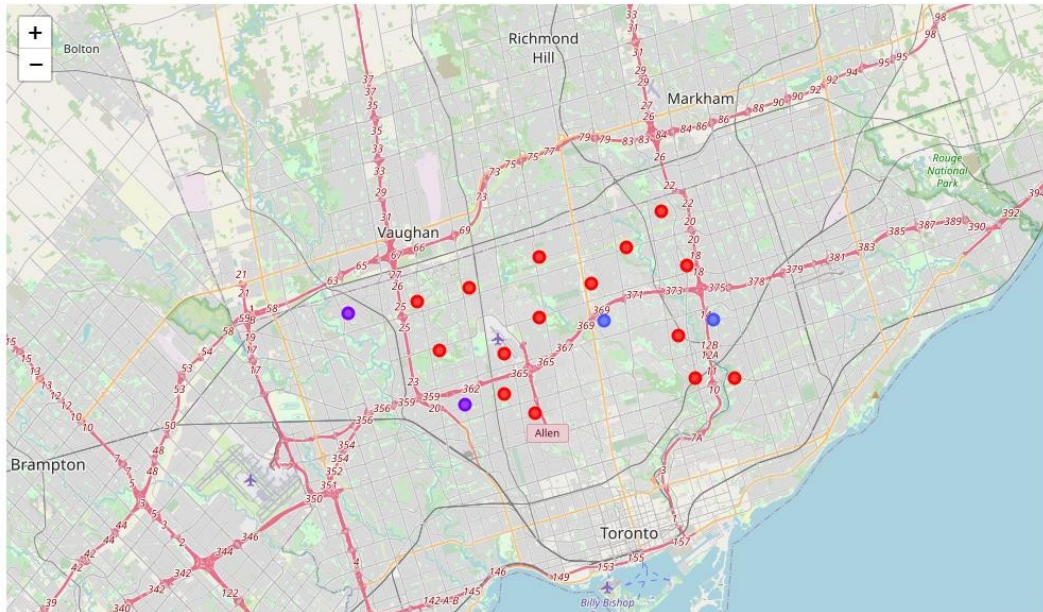
The Location: North York is an eclectic, multicultural district home to the hands-on Ontario Science Centre and the Aga Khan Museum, with exhibits on Islamic culture in a striking modern building. In the area's north, Black Creek Pioneer Village is an 1800s living museum. Sprawling Downsview Park includes a lake, event spaces, and a flea and farmers' market, while Edwards Gardens has a greenhouse, fountains, and botanic gardens

Foursquare API: This project has used Four-square API as its prime data gathering source as it has a database of millions of places, especially their places API which provides the ability to perform location search, location sharing and details about a business.

5. Discussion Section

Problem Which Tried to Solve: The major purpose of this project, is to suggest a better neighborhood in a new city for the person who are shiffting there. Social presence in society in terms of like minded people. Connectivity to the airport, bus stand, city center, markets and other daily needs things nearby.

6. Conclusion Section

In this project, using k-means cluster algorithm I separated the neighborhood into 10(Ten) different clusters and for 103 different lattitude and logitude from dataset, which have very-similar neighborhoods around them. Using the charts above results presented to a particular neighborhood based on average house prices and school rating have been made.

Future Works: This project can be continued for making it more precise in terms to find best house in North York. Best means on the basis of all required things (daily needs or things we need to live a better life) around and also in terms of cost effective.

IBM Data Science

Made by: Anirudh Jilla