The Battle of Neighbourhoods (Week 2)

Introduction

Toronto is the provincial capital of Ontario. With a recorded population of 2,731,571 in 2016, it is the most populous city in Canada and the fourth most populous city in North America. It is a prominent centre for music, theatre, motion picture production, and television production, and is home to the headquarters of Canada's major national broadcast networks and media outlets. Its varied cultural institutions, which include numerous museums and galleries, festivals and public events, entertainment districts, national historic sites, and sports activities, attract over 43 million tourists each year. Toronto is known for its many skyscrapers and high-rise buildings, in particular the tallest free-standing structure in the Western Hemisphere, the CN Tower. Postal codes beginning with M are located within the city of Toronto in the province of Ontario.

Business Problem:

With diverse culture and populous area, the city of Toronto has potential for expansion and buildout. There is window of opportunity for construction companies, businessmen, entrepreneur and other possible developments. Construction companies need site suitable for establishments like gym, sport complex, malls, etc or housing and apartments near workplaces or colleges. Start-up companies look for spot to open their office or shops. These are the some of possible clients or people who would be interested in this project.

In this project, we try to find answers to following questions:

1. What type of venues are available in each neighbourhood in Toronto?
2. List the places not available in neighbourhood and can be build for development.
3. Find area having potential to build a chain of store and type of establishment.
4. Find type or common venues in neighbourhood.
5. Search or locate particular venues in neighbourhood.

Data Section:

Toronto's demographics show that it is a large and ethnically diverse metropolis. There are many venues and of all kinds. Some neighbourhood are populous and congested while other have potential for development.

For the project we need the following data:

- Toronto data that contain lists of boroughs and neighbourhoods along with their latitude and longitude
  - Data Source: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
  - Data Description: Data source contain table with list of Postal Code in the city and the borough and neighbourhood assigned to it.

- List and map of venues and venue category in each neighbourhood
  - Data source: Foursquare API and Python packages
  - Data Description: Foursquare API give list of venues and category. Packages like geopy will give us latitude and longitude of any location while folium helps in visualizing everything on map with labels.

Methodology:

1. Data of Toronto City have already been collected, preprocessed, cleaned and structured in csv file.

	Postal Code	Borough	Neighborhood	Latitude	Longitude
0	M3A	North York	Parkwoods	43.753259	-79.329656
1	M4A	North York	Victoria Village	43.725882	-79.315572
2	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636
3	M6A	North York	Lawrence Manor, Lawrence Heights	43.718518	-79.464763
4	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government	43.662301	-79.389494

2. Group the data and find the number of neighbourhoods present in each borough.

Borough	Postal Code	Neighborhood	Latitude	Longitude
Central Toronto	9	9	9	9
Downtown Toronto	19	19	19	19
East Toronto	5	5	5	5
East York	5	5	5	5
Etobicoke	12	12	12	12
Mississauga	1	1	1	1
North York	24	24	24	24
Scarborough	17	17	17	17
West Toronto	6	6	6	6
York	5	5	5	5

This shows that the North York has highest number of neighbourhoods while Mississauga has least. We will split the data and make separate dataframe for each borough so that it will be easy for us to analyse the data and get insights for our project.

3. List all the nearby venues in the neighbourhoods and store for each borough.
   Find all categories of venue existing and this will give the list of venues not available nearby in the neighbourhood. This list will be great helpful for our clients and for comprehension.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Regent Park, Harbourfront	43.654260	-79.360636	(Roselle Desserts,)	(43.653446723052674,)	(-79.3620167174383,)	Bakery
1	Regent Park, Harbourfront	43.654260	-79.360636	(Tandem Coffee,)	(43.65355870959944,)	(-79.36180945913513,)	Coffee Shop
2	Regent Park, Harbourfront	43.654260	-79.360636	(Cooper Koo Family YMCA,)	(43.65324910177244,)	(-79.35800826343677,)	Distribution Center
3	Regent Park, Harbourfront	43.654260	-79.360636	(Body Blitz Spa East,)	(43.65473505045365,)	(-79.35987433132891,)	Spa
4	Regent Park, Harbourfront	43.654260	-79.360636	(Dominion Pub and Kitchen,)	(43.65691857501867,)	(-79.35896684476664,)	Pub
5	Regent Park, Harbourfront	43.654260	-79.360636	(Corktown Common,)	(43.655617799749734,)	(-79.3562113397429,)	Park
6	Regent Park, Harbourfront	43.654260	-79.360636	(Impact Kitchen,)	(43.65636850543279,)	(-79.35697968750694,)	Restaurant
7	Regent Park, Harbourfront	43.654260	-79.360636	(Morning Glory Cafe,)	(43.653946942635294,)	(-79.36114884214422,)	Breakfast Spot
8	Regent Park, Harbourfront	43.654260	-79.360636	(The Extension Room,)	(43.65331304337331,)	(-79.35972538802777,)	Gym / Fitness Center
9	Regent Park, Harbourfront	43.654260	-79.360636	(The Distillery Historic District,)	(43.65024435658077,)	(-79.35932278633118,)	Historic Site

4. Analyse the common venues in each neighbourhood and their frequency of occurrence in that location will aid in getting the environment of surrounding and localities.
   Eg. Venues in Downtown's neighbourhood
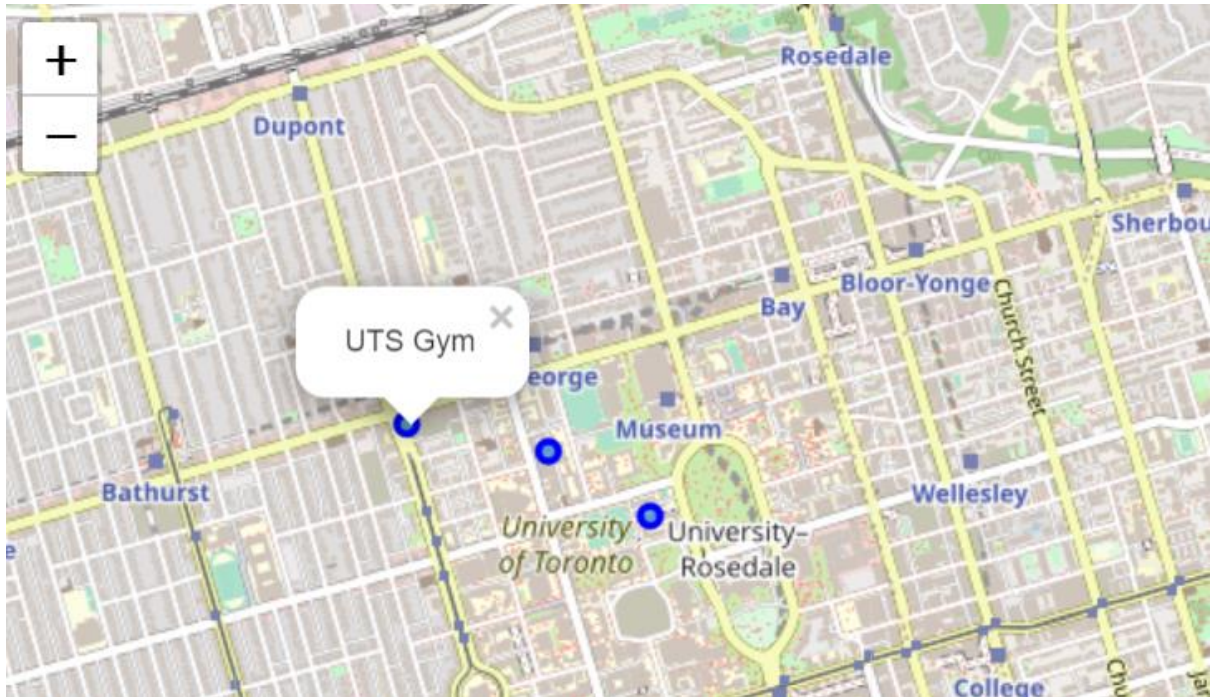
```
----Caledonia-Fairbanks----
           venue  freq
0           Park  0.50
1           Pool  0.25
2  Women's Store  0.25
3            Bar  0.00
4    Coffee Shop  0.00


----Del Ray, Mount Dennis, Keelsdale and Silverthorn----
              venue  freq
0               Bar  0.2
1       Coffee Shop  0.2
2  Convenience Store  0.2
3    Discount Store  0.2
4    Sandwich Place  0.2
```

5. Client have option to query for particular venue in particular location as they like.

Eg. Search result for gym near University of Toronto

Results:

We try to answer to the problem taken earlier:

1. We have table for each borough tabulating each venue along with its address, latitude, longitude, venue category and its neighbourhood detail.
2. We find that 'Field, Farm, River, College Stadium' are some of the categories not present in most of the borough. There are more than 250 categories of venue and most of borough have less than 100 categories.
3. We find that Mississauga and York has least number of venues. All neighbourhoods except some in Downtown Toronto and North York has potential for construction and developments on large scale.
4. We find that there are certain sites present in most of the neighbourhoods like Coffee shop, bar, restaurants, etc.
5. It is easy for a person to query nearby venues entering its exact address.

Conclusion:

The occupancy of any venue highly depends on locality e.g., bar and diner are nearby, café near offices or adjacent stores, etc. However, the categories of venue is unclear and mixed, it complicates, jumbles and gives erroneous and inaccurate result. It is highly recommended to include all similar meaning words in search query.