Capstone Project - The Battle of Neighborhoods

2. Data acquisition and cleaning

2.1 Data sources

Neighbourhood data are from Wikipedia for Postcode, Borough and Neighbourhood.

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

	Postcode	Borough	Neighbourhood
0	M1A	Not assigned	Not assigned\n
1	M2A	Not assigned	Not assigned\n
2	M3A	North York	Parkwoods\n
3	M4A	North York	Victoria Village\n
4	M5A	Downtown Toronto	Harbourfront\n

Geospatial Coordinates data are from the CSV file.

https://cocl.us/Geospatial_data

	Postal Code	Latitude	Longitude
0	M1B	43.806686	-79.194353
1	M1C	43.784535	-79.160497
2	M1E	43.763573	-79.188711
3	M1G	43.770992	-79.216917
4	M1H	43.773136	-79.239476

Amenity data are from Foursquare by API request.

url = 'https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&v={}&ll={},{}&radius={}&limit={}'.format(CLIENT_ID, _SECRET, VERSION, lat, lng, radius, LIMIT)

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Harbourfront, Regent Park	43.65426	-79.360636	Roselle Desserts	43.653447	-79.362017	Bakery
1	Harbourfront, Regent Park	43.65426	-79.360636	Tandem Coffee	43.653559	-79.361809	Coffee Shop
2	Harbourfront, Regent Park	43.65426	-79.360636	Toronto Cooper Koo Family Cherry St YMCA Centre	43.653191	-79.357947	Gym / Fitness Center
3	Harbourfront, Regent Park	43.65426	-79.360636	Body Blitz Spa East	43.654735	-79.359874	Spa
4	Harbourfront, Regent Park	43.65426	-79.360636	Morning Glory Cafe	43.653947	-79.361149	Breakfast Spot

2.2 Data cleaning

Neighbourhood data are available on the web. I got them by using BeautifulSoup and put them into a dataframe. Then read the CSV file with Geospatial Coordinates into another dataframe. As both of the two dataframe have postal code, I can use the postal code as keys to consolidate two dataframes into one dataframe with Neighbourhood and Coordinates. Then I use the Coordinates to call Foursqare API to get Amenity data.