Applied Data Science Capstone project

The Battle of Neighborhoods – Week 1 – Part 2

Data

As part of this project, we would be comparing neighborhoods in New York & Toronto cities. We would need the Boroughs, neighborhoods and latitude & longitude coordinates of each neighborhood in these two cities.

We will use Four Square api to explore each neighborhood and will be getting top 100 venues within 500m radius

New York city data

New York has a total of 5 boroughs and 306 neighborhoods. We need a dataset that contains the 5 boroughs and the neighborhoods that exist in each borough as well as the latitude and longitude coordinates of each neighborhood.

This dataset is available in the below link

https://cocl.us/new_york_dataset

We will be using the below command to get the dataset

!wget -q -O 'newyork_data.json' https://cocl.us/new_york_dataset

We will transform the data that we get from the above link into a pandas data frame and then use Foursquare api to get venue details for each neighborhood

Toronto city data

We will scrape the below Wikipedia page using BeautifulSoup api in order to obtain Borough, Neighborhood and postal codes in Toronto and then transform the data into a pandas data frame

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

We will use the csv file present in the below link to get latitude & longitude of each neighborhood of Toronto using Postal Code as the Key

http://cocl.us/Geospatial_data

Foursquare API

We will use the below Four Square API URL to fetch the top 100 venues within a radius of 500m for each neighborhood

https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&v={}&ll={},{}&radius={}&limit={}