

## Data :

### Requirements:

The required data for this project is the Toronto dataset and the co-ordinates of latitudes and longitudes. For the given scenarios we need the restaurants data set of Toronto and the location of those restaurants. Similarly, we need the details of the restaurant such as what type of food it serves such as famous cuisine etc,

### Data acquisition:

We can collect the data from [https://geo.nyu.edu/catalog/nyu\\_2451\\_34572](https://geo.nyu.edu/catalog/nyu_2451_34572) .

This dataset has neighborhood which has a total of 5 boroughs and 306 neighborhoods. We can get the location such as longitude and latitude co-ordinates of the restaurants by using “geopy library”. After removing unwanted data the dataset looks like this:

	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
5	M9A	Etobicoke	Islington Avenue	43.667856	-79.532242
6	M1B	Scarborough	Malvern, Rouge	43.806686	-79.194353
7	M3B	North York	Don Mills	43.745906	-79.352188
8	M4B	East York	Parkview Hill, Woodbine Gardens	43.706397	-79.309937
9	M5B	Downtown Toronto	Garden District, Ryerson	43.657162	-79.378937
10	M6B	North York	Glencairn	43.709577	-79.445073
11	M9B	Etobicoke	West Deane Park, Princess Gardens, Martin Grov...	43.650943	-79.554724
12	M1C	Scarborough	Rouge Hill, Port Union, Highland Creek	43.784535	-79.160497
13	M3C	North York	Don Mills	43.725900	-79.340923

We should append the two datasets such as the Toronto dataset and the location dataset which we got by using geopy library. So this dataset contains the Postal Code, Borough, Neighborhood and geographical coordinates.

### Foursquare API:

We use Foursquare API to fetch the nearest venue so that we can use it to form a cluster. In this project I will be using Foursquare API to find the venues within 500 meters radius.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Regent Park, Harbourfront	43.65426	-79.360636	Roselle Desserts	43.653447	-79.362017	Bakery
1	Regent Park, Harbourfront	43.65426	-79.360636	Tandem Coffee	43.653559	-79.361809	Coffee Shop
2	Regent Park, Harbourfront	43.65426	-79.360636	Morning Glory Cafe	43.653947	-79.361149	Breakfast Spot
3	Regent Park, Harbourfront	43.65426	-79.360636	Cooper Koo Family YMCA	43.653249	-79.358008	Distribution Center
4	Regent Park, Harbourfront	43.65426	-79.360636	Body Blitz Spa East	43.654735	-79.359874	Spa