

Exploring the neighbourhoods in Toronto, Canada.

Data acquisition

There are different sources from which I will collect the data for different purpose.

1. List of postal Codes for Canada :-

- I will get the postal code of the neighbourhoods in Canada from Wikipedia.
- Link - https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

2. Geographical Co-ordinates :-

- I will use a csv file which consist latitude and longitude of the neighbourhoods in Canada.
- We can use geocoder for the same purpose, but it is not persistent sometime. So I choose the csv file instead using geocoder.
- Link for csv - http://cocl.us/Geospatial_data

3. Fetching Details of the venue :

- I will use Foursquare API for fetching the details and location of the venues.
- I will use venue ratings as a threshold. And finally visualize using Folium.

Data Cleaning

Cleaning the Postal Code data

- The dataframe will consist of three columns: PostalCode, Borough, and Neighborhood
- Only process the cells that have an assigned borough. Ignore cells with a borough that is **Not assigned**.
- More than one neighborhood can exist in one postal code area. For example, in the table on the Wikipedia page, you will notice that **M5A** is listed twice and has two neighborhoods: **Harbourfront** and **Regent Park**. These two rows will be combined into one row with the neighborhoods separated with a comma as shown in **row 11** in the above table.
- If a cell has a borough but a **Not assigned** neighborhood, then the neighborhood will be the same as the borough.

	PostalCode	Borough	Neighborhood
0	M5G	Downtown Toronto	Central Bay Street
1	M2H	North York	Hillcrest Village
2	M4B	East York	Parkview Hill, Woodbine Gardens
3	M1J	Scarborough	Scarborough Village
4	M4G	East York	Leaside
5	M4M	East Toronto	Studio District
6	M1R	Scarborough	Wexford, Maryvale
7	M9V	Etobicoke	South Steeles, Silverstone, Humbergate, Jamest...
8	M9L	North York	Humber Summit
9	M5V	Downtown Toronto	CN Tower, King and Spadina, Railway Lands, Har...
10	M1B	Scarborough	Malvern, Rouge
11	M5A	Downtown Toronto	Regent Park, Harbourfront

DataFrame of Postal Codes in Canda.

Adding geographical Co-ordinates to the Neighbourhoods

	PostalCode	Borough	Neighborhood	Latitude	Longitude
0	M5G	Downtown Toronto	Central Bay Street	43.657952	-79.387383
1	M2H	North York	Hillcrest Village	43.803762	-79.363452
2	M4B	East York	Parkview Hill, Woodbine Gardens	43.706397	-79.309937
3	M1J	Scarborough	Scarborough Village	43.744734	-79.239476
4	M4G	East York	Leaside	43.709060	-79.363452
5	M4M	East Toronto	Studio District	43.659526	-79.340923
6	M1R	Scarborough	Wexford, Maryvale	43.750071	-79.295849
7	M9V	Etobicoke	South Steeles, Silverstone, Humbergate, Jamest...	43.739416	-79.588437
8	M9L	North York	Humber Summit	43.756303	-79.565963
9	M5V	Downtown Toronto	CN Tower, King and Spadina, Railway Lands, Har...	43.645711	-79.392732
10	M1B	Scarborough	Malvern, Rouge	43.806686	-79.194353
11	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636

DataFrame of geographical co-ordinates with neighbourhood.

Get the location and other details of the venue using foursquare API

	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
Berczy Park	56	56	56	56	56	56
Brockton, Parkdale Village, Exhibition Place	23	23	23	23	23	23
Business reply mail Processing Centre, South Central Letter Processing Plant Toronto	16	16	16	16	16	16
CN Tower, King and Spadina, Railway Lands, Harbourfront West, Bathurst Quay, South Niagara, Island airport	18	18	18	18	18	18

DataFrame of venue details.

Segmenting and Clustering

Clustering the neighbourhoods in Toronto.

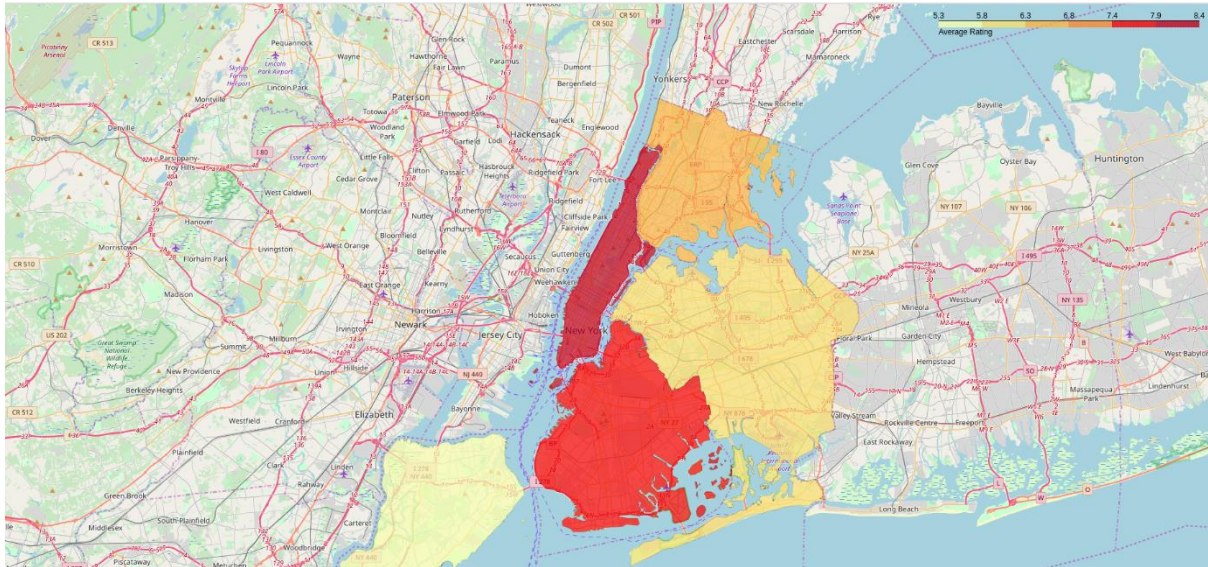
```
# set number of clusters
kclusters = 5

toronto_grouped_clustering = toronto_grouped.drop('Neighborhood', 1)

# run k-means clustering
kmeans = KMeans(n_clusters=kclusters, random_state=0).fit(toronto_grouped_clustering)

# check cluster labels generated for each row in the dataframe
kmeans.labels_[0:10]
```

Visualising Choropleth map based on average rating of Indian Restaurants



Note : - This is Choropleth Map of New York not Toronto

Then Concluding the results.

Limitations

- The ranking is purely on basis of rating of restaurants
- The accuracy of data depends purely depends on the data provided by Foursquare API.

By

Siddharaj Maramwar