

Date:25/05/2020

Final Report | Capstone Project – The Battle of Neighborhoods Finding Common Places in Toronto

1. Introduction:

The purpose of this Project is to help people avoid common places for their safety and family due to Covid-19 crisis. It will help people making smart and efficient decision on selecting neighborhood places out of numbers of other neighborhoods in Toronto.

Lots of people movement around the common places might be infectious for persons to travel who might get infected and spread the virus to friends & family members. And these common places are at high risk and prone to spread infection as cluster. And its good to avoid such places to have or stop Covid-19 virus infection.

This Project aim to create an analysis of most visited common places for people to take precautionary steps before passing through or avoid such places to buy their groceries and vegetables.

It will help people to get awareness of the area and neighborhood before moving out for their daily chores.

2. Data Section

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

Will use Toronto,Scarborough dataset which we scrapped from wikipedia on Week 3. Dataset consisting of latitude and longitude, zip codes.

Foursquare API Data:

Foursquare API Data

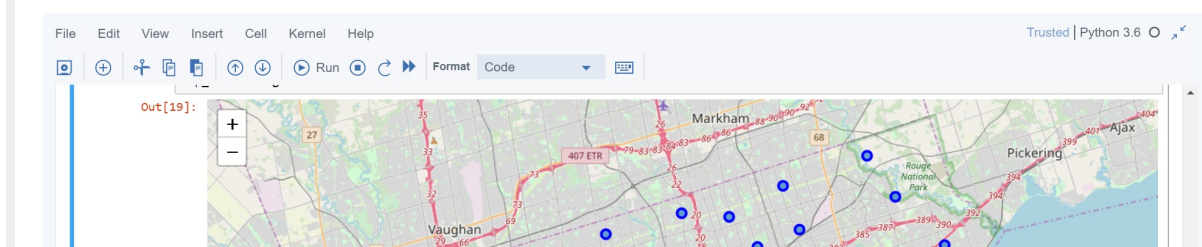
We will need data about different venues in and around the neighborhoods for identifying the common places. In order to gain that information we will use "Foursquare" locational information. Foursquare is a location data provider with information about all manner of venues and events within an area of interest. Such information includes venue names, locations, menus and even photos. As such, the foursquare location platform will be used as the sole data source since all the stated required information can be obtained through the API.

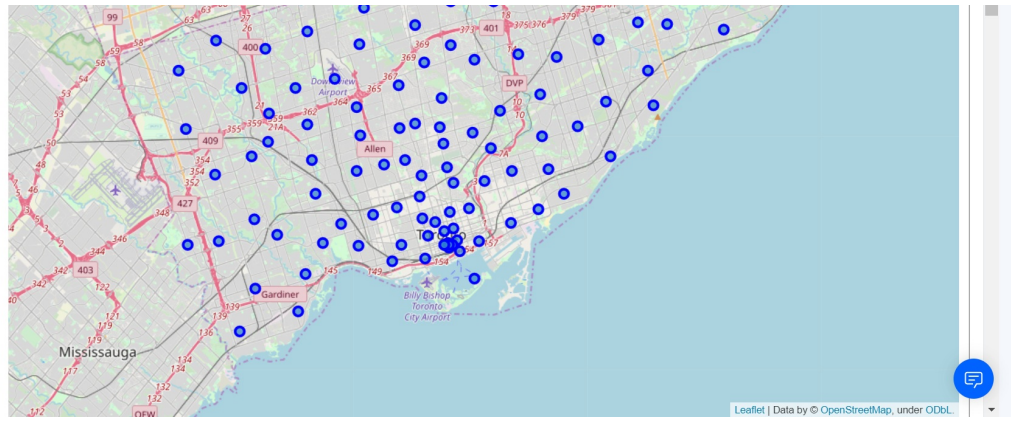
After finding the list of neighborhoods, we then connect to the Foursquare API to gather information about venues inside each and every neighborhood.

The data retrieved from Foursquare contained information of venues within a specified distance of the longitude and latitude of the postcodes. The information obtained per venue as follows:

1. Neighborhood
2. Neighborhood Latitude
3. Neighborhood Longitude
4. Venue
5. Name of the venue e.g. the name of a store or restaurant
6. Venue Latitude
7. Venue Longitude
8. Venue Category

Map of Toronto





3. Methodology Section

Clustering Approach:

To explore the neighborhoods, segment them, and group them into clusters of common places most visited by people in Toronto. And to be able to do that, we need to cluster data which is a form of unsupervised machine learning: k-means clustering algorithm.

Most Common venues near Neighborhood

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Berczy Park	Coffee Shop	Cocktail Bar	Café	Beer Bar	Bakery	Seafood Restaurant	Cheese Shop	Restaurant	Pub	Farmers Market
1	Brockton, Parkdale Village, Exhibition Place	Café	Breakfast Spot	Nightclub	Coffee Shop	Climbing Gym	Burrito Place	Italian Restaurant	Intersection	Stadium	Restaurant
2	Business reply mail Processing Centre, South C...	Yoga Studio	Pizza Place	Smoke Shop	Brewery	Spa	Burrito Place	Farmers Market	Fast Food Restaurant	Restaurant	Light Rail Station
3	CN Tower, King and Spadina, Railway Lands, Har...	Airport Lounge	Airport Service	Airport Terminal	Harbor / Marina	Sculpture Garden	Coffee Shop	Boat or Ferry	Boutique	Bar	Airport Gate
4	Central Bay Street	Coffee Shop	Italian Restaurant	Sandwich Place	Café	Burger Joint	Ice Cream Shop	Japanese Restaurant	Thai Restaurant	Bar	Bubble Tea Shop

Using K-Means Clustering Approach

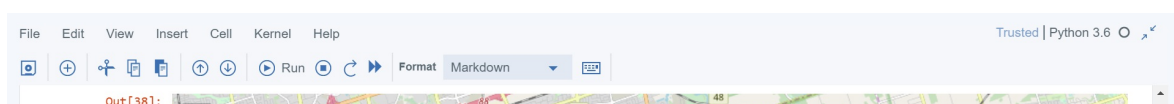
	Postalcode	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue
37	M4E	East Toronto	The Beaches	43.676357	-79.293031	0	Trail	Pub	Health Food Store	Distribution Center	Department Store	Dessert Shop	Dim Sum Restaurant	
41	M4K	East Toronto	The Danforth West, Riverdale	43.679557	-79.352188	0	Greek Restaurant	Coffee Shop	Italian Restaurant	Ice Cream Shop	Furniture / Home Store	Restaurant	Bubble Tea Shop	
42	M4L	East Toronto	India Bazaar, The Beaches West	43.668999	-79.315572	0	Fast Food Restaurant	Steakhouse	Burrito Place	Pub	Restaurant	Intersection	Fish & Chips Shop	Re
43	M4M	East Toronto	Studio District	43.659526	-79.340923	0	Café	Coffee Shop	Gastropub	Bakery	Brewery	American Restaurant	Yoga Studio	Re
44	M4N	Central Toronto	Lawrence Park	43.728020	-79.388790	3	Park	Dim Sum Restaurant	Bus Line	Swim School	Event Space	Electronics Store	Eastern European Restaurant	

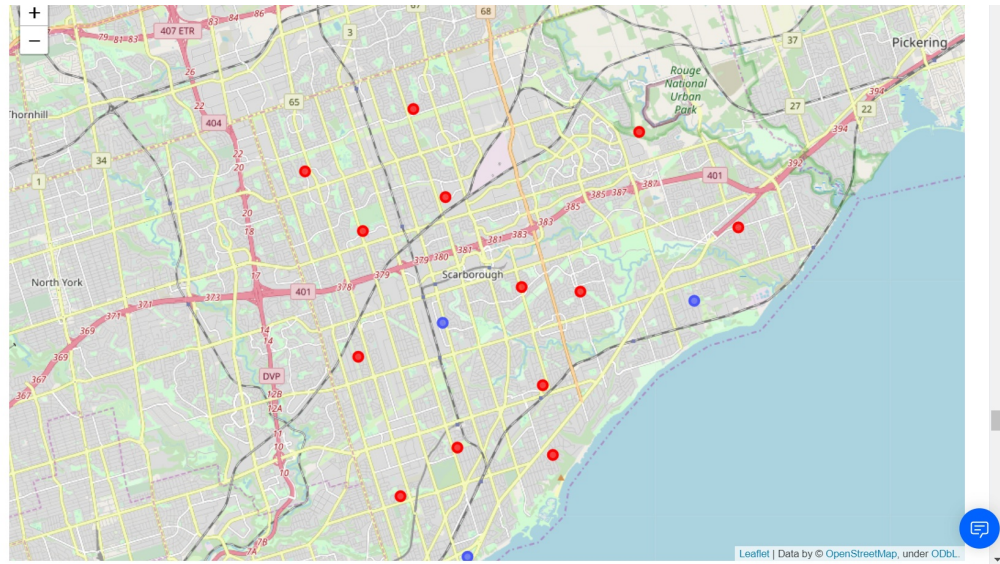
Work Flow:

Using credentials of Foursquare API features of near-by places of the neighborhoods would be mined. Due to http request limitations the number of places per neighborhood parameter would reasonably be set to 100 and the radius parameter would be set to 500.

4. Results Section

Map of Clusters in Toronto





The Location:

Toronto area in Canada has most people movement in and around the city to get their daily needs to be transported or purchase the groceries and vegetables. Due Covid-19 crisis many governments asking people not to move around the city to avoid the Covid-19 infection spread. This has become hot topic to restrict the people movement in common places and had to shift essential goods or stores to other places in order to maintain social distancing.

Foursquare API:

This project have used Four-square API as its prime data gathering source as it has a database of millions of places, especially their places API which provides the ability to perform location search, location sharing and details about a business.

5. Discussion Section

Problem Which Tried to Solve:

The major purpose of this project, is to suggest or identify common places which has more people movement in and around the neighborhood city prone to Covid-19 infection spread. Social distancing and avoiding the common places such as metros, airport, bus stand, beaches, markets and other more people movement areas. Identify and sort top 10 places prone to Covid-19 infection spread.

6. Conclusion Section

In this project, using k-means cluster algorithm I separated the neighborhood into 10(Ten) different clusters and for 103 different latitude and longitude from dataset, which have very-similar neighborhoods around them. Using the above results presented to a particular neighborhood based on common places which can be avoided by the public during the Covid-19 crisis.

I feel rewarded with the efforts and believe this course with all the topics covered is well worthy of appreciation. This project has shown me a practical application to resolve a real situation that has impacting personal and financial impact using Data Science tools. During pandemic situation this is very powerful technique to consolidate information on common places and make the analysis and take better decision.

Future Works:

This project can be continued for making it more precise in terms categories which will be easy to find alternatives in Toronto. This will fulfill their daily needs or things we need to live a better life during the Pandemic crisis.

Libraries Which are Used to Develop the Project:

Pandas: For creating and manipulating dataframes.

Folium: Python visualization library would be used to visualize the neighborhoods cluster distribution of using interactive leaflet map.

Scikit Learn: For importing k-means clustering.

JSON: Library to handle JSON files.

XML: To separate data from presentation and XML stores data in plain text format.

Geocoder: To retrieve Location Data.

Beautiful Soup and Requests: To scrap and library to handle http requests.

Matplotlib: Python Plotting Module.

In []:

In []: