New Restaurant - Toronto

08 March 2020 01:34

Background

Toronto City as the provincial capital of Ontario and the most populous city in Canada, with a population of 2,731,571 in 2016. Current to 2016, the Toronto census metropolitan area (CMA), of which the majority is within the Greater Toronto Area (GTA), held a population of 5,928,040, making it Canada's most populous CMA. The city is the anchor of the Golden Horseshoe, an urban agglomeration of 9,245,438 people (as of 2016) surrounding the western end of Lake Ontario. Toronto is an international center of business, finance, arts, and culture, and is recognized as one of the most multicultural and cosmopolitan cities in the world.

Problem Description

Now let me explain the context of this Capstone project through a scenario. Say you live on the west side of the city of Toronto in Canada. You love your neighbourhood, mainly because of all the great amenities and other types of venues that exist in the neighbourhood, such as gourmet fast food joints, pharmacies, parks, graduate schools and so on.
The aim of this report is to study and analyse the neighbourhoods of Toronto city and group them into similar clusters and, to analyse those clusters to gather meaningful information.
This information provided by this report would be useful for people who are interested in opening the new restaurant. We shall see the picture of the existing situation which will tell in what kind of restaurant people like to visit and maybe to define the missing values-gaps which can be the opportunity to run the new business in a chosen neighbourhood.

Data Description

Data sources used for the analysis:

- **Toronto Neighbourhood Data**: The following Wikipedia page was scraped to pull out the necessary information: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

  The information obtained i.e. the table of postal codes was transformed into a pandas data frame for further analysis.

- **Coordinate data for each Neighbourhood in Toronto**: The following csv file gave us the geographical coordinates of each postal code: http://cocl.us/Geospatial_data

- **Using Foursquare API for identifying all venues for each neighbourhood.**

Methodology

- **Scrape the Wikipedia page and gathering data into a Pandas dataframe**
  To start with our analysis, we used the BeautifulSoup package to transform the data in the table on the Wikipedia page into the below pandas dataframe.

	Postalcode	Borough	Neighborhood
0	M3A	North York	Parkwoods
1	M4A	North York	Victoria Village
2	M5A	Downtown Toronto	Harbourfront
3	M6A	North York	Lawrence Heights, Lawrence Manor
5	M7A	Downtown Toronto	Queen's Park

We also fetched the coordinate data for all the neighbourhoods in Toronto using the csv file and put it into a dataframe.

	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

Next, we combine both the dataframes i.e. adding the coordinate data to the original dataframe.

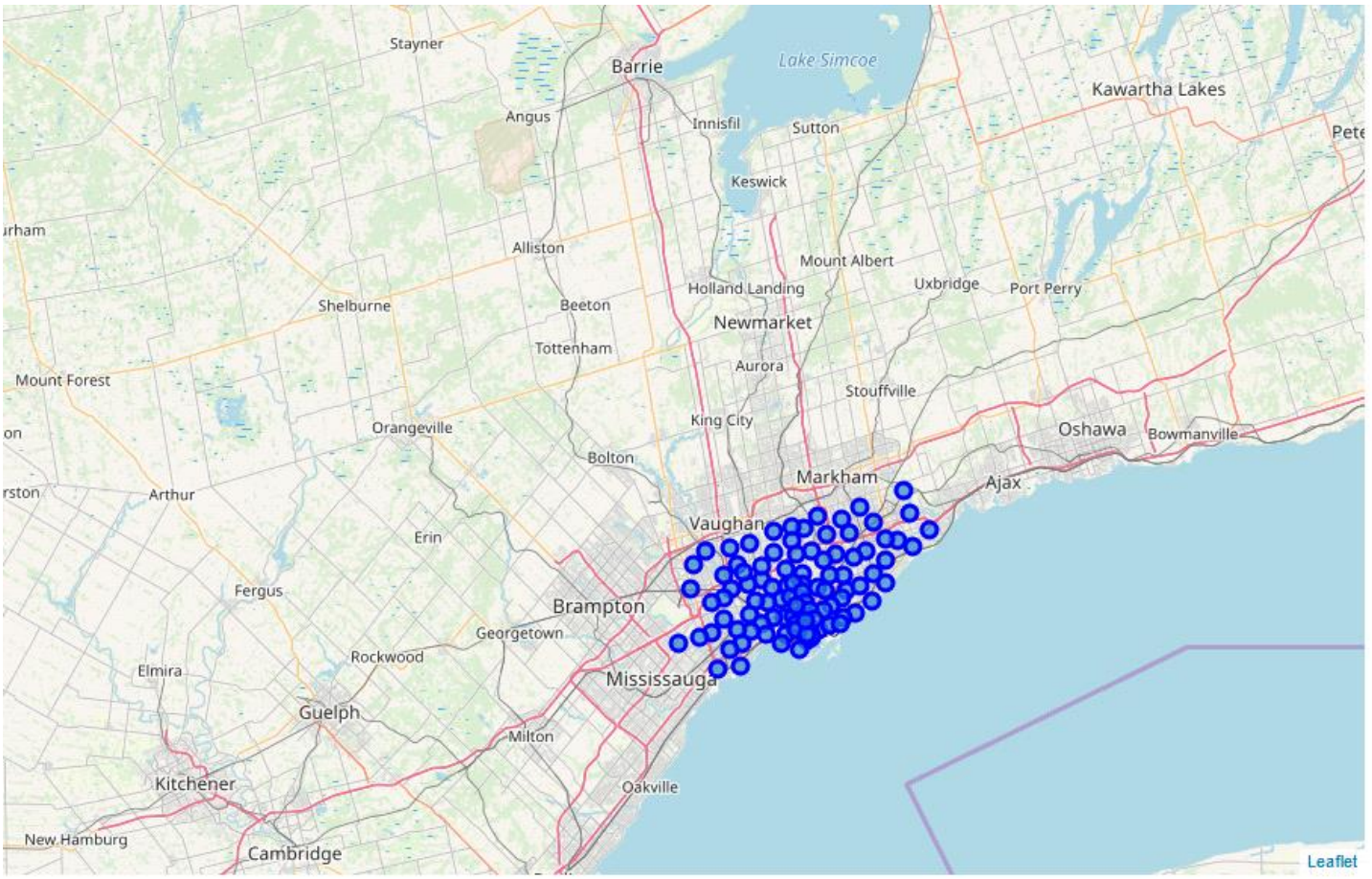
	Postalcode	Latitude	Longitude	Borough	Neighborhood
0	M1B	43.806686	-79.194353	Scarborough	Rouge, Malvern
1	M1C	43.784535	-79.160497	Scarborough	Highland Creek, Rouge Hill, Port Union
2	M1E	43.763573	-79.188711	Scarborough	Guildwood, Morningside, West Hill
3	M1G	43.770992	-79.216917	Scarborough	Woburn
4	M1H	43.773136	-79.239476	Scarborough	Cedarbrae

Generating a map of Toronto and plotting the Neighbourhood data on it

We first filter the data to find boroughs containing the word “Toronto”,

	Postalcode	Borough	Neighborhood	Latitude	Longitude
37	M4E	East Toronto	The Beaches	43.676357	-79.293031
41	M4K	East Toronto	The Danforth West, Riverdale	43.679557	-79.352188
42	M4L	East Toronto	The Beaches West, India Bazaar	43.668999	-79.315572
43	M4M	East Toronto	Studio District	43.659526	-79.340923
44	M4N	Central Toronto	Lawrence Park	43.728020	-79.388790

We then use the python **folium** library to visualize geographic details of Toronto and its boroughs. I created a map of Toronto with boroughs superimposed on top using the latitude and longitude values to get the visual as below:



Results :

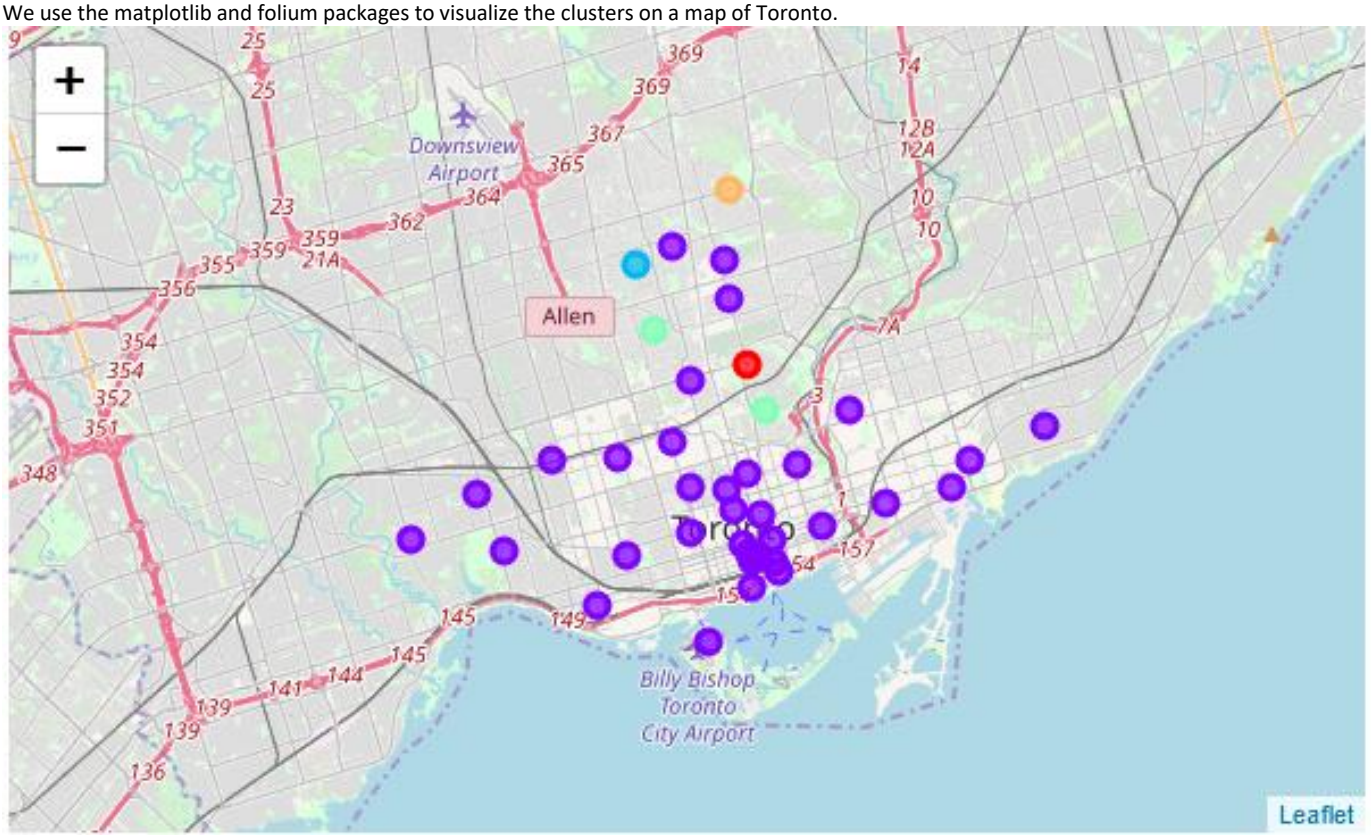
Adding the Cluster Labels to the Venue Data

The below table depicts the clustered data along with the top 10 most common venues in that cluster.

	Cluster Labels	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
0	1	Adelaide, King, Richmond	Coffee Shop	Thai Restaurant	Bar	Restaurant	Café	Steakhouse
1	1	Berczy Park	Coffee Shop	Cocktail Bar	French Restaurant	Cheese Shop	Café	Restaurant
2	1	Brockton, Exhibition Place, Parkdale Village	Café	Coffee Shop	Breakfast Spot	Gym	Convenience Store	Restaurant
3	1	Business Reply Mail Processing Centre 969 Eastern	Yoga Studio	Auto Workshop	Park	Pizza Place	Recording Studio	Restaurant
4	1	CN Tower, Bathurst Quay, Island airport, Harbo...	Airport Lounge	Airport Service	Sculpture Garden	Boat or Ferry	Airport	Airport Food Court

Visualizing the resulting Clusters

We use the matplotlib and folium packages to visualize the clusters on a map of Toronto.



Discussion :

What we can see is such analysis are the common types of restaurants within various

neighbourhoods in different parts of the city. This should help us decide how people like to spend the time, what to eat and maybe bring and idea what could still missing and has chance to bring new value to the people leaving in the particular neighbourhood.

- **Conclusion:**

Like seen in the example above, data was used to cluster neighbourhoods in Toronto based on the most common venues in those neighbourhoods.