BATTLE OF NEIGHBORHOODS- TORONTO

Introduction:

Toronto is the capital of Ontario and is the most populous city in Canada. Toronto is the most ethically diverse city in Canada thereby it has unique food culture. Toronto's landscape is distinctly different than the rest of the country. There are various restaurants of local cuisine and many restaurants of various international cuisines such as Italian, Indian, Chinese, Mexican, Thai etc. According to the "The Canadian Restaurant Industry Landscape" there are more European menu type restaurants in Toronto than in the rest of Canada, making up 5.3 percent of the Toronto landscape, compared to 3.2 nationwide. According to the study 9.6 percent restaurants in Toronto are Asian, whereas Asian restaurants only account for 6.6 percent nationwide. One of the USA's favorite menu type 'hamburger,' does not have as strong of a presence in Canada and is even less predominant is Toronto. As a nation, approximately 6.1 percent of Canadian restaurants serve a 'hamburger' menu type. So we can say Toronto is the best place for foodies.

Problem description:

Let's say I travel to different places in Toronto and it is very difficult to me to identify the best restaurants in that area. In these situations, we need to find the right place to eat with reasonable cost. So the possible analysis can be done on these:
1.   Which is the most nearest restaurant?
2.   Which is the most visited restaurant?
3.   What is the most famous cuisine in that area?
So in this project I will be analyzing the Toronto data to find the solutions of the above problem statements.

Target Audience:

A visitor who wants to find the best restaurant nearby

 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 . 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	Longitud
0	M3A	North York	Parkwoods	43.753259	-79.32965
1	M4A	North York	Victoria Village	43.725882	-79.31557
2	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.36063
3	M6A	North York	Lawrence Manor, Lawrence Heights	43.718518	-79.46476
4	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government	43.662301	-79.38949
5	M9A	Etobicoke	Islington Avenue	43.667856	-79.53224
6	M1B	Scarborough	Malvern, Rouge	43.806686	-79.19435
7	M3B	North York	Don Mills	43.745906	-79.35218
8	M4B	East York	Parkview Hill, Woodbine Gardens	43.706397	-79.30993
9	M5B	Downtown Toronto	Garden District, Ryerson	43.657162	-79.37893
10	M6B	North York	Glencairn	43.709577	-79.44507
11	M9B	Etobicoke	West Deane Park, Princess Gardens, Martin Grov...	43.650943	-79.55472
12	M1C	Scarborough	Rouge Hill, Port Union, Highland Creek	43.784535	-79.16049
13	M3C	North York	Don Mills	43.725900	-79.34092

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	Ven Catego
0	Regent Park, Harbourfront	43.65426	-79.360636	Roselle Desserts	43.653447	-79.362017	Bak
1	Regent Park, Harbourfront	43.65426	-79.360636	Tandem Coffee	43.653559	-79.361809	Coffee Sh
2	Regent Park, Harbourfront	43.65426	-79.360636	Morning Glory Cafe	43.653947	-79.361149	Breakfast S
3	Regent Park, Harbourfront	43.65426	-79.360636	Cooper Koo Family YMCA	43.653249	-79.358008	Distributi Cen
4	Regent Park, Harbourfront	43.65426	-79.360636	Body Blitz Spa East	43.654735	-79.359874	S

Methodology:

Exploratory analysis:

For this analysis we need to scrap the data from the following link https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M .

The data which is acquired consists of the neighborhoods, boroughs of the city Toronto. I am using Beautiful Soup to get the data from the given Wikipedia link. The acquired dataset is in the following form:

	Postal Code	Borough	Neighborhood
0	M1A	Not assigned	NaN
1	M2A	Not assigned	NaN
2	M3A	North York	Parkwoods
3	M4A	North York	Victoria Village
4	M5A	Downtown Toronto	Regent Park, Harbourfront

Since we are interested in the data whose Neighborhood is assigned we should vomit the records whose Neighborhood is not defined.

Then the dataset looks like this:

	Postal Code	Borough	Neighborhood
0	M3A	North York	Parkwoods
1	M4A	North York	Victoria Village
2	M5A	Downtown Toronto	Regent Park, Harbourfront
3	M6A	North York	Lawrence Manor, Lawrence Heights
4	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government

After this we need to append the respective latitude and longitude coordinates to to the neighborhoods. This can be achieved through "geopy library"

Then the resulting dataset looks like this:

	Postal Code	Borough	Neighborhood	Latitude	Longitude
0	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636
1	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government	43.662301	-79.389494
2	M5B	Downtown Toronto	Garden District, Ryerson	43.657162	-79.378937
3	M5C	Downtown Toronto	St. James Town	43.651494	-79.375418
4	M4E	East Toronto	The Beaches	43.676357	-79.293031

Now using Folium the data in the dataframe can be visualized :
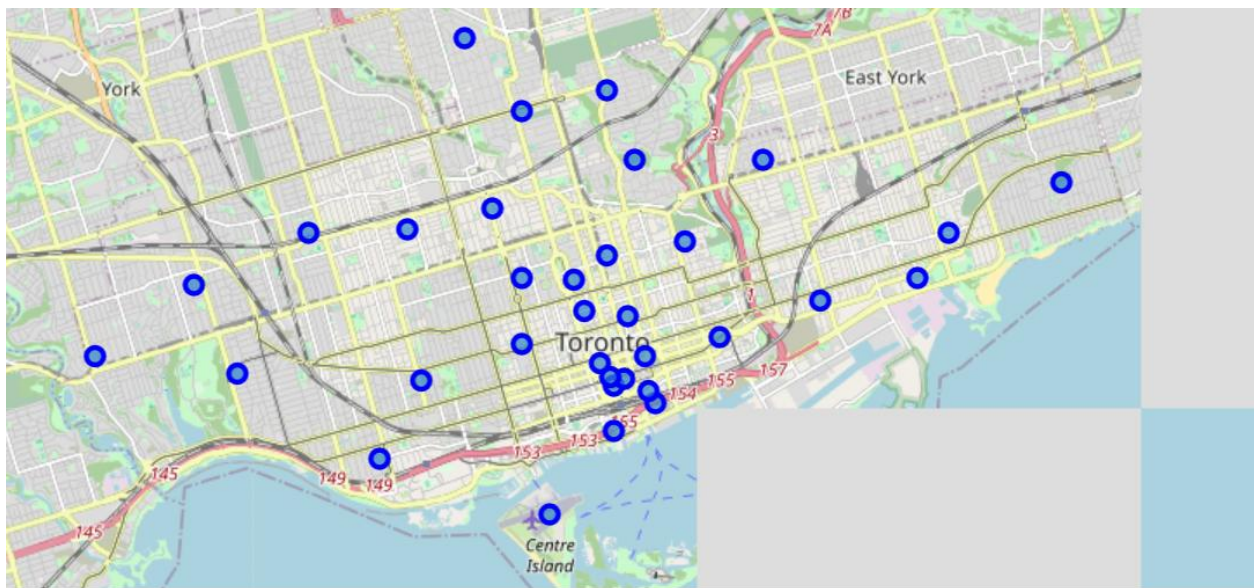
Inferential Analysis:

By using K-means clustering we can divide the area into 'n' number of clusters. Here I have taken n=5.

Result:

By using Foursquare we can fetch the list of venues in each cluster.

The list of most common venues in the cluster 0 are as follows:

	Borough	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	9th Most Common Venue	10th Most Common Venue
0	Downtown Toronto	0	Coffee Shop	Bakery	Pub	Park	Breakfast Spot	Restaurant	Café	Theater	Yoga Studio	Mexican Restaurant
1	Downtown Toronto	0	Coffee Shop	Sushi Restaurant	Gym	Diner	Park	Mexican Restaurant	Japanese Restaurant	Italian Restaurant	Hobby Shop	Wings Joint
2	Downtown Toronto	0	Clothing Store	Coffee Shop	Middle Eastern Restaurant	Restaurant	Bubble Tea Shop	Café	Japanese Restaurant	Italian Restaurant	Cosmetics Shop	Bookstore
3	Downtown Toronto	0	Coffee Shop	Café	Cocktail Bar	Gastropub	American Restaurant	Gym	Beer Bar	Italian Restaurant	Lingerie Store	Department Store
5	Downtown Toronto	0	Coffee Shop	Cocktail Bar	Bakery	Cheese Shop	Café	Restaurant	Beer Bar	Seafood Restaurant	French Restaurant	Pub

Since we are more interested in finding a restaurant, the most visited in Downtown Toronto is Sushi Restaurant. So the most famous cuisine in that area is "Japanese".

We have several coffee shops nearby in that area.

Discussion :

Before concluding the most famous venue nearby user ratings can also be included for the better user experience.

Conclusion:

The most visited restaurant near by is "Sushi restaurant" which is Japanese Cuisine.