

# Data Science - Capstone Project: Burger restaurant data analysis of Toronto city

## A. Introduction

### A.1. Description & Business problem

Toronto is the largest city in Canada. In 2017, sales of the Canadian restaurant industry amounted to approximately 85 billion Canadian dollars, up from 68.4 billion the previous year.

Although certain dishes may be identified as "Canadian" due to the ingredients used or the origin of its inception, an overarching style of Canadian cuisine is more difficult to define. Some Canadians such as the former Canadian believe that Canadian cuisine is a collage of dishes from the cuisines of other cultures. Clark himself has been paraphrased to have noted: "Canada has a cuisine of cuisines."

While the immense size of Canada and diversity of its inhabitants compounds the difficulty in identifying a specific Canadian food identity, Hersch Jacobs acknowledges that the lack of a hegemonic definition does not preclude the existence of a Canadian cuisine.

Restaurant sales in Canada have grown over the past five years from 61 billion Canadian dollars in 2010 to 85 billion Canadian dollars in 2017. To open a new restaurant, there are so many things that needs to be considered such as its location, finance, naming, menus and so many other things. Location being a major factor,

As part of this assignment, I will focus on the Burger restaurant businesses operating around the neighborhoods of Toronto. The first step in opening a new restaurant is to decide what type of restaurant it is going to be, whether a fine-dining restaurant, a casual one or a specific cuisine based. We will be focusing on the burger restaurants that are being present in a neighborhood and see if any of it suits as a good option to open a new burger restaurant. It too depends on what kind of burger cuisine is popular in that area.

## B. Target Audience:

This assignment aims in recommending a suitable location for stakeholders who are planning to open a new restaurant in the neighborhoods of Toronto.

## C. Data Description

We will use the neighborhood data of Toronto by scraping the wikipedia page of Canada [https://en.wikipedia.org/wiki/List\\_of\\_postal\\_codes\\_of\\_Canada:\\_M\\_.1](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M_.1)

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

The latitude and longitude of each neighborhood was defined. To get this we will be using the csv file that has the geographical coordinates of each postal code [http://cocl.us/Geospatial\\_data](http://cocl.us/Geospatial_data). This link is downloaded as 'Geospatial\_Coordinates.csv'. The data is stored into a pandas 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

The data frame will be merged with our postal code/neighborhood data frame to get a combined version of data. After that, we will filter out only neighbourhoods that are part of Toronto.

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

The data above used to fetch the restaurants in the neighborhood within 500 meters. The data will look like as below:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Regent Park, Harbourfront	43.654260	-79.360636	Impact Kitchen	43.656369	-79.356980	Restaurant
1	Regent Park, Harbourfront	43.654260	-79.360636	Cluny Bistro & Boulangerie	43.650565	-79.357843	French Restaurant
2	Regent Park, Harbourfront	43.654260	-79.360636	El Catrin	43.650801	-79.358820	Mexican Restaurant
3	Queen's Park, Ontario Provincial Government	43.662301	-79.389494	Nando's	43.661728	-79.386391	Portuguese Restaurant
4	Queen's Park, Ontario Provincial Government	43.662301	-79.389494	Mercatto	43.660391	-79.387664	Italian Restaurant

## 1. Methodology

### Business Understanding

The main goal of this assignment is to guide stakeholders who wants to start new restaurant business by suggesting a suitable location among Toronto neighborhoods.

### Analytical Approach

Clustering of the neighborhoods was performed based on the restaurants that are present within each neighborhood.

## Exploratory Data Analysis

### 1.1. Neighborhood Data

The dataset with Toronto neighborhood details is not readily available. Therefore, I scraped a wikipedia page that has postal codes of Canada and loaded that into a panda's data frame. It had Postal Codes, Borough and Neighborhood data. We identified some postal codes as with Not Assigned Boroughs. To make the analysis easy, those rows were removed from the dataset. Then I used a csv file that has the geographical coordinates of each postal codes and loaded it into a data frame. After this both the data frames were merged to get a combined dataset. Since we have data for the whole Canada, to make it only to Toronto, we filtered out Toronto rows as our final dataset for further analysis. Then this data was used to get the nearby restaurants within 500 meters using foursquare location data. Then, I applied a k-means clustering algorithm on this neighborhood and clustered them based on the restaurants.

### 1.2. Visualizing the restaurants in Toronto neighborhood



### 1.3. Total number of restaurants in each neighbourhood

Analysis on the number of restaurants located in each neighbourhood and it came around 30 restaurants in most of the neighbourhood within 500 meters and a total of 727 restaurants.

Neighborhood	
Berczy Park	30
Brockton, Parkdale Village, Exhibition Place	14
Business reply mail Processing Centre, South Central Letter Processing Plant Toronto	5
CN Tower, King and Spadina, Railway Lands, Harbourfront West, Bathurst Quay, South Niagara, Island airport	2
Central Bay Street	30
Christie	7
Church and Wellesley	30
Commerce Court, Victoria Hotel	30
Davisville	30
Davisville North	4
Dufferin, Dovercourt Village	8
First Canadian Place, Underground city	30
Forest Hill North & West, Forest Hill Road Park	4
Garden District, Ryerson	30
Harbourfront East, Union Station, Toronto Islands	30
High Park, The Junction South	15
India Bazaar, The Beaches West	12
Kensington Market, Chinatown, Grange Park	30
Lawrence Park	1
Little Portugal, Trinity	30
Moore Park, Summerhill East	2
North Toronto West, Lawrence Park	9
Parkdale, Roncesvalles	8
Queen's Park, Ontario Provincial Government	29
Regent Park, Harbourfront	21
Richmond, Adelaide, King	30
Rosedale	1
Runnymede, Swansea	25
St. James Town	30
St. James Town, Cabbagetown	28
Stn A PO Boxes	30
Studio District	26
Summerhill West, Rathnelly, South Hill, Forest Hill SE, Deer Park	9
The Annex, North Midtown, Yorkville	14
The Beaches	3

### Top 10 restaurants in each 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	Bakery	Restaurant	French Restaurant	Café	Seafood Restaurant	Greek Restaurant	Diner	Italian Restaurant	Irish Pub	Indian Restaurant
1	Brockton, Parkdale Village, Exhibition Place	Café	Breakfast Spot	Bakery	Sandwich Place	Burrito Place	Vietnamese Restaurant	Italian Restaurant	Japanese Restaurant	Restaurant	Diner
2	Business reply mail Processing Centre, South C...	Fast Food Restaurant	Pizza Place	Restaurant	Burrito Place	Dumpling Restaurant	Creperie	Cuban Restaurant	Deli / Bodega	Dim Sum Restaurant	Diner
3	CN Tower, King and Spadina, Railway Lands, Har...	American Restaurant	Tapas Restaurant	Ethiopian Restaurant	Cuban Restaurant	Deli / Bodega	Dim Sum Restaurant	Diner	Donut Shop	Dumpling Restaurant	Eastern European Restaurant
4	Central Bay Street	Café	Italian Restaurant	Japanese Restaurant	Chinese Restaurant	Sushi Restaurant	Sandwich Place	New American Restaurant	Portuguese Restaurant	Donut Shop	Diner

## 2. Results

Around 30 restaurants in each neighborhood using Foursquare API and found top 10 out of them. To cluster these neighbourhoods, we applied a k-means clustering algorithm with 5 clusters. Each cluster is different from one another based on type of restaurants. The cluster separation can be seen in the below map.



We can see from majority of the neighbourhoods belongs to **cluster 1 (marked red)**

	Neighbourhood	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
4	Regent Park, Harbourfront	Café	Bakery	Restaurant	Breakfast Spot	Greek Restaurant	Seafood Restaurant	Italian Restaurant	Japanese Restaurant	Mexican Restaurant	Deli / Bodega
6	Queen's Park, Ontario Provincial Government	Sushi Restaurant	Burrito Place	Diner	Wings Joint	Burger Joint	Café	Chinese Restaurant	Creperie	Portuguese Restaurant	Falafel Restaurant
13	Garden District, Ryerson	Middle Eastern Restaurant	Café	Fast Food Restaurant	Italian Restaurant	Japanese Restaurant	Ramen Restaurant	Restaurant	Ethiopian Restaurant	Gastropub	Diner
22	St. James Town	Café	Restaurant	Gastropub	Italian Restaurant	Japanese Restaurant	American Restaurant	Diner	Middle Eastern Restaurant	Molecular Gastronomy Restaurant	New American Restaurant
31	Berczy Park	Bakery	Restaurant	French Restaurant	Café	Seafood Restaurant	Greek Restaurant	Diner	Italian Restaurant	Irish Pub	Indian Restaurant
40	Central Bay Street	Café	Italian Restaurant	Japanese Restaurant	Chinese Restaurant	Sushi Restaurant	Sandwich Place	New American Restaurant	Portuguese Restaurant	Donut Shop	Diner
41	Christie	Café	American Restaurant	Diner	Italian Restaurant	Restaurant	Eastern European Restaurant	Deli / Bodega	Dim Sum Restaurant	Donut Shop	Dumpling Restaurant
49	Richmond, Adelaide, King	American Restaurant	Café	Seafood Restaurant	Steakhouse	Sushi Restaurant	Asian Restaurant	Breakfast Spot	Burrito Place	Brazilian Restaurant	Pizza Place
50	Dufferin, Dovercourt Village	Pizza Place	Bakery	Portuguese Restaurant	Middle Eastern Restaurant	Café	Brazilian Restaurant	Eastern European Restaurant	Deli / Bodega	Dim Sum Restaurant	Diner
56	Harbourfront East, Union Station, Toronto Islands	Café	Fried Chicken Joint	Italian Restaurant	Restaurant	Pizza Place	New American Restaurant	Sandwich Place	Mexican Restaurant	Chinese Restaurant	Deli / Bodega
59	Little Portugal, Trinity	Asian Restaurant	Vegetarian / Vegan Restaurant	Bakery	New American Restaurant	Pizza Place	Vietnamese Restaurant	Greek Restaurant	Café	Deli / Bodega	Cuban Restaurant

Cluster 2 and cluster3, each cluster has only one neighbourhood.

	Neighbourhood	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
93	Lawrence Park	Dim Sum Restaurant	Wings Joint	Ethiopian Restaurant	Cuban Restaurant	Deli / Bodega	Diner	Donut Shop	Dumpling Restaurant	Eastern European Restaurant	Falafel Restaurant

	Neighbourhood	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
147	Rosedale	Japanese Restaurant	Wings Joint	Falafel Restaurant	Deli / Bodega	Dim Sum Restaurant	Diner	Donut Shop	Dumpling Restaurant	Eastern European Restaurant	Ethiopian Restaurant

If you compare cluster2 and cluster3, they become different because of two restaurant types; a Cuban Restaurant in cluster2 which is not in cluster3 and a Japanese Restaurant in cluster3 which is not in cluster2. Cluster 4 has two neighbourhood and what makes it different from cluster2 & 3 is an Asian restaurant, Breakfast Spot and a Burger Joint.

	Neighbourhood	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
30	The Beaches	Pizza Place	Asian Restaurant	Burger Joint	Ethiopian Restaurant	Cuban Restaurant	Deli / Bodega	Dim Sum Restaurant	Diner	Donut Shop	Dumpling Restaurant
102	Davisville North	Pizza Place	Breakfast Spot	Asian Restaurant	Sandwich Place	Cuban Restaurant	Deli / Bodega	Dim Sum Restaurant	Diner	Donut Shop	Dumpling Restaurant

Cluster 5 also has one neighbourhood and what makes it specific is having an Italian Restaurant and a normal restaurant.

	Neighbourhood	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
129	Moore Park, Summerhill East	Italian Restaurant	Restaurant	Wings Joint	Eastern European Restaurant	Cuban Restaurant	Deli / Bodega	Dim Sum Restaurant	Diner	Donut Shop	Dumpling Restaurant

### 3. Discussions

I can suggest that opening burger restaurant in cluster 2, cluster 3 and cluster 5 could be a good option as it is clearly seen on the result section. There is no Indian restaurant within 500 meters in any of the cluster, so a person who wants to start an Indian restaurant can preferably choose any of the cluster. It is also observed that Italian, Dumpling, Eastern European, Japanese cuisines are already much in count. As a result, opening any such would make a risk to the business. In cluster 2 and 3, there is no breakfast option, which makes a way for a breakfast spot.

### 4. Conclusion

We can clearly see that there are a lot of scope in all clusters except cluster 1. As the data indicated the analysis is performed on small set of data within 500 meters, so that better vision can be achieved by increasing the radius.