Battle of Neighbourhoods Toronto

Shohra Afaque

Introduction:

Toronto is 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 diverse population of Toronto reflects its current and historical role as an important destination for immigrants to Canada. More than 50 percent of residents belong to a visible minority population group and over 200 distinct ethnic origins are represented among its inhabitants. While the majority of Torontonians speak English as their primary language, over 160 languages are spoken in the city.

Description:

Being so culturally diverse and acting as a business hub Toronto offers many business opportunities for new bees. A careful consideration of analysis for the business expansion is so much important as it directly proportional to the cost of the business. The analysis from this report help for the new businesses strategically target the market and help in high return on investment so is low risk.

First step in opening a new restaurant is deciding what type of restaurant it is going to be? Is it a high end fine dining restaurant or casual style dinner or a specific type of cuisine in mind such as French , Italian or Indian. Choosing a restaurant location in a busy area with plenty of foot traffic, or is there enough parking or are there any other restaurants opened or closed in the same spot?

Target Audience:

- The primary objective of this project is to recommend which neighbourhood is best to open a new restaurant?
- We want to use data & find patterns in the existing data, to help the investors in making right decision toward their investments.
- A data driven argument towards a proposed recommendation.
- A data driven argument towards a recommendation.

Question to Answer:

- What is best location in Toronto for a new restaurant ?
- Which areas have potential for a new restaurant ?
- Which all areas lack specific cuisine restaurant ?

Data:

There is no second opinion that understand the problem and business requirement is extremely important. As important it is to understand the problem, choosing the data science approach to give a solution it is important to understand the data. Once we start analysing the data then we may feel the need to gather more data and the quest continues. Being part of this internet era we are surrounded by data, data is flowing all around us. It is everywhere, gathering data has never been so easy.

Following are the different source of data with sample:-

- *List of Postal Codes & Neighbourhoods (Wikipedia)*
  - We will be using list of Toronto Neighbourhood data scrapped form Wikipedia (https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M) for analysis. There are total 180 unique postal codes with 208 unique neighbourhoods in Toronto. This will be a great starting point for us. Starting from here we will gather more data online.

  - Sample record from Wikipedia

Postcode	Borough	Neighbourhood
M4N	Central Toronto	Lawrence Park
M5N	Central Toronto	Roselawn
M4P	Central Toronto	Davisville North
M5P	Central Toronto	Forest Hill North
M5P	Central Toronto	Forest Hill West
M4R	Central Toronto	North Toronto West
M5R	Central Toronto	The Annex
M5R	Central Toronto	North Midtown

- *Coordinates of each Postcode:-*
  - We can use the geopy library in python to get the latitude and longitude information for each neighbourhood.
  - We can then use the coordinates to get the 100 nearby restaurants within 3000 meters using foursquare location data.
  - Once the coordinates are obtained by geopy library they can be joined with the neighbourhood table that we obtained from the Wikipedia.
  - Sample record of joining coordinates with the neighbourhood table.

	Postcode	Borough	Neighbourhood	Latitude	Longitude
0	M7R	Mississauga	Canada Post Gateway Processing Centre	43.636966	-79.615819
1	M9R	Etobicoke	Kingsview Village, Martin Grove Gardens, Richv...	43.688905	-79.554724
2	M5V	Downtown Toronto	CN Tower, Bathurst Quay, Island airport, Harbo...	43.628947	-79.394420
3	M4B	East York	Woodbine Gardens, Parkview Hill	43.706397	-79.309937
4	M2R	North York	Willowdale West	43.782736	-79.442259

- *Coordinates and Venues (Foursquare API)*

- We will then use this Toronto geographical data as input to the foursquare location data and fetch top 100 restaurants nearby to each neighbourhood within 3000 meters.
- Once we have all of these data then we can start making preparing our clustering model. Once the clustering model is trained we can start making decision examining each cluster of neighbourhoods.
- Sample response from foursquare API can be found at this URL, please have a look: Foursquare API sample response (click me)
- The sample after joining the category of venue it with the postal code looks like this

	Postal Code	Neighborhood Latitude	Neighborhood Longitude	Neighbourhood	Venue Latitude	Venue Longitude	Venue Category
0	M5V	43.628947	-79.39442	Billy Bishop Toronto City Airport (YTZ) (Billy...	43.631579	-79.395605	Airport
1	M5V	43.628947	-79.39442	Porter Lounge	43.630680	-79.395756	Airport Lounge
2	M5V	43.628947	-79.39442	Toronto Harbour	43.633045	-79.396484	Harbor / Marina
3	M5V	43.628947	-79.39442	Billy Bishop Café	43.631132	-79.396139	Airport Food Court
4	M5V	43.628947	-79.39442	Air Canada Check-In Counter	43.631226	-79.395987	Airport Terminal

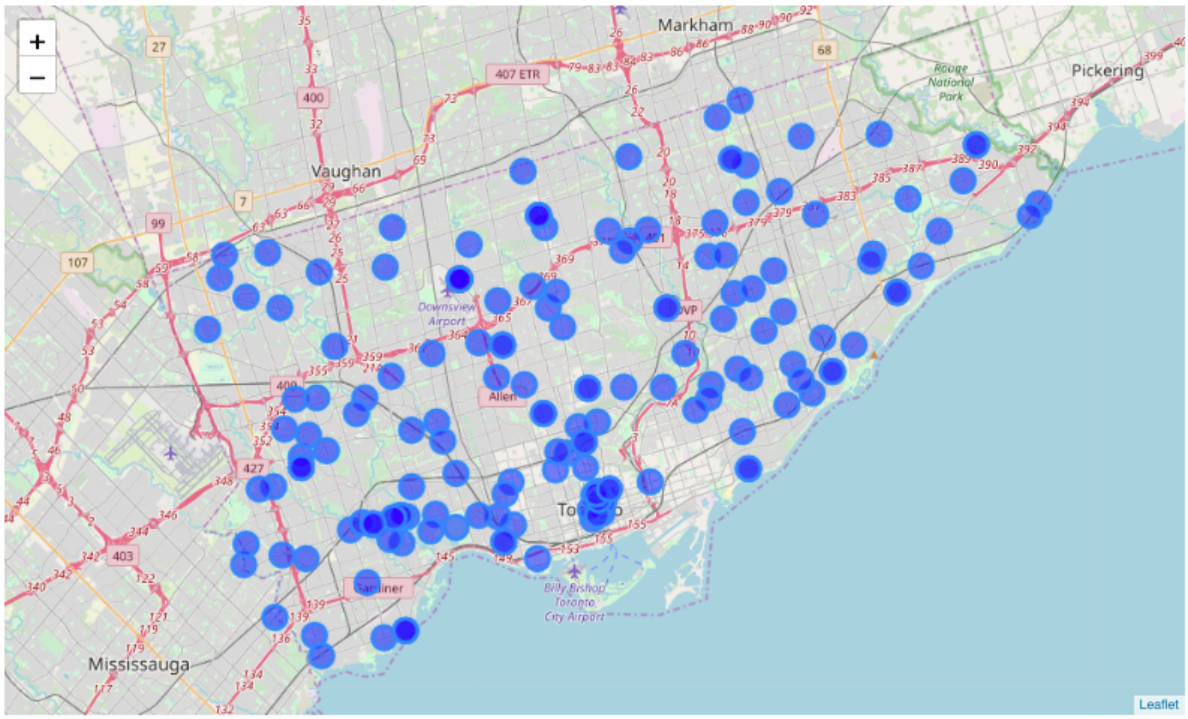
Methodology:

- **Business Understanding:**
  The objective of this project is to find which neighbourhood of Toronto is a good choice for a new restaurant business to open. i.e. finding the best location for a new restaurant to start.

- **Analytical Approach:**
  Toronto City has total of 28 unique Neighbourhoods and our goal in this project is to cluster these neighbourhood based on the similarities between neighbourhoods based on the restaurants present in the neighbourhood.

- **Exploratory Data Analysis:**
  - **Toronto Neighbourhood data:**
    1. We first scrap the neighbourhood information from the Wikipedia and load that into a data frame dropping unnecessary columns.
    2. Then we perform data wrangling to convert the data into analysis ready form
    3. Then we use geopy library to get the latitude and longitude information for each neighbourhood
    4. Then we use this data to get the 100 nearby restaurants within 3000 meters using foursquare location data.
    5. Calculate and perform unique restaurant categories in each neighbourhood and find the top 10 most common restaurants.
    6. Then we will cluster all the restaurants in all the neighbourhoods using k-means clustering algorithm.

- **Toronto city Neighbourhood visualization:**



- **Total number of venues in each Neighbourhood:**
  There are around 100 different types of restaurants in all the neighbourhood and total of 2863 restaurants.

- **Find 10 most common restaurants in each neighbourhood:**

```
In [30]: 1 neighborhoods_venues_sorted
```

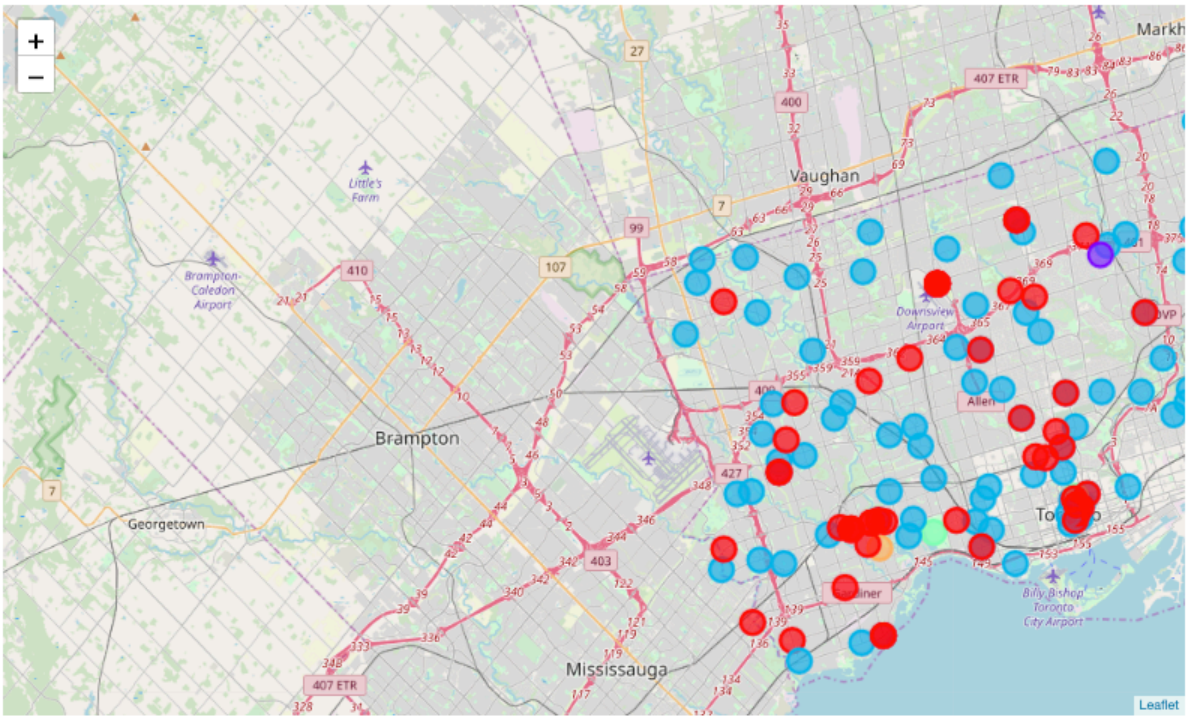
Out[30]:

	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	Agincourt	Chinese Restaurant	Vietnamese Restaurant	Bakery	Caribbean Restaurant	Hotpot Restaurant	Dumpling Restaurant	Restaurant	Sandwich Place	Mexican Restaurant	Pizza Place
1	Agincourt North	Chinese Restaurant	Fast Food Restaurant	BBQ Joint	Sandwich Place	Noodle House	Indian Restaurant	Fried Chicken Joint	Cantonese Restaurant	Food Court	Restaurant
2	Albion Gardens	Fast Food Restaurant	Indian Restaurant	Café	Sandwich Place	Chinese Restaurant	Pizza Place	Bakery	Italian Restaurant	Japanese Restaurant	Restaurant
3	Alderwood	Fast Food Restaurant	Pizza Place	Sandwich Place	Burger Joint	Burrito Place	South American Restaurant	Sushi Restaurant	Breakfast Spot	Food Court	Restaurant
4	Bathurst Manor	Pizza Place	Sandwich Place	Restaurant	Middle Eastern Restaurant	Fast Food Restaurant	Sushi Restaurant	Japanese Restaurant	Café	Breakfast Spot	Chinese Restaurant
5	Bedford Park	Pizza Place	Asian Restaurant	Restaurant	Bakery	Thai Restaurant	Café	Middle Eastern Restaurant	Sushi Restaurant	Korean Restaurant	Seafood Restaurant

Results:

From the foursquare location data, we got 100 restaurants in each neighbourhood and found most common restaurants. Performing the k-means clustering on the most common restaurants for each neighbourhood will partition the data set into 5 clusters. The 5 clusters are partitioned based on similar type of restaurants that belong to neighbourhoods.

The folium representation of the clustered data will look like below.



Examining the Clusters:

We can examine each cluster and determine the discriminating types of restaurants that distinguish each cluster.

- **Cluster 1:**

	Neighborhood	Relative_Location	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 Comm Ven
2	Parkwoods	North York, Toronto	43.758800	-79.320197	2	Pizza Place	Fast Food Restaurant	Middle Eastern Restaurant	Chinese Restaurant	Japanese Restaurant	Sandwich Place	Restaurant	Seafo Restaur
3	Victoria Village	North York, Toronto	43.732658	-79.311189	2	Sandwich Place	Fast Food Restaurant	Middle Eastern Restaurant	Japanese Restaurant	Burger Joint	Italian Restaurant	Restaurant	Americ Restaur
6	Lawrence Heights	North York, Toronto	43.722778	-79.450933	2	Fast Food Restaurant	Sandwich Place	Italian Restaurant	Pizza Place	Vietnamese Restaurant	Restaurant	American Restaurant	C
7	Lawrence Manor	North York, Toronto	43.722079	-79.437507	2	Italian Restaurant	Sushi Restaurant	Bakery	Japanese Restaurant	Burger Joint	Café	Middle Eastern Restaurant	Steakhou
10	Islington Avenue	Etobicoke, Toronto	43.678207	-79.538269	2	Pizza Place	Café	Sandwich Place	Restaurant	Bakery	Chinese Restaurant	Diner	As Restaur

- **Cluster 2:**

	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
2	Parkwoods	Pizza Place	Fast Food Restaurant	Middle Eastern Restaurant	Chinese Restaurant	Japanese Restaurant	Sandwich Place	Restaurant	Seafood Restaurant	Burger Joint	Sushi Restaurant
3	Victoria Village	Sandwich Place	Fast Food Restaurant	Middle Eastern Restaurant	Japanese Restaurant	Burger Joint	Italian Restaurant	Restaurant	American Restaurant	Pizza Place	Chinese Restaurant
6	Lawrence Heights	Fast Food Restaurant	Sandwich Place	Italian Restaurant	Pizza Place	Vietnamese Restaurant	Restaurant	American Restaurant	Café	Bagel Shop	Bakery
7	Lawrence Manor	Italian Restaurant	Sushi Restaurant	Bakery	Japanese Restaurant	Burger Joint	Café	Middle Eastern Restaurant	Steakhouse	Gastropub	Sandwich Place
10	Islington Avenue	Pizza Place	Café	Sandwich Place	Restaurant	Bakery	Chinese Restaurant	Diner	Asian Restaurant	Japanese Restaurant	Breakfast Spot
11	Rouge	Fast Food Restaurant	Pizza Place	Fried Chicken	Breakfast Spot	Sandwich Place	Burger Joint	Bakery	Restaurant	Mediterranean Restaurant	Chinese Restaurant

- **Cluster 3:**

	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
2	Parkwoods	Pizza Place	Fast Food Restaurant	Middle Eastern Restaurant	Chinese Restaurant	Japanese Restaurant	Sandwich Place	Restaurant	Seafood Restaurant	Burger Joint	Sushi Restaurant
3	Victoria Village	Sandwich Place	Fast Food Restaurant	Middle Eastern Restaurant	Japanese Restaurant	Burger Joint	Italian Restaurant	Restaurant	American Restaurant	Pizza Place	Chinese Restaurant
6	Lawrence Heights	Fast Food Restaurant	Sandwich Place	Italian Restaurant	Pizza Place	Vietnamese Restaurant	Restaurant	American Restaurant	Café	Bagel Shop	Bakery
7	Lawrence Manor	Italian Restaurant	Sushi Restaurant	Bakery	Japanese Restaurant	Burger Joint	Café	Middle Eastern Restaurant	Steakhouse	Gastropub	Sandwich Place
10	Islington Avenue	Pizza Place	Café	Sandwich Place	Restaurant	Bakery	Chinese Restaurant	Diner	Asian Restaurant	Japanese Restaurant	Breakfast Spot
11	Rouge	Fast Food Restaurant	Pizza Place	Fried Chicken	Breakfast Spot	Sandwich Place	Burger Joint	Bakery	Restaurant	Mediterranean Restaurant	Chinese Restaurant

- **Cluster 4:**

	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
85	Union Station	Italian Restaurant	Xinjiang Restaurant	Food Court	Donut Shop	Dumpling Restaurant	Eastern European Restaurant	Egyptian Restaurant	Empanada Restaurant	Ethiopian Restaurant	Falafel Restaurant
159	High Park	Sushi Restaurant	Deli / Bodega	Italian Restaurant	Food Court	Donut Shop	Dumpling Restaurant	Eastern European Restaurant	Egyptian Restaurant	Empanada Restaurant	Ethiopian Restaurant
193	Tam O'Shanter	Deli / Bodega	Xinjiang Restaurant	Diner	Donut Shop	Dumpling Restaurant	Eastern European Restaurant	Egyptian Restaurant	Empanada Restaurant	Ethiopian Restaurant	Falafel Restaurant

- **Cluster 5:**

	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
44	West Hill	Restaurant	Pizza Place	Xinjiang Restaurant	Diner	Donut Shop	Dumpling Restaurant	Eastern European Restaurant	Egyptian Restaurant	Empanada Restaurant	Ethiopian Restaurant
165	Wexford	Pizza Place	Xinjiang Restaurant	Food	Donut Shop	Dumpling Restaurant	Eastern European Restaurant	Egyptian Restaurant	Empanada Restaurant	Ethiopian Restaurant	Falafel Restaurant
266	Humber Bay	Pizza Place	Café	Xinjiang Restaurant	Diner	Donut Shop	Dumpling Restaurant	Eastern European Restaurant	Egyptian Restaurant	Empanada Restaurant	Ethiopian Restaurant

Discussions:

1. From the geographical presentation of the clusters we can say that there is a scope to open a new Indian restaurant in West Hill from 5th Cluster.
2. From the results we can suggest that opening a new Japanese Cuisine restaurant in North York, Toronto from 1st cluster will be beneficial, because mostly Italian cuisine restaurants are present.

3. It can also be suggested that opening a restaurant near East York, Toronto has low profile there is high competition and highly crowded in all the cuisine.

Conclusion:

As the analysis is performed on small set of data, we can achieve better results by increasing the neighbourhood information. Toronto has scope to expand different types restaurants at different places. Based on the current results it can be observed that North York & West Hill of Downtown Toronto has many different types of new restaurant business to offer. It can also be explored further with careful consideration of neighbourhoods at East and South sides of Toronto.