

Introduction

Immigrating to a new country like Canada comes with much excitement as well as some fear of uncertainty. Our clients who is moving to Canada from Africa has been exploring ways of getting the best from his soon to be home. As a business minded person, he wants to know the neighborhood in Toronto where Restaurants are located and not only that, he wants to know what kind of restaurants exist in each location within a particular Borough or Neighborhood. He is convinced that having that information will help him make the right decisions as to what part of the city of Toronto he should reside or set up his business. Therefore, in this project, our goal is to help our client answer the following questions and give more vital insights if possible:

- Where are restaurants located in the city of Toronto?
- What are the top restaurants in that Borough?
- How can I know the names or types of restaurants in specific address within the Borough?

Tools and Libraries

Python Libraries:

Pandas: Pandas is a python library that makes the exploration of data of different format easy. It supports the reading of data from format such as excel, csv, table, html etc.

Request: Request is a python library that allows you to make HTTP requests and interact with API endpoints. According to the pypi.org, “requests is one of the most downloaded python packages”.

Geopy: This python library provides the ability to locate coordinates of addresses, cities, and various locations across the globe. In this project, we used this library to get the coordinates of all the neighborhood in our Toronto dataset.

Folium: Folium is a powerful python library that allows us to visualize our location coordinates in form of a leaflet map. With Folium, you can build an interactive map that can also be rendered on the browser.

Data Description

The initial dataset was collected from Wikipedia (https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M). It contains list of postal codes in the city of Toronto as well as the Borough and Neighborhood that falls within that postal code. Our initial dataset contains 288 rows and 3 columns. From our initial dataset, there was no information about the restaurants in any Borough. In order to get data containing restaurants in the Boroughs, first we, we used the geopy python library to get the coordinates for each Borough. Secondly, we used the requests library to fetch more data about the Borough such as venues, and other useful data from the Foursquare API. Having collected all these other data, we combine them to form a new dataset that necessary information which can help us answer the questions before us.

Data Analysis and Exploration

Data Preprocessing - We discovered that our dataset contains missing values ('Not assigned), duplicates postal code. In order to clean up our data, we did the following:

- Dropped rows where values were missing ('Not assigned) for both Borough and Neighborhood (77 rows in total).

- Combined rows with duplicate postal code as one. For example, postal code “M1C” below:

Duplicated postal code “M1C”

	Postcode	Borough	Neighborhood
27	M1C	Scarborough	Highland Creek
28	M1C	Scarborough	Rouge Hill
29	M1C	Scarborough	Port Union

Combined postal code “M1C”

1	M1C	Scarborough	Highland Creek,Rouge Hill,Port Union
---	-----	-------------	--------------------------------------

Data Visualization: After cleaning up our data, we visualized our data in order to get general information from it which we later guide us on the kind of insights we can get. First, we plotted the neighborhoods against the boroughs. The North York borough has the highest number of neighborhoods as shown in the fig.1 below. However, our interest is in getting insights and discovering restaurants.

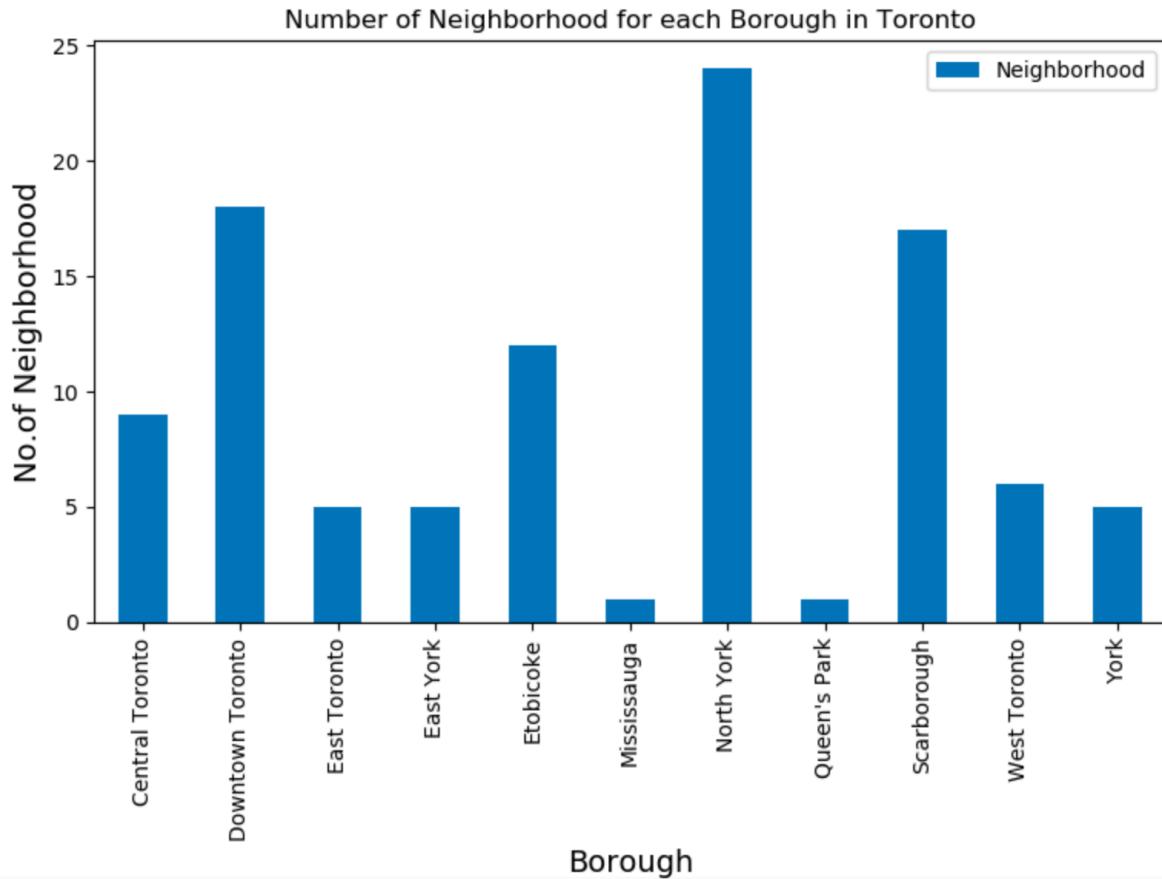


Fig 1: Neighborhoods in each borough

Data exploration – How can we discover where the restaurants are located in the city of Toronto? To answer this question, we leverage on the coordinates of each borough. First, we used postal code for each borough to fetch the coordinates by supplying it as an argument to the geopy library. Then we made a call (query) to the foursquare API and provided the coordinates for each borough programmatically.

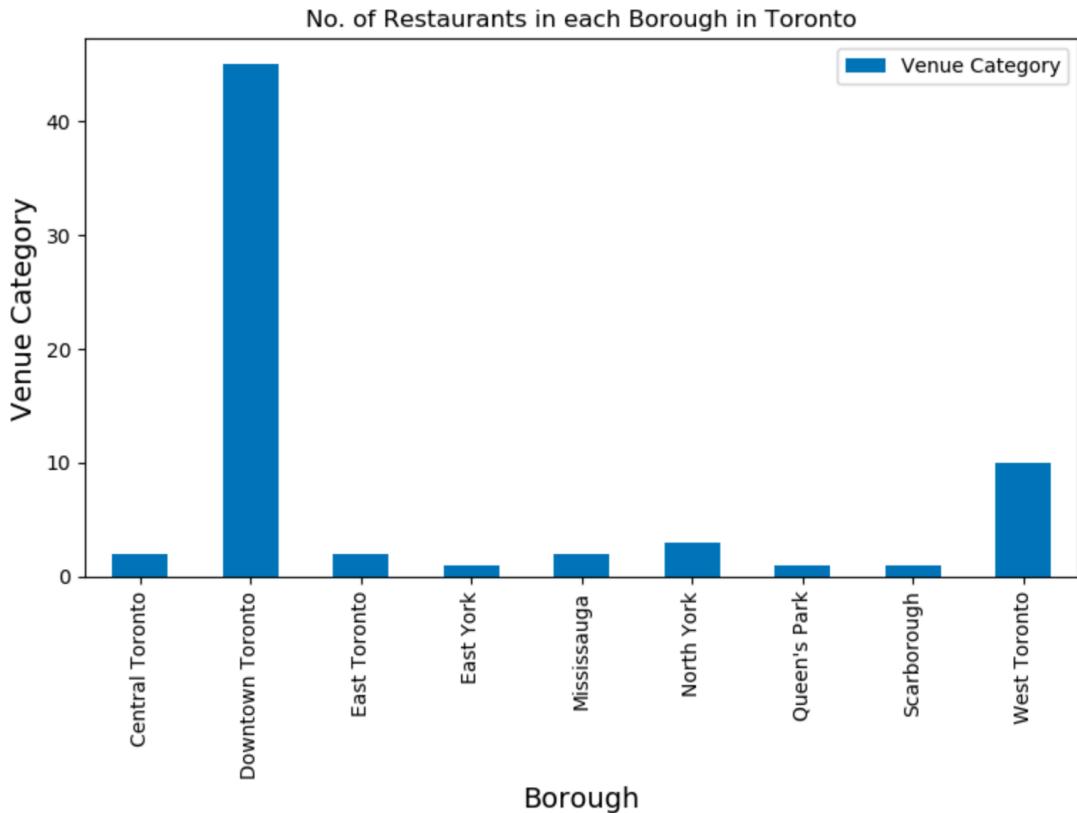


Fig 2: Restaurants in each borough

Now we have the information about borough with restaurants. However, this gives rise to another question. How do we know what kind of restaurants are available in these boroughs? In order to get that, we define a python function, to return for us the name of the restaurants as well as the physical address and location coordinates of the restaurants in the boroughs where a restaurant is present.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	ID	Address	Venue	Venue Latitude	Venue Longitude	Venue Category
11	Scarborough Village	43.743125	-79.231750	4bb3e779f187a593b89e13f8	2990 Eglinton Ave E	Swiss Chalet Rotisserie & Grill	43.740378	-79.232656	Restaurant
118	Fairview, Henry Farm, Oriole	43.780880	-79.347796	5650fa8e498e405c38a5e6d4	1800 Sheppard Ave E, Fairview Mall	Thai Express	43.777990	-79.344091	Restaurant
320	Leaside	43.709765	-79.363901	53c95181498ec1d30073384e	190 Laird Drive (Wicksteed Avenue)	The Leaside Pub	43.710429	-79.363547	Restaurant
474	Davisville	43.703395	-79.385964	58a881442520ae596a926970	505 Mount Pleasant Road	Starving Artist	43.701538	-79.387240	Restaurant
500	Cabbagetown, St. James Town	43.668160	-79.366602	5412007d498e1e2d959ac35e	568 Parliament St. (Wellesley)	Murgatroid	43.667381	-79.369311	Restaurant

Table with the name, address, coordinates of restaurants in Toronto

- **Clustering using k-means** – This is the final phase of analysis, where we group restaurants into a cluster based on dissimilarities. In our project we focused on the Downtown Toronto borough as it appeals well to our client's need. So we grouped the restaurants in Downtown Toronto boroughs using a k-means algorithm. Since our dataset is unlabeled, we need to know what kinds of restaurants you are likely to find in a particular location within the Downtown Toronto borough. Should our client decide to open his restaurant business, it will be easier to make a proper decision where to locate his business in the heart of the city of Toronto.

Results/Findings and Discussions

At the end of our Data Analysis, exploration, and clustering, these are our findings:

- The neighborhood with the highest number of restaurants is Downtown Toronto. Being the heart of the city of Toronto, restaurants are more likely to open up in the area because of the many activities that go on there. In addition, with the number of businesses in that area, people who work there will patronize the restaurants.

- Our findings also show that within the Downtown Toronto boroughs, two cities seem to have the highest number of restaurants. They are St. James Town, and Commerce Court Victoria Hotel as shown in the fig below.

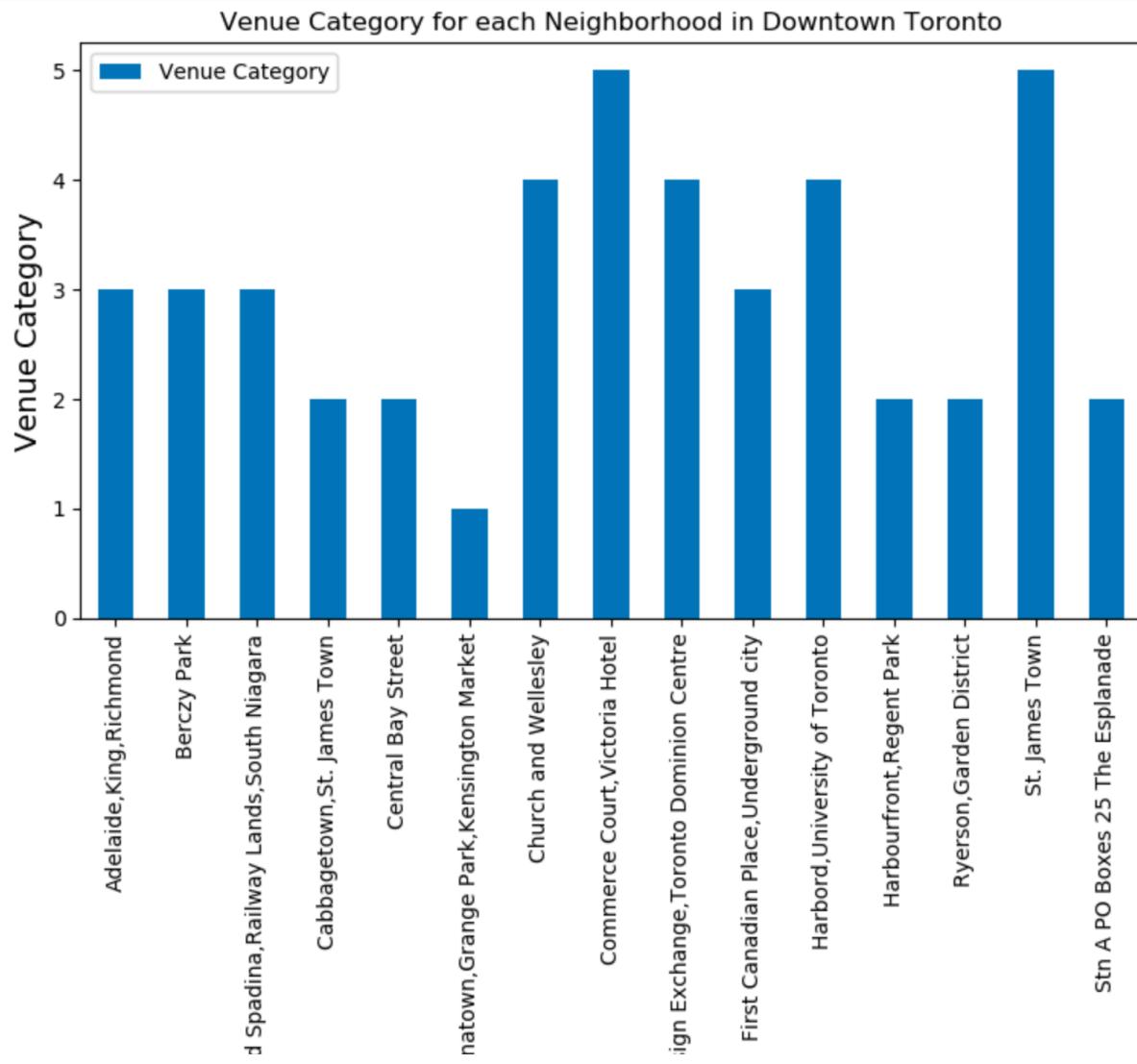


Fig 3. Venue category (Restaurants) in each Neighborhood

We discovered as well that for each neighborhood within Toronto downtown, that there are venues (restaurants) that are most common. For example, in “Adelaide, King, Richmond”

Neighborhood, Bannock restaurant is the most common in that neighborhood. The table below shows the most common restaurants in a given neighborhood.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Adelaide,King,Richmond	Bannock	Bymark	Canoe	Victoria's Restaurant	Bar Mercurio
1	Berczy Park	The Works Gourmet Burger Bistro	Real Mo-Mo's	Marché Mövenpick	Wish	Impact Kitchen B
2	CN Tower,Bathurst Quay,Island airport,Harbour...	Big Smoke Burger	Colette Grand Café	Portland Variety	Wish	JOEY Eaton Centre B
3	Cabbagetown,St. James Town	The Pear Tree	Murgatroid	Wish	JOEY Eaton Centre	Bar Mercurio
4	Central Bay Street	Jack Astor's Bar & Grill	JOEY Eaton Centre	Victoria's Restaurant	Bar Mercurio	Big Smoke Burger

In addition, we used the k-means algorithm to cluster the restaurants in Downtown Toronto neighborhood. With a cluster size of 5, we got four cluster labels [0, 1, 3 ,4]. We show below the neighborhood and the restaurants in each of the cluster. In cluster 0, there are twenty-one (21) neighborhoods. Cluster 1, there are seventeen (17) neighborhoods. In cluster 3, there are five (5) neighborhoods. Cluster 4 has just one neighborhood.

Cluster 0:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
4	Cabbagetown,St. James Town	The Pear Tree	Murgatroid	Wish	JOEY Eaton Centre	Bar Mercurio
5	Cabbagetown,St. James Town	The Pear Tree	Murgatroid	Wish	JOEY Eaton Centre	Bar Mercurio
6	Church and Wellesley	Wish	Sushi Shop	Fabarnak	O. Noir	JOEY Eaton Centre
7	Church and Wellesley	Wish	Sushi Shop	Fabarnak	O. Noir	JOEY Eaton Centre
8	Church and Wellesley	Wish	Sushi Shop	Fabarnak	O. Noir	JOEY Eaton Centre

Cluster 1:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
24	Adelaide,King,Richmond	Bannock	Bymark	Canoe	Victoria's Restaurant	Bar Mercurio
25	Adelaide,King,Richmond	Bannock	Bymark	Canoe	Victoria's Restaurant	Bar Mercurio
26	Adelaide,King,Richmond	Bannock	Bymark	Canoe	Victoria's Restaurant	Bar Mercurio
27	Design Exchange,Toronto Dominion Centre	Bymark	Canoe	Marché Mövenpick	Kellys Landing	Wish

Cluster 3:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
12	Ryerson,Garden District	Jack Astor's Bar & Grill	JOEY Eaton Centre	Victoria's Restaurant	Bar Mercurio	Big Smoke Burger
13	Ryerson,Garden District	Jack Astor's Bar & Grill	JOEY Eaton Centre	Victoria's Restaurant	Bar Mercurio	Big Smoke Burger
22	Central Bay Street	Jack Astor's Bar & Grill	JOEY Eaton Centre	Victoria's Restaurant	Bar Mercurio	Big Smoke Burger
23	Central Bay Street	Jack Astor's Bar & Grill	JOEY Eaton Centre	Victoria's Restaurant	Bar Mercurio	Big Smoke Burger

Cluster 4:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
42	Chinatown,Grange Park,Kensington Market	Helena's Magic Kitchen	Wish	Victoria's Restaurant	Bar Mercurio	Big Smoke Burger

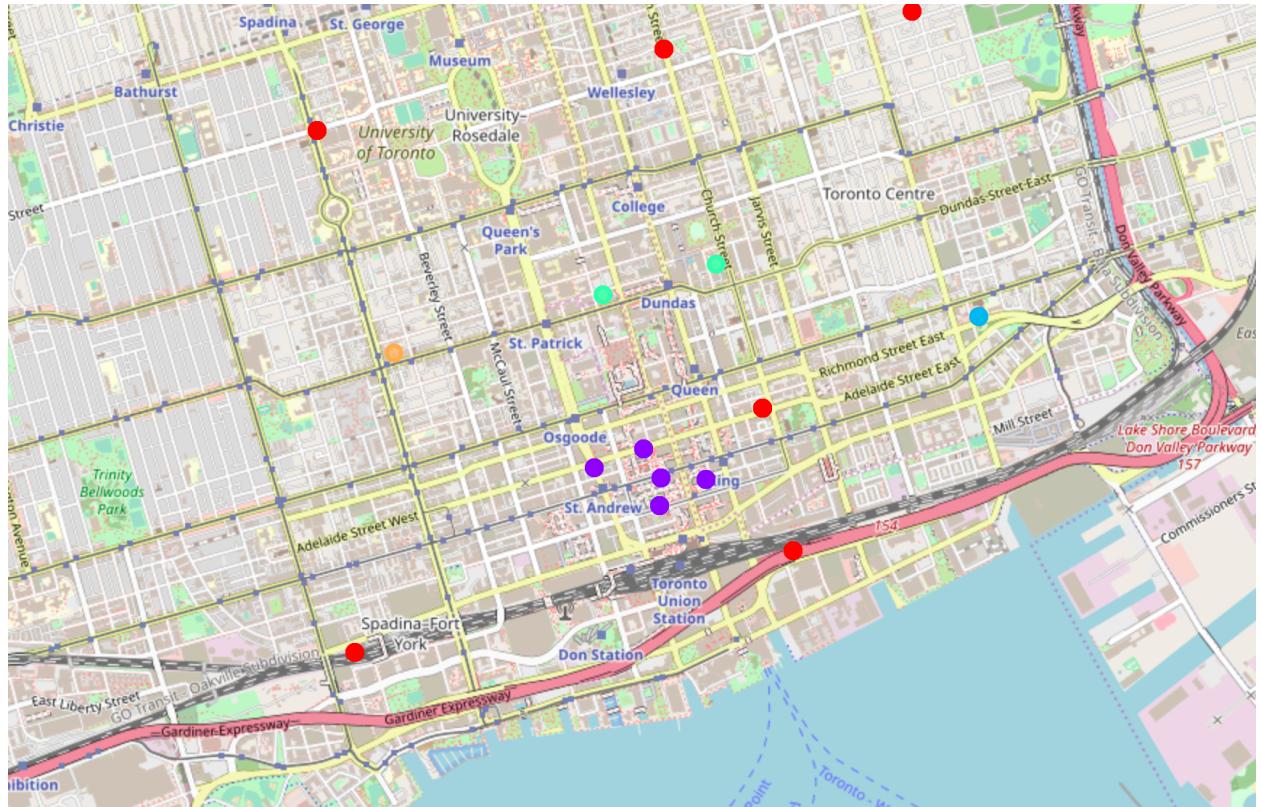


Fig 4. Leaflet map of the Downtown Toronto Neighborhoods and the cluster

Conclusion

There are so many insights that can be drawn from the dataset we collected such as the ratings of each restaurants, the likes, the tips etc. But for the scope of this project we have limited our explorations to finding the restaurants that are within the Downtown Toronto neighborhood. We have discovered that there are many restaurants situated in Downtown Toronto borough. And also, we have been able to identify those restaurants and the physical address within the neighborhood. In future work, we would like to explore our dataset and see what the major factor that determines why a restaurant would prefer to open up in a particular neighborhood.