

Capstone Project

Introduction

The city of Toronto is one of the major cities of Canada. Therefore it might be great opportunity to open a restaurant in Toronto. As there are many restaurants in Toronto already opening a new can be a challenging task.

Business Problem

The objective of the capstone project is to find a most suitable for the business owner to open a new restaurant in Toronto. And what type of restaurant will it be and where it should be opened.

Target Audience

For any business owner/entrepreneur who wants to open a new restaurant in Toronto.

Data

The following data will be required

- List of neighborhoods in Toronto, Canada.
https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

- Latitude and Longitude of these locations.
- Venue data related to restaurants from Foursquare API

Extracting the Data

- The scrapping of Toronto neighborhoods via Wikipedia
- Getting Latitude and Longitude data of these neighborhoods via Geocoder package
- Using Foursquare API to get venue data related to these neighborhoods

Methodology

First, I need to get the list of neighborhoods in Toronto, Canada. This is possible by extracting the list of neighborhoods from Wikipedia:

https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

I did the web scraping by utilizing pandas HTML table scraping method as it is easier and more convenient to pull tabular data directly from a web page into the data frame. However, it is only a list of neighborhood names and postal codes. I need to get their coordinates to utilize Foursquare to pull the list of venues near these neighborhoods. To get the coordinates, I tried using Geocoder Package but it was not working so I used the CSV file provided by IBM team to match the coordinates of Toronto neighborhoods. After gathering these coordinates, I

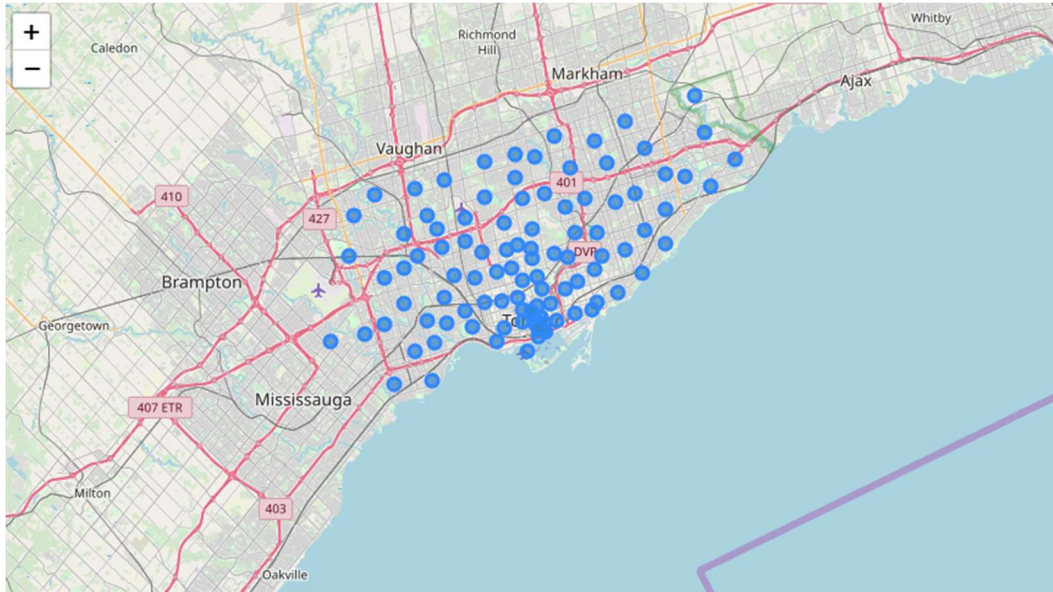
visualize the map of Toronto using Folium package to verify whether these are correct coordinates. Next, I use Foursquare API to pull the list of top 100 venues within 500 meters radius. I have created a Foursquare developer account in order to obtain account ID and API key to pull the data. From Foursquare, I am able to pull the names, categories, latitude, and longitude of the venues. With this data, I can also check how many unique categories that I can get from these venues.

Then, I analyze each neighborhood by grouping the rows by neighborhood and taking the mean on the frequency of occurrence of each venue category. This is to prepare clustering to be done later.

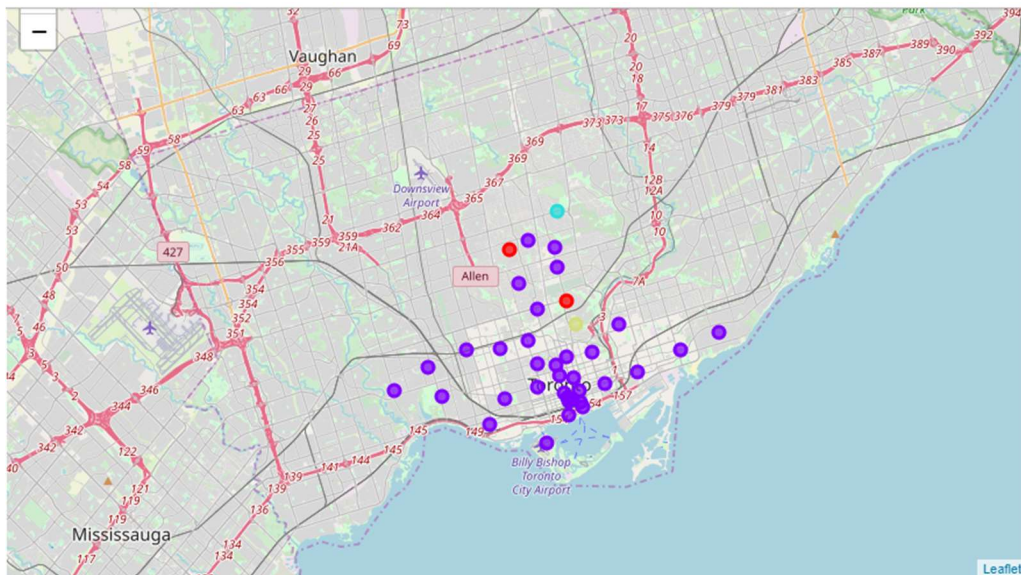
Lastly, I performed the clustering method by using k-means clustering. K-means clustering algorithm identifies k number of centroids, and then allocates every data point to the nearest cluster while keeping the centroids as small as possible. It is one of the simplest and popular unsupervised machine learning algorithms and it is highly suited for this project as well. I have clustered the neighborhoods in Toronto into 4 clusters based on their frequency of occurrence for restaurants. Based on the results (the concentration of clusters), I will be able to recommend the ideal location to open the restaurant.

Results

Visualizing the map of Toronto



Visualizing the clusters



Analyzing the result of clusters

Cluster 1

	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
count	2	2.0	2	2	2	2	2	2	2	2	2	2
unique	1	NaN	1	2	2	2	2	2	2	2	2	2
top	Central Toronto	NaN	Restaurant	French Restaurant	Sushi Restaurant	Wings Joint	Falafel Restaurant	Diner	Doner Restaurant	Donut Shop	Dumpling Restaurant	Donut Shop
freq	2	NaN	2	1	1	1	1	1	1	1	1	1

Here the most common venue: Restaurant

The 10th most common restaurant : Donut Shop

Cluster 2

	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
count	34	34.0	34	34	34	34	34	34	34	34	34	34
unique	4	NaN	10	13	19	21	19	22	23	23	25	26
top	Downtown Toronto	NaN	Café	Restaurant	Restaurant	Pizza Place	Italian Restaurant	Bakery	Gastropub	Burger Joint	Steakhouse	Doner Restaurant
freq	18	NaN	11	10	6	6	6	5	3	4	4	3

Here the most common venue: Café

The 10th most common restaurant : Doner Restaurant

Cluster 3

	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
count	1	1.0	1	1	1	1	1	1	1	1	1	1
unique	1	NaN	1	1	1	1	1	1	1	1	1	1
top	Central Toronto	NaN	Dim Sum Restaurant	Wings Joint	Cuban Restaurant	Diner	Doner Restaurant	Donut Shop	Dumpling Restaurant	Eastern European Restaurant	Ethiopian Restaurant	Falafel Restaurant
freq	1	NaN	1	1	1	1	1	1	1	1	1	1

Here the most common venue: Dim sum Restaurant

The 10th most common restaurant : Falafel restaurant

Cluster 4

	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
count	1	1.0	1	1	1	1	1	1	1	1	1	1
unique	1	NaN	1	1	1	1	1	1	1	1	1	1
top	Downtown Toronto	NaN	Japanese Restaurant	Wings Joint	Fast Food Restaurant	Dim Sum Restaurant	Diner	Doner Restaurant	Donut Shop	Dumpling Restaurant	Eastern European Restaurant	Ethiopian Restaurant
freq	1	NaN	1	1	1	1	1	1	1	1	1	1

Here the most common venue: Japanese Restaurant

The 10th most common restaurant : Ethiopian restaurant

The potential areas are:

- Rosedale, Downtown Toronto
- Forest Hill North, Forest Hill West, Central Toronto
- Moore Park, Central Toronto
- Lawrence Park, Central Toronto

As there are very few restaurants in these locations a new restaurant can be opened here. In our recommendation we advise selecting from the 10th or 9th positions. The selection is a reasonable balance between being too popular and having no customers.

Conclusion

- In this report we established a methodology where a restaurant should be opened.
- This type of analysis can be applied to any city of your choice which has available geospatial information.
- This type of analysis can be applied to any type of venue (shopping malls, clubs, etc.) That is available in the foursquare database.