Geospatial Analysis and K-means Clustering for Business Decision Making

Table of contents

A. Introduction

As the provincial capital of Ontario and the most populous city in Canada, Toronto is an international centre of business, finance, arts, and culture, and is recognized as one of the most multicultural and cosmopolitan cities in the world.

In this project, we will try to find an optimal location for a restaurant. Specifically, this report will be targeted to stakeholders interested in opening a Chinese restaurant in Toronto, Canada. Since there are lots of restaurants in Toronto, we will try to detect locations that are not already crowded with restaurants. We are also particularly interested in areas with no Chinese restaurants in the vicinity. We would also prefer locations as close to the city centre as possible, assuming that the first two conditions are met.

We will use our data science powers to generate a few most promising neighbourhoods based on this criterion. Advantages of each area will then be clearly expressed so that the best possible final location can be chosen by stakeholders.

B. Data Source

To consider the problem we can list the data required as below:

- number of existing restaurants in the neighbourhood (any type of restaurant)
- number of and distance to Chinese restaurants in the neighbourhood, if any

Following data sources will be needed to extract/generate the required information:

- The postal code and name of the neighbourhood in Toronto from **Wikipedia** (https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M)
- number of restaurants and their type and location in every neighbourhood will be obtained using the **Foursquare API**
- coordinate of each neighbourhood centre will be obtained from http://cocl.us/Geospatial_data

C. Methodology

First, the exploratory analysis will be conducted to reveal the general information like how many venues in each neighbourhood and the density of the restaurant, we will need to detect areas that have low restaurant density, particularly those with the low number of Chinese restaurants. The main data has the components **PostalCode, Borough, Neighborhood, Latitude and Longitude.**

Table 1 Geospatial Information of the neighbourhood in Toronto

	PostalCode	Borough	Neighborhood	Latitude	Longitude
0	M3A	North York	Parkwoods	43.753259	-79.329656
1	M4A	North York	Victoria Village	43.725882	-79.315572
2	M5A	Downtown Toronto	Regent Park / Harbourfront	43.654260	-79.360636
3	M6A	North York	Lawrence Manor / Lawrence Heights	43.718518	-79.464763
4	M7A	Queen's Park / Ontario Provincial Government	Queen's Park / Ontario Provincial Government	43.662301	-79.389494
5	M9A	Etobicoke	Islington Avenue	43.667856	-79.532242
6	M1B	Scarborough	Malvern / Rouge	43.806686	-79.194353
7	M3B	North York	Don Mills	43.745906	-79.352188
8	M4B	East York	Parkview Hill / Woodbine Gardens	43.706397	-79.309937
9	M5B	Downtown Toronto	Garden District, Ryerson	43.657162	-79.378937

By using the Foursquare API, I explore the boroughs and segment them and set the limit as **100 venues** and the radius **500 meters** for each borough from their given latitude and longitude information. Below is a part of the list of neighbourhoods, Venues name, category, latitude and longitude information from Foursquare API.

Table 2 Cleaned Geospatial Information of the neighbourhood in Toronto

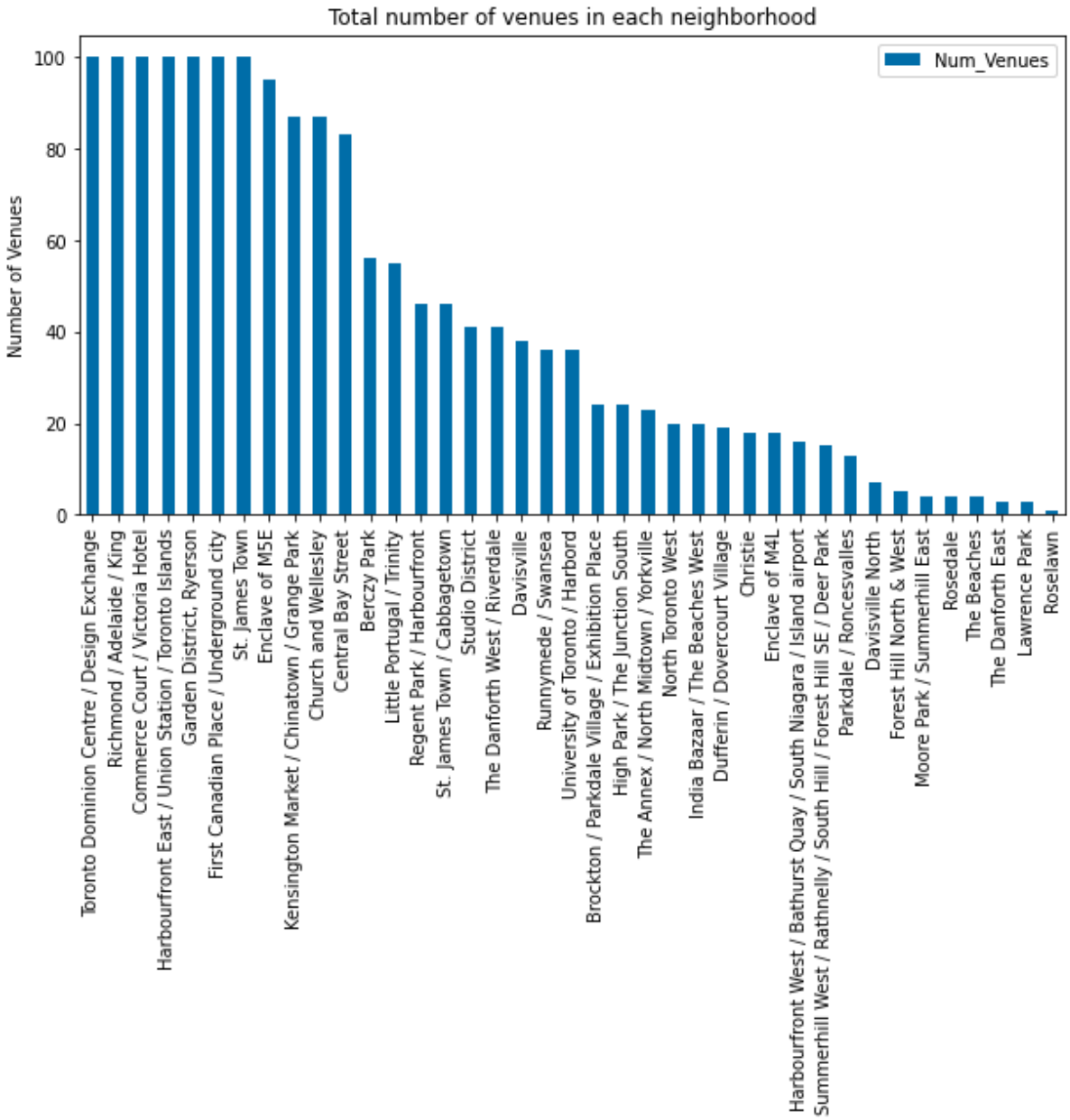
	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Regent Park / Harbourfront	43.65426	-79.360636	Roselle Desserts	43.653447	-79.362017	Bakery
1	Regent Park / Harbourfront	43.65426	-79.360636	Tandem Coffee	43.653559	-79.361809	Coffee Shop
2	Regent Park / Harbourfront	43.65426	-79.360636	Cooper Koo Family YMCA	43.653249	-79.358008	Distribution Center
3	Regent Park / Harbourfront	43.65426	-79.360636	Body Blitz Spa East	43.654735	-79.359874	Spa
4	Regent Park / Harbourfront	43.65426	-79.360636	Impact Kitchen	43.656369	-79.356980	Restaurant

The second step in our analysis will focus on most promising areas and within those create **clusters of locations that meet some basic requirements** with Machine Learning. For instance, the area with lots of venues but less Chinese restaurants or Asian restaurants. We will present a map of all such locations but also create clusters (using **k-means clustering**) of those locations to identify general zones / neighbourhoods which should be a starting point for final exploration and search for optimal venue location by stakeholders.

D. Data Analysis

There are 15 Borough and 98 Neighbourhoods in Toronto. However, since we only consider the central Toronto, Downtown, East, North and West Toronto has been analysed at this stage. There are 39 neighbourhoods and 240 unique categories of venues in Central Toronto. The figure below indicates the number of venues in each neighbourhood. It should be noted this result depends on the available Latitude and Longitude information. More neighbourhood's information could be revealed if given more Latitude and Longitude information.

Figure 1 Number of venues in each neighbourhood



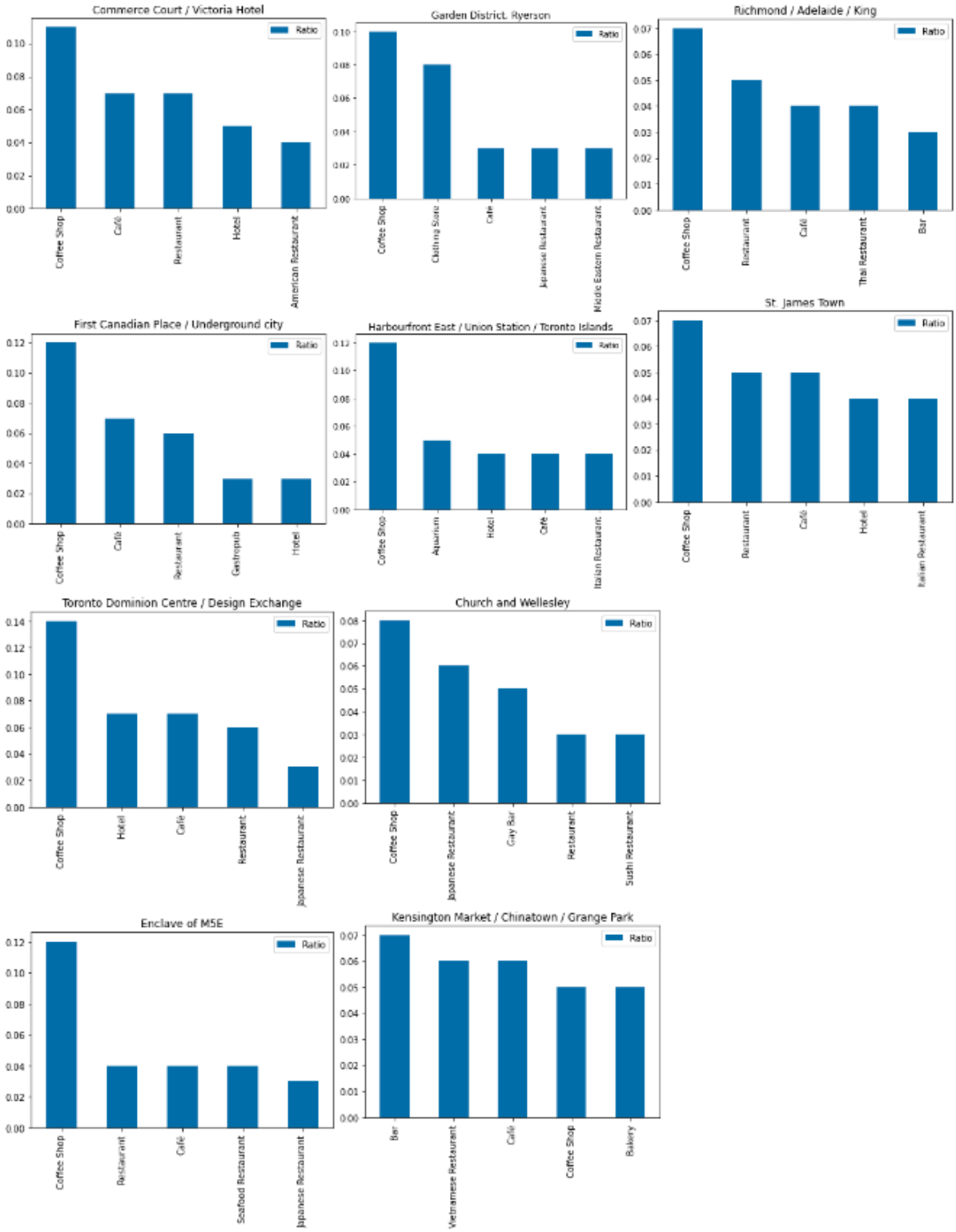
More detailed information about the top 10 venue category for each neighbourhood is shown below. It clear that each neighbourhood has various functions. For instance, CN Tower is the transportation hub while Central Bay Street could be a tourist spot since the majority of the venues are restaurants.

Table 3 Detailed information of top 10 venues in each neighbourhood

	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	Coffee Shop	Cocktail Bar	Bakery	Cheese Shop	Café	Restaurant	Farmers Market	Beer Bar	Seafood Restaurant	Museum
1	Brockton / Parkdale Village / Exhibition Place	Nightclub	Coffee Shop	Café	Breakfast Spot	Gym / Fitness Center	Bakery	Performing Arts Venue	Pet Store	Climbing Gym	Restaurant
2	CN Tower / King and Spadina / Railway Lands / ...	Airport Lounge	Airport Service	Airport Terminal	Plane	Airport	Airport Food Court	Airport Gate	Sculpture Garden	Harbor / Marina	Boutique
3	Central Bay Street	Coffee Shop	Italian Restaurant	Japanese Restaurant	Burger Joint	Ice Cream Shop	Sandwich Place	Gym / Fitness Center	Department Store	Restaurant	Middle Eastern Restaurant
4	Christie	Grocery Store	Café	Park	Gas Station	Italian Restaurant	Restaurant	Diner	Baby Store	Athletics & Sports	Nightclub
5	Church and Wellesley	Coffee Shop	Japanese Restaurant	Gay Bar	Sushi Restaurant	Restaurant	Gastropub	Hotel	Pub	Bubble Tea Shop	Men's Store

To close investigate the neighbourhood that we interest, top 5 venues for the neighbourhood with more than 80 venue categories have been plotted as below.

Figure 2 Top 5 venues for the neighbourhood with more than 80 venue categories



Regarding the restaurant information, there is 18% neighbourhood in Central Toronto do not have a restaurant which is because these areas do not have enough

venues to attract more visitors. There is a 77% neighbourhood does not have a Chinese restaurant. It is surprising that there are 7 areas with more than 50 category venues do not have even one Chinese restaurant. However, some of these places have similar Asian restaurants.

It turns out among the neighbourhoods that have more than 50 category venues, 'Enclave of M5E' and 'Berczy Park' do not have any Asian restaurant or Chinese restaurants while 'St. James Town' and 'Commerce Court / Victoria Hotel' only have one Asian restaurant. Therefore, these four neighbourhoods could be the top choices of the location to open the new Chinese restaurant.

Figure 3 Number of Restaurant vs Number of venues

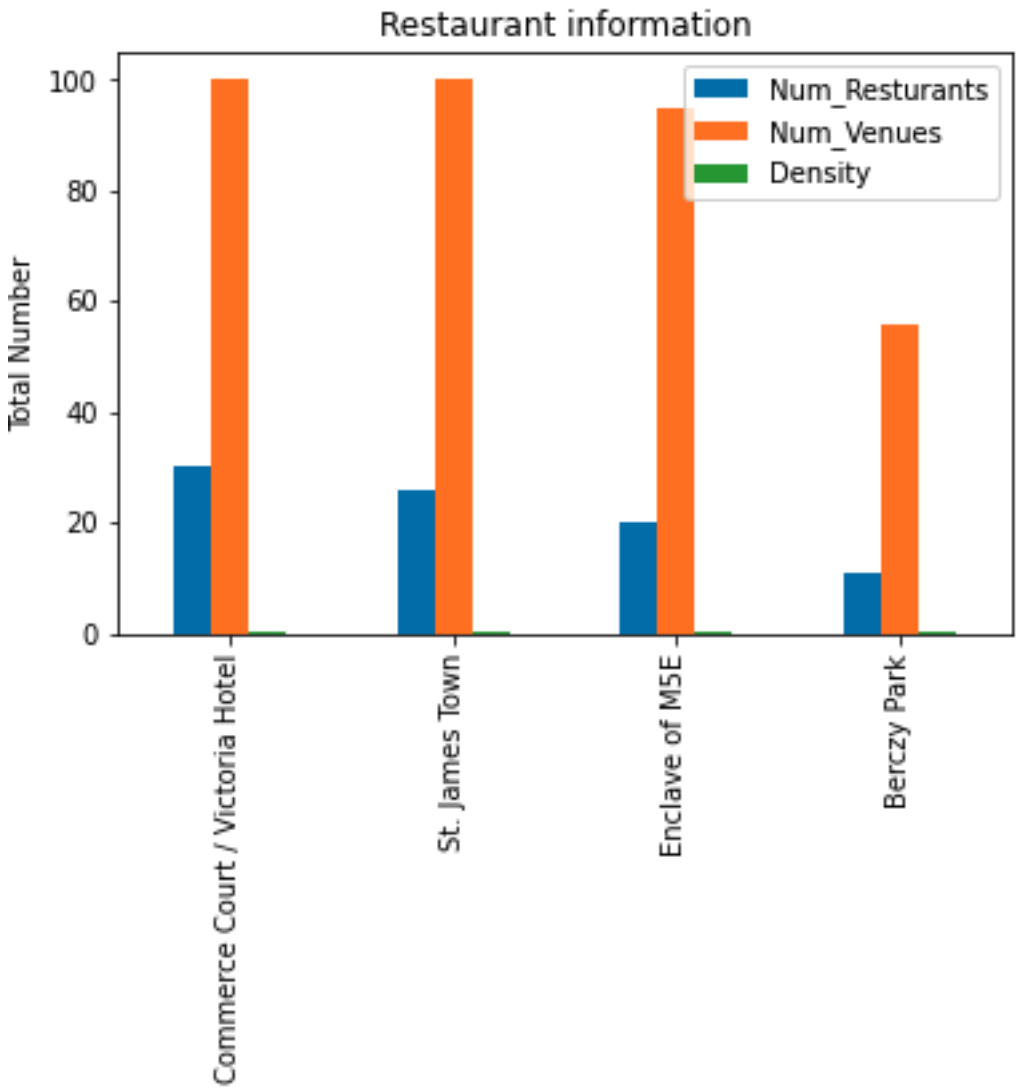
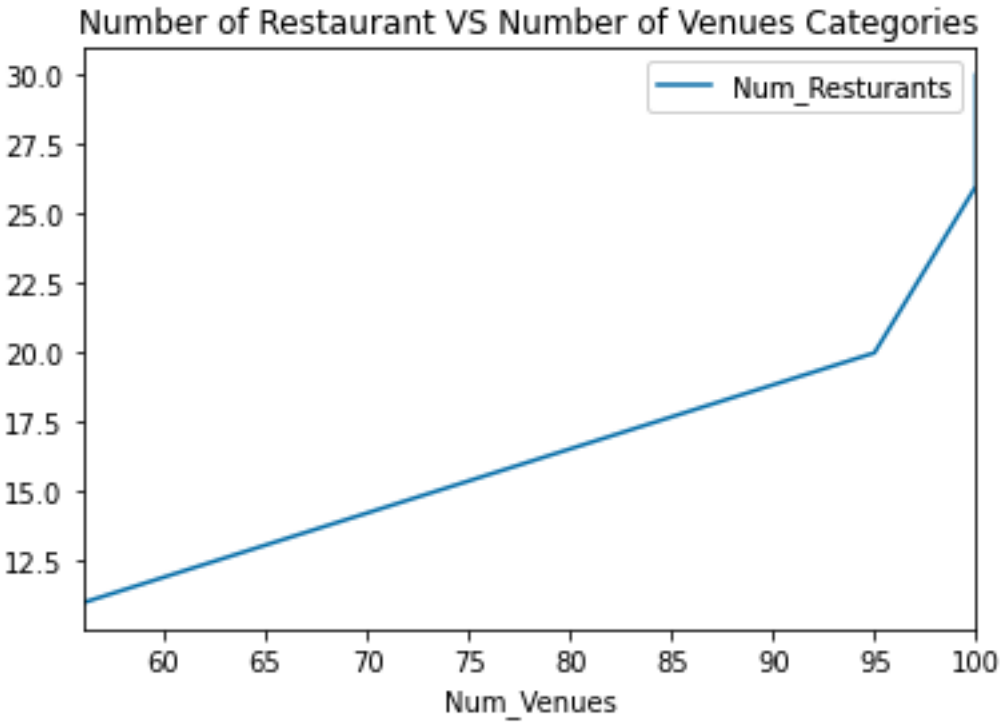


Table 4 Number of Restaurant vs Number of venues

	Neighborhood	Num_Resturants	Num_Venues	Density
0	Commerce Court / Victoria Hotel	30	100	0.300000
1	St. James Town	26	100	0.260000
2	Enclave of M5E	20	95	0.210526
3	Berczy Park	11	56	0.196429

Figure 4 Number of Restaurant vs Number of venues



It is clear that the number of restaurants is highly related to the number of venue categories. It seems like 'Enclave of M5E' could be the best choice since it has more venue categories but with relatively low density.

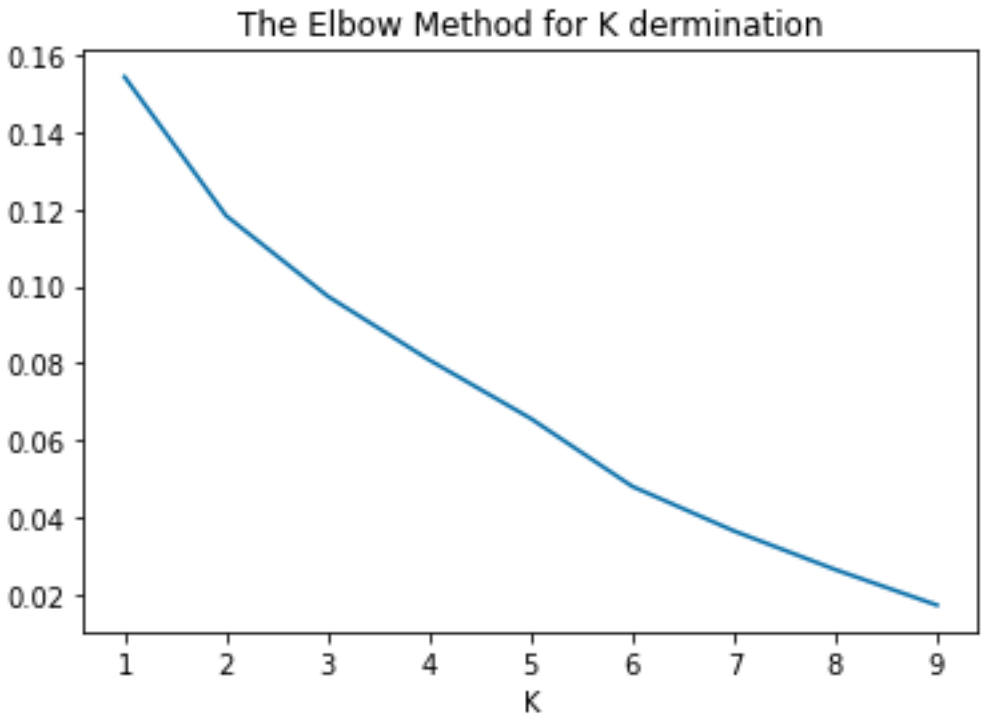
E. Machine Learning

After the exploratory analysis, it would be great to apply machine learning to reveal the insight based on the common venue categories in boroughs. As one of

the most common cluster methods, unsupervised learning **K-means Clustering** is applied to cluster the neighbourhood with the venues more than 50. The aim was to find the other potential locations similar to the three places based on the category of the venues.

First, optimal K for K-Means with Elbow Method is determined by calculating the Within-Cluster-Sum of Squared Errors (WSS) for different values of k. As shown by the figure below, 6 clusters have been chosen.

Figure 5 Elbow method for k determination



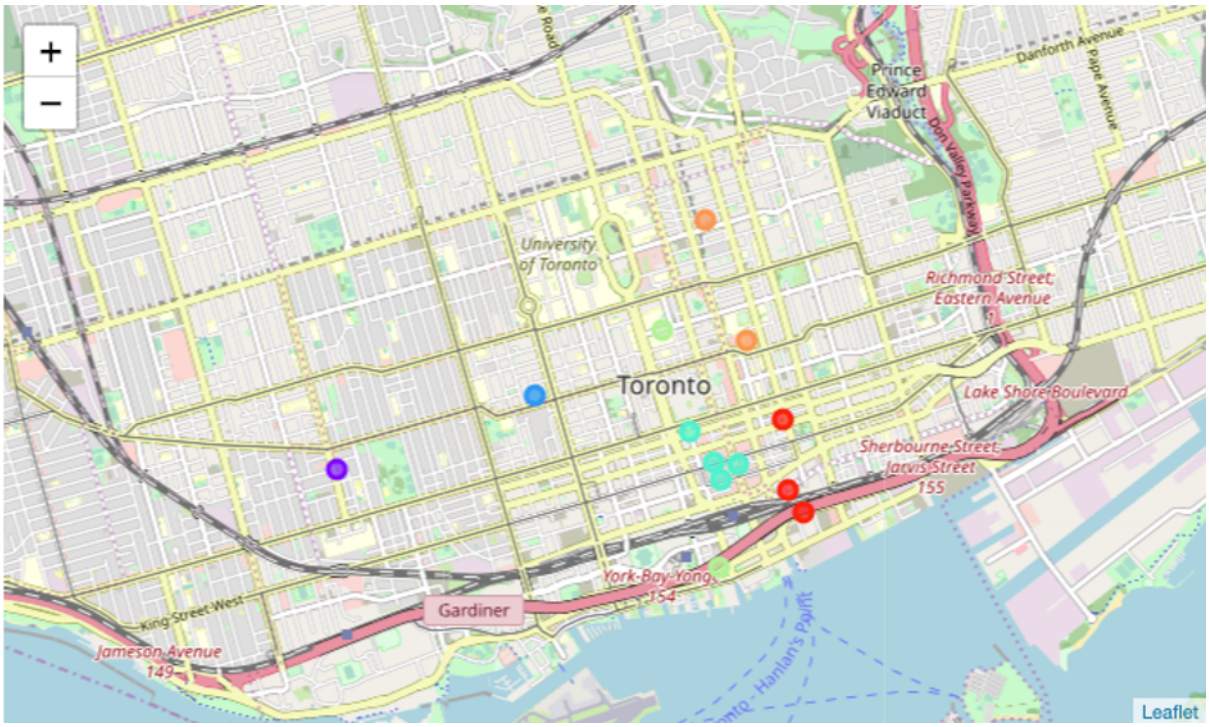
Here is part of a merged table with cluster labels for each borough. It is clear that the algorithm is able to label the neighbourhood with similar venues together. It is surprising that the algorithm clusters 'Enclave of M5E', 'Berczy Park' and 'St. James Town' together (with label 0) which are happened to be the three choices to open a Chinese restaurant based on the previous analysis.

Table 5 Illustration of clustering results

	Borough	Neighborhood	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 Common Venue	9th Most Common Venue	10th Most Common Venue	Num_Venues
	Downtown Toronto	St. James Town	43.651494	-79.375418	0	Coffee Shop	Café	Restaurant	Hotel	Italian Restaurant	Bakery	Cosmetics Shop	Diner	Beer Bar	Breakfast Spot	100.0
	Downtown Toronto	Berczy Park	43.644771	-79.373306	0	Coffee Shop	Cocktail Bar	Bakery	Cheese Shop	Café	Restaurant	Farmers Market	Beer Bar	Seafood Restaurant	Museum	56.0
	Downtown TorontoStn A PO Boxes25 The Esplanade	Enclave of M5E	43.646435	-79.374846	0	Coffee Shop	Restaurant	Café	Seafood Restaurant	Beer Bar	Japanese Restaurant	Hotel	Breakfast Spot	Farmers Market	Creperie	95.0

It is clearer to see the results by the map below. It is interesting that similar neighbourhoods are located close to each other. The neighbourhood with the label 3 (Cyan) located in the middle of the central Toronto while the revealed three potential choices with label 0(red) are located at the east side. It is understandable that the venues distributed in each neighbourhood are highly related to their locations.

Figure 6 Illustration of clustering results with map



F. Results and Discussion

Our analysis fist captured the top 10 common venues in each neighbourhood which shows that the neighbourhood with more than 50 venues have a coffee shop, cafe, bar and restaurant since these areas attract more tourist. While for the other neighbourhood with fewer venues for instance 'Parkdale,Roncesvalles' has the book shop and movie theatre as the top venue category. Although there is about 15% area do not have any restaurant, these places have not enough venues to attract tourists therefore they are not the ideal place to open a Chinese restaurant. While there is 74% neighbourhood in Central Toronto do not have the Chinese restaurant does not mean we have a lot of options. Because the areas that do not have Chinese restaurant may have similar Asian restaurant which increases the competition as well. Finally, we were able to find among the neighbourhood that with more than 50 venues, 'Enclave of M5E' and 'Berczy Park' do not have any Asian restaurant or Chinese restaurants; 'St. James Town' and 'Commerce Court / Victoria Hotel' only have one Asian restaurant. After calculating the density (number of restaurant/numbers of venue), "Enclave of M5E" would be a better choice since it has a relatively high number of venues and low restaurant density.

The results, of course, does not imply that those zones are actually optimal locations for a new restaurant! Purpose of this analysis was to only provide information on areas in the Toronto centre but not crowded with existing restaurants (particularly Chinese or Asian restaurants). The recommended zones should, therefore, be considered only as a starting point. A more detailed analysis should be conducted to find the location which has not only no nearby competition, but also other factors taken into account and all other relevant conditions met. For instance: 1) The spot should have the visibility and accessibility to make sure the restaurant can be seen by those driving or walking

by. 2) Ensure the target market of your restaurant matches the demographics of the area. 3) To consider the leasing and labour cost.