



Battle of Neighborhood Restaurant Business in Toronto

IBM Data Science Professional Certificate – Capstone Project

Date : 12 May 2021
SHROWTHI Bharadwaja

Introduction

- ❖ Toronto is the capital city of the Canadian province of Ontario. It is the most populous city in Canada and the fourth most populous city in North America.
- ❖ The cuisine of Toronto reflects Toronto's size and multicultural diversity.
- ❖ Canadian cuisine varies widely depending on the regions of the nation. The four earliest cuisines of Canada have indigenous, English, Scottish and French roots. The traditional cuisine of English Canada is closely related to British cuisine.
- ❖ Overtime, with subsequent waves of immigration in the 19th and 20th centuries, Canadian food has been shaped and impacted by those of indigenous people, settlers, and immigrants.
- ❖ Different ethnic neighborhoods throughout the city focus on variety of cuisines. Examples: Chinese, Indian, Italian, Japanese, Caribbean, Jewish, Vegetarian/Vegan, American, Mediterranean, Fast Food Centers etc.
- ❖ A number of culinary festivals take place in Toronto each year. Any trip to Toronto is incomplete without checking out its food tours. Going on any one of these, will without a doubt, leave you satiated.



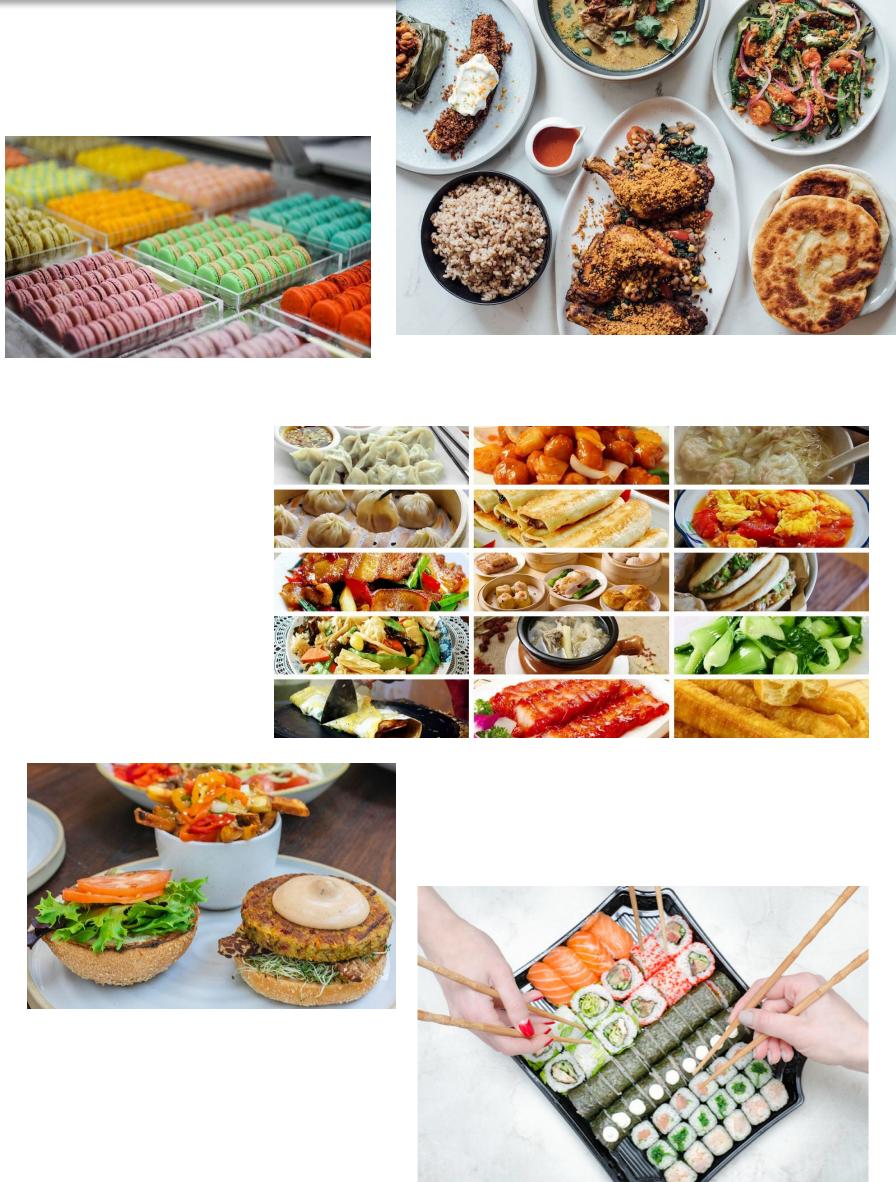
References:

https://en.wikipedia.org/wiki/Cuisine_in_Toronto

<https://tastytourstoronto.com>

Business Problem

- ❖ Toronto is one of the most densely populated areas in Canada with an estimated population of ~ 6 million.
- ❖ The city is home of the finest and most diverse haute cuisine restaurants in the world.
- ❖ Toronto is well known for its great food. Canadian culinary includes an array of international cuisines influenced by multiculturalism of the town.
 - Example: Chinese, Indian, Italian, Japanese, Caribbean, Jewish, Vegetarian/Vegan, American, Mediterranean, Fast Food Centers etc.
- ❖ The objective of this project is:
 - to use Foursquare location data and regional clustering of venue information to determine the ‘best’ neighborhood in Toronto to open a new restaurant with a specific cuisine style.
 - to evaluate the most suitable neighborhood in Toronto for an entrepreneur to invest in a restaurant business. This analysis will provide vital information that can be used by the target audience (Entrepreneurs).



❖ Objective:

1. To recommend the best neighborhood for a restaurant business in Toronto
1. To understand the similarities and differences between the neighborhoods using Unsupervised K-Mean Clustering Algorithm.

❖ Data Sources:

1. Neighborhoods of the city of Toronto for it's Boroughs and Venues.
[https://en.wikipedia.org/w/index.php?title=List_of_postal_codes_of_Canada:_M&oldid=1008658627'](https://en.wikipedia.org/w/index.php?title=List_of_postal_codes_of_Canada:_M&oldid=1008658627)
2. Geo-coordinates data for each Neighborhood in Toronto is collected from the following csv file:
https://cocl.us/Geospatial_data
3. Foursquare Venues data for Restaurants and their geographical location using Client credentials.

❖ Data Cleaning (or preprocessing):

- Foursquare API gives latitude and longitude values for specified venues covered within a given circle of radius.
- The foursquare API data will be merged with Toronto's Borough's and Neighborhoods data frame.
- All the NaN for venues are not considered in this analysis
- Boroughs with NaN are replaced with Neighborhoods names.

❖ Feature Selection:

- Toronto has 10 unique Boroughs and has 103 Neighborhoods.
List of Boroughs: 'Scarborough', 'North York', 'East York', 'East Toronto', 'Central Toronto', 'Downtown Toronto', 'York', 'West Toronto', 'Mississauga', 'Etobicoke'
- Out of these, only Boroughs that contains “Toronto” word are chosen for this analysis
- The top venues are retrieved using “group by neighborhoods” and Foursquare API . This list is further filtered for restaurants with the cuisine style.

❖ Analytical Method:

- Unsupervised K-Mean Clustering Algorithm is implemented to segregate the neighborhoods and the respective restaurants distribution per neighborhood.
- Total 5 clusters are considered for this analysis.
- Once the data frame is clustered, only top 10 most common restaurant venues are considered for their geographic distribution per cluster
- Based on the restaurant and their cuisine style distribution per neighborhood, the entrepreneur can consider the suitable location to start a restaurant. The new restaurant cuisine style can be decided based on the existing restaurant distribution. The choice of cuisine style is an option for the entrepreneur investment strategy.

Link: [Capstone Project: Python Code on Github](#)

Python Packages and Dependencies

❖ The following python libraries (but not limited) are mainly used for the analysis:

- | | |
|----------------|---|
| Pandas | - Library for Data Analysis |
| NumPy | - Library to handle data in a vectorized manner |
| JSON | - Library to handle JSON files |
| Geopy | - To retrieve Location Data |
| Requests | - Library to handle http requests |
| Matplotlib | - Python Plotting Module |
| Sklearn | - Python machine learning Library |
| Folium | - Map rendering Library |
| Beautiful Soup | - Web Scraping and Data Wrangling |
| Foursquare API | - To collect the top trending venues |

Link: [Capstone Project: Python Code on Github](#)

- ❖ Canadian Postal Codes and Neighborhood are scraped from Wikipedia website using BeautifulSoup

Link: [Canadian Postal Codes and Neighborhoods](#)

- ❖ All the Neighborhoods that are not assigned are removed
- ❖ Combined the neighborhoods with same postal codes
- ❖ Neighborhoods that are not assigned are replaced with their Borough's name
- ❖ The resultant Pandas data frame size is:

103 rows x 3 columns

	Postal Code	Borough	Neighborhood
0	M3A	North York	Parkwoods
1	M4A	North York	Victoria Village
2	M5A	Downtown Toronto	Regent Park, Harbourfront
3	M6A	North York	Lawrence Manor, Lawrence Heights
4	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government
...
98	M8X	Etobicoke	The Kingsway, Montgomery Road, Old Mill North
99	M4Y	Downtown Toronto	Church and Wellesley
100	M7Y	East Toronto	Business reply mail Processing Centre, South C...
101	M8Y	Etobicoke	Old Mill South, King's Mill Park, Sunnylea, Hu...
102	M8Z	Etobicoke	Mimico NW, The Queensway West, South of Bloor,...

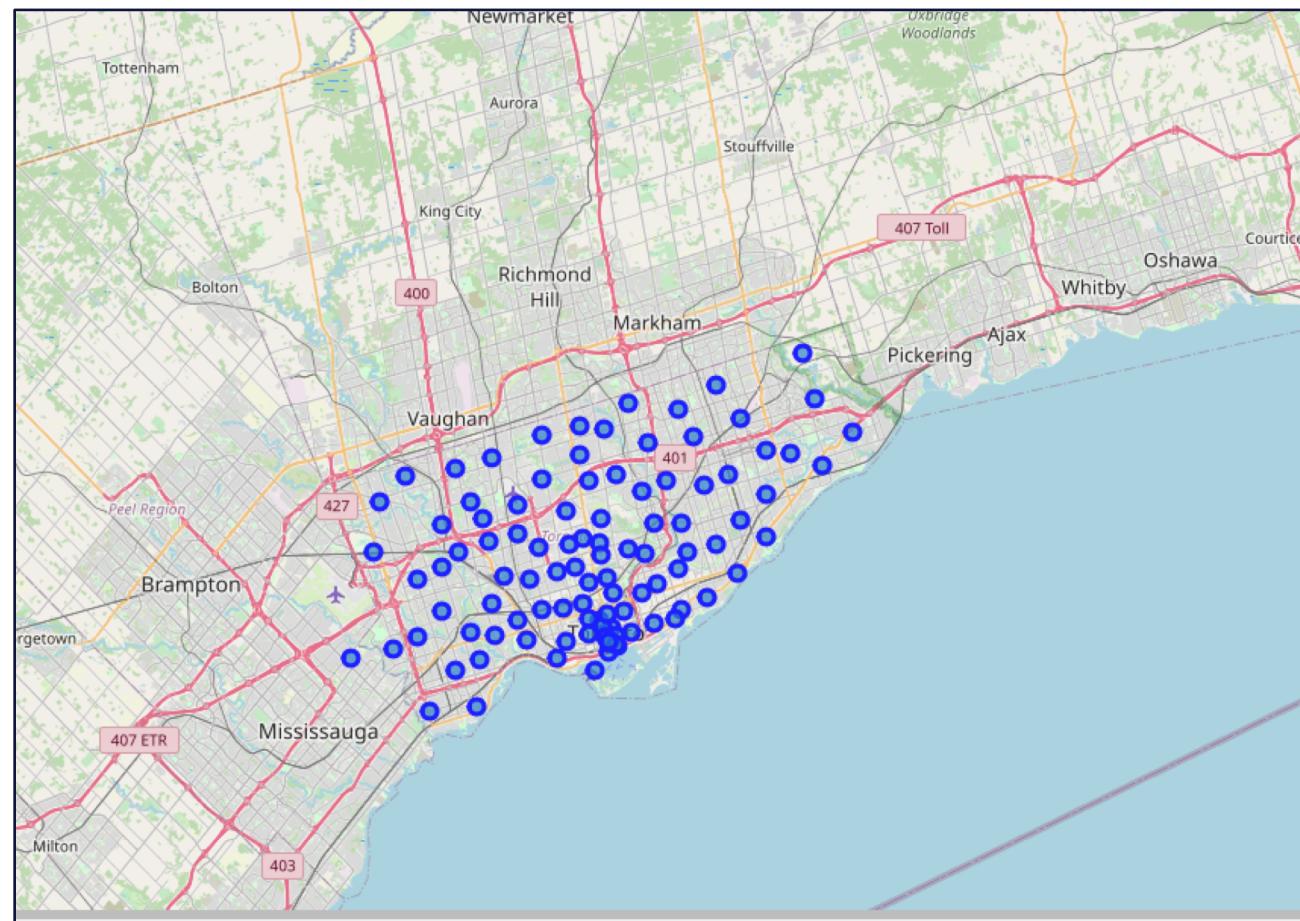
Merge the Canadian Neighborhood data frame with respective latitudes and longitudes

- ❖ Geospatial data (longitudes and latitudes) for Canadian provinces are combined with respective postal codes.
- ❖ The geospatial data is collected from
Link: [Canadian Geospatial Data](#)
- ❖ The resultant Pandas data frame size is: 103 rows x 5 columns

	Postal Code	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	Downtown Toronto	Queen's Park, Ontario Provincial Government	43.662301	-79.389494
...
98	M8X	Etobicoke	The Kingsway, Montgomery Road, Old Mill North	43.653654	-79.506944
99	M4Y	Downtown Toronto	Church and Wellesley	43.665860	-79.383160
100	M7Y	East Toronto	Business reply mail Processing Centre, South C...	43.662744	-79.321558
101	M8Y	Etobicoke	Old Mill South, King's Mill Park, Sunnylea, Hu...	43.636258	-79.498509
102	M8Z	Etobicoke	Mimico NW, The Queensway West, South of Bloor,...	43.628841	-79.520999

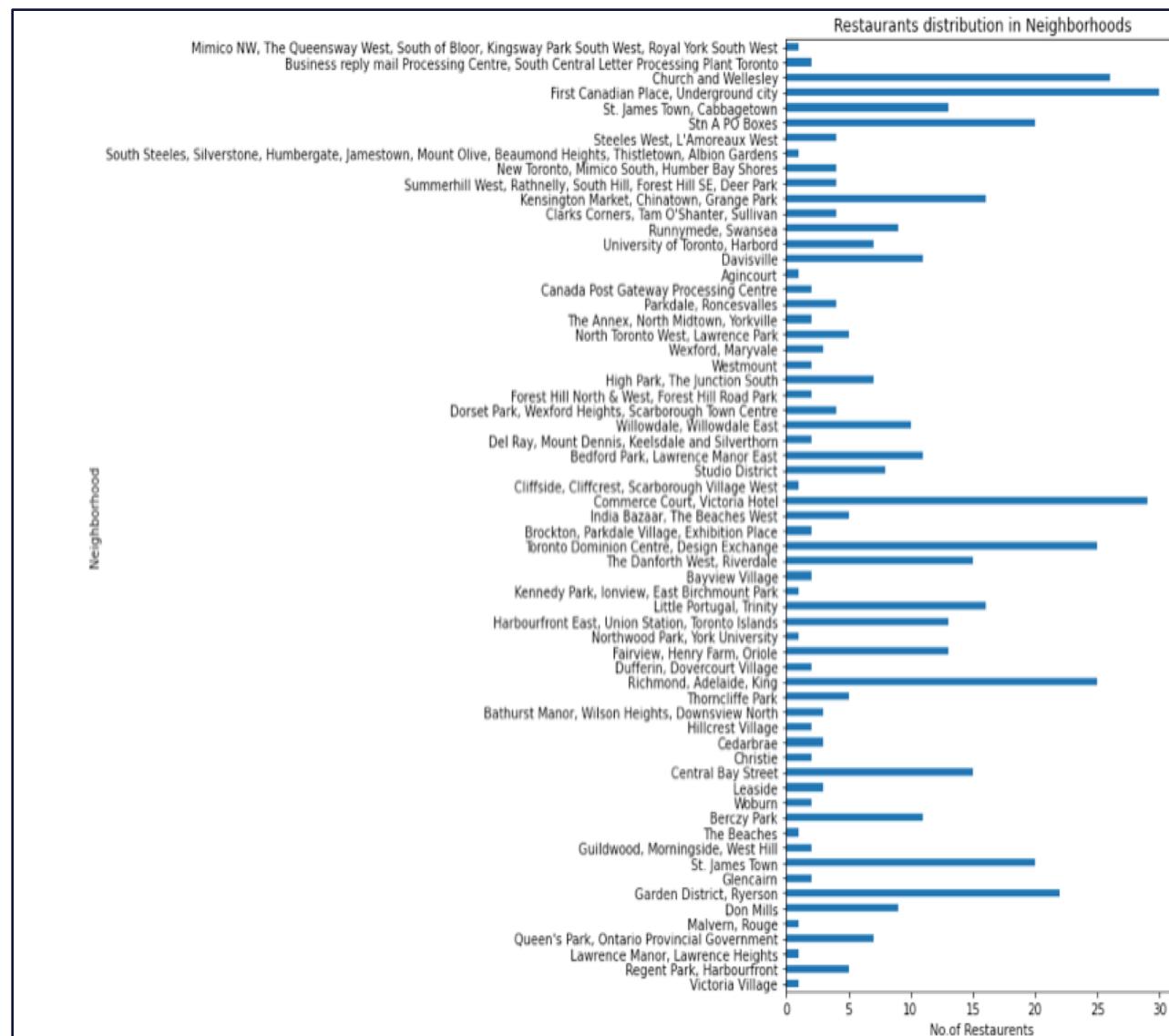
Analyzing Toronto Geographical Coordinates and Neighborhoods

- ❖ Toronto's geographical coordinates are extracted using google geocoder with latitudes and longitudes.
- ❖ A map of Toronto with Toronto's neighborhoods, Boroughs is visually analyzed with their geospatial data.
- ❖ Total neighborhoods in 10 Boroughs are: 103
- ❖ Folium library is used for this purpose



Restaurant Distribution Frequency per Neighborhood in Toronto

- ❖ Foursquare API is used with required client credentials to exact 200 top venues and their categories around Canadian Neighborhoods within 5 km radius from its center.
- ❖ The data is stored into an url format for subsequent analysis:
Link: '[Top 200 venues within 5000 meters radius](#)'
- ❖ Data is filtered for “Restaurant” venue category per neighborhood (grouped) after necessary data pre-processing steps.
- ❖ The resultant data frame size: 63 x 49



Top 10 Restaurant Venues and Cuisine Style

❖ Top 10 restaurant venues per neighborhood are sorted with one hot encoding method.

	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	Latin American Restaurant	Vietnamese Restaurant	Hakka Restaurant	Gluten-free Restaurant	German Restaurant	French Restaurant	Filipino Restaurant	Fast Food Restaurant	Falafel Restaurant	Ethiopian Restaurant
1	Bathurst Manor, Wilson Heights, Downsview North	Sushi Restaurant	Restaurant	Middle Eastern Restaurant	Vietnamese Restaurant	Doner Restaurant	German Restaurant	French Restaurant	Filipino Restaurant	Fast Food Restaurant	Falafel Restaurant
2	Bayview Village	Japanese Restaurant	Chinese Restaurant	Vietnamese Restaurant	Dumpling Restaurant	Gluten-free Restaurant	German Restaurant	French Restaurant	Filipino Restaurant	Fast Food Restaurant	Falafel Restaurant
3	Bedford Park, Lawrence Manor East	Restaurant	Italian Restaurant	Indian Restaurant	Comfort Food Restaurant	Fast Food Restaurant	Greek Restaurant	American Restaurant	Sushi Restaurant	Thai Restaurant	Sri Lankan Restaurant
4	Berczy Park	Seafood Restaurant	Restaurant	Greek Restaurant	Vegetarian / Vegan Restaurant	Thai Restaurant	French Restaurant	Japanese Restaurant	Comfort Food Restaurant	Eastern European Restaurant	Dumpling Restaurant

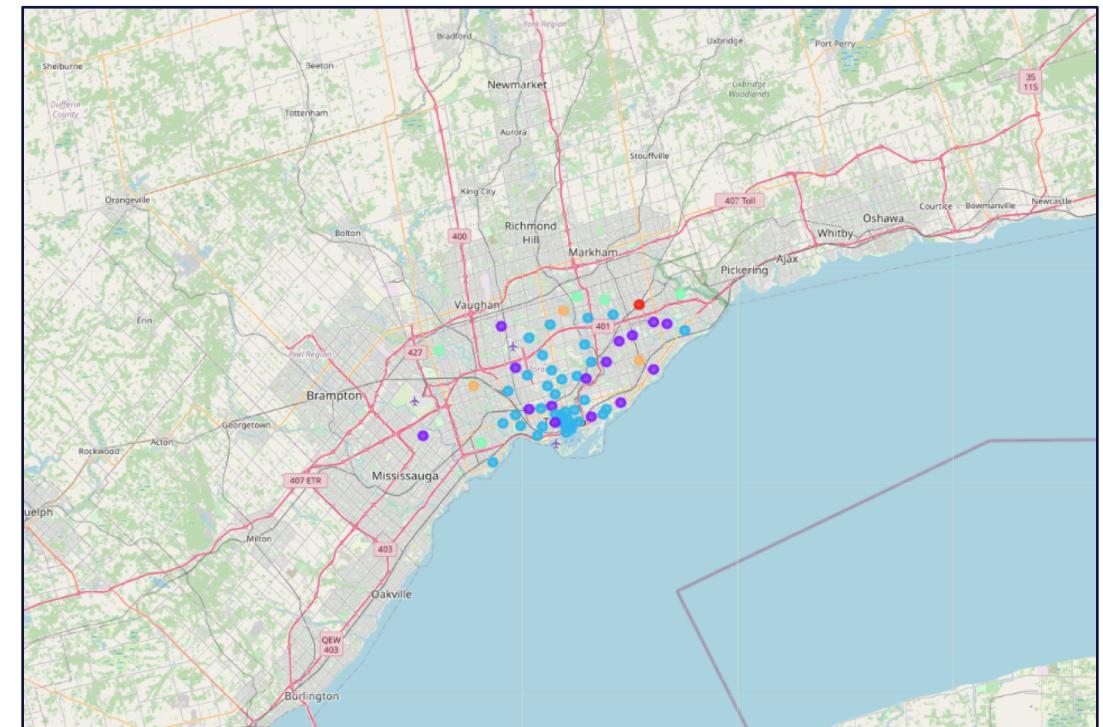
❖ Grouping restaurants based on their cuisine style per Neighborhood in Toronto

Postal Code	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	
0	M3A	North York	Parkwoods	43.753259	-79.329656	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
1	M4A	North York	Victoria Village	43.725882	-79.315572	1.0	Portuguese Restaurant	Vietnamese Restaurant	Doner Restaurant	Gluten-free Restaurant	German Restaurant	French Restaurant	Filipino Restaurant	Fast Food Restaurant	Falafel Restaurant	Ethiopian Restaurant
2	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636	2.0	Restaurant	Asian Restaurant	French Restaurant	Mexican Restaurant	Vietnamese Restaurant	Dumpling Restaurant	Gluten-free Restaurant	German Restaurant	Filipino Restaurant	Fast Food Restaurant
3	M6A	North York	Lawrence Manor, Lawrence Heights	43.718518	-79.464763	1.0	Vietnamese Restaurant	Vegetarian / Vegan Restaurant	Greek Restaurant	Gluten-free Restaurant	German Restaurant	French Restaurant	Filipino Restaurant	Fast Food Restaurant	Falafel Restaurant	Ethiopian Restaurant
4	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government	43.662301	-79.389494	2.0	Sushi Restaurant	Vegetarian / Vegan Restaurant	Italian Restaurant	Japanese Restaurant	Portuguese Restaurant	Mexican Restaurant	Vietnamese Restaurant	Dumpling Restaurant	French Restaurant	Filipino Restaurant

This is the required data frame for Unsupervised K Means Clustering of Restaurant Business in Toronto

Cluster Analysis on Neighborhood vs. Restaurant Distribution

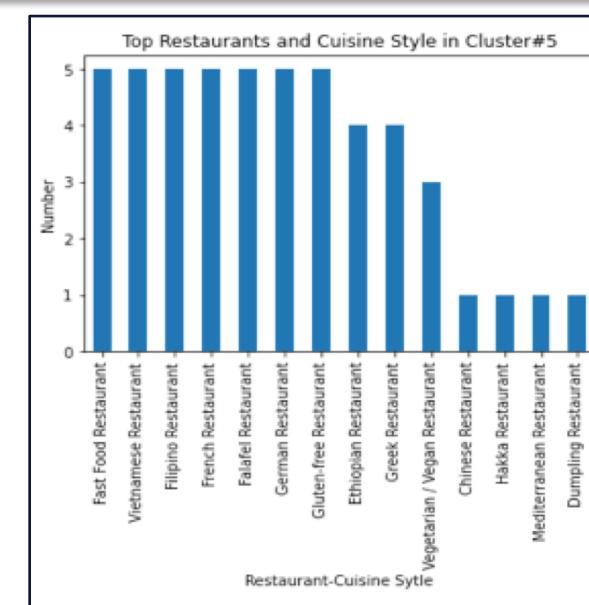
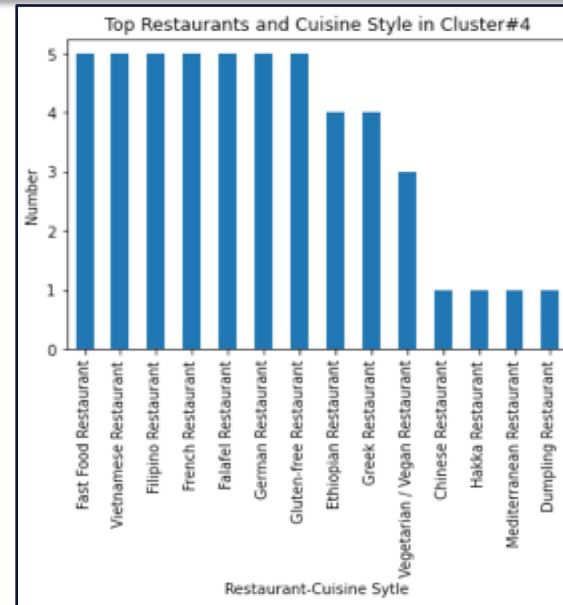
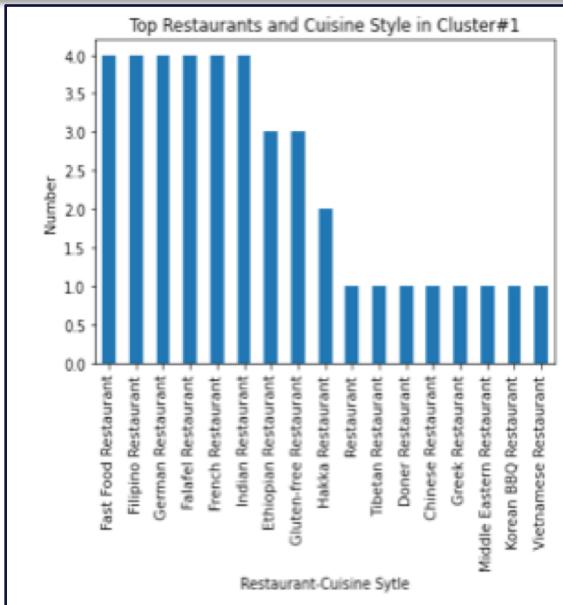
- ❖ K Means Cluster Analysis algorithm is used for cluster analysis
- ❖ Cluster number is fixed at 5
- ❖ These clusters are visualized on Toronto City map using Folium library



Cluster Labels	Postal Code	Borough	Neighborhood	Latitude	Longitude	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
	0.0	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1.0	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15
2.0	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40
3.0	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
4.0	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3

Neighborhood vs. Restaurant Distribution

Cluster Analysis -1

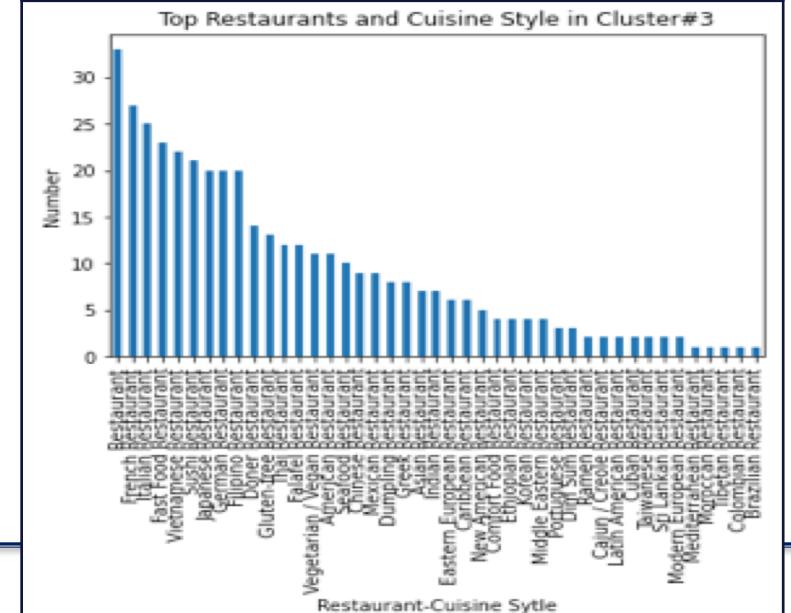
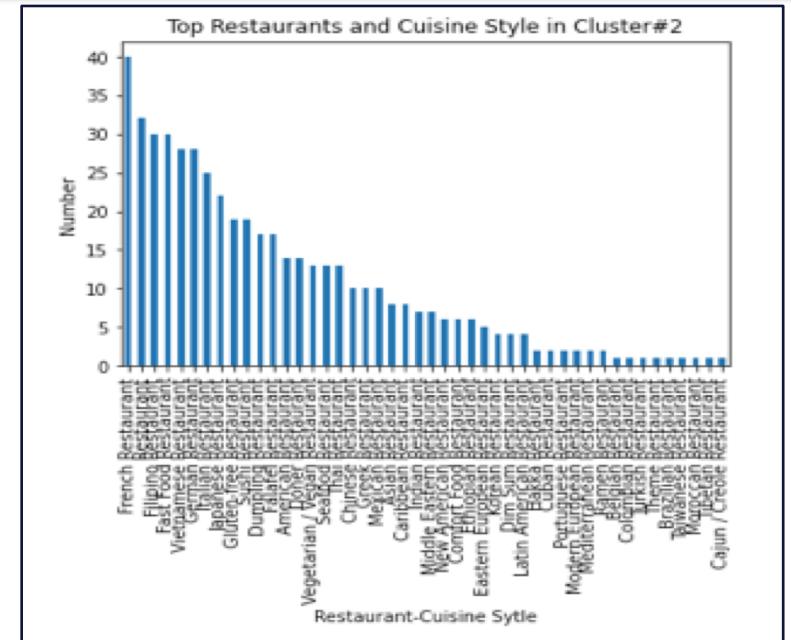


- ❖ Cluster#1, #4, and #5 has small number of restaurants. The competition can be low from other restaurants to start a new restaurant these areas.
- ❖ High standards for food quality and customer treatment can be implemented for a successful business.
- ❖ However, for a successful restaurant business customer visitations are very important. Further analysis of local attractions, events, demographic distribution, and other businesses are very important for customer turn around frequency in these areas. An thorough analysis considering these factors can enable an entrepreneur to take a decision to start a business for better return on investment (ROI).

Neighborhood vs. Restaurant Distribution Cluster Analysis -2

- ❖ Cluster#2 and 3: Both these clusters have high number of restaurants. Probably, these are commercially busy areas for tourists and customers. The prospects for a successful new business can be high in these two clusters.
- ❖ There are multiple cuisine styles in these two clusters. Customers can enjoy the variety of cuisines available.
- ❖ The competition can be high for an entrepreneur to start a new restaurant business
- ❖ However, the entrepreneur can choose a cuisine style that has low frequency in these clusters to attract customers. **Demographic distribution also plays a role**
- ❖ In addition to cuisine style, **the quality, ambience and customer service play a huge role to choose the cuisine style and to start a new restaurant business.**

Link: [Capstone Project: Python Code on Github](#)



- ❖ The primary objective of this analysis is to recommend a suitable neighborhood in Toronto City for a new restaurant business. Using unsupervised K Means cluster algorithm, it is shown that restaurant business and choice of cuisine style is determined by local restaurant competition.
- ❖ This analysis can give a probable cluster to start a new restaurant with specific cuisine style. Some areas can have high number of restaurants with a wide choice of cuisine style. In such areas, choosing a cuisine style can play a good role. Of course this analysis also requires local demographic distribution for regular customers.
- ❖ Cluster#2 and #3 have large number of restaurants. Here, quality of service and food taste at a affordable price can play a big role.
- ❖ Cluster#1,#4,#5 have small number of restaurants. Even though the local competition is small in these areas, it is important to consider customer visits based on local situation.
- ❖ At the end it is imperative quality of service and food taste matters for successful restaurant business in addition to customer turn around frequency, local businesses, events, attractions. The analysis can be extended considering these sets of data for better analysis.

Link: [Capstone Project: Python Code on Github](#)

THANK YOU