Capstone Project – Battle of Neighborhoods (Week 2): Final Report

Determining the top Japanese Restaurants in Toronto using K-means clustering algorithm

By Hoo Chee Hau

4 MAY 2020

1. INTRODUCTION

1.1 Background

Toronto is by far the largest city in Canada with a population of 2,731,571, based on 2016 census conducted by Statistics Canada, which has seen an increase of 4.5% in population growth since 2011. Being a provincial capital of Ontario, Toronto is geographically located on the north-western shore of Lake Ontario.

Major districts in Toronto include Downtown, Central, East, West and North Toronto that are used to be those of the Old Toronto, and the result of amalgamation and expansion has seen inclusion of surrounding districts, such as East York, Etobicoke, North York, York, and Scarborough. Toronto Pearson International Airport is located at Mississauga, a suburban municipality adjacent to Etobicoke.

The city of Toronto is considered as an important international business and financial center in Canada where its economic strength is underscored by the Toronto Stock Exchange, the 7th largest stock exchange bourse in the world by market capitalization.

Toronto is home to a sizeable minority groups of diverse ethnic origins stemmed from active immigration from different parts of the world since 1980s with more than 150 different languages spoken and therefore, this city is recognized as one of the most cosmopolitan and vibrant cities in North America.

According to an article written by Corey Mintz in Globe and Mail dated 3 January 2020, Japanese cuisine has blossomed in cosmopolitan Toronto since 2012, picking up the trend from Vancouver, where Japanese businesses have been flourishing and expanding to Toronto, taking advantage of its bigger market. In recent years, Toronto has seen Japanese izakaya groups such as Guu and Kingyo set up business, with former employees breaking off to open places such as Imanishi and Ramen Ishin.

1.2 Business Problem

1.2.1 Target Group

This study focussed primarily on the needs of a specific target group, namely - business travellers, tourists and short-term visitors to Toronto who are not familiar with the city and are looking for general recommendations on Japanese restaurants to dine-in.

1.2.2 Objective

Hence, the objective of this study was to provide reliable and well-researched information in addressing the requirements of the aforementioned target group outlined as below:-

a. List of Japanese restaurants in Toronto.
b. Boroughs in Toronto that have a significant presence of Japanese restaurants and the borough that has the highest number of Japanese restaurants.
c. Ranking of the Japanese restaurants in Toronto based on customers' sentiments.
d. Recommendation on the top 3 Japanese restaurants in Toronto.

By and large, this study endeavored a novel approach of applying data science and machine learning techniques, which is the showcase of this capstone project, to achieve its objective.

This project was done entirely in Anaconda Jupyter Notebook with Python 3.0 coding.

2. DATA ACQUISITION & CLEANING

In this section, the description on how the data was acquired, with the source of data from where it was obtained and how the data was cleansed, is detailed in the subsections as below:-

2.1 Data sources

The first dataset that was acquired was Toronto's FSA (Forward Sortation Area) 3-letter codes which can be found from Wikipedia's webpage titled - 'List of postal codes of Canada: M' with the URL given as:-https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M. According to Canada Post, Toronto's FSA code begins with the letter 'M'.

A process known as webscraping was performed to scrape the relevant data related to Toronto's FSA postal codes and the corresponding names of boroughs and neighborhoods from this Wikipedia webpage and convert those to a Pandas dataframe in the project. A preview of this raw dataset in shown in Figure 1. This dataframe contains 180 rows and 3 columns.

	Postal code	Borough	Neighborhood
0	M1A	Not assigned	NaN
1	M2A	Not assigned	NaN
2	M3A	North York	Parkwoods
3	M4A	North York	Victoria Village
4	M5A	Downtown Toronto	Regent Park / Harbourfront
5	M6A	North York	Lawrence Manor / Lawrence Heights
6	M7A	Downtown Toronto	Queen's Park / Ontario Provincial Government
7	M8A	Not assigned	NaN
8	M9A	Etobicoke	Islington Avenue
9	M1B	Scarborough	Malvern / Rouge
10	M2B	Not assigned	NaN

***Figure 1**: Preview of dataframe with columns: Toronto postal code FSA, Borough and Neighborhood scraped from Wikipedia*

Several issues were found during the initial inspection of the data as described below:-

a. There were considerable number of rows containing 'Not assigned' for Boroughs and 'NaN' for Neighborhood
b. Several neighborhood names in Neighborhood had delimiters of '/' and had to be replaced with ','.

These data issues were dealt with during the data cleaning process and that will be described in detail in the next subsection.

The second acquired dataset was the geographical coordinates in Longitude and Latitude corresponding to Toronto's postal code FSA which were downloaded from : http://cocl.us/Geospatial_data in comma delimited text format. A preview of this raw dataset is shown in Figure 2. This dataframe contains 103 rows and 3 columns.

	Postal Code	Latitude	Longitude
0	M1B	43.806686	-79.194353
1	M1C	43.784535	-79.160497
2	M1E	43.763573	-79.188711
3	M1G	43.770992	-79.216917
4	M1H	43.773136	-79.239476
5	M1J	43.744734	-79.239476
6	M1K	43.727929	-79.262029
7	M1L	43.711112	-79.284577
8	M1M	43.716316	-79.239476
9	M1N	43.692657	-79.264848

***Figure 2**: Preview of dataframe with columns: Toronto postal code FSA, Latitude & Latitude table imported from CSV file*

This dataset was found to have no data issues.

2.2 Data cleaning and processing

In the first dataframe, rows containing 'Not assigned' for Boroughs and 'NaN' for Neighborhood were dropped mainly because these data were irrelevant in the study.

In the same dataframe, neighborhood names in Neighborhood which had delimiters of '/' were replaced with ','. A comma ',', is considered as a proper delimiter compared with a slash '/'.

A preview of the first dataframe after data cleaning process is shown in Figure 3.

	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
5	M9A	Etobicoke	Islington Avenue
6	M1B	Scarborough	Malvern , Rouge
7	M3B	North York	Don Mills
8	M4B	East York	Parkview Hill , Woodbine Gardens
9	M5B	Downtown Toronto	Garden District, Ryerson
10	M6B	North York	Glencairn

***Figure 3**: The first dataframe after data cleaning process*

In the next step, the two aforementioned dataframes were merged to create a new expanded dataframe that contains columns as the following: Postcode, Borough, Neighborhood, Longitude & Latitude. The preview of this new dataframe is shown in Figure 4.

	Postcode	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
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

***Figure 4**: Preview of the merged dataframe*

This merged dataframe contains 10 Boroughs and 103 Neighborhoods of Toronto.

3. METHODOLOGY

3.1 Exploratory Data Analysis

The 103 neighborhoods with the corresponding latitudes and longitudes from the merged dataframe are overlain on Toronto map created from python folium library for visualization as shown in Figure 5.

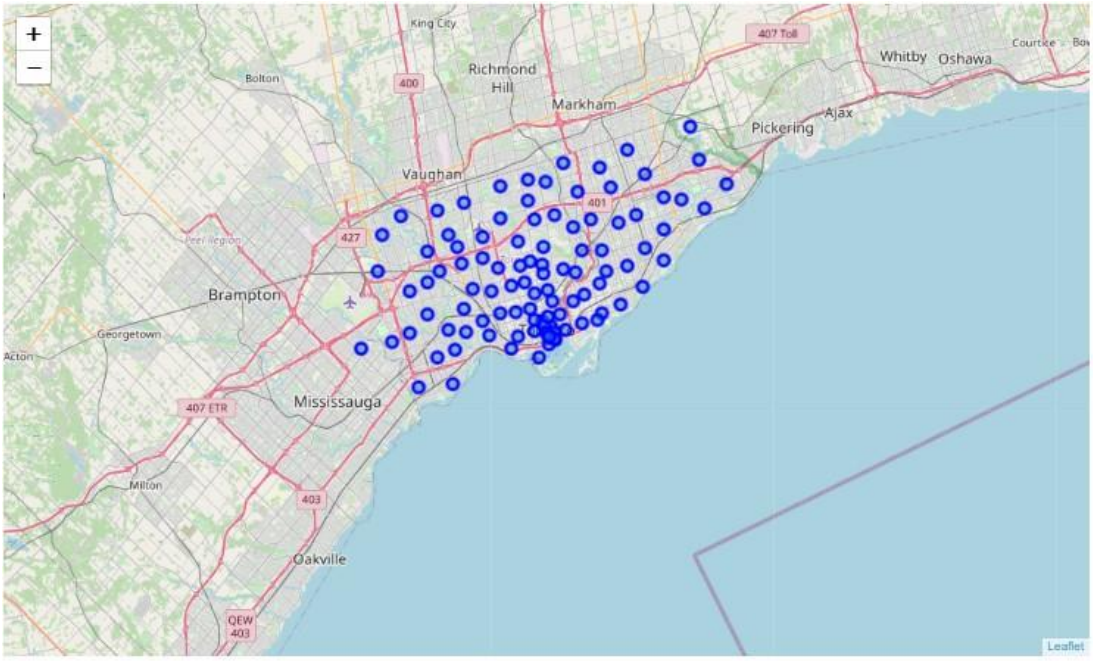

***Figure 5**: 103 neighborhoods with coordinates in Latitude and Longitude are overlain on Toronto map*

Next, the numbers of the neighborhoods for each borough were analyzed. Boroughs that have the highest number of neighborhoods were determined as shown in Figure 6.

Borough	Neighborhood Count
North York	24
Downtown Toronto	19
Scarborough	17
Etobicoke	12
Central Toronto	9
West Toronto	6
East Toronto	5
East York	5
York	5
Mississauga	1

***Figure 6**: Table showing Boroughs with Neighborhood Count sorted in descending order*

From Figure 6, North York has the highest neighborhood count (24), followed by Downtown Toronto (19), and Scarborough (17). In terms of percentage, North York accounts for 23.3%, Downtown Toronto 18.4% while Scarborough 16.5%.

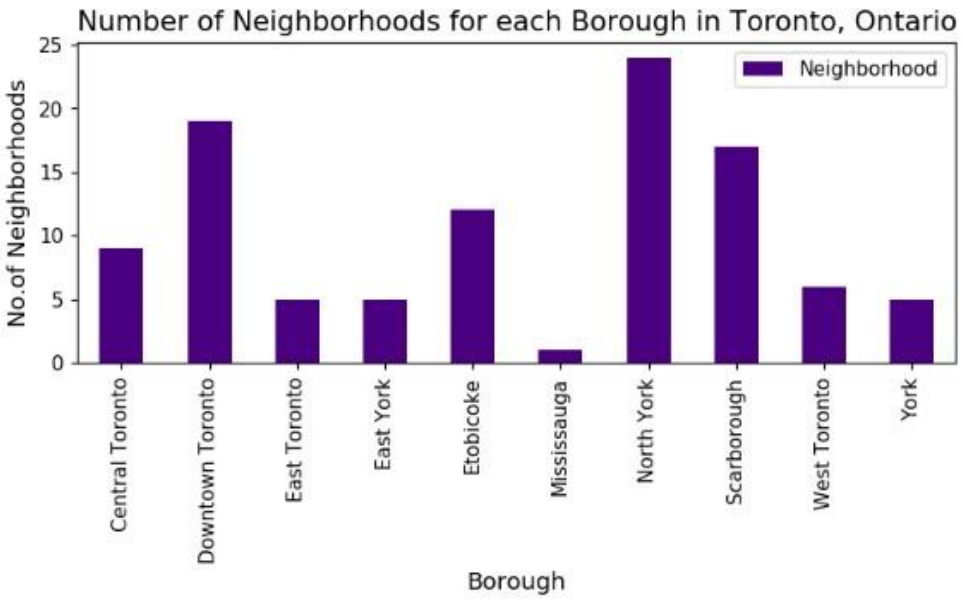

***Figure 7**: Bar chart showing number of neighborhoods for each Borough in Toronto*

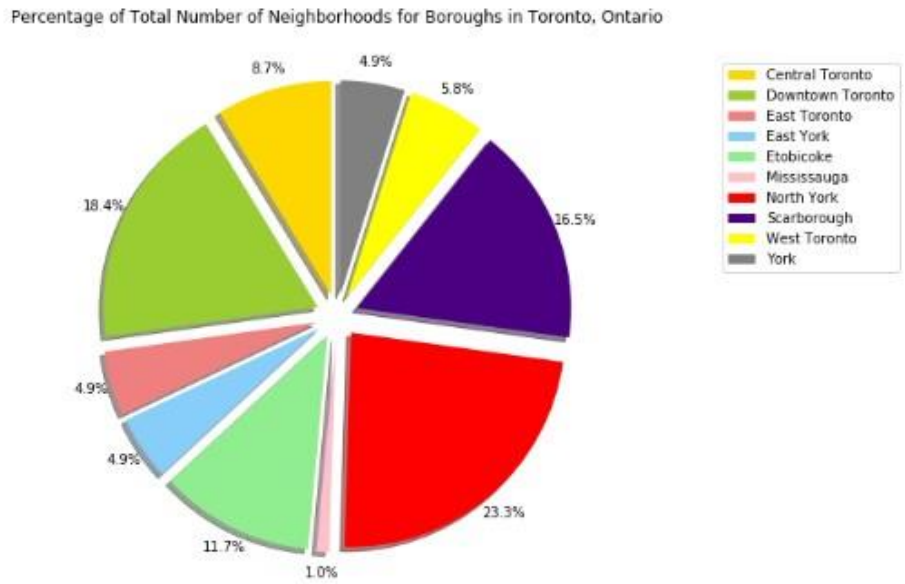

***Figure 8**: Pie chart showing the percentage of the total number of boroughs in Toronto*

3.2 Venues search using Foursquare API

Foursquare API (https://api.foursquare.com/v2/) was utilized extensively to perform venues search and to retrieve the name and id of the respective Japanese restaurants in each neighborhood based on its geographical coordinates in longitude and latitude. Foursquare is a popular social location service that requires login authentication.

The search radius is set to 1 km, a reasonable distance to have a fairly good area coverage with not too much of an overlap especially in neighborhoods such as Downtown Toronto. Figure 9 shows the results of venues search performed by utilizing Foursquare API, returning the name and id of the respective Japanese restaurants based on 1 km radius search around the neighborhoods' geographical coordinates.

	Borough	Neighborhood	ID	Name	Latitude	Longitude
0	Downtown Toronto	Queen's Park , Ontario Provincial Government	4ad4c061f964a520abf720e3	Tokyo Grill	43.662301	-79.389494
1	Downtown Toronto	Queen's Park , Ontario Provincial Government	4b2bd898f964a52042bc24e3	Kinka Izakaya Original	43.662301	-79.389494
2	Downtown Toronto	Queen's Park , Ontario Provincial Government	4c193c77838020a1e768e561	Kawa Sushi	43.662301	-79.389494
3	North York	Don Mills	53bafb4f498eb927faa3cd9e	Matsuda Japanese Cuisine & Teppanyaki	43.745906	-79.352188
4	North York	Don Mills	4b639286f964a5203c842ae3	Gonoe Sushi	43.745906	-79.352188
5	North York	Don Mills	4bd5bd1a29eb9c7434ce93e1	Teriyaki Experience	43.745906	-79.352188
6	Downtown Toronto	Garden District, Ryerson	4b2bd898f964a52042bc24e3	Kinka Izakaya Original	43.657162	-79.378937
7	Downtown Toronto	Garden District, Ryerson	574ad72238fa943556d93b8e	Gyu-Kaku Japanese BBQ	43.657162	-79.378937
8	Downtown Toronto	Garden District, Ryerson	5773f01f498e98371390bdfd	Rolltation	43.657162	-79.378937
9	Downtown Toronto	Garden District, Ryerson	50e8b903e4b04fa56858ecfe	Zakkushi ざっ串	43.657162	-79.378937
10	Downtown Toronto	Garden District, Ryerson	4c193c77838020a1e768e561	Kawa Sushi	43.657162	-79.378937

***Figure 9**: Results from venues search performed by utilizing Foursquare API returning name and id of the respective Japanese restaurants based on 1km radius search around the neighborhoods' geographical coordinates.*

From the venues search results, the total number of Japanese restaurants is 53, after filtering the duplicate results of the same restaurants due to overlapping of the search radius of 1 km with the adjacent neighborhoods during the venues search process. Figure 10 shows Boroughs and the corresponding number of Japanese Restaurants which are sorted in descending order in a table while Figure 11 shows the number of Japanese restaurants in each Borough represented by a bar chart.

	Borough	Japanese Restaurant Count
1	Downtown Toronto	19
5	North York	17
0	Central Toronto	5
2	East Toronto	5
7	West Toronto	3
3	East York	1
4	Mississauga	1
6	Scarborough	1
8	York	1

***Figure 10**: Table showing Boroughs and number of Japanese restaurants sorted in descending order*

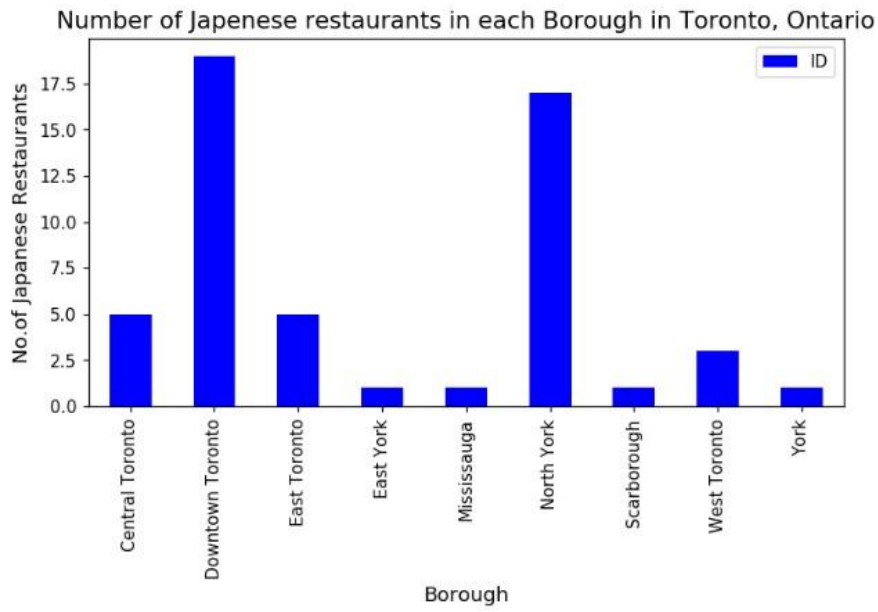

***Figure 11**: Number of Japanese Restaurants in each Borough from the venues search results*

From the venues search results, Downtown Toronto has the highest number of the Japanese restaurants with nineteen (19) restaurants, followed by North York with seventeen (17).

In terms of percentage, Downtown Toronto has the largest share of Japanese restaurants with 35.8%, followed by North York at 32.1%. Both boroughs account for almost two-thirds of the total number of Japanese restaurants in Toronto. Figure 12 shows the percentage of the total number of Japanese restaurants for each Borough represented by a pie chart.

It was also observed that Scarborough has only one (1) Japanese restaurant while Etobicoke has none (0), despite Scarborough having the third highest number of neighborhoods (17) behind North York (24) and Downtown Toronto (19) while Etobicoke ranks fourth with twelve (12) neighborhoods.

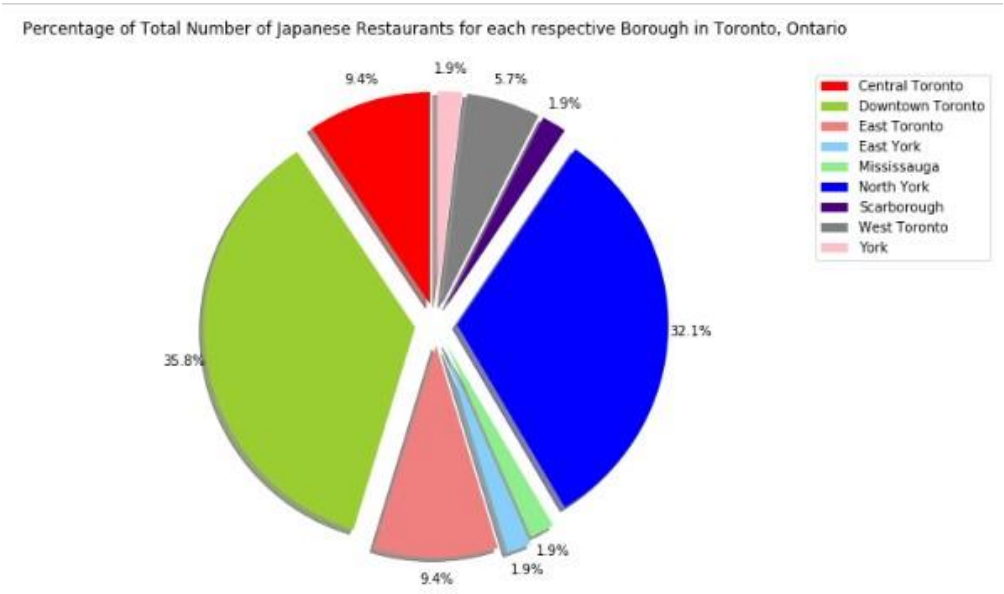

***Figure 12:** Pie chart showing the percentage of total number of Japanese restaurants for each Borough in Toronto*

3.3 Venue details search using Foursquare API

Next, Foursquare API premium search was utilized to perform venue details search to retrieve venue details information, such as ratings, number of likes, number of tips and price tiers based on the id of each Japanese restaurant, retrieved from the previous venues search. From the venue details search results, a total of 4 Japanese restaurants were found not having any data and were therefore dropped from the final list which has 50 Japanese restaurants. A preview of the results of the venue details search is shown in Figure 13. The final count of Japanese restaurants in each borough is shown in Figure 14.

	Borough	Neighborhood	ID	Name	Likes	Rating	Tips	PriceTier	Longitude	Latitude
0	Central Toronto	Davisville North	589f58fbd0bb3e25a8e5a88a	Rolltation	13	7.5	4	2	-79.390197	43.712751
1	Central Toronto	Forest Hill North & West	51ce2935498e66d5290f45c9	Sake Bar Kushi	29	7.7	14	2	-79.411307	43.696948
2	Central Toronto	Moore Park , Summerhill East	4b69b8ddf964a52000b02be3	Suki Japanese Cuisine	6	6.5	12	2	-79.383160	43.689574
3	Central Toronto	Roselawn	4b107e81f964a520b07123e3	EDO	11	7.1	9	3	-79.416936	43.711695
4	Central Toronto	The Annex , North Midtown , Yorkville	59f7bad335811b13a241e498	Gyubee	15	8.0	5	2	-79.405678	43.672710
5	Downtown Toronto	Berczy Park	56201ed4498e7f700c462170	Miku	142	9.0	47	2	-79.373306	43.644771
6	Downtown Toronto	Berczy Park	5a4fdf56772fbc5e9fa73c7f	Chotto Matte	11	8.0	2	2	-79.373306	43.644771

***Figure 13**: A preview of the results from the venue details search showing Likes, Rating, Tips and Price Tier for the corresponding Japanese restaurants*

	Borough	Japanese Restaurant Count
1	Downtown Toronto	19
4	North York	16
0	Central Toronto	5
2	East Toronto	5
5	West Toronto	3
3	Mississauga	1
6	York	1

***Figure 14**: Final number of Japanese restaurants for each borough*

From the final list of Japanese restaurants shown in Figure 14, Downtown Toronto has the highest number of Japanese restaurants with nineteen (19) followed by North York which has sixteen (16).

The retrieved venue details: ratings, number of likes, and number of tips for each restaurant were used as input to the K-means clustering algorithm to rank these Japanese restaurants. This workflow will be explained in detail in the next subsection.

3.4 K-means clustering algorithm

3.4.1 Introduction to K-means clustering

K-means clustering algorithm is one of the unsupervised machine learning techniques commonly used in data science projects. Unsupervised learning draws inferences from datasets consisting of input data without any labeled responses. The objective of K-means is to group similar data points together and to unravel underlying patterns. To achieve this objective, a fixed number of clusters, *k*, needs to be defined in a dataset. A cluster refers to a grouping of aggregated data points that have certain similarities. Each cluster has a centroid or center point.

The K-means algorithm starts with randomly selected centroids, which are used as the beginning points for every cluster, and then performs iterative calculations to optimize the positions of the centroids until those centroids have stabilized where there is no change in their values or until the defined number of iterations has been achieved.

3.4.2 Application of K-means clustering algorithm

The decision to use K-means clustering to rank the Japanese restaurants is based on several strengths of this algorithm as listed below:-

a. Relatively fast and easy to implement
b. Guarantees convergence
c. Able to group datapoints similar to each other
d. Unravels underlying trends or patterns
e. Easier outlier detection

For the selection of *k*, the number of clusters, there is no definitive hard rule to determine the optimum *k*. In this study, *k*=6 was chosen as a reasonable number of clusters to reduce the complexity during interpretation and ranking of the results.

To define the input dataset for K-means clustering, from the list of 50 Japanese restaurants, only 3 features were selected – number of likes, rating and number of tips. Price tier was dropped because it was not related to customers' sentiment. Figure 15 shows the preview of the number of likes, rating and number of tips for each Japanese restaurant used as input to K-means clustering.

	Likes	Rating	Tips
0	13	7.3	4
1	29	7.4	14
2	6	6.6	12
3	11	7.3	9
4	15	7.6	5
5	93	7.7	60
6	142	9.0	47
7	104	8.0	66
8	10	7.0	4
9	23	7.0	9

***Figure 15**: Preview of the number of likes, rating and number of tips for each Japanese restaurant used as input to K-means clustering*

4. RESULTS

4.1 ANALYSIS OF RESULTS

The results of the K-means clustering (k=6) which produced 6 different clusters were analyzed. Figure 16 shows color-coded clusters of the Japanese restaurants overlain on the map of Toronto.

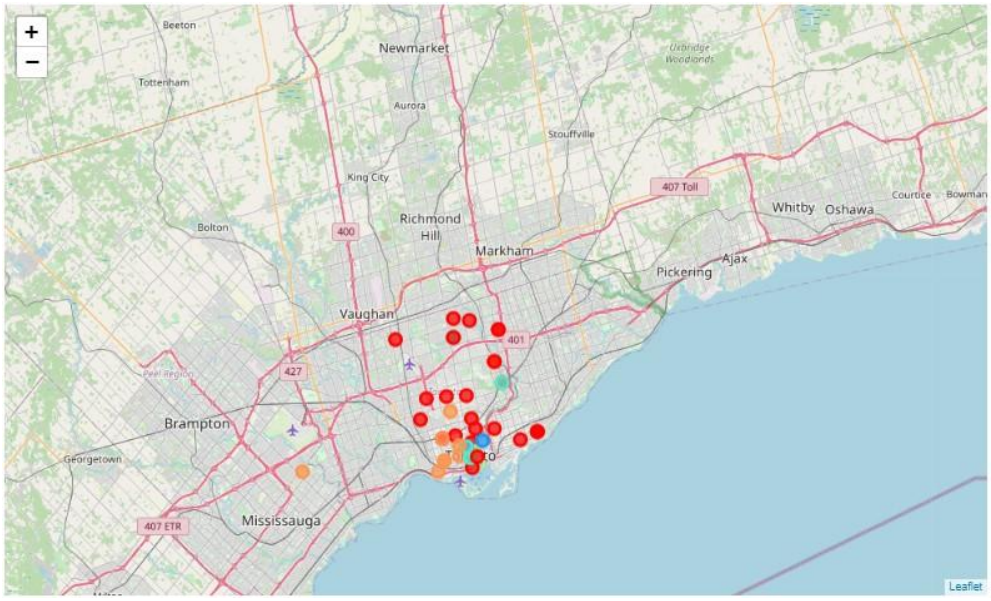

***Figure 16**: Map showing color-coded clusters of the Japanese restaurants as the result of K-means clustering*

Cluster 1 has 28 restaurants. A preview of Cluster 1 (only displaying the first 20 restaurants) is shown in Figure 17.

	Borough	Neighborhood	ID	Name	Likes	Rating	Tips	PriceTier	Longitude	Latitude
0	Central Toronto	Davisville North	589f58fbd0bb3e25a8e5a88a	Rolltation	13	7.3	4	2	-79.390197	43.712751
2	Central Toronto	Moore Park , Summerhill East	4b69b8ddf964a52000b02be3	Suki Japanese Cuisine	6	6.6	12	2	-79.383160	43.689574
3	Central Toronto	Roselawn	4b107e81f964a520b07123e3	EDO	11	7.3	9	3	-79.416936	43.711695
4	Central Toronto	The Annex , North Midtown , Yorkville	59f7bad335811b13a241e498	Gyubee	15	7.6	5	2	-79.405678	43.672710
8	Downtown Toronto	Christie	5884fa852f91cb3bc7e8ae95	Japanhako	10	7.0	4	2	-79.422564	43.669542
11	Downtown Toronto	Church and Wellesley	59c54d4f2d2fd97564d4cfc8	Onnki Donburi	13	7.6	3	2	-79.383160	43.665860
13	Downtown Toronto	Garden District, Ryerson	599cb69eb5461876d54ab713	Katsuya	14	7.4	3	2	-79.378937	43.657162
15	Downtown Toronto	Harbourfront East , Union Station , Toronto Is...	5a2c772450a6f0229d83d923	Akira Back	9	8.4	6	4	-79.381752	43.640816
20	Downtown Toronto	Rosedale	4b50f565f964a5209b3a27e3	Nijo Japanese Restaurant	13	6.6	11	2	-79.377529	43.679563
21	Downtown Toronto	St. James Town	5a4fdf56772fbc5e9fa73c7f	Chotto Matte	11	8.0	2	2	-79.375418	43.651494
24	East Toronto	India Bazaar , The Beaches West	4ad4c05df964a52048f620e3	Aki Da Japanese Seafood House	5	6.4	12	2	-79.315572	43.668999
25	East Toronto	The Beaches	4c10317981e976b075d10eeb	Bikkuri Japanese Restaurant	4	6.4	9	2	-79.293031	43.676357
26	East Toronto	The Beaches	587bae8728680458 4acd2a7a	Isabella's Boutique Restaurant	7	8.3	3	2	-79.293031	43.676357
27	East Toronto	The Beaches	4aff4c2df964a520103722e3	Yumei Sushi	13	7.2	12	3	-79.293031	43.676357
28	East Toronto	The Danforth West , Riverdale	4cec56edb80da09371e51b00	Sushi Friends	16	6.9	9	2	-79.352188	43.679557
30	North York	Bayview Village	5529a011498eddb919b0f2f5	Kori Sushi	9	7.2	1	2	-79.385975	43.786947
31	North York	Don Mills	4b639286f964a5203c842ae3	Gonoe Sushi	11	6.8	14	4	-79.352188	43.745906
32	North York	Don Mills	4b36ac26f964a520623a25e3	Kaiseki Yu-Zen Hashimoto	5	8.1	6	2	-79.340923	43.725900
35	North York	Don Mills	4bd5bd1a29eb9c7434ce93e1	Teriyaki Experience	3	6.0	3	2	-79.352188	43.745906
36	North York	Fairview , Henry Farm , Oriole	5952e5ed491be7022cf5d19f	Heart Sushi	5	6.3	3	2	-79.346556	43.778517
37	North York	Fairview , Henry Farm , Oriole	4c296a65e19720a11990f958	Tekka Sushi	4	7.0	6	2	-79.346556	43.778517
38	North York	Glencairn	4fca732be4b07c8e3e198c27	Miyako Sushi Restaurant	8	6.9	3	2	-79.445073	43.709577

***Figure 17:** A Preview of Cluster 1 with 28 Japanese restaurants*

Cluster 2 has only 1 restaurant as shown in Figure 18.

	Borough	Neighborhood	ID	Name	Likes	Rating	Tips	PriceTier	Longitude	Latitude
17	Downtown Toronto	Queen's Park , Ontario Provincial Government	4b2bd898f964a52042bc24e3	Kinka Izakaya Original	387	8.3	223	2	-79.389494	43.662301

***Figure 18**: Cluster 2 with 1 Japanese restaurant*

Cluster 3 has 3 Japanese restaurants as shown in Figure 19.

	Borough	Neighborhood	ID	Name	Likes	Rating	Tips	PriceTier	Longitude	Latitude
5	Downtown Toronto	Berczy Park	4ae73054f964a5203ca921e3	Ki Modern Japanese + Bar	93	7.7	60	3	-79.373306	43.644771
7	Downtown Toronto	Central Bay Street	4ad9f607f964a520691c21e3	Manpuku まんぷく	104	8.0	66	2	-79.387383	43.657952
22	Downtown Toronto	St. James Town , Cabbagetown	50cc21c37ab4d9c0594b6994	Kingyo Toronto	113	8.7	54	2	-79.367675	43.667967

***Figure 19**: Cluster 3 with 3 Japanese restaurants*

Cluster 4 has 5 Japanese restaurants as shown in Figure 20.

	Borough	Neighborhood	ID	Name	Likes	Rating	Tips	PriceTier	Longitude	Latitude
18	Downtown Toronto	Queen's Park , Ontario Provincial Government	4ad4c061f964a520abf720e3	Tokyo Grill	35	7.7	29	1	-79.389494	43.662301
19	Downtown Toronto	Richmond , Adelaide , King	4afa1f33f964a520e91622e3	Fune Japanese Restaurant	53	8.0	27	2	-79.384568	43.650571
34	North York	Don Mills	4aec790bf964a5208fc721e3	Memories of Japan	35	6.7	32	2	-79.340923	43.725900
42	North York	Willowdale	4fc6d475e4b0785353957fe7	Nomé Izakaya	65	6.7	31	2	-79.408493	43.770120
43	North York	Willowdale	4c76920adb52b1f7a0c97bdc	Sushi One	43	7.3	35	3	-79.408493	43.770120

***Figure 20**: Cluster 4 with 5 Japanese restaurants*

Cluster 5 has 2 Japanese restaurants as shown in Figure 21.

	Borough	Neighborhood	ID	Name	Likes	Rating	Tips	PriceTier	Longitude	Latitude
6	Downtown Toronto	Berczy Park	56201ed4498e7f700c462170	Miku	142	9.0	47	2	-79.373306	43.644771
14	Downtown Toronto	Garden District, Ryerson	50e8b903e4b04fa56858ecfe	Zakkushi ざっ串	125	8.5	34	2	-79.378937	43.657162

***Figure 21**: Cluster 5 with 2 Japanese restaurants*

The last cluster, Cluster 6 has 11 restaurants as shown in Figure 22.

	Borough	Neighborhood	ID	Name	Likes	Rating	Tips	PriceTier	Longitude	Latitude
1	Central Toronto	Forest Hill North & West	51ce2935498e66d5290f45c9	Sake Bar Kushi	29	7.4	14	2	-79.411307	43.696948
9	Downtown Toronto	Christie	5317b12e11d2ed4e5d1b13ef	KINTORI YAKITORI	23	7.0	9	2	-79.422564	43.669542
10	Downtown Toronto	Church and Wellesley	4c193c77838020a1e768e561	Kawa Sushi	27	7.8	13	2	-79.383160	43.665860
12	Downtown Toronto	Garden District, Ryerson	574ad72238fa943556d93b8e	Gyu-Kaku Japanese BBQ	42	8.6	11	2	-79.378937	43.657162
16	Downtown Toronto	Kensington Market , Chinatown , Grange Park	503fb59ce4b050691ba405a3	Gushi	21	7.5	9	2	-79.400049	43.653206
23	Downtown Toronto	University of Toronto , Harbord	5362c366498e602fbe1db395	Yasu	46	8.7	12	2	-79.400049	43.662696
29	Mississauga	Canada Post Gateway Processing Centre	4bc895eeaf07a5935528812d	Masamune	22	8.1	15	2	-79.615819	43.636966
33	North York	Don Mills	53bafb4f498eb927faa3cd9e	Matsuda Japanese Cuisine & Teppanyaki	37	7.2	10	2	-79.352188	43.745906
46	West Toronto	Brockton , Parkdale Village , Exhibition Place	5792a2e4cd10f4e248ad94e2	Guu Izakaya	29	8.6	2	2	-79.428191	43.636847
47	West Toronto	Little Portugal , Trinity	51049684e4b06b57cac4e3e8	Bazara	28	8.2	15	2	-79.419750	43.647927
48	West Toronto	Little Portugal , Trinity	56418c69498ec00425302f6a	Imanishi Japanese Kitchen	48	9.0	9	2	-79.419750	43.647927

***Figure 22**: Cluster 6 with 11 Japanese restaurants*

From the results, each one of the six (6) clusters was analyzed with observations described as below:-

a. Cluster 1: This cluster by far the largest amongst the 6 clusters, has 28 restaurants, with locations in Central Toronto, Downtown Toronto, East Toronto, North York, and York. This cluster represents more than half of the total number (50) of restaurants on the final list. The maximum number of likes is 16 and minimum is 3, while the maximum number of tips is 14 and the minimum is 1. The rating ranges from 6.0 to 8.8.
b. Cluster 2: This cluster only has 1 restaurant, located in Downtown Toronto. This restaurant has apparently the highest number of likes that is 387 and the highest number of tips which stands at 223. The rating for this restaurant is 8.3.
c. Cluster 3: This cluster has 3 restaurants, all located in Downtown Toronto with the number of likes ranges from 93 to 113, while number of tips is from 54 to 60. The rating ranges from 7.7 to 8.7.
d. Cluster 4: This cluster has 5 restaurants, located in Downtown Toronto and North York, with the number of likes ranges from 45 to 65, while the number of tips is from 27 to 35. The rating ranges from 6.7 to 8.0
e. Cluster 5: This cluster has only 2 restaurants, both are in Downtown Toronto with the numbers of likes are 142 and 125 respectively, while the numbers of tips are 47 and 34 respectively. The ratings are 9.0 and 8.5 respectively.
f. Cluster 6: The last cluster has 11 restaurants, located in various locations - Central Toronto, Downtown Toronto, West Toronto, North York, and Mississauga with numbers of likes from 21 to 48, while the number of tips ranges from 9 to 15. Ratings are from 7.0 to 9.0.

4.2 INTERPRETATION OF RESULTS & RANKING

From the analysis on the 6 different clusters, the results of the K-means clustering were interpreted and ranked as the following:-

a. The only restaurant in Cluster 2, has apparently the highest number of likes (387) and the highest number of tips (223). The rating of this restaurant is 8.3. Undoubtedly, this gives indication that this restaurant is the most popular in Toronto and should be ranked the top of the list. This restaurant is in Downtown Toronto.
b. The cluster in which the restaurants have the second and third number of likes (142 & 125 respectively) is Cluster 5. Though the number of tips is 47 and 34 respectively, however the rating scores are high: 9.0 and 8.5 respectively. This cluster of 2 restaurants was ranked second on the list, below the Cluster 2 restaurant. Both restaurants are in Downtown Toronto.
c. The cluster that was ranked third on the list, behind Cluster 5 is Cluster 3. It has 3 restaurants, all located in Downtown Toronto with the number of likes ranges from 93 to 113, which are lower than those of Cluster 5 while number of tips ranges from 54 to 60, which are slightly higher than those of Cluster 5. The rating ranges from 7.7 to 8.7, which are lower than Cluster 5 in general.
d. The cluster that was ranked fourth on the list is Cluster 4 which has 5 restaurants, with the number of likes ranges from 45 to 65, while the number of tips is from 27 to 35. The rating ranges from 6.7 to 8.0. The restaurants are in Downtown Toronto and North York.
e. Next on the list, Cluster 6 was ranked fifth, which has 11 restaurants found in various locations - Central Toronto, Downtown Toronto, West Toronto, North York, and Mississauga with numbers of likes from 21 to 48, while the number of tips ranges from 9 to 15. Ratings are from 7.0 to 9.0.

f. Bottom on the list is Cluster 1 that was ranked sixth. This cluster has 28 restaurants in various locations including Central Toronto, Downtown Toronto, East Toronto, North York, and York. The maximum number of likes is 16 and the minimum is 3, while the maximum number of tips is 14 and the minimum is 1. The rating ranges from 6.0 to 8.8. Cluster 1 has the lowest number of likes and the lowest number of tips amongst the six clusters.

The result of the ranking exercise is summarized in Figure 23.

Ranking	Cluster Number	Restaurant Count	Number of Likes (Range)	Number of Tips (Range)	Rating (Range)	Boroughs
1	2	1	387	233	8.3	DT
2	5	2	125-142	34-47	8.5-9.0	DT
3	3	3	93-113	54-60	7.7-8.7	DT
4	4	5	45-65	27-35	6.7-8.0	DT, NY
5	6	11	21-48	9-15	7.0-9.0	CT, DT, WT, NY, M
6	1	28	3-16	1-14	6.0-8.8	CT, DT, ET, NY, Y

Abbreviation Key for Borough Names

CT	DT	ET	WT	NY	Y	M
Central Toronto	Downtown Toronto	East Toronto	West Toronto	North York	York	Mississauga

***Figure 23**: Table of the ranking result for the 6 different clusters*

4.3 SELECTING THE TOP 3 RESTAURANTS

The restaurants from Cluster 2 and Cluster 5, which are the top two clusters from the ranking list were selected to be the top 3 Japanese restaurants in Toronto. The names of these restaurants are – Kinka Izakaya Original from Cluster 2, Miku and Zakkushi are from Cluster 5. These restaurants are all located in Downtown Toronto. Further venue details search was performed using Foursquare API premium search to retrieve details such as Address, Phone Number, Opening Hours and actual location coordinates in Latitude and Longitude for the purpose of map visualization.

The retrieved venue details for the top 3 Japanese restaurants are shown below in Figure 24 :-

	Borough	Neighborhood	ID	Name	Address	PhoneNumber	Hours	Latitude	Longitude
0	Downtown Toronto	Queen's Park , Ontario Provincial Government	4b2bd898f964a52042bc24e3	Kinka Izakaya Original	398 Church St	4169770999	Closed until 11:30 AM	43.660596	-79.378891
1	Downtown Toronto	Berczy Park	56201ed4498e7f700c462170	Miku	10 Bay St	6473477347	Closed until 11:30 AM	43.641374	-79.377531
2	Downtown Toronto	Garden District, Ryerson	50e8b903e4b04fa56858ecfe	Zakkushi ざっ串	193 Carlton St.	6473529455	Closed until 5:30 PM	43.663618	-79.370670

***Figure 24**: Venue details for the top 3 Japanese restaurants which include Address, Phone Number, Opening Hours and Actual Latitude and Longitude.*

The locations of these top 3 Japanese restaurants are plotted on the map for visualization purpose as shown in Figure 25.

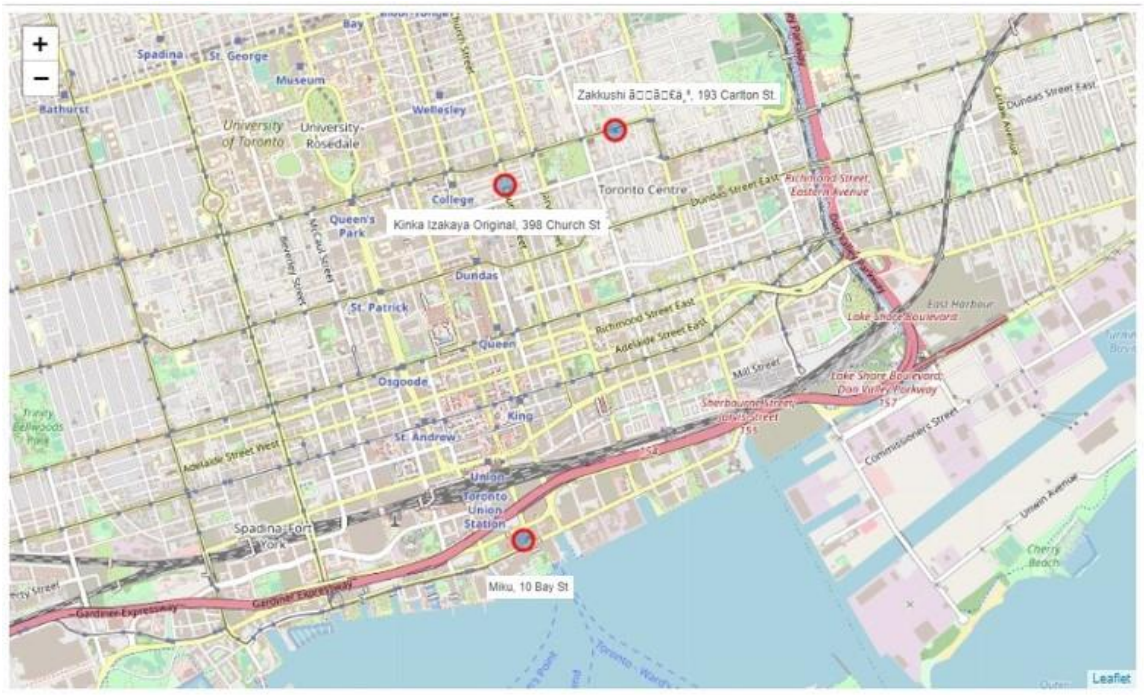

***Figure 25**: The locations of the top 3 Japanese restaurants are shown on the map*

5. DISCUSSION

5.1 Observations

From the analysis, interpretation and ranking of the K-means clustering results, several observations were made and noted. These observations are discussed as below:-

a. K-means clustering algorithm is an effective technique that is able to distinguish the top Japanese restaurants from the rest of the group based on the quantitative customers' sentiments such as number of likes, number of tips and ranking as input data.
b. Cluster 1 has the lowest number of likes and the lowest number of tips amongst the six clusters, but it is the largest cluster which accounts more than half of the total number of Japanese restaurants on the final list. This is indicative that these restaurants are the least popular amongst customers. Customers in Toronto seem to be quite selective when choosing their preferred choice of Japanese restaurants.
c. All the top 3 Japanese restaurants have Price Tier of 2 which is on the less expensive side of the scale. In Toronto, less expensive Japanese restaurants seem to be preferred by customers, in general.
d. The Foursquare API premium search does not have complete venue details data in its database. When performing venue details search for this study, 3 out of the initial list of 53 Japanese restaurants were found not to have the required venue details data and as a result, those 3 restaurants had to be dropped from the list of Japanese restaurants to be used as input for K-means clustering. With a small dataset size of 50 restaurants, even 3 restaurants could substantially affect the outcome of the results. Ideally, after data cleaning process, all the 53 restaurants should have been maintained to be used as input for K-means clustering.

5.2 Recommendations

From this study, several recommendations for further improvements are discussed as below:-

a. The workflow implemented in this study which utilized K-means clustering algorithm can be replicated to be used in other similar data science projects, for example – determining the top Italian, Korean, Indian, Indonesian restaurants et cetera in certain cities of interest.
b. An alternative to Foursquare API such as Google Map API should be explored although its pricing could be a drawback. This is because Foursquare may not have the complete information or data on the venue details which may affect the results of the study if the dataset size is small.
c. The lack of free geospatial data in terms of Latitude and Longitude in relation to Canadian postal codes may be a challenge for the projects which crucially need those data to perform venues search or any geospatial analysis work related to Canada. In view of this, Canadian government agencies or universities should make available this type of data in the public domain for free.

6. CONCLUSION

The results of this study which utilized the K-means clustering algorithm, have successfully achieved the objective to provide reliable and well-researched information to the target group comprising business travellers, tourists and short-term visitors to Toronto in addressing their requirements as below:-

a. List of Japanese restaurants in Toronto: A final list of 50 Japanese restaurants which contains venue details such as number of likes, number of tips, ratings, and price tiers.
b. Boroughs in Toronto that have a significant presence of Japanese restaurants and the borough that has the highest number of Japanese restaurants: Downtown Toronto has the highest number of Japanese restaurants (19) followed by North York (16). Travelers can choose their hotel accommodation in these 2 boroughs for the proximity to a wide selection of Japanese restaurants.
c. Ranking of the Japanese restaurants in Toronto based on customers' sentiments: Ranking exercise was done to rank 6 different clusters from the results of K-means clustering. Cluster 2 with 1 restaurant was ranked top, cluster 5 with 2 restaurants was ranked second, followed by Cluster 3 with 3 restaurants. Cluster 4 with 5 restaurants was ranked fourth, followed by Cluster 6 with 11 restaurants. Cluster 1 with 28 restaurants was ranked bottom of the list.
d. Recommendation on the top 3 Japanese restaurants in Toronto: The 3 top Japanese restaurants identified are from Cluster 2 and Cluster 5. These restaurants are - Kinka Izakaya Original, Miku, and Zakkushi, which are all in Downtown Toronto.