Japanese Restaurant

CAPSTONE PROJECT - BATTLE OF NEIGHBORHOODS

Determining the Top Japanese Restaurants in Toronto using K-means algorithm

BY : **HOO CHEE HAU**

Photo credit: dreamstime.com

Introduction

- **Toronto** is the largest city in Canada
- Located on northwestern shore of Lake Ontario
- An important international business and financial center in Canada
- One of the most cosmopolitan and vibrant cities in North America
- Japanese cuisine has blossomed in cosmopolitan Toronto since 2012

Japanese Restaurant

Business Problem

- **Business travellers**, **visitors** & **tourists** to Toronto who need well-researched information on:-

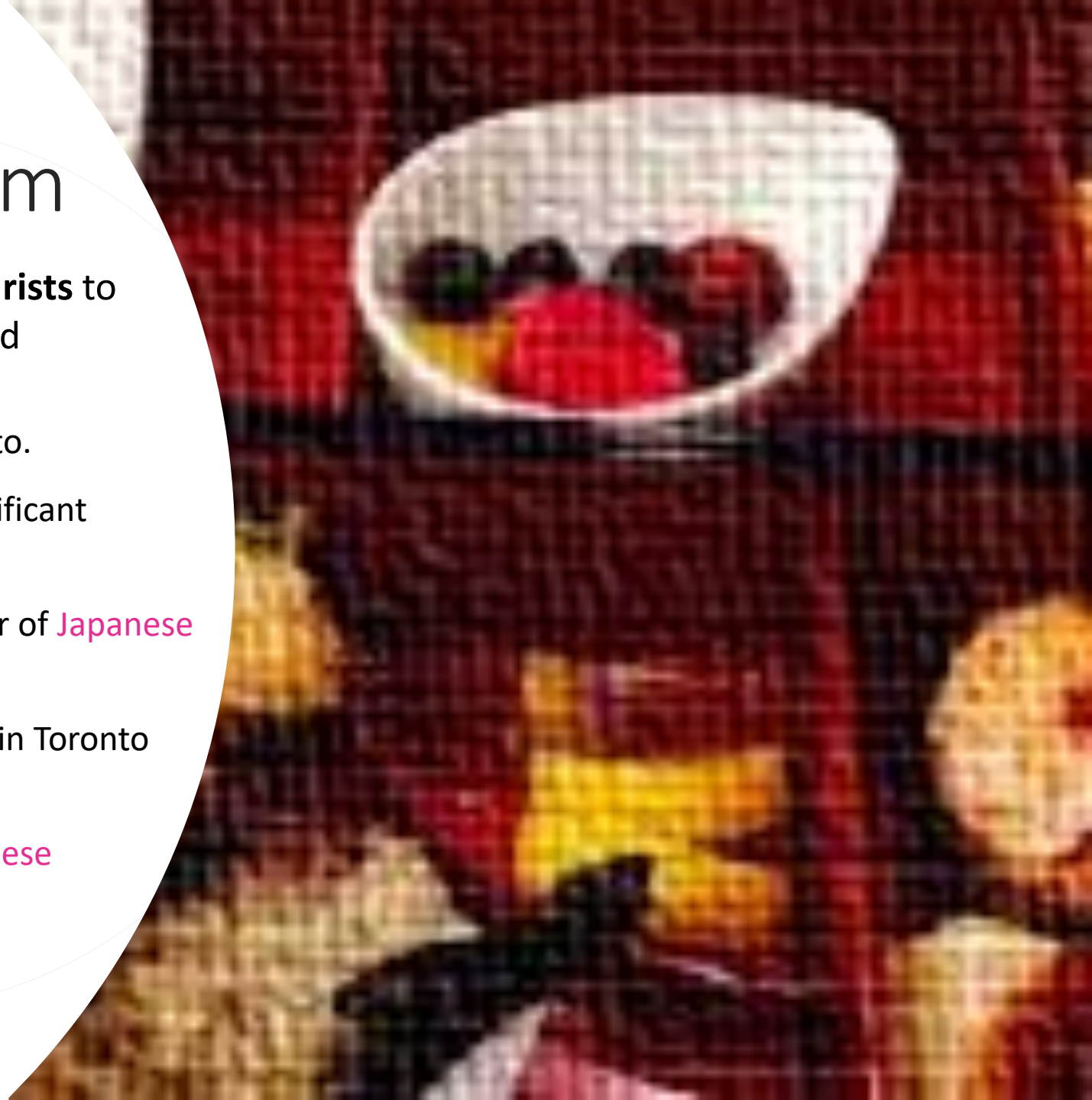
a. List of Japanese restaurants in Toronto.

b. Boroughs in Toronto that have a significant presence of Japanese restaurants

c. Borough that has the highest number of Japanese restaurants.

d. Ranking of the Japanese restaurants in Toronto based on customers' sentiments.

e. Recommendation on the top 3 Japanese restaurants in Toronto.



Data Acquisition & Cleaning

Japanese Restaurant

- Toronto's Postal Code **FSA** (Forward Sortation Area) with **Borough** & **Neighborhood** details were webscraped from Wikipedia website:- https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
- Data with 'NaN' records were removed and '/' delimiters were replaced with ','
- After data cleaning, total **Borough**s is 10 and total **Neighborhood**s is 103

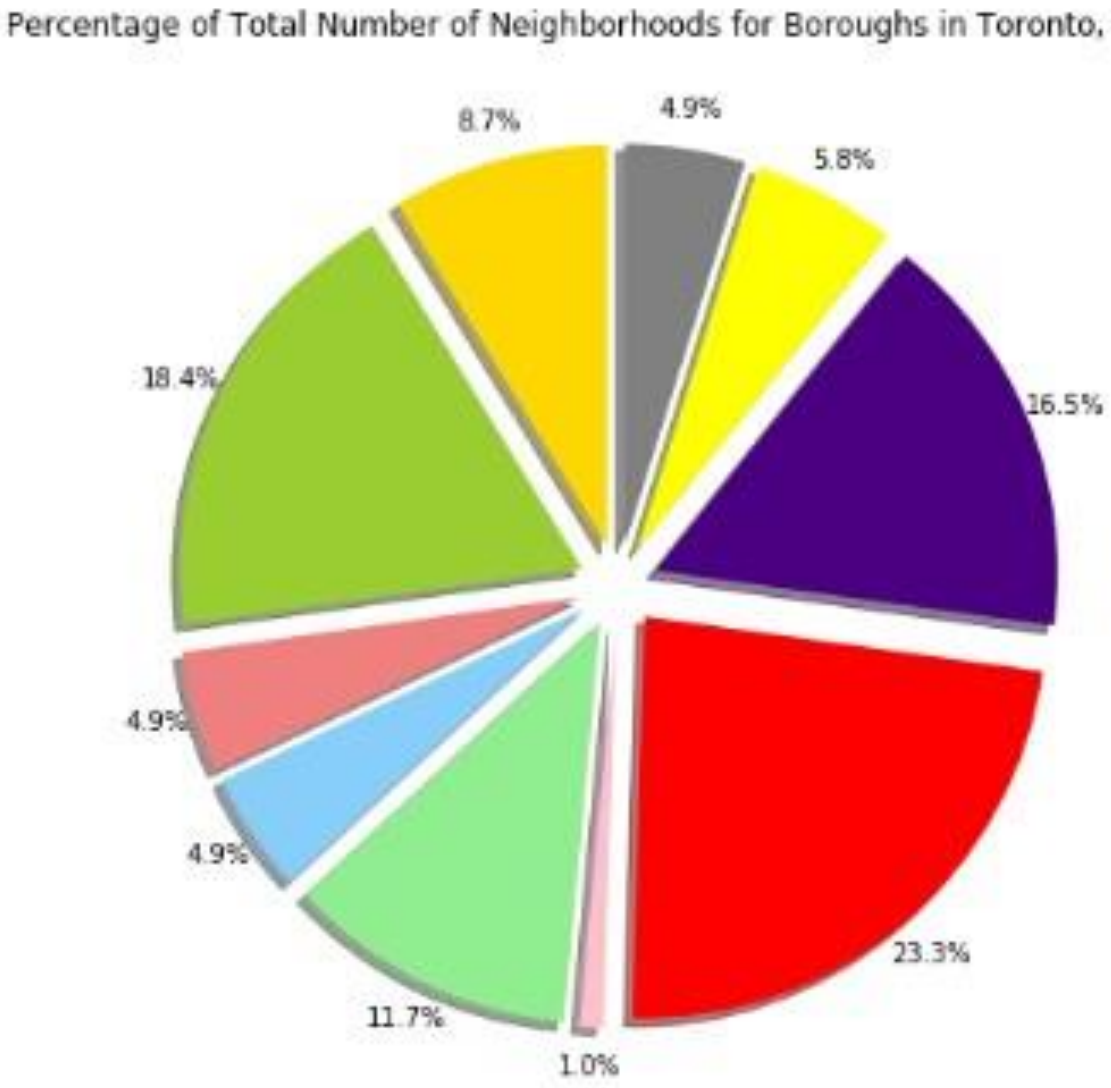
- Toronto's Postal Code **FSA** with geographically coordinates in **Latitude** & **Longitude** were downloaded from: http://cocl.us/Geospatial_data

- Cleaned data containing 10 Boroughs & 103 Neighborhoods were merged with Toronto's **FSA** geospatial data to obtain the Neighborhoods' geographical coordinates

- **Foursquare API** (https://api.foursquare.com/v2/) was utilized to perform **get venues** search on Japanese restaurants based on the Neighborhoods' geographical coordinates.
- Total Japanese restaurants found after removing duplicates was 53

Methodology: Analysis on Toronto Boroughs & Neighborhoods

Japanese Restaurant

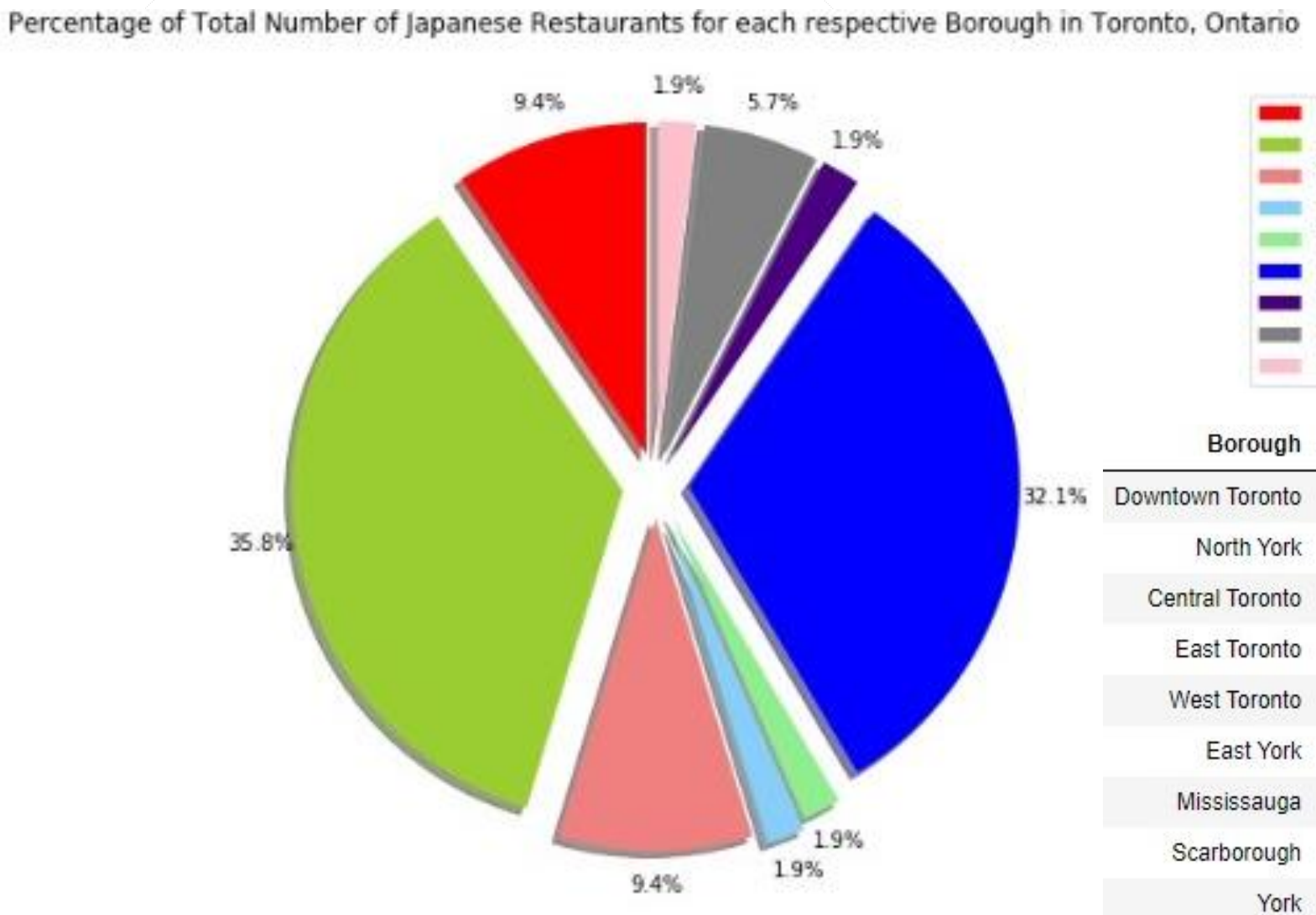


Borough	Neighborhood Count
North York	24
Downtown Toronto	19
Scarborough	17
Etobicoke	12
Central Toronto	9
West Toronto	6
East Toronto	5
East York	5
York	5
Mississauga	1

North York has the highest number of neighborhoods, followed by **Downtown Toronto** and **Scarborough**

Methodology: Analysis on Japanese Restaurants from Foursquare API get venues search

Japanese Restaurant



Downtown Toronto has the highest number of Japanese restaurants followed by **North York**

Borough	Japanese Restaurant Count
Downtown Toronto	19
North York	17
Central Toronto	5
East Toronto	5
West Toronto	3
East York	1
Mississauga	1
Scarborough	1
York	1

Methodology: Foursquare API venue details search results

- Venue details search based on venue id to retrieve **number of likes**, **rating**, **number of tips** and **price tier** of each Japanese restaurant
- Final count of the Japanese restaurants which have venue details data is 50

	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

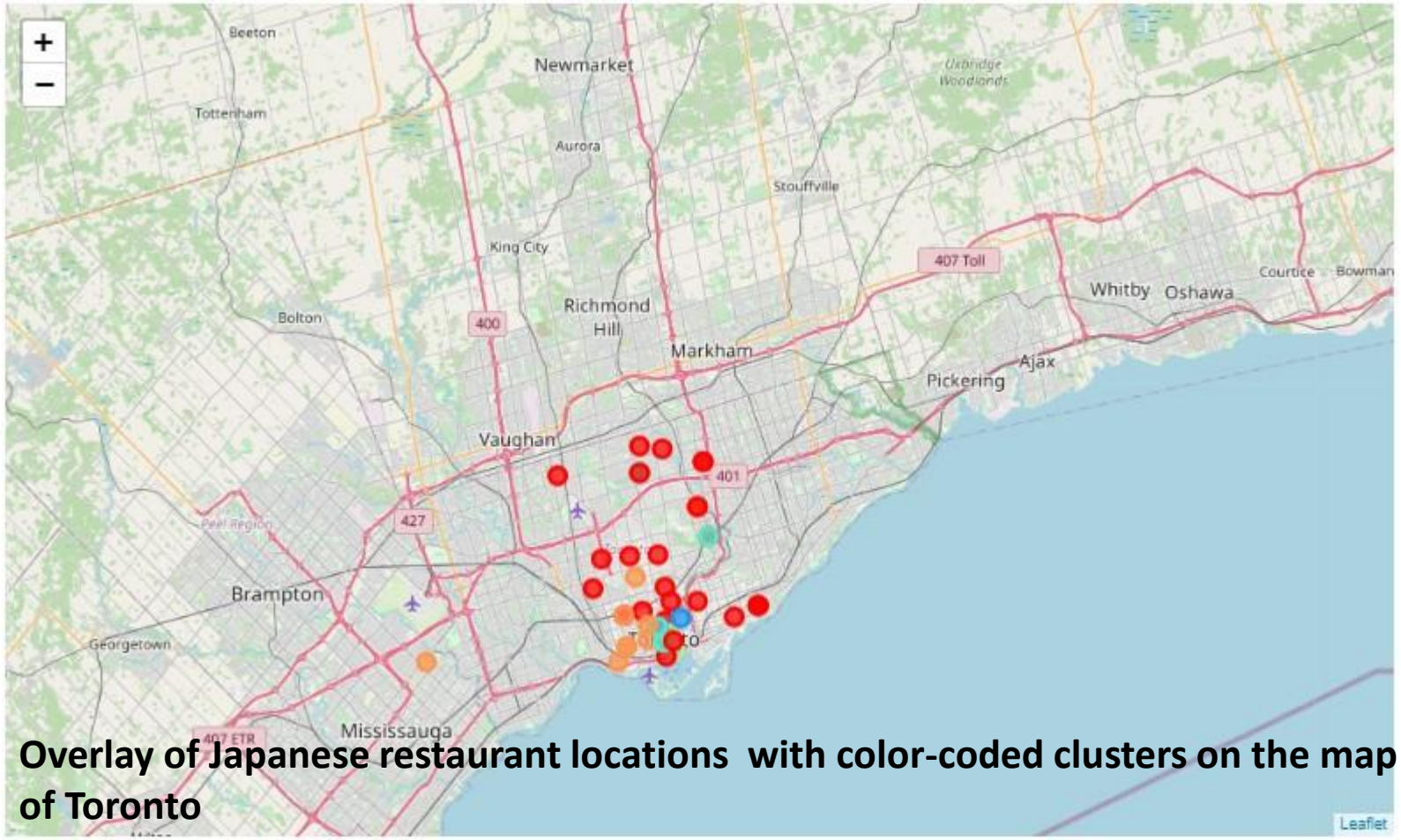
Methodology: K-means clustering algorithm

	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

- K-means clustering algorithm is **unsupervised machine learning** technique
- The **objective** of K-means clustering is to **group** similar data points together and to unravel **underlying patterns**
- Input data to K-means clustering: **number of likes**, **rating** and **number of tips**
- **Number of Clusters (k) =6**

Results: Analysis on the clusters

- 6 different clusters as the result of K-means clustering were analyzed and ranked



Overlay of Japanese restaurant locations with color-coded clusters on the map of Toronto

Cluster Number	Number of Restaurants
1	28
2	1
3	3
4	5
5	2
6	11

Results: Ranking of the clusters

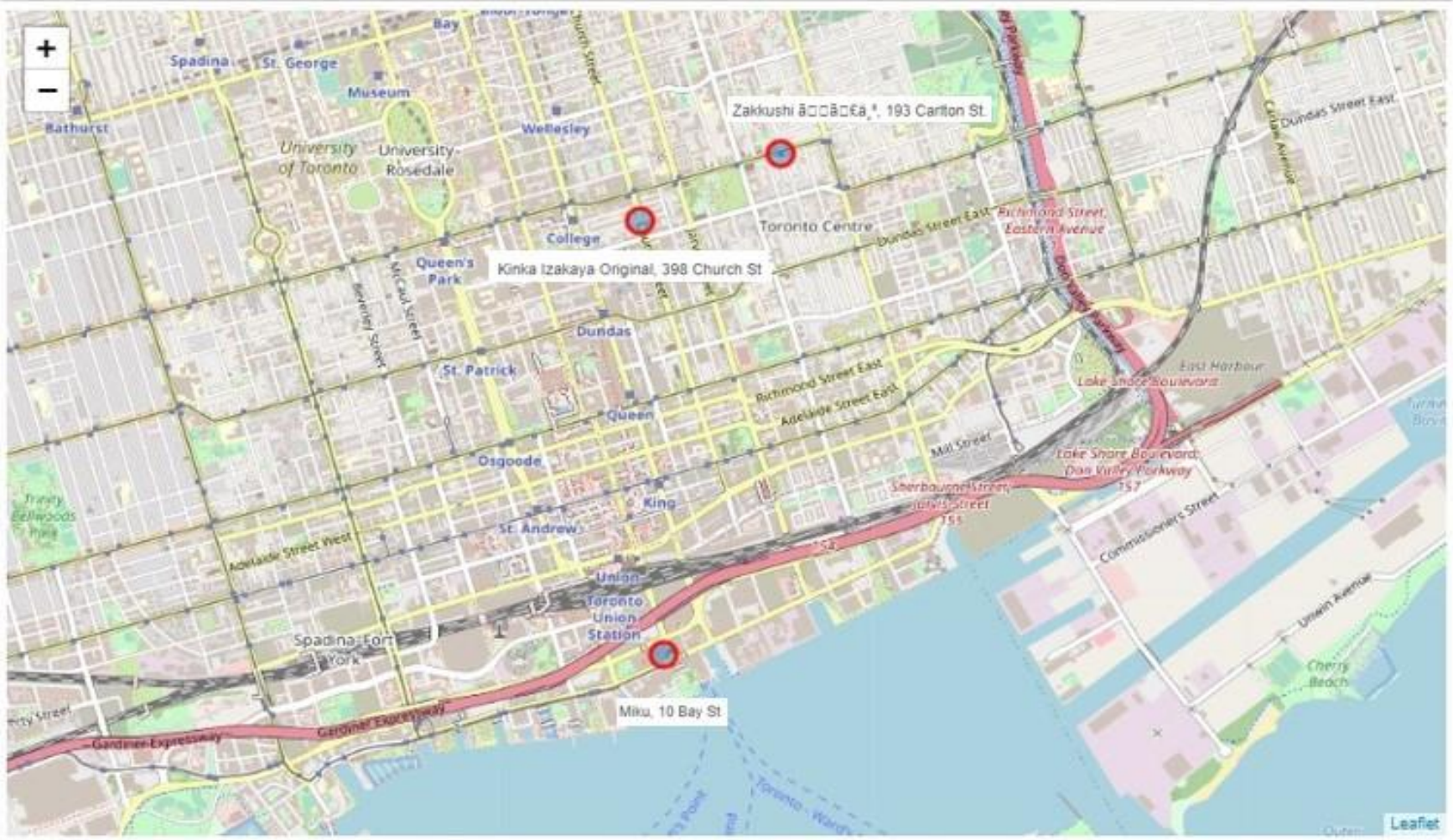
Ranking	Cluster Number	Restaurant Count	Number of Likes (Range)	Number of Tips (Range)	Rating (Range)	Boroughs
1	2	1	387	233	8.3	DT
2	5	2	125-142	34-47	8.5-9.0	DT
3	3	3	93-113	54-60	7.7-8.7	DT
4	4	5	45-65	27-35	6.7-8.0	DT, NY
5	6	11	21-48	9-15	7.0-9.0	CT, DT, WT, NY, M
6	1	28	3-16	1-14	6.0-8.8	CT, DT,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

Results: The 3 Top Japanese Restaurants

- The 3 top Japanese Restaurants in Toronto are **: Kinka Izakaya Original** , **Miku** & **Zakkushi**



Discussion: Observations

- **K-means clustering algorithm** has successfully distinguished 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

- **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.

- All the **top 3 Japanese restaurants** have Price Tier of 2 which is on the **less expensive** side of the scale which indicate less pricey Japanese restaurants are preferred by customers.
- The **Foursquare API** premium search does **not** have **complete venue details** data in its database. This could substantially affect the outcome of the study if the dataset size is small.

Discussion: Recommendations

- The **workflow** implemented in this study which utilized **K-means clustering** algorithm can be **replicated** to be used in other similar data science projects.

- An **alternative** to Foursquare API such as **Google Map API** should be explored although its pricing could be a drawback.

- The lack of **free geospatial data** in relation to Canadian **postal codes** may be a challenge which some projects crucially need for geospatial analyses. Canadian government agencies or universities should make available this type of data in the **public domain** for **free**.

Conclusion

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

- **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.

- **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 (19), followed by **North York** (16).

- **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**.

- **Recommendation** on the **top 3 Japanese restaurants** in Toronto: **Kinka Izakaya Original**, **Miku**, and **Zakkush**i, which are all in **Downtown Toronto**.

THANK YOU

Coursera Applied Data Science Capstone Project