Opening a Japanese Restaurant in Toronto

Background

- Toronto is the capital city of the province of Ontario, one of the largest cities in Canada by population with 2,731,571 residents as of
- The majority of the Ontario residents live the the Greater Toronto Area which makes it Canada's most populous city
- With the wide diversity in Canada and in the City of Toronto, there are many restaurants that offer almost every cuisine that exists on the planet
- However, considering the increasing popularity of the oriental food, it is always a best idea to open an authentic Japanese restaurant

Problem

- Opening a new restaurant in the city would require a good understanding of the geography, neighborhoods and diversity of the city and the distribution of the restaurants on each side of the city that ranges from east to west and the center

Interest

- Our study would require filtering all the venues to only work on venues that serve food, fast food or are just diner places or restaurants
- Also the information will be filtered in order to see how many Japanese restaurants are in each area
- Our study should decide which area would be best to open an Japanese restaurant

Data sources

- For the Toronto neighborhood data, I will use the following Wikipedia page: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
- I will scrape the Wikipedia page by using BeautifulSoup and wrangle the data, clean it, and then read it into a pandas dataframe
- The geographical coordinates of each postal code will be read in a pandas dataframe as well from: https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkillsNetwork-DS0701EN-SkillsNetwork/labs_v1/Geospatial_Coordinates.csv

Data cleaning

- The both dataframes obtained will be merged together so that each neighborhood has its geographical coordinates assigned
- I will use Foursquare APIs to fetch the data from all the venues in Toronto
- Then I will filter the data to get only the restaurants/diners or food related venues to work on
- The venue data obtained will help to find out which area is the best one to open a Japanese restaurant

Methodology

- The data was scraped using BeautifulSoup and put into a dataframe as shown in Fig.1.
- The second source of data provided us with the Geographical coordinates of the neighborhoods with the respective Postal Codes (Fig.2)
- Both dataframes obtained were merged into one, so that each neighborhood has its geographical coordinates assigned (Fig. 3)
- The retrieval of the location, name, category and food type about the restaurants in Toronto was collected through the Foursquare explore API
- To obtain the data, it was required to make an account where it would provide a ‘Secret Key’ as well as a ‘Client ID’ which would allow me to pull any data

[6]: `df_Toronto.head(10)`

[6]:

	PostalCode	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	Queen's Park	Ontario Provincial Government
5	M9A	Etobicoke	Islington Avenue
6	M1B	Scarborough	Malvern, Rouge
7	M3B	North York	Don Mills North
8	M4B	East York	Parkview Hill, Woodbine Gardens
9	M5B	Downtown Toronto	Garden District, Ryerson

Fig. 1

`df_geo_coor.head()`

[7]:

	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

[ ]:

Fig. 2

`df_toronto.head()`

[8]:

	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	43.662301	-79.389494

Fig. 3

- We created a map using folium and color coded the restaurants: the Japanese ones are marked with red
- We then filtered the Japanese restaurants and then created a map with all of them in the entire Toronto

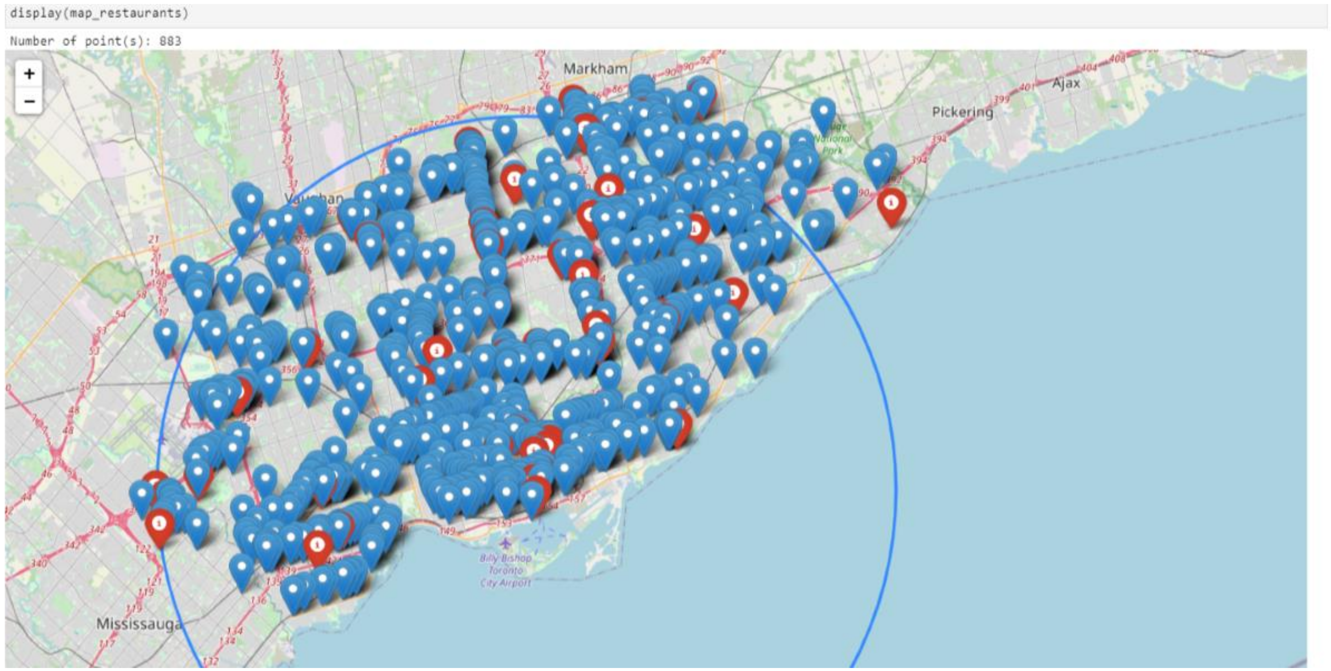


Fig.4

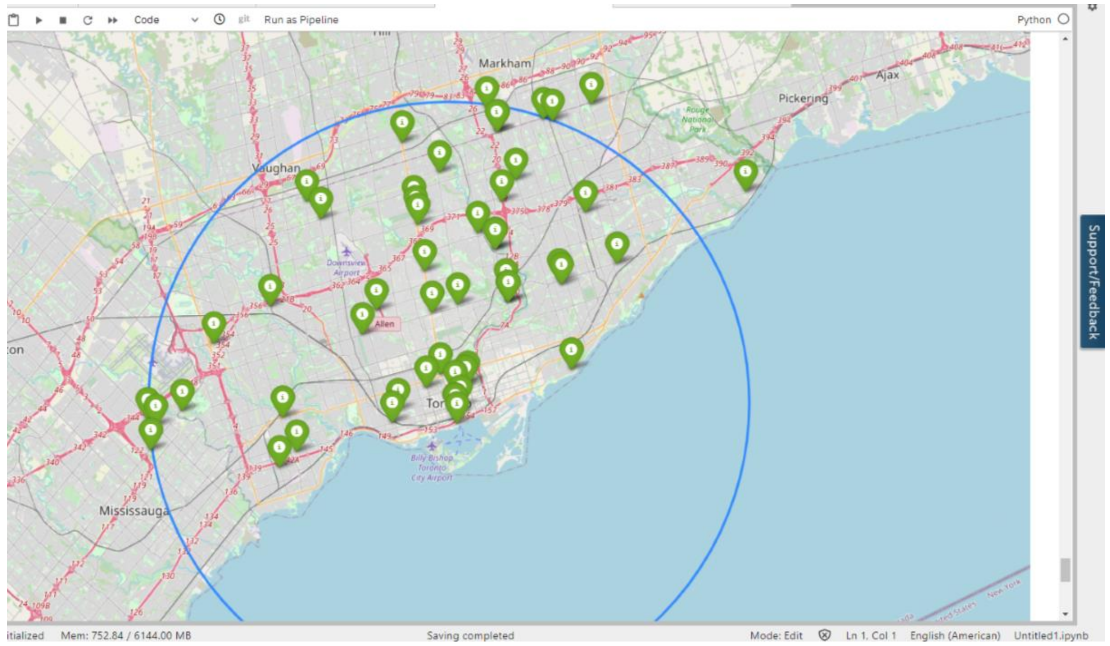


Fig. 5

- To analyze the data we performed a technique in which Categorical Data is transformed into Numerical Data for Machine Learning algorithms
- This technique is called One hot encoding
- For each of the neighborhoods, individual restaurants were turned into the frequency at how many of those restaurants were located in each neighborhood

```
neighborhood_grouped = neighborhood_onehot.groupby('Neighborhood').mean().reset_index()
neighborhood_grouped
```

[27]:

	Neighborhood	Afghan	African	American	Asian	Brazilian	Cajun / Creole	Cantonese	Caribbean	Caucasian	...	Szechuan	Taiwanese	Tapas	Thai	Theme Restaurant	Tibetan	Turkish	Vegetarian / Vegan	Vietnamese
0	Agincourt	0.0	0.0	0.000000	0.071429	0.0	0.0	0.071429	0.071429	0.0	...	0.0	0.0	0.0	0.000000	0.0	0.0	0.000000	0.0	0.071429
1	Alderwood, Long Branch	0.0	0.0	0.066667	0.000000	0.0	0.0	0.000000	0.000000	0.0	...	0.0	0.0	0.0	0.000000	0.0	0.0	0.000000	0.0	0.000000
2	Bathurst Manor, Wilson Heights, Downsview North	0.0	0.0	0.000000	0.038462	0.0	0.0	0.000000	0.000000	0.0	...	0.0	0.0	0.0	0.076923	0.0	0.0	0.000000	0.0	0.000000
3	Bayview Village	0.0	0.0	0.000000	0.000000	0.0	0.0	0.000000	0.000000	0.0	...	0.0	0.0	0.0	0.058824	0.0	0.0	0.000000	0.0	0.000000
4	Bedford Park, Lawrence Manor East	0.0	0.0	0.000000	0.111111	0.0	0.0	0.000000	0.000000	0.0	...	0.0	0.0	0.0	0.000000	0.0	0.0	0.111111	0.0	0.000000
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
78	Willowdale West	0.0	0.0	0.000000	0.000000	0.0	0.0	0.000000	0.000000	0.0	...	0.0	0.0	0.0	0.000000	0.0	0.0	0.000000	0.0	0.100000
79	Willowdale, Newtonbrook	0.0	0.0	0.000000	0.000000	0.0	0.0	0.040000	0.000000	0.0	...	0.0	0.0	0.0	0.000000	0.0	0.0	0.000000	0.0	0.000000
80	Woburn	0.0	0.0	0.071429	0.000000	0.0	0.0	0.000000	0.071429	0.0	...	0.0	0.0	0.0	0.071429	0.0	0.0	0.000000	0.0	0.000000
81	Woodbine Heights	0.0	0.0	0.117647	0.117647	0.0	0.0	0.000000	0.058824	0.0	...	0.0	0.0	0.0	0.058824	0.0	0.0	0.058824	0.0	0.000000
82	York Mills, Silver Hills	0.0	0.0	0.000000	0.000000	0.0	0.0	0.000000	0.000000	0.0	...	0.0	0.0	0.0	0.000000	0.0	0.0	0.000000	0.0	0.000000

83 rows × 62 columns

Support/Feedback

Fig. 6

- We then searched for the most commons restaurants for each neighborhood and overall. We can see that overall the Japanese Restaurants are on the 5th position.

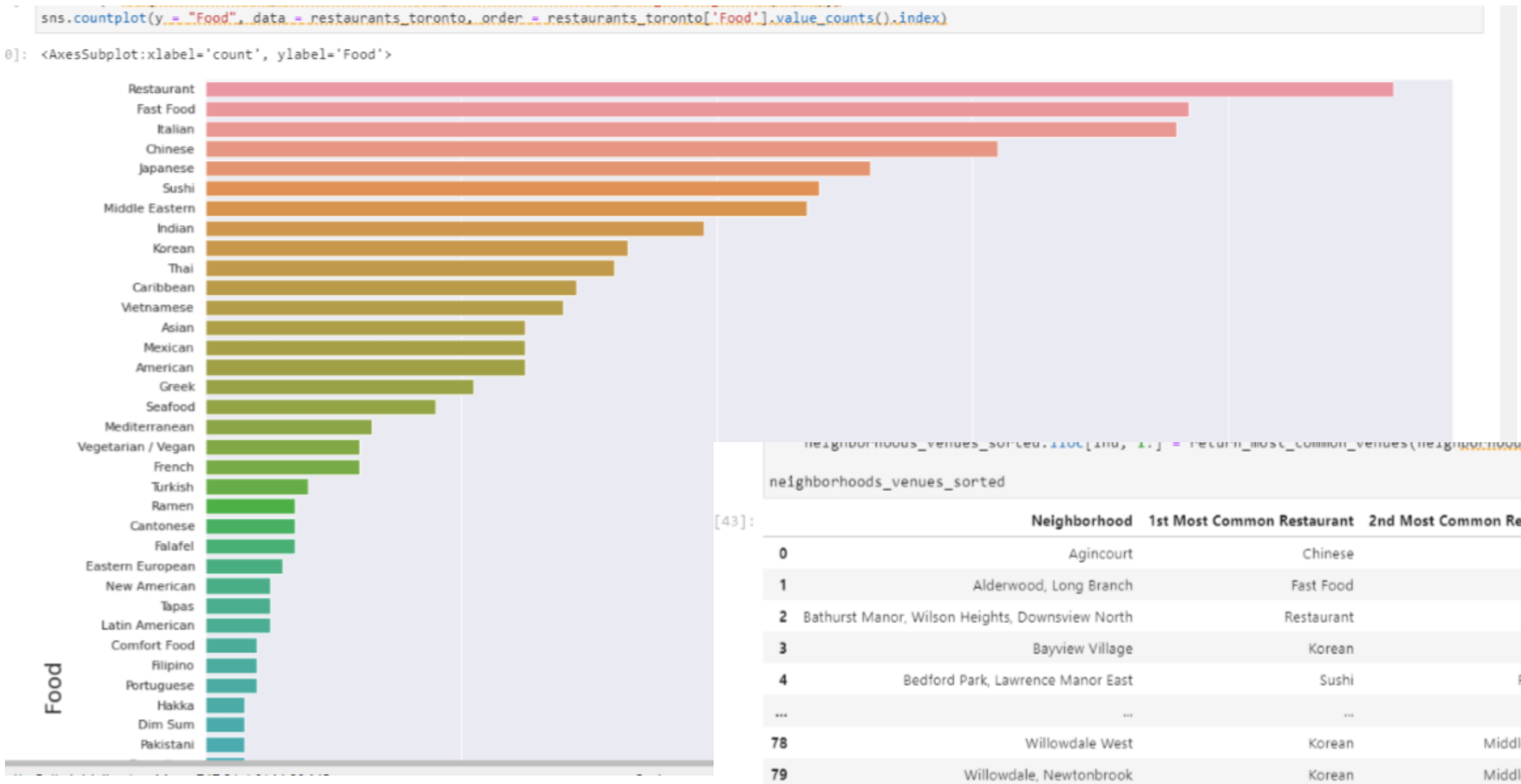


Fig. 7

Fig. 8

```
neighborhoods_venues_sorted
```

	Neighborhood	1st Most Common Restaurant	2nd Most Common Restaurant	3rd Most Common Restaurant	4th Most Common Restaurant	5th Most Common Restaurant
0	Agincourt	Chinese	Indian	Fast Food	Cantonese	Restaurant
1	Alderwood, Long Branch	Fast Food	Seafood	Restaurant	Chinese	Middle Eastern
2	Bathurst Manor, Wilson Heights, Downsview North	Restaurant	Sushi	Middle Eastern	Japanese	Ramen
3	Bayview Village	Korean	Sushi	Ramen	Middle Eastern	Seafood
4	Bedford Park, Lawrence Manor East	Sushi	Fast Food	Turkish	Asian	Japanese
...	...	...	...	...	...	...
78	Willowdale West	Korean	Middle Eastern	Sushi	Vietnamese	Restaurant
79	Willowdale, Newtonbrook	Korean	Middle Eastern	Sushi	Fast Food	Indian
80	Woburn	Indian	Fast Food	Restaurant	Caribbean	Hakka
81	Woodbine Heights	Indian	American	Asian	Caribbean	Fast Food
82	York Mills, Silver Hills	Restaurant	Italian	Chinese	Indian	Middle Eastern

83 rows × 6 columns

- We also wanted to cluster the neighborhoods to see which has the most Japanese Restaurants
- In this technique we ran a test with different number of K values and measured the accuracy and then chose the best K value

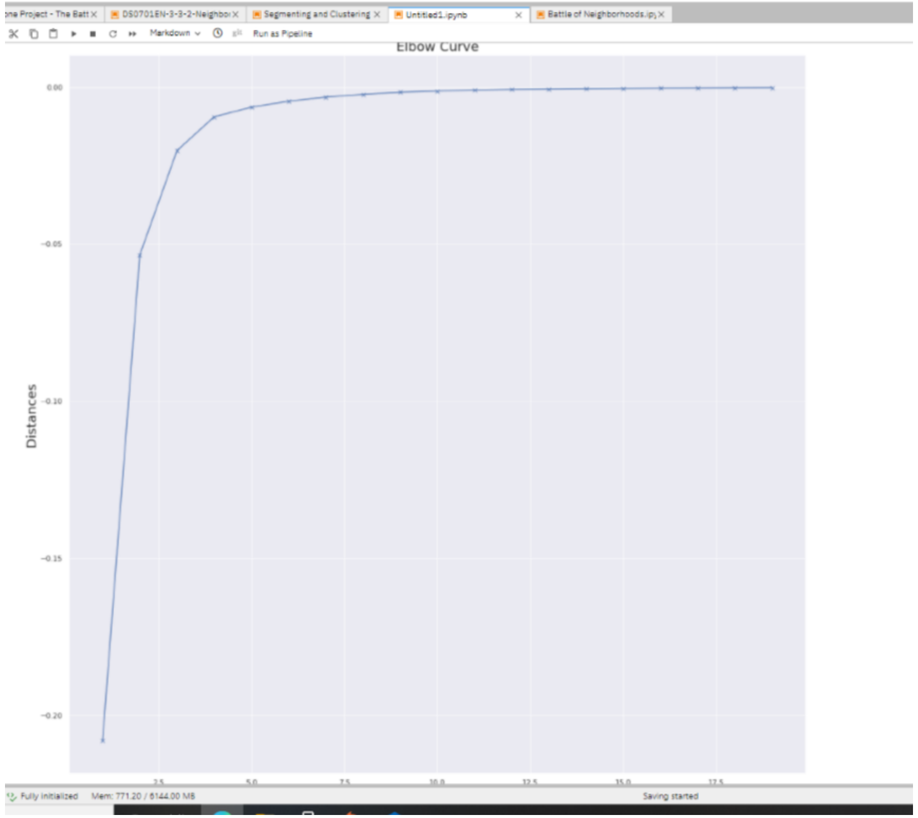
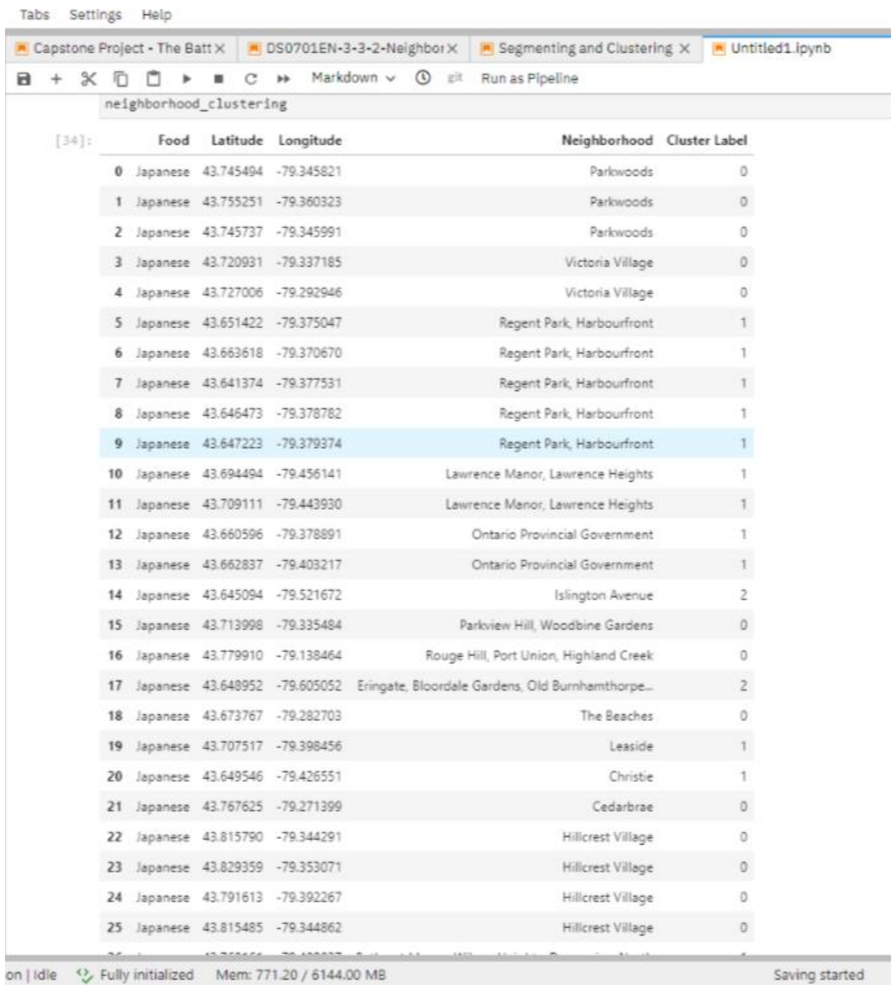


Fig. 9

- We see that the Elbow is at K=3
- Moreover, in K-Means clustering, objects that are similar based on a certain variable are put into the same cluster
- Neighborhoods that had similar mean frequency of Japanese Restaurants were divided into 3 clusters
- Each of these clusters were labelled from 0 to 3 as the indexing of labels begin with 0 instead of



	Food	Latitude	Longitude	Neighborhood	Cluster Label
0	Japanese	43.745494	-79.345821	Parkwoods	0
1	Japanese	43.755251	-79.360323	Parkwoods	0
2	Japanese	43.745737	-79.345991	Parkwoods	0
3	Japanese	43.720931	-79.337185	Victoria Village	0
4	Japanese	43.727006	-79.292946	Victoria Village	0
5	Japanese	43.651422	-79.375047	Regent Park, Harbourfront	1
6	Japanese	43.663618	-79.370670	Regent Park, Harbourfront	1
7	Japanese	43.641374	-79.377531	Regent Park, Harbourfront	1
8	Japanese	43.646473	-79.378782	Regent Park, Harbourfront	1
9	Japanese	43.647223	-79.379374	Regent Park, Harbourfront	1
10	Japanese	43.694494	-79.456141	Lawrence Manor, Lawrence Heights	1
11	Japanese	43.709111	-79.443930	Lawrence Manor, Lawrence Heights	1
12	Japanese	43.660596	-79.378891	Ontario Provincial Government	1
13	Japanese	43.662837	-79.403217	Ontario Provincial Government	1
14	Japanese	43.645094	-79.521672	Islington Avenue	2
15	Japanese	43.713998	-79.335484	Parkview Hill, Woodbine Gardens	0
16	Japanese	43.779910	-79.138464	Rouge Hill, Port Union, Highland Creek	0
17	Japanese	43.648952	-79.605052	Eringate, Bloordale Gardens, Old Burnhamthorpe...	2
18	Japanese	43.673767	-79.282703	The Beaches	0
19	Japanese	43.707517	-79.398456	Leaside	1
20	Japanese	43.649546	-79.426551	Christie	1
21	Japanese	43.767625	-79.271399	Cedarbrae	0
22	Japanese	43.815790	-79.344291	Hillcrest Village	0
23	Japanese	43.829359	-79.353071	Hillcrest Village	0
24	Japanese	43.791613	-79.392267	Hillcrest Village	0
25	Japanese	43.815485	-79.344862	Hillcrest Village	0

Fig. 10

After, we merged the venue data with the table on the left creating a new table which would be the basis for analyzing new opportunities for opening a new Japanese Restaurant in Toronto. Then we created a map using the Folium package in Python and each neighborhood was colored based on the cluster label. For example, cluster 1 was purple and cluster 2 was green.

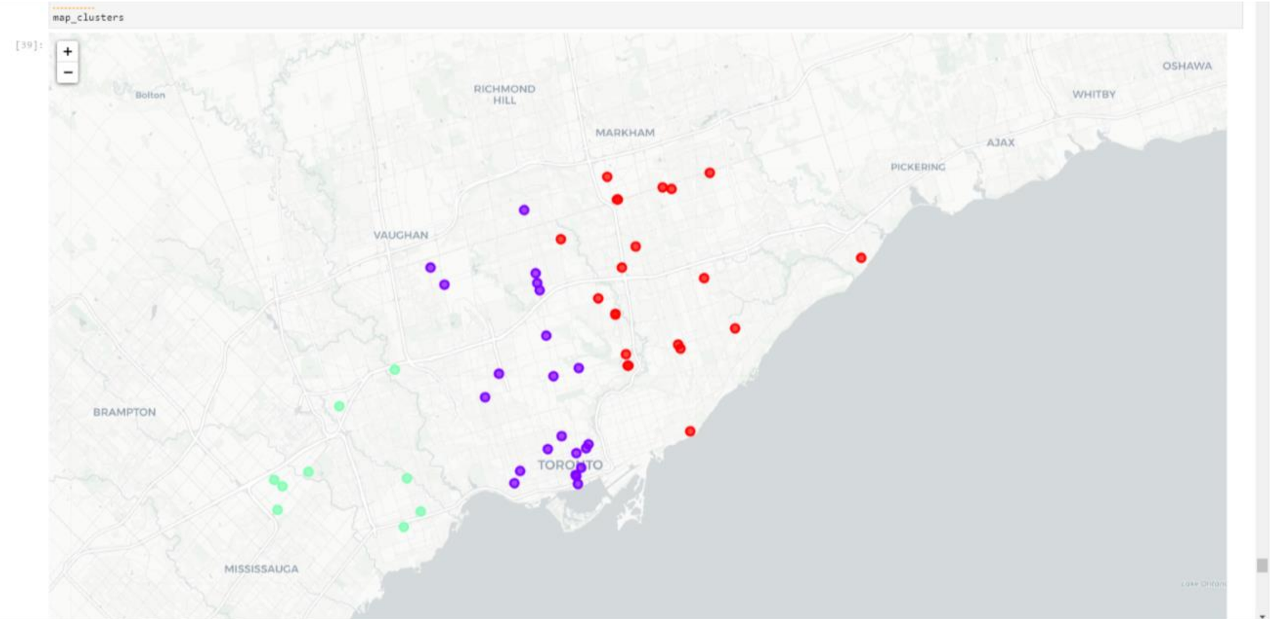


Fig. 11

```
[40]: ## Cluster 1 in neighborhood_merged
neighborhood_cluster1 = neighborhood_clustering[neighborhood_clustering['Cluster Label'] == 0]
neighborhood_cluster1
```

[40]:

	Food	Latitude	Longitude	Neighborhood	Cluster Label
0	Japanese	43.745494	-79.345821	Parkwoods	0
1	Japanese	43.823043	-79.306446	Milliken, Agincourt North, Steeles East, L'Amo...	0
2	Japanese	43.831832	-79.266304	Milliken, Agincourt North, Steeles East, L'Amo...	0
3	Japanese	43.822372	-79.298905	Milliken, Agincourt North, Steeles East, L'Amo...	0
4	Japanese	43.724299	-79.290987	Kennedy Park, Ionview, East Birchmount Park	0
5	Japanese	43.774171	-79.340401	Fairview, Henry Farm, Oriole	0
6	Japanese	43.787145	-79.328940	Fairview, Henry Farm, Oriole	0
7	Japanese	43.737003	-79.244750	Scarborough Village	0
8	Japanese	43.714044	-79.335000	Thorncliffe Park	0
9	Japanese	43.791613	-79.392267	Hillcrest Village	0
10	Japanese	43.829359	-79.353071	Hillcrest Village	0
11	Japanese	43.815790	-79.344291	Hillcrest Village	0
12	Japanese	43.767625	-79.271399	Cedarbrae	
13	Japanese	43.673767	-79.282703	The Beaches	
14	Japanese	43.779910	-79.138464	Rouge Hill, Port Union, Highland Creek	
15	Japanese	43.815485	-79.344862	Hillcrest Village	
16	Japanese	43.755251	-79.360323	Parkwoods	
17	Japanese	43.745737	-79.345991	Parkwoods	
18	Japanese	43.720931	-79.337185	Victoria Village	
19	Japanese	43.727006	-79.292946	Victoria Village	
20	Japanese	43.713998	-79.335484	Parkview Hill, Woodbine Gardens	

Fig. 12 Cluster 1

```
[41]: ## Cluster 2 in neighborhood_merged
neighborhood_cluster2 = neighborhood_clustering[neighborhood_clustering['Cluster Label'] == 1]
neighborhood_cluster2
```

[41]:

	Food	Latitude	Longitude	Neighborhood	Cluster Label
21	Japanese	43.641846	-79.431086	Little Portugal, Trinity	1
22	Japanese	43.670683	-79.391056	Moore Park, Summerhill East	1
23	Japanese	43.712637	-79.376914	Lawrence Park	1
24	Japanese	43.774224	-79.501720	Downsview Northwest	1
25	Japanese	43.763523	-79.490136	Downsview Northwest	1
26	Japanese	43.732562	-79.404147	Bedford Park, Lawrence Manor East	1
27	Japanese	43.809168	-79.423015	Willowdale, Newtonbrook	1
28	Japanese	43.651422	-79.375047	Regent Park, Harbourfront	1
29	Japanese	43.663618	-79.370670	Regent Park, Harbourfront	1
30	Japanese	43.665895	-79.368415	Rosedale	1
31	Japanese	43.646473	-79.378782	Regent Park, Harbourfront	1
32	Japanese	43.764412	-79.411880	Bathurst Manor, Wilson Heights, Downsview North	1
33	Japanese	43.770806	-79.413138	Bathurst Manor, Wilson Heights, Downsview North	1
		3.760161	-79.409827	Bathurst Manor, Wilson Heights, Downsview North	1
		3.647223	-79.379374	Regent Park, Harbourfront	1
		3.694494	-79.456141	Lawrence Manor, Lawrence Heights	1
		3.709111	-79.443930	Lawrence Manor, Lawrence Heights	1
		3.660596	-79.378891	Ontario Provincial Government	1
		3.641374	-79.377531	Regent Park, Harbourfront	1
		3.662837	-79.403217	Ontario Provincial Government	1
		3.707517	-79.398456	Leaside	1
42	Japanese	43.649546	-79.426551	Christie	1

in neighborhood_merged

Fig. 13 Cluster 2

```
[42]: ## Cluster 3 in neighborhood_merged
neighborhood_cluster3 = neighborhood_clustering[neighborhood_clustering['Cluster Label'] =
neighborhood_cluster3
```

[42]:

	Food	Latitude	Longitude	Neighborhood	Cluster Label
43	Japanese	43.615389	-79.524415	New Toronto, Mimico South, Humber Bay Shores	2
44	Japanese	43.648952	-79.605052	Eringate, Bloordale Gardens, Old Burnhamthorpe...	2
45	Japanese	43.640005	-79.626960	Enclave of L4W	2
46	Japanese	43.625661	-79.630920	Enclave of L4W	2
47	Japanese	43.644181	-79.633899	Enclave of L4W	2
48	Japanese	43.711507	-79.531797	Humberlea, Emery	2
49	Japanese	43.624728	-79.509904	New Toronto, Mimico South, Humber Bay Shores	2
50	Japanese	43.689193	-79.578441	Kingsview Village, St. Phillips, Martin Grove ...	2
51	Japanese	43.645094	-79.521672	Islington Avenue	2

Fig. 14 Cluster 3

Discussion

- Most of the Japanese Restaurants are in cluster 1 and cluster 2 represented by the red and purple clusters. The neighborhoods in these areas that have the highest average of Japanese Restaurants are Hillcrest Village and Regent Park.
- We can see that in the cluster 3 there are significantly fewer Japanese restaurants. However, in this area the most restaurants are in Enclave of L4W.
- Looking at the nearby venues, the optimum place to put a new Japanese restaurant is in the cluster 3 area as there are many Neighborhoods in the area but little to no Japanese restaurants.

Conclusion

- In conclusion, to end off this project, we had an opportunity on a business problem, and it was tackled in way that it was similar to how a genuine data scientist would do
- We utilized numerous Python libraries to fetch the information , to control the content and to break down and visualize those datasets
- We have utilized Foursquare API to investigate the settings in neighborhoods of Toronto, get great measure of data from Wikipedia which we scraped with the Beautifulsoup Web scraping Library
- We also visualized utilizing different plots present in seaborn and matplotlib libraries