Capstone Project – The Battle of Neighborhoods

Final Report

Version: 1.0
February 2020

Johnnie Ng

Table of Contents

1 BACKGROUND

Toronto is the provincial capital of Ontario and the most populous city in Canada, with a population of 2,954,024 as of July 2018. 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. People have travelled through and inhabited the Toronto area for more than 10,000 years. The diverse population of Toronto reflects its current and historical role as an important destination for immigrants to Canada. For Chinese people who are new to Toronto, how do they know what the best Chinese restaurants are worth going to and where they are?

2 PROBLEM

This project aims to create a simple Chinese Restaurants Guide based on Foursquare Rating, restaurant category and geographic location data for restaurants in Toronto. Finally, these restaurants will be clustered based on their similarities.

3 INTEREST

Obviously, Chinese people who will consider visiting Toronto will be interested to explore Chinese restaurants to get something to eat. Others who want to try the Chinese food may also be interested.

4 DATA ACQUISTION AND CLEANING

(a) Scrape the following Wikipedia page

https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

in order to obtain the data that is in the table of postal codes and to transform the data into a pandas dataframe.

(b) Get geolocator latitude and longitude coordinates for Toronto.

(c) Use Foursquare API to get a list of venues in Toronto.

(d) Use Foursquare API to get restaurant name, ID, location, category and rating.

5 METHODOLOGY

(a) Use Beautiful Soup library in python to scrape the Wikipedia page to extract the data in the tabular format as shown in the website.

```
In [43]: print(tabulate(df2, headers='keys', tablefmt='psql') )

|  41 | M4K          | East Toronto     | The Danforth West, Riverdale
|  42 | M4L          | East Toronto     | The Beaches West, India Bazaar
|  43 | M4M          | East Toronto     | Studio District
|  44 | M4N          | Central Toronto  | Lawrence Park
|  45 | M4P          | Central Toronto  | Davisville North
|  46 | M4R          | Central Toronto  | North Toronto West
|  47 | M4S          | Central Toronto  | Davisville
|  48 | M4T          | Central Toronto  | Moore Park, Summerhill East
|  49 | M4V          | Central Toronto  | Deer Park, Forest Hill SE, Rathnelly, South Hill, Summerhill West
|  50 | M4W          | Downtown Toronto | Rosedale
|  51 | M4X          | Downtown Toronto | Cabbagetown, St. James Town
|  52 | M4Y          | Downtown Toronto | Church and Wellesley
|  53 | M5A          | Downtown Toronto | Harbourfront
|  54 | M5B          | Downtown Toronto | Ryerson, Garden District
|  55 | M5C          | Downtown Toronto | St. James Town
|  56 | M5E          | Downtown Toronto | Berczy Park
|  57 | M5G          | Downtown Toronto | Central Bay Street
|  58 | M5H          | Downtown Toronto | Adelaide, King, Richmond
|  59 | M5J          | Downtown Toronto | Harbourfront East, Toronto Islands, Union Station
|  60 | M5K          | Downtown Toronto | Design Exchange, Toronto Dominion Centre
```

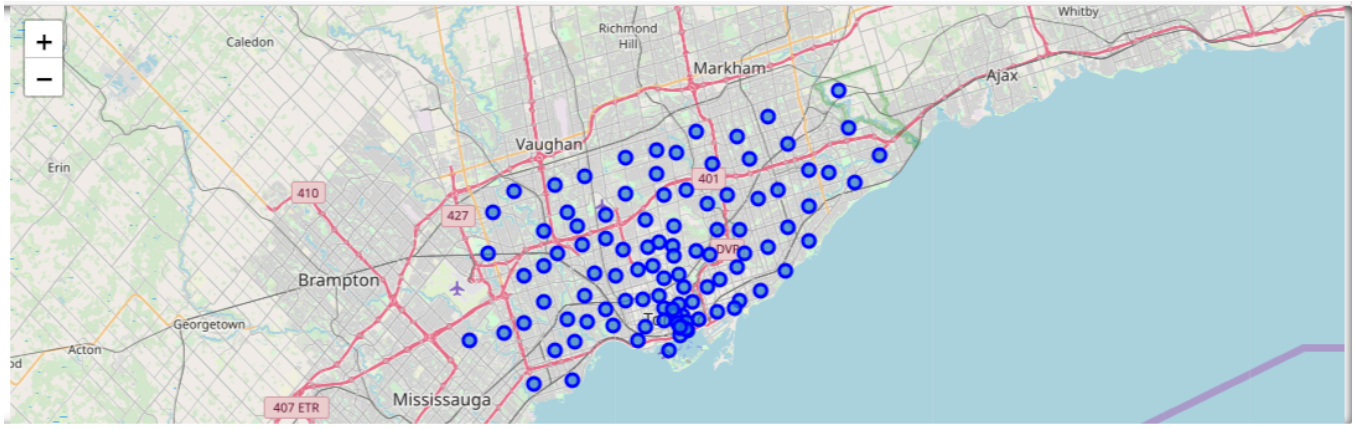
(b) Use the csv file to create the dataframe with Latitude and Longitude and Merge two data sets on the Postal Code to form a new data set for visualization.

```
In [98]: df3 = pd.read_csv("http://cocl.us/Geospatial_data")
         df2['Latitude'] = df3['Latitude'].values
         df2['Longitude'] = df3['Longitude'].values
         df2
```

Out[98]:

	PostalCode	Borough	Neighborhood	Latitude	Longitude
0	M1B	Scarborough	Rouge, Malvern	43.806686	-79.194353
1	M1C	Scarborough	Highland Creek, Rouge Hill, Port Union	43.784535	-79.160497
2	M1E	Scarborough	Guildwood, Morningside, West Hill	43.763573	-79.188711
3	M1G	Scarborough	Woburn	43.770992	-79.216917
4	M1H	Scarborough	Cedarbrae	43.773136	-79.239476
5	M1J	Scarborough	Scarborough Village	43.744734	-79.239476
6	M1K	Scarborough	East Birchmount Park, Ionview, Kennedy Park	43.727929	-79.262029
7	M1L	Scarborough	Clairlea, Golden Mile, Oakridge	43.711112	-79.284577
8	M1M	Scarborough	Cliffcrest, Cliffside, Scarborough Village West	43.716316	-79.239476
9	M1N	Scarborough	Birch Cliff, Cliffside West	43.692657	-79.264848
10	M1P	Scarborough	Dorset Park, Scarborough Town Centre, Wexford ...	43.757410	-79.273304

(c) Use geopy library to get the latitude and longitude values of Toronto and create a map of Toronto with neighborhoods superimposed on top.

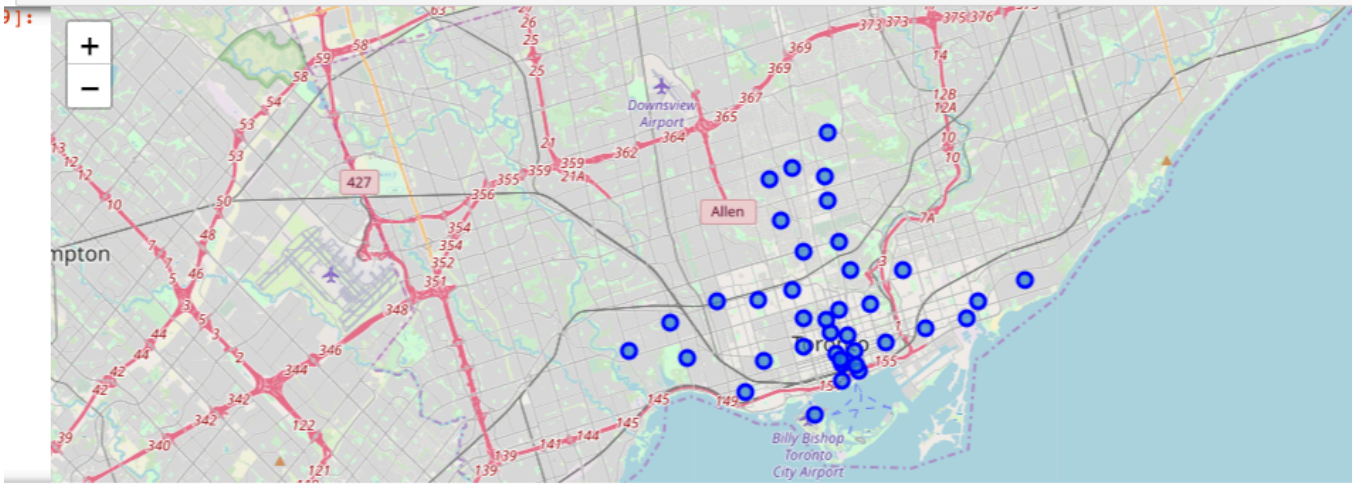


(d) Work with only boroughs that contain the word Toronto.

```
# create map of Tornoto using latitude and longitude values
map_tornoto = folium.Map(location=[latitude, longitude], zoom_start=11)

# add markers to map
for lat, lng, label in zip(toronto_data['Latitude'], toronto_data['Longitude'], toronto_data['Neighborhood']):
    label = folium.Popup(label, parse_html=True)
    folium.CircleMarker(
        [lat, lng],
        radius=5,
        popup=label,
        color='blue',
        fill=True,
        fill_color='#3186cc',
        fill_opacity=0.7,
        parse_html=False).add_to(map_tornoto)

map_tornoto
```



(e) Use Foursquare API to explore all Chinese Restaurants in Toronto with radius = 500.

```
In [81]: toronto_venues = toronto_venues[toronto_venues['Venue Category'].str.contains("Chinese Restaurant")]
         print(toronto_venues.shape)
         toronto_venues
```

```
(13, 8)
```

Out[81]:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue ID	Venue Latitude	Venue Longitude	Venue Category
122	North Toronto West	43.715383	-79.405678	C'est Bon	4aec79e5f964a5209fc721e3	43.716785	-79.400406	Chinese Restaurant
234	Cabbagetown, St. James Town	43.667967	-79.367675	China Gourmet	4bca992068f976b017d35f83	43.664180	-79.368359	Chinese Restaurant
296	Church and Wellesley	43.665860	-79.383160	Crown Princess Fine Dining 伯爵名宴	4c792e4981bca093af5efc14	43.666455	-79.387698	Chinese Restaurant
442	Ryerson, Garden District	43.657162	-79.378937	GB Hand-Pulled Noodles	58e91d60e0adac258fcea481	43.656434	-79.383783	Chinese Restaurant
649	Central Bay Street	43.657952	-79.387383	Yueh Tung Chinese Restaurant	52a7ae41498eed3af4d0a3fa	43.655281	-79.385337	Chinese Restaurant
650	Central Bay Street	43.657952	-79.387383	GB Hand-Pulled Noodles	58e91d60e0adac258fcea481	43.656434	-79.383783	Chinese Restaurant
836	Harbourfront East, Toronto Islands, Union Station	43.640816	-79.381752	Pearl Harbourfront	4ae33054f964a520759121e3	43.638157	-79.380688	Chinese Restaurant
1162	Harbord, University of Toronto	43.662696	-79.400049	River Tai Restaurant	4b340703f964a5200d2425e3	43.662902	-79.403167	Chinese Restaurant
1213	Chinatown, Grange Park, Kensington Market	43.653206	-79.400049	Swatow Restaurant 汕頭小食家	4ae29812f964a520288f21e3	43.653866	-79.398334	Chinese Restaurant
1225	Chinatown, Grange Park, Kensington Market	43.653206	-79.400049	Rosewood Chinese Cuisine	4fd3ea3ee4b0191b9c2b1aff	43.653171	-79.396710	Chinese Restaurant
1226	Chinatown, Grange Park, Kensington Market	43.653206	-79.400049	New Sky Restaurant 小沙田食家	4b074bb1f964a52077fb22e3	43.655337	-79.398897	Chinese Restaurant
1242	Chinatown, Grange Park, Kensington Market	43.653206	-79.400049	Asian Legend 味香村	4adb5472f964a520fc2521e3	43.653603	-79.395047	Chinese Restaurant
1682	Queen's Park	43.662301	-79.389494	Crown Princess Fine Dining 伯爵名宴	4c792e4981bca093af5efc14	43.666455	-79.387698	Chinese Restaurant

(f) Use Foursquare API to get rating of Chinese Restaurants based on venue id.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue ID	Venue Latitude	Venue Longitude	Venue Category	Rating
122	North Toronto West	43.715383	-79.405678	C'est Bon	4aec79e5f964a5209fc721e3	43.716785	-79.400406	Chinese Restaurant	7.4
234	Cabbagetown, St. James Town	43.667967	-79.367675	China Gourmet	4bca992068f976b017d35f83	43.664180	-79.368359	Chinese Restaurant	6.0
296	Church and Wellesley	43.665860	-79.383160	Crown Princess Fine Dining 伯爵名宴	4c792e4981bca093af5efc14	43.666455	-79.387698	Chinese Restaurant	7.5
442	Ryerson, Garden District	43.657162	-79.378937	GB Hand-Pulled Noodles	58e91d60e0adac258fcea481	43.656434	-79.383783	Chinese Restaurant	7.8
649	Central Bay Street	43.657952	-79.387383	Yueh Tung Chinese Restaurant	52a7ae41498eed3af4d0a3fa	43.655281	-79.385337	Chinese Restaurant	7.9
650	Central Bay Street	43.657952	-79.387383	GB Hand-Pulled Noodles	58e91d60e0adac258fcea481	43.656434	-79.383783	Chinese Restaurant	7.8
836	Harbourfront East, Toronto Islands, Union Station	43.640816	-79.381752	Pearl Harbourfront	4ae33054f964a520759121e3	43.638157	-79.380688	Chinese Restaurant	8.2
1162	Harbord, University of Toronto	43.662696	-79.400049	River Tai Restaurant	4b340703f964a5200d2425e3	43.662902	-79.403167	Chinese Restaurant	6.6
1213	Chinatown, Grange Park, Kensington Market	43.653206	-79.400049	Swatow Restaurant 汕頭小食家	4ae29812f964a520288f21e3	43.653866	-79.398334	Chinese Restaurant	7.7
1225	Chinatown, Grange Park, Kensington Market	43.653206	-79.400049	Rosewood Chinese Cuisine	4fd3ea3ee4b0191b9c2b1aff	43.653171	-79.396710	Chinese Restaurant	7.7
1226	Chinatown, Grange Park, Kensington Market	43.653206	-79.400049	New Sky Restaurant 小沙田食家	4b074bb1f964a52077fb22e3	43.655337	-79.398897	Chinese Restaurant	7.6
1242	Chinatown, Grange Park, Kensington Market	43.653206	-79.400049	Asian Legend 味香村	4adb5472f964a520fc2521e3	43.653603	-79.395047	Chinese Restaurant	7.8
1682	Queen's Park	43.662301	-79.389494	Crown Princess Fine Dining 伯爵名宴	4c792e4981bca093af5efc14	43.666455	-79.387698	Chinese Restaurant	7.5

(g) Use k-means clustering algorithm that groups Chinese Restaurants into 4 clusters.

6 RESULTS

(a) Cluster 1

Cluster Labels	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue ID	Venue Latitude	Venue Longitude	Venue Category	Rating	Rating_cat
0	North Toronto West	43.715383	-79.405678	C'est Bon	4aec79e5f964a5209fc721e3	43.716785	-79.400406	Chinese Restaurant	7.4	above avg
0	Church and Wellesley	43.665860	-79.383160	Crown Princess Fine Dining 伯爵名宴	4c792e4981bca093af5efc14	43.666455	-79.387698	Chinese Restaurant	7.5	above avg
0	Ryerson, Garden District	43.657162	-79.378937	GB Hand-Pulled Noodles	58e91d60e0adac258fcea481	43.656434	-79.383783	Chinese Restaurant	7.8	above avg
0	Central Bay Street	43.657952	-79.387383	Yueh Tung Chinese Restaurant	52a7ae41498eed3af4d0a3fa	43.655281	-79.385337	Chinese Restaurant	7.9	above avg
0	Central Bay Street	43.657952	-79.387383	GB Hand-Pulled Noodles	58e91d60e0adac258fcea481	43.656434	-79.383783	Chinese Restaurant	7.8	above avg
0	Chinatown, Grange Park, Kensington Market	43.653206	-79.400049	Swatow Restaurant 汕頭小食家	4ae29812f964a520288f21e3	43.653866	-79.398334	Chinese Restaurant	7.7	above avg
0	Chinatown, Grange Park, Kensington Market	43.653206	-79.400049	Rosewood Chinese Cuisine	4fd3ea3ee4b0191b9c2b1aff	43.653171	-79.396710	Chinese Restaurant	7.7	above avg
0	Chinatown, Grange Park, Kensington Market	43.653206	-79.400049	New Sky Restaurant 小沙田食家	4b074bb1f964a52077fb22e3	43.655337	-79.398897	Chinese Restaurant	7.6	above avg
0	Chinatown, Grange Park, Kensington Market	43.653206	-79.400049	Asian Legend 味香村	4adb5472f964a520fc2521e3	43.653603	-79.395047	Chinese Restaurant	7.8	above avg
0	Queen's Park	43.662301	-79.389494	Crown Princess Fine Dining 伯爵名宴	4c792e4981bca093af5efc14	43.666455	-79.387698	Chinese Restaurant	7.5	above avg

(b) Cluster 2

Cluster Labels	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue ID	Venue Latitude	Venue Longitude	Venue Category	Rating	Rating_cat
1	Harbourfront East, Toronto Islands, Union Station	43.640816	-79.381752	Pearl Harbourfront	4ae33054f964a520759121e3	43.638157	-79.380688	Chinese Restaurant	8.2	great

(c) Cluster 3

Cluster Labels	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue ID	Venue Latitude	Venue Longitude	Venue Category	Rating	Rating_cat
2	Harbord, University of Toronto	43.662696	-79.400049	River Tai Restaurant	4b340703f964a5200d2425e3	43.662902	-79.403167	Chinese Restaurant	6.6	below avg

(d) Cluster 4

Cluster Labels	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue ID	Venue Latitude	Venue Longitude	Venue Category	Rating	Rating_cat
3	Cabbagetown, St. James Town	43.667967	-79.367675	China Gourmet	4bca992068f976b017d35f83	43.66418	-79.368359	Chinese Restaurant	6.0	poor

7 DISCUSSION

The aim of this project is to help people who want to visit Toronto and choose the Chinese restaurants based on the rating recommended in Foursquare. For example, if a person is looking for a Chinese restaurant with good rating, we can see that Cluster 1 and 2 have rating above 7.4. If a person is looking for a Chinese restaurant with the highest rating, the Chinese restaurant in the second cluster is suitable. For budget consideration, Chinese restaurants in Cluster 3 and 4 are more suitable and the rating is around 6.0. The preference of Chinese restaurants may vary from person to person, they can select a restaurant based on their own tastes and the proximity.

8 CONCLUSION

This project helps a person get a better understanding of a specific type of restaurant in one area. It is always helpful to make use of technology to stay one step ahead, i.e. finding out more about rating before choosing the place to eat. We have just taken rating as a primary concern to shortlist the best Chinese restaurant in Toronto. The future of this project includes taking other factors such as pricing and number of likes into consideration to filter restaurants based on a predefined budget and other comments.

- END -