Location selection for new restaurant in Toronto

Sijun Mao

March 22, 2020

I. Introduction

1. Background

Toronto is the provincial capital of Ontario and the most populous city in Canada. Toronto is located in the southern Ontario area on the northwestern shore of Lake Ontario, adjacent to New York in the southeast and Ottawa in the northeast, covering an area of 7,125 square kilometers. Half of the residents in the urban area are immigrants from more than 100 nationalities in various countries around the world. With its diverse ethnic characteristics, it brings together more than 140 languages in the world, making it one of the most diverse cities in the world.[1]

Therefore, Toronto also gathers cuisines from various nationalities.

The restaurant industry is not only an important source of jobs and careers, it is vitally important to the success of many other industries in the economy. Restaurant industry offers significant opportunities for advancement and entrepreneurship.

In today's time, besides great food, restaurants are important for meeting friends, relatives; spending some time office meetings.

As a resident of Toronto and a food lover, I would like to explore something about the Restaurant Industry in Toronto.

2. Business Problem

The location selection is one of the most important factors for entrepreneurs who want to open a new restaurant. A good location could not only make the restaurant be easily visible to consumers, they must also be easily accessible. The location could sometime determine the fate of an entrepreneurial restaurant. However, the site selection is arduous and complicated. Even experienced diners will inevitably make mistakes.

Data analysis could be a useful tool that could be of help to location selection for restaurants.

3. Target audience

For most restaurant entrepreneurs, location selection is their top priority. Obviously, current restaurants owners who want to open a branch will also be interested.

II. Data

1. Data Sources

a) List of Neighborhoods in Toronto. This comes from the Wikipedia webpage https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

b) The latitude and the longitude coordinates of each neighborhood. Here we use the Geocoder Python package to get the geographical coordinates of a given postal code.
c) Venue exploration. Here Forsquare API is utilized to get the information of restaurants and shopping malls in each neighborhood, including the name, category, latitude and longitude.

2. Data Cleaning

a) Toronto Neighborhoods scraping from the Wikipedia page.

The BeautifulSoup package, a website scraping library, is used here. The table on the website is transformed to a pandas dataframe as shown in Table 1.

Table 1. Borough and Neighborhoods of Toronto

	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	Queen's Park / Ontario Provincial Government

b) Geographical coordinates of Neighborhoods

The Geocoder Python package is used to get the latitude and the longitude of a given postal code. Then a new dataframe is formed. As table 2 shows.

Table 2. Latitude and Longitude of Neighborhoods

	PostalCode	Borough	Neighborhood	Latitude	Longitude
0	M3A	North York	Parkwoods	43.752420	-79.329242
1	M4A	North York	Victoria Village	43.730600	-79.313265
2	M5A	Downtown Toronto	Regent Park, Harbourfront	43.650295	-79.359166
3	M6A	North York	Lawrence Manor, Lawrence Heights	43.723270	-79.451286
4	M7A	Queen's Park / Ontario Provincial Government	Queen's Park / Ontario Provincial Government	43.661150	-79.391715

c) Venues of Neighborhoods

Foursquare API is utilized to explore the venues in each neighborhood. Here we get up to 100 venues within 2000 meters radius around the coordinate of each neighborhood. The API returns the JSON file. Then I transform it into a dataframe including the Venue name, Latitude, Longitude and Category. As shown in Table3.

Table 3. Venues of Neighborhoods

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Parkwoods	43.752420	-79.329242	Brookbanks Park	43.751976	-79.332140	Park
1	Parkwoods	43.752420	-79.329242	Variety Store	43.751974	-79.333114	Food & Drink Shop
2	Victoria Village	43.730600	-79.313265	Wigmore Park	43.731023	-79.310771	Park
3	Victoria Village	43.730600	-79.313265	Memories of Africa	43.726602	-79.312427	Grocery Store
4	Regent Park, Harbourfront	43.650295	-79.359166	The Distillery Historic District	43.650244	-79.359323	Historic Site

d) Restaurants and shopping malls

Here, I extract all the restaurant from the venues, and change their category value to Restaurant. Because there are various restaurants before correction. This process is for the convenience of later one-hot processing.

Table 4. DataFrame of Restaurants

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
11	Regent Park, Harbourfront	43.650295	-79.359166	Cluny Bistro & Boulangerie	43.650565	-79.357843	Restaurant
16	Regent Park, Harbourfront	43.650295	-79.359166	El Catrin	43.650601	-79.358920	Restaurant
20	Regent Park, Harbourfront	43.650295	-79.359166	Izumi	43.649970	-79.360153	Restaurant
56	Lawrence Manor, Lawrence Heights	43.723270	-79.451286	RH Courtyard Café	43.724874	-79.455536	Restaurant
59	Lawrence Manor, Lawrence Heights	43.723270	-79.451286	JOEY	43.724131	-79.454042	Restaurant

Here we also get all the shopping malls considering about many restaurants would locate in or around shopping malls. Hence, there could be some relationship.

Table 5. Dataframe of shopping malls in Toronto

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
44	Lawrence Manor, Lawrence Heights	43.723270	-79.451286	Yorkdale Shopping Centre	43.725482	-79.452736	Shopping Mall
153	Islington Avenue	43.662299	-79.528195	Thorncrest Plaza	43.662620	-79.532146	Shopping Mall
204	Garden District, Ryerson	43.657363	-79.378180	CF Toronto Eaton Centre	43.654540	-79.380677	Shopping Mall
426	Eringate, Bloordale Gardens, Old Burnhamthorpe...	43.648573	-79.578250	Renforth Plaza	43.651194	-79.579612	Shopping Mall
488	Berczy Park	43.645160	-79.373675	Brookfield Place	43.646791	-79.378769	Shopping Mall
527	Leaside	43.709495	-79.363989	Leaside Village	43.705682	-79.360777	Shopping Mall
579	Central Bay Street	43.656091	-79.384930	CF Toronto Eaton Centre	43.654540	-79.380677	Shopping Mall
794	Fairview, Henry Farm, Oriole	43.780970	-79.347813	CF Fairview Mall	43.777994	-79.343665	Shopping Mall
979	Toronto Dominion Centre, Design Exchange	43.647100	-79.381531	Brookfield Place	43.646791	-79.378769	Shopping Mall
1204	Commerce Court, Victoria Hotel	43.648395	-79.378865	Brookfield Place	43.646791	-79.378769	Shopping Mall
1268	Willowdale, Newtonbrook	43.791475	-79.413605	Iranian Plaza	43.791380	-79.418164	Shopping Mall
1389	Willowdale South	43.768165	-79.407420	Empress Walk	43.768540	-79.412671	Shopping Mall
1425	Downsview Northwest	43.755371	-79.519590	Yorkgate Mall	43.758664	-79.519641	Shopping Mall
1669	Agincourt	43.793940	-79.267976	Dragon Centre	43.791410	-79.272081	Shopping Mall
1673	Agincourt	43.793940	-79.267976	Chartwell Shopping Centre 集友商場	43.797768	-79.270853	Shopping Mall
1820	Clarks Corners, Tam O'Shanter, Sullivan	43.784725	-79.299066	Eaton Centre USA	43.783572	-79.304916	Shopping Mall
2258	First Canadian Place, Underground city	43.648280	-79.381461	Brookfield Place	43.646791	-79.378769	Shopping Mall

e) Onehot encoding

Because some machine learning algorithm can work with categorical data directly, they require all input variables and output variables to be numeric. Therefore, we should do the onehot encoding here.

One hot encoding is a process by which categorical variables are converted into a form that could be provided to ML algorithms to do a better job in prediction.

Category column is transformed into numeric data as shown in table 6. And the particular restaurants data after one hot encoding is also extracted, as shown in table 7.

Table 6. Onehot of all catorgaries

	Neighborhood	Yoga Studio	Accessories Store	Adult Boutique	Airport	Antique Shop	Art Gallery	Art Museum	Arts & Crafts Store	Athletics & Sports	...	Trail	Train Station	Tram Station	Transportation Service	Video Game Store	Video Store	Wine Bar
0	Agincourt	0.000000	0.0	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.00	0.0	0.000000	0.000000	0.0000	0.000000
1	Alderwood, Long Branch	0.000000	0.0	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.00	0.0	0.000000	0.000000	0.0000	0.000000
2	Bayview Village	0.000000	0.0	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.250000	0.00	0.0	0.000000	0.000000	0.0000	0.000000
3	Bedford Park, Lawrence Manor East	0.000000	0.0	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.00	0.0	0.000000	0.000000	0.0000	0.000000
4	Berczy Park	0.000000	0.0	0.000000	0.00	0.000000	0.016129	0.000000	0.000000	0.000000	...	0.000000	0.00	0.0	0.000000	0.000000	0.0000	0.000000
5	Birch Cliff, Cliffside West	0.000000	0.0	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.00	0.0	0.000000	0.000000	0.0000	0.000000
6	Brockton, Parkdale Village, Exhibition Place	0.000000	0.0	0.000000	0.00	0.000000	0.013889	0.000000	0.013889	0.000000	...	0.000000	0.00	0.0	0.000000	0.000000	0.0000	0.000000
7	CN Tower, King and Spadina, Railway Lands, Har...	0.014925	0.0	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.00	0.0	0.000000	0.000000	0.0000	0.000000
8	Caledonia-Fairbanks	0.000000	0.0	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.00	0.0	0.000000	0.000000	0.0000	0.000000

Table 7. Onehot of Restaurant for clustering

	Neighborhood	Restaurant
0	Agincourt	0.352941
1	Alderwood, Long Branch	0.000000
2	Bayview Village	0.000000
3	Bedford Park, Lawrence Manor East	0.391304
4	Berczy Park	0.225806

Now we can use these data to do the clustering process and find the density of restaurants.

III. Methodology

1. We can find something from the location of existing restaurants. These historical location data could help us to understand where is a good location for restaurants, or at least, we could just choose a location around them.
2. According to experience, many restaurants would open in or around shopping malls where are places with a lot customer flows. We can take an exploration to the relationship between the location of shopping malls and restaurants.
3. The density of population is also an important factor for restaurants industry.
4. Clustering algorithm such as k-means could be a good method to find these relationships. This worth a try.

IV. Results

1. Locations of existing restaurants

First of all, locations of restaurants are visualized to a map. Here the folium library is used. Figure 1 shows the location of Neighborhoods.

Figure 2 shows the number of restaurants in each neighborhood. Commerce Court, Victorial Hotel, Richmond, Adelaid, King, First Candadian Place, Underground city, Enclave of M5E, M4L, L4W have the the largest number of restaurants.

As figure 3 shows, locations of restaurants are marked as red dots, locations of shopping malls are marked as blue circle.

In general, we can find that, most restaurants are gathering around downtown area of City Toronto where is a place with the largest human traffic and the most developed business. Also, some restaurants are located around the shopping malls. Last but not least, most restaurants spread along the main streets.

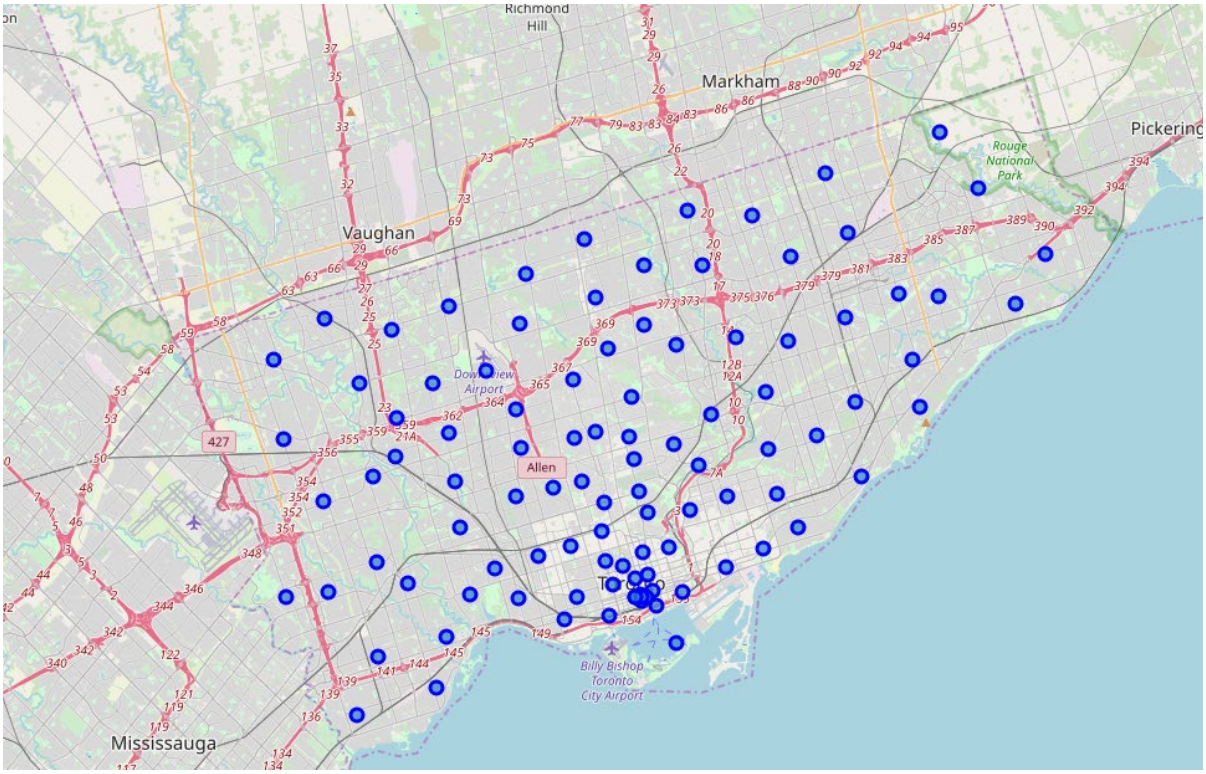

Figure 1. Map of Toronto's Neighborhoods

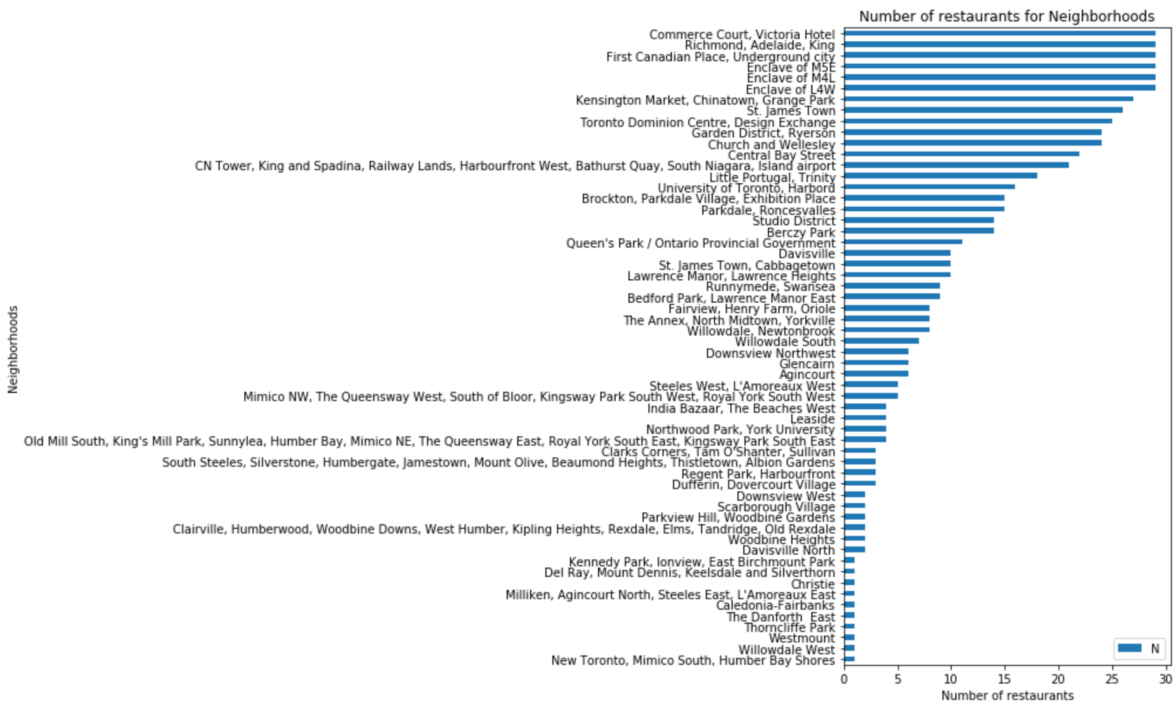

Figure 2. Number of restaurants for Neighborhoods

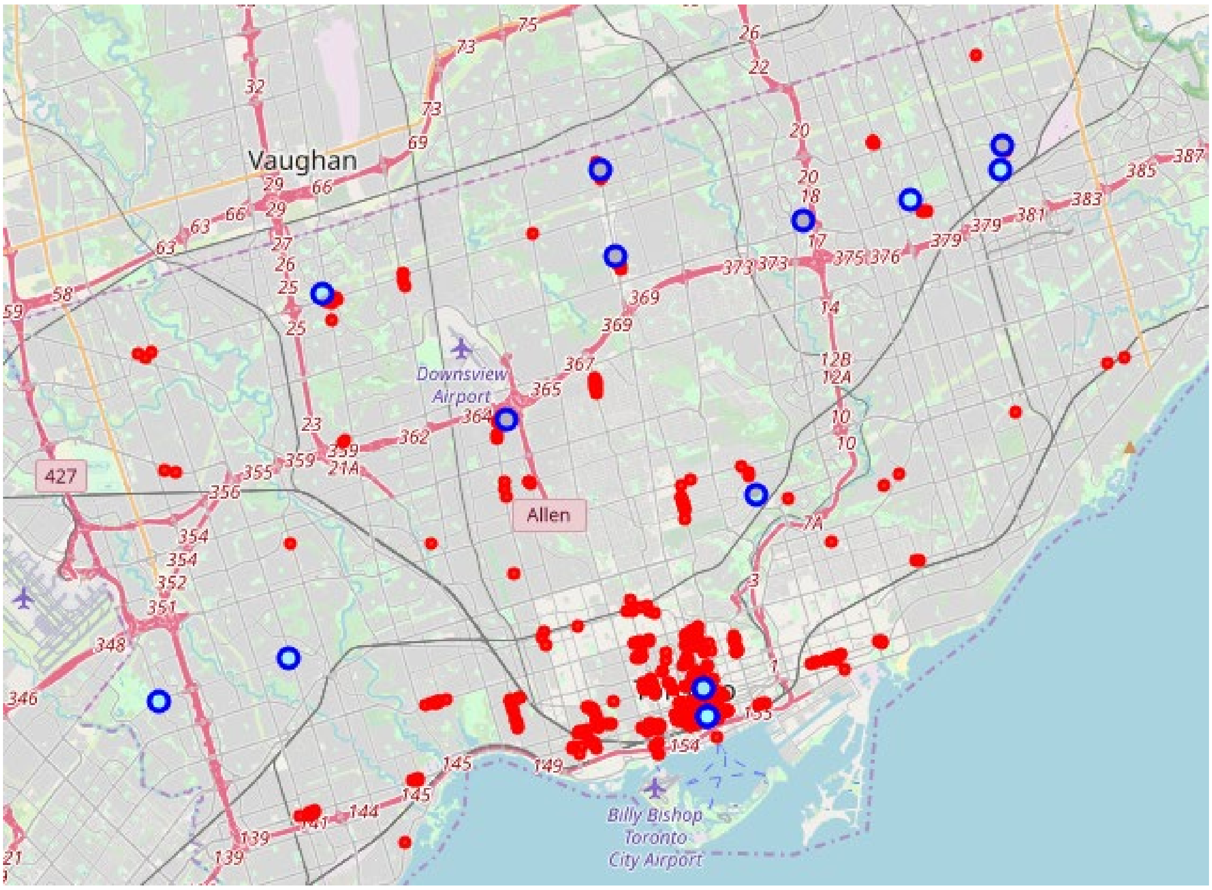

Figure 3. The spread of Restaurants and Shopping Malls

2. Clustering

k-means clustering algorithm is used here. To find the best num of k, the elbow method is implemented. As shown in the figure4, the best number of k is 4. Therefore, the clustering result when k=4 is employed to analyze the spreading of restaurants.

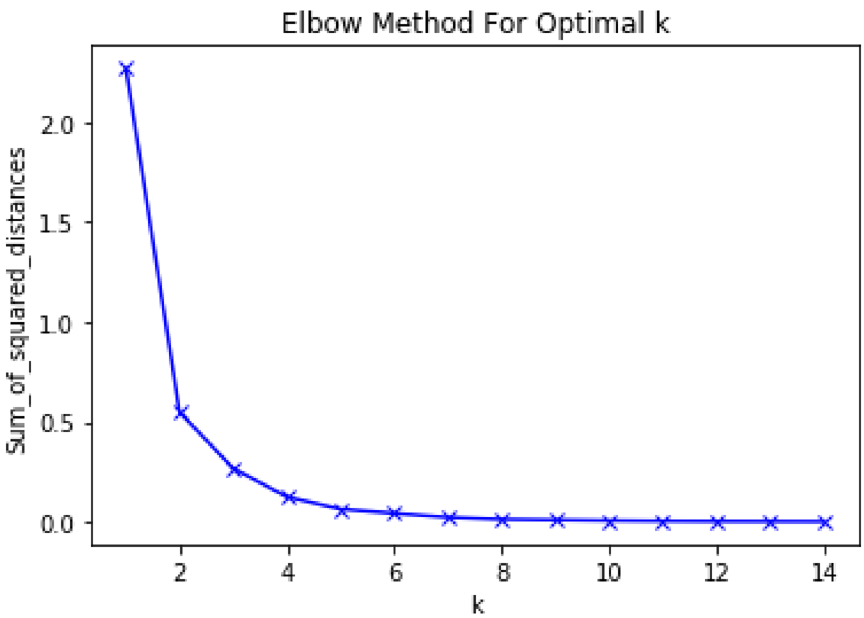

Figure4. elbow method to find the best K

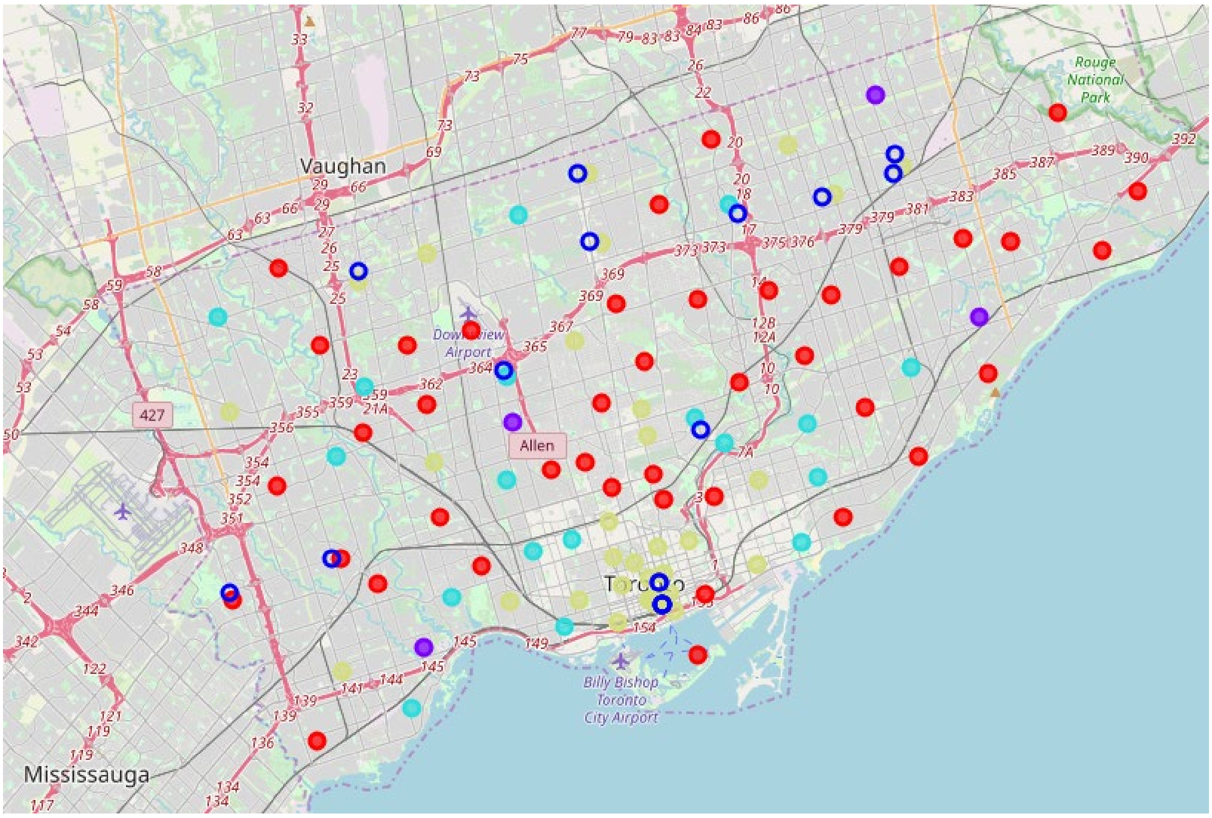

Figure 5. resulting clusters

As shown in figure 5, we get 4 clusters from the k-means algorithm.

1. Cluster 0, marked as red solid circle, represents neighborhoods with no restaurants.
2. Cluster 1, marked as purple solid circle, represents neighborhoods with high density of restaurants.
3. Cluster 2, marked as light blue solid circle, represents neighborhoods with low density of restaurants.
4. Cluster 3, marked as light green solid circle, represents neighborhoods with moderate density of restaurants.

Shopping malls are marked as blue empty circle. The detail results for each cluster are show in Table 8, 9, 10, 11 respectively.

Table 8. Results of Cluster 0

	Neighborhood	Restaurant	Cluster Labels	Latitude	Longitude
1	Alderwood, Long Branch	0.000	0	43.601131	-79.538785
2	Bayview Village	0.000	0	43.781015	-79.380529
5	Birch Cliff, Cliffside West	0.000	0	43.696770	-79.259967
9	Cedarbrae	0.000	0	43.769688	-79.239440
15	Cliffside, Cliffcrest, Scarborough Village West	0.000	0	43.724235	-79.227925
20	Don Mills North	0.000	0	43.749055	-79.362227
21	Don Mills South	0.000	0	43.721425	-79.343454
22	Dorset Park, Wexford Heights, Scarborough Town...	0.000	0	43.759975	-79.268974
23	Downsview Central	0.000	0	43.733690	-79.496740
24	Downsview East	0.000	0	43.738685	-79.467320
31	Eringate, Bloordale Gardens, Old Burnhamthorpe...	0.000	0	43.648573	-79.578250
34	Forest Hill North & West	0.000	0	43.694785	-79.414405
37	Golden Mile, Clairlea, Oakridge	0.000	0	43.713054	-79.285055
38	Guildwood, Morningside, West Hill	0.000	0	43.765815	-79.175193

Table 9. Results of Cluster 1

	Neighborhood	Restaurant	Cluster Labels	Latitude	Longitude
36	Glencairn	0.500000	1	43.707990	-79.448367
55	Milliken, Agincourt North, Steeles East, L'Amo...	0.500000	1	43.817685	-79.280187
62	Old Mill South, King's Mill Park, Sunnylea, Hu...	0.666667	1	43.632835	-79.489550
73	Scarborough Village	0.500000	1	43.743125	-79.231750

Table 10. Results of Cluster 2

	Neighborhood	Restaurant	Cluster Labels	Latitude	Longitude
6	Brockton, Parkdale Village, Exhibition Place	0.208333	2	43.639410	-79.424362
8	Caledonia-Fairbanks	0.111111	2	43.688640	-79.451018
11	Christie	0.090909	2	43.668781	-79.420710
26	Downsview West	0.166667	2	43.720140	-79.516980
27	Dufferin, Dovercourt Village	0.142857	2	43.665087	-79.438705
32	Fairview, Henry Farm, Oriole	0.153846	2	43.780970	-79.347813
45	India Bazaar, The Beaches West	0.200000	2	43.667965	-79.314667
47	Kennedy Park, Ionview, East Birchmount Park	0.200000	2	43.726276	-79.263625
50	Lawrence Manor, Lawrence Heights	0.181818	2	43.723270	-79.451286
52	Leaside	0.111111	2	43.709495	-79.363989
58	New Toronto, Mimico South, Humber Bay Shores	0.125000	2	43.612200	-79.495146
64	Parkview Hill, Woodbine Gardens	0.133333	2	43.707535	-79.311773
71	Runnymede, Swansea	0.209302	2	43.649620	-79.476141
74	South Steeles, Silverstone, Humbergate, Jamest...	0.187500	2	43.743145	-79.584664
85	Thorncliffe Park	0.125000	2	43.701240	-79.349825
89	Westmount	0.200000	2	43.696505	-79.530252
93	Willowdale West	0.125000	2	43.777695	-79.445797
96	Woodbine Heights	0.111111	2	43.689640	-79.306874

Table 11. Results of Cluster 3

	Neighborhood	Restaurant	Cluster Labels	Latitude	Longitude
0	Agincourt	0.352941	3	43.793940	-79.267976
3	Bedford Park, Lawrence Manor East	0.391304	3	43.735460	-79.419148
4	Berczy Park	0.225806	3	43.645160	-79.373675
7	CN Tower, King and Spadina, Railway Lands, Har...	0.313433	3	43.640815	-79.399538
10	Central Bay Street	0.229167	3	43.656091	-79.384930
12	Church and Wellesley	0.279070	3	43.666585	-79.381302
13	Clairville, Humberwood, Woodbine Downs, West H...	0.250000	3	43.711740	-79.579181
14	Clarks Corners, Tam O'Shanter, Sullivan	0.250000	3	43.784725	-79.299066
16	Commerce Court, Victoria Hotel	0.290000	3	43.648395	-79.378865
17	Davisville	0.333333	3	43.703395	-79.385964
18	Davisville North	0.222222	3	43.712755	-79.388514
19	Del Ray, Mount Dennis, Keelsdale and Silverthorn	0.250000	3	43.694530	-79.484489
25	Downsview Northwest	0.272727	3	43.755371	-79.519590
28	Enclave of L4W	0.290000	3	43.648690	-79.385440
29	Enclave of M4L	0.290000	3	43.648690	-79.385440
30	Enclave of M5E	0.290000	3	43.648690	-79.385440

V. Discussion

1. Recommendations

Through the analysis of the results, neighborhoods in cluster 2 and 3 could be good choices of locations. Since cluster 2 and 3 have reasonable density of restaurants, which means there could be reasonable amount of competition between restaurants, and also, these locations have been proven to be profitable for restaurants so that they could survive.

We can also find that there is no restaurant around the two shopping malls, Renforth Plaza and Thorncrest Plaza. There are probable opportunities for opening new restaurants.

2. Limitations

There are many other factors can influence the location selection for restaurants. Such as convenient transportation and easy parking. Generally, restaurants choose places with a large population and places with a large number of passengers, such as integrated commercial centers, schools, pedestrian streets, shopping plazas, office areas, and so on.

However, it is not easy to catch complete data for specific neighborhoods from open source datasets.

Therefore, the recommendation in this project is just a rough prediction. I believe the recommendation process could reach a high level if there are sufficient related data available in the future.

VI. Conclusion

In this project, I analyzed the historical location data of existing restaurants and the relation with shopping malls. I built a clustering model to recognize a reasonable location for new restaurants. It is feasible to do the location recommendation for opening a new restaurant through analyzing data. Although this model is not accurate so far, it can definitely be improved after more factor data involved.