Capstone Project Report

--The Battle of Neighborhoods

Open a restaurant in Toronto

Chenxin Xu

Introduction:

Toronto, the capital of the province of Ontario, is a major Canadian city along Lake Ontario's northwestern shore. The GDP of Toronto in 2016 is 385,934 dollars and the overall population in 2016 is 27,31,571. With this huge market and great geography, opening a restaurant in Toronto can gain huge success. However, based on the data found in yelp.com, which is the main website containing most local restaurant information, there are over 8,000. It also indicates serious competition. Hence, to maintain profits, opening a restaurant in Toronto needs some strategies.

This project aims to help people who would like to open a restaurant in Toronto. To make the final decision, it will be valued from three different perspectives. The first one is the population density in neighborhoods. A restaurant in the neighborhood with more people can create more cash flow and make profits. Second, the decision will consider the ethic in Toronto. Canada is well known for its open immigration policy. Hence, people in Toronto have different backgrounds and preferences for food. Moreover, this project will also value the proportion and the increase of the ethic to ultimate the decision. Finally, other facilities such as shopping malls, supermarkets, and parking lots are also a factor to impact the restaurant. The project will analyze the cluster of the markets in those areas. Another factor that will increase the cost of opening a restaurant is the rent fee. The project will compare the Rent-Geared-to-Income (RGI) which is a housing subsidy or benefit offered by the City of Toronto to make rent affordable for households. After the analysis in those perspectives, it will drive a conclusion that will provide people some advice to open a restaurant. It includes the type of the restaurant, the location, and the operational strategies.

Data:

The data used in this project is the neighborhood profile from the website published by the Toronto government, which contains the population of each neighborhood in 2016, the range of age, the ethic, income, etc. The geographical information also downloads from this website. Data used to find the cluster of the market and sort the RGI is also published by the Toronto government.

Methodology:

This project will visualize and analyze the demography to create an easy and strict view. Moreover, it will use k-means clustering to find the center of the neighborhoods and derive the results for the project goal.

1. Population Density Analysis:

This section will analyze the demography including the population in each neighborhood and the increasing rate. Identifying the density of each neighborhood helps to make the decision. Usually, the restaurant that opens in a popular area can gain a large cash flow; therefore, it is most likely to make a profit and keep running. With knowing that, the analysis of the population density came to the first section of this project.

The data derived from the website published by the Toronto government is the neighborhood profile in 2016. According to the data, it shows that the Toronto government publishes the data once five years. Therefore, the project can only study the data in 2016. The dataset contains lots of information. Toronto has 140 neighborhoods; the dataset has 2383 rows including population, income, education, ethic, etc. Hence, the first step is to slice the data. By using pandas, the data was reduced to contain only the population in 140 neighborhoods and the name of that. This section aims to figure out the neighborhoods that have the most 20 population, which is shown in Figure 1. Since the data type of the original dataset is string when imported to the notebook, it needs to transfer to an integer. Here, I defined a function, which replaces the "," with an empty string and transferred the string to an integer. After converting, the data frame can be sorted in descending order. As Figure 1 shown, Waterfront Communities-The Island has the most population following by Woburn, Willowdale East, and Rouge. To make it more direct to see, I used the folium package to apply a choropleth map. To obtain the map information, I downloaded a GEOJSON file from the government website, which contain the same area ID to

match the profile data. From Figure 2, it clearly shows the area that has the most population with the deeper color.

[89]:

	id	area_id	Population	neighborhood
122	10472	2480154	65913	Waterfront Communities-The Island
132	10441	2480096	53485	Woburn
129	10500	2480065	50434	Willowdale East
105	10477	2480078	46496	Rouge
66	10461	2480158	43993	L'Amoreaux
58	10476	2480079	43965	Islington-City Centre West
73	10459	2480160	43794	Malvern
32	10465	2480156	36625	Dovercourt-Wallace Emerson-Junction
33	10368	2480137	35052	Downsview-Roding-CFB
95	10434	2480103	34805	Parkwoods-Donalda
77	10453	2480164	33964	Mimico (includes Humber Bay Shores)
124	10385	2480131	33312	West Humber-Clairville
81	10395	2480050	32954	Mount Olive-Silverstone-Jamestown
23	10468	2480083	31340	Church-Yonge Corridor
87	10438	2480099	31180	Niagara
3	10362	2480140	30526	Annex
45	10402	2480121	30491	Glenfield-Jane Heights
11	10447	2480092	29960	Bendale
83	10383	2480051	29658	Mount Pleasant West
0	10417	2480178	29113	Agincourt North

Figure 1: Neighborhoods with Top 20 Population

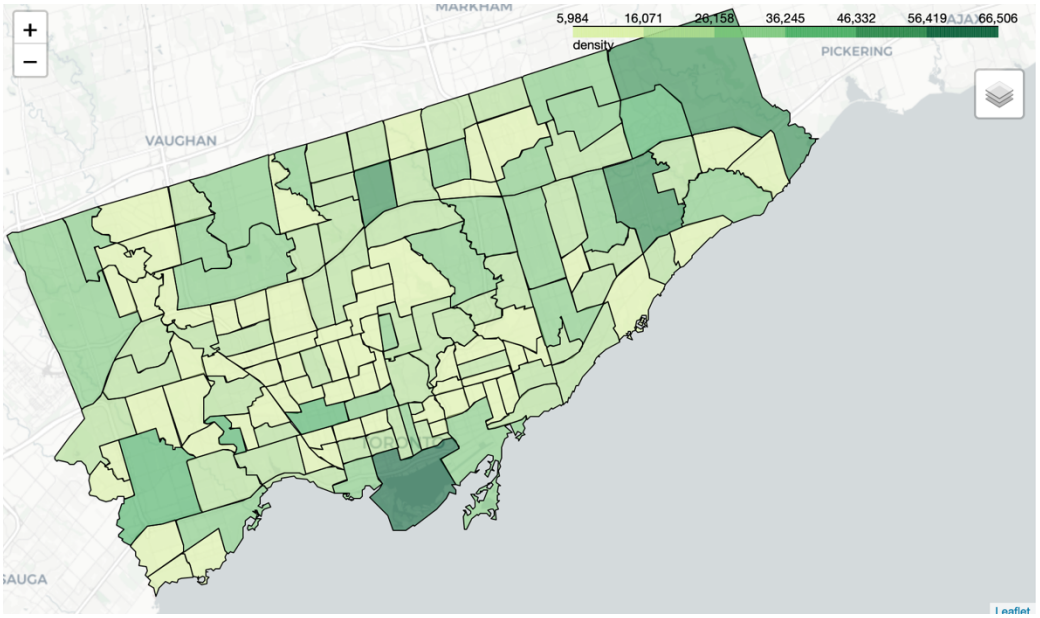

Figure 2: Choropleth Map with Population

However, as mentioned before, this data represents the population in 2016, which seems outdated in 2021. Therefore, I added a tendency analysis of population, which in a way indicates the trends of the population in the future and increases the stability of the whole project.

To study this point, I created a new data frame which scrapped from the original dataset. This data frame contains the population in 2016, the population in 2011, the change in percentage between these two years, and the names of neighborhoods. Figure 3 shows the line plot which compared the population in 2011 and 2016. It clearly shows that some areas increased their population dramatically. Then, I applied the same function to convert the string to integer and sorted the data by the change in percentage. The result is shown in Figure 4. The neighborhood, Waterfront Communities-The Island increases its population dramatically followed by Niagara and Henry Farm.

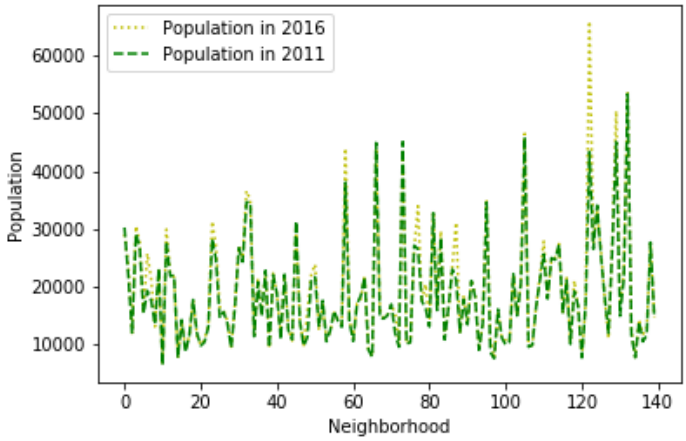

Figure 3: Line Plot Compared the Population in 2011 and 2016

	Population in 2016	Population in 2011	Change	Neighborhood
122	65913	43361	52.0	Waterfront Communities-The Island
87	31180	21274	46.6	Niagara
48	15723	11333	38.7	Henry Farm
6	25797	19348	33.3	Bay Street Corridor
71	15559	12050	29.1	Little Portugal
77	33964	26541	28.0	Mimico (includes Humber Bay Shores)
79	20506	16306	25.8	Moss Park
96	10722	8710	23.1	Pelmo Park-Humberlea
7	21396	17671	21.1	Bayview Village
58	43965	38084	15.4	Islington-City Centre West
130	16936	15004	12.9	Willowdale West
25	16472	14612	12.7	Clanton Park
129	50434	45041	12.0	Willowdale East
136	11817	10578	11.7	Yonge-Eglinton
23	31340	28349	10.6	Church-Yonge Corridor
67	16164	14642	10.4	Lansing-Westgate
50	23925	21740	10.1	High Park-Swansea
118	21108	19225	9.8	Thorncliffe Park
24	26984	24770	8.9	Clairlea-Birchmount
110	27876	25642	8.7	South Riverdale

Figure 4: Neighborhoods with Top 20 Change in Population

By adding the examination of the trends in population, this section would be more stable to derive the conclusion. It contains the current status and the possible situation in future.

2. Ethic Analysis:

Canada is well known for its diversity. People immigrated from different countries and have different ethic. Hence, analyzing the ethic in Toronto helps to determine the food type of the restaurant. Here I split the original data into a new data set, which contains the ethic and the total number of people with that ethic in Toronto. In this section, the analysis did not divide into each neighborhood. Considering people in Toronto can drive to different areas, the analysis would be more accurate to maintain at the level of the whole city. Simply converting the data from string to integer and sorting by the number of people, the result is shown in Figure 5. The first seven rows represent the continents. Then, starting with row eight represents each country. It shows that most people are of European origins and most people have Chinese ethic.

	Characteristic	City of Toronto
1365	European origins	1288850
1550	Asian origins	1079290
1366	British Isles origins	597295
1598	East and Southeast Asian origins	586515
1414	Southern European origins	441485
1582	South Asian origins	350040
1354	Other North American origins	345710
1601	Chinese	332830
1369	English	331895
1357	Canadian	323175
1397	Eastern European origins	302485
1370	Irish	262965
1586	East Indian	202675
1379	Western European origins (except French origins)	187190
1421	Italian	182500

Figure 5: Top 20 ethic in Toronto

3. Finance Analysis:

This section contains two parts and studies the economic considerations of Toronto. The first one is the cluster of markets. The cluster of markets in a way can represent the cluster of people. To open a restaurant in those areas where markets cluster can enhance the number of customers in the restaurant and create more cash flow. The second analysis studies the RGI, which is the benefits offered by the government to reduce the rent. This can decrease the cost to open a restaurant.

3.1 Cluster of markets:

To learn the cluster of the markets in Toronto, the first step is to separate the data. The original data is downloaded from the government website, which includes the number of markets in each area, RGI data, the neighborhood code, and geometry information. To run the clustering algorithm, I had to separate the latitude and longitude from the geometry information. Hence, I programmed a function to split the geometry message and convert the string to a float number. Then, I dropped the row where the market unit equals zero. After finishing the data set, I created the K-Mean clustering with the number of clusters of five and added the clusters label to the data frame. To visualize the result, I applied the data to the map using folium. The circles with different colors represent each cluster. As Figure 6 shown, the purple and orange circles are in higher density. By applying the function, it can be easily derived the center of each cluster. The larger and dark circles represent the centers of each cluster.

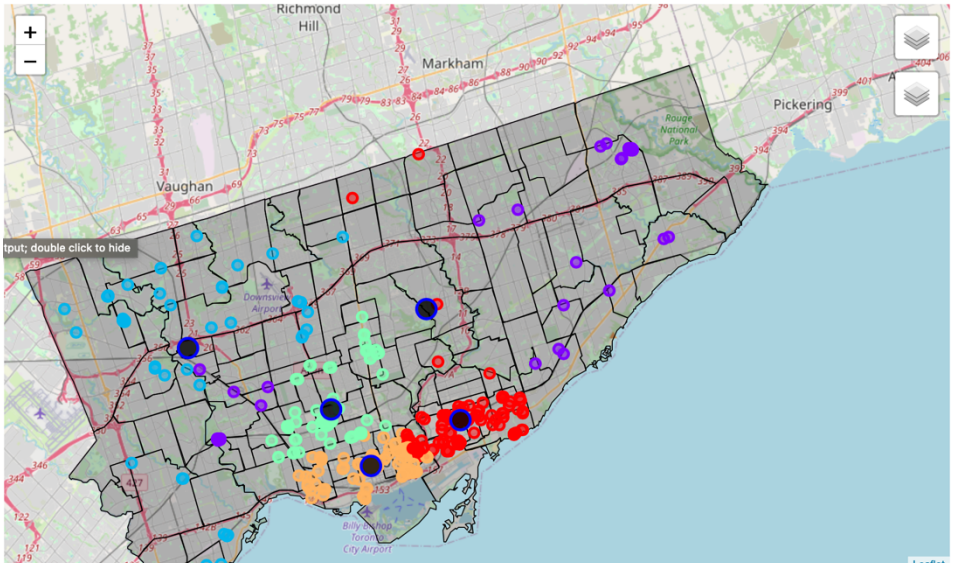

Figure 6: The clusters of Markets Visualized in Folium Map

3.2 Study of RGI:

This part learned the financial support from the government based on the study of RGI. The result can be simply derived by applying the sort method based on RGI. Figure 7 shows the tail of the data set in ascending order. It shows that the neighborhood with code 74 has the highest RGI, where this code represents the neighborhood, North St. James Town. To have a stricter view, I applied the data to the choropleth map, which is shown in Figure 8.

]:

	NGHBRHD_NUM	MRKT_UNIT	RGI_UNITS	geometry
893	75	9	331	{u'type': u'Point', u'coordinates': (-79.38228...
903	136	2	343	{u'type': u'Point', u'coordinates': (-79.18533...
922	14	2	348	{u'type': u'Point', u'coordinates': (-79.52777...
905	66	1	349	{u'type': u'Point', u'coordinates': (-79.33398...
889	122	0	350	{u'type': u'Point', u'coordinates': (-79.25475...
414	91	0	352	{u'type': u'Point', u'coordinates': (-79.45777...
902	136	3	372	{u'type': u'Point', u'coordinates': (-79.18875...
476	24	0	374	{u'type': u'Point', u'coordinates': (-79.50788...
901	113	13	378	{u'type': u'Point', u'coordinates': (-79.51606...
894	137	0	378	{u'type': u'Point', u'coordinates': (-79.23181...
662	85	0	384	{u'type': u'Point', u'coordinates': (-79.43360...
895	35	0	389	{u'type': u'Point', u'coordinates': (-79.44617...
930	48	1	396	{u'type': u'Point', u'coordinates': (-79.36147...
911	137	0	400	{u'type': u'Point', u'coordinates': (-79.22189...
1251	136	0	419	{u'type': u'Point', u'coordinates': (-79.19314...
480	124	0	421	{u'type': u'Point', u'coordinates': (-79.25156...
787	88	0	439	{u'type': u'Point', u'coordinates': (-79.46739...
703	95	10	450	{u'type': u'Point', u'coordinates': (-79.39824...
171	72	0	469	{u'type': u'Point', u'coordinates': (-79.35764...
719	74	8	711	{u'type': u'Point', u'coordinates': (-79.37361...

Figure 7: Neighborhood with Top 20 RGI in Ascending Order

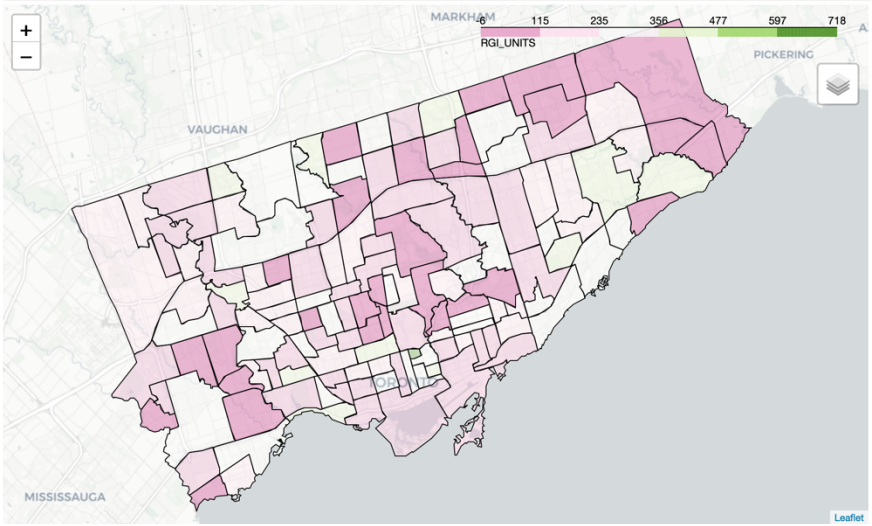

Figure 8: Visualized RGI Data in Choropleth Map

Result:

From the data and analysis did above, the decisions can be derived. The population density analysis would help to identify the best neighborhoods to open a restaurant. From the density analysis, Waterfront Communities-The Island, Woburn, and Willowdale East have the highest density. However, considering the trends of the population, Niagara and Mimico also have the potential possibility to attract a large number of people. Meanwhile, the population of Niagara and Mimico also ranked in the top 20. Hence, the final decision for the neighborhoods' choice would be Waterfront Communities-The Island (Neighborhood Number: 77), Niagara (Neighborhood Number: 82), and Mimico (Neighborhood Number: 17). To convince this decision, RGI analysis can be a support. Figure 9 represents the map of the neighborhood with their number. Compared with this map and Figure 8, these three neighborhoods have the financial support at the upper-middle level of these 140 neighborhoods. Another strong evidence to support this decision is the cluster examination. The markets are mostly clustered in these three areas. Therefore, overall speaking, these three neighborhoods can be considered as a great place to open a restaurant in Toronto.

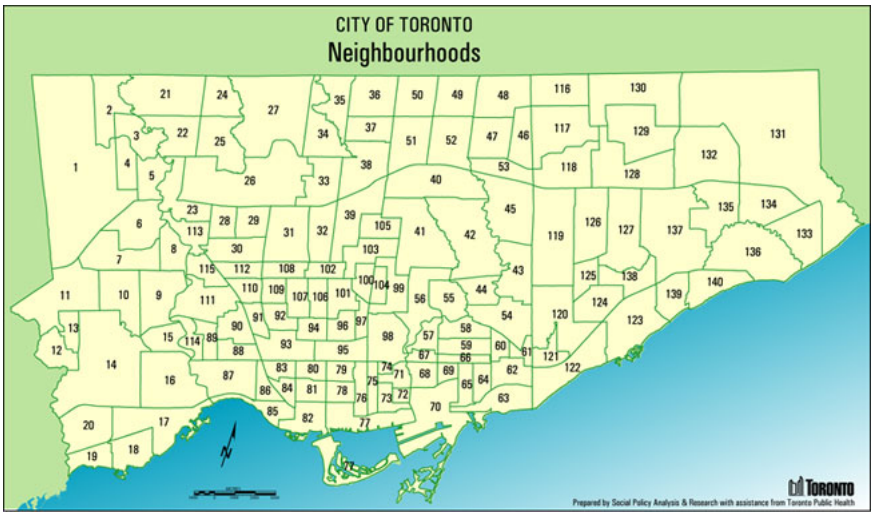

Figure 9: Toronto Neighborhood Map with Neighborhood Number

After deciding the location of the restaurant, another critical thing that should be determined is the type of the restaurant. The type of restaurant decides the targeted group of people; therefore, this decision would base on the ethical analysis. It clearly shows that the largest group of people in Toronto are European, Chinese, and Canadian. Hence, it would be a great idea to open a restaurant in Canadian. Moreover, considering the universities in Toronto,

which attracts many international students and requires food prepared in a short time for the lunch break, the restaurant can also open in Chinese and fast food.

Last but not least, the cluster study can provide some strategies. For example, running a restaurant at the beginning needs to enhance the recognition. Hence, to make advertisement in those centers of the clusters can be a good way, which decreases the time and cost but at the same time increases the effectiveness and efficiency. Besides, when choosing the detailed location for the restaurant, the distance between that and the centers can be evaluated to help find the best place.

Discussion:

Based on the analysis, the project finally derived three suggestions to open a restaurant in Toronto. However, there are also some limitation for this analysis. First is the time of the data. Most data in this report recorded by 2016, which seems outdated in 2021. It might be a huge difference from the past to nowadays. Hence, I included a trends analysis for the population, which can in a way decrease the impact of that. Second, opening a restaurant can be complex project. This report only provides the first step. There are still some other factors needs to be studied. For example, the salary in Toronto should be examined if planning to hire someone else. Besides, thousands of restaurants have already built their reputation locally. The competitiveness with other restaurants should also be evaluated. Moreover, lots of details should be considered such as the range of the price, the operation cost. This report only provides a brief analysis of the overall marketing in Toronto.

Conclusion:

This report concludes the project that provides some suggestions to open a restaurant in Toronto. By studying the neighborhood profiles including the population, the ethic, it finally derives the result that the restaurant can open in any of these three neighborhoods, Waterfront Communities-The Island, Niagara, and Mimico in the type of Chinese, Canadian or fast food. Besides, according to the clustering analysis, some strategics to increase the possibility of success can be taken such as the detailed location in the neighborhood and the location where to advertise. Even though this study has some limitations to the real situation, it still provides a stable and viable plan to move a step forward for opening a restaurant in Toronto.

Citation:

Toronto GDP: Statistics Canada. Table 36-10-0468-01 Gross domestic product (GDP) at basic prices, by census metropolitan area (CMA) (x 1,000,000)

Toronto Neighborhood Profile: Statistics Canada. 2016 Neighborhood Profiles