

Capstone Project – The Battle of the Neighborhoods (Week 2)

Topic: Finding an Ideal Location for the Opening of a Chinese Restaurant in Toronto

Background Information

Toronto is a vibrant multicultural city in Canada. According to information from Statistics Canada, Toronto had a recorded population of 2,731,571 in 2016. Toronto is thus the most populous city in Canada. This large city is home to many ethnic groups - more than 50% of the residents belong to a visible minority population group. A person who lives in Toronto, as this author has, will definitely be exposed to many cultural things - particularly cuisine. Since Toronto's population has exposure to various cultures, many people are willing to try cuisines that are not of their own. At the same time, many of Toronto's immigrants who have come to Toronto in their adulthood will have great affinity to their home cuisine. There is definitely great opportunity for someone to open a restaurant of ethnic cuisine. In this Capstone Project, the opportunity to open a Chinese restaurant will be explored.

Business Problem

Although there are many potential customers for a Chinese restaurant, there also comes great competition. There are already two Chinatowns in Toronto with many restaurants in these neighbourhoods. Adding onto that are many other areas, such as Scarborough, where many Chinese immigrants had immigrated to in the 1980s and 1990s, have many Chinese restaurants. A good location will be imperative to start a successful Chinese restaurant.

The business will have an advantage if the chosen location has:

- sufficient income to spend at restaurants
- sufficient population growth to sustain and grow clientele
- dining out is popular with the population
- few Chinese restaurants to compete with

Data

The first data that was required to complete the analysis was:

1. Postal codes of various neighborhoods
2. Geospatial data of Toronto neighborhoods

The postal codes can be found on Wikipedia (The postal codes can be found on Wikipedia (https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M)) and the geospatial data can be found from a previous lab. https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkillsNetwork-DS0701EN-SkillsNetwork/labs_v1/Geospatial_Coordinates.csv)

Income and population growth data and be found in the City of Toronto's Open Data Catalogue (<https://open.toronto.ca/>). The data is categorized by wards instead of neighborhoods. Thus, a separate dataset that has matching postal codes and wards was also downloaded from City of Toronto's Open Data Catalogue.

Data Cleaning

The organization of the data from the various sources are not consistent with each other. The data had to be gleaned and reorganized as Python dataframes in order to conduct the analysis. Postal codes from Wikipedia were gleaned with BeautifulSoup. Geospatial data was fortunately already prepared from a previous Capstone Course Lab. The open data on the City of Toronto's website was well presented, but it was necessary to reorganize it to connect the various tables together.

The postal codes and geospatial data were merged together as shown in a previous lab. The census data (which shows average income and population) in the open data from the City of Toronto was not categorized by postal codes and only wards. Therefore, a document that showed postal codes and corresponding wards was required and fortunately was available from another dataset in the open data. After data cleaning, a resulting table combining all the required data was generated (as shown below):

postal_code	ward_number	Postal Code	Borough	Neighborhood	Latitude	Longitude	population_change_pct	avg_income	
0	M4T	11	M4T	Central Toronto	Moore Park, Summerhill East	43.689574	-79.383160	0.0578571	98820
1	M6G	11	M6G	Downtown Toronto	Christie	43.669542	-79.422564	0.0578571	98820
2	M1L	20	M1L	Scarborough	Golden Mile, Clairlea, Oakridge	43.711112	-79.284577	0.0332615	38546
3	M6J	9	M6J	West Toronto	Little Portugal, Trinity	43.647927	-79.419750	0.0596913	40586
4	M2R	6	M2R	North York	Willowdale West	43.782736	-79.442259	0.0403391	41225
...
91	M4P	15	M4P	Central Toronto	Davisville North	43.712751	-79.390197	0.0269485	109887
92	M6L	5	M6L	North York	North Park, Maple Leaf Park, Upwood Park	43.713756	-79.490074	0.00064317	33742
93	M5G	10	M5G	Downtown Toronto	Central Bay Street	43.657952	-79.387383	0.400546	66906
94	M9N	5	M9N	York	Weston	43.706876	-79.518188	0.00064317	33742
95	M8X	3	M8X	Etobicoke	The Kingsway, Montgomery Road, Old Mill North	43.653654	-79.506944	0.118205	59192

Figure 1 - Dataframe of Combined Data

Data Analysis

From the income data chart (Figure 2) and population change (Figure 3) chart, it was possible to do a simple analysis to see which wards would be favourable for businesses. Areas with a positive change in population growth and relatively high income was considered as favourable circumstances, as it generally indicates that there is a growing number of potential customers and that there is available disposable income to spend on dining.

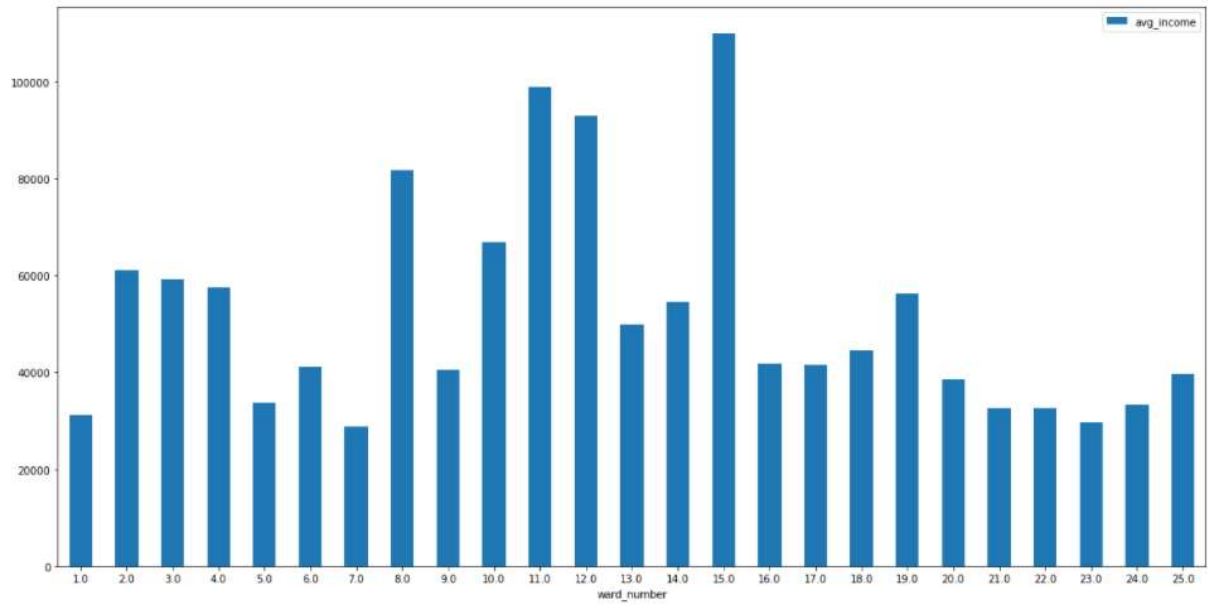


Figure 2 - Income Chart Based on Ward

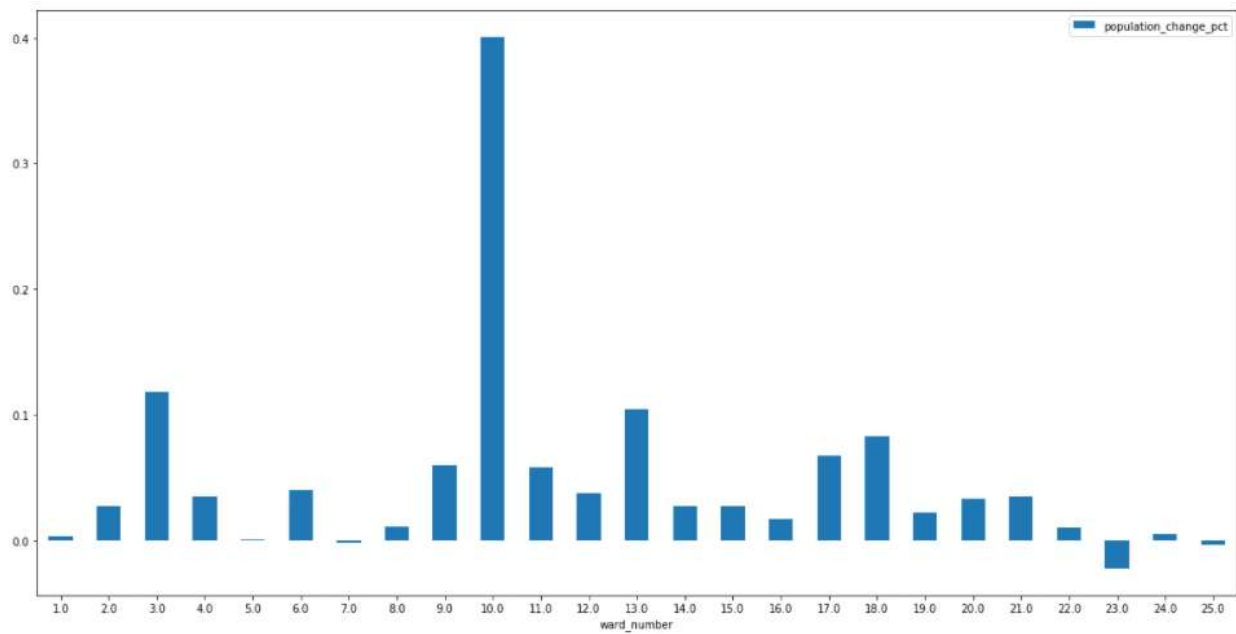


Figure 3 - Population Change By Ward

Foursquare Data

Foursquare data was extracted to determine what restaurants/eateries were in Toronto neighbourhoods. The objective here was to determine whether dining out is a popular activity in a neighborhood (based on the most popular type of restaurant type). In the Foursquare data, there are locations, such as grocery stores, cafés, pizza places, fast food places, etc., included in the restaurant data. Since the target restaurant does not belong to these types of places, if the more common restaurants fell in these categories, it was considered that the population will not likely dine in a Chinese restaurant.

Conclusions

Based on the clustering and generated dataframes, it was determined that Ward 15, namely Thorncliffe Park or Lawrence Park neighborhoods, would be the best location to start a location based on its high average income and types of restaurants. Out of the five wards chosen based on income and population change, it has an Indian Restaurant and Dim Sum Restaurant as most popular restaurant type, showing that people are willing to spend money at restaurants. Although Dim Sum is part of Chinese cuisine and dim sum restaurants may be competitors, it is possible to focus on another part of Chinese cuisine and become successful.

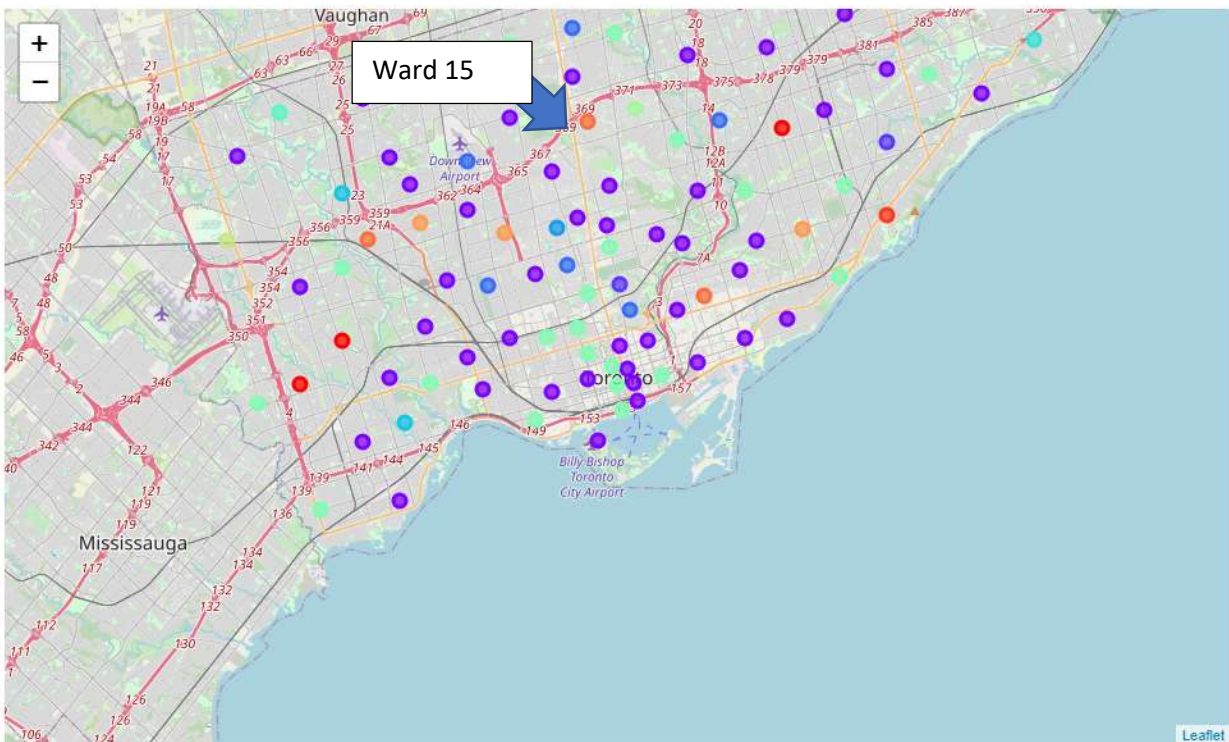


Figure 4 - Representation of Clustered Data

Cluster Ward 15 (Chosen Based on Income)												
cluster_ward_15=toronto_w_clusters_res.loc[toronto_w_clusters_res['ward_number'] == 15] cluster_ward_15												
	postal_code	ward_number	Postal Code	Borough	Neighborhood	Latitude	Longitude	population_change_pct	avg_income	Cluster Labels	1st Most Common Restaurant Type	2nd Most Common Restaurant Type
14	M2L	15	M2L	North York	York Mills, Silver Hills	43.757490	-79.374714	0.0269485	109887	9	Martial Arts School	Yoga Studio
17	M4G	15	M4G	East York	Leaside	43.709060	-79.363452	0.0269485	109887	1	Coffee Shop	Sporting Goods Shop
78	M4H	15	M4H	East York	Thorncliffe Park	43.705369	-79.349372	0.0269485	109887	1	Indian Restaurant	Yoga Studio
83	M4N	15	M4N	Central Toronto	Lawrence Park	43.728020	-79.388790	0.0269485	109887	1	Dim Sum Restaurant	Pizzeria
88	M2P	15	M2P	North York	York Mills West	43.752758	-79.400049	0.0269485	109887	13	Park	Convenience Store
91	M4P	15	M4P	Central Toronto	Davisville North	43.712751	-79.390197	0.0269485	109887	1	Pizza Place	Breakfast Spot

Figure 5 - Dataframe of Ward 15

Assumptions

There have been assumptions while coming to the above conclusion. It is assumed that the income and population change within a ward is spread homogeneously throughout neighborhoods. In reality, it is likely there is some disparity between neighborhoods. It was also assumed that population change continues to progress as in the past, however, this is not guaranteed.