Finding the best location to establish a new Chinese restaurant

Coursera Capstone Project – Week 5

By: Si Tan Zhang

1.0 Introduction / Business Problem

1.1 Background:

Toronto is Canada's largest city by population (4th largest city by population in North America) and is recognized as one of the most multicultural and cosmopolitan cities in the world. The diverse population of Toronto reflects its current and historical role as an important destination for immigrants to Canada. 50%+ of residents belong to a visible minority population group.

The diversity of the cuisine available is reflective of Toronto's multiculturalism and is key to maintaining Toronto's reputation as one of the most immigration friendly and ethnically diverse cities in Canada. Toronto can be divided into many ethnic neighborhoods mainly serving different kinds of ethnic foods where residents and tourists alike know where to get the best selection of ethnic foods to satisfy all of their food cravings.

Opening a new restaurant (i.e, Chinese restaurant) is often an immigrant/resident's dream and could become their main/only source of income. As with any business decision, opening a new restaurant requires serious consideration of many complex factors. Specifically, it is critical for the owner to accurately predict the best location to establish the restaurant to ensure the long-term success of the business. One of the key considerations in deciding on a location could be looking at where the other Chinese restaurants are located in the city.

1.2 Problem / Objective:

The objective of this project is to analyze and pick the best locations to open a new Chinese restaurant in Toronto. Factors critical in determining the best location to open a Chinese restaurant could include:

- Boroughs/neighborhoods with the highest/lowest number of similar restaurants
- Boroughs/neighborhoods with the highest rated and lowest rated Chinese restaurants
- Average cost of Chinese dishes by borough/neighborhood

Applying data science methodologies on Foursquare location data, the project aims to provide solutions to the following question: Where should a new business owner look to open a Chinese restaurant in the city of Toronto?

1.3 Target Audience / Interested Parties:

The main stakeholder/target audience would be the business owner who is looking to make his/her main/only source of income by establishing a new Chinese restaurant at an optimal location. Other interested parties include residents and new immigrants who want to try out a new Chinese restaurant/make this their regular dining out spot, tourists who come from all over the world to dine at the best Toronto restaurants, essentially everyone who likes Chinese food/is interested in trying out Chinese food.

2.0 Data Description / Data Sources:

Data required for this project includes:

- List of boroughs/neighborhoods in Toronto, as the location of the project is confined to the city of Toronto in the country of Canada (Wikipedia)

- Latitude and longitude coordinates of these boroughs. This is required to plot the map as well as retrive the venue data (Geocoder package on Python)
- Venue data, specifically data related to Chinese restaurants. This data will be used to get a list of Chinese restaurants in different neighborhoods (Foursquare)

Sources of data and methods for extraction:

- A list of boroughs/neighbourhoods in Toronto sorted by postcode can be found on Wikipedia (https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M). To extract the data, Python has built-in functions to turn the data into table form.
- Then the Python Geocoder package will be used to get the geographical coordinates (latitude and longitude coordinates) of the neighborhoods.
- The Foursquare API will then be used to get the venue data for the neighborhoods. The Foursquare API can provide many categories of venue data but my area of interest is in the Chinese restaurants category.

3.0 Methodology:

Step 1: import the neighborhood data into a data frame

- In this step, I web scrapped Wikipedia into a data frame containing postcode, borough and neighborhood info for the city of Toronto which is one of my main data sources. The head of the data is as follows:

	Postcode	Borough	Neighborhood
2	M3A	North York	Parkwoods
3	M4A	North York	Victoria Village
4	M5A	Downtown Toronto	Harbourfront
5	M6A	North York	Lawrence Heights
6	M6A	North York	Lawrence Manor

Step 2: clean the data

- I then cleaned the data so that each unique postal code corresponds to a borough and a list of neighborhoods. Postcodes without an assigned borough are also deleted. Head of the cleaned data frame is as follows:

	Postcode	Borough	Neighborhood
0	M1B	Scarborough	Rouge, Malvern
1	M1C	Scarborough	Highland Creek, Rouge Hill, Port Union
2	M1E	Scarborough	Guildwood, Morningside, West Hill
3	M1G	Scarborough	Woburn
4	M1H	Scarborough	Cedarbrae

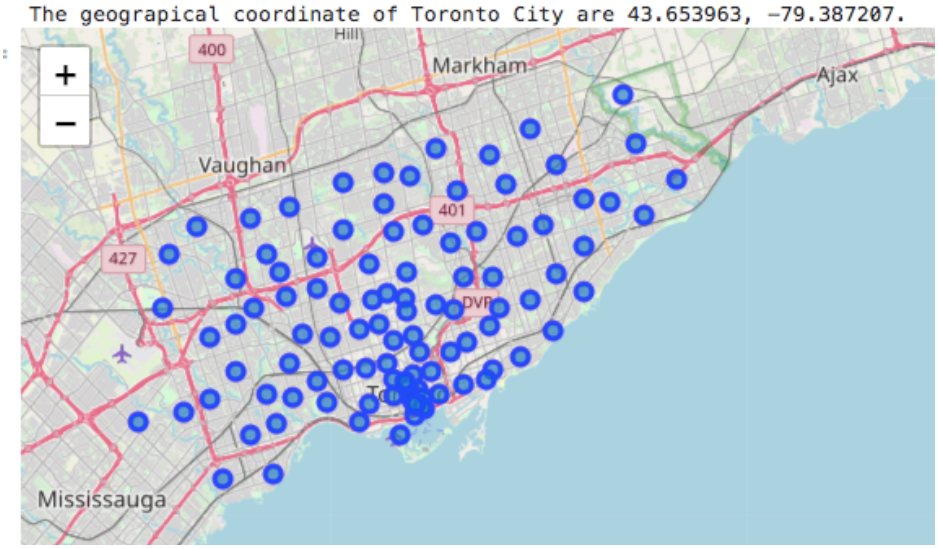
Step 3: add/append geographic coordinates

- In this step, I used the geospatial data containing latitude and longitude data provided by Coursera in order to add/append coordinates onto my earlier database so that it now contain the coordinates and can be plotted. Head of the appended data frame is as follows:

	Postcode	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

Step 4: explore and cluster neighborhoods then superimpose the Toronto map on top

- I then used python Folium to visualize geographic details of Toronto and its boroughs and created a map of Toronto with boroughs and neighborhoods superimposed on tip. I used latitude and longitude values from geocoder to get the visuals as below:



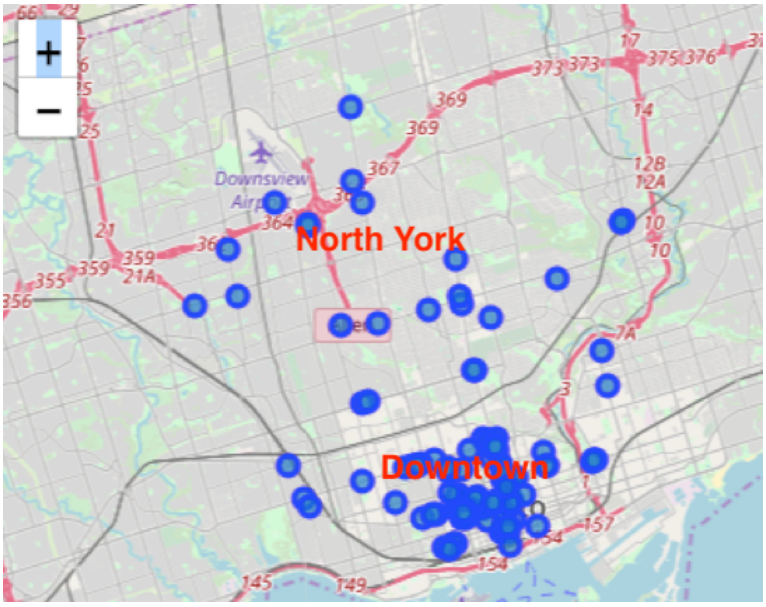
Step 5: using Foursquare API to screen out all Chinese restaurants in Toronto

- I then used the Foursquare API to find Chinese restaurants in each borough. I designed the limit as 1,000 venues and the radius as 6,000 meters from center of Toronto as determined by Foursquare based on latitude and longitude information. The category ID corresponding to Chinese restaurants is also input to limit the search to Chinese restaurants only.
- Below is a head list containing venue name, category (in this case Chinese restaurants only), latitude and longitude information from Foursquare API noting there are 157 Chinese restaurant venues in total returned by Foursqaure:

	name	categories	lat	lng
0	Swatow Restaurant 汕頭小食家	Chinese Restaurant	43.653866	-79.398334
1	C'est Bon	Chinese Restaurant	43.716785	-79.400406
2	GB Hand-Pulled Noodles	Chinese Restaurant	43.656434	-79.383783
3	Cho Cho Cho	Chinese Restaurant	43.680139	-79.432895
4	Crown Princess Fine Dining 伯爵名宴	Chinese Restaurant	43.666455	-79.387698
5	Asian Legend 味香村	Chinese Restaurant	43.653603	-79.395047
6	Lee	Chinese Restaurant	43.644375	-79.400425
7	Sea-Hi	Chinese Restaurant	43.731630	-79.433014
8	Yueh Tung Chinese Restaurant	Chinese Restaurant	43.655281	-79.385337
9	House of Gourmet 滿庭芳	Chinese Restaurant	43.653273	-79.397230

4.0 Results:

I first mapped the Chinese restaurants (see below) and noticed they are fairly scattered, but seem to be more concentrated in Downtown and North York Toronto (labeled in red).

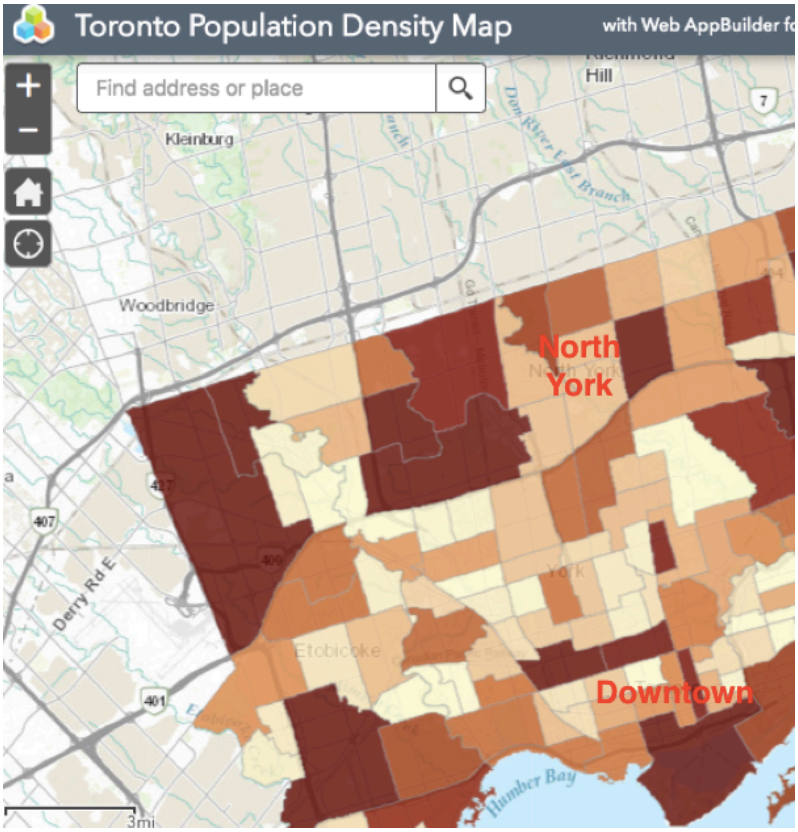


Then, by mapping the coordinates of the restaurants to the closest borough and displaying it in a data frame per below, we can further confirm that downtown Toronto has the highest number of Chinese restaurants, followed by North York.

borough	name	categories
Central Toronto	9	9
Downtown Toronto	119	119
East Toronto	5	5
East York	2	2
North York	13	13
West Toronto	6	6
York	3	3

5.0 Discussion:

Toronto is a big city with pockets of high population density, with Downtown and North York both having pockets of high density per the below density map.



I first visualized the Foursquare Chinese restaurant data and was able to see there are high concentrations of Chinese restaurants in Downtown and North York Toronto.

Then, I clustered the restaurants into their corresponding boroughs. In order to do this I applied a mathematical algorithm to assign restaurants to the boroughs with a central coordinate closest to the

restaurant as measured by a straight line. After that I put this data into a data frame which further confirmed that Downtown Toronto has the highest number of Chinese restaurants by far (119) vs North York which has the second highest number of Chinese restaurants (13) and other neighbourhoods. Chinatown is also located in Downtown Toronto. So it is safe to say that when people want Chinese food, they know they can best find them in Downtown and North York.

Based on these results, I would recommend opening a Chinese restaurant in a location with the highest population density and number of Chinese restaurants, which would be Downtown Toronto. My rationale includes the following:

- Areas with high population density will have more mouths to feed. As a result, there will be higher demand for restaurants in those areas where residents are looking to eat something convenient and delicious without having to travel far.
- Toronto has one Chinatown located Downtown, it is the most popular place residents and tourists go when they are craving Chinese food since it has the greatest and best selection. A Chinese restaurant set up in this area would get a lot of foot traffic from residents and tourists alike who are just passing buy and want to try out a new place.

6.0 Conclusion:

To conclude, Downtown Toronto, specifically Chinatown, is known for its high population density and for being a tourist attraction for Toronto and foreign residents alike who want to try out different types of Chinese cuisine. As a result, this neighborhood also has the highest number of Chinese restaurants.

I recommend opening a new Chinese restaurant in this location since it is the go to neighborhoods for Chinese cuisine. When the restaurant first opens it will gain a lot of foot traffic and attention from passing residents and tourists alike, as long as food quality, reasonable prices (covering costs) and great service can be ensured, the good word will get out and the restaurant can expect to prosper in the long term!