Prospects of opening Chinese Restaurant in Toronto, Canada.

Background:

Toronto, capital of the province of Ontario, is a major Canadian city along Lake Ontario's north-western shore. It's a dynamic metropolis with a core of soaring skyscrapers, all dwarfed by the iconic, free-standing CN Tower. Toronto also has many green spaces, from the orderly oval of Queen's Park to 400-acre High Park and its trails, sports facilities and zoo.

In the 21st Century, Toronto has integrated the core and the suburbs under one government, although many bylaws enacted by the former municipalities remain in effect. A division has persisted between the interests of those who live in the former suburbs and those of the central core. The central core has seen unprecedented office growth and residential growth, particularly of condominium apartments, while the former suburbs and further outlying suburbs have seen the bulk of new industrial investment. A major metropolis of just over 2.8 million people, Toronto is also one of the most ethnically diverse in the world. All of this growth took place on the lands of the original Toronto Purchase, of which final agreement was only finally reached between the Mississauga and the Government of Canada in 2010. With many office and residential growth in the city, there would be a huge demand of food, a good business chance to invest a restaurant in the city. The target audience would be the residents living in the area/working in the area and the tourists visiting Toronto. We will be looking at the prospects of opening a Chinese restaurant in Toronto.

Who will be more interested in this project? What type of clients or a group of people would be benefitted?

- Business personnel who want to invest or open a Chinese restaurant in Toronto. This analysis will give an idea, how beneficial it is to open a restaurant and what are the pros and cons of this business.
- Chinese tourists who are visiting Toronto but still would like a taste of their homeland dishes
- Business Analyst or Data Scientists, who wish to analyse the neighbourhoods of Toronto using Exploratory Data Analysis and other statistical & machine learning techniques to obtain all the necessary data, perform some operations on it and, finally be able to tell a story out of it.

Data acquisition and cleaning

Data Sources

a) "List of Postal code of Canada: M" (https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M) Wikipedia page is used to get all the information about the neighbourhoods present in Toronto. This page has the postal code, borough & the name of all the neighbourhoods present in Toronto.

b) "https://cocl.us/Geospatial_data" csv file to get all the geographical coordinates of the neighbourhoods for Toronto

c) "Demographics of Toronto" (https://en.m.wikipedia.org/wiki/Demographics_of_Toronto#Ethnic_diversity) Wikipedia page to get the diversity culture in Toronto. Using this page I'm going to identify the neighbourhoods which are densely populated with Chinese as it might be helpful in identifying the suitable neighbourhood to open a new Chinese restaurant.

d) To get location and other information about various venues in Toronto I'm using Foursquare's explore API. Using the Foursquare's explore API (which gives venues recommendations), I'm fetching details about the venues up present in Toronto and collected their names, categories and locations (latitude and longitude). From Foursquare API (https://developer.foursquare.com/docs), I retrieved the following for each venue:

Name: The name of the venue.

Category: The category type as defined by the API.

Latitude: The latitude value of the venue.

Longitude: The longitude value of the venue.

Data Cleaning

a) Scraped the following Wikipedia page, “List of Postal code of Canada: M” in order to obtain the data about the Toronto & the Neighbourhoods in it.

b) Scraped the Demographics of Toronto Wikipedia page to obtain the data about Toronto and ethnicity/diversity culture

Dataframe will consist of three columns: Postal Code, Borough, and Neighbourhood

Only the cells that have an assigned borough will be processed. Borough that is not assigned are ignored.

More than one neighbourhood can exist in one postal code area. For example, in the table on the Wikipedia page, you will notice that M5A is listed twice and has two neighbourhoods: Harbour front and Regent Park. These two rows will be combined into one row with the neighbourhoods separated with a comma as shown in row 11 in the above table.

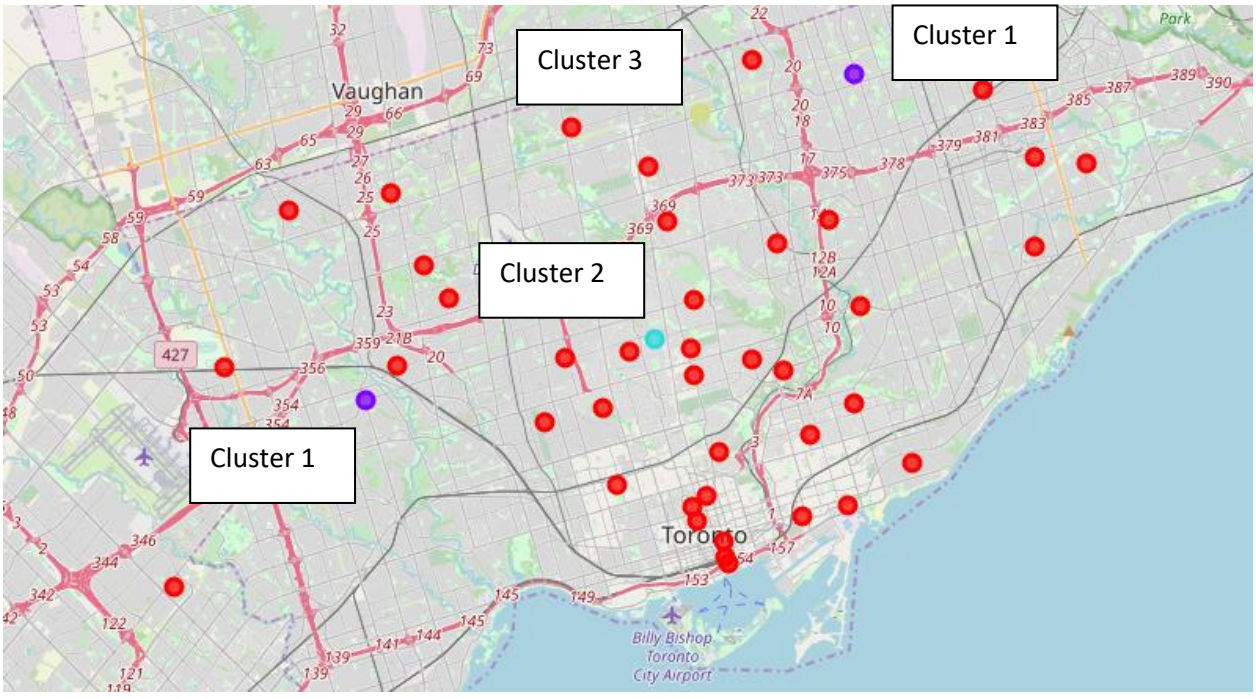
If a cell has a borough but a Not assigned neighbourhood, then the neighbourhood will be the same as the borough.

Methodology

1. Download the postal code of Canada data from the Wikipedia page, clean and format the data frame for merging
2. Download the latitude and longitude information for the Toronto and merge them onto the postal code dataframe
3. Download the diversity tables from the demographics of toronto Wikipedia page, clean and format the dataframe for merging
4. Use foursquare API to collect venue information in toronto, reformat and clean the data to output it as a data frame, filter for only venues that are restaurants in toronto
5. Merge the data frame with restaurants to the postal code data frame by their neighborhoods
6. Collect data from Wikipedia pages on the neighbourhoods in each riding and merge them to the demographics data frame , calculate the Chinese population in each riding and merge the data frame to the postal code data frame that contains the restaurant data.

Results

<u>Clustering of neighbourhoods with Chinese Restaurants</u>



Please see Appendix for the table of neighbourhoods by their cluster labels and average number of Chinese restaurants in the borough

Using the data from web resources like Wikipedia, geospatial coordinates of Toronto neighbourhoods, and Foursquare API, to set up a very realistic data-analysis scenario. We have found out that — In all the ridings, Scarborough-Agincourt, Scarborough North, Don Valley North , Willowdale, Spadina-Fort York,Toronto-Danforth ,University-Rosedale, Toronto Centre , Don Valley West , Don Valley East, Scarborough Centre ,Scarborough Southwest,Scarborough-Guildwood are the densely populated with Chinese crowd ridings.

With the help of clusters examing looks like Don Valley north, Toronto Centre, Scarborough centre and Scarborough southwest are already densely populated with Chinese restaurants. So it is better idea to leave those boroughs out and consider only Don Valley north, Toronto Centre, Scarborough centre and Scarborough southwest for the new restaurant's location.

After careful consideration it is a good idea to open a new Chinese restaurant in Scarborough borough since it has high number of Chinese population which gives a higher number of customers possibility and lower competition since very less Chinese restaurants in the neighbourhoods.

Discussion

According to this analysis, Scarborough borough will provide least competition for the new upcoming Chinese restaurant as there are very little Chinese restaurants spread or no Chinese restaurants in neighbourhoods.

Also looking at the population distribution looks like it is densely populated with Chinese crowd which helps the new restaurant by providing high customer visit possibility. So, definitely this region could potentially be a perfect place for starting quality Chinese restaurants.

Some of the drawbacks of this analysis are — the clustering is completely based only on data obtained from Foursquare API. Also the Chinese population distribution in each neighbourhood is also based on the 2016 census which is not up-to date. Thus population distribution would have definitely changed by 2020 given 4 years gap in the data.

Since population distribution of Chinese crowd in each neighbourhood & number of Chinese restaurants are the major feature in this analysis and it is not fully up-to date data, this analysis is definitely not far from being conclusory & it has lot of areas where it can be improved. We also noticed that there were no indication or restaurants in the neighbourhoods that contain the 'Chinatown' area, even though google maps show that there are a bunch of restaurants there. This could be indicating that the restaurants there are labelled as 'Asian' or some form of Asian culture diversity and not specifically 'Chinese', which causes our data to be looking at a different area for opening up a business. Perhaps expanding to look at all Asian restaurants will give a more accurate representation to select where the new business should start up.

However, it certainly provides us with some good insights, preliminary information on possibilities & a head start into this business problem by setting the step stones properly. Furthermore, this may also potentially vary depending on the type of clustering techniques that we use to examine the data.

Conclusion

Finally to conclude this project, we have got a chance to on a business problem like how a real like data scientists would do. Using python libraries to fetch, clean and format the data, to manipulate the contents & to analyse and visualize those datasets. We have made use of Foursquare API to explore the venues in neighbourhoods of Toronto; to imported data from Wikipedia visualized using various plots present in seaborn & matplotlib for graphs.

We also applied machine learning technique to predict the output given the data and used Folium to visualize it on a map. Also, some of the drawbacks or areas of improvements shows us that this analysis can further be improved with help more data and different machine learning technique. Similarly we can use this project to analysis any scenario such opening a different cuisine or success of opening a new gym and etc. Hopefully, this project helps acts as initial guidance to take more complex real-life challenges using data-science.

Appendix A: Tables

Table 1: Neighbourhoods by their cluster labels and Chinese Restaurant value

Borough	Neighbourhood	Cluster Labels	Chinese Restaurant
North York	Bayview Village	3	0.25
Central Toronto	North Toronto West	2	0.058824
Scarborough	L'Amoreaux West	1	0.142857
Etobicoke	Westmount	1	0.125
Scarborough	Woburn	0	0
Scarborough	Cedarbrae	0	0
Scarborough	Scarborough Village	0	0
Scarborough	Agincourt	0	0
North York	Hillcrest Village	0	0
North York	Willowdale South	0	0
North York	York Mills West	0	0
North York	Willowdale West	0	0
North York	Parkwoods	0	0
North York	Don Mills North	0	0
North York	Downsview West	0	0
North York	Downsview Central	0	0
North York	Downsview Northwest	0	0
North York	Victoria Village	0	0
East York	Woodbine Heights	0	0
East Toronto	The Beaches	0	0
East York	Leaside	0	0
East York	Thorncliffe Park	0	0
East York	East Toronto	0	0
East Toronto	Studio District	0	0
Central Toronto	Lawrence Park	0	0
Central Toronto	Davisville North	0	0
Central Toronto	Davisville	0	0
Downtown Toronto	Rosedale	0	0
Downtown Toronto	Church and Wellesley	0	0
Downtown Toronto	St. James Town	0	0
Downtown Toronto	Berczy Park	0	0
Downtown Toronto	Central Bay Street	0	0.012821
Central Toronto	Roselawn	0	0
Downtown Toronto	Stn A PO Boxes 25 The Esplanade	0	0
North York	Glencairn	0	0
York	Humewood-Cedarvale	0	0
York	Caledonia-Fairbanks	0	0
Downtown Toronto	Christie	0	0
Queen's Park	Not assigned	0	0
Mississauga	Canada Post Gateway Processing Centre	0	0

East Toronto	Business Reply Mail Processing Centre 969 Eastern	0	0
North York	Humber Summit	0	0
York	Weston	0	0
Etobicoke	Northwest	0	0