Coursera IBM Data Science Capstone Project :

Opening a new
Bermese Restaurant in Toronto

Report Prepared by :

Abhishek Baidya

Introduction

For this Capstone project, I am creating a hypothetical scenario for a concept Burmese restaurateur who wants to explore opening an authentic Burmese restaurant in Toronto area.

The idea behind this project is that there may not be enough Burmese restaurants in Toronto and it might present a great opportunity for this entrepreneur who is based in Canada. As Burmese food is very similar to other Asian cuisines, this entrepreneur is thinking of opening this restaurant in locations where Asian food is popular (aka many Asian restaurants in the neighborhood).

Business Problem

The objective of this capstone project is to find the most suitable location for the entrepreneur to open a new Burmese restaurant in Toronto, Canada. By using data science methods and machine learning methods such as clustering, this project aims to provide solutions to answer the business question:

In Toronto, if an entrepreneur wants to open a Burmese restaurant,where should they consider opening it?

Target Audience:

The entrepreneur who wants to find the location to open authentic Burmese restaurant

Data

To solve this problem, I will need below data:

- List of neighborhoods in Toronto, Canada.

- Latitude and Longitude of these neighborhoods for Location.

- Venue data related to Asian restaurants. This will help us find the neighborhoods that are most suitable to open a Burmese restaurant.

Extracting Data

Scrapping of Toronto neighborhoods via Wikipedia

Getting Latitude and Longitude data of these neighborhoods via Geocoder package

Using Foursquare API to get venue data related to these neighborhoods

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Adelaide, King, Richmond	100	100	100	100	100	100
Berczy Park	56	56	56	56	56	56
Brockton, Exhibition Place, Parkdale Village	23	23	23	23	23	23
Business Reply Mail Processing Centre 969 Eastern	16	16	16	16	16	16
CN Tower, Bathurst Quay, Island airport, Harbourfront West, King and Spadina, Railway Lands, South Niagara	16	16	16	16	16	16
Cabbagetown, St. James Town	47	47	47	47	47	47
Central Bay Street	83	83	83	83	83	83
Chinatown, Grange Park, Kensington Market	87	87	87	87	87	87
Christie	19	19	19	19	19	19
Church and Wellesley	82	82	82	82	82	82
Commerce Court, Victoria Hotel	100	100	100	100	100	100
Davisville	32	32	32	32	32	32
Davisville North	8	8	8	8	8	8
Deer Park, Forest Hill SE, Rathnelly, South Hill, Summerhill West	15	15	15	15	15	15
Design Exchange, Toronto Dominion Centre	100	100	100	100	100	100
Dovercourt Village, Dufferin	16	16	16	16	16	16
First Canadian Place, Underground city	100	100	100	100	100	100
Forest Hill North, Forest Hill West	4	4	4	4	4	4
Harbord, University of Toronto	37	37	37	37	37	37
Harbourfront	47	47	47	47	47	47
Harbourfront East, Toronto Islands, Union Station	100	100	100	100	100	100
High Park, The Junction South	24	24	24	24	24	24
Lawrence Park	3	3	3	3	3	3

Methodology

- First, I need to get the list of neighborhoods in Toronto, Canada. This is possible by extracting the list of neighborhoods from Wikipedia page (" https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M ") I did the web scraping by utilizing pandas html table scraping method as it is easier and more

- I will need to get their coordinates to utilize Foursquare to pull the list of venues near these neighborhoods. After gathering all these coordinates, I visualized the map of Toronto using Folium package to verify whether these are correct coordinates. Next, I use Foursquare API to pull the list of top 100 venues within 500 meters radius. From Foursquare, I am able to pull the names, categories, latitude and longitude of the venues. With this data, I can also check how many unique categories that I can get from these venues. Then, I analyze each neighborhood by grouping the rows by neighborhood and taking the mean on the frequency of occurrence of each venue category.

- Previously, when I ran the model, I was looking for "Asian restaurants" but there are very few results (maybe due to Foursquare categorization) so I looked for the restaurants closest to Burmese cuisine taste (side note: Burmese food and Thai food are very similar in taste, so my justification is that if there are people who enjoyed Thai food, they likely are going to enjoy Burmese food too)

- Lastly, I performed the clustering method by using k-means clustering. K-means clustering algorithm identifies k number of centeriods, and then allocates every data point to the nearest cluster, while keeping the centroids as small as possible. It is one of the simplest and popular unsupervised machine learning algorithms and it is highly suited for this project as well. I have clustered the neighborhoods in Toronto into 3 clusters based on their frequency of occurrence for "Thai food". Based on the results (the concentration of clusters), I will be able to recommend the ideal location to open the restaurant.

Results

	Neighborhoods	Afghan Restaurant	Airport	Airport Food Court	Airport Gate	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	Antique Shop	...	Theme Restaurant	Toy / Game Store	Trail	Train Station	Vegetarian / Vegan Restaurant	Video Game Store	Vietnamese Restaurant	Wine Bar	Women's Store	Yoga Studio
0	The Beaches	0	0	0	0	0	0	0	0	0	...	0	0	1	0	0	0	0	0	0	0
1	The Beaches	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
2	The Beaches	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
3	The Beaches	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
4	The Beaches	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0

|	Neighborhoods	Afghan Restaurant	Airport	Airport Food Court	Airport Gate	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	Antique Shop	...	Theme Restaurant	Toy / Game Store	Trail	Train Station	Vegetarian / Vegan Restaurant	Video Game Store	Vietnamese Restaurant	Wine Bar	Women's Store	Yoga Studio
0	Adelaide, King, Richmond	0.000000	0.0000	0.0000	0.0000	0.000	0.000	0.000	0.020000	0.000000	...	0.000000	0.00000	0.00000	0.00	0.020000	0.000000	0.000000	0.010000	0.01	0.000000
1	Berczy Park	0.000000	0.0000	0.0000	0.0000	0.000	0.000	0.000	0.000000	0.000000	...	0.000000	0.00000	0.00000	0.00	0.017857	0.000000	0.000000	0.000000	0.00	0.000000
2	Brockton, Exhibition Place, Parkdale Village	0.000000	0.0000	0.0000	0.0000	0.000	0.000	0.000	0.000000	0.000000	...	0.000000	0.00000	0.00000	0.00	0.000000	0.000000	0.000000	0.000000	0.00	0.043478
3	Business Reply Mail Processing Centre 969 Eastern	0.000000	0.0000	0.0000	0.0000	0.000	0.000	0.000	0.000000	0.000000	...	0.000000	0.00000	0.00000	0.00	0.000000	0.000000	0.000000	0.000000	0.00	0.000000
4	CN Tower, Bathurst Quay, Island airport, Harbo...	0.000000	0.0625	0.0625	0.0625	0.125	0.125	0.125	0.000000	0.000000	...	0.000000	0.00000	0.00000	0.00	0.000000	0.000000	0.000000	0.000000	0.00	0.000000
5	Cabbagetown, St. James Town	0.000000	0.0000	0.0000	0.0000	0.000	0.000	0.000	0.021277	0.000000	...	0.000000	0.00000	0.00000	0.00	0.000000	0.000000	0.000000	0.000000	0.00	0.000000
6	Central Bay Street	0.000000	0.0000	0.0000	0.0000	0.000	0.000	0.000	0.012048	0.000000	...	0.000000	0.00000	0.00000	0.00	0.012048	0.000000	0.000000	0.012048	0.00	0.012048
7	Chinatown, Grange Park, Kensington Market	0.000000	0.0000	0.0000	0.0000	0.000	0.000	0.000	0.000000	0.000000	...	0.000000	0.00000	0.00000	0.00	0.045977	0.000000	0.068966	0.011494	0.00	0.000000
8	Christie	0.000000	0.0000	0.0000	0.0000	0.000	0.000	0.000	0.000000	0.000000	...	0.000000	0.00000	0.00000	0.00	0.000000	0.000000	0.000000	0.000000	0.00	0.000000
9	Church and Wellesley	0.012195	0.0000	0.0000	0.0000	0.000	0.000	0.000	0.012195	0.000000	...	0.012195	0.00000	0.00000	0.00	0.000000	0.000000	0.000000	0.000000	0.00	0.012195
10	Commerce Court, Victoria Hotel	0.000000	0.0000	0.0000	0.0000	0.000	0.000	0.000	0.030000	0.000000	...	0.000000	0.00000	0.00000	0.00	0.020000	0.000000	0.000000	0.010000	0.00	0.000000
11	Davisville	0.000000	0.0000	0.0000	0.0000	0.000	0.000	0.000	0.000000	0.000000	...	0.000000	0.03125	0.00000	0.00	0.000000	0.000000	0.000000	0.000000	0.00	0.000000
12	Davisville North	0.000000	0.0000	0.0000	0.0000	0.000	0.000	0.000	0.000000	0.000000	...	0.000000	0.00000	0.00000	0.00	0.000000	0.000000	0.000000	0.000000	0.00	0.000000
13	Deer Park, Forest Hill SE, Rathnelly, South Hi...	0.000000	0.0000	0.0000	0.0000	0.000	0.000	0.000	0.066667	0.000000	...	0.000000	0.00000	0.00000	0.00	0.000000	0.000000	0.066667	0.000000	0.00	0.000000
14	Design Exchange, Toronto Dominion Centre	0.000000	0.0000	0.0000	0.0000	0.000	0.000	0.000	0.020000	0.000000	...	0.000000	0.00000	0.00000	0.01	0.010000	0.000000	0.000000	0.010000	0.00	0.000000
15	Dovercourt Village, Dufferin	0.000000	0.0000	0.0000	0.0000	0.000	0.000	0.000	0.000000	0.000000	...	0.000000	0.00000	0.00000	0.00	0.000000	0.000000	0.000000	0.000000	0.00	0.000000
16	First Canadian Place, Underground city	0.000000	0.0000	0.0000	0.0000	0.000	0.000	0.000	0.020000	0.000000	...	0.000000	0.00000	0.00000	0.01	0.010000	0.000000	0.000000	0.010000	0.00	0.000000
17	Forest Hill North, Forest Hill West	0.000000	0.0000	0.0000	0.0000	0.000	0.000	0.000	0.000000	0.000000	...	0.000000	0.00000	0.25000	0.00	0.000000	0.000000	0.000000	0.000000	0.00	0.000000

Clusters

The results from k-means clustering show that we can categorize Toronto neighborhoods into 3 clusters based on how many Thai restaurants are in each neighborhood:
Cluster 0: Neighborhoods with little or no Thai restaurants
Cluster 1: Neighborhoods with no Thai restaurants
Cluster 2: Neighborhoods with high number of Thai restaurants
The results are visualized in the above map with Cluster 0 in red color, Cluster 1 in purple
Color and Cluster 2 in light green color.
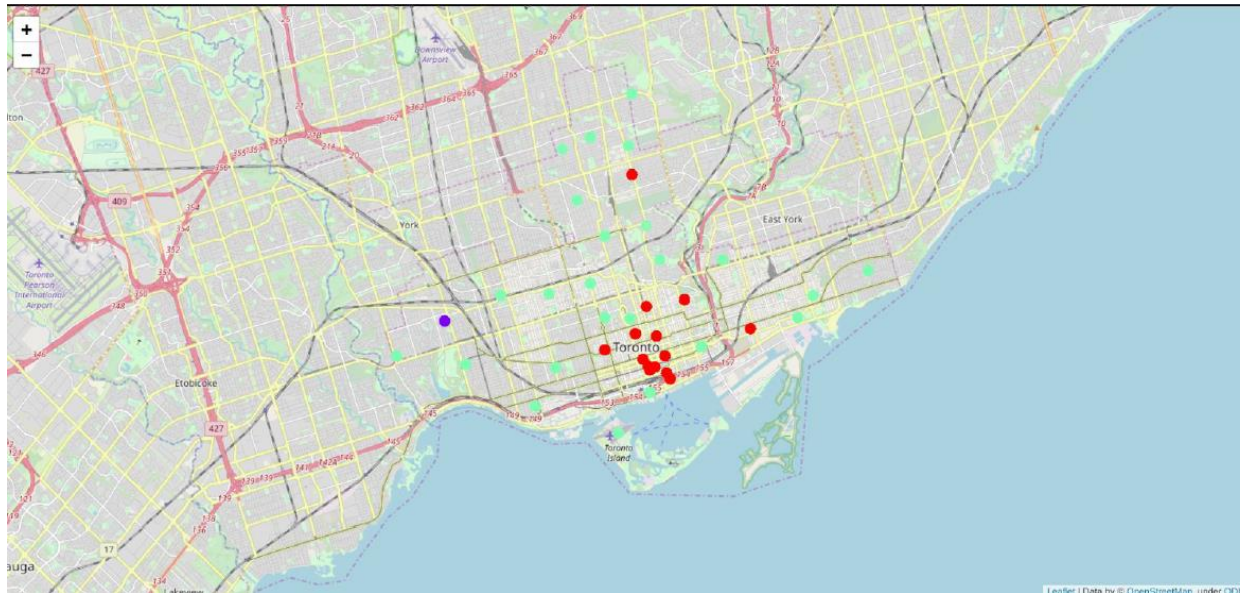
Recommendation

Most of Thai restaurants are in Cluster 2 which is around Adelaide, King, Richmond areas and lowest (close to zero) in Cluster 1 areas which are North Toronto West and Parkdale areas.

Also, there are good opportunities to open near Chinatown, St James town as the competition seems to be low. Looking at nearby venues, it seems Cluster 1 might be a good location as there are not a lot of Asian restaurants in these areas. Therefore, this project recommends the entrepreneur to open an authentic Burmese restaurant in these locations with little to no competition. Nonetheless, if the food is authentic, affordable and good taste, I am confident that it will have great following everywhere.

•

Conclusion

In this project, we have gone through the process of identifying the business problem, specifying the data required, extracting and preparing the data, performing the machine learning by utilizing k-means clustering and providing recommendation to the stakeholder.

References:

List of neighborhoods in Toronto:

https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

Foursquare Developer Documentation: https://developer.foursquare.com/docs