IBM Data Science Capstone Report

Leslie Chiu

August 22, 2020

Abstract

This report use data from Foursqaure to help a business owner make the decision about where to open the restaurant

1 Introduction

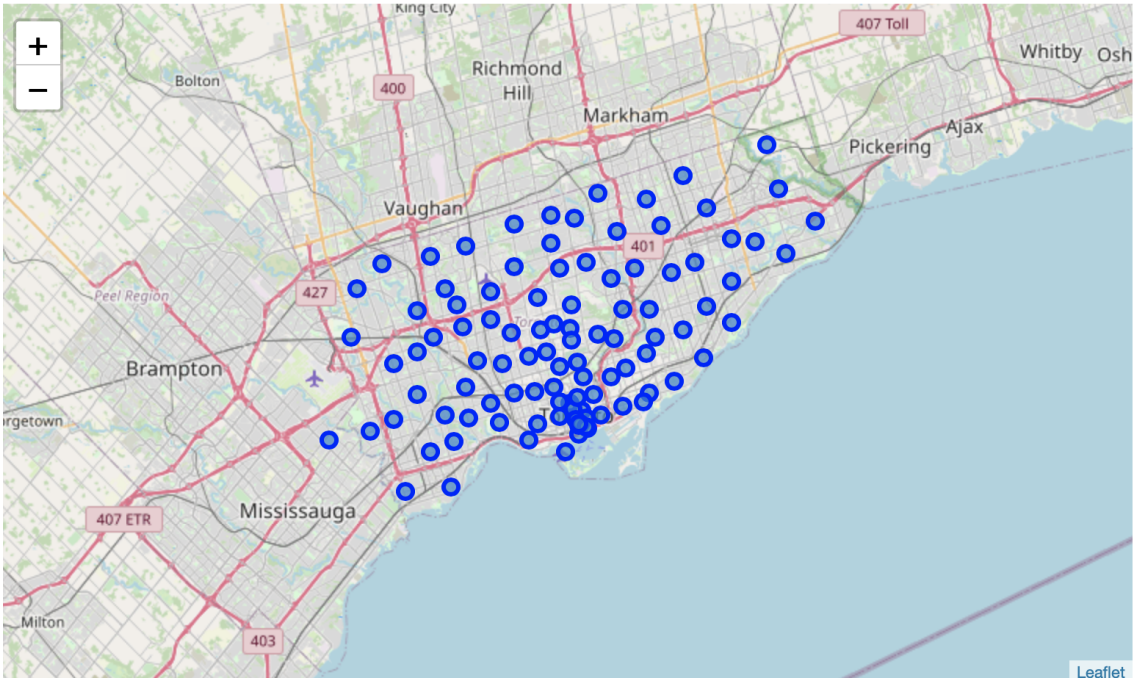
Mr. Sammon is a Korean immigrant and he is good at cooking. He wants to establish his own business by openning a Korean style restaurant in Toronto. However, he doesn't know the business environment in Toronto and people's lifestyles, thus has been struggling deciding where to open his business. We would like to conduct some analysis in this report and offer him some advice regarding this question.

2 Data

We use three different datasets, first two are what we have been using in the last week's project, which includes lists of postal codes of Canada and the geographical information. The third data set we will use in addition is the lists of venues obtained from Foursquare. We will combine these datasets and use them to see in which neighborhood the competition is not too harsh so that Mr. Sammon has a chance to make a profit.

3 Methodology

I first do some data visualization. I restrict my attention to downtown Toronto and plot all the neighborhoods in the following map to visually show their distributions.



Then I extract all the venues from FourSquare dataset and group them by venue type as well as neighbors as follows

	Neighborhood	Yoga Studio	Airport	Airport Food Court	Airport Gate	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	Antique Shop	...	Theater
0	Berczy Park	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000
1	CN Tower, King and Spadina, Railway Lands, Har...	0.000000	0.058824	0.058824	0.058824	0.117647	0.176471	0.117647	0.000000	0.000000	...	0.000000
2	Central Bay Street	0.015873	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000
3	Christie	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000
4	Church and Wellesley	0.026667	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.013333	0.000000	...	0.013333
5	Commerce Court, Victoria Hotel	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.040000	0.000000	...	0.000000
6	First Canadian Place, Underground city	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.030000	0.000000	...	0.010000
7	Garden District, Ryerson	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.020000
8	Harbourfront East, Union Station, Toronto	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.010000

I also make a rank based on the most popular type of venues within each neighborhood, this is similar to what we did for New York

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	..
0	Berczy Park	Coffee Shop	Café	Cheese Shop	Bakery	Cocktail Bar	Seafood Restaurant	Beer Bar	Restaurant	Farmers Market	..
1	CN Tower, King and Spadina, Railway Lands, Har...	Airport Service	Airport Lounge	Airport Terminal	Boutique	Rental Car Location	Airport	Airport Food Court	Airport Gate	Harbor / Marina	..
2	Central Bay Street	Coffee Shop	Sandwich Place	Café	Italian Restaurant	Japanese Restaurant	Salad Place	Bubble Tea Shop	Burger Joint	Modern European Restaurant	..
3	Christie	Grocery Store	Café	Park	Athletics & Sports	Diner	Restaurant	Italian Restaurant	Candy Store	Baby Store	..
4	Church and Wellesley	Coffee Shop	Japanese Restaurant	Sushi Restaurant	Gay Bar	Restaurant	Yoga Studio	Men's Store	Café	Bubble Tea Shop	..
5	Commerce Court, Victoria Hotel	Coffee Shop	Restaurant	Café	Hotel	Gym	American Restaurant	Italian Restaurant	Seafood Restaurant	Japanese Restaurant	..
6	First Canadian Place, Underground city	Coffee Shop	Café	Hotel	Restaurant	Gym	Japanese Restaurant	Asian Restaurant	Steakhouse	Seafood Restaurant	..
7	Garden District, Ryerson	Clothing Store	Coffee Shop	Café	Bubble Tea Shop	Cosmetics Shop	Japanese Restaurant	Lingerie Store	Italian Restaurant	Hotel	..
8	Harbourfront East, Union Station, Toronto Islands	Coffee Shop	Aquarium	Hotel	Café	Restaurant	Scenic Lookout	Fried Chicken Joint	Brewery	Italian Restaurant	..

In the next step, I conduct the main analysis, I first compute the density of restaurants that serve east Asian style food (including Asian Restaurant, Chinese Restaurant, Japanese Restaurant, Korean Restaurant, Sushi Restaurant, Taiwanese Restaurant, Thai Restaurant and Vietnamese Restaurant). The assumption I make is that the east Asian style food are similar so that people who are interested in other types of east Asian style food should have a higher probability to be also interested in Korean style restaurants. Thus I calculate the difference between the density of east Asian style restaurant and Korean restaurant: since the former density is an approximate how many people in this neighborhood might be interested in Korean restaurant and the later density is the current number of Korean restaurant, the difference is a good indicator of how profitable openning a Korean restaurant will be in this neighborhood.

4 Results and Discussion

	Neighborhood	diff
0	Berczy Park	0.053571
1	CN Tower, King and Spadina, Railway Lands, Har...	0.000000
2	Central Bay Street	0.063492
3	Christie	0.000000
4	Church and Wellesley	0.146667
5	Commerce Court, Victoria Hotel	0.060000
6	First Canadian Place, Underground city	0.100000
7	Garden District, Ryerson	0.070000
8	Harbourfront East, Union Station, Toronto Islands	0.030000
9	Kensington Market, Chinatown, Grange Park	0.057971
10	Queen's Park, Ontario Provincial Government	0.088235
11	Regent Park, Harbourfront	0.000000
12	Richmond, Adelaide, King	0.070000
13	Rosedale	0.000000
14	St. James Town	0.035294
15	St. James Town, Cabbagetown	0.086957
16	Stn A PO Boxes	0.051546
17	Toronto Dominion Centre, Design Exchange	0.070000
18	University of Toronto, Harbord	0.114286

The difference of density of east Asian restaurants and Korean restaurants is shown in the table above. As we can see, the largest difference appear in neighborhood Church and Wellesley. Since the former density is an approximate how many people in this neighborhood might be interested in Korean restaurant and the later density is the current number of Korean restaurant, the difference is a good indicator of how profitable openning a Korean restaurant will be in this neighborhood, thus I think it's a good idea for Mr. Sammon should open his Korean restaurant in the neighborhood of Church and Wellesley.

Admittedly, this analysis is pretty naive and preliminary, if more data available, it would be a great idea to analyze the population demographic structure in each neighborhood and also the profitability of existed restaurants to help Mr. Sammon make better decisions.

5 Conclusion

Mr. Sammon should open his Korean restaurant in the neighborhood of Church and Wellesley

6 References