Setting up a Restaurant in a Multicultural Country

Introduction

Background (Problem Statement)

Food touches people's soul and the demands of different varieties of cuisine in multicultural cities increase. This leads to business opportunities to appeal the multicultural audience. The aim of this project is to find a location for opening a profitable restaurant by studying the popularity of a variety of cuisines in a multicultural. The popularity of a type of restaurant is gauge by the number of similar type of restaurant in the area. Here, we focus on two similar size and population density multi-cultural cities – Singapore and Toronto. Singapore has a city area of 728.3km^2 with density population 7804/km^2 from 2019 estimate. Toronto city area, on the other hand, is 630.2 km^2 with density population 4334/km^2.

Data Overview

We obtain the list of neighborhoods in Toronto and Singapore via Wikipedia. Using Geocoder package, we can obtain the geographical location of the neighborhoods. Then, with Foursquare, we can explore the venues in these locations pertaining to restaurants. We filter the restaurants out of the top 10 venues to find out the cuisines offered by these restaurants

Data Acquisition

Step 1: List of Areas in Both Countries via Wikipedia

Singapore	Toronto
https://en.wikipedia.org/wiki/Regions_of_Singapore	https://en.wikipedia.org/w/index.php?title=List_of_postal_codes_of_Canada:_M&oldid=1011037969
Table 1: The URL where data is extracted.	

The names of area from each city are extracted from these sources for the respective countries.

Step 2: Geocoder package to Find Geographical Location

Singapore			Toronto				
Region	**Latitude**	**Longitude**	**PostalCode**	**Borough**	**Neighborhood**	**Latitude**	**Longitude**
The City	1.333108	103.819499	M3A	North York	Parkwoods	43.75245	-79.32991
Tampines	1.333108	103.943571	M4A	North York	Victoria Village	43.73057	-79.31306
Woodlands	1.333108	103.786216	M5A	Downtown Toronto	Regent Park, Harbourfront	43.65512	-79.36264
Seletar	1.333108	103.877379	M6A	North York	Lawrence Manor, Lawrence Heights	43.72327	-79.45042
Jurong East	1.333108	103.742294	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government	43.66253	-79.39188
			...	...	...	...	...
			M8X	Etobicoke	The Kingsway, Montgomery Road, Old Mill North	43.65319	-79.51113
			M4Y	Downtown Toronto	Church and Wellesley	43.66659	-79.38133
			M7Y	East Toronto	Business reply mail Processing Centre, South C...	43.64869	-79.38544
			M8Y	Etobicoke	Old Mill South, King's Mill Park, Sunnylea, Hu...	43.63278	-79.48945
			M8Z	Etobicoke	Mimico NW, The Queensway West, South of Bloor,...	43.62513	-79.52681

Table 2: Adding latitude and longitude to the data frame

With Geocoder and the data extracted from step 1, we find the latitude and longitude of the geographical locations

Step 3: Foursquare to Extract Venue and Venue Category from Geographical Location

Singapore					Toronto				
name	**categories**	**lat**	**lng**		**name**	**categories**	**lat**	**lng**	
Mr Teh Tarik Eating House	Food Court	1.378301	103.836725		Philosopher's Walk	Park	43.666894	-79.395597	
Charcos The Flaming Chicken	Australian Restaurant	1.378220	103.836871		Koerner Hall	Concert Hall	43.667983	-79.395962	
Selera Rasa Nasi Lemak	Malay Restaurant	1.380600	103.835820		Tiffany & Co.	Jewelry Store	43.669135	-79.393484	
Coffee Sense 咖啡鄉	Food Court	1.377207	103.837049		Royal Ontario Museum	Museum	43.668367	-79.394813	
Thomson Green Playground	Playground	1.378758	103.832444		ROM Museum Store	Gift Shop	43.668514	-79.394879	

Table 3: Add venue and venue category in the data frame

Foursquare provides the details of the locations based on the latitude and longitude information obtained in step 2.

Data Clean up

All the data is extracted and organized into data frames in step 1 (Table 2). The Toronto data frame requires additional data clean up before analyzing the trend.

Assumptions in Toronto Data

1. Borough with not assigned are omitted in data analysis
2. Extract boroughs contain 'Toronto' in it

With the cleaned data from step 1, we use the Geocoder package to obtain the latitude and longitude of the locations in both cities. We add two latitude and longitude columns in each of the data frame and use map folium to display the location of interest on the map (Figure 1).

Map Around the Region of Interest	
Singapore	Toronto
Figure 1	
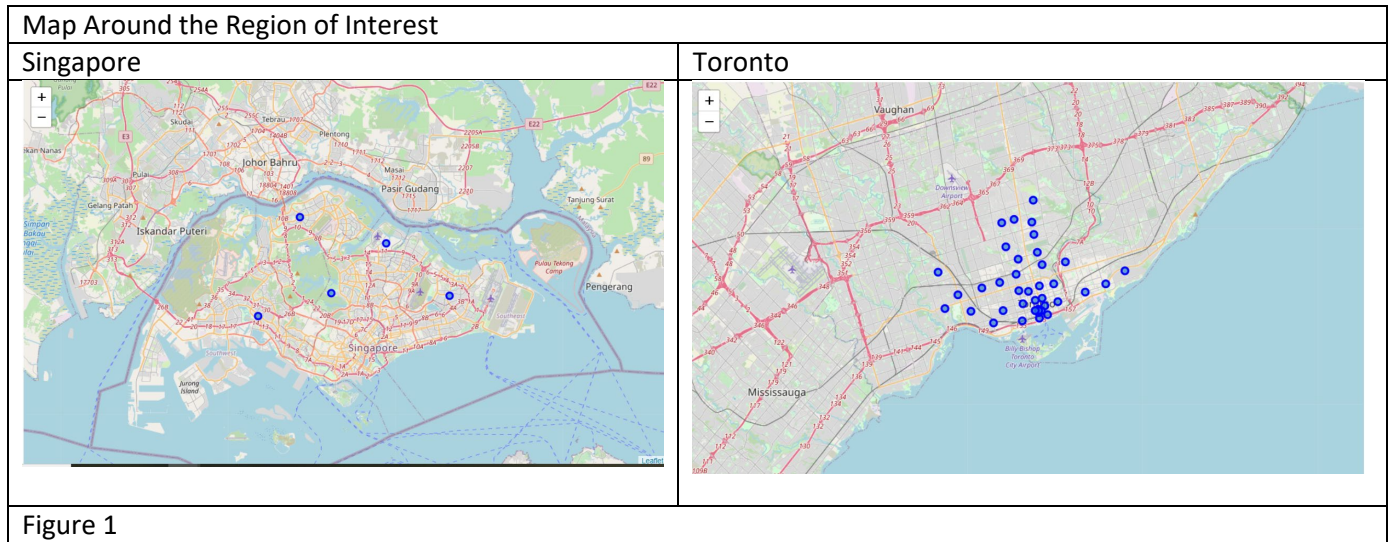
Next, we use the Foursquare API to provide the local venue and its corresponding categories within 500-meter radius for Toronto (5000-meter radius for Singapore)*.

* After this initial investigation, I discover there is another data set that split Singapore into more neighborhood from the initial dataset used. For the interest in time, I shall use the initial data set. This project can be improved later.

Data Exploration

After organizing the data, we can analyze it based on the venue category. First, we convert the categorical data to numerical data for Machine Learning purposes using "One hot encoding".

	Yoga Studio	American Restaurant	Antique Shop	Aquarium	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	BBQ Joint	...	Theater	Theme Restaurant	Toy / Game Store	Trail	Train Station	Vegetarian / Vegan Restaurant
0	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0
3	1	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0

Table 4: A data frame (Toronto example) containing the frequency of the venue corresponding where the index indicates the corresponding row of cleaned data frame as shown in Table 3

	Neighborhood	Yoga Studio	American Restaurant	Antique Shop	Aquarium	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	...	Theater	Theme Restaurant	Toy / Game Store
0	Berczy Park	0.000000	0.000000	0.000000	0.000000	0.033333	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.000000
1	Brockton, Parkdale Village, Exhibition Place	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.033333	0.000000	0.000000	...	0.000000	0.000000	0.000000
2	Business reply mail Processing Centre, South C...	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.033333	0.000000	...	0.066667	0.000000	0.000000
3	CN Tower, King and Spadina, Railway Lands,	0.033333	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.033333	0.000000	...	0.000000	0.000000	0.000000

Table 5: Create a data frame with the neighborhood with its corresponding mean of the number of occurrenc for each cater gory.

Then, we group the rows by neighborhood and take the mean of the number of occurrences for each category (Table 5)

	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	10th Most Common Venue
0	Berczy Park	Seafood Restaurant	Farmers Market	Cocktail Bar	Beer Bar	Coffee Shop	Café	Bistro	Restaurant	Jazz Club	Pharmacy
1	Brockton, Parkdale Village, Exhibition Place	Gift Shop	Coffee Shop	Italian Restaurant	Furniture / Home Store	Restaurant	Bakery	Sandwich Place	Japanese Restaurant	Breakfast Spot	Café
2	Business reply mail Processing Centre, South C...	Coffee Shop	Café	Concert Hall	Theater	Restaurant	Sushi Restaurant	Mediterranean Restaurant	Lounge	Opera House	Japanese Restaurant
3	CN Tower, King and Spadina, Railway Lands, Har...	Italian Restaurant	Park	Gym / Fitness Center	Restaurant	Yoga Studio	Sandwich Place	Brewery	Ramen Restaurant	Café	Caribbean Restaurant
4	Central Bay Street	Coffee Shop	Clothing Store	Plaza	Pizza Place	Park	Sandwich Place	Bubble Tea Shop	Ramen Restaurant	Poke Place	Café

Table 6: A data frame with the 10 top-most common venue for each neighborhood

Finally, we look at the types of venues for the top 10 most common venue for each neighborhood