Toronto Business Analysis of Tea shop

Author: Lin Cen

Introduction

1) Description of the problem and a discussion of the background.

Toronto, the most populous city in Canada, with a population of 2,731,571 as of 2016. Current to 2016, the Toronto census metropolitan area (CMA), of which the majority is within the Greater Toronto Area (GTA), held a population of 5,928,040, making it Canada's most populous CMA.The Toronto metropolitan area was the second-fastest-growing metropolitan area in North America, adding 125,298 persons.

The Toronto region's Food and Beverage sector employs more than 64,000 workers with annual wages totaling $3.2B. Sector businesses located within the city of Toronto account for more than 50 percent of this workforce.

The amount of tea that Canadians drink has doubled in the past few decades. As of 2015, Canadians drink 85 litres per person per year, up from 79.4 litres in 2008, and only 36 litres in 1991! Tea surpassed both soft drinks and bottled water, and is only slightly less than the amount of coffee that Canadians drink. However it's also interesting to note that the average amount of coffee drank by Canadians decreased from the previous year, while tea rose by several litres.

There are the same neighborhoods, some places have a high demand while there are others hardly getting any profit. So if I wanna open a tea store, where would be a nice place to open my first store? The objective of the following analysis is to check where would be popular in a determined neighborhood, and which neighborhood has a better acceptance for that kind. This capstone would be interesting for those who wanna open a tea store and have no idea where to open.

2) Description of the data and how it will be used to solve the problem.

The following data will be used to solve the problem:

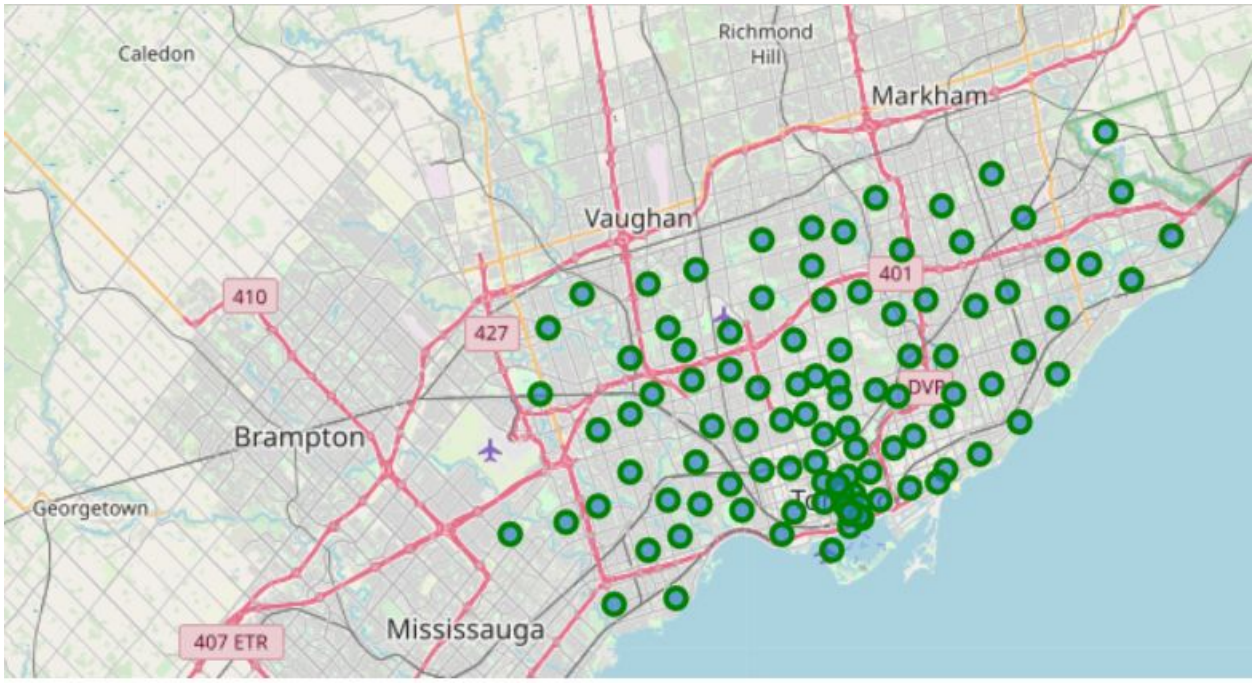
1. From Wikipedia, I have found a complete list of postal codes of Toronto, followed by theirs borough and neighborhood name: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
2. I used the Foursquare API to explore the top venues of a given neighborhood.
3. CSV file "Geospatial_Coordinates.csv" given by IBM

Methodology

Import table from URL Wikipedia, set dataframe with python and then insert the latitude and longitude of each neighborhood into the table.

	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

Using Python Folium library to create a geographical visualization of teh dataframe.



After that, with the help of the Foursquare API, I gathered the top 20* venues within a radius of 1 km from each neighborhood. Getting approximately 1800 popular venues, I created the following Data frame.

* Not all neighborhoods have returned 20 venues, some of them returned less than 10 results.

	name	categories	lat	lng
0	Wendy's	Fast Food Restaurant	43.802008	-79.198080
1	Wendy's	Fast Food Restaurant	43.807448	-79.199056
2	Staples Morningside	Paper / Office Supplies Store	43.800285	-79.196607
3	Harvey's	Restaurant	43.800020	-79.198307
4	Caribbean Wave	Caribbean Restaurant	43.798558	-79.195777

Since the focus is tea shop related venues, I cleaned this data frame to show results only for the tea related categories. And took the mean frequency of occurrence of each category in each neighborhood. Resulting in a data frame that could measure how popular is each of those categories in every neighborhood:

	Neighborhood	Breakfast Spot	Bubble Tea Shop	Café	Cantonese Restaurant	Chinese Restaurant	Coffee Shop	Tea Room
0	Adelaide, King, Richmond	0.000000	0.00000	0.050000	0.00	0.000000	0.000000	0.00
1	Agincourt	0.050000	0.00000	0.000000	0.05	0.150000	0.000000	0.00
2	Agincourt North, L'Amoreaux East, Milliken, St...	0.000000	0.00000	0.000000	0.00	0.150000	0.050000	0.00
3	Albion Gardens, Beaumond Heights, Humbergate, ...	0.000000	0.00000	0.000000	0.00	0.000000	0.052632	0.00
4	Alderwood, Long Branch	0.000000	0.00000	0.000000	0.00	0.000000	0.050000	0.00
5	Bathurst Manor, Downsview North, Wilson Heights	0.000000	0.00000	0.000000	0.00	0.000000	0.100000	0.00
6	Bayview Village	0.000000	0.00000	0.066667	0.00	0.066667	0.000000	0.00
7	Bedford Park, Lawrence Manor East	0.000000	0.00000	0.050000	0.00	0.000000	0.150000	0.00
8	Berczy Park	0.000000	0.00000	0.000000	0.00	0.000000	0.000000	0.05
9	Birch Cliff, Cliffside West	0.000000	0.00000	0.090909	0.00	0.000000	0.000000	0.00
10	Bloordale Gardens, Eringate, Markland Wood, Ol...	0.000000	0.00000	0.055556	0.00	0.000000	0.111111	0.00

I then added a ´Total´ column, which is the sum of the popularity of each category of interest, so I can know how popular a venue of interest is in each neighborhood. Finally, I used an unsupervised machine learning algorithm to cluster each region based on the popularity. The algorithm used was the K-means algorithm, which is one of the most common algorithms for clustering.The ideal number of clusters (3) was found using the Elbow method.

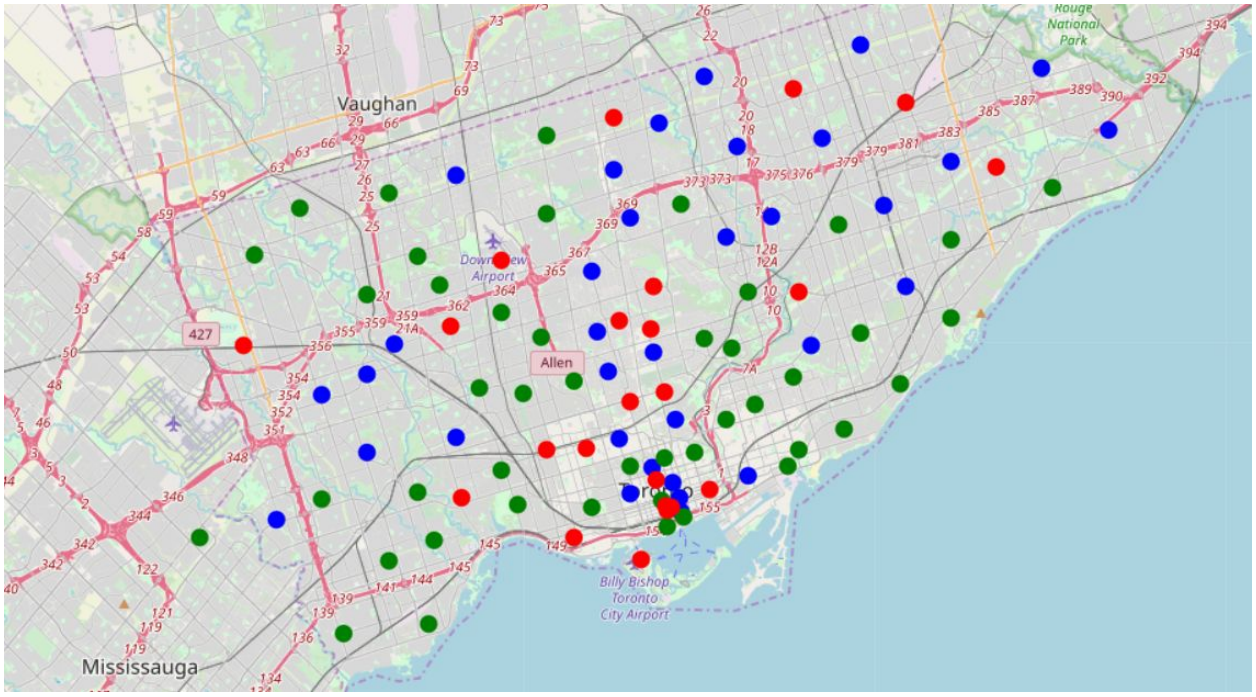
Results

Now that we have the score of popularity for each neighborhood. We can now set the area where popularity of this type of shop is greater than 0.2

	Postcode	Borough	Neighborhood	Latitude	Longitude	Breakfast Spot	Bubble Tea Shop	Café	Cantonese Restaurant	Chinese Restaurant	Coffee Shop	Tea Room	Total
60	M5L	Downtown Toronto	Commerce Court, Victoria Hotel	43.648198	-79.379817	0.00	0.00000	0.250000	0.00	0.000000	0.100000	0.05	0.400000
44	M4P	Central Toronto	Davisville North	43.712751	-79.390197	0.05	0.00000	0.150000	0.00	0.000000	0.100000	0.05	0.350000
52	M5A	Downtown Toronto	Harbourfront	43.654260	-79.360636	0.10	0.00000	0.000000	0.00	0.000000	0.250000	0.00	0.350000
75	M6H	West Toronto	Dovercourt Village, Dufferin	43.669005	-79.442259	0.00	0.00000	0.200000	0.00	0.000000	0.150000	0.00	0.350000
69	M5X	Downtown Toronto	First Canadian Place, Underground city	43.648429	-79.382280	0.00	0.00000	0.200000	0.00	0.000000	0.150000	0.00	0.350000
59	M5K	Downtown Toronto	Design Exchange, Toronto Dominion Centre	43.647177	-79.381576	0.00	0.00000	0.150000	0.00	0.000000	0.200000	0.00	0.350000
3	M1G	Scarborough	Woburn	43.770992	-79.216917	0.00	0.00000	0.000000	0.00	0.111111	0.222222	0.00	0.333333
67	M5V	Downtown Toronto	CN Tower, Bathurst Quay, Island airport, Harbo...	43.628947	-79.394420	0.00	0.00000	0.125000	0.00	0.000000	0.187500	0.00	0.312500
15	M1W	Scarborough	L'Amoreaux West	43.799525	-79.318389	0.05	0.05000	0.000000	0.00	0.100000	0.100000	0.00	0.300000
43	M4N	Central Toronto	Lawrence Park	43.728020	-79.388790	0.00	0.00000	0.200000	0.00	0.000000	0.100000	0.00	0.300000
83	M6S	West Toronto	Runnymede, Swansea	43.651571	-79.484450	0.00	0.00000	0.150000	0.00	0.000000	0.100000	0.05	0.300000
56	M5G	Downtown Toronto	Central Bay Street	43.657952	-79.387383	0.00	0.05000	0.000000	0.00	0.000000	0.200000	0.05	0.300000
78	M6L	North York	Downsview, North Park, Upwood Park	43.713756	-79.490074	0.00	0.00000	0.000000	0.00	0.090909	0.181818	0.00	0.272727
12	M1S	Scarborough	Agincourt	43.794200	-79.262029	0.05	0.00000	0.000000	0.05	0.150000	0.000000	0.00	0.250000

With the final result of each cluster:

- High means Green
- Medium means Blue
- Low means Red



Conclusion

As a result, the neighborhoods with green and blue colors would be more interesting for new business related to tea since the popularity of similar venues are high on those areas.

Reference

[1] https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
[2] https://developer.foursquare.com/
[3] CSV file "Geospatial_Coordinates.csv" given by IBM