Find a Location for a New Cafe in Toronto

Coursera Capstone Project

Introduction

Toronto is a city filled with coffee lovers. There are big chain coffee shops like Tim Hortons and Starbucks, and indie cafés like Balzac's Coffee and Sam James Coffee Bar.

In this project I am going to find an optimal location for a cafés. Since there are lots of cafés in Toronto, I will try to detect locations that are not already crowded with cafés and in Downtown Toronto if possible. The target audience of this project would be someone who would like to open a new café in Toronto.

I will use the data science power to generate a few most promising neighborhoods based on these criteria. Advantages of each area will then be clearly expressed so that best possible final location can be chosen by stakeholders.

Data

Based on the definition of our problem, factors that will influence decision are:
- Number of existing cafés in the neighborhood
- Number of existing big chain coffee shops in the neighborhood
- If the neighborhood is in Downtown Toronto or not

Following data sources will be needed to extract/generate the required information:
- The postal codes and neighborhoods information will be scraped from a Wikipedia page: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
- Number of cafés and location in every neighborhood will be obtained using Foursquare API

Methodology

In the first part of this project, I am going to create a dataframe with all postal codes, boroughs, neighborhoods and geographical coordinates (latitudes and longitudes) of Toronto.

Firstly, I am going to scrape a Wikipedia page (https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M) using a Python package called BeautifulSoup, and get the postal codes and corresponding boroughs and neighborhoods in Toronto.

After cleaning the table, I will merge this dataframe together with a dataframe with postal codes and geographical coordinates in Toronto. Then I will get the dataframe as shown on the next slide.

	PostalCode	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
5	M1J	Scarborough	Scarborough Village	43.744734	-79.239476
6	M1K	Scarborough	East Birchmount Park, Ionview, Kennedy Park	43.727929	-79.262029
7	M1L	Scarborough	Clairlea, Golden Mile, Oakridge	43.711112	-79.284577
8	M1M	Scarborough	Cliffcrest, Cliffside, Scarborough Village West	43.716316	-79.239476
9	M1N	Scarborough	Birch Cliff, Cliffside West	43.692657	-79.264848

The first 10 rows of the dataframe. There are 11 boroughs and 103 neighborhoods in total.
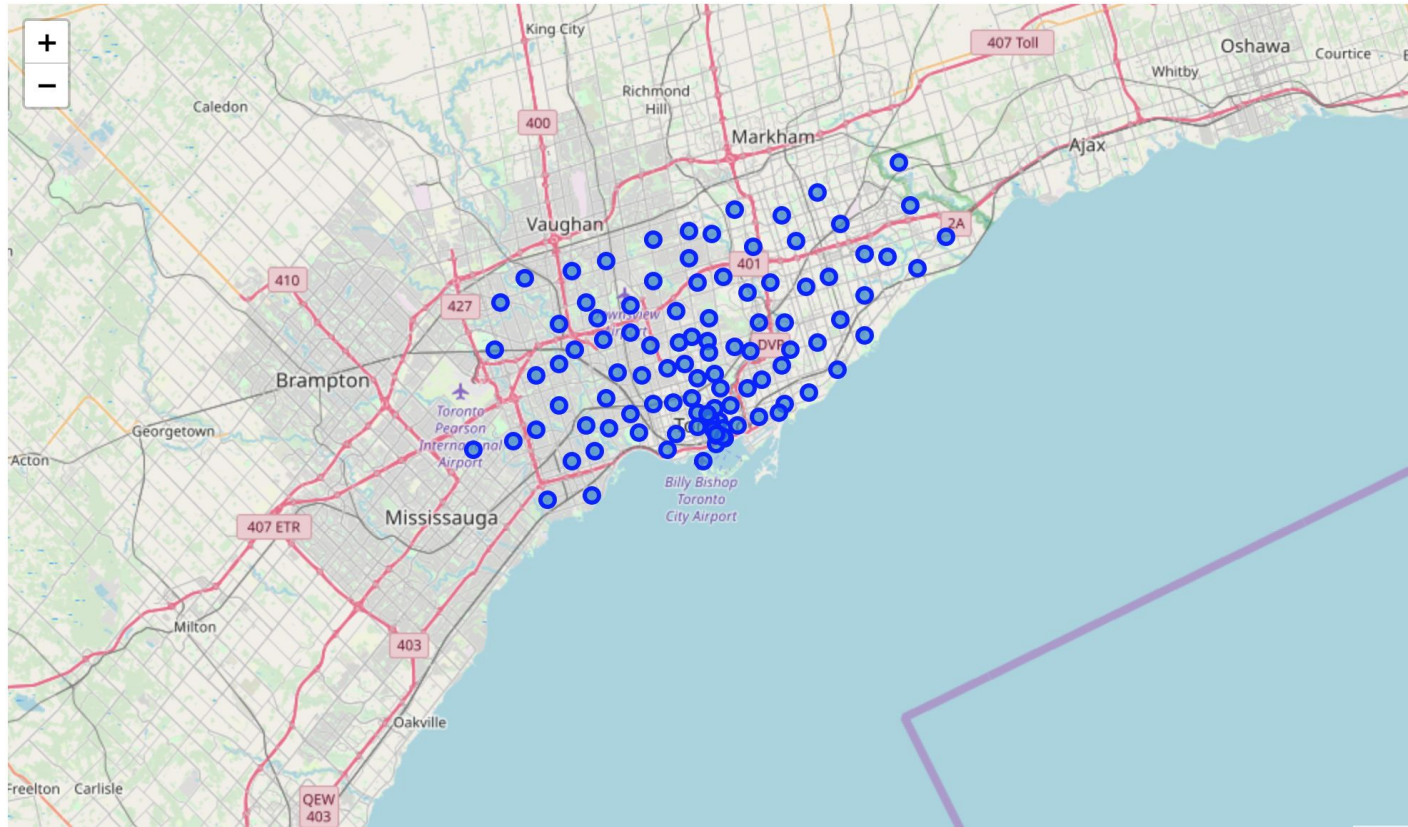
Use geopy library in Python to get the latitude and longitude values of Toronto and use folium library to create a map of Toronto with neighborhoods superimposed on top

In the second part, I am going to explore the neighborhoods in Toronto using Foursquare API.

Foursquare is a technology company that built a massive dataset of location data. Foursquare powers location data for Apple Maps, Uber, Snapchat, Twitter, and many others. Their API and location data are currently being used by over 100,000 develops.

I create a function to get nearby venues in neighborhoods in Toronto and use it to get all cafés within 1000 metres in all neighborhoods. The result contains other venue categories except Coffee Shop and Café. Those venues may offer coffee, but in this project I will focus on coffee shops and cafés only. So the next step is to get rid of venues other than coffee shops and cafés.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Guildwood, Morningside, West Hill	43.763573	-79.188711	Tim Hortons	43.769867	-79.186883	Coffee Shop
1	Cedarbrae	43.773136	-79.239476	Coffee Culture	43.776278	-79.234026	Café
4	Birch Cliff, Cliffside West	43.692657	-79.264848	The Birchcliff	43.691666	-79.264532	Café
5	Dorset Park, Scarborough Town Centre, Wexford ...	43.757410	-79.273304	Nova Ristorante	43.749434	-79.277630	Café
7	Agincourt	43.794200	-79.262029	Francesca Bakery	43.787716	-79.256852	Café
8	Agincourt North, L'Amoreaux East, Milliken, St...	43.815252	-79.284577	Starbucks	43.816165	-79.293562	Coffee Shop
11	Fairview, Henry Farm, Oriole	43.778517	-79.346556	Ten Ren's Tea Time	43.777158	-79.344531	Café
12	Fairview, Henry Farm, Oriole	43.778517	-79.346556	Aroma Espresso Bar	43.777700	-79.344652	Café
13	Fairview, Henry Farm, Oriole	43.778517	-79.346556	Teaopia	43.778743	-79.344512	Café
14	Newtonbrook, Willowdale	43.789053	-79.408493	Red Rose Pâtisserie	43.793285	-79.419027	Café

First 10 results are shown but there are 3227 different cafes in total. That's a lot!

Finally, in the last part, I am going to choose the optimal neighborhoods which satisfy all the criteria.

Firstly, I use groupby function in Python to get the top 10 neighborhoods with most cafés and top 10 cafés with most stores in Toronto, then add these values to the dataframe with all neighborhood data, and name the column Café and Big Café.

Secondly, it would be an advantage if a neighborhood is in Downtown Toronto, so I will add a column for this too. If the neighborhood is in Downtown Toronto, I will assign a value of 1.0 for this neighborhood. If not, a value of 0 will be assigned.

	PostalCode	Borough	Neighborhood	Latitude	Longitude	Café	Big Café	Downtown
0	M1B	Scarborough	Rouge, Malvern	43.806686	-79.194353	3.0	2.0	0
1	M1C	Scarborough	Highland Creek, Rouge Hill, Port Union	43.784535	-79.160497	1.0	0.0	0
2	M1E	Scarborough	Guildwood, Morningside, West Hill	43.763573	-79.188711	3.0	3.0	0
3	M1G	Scarborough	Woburn	43.770992	-79.216917	4.0	2.0	0
4	M1H	Scarborough	Cedarbrae	43.773136	-79.239476	4.0	2.0	0
5	M1J	Scarborough	Scarborough Village	43.744734	-79.239476	2.0	1.0	0
6	M1K	Scarborough	East Birchmount Park, Ionview, Kennedy Park	43.727929	-79.262029	5.0	3.0	0
7	M1L	Scarborough	Clairlea, Golden Mile, Oakridge	43.711112	-79.284577	4.0	3.0	0
8	M1M	Scarborough	Cliffcrest, Cliffside, Scarborough Village West	43.716316	-79.239476	0.0	0.0	0
9	M1N	Scarborough	Birch Cliff, Cliffside West	43.692657	-79.264848	3.0	0.0	0

The first 10 rows of the final dataframe

Results

After getting the dataframe with all information needed, I will define the weights for all criteria.

- Assign the weight of -1 to Cafe
- Assign the weight of -2 to Big Cafe
- Assign the weight of 10 to Downtown

Then calculate the score for each neighborhood, with the starting score of 100.

	Borough	Neighborhood	Score
31	North York	Downsview West	99.0
80	York	Del Ray, Keelesdale, Mount Dennis, Silverthorn	99.0
32	North York	Downsview Central	99.0
100	Etobicoke	Kingsview Village, Martin Grove Gardens, Richv...	99.0
23	North York	York Mills West	99.0
1	Scarborough	Highland Creek, Rouge Hill, Port Union	99.0
93	Etobicoke	Islington Avenue	99.0
19	North York	Bayview Village	98.0
24	North York	Willowdale West	97.0
94	Etobicoke	Cloverdale, Islington, Martin Grove, Princess ...	97.0

In the above dataframe showing the first 10 neighborhoods of the results, 7 of them score 99.0, and 5 of them are in North York borough. Therefore, I will choose neighborhoods scored 99.0 and in North York borough as the best locations: Downsview West, Downsview Central and York Mills West.

Discussion

Due to time constraints, there are some problems need to be done in this project.

Firstly, I only take three factors to make decisions. There are many more factors that can be taken into account, such as the population, transportation, parking availability in the neighborhood.

Secondly, although there are no duplicate venues in the dataframe, 3227 coffee shops seem to be too many in a city, so there may be some problem with the data.

Lastly, there are 6 neighborhoods with no cafés, so they do not have a score in this battle.

Conclusion

From the research throughout this project, three neighborhoods are chosen to be the optimal locations for a new cafe opening in Toronto, which are Downsview West, Downsview Central and York Mills West. However, I have not take all factors into account to make a decision. Improving the results will help the future coffee shop owner make a better decision.