Coursera - Applied Data Science Capstone Project

Introduction

Cycling is a common mode of transportation within the city and is a great recreational activity. For this capstone project, we'll be looking into where to open a new cycling store in Toronto.

Business Problem

When trying to open a new business, a critical aspect is finding the right location. When assessing whether a location is suitable it's important to assess whether there is a community in that area that could benefit from the services that your business will be offering. If there is a community, it's also important to understand the local competition. What other businesses are existing in the local area which would offer similar services and act as competition to your business - it may not be ideal to open a business in a section of the city where the market may already be saturated.

Data

To perform the analysis for the two considerations above, we'll need to obtain supporting data which can be reviewed to assess the identified conditions.

Using Wikipedia, we can gather information about different neighbourhoods within the Toronto area. (https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M) With this information, GeoPy can be utilized to gather positioning data (Latitude and Longitude) to plot out the areas on a map to help with a visual assessment of where each neighbourhood is.

Through the use of FourSquare's API, for each neighbourhood location - a listing of local businesses that offer similar services can be gathered. This can help provide details as to how many existing businesses are within a specific radius to these neighbourhoods.

The City of Toronto provides open data to the public which can help us understand whether Cycling infrastructure exists within our neighbourhoods and how much. The amount of bicycling parking infrastructure that exists in each neighbourhood may help in evaluating the prevalence of cycling in that neighbourhood.
(https://open.toronto.ca/dataset/bicycle-parking-high-capacity-outdoor/)

Methodology

Generating an initial set of data within the environment was the first action taken. Data was extracted from wikipedia using BeautifulSoup and then placed into a Panda's Dataframe. The

data was reviewed and cleaned up. Locational data (latitude and longitude) data was then added to the data frame using a csv provided as part of the course. The final dataset was then adjusted to only include boroughs from Toronto.

Following this, Bicycle Parking data from the City of Toronto Open Data was downloaded and imported into the environment. That data was adjusted to group based on Forward Sortation Area (First three characters of the Postal Code). The data was merged with the original dataset providing a data frame which now included Postal Codes, Borough, Neighbourhood, Longitude, Latitude and Bicycle Capacity (Parking) information. This information was then mapped out to help provide a visual representation of the neighbourhoods and their associated parking capacities.

To gather more information on our neighbourhoods, data from FourSquare was pulled. For each neighbourhood record, data was pulled from FourSquares API which provided details of venues nearby (within a radius of 500m) which was categorized by FourSquare as a 'Bike Shop' (CategoryID: 4bf58dd8d48988d115951735)

Venue location was then mapped to help provide a visual representation of the distribution of the Bike Shops in the areas.

This information was grouped and merged into our original dataframe to provide a count of the number of nearby venues which were classified as 'Bike Shops' for each of our neighbourhood records. This now left us with information which represented the count of Bike Parking spots and Bike Shops in our neighbourhoods.

To help provide further details on a wide level than individual neighbourhoods, Clustering was used to help group together nearby neighbourhoods. 4 Clusters were used which helped split the city into Central, North, East and West. (South of the downtown core is the lake).

This data was mapped to help understand our 4 major areas (clustered). Using this data provides further details and representation as the grouped distribution of shops and parking capacity based on major subsections of the city.

Results

There were 22 neighbourhoods in Toronto being reviewed. The neighbourhood locations can be seen on the map below:

The final dataframe by Postal Code (Neighbourhoods) helps provide an indication of areas where there was a high availability of bicycle parking and a lower amount of competition. Two specific areas were the Richmond, Adelaide and King neighbourhood and the neighbourhood around the CN Tower. These area had a substantial amount of bicycle parking spaces, 144 and 150 respectively, without any nearby businesses that were categorized as a bike shop.

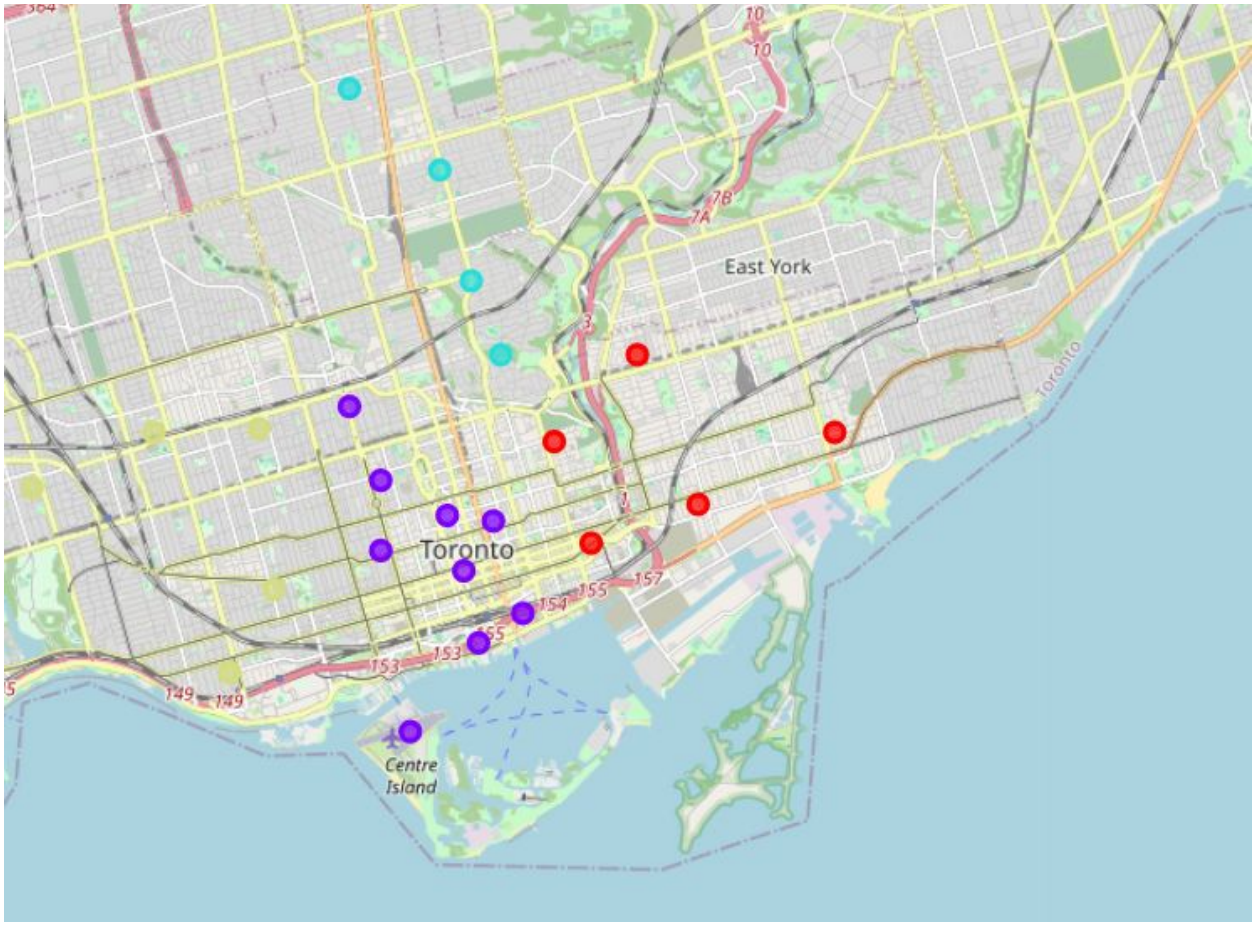
Neighbourhood	Postal_Code	Borough	Latitude	Longitude	Bike_Rack_Capacity	Store_Count
Regent Park, Harbourfront	M5A	Downtown Toronto	43.654260	-79.360636	78	0.0
Garden District, Ryerson	M5B	Downtown Toronto	43.657162	-79.378937	48	0.0
Berczy Park	M5E	Downtown Toronto	43.644771	-79.373306	8	0.0
Central Bay Street	M5G	Downtown Toronto	43.657952	-79.387383	80	2.0
Christie	M6G	Downtown Toronto	43.669542	-79.422564	112	1.0
Richmond, Adelaide, King	M5H	Downtown Toronto	43.650571	-79.384568	144	0.0
Dufferin, Dovercourt Village	M6H	West Toronto	43.669005	-79.442259	80	2.0
Harbourfront East, Union Station, Toronto Islands	M5J	Downtown Toronto	43.640816	-79.381752	102	1.0
Little Portugal, Trinity	M6J	West Toronto	43.647927	-79.419750	93	2.0
The Danforth West, Riverdale	M4K	East Toronto	43.679557	-79.352188	32	1.0
Brockton, Parkdale Village, Exhibition Place	M6K	West Toronto	43.636847	-79.428191	16	0.0
India Bazaar, The Beaches West	M4L	East Toronto	43.668999	-79.315572	8	1.0
Studio District	M4M	East Toronto	43.659526	-79.340923	40	2.0
High Park, The Junction South	M6P	West Toronto	43.661608	-79.464763	42	3.0
North Toronto West, Lawrence Park	M4R	Central Toronto	43.715383	-79.405678	16	0.0
The Annex, North Midtown, Yorkville	M5R	Central Toronto	43.672710	-79.405678	16	1.0
Davisville	M4S	Central Toronto	43.704324	-79.388790	24	0.0
University of Toronto, Harbord	M5S	Downtown Toronto	43.662696	-79.400049	32	3.0
Moore Park, Summerhill East	M4T	Central Toronto	43.689574	-79.383160	32	0.0
Kensington Market, Chinatown, Grange Park	M5T	Downtown Toronto	43.653206	-79.400049	98	3.0
CN Tower, King and Spadina, Railway Lands, Har...	M5V	Downtown Toronto	43.628947	-79.394420	150	0.0
Rosedale	M4W	Downtown Toronto	43.679563	-79.377529	40	0.0
St. James Town, Cabbagetown	M4X	Downtown Toronto	43.667967	-79.367675	8	0.0

However, once clustered, both neighbourhoods fell within cluster grouping 1 which included the highest number of store counts overall. While each neighbourhood did not have a venue within the specified radius - they were in a clustered zone which provided competing services.

Cluster Labels	Bike_Rack_Capacity	Store_Count
0	166	4.0
1	678	10.0
2	112	0.0
3	343	8.0

	Cluster Labels	Neighbourhood	Postal_Code	Borough	Latitude	Longitude	Bike_Rack_Capacity	Store_Count
0	0	Regent Park, Harbourfront	M5A	Downtown Toronto	43.654260	-79.360636	78	0.0
1	1	Garden District, Ryerson	M5B	Downtown Toronto	43.657162	-79.378937	48	0.0
2	1	Berczy Park	M5E	Downtown Toronto	43.644771	-79.373306	8	0.0
3	1	Central Bay Street	M5G	Downtown Toronto	43.657952	-79.387383	80	2.0
4	3	Christie	M6G	Downtown Toronto	43.669542	-79.422564	112	1.0
5	1	Richmond, Adelaide, King	M5H	Downtown Toronto	43.650571	-79.384568	144	0.0
6	3	Dufferin, Dovercourt Village	M6H	West Toronto	43.669005	-79.442259	80	2.0
7	1	Harbourfront East, Union Station, Toronto Islands	M5J	Downtown Toronto	43.640816	-79.381752	102	1.0
8	3	Little Portugal, Trinity	M6J	West Toronto	43.647927	-79.419750	93	2.0
9	0	The Danforth West, Riverdale	M4K	East Toronto	43.679557	-79.352188	32	1.0
10	3	Brockton, Parkdale Village, Exhibition Place	M6K	West Toronto	43.636847	-79.428191	16	0.0
11	0	India Bazaar, The Beaches West	M4L	East Toronto	43.668999	-79.315572	8	1.0
12	0	Studio District	M4M	East Toronto	43.659526	-79.340923	40	2.0
13	3	High Park, The Junction South	M6P	West Toronto	43.661608	-79.464763	42	3.0
14	2	North Toronto West, Lawrence Park	M4R	Central Toronto	43.715383	-79.405678	16	0.0
15	1	The Annex, North Midtown, Yorkville	M5R	Central Toronto	43.672710	-79.405678	16	1.0
16	2	Davisville	M4S	Central Toronto	43.704324	-79.388790	24	0.0
17	1	University of Toronto, Harbord	M5S	Downtown Toronto	43.662696	-79.400049	32	3.0
18	2	Moore Park, Summerhill East	M4T	Central Toronto	43.689574	-79.383160	32	0.0
19	1	Kensington Market, Chinatown, Grange Park	M5T	Downtown Toronto	43.653206	-79.400049	98	3.0
20	1	CN Tower, King and Spadina, Railway Lands, Har...	M5V	Downtown Toronto	43.628947	-79.394420	150	0.0
21	2	Rosedale	M4W	Downtown Toronto	43.679563	-79.377529	40	0.0
22	0	St. James Town, Cabbagetown	M4X	Downtown Toronto	43.667967	-79.367675	8	0.0

Factoring in clusters, it becomes evident that Cluster 2 (North) was an area where there were no competitors within the cluster. The area did have the lowest capacity for bicycle parking, however it was still present.



In the above map, our neighbourhoods are divided into 4 clusters. Within Cyan is cluster 2, our north grouping which presents zero competing businesses to the proposed bike shop. Within Purple is Cluster 1, located in the downtown core of Toronto. This cluster included two neighbourhoods with the highest amount of available bike parking within zero competitors, however the larger clustered area does include several competing businesses. (Represented in Red is cluster 0 and within a mustard yellow is cluster 3)

Discussion

When determining the location to open a new business there are many factors that must be assessed to determine the ideal location. Using existing available data, data science can help provide tools to make an educated decision. Ultimately, the decision and factors which are important to a business owner up to the owners discretion as there is no single correct answer.

The assessment above reviews two factors which can help drive a decision, however there are other factors which may be important when making a decision. These could include demographic details as to age and overall population. Operational consideration such as the

cost of doing business such as rent and utilities may be useful variables. It is also important to consider other cycling activities such as trails and bike parks within the area.

Based on the investigation performed, my recommendation would be research and outline the business factors which are perceived as critical to making this decision. Within those factors, gather data sources and determine a weighting. With additional data and identified weighted criteria - an extended investigation may provide further insights on where to open a new bike shop. However, based on the data at hand I would recommend opening to the North where there is not existing competition but still a significant potential client base.

Conclusion

When opening a new Bike Store, there would be the lowest amount of competition if opening a new store to the North of downtown. However, based on bike rack concentration and clustering, there may be a higher need within the Richmond, Adelaide and King neighbourhood or the CN Tower neighbourhood.