

Applied Data Science Capstone Final Project:

The Battle of Neighbourhoods

Business in Toronto

Introduction

Problem background

Nowadays, people tend to search online before they make decisions. There are many websites summarizing the most popular business types that one can consider to develop in a city. It is always good to know where in the city is the best place to start. It is found on three webpages that the top three business investment opportunities that one may go for in Toronto are:

Rank	Profitableventure	OPSTART*	Business Insider
1	Gluten-free restaurants and bakeries	Real Estate	Alcohol delivery service
2	Daily-care and Babysitting Services	Pet Services	Café dinning
3	Patient Advocacy	Financial Consulting	Fitbit

*Noted that the second website rank “E-commerce” as the most promising business but it is not put in the table because E-commerce may not be recorded as neighbourhood business.

This project aims to recommend the suitable district/neighbourhood to start the business which ranks top in the three websites shown above based on the neighbourhood study in Toronto.

Target client

The target group will be individuals or companies who are looking for potential small business opportunities in Toronto. This project will provide suggestions on potential business types and where to develop them in Toronto.

Data description

The data that will be used to solve the problem and the source of the data are described in the table below.

NO	Data	Type	Source
1	Top 20 Small Business Investment Opportunities in Toronto	Exploratory	https://www.profitableventure.com/toronto-ontario/
2	The Most Promising Business Opportunities In Canada	Exploratory	https://www.opstart.ca/promising-business-opportunities-canada/

3	The 21 Coolest Small Businesses In Toronto	Exploratory	https://www.businessinsider.com/coolest-small-businesses-in-toronto-2014-7?IR=T
4	List of postal codes of Canada: M	Exploratory	https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
5	Geospatial data of the neighbourhoods	Geo-data	http://cocl.us/Geospatial_data
6	Toronto exploring	Request	Foursquare

The first three are informatic data which provide the ranking of promising small business in Toronto based on social survey or research. The fourth source is the Wikipedia page which includes that postal codes of Toronto districts. The fifth source contains the geo-coordinates of the neighbourhoods. Based on source 4 and 5, exploratory request will be made for the districts in Toronto to the last source, Foursquare, to get the detailed neighbourhood information for each district. Statistical analysis will be done to the neighbourhood information to find out the optimal area to developed the highly ranked business types found in the first three sources.

Methodology

Data collection

Neighbourhood information is extracted from source 4 (refer to the table in data description) through a python library “Beautifulsoup”. Part of the organised neighbourhood data is shown as below:

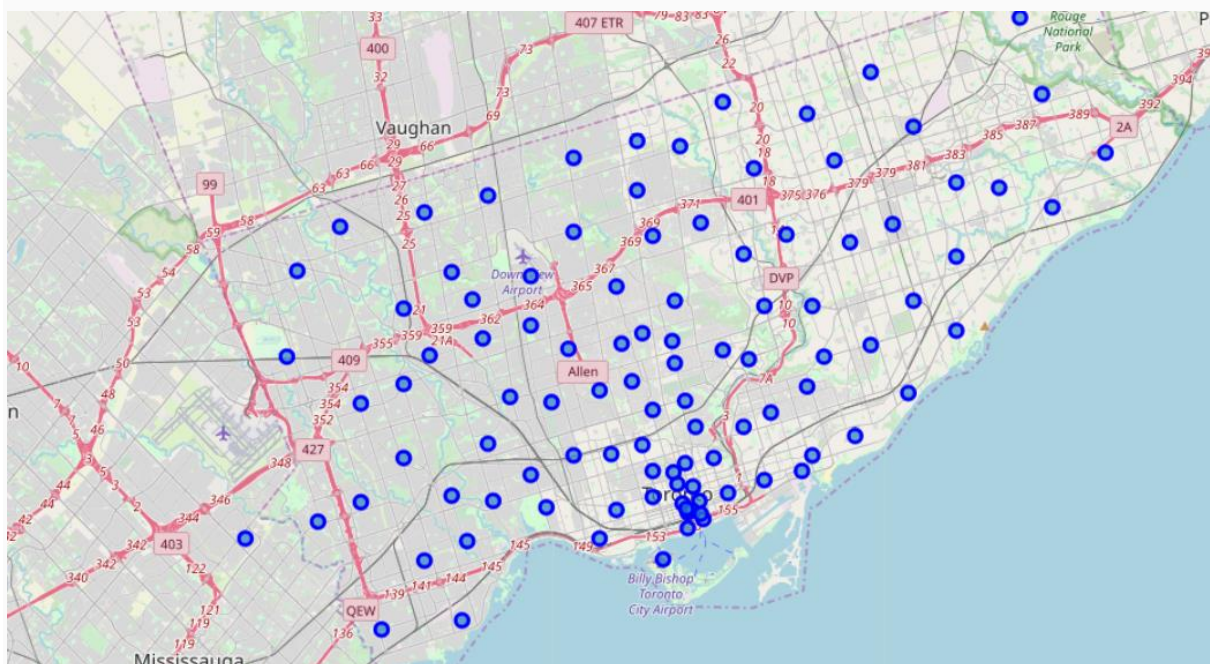
	Postcode	Borough	Neighborhood
0	M1B	Scarborough	Rouge,Malvern
1	M1C	Scarborough	Highland Creek,Rouge Hill,Port Union
2	M1E	Scarborough	Guildwood,Morningside,West Hill
3	M1G	Scarborough	Woburn
4	M1H	Scarborough	Cedarbrae
5	M1J	Scarborough	Scarborough Village
6	M1K	Scarborough	East Birchmount Park,Ionview,Kennedy Park
7	M1L	Scarborough	Clairlea,Golden Mile,Oakridge
8	M1M	Scarborough	Cliffcrest,Cliffside,Scarborough Village West
9	M1N	Scarborough	Birch Cliff,Cliffside West
10	M1P	Scarborough	Dorset Park,Scarborough Town Centre,Wexford He...

Then the geo-coordinates of the neighbourhoods sorted from the Wikipedia page are found from source 5 based on their postal codes. Part of the results is shown as below:

	PostalCode	Latitude	Longitude	
0	M1B	43.806686	-79.194353	
1	M1C	43.784535	-79.160497	
2	M1E	43.763573	-79.188711	
3	M1G	43.770992	-79.216917	
4	M1H	43.773136	-79.239476	

	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

The neighbourhoods are plotted onto the map of Toronto for better visualization.



Other python libraries like Pandas and Numpy are used wherever needed. More details about the data collection process can be seen in the Jupiter notebook.

Data mining

With the location of each neighbourhood, it is no ready to make request to Foursquare to get the business information with a searching radius of 500 m. Part of the information obtained is shown as below:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Rouge,Malvern	43.806686	-79.194353	Wendy's	43.807448	-79.199056	Fast Food Restaurant
1	Highland Creek,Rouge Hill,Port Union	43.784535	-79.160497	Royal Canadian Legion	43.782533	-79.163085	Bar
2	Guildwood,Morningside,West Hill	43.763573	-79.188711	Swiss Chalet Rotisserie & Grill	43.767697	-79.189914	Pizza Place
3	Guildwood,Morningside,West Hill	43.763573	-79.188711	G & G Electronics	43.765309	-79.191537	Electronics Store
4	Guildwood,Morningside,West Hill	43.763573	-79.188711	Marina Spa	43.766000	-79.191000	Spa

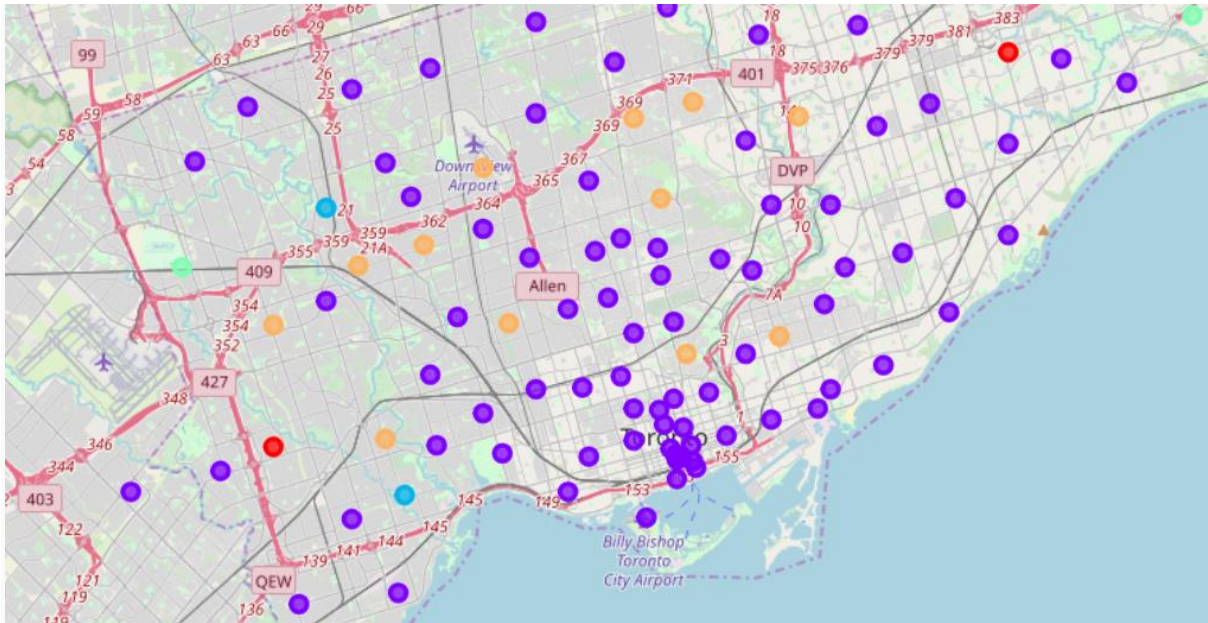
In total, 274 unique venue categories are found in Toronto, and the top venues for each neighbourhood are found and listed. An example is as below:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Adelaide,King,Richmond	Coffee Shop	Café	Thai Restaurant	American Restaurant	Steakhouse
1	Agincourt	Lounge	Breakfast Spot	Sandwich Place	Skating Rink	Dim Sum Restaurant
2	Agincourt North,L'Amoreaux East,Milliken,Steel...	Park	Coffee Shop	Playground	Dog Run	Deli / Bodega
3	Albion Gardens,Beaumont Heights,Humbergate,Jam...	Grocery Store	Beer Store	Fried Chicken Joint	Fast Food Restaurant	Pharmacy
4	Alderwood,Long Branch	Pizza Place	Gym	Coffee Shop	Athletics & Sports	Skating Rink

Clustering

To find the optimal area in Toronto to develop a specific business, one must first know how many areas there are and what the business environment is in each area. Hence, the city area should be divided into neighbourhoods with characteristics which are distinct to each other to some extents. To do so, clustering, one of the machine learning techniques will be useful. Hence, for this project, k-means clustering is used to split up Toronto city into 5 neighbourhoods, labelled as cluster 0 to 4 as shown in the figure below.

	PostalCode	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
0	M1B	Scarborough	Rouge,Malvern	43.806686	-79.194353	4	Fast Food Restaurant	Women's Store	Doner Restaurant
1	M1C	Scarborough	Highland Creek,Rouge Hill,Port Union	43.784535	-79.160497	3	Bar	Women's Store	Donut Shop
2	M1E	Scarborough	Guildwood,Morningside,West Hill	43.763573	-79.188711	1	Breakfast Spot	Medical Center	Rental Car Location
3	M1G	Scarborough	Woburn	43.770992	-79.216917	1	Coffee Shop	Insurance Office	Korean Restaurant
4	M1H	Scarborough	Cedarbrae	43.773136	-79.239476	0	Hakka Restaurant	Fried Chicken Joint	Thai Restaurant



Cluster 0: red

	Borough	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
4	Scarborough	0	Hakka Restaurant	Fried Chicken Joint	Thai Restaurant
94	Etobicoke	0	Bank	Golf Course	Donut Shop

Cluster 1: purple (showing only part of the table)

	Borough	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
2	Scarborough	1	Breakfast Spot	Medical Center	Rental Car Location
3	Scarborough	1	Coffee Shop	Insurance Office	Korean Restaurant
5	Scarborough	1	Playground	Convenience Store	Women's Store
6	Scarborough	1	Bus Station	Department Store	Coffee Shop
7	Scarborough	1	Bakery	Bus Line	Soccer Field

Cluster 2: blue

	Borough	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
91	Etobicoke	2	Baseball Field	Women's Store	Donut Shop
97	North York	2	Construction & Landscaping	Baseball Field	Drugstore

Cluster 3: green

	Borough	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
1	Scarborough	3	Bar	Women's Store	Donut Shop
102	Etobicoke	3	Rental Car Location	Bar	Drugstore

Cluster 4: orange (showing only part of the table)

	Borough	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
0	Scarborough	4	Fast Food Restaurant	Women's Store	Doner Restaurant
14	Scarborough	4	Park	Coffee Shop	Playground
20	North York	4	Park	Cafeteria	Women's Store
23	North York	4	Bank	Park	Electronics Store
25	North York	4	Park	Fast Food Restaurant	Food & Drink Shop

These clustering results are then compared against the findings on the three websites. Excel and python coding are used wherever necessary. The results and discussion are presented in later sections.

Results

Based on the most common venues found in the tables above, the three most popular business types (venues with highest frequency in the table column “1st most common venue”. If there are less than 3 entries in the table, “2nd most common venue” is considered) are summarized in the table below for each cluster. Note that for normalised comparison, venue names are simplified into a type. For example, all dining places are presented as restaurant in the table below.

Cluster	1 st popular business	2 nd popular business	3 rd popular business
0	Restaurant /Bank	Restaurant /Golf course	Restaurant/Donut shop
1	Coffee shop	Cafe	Grocery store
2	Baseball field/construction & landscaping	Baseball field/Women's store	Drug store/Donut shop
3	Bar/rental car location	Women's store/bar	Drug store/Donut shop
4	Park	Restaurant /construction & landscaping/bank	Women's store

Discussion

Recall the ranking in the three websites.

Rank	Profitableventure	OPSTART	Business Insider
1	Gluten-free restaurants and bakeries	Real Estate	Alcohol delivery service
2	Daily-care and Babysitting Services	Pet Services	Café dinning
3	Patient Advocacy	Financial Consulting	Fitbit

By comparing the results of clustering and the findings on the websites, the conclusion can be drawn as follows.

Rank	Profitableventure	Recommend developing zone in Toronto	OPSTART	Recommend developing zone in Toronto	Business Insider	Recommend developing zone in Toronto
1	Gluten-free restaurants and bakeries	Cluster 0	Real Estate	Cluster 2	Alcohol delivery service	Cluster 3
2	Daily-care and Babysitting Services	Cluster 1	Pet Services	Cluster 1*	Café dinning	Cluster 1
3	Patient Advocacy	Cluster 2 or 3	Financial Consulting	Cluster 0	Fitbit	Cluster 4

*Pet service related venues are not shown in the result table, but it relatively more popular in cluster 1 compared to other clusters based on the full table.

Conclusion

For the most popular business types ranked on these three websites, namely Profitableventure, OPSTART and Business Insider, the recommended developing areas in Toronto are summarized in the table above. Note that the conclusion of this paper is purely based on appearing frequency of various business types, without considering other factors like business competition or financing limit.

It can only serve as a preliminary research. A more complex study is needed to provide high-quality recommendations to business developers.