

#### INTRODUCTION

Background: in this assignment, I will develop a data science tool to support entrepreneurs who want to start a new horeca business (hotel, restaurant catering) with information about:

- Which neighbourhood in Toronto is the most suitable to put their business location i.E. Shops/restaurants.
- Compare their business to other businesses that can potentially create a good niche for their business to grow.



#### DATA PREPARATION

For this assignment, I will use location data from foursquare to develop insights to support entrepreneurs to choose a neighbourhood for their new horeca business. The data will focus on the city of toronto, canada. The following categories of data will be used:

- Neighbourhood name
- Venue name
- Location (lat and lon data)
- Business types

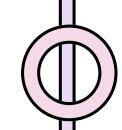
Standard data cleansing methods will be applied, including removing nas, removing 'not specified' entities, merge similar entries.



#### RESULTS

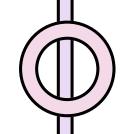
THE NEIGHBOURHOODS OF TORONTO ARE SCRAPPED FROM WIKIPEDIA AND CLEANSED AND SAVE IN A DATAFRAME:

	Postal Code	Borough	Neighbourhood	
0	M1B	Scarborough	Malvern, Rouge	
1	M1C	Scarborough	Rouge Hill, Port Union, Highland Creek	
2	M1E	Scarborough	Guildwood, Morningside, West Hill	
3	M1G	Scarborough	Woburn	
4	M1H	Scarborough	Cedarbrae	
5	M1J	Scarborough	Scarborough Village	
6	M1K	Scarborough	Kennedy Park, Ionview, East Birchmount Park	
7	M1L	Scarborough	Golden Mile, Clairlea, Oakridge	
8	M1M	Scarborough	Cliffside, Cliffcrest, Scarborough Village West	
9	M1N	Scarborough	Birch Cliff, Cliffside West	
10	M1P	Scarborough	Dorset Park, Wexford Heights, Scarborough Town	
11	M1R	Scarborough	Wexford, Maryvale	
12	M1S	Scarborough	Agincourt	
13	M1T	Scarborough	Clarks Corners, Tam O'Shanter, Sullivan	
14	M1V	Scarborough	Milliken, Agincourt North, Steeles East, L'Amo	
15	M1W	Scarborough	Steeles West, L'Amoreaux West	
16	M1X	Scarborough	Upper Rouge	
17	M2H	North York	Hillcrest Village	
18	M2J	North York	Fairview, Henry Farm, Oriole	
19	M2K	North York	Bayview Village	
20	M2L	North York	York Mills, Silver Hills	



## RESULTS

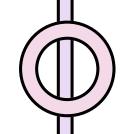
LOCATION DATA (LAT AND LON) ARE RETRIEVED ONLINE, AND ADDED TO THE DATAFRAME:



	Postal Code	Borough	Neighbourhood	Latitude	Longitude
0	M1B	Scarborough	Malvern, Rouge	43.806686	-79.194353
1	M1C	Scarborough	Rouge Hill, Port Union, Highland Creek	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

## RESULTS

LOCATION DATA (LAT AND LON) ARE RETRIEVED ONLINE, AND ADDED TO THE DATAFRAME:



	Postal Code	Borough	Neighbourhood	Latitude	Longitude
0	M1B	Scarborough	Malvern, Rouge	43.806686	-79.194353
1	M1C	Scarborough	Rouge Hill, Port Union, Highland Creek	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

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most ( Venue
0 4	Aginolunt	ge ge	Latin American Restaurant	Clothing Store	Breakfast Spot	Skating Ri
1	Alderw		Gym	Coffee Shop	Skating Rink	Sandwich
2	R E S U L	T S	Coffee Shop	Diner	Supermarket	Sushi Res
3	7		Japanese Restaurant	Bank	Chinese Restaurant	Women's
4	K-MEANS CLUSTER ARE PREFORMED.	ING W	Coffee Shop	Sandwich Place	Liquor Store	Thai Resta
5	ARE PREFORMED, THE MOST COMMO VENUES ARE LISTE	Cocktail Bar	Cheese Shop	Restaurant	Beer Bar	
6	EACH NEIGHBOURI	General Entertainment	Café	Skating Rink	Dim Sum I	
7	THIS INFORMATIO HELP ENTREPRENE DECIDE WHICH	Nightclub	Breakfast Spot	Coffee Shop	Performing	
8	NEIGHBOURHOOD SUITABLE TO STAF THEIR BUSINESS.	Yoga Studio	Spa	Burrito Place	Skate Parl	
9	CN Tow Lands, Hal.		Airport Service	Airport Terminal	Boat or Ferry	Boutique
10	Caledonia-Fairban.		Women's Store	Pool	Doner Restaurant	Dim Sum I
11	Canada Post Gateway Processing Centre	Coffee Shop	Hotel	Gym	Gas Station	Food Truc
12	Cedarbrae	Hakka Restaurant	Bakery	Lounge	Athletics & Sports	Bank
13	Central Bay Street	Coffee Shop	Café	Sandwich Place	Italian Restaurant	Bubble Tea

# DISCUSSION AND CONCLUSION

The developed tool based on data scrapping and k-means clustering could help to support decision making of entrepreneurs who want to start a business in a new neighbourhood by pointing out what kind of business is already available in that neighbourhood, and show whether they can find synergies with existing businesses to create a niche.

Further work is suggested to develop a automatic recommending system, where entrepreneurs can provide in a table their business information using keywords. This table is then used to compare against the table developed for the neighbourhood, and result in a list of top 5 most suitable neighbourhoods.



