LOCATION LOCATION LOCATION

Clustering and Analysing Location Data in Toronto

Justin Charette

1. INTRODUCTION

THE PROBLEM

Starting a business is a risky endeavour, and one that involves a significant investment of time and resources. That investment is especially significant for a small business owner in a high-turnover industry. It is therefore advantageous to minimize potential risk by choosing a suitable location.

In this scenario, an entrepreneur wants to open a cafe somewhere in Toronto. The goal is to find a location in which a cafe is likely to be successful, which will require identifying neighbourhoods in which one can find similar types of businesses, such as bars, restaurants, shops, etc.; but preferably one that is not already saturated with cafes.

We can tackle this problem using geospatial data and machine learning techniques.

2. DATA

For this problem I will be using geospatial information on Toronto, which will include:

- -Borough name
- -Neighbourhood name
- -Neighbourhood postal code
- -Latitude and longitude associated with the neighbourhood postal code

I will also use the Foursquare API to obtain location information on cafes and other businesses in each neighbourhood. The relevant features will include:

- -Venue name
- -Venue category
- -Latitude and Longitude

This information can then be used to cluster the neighbourhoods using K-means clustering based on the types of businesses common to the different neighbourhoods.

3. METHODOLOGY

We begin by acquiring the list of neighbourhoods and corresponding postal codes for Toronto from Wikipedia, then matching the postal codes to GPS coordinates. We can plot these points on a map of Toronto to visualize the different neighbourhoods.

Using the Foursquare API, we can get information on the types of venues in the neighbourhoods we have mapped. We can then find the most common types of venues for each neighbourhood, and group them using K-Means Clustering.

We will be able to categorize the different clusters based on their most common types of venues, and and determine which areas might be most suitable.

We can also use the foursquare API to search venues by category ID to get the location and count of coffee shops and cafes for each neighbourhood and discard the areas that have the highest density.

The major steps are outlined and explained in detail below.

Plotting the Neighbourhoods

After importing and installing all of the necessary packages and libraries, the first step is acquiring the neighbourhood information. I used Beautiful Soup to scrape this information from Wikipedia (https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M) and read it into a pandas dataframe shown in Fig. 3.1.

[2]:

	PostalCode	Borough	Neighborhood
0	M3A	North York	Parkwoods
1	M4A	North York	Victoria Village
2	M5A	Downtown Toronto	Regent Park, Harbourfront
3	M6A	North York	Lawrence Manor, Lawrence Heights
4	M7A	Queen's Park	Ontario Provincial Government

Fig. 3.1

Next, I read in a .csv file of the postal codes and coordinates into another pandas dataframe, I then merged these using an inner join to create a single dataframe with all of the necessary neighbourhood data, Fig. 3.2.

[6]:
```
df_merged = pd.merge(df, df_coords, on='PostalCode', how='inner')
df_merged.head()
```

[6]:

	PostalCode	Borough	Neighborhood	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

Fig. 3.2

Next, I used Folium to create a map of Toronto and plot the latitude and longitude of each neighbourhood as a blue circle, shown in Fig 3.3.

Fig. 3.3

Searching Venues

Now we can begin to explore the different neighbourhoods using the Foursquare API. For this I used a function that returned a list of venues within a 500 meter radius around the latitude and longitude points of each neighbourhood.

The new dataframe, shown in Fig. 3.4 contains the relevant information about the venues retrieved from the Foursquare API and the neighbourhood information.

```
[12]: print(toronto_venues.shape)
      toronto_venues.head()
```

```
(2146, 7)
```

[12]:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Malvern, Rouge	43.806686	-79.194353	Wendy's	43.807448	-79.199056	Fast Food Restaurant
1	Rouge Hill, Port Union, Highland Creek	43.784535	-79.160497	RIGHT WAY TO GOLF	43.785177	-79.161108	Golf Course
2	Rouge Hill, Port Union, Highland Creek	43.784535	-79.160497	Royal Canadian Legion	43.782533	-79.163085	Bar
3	Guildwood, Morningside, West Hill	43.763573	-79.188711	RBC Royal Bank	43.766790	-79.191151	Bank
4	Guildwood, Morningside, West Hill	43.763573	-79.188711	G & G Electronics	43.765309	-79.191537	Electronics Store

Fig. 3.4

One-Hot Encoding

We are going to cluster our neighbourhoods according the types of businesses present. To do this, we must determine the frequency of each venue category. First, we will group the venues dataframe by neighbourhood and obtain the count, then we will perform one-hot encoding. This process will assign a binary numèric “dummy” variable to each unique venue category.

Then we will find the mean of the values in the one-hot dataframe to determine the frequency of each venue category by neighbourhood, Fig. 3.5.

	Neighborhood	Accessories Store	Afghan Restaurant	Airport	Airport Food Court	Airport Gate	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	Antique Shop	Aquarium	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant
0	Agincourt	0.000000	0.000000	0.0000	0.0000	0.0000	0.0000	0.000	0.000	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000
1	Alderwood, Long Branch	0.000000	0.000000	0.0000	0.0000	0.0000	0.0000	0.000	0.000	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000
2	Bathurst Manor, Wilson Heights, Downsview North	0.000000	0.000000	0.0000	0.0000	0.0000	0.0000	0.000	0.000	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000
3	Bayview Village	0.000000	0.000000	0.0000	0.0000	0.0000	0.0000	0.000	0.000	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000
4	Bedford Park, Lawrence Manor East	0.000000	0.000000	0.0000	0.0000	0.0000	0.0000	0.000	0.000	0.038462	0.000000	0.00	0.000000	0.000000	0.000000	0.000000
5	Berczy Park	0.000000	0.000000	0.0000	0.0000	0.0000	0.0000	0.000	0.000	0.016667	0.000000	0.00	0.016667	0.000000	0.000000	0.000000
6	Birch Cliff, Cliffside West	0.000000	0.000000	0.0000	0.0000	0.0000	0.0000	0.000	0.000	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000
7	Brockton, Parkdale Village, Exhibition Place	0.000000	0.000000	0.0000	0.0000	0.0000	0.0000	0.000	0.000	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000

Fig. 3.5

Finding the Most Common Venues

We will sort the venues in descending order and then create a new dataframe featuring the top 10 most common venue types for each neighbourhood, Fig. 3.6

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Agincourt	Lounge	Breakfast Spot	Skating Rink	Latin American Restaurant	Clothing Store	Yoga Studio	Drugstore	Dive Bar	Dog Run	Doner Restaurant
1	Alderwood, Long Branch	Pizza Place	Pharmacy	Skating Rink	Dance Studio	Coffee Shop	Pub	Sandwich Place	Gym	Airport Terminal	Farmers Market
2	Bathurst Manor, Wilson Heights, Downsview North	Coffee Shop	Bank	Frozen Yogurt Shop	Bridal Shop	Sandwich Place	Diner	Restaurant	Middle Eastern Restaurant	Supermarket	Intersection
3	Bayview Village	Café	Japanese Restaurant	Bank	Chinese Restaurant	Discount Store	Distribution Center	Dive Bar	Dog Run	Doner Restaurant	Yoga Studio
4	Bedford Park, Lawrence Manor East	Coffee Shop	Sandwich Place	Italian Restaurant	Grocery Store	Cupcake Shop	Indian Restaurant	Pub	Sushi Restaurant	Restaurant	Café
5	Berczy Park	Coffee Shop	Bakery	Beer Bar	Farmers Market	Cocktail Bar	Restaurant	Pharmacy	Seafood Restaurant	Cheese Shop	Breakfast Spot
6	Birch Cliff, Cliffside West	College Stadium	Skating Rink	General Entertainment	Café	Electronics Store	Eastern European Restaurant	Ethiopian Restaurant	Dumpling Restaurant	Drugstore	Dim Sum Restaurant
7	Brockton, Parkdale Village, Exhibition Place	Café	Breakfast Spot	Nightclub	Coffee Shop	Grocery Store	Burrito Place	Italian Restaurant	Bar	Bakery	Intersection

Fig. 3.6

Clustering

As we will be using K-Means clustering, we must first determine the optimal number of clusters (k).
Performing the elbow method returns a value of k = 6. Fig. 3.7.

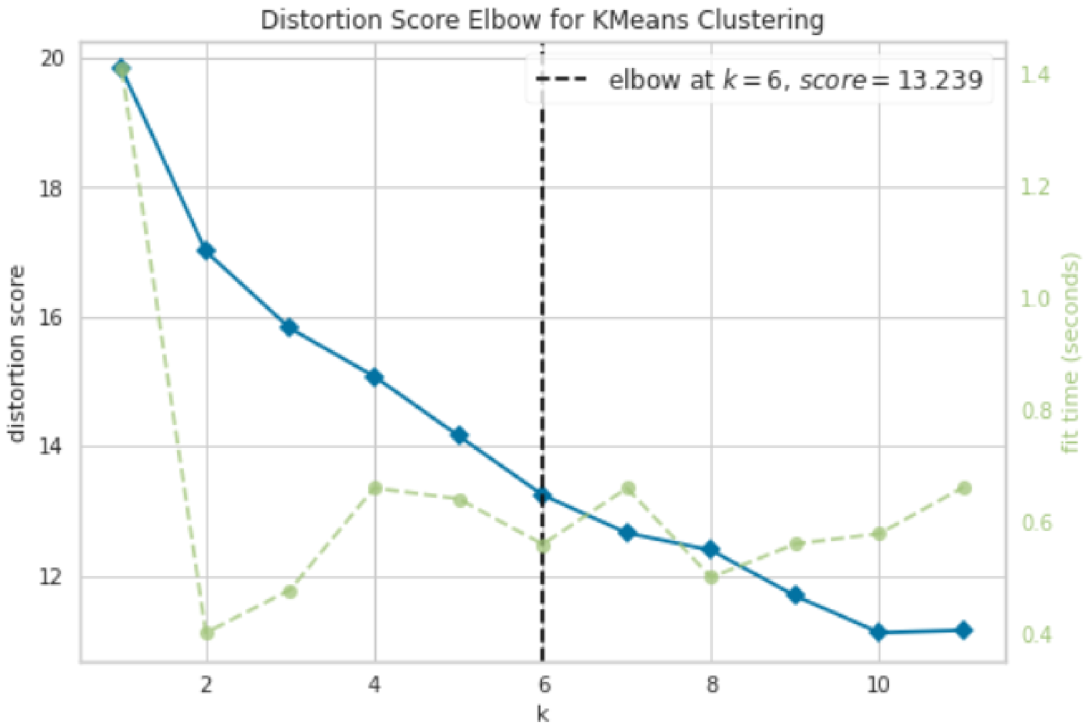

Fig. 3.7

After creating a KMeans object, fitting the model, adding cluster labels to the dataframe with the most common venues, we then merge the dataframe with the dataframe containing the neighbourhood data, Fig. 3.8.

Borough	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
carborough	Malvern, Rouge	43.806686	-79.194353	4	Fast Food Restaurant	Donut Shop	Diner	Discount Store	Distribution Center	Dive Bar	Dog Run	Doner Restaurant	Yoga Studio	Dim Sum Restaurant
carborough	Rouge Hill, Port Union, Highland Creek	43.784535	-79.160497	1	History Museum	Bar	Yoga Studio	Drugstore	Distribution Center	Dive Bar	Dog Run	Doner Restaurant	Donut Shop	Dumpling Restaurant
carborough	Guildwood, Morningside, West Hill	43.763573	-79.188711	1	Mexican Restaurant	Donut Shop	Restaurant	Rental Car Location	Bank	Intersection	Medical Center	Breakfast Spot	Electronics Store	Dumpling Restaurant
carborough	Woburn	43.770992	-79.216917	1	Coffee Shop	Other Repair Shop	Convenience Store	Korean BBQ Restaurant	Yoga Studio	Donut Shop	Distribution Center	Dive Bar	Dog Run	Doner Restaurant
carborough	Cedarbrae	43.773136	-79.239476	1	Hakka Restaurant	Bakery	Thai Restaurant	Caribbean Restaurant	Athletics & Sports	Gas Station	Bank	Fried Chicken Joint	Dog Run	Dive Bar

Fig. 3.8

Folium can be used again to generate another map, and each neighbourhood marker is assigned a diferent colour depending on which cluster it belongs to, Fig. 3.9.

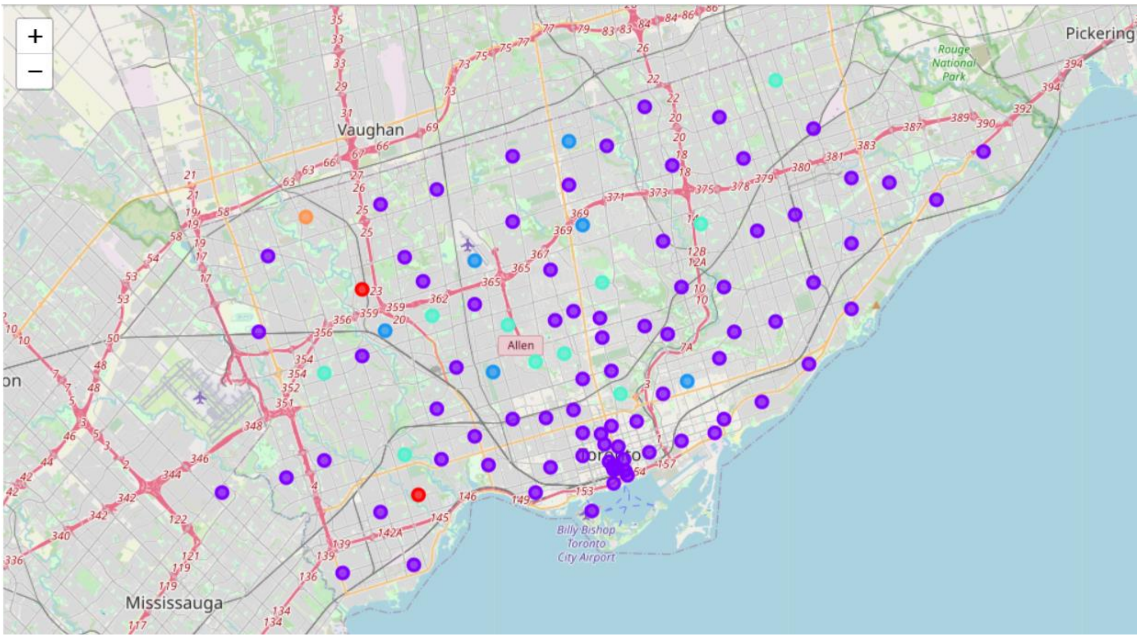

Fig. 3.9

4. RESULTS

After generating the clusters, we can look more closely at each one and classify each in terms of what kinds of venues they are composed of. We can view the information in the merged dataframe by separating out each cluster label.

Cluster 1:

	Borough	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
91	Etobicoke	0	Pool	Baseball Field	Yoga Studio	Donut Shop	Discount Store	Distribution Center	Dive Bar	Dog Run	Doner Restaurant	Drugstore
97	North York	0	Baseball Field	Yoga Studio	Discount Store	Distribution Center	Dive Bar	Dog Run	Doner Restaurant	Donut Shop	Drugstore	Fast Food Restaurant

Though this is a small cluster, based on the position of “pool”, “baseball field”, and “yoga Studio” in the top most common venues, I will classify this cluster as recreation.

Cluster 2 (As the full dataframe is quite large, I have only included portions below. It can be seen in it’s entirety in the notebook):

	Borough	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
1	Scarborough	1	History Museum	Bar	Yoga Studio	Drugstore	Distribution Center	Dive Bar	Dog Run	Doner Restaurant	Donut Shop	Dumpling Restaurant
2	Scarborough	1	Mexican Restaurant	Donut Shop	Restaurant	Rental Car Location	Bank	Intersection	Medical Center	Breakfast Spot	Electronics Store	Dumpling Restaurant
3	Scarborough	1	Coffee Shop	Other Repair Shop	Convenience Store	Korean BBQ Restaurant	Yoga Studio	Donut Shop	Distribution Center	Dive Bar	Dog Run	Doner Restaurant
4	Scarborough	1	Hakka Restaurant	Bakery	Thai Restaurant	Caribbean Restaurant	Athletics & Sports	Gas Station	Bank	Fried Chicken Joint	Dog Run	Dive Bar
5	Scarborough	1	Women's Store	Playground	Drugstore	Discount Store	Distribution Center	Dive Bar	Dog Run	Doner Restaurant	Donut Shop	Dumpling Restaurant
6	Scarborough	1	Train Station	Department Store	Coffee Shop	Hobby Shop	Chinese Restaurant	Yoga Studio	Discount Store	Distribution Center	Dive Bar	Dog Run
7	Scarborough	1	Bakery	Metro Station	Intersection	Ice Cream Shop	Bus Line	Soccer Field	Park	Eastern European Restaurant	Dumpling Restaurant	Drugstore
8	Scarborough	1	American Restaurant	Motel	Yoga Studio	Discount Store	Distribution Center	Dive Bar	Dog Run	Doner Restaurant	Donut Shop	Drugstore
9	Scarborough	1	College Stadium	Skating Rink	General Entertainment	Café	Electronics Store	Eastern European Restaurant	Ethiopian Restaurant	Dumpling Restaurant	Drugstore	Dim Sum Restaurant
49	Central Toronto	1	Coffee Shop	American Restaurant	Light Rail Station	Restaurant	Supermarket	Bank	Sushi Restaurant	Bagel Shop	Pub	Fried Chicken Joint
51	Downtown Toronto	1	Park	Café	Bakery	Coffee Shop	Restaurant	Pub	Italian Restaurant	Pizza Place	Convenience Store	Sandwich Place
52	Downtown Toronto	1	Coffee Shop	Japanese Restaurant	Sushi Restaurant	Gay Bar	Restaurant	Yoga Studio	Café	Fast Food Restaurant	Hotel	Men's Store
53	Downtown Toronto	1	Coffee Shop	Café	Park	Pub	Bakery	Theater	Farmers Market	Restaurant	Event Space	Spa
54	Downtown Toronto	1	Coffee Shop	Clothing Store	Café	Cosmetics Shop	Japanese Restaurant	Bubble Tea Shop	Pizza Place	Fast Food Restaurant	Theater	Electronics Store
55	Downtown Toronto	1	Coffee Shop	Café	Restaurant	Cocktail Bar	Cosmetics Shop	Italian Restaurant	Seafood Restaurant	Clothing Store	Park	Moroccan Restaurant
56	Downtown Toronto	1	Coffee Shop	Bakery	Beer Bar	Farmers Market	Cocktail Bar	Restaurant	Pharmacy	Seafood Restaurant	Cheese Shop	Breakfast Spot
57	Downtown Toronto	1	Coffee Shop	Café	Sandwich Place	Restaurant	Bank	Bubble Tea Shop	Salad Place	Burger Joint	Italian Restaurant	Japanese Restaurant
58	Downtown Toronto	1	Coffee Shop	Café	Hotel	Restaurant	Gym	Thai Restaurant	Sushi Restaurant	Clothing Store	Bookstore	Salad Place
59	Downtown Toronto	1	Coffee Shop	Aquarium	Hotel	Café	Fried Chicken Joint	Sporting Goods Shop	Brewery	Restaurant	Scenic Lookout	Bar
78	West Toronto	1	Café	Breakfast Spot	Nightclub	Coffee Shop	Grocery Store	Burrito Place	Italian Restaurant	Bar	Bakery	Intersection
80	York	1	Discount Store	Sandwich Place	Skating Rink	Bar	Donut Shop	Distribution Center	Dive Bar	Dog Run	Doner Restaurant	Yoga Studio
81	York	1	Pizza Place	Grocery Store	Bus Line	Brewery	Donut Shop	Discount Store	Distribution Center	Dive Bar	Dog Run	Doner Restaurant
82	West Toronto	1	Café	Mexican Restaurant	Thai Restaurant	Grocery Store	Speakeasy	Fried Chicken Joint	Bakery	Bar	Italian Restaurant	Flea Market
83	West Toronto	1	Breakfast Spot	Gift Shop	Restaurant	Cuban Restaurant	Bar	Bookstore	Dessert Shop	Movie Theater	Dog Run	Italian Restaurant
84	West Toronto	1	Coffee Shop	Café	Italian Restaurant	Restaurant	Pub	Pizza Place	Sushi Restaurant	Fish & Chips Shop	Falafel Restaurant	Spa
85	Queen's Park	1	Coffee Shop	Café	Yoga Studio	Burger Joint	Mexican Restaurant	Bar	Bank	Sushi Restaurant	Diner	Beer Bar
86	Mississauga	1	Coffee Shop	Hotel	Gym	American Restaurant	Fried Chicken Joint	Gas Station	Sandwich Place	Burrito Place	Mediterranean Restaurant	Intersection
87	East Toronto Business	1	Park	Garden Center	Skate Park	Farmers Market	Fast Food Restaurant	Light Rail Station	Burrito Place	Butcher	Café	Restaurant
88	Etobicoke	1	Coffee Shop	Café	Pharmacy	Bakery	Fast Food Restaurant	Liquor Store	Pet Store	Pizza Place	Restaurant	Gym

Examining this cluster, we can see many neighbourhoods in which "cafe" or "coffee shop" appears high in the top 10 venues. Looking closer, we can also see a variety of restaurants, shops, and other businesses, giving the impression that this cluster represents diverse commercial neighbourhoods; i.e. mixed-use neighbourhoods that receive high levels of foot traffic and are likely to draw in people from other parts of the city, This type of neighbourhood profile seems like the ideal sort of area to open a new cafe. This cluster is represented on the map in Fig. 3.9 by purple markers.

Cluster 3:

	Borough	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
21	North York	2	Park	Yoga Studio	Drugstore	Discount Store	Distribution Center	Dive Bar	Dog Run	Doner Restaurant	Donut Shop	Dumpling Restaurant
23	North York	2	Park	Convenience Store	Yoga Studio	Drugstore	Discount Store	Distribution Center	Dive Bar	Dog Run	Doner Restaurant	Donut Shop
30	North York	2	Airport	Park	Yoga Studio	Drugstore	Discount Store	Distribution Center	Dive Bar	Dog Run	Doner Restaurant	Donut Shop
40	East York/East Toronto	2	Park	Convenience Store	Yoga Studio	Drugstore	Discount Store	Distribution Center	Dive Bar	Dog Run	Doner Restaurant	Donut Shop
74	York	2	Park	Women's Store	Pool	Yoga Studio	Donut Shop	Diner	Discount Store	Distribution Center	Dive Bar	Dog Run
98	York	2	Park	Yoga Studio	Drugstore	Discount Store	Distribution Center	Dive Bar	Dog Run	Doner Restaurant	Donut Shop	Dumpling Restaurant

Cluster 4:

	Borough	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
14	Scarborough	3	Playground	Park	Intersection	Yoga Studio	Doner Restaurant	Diner	Discount Store	Distribution Center	Dive Bar	Dog Run
25	North York	3	Fast Food Restaurant	Park	Food & Drink Shop	Ethiopian Restaurant	Electronics Store	Event Space	Eastern European Restaurant	Dumpling Restaurant	Dim Sum Restaurant	Drugstore
44	Central Toronto	3	Business Service	Park	Bus Line	Swim School	Yoga Studio	Doner Restaurant	Distribution Center	Dive Bar	Dog Run	Donut Shop
50	Downtown Toronto	3	Park	Playground	Trail	Yoga Studio	Doner Restaurant	Diner	Discount Store	Distribution Center	Dive Bar	Dog Run
64	Central Toronto	3	Park	Sushi Restaurant	Trail	Jewelry Store	Yoga Studio	Doner Restaurant	Discount Store	Distribution Center	Dive Bar	Dog Run
72	North York	3	Pizza Place	Park	Bakery	Japanese Restaurant	Doner Restaurant	Discount Store	Distribution Center	Dive Bar	Dog Run	Donut Shop
73	York	3	Field	Park	Hockey Arena	Trail	Doner Restaurant	Diner	Discount Store	Distribution Center	Dive Bar	Dog Run
79	North York	3	Park	Bakery	Construction & Landscaping	Trail	Yoga Studio	Doner Restaurant	Discount Store	Distribution Center	Dive Bar	Dog Run
90	Etobicoke	3	River	Park	Pool	Yoga Studio	Doner Restaurant	Diner	Discount Store	Distribution Center	Dive Bar	Dog Run

Both clusters 3 and 4 appear similar, they both feature venues like parks and yoga studios prominently in the top venues. They also both feature venues like dive bars and dog runs further to the right side of the dataframe. Cluster 3, however, features more venues like drugstores and convenience stores in the top 5 positions, on the left half of the dataframe, and so I think cluster 3 would best categorized as a primarily residential cluster. Cluster 4, on the other hand, features a lot more parks on the left half of the dataframe than cluster 3, plus we also see other venues like playgrounds, trails, hockey arenas and swim studios. Based on this difference, I think cluster 4 would be better characterized as recreational/activity oriented.

Cluster 5:

	Borough	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Scarborough	4	Fast Food Restaurant	Donut Shop	Diner	Discount Store	Distribution Center	Dive Bar	Dog Run	Doner Restaurant	Yoga Studio	Dim Sum Restaurant

Cluster 6:

	Borough	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
96	North York	5	Pizza Place	Donut Shop	Diner	Discount Store	Distribution Center	Dive Bar	Dog Run	Doner Restaurant	Drugstore	Harbor / Marina

Clusters 5 and 6 are quite similar to each other as well. It's possible the model is over-fitting; however, múltiple iterations of the K-means algorithm produced similar single-unit clusters. Most importantly, though, is that the cluster that we are interested in for this problem (that is cluster 2, the "cafè cluster") remained relatively consistent in size and composition of its most common venues. This consistency was present even in iterations in which I used different values of K.

5. DISCUSSION

It will be useful narrow down our results further by looking at the count of cafes. We can do this by making another request to the API using the category ID for cafès and coffee shops and applying it to the neighbourhoods in the cluster 2, Fig. 5.1.

	PostalCode	Borough	Neighborhood	Latitude	Longitude	Number of Cafes
1	M1C	Scarborough	Rouge Hill, Port Union, Highland Creek	43.784535	-79.160497	0
2	M1E	Scarborough	Guildwood, Morningside, West Hill	43.763573	-79.188711	0
3	M1G	Scarborough	Woburn	43.770992	-79.216917	0
4	M1H	Scarborough	Cedarbrae	43.773136	-79.239476	2
5	M1J	Scarborough	Scarborough Village	43.744734	-79.239476	0
6	M1K	Scarborough	Kennedy Park, Ionview, East Birchmount Park	43.727929	-79.262029	0
7	M1L	Scarborough	Golden Mile, Clairlea, Oakridge	43.711112	-79.284577	1
8	M1M	Scarborough	Cliffside, Cliffcrest, Scarborough Village West	43.716316	-79.239476	0
9	M1N	Scarborough	Birch Cliff, Cliffside West	43.692657	-79.264848	1
10	M1P	Scarborough	Dorset Park, Wexford Heights, Scarborough Town...	43.757410	-79.273304	0
11	M1R	Scarborough	Wexford, Maryvale	43.750072	-79.295849	0
12	M1S	Scarborough	Agincourt	43.794200	-79.262029	0
13	M1T	Scarborough	Clarks Corners, Tam O'Shanter, Sullivan	43.781638	-79.304302	0
15	M1W	Scarborough	Steeles West, L'Amoreaux West	43.799525	-79.318389	1
17	M2H	North York	Hillcrest Village	43.803762	-79.363452	0
18	M2J	North York	Fairview, Henry Farm, Oriole	43.778517	-79.346556	8
19	M2K	North York	Bayview Village	43.786947	-79.385975	1
22	M2N	North York	Willowdale South	43.770120	-79.408493	12

Fig. 5.1

The results, if we examine the full dataframe, show that the different neighbourhoods contain a wide range of number of cafes, from 50 to 0, with a mean of 10.3, and a median of 2.5.

Many of the neighbourhoods in the Downtown borough have a high count. It can be concluded that cafes are a very popular venue type in these areas; however, it also means that a new business will face considerable competition, and the risk of turnover will also be higher.

On the other hand, there are also many neighbourhoods with very few cafes, or none, retrieved by the API. These areas may appear to be ripe opportunities, but it may also be the case that there is simply little demand for such venues, and therefore might prove to be less advantageous.

The best option then might be to look at the neighbourhoods in which the count is closer to the mean of 10.5. The chart below shows the count of each neighbourhood in cluster 2 sorted in descending order, Fig. 5.2.

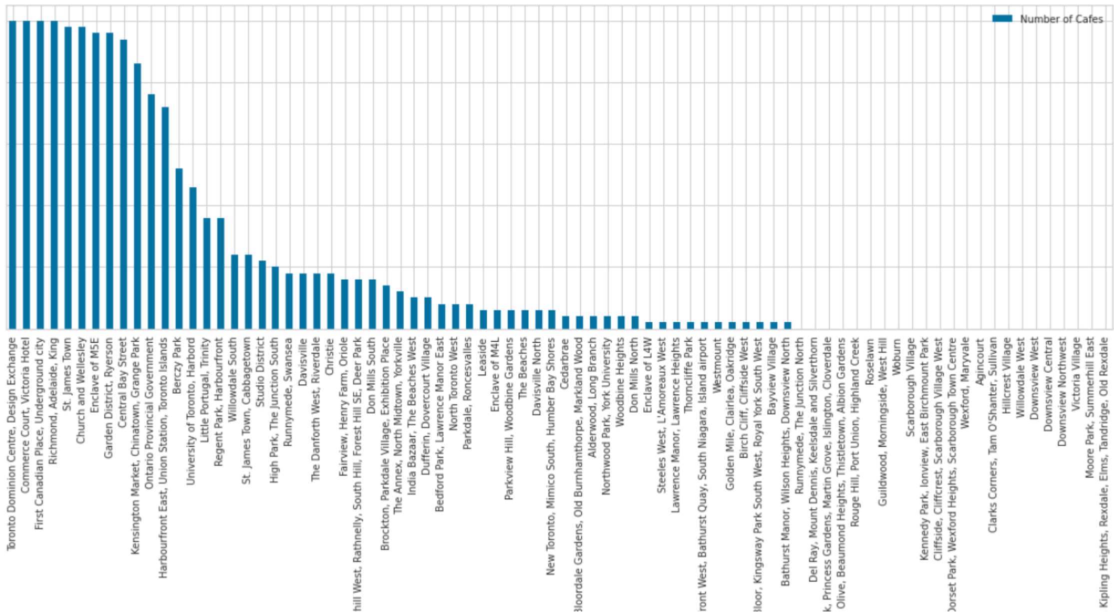

Fig. 5.2

6. CONCLUSION

There are many factors that contribute to the success of a business, and there are many associated risks, especially for an independent without a recognized brand and established customer base. It is therefore important to choose a location that has the best chance of maximizing success. An important thing to conisder, then, is choosing a location where there is a market for your business, in this case, a cafe. Using K-means clustering with our location data acquired from the Foursquare API, we can identify neighbourhoods of Toronto in which cafès are a viable business. These neighbourhoods also commonly have a variety of different types of restaurants, shops, and other things to do. These types of areas are likely to have a lot of foot-traffic, and will draw in people from outside of that neighbourhood, and therefore present more opportunities to find customers.

However, it is also wise to avoid areas where the market is saturated, or where there may not be suficient demand. To that end, we can obtain a count of the cafès in each neighbourhood, and rule out areas that are too heavily populated with cafès, or where there are none. The final decision will depend how cautious one wants to be with regards to what neighbourhoods constitute excessive risk.