1. Introduction

The purpose of this study is to compare the cities of New York and Toronto and by means of an exploratory analysis and Foursquare location data, we will explore the venues of the neighborhoods of these two cities in order to find insights for business opportunities.

1.1 Background

Both New York City and Toronto are very diverse and are the financial capitals of their respective countries. Neighborhood of New York City has a total of 5 boroughs and 306 neighborhoods while Toronto has a total of 6 boroughs and 140 neighborhoods.

1.2 Business Problem

In this study we found that the first neighborhood in Scarborough in Toronto has very few venues than the first neighborhood in Brooklyn in New York City. In fact, we only found just one venue in the first neighborhood in Scarborough in Toronto. Compare to the first neighborhood of Brooklyn in New York City, which has many venues, any person interested in starting a new business in Scarborough can open almost any kind of business.

1.3 Interest

This study in intended for individuals who want to start a new business in the in Scarborough. A new business could be a coffee shop, bar, bakery, gym, or restaurant. The choice is up to the customer preference.

2. Data acquisition and cleaning

2.1. New York City – Brooklyn.

In order to segment the neighborhoods and explore them, we will essentially need a dataset that contains the 5 boroughs and the neighborhoods that exist in each borough as well as the latitude and longitude coordinates of each neighborhood.

This dataset exists for free on the web. On the other hand, for our convenience, the file was placed it on the server, so we can simply run a 'wget' command and access the data.

After downloading this dataset, we will transform it into a pandas dataframe and use the geopy library to get the latitude and longitude values of New York City to create a map with

its neighborhoods. Then, we will cluster and segment the neighborhoods in Brooklyn to find the first one for the analysis.

Being Bay Ridge the first neighborhood, we will create the GET request URL to find nearest venues. Also, we will determine which neighborhood in Brooklyn has more venues and we will group them as per category. Then, we will find the top 5 venues in Brooklyn in order to find insights.

2.2. Scarborough

Unlike New York City, the neighborhood data is not readily available on the internet. For the Toronto neighborhood data, a Wikipedia page exists that has all the information we need to explore and cluster the neighborhoods in Toronto.

We will scrape the Wikipedia page and wrangle the data, clean it, and then read it into a pandas dataframe. Once the data is in a structured format, we can replicate the analysis that we did to the New York City dataset to explore and cluster the neighborhoods in the city of Toronto.

After the mention procedure, we will use the geopy library to get the latitude and longitude values of Toronto to create a map with its neighborhoods. Then, we will cluster and segment the neighborhoods in Scarborough to find the first postal code for the analysis.

Being Malvern and Rouge the neighborhoods of the first postal code M1B, we will create the GET request URL to find nearest venues of this postal code. On the other hand, we will determine which neighborhood in Scarborough has more venues and we will group them as per category. Finally, we will find the top 5 venues in Scarborough in order to find insights.
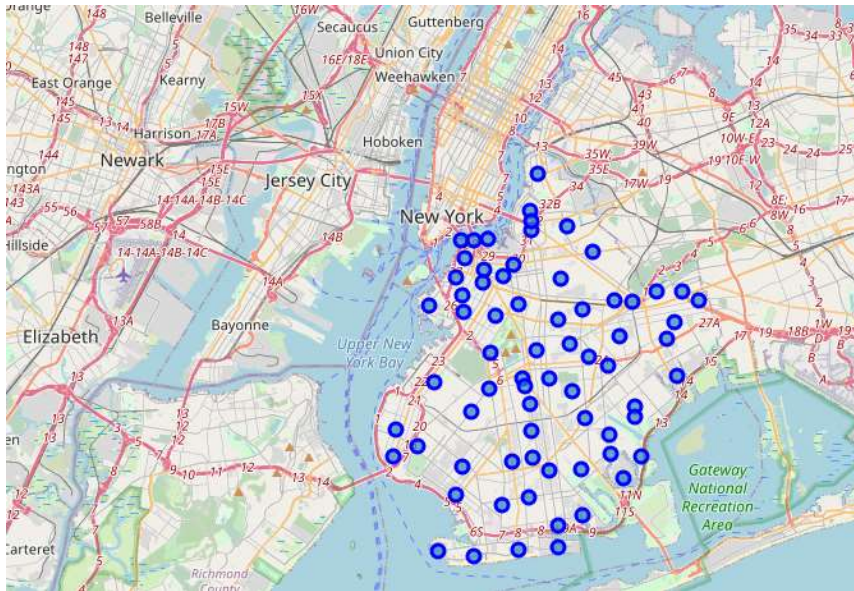
2.3. Brooklyn vs Scarborough

Finally, by means of the insights founds in Brooklyn and Scarborough, we will determine if we can open a new business in terms of the people needs.

3. Exploratory Data Analysis

3.1 Neighborhoods in New York City

3.2 Clustering and segmentation the neighborhoods in Brooklyn.

3.3 Exploring the first neighborhood of Brooklyn in our dataframe.
Bay Ridge

	Borough	Neighborhood	Latitude	Longitude
0	Brooklyn	Bay Ridge	40.625801	-74.030621
1	Brooklyn	Bensonhurst	40.611009	-73.995180
2	Brooklyn	Sunset Park	40.645103	-74.010316
3	Brooklyn	Greenpoint	40.730201	-73.954241
4	Brooklyn	Gravesend	40.595260	-73.973471

3.4 Brooklyn Venues Dataframe

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Bay Ridge	40.625801	-74.030621	Pilo Arts Day Spa and Salon	40.624748	-74.030591	Spa
1	Bay Ridge	40.625801	-74.030621	Bagel Boy	40.627896	-74.029335	Bagel Shop
2	Bay Ridge	40.625801	-74.030621	Leo's Casa Calamari	40.624200	-74.030931	Pizza Place
3	Bay Ridge	40.625801	-74.030621	Pegasus Cafe	40.623168	-74.031186	Breakfast Spot
4	Bay Ridge	40.625801	-74.030621	The Bookmark Shoppe	40.624577	-74.030562	Bookstore

There are total 2296 venues in the neighborhoods of Brooklyn in our dataset. The dataframe generated gives interesting information about Venue names like The *Bookmark Shoppe* or *Pegasus Cafe* and their corresponding categories like *Bookstore* or *Breakfast Spot.*

We can use this interesting information to categorize these neighborhoods using *k_*Means clustering to get insights about business or real estate investments.

3.5 Neighborhoods in Brooklyn with more Venues

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
North Side	100	100	100	100	100	100
South Side	100	100	100	100	100	100
Downtown	100	100	100	100	100	100
Carroll Gardens	98	98	98	98	98	98
Brooklyn Heights	93	93	93	93	93	93
East Williamsburg	82	82	82	82	82	82
Clinton Hill	81	81	81	81	81	81
Greenpoint	77	77	77	77	77	77
Cobble Hill	71	71	71	71	71	71

So we can see from the above dataframe that the Top 5 neighborhoods: South Side, Downtown, Noth Side, Carroll Gardens, and Brooklyn Heights are the neighborhoods in Brooklyn with more venues.

3.6 Popular Venue Categories in Brooklyn.

Venue Category	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude
Pizza Place	101	101	101	101	101	101
Coffee Shop	85	85	85	85	85	85
Bar	69	69	69	69	69	69
Deli / Bodega	65	65	65	65	65	65
Bakery	55	55	55	55	55	55
Grocery Store	53	53	53	53	53	53
Chinese Restaurant	49	49	49	49	49	49
Italian Restaurant	48	48	48	48	48	48
Mexican Restaurant	44	44	44	44	44	44

We can see from above dataframe that the Top 5 Venue Categories: Pizza Place, Coffee Shop, Deli / Bodega, Bar, and Bakery are the most frequently observed venue categories in Brooklyn. The occurrence of any of these venues in any neighborhood in Brooklyn will define the category of that neighborhood. Hence, we will cluster the neighborhoods on the basis of the frequency of the venue categories.

3.7 Top 10 Venues for each Neighborhood in Brooklyn Dataframe.

	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	Bath Beach	Gas Station	Italian Restaurant	Pharmacy	Middle Eastern Restaurant	Rental Car Location	Diner	Surf Spot	Sushi Restaurant	Check Cashing Service	German Restaurant
1	Bay Ridge	Spa	Italian Restaurant	Greek Restaurant	Thai Restaurant	Ice Cream Shop	Chinese Restaurant	Pizza Place	Sushi Restaurant	Grocery Store	Diner
2	Bedford Stuyvesant	Coffee Shop	Bus Station	Café	Bar	Thrift / Vintage Store	Bagel Shop	Boutique	New American Restaurant	Gift Shop	Juice Bar
3	Bensonhurst	Chinese Restaurant	Italian Restaurant	Asian Restaurant	Factory	Grocery Store	Sushi Restaurant	Pharmacy	Cosmetics Shop	Liquor Store	Park
4	Bergen Beach	Harbor / Marina	Baseball Field	Athletics & Sports	Donut Shop	Playground	Farm	Farmers Market	Fast Food Restaurant	Field	Filipino Restaurant

3.8 Neighborhoods in Toronto

3.9 Clustering and segmentation the neighborhoods in Scarborough.

3.10 Exploring the first neighborhood of Scarborough in our dataframe.

Malvern, Rouge

	Postal Code	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
5	M1J	Scarborough	Scarborough Village	43.744734	-79.239476

3.11 Scarborough Venues Dataframe.

	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	Royal Canadian Legion	43.782533	-79.163085	Bar
2	Rouge Hill, Port Union, Highland Creek	43.784535	-79.160497	SEBS Engineering Inc. (Sustainable Energy and ...	43.782371	-79.156820	Construction & Landscaping
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
5	Guildwood, Morningside, West Hill	43.763573	-79.188711	Big Bite Burrito	43.766299	-79.190720	Mexican Restaurant
6	Guildwood, Morningside, West Hill	43.763573	-79.188711	Enterprise Rent-A-Car	43.764076	-79.193406	Rental Car Location

There are total 71 venues in the neighborhoods of Scarborough in our dataset. The dataframe generated gives interesting information about Venue names like *Wendy's* or *Starbucks* and their corresponding categories like *Fast Food Restaurant* or *Coffee Shop.*

We can use this interesting information to categorize these neighborhoods using *k*_Means clustering to get insights about business or real estate investments.

3.12 Neighborhoods in Scarborough with more Venues.

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Steeles West, L'Amoreaux West	14	14	14	14	14	14
Cedarbrae	7	7	7	7	7	7
Clarks Corners, Tam O'Shanter, Sullivan	6	6	6	6	6	6
Golden Mile, Clairlea, Oakridge	6	6	6	6	6	6
Guildwood, Morningside, West Hill	5	5	5	5	5	5
Kennedy Park, Ionview, East Birchmount Park	5	5	5	5	5	5
Birch Cliff, Cliffside West	4	4	4	4	4	4
Agincourt	3	3	3	3	3	3

So we can see from the above dataframe that the Top neighborhoods: Steeles West, L'Amoreaux West; Cedarbrae; Clarks Corners, Tam O'Shanter, Sullivan; Golden Mile, Clairlea, Oakridge; and Guildwood, Morningside, West Hill are the neighborhoods in Scarborough with more venues.

3.13 Popular Venue Categories in Scarborough.

Venue Category	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude
Fast Food Restaurant	4	4	4	4	4	4
Bakery	4	4	4	4	4	4
Bank	4	4	4	4	4	4
Coffee Shop	4	4	4	4	4	4
Playground	2	2	2	2	2	2
Gas Station	2	2	2	2	2	2

We can see from above that Fast Food Restaurants, Bakery, Bank, Coffee Shop, and Playground are the most frequently observed venue categories in Scarborough. The occurrence of any of these venues in any neighborhood will define the category of that neighborhood. Hence we will cluster the neighborhoods on the basis of the frequency of the venue categories.

3.14 Top 10 Venues for each Neighborhood in Scarborough Dataframe.

	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	Breakfast Spot	Latin American Restaurant	Lounge	Vietnamese Restaurant	Chinese Restaurant	Electronics Store	Discount Store	Department Store	Convenience Store	Construction & Landscaping
1	Birch Cliff, Cliffside West	Café	Skating Rink	College Stadium	General Entertainment	Chinese Restaurant	Fast Food Restaurant	Electronics Store	Discount Store	Department Store	Convenience Store
2	Cedarbrae	Caribbean Restaurant	Bank	Hakka Restaurant	Gas Station	Thai Restaurant	Athletics & Sports	Bakery	Bar	Construction & Landscaping	Fast Food Restaurant
3	Clarks Corners, Tam O'Shanter, Sullivan	Thai Restaurant	Fast Food Restaurant	Shopping Mall	Bank	Gas Station	Coffee Shop	Chinese Restaurant	Electronics Store	Discount Store	Department Store
4	Cliffside, Cliffcrest, Scarborough Village West	American Restaurant	Motel	Athletics & Sports	Chinese Restaurant	Fast Food Restaurant	Electronics Store	Discount Store	Department Store	Convenience Store	Construction & Landscaping

4. Results

The first neighborhood of Brooklyn is Bay Ridge. The first neighborhood of Scarborough is Malvern, Rouge. Bay Ridge in Brooklyn has 63 venues. Malvern and Rouge in Scarborough have only 1 venue; which is Wendy's.

5. Discussion

Given the above results, I would suggest starting an Italian Food Restaurant in Scarborough would be very interesting since this type cuisine is not yet available. The people in Scarborough will appreciate this restaurant because it will be the only restaraunt with a unique type of cuisine.

6. Conclusion

Foursquare location is a very strong tool for Data Scientists. Among other things, we can explore cities in order to find insights for new or actual business.