CAPSTONE DATA SCIENCE PROJECT

New York VS Toronto

CLAUDE BERNARD BADJAGOU

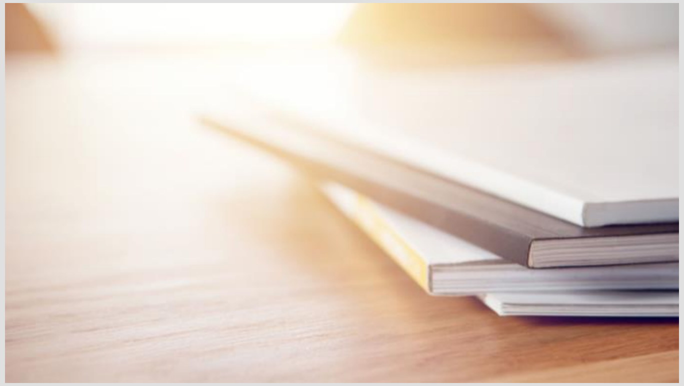
1. Introduction

1.1 Background

The geographical location of a business has a great influence on its accessibility, revenues, profits and survival, etc. Thus, for various reasons, it is essential to conduct a study on the characteristics of certain geographical areas to determine the location of a business.

1.2 Problem

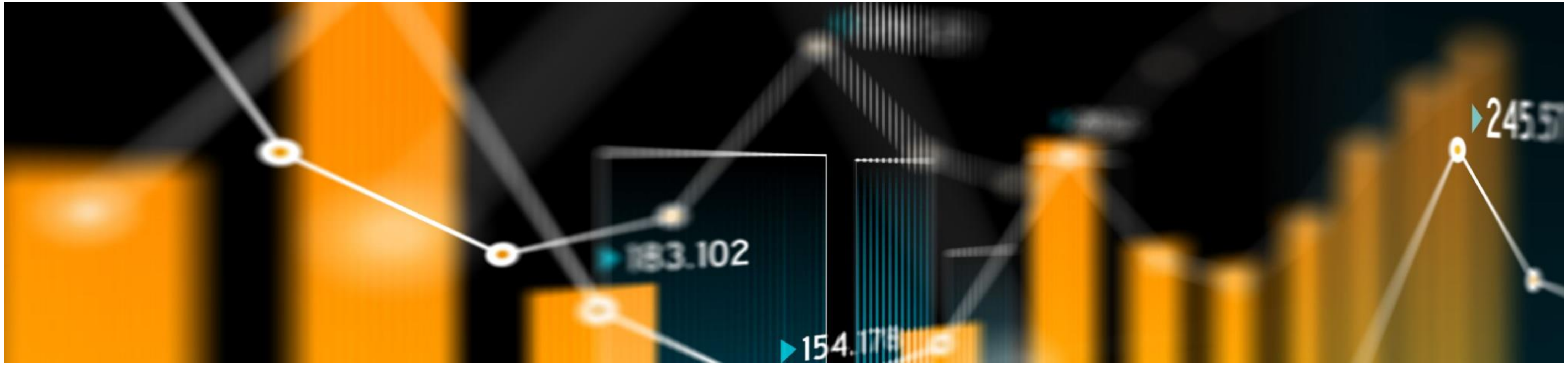
For this work, we want to choose the strategic location of a café restaurant in a district of the capital of a country. To do this, we will first study the characteristics of two different cities in each of the two countries (New York in America and Toronto in Canada). Then, we will choose the most appropriate area of the city that will allow the company to achieve its objectives.



1.3 Interest

First of all, globalization has favored immigration, and today we can find a diversity of cuisines and many coffee shops in all cities of the world such as New York, Toronto, etc. Therefore, in order to choose the location of the coffee shop in New York or Toronto, several characteristics should be explored such as The population of the neighborhoods The frequency of coffee restaurants and other places. Population density of the neighborhoods Availability of markets to purchase quality ingredients at a reasonable cost The availability of places of affluence (gyms, entertainment, education, parks, etc. nearby) The list is not exhaustive and can be expanded upon. Therefore, the choice of location is very important for the company before starting its activities. Thus, the first thing to do is to identify in both cities a populated area with high coffee consumption and low frequency of coffee restaurants. Then, we will compare the ratios of the characteristics of the two identified neighborhoods. Finally, we will choose the neighborhood that will favor the business in this exchange market.

2. Data acquisition and cleaning



2.1 Data sources

I will use the geographic location data of "New York City, USA" and "Toronto, Canada" available on Foursquare. I will also use postal codes of Canada table available on
https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

I will import a csv file named "Geospatial_Coordinates.csv" with the latitude and longitude coordinates of the Toronto boroughs available on
Coursera_Capstone/Geospatial_Coordinates.csv at main · lf-B7/Coursera_Capstone (github.com)
I will also import a json file with geographic information about New York City available at the following link:
https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkillsNetwork-DS0701EN-SkillsNetwork/labs/newyork_data.json

2.2 Data cleaning

Before cleaning the data, we must first download it into the working environment. To do this, we need to download or import the required libraries. Here we download and import the various functions and libraries we will need for the dataset operations.

```
# Librairies and functions
!pip install plotly
!pip install wordcloud
from bs4 import BeautifulSoup # this module helps in web scrapping.
import requests  # this module helps us to download a web page

import pandas as pd # library for data analsysis
pd.set_option('display.max_columns', None)
pd.set_option('display.max_rows', None)
pd.set_option("max_columns", 30)

import json # library to handle JSON files
from pandas.io.json import json_normalize # tranform JSON file into a pandas dataframe

from geopy.geocoders import Nominatim # convert an address into latitude and longitude values

import folium # map rendering library

# Matplotlib and associated plotting modules
import matplotlib.pyplot as plt
import matplotlib.cm as cs
import matplotlib.colors as colors

import numpy as np

import plotly.express as px
import plotly.graph_objects as go
print('Libraries imported.')

from wordcloud import WordCloud
from PIL import Image
```

```
Libraries imported.
```

Let's first view a part of Toronto data table from the source (https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M) using "BeautifulSoup"

	0	1	2	3	4	5	
0	M1ANot assigned	M2ANot assigned	M3ANorth York(Parkwoods)	M4ANorth York(Victoria Village)	M5ADowntown Toronto(Regent Park / Harbourfront)	M6ANorth York(Lawrence Manor / Lawrence Heights)	M7AQuee Park(Ontario Provinc Governme
1	M1BScarborough(Malvern / Rouge)	M2BNot assigned	M3BNorth York(Don Mills)North	M4BEast York(Parkview Hill / Woodbine Gardens)	M5BDowntown Toronto(Garden District, Ryerson)	M6BNorth York(Glencairn)	M7BNot assigr
2	M1CScarborough(Rouge Hill / Port Union / Highl...	M2CNot assigned	M3CNorth York(Don Mills)South(Flemingdon Park)	M4CEast York(Woodbine Heights)	M5CDowntown Toronto(St. James Town)	M6CYork(Humewood-Cedarvale)	M7CNot assigr
3	M1EScarborough(Guildwood / Morningside / West ...	M2ENot assigned	M3ENot assigned	M4EEast Toronto(The Beaches)	M5EDowntown Toronto(Berczy Park)	M6EYork(Caledonia-Fairbanks)	M7ENot assigr
4	M1GScarborough(Woburn)	M2GNot assigned	M3GNot assigned	M4GEast York(Leaside)	M5GDowntown Toronto(Central Bay Street)	M6GDowntown Toronto(Christie)	M7GNot assigr
5	M1HScarborough(Cedarbrae)	M2HNorth York(Hillcrest Village)	M3HNorth York(Bathurst Manor / Wilson Heights ...	M4HEast York(Thorncliffe Park)	M5HDowntown Toronto(Richmond / Adelaide / King)	M6HWest Toronto(Dufferin / Dovercourt Village)	M7HNot assigr
6	M1JScarborough(Scarborough Village)	M2JNorth York(Fairview / Henry Farm / Oriole)	M3JNorth York(Northwood Park / York University)	M4JEast YorkEast Toronto(The Danforth East)	M5JDowntown Toronto(Harbourfront East / Union ...	M6JWest Toronto(Little Portugal / Trinity)	M7JNot assigr
7	M1KScarborough(Kennedy Park / Ionview / East B...	M2KNorth York(Bayview Village)	M3KNorth York(Downsview)East (CFB Toronto)	M4KEast Toronto(The Danforth West / Riverdale)	M5KDowntown Toronto(Toronto Dominion Centre / ...	M6KWest Toronto(Brockton / Parkdale Village / ...	M7KNot assigr
8	M1LScarborough(Golden Mile / Clairlea / Oakridge)	M2LNorth York(York Mills / Silver Hills)	M3LNorth York(Downsview)West	M4LEast Toronto(India Bazaar / The Beaches West)	M5LDowntown Toronto(Commerce Court / Victoria ...	M6LNorth York(North Park / Maple Leaf Park / U...	M7LNot assigr
9	M1MScarborough(Cliffside / Cliffcrest / Scarbo...	M2MNorth York(Willowdale / Newtonbrook)	M3MNorth York(Downsview)Central	M4MEast Toronto(Studio District)	M5MNorth York(Bedford Park / Lawrence Manor East)	M6MYork(Del Ray / Mount Dennis / Keelsdale and...	M7MNot assigr

We removed incomplete information with "*Not assigned*" which is difficult to replace. We also renamed some "borough" after deleting some characters to create the following dataframe with the previous data toronto table which now contains in column postalcode, borough, neighborhood. And to get the full dataframe, we attached the data frame saved in the csv file to complete the latitudes and longitudes of the **Neighborhood**. (this is the head)

	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

New York data : the following is the content of the json file but not complete


```
{'type': 'FeatureCollection', 'totalFeatures': 306, 'features': [{'type': 'Feature', 'id': 'nyu_2451_34572.1', 'geometry': {'ty
pe': 'Point', 'coordinates': [-73.84720052054902, 40.89470517661]}, 'geometry_name': 'geom', 'properties': {'name': 'Wakefiel
d', 'stacked': 1, 'annoline1': 'Wakefield', 'annoline2': None, 'annoline3': None, 'annoangle': 0.0, 'borough': 'Bronx', 'bbox':
[-73.84720052054902, 40.89470517661, -73.84720052054902, 40.89470517661]}}, {'type': 'Feature', 'id': 'nyu_2451_34572.2', 'geom
etry': {'type': 'Point', 'coordinates': [-73.82993910812398, 40.87429419303012]}, 'geometry_name': 'geom', 'properties': {'nam
e': 'Co-op City', 'stacked': 2, 'annoline1': 'Co-op', 'annoline2': 'City', 'annoline3': None, 'annoangle': 0.0, 'borough': 'Bro
nx', 'bbox': [-73.82993910812398, 40.87429419303012, -73.82993910812398, 40.87429419303012]}}, {'type': 'Feature', 'id': 'nyu_2
451_34572.3', 'geometry': {'type': 'Point', 'coordinates': [-73.82780644716412, 40.887555677350775]}, 'geometry_name': 'geom',
'properties': {'name': 'Eastchester', 'stacked': 1, 'annoline1': 'Eastchester', 'annoline2': None, 'annoline3': None, 'annoangl
e': 0.0, 'borough': 'Bronx', 'bbox': [-73.82780644716412, 40.887555677350775, -73.82780644716412, 40.887555677350775]}}, {'typ
e': 'Feature', 'id': 'nyu_2451_34572.4', 'geometry': {'type': 'Point', 'coordinates': [-73.90564259591682, 40.89543742690383]},
'geometry_name': 'geom', 'properties': {'name': 'Fieldston', 'stacked': 1, 'annoline1': 'Fieldston', 'annoline2': None, 'annoli
ne3': None, 'annoangle': 0.0, 'borough': 'Bronx', 'bbox': [-73.90564259591682, 40.89543742690383, -73.90564259591682, 40.895437
42690383]}}, {'type': 'Feature', 'id': 'nyu_2451_34572.5', 'geometry': {'type': 'Point', 'coordinates': [-73.9125854610857, 40.
890834493891305]}, 'geometry_name': 'geom', 'properties': {'name': 'Riverdale', 'stacked': 1, 'annoline1': 'Riverdale', 'annoli
ne2': None, 'annoline3': None, 'annoangle': 0.0, 'borough': 'Bronx', 'bbox': [-73.9125854610857, 40.890834493891305, -73.912585
4610857, 40.890834493891305]}}, {'type': 'Feature', 'id': 'nyu_2451_34572.6', 'geometry': {'type': 'Point', 'coordinates': [-7
3.90281798724604, 40.88168737120521]}, 'geometry_name': 'geom', 'properties': {'name': 'Kingsbridge', 'stacked': 1, 'annoline
1': 'Kingsbridge', 'annoline2': None, 'annoline3': None, 'annoangle': 0.0, 'borough': 'Bronx', 'bbox': [-73.90281798724604, 40.
88168737120521, -73.90281798724604, 40.88168737120521]}}, {'type': 'Feature', 'id': 'nyu_2451_34572.7', 'geometry': {'type': 'P
oint', 'coordinates': [-73.91065965862981, 40.87655077879964]}, 'geometry_name': 'geom', 'properties': {'name': 'Marble Hill',
'stacked': 2, 'annoline1': 'Marble', 'annoline2': 'Hill', 'annoline3': None, 'annoangle': 0.0, 'borough': 'Manhattan', 'bbox':
[-73.91065965862981, 40.87655077879964, -73.91065965862981, 40.87655077879964]}}, {'type': 'Feature', 'id': 'nyu_2451_34572.8',
'geometry': {'type': 'Point', 'coordinates': [-73.86731496814176, 40.89827261213805]}, 'geometry_name': 'geom', 'properties':
{'name': 'Woodlawn', 'stacked': 1, 'annoline1': 'Woodlawn', 'annoline2': None, 'annoline3': None, 'annoangle': 0.0, 'borough':
'Bronx', 'bbox': [-73.86731496814176, 40.89827261213805, -73.86731496814176, 40.89827261213805]}}, {'type': 'Feature', 'id': 'n
yu_2451_34572.9', 'geometry': {'type': 'Point', 'coordinates': [-73.8793907395681, 40.87722415599446]}, 'geometry_name': 'geo
m', 'properties': {'name': 'Norwood', 'stacked': 1, 'annoline1': 'Norwood', 'annoline2': None, 'annoline3': None, 'annoangle':
0.0, 'borough': 'Bronx', 'bbox': [-73.8793907395681, 40.87722415599446, -73.8793907395681, 40.87722415599446]}}, {'type': 'Feat
ure', 'id': 'nyu_2451_34572.10', 'geometry': {'type': 'Point', 'coordinates': [-73.85744642974207, 40.88103887819211]}, 'geomet
ry_name': 'geom', 'properties': {'name': 'Williamsbridge', 'stacked': 1, 'annoline1': 'Williamsbridge', 'annoline2': None, 'ann
oline3': None, 'annoangle': 0.0, 'borough': 'Bronx', 'bbox': [-73.85744642974207, 40.88103887819211, -73.85744642974207, 40.881
```

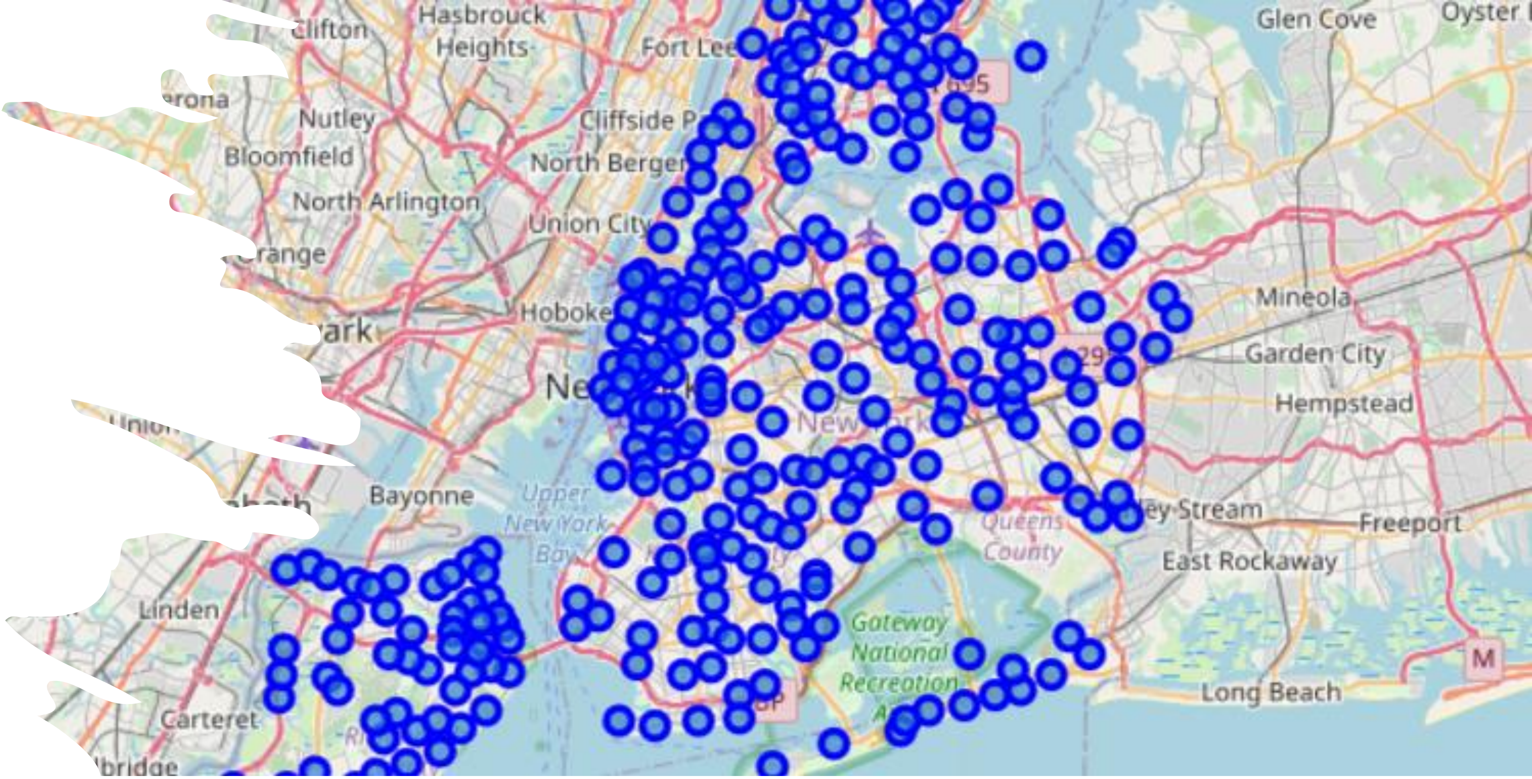

We have organized the data according to its content to obtain the following dataframe (this is the head):

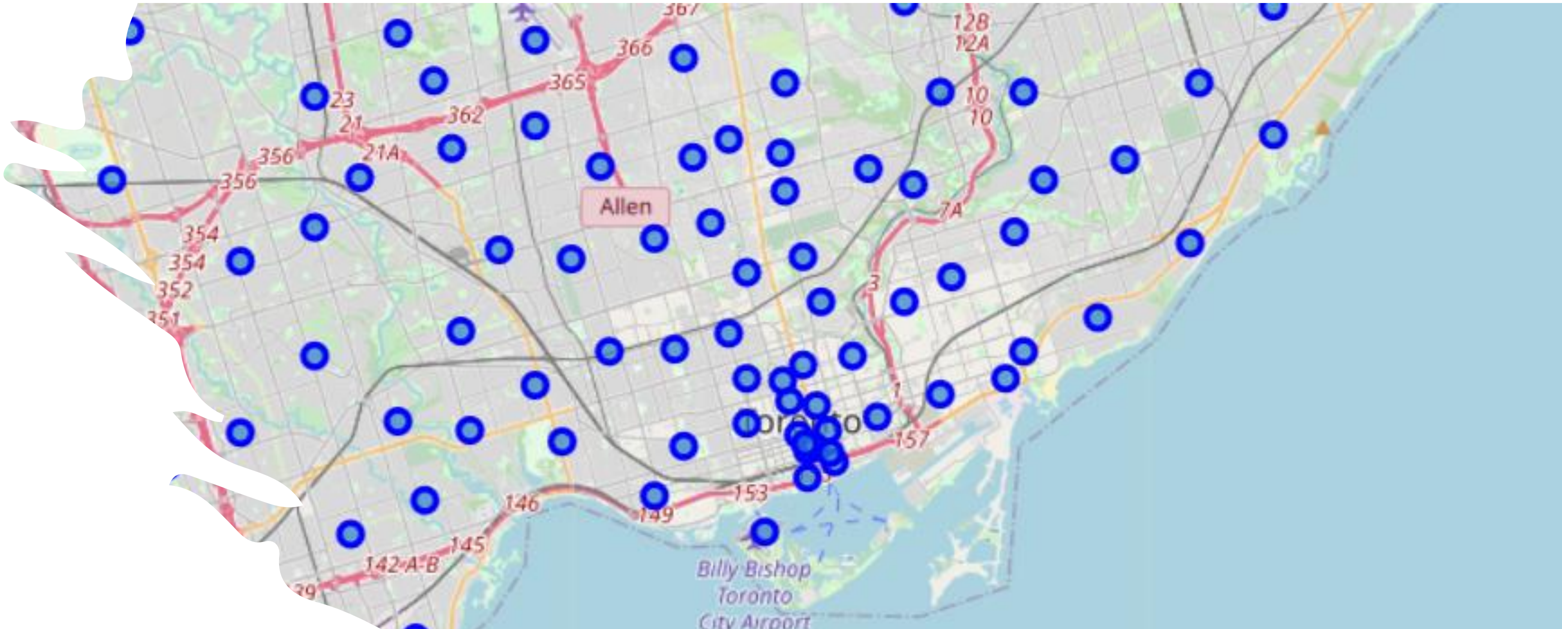
	Borough	Neighborhood	Latitude	Longitude
0	Bronx	Wakefield	40.894705	-73.847201
1	Bronx	Co-op City	40.874294	-73.829939
2	Bronx	Eastchester	40.887556	-73.827806
3	Bronx	Fieldston	40.895437	-73.905643
4	Bronx	Riverdale	40.890834	-73.912585

2.3 Feature selection

It is the different neighborhoods and their characteristics (locations and venues) that interest us here.

To do this we first created and explored the maps of New York and Toronto. The first map is for New York and the second is for Toronto.





We then used the Foursquare API to explore neighborhoods to extract the location category in all neighborhoods in New York and Toronto. Given the characteristics, we were interested in Downtown, TORONTO and Brooklyn, NEW YORK whose venues are respectively organized in the following dataframes:

```
30  downtown_neighborhoods_venues_sorted.head()
```

	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	Berczy Park	Coffee Shop	Cocktail Bar	Bakery	Beer Bar	Farmers Market	Pharmacy	Restaurant	Cheese Shop	Seafood Restaurant	Basketball Stadium
1	CN Tower, King and Spadina, Railway Lands, Har...	Airport Service	Airport Lounge	Airport Terminal	Bar	Boat or Ferry	Airport	Airport Food Court	Harbor / Marina	Rental Car Location	Sculpture Garden
2	Central Bay Street	Coffee Shop	Café	Sandwich Place	Italian Restaurant	Burger Joint	Japanese Restaurant	Thai Restaurant	Salad Place	Bank	Bubble Tea Shop
3	Christie	Grocery Store	Café	Park	Baby Store	Italian Restaurant	Candy Store	Nightclub	Coffee Shop	Restaurant	Dog Run
4	Church and Wellesley	Coffee Shop	Japanese Restaurant	Sushi Restaurant	Restaurant	Gay Bar	Yoga Studio	Men's Store	Mediterranean Restaurant	Hotel	Fast Food Restaurant

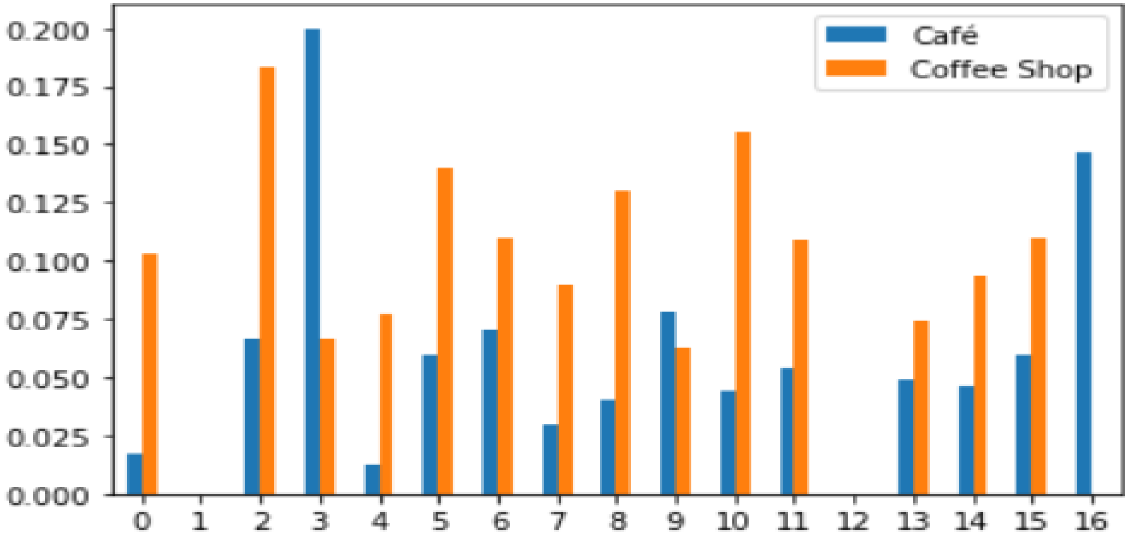
```
30  brooklyn_neighborhoods_venues_sorted.head()
```

	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	Chinese Restaurant	Pizza Place	Bubble Tea Shop	Italian Restaurant	Gas Station	Fast Food Restaurant	Donut Shop	Pharmacy	Ice Cream Shop	Burger Joint
1	Bay Ridge	Italian Restaurant	Pizza Place	Spa	Bagel Shop	Greek Restaurant	American Restaurant	Bar	Chinese Restaurant	Pharmacy	Playground
2	Bedford Stuyvesant	Bar	Coffee Shop	Deli / Bodega	Café	Pizza Place	Thrift / Vintage Store	Juice Bar	Bagel Shop	Gourmet Shop	Fruit & Vegetable Store
3	Bensonhurst	Chinese Restaurant	Ice Cream Shop	Park	Grocery Store	Italian Restaurant	Donut Shop	Sushi Restaurant	Pizza Place	Coffee Shop	Cosmetics Shop
4	Bergen Beach	Harbor / Marina	Baseball Field	Park	Playground	Athletics & Sports	Farmers Market	Event Service	Event Space	Factory	Falafel Restaurant

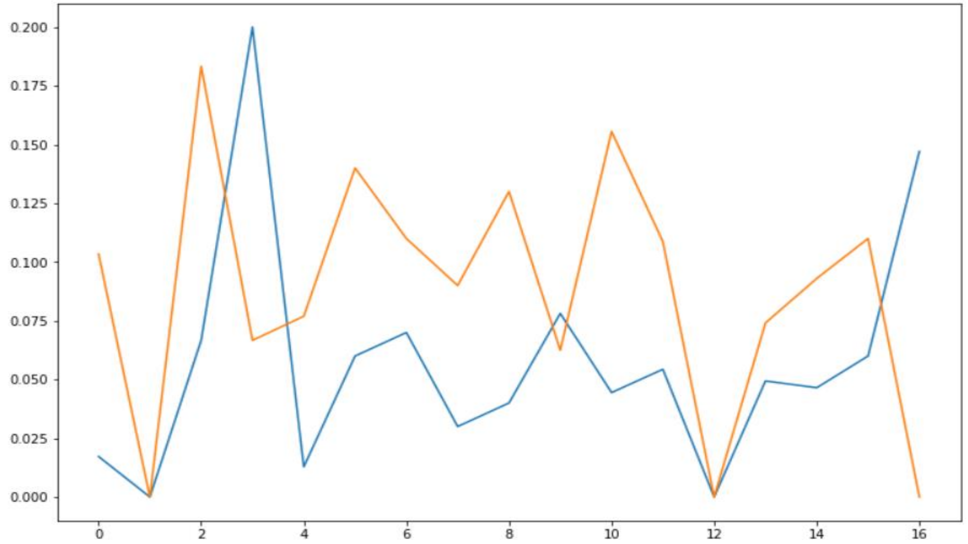
3. Exploratory and Data Analysis

The frequency and layout of the venues like "Café", "Coffee Shop" in "Downtown, Toronto" motivated us to choose this location to further research the location for our "Café" Restaurant.

This is the Bar plot of "Café" and "Coffee Shop" in Downtown

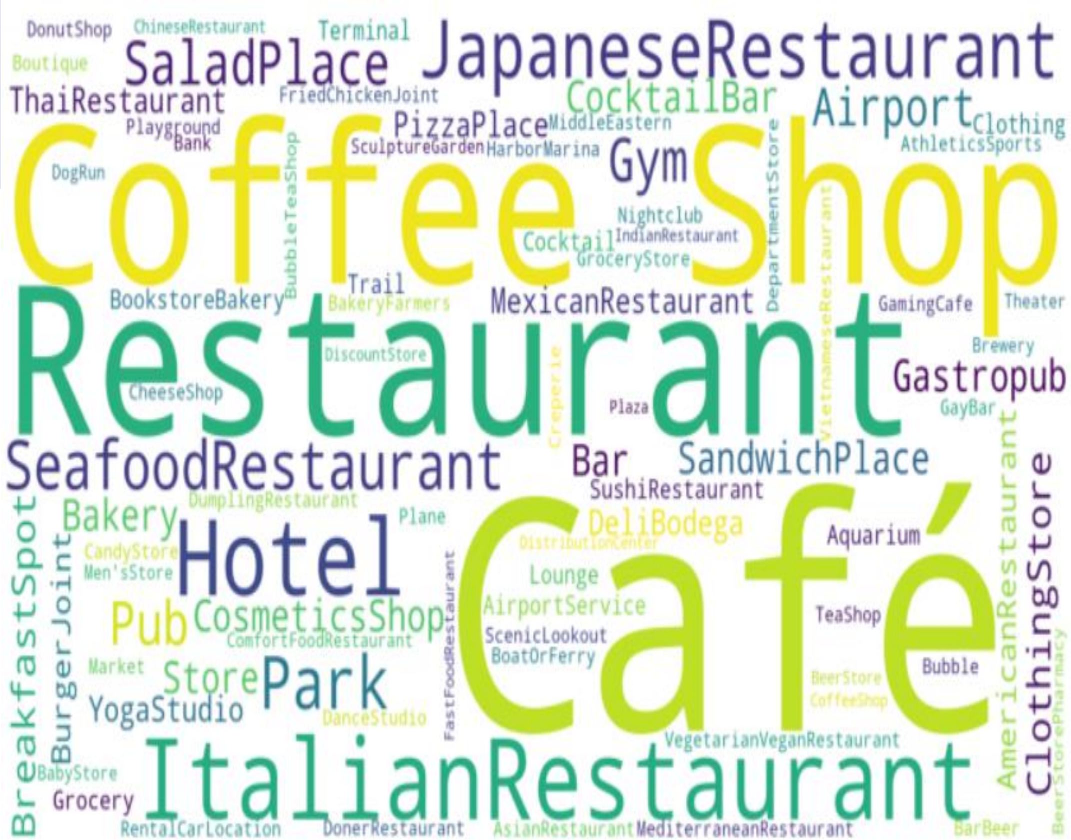


This is the Line plot of "Café" and "Coffee Shop" in Downtown



This shows that "Café" or even "Coffee Shop" is more frequent in some regions than in others. Also in the same region, "Coffee Shop" is almost more frequent than "Café".

The main characteristics of the downtown, Toronto areas



The first figure (word cloud) shows that "Coffee shop", "Café" and "Restaurant", which are more visible, are more frequent than other features like "Dog run", "Plane", etc. which are less visible.

The center is called "Downtown", and around the center we have "the different zones". The features closest to the center are more frequent in the area.

Dans ce cadre de données, nous avons gardé "Coffee" et "Coffee Shop" et supprimé les autres caractéristiques pour mieux voir l'ordre de présence de ces deux caractéristiques dans les zones.

	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	Berczy Park	Coffee Shop	-	-	-	-	-	-	-	-	-
1	CN Tower, King and Spadina, Railway Lands, Har...	-	-	-	-	-	-	-	-	-	-
2	Central Bay Street	Coffee Shop	Café	-	-	-	-	-	-	-	-
3	Christie	-	Café	-	-	-	-	-	Coffee Shop	-	-
4	Church and Wellesley	Coffee Shop	-	-	-	-	-	-	-	-	-
5	Commerce Court, Victoria Hotel	Coffee Shop	-	Café	-	-	-	-	-	-	-
6	First Canadian Place, Underground city	Coffee Shop	Café	-	-	-	-	-	-	-	-
7	Garden District, Ryerson	Coffee Shop	-	-	-	-	-	-	Café	-	-
8	Harbourfront East, Union Station, Toronto Islands	Coffee Shop	-	Café	-	-	-	-	-	-	-
9	Kensington Market, Chinatown, Grange Park	Café	-	Coffee Shop	-	-	-	-	-	-	-
10	Regent Park, Harbourfront	Coffee Shop	-	-	-	-	-	Café	-	-	-
11	Richmond, Adelaide, King	Coffee Shop	Café	-	-	-	-	-	-	-	-
12	Rosedale	-	-	-	-	-	-	-	-	-	-
13	St. James Town	Coffee Shop	-	Café	-	-	-	-	-	-	-
14	St. James Town, Cabbagetown	Coffee Shop	-	-	Café	-	-	-	-	-	-
15	Toronto Dominion Centre, Design Exchange	Coffee Shop	-	Café	-	-	-	-	-	-	-
16	University of Toronto, Harbord	Café	-	-	-	-	-	-	-	-	-

4. Results

We can see that "Coffee" is LESS frequented in the following neighborhoods: "Berczy Park", "CN Tower", "King and Spadina", "Railway", "Lands", "Rosedale", "Church and Wellesley", "Regent Park" and "Harbourfront".

So we'd better choose a coffee corner in one of these neighborhoods where "Coffee" is not among the 10 most frequented places. But before making a choice, we will first check the frequency of "Coffee Shop" in the neighborhood to choose.

"Coffee Shop" is less frequent in neighborhoods like ("CN Tower", "King and Spadina", "Railway", "Lands", "Rosedale", "University of Toronto" and "Harbord"). We can afford to eliminate the neighborhoods where "Coffee" and "Coffee Shop" are less frequent. This operation therefore consists of eliminating the following neighborhoods: ["CN Tower", "King and Spadina", "Railway", "Lands", "Rosedale"].

We can also see that "Coffee Shop" is MORE frequent in almost all neighborhoods, except those mentioned in parentheses. The high frequency of "Coffee Shop" would put them in competition. Competition that would make the price of ingredients affordable.

The second operation is to look at the other neighborhoods that remain in parentheses, presenting those where "Coffee Shop" is LESS frequent after eliminating the neighborhoods listed in parentheses. From this new list of neighborhoods, we must choose "Berczy Park", "Church and Wellesley", "Regent Park" and "Harbourfront".

Our choice should be the neighborhood with a low frequency of "Coffee" and a high frequency of "Coffee Shop". We can afford to eliminate "Regent Park" and "Harbourfront" which have a higher frequency of "Coffee Shop" (we have "Coffee Shop" as the 7th most frequent place on its line while "Berczy Park", "Church and Wellesley" have nothing until the 10th most frequent place.

The "Berczy Park" neighborhood seems to be one of the best candidates for the coffee corner location. But after consulting the map, it seems that the area is not favorable for the location of the restaurant "Café". It would be better to choose the location of the restaurant "Café" in the area of "Church and Wellesley".

5. Discussion

What would you like me to add or remove from this project ?

To see the python code of this project you can click on the following link:
[Coursera_Capstone/The Battle of Neighborhoods (Week 2).ipynb at main · lf-B7/Coursera_Capstone (github.com)]

We have used the Foursquare API which can however generate additional data to those we have used in this project. So our conclusion may become biased against our own basic criteria.

6. Conclusion

Our criteria and analysis point to downtown Toronto and specifically to "Church and Wellesley" instead of "Berczy Park". If we change the analysis criteria, we could end up in another neighborhood area in Toronto or New York. Like other factors and or criteria that may come into play in the choice of the coffee corner, it would be preferable to choose "Church and Wellesley" as the area emplacement of our "Café" Restaurant.

THANK YOU FOR READING THIS PROJECT "The Battle of Neighborhoods" TO THE END !!!

- ✓ Please add your comments