IBM Capstone Project - The Battle of Neighborhoods

Winson Yeung, 10 May 2021

Introduction

- New York and Toronto are the financial capitals of their respective countries and are very diverse. Both cities offer a wide variety of world’s delicious food.

Target Audience

- The task is to compare both cities and provides insights to people who would like to open a restaurant in these two cities.

Data Set

In order to compare New York City and Toronto, data such as geographical data and venue data were collected. Geographical data include boroughs, neighbourhoods, latitude and longitude for both cities to draw maps.

To compare different public venues in the neighbourhoods of both cities, data were collected from Foursquare api and appended into the geographical data.

- A .json file to fetch data of New York city: https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkillsNetwork-DS0701EN-SkillsNetwork/labs/newyork_data.json
- A Wikipedia Page to fetch data of Toronto city: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
- Foursquare api to fetch different public venues in the vicinity of the neighbourhood

Methodology

Data downloaded or scraped from multiple sources were combined into one table for New York City and Toronto respectively. Geopy library were used to get the latitude and longitude values of New York City and Toronto. New York City were divided into 5 boroughs with 306 neighbourhoods while Toronto were divided into 15 boroughs with 103 neighbourhoods.

```
neighborhoods.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

```
print('The dataframe has {} boroughs and {} neighborhoods.'.format(
        len(neighborhoods['Borough'].unique()),
        neighborhoods.shape[0]
    )
)
```

```
The dataframe has 5 boroughs and 306 neighborhoods.
```

```
toronto = pd.merge(df, geotable, left_on='Postal Code', right_on='Postal Code', left_index=F
toronto.head()
```

	Postal Code	Borough	Neighborhood	Latitude	Longitude
0	M3A	North York	Parkwoods	43.753259	-79.329656
1	M4A	North York	Victoria Village	43.725882	-79.315572
2	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636
3	M6A	North York	Lawrence Manor, Lawrence Heights	43.718518	-79.464763
4	M7A	Queen's Park	Ontario Provincial Government	43.662301	-79.389494

```
print('The dataframe has {} boroughs and {} neighborhoods.'.format(
        len(toronto['Borough'].unique()),
        toronto.shape[0]
))
```

```
The dataframe has 15 boroughs and 103 neighborhoods.
```

Foursquare API was applied to explore the neighbourhoods and segment them by accessing and acquiring the venue data such as venue name, venue unique ID, venue category, venue location (latitude and longitude) etc. A Foursquare Developer account was registered to obtain credentials (ie. client ID and client Secret key).

```
: CLIENT_ID = 'DE0OR0AXUKDZFS25UM4DALLMJ5PST4OBRRDRYYVP1TATKE3A' # your Foursquare ID
  CLIENT_SECRET = 'SNC5NNSSP1LTTE23U4VH2ZIF3IV5GSWZJFBVIY4UYRRUNCHS' # your Foursquare Secret
  VERSION = '20180605' # Foursquare API version
  LIMIT = 100 # A default Foursquare API limit value

  print('Your credentails:')
  print('CLIENT_ID: ' + CLIENT_ID)
  print('CLIENT_SECRET:' + CLIENT_SECRET)
```

```
Your credentails:
CLIENT_ID: DE0OR0AXUKDZFS25UM4DALLMJ5PST4OBRRDRYYVP1TATKE3A
CLIENT_SECRET:SNC5NNSSP1LTTE23U4VH2ZIF3IV5GSWZJFBVIY4UYRRUNCHS
```

Each Neighbourhood was analysed and top 10 most common venues with their frequency for each neighbourhood were displayed.

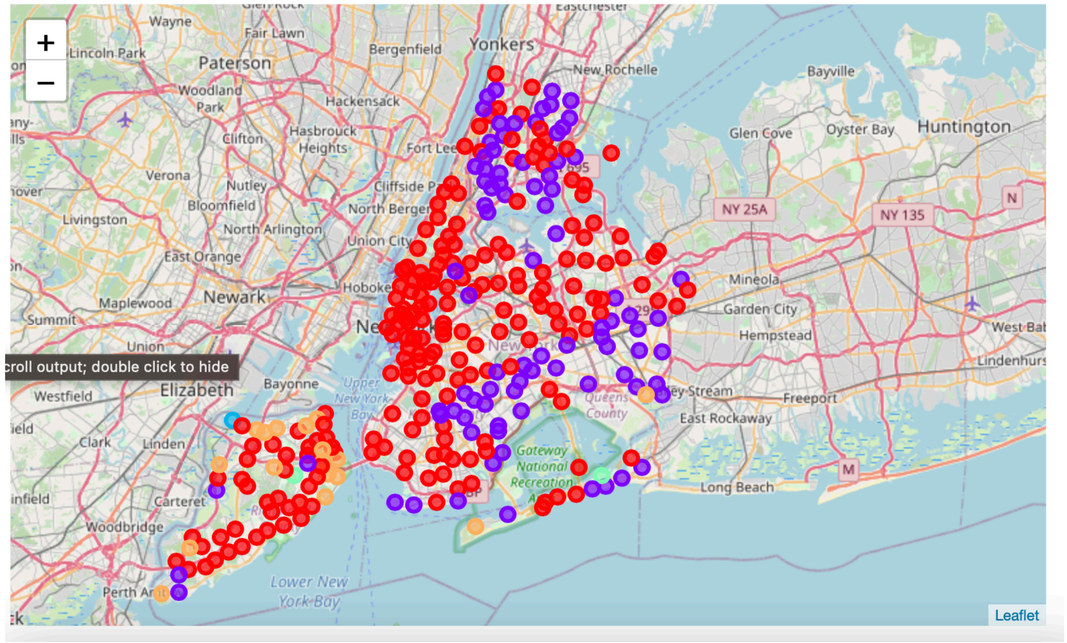
	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10 C
0	Allerton	Pizza Place	Supermarket	Chinese Restaurant	Deli / Bodega	Grocery Store	Cosmetics Shop	Pharmacy	Check Cashing Service	Mexican Restaurant	
1	Annadale	Pizza Place	Dance Studio	Deli / Bodega	Restaurant	Diner	Train Station	Flower Shop	Falafel Restaurant	Farm	
2	Arden Heights	Coffee Shop	Pizza Place	Playground	Pharmacy	Yemeni Restaurant	Flea Market	Factory	Falafel Restaurant	Farm	
3	Arlington	Deli / Bodega	ATM	Coffee Shop	American Restaurant	Intersection	Food	Falafel Restaurant	Farm	Farmers Market	Fa Re
4	Arrochar	Pizza Place	Bus Stop	Italian Restaurant	Deli / Bodega	Polish Restaurant	Outdoors & Recreation	Sandwich Place	Liquor Store	Bagel Shop	At

```
toronto_neighborhoods_venues_sorted.head()
```

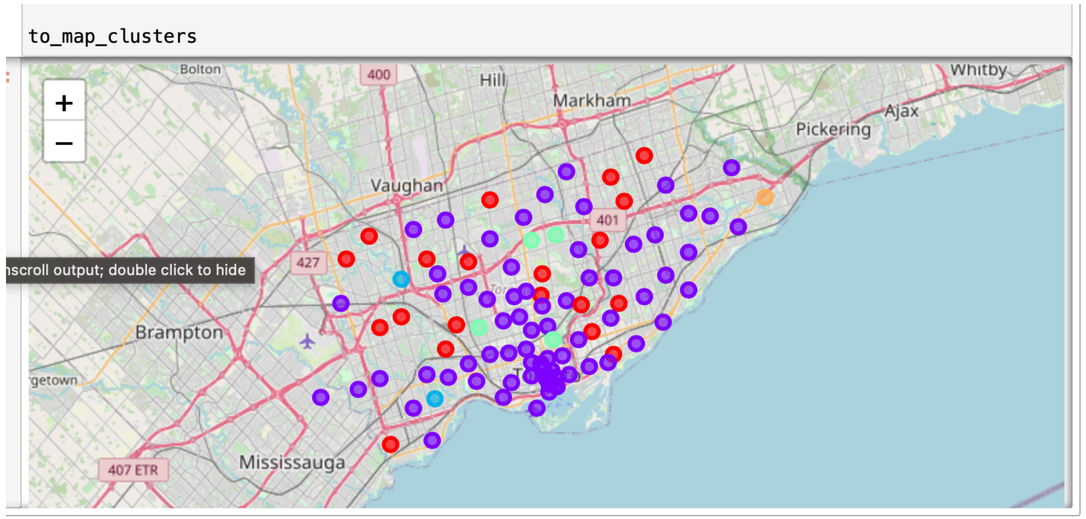

	Ne	1st Most Venue	2nd Most Venue	3rd Most ommon Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	1(
0	Agincourt	Lounge	Latin American Restaurant	Breakfast Spot	Clothing Store	Donut Shop	Diner	Discount Store	Distribution Center	Dog Run	R
1	Alderwood, Long Branch	Pizza Place	Coffee Shop	Skating Rink	Playground	Pub	Sandwich Place	Gym	Airport Lounge	Department Store	R
2	Bathurst Manor, Wilson Heights, Downsview North	Bank	Coffee Shop	Mobile Phone Shop	Health Food Store	Supermarket	Ice Cream Shop	Restaurant	Deli / Bodega	Fried Chicken Joint	
3	Bayview Village	Café	Japanese Restaurant	Bank	Chinese Restaurant	Women's Store	Doner Restaurant	Discount Store	Distribution Center	Dog Run	
4	Bedford Park, Lawrence Manor East	Coffee Shop	Sandwich Place	Italian Restaurant	Restaurant	Hobby Shop	Juice Bar	Fast Food Restaurant	Butcher	Indian Restaurant	

Performing K-means to cluster the neighbourhoods into 5 clusters. Then each cluster was examined to discriminate venue categories that distinguish each cluster. Folium was used to help create maps of New York and Toronto with clustered neighbourhoods on top for visualisation purpose.

ny_map_clusters



to_map_clusters



Results

The results allow the target audience to identify which neighbourhoods have higher concentration of restaurants while which have fewer.

```
newyork_merged.loc[newyork_merged['Cluster Labels'] == 0, newyork_merged.columns[[1] + list(
```

53	Manhattan Terrace	Pizza Place	Bagel Shop	Ice Cream Shop	Donut Shop	Mobile Phone Shop	Grocery Store	Convenience Store	Cosm
54	Flatbush	Deli / Bodega	Coffee Shop	Caribbean Restaurant	Plaza	Pharmacy	Fast Food Restaurant	Bank	Ba
55	Crown Heights	Pizza Place	Museum	Bookstore	Bakery	Café	Bagel Shop	Playground	Candy
57	Kensington	Thai Restaurant	Grocery Store	Ice Cream Shop	Pizza Place	Music Venue	Spa	Mexican Restaurant	Mobile P
58	Windsor Terrace	Diner	Deli / Bodega	Plaza	Park	Grocery Store	Sushi Restaurant	Bar	Vintage
59	Prospect Heights	Bar	Mexican Restaurant	Coffee Shop	Thai Restaurant	Wine Shop	Bakery	Restaurant	Cocktai
61	Williamsburg	Pizza Place	Bagel Shop	Coffee Shop	Breakfast Spot	Lounge	Bar	Middle Eastern Restaurant	Resta

```
toronto_merged.loc[toronto_merged['Cluster Labels'] == 1, toronto_merged.columns[[1] + list(
```

	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	
1	North York	1.0	Coffee Shop	Portuguese Restaurant	Financial or Legal Service	Hockey Arena	Women's Store	Doner Restaurant	Dim Sum Restaurant	
2	Downtown Toronto	1.0	Coffee Shop	Park	Bakery	Café	Pub	Breakfast Spot	Theater	Yo
3	North York	1.0	Clothing Store	Furniture / Home Store	Accessories Store	Vietnamese Restaurant	Coffee Shop	Carpet Store	Miscellaneous Shop	
4	Queen's Park	1.0	Coffee Shop	Sushi Restaurant	Yoga Studio	Diner	Bank	Bar	Restaurant	R
6	Scarborough	1.0	Fast Food Restaurant	Print Shop	Women's Store	Donut Shop	Dim Sum Restaurant	Diner	Discount Store	Di
7	North York	1.0	Gym	Baseball Field	Café	Caribbean Restaurant	Japanese Restaurant	Discount Store	Distribution Center	