Capstone Project - The Battle of Neighborhoods

Introduction

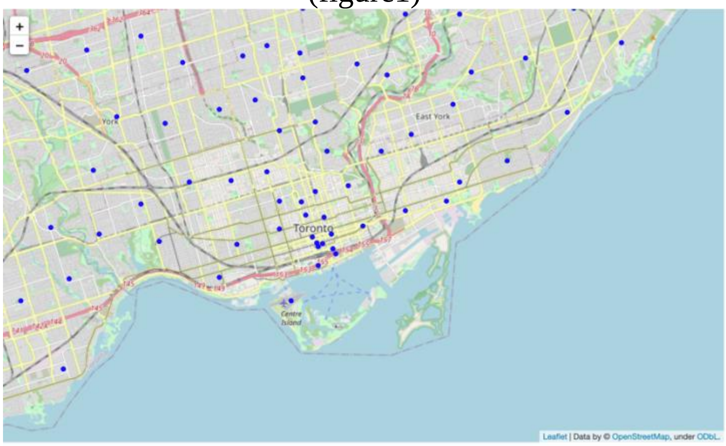
When a person is relocating to another city/neighbors due to various reasons like career development, family gathering and/or other personal reasons, it is of his/her preference to find the similar living evironment as previous live. Therefore, in this task, I will use the two cities (New York and Toronto)'s imformation to try to find out the similar neigbors/boroughs across cities or boroughs. Therefore, in this capstone project, I will use the data of both Toronto and New York (boroughs, neighborhoods, latitude and longitude, venues, etc), which extracted from the Wikipedia and the venues information by using API from Foursquare to help further analysis. This analysis will finally came to the conclusion that the similar neighbors/boroughs of these two cities (in case of searching shops for coffee, deserts) in order for the people who has the previous described situation's selection.

Data

At first, for the Toronto data, it is extract from the following Wikipedia page, https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M, which includes the neighbors, postcode and boroughs of Toronto. By using the methods of web scrap of Beautiful soup, the data can be extracted as the pandas data frame for further usage. The shape of the data is (180, 3), which means that the data has 180 different categories with three dimensions. For further usage, I conducted some adjustment and assigned the postcode as the keys of the data frame, which enables the (103, 3). Therefore, the number of keys are 103. Further, I use the Geospatial data of the Toronto and append it into the previous data frame to enrich it with the latitude and longitude information.

	Postal Code	Borough	Neighbourhood
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	Downtown Toronto	Queen's Park, Ontario Provincial Government

(figure1)

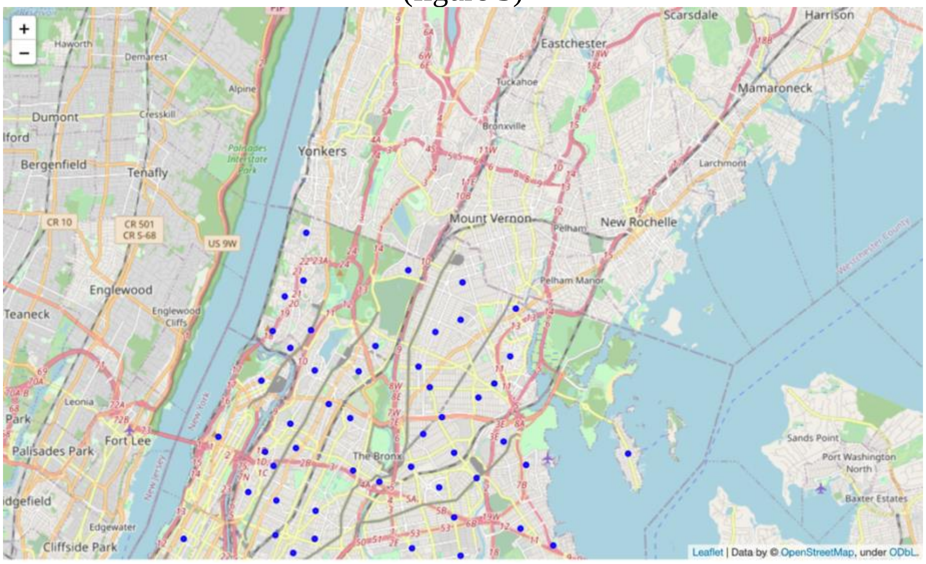


(figure 2)

For the New York data, I use the New York data provided by the project and extracted it into the data frame with the similar format of the Toronto.

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

(figure 3)



(figure 4)

After extracted the data into the above data frame, I use the Foursquare API to extract the venues information of the two cities. The below graph is the data frame include the venue information of New York and Toronto.

	Neighborhood	Borough	Latitude	Longitude	VenueName	VenueLatitude	VenueLongitude	VenueCategory
0	Wakefield	Bronx	40.894705	-73.847201	Lollipops Gelato	40.894123	-73.845892	Dessert Shop
1	Wakefield	Bronx	40.894705	-73.847201	Ripe Kitchen & Bar	40.898152	-73.838875	Caribbean Restaurant
2	Wakefield	Bronx	40.894705	-73.847201	Ali's Roti Shop	40.894036	-73.856935	Caribbean Restaurant
3	Wakefield	Bronx	40.894705	-73.847201	Jackie's West Indian Bakery	40.889283	-73.843310	Caribbean Restaurant
4	Wakefield	Bronx	40.894705	-73.847201	Rite Aid	40.889062	-73.842993	Pharmacy
5	Wakefield	Bronx	40.894705	-73.847201	Carvel Ice Cream	40.890487	-73.848568	Ice Cream Shop
6	Wakefield	Bronx	40.894705	-73.847201	Jimbo's	40.891740	-73.858226	Burger Joint
7	Wakefield	Bronx	40.894705	-73.847201	Dunkin'	40.890459	-73.849089	Donut Shop
8	Wakefield	Bronx	40.894705	-73.847201	Walgreens	40.896528	-73.844700	Pharmacy
9	Wakefield	Bronx	40.894705	-73.847201	Rite Aid	40.896649	-73.844846	Pharmacy
10	Wakefield	Bronx	40.894705	-73.847201	Subway	40.890468	-73.849152	Sandwich Place
11	Wakefield	Bronx	40.894705	-73.847201	Walgreens	40.898757	-73.854446	Pharmacy
12	Wakefield	Bronx	40.894705	-73.847201	Shell	40.894187	-73.845862	Gas Station
13	Wakefield	Bronx	40.894705	-73.847201	T-Mobile	40.893216	-73.857251	Mobile Phone Shop
14	Wakefield	Bronx	40.894705	-73.847201	Popeyes Louisiana Kitchen	40.898292	-73.854719	Fried Chicken Joint

(figure 5)

	Neighborhood	Borough	Latitude	Longitude	VenueName	VenueLatitude	VenueLongitude	VenueCategory
0	Parkwoods	North York	43.753259	-79.329656	Allwyn's Bakery	43.759840	-79.324719	Caribbean Restaurant
1	Parkwoods	North York	43.753259	-79.329656	Brookbanks Park	43.751976	-79.332140	Park
2	Parkwoods	North York	43.753259	-79.329656	Tim Hortons	43.760668	-79.326368	Café
3	Parkwoods	North York	43.753259	-79.329656	Bruno's valu-mart	43.746143	-79.324630	Grocery Store
4	Parkwoods	North York	43.753259	-79.329656	High Street Fish & Chips	43.745260	-79.324949	Fish & Chips Shop

(figure 6)

Methodology

In this analysis, I first identify the venues category with the name included shops and see the distribution of these shop in both cities. In New York, the unique venue category include 'shop' are in figure 7, while for Toronto, the unique shops include in figure 8.

```
array(['Dessert Shop', 'Ice Cream Shop', 'Donut Shop',
       'Mobile Phone Shop', 'Food & Drink Shop', 'Flower Shop',
       'Shopping Mall', 'Bagel Shop', 'Coffee Shop', 'Miscellaneous Shop',
       'Sporting Goods Shop', 'Cosmetics Shop', 'Wine Shop',
       'Supplement Shop', 'Automotive Shop', 'Shop & Service',
       'Frozen Yogurt Shop', 'Gourmet Shop', 'Smoke Shop',
       'Fish & Chips Shop', 'Gift Shop', 'Hobby Shop', 'Cupcake Shop',
       'Print Shop', 'Cheese Shop', 'Optical Shop', 'Bubble Tea Shop',
       'Bike Shop', 'Record Shop', 'Other Repair Shop', 'Comic Shop',
       'Antique Shop', 'Pie Shop', 'Motorcycle Shop', 'Souvenir Shop',
       'Shopping Plaza', 'Board Shop', 'Tailor Shop', 'Chocolate Shop'],
      dtype=object)
```

(figure 7)

```
: array(['Fish & Chips Shop', 'Food & Drink Shop', 'Coffee Shop',
         'Shopping Mall', 'Cosmetics Shop', 'Shop & Service',
         'Chocolate Shop', 'Dessert Shop', 'Ice Cream Shop', 'Cheese Shop',
         'Miscellaneous Shop', 'Bubble Tea Shop', 'Hobby Shop',
         'Smoke Shop', 'Gift Shop', 'Gourmet Shop', 'Comic Shop'],
        dtype=object)
```

(figure 8)

After filtering the category of venues by 'shops', I sort the result and create one hot dummies of the two data and drop all columns besides the dummies, that illustrate the shop distribution of the different area in case of further analysis. The result is as below that New York is shown as figure 9 and Toronto is figure 10.

	Antique Shop	Automotive Shop	Bagel Shop	Bike Shop	Board Shop	Bubble Tea Shop	Cheese Shop	Chocolate Shop	Coffee Shop	Comic Shop	Cosmetics Shop	Cupcake Shop	Dessert Shop	Don Sh
0	1	0	0	0	0	0	0	0	0	0	0	0	0	
1	1	0	0	0	0	0	0	0	0	0	0	0	0	
2	1	0	0	0	0	0	0	0	0	0	0	0	0	
3	1	0	0	0	0	0	0	0	0	0	0	0	0	
4	0	1	0	0	0	0	0	0	0	0	0	0	0	

(figure 9)

	Bubble Tea Shop	Cheese Shop	Chocolate Shop	Coffee Shop	Comic Shop	Cosmetics Shop	Dessert Shop	Fish & Chips Shop	Food & Drink Shop	Gift Shop	Gourmet Shop	Hobby Shop	Ice Cream Shop	Miscellane s
0	1	0	0	0	0	0	0	0	0	0	0	0	0	
1	1	0	0	0	0	0	0	0	0	0	0	0	0	
2	0	1	0	0	0	0	0	0	0	0	0	0	0	
3	0	0	1	0	0	0	0	0	0	0	0	0	0	
4	0	0	0	1	0	0	0	0	0	0	0	0	0	

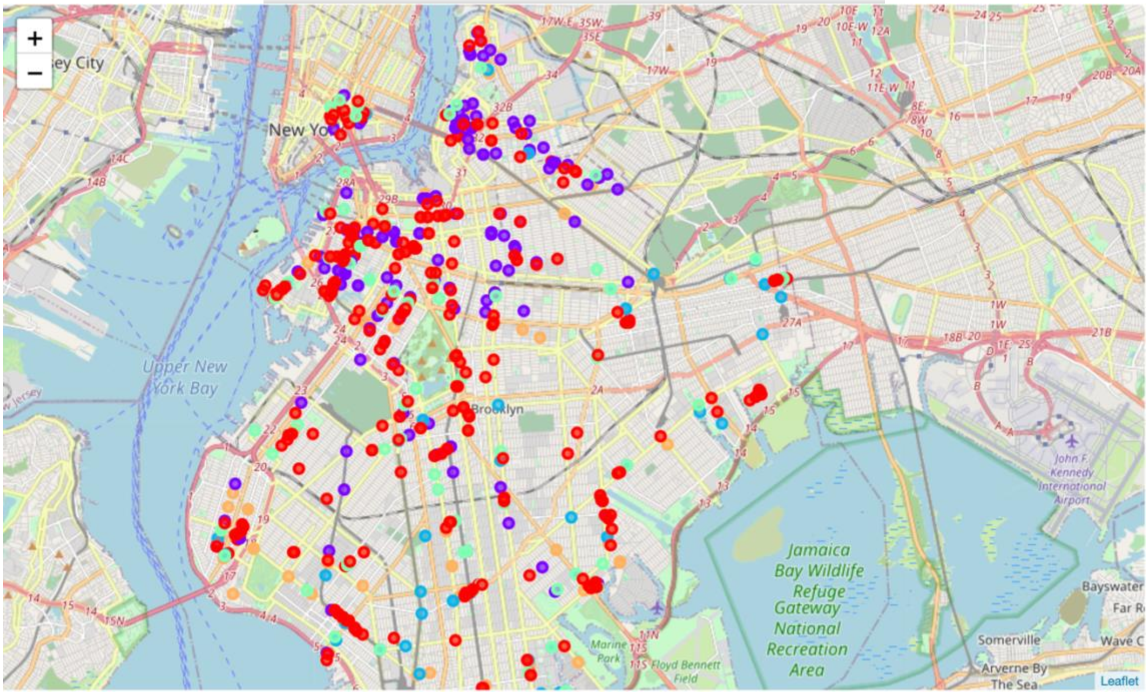
(figure 10)

After these preparation, I use the machine learning method of K-means in order to cluster the area of different 'shops' in the cities. The result can be used to identify the neighbors which has similar shops around in these two cities.

Result

It is shown on the below map that the distribution of shops in these two cities that figure is New York and figure 12 is Toronto. It is clearly that the distribution is more crowded in New York than Toronto. In figure 11, the purple display the 'coffee shop' as cluster 1, light green is the ‘ice cream shop’ as cluster 3, orange is the ‘bagel shop’ as cluster 4 and the red is the all the rest shops as cluster 0.

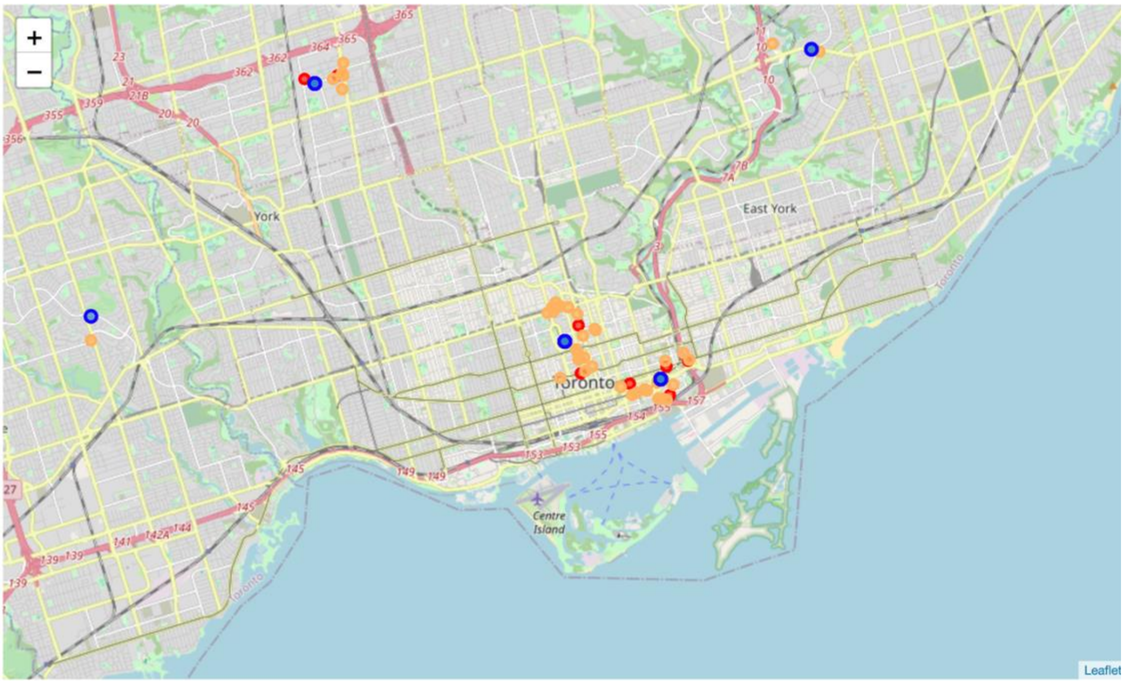
```
Hobby Shop                 2
Comic Shop                 2
Other Repair Shop          2
Motorcycle Shop            2
Shop & Service             2
Shopping Plaza             1
Print Shop                 1
Chocolate Shop             1
Souvenir Shop              1
Tailor Shop                1
Name: VenueCategory, dtype: int64
Coffee Shop    252
Name: VenueCategory, dtype: int64
Donut Shop    187
Name: VenueCategory, dtype: int64
Ice Cream Shop    125
Name: VenueCategory, dtype: int64
Bagel Shop    95
Name: VenueCategory, dtype: int64
```



(figure 11)

In the figure 12, the orange the rest of the shop as the cluster 0, blue is the ‘coffee shop’ as cluster 1, ‘cosmetic shop’ as cluster 2, the ‘shopping mall as cluster 3 and ‘ice cream shop’ as cluster 4.

```
Dessert Shop             3
Miscellaneous Shop       2
Bubble Tea Shop          2
Comic Shop               1
Cheese Shop              1
Hobby Shop               1
Gift Shop                1
Food & Drink Shop        1
Shop & Service           1
Fish & Chips Shop        1
Chocolate Shop           1
Gourmet Shop             1
Smoke Shop               1
Name: VenueCategory, dtype: int64
Coffee Shop     31
Name: VenueCategory, dtype: int64
Cosmetics Shop     4
Name: VenueCategory, dtype: int64
Shopping Mall     4
Name: VenueCategory, dtype: int64
Ice Cream Shop     3
Name: VenueCategory, dtype: int64
```



(figure 12)

Discussion

The map and analysis shown above is shows the distribution of shops around the city of Toronto and New York. It is clearly that, the shop distribution of these two cities are varied, that New York has more shops than Toronto. For example, if the person from New York would like to move to the Toronto and has as many as coffee shop as New York, it is better to live in the neighbor that the blue color highlighted.

Conclusion

It is worthwhile to see the difference in distribution of shops of these two cities if a person needs to re-locate within these two cities. It is clearly that if the person is accustomed to the happy hour of the New York with its flourish shops, it is better to move to the Toronto that is around downtown since it has the most crowded shops distribution there. The location selection is based on the person's needs.