PREDICT SIMILAR BETWEEN NEIGHBOURHOOD

I. Introduction

Background

Nowaday, many people must live far away from home for working, but they want to get a place similar to their old house/apartment. This is very important because selecting a uncomfortable place may effect to people's working, studying, sleeping and many things else. In present, many group, company created a lot applications which help user to select new accommodations quickly depend on customer's interesting: they may want to live near many restaurants, or gyms, or markets, milktea shop, coffeehouse…

Problem

But sometime, customers may not understand what they need or customers just need accommodations which has the most similar to their current house/apartment. In this point, I will create an application which create accordant places for living depend on user's current accommodations. When using this app, users just input their current accommodations and district-city they are going to live, user will get list of places which are similar to. In this project, I am going to use this instance: a person who live in Bronx, NewYork City, NY want to live in Manhattan, NewYork City, NY. Where is the most benefit neighbourhood for him?

II. Data description

To do solve this problem, I need informations about:

1. The popular venues in Bronx.
2. The popular venues each neighbourhood in Manhattan.
3. Dataset about neigbourhood in NewYork City.

About topic (1) and (2), I used Foursquare API to get the most venues of given places. Topic (3) is provided in this source: https://geo.nyu.edu/catalog/nyu_2451_34572

III. Methodology

1. Draw map each Borough

I use folium library for draw map about Manhattan's neighbourhood and Bronx's neighbourhood for detail and help user distinguishing them.

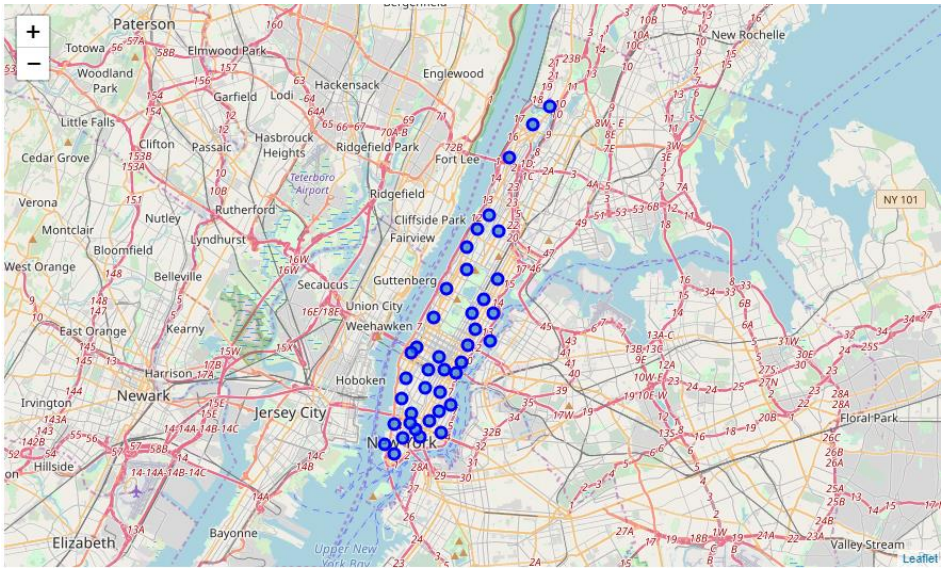

Figure 1 Manhattan map

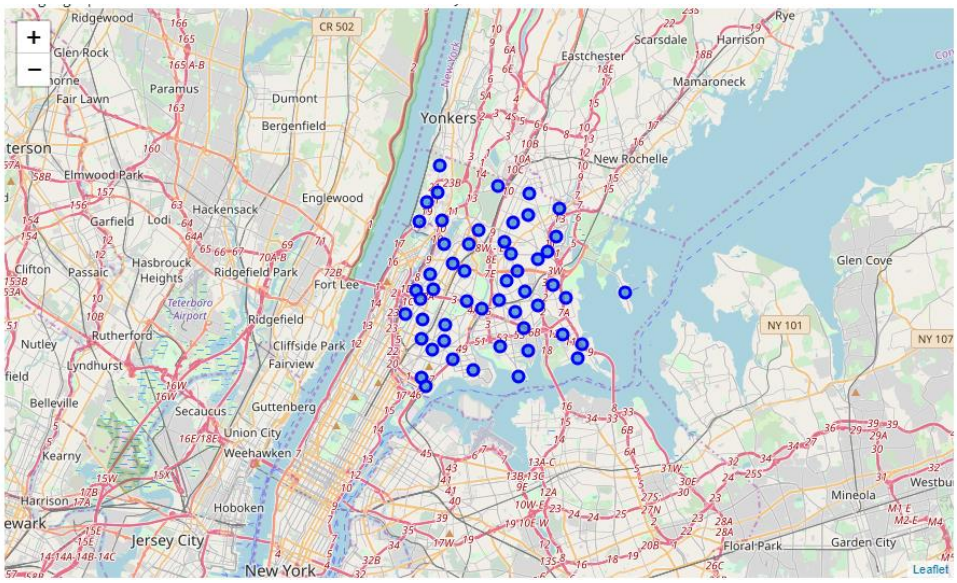

Figure 2 Bronx map

2. Data preproccessing

First, I get top 100 venues each neighbourhood in Bronx and Manhattan from Foursquare. Each venue has Latitude, Longtitude and its category. Then, I merge this table with information about neigbourhood in NewYork City in order to get coordinates of the center neigbourhoods, import it to dataframe. At this time, dataframe has 364 unique categories in venues. In nextstep, I use one-hot encoding for data description. Finally, my method groups dataframe by ['borough' and 'neigbourhood'] before use model for clustering.

	Borough	Neighborhood	Accessories Store	Adult Boutique	Afghan Restaurant	African Restaurant	American Restaurant	Antique Shop	Arcade	Arepa Restaurant	Argentinian Restaurant	Art Gallery	Art Museum
0	Bronx	Allerton	0.000000	0.00	0.00	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000
1	Bronx	Baychester	0.000000	0.00	0.00	0.000000	0.000000	0.00	0.041667	0.000000	0.000000	0.000000	0.000000
2	Bronx	Bedford Park	0.000000	0.00	0.00	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000
3	Bronx	Belmont	0.000000	0.00	0.00	0.000000	0.010204	0.00	0.000000	0.000000	0.000000	0.000000	0.000000
4	Bronx	Bronxdale	0.000000	0.00	0.00	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000
5	Bronx	Castle Hill	0.000000	0.00	0.00	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000
6	Bronx	City Island	0.000000	0.00	0.00	0.000000	0.034483	0.00	0.000000	0.000000	0.000000	0.034483	0.000000
7	Bronx	Claremont Village	0.000000	0.00	0.00	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000
8	Bronx	Clason Point	0.000000	0.00	0.00	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000
9	Bronx	Co-op City	0.000000	0.00	0.00	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000
10	Bronx	Concourse	0.000000	0.00	0.00	0.038462	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.038462
11	Bronx	Concourse Village	0.000000	0.00	0.00	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000
12	Bronx	Country Club	0.000000	0.00	0.00	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000
13	Bronx	East Tremont	0.000000	0.00	0.00	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000
14	Bronx	Eastchester	0.000000	0.00	0.00	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000
15	Bronx	Edenwald	0.000000	0.00	0.00	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000

Figure 3 Apart of Dataset after preproccessing

3. Modeling

My model is K-means cluster. After fitting, I get labels for each neigbouhood in 2 borough Bronx and Manhattan. If two neighbourhoods have same label, its would have more similar between.

4. Drawing map

Lately, I draw a map by folium library and describe neighbourhoods and its clusters in map.

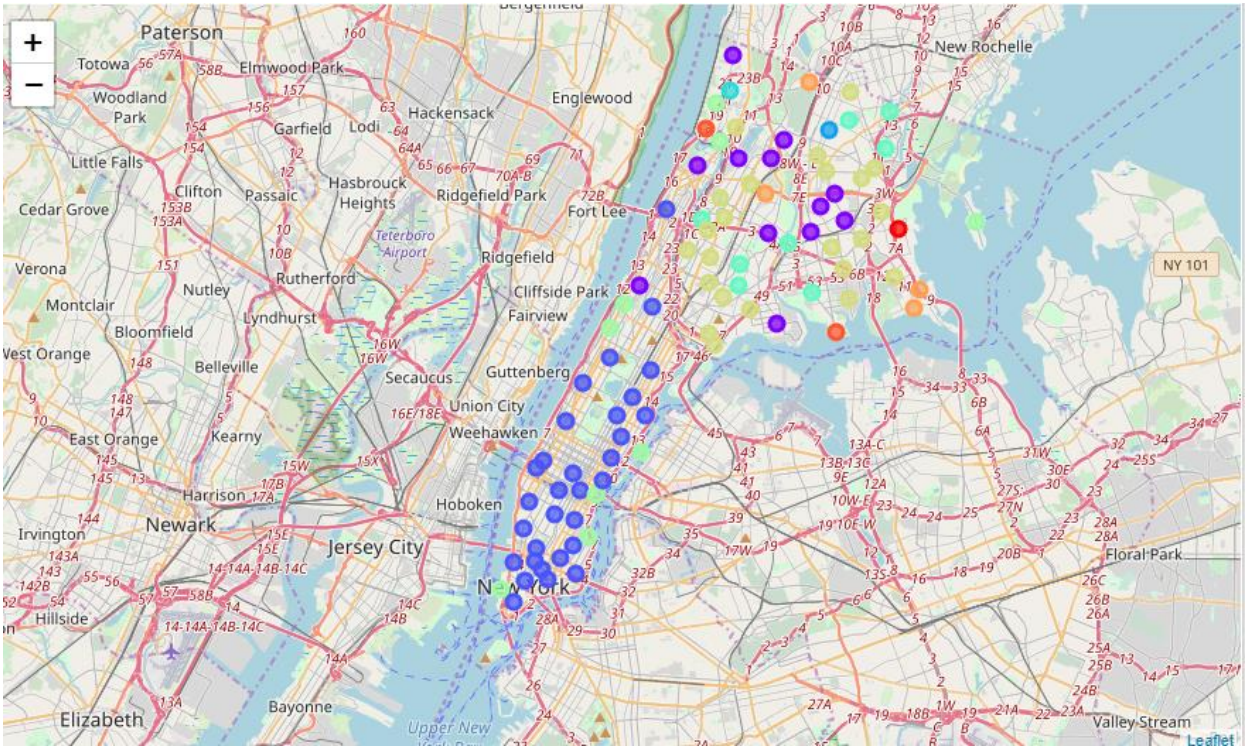

Figure 4 Map after clustering

IV. Result

Now, I can give recommend place for living when customers want to move from Bronx to Manhattan. Look at the map in III-4, circles which have the same color, have the same characteristics. For more details, users should use the list below. If they want to get the same environment place X as their old place, they should take them from the same cluster as place X.

Cluster_labels	Borough	Neighborhood	Latitude	Longitude
7	Bronx	Allerton	40.86579	-73.8593
7	Bronx	Baychester	40.86686	-73.8358
1	Bronx	Bedford Park	40.87019	-73.8855
8	Bronx	Belmont	40.85728	-73.8885
1	Bronx	Bronxdale	40.85272	-73.8617
7	Bronx	Castle Hill	40.81901	-73.848
6	Bronx	City Island	40.84725	-73.7865
5	Bronx	Claremont Village	40.83143	-73.9012
9	Bronx	Clason Point	40.80655	-73.8541

5	Bronx	Co-op City	40.87429	-73.8299
7	Bronx	Concourse	40.83428	-73.9156
7	Bronx	Concourse Village	40.82478	-73.9158
0	Bronx	Country Club	40.84425	-73.8241
1	Bronx	East Tremont	40.8427	-73.8874
5	Bronx	Eastchester	40.88756	-73.8278
5	Bronx	Edenwald	40.88456	-73.8481
8	Bronx	Edgewater Park	40.82199	-73.8139
4	Bronx	Fieldston	40.89544	-73.9056
7	Bronx	Fordham	40.861	-73.8964
7	Bronx	High Bridge	40.83662	-73.9261
1	Bronx	Hunts Point	40.80973	-73.8833
7	Bronx	Kingsbridge	40.88169	-73.9028
1	Bronx	Kingsbridge Heights	40.87039	-73.9015
7	Bronx	Longwood	40.8151	-73.8958
7	Bronx	Melrose	40.81975	-73.9094
5	Bronx	Morris Heights	40.8479	-73.9197
1	Bronx	Morris Park	40.84755	-73.8504
5	Bronx	Morrisania	40.82359	-73.9015
7	Bronx	Mott Haven	40.80624	-73.9161
7	Bronx	Mount Eden	40.84383	-73.9166
7	Bronx	Mount Hope	40.84884	-73.9083
1	Bronx	North Riverdale	40.90854	-73.9045
1	Bronx	Norwood	40.87722	-73.8794
7	Bronx	Olinville	40.87137	-73.8633
7	Bronx	Parkchester	40.83794	-73.856
7	Bronx	Pelham Bay	40.85064	-73.8321
7	Bronx	Pelham Gardens	40.86297	-73.8416
1	Bronx	Pelham Parkway	40.85741	-73.8548
7	Bronx	Port Morris	40.80166	-73.9132
6	Bronx	Riverdale	40.89083	-73.9126
7	Bronx	Schuylerville	40.82658	-73.8262
5	Bronx	Soundview	40.82101	-73.8657
9	Bronx	Spuyten Duyvil	40.88139	-73.9172
8	Bronx	Throgs Neck	40.81511	-73.8164
7	Bronx	Unionport	40.82977	-73.8505
7	Bronx	University Heights	40.85573	-73.9104
1	Bronx	Van Nest	40.84361	-73.8663
7	Bronx	Wakefield	40.89471	-73.8472
5	Bronx	West Farms	40.83948	-73.8777
7	Bronx	Westchester Square	40.84062	-73.8422
3	Bronx	Williamsbridge	40.88104	-73.8574
8	Bronx	Woodlawn	40.89827	-73.8673
6	Manhattan	Battery Park City	40.71193	-74.0169
2	Manhattan	Carnegie Hill	40.78268	-73.9533
2	Manhattan	Central Harlem	40.81598	-73.9432
2	Manhattan	Chelsea	40.74403	-74.0031

2	Manhattan	Chinatown	40.71562	-73.9943
2	Manhattan	Civic Center	40.71523	-74.0054
2	Manhattan	Clinton	40.7591	-73.9961
2	Manhattan	East Harlem	40.79225	-73.9442
2	Manhattan	East Village	40.72785	-73.9822
2	Manhattan	Financial District	40.70711	-74.0107
2	Manhattan	Flatiron	40.73967	-73.9909
2	Manhattan	Gramercy	40.73721	-73.9814
2	Manhattan	Greenwich Village	40.72693	-73.9999
1	Manhattan	Hamilton Heights	40.8236	-73.9497
2	Manhattan	Hudson Yards	40.75666	-74.0001
1	Manhattan	Inwood	40.86768	-73.9212
2	Manhattan	Lenox Hill	40.76811	-73.9589
2	Manhattan	Lincoln Square	40.77353	-73.9853
2	Manhattan	Little Italy	40.71932	-73.9973
2	Manhattan	Lower East Side	40.71781	-73.9809
2	Manhattan	Manhattan Valley	40.79731	-73.9643
6	Manhattan	Manhattanville	40.81693	-73.9574
6	Manhattan	Marble Hill	40.87655	-73.9107
2	Manhattan	Midtown	40.75469	-73.9817
2	Manhattan	Midtown South	40.74851	-73.9887
6	Manhattan	Morningside Heights	40.808	-73.9639
2	Manhattan	Murray Hill	40.7483	-73.9783
2	Manhattan	Noho	40.72326	-73.9884
6	Manhattan	Roosevelt Island	40.76216	-73.9492
2	Manhattan	Soho	40.72218	-74.0007
6	Manhattan	Stuyvesant Town	40.731	-73.9741
2	Manhattan	Sutton Place	40.76028	-73.9636
2	Manhattan	Tribeca	40.72152	-74.0107
6	Manhattan	Tudor City	40.74692	-73.9712
2	Manhattan	Turtle Bay	40.75204	-73.9677
2	Manhattan	Upper East Side	40.77564	-73.9605
2	Manhattan	Upper West Side	40.78766	-73.9771
2	Manhattan	Washington Heights	40.8519	-73.9369
2	Manhattan	West Village	40.73443	-74.0062
2	Manhattan	Yorkville	40.77593	-73.9471

V. Discussion

This is just a simple way to solve “Finding accommodations problems”. As you know, there are many independent variables which are not using in my method, such as prices of apartments, quality of atmosphere or characteristics of people living around you. This is the first method I building, and I am going to upgrade this method in the future.

VI. References