

Similar Neighbors in Two Citys

Weizhuang Peng

Moving to a new city

- Often times, we need to move to a new city and we want to move to a similar neighbors as the one we currently living in.
- By similar, I mean, maybe similar restaurant nearby, or a shopping mall nearby
- Ideas : can use the cluster algorithms in Machine Learning to group similar neighbors into the same cluster using there nearby venues information

Tools used to achieve this

- We need a map to explore the nearby venues of a city, in this project, we choose to use Foursquare API since it's has the least limitations. We also need the neighborhood information of Nashville and New York, which can be found online
- Since there are too many neighbors in New York City, we only choose one part, the Manhattan part

Some sample data

	Neighborhood	Latitude	Longitude
0	Antioch	36.060060	-86.672219
1	Bellevue	36.064782	-86.939446
2	Donelson	36.162557	-86.669997
3	East Nashville	36.172556	-86.759721
4	Germantown	36.279498	-86.873611
5	Green Hills	36.103670	-86.816666

This is the Nashville neighbors information : the neighborhood name and it's latitude and longitude used by the Foursquare API to get the nearby venues.

Some nearby venues in Nashville

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Antioch	36.060060	-86.672219	Mill Creek Greenway Trail	36.056369	-86.672999	Trail
1	Antioch	36.060060	-86.672219	Ford Ice Center	36.053014	-86.656621	Skating Rink
2	Antioch	36.060060	-86.672219	Casa Fiesta	36.045180	-86.663239	Mexican Restaurant
3	Antioch	36.060060	-86.672219	King Market, Laos Thai Cafe	36.071025	-86.685031	Asian Restaurant
4	Antioch	36.060060	-86.672219	Hai Woon Dai	36.068354	-86.684317	Korean Restaurant
...
1285	West Nashville	36.158946	-86.846389	Bordeaux Trailway	36.180683	-86.843791	Trail
1286	West Nashville	36.158946	-86.846389	Gridiron9	36.167602	-86.820488	American Restaurant
1287	West Nashville	36.158946	-86.846389	Robertson Ave Market	36.157412	-86.874422	Smoke Shop
1288	West Nashville	36.158946	-86.846389	Roll With It	36.145268	-86.871461	Smoke Shop
1289	West Nashville	36.158946	-86.846389	Centenial Park Volleyball	36.149752	-86.816906	Park

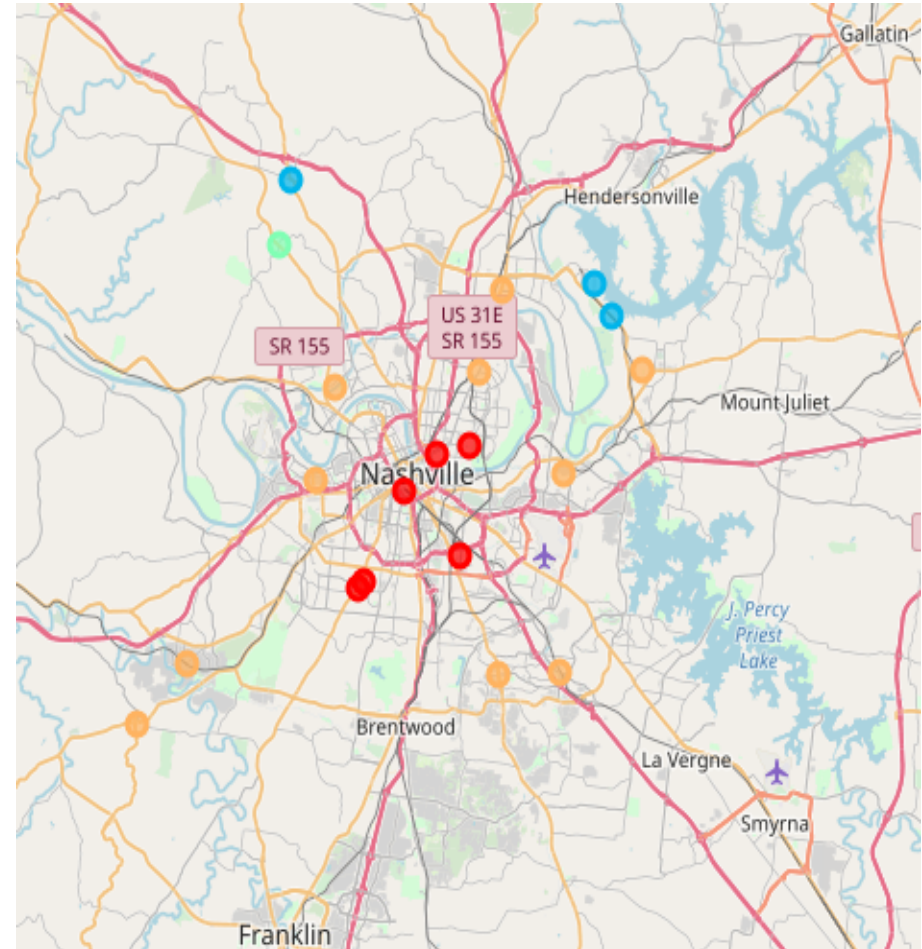
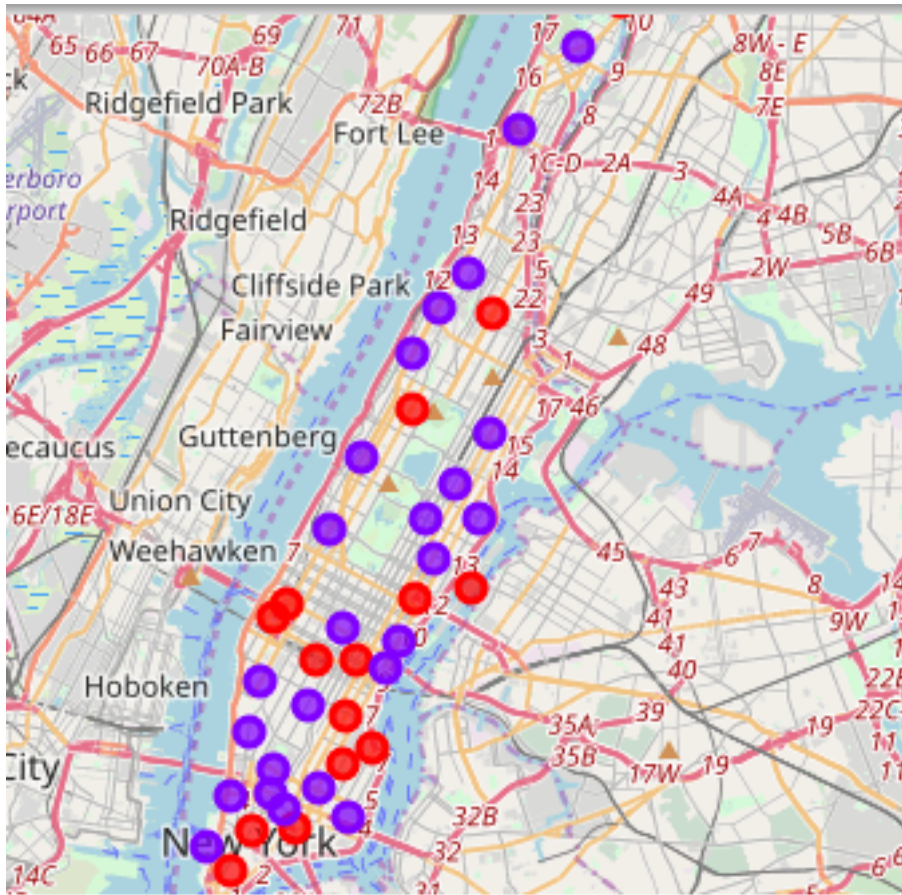
The above figure shows the dataframe showing the nearby venues in Nashville neighbors.

One hot encoding of venues

- To use the k means algorithms, we one-hot encoded the data and feed them into the k-means algorithm. The following figure is a snapshot of the averaged one-hot encoding result

	Neighborhood	Accessories Store	American Restaurant	Antique Shop	Arcade	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	Auto Workshop	...	Vegetarian / Vegan Restaurant	Video Game Store	Video Store	Vietnamese Restaurant
0	Antioch	0.000000	0.000000	0.000000	0.014706	0.000000	0.000000	0.000000	0.029412	0.000000	...	0.000000	0.000000	0.058824	0.000000
1	Battery Park City	0.000000	0.013514	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.000000	0.000000
2	Bellevue	0.000000	0.010526	0.000000	0.000000	0.000000	0.000000	0.021053	0.000000	0.000000	...	0.000000	0.010526	0.010526	0.000000
3	Carnegie Hill	0.000000	0.010989	0.000000	0.000000	0.000000	0.010989	0.000000	0.000000	0.000000	...	0.010989	0.000000	0.000000	0.021978
4	Central Harlem	0.000000	0.054054	0.000000	0.000000	0.027027	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.000000	0.000000
5	Chelsea	0.000000	0.037037	0.000000	0.000000	0.024691	0.000000	0.000000	0.012346	0.000000	...	0.012346	0.000000	0.000000	0.000000
6	Chinatown	0.000000	0.053333	0.000000	0.000000	0.000000	0.000000	0.000000	0.026667	0.000000	...	0.013333	0.000000	0.000000	0.040000
7	Civic Center	0.000000	0.037975	0.012658	0.000000	0.000000	0.000000	0.000000	0.012658	0.000000	...	0.000000	0.000000	0.000000	0.000000
8	Clinton	0.000000	0.048780	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.012195	0.000000	0.000000	0.000000
9	Donelson	0.000000	0.060000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.000000	0.000000

Results



Discussion

- In the previous slide, we showed the map for the clustered neighbors in Manhattan and Nashville, we found that there are some similar labels between the two maps, basically, downtown Nashville area are being labeled the same as many of neighbors in Manhattan. This is pretty reasonable.
- In the next slides we will have a look at the most frequent venues in some neighbors in the same cluster

Discussion

----East Nashville----

	venue	freq
0	Bar	0.13
1	American Restaurant	0.05
2	Music Venue	0.05
3	Hotel	0.04
4	Coffee Shop	0.03
5	Cocktail Bar	0.03

----Green Hills----

	venue	freq
0	Clothing Store	0.07
1	Cosmetics Shop	0.06
2	American Restaurant	0.05
3	Coffee Shop	0.05
4	Spa	0.05
5	Furniture / Home Store	0.05

----Marble Hill----

	venue	freq
0	Sandwich Place	0.10
1	Coffee Shop	0.10
2	Ice Cream Shop	0.05
3	American Restaurant	0.05
4	Department Store	0.05
5	Diner	0.05

The above three tables shows the three neighbors in Nashville (first two) and Manhattan (the third) with the first 6 most frequent venues nearby, we can see there are two are the same in both 3 neighbors. So this cluster results makes sense.