The Battle of Neighborhoods

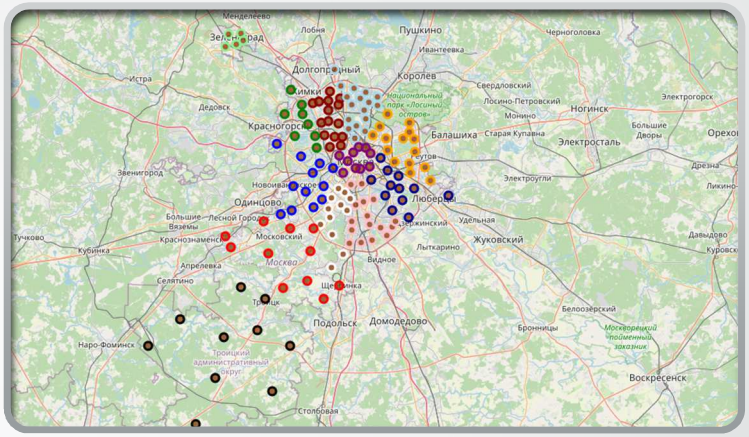
Introduction

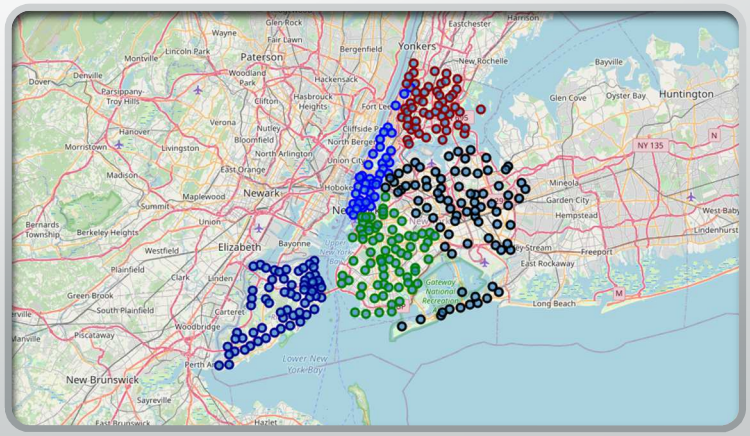
Sometimes we have to change our location because of many reasons: work, study, family… This project will help someone to understand how similar two areas from different cities are. As an example, we will try to compare two areas of two cities: Moscow and New York. Similarity of two areas can help to make a decision about migration from one city to another or about business expansion or just can provide some interesting information about two areas from different sides of the globe.

Data

All data for this project was taken from free and open sources on the internet. Most of initial information was scraped from Wikipedia, OpenStreetMap. Most detailed information about every location was obtained with Foursquare API. Data was cleaned and processed locally.

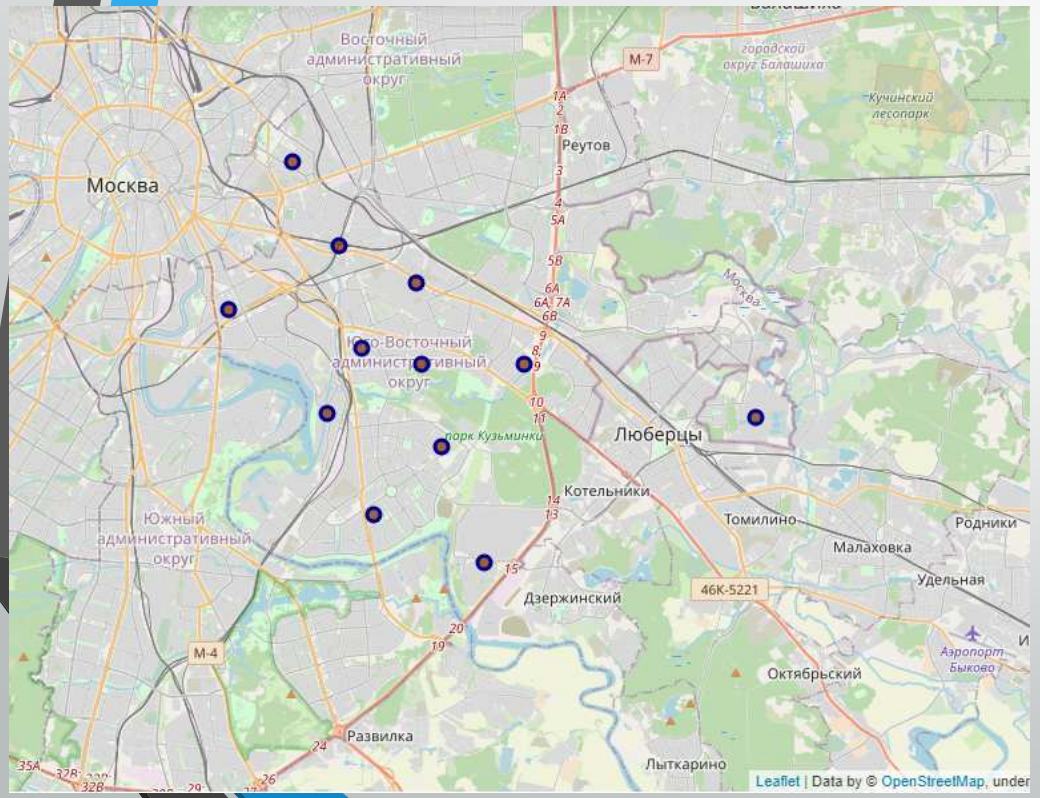
Everything begins from understanding territory division of every city. Moscow divided in 12 Administrative Okrug and New York City divided in 5 Borough. After getting some information about every Administrative Okrug and Borough we can decide which areas we will use for comparison. There are districts and neighborhoods inside of each Administrative Okrug and Borough.

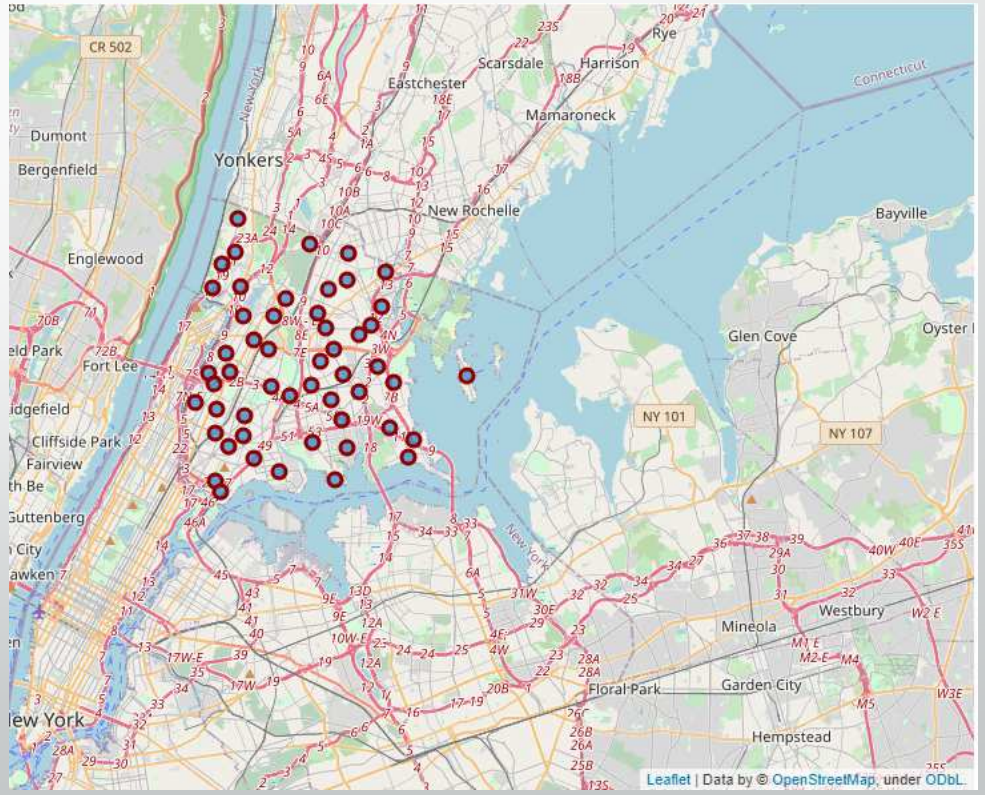




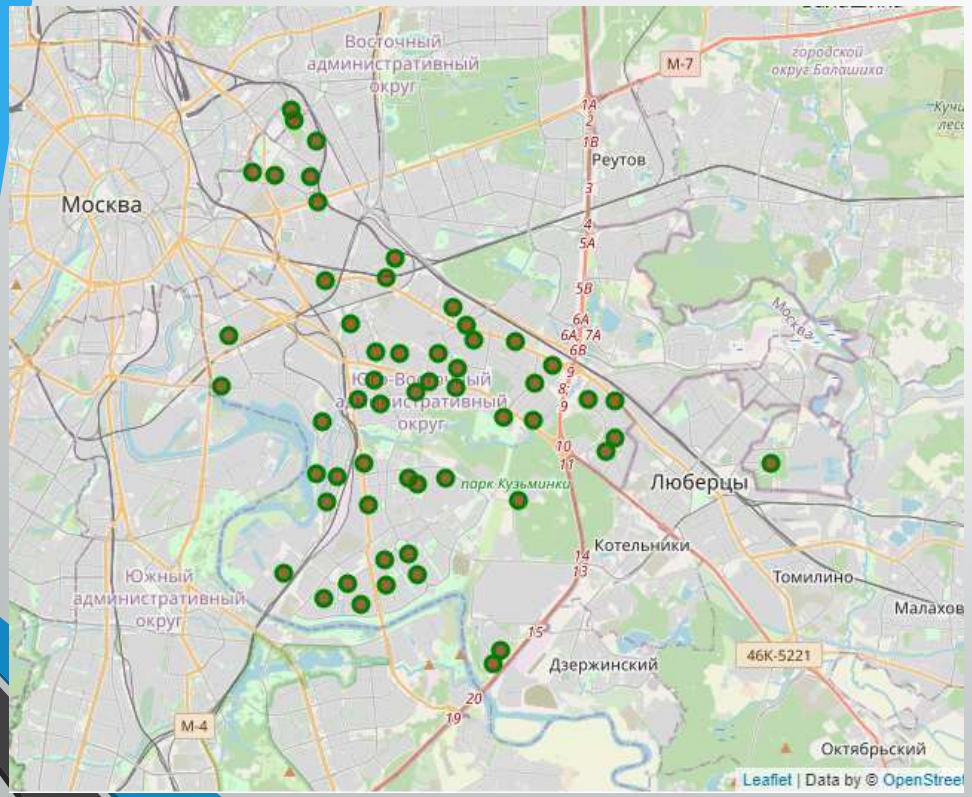
From Wikipedia data the most similar areas are:

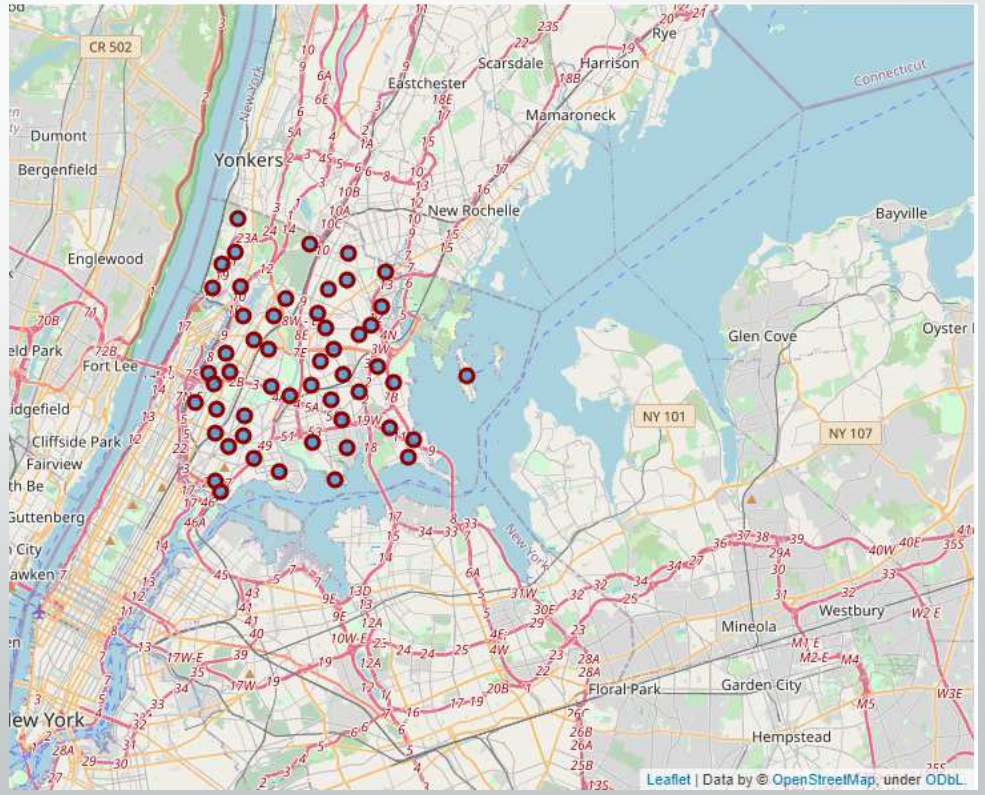
Name	Land area km²	% of all area	Place by area	Population	% of all population	Place by population	Density pers/km²	Place by density
South-Eastern	117.56	23833	6	1433828	11628	5	12196.59	4
The Bronx	109.04	13.91	4	1418207	43847	4	13006.00	3





South-Eastern area in Moscow has 12 districts and Bronx Area from New York has 52 neighborhoods. Comparison of 12 and 52 areas will not be fair. So let's use postal offices coordinates from these 12 districts in Moscow. It will let us to increase number of points inside study area.





Methodology

We are trying to compare two areas of biggest cities to understand how similar these areas are. Such information will be useful when you are trying to change living location to be sure that familiar things are nearby, or for example to understand is it possible to expand your business there - open a cafe, bar or gym.

By now we have collected some data about cities. We have learned about land area and population of districts inside cities. Based on this knowledge we have chosen two districts similar by land area and population.

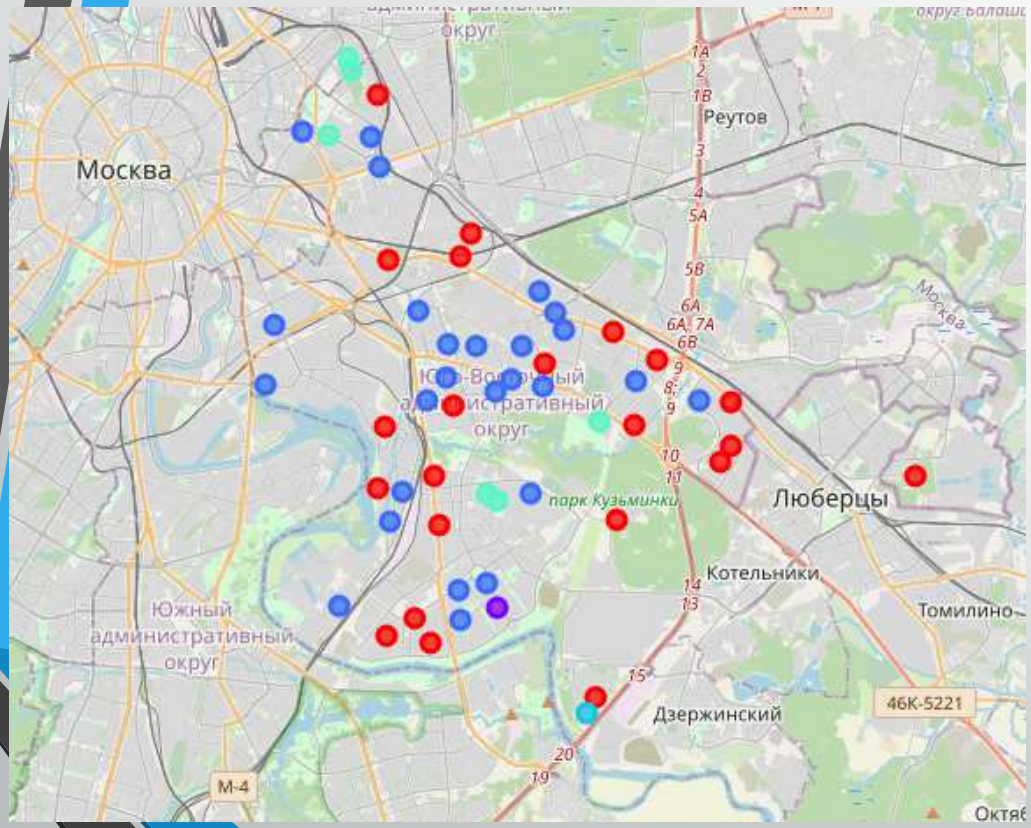
Next we will try to use Foursquare API to get more knowledge about chosen areas - venues around every point. Similarity of these areas will be evaluated after venue clusterization after getting data from Foursquare API.

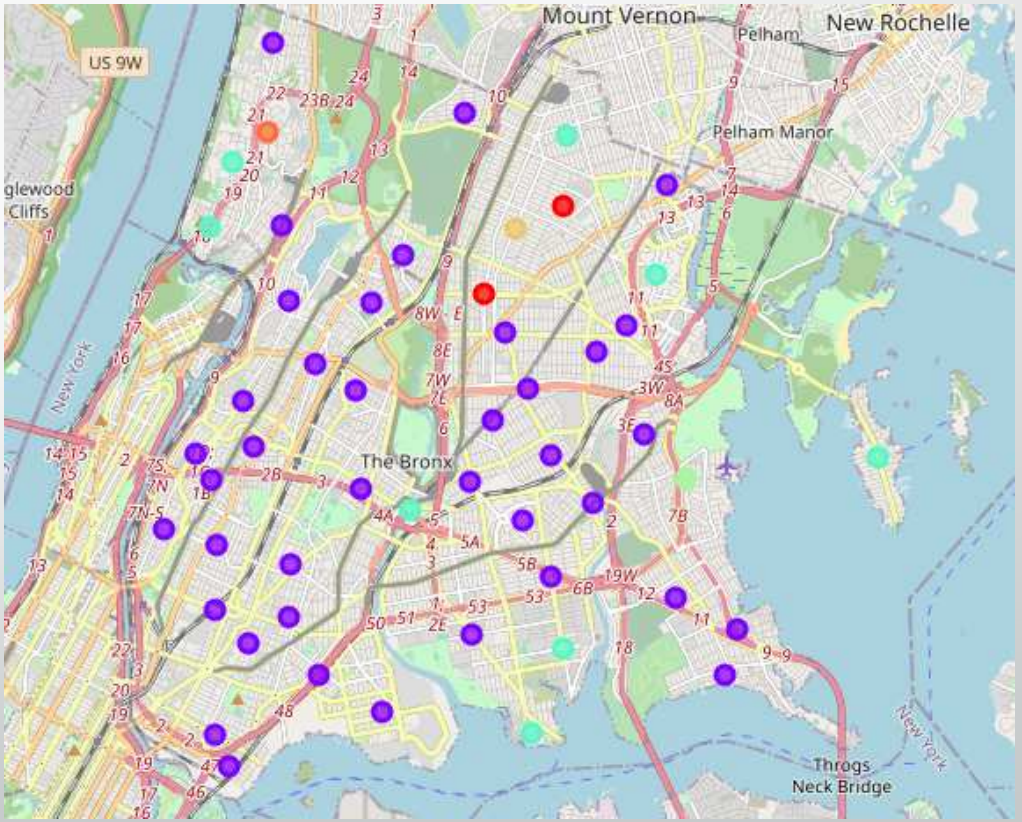
Analysis

Analysis of venues in 500 meters radius of every point gave us 243 unique categories. One hot encoding and grouping rows by neighborhood by taking the mean of the frequency of occurrence of each category, then sorting the venues in descending order and using only first five venues gives us such results:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Allerton	Pizza Place	Cosmetics Shop	Deli / Bodega	Bus Station	Supermarket
1	Baychester	Donut Shop	Spanish Restaurant	Gym / Fitness Center	Bus Station	Mattress Store
2	Bedford Park	Mexican Restaurant	Diner	Deli / Bodega	Pizza Place	Spanish Restaurant
3	Belmont	Italian Restaurant	Pizza Place	Deli / Bodega	Bakery	Bank
4	Bronxdale	Pizza Place	Bank	Performing Arts Venue	Paper / Office Supplies Store	Chinese Restaurant

Clusterization time: we will use k-means clustering with number of clusters equal to 8 (almost random number, not so big and not so small).





If we will try to compare results by colors - we can see that these two areas almost have nothing similar.

Most popular clusters in NY are 1 and 4 (and it's 39+8=47 places (47/52*100% = 90,4% of all)) and most popular clusters in Moscow are 2 and 0 (it's 26+22=48 places (85,7% of all), very similar quantity to NY).

Comparison of top 15 places of both clusters for every city will look like this:

New York City		Moscow	
Cluster	Q-ty	Cluster	Q-ty
1	39	2	26
4	8	0	22
0	2	4	6
7	1	3	1
6	1	1	1
5	1		

	Moscow Place	Moscow counts	NY Place	NY counts
0	Supermarket	29	Pizza Place	28
1	Pizza Place	11	Deli / Bodega	14
2	Gym / Fitness Center	11	Pharmacy	13
3	Food & Drink Shop	9	Donut Shop	12
4	Cosmetics Shop	7	Grocery Store	11
5	Fast Food Restaurant	7	Bus Station	11
6	Park	7	Bank	10
7	Bus Stop	7	Spanish Restaurant	7
8	Convenience Store	6	Italian Restaurant	6
9	Pharmacy	6	Sandwich Place	6
10	Mobile Phone Shop	5	Park	6
11	Café	5	Fast Food Restaurant	6
12	Sushi Restaurant	5	Chinese Restaurant	6
13	Coffee Shop	4	Supermarket	5
14	Soccer Field	4	Bakery	5

Results and discussion

Achieved results show us that two areas a very different with only 30% of similar places in top 15 places of each area. They are different after comparison of most common places in almost the same quantity of small areas inside of each big area.

Conclusion

As we can see - two areas with similar land area, population and density from two biggest cities from different sides of the globe are very different. From top 15 places of more than 85% of all common places for these areas only pizza places, pharmacy, bus stops, grocery stores and parks are present in both areas in different proporions. Such similary will allow you to eat pizza, buy pills and go to park if you will decide to change your location. But you will not be able to visit fitness center and cosmetics shop if you will migrate from Moscow to NYC, and you will not be able to buy donuts and visit spanish restaurant in case of NYC to Moscow migration. In conclusion I would like to say that location change from South-Eastern part of Moscow to The bronx in New York or vice versa will not be the easiest thing with familiar places in not familiar distance. But in the same time it's an ability for donut expansion in Moscow or ability to open a gym in New York.