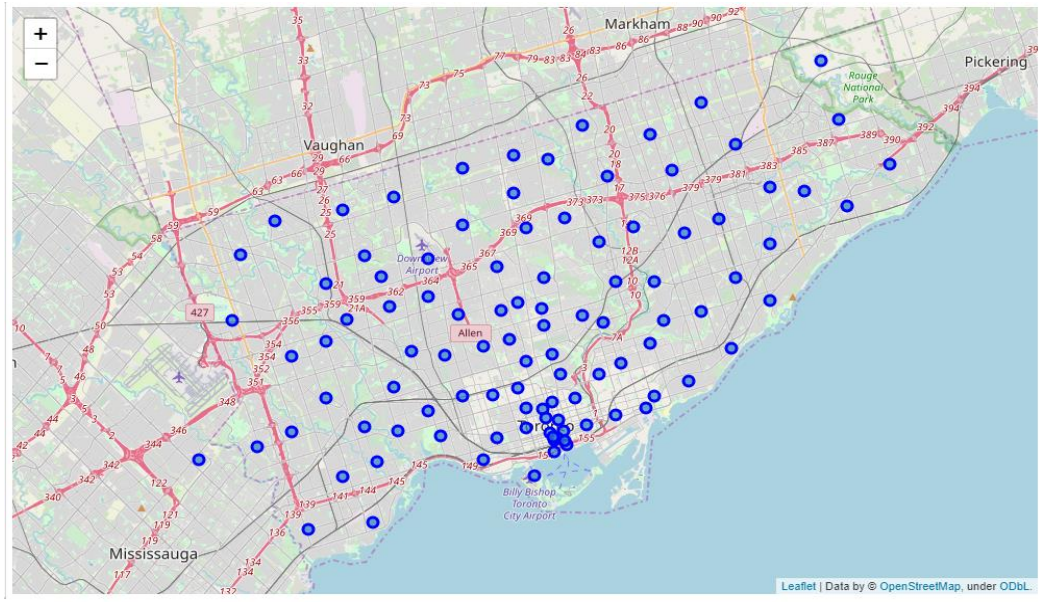
Battle of Neighborhoods

by Nikolay Krysko

This project contains:

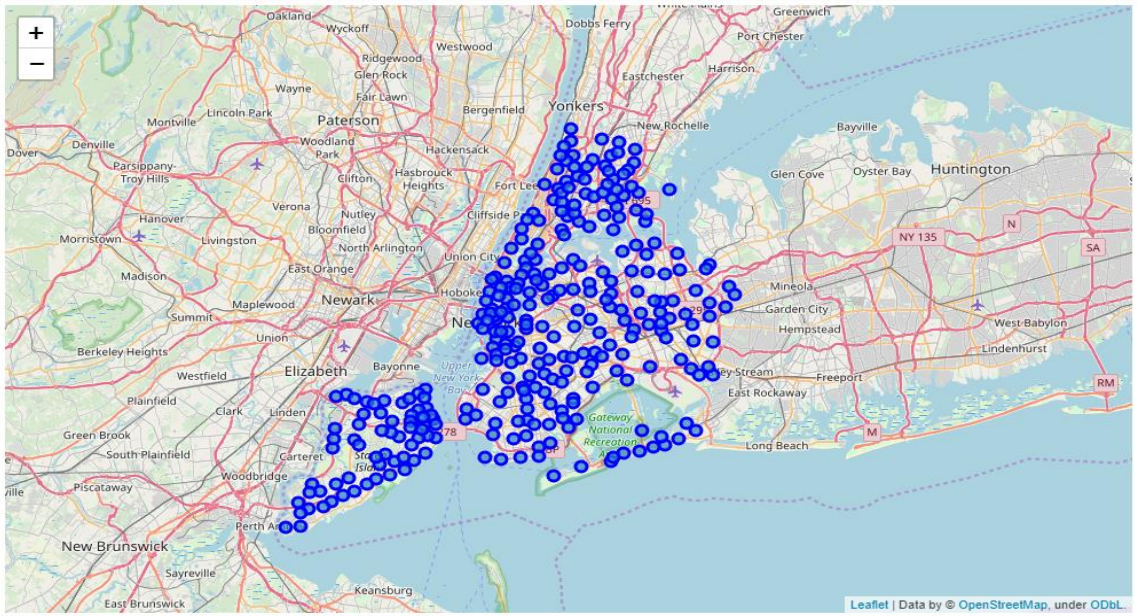
- Getting the data from Internet about the neighborhoods at Toronto and New York;
- Data preparation using Python Pandas library;
- Getting the neighborhoods venue data from Forsquare API;
- Prepare the data to clustering;
- Clustering the data by k-meant machine learning algorithm (k=5) using Python Scikit - learn library;
- Visualize clusters on map using Python Matplotlib anf Folium libraries;
- Analyze cluster sizes.

Toronto

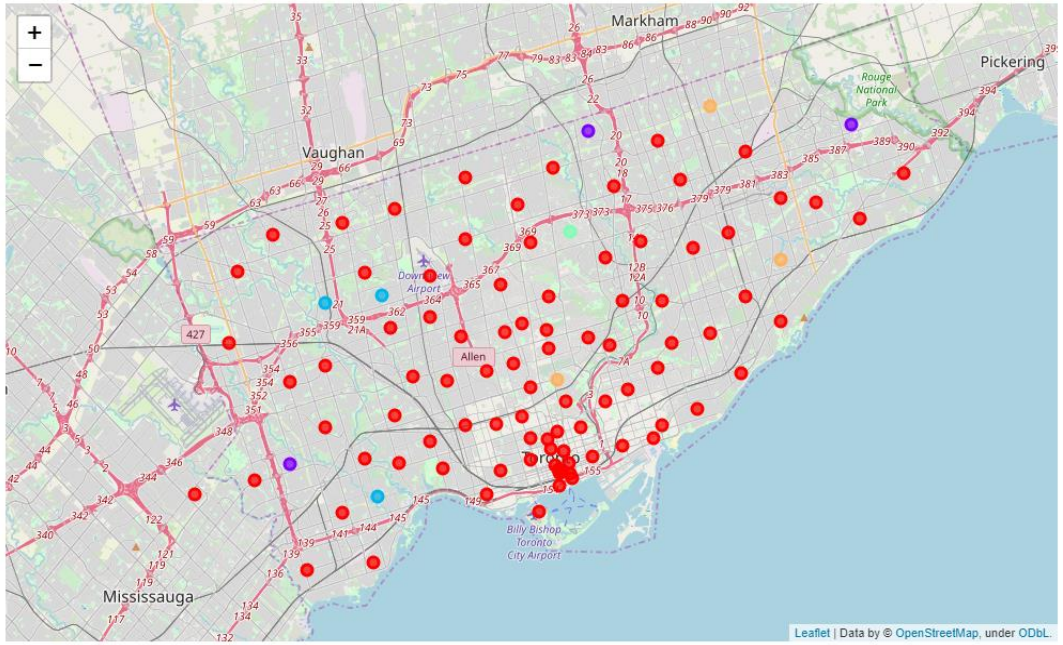


Before clustering

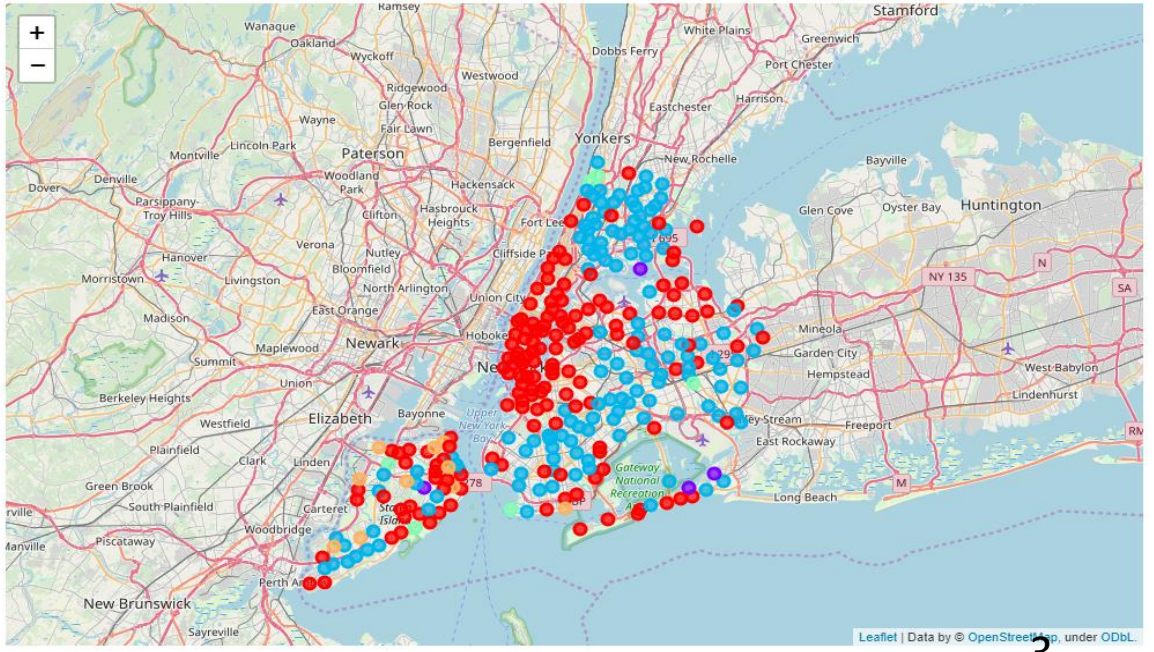
New York



Before clustering



After clustering



After clustering

Results

Number of cluster	Toronto		New York	
	Amount of Neighborhoods	Percentage	Amount of Neighborhoods	Percentage
1	90	90.0%	149	48.9%
2	3	3.0%	134	43.9%
3	3	3.0%	12	3.9%
4	3	3.0%	6	2.0%
5	1	1.0	4	1.3%

Discussion

It can be seen that Toronto has one big cluster (90.0% of the neighborhoods) and other are much smaller (about 3.0%). For New York, there are two big clusters (48.9% and 43.9% of the neighborhoods) and other are much smaller (about 3.0% too). Segmentation of two cities are different. Toronto has more uniform neighborhood type. New York has much more varieties.

Conclusion

As it was said in discussion there are two big clusters for New York and one large cluster for Toronto. If you are going to visit two cities and don't know what to visit first it is better to first visit Toronto because it has more uniform neighborhood type as were determined in this report.