

Finding of differences between Toronto and Manhattan

Description of the problem:

Goal of this research is to find differences or similarities across Toronto and Manhattan.

For solving this question the following datasources being used:

- Data on NY neighborhoods from Week 3 lab
- Data on Toronto from Week 3 assignment
- Data from Foursquare on types of venues in both regions.

This project could be interesting for individuals travelling to respective regions or to the individuals searching for employment in the regions. Comparison of the cities specifics could be additional pros or cons.

Methodology

As step 1 of the research we need to obtain data on both regions.

I will do this by loading the respective data and transforming it to dataframes.

I will utilize power of folium to show initial location of each neighborhood and clasterization of the later stages.

Having the data obtained we will append it with the Foursquare's data on type of venues in the neighborhoods of each region.

To do k-means clustering algorithm on the data we need to:

- do one hot encoding with mean frequency of occurrence of each category of venues in the cluster;
- take top-10 venues in each neighborhood;
- on this data do the clustering algorithm;

Results

Table below represent fraction of venues type from all the venues.

	Toronto	Manhattan
Coffee Shop	8.834586	4.620462
Café	5.639098	2.541254
Restaurant	3.508772	NaN
Italian Restaurant	2.631579	3.300330
Park	2.130326	2.376238
Japanese Restaurant	2.067669	NaN
Hotel	1.942356	2.046205
Bakery	1.817043	2.112211
Bar	1.691729	NaN
Gym	1.629073	1.815182
American Restaurant	NaN	2.013201
Gym / Fitness Center	NaN	1.947195
Pizza Place	NaN	2.541254

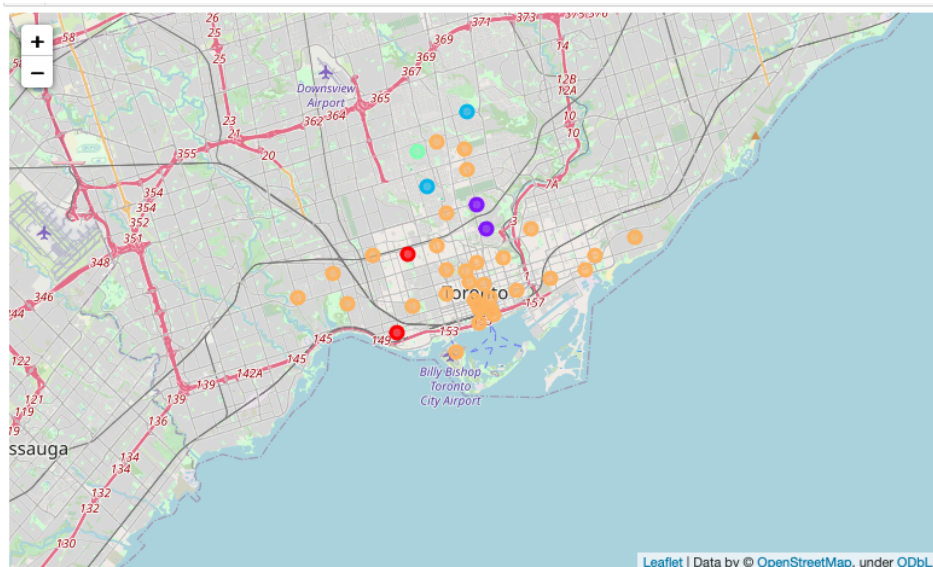
As could be seen from the table above, Toronto has more Coffee Shops and Cafe as percentage from total categories value, therefore we could assume that Toronto's citizen like to drink more coffee.

Additionally Bars and Japanese restaurants are not as popular on Manhattas as in Toronto which could be an indicator of different structure of the population.

On the opposite side, number of Gym and Fitness Centres are higher on Manhattan => healthier lifestyle could be expected.

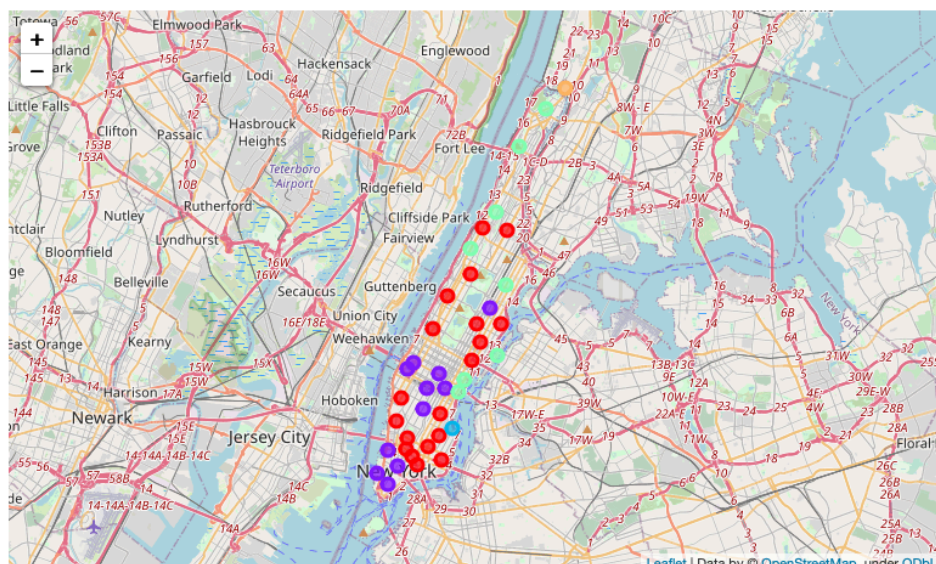
However both of the prepositions requires additional research.

Clusters of Toronto



Majority of neighborhoods are belong to "cluster 4" (marked orange) which characterized by Coffee Shop, Cafe or Bars.

Clusters of Manhattan



Conclusion

Presenter research clearly shown differentiation of regions. If you prefer to drink coffee and go to bars, you should probably take Toronto.

On the opposite, if you prefer healthy lifestyle, Manhattan is your choice.

Within the scope of this research.