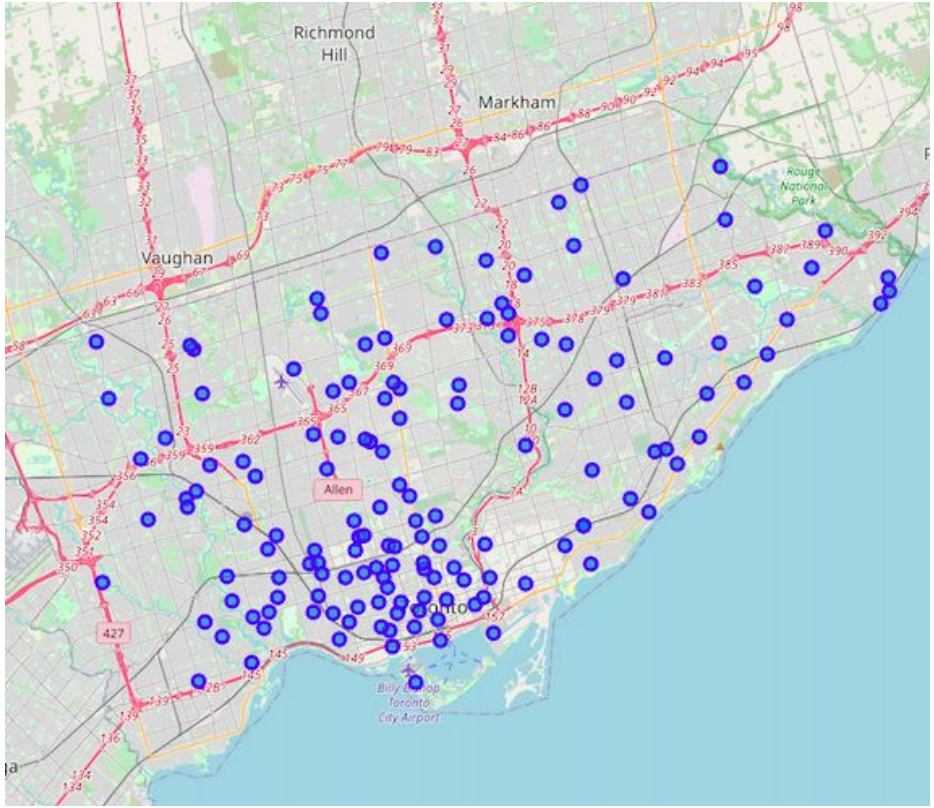
IBM_C9_Capstone_Toronto_restaurant

Boris Yushenkov, 2021-02, St-Petersburg, Russia

Table Of Contents



The IDEA

Let's assume that someone is looking to open a restaurant in Toronto, Ontario, Canada. What exact location should I recommend this person to choose?

A simple answer is to find a specific Neighborhood where:

- market segment competition is lowest,
- population density is highest,
- average income is highest.

The idea of the Project is to collect nesessary data and to use K-means approach to analyse the location selected.

DATA

"Demographics_of_Toronto_neighborhoods" wiki site: there you can find a lot of interesting data for each Toronto neighborhood. A list of these datatypes can be found as the names of the columns in the table below. Geolocation of neighborhoods was carried out through the address string, using the "geocode" method from **"geolocator"** library.

The second data source was **Forsquare.com**. With "food" search query one can get a plenty of information. After deleting matches, a list of 684 venues remains. While cleaning dataset we left only id, name, category and location of venues.

	Neighborhood	Population	Land area (km2)	Density (people/km2)	Change in Population since 2001 (%)	Average Income (CAD)	Transit Commuting %	Renters (%)	Second language after English	Latitude	Longitude
0	Agincourt	44.577	12.45	3580.0	4.6	25.750	11.1	5.9	Cantonese (19.3%)	43.785353	-79.278549
1	Alexandra Park	4.355	0.32	13609.0	0.0	19.687	13.8	28.0	Cantonese (17.9%)	43.650787	-79.404318
2	Allenby	2.513	0.58	4333.0	-1.0	245.592	5.2	3.4	Russian (1.4%)	43.712849	-79.547065
3	Amesbury	17.318	3.51	4934.0	1.1	27.546	16.4	19.7	Spanish (6.1%)	43.706162	-79.483492
4	Armour Heights	4.384	2.29	1914.0	2.0	116.651	10.8	16.1	Russian (9.4%)	43.743944	-79.430851
...	...	...	...	...	...	...	...	...	...	...	...

METHODOLOGY

A very simple approach was used in this research. **The choice of the restaurant location in the first approximation depends on two things**:

- competition in the nearest environment
- purchasing power of neighborhood's residents

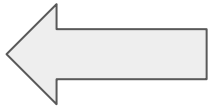
	Clusters	Neighborhood	PP	Latitude	Longitude
2	2	Allenby	1064150.0	43.712849	-79.547065
7	2	Bay Street Corridor	1766744.0	43.668865	-79.389126
24	2	Church and Wellesley	917152.0	43.665524	-79.383801
32	2	Deer Park	838272.0	43.688090	-79.394094
88	2	North York City Centre	1278415.0	43.739396	-79.513131
114	2	St. James Town	1424574.0	43.669403	-79.372704
145	2	Yorkville	1136055.0	43.671386	-79.390168

The level of population purchasing power **was estimated using the k-means method** in the three-dimensional space of parameters:

- the first two parameters are the latitude and longitude of the neighborhood location
- the third parameter is the purchasing power of the population of the neighborhood, obtained as the product of the density of residents per square kilometer and their average income.

Based on the results of k-means method, the resulting clusters were grouped according to the average purchasing power of all neighbors in the cluster in descending order.

Here you can see the Neighborhoods listed in the prime classifier, i.e. classifier with the largest value of persuasing power:



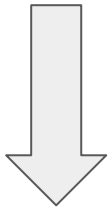
RESULTS

The analysis shows that despite the large number of restaurants in Toronto (684), there are two neighborhoods in which the density of restaurants is low, and the purchasing power of the population in autumn is high in comparison with other neighborhoods. This neighborhood are Allenby and North York City Center. They are located in the North-Wester part of the map, shown below.

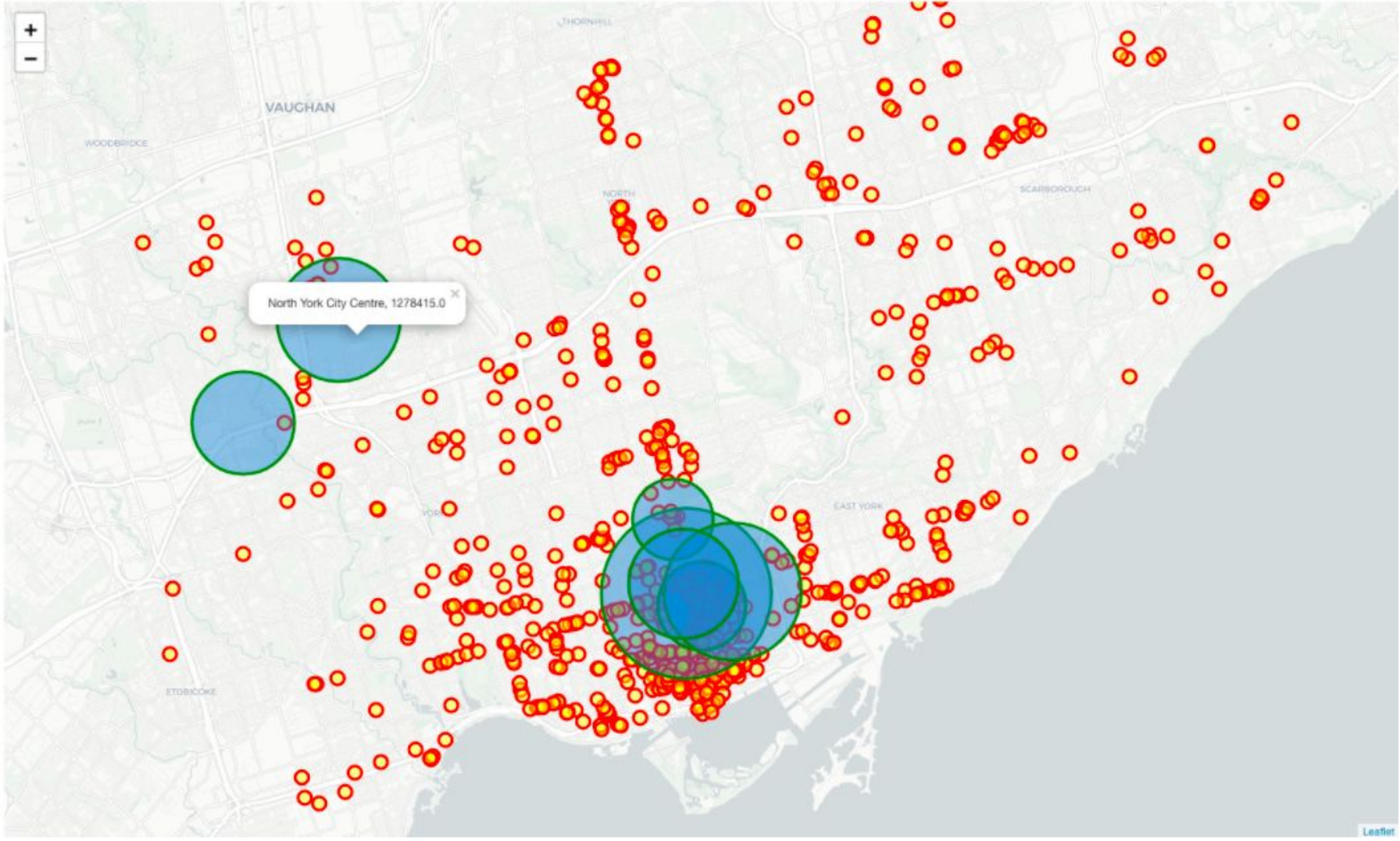
My recommendation would be to locate the new restaurant in these neighborhoods as the most suitable for this purpose.

They are:

- Allenby
- North York City Center



MAP ON THE NEXT PAGE


VAUGHAN
WOODBRIDGE
THORNHILL
NORTH
SCARBOROUGH
EAST YORK
ETOBICOKE
North York City Centre, 1278415.0
Leaflet