Capstone Project - The Battle of the Neighborhoods (Week 2)

Applied Data Science Capstone by IBM/Coursera

Created by RESHETNIKOV Ruslan

June 27, 2021

Introduction

In such a large and multicultural city like Toronto, it is very difficult to choose a place to open a restaurant. People have different preferences. The level of competition differs from area to area. Data Science will help us do this

Business Problem

For an investor who would like to invest their money in the restaurant business, it may be important to know the level of competition in the area where he plans to open a restaurant. Using open data, it is possible to determine the number of restaurants per 1000 people. I will be using Toronto's public statistics. I will determine the population of the city districts. Then I will pull data from Foursquare to determine the number of restaurants. Using this data, you can determine the level of competition in each borough of Toronto. This will allow the investor to more easily decide in which district of the city to open a restaurant.

Data

Download the data from http://map.toronto.ca/wellbeing/ The downloaded file contains data about

- the neighbourhood
- the population

in these neighbourhood. The Geopsy service was used to determine the coordinates of the city's districts. Geopy is a Python client for several popular geocoding web services. Geopy makes it easy for Python developers to locate the coordinates of addresses, cities, countries, and landmarks across the globe using third-party geocoders and other data sources.

Next, we combine this data with the data that we received using the Foursquare API.

	Neighborhood	Neighbourhood Id	Total Population
0	West Humber-Clairville	1.0	33312.0
1	Mount Olive-Silverstone-Jamestown	2.0	32954.0
2	Thistletown-Beaumond Heights	3.0	10360.0
3	Rexdale-Kipling	4.0	10529.0
4	Elms-Old Rexdale	5.0	9456.0
...	...	...	...
135	West Hill	136.0	27392.0
136	Woburn	137.0	53485.0
137	Eglinton East	138.0	22776.0
138	Scarborough Village	139.0	16724.0
139	Guildwood	140.0	9917.0

We connect to the geopsy server to determine the coordinates.

We attach two Datagrams and remove the lines containing NaN. We determine the coordinates of the canter of Toronto for the visualization of the Folium map.

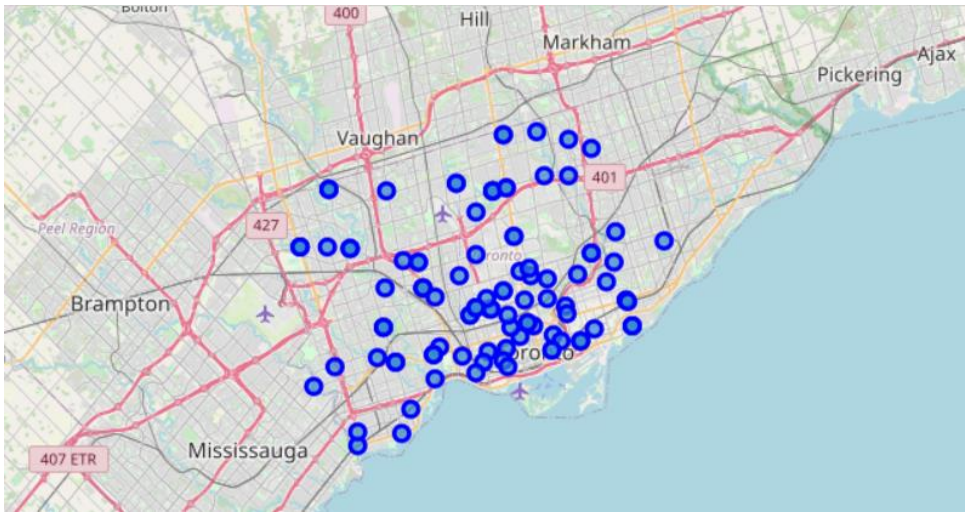

Figure 1. Neighbourhoods of Toronto

Methodology

In this project, we will focus our efforts on finding the areas of Toronto with the fewest restaurants per 1000 residents of the district. At the first stage, we collected the necessary data: the location and type (category) of each Toronto restaurant. We have identified all places of public catering, including such as cafeterias, pizzerias and bistros (according to the Foursquare classification). The second step in our analysis will be to calculate and study the "density of restaurants" for every 1000 people living in different areas of Toronto - we will use Folium maps to determine the Top 10 promising areas with a small density of restaurants per 1000 people. At the third and last stage, we will visualize a map of the most promising areas and within them we will create clusters of locations that meet some basic requirements. We connect to Foursquare to get the data of venue.

We select Venue Category containing words: Restaurant, Sandwich, Coffee, Pizza, Bar, Diner, Café, Breakfast, Pub, Cafeteria and Bistro.

Analysis

Let's calculate how many public catering places there are in each district of the city of Toronto.:

Neighborhood	Venue Category
Agincourt South-Malvern West	39
Alderwood	9
Annex	27
Bay Street Corridor	2
Bayview Woods-Steeles	4
...	...
Woodbine Corridor	6
Woodbine-Lumsden	3

Wychwood		10
Yonge-St.Clair		27
York University Heights		5

We calculate the number of restaurants for 1000 people. To do this, you need to divide the number of residents of the district by 1000 (pop_div1000). Then we divide the number of public catering places in the area by pop_div1000and get the required values per1000

Using sorting, we get the Top 10 districts with the smallest density of restaurants per 1000 people.

	Neighborhood	Total Population	per1000
46	Mimico	33964.0	0.029443
13	Church-Yonge Corridor	31340.0	0.031908
24	Flemingdon Park	21933.0	0.045593
19	Downsview-Roding-CFB	35052.0	0.057058
77	Westminster-Branson	26274.0	0.076121
3	Bay Street Corridor	25797.0	0.077528
23	Etobicoke West Mall	11848.0	0.084402
51	New Toronto	11463.0	0.087237
8	Black Creek	21737.0	0.092009
45	Markland Wood	10554.0	0.094751

Average number of restaurants per 1,000 people in Toronto: 0.8856235373773195

Let's build a plot.box() of the number of restaurants per 1000 people

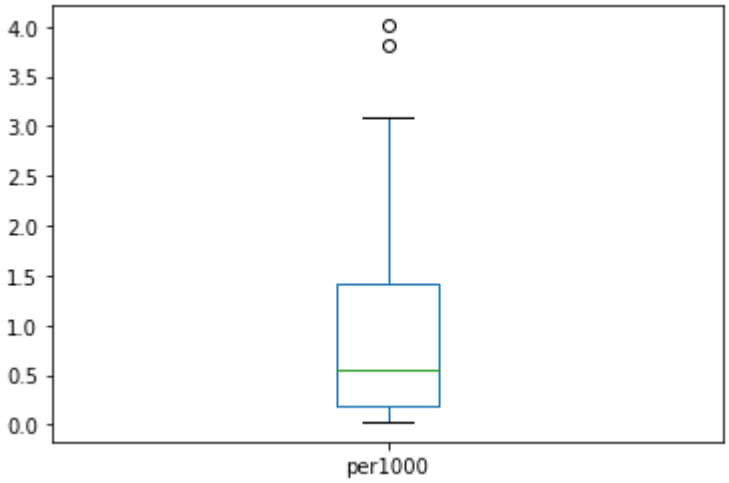

Figure 2. Box plot of the number of restaurants per 1000 people.

Now we will attach the Datafram nd1 with coordinates for further visualization on Folium maps.

For a more correct display of the size of circles with the number of restaurants per 1000 people, we will create a new column per1000minus, the values in which we will calculate by subtracting from the maximum possible number of restaurants (that is, 4) the total number of restaurants per1000. Thus, we make an inversion of the per1000 value.

Now we will perform clustering of data on the Folium map:

- The circles of green color show areas in which the number of restaurants per 1000 people is less than the average throughout the city of Toronto.
- The circles of red color show areas in which the number of restaurants per 1000 people is more than the average throughout the city of Toronto.

- The radius of the circles is formed from the per1000minus column, that is, the larger the radius, the more attractive the area for opening a restaurant.

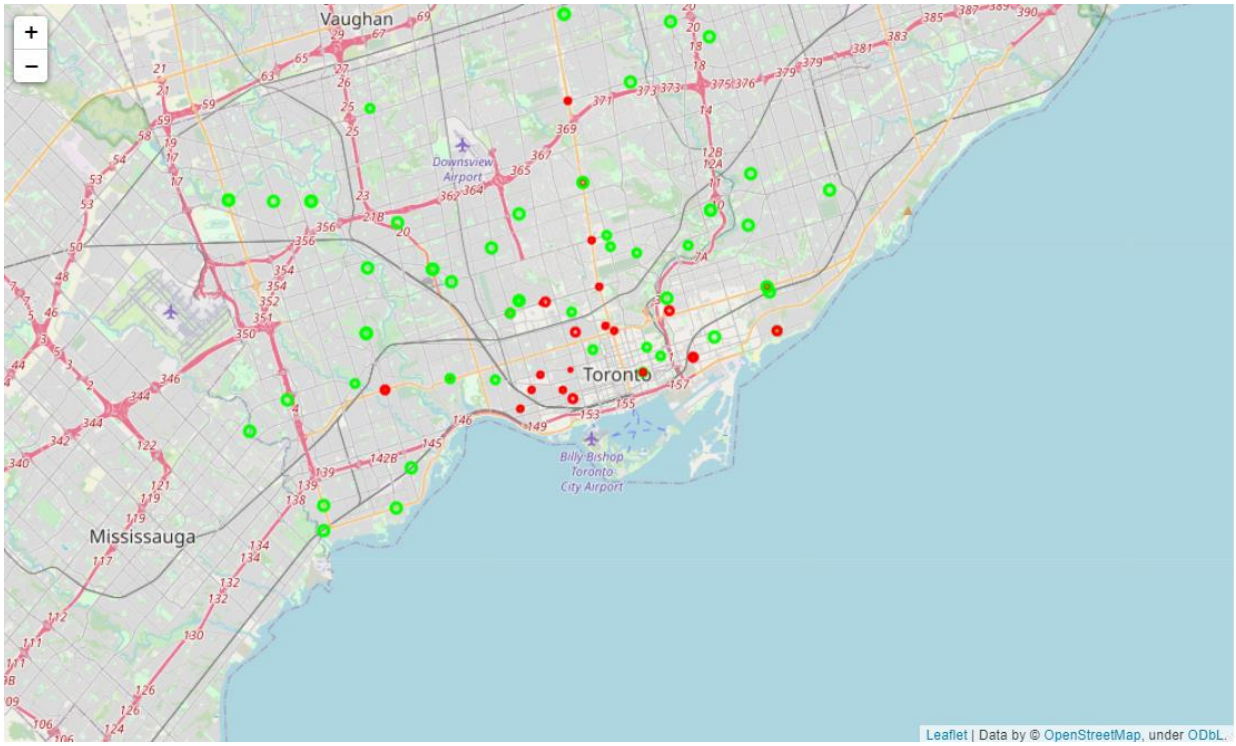

Figure 3. A map of Toronto neighbourhoods showing the number of restaurants per 1000 people.

Hierarchical Clustering

Determine which areas of Toronto are most similar, that is, how you can group the areas, which areas are most similar to each other in terms of the number of restaurants per 100 people. Our goal here is to use clustering techniques to find the most distinctive groups of Toronto neighborhoods. This will help investors make a decision about opening a restaurant in a certain area.

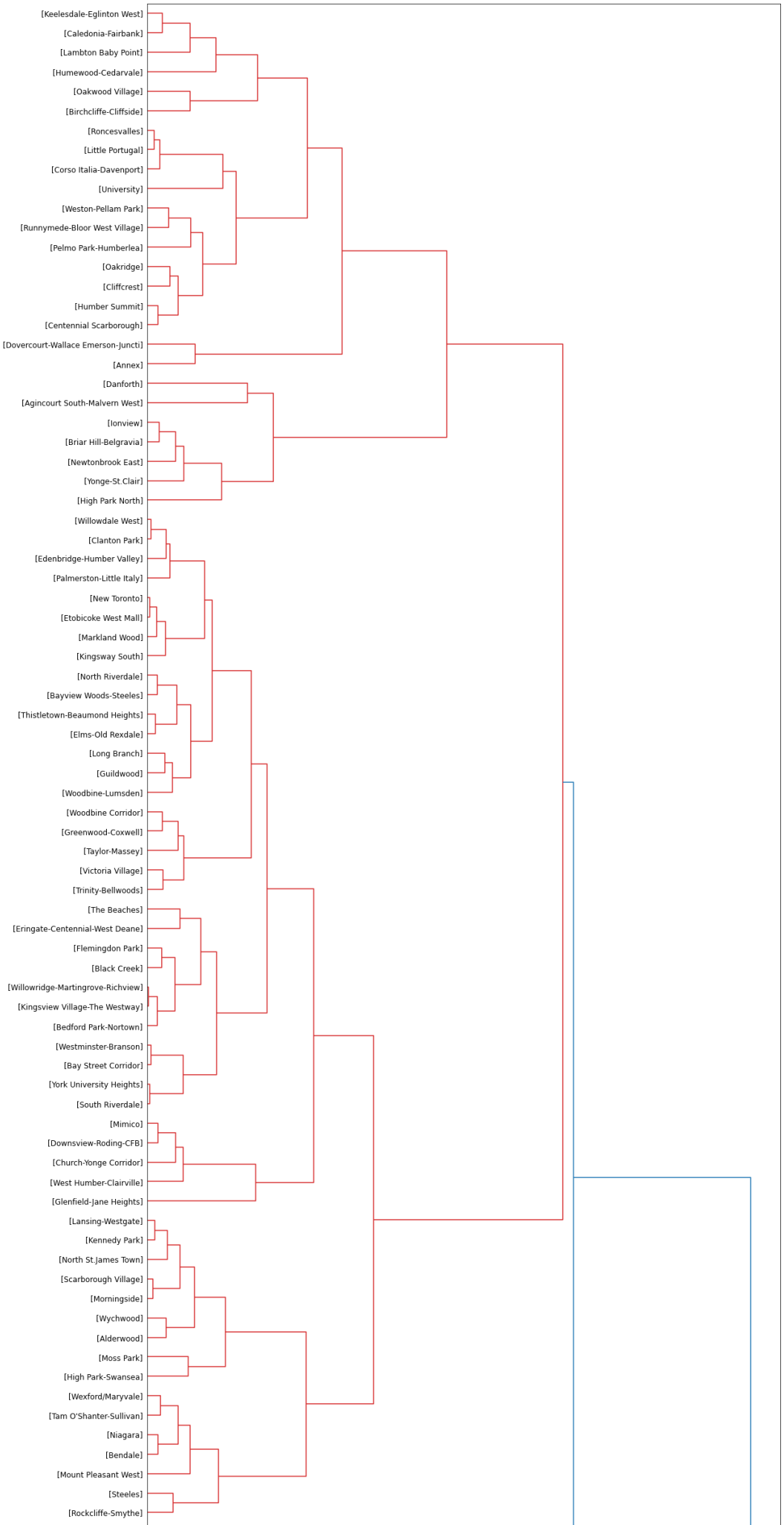
[Keelesdale-Eglinton West]
[Caledonia-Fairbank]
[Lambton Baby Point]
[Humewood-Cedarvale]
[Oakwood Village]
[Birchcliffe-Cliffside]
[Roncesvalles]
[Little Portugal]
[Corso Italia-Davenport]
[University]
[Weston-Pellam Park]
[Runnymede-Bloor West Village]
[Pelmo Park-Humberlea]
[Oakridge]
[Cliffcrest]
[Humber Summit]
[Centennial Scarborough]
[Dovercourt-Wallace Emerson-Juncti]
[Annex]
[Danforth]
[Agincourt South-Malvern West]
[Ionview]
[Briar Hill-Belgravia]
[Newtonbrook East]
[Yonge-St.Clair]
[High Park North]
[Willowdale West]
[Clanton Park]
[Edenbridge-Humber Valley]
[Palmerston-Little Italy]
[New Toronto]
[Etobicoke West Mall]
[Markland Wood]
[Kingsway South]
[North Riverdale]
[Bayview Woods-Steeles]
[Thistletown-Beaumond Heights]
[Elms-Old Rexdale]
[Long Branch]
[Guildwood]
[Woodbine-Lumsden]
[Woodbine Corridor]
[Greenwood-Coxwell]
[Taylor-Massey]
[Victoria Village]
[Trinity-Bellwoods]
[The Beaches]
[Eringate-Centennial-West Deane]
[Flemingdon Park]
[Black Creek]
[Willowridge-Martingrove-Richview]
[Kingsview Village-The Westway]
[Bedford Park-Nortown]
[Westminster-Branson]
[Bay Street Corridor]
[York University Heights]
[South Riverdale]
[Mimico]
[Downsview-Roding-CFB]
[Church-Yonge Corridor]
[West Humber-Clairville]
[Glenfield-Jane Heights]
[Lansing-Westgate]
[Kennedy Park]
[North St.James Town]
[Scarborough Village]
[Morningside]
[Wychwood]
[Alderwood]
[Moss Park]
[High Park-Swansea]
[Wexford/Maryvale]
[Tam O'Shanter-Sullivan]
[Niagara]
[Bendale]
[Mount Pleasant West]
[Steeles]
[Rockcliffe-Smythe]

Figure 4. Hierarchical Clustering of Toronto neighbourhoods.

Results and Discussion

Our analysis shows that, although Toronto has a relatively small number of restaurants for every 1000 people, there are areas with a high density of restaurants. The largest number of restaurants for every 1000 people are located in coastal areas and areas adjacent to Yonge Street. Areas remote from the coast and located on both sides of Yonge Street are of the greatest interest to investors in the restaurant business. The TOP 10 most attractive areas for opening a restaurant were determined. These are such areas as Mimico, Church-Yonge Corridor, Flemingdon Park and others.

We performed hierarchical clustering to determine which areas of Toronto are most similar, that is, how to group areas, which areas are most similar to each other in terms of the number of restaurants per 100 people.

This, of course, does not mean that these areas are actually the optimal place for a new restaurant! The purpose of this analysis was to provide information only about areas that are not crowded with existing restaurants - it is quite possible that there is a very good reason for a small number of restaurants in any of these areas,reasons why they are not suitable for a new restaurant, despite the lack of competition in this area. Therefore, the recommended zones should be considered only as a starting point for a more detailed analysis, which can eventually lead to a location where not only there are no competitors nearby, but also other factors are taken into account and all other relevant conditions are met.

Conclusion

The aim of this project was to identify areas with the least number of restaurants per 1000 people to help investors in the restaurant business narrow down the search for the optimal place to open a new restaurant. By calculating the distribution of restaurant density based on Foursquare data, we have identified the 10 most attractive areas for investors. Clustering and visualization of districts showed areas with a high number of restaurants for every 1000 residents, in which it is not advisable to open new restaurants.