Finding the best place for a restaurant in Toronto

Vladimir Volkov

19/10/19

1. Introduction

Toronto has an estimated population of just over 2.8 million in 2016, which makes it the 4th most populous city in North America the most populous Great Lakes city. Over the next 20 years, Toronto is expected to continue its moderate growth, surpassing 3 million by 2026, and reaching nearly 3.2 million in 2036.

Toronto is a global city filled with vast opportunity and is home to an array of distinctive and dynamic neighbourhoods that reflect the diversity of its population.

As the population of Toronto is constantly growing, this is a good place to open a restaurant business.

Usually restaurants are concentrated in developed areas.

The easiest way to find the best place for a restaurant is to search for the area where the largest number of restaurants are already concentrated. It will also be useful to identify the most popular types of restaurants Toronto.

2. Question

What types of restaurants are the most popular in Toronto? What neighborhood has maximum density of restaurants?

3. Data acquisition and cleaning

3.1 Data sources
I'll use Wikipedia page with the list of Neigborhoods in Toronto to create the dataframe with the list of Toronto neighborhoods. https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
To obtain geographic coordinates, the file will be used at the following link: https://cocl.us/Geospatial_data

Forsquare base will be used to search for venues in Toronto neighborhoods, including restaurants.

3.2 Data cleaning
Scraped data from the Wikipedia page will be cleaned in order to remove "Not assigned" postal codes.
It is also needed to replace missing data in column "Neigbourhood".

Aggregated by postal code Data will be combined with geographic coordinates.

To get a list of restaurants, first I need to get a list of all the places in Toronto from the Forsquare database. Then this list will be filtered by category of the venue including string "restaurant" in its name.

The final dataframe for analysis will include following information: Neighborhood, Neighborhood Latitude, Neighborhood Longitude, Venue (Name of a restaurant), Venue Latitude, Venue Longitude and Venue Category (Type of a restaurant).

3.3 Feature selection

Full list of Toronto venues includes 2225 venues. But we only need information about restaurants.

The final dataframe for analysis will include following information: Neighborhood, Neighborhood Latitude, Neighborhood Longitude, Venue (Name of a restaurant), Venue Latitude, Venue Longitude and Venue Category (Type of a restaurant).

|	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Rouge, Malvern	43.806686	-79.194353	Wendy's	43.807448	-79.199056	Fast Food Restaurant
1	Guildwood, Morningside, West Hill	43.763573	-79.188711	Big Bite Burrito	43.766299	-79.190720	Mexican Restaurant
2	Woburn	43.770992	-79.216917	Korean Grill House	43.770812	-79.214502	Korean Restaurant
3	Cedarbrae	43.773136	-79.239476	Drupati's Roti & Doubles	43.775222	-79.241678	Caribbean Restaurant
4	Cedarbrae	43.773136	-79.239476	Federick Restaurant	43.774697	-79.241142	Hakka Restaurant

```
restaurants.shape
```
```
(510, 7)
```

So, we can see that there are 510 restaurants in Toronto according to Forsquare database.

Geographic coordinates of the venues can be used to define which Neighborhoods have the highest density of restaurants.

The map of restaurants in Toronto.
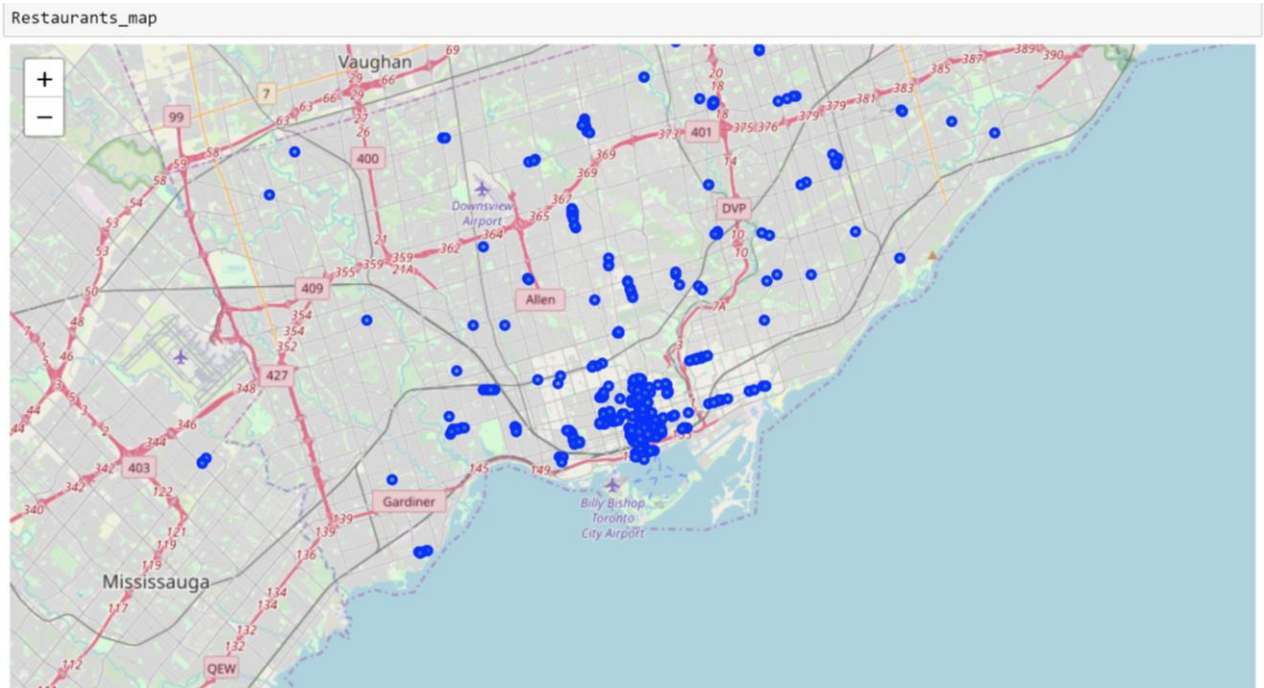
We can make a bar plot to see, what types of restaurants are popular in Toronto. The most popular types of restaurants:
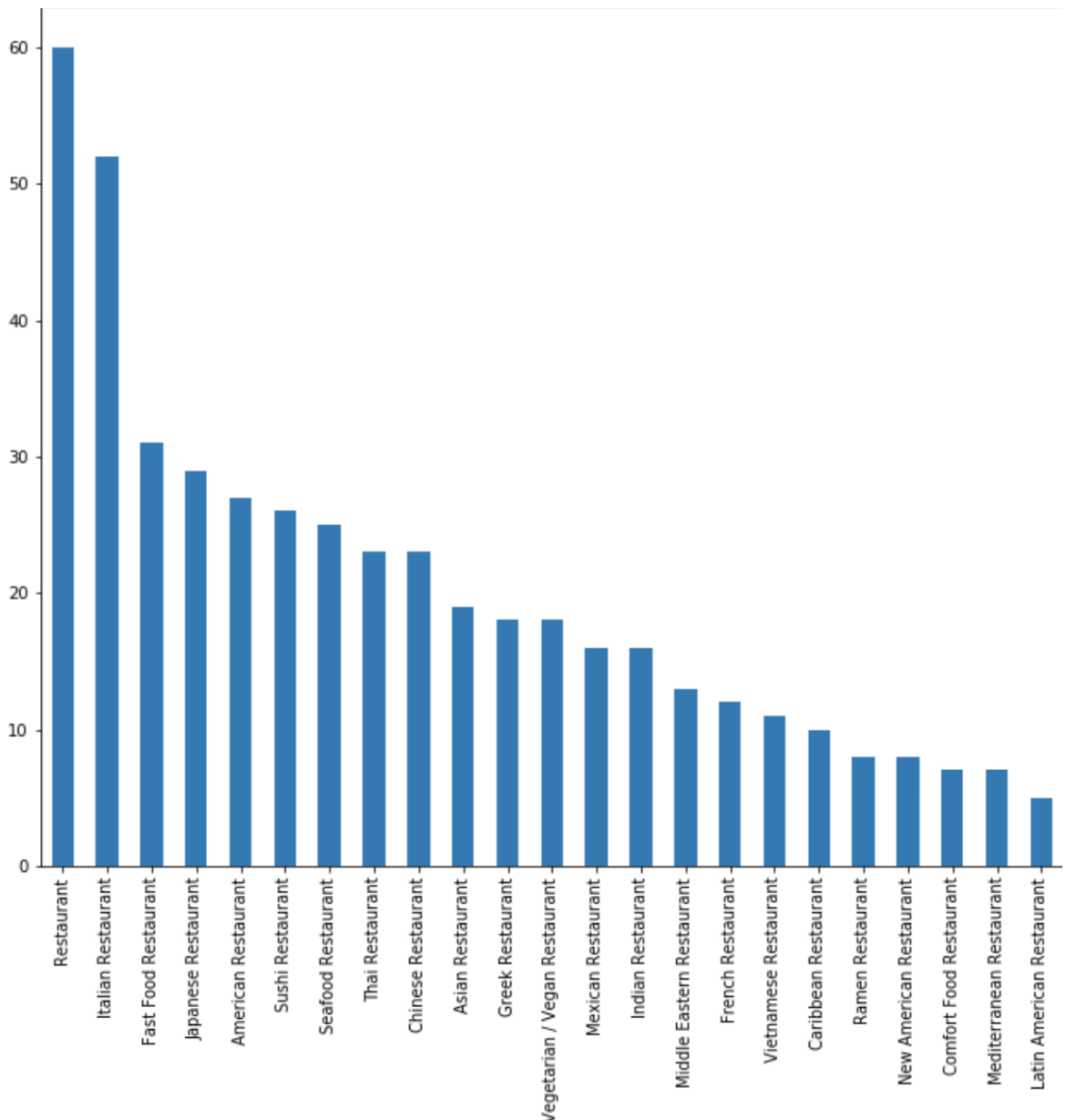
According to the plot, the most popular types of restaurants in Toronto are: "Restaurant", "Italian restaurant", "Fast food restaurant" and "Japanese restaurant".

4. Methodology and modelling

To search for areas with the maximum density of restaurants I will use the DBSCAN model.
This model has two main parameters: epsilon and min samples.
The epsilon parameter is the max distance between points of cluster (0.1 km in my example).

The min_samples parameter is the minimum cluster size (everything else gets classified as noise).

I'll set min_samples of restaurants in cluster to 8. Other areas will be classified as noise.

I use the haversine metric and ball tree algorithm to calculate great circle distances between points. Notice my epsilon and coordinates get converted to radians, because scikit-learn's haversine metric needs radian units.

After training the model we receive following map of clusters with maximum density of restaurants in Toronto. Each point is a restaurant in a cluster.
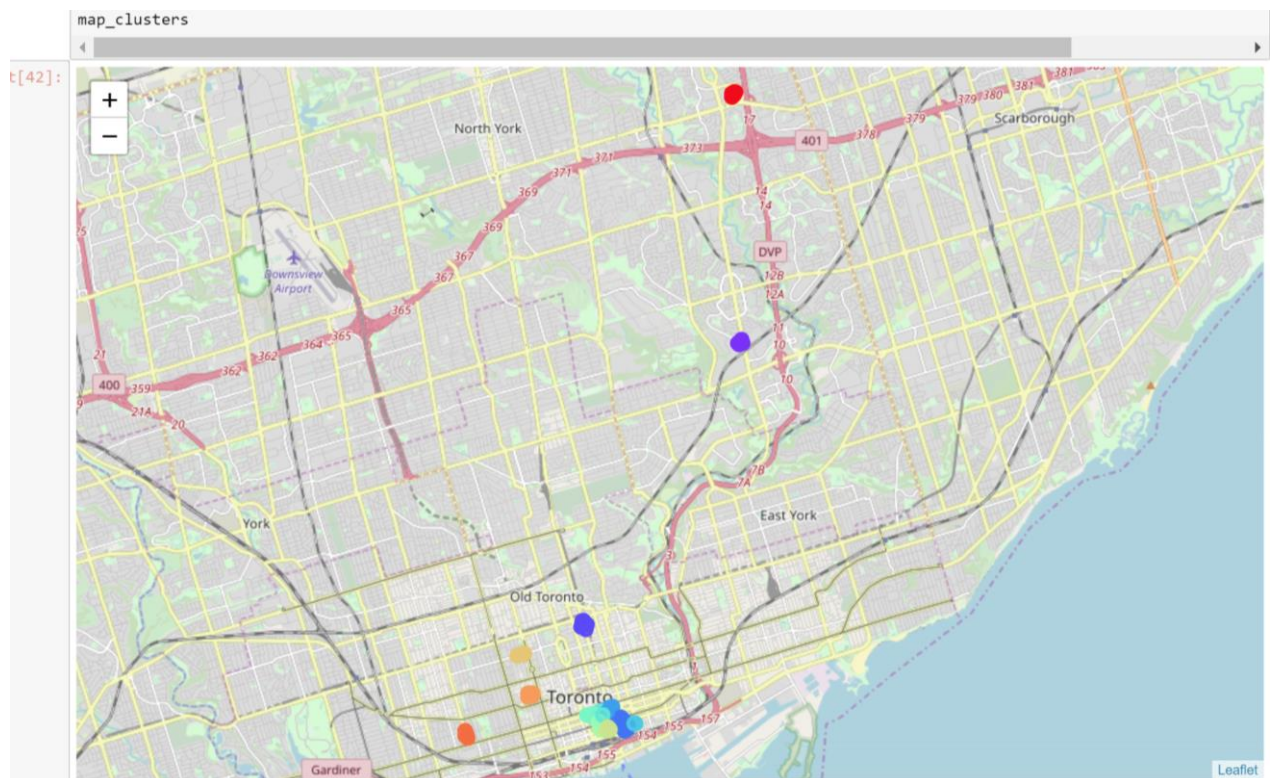

5. Results

After we have received the list of restaurants included in the clusters with the highest density, we can create a list of areas most suitable for opening a restaurant.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude
0	Fairview, Henry Farm, Oriole	43.778517	-79.346556
1	Flemingdon Park, Don Mills South	43.725900	-79.340923
2	Church and Wellesley	43.665860	-79.383160
3	St. James Town	43.651494	-79.375418
4	Berczy Park	43.644771	-79.373306
5	Adelaide, King, Richmond	43.650571	-79.384568
6	Design Exchange, Toronto Dominion Centre	43.647177	-79.381576
7	Commerce Court, Victoria Hotel	43.648198	-79.379817
8	Harbord, University of Toronto	43.662696	-79.400049
9	Chinatown, Grange Park, Kensington Market	43.653206	-79.400049
10	Stn A PO Boxes 25 The Esplanade	43.646435	-79.374846

And below is the map of best places for a restaurant in Toronto.
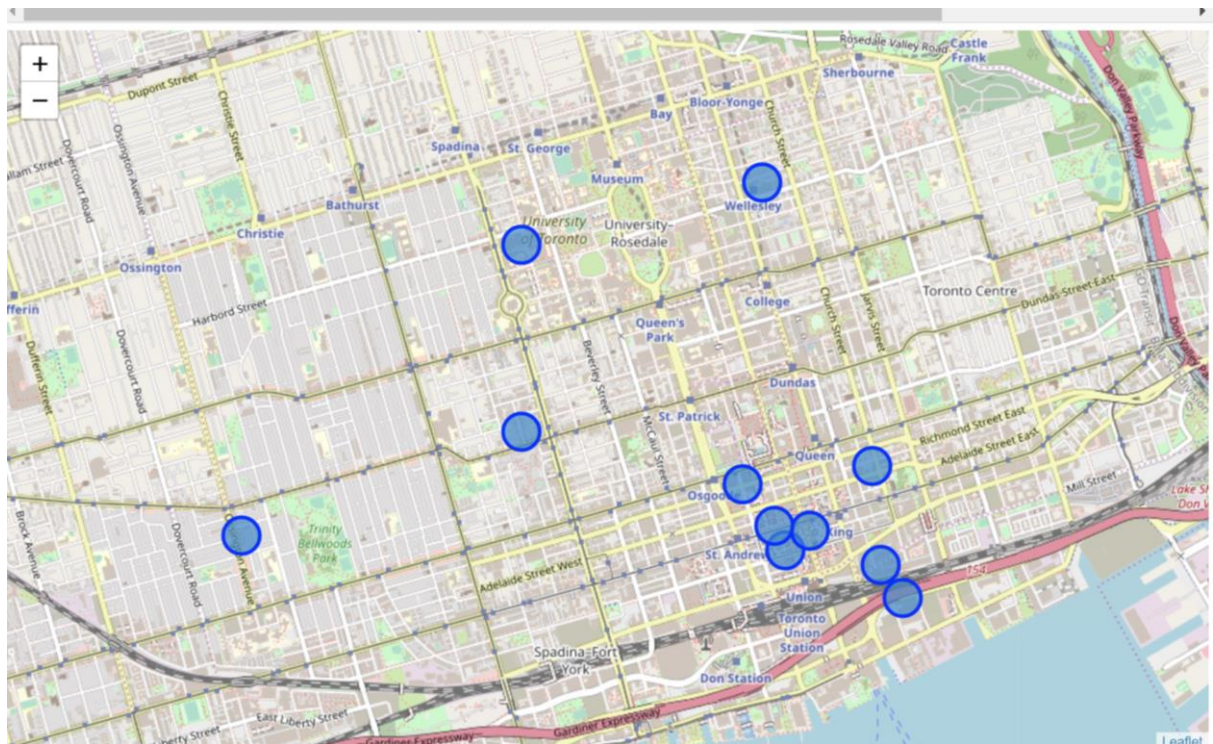
And according to the previous research, the most popular types of restaurants in Toronto are: "Restaurant", "Italian restaurant", "Fast food restaurant" and "Japanese restaurant".

6. Conclusions

In this study, I analyzed the areas of Toronto with the maximum density of restaurants. These areas are good for placing a new restaurant. Also, I defined which type of restaurants are most popular in Toronto.

7. Future directions

In future, it will be useful to explore some correlation between restaurants and other type of venues, for example, offices, trade centers, hotels and so on.