Coursera capstone

The Battle of Neighborhoods

A Restaurant in Toronto

By Cheng shuiqi

Problem Analyses

- **Location is of vital importance to a restaurant**
- **Density of the population in the area**:

if the area holds the enough population so that our business is more likely to trait more consumers.

- **Distribution of other restaurants**:

if the area holds the enough population so that our business is more likely to trait more consumers.

- **Target audience**:

this project is useful for someone who is about to open a restaurant in Toronto.

Experimental Data

- **Borough-Neighborhood information**

in order to obtain the density of population, we may as well assume the neighborhood information can represent the population information in some way.

- **position information**:

we need to transform form them into latitude and longitude. we can simply use the Google Maps Geocoding API.

- **restaurant distribution information**:

use the foursquare API and the position information combined with the folium lib, we can visualize the position of the restaurants and shops in the area

Modeling Method

- **Obtain the Data Frame**

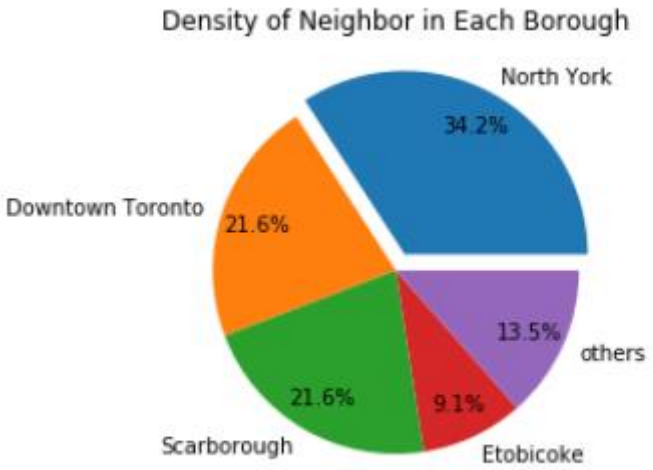
In this part, we can utilize the re and Beautifulsoup lib to get the data from the data source. With the help of our csv document, we can add the coordinate into our dataframe.

	PostalCode	Borough	Neighborhood	Latitude	Longitude
0	M5G	Downtown Toronto	Central Bay Street	43.657952	-79.387383
1	M2H	North York	Hillcrest Village	43.803762	-79.363452
2	M4B	East York	Parkview Hill,Woodbine Gardens	43.706397	-79.309937
3	M1J	Scarborough	Scarborough Village	43.744734	-79.239476
4	M4G	East York	Leaside	43.709060	-79.363452
5	M4M	East Toronto	Studio District	43.659526	-79.340923
6	M1R	Scarborough	Wexford,Maryvale	43.750072	-79.295849
7	M9V	Etobicoke	South Steeles,Silverstone,Humbergate,Jamestown...	43.739416	-79.588437
8	M9L	North York	Humber Summit	43.756303	-79.565963
9	M5V	Downtown Toronto	CN Tower,King and Spadina,Railway Lands,Harbou...	43.628947	-79.394420
10	M1B	Scarborough	Malvern,Rouge	43.806686	-79.194353
11	M5A	Downtown Toronto	Regent Park,Harbourfront	43.654260	-79.360636

Modeling Method

- **Neighborhood Distribution**

we can focus on the number of neighbors each borough holds. Then we calculate the number and plot the pie chart. And we can see that the North York holds a relatively large population.

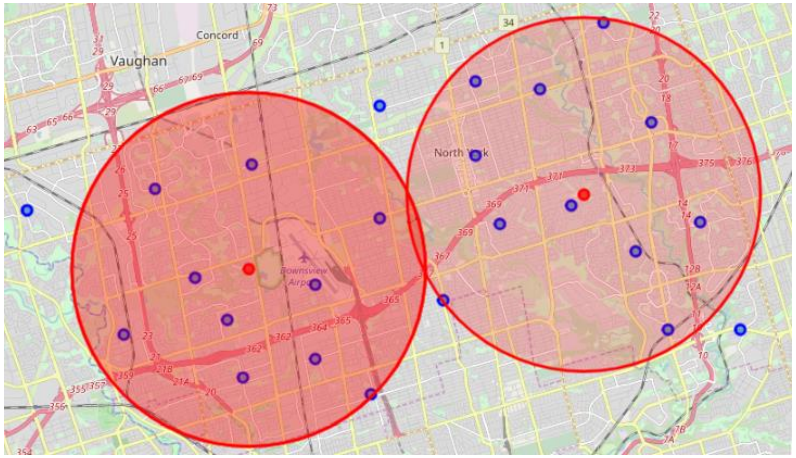


Modeling Method

- **Neighborhood Distribution**

Select the data and use the folium lib to plot the map figure. With the help of joint plot , it is reasonable to divide this area into two clusters.

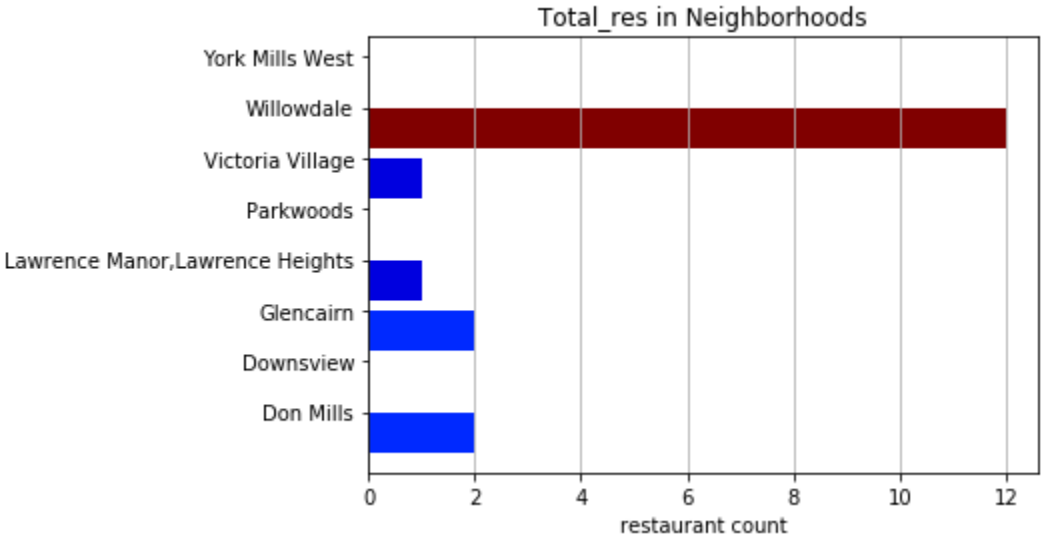
Simply utilize the Kmeans method, we can find 2 centers of the area.



Modeling Method

- **Nearby Restaurants**

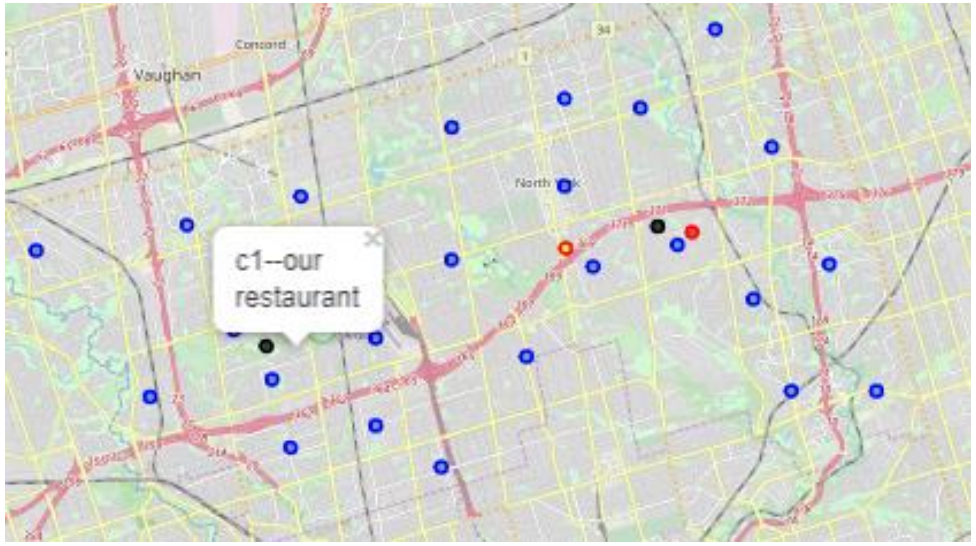
From the Foursqaure API we get the restaurant information. By using the weight method, we can find a center point of restaurants known as the restaurant center.



Modeling Method

- **Co-Analyze**

we can find the best location by considering both the two aspects, as we know, Neighborhood Distribution and Nearby Restaurants.



Results

In this model, the best position to open a restaurant in Toronto is at (43.73575162, -79.4973781). As is analyzed, we are more likely to have relatively more customers with correspondingly lower competition with other restaurants.

Discussion

We can find that most of the restaurants are located in Willowdale in the right cluster. Oppositely, few of restaurants are clustered into the left cluster, while they have almost the same population density. This provide the opportunity for us to open the restaurant here and have more space to develop.

Conclusion

Although, there's some features that we didn't take into consideration such as the income information in each area and the accurate population distribution.

In terms of the existing data we've gathered, the result has relatively high credibility, since we just focus on the specific area, Toronto.

THANKS