Open a new restaurant in Toronto

Wen Gao

May 16, 2021

1. Introduction

1. There are many good restaurants in Toronto, and it is not easy to choose a great location, if a businessman wants to open a new restaurant in Toronto.
2. This new location needs to be in a popular neighborhood, but there should not be many restaurants already, so the market of restaurants is not saturated yet.
3. In this case, we can use data science to make a solution to this problem, and provide suggestions to the businessman of where to open the new restaurant, and can help the businessman to have more chance to earn money from this new restaurant.

2. Data acquisition and cleaning

There are 3 data sources in this project:

1. To get the whole overview of neighborhoods info: Use Wikipedia URL
2. To match the neighborhoods with accurate latitude and longitude: Use IBM data science course resource.
3. To get the top 10 venues in each neighborhoods in Toronto, using Foursqaure API.

2. Data acquisition and cleaning

In total, we will combine different data into 1 dataframe, and then apply cluster analysis on it.

1. Get the neighborhoods data of Toronto from wikipage, which will be in HTML format. Then transfer the HTML data into pandas dataframe by using python soup function.
2. For the CSV file of latitude and longitude of neighborhoods in Toronto, it is already in clean structure, and the only thing we need to do it to combine the data with the 1st dataframe, and get a new dataframe out of it.
3. For the response from Foursquare data, it is in json format. In this case, we need to transfer the json data into pandas dataframe.

The output dataframe looks as below:

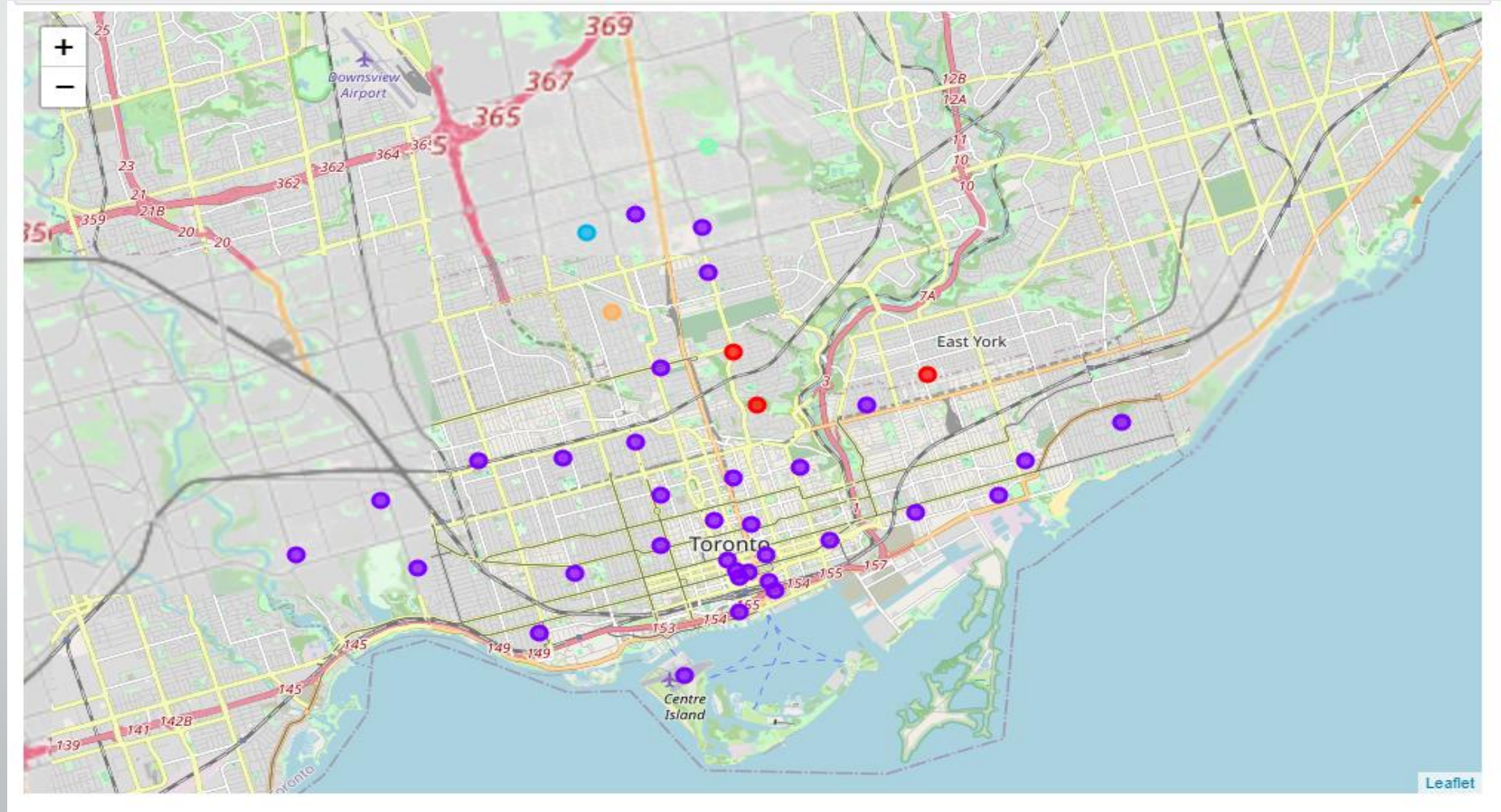
	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Regent Park, Harbourfront	43.65426	-79.360636	Roselle Desserts	43.653447	-79.362017	Bakery
1	Regent Park, Harbourfront	43.65426	-79.360636	Tandem Coffee	43.653559	-79.361809	Coffee Shop
2	Regent Park, Harbourfront	43.65426	-79.360636	Cooper Koo Family YMCA	43.653249	-79.358008	Distribution Center
3	Regent Park, Harbourfront	43.65426	-79.360636	Body Blitz Spa East	43.654735	-79.359874	Spa
4	Regent Park, Harbourfront	43.65426	-79.360636	Impact Kitchen	43.656369	-79.356980	Restaurant

3. Methodology

1. Use cluster modeling as the method.
2. Get each neighborhood labeled a cluster mark, then calculate the counts of venues for each neighborhood, and the counts of restaurants for each neighborhood.
3. Use the value of "counts of venues for each neighborhood" to divide "counts of restaurants for each neighborhood", in this case, we get the percentage of the restaurants among the venues for each neighborhood.
4. Since the neighborhoods in the same clustered group should have similar attributes, we can assume that within the same group, the less percentage a neighborhood has, the better that this location is great for opening a new restaurant.

3. Methodology

- This is the clustering map overview:



4. Results

- In cluster 0, both neighborhoods Rosedale and The Danforth East are good choices to open a new restaurant, as there is no restaurant in top 10 venues yet. The overview is as below:

	Neighborhood	total_venues_count_per_neighboorhood	is_restaurant	percentage
0	Rosedale	4	0	0.0
1	The Danforth East	4	0	0.0

4. Results

- In cluster 1, the top 5 neighborhoods for the choices to open a new restaurant are:

1) "CN Tower, King and Spadina, Railway Lands, Harbourfront West, Bathurst Quay, South Niagara, Island airport",

2) "The Beaches",

3) "Davisville North"

4) "Dufferin, Dovercourt Village"

5) "Brockton, Parkdale Village, Exhibition Place"

And here is the overview:

	Neighborhood	total_venues_count_per_neighboorhood	is_restaurant	percentage
2	CN Tower, King and Spadina, Railway Lands, Har...	17	0	0.000000
29	The Beaches	4	0	0.000000
8	Davisville North	8	0	0.000000
9	Dufferin, Dovercourt Village	15	1	0.066667
1	Brockton, Parkdale Village, Exhibition Place	22	2	0.090909

5. Discussion

There is still space to improve for this project.

1. Data concerning: In this project, we don't have enough big data to provide a very accurate result. Both data of neighborhoods and top venues are not big enough.
2. Methodology concerning: The dimension to analyze this project is very single, but in real life, this problem has multiple factors that we need to take into consideration. So actually we need more factors analysis to provide a better result.

6. Conclusions

- In all, from this project, we can have a primary prediction of which neighborhoods in Toronto are more suitable to open a new restaurant. The best neighborhoods to open a new restaurant in Toronto that I suggested are:

1) "CN Tower, King and Spadina, Railway Lands, Harbourfront West, Bathurst Quay, South Niagara, Island airport",

2) "The Beaches",

3) "Davisville North"

4) "Dufferin, Dovercourt Village"

5) "Brockton, Parkdale Village, Exhibition Place"

6) "Rosedale"

7) "The Danforth East"