The Battle of the Neighborhoods

Xiaomeng Shang

November 19, 2020

Table of Contents

The Battle of the Neighborhoods

1. Introduction

1.1 Background

Toronto is Canada's largest city, the most populous city in Canada, and home to a diverse population of about 2.9 million people.

The city is ranked as one of the top destinations around the globe. It boasts world-class restaurants, cultural attractions as varied as the cultures themselves.

Moreover, Toronto is recognized for being Canada’s commercial capital and for its excellence in a number of sectors including life sciences, technology, and education. Thus, the outstanding opportunities attract investors all around the world.

1.2 Problem

A group of stakeholders have experience in running restaurant. They are attracted by culture diversity of Toronto and want to expand restaurant business.

They intend to open a Chinese restaurant in downtown Toronto, the main central business district of Toronto.

1.3 Interest

The location will make an impact on succeed of the restaurant. We particularly interested in:

1) areas with no Chinese restaurants;
2) areas which are not crowded with restaurants.

We are going to analysis location information of the restaurants in downtown Toronto and find an optimal location for stakeholders’ new restaurant.

2. Data acquisition and cleaning

2.1 Data sources

Information of Neighborhoods of Toronto can be found in a Wikipedia page here. A table in this page list postal code, borough and neighborhood name.

In week 3, the course provides a link of a csv document here, through which we can obtain the geographical coordinate conveniently.

Then, we can use Foursquare API to get venues' information in each neighborhood.

2.2 Data cleaning

1) Neighborhood information

We can use lxml package to scrape the table from Wikipedia page.

We only process with cells with valid values. Thus, we delete rows, which have Borough with "Not assigned" value. After checking Neighborhood column, we find that all the values are valid.

Then, we reset the index and get the data frame of Neighborhood information, including postal code, borough and neighborhood.

	Postal Code	Borough	Neighborhood
0	M3A	North York	Parkwoods
1	M4A	North York	Victoria Village
2	M5A	Downtown Toronto	Regent Park, Harbourfront
3	M6A	North York	Lawrence Manor, Lawrence Heights
4	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government

2) Geographical information

To get the data efficiently, we use the csv file to get the geographical information and load the data into a data frame.

	Postal Code	Latitude	Longitude
0	M1B	43.806686	-79.194353
1	M1C	43.784535	-79.160497
2	M1E	43.763573	-79.188711
3	M1G	43.770992	-79.216917
4	M1H	43.773136	-79.239476

Join the two data frames, we get a new data frame, which combines the neighborhood information and geographical information.

	PostalCode	Borough	Neighborhood	Latitude	Longitude
0	M3A	North York	Parkwoods	43.753259	-79.329656
1	M4A	North York	Victoria Village	43.725882	-79.315572
2	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636
3	M6A	North York	Lawrence Manor, Lawrence Heights	43.718518	-79.464763
4	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government	43.662301	-79.389494

3) Venues' information

We can use Foursquare API to get venues' information in each neighborhood of downtown Toronto.

At first, we write a function to get the top 100 venues in a radius of 500 meters in every neighborhood.

Then, we get venues information. There are 1253 venues.

	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

2.3 Feature selection

As we care about the category of each venue. After checking, we found that there are 214 unique categories.

We should focus on category 'restaurant'. However, there are so many categories

related to restaurant, in another words, they are specific restaurant categories, such as 'French Restaurant', 'Mexican Restaurant', 'Portuguese Restaurant', 'Italian Restaurant'.

Those categories should be taken into consideration as well. We examined every category and extracted categories, whose name contains ‘Restaurant’. We regard them as competitor category.

In total, there are 43 competitor categories.

```
['Restaurant', 'French Restaurant', 'Mexican Restaurant', 'Italian Restaurant', 'Portuguese Restaurant', 'Sushi Restaurant', 'C
hinese Restaurant', 'Ramen Restaurant', 'Thai Restaurant', 'New American Restaurant', 'Fast Food Restaurant', 'Japanese Restaur
ant', 'Middle Eastern Restaurant', 'Modern European Restaurant', 'Ethiopian Restaurant', 'Seafood Restaurant', 'Vietnamese Rest
aurant', 'American Restaurant', 'Latin American Restaurant', 'Vegetarian / Vegan Restaurant', 'German Restaurant', 'Comfort Foo
d Restaurant', 'Asian Restaurant', 'Moroccan Restaurant', 'Belgian Restaurant', 'Greek Restaurant', 'Eastern European Restauran
t', 'Falafel Restaurant', 'Indian Restaurant', 'Korean Restaurant', 'Colombian Restaurant', 'Mediterranean Restaurant', 'Brazil
ian Restaurant', 'Gluten-free Restaurant', 'Caribbean Restaurant', 'Dumpling Restaurant', 'Doner Restaurant', 'Filipino Restaur
ant', 'Dim Sum Restaurant', 'Molecular Gastronomy Restaurant', 'Taiwanese Restaurant', 'Theme Restaurant', 'Afghan Restaurant']
```

As we explored further, we discovered the top 10 popular categories:

Venue Category	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue
Restaurant	44	44	44	44
Japanese Restaurant	31	31	31	31
Italian Restaurant	22	22	22	22
Seafood Restaurant	20	20	20	20
Sushi Restaurant	17	17	17	17
American Restaurant	16	16	16	16
Thai Restaurant	15	15	15	15
Vegetarian / Vegan Restaurant	14	14	14	14
Asian Restaurant	9	9	9	9
Mexican Restaurant	9	9	9	9

Obviously, Japanese and Italian restaurant are particularly popular.

We drew a map to show how these restaurants distribute in downtown Toronto.

In the next section, we will use methodology to cluster the competitor restaurant.

3. Methodology

4. Results

5. Discussion

6. Conclusion