

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 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 areas with no Chinese restaurants. We prefer to areas as well, which are not crowded with restaurants.

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.

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.

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	Impact Kitchen	43.656369	-79.356980	Restaurant
4	Regent Park, Harbourfront	43.65426	-79.360636	Body Blitz Spa East	43.654735	-79.359874	Spa

2.3 Feature selection

As we care about the category of each venue, we examined category amount and categories related to restaurant.

There are 214 unique categories and 43 categories related to restaurant, which are different types of restaurants.

We extract these restaurant data and structure them into a new data frame.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
3	Regent Park, Harbourfront	43.654260	-79.360636	Impact Kitchen	43.656369	-79.356980	Restaurant
21	Regent Park, Harbourfront	43.654260	-79.360636	Cluny Bistro & Boulangerie	43.650565	-79.357843	French Restaurant
26	Regent Park, Harbourfront	43.654260	-79.360636	El Catrin	43.650601	-79.358920	Mexican Restaurant
46	Queen's Park, Ontario Provincial Government	43.662301	-79.389494	Nando's	43.661728	-79.386391	Portuguese Restaurant
47	Queen's Park, Ontario Provincial Government	43.662301	-79.389494	Mercatto	43.660391	-79.387664	Italian Restaurant