**The Battle of the Neighborhoods**



Xiaomeng Shang

November 19, 2020

Table of Contents

[1. Introduction 1](#_Toc56709312)

[1.1 Background 1](#_Toc56709313)

[1.2 Problem 1](#_Toc56709314)

[1.3 Interest 1](#_Toc56709315)

[2. Data acquisition and cleaning 1](#_Toc56709316)

[2.1 Data sources 1](#_Toc56709317)

[2.2 Data cleaning 2](#_Toc56709318)

[2.3 Feature selection 3](#_Toc56709319)

[3. Methodology 5](#_Toc56709320)

[4. Results 5](#_Toc56709321)

[5. Discussion 5](#_Toc56709322)

[6. Conclusion 5](#_Toc56709323)

**The Battle of the Neighborhoods**

# Introduction

## 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.

## 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.

## 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.

# Data acquisition and cleaning

## 2.1 Data sources

Information of Neighborhoods of Toronto can be found in a Wikipedia page [here](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M). 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](http://cocl.us/Geospatial_data), through which we can obtain the geographical coordinate conveniently.

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

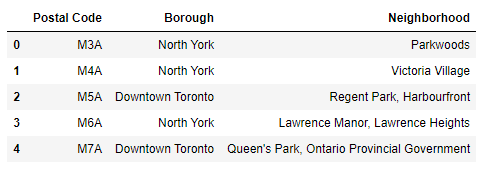
## 2.2 Data cleaning

1. Neighborhood information

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

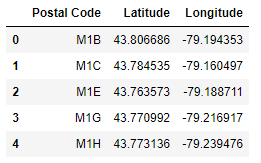
We only process with cells with valid values. Thus, we delete rows, which have Borough with “Not assigned” value. After checking Neighborhood column, we find that all the values are valid.

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

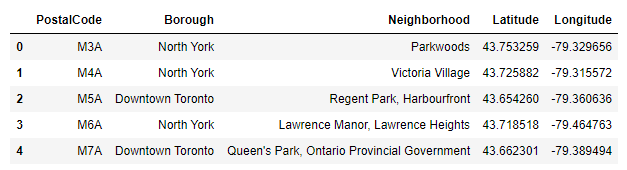


1. Geographical information

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.

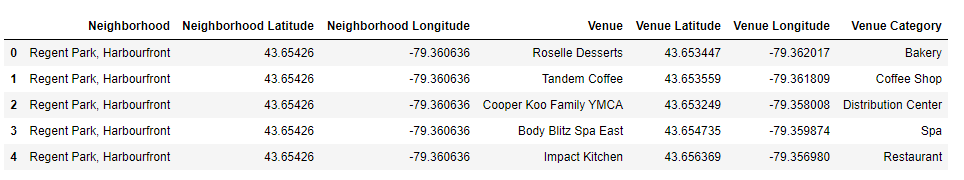


1. Venues’ information

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.



## 2.3 Feature selection

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

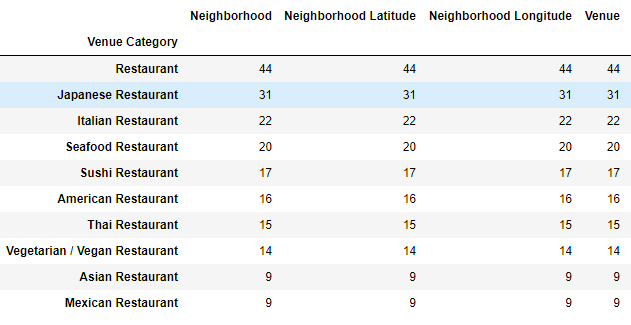
We should focus on category 'restaurant'. However, there are so many categories related to restaurant, in another words, they are specific restaurant categories, such as 'French Restaurant', 'Mexican Restaurant', 'Portuguese Restaurant', 'Italian Restaurant'.

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.



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



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.

# Methodology

# Results

# Discussion

# Conclusion