

Find the optimal location to open a bubble tea shop in Toronto

Iris Li

July 20, 2020

1. Introduction

1.1 Background

There are many factors to determine whether a new restaurant will thrive such as fund, brand recognition and quality of food, while geographical location is one of them that plays an essential role of the success. Thus, finding an optimal placement for a new shop can help the business gain customer base, brand recognition and more profit in the near future when it is opened.

1.2 Business Problem

This project aims to find an optimal place to open a brand-new bubble tea shop at North York district in Toronto. The goal is to find the best neighbourhood to open the restaurant by considering two essential factors. First, the market competition around each neighborhood will be examined to determine the market saturation and opportunity to gain customers. Second, I will find shopping malls as well as Asian restaurants in each neighborhood with which the bubble tea store can create marketing synergy and cooperation.

Interest

People who plan to open a food service business in Toronto area especially North York will be interested in this project since location is a necessary factor to be considered when someone wants to start a service business.

2. Data acquisition and cleaning

2.1 Data sources

The postal code, borough and neighborhood of Toronto are found in [Wikipedia](#). The latitude and longitude of each neighborhood can be obtained from this [CSV](#) file. I also use `explore` function in [Foursquare API](#) to get the venues around each neighborhood. By combining these three datasets, I am able to make a conclusion on the optimal location to open a bubble tea shop in North York, ON.

2.2 Data cleaning

All these datasets need to be cleaned and combined into one table. There is a table on the website of Wikipedia that will be used in the analysis. It contains postal code, borough, and neighbourhood of GTA area. To make it useable, I used the *Beautiful Soup* code to obtain the table from the website. I made three assumptions:

- Remove all the missing value in borough (not assigned value)
- Remove any duplicates
- If a cell has a value in borough but a not assigned in neighborhood, then the neighborhood will be the same as the borough.

I also read the CSV file with the latitude and longitude of Toronto into my code. And then use the *merge* function to combine these two table with columns of postal code, borough, neighbourhood, latitude and longitude. The first five rows look like this:

	Postal Code	Borough	Neighbourhood	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

Because I only want to examine neighborhood in North York, I created a new table which only contains the borough with the word "North York". This new table has 24 rows representing the 24 neighborhoods in North York. The first five rows show below:

	Postal Code	Borough	Neighbourhood	Latitude	Longitude
0	M3A	North York	Parkwoods	43.753259	-79.329656
1	M4A	North York	Victoria Village	43.725882	-79.315572
2	M6A	North York	Lawrence Manor, Lawrence Heights	43.718518	-79.464763
3	M3B	North York	Don Mills	43.745906	-79.352188
4	M6B	North York	Glencairn	43.709577	-79.445073

3. Exploratory Data Analysis

3.1