

The Battle of the Neighborhoods: Toronto

Anita Rafat

4/16/2021

1. Introduction

1.1 Background

Toronto is the capital city of the Canadian province of Ontario. It is the most populous city in Canada with a population of over 2.9 million. Toronto is an international center of business, finance, arts, and culture. The city has a high concentration of banks and brokerage firms and is considered the financial and industrial capital of Canada. The city is also recognized as one of the most multicultural and cosmopolitan cities in the world.

1.2 Business Problem

A client is looking to open a sushi restaurant in Toronto. Sushi menus tend to be expensive, however, sushi lovers are willing to pay a premium price for an expertly crafted roll. Sushi thrives in urban centers with an active nightlife and in an area with a concentration of working professionals. A sushi restaurant that is located within a walking distance of office workers can bring a booming business during the day. If the restaurant is also located near a thriving bar, it can bring customers in for dinner.

2. Data Acquisition and Cleaning

2.1 Data Sources

Foursquare data will be used to determine the best location for opening a sushi restaurant in Toronto. I used the Wikipedia page [here](#) to get the postal codes of Toronto. I also used the csv file Geospatial_data.csv to get the geographical coordinates of each postal zip code in order to utilize the Foursquare location data. I also scraped data from this webpage [here](#) which includes the information on average real estate prices of each postal code in Toronto.

2.2 Data Cleaning

Data from Wikipedia was scraped using Beautiful Soup. All the missing values were removed. Some of the neighborhood names had parenthesis which also needed to be removed. An empty data frame was created and populated with the postal codes data. The postal codes coordinates downloaded from the Geospatial_data.csv file was combined into one table.

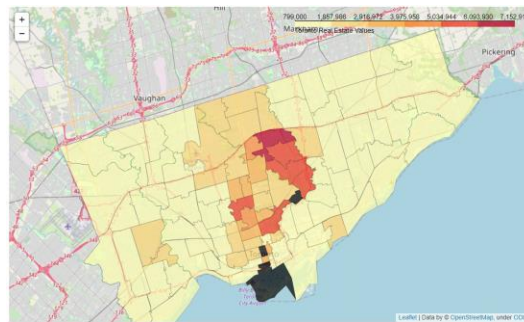
Data of the average real estate prices of Toronto included several redundant or unneeded columns. Since I only needed the postal code and average price information, just these two

columns were selected to create a new data frame. There were some missing values that were removed. The data frame was merged with the Geospatial_data.csv data in order to obtain the coordinates. I removed the dollar sign and commas in the Average Price column to convert the column from object to integer. It was essential to do this in order to create a choropleth map using this data frame.

3. Exploratory Data Analysis

3.1 Average Real Estate Prices

The average real estate prices are shown in the choropleth map below. The highest real estate prices are located at the center of Toronto. Since sushi restaurants tend to be expensive and attract more affluent customers, the ideal location to open a restaurant will be not too far from the center of Toronto.



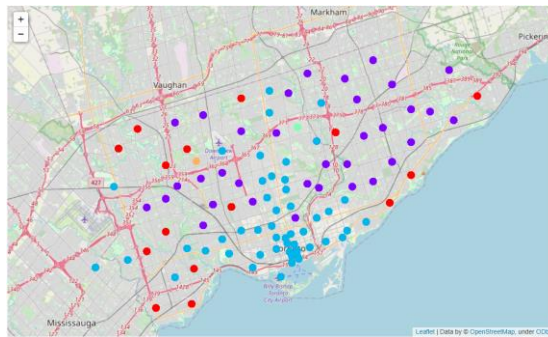
3.2 Foursquare

I utilized Foursquare to find locations in Toronto that have lots of venues including coffee shops and bars but don't have sushi restaurants. In the first step, I collected the required data: The neighborhoods and their locations and also the venues in each of these neighborhoods to see density. I identified sushi restaurants, coffee shops and other venues using Foursquare. I used maps to identify a few promising areas close to center with low number of sushi restaurants and high number of other venues such as tapas restaurants, coffee shops, and bars.



3.3 K-Means Clustering

In the third and final step, I used the most promising areas to create clusters of locations (using k-means clustering). The cluster map shows a total of five clusters. The last two clusters only had a few locations, so I focused on just the first three clusters. Cluster 1 includes mostly parks, pharmacies, gyms, shopping malls, etc. These locations are near residential areas. Cluster 2 is similar to Cluster 1. There are mostly parks, pharmacies, grocery stores, gyms, etc. However, this cluster also includes several Portuguese restaurants. These locations are also near residential areas. Cluster 3 makes up the majority of the neighborhoods fall into this cluster. There are mostly business areas with coffee shops, pizza places, restaurants, bars, etc. There are also social activity venues. Many of the neighborhoods are located near the center of Toronto and some of the neighborhoods are close to the University of Toronto.



4. Conclusion

The purpose of this project was to identify areas close to the center of Toronto that have many businesses and venues in order to help our client narrow down the search for the optimal location to open a sushi restaurant. By analyzing the venues in Toronto using Foursquare, I have identified the locations that have many venues and don't already have a sushi restaurant. Based on my findings, neighborhoods in Cluster 3 will be ideal locations for the sushi restaurant. The neighborhoods in Cluster 3 that I believe are optimal for opening a sushi restaurant are 1) Don Mills North; 2) Kensington, Chinatown, Grange Park; and 3) The Annex, North Midtown, Yorkville. These neighborhoods have many coffee shops and business areas as well as bars and nightlife. These three neighborhoods are also ideal because they don't have a sushi restaurant. According to the real estate data gathered, these locations have higher than average real estate prices which indicates that these neighborhoods are optimal for a sushi restaurant. Final decision on optimal restaurant location will be made by my client based on specific characteristics of neighborhoods and locations in every recommended zone, taking into consideration additional factors such as attractiveness of each location (proximity to park or water), levels of noise / proximity to major roads, real estate availability, prices, social and economic dynamics of every neighborhood, etc.