Site Selection for Japanese Restaurants in Toronto

IBM Data Science Capstone
Liang Nie

Introduction

Japanese cuisine is extremely popular throughout North America, primarily in suburban and urban metros like Toronto. Diners frequently patronize Japanese restaurants for Japanese-themed ambience and quality food including sushi, ramen, udon, etc. A good location is crucial to the new restaurant's performance as it will have a great impact on customer volume and customer type. This study explores all the neighborhoods in Toronto in terms of the frequency of occurrence of Japanese Restaurants. Cluster analysis is applied to group similar neighborhoods together. Through visualizing the resulting clusters on the map, we are able to gain a clearer understanding of how the Japanese restaurants are distributed across all the neighborhoods in Toronto. This project aims to help solve the business problem: Where's suitable to open a Japanese restaurant in Toronto?

This study targets entrepreneurs who are thinking about starting a Japanese restaurant in Toronto. Opening a Japanese restaurant is a challenging yet potentially rewarding venture given the popularity of Japanese cuisine in metro cities. Entrepreneurs would be interested to know how the Japanese restaurants are saturated across neighborhoods in Toronto and find out how many competitors are nearby.

Data

To get the list of all the neighborhoods in Toronto, we parsed the Wikipedia page (https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M). Since we are only interested in urban area, the final list contains 38 neighborhoods in 4 boroughs . Latitudes and longitudes of these boroughs could be obtained using Python Geocoder package, but a csv file contained all of the geographical coordinates was already provided by the course.

Venue data for each neighborhood were gathered using Foursquare API. Foursquare's global database contains 105 million places, which empowers users to explore points-of-interest conveniently and accurately. In the Foursquare dataset, "Japanese Restaurant" and "Sushi Restaurant" are two separate categories. However, in our project, we would like to consider both as "Japanese Restaurant".

Methodology

We began this project with scraping a Wikipedia page that contains all the borough and neighborhood information of Toronto. We used BeautifulSoup package to parse the page and wrangled the data, cleaned it, and then read it into a pandas dataframe. In order to

utilize the Foursquare location data, we obtained longitudes and latitudes of each neighborhood from a csv file that was prepared by the course already. Python Geocoder package would do the same work.

The original data contained 11 boroughs , but since we are only interested in urban Toronto, we narrowed the range down to 4 boroughs including "Downtown Toronto", "Central Toronto", "West Toronto", "East Toronto". The resulting number of neighborhoods is 38. We got the latitude and longitude value of urban Toronto through geopy library, which are 43.653963 and -79.387207. To make sure all the coordinates were correctly returned, we created a map of Toronto with 38 neighborhoods superimposed on top.

Next, we started utilizing the Foursquare API to explore the neighborhoods. We got the top 100 nearby venues for each neighborhood within a radius of 2500 meters. A total of 3800 venues were returned for all of the neighborhoods and there were 221 unique venue categories. To get ready for the clustering analysis, we used one hot encoding to transform the data and grouped rows by neighborhood and by taking the mean of the frequency of occurrence of each category. Since we are mainly focused on "Japanese Restaurant", we filtered the result with only "Japanese Restaurant" and "Sushi Restaurant" columns remained. To simplify the analysis, we just added the row values of these two columns together. The new value represents the frequency of Japanese Restaurant for each neighborhood.

Finally, we ran k-means clustering to cluster the 38 neighborhoods into 3 clusters based on the frequency of occurrence of Japanese restaurant. k-means is a simple and popular unsupervised machine learning algorithm. It aims to group similar observations together and thus discover underlying patterns. The result allowed us to identify neighborhoods with high, medium and low concentration of Japanese restaurants and would help us to locate the suitable site for the new restaurant.

Results

The resulting three clusters are as follows:

- Cluster 0: Neighborhoods with moderate concentration of Japanese restaurants
- Cluster 1: Neighborhoods with relatively low concentration of Japanese restaurants
- Cluster 2: Neighborhoods with relatively high concentration of Japanese restaurants

We visualized the resulting clusters on the map below with Cluster 0 marked as red, Cluster 1 purple and Cluster 2 green.
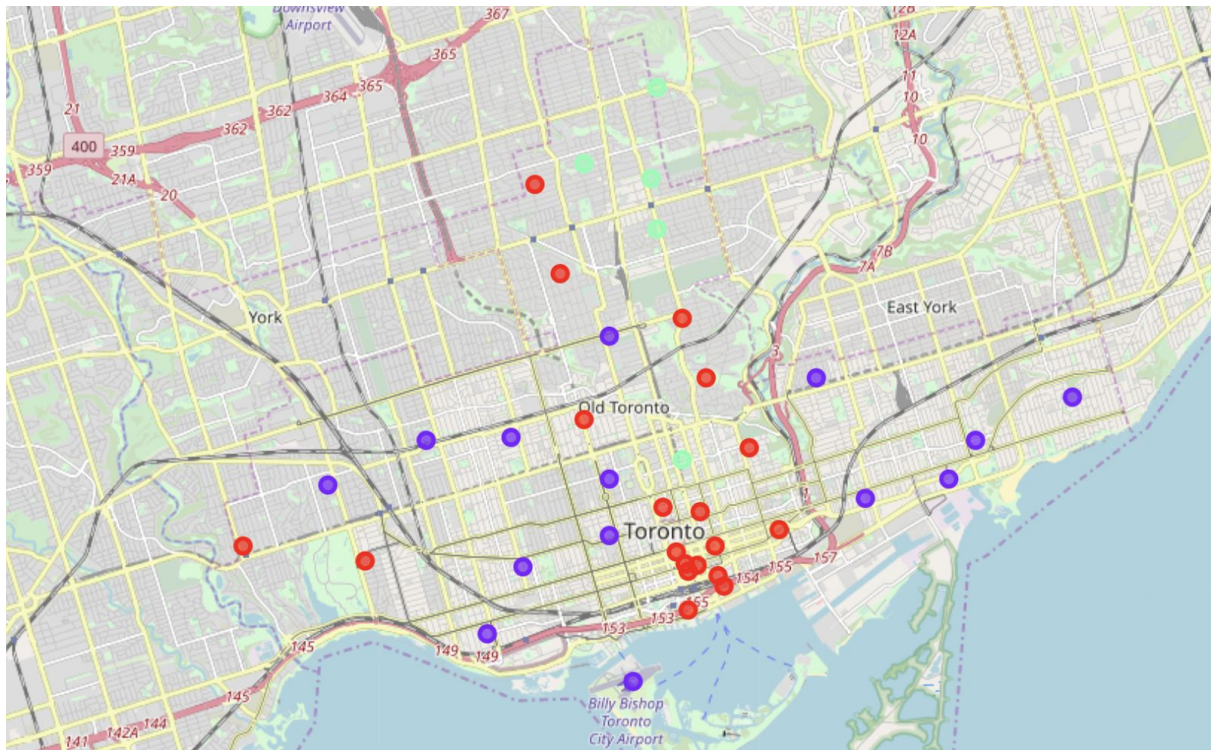
Discussion

Four out of five Cluster 2 neighborhoods with the most concentration of Japanese restaurants are in Central Toronto area. Japanese restaurants in this area are very likely to suffer from intense competition.

Most of the Cluster 0 neighborhoods with moderate saturation of Japanese restaurants are in Downtown Toronto. Downtown is always crowded with large population density with plenty of business opportunities. However, there are some immediate concerns when locating in a downtown area such as leasing cost and parking availability.

Cluster 1 neighborhoods, mostly locate in East Toronto and West Toronto, have relatively low saturation of Japanese restaurants. Among these neighborhoods, two spots which are located in Downtown Toronto especially stand out. They are "Chinatown, Grange Park, Kensington Market" and "Harbord, University of Toronto". The superior location with little concentration of Japanese restaurants may create advantageous opportunity for developing a new restaurant.

This project only examines the frequency of occurrence in relation to site selection. However, there are many other factors such as customer demographics, proximity to workforce and parking availability that should be considered in the real world. For the most part, restaurants target local customers. Future research could gather more data to determine if the community have the right demographics to support the new business.

Conclusion

In this project, we gathered location data of Toronto by web scraping and by utilizing Foursquare API. We mainly applied k-means clustering, a popular machine learning technique, to group similar neighborhoods in terms of the frequency of occurrence of nearby Japanese restaurants. Based on the clustering results, we recommended a few ideal spots for opening a Japanese restaurant. The findings will help the stakeholders to capitalize on the locations with the most business potentials.