IBM DATA SCIENENCE CAPSTONE PROJECT

The best location in Toronto to open a new Japanese Restaurant

Produced by Kaito Kobayashi

Contents

1. <u>INTRODUCTION</u>

This is the last project of IBM Data Science. In this project, I am pretending that I am consulting for an entrepreneur who wants to open a new Japanese restaurant in Toronto and try to decide which location to open it. I am a Japanese who is in the United States studying business. Location of restaurant is a critical factor of success in restaurant business, which is why this project result will be insightful for people who want to open Japanese restaurant in Toronto.

2. <u>BUSINESS PROBLEM</u>

The objective of this capstone project is to find out the most suitable location so that I will be able to recommend the location to the entrepreneur to open a new Japanese restaurant in Toronto. By using machine learning, specifically clustering, I will identify groups of locations that have similar attributes and spot the best group of locations to open a new Japanese restaurant.

3. <u>TARGET AUDIENCE</u>

Entrepreneurs who wants to open a new Japanese restaurant in Toronto and is wondering which location they should choose to open it.

4. <u>Data</u>

The data I used for the analysis is following:

- List of neighborhoods in Toronto, Canada (from Wikipedia)
- Latitude and Longitude of those locations (https://cocl.us/Geospatial_data)
- Venue data of Toronto (from FourSquare)

5. <u>Methodology</u>

First, I obtained the list of neighborhoods in Toronto, Canada. I used Wikipedia as a source of the data (https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M). I executed the web scraping using Beautiful Soup package. The neighborhood data include a list of postal codes, boroughs, and neighbors in Toronto.

Since I needed coordination data to utilize FourSquare data, I used the data prepared for this course to get each coordinate of the location in Toronto, and I merged the coordinate data to the first data. The figure below shows each location included in the data.
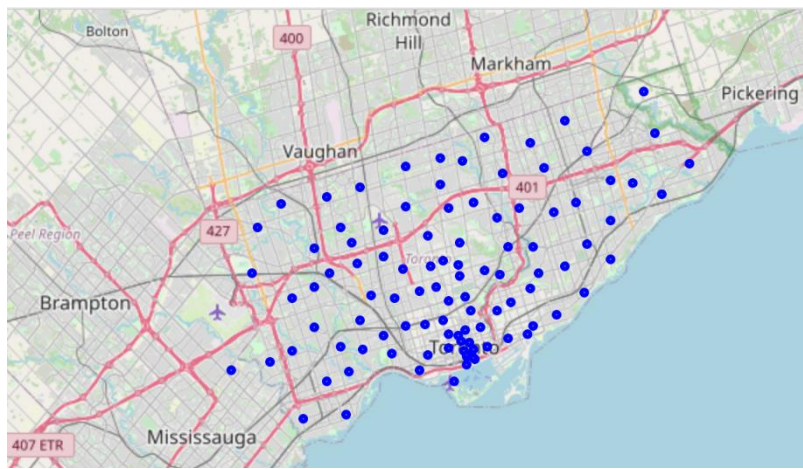

Next, I used Foursquare API to pull the list of top 30 venues. The data from FoourSquare includes the names, categories, latitude, and longitude of the venues. Since I am curious about food and restaurant industry, I only kept the categories that are related to restaurant. Then I used k-means clustering to categorize neighbors into five groups.

6. <u>Results</u>

The figure below shows the result of k-means clustering which shows five groups of neighbors.

- Cluster 1 (purple): This group of neighbors have various restaurants, bars, dessert shops, indicating that those places are the destination for foodies. Some location is famous for Japanese restaurants and Sushi restaurants.
- Cluster 2 (green): This group of neighbors is especially famous for coffee shops, American restaurants, and sea food restaurant.
- Cluster 3 (red): This group is famous for coffee shops but there are not many restaurants unlike cluster 2. However, there is one neighbor that has famous Sushi restaurants.

- Cluster 4 (orange): This neighbor is known for other countries' food such as Greek restaurants, Italian restaurants, and Caribbean restaurants.
- Cluster 5 (blue): This neighbor does not have much restaurants but is famous for ice cream shops.
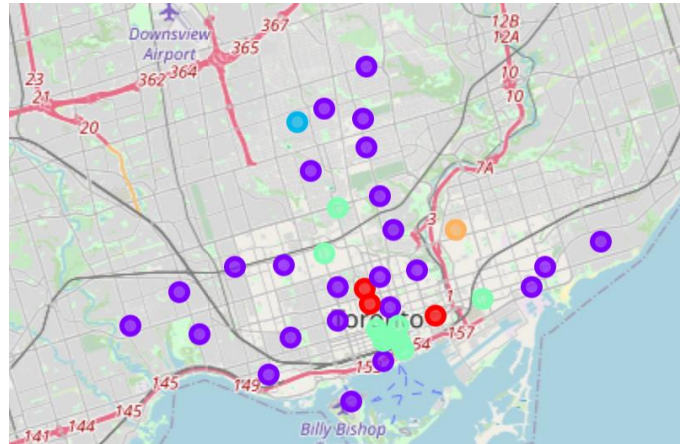
7. <u>Recommendation</u>

I suggest that entrepreneurs open a new Japanese restaurant at location in either cluster 1 or 2. Those locations are the destination for lunch and dinner. Therefore, it is relatively easy for a new restaurant to be noticed by customers without doing much effort, comparing to other groups of locations. One disadvantage of these cluster is that there are many competitors in those locations.