Report on Japanese Restaurant in the Battle of Neighborhoods

Report on Japanese Restaurant in the Battle of Neighborhoods

Introduction

A Japanese restaurant owner wants to explore opening an authentic Japanese restaurant in Toronto, . As Japanese food is very similar to other Japanese cuisines, this business owner is thinking of opening this restaurant in locations where Japanese food is popular (aka many Japanese restaurants in the neighborhood). With the purpose in mind, finding the location to open such a restaurant is one of the most important decisions for this business owner and I am designing this project to help him find the most suitable location.

Business Problem or Objective:

The objective is to find the most suitable location for the restaurant owner to open a new Japanese restaurant in Toronto, . With the help of machine learning methods such as clustering, this project aims to provide solutions to answer the business question: what is the best location to open a authentic Japanese restaurant.

Data Required to solve the problem:

- Data related to Japanese restaurants to help find the right place to establish.
- Scrapping the total list of neighborhoods in Toronto area through Wikipedia
- Acquiring coordinates of these neighborhoods through the available Geocoder package
- Foursquare to get data related to these neighborhoods

Initial step is to get the list of neighborhoods in Toronto. This can be done by extracting the list of neighborhoods from Wikipedia page. I did the web scraping by utilizing pandas html table scraping method as it is easier and more convenient to pull tabular data directly from a web page into data frame.

However, it is only a list of neighborhood names and postal codes. I will need to get their coordinates to utilize Foursquare to pull the list of venues near these neighborhoods. To get the coordinates, I tried using Geocoder package but it was not working so I used the csv file provided by IBM team to match the coordinates of Toronto neighborhoods. After gathering all these coordinates, I visualized the map of Toronto using Folium package to verify whether these are correct coordinates.

The next step in the methodology is to use Foursquare API to pull the list of top 100 venues within 500 meters radius. I have created a Foursquare developer account in order to obtain account ID and API key to pull the data. From Foursquare, I am able to pull the names, categories, latitude and longitude of the venues. With this data, I can also check how many unique categories that I can get from these venues. Then, I analyze each neighborhood by grouping the rows by neighborhood and taking the mean on the frequency of occurrence of each venue category. This is to prepare clustering to be done later.

Here, I made a justification to specifically look for "Japan restaurants". Previously, when I ran the model, I was looking for "Japan restaurants" but there are very few results (maybe due to Foursquare categorization) so I looked for the restaurants closest to Japanese cuisine taste (side note: Japanese food and Japan food are very similar in taste, so my justification is that if there are people who enjoyed Japan food, they likely are going to enjoy Japanese food too!)

Lastly, I performed the clustering method by using k-means clustering. K-means clustering algorithm identifies k number of centeriods, and then allocates every data point to the nearest cluster, while keeping the centroids as small as possible. It is one of the simplest and popular unsupervised machine learning algorithms and it is highly suited for this project as well. I have clustered the neighborhoods in Toronto into 3 clusters based on their frequency of occurrence for "Japan food". Based on the results (the concentration of clusters), I will be able to recommend the ideal location to open the restaurant.

Final Results :

These are the clusters.
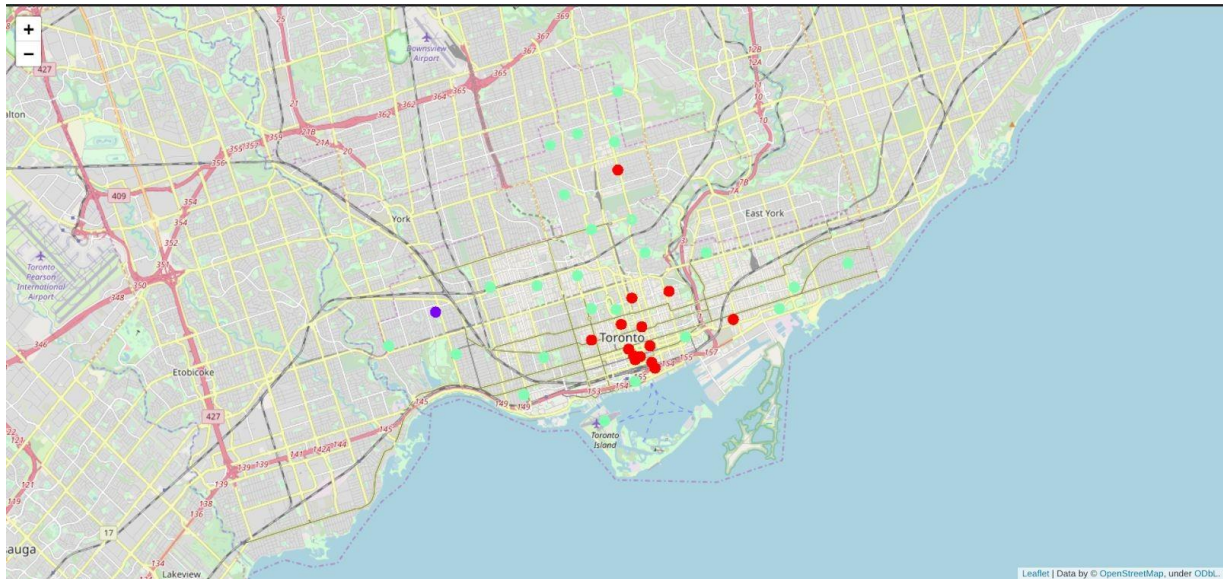
The results from k-means clustering show that we can categorize Toronto neighborhoods into 3 clusters based on how many Japan restaurants are in each neighborhood:
- Cluster 0: Neighborhoods with little or no Japan restaurants
- Cluster 1: Neighborhoods with no Japan restaurants
- Cluster 2: Neighborhoods with high number of Japan restaurants

The results are visualized in the above map with Cluster 0 in red color, Cluster 1 in purple color and Cluster 2 in light green color.

Recommendations :

Most of Japan restaurants are in Cluster 2 which is around Adelaide, King, Richmond areas and lowest (close to zero) in Cluster 1 areas which are North Toronto West and Parkdale areas. Also, there are good opportunities to open near Chinatown, St James town as the competition seems to be low. Looking at nearby venues, it seems Cluster 1 might be a good location as there are not a lot of Japan restaurants in these areas. Therefore, this project recommends the entrepreneur to open an authentic Japanese restaurant in these locations with little to no competition.

Conclusion :

In this project, we have gone through the process of identifying the business problem,
specifying the data required, extracting and preparing the data, performing the
machine
learning by utilizing k-means clustering and providing recommendation to the stakeholder.

List of References :

List of neighborhoods in Toronto:
https://en.wikipedia.org/wiki/List_of_postal_codes_of_:_M

Foursquare Developer Documentation: https://developer.foursquare.com/docs