www.github.com/ritchatterjee

IBM Data Science Capstone Report: Opening a new Sushi Restaurant in Toronto

1. Introduction

Over the last few years, succeeding in the Restaurant business has become a notoriously difficult task to achieve. Not only is it a relatively capital and employee-intensive business, restaurants are also highly regulated (especially under the post COVID-19 scenario), have low profit margins and in most neighborhoods have a plethora of competition to deal with. To build a niche and remain relevant, prospective, as well as established, restaurants have to hedge their bets with well-rounded and well-directed dataset. Data science provides valuable insights regarding market trends and evolving consumer lifestyles so that restaurateurs can better address and meet public demand.

In today's market, no particular restaurant is ever going to appeal to everyone. As the growth in the restaurant industry has flourished, so has the wide ranges and taste profiles for the consumer.

2. Research Question

My client is a very popular Sushi restaurant chain originating from the UK that specializes in a conveyor belt-style sushi for dine in along with home delivery options. They have already established themselves as one of the marquee restaurant chains in the UK and Ireland and are now looking to make inroads into the Canadian Market.

The objective of this research is to find and recommend which neighborhood in Toronto would be the best choice to open their first restaurant in Canada.

3. Data

In order to generate insights that solve this question, I have used the following datasets:

- A list of Neighborhoods grouped by boroughs for the city of Toronto
- The corresponding geolocation data (latitudes and longitudes) to create the centroids for these neighborhoods.
- Location venue data pertaining to Japanese (and in particular, Sushi Restaurants). This will help us identify what particular neighborhoods are most suitable to open a new sushi restaurant.

	Postalcode	Borough	Neighbourhood	Latitude	Longitude
37	M4E	East Toronto	The Beaches	43.676357	-79.293031
41	M4K	East Toronto	The Danforth West, Riverdale	43.679557	-79.352188
42	M4L	East Toronto	India Bazaar, The Beaches West	43.668999	-79.315572
43	M4M	East Toronto	Studio District	43.659526	-79.340923
44	M4N	Central Toronto	Lawrence Park	43.728020	-79.388790

Figure 1: Neighbourhood Data w/coordinates for Toronto

The Neighborhood data is scrapped from a Wikipedia Page[1] using the Beautiful Soup python package. The geolocation data is extracted from a cognitive class (cocl)[2] dataset and the venue data has been extracted from Foursquare.com using their API[3].

4. Methodology

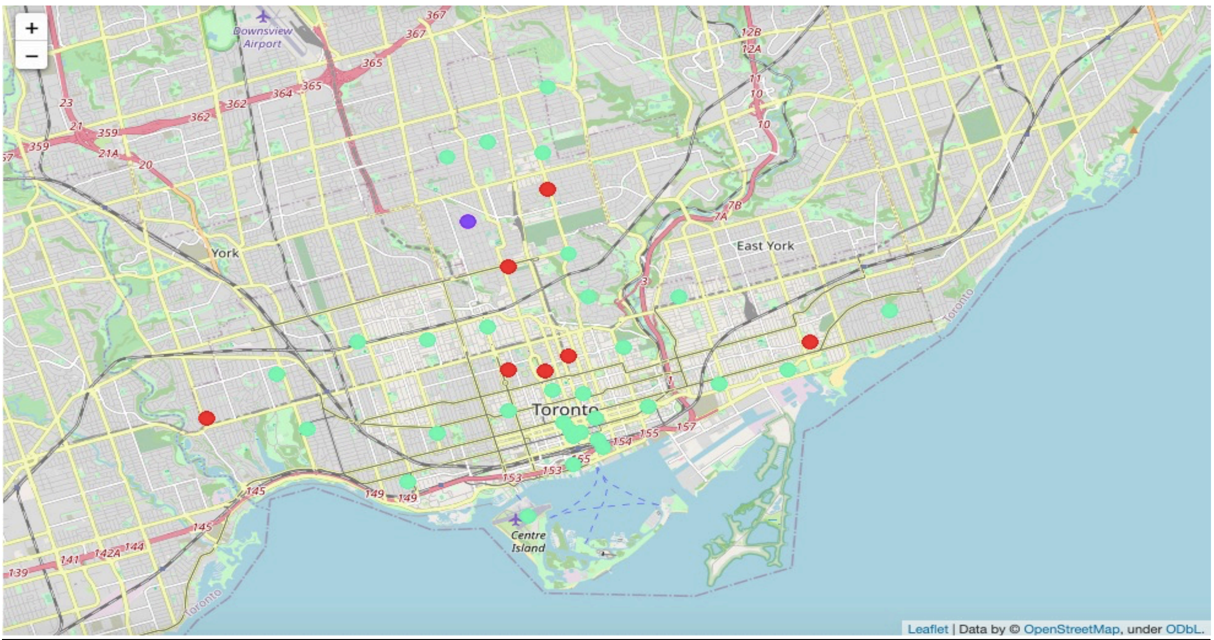
The first task was to extract the neighborhood data using beautiful soup. After scraping the data from the web, I structured it into a pandas data frame – this gave me a dataset of neighborhoods along with their corresponding boroughs. In order to leverage the use of the venue data, I would have to extract the geolocation data from the cognitive class dataset. I then extracted this data and merged it with the existing dataframe of neighborhoods. I could then use the folium package to verify whether the correct geolocation data was indeed accurate.

The next step was to use the Foursquare API to generate the list of venues. I chose to extract the top 100 venues within a radius of 500 meters. In order to generate the API key, I created a foursquare developer account. The Foursquare API data gave me a complete list of venues by name, address, coordinates and venue category (restaurants, bars, gyms etc.). I could use this data to see the unique categories of venues present. I then grouped this data by neighborhood to see the frequency each venue category occurring within a neighborhood. This step was required to help me define the criteria for the next step.

I then clustered the data based on the frequency of "Sushi Restaurants" in each neighborhood using K-Means clustering. The K-Means algorithm identifies k clusters and allocates every datapoint to the nearest cluster based upon the frequency of occurrence of "Sushi Restaurants". Based on these results, I will be able to recommend potential locations to open up new restaurant(s).

5. Results

The results of the K-means clustering are shown below:



These results show that there are 23 neighborhoods in Toronto having at least 1 or more Sushi Restaurants. There are 3 clusters that we can group these neighborhoods into based on how many Sushi Restaurants are there in each neighborhood.

- Cluster 0 (Red): Neighborhoods with little to no Sushi Restaurants.
- Cluster 1 (Purple): Neighborhoods with no Sushi Restaurants
- Cluster 2 (Light Green): Neighborhoods with a moderate to high number of Sushi Restaurants.

6. Evaluation

These results show that there are 23 neighborhoods in Toronto having at least 1 or more Sushi Restaurants. From this analysis, we can see that there is a high concentration of Sushi Restaurants near south-central Toronto which include areas of Downtown Toronto such as Commerce Court – Victoria Hotel and Regent Park. However, we do find a pocket of area starting from Central Bay Street up to North Midtown that has a relatively low presence of Sushi Restaurants and where competition will not be as intense. Furthermore, we also believe that looking further east, towards India Bazaar also presents particular opportunities as there are many non-Japanese Asian restaurants in that area and thus attracts a consumer who is particularly interested in Asian food. Furthermore, these areas are far closer to apartment complexes and other residential areas making it easier and cheaper to order Sushi at home.

7. Limitations and Scope for Future Research

This project primarily takes one factor consideration - the presence of Japanese (and Asian) restaurants in each Neighborhood. However, there are many other factors that can be incorporated into this search such as:

- Foot-traffic Data
- Population Density and Demographic Distribution
- Mean, median and standard deviations of income of residents in each neighborhood
- Consumer Survey Data to identify the specific demand in the market
- Location Rentals and Capital Leasing costs to justify the financial feasibility of opening the restaurant

To analyze the merits for each neighborhood with all the above data is not only a long but also an expensive process. However, utilizing the above data sources will lead to a much more informed and complete recommendation that will better suit the client.

8. Conclusion

In this project we have gone through the process of defining a specific business problem for a client, identifying and gathering data from a variety of web sources, performing k-means clustering to generate insights and recommendations that we have delivered to the client.

9. Sources

[1] https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

[2] http://cocl.us/Geospatial_data

[3] http://developer.foursquare.com/docs