

# THE BEST PLACE

## FINAL REPORT- CAPSTONE PROJECT

(The Battle of Neighborhoods)

Student:

KSHITIJ PAL

Course:

APPLIED DATA SCIENCE CAPSTONE (IBM)

# Introduction

Considering a case where an entrepreneur is looking to open up a Japanese restaurant in the Toronto city, since Japanese culture and food are pretty much famous all over the world, but opening a restaurant from scratch is a big deal. Therefore, selecting the right place to for it is first of many important steps.

Focusing on that point, it should be somewhere where the actual Japanese (or Asian) people reside in the city. So, for finding the suitable location to open up the restaurant is the aim of this project.

It will help entrepreneurs to decide where to invest actually before starting the actual thing.

## Problem

The objective of this problem is to find the most suitable location for the entrepreneur to open a new Japanese restaurant in Toronto, Canada. By using the Data Science methods, and tools along Machine Learning algorithms such as clustering, this project will provide solution to the business question:

In Toronto, if an Entrepreneur wants to open a Japanese restaurant, where should they consider opening it?

# Data Section

To solve this problem, the different kind of data required is:

- List of neighborhoods in Toronto, Canada
- Latitude and Longitude of these neighborhoods
- Venue data related to Japanese restaurants. This will help in finding the optimal place for opening a new Japanese restaurant.

## Data sources

[https://en.wikipedia.org/wiki/List\\_of\\_postal\\_codes\\_of\\_Canada: M](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada: M)

[http://cocl.us/Geospatial\\_data](http://cocl.us/Geospatial_data)

## Extracting the data

- The scraping of Toronto neighborhoods from Wikipedia
- Getting latitude and longitude data of these neighborhoods via Geocoder package
- Using Foursquare API to get venue data related to these neighborhoods

After finding the list of neighborhoods, we then connect to the Foursquare API to gather information about venues inside each and every neighborhood. For each neighborhood, we have chosen the radius to be 100 meters.

The data retrieved from Foursquare contained information of venues within a specified distance of the longitude and latitude of the postcodes.

# Methodology

First web scraping was done, by utilizing pandas HTML table scraping method as it is easier and more convenient to pull tabular data directly from a web page into the data frame. Since it is just the list of neighborhoods, therefore coordinates were gathered from the Geocoder package according to the Postal codes.

After that, using folium package, the map of Toronto was generated. Then, it was required to generate the list of venues, using the foursquare API, the requirement was an account on the foursquare website, a key was given, using that API call were made to get the list of venues in the Toronto according to the latitude and longitude.

Then comes the analyzing part by grouping the entire data according to some neighborhood and taking the mean on the frequency of occurrence of each venue category.

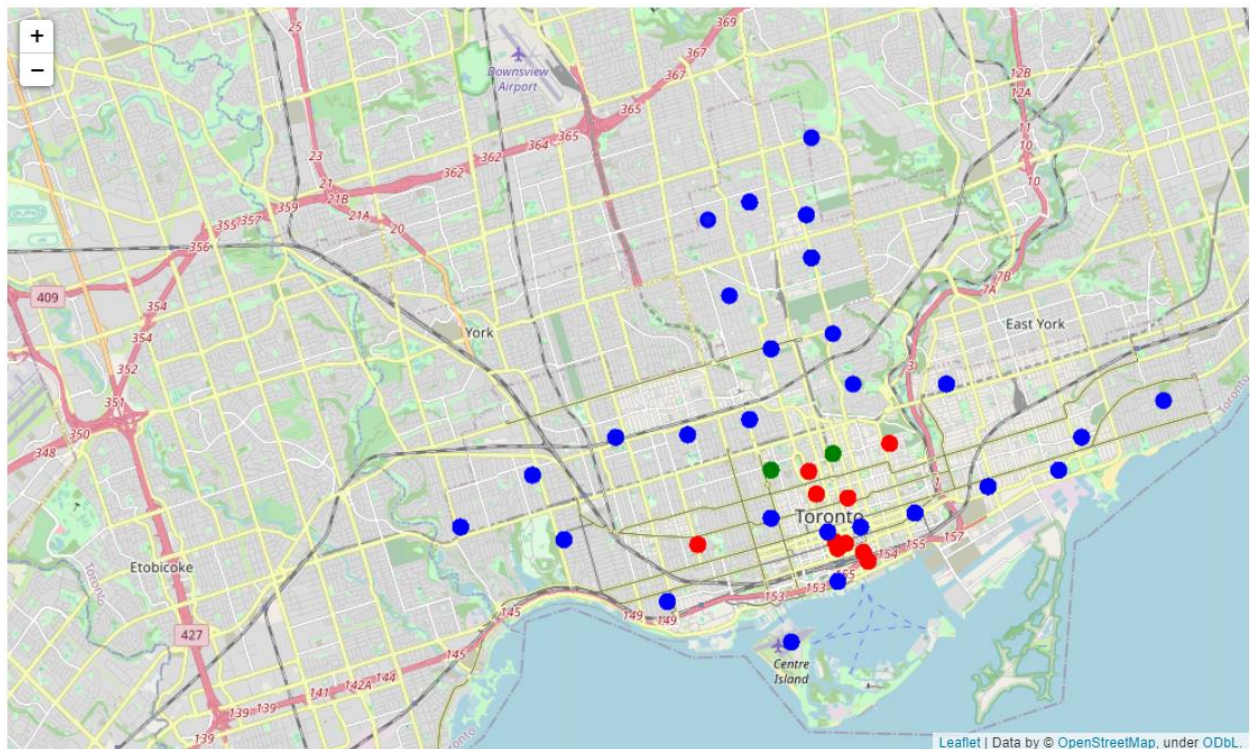
Since the need is to look for “Japanese restaurant”, therefore clustering was performed using K-means. As it identifies K number of centroids, allocating each data point to the nearest cluster. Based on the results, one can recommend the optimal place for opening a new Japanese restaurant.

```
In [49]: 1 toronto_venues['Venue Category'].unique()[:100]
```

```
Out[49]: array(['Trail', 'Health Food Store', 'Pub', 'Neighborhood',  
               'Cosmetics Shop', 'Ice Cream Shop', 'Greek Restaurant',  
               'Italian Restaurant', 'Brewery', 'Juice Bar', 'Yoga Studio',  
               'Fruit & Vegetable Store', 'Dessert Shop', 'Pizza Place',  
               'Restaurant', 'Bookstore', 'Grocery Store',  
               'Furniture / Home Store', 'Spa', 'Bubble Tea Shop',  
               'Caribbean Restaurant', 'Coffee Shop', 'Bakery',  
               'Indian Restaurant', 'Café', 'Frozen Yogurt Shop', 'Lounge',  
               'American Restaurant', 'Liquor Store', 'Sushi Restaurant', 'Gym',  
               'Fish & Chips Shop', 'Fast Food Restaurant', 'Park',  
               'Burrito Place', 'Pet Store', 'Steakhouse', 'Movie Theater',  
               'Sandwich Place', 'Food & Drink Shop', 'Fish Market',  
               'Seafood Restaurant', 'Gay Bar', 'Cheese Shop',  
               'Middle Eastern Restaurant', 'Comfort Food Restaurant',  
               'Stationery Store', 'Thai Restaurant', 'Coworking Space',  
               'Wine Bar', 'Latin American Restaurant', 'Bar', 'Gastropub',  
               'Gym / Fitness Center', 'Bank', 'Clothing Store',  
               'Convenience Store', 'Diner', 'Swim School', 'Bus Line',  
               'Breakfast Spot', 'Department Store', 'Hotel',  
               'Chinese Restaurant', 'Salon / Barbershop', 'Mexican Restaurant',  
               'Sporting Goods Shop', 'Health & Beauty Service', 'Shoe Store',  
               'Toy / Game Store', 'Gas Station', 'Farmers Market',  
               'Gourmet Shop', 'Pharmacy', 'Discount Store', 'Tennis Court',  
               'Dance Studio', 'Summer Camp', 'Playground', 'Supermarket',  
               'Sports Bar', 'Fried Chicken Joint', 'Vietnamese Restaurant',  
               'Light Rail Station', 'Bagel Shop', 'Japanese Restaurant',  
               'Jewelry Store', 'General Entertainment', 'Butcher',  
               'Deli / Bodega', 'Taiwanese Restaurant', 'Gift Shop', 'Market',  
               'Beer Store', 'Snack Place', 'Theme Restaurant', 'Beer Bar',  
               'Ramen Restaurant', 'Creperie', 'Men's Store'], dtype=object)
```

# Results

Clusters:



The results from k-means clustering show that we can categorize Toronto neighborhoods into 3 clusters based on how many Indian restaurants are in each neighborhood:

The green one represents cluster 1, with a significantly less restaurants.

The blue one represents cluster 0, with highest number of restaurants.

At last the red one represents cluster 2, with somewhere in the middle of the above two.

# Conclusion

Most of the Japanese restaurants are in Cluster 0, which is around :

Garden city, Ryerson, Underground city, Cabbagetown, Central Bay street, Berczy park and Victoria hotel

And the least are in cluster 1, which is around :

St. James town, Adelaide, Toronto island, Grange park and Chinatown

So it will be beneficial for the new entrepreneurs to set up restaurants in the Cluster 1, because there will be less competition and more profit in that location.

## Libraries Which are Used to Develop the Project:

Pandas: For creating and manipulating dataframes.

Folium: Python visualization library would be used to visualize the neighborhoods cluster distribution of using interactive leaflet map.

Scikit Learn: For importing k-means clustering.

JSON: Library to handle JSON files.

XML: To separate data from presentation and XML stores data in plain text format.

Geocoder: To retrieve Location Data.

Beautiful Soup and Requests: To scrap and library to handle http requests.

Matplotlib: Python Plotting Module.