

Capstone Project - The Battle of Neighborhoods (Week 2)

Venue Suggestion to Opening a Japanese Sushi Restaurant in Toronto, Canada

Yui Tak Cheung

1. Introduction

1.1 Background

This project aims to help us in finding an optimal location to open a Japanese Sushi Restaurant in Toronto, Canada. In other words, this project will be delivering a reference venue for people who are interested in opening a Sushi store in Toronto, Canada.

Toronto is one of the diversified and wealthiest cities in Canada. The multicultural city attracts many restaurants with different cuisines. In this project, we will only focus on all areas in Toronto. We will conclude the existing Sushi restaurants compared to all other restaurants, and will use clustering to find similar areas in Toronto regarding restaurants of each borough. The preferred area shall be distant from existing restaurants. We will use data science tools to fetch the raw data, visualize it then generate a few most promising areas based on the above criteria. We will also explain the advantages and traits for the candidates, so that stakeholders can make the final decision based on the analysis.

1.2 Data Requirement

Based on the definition of our problem, factors that may impact our decision are the number of existing restaurants in the neighborhood and nearby. Restaurant data in every neighborhood will be obtained using Foursquare API. We will extract the coordinates of all restaurants. With the information about venues around a given neighborhood, we can create a dataframe where we can see what kind of venue appears around the neighborhoods in that city and their frequency of appearance. These frequencies are weighted. We can then weight the frequency of these venues in our dataframe for each neighborhood.

2. Methodology

2.1 Data sources

First we obtain the data of postal codes in Toronto by using BeautifulSoup constructor to find the necessary html table in wikipedia. Next we gather the coordinates (latitude and longitude) of each Postal code from the csv file provided from the following link:

https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDriverSkillsNetwork-DS0701EN-SkillsNetwork/labs_v1/Geospatial_Coordinates.csv

2.2 Foursquare

At the same time, we will download and import the necessary libraries that we will use for the project

We then use Foursquare API to get the information on food venues in each neighborhood. As we know that Japanese foods are diverse and separated between each specific cuisine, for example, a Sushi restaurant can be considered different from a Ramen restaurant. So we are only interested in venues in the 'Sushi Restaurant' category only. and will not consider other kinds of Japanese restaurants.

We will generate Asian restaurant venues based on the distance to the neighborhoods. We will take the radius as 2000 from the centre of the neighborhood and limit it to 2000 restaurants for each neighborhood.

Finally, we will extract the name, latitude and longitude, postal code and distance of the food venue from the json file we received using the foursquare api and collect it in a dataframe.

2.2 Data Wrangling

Data downloaded or scraped from multiple sources were combined into one table. In this section, we will do data wrangling of the data we gathered for the purpose of determining the density of food venues in each of Toronto's neighborhoods, and the type of food venues that are the most common in the neighborhoods. We verify with the collected data in terms of duplications, un-matched addresses by their name, location, category, postal code and distance of every food venue in the city of Toronto.

2.3 Data Analysis and Modelling

In the analysis and modelling section, we will present a heatmap to show the density of neighborhoods in terms of food venues and focus on what implications these can have for stakeholders. In addition, we will provide basic information to the stakeholders by considering the most common types of food venues in these neighborhoods. This will allow stakeholders to choose a starting point at the

neighborhood level. And finally, we will create clustering (using k-means clustering) to offer stakeholders different alternatives in terms of neighborhood.

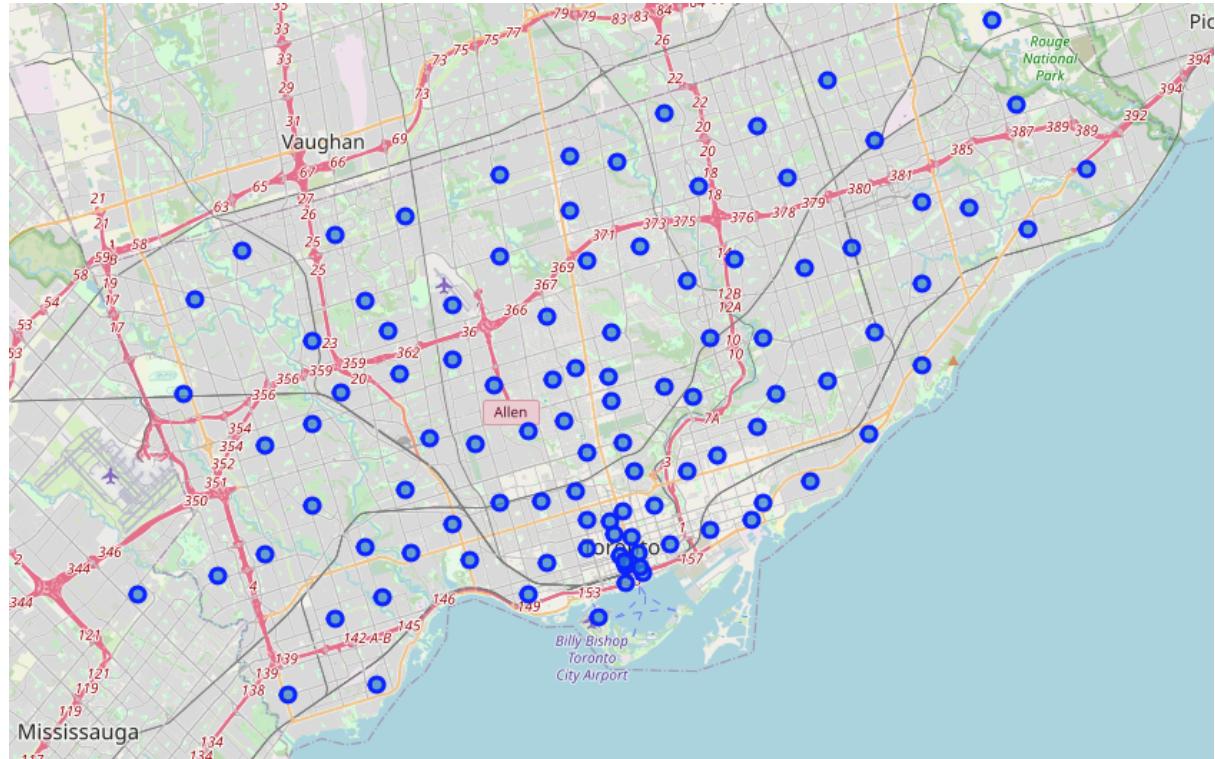
3. Analysis

3.1 Data Cleaning

We firstly found 15 of Boroughs in Toronto and 103 of Neighborhoods in Toronto, and their corresponding coordinates.

	Postal Code	Borough	Neighborhood	Latitude	Longitude
0	M3A	North York	Parkwoods	43.753259	-79.329656
1	M4A	North York	Victoria Village	43.725882	-79.315572
2	M5A	North York	Regent Park, Harbourfront	43.654260	-79.360636
to scroll output; double click to hide			Lawrence Manor, Lawrence Heights	43.718518	-79.464763
3	M6A	North York	Ontario Provincial Government	43.662301	-79.389494
4	M7A	Queen's Park			

Then we can visualize the data we have the geographical coordinate of Toronto are 43.6534817, -79.3839347.



Next, we extract the name, latitude and longitude, postal code and distance of the food venue from the json file we received using the foursquare api and collect it in a dataframe. We have found 1125 Asian restaurants in Toronto in which there are 184 Sushi restaurants. We also know which food venues exactly are in the vicinity of every neighborhood. Now we have the required data for our analysis.

	Neighborhood Postal Code	Borough	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Postal Code	Venue Distance	Venue Category	VPostalcode
0	M3A	North York	Parkwoods	43.753259	-79.329656	Dragon Pearl Buffet 龍珠	43.753693	-79.349730	M3B	1614	Asian Restaurant	M3B
1	M3A	North York	Parkwoods	43.753259	-79.329656	Gonoe Sushi	43.745737	-79.345991	M3B	1557	Japanese Restaurant	M3B
2	M3A	North York	Parkwoods	43.753259	-79.329656	Thai One On	43.754583	-79.351543	M3B	1765	Thai Restaurant	M3B
3	M3A	North York	Parkwoods	43.753259	-79.329656	China Gourmet	43.755189	-79.348382		1520	Asian Restaurant	
4	M3A	North York	Parkwoods	43.753259	-79.329656	Matsuda Japanese Cuisine & Teppanyaki	43.745494	-79.345821	M3B	1561	Japanese Restaurant	M3B

	Neighborhood Postal Code	Borough	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Postal Code	Venue Distance	Venue Category
6	M3A	North York	Parkwoods	43.753259	-79.329656	Sushi Ichiban	43.758912	-79.310811		1640	Sushi Restaurant
12	M3A	North York	Parkwoods	43.753259	-79.329656	Robo Sushi	43.754837	-79.349603	M3B 1Y6	1613	Sushi Restaurant
28	M4A	North York	Victoria Village	43.725882	-79.315572	Bento Sushi	43.722265	-79.337631		1819	Sushi Restaurant
46	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636	Qi sushi	43.662552	-79.364258		968	Sushi Restaurant
53	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636	Tachi	43.650596	-79.383396	M5H 2G4	1878	Sushi Restaurant

There are in total 3915 rows of Asian restaurants found in Toronto. Since there is overlap of restaurants between neighborhoods, we will first make matches using postal codes and make them the same type. Afterwards, we compare postal codes and mark which venue is located in the right neighborhood or not. We shall not consider restaurants that do not have postal codes.

2817 restaurants with postal codes remain. Now we can remove the duplicated ones by matching the same postal code and the same restaurant name.

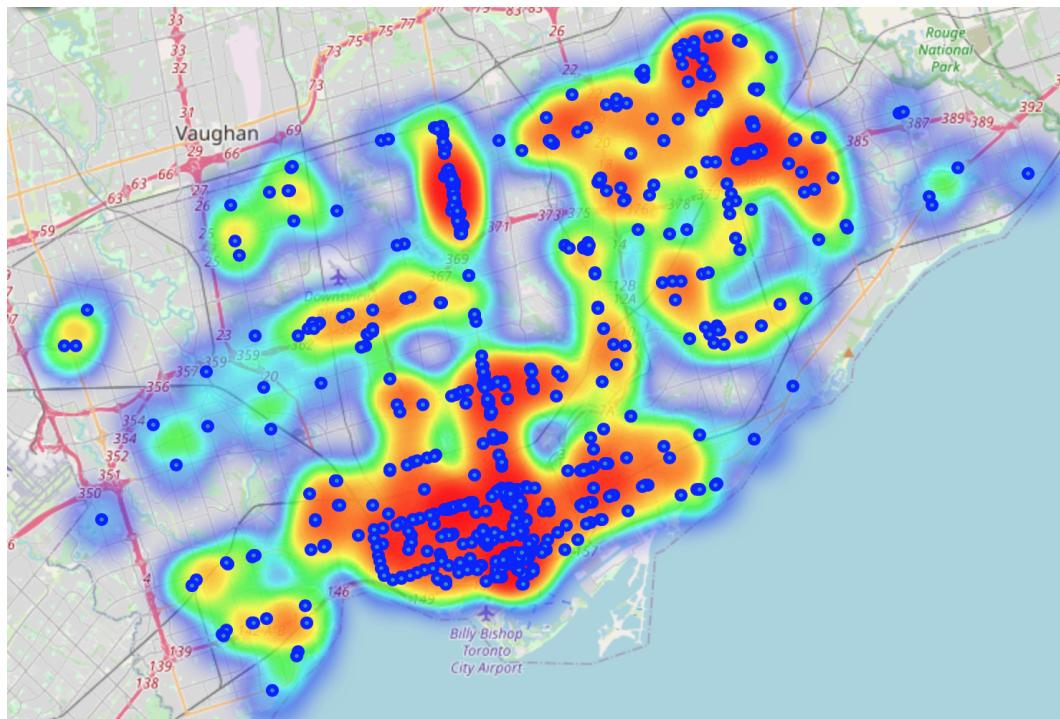
We then remove duplicated restaurant names and mark restaurants with the same postal code as the neighborhood postal code as True, otherwise as False.

```
False    2121
True     696
Name: Flag, dtype: int64
```

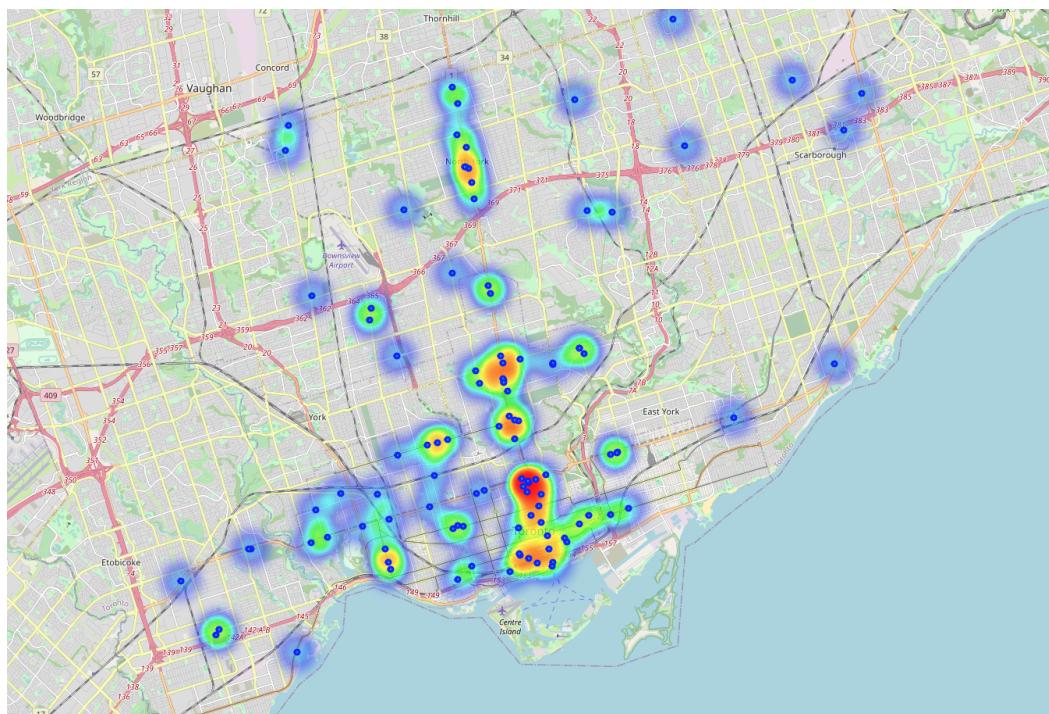
We can now finalize 696 Asian restaurants, with 107 Sushi restaurants in it, for further analysis

3.2 Heatmap

We can create a map showing the heat map / density of food venues and try to extract meaningful information from it. Also let's show the neighborhoods on our map for comparisons.



Looks like a few pockets of high restaurant density apart from the city center can be found on the north and on the east. Below is another heatmap showing the density of Sushi restaurants only.



Showing only sushi restaurants, the map looks less hot, and we can see that except the neighborhoods at Downtown, Willowdale, North York has a good density of Asian restaurants as well as Sushi restaurants. As can be understood from here, this area is one of the major food/dining areas.

By sorting the number of venues by neighborhood, It is found out that Willowdale, North York is not the neighborhood that owns the most Asian restaurants, but also the most Sushi restaurants in Toronto.

	Neighborhood Postal Code	Borough	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Postal Code	Venue Distance	Venue Category	VPostalcode	Flag
Neighborhood												
Willowdale South	60	60	60	60	60	60	60	60	60	60	60	60
Milliken, Agincourt North, Steeples East, L'Amoreaux East	47	47	47	47	47	47	47	47	47	47	47	47
Christie	39	39	39	39	39	39	39	39	39	39	39	39
Agincourt	37	37	37	37	37	37	37	37	37	37	37	37
CN Tower, King and Spadina, Railway Lands, Harbourfront West, Bathurst Quay, South Niagara, Island airport	26	26	26	26	26	26	26	26	26	26	26	26

	Neighborhood Postal Code	Borough	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Postal Code	Venue Distance	Venue Category	VPostalcode	Flag
Neighborhood												
Willowdale South	7	7	7	7	7	7	7	7	7	7	7	7
High Park, The Junction South	5	5	5	5	5	5	5	5	5	5	5	5
Davisville	5	5	5	5	5	5	5	5	5	5	5	5
CN Tower, King and Spadina, Railway Lands, Harbourfront West, Bathurst Quay, South Niagara, Island airport	5	5	5	5	5	5	5	5	5	5	5	5
Christie	4	4	4	4	4	4	4	4	4	4	4	4

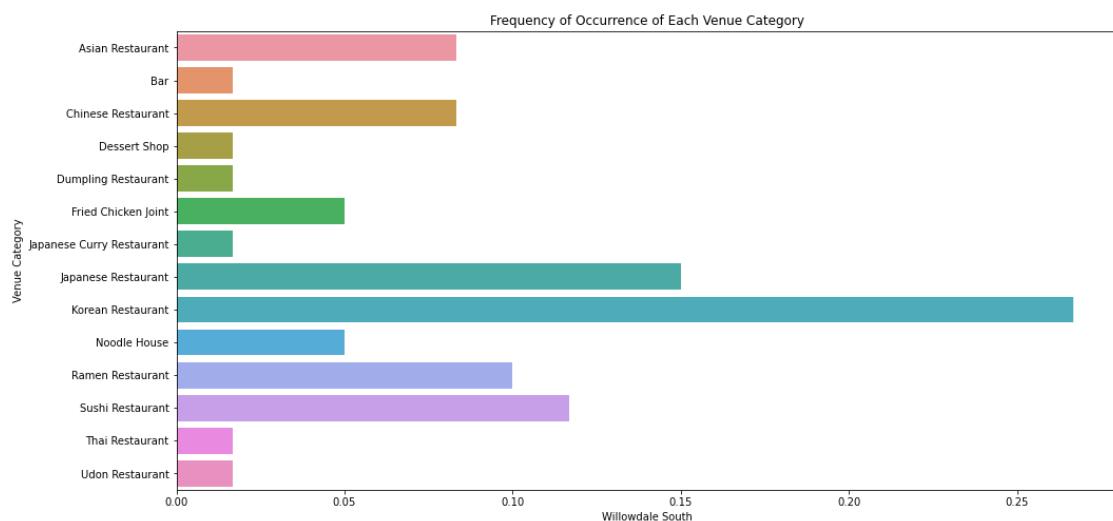
Next, let's group rows by neighborhood and by taking the mean of the frequency of occurrence of each category

	Neighborhood	Asian Restaurant	BBQ Joint	Bakery	Bar	Breakfast Spot	Bubble Tea Shop	Buffet	Café	Cambodian Restaurant	...	Sushi	Szechuan	Taiwanese Restaurant	Tea Room	Thai Restaurant	Til Resta
0	Agincourt	0.162162	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.054054	0.0	0.0	0.0	0.0	
1	Bathurst Manor, Wilson Heights, Downsview North	0.250000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.250000	0.0	0.0	0.0	0.0	
2	Bayview Village	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.000000	0.0	0.0	0.0	0.0	
3	Bedford Park, Lawrence Manor East	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.500000	0.0	0.0	0.0	0.5	
4	Berczy Park	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.666667	0.0	0.0	0.0	0.0	

Now let's create the new dataframe and display the top 10 venues for each neighborhood.

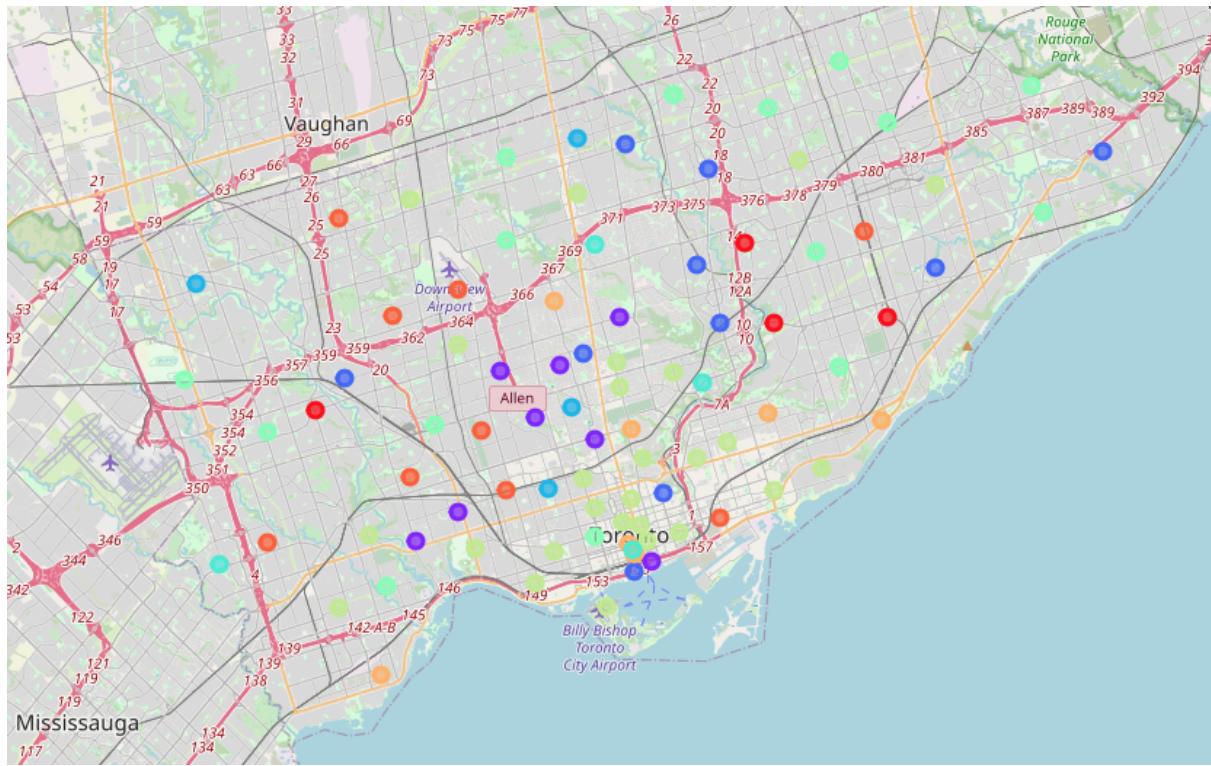
	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Agincourt	Chinese Restaurant	Asian Restaurant	Cantonese Restaurant	Vietnamese Restaurant	Sushi Restaurant	Hong Kong Restaurant	Noodle House	Filipino Restaurant	Shanghai Restaurant	Diner
1	Bathurst Manor, Wilson Heights, Downsview North	Asian Restaurant	Filipino Restaurant	Sushi Restaurant	Chinese Restaurant	Comfort Food Restaurant	Hong Kong Restaurant	Hakka Restaurant	Grocery Store	Fried Chicken Joint	Fast Food Restaurant
2	Bayview Village	Japanese Restaurant	Chinese Restaurant	Xinjiang Restaurant	Dessert Shop	Hong Kong Restaurant	Hakka Restaurant	Grocery Store	Fried Chicken Joint	Filipino Restaurant	Fast Food Restaurant
3	Bedford Park, Lawrence Manor East	Thai Restaurant	Sushi Restaurant	Xinjiang Restaurant	Coffee Shop	Hakka Restaurant	Grocery Store	Fried Chicken Joint	Filipino Restaurant	Fast Food Restaurant	Dumpling Restaurant
4	Berczy Park	Sushi Restaurant	Chinese Restaurant	Xinjiang Restaurant	Comfort Food Restaurant	Hong Kong Restaurant	Hakka Restaurant	Grocery Store	Fried Chicken Joint	Filipino Restaurant	Fast Food Restaurant

The Willowdale South neighborhood is the area with the most venues. As seen in the graph, the third common types of food venues in this region are Sushi Restaurants. However this region is very valuable for food venues, it may not be a good idea for a new business to open one of these types of venues, which are already in large numbers here.



3.3 Cluster Neighborhoods

Now let's cluster neighborhoods together to see similar neighborhoods. This will enable us to offer stakeholders different neighborhoods as an alternative.



4. Results and Discussion

Our analysis shows that the most popular area in the city of Toronto in terms of Asian food venues is the neighborhood Willowdale South in the North York borough. From this analysis can we understand that this area has great capability for Asian food venues. However, there are many things to consider when choosing a place for the food venue. For example it may be a valuable place which may not be an optimal choice for some food venues.

After directing our attention to this more narrow area of interest, we focus only on the competition and restaurant density. In order to evaluate this factor, we presented the neighborhood-based venues we collected and how often they were in which neighborhood as a result of our analysis. For example, although the area consisting of the neighborhood seems like an attractive location, it may not be a good idea for venues. If we are looking for a venue in similar style, we can also consider the neighborhood with the same cluster number, (i.e. 1), such as Don Mills North, The Beaches, Bayview Village etc, at the nearby areas.

Finally, the fact that this study is not inclusive may reveal some drawbacks. One of them is that the information we provide is far from offering pinpoint options for stakeholders. The reason for this is that our analyzes are far from the real field. Therefore, it should not be forgotten that the information we provide will only help stakeholders to choose a starting point in finding a new location for their business.

6. Conclusion

The aim of this project is to provide suggestions in venues to assist stakeholders who are considering opening a Sushi restaurant venue in the city of Toronto. There are many factors that affect the location of a food venue. In this study, analyzes were made considering only the competition and density.

By calculating Asian restaurant density distribution from Foursquare data we have first identified general neighborhoods. As a result of the analysis made afterwards, the density of asian food venues in the city was revealed. Thus, the competition factor at the general level was examined. Then, going a little more specific, the frequency of food venues in the neighborhoods was revealed as a result of analysis. Clustering of those locations was then performed in order to create major zones of interest; those zone centers were created to be used as starting points for final exploration by stakeholders.

Finally, it should be noted that the stakeholders should pay attention to all factors when choosing the location for the food venue. We can list these factors as follows: demography, labor costs and minimum wage, access, proximity to suppliers, competition, visibility, future growth, health regulations and zoning, security / crime rates. When the stakeholders consider all these factors, they will be able to choose the best location for their food venues.

7. Reference

- https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDriverSkillsNetwork-DS0701EN-SkillsNetwork/labs_v1/Geospatial_Coordinates.csv
- https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
- <https://developer.foursquare.com>