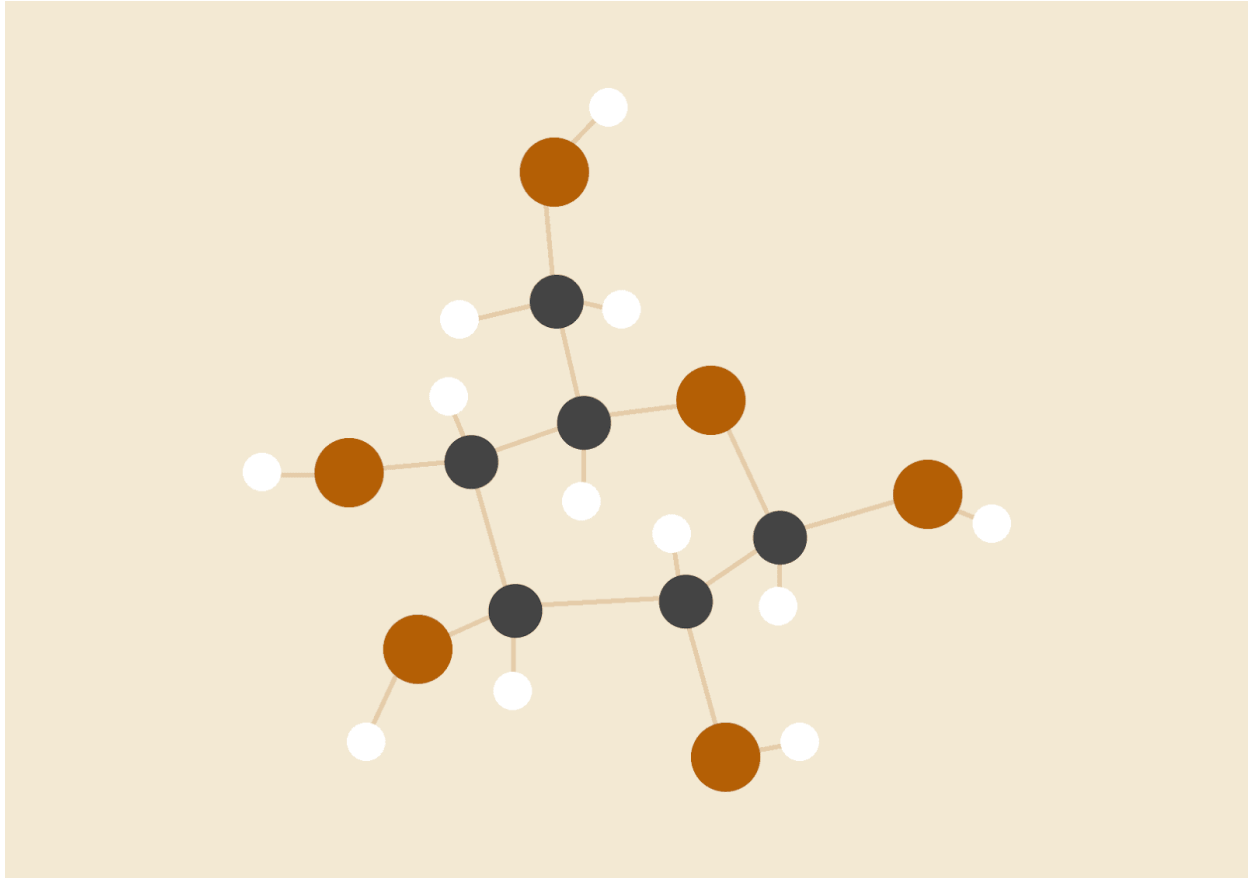


DATA SCIENCE REPORT

Site selection to open Sushi place



Dawit H. Hailu

dawit.hailu@mail.huji.ac.il

December, 2018

Coursera-IBM

INTRODUCTION

Imagine a friend of yours, who is also business person, has asked your help for site selection. The business she wants to open is a Sushi place somewhere in Toronto. After going through the budgetary concerns, the next issue --- and where your input as a data scientist is sought --- is related with the choice of location selection for the business as it is one of the biggest concerns she has. She is asking you _where is the best neighborhood to start the business?_. In this project we will use data science tools in helping us choose the best site.

Although this is a mini project aimed at addressing your friend's location choice, yet it is implied that anyone who interested to start a Sushi business in Toronto, and are not sure of where to open, can also be considered as our target audience.

According to wikipedia [1]

Sushi (すし, 寿司, 鮓, pronounced [su⁺ɕi] or [suɕi⁺]) is a Japanese dish of specially prepared vinegared rice (鮓飯 sushi-meshi), usually with some sugar and salt, combined with a variety of ingredients (ネタ neta), such as seafood, vegetables, and occasionally tropical fruits. Styles of sushi and its presentation vary widely, but the key ingredient is "sushi rice", also referred to as shari (しゃり), or sumeshi (酢飯). The term sushi is no longer used in its original context and literally means "sour-tasting"..

DATA ACQUISITION

This section presents the datasets used for this study and details of the methods used to perform our analysis. As the Toronto data that we need to explore, segment, and cluster the neighborhoods in the city of Toronto is not readily available, we need to scrape the Wikipedia page [2] and wrangle the data, clean it, and then read it into a pandas dataframe so that it is in a structured format ready for further use. Next because the data from wikipedia page lacks coordinates (i.e latitudes and longitudes), we will have to import the Geospatial_data from [3] and then Toronto neighborhood addresses are converted into their equivalent latitude and longitude values retrieved from Geospatial_data using Pandas library in Python. Table 1 shows the first 12 examples of a dataset comprising of Borough, Neighborhood, Latitude, and Longitude of Toronto neighborhoods.

	PostalCode	Borough	Neighbourhood	Latitude	Longitude
0	M3A	North York	Parkwoods	43.753259	-79.329656
1	M4A	North York	Victoria Village	43.725882	-79.315572
2	M5A	Downtown Toronto	Harbourfront,Regent Park	43.654260	-79.360636
3	M6A	North York	Lawrence Heights,Lawrence Manor	43.718518	-79.464763
4	M7A	Queen's Park	Queen's Park	43.662301	-79.389494
5	M9A	Etobicoke	Islington Avenue	43.667856	-79.532242
6	M1B	Scarborough	Rouge,Malvern	43.806686	-79.194353
7	M3B	North York	Don Mills North	43.745906	-79.352188
8	M4B	East York	Woodbine Gardens,Parkview Hill	43.706397	-79.309937
9	M5B	Downtown Toronto	Ryerson,Garden District	43.657162	-79.378937
10	M6B	North York	Glencairn	43.709577	-79.445073
11	M9B	Etobicoke	Cloverdale,Islington,Martin Grove,Princess Gar...	43.650943	-79.554724

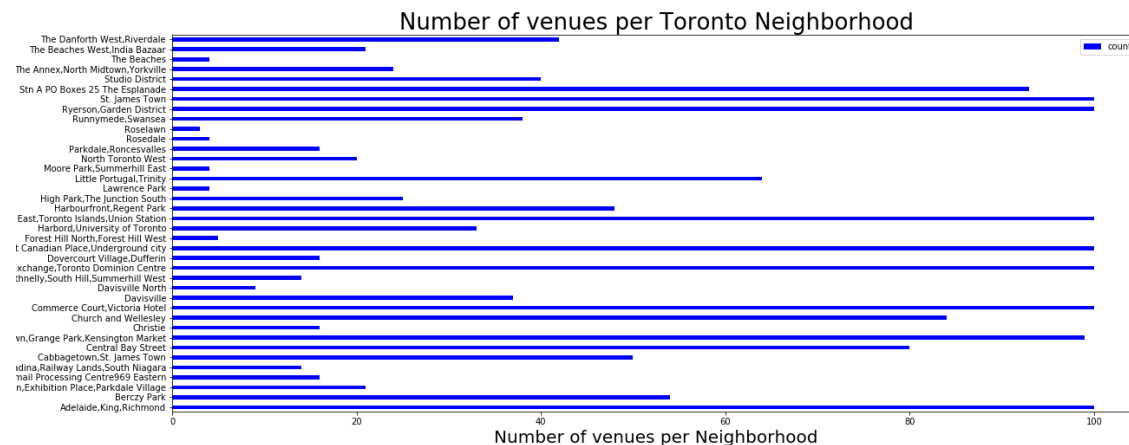
We explore the neighborhoods popular venues and the likes via data acquired using Foursquare [4]. The Foursquare API is used to explore neighborhoods in Toronto. Using the explore function, the most common venue categories within a radius of 500 meters is extracted for each neighborhood and the number of returned venues by Foursquare API is limited to 100. Table 2 presents the first 7 venues of dataset returned.

	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
0	Adelaide,King,Richmond	Coffee Shop	Café	Steakhouse	American Restaurant	Thai Restaurant	Gym	Clothing Store
1	Berczy Park	Coffee Shop	Cocktail Bar	Restaurant	Farmers Market	Bakery	Steakhouse	Cheese Shop
2	Brockton,Exhibition Place,Parkdale Village	Coffee Shop	Performing Arts Venue	Café	Breakfast Spot	Pet Store	Stadium	Grocery Store
3	Business reply mail Processing Centre969 Eastern	Yoga Studio	Auto Workshop	Comic Shop	Park	Pizza Place	Butcher	Recording Studio
4	CN Tower,Bathurst Quay,Island airport,Harbourf...	Airport Lounge	Airport Terminal	Airport Service	Harbor / Marina	Sculpture Garden	Boutique	Boat or Ferry
5	Cabbagetown,St. James Town	Coffee Shop	Restaurant	Bakery	Indian Restaurant	Italian Restaurant	Café	Pub
6	Central Bay Street	Coffee Shop	Café	Sandwich Place	Italian Restaurant	Bubble Tea Shop	Bar	Burger Joint

In this project, we will just use the location data from the Foursquare. By analyzing the venue categories of each neighborhood in Toronto, we will decide if the neighborhood is suitable to open a new Sushi place.

As the Foursquare provide us data for each neighborhood in Toronto. We will use the data acquired to examine different questions such as types, numbers, and distance of each venue. Thereby make a decision as to whether the identified location is good or bad for a Sushi business.

The following bar chart summarizes the result obtained from Foursquare by querying for popular venues in each neighborhood and the convert the output into data frame ready to produce the chart



METHODOLOGY

In this project, we will use the Foursquare API to explore neighborhoods in Toronto city. For simplicity we will work with those Borough whose name contains Toronto in it. We will use the **explore** function to get the most common venue categories in each neighborhood, and then use this feature to group the neighborhoods into clusters. We will use the k-means clustering algorithm to complete the grouping of neighborhoods into clusters. Finally, I will use the Folium library to visualize the neighborhoods in Toronto City and their emerging clusters.

As Sushi places can have different names we widen our search of existing Sushi places by using terms like: Sushi Restaurant, Japanese Restaurant, Asian Restaurant. To learn from the experience of the existing Sushi places we need to know what are the most common nearby venues around the existing Sushi places in each neighborhood and the result is displayed as list

```

Coffee Shop          98
Café                 60
Restaurant           36
Hotel                33
Bar                  26
Japanese Restaurant  24
Italian Restaurant   23
Pizza Place          23
Gym                  18
American Restaurant  18
dtype: int64

```

After we examine the dataset history and see what are the popular venues around a Sushi restaurant, next we examine the average number of Sushi place for each neighborhood. Then we are looking for all the neighborhoods, which have the top 5 venues categories. We choose the least number of Sushi place in the vicinity as the suggested neighborhood.

RESULTS

Firstly, we point out that for all the neighborhoods which has at least one Sushi place, the top venues categories (put aside Japanese restaurant) are: coffee shop and café, Restaurant, Hotel, Bar, Italian Restaurant, Pizza place, and Gym. As I consider Japanese Restaurant to be part of the category of Sushi place, I omit from it from the list of the top categories. I do so because from this experience we learn that the optimal neighborhoods to start a Sushi business should have the following five venues categories: Coffee shop and café, Restaurant, Hotel, Bar, Pizza Place respectively.

Furthermore, we calculated the average number of Sushi near those neighborhoods to be 3.5. With the two conditions. First the top five categories of venues in a neighborhood and second we demand less than 4 Sushi places in a neighborhood,

We next identified neighborhoods to have met the conditions. Finally, we calculate the total number of Sushi places in those neighborhoods. Here are the results:

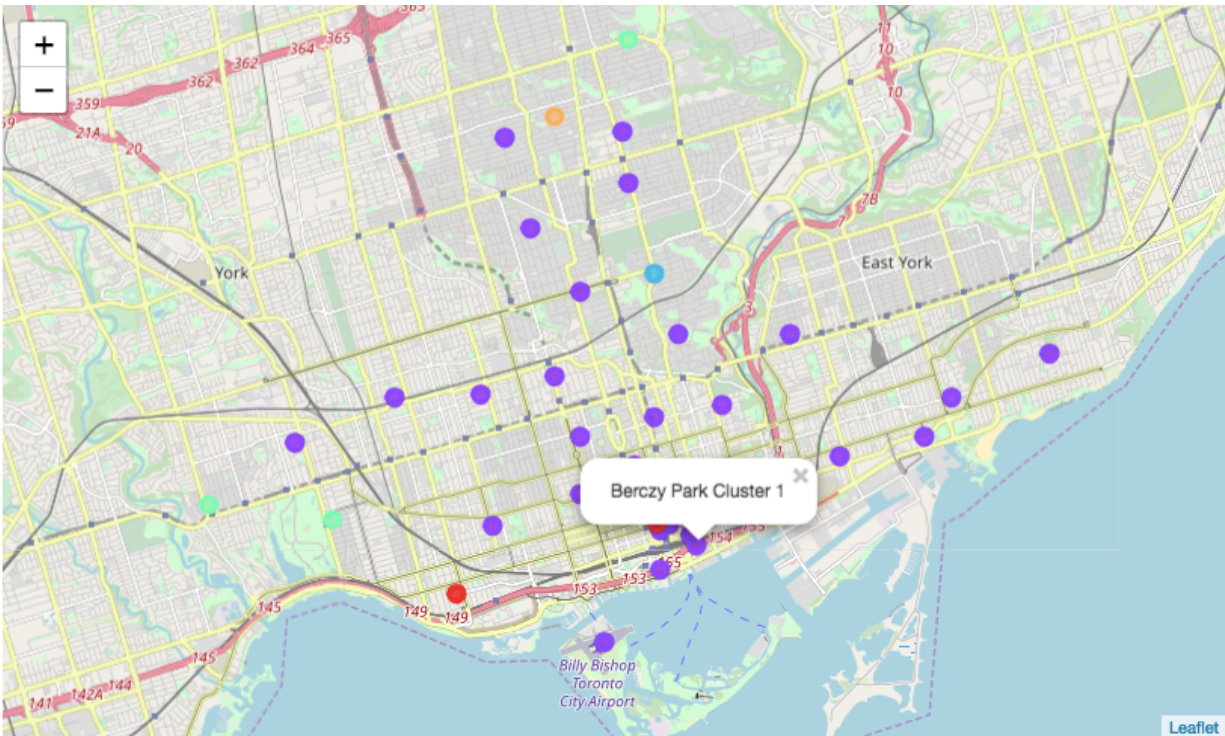
```

Neighborhood
Berczy Park                                0
Dovercourt Village,Dufferin                0
High Park,The Junction South                0
Studio District                            0
The Danforth West,Riverdale                 0
Cabbagetown,St. James Town                  1
Chinatown,Grange Park,Kensington Market    1
Commerce Court,Victoria Hotel              1
Harbourfront,Regent Park                   1
Harbourfront East,Toronto Islands,Union Station 2
Little Portugal,Trinity                     2
St. James Town                             2
Stn A PO Boxes 25 The Esplanade             2
Design Exchange,Toronto Dominion Centre    3
Harbord,University of Toronto              3
dtype: int64

```

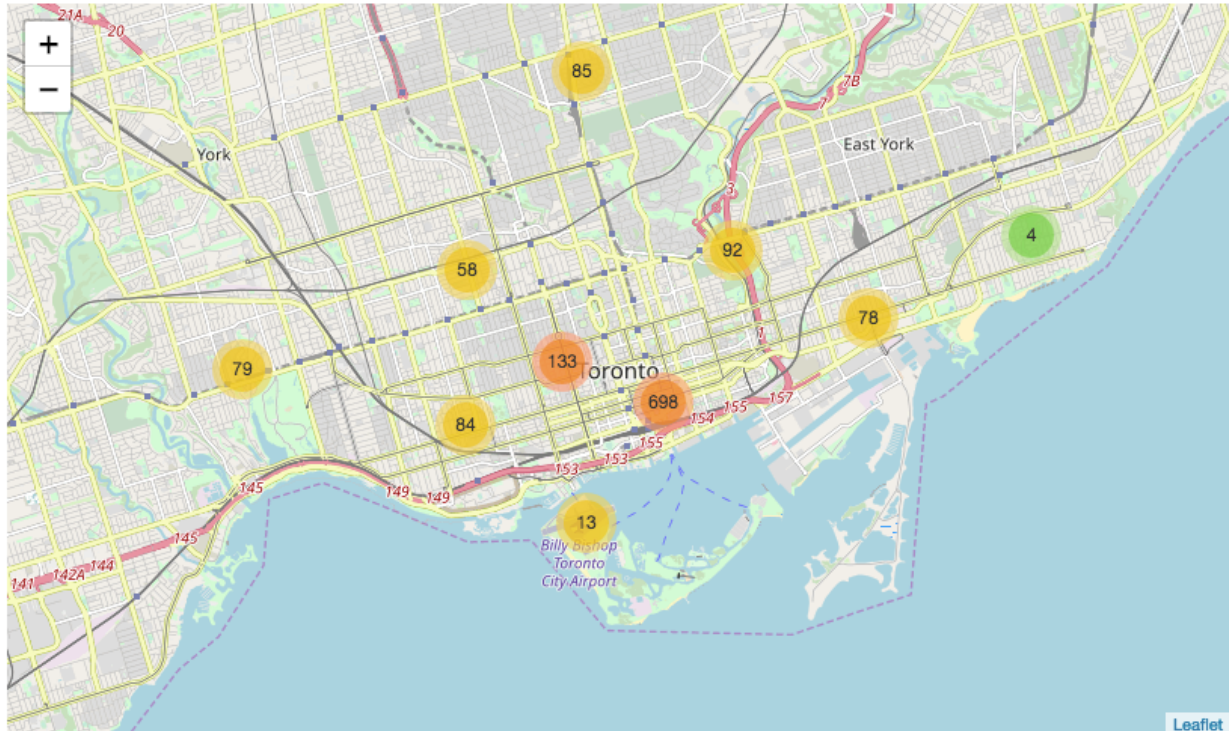
We next identified neighborhoods to have met the conditions. Finally, we calculate the total number of Sushi places in those neighborhoods.

It is reasonable to choose the one with the minimum number as the suggested neighborhood, but as we can see we have five neighborhood that are equally plausible choice for opening Sushi place. At this point it is also an option to provide the five options and see if the owners have any preference.



Map 1: Berczy Park Neighborhood

However as an alternative we further refine our choice by addressing the question that which of these will be able to cover the most number of neighborhoods, i.e proximity to other neighborhoods. When we cluster our dataset based on these criteria we have obtained the following result and with this we chose the best option. Our recommendation is thus Berczy Park with the coordinate location of [43.6447708, -79.3733064] and it belongs to the orange cluster group shown numbered 698 Map 2.



Map 2: Cluster of neighborhoods

DISCUSSION

From the analysis above, if one can conclude, solely, by considering the locations, Berczy Park should be a good start to explore as a candidate for the site. On top of this, these neighborhood fulfills the top 5 venues that were seen to usually accompany Sushi restaurants' surrounding. Furthermore we see from the results that opening in such neighborhood avoids competitions, which in itself is an attractive feature.

This study is limited in its scope in that it has explored Boroughs in Toronto whose name included Toronto in it, for instance Downtown Toronto. Another shortage of the project is it has not taken into consideration the size of population in each neighborhood.

As a next stage it will be interesting to see what results we get if we include the size of population in each neighborhood.

CONCLUSION

In this project we managed to find the best neighborhood for a Sushi place. In order to identify the optimal location factors we took into consideration are: (a) learn from experience by studying the popular venues around existing Sushi places and (b) assess the existence, or lack thereof, Japanese restaurant/ Sushi restaurant/ Asian restaurant. We have omitted seafood restaurants as they solely are not specialized in Sushi.

One limitation of the study worth mentioning is that we haven't included the population density of the neighborhoods, as we did not have ready made dataset for population in each neighborhood. But as remedy we have further tried to cluster the neighborhoods' proximity to each other. The more neighborhood dense clusters our site belongs to the better location it is, as it will be easily accessible from those neighborhoods.

ACKNOWLEDGEMENT

I thank my instructors at Coursera IBM Data Science, I also thank course attendees who have patiently reviewed my peer assignments and projects.

REFERENCES

- [1] <https://en.wikipedia.org/wiki/Sushi>
- [2] https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
- [3] https://cocl.us/Geospatial_data
- [4] <https://foursquare.com/>