The battle of neighborhood

Investigate the promising location and type of cuisine of opening up a restaurant in Vancouver

(Croft, 2010; Brend, 2020)

Dongning Sun

IBM Data Science Capstone Project

04/05/2020

Table of Content

Executive Summary

This report is a part of the IBM Data Scientist Certificate Capstone Project.

The target reader of this report could be people who are new to Vancouver and interested in opening up a restaurant in the city, with no personal preference on the type or the location of the restaurant. (after COVID-19 dies down of course!)

The purpose of this project is to investigate the promising location as well as the type of cuisine of opening a restaurant in Vancouver. Due to certain limitations in the type and amount of data, as well as the scope of this project, such as unavailability of restaurant ratings, target readers' existing resources, the restaurant type in more depth and detail, the demographic information of neighborhood such as their ages, occupations, etc. The investigation and proposal will only serve as an initial reference, and the related personnel will need to investigate further for the final selection and business decision.

The methodology employed is mainly based on the distribution of different types of restaurants in Vancouver's neighborhood. The restaurant categories and geospatial information is retrieved by Four Square API. Neighborhood data of Vancouver is realized from the combination of wikipedia's 'List of postal codes of Canada: V' and a free web resource geonames.org providing corresponding geo coordinates.

It is assumed that the neighborhoods containing the most restaurant of the specific category will be the ideal location, and the category will be the ideal cuisine type.

Via optimized-k-numbered K-means clustering and ranking the popular restaurant type in certain neighborhood areas, it was proposed to consider Japanese cuisine in Richmond Central area. Despite the limitation stated above, the proposal provides consistent logic with existing Vancouver demographic condition, and encourages a more in-depth investigation for the restaurant location and type selection.

Introduction

Background information and problem statement

Vancouver is a seaport city in western Canada. It is the most populous city in the province of British Columbia, and the third-largest metropolitan area in Canada. The Greater Vancouver area had a population in 2,463,431 in 2016, with a population density of over 5.4 k people per square kilometer. It is the most ethnically diverse cities in Canada, with 50.6% of residents belonging to visible minority groups. (StatCan, 2016)

Vancouver has the highest population density, encouraging more business opportunities for service industry. Being located near the sea provides availability of fresh seafood, and the variety of population opens up the alternatives of various cuisine type for restaurant owners to choose from. In short, Vancouver is a promising city to open up a restaurant given its demographic information. However, more information is needed in order to open up a restaurant. Location is not only the "three" important things in real estate, but also for restaurant business. This report investigates the neighborhoods of Great Vancouver Area, and based on the restaurant information of Four Square, provide basic and simple analysis for the purpose of providing a promising location and cuisine type for any new-comer of Vancouver who is interested in opening up a restaurant in this city.

Data Description

Four square is a local search-and-discovery mobile app developed by Foursquare Labs Inc. Its developer version allows users to retrieve places (called venues in Foursquare terminologies) by providing authorization information given an exact location of latitude and longitude. Given its popularity in Canada, the Foursquare contains most running restaurants' information, and will be the main api used in this project for restaurant information retrieving.

Vancouver's neighborhood information is composed of two parts relevant for this project: names and point location of neighborhoods. Postal code is used as the link between neighborhood names and neighborhood geographical information (i.e. coordinates). Neighborhoods' list with their corresponding postal code is available at wiki list of postal codes of Canada (Wikipedia, N.D.),

and the postal codes with corresponding geographic coordinates is available at an open resource websites called geonames (geonames.org, 2011).

Methodology

The methodology is divided into three parts.

- Data Preparation: presents how the dataset was obtained, where the geographical data was obtained from wiki list of postal codes of Canada, and geo coordinates was obtained from the geonames.org; the restaurant (venue) information was retrieved by API calls made to Four Square;

- Data analysis:
  - Clustering: K-means clustering of different groups of venues was used after determining the best k;
  - Group by and Ranking: top 6 clusters' feature (location and cuisine type) were analyzed

- Results, Discussion, and Conclusion: based on the analysis of top 6 clusters, the suggestions of location and type of restaurant was made, with a discussion of limitations of existing data, assumptions, and further possible investigations

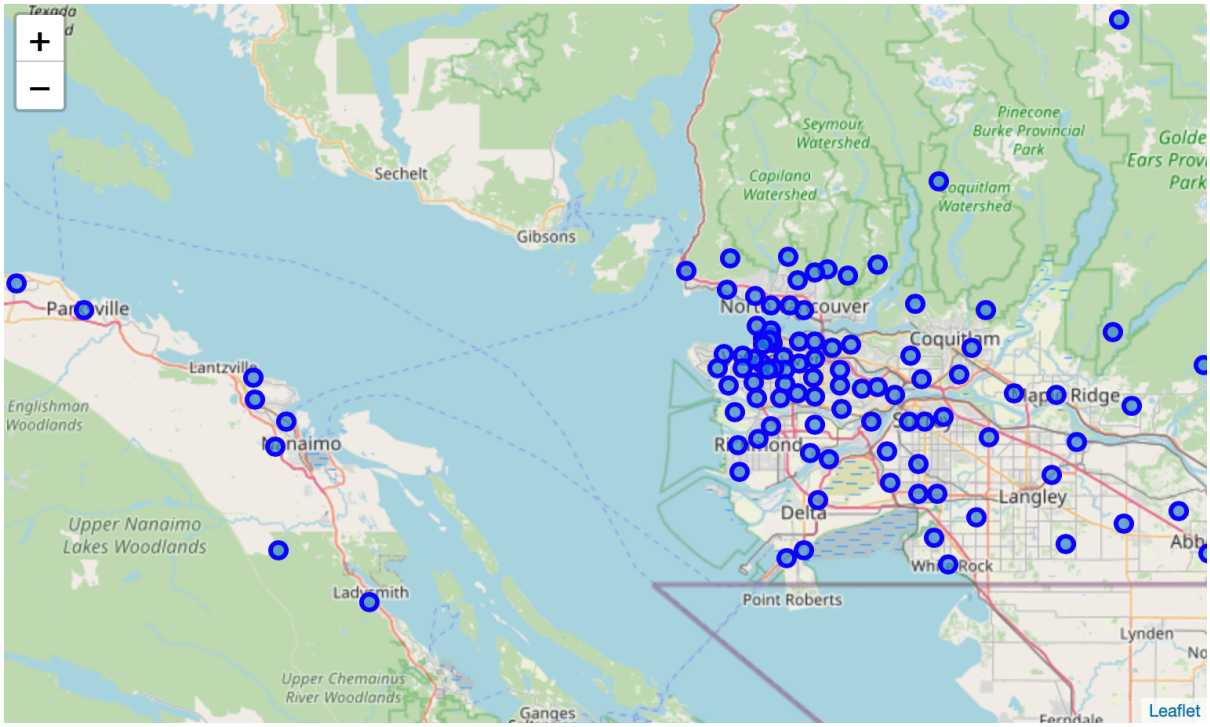
Results and Discussion

Neighborhood information

Processed and cleaned Vancouver Neighborhood with postal code, latitude, and longitude information. There are 165 unique postal codes area in Greater Vancouver area.

	Postal code	Neighborhood	Latitude	Longitude
0	V1B	VernonEast	50.084	-118.938
1	V2B	KamloopsNorthwest	50.8869	-120.736
2	V3B	Port CoquitlamCentral	49.274	-122.765
3	V4B	White Rock	49.0259	-122.806
4	V5B	Burnaby(Parkcrest-Aubrey / Ardingley-Sprott)	49.2769	-122.976

It is distributed as below:



Due to the limit of neighborhood population and demographic data, the distribution of neighborhood areas could only serve as a a basic understanding of potential areas to open up the restaurant, i.e. the main Vancouver city.

FourSquare API Call

By setting the radius of 1000m and limit of 20, a total of 4138 venues were retrieved of the Vancouver area. Among which, 1064 venues are restaurants.

	Neighborhood	Latitude	Longitude	Venue Name	Venue Latitude	Venue Longitude	Venue Category
1	Port CoquitlamCentral	49.2740	-122.7649	Pho Tan Phat	49.271024	-122.755339	Vietnamese Restaurant
2	Port CoquitlamCentral	49.2740	-122.7649	San Remo Pizza	49.267577	-122.773518	Italian Restaurant
27	White Rock	49.0259	-122.8058	Uli's Restaurant	49.021369	-122.805943	Restaurant
28	White Rock	49.0259	-122.8058	Taka's Sushi	49.023651	-122.800515	Japanese Restaurant
30	White Rock	49.0259	-122.8058	Kappa Japanese Restaurant	49.028604	-122.799631	Japanese Restaurant

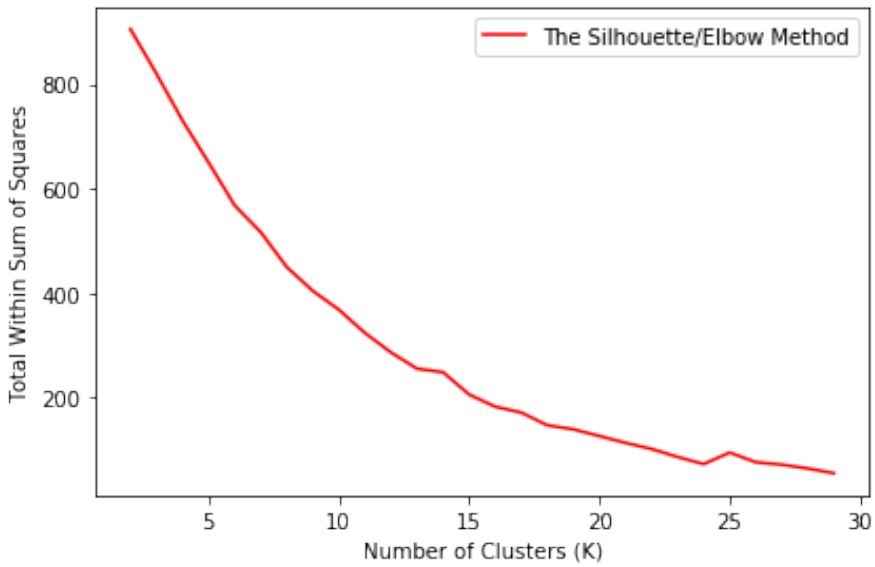
```
df1.shape
```

```
(1064, 8)
```

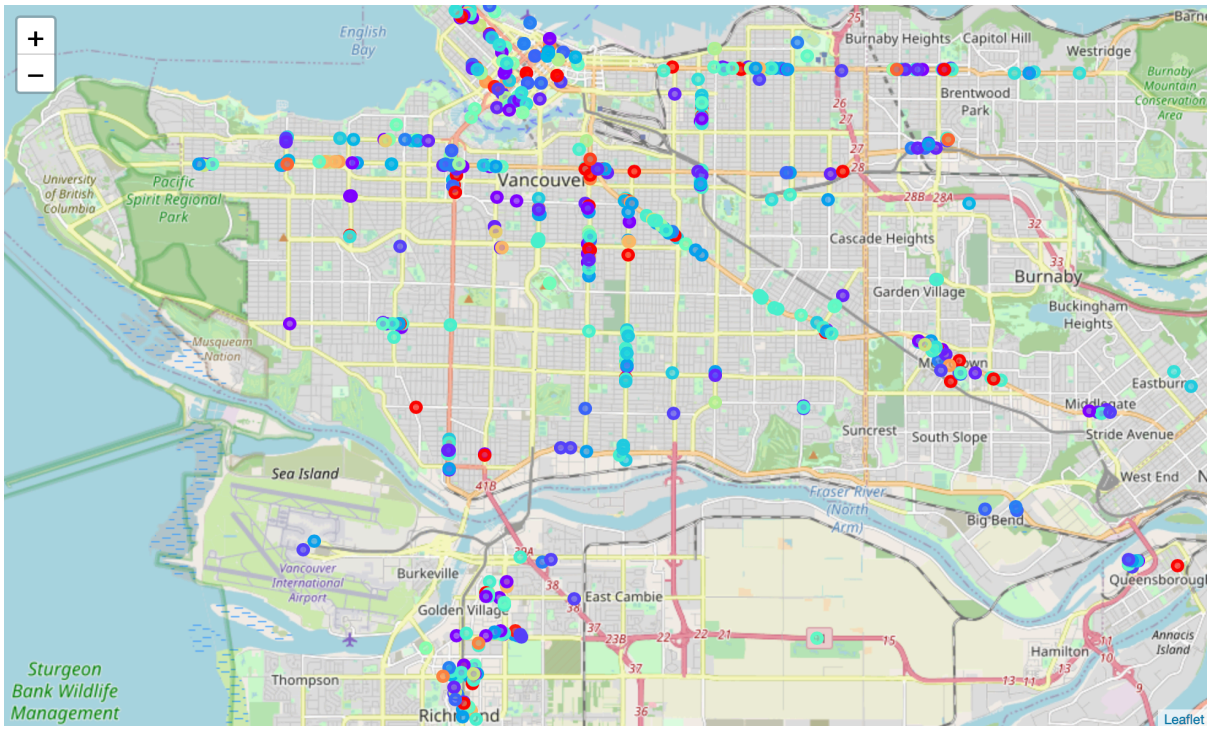
A limitation is the detail level of venue category. While some venues categories are clearly listed as “Vietnamese Restaurant” or “Japanese Restaurant”, Uli’s Restaurant only provides Category as “Restaurant”, adding obstacles to analyze the data.

Clustering of Neighborhood

K-Means machine learning algorithm was used for the clustering of the retrieved and parameterized restaurants’ information. By elbow method, an optimized number k was achieved as 24.



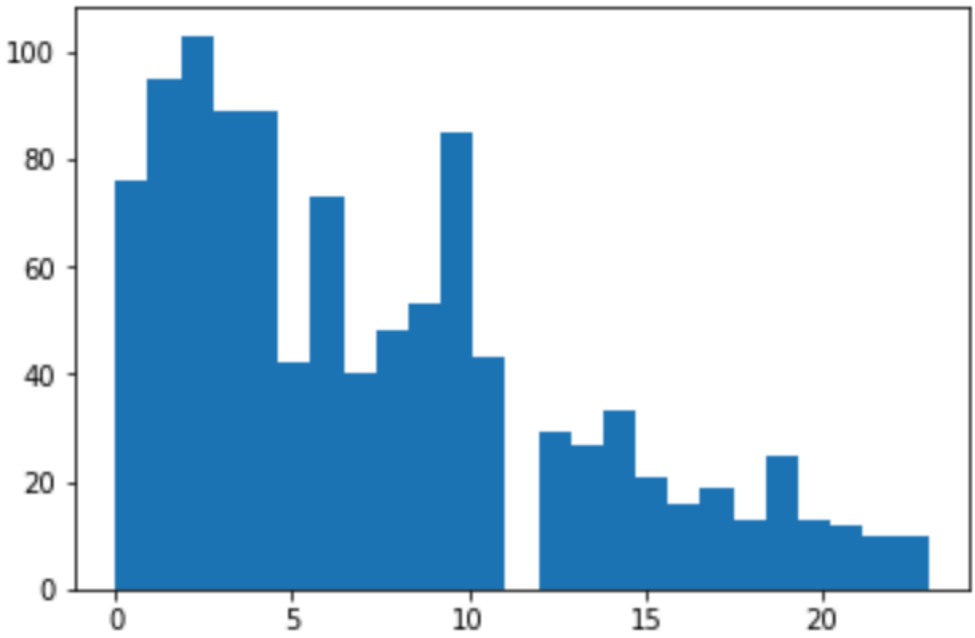
The correponding 24 clusters of restaurants are distributed in the Greater Vancouver Area as below:



The initial layout of venues suggest three locations where restaurants are most densely located: Vancouver Downtown, Vancouver Central, and Bournaby. However, the lack of demographic information in neighborhoods cannot identify whether the restaurant business is already saturated in the above-mentioned areas; nor can it suggest business opportunities in neighborhoods with less restaurants. This report assumes the more a certain type of restaurant in certain area, the higher rate of return for opening up the specific restaurant in this area.

Evaluating the clusters

Clustering, being an unsupervised learning algorithm, can only make sense after a closer look at the clustered data features. Below is the histogram of cluster number with the amount of each bins.



A further investigation of the bins with most venues number are cluster 0, 1, 2, 3, 4, 10 with a sum of 537 venues included. The total number of restaurant venues is 1064, 537 > (1064/2=532) venues clusters. The inclusion of investigated clusters is reasonable.

```
(1.84, 2.76]      103
(0.92, 1.84]       95
(2.76, 3.68]       89
(3.68, 4.6]        89
(9.2, 10.12]       85
(-0.024, 0.92]     76
```

The six clusters and its corresponding venue categories are listed as below

```
1 Japanese          103
2 Sushi              95
3 Fast Food          89
4 Restaurant         89
10 Chinese           85
0 Various type       76
```

This is where the problem of "Restaurant" as venue categories causing trouble. Nevertheless, based on the assumption, it is seen that the most popular food is Japanese (including sushi) of a total 198 restaurants available, fast-food comes next with a total of 89 restaurants, and Chinese is also popular, with 85 restaurant.

The three cuisine types are further investigated in order to see their distribution and relationship with neighborhood. Fast-food restaurant is distributed quite sparsely, and the hot number of Chinese and Japanese restaurant and their belonging neighborhoods are listed as below:

	Neighborhood	No. of Jap Restaurant
0	RichmondCentral	8
1	Vancouver(SE West End / Davie Village)	8
2	Burnaby(Burnaby Heights / Willingdon Heights /...	8
3	Vancouver(West Kitsilano / West Point Grey / J...	7
4	Vancouver(North Grandview-Woodland)	7
5	Vancouver(SW Downtown)	7
6	Vancouver(West Fairview / Granville Island / N...	7
7	Vancouver(West Mount Pleasant / West Riley Par...	7

double click to hide

	Neighborhood	No. of Chinese Restaurant
0	RichmondCentral	9
1	Vancouver(South Renfrew-Collingwood)	8
2	Vancouver(SE Riley Park-Little Mountain / SW K...	7
3	RichmondNorth	7
4	Vancouver(West Kensington-Cedar Cottage / NE R...	6

Conclusion

Despite the limitations in detailing of venue categories and neighborhood demographic information, and the inherent uncertainty of assumption, it is seen that Chinese restaurant in Richmond central might be the most promising location, however, the proposal shall consider future possibility of relocation or expansion, moreover, there is only one restaurant deviation comparing Chinese with Japanese. It is noted Japanese restaurant is the most popular cuisine type in Toronto, thus, a sound proposal is to open up a Japanese restaurant in Richmond Central.

Reference

Brend, Y. (2020). Restaurant Eye 'Soft' Reopening and Warn another Hard Close could Shut Doors for Good. CBC News. Published on 21.04.2020, available at <https://www.cbc.ca/news/canada/british-columbia/restaurant-reopen-covid-vancouver-economy-1.5539693>. Date Retrieved: 02/05.2020.

Crofit, E. (2010). *Best Food in Vancouver*. Available at <http://www.ellecroft.com/blog/2010/04/best-of-vancouver-food-and-drink/> Date retrieved: 02.05.2020.

Wikipedia, (N.D.) *List of postal codes of Canada: V*. Available at <https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_V> Date retrieved: 02.05.2020.

Geonames (2011) *Zip Code List.* Available at <(geonames.org> Date retrieved: 02.05.2020