

# The Battle of the Cities - Great Vancouver, Canada

## IBM Data Science Capstone project by Kyung-A Chang

### 1. Introduction : Business Problem

In this project we will try to find an optimal location for a **Korean restaurant** in the **Greater Vancouver**, British Columbia, Canada.

Vancouver is located in the west coast Canada and has the highest population density in Canada, with over 5,400 people per square kilometre which makes it the fifth-most densely populated city with over 250,000 residents in North America. The population of Vancouver and Greater Vancouver is about **2.5 million**. Vancouver is one of the most ethnically and linguistically diverse cities in Canada. 52% of its residents are not native English speakers, 48.9% are native speakers of either English or French, and 50.6% of residents belong to visible minority groups. The Chinese are the largest visible ethnic group in the city(26.5%). Other significant Asian ethnic groups in Vancouver are South and West Asian (7.4%), Filipino (5.9%), Japanese (1.7%), Korean (1.5%), as well as sizeable communities of Vietnamese, Indonesians, and Cambodians.

Being a high populated and ethnically diversified city, Vancouver offers multicultural restaurants in many neighbourhoods. To reflect the high Asian population, you can easily find the Asian restaurants in most of neighbourhoods.

Since there are lots of restaurants in Vancouver we will try to select locations that **are already crowded with existing Asian restaurants but not too many Korean restaurants in the neighbourhood**.

This report will be targeted for any **clients who are new to the city and have interests to set up a new Korean restaurant** or any **clients who already own a restaurant but search for a new optimal location**.

We will use our data science technique to generate a few most promising neighbourhoods based on this criteria. Advantages of each area will then be clearly expressed so that best possible final location can be chosen by stakeholders.

### 2. Data Understanding

Based on our Business problem, factors that will influence our decision are:

- number of existing Asian restaurants in the neighbourhood
- number of and distance to Korean restaurants in the neighbourhood, if any

Greater Vancouver is divided by many sub cities. There are over 1000 Asian restaurants in the greater Vancouver area. With Foursquare API's restriction returning top 100, we will loop thru the below coordinates using Foursquare API to extract the full list of Asian restaurants.

- latitude : 49.00N ~ 49.36N
- longitude : -123.18W ~ -122.7W
- step : 0.05

## Google Maps API reverse geocoding

After find the centre of each clusters, the city name will be obtained using the coordinates of the centre points.

## Foursquare : Venues Data

Foursquare has venue data with categories and geo locations for food and type of restaurants. Data can be searched by area and retrieved in GeoJSON format. The Foursquare data is crowdsourced, so accuracy is not guaranteed and may have duplicate entries for the same venue and incorrect category assignments.

<https://api.foursquare.com/v2/venues/explore?>

- categoryId = {"Asian Restaurant" : 4bf58dd8d48988d142941735}
- intent = {indicating your intent in performing the search : browse}
- sw = { limits results to the bounding box by south-west corner: 49.00, -123.18}
- ne = { limits results to the bounding box by north-east corner : 49.36, -122.7}

After initial evaluation of results from Foursquare API before data cleaning, we have 42 different types and 1061 Asian restaurants.

Sushi Restaurant	226
Chinese Restaurant	190
Japanese Restaurant	145
Vietnamese Restaurant	127
Asian Restaurant	112
Korean Restaurant	80
Thai Restaurant	58
Noodle House	26
Ramen Restaurant	22
Malay Restaurant	17
Dim Sum Restaurant	16
Taiwanese Restaurant	10
Cantonese Restaurant	9
Szechuan Restaurant	6
Filipino Restaurant	5
Shanghai Restaurant	5
Coffee Shop	5
Hotpot Restaurant	4
Vegetarian / Vegan Restaurant	4

Food Court	4	
Bakery	2	
Café	2	
Food Truck	2	
Seafood Restaurant		2
BBQ Joint	2	
Bubble Tea Shop		2
Shabu-Shabu Restaurant		2
Burger Joint	1	
Taco Place	1	
Fast Food Restaurant		1
Donut Shop	1	
Dumpling Restaurant		1
Diner	1	
Hot Dog Joint	1	
Indian Chinese Restaurant		1
Tea Room	1	
Middle Eastern Restaurant		1
Lounge	1	
Ice Cream Shop		1
Restaurant	1	
Japanese Curry Restaurant		1
Dessert Shop	1	

This is sufficient number of restaurants to group compare restaurant density thru the greater Vancouver area.

### 3. Methodology:

In this project, I focus on detecting city/neighbourhood in the Greater Vancouver that have high density of Asian restaurants but low number of Korean restaurants.

- I have to figure out what Foursquare API option I use for this project. As Foursquare API limits the number of returning query 100 for explore option and 50 for search option, I try with the explore option. After run a few initial Foursquare queries, my result sets are over 100 for each cities. The result sets were reaching 100 every time. Then I discover that bounding box option can be used to narrow down the search query if I select the smaller area. The bounding box is required to have two points, SW and NE, then it pull the venue results within that boundary.
- As I am using Foursquare API bounding box option, I need to figure out the correct coordinates that cover the full Greater Vancouver region. Since Greater Vancouver does not have defined coordinates, I had to go thru the all the cities in the boundary to find the right coordinate using google map.
- After collect the required data, I need to check for the duplicated entries since I never used bounding box options before.

- Foursquare data is crowded data so the venue categories were not very consistently populated. I have to merge, join, clean the data before using it. Also I had to search some of restaurant's name to determine the correct category for ambiguous entries.
- I use K-Means Clustering algorithm to cluster the restaurants by coordinates. K-Means algorithm is one of the most common cluster method of unsupervised learning. Latitude and Longitude are used as the features. First I find the optimal K with K-Means with elbow method.
- Finally I focus on the most promising area where high populated with Asian restaurants but lower number of Korean restaurant to meet my project's requirement.

#### 4. Result and Discussion:

- It's worth noting that Foursquare API only returned 25 Korean restaurants without specifying "intent=browse". The default is "intent=checkin" which does not return restaurants if you are outside of Foursquare defined check-in area
- This analysis shows that although there is over 1000 Asian restaurants in the Greater Vancouver area, **Vancouver** city offers the best location for the new Korean restaurant venue. Burnaby and Coquitlam shows very similar number of total Asian restaurants but Coquitlam has 3.4 times more Korean restaurants than Burnaby.

Burnaby	Chinese Restaurant	95
	Japanese Restaurant	76
	Vietnamese Restaurant	38
	Asian Restaurant	26
	Korean Restaurant	14
	Thai Restaurant	10
	Malay Restaurant	3
	Noodle House	3
Coquitlam	Japanese Restaurant	90
	Chinese Restaurant	51
	Korean Restaurant	48
	Vietnamese Restaurant	32
	Asian Restaurant	20
	Thai Restaurant	19
	Malay Restaurant	3
	Noodle House	2

- I used the K-Means algorithm as part of this cluster studying. When I tested the K-Means with Elbow method, the optimal K can be 6 to 10. After I tested with different value, the results were not too different. Therefore I selected 6 as my optimal K.

- I ended up the study by visualizing the Vancouver clustering information on the Greater Vancouver map. For the future study, the distance between Korean restaurants/Asian restaurants can be calculated and considered as another feature.

## **5. Conclusion**

Purpose of this project was to identify an optimal city/neighbourhood in the Greater Vancouver region for a new Korean restaurant. By calculating restaurant density distribution from Foursquare data, we narrowed down the candidates into 6 cities. Then further study of the clustering data, we selected the final location.

Final decision on optimal restaurant location will be made by stakeholders based on specific characteristics of neighbourhoods and locations in every recommended zone, taking into consideration additional factors like attractiveness of each location (proximity to park or water), levels of noise / proximity to major roads, real estate availability, prices, social and economic dynamics of every neighbourhood etc.