Finding Locations with Similar Cuisines Nearby

Dakota Chang

July 25, 2021

1. Introduction

1.1 Background

Canada accepts 300,000 immigrants every year, with the British Columbia province accepting a whopping 15% of them. With Vancouver boasting one of the most diverse arrays of cuisines and plenty of recreational venues, there are a plethora of choices for the average person moving to BC. In the event of moving to another country, people would inevitably miss foods from their home country. However, as can be expected, when moving to an unfamiliar place, most of them would not know where to settle down and find food that reminds them of home. In order to promote more consumer activity and help the new citizens feel accepted, it is vital to provide a comprehensive guide to the cuisine distribution of the BC province. The map created can guide the new inhabitants to explore the BC province and promote local business all in one. It can also be used for deciding where to open a restaurant and which cuisine is the most popular or over-saturated in the area.

1.2 Problem

Data that might contribute to the construction of this map would include the number and category of food-related venues in the BC province. It is also vital to use the postal codes of each area to determine the proximity of each cluster point, as the postal codes would be reflective of the population nearby and act as a marker to different areas. This project aims to cluster locations based on the selection of cuisines near them.

1.3 Interest

This map can potentially gain the interests of business owners looking for an analysis of a location’s cuisine profile, citizens looking to explore the province further, and tourists looking into experiencing the local culture and support local restaurants.

2. Data acquisition and cleaning

2.1 Data Sources

Most postal codes, latitudes, and longitudes of locations in the BC province can be scraped off the geonames.org database. This data set provides the name of the location, the postal code, country, one or two admins of the province, and latitudes and longitudes of each location. Information about nearby venues was also sourced from Foursquare using a personal account. The information provided by Foursquare provided information of venues of considerable proximity, the distance of each of the provided venues, and the category of each venue (i.e., lake, park, bakery, Chinese restaurant, and more).

2.2 Data Cleaning

Due to the nature of the data frame on the geonames website and the variability of the number of admins for each postal area, it was particularly hard to process each postal code's data frame and location names. However, since we are in the BC section of Canada, it is more accurate to only take into account the postal codes and latitudes and longitudes of each postal area. While the inclusion of names of the site could be helpful, most of the names were relatively generic and repetitive. Therefore, the inclusion of postal codes only would operate for our current needs. The postal codes were used as identifiers of the unique locations in this case, with the latitudes and longitudes saved as floats and acting as a marker for the Foursquare API and map generation.

In the Foursquare API, I have found a lack of information on restaurants in Vancouver, and that I must expand the search radius to reflect the cuisines surrounding each location. I also

set the scrape limit to 200 venues to include as many food-related venues as possible without overwhelming the map and clusters. There were also many discrepancies for each venue's category names, meaning that I must clean up the data and filter out the non-food-related venues such as lakes and recreational parks.

After fixing these problems, I checked for outliers and duplicates and removed them from the dataset. I also removed any locations with no available data on their nearby venues.

2.3 Feature Selection

After data cleaning, there were still many venue categories left (fig. 1a) as there was information of 1795 venues collected. Upon examining these labels, only food or drink-related venues were kept. All venues not relating to consumable items being served in real-time were eliminated (fig. 1b), and after assigning these venues to each location, locations with no food-related stores nearby were eliminated and removed from the data frame. The act of filtering removed 95 locations, resulting in a dataset of 97 locations and 102 one-hot encoded values of venues nearby.

['Lake', 'Pub', 'Brewery', 'Coffee Shop', 'Gym / Fitness Center', 'Convenience Store', 'Vietnamese Restaurant', 'Sandwich Place', 'Gym', 'Grocery Store', 'Japanese Restaurant', 'Bistro', 'Elementary School', 'Waterfront', 'Café', 'Seafood Restaurant', 'Greek Restaurant', 'Gastropub', 'Health Food Store', 'Gas Station', 'Steakhouse', 'American Restaurant', 'Thai Restaurant', 'Mobility Store', 'Museum', 'Ice Cream Shop', 'Dessert Shop', 'Breakfast Spot', 'Restaurant', 'Pizza Place', 'Fast Food Restaurant', 'Pharmacy', 'Supermarket', 'Burger Joint', 'Factory', 'Liquor Store', 'Diner', 'Farm', 'German Restaurant', 'Plaza', 'Bus Line', 'Business Service', 'Construction & Landscaping', 'Campground', 'Sushi Restaurant', 'Inn', 'Thrift / Vintage Store', 'Baseball Field', 'Hotel', 'Motel', 'Moving Target', 'Shop & Service', 'Playground', 'Smoothie Shop', 'Food Truck', 'Bakery', 'Park', 'Hockey Arena', 'Food & Drink Shop', 'Flower Shop', 'Golf Course', 'Bank', 'Clothing Store', 'Mountain', 'Bus Stop', 'Movie Theater', 'Lounge', 'Furniture / Home Store', 'Bowling Alley', 'Discount Store', 'Department Store', 'Comfort Food Restaurant', 'Tea Room', 'Deli / Bodega', 'Women's Store', 'Bar', 'Sports Club', 'Sports Bar', 'Skate Park', 'Dog Run', 'Pool', 'Pet Store', 'Art Gallery', 'Indian Restaurant', 'Falafel Restaurant', 'Chinese Restaurant', 'Malay Restaurant', 'Donut Shop', 'Bookstore', 'Food Court', 'Sporting Goods Shop', 'Library', 'Paper / Office Supplies Store', 'Filipino Restaurant', 'Music Store', 'Bubble Tea Shop', 'Toy / Game Store', 'Big Box Store', 'Hardware Store', 'Asian Restaurant', 'Dim Sum Restaurant', 'Shopping Mall', 'Light Rail Station', 'Bus Station', 'Korean Restaurant', 'Electronics Store', 'Antique Shop', 'Burrito Place', 'Candy Store', 'Market', 'City', 'Home Service', 'Skating Rink', 'Memorial Site', 'Noodle House', 'Middle Eastern Restaurant', 'Fish Market', 'Motorcycle Shop', 'Post Office', 'Fried Chicken Joint', 'Fish & Chips Shop', 'Farmers Market', 'Juice Bar', 'Salon / Barbershop', 'Outdoor Sculpture', 'Ethiopian Restaurant', 'Music Venue', 'Cocktail Bar', 'Field', 'Harbor / Marina', 'Cantonese Restaurant', 'Building', 'Cafeteria', 'Financial or Legal Service', 'Taco Place', 'Spa', 'Cosmetics Shop', 'Hotel Bar', 'Yoga Studio', 'Leather Goods Store', 'Scenic Lookout', 'Australian Restaurant', 'Mediterranean Restaurant', 'New American Restaurant', 'Chocolate Shop', 'Irish Pub', 'Poke Place', 'Hot Dog Joint', 'Men's Store', 'Airport Terminal', 'Train Station', 'Vegetarian / Vegan Restaurant', 'Boat or Ferry', 'Airport', 'Cruise Ship', 'Office', 'French Restaurant', 'Gay Bar', 'Italian Restaurant', 'South Indian Restaurant', 'Ramen Restaurant', 'Miscellaneous Shop', 'Wine Shop', 'Mexican Restaurant', 'Hawaiian Restaurant', 'Nightclub', 'Himalayan Restaurant', 'Hostel', 'Garden', 'Trail', 'Historic Site', 'Physical Therapist', 'Tennis Court', 'Hotpot Restaurant', 'Frozen Yogurt Shop', 'Bagel Shop', 'Southern / Soul Food Restaurant', 'Caribbean Restaurant', 'Massage Studio', 'Gift Shop', 'Dumpling Restaurant', 'Taiwanese Restaurant', 'Flea Market', 'Shoe Store', 'Portuguese Restaurant', 'Lebanese Restaurant','Indie Movie Theater', 'Beach', 'Airport Lounge', 'Rental Car Location', 'Airport Service', 'Travel Lounge', 'Public Art', 'Duty-free Shop', 'Wine Bar', 'Shopping Plaza', 'Rock Climbing Spot', 'Butcher', 'Resort', 'History Museum', 'Gaming Cafe', 'Tapas Restaurant', 'Comic Shop', 'Turkish Restaurant', 'BBQ Joint', 'Modern European Restaurant', 'Record Shop', 'Indonesian Restaurant', 'Theater', 'Cajun / Creole Restaurant', 'Ski Chalet', 'Ski Trail', 'Ski Chairlift', 'Concert Hall', 'Mobile Phone Shop', 'Japanese Curry Restaurant', 'Lingerie Store', 'Jewelry Store', 'Dance Studio', 'Optical Shop', 'Performing Arts Venue', 'Moroccan Restaurant', 'Board Shop', 'Tourist Information Center', 'Apres Ski Bar', 'Mongolian Restaurant', 'Ski Area', 'North Indian Restaurant', 'Casino', 'Eye Doctor', 'Church', 'Auto Garage', 'Video Game Store', 'Distribution Center', 'Salad Place', 'Garden Center', 'Fair', 'Theme Park Ride / Attraction', 'Theme Park', 'Soccer Field', 'Beer Garden', 'Amphitheater', 'Event Space', 'Stadium', 'Arts & Crafts Store', 'Athletics & Sports', 'Event Service', 'Hobby Shop', 'Metro Station', 'Storage Facility', 'Child Care Service', 'Lawyer', 'Bed & Breakfast', 'Mattress Store', 'Auto Workshop', 'Arts & Entertainment', 'Public Bathroom', 'Marijuana Dispensary', 'Food Service', 'Kids Store', 'Pie Shop', 'Cheese Shop', 'Belgian Restaurant', 'Trade School', 'Warehouse Store', 'Hookah Bar','Golf Driving Range']

Fig. 1a: All original categories of venues.

```
food_venues = ['Pub', 'Brewery', 'Coffee Shop', 'Vietnamese Restaurant', 'Sandwich Place', 'Japanese Restaurant',
'Bistro', 'Seafood Restaurant', 'Greek Restaurant', 'Gastropub', 'Steakhouse', 'American Restaurant', 'Thai Restaurant', 'Ice Cream Shop',
'Dessert Shop', 'Breakfast Spot', 'Pizza Place',
'Fast Food Restaurant', 'Burger Joint', 'Liquor Store', 'Diner', 'German Restaurant',
'Sushi Restaurant', 'Smoothie Shop', 'Food Truck', 'Bakery',
'Food & Drink Shop','Comfort Food Restaurant',
'Tea Room', 'Deli / Bodega', 'Bar', 'Sports Bar',
'Indian Restaurant', 'Falafel Restaurant', 'Chinese Restaurant', 'Malay Restaurant',
'Donut Shop', 'Filipino Restaurant', 'Bubble Tea Shop',
'Asian Restaurant', 'Dim Sum Restaurant', 'Korean Restaurant', 'Burrito Place', 'Candy Store',
'Noodle House', 'Middle Eastern Restaurant', 'Fish Market', 'Fried Chicken Joint',
'Fish & Chips Shop', 'Farmers Market', 'Juice Bar', 'Ethiopian Restaurant',
'Cocktail Bar', 'Cantonese Restaurant',
'Taco Place', 'Australian Restaurant', 'Mediterranean Restaurant', 'New American Restaurant',
'Chocolate Shop', 'Irish Pub', 'Poke Place', 'Hot Dog Joint',
'Vegetarian / Vegan Restaurant', 'French Restaurant', 'Gay Bar',
'Italian Restaurant', 'South Indian Restaurant',
'Ramen Restaurant', 'Wine Shop',
'Mexican Restaurant', 'Hawaiian Restaurant',
'Himalayan Restaurant', 'Hotpot Restaurant', 'Frozen Yogurt Shop', 'Bagel Shop',
'Southern / Soul Food Restaurant', 'Caribbean Restaurant', 'Dumpling Restaurant',
'Taiwanese Restaurant', 'Portuguese Restaurant', 'Lebanese Restaurant',
'Wine Bar', 'Gaming Cafe', 'Tapas Restaurant',
'Turkish Restaurant', 'BBQ Joint', 'Modern European Restaurant',
'Indonesian Restaurant', 'Cajun / Creole Restaurant',
'Japanese Curry Restaurant',  'Moroccan Restaurant',
'Apres Ski Bar', 'Mongolian Restaurant', 'North Indian Restaurant',
'Salad Place', 'Beer Garden', 'Food Service','Pie Shop', 'Cheese Shop', 'Belgian Restaurant','Hookah Bar']
```

Fig. 1b: All categories left after filtering.

3. Exploratory Data Analysis

3.1 Clustering of Nearby Venue Profiles

Based on the graphed data, the hypothesis here is that while each location might be in close proximity of each other, they would each have distinct make-up of food-related venue and not overlap each other and can be filtered into a wide range of clusters in order to reflect the quality of each location (fig. 4). The filtered data was also shown to reflect the population density and economic activity in the area (fig. 3). There are also certain locations with more diversity to its nearby venue make-up so that diversity will be taken into account for clustering data (fig. 2).

Postal Code	Postal Area Latitude	Postal Area Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
V0A	1	1	1	1	1	1
V1A	3	3	3	3	3	3
V1E	1	1	1	1	1	1
V1J	4	4	4	4	4	4
V1K	1	1	1	1	1	1
...	...	...	...	...	...	...
V9K	2	2	2	2	2	2
V9L	10	10	10	10	10	10
V9S	4	4	4	4	4	4
V9Y	3	3	3	3	3	3
V9Z	1	1	1	1	1	1

Fig. 2: Data frame of postal codes and information about them.

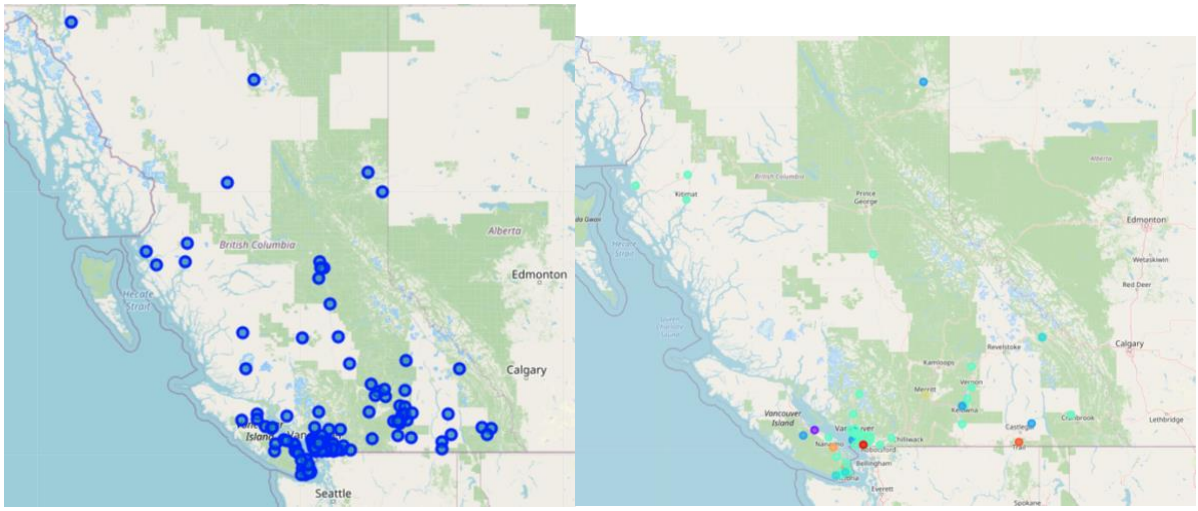

Fig. 3: Map before and after filtering of venues and locations.

	Postal Code	1V Most Common Venue	2th Most Common Venue	3th Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	V0A	Bakery	Wine Shop	Fish & Chips Shop	Dessert Shop	Dim Sum Restaurant
1	V1A	American Restaurant	Fast Food Restaurant	German Restaurant	Asian Restaurant	Fish & Chips Shop
2	V1E	Modern European Restaurant	Wine Shop	Filipino Restaurant	Dessert Shop	Dim Sum Restaurant
3	V1J	Coffee Shop	Ice Cream Shop	Liquor Store	Filipino Restaurant	Dessert Shop
4	V1K	Sandwich Place	Wine Shop	Coffee Shop	Deli / Bodega	Dessert Shop

Fig. 4: Example of the venue profiles of each location.

4. Clustering Analysis

4.1 Choosing Machine Learning Model

I applied the KNN (k-nearest neighbor) algorithm to the collected data. The KNN is a supervised machine learning algorithm that is often used for solving regression or classification problems. In this case, it is used to classify each location based on the nearby venues to identify trends amongst the location.

4.2 Performance of KNN with Different Parameters

Initially, the number of clusters to be generated by the algorithm was set to be 5 clusters to maximize efficiency and navigability. However, upon inspection of the 5 clusters, the clusters were separated by generic categories such as ‘Restaurant’, ‘Coffee shop’, ‘Fast Food Chain’, skewing the data due to the sheer number of these venues. I iterated back to the data engineering step and further filtered it down. I also increased the cluster number to 10 to accommodate the wide variety of cuisine profiles in BC province.

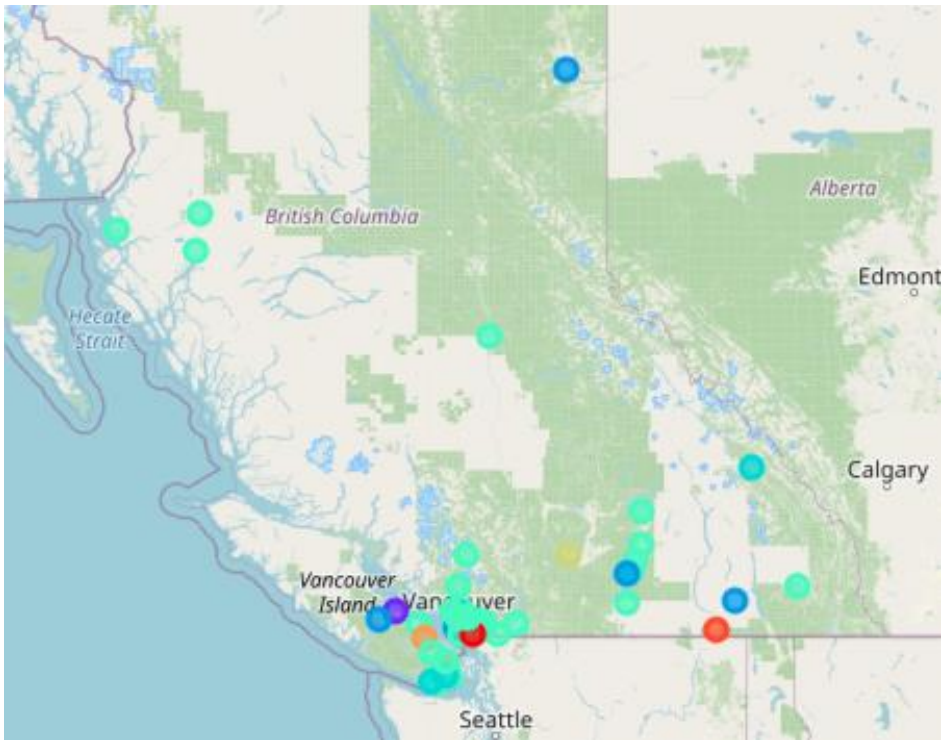

Fig. 5: Overall distribution of cuisine profiles in the BC province sorted based on color.

5. Conclusions

Upon inspection of some samples of each cluster, the main criteria of clustering were identified in figure 6. The data collected showed that many locations within the BC province share a similar cuisine profile, despite certain areas being more well known for a specific cuisine. It is also identified that many restaurants serving Asian food, and most locations in Vancouver have a wide variety of choices around the

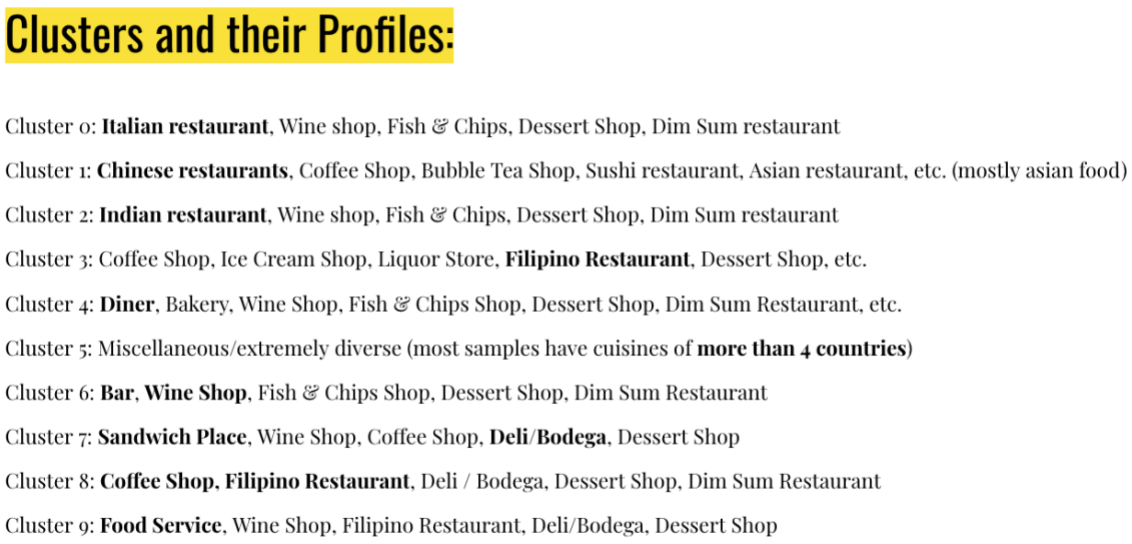

Figure 6: The pattern as found from each cluster, with the most commonly appearing venue bolded.

6. Future Directions

When using Foursquare, I realised many venues are not updated onto their map, and therefore missed a lot of potential nearby food venues. There is also room for improvement when filtering the food venues, and some restaurants can be re-categorized to lessen the specificity of each cluster.