Capstone Project: Battle of the Neighbourhoods

Dakota Chang

Introduction

Discussion of the business problem and the audience who would be interested in this project.

Description of the Problem and Background

Scenario/Target Audience:

My friend will be moving to Vancouver soon, but she is not quite familiar with the city. She is a lover of food and would like to learn about the neighbourhoods and distribution of international foods within them. Since Vancouver has a diverse population and is a popular choice for a lot of immigrants, this project will be helpful to a lot of other people who also love trying different foods and are moving to Vancouver.

Business Problem:

The challenge is to find locations with a similar profile of international cuisine in close proximity to it and label it so it can be easily comprehended.

Data Acquisition:

Description of the data and its sources that will be used to solve the problem

Description of the Data

Sources:

- Postal Codes and Long Lat of Locations in BC:
  - This website will be scraped and engineered into a dataframe of places in Vancouver.

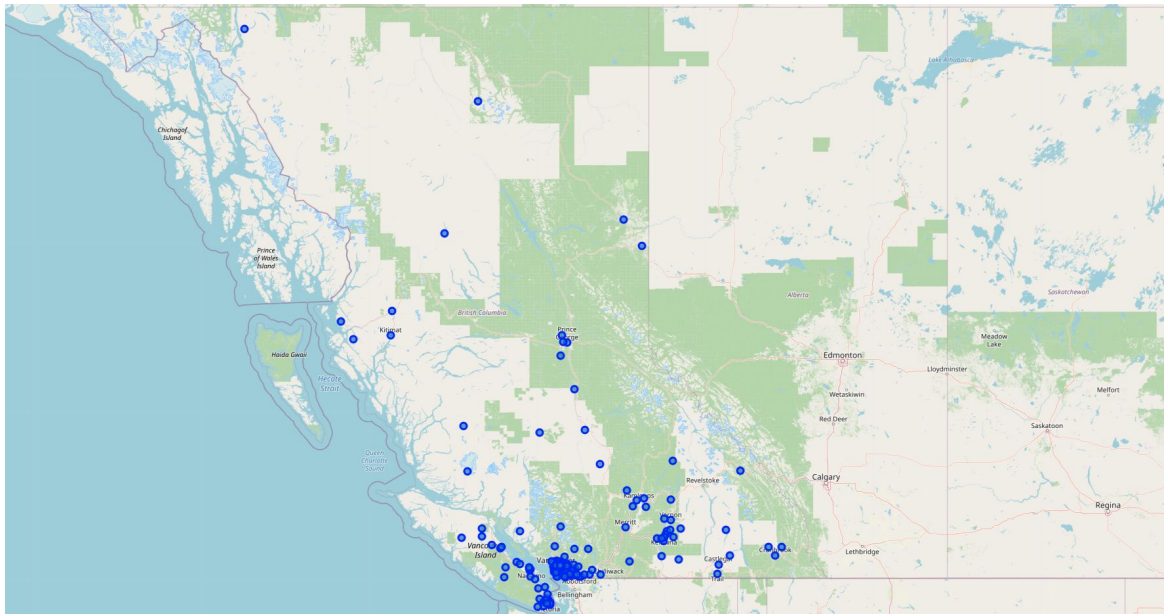
- Foursquare:
  - Foursquare will be scraped for the nearby venues of each location.

Data Engineering:

Steps of Data Cleaning and Engineering:

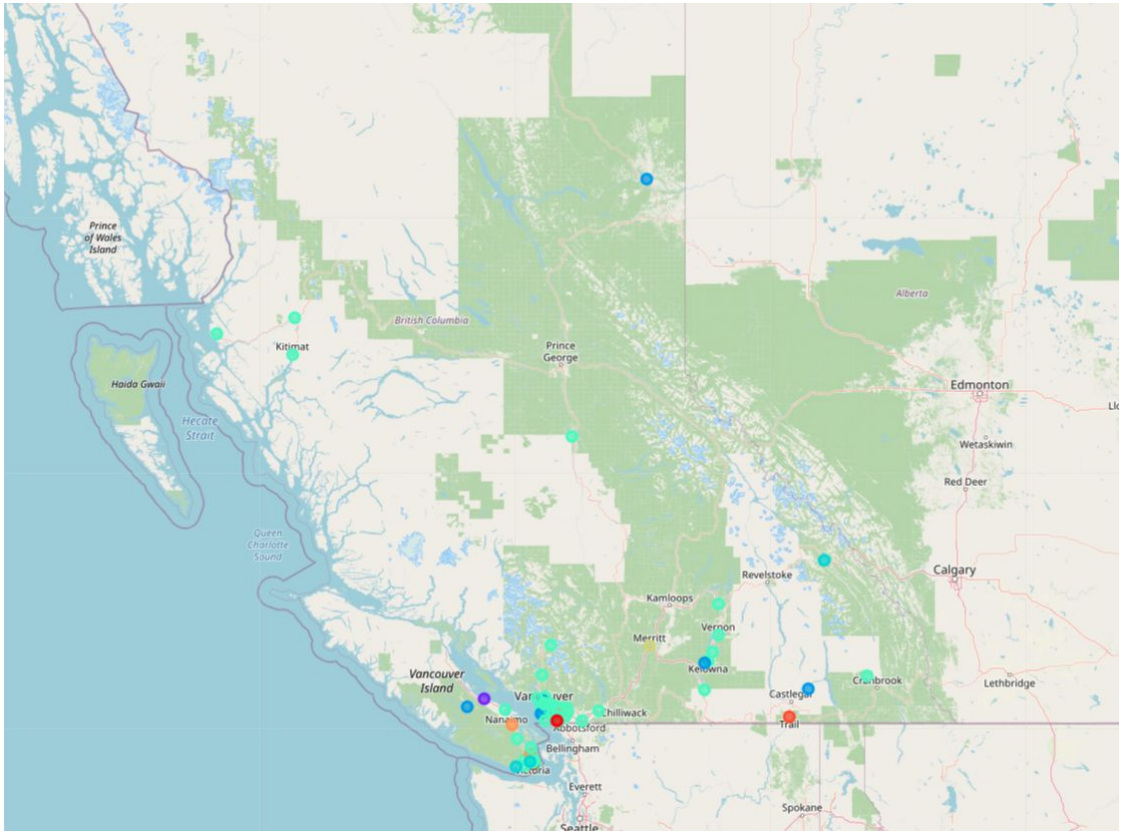
1. Scrape data from geonames.org in order to create a dataframe with postal codes, longitude and latitude of locations in BC province, Canada
2. Engineer the data such that the data frame would consist of the right postal codes, longitudes and latitudes.
3. Match up the longitudes and latitudes of the collected data and use the Foursquare API to find the nearest venues of all locations
4. Remove unrelated venues from the data frame (i.e. keep only restaurants/international cuisines)
5. Apply one-hot encoding to the data frame so it can be easily processed

Potential Locations for Analyzation



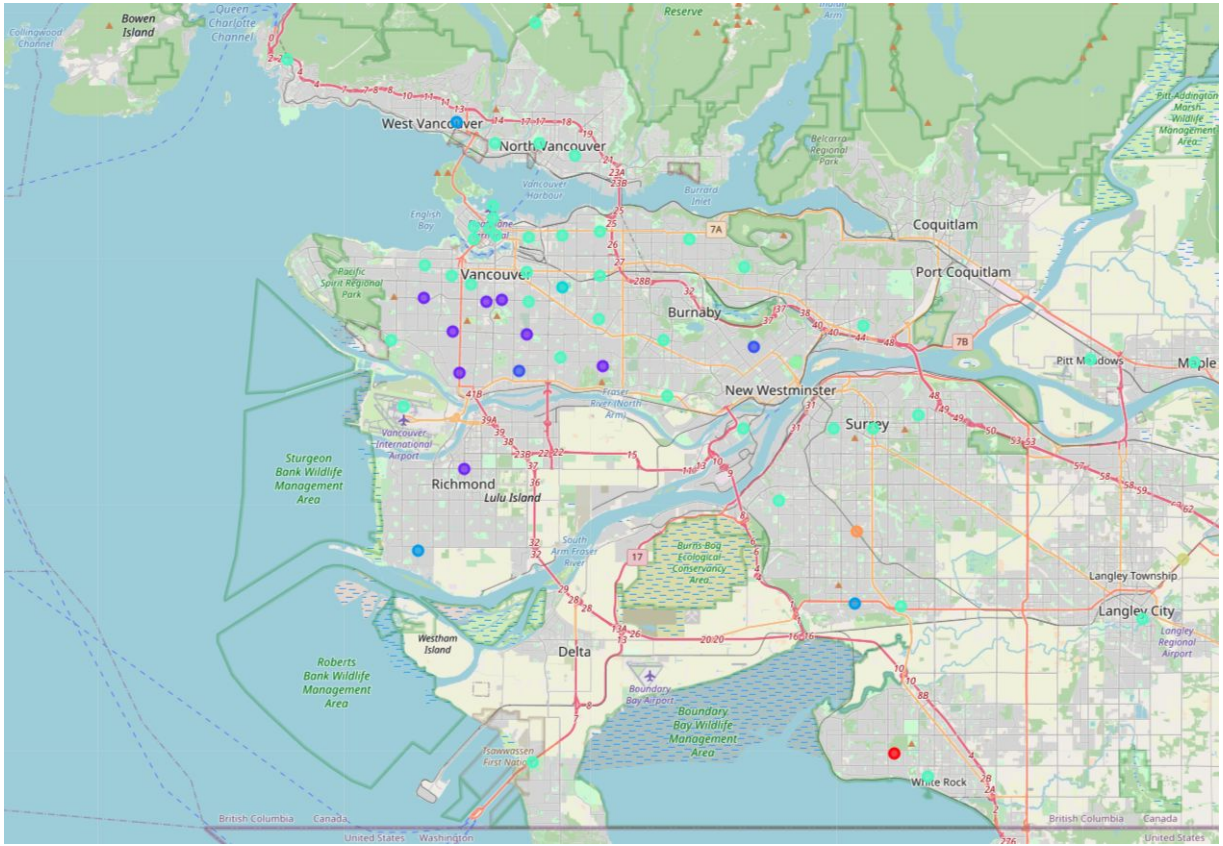
Blue dots represent potential locations for clustering. Locations are based on postal codes and their corresponding longitudes and latitudes.

Overall Map of Clustered Areas



Based on each location's nearby 5 venues, the machine learning algorithm (KNN) put the locations into 10 categories, each signifying a different demographic profile of food venues.

Zoomed in to Vancouver



A zoomed-in version for more straightforward observation. A lot of locations have similar profiles due to the demographic of people living in each area. For example, the clusters of purple dots are due to a similar profile of Chinese and Asian(overall) food venues surrounding it. This area is also heavily populated by Asian people, so it tracks the decision to cluster these together.

A More Detailed Look

The map is interactive, with the labels showing their corresponding cluster and postal code.

Clusters and their Profiles:

Cluster 0: **Italian restaurant**, Wine shop, Fish & Chips, Dessert Shop, Dim Sum restaurant

Cluster 1: **Chinese restaurants**, Coffee Shop, Bubble Tea Shop, Sushi restaurant, Asian restaurant, etc. (mostly asian food)

Cluster 2: **Indian restaurant**, Wine shop, Fish & Chips, Dessert Shop, Dim Sum restaurant

Cluster 3: Coffee Shop, Ice Cream Shop, Liquor Store, **Filipino Restaurant**, Dessert Shop, etc.

Cluster 4: **Diner**, Bakery, Wine Shop, Fish & Chips Shop, Dessert Shop, Dim Sum Restaurant, etc.

Cluster 5: Miscellaneous/extremely diverse (most samples have cuisines of **more than 4 countries**)

Cluster 6: **Bar**, **Wine Shop**, Fish & Chips Shop, Dessert Shop, Dim Sum Restaurant

Cluster 7: **Sandwich Place**, Wine Shop, Coffee Shop, **Deli/Bodega**, Dessert Shop

Cluster 8: **Coffee Shop, Filipino Restaurant**, Deli / Bodega, Dessert Shop, Dim Sum Restaurant

Cluster 9: **Food Service**, Wine Shop, Filipino Restaurant, Deli/Bodega, Dessert Shop

Conclusions and Future Directions

Conclusion:

The data collected showed that many locations within BC province share a similar cuisine profile, despite certain areas being more well known for a specific cuisine. The map has also clustered these locations and allowed users to explore these streets based on their personal preferences.

Improvements:

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.