Capstone Project Report

- The Battle of Neighborhoods

Yu Zhao

Introduction/Business Problem

The goal of this project is to help BC residents find neighborhoods that are convenient to live without having a car. Car insurance in BC is very expensive, which makes it challenging for many BC residents to own a car. It's also challenging for people to rely on public transit in BC because of the long wait for buses and the limited coverage of Skytrain. A convenient neighborhood to live in without a car needs to have enough venues within walking distance. We can get venue data near a neighborhood by using FourSquare's API.

Target audience

The target audiences of this study are

- Current or future residents in BC who don't have a car: the result of this study can help them choose a convenient area to purchase or rent a property
- Realtors who help clients without a car: this study can help inform realtors to make better suggestions to their clients who enjoy the lifestyle without a car
- Residential building developers who want to develop transit-oriented residential buildings: there are many residential buildings target at buyers who are transit-oriented in

metro-Vancouver. The model and outcome of this study can help residential building developers choose better locations for their transit-oriented projects.

Data

To solve this problem, I will need below data:

- List of neighborhoods in Metro Vancouver, BC, Canada:
    - The data can be obtained from wikipedia page: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_V
    - This page contains the city/neighborhood name, with the postal code. A city/neighborhood can have multiple postal codes.
- Latitude and Longitude of these neighborhoods
    - This data can be obtained by using the postal code to fetch latitude and longitude data from https://geocoder.ca/ API. The latitude and longitude information for each neighbourhood can be used to retrieve nearby places information from FourSquare API
- Venue data near neighborhoods from FourSquare
    - FourSquare's place API can return nearby places for a pair of latitude and longitude data. The places information they return has a list of attributes, which can be found on this documentation page. The features of place we will retrieve from their API are:
        - Name
        - Location
        - Categories

- Description

- Phrases:List of phrases commonly seen in this venue's tips, as well as a sample tip snippet and the number of tips this phrase appears in.

- Tips: Contains the total count of tips and groups with friends and others as groupTypes. Groups may change over time.

- Attributes: Attributes associated with the venue, such as price tier, whether the venue takes reservations, and parking availability.

3. Methodology

The analysis and the strategy:

The strategy is based on mapping the above-described data in section 2.0, in order to facilitate the choice of at least two candidate places to live. The choice is made based on the demands imposed - the number of venues within walking distance. This visual approach and maps with popups labels allow quick identification of location and feature, thus making the selection very easy.

The processing of these DATA and its mapping will allow answering the key questions to make a decision:

- Which neighborhood has the most venus to live?
- How venues distribute among the Metro Vancouver area?
- Are there tradeoffs between size and price and location?

Exploratory Analysis

Vancouver Geographical data

From the Vancouver postal code Wikipedia page, 87 neighborhoods in Vancouver are collected.
It's worth noting that the neighborhood name in Vancouver is arranged differently with Toronto.
The neighborhood in Vancouver is further classified into 179 communities. Each community has
a unique postal code. The area is visualized in the map:
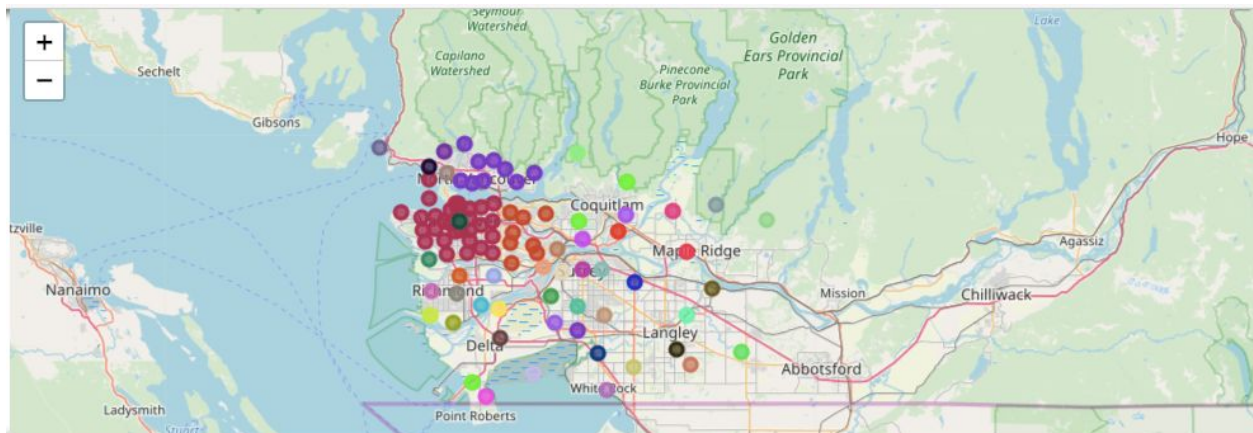
By fetching the latitude and longitude coordinates using geo-coder and the postal code scraped
from the Wikipedia page, a data frame is created to store the postal code, neighborhood,
community, latitude and longitude coordinates:

Out[15]:

	postalCode	neighbourhood	community	Latitude	Longitude
30	V5E	Burnaby	Lakeview-Mayfield / Richmond Park / Kingsway-B...	49.226625	-122.954117
111	V5R	Vancouver	South Renfrew-Collingwood	49.239335	-123.041105

The data frame is also saved in a CSV file for faster data analysis in the future.

Vancouver Venues

To determine whether a place is convenient to live without a car, we get the venues for each neighborhood using Four Square's Places API. Because we wanted to count the venues that are accessible to residents by foot, we set the radius to be 200m when searching for nearby venues.

597 nearby venues are returned by Foursquare API for the 179 communities. These venues are distributed among 104 venue categories.

The top 3 venue categories are:

1. Chinese Restaurant, 91 venues

2. Liquor Store, 89 venues

3. Bubble Tea Shop, 87 venues

The 3 communities with the most venues within 200 meters are:

1. Vancouver, Waterfront / Coal Harbour / Canada Place - 27 venues

2. Vancouver, West Kensington-Cedar Cottage / NE Riley Park-Little Mountain - 21 venues

3. North Vancouver, district municipality - 20 venues

The result is consistent with my understanding of the neighborhoods.

After getting all the nearby neighborhood data, one-hot encoding is performed to count the venues under each venue category for each community. The resulted data frame has the following structure:

(133, 105)

|	Postcode	American Restaurant	Art Gallery	Arts & Crafts Store	Asian Restaurant	Australian Restaurant	Auto Garage	Bakery	Bank	Bar	...	Sushi Restaurant	Taco Place	Tea Room	Thai Restaurant	Theater	Thrift / Vintage Store	Trail	Vegetarian / Vegan Restaurant	Vietnamese Restaurant	Yoga Studio
0	V1B	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1	V1C	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	V1E	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3	V1G	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4	V1H	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
5	V1J	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
6	V1K	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
7	V1L	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

This data frame has 179 rows for each community, 1 column for postcode, and 104 columns for each venue category.

Machine Learning

K-means clustering is used to cluster the 179 communities based on their number of venues in each venue category. Each community can be considered as having 104 attributes, which are the number of venues per category. 5 clusters are generated and visualized. The results are discussed in the next section.

Results

The map below shows the clustering result. Most of the communities are clustered into one cluster. We will examine each cluster in detail.
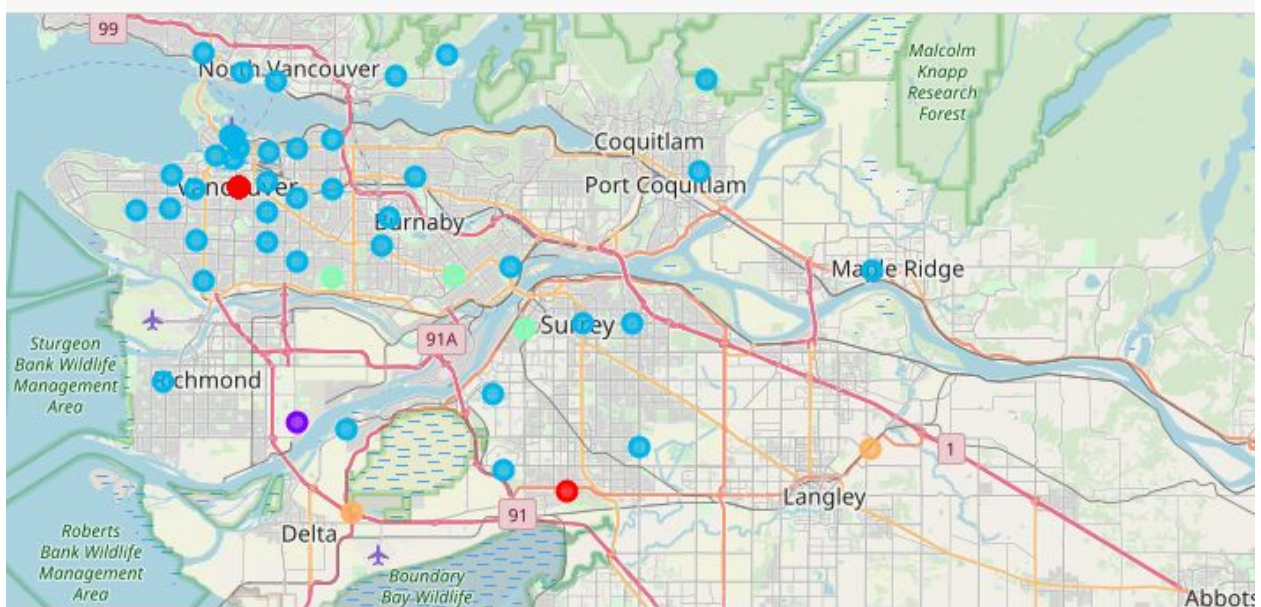
Cluster 1

Cluster 1 has 86 communities, which are represented by the red dots on the map above. Because these places are fairly scattered on the map, they are not all included in the map screenshot. These are rural areas that have very few venues nearby and need to be navigated around by car.

Cluster 2

Cluster 2 only has one community in it, which is Richmond Sourtheast represented by the purple dot on the map. The venues returned for this area is only a Garden centre. It's not a convenient place to live without a car.

Cluster 3

Cluster 3 returned 40 communities. These communities are represented by the blue dots on the map. They near the city centre and have relatively more venues than other communities. These places are convenient places to live with car.

Cluster 4

Cluster 4 returned 3 communities. They are represented by the green dots on the map. These three places are near bus stops and bus stations. However, there are not any other venues near walking distance.

Cluster 5

Cluster 5 returned 2 communities. These two areas are in the same clusters because the only venue categories returned for them are construction & landscaping. These two areas are not convenient to live.

Discussion

Communities in Cluster 3 have the most venues near walking distance. However, because there are 40 venues returned in this cluster, we need to take a closer look into them to narrow down the options.

Among the venues in Cluster 3, the top 5 communities that have the most venues are:

1. Vancouver, Waterfront / Coal Harbour / Canada Place - 27 venues

2. Vancouver, West Kensington-Cedar Cottage / NE Riley Park-Little Mountain - 21 venues

3. North Vancouver, district municipality - 20 venues

4. Vancouver, SE Kerrisdale/SW Oakridge/West Marpole - 18 venues

5. Vancouver, Bentall Centre - 17 venues

The above 5 communities are the most recommended for people to live without a car based on this analysis.

Conclusion

The analysis recommended five areas in Metro-Vancouver for people to live without a car. The result of the analysis is consistent with my knowledge of Metro-Vancouver. There are a few enhancements we can do to make better recommendations for people.

The first improvement is only considering the venue categories that the users are interested in to cluster the neighbourhoods. For example, some people care about Chinese Restaurants, whereas some other people want to live near Greek Restaurants. It would be helpful if we can first collect the venue categories that the users are interested in and then only cluster the neighbourhoods based on their needs.

The second improvement is to take price factors in as housing and rental cost in Vancouver are very high. Even though some places are convenient to live in, it's not affordable for many people.