Introduction

Here I will discuss the business problem and who would be interested in this project.

A new student is moving to Vancouver, Canada to study at UBC. To help her decide where to live, s/he would like to know which neighborhood has which characteristics and how each neighborhood differs from each other. Other potential interested parties may include real estate agents, housing developers, etc.

Data

Here I will describe the data that will be used to solve the problem and the source of the data.

In courtesy to GeoNames (https://www.geonames.org), a downloadable csv file is available and contains all Canadian postal codes and corresponding geospatial data. Only the first letters of the full postal codes are provided for copyright reasons. The data format is tab-delimited text in utf8 encoding, with the following fields:

- country code: iso country code, 2 characters
- postal code: varchar(20)
- place name: varchar(180)
- admin name1: 1st order subdivision (state) varchar(100)
- admin code1: 1st order subdivision (state) varchar(20)
- admin name2: 2nd order subdivision (county/province) varchar(100)
- admin code2: 2nd order subdivision (county/province) varchar(20)
- admin name3: 3rd order subdivision (community) varchar(100)
- admin code3: 3rd order subdivision (community) varchar(20)
- latitude: estimated latitude (wgs84)
- longitude: estimated longitude (wgs84)
- accuracy: accuracy of lat/lng from 1=estimated, 4=geonameid, 6=centroid of addresses or shape

We also utilize the Foursquare, which is the location data provider we practiced using in this course, API to explore the neighborhoods and segment them.

Methodology and Results

Here I will discuss and describe exploratory data analysis that I did, inferential statistical testing that I performed, and what machine learnings were used and why.

We retrieved our target locations' geospatial data by postal codes and, neighborhood, latitude, and longitude:

|	Postal Code	Neighborhood	Latitude	Longitude
0	V5K	Vancouver (North Hastings-Sunrise)	49.2807	-123.0397
1	V5L	Vancouver (North Grandview-Woodlands)	49.2795	-123.0667
2	V5M	Vancouver (South Hastings-Sunrise / North Renf...	49.2600	-123.0398
3	V5N	Vancouver (South Grandview-Woodlands / NE Kens...	49.2551	-123.0667
4	V5P	Vancouver (SE Kensington / Victoria-Fraserview)	49.2220	-123.0683

We visualized the target neighborhoods:
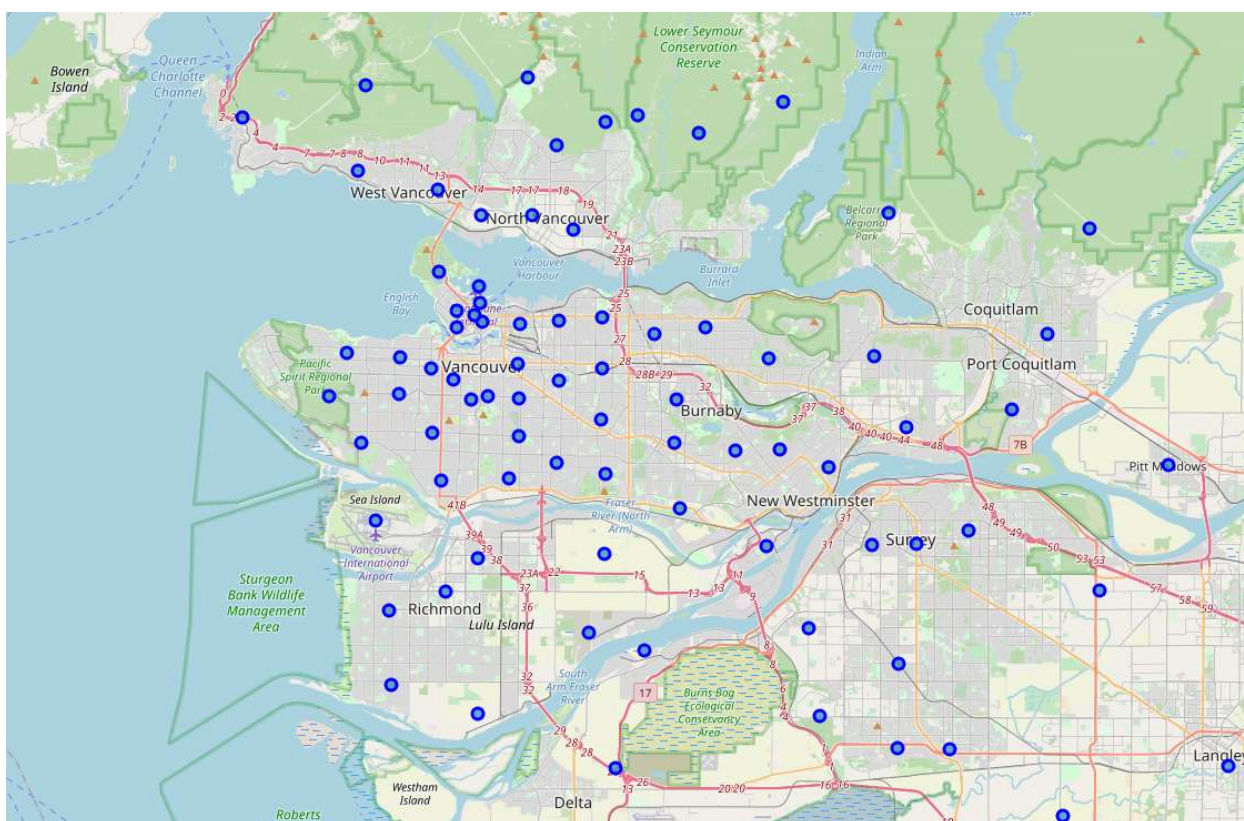
We retrieved top 20 venues within 1000-meter radius of each neighborhood in Greater Vancouver. We then checked how many venues were returned for each neighborhood. From this we got to now there are 1,223 venues in 219 unique categories across our target 88 neighborhoods: ranging from ATM, various types of restaurants, public facilities, to yoga studios, etc.

Next, we grouped results by each neighborhood by taking the mean of the frequency of occurrence of each venue category. Thus, we can show top 10 venue categories in each neighborhood.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Burnaby (Burnaby Heights / Willingdon Heights ...	Café	French Restaurant	Sushi Restaurant	Sporting Goods Shop	Liquor Store
1	Burnaby (Cascade-Schou / Douglas-Gilpin)	Sandwich Place	Insurance Office	Vegetarian / Vegan Restaurant	Park	Auto Garage
2	Burnaby (East Big Bend / Stride Avenue / Edmon...	Market	Park	Burger Joint	Chinese Restaurant	Motorcycle Shop
3	Burnaby (Government Road / Lake City / SFU / B...	Burger Joint	Café	Golf Course	Golf Driving Range	Pizza Place
4	Burnaby (Lakeview-Mayfield / Richmond Park / K...	Sushi Restaurant	Yoga Studio	Liquor Store	Bus Stop	Frozen Yogurt Shop
5	Burnaby (Maywood / Marlborough / Oakalla / Win...	Bookstore	Portuguese Restaurant	Fast Food Restaurant	Shanghai Restaurant	Shopping Mall
6	Burnaby (Parkcrest-Aubrey / Ardingley-Sprott)	Italian Restaurant	Bank	Coffee Shop	Pizza Place	Pharmacy

To understand how each neighborhood differs from each other, we ran *k*-means clustering to cluster the neighborhoods into 6 clusters. Finally, we show visualization of the resulting clusters.
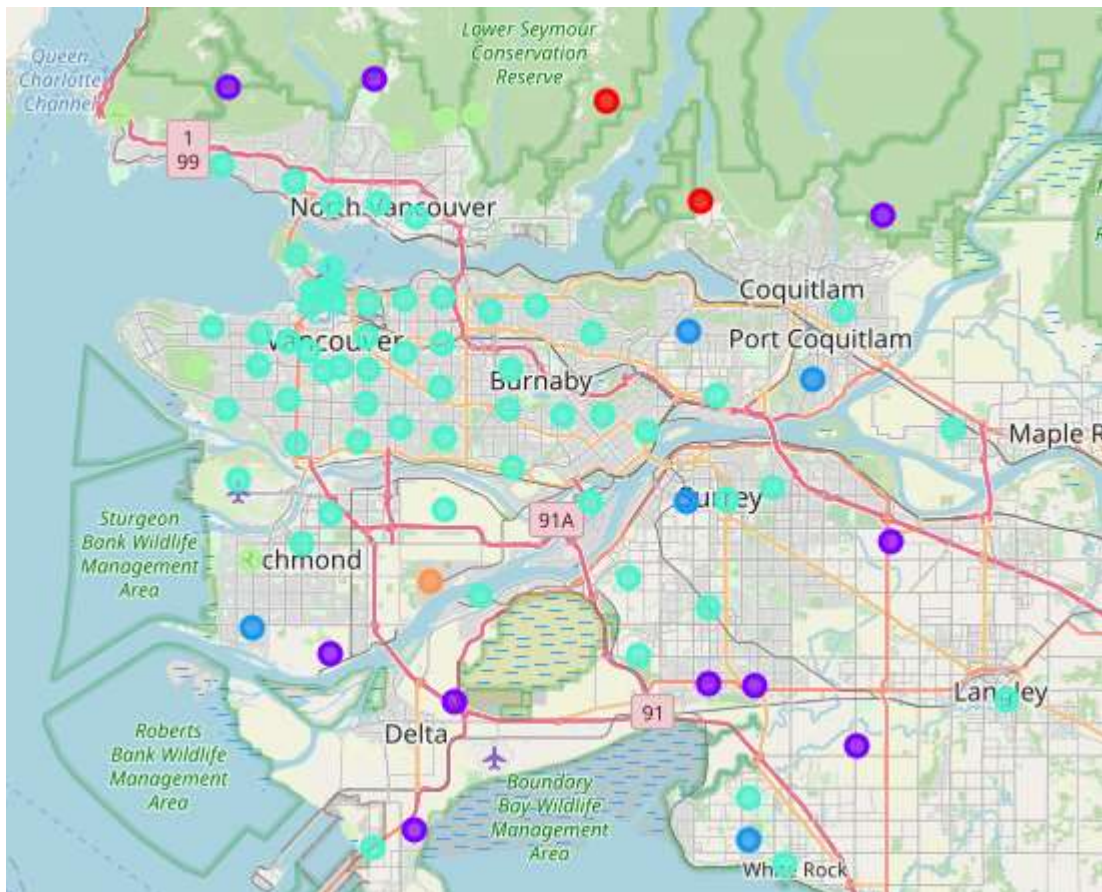
Conclusion and Discussion

Here I will discuss observations and recommendations based on the results.

Neighborhood clustering based on Foursquare venues data effectively display each neighborhood's venues or lifestyle characteristics.

For potential client interest, we could enhance neighborhood segmentation by including other neighborhood-specific socio-economic data. E.g. income level, education level, crime rate, housing prices, etc.