IBM Data Science Capstone Project

Part of IBM Data Science Professional Certificate

Opening a craft beer tap room in Vancouver

By: Agnieszka Jugowicz

July 2020

Table of Contents

1. Introduction

Vancouver is a coastal seaport city in the mainland of British Columbia, Canada. Vancouver is consistently ranked one of the most livable cities in the world. Its location near ocean, tucked up against mountains, makes the city a go-to location for year-round exploration.

Vancouver's metropolitan area, with population of over 600,000 inhabitants, is an ideal place for investment, especially in the hospitality industry. The tourism is flourishing, as the city combines an urban getaway with outdoor adventures. Vancouver has also a rich cultural background and is third-largest film production center in North America, holding the moniker of "Hollywood North".

This project will attempt to explore the patterns of neighbourhoods within Vancouver by categorizing them into clusters in order to identify existing trends within each neighbourhood. Using data science and machine learning techniques, this project aims to provide an answer to the business question: If an investor is looking to open a new craft beer tap room, where should it be opened?

The recommendation is aimed at investors wanting to open or expand their business in the hospitality sector of Vancouver.

2. Data

In order to analyze the trends within each neighbourhoods of Vancouver, the list of forward sortation areas (FSAs) was downloaded from the website https://www.geonames.org/postalcode-search.html?q=vancouver&country=CA as CSV file. Forward sortation area is a geographical region in which postal codes start with the same three letters. It enables a geographical grouping of city areas, creating neighbourgoods.

The database, as seen below, consists of the following columns: *CountryCode*, *FSA*, *Neighbourhoods*, *Province*, *Latitude*, and *Longitude*.

	CountryCode	FSA	Neighbourhoods	Province	Latitude	Longitude
0	CA	V5K	Vancouver (North Hastings-Sunrise)	British Columbia	49.2807	-123.0397
1	CA	V5L	Vancouver (North Grandview-Woodlands)	British Columbia	49.2795	-123.0667
2	CA	V5M	Vancouver (South Hastings-Sunrise / North Renf...	British Columbia	49.2600	-123.0398
3	CA	V5N	Vancouver (South Grandview-Woodlands / NE Kens...	British Columbia	49.2551	-123.0667
4	CA	V5P	Vancouver (SE Kensington / Victoria-Fraserview)	British Columbia	49.2220	-123.0683

Venue queries will be performed using FourSquare APIs. The results of the venue queries will then be used to observe commonalities between neighbourhoods. The clusters will then provide insights on which cluster is most suitable for the investment. K-means clustering algorithm will be used to find patterns between the neighbourhoods.

3. Methodology

Firstly, the data retrieved from www.geonames.org website will be loaded into pandas dataframe and then visualized in a map using Folium package in order to validate the correctness of data. Next, the FourSquare API will be used to obtain the venues data. API calls to FourSquare will be made, passing in the geographical coordinates of the neighbourhoods in a Python loop.

The K-means clustering algorithm will be used to categorize neighbourhoods within Vancouver. A one-hot encoding will be performed on the venue dataframe and the dataframe will be grouped by each neighbourhood. The encoding will provide information on venue categories (in columns) per neighbourhood. This data will then be grouped to provide weighting of venue type occurrence per neighbourhood. Next, the encoded dataframe will be filtered into top venues, before K-means algorithm is implemented.

4. Analysis

Vancouver consists of over 40 different FSA codes, which divide the city into neighbourhoods. A map visualizing the neighbourhoods can be seen below.
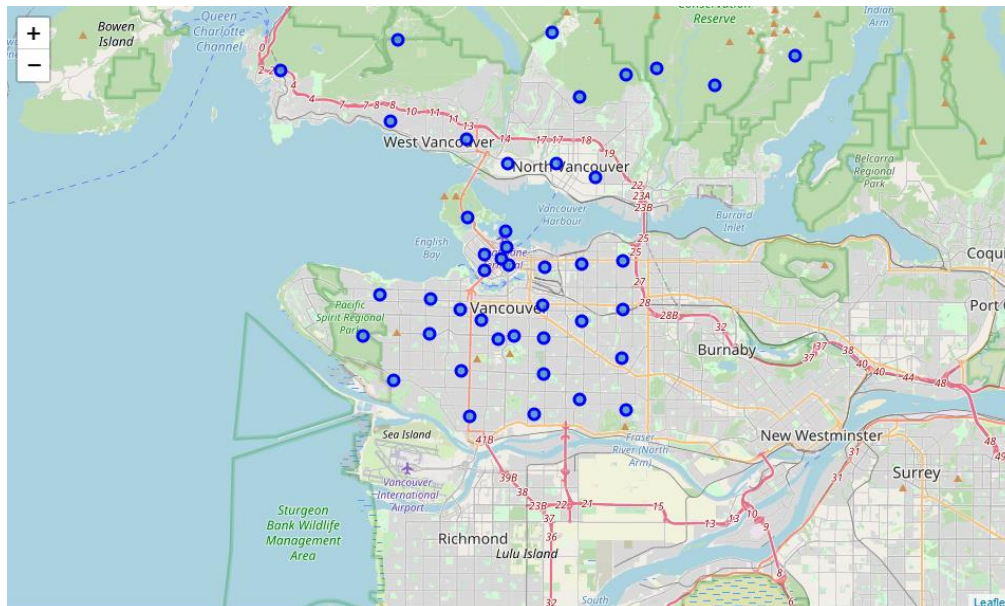
The FourSquare analysis of venues resulted in finding 249 unique venue categories in the analyzed dataset. These venues were grouped per neighbourhood, and most common venues per neighbourhood were selected. Below a least of first 5 rows of the newly created dataframe.

	Neighbourhoods	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	North Vancouver East Central	Trail	Park	Lake	Mountain	Coffee Shop	Scenic Lookout	French Restaurant	Fried Chicken Joint	Food Truck	Filipino Restaurant
1	North Vancouver Inner East	Ski Area	Trail	Yoga Studio	Food & Drink Shop	Financial or Legal Service	Fish & Chips Shop	Flower Shop	Food	Food Court	Field
2	North Vancouver North Central	Trail	Park	Scenic Lookout	Coffee Shop	Mountain	Lake	Convenience Store	Food Court	Fish & Chips Shop	Flower Shop
3	North Vancouver Northwest	Trail	Scenic Lookout	Ski Chairlift	Coffee Shop	Mountain	Ski Chalet	Restaurant	Fast Food Restaurant	Dessert Shop	Sporting Goods Shop
4	North Vancouver Northwest Central	Trail	Park	Coffee Shop	Convenience Store	Chinese Restaurant	Shopping Mall	Grocery Store	Bar	Gym / Fitness Center	Sandwich Place

The k-means clustering algorithm presented a list of 4 clusters:
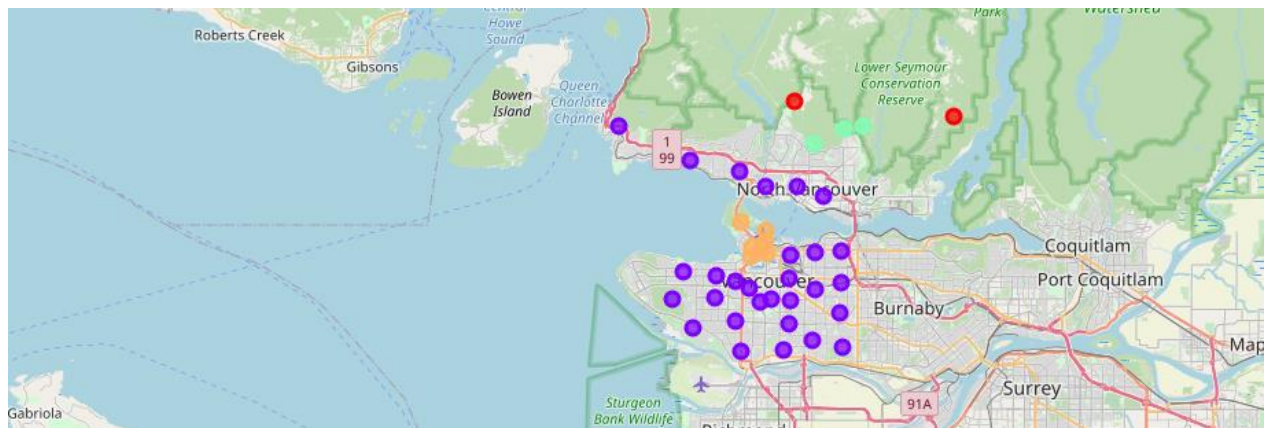
Cluster 1

Includes neighbourhoods located in the center of Vancouver and on the north coast line of Vancouver harbor. This cluster includes the majority of neighbourhoods and contains a higher concentration of restaurants and coffee shops. It can be observed that the hospitality infrastructure is well-developed.

	Neighbourhoods	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue
0	Vancouver (North Hastings-Sunrise)	49.2807	-123.0397	1	Vietnamese Restaurant	Café	Coffee Shop	Park	Bakery	French Restaurant	Chinese Restaurant	Pub
1	Vancouver (North Grandview-Woodlands)	49.2795	-123.0667	1	Coffee Shop	Brewery	Café	Pizza Place	Sushi Restaurant	Italian Restaurant	Indian Restaurant	Grocery Store
2	Vancouver (South Hastings-Sunrise / North Renf...	49.2600	-123.0398	1	Coffee Shop	Park	Bus Stop	Pizza Place	Grocery Store	Convenience Store	Burger Joint	Bus Station
3	Vancouver (South Grandview-Woodlands / NE Kens...	49.2551	-123.0667	1	Vietnamese Restaurant	Vegetarian / Vegan Restaurant	Coffee Shop	Pizza Place	Italian Restaurant	Café	Mexican Restaurant	Indian Restaurant
4	Vancouver (SE Kensington / Victoria-Fraserview)	49.2220	-123.0683	1	Chinese Restaurant	Vietnamese Restaurant	Japanese Restaurant	Coffee Shop	Sandwich Place	Asian Restaurant	Gas Station	Bank
5	Vancouver (South Renfrew-Collingwood)	49.2397	-123.0407	1	Chinese Restaurant	Coffee Shop	Vietnamese Restaurant	Asian Restaurant	Pizza Place	Japanese Restaurant	Park	Café
6	Vancouver (Killarney)	49.2175	-123.0380	1	Coffee Shop	Park	Golf Course	Bus Stop	Convenience Store	Sandwich Place	Chinese Restaurant	Café

Cluster 2

Second cluster includes three neighbourhoods on the north site of the city. The most common venues include trails, parks and other hiking destinations. The hospitality infrastructures is also present, however, not to such extend as in cluster 1.

	Neighbourhoods	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10 C
31	North Vancouver East Central	49.3622	-123.0178	3	Trail	Park	Lake	Mountain	Coffee Shop	Scenic Lookout	French Restaurant	Fried Chicken Joint	Food Truck	Re:
32	North Vancouver North Central	49.3597	-123.0377	3	Trail	Park	Scenic Lookout	Coffee Shop	Mountain	Lake	Convenience Store	Food Court	Fish & Chips Shop	
35	North Vancouver Northwest Central	49.3500	-123.0679	3	Trail	Park	Coffee Shop	Convenience Store	Chinese Restaurant	Shopping Mall	Grocery Store	Bar	Gym / Fitness Center	S:

Cluster 3

Third cluster includes locations concentrated on the area between Vancouver Harbour and English Bay. This cluster has a strong hospitality infrastructure, with many hotels and restaurants. It is a tourism center.

	Neighbourhoods	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	
14	Vancouver (NE Downtown / Harbour Centre / Gast...	49.2788	-123.1139	4	Hotel	Restaurant	Bakery	Dessert Shop	Sandwich Place	Coffee Shop	Park	Japanese Restaurant	Breakfast Spot	T
15	Vancouver (Waterfront / Coal Harbour / Canada ...	49.2866	-123.1158	4	Hotel	Restaurant	Dessert Shop	Café	Sandwich Place	Breakfast Spot	Coffee Shop	Plaza	Concert Hall	
16	Vancouver (South West End)	49.2833	-123.1298	4	Hotel	Dessert Shop	Bakery	Restaurant	Japanese Restaurant	Park	Seafood Restaurant	Sandwich Place	Ramen Restaurant	
17	Vancouver (North West End / Stanley Park)	49.2990	-123.1408	4	Coffee Shop	Scenic Lookout	Seafood Restaurant	Park	Sculpture Garden	Restaurant	Ramen Restaurant	Trail	Beach	
28	Vancouver (SW Downtown)	49.2767	-123.1300	4	Hotel	Seafood Restaurant	Bakery	Restaurant	Park	Dessert Shop	Italian Restaurant	Trail	Coffee Shop	F
42	Vancouver (Bentall Centre)	49.2935	-123.1162	4	Hotel	Coffee Shop	Dessert Shop	Sandwich Place	Restaurant	Café	Plaza	Park	Ramen Restaurant	
43	Vancouver (Pacific Centre)	49.2819	-123.1190	4	Hotel	Restaurant	Dessert Shop	Sandwich Place	Bakery	Coffee Shop	Italian Restaurant	Japanese Restaurant	Café	

Cluster 4

Last cluster includes areas furthest north, right next to hiking trails and ski chair lifts. These are recreational areas with gastronomy infrastructure.

	Neighbourhoods	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	C
29	North Vancouver Outer East	49.3678	-122.9278	5	Trail	Ski Chairlift	Ski Lodge	Outdoors & Recreation	Restaurant	Cafeteria	Ski Area	Scenic Lookout	Filipino Restaurant	F
37	North Vancouver Northwest	49.3775	-123.0862	5	Trail	Scenic Lookout	Ski Chairlift	Coffee Shop	Mountain	Ski Chalet	Restaurant	Fast Food Restaurant	Dessert Shop	

5. Result and Discussion

A new craft beer tap room should be located in an area with existing hospitality infrastructure. Preferably, it should be an area in the city center that many tourists may easily access, with hotels in close proximity.

K-means clustering presented 4 distinctive clusters: Cluster 1 grouped residential areas, which have a strong presence of facilities such as cafes, restaurants. Cluster 2 grouped areas that can be characterized by prevalence of hiking trails that also have other touristic amenities such as coffee shops and restaurants. Cluster 3 grouped postal codes located in the center of the city with many hotels and touristic facilities. Finally, cluster 4 grouped neighbourhoods with prevalence of hiking trails and skying infrastructure.

It is worth to highlight that FourSquare API limit of 100 venues may result in the skew of results.

6. Conclusion

The analysis showed that opening a new craft beer tap room should be considered in cluster 3. This area consists of neighbourhoods with the strongest prevalence of touristic facilities: hotels, bars and restaurants. Even though there may be a stronger competition, this location guarantees the inflow of new customers, either visiting, or those being city's residents.