Capstone Project – 'The Battle of Neighborhoods'

Prologue

On January 1, 2020, I moved to Canada carrying dreams in my eyes to work, travel and live my life. Never knew that many of my plans would be cancelled with the surge of Corona Virus disease, profoundly affecting the life around the world. Given the life-threatening and economic challenging crisis people facing right now, one can only hope to stay positive and test negative. The isolation and shutdown enforced me to devote my time in understanding the value of learning and upgrading my skills. Having research experience in data analyses using python, knowing the basics paved the way to leverage my skills by learning the fundamentals of data science and various tools and methods in practice such as exploratory data analysis, machine learning etc.

I would like to recommend the IBM Data Science Professional Certificate Course. The final assignment of this course is the so-called "Capstone Project" in which many of the tools and methods learned during the course are applied in a "Battle of Neighbourhoods" challenge. In this challenge, I am asked to choose a city and design a problem that can be solved using location data in addition to other datasets. I choose the city of Vancouver as a case. I hope you like my blogpost on this capstone project.

Introduction/Busines Problem

Canada has built a reputation of welcoming a number of immigrants every year and valuing multiculturalism. Vancouver is among the topmost cities that offers great opportunities to new immigrants in terms of work opportunities, education, lifestyle, and food choices. Vancouver is known for its sushi, fresh seafood, the finest house-made charcuterie, the most delicious taco truck, and the B.C. wines. For an immigrant like me, who prefer vegetarian food, finding a right place to eat can be a daunting challenge.

Thus, the goal of this project is to explore, segment, and cluster the neighborhoods in the city of Vancouver and provide recommendations to my friend's start-up to develop an app for food-delivery based on the common restaurant categories in different neighborhoods. The research questions specific to this business problem are: In which neighbourhood of the city will you find large number of restaurants? Which neighbourhood have the higher concentration of restaurants depending on its type? Where to find vegan-friendly restaurant? Where to find restaurants with Indian, or mediterranen cuisine? The target population for my project is the new immigrants who recently moved to Vancouver.

Description of the data

For the Vancouver neighbourhood data, a Wikipedia page exists that has all the information that can be used to explore and cluster the 20 official neighbourhoods in the city of Vancouver. The corresponding coordinates are obtained using geocoder.
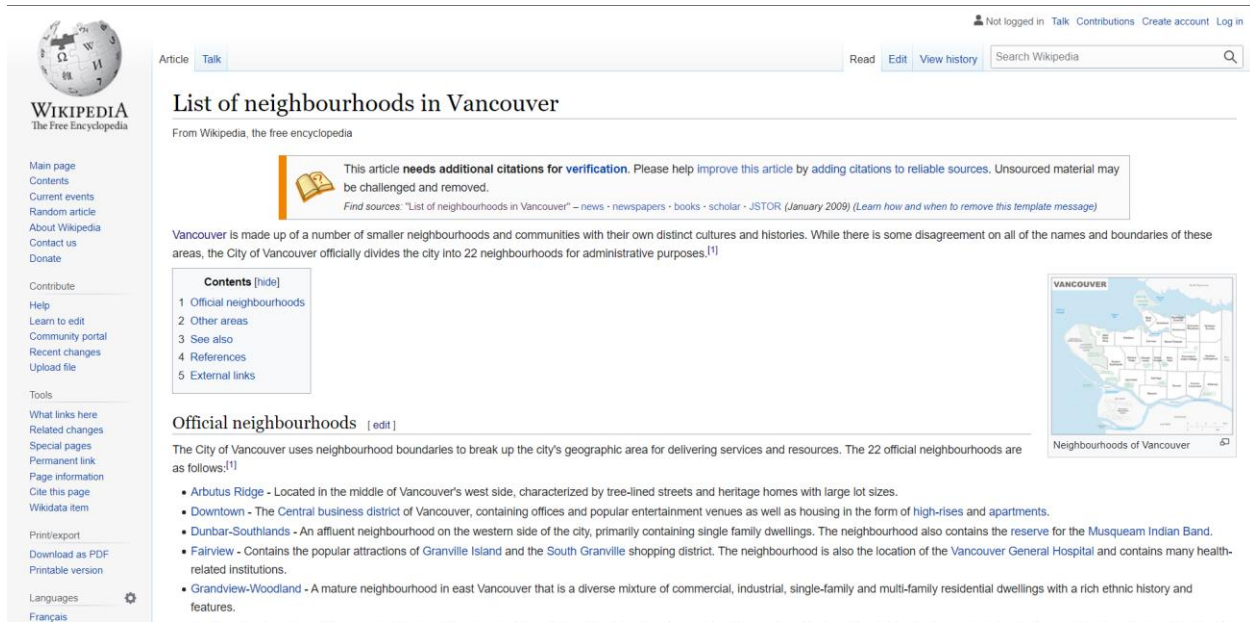
Figure 1 List of neighborhoods in Vancouver (Source: Wikipedia)

As mandated in the challenge instructions, I will use the Foursquare location data about restaurants in Vancouver. Foursquare is a US technology company that provides recommendations of the places via a location technology platform offering business solutions and consumer products near a user's specified location based on users' previous browsing history and check-in history. I will use foursquare data such as the restaurant name, ID, location and category of food.

I will use the neighbourhoods in Vancouver, geocoder, and data about the restaurants in these neighbourhoods from foursquare to show the density of restaurants in them. It is noteworthy to mention that while working on this project, most of the Canadian cities are following strict lockdown measures and several restaurants are only open for take-aways. This limitation may affect the sample size of this project for analysis. However, this project can be implemented for future use as well.

Methodology

In this section, I will describe about the procedure used to collect data from various sources, converting into data into tabular form for pre-processing and exploratory data analysis. I started out by scrapping the data from Wikipedia to create a dataframe with the neighborhood of the city of Vancouver. I used various python libraries such as pandas, requests, BeautifulSoup. I cleaned the data to keep the required information such as neighborhood's names and discard the rest description. The next step is to get the coordinates of all the neighborhood using geocoder and nominatim function to add geospatial data coordinates- latitudes and longitudes. Further, I append geospatial coordinates to the dataframe which resulted in the below table:

	Neighborhood	Latitude	Longitude
0	Grandview-Woodland	49.28296	-123.06934
1	Hastings-Sunrise	49.28021	-123.02869
2	Kensington-Cedar Cottage	49.24973	-123.06986
3	Kerrisdale	49.22097	-123.15073
4	Killarney	49.22670	-123.03657
5	Kitsilano	49.26833	-123.16542
6	Marpole	49.20067	-123.13053
7	Mount Pleasant	49.26178	-123.09608
8	Oakridge	49.23004	-123.12117
9	Renfrew-Collingwood	49.24970	-123.02975
10	Riley Park	49.25028	-123.10069
11	Shaughnessy	49.25039	-123.12981
12	South Cambie	49.25010	-123.11999
13	Strathcona	49.28009	-123.08024
14	Sunset	49.22026	-123.08023
15	Victoria-Fraserview	49.22027	-123.06941
16	West End	49.28619	-123.13373
17	West Point Grey	49.27110	-123.20229
18	Yaletown	49.27432	-123.12180
19	Coal Harbour	49.29120	-123.12965

Figure 2 Neighborhood dataframe with geospatial corrdiantes

I visualized the map of Vancouver with neighborhoods denoted by blue circles using folium package and my dataframe as shown in figure 3.

Then, I utilized the Foursquare API to explore the neighborhoods and segment them. I retrieved the foursquare data for all venues in Vancouver on foursquare within a radius of 500 m and limited the number of venues to 100 in each neighborhood that resulted in 560 venues in total. I explored the neighborhood with the number of venues in each and later, analyzed for restaurant only venues. There was a total of 36 unique restaurant categories. Given the ethnicity and culturally diverse population of the city, there is a diversity in the specialized cuisine restaurants. The dataframe was further analyzed by segregating the restaurants using onehot encoding and listed the top 10 most common restaurants in each neighborhood.

Furthermore, I ran an unsupervised machine learning algorithm. I clustered the neighborhoods into 5 categories using k-means clustering and labelled these clusters depending on the commonality of cuisines in those clusters. I also tried with k value of 3 and 6, but I fixed my analysis for k value to be 5 considering k=5 is the most general case used in clustering method. These clusters are further analyzed to answer our research questions and discussed in the subsequent sections.
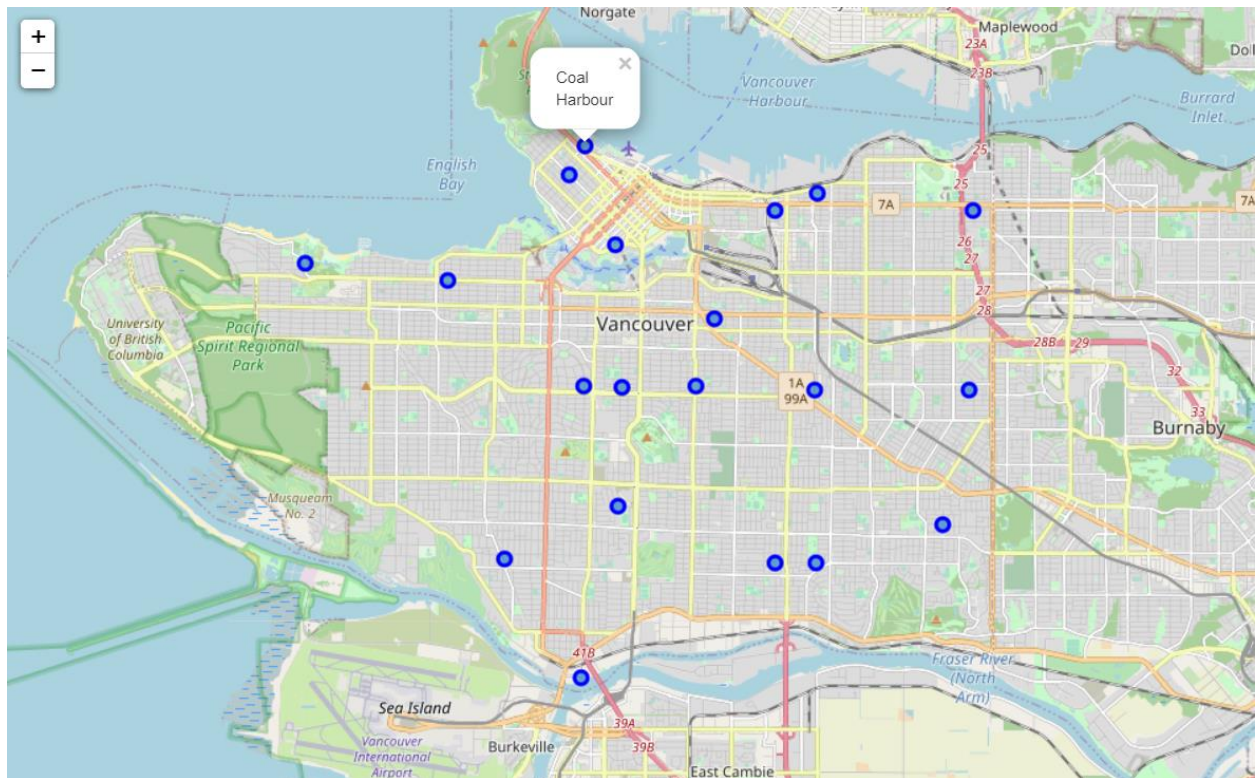
Figure 3 Vancouver City map visualized using folium

Results

From the foursquare data, we figured that there are 36 unique categories in the restaurant venues in Vancouver. We analyzed the top 10 most frequently occurring restaurants' categories as shown in figure 4.
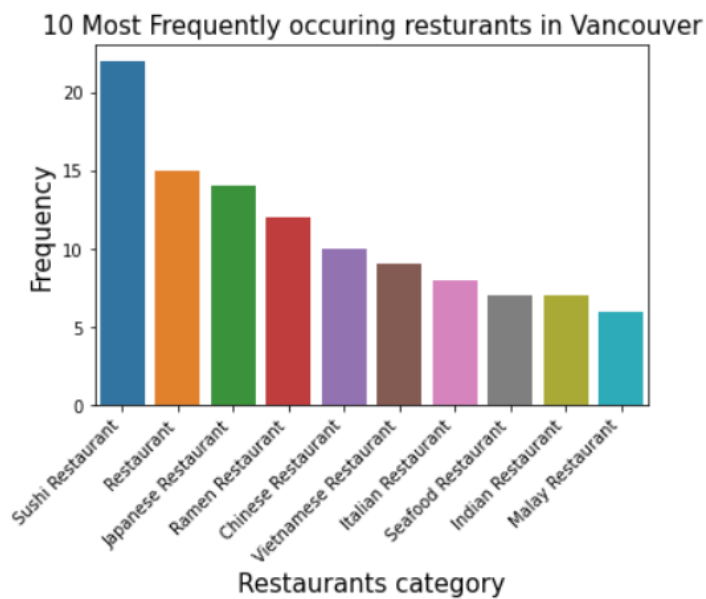

Figure 4 Most frequently occurring restaurants in Vancouver

Restaurants serving sushi cuisine are the winners in the list of topmost frequently occurring restaurants in Vancouver. As my project is looking more for Vegetarian/Vegan restaurants or Indian or Italian restaurants, I am amazed to see theses cuisine in the top 10 list of restaurants. While this is interesting, I would like to dig deeper to know which neighborhoods have the higher number of aforementioned cuisines.
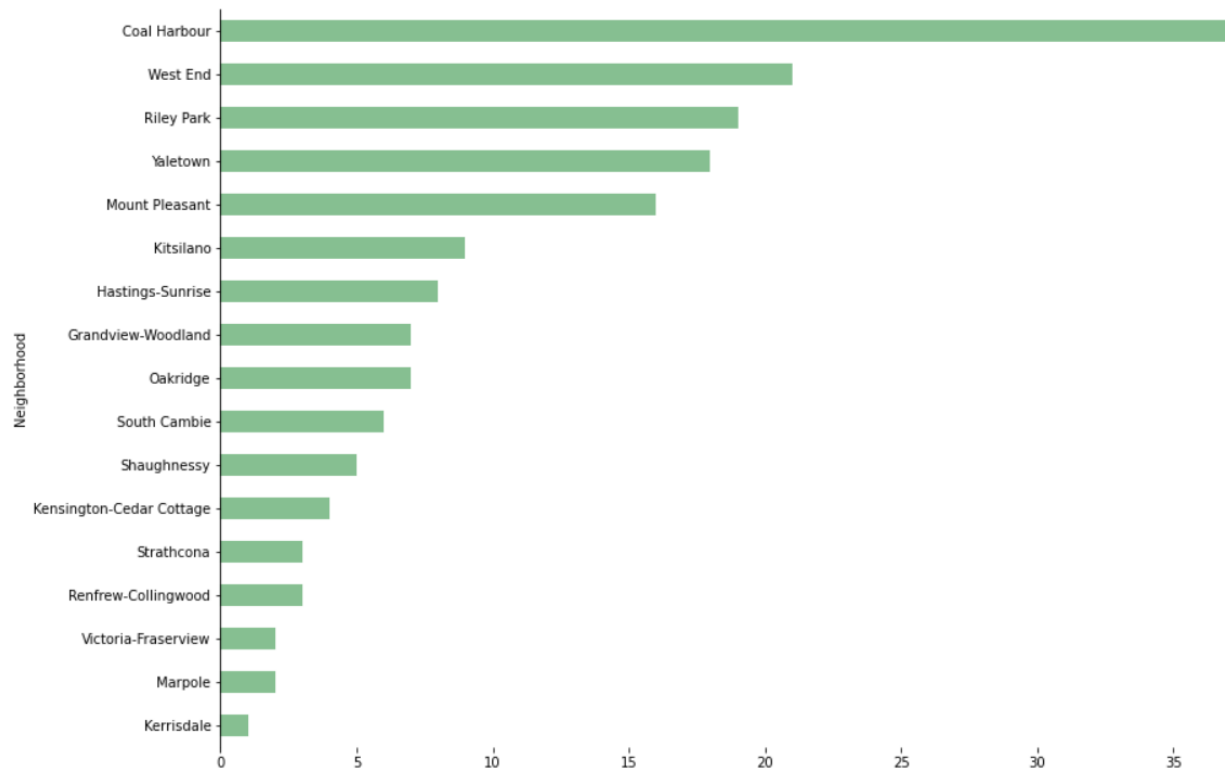

Figure 5 Number of restaurants in each neighborhood in Vancouver

Out of total 168 restaurant venues, Coal Harbour seems to be the most popular neighborhood for restaurants, followed by West End as seen in figure 5. Riley Park, Yaletown and Mount Pleasant are almost equally popular with 19, 18 and 16 number of restaurant venues, respectively. Kerrisdale is having the least number of restaurants.

I sorted the list to display the neighborhood and its 10 most common restaurant venues as shown in figure 6 which is further used for clustering. Coal Harbor undoubtedly topped the chart with Japanese cuisine and seafood in top 5 most common venues, claiming the fact that Vancouver is indeed known for sushi and seafood. Nevertheless, I can view a variety in commonality of other cuisines in other neighborhoods. I used the k-means clustering machine learning algorithm to make clusters of all neighborhoods in the city.

	Cluster Labels	Neighborhood	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	1	Coal Harbour	Japanese Restaurant	Ramen Restaurant	Sushi Restaurant	Korean Restaurant	Seafood Restaurant	Restaurant	Italian Restaurant	Vietnamese Restaurant	Malay Restaurant	Asian Restaurant
1	1	Grandview-Woodland	Fast Food Restaurant	Italian Restaurant	Asian Restaurant	Sushi Restaurant	Chinese Restaurant	Middle Eastern Restaurant	French Restaurant	Vietnamese Restaurant	Indian Restaurant	Hotpot Restaurant
2	1	Hastings-Sunrise	Restaurant	Chinese Restaurant	Malay Restaurant	Fast Food Restaurant	Seafood Restaurant	Italian Restaurant	Sushi Restaurant	Asian Restaurant	Caribbean Restaurant	Cantonese Restaurant
3	1	Kensington-Cedar Cottage	Vietnamese Restaurant	Indian Restaurant	Seafood Restaurant	Asian Restaurant	Cantonese Restaurant	Caribbean Restaurant	Chinese Restaurant	Eastern European Restaurant	Ethiopian Restaurant	Korean Restaurant
4	2	Kerrisdale	Spanish Restaurant	Vietnamese Restaurant	Falafel Restaurant	Italian Restaurant	Indian Restaurant	Hotpot Restaurant	Greek Restaurant	French Restaurant	Fast Food Restaurant	Ethiopian Restaurant
5	1	Kitsilano	Vietnamese Restaurant	Southern / Soul Food Restaurant	Vegetarian / Vegan Restaurant	Italian Restaurant	Chinese Restaurant	Restaurant	Sushi Restaurant	Thai Restaurant	Indian Restaurant	Asian Restaurant
6	3	Marpole	Restaurant	Vietnamese Restaurant	Korean Restaurant	Italian Restaurant	Indian Restaurant	Hotpot Restaurant	Greek Restaurant	French Restaurant	Fast Food Restaurant	Falafel Restaurant
7	1	Mount Pleasant	Sushi Restaurant	Vietnamese Restaurant	Indian Restaurant	Restaurant	Chinese Restaurant	Ethiopian Restaurant	Mexican Restaurant	New American Restaurant	Peruvian Restaurant	Ramen Restaurant
8	1	Oakridge	Sushi Restaurant	Fast Food Restaurant	Vietnamese Restaurant	Cantonese Restaurant	Restaurant	Falafel Restaurant	Italian Restaurant	Indian Restaurant	Hotpot Restaurant	Greek Restaurant
9	0	Renfrew-Collingwood	Malay Restaurant	Sushi Restaurant	Chinese Restaurant	Vietnamese Restaurant	Fast Food Restaurant	Italian Restaurant	Indian Restaurant	Hotpot Restaurant	Greek Restaurant	French Restaurant
10	1	Riley Park	Restaurant	Japanese Restaurant	Thai Restaurant	Sushi Restaurant	Vegetarian / Vegan Restaurant	Hotpot Restaurant	Caribbean Restaurant	Chinese Restaurant	Seafood Restaurant	Malay Restaurant
11	0	Shaughnessy	Sushi Restaurant	Malay Restaurant	Cantonese Restaurant	Greek Restaurant	Vietnamese Restaurant	Fast Food Restaurant	Italian Restaurant	Indian Restaurant	Hotpot Restaurant	French Restaurant
12	0	South Cambie	Chinese Restaurant	Vietnamese Restaurant	Malay Restaurant	Sushi Restaurant	Cantonese Restaurant	Fast Food Restaurant	Italian Restaurant	Indian Restaurant	Hotpot Restaurant	Greek Restaurant

Figure 6 Neighborhoods and their 10 most common venues

As you can see from the table in figure 6, the neighborhoods in the city of Vancouver and their most common venues are assigned five cluster labels from 0 to 4. These clusters are further analyzed and labelled depending on the most common cuisine venue in that neighborhood.
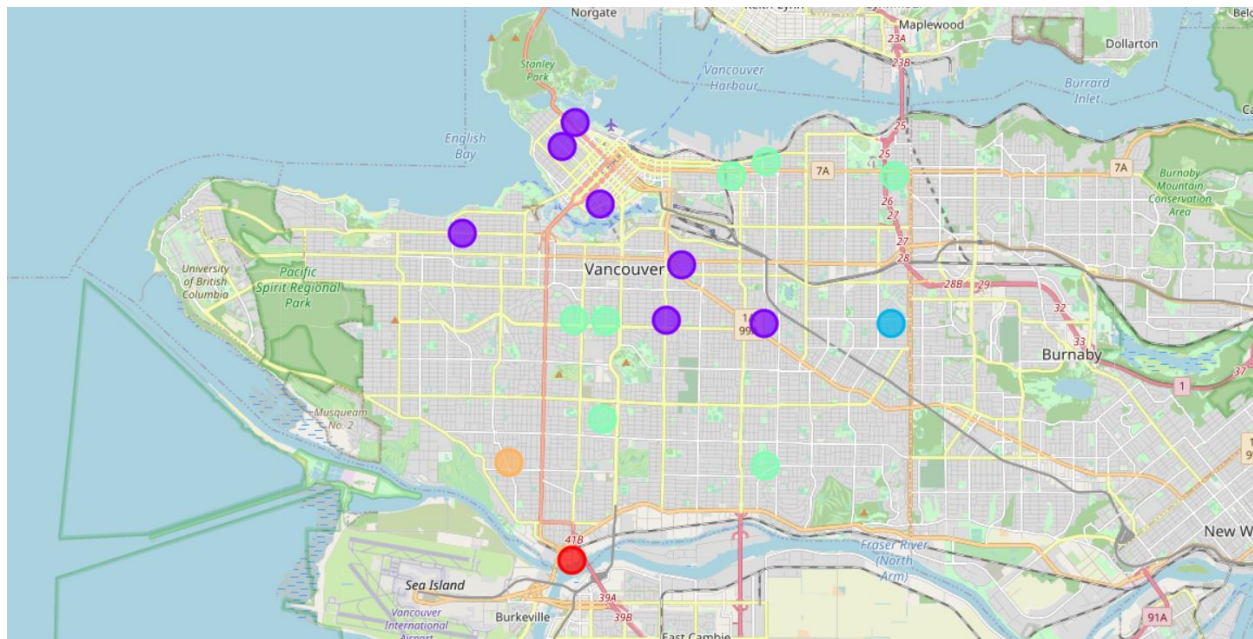

Figure 7 Neighborhood clusters in the city of Vancouver

Marpole food cluster

Contemplating the diversity in terms of population density and the residential areas because of its proximity to Airport, this cluster (in red) has most venues from Vietnamese and Korean cuisines. There are some restaurants serving Italian and Indian cuisine available too.

Downtown food cluster

The cluster 2 (in purple) included 7 neighborhoods mostly in the downtown and surrounding areas in the city. While Vietnamese and Japanese restaurant were among topmost common venues, mostly all the neighborhoods showed Vegetarian/Vegan, or Indian restaurants venues except Coal Harbor. Coal harbor as stated earlier is more for sushi and seafood lovers.

Mixed food cluster

The cluster 3 (in blue), cluster 4 (in green) and cluster 5 (in orange) included neighborhoods having residential dwellings: home, apartments, single- and multi-family residences. If we clearly analyze the most common venues in these neighborhoods, Fast food restaurants and cuisine from Asian countries top the chart. I rename this cluster as the 'Fast food cluster'.

Discussion

Coal Harbor has the highest number of restaurants with unique categories. This makes more sense as it is the hub of Vancouver and closer to downtown. There are not enough Vegetarian/Vegan friendly options available in Coal Harbor. One can find many restaurants specialised in sushi and seafood, serving the best Vancouver is known for. The surrounding areas close to downtown including Yaletown, Kensington-Cedar Cottage, Mount Pleasant and Riley Park has variety of restaurants specialised in Indian and Mediterranean cuisines. The residential areas are categorized for mixed food cluster owing to diversity in the population and multi-culturalism. Kerrisdale is the only exceptional neighborhood that excel in Spanish cuisine. Vegetarian/Vegan restaurants are more common in the neighborhoods of Kensington-Cedar Cottage, Kitsilano, Riley Park and Yaletown. As most of the immigrants would prefer to live in residential areas listed in cluster 3,4 or 5, I would recommend the start-up to strategize some ideas to deliver Vegetarian/Vegan food to the neighborhoods in the mixed food cluster.

Conclusion

In this project titled 'The Battle of Neighborhoods', my aim was to explore, segment and cluster the neighborhoods in the city of Vancouver. Undeniably, Vancouver is known as the "sushi capital of North America" and famous sushi rolls. In the past years, there has been a rising popularity in Vegetarian/vegan food preferences over meat-based/seafood. Through this project, I used the ML k-means clustering algorithm to find the most common venues in each neighborhood and was able to configure the popular neighborhoods with most Vegan friendly restaurants. Start-up company can use this analysis to develop business ideas with delivery options to target new immigrants looking for specific cuisine.

Acknowledgement

I am grateful to several cheat sheets found on github, stackoverflow and medium to assist me in working out this project and writing full report. I also like to acknowledge this article for inspiring me to design my business problem. The course of the IBM Data Science Professional Certification has immensely supported me in my journey of learning data science. For more details on this project, please follow the github. Feel free to contact me if you have any questions or comments. Looking forward to talking to you about data science!