IBM

Coursera Capstone Project Report

By Alex Li

Introduction:

Vancouver is one of the most beautiful and ethnically diverse cities in Canada. The population of this city is constantly increasing. From a business perceptive, there are lots of business opportunities and interest in Vancouver. Locations and competitions are important factors concerning the entrepreneurs and investors before investing into a business. This project will implement basis data analysis on the neighborhood location, population and common business venues to investigate the potential of starting a business in Vancouver and be interest of investors.

Data Collection:

To perform the data analysis for this problem, data is collected from the following data sources and used to determine the potential of opening a business in an area.

1- List of venues extracted from FourSquare API database. By Using the geo coordinates of the neighborhood, common venues nearby the neighborhood area extracted from API Servers.
    a. http://foursquare.com/
2- List of neighborhoods in Vancouver
    a. https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Vancouver
3- Geo-Coordinates of the Neighborhoods in Vancouver using Geocoder in Python
4- Population of Vancouver in 2016 is uploaded onto Python using panda library.
    a. https://webtransfer.vancouver.ca/opendata/csv/CensusLocalAreaProfiles2016.csv
5- Vancouver Geojson file from GitHub open source
    a. https://github.com/codeforamerica/click_that_hood/edit/master/public/data/vancouver.geojson%22

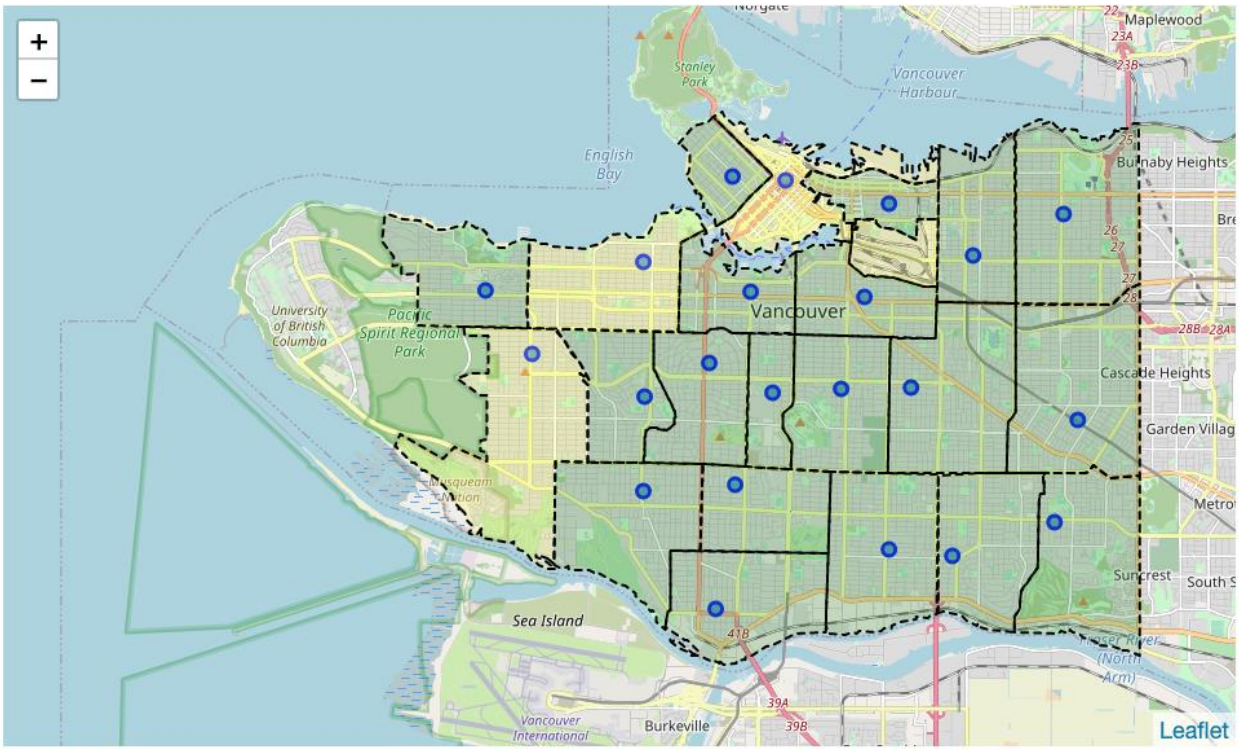
Methodology:

Data are extracted from different data resources and stored in different variables. Population of each neighborhood is downloaded into a CSV file and uploaded onto Jupyter. The neighbourhood is also verified using Wikipedia link. Using the population and the neighborhood name, the location of each neighborhood is obtained from Geocoder library. The data is organized by applying exploratory data analysis and data wrangling which is shown below.

	Neighborhood	Population	Latitude	Longitude
0	Arbutus-Ridge	15295.0	49.240968	-123.167001
1	Downtown	62030.0	49.283393	-123.117456
2	Dunbar-Southlands	21425.0	49.253460	-123.185044
3	Fairview	33620.0	49.264113	-123.126835
4	Grandview-Woodland	29175.0	49.270559	-123.067942
5	Hastings-Sunrise	34575.0	49.277594	-123.043920
6	Kensington-Cedar Cottage	49325.0	49.247632	-123.084207
7	Kerrisdale	13975.0	49.234673	-123.155389
8	Killarney	29325.0	49.224274	-123.046250
9	Kitsilano	43045.0	49.269410	-123.155267
10	Marpole	24460.0	49.209223	-123.136150
11	Mount Pleasant	32955.0	49.263330	-123.096588
12	Oakridge	13030.0	49.230829	-123.131134
13	Renfrew-Collingwood	51530.0	49.242024	-123.057679
14	Riley Park	22555.0	49.247438	-123.102966
15	Shaughnessy	8430.0	49.251863	-123.138023
16	South Cambie	7970.0	49.246685	-123.120915
17	Strathcona	12585.0	49.279554	-123.089979
18	Sunset	36500.0	49.219593	-123.090239
19	Victoria-Fraserview	31065.0	49.218416	-123.073287
20	West End	47200.0	49.284131	-123.131795
21	West Point Grey	13065.0	49.264484	-123.185433

Table 1: Neighborhood of Vancouver with geo coordinates and population

A folium map of Vancouver is created and each neighbourhood is imposed onto the map. Vancouver Geojson file is extracted from GitHub database and is used to outline the boundary of each neighbourhood.



Each neighbourhood has different venues. By using API Foursquare and setting the radius of 650 meter from the geo coordinates, the venues in the neighborhood and its corresponding categories are extracted. One Hot encoding method is also used to separate each individual category.

	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	Arbutus-Ridge	Burger Joint	Bakery	Seafood Restaurant	Dance Studio	Sandwich Place	Lounge	Liquor Store	Event Space	Discount Store	Fast Food Restaurant
1	Downtown	Hotel	Café	Coffee Shop	Sandwich Place	Seafood Restaurant	Dessert Shop	Steakhouse	Bookstore	Concert Hall	Clothing Store
2	Dunbar-Southlands	Sushi Restaurant	Ice Cream Shop	Italian Restaurant	Indian Restaurant	Coffee Shop	Ethiopian Restaurant	Filipino Restaurant	Fast Food Restaurant	Farmers Market	Falafel Restaurant
3	Fairview	Coffee Shop	Park	Asian Restaurant	Japanese Restaurant	Sandwich Place	Malay Restaurant	Salon / Barbershop	Restaurant	Camera Store	Sushi Restaurant
4	Grandview-Woodland	Coffee Shop	Italian Restaurant	Japanese Restaurant	Sushi Restaurant	Café	Park	Indian Restaurant	Pizza Place	Bakery	Record Shop
5	Hastings-Sunrise	Vietnamese Restaurant	Gas Station	Inn	Park	Coffee Shop	Food Truck	Pharmacy	Event Space	Bakery	Sushi Restaurant
6	Kensington-Cedar Cottage	Chinese Restaurant	Coffee Shop	Bus Stop	Filipino Restaurant	Vietnamese Restaurant	Sandwich Place	Convenience Store	Ice Cream Shop	Pizza Place	Café

Figure 2: Parts of Common Venue Table based on Neighborhood.

K means approach is a unsupervised machine learning algorithm. The tables are organized and grouped together based on population and the frequency of the common venues in each neighborhood for the machine K means clustering analysis. The optimum number of cluster is determined based on the common venues and the population of each neighborhood. A graph is generated showing the optimum number of cluster.

	American Restaurant	Art Gallery	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	Australian Restaurant	BBQ Joint	Bagel Shop	Bakery	Bank	Bar	Beach	Bistro	Board Shop
0	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.071429	0.000000	0.000000	0.000000	0.000000	0.000000
1	0.010000	0.010000	0.000000	0.000000	0.000000	0.01	0.000000	0.000000	0.000000	0.000000	0.010000	0.000000	0.000000	0.000000
2	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
3	0.000000	0.000000	0.000000	0.076923	0.000000	0.00	0.038462	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
4	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.014286	0.000000	0.028571	0.014286	0.000000	0.000000	0.000000	0.000000
5	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.076923	0.000000	0.000000	0.000000	0.000000	0.000000
6	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
7	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.066667	0.000000	0.000000	0.000000	0.000000	0.000000

Figure 3: Frequency of Common Venue Table based on Neighborhood.

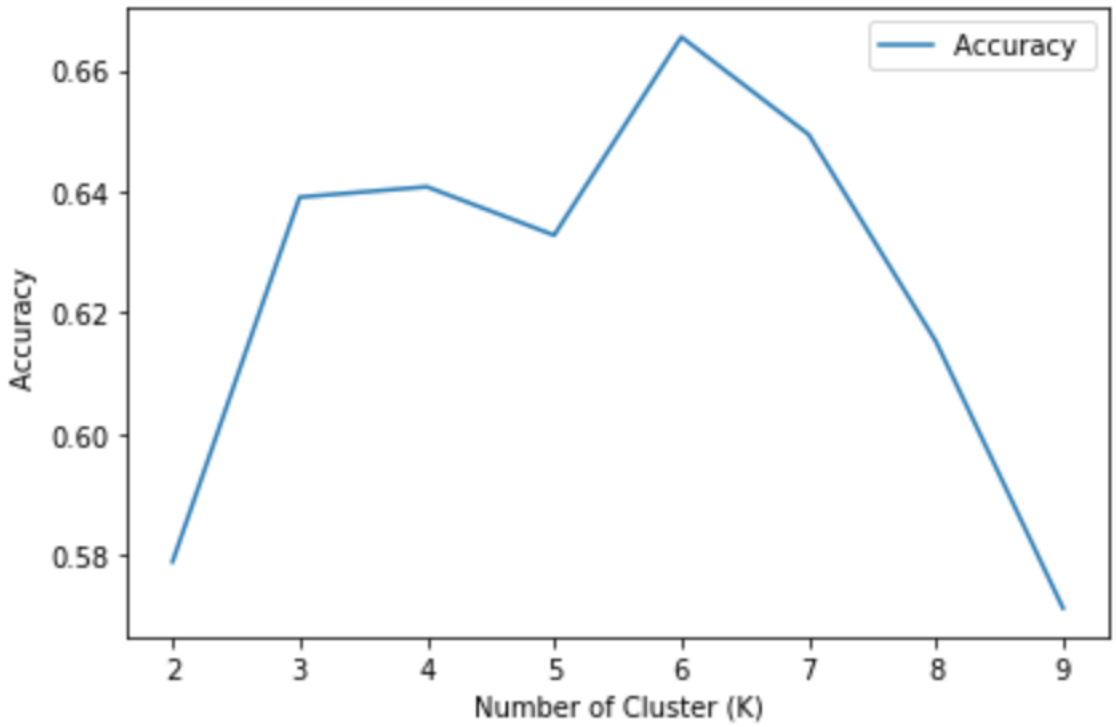


Figure 4: Graph of number of clusters for the best accuracy.

Results

The K clustering approach suggests to use 6 clusters based on the common venues and the population of the neighborhood. A map of each cluster and neighborhood is generated to show the differences.

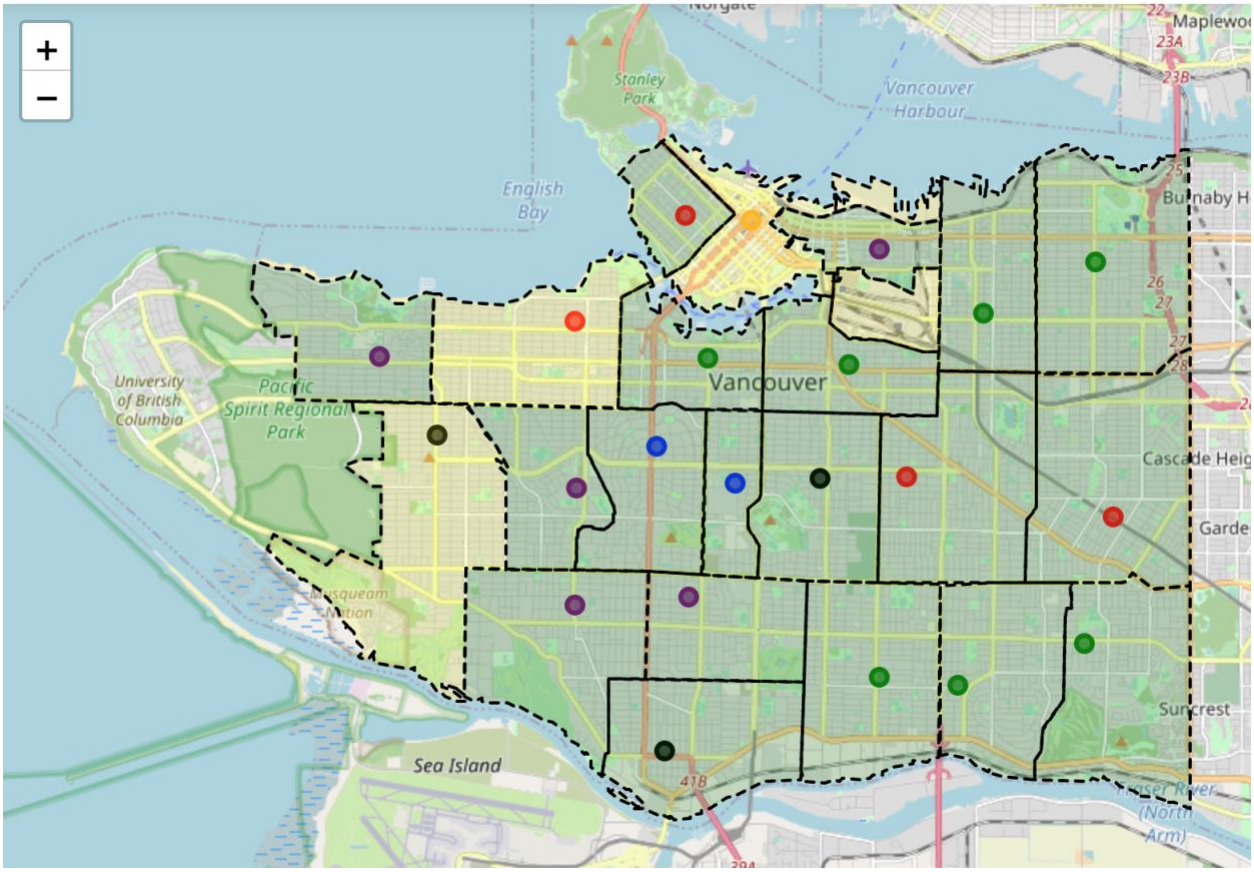


Figure 5 Map of clusters with neighborhood borders.

Each cluster is organized by its common venues and is shown below.

	Neighborhood	Population	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	10th Most Common Venue
6	Kensington-Cedar Cottage	49325.0	0	Chinese Restaurant	Coffee Shop	Bus Stop	Filipino Restaurant	Vietnamese Restaurant	Sandwich Place	Convenience Store	Ice Cream Shop	Pizza Place	Café
9	Kitsilano	43045.0	0	Bakery	American Restaurant	Coffee Shop	Japanese Restaurant	Sushi Restaurant	Ice Cream Shop	French Restaurant	Food Truck	Thai Restaurant	Restaurant
13	Renfrew-Collingwood	51530.0	0	Bus Stop	Metro Station	Bar	Bus Station	Park	Flower Shop	Filipino Restaurant	Fast Food Restaurant	Farmers Market	Falafel Restaurant
20	West End	47200.0	0	Bakery	Japanese Restaurant	Greek Restaurant	Sushi Restaurant	Dessert Shop	Farmers Market	Breakfast Spot	Restaurant	Ramen Restaurant	Bubble Tea Shop

Figure 6 Cluster label 0 (Red Color)

	Neighborhood	Population	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	10th Most Common Venue
2	Dunbar-Southlands	21425.0	1	Sushi Restaurant	Ice Cream Shop	Italian Restaurant	Indian Restaurant	Coffee Shop	Ethiopian Restaurant	Filipino Restaurant	Fast Food Restaurant	Farmers Market	Falafel Restaurant
10	Marpole	24460.0	1	Sushi Restaurant	Bubble Tea Shop	Dessert Shop	Vietnamese Restaurant	Pizza Place	Chinese Restaurant	Dim Sum Restaurant	Falafel Restaurant	Café	Bus Stop
14	Riley Park	22555.0	1	Restaurant	Japanese Restaurant	Coffee Shop	Café	Pub	Chinese Restaurant	Sushi Restaurant	Thai Restaurant	Lounge	Grocery Store

Figure 7 Cluster label 1 (Black Color)

	Neighborhood	Population	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	10th Most Common Venue
15	Shaughnessy	8430.0	2	French Restaurant	Bus Stop	Park	Ethiopian Restaurant	Flower Shop	Filipino Restaurant	Fast Food Restaurant	Farmers Market	Falafel Restaurant	Event Space
16	South Cambie	7970.0	2	Coffee Shop	Gift Shop	Cantonese Restaurant	Bus Stop	Park	Vietnamese Restaurant	Malay Restaurant	Grocery Store	Sushi Restaurant	Bank

Figure 8 Cluster label 2 (Blue Color)

	Neighborhood	Population	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	10th Most Common Venue
3	Fairview	33620.0	3	Coffee Shop	Park	Asian Restaurant	Japanese Restaurant	Sandwich Place	Malay Restaurant	Salon / Barbershop	Restaurant	Camera Store	Sushi Restaurant
4	Grandview-Woodland	29175.0	3	Coffee Shop	Italian Restaurant	Japanese Restaurant	Sushi Restaurant	Café	Park	Indian Restaurant	Pizza Place	Bakery	Record Shop
5	Hastings-Sunrise	34575.0	3	Vietnamese Restaurant	Gas Station	Inn	Park	Coffee Shop	Food Truck	Pharmacy	Event Space	Bakery	Sushi Restaurant
8	Killarney	29325.0	3	Pool	Italian Restaurant	Track	Gym	Deli / Bodega	Dessert Shop	Dance Studio	Dim Sum Restaurant	Diner	Discount Store
11	Mount Pleasant	32955.0	3	Coffee Shop	Diner	Breakfast Spot	Sushi Restaurant	Sandwich Place	Hotel	Thrift / Vintage Store	Lounge	Brewery	Indian Restaurant
18	Sunset	36500.0	3	Cosmetics Shop	Dessert Shop	Home Service	Ethiopian Restaurant	Flower Shop	Filipino Restaurant	Fast Food Restaurant	Farmers Market	Falafel Restaurant	Event Space
19	Victoria-Fraserview	31065.0	3	Convenience Store	Gas Station	Fast Food Restaurant	Sandwich Place	Pizza Place	Deli / Bodega	Dessert Shop	Dim Sum Restaurant	Diner	Discount Store

Figure 9 Cluster label 3 (Green Color)

	Neighborhood	Population	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	10th Most Common Venue
0	Arbutus-Ridge	15295.0	4	Burger Joint	Bakery	Seafood Restaurant	Dance Studio	Sandwich Place	Lounge	Liquor Store	Event Space	Discount Store	Fast Food Restaurant
7	Kerrisdale	13975.0	4	Bus Stop	Chinese Restaurant	Grocery Store	Bakery	Café	Pizza Place	Bubble Tea Shop	Gift Shop	Pharmacy	Gym Pool
12	Oakridge	13030.0	4	Israeli Restaurant	Sushi Restaurant	Pharmacy	Sandwich Place	Café	Fast Food Restaurant	Convenience Store	Vietnamese Restaurant	Gym Pool	Farmers Market
17	Strathcona	12585.0	4	Coffee Shop	Café	Sandwich Place	Park	Vietnamese Restaurant	Brewery	Deli / Bodega	Pub	Cheese Shop	Restaurant
21	West Point Grey	13065.0	4	Pool	Optical Shop	Watch Shop	Filipino Restaurant	Fast Food Restaurant	Farmers Market	Falafel Restaurant	Event Space	Ethiopian Restaurant	Electronics Store

Figure 10 Cluster label 4 (Purple Color)

	Neighborhood	Population	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	10th Most Common Venue
1	Downtown	62030.0	5	Hotel	Café	Coffee Shop	Sandwich Place	Seafood Restaurant	Dessert Shop	Steakhouse	Bookstore	Concert Hall	Clothing Store

Figure 11 Cluster label 5 (Orange Color)

Discussion:

By looking at the cluster data mentioned above, Cluster 0 (Red Color) have lots of bakery business and coffee shops and bus stops in those areas. Cluster 1 (Black Color) have lots of sushi restaurant and ice cream shops. Cluster 2 (Blue Color) and cluster 4 (Purple Color)consists a variety of business. Cluster 3 (Green Color) has more coffee shops and recreational centers for activities. Cluster 5 is a standalone cluster with Hotel business being the most common venues. It is because Downtown Vancouver is the most vibrant part of the cities with lots of tourist. From the business perceptive, business owner can determine the location base on the common venues (competitions) and the populations around the neighbourhood.

Conclusion

This project discusses the process of coming up an approach for business owners to open a business in Vancouver, whether it is a coffee shop or a restaurant. The analysis is performed based on population of each neighborhood of Vancouver. The output of the analysis provided a general view of potential customers and competitions in each neighborhood in Vancouver.