IBM Capstone Project

Report: An Analysis of the Neighbourhoods of the City of Vancouver to Determine the Best Location(s) to Establish a New Restaurant or Bar

Table of Contents

Acknowledgements……2

Abstract……..3

Introduction……3

Data……4

Methodology……...9

Results and Discussion……..15

Conclusion……18

Acknowledgements

My thanks to IBM for running a fascinating nine-course Professional Certificate in Data Science, inspiring a passion and (hopefully!) a career in the field in the process.

Abstract

Vancouver is a city on the west cost of Canada with a busy and competitive restaurant/bar industry. The city is diverse, affluent, passionate about a wide range of food and drinks, and visited frequently by tourists, making it an optimum location to establish a new restaurant or bar. However, different neighbourhoods within the city present different advantages and disadvantages to a prospective restauranteur or publican. This report aims to determine which neighbourhoods in Vancouver are the best locations for a new restaurant or bar using first quantitative analysis through Python and Pandas before drawing conclusions through inductive analysis of the quantitative analysis.

1. Introduction

Vancouver is well a developed Canadian city with a highly competitive services industry. In particular, the restaurant and bar scene is particularly vibrant, if not saturated, and populated generally by chains or independent locations. Though a tough market to get in to, the industry in Vancouver is such that a restaurant or bar can be financially very successful if it gains a foothold in the scene. These services can thrive off the city's remarkably high income, ethnically diverse population, and passionate food/craft beer scene.

This project will attempt to explore patterns of suburbs within Vancouver by categorizing them into clusters in order to identify existing trends within neighborhoods. Recommendations can be made on which category of neighborhood will be most suitable for a certain type of venue to be opened.

The result of this project may be most useful for entrepreneurs in the food and beverage sector given that location can be the deciding factor for a success.

2. Data

To analyze trends in Vancouver's neighbourhoods, the list of neighbourhoods is obtained from the "Local Area Boundary" dataset from opeandata.vancouver.ca. The dataset contained details of all the neighbourhoods within the city limits of Vancouver. It can be accessed at: https://opendata.vancouver.ca/explore/dataset/local-area-boundary/table/?dataChart=eyJxdWVyaWVzIjpbeyJjb25maWciOnsiZGF0YXNldCI6ImxvY2FsLWFyZWEtYm91bmRhcnkiLCJvcHRpb25zIjp7fX0sImNoYXJ0cyI6W3siYWxpZ25Nb250aCI6dHJ1ZSwidHlwZSI6ImNvbHVtbiIsImZ1bmMiOiJDT1VOVCIsInNjaWVudGlmaWNEaXNwbGF5Ijp0cnVlLCJjb2xvciI6IiMwMjc5QjEifV0sInhBeGlzIjoibmFtZSIsIm1heHBvaW50cyI6NTAsInNvcnQiOiIifV0sInRpbWVzY2FsZSI6IiIsImRpc3BsYXlMZWdlbmQiOnRydWUsImFsaWduTW9udGgiOnRydWV9&location=11,49.32602,-123.12412
As the file rendered poorly as a json or csv, I took it as an excel sheet and cleaned it, removing some unwanted columns.

Venue queries will then be made by neighbourhoods using FourSquare APIs. The resulting data regarding venue category will be used to observe commonality between neighbourhoods. The commonality clusters can then provide insight on which type of venue will thrive better on which cluster. K-means clustering algorithm will be used to find pattern between the neighbourhoods.

In summary, the following data is required to conduct the analysis:

The neighbourhoods of Vancouver, the coordinates of these neighbourhoods, trending venues in the area, and venue categories.

This project uses data from the page above. It will be scraped, preprocessed, and finally analyzed in conjunction with FourSquare location data to see the optimum locations to open a bar or restaurant in the city of Vancouver.

Using the dataset, the coordinates for each neighbourhood was established and used for the creation of maps and clustering throughout the analysis.

	MAPID	Neighborhood	Latitude	Longitude
0	AR	Arbutus-Ridge	49.246805	-123.161669
1	CBD	Downtown	49.280747	-123.116567
2	FAIR	Fairview	49.264540	-123.131049
3	GW	Grandview-Woodland	49.276440	-123.066728
4	HS	Hastings-Sunrise	49.277934	-123.040270

Fig. 1. Table showing the details of the first five neighbourhoods in the dataset.

The first map created shows all the city's neighbourhoods.

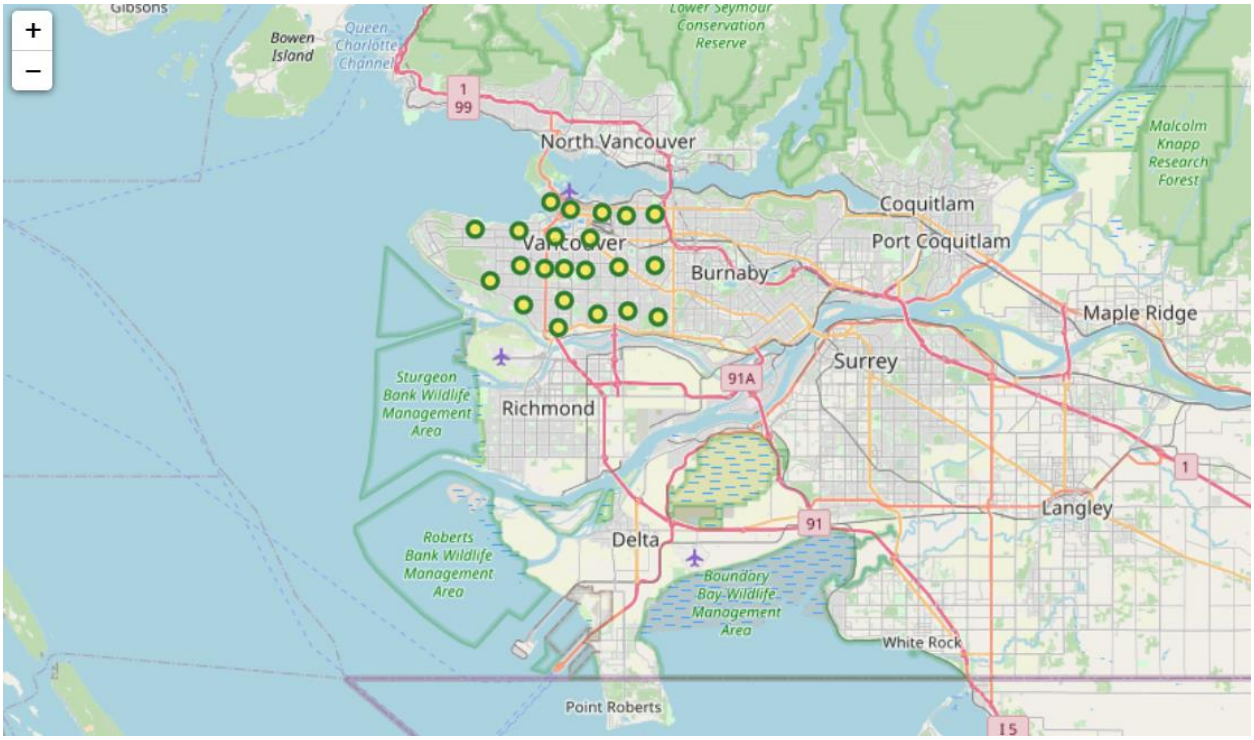

Fig. 2. Rendered map marking the neighbourhoods within the city of Vancouver.

FourSquare API was then integrated to create a function that pulls recommended venues from each neighbourhood. Doing this allows for the examination the distribution of venues in each via a map to see which neighbourhoods would be a good place to open a new one.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Arbutus-Ridge	49.246805	-123.161669	valley bike lane	49.249531	-123.159336	Bike Rental / Bike Share
1	Arbutus-Ridge	49.246805	-123.161669	Bus Stop 51498 (25)	49.247906	-123.167012	Bus Stop
2	Arbutus-Ridge	49.246805	-123.161669	Total Corporate Learning Inc.	49.242441	-123.161609	Business Service
3	Arbutus-Ridge	49.246805	-123.161669	Triangle Park	49.245061	-123.167914	Park
4	Downtown	49.280747	-123.116567	L'Hermitage	49.280139	-123.117480	Hotel
...	...	...	...	...	...	...	...
470	Victoria-Fraserview	49.220012	-123.064135	7-Eleven	49.221115	-123.065350	Convenience Store
471	Victoria-Fraserview	49.220012	-123.064135	East Side Re-Rides	49.219511	-123.066045	Motorcycle Shop
472	Victoria-Fraserview	49.220012	-123.064135	Panago Pizza	49.219231	-123.066215	Pizza Place
473	Victoria-Fraserview	49.220012	-123.064135	Bosley's	49.220926	-123.065382	Pet Store
474	Victoria-Fraserview	49.220012	-123.064135	Lim Kee	49.221016	-123.066017	Asian Restaurant

475 rows × 7 columns

Fig. 3. Table showing details of neighbourhoods and each registered venue.

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Arbutus-Ridge	4	4	4	4	4	4
Downtown	77	77	77	77	77	77
Dunbar-Southlands	6	6	6	6	6	6
Fairview	27	27	27	27	27	27
Grandview-Woodland	35	35	35	35	35	35
Hastings-Sunrise	14	14	14	14	14	14
Kensington-Cedar Cottage	13	13	13	13	13	13
Kerrisdale	4	4	4	4	4	4
Killarney	18	18	18	18	18	18
Kitsilano	55	55	55	55	55	55
Marpole	7	7	7	7	7	7
Mount Pleasant	76	76	76	76	76	76
Oakridge	10	10	10	10	10	10
Renfrew-Collingwood	4	4	4	4	4	4
Riley Park	58	58	58	58	58	58
Shaughnessy	7	7	7	7	7	7
South Cambie	13	13	13	13	13	13
Strathcona	10	10	10	10	10	10
Sunset	1	1	1	1	1	1
Victoria-Fraserview	8	8	8	8	8	8
West End	22	22	22	22	22	22
West Point Grey	6	6	6	6	6	6

Fig. 4. Table displaying the total number of venues in each neighbourhood.

In fig. 4, the number across each row represents how many venues are registered in each neighbourhood. Certain neighbourhoods are primarily residential and are also less accessible vis-a-vis more central neighbourhoods, and so have a low number of venues. We can exclude neighbourhoods from our analysis that have less than 10 venues as these would seem unlikely and unpopular spots to establish venues. The assumption made for neighbourhoods with a low (less than ten) number of venues is that they have low footfall and out-of-neighbourhood visitors and are, therefore, not optimum places to do business. Thus, neighbourhoods with less than ten venues are not considered henceforth.

The number of unique venue categories was then derived. It is worth acknowledging that the goal of this analysis is to find out optimum areas to open a new restaurant or bar, which would fall in to one of the 146 listed categories of venues. The assumption made in not filtering out all other categories is that other

venues are indicative of an area that has activity and has thriving businesses. Moreover, it must be considered that other venues can attract people to a restaurant or bar; for example, someone may decide to go for a meal or a drink after shopping, doing errands, etc. in the neighbourhood. Opening a venue in area with other venues nearby allows for the spontaneous 'walk-in' visitor, rather than relying solely on those who visit the neighbourhood with the preplanned intention of going to a new venue. The map shows neighbourhoods and their venues (marked with green circles). Those neighbourhoods with less than ten venues are displayed but their venues are not.

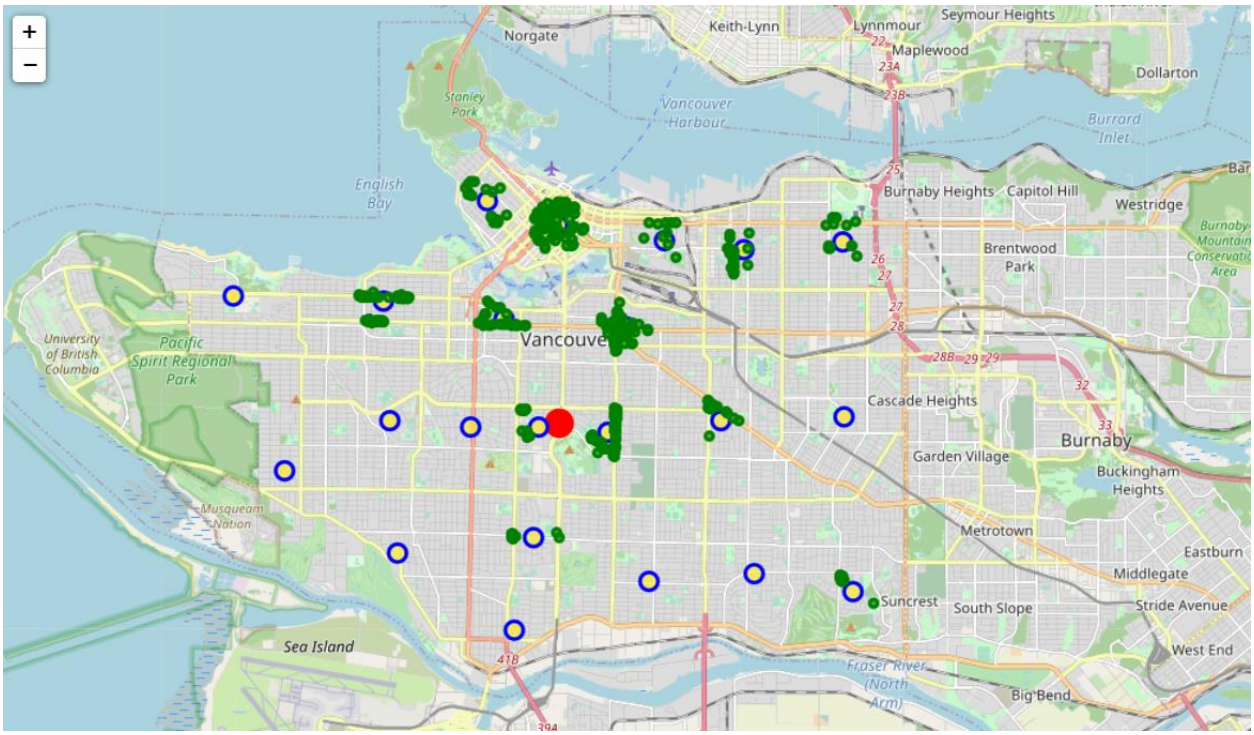

Fig. 5. Map showing neighbourhoods with ten or more venues.

Intuitively, three observations can be made: 1) The neighbourhoods marked that have few (< 10) venues do not have their venues displayed. We can identify these blue markers as poor candidates in which to establish a new restaurant/bar venue. 2) Certain areas of the city have large clusters of venues. For example, if we look at the top of the map, it's apparent that, unsurprisingly, downtown is densely populated with venues. Other areas, such as Mount Pleasant and Riley Park, also seem to be areas of activity. 3) The further away we get from downtown, the less dense the clusters of venues seem to be.

3. Methodology

Since the objective of this analysis is to categorize the neighbourhoods, a K-means clustering algorithm is deployed to categorize each of the neighbourhoods within Vancouver.

A one-hot encoding will be done on the venue dataframe and it will be grouped by neighbourhoods. The encoding will return venue categories as column per neighbourhood, which will then be grouped to provide weighting of venue type occurrence on each neighbourhood.

The encoded dataframe will be further filtered into top venues before the K-means clustering algorithm will be run over it. This will return cluster labels over the neighbourhoods. The clusters will be observed one by one manually to determine its content.

Recommendation will be made based on the clustering.

	Neighborhood	American Restaurant	Art Gallery	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	BBQ Joint	Bagel Shop	Bakery	Bank	...	Theater	Theme Park Ride / Attraction	Thrift / Vintage Store	Tiki Bar	Toy / Game Store	Vegetarian / Vegan Restaurant	Viet Re
4	Downtown	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	
5	Downtown	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	
6	Downtown	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	
7	Downtown	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	
8	Downtown	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	

5 rows × 147 columns

Fig. 6. Table displaying data after one-hot coding methodology is employed.

The occurrence of each venue in the dataframe is then grouped according to neighbourhood. Then, the top five venues in each neighbourhood is derived.

	Neighborhood	American Restaurant	Art Gallery	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	BBQ Joint	Bagel Shop	Bakery	Bank	...	Theater	Theme Park Ride / Attraction	Thrift / Vintage Store	Tiki Bar
0	Downtown	0.000000	0.025974	0.000000	0.000000	0.000000	0.000000	0.000000	0.012987	0.000000	...	0.025974	0.000000	0.000000	0.000000
1	Fairview	0.000000	0.000000	0.037037	0.074074	0.000000	0.037037	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.000000	0.000000
2	Grandview-Woodland	0.000000	0.000000	0.000000	0.028571	0.028571	0.028571	0.000000	0.028571	0.000000	...	0.057143	0.000000	0.000000	0.000000
3	Hastings-Sunrise	0.000000	0.000000	0.071429	0.000000	0.000000	0.071429	0.000000	0.000000	0.000000	...	0.000000	0.142857	0.000000	0.000000
4	Kensington-Cedar Cottage	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.076923	...	0.000000	0.000000	0.000000	0.000000
5	Killarney	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.055556	0.055556	...	0.000000	0.000000	0.000000	0.000000
6	Kitsilano	0.018182	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.036364	0.036364	...	0.000000	0.000000	0.018182	0.000000
7	Mount Pleasant	0.000000	0.000000	0.026316	0.000000	0.000000	0.000000	0.013158	0.000000	0.000000	...	0.000000	0.000000	0.026316	0.000000
8	Oakridge	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.000000	0.000000
9	Riley Park	0.000000	0.017241	0.017241	0.000000	0.017241	0.000000	0.000000	0.000000	0.017241	...	0.000000	0.000000	0.000000	0.017241
10	South Cambie	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.076923	...	0.000000	0.000000	0.000000	0.000000
11	Strathcona	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.000000	0.000000
12	West End	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.045455	0.000000	...	0.000000	0.000000	0.000000	0.000000

13 rows × 147 columns

Fig. 7. Table of the occurrence of each venue in the dataframe grouped by neighbourhood.

```
----Downtown----
                venue  freq
0               Hotel  0.08
1          Restaurant  0.06
2  Seafood Restaurant  0.04
3         Coffee Shop  0.04
4                Café  0.04


----Fairview----
                 venue  freq
0          Coffee Shop  0.11
1    Asian Restaurant  0.07
2  Japanese Restaurant  0.07
3                 Park  0.07
4       Breakfast Spot  0.07
```

Fig. 8. Sample of output from code returning the top five venues in a neighbourhood, with frequency of occurrence.

A function then was created to return the most common venues, along with a new dataframe that displays the top 10 venues for each neighborhood.

	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	Downtown	Hotel	Restaurant	Seafood Restaurant	Café	Coffee Shop	Gastropub	Breakfast Spot	Steakhouse	Sandwich Place	Concert Hall
1	Fairview	Coffee Shop	Park	Asian Restaurant	Breakfast Spot	Japanese Restaurant	Indian Restaurant	Korean Restaurant	Pet Store	Pharmacy	Nail Salon
2	Grandview-Woodland	Coffee Shop	Pizza Place	Indian Restaurant	Vegetarian / Vegan Restaurant	Theater	Deli / Bodega	Brewery	Park	Pub	Record Shop
3	Hastings-Sunrise	Theme Park Ride / Attraction	Pizza Place	Gun Shop	Farmers Market	Bridal Shop	Event Space	Gas Station	Office	Portuguese Restaurant	Café
4	Kensington-Cedar Cottage	Vietnamese Restaurant	Indian Restaurant	Bank	Burger Joint	Sandwich Place	Café	Breakfast Spot	Supermarket	Seafood Restaurant	Bus Stop

Fig 9. Table showing the top ten venues in each neighbourhood.

Neighbourhoods were then clustered using k-means coding. A new dataframe with the clusters was created that merged the first dataframe (containing neighbourhood details) with the ten most common venues in each. The clusters were then rendered in another folium map.

	MAPID	Neighborhood	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 M Comm Ve
0	CBD	Downtown	49.280747	-123.116567	2	Hotel	Restaurant	Seafood Restaurant	Café	Coffee Shop	Gastropub	Breakfast Spot	Steakhouse	Sandw P
1	FAIR	Fairview	49.264540	-123.131049	2	Coffee Shop	Park	Asian Restaurant	Breakfast Spot	Japanese Restaurant	Indian Restaurant	Korean Restaurant	Pet Store	Pharm
2	GW	Grandview-Woodland	49.276440	-123.066728	2	Coffee Shop	Pizza Place	Indian Restaurant	Vegetarian / Vegan Restaurant	Theater	Deli / Bodega	Brewery	Park	
3	HS	Hastings-Sunrise	49.277934	-123.040270	5	Theme Park Ride / Attraction	Pizza Place	Gun Shop	Farmers Market	Bridal Shop	Event Space	Gas Station	Office	Portugu Restau
4	RP	Riley Park	49.244766	-123.103147	2	Japanese Restaurant	Coffee Shop	Vietnamese Restaurant	Farmers Market	Skating Rink	Restaurant	Sporting Goods Shop	Café	Chir Restau

Fig. 10. Dataframe containing clusters, neighbourhood details, and most top ten common venues.

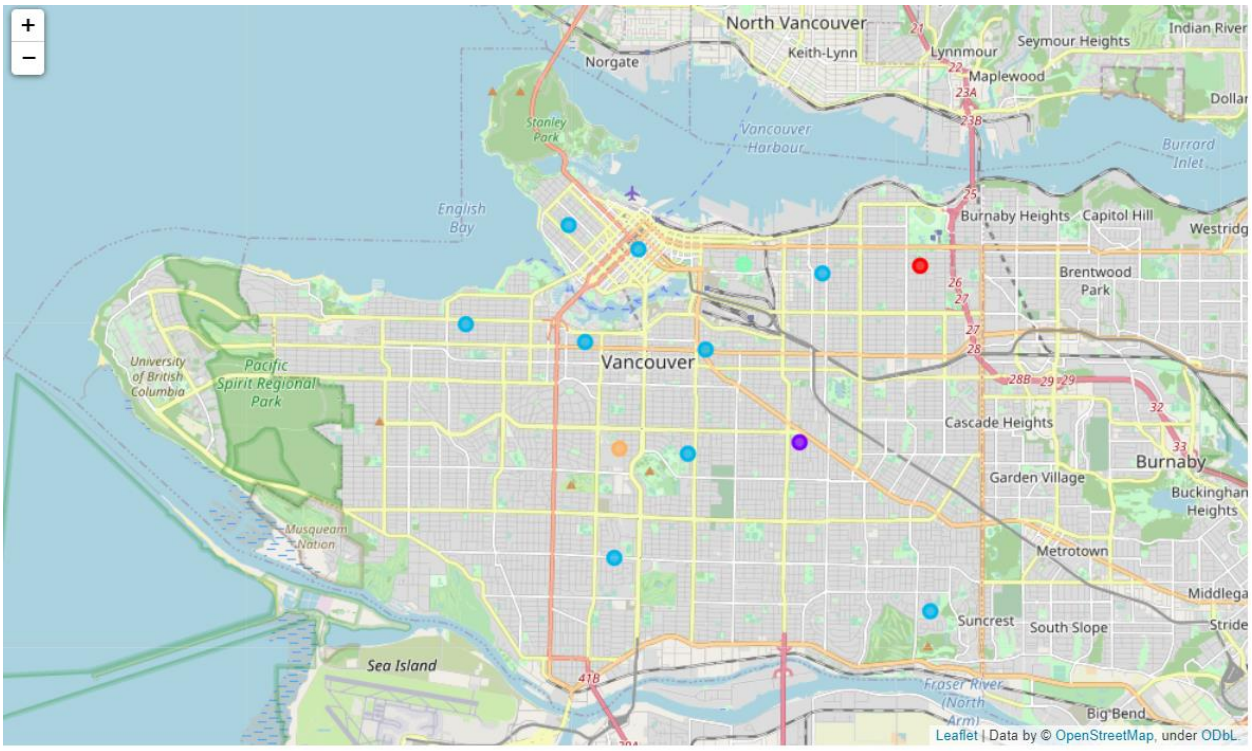

Fig. 11. Map showing clustered neighbourhoods. Key: Purple markers = cluster 1, blue markers = cluster 2, green markers = cluster 3, orange markers = cluster 4, red marker = cluster 5.

The final step of the analysis was to manually observe each cluster and derive recommendations.

Cluster 1:

	MAPID	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
7	KC	1	Vietnamese Restaurant	Indian Restaurant	Bank	Burger Joint	Sandwich Place	Café	Breakfast Spot	Supermarket	Seafood Restaurant	Bus Stop

Fig. 12. Table showing cluster details of Cluster 1.

It is observable that cluster 1 has a high incidence of Vietnamese and Indian restaurants - these could be good types of restaurants to open. It appears that these cuisines are popular, and so another one may cater to local taste. Moreover, restaurants/bars seem to do well here; 7 out 10 of the most common venues are places to dine in. However, the cluster one consists only of one neighbourhood (MAPID: KC = Kensington-Cedar Cottage), so it has less venues and activity. This maps with the cluster being quite far

away from Downtown and closer to residential suburbs. There may be other clusters that are busier and therefore better places to open a restaurant/bar.

Cluster 2:

	MAPID	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	CBD	2	Hotel	Restaurant	Seafood Restaurant	Café	Coffee Shop	Gastropub	Breakfast Spot	Steakhouse	Sandwich Place	Concert Hall
1	FAIR	2	Coffee Shop	Park	Asian Restaurant	Breakfast Spot	Japanese Restaurant	Indian Restaurant	Korean Restaurant	Pet Store	Pharmacy	Nail Salon
2	GW	2	Coffee Shop	Pizza Place	Indian Restaurant	Vegetarian / Vegan Restaurant	Theater	Deli / Bodega	Brewery	Park	Pub	Record Shop
4	RP	2	Japanese Restaurant	Coffee Shop	Vietnamese Restaurant	Farmers Market	Skating Rink	Restaurant	Sporting Goods Shop	Café	Chinese Restaurant	Lounge
6	WE	2	Café	Gay Bar	Farmers Market	Coffee Shop	Diner	Lingerie Store	Grocery Store	Sandwich Place	Bakery	Noodle House
8	MP	2	Coffee Shop	Diner	Breakfast Spot	Sandwich Place	Sushi Restaurant	Lounge	Brewery	Indian Restaurant	Vietnamese Restaurant	Thrift / Vintage Store
9	OAK	2	Bubble Tea Shop	Light Rail Station	Pizza Place	Vietnamese Restaurant	Bus Station	Coffee Shop	Sporting Goods Shop	Sandwich Place	Fast Food Restaurant	Sushi Restaurant
10	KIL	2	Coffee Shop	Juice Bar	Shopping Mall	Sandwich Place	Salon / Barbershop	Recreation Center	Liquor Store	Sushi Restaurant	Farmers Market	Chinese Restaurant
11	KITS	2	Coffee Shop	Pizza Place	Japanese Restaurant	Wine Shop	Bakery	Bank	Toy / Game Store	Food Truck	Optical Shop	Grocery Store

Fig. 13. Table showing cluster details of Cluster 2.

Cluster 2 consists of 10 neighbourhoods that may be suitable candidates for opening our new restaurant/bar. All the neighbourhoods contained within are known as areas popular with restaurant-goers. Importantly, the top 6 clusters in the north of the map are densely populated and known cultural areas, suggesting there is already a culture of places to eat and drink at - although this could mean these areas are competitive and a new venture could struggle to become established. On the other hand, most are located close to Downtown or are situated on good transport lines. We can see that all have a good number of bars or restaurants within them, along with a selection of other businesses that might draw hungry/thirsty customers. Moreover, there is a good distribution of different types of restaurants and bars, suggesting any style, from a cafe to a vegan spot, could thrive.

Cluster 3:

	MAPID	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
5	STR	3	Park	Deli / Bodega	Cheese Shop	Food Truck	Brewery	Sandwich Place	Restaurant	Pub	Coffee Shop	Dive Bar

Fig. 14. Table showing cluster details of Cluster 3.

Cluster 3 consists of one neighbourhood (Strathcona). It is known as a hub for brewing and the craft beer scene in the city. Although there are bars/pubs and some venues to grab a quick bite to eat, it appears to be lacking in a range of sit-down restaurants. On the other hand, this area is quite industrial, is not the most well-linked area by public transport, and lacks a diversity of cultural venues that would attract visitors and customers. Venues here may also face competition from the nearby downtown area in Cluster 2, which boasts well established bars/restaurants and other venues.

Cluster 4:

	MAPID	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
12	SC	4	Coffee Shop	Café	Vietnamese Restaurant	Malay Restaurant	Cantonese Restaurant	Grocery Store	Bank	Liquor Store	Gift Shop	Sushi Restaurant

Fig. 15. Table showing cluster details of Cluster 4.

Cluster 4 contains one neighbourhood (South Cambie). It appears to have a mix of practical amenities such as banks, grocery stores, etc., tracking with its status as a fairly residential area. We can also see that 4 of the ten most common venues are Asian-style cuisines. Cambie and much of the south of the city has an ethnically diverse population, especially Asian immigrants and their descendants. These restaurants could be common as they are catering to local taste. A restaurant of a different style could face challenges cutting in to this market. On the other hand, bars do not seem to be common here, which could represent an opportunity for one to set up.

Cluster 5:

	MAPID	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	HS	5	Theme Park Ride / Attraction	Pizza Place	Gun Shop	Farmers Market	Bridal Shop	Event Space	Gas Station	Office	Portuguese Restaurant	Café

Fig. 16. Table showing cluster details of Cluster 5.

Cluster 5 consists of the Hastings-Sunrise neighbourhood, an area that presents opportunities and challenges for a new restaurant or bar. On one hand, the area has a theme parks (namely the PNE) which draws out-of-town visitors, and restaurants and bars seem to be not common. The area is also very close to downtown. However, it is the accessibility of downtown which may draw guests away from the area to more established areas. Moreover, Hastings-Sunrise has perhaps the highest levels of social deprivation in the city and a reputation it has struggled to shake off. A new restaurant or bar may be welcome here but may struggle to gain traction owing to the area's reputation, relative lack of venue diversity, and competition from venues in the downtown area.

4. Results and Discussion

As discussed above, there are several factors to consider when opening a restaurant/bar in Vancouver city. Groupings as a result of K-means clustering algorithm tallies with demographic and social trends in the city and we can make our recommendations based of these and induction from experience living and working in the city. Ranked in order of suitability for a new establishment, the optimum clusters in which to set up a new restaurant are:

1. Cluster 2
2. Cluster 3
3. Cluster 1
4. Cluster 4
5. Cluster 5

Much of the reasons for the suitability of each cluster have been discussed above. However, more detailed analysis can be derived:

Cluster 5, the Hastings-Sunrise neighbourhood, appears the least suited for a new bar/restaurant as its lack of venues of cultural/social significance will fail to bring in many visitors to the neighbourhood itself, though a busy summer's day at the PNE theme park may preclude this. Moreover, Cluster 5's the proximity to downtown and even other neighbourhoods in Cluster 2 will likely mean both locals and tourists will choose to visit bars/restaurants elsewhere. In addition, true or not, the reputation the area has as one of social and economic deprivation will drive away visitors and, therefore, prospective customers. For example, tourists arriving in the downtown area of the city are often encouraged by city residents to avoid what is colloquially known as 'Eastside'.

Cluster 4, consisting of the South Cambie neighbourhood, is seemingly lacking in restaurants and some particular cuisines, and the ethnically diverse population in a residential area may attract customers to a new and exciting establishment. A well-planned venture focusing on local residents could be a hit, especially among those who would like cuisines not frequently on offer. However, I believe it remains a relatively poor choice because the area is primarily residential, relatively far away from the downtown core, and lacks other venues to attract visitors. A new establishment here should not expect visitors from out of the neighbourhood or indeed the city, reducing the chance of spontaneous walk-in customers. An establishment here will have to, therefore, gain a formidable reputation to attract guests from other areas of the city. However, as noted, doing so could give it a large local clientele.

Cluster 1, consisting of the Kensington-Cedar Cottage neighbourhood, is a mid-tier candidate for opening a new restaurant/bar. Geographically, it faces challenges as it is situated relatively far away from the downtown core and its surrounding neighbourhoods and more towards the suburbia of the south and east of the city. However, while restaurants are prevalent venues (7 out of 10 of the most common venues are restaurants and cafes) , they do not totally dominate, suggesting that these establishments can be successful and that there may be space for more. The results of the most common venues analysis also shows that there is a myriad of other common venues, such as banks, grocery stores, etc. that suggest that the area has activity and functions more than just an area for people to live before they commute elsewhere.

Cluster 3, the Strathcona neighbourhood, presents several opportunities for a new restaurant/bar. Doing so will encounter some of the problems characteristic of the Eastside of the city that Cluster 5 also faces. Cluster 3 is located directly adjacent to the struggling 'Downtown Eastside' area and the area itself is

industrial. However, the area is still bustling, with a reputation for having many famous breweries that attract visitors from all over the city. A restaurant or gastropub could succeed here by tapping in to this market and using its proximity to downtown as a draw for visitors.

Cluster 2 appears to contain the best areas to establish a new bar or restaurant. All but one of the neighbourhoods within it are near the downtown area, which are known as places to visit, eat, and drink. The only exception to this is one neighbourhood in the cluster which is on the city boundary line, but is close to another city that begins where Vancouver ends. Neighbourhood clusters such as Downtown, Kitsilano, and Riley Park all encompass bustling areas of economic and social activity that residents and tourists from around the city will come to visit based on the reputation of these areas and their venues. As a result, it is unsurprising to see that all clusters have a good distribution of restaurants and bars and, crucially, other venues that bring customers and visitors. Choosing one of these neighbourhoods, such as the West End (MAP ID: WE), that has a high number of venues for a "casual bite" but not sit-down restaurants or bars could prove successful for meeting demand in a busy area. Although the scene in all of the neighbourhoods in this cluster will undoubtedly be very competitive, the distribution of restaurants and bars suggest they can be successful in areas that rely on their reputation as areas to explore, eat, and drink to draw customers and visitors.

In terms of the data analysis, having reviewed the code used, it seems there may be some issues regarding how FourSquare API registers and venues and decides how frequent venues are in a given area. For example, many known venues in the city do not appear on this map. More importantly, however, a minority of the venues returned in most common venues appear to be incorrect. For example, Cluster 2 shows that theme parks and gun shops occupy the first and second most common venues respectively in the area. However, this is demonstrably incorrect; Cluster 2 has many more cafes and restaurants than gun shops, for instance, while there is only one (though significant) theme park in Vancouver, the PNE. It is not plausible to say that there are more theme parks here than gas stations, bus stops, offices and other venues that are considered by the API. Although I do not believe this significantly impacted my analysis, it does appear to be a problem that could negatively skew less inductive future analyses.

5. Conclusion

Based on the analysis above, I conclude that a Cluster 2 contains the best locations for a prospective new restaurant or bar, owing to reasons of reputation, proximity to the city's downtown area, and precedent of other similar, successful venues.