Marketing Opportunity Analysis for Vancouver

1. Introduction

1.1. Background

Vancouver

Vancouver is a coastal seaport city in western Canada, located in the Lower Mainland region of British Columbia. Classed as a Global Beta City, Vancouver is consistently named as one of the top five worldwide cities for liveability and quality of life.
In recent years, Vancouver has become a centre for software development, biotechnology, aerospace, video game development, animation studios and television production and film industry. The city's strong focus on lifestyle and health culture also makes it a hub for many lifestyle brands with Lululemon, Arc'teryx, Kit and Ace, Mountain Equipment Co-op, Herschel Supply Co., Aritzia and so on.

Marketing Opportunity

For this project, I have classified marketing opportunity into two aspects based on the business need and presence,

a) **Existing**
   Existing opportunities are with respect to businesses that already have a presence in Vancouver and are looking at locations to focus their marketing efforts.
   e.g. A health & lifestyle company looking for places to launch products, conduct community events

b) **New**
   New opportunities are with respect to business that have no presence and are looking to expand in Vancouver. They also include potential locations for existing business to expand.
   e.g. A restaurant chain looking to expand into Vancouver

1.2. Business Problem

As a Marketing consultant, I have considered the most common marketing strategy requirements that businesses and municipalities would face in a location stressed city such as Vancouver,

- Where do we launch our products?
- Where do we conduct community events?
- Where do we expand our business into?

The underlying business need is to identify the geographical locations in Vancouver based on a company's focus area where marketing tactics would be most effective. Effective use of marketing budget is a critical need for businesses in order to ensure positive ROI and business growth.

1.3. Target Audience

The target audience for this analysis would be ***Chief Marketing Officers, Marketing directors and other marketing leaders to Campaign managers who are responsible for strategy and execution of marketing campaigns in Vancouver***.
This analysis would provide a data driven guide to decide on potential opportunities for marketing. It can be further refined with first party data to increase the level of accuracy and gather further insights.

2. Data Acquisition and Cleaning

2.1. Data sources

The following open data sources and API's were used for this project,

i. Geographical coordinates such as latitude and longitude by Postal Codes for Canada sourced from Geonames
ii. Foursquare developer API was used to pull information on venues and coordinates
iii. Geocoder API was used to pull coordinates for Vancouver

2.2. Data clean up

The data sourced from Geonames included the entire dataset for Canada. This was converted to a csv file and cleaned up. The data was sliced down to Vancouver as the city of concern. The columns such as province, province code, admin areas and accuracy analysis were dropped as they were redundant. The neighborhood data was edited for consistency by removing extraneous characters and indexed.
The dataset had **44** rows or neighborhoods split by postal codes.

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

The data pulled from Foursquare API included the venue name, venue category and location coordinates. This did not require any clean up.
The data pulled from the Geocoder did not require any clean up.

2.3. Feature selection

The features selected for this analysis were venue category, latitude and longitude of venue for each neighborhood (defined by postal code). The individual venues were considered by their venue category and grouped by the neighborhood limited to a radius of 100 m from the latitude and longitude of the neighborhood**.** The number of venues were limited to 100.
The total size of the dataset was 854 rows by 7 features.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	North Hastings-Sunrise	49.2807	-123.0397	The Fair at the PNE	49.282971	-123.042109	Fair
1	North Hastings-Sunrise	49.2807	-123.0397	Wooden Roller Coaster	49.281744	-123.035128	Theme Park Ride / Attraction
2	North Hastings-Sunrise	49.2807	-123.0397	Livestock Barns	49.284037	-123.039278	Farm
3	North Hastings-Sunrise	49.2807	-123.0397	Playland	49.281924	-123.036258	Theme Park
4	North Hastings-Sunrise	49.2807	-123.0397	Hastings Community Centre	49.280778	-123.039176	Event Space

Focus Area

Focus area lists were created based on the unique venue categories collected for Vancouver. The rationale behind the focus areas being that businesses would be interested in certain types of categories which would hold more relevance to them. This distribution was done based on the industry and associated activities with that field.
The lists are non-exclusive and hard coded in this iteration.
The following lists were hard coded and created,

Focus Area	Venue categories	Target Audience
Food & Beverages	86	Restaurant chains, Cafes e.g. Tim Hortons, Subway
Shopping	39	Retail stores e.g. 7-Eleven, Artizia
Healthy Living	19	Health focused apparel companies e.g. Lululemon, MEC
Entertainment	26	Theatre companies, record labels e.g. 1080p, 604 records
Services	18	Civic services e.g. City of Vancouver

Example of Healthy Living list = 'Park', 'Gym', 'Gym / Fitness Centre', 'Skating Rink', 'Lake', 'Beach', 'Field', 'Sporting Goods Shop', 'Soccer Stadium', 'Spa', 'Hockey Arena', 'Athletics & Sports', 'Yoga Studio', 'Trail', 'Tennis Court', 'Baseball Field', 'Ski Trail', 'Ski Chairlift', 'Track', 'Health Food Store'

3. Methodology

3.1. Approach overview

First, I collected the geographical location for all the neighborhoods as defined by their postal codes in Vancouver city. For this data, I collected the nearby venues along with their geographical location data and prepared a master data set.
Second, I looked at the unique venue categories available for the entire data set. I created multiple focus area lists which were used to split the master data set into subsets.
Third, these subsets were hot encoded and clustered using K-means clustering into **5** clusters.
Fourth, the clusters were analysed in detail and viable marketing opportunities were provided for the considered focus area.

3.2. Exploratory Data Analysis

A total of 843 venues were considered across 40 neighborhoods and 188 unique categories.
The distribution of the venues considered by focus area is given below,

Focus Area	Total Venues considered
Food & Beverages	560
Shopping	117
Healthy Living	72
Entertainment	61
Services	36

The most common venues by neighborhood were identified and collected.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Bentall Centre	American Restaurant	Gastropub	Irish Pub	Breakfast Spot	Filipino Restaurant
1	Central Kitsilano	Coffee Shop	Pub	Pizza Place	Café	Wine Shop
2	Dunbar-Southlands / Musqueam	Vietnamese Restaurant	Fast Food Restaurant	Wine Shop	Filipino Restaurant	Deli / Bodega
3	East Fairview / South Cambie	Coffee Shop	Sushi Restaurant	Malay Restaurant	Vietnamese Restaurant	Liquor Store
4	East Mount Pleasant	Sushi Restaurant	Vietnamese Restaurant	Ethiopian Restaurant	Sandwich Place	Sports Bar

The neighborhoods were analysed by using “one hot” encoding and the data normalized by division with the mean.

	Neighborhood	Athletics & Sports	Beach	Field	Gym	Gym / Fitness Center	Health Food Store	Hockey Arena	Lake	Park	Skating Rink	Ski Chairlift	Ski Trail	Soccer Stadium	Spa	Sporting Goods Shop	Tennis Court	Track	Trail
0	Bentall Centre	0.0	0.0	0.0	1.0	0.00	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.00	0.0	0.0	0.0	0.0
1	Central Kitsilano	0.0	0.0	0.0	0.0	0.25	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.25	0.0	0.0	0.0	0.0
2	Chaldecutt / South University Endowment Lands	0.0	0.0	0.0	0.0	0.00	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.00	0.0	0.0	0.0	0.0
3	East Fairview / South Cambie	0.0	0.0	0.0	0.0	0.00	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.00	0.0	0.0	0.0	0.0
4	East Mount Pleasant	0.0	0.0	0.0	0.0	0.00	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.00	0.0	0.0	0.0	0.0

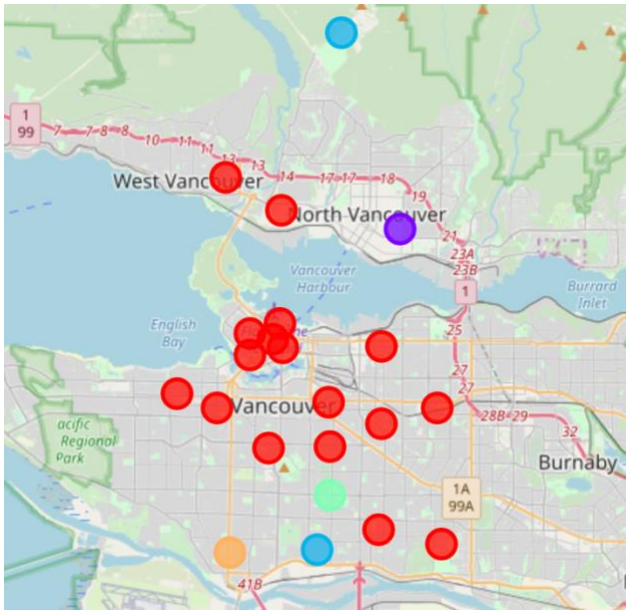
3.3. Machine Learning Techniques

The **K-means clustering algorithm** was then employed on the overall focus area datasets after dropping the ‘Neighborhood’ identification.
K-means clustering is a type of unsupervised learning, which is used when you have unlabelled data (i.e., data without defined categories or groups). The goal of this algorithm is to find groups in the data, with the number of groups represented by the variable K. The algorithm works iteratively to assign each data point to one of K groups based on the features that are provided. Data points are clustered based on feature similarity.

The number of clusters designated was **5**. This was arbitrarily chosen after a few trials.
The geographical coordinates were then matched back to the clustered data and the clusters were plotted on the map of Vancouver and labelled.

4. Results

4.1. Focus Area – Shopping



Cluster Label	Colour	Characteristics
0	Red	Liquor stores, Pharmacies, Grocery stores - Downtown focused
1	Purple	Butcher – Single point*
2	Blue	Sporting goods shop, women's store – Edges of city
3	Green	Cosmetics shop – Single point*
4	Orange	Video store – Single point*

Marketing Opportunities

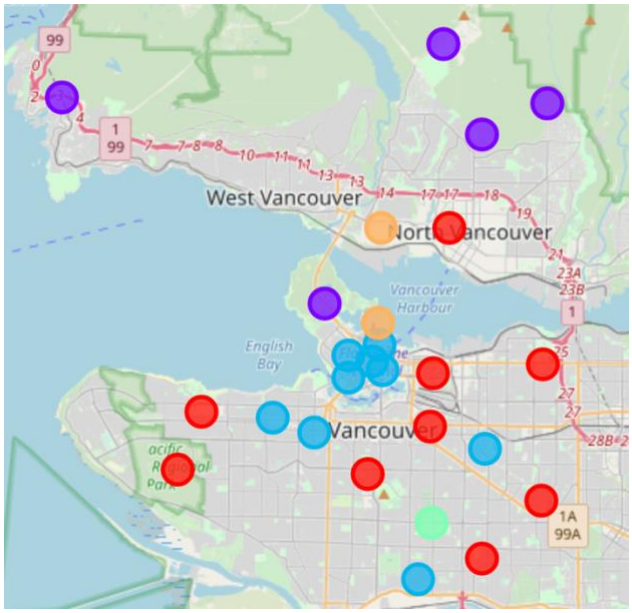
Existing

a) Ad spend targeting buy stage of the consumer journey can be focused on a 7km radius around downtown Vancouver
b) Community based events can be executed in dense cluster areas
c) Competitor influence analysis can be done by analysing store distances

New

a) Possible locations to open pop-up stores include south western and western regions not covered by clusters

4.2. Focus Area – Healthy Living



Cluster Label	Colour	Characteristics
0	Red	Park, Yoga studio, Beach, Field, Gym – Suburbs
1	Purple	Trail, Yoga studio, Park, Beach – North of the city
2	Blue	Spas, Sporting goods store, Yoga Studio – Downtown focused
3	Green	Field – Single point*
4	Orange	Gym, Sporting goods store – Single point*

Marketing Opportunities

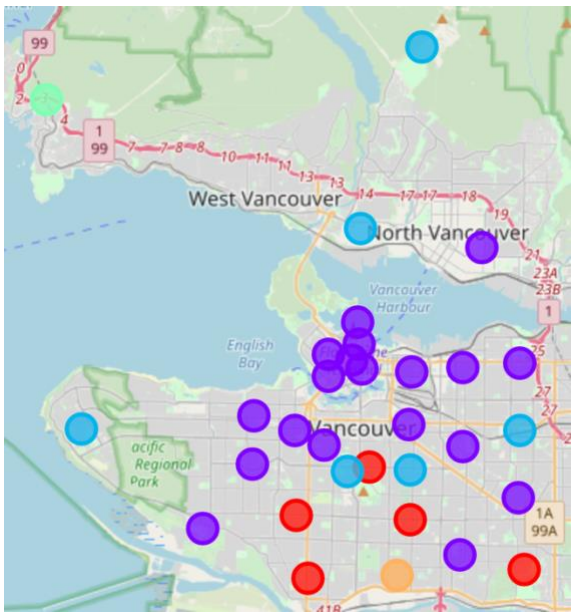
Existing

a) Products could be grouped according to the needs of the cluster in stores falling under the covered region
b) Yoga studios are one of the most popular venue categories across clusters. They are a potential candidate for partnership and affiliate marketing

New

a) Possible locations to conduct community events include southern regions not covered by clusters

4.3. Focus Area – Food & Beverages



Cluster Label	Colour	Characteristics
0	Red	Chinese, Asian and Indian restaurants – Suburbs
1	Purple	Cafes, Coffee shops, Fast food restaurants – Spread across city
2	Blue	Coffee shops, Liquor stores – Educational institutions nearby
3	Green	Tapas restaurant – Single point*
4	Orange	Indian restaurant – Single point*

Marketing Opportunities

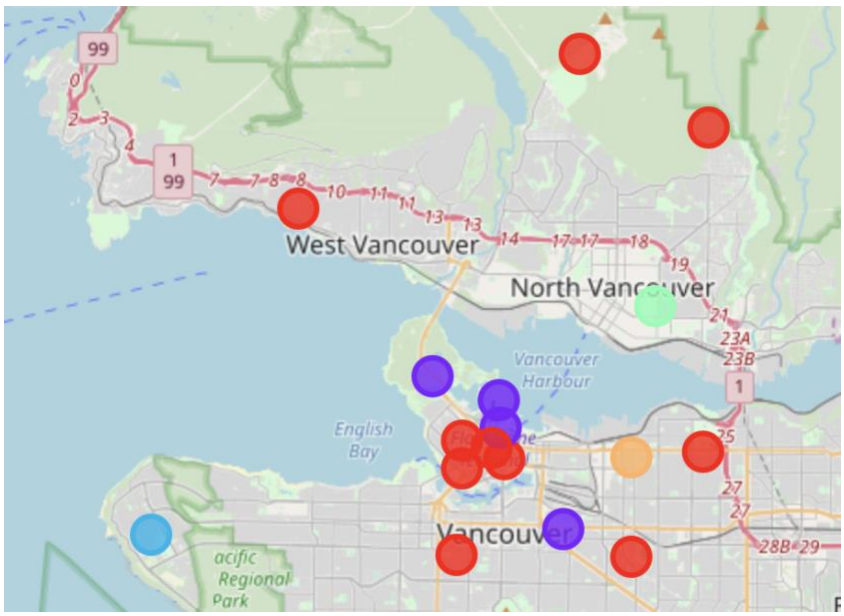
Existing

a) Competitor influence analysis can be done by analysing restaurant distances and similar type density
b) Takeaway type restaurants seem to dominate the city as evidenced by the spread of cluster 1. This could be incorporated in restaurants covered by the cluster not offering the service currently

New

a) Possible locations to open up new restaurants include north western and western regions not covered by clusters
b) Possible locations to provide free tasters of restaurant in the downtown region

4.4. Focus Area – Entertainment



Cluster Label	Colour	Characteristics
0	Red	Theatres, Plaza – North of the city and Downtown focused
1	Purple	Event space, Plaza – Harbour area
2	Blue	Theme Park – Educational institution – Single point*
3	Green	Mountain – Single point*
4	Orange	Theatre – Single point*

Marketing Opportunities

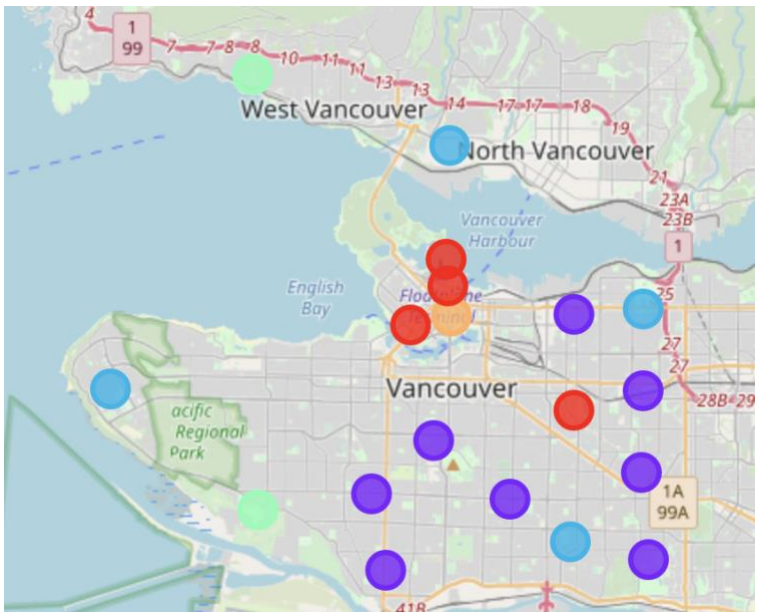
Existing

a) Ad spend targeting buy stage of the consumer journey can be focused on a 3km radius around downtown Vancouver
b) Location specific marketing around the Coal Harbour area would yield higher returns

New

a) Possible locations to conduct community events include regions not covered by clusters

4.5. Focus Area – Services



Cluster Label	Colour	Characteristics
0	Red	Dog Run, Church – Downtown focused
1	Purple	Bus Stop, Railway stations, Ferry – Transport system
2	Blue	Bus Stop, Railway stations, Church – Edges of the city
3	Green	Home service, train station
4	Orange	Trade school – Single point

Marketing Opportunities

Existing

a) Potential locations for broadcasting community messages would be dog parks and churches
b) Transport system covers most of the city and would prove to be an effective brand medium for the city civic services

New

a) Possible locations to conduct community events include regions not covered by clusters

5. Conclusion

This project highlights the opportunities available to define marketing strategy and go-to market plans for companies in the areas of Food & Beverage, Healthy Living, Shopping and Entertainment.

Future Growth

The analysis performed here could be refined with the addition of first party data from the marketing systems of these companies to increase the accuracy and yield further insights. Census data when included would provide more effective buyer persona profiling.
This project could also be enhanced by using various other machine learning techniques and built into a web application which would prove to be an effective market research and planning tool.

6. Notes

(*) – Single point data, the cluster has just one data point which can be considered an outlier
Venues limited to a maximum of 100 based on the API limit for a sandbox account in Foursquare

7. References

Github Link
LinkedIn Profile
Geonames
Foursquare developer API
Geocoder API