Neighbourhood of Vancouver

Abhishek Mahesh Kurhekar
May 10, 2020

1. Introduction

Vancouver is a seaport city in western Canada, located in the Lower Mainland region of British Columbia. As the most populous city in the province, the 2016 census recorded 631,486 people in the city, up from 603,502 in 2011 It is multicultural. It provides a lot of business opportunities and a business friendly environment. It has attracted many different players into the market.

Vancouver has hosted many international conferences and events, and Vancouver's scenic location makes it a major tourist destination. Over 10.3 million people visited Vancouver in 2017. Annually, tourism contributes approximately $4.8 billion to the Metro Vancouver economy and supports over 70,000 jobs.

This means that the market for restaurants is highly competitive. As it is a highly developed city so cost of doing business is also one of the highest. Thus, any new business venture or expansion needs to be analysed carefully. The insights derived from analysis will give a good understanding of the business environment which help in strategically targeting the market. This will help in reduction of risk. And the Return on Investment will be reasonable.

1.1 Business problem

Vancouver has a food culture which includes an array of international cuisines influenced by the city's immigrant history.

There are various types of restaurants but it's tough to determine which place is suitable for which cuisines as it varies by area. So we would like to find out the popular category of business in a particular area.

1.2 Target Audience

The clientele includes any person who wants to open his restaurant of any type in Vancouver, so I only focus on Vancouver city for my analysis. The objective is to recommend to the management which neighborhood of Vancouver city will be the best choice to start business in the area of their choice.

2. Data

2.1 Source of Data

For the project we required various types of data which were not available in ready to use form. So we need to use web mining, API, and libraries.

We need the data about restaurants in various areas of Vancouver, so we would need names of areas of Vancouver, coordinates of those areas and the data about restaurants in those areas.

Names of Area:

The names of area are fetched from the wikipedia page of postal codes : https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_V

Coordinates of Area:

To get the latitude and longitude of the area we used python library 'geopy'.

Data on Restaurants:

We use Foursquare api to fetch data regarding restaurants in a particular area.

2.2 Data Cleaning

Data Collected from the wikipedia was in raw form i.e in html page, so we need to fetch the table containing area names of Vancouver city. The table contains the postal codes for British Columbia province which has various codes allocated for various neighbourhoods in Vancouver city, from which we fetch the names of neighbourhoods of Vancouver city.

After getting coordinates for those neighbourhoods using ‘geopy’ library, we fetch data on venues in the area and filter the data for keyword ‘restaurants’.

When we generate data using api’s we need to make sure that data generated is of good quality as there can be missing values and duplicate data.

So when we generated data for coordinates of neighbourhoods we found that there were areas which had the same coordinates so we selected only one of them.

When fetching the data for restaurants for a neighbourhood we found that same restaurant was also allocated to another neighbourhood, which would result in duplicate data so we allocated the restaurant to the neighbourhood which is nearest to the restaurant.

3. Methodology

To select the best possible type of restaurant in a particular area can be found by finding the most common type of restaurant in that area, as it will be easier to attract the customer from the competitor restaurants that have customers with similar choices.

Due to the cost of doing business in a particular area can be tough we can find alternative areas with similar opportunities. This can be achieved by clustering the similar areas which have similar distribution of restaurants.

We will consider the top 10 most common types of restaurants, which we will use for analyzing the similar areas in the city for business. Then we will also find the top 5 most common restaurants in a particular area to get more insights.

So first we fetch the data for areas in the Vancouver city, then we use the ‘Geopy’ library to find the coordinates of the areas of the Vancouver city. On analyzing the data for coordinates we found that the library gave similar values for areas like NE Downtown and SW Downtown, so we group the area having similar coordinates and gave a common name like Downtown from the above example.

Then we fetch the data for restaurants in the areas we set the radius to 1km so there were many restaurants which were in more than one area so we allocated the restaurant to the nearest area.

After that we prep the data for kMeans Clustering as the data was categorical we need to transform the data suitable for the algorithm, so we use one hot encoding in which we create one column for each category.

	Area	African Restaurant	American Restaurant	Asian Restaurant	Belgian Restaurant	Brazilian Restaurant	Cajun / Creole Restaurant	Cantonese Restaurant	Caribbean Restaurant	Chinese Restaurant	...	South American Restaurant	South Indian Restaurant
0	Bentall Centre	0	0	0	0	0	0	0	0	0	...	0	0
1	Mount Pleasant	0	0	0	0	0	0	0	0	0	...	0	0
2	Mount Pleasant	0	0	0	0	0	0	0	0	0	...	0	0
3	Mount Pleasant	0	0	0	0	0	0	0	0	1	...	0	0
4	Grandview-Woodland	0	0	0	0	0	0	0	0	0	...	0	0

5 rows × 52 columns

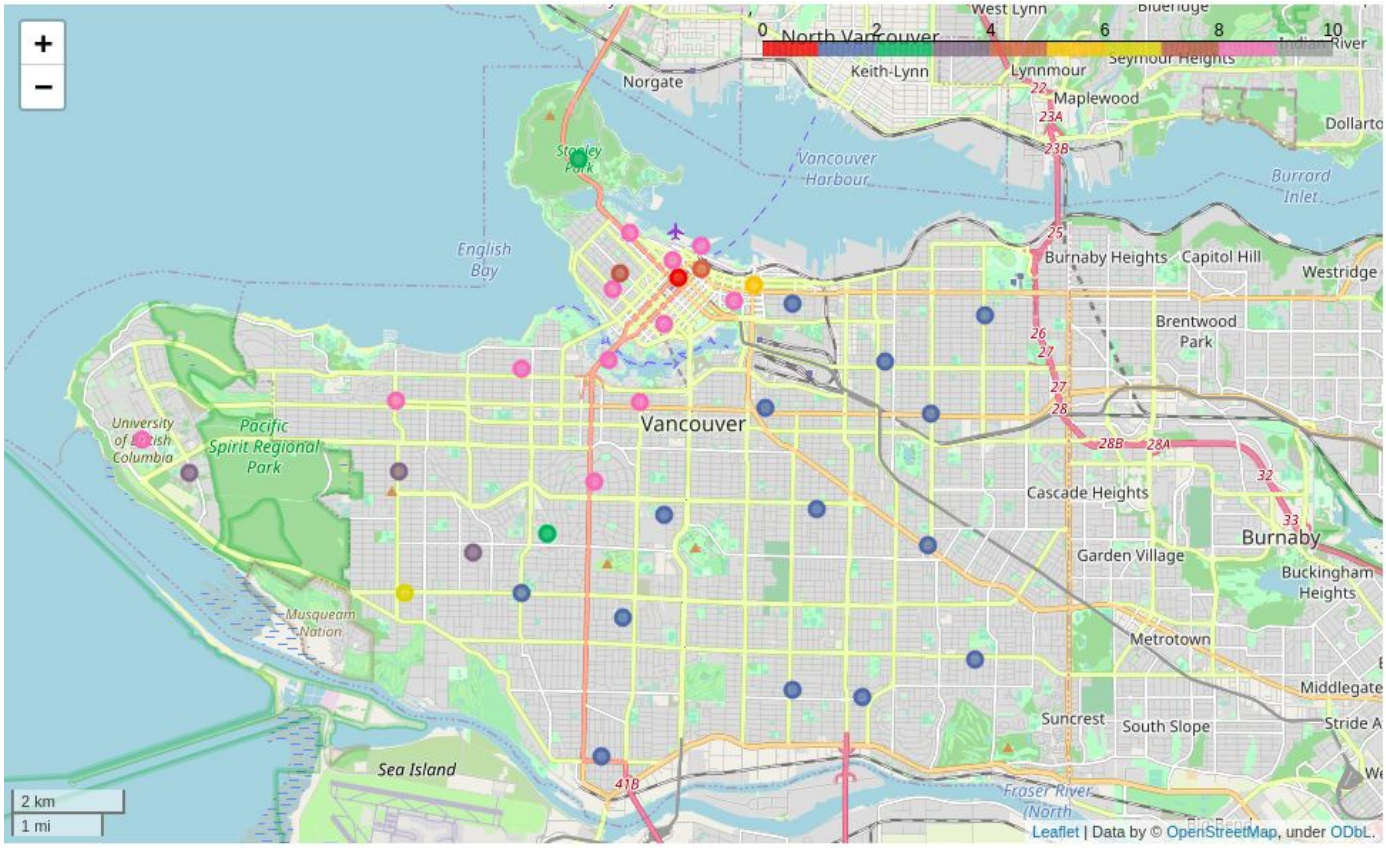
Then we use groupby to get the number and type of restaurants in an area. Then use the data for the clustering algorithm with the value of k set to 9.

4. Result

After trying out various values for k in K-Means Clustering we found the best possible value for k i.e 9.Which result in 3 major clusters.

The area in clusters has a similar distribution of restaurants. The outcome of the clustering is:

Cluster 1:- Simple Restaurant
Cluster 2:- Vietnamese, Chinese, Japanese, Sushi, Indian Restaurants
Cluster 3:- Fast Food Restaurant
Cluster 4:- Sushi, Indian, Italian, Simple Restaurant, Mediterranean Restaurants
Cluster 5:- Mexican Restaurant
Cluster 6:- Asian Restaurant
Cluster 7:- Japanese Restaurant
Cluster 8:- Hawaiian Restaurant, Malay Restaurant
Cluster 9:- Japanese, Simple, Sushi, Seafood, Vegetarian / Vegan Restaurants



We also got the top 5 restaurants in every area.

```
Area :  Arbutus Ridge
                 Restaurant Count Frequency
0          Sushi Restaurant     4  0.571429
1          Asian Restaurant     1  0.142857
2  Mediterranean Restaurant     1  0.142857
3        Italian Restaurant     1  0.142857
4     Vietnamese Restaurant     0         0
============================================================
Area :  Bentall Centre
                      Restaurant Count  Frequency
0            Japanese Restaurant     4   0.166667
1                     Restaurant     3      0.125
2             Italian Restaurant     3      0.125
3  Vegetarian / Vegan Restaurant     2  0.0833333
4        New American Restaurant     2  0.0833333
============================================================
Area :  Cambie
                      Restaurant Count  Frequency
0          Vietnamese Restaurant     2   0.181818
1               Sushi Restaurant     2   0.181818
2             Chinese Restaurant     2   0.181818
3  Vegetarian / Vegan Restaurant     1  0.0909091
4               Greek Restaurant     1  0.0909091
============================================================
```

5. Discussion

So from above we can say that the prominent clusters are :
Cluster 2, Cluster 3 which consists of only fast food as it is located near a park where fast food would be very popular, Cluster 9 which consists of Japanese, Sushi and Seafood which are located near water bodies as it would be easier to get raw fish from the water bodies.

We could cluster restaurants on other factors like rating, cost of food, availability etc.

6. Conclusion

From the data now we can choose which type of restaurant to open in a particular area. We also found out that the asian cuisines like Chinese and Japanese are very popular among the people of Vancouver city.