

PROJECT REPORT: Predicting Neighbourhoods for opening Indian Restaurant in British Columbia, Canada

Ajay Kumar Rabidas

July 4, 2019 (Capstone Week 2)

1. Introduction

1.1 Background

We all know that Indians could be found in every part of the world, but we are going to focus our attention towards British Columbia, Canada is one of the countries which allows a lot of immigrants to become a permanent resident or work there. This creates a big opportunity of diverse businesses to flourish. A restaurant business is one of them. Every neighbourhood in every country in the entire world has some type of restaurant. Because people of different countries and cultures live in Canada and according to a survey of the Canadian government, most immigrants are from Asian countries. Another survey shows that the second most recognized language spoken in British Columbia is Punjabi. This makes it obvious that there would be a huge requirement of Asian eat outs there.

1.2 Problem Statement

In cities across the world there are neighbourhoods that concentrate most of the activity and other neighbourhoods that concentrate most of the housing space. In this project I want to explore the differences between areas of a city, analysing the kind of activities that create the most important leisure centres and exploring the changes, new businesses or activities, required to be created in some regions to make them more similar to the areas that concentrate the activity. For this instance, we are concentrating on opening an Indian restaurant in a neighbourhood that doesn't have many, which could comfort the native Indians and hint them to consider choosing that place for living. Which could result in spreading out the Asian immigrant population to less dense areas and ultimately proper development across all areas of British Columbia.

We will start with the most populated and active place and try diverging from there. Once the boundary is broken the spread could be unlimited. There wouldn't be a restriction on immigrants' mind that Indian should prefer Victoria or Vancouver only for immigrating.

Looking at the background it seems very easy to judge that an Indian restaurant business is going to flourish in British, Columbia. But the problem here is to find a neighbourhood for opening that restaurant. British Columbia is a big place to choose from. My problem statement for this project would be to analyse the neighbourhood of British Columbia and try to find a neighbourhood that is best suited for opening a new Indian restaurant business and in the process figure out if a neighbourhood is good for moving in.

2. Data acquisition and cleaning

2.1 Data sources

Based on definition of our problem, factors that will influence our decision are:

- Number of existing any type of restaurants in the neighbourhood
- Number of Indian restaurants in the neighbourhood

Following data sources will be needed to extract/generate the required information:

- Centres of candidate co-ordinates would be extracted from a webpage URL: <http://www.geonames.org/postal-codes/CA/BC/british-columbia.html>
- Number of restaurants and their type and location in every neighbourhood will be obtained using Foursquare API (<https://en.foursquare.com/>)

Since the [geonames.org](http://www.geonames.org) website does not follow the best html standards the first step would be to extract information from it. I would be using a web scraping framework to load the relevant data to a pandas data-frame.

Secondly, cleaning the data is important because the neighbourhood names and latitudes and longitudes are inappropriately formatted.

Thirdly, I will be using foursquare to get active venues of all the neighbourhoods.

All the data is combined into a data-frame that contains all the explicative variables for every section of the city. The information is then vectorised to include a count of the number of services offered in every section, that is, the total number of parking slots, total number of gardens, total number of parks is computed. This information is complemented with the information obtained from Foursquare, thus including information about the number of cafeterias, restaurants, clothing shops and other businesses in every section. The so formed dataset will be the working dataset for the project.

A clustering algorithm will then be used to classify the different section into groups of similar sections. The total number of clusters will be selected in such a way that the clustering is compatible with the field experience, that is, ensuring that each cluster really represent a quantity and quality of activity that can be observed on the field. If discrepancies are observed between the clustering algorithm output and the classification based on experience, different machine learning techniques will be tested to map the section characteristics to the group assigned by experience and field work. The objective of the previous task is to construct the decision boundaries that separate the different classes. Once these boundaries have been constructed, it will be easy to compute the minimal set of changes that should be implemented into one section of the city to transform it into a different type, that is, how to modify one section to change the overall level of activity and make it more similar to the objective that we may be pursuing as decision-makers.

2.2 Data Cleansing

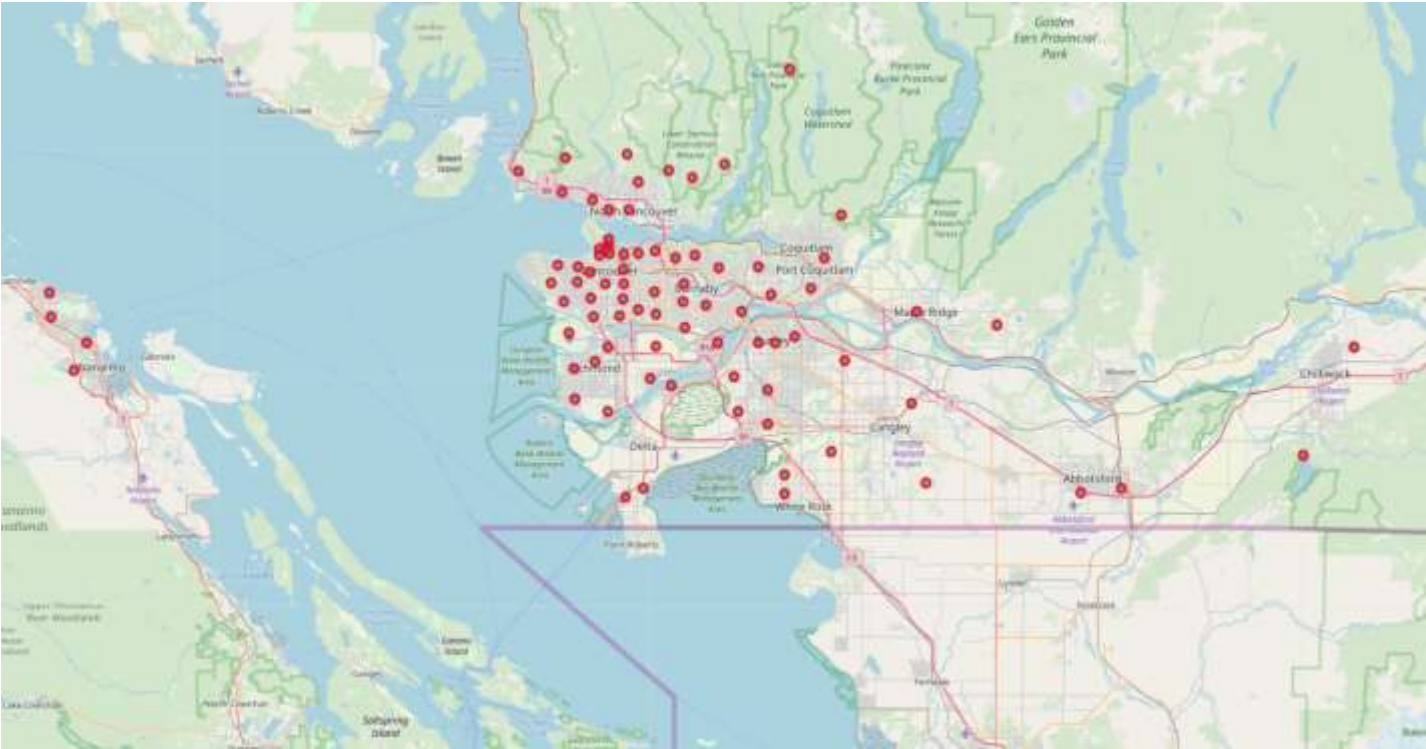
The dataframe obtained from the above mentioned procedures are not directly usable with any analytical algorithm. Here is a brief description of the data received from the webpage and preliminary corrections:

Index	Postal Code	Borough	Neighbourhood (to be fixed)	Location (to be fixed)
86	V5K	Vancouver	Vancouver (North/East Hastings-Sunrise)	\xa0\xa0\xa049.281/-123.04
87	V5L	Vancouver	Vancouver (Northern Grandview-Woodlands / Western Grandview-Woodlands/ Grandview-Woodlands downtown)	\xa0\xa0\xa049.279/-123.067
94	V5V	Vancouver	Vancouver (Northern Kensington / Eastern Kensington)	\xa0\xa0\xa049.248/-123.091
95	V5W	Vancouver	Vancouver (SE Riley Park-Little Mountain)	\xa0\xa0\xa049.233/-123.092
94	V5V	Vancouver	Vancouver (South-east Kensington)	\xa0\xa0\xa049.248/-123.093
94	V5V	Vancouver	Vancouver (Kensington / Kensington Garden Park)	\xa0\xa0\xa049.248/-123.090

As we can see, the location data obtained from the webpage is not usable, and multiple Neighbourhood names are duplicate if we remove north, south, east, and west from their names. Foursquare API was unable to work with such names. Multiple other records didn't have a Borough name. After cleansing, the final data frame looks like this:

Index	PostalCode	Borough	Neighbourhood (fixed)	Latitude (fixed)	Longitude (fixed)
86	V5K	Vancouver	Hastings-Sunrise	49.281	-123.04
87	V5L	Vancouver	Grandview-Woodlands	49.279	-123.067
90	V5P	Vancouver	SE Kensington	49.222	-123.068
91	V5R	Vancouver	Renfrew-Collingwood	49.24	-123.041
92	V5S	Vancouver	Killarney	49.218	-123.038
93	V5T	Vancouver	Mount Pleasant	49.262	-123.092
94	V5V	Vancouver	Kensington	49.248	-123.091
95	V5W	Vancouver	SE Riley Park-Little Mountain	49.233	-123.092

Here is how the data looks when plotted on a folium map:



Now we use, Foursquare API to get Venue Data per Neighbourhood, we receive the fields highlighted in yellow:

Postal Code	Borough	Neighbourhood	Borough Latitude	Borough Longitude	Venue Name	Venue Latitude	Venue Longitude	Venue Category
V5K	Vancouver	Hastings-Sunrise	49.281	-123.04	The Fair at the PNE	49.282971	-123.042109	Fair
V5K	Vancouver	Hastings-Sunrise	49.281	-123.04	Wooden Roller Coaster	49.281744	-123.035128	Theme Park Ride / Attraction
V5K	Vancouver	Hastings-Sunrise	49.281	-123.04	Livestock Barns	49.284037	-123.039278	Farm
V5K	Vancouver	Hastings-Sunrise	49.281	-123.04	Playland	49.281924	-123.036258	Theme Park
V5K	Vancouver	Hastings-Sunrise	49.281	-123.04	Hastings Community Centre	49.280778	-123.039176	Event Space
V5K	Vancouver	Hastings-Sunrise	49.281	-123.04	PNE Amphitheatre	49.283777	-123.037242	Stadium

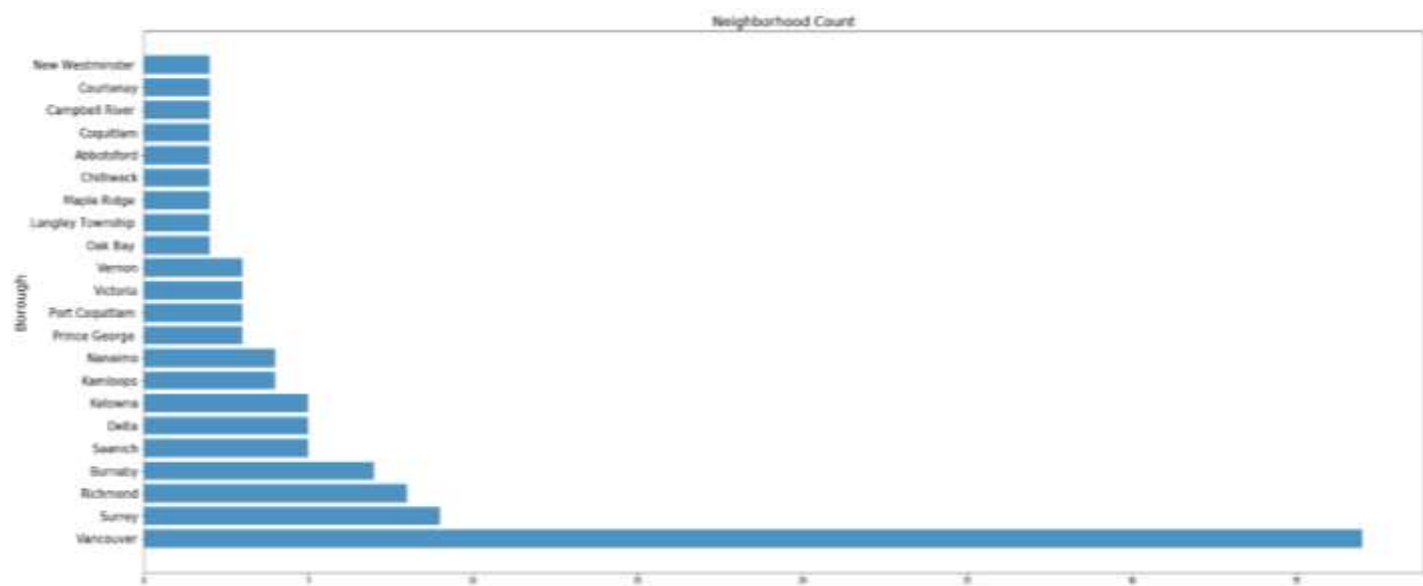
3. Explanatory Data Analysis:

3.1 Feature Selection:

Keeping only the following columns and dropping others as it wouldn't be required in our analysis

Postal Code	Borough	Neighbourhood	Borough Latitude	Borough Longitude	Venue Name	Venue Latitude	Venue Longitude	Venue Category
-------------	---------	---------------	------------------	-------------------	------------	----------------	-----------------	----------------

Using the `value_counts()` function `neighborhoodDF['Borough'].value_counts()` we can see that **Vancouver** has the maximum number of neighbourhood data available and since it is also the most populated and famous city for immigrants. Since our plan is to start from the most populated city and extend outwards, which means if we open up an Indian restaurant in less restaurant dense areas of Vancouver we could be able to attract Asian population thereby increase our chance of profitability. Also more number of immigrants would get the hint of their community presence there which would help immigrant population spread out of the populated area.



3.2 Methodology:

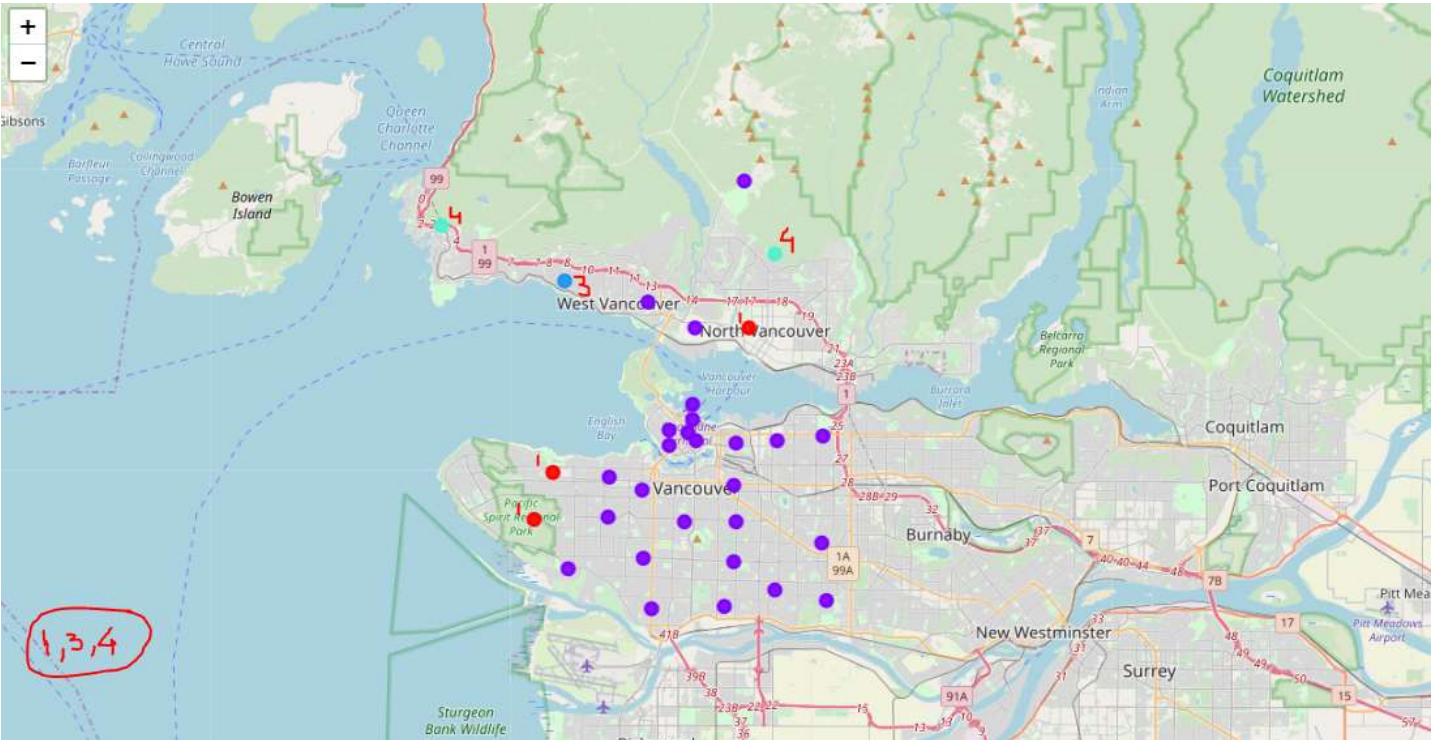
On directing our efforts on detecting areas of Vancouver that have low restaurant density. In first step we have collected the required **data: location and type (category) of every restaurant in from Vancouver centre**. We have also **identified Asian restaurants** (according to foursquare categorization). Second step in our analysis is exploration of different areas of Vancouver - we will calculate the count of each venue category in a given area, which gives us a basic understanding of the type of the area. Third step is to rank top 10 venues in every neighbourhood to get a proper view of the neighbourhood. This is what the data looks at this stage: showing only to 5 here

Labels	Borough	Neighborhood	latitude	longitude	Venue Rank-1	Venue Rank-2	Venue Rank-3	Venue Rank-4	Venue Rank-5
2	Vancouver	Bentall Centre	49.293	-123.116	Airport Terminal	Airport	Irish Pub	Outdoor Sculpture	Breakfast Spot
2	Vancouver	Central Kitsilano	49.265	-123.165	Coffee Shop	Pizza Place	Yoga Studio	Pub	Vegetarian / Vegan Restaurant
2	Vancouver	Dunbar-Southlands	49.23	-123.189	Construction & Landscaping	Vietnamese Restaurant	Home Service	Fast Food Restaurant	Yoga Studio
2	Vancouver	End	49.283	-123.13	Japanese Restaurant	Coffee Shop	Gay Bar	Hotel	Bakery
2	Vancouver	Fairview	49.248	-123.121	Bus Stop	Coffee Shop	Chinese Restaurant	Sushi Restaurant	Bubble Tea Shop
2	Vancouver	Grandview-Woodlands	49.279	-123.067	Asian Restaurant	Theater	Italian Restaurant	Grocery Store	Brewery
2	Vancouver	Hastings-Sunrise	49.281	-123.04	Theme Park Ride / Attraction	Stadium	Beer Garden	Event Space	Theme Park
2	Vancouver	Kensington	49.248	-123.091	Coffee Shop	Bus Stop	Chinese Restaurant	Vietnamese Restaurant	Ice Cream Shop
2	Vancouver	Killarney	49.218	-123.038	Chinese Restaurant	Bus Stop	Pharmacy	Bank	Farmers Market
2	Vancouver	Mount Pleasant	49.262	-123.092	Sushi Restaurant	Hotel	Vietnamese Restaurant	Grocery Store	Ethiopian Restaurant
2	Vancouver	NE Downtown	49.279	-123.114	Hotel	Coffee Shop	Restaurant	Breakfast Spot	Italian Restaurant
2	Vancouver	NW Arbutus Ridge	49.25	-123.166	Bakery	Bike Rental / Bike Share	Italian Restaurant	Caribbean Restaurant	Farm
2	Vancouver	NW Shaughnessy	49.26	-123.146	Pharmacy	Restaurant	Sushi Restaurant	Furniture / Home Store	Tennis Court
2	Vancouver	Pacific Centre	49.282	-123.119	Hotel	Food Truck	Lounge	Café	Steakhouse
2	Vancouver	Renfrew-Collingwood	49.24	-123.041	Bar	Hotel	Fish & Chips Shop	Bus Stop	Asian Restaurant
2	Vancouver	SE Kensington	49.222	-123.068	Pizza Place	Noodle House	Middle Eastern Restaurant	Pharmacy	Fish Market
2	Vancouver	SE Kerrisdale	49.215	-123.14	Chinese Restaurant	Dessert Shop	Sushi Restaurant	Field	Thai Restaurant
2	Vancouver	SE Oakridge	49.216	-123.098	Sporting Goods Shop	Indian Restaurant	Restaurant	Park	Yoga Studio
2	Vancouver	SE Riley Park-Little Mountain	49.233	-123.092	Chinese Restaurant	Dessert Shop	Bakery	Field	Pizza Place
2	Vancouver	SW Downtown	49.277	-123.13	Hotel	Mexican Restaurant	Café	Diner	Sushi Restaurant
2	Vancouver	Shaughnessy	49.234	-123.145	Bus Stop	Chinese Restaurant	Sushi Restaurant	Asian Restaurant	Electronics Store
2	Vancouver	Strathcona	49.278	-123.091	Park	Pie Shop	Pub	Cheese Shop	Sandwich Place
2	Vancouver	Vancouver Northwest	49.378	-123.086	Trail	Restaurant	Ski Chalet	Mountain	Movie Theater
2	Vancouver	Vancouver Southwest	49.322	-123.115	Coffee Shop	Gym	Furniture / Home Store	Liquor Store	Greek Restaurant
2	Vancouver	Waterfront	49.287	-123.116	Hotel	Coffee Shop	Café	American Restaurant	Restaurant

In third and final step we will focus on most promising areas by creating **clusters of locations that are of similar type**: we will take into consideration locations with no more than two restaurants, and we want locations **without Indian restaurants**. We will create clusters (using **k-means clustering**) and present map of all such locations.

3. Result and Discussion:

Looking at the cluster markings on the map below gives us an idea that the Purple areas are of similar type and Vancouver being one of the most developed and populated cities in British Columbia would be difficult to consider opening a new restaurant as it would already be crowded of many others. We see the same looking at the dataset for the cluster label 1 which is purple on the map. It confirms our assumption. It has a huge number of Asian restaurants.



The neighbourhoods marked as 1, 3 and 4 in red, cyan and sky seems to be the best place to open a new Indian restaurant business as they have a very low density on restaurants.

To verify our findings we can look at the data below, we don't have a competition in this region and it also appears to be a residential area:

Cluster Labels	Borough	Neighborhood	latitude	longitude	Venue Rank-1	Venue Rank-2	Venue Rank-3	Venue Rank-4	Venue Rank-5	V
1	Vancouver	Chaldecutt	49.249	-123.209	Park	Yoga Studio	Fair	Fish & Chips Shop	Financial or Legal Service	
1	Vancouver	Kitsilano	49.267	-123.198	Park	Yoga Studio	Fair	Fish & Chips Shop	Financial or Legal Service	
1	Vancouver	Vancouver Southwest Central	49.322	-123.083	Park	Baseball Field	Yoga Studio	Falafel Restaurant	Fish Market	
3	Vancouver	Vancouver South	49.34	-123.191	Art Gallery	Yoga Studio	Fair	Fish Market	Fish & Chips Shop	1
4	Vancouver	Vancouver Northwest Central	49.35	-123.068	Trail	Paper / Office Supplies Store	Yoga Studio	Fair	Fish & Chips Shop	1
4	Vancouver	Vancouver West	49.361	-123.263	Tapas Restaurant	Trail	Yoga Studio	Fair	Fish & Chips Shop	1

4. Conclusion:

We started this project to explore the British Columbia for opening an Indian restaurant business. Starting from a huge country, to narrowing down to 4 specific cities was really interesting. Going through the project has made me realise the power of data science tools. The methodology however may be applied to larger cities, where the larger amount of population and the larger space to cover, can generate opportunities to improve the presence of important commercial activities for the section or the neighbourhood.