

Predicting Neighborhoods for opening Indian Restaurant in British Columbia, Canada

Ajay Kumar Rabidas
July 4, 2019 (Capstone Week 1)

1. Introduction

1.1 Background

1.1.1 About British Columbia

British Columbia is the western-most province in Canada, bordering the Pacific Ocean. Internationally known for its stunning natural beauty (there are six national parks), British Columbia is home to more than 4.6 million people. While most of these residents live in the Lower Mainland area, British Columbia's land area is the size of France, Germany and the Netherlands combined.

British Columbia is well known for its high standard of living, with the city of Vancouver consistently ranked among the top cities in the world to live and work in. Vancouver provides economic opportunity and support networks of existing immigrant communities, all within safe communities and natural beauty. Likewise, Victoria and interior B.C. communities such as Kelowna provide a similar array of qualities.

1.1.2 British Columbia for an Indian

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 neighborhood 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 neighborhoods that concentrate most of the activity and other neighborhoods that concentrate most of the housing space. In this project I want to explore the differences between areas of a city, analyzing the kind of activities that create the most important leisure centers 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 neighborhood 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 denser 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 neighborhood for opening that restaurant. British Columbia is a big place to choose from. My problem statement for this project would be to analyze the neighborhood of British Columbia and try to find a neighborhood that is best suited for opening a new Indian restaurant business and in the process figure out if a neighborhood is good for moving in.

2. Data acquisition and cleaning

Data sources

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

- number of existing any type of restaurants in the neighborhood
- number of Indian restaurants in the neighborhood

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

- Centers 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 neighborhood will be obtained using Foursquare API (<https://en.foursquare.com/>)

I would be using the website <http://www.geonames.org/postal-codes/CA/BC/british-columbia.html> to get the list of neighborhoods and its location.

Since the 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 neighborhood names and latitudes and longitudes are inappropriately formatted. Thirdly, I will be using foursquare to get active venues of all the neighborhoods.

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 vectorized 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.

The information collected divides British Columbia into several sections that are shown in the figure below.