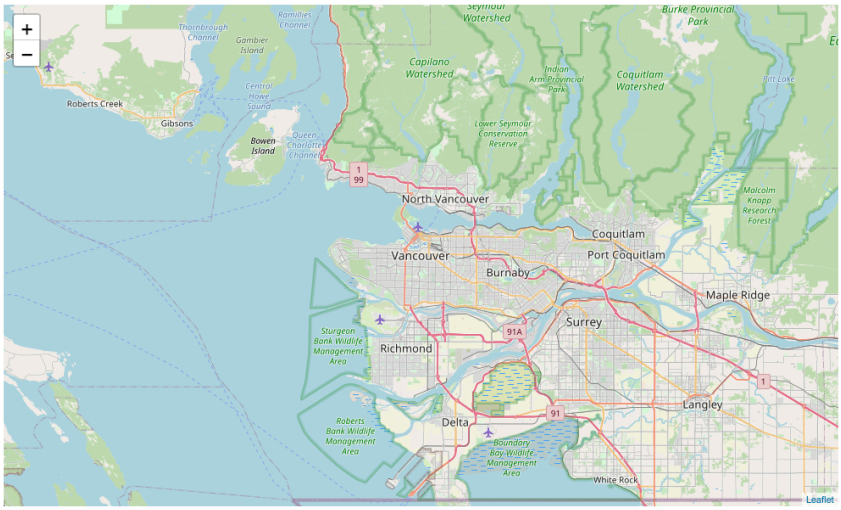
Regional Chinese Restaurant Saturation in Greater Vancouver, Canada

1. Introduction

This project aims to assist entrepreneurs in the restaurant sector, and in particular entrepreneurs interested in opening a new Chinese restaurant within the Greater Vancouver area. Vancouver, sits on the west coast of Canada and as such has been an attractive immigration destination for individuals, families, and investors from Asia. There are many restaurants throughout Greater Vancouver 's various regions, and as a result of the aforementioned immigration, also a substantial number of Chinese, Japanese, Vietnamese, and other Asian food serving restaurants.

An entrepreneur looking to open a new Chinese restaurant in Vancouver would want to know which location would have the potential of maximizing customer entry from local footfall, while also minimizing direct competition from other Chinese restaurants in the area.

This project therefore aims to assist such entrepreneurs by analyzing venues within the various regions of the city of Vancouver in order to identify which regions have a high number of restaurants (making them a drawing point for clients interested in eating out) while at the same time have a low number of popular Chinese restaurants, which would lead to less direct competition for a new potential Chinese restaurant opening in the region. The main research question evaluated is, which region in the Greater Vancouver area has both a high number of restaurants and at the same time a low number of Chinese restaurants?



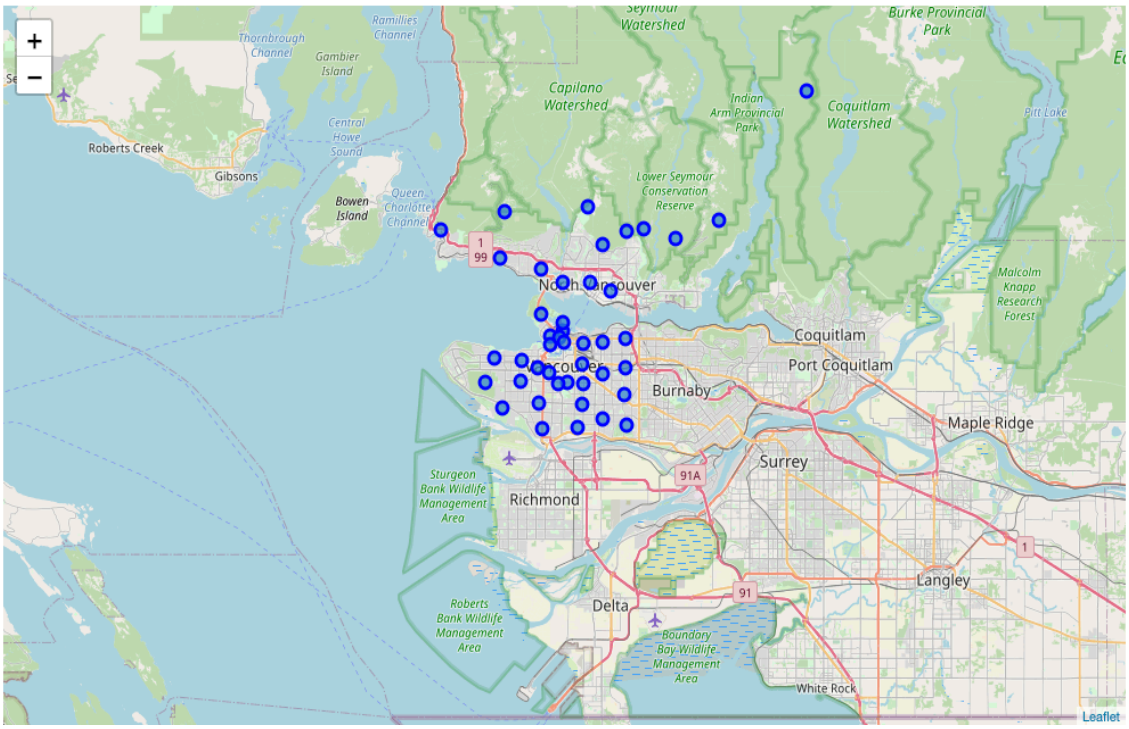
Map of Greater Vancouver, BC

2. Data

The data that this project relies on is a combination of 2 key data sources:

1. A structured dataset that has been scraped from a website containing postal code information for regions throughout the entire province of British Columbia, Canada, in which the city of Vancouver is located. This data is then cleaned to focus only on regions within the Greater Vancouver area.

2. A dataset of 100 top venues within a 1000 meter proximity of each region of Greater Vancouver that have been requested via the Foursquare location data API.



Map of Greater Vancouver, BC, with key regions superimposed in purple markers

3. Methodology

In this project I utilised a Clustering algorithm, which is an *Unsupervised Learning* approach. It is unsupervised due to the fact that I do not have pre-existing labels for the various regions and would therefore like to rely on machine-learning to identify potential patterns and similarities within the various regions and venues included in this dataset.

By using the Clustering algorithm, the regions will be split into subgroups based on patterns and similarities that are found in the data. Observations within one cluster will have shared similarities with other members of its cluster and will differ from those of other clusters.

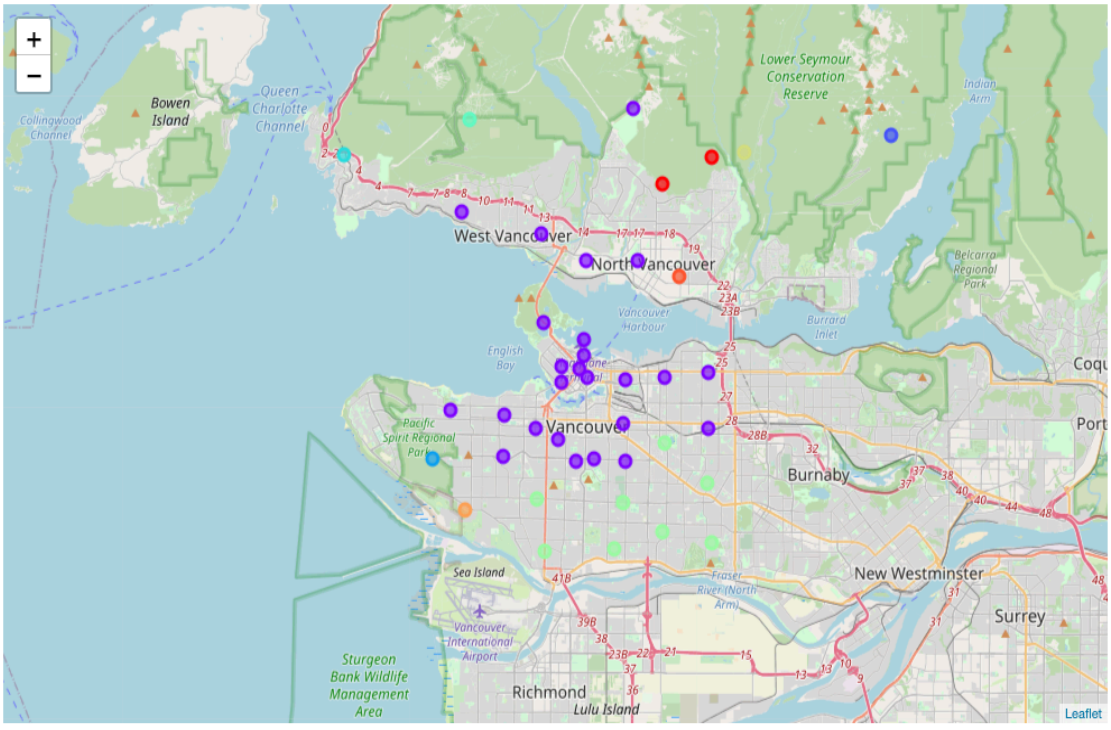
The clustering algorithm selected in this case is the K-Means algorithm. K is a pre-specified number of subgroups into which subgroups will be populated. It is therefore important in some cases (when possible) to rely on the data at hand to determine the target number of clusters.

The methodology for this project followed the following steps: Scraping and wrangling the initial dataset of postcode-based regions, merging with regional venue-data from the Foursquare API, utilizing one-hot coding to classify venue frequency at the regional level, and finally, splitting the regions into subgroups of similar characteristics using the K-Means clustering algorithm.

4. Results

When running the algorithm using K-Means, where K = 10, I find that the largest 2 clusters comprise of regions within the northern and southern parts of central Vancouver. These 2 clusters are of the most urban nature and each have more urban amenities and venues within their top venues results. These include restaurants, coffee shops, and hotels in the top most common venue spots. Interestingly, the cluster representing the south part of metro-Vancouver (light-green) sees a substantially higher number of Indo-Chinese restaurants in their top 2 most common venue spots.

Other clusters contain considerably smaller subgroups and tend to be on the outskirts of the city. These comprise of venues such as parks, trails, zoos, golf courses, and Ski areas depending on their geographical location and characteristics.



Greater Vancouver with clusters superimposed and color-coded

5. Discussion

It is important to note that these results are based on an initially chosen number of clusters of K=10. Additional experimentation may lead to identifying a more optimal number of clusters that could result in a better segmentation of subgroups. Initial attempts made with a smaller K resulted in only 1 major cluster comprising of the entire Vancouver metro area. This is what led to an eventual clustering with K=10.

It is important to note that more data could be taken into consideration in this project. This would allow me to carry out a more in-depth analysis and arrive at higher quality recommendations. For example, factors such as commercial rent prices in each region, average spend on eating out in each region per individual, local business tax rates, and socio-economic status of local residents at each region, could further benefit the outcome of this project.

Based on these results I would recommend the prospective entrepreneur focus on the northern part of metro Vancouver (purple cluster) that has a large number of restaurants but competing Chinese restaurants are less common than they are in the south part of the city. But within this subgroup, I would recommend avoiding areas such as Chinatown, West Mount Pleasant, East Fairview, West Kensington, and East Mount Pleasant which all have a strong presence of pre-existing Chinese restaurants. In order to further narrow down the final selected location, I recommend the additional factors noted above should be considered for each of the remaining regions within this target subgroup.

6. Conclusion

In this report, I have clustered the regions of the Greater Vancouver city area. I have observed interesting differences between clusters of regions that were of a primarily urban environment and clusters of regions that lay on the outskirts of the city and those that lay within the more mountainous area to the north.

This project does not take additional factors into consideration beyond the presence of common venues in various regions, and as such does cannot provide a final recommendation as to the best location for opening a new Chinese restaurant. It does however prove useful in narrowing down the search to a smaller number of key prospective regions which should then undergo further due diligence.