Vancouver Restaurant Market Analysis

Jan 11th, 2021

Introduction

Restaurants are constantly creating bucket loads of data, which, if used correctly, can prove to be a gold mine for a restaurant's success. The objective of this project is to analyze various Vancouver restaurant data to help determine the optimal location and the type of cuisine to specialize in for a new restaurant in Vancouver, BC. Considering the possibility of expanding the business into multiple locations in the future, we also examined chain restaurant data. Below are the questions we want to answer:

- What are the most common cuisines in each Vancouver neighbourhood?
- Is there a relationship between the population of a region and its number of chain restaurants?
- Does a chain restaurant's geographical location influence its rating?
- Which chain restaurants could be our potential competitors?
- How many clusters of chain restaurants are there in Vancouver?

Data Acquisition

We collected two different restaurant datasets in order to answer the above questions. One includes an unbiased sample of all restaurants in Vancouver, and the other consists of a representative sample of Vancouver chain restaurants.

For the all restaurant dataset, we scraped the restaurant data from Google Maps for each neighbourhood in Vancouver using Selenium. And to improve data quality, we included a list of restaurants from the Foursquare API. For the chain restaurant dataset, we queried searches on Google Maps using a carefully chosen list of chain restaurants, and scraped data from those search results.

We also downloaded datasets from Vancouver's Open Data Portal to gather detailed information about the geographic coordinate, ethnicities, area boundaries, and annual household incomes for each neighbourhood in Vancouver.

Data Cleaning

The raw restaurant data contains many places that are not considered to be restaurants, such as shopping malls or real estate agencies. Therefore, we had to remove those manually before analyzing. Furthermore, any duplicates and restaurants with an invalid address or outside of Vancouver were also removed.

Unlike the data from the Foursquare API, the data we scraped from Google contained neither the geographic coordinate nor the neighbourhood of each restaurant. Therefore, we had to use Geocode to find out those missing latitudes and longitudes. And then proceeded to determine the neighbourhood of each restaurant using Shapely.

	name	address	cuisine	lat	lng	neighbourhood
0	BIGSBY THE BAKEHOUSE	4894 MacKenzie St	Restaurant	49.242537	-123.170078	Arbutus-Ridge
1	One Bowl	5687 Balsam St	Restaurant	49.235075	-123.161575	Arbutus-Ridge
2	Moki's Pizza	4027 Macdonald St	Pizza Delivery	49.250606	-123.168017	Arbutus-Ridge
3	Secret Garden Tea Company	2138 W 40th Ave	Restaurant	49.235447	-123.156045	Arbutus-Ridge
4	Subway	4665 Arbutus St	Sandwich	49.244540	-123.153953	Arbutus-Ridge
5	Tera V Burger	2961 W Broadway	Fast Food	49.264299	-123.171709	Kitsilano
6	Shota Sushi & Grill	5688 Yew St	Sushi	49.234952	-123.157502	Arbutus-Ridge
7	Bishop's	2183 W 4th Ave	Pacific Northwest	49.268319	-123.154697	Kitsilano

The chain restaurant data was handled with further filtering; those with names that do not match the keyword searched on Google Maps were filtered out using difflib. In addition, only the chain restaurants with a valid rating and review count were kept to prevent an invalid conclusion of whether a chain restaurant's geographical location influences its rating.

Data Analysis & Findings

Most Common Cuisines

Surprisingly, the majority of restaurants in Vancouver originate in some other country.

Perhaps due to the convenience and affordability of fast food restaurants, they are currently the most common cuisines in most of the Vancouver neighbourhoods.

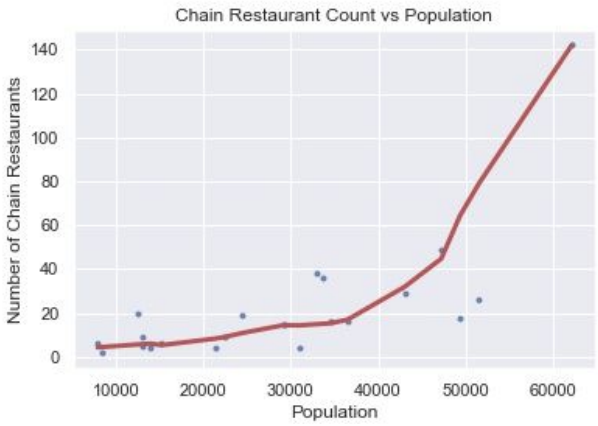
Besides fast food restaurants, the list of most common types of cuisine in each Vancouver neighbourhood usually includes Cafe, Japanese, and Chinese restaurants. Vietnamese restaurants that mainly serve pho are also quite popular.

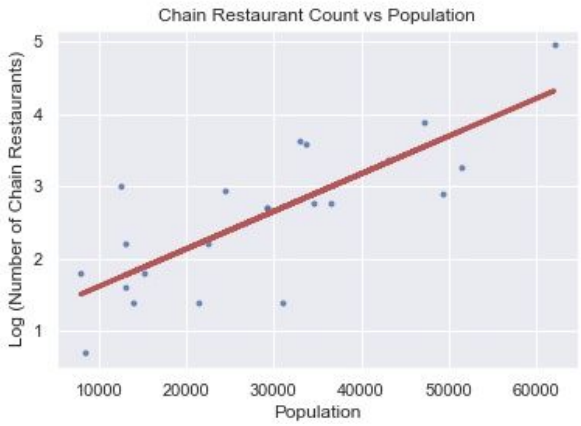
We suspect that there is a relationship between the population and income of a neighbourhood and its most common cuisines. However, further statistical data analysis is required to draw a conclusion.

	neighbourhood	first	second	third	fourth	fifth
0	South Cambie	Japanese	Chinese	Sandwich	Fast Food	Bakery
1	Shaughnessy	Asian	Coffee	Fast Food	Japanese	Sandwich
2	Strathcona	Cafe	Pub	Chinese	Fast Food	Sandwich
3	Oakridge	Fast Food	Japanese	Bakery	Cafe	Chinese
4	West Point Grey	Japanese	Fast Food	Cafe	Bakery	Asian
5	Kerrisdale	Chinese	Japanese	Bakery	Cafe	Fast Food
6	Arbutus-Ridge	Japanese	Chinese	Fast Food	Sandwich	Noodle
7	Dunbar-Southlands	Japanese	Cafe	Fast Food	Breakfast	British
8	Riley Park	Japanese	Chinese	Vietnamese	Fast Food	Bakery
9	Marpole	Chinese	Fast Food	Japanese	Vietnamese	Sandwich
10	Grandview-Woodland	Fast Food	Japanese	Cafe	Italian	Breakfast
11	Killarney	Chinese	Fast Food	Japanese	Korean	Bakery
12	Victoria-Fraserview	Chinese	Fast Food	Japanese	Vietnamese	Bakery
13	Mount Pleasant	Fast Food	Japanese	Cafe	Bakery	Mexican
14	Fairview	Japanese	Fast Food	Cafe	Chinese	Breakfast
15	Hastings-Sunrise	Fast Food	Japanese	Vietnamese	Bakery	Cafe
16	Sunset	Indian	Fast Food	Chinese	Japanese	Vietnamese
17	Kitsilano	Japanese	Fast Food	Cafe	Thai	Mexican
18	West End	Japanese	Fast Food	Chinese	Italian	Indian
19	Kensington-Cedar Cottage	Vietnamese	Chinese	Fast Food	Japanese	Cafe
20	Renfrew-Collingwood	Chinese	Fast Food	Japanese	Cafe	Filipino
21	Downtown	Fast Food	Japanese	Mexican	Cafe	Italian

Relationship between the Population and Number of Chain Restaurants

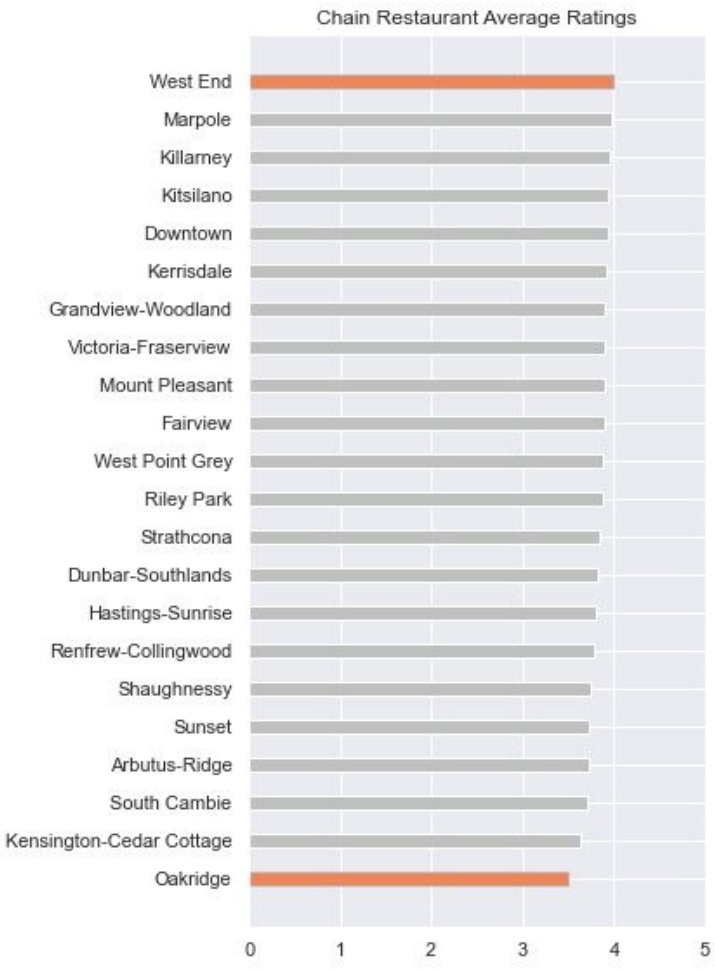
Ordinary Least Squares Regression was performed to estimate the relationship between the population of a region and its number of chain restaurants. We transformed right-skewed restaurant data in order to meet the test assumption that the relationship between the variables is linear. All other test assumptions, such as the normality and independence of the residuals were also tested to avoid drawing an incorrect conclusion from the results of our analysis. As a result, we are 95% confident that there is a relationship between the population of a region and its number of chain restaurants.





How Geographical Location Influence the Rating

We grouped chain restaurant data by its region, calculated their mean ratings and found that with a 5-star rating system, and found out that only 0.51 stars separate the neighbourhoods with the highest and the lowest average ratings (Figure on the right). Thus we can reasonably conclude that a chain restaurant's geographical location does not affect its rating.

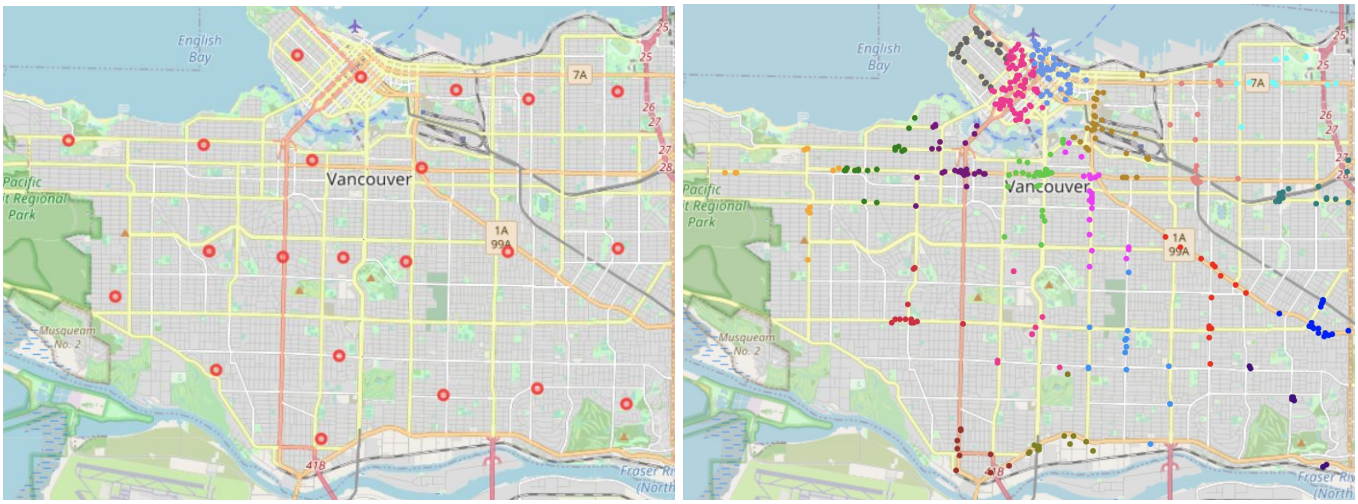


Potential Chain Competitors

Sklearn.neighbors and Geocode were used to find various information about potential nearby chain competitors. Upon entering an valid address in Vancouver, the algorithm would find its accurate geographic coordinate using forward geocoding. It would then display information on the specified number of nearby chain competitors.

Clustering of Chain Restaurants

We plotted the center locations for each neighbourhood using Folium to get a visual representation of Vancouver neighbourhoods (Figure below on the left). And then performed Affinity Propagation, a clustering algorithm based on the concept of “message passing” between data points, which predicts the cluster each chain restaurant belongs to. The number of clusters the algorithm returned was almost identical to the number of neighbourhoods within Vancouver. Furthermore, most clusters of chain restaurants are scattered around the central location of a neighbourhood.



Limitations

Lack of Data

The available restaurant datasets online do not contain enough data about Vancouver local restaurants. As a result, we had to carry out all the restaurant data collection ourselves. Moreover, the best dataset we could find for the Vancouver neighbourhoods were collected around 2016, which may not be representative in today’s time.

Data Inconsistency

Another issue we came across was that the cuisine labels provided by Google Maps were inconsistent. One restaurant would be labeled "Chinese Takeaway" whereas another would be "Chinese BBQ". And consolidating these labels into its "base" cuisine through manual cleaning requires tremendous effort and may be subjective.

Lack of Domain Knowledge

Without the right domain understanding, it is often difficult to get the right results. We only analyzed a small portion of the factors that we believed would contribute the most to a new restaurant’s success in Vancouver, BC. But perhaps the crime rates in the local area, the space and cost of nearby parking lots, and the current food trends would all be a determining factor of the restaurant’s success.
