Developing a new restaurant in the Calgary Zone

Written by: Lindsay D.

Nov 24, 2020

Introduction

Eating out at restaurants is a very common and popular activity in the city of Calgary. Calgary is home to many ethical backgrounds, which drives the city's diverse style of restaurants – these include restaurants serving Chinese, Italian, and Korean cuisines, to name a few. New restaurants are continually popping up around the city to meet high demand. Although eating out is a popular activity, there are also restaurants that are unfortunately unsuccessful in implementing their business – perhaps due to a poor choice in restaurant location. When opening a new restaurant, there are many factors to consider before selecting where you set up shop.

Problem

This research project aims to provide a recommendation for the style and location of a new restaurant in Calgary. There are many factors to consider when creating this recommendation such as: the number of existing restaurants in Calgary areas and their popularity, the popularity of different types of cuisines offered in Calgary, and information on the customers being served such as age and income.

Data Acquisition and Utilization

Data Sources

To explore the city of Calgary and get an understanding of the neighborhoods it includes, a table of neighborhoods and their corresponding locations was scraped from Wikipedia [1]. The following table provides a snapshot of the data collected:

Postal Code	Borough	Neighborhood	Latitude	Longitude
T1A	Medicine Hat	Central Medicine Hat	50.036460	-110.679250
T2A	Calgary	Penbrooke Meadows, Marlborough	51.049680	-113.964320
T3A	Calgary	Dalhousie, Edgemont, Hamptons, Hidden Valley	51.126060	-114.143158

Postal Code	Borough	Neighborhood	Latitude	Longitude
T4A	Airdrie	East Airdrie	51.272450	-113.986980
T5A	Edmonton	West Clareview, East Londonderry	53.5899	-113.4413
T6A	Edmonton	North Capilano	53.5483	-113.408
T7A	Drayton Valley	Not assigned	53.2165	-114.9893
T8A	Sherwood Park	West Sherwood Park	53.519	-113.3216

Table 1: Calgary neighborhood data scraped from Wikipedia

FourSquare information services were also utilized, via API connection, to retrieve data on popular venues across Calgary. This dataset returned all recorded venues in the searched proximity, including restaurants, gas stations, convenience stores, bars, etc.

Calgary community demographic information has also been scraped from a *Great News* demographics page [2]. This web page provides census data in different communities in Calgary, including: median household income, and median age. A sample of the information scraped from this webpage is provided in Table 2 below.

Community	Median Household Income	Median Age	City Quadrant	Median Home Sale Price
Abbeydale	$55,345	34	SE	$305,000
Acadia	$46,089	42	SE	$447,000
Albert Park / Radisson Heights	$38,019	37	SE	$349,900
Altadore	$53,786	37	SW	$925,000
Applewood Park	$65,724	33	SE	$380,000

Table 2: Calgary demographic data scraped from Great-News.ca

How the data will be used

The Calgary neighborhood data scraped from Wikipedia will be utilized to retrieve venue information from the FourSquare API. With this API, we can easily match neighborhoods with available venues, and filter out the venues we are interested in – restaurant venues.
With the FourSquare data, we can understand the different types of restaurants that currently exist in Calgary, which neighborhoods have the most and least number of restaurants, and which restaurants are the most popular based on occurrence frequency. This information allows us to cluster popular restaurants, and get an idea of how restaurant tastes vary across the city quadrants (NE, SE, SW and NW).

City census data will also be used to get a better understanding of the population throughout the city, and analyze metrics that might affect the constituent's frequency of eating out. Analysis may include looking at the median household income, and median resident age across different neighborhoods. It has been found that younger populations are more likely to spend their time eating out at restaurants than are older constituents that are more likely to have families at home. Additionally, as Calgary has been in an economic downfall for the past five years, much of the population has increased the portion of their disposable income that they reserve for savings, and limiting spending on luxuries, such as eating out at restaurants.

Methodology

The data collected was cleaned and organized into formats that could be applied in exploratory data analysis methods. A number of bar graphs were produced in order to visualize the data, and help guide the exploratory data analysis in order to draw useful results from the data.

Data Cleaning and Exploratory Data Analysis Completed

The raw Calgary Neighborhood data scraped from the Wikipedia page came from a table that also included information on other cities and towns in Alberta.

[2]:

	Postal Code	Borough	Neighborhood	Latitude	Longitude
0	T1A	Medicine Hat	Central Medicine Hat	50.036460	-110.679250
1	T2A	Calgary	Penbrooke Meadows, Marlborough	51.049680	-113.964320
2	T3A	Calgary	Dalhousie, Edgemont, Hamptons, Hidden Valley	51.126060	-114.143158
3	T4A	Airdrie	East Airdrie	51.272450	-113.986980
4	T5A	Edmonton	West Clareview, East Londonderry	53.5899	-113.4413

Figure 1: Raw data scraped from Wikipedia, prior to any filtering or cleaning

The first data cleaning step taken was to filter the dataset to remove information not pertaining to the city of Calgary.

The first data cleaning step taken was to filter the dataset to remove information not pertaining to the city of Calgary (represented in the Borough column in Figure 1). The Borough column was removed from

the dataset for clarity.

[3]:

	Postal Code	Neighborhood	Latitude	Longitude
1	T2A	Penbrooke Meadows, Marlborough	51.04968	-113.964320
2	T3A	Dalhousie, Edgemont, Hamptons, Hidden Valley	51.12606	-114.143158
10	T2B	Forest Lawn, Dover, Erin Woods	51.03180	-113.978600
11	T3B	Montgomery, Bowness, Silver Springs, Greenwood	51.08090	-114.161600
19	T2C	Lynnwood Ridge, Ogden, Foothills Industrial, G...	50.98780	-114.000100

Figure 2: Calgary Neighborhood dataset, cleaned from Figure 1

This dataset was explored on a map using the Folium library in order to get a better understanding of the number of neighborhoods in the City of Calgary, and their distribution in the city.

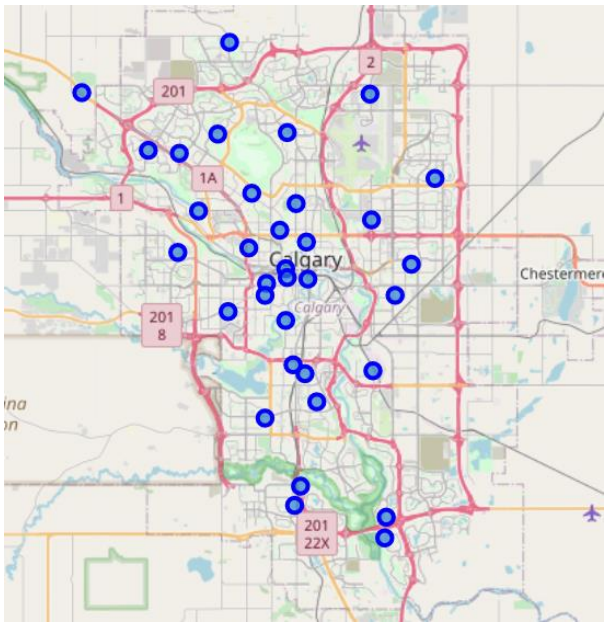

Figure 3: Calgary neighborhood exploratory data analysis with Folium

The Foursquare API was called upon to retrieve information on venues local to each of these neighborhoods.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
	Dalhousie, Edgemont, Hamptons, Hidden Valley	51.12606	-114.143158	Petro-Canada	51.128068	-114.138057	Gas Station
	Dalhousie, Edgemont, Hamptons, Hidden Valley	51.12606	-114.143158	Edgemont City	51.126473	-114.138997	Asian Restaurant
	Dalhousie, Edgemont, Hamptons, Hidden Valley	51.12606	-114.143158	Friends Cappuccino Bar & Bake Shop	51.126370	-114.138676	Café
	Dalhousie, Edgemont, Hamptons, Hidden Valley	51.12606	-114.143158	Mac's	51.128309	-114.137902	Convenience Store
	Forest Lawn, Dover, Erin Woods	51.03180	-113.978600	Bonasera Pizza And Sports Bar	51.029893	-113.982543	Bar

Figure 4: Calgary neighborhood venues retrieved utilizing the Foursquare API

Linking this dataset back to the business problem of determining a suitable location for a restaurant, I filtered this dataset to only include venues that included the term 'Restaurant'.

I utilized the 'group-by' data analysis method to understand how many different 'Restaurant' venues there are in the city of Calgary, and how they are distributed among the various neighborhoods. I utilized a bar graph visual to enhance my understanding.

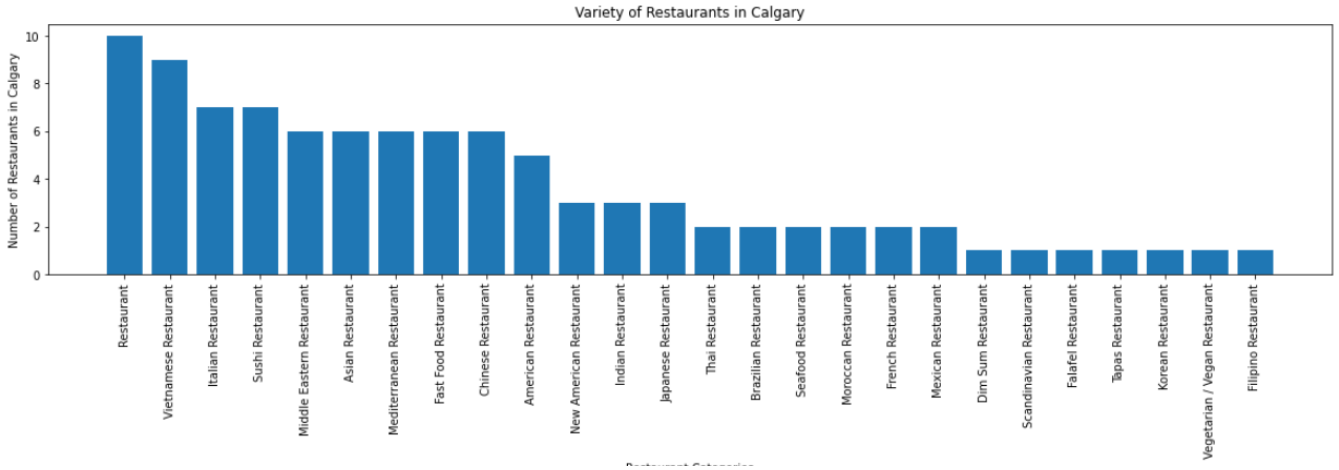

Figure 5: Variety and number of restaurants in all Calgary neighborhoods, as retrieved by the Foursquare API

This data was also grouped by Neighborhood name, to provide an understanding of the number of existing restaurants in each neighborhood.

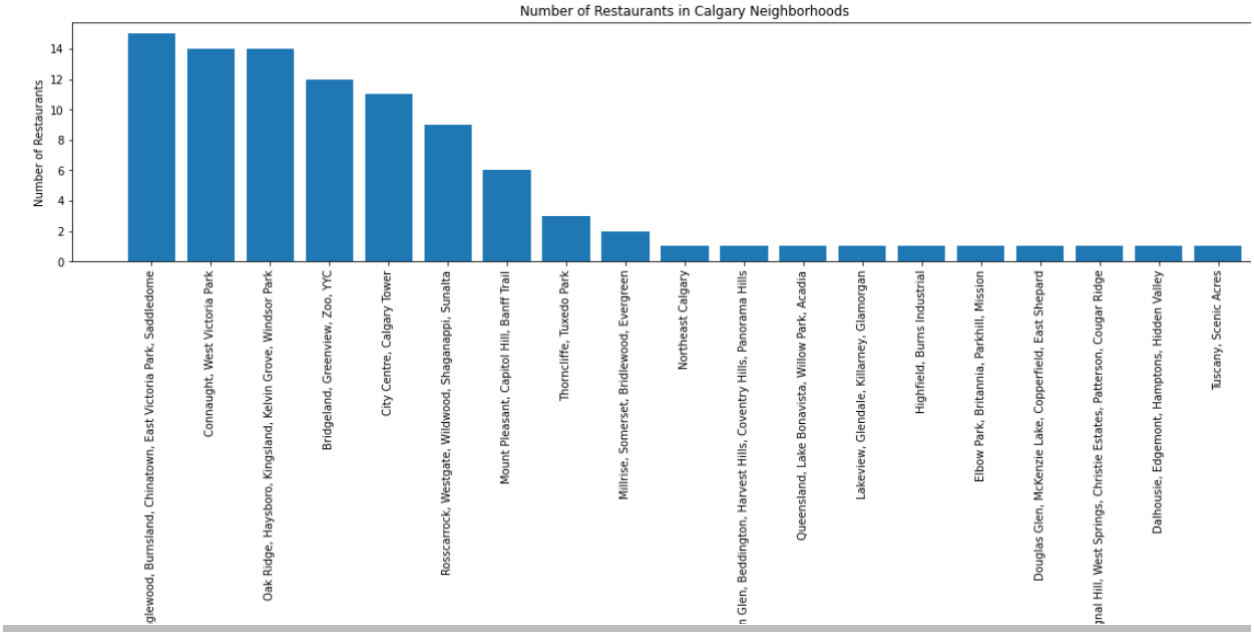

Figure 6: Number of existing restaurants in each Calgary neighborhood, as retrieved by the Foursquare API

Further examining the restaurants in each neighborhood, the dataset was further refined to display the frequency of each type of restaurant in each community. This information is critical to understand when answering both *Where to open a new restaurant* and *What type of restaurant to open.*

The one-hot encoding methodology was applied in order to sort the dataset. The five most popular restaurant types in each neighborhood were displayed. Figure 7 below represents an example for the communities of Bridgeland and Greenview, which are treated as the same neighborhood area due to close proximity and identical postal codes:

```
----Bridgeland, Greenview, Zoo, YYC----
                       venue  freq
0     Vietnamese Restaurant  0.19
1                 Restaurant  0.11
2           Sushi Restaurant  0.11
3  Mediterranean Restaurant  0.11
4         Chinese Restaurant  0.11
```

Figure 7: Example of the five most popular types of restaurants in one Calgary neighborhood area

The Clustering machine learning technique was utilized to group Calgary neighborhoods based on the top ten most popular restaurants in each neighborhood. Five cluster categories were used for simplicity.

Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	
Dalhousie, Edgemont, Hamptons, Hidden Valley	51.12606	-114.143158	2	Thai Restaurant	Asian Restaurant	Middle Eastern Restaurant	Indian Restaurant	Vietnamese Restaurant	Japanese Restaurant	R
Forest Lawn, Dover, Erin Woods	51.03180	-113.978600	3	Vietnamese Restaurant	Indian Restaurant	Mediterranean Restaurant	Japanese Restaurant	Asian Restaurant	Brazilian Restaurant	R
Montgomery, Bowness, Silver Springs, Greenwood	51.08090	-114.161600	1	Mexican Restaurant	Restaurant	Fast Food Restaurant	Vietnamese Restaurant	Japanese Restaurant	Asian Restaurant	R
Lynnwood Ridge, Ogden, Foothills	50.98780	-114.000100	0	Vietnamese Restaurant	Asian Restaurant	Sushi Restaurant	Seafood Restaurant	Japanese Restaurant	Brazilian Restaurant	R

Figure 8: Cluster labels applied to dataset of most popular restaurant types in Calgary neighborhoods

The dataset was later separated by cluster, in order to try and understand the data attributes that make up each cluster group:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
19	Lynnwood Ridge, Ogden, Foothills Industrial, G...	Vietnamese Restaurant	Asian Restaurant	Sushi Restaurant	Seafood Restaurant	Japanese Restaurant	Brazilian Restaurant	Chinese Restaurant	Dim Sum Restaurant	Falafel Restaurant	Fast Food Restaurant
20	Rosscarrock, Westgate, Wildwood, Shaganappi, S...	Restaurant	Italian Restaurant	American Restaurant	Asian Restaurant	Fast Food Restaurant	New American Restaurant	Middle Eastern Restaurant	Vegetarian / Vegan Restaurant	Chinese Restaurant	Dim Sum Restaurant
28	Bridgeland, Greenview, Zoo, YYC	Vietnamese Restaurant	Chinese Restaurant	Sushi Restaurant	Mediterranean Restaurant	Restaurant	Moroccan Restaurant	Fast Food Restaurant	Italian Restaurant	American Restaurant	Japanese Restaurant
37	Inglewood, Burnsland, Chinatown, East Victoria...	Sushi Restaurant	Restaurant	Mediterranean Restaurant	Asian Restaurant	Brazilian Restaurant	French Restaurant	Vietnamese Restaurant	Chinese Restaurant	Fast Food Restaurant	Filipino Restaurant

Figure 9: Cluster 0

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
11	Montgomery, Bowness, Silver Springs, Greenwood	Mexican Restaurant	Restaurant	Fast Food Restaurant	Vietnamese Restaurant	Japanese Restaurant	Asian Restaurant	Brazilian Restaurant	Chinese Restaurant	Dim Sum Restaurant	Falafel Restaurant

Figure 10: Cluster 1

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
2	Dalhousie, Edgemont, Hamptons, Hidden Vallev	Thai Restaurant	Asian Restaurant	Middle Eastern Restaurant	Indian Restaurant	Vietnamese Restaurant	Japanese Restaurant	Brazilian Restaurant	Chinese Restaurant	Dim Sum Restaurant	Falafel Restaurant

Figure 11: Cluster 2

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
10	Forest Lawn, Dover, Erin Woods	Vietnamese Restaurant	Indian Restaurant	Mediterranean Restaurant	Japanese Restaurant	Asian Restaurant	Brazilian Restaurant	Chinese Restaurant	Dim Sum Restaurant	Falafel Restaurant	Fast Food Restaurant

Figure 12: Cluster 3

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
29	Lakeview, Glendale, Killarney, Glamorgan	Chinese Restaurant	Scandinavian Restaurant	Middle Eastern Restaurant	Italian Restaurant	Vietnamese Restaurant	Japanese Restaurant	Asian Restaurant	Brazilian Restaurant	Dim Sum Restaurant	Falafel Restaurant

Figure 13: Cluster 4

Again utilizing the Folium library, these clusters were displayed on a map of Calgary to gain an appreciation for any geographical influences on this dataset. Note: red circles represent Cluster 0; purple circles represent Cluster 1, blue circles represent Cluster 2, light green circles represent Cluster 3 and yellow circles represent Cluster 4.

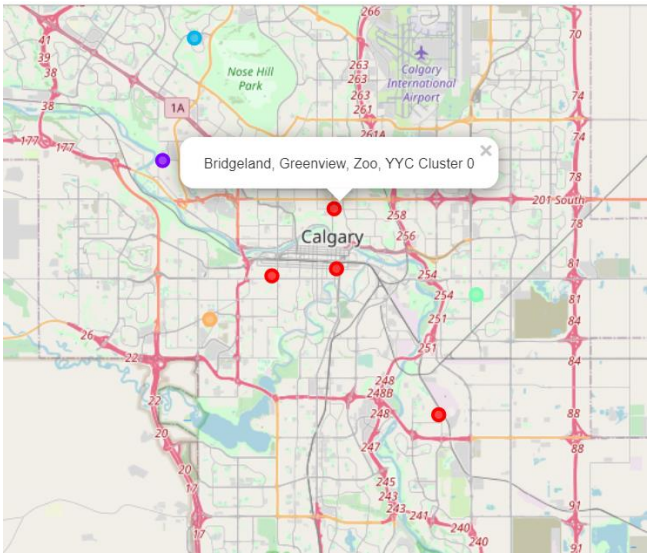

Figure 14: Clustered map of the most popular restaurants among Calgary neighborhoods

Looking to gain a better appreciation for the citizens living each of these neighborhoods in Calgary, the Calgary Census data was scraped from *Great-News.ca:*

Community	Newsletter Name	Median Household Income	Median Age	Population 2014	Dwellings 2014	City Quadrant	Median Home Sale Price
Abbeydale	-	$55,345	34	6071	2031	SE	$305,000
Acadia	Acadia	$46,089	42	10969	5067	SE	$447,000
Albert Park / Radisson Heights	-	$38,019	37	6529	2936	SE	$349,900
Altadore	The Source	$53,786	37	9518	4537	SW	$925,000
Applewood Park	-	$65,724	33	6864	2228	SE	$380,000
...	...	...	...	...	...	...	...
Willow Park	Your Willow Park Maple Ridge	$63,588	45	5440	2284	SE	$705,000
Windsor Park	The Elbow Scene	$39,425	37	4417	2453	SW	$675,000
Winston Heights/Mountview	Community Matters	$41,065	42	3845	1899	NE	$625,000

Figure 15: Sample of raw census data scraped from Great-news.ca

Fields identified for further investigation and cleaning were *Median Household Income, Median Age, Population* and *Dwellings*. For simplicity, this data was cleaned and grouped by city quadrant:

City Quadrant	Median Household Income	Median Age	Population 2014	Dwellings 2014
NE	56608.250000	35.850000	9006.500000	3211.200000
NW	66574.647059	38.627451	7075.254902	2734.215686
SE	66750.771429	37.885714	7009.514286	2724.314286
SW	75388.391892	40.540541	5059.202703	2207.405405

Figure 16: Cleaned Calgary census data, sorted by city quadrant

A number of pie charts were created from this data:

Median Household Income among City Quadrants

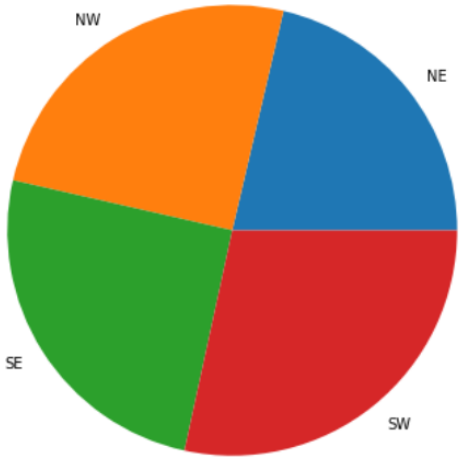

Figure 17: Median Household Income distribution among city quadrants

2014 Population among City Quadrants

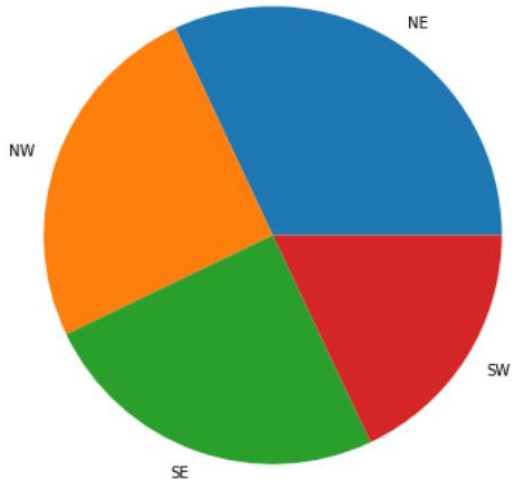

Figure 18: 2014 Population among City Quadrants

A bar chart was also produced to compare the Median Age of residents among City Quadrants:

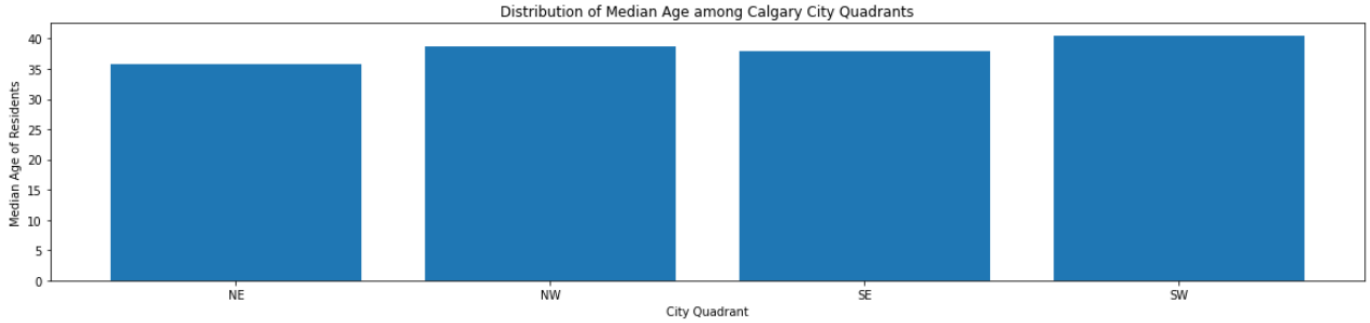

Figure 19: Comparison of median resident age among city quadrants

Results

From the data presented and analyzed, it is clear that there is a demand for Category 0 – type restaurants, which are most likely to serve Eastern-Asian cuisine such as Vietnamese, Chinese, Korean cuisines, or sushi options. As demonstrated in Figure 14, four of the eight clustered neighborhoods were classified as Category 0-type restaurants.

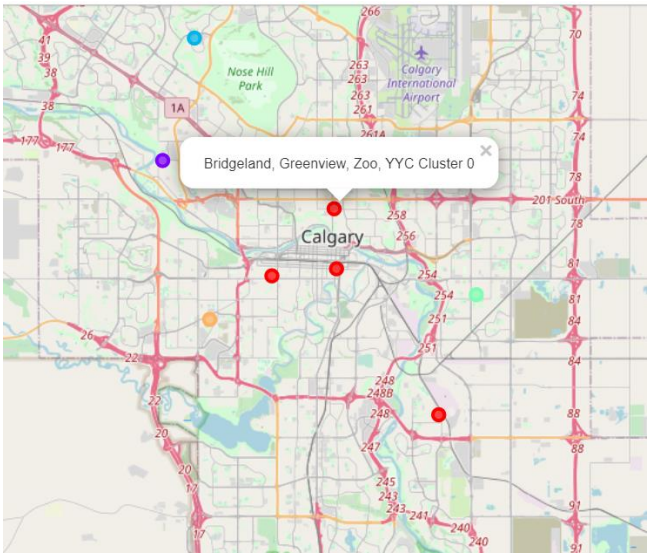

Figure 20: Clustered map of the most popular restaurants among Calgary neighborhoods

There is an obvious lack of restaurants in the North-East quadrant of the city. This is an odd finding, as the North-East quadrant also includes the greatest percentage of the city population.

From these results, it has been recommended that an Eastern-Asian cuisine restaurant is opened in North East Calgary.

Discussion

Most of the neighborhoods in Calgary that had FourSquare restaurant data associated with them appear to have been categorized into Cluster 0. Based off of Figure 9, this cluster appears to show a high frequency of eastern-Asian cuisine options, including Vietnamese, Asian, Sushi and Chinese. Fast food restaurant options are typically among some of the lesser frequency restaurants in these neighborhoods, if present at all.

Analyzing Cluster 0 neighborhoods on Figure 14 (red circles), it becomes apparent that not only are Cluster 0 restaurants the most popular within and among neighborhoods in Calgary, but that they are centralized around the downtown core. This has been attributed to a few findings: a) the downtown core has the greatest restaurant density per area, making it a popular location for existing restaurants of any type, and b) the downtown core contains different ethical towns, including a named China Town and Korea Town.

Although there are many neighborhoods in North East Calgary, as confirmed by Figure 3, none of these neighborhoods have been identified as having popular restaurants, as identified by FourSquare.

For example, the neighborhood identified as Northeast Calgary contains only one restaurant venue:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
188	Northeast Calgary	51.1494	-114.0019	Pacini	51.150375	-114.000692	American Restaurant
189	Northeast Calgary	51.1494	-114.0019	Acclaim Hotel	51.150384	-114.000140	Hotel

Figure 21: All venues identified in the neighborhood of Northeast Calgary

No restaurant venues were found in the north-eastern neighborhoods of Forest Lawn, Dover, or Erin Woods:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
4	Forest Lawn, Dover, Erin Woods	51.0318	-113.9786	Bonasera Pizza And Sports Bar	51.029893	-113.982543	Bar
5	Forest Lawn, Dover, Erin Woods	51.0318	-113.9786	7-Eleven	51.029839	-113.982060	Convenience Store
6	Forest Lawn, Dover, Erin Woods	51.0318	-113.9786	Foggy Gorilla Vaping Co.	51.030038	-113.972642	Smoke Shop

Figure 22: All venues identified in the north-eastern neighborhoods of Forest Lawn, Dover, and Erin Woods

Median Household Income was thought to shed some light on how financially well-off residents of different neighborhoods are, and consequently how likely or unlikely they may be to eat out at restaurants (driving the demand chain to open more restaurants). As demonstrated in Figure 17, the Median Household Income does not greatly differ across the city quadrants. The neighborhoods in North East Calgary are slightly less well-off financially, while the neighborhoods in South West Calgary are primarily more financially stable. This may indicate that residents in the SW neighborhoods are more likely to eat out at restaurants than are residents in NE neighborhoods.

It is apparent from Figure 18 that the North East neighborhoods are home to a higher percentage of the city's population. As the data used in this analysis was collected in 2014, it is assumed that the city has experienced even population growth among all of its neighborhoods.

Figure 19 does not present any obvious contrasts in the median age of residents living in the various quadrants of the city. For the purpose of this analysis, this data is deemed to not have an affect on the recommendation provided.

Conclusion

The Wikipedia Calgary neighborhood dataset, Calgary census dataset, and Foursquare API venue dataset provided a number of insights into Calgary's current population and appetite to eat out at restaurants. The existing restaurants and their locations were evaluated and compared, and the most popular restaurant venue types were discovered.

Through visual data analysis and clustering machine learning techniques, the user concluded that there is a strong business opportunity to open a restaurant of East-Asian cuisine in any Northeastern Calgary neighborhood.

References

[1]: Wikipedia: "*List of postal codes of Canda: T"*, Table 2: *"Alberta – 157 FSAs"*. Retrieved: Nov 20, 2020. URL: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_T

[2]: Great News: "*Calgary Community Demographics".* Retrieved Nov 23, 2020. URL: https://great-news.ca/demographics/