

The Battle of the Neighborhoods

-Where to start a Sushi restaurant in Calgary, Alberta?

1. Business problem

Calgary is a city in the western Canadian province of Alberta and it is situated about 80 km east of the Canadian Rockies. Calgary has a population of 1.3 million, making it Alberta's biggest city and the second largest in western Canada after only Vancouver.

As a multi-cultural city, Calgary has numerous cuisines from all over the world. Personally, I have been in Calgary for 15 years and love to try different cuisines and experience different flavours. Among them, I love the Sushi best.

Thus, the goal of this project is to study the neighborhoods in Calgary and determine the right places to start a Sushi restaurant. The project can be very interesting to many business owners who are looking to open up a new Sushi restaurant in Calgary.

To solve this problem, we need to acquire enough demographic data and food venue data of each neighborhoods and analyze these data using some machine learning algorithms I just learned to provide recommendations for interested business owner and entrepreneurs.

2. Data Acquisition and ways to solve the problem

2.1 Data sources

To collect all the data, we will start with the community demographics data. I will use two data sources to get these data. Pandas web scraper library will be used to get these publicly available data from the web.

1. Neighborhood and area from Wikipedia

It seems not easy to get the community population density data after many unsuccessful tries and I decided to calculate it using the population and the area of a community. The below link provides the community and its area(km²) in Calgary:

https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Calgary

	Community	Area
0	Abbeydale	1.7
1	Acadia	3.9
2	Albert Park / Radisson Heights	2.5
3	Altadore	2.9
4	Alyth/Bonnybrook	3.8
5	Applewood Park	1.6
6	Arbour Lake	4.4
7	Aspen Woods	3.8
8	Auburn Bay	4.5
9	Aurora Business Park	2.4

2. Neighborhood and demographics data from great-news.ca

Great-news.ca has a fairly complete dataset for Calgary community demographics data. The link here: <https://great-news.ca/demographics/>

	Community	Median Household Income	Population	Area	PopulationDensity
0	Abbeydale	55345.0	6071	1.7	3571.176471
1	Acadia	46089.0	10969	3.9	2812.564103
2	Albert Park / Radisson Heights	38019.0	6529	2.5	2611.600000
3	Altadore	53786.0	9518	2.9	3282.068966
4	Applewood Park	65724.0	6864	1.6	4290.000000
5	Arbour Lake	70590.0	10987	4.4	2497.045455
6	Aspen Woods	133939.0	7496	3.8	1972.631579
7	Auburn Bay	84350.0	11127	4.5	2472.666667
8	Banff Trail	49996.0	4204	1.5	2802.666667
9	Bankview	32474.0	5416	0.7	7737.142857

3. Geo data from Geocoder

We also need coordinates for each neighborhood to explore its venues with foursquare API. We will use geocoder library to get all the coordinates of each neighborhood.

	Community	Median Household Income	Population	Area	PopulationDensity	Latitude	Longitude
0	Abbeydale	55345.0	6071	1.7	3571.176471	51.05976	-113.92546
1	Acadia	46089.0	10969	3.9	2812.564103	50.97227	-114.05882
2	Albert Park / Radisson Heights	38019.0	6529	2.5	2611.600000	51.04200	-113.99683
3	Altadore	53786.0	9518	2.9	3282.068966	51.01601	-114.10558
4	Applewood Park	65724.0	6864	1.6	4290.000000	51.04544	-113.92513
5	Arbour Lake	70590.0	10987	4.4	2497.045455	51.13364	-114.20307
6	Aspen Woods	133939.0	7496	3.8	1972.631579	51.04519	-114.21160
7	Auburn Bay	84350.0	11127	4.5	2472.666667	50.88976	-113.96397
8	Banff Trail	49996.0	4204	1.5	2802.666667	51.07472	-114.11297
9	Bankview	32474.0	5416	0.7	7737.142857	51.03412	-114.10044

2.2 Food venues in neighborhood

With the above community dataset, we are ready to explore all the venues of a neighborhood using foursquare API.

Since our goal is to identify a Sushi restaurant location, we will focus on the food venues within 1500m (a reasonable walking distance) of a community.

To set the section parameter for foursquare API, we will use keyword **section** and follow this:

```
url =
'https://api.foursquare.com/v2/venues/explore?&client_id={ }&client_secret={ }&v={ }&ll={ },{ }&radius={ }&section={ }&limit={ }'.format(CLIENT_ID, CLIENT_SECRET, VERSION, lat, lng, radius, SECTION,LIMIT)
```

	Community	Community Latitude	Community Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Abbeydale	51.05976	-113.92546	Atlas Pizza and Sports Bar	51.052481	-113.941859	Pizza Place
1	Abbeydale	51.05976	-113.92546	A&W	51.068291	-113.933571	Fast Food Restaurant
2	Abbeydale	51.05976	-113.92546	Subway	51.059239	-113.934423	Sandwich Place
3	Abbeydale	51.05976	-113.92546	Subway	51.069623	-113.932907	Sandwich Place
4	Abbeydale	51.05976	-113.92546	Subway	51.052786	-113.942449	Sandwich Place

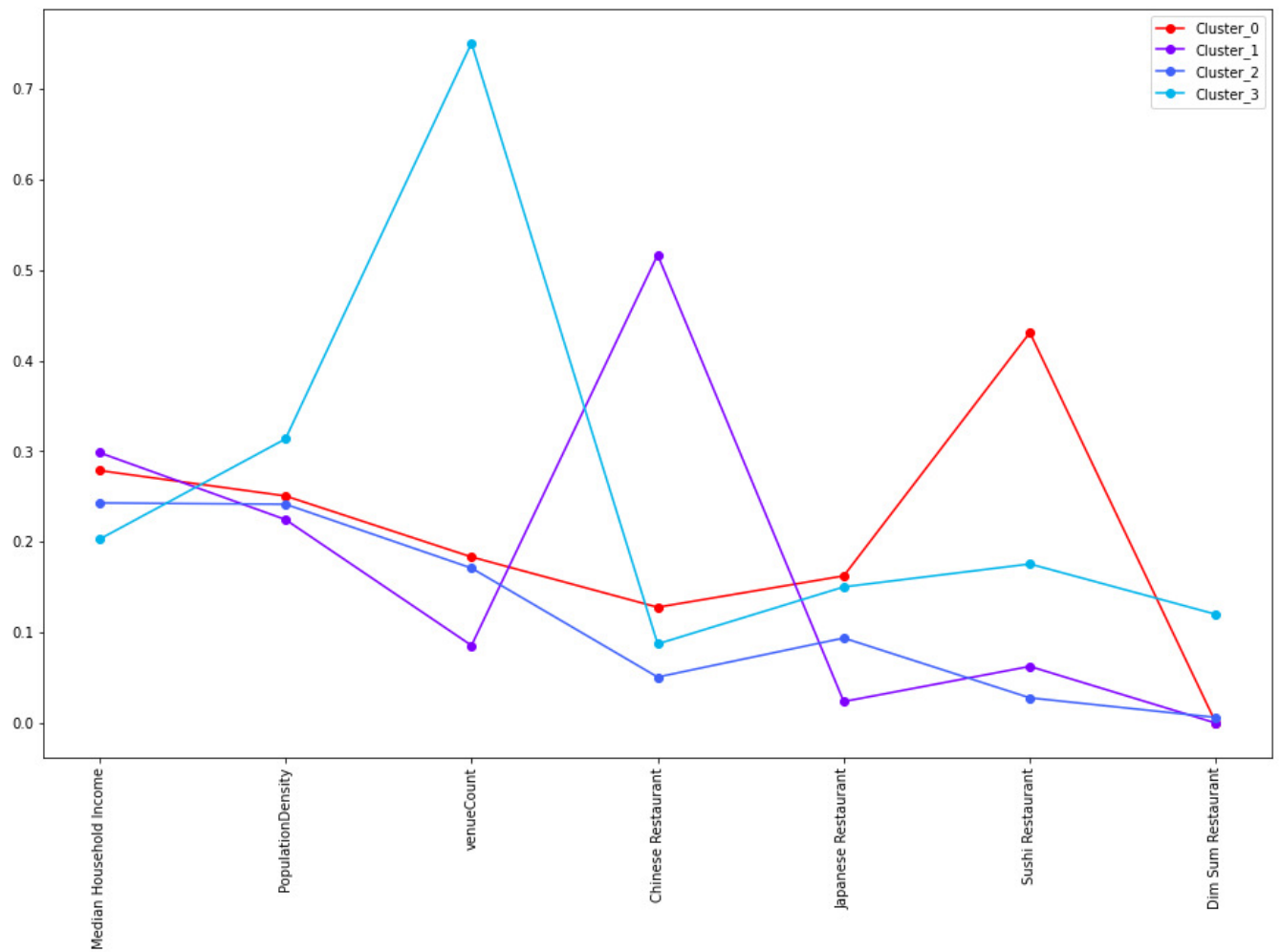
2.3 Use the data to solve the business problem

To identify the best location of a Sushi restaurant, we need to analyze the data, such as, median household income, population density, total venues (which can largely indicate if the neighborhood is fully developed, well developed, or underdeveloped).

Since the competition to a Sushi restaurant is mostly from East Asian cuisines, we will also analyze the community coverage of other East Asian restaurants, such as Chinese restaurants, Japanese restaurants, Dim Sum restaurants, and etc.

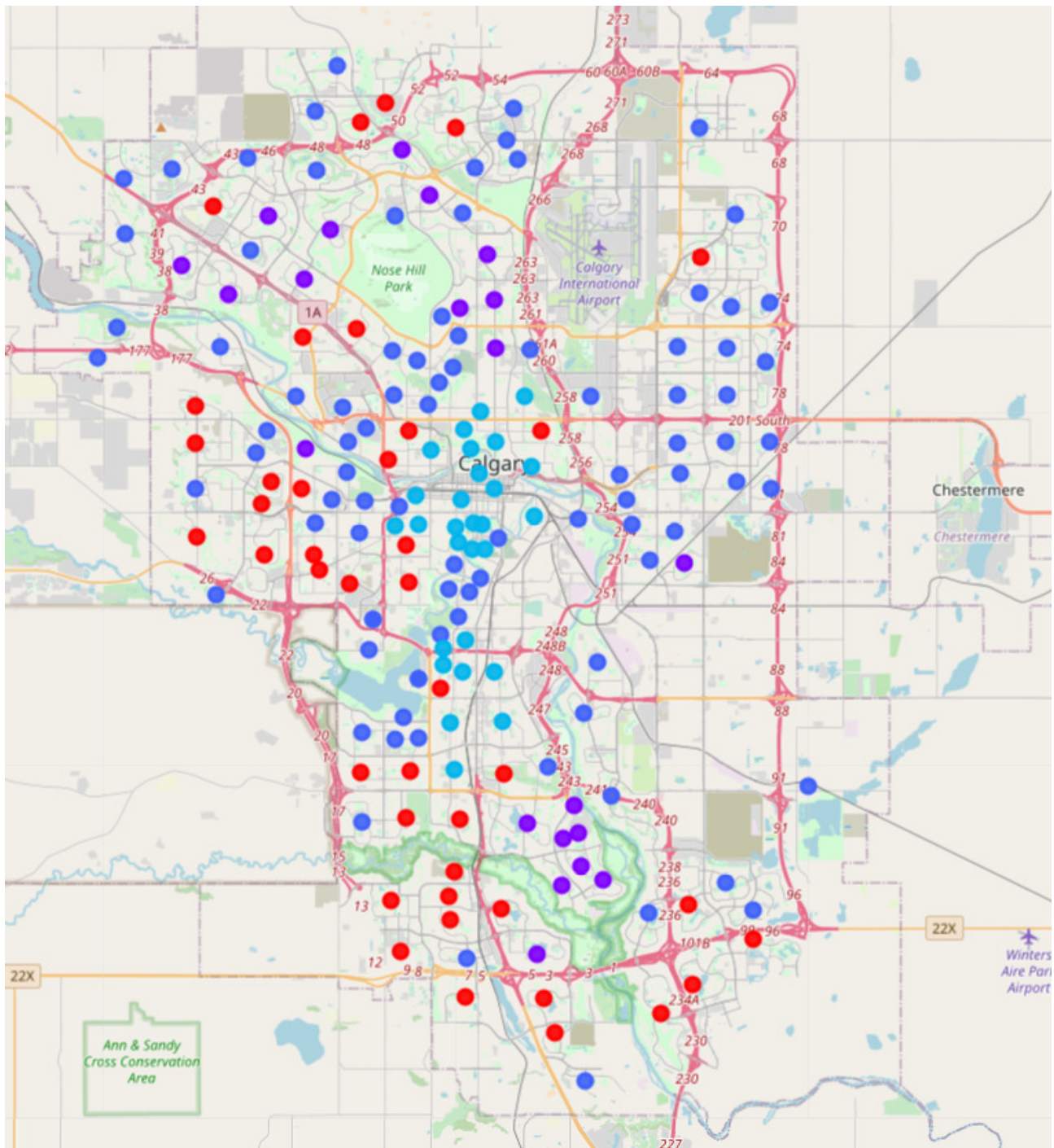
	Community	Median Household Income	PopulationDensity	venueCount	Chinese Restaurant	Japanese Restaurant	Sushi Restaurant	Dim Sum Restaurant
0	Abbeydale	55345.0	3571.176471	12.0	0.083333	0.0	0.00	0.00
1	Acadia	46089.0	2812.564103	50.0	0.020000	0.0	0.04	0.02
2	Albert Park / Radisson Heights	38019.0	2611.600000	27.0	0.000000	0.0	0.00	0.00
3	Altadore	53786.0	3282.068966	20.0	0.050000	0.0	0.15	0.00
4	Applewood Park	65724.0	4290.000000	8.0	0.000000	0.0	0.00	0.00

With all the data of each neighborhood, we will employ an unsupervised machine learning algorithm to cluster all the neighborhoods in Calgary and analyze the produced clusters to make recommendations for starting a Sushi restaurant.



Cluster graphs show how each cluster performs on each measuring matrix on X-axis

We will also show the clustering result on the map.



By analyzing the clustering result, we could provide different recommendations for different business goals. For example, a low-end fast food like Sushi restaurant could be started at a busy and downtown hub area; in contrast, a high-end Sushi restaurant might be ideally situated in a less crowded rural community with a high median household income.