

## **IBM Data Science Capstone Project**

### **Scoping out the most desirable Postal code to open a new restaurant in Calgary, AB**

Ian Austin

September 2021

#### **Introduction**

Opening a restaurant, or any business for that matter, is one of the bigger decisions that a person can make in their lives. It is a huge decision, and decisions made during set up can make or break the business. An example of such decisions includes:

- Restaurant theme
- Restaurant location
- Timing of opening

For this study, we will focus on location, namely within which of the postal code areas (i.e.: T2H vs T1Y), and various other parameters (population and income measures) to create clusters of postal code areas to assist in making this decision a little easier.

According to the following CNBC article, approximately 60% of all restaurants fail within the first year, and 80% fail prior to their 5<sup>th</sup> (<https://www.cnbc.com/2016/01/20/heres-the-real-reason-why-most-restaurants-fail.html>). The number one reason cited is choice of location – in this case, that choice really does seem to be ‘do or die’.

In the city of Calgary, Alberta, a shift in overall person and business traffic flow since the oil industry downturn (beginning in late 2014) has further complicated this decision. Historically, the downtown core was the hot spot for openings of new restaurants, with revenue fed by a seemingly never-ending flow of clients with large expense accounts. That was when office vacancy rates were approximately 3%. Today they are hovering around 30%, with at least 100,000 fewer people making the daily trip to the downtown core.

Additionally, there have been other forces at play that complicate the overall picture. Some have been sparked by the COVID-19 pandemic, others by general changes to the City of Calgary’s overall community planning philosophy. These considerations include:

- Community Population/Density: Calgary is historically a city of suburb communities, full of single-family homes. In recent years, due to the ever expanding ‘urban sprawl’, there has been an increasing push for multi-family densification. These demographic changes should be considered in any analysis
- Walkability/ride-ability: another push for increased walkability of communities, and an increasing number of cycling lanes on community roads (with the corresponding reduction in the ease of driving and finding parking) should be considered when evaluating communities to open in, especially when considering serving alcohol

It should be noted that cost of operation is certainly a big consideration. However, by community area, there is no repository for property tax and rental rates published within Calgary. For this reason, the

analysis will focus solely on the potential for revenue generation, with the assumption that a savvy business owner will be able to control costs to achieve profitability.

## **The Data**

### **Data Sources**

To create this evaluation/model, the following data will be used to cluster neighborhoods:

- Calgary Postal code (and contained neighborhood) data, complete with the geographic center of each postal code area. Note there are 34 postal code areas located within the City of Calgary proper ([https://en.wikipedia.org/wiki/List\\_of\\_postal\\_codes\\_of\\_Canada:\\_T](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_T))
- Population by postal code area: taken from the 2016 census analysis (StatsCan) <https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/hlt-fst/pd-pl/Table.cfm?Lang=Eng&T=1201&SR=1&S=22&O=A&RPP=9999&PR=0>
- Income data by community area: although I was unable to get income data per postal code area, I was able to get the data from the Calgary Mortgage and Housing corporation (CMHC). Since it was by area, I had to assign each community a postal code and group by postal codes to average the income per community. Any postal codes with no reported income metrics were assigned the median of all the Calgary communities/postal code areas

For the analysis, I chose to focus on after tax income, so kept only median household income after tax. A link to the data follows: <https://www03.cmhc-schl.gc.ca/hmip-pimh/en/TableMapChart/TableMatchingCriteria?GeographyType=MetropolitanMajorArea&GeographyId=0140&CategoryLevel1=Population%2C%20Households%20and%20Housing%20Stock&CategoryLevel2=Household%20Income&ColumnField=HouseholdIncomeRange&RowField=Neighbourhood&SearchTags%5B0%5D.Key=Households&SearchTags%5B0%5D.Value=Number&SearchTags%5B1%5D.Key=Statistics&SearchTags%5B1%5D.Value=AverageAndMedian>

- Community Walkability score: since Calgary publishes this index per community, communities will need to be grouped into postal code areas, like what was done for the income analysis, and a median walkability score per postal code area will need to be stored. A link follows: <https://www.walkscore.com/CA-AB/Calgary>
- Foursquare API: the Foursquare API will be utilized to provide information on the number and type of restaurants that are within a specified distance from the center of the postal code area. This will be used to determine the number and type of restaurants already present, and as part of the information used to cluster the communities by their potential suitability for a new restaurant

## **Data Cleaning & Preparation**

Asides from the data retrieved, all data for the analysis was web scraped from the sources listed above. There were several problems when attempting to integrate the data sources into a single final data frame for analysis, including:

- Some of the data (income and walk score) were collected in a different manner than the postal code data, necessitating grouping communities within postal codes and taking median values of the measures
- Some postal codes were not represented in the income analysis, due to the data being scrapped together from different data sources. To compensate, these postal codes were assigned median values of all the other postal codes for income

For a future, more detailed analysis, it would help to look for additional (probably pay-for analysis) data, grouped by neighborhood, with a consistent list of neighborhoods.

An additional issue was the time frame of the collected data. Since it came from disparate sources, some data (Walk Score) was more up to date than data that came from Federal Government sources (Income Data, last analyzed for the 2016 Census). There is currently no way around this, when scraping from free/public data sets, other than assuming that on an area's basis (tens of thousands of residents) that the median of the data will not change greatly over short periods of time.

The data was first scraped from the various sources and in sequence combined, with the resulting single data frame for final analysis. Data was scraped and combined in the following order:

- List of Calgary postal codes and neighborhoods
- Population by postal code
- Income data by neighborhood and postal code
  - o Note that for this portion, communities were grouped into specific postal codes (unfortunately, this was a manual process due to data availability and layout) and the median value per postal code was taken for analysis
- Walk Scores per neighborhood and postal code
  - o Note that for this portion, communities were grouped into specific postal codes (unfortunately, this was a manual process due to data availability and layout) and the median value per postal code was taken for analysis
- Foursquare API Data containing restaurants
  - o The data was collected within 10 km of the center of each postal code to ensure every restaurant was included for analysis. Note this does not mean we believe 10 km to be an acceptable walking distance

The resulting final data frame, including counts of restaurant type (which was not further analyzed due to the scoping nature of the analysis) follows in Figure 1:

[41]:

	Postal Code	Community	Neighbourhood	Latitude	Longitude	Population (2016)	Total Private Dwellings (2016)	Median Household Income After Taxes	Walk Score	Venue Category	Count	Total Restaurants	Restaurants/1000 residents	Restaurants/1000 residences
0	T2A	Calgary	Penbrooke Meadows, Marlborough	51.04968	-113.96432	59641	22652	62824.5	50.4	American Restaurant	2	31	0.519777	1.368533
1	T2A	Calgary	Penbrooke Meadows, Marlborough	51.04968	-113.96432	59641	22652	62824.5	50.4	Asian Restaurant	3	31	0.519777	1.368533
2	T2A	Calgary	Penbrooke Meadows, Marlborough	51.04968	-113.96432	59641	22652	62824.5	50.4	Cajun / Creole Restaurant	1	31	0.519777	1.368533
3	T2A	Calgary	Penbrooke Meadows, Marlborough	51.04968	-113.96432	59641	22652	62824.5	50.4	Chinese Restaurant	1	31	0.519777	1.368533
4	T2A	Calgary	Penbrooke Meadows, Marlborough	51.04968	-113.96432	59641	22652	62824.5	50.4	Falafel Restaurant	1	31	0.519777	1.368533
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
508	T2Z	Calgary	Douglas Glen, McKenzie Lake, Copperfield, East...	50.90230	-113.98730	68438	24669	101028.0	31.4	Middle Eastern Restaurant	1	25	0.365294	1.013418
509	T2Z	Calgary	Douglas Glen, McKenzie Lake, Copperfield, East...	50.90230	-113.98730	68438	24669	101028.0	31.4	Restaurant	6	25	0.365294	1.013418
510	T2Z	Calgary	Douglas Glen, McKenzie Lake, Copperfield, East...	50.90230	-113.98730	68438	24669	101028.0	31.4	Seafood Restaurant	1	25	0.365294	1.013418
511	T2Z	Calgary	Douglas Glen, McKenzie Lake, Copperfield, East...	50.90230	-113.98730	68438	24669	101028.0	31.4	Sushi Restaurant	4	25	0.365294	1.013418
512	T2Z	Calgary	Douglas Glen, McKenzie Lake, Copperfield, East...	50.90230	-113.98730	68438	24669	101028.0	31.4	Vietnamese Restaurant	3	25	0.365294	1.013418

513 rows x 14 columns

**Figure 1: Final Data frame for Analysis**

## **Features/Parameters**

As previously mentioned, the analysis was inspired by changes to city design philosophy within Calgary, coupled with changes in traffic flows due to the COVID-19 pandemic (and some of the more permanent changes, such as work from home).

The parameters to evaluate, based on analysis were, on a postal code basis:

- Median After Tax Income
- Median Walk Score
- The current number of restaurants per 1000 residents (normalized to account for total restaurants per population)

## **Exploratory Data Analysis**

A short exploratory analysis was conducted to qualify and quantify a relationship between postal code and the various other variables:

- Median After Tax Income
- Median Walk Score
- The current number of restaurants per 1000 residents (normalized to account for total restaurants per population)

Separate bar charts were produced to examine relationships. They follow in Figures 2 through 4:

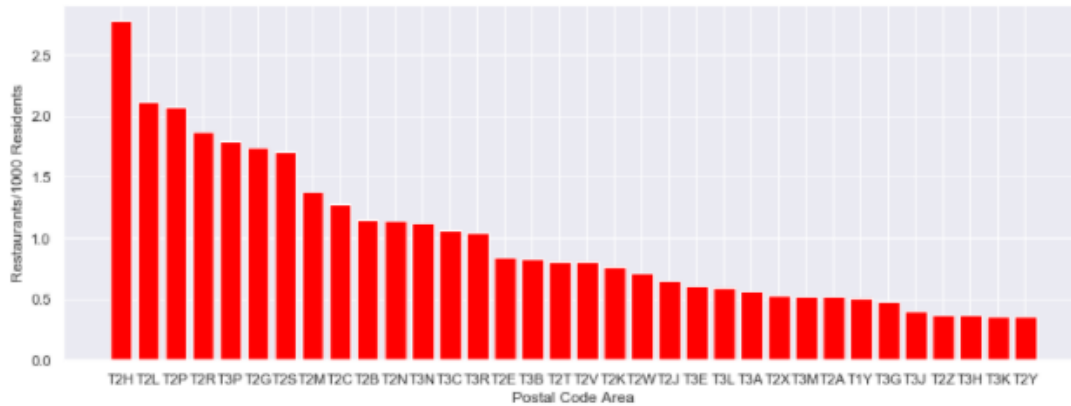


Figure 2: Restaurants/1000 Residents by Postal Code

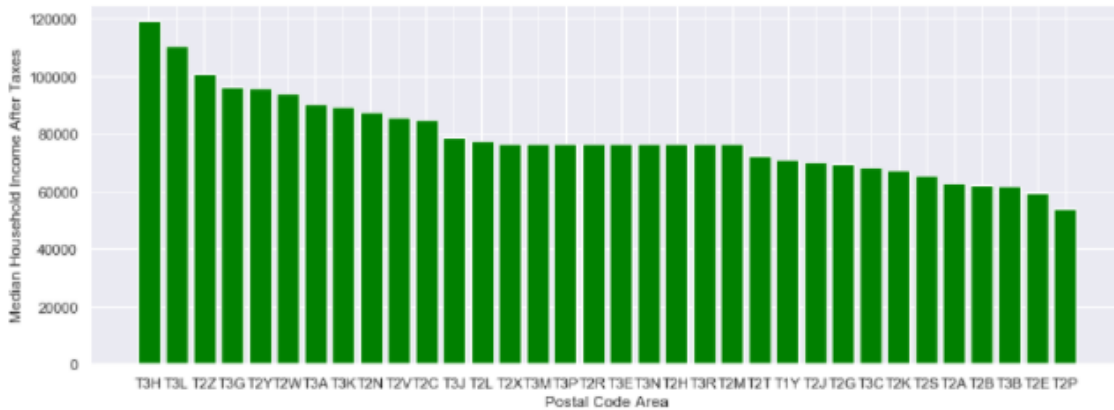


Figure 3: Median Income by Postal Code

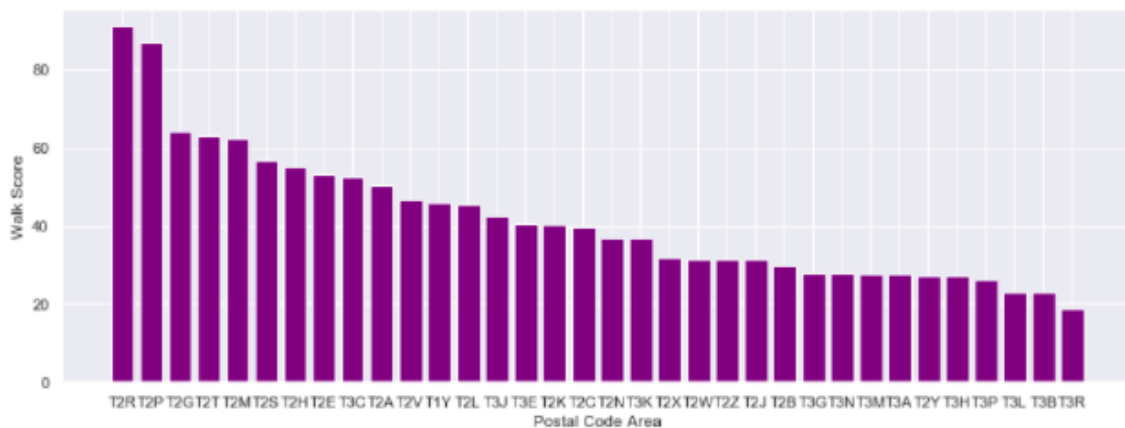


Figure 4: Median Walk Score by Postal Code

In examining the bar charts, it was apparent that visually there was a negative correlation between median income and restaurant density – the postal codes with higher incomes in general had fewer restaurants/1000 residents. Examining walkability, the correlation seemed less strong, at least visually. However, it could be observed that the number of restaurants in a postal code increased with increasing walkability.

Based on these observations, the analysis was taken further, and a correlation matrix was built to relate correlations between median income, restaurant density and walkability. It follows in Figure 5:

[47]:

	Median Household Income After Taxes	Walk Score	Restaurants/1000 residents
Median Household Income After Taxes	1.000000	-0.454394	-0.366446
Walk Score	-0.454394	1.000000	0.556149
Restaurants/1000 residents	-0.366446	0.556149	1.000000

**Figure 5: Correlation Matrix for Median Income, Restaurant Density and Walkability**

From the correlation matrix it was observed that the visual finds held true mathematically:

- Moderate negative correlation between Median Income and both Walkability and Restaurant Density
- Moderate positive correlation between Walkability and Restaurant Density

### **Modeling (Cluster Analysis)**

Given the type of analysis we're attempting to undertake:

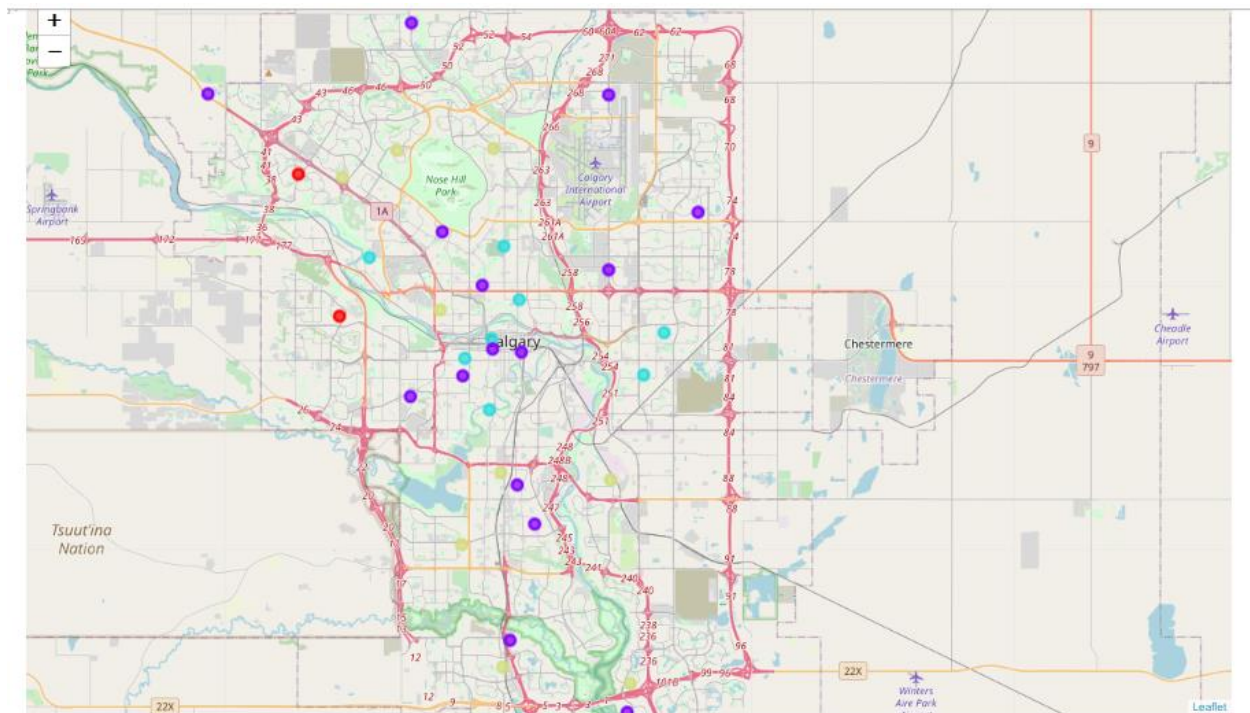
- Analysis: scope out the most promising postal codes in Calgary, AB to open a new restaurant in
- Target Data: None. It is a fresh analysis taking into certain chosen characteristics of each neighborhood/postal code, such as:
  - o Median After Tax Income
  - o Median Walk Score
  - o The current number of restaurants
  - o The population

It was decided that a clustering algorithm fit the analysis best. This way, we could cluster together postal codes based on the above measures, filter through them, and based on criteria determine what made a postal code desirable.

It should be noted that the lack of target data, for example whether a restaurant was successful or not, lead to the clustering conclusion. If the data was available for restaurant starts/failures in a community over a certain period, then a classification algorithm may have been more appropriate.

As mentioned previously, the data was be grouped into a single data frame for analysis. Multiple different cluster numbers were selected for evaluation, to determine the model that best describes the data. Analysis with 3 to 7 clusters were evaluated – those over 4 yielded too many clusters that were too small

and loosely descriptive. Those less than 4 had too few clusters and too many different postal codes per cluster, to the point that it again was difficult to develop a theme for the cluster. Thus, the number of clusters were set to 4 for analysis. Figure 6 shows a cluster map from the first analysis:



**Figure 6: Cluster Map – first cluster analysis**

### **First Analysis**

After the first analysis, the clusters were classified according to themes. The following themes were identified:

**Cluster 1:** West side of Calgary communities. On average higher income, lower walkability, and relatively low restaurant density. From this cluster of 2 postal codes, the T3H community was selected for further analysis.

[58]:

Cluster Labels	Postal Code	Community	Neighbourhood	Latitude	Longitude	Population (2016)	Total Private Dwellings (2016)	Median Household Income After Taxes	Walk Score	Total Restaurants	Restaurants/1000 residents	Restaurants/1000 residences	
11	0	T3H	Calgary	Discovery Ridge, Signal Hill, West Springs, Ch...	51.0566	-114.1815	71251	25098	119018.5	27.3	26	0.364907	1.035939
17	0	T3L	Calgary	Tuscany, Scenic Acres	51.1162	-114.2089	30168	10093	110614.0	23.0	18	0.596659	1.783414

**Cluster 2:** The largest cluster. Consisting largely of medium income communities, spread out all over the city. This cluster contains a full range of walk scores and in general medium – high restaurant density. None of the postal codes within this cluster scored particularly high in any measure and were in general middling across the board. The decision was made not to proceed with any postal codes from this cluster for further analysis

[59]:

Cluster Labels	Postal Code	Community	Neighbourhood	Latitude	Longitude	Population (2016)	Total Private Dwellings (2016)	Median Household Income After Taxes	Walk Score	Total Restaurants	Restaurants/1000 residents	Restaurants/1000 residences	
7	1	T3E	Calgary	Lakeview, Glendale, Killarney, Glamorgan	51.0227	-114.1342	41415	18806	76644.0	40.700000	25	0.603646	1.329363
8	1	T2G	Calgary	Inglewood, Burnside, Chinatown, East Victoria...	51.0415	-114.0599	14370	8593	69418.0	64.200000	25	1.739736	2.909345
10	1	T2H	Calgary	Highfield, Burns Industrial	50.9857	-114.0631	9372	4395	76644.0	55.000000	26	2.774221	5.915813
12	1	T2J	Calgary	Queensland, Lake Bonavista, Willow Park, Acadia	50.9693	-114.0514	46394	18714	70152.0	31.400000	30	0.646635	1.603078
13	1	T3J	Calgary	Martindale, Taradale, Falconridge, Saddle Ridge	51.0999	-113.9422	77605	20910	78960.0	42.166667	31	0.399459	1.482544
16	1	T2L	Calgary	Brentwood, Collingwood, Nose Hill	51.0917	-114.1127	14214	6539	77548.0	45.333333	30	2.110595	4.587857
18	1	T2M	Calgary	Mount Pleasant, Capitol Hill, Banff Trail	51.0696	-114.0862	21086	9925	76435.0	62.285714	29	1.375320	2.921914
19	1	T3M	Calgary	Cranston, Auburn Bay, Mahogany	50.8902	-113.9892	40268	14202	76644.0	27.666667	21	0.521506	1.478665
21	1	T3N	Calgary	Northeast Calgary	51.1494	-114.0019	12483	4042	76644.0	28.000000	14	1.121525	3.463632
23	1	T3P	Calgary	Symons Valley	51.1793	-114.1333	15578	5038	76644.0	26.000000	28	1.797407	5.557761
24	1	T2R	Calgary	Connaught, West Victoria Park	51.0426	-114.0791	13879	10715	76644.0	91.000000	26	1.873334	2.426505
25	1	T3R	Calgary	Northwest Calgary	51.1497	-114.2695	23866	8045	76644.0	18.666667	25	1.047515	3.107520
27	1	T2T	Calgary	South Calgary (Altdore / Bankview / Richmond)	51.0316	-114.0994	29706	15860	72462.5	63.166667	24	0.807918	1.513241
30	1	T2X	Calgary	Midnapore, Sundance	50.9204	-114.0674	43339	15029	76644.0	31.600000	23	0.530700	1.530375
31	1	T1Y	Calgary	Rundle, Whitehorn, Monterey Park	51.0759	-114.0015	57237	19139	70815.5	45.800000	29	0.506665	1.515231

**Cluster 3:** This cluster contained many communities that were closer to Calgary city center on a map. It was predominantly made up of lower to medium income families, low – medium walkability and medium restaurant density. Of all clusters this one stood out as the least desirable to analyze further

[60]:

Cluster Labels	Postal Code	Community	Neighbourhood	Latitude	Longitude	Population (2016)	Total Private Dwellings (2016)	Median Household Income After Taxes	Walk Score	Total Restaurants	Restaurants/1000 residents	Restaurants/1000 residences	
0	2	T2A	Calgary	Penbrooke Meadows, Marlborough	51.04968	-113.96432	59641	22652	62824.5	50.400000	31	0.519777	1.368533
2	2	T2B	Calgary	Forest Lawn, Dover, Erin Woods	51.03180	-113.97860	23418	9953	62116.0	29.500000	27	1.152959	2.712750
3	2	T3B	Calgary	Montgomery, Bowness, Silver Springs, Greenwood	51.08090	-114.16160	37348	15638	61978.0	23.000000	31	0.830031	1.982351
5	2	T3C	Calgary	Rosscarrock, Westgate, Wildwood, Shaganappi, S...	51.03880	-114.09800	23519	12429	68201.0	52.571429	25	1.062970	2.011425
6	2	T2E	Calgary	Bridgeland, Greenview, Zoo, YYC	51.06320	-114.06140	34729	17958	59445.0	53.250000	29	0.835037	1.614879
14	2	T2K	Calgary	Thornccliffe, Tuxedo Park	51.08570	-114.07140	35544	15376	67183.5	40.125000	27	0.759622	1.755983
22	2	T2P	Calgary	City Centre, Calgary Tower	51.04720	-114.08020	13056	8723	53741.0	87.000000	27	2.068015	3.095265
26	2	T2S	Calgary	Elbow Park, Britannia, Parkhill, Mission	51.01710	-114.08120	14637	8485	65239.0	56.875000	25	1.708000	2.946376

**Cluster 4:** This cluster consisted of higher income communities, again spread geographically throughout the city, with low – medium walkability and a full spectrum of restaurant density. This cluster was determined to be worthy of further analysis.



[61]:

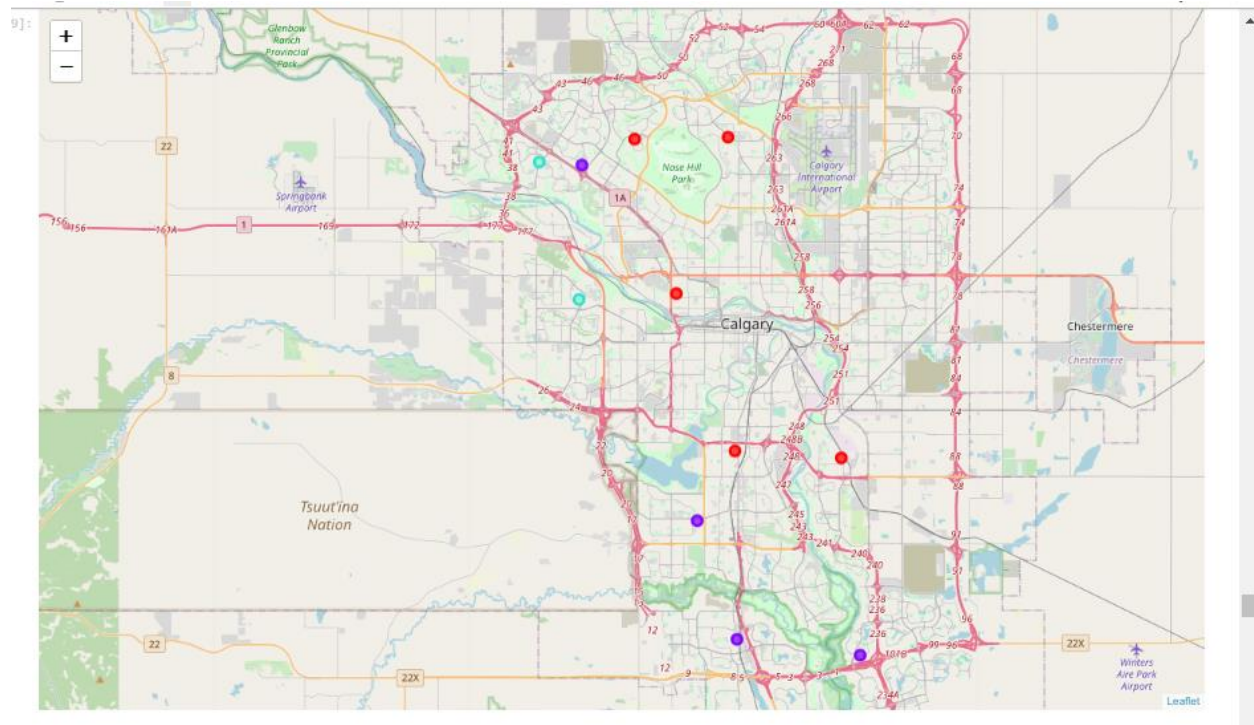
	Cluster Labels	Postal Code	Community	Neighbourhood	Latitude	Longitude	Population (2016)	Total Private Dwellings (2016)	Median Household Income After Taxes	Walk Score	Total Restaurants	Restaurants/1000 residents	Restaurants/1000 residences
1	3	T3A	Calgary	Dalhousie, Edgemont, Hamptons, Hidden Valley	51.12606	-114.143158	53224	19393	90225.0	27.666667	30	0.563655	1.546950
4	3	T2C	Calgary	Lynnwood Ridge, Ogden, Foothills Industrial, G...	50.98780	-114.000100	20406	8332	84791.0	39.500000	26	1.274135	3.120499
9	3	T3G	Calgary	Hawkwood, Arbour Lake, Citadel, Ranchlands, Ro...	51.11470	-114.179600	59025	21018	96005.0	28.000000	28	0.474375	1.332191
15	3	T3K	Calgary	Sandstone, MacEwan Glen, Beddington, Harvest H...	51.12700	-114.078700	80792	27353	89515.0	36.714286	29	0.358946	1.060213
20	3	T2N	Calgary	Kensington, Westmont, Parkdale, University	51.05910	-114.114600	26342	12601	87601.0	36.800000	30	1.138866	2.380763
28	3	T2V	Calgary	Oak Ridge, Haysboro, Kingsland, Kelvin Grove, ...	50.99090	-114.074000	33450	15417	85467.0	46.500000	27	0.807175	1.751313
29	3	T2W	Calgary	Braeside, Cedarbrae, Woodbine	50.96040	-114.100100	42377	17227	94028.0	31.428571	30	0.707931	1.741452
32	3	T2Y	Calgary	Millrise, Somerset, Bridlewood, Evergreen	50.90930	-114.072100	61344	21111	95756.0	27.333333	22	0.358633	1.042111
33	3	T2Z	Calgary	Douglas Glen, McKenzie Lake, Copperfield, East...	50.90230	-113.987300	68438	24669	101028.0	31.400000	25	0.365294	1.013418

## Second Analysis

The postal codes from Cluster 1 and 4 were determined desirable for further analysis. At this point a second clustering analysis was undertaken on a data set containing data only from these clusters, to determine if additional insight could be uncovered and if any additional clusters could be made from the two survivors.

During this analysis, the attempt was made to split into 4 clusters again. However, that proved to be a poor choice, as the 2 postal codes in cluster 1 from the previous analysis just ended up being split into their own single-element clusters. Adjusting the analysis to a 3-cluster analysis yielded the exact same cluster with the West Calgary communities, with Cluster 4 being split into 2 different clusters.

The clusters were again plotted and grouped, with postal codes taken out of each cluster for final analysis. A cluster map of the second clustering analysis is in Figure 7:



**Figure 7: Cluster Map – second cluster analysis**

Again, after the second analysis, the clusters were grouped and themed, with the following identified:

**Cluster 1:** New cluster, made up of communities in Cluster 4 or the original analysis. It consists of higher income communities with low – medium walkability and the full spectrum of restaurant density. Looking further into this cluster, the postal code T2V (high income, moderate walk score, moderate restaurant density) looked best for a final analysis

[70]:

	Cluster Labels	Postal Code	Community	Neighbourhood	Latitude	Longitude	Population (2016)	Total Private Dwellings (2016)	Median Household Income After Taxes	Walk Score	Total Restaurants	Restaurants/1000 residents	Restaurants/1000 residences
2	0	T3A	Calgary	Dalhousie, Edgemont, Hamptons, Hidden Valley	51.12606	-114.143158	53224	19393	90225.0	27.666667	30	0.563655	1.546950
3	0	T2C	Calgary	Lynnwood Ridge, Ogden, Foothills Industrial, G...	50.98780	-114.000100	20406	8332	84791.0	39.500000	26	1.274135	3.120499
5	0	T3K	Calgary	Sandstone, MacEwan Glen, Beddington, Harvest H...	51.12700	-114.078700	80792	27353	89515.0	36.714286	29	0.358946	1.060213
6	0	T2N	Calgary	Kensington, Westmont, Parkdale, University	51.05910	-114.114600	26342	12601	87601.0	36.800000	30	1.138866	2.380763
7	0	T2V	Calgary	Oak Ridge, Haysboro, Kingsland, Kelvin Grove, ...	50.99090	-114.074000	33450	15417	85467.0	46.500000	27	0.807175	1.751313

**Cluster 2:** Another new cluster, made up of communities in cluster 4 of the original analysis. Now contains high income communities, with low walkability and low restaurant density. It was difficult to separate many of the communities in this cluster, so all proceeded to final analysis based on high income and low restaurant density.

[71]:

Cluster Labels	Postal Code	Community	Neighbourhood	Latitude	Longitude	Population (2016)	Total Private Dwellings (2016)	Median Household Income After Taxes	Walk Score	Total Restaurants	Restaurants/1000 residents	Restaurants/1000 residences	
4	1	T3G	Calgary	Hawthwood, Arbour Lake, Citadel, Ranchlands, Ro...	51.1147	-114.1796	59025	21018	96005.0	28.000000	28	0.474375	1.332191
8	1	T2W	Calgary	Braeside, Cedarbrae, Woodbine	50.9604	-114.1001	42377	17227	94028.0	31.428571	30	0.707931	1.741452
9	1	T2Y	Calgary	Millrise, Somerset, Bridlewood, Evergreen	50.9093	-114.0721	61344	21111	95756.0	27.333333	22	0.358633	1.042111
10	1	T2Z	Calgary	Douglas Glen, McKenzie Lake, Copperfield, East...	50.9023	-113.9873	68438	24669	101028.0	31.400000	25	0.365294	1.013418

**Cluster 3:** the exact same cluster as Cluster 1 from the first analysis. The T3H postal code was chosen for further analysis from this cluster due to the slightly better walk score and lower restaurant density

[58]:

Cluster Labels	Postal Code	Community	Neighbourhood	Latitude	Longitude	Population (2016)	Total Private Dwellings (2016)	Median Household Income After Taxes	Walk Score	Total Restaurants	Restaurants/1000 residents	Restaurants/1000 residences	
11	0	T3H	Calgary	Discovery Ridge, Signal Hill, West Springs, Ch...	51.0566	-114.1815	71251	25098	119018.5	27.3	26	0.364907	1.035939
17	0	T3L	Calgary	Tuscany, Scenic Acres	51.1162	-114.2089	30168	10093	110614.0	23.0	18	0.596659	1.783414

## Final Analysis

For the final analysis, data from the most promising postal codes of each cluster were compared. Note that due to the inability to distinguish easily that all of Cluster 2 was included here, leaving a total of 6 postal codes to compare.

A weighted comparison between the 3 measures (income, walkability, and restaurant density) was made to narrow the list down from 6 to 3 postal codes. It was determined that a general scoping tool, due to the inexact nature of the analysis, should give a ranking of multiple choices as opposed to giving a definitive “this is the best” answer.

To begin with, since income, walkability, and restaurant density varied by orders of magnitude, the values of each were normalized to the mean value (mean/standard deviation scaling) using the scikit-learn StandardScaler. Each measure was then given a relative weight, as determined by the analyst-determined importance of each measure. Note that a small sensitivity analysis was carried out on measure weighting, with no change in the overall result. A table follows, showing the weighting of each measure with explanation:

Measure	Weighting	Weighting Sensitivity	Explanation
Income	0.4	0.33 – 0.5	Income given higher weighting, as disposable income is the highest predictor as to whether a person will spend money in a restaurant
Walkability	0.4	0.33 – 0.4	Walkability given high weighting, as with changes in traffic flows people will look to eat (when possible) closer to

			home when going out. Also considered traffic from take out services (Doordash etc)
Restaurant Density	0.2	0.1 – 0.33	Given slightly lower weighting. Areas with high density have been filtered out in previous steps. The assumption is the areas left can absorb another restaurant under the right conditions

**Table 1: Measure Weighing for Final Analysis**

The final analysis was conducted by computing a final score. Each normalized measure was multiplied by its relative weighting to get a weighted score. The weighted score was then totaled as Income + Walk Score – Restaurant Density (the – in restaurant density was to account for the fact that lower density is better). The results are presented as follows:

[87]:	Postal Code	Neighbourhood	Income - Weighted	Walk Score - Weighted	Restaurants/1000 residents - Weighted	Total Score	Latitude	Longitude	Rank
0	T3H	Discovery Ridge, Signal Hill, West Springs, Ch...	0.798361	-0.279459	-0.164943	0.683844	51.0566	-114.1815	1
1	T2Z	Douglas Glen, McKenzie Lake, Copperfield, East...	0.096639	-0.035346	-0.164512	0.225805	50.9023	-113.9873	2
2	T2V	Oak Ridge, Haysboro, Kingsland, Kelvin Grove, ...	-0.510321	0.863706	0.327472	0.025914	50.9909	-114.0740	3

## **Conclusions & Further Analysis**

This analysis determined that the three most promising postal codes, as weighted by income, walkability, and restaurant density, to open a new restaurant in are:

- T3H: Discovery Ridge, Signal Hill, West Springs
- T2Z: Douglas Glen, McKenzie Lake, Copperfield
- T2V: Oak Ridge, Haysboro, Kingsland

It should be noted that this is considered only a start point to scope out a general area. Based on the availability of data, and an understanding of industrial activity in each area, a further analysis can be undertaken to narrow down each postal code to the community level.

Further analysis can be performed on this problem/issue. However, based on my research on the availability of data, it may require acquiring non-open-source data sets, likely from for-profit services. With that sort of data, further analysis could be undertaken to improve this study in areas such as:

- Neighborhood level analysis: a more granular analysis of the current topic
  - o Includes being able to look at things like walk scores and income on a neighborhood, rather than multi-neighborhood/postal code levels

- More up to date data: government data typically lags by quite a bit (see population and income data reported from the 2016 census, and it being the best source available in 2021)
- Demographic data per neighborhood, so a more detailed analysis of restaurant type can be conducted based on neighborhood demographic, as opposed to restaurants currently in place
-