



# A New Restaurant in Calgary, AB

Prepared for: Applied Data Science - Final Project

Prepared by: Manny Gill

---

---

# 1. INTRODUCTION

## 1.1 Background

Opening a new restaurant can be a large undertaking and seem daunting at times. One of the most important factors for a successful business is Location. It is often argued that location is the most important factor in opening a new restaurant. Using a data we can recommend a potential location in Calgary, Alberta, Canada. Calgary sits in the sunny eastern foothills of Canada's Rocky Mountains, where the Bow and Elbow rivers meet. It is the major urban centre for the entire southern half of the province of Alberta, and is surrounded by an area of profound beauty with an unspoiled, resource-rich natural environment. It is no surprise to Calgarians that their home has been ranked the 5th most liveable city in the world by The Economist Intelligence Unit from 2009-2017. In 2018, Calgary rose to become the 4th most liveable city in the world. Population density is low, quality of life is high and housing abundant. Find a new location for a restaurant can define its continued success.

## 1.2 Business Problem

In this report we will find communities in Calgary and determine which community would be an optimal location for a new restaurant. Information that would help us in knowing the current venues in the communities and related populations. Factors that could help determine a location include; population density, income, venues in the area.

What community in Calgary should a new restaurant be opened?

## 1.3 Interest

In the this report we aim help new restauranteurs to navigate the communities in Calgary, Alberta, Canada to find the best location to open a new restaurant. The information can help communities attract new restauranteurs to their communities. Others that man benefit are current restaurant owners, politicians and investors.

---

## 2. DATA

### 2.1 Sources

We will be using data from a variety of sources to get the information required;

- Wikipedia( [https://en.wikipedia.org/wiki/List\\_of\\_neighbourhoods\\_in\\_Calgary](https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Calgary)) : Using the data we will be able to get the neighbourhoods, population and population density.
- Foursquare API: Using the information from foursquare we will get the venues, restaurants for the neighbourhoods.
- Income Data: For all intensive purposes we will use the data from Great News (<https://great-news.ca/demographics/>) and assume the information is representative of Venue Count in Calgary.

### 2.2 Cleaning

Data will need to be scraped from the sources mentions in section 2.1 and combined. Missing data or incomplete data was present for some of the income figures related to neighbourhoods. We will make a few assumptions;

- Only neighbourhoods with assigned data for median household will be processed.
- Neighbourhoods missing data will be dropped.
- Income values from given year will be used to characterize neighbourhood spending.

Examples of Data;

Name <sup>[9]</sup>	Quadrant	Sector <sup>[10]</sup>	Ward <sup>[11]</sup>	Type <sup>[10]</sup>	2012 Population Rank	Population (2012) <sup>[9]</sup>	Population (2011) <sup>[9]</sup>	% change	Dwellings (2012) <sup>[9]</sup>	Area (km <sup>2</sup> ) <sup>[10]</sup>	Population density
Abbeyle	NE/SE	Northeast	10	Residential	82	5,917	5,700	3.8	2,023	1.7	3,480.6
Acadia	SE	South	9	Residential	27	10,705	10,615	0.8	5,053	3.9	2,744.9
Albert Park/Radisson Heights	SE	East	10	Residential	75	6,234	6,217	0.3	2,709	2.5	2,493.6
Altadore	SW	Centre	11	Residential	39	9,116	8,907	2.3	4,486	2.9	3,143.4
Alyth/Bonnybrook	SE	Centre	9	Industrial	208	16	17	-5.9	14	3.8	4.2
Applewood Park	SE/NE	East	10	Residential	69	6,498	6,404	1.5	2,215	1.6	4,061.3

Fig 2.2.1 : Neighbourhood Data

Community	Newsletter Name	Median Household Income	Median Age	Population 2014	Dwellings 2014	City Quadrant	Median Home Sale Price
Abbeyle	-	\$55,345	34	6,071	2,031	SE	\$305,000
Acadia	Acadia	\$46,089	42	10,969	5,067	SE	\$447,000
Albert Park / Radisson Heights	-	\$38,019	37	6,529	2,936	SE	\$349,900
Altadore	The Source	\$53,786	37	9,518	4,537	SW	\$925,000
Applewood Park	-	\$65,724	33	6,864	2,228	SE	\$380,000

Fig 2.2.2 : Neighbourhood Income data

### 3. METHODOLOGY

#### 3.1 Exploratory Data Analysis

One of the first steps we will undertake is to explore the data. In order to become more familiar with the data we will be perform some correlation studies, mapping the data, and viewing some scatter plots. Once taking the data from the various sources I was able to compile the information into a single data-frame. Given the venue count is a target value as we looking to increase that by adding a restaurant to a given community. We were interested to know the correlation of other variables such as Venue Count and population density as they relate to venue count. Also it is important to note the relationship between Population Density and Median household income.

	Community	Community Latitude	Community Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
count	3598	3598.000000	3598.000000	3598	3598.000000	3598.000000	3598
unique	174	NaN	NaN	1640	NaN	NaN	251
top	CLIFF BUNGALOW	NaN	NaN	Tim Hortons	NaN	NaN	Coffee Shop
freq	100	NaN	NaN	91	NaN	NaN	260
mean	NaN	51.035618	-114.080074	NaN	51.035788	-114.079739	NaN
std	NaN	0.061713	0.057455	NaN	0.061438	0.057038	NaN
min	NaN	50.857617	-114.260486	NaN	50.858113	-114.267232	NaN
25%	NaN	51.005040	-114.106353	NaN	51.005123	-114.107874	NaN
50%	NaN	51.037271	-114.078200	NaN	51.038763	-114.074253	NaN
75%	NaN	51.062712	-114.057466	NaN	51.066255	-114.061070	NaN
max	NaN	51.175690	-113.924528	NaN	51.184048	-113.915710	NaN

Fig 3.1.1: Venue description

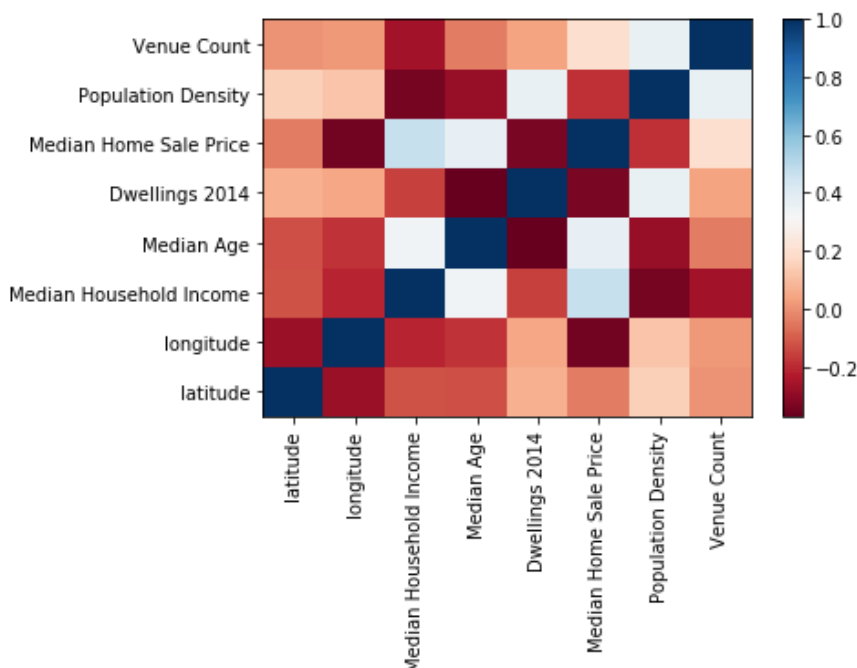


Fig 3.1.2 : Correlation Study, Calgary, Foursquare data.

We can see the variables venue count, population density and median household income are related. See Values below;

The Pearson Correlation Coefficient is 0.3688238884606182 with a P-value of  $P = 5.507276499715131e-07$  For Population Density vs. Venue Count

The Pearson Correlation Coefficient is -0.26250677026798164 with a P-value of  $P = 0.00046645289590050543$  For Median Household Income vs. Venue Count

The Pearson Correlation Coefficient is -0.34131725946078717 with a P-value of  $P = 4.046037043666337e-06$  For Population Density vs. Median Household Income

Based on the correlation and P-value we will proceed with using these factors to determine a community. Additionally we would like to get a sense of where the population density and median household incomes are relative to each other. Fig 3.1.1 : Correlation Study, Calgary, Foursquare data.

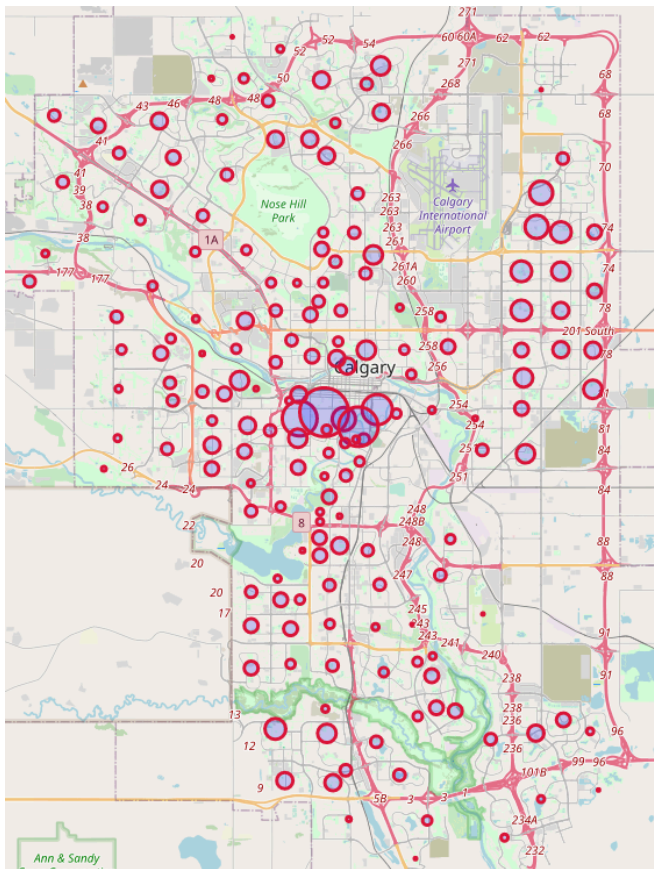


Fig 3.1.3 :Population Density bubble map

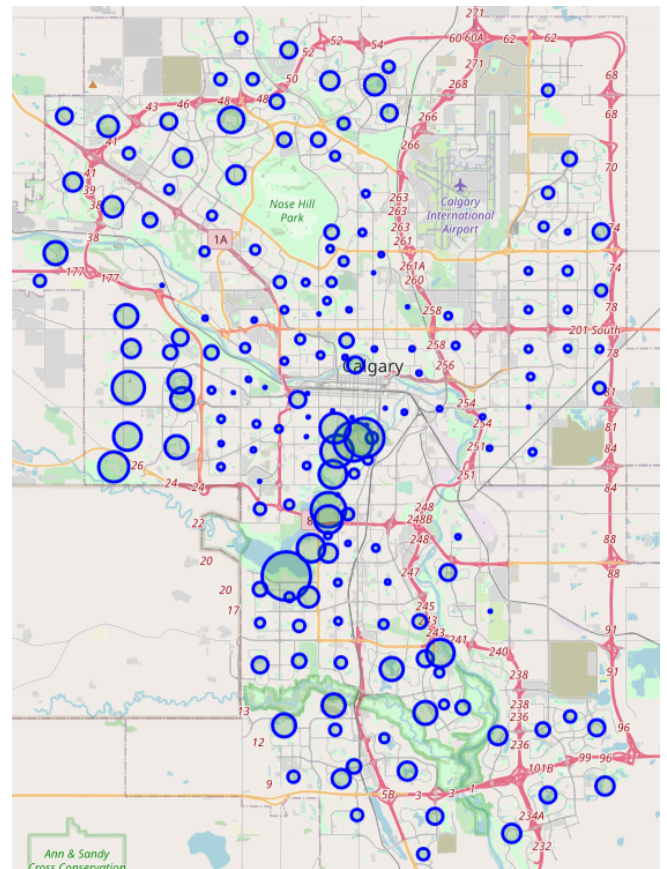


Fig 3.1.4 :Median Household Income bubble map



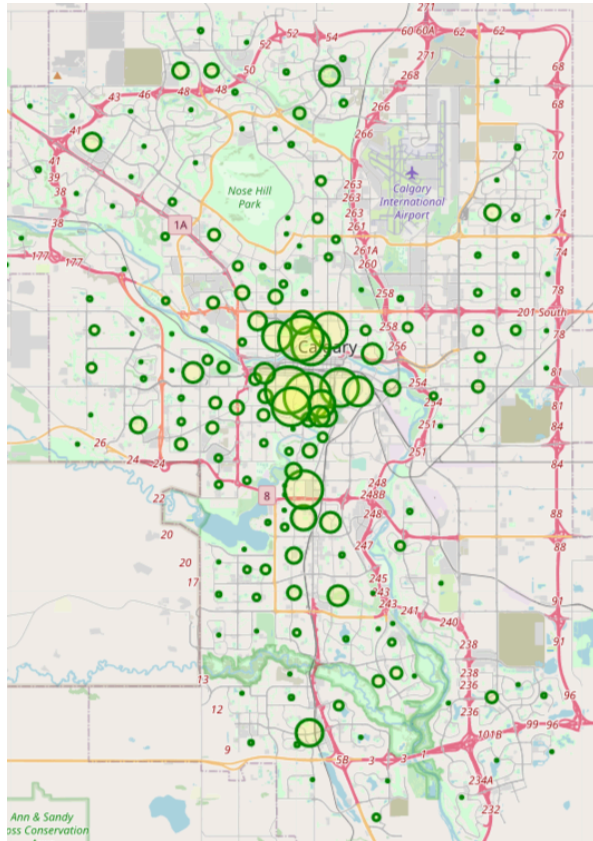
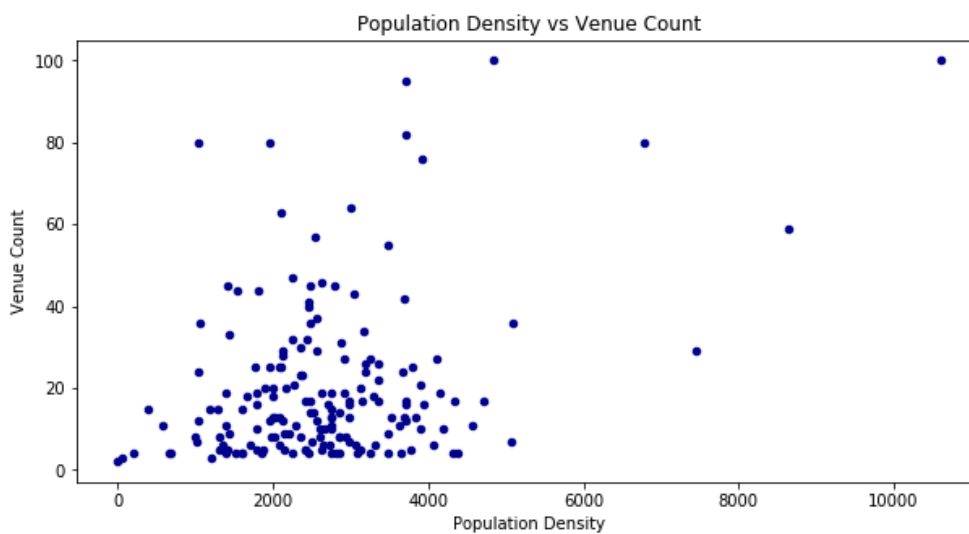


Fig 3.1.5 :Venue Count bubble map

### 3.2 Relationship between Population Density vs Venue Count

We can infer this would be a given relationship for established neighbourhoods as it indicates the venue could will go up as the population density increases. Plotting the data on a scatter plot we can see this relationship. However given communities with shopping centres (Malls) we can see a large jump in a venue count.

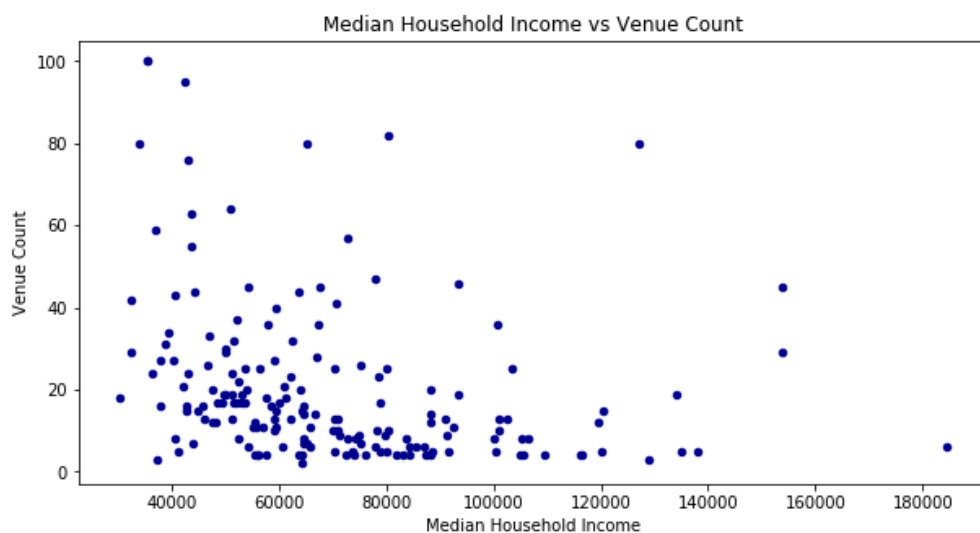


3.2.1 Fig: Population Density Vs Venue Count

---

### 3.3 Relationship between Median Household Income vs Venue Count

We are looking to see the relationship between Median household income and venue count. Using the idea the higher the Median household income the more spending power the household has. See the scatter plot below we can infer that as median household income increases the venue count goes down



3.3.1 Fig: Median Household Income Vs Venue Count

---

### 3.4 Using DBSCAN Clustering

We were able to classify the communities with the given median household income, population density, and venue count into clusters. See the breakdown of the clusters below. We used machine learning to

	Median Household Income	Population Density	Venue Count
Clus_Db			
-1	71371.239437	2871.028169	30.690141
0	65774.925926	2569.425926	11.814815
1	69144.000000	2465.000000	43.000000
2	70481.000000	3806.666667	14.000000
3	39497.000000	3098.333333	31.000000
4	102903.666667	2673.000000	10.000000
5	76377.333333	3630.333333	6.333333
6	68637.500000	2041.000000	26.500000
7	117599.000000	1465.000000	4.666667
8	47176.666667	3956.333333	17.500000
9	53281.500000	2408.500000	34.750000
10	55145.500000	2722.750000	5.500000
11	100590.500000	2012.500000	10.000000
12	73366.000000	840.500000	6.500000
13	43851.500000	1236.500000	15.000000
14	89504.500000	3622.000000	12.500000
15	90844.000000	1899.500000	21.000000
16	64582.250000	655.500000	9.000000
17	105663.000000	1411.500000	6.000000

3.4.1 Fig: DBSCAN median cluster values

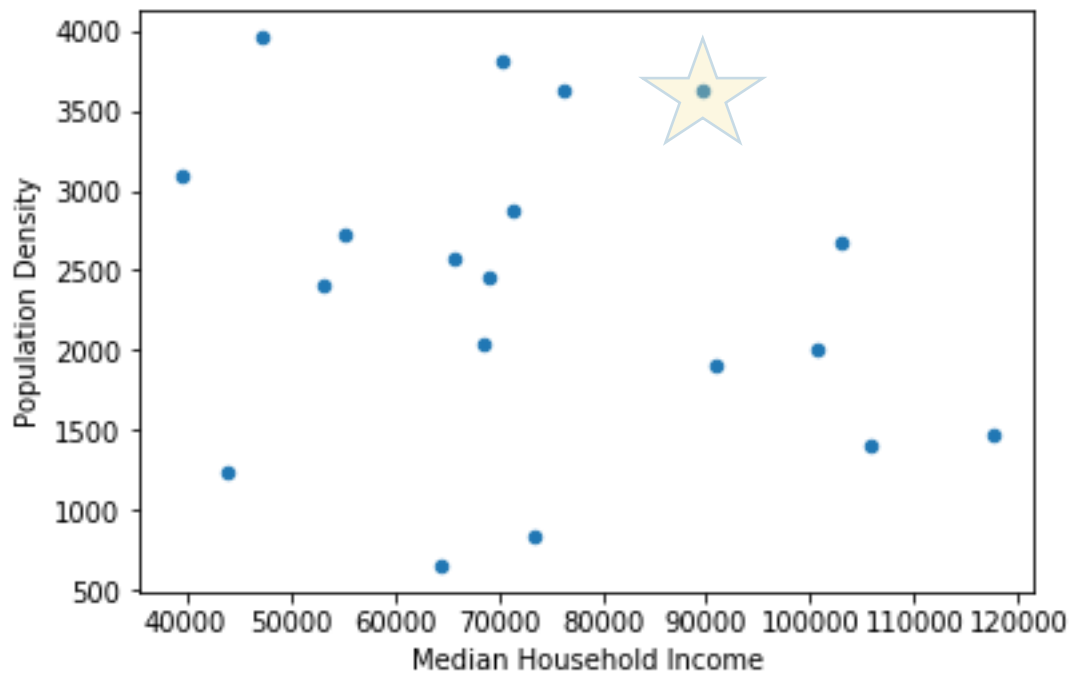


## 4. RESULTS

Given the variety of venues by getting an average venue count we can get an idea of neighbourhoods. Venues with average venue count are a good balance between optimal competition as well as gaining traffic to the restaurant. We can see that some neighbourhoods have very high venue counts, however reading into the results we see these neighbourhoods have Malls. A restaurant location inside a Mall is not part of this new restaurant location selection

	latitude	longitude	Median Household Income	Median Age	Dwellings 2014	Median Home Sale Price	Population Density	Venue Count	Clus_Db	Clus_km
count	174.000000	174.000000	174.000000	174.000000	174.000000	1.740000e+02	174.000000	174.000000	174.000000	174.000000
mean	51.034690	-114.081681	69529.178161	38.896552	2585.954023	5.976602e+05	2669.850575	20.678161	2.057471	2.448276
std	0.074565	0.075531	26484.943723	5.823084	1853.100881	2.283349e+05	1340.214248	19.908167	4.890374	1.530111
min	50.857617	-114.260486	30411.000000	29.000000	110.000000	3.050000e+05	0.000000	2.000000	-1.000000	0.000000
25%	50.985098	-114.129630	51258.250000	34.250000	1153.750000	4.402500e+05	1956.750000	7.000000	-1.000000	1.000000
50%	51.041900	-114.085021	64164.000000	38.000000	2402.000000	5.480000e+05	2544.500000	14.000000	0.000000	3.000000
75%	51.084942	-114.036100	81422.500000	42.000000	3516.500000	6.876250e+05	3238.250000	25.000000	2.750000	4.000000
max	51.175690	-113.924528	184383.000000	58.000000	15243.000000	1.701000e+06	10600.000000	100.000000	17.000000	4.000000

4.1 Fig: Description of results from clustering



4.2 Fig: DBSCAN  
median value  
plotting

---

By selecting a cluster that is above the average Density and the largest Median Household Income we can see the Cluster #14.

- Panorama Hills: Population Density: 3,531, Median Household Income: \$90,851, Venue Count: 13

## 4. DISCUSSION

Once a new restaurant would be added to the neighbourhood we can see that venue count would be in the 50% percentage value of venue count per neighbourhood.

Looking into the neighbourhood further we see the community is a diverse neighbourhood with 48% Immigrant population. Additionally from an article from the time of the data we can see it is also pegged as the largest community in Calgary.

### **Calgary's largest communities by population:**

- **Panorama Hills — 25,993**
- **Beltline — 21,939**
- **Evergreen — 21,700**
- **Tuscany — 19,737**
- **Taradale — 19,223**

Article: Calgary Herald,

<https://calgaryherald.com/news/local-news/is-calgary-still-growing-city-releases-new-census-figures/>

---

---

## 5. CONCLUSION & FUTURE DIRECTION

All in all, the community we selected was Panorama Hills: with a Population Density: 3,531, Median Household Income: \$90,851, Venue Count: 13. We came to this conclusion by comparing population density, median household income, four square data. By using classification systems we were able to come up with certain classes of neighbourhoods that would be ideal for a new restraint location. Although we concluded with one specifically, we would need to dive deeper into other metrics to come to answer the much boarder question.

We were able to establish the bases for larger study in opening a new restaurant. Additional factors can be taken into account into establishing a successful restaurant. Factors can include the leasing costs, restaurant types, demographics, established cultural district and so on. However as said in the industry Location, Location, Location.

---