Where to Open a Coffee Shop in Calgary, Alberta, Canada?

# Steve Ennis

# May 21, 2020

## **Introduction**

Captain Joe Columbian is a lifelong coffee aficionado. He fervently believes he can source, roast, and serve the best coffee in Calgary. He even has a name for the first coffee shop that will start his coffee revolution that will eventually lead to world domination, ‘Cappa Joes’. Captain Joe now needs to find the perfect spot to begin his coffee empire.

## **Cappa Joes Business Problem**

Through Captain Joe’s connections in the Central and South American coffee growing regions, he able to find the rarest and most complex tasting coffee beans in the world. The problem is that these beans are expensive, and he will have to place Cappa Joes in an area where the clientele are more likely to appreciate gourmet coffee and more importantly pay gourmet prices.

## **Cappa Joes Target Clientele**

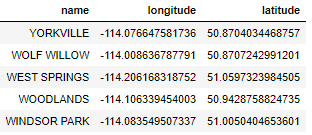
Joe has found 33 amazing coffee statistics [online](https://disturbmenot.co/coffee-statistics/). One statistic that will drive his business model is that ‘48% of millennials consume gourmet coffee’(Source [Reuters](https://www.reuters.com/article/us-coffee-conference-survey/americans-are-drinking-a-daily-cup-of-coffee-at-the-highest-level-in-six-years-survey-idUSKCN1GT0KU), [CreditDonkey](https://www.creditdonkey.com/coffee-drinking-statistics.html) ). Another statistic for coffeeshop consideration is that ‘79% of Americans make coffee at home’(Source [Reuters](https://www.reuters.com/article/us-coffee-conference-survey/americans-are-drinking-a-daily-cup-of-coffee-at-the-highest-level-in-six-years-survey-idUSKCN1GT0KU)). Cappa Joes needs to target millennials where they live and develop a thriving take out business.

One final selection criterion to keep in mind is market saturation. The ideal number of coffeeshops per person is unknown but for this exercise we will look to the list top 10 most coffee shops of per capita in the United States (Source [MatadorNetwork](https://matadornetwork.com/read/us-cities-coffee-shops-per-capita/)). The number 10 city is Ann Arbor, Michigan with 1 coffee related business per 2825 people. For this study, the threshold will be set at 1 coffeeshop per 3000 people.

## **The Data Sources**

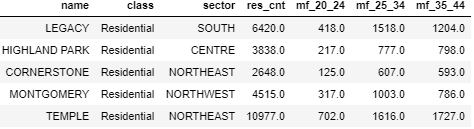
[Open Calgary](https://data.calgary.ca/) is an excellent source for open data about the city of Calgary. Location data for the various communities or neighborhoods within Calgary can be found [here](https://data.calgary.ca/Base-Maps/Wards-and-Communities-Data-Lens/9p3u-cwgk). This source will provide the latitude and longitude data for each neighborhood. Figure 1 is an example of the data obtained from the above-mentioned source.

**Figure 1. Location Data**



The next data source can also be found on the Open Calgary data portal. Demographic data can be found [here](https://data.calgary.ca/Base-Maps/Wards-and-Communities-Data-Lens/9p3u-cwgk). The Census by Community for 2019 information will be used to extract the population by neighborhood. Figure 2 is a sample of information taken from the census data.

**Figure 2. Demographic Data**



Wikipedia defines [millennials](https://en.wikipedia.org/wiki/Millennials) as being born in the years 1980-1996. This put the age bracket of interest at 24-40 years old. This data is also found in the census information mentioned above.

The final data source to be used is Foursquare location data. This source will provide an understanding of the neighborhoods Cappa Joes is considering opening a coffeeshop. From Foursquare data stats Cappa Joe will be able to understand the amenities in the prospective neighborhoods. Who is the competition? Are there complementary businesses that bring foot traffic such as bookstores, parks, and clothing stores? Figure 3 is an example of data obtained from Foursquare.

**Figure 3. Foursquare API data**

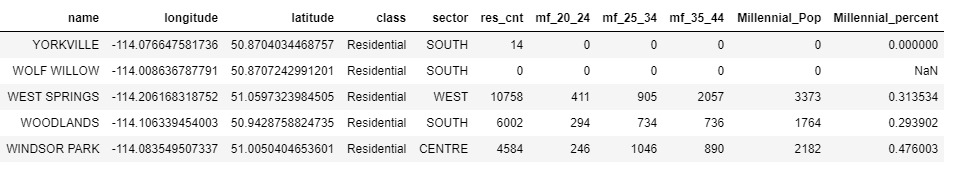


## **Data Acquisition and Wrangling**

When scraping the data from both City of Calgary data sources, the Socrata library from sodapy was used. The location data was limited to neighborhood name and the latitude and longitude.

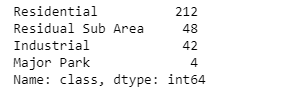
The demographic data was a much larger set which consisted of 142 columns. After examining the full data set it was determined that for this project to limit the data to neighborhood, neighborhood class, city sector and population data. The population data was further limited to the total resident count and the age brackets which millennial’s fall (mf\_20\_4, mf\_25\_34, and mf\_35\_44). It was then decided to the add the brackets together to produce a column for the millennial population. An analysis of the data type from the demographic data revealed that the population columns were in ‘object’ datatypes. The columns were promptly converted to integer so that the column could be completed. A column for the percent of Millennials in a neighborhood was also calculated and the demo graphic data was merged with the location data. Figure 4 show the result of this data preparation.

**Figure 4. Merged Demographic and Location data**

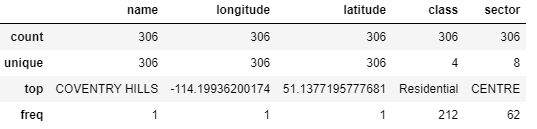


**Statistics of the merged dataset were examined, and it was noted that 42 neighborhoods had classes listed as Industrial. Figure 5 shows the count of the different Neighborhood classes contained within the data set. And analysis of the Object datatypes in the dataset and the results can be found in Figure 6. Figure 7 shows the stats of the numeric datatypes found within the dataset.**

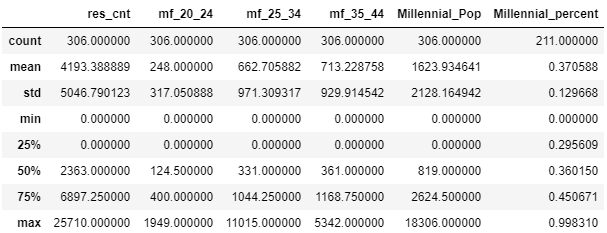
**Figure 5. Neighborhood Class count**



**Figure 6. Statistics of Object datatypes**

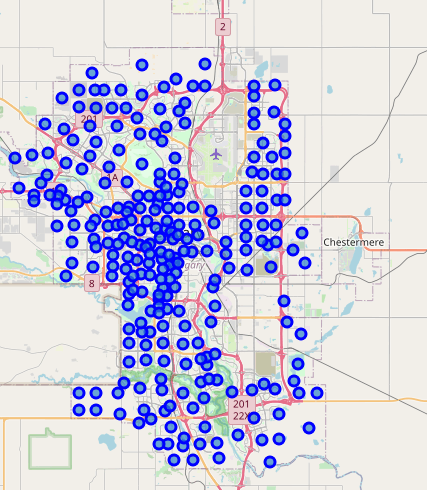


**Figure 7. Statistics of Object datatypes**



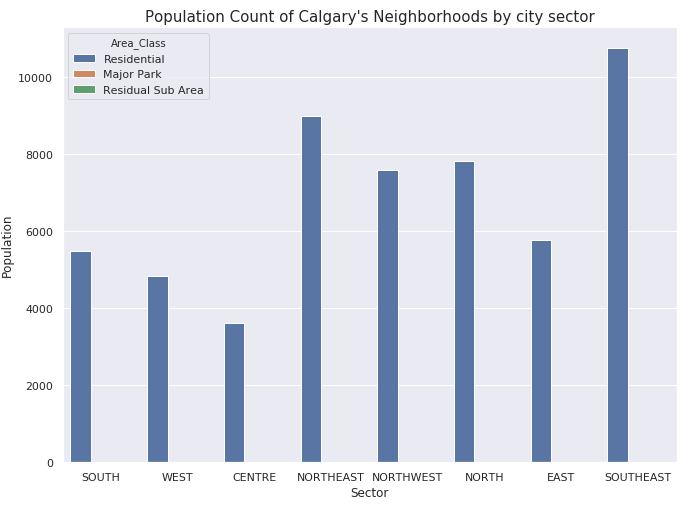
**These industrial classed areas were filtered out and the results were plotted in a folium map (see Map 1)**

**Map 1. Map of Non-Industrial areas of Calgary**



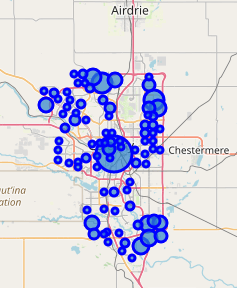
**The population data was then plotted in a bar graph by city sector with bars coloured by the remaining area classes. This analysis revealed that the Residential class contained all the population. Figure 8 highlights this. At this point it was decided to filter the data to Area Class equal to Residential and to areas with a millennial population greater than or equal to 2500 people.**

**Figure 8. Population Count of Calgary Neighborhoods by City Sector and Class**



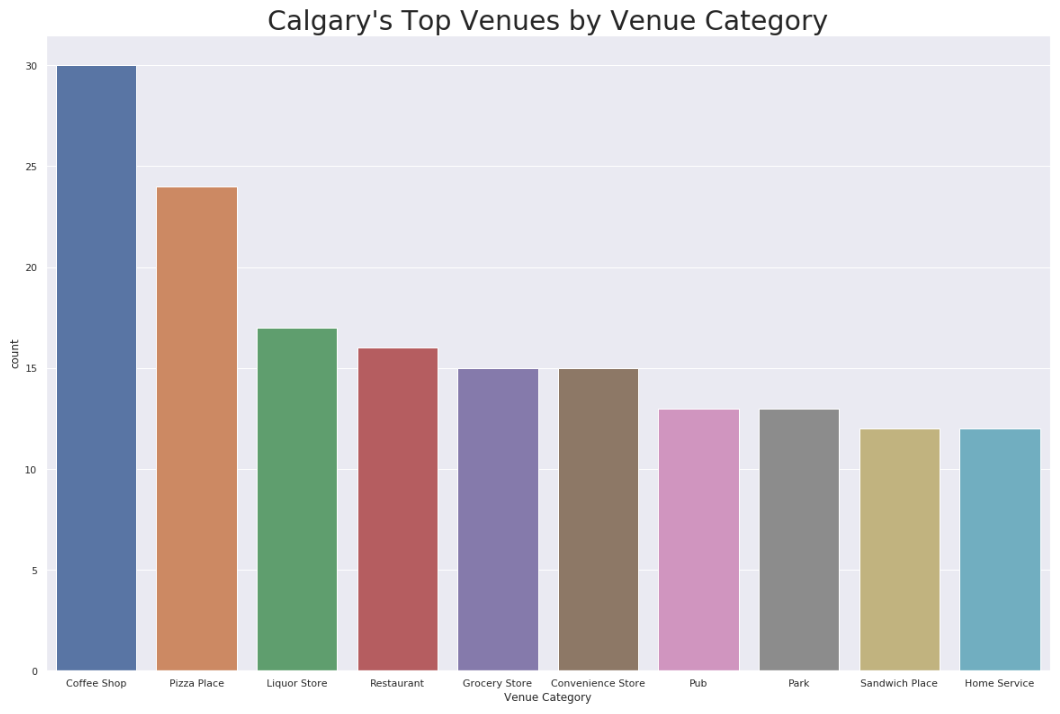
**The remaining dataset of 88 communities was plotted on Map 2 with the circle marker set to reflect size of the millennial population.**

**Map 2. Map of Residential areas of Calgary sized by Millennial Population**



**The location data was then used to extract venue data from the Foursquare API. A count of the Venue Category was done, and it revealed that Calgarians love their coffee as coffee shop was the number 1 venue in the Foursquare dataset.**

**Figure 9. Count of Calgary’s Top 10 Venues by Venue Category**



## **Data Analysis and Methodology**

At this point a decision was made to limit the data to Neighborhoods with coffee shops. Cappa Joes is a start up with a small budget and it was decided that coffee giants like Starbucks, Second Cup and Tim Horton’s have researched the ideal locations regarding foot and vehicle traffic. Cappa Joes decided focus on these areas in hopes of luring coffee aficionados from the big coffee chains with its higher quality product. Cappa Joes needs the visibility of being near high coffee traffic areas.

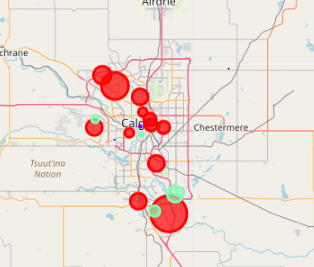
The venue data was filtered to Coffee Shops and Cafes. Café’s were replaced with coffee shop in the venue data. The data was grouped, and hot encoding was done to get a listing of the neighborhoods with the number of coffee shops (see Figure 10).

**Figure 10. Count of Coffee Shops in Neighborhoods with Millennial Population over 2500**



The above-mentioned data was then merged with the demographic data wrangled earlier and the resulting data set was iterated several times though a K Means clustering algorithm. It was determined that the optimum clustering number was 3. The results were then represented on Map 3. The circle markers were sized by the Population per existing Coffee Shop which represents areas where the population may be able to support more coffee shop saturation.

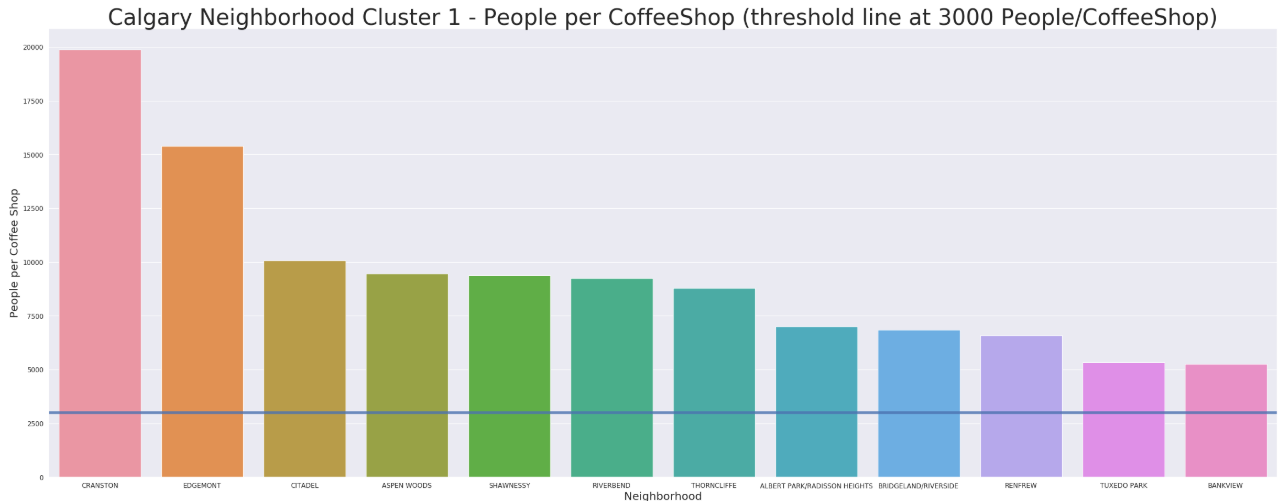
**Map 3. Map of Residential Clusters sized by Population per Coffee Shop**



## **Cluster 1 Analysis**

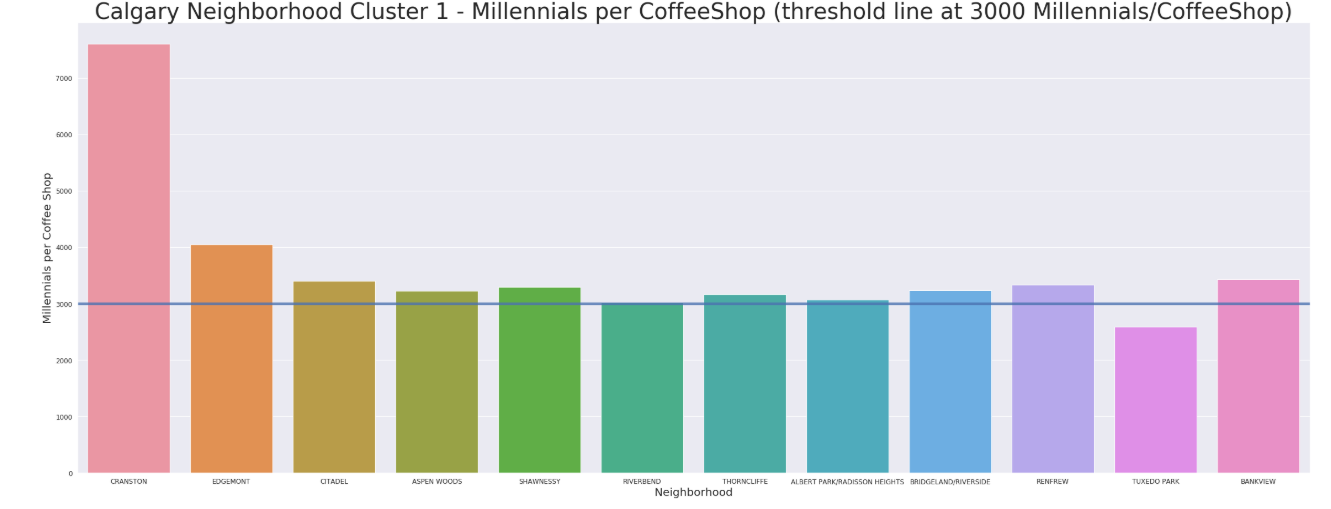
Cluster 1 provides the greatest opportunity for opening a coffee shop. The threshold of 3000 persons per coffee shop that was mentioned at the beginning of the report was represented by a line on a bar graph at 3000. As can be seen in Figure 11, 12 Neighborhoods in Cluster 1 pass this threshold.

**Figure 11. Cluster 1 – People Per Coffee Shop**



Deeper analysis of the millennial population per coffee shop in Cluster 1 is also encouraging. 8 of the 12 Neighborhoods in Cluster 1 are over the 3000 millennials per coffee shop threshold (see Figure 12).

**Figure 12. Cluster 1 – Millennials Per Coffee Shop**

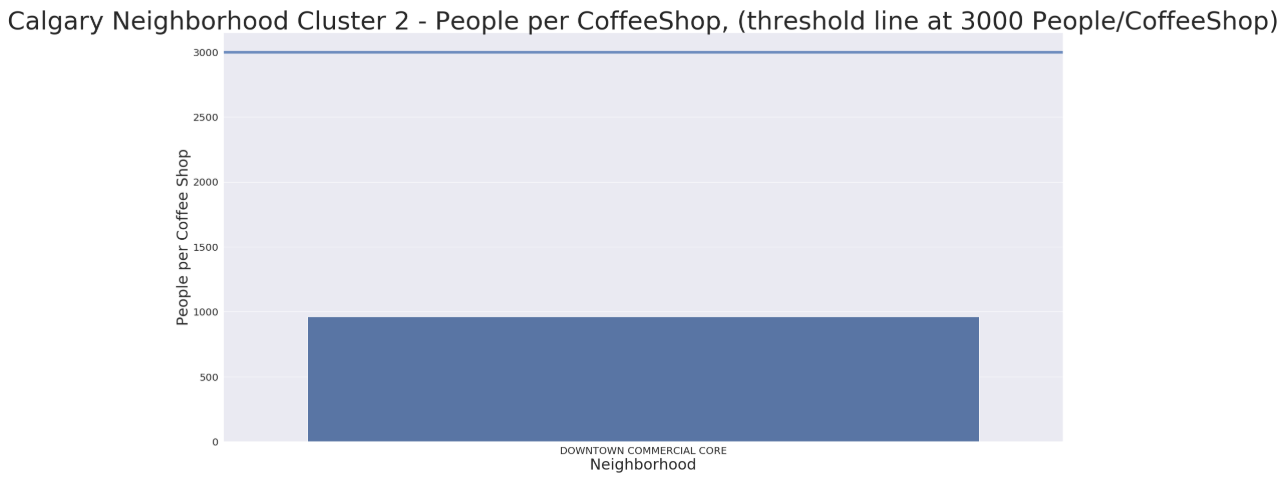


In summary, Cluster 1 contains several Neighborhoods that may provide ideal locations to open a high-end coffee shop.

## **Cluster 2 Analysis**

At first glance, Cluster 2 looks to be over saturated in coffee shops. There are 9 coffee shops in the downtown core. This means there are 964 residents (see Figure 13) and 619 millennials per coffee shop. This is well below the 3000 per coffee shop threshold. More analysis of needs to be done on this area because of the daily influx of workers in the downtown core. For this report the worker population has not been considered and the area may indeed be able to support additional coffee shops. The high cost of leasing or renting space for a coffee shop in this area must also be taken into consideration. Much more research must be done before opening a coffee shop in this area.

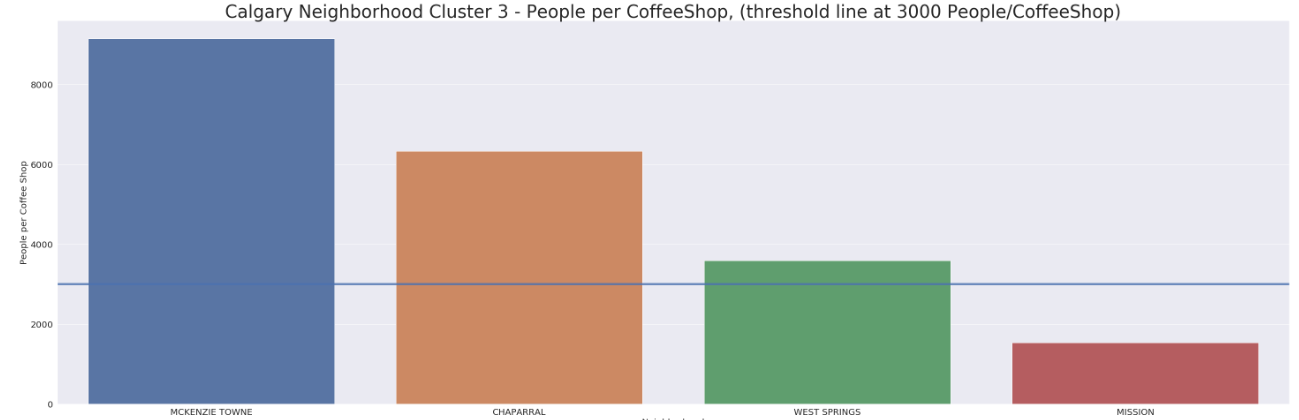
**Figure 13. Cluster 2 – People Per Coffee Shop**



## **Cluster 3 Analysis**

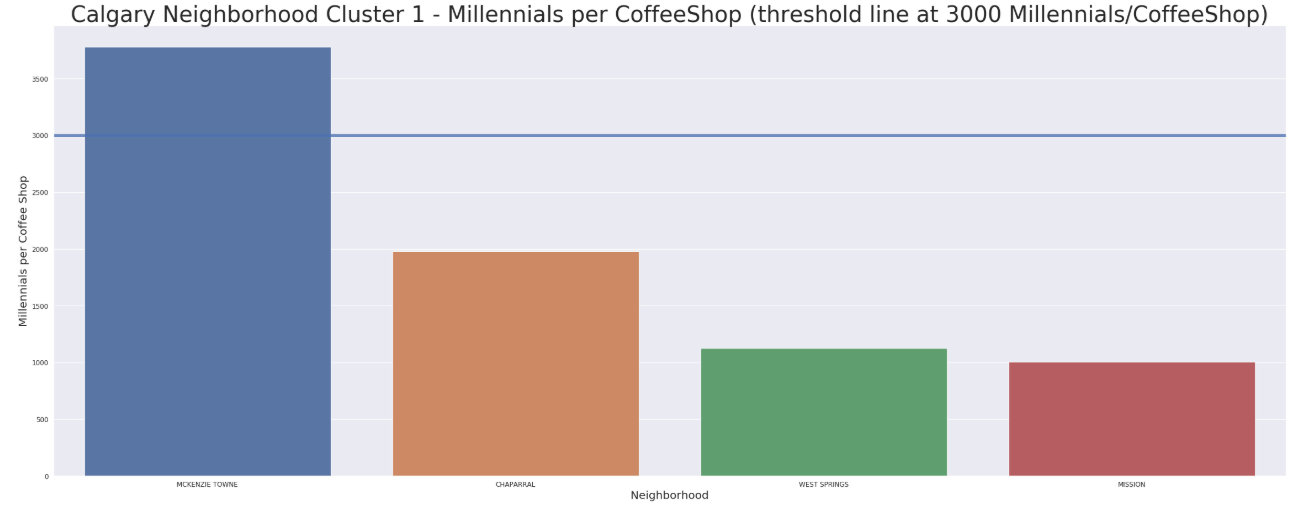
Cluster 3 has 3 neighborhoods that pass the 3000 people per coffeeshop threshold (Figure 14).

**Figure 14. Cluster 3 – People Per Coffee Shop**



McKenzie Town is the only neighborhood in Cluster 3 that passes the 3000 millennials per coffee shop threshold (Figure 15). McKenzie Town may be the best option in this Cluster of neighborhoods to start a coffee shop.

**Figure 15. Cluster 3 – Millennials Per Coffee Shop**



## **Results and recommendations**

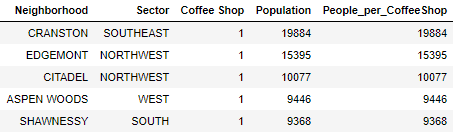
The data was then ranked in descending order by millennials per coffee shop. A list of the Top 5 locations with most millennials per coffeeshop location was put into a summary table (Figure 16). This table contains the neighborhoods where Cappa Joes should consider opening a coffee shop to capture the millennial population and their gourmet coffee tastes.

**Figure 16. Top 5 Millennial Coffee Shop locations**



To give a business its best chance of succeeding, it is also wise to look at the coffee shop density regarding the general population. Figure 17 summarizes this list of top coffee shop prospective locations and fortunately it is the same as the millennial list. Further analysis of the population demographics should be conducted to understand whether a younger age cohort will be moving into the desired age brackets. These 5 recommended neighborhoods are the best candidates for opening Cappa Joes first coffee shop.

**Figure 17. Top 5 Millennial Coffee Shop locations**



## **Conclusion**

The neighborhoods recommended in this report are a first step. On population data alone the 5 neighborhoods appear to be able to support more coffee shop density. Further analysis of the traffic patterns, proximity to other businesses and demographic data must also be completed to get a more complete picture of the viability of a prospective location.