# Analyzing the Neighborhood Coffee Shops in Calgary

## Introduction

Calgary City is situated in the western province of Canada in Alberta. According to the 2019 statistics the city had a population of 1.286 million, making it Alberta’s most populated city (1). Although the winters in Calgary city can get a bit cold at times, there are plenty of perks about living in Calgary city. According to Economist’s Global Livability Ranking, Calgary city has been named one of the top five most livable cities for the past five years. The ranking assesses livability based on stability, healthcare, culture, environment, education, and infrastructure (2).

Asides from the livability ranking, according to the 2018 census for the metropolitan area, Calgary has the highest concentration of head offices in Canada and is home to 115 or approximately 1 in 7 of Canada’s 800 largest corporate headquarters (Ly, 2018). With the numerous head offices and corporate offices situated in Calgary, the number of coffee shops located in Calgary is also abundant. This report will focus on analyzing the coffee venues in Calgary city to offer an assessment of how the location and surrounding venues might affect the overall performance of the coffee shops. By exploring the neighborhood venues of the coffee shops, this report will provide insight to business owners on potentially where to open a new coffee shop in Calgary.

## Data Sources and Preparation

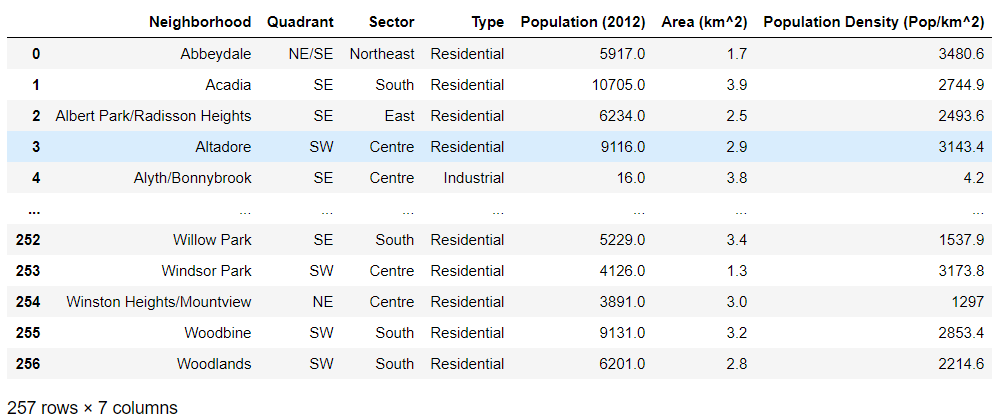
Based on the problem definition, the different factors to be considered in the analysis include:

* Calgary neighborhood and demographic information
* Number of different coffee shop venues in Calgary neighborhoods
* The types of venues surrounding the coffee shops
* Coffee shop ratings, likes, photos, tips, and price

The data sources and preparation methods are listed below:

#### Data Sources

* Calgary Neighbourhood and Census Dataset: <https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Calgary#cite_note-opendatacommunities-11>



* Calgary Neighborhood Geojson File: <https://raw.githubusercontent.com/blackmad/neighborhoods/master/calgary.geojson>

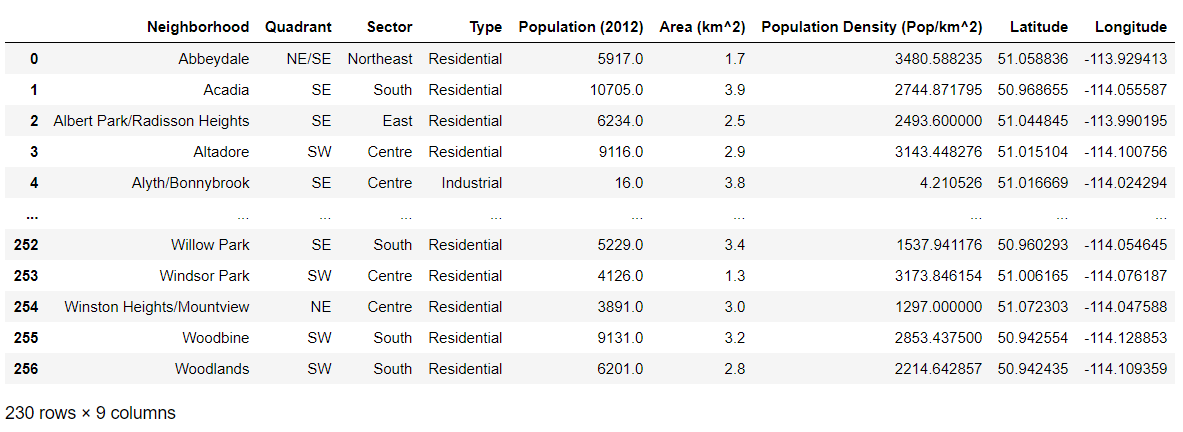
#### Data Preparation:

* *Geopy*—is used to scrape the Calgary Neighbourhood and Census Dataset to generate latitude and longitude information for further analysis
* *Foursquare API*—is used to explore the Calgary neighborhoods to generate the list of venues
* *Foursquare API*—is also used to obtain detailed information for each venue including ratings, likes, photos, tips, and price

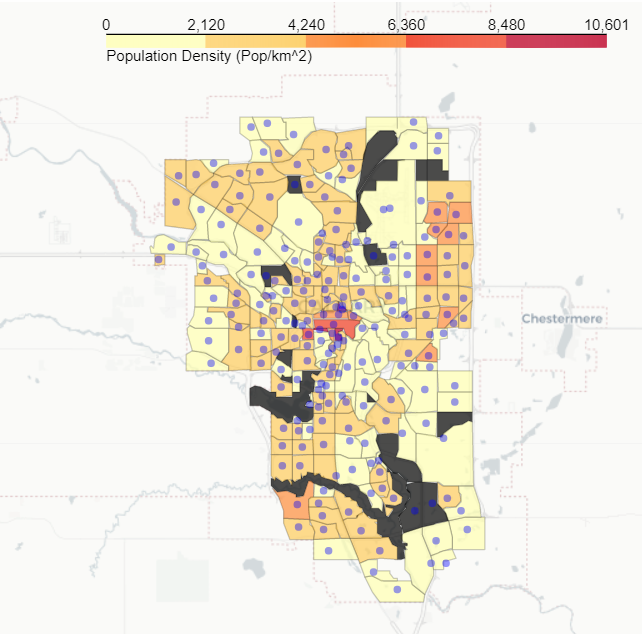
## Methodology

This report will focus on analyzing the coffee venues in Calgary city to offer an assessment of how the location and surrounding venues might affect the overall performance of the coffee shops. By exploring the neighborhood venues of the coffee shops, this report will provide insight to business owners on potentially where to open a new coffee shop in Calgary.

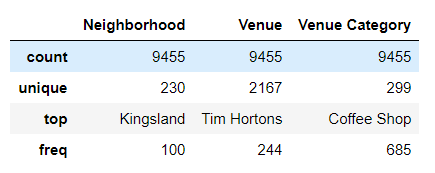
To obtain data that could be used for the assessment of coffee venues, first, the Calgary Neighbourhood and Census Dataset were scraped from Wikipedia webpage using pandas read\_html( ) method. The dataset is then cleaned and Geopy was used to obtain the geological coordinates for each neighborhood in Calgary, the resulting dataframe is shown below.

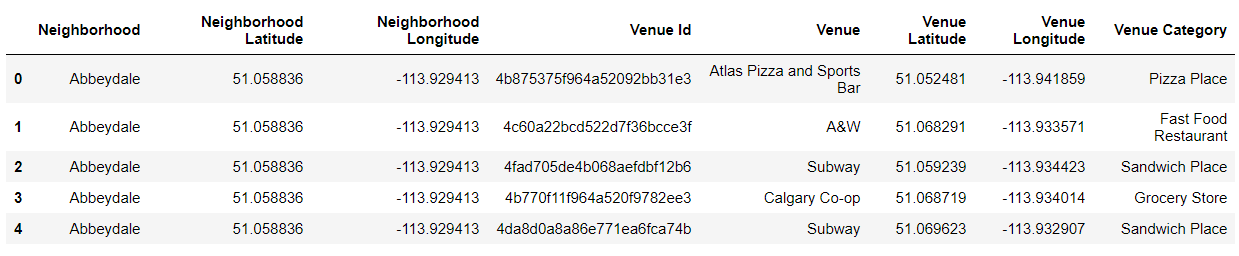


After the geological coordinates for each neighborhood were obtained, Folium was utilized to generate a choropleth map to visualize population density for each neighborhood. To produce the choropleth map, geojson data for Calgary city was utilized along with the census information obtained from the previous dataset. From the map below, it could be seen that the Calgary downtown area are the most densely populated, this area also contains the most head offices and corporate offices situated in Calgary.

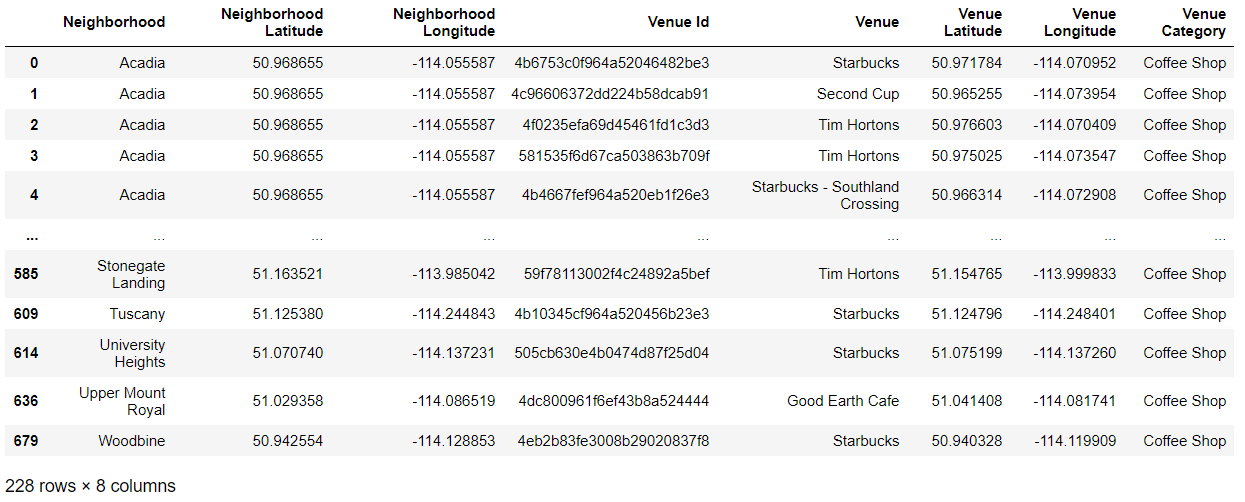


Foursquare API was utilized to generate a list of venues using the latitude and longitude information for each neighborhood. The parameters used for venue limit and radius were set to 100 and 1500m respectively. In this step, it is desired to obtain the maximum number of venues for each neighborhood to ensure that all the possible coffee venues in Calgary are included in the subsequent analysis. The resulting dataframe and summary statistics are shown below. A total of 9455 venues was returned with coffee shops accounting for the most frequent venue category.

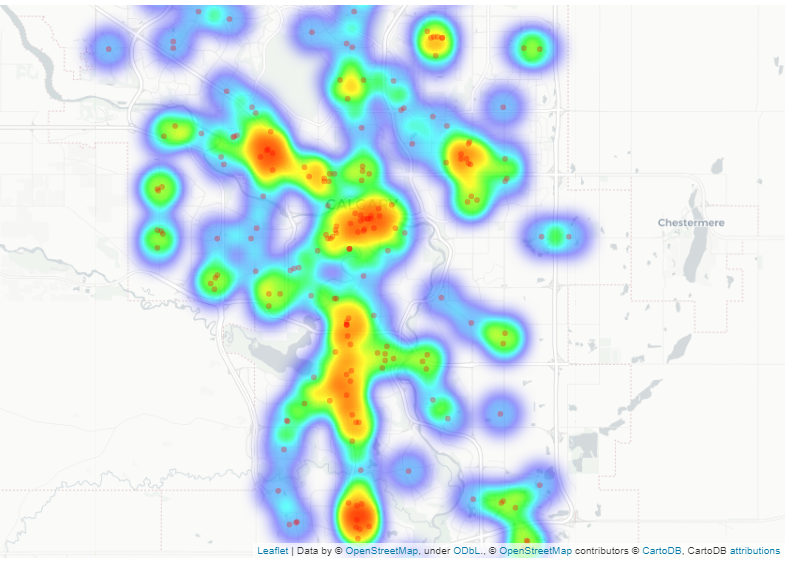




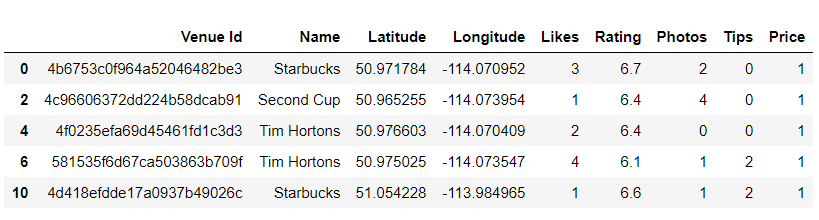
The venues list generated above could contain possible duplicates due to the broad radius parameter (1500m) specified. For subsequent analysis, unique venue Ids for coffee shops are required. The dataset is cleaned to generate a dataframe containing unique venues. It could be seen that there a total of 228 unique coffee shops in Calgary surrounding the neighborhoods that were explored earlier.



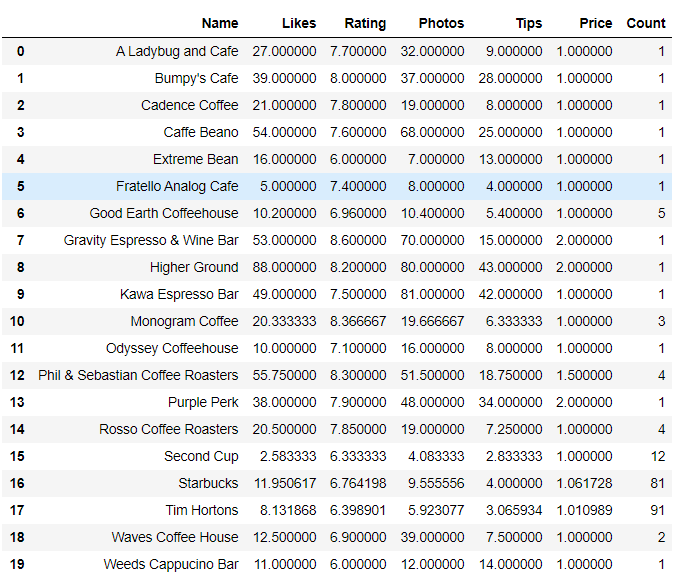
A heatmap was also generated to provide visualization on the density and spread of the coffee shops located in Calgary. From the map generated, it could be seen that there are hotspots for coffee shops in some neighborhoods while in others the coffee shops are more sparsely situated.



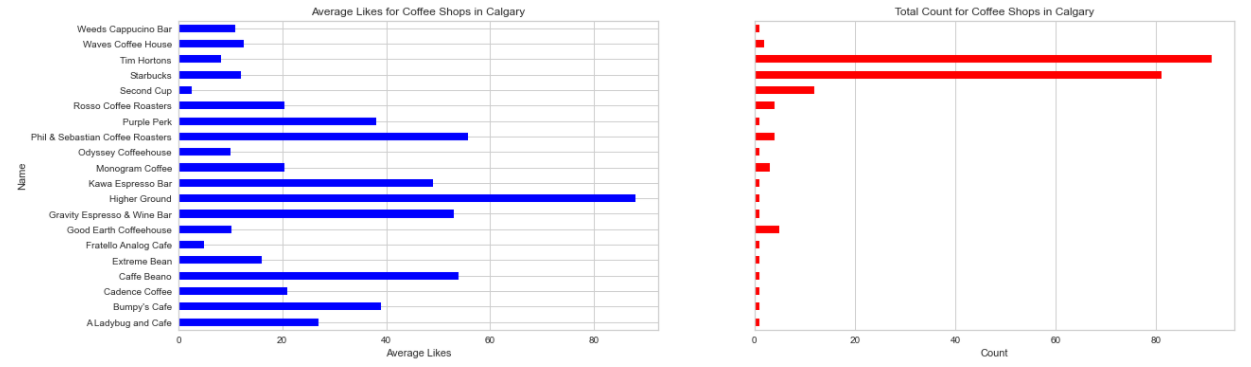
To get a better understanding of the coffee shops generated, Foursquare API was utilized again to pull the detailed information for each coffee shop venue including ratings, likes, photos, tips as well as price. The resulting data frame is shown below.



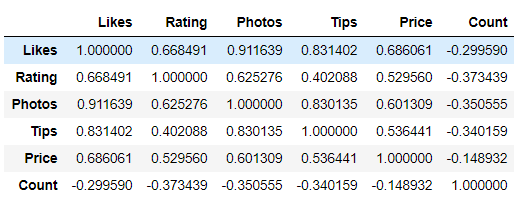
The dataset is then grouped by coffee shop names to get summary statistics on the average parameters for each unique coffee shops. From the data frame below it could be seen that Tim Hortons and Starbucks are the most common coffee shops in Calgary with counts of over 91 and 81 respectively.



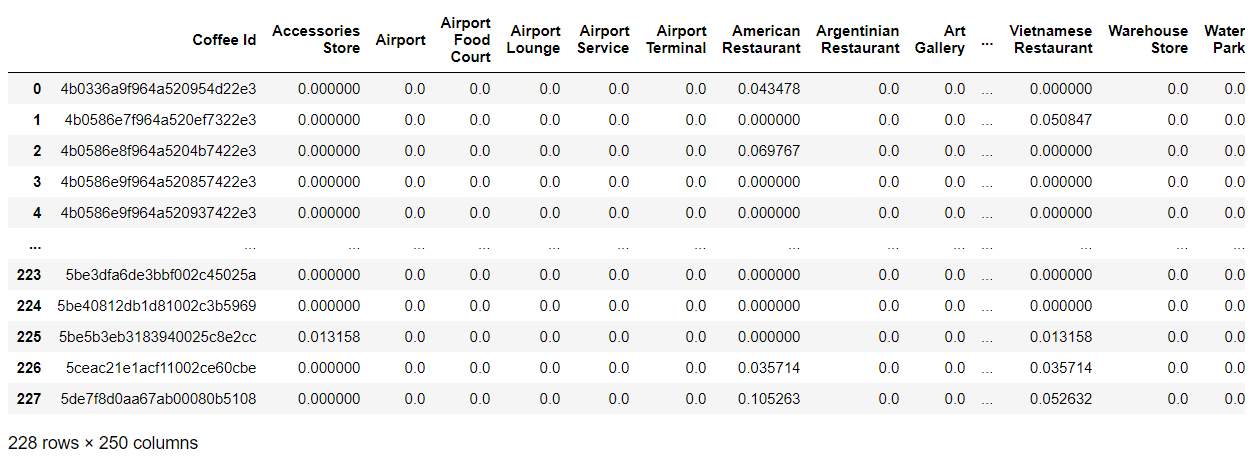
Matplotlib was used to graph the dataset to provide better visualization. The graph below shows the average likes correlated with the total count of the coffee shops. As could be seen below, coffee shops with higher occurrences tend to have lower likes due to the various coffee shop venue ratings being averaged together. While some coffee venues with single occurrences have an exceptionally high number of likes. This shows that the number of likes is not a good metric to use for further analysis.



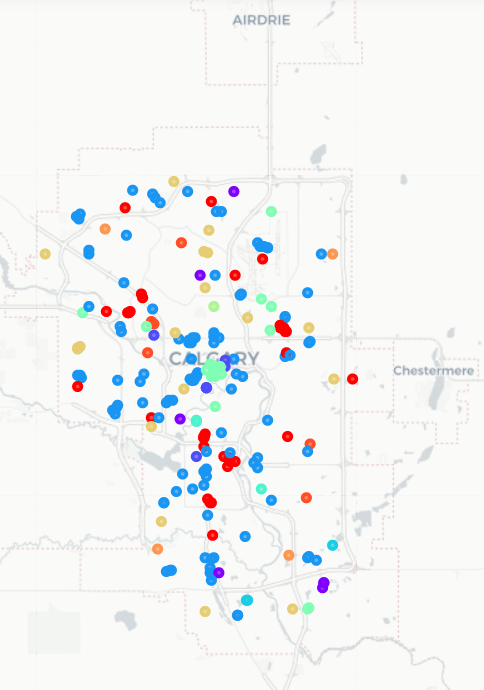
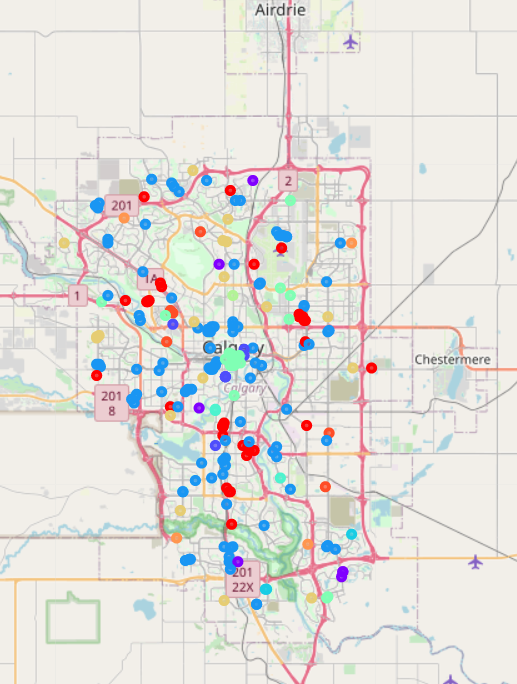
A correlation analysis was also conducted on the dataset to gauge how closely the different parameters correlate to each other. Likes are highly correlated with photos and tips, meaning that when a user like a venue, they are more likely to leave tips and photos regarding the venue. Ratings on the other hand are not as strongly correlated with likes and photos; this makes sense since the dataset contains the average ratings for all the coffee shops combined. Users who leave a good rating but do not click on likes or post photos and tips might have caused this lower correlation. However, since we are interested in the general performance of the coffee shops, rating is a good metric to use in this scenario to gauge the average consensus amongst userbase on the popularity/likeability of each coffee shop.



To answer the original question: whether neighborhood venues influence the coffee shop ratings and performance. Foursquare API was utilized to pull the surrounding venues for the list of unique coffee shops generated previously. This time the radius was set to 500m to gauge the effects of venues within 500m of each coffee shop on the coffee shop performance. The dataset is then formatted, and one-hot encoding was performed on the venue category. The resulting table is shown below:

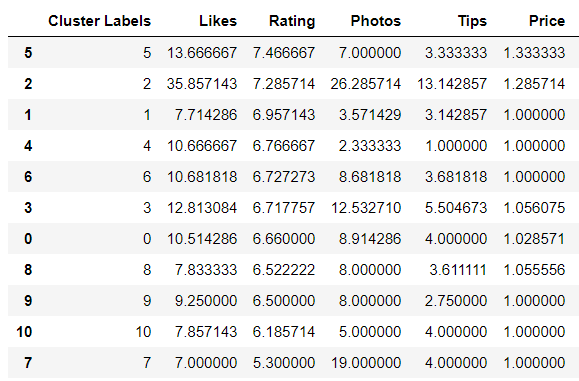


After encoding, sklearn was utilized to cluster the different coffee shops together based on their surrounding venues, k =11 was selected for this analysis. This approach was taken to group coffee shops with similar surrounding venues together. Once each cluster is generated, summary statistics analysis could then be conducted on each cluster to find out what the top venues for each cluster are and whether the top venues influence the cluster ratings. The resulting clusters are then visualized using Folium.

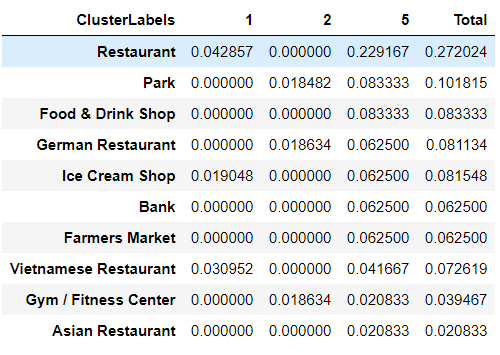
 

## Results

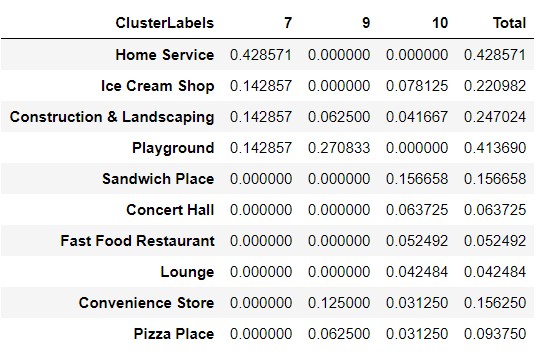
After k-means clustering was employed, the dataset is grouped by individual cluster labels and sorted by ratings in descending order. It could be seen that cluster labels 5 and 2 have the highest average rating and pricing, while cluster labels 10, 7, and 9 have the lowest average rating and price points.



To answer the question posed earlier on how the location and surrounding venues might affect the overall performance of the coffee shops. The top three clusters and bottom three clusters are further examined. By exploring the neighborhood venues of the coffee shops, this report will provide insight to business owners on potentially where to open a new coffee shop in Calgary. The top three cluster venues (5, 2, 1) are grouped and shown below, from the table below it could be seen that the highest-rated coffee shops are typically opened next to restaurants, parks, and other food and drink shops.



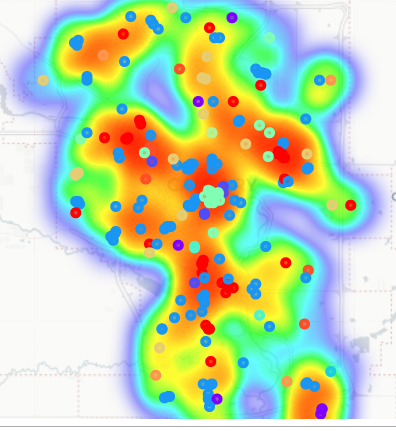
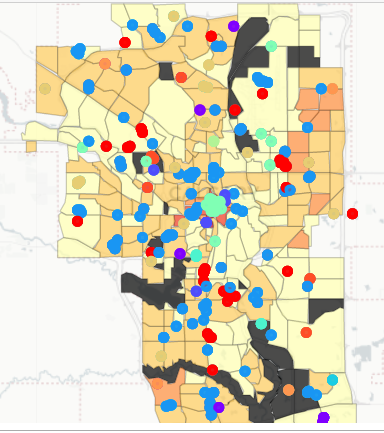
The bottom three cluster venues (7, 10, 9) are grouped and shown below. The lowest rated coffee shops are shown to be within close proximity to home services, playground, construction and landscaping, concert hall, convenience store, etc.



From the above dataset and analysis, it could be seen that coffee shops with good ratings tend to be located within proximity to restaurants, parks, and other food and drink shops. These surrounding venues provide a leisure environment for customers to enjoy the coffee in, and as a result, the coffee shops tend to have a higher average rating than the rest. On the other hand, coffee shops located near home services, construction and landscaping, playground, and convenience stores tend to have a poor rating. This could be due to the category of the surrounding venues being more utility based, and customers who visit these coffee shops are typically more rushed and thus would have less time to enjoy the coffee.

## Discussion

To provide insight to business owners on potentially where to open a new coffee shop in Calgary. The neighborhood venues of the coffee shops are clustered and plotted below in a choropleth map and heat map using Folium. Based on the population density and heat map, the optimal location to open new coffee shops is in a populated neighborhood within proximity to leisure venues such as restaurants, parks, and other food and drink shops. From the coffee shop density heat map, a location that meets the above criteria and also does not contain a large number of coffee shops can be selected.



## Conclusion

In this report, the coffee venues in Calgary city was analyzed, clustered, and grouped into different categories based on their surrounding venues. Ratings of coffee shops are then used to select the most optimal and least optimal cluster venues. However, this report only considered a few factors that might affect coffee shop ratings. For a more robust analysis, many more aspects need to be taken into account.

For instance, one problem not addressed in this analysis is that there are many different types of coffee shops and some coffee shops could have higher ratings than others based on factors other than location and surrounding venues. In addition, the rating data obtained from Foursquare is limited and only reflects user feedback from one platform. Rating also cannot be directly correlated to how well a coffee shop is operating and how much revenue it is generating. This report only serves as a preliminary analysis to gain an understanding of how different surrounding venues might affect the overall coffee shop ratings.

# Bibliography

1. Calgary. *Wikipedia.* [Online] [Cited: 11 25, 2020.] https://en.wikipedia.org/wiki/Calgary.

2. Global Liveability Ranking. *Wikipedia.* [Online] [Cited: 11 26, 2020.] https://en.wikipedia.org/wiki/Global\_Liveability\_Ranking.

3. Ly, P. U. B. L. I. S. H. E. D. J. U. (2018). Fact Sheet : Calgary Head Offices, 1–5.