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

	Neighborhood	Quadrant	Sector	Type	Population (2012)	Area (km^2)	Population Density (Pop/km^2)
0	Abbeydale	NE/SE	Northeast	Residential	5917.0	1.7	3480.6
1	Acadia	SE	South	Residential	10705.0	3.9	2744.9
2	Albert Park/Radisson Heights	SE	East	Residential	6234.0	2.5	2493.6
3	Altadore	SW	Centre	Residential	9116.0	2.9	3143.4
4	Alyth/Bonnybrook	SE	Centre	Industrial	16.0	3.8	4.2
...	...	...	...	...	...	...	...
252	Willow Park	SE	South	Residential	5229.0	3.4	1537.9
253	Windsor Park	SW	Centre	Residential	4126.0	1.3	3173.8
254	Winston Heights/Mountview	NE	Centre	Residential	3891.0	3.0	1297
255	Woodbine	SW	South	Residential	9131.0	3.2	2853.4
256	Woodlands	SW	South	Residential	6201.0	2.8	2214.6

257 rows × 7 columns

- 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

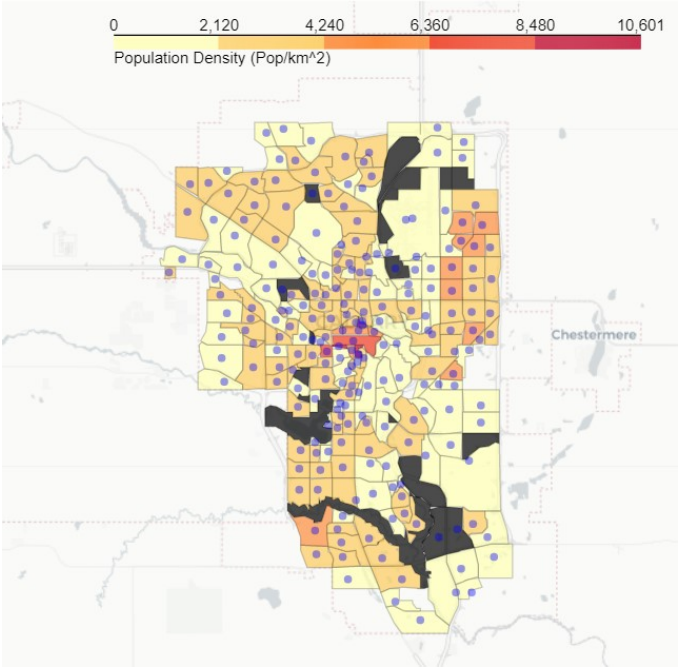
This report will focus on analyzing the coffee venues in Calgary city to offer an assessment of how the location and surrounding venues might affect the overall performance of the coffee shops. By exploring the neighborhood venues of the coffee shops, this report will provide insight to business owners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

	Neighborhood	Quadrant	Sector	Type	Population (2012)	Area (km^2)	Population Density (Pop/km^2)	Latitude	Longitude
0	Abbeydale	NE/SE	Northeast	Residential	5917.0	1.7	3480.588235	51.058836	-113.929413
1	Acadia	SE	South	Residential	10705.0	3.9	2744.871795	50.968655	-114.055587
2	Albert Park/Radisson Heights	SE	East	Residential	6234.0	2.5	2493.600000	51.044845	-113.990195
3	Altadore	SW	Centre	Residential	9116.0	2.9	3143.448276	51.015104	-114.100756
4	Alyth/Bonnybrook	SE	Centre	Industrial	16.0	3.8	4.210526	51.016669	-114.024294
...	...	...	...	...	...	...	...	...	...
252	Willow Park	SE	South	Residential	5229.0	3.4	1537.941176	50.960293	-114.054645
253	Windsor Park	SW	Centre	Residential	4126.0	1.3	3173.846154	51.006165	-114.076187
254	Winston Heights/Mountview	NE	Centre	Residential	3891.0	3.0	1297.000000	51.072303	-114.047588
255	Woodbine	SW	South	Residential	9131.0	3.2	2853.437500	50.942554	-114.128853
256	Woodlands	SW	South	Residential	6201.0	2.8	2214.642857	50.942435	-114.109359

230 rows × 9 columns

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.

	Neighborhood	Venue	Venue Category
count	9455	9455	9455
unique	230	2167	299
top	Kingsland	Tim Hortons	Coffee Shop
freq	100	244	685

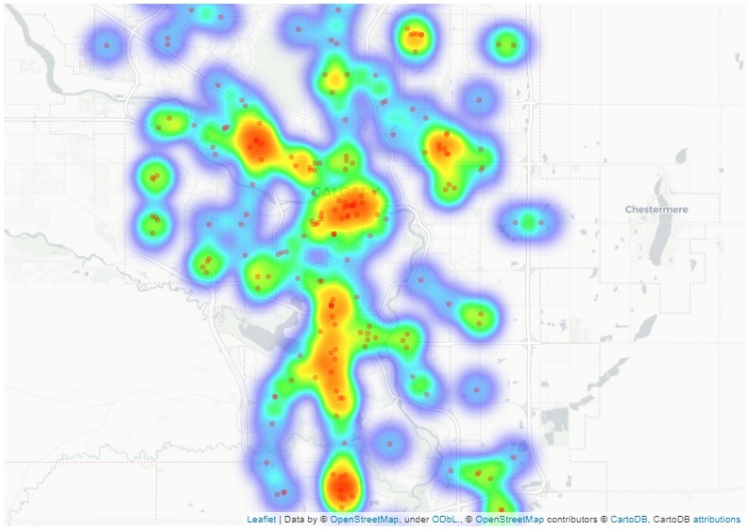
	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue Id	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Abbeydale	51.058836	-113.929413	4b875375f964a52092bb31e3	Atlas Pizza and Sports Bar	51.052481	-113.941859	Pizza Place
1	Abbeydale	51.058836	-113.929413	4c60a22bcd522d7f36bcce3f	A&W	51.068291	-113.933571	Fast Food Restaurant
2	Abbeydale	51.058836	-113.929413	4fad705de4b068aefdbf12b6	Subway	51.059239	-113.934423	Sandwich Place
3	Abbeydale	51.058836	-113.929413	4b770f11f964a520f9782ee3	Calgary Co-op	51.068719	-113.934014	Grocery Store
4	Abbeydale	51.058836	-113.929413	4da8d0a8a86e771ea6fca74b	Subway	51.069623	-113.932907	Sandwich Place

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.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue Id	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Acadia	50.968655	-114.055587	4b6753c0f964a52046482be3	Starbucks	50.971784	-114.070952	Coffee Shop
1	Acadia	50.968655	-114.055587	4c96606372dd224b58dcab91	Second Cup	50.965255	-114.073954	Coffee Shop
2	Acadia	50.968655	-114.055587	4f0235efa69d45461fd1c3d3	Tim Hortons	50.976603	-114.070409	Coffee Shop
3	Acadia	50.968655	-114.055587	581535f6d67ca503863b709f	Tim Hortons	50.975025	-114.073547	Coffee Shop
4	Acadia	50.968655	-114.055587	4b4667fef964a520eb1f26e3	Starbucks - Southland Crossing	50.966314	-114.072908	Coffee Shop
...	...	...	...	...	...	...	...	...
585	Stonegate Landing	51.163521	-113.985042	59f78113002f4c24892a5bef	Tim Hortons	51.154765	-113.999833	Coffee Shop
609	Tuscany	51.125380	-114.244843	4b10345cf964a520456b23e3	Starbucks	51.124796	-114.248401	Coffee Shop
614	University Heights	51.070740	-114.137231	505cb630e4b0474d87f25d04	Starbucks	51.075199	-114.137260	Coffee Shop
636	Upper Mount Royal	51.029358	-114.086519	4dc800961f6ef43b8a524444	Good Earth Cafe	51.041408	-114.081741	Coffee Shop
679	Woodbine	50.942554	-114.128853	4eb2b83fe3008b29020837f8	Starbucks	50.940328	-114.119909	Coffee Shop

228 rows × 8 columns

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.



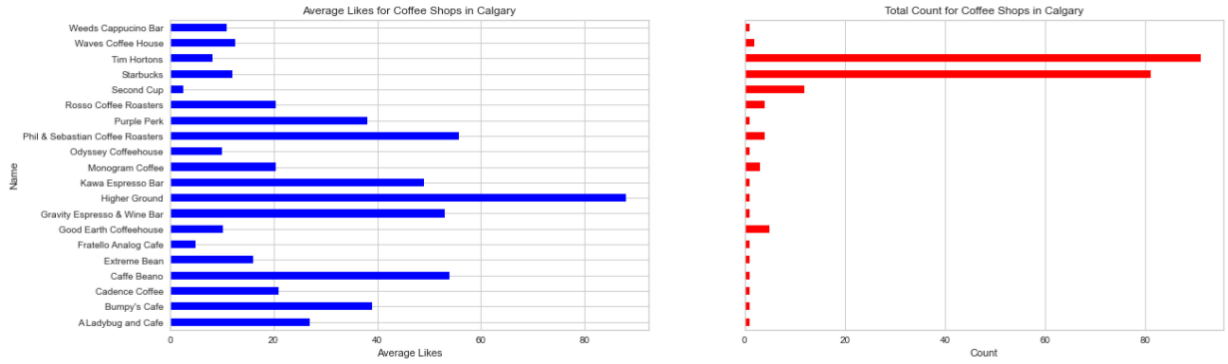
To get a better understanding of the coffee shops generated, Foursquare API was utilized again to pull the detailed information for each coffee shop venue including ratings, likes, photos, tips as well as price. The resulting data frame is shown below.

	Venue Id	Name	Latitude	Longitude	Likes	Rating	Photos	Tips	Price
0	4b6753c0f964a52046482be3	Starbucks	50.971784	-114.070952	3	6.7	2	0	1
2	4c96606372dd224b58dcab91	Second Cup	50.965255	-114.073954	1	6.4	4	0	1
4	4f0235efa69d45461fd1c3d3	Tim Hortons	50.976603	-114.070409	2	6.4	0	0	1
6	581535f6d67ca503863b709f	Tim Hortons	50.975025	-114.073547	4	6.1	1	2	1
10	4d418efdde17a0937b49026c	Starbucks	51.054228	-113.984965	1	6.6	1	2	1

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.

	Name	Likes	Rating	Photos	Tips	Price	Count
0	A Ladybug and Cafe	27.000000	7.700000	32.000000	9.000000	1.000000	1
1	Bumpy's Cafe	39.000000	8.000000	37.000000	28.000000	1.000000	1
2	Cadence Coffee	21.000000	7.800000	19.000000	8.000000	1.000000	1
3	Caffe Beano	54.000000	7.600000	68.000000	25.000000	1.000000	1
4	Extreme Bean	16.000000	6.000000	7.000000	13.000000	1.000000	1
5	Fratello Analog Cafe	5.000000	7.400000	8.000000	4.000000	1.000000	1
6	Good Earth Coffeehouse	10.200000	6.960000	10.400000	5.400000	1.000000	5
7	Gravity Espresso & Wine Bar	53.000000	8.600000	70.000000	15.000000	2.000000	1
8	Higher Ground	88.000000	8.200000	80.000000	43.000000	2.000000	1
9	Kawa Espresso Bar	49.000000	7.500000	81.000000	42.000000	1.000000	1
10	Monogram Coffee	20.333333	8.366667	19.666667	6.333333	1.000000	3
11	Odyssey Coffeehouse	10.000000	7.100000	16.000000	8.000000	1.000000	1
12	Phil & Sebastian Coffee Roasters	55.750000	8.300000	51.500000	18.750000	1.500000	4
13	Purple Perk	38.000000	7.900000	48.000000	34.000000	2.000000	1
14	Rosso Coffee Roasters	20.500000	7.850000	19.000000	7.250000	1.000000	4
15	Second Cup	2.583333	6.333333	4.083333	2.833333	1.000000	12
16	Starbucks	11.950617	6.764198	9.555556	4.000000	1.061728	81
17	Tim Hortons	8.131868	6.398901	5.923077	3.065934	1.010989	91
18	Waves Coffee House	12.500000	6.900000	39.000000	7.500000	1.000000	2
19	Weeds Cappucino Bar	11.000000	6.000000	12.000000	14.000000	1.000000	1

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.

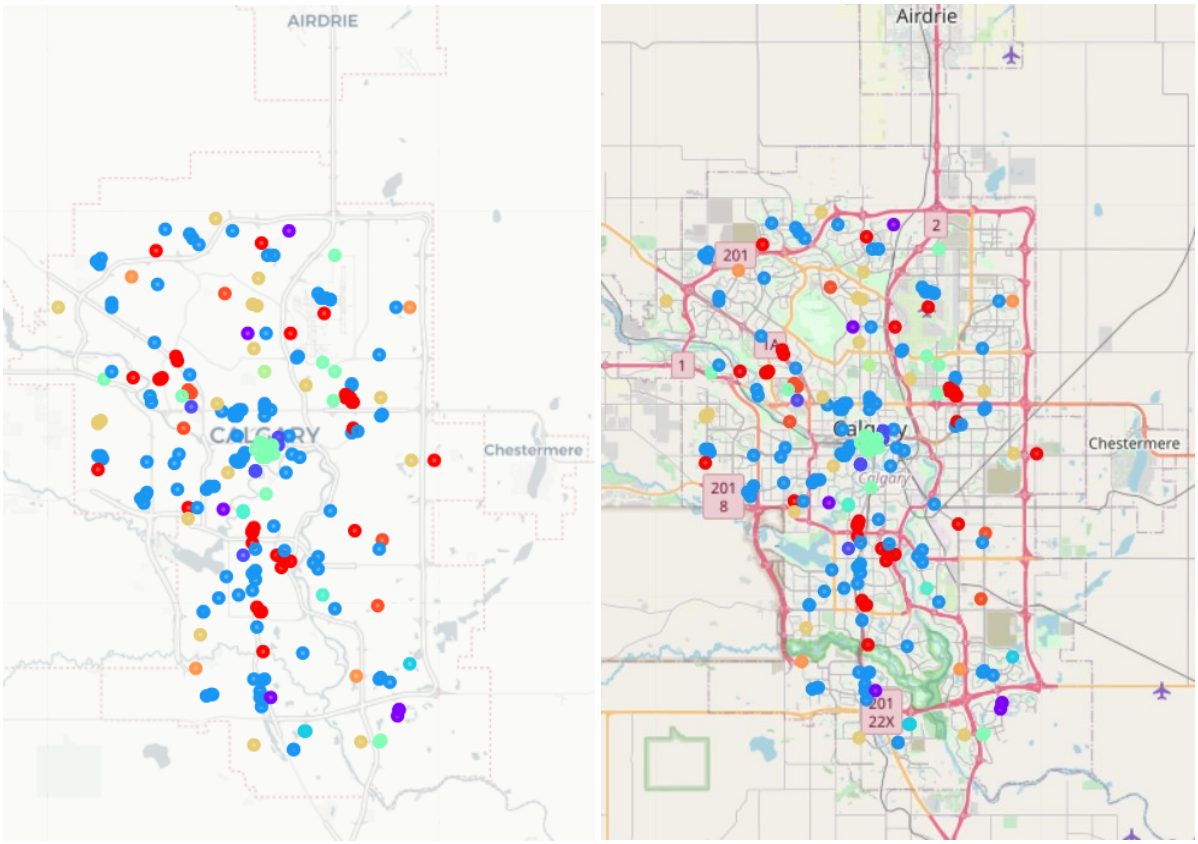
	Likes	Rating	Photos	Tips	Price
Likes	1.000000	0.668491	0.911639	0.831402	0.686061
Rating	0.668491	1.000000	0.625276	0.402088	0.529560
Photos	0.911639	0.625276	1.000000	0.830135	0.601309
Tips	0.831402	0.402088	0.830135	1.000000	0.536441
Price	0.686061	0.529560	0.601309	0.536441	1.000000
Count	-0.299590	-0.373439	-0.350555	-0.340159	-0.148932

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:

	Coffee Id	Accessories Store	Airport	Airport Food Court	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	Argentinian Restaurant	Art Gallery	...	Vietnamese Restaurant	Warehouse Store	Water Park
0	4b0336a9f964a520954d22e3	0.000000	0.0	0.0	0.0	0.0	0.0	0.043478	0.0	0.0	...	0.000000	0.0	0.0
1	4b0586e7f964a520ef7322e3	0.000000	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	...	0.050847	0.0	0.0
2	4b0586e8f964a5204b7422e3	0.000000	0.0	0.0	0.0	0.0	0.0	0.069767	0.0	0.0	...	0.000000	0.0	0.0
3	4b0586e9f964a520857422e3	0.000000	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	...	0.000000	0.0	0.0
4	4b0586e9f964a520937422e3	0.000000	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	...	0.000000	0.0	0.0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
223	5be3dfa6de3bbf002c45025a	0.000000	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	...	0.000000	0.0	0.0
224	5be40812db1d81002c3b5969	0.000000	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	...	0.000000	0.0	0.0
225	5be5b3eb3183940025c8e2cc	0.013158	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	...	0.013158	0.0	0.0
226	5ceac21e1acf11002ce60cbe	0.000000	0.0	0.0	0.0	0.0	0.0	0.035714	0.0	0.0	...	0.035714	0.0	0.0
227	5de7f8d0aa67ab00080b5108	0.000000	0.0	0.0	0.0	0.0	0.0	0.105263	0.0	0.0	...	0.052632	0.0	0.0

228 rows × 250 columns

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.

	Cluster Labels	Likes	Rating	Photos	Tips	Price
5	5	13.666667	7.466667	7.000000	3.333333	1.333333
2	2	35.857143	7.285714	26.285714	13.142857	1.285714
1	1	7.714286	6.957143	3.571429	3.142857	1.000000
4	4	10.666667	6.766667	2.333333	1.000000	1.000000
6	6	10.681818	6.727273	8.681818	3.681818	1.000000
3	3	12.813084	6.717757	12.532710	5.504673	1.056075
0	0	10.514286	6.660000	8.914286	4.000000	1.028571
8	8	7.833333	6.522222	8.000000	3.611111	1.055556
9	9	9.250000	6.500000	8.000000	2.750000	1.000000
10	10	7.857143	6.185714	5.000000	4.000000	1.000000
7	7	7.000000	5.300000	19.000000	4.000000	1.000000

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.

ClusterLabels	1	2	5	Total
Restaurant	0.042857	0.000000	0.229167	0.272024
Park	0.000000	0.018482	0.083333	0.101815
Food & Drink Shop	0.000000	0.000000	0.083333	0.083333
German Restaurant	0.000000	0.018634	0.062500	0.081134
Ice Cream Shop	0.019048	0.000000	0.062500	0.081548
Bank	0.000000	0.000000	0.062500	0.062500
Farmers Market	0.000000	0.000000	0.062500	0.062500
Vietnamese Restaurant	0.030952	0.000000	0.041667	0.072619
Gym / Fitness Center	0.000000	0.018634	0.020833	0.039467
Asian Restaurant	0.000000	0.000000	0.020833	0.020833

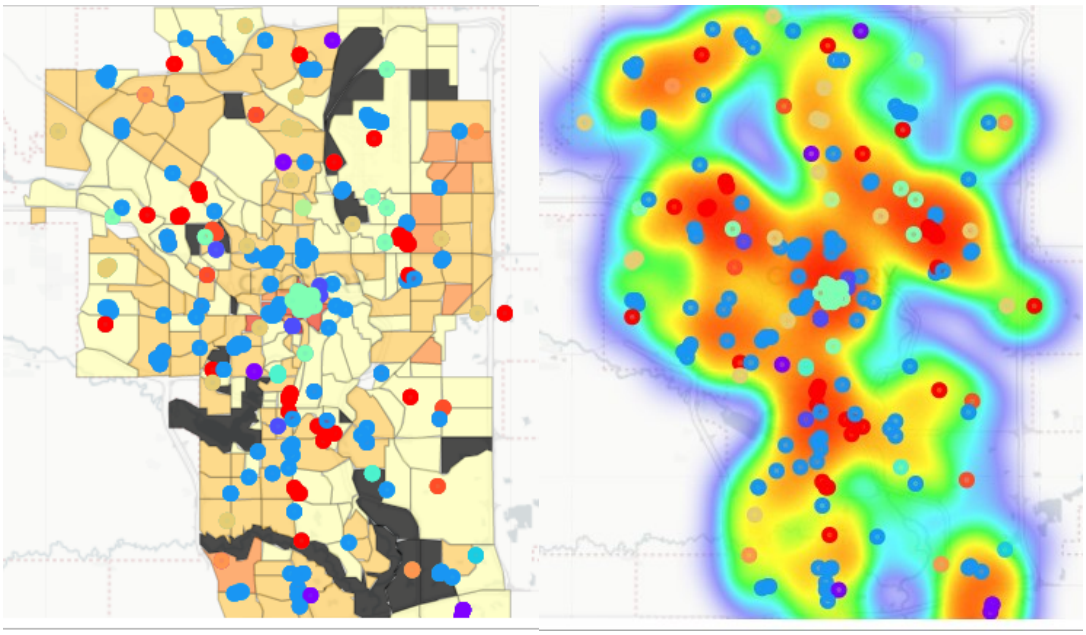
The bottom three cluster venues (7, 10, 9) are grouped and shown below. The lowest rated coffee shops are shown to be within close proximity to home services, playground, construction and landscaping, concert hall, convenience store, etc.

ClusterLabels	7	9	10	Total
Home Service	0.428571	0.000000	0.000000	0.428571
Ice Cream Shop	0.142857	0.000000	0.078125	0.220982
Construction & Landscaping	0.142857	0.062500	0.041667	0.247024
Playground	0.142857	0.270833	0.000000	0.413690
Sandwich Place	0.000000	0.000000	0.156658	0.156658
Concert Hall	0.000000	0.000000	0.063725	0.063725
Fast Food Restaurant	0.000000	0.000000	0.052492	0.052492
Lounge	0.000000	0.000000	0.042484	0.042484
Convenience Store	0.000000	0.125000	0.031250	0.156250
Pizza Place	0.000000	0.062500	0.031250	0.093750

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.