Battle of Neighbourhoods: Neighbourhood Analysis of the City of Calgary

By: Ayda Takehei

1. Introduction

1.1 Background

Moving to another city is extremely difficult when you don't know much about the place. Which neighbourhood is the safest? Where are the majority of cafes/restaurants or parks located? Which neighbourhood has the most amount of traffic? Which is the quietest? These are amongst the many questions that cross people's minds when moving to another city. Being well prepared and knowing what to expect can result in a smooth relocation. One of the biggest challenges in this journey is finding a suitable new home that meets a person's interests and expectations.

1.2 Business Problem

Calgary, a city located in western Canada, was ranked the most livable city in North America ("Calgary Facts | Life in Calgary", 2021). This extremely sunny city has diverse communities and endless opportunities. It is no surprise that this gateway to the beautiful Rocky Mountains experiences high population growth. As new families move into this beautiful city, the biggest question that arises is which neighbourhood could they call home? A real estate company has asked us to study the market to help buyers decide which neighbourhoods would better suit their needs and interests. The main task is to analyze neighbourhoods in Calgary using various factors that play into the decision making process such as type of venues/amenities, population density and crime rate. Combined, these data can help buyers/migrants decide which neighbourhood to settle in.

The target audience for this project includes people who are interested in moving to the city of Calgary or those who would like to become homeowners.

2.Data

2.1 Data Sources

Data required for this study was retrieved from the sources listed below.

1. The list of Calgary neighborhoods and their population densities were scraped from a Wikipedia page using the pandas library method. The following list was then converted into a pandas dataframe.
2. The geographical coordinates of each neighbourhood were extracted using the python GeoPy library. This data was then used to plot the map of Calgary containing markers for each neighbourhood.
3. The Foursquare API was used to obtain data of the top 100 venues that are within a 500m radius from each neighborhood marker.
4. The list of Calgary Neighbourhood crime data for the year 2021 was retrieved from the 'Open Calgary' website using the GET request in HTTP methods.

2.2 Data Acquisition and Wrangling

2.21 Calgary neighbourhood data

The pandas.read_html() function was used to scrape a table containing all the neighbourhoods in Calgary from the Wikipedia website. This function uses the scraping library Urllib to return the table as the following dataframe:

	Name[10]	Quadrant	Sector[11]	Ward[12]	Type[11]	2012 PopulationRank	Population(2012)[10]	Population(2011)[10]	% change	Dwellings(2012)[10]	Area(km2)[11]	Populationdensity
0	Abbeydale	NE/SE	Northeast	10	Residential	82	5917.0	5700.0	3.8	2023.0	1.7	3480.6
1	Acadia	SE	South	9	Residential	27	10705.0	10615.0	0.8	5053.0	3.9	2744.9
2	Albert Park/Radisson Heights	SE	East	10	Residential	75	6234.0	6217.0	0.3	2709.0	2.5	2493.6
3	Altadore	SW	Centre	11	Residential	39	9116.0	8907.0	2.3	4486.0	2.9	3143.4
4	Alyth/Bonnybrook	SE	Centre	9	Industrial	208	16.0	17.0	–5.9	14.0	3.8	4.2
5	Applewood Park	SE/NE	East	10	Residential	69	6498.0	6404.0	1.5	2215.0	1.6	4061.3
6	Arbour Lake	NW	Northwest	2	Residential	26	10836.0	10762.0	0.7	3918.0	4.4	2462.7
7	Aspen Woods	SW	West	6	Residential	92	5271.0	4469.0	17.9	2281.0	3.8	1387.1
8	Auburn Bay	SE	Southeast	12	Residential	60	7193.0	5769.0	24.7	2808.0	4.5	1598.4
9	Aurora Business Park	NE	North	3	Industrial	237	0.0	0.0	—	0.0	2.4	0

Table 1: Raw neighbourhood data retrieved from the Wikipedia website.

Many irrelevant columns were omitted from the dataframe as shown below.

	Neighbourhoods	Population Density	Latitude	Longitude
0	Abbeydale	3480.6	51.058836	-113.929413
1	Acadia	2744.9	50.968655	-114.055587
2	Albert Park/Radisson Heights	2493.6	51.044845	-113.990195
3	Altadore	3143.4	51.015104	-114.100756
4	Alyth/Bonnybrook	4.2	51.016669	-114.024294
5	Applewood Park	4061.3	51.044658	-113.928931
6	Arbour Lake	2462.7	51.136786	-114.202355
7	Aspen Woods	1387.1	51.043119	-114.210185
8	Auburn Bay	1598.4	50.890605	-113.959565
10	Banff Trail	2558	51.071462	-114.111647

Table 2: Clean neighbourhood data.

2.22 Geographical Neighbourhood Coordinates

The geocoding web service from geopy is used to geolocate a query, the list of Calgary neighbourhoods in our case, to coordinates. First, geocoding was used to obtain the coordinates of the city of Calgary (Figure 1).

```
geolocator = Nominatim(user_agent="Mozilla/76.0")
location = geolocator.geocode("Calgary")
latitude = location.latitude
longitude = location.longitude
print(f"Coordinates of Calgary are {latitude}, {longitude}")
```

```
Coordinates of Calgary are 51.0460954, -114.065465
```

Figure 1: Latitude and longitude coordinates of the city of Calgary.

Next, the coordinates of the different neighbourhoods were obtained. However since there are too many neighbourhoods to locate, RteLimiter() function was used to prevent the service from overloading as it allows bulk operations to perform while gracefully handling error responses and adding delays when needed ("Welcome to GeoPy's documentation! — GeoPy 2.2.0 documentation", 2021). A delay of 1 second was made between calls to the geocoding service and the tqdm function was used to display the progress bar. This is all seen in Figure 2.

```
from tqdm import tqdm
from geopy.extra.rate_limiter import RateLimiter

tqdm.pandas()
geocode = RateLimiter(geolocator.geocode, min_delay_seconds=1)
coords = (calgary_raw['Neighbourhoods'] + ' suburb, Calgary').progress_apply(geocode)
```

```
100%|██████████| 258/258 [04:26<00:00,  1.03s/it]
```

Figure 2: Display of coding used to obtain coordinates of neighbourhoods in Calgary.

The geographical coordinates were then added to the dataframe and used to create a map of Calgary with neighborhoods superimposed on top (Figure 3). This was achieved using the Folium library.

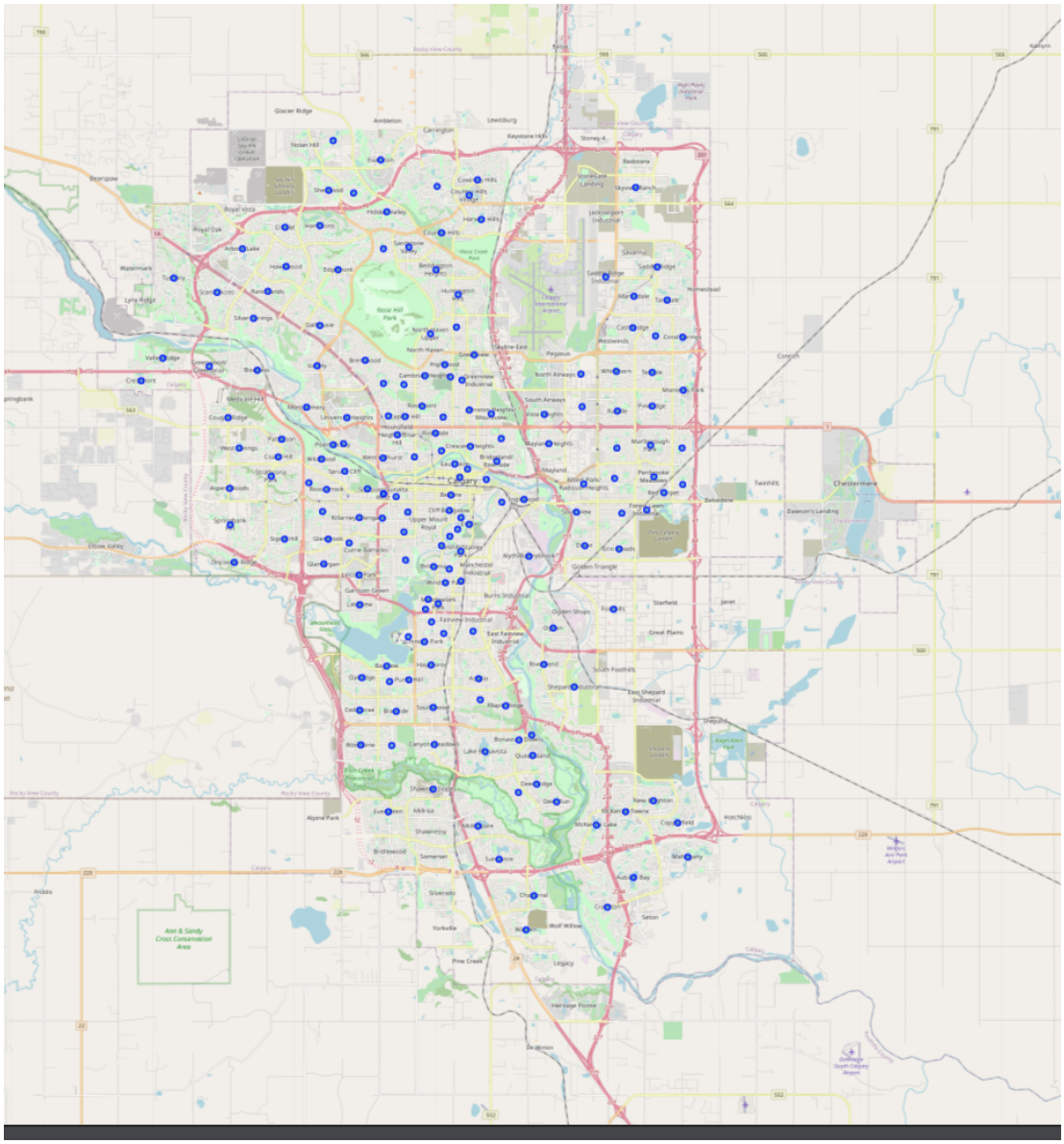


Figure 3: Map of the city of Calgary with neighbourhood coordinates superimposed on top.

2.23 Neighbourhood Venue Data

To extract the top 100 venues within a 500m radius from each neighbourhood marker, the explore function in the Foursquare API was used.Foursquare is a mobile app developed to provide users with accurate location-aware recommendations of places to visit. A foursquare developer account needs to be registered to obtain the Foursquare ID and secret key. After, API calls to Foursquare are made with the geographical coordinates of the neighbourhoods and as a result a venue data in JSON format is returned. Along with the venue name, the venue category and venue latitude and longitude are also extracted from the API, then stored as a new dataframe as seen in

Table 3. Knowledge of the different type and number of venues in each neighbourhood provides a general overview of each neighbourhood, allowing citizens to choose one based on their needs and preference. That being the case, a bar chart displaying the number of venues within each neighbourhood was created for visualization purposes (Figure 4).

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Abbeydale	51.058836	-113.929413	Subway	51.059239	-113.934423	Sandwich Place
1	Abbeydale	51.058836	-113.929413	Mac's	51.059376	-113.934425	Convenience Store
2	Abbeydale	51.058836	-113.929413	roadside pub	51.059277	-113.934529	Wings Joint
3	Abbeydale	51.058836	-113.929413	Ryan Goeseels	51.054479	-113.927731	Home Service
4	Acadia	50.968655	-114.055587	Eagles Club	50.968756	-114.052836	Pub

Table 3: Dataframe displaying the venue, venue coordinates and category in each neighbourhood.

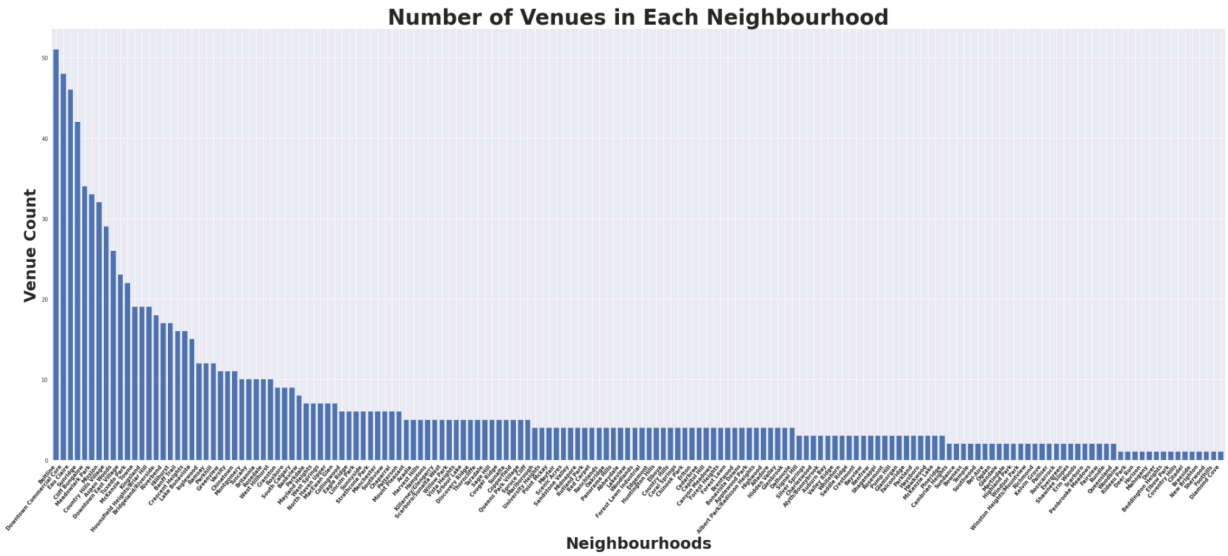


Figure 4: Bar graph displaying the venue count in each neighbourhood.

2.24 Neighborhood Crime Data

To determine the crime statistics in the neighbourhoods of Calgary, the GET request in HTTP methods was used to extract the number of crime incidents found in each neighbourhood in Calgary in the year 2021. Once this was extracted, all types of crime that occured in each neighbourhood were summed up and all unnecessary data was dropped from the table, shown below.

	community_name	year	long	lat	total
0	10E	2021	-113.91623374150296	51.081453566081535	1
1	12A	2021	-113.89321145110586	50.96505334266445	2
2	12J	2021	-113.88346245083656	50.883575760532	1
3	13F	2021	-114.17505002464321	50.89941005804682	1
4	ABBEYDALE	2021	-113.92779391882777	51.05952012318639	15
5	02K	2020	-114.18058839890571	51.19042746687576	1
6	ACADIA	2021	-114.05369821963635	50.972409209651765	22
7	ALBERT PARK/RADISSON HEIGHTS	2021	-113.99678382147289	51.04454101383472	53
8	ALTADORE	2021	-114.10079474153764	51.015948466883415	28
9	ALYTH/BONNYBROOK	2021	-114.02248220564552	51.02285932043044	4
10	APPLEWOOD PARK	2021	-113.92785764420194	51.044981778174524	18

Table 4: Cleaned up crime data extracted off of ‘Open Calgary’ website.

3.Data Analyses

3.1 Exploratory Data Analyses

To better analyze the neighbourhoods, the venue's data must be organized into unique categories to get a grasp of concepts in each neighbourhood. Then, one-hot encoding is performed on the dataset to convert the categorical features to binary values. Thus, a value of 1 indicates that this category is found in the neighbourhood and 0 indicates that it is not. Then, the venues are grouped by neighbourhoods and the mean of the frequency of occurrence of each venue is computed (Table 5). Based on the mean frequency, the top 10 most common venues will be computed for each neighbourhood (Table 6). This data is of great significance as it will be used in preparation of clustering as it allows for the classification of each neighbourhood.

```
----Abbeydale----
                venue  freq
0         Wings Joint  0.25
1  Convenience Store  0.25
2     Sandwich Place  0.25
3       Home Service  0.25
4           Pharmacy  0.00
```

Table 5: The mean frequency of occurrence of venues in the first neighbourhood.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Abbeydale	Wings Joint	Convenience Store	Sandwich Place	Home Service	Pharmacy	Optical Shop	Movie Theater	Moving Target	Museum	Music Store
1	Acadia	Construction & Landscaping	Pub	Furniture / Home Store	Athletics & Sports	Other Great Outdoors	Moving Target	Museum	Music Store	Music Venue	New American Restaurant
2	Albert Park/Radisson Heights	Pizza Place	Asian Restaurant	Convenience Store	Light Rail Station	Yoga Studio	Other Great Outdoors	Movie Theater	Moving Target	Museum	Music Store
3	Altadore	Dog Run	Greek Restaurant	Massage Studio	Coffee Shop	Yoga Studio	Other Great Outdoors	Moving Target	Museum	Music Store	Music Venue
4	Alyth/Bonnybrook	Gas Station	Transportation Service	Restaurant	Yoga Studio	Optical Shop	Movie Theater	Moving Target	Museum	Music Store	Music Venue
5	Applewood Park	Liquor Store	Pizza Place	Park	Home Service	Yoga Studio	Movie Theater	Moving Target	Museum	Music Store	Music Venue
6	Arbour Lake	Bus Station	Lake	Residential Building (Apartment / Condo)	Grocery Store	Construction & Landscaping	Yoga Studio	Other Great Outdoors	Museum	Music Store	Music Venue
7	Aspen Woods	Coffee Shop	Restaurant	Yoga Studio	Boutique	Electronics Store	Pharmacy	Pet Store	Park	Cupcake Shop	Salon / Barbershop
8	Auburn Bay	Tennis Court	Lake	Other Repair Shop	Yoga Studio	Motorcycle Shop	Movie Theater	Moving Target	Museum	Music Store	Music Venue
9	Banff Trail	Hotel	Liquor Store	Ice Cream Shop	Steakhouse	Breakfast Spot	Park	BBQ Joint	Pizza Place	Seafood Restaurant	Vietnamese Restaurant
10	Bankview	Pizza Place	Convenience Store	Spa	Candy Store	Health & Beauty Service	Tattoo Parlor	Coffee Shop	Park	Paper / Office Supplies Store	Moving Target

Table 6: Table displaying the top ten most common venues of ten Calgary neighbourhoods.

3.2 Machine Learning Algorithm: K-Means Clustering

Unsupervised learning is a machine learning technique used to analyze and cluster a set of unlabelled data, without the need of human intervention (Education, 2021). This technique uses machine learning algorithms to discover similarities and differences in a given dataset clustering them accordingly.

K-mean clustering is one of the most popular machine learning algorithms used in exploratory data analysis. K-means clustering groups similar data together into a cluster, which is a collection of data aggregated together based on similarities ("Understanding K-means Clustering in Machine Learning", 2021). The algorithm looks for a fixed number of clusters (k) or centroids, a data point that is in the center of a cluster, in a given dataset. Every data point within the dataset is then allocated to a cluster while minimizing the intra-cluster distances.

The optimal value of k can be determined using the elbow method number for k by plotting variation as a function of the number of clusters in a dataset. Once a curve has been plotted, the elbow of the curve will be the k value. In Figure 5, the elbow occurs at a k value of 2. The 2 centroids are then chosen, and iterative calculations are performed to optimize the positions of the centroids until the centroids no longer move.

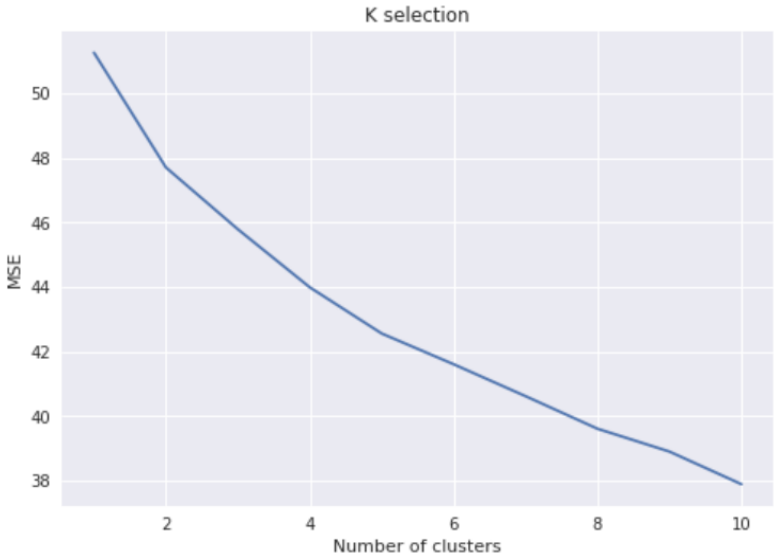


Figure 5: Line graph of variation as a function of the number of clusters.

4. Results

4.1 K-means Clusters

Using the k-means clustering method, the optimal value of k was discovered to be 2 so a total of 2 clusters were identified. Both clusters with the top ten venues in each neighbourhood are seen in the following tables in Figure 6. In addition, a station column was added to the data to indicate the type of station found in each neighbourhood, in support of determining the degree of reliance on transportation. To better visualize the clusters, a bar graph was made to show the vast difference in frequency of occurrence of venues in both clusters (Figure 7).

a.

	Neighbourhoods	Population Density	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue	Station
0	Abbeydale	3480.6	51.058836	-113.929413	0	Wings Joint	Convenience Store	Sandwich Place	Home Service	Pharmacy	Optical Shop	Movie Theater	Moving Target	Museum	Music Store	No
1	Acadia	2744.9	50.968655	-114.055587	0	Construction & Landscaping	Pub	Furniture / Home Store	Athletics & Sports	Other Great Outdoors	Moving Target	Museum	Music Store	Music Venue	New American Restaurant	No
2	Albert Park/Radisson Heights	2493.6	51.044845	-113.990195	0	Pizza Place	Asian Restaurant	Convenience Store	Light Rail Station	Yoga Studio	Other Great Outdoors	Movie Theater	Moving Target	Museum	Music Store	Yes
3	Altadore	3143.4	51.015104	-114.100756	0	Dog Run	Greek Restaurant	Massage Studio	Coffee Shop	Yoga Studio	Other Great Outdoors	Moving Target	Museum	Music Store	Music Venue	No
4	Alyth/Bonnybrook	4.2	51.016669	-114.024294	0	Gas Station	Transportation Service	Restaurant	Yoga Studio	Optical Shop	Movie Theater	Moving Target	Museum	Music Store	Music Venue	Yes
5	Applewood Park	4061.3	51.044658	-113.928931	0	Liquor Store	Pizza Place	Park	Home Service	Yoga Studio	Movie Theater	Moving Target	Museum	Music Store	Music Venue	No
6	Arbour Lake	2462.7	51.136786	-114.202355	0	Bus Station	Lake	Residential Building (Apartment / Condo)	Grocery Store	Construction & Landscaping	Yoga Studio	Other Great Outdoors	Museum	Music Store	Music Venue	Yes
7	Aspen Woods	1387.1	51.043119	-114.210185	0	Coffee Shop	Restaurant	Yoga Studio	Boutique	Electronics Store	Pharmacy	Pet Store	Park	Cupcake Shop	Salon / Barbershop	No
8	Auburn Bay	1598.4	50.890605	-113.959565	0	Tennis Court	Lake	Other Repair Shop	Yoga Studio	Motorcycle Shop	Movie Theater	Moving Target	Museum	Music Store	Music Venue	No
10	Banff Trail	2558	51.071462	-114.111647	0	Hotel	Liquor Store	Ice Cream Shop	Steakhouse	Breakfast Spot	Park	BBQ Joint	Pizza Place	Seafood Restaurant	Vietnamese Restaurant	No

b.

	Neighbourhoods	Population Density	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue	Station
71	Erin Woods	4313.1	51.019345	-113.968367	1	Park	Yoga Studio	Optical Shop	Motorcycle Shop	Movie Theater	Moving Target	Museum	Music Store	Music Venue	New American Restaurant	No
72	Erlton	2468	51.029086	-114.061276	1	Park	Café	History Museum	Yoga Studio	Optical Shop	Movie Theater	Moving Target	Museum	Music Store	Music Venue	No
84	Glamorgan	3188	51.013442	-114.151697	1	Construction & Landscaping	Park	Yoga Studio	Other Great Outdoors	Movie Theater	Moving Target	Museum	Music Store	Music Venue	New American Restaurant	No
129	McKenzie Lake	2754.9	50.911282	-113.982636	1	Lake	Park	Yoga Studio	Optical Shop	Motorcycle Shop	Movie Theater	Moving Target	Museum	Music Store	Music Venue	No
139	New Brighton	2863.1	50.920726	-113.947085	1	Park	Yoga Studio	Optical Shop	Motorcycle Shop	Movie Theater	Moving Target	Museum	Music Store	Music Venue	New American Restaurant	No
182	Rideau Park	2120	51.024344	-114.073599	1	Park	Yoga Studio	Optical Shop	Motorcycle Shop	Movie Theater	Moving Target	Museum	Music Store	Music Venue	New American Restaurant	No
186	Rosemont	2482	51.075345	-114.090674	1	Dog Run	Athletics & Sports	Park	Yoga Studio	Other Great Outdoors	Movie Theater	Moving Target	Museum	Music Store	Music Venue	No
202	Shaganappi	1035.3	51.042939	-114.124599	1	Golf Course	Park	Light Rail Station	Yoga Studio	Optical Shop	Movie Theater	Moving Target	Museum	Music Store	Music Venue	Yes
206	Sherwood	1051.3	51.159564	-114.148693	1	Park	Yoga Studio	Optical Shop	Motorcycle Shop	Movie Theater	Moving Target	Museum	Music Store	Music Venue	New American Restaurant	No
207	Signal Hill	2484.6	51.023313	-114.176233	1	Business Service	Park	Coffee Shop	Yoga Studio	Optical Shop	Movie Theater	Moving Target	Museum	Music Store	Music Venue	Yes
208	Silver Springs	1786.6	51.109649	-114.195402	1	Spa	Park	Home Service	Yoga Studio	Noodle House	Motorcycle Shop	Movie Theater	Moving Target	Museum	Music Store	No

Figure 6: Tables displaying the top ten venues of ten neighbourhoods in a) Cluster 1 and b) Cluster 2.

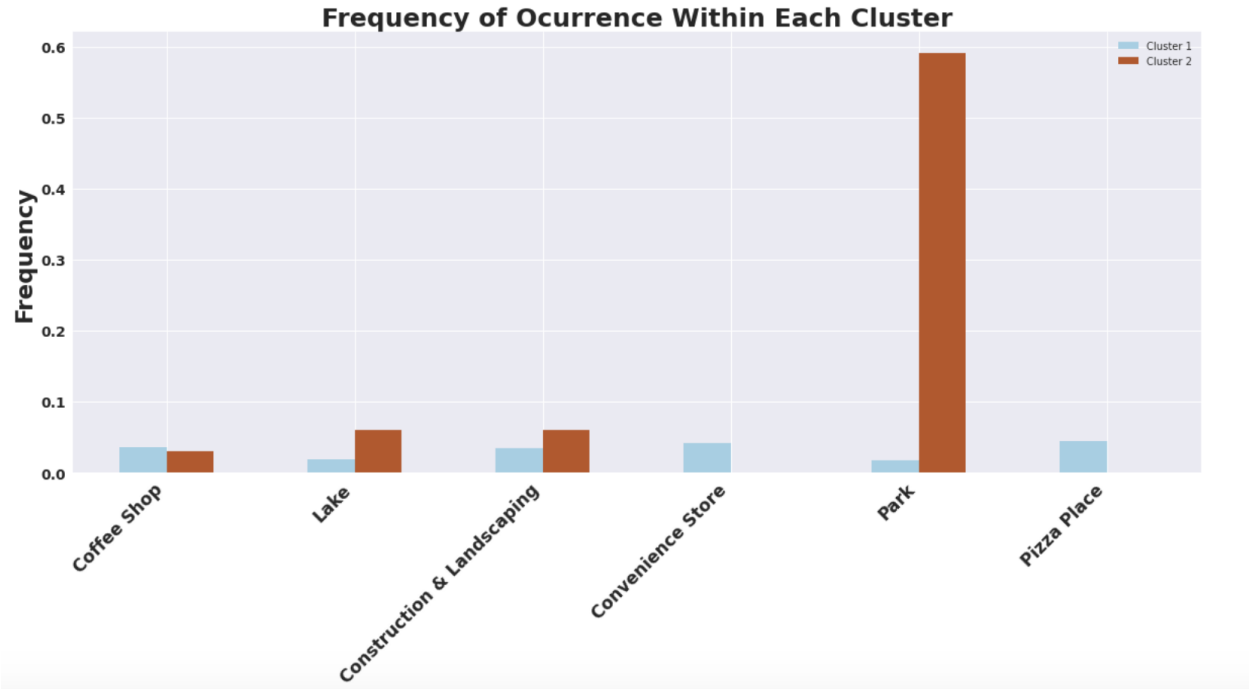


Figure 7: Bar graph displaying the frequency of occurrence of types of venue within each cluster.

4.2 Data Visualization

4.21 Cluster Visualization

To better visualize both clusters and their respective locations, a cluster map was created using folium. Each cluster is color-coded on the map to provide a quick overview of where the venue categories are located. The cluster map is found below.

Legend:
Red: Cluster 1
Purple: Cluster 2

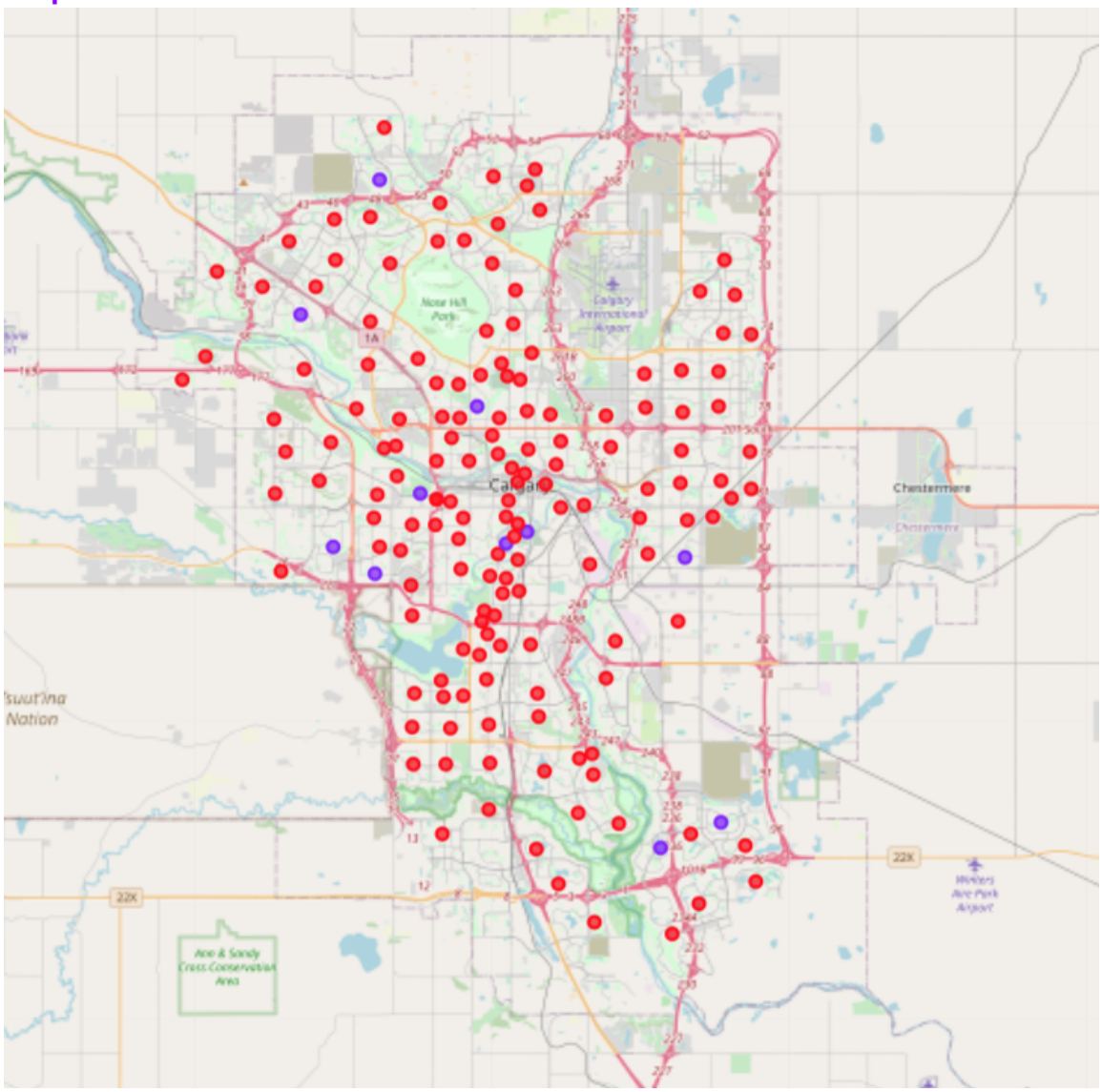


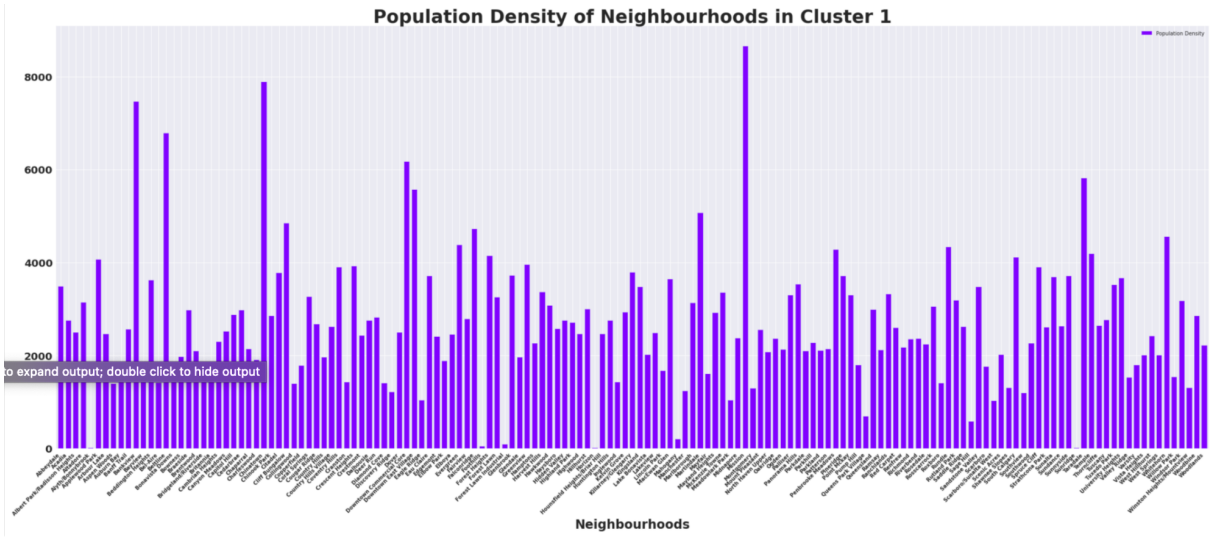
Figure 8: Map of the city of Calgary with both clusters color-coded and superimposed on top.

4.22 Population Density Visualization

Aside from the type of amenities and venues, population density is also a major factor that many people consider when determining which neighbourhood to settle in. Population density is the concentration of individuals within a specific geographic locale (Society, 2021). It can be used to assess relationships with human habits,infrastructure and help determine factors for businesses. Some individuals prefer to live in densely populated regions where foot traffic is highest whereas others prefer the quiet, low

densely populated regions. As a result, population density bar graphs were plotted for each cluster and their respective neighbourhoods as seen below.

a.



b.

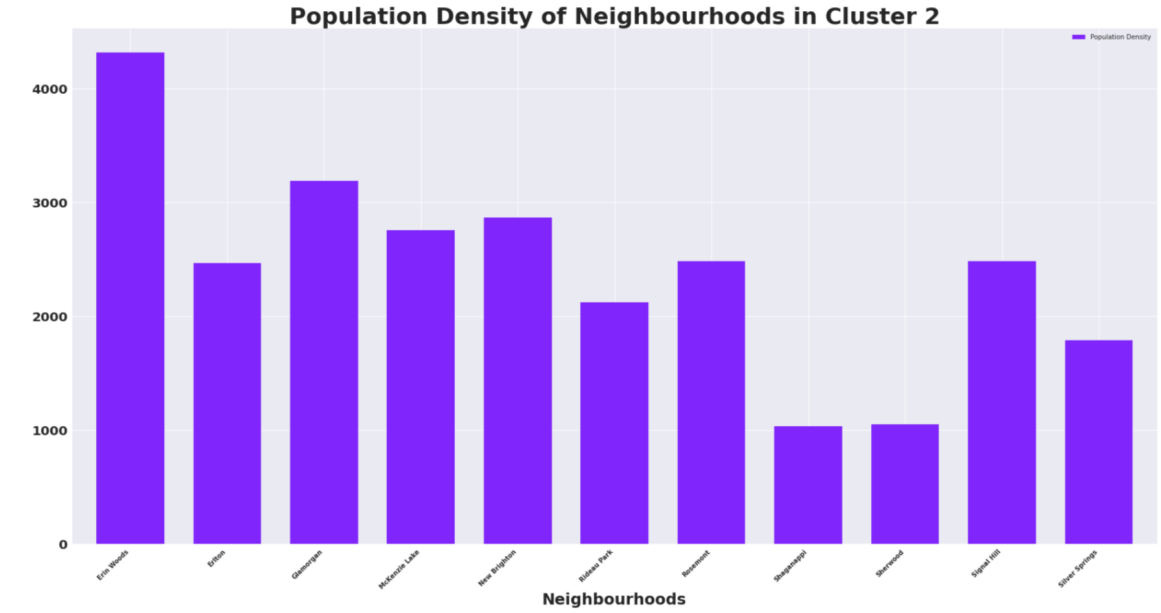


Figure 9: Population Density bar graphs of a) Cluster 1 and b) Cluster 2.

4.23 Crime Data Visualization through Choropleth Map

A Choropleth map is an extremely useful visualization method that is used to represent statistical data through various shading patterns on geographical areas ("Story Map Journal", 2021). Utilization of choropleth maps makes it easier for users to quickly

determine a general hypothesis of the studied data through statistical analysis. Therefore, in this study, a choropleth map was used to determine the geographical variation in crime in the city of Calgary.

Using the crime data gathered, a choropleth map was plotted using Folium. To do so, the following are required:

- A shape file of Calgary containing the longitude and latitude coordinates of Calgary (Figure 10). This geojson file was taken off of a github user account.
- A dataset containing the crime statistics of the neighbourhoods in Calgary. The data was extracted from the ‘Open Calgary' website stated earlier.

Merging the shape file and crime data, a choropleth map was plotted using the choropleth function in Folium (Figure 11).

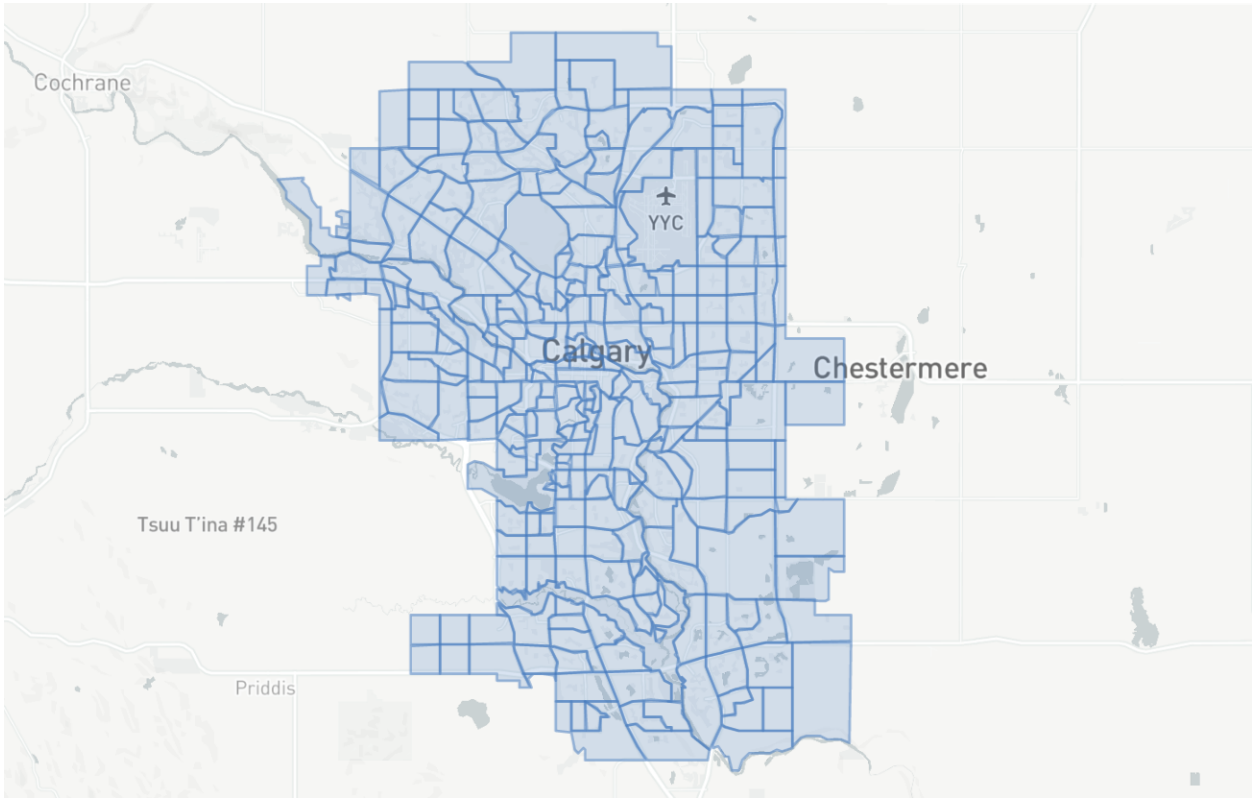


Figure 10: Geojson map of Calgary.

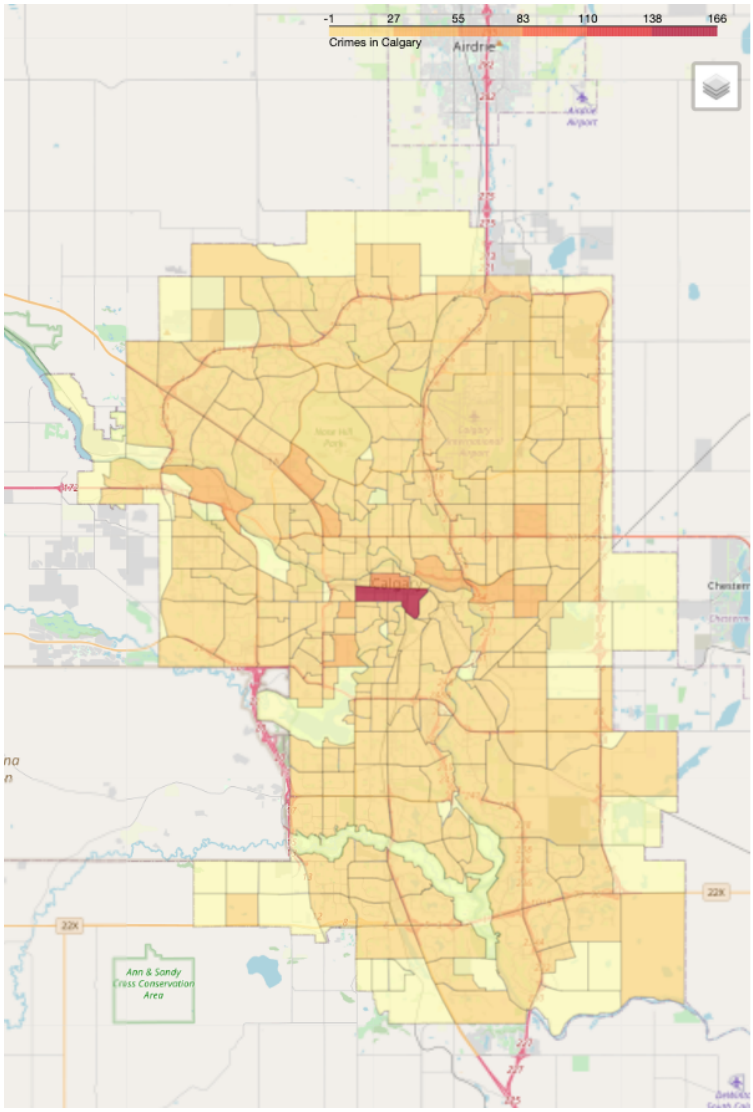


Figure 11: Choropleth crime map of Calgary.

5. Discussion

Through the use of unsupervised learning algorithms, although not always deemed reliable and predictable, venue clusters were created for the city of Calgary based on available data. Through iterative calculations Calgary was split into two clusters based on the venue's data. The low k value of 2 indicates large clusters with less granularity which is exactly depicted in the results section of this report. Despite this, the logistics of each neighbourhood, and even further, the sectors could be determined to get a general understanding of the city.

Cluster 1 is highly versatile with a wide variety of restaurants,cafes, recreational services, entertainment services and parks, whereas cluster 2 consists mostly of parks with entertainment services, lakes and recreation coming in second. The station data

suggests that cluster 1 is directed to individuals with flexibility in transportation services in contrast to the little number of stations found in cluster 2. Thus, cluster 2 would be appropriate for individuals that possess cars and have little transportation expectations. The population density data indicates a high traffic count in cluster 1 as majority of the neighbourhoods exceed the 3000 mark, a mark that is barely exceeded in cluster 2, and even reaching the 8000 mark in some. Therefore, cluster 1 would be appropriate for individuals looking to maintain an engaging lifestyle that is always on the go, filled with a variety of different places to see and discover. On the contrary, cluster 2 would seek individuals looking for a relatively quiet, active and routinely based lifestyle.

One of the major deciding factors that play a role in neighbourhood hunting is crime rate. The Calgary choropleth crime map indicates a fairly uniform crime rate throughout the city, with the highest occuring in the city centre and lowest on the outskirts of Calgary. The randomly dispersed clusters indicate zero correlation between the two variables, crime rate and clusters. Thus, little information can be taken to differentiate between the two clusters but could still be used to further differentiate the neighbourhoods within each cluster. This being said, the neighbourhood with the highest overall crime rate occurred in the Beltline, a central neighbourhood part of cluster 1.

5.1 Limitations and Suggestions for Future Studies

This study only focuses on certain criteria necessary to help decide in which neighbourhood to reside in. Aside from venues, population density, and crime data, other aspects may also need to be factored into the decision making process. These may include data on property value and census by age. However, such data are not readily available and attainable to the neighbourhood level required, making it inappropriate for the scope of this project. Furthermore, the “free” foursquare plan, Sandbox Tier account, used to extract venue information only allowed a certain number of API calls to be made, thus limiting our ability to fully capture precise data.

6. Conclusion

In this project, required data were identified, extracted and prepared to provide cluster analysis on neighbourhoods based on venue data, using an unsupervised learning algorithm. Based on similarities, two clusters were identified and further used in combination with crime and population density data, to provide recommendations to home buyers and/or migrants regarding the most appropriate location to settle in based on preferences. The findings indicate that the two clusters identified, along with the other two sets of data have little correlation and thus can not be used exclusively to differentiate clusters. However, a general assumption could be made and that is: Cluster 1 is appropriate for younger individuals seeking an active lifestyle open to new experiences and cluster 2 for families who prefer the quiet and conventional lifestyle.

7. References

Calgary Facts | Life in Calgary. (2021). Retrieved 19 August 2021, from https://www.lifeincalgary.ca/about-calgary/calgary-facts 2021)

Education, I. (2021). What is Unsupervised Learning?. Retrieved 19 August 2021, from https://www.ibm.com/cloud/learn/unsupervised-learning

Society, N. (2021). Population Density. Retrieved 19 August 2021, from https://www.nationalgeographic.org/encyclopedia/population-density/?utm_source=BibblioRCM_Row

Story Map Journal. (2021). Retrieved 19 August 2021, from https://www.arcgis.com/apps/MapJournal/index.html?appid=75eff041036d40cf8e70df99641004ca#:~:text=Choropleth%20maps%20are%20popular%20thematic,desired%20measurement%2C%20across%20a%20region.

Understanding K-means Clustering in Machine Learning. (2021). Retrieved 19 August 2021, from https://towardsdatascience.com/understanding-k-means-clustering-in-machine-learning-6a6e67336aa1

Welcome to GeoPy's documentation! — GeoPy 2.2.0 documentation. (2021). Retrieved 19 August 2021, from https://geopy.readthedocs.io/en/stable/