CAPSTONE PROJECT

THE BATTLE OF THE NEIGHBORHOODS REPORT

INTRODUCTION

Background

Minneapolis is the largest and most populous city in the state of Minnesota. as of 2019, there are an estimated 429,606 residents making in the 46th largest city in the United States and 8th largest in the Midwest. Minneapolis is also a neighbor to Saint Paul that make up what is called the "Twin Cities", with Minneapolis being larger. The Twin Cities metro area is also home to roughly 3.6 million people.

The Twin Cities was ranked by Business Insider as the 6th best city to live in the United States. They have many amenities like museums, sports stadiums, and metro transit with a "Midwestern feel". Local residents are also accustomed to the changing seasons with associated activities, such as ice fishing and cross-country skiing in the winter, music festivals and baseball games in the spring and summer, and the "Great Minnesota Get Together" at the Minnesota state fair towards the end of summer. Minneapolis and Saint Paul also have jobs available in science, manufacturing, agriculture, service sectors, and more. Some fortune 500 companies also have their headquarters based in the metro area. These would include companies like Target, Best Buy, 3M, Medtronic, Ecolab, Cargill, and United Health Group to just name a few.

This is why more people are moving to the Twin Cities area.

Problem

Since more people are moving to the Twin Cities area, it's important for them to decide which neighborhood to live in. A key determining factor on selecting the right neighborhood is the safety of the community. The safety of the community depends on the level of crime for the various neighborhoods of Minneapolis.

This project aims to identify the top three safest places to live in the city of Minneapolis. This will be determined based on the crime report statistics found in the Open Data Minneapolis data set found on Kaggle. This project will also reveal the most common venues of the three safest neighborhoods by clustering neighborhoods and using k-mean clustering.

This report will be targeted people who are looking to relocate to the city of Minneapolis. In order to conclude a neighborhood to look for an apartment, safety should be considered as a top priority when moving to a new place. You will not be able to enjoy your new apartment and the surrounding venues if you do not feel safe. The crime statistics in this report will provide an insight into this issue.

DATA

Data Acquisition

Based on the definition of our problem, the factors that influence the decision are:

- The total number of crimes committed in each neighborhood
- The most common venues of the selected neighborhood deemed the safest
- The data sources below will be needed in order to extract the required information:

https://www.kaggle.com/mrisdal/minneapolis-incidents-crime/data (This data set contains the crime and incident statistics of each neighborhood in Minneapolis.)

Foursquare API Data

Foursquare is a location data provider with information about venues and events within an area. This will provide us with information that includes different venues.

Libraries

The libraries in this project would consist of:

- Pandas: For creating and manipulating data frames.
- Folium: Python visualization library would be used to visualize the neighborhoods cluster distribution of using interactive leaflet map.
- Matplotlib: Python Plotting Module.
- XML: To separate data from presentation and XML stores data in plain text format.
- Scikit Learn: For importing k-means clustering.
- JSON: Library to handle JSON files.
- Geocoder: To retrieve Location Data.

Data Cleaning

The data preparation from the source data is cleaned for preprocessing and NaN values are dropped. The offense column contains the type of criminal offense that is reported. Since this information is relevant for our problem, the non-number values needed to be generalized to fit the data frame (figure 2.1)

	BeginDate	Time	Lat	Long	Neighborhood	Precinct	Offense	Description
13305	2010-01-01T00:00:00.000Z	00:00:00	44.896919	-93.280083	WINDOM	5	0	Other Theft
3978	2010-01-01T00:00:00.000Z	00:00:00	45.048050	-93.303879	SHINGLE CREEK	4	0	Other Theft
99151	2010-01-01T00:00:00.000Z	00:00:00	45.014889	-93.260170	HOLLAND	2	0	Other Theft
7229	2010-01-01T00:00:00.000Z	00:00:00	44.924450	-93.288727	EAST HARRIET	5	1	Theft From Motr Vehc
8473	2010-01-01T00:00:00.000Z	00:00:00	44.943840	-93.299957	ECCO	5	0	Other Theft

(*figure 2.1: First 5 rows of crime data after data cleaning)*

Once we were able to assign an integer value to the offense column, we were now able to see how many crimes were committed and what types of crimes were committed in a given neighborhood. The total number of crimes for each neighborhood are shown in figure 2.2.

(Figure 2.2: Total types of crimes committed in Minneapolis)

Other Theft	31521
Burglary Of Dwelling	19413
Theft From Motr Vehc	18867
Motor Vehicle Theft	9017
Shoplifting	4848
Robbery Per Agg	4451
Robbery Of Person	3925
Burglary Of Business	3774
Asslt W/dngrs Weapon	3325
Bike Theft	3135
Theft From Person	1940
Crim Sex Cond-rape	1475
Domestic Assault/Strangulation	1371
2nd Deg Domes Aslt	1228
Theft By Swindle	1071
Aslt-sgnfcnt Bdly Hm	999
Theft-motr Veh Parts	628
Robbery Of Business	617
Arson	607
Aslt-police/emerg P	526
Theft From Building	481
3rd Deg Domes Aslt	351
Murder (general)	133
Aslt-great Bodily Hm	118
Other Vehicle Theft	91
Gas Station Driv-off	69
Theft/coinop Device	52
Adulteration/poison	27
Theft By Computer	21
Scrapping-Recycling Theft	16
1st Deg Domes Asslt	15
Pocket-picking	13
On-line Theft	10
Disarm a Police Officer	5
Looting	1

Name: Description, dtype: int64

(Figure 2.3: Most crimes committed in Minneapolis neighborhoods)

DOWNTOWN WEST	13220
WHITTIER	4705
JORDAN	3619
LONGFELLOW	3373
NEAR - NORTH	3250
MARCY HOLMES	3192
WILLARD - HAY	2775
HAWTHORNE	2628
LOWRY HILL EAST	2559
POWDERHORN PARK	2479
LORING PARK	2468
FOLWELL	2329
VENTURA VILLAGE	2294
ELLIOT PARK	2241
SEWARD	2177
MIDTOWN PHILLIPS	2146
CENTRAL	2101
LYNDALE	2009
NORTH LOOP	1884
EAST PHILLIPS	1817

Name: Neighborhood, dtype: int64

(Figure 2.4: Least crimes committed in Minneapolis neighborhoods)

BRYANT	512
BRYN - MAWR	509
REGINA	453
ST. ANTHONY EAST	422
ST. ANTHONY WEST	420
FIELD	407
WEST CALHOUN	388
ARMATAGE	383
MORRIS PARK	378
SUMNER - GLENWOOD	367
MID - CITY INDUSTRIAL	365
MARSHALL TERRACE	348
BELTRAMI	327
COLUMBIA PARK	309
HALE	279
KENNY	277
KENWOOD	228
PAGE	178
CAMDEN INDUSTRIAL	135
HUMBOLDT INDUSTRIAL AREA	39

Name: Neighborhood, dtype: int64

The important thing to note in figure 2.3 is the difference in crime incidents from Downtown West and Whittier. The Downtown West neighborhood has almost three times more crimes committed than the second most of the neighborhood Whittier.

Since we are trying to find the safest place to live in Minneapolis, we have identified the neighborhoods we will avoid in our analysis in figure 2.3.

We can conduct a similar search to show the neighborhoods with the least amount of crime. Figure 2.4 shows the neighborhoods with the least reported crimes committed. Its important to note that industrial areas will not be considered as neighborhoods since there are no residents living in these areas. The safest neighborhood in Minneapolis starts with the Page neighborhood as shown in figure 2.3.

METHODOLOGY

Exploratory Data Analysis

Visualize the crimes rates for the Minneapolis neighborhoods that are the safest. Within the selected neighborhoods we will find the most common venues using the Foursquare API.

Modeling

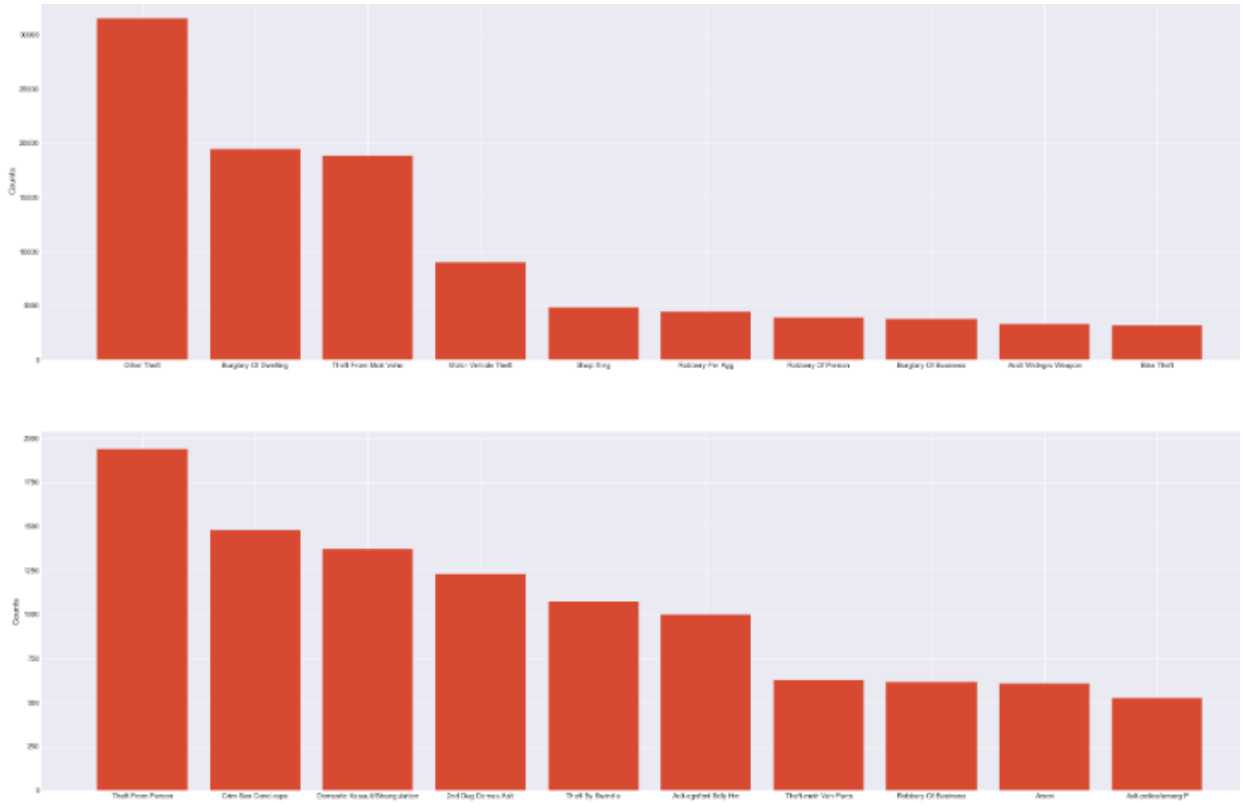
We will be using K-means clustering to find similar neighborhoods to our initial selection so that people who move to Minneapolis will have options when finding the perfect place for them to live.

First, we will get a statistical summary of our crime statistics of Minneapolis. When we call the describe function, we are returned with a mean, standard deviation, minimum, maximum, 1sr quartile (25%), 2nd quartile, (50%), 3rd quartile (75%) for each of the major crime categories. Figure 3.1.

	Lat	Long	Precinct	Offense
count	114141.000000	114141.000000	114141.000000	114141.000000
mean	44.969877	-93.269605	3.071517	5.076598
std	0.032792	0.027013	1.412271	6.974375
min	44.890629	-93.329086	1.000000	0.000000
25%	44.948372	-93.290138	2.000000	0.000000
50%	44.970032	-93.272751	3.000000	2.000000
75%	44.991798	-93.249451	4.000000	9.000000
max	45.051239	-93.197487	18.000000	35.000000

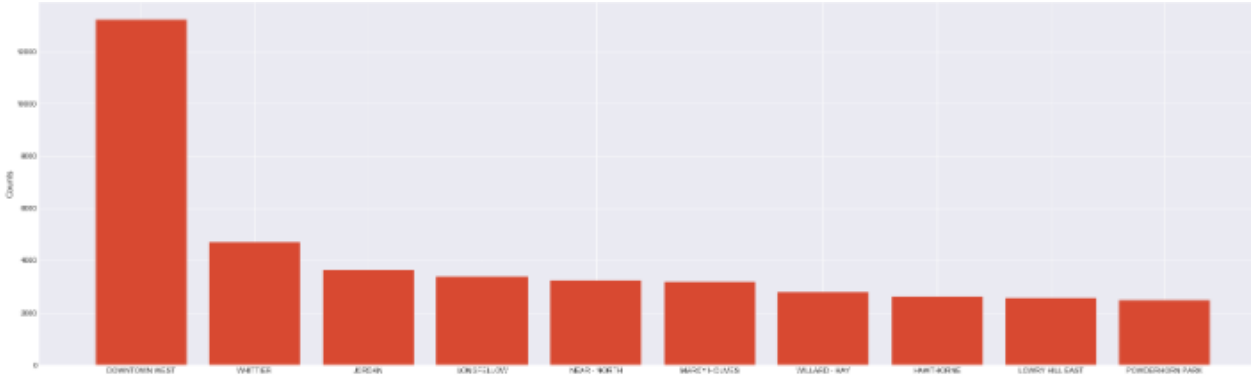
(Figure 3.1: Describe the crime statistics)

Next, we look to plot the crimes count based on the type of crime that is committed (figure 3.2). This figure shows that the majority of crimes that are committed in the city of Minneapolis are other types of crimes. The next most common are burglary of dwelling, theft from vehicle, shoplifting, robbery, etc.



(figure 3.2: Total types of crimes committed)

We look at how these crimes are attributed to the neighborhoods of Minneapolis to see where most of the crimes are taking place. These will be the areas we need to avoid in our analysis. (figure 3.3). As we can see in this graph, the neighborhoods with the most crimes committed are: Downtown West, Whittier, Jordan, Longfellow, etc.



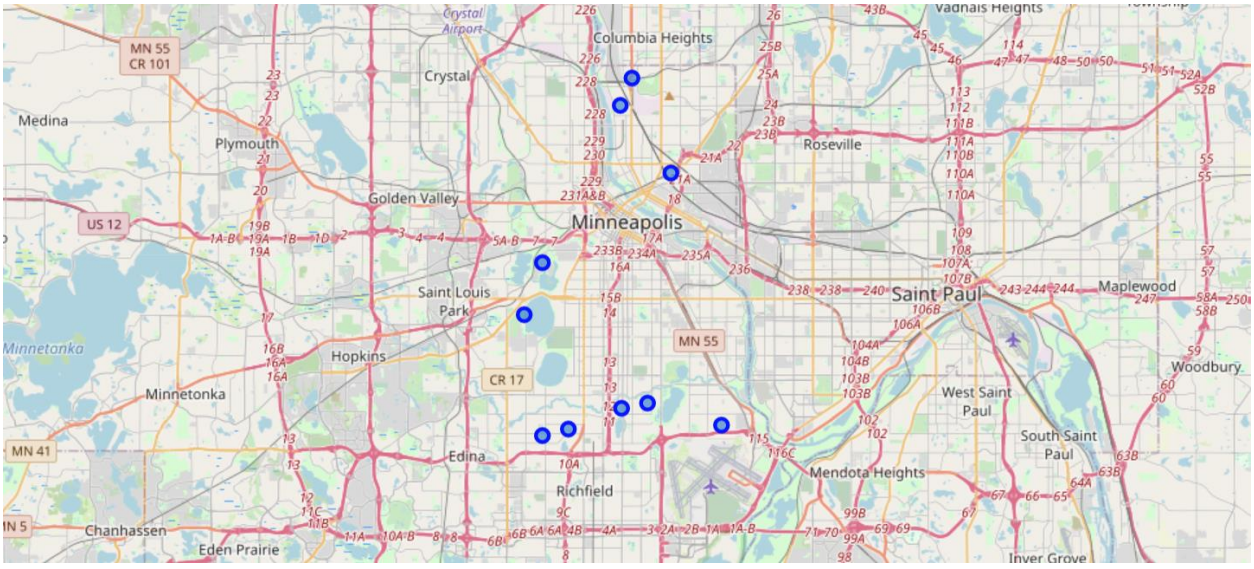
(Figure 3.3: Crimes per neighborhood)

Next we will determine the safest neighborhoods in Minneapolis and visualize them on a map. In order to do this, we will need to create a new data set that contains the safest neighborhoods. We will also need to generate the coordinates of the safety neighborhoods so they can be placed on a map of Minneapolis. Figure 3.4 shows the new data set with the filled in coordinates of the chosen safe neighborhoods determined in figure 2.4.

	Neighborhood	Latitude	Longitude
0	Page	44.907684	-93.268218
1	Kenwood	44.962087	-93.310523
2	Kenny	44.899857	-93.296505
3	Hale	44.909157	-93.254766
4	Columbia Park	45.030713	-93.263177
5	Beltrami	44.995352	-93.242513
6	Marshall Terrace	45.020870	-93.268819
7	Morris Park	44.900982	-93.215862
8	Armatage	44.897236	-93.310342
9	West Calhoun	44.942516	-93.319656

(Figure 3.4: New data set of safe neighborhoods with coordinates)

The new coordinates will allow us to pinpoint and visualize where the neighborhoods are located in Minneapolis. Figure 3.5 shows the map of Minneapolis with the safest neighborhoods.



(Figure 3.5: Map of Minneapolis with the safest neighborhoods)

Using the new data set for safest neighborhoods in Minneapolis, along with the coordinates, we can find all nearby venues within a 500-meter radius of each neighborhood. This will be executed using the foursquare API that will return a json file containing all the venues each neighborhood has. This can be converted into a pandas data frame for easy visualization. This data frame contains all the venues along with their coordinates and category. (figure 3.6) The amount of venues related to the associated neighborhood can be seen in figure 3.7.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Page	44.907684	-93.268218	Cottontail On The Trail	44.906506	-93.265617	Sculpture Garden
1	Page	44.907684	-93.268218	Harmon Killebrew Field	44.906264	-93.269500	Baseball Field
2	Page	44.907684	-93.268218	Pearl Park Soccer Fields	44.907773	-93.269200	Soccer Field
3	Page	44.907684	-93.268218	Diamond Lake	44.906503	-93.265644	Lake
4	Page	44.907684	-93.268218	Pearl Park	44.904679	-93.267275	Park

(Figure 3.6: Venue details of each neighborhood)

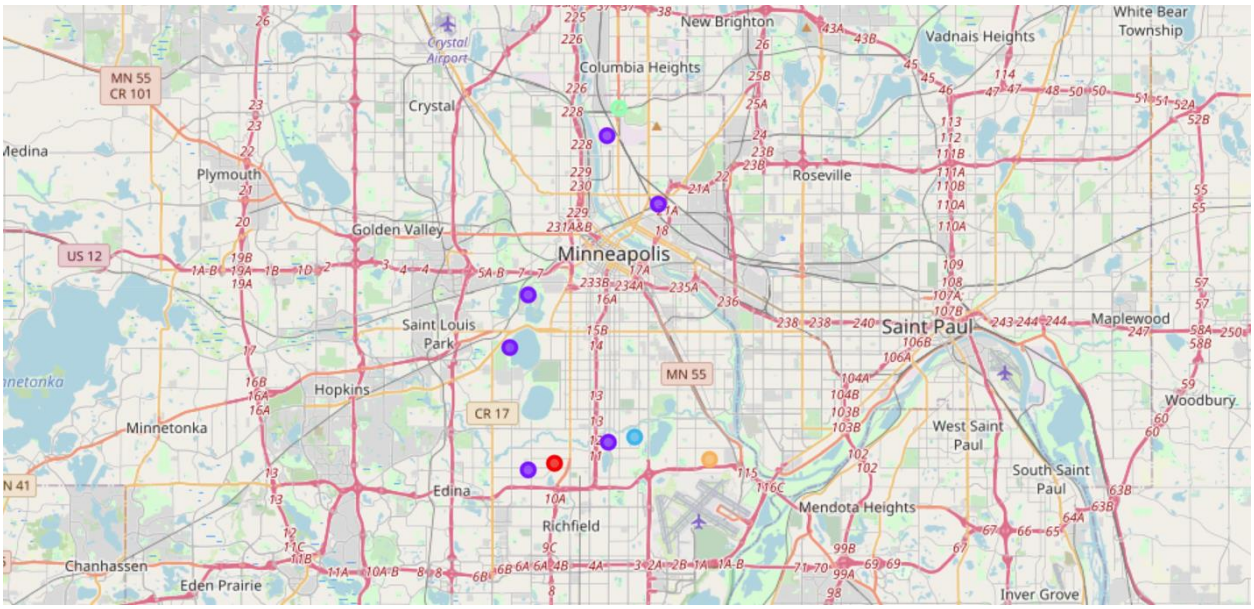
Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Armatage	8	8	8	8	8	8
Beltrami	10	10	10	10	10	10
Columbia Park	5	5	5	5	5	5
Hale	4	4	4	4	4	4
Kenny	4	4	4	4	4	4
Kenwood	11	11	11	11	11	11
Marshall Terrace	6	6	6	6	6	6
Morris Park	3	3	3	3	3	3
Page	10	10	10	10	10	10
West Calhoun	9	9	9	9	9	9

(Figure 3.7: Count of venues and Neighborhoods)

One hot encoding is done on the venue data. The venues data is then grouped by neighborhoods and the mean of the venues are calculated. The 10 most common venues are calculated for each of the neighborhoods. To help people find similar neighborhoods, we will be creating a cluster of the safe neighborhoods using K-means clustering. This is a form of unsupervised machine learning that clusters data based on predefined cluster size. We will use a cluster size of 5 for this project.

RESULTS

After running K-Means clustering, we can access each cluster that we created and reveal similarities and differences. We have to attach cluster labels and sort the new data frame in order to visualize is on the map. (figure 3.8)



(Figure 3.8: Map of clusters)

We first examine the first cluster:

	Neighborhood	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
2	Kenny	44.899857	-93.296505	0	Gym / Fitness Center	Construction & Landscaping	Park	Garden	Volleyball Court	Coffee Shop	Frozen Yogurt Shop	Food Truck	Food	Flower Shop

The first cluster contains only one neighborhood with venues such as gyms, parks, and gardens.

Second cluster:

	Neighborhood	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
0	Page	44.907684	-93.268218	1	Sculpture Garden	Park	Travel & Transport	Asian Restaurant	Athletics & Sports	Soccer Field	Café	Baseball Field	Lake
1	Kenwood	44.962087	-93.310523	1	American Restaurant	Bakery	Lawyer	Café	Park	Bookstore	Baseball Field	New American Restaurant	Skating Rink
5	Beltrami	44.995352	-93.242513	1	Gay Bar	Playground	Food Truck	Liquor Store	Convenience Store	Coffee Shop	Bus Station	Brewery	Café
6	Marshall Terrace	45.020870	-93.268819	1	Event Service	Brewery	Miscellaneous Shop	Park	Paper / Office Supplies Store	Music Venue	Volleyball Court	Clothing Store	Food Truck
8	Armatage	44.897236	-93.310342	1	Trail	Food	Flower Shop	Home Service	Skate Park	Pizza Place	Convenience Store	Park	Volleyball Court
9	West Calhoun	44.942516	-93.319656	1	Trail	Volleyball Court	Performing Arts Venue	Event Space	Moving Target	Burger Joint	Bus Station	Frozen Yogurt Shop	Bookstore

The second cluster is the largest cluster that contains 6 out of the 10 safest neighborhoods. We can see that most venues in this cluster contain restaurants, parks, and shopping places.

Third cluster:

	Neighborhood	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
3	Hale	44.909157	-93.254766	2	Furniture / Home Store	Thrift / Vintage Store	Playground	Breakfast Spot	Clothing Store	Frozen Yogurt Shop	Food Truck	Food	Flower Shop	Farmers Market

The third cluster contains only one neighborhood with venues such as stores, restaurants, and playgrounds.

Fourth cluster:

	Neighborhood	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
4	Columbia Park	45.030713	-93.263177	3	Gym	Arts & Crafts Store	Skate Park	Motorcycle Shop	Volleyball Court	Coffee Shop	Food Truck	Food	Flower Shop	Farmers Market

The fourth cluster contains only one neighborhood with venues such as gym, parks, and stores.

Fifth cluster:

	Neighborhood	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
7	Morris Park	44.900982	-93.215862	4	Clothing Store	Hotel	Playground	Volleyball Court	Frozen Yogurt Shop	Food Truck	Food	Flower Shop	Farmers Market	Event Space

The fifth and last cluster contains only one neighborhood with venues such as shops, playgrounds, and markets.

DISCUSSION

The goal of this project was to help people decide on the safest neighborhood to move to in the city of Minneapolis. The analysis shows that there are 10 safe neighborhoods with access to nearby venues and amenities. For example, if a person wants to live relatively close to a gym, we can see that cluster 1 and 4 has access to the nearest gym. If a person was looking to live closer to restaurants and shopping centers, any neighborhood in cluster 2 will suffice. Families that want to live closer to parks for their children should consider moving to neighborhoods in clusters 1,2, or 5 that are close to playgrounds, gardens, and parks.

CONCLUSION

This project gives a clear understanding for someone who is looking to move to the Minneapolis area. The information presented in this project will give them the information they need to decide on the best neighborhood to move to based on their interest. It's important to determine the desired neighborhood to live in and have some options in case there are no places to move to in certain neighborhoods. The future work of this project will consist of housing and renting prices as well as other community factors, such as school quality, cost of living, employment opportunities, etc. Attributing more factors and variables on top of the crime data will develop a more in-depth analysis on the desired place to live in the city of Minneapolis.