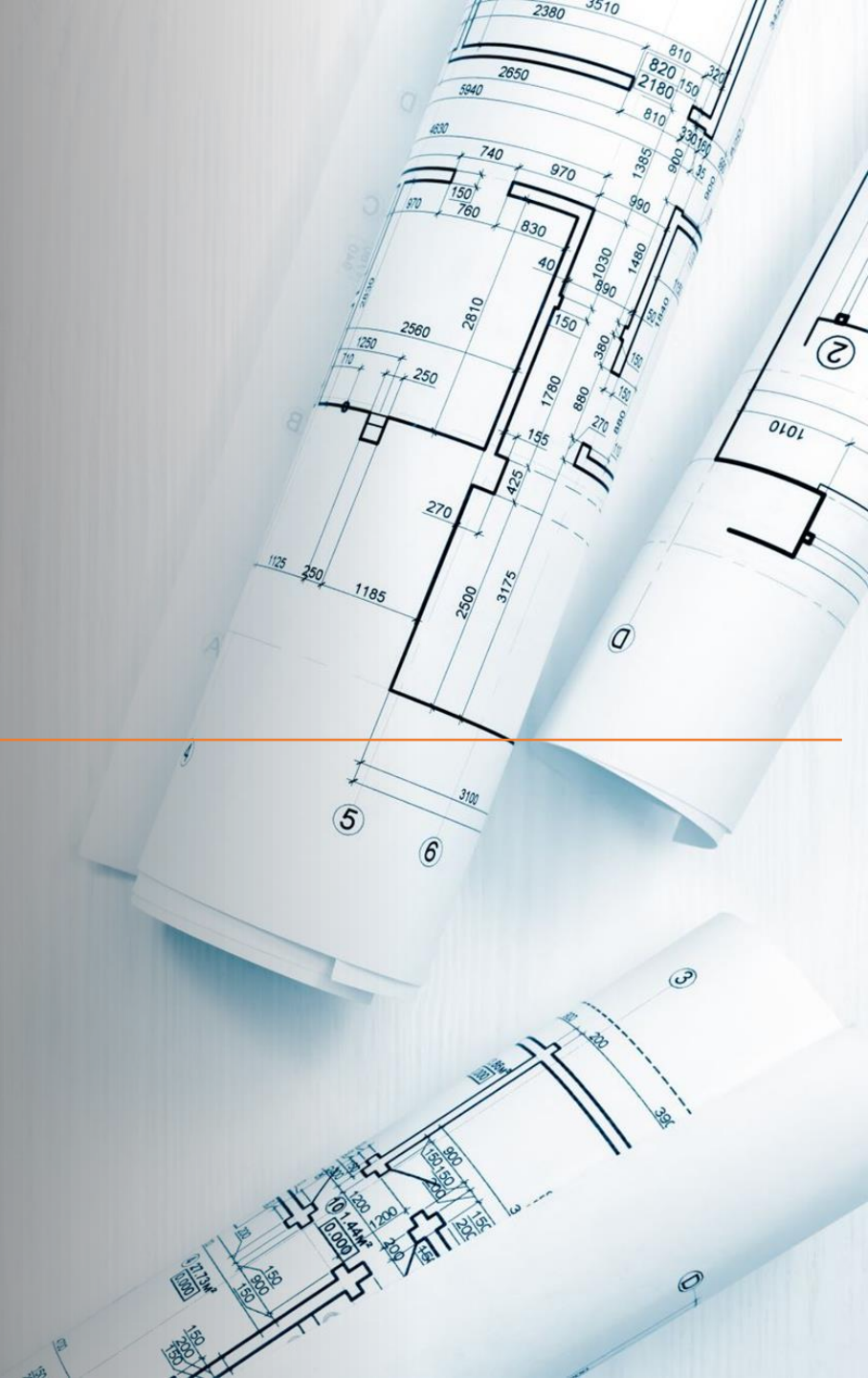
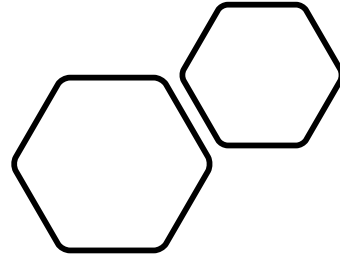


Capstone Project

Battle of the neighborhoods



Introduction



Background

- Minneapolis is most populated city in MN
- 8th largest in Midwest
- 6th best city to live in the US

Problem

- More people are moving to Minneapolis
- Important for new residents to select the safest neighborhood to move into

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.)

Data – Foursquare API

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.



Data - Libraries



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.

	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

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.

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.

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

Data Cleaning

- The important thing to note 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

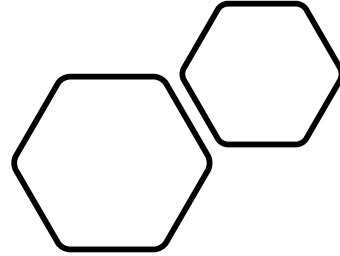
JORDAN	3619
LONGFELLOW	3373
NEAR - NORTH	3250
MARCY HOLMES	3192
WILLARD - HAY	2775
HAWTHORNE	2628
LOWRY HILL EAST	2559
POWDERHORN PARK	2479
LORING PARK	2468
FOLWELL	2329
VENTURA VILLAGE	2294
ELLIOT PARK	2241
SEWARD	2177
MIDTOWN PHILLIPS	2146
CENTRAL	2101
LYNDALE	2009
NORTH LOOP	1884
EAST PHILLIPS	1817

Data Cleaning

- We can conduct a similar search to show the neighborhoods with the least amount of crime. 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.

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	

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.

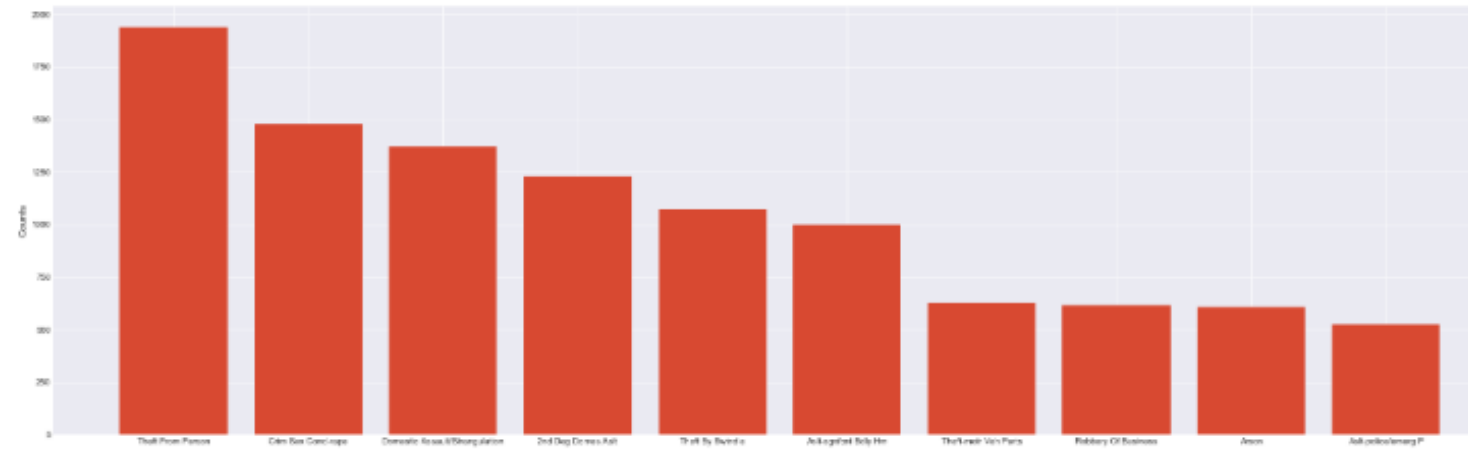
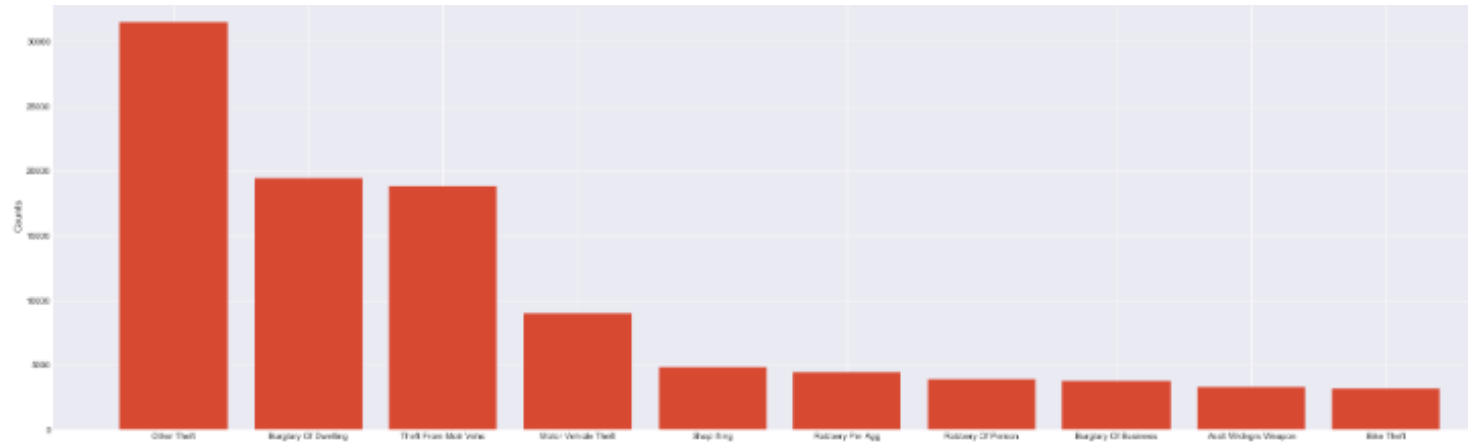
Methodology

Modeling

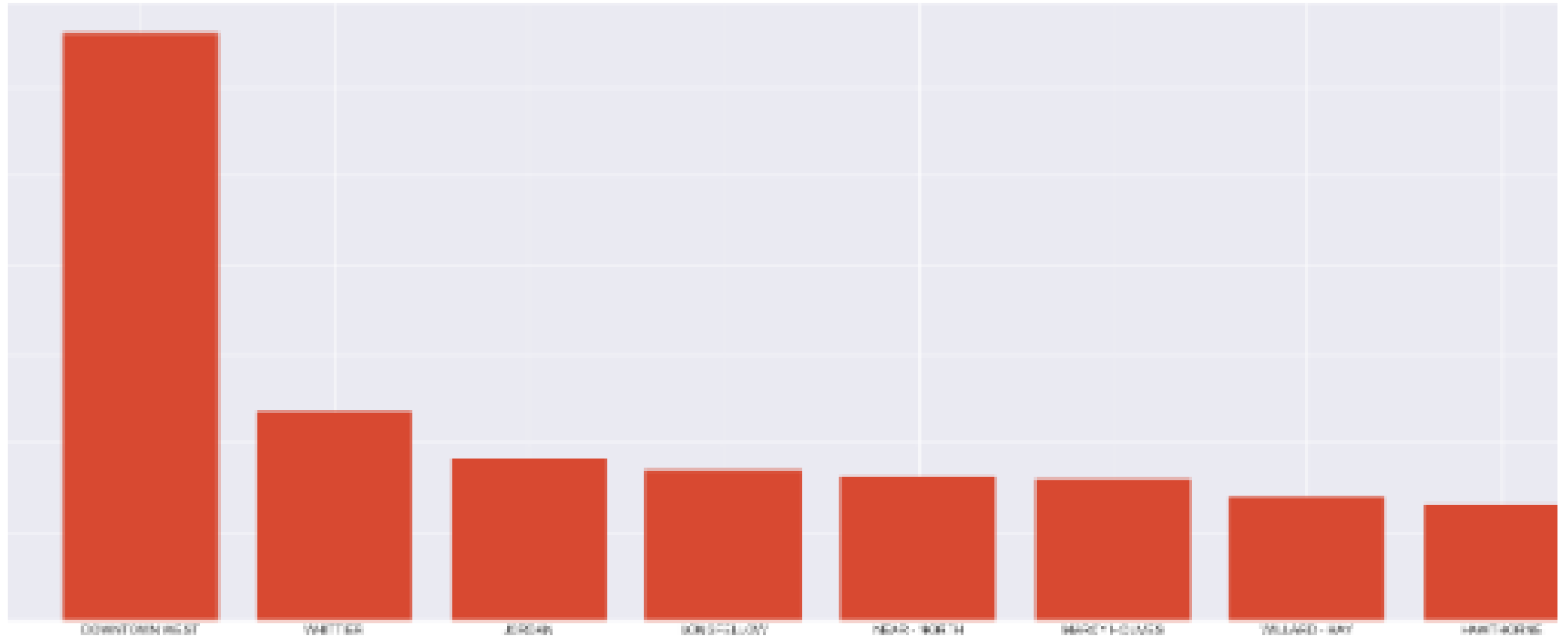


We will be using K-means clustering to find similar neighborhoods to our initial selection so that people who move to Minneapolis will have options when finding the perfect place for them to live.

Crimes Count



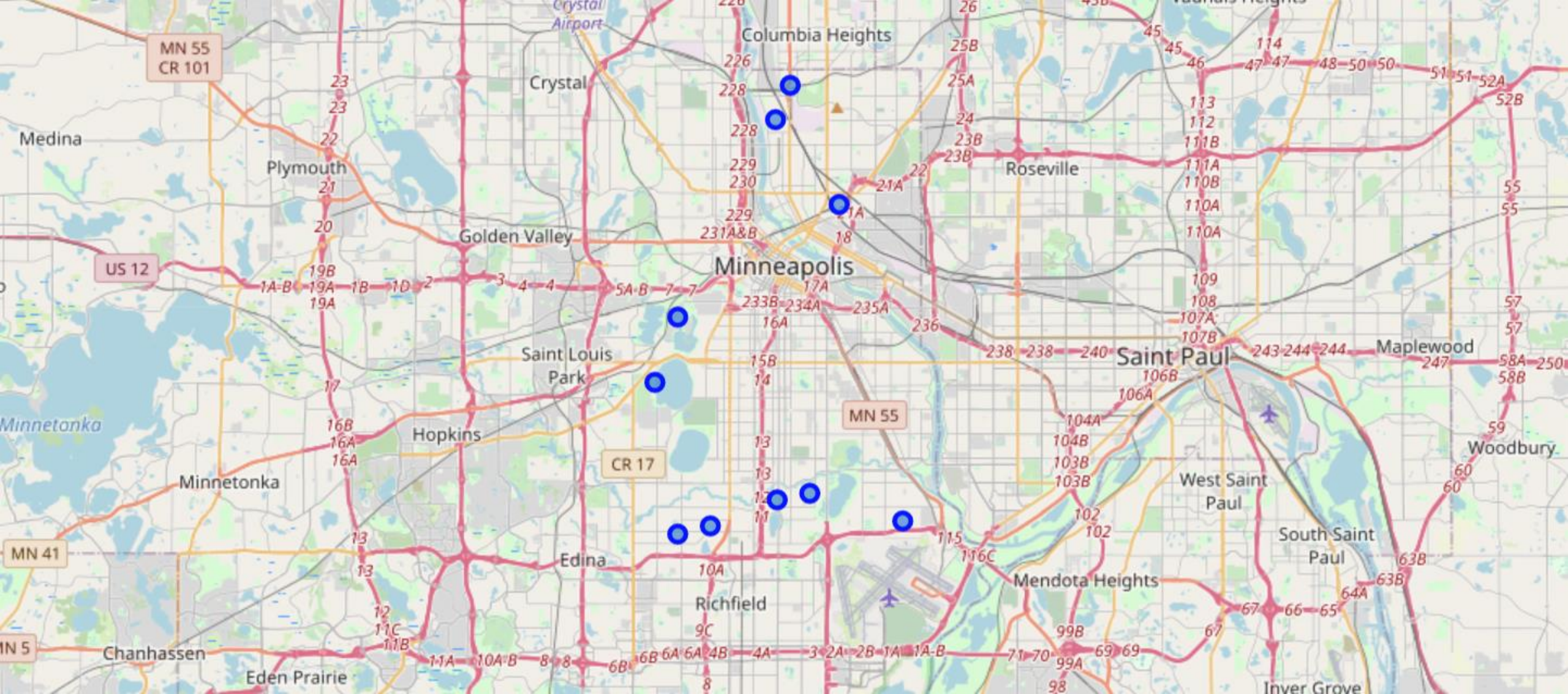
Crimes per neighborhood



Determine Coordinates for safest neighborhoods

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

Neighborhood	Latitude	Longitude
Page	44.907684	-93.268218
Kenwood	44.962087	-93.310523
Kenny	44.899857	-93.296505
Hale	44.909157	-93.254766
Columbia Park	45.030713	-93.263177
Beltrami	44.995352	-93.242513
Marshall Terrace	45.020870	-93.268819
Morris Park	44.900982	-93.215862
Armatage	44.897236	-93.310342
West Calhoun	44.942516	-93.319656



Map of safe neighborhoods

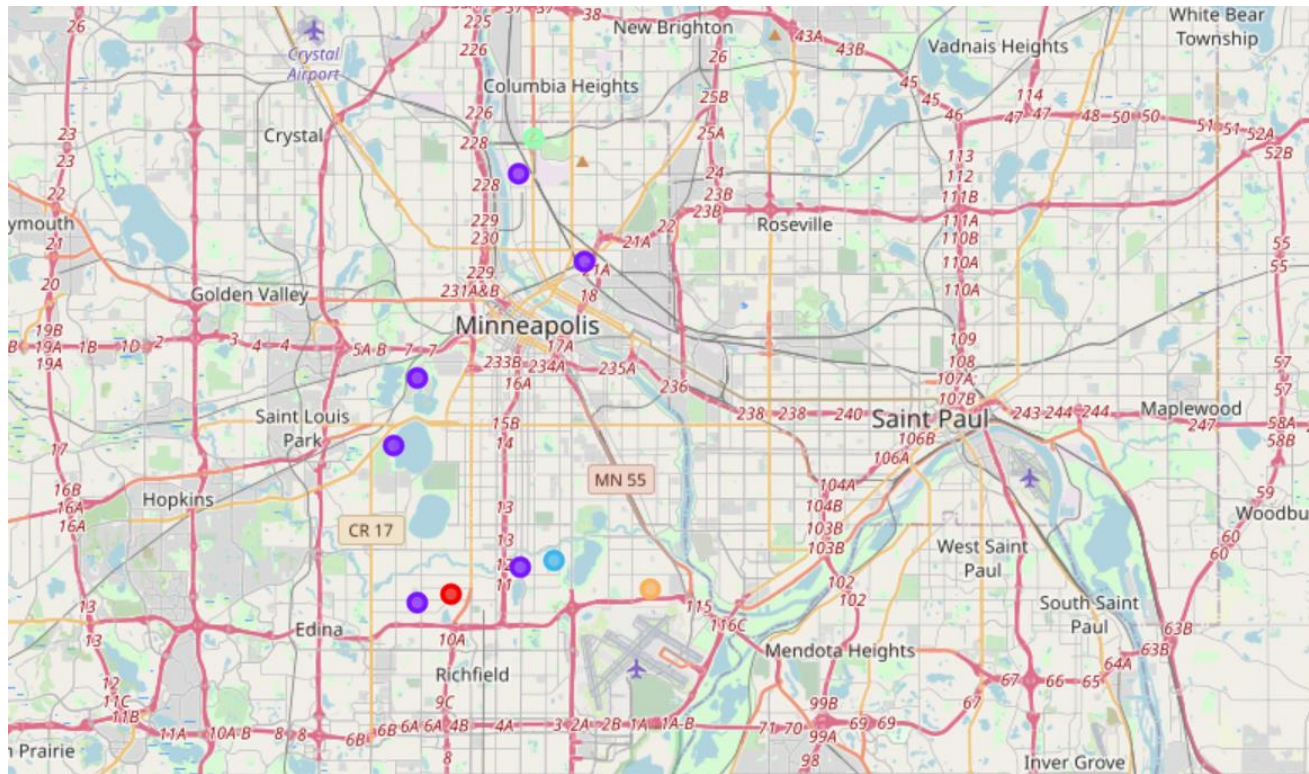
	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Page	44.907684	-93.268218	Cottontail On The Trail	44.906506	-93.265617	Sculpture Garden
1	Page	44.907684	-93.268218	Harmon Killebrew Field	44.906264	-93.269500	Baseball Field
2	Page	44.907684	-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

Use foursquare API to get coordinates of nearby venues

One hot encoding

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.

K-means clustering



- 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

First cluster

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

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

Second cluster

- 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.

	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

Third cluster

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

	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

Fourth cluster

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

	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

Fifth cluster

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

	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

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.