# Capstone Project - The Battle of Neighborhoods

# Report

# 1. Introduction

## 1.1 Background

The people will relocate to different places as and when they change jobs or their business. The average American relocate around 11 times in their lifetime. When people need to relocate, they search for lot of information about locations in the city and different types of venues in nearby to the location.

As relocation is a tidy, expensive and most difficult task, people will not take a chance by relocate first and check for safety and venues nearby. People will look for the information whether the place is good in terms of safety and venues nearby.

## 1.2 Problem

The crime statistics dataset of Denver on Kaggle has crimes in each Neighborhood of Denver from 2016 to 2019. We will consider the year 2019 which is latest for this analysis.

This project aims to identify the safest cluster in Denver based on total crimes in the neighborhoods, nearby venues including 10 most common venues in each neighborhood. And finally Cluster the venues & neighborhoods using k-mean clustering.

## 1.3 Interest

This will help the people who are interested in relocating to Denver in identifying the safest Neighborhood and explore nearest venues for each of the Neighborhoods.

## 2. Data Acquisition and Cleaning

### 2.1 Data Acquisition

The first data source of the project uses Denver crime data from kaggle website that shows the crime data in each Neighborhood in London.

The dataset contains the following columns:

- OFFENSE_TYPE_ID: which gives which type of offense it is.
- NEIGHBORHOOD_ID: This will give the Neighborhood in which offense reported
- IS_CRIME : This showing the offense is a crime
- Is_TRAFFIC: which shows if the offense is traffic offense
- REPORTED_DATE: This include crime reported date & time

The second data source of data Latitude & Longitude  is extracted using geocoder for each Neighborhood.

 The third data source is the list of nearby venues for Neighborhoods of the Denver and Venues information using Foursquare.

Leveraged Foursquare to extract the following data.

- Venue Name
- Venue Latitude
- Venue Longitude
- Venue category

## 2.2 Data Cleaning

From the Denver crime data, the crimes during the most recent year 2019 are only selected.

| | OFFENSE_CODE | OFFENSE_TYPE_ID | OFFENSE_CATEGORY_ID | NEIGHBORHOOD_ID | IS_CRIME | year |
|---|---|---|---|---|---|---|
| 1568 | 1208 | robbery-residence | robbery | east-colfax | 1 | 2019 |
| 23326 | 5499 | traf-other | all-other-crimes | jefferson-park | 1 | 2019 |
| 23327 | 2404 | theft-of-motor-vehicle | auto-theft | west-highland | 1 | 2019 |
| 23328 | 2303 | theft-shoplift | larceny | city-park | 1 | 2019 |
| 23329 | 2999 | criminal-mischief-mtr-veh | public-disorder | montbello | 1 | 2019 |

The major categories of crime are pivoted to get the total crimes per Neighborhood as per category.

| | Neighborhood | aggravated-assault | all-other-crimes | Earson | auto-theft | burglary | drug-alcohol | larceny | murder | other-crimes-against-persons | public-disorder | robbery | sexual-assault | theft-from-motor-vehicle | white-collar-crime | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | athmar-park | 33 | 207 | 1 | 63 | 28 | 98 | 58 | 1 | 48 | 81 | 16 | 9 | 86 | 21 | 750 |
| 1 | auraria | 15 | 271 | 1 | 19 | 24 | 28 | 97 | 0 | 52 | 100 | 11 | 8 | 35 | 11 | 672 |
| 2 | baker | 28 | 227 | 6 | 72 | 49 | 120 | 262 | 0 | 55 | 133 | 21 | 7 | 123 | 20 | 1123 |
| 3 | barnum | 35 | 212 | 2 | 40 | 27 | 47 | 31 | 1 | 46 | 91 | 17 | 10 | 51 | 11 | 621 |
| 4 | barnum-west | 13 | 100 | 3 | 41 | 14 | 42 | 11 | 1 | 16 | 60 | 2 | 4 | 38 | 4 | 349 |

The second data of Neighborhood Latitude and Longitude data extracted using Geocoder.

| neighborhood | aggravated-assault | all-other-crimes | Earson | auto-theft | burglary | drug-alcohol | larceny | murder | other-crimes-against-persons | public-disorder | robbery | sexual-assault | theft-from-motor-vehicle | white-collar-crime | Total | latitude |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| wellshire | 1 | 4 | 0 | 6 | 12 | 3 | 5 | 0 | 4 | 13 | 0 | 0 | 18 | 2 | 68 | 39.658484 |
| indian-creek | 1 | 6 | 0 | 9 | 4 | 1 | 11 | 1 | 5 | 13 | 3 | 0 | 26 | 4 | 84 | 39.684761 |
| country-club | 3 | 22 | 0 | 12 | 34 | 2 | 17 | 0 | 3 | 11 | 3 | 0 | 33 | 1 | 141 | 39.720100 |
| skyland | 6 | 69 | 1 | 13 | 11 | 8 | 16 | 2 | 11 | 22 | 1 | 3 | 26 | 5 | 194 | 39.755713 |
| rosedale | 4 | 30 | 0 | 17 | 19 | 21 | 26 | 0 | 7 | 23 | 1 | 4 | 49 | 5 | 206 | 39.739236 |

The third data of nearby venues of Neighborhood using Foursquare API.

| | Neighborhood | Neighborhood Latitude | Neighborhood Longitude | Venue | Venue Latitude | Venue Longitude | Venue Category | Cluster Label | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Mo Comm Ven |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | wellshire | 39.658484 | -104.941353 | Wellshire Golf Course | 39.658276 | -104.941329 | Golf Course | 2 | Golf Course | Athletics & Sports | Playground | Park | Yc Stu |
| 1 | wellshire | 39.658484 | -104.941353 | Mamie D. Eisenhower Park | 39.658946 | -104.937348 | Park | 2 | Golf Course | Athletics & Sports | Playground | Park | Yc Stu |
| 2 | wellshire | 39.658484 | -104.941353 | Wellshire Inn | 39.658247 | -104.941277 | Golf Course | 2 | Golf Course | Athletics & Sports | Playground | Park | Yc Stu |
| 3 | wellshire | 39.658484 | -104.941353 | Eisenhower Rec Center | 39.659563 | -104.937773 | Athletics & Sports | 2 | Golf Course | Athletics & Sports | Playground | Park | Yc Stu |
| 4 | wellshire | 39.658484 | -104.941353 | Eisenhower Park | 39.659608 | -104.936572 | Playground | 2 | Golf Course | Athletics & Sports | Playground | Park | Yc Stu |

## 3. Methodology

### 3.1 Exploratory Data Analysis

### 3.1.1 Statistical summary of crimes

The describe function in python is used to get statistics of the Denver crime data, this returns the mean, standard deviation, minimum, 1st quartile (25%), 2nd quartile(50%), and 3rd quartile (75%) for each of the major categories of crime (See fig 3.1.1)

| | aggravated-assault | all-other-crimes | Earson | auto-theft | burglary | drug-alcohol | larceny | murder | other-crimes-against-persons | public-disorder | robbery | sexual-assault | the |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| count | 78.000000 | 78.000000 | 78.00000 | 78.000000 | 78.000000 | 78.000000 | 78.000000 | 78.000000 | 78.000000 | 78.000000 | 78.000000 | 78.000000 | 78 |
| mean | 32.294872 | 187.717949 | 1.25641 | 68.910256 | 51.589744 | 64.576923 | 125.692308 | 0.897436 | 59.705128 | 108.333333 | 15.294872 | 10.974359 | 104 |
| std | 36.949284 | 205.284212 | 1.59085 | 52.556756 | 33.051096 | 83.519267 | 165.732715 | 1.123419 | 69.044332 | 96.657068 | 15.843479 | 10.813621 | 71 |
| min | 1.000000 | 4.000000 | 0.00000 | 6.000000 | 4.000000 | 1.000000 | 5.000000 | 0.000000 | 2.000000 | 11.000000 | 0.000000 | 0.000000 | 18 |
| 25% | 12.250000 | 59.000000 | 0.00000 | 38.250000 | 32.250000 | 16.250000 | 49.250000 | 0.000000 | 22.000000 | 52.000000 | 6.000000 | 5.000000 | 65 |
| 50% | 21.000000 | 102.500000 | 1.00000 | 54.500000 | 44.000000 | 37.000000 | 70.500000 | 1.000000 | 40.000000 | 79.500000 | 10.500000 | 7.000000 | 87 |
| 75% | 37.250000 | 228.500000 | 2.00000 | 82.000000 | 60.750000 | 79.250000 | 131.500000 | 1.000000 | 57.500000 | 120.000000 | 19.750000 | 12.500000 | 124 |
| max | 243.000000 | 1013.000000 | 7.00000 | 259.000000 | 202.000000 | 475.000000 | 959.000000 | 5.000000 | 419.000000 | 626.000000 | 87.000000 | 65.000000 | 483 |

Fig 3.1.1 Statistical description of the Denver crimes

The count for each of the major categories of crime returns the value 78 which is the number of Denver neighborhoods. 'larceny' is the highest reported crime during the year 2019 followed by 'public-disorder' and 'theft-from-motor-vehicle' offenses. The lowest recorded crimes are 'murder', 'Earson' & 'sexual-assault' offenses

### 3.1.2 Neighborhoods with the highest crime rates

Comparing five neighborhoods with the highest crime rate during the year 2019 it is evident that 'five-points' has the highest crimes recorded followed by stapleton, cbd, capitalhill & montbello. (see fig 3.1.2)
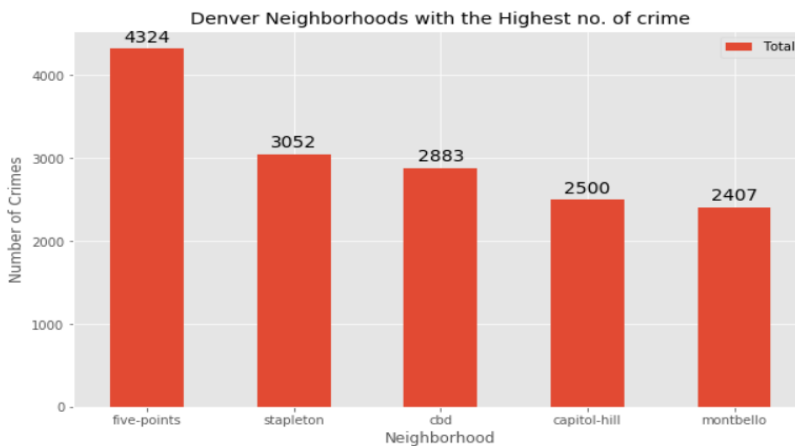
Fig 3.1.2 Neighborhoods with the highest crime rates

### 3.1.3 Neighborhoods with the lowest crime rates

Comparing five neighborhoods with the lowest crime rate during the year 2019, 'Wellshire' has the lowest recorded crimes followed by indian-creek, country-club, skyland and rosedale... (see fig 3.1.3)
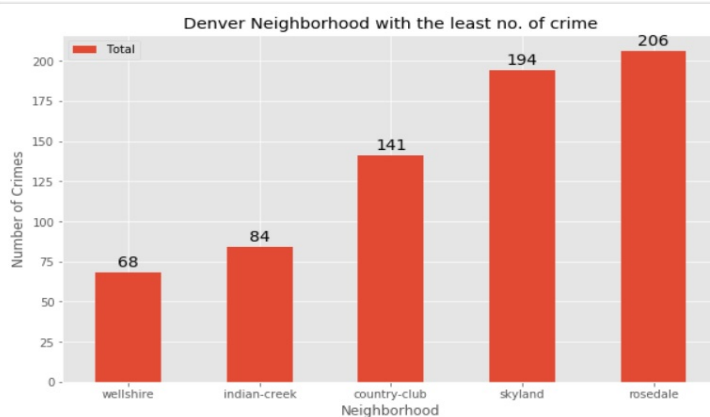
Fig 3.1.3 Neighborhoods with lowest crime rates

Both 'Wellshire' and indian-creeks' have the significantly lower crime rate compare to other neighborhoods.

**3.1.4 The lowest crime neighborhood visualized on a map using folium on python (see fig 3.1.4)**
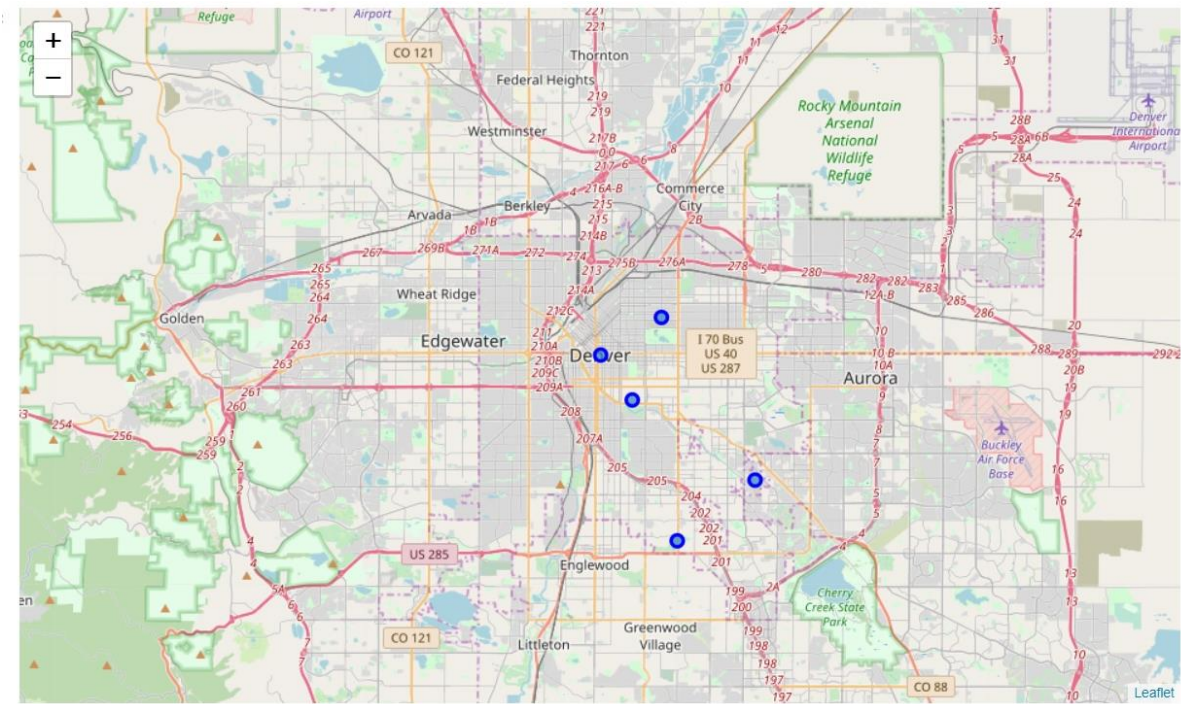


Fig 3.1.4 Lowest crime neighborhoods in Denver

**3.2 Modelling**

Using the final dataset containing the neighborhoods in Denver along with latitude and longitude, we can find all the venues within a 500 meter radius of each neighborhood by connecting to the Foursquare API. This returns a json file containing all the venues in each neighborhood which is converted to a pandas data frame. This data frame contains all the venues along with their coordinates and category (see fig 3.2.1)

| | Neighborhood | Neighborhood Latitude | Neighborhood Longitude | Venue | Venue Latitude | Venue Longitude | Venue Category | Cluster Label | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Mo Comm Ven |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | wellshire | 39.658484 | -104.941353 | Wellshire Golf Course | 39.658276 | -104.941329 | Golf Course | 2 | Golf Course | Athletics & Sports | Playground | Park | Yc Stu |
| 1 | wellshire | 39.658484 | -104.941353 | Mamie D. Eisenhower Park | 39.658946 | -104.937348 | Park | 2 | Golf Course | Athletics & Sports | Playground | Park | Yc Stu |
| 2 | wellshire | 39.658484 | -104.941353 | Wellshire Inn | 39.658247 | -104.941277 | Golf Course | 2 | Golf Course | Athletics & Sports | Playground | Park | Yc Stu |
| 3 | wellshire | 39.658484 | -104.941353 | Eisenhower Rec Center | 39.659563 | -104.937773 | Athletics & Sports | 2 | Golf Course | Athletics & Sports | Playground | Park | Yc Stu |
| 4 | wellshire | 39.658484 | -104.941353 | Eisenhower Park | 39.659608 | -104.936572 | Playground | 2 | Golf Course | Athletics & Sports | Playground | Park | Yc Stu |

Fig 3.2.1 Venue details of each Neighborhood

One hot encoding is done on venues data. (One hot encoding is a process by which categorical variables are converted into a form that could be provided in ML algorithms to do a better job in prediction). The Venues data is then grouped by Neighborhood and the mean of venues are calculated, finally the 10 common venues are calculated for each of the neighborhoods.

To help people find simlar neighborhoods in the safest borough we will be clustering similar neighborhoods using K-means clustering which is a form of unsupervised machine learning algorithm that clusters data based on predefined cluster size. We will use cluster size of 4 for this project. The reason to conduct a K-means clustering is to cluster neighborhoods with similar venues together so that people can shortlist the area of their interests based on the venues/amenities around each neighborhood.

## 4. Results

After running the K-means clustering we can access each cluster created to see which neighborhoods were assigned to each 4 clusters. Looking into the neighborhoods in the first cluster (see fig 4.1)

| | Neighborhood | Neighborhood Latitude | Neighborhood Longitude | Venue | Venue Latitude | Venue Longitude | Venue Category | Cluster Label | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 5 | country-club | 39.7201 | -104.9667 | Denver Country Club | 39.717805 | -104.966252 | Golf Course | 0 | Health & Beauty Service | Golf Course | Yoga Studio | Carpet Store | Gym | |
| 6 | country-club | 39.7201 | -104.9667 | Namaste Health Care | 39.718477 | -104.969934 | Health & Beauty Service | 0 | Health & Beauty Service | Golf Course | Yoga Studio | Carpet Store | Gym | |

Fig 4.1 Cluster 1

The first Neighborhood cluster has two Venues which are Golf Course and Health & Beauty Service

Looking into the venues in the second cluster (See Fig 4.2)

| | Neighborhood | Neighborhood Latitude | Neighborhood Longitude | Venue | Venue Latitude | Venue Longitude | Venue Category | Cluster Label | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th M Comm Ver |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 11 | rosedale | 39.739236 | -104.984862 | Sassafras American Eatery | 39.739949 | -104.982756 | Breakfast Spot | 1 | Yoga Studio | Noodle House | Marijuana Dispensary | Art Museum | Bak |
| 12 | rosedale | 39.739236 | -104.984862 | Phở-natic | 39.740081 | -104.984111 | Noodle House | 1 | Yoga Studio | Noodle House | Marijuana Dispensary | Art Museum | Bak |
| 13 | rosedale | 39.739236 | -104.984862 | City, O' City | 39.736724 | -104.984669 | Vegetarian / Vegan Restaurant | 1 | Yoga Studio | Noodle House | Marijuana Dispensary | Art Museum | Bak |
| 14 | rosedale | 39.739236 | -104.984862 | Capitol Hill Books | 39.739979 | -104.983472 | Bookstore | 1 | Yoga Studio | Noodle House | Marijuana Dispensary | Art Museum | Bak |
| 15 | rosedale | 39.739236 | -104.984862 | Shish Kabob Grill | 39.740246 | -104.983633 | Middle Eastern Restaurant | 1 | Yoga Studio | Noodle House | Marijuana Dispensary | Art Museum | Bak |
| 16 | rosedale | 39.739236 | -104.984862 | Kindness Yoga | 39.736721 | -104.984407 | Yoga Studio | 1 | Yoga Studio | Noodle House | Marijuana Dispensary | Art Museum | Bak |
| 17 | rosedale | 39.739236 | -104.984862 | Good Chemistry - Denver Dispensary | 39.739907 | -104.982668 | Marijuana Dispensary | 1 | Yoga Studio | Noodle House | Marijuana Dispensary | Art Museum | Bak |
| 18 | rosedale | 39.739236 | -104.984862 | CorePower Yoga | 39.737070 | -104.983057 | Yoga Studio | 1 | Yoga Studio | Noodle House | Marijuana Dispensary | Art Museum | Bak |
| 19 | rosedale | 39.739236 | -104.984862 | City Grille | 39.740165 | -104.982861 | Burger Joint | 1 | Yoga Studio | Noodle House | Marijuana Dispensary | Art Museum | Bak |
| 20 | rosedale | 39.739236 | -104.984862 | Civic Center Park | 39.739370 | -104.988776 | Park | 1 | Yoga Studio | Noodle House | Marijuana Dispensary | Art Museum | Bak |

Fig 4.2 Cluster 2

The second Neighborhood cluster is the biggest cluster with consists of 40 venues nearby. the venues includes Restaurants, Yoga centers, nightclubs, Parks....etc

| | Neighborhood | Neighborhood Latitude | Neighborhood Longitude | Venue | Venue Latitude | Venue Longitude | Venue Category | Cluster Label | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th M Comm Ven |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | wellshire | 39.658484 | -104.941353 | Wellshire Golf Course | 39.658276 | -104.941329 | Golf Course | 2 | Golf Course | Athletics & Sports | Playground | Park | Yc Stu |
| 1 | wellshire | 39.658484 | -104.941353 | Mamie D. Eisenhower Park | 39.658946 | -104.937348 | Park | 2 | Golf Course | Athletics & Sports | Playground | Park | Yc Stu |
| 2 | wellshire | 39.658484 | -104.941353 | Wellshire Inn | 39.658247 | -104.941277 | Golf Course | 2 | Golf Course | Athletics & Sports | Playground | Park | Yc Stu |
| 3 | wellshire | 39.658484 | -104.941353 | Eisenhower Rec Center | 39.659563 | -104.937773 | Athletics & Sports | 2 | Golf Course | Athletics & Sports | Playground | Park | Yc Stu |
| 4 | wellshire | 39.658484 | -104.941353 | Eisenhower Park | 39.659608 | -104.936572 | Playground | 2 | Golf Course | Athletics & Sports | Playground | Park | Yc Stu |

Fig 4.3 Cluster 3

The third Neighborhood cluster has five venues nearby which include Golf Course, Park, Rec Center & Play Ground.

| | Neighborhood | Neighborhood Latitude | Neighborhood Longitude | Venue | Venue Latitude | Venue Longitude | Venue Category | Cluster Label | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 7 | skyland | 39.755713 | -104.950224 | Street Food Warriors | 39.754849 | -104.947826 | Food Truck | 3 | Carpet Store | Shop & Service | Food Truck | Nightlife Spot | Yoga Studio |
| 8 | skyland | 39.755713 | -104.950224 | Owl Club of Denver | 39.757190 | -104.946836 | Nightlife Spot | 3 | Carpet Store | Shop & Service | Food Truck | Nightlife Spot | Yoga Studio |
| 9 | skyland | 39.755713 | -104.950224 | East 28th Liquor | 39.757153 | -104.946129 | Shop & Service | 3 | Carpet Store | Shop & Service | Food Truck | Nightlife Spot | Yoga Studio |
| 10 | skyland | 39.755713 | -104.950224 | Modern and Custom Rugs by Doris Leslie | 39.757791 | -104.955253 | Carpet Store | 3 | Carpet Store | Shop & Service | Food Truck | Nightlife Spot | Yoga Studio |

Fig 4.4 Cluster 4

The fourth Neighborhood cluster has four venues nearby which include Street Food, Nightlife spot, shop & a carpet store .

Visualizing the clustered neighborhoods on a map using the folium library (see fig 4.6).
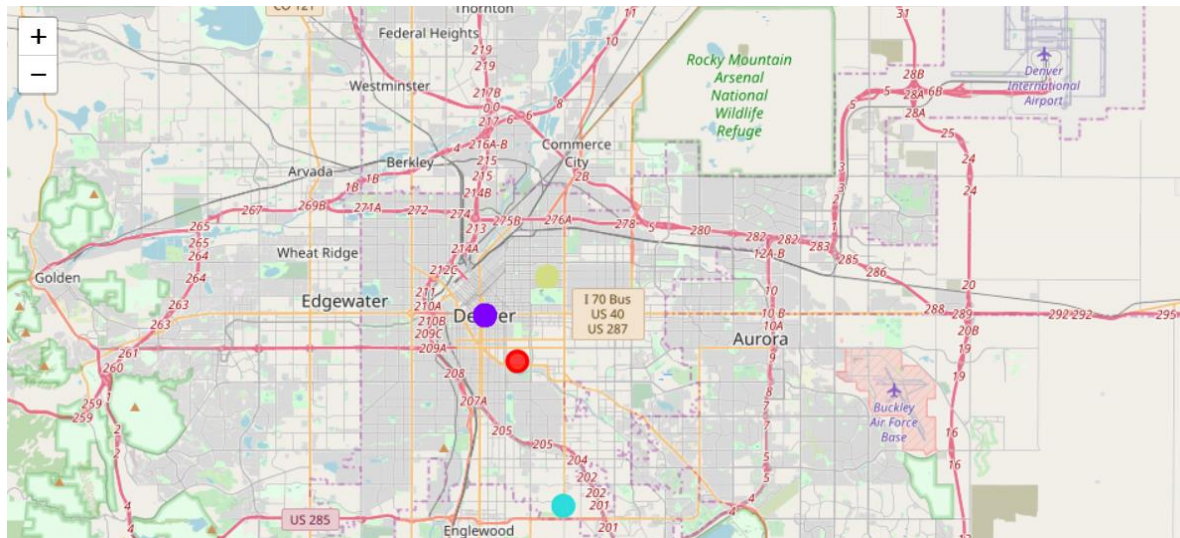


4.6 Clustered neighborhoods in Denver

Each cluster is color coded for ease of presentation. Four clusters have four different color codes.

## 5. Results Discussion

The aim of this project is to help people who want to relocate to the safest neighborhood in Denver, expats can choose the neighborhoods to which they want to relocate based on the most common venues in it. For example, if a person is looking for a neighborhood with good connectivity and public transportation, we can see that Cluster 2 has the most common venues. If a person is looking for a neighborhood with stores and sports centers in a close proximity then the neighborhoods in the third cluster is suitable.

## 6. Conclusion

This project helps a person get a better understanding of the neighborhoods with respect to the most common venues in that neighborhood. It is always helpful to make use of technology to stay one step ahead i.e. finding out more about places before moving into a neighborhood. We have just taken safety as a primary concern to shortlist the Neighborhoods of Denver. The future of this project includes taking other factors such as cost of living in the areas into consideration to shortlist the neighborhoods based on safety and a predefined budget