

# Capstone Project - The Battle of Neighborhoods Report

---

Rama Krishna .K

# 1. Introduction

- **Background** : Safety is a top concern when relocating to new area. If you don't feel safe in the area where you are living, you will not be able to enjoy living there.
- **Problem** : This project aims to select the safest neighborhood in Denver based on total crimes, explore nearby venues for each of the neighborhood to find the 10 most common venues in each neighborhood and finally cluster the neighborhoods using k-mean clustering.
- **Interest** : Expats who are considering to relocate to Denver will be interested to identify the safest neighborhood in Denver and explore the nearby venues for each neighborhood.

## 2. Data Acquisition and Cleaning

### **Data Acquisition:**

The data acquired for this project is a combination of data from three sources.

- 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 second data source is both Latitude and Longitude data for Neighborhoods using python geocoder.
- The third data source is the list of nearby venues for Neighborhoods of the Denver and Venues information using Foursquare.

## **Data Cleaning:**

The data cleaning process for each of the three sources of data are done separately.

- From the Denver crime data, the crimes during the most recent year (2019) are only selected. The major categories of crime are pivoted to get the total crimes per the Neighborhood.
- Sorted the data to find the top 5 and bottom 5 neighborhoods of Denver in terms of number of crimes.
- After sorting the data, extracted the Latitude and Longitude data of each Neighborhood of Denver and Visualized the location of 5 Neighborhoods which have less crime rate.
- The new data set is used to generate the 10 most common venues for each neighborhood using the Foursquare API, finally using k means clustering algorithm to cluster the neighborhoods.

# 3. Methodology

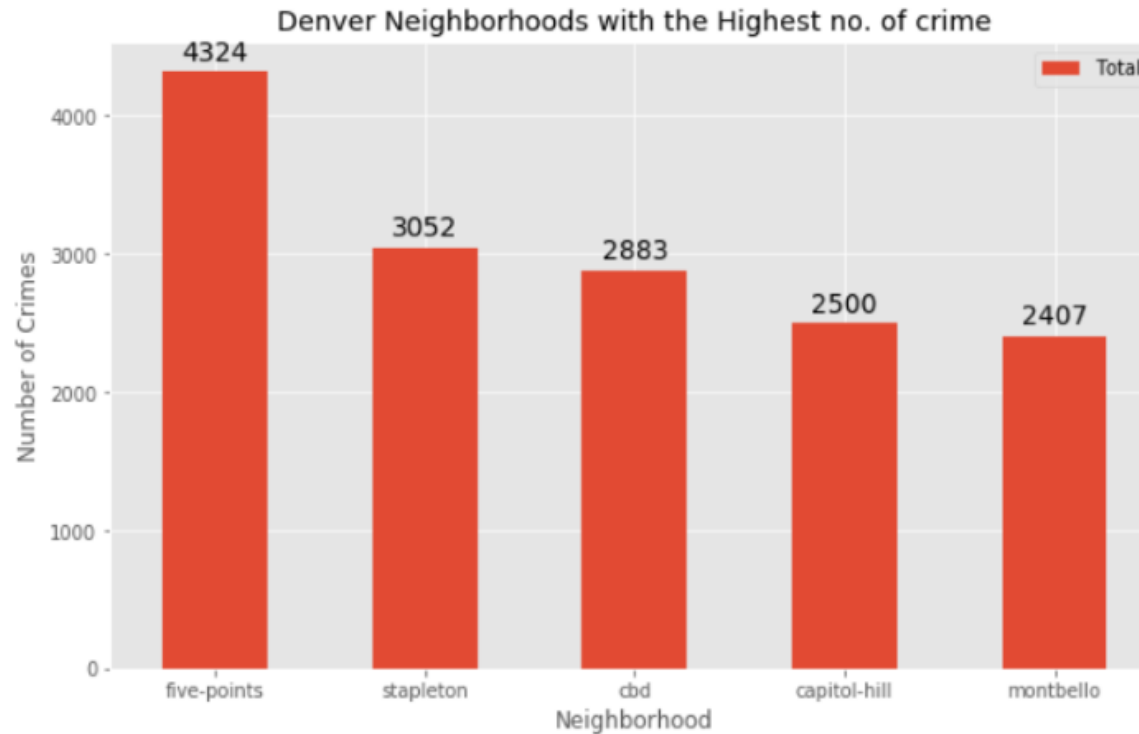
## Exploratory Data Analysis

### Statistical summary of crimes

	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

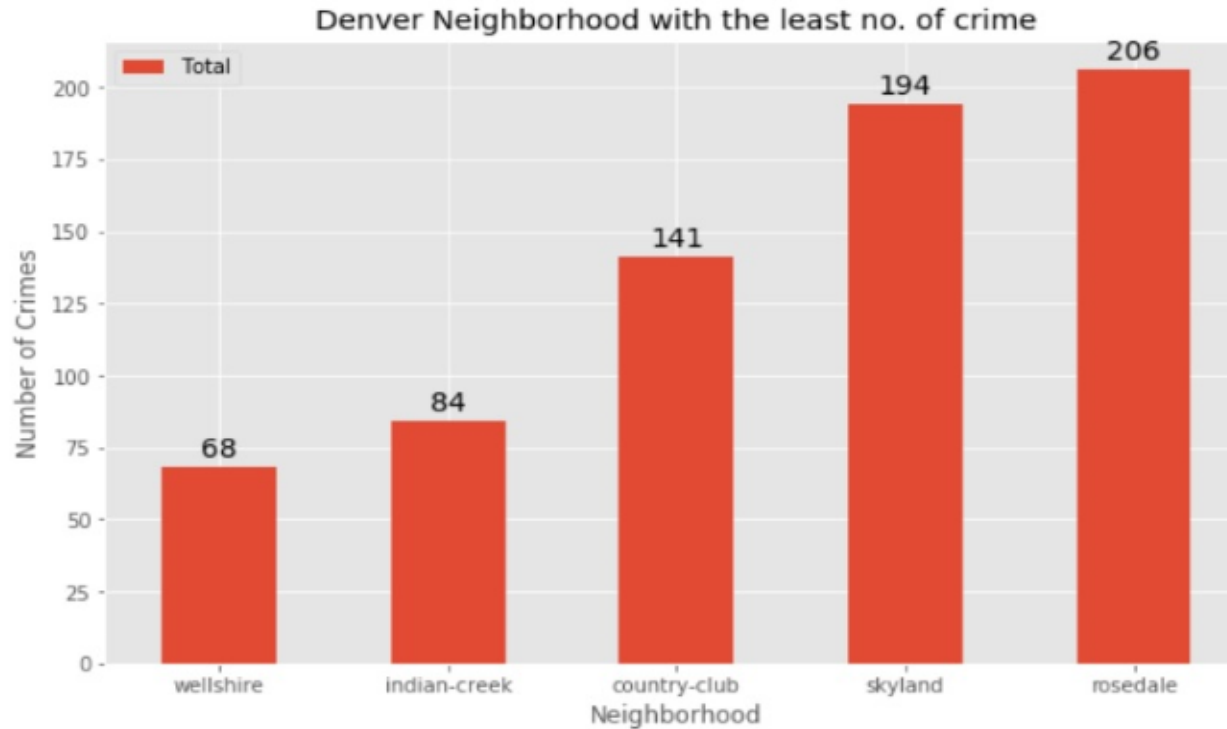
The count for each of the major categories of crime returns the value 78 which is the number of Denver neighborhoods. 'larceny' is the highest reported crime during the year 2019 followed by 'public-disorder' and 'theft-from-motor-vehicle' offenses. The lowest recorded crimes are 'murder', 'Earson' &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

## Neighborhoods with highest crime rate



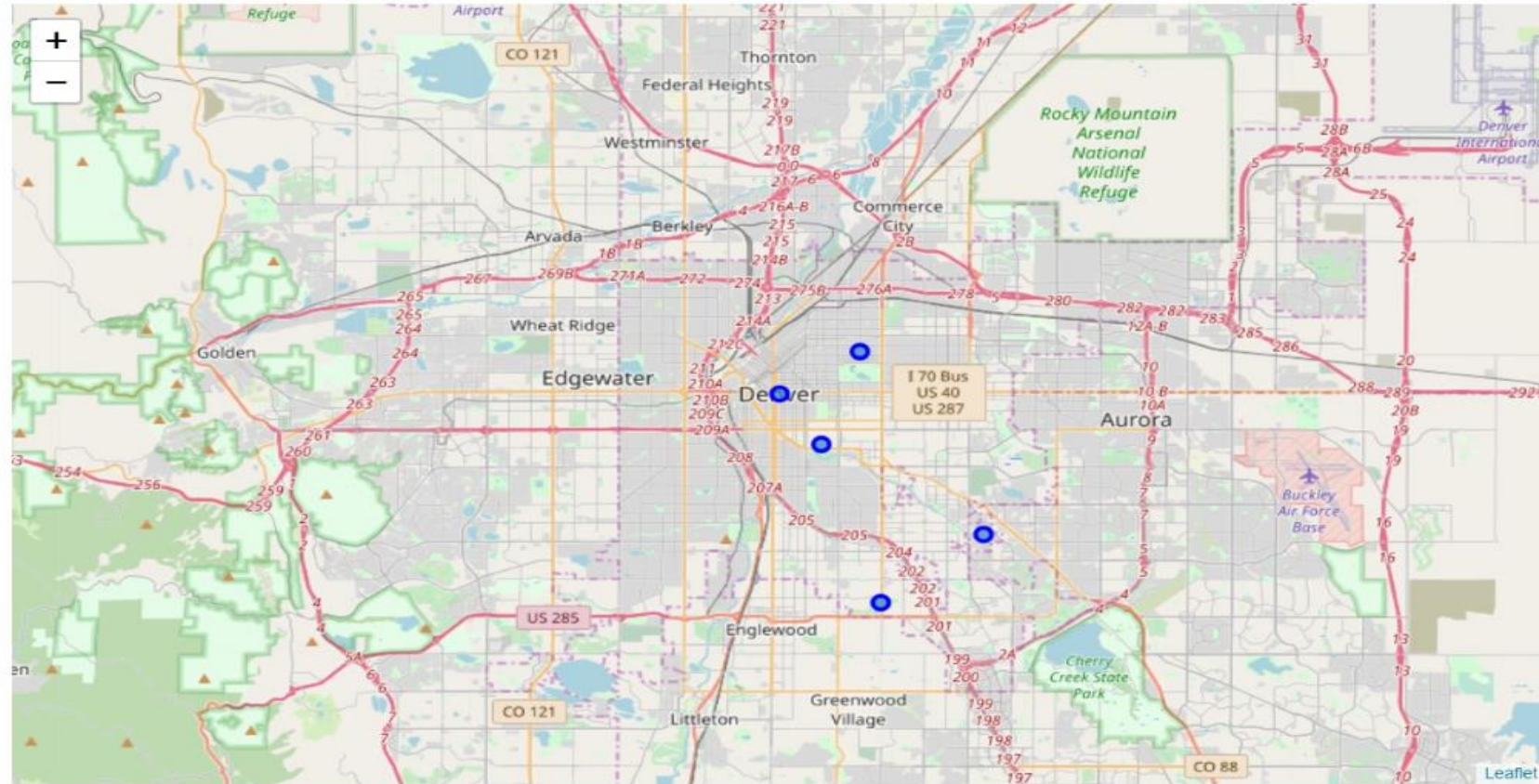
Comparing five neighborhoods with the highest crime rate during the year 2019 it is evident that 'five-points' has the highest crimes recorded followed by staple, cbd, capitalhill & montbello.

## Neighborhoods with lowest crime rate



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

## Neighborhoods with less crime rate



Visualization of the Neighborhoods with least number of crimes reported in 2019 in Denver



## Modeling:

- 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.

	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
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

- 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 similar 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.

## Results:

### Cluster 1: Looking to venues of neighborhood of first cluster

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

	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

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

## Cluster 2: Looking into venues of neighborhood in the second cluster

	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
11	rosedale	39.739236	-104.984862	Sassafras American Eatery	39.739949	-104.982756	Breakfast Spot	1	Yoga Studio	Noodle House	Marijuana Dispensary	Art Museum	Ballroom
12	rosedale	39.739236	-104.984862	Phở-natic	39.740081	-104.984111	Noodle House	1	Yoga Studio	Noodle House	Marijuana Dispensary	Art Museum	Ballroom
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	Ballroom
14	rosedale	39.739236	-104.984862	Capitol Hill Books	39.739979	-104.983472	Bookstore	1	Yoga Studio	Noodle House	Marijuana Dispensary	Art Museum	Ballroom
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	Ballroom
16	rosedale	39.739236	-104.984862	Kindness Yoga	39.736721	-104.984407	Yoga Studio	1	Yoga Studio	Noodle House	Marijuana Dispensary	Art Museum	Ballroom
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	Ballroom
18	rosedale	39.739236	-104.984862	CorePower Yoga	39.737070	-104.983057	Yoga Studio	1	Yoga Studio	Noodle House	Marijuana Dispensary	Art Museum	Ballroom
19	rosedale	39.739236	-104.984862	City Grille	39.740165	-104.982861	Burger Joint	1	Yoga Studio	Noodle House	Marijuana Dispensary	Art Museum	Ballroom
20	rosedale	39.739236	-104.984862	Civic Center Park	39.739370	-104.988776	Park	1	Yoga Studio	Noodle House	Marijuana Dispensary	Art Museum	Ballroom

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

### Cluster 3: Looking into venues of neighborhood in the third cluster

	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
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

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

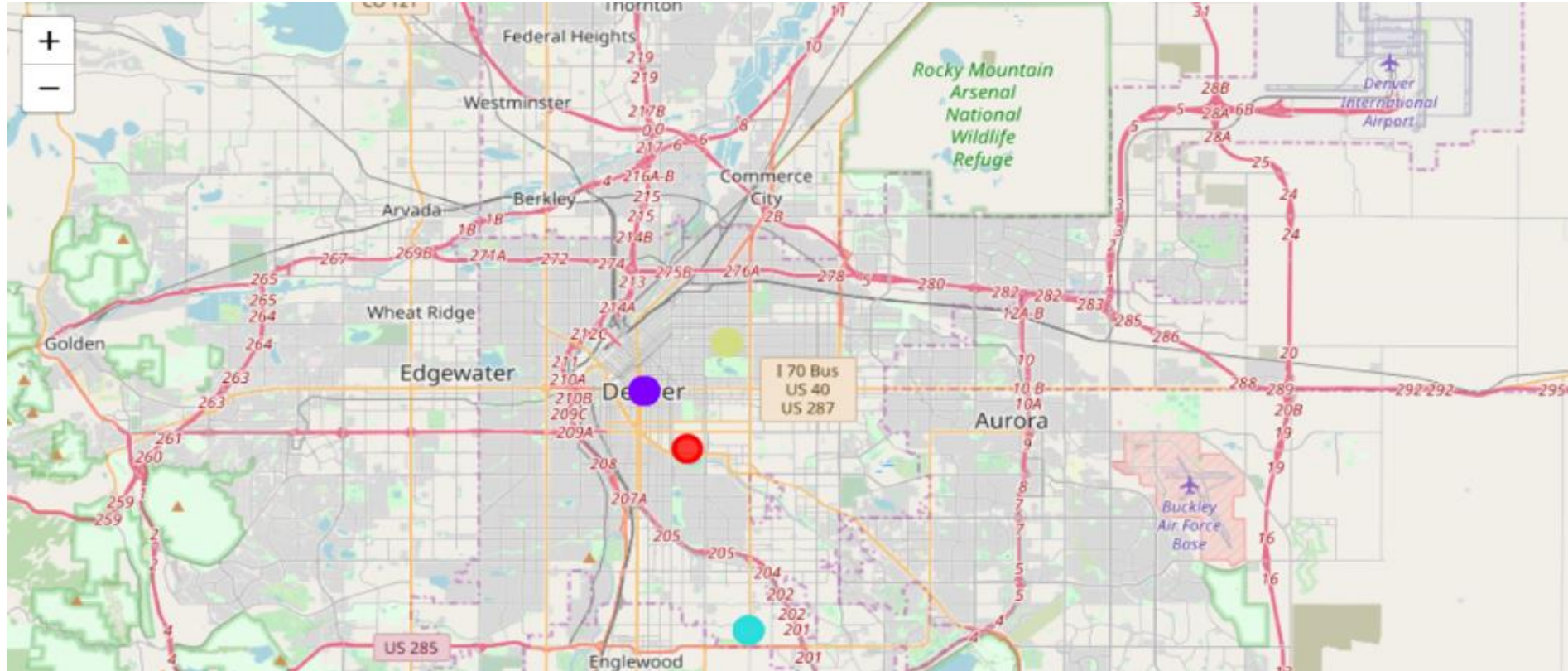
## Cluster 4: Looking into venues of neighborhood in the fourth cluster

	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 Leal	39.757791	-104.955253	Carpet Store	3	Carpet Store	Shop & Service	Food Truck	Nightlife Spot	Yoga Studio

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



After running the K-means clustering we can access each cluster created to see which neighborhoods were assigned to each 4 clusters. Visualizing the clustered neighborhoods on a map using folium library.



Each cluster is color coded for ease of presentation.

## 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.