Capstone Project - The Battle of Neighborhoods

Abdulrahman Falyoun

Introduction

- **Background:** The most important thing that people concerns is their lives and their safety, so when moving to a new city, country you need to check if it's safe or not to live in.
- **Problem:** This project aims to build a system where the citizen is able to select the safest borough in Chicago based on the total crimes between 2010 and 2018.

  The system also allows to explore the neighborhoods of that borough and find the 10 most common venues in each neighborhood and finally cluster the neighborhoods using k-mean clustering.

  This report will be targeted to people who are looking to relocate to Chicago

  We will focus on the safest borough so that any person can choose best neighborhood suited to him.
- **Interest:** People who are looking to relocate to Chicago will be the targeted people, they will be able to identify the safest borough in Chicago and explore its neighborhoods and common venues around each neighborhood

Data Acquisition and cleaning

- **Acquiring data:** The data used for this project is available through:
- A scrapped wikipedia page that contains Chicago boroughs and some related information
- List of neighborhoods acquired using foursquare for least crime rate borough
- The Chicago crimes dataset from ibm box:
  - Https://ibm.box.com/shared/static/svflyugsr9zbqy5bmowgswqemfpm1x7f.csv.

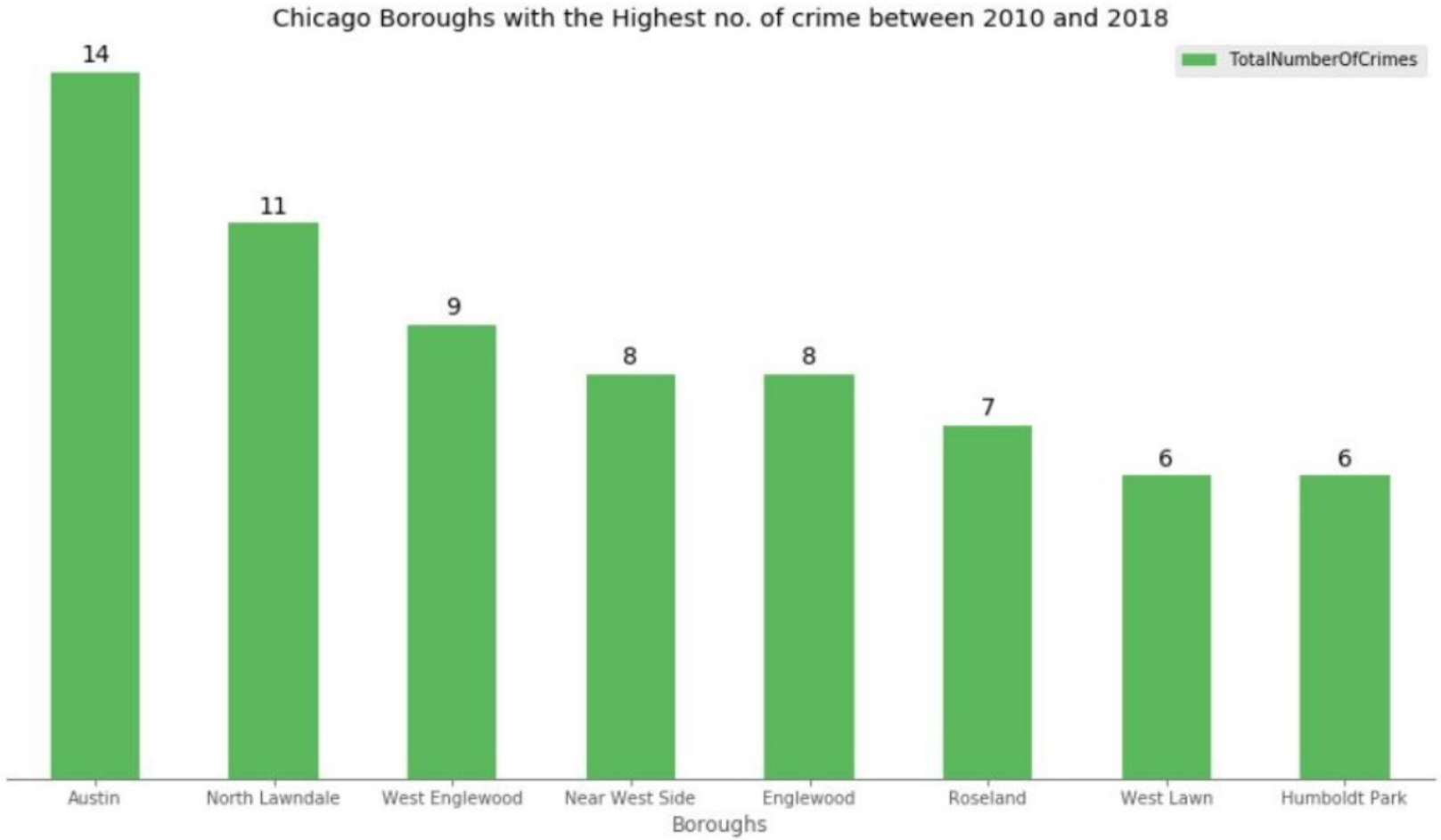
- **Data cleaning:** for each source there is a separate process of cleaning the data
- For scrapped wiki page, we're gonna use the Beautiful Soup library in python.
- For generating the location of neighborhoods around least crime borough rate, we're gonna use Forsquare API and Folium module to visualize the data within a map, as well as we're gonna use k means clustering algorithm to cluster similar neighborhoods together
- For data available from ibm box, we're gonna use Panda, Numpy to get the needed table of crimes information, as well as we will use matplotlib for visualizaion process

Methodology

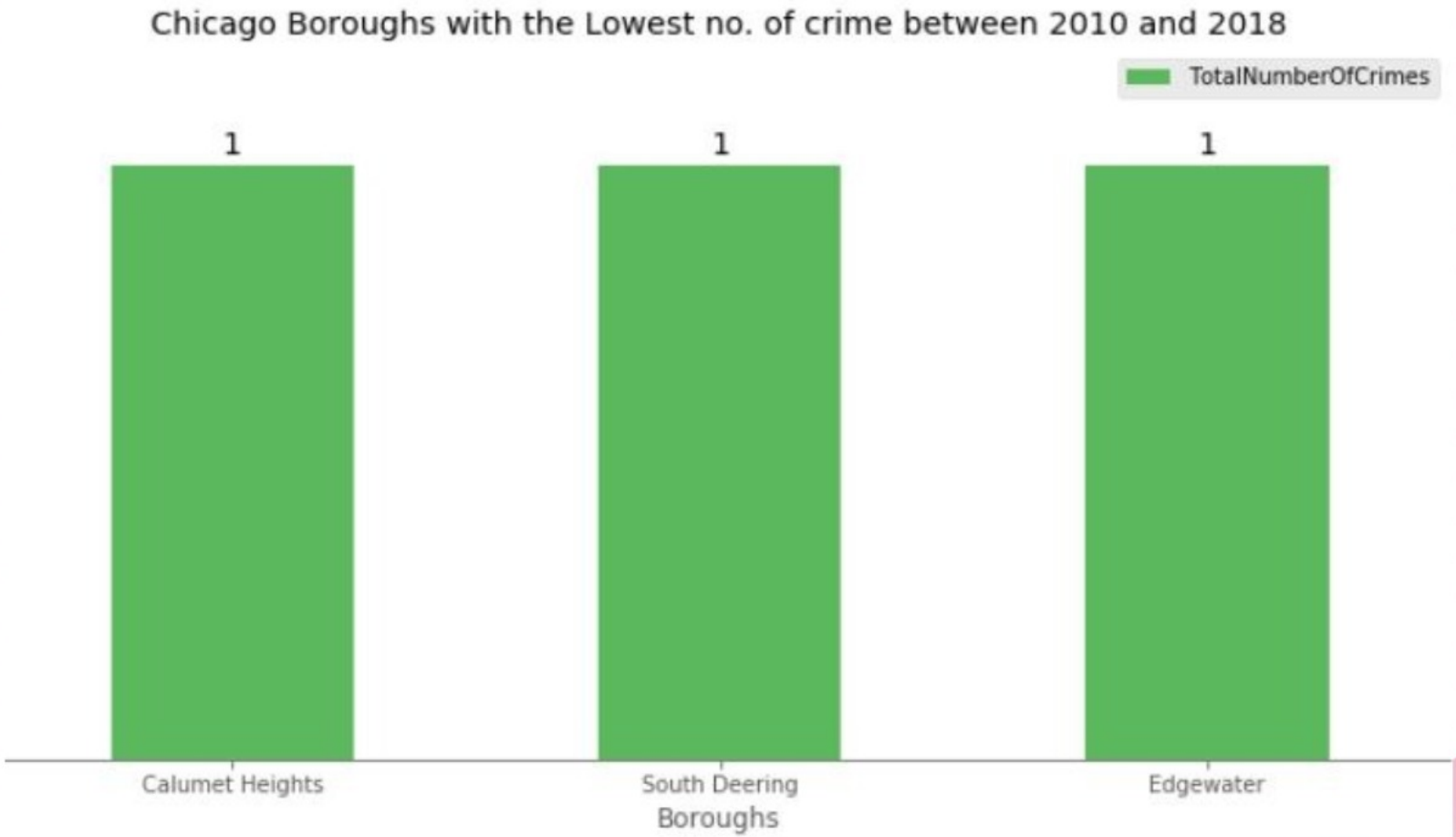
- Statical summary of crimes

	COMMUNITY_AREA_NUMBER	LATITUDE	LONGITUDE	TotalNumberOfCrimes	Boroughs
0	25.0	41.875659	-87.762569	14	Austin
1	29.0	41.854811	-87.716310	11	North Lawndale
2	67.0	41.787324	-87.664412	9	West Englewood
3	28.0	41.865761	-87.646876	8	Near West Side
4	68.0	41.779794	-87.631827	8	Englewood
...	...	...	...	...	...
61	35.0	41.828502	-87.619279	1	Douglas
62	39.0	41.807068	-87.603844	1	Oakland
63	48.0	41.737116	-87.571657	1	Calumet Heights
64	51.0	41.721774	-87.578985	1	South Deering
65	77.0	41.981153	-87.669750	1	Edgewater

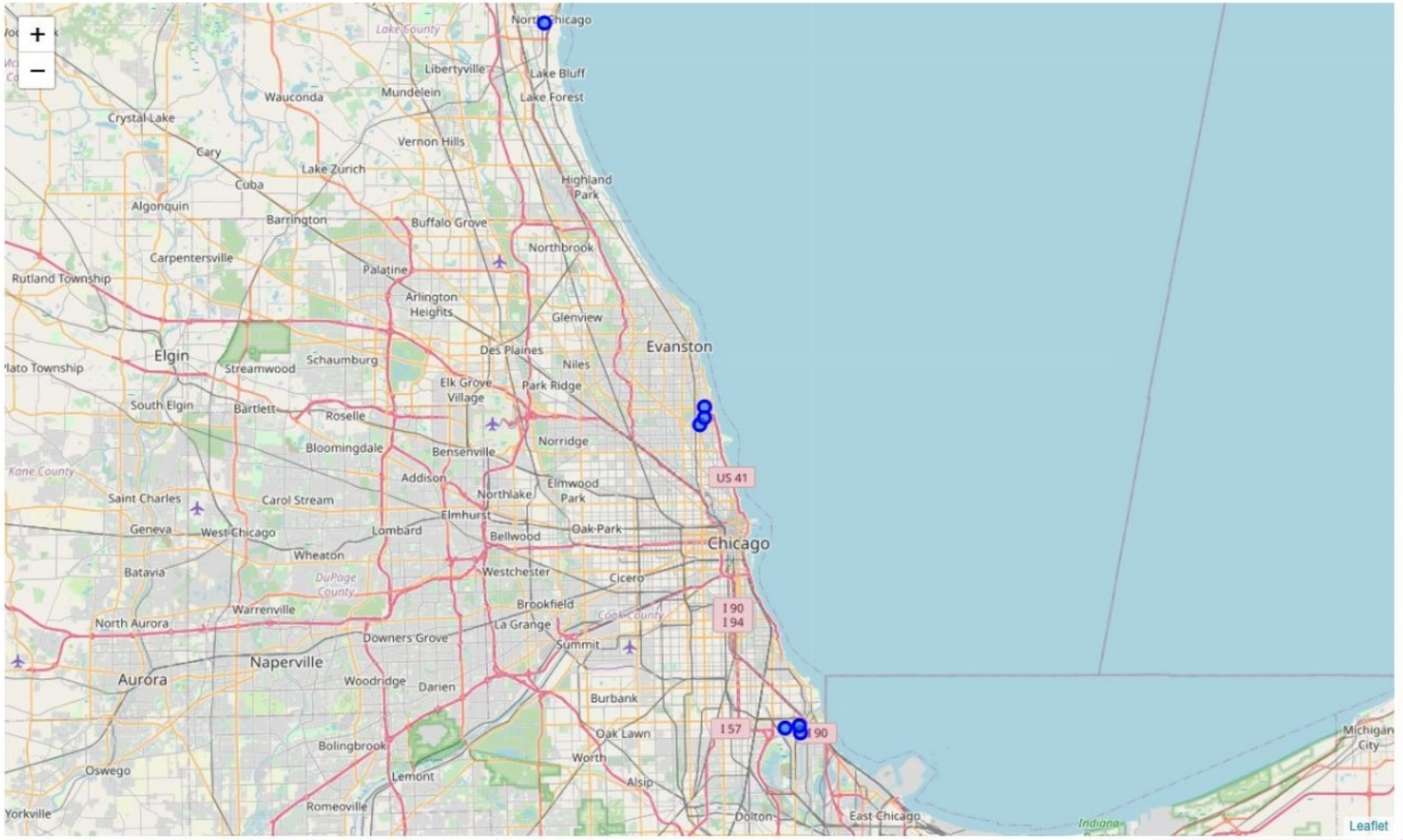
Boroughs with highest number of crime between 2010 and 2018



Boroughs with lowest number of crime between 2010 and 2018



Neighborhoods around the least crime rate borough

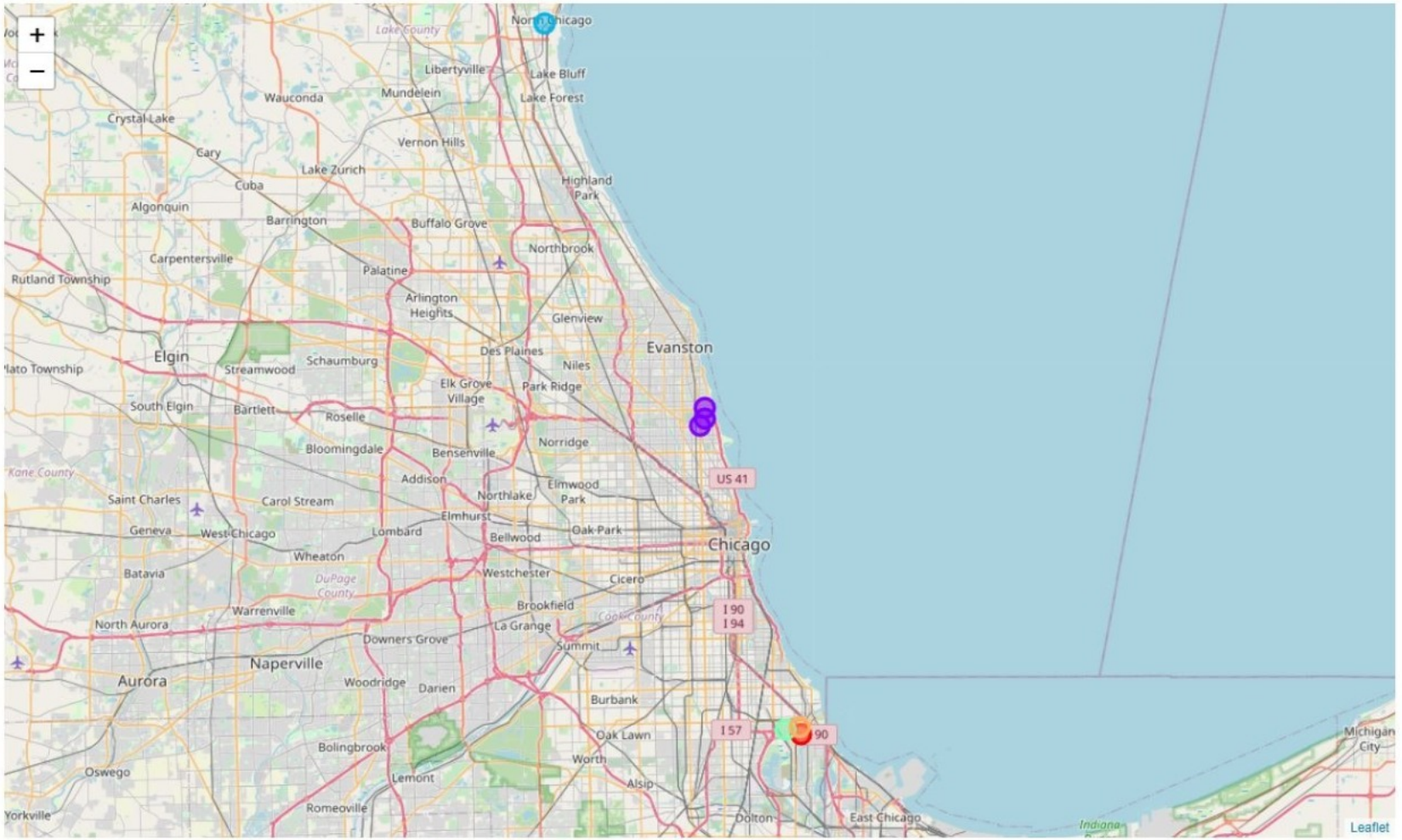


- **Modeling**: After we obtained the safest neighborhoods, we can find all the venues within 500 meters radius of each neighborhood on its own by using Foursquare API.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Andersonville	41.977139	-87.669273	Lost Larson Bakery	41.978617	-87.668411	Bakery
1	Andersonville	41.977139	-87.669273	Taste of Lebanon	41.976151	-87.668847	Middle Eastern Restaurant
2	Andersonville	41.977139	-87.669273	Middle East Bakery & Grocery & Café	41.976213	-87.668610	Grocery Store
3	Andersonville	41.977139	-87.669273	Kopi Café	41.978612	-87.668298	Café
4	Andersonville	41.977139	-87.669273	George's Ice Cream & Sweets	41.978299	-87.668468	Ice Cream Shop

- One hot encoding is done on the venues data shown above in the table, the venues then are grouped by the neighborhood and the mean of the venues calculated, finally the 10 most common venues are calculated for each neighborhood.
- K-Means is used to find similar neighborhoods in the safest borough, then clusters will be constructed

Result


Lake County
Libertyville
Lake Bluff
Lake Forest
Wauconda
Mundelein
Crystal Lake
Vernon Hills
Gary
Lake Zurich
Cuba
Highland Park
Algonquin
Barrington
Buffalo Grove
Northbrook
Carpentersville
Palatine
Rutland Township
Arlington Heights
Glenview
Evanston
Elgin
Schaumburg
Des Plaines
Niles
Streamwood
Elk Grove Village
Park Ridge
South Elgin
Bartlett
Roselle
Norridge
Bloomingdale
Bensenville
Kane County
Addison
US 41
Saint Charles
Carol Stream
Northlake
Elmwood Park
Elmhurst
Geneva
West Chicago
Lombard
Bellwood
Oak Park
Chicago
Wheaton
Batavia
Westchester
Cicero
DuPage County
Brookfield
Warrenville
Cook County
190
194
North Aurora
La Grange
Downers Grove
Summit
Naperville
Aurora
Woodridge
Darien
Burbank
157
90
Oak Lawn
Bolingbrook
Worth
Oswego
Lemont
Alsip
Romeoville
Yorkville
Dolton
East Chicago
Indiana
Leaflet

First three clusteres

```
ven_merged[ven_merged['Cluster Labels'] == 0]
```

	Neighborhood	Boroughs	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
0	Irondale	South Deering	41.710591	-87.552268	0	American Restaurant	Bar	Food & Drink Shop	Wings Joint	Deli / Bodega	Hot Dog Joint	Hobby Shop	Gym / Fitness Center	Gym	Grocery Store

```
ven_merged[ven_merged['Cluster Labels'] == 1]
```

	Neighborhood	Boroughs	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
0	Andersonville	Edgewater	41.977139	-87.669273	1	Middle Eastern Restaurant	Indie Theater	Pet Store	Grocery Store	Gym	Comic Shop	Ice Cream Shop	Coffee Shop	Wine Shop	Irish Pub
1	Edgewater Glen	Edgewater	41.992390	-87.664046	1	Sandwich Place	Wine Bar	Thai Restaurant	Asian Restaurant	Gym / Fitness Center	Mexican Restaurant	Grocery Store	Wings Joint	Wine Shop	Ice Cream Shop
3	Edgewater Beach	Edgewater	41.983369	-87.663952	1	Asian Restaurant	Gym / Fitness Center	Coffee Shop	Sushi Restaurant	Greek Restaurant	Korean Restaurant	Deli / Bodega	Mexican Restaurant	Indian Restaurant	Chinese Restaurant

```
ven_merged[ven_merged['Cluster Labels'] == 2]
```

	Neighborhood	Boroughs	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	Broadway	Edgewater	42.321528	-87.849368	2	Discount Store	Wings Joint	Comic Shop	Hot Dog Joint	Hobby Shop	Gym / Fitness Center	Gym	Grocery Store	Greek Restaurant	Gift Shop

```
ven_merged[ven_merged['Cluster Labels'] == 3]
```

	Neighborhood	Boroughs	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
1	Jeffery Manor	South Deering	41.715035	-87.570046	3	Park	Basketball Court	Music Venue	Wings Joint	Deli / Bodega	Hobby Shop	Gym / Fitness Center	Gym	Grocery Store	Greek Restaurant

Discussion

- The system aims to help people who are looking to relocate to on of Chicago boroughs, so they can pick the safest borough, as well as they are able to pick the best venue fits to their needs, so people will look for each cluster and pick what they needs, for example:
- If the person is looking for American restaurant he might pick the cluster with label 0
- If the person is looking for sandwich place or Asian restaurant he might pick the cluster with label 1, etc...

Conclusion

- The system enables people want to relocate to pick the safest borough in Chicago to live in, taking into consideration the top common venues in the safest areas.
- The system was build using the factor as the total number of crimes happened in Chicage neighborhood between 2010 and 2018.