

# **Suggesting Neighborhoods in Chicago: Per Capita Income and Lower Crime Rate**

**Final Report – IBM Data Science Project**

**Garima Singhal**

**06/12/2020**

## **INTRODUCTION:**

Chicago, Illinois is one of the most populous big cities in the United States. It is famous for its bold architecture, consisting of skyscrapers such as Willis Tower, John Hancock Center, and the Tribune Tower. The city is renowned for its museums, and art galleries. It is also one of the greatest hubs for business, education, industry, culture, transportation, and a lot more. The city is highly diversified with people from all backgrounds and cultures making it the most balanced economy in the United States.

After moving to United States about 5 years ago, I always had the dream of working for a firm in a big city. So, when I got an offer from a technology company in Chicago, I was excited to move. However, when I was looking for places to live in Chicago, I found out that not every neighborhood in the city is considered safe and sound. This led me to do intensive research for finding a good and safe neighborhood in the city.

## **BUSINESS PROBLEM:**

Chicago being one of the cities embodying highly paid corporate jobs, we see that more people are moving into the city every day. Considering Chicago, which has an overall crime rate higher than the US average, it becomes challenging to find a good and safe neighborhood.

In 2016, the city saw a surge in gun violence with 762 murders, 3550 shooting incidents, and 4331 shooting victims which was more than the number of murders in New York City and Los Angeles, combined. The estimated number of homicides in Chicago increased by 52% in 2016.

Most of these killings happened in five mostly black and Latino neighborhoods on the south and the west side of city.

To buy an apartment or a house, deciding which neighborhood you should choose is one of the most important decisions. Safety is the foremost priority when it comes to finding the right neighborhood and income being the second most important. As I have seen from my personal experience the process of finding a safe neighborhood based on your annual income can be tiring.

The aim of this project is to find a safe neighborhood based on the crime rate and per capita income in various neighborhoods across the city of Chicago. The goal of this project is to help new people move into the city and help them find a neighborhood which is safe, has a low crime rate and fits into their budget.

## **METHODOLOGY:**

### **Data Acquisition:**

The data required for this project is a combination of three data sources. The first source of data is a Wikipedia Page that contains the list of the Chicago community areas. The dataset consists of following columns:

| Column Name         | Description                                    | Type       |
|---------------------|--|------------|
| Serial Number       |  | Number     |
| Community Area Name |  | Plain Text |
| Neighborhood        | Name of the neighborhood in the Community area | Plain Text |

The second data source for the project will use the Chicago Crime Data that shows the crime per community area in Chicago. The dataset consists of the following columns:

| Column Name | Description                       | Type   |
|-------------|-----------------------------------|--------|
| ID          | Unique identifier for the record. | Number |

|                |  |             |
|----------------|--|-------------|
| Case Number    | The Chicago Police Department Record Number  | Plain Text  |
| Date           | Date when the incident occurred.   | Date & Time |
| Block          | The partially redacted address   | Plain Text  |
| IUCR           | The Illinois Uniform Crime Reporting code.   | Plain Text  |
| Primary Type   | The primary description of the IUCR code.  | Plain Text  |
| Description    | The secondary description of the IUCR code.  | Plain Text  |
| Location       | Description of the location  | Plain Text  |
| Arrest         | Indicates whether an arrest was made.  | Checkbox    |
| Domestic       | Indicates whether the incident was domestic related as defined by the Illinois Domestic Violence Act.          | Checkbox    |
| Beat           | Indicates the beat where the incident occurred.  | Plain Text  |
| District       | Indicates the police district where the incident occurred.   | Plain Text  |
| Ward           | The ward (City Council district) where the incident occurred.  | Number      |
| Community Area | Indicates the community area where the incident occurred.  | Plain Text  |
| FBI Code       | Indicates the crime classification as outlined in the FBI's National Incident-Based Reporting System (NIBRS).  | Plain Text  |
| X Coordinate   | The x coordinate of the location where the incident occurred in State Plane Illinois East NAD 1983 projection. | Number      |
| Y Coordinate   | The y coordinate of the location where the incident occurred in State Plane Illinois East NAD 1983 projection. | Number      |

|            |  |                |
|------------|--|----------------|
| Year       | Year the incident occurred.  | Number         |
| Updated On | Date and time the record was last updated.   | Date &<br>Time |
| Latitude   | The latitude of the location where the incident occurred.  | Number         |
| Longitude  | The longitude of the location where the incident occurred.   | Number         |
| Location   | The location where the incident occurred in a format that allows for creation of maps and other geographic operations on this data portal. | Location       |

The third data source, the Chicago Census Data – Selected socioeconomic indicators in Chicago, 2008 – 2012 will be used. This dataset contains a selection of six socioeconomic indicators of public health significance and a “hardship index”, for each community area. The dataset consists of the following columns:

| Column Name                                     | Description   | Type          |
|---|---|---------------|
| Community Area Number                           |   | Number        |
| COMMUNITY AREA NAME                             |   | Plain<br>Text |
| PERCENT OF HOUSING<br>CROWDED                   | Percent occupied housing units with more than one person per room           | Number        |
| PERCENT HOUSEHOLDS BELOW<br>POVERTY             | Percent of households living below the federal poverty level                | Number        |
| PERCENT AGED 16+<br>UNEMPLOYED                  | Percent of persons over the age of 16 years that are unemployed             | Number        |
| PERCENT AGED 25+ WITHOUT<br>HIGH SCHOOL DIPLOMA | Percent of persons over the age of 25 years without a high school education | Number        |

|                                     |  |        |
|-------------------------------------|--|--------|
| PERCENT AGED UNDER 18 OR<br>OVER 64 | Percent of the population under 18 or over 64 years of age (i.e.,<br>dependency)   | Number |
| PER CAPITA INCOME                   | Community Area Per capita income is estimated as the sum of tract<br>level aggregate incomes divided by the total population | Number |
| HARDSHIP INDEX                      | Score that incorporates each of the six selected socioeconomic<br>indicators<br>(see dataset description)                    | Number |

We will use web scraping technique to extract the data from the Wikipedia page, with the help of python requests, and BeautifulSoup packages. Then we will get the geographical coordinates of the neighborhoods using Python Geocoder package which will provide us the latitude and longitude coordinates of all the neighborhoods.

The next step in the project will be using the Foursquare API to get the venue data for those neighborhoods. Foursquare consists of one of the largest databases of 105+ million places and is used by almost 125,000 developers. This project will use multiple data science skills, such as Web Scraping, working with Foursquare API, data cleaning, data wrangling, machine learning algorithm: K-means clustering, and data visualization using the Folium package.

### **Data Cleaning and Processing:**

As discussed in the data acquisition section, the data selected for this project comes from three different sources. Data Preparation of all three sources is done separately. The data preparation includes data cleaning and data processing as it is to be used for the project.

The first data source: Wikipedia page (Link 1) is web scraped using the BeautifulSoup Package library. Using this, we extract the data in a tabular format as it appears on the website. After this step, we used string manipulation to get the Community areas name in the correct form (Figure 1). This is important because another dataset is merged to this later.

|   | Neighborhood    | Community Area Name |
|---|-----------------|---------------------|
| 0 | Albany Park     | Albany Park         |
| 1 | Altgeld Gardens | Riverdale           |
| 2 | Andersonville   | Edgewater           |
| 3 | Archer Heights  | Archer Heights      |
| 4 | Armour Square   | Armour Square       |

Figure 1: Web scraped data

The second data source is from the Chicago Crime Data (Link 2) from which the crimes during the most recent year (2020) are only selected. The major categories of crime are segregated by community area number to get the total crime cases per community area (Figure 2).

| ID | Case Number       | Date                   | Block                 | IUCR | Primary Type       | Description               | Location Description | Arrest | Domestic | Beat | District | Ward | Community Area Number | FBI Code | X Coordinate | Y Coordinate | Year | Updated On             | Latitude  | Longit  |
|----|-------------------|------------------------|-----------------------|------|--------------------|---------------------------|----------------------|--------|----------|------|----------|------|-----------------------|----------|--------------|--------------|------|------------------------|-----------|---------|
| 1  | 12070768 JD255590 | 05/31/2020 04:00:00 PM | 001XX W 95TH ST       | 0810 | THEFT              | OVER \$500                | SMALL RETAIL STORE   | False  | False    | 634  | 6        | 21.0 | 49                    | 06       | 1177048.0    | 1841963.0    | 2020 | 06/10/2020 03:46:57 PM | 41.721682 | -87.627 |
| 2  | 12067497 JD252152 | 06/03/2020 09:55:00 AM | 012XX N CLYBOURN AVE  | 0320 | ROBBERY            | STRONG ARM - NO WEAPON    | BANK                 | False  | False    | 1821 | 18       | 27.0 | 8                     | 03       | 1172996.0    | 1908394.0    | 2020 | 06/10/2020 03:49:15 PM | 41.904065 | -87.639 |
| 3  | 12069889 JD253440 | 05/31/2020 04:00:00 AM | 006XX S WELLS ST      | 1310 | CRIMINAL DAMAGE    | TO PROPERTY               | OTHER (SPECIFY)      | False  | False    | 123  | 1        | 25.0 | 32                    | 14       | 1174818.0    | 1897412.0    | 2020 | 06/10/2020 03:46:57 PM | 41.873889 | -87.633 |
| 4  | 12067925 JD252507 | 06/03/2020 01:30:00 PM | 009XX W FULLERTON AVE | 1210 | DECEPTIVE PRACTICE | THEFT OF LABOR / SERVICES | CTA STATION          | False  | False    | 1812 | 18       | 43.0 | 7                     | 11       | 1169577.0    | 1916141.0    | 2020 | 06/10/2020 03:49:15 PM | 41.925398 | -87.652 |
| 5  | 25162 JD249292    | 06/03/2020 12:08:00 AM | 003XX W 64TH ST       | 0110 | HOMICIDE           | FIRST DEGREE MURDER       | HOSPITAL             | False  | False    | 722  | 7        | 20.0 | 68                    | 01A      | 1175044.0    | 1882506.0    | 2020 | 06/10/2020 03:49:15 PM | 41.778099 | -87.633 |

Figure 2: Chicago Crime Data

The third data source, Chicago Census Data – Selected socioeconomic indicators in Chicago, 2008-2012 (Link 3) from which the non-desirable columns, and null values are dropped. Here is what the data looks like after processing (Figure 3).

|   | Community Area Number | Community Area Name | Per_Capita_Income |
|---|-----------------------|---------------------|-------------------|
| 1 | 1                     | Rogers Park         | 23939             |
| 2 | 2                     | West Ridge          | 23040             |
| 3 | 3                     | Uptown              | 35787             |
| 4 | 4                     | Lincoln Square      | 37524             |
| 5 | 5                     | North Center        | 57123             |

Figure 3: Chicago Census Data

The Crime and the Census datasets are merged on the Community Area Number to form a new dataset (Figure 4). The purpose of this dataset is to visualize the distribution of crime and per capita income across community areas and identify the community areas with the least crime records and higher per capita income during the year 2020.

|   | Community Area Number | Community Area Name | Per_Capita_Income | Total_Cases |
|---|-----------------------|---------------------|-------------------|-------------|
| 0 | 1                     | Rogers Park         | 23939             | 1279        |
| 1 | 2                     | West Ridge          | 23040             | 1125        |
| 2 | 3                     | Uptown              | 35787             | 1092        |
| 3 | 4                     | Lincoln Square      | 37524             | 682         |
| 4 | 5                     | North Center        | 57123             | 451         |

Figure 4: Chicago Crime and Census Data merged

Range of per capita income is distributed randomly between \$8201 to \$88,669 with an average of \$25,597 (Figure 5).

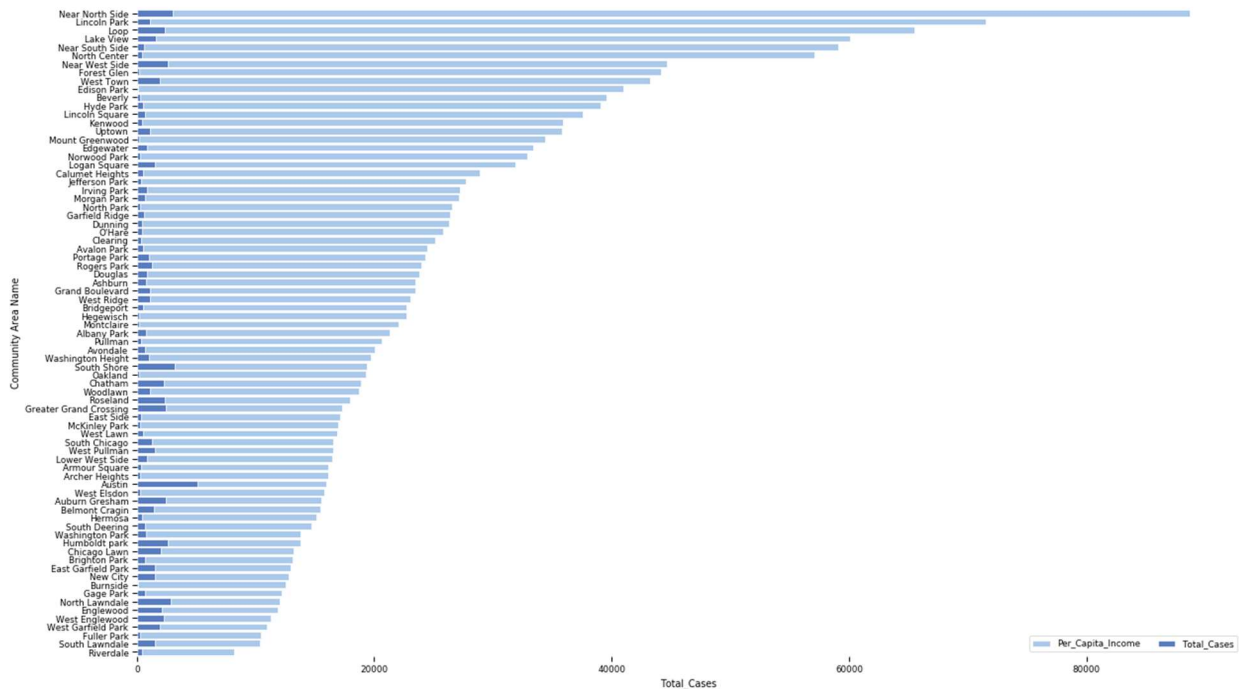


Figure 5: Distribution of Per Capita Income vs Community Area Name

The description of Per Capita Income and Total number of crime cases community area wise is given as below (Figure 6):

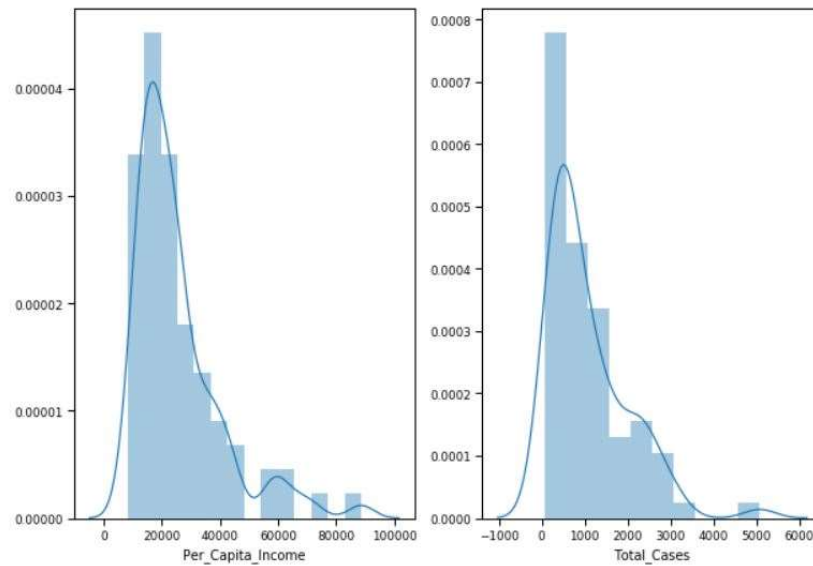


Figure 6: Histogram of Per Capita Income and Total Cases

As per the above plot distribution, total number of crime cases are distributed between 5084 to 79 with a mean of 1082 and median of 745 (Figure 7).

|       | Community Area Number | Per_Capita_Income | Total_Cases |
|-------|-----------------------|-------------------|-------------|
| count | 77.000000             | 77.000000         | 77.000000   |
| mean  | 39.000000             | 25563.168831      | 1082.831169 |
| std   | 22.371857             | 15293.098259      | 928.633042  |
| min   | 1.000000              | 8201.000000       | 79.000000   |
| 25%   | 20.000000             | 15754.000000      | 417.000000  |
| 50%   | 39.000000             | 21323.000000      | 745.000000  |
| 75%   | 58.000000             | 28887.000000      | 1533.000000 |
| max   | 77.000000             | 88669.000000      | 5084.000000 |

Figure 7: Description of the Data frame

For selection of the community area with lower crime rate and higher per capita income, initial top 50 community areas are selected among all the areas. After filtering, this dataset is merged with the associate neighborhood. This new dataset is created from scratch, consisting of names of



the neighborhoods and the community areas. The coordinates (Latitude, Longitude) of the neighborhoods are fetched using the Geocoder package to create a final consolidated dataset of the neighborhoods, along with the community areas, total crime cases, and per capita income (Figure 8).

|    | Neighborhood     | Community Area Name | Latitude  | Longitude  | Per_Capita_Income | Total_Cases |
|----|------------------|---------------------|-----------|------------|-------------------|-------------|
| 0  | Albany Park      | Albany Park         | 41.968290 | -87.723380 | 21323             | 767         |
| 1  | Mayfair          | Albany Park         | 41.691450 | -87.708300 | 21323             | 767         |
| 2  | North Mayfair    | Albany Park         | 41.979590 | -87.904460 | 21323             | 767         |
| 3  | Ravenswood Manor | Albany Park         | 41.973512 | -87.865461 | 21323             | 767         |
| 4  | Ashburn          | Ashburn             | 41.747850 | -87.709950 | 23482             | 745         |
| 5  | Ashburn Estates  | Ashburn             | 41.941674 | -88.198809 | 23482             | 745         |
| 6  | Beverly View     | Ashburn             | 41.695888 | -87.649990 | 23482             | 745         |
| 7  | Crestline        | Ashburn             | 41.843090 | -87.627830 | 23482             | 745         |
| 8  | Parkview         | Ashburn             | 41.816538 | -87.619778 | 23482             | 745         |
| 9  | Scottsdale       | Ashburn             | 42.007122 | -87.675720 | 23482             | 745         |
| 10 | Avalon Park      | Avalon Park         | 41.745070 | -87.588160 | 24454             | 483         |
| 11 | Marynook         | Avalon Park         | 41.690390 | -87.665990 | 24454             | 483         |

*Figure 8: Consolidated dataset of neighborhoods along with its Geographical Location, Crime Data, and Census Data*

Neighborhoods with low crime rate and high per capita income are selected. The neighborhoods that satisfy the criteria are visualized below using the Folium library in Python (Figure 9).



Figure 9: Visualization of the Selected Neighborhoods

## Modelling:

Using the final dataset containing the selected neighborhoods along with their latitude and longitude, we then find all the venues within 500-meter radius of each neighborhood by connecting to the Foursquare API. This returns a json file containing all the venues along with their coordinates and categories that they belong to (Figure 10).

|   | Neighborhood | Latitude | Longitude | VenueName                | VenueLatitude | VenueLongitude | VenueCategory      |
|---|--------------|----------|-----------|--------------------------|---------------|----------------|--------------------|
| 0 | Albany Park  | 41.96829 | -87.72338 | Lawrence Fish Market     | 41.968280     | -87.726250     | Seafood Restaurant |
| 1 | Albany Park  | 41.96829 | -87.72338 | Chicago Kalbi Korean BBQ | 41.968314     | -87.722771     | Korean Restaurant  |
| 2 | Albany Park  | 41.96829 | -87.72338 | Starbucks                | 41.968911     | -87.728817     | Coffee Shop        |
| 3 | Albany Park  | 41.96829 | -87.72338 | Rojo Gusano              | 41.968425     | -87.724549     | Taco Place         |
| 4 | Albany Park  | 41.96829 | -87.72338 | El Gallo Bravo #6        | 41.968324     | -87.721338     | Mexican Restaurant |

Figure 10: Venue Details of each neighborhood

For analyzing each neighborhood, we use the process of one hot encoding. One hot encoding is a process by which categorical variables are converted into a form that could be provided to Machine Learning algorithms for a better prediction. This process is performed on the venues data. The venues data is then grouped by the neighborhood and the mean of the venues is calculated. Finally, we calculate the top 10 common venues for each neighborhood (Figure 11).

|   | Neighborhood    | 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 | Albany Park     | Mexican Restaurant    | Bus Station           | Korean Restaurant     | Pharmacy              | Pet Store             | Park                  | Coffee Shop           | Discount Store            | Dive Bar              | Seafood Restaurant     |
| 1 | Ashburn         | Cosmetics Shop        | Martial Arts Dojo     | Nightclub             | Bar                   | Light Rail Station    | Bus Station           | Automotive Shop       | Snack Place               | Fast Food Restaurant  | Field                  |
| 2 | Ashburn Estates | Mexican Restaurant    | Gym                   | Yoga Studio           | Farmers Market        | Ethiopian Restaurant  | Event Service         | Exhibit               | Falafel Restaurant        | Farm                  | Fast Food Restaurant   |
| 3 | Avalon Park     | Pizza Place           | Burger Joint          | Fast Food Restaurant  | ATM                   | Grocery Store         | Diner                 | Sandwich Place        | Cajun / Creole Restaurant | Business Service      | Boutique               |
| 4 | Avondale        | Food Truck            | Chinese Restaurant    | Hot Dog Joint         | Diner                 | Brewery               | Soccer Field          | Bus Line              | Bus Station               | Storage Facility      | Supermarket            |

Figure 11: Ten most common venues in each neighborhood

To help people find similar neighborhoods in the safest community area, we cluster similar neighborhoods using K-means clustering algorithm which is a form of unsupervised machine learning algorithm that clusters data based on the predefined cluster size. A cluster size of 5 clusters all the selected neighborhoods into 5 separate clusters. The reason to conduct a K-means clustering is to cluster neighborhoods with similar venues together so that people can shortlist the neighborhoods of their interests based on the venues and facilities provided in each neighborhood.

## RESULTS:

After running the K-means clustering algorithm, we can access each cluster created to see which neighborhoods were assigned to each of the five clusters. Looking into the neighborhoods in the first cluster (Figure 12). This cluster consists of the maximum number of neighborhoods spared across Albany Park, Edison Park etc. Upon closely examining these neighborhoods, we see that the most common venues in these neighborhoods are food joints, bar, café, bike shop, gym/yoga studio, pharmacy, grocery stores, parks etc.

|    | Neighborhood    | Per_Capita_Income | Total_Cases | 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  | Albany Park     | 21323             | 767         | 0              | Mexican Restaurant    | Bus Station                     | Korean Restaurant     | Pharmacy              | Pet Store             | Park                  | Coffee Shop               | Discount Store        | Dive Bar              | Seafood Restaurant     |
| 55 | South Edgebrook | 44164             | 189         | 0              | Dog Run               | Southern / Soul Food Restaurant | Yoga Studio           | Farmers Market        | Ethiopian Restaurant  | Event Service         | Exhibit                   | Falafel Restaurant    | Farm                  | Fast Food Restaurant   |
| 54 | Sauganash       | 44164             | 189         | 0              | Mexican Restaurant    | Ice Cream Shop                  | American Restaurant   | Grocery Store         | Italian Restaurant    | Yoga Studio           | Farm                      | Ethiopian Restaurant  | Event Service         | Exhibit                |
| 53 | Old Edgebrook   | 44164             | 189         | 0              | Dive Bar              | Ice Cream Shop                  | Bike Shop             | Fried Chicken Joint   | Taco Place            | Caribbean Restaurant  | Latin American Restaurant | Bar                   | Cocktail Bar          | Coffee Shop            |
| 52 | Forest Glen     | 44164             | 189         | 0              | Park                  | Bus Station                     | Baseball Field        | Food                  | Chinese Restaurant    | Department Store      | Train Station             | Falafel Restaurant    | Event Service         | Exhibit                |
| 51 | Edgebrook       | 44164             | 189         | 0              | Mexican Restaurant    | Sushi Restaurant                | Indian Restaurant     | Asian Restaurant      | Antique Shop          | Bakery                | Yoga Studio               | Coffee Shop           | Spa                   | Boutique               |
| 50 | East Side       | 17104             | 355         | 0              | Mexican               | Pizza Place                     | Café                  | Bar                   | Sushi                 | Coffee Shop           | Critical Shop             | Dive Bar              | Gym                   | Ice Cream              |

Figure 12: Cluster 1

The second cluster (Figure 13) consists of two neighborhoods Old Norwood and Beverly. The most common venues in these neighborhoods are parks, breakfast places, farmers market, electronics store, Ethiopian restaurant.

|    | Neighborhood | Per_Capita_Income | Total_Cases | 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 |
|----|--------------|-------------------|-------------|----------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|------------------------|
| 28 | Old Norwood  | 32875             | 291         | 1              | Park                  | Mini Golf             | Breakfast Spot        | Farmers Market        | Ethiopian Restaurant  | Event Service         | Exhibit               | Falafel Restaurant    | Farm                  | Fast Food Restaurant   |
| 19 | Beverly      | 39523             | 296         | 1              | Flower Shop           | Platform              | Park                  | Farm                  | Electronics Store     | Ethiopian Restaurant  | Event Service         | Exhibit               | Falafel Restaurant    | Farmers Market         |

Figure 13: Cluster 2

The third, fourth, and fifth clusters (Figure 14, 15, 16) consist of just one neighborhood each. This is because of the unique venues in each of the neighborhoods, hence they could not be clustered into the similar neighborhoods.

The most common venues in cluster three are a Mexican restaurant, gym/yoga studio, farmers market.

|   | Neighborhood    | Per_Capita_Income | Total_Cases | 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 |
|---|-----------------|-------------------|-------------|----------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|------------------------|
| 5 | Ashburn Estates | 23482             | 745         | 2              | Mexican Restaurant    | Gym                   | Yoga Studio           | Farmers Market        | Ethiopian Restaurant  | Event Service         | Exhibit               | Falafel Restaurant    | Farm                  | Fast Food Restaurant   |

Figure 14: Cluster 3

The most common venues in cluster four are a convenience store, yoga studio, Ethiopian restaurant.

|    | Neighborhood       | Per_Capita_Income | Total_Cases | 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 |
|----|--------------------|-------------------|-------------|----------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|------------------------|
| 35 | Ravenswood Gardens | 37524             | 682         | 3              | Convenience Store     | Intersection          | Yoga Studio           | Farmers Market        | Ethiopian Restaurant  | Event Service         | Exhibit               | Falafel Restaurant    | Farm                  | Fast Food Restaurant   |

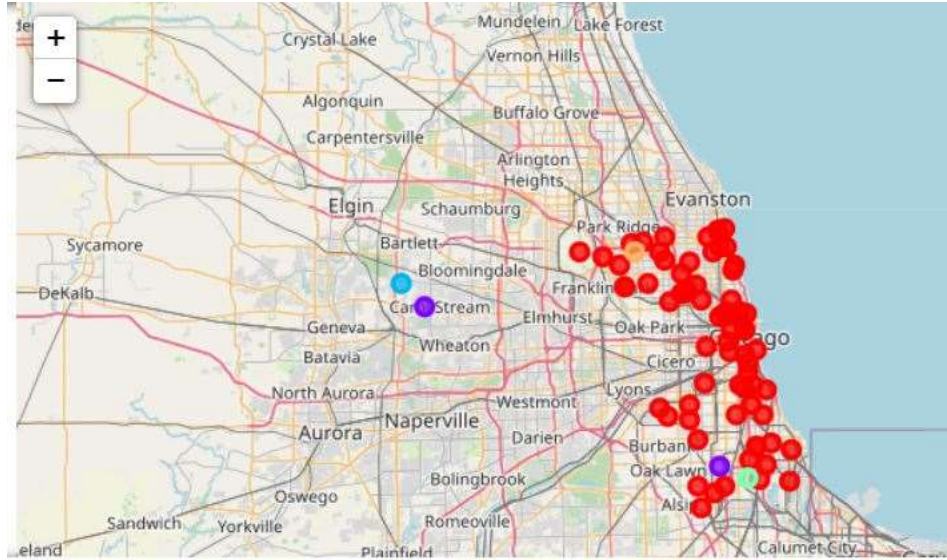
Figure 15: Cluster 4

The most common venues in cluster five are football stadiums, parks, yoga studio, farms.

|    | Neighborhood | Per_Capita_Income | Total_Cases | 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 |
|----|--------------|-------------------|-------------|----------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|------------------------|
| 75 | O'Hare       | 25828             | 436         | 4              | Football Stadium      | Park                  | Yoga Studio           | Farm                  | Electronics Store     | Ethiopian Restaurant  | Event Service         | Exhibit               | Falafel Restaurant    | Farmers Market         |

Figure 16: Cluster 5

Visualization of the clustered neighborhoods on a map using a Folium library in Python (Figure 17).



*Figure 17: Visualization of the clustered neighborhoods*

Each cluster is color coded for the ease of presentation; we can see that most of the neighborhoods falls in the red cluster which is the first cluster. The purple cluster consists of 2 neighborhoods which is represented by the second cluster. The blue, orange, and green clusters consists of 1 neighborhood each representing the third, fourth, and fifth cluster, respectively.

## **DISCUSSION:**

We explored the city of Chicago, Illinois to find a best neighborhood in the city where the crime rate is lowest. We worked on Chicago crime data to understand various kinds of crimes in each community area of Chicago and later segregated them based on per capita income. This strategy helped us in selecting community areas with lower crime rate and higher per capita income. Once the community areas were short listed based on the lower crime rate and higher per capita income, consideration of neighborhoods became easier as the number of neighborhoods reduced. We further shortlisted the neighborhoods based on common venues, to choose a neighborhood which best suits the problem.

## **CONCLUSION:**

The objective of this project was to find a safe neighborhood in the city of Chicago, Illinois based on low crime rate and high per capita income. This was achieved by analyzing the Chicago crime data to find a safe community area and by analyzing the Chicago Census Data for determining the per capita income of each of the community areas. After the selection of the community area based on the two factors: lower crime rate and higher per capita income, it was vital to choose a neighborhood where an individual can look for a place to live. We accomplished this by grouping the neighborhoods into clusters to assist an individual with finding a safe place by providing them with relevant data about total crime cases, per capita income, and common venues around a given neighborhood.

## **APPENDIX:**

Link 1: [https://en.wikipedia.org/wiki/List\\_of\\_neighborhoods\\_in\\_Chicago](https://en.wikipedia.org/wiki/List_of_neighborhoods_in_Chicago)

Link 2: <https://data.cityofchicago.org/api/views/qzdf-xmn8/rows.csv?accessType=DOWNLOAD>

Link 3: <https://ibm.box.com/shared/static/05c3415cbfbtfnr2fx4atenb2sd361ze.csv>