

# CAPSTONE PROJECT - THE BATTLE OF NEIGHBORHOODS (WEEK 2)

Favorable community in crime city of Chicago

## **PROBLEM BACKGROUND:**

- Chicago had a murder rate of 18.5 per 100,000 residents in 2012, ranking 16th among US cities with 100,000 people or more. This was higher than in New York City and Los Angeles, the two largest cities in the United States, which have lower murder rates and lower total homicides.
- Department reported that the city experienced a dramatic rise in gun violence, with 4,331 shooting victims.
- The department also reported 762 murders in Chicago for the year 2016, a total that marked a 62.79% increase in homicides from 2015.

## **PROBLEM DESCRIPTION:**

The life in Chicago is like Russian roulette. And the main question is:

How citizens or immigrants can survive in this city?

The first thing that comes to mind is to choose most safe area to live.

# DATA

Country: USA

State: Illinois

City: Chicago

Population: 2.74936 Million

Number of community areas: 77

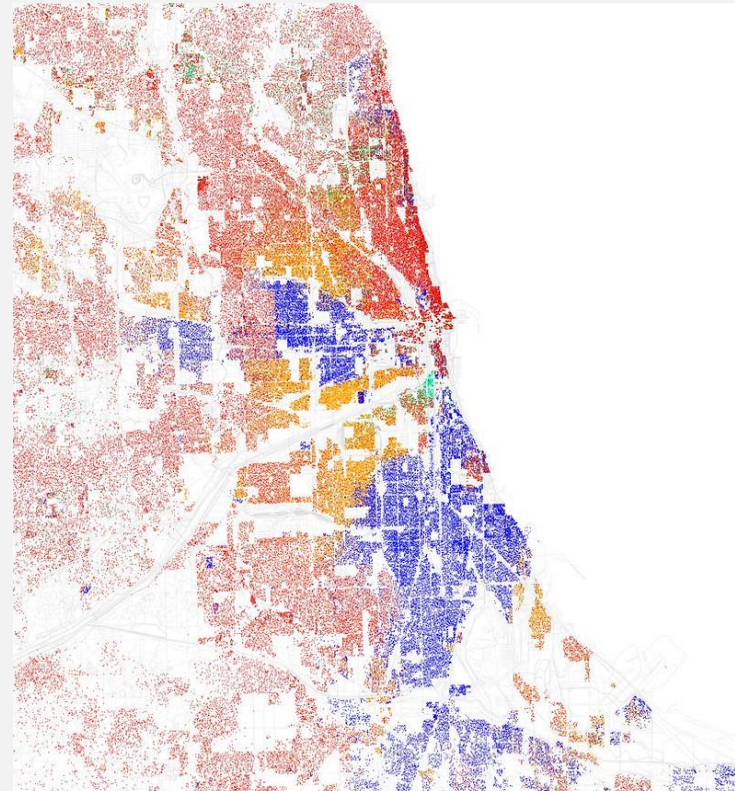
**Data 1:** Crime statistic. It is large (about 7 millions rows) dataset that consist of all kinds of crime since 2001 year till our days.

**Data 2:** Geo data and List of Community Areas. This datasets consist of 77 community areas, geodata and their names.

**Data 3:** Geo data from Foursquare API to explore favorable community areas in the city of Chicago!

# METHODOLOGY

- **Data pre-processing.**
- **Data visualization.**
- **Approach.**
- **Community coordinates.**
- **Community analyzing.**



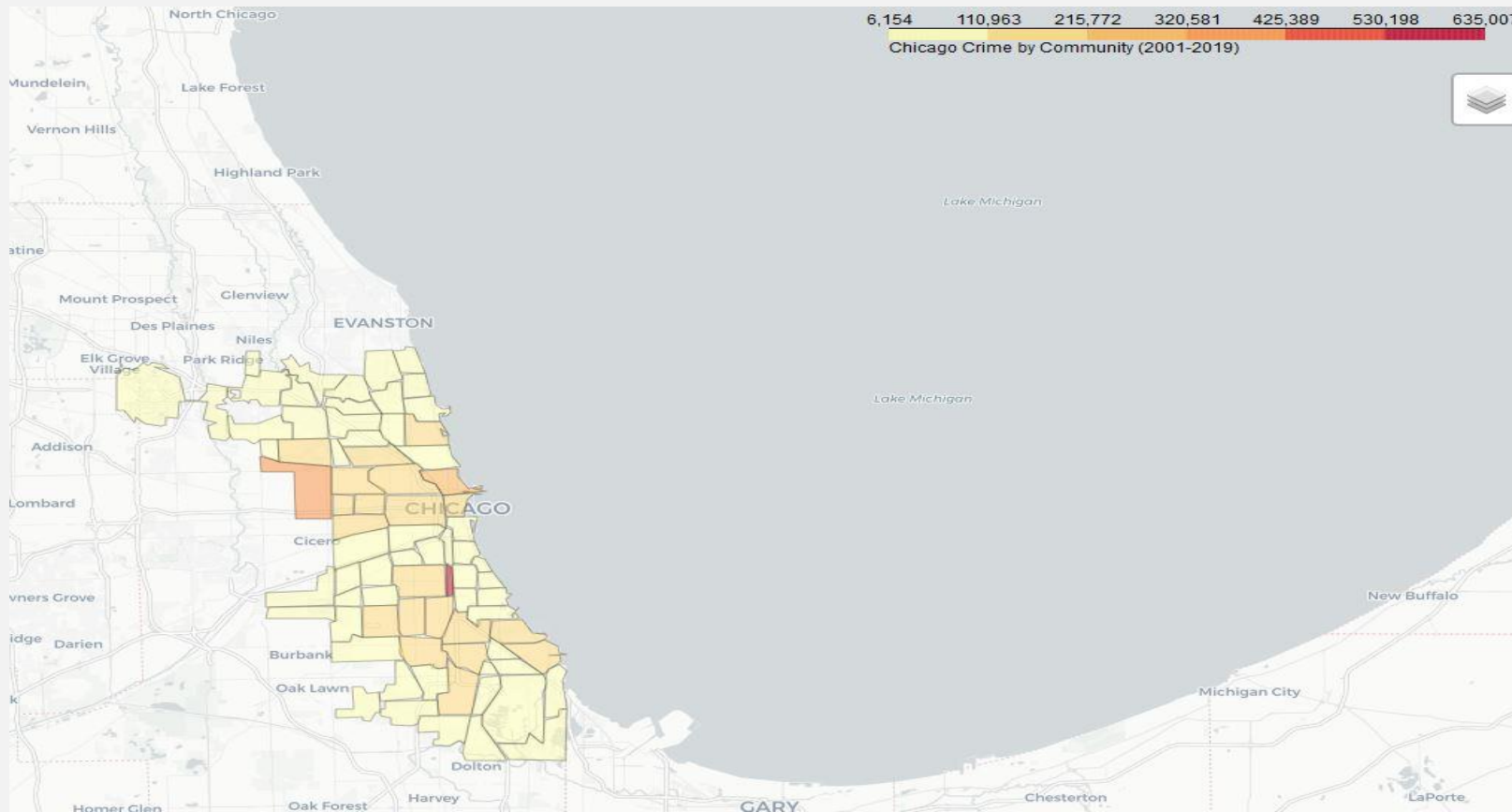
# ANALYSIS WITH PYTHON

```
1 import os
2 import folium
3 import numpy as np
4 import pandas as pd
5
6 pd.set_option('display.max_columns', None)
7 pd.set_option('display.max_rows', None)
8
9 community_Chicago = pd.read_csv('CommAreas.csv')
10 crimes_chicago = pd.read_csv('Crimes_2001_to_present.csv')
11 geo_Chicago = pd.read_json('https://data.cityofchicago.org/resource/igwz-8jzy.json')
12 # -----
13 # get Chicago community areas geo data
14 areas = geo_Chicago
15 # drop unused columns
16 areas = areas.drop(['area', 'area_num_1', 'comarea', 'comarea_id',
17                    'shape_area', 'shape_len', 'perimeter'], axis=1)
18 # rename area_numbe and convert it to int type
19 # for crime dataframe merge on community area # later
20 areas = areas.rename(columns={'area_numbe': 'Community Area'})
21 areas['Community Area'] = areas['Community Area'].astype(np.int64)
22 # -----
23 # load crimes parquet data into dask df
24 crimes = crimes_chicago
25 # drop unused columns and let me take information for last 5 years
26 crimes = crimes.drop(['ID', 'Case Number', 'Date', 'Block', 'IUCR',
27                       'Description', 'Location Description', 'Arrest', 'Domestic', 'Beat',
28                       'District', 'Ward', 'FBI Code', 'X Coordinate',
29                       'Y Coordinate', 'Updated On',
30                       'Historical Wards 2003-2015', 'Zip Codes',
31                       'Community Areas', 'Census Tracts', 'Wards', 'Boundaries - ZIP Codes',
```

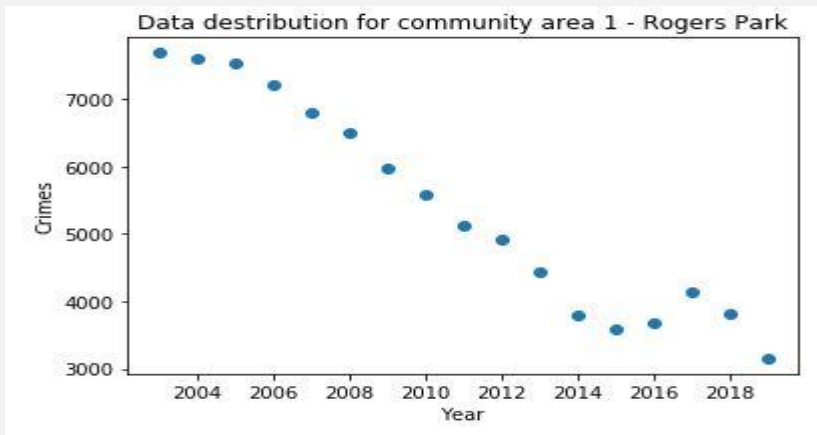
	ID	Case Number	Date	Block	IUCR	Primary Type	Description	Location Description	Arrest	Domestic	Beat	District	Ward	Community Area	FBI Code	X Coordinate	Y Coordinate	Year	Updated
0	11034701	JA366925	01/01/2001 11:00:00 AM	016XX E 86TH PL	1153	DECEPTIVE PRACTICE	FINANCIAL IDENTITY THEFT OVER \$ 300	RESIDENCE	False	False	412	4.0	8.0	45.0	11	NaN	NaN	2001	08/05/2003:50:1
1	11227287	JB147188	10/08/2017 03:00:00 AM	092XX S RACINE AVE	0281	CRIM SEXUAL ASSAULT	NON-AGGRAVATED	RESIDENCE	False	False	2222	22.0	21.0	73.0	02	NaN	NaN	2017	02/11/2003:57:1
2	11227583	JB147595	03/28/2017 02:00:00 PM	026XX W 79TH ST	0620	BURGLARY	UNLAWFUL ENTRY	OTHER	False	False	835	8.0	18.0	70.0	05	NaN	NaN	2017	02/11/2003:57:1
3	11227293	JB147230	09/09/2017 08:17:00 PM	060XX S EBERHART AVE	0810	THEFT	OVER \$500	RESIDENCE	False	False	313	3.0	20.0	42.0	06	NaN	NaN	2017	02/11/2003:57:1
4	11227634	JB147599	08/26/2017 10:00:00 AM	001XX W RANDOLPH ST	0281	CRIM SEXUAL ASSAULT	NON-AGGRAVATED	HOTEL/MOTEL	False	False	122	1.0	42.0	32.0	02	NaN	NaN	2017	02/11/2003:57:1

	community	Total
0	DOUGLAS	71028
1	OAKLAND	14114
2	FULLER PARK	635007
3	GRAND BOULEVARD	88207
4	KENWOOD	36198

# CRIME MAP (2001-2019) WITH FOLIUM.

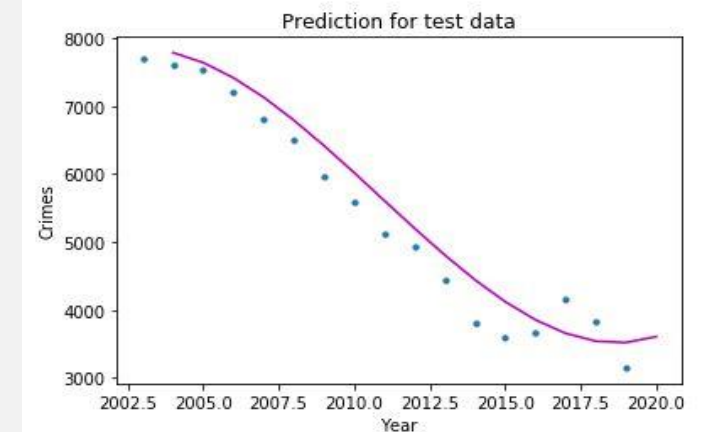
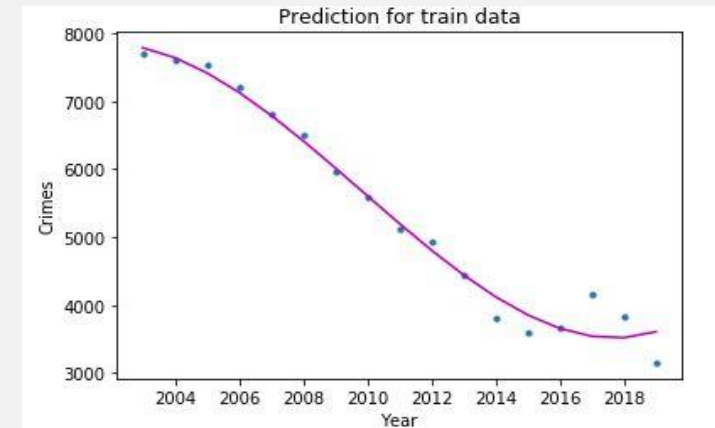


# CRIME PREDICTION WITH POLYNOMIAL REGRESSION



R2 score: 0.9778812360587088  
RMSE: 229.7479808240205

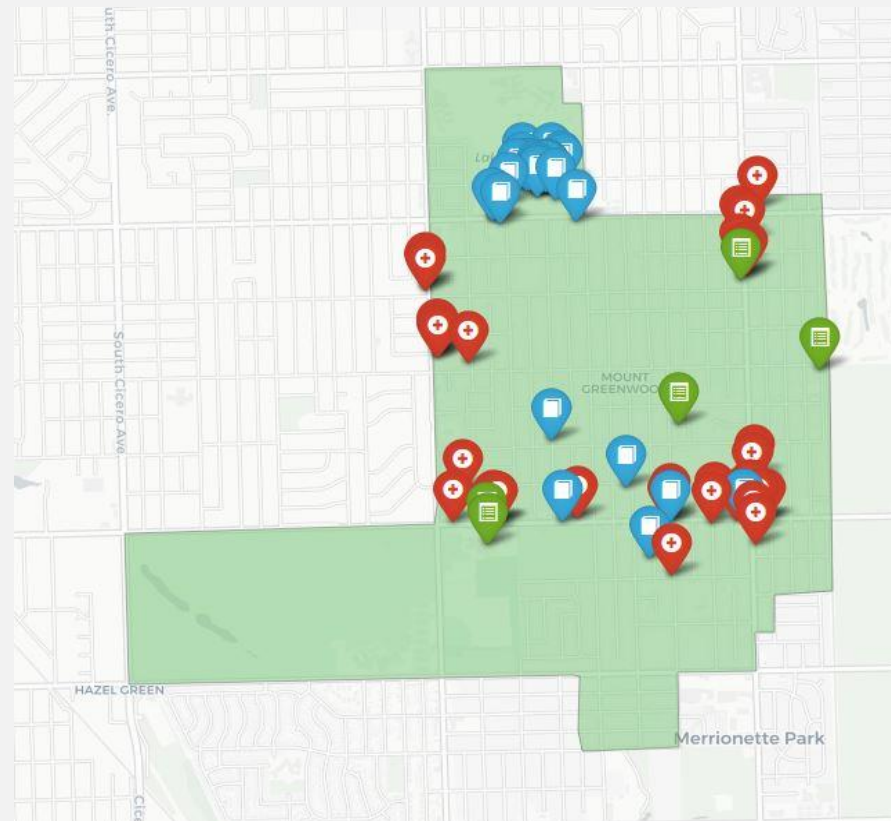
	community	Community Area	Year	Total
76	EDISON PARK	9	2020	233
42	BURNSIDE	47	2020	357
72	MOUNT GREENWOOD	74	2020	435
11	FOREST GLEN	12	2020	459
53	HEGEWISCH	55	2020	472





# CHICAGO MOST FAVORABLE COMMUNITY COORDINATES WITH GEOPY AND FOURSQUARE API

	community	Medicine	EducationH	EducationL
0	EDISON PARK	28	1	14
1	BURNSIDE	38	2	8
2	MOUNT GREENWOOD	35	28	5
3	FOREST GLEN	40	2	4
4	HEGEWISCH	3	1	3



# CONCLUSION

- In data from Foursquare we can see a lot of doctors, dentists clinics in our community.
- It also concerns educational infrastructure

**We answered on the main question and task of our project. Mount Greenwood will be the safest and most lively area in 2020. Main goal is achieved.**