Capstone Project - The Battle of Neighborhoods (Week 2) Favorable community in crime city of Chicago

Introduction

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. However, it was less than in many smaller American cities, including New Orleans, Newark, and Detroit, which had 53 murders per 100,000 residents in 2012. The 2015 year-end crime statistics showed there were 468 murders in Chicago in 2015 compared with 416 the year before, a 12.5% increase, as well as 2,900 shootings—13% more than the year prior, and up 29% since 2013. Chicago had more homicides than any other city in 2015 in total but not on per capita basis, according to the Chicago Tribune.In its annual crime statistics for 2016, the Chicago Police 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. In June 2017, the Chicago Police Department and the Federal ATF announced a new task force, similar to past task forces, to address the flow of illegal guns and repeat offenses with guns.

According to reports in 2013, "most of Chicago's violent crime comes from gangs trying to maintain control of drug-selling territories", and is specifically related to the activities of the Sinaloa Cartel, which is active in several American cities. By 2006, the cartel sought to control most illicit drug sells. Violent crime rates vary significantly by area of the city, with more economically developed areas having low rates, but other sections have much higher rates of crime. In 2013, the violent crime rate was 910 per 100,000 people; the murder rate was 10.4 – while high crime districts saw 38.9, low crime districts saw 2.5 murders per 100,000.

(c) wikipedia - https://en.wikipedia.org/wiki/Chicago#Crime

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. I will work in that way, I should go through the crime statistic and try to give you answers on these questions:

What would be the most safe area in city of Chicago in 2020? Which area is better regarding its residential infrastructure? Residential infrastructure refers to public services and facilities necessary to support residential development including sewer and water lines, waste water treatment plants, waterworks, schools, parks, streets, street lights, and storm water management facilities.

By residential infrastructure, important for family life, we will mean the presence of medical institutions (hospitals, clinics and other) and educational institutions (schools, universities and other)

Target Audience:

Point 1. Citizens who wants to buy a new flat in safe for their lives area. Immigrants.

Point 2. Small or middle business whose main auditory is in the first point. For example it could be companies like: restaurants, cafe, market shops, gyms and others.

Success Criteria:

The success criteria of the project will be a recommendation of most safe area in Chicago in 2020 year.

Data

Country: USA
State: Illinois
City: Chicago
Population: 2.74936 Million
Number of community areas: 77

For our analyzing we will be using below datasets:

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.
You can find it on the Chicago's public web site: https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-present/ijzp-q8t2

Data 2: Geo data and List of Community Areas. This datasets consist of 77 community areas, geodata and their names.
You can find it on the Chicago's public web site: https://data.cityofchicago.org/Facilities-Geographic-Boundaries/Boundaries-Community-Areas-current-/cauq-8yn6

Data 3: Geo data from Foursquare API to explore favorable community areas in the city of Chicago!
To learn more about Foursquare go by link: https://api.foursquare.com

Methodology

This Project will consist of 5 parts.

1. Data pre-processing

We need to download all data what can help us. The output will big dataset with more than 30 millions values. We need to choose data what we will work with. After this step we need to prepare this data frames and choose what libraries we will use in this project.

2. Data visualization.

The second step is data visualization. We need to do it for more understanding of problem. In this project I chose Choropleth map with folium.

3. Approach.

In this section we need to understand what model we need and how can we give the answer on the main question. I chose to use prediction model with "Polynomial Regression".

4. Community coordinates.

The fourth step will be location view, we need to find all interesting venues in our communities. For this step I chose to use geopy and Foursquare API.

5. Community analyzing.

At the last step we need to determine most favorable community regarding to less crime and infrastructure.

Analysis with Python

Data pre-processing

Crime statistic. It is large (about 7 millions rows) dataset that consist of all kinds of crime since 2001 year till our days. So, we need do prepare data.

	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	Updat
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	08/05/20 03:50:
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	02/11/20 03:57:
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	02/11/20 03:57:
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	02/11/20 03:57:
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	02/11/20 03:57:

1. Use mean method to NA values
2. Drop unused columns
3. Merge with geo data to fill coordinates (we will use it in another dataset)
4. Convert values

The result of this work is:

	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.
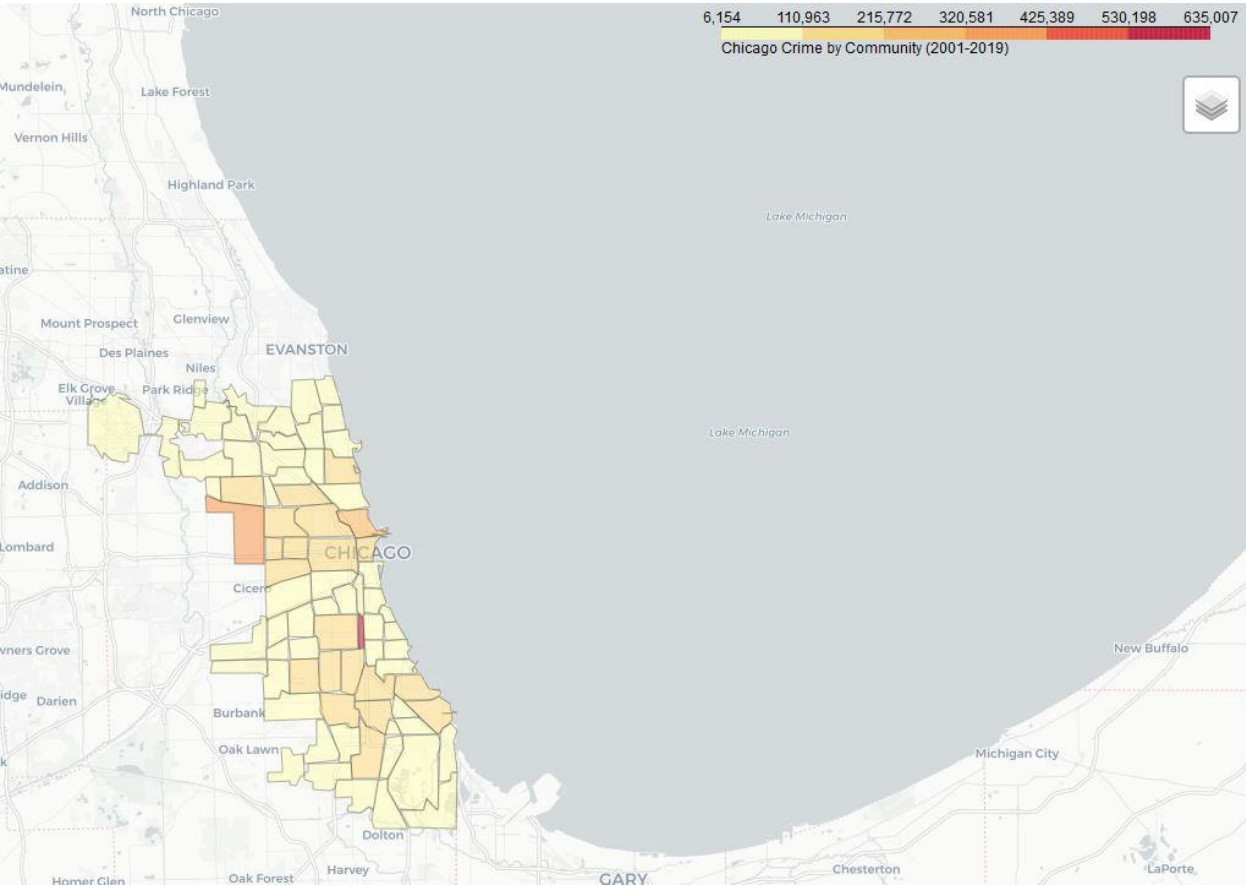
Crime prediction with Polynomial Regression (2020)

I try to understand our data and plot the following graphic.

Data distribution for community area 1 - Rogers Park
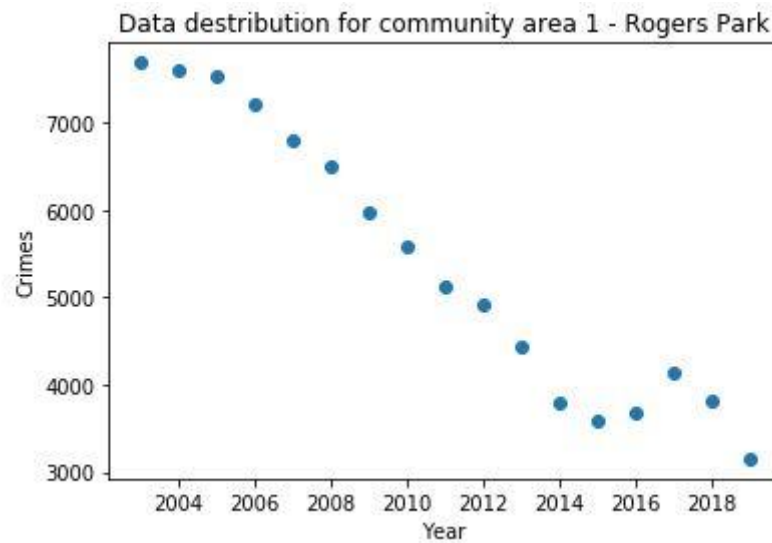
My decision was to use Polynomial Regression to fit our model.

Prediction for train data

Prediction for test data
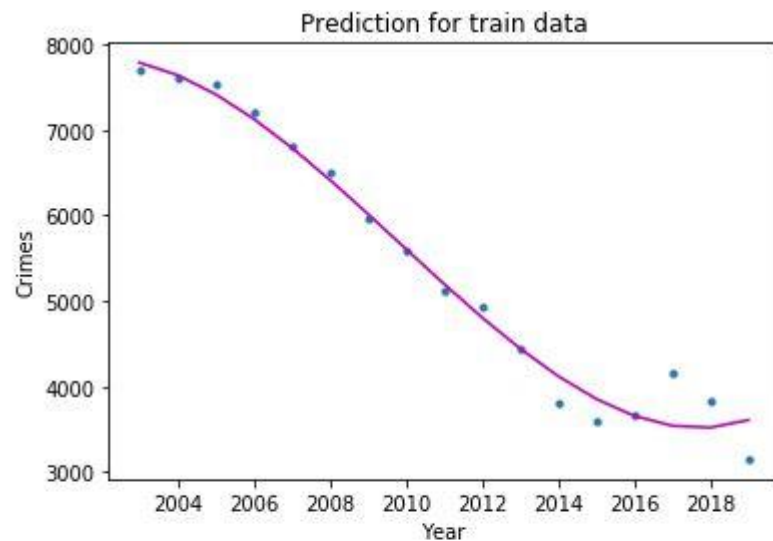
After this I have to estimate accuracy of our model:

```
R2 score:  0.9778812360587088
RMSE:  229.7479808240205
```

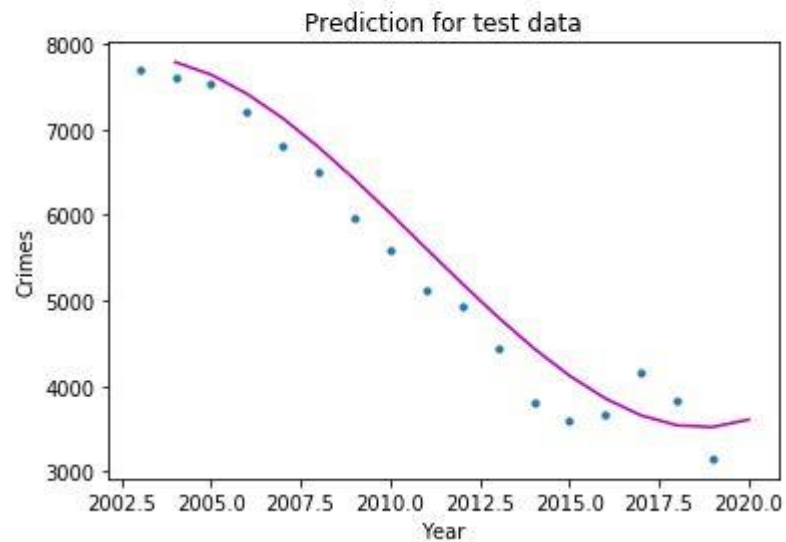In my opinion this is good result. So the next step is to build new data frame with top 5 communities with less crimes in 2020 year.

	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 community coordinates with geopy and Foursquare API

Foursquare City Guide, commonly known as Foursquare, is a local search-and-discovery mobile app developed by Foursquare Labs Inc. The app provides personalized recommendations of places to go near a user's current location based on users' previous browsing history and check-in history.

To find venues coordinates I was connecting to Foursquare API .

Learn more on the Foursquare web site.

As result was this table:

	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

.

Community analyzing.

In my opinion regarding to this table the most favorable community is Mount Greenwood community. Because there is a well-developed medical and social infrastructure.
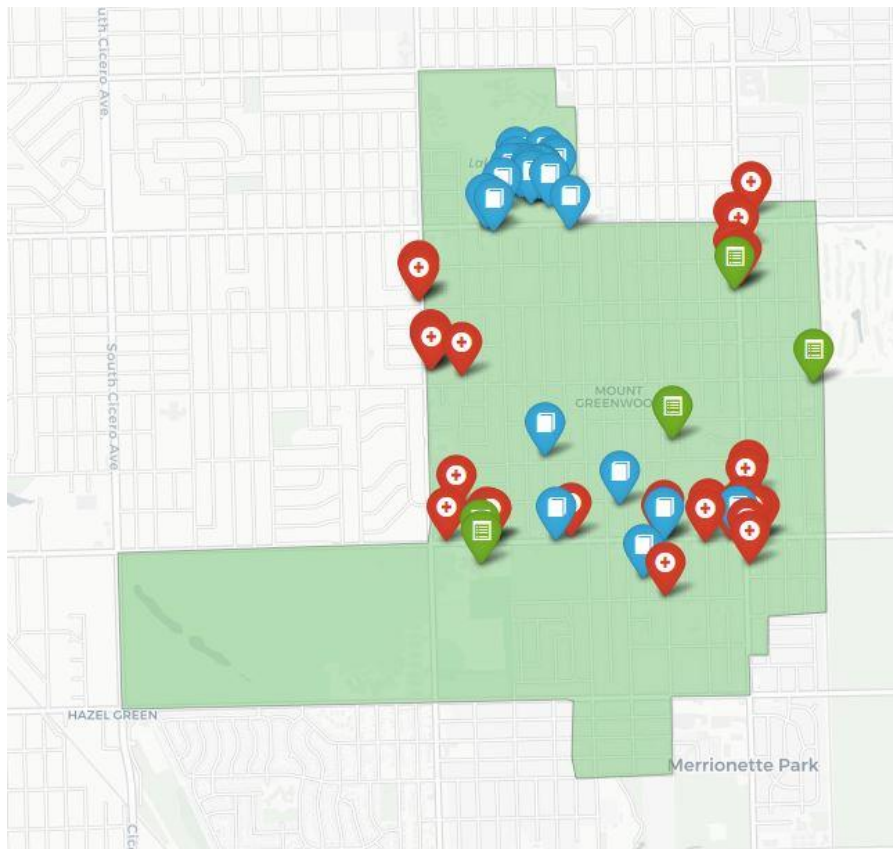
Results and Discussion

Infrastructure.

Medicine and Education.
In data from Foursquare we can see a lot of doctors, dentists clinics in our community.

```
<bound method NDFrame.describe of 0        Davita Dialysis At Mt Greenwood
1                   Dr. Sharon Malinowski
2                        The Healing House
3                          Davita Dialysis
4                    Midwest Express Clinic
5                     Vista Family Medicine
6                        Reed Chiropractic
7                                  DR BUSH
8              Southwest Medical Consultants
9                   Vascular Access Center
10                               Dr. Feria
11                  Advocate Medical Group
12                               South Care
13                 Southcare Family Practice
14                          Dr Thomas Knox
15                           Kathleen Daly
16                         The Chicago FIX
17                       Dr. Bob (Dentist)
18                 Timeless Smiles Dental Pc
19                           Dr Clodfelter
20                        Kathleen Daly DPM
21                         McNicholas Dental
22                    My. Greenwood Dialysis
23                               Little You
24               Chicagoland Foot And Ankle
25                           Baptist Dental
26                          Family Eye Care
27            Chicago Foot & Ankle Specialists
28           Midwest Chiropractic Care Center
29                 Salerno Family Dentistry
30                   Southwest Dental Center
31          Gellman Family Therapy Associates
32                     Chris Nikolovski, DDS
33            Greenwood Chiropractic Wellness
34                            Dr Nora Bellosa
Name: 3, dtype: object>
```

As you have noticed, it also concerns educational infrastructure.

```
0                    Wright College - S320          0                  Ebinger Elementary School
1                  Chicago State University          1              Montessori Learning Center
2                          Kwame N'Krumah            2                           Ebinger school
3                            Mustard Seed            3                    Expert Driving School
4                          Macro Economics           4                          Childrens Campus
5       Saint Xavier University O'Grady Center       5                         Stock Pre School
6            Saint Xavier University (111th)          6                 St. Paul of the Cross School
7                    Warde Academic Center           7                        St. Juliana School
8                              Philosophy            8                      10th Planet Jiu Jitsu
9                            Stump Library           9              Ravenswood Ridge Network Office
10            St. Xavier University (Graham)        10   Lutheran Unity School Association East & West ...
11   Saint Xavier University (Warde Academic Center) 11                  Edison Park Elementary
12       Saint Xavier University (O'Grady Center)    12                              Stock School
13       Saint Xavier University Health Center       13     CPS North/Northwest Network Office
14                             SXU Diner             14                  Harold Washington School
15   Saint Xavier University (Andrew Conference Cen... 15                         Camp Invention
16                      La Oficina de Jose           16              Burnside Scholastic Academy
17                     SXU Information Desk          17                     Making A Difference
18                        SXU Old Gym/RPS            18   Mary E. McDowell Elementary School Of Internat...
19            SXU Student Organization Office        19              Jenkins Institute of Love
20                       Campus Bookstore            20                    Arthur Ashe Elementary
21                  Graham School of Management      21       Chatham Academy Alternative Charter School
22        Saint Xavier University (Alumni House)     22           Mount Greenwood Elementary School
23                        SXU Math Offices           23           Chicago Agricultural High School
24                         The Meadowlark            24             Keller Regional Gifted Center
25                       SXU Shannon Center          25                   Garden Center Services
26       Saint Xavier University (Shannon Center)    26   Chicago High School of Agricultural Sciences -...
27                       SXU Student Lounge          27                  Midwest Fencing Academy
28                           Chicago Pulse           28               Queen of All Saints School
29          Saint Xavier University S108 Lab         29                                  QAS gym
30                  St. Xavier University            30                 Edgebrook Elementary School
31               Sauganash Elementary School        31             Henry Clay Elementary School
32                                Spokane            32                   St. Florian Gymnasium
33                              St Florian           33             Virgil Grissom Elementary School
```

The only thing I want change is to optimize our model so that the system works more accurately.

Conclusion

In my opinion, we answered the main question and task of our project. Mount Greenwood will be the safest and most lively area in 2020. Main goal is achieved.