Capstone Project - The Battle of Neighborhoods

1. Introduction/Business Problem

Chicago saw a major rise in violent crime starting in the late 1960s. Murders in the city peaked in 1974, with 970 homicides when the city's population was over three million, resulting in a murder rate of around 29 per 100,000, and again in 1992, with 943 murders when the city had fewer than three million people, resulting in a murder rate of 34 murders per 100,000 citizen. Chicago reduced a murder rate in 2018. The municipal authorities want to keep this trend. They outsourced to prepare a report which is to identify crime patterns in Chicago.

The report should provide answers to below questions and explain the difference between overall crimes and homicides including the following aspects:

- When crimes are committed?
- Which community areas are the most dangerous? (number of crimes per 1000 inhabitants)
- Are police departments properly located? (Foursquare location)

2. Data

The sources of data are official websites:

https://data.cityofchicago.org – juxtaposition of crimes in 2018

 https://www.chicago.gov – Chicago census data

Data were loaded from csv files using pandas package. Both dataframes were combined and cleansed from missing data. Total number of rows was reduced by 1.5% due to lack of locations.

Finally two tables were created:

- total - containing all crimes

	Case Number	Season	Part of the day	Primary Type	Description	Block	Location Description	Arrest	Beat	District	Community Area	Community Area Description	Community Area Population	Latitude	Longitude
0	JC104662	Winter	Night	CRIMINAL DAMAGE	TO VEHICLE	112XX S SACRAMENTO AVE	STREET	False	2211	22	74	Mount Greenwood	19093	41.689079	-87.696064
1	JC110056	Winter	Night	DECEPTIVE PRACTICE	FINANCIAL IDENTITY THEFT OVER $ 300	013XX W 72ND ST	OTHER	False	734	7	67	West Englewood	35505	41.763181	-87.657709
2	JC100043	Winter	Night	CRIMINAL DAMAGE	TO PROPERTY	084XX S SANGAMON ST	APARTMENT	False	613	6	71	Auburn Gresham	48743	41.740521	-87.647391
3	JC100006	Winter	Night	BATTERY	AGG: HANDS/FIST/FEET NO/MINOR INJURY	018XX S ALLPORT ST	OTHER	True	1233	12	31	Lower West Side	35769	41.857068	-87.657625
4	JC100031	Winter	Night	BATTERY	DOMESTIC BATTERY SIMPLE	078XX S SANGAMON ST	APARTMENT	False	621	6	71	Auburn Gresham	48743	41.751914	-87.647717

Graph1. Initial data for total crimes (5 rows from 263'425)

- homicide - containing this kind of crime

	Case Number	Season	Part of the day	Primary Type	Description	Block	Location Description	Arrest	Beat	District	Community Area	Community Area Description	Community Area Population	Latitude	Longitude
315	JB573872	Winter	Afternoon	HOMICIDE	FIRST DEGREE MURDER	002XX W 95TH ST	STREET	False	634	6	49	Roseland	44619	41.721658	-87.628972
428	JB573562	Winter	Morning	HOMICIDE	FIRST DEGREE MURDER	003XX E 53RD ST	STREET	True	231	2	40	Washington Park	11717	41.798450	-87.616858
1098	JB572372	Winter	Morning	HOMICIDE	FIRST DEGREE MURDER	066XX S LAFLIN ST	ABANDONED BUILDING	True	725	7	67	West Englewood	35505	41.773350	-87.661613
3523	JB567847	Winter	Afternoon	HOMICIDE	FIRST DEGREE MURDER	057XX W 64TH PL	HOUSE	True	812	8	64	Clearing	23139	41.775264	-87.765095
4047	JB566817	Winter	Afternoon	HOMICIDE	FIRST DEGREE MURDER	024XX S WHIPPLE ST	STREET	False	1033	10	30	South Lawndale	79288	41.847238	-87.701441

Graph2. Initial data for homicides (5 rows from 590)

Both tables consist of 15 columns which determine:

- time of crime - Season, Part of the day
- place of crime - Block, Location description, Longitude, Latitude, Community Area, Community Area Description,
- crime classification - Case number, Primary Type, Arrest
- police assignment - Beat, District

3. Methodology

In order to check if there are any time patterns for when crimes were committed, data were grouped based on season and part of the day.
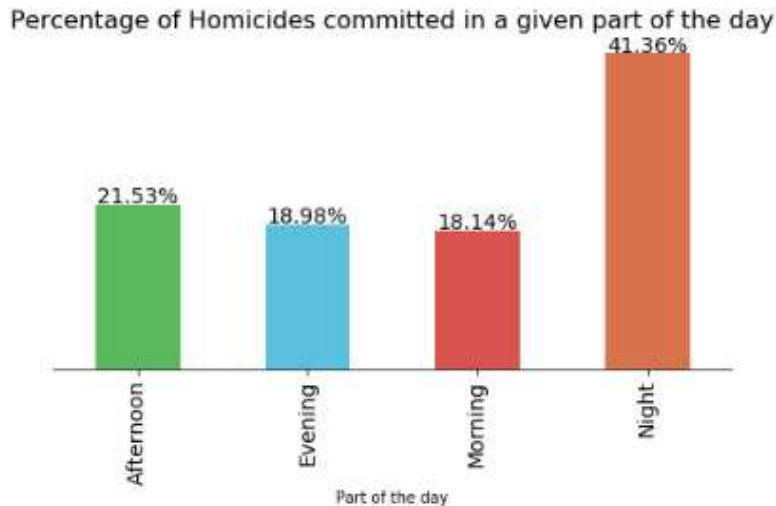
Season	Part of the day	Number of crimes	Percentage
Fall	Afternoon	16451	6.25
	Evening	14548	5.52
	Morning	15188	5.77
	Night	18750	7.12
	Total	64937	24.66
Spring	Afternoon	17321	6.58
	Evening	14637	5.56
	Morning	15649	5.94
	Night	18812	7.14
	Total	66419	25.22
Summer	Afternoon	18532	7.04
	Evening	16297	6.19
	Morning	16242	6.17
	Night	22601	8.58
	Total	73672	27.98
Winter	Afternoon	15281	5.8
	Evening	13275	5.04
	Morning	13614	5.17
	Night	16227	6.16
	Total	58397	22.17

Graph3. Grouped Total Crimes by time

Season	Part of the day	Number of crimes	Percentage
Fall	Afternoon	38	6.44
	Evening	28	4.75
	Morning	30	5.08
	Night	59	10
	Total	155	26.27
Spring	Afternoon	24	4.07
	Evening	28	4.75
	Morning	16	2.71
	Night	61	10.34
	Total	129	21.87
Summer	Afternoon	38	6.44
	Evening	36	6.1
	Morning	34	5.76
	Night	69	11.69
	Total	177	29.99
Winter	Afternoon	27	4.58
	Evening	20	3.39
	Morning	27	4.58
	Night	55	9.32
	Total	129	21.87

Graph4. Grouped Homicides by time

Graph5. Percentage of Homicides committed in a given part of the day

The geographical classification was defined using a crime rate which was calculated as the number of crimes per 1000 inhabitants. The main indicator was determined as a crime rate normalized which is the mean of two normalized vectors: number of homicides and crime rate to avoid the overrepresentation of sparsely populated community areas.

	Community Area	Community Area Description	Community Area Population	Number of Homicides	Crime Rate	Crime Rate Normalized	Postion in Total Crimes
1	68	Englewood	3654	33	9.03	0.79	3
2	27	East Garfield Park	2567	22	8.57	0.67	2
3	25	Austin	98514	56	0.57	0.53	4
4	29	North Lawndale	35912	32	0.89	0.33	6
5	26	West Garfield Park	18001	26	1.44	0.31	15
6	53	West Pullman	29651	26	0.88	0.27	25
7	43	South Shore	49767	25	0.50	0.24	8
8	61	New City	44377	23	0.52	0.23	27
9	67	West Englewood	35505	23	0.65	0.23	11
10	71	Auburn Gresham	48743	23	0.47	0.22	12
11	54	Riverdale	6482	13	2.01	0.22	38
12	49	Roseland	44619	22	0.49	0.22	14
13	45	Avalon Park	1185	4	3.38	0.21	10
14	23	Humboldt Park	56323	21	0.37	0.20	9
15	66	Chicago Lawn	55628	20	0.36	0.19	19

Graph6. Homicide rate for community areas (15 rows from 77)

The next step was to cluster community areas based on a crime rate normalized. Four clusters were assigned based on k-Means method. Clusters were redesigned to allocate the high number to the most dangerous place.

	Community Area	Community Area Description	Community Area Population	Number of Homicides	Crime Rate	Crime Rate Normalized	Postion in Total Crimes	Community Area Homicide Cluster
1	68	Englewood	3654	33	9.03	0.79	3	3
2	27	East Garfield Park	2567	22	8.57	0.67	2	3
3	25	Austin	98514	56	0.57	0.53	4	3
4	29	North Lawndale	35912	32	0.89	0.33	6	2
5	26	West Garfield Park	18001	26	1.44	0.31	15	2
6	53	West Pullman	29651	26	0.88	0.27	25	2
7	43	South Shore	49767	25	0.50	0.24	8	2
8	61	New City	44377	23	0.52	0.23	27	2
9	67	West Englewood	35505	23	0.65	0.23	11	2
10	71	Auburn Gresham	48743	23	0.47	0.22	12	2
11	54	Riverdale	6482	13	2.01	0.22	38	2
12	49	Roseland	44619	22	0.49	0.22	14	2
13	45	Avalon Park	1185	4	3.38	0.21	10	2
14	23	Humboldt Park	56323	21	0.37	0.20	9	2
15	66	Chicago Lawn	55628	20	0.36	0.19	19	2

Graph7. Community areas clustered based on crime rate (15 rows from 77)

The following issue was to analyze the neighborhood of crimes. Consequently, the function was built to catch the location of all the police departments within a specific crime site. The search query was "Police". The location was gained using Foursquare API.

	Case Number	Latitude	Longitude	Police Department	Latitude Police	Longitude Police
0	JB573872	41.721658	-87.628972	Chicago Police Department - 4th District	41.707732	-87.585895
1	JB573872	41.721658	-87.628972	Chicago Police Station 3rd District	41.766998	-87.605839
2	JB573872	41.721658	-87.628972	Chicago Police 22nd District	41.691507	-87.668432
3	JB573872	41.721658	-87.628972	Chicago Police Department - District 6	41.751796	-87.644123
4	JB573872	41.721658	-87.628972	Chicago Police Department 5th District	41.692799	-87.604573

Graph8. Police Department location (5 rows from 9727)

The distance between a crime site and the police station was calculated based on geopy library.

	Case Number	Latitude	Longitude	Police Department	Latitude Police	Longitude Police	Distance [m]
0	JB573872	41.721658	-87.628972	Chicago Police Department - 4th District	41.707732	-87.585895	3904.0
1	JB573872	41.721658	-87.628972	Chicago Police Station 3rd District	41.766998	-87.605839	5391.0
2	JB573872	41.721658	-87.628972	Chicago Police 22nd District	41.691507	-87.668432	4690.0
3	JB573872	41.721658	-87.628972	Chicago Police Department - District 6	41.751796	-87.644123	3577.0
4	JB573872	41.721658	-87.628972	Chicago Police Department 5th District	41.692799	-87.604573	3794.0

Graph9. Distance between crime site and police station (5 rows from 9727)

Clusters were assigned to crime sites taking into consideration the nearest police station. As a method of clustering k-Means was used.

	Case Number	Distance [m]	Police Department Cluster
0	D439410	3204.0	3
1	F780368	778.0	0
2	HS260050	2034.0	1
3	HW249839	759.0	0
4	HY370531	535.0	0

Graph10. Cluster based on distance to police department (5 rows from 567)

The next step was to reassign the cluster basing on distance so as the longest length corresponded to the furthest cluster.

	min_Distance	max_Distance	Police Department Cluster	Police Department Cluster new
0	28.0	941.0	0	0
1	955.0	1646.0	2	1
2	1665.0	2412.0	1	2
3	2460.0	4123.0	3	3

Graph11. Cluster based on nearest police department to a crime

The final table was created thanks to merging all the necessary dataframes.

|	Case Number	Season	Part of the day	Primary Type	Description	Block	Location Description	Arrest	Beat	District	Community Area	Community Area Description	Community Area Population	Latitude	Longitude	Community Area Homicide Cluster	Distance [m]	Police Department Cluster
0	J8573872	Winter	Afternoon	HOMICIDE	FIRST DEGREE MURDER	002XX W 95TH ST	STREET	False	634	6	49	Roseland	44619	41.721658	-87.628972	2	2685.0	3
1	J8573562	Winter	Morning	HOMICIDE	FIRST DEGREE MURDER	003XX E 53RD ST	STREET	True	231	2	40	Washington Park	11717	41.798450	-87.616858	1	1108.0	1
2	J8572372	Winter	Morning	HOMICIDE	FIRST DEGREE MURDER	066XX S LAFLIN ST	ABANDONED BUILDING	True	725	7	67	West Englewood	35505	41.773350	-87.661613	2	700.0	0
3	J8567847	Winter	Afternoon	HOMICIDE	FIRST DEGREE MURDER	057XX W 64TH PL	HOUSE	True	812	8	64	Clearing	23139	41.775264	-87.765095	0	486.0	0
4	J8566817	Winter	Afternoon	HOMICIDE	FIRST DEGREE MURDER	024XX S WHIPPLE ST	STREET	False	1033	10	30	South Lawndale	79288	41.847238	-87.701441	1	1197.0	1

Graph12. Final table (5 rows from 590)

Due to the fact that json file contained only zip -code, zip codes were assigned to community areas.

	Zip code	Community Area	Community Area Description	Community Area Homicide Cluster
105	60624	27	East Garfield Park	3
198	60620	72	Beverly	3
97	60624	25	Austin	3
98	60639	25	Austin	3
100	60651	25	Austin	3

Graph13. Zip code assigned to community area (5 rows from 210)

Derived from the final table, a map was created. The most dangerous community areas are colored in red. The map contains all sites of homicides which are marked: as green in case where the police station was located between 28 and 941 meters from the crime site, in yellow - between 955 and 1646 m, orange markers were found between 1665 and 2412 m. The red ones were assigned the furthest located.
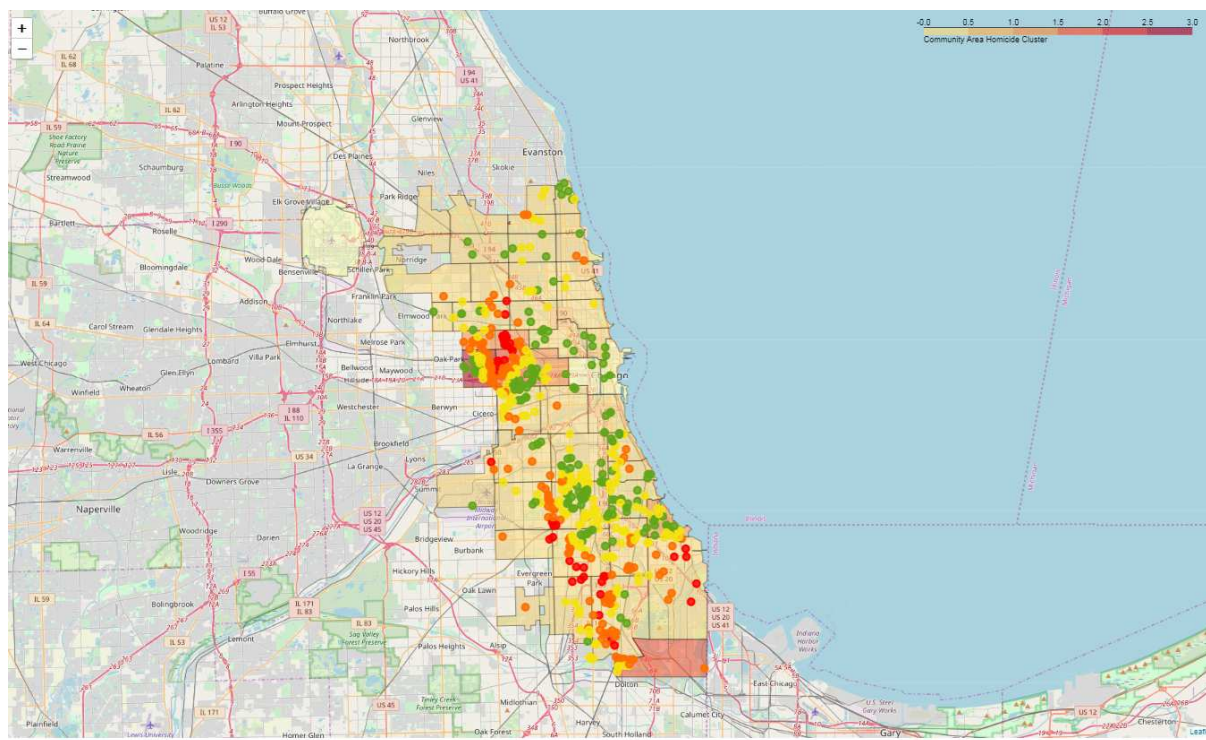
Graph14. Final Map with all homicide sites.

4. Results

There was found a slight dependence of the season when the crime was committed (graph 1 and 2). Both homicides and other types of crimes are a bit more often committed in summer comparing to winter. The significant relation was defined in case of time of the crime. Murders are committed mainly during night (graph 3). There was found such a correlation for all crimes.

The most dangerous community areas were defined based on crime rate which included population and number of crimes. Interestingly, the level of danger for community areas is different for homicides and total crimes. "Near North Side" which is at first place for all crimes is out of the top 15 in case of homicides. The second place at both lists is for "East Garfield Park". The highest homicide rate is in Englewood which is 3rd in total crimes. The community areas were clustered into 4 groups (low risk, medium risk, high risk and very high risk). The most hazardous areas are middle western parts of the city and on the most southern one (graph 14). Each homicide site was assigned to one of four clusters based on distance between site and police station

(short, medium, distant, very distant). It was found that there was significant number of distant sites at the high risk area at the middle western part of the city. The number of cases based on distance from police station and grouped by community area category are shown in the below table.

Community Area Homicide Cluster	Police Department Cluster	Case Number
0	0	33
	1	28
	2	24
	3	8
1	0	50
	1	50
	2	20
	3	8
2	0	55
	1	96
	2	78
	3	29
3	0	34
	1	47
	2	30

Graph15. The number of cases based on distance from police station grouped by community area category.

5. Discussion

The recommendation is to increase the number of police patrols at night in 70 specific blocks defined by below addresses. The blocks were selected according to the following rules:

- 27 out of 70: community areas cluster-2 and police department cluster-3

    002XX W 95TH ST , 012XX N KILDARE AVE , 008XX N HARDING AVE , 119XX S LA SALLE ST , 111XX S NORMAL AVE , 007XX N SPRINGFIELD AVE , 042XX W AUGUSTA BLVD , 045XX W ADAMS ST , 044XX W WEST END AVE , 073XX S CLAREMONT AVE ,   023XX W 72ND ST , 097XX S HARVARD AVE , 045XX W LAKE ST , 001XX E 124TH ST , 012XX N KEELER AVE , 122XX S STATE ST , 041XX W POTOMAC AVE , 009XX N KEDVALE AVE , 088XX S PAULINA ST , 042XX W CARROLL AVE , 072XX S WESTERN AVE , 045XX W WEST END AVE , 013XX N KILDARE AVE , 042XX W HADDON AVE , 010XX N PULASKI RD ,   073XX S BELL AVE , 042XX W HIRSCH

- 27 out of 70: community areas cluster-3 and police department cluster-2

002XX W 95TH ST , 012XX N KILDARE AVE , 008XX N HARDING AVE , 119XX S LA SALLE ST , 111XX S NORMAL AVE , 007XX N SPRINGFIELD AVE , 042XX W AUGUSTA BLVD , 045XX W ADAMS ST , 044XX W WEST END AVE , 073XX S CLAREMONT AVE , 023XX W 72ND ST , 097XX S HARVARD AVE , 045XX W LAKE ST , 001XX E 124TH ST , 012XX N KEELER AVE , 122XX S STATE ST , 041XX W POTOMAC AVE , 009XX N KEDVALE AVE , 088XX S PAULINA ST , 042XX W CARROLL AVE , 072XX S WESTERN AVE , 045XX W WEST END AVE , 013XX N KILDARE AVE , 042XX W HADDON AVE , 010XX N PULASKI RD , 073XX S BELL AVE , 042XX W HIRSCH ST

- 16 out of 70: police department cluster-3 with the high density

012XX N KILDARE AVE , 008XX N HARDING AVE , 007XX N SPRINGFIELD AVE , 042XX W AUGUSTA BLVD , 012XX N KEELER AVE , 041XX W POTOMAC AVE , 009XX N KEDVALE AVE , 013XX N KILDARE AVE , 042XX W HADDON AVE , 010XX N PULASKI RD , 042XX W HIRSCH ST , 014XX W 91ST ST , 012XX W 96TH ST , 102XX S EGGLESTON AVE , 091XX S ABERDEEN ST , 095XX S VINCENNES AVE

6.   Conclusion

Chicago is a city with high homicide rate. In order to decrease the number of serious crimes it is recommended to organize sustained police patrols at given addresses.
It is justified to build a new police station in the middle western part of the city.