

IDS 462

Chicago Crime Data Analysis (2015-2016)

Group 1

Aleksandra Zdzenicka, Gurneha Naggi, Aditya Prithvinath,
Palak Vora, Danqing Zhu, Charlie Pranckus



BUSINESS PROBLEM

The most general objective of this project is to conduct an analysis of data that would help us come up with possible solutions to help make the city of Chicago a safer place overall. Chicago Police Department (CPD) has limited law enforcement resources (police officers, police cars etc.). With the crime in Chicago at an all-time high, with reports of people being killed every day, we are trying to determine how to best allocate these resources (where and when to dispatch officers) as well as enable the officials with specialised training according to the most prevalent crime, in order to make Chicago a safer place.

DATA

In order to accomplish this objective, we researched and analyzed the crime from the City of Chicago Data Portal. Direct source link to the data we are discussing can be found under the "Sources" part of this report. We have extracted the data from 1st october 2015 to 1st October 2016 for our analysis into a Microsoft Office Excel file. The file consists of data such as the date and time of the crime occurrence, the location, the type of crime that took place, as well as some less relevant statistics such as the FBI code for the specific misconducts that took place. We will mainly focus on: the primary types of crimes, the specific location of the crime and the time and date of the incidents, arrests etc.

SOFTWARE

The software we used for analyzing our data were Microsoft Office Excel (for extracting the data), and the SAS 9.4 Software(SAS programming language) for cleaning up the data and for the analysis itself.

QUESTIONS FOR ANALYSIS

In order to effectively answer our business problem, we needed to ask the right questions that would help us determine ways in which the Chicago Police Department could allocate their resources most effectively and efficiently. These questions were as follows:

1. Which neighborhoods in Chicago are most dangerous?
2. When are crimes most likely to occur according to month, day of week, time of day?
3. Where do most crimes occur (street, household)?
4. What types of crimes result in an arrest and when do most arrests occur?
5. Which are the top crimes that occur in Chicago?

We needed to answer these questions in order to figure out our Business problem : how to most effectively and efficiently allocate the limited CPD resources. We wanted to know where and when to dispatch the largest number of officers. We also thought that since one of the main resources the CPD has is their money, we wanted to help them figure out a way to best allocate it. We wanted to find out what areas to train the officers in. We believe it would be helpful for the CPD to know which training programs are going to be beneficial based on the crimes the police officers deal with the most often. In order to address these issues we asked:

DATA PREPARATION

Prior to starting our analysis, we needed clean up the data in order to run our designated SAS PROCs without any errors.

- The first step of our data preparation was procuring the dataset from the City of Chicago Data Portal. With the help of the portal's API, we filtered the data prior to export. For our analysis we have considered only the past 1 year's data i.e from Oct 1st 2015 to 1st Oct. 2016.
- In the second step the data was imported into the SAS environment using **Proc Import** which validated the CSV with the original data for its format. We also included the delimiter-sensitive-data (DSD) option in the Proc Import statement step to remove the errors caused by missing data. We also checked the contents of the dataset using proc contents to check the data imported against that present in the CSV file.
- In the Third step we also checked and ensured that there no missing values for character variables. We dropped columns which were irrelevant for our business case analysis. For example: X Coordinate and Y-Coordinate, Year, and Updated_On.
- We also created new columns extracted from DATETIME variable namely: Hour, Month Date ,Weekday, Day_of_month and we converted the **Arrest** and **Domestic** variables to binary. Here are screenshots to show our data variables before cleaning and after cleaning:

1.PRIOR

2. AFTER

Alphabetic List of Variables and Attributes					
#	Variable	Type	Len	Format	Informat
9	Arrest	Char	5	\$5.	\$5.
11	Beat	Num	8	BEST12.	BEST32.
4	Block	Char	21	\$21.	\$21.
2	Case_Number	Char	8	\$8.	\$8.
14	Community_Area	Num	8	BEST12.	BEST32.
3	Date	Num	8	DATETIME.	ANYDTDTM40.
7	Description	Char	35	\$35.	\$35.
12	District	Num	8	BEST12.	BEST32.
10	Domestic	Char	5	\$5.	\$5.
15	FBI_Code	Char	3	\$3.	\$3.
1	ID	Num	8	BEST12.	BEST32.
5	IUCR	Char	10	\$10.	\$10.
20	Latitude	Num	8	BEST12.	BEST32.
22	Location	Char	31	\$31.	\$31.
8	Location_Description	Char	31	\$31.	\$31.
21	Longitude	Num	8	BEST12.	BEST32.
6	Primary_Type	Char	20	\$20.	\$20.
19	Updated_On	Num	8	DATETIME.	ANYDTDTM40.
13	Ward	Num	8	BEST12.	BEST32.
16	X_Coordinate	Num	8	BEST12.	BEST32.
17	Y_Coordinate	Num	8	BEST12.	BEST32.
18	Year	Num	8	BEST12.	BEST32.

Alphabetic List of Variables and Attributes					
#	Variable	Type	Len	Format	Informat
14	Arrest_B	Num	8		
9	Beat	Num	8	BEST12.	BEST32.
4	Block	Char	21	\$21.	\$21.
2	Case_Number	Char	8	\$8.	\$8.
12	Community_Area	Num	8	BEST12.	BEST32.
17	Crime_Date	Num	8		
3	Date	Num	8	DATETIME.	ANYDTDTM40.
19	Day_of_Month	Num	8		
7	Description	Char	35	\$35.	\$35.
10	District	Num	8	BEST12.	BEST32.
15	Domestic_B	Num	8		
13	FBI_Code	Char	3	\$3.	\$3.
16	Hour	Num	8		
1	ID	Num	8	BEST12.	BEST32.
5	IUCR	Char	10	\$10.	\$10.
8	Location_Description	Char	31	\$31.	\$31.
20	Month	Num	8		
6	Primary_Type	Char	20	\$20.	\$20.
11	Ward	Num	8	BEST12.	BEST32.
18	Weekday	Num	8		

STATISTICAL METHODS

With the above preliminary data preparation steps executed, our dataset was ready for our business analysis. The statistical methods we utilized in order to analyze our data set for this project were as follows:

1) PROC FREQ:

This proc measures to frequency of an occurrence. We used this particular PROC to analyze the following:

- The top crimes that are most likely to occur in Chicago
- Types of crimes that are most likely to lead to an arrest
- Top months when most crimes have occurred

- Top months in which most arrests have occurred
- Using **Chi square** option to test for equal proportions we determined the **Day of the month, Day of week and Time of day** when most arrests have been made (what day of the month corresponds to the highest number of arrests happening in Chicago)
- Using **Chi square** option to test for equal proportions we determined the **Day of the month, Day of week and Time of day** when most crimes have occurred.
- Top 15 most dangerous communities.

2. CHI-SQUARE TEST:

A chi-square test is used to examine the association between categorical variables. We have utilized it in order to determine:

- Association between the Different months ,time of day the crime has occurred and the day of week.
- Comparison between top crime rates and top communities.

3. ANOVA:

ANOVA Multiple Comparison Test with Tukey Post hoc analysis

- To study the variance in crime percentage across month and time of day.

SYNTAX

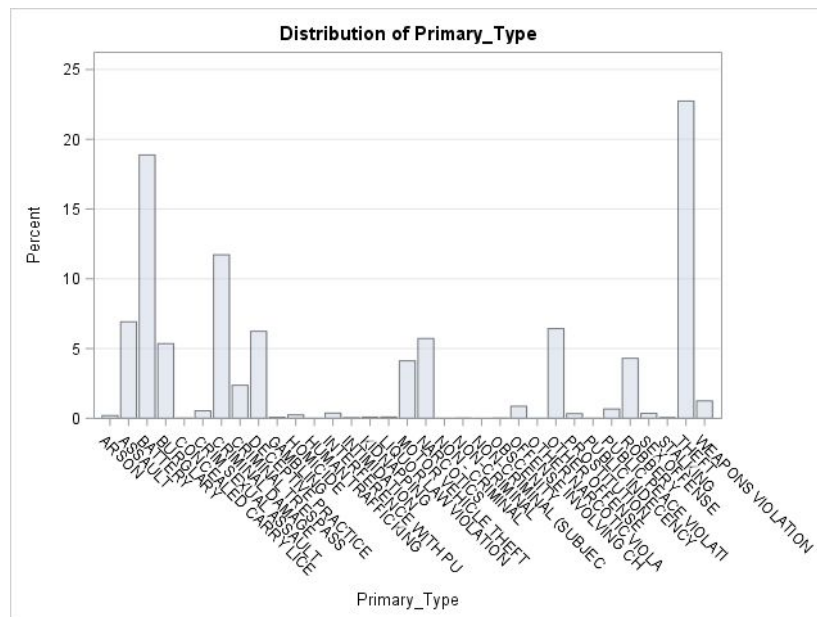
The SAS syntax we have used in order to perform the analysis will be included in the zip file as a .sas file.

RESULTS/FINDINGS/APPLICATION

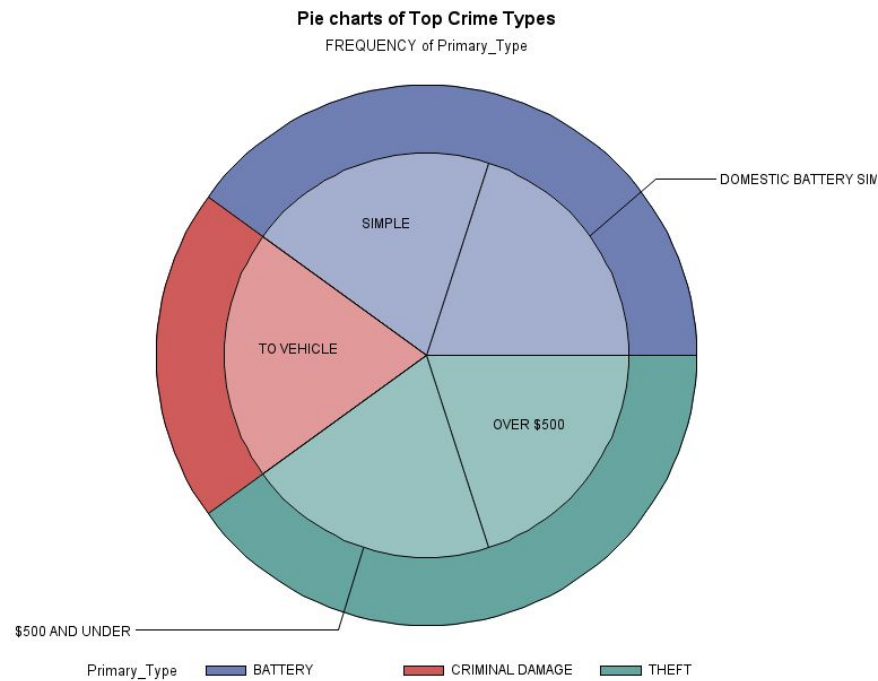
Having run all of our above mentioned code, we have found the following results:

1. TOP 5 TYPES OF CRIMES:

- Theft (22.74%)
- Battery (18.88%)
- Criminal damage (11.72%)
- Assault (6.92%)
- Other offences (6.43%)



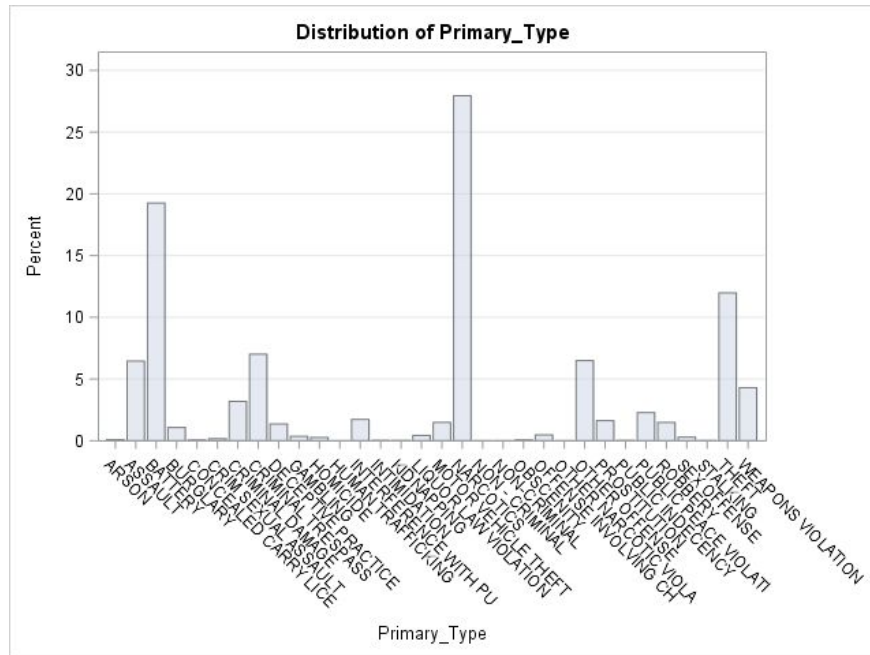
Within the top 5 Categories of crimes we chose top 3 and we went further deep and found primary category of crime type.



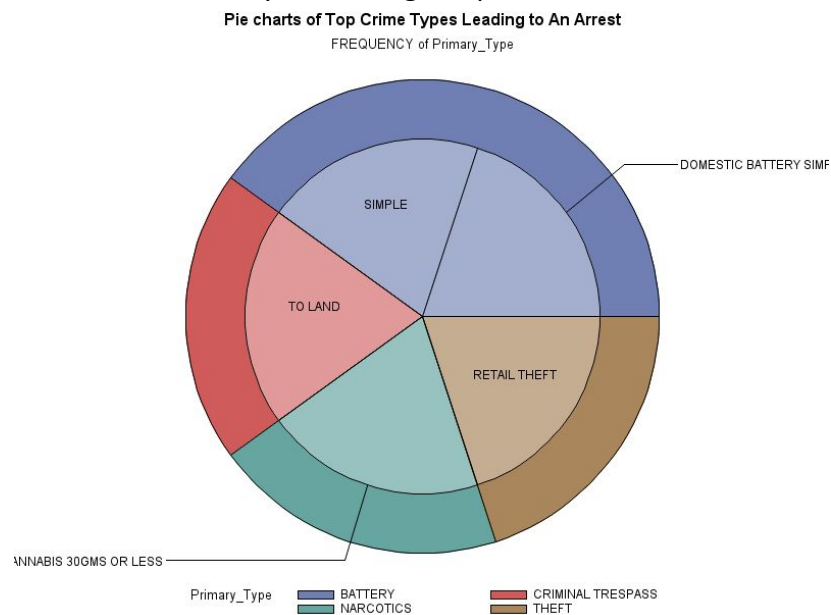
Recommendation : The top 5 crimes most prevalent in Chicago help us in determining the area in which officers should be provided specialised trainings to reduce and tackle these crimes efficiently.

2. TOP 4 TYPES OF CRIMES LEADING TO AN ARREST: Secondly, we studied which types of crimes were most likely to lead to an arrest. Here are our results:

- Narcotics (27.95%)
- Battery (19.26%)
- Theft (11.98%)
- Criminal trespassing (7.01%)



This helps us determine the intensity of training the police officers need.

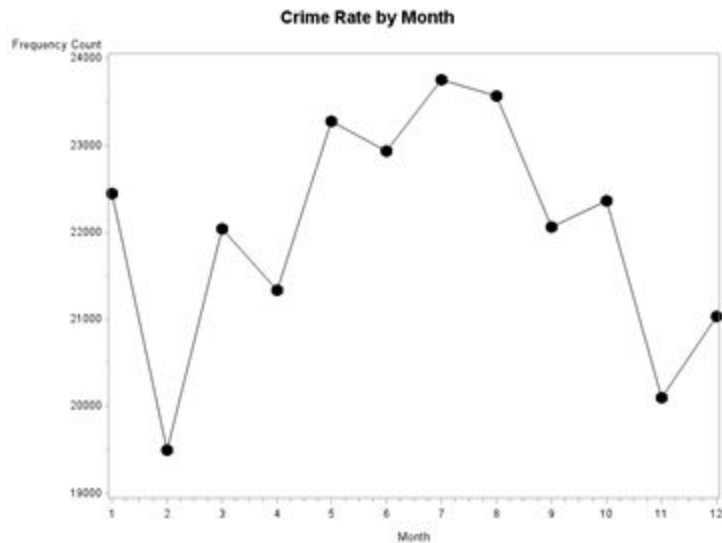


Recommendation : Theft and criminal damage which are a part of top 5 most prevalent crime has lower percentage of arrests that clearly means that our officers should focus more on these crimes areas.

Primary concentration of the Police department seems to be on cutting down crimes related to narcotics and hence leading to the most number of arrests.

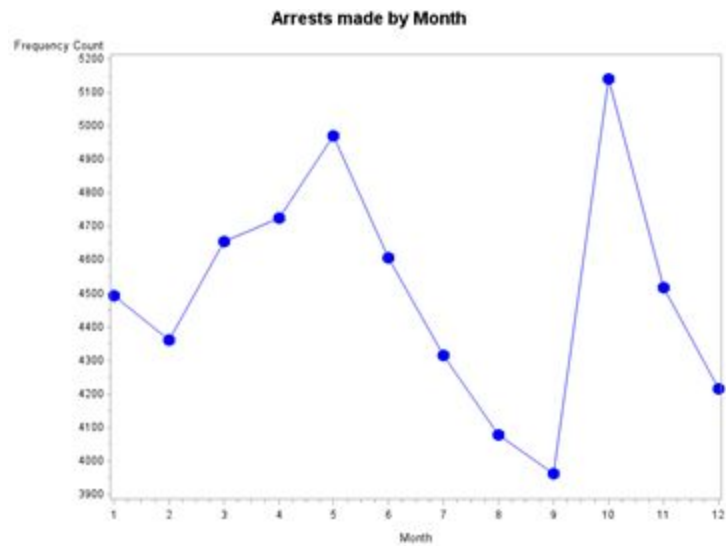
3. IN WHICH MONTH DO MOST CRIMES HAPPEN (TOP 5):

JULY, AUGUST, MAY, JUNE, JANUARY(from the period October 2015 to October 2016).



Recommendation : From graph below we can conclude that officers should be more alert in the months of July, August, May, June and January as these months show most crime percentage in decreasing order. Hence it can be concluded that summer months from May to August experience more crimes and hence our police forces should be more alert and active during that period specially.

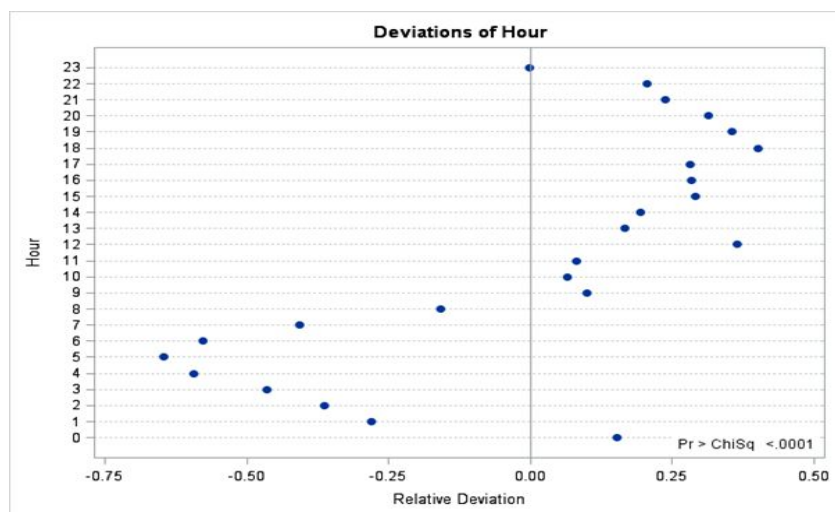
4. IN WHICH MONTH DO MOST ARREST OCCUR: OCTOBER, MAY, APRIL, MARCH, JUNE.



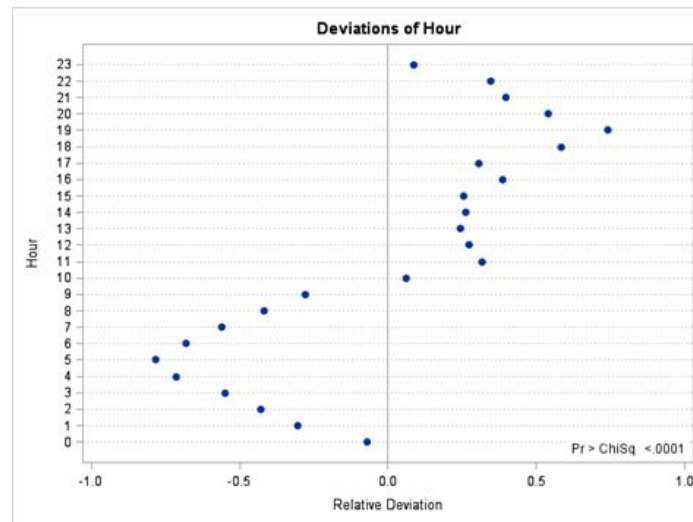
Recommendation : From our previous analysis we have seen that the top 5 months when most crimes occurs is July, August, May, June and January whereas the highest percentage of arrests are in October , May, April, March and June. Hence the police should probably analyze why there are lesser arrests when the crimes are highest are in months of July, August, June and January and probably consider in increasing the police force or patrolling during these months.

5. DEVIATIONS OF CRIME RATES ACROSS DIFFERENT TIMES IN A DAY:

Most crimes occur in between 3:00pm and 10:00pm. The least number of crimes occurs in hours between 1:00am and 8:00am.

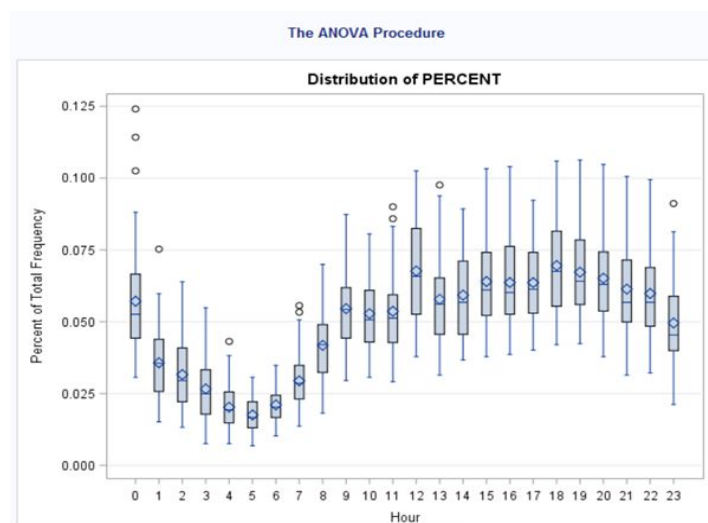


The deviation in the Arrests made by time of day is done using CHI SQUARE test of equal proportion.



The number of arrests when checked using the Chi-Square test of equal proportions reveals that the max number of arrest are made from 6 pm to 10 pm. This Recommendation is coherent with the corresponding spike in crime rate.

We further use the **Proc Anova** and the post hoc test (tukey) to ascertain at the statistical significance of the difference in crime occurrence by time of day. The results from the tests are denoted below:



Post hoc Tukey test was conducted to check for the variation in crime percentage for time of day.

Means with the same letter are not significantly different.																			
Tukey Grouping		Mean	N	Hour	F	B	D	E	C	0.059810	84	22	I						
	A	0.069540	84	18	F		D	E	C				I						
	A				F		D	E	C	0.059251	84	14	I	J		0.035693	84	1	
B	A	0.067644	84	12	F		D	E	C					J					
B	A				F		D	E	C	0.057810	84	13	K	J		0.031591	84	2	
B	A	0.067243	84	19	F		D	E					K	J					
B	A				F	G	D	E		0.057126	84	0	K	J		0.029403	84	7	
B	A C	0.065127	84	20	F	G		E					K						
B	A C				F	G		E		0.054523	84	9	K						
B	D A C	0.063984	84	15	F	G		E					K	L		0.026607	84	3	
B	D A C				F	G		E		0.053605	84	11		L					
B	D A C	0.063637	84	16	F	G							M	L		0.021023	84	6	
B	D A C				F	G				0.052821	84	10	M	L					
B	D A C	0.063605	84	17		G							M	L		0.020231	84	4	
B	D C					G		H		0.049548	84	23	M						
B	D E C	0.061363	84	21				H											
B	D E C				I		H			0.041767	84	8	M			0.017525	84	5	

Clearly from the results above, 6 pm has the highest crime rate and 1 am to 6 am see the lowest occurrence of crime.

Recommendation : From the above analysis more Police force should be in the evening hours and beats should be increased in those hours. The hours during afternoon which see lower number of crimes can have lesser police officers deployed during that period.

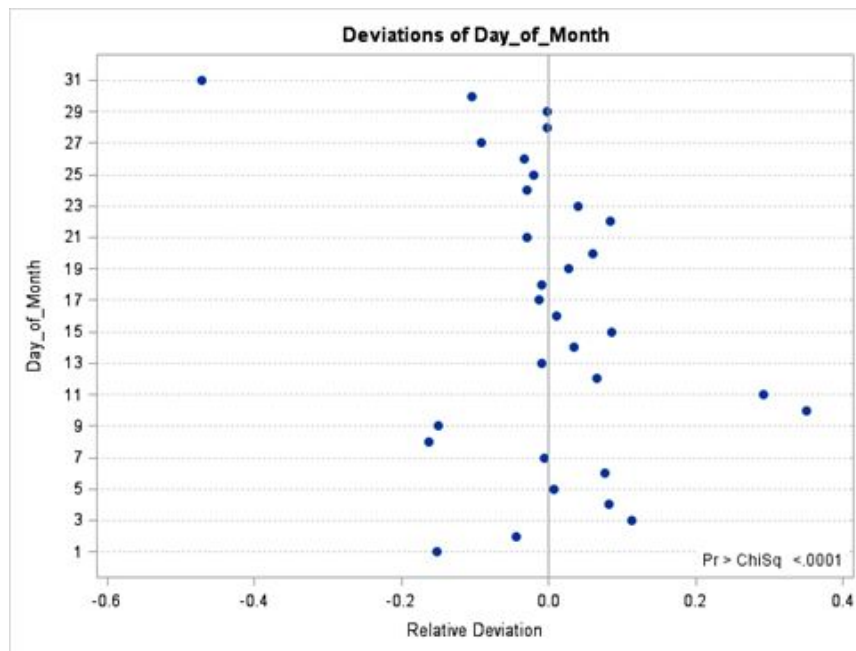
6. WEEKDAY WHEN MOST ARRESTS WERE MADE ACCORDING TO MONTH:

Jan	Feb	Mar	Apr	May	June	Jul	Aug	Sep	Oct	Nov	Dec
Sat	Tues	Wed	Fri	Mon	Thur	Fri	Mon	Thur	Sat	Tues	Thur

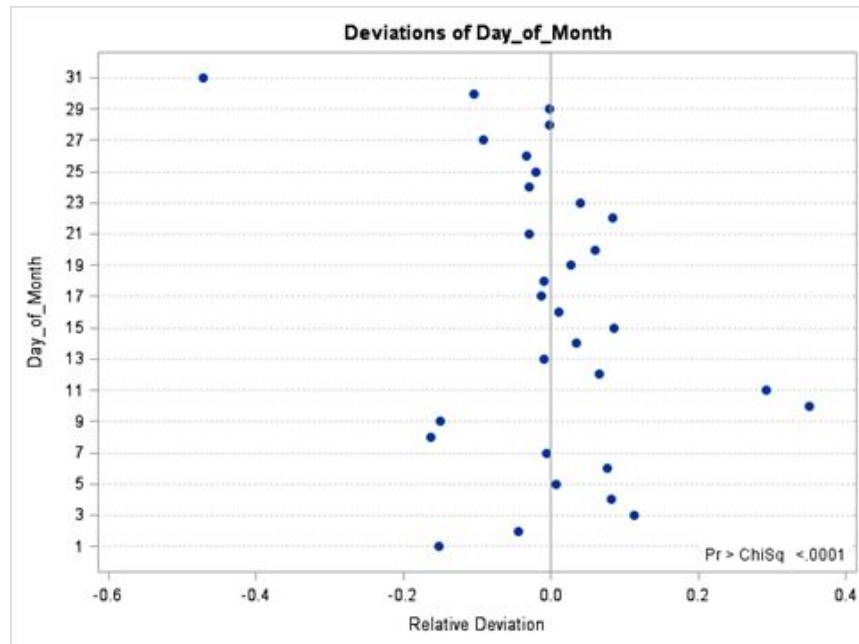
7. ARRESTS MADE BY DAY OF MONTH:

We also have found that the 10th and 11th day of each month is when most arrest have been made. 4.36% of all arrest made happened on the 10th and 4.17% happened on the 11th. This is also reflective of the high crime rate on these days of the month in the given duration of the study.

Crime rate by day of month



Arrests made by day of month



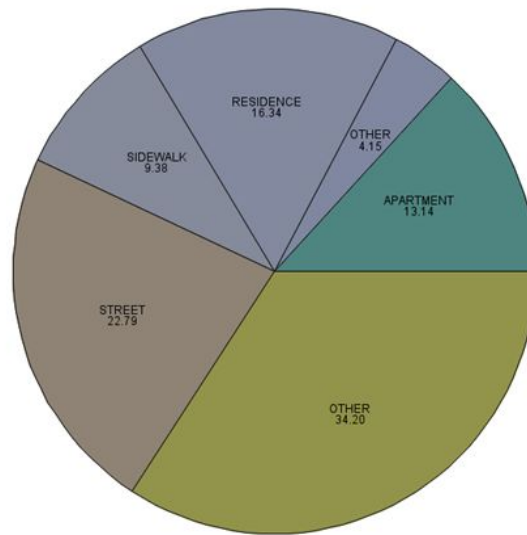
8. CRIME RATE BY LOCATION

To continue, we have found statistics that indicate where the most crimes occur (Street, Apartment, Residence, Sidewalk etc.). Here are our findings:

- Other (38.35%)
- Street (22.79%)
- Residence (16.34%)
- Apartment (13.14%)
- Sidewalk (9.38%)

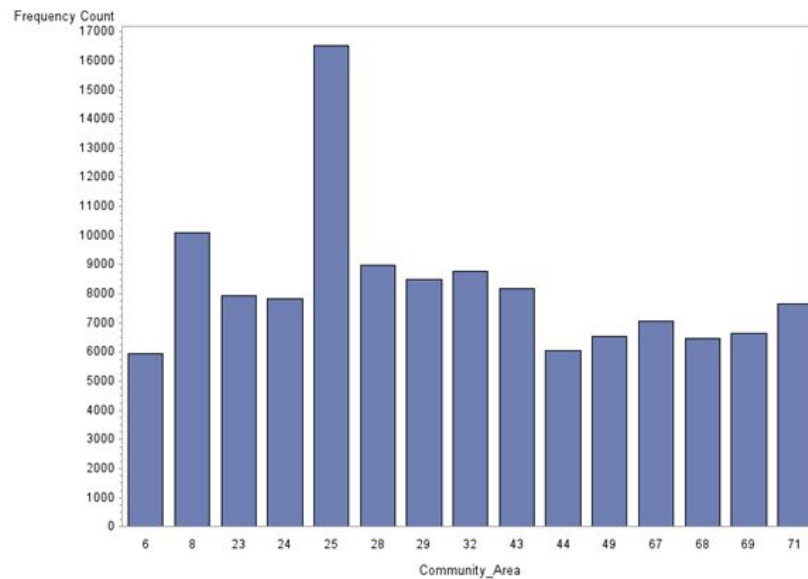
Recommendation : Street and Sidewalk combined consists of 33% and hence police should continue patrolling and increase their beats and force during peak hours of crime according. The crime percentage at Other location is 38% and is the highest. We don't have much data about "other" to specify the type of location and give recommendation.

Crime Rate by Location



9. NEIGHBORHOODS WITH THE HIGHEST CRIME RATE (According to community area):

Top 15 Dangerous Communities

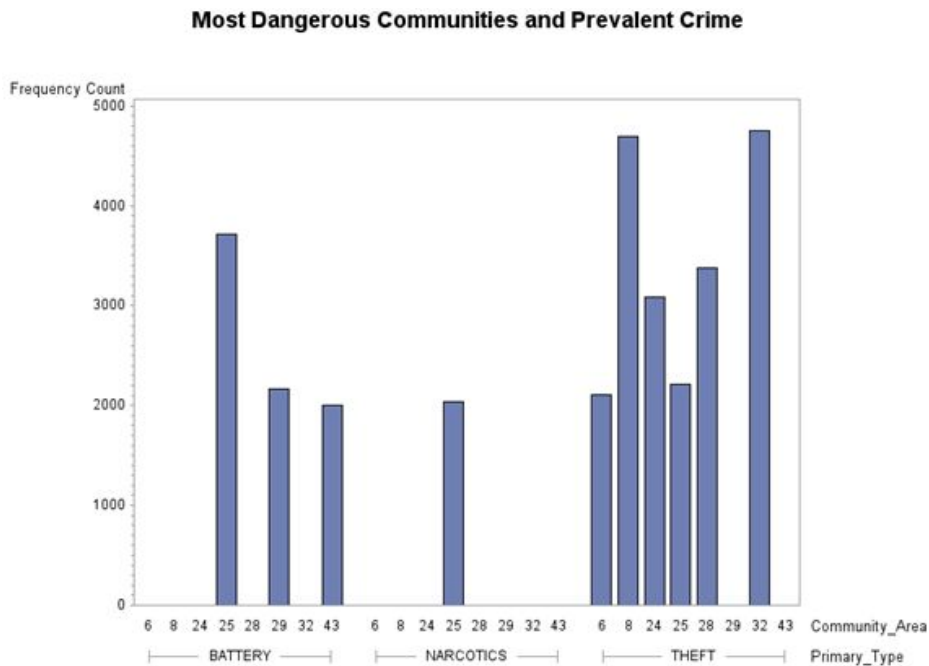


COMMUNITY CODE	COMMUNITY
6	Lake view
8	Near North side
23	Humboldt Park
24	West Town
25	Austin
28	Near West side
29	North Lawndale
32	Loop
43	South shore
44	Chatham
49	Roseland
67	West Englewood
68	Englewood
69	Greater Grand crossing
71	Auburn Gresham

Recommendation : The areas with highest crime rates should have more police beats and force deployed during peak hours and days of crime.

10. COMPARISON BETWEEN TOP CRIMES AND TOP COMMUNITIES:

1. Narcotics - 25(AUSTIN)
2. Theft: 32(LOOP), 8 (NORTH SIDE)
3. Battery: 25 (AUSTIN), 29(NORTH LAWNDALE)

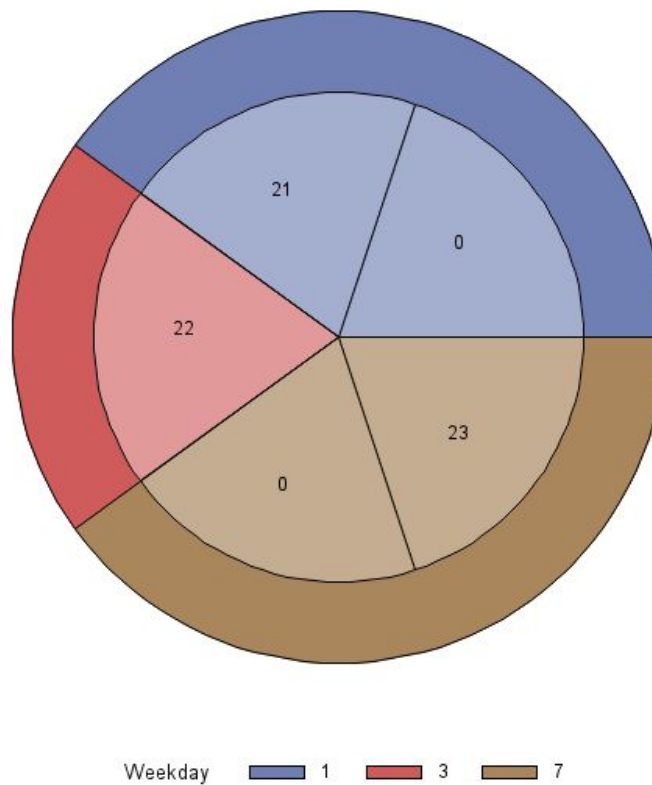


11. DOMESTIC CRIME

A descriptive study of the occurrence of domestic crime was analyzed. The data reveals that most of the domestic crime occur either on the weekend (Day 1-Sunday and Day 7-Saturday) or Tuesday and also they predominantly happen late nights. The slice inside the pie chart is the time of day the crime occurs(9 pm to 12 am) . In order, to prevent such crimes or help victims of such crime, the patrolling in residential areas could be increased so that the response time is reduced.

Top Domestic Crime Types By Day of Week and Time of Day

FREQUENCY of Weekday



RECOMMENDATIONS

Based on the statistical analysis above, we have the following recommendations that will help answer the major questions we have raised:

- We would suggest that the Chicago Police department spends the largest portion of their training budget on theft and battery education since these two crime types are the most common in the City of Chicago.
- We have also concluded that officers that specialize in busting drug gangs will need to be more prepared to conduct an arrest more than the officers who deal with theft or criminal trespassing.

- CPD needs to make sure they have enough staff to deal with the large number of crimes occurring. We suggest conducting major recruiting twice a year - April and September, since most crimes happen on the cusp of summer and fall months.
- Since most crimes occur in between **3:00pm and 10:00pm** and the least number of crimes occurs in hours between **1:00am and 8:00 am**, the CPD officers need to staff/dispatch the highest number of officers in the afternoon hours, rather than the morning hours.
- CPD needs to study the table that shows the “Weekday When Most Arrests Were Made According to Month.” By doing so they can get a better idea of when their jails are going to be reaching full capacity.
- From our analysis we were able to see that most arrests occur in the evening and night hours of the day. Again, this gives the police department better idea about their staffing and dispatch needs.
- Finding which neighborhoods have the highest crime rate let the CPD know which police precincts in the Chicagoland area need the most staffing. This information is useful and crucial for many different departments of the law enforcement field. For example, it lets the Finance team know which locations need the most funding in order to finance the salaries and wages of the police officers as well as which locations need the largest number of squad cars.
- Last but not least, the comparison between top crimes and top communities shows the relationship between the most likely type of crime to occur and the neighborhood in which the crime is likely to occur. This is a very useful piece of information in that it

informs the Chicago Police about what crime are the different police departments around the city most likely to see. This again shows the best division of training programs. For example, the police officers stationed near South Lawndale and Ashburn need a more extensive information and training on narcotics. This will ensure that the right officers are placed in these designated areas. It is better to dispatch officers that already have some experience and knowledge on this type of crime, rather than someone who's been dealing mostly with batteries, for example.

We are hoping that our findings will help the Chicago Police Department understand the patterns of crime occurrences. By analyzing and understanding the data statistically, instead of solely looking at the extensive amount of data on a spreadsheet, the CPD can ensure that Chicago will be a safer place, after all.

SOURCES

DATA: <https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-present/ijzp-q8t2>