ANALYZING THEFT PATTERN IN CHICAGO USING SAS ENTERPRISE MINER 14.1

Shikha Prasad, MS in Business Analytics, Oklahoma State University
Kunal Parekh, MS in MIS, Oklahoma State University

ABSTRACT

Chicago is one of the big cities in the United States that has witnessed a dramatic change in the crime rates in the past few decades. The objective of this paper is to find specific patterns for crimes such as larceny and motor vehicle theft versus other crimes that include battery, homicide, criminal damage, assault, etc. using Base SAS 9.4 and SAS Enterprise Miner 14.1. The purpose is to identify locations in the city that have the most frequent instances of these two types of thefts reported since 2012.

I. INTRODUCTION

Data driven crime analysis is used by police agencies and criminologists around the world to track the occurrences of criminal activities so that they can be prevented. Data from several sources such as police departments, census bureau and city data portals can be consolidated to perform predictive crime analysis and estimate the likelihood of a crime to happen in the future. Crime mapping is an application of crime analysis, which in conjunction with Global Information Systems (GIS) can be used to identify various crime zones on a map and find the association of a crime with factors that influence the occurrence of that crime. In most cases, studies have shown that crime incidents are often associated not only with the location but also with the time of the day, day of the week and month of the year.

The population of Chicago as estimated in 2015 was almost a third of the size of New York City, making it the third largest US city by population after New York and Los Angeles [1]. On comparing Chicago to other communities of similar population, its crime rate per thousand residents stands out as higher than most. Chicago's rate for property crime is 30 per thousand residents. This makes Chicago a place where there is an above average chance of becoming a victim of a property crime, when compared to all other communities in America of all population sizes. Property crimes are motor vehicle theft, arson, larceny and burglary. The chance of becoming a victim of any of these crimes in Chicago is one in 33. Based on FBI crime data, Chicago is one of the most dangerous communities in America. It also has one of the highest rates of motor vehicle theft in the nation when looking at communities of all sizes, from the smallest to the largest, with one in 266 chance of getting a car stolen [2].

II. LITERATURE REVIEW

According to the study performed by Andrew Hovel at College of Saint Benedict/Saint John's University, clustering of high property crime is mainly on the northeast side of Chicago whereas violent crime is more concentrated on the southern side as well as the northwest side. The thesis also provides correlation of crime with population density, employment rate, literacy rate, etc. [3].

In a recent case study conducted by Dr. Meead Saberi on crime statistics and travel behavior, it was found that spatial distribution of crime depends on the crime type. While theft is more concentrated in higher income areas of North Chicago, robbery and battery have higher concentration in low income areas of South and West Chicago. The downtown area of the city appears as a crime hot spot regardless of the crime type [4].

The Federal Bureau of Investigation's Uniform Crime Reporting (UCR) Program defines larceny-theft as the unlawful taking, carrying, leading, or riding away of property from the possession or constructive possession of another. Examples are thefts of bicycles, motor vehicle parts and accessories, shoplifting, pocket-picking, or the stealing of any property or article that is not taken by force and violence or by fraud [5]. Therefore, although such types of thefts may not always be life threatening, but they can result in loss of personal property due to asportation. In 2016, Chicago reported the highest number of larceny-theft incidents out of all crimes, making it the most prevalent type of crime in the city. Therefore, there is a need to identify community areas in the city of Chicago that report the highest proportion of larceny-thefts. The results from this paper can help in answering questions such as: What are the specific locations where most of these thefts are committed? What could be the possible reasons for the frequency of thefts being higher in these locations than other crimes? Preventive patrol in community areas that are most affected by larceny and motor vehicle thefts can act as a deterrent, making them safer by reducing the number of such crime incidents in the future.

III. DATA BACKGROUND

This dataset reflects reported incidents of crime that occurred in the city of Chicago from 2012 to 2017. Data is extracted from the Chicago Police Department's CLEAR (Citizen Law Enforcement Analysis and Reporting) system. To protect the privacy of crime victims, addresses are available at the block level, but specific locations are not identified [6].

IV. DATA PREPARATION

The data for the year 2012 through 2017 originally had 23 variables and 1,456,714 observations. For preparing the data to build a predictive model in order to predict the chances of theft versus other crimes, some transformations were done. In Base SAS 9.4, a new variable was created to obtain the month from the date variable which had date time stamp.

```
LIBNAME sp "C:\Users\Shikha Prasad\Downloads\Chicago_Crime_Latest\crimes-in-chicago";

OPTIONS MCOMPILENOTE = all;
OPTIONS MPRINT;
OPTIONS SYMBOLGEN;

%MACRO sp(yr);

DATA sp.cc_&yr;
LENGTH mon $3.;
SET sp.Cc_2012_to_2017;
WHERE year = &yr;
date = date*1;
date_new= datepart(date);
month= month(date_new);
if month = 1 then mon = "Jan";
if month = 2 then mon = "Feb";
if month = 3 then mon = "Mar";
if month = 4 then mon = "Apr";
if month = 5 then mon = "May";
if month = 6 then mon = "Jun";
if month = 7 then mon = "Jul";
if month = 8 then mon = "Aug";
if month = 9 then mon = "Sep";
if month = 10 then mon = "Oct";
if month = 11 then mon = "Nov";
if month = 12 then mon = "Dec";
run;

%MEND sp;
```

The macro was called for the six years – 2012 through 2017.

```
%sp(2012)
%sp(2013)
```

```
.
.
%sp(2017)
```

Yearly data was primarily generated to meet two purposes:

- 2017 had observations only for the month of January. Therefore, data for the latest available year, 2016 was needed to perform descriptive analytics on the 12 months of information
- 2012 data was required to identify a connection between community areas that have highest proportion of theft crimes and the socio-economic factors such as hardship index and per-capita income

In another data step, these yearly datasets were appended to create a consolidated data to be used for model development at a later stage.

```
DATA sp.cc_2012_2017_2;
SET
sp.cc_2012
sp.cc_2013
sp.cc_2014
sp.cc_2015
sp.cc_2016
sp.cc_2017;
RUN;
```

Further, FBI Code 06 (Larceny) and 07 (Motor Vehicle Theft) were assigned a value of 1 while all other crimes were assigned a value of 0 to create a binary target variable, Theft.

```
DATA CRIME.ANALYSIS1(RENAME=(Community_Area = Community_Area_Number));
SET CRIME.Cc_2012_2017_2;
IF fbi_code = "06" OR fbi_code = "07" THEN Theft = 1;
ELSE Theft = 0;
RUN;
```

The prepared data was merged with census data in order to incorporate socioeconomic factors that could help in providing useful insights.

```
PROC SORT DATA = CRIME.ANALYSIS1;
BY Community_Area_Number;
RUN;

DATA CRIME.census_socieconimic_data1;
SET CRIME.census_socieconimic_data;
IF Community_Area_Number = . THEN DELETE;
RUN;

PROC SQL;
CREATE TABLE CRIME.merge_1
AS
SELECT A.*, B.* FROM CRIME.ANALYSIS1 A LEFT JOIN CRIME.census_socieconimic_data1 B
ON A.Community_Area_Number = B.Community_Area_Number;
QUIT;
```

Chicago Data Portal Census Data*

The data obtained from Chicago Data Portal contains a selection of six socioeconomic indicators of public health significance and a “hardship index,” by community area, for the years 2008 – 2012 [7].

Community Area Number	COMMUNITY AREA NAME	PERCENT OF HOUSING CROWDED	PERCENT HOUSEHOLDS BELOW POVERTY	PERCENT AGED 16+ UNEMPLOYED	PERCENT AGED 25+ WITHOUT HIGH SCHOOL DIPLOMA	PERCENT AGED UNDER 18 OR OVER 64	PER CAPITA INCOME	HARDSHIP INDEX
8	Near North Side	1.9	12.9	7	2.5	22.6	88669	1
7	Lincoln Park	0.8	12.3	5.1	3.6	21.5	71551	2
32	Loop	1.5	14.7	5.7	3.1	13.5	65526	3
6	Lake View	1.1	11.4	4.7	2.6	17	60058	5
5	North Center	0.3	7.5	5.2	4.5	26.2	57123	6
33	Near South Side	1.3	13.8	4.9	7.4	21.8	59077	7
9	Edison Park	1.1	3.3	6.5	7.4	35.3	40959	8
24	West Town	2.3	14.7	6.6	12.9	21.7	43198	10
12	Forest Glen	1.1	7.5	6.8	4.9	40.5	44164	11
72	Beverly	0.9	5.1	8	3.7	40.5	39523	12
41	Hyde Park	1.5	18.4	8.4	4.3	26.2	39056	14
28	Near West Side	3.8	20.6	10.7	9.6	22.2	44689	15
74	Mount Greenwood	1	3.4	8.7	4.3	36.8	34381	16
4	Lincoln Square	3.4	10.9	8.2	13.4	25.5	37524	17
77	Edgewater	4.1	18.2	9.2	9.7	23.8	33385	19
3	Uptown	3.8	24	8.9	11.8	22.2	35787	20
10	Norwood Park	2	5.4	9	11.5	39.5	32875	21
22	Logan Square	3.2	16.8	8.2	14.8	26.2	31908	23
76	O'Hare	3.6	15.4	7.1	10.9	30.3	25828	24
11	Jefferson Park	2.7	8.6	12.4	13.4	35.5	27751	25
39	Kenwood	2.4	21.7	15.7	11.3	35.4	35911	26
17	Dunning	5.2	10.6	10	16.2	33.6	26282	28
64	Clearing	2.7	8.9	9.5	18.8	37.6	25113	29
75	Morgan Park	0.8	13.2	15	10.8	40.3	27149	30
56	Garfield Ridge	2.6	8.8	11.3	19.3	38.1	26353	32
13	North Park	3.9	13.2	9.9	14.4	39	26576	33
16	Irving Park	6.3	13.1	10	22.4	31.6	27249	34
15	Portage Park	4.1	11.6	12.6	19.3	34	24336	35
70	Ashburn	4	10.4	11.7	17.7	36.9	23482	37
48	Calumet Heights	2.1	11.5	20	11	44	28887	38
1	Rogers Park	7.7	23.6	8.7	18.2	27.5	23939	39
45	Avalon Park	1.4	17.2	21.1	10.6	39.3	24454	41
21	Avondale	6	15.3	9.2	24.7	31	20039	42
60	Bridgeport	4.5	18.9	13.7	22.2	31.3	22694	43
55	Hegewisch	3.3	17.1	9.6	19.2	42.9	22677	44
2	West Ridge	7.8	17.2	8.8	20.8	38.5	23040	46
35	Douglas	1.8	29.6	18.2	14.3	30.7	23791	47
73	Washington Height	1.1	16.9	20.8	13.7	42.6	19713	48
18	Montclaire	8.1	15.3	13.8	23.5	38.6	22014	50
50	Pullman	1.5	21.6	22.8	13.1	38.6	20588	51
49	Roseland	2.5	19.8	20.3	16.9	41.2	17949	52
14	Albany Park	11.3	19.2	10	32.9	32	21323	53
43	South Shore	2.8	31.1	20	14	35.7	19398	55
65	West Lawn	5.8	14.9	9.6	33.6	39.6	16907	56
38	Grand Boulevard	3.3	29.3	24.3	15.9	39.5	23472	57
42	Woodlawn	2.9	30.7	23.4	16.5	36.1	18672	58
44	Chatham	3.3	27.8	24	14.5	40.3	18881	60
59	McKinley Park	7.2	18.7	13.4	32.9	35.6	16954	61
53	West Pullman	3.3	25.9	19.4	20.5	42.1	16563	62
52	East Side	6.8	19.2	12.1	31.9	42.8	17104	64
51	South Deering	4	29.2	16.3	21	39.5	14685	65
69	Greater Grand Crossing	3.6	29.6	23	16.5	41	17285	66
57	Archer Heights	8.5	14.1	16.5	35.9	39.2	16134	67
62	West Elsdon	11.1	15.6	16.7	37	37.7	15754	69
19	Belmont Cragin	10.8	18.7	14.6	37.3	37.3	15461	70
20	Hermosa	6.9	20.5	13.1	41.6	36.4	15089	71
25	Austin	6.3	28.6	22.6	24.4	37.9	15957	73
71	Auburn Gresham	4	27.6	28.3	18.5	41.9	15528	74
46	South Chicago	4.7	29.8	19.7	26.6	41.1	16579	75
31	Lower West Side	9.6	25.8	15.8	40.7	32.6	16444	76
36	Oakland	1.3	39.7	28.7	18.4	40.4	19252	78
47	Burnside	6.8	33	18.6	19.3	42.7	12515	79
66	Chicago Lawn	7.6	27.9	17.1	31.2	40.6	13231	80
34	Armour Square	5.7	40.1	16.7	34.5	38.3	16148	82
27	East Garfield Park	8.2	42.4	19.6	21.3	43.2	12961	83
58	Brighton Park	14.4	23.6	13.9	45.1	39.3	13089	84
23	Humboldt park	14.8	33.9	17.3	35.4	38	13781	85
29	North Lawndale	7.4	43.1	21.2	27.6	42.7	12034	87
40	Washington Park	5.6	42.1	28.6	25.4	42.8	13785	88
67	West Englewood	4.8	34.4	35.9	26.3	40.7	11317	89
61	New City	11.9	29	23	41.5	38.9	12765	91
26	West Garfield Park	9.4	41.7	25.8	24.5	43.6	10934	92
63	Gage Park	15.8	23.4	18.2	51.5	38.8	12171	93
68	Englewood	3.8	46.6	28	28.5	42.5	11888	94
30	South Lawndale	15.2	30.7	15.8	54.8	33.8	10402	96
37	Fuller Park	3.2	51.2	33.9	26.6	44.9	10432	97
54	Riverdale	5.8	56.5	34.6	27.5	51.5	8201	98

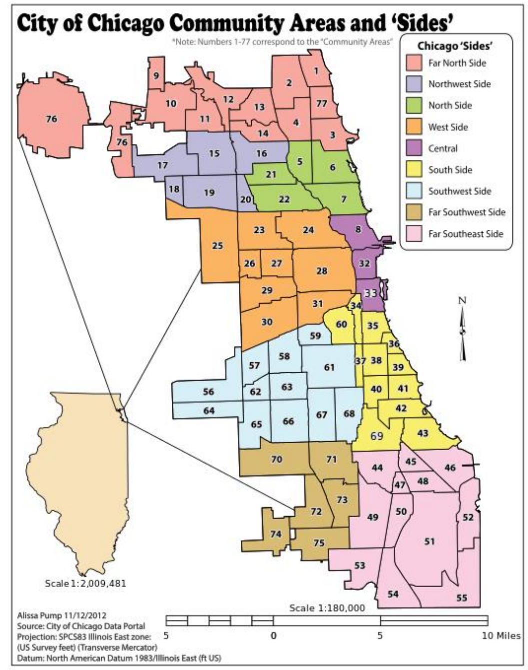

Figure 1: Chicago City Community Area Socioeconomic details and Map

**Image source: Internet*

V. DATA EXPLORATION

The data for 2012 was explored to perform some initial descriptive analysis. It was observed that there are certain community areas in the city that report higher percentage of crimes. These are community area number 25 (Austin), number 8 (Near North Side) and number 43 (South Shore). The proportion of thefts out of all crimes is highest at 54% in area number 32 (Loop), followed by number 7 (Lincoln Park) and number 8 (Near North Side), where thefts account to 50% of all types of crimes. These are the community areas that are listed as top three communities of the city in the socioeconomic data from 2008 – 2012.

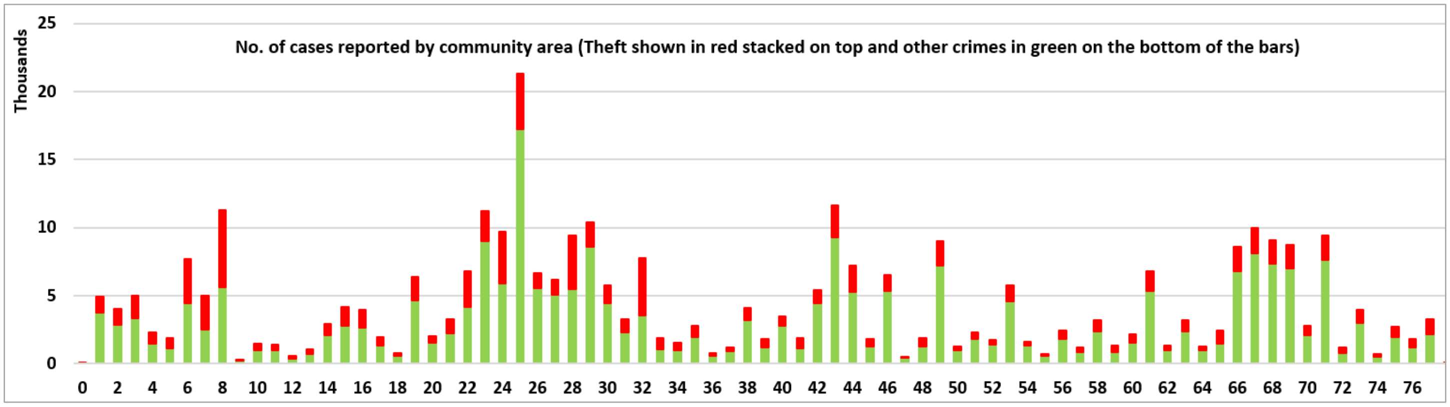

Figure 2: Frequency of crimes by Community Area - 2012

The graph below shows the distribution of thefts by month for the year 2016. It can be seen that most of the thefts are reported in the months of July and August. The reported figures are comparatively lesser from November through April.

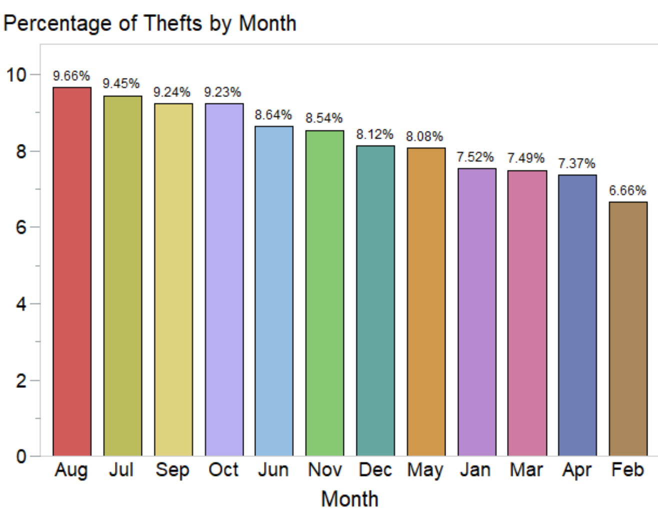

Figure 3: Descending order of thefts reported by Month - 2016

In the consolidated data for all the years, most of the theft cases (89% of all thefts) do not result in arrests:

Theft=1

Arrest	Frequency	Percent	Cumulative Frequency	Cumulative Percent
False	349671	89.52	349671	89.52
True	40927	10.48	390598	100.00

Table 1: Arrests made for larceny and motor vehicle theft – 2012 to 2017

The percentage of thefts out of all crimes was observed to be highest in locations such as athletic clubs and department stores.

Location Description	Other (%)	Theft (%)
ATHLETIC CLUB	16.47	83.53
DEPARTMENT STORE	18.13	81.87
DELIVERY TRUCK	24.24	75.76
AIRPORT TERMINAL LOWER LEVEL - SECURE AREA	24.53	75.47
VEHICLE - DELIVERY TRUCK	25	75
AIRPORT BUILDING NON-TERMINAL - SECURE AREA	29.14	70.86
GROCERY FOOD STORE	32.1	67.9
DRUG STORE	34.17	65.83
CTA TRAIN	42.11	57.89
SMALL RETAIL STORE	43.12	56.88
CONSTRUCTION SITE	43.23	56.77
AIRPORT BUILDING NON-TERMINAL - NON-SECURE AREA	43.96	56.04
FIRE STATION	44	56
AIRPORT EXTERIOR - SECURE AREA	44.31	55.69
VEHICLE-COMMERCIAL	44.82	55.18
AIRPORT TERMINAL MEZZANINE - NON-SECURE AREA	45	55
AIRPORT TRANSPORTATION SYSTEM (ATS)	45.28	54.72
DRIVEWAY - RESIDENTIAL	48.41	51.59
MOVIE HOUSE/THEATER	48.9	51.1

Table 2: Common places where thefts are reported – 2012 to 2017

Very small percentage of the thefts were categorized as domestic thefts.

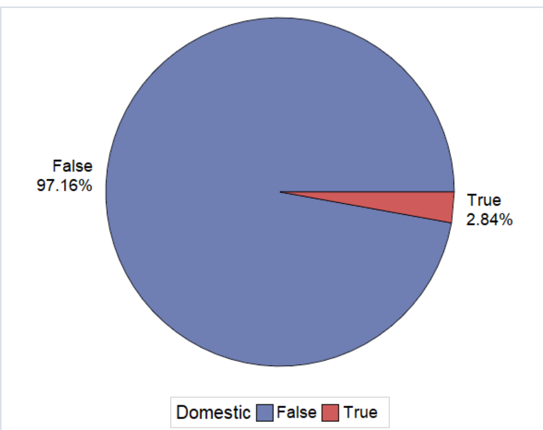

Figure 4: Domestic vs Non-Domestic – 2012 to 2017

VI. MODEL BUILDING

To predict the outcome of a crime being a theft or not, several models were built using SAS Enterprise Miner 14.1.

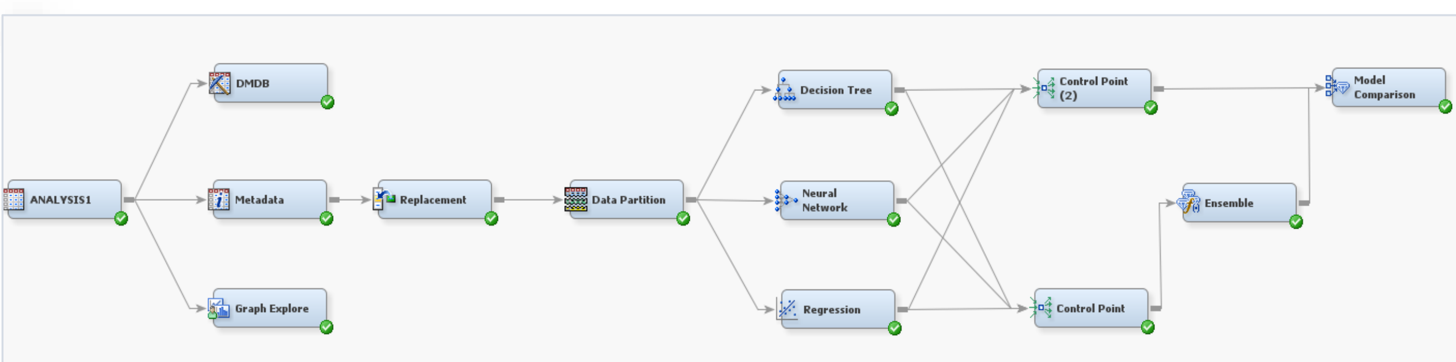

Figure 5: SAS EM Process Flow Diagram

The following results were obtained from the Data Mining Database (DMDB) node, which shows the missing values for all the variables in the dataset (here, the maximum number of class levels was set at 25 so the output does not show all the levels for the nominal variables):

Interval Variable Summary Statistics

Variable	Label	Missing	N	Minimum	Maximum	Mean	Standard Deviation	Skewness	Kurtosis
Latitude		37083	1419631	36.62	42.02	41.84	0.09	-9.2508	508.512
Longitude		37083	1419631	-91.69	-87.52	-87.67	0.07	-12.0766	712.770
X_Coordinate	X Coordinate	37083	1419631	0.00	1205119.00	1164397.91	18508.35	-13.6976	848.655
Y_Coordinate	Y Coordinate	37083	1419631	0.00	1951573.00	1885523.19	34247.75	-9.0912	496.749
date_new		0	1456714	18993.00	20837.00	19867.96	537.54	0.1160	-1.223

Class Variable Summary Statistics

Variable	Label	Type	Number of Levels	Missing
Arrest		C	2	0
Beat		N	26	0
Block		C	26	0
Case_Number	Case Number	C	26	0
Community_Area_Number	Community Area	N	25	40
Description		C	26	0
District		N	24	1
Domestic		C	2	0
F1		N	26	0
FBI_Code	FBI Code	C	26	0
IUCR		C	26	0
Location		C	25	37083
Location_Description	Location Description	C	26	0
Primary_Type	Primary Type	C	26	0
Theft		N	2	0
Ward		N	25	14
Year		N	6	0
mon		C	12	0
month		N	12	0

Figure 6: Results from DMDB node

The data had missing values for the variables – Latitude, Longitude, Location, Community Area Number, Ward, etc. For the analysis purpose, 40 missing values for Community Area Number were replaced by a value of "99" using a replacement node. Similar replacements were done for Ward, which had 14 missing values and District, which had 1 missing value. This was done so that observations with the missing values were not rejected as incomplete cases by the regression model. Interval variables, with a large percent of missing values such as latitude, longitude, X and Y coordinates were rejected from the analysis and imputation or replacements were not performed for these variables. The metadata node was used to set the measurement levels and input roles for the variables:

Name	Hidden	Hide	Role	New Role
Year	N	Default	Input	Default
Community	N	Default	Input	Default
Domestic	N	Default	Input	Default
ID	N	Default	ID	Default
Arrest	N	Default	Input	Default
Location_Des	N	Default	Input	Default
mon	N	Default	Input	Default
Updated_On	N	Default	Time ID	Rejected
Longitude	N	Default	Input	Rejected
Primary_Typ	N	Default	Input	Rejected
date_new	N	Default	Input	Rejected
Y_Coordinate	N	Default	Input	Rejected
month	N	Default	Input	Rejected
Ward	N	Default	Input	Rejected
X_Coordinate	N	Default	Input	Rejected
Description	N	Default	Input	Rejected
Date	N	Default	Time ID	Rejected
District	N	Default	Input	Rejected
Block	N	Default	Input	Rejected
Beat	N	Default	Input	Rejected
Case_Numbe	N	Default	Input	Rejected
IUCR	N	Default	Input	Rejected
Location	N	Default	Input	Rejected
Latitude	N	Default	Input	Rejected
F1	N	Default	Input	Rejected
FBI_Code	N	Default	Input	Rejected
Theft	N	Default	Input	Target

Figure 7: Metadata settings

Description was rejected as it had specifications related to a particular offense, which could mask the significance of other relevant variables in the model. Primary Type was also rejected for the same reason. Date, FBI Code and F1 were some of the other variables that were rejected as these were considered irrelevant in explaining the response. Certain other variables like Block, Beat and District were rejected in order to make the model simpler and use the Community Area Number as one input explanatory variable instead of these three variables.

Next, the data was partitioned into 60% train and 40% validation for obtaining proper model assessment. In order to predict the response, three models were built using the Regression, Decision Tree and Neural Network nodes. An ensemble model was also built to combine predictions from these models. The Model Selection node was used to select the model with the least value of validation misclassification rate.

From the results of the regression model, it was found that the variables that are significant in predicting whether a crime will be identified as theft or non-theft are Arrest, Domestic, Location Description, Community Area Number, Year and Month:

Type 3 Analysis of Effects

Effect	DF	Wald Chi-Square	Pr > ChiSq
Arrest	1	43161.3654	<.0001
Domestic	1	15842.3291	<.0001
Location_Description	132	78091.5729	<.0001
REP_Community_Area_Number	78	12849.4684	<.0001
Year	5	1003.7019	<.0001
mon	11	132.2187	<.0001

Figure 8: Results from Logistic Regression

From the Decision Tree model, locations with higher chances of thefts as compared to other crimes can be identified based on certain conditions. The screenshot below shows a partial output from the default Decision Tree node. The English node rules help in explaining that the probability of a crime being reported as Theft is 94% for the conditions based on Community Area Number, Location Description, Domestic and Arrest Flags.

```
*------------------------------------------------------------*
 Node = 59
*------------------------------------------------------------*
if Replacement: Community Area IS ONE OF: 1, 2, 3, 4, 5, 6, 7, 8, 13, 15, 16, 18, 19, 20, 21, 22, 23, 24, 25, 28, 32, 35, 37, 39, 43, 44, 45, 46, 50, 51, 55, 57, 59, 61, 63, 65, 66, 69, 71, 75, 77
AND Location Description IS ONE OF: DEPARTMENT STORE
AND Domestic IS ONE OF: FALSE or MISSING
AND Arrest IS ONE OF: TRUE
then
 Tree Node Identifier   = 59
 Number of Observations = 6983
 Predicted: Theft=1 = 0.94
 Predicted: Theft=0 = 0.06
```

Figure 9: Decision Tree node rules

The Neural Network and Ensemble model generated the best prediction with maximum validation ROC index – 82.2% area under the curve. ROC was chosen as the metric for model assessment because it is insensitive to the bias present in the data – there are more number of other crimes (response = 0) than the number of thefts (response = 1).

Selected Model	Predecessor Node	Model Node	Model Description	Target Variable	Selection Criterion: Valid: Roc Index
Y	Neural2	Neural2	Neural Network	Theft	0.822
	Ensmbl	Ensmbl	Ensemble	Theft	0.822
	Reg	Reg	Regression	Theft	0.816
	Tree	Tree	Decision Tree	Theft	0.814

Theft	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	1066116	73.19	1066116	73.19
1	390598	26.81	1456714	100.00

Table 3 and 4: Model assessment results and proportion of thefts in the consolidated data

VII. CONCLUSION

1. Community areas that are likely to report higher occurrences of thefts in Chicago are Lincoln Park (Community Area #7), Near North Side (Community Area #8) and Loop (Community Area #32). These are located in the North and Central regions of the city. The interesting fact here is that these three areas can be traced on the map, just one below the other.
2. Using census data obtained from Chicago Data Portal, it can be concluded that the areas where thefts are most prevalent have low value of hardship index reported. This means that thefts occur primarily in those areas which are more developed and where the per capita income is greater than $65,000.
3. Only about 2.8% of all the thefts are reported as Domestic, which means that most of the theft occurrences are not related to households but instead take place outside homes in locations such as street, sidewalk and parking lot, etc.
4. Thefts in general, result in lesser number of arrests made as compared to other crimes.
5. Specific places where thefts are more reported include athletic clubs, department stores, food and drug stores, airport terminals, delivery truck and commercial vehicle locations and residential driveways.

A good predictive model that is capable of generating the probability of theft occurrences based on inputs such as community area and location can be useful for better law enforcement and effective policing. This in turn would be helpful in reducing the number of potential thefts, thereby preventing financial losses that the crime victim has to suffer.

VIII. LIMITATIONS OF STUDY

The analysis does not classify property crimes such as robbery and burglary as target event, i.e., theft = 1. These crimes may involve the use of force and/or putting the victim in fear and are therefore different from larceny or automobile thefts. The results might vary if all the property crimes are taken into consideration.

IX. FURTHER WORK

The scope of this paper could be further extended to perform more advanced spatial analysis by taking geography (location details such as latitude and longitude) into consideration.

X. ACKNOWLEDGEMENT

We thank Dr. Goutam Chakraborty, SAS® Professor of Marketing Analytics and Dr.Miriam McGaugh, Clinical Professor at Oklahoma State University for their guidance and support throughout the research.

XI. REFERENCES:

[1] List of United States cities by population. (2017, September 12). In *Wikipedia, The Free Encyclopedia*.
, from https://en.wikipedia.org/w/index.php?title=List_of_United_States_cities_by_population&oldid=800217305

[2] Neighborhoodscout.com. (2017). *Chicago Crime Rates and Statistics - NeighborhoodScout*.: https://www.neighborhoodscout.com/il/chicago/crime

[3] Hovel, A. (2014). *Crime, Income Inequality, and Density at the Neighborhood Level*. [online] Digitalcommons.csbsju.edu. : http://digitalcommons.csbsju.edu/cgi/viewcontent.cgi?article=1040&context=honors_theses

[4] CityX | Dr. Meead Saberi. (2017). *Crime Statistics and Travel Behavior*: http://www.cityxlab.com/crime-statistics-and-travel-behavior.html

[5] U.S. Department of Justice, FBI. (2017). *Larceny-theft*.: https://ucr.fbi.gov/crime-in-the-u.s/2010/crime-in-the-u.s.-2010/property-crime/larcenytheftmain

[6] Kaggle.com. (2017). *Crimes in Chicago | Kaggle*.: https://www.kaggle.com/currie32/crimes-in-chicago

[7] Data.cityofchicago.org. *Census Data - Selected socioeconomic indicators in Chicago, 2008 – 2012 | City of Chicago | Data Portal*.: https://data.cityofchicago.org/Health-Human-Services/Census-Data-Selected-socioeconomic-indicators-in-C/kn9c-c2s2

XII. CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the authors at:

1. Shikha Prasad, Oklahoma State University, Stillwater OK
   E-mail: Shikha.prasad@okstate.edu
2. Kunal Parekh, Data Scientist, Verisk Analytics, San Francisco (OSU Alumni)
   E-mail: kunalparekh78@gmail.com