Project Proposal: Chicago Crime Data Analysis

MIDS -INFO- S18 Python Fundamental
Project 2: Data Analysis
Angshuman Paul, Anuradha Sankaranarayanan, Yue Hu, Kuangwei Huang

Summary:

The analysis of Crimes in Chicago has been an area of interest in most studies related to crimes, especially those in the United States of America. Since the beginning of the 20th century, the crime rate in Chicago has been above the US average[1]. We will be analyzing the crime data made available by the city of chicago and try to uncover some insights related to the crimes that were committed in the city between 2015 and 2017. We will also be using weather data as a supplementary dataset to understand any bearing that weather may have had on these crimes.

[1] - https://en.wikipedia.org/wiki/Crime_in_Chicago

Datasets:

1. **Chicago Crime Data Since 2001.** This dataset reflects reported incidents of crime that occurred in the City of Chicago from 2001 to present, minus the most recent seven days. We will slice this dataset to focus on a more manageable subset spanning 3 years from January 1, 2015 through to December 30, 2017.
   Dataset Link: https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-present/ijzp-q8t2

2. **Weather Data**. In order to study the crime rate with weather, the weather data in Chicago from January 1, 2015 through to December 30, 2017 are downloaded from Weather Underground website and parsed into a flat CSV file.

   Python Code Link: https://github.com/fivethirtyeight/data/tree/master/us-weather-history (wunderground_scraper.py and wunderground_parser.py are used to download and store weather data in CSV file)

   Dataset link: https://drive.google.com/open?id=1aqqyxW77aafZINpV4fHIXVN-SKag2sACzT67DY66odY

Dataset structure:

Chicago Crime Data:

- Full data set total lines (Since 2001): 6,563,183,
- 3-year subset total lines(2015 - 2017): 798,748
- Number of columns: 17

Weather Data:

- Data set total lines (January 1, 2015 - December 30, 2017): 1,096
- Data set total columns: 13

Initial data exploration:

Figure 1: Countplot of different types of crimes between 2015 - 2017

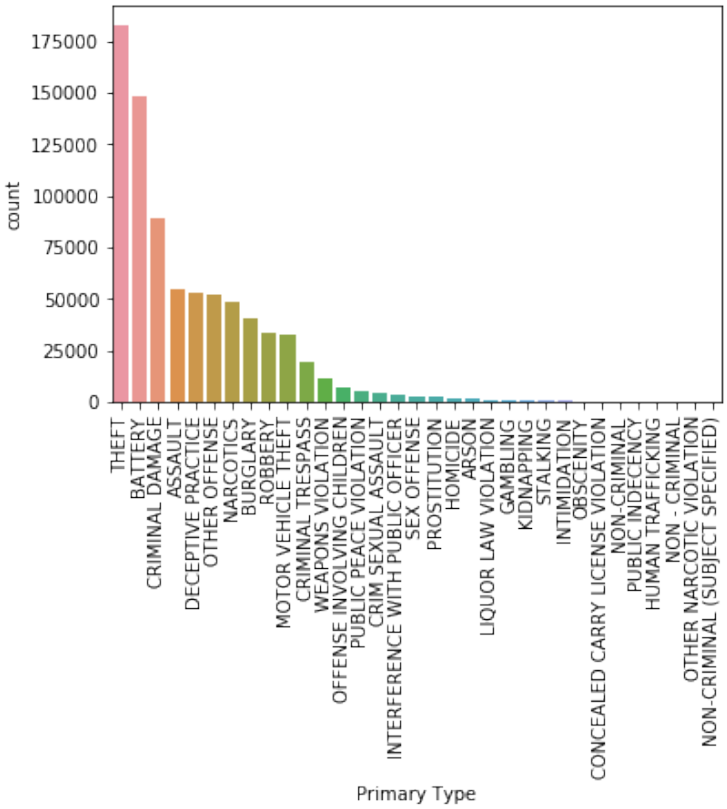
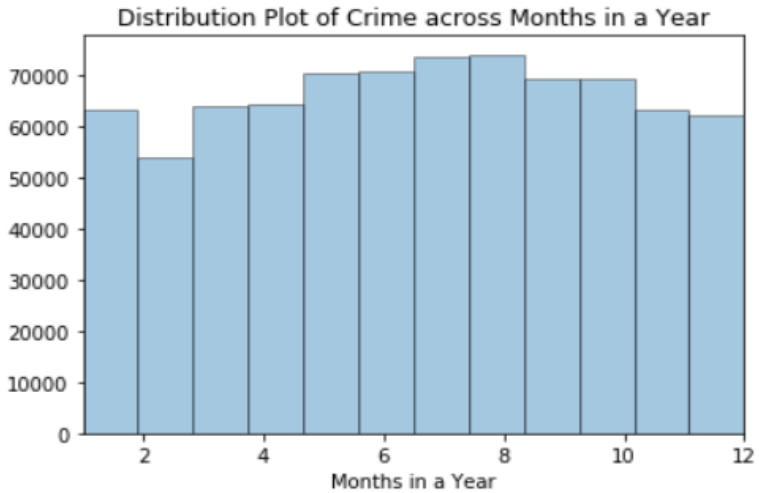


Figure 2: Distribution Plot of Crime across Months in a Year



Dataset Stats:

Crime Dataset Stats:

	Date	Block	IUCR	Primary Type	Description	Location Description	Arrest	Domestic
count	798748	798748	798748	798748	798748	798748	798748	798748
unique	342113	31884	353	33	329	144	2	2
top	2015-01-01 00:01:00	001XX N STATE ST	820	THEFT	SIMPLE	STREET	FALSE	FALSE
freq	150	2536	73379	183051	84790	181658	624787	671585

	Beat	District	Ward	Community Area	FBI Code	Year	Latitude	Longitude	YrMth
count	798748	798748	798748	798748	798748	798748	798748	798748	798748
unique	274	24	51	78	26	3	264443	264353	36
top	1834	11	42	25	6	2016	nan	nan	201608
freq	7720	55925	44548	49356	183051	268428	8570	8570	24817

Weather Dataset Stats:

	actual_mean_temp	actual_min_temp	actual_max_temp	actual_precipitation
count	1095	1095	1095	1095
mean	52.005479	43.099543	60.395434	0.10853
std	19.690428	18.913837	21.019921	0.295517
min	-2	-13	4	0
25%	37.5	30	43	0
50%	54	44	63	0
75%	69	59.5	79	0.05
max	84	76	95	4.19

What we plan to cover in the final report:

As a part of our analysis project some of the questions we will try to gather insights on are:

1. How does crime vary across the years from 2015 through 2017?
2. How does crime in Chicago relate to the weather?
3. Are there any high incidence areas whereby the police should increase patrols during different weather, times of the year?
4. What type of crime is the most frequently committed (and in which season)?
5. Are domestic crimes more or less violent than the non-domestic ones in general? Are they more or less frequent?
6. Is any part of the city more susceptible to crimes involving firearms?
7. Is there any correlation between arrests and the type, location or time of crime?
8. Is there any correlation between the type of crime and the type of location where they are committed? For example, are thefts more common in retail outlets and gas stations as opposed to apartment or residences?
9. How does theft vary during the holiday season (Thanksgiving to New Year) compared to the rest of the year?