Weather on crime

Data Analysis Report

CS5163 - Introduction to Data Science
Fall 2018

Anusha Meka
Department of Computer Science
University of Texas at San Antonio
One UTSA Circle, San Antonio, TX 78249
Email: bul924@cs.utsa.edu

Manasi Subhedar
Department of Computer Science
University of Texas at San Antonio
One UTSA Circle, San Antonio, TX 78249
Email: ikl598@cs.utsa.edu

Abstract:

The main idea behind this data analysis is the new gun policy in chicago that has been introduced in early 2013. Other idea is to dig the crime data and analyze whether there is any relation between crime rate and temperature changes ie., seasonal changes. The idea to link crime data with whether data came from the notion in our hometown that crime rate is high during summer season because people do not find a job and they prefer to do some or other type of crime and earn money. Does this rule follows the chicago crime data as well? The dataset we picked reflects reported crime incidents that happened in the City of Chicago from 2001 to present. Dataset is exported from the Chicago Police Department's CLEAR (Citizen Law Enforcement Analysis and Reporting) framework. So as to secure the protection of crime unfortunate casualties, addresses are appeared at the square dimension just and explicit areas are not distinguished. The main questions covered in the report are - 1. How can we attribute a fall in violent crime in Chicago to its new conceal and carry laws? 2. How does this policy contributed to the murder rate promptly [falling] to 1958 levels? 3. Does weather has impact on crimes in Chicago? If so, on which types of crimes?

The dataset has 32 different types of crimes from which we have classified them into personal and property crimes. We have tried to find out the correlation between various types of crimes and temperature. We found the Relationship crimes has with environmental factors like temperature as well as temporal factors like the hour of the day and day of the week. Build a fitting curve for the temperature data for one year. We have checked whether the same statistical modeling (curve fitting) of temperature is applicable for personal and property crimes. With the help of the modeling we have predicted the crime rate from 2018 to 2021. We have Found the evidence of substantial differences in reporting crime data since 2013.

Introduction:

The seasonal cycle of Chicago crime peaks during summer months, but not all types of crime are equally susceptible to changes in temperature. Data show that it isn't necessarily that more crimes are linked with the summer. The connection is the warm weather in general, no matter what time of year. A lot of data is available for Chicago, so we looked at some graphs focusing on the third largest city i.e. Chicago in the U.S. The weather doesn't cause crime, of course. It is people's actions that lead to violence. Another reason for the increase in crime could just be that there are more chances with the population increasing outdoors. However, many of these studies were done with climate change in mind.

We have basically main four things:

1. To start with, we demonstrated the relationship crime has with temporal variables like temperature and additionally fleeting elements like the hour of the day and day of the week. We have recognized that criminal activity follows a yearly pattern, as well as examples by day of the week.
2. Second, we estimated a simple statistical model dependent on the discoveries above. This model consolidates temperature, the day of the week, the seven day stretch of the year, and longer-term chronicled patterns and in spite of its straightforwardness, completes a great job clarifying the elements of crimes in Chicago in the course of recent years.
3. Third, we have utilized this statistical model to make predictions about crime rate for left

of 2014. If there's a significant fall-off in violent crime following the introduction of the conceal and carry policy in March 2014, this could be proof of its success as an deterrent. But if the real crime information coordinates the model's forecasted patterns, it recommends the new hide and convey strategy has had no impact.

4. Fourth, we discovered proof of substantial discrepancies in the reporting crime data since 2013. we performed some extra investigations to reveal which violations and detailing areas are driving this discrepancy and in addition how extreme this discrepancy is.

Materials:

We have two different datasets - 1. Crime Dataset includes crime reports of Chicago from 2001 to present. The dimension of the data set is 6.74M records * 22 Columns. It is collected from https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-present/ijzp-q8t2/data.
2. Weather Dataset is collected from Google which includes yearly data from 2001 to 2014 and combined it into single dataset. The dimension of Weather data set is 4850 records * 23 cols.

We have combined each year's weather CSV files into a master CSV file by loading each CSV and saving it into the weather list of DataFrames and then using pd.concat to combine them all together. Then we save the resulting data as weather.csv to load up later.

In crime dataset we have a column with date and time when the crime was happened. We modified that column into 5 columns like weekend, weekday, hour of the day, month of the year, year. In weather dataset we have a column named precipitation with missing values which we have changed those into NaN.

Results and discussion:

We have considered Crime as a function of Temperature. That is how is the distribution of personal and property crimes with temperature. Binned mean temperature data. For instance, if the temperature is between 10-20 then we made it as 20. Similarly, whole temperature column is transformed. Personal and Property crimes data has been normalized before plotting.

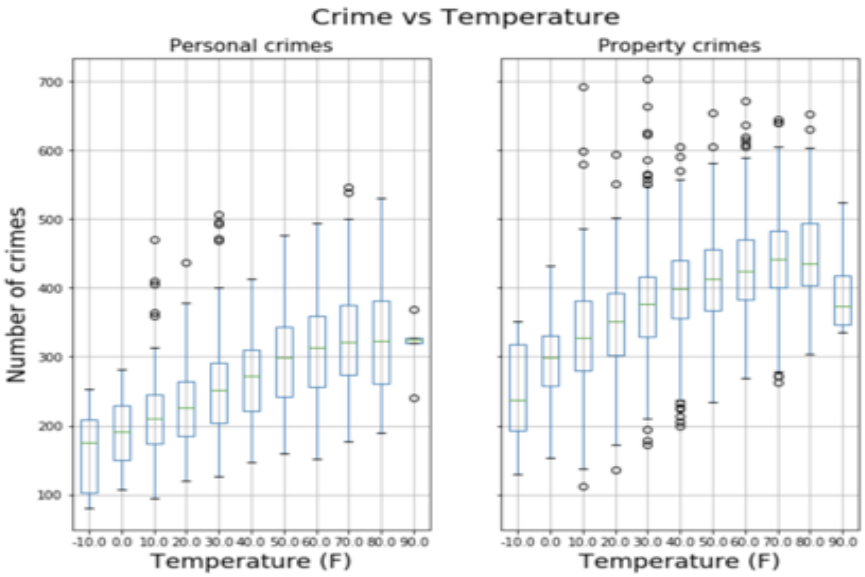

Fig-1 : Distribution of personal & property crimes with temperature

From Fig-1, we can say that mean of the crimes is increasing with temperature. And so we can come to partial conclusion that crimes are high during summer season. Seasonal changes can affect the crime rate. The question is does all types of crimes follow the same? The below analysis can give us some answers.

Next we have explored the correlation between temperature and some of personal and property crimes. From Fig-2, we can clearly state that there is high correlation between Battery and Assault, both of them fall into personal crimes, which indicates when batteries are high assaults are also high There is weakest correlation between Temperature and Humidity. We have hard-coded the diagonal values to zero that is Arson is perfectly correlated with Arson.

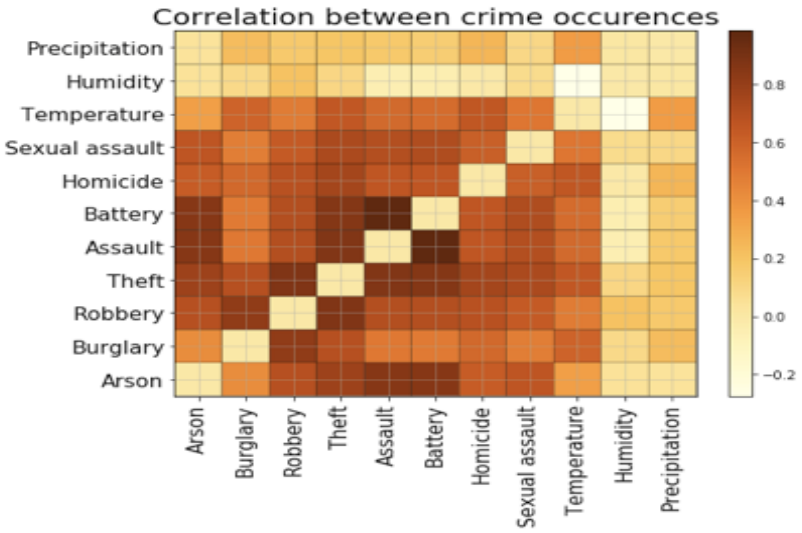

Fig-2 : Correlation between Temperature and few crimes

Let us know see the Relationship crime has with temperature as well as temporal factors hour of the day, day of the week, month of the year for both personal and property crimes. Fig-3 indicates that crimes are peak during noon hours, Fig-4 indicates that property crimes fall and personal crimes rise over the weekend, Fig-5 indicates that crime rate is high during summer months and fall during rest of the months.

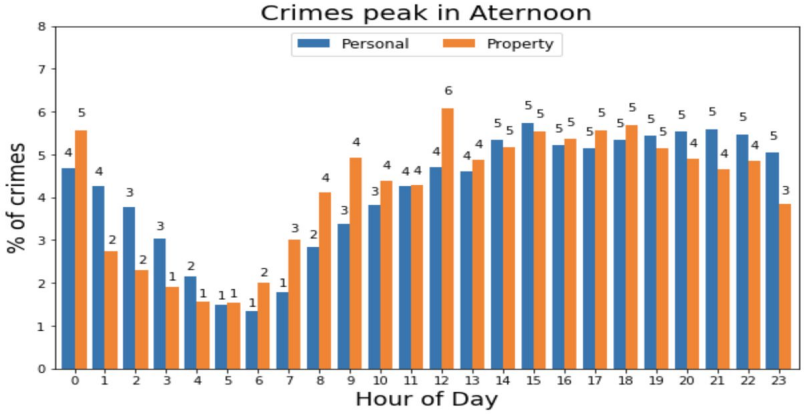

Fig-3 : Crime rates vs Hour of the day

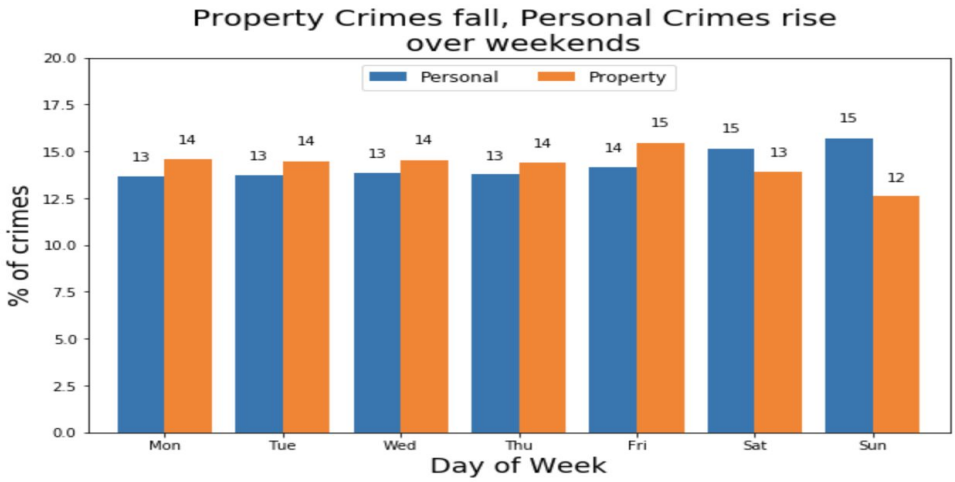

Fig-4 : Crime rates vs Day of week

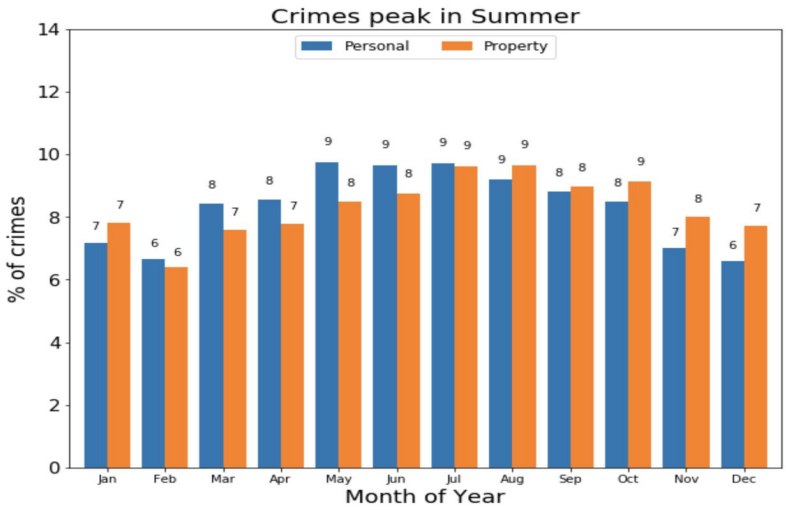

Fig-5 : Crime rates vs Month of year

As there is high correlation between Battery and Assaults, we shall dig deeper on both.Assaults are high during noon hours.High over the weekdays and less over the weekends

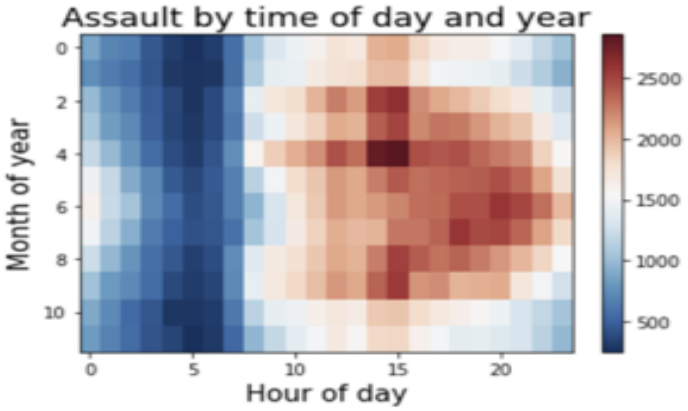

Fig-6 : Assaults with Month of Year vs Hour of Day

Now let us see the Battery crime rate with month of year versus hour of day. Fig-7 shows that crimes are high during early night hours starting from noon hours.

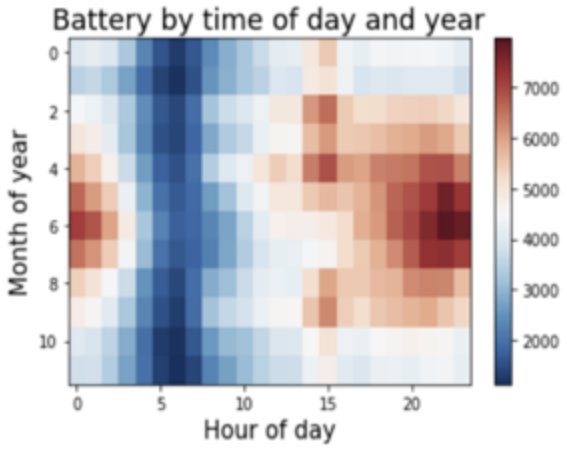

Fig-7 : Battery with Month of Year vs Hour of Day

Next, let us see the linear regression analysis for the recorded temperatures of each day and draw a fitting curve. The reason behind doing this is to check whether the same model fits the crimes data across the years and use the same model to do predictions for rest of the years.

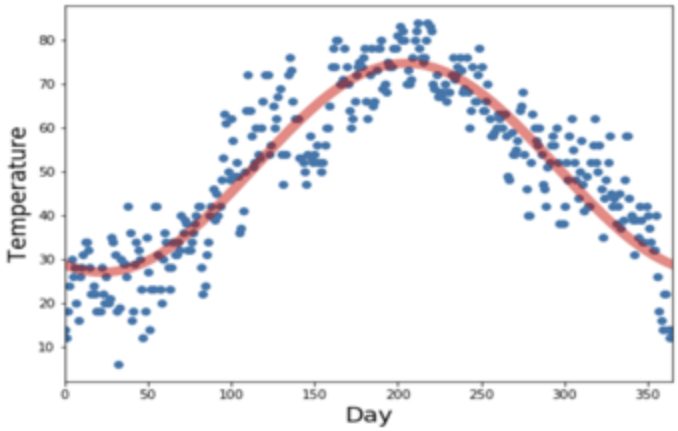

Fig-8 : Statistical model of temperature vs day of the year

Now, let us see how far the same model fits the observed crime data. From Fig-9, we can say that the model fits the observed crime data and as mentioned before crime rate is varying in accordance with temperature that is high during summers. The shaded portion of the graph indicates the prediction of crimes for the years 2014 to 2021. There is a continuous descend in the crime rate from 2000 to 2014 and we expect sudden fall in crime rate after the gun policy came into existence from 2013. Unlike our expectation there is no sudden fall in occurrence of crimes. I have chosen a assault crime and saw whether the model fits the data. The model is fitting the data but not overfitting.

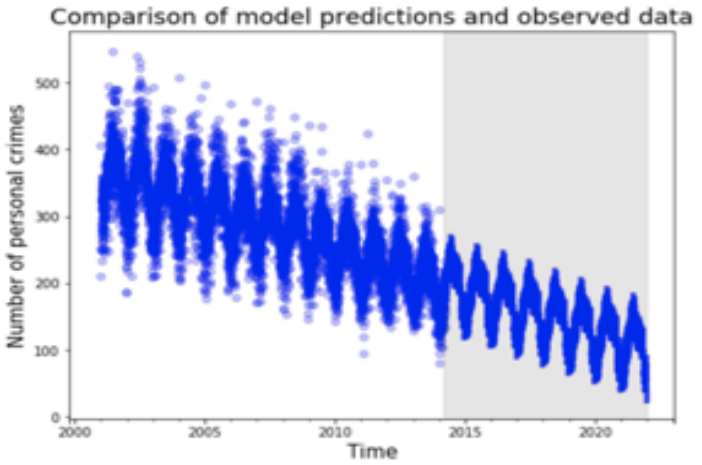

Fig-9 : Statistical predictions and Observed personal crime data

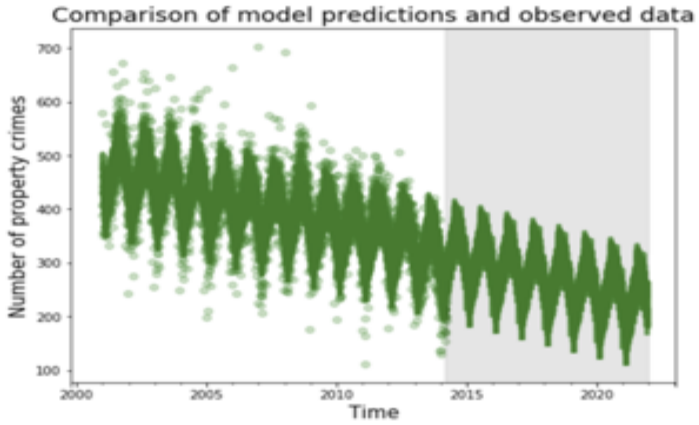

Fig-10 : Statistical predictions and Observed property crime data

Next step is to find the evidence of substantial differences in reporting crime data since 2013 for all the districts in chicago. From Fig-11 and Fig-12, we can say that both personal & property crimes rates across districts fall faster in 2013 when compared to 2010-12 percentage change.

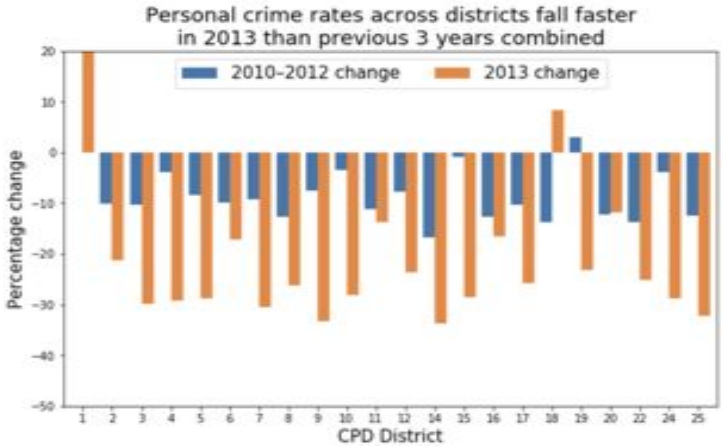

Fig -11 : Personal crime rates across districts

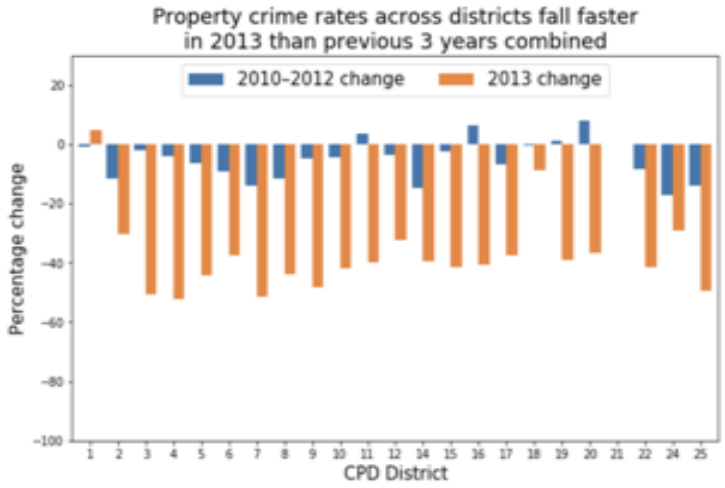

Fig -12 : Property crime rates across districts

There is a huge difference for district 15 for both personal and property crimes. Does that mean district 15 is safe district to live in chicago? Let us perform **ttest_1samp** on each district with one column having percentage change for years 2010 - 12 and other column with percentage change from year 2012 to 2013 and confirm that. Statistically it is shown that district 15 records faster fall in number of crimes. Does the p-values support the same? In this case, ttest results too conveyed the same i.e., district 15 has lower p-value and so we can say that district 15 in chicago is safe to live for most type of crimes when compared with others. Below are the results of the ttest_1samp of each district listed according to chicago police.

1 [2.49591088 0.03717818]
2 [1.83446617 0.10392335]
3 [4.63480444e+00 1.67751911e-03]
4 [4.02234258e+00 3.82864659e-03]
5 [4.41494397e+00 2.24156286e-03]
6 [-0.23363273 0.82113759]
7 [4.11362157e+00 3.37373940e-03]
8 [2.84320227 0.02170478]
9 [4.40683339e+00 2.26596900e-03]
10 [5.69487012e+00 4.57076333e-04]
11 [8.83196796e+00 2.12806818e-05]
12 [4.71291676e+00 1.51595611e-03]
14 [4.16585517e+00 3.13989412e-03]
15 [1.03501082e+01 6.56125307e-06]
16 [3.31174059 0.01067051]
17 [7.36932476e+00 7.84683815e-05]
18 [4.43391307e+00 2.18558593e-03]
19 [5.72493732e+00 4.41519622e-04]
20 [6.61927931e+00 1.66020060e-04]
22 [4.06783246e+00 3.59419989e-03]
24 [2.60793459 0.03123036]
25 [4.40466529e+00 2.27254184e-03]

Conclusions:

From the data analysis we performed, we can conclude that crimes are high in summer season which satisfies our initial hypothesis. From the gun policy change in Chicago since 2013 there is a decreasing trend in the crime rate but not substantial. It is evident from the statistical modeling done across years from 2001 to 2021. Crime rate fall faster in 2013 when compared to previous three years in which we have noticed there is huge change for 15th district. And ttest results show the same and from which we stated that 15th district is safe to live in Chicago city.

We have performed data analysis as much as we can do which convey same results for a layman person on first glance of figures in results section. If we would have more time, then we may dig deeper into the data. From this project, one can get familiar with pandas, matplotlib,

numpy, sklearn, scipy, regression analysis - statistical modelling - curve fitting, hypothesis testing,

References:

1. https://blog.weatherops.com/how-does-the-weather-affect-crime-rates
2. https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-present/ijzp-q8t2
3. https://www.chicagotribune.com/news/data/ct-crime-heat-analysis-htmlstory.html
4. https://stackoverflow.com/questions/35788140/scipy-stats-ttest-1samp-hypothesis-testing-for-comparing-previous-performance-to
5. http://blog.minitab.com/blog/adventures-in-statistics-2/how-to-correctly-interpret-p-values