ANALYZING MINNEAPOLIS CRIME DATA OVER TIME

Lavanya Basava Raju, Sridevi Bayyapuneedi, Heming Chen, Yuefeng Pan, Ryan Borowicz
August 24, 2016

Contents

1.0 Introduction

1.1. Background

The topic of crime and public safety casts a wide net, touching upon the lives of all citizens in an area. We each pay taxes at a local and state level to fund our police departments and their efforts in the hopes that crime will be reduced and our neighborhoods will be safe. New citizens to the area have a unique challenge, without previous knowledge of the area or its history. In an effort to inform new and current citizens of the historical crime patterns as well as provide predictive measures of future crime, we set out to perform an analysis of a variety of crime related factors. Through this analysis we will look to present visual and numeric historical crime trends, provide a tool for citizens to use to rate crime in an area, and describe a predictive model that can be used to forecast future crimes by area.

1.2. Research Questions

To gain a better understanding of the crime data we first wanted to analyze the historical crime trends, and identify any relevant patterns important to consider in the analysis. With these identified, we moved into predicting future crime rates, and incorporating a crime scoring system that citizens can use to gauge the safety of a neighborhood. Our final goal is to utilize social media and text analytics to gain insight into crime patterns. Below we detail the specific research questions that will be of interest in this paper.

1.2.1. How has crime varied over time in Minneapolis?

For this question we want to look broadly at how crime has changed over the past five years by crime type.

1.2.2. What types of crimes are most prevalent in each neighborhood?

This question looks to evaluate the frequency of certain crime types by neighborhood.

1.2.3. Is there more crime during the week or on weekends?

We want to identify whether there are differences in crime rates between weekdays and weekends, with weekdays being defined as Monday-Thursday and weekends Friday-Sunday.

1.2.4. Does crime vary by season of the year?

We would like to identify any seasonal discrepancies in crime volume, with the seasons being classified as Fall (September – November), Winter (December – February), Spring (March – May), and Summer (June – August).

1.2.5. Can we create a crime score rating system to allow users to compare neighborhood safety?

This question involves the creation of a crime score rating system that will allow users to input their preferences and return an overall crime score based on their entries.

1.2.6. Can we predict future crime rate types and locations?

We will look to build a forecasting model that can predict crime amounts by type, and will provide results for the next six months as a test baseline.

1.2.7. What insights can we gain by analyzing crime through social media sentiment analysis?

Utilizing social media text as the source, we will try and glean insights from the data, and determine whether negative sentiment is associated with increased crime rates.

1.3. Data Description

Our primary data source was Open Data Minneapolis, which is a publicly available website that has yearly spreadsheets of all crimes occurring in Minneapolis neighborhoods dating back to 2011. The data is at a detail level, and provides information about the date, time, location, and type of crime committed. Some of the primary dataset values we included in our analysis include:

- **BeginDate** – This is the actual date/time of the crime incident
- **ReportedDate** – This is the date/time the crime was reported
- **Offense** – This provides the type of crime (i.e. SHOPLF = shoplifting, AUTOTH = auto theft)
- **Precinct** – This is the number of the precinct of the officer responding to the incident
    - Minneapolis has five precincts broken out by city geography (1 – Downtown, 2 – Northeast, 3 – Southeast, 4 – Northwest, 5 - Southwest
- **UCR Code** – Summary reporting classification of the type of crime
- **Neighborhood** – Minneapolis neighborhood the crime was committed in

The final dataset was consolidated from the yearly spreadsheets into one Microsoft Excel CSV file prior to loading in Python. We submitted an open data request to obtain information regarding the specific number of police officers assigned to each precinct, sector, and neighborhood, but did not receive results prior to the submission of this paper.

2.0 Related Work

Crime rate is a major public safety concern, and many scholars and researchers have been doing related studies to reveal the scientific way to analyze and predict the future crime rates. A few key relevant studies are highlighted below,

Analyze crime trends over times and neighborhood

Back in 1979, two sociological scholars from University of Illinois, Lawrence Cohen and Marcus Felson (Cohen and Felson, 1979), worked out a theory called "Routine Activity" approach, for analyzing crime rate trends and cycles. The model predicted that criminal events result from suitable targets, capable defenders, and likely offenders in a nonrandom time and space. However, the previous research was only relying on individual- and household-level data to test that premise. Our model differs by being built heavily upon exploratory and time-series analyses from government disclosure data source. In addition, the results prove the fact that crime distribution is not random, and that time and space factors are highly correlated with the crime.

Analyze crime varied by day of the week and season

Cohn and Rotton (2000) analyzed the relationship between day of the week, seasonal trends and property crimes in Minneapolis between 1987-1988. Only three types of crime are analyzed in his work: theft, burglary and robbery, which are also included in our analysis. Cohn and Rotton applied hierarchical regression analysis as their main methodology, and the result indicates that for these three types of crimes, weekends had an average higher frequency than weekdays, however, the frequency decreased sharply on Sunday, which is consistent with our result. One difference in our study is that we treated Friday as a weekend day in our analysis, and we did find some significant results for that day. In regards to the seasonal impacts, Cohn and Rotton list summer as the most dangerous season as it witnesses the highest number of any season, with winter having the lowest number. The rank of seasons in their report is consistent with the analysis we performed.

Perform sentiment analysis on social media data to identify patterns that can be used in crime prediction

Scientists from Google Inc., and Stony Brook University, Namrata Godbole, Manjunath Srinivasaiah and Steven Skiena, have established the large-scale sentiment analysis theory for news and blogs (Godbole, Srinivasaiah and Skiena, 2007). By presenting a system which assigns scores to positive and negative words and opinions to each elements in the news and blogs, it's getting viable to quantify the optimism and pessimism feeling through words. This rating system built the foundation of our study, which intends to learn about the public safety in the Minneapolis area based on the comments and articles posted on the local blogs. By assigning each comment a grade based on the positive and negative words through year 2011 to 2015 on the blog, we could track the yearly trend of opinion from residents in Minneapolis area about public safety.

3.0 Process and Results

3.1. Data Preparation/Cleansing

The Open Data Minneapolis spreadsheets were relatively clean datasets, but there were still a number of data cleaning tasks that needed to occur prior to analyzing and working with the data. These steps included:

- Formatting Date and Time variables to suit the analysis
- Creating segments from the crime types in an effort to ease the interpretation of the results.
    - The data had 36 crime types, which we classified into 8 summary types based on the description of the crime types. The final 8 types included:
        - Homicide
        - Rape
        - Domestic assault
        - Theft of vehicle
        - Theft of place/property
        - Theft of person

- Other theft
- Other crimes

- Creating a variable to identify crimes by weekdays and weekends, which we classified as Monday through Thursday as Weekdays and Friday through Sunday as Weekends
- Creating a variable to identify crimes by season, classifying December through February as Winter, March through May as Spring, June through August as Summer, and September through November as Fall.

From this cleaned dataset we were able to begin analyzing the data.

3.2. Process and Results

How has crime varied over time in Minneapolis?

We first analyzed the crime rate trends over the last five years to obtain a general picture of the types of crimes and the frequencies at which they occurred. This can best be viewed in the figure below, which visually describes the up and down nature of crime rates over the years.

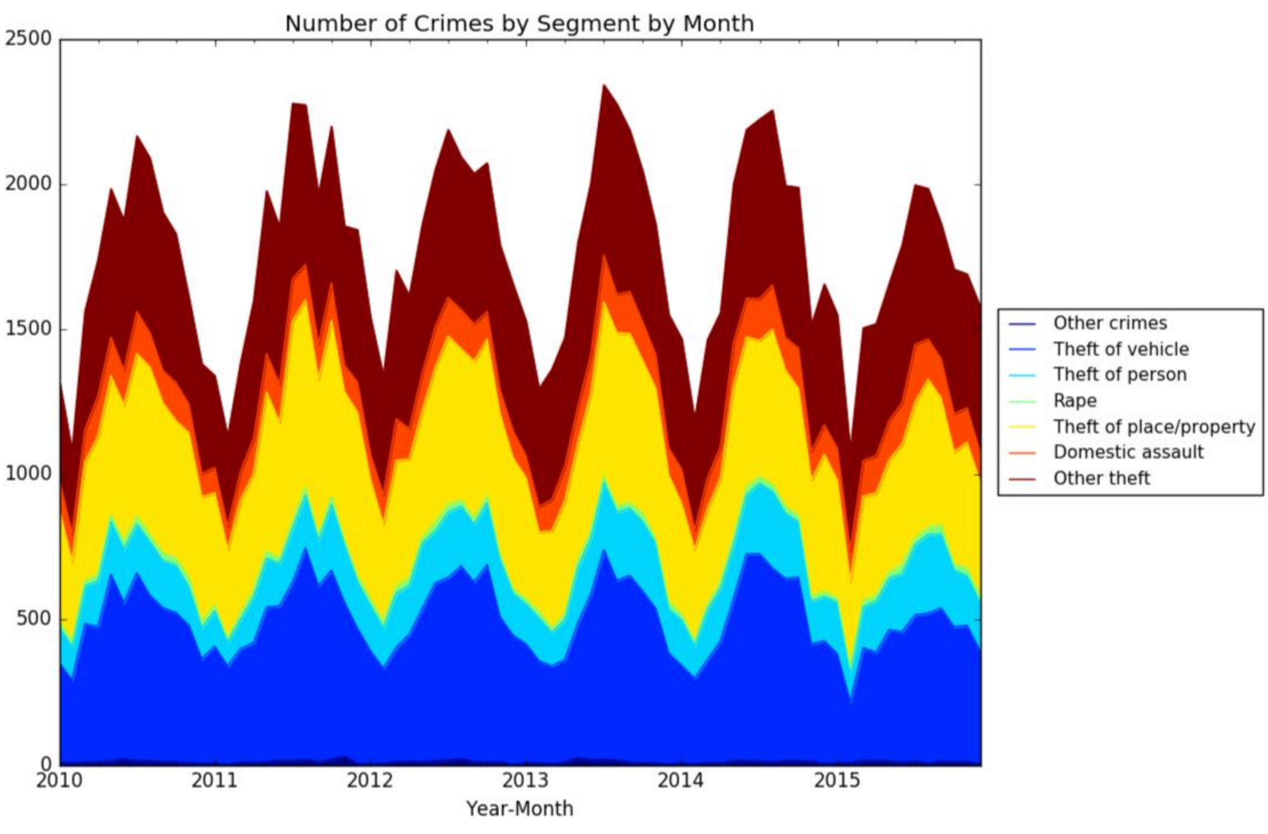

FIGURE 1 - NUMBER OF CRIMES BY CRIME TYPE

What is clear from the graph is that overall crime rates have a cyclical pattern, with surprising similarities in the curves from year-to-year. The interesting follow-up question that arrives is what is the underlying reason for these striking similarities in year-over-year crimes? There are clearly seasonal and monthly variations within each type that we will hope to explain in our further analysis below.

What types of crimes are most prevalent in each neighborhood?

There are 86 different neighborhoods that the Minneapolis Police Department is responsible for. The police force is divided into five precincts separated by geographic area, individually covering the downtown, northeast, northwest, southeast, and southwest areas of the city. These precincts are then further divided into 15 sectors which include the 86 neighborhoods. The original data is a bit misleading, however, due to the way it is classified. The police report this data based on the precinct of the officer who ends up responding to the incident. This could be an officer from the same precinct as the incident, but often it is an officer from a separate precinct that is just closer to the location at the time. To address this, we created a mapping of precinct to sector to neighborhood that would provide us the geographic information we needed. To visually analyze the changes over time we then summarized to the precinct level, and created the heat map shown below.

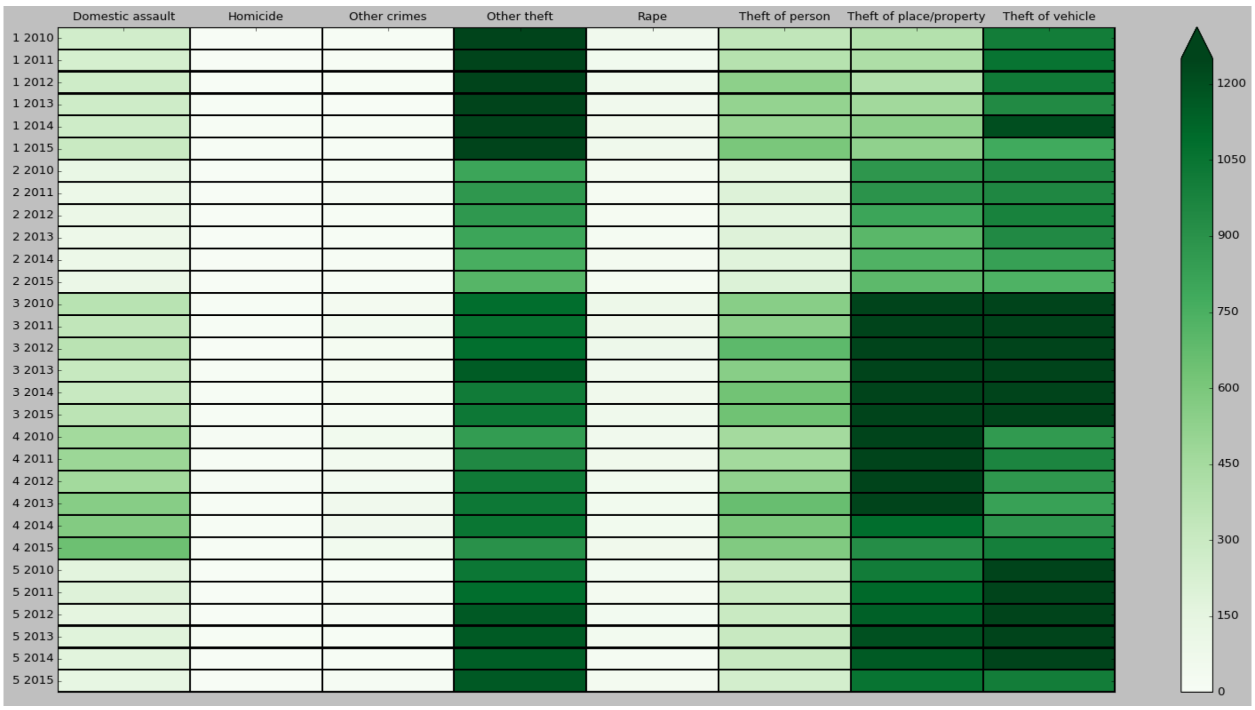

FIGURE 2 - CRIME TYPES BY PRECINCT

This map indicates higher crime rates in darker colors, showing the higher crime rate figures for "Other Theft" in precinct 1, and theft of vehicle/place/property in precincts 3,4, and 5.

Drilling down into the detail we can see that the neighborhoods with the highest crime rates leading to the overall results include Downtown West (precinct 1 - Downtown), Whittier (precinct 5 – SW), Jordan (precinct 4 – NW), Longfellow (precinct 3 – SE), Near-North (precinct 4 – NW), and Marcy Holmes (precinct 2 – NE). Downtown West has the highest crime rates, with total crimes over three times greater than the next closest community, Whittier.

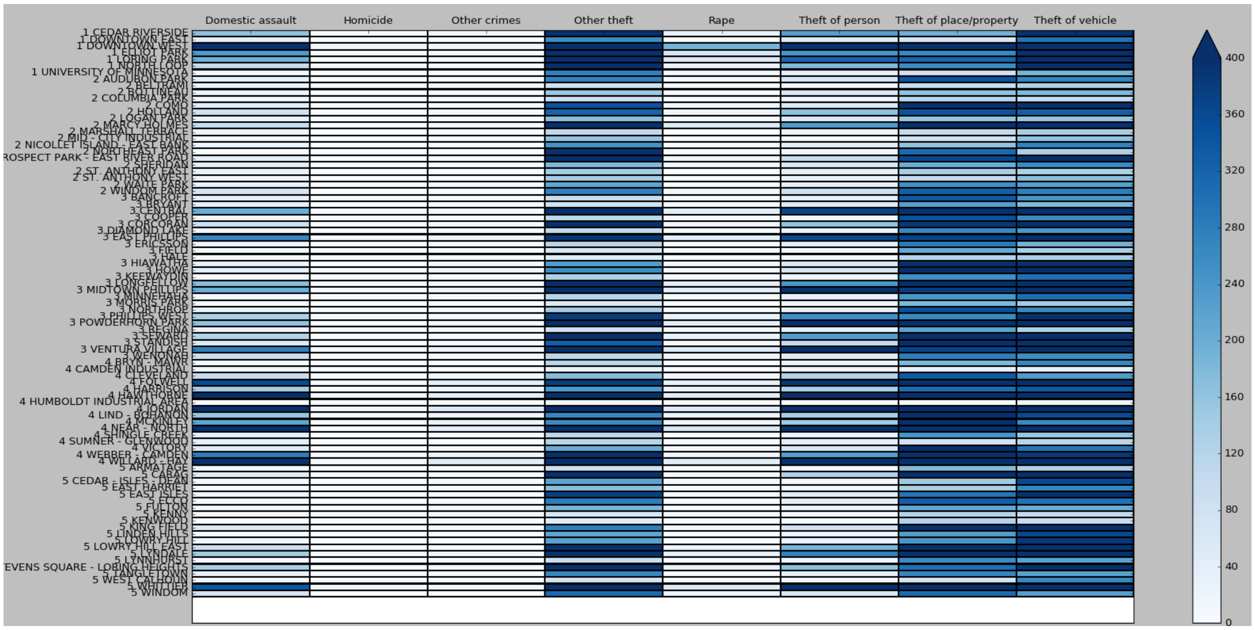

FIGURE 3 - CRIMES BY NEIGHBORHOOD

One interesting finding is that of the neighborhoods with the highest crime rates each seem to have a unique type of crime that is disproportionately associated with the area. For example, Downtown West has an extremely high amount of other theft, Whittier motor vehicle theft, Jordan burglary of dwelling, Longfellow shoplifting, and Marcy Holmes bike theft.

Is there more crime during the week or on weekends?

We initially hypothesized that there would be more crime on weekends relative to weekdays. This was proved partially correct, as the highest crime rates across all crime types were seen on Fridays, particularly for theft and assault. However, Sunday produced the lowest crime rates, so the overall weekend numbers average out similarly to the weekdays.

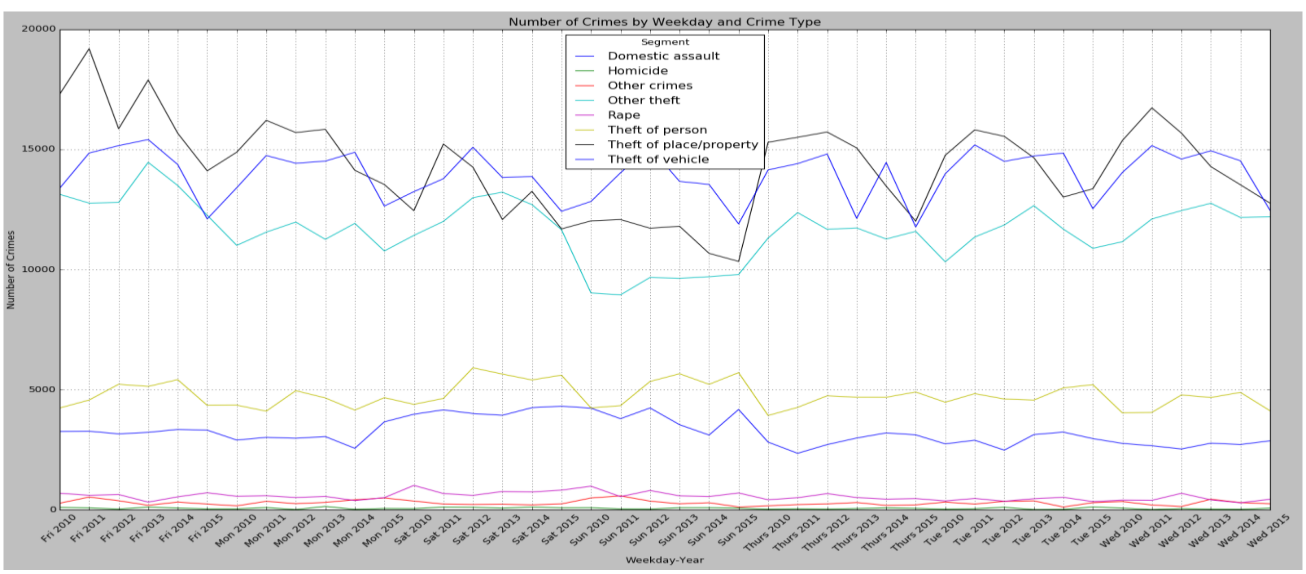

FIGURE 4 - CRIMES BY DAY OF WEEK

Does crime vary by season of the year?

The data indicate a clear relationship between the season of the year and crime rates. Ranking in order of highest crimes are:

1) Summer
2) Fall
3) Spring
4) Winter

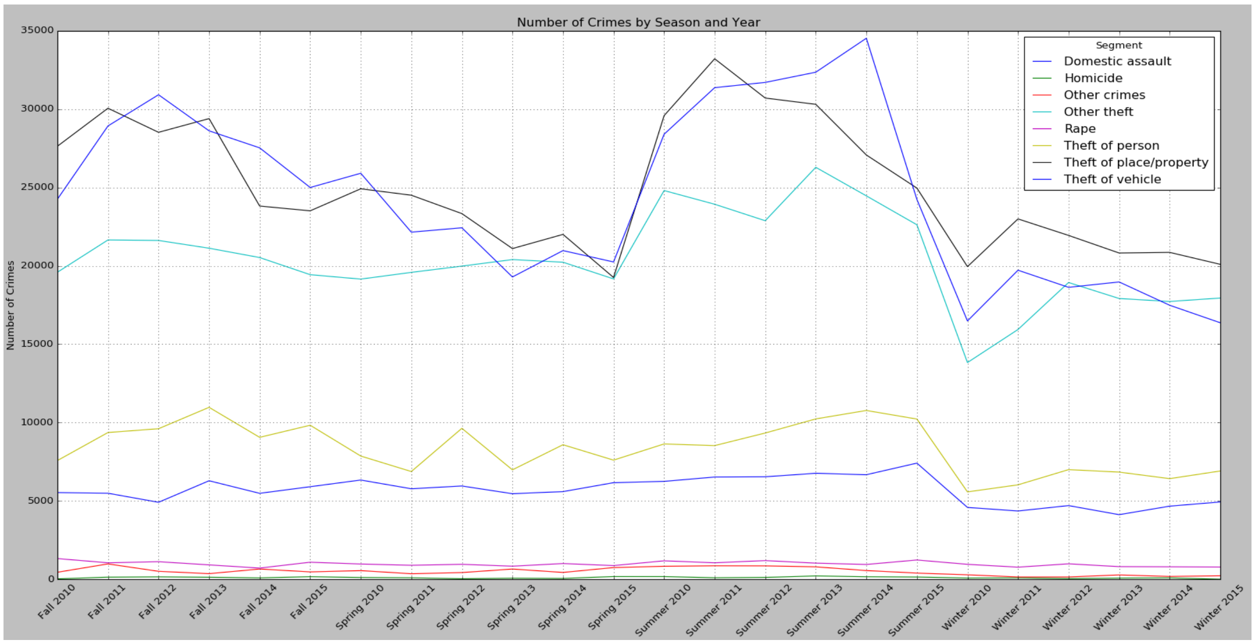

FIGURE 5 - CRIMES BY SEASON

Can we create a crime score rating system to allow users to compare neighborhood safety?

To create a crime score rating system, we needed to consider the factors that determine people's feelings of safety. The best indicator we identified was the total number of crimes and the types of crimes in a particular neighborhood. To address these issues, we first normalized the volume of crimes to give equal weighting to each type, and provided the user an option to input the specific values that are important to them. For example, let's take the example of someone who is only concerned about vehicle theft. This person would enter in the weightages per below:

```
Find below the weightage associated with crime types for the Neighbourhood score :
Homicide - 0.2
Rape - 0.2
Domestic assault - 0.15
Theft of vehicle - 0.1
Theft of place/property - 0.1
Theft of person - 0.1
Other theft - 0.1
Other crimes - 0.05

Enter the weightages :

Enter weightage for Homicide (0.2) :0

Enter weightage for Rape (0.2) :0

Enter weightage for Domestic assault (0.15) :0

Enter weightage for Theft of vehicle (0.1) :1

Enter weightage for Theft of place/property (0.1) :0

Enter weightage for Theft of person (0.1) :0

Enter weightage for Other theft (0.1) :0

Enter weightage for Other crimes (0.05) :0
```

After the user inputs their preferences, our program will produce a separate visual display of the top 10 safest and dangerous neighborhoods according to our base values and to their user defined preferences, similar to the graphs below.

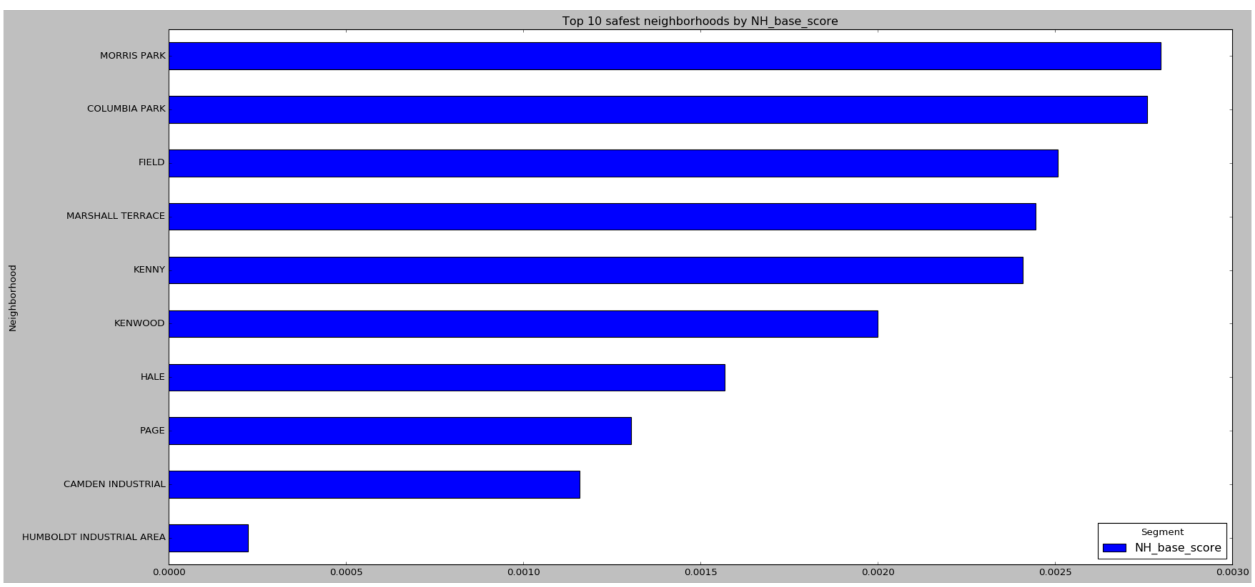

FIGURE 6 - BASE SCORE - SAFEST NEIGHBORHOODS

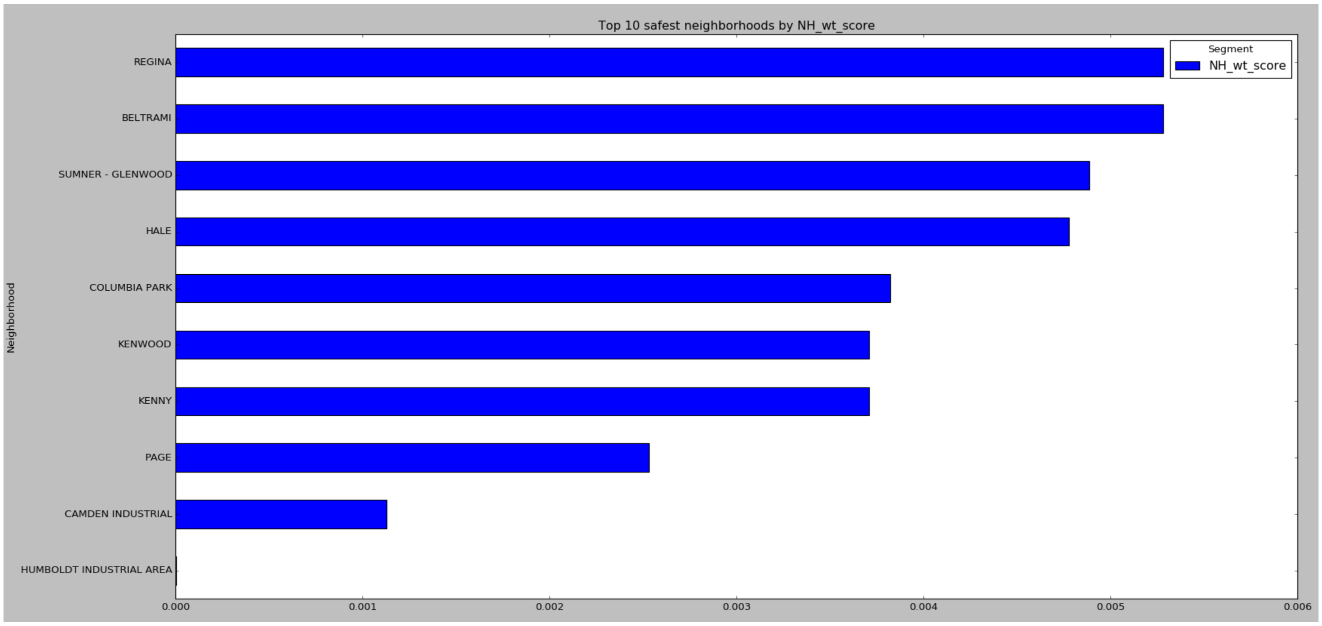

FIGURE 7 - WEIGHTED SCORES - SAFEST NEIGHBORHOODS

Similarly, the user can see the most dangerous neighborhoods for both the base score and their weighted score.

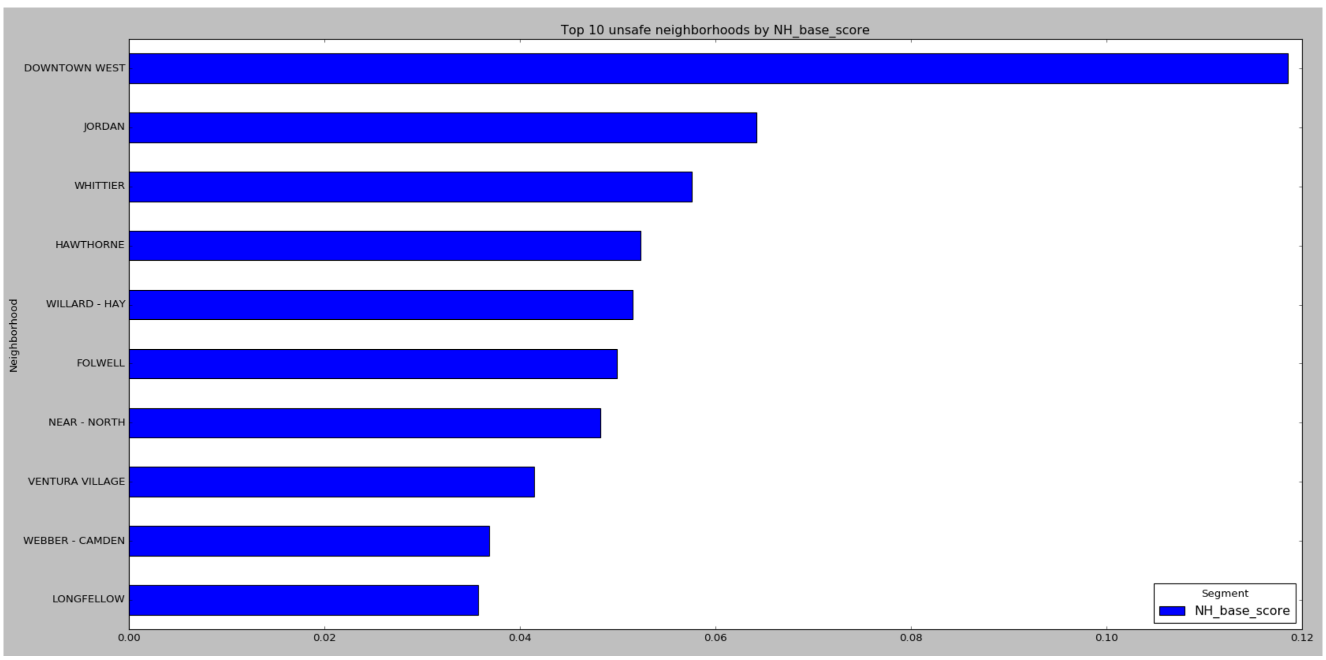

FIGURE 8 - BASE SCORE - MOST DANGEROUS NEIGHBORHOODS

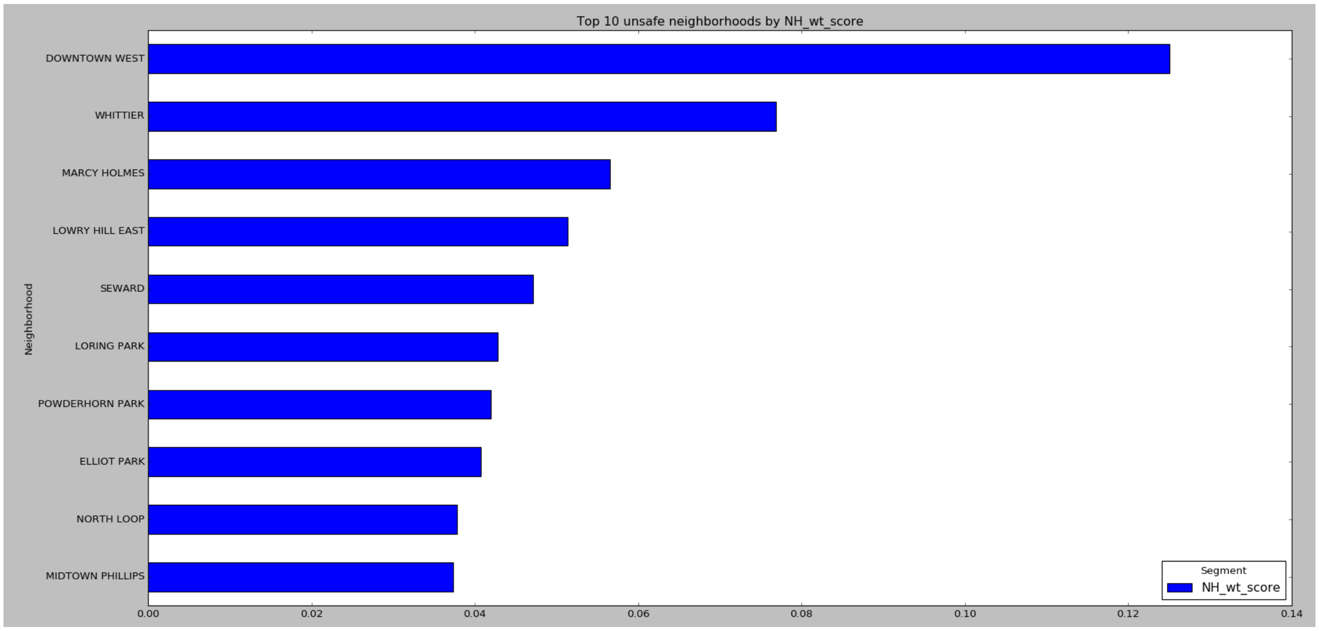

FIGURE 9 - WEIGHTED SCORE - MOST DANGEROUS NEIGHBORHOODS

Can we predict future crime rate types and locations?

The nature of predictive modeling is to learn from the past and see into the future. Essentially, we use statistical models so that we can fit the data into a pattern of behavior and anticipate future results.

Several predictive policing methods are currently in use in law enforcement agencies across the United States, and they allow police to work more proactively with limited resources. The objective of these methods is to develop effective strategies that will prevent crime or make investigation efforts more effective.

The objective of our research and project was to understand the pattern in the number of crimes in Minneapolis and predict the number of crimes for a particular crime type. We created a predictive model to utilize the historic crime numbers and were able to establish a significant relationship between time and the crime rate at month level. The best fitted model we were able to derive for the crime type – "Theft of Vehicle" at monthly level was,

Number of Crimes = 230 + (-58) (February) + 44 (March) + 40 (April) + 147 (May) + 148 (June) + 203 (July) + 168 (August) + 137 (September) + 148 (October) + 45 (November) + 8 (December) + 1.39 (Trend/Time index) + (-0.025) (Square of Trend) + 0.35 (No of Crimes – Previous Month)

The above model equation translates to the following,

1. In the presence of none of the other factors (Trend, Previous crimes, etc.), Minneapolis will have 230 auto thefts.
2. The month of February has 58 less crimes than the month of January, the month of March has 44 more crimes than the month of January, and so on; considering all other factors fixed/constant.

3. There is a second order relationship between the number of crimes and time index (square of Trend) which means the number of crimes follows a curvilinear relationship with respect to time rather than strictly a linear relationship, considering all other factors fixed/constant.
4. There is a significant relationship associated with the number of auto thefts that happened in the previous month.

We can see that number of crimes from November to April is much lower than the crimes between May and October. The number of crimes are correlated to the number of crimes from previous months. The model explains 86% of the variance in the number of crimes. The graph below shows the actual number of auto thefts (blue) against the predicted number of auto thefts (red).

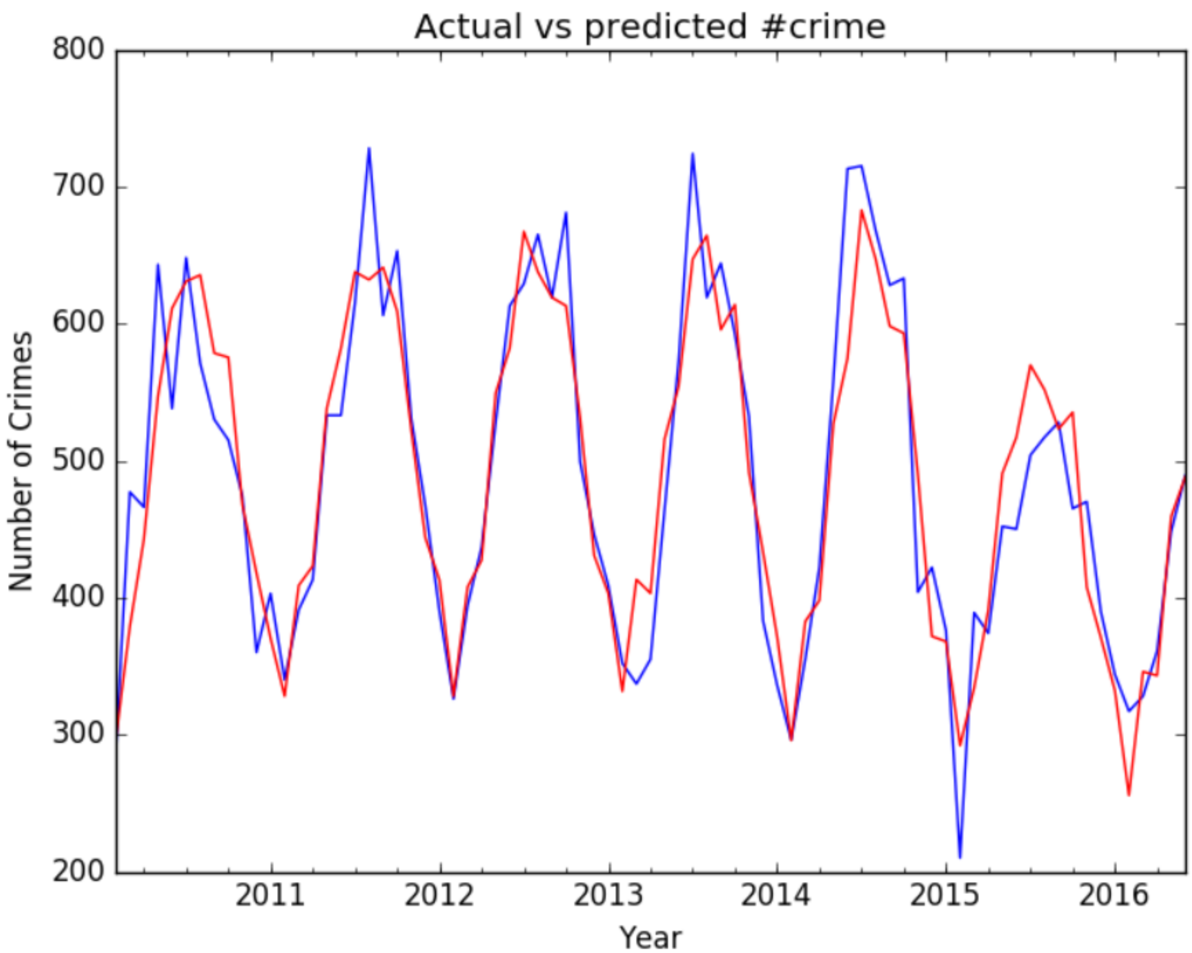

FIGURE 10 - CRIMES - ACTUAL VS. PREDICTED

We can see that the predicted values are not exactly overlapping with the actual values. This indicates that there is still 14% of the information in the number of crimes that we are unable to explain with the existing model. This might be because of additional factors not included in the analysis such as weather, neighborhood, economic factors, etc.

But from the existing model we can forecast the number of auto thefts for the months after June 2016,

	Year-Month	Crimes	Month	Trend	Lag_1_Crimes
Actual	2016-01	344	1	73	391
	2016-02	318	2	74	344
	2016-03	328	3	75	318
	2016-04	361	4	76	328
	2016-05	449	5	77	361
	2016-06	489	6	78	449
Forecasted	2016-07	**558**	7	79	489
	2016-08	**544**	8	80	558
	2016-09	**506**	9	81	544
	2016-10	**501**	10	82	506
	2016-11	**394**	11	83	501
	2016-12	**316**	12	84	394

FIGURE 11 - CRIME FORECAST

We can use this methodology to predict number of crimes by each crime type and by neighborhood as well.

What insights can we gain by analyzing crime through social media sentiment analysis?

To analyze crime sentiment through social media we first needed to identify our information sources, then create a set of dictionaries to classify words based on their crime sentiment, and finally come up with a programmatic approach to analyzing the data and providing results. We considered a number of different sources of information, including Twitter, Facebook, and relevant Minneapolis blogs. Twitter was our preferred approach, but due to the costs of obtaining historical Twitter data, this was deemed out of scope for this project. The dataset we decided on was the topix.com Minneapolis public blog. We wanted to test our hypothesis that negative sentiment was associated with increased crime rates. Our test dataset included the first ten posts and included threads from the blog for the months of May 2015 and May 2016. Comparing year over year, we can see that sentiment is far more negative in 2016 than 2015. A lot of this can be attributed to the posts related to the political events and the Minneapolis police shootings occurring at that time. While overall crime was down in May 2016 compared to 2015, there are certain neighborhoods and crime types that experienced a rise. These included the number of aggravated assaults increasing by 27, and total crime in the Downtown, Near-North, and East Phillips rising by double digits. While there could be a correlation between the negative sentiment and these increases, we believe additional analysis is needed to make such a causal conclusion.

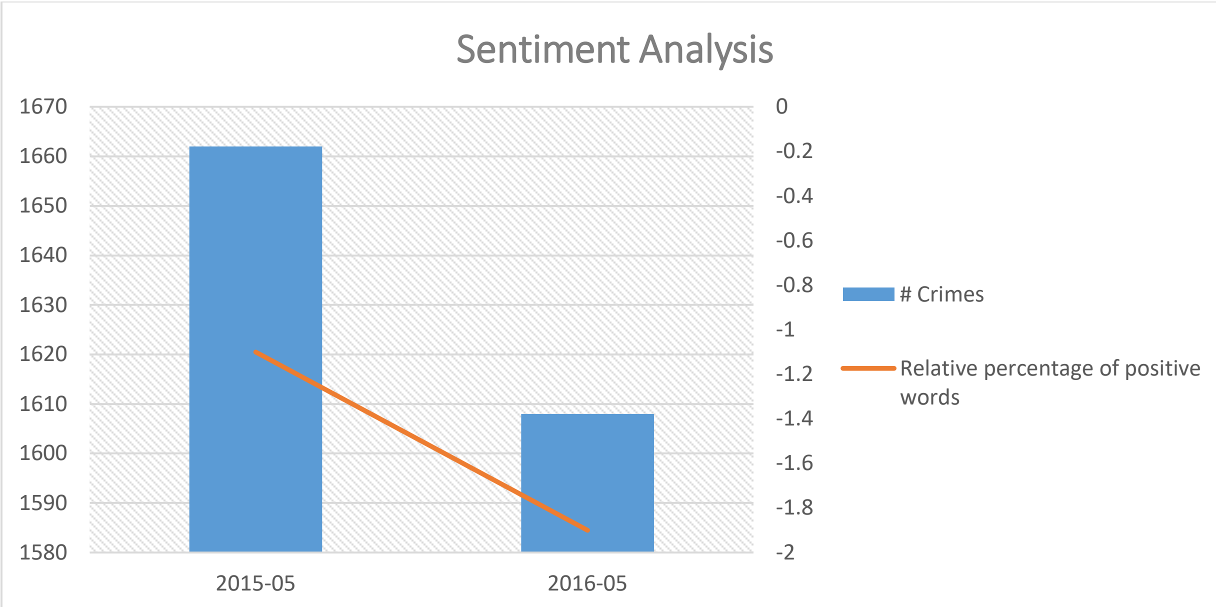

FIGURE 12 - SENTIMENT ANALYSIS - CRIME VS. SENTIMENT

4.0 Conclusions

Our analysis of Minneapolis historical crime data produced a number of interesting findings. The key conclusions related to our research are detailed below.

4.1. Minneapolis crime volume is cyclical, and has remained steady over time

4.2. Neighborhoods typically have one type of crime that is disproportionately experienced in that area

4.3. Crime rates are generally highest on Fridays, and lowest on Sundays

4.4. Crime rates are typically highest in the summer, followed by fall, spring, and winter

4.5. Our crime score rating system can help people decide which Minneapolis neighborhood is safest for them based on their preferences

4.6. Our forecasting model can explain 86% of the variation in crime volume, but could be improved through capturing additional factors

4.7. Sentiment analysis can potentially provide meaningful insights into future crime rates, but more research is needed to justify a causal conclusion

Despite the conclusions above, there were a number of areas that we weren't able to fully detail during the course of this project that could be further researched. Specifically, we would like to look into a procedure for automating the import of website blog data into Python to avoid having to manually create the text files. This would greatly reduce the processing time needed to analyze the data. In addition, we could set up a procedure to store Twitter data over time and use this data to compare against historical crime rates. One additional research area that piqued our interest was that of applying Benford's law to the monthly crime statistic numbers published by the Minneapolis police department. It seemed strange that crime patterns would be so consistent each year across months, and it could be hypothesized that police departments have similar incentives to CEO's in keeping crime numbers flat or slightly decreasing in a similar fashion to companies always reporting slight increases in earnings each quarter.

5.0 References

Corman, H., & Mocan, H. N. (2000). A Time-Series Analysis of Crime, Deterrence, and Drug Abuse in New York City. American Economic Review, 90(3), 584-604. doi:10.1257/aer.90.3.584

Police Incidents 2015. (2016). Retrieved August 6, 2016, from http://opendata.minneapolismn.gov/

Godbole, N., Srinivasaiah?, M., & Skiena, S. (2007). Large-Scale Sentiment Analysis for News and Blogs. Retrieved August 21, 2016, from http://www.uvm.edu/~pdodds/files/papers/others/2007/godbole2007a.pdf

Criminal Justice Data Reporting. (n.d.). Retrieved August 20, 2016, from https://dps.mn.gov/divisions/bca/bca-divisions/mnjis/Pages/uniform-crime-reports.aspx

Cohn, E. G., & Rotton, J. (2000). Weather, Seasonal Trends And Property Crimes In Minneapolis, 1987–1988. A Moderator-Variable Time-Series Analysis Of Routine Activities. Journal of Environmental Psychology, 20(3), 257-272. doi:10.1006/jevp.1999.0157

(2016). Retrieved August 13, 2016, from https://raidsonline.com/?rms=Minneapolis_Crime_Map

Various. (n.d.). Retrieved August 17, 2016, from http://www.topix.com/forum/city/minneapolis-mn