Analysis and Visualization of City Crimes

Mohamed Badawy, Jared Cirelli, Howie Setyono and Faisal Aqlan*
Industrial Engineering Department
Penn State Erie, The Behrend College
Erie, PA 16509, USA
*FUA11@psu.edu

Abstract

Crimes are rampant across cities and towns throughout the world. Using crime analytics allows law enforcement agencies to pinpoint areas with high crime rates and determine methods to reduce them. Analytical models can be used to predict and visualize crimes so that they can be prevented, before they happen. This paper presents a case study on crime analysis and visualization in Erie City, Pennsylvania, USA. Crime data was obtained by the Erie Police Department. Data was pre-processed to remove the outliers, fix invalid addresses, and calculate the longitudes and latitudes. Descriptive analytics was developed to analyze the crimes per crime type and region and develop heat maps for the crime distribution. Two specific areas that have high crime rates were further investigated. The results provide decision makers with valuable insights into crime prediction and prevention. Cameras were installed in the areas with high crimes rate.

Keywords
Crime analysis, crime visualization, data mining, heat maps

1. Introduction

Crime in short is breaking the law. The law, is made up by the government that the citizens must adhere to. When a citizen breaks the law, they become a criminal and will be tried in court and punished according to the severity of the crime. Crime detection process' focuses on uncovering criminal activity or verifying reported crime and acquiring evidence to identify and prosecute its perpetrators. Crime detection can happen through different ways of data gathering such as phone tapping, DNA and fingerprints analysis, etc. The common methods that are known to be still effective in reducing crimes are: educating the public on crime prevention, balancing between prevention and repression, and helping former prisoners adjust to a normal life. As technology advances every day, it becomes a double-edged sword for both criminals and law enforcements to use. Furthermore, new crimes that were considered unheard of many years ago have recently been on the rise. Many of these crimes fall under the category of cybercrimes which are crimes committed through the internet or a computer network. With these "21st century crimes" becoming more prevalent, criminals are getting smarter and more elusive. Law enforcement officers seem to have a few solutions to combat this growing issue. Data analytics and data mining techniques can effectively be utilized in crime detection and prevention. Data mining uses sophisticated mathematical algorithms to determine relationships and patterns in the data. This helps in creating several predictions that could offer a way to prevent future crimes from happening. Hot-spot map (also known as heat map) is a traditional method for crime analysis and visualization. In hot-spot maps, crimes are represented by color codes on geographic maps. Several other techniques are used to analyze and predict crimes such as linkage analysis, statistical analysis, profiling and spatial analysis. Linkage analysis identifies different behavioral patterns associated with crimes. Profiling and spatial analysis identify geographical locations and analyze human behavior patterns.

Several studies in the literature have discussed the use of data mining and analytics techniques to study and predict crimes. For example, spatiotemporally tagged Tweets were used for crime prediction in Gerber (2015). The study used Twitter-specific linguistic analysis and statistical topic modeling to automatically identify crimes. The relationship between happiness, crime victimization and neighborhood safety were investigated in Cheng and Smyth (2015). The study found that being a victim of crime and having an acquaintance who is a victim of crime, have a

negative effect on happiness. The relationship between crime and place in the countryside was studied in Mawby (2015). It was found that locational status of victims and offenders are key to understanding rural crimes.

Other studies discussed the use of data mining techniques to model the performance of crime scene investigators (Adderley et al., 207). The use of text mining for crime investigation was discussed in Tseng et al. (2012). The study presented and integrated term-relationship mining and exploration approach for crime exploration and investigation. Phillips and Lee (2012) used data mining techniques to investigate the co-distribution patterns in large crime datasets. A graph-based dataset representation was developed to allow for extracting patterns from heterogeneous areal aggregated datasets and visualize the resulting patterns efficiently. A study was performed to analyze and predict crimes using clustering and classification techniques (Kiani et al., 2015). Keyvanpour et al. (2011) developed a general crime matching framework to detect and predict crime based on data mining techniques.

In this study, we utilize data mining and analytics to analyze and visualize crimes. Analysis of crimes per crime type, time, and date is performed. Furthermore, heat maps are used to visualize the crimes in different areas. We also focused on two specific areas as suggested by the Erie Police Department.

2. Research Framework

The proposed framework for crime analysis and visualization is shown in Figure 1. Crime data usually includes a serial number for the crime as well as a set of crime attributes that include address, situation found, status, call-for-service number (CFS#), primary code, offensive code, disposition, and report date. A pre-processing step is required to remove (or impute) any missing data, identify and remove the outliers (if any), and calculate the longitudes and the latitudes of the crime addresses. The next step in the framework is to analyze the data and identify patterns. This includes identifying the number and percentage of crimes per location, crime types, disposition, etc. Heat maps are then developed to visualize the crimes. Heat map is a good tool for identifying the areas with high crime rates. The last step is to provide recommendations for reducing the crime rates.

Figure 1. Crime analytics framework

3. Case Study

Located in Northwestern Pennsylvania (see Figure 2) and named after the lake as well as the Native American tribe that settled at the south shore, Erie is the state's fourth-largest city. With just an area of 19.1 square miles on land, Erie has a population of approximately 99 thousand citizens (Duda, 2015). Erie's climate is like cities near the Great Lakes and is located in the Snowbelt which results in cold winters with heavy lake effect snow. Erie was named the snowiest city in the US during the 2013-2014 winter season (Duda, 2015). During the 20th century, Erie was an important city for iron and steel manufacturing. Due to multiple financial crises, many companies have either moved their facilities elsewhere or declared bankruptcy (ERIE, 2013). Today, however, companies such as General Electric and Erie Insurance are the only remaining major employers of the city's workforce (Wertz and Wood, 2013).

When compared to similar sized cities in Pennsylvania, New York and Ohio, Erie's violent crime rate is the lowest. However, when compared to the national average, Erie's crime rate is currently higher (see fig. 3). One reason often cited for the higher crimes in Erie is the economic stress which can lead to high unemployment and endemic poverty. Another reason that can be attributed to the higher crimes may be the size of Erie's police force. In 2003, Erie's police force had 214 officers. The number of officers went down to 167 in 2006 and then 161 in 2008. This may be the cause that Erie's crime rate exceeded the national average in 2006 and went up to a maximum value in 2008 (Jefferson Educational Society, 2015).

The data for the presented case study was obtained from the City of Erie Police Department. The data set represents reports where the police were called in, regardless if an arrest was made or not. The data includes over sixteen thousand data points for the period January 1st 2015 to March 28th 2016. The crime attributes include location of the

incidents, incident number, situation code, primary, offensive code, disposition, reporting date and time, and reporting code. Location of the incident is the address of the crime which includes apartment number and bar location. Incident number is the serial number of the crime. Latitude and longitude values were calculated based on the addresses. Situation code is a three or four letter code that represents the type of the crime, for example AAH is the situation code for harassment. Primary codes represent the general categories of the crimes. Offensive codes represent crimes from the Pennsylvania Crimes Code or are City of Erie city ordinance violations. Disposition is the final settlement of the crimes. The main crime disposition codes are: C for Closed by Arrest, E for Exceptional Clearance, CJ for Closed by Juvenile Arrest, U for Unfounded, and O for Open. Reporting codes are the codes of the officers who report the crimes.

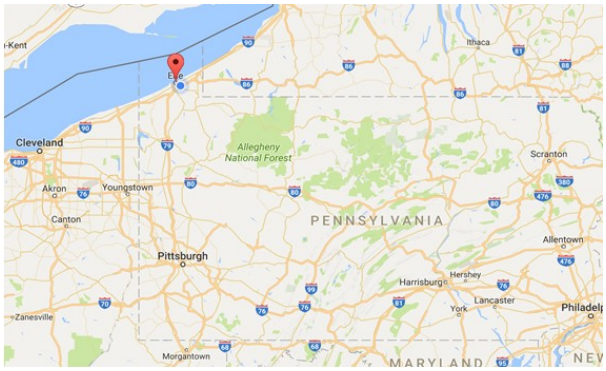

Figure 2. A google map showing location of Erie

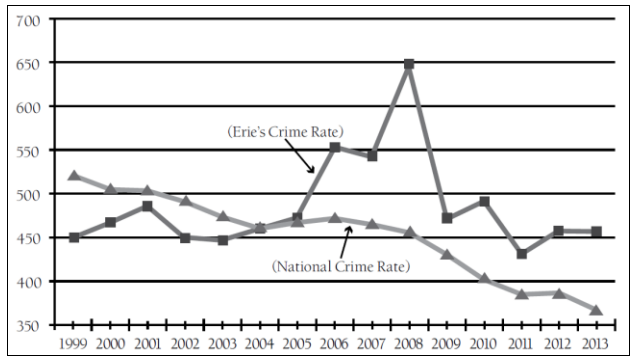

Figure 3. Erie's Crime Rate vs. National Average (Jefferson Educational Society, 2015)

The pre-processing step of the data consists of calculating the latitude and longitude of the addresses, removing the missing data, fixing or removing invalid addresses, and identifying the correct data measurement scale for each attribute. We used IBM SPSS Modeler and R software for analyzing the data and developing the visualization models. Fig. 5 shows the crime distribution per day of the week and fewer number of crimes occurs during the weekend. The reason for this is that all auto-incidents are considered as crimes and their frequency during the weekend is relatively small compared to the rest of the week.

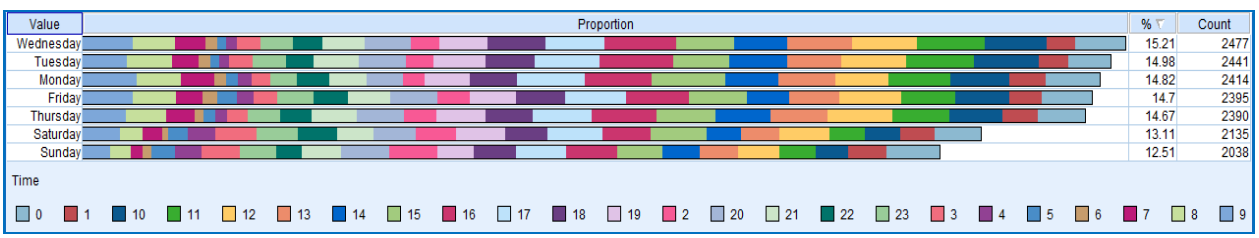

Figure 4. Crime distribution per day (colors represent crimes per hour of the day)

Figure 5 shows the distribution of the crimes per crime type and disposition. Most auto-accident crimes are not assigned disposition codes.

It can be noted that most of the theft crimes are "Open" whereas most of the drunkenness crimes are "Closed". The codes of the crime deposition are as follows:

- C → Closed by Arrest
- E → Exceptional Clearance
- CJ → Closed by Juvenile Arrest
- U → Unfounded
- → Open

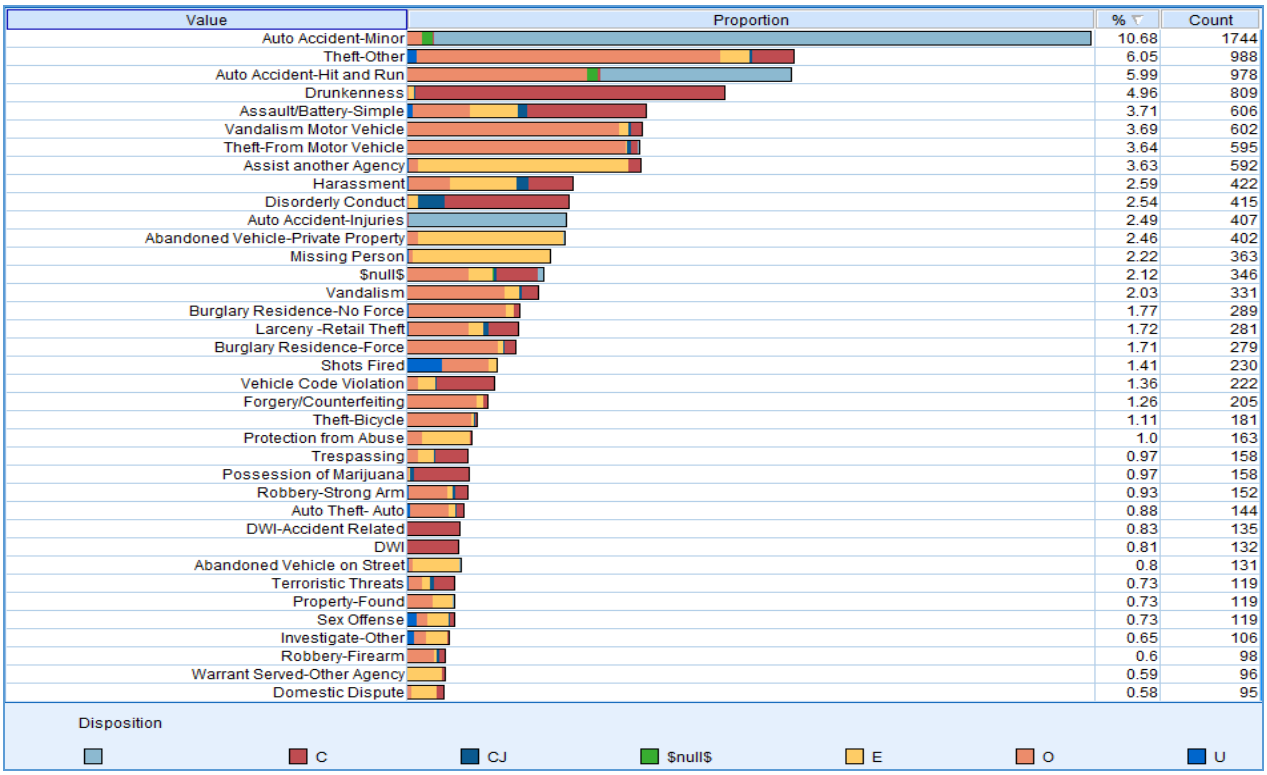

Figure 5. Distribution of crimes per crime type (showing the top 80%, color shows crime disposition)

Figure 6 shows the crime distribution per hour and day. The graph shows the top 80% crime types. The first column represents the hour of the day during which crimes occurred. 50% of the crimes occur between 10:00am and 6:00pm. In Figure 7, the distribution of the crimes per month is presented. It is noticed that higher number of crimes occur during the summer time.

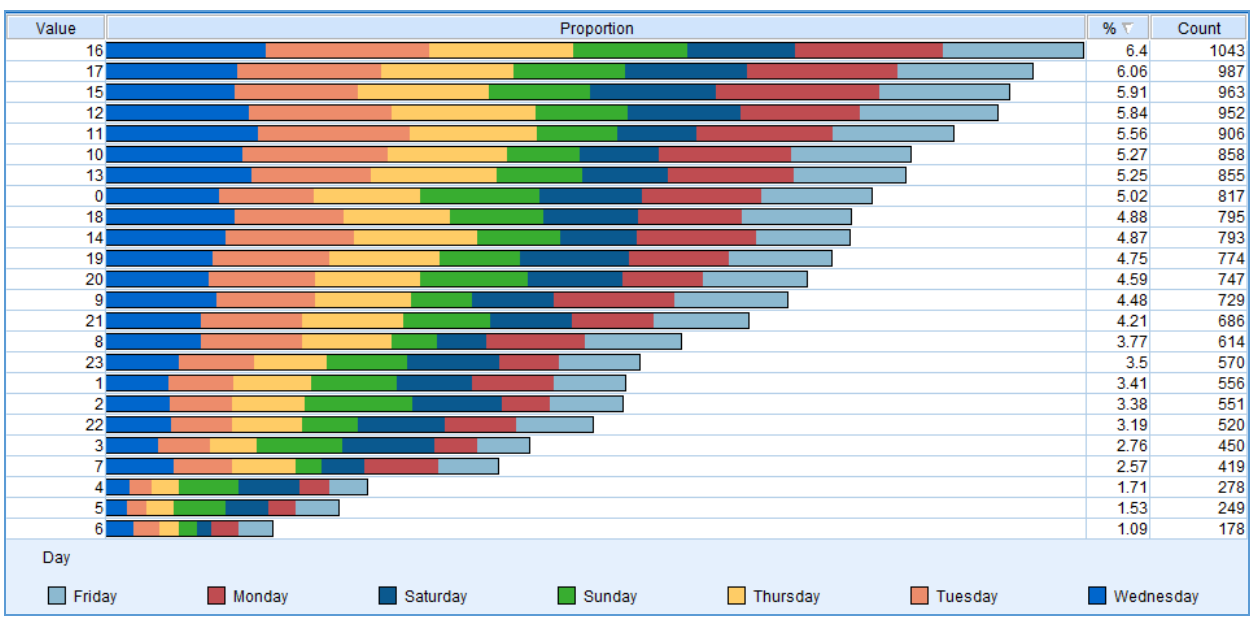

Figure 6. Crime distribution per hour of the day (colors represent the distribution per day)

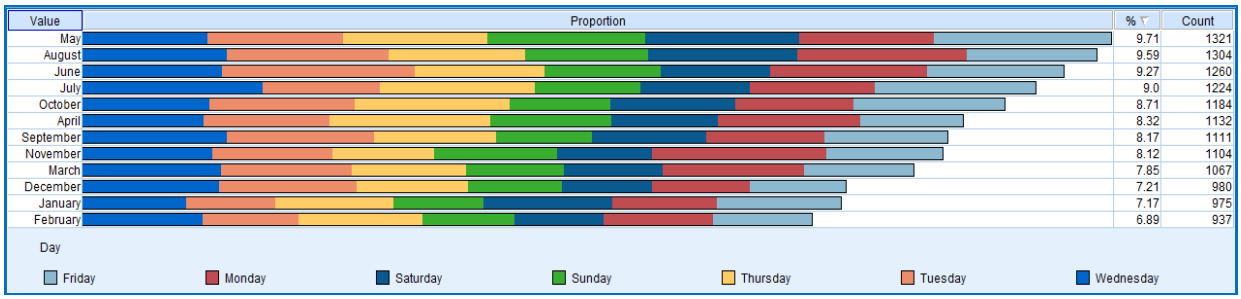

Figure 7. Crime distribution per month (2015 data, colors represent the days)

Figures 8 and 9 show the heat maps for the crime distribution for 'theft from motor vehicle' and 'theft-other', respectively. The figures show examples of how crimes can be visualized using the heat maps. Figure 8 shows the crime distribution for "theft from motor vehicle" and it is noted that most of these crimes occur in the suburbs of the city, with a major focus in the downtown area and the industrial area of the city. In Figure 9, most of the 'theft-other' crimes are concentrated in the downtown area.

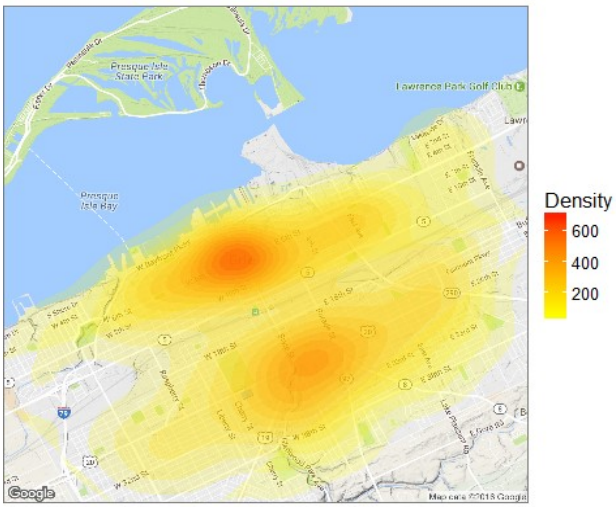

Figure 8. Crime distribution for "Theft-From-Motor Vehicle"

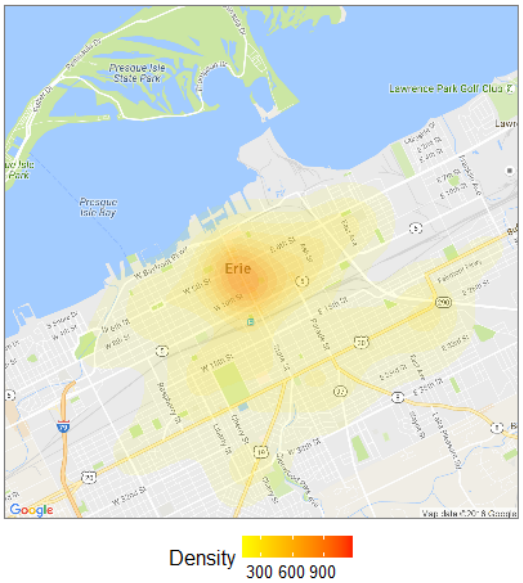

Figure 9. Crime distribution for "Theft-Other"

Figure 10 shows the drunkenness crimes; these crimes are concentrated in the lower part of State Street which has a higher density of direct liquor serving bars and restaurants compared to other parts of Erie. Missing person crimes, which are shown in Figure 11, are concentrated in the suburb area of the city, where there are schools and a higher incidence of kidnapping. More analysis of the causes of crime concentration will be done in the future work.

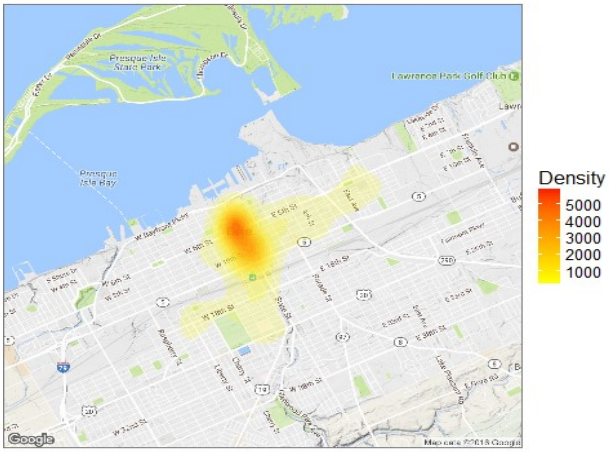

Figure 10. Heat map for drunkenness crimes

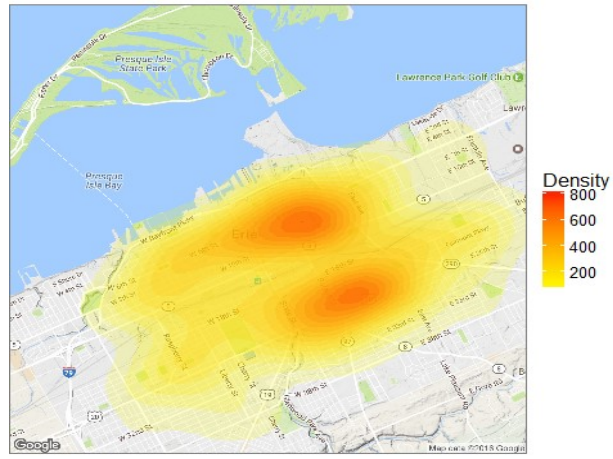

Figure 11. Heat map for "missing person"

3.1. Crime Analysis in East Side Area

East Side area is one of the areas that have high crime rates and needs further analysis to identify crime types and provide recommendations to reduce the crimes. Figure 12 shows the map of the East Side area along with the longitudes and latitudes. Figure 13 shows the distribution of the crimes per crime type and disposition. For the crime disposition, most of the abandoned vehicle related crimes in a private property have been exceptionally cleared. Most of the crimes related to minor auto accidents were not disposed as these crimes can be resolved without going to the court. On the other hand, most of crimes related to vandalism of motor vehicles and thefts are still open. Figure 14 shows the visualization of the crimes in the East Side area. Figure 15 shows the visualization of the simple assault and battery crimes. The crimes are concentrated in some areas.

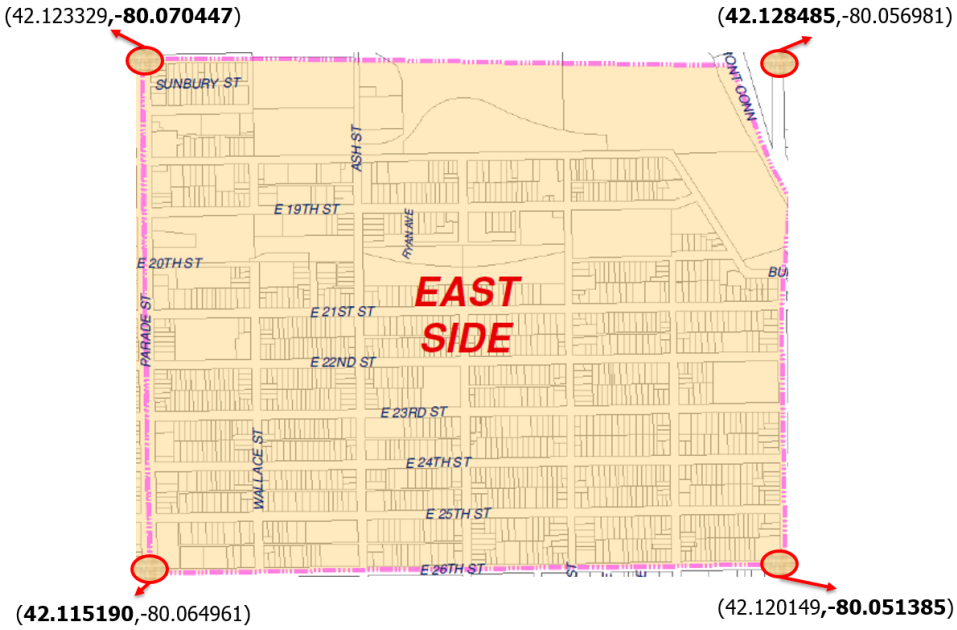

Figure 12. Longitudes and latitudes for East Side area

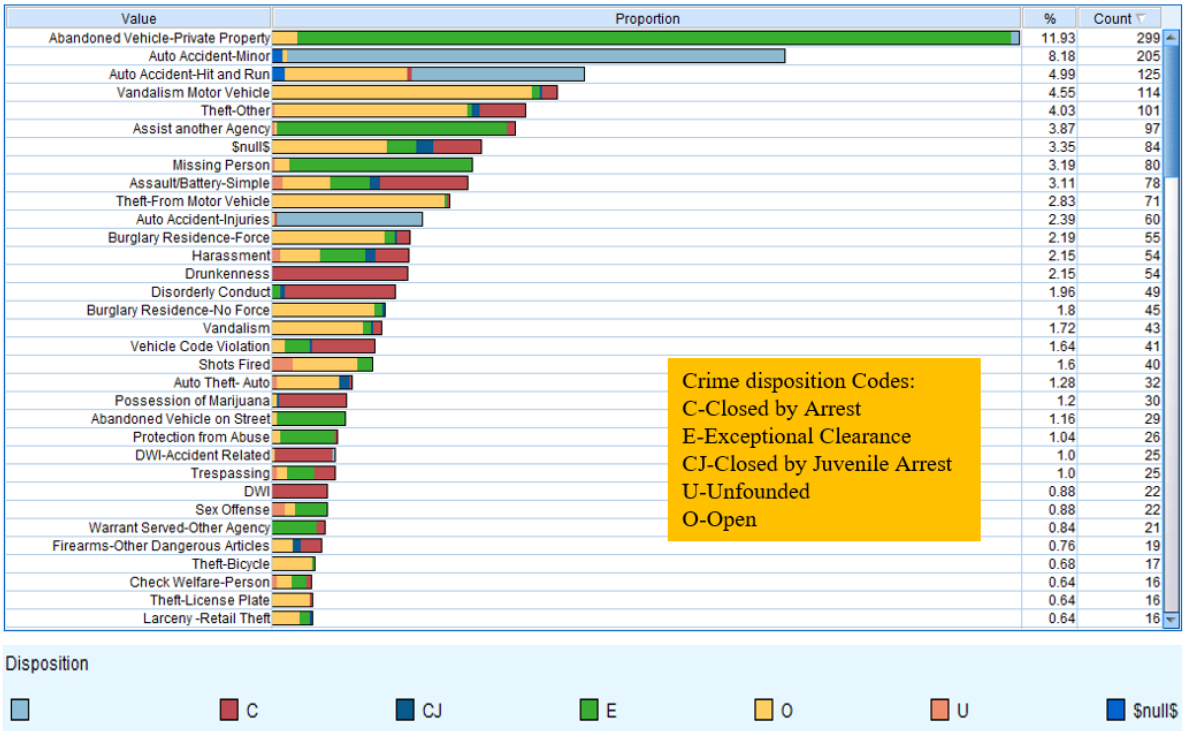

Figure 13. Analysis of crimes in East Side area

Figure 14. Heat map for all crimes in East Side area

Figure 15. Heat map for assault/battery simple crimes

3.2. East Side Eagles Area

The Eastside Eagles area is another area that has high crime rates and the Police Department has received many complains about the high crime rates in this area. Fig. 18 shows the coordinates for the Eastside Eagles area. Analysis of the crimes in the area is shown in Figure 16. The figure shows the crime distribution per crime type and disposition. Assault crimes represent the highest percentage followed by theft crimes.

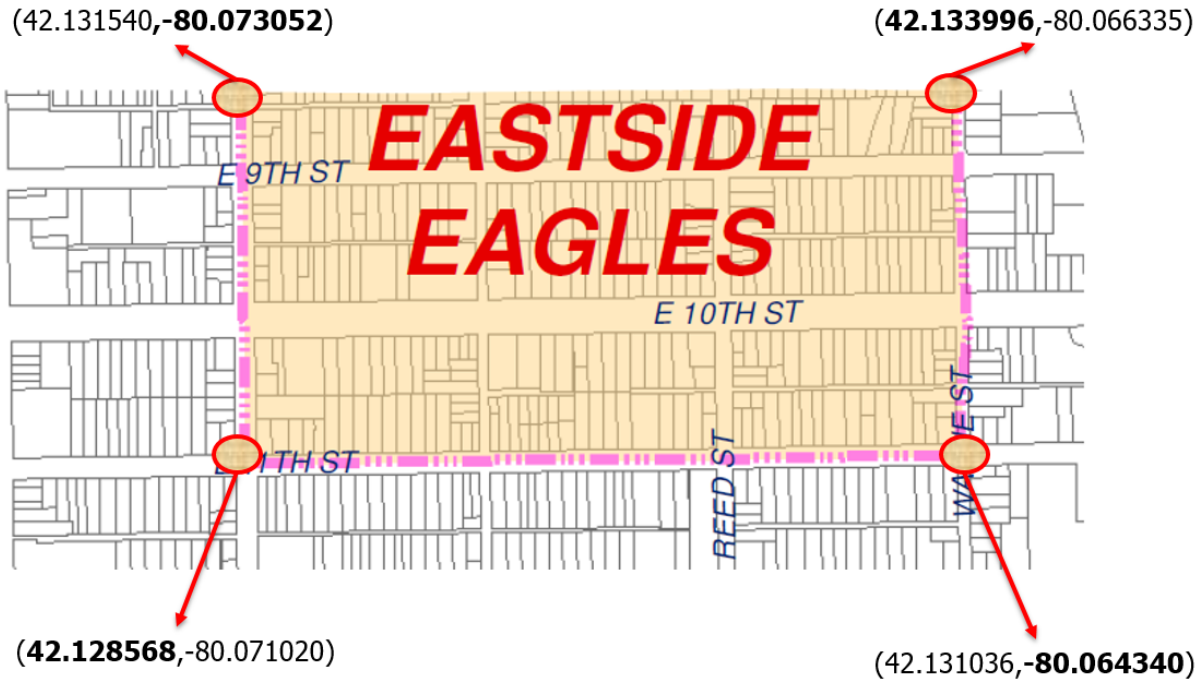

Figure 16. Longitudes and latitudes for East Side area

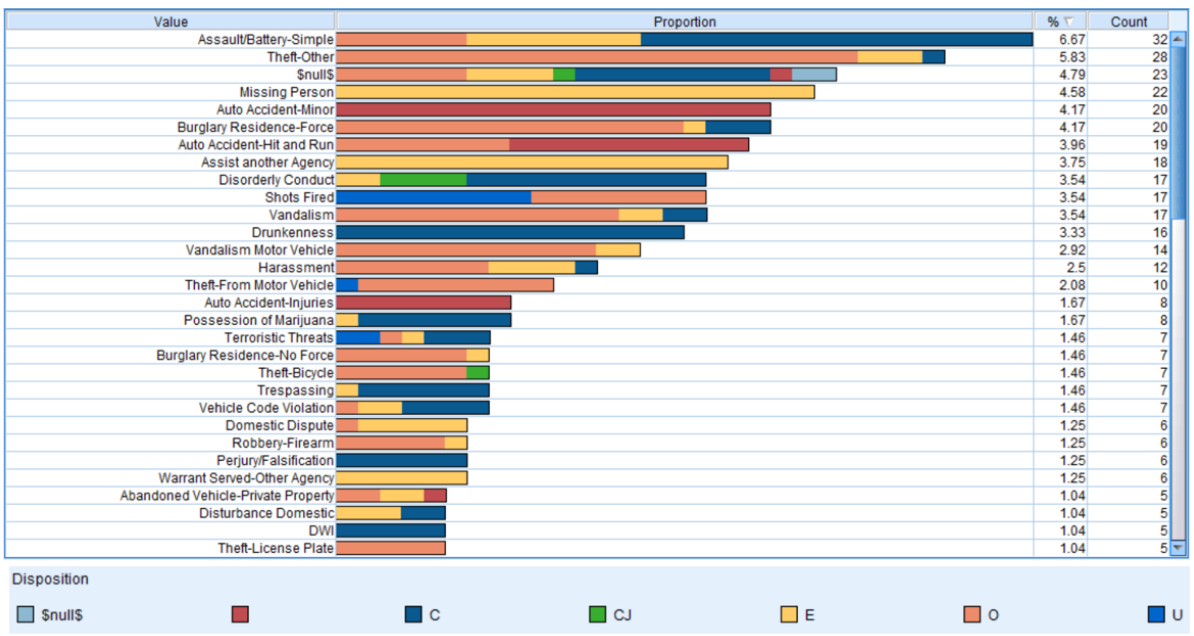

Figure 17. Analysis of crimes in East Side area

Figure 18 show the crime visualization for all crimes types. Figure 19 shows the visualization of the assault crimes. These crimes are concentrated in the middle area. It was recommended that more cameras are installed in these areas to reduce the crime rate,

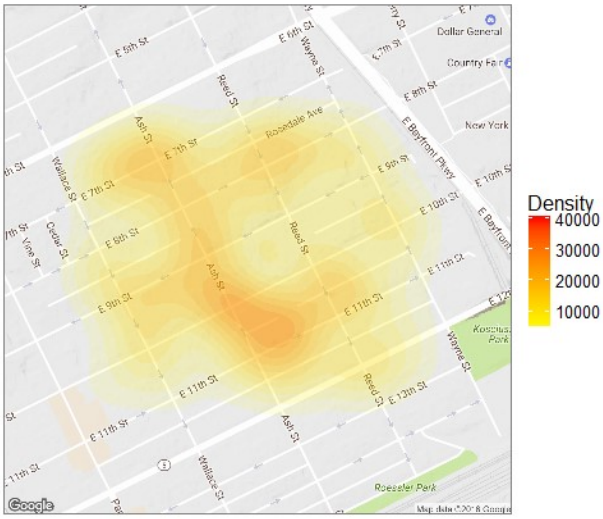

Figure 18. Heat map for all crimes in East Eagles area

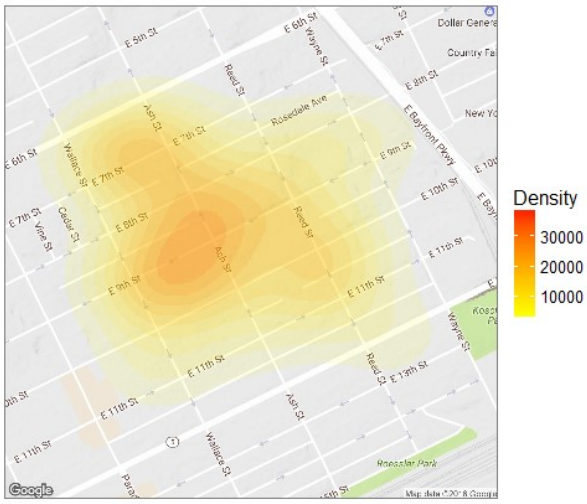

Figure 19. Heat map for assault/battery simple crimes

4. Conclusions

In this paper, we discussed a case study on crime analysis and visualization in Erie City, Pennsylvania, USA. It was shown that the months of May, August, June, July, and October have a higher incidence of crimes. Typically, in these months the weather is warm which might explain why there is an increased crime rate. Heat maps also show that auto accidents are spread-out around the city of Erie and these accidents are focused in the downtown area which has many restaurants. Drunkenness crimes, however, are focused in the street that has the bars. Cameras are installed in the streets that have high crime rates.

Future work will focus on collecting more data and developing predicative analytics models for the crimes. Moreover, text analytics will be developed using data collected from different social media websites such as Twitter and Facebook.

References

Gerber, M.S., Predicting crime using Twitter and kernel density estimation, *Decision Support Systems*, vol. 61, pp. 115-125, 2015.

Cheng, Z., and Smyth, R., Crime victimization, neighborhood safety and happiness in China, *Economic Modelling*, vol. 51, pp. 424-435, 2015.

Mawby, R.I., Exploring the relationship between crime and place in the countryside, *Journal of Rural Studies*, vol. 39, pp. 262-270, 2015.

Adderley, R., Townsley, M., and Bond, J., Use of data mining techniques to model crime scene investigator performance, *Knowledge-Based Systems*, vol. 20, pp. 170-176, 2007.

Tseng, Y.H., Ho, Z.P., Yang, K.S., and Chen, C.C., Mining term networks from text collections for crime investigation, *Expert Systems with Applications*, vol. 39, pp. 10082-10090, 2012.

Phillips, P., and Lee, I., Mining co-distribution patterns for large crime datasets, *Expert Systems with Applications*, vol. 39, pp. 11556-11563, 2012.

Kiani, R., Mahdavi, S., and Keshavarzi, A., Analysis and prediction of crimes by clustering and classification, *International Journal of Advanced Research in Artificial Intelligence*, vol. 4, no. 8, pp. 11-17, 2015.

Keyvanpour, M., Javideh, M., and Ebrahimi, M.R., Detecting and investigating crime by means of data mining: a general crime matching framework, *Procedia Computer Science*, vol. 3, pp. 872-880, 2011.

Duda, S., Erie, Pa and Presque Isle State Park Facts and Legends, Xlibris US, 2015.

Economic Research Institute of Erie (ERIE), The ERIE guide to the Erie economy, 4th Edition, Penn State Behrend, Erie, PA, 2013.

Wertz, J., and Wood, P., Erie's Advanced Industries,Jefferson Educational Society, Erie, PA, 2013.

Jefferson Educational Society, Is Erie a safety city? Perception, realty, recommendation, Erie, PA, 2015.

Biographies

Mohamed Badawy, Jared Cirelli, and Howie Setyono are undergraduate students in the Industrial Engineering department at Penn State Erie, The Behrend College.

Faisal Aqlan is currently an assistant professor of Industrial Engineering and Master of Manufacturing Management (MMM) at Penn State Behrend. He earned his Ph.D. in Industrial and Systems Engineering from the State University of New York at Binghamton in 2013. Aqlan has worked on industry projects with Innovation Associates Company and IBM Corporation. His work has resulted in both business value and intellectual property. He is a certified Lean Silver and Six Sigma Black Belt. He is a senior member of the Institute of Industrial and Systems Engineers (IISE) and currently serves as the president of IISE Logistics and Supply Chain Division, director of Young Professionals Group, and founding director of Modeling and Simulation Division. Aqlan is also a member of American Society for Quality (ASQ), Society of Manufacturing Engineers (SME), and Industrial Engineering and Operations Management (IEOM) Society. He has received numerous awards including the IBM Vice President award for innovation excellence, Penn State Behrend's School of Engineering Distinguished Award for Excellence in Research, and the Penn State Behrend's Council of Fellows Faculty Research Award. Aqlan is the Principal Investigator and Director of the NSF RET Site in Manufacturing Simulation and Automation at Penn State Behrend.