

Crime Analytics

Arjun Hande Hariharapura
Narasimha
NYU
Brooklyn
ahn264@nyu.edu

Vamshidhar Reddy Reddygari
NYU
Brooklyn
vr882@nyu.edu

Shashank Pavan Segu
NYU
Brooklyn
sps452@nyu.edu

Abstract — This paper examines the study on crime incidents grouping them together based on localities in a city along with other factors like legality of narcotics, increased availability of guns. Data from the cities of Chicago, Detroit and New York are used to examine the hypotheses of the inter relation of factors. The result of the project will help government to determine the kind of laws to be implemented based on the crime intensity value. It also helps citizens to get a sense of how safe a location is in a city.

I. INTRODUCTION

The Crimes occurring in the cities of United States are analyzed based on locality and crime types along with other factors like legality of narcotics, unemployment, increased availability of guns etc., With the help of this analytics, the government in those cities can frame the laws accordingly to reduce the crime rate. The police can use these analytics to provide higher security where crime rates are higher and reduce the number of crimes happening in that city. This analysis helps us to get to know whether crimes are increasing or decreasing gradually year by year with changes in law, unemployment rate etc., Lastly we will compare the data of US cities with some other international cities like Amsterdam and Singapore where the regulations are completely different.

The motivation behind the study is to understand the increase in crime activities related to gun and violence in certain regions of a city. To do this there is a need to analyze the data scientifically.

II. RELATED WORK

A. Survey of Crime Analysis and Prediction

The paper discusses various factors that contributes to crime in a locality and also suggests focusing on those areas where crime rate is relatively high. The factors that contribute to crime are increased incarceration, more police, usage of street lights, and the decline of crack. Paper tells that there is a correlation between the characteristics of population and the rate of violent crimes.

Analytics for crime data can be performed by defining **hotspots**. Hotspots are defined as the areas that have a higher concentration of crime events. Hotspots are identified by calculating crime rate over fixed area. It is better to perform crime analytics for smaller hotspots than an entire city. It is because, when entire city is considered then we may lose lots of information. Therefore, the size of hotspots must be small in other words the size of the area where analytics are done must

be small. Results from the analytics can be used to build a prediction model and also by performing analytics, crime rate can be brought down in a locality by taking appropriate measures.

B. Narcotics and Criminality

The paper discusses the criminality with narcotics problem in detail. The author gives citation to The Uniform Crime Reports and the data collected in the course of the Chicago Narcotics Survey, 1951-53, both indicate a trend, in evidence since the early thirties and interrupted only during the war years, towards not only an increasing number of, but also a correspondingly increased involvement of young persons in narcotic drug law violations. The problems associated with data 1952 onwards because of the change in basis of reporting arrests is also discussed.

The author then looks into how people at an adolescent age get into this drug addiction. The disadvantaged areas problems and terms like street-corner society have been dealt with in detail. The paper concludes with the limitations of both the correctional and the medical approaches, particularly as applicable to the young addict, should be recognized. He concentrates more on the the rehabilitation of a young addict.

C. More Guns, More Crime

The paper starts with a question of how more gun ownership can increase or even decrease the crime rate with an instance for each. Then he goes on to compare the previous works that attempted to answer this. The previous works by different researchers used a broader set of data which was at best at a national, state or city level. This was because reliable data on gun ownership were available at only the national level, researchers constructed proxies, such as the fraction of crimes committed with a gun, to estimate the level of gun ownership in an area. The author also argues any significant statistical relationship between guns and crime could have been driven by reverse causation or omitted variables.

For all these reasons the author proposes a new way to measure change in gun ownership within an area. The method uses state- and county-level sales data for one of the nation's largest gun magazines, Guns & Ammo, which provides a much more accurate way to measure both the level and the change in gun ownership within an area. He uses five methods to test the validity of this assumption. His findings reveal that changes in homicide and gun ownership are significantly positively related.

He comes to a conclusion after a series of studies that the decline in gun ownership can explain approximately one-third of the differential decline in gun homicides from 1993 to 1998, with the largest declines occurring in areas with the largest reductions in ownership of firearms.

The author concludes that there is a substantial increase in the overall homicide rate with increase in gun ownership. He also asserts that these are not driven by reverse causation or by omitted variables. Another important thing to note is the relationship between changes in gun ownership and changes in all other crime categories is weaker and typically insignificant, suggesting that guns influence crime primarily by increasing the homicide rate. The author also provides several statistical numbers to prove his argument and suggests to future researchers on similar topics and lawmakers that recent reductions in firearm ownership have caused the reduction in the nation's suicide rate.

D. A Predictive Model for Mapping Crime using Big Data Analytics

The paper describes the forecasting framework developed on the Hadoop platform which will be able to predict the near likely crimes and assist the police to properly plan and utilize the resources. Here the hotspot clusters are defined for different localities and are used as a basic thing for predictive algorithm. This hotspot visualization helps in prediction of crime. The framework uses the capability of Hadoop to process large amount of data in very less time. This model has 3 stages in the approach of predicting crime. The first one is geographic distribution of crime data analysis that identifies clusters with higher risk of crime. Here, the data is collected from past records and it is mapped. Kernel density estimation is used to make clusters from mapped data. In second stage, a clustering algorithm is used to decide the quality of each identified cluster. Gamma test helps in estimating how much potential each cluster has to facilitate prediction. Finally, in last stage Artificial Neural Network model predicts the crime based on classification and regression tree predictive specification.

E. Big Data Pre-Processing: A Quality Framework

Pre-processing is an important stage in analytics where unwanted data is removed and it involves Data Cleansing, Data Integration, Data Enhancement and Enrichment, Data Transformation, Data reduction and Data discretization. Data Cleansing is a process of identifying, searching and removing errors in datasets. Main aim of Data Cleansing is to identify inaccurate, incomplete, or unreasonable data and then updates, repairs or deletes these data to improve quality. There are several rules in Data Cleansing: Duplicate instance rule, Illegal value rule and Functional Dependency rules. Every analytics project must undergo Data Pre-processing stage to improve data quality of datasets. With improved data quality better and meaningful results can be achieved.

F. City Crime Profiling Using Cluster Analysis

The paper describes an interdisciplinary approach between computer science and criminal justice to develop a data mining paradigm that aims to help solve crimes quicker. Specifically, the clustering based models are used to help in identification of crime profiles in the cities. In the first stage, Clustering is done

based on type of crime. Here, we get to know the most happening crime types in the city. In second stage, crime profiling is done based on area category i.e. slums, residential, commercial, VIP zones, travel points and markets. From this, we get to know the localities with most crimes in the city. In final stage, profiling is done based on particular type of crime and particular type of area. This gives us the distribution of crime types in each location. Conclusively, this paper provides information about crime that is how much percentage of which crime type occurs in particular city and which area types are more sensitive to crime and percentage of each particular type of crime occurs in which area.

III. DESIGN

The number of crime incidences occurring in a particular locality (a rectangle based on latitude and longitude from dataset) over a period of time (timestamps from dataset) is what we are interested in. This can be achieved by processing the data based on time period first and then classifying it based on the locality information in the processed data. Multiple map phases will be able to take care of the processing that is required. The reduce phase will sum the number of occurrences in this multi-processed data.

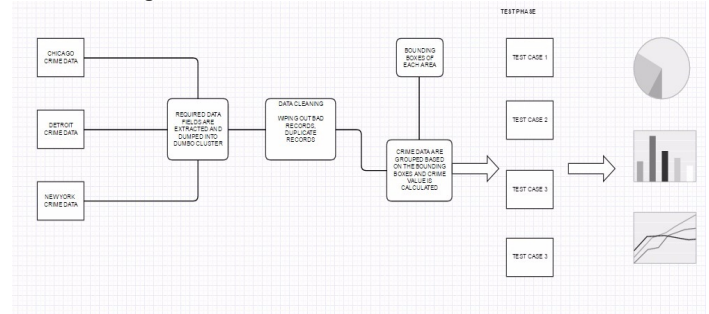


fig 1. Design diagram

The data that is used in the analytics above is cleaned first by looking for duplicate data records and data outliers which cannot be valid value for a particular field. Duplicate records are removed by filtering based on timestamp and crime_id of the records. Data outliers are taken care based on the criticality of the field.

Analytics of the data can be performed after preprocessing and data cleaning. Weights for different types of crimes are assigned with in the range from 0.0 to 1.0 based on crime severity. These weights are used to calculate the crime value of a location(precinct, cities in our case). For each precinct, the crime value is calculated by adding weights corresponding to the crime for a particular year. Similarly, crime value is also calculated for three cities(Detroit, New York City and Chicago) for a particular year.

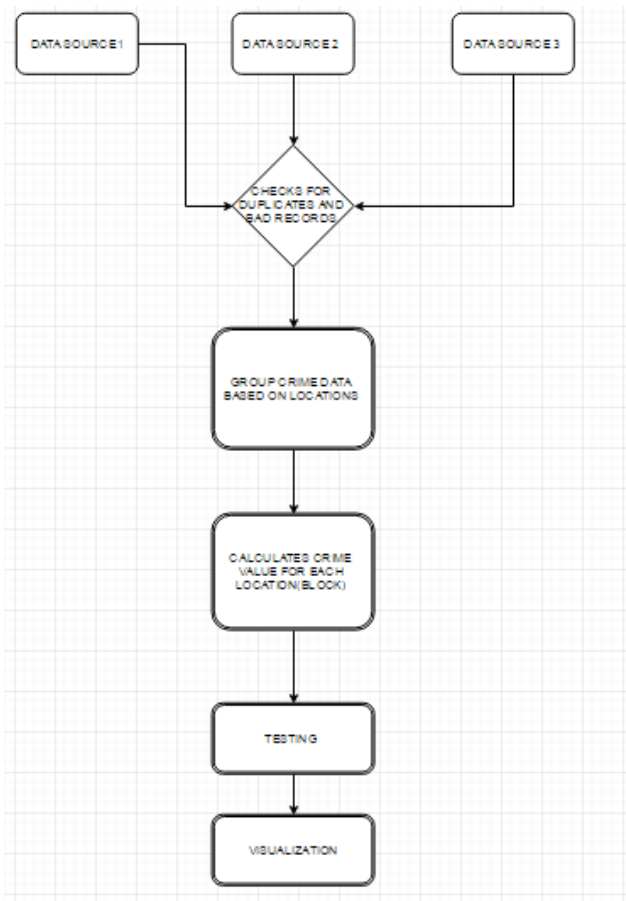


fig 2. Flow diagram

IV. RESULTS

The crime intensities were calculated by joining the crime dataset and precinct information dataset in Hive. Intensity values for individual precincts were calculated by grouping the crime indices by precinct. This data is visualized on a geographic map for better presentation to the target audience.

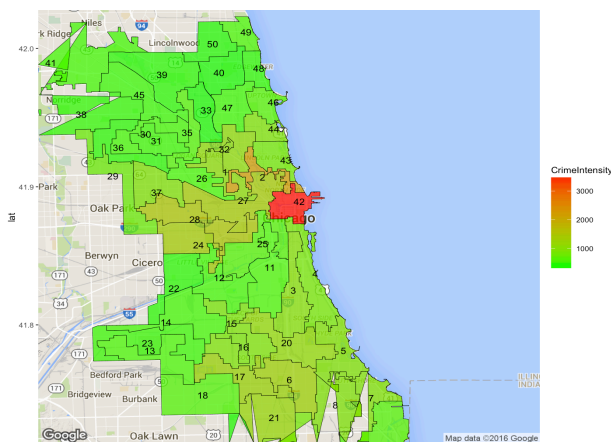


fig 3a. Chicago crime heat-map for year 2015

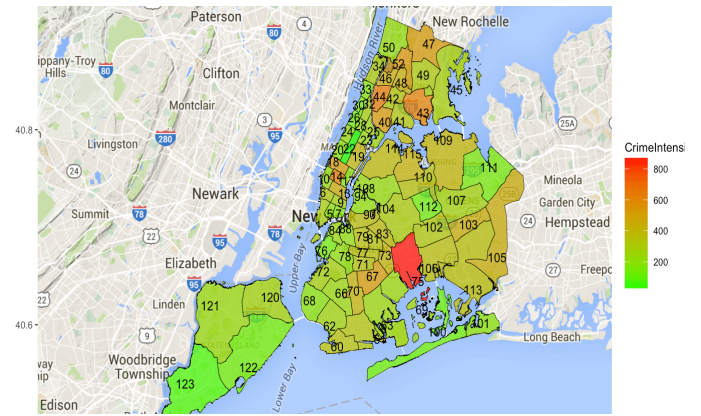


fig 3b. New York City crime heat-map for year 2015

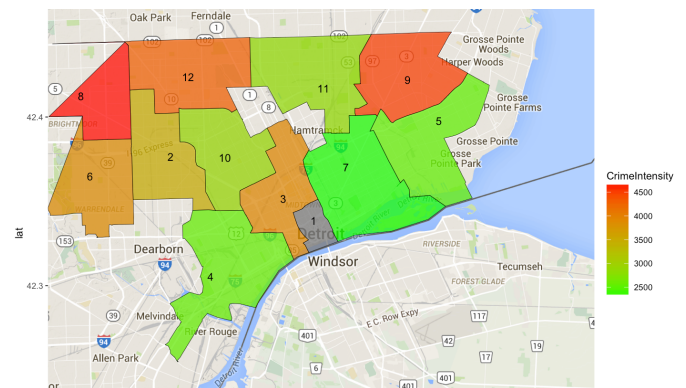


fig 3c. Detroit crime heat-map for year 2015

The probability of a person being a victim of a crime incident is also calculated by taking into consideration the population and number of incidences along with it's weight.

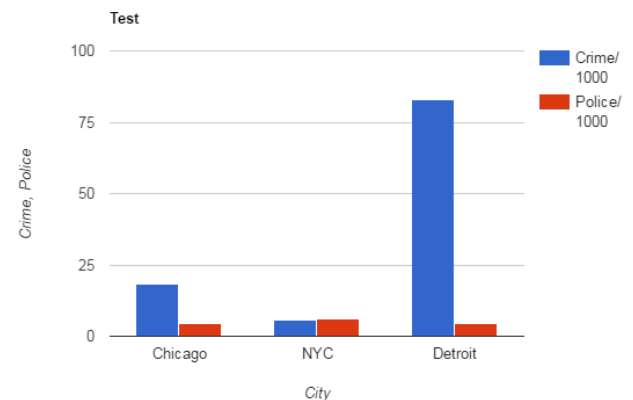


fig 4. Crime and police density per 1000 persons

The situation of cities in terms of crime rate over past several years is calculated by filtering the data by year. The data of past four years is taken into consideration for this result.

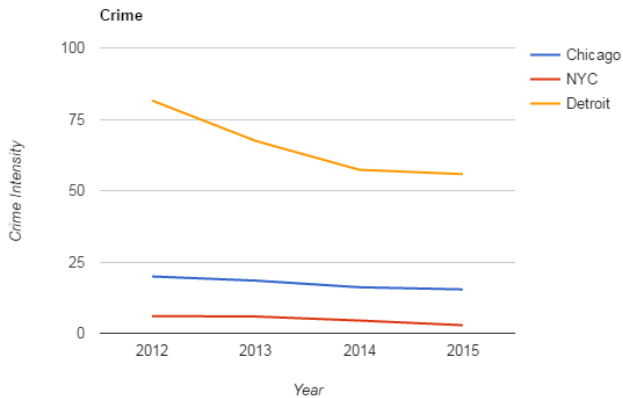


fig 5. Year over year change in Crime Intensity

V. CONCLUSION

The contributions of the project are summarized as follows: First, based on the calculated crime values the police can understand the neighborhoods and deploy more resources as and when required to ensure the safety of citizens.

From the above heat-map visualizations it is evident that ward 42 of Chicago downtown needs more attention of the police as the crime intensity is the highest in that ward. Similarly, in New York the crime intensity of precinct 75 located in the East New York section of Brooklyn needs more attention. The same argument applies for the precinct 8 which is the North-West Detroit.

Second, a common individual can use the probability of being a victim data to decide on which city or neighborhood he wants to reside. Government bodies can use these values to frame stricter laws which will result in bringing the crime rate down of the entire city.

The bar graph in the fig 4 of the Results section clearly indicates an individual can feel more safe in New York City than the other two cities as the number of police per thousand persons is highest and crime intensity per thousand persons is the least

The line graph in the fig 5 shows a trend of decreasing crime rate in all the cities from 2012 to 2015. From this result it can be concluded that law and order situation in all three cities is improving over time in recent years. The steep decrease in the rate of Detroit crime rate gives more hope to the people of that city.

FUTURE WORK

In future research, we will adapt our analytics for a more flexible geographic location within a city like individual blocks and neighborhoods. This level detailed analytic will give more clarity to people and Police department as to which street and which tiny pocket of a city has more problems.

For more information on crime dependencies, a detailed dataset on gun ownership with precincts and exact location of

the owner can be analyzed along with crime intensities in a particular geographic region.

The data of the US cities can be compared with other international cities where the law and system will be relatively different. By this sort of comparison, if an international city has a very low rate of crime that means there is something to learn from the system present there.

ACKNOWLEDGEMENT

We would like to thank Prof. Suzanne McIntosh, NYU Tandon School of Engineering and her staff for the guidance and inputs for the project. We would also like to thank the governments of NYC, Chicago and Detroit for making the data available to general public.

REFERENCES

- [1] Lenin Mookiah, William Eberle and Ambareen Siraj, "Survey of Crime Analysis and Prediction,".
- [2] HAROLD Finestone, "NARCOTICS AND CRIMINALITY,".
- [3] Mark Duggan, "More Guns, More Crime,".
- [4] Saumya1 , Anurag Singh Baghel2, "A Predictive Model for Mapping Crime using Big Data Analytics,".
- [5] Ikbal Taleb, Rachida Dssouli and Mohamed Adel Serhani, "Big Data Pre-Processing: A Quality Framework,".
- [6] Priyanka Gera, Rajan Vohra, "City Crime Profiling Using Cluster Analysis,".