Safety Indexing Using Big Data Techniques in New York City

Mehnaz Tabassum Mahin, Muhammad Shihab Rashid,
Madhurima Chakraborty, Shamali Shinde and Urja Parekh

1 Introduction

Humanity is a predominantly urban species, with over 56% of the population living in cities. This number will go upto 68% by 2050, reflecting a speed of urbanization even faster than previously predicted according to an article by the Economist. With this rate of increase in urbanization there is an increase in certain other factors like crime rates. In 2016, it was reported that Americans' level of concern about crime and violence was at its highest point in 15 years. The rise in Americans' level of concern about crime reflects actual, albeit modest, increases in crime, as well as increasing media coverage of it. Another factor that accounts for increasing crime rates is media, specially in some cities there is a raised broad concerns about decreasing levels of public safety. But, most crime analysis in present times produces maps and statistics for crimes that have already occurred. The Los Angeles Times has an article about predictive policing which tends to analyze when and where crimes are most likely to occur.

With continually increasing crime rates everywhere, now-a-days, a primary concern of people is safety, and they don't feel comfortable to roam around any area that is crime-prone. Therefore, this is really necessary for governments and law enforcement agencies to keep the citizens of the country safe from crimes. With the increasing population, crimes and crime rate analyzing related data is a huge issue for governments to make strategic decisions so as to maintain law and order. Even though safety, security, and justice are seen as moral rights and intrinsic to development ensuring it is getting difficult day by day. The best place to look up to find room for improvement is the voluminous raw data that is generated on a regular basis from various sources. But, such voluminous data becomes tough to process using traditional approaches. Hence, Big Data Analytics (BDA) comes into the picture which helps to analyze certain trends that must be discovered, so that law and order can be maintained properly and there is a sense of safety and well-being among the citizens of the country. Spark is an excellent and robust analytics platform for Big Data which can process huge data sets at a really quick speed by providing scalability. It can manage all aspects of Big Data such as volume, velocity and variety by storing and processing the data over a cluster of nodes. We plan to address this problem through the applications of big data technologies to protect users from possible crime and provide them psychological security.

The basic idea is to assign each community with a safety index based on the active crime rate in that area. Also, the law enforcement agencies need to model spatio-temporal patterns of crime to ensure public safety. The objective of our project is to provide the safety advocates, policy makers, government institutions and the people in general with a safety index to understand and prioritize their actions to improve safety around the communities. In addition to this we plan to build an interactive tool that allows users to browse the exclusive safety index of each neighborhood. Although the accuracy of the information of the tool is limited to the validity and accuracy of available data, the goal of the project is to include all available resources to develop the most complete and reliable tool for indexing a neighborhood in terms of safety.

The next sections are organized as follows. Section 2 presents the problem formulation of our project. Section 3 introduces the related works. In Section 4, we discuss the data collection and preparation process to make ready the collected NYPD dataset for our project to measure the safety index. We give an overview of our proposed framework in details in Section 5, and present the final outcomes of the project in Section 6. Section 7 presents the possible applications of our proposed framework and associated limitations to measure the accurate safety index. Finally, in Section 8 we conclude the paper with future works.

2 Problem Formulation

The objective of the project is to analyze how safe a community is with respect to all other communities in the same city, and provide users a measure which helps them to identify a safe community of any area. We envision that users will be able to remain safe and get psychological security using the safety index of a neighborhood of an area of interest. The problem being tackled in this project can be best explained in two distinct parts:

1. Performing exploratory analysis of the data to determine the relative state of safety in a particular neighborhood:
    - We analyze the spread and impact of all forms of crime to determine the safety within different neighborhoods in a city.
    - We utilize the historical crime dataset collected by the NYPD and perform exploratory analysis on it, to observe existing patterns of all forms of crime throughout the city of New York.
    - We study the crime spread in the city based on the geographical location of each crime, the possible areas of victimization in the neighborhoods, changes in the crime rate and the type, and the variations in crime from one neighborhood to another.
2. Building a visual model to depict the state of safety across all neighborhoods in a city:
    - After observing the patterns of crime from the historical data as explained above, we have to realize their pattern with respect to geographical position of their occurrence.
    - We aim to build an interactive model that treats this problem as a classification problem, where each neighborhood is assigned a number based on how safe they are with respect to all other neighborhoods in the city.
    - This is expected to help the police plan their patrol and effectively contribute to building a smarter city.

For the first part, we will make use of various data analytics tools along with the big data platform Apache Spark for initial data processing and computation, to analyze the spread of the crime in the city. For the second part, in order to build a visual model, we plan to use PySpark. In a summary, we plan to present the experimental results using heat maps, graphs and statistics.

3 Related Works

Many researches have been performed on the past crime data to identify the geographical areas of future crime and leverage the reliability of crime hotspot prediction. This influences us to analyze the crime incidents of an area of interest to estimate the safety index of its neighborhood based on past crime incidents and alert the users of the influence of crime in the area. We envision that users will be able to identify safer communities with the help of the safety index of an area of interest.

3.1 Crime Mapping and Safety Indexing

Crime mapping [1] is a popular analytical technique to identify crime-prone areas and predict the future crime incidents. Crime mapping is employed in different forms to identify and predict future crime incidents. On the other hand, a safety index of a neighborhood can be used to detect the areas with high crime volume known as 'hotspots' [1, 2]. In the 'hotspot' model [2], the current crime incidents are collected to develop clusters, i.e., hotspots for crimes and future crime incidents may possibly occur in the same areas. Also, it is an efficient indicator of the nature and severity of crime in a particular area. In [3], the authors detected the policing hotspots of crimes as a proactive crime reduction strategy.

3.2 Data Analysis for Safety Indexing

Rapidly growing social media context and the availability of social media data for analysis influence many researchers to focus on social media data for crime analysis. In [4, 5, 6, 7, 8], the authors used the social media data (Twitter) to predict and detect future crimes. Bendler et. al [5] leveraged the correlation between Twitter communicative activity and crime rates in this regard. Mathew et. al [4] examined the affordances and limitations of big data for crime analysis. They analyzed the social media data and estimated the offline crime pattern using the textual content of social media data. The authors found that disorder-related posts [9] on Twitter are the major indications for crime events. However, usage of social media data for crime prediction has some limitations.

The social media data can capture the recent crime incidents to some extent. But the major drawback of using social media data to estimate the future crimes [4, 10] is that they have to be used together with the conventional static and trusted data. On the other hand, the existing historic data are more reliable in crime indexing of a neighborhood. Wang et. al [11] proposed to use the prior crime incidents for supervised training within the predictive model. Liu et. al [12], Xue et. al [13] conducted the spatial analysis of past criminal incidents to predict future ones. All of these works prove higher accuracy in predicting crime using the historic crime data.

The drawbacks of social media data and at the same time the trustworthiness of historic data lead us to analyze the historic data from year 2006 to 2017 of the New York City (NYC) collected by the New York City Police Department (NYPD) [14]. We aim to cover from small areas (e.g., ZIP codes) to large areas (e.g., boroughs) of the NYC for this work. Though we provide the results using the NYC data, it can be generalized for any area of a city or a country.

3.3 Features for Crime Data Analysis

Most of the social media analysis techniques include the keyword volume analysis using text pattern mining and sentiment analysis. Specially, the geo-tagged tweets and keywords [4, 5, 6, 7, 10, 15] are most popular in the context of social media data analysis, including crime data analysis. In [5], the authors used Twitter's geo-tagging functionality and the location of different crime types to detect the probable crime area. The authors of [4] considered the 'Broken windows' theory [16] that relates visible forms of disorder to crime incidents. In [6, 7], the authors aimed to predict the time and location of a crime with sentiment analysis using social media data. In [10], the authors focused on the Twitter keywords with Google trend keywords to determine the similarity and predict the crime incidents. Though text mining and keyword analysis can be useful only for social data analysis, the locations and types of crime incidents are common for all platforms.

Moreover, with respect to the static historical data, along with the crime types and locations, time of crime incidents play an important role to estimate the future crime incidents. Hotspot model, the basic crime prediction model using historic data, also utilizes these features to predict the crime. However, only the types, locations and time of crime incidents do not suffice to detect the changing patterns of crime incidents. To address this problem, the spatial density of crime incidents and geographic features are taken into account by many researchers in recent years [11, 12, 13, 17]. Similarly, in our work, we focus on the crime types, locations and time of the crime incidents, spatial influence of crime along with the impact of a crime in an area. Based on these features, we compute the safety index of any area of interest in the NYC using static data of previous crime incidents of the NYC.

3.4 Existing Techniques for Crime Data Analysis

In the field of crime data analysis, crime prediction and identifying crime hotspots have major impacts on our day-to-day life to remain safe in a neighborhood. Mapping crime from past crime incidents is a very popular technique for these purposes. Hotspot mapping [2] is used as a basic form of crime prediction technique based on past historic data. A number of different mapping techniques are used to identify the crime hotspots. Some popular and widely used mapping techniques are point mapping [12, 17], thematic mapping [18, 19], heat map, kernel density estimation (KDE) [6, 15] etc. These techniques use the location of crimes and orientation of clusters of crime incidents to identify the future ones. In [10], the authors presented descriptive maps with the relative intensity of crime occurrence.

In this work, we measure the safety index of a neighborhood of an area based on the historic data. We measure the safety index of a location based on the crime rate detected from the historic crime data, and also considering the impact of a crime type in a community. Though we develop a framework based on historic data, the accuracy and reliability of our framework largely depends on the frequent update in data collection. The availability of large collections of crime data allows our framework to remain up-to-date. As the periodical update is required, we have to make our framework compatible with such huge periodical data flow. The big data techniques allow us to make it feasible by updating with new incoming huge data.

3.5 Existing Tools and Applications

There are some existing tools for crime hotspot maps using crime data collected from law enforcement agencies. Among these existing systems, Crimespotting[1], CrimeMapping[2] and SpotCrime[3] are specially designed for analyzing criminal activities. These systems collect crime data from the local police department and visualize them on an interactive map. These systems are useful for crime analytics. However, these systems rely on the past crime incidents of about the past six months. On the other hand, we aim to a user-friendly framework that will give the users to view the safety index of any area of interest.

4 Data Collection and Preparation

In this project, we analyze the historic dataset of New York City Police Department (NYPD) [14] to measure the safety index of an area of interest in New York City (NYC). To make it user-friendly, the project is designed to divide the NYC into boroughs, neighborhoods and ZIP codes so that the users can get the safety index of a particular area which they are interested in. It will enable the users to visualize and understand how much the particular area is safe.

The NYPD dataset contains about 6.5M crime entries with 35 features, and includes the reported crime incidents of the NYC spanning from 2006 to 2017. The crime data for each incident captures the description of the suspect and the victim along with the main features like the crime type, location (latitude, longitude), borough and timestamp of the incident. Indexing and processing this huge volume of data using big data tools and techniques can help us to measure the safety index of any area in real time more efficiently than the traditional database management systems.

However, the NYPD dataset contains the borough name, whereas, we need the name of neighborhoods and the ZIP codes along with the borough names. To map the crime location to the corresponding neighborhoods and ZIP codes, we have collected two more datasets of

[1] http://stamen.com/work/crimespotting/
[2] http://www.crimemapping.com/
[3] https://www.spotcrime.com/

the NYC. The NYC Neighborhood and ZIP code dataset [20] contains the ZIP codes with corresponding neighborhoods of boroughs in the NYC, and the NYC ZIP code dataset [21] contains the latitude and longitude of the ZIP codes of the NYC. From these two datasets, we have generated the NYC Location dataset of ⟨borough, neighborhood, zip code, latitude, longitude⟩ tuples. Using the NYC Location dataset, we have mapped each crime entry of the NYPD dataset to the corresponding neighborhood and ZIP code such that the location (latitude, longitude) of the ZIP code is the closest one from the location of the crime incident and the assigned neighborhood includes the ZIP code. Thus, we have prepared a dataset to analyze the safety index that contains each crime entry with crime type, latitude, longitude, borough, neighborhood, ZIP code and timestamp of the crime incident.

Crime ID	**Crime Type**	**Time stamp**	**Borough**	**Latitude**	**Longitude**
348	Vehicle and Traffic Laws	20:10:00	Manhattan	40.81077	-73.9526

(a) The excerpt of a crime entry in the NYPD Crime Dataset

ZIP Code	**Latitude**	**Longitude**	**Neighborhood**	**Borough**
10026	40.802381	-73.952681	Central Harlem	Manhattan
10027	40.811407	-73.953060	Central Harlem	Manhattan
10029	40.791763	-73.943970	East Harlem	Manhattan
10030	40.818267	-73.942836	Central Harlem	Manhattan
10128	40.781432	-73.950013	Upper East Side	Manhattan
10280	40.708538	-74.016650	Lower Manhattan	Manhattan

(b) The entries in the NYC Location Dataset where Borough = "Manhattan"

Crime ID	**Crime Type**	**Time stamp**	**Borough**	**Latitude**	**Longitude**	**ZIP Code**	**Neighbor-hood**
348	Vehicle and Traffic Laws	20:10:00	Manhattan	40.81077	-73.9526	**10027**	**Central Harlem**

(c) The NYPD crime entry with ZIP code and neighborhood

Figure 1: Data preparation

For example, we can consider the excerpt of a crime entry in the NYPD crime dataset as shown in Figure 1a, which contains "Manhattan" as a borough. Now, we want to assign the ZIP code and neighborhood to the crime entry using the NYC Location dataset. To speed up the process and at the same time to increase the accuracy while assigning the ZIP code and neighborhood to the crime entry in the NYPD dataset, we will consider only the entries of the NYC Location dataset that contain "Manhattan" as a borough. Figure 1b shows such entries of the NYC Location dataset which we can consider for this purpose, and highlights the location of the closest ZIP code and neighborhood to the location of the crime entry of Figure 1a. Figure 1c shows the crime entry in the NYPD crime dataset with the assigned ZIP code and neighborhood.

5 Proposed Framework

The birds eye view of our proposed framework is shown in Figure 2.
In this section, the detailed description of the proposed framework is explained.

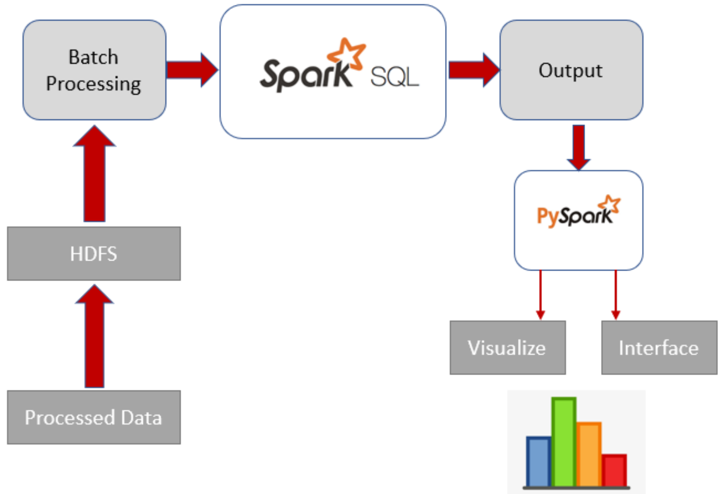


Figure 2: Overview of Proposed Framework

5.1 Storing and Loading

After the data has been cleaned of the redundant columns and the datasets have been merged, the final CSV file will be stored in a distributed file environment so that batch processing is enabled. The final CSV file will be put on Hadoop distributed file system, where we have two slave nodes and one driver machine. The driver machine will run PySpark and it will send job process to the slave nodes. It will not run any queries on the dataset. The slave nodes will run SparkSQL engine to process the queries. It will send back the results to the driver machine which will in turn visualize the results.

5.2 Spark Engine

We use Apache Spark engine to process the batch inputs of data. For this, we are using SparkSQL to run queries on our data. The data will be aggregated, mapped and grouped together. It will be reduced and counted. Safety index values will be calculated for each neighborhood of the NYC. As we are using a weighted index, simple count will not be useful. For example, 'murder' and 'robbery' do not have the same weights (i.e., seriousness) assigned to it. Murder is definitely a more serious crime. So while counting the rows for each neighborhood, a heavier weight will be assigned if the type of crime is 'murder'. After the weight calculation is done, the final output of the SparkSQL will be sent back to driver machine.

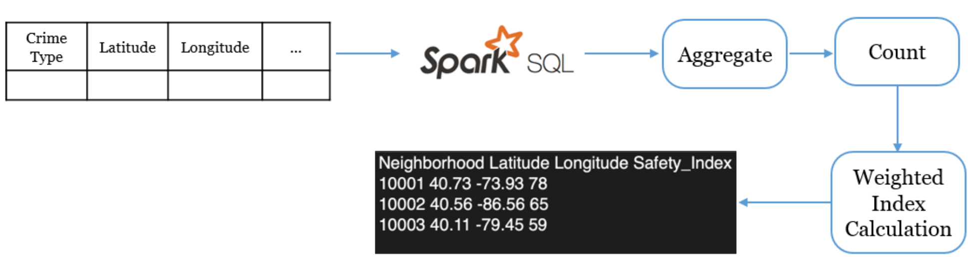


Figure 3: Architecture of SparkSQL Engine

The architecture overview of the SparkSQL engine is shown in Figure 3. We see from the output that it contains safety index value (which is out of 100) of each neighborhood. It has also the latitude and longitude values to help PySpark to visualize it better.

5.3 Safety Index

The basic idea to assign each community with a safety index is based on the active crime rate in that area. A safety index essentially indicates how safe a community is, with respect to all other communities in the same city. To measure the safety index of an area of interest, we have considered a similar equation derived by David et. al [22] which was used to deduce the quantification of the relative state of safety across 187 countries at the industrial level.

Let k be the total number of crimes of the NYC, k_i be the number of crimes of an area i ($k_i \leq k$) and W_n be the crime seriousness scale for each crime n of k crimes, according to the global crime seriousness scale [4]. Also, let X_n be the number of occurrences of a crime type n in any area across all over the city, and $X_{n,i}$ be the number of occurrences of a crime type n in a particular area i. The safety index SI_i of an area i can be calculated as follows:

$$SI_i = (1 - \frac{1}{k_i} \times \sum_{n=1}^{k} \frac{X_{n,i} - \min X_n}{\max X_n - \min X_n} \times W_n) \times 100$$

The objective of the safety index SI_i is to find the position of an area i on a scale of 0 to 100, and to determine how safe the area is with respect to all other areas in the same city. Since such safety index is subject to change from time to time as the rate of crime varies, this may need to be re-calculated from time to time, at least a year we suggest, to identify the true scenario at any point.

5.4 Analysis and Visualization

After getting the final output from the SparkSQL engine, the driver machine will produce a heat map of the NYC areas, where darker colors indicate the neighborhoods with higher safety indexes. Apart from producing a heat map of index values, different statistical analysis will be generated to show the number of crimes for each crime types. Similar analysis can be done to show the number of crimes for each neighborhoods. To generate such types of images, we use some useful libraries like "Matplotlib", "geoplot", "geopandas" of Python. The purpose of these analysis is that it will give our users an overview of crime types and their rates in any particular area. These insights will be valuable to users who want to relocate or tourists so that they can be aware of dangerous areas in the city.

6 Project Outcomes

For this project, we have used Apache Spark engine on top of the HDFS. We have run our project using the *provided cluster* of two nodes. In the *cluster mode*, the driver machine have submitted the SparkSQL jobs to the cluster, the slave nodes have calculated the safety index (as shown in Section 5.3) and generated the corresponding output which contains the safety indexes of each ZIP codes of the NYC along with the geometry of the ZIP codes.

After processing the datasets and calculating the safety indexes in the *cluster mode*, we have collected the results and done the visualization part in the local machine. The heat map generation requires some specific Python libraries, as mentioned in Section 5.4. However, the required software (python3) and the libraries are not available at the cluster nodes, and also there were limited access to the cluster nodes to install these libraries. So we have to perform the visualization in the local machine.

6.1 Visualization of Safety Index

After we get the results from Spark Engine which contains neighborhood, ZIP code, latitude, longitude and safety index value (out of 100), we have visualized it with Python's geoplot and geopandas libraries. The libraries take "geojson" file as input, where the boundaries of

[4] https://www.global-regulation.com/law/united-states/990899/crime-seriousness-scale.html

each NYC ZIP codes are needed. So we have collected the NYC ZIP code boundaries data and integrated with our final output. To do this, we have merged the final output CSV to boundaries CSV and then converted the unified CSV to geojson. This allowed the Python libraries to visualize the safety index values. In the heat map generated by geoplot, the darker the color, the safer the area. From this heat map, we get an overall idea about all types of crimes happening in each NYC area.

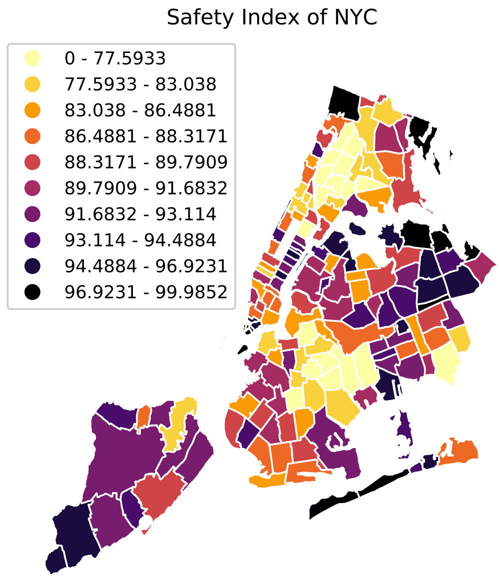


Figure 4: Heat map of the Safety Indexes of the NYC

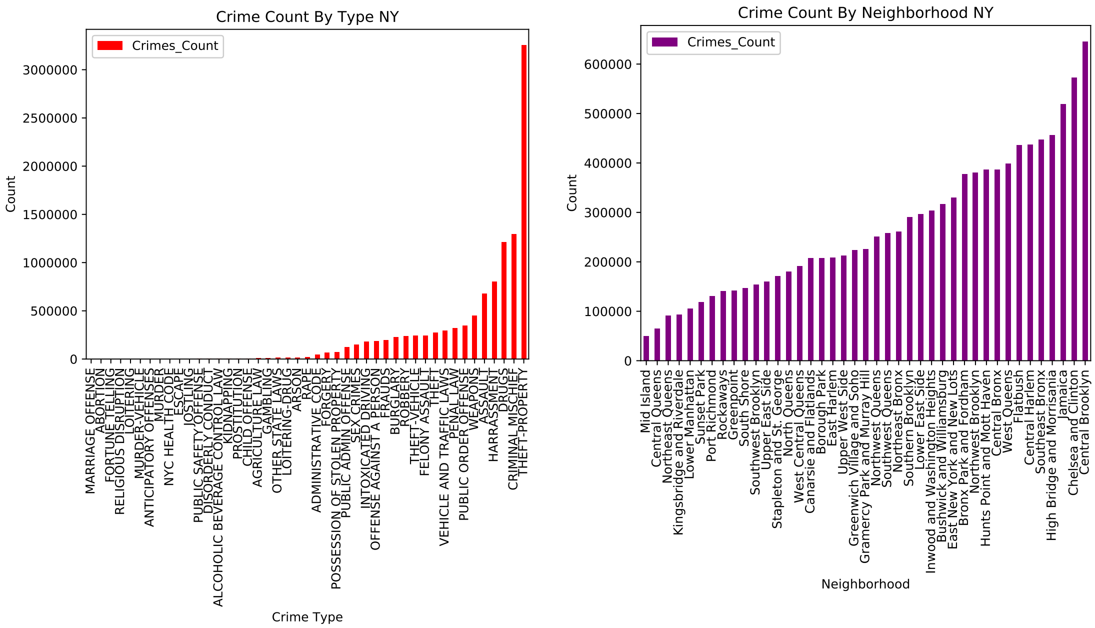
Figure 4 shows the heat map of the calculated safety indexes for each ZIP codes of the NYC. From the figure, we can see that in Central Manhattan of the NYC, it is the least safe. So police forces can enforce more personnel in that area to ensure a safer district.

6.2 Statistical Analysis

Apart from producing a heat map of safety index values, different statistical analysis are shown such as number of crimes grouped by types of crimes or whether any particular areas is showing an increasing number of crimes in recent years etc. Figure 5 shows the statistical crime analyses with the number of crimes per neighborhood and per crime types. From Figure 5a, we can see that "Property theft" has the highest count among all types of crimes. So extra laws related to property theft can be enforced to make sure this count stays low. Also, from Figure 5b, we can see that Central Brooklyn of the NYC has significant amount of crimes than the other areas. Python's "Matplotlib" library was used to show these statistical analysis.

7 Applications and Limitations

In this project, we analyze the safety index of an area of interest in the New York City, and provide the users a user-friendly framework to visualize the safety index of that area. There are many applications where the measure "Safety index" can be used as an efficient tool as described in Section 7.1. However, there are some limitations to compute the safety index of any community, as mentioned in Section 7.2.



(a) Number of Crimes per Crime Type

(b) Number of Crimes per Neighborhood

Figure 5: Statistical Crime Analysis of the NYC

7.1 Applications

- It can be used as an excellent resource for “Expert System creation” and for plotting “At Risk” crime rates and its indices.
- It may be immensely helpful for social-economic analysis while studying or creating crime reporting tools and designs.
- One can plan safe tours once being aware of the dangers around and travel accordingly.
- This can serve as a solid background work of a region, thus helping in buying, selling or renting places.

7.2 Limitations

- Since the standards of crime vary from region to region and from the FBI, not all the records from a single source like NYPD can be termed as classified standardized crimes. For instance, the data provided to the FBI contain only those crimes reported or known to law enforcement as opposed to all crime that has actually occurred.
- Additionally, transparency of crime in report is vital. But unfortunately, clear facts and figures are not exposed enough to scale and map them and therefore can at times be less of a help in visualization of crime deduction.

8 Conclusion and Future Work

In this work, we have proposed and developed a framework to compute the safety index of an area of interest in the NYC. Safety index of an area indicates how the area is safe for moving and living. The higher is the safety index, the lower the crime rate in the area is. We also provide a simple visualization which can help the users to get a better understanding at a single glance. To process a huge crime dataset and calculate the safety index of a large community such as the NYC efficiently and in real time, we have used the big data techniques provided by Apache Spark. In future, along with the crime types and the areas, the crime density and population of the areas can be considered to get more accurate results. This project can also be further extended by adding an active visualization where a user can input the fields into a UI and view a heat map or graph of any area of the NYC based on the safety indexes. Though we have considered the NYC crime records to measure the safety index of the NYC, it also can be applicable for any city or country.

References

[1] J. Eck, S. Chainey, J. Cameron, M. Leitner, and R. Wilson, "Mapping crime: Understanding hot spots," 08 2005.

[2] C. R. Block, "Stac hot spot areas: A statistical tool for law enforcement decisions 1," 1993.

[3] A. Braga, A. Papachristos, and D. Hureau, "Hot spots policing effects on crime," *Campbell Systematic Reviews*, vol. 8, no. 1, pp. 1–96, 2012.

[4] M. L. Williams, P. Burnap, and L. Sloan, "Crime Sensing With Big Data: The Affordances and Limitations of Using Open-source Communications to Estimate Crime Patterns," *The British Journal of Criminology*, vol. 57, pp. 320–340, 03 2016.

[5] J. Bendler, T. Brandt, S. Wagner, and D. Neumann, "Investigating crime-to-twitter relationships in urban environments - facilitating a virtual neighborhood watch," in *ECIS*, 2014.

[6] X. Chen, Y. Cho, and S. Y. Jang, "Crime prediction using twitter sentiment and weather," in *2015 Systems and Information Engineering Design Symposium*, pp. 63–68, April 2015.

[7] R. D. Flores, "Do anti-immigrant laws shape public sentiment? a study of arizona's sb 1070 using twitter data," *American Journal of Sociology*, vol. 123, no. 2, pp. 333–384, 2017.

[8] X. Wang, M. S. Gerber, and D. E. Brown, "Automatic crime prediction using events extracted from twitter posts," in *Social Computing, Behavioral - Cultural Modeling and Prediction* (S. J. Yang, A. M. Greenberg, and M. Endsley, eds.), pp. 231–238, 2012.

[9] R. J. Sampson and S. W. Raudenbush, "Seeing disorder: Neighborhood stigma and the social construction of "broken windows"," *Social Psychology Quarterly*, vol. 67, no. 4, pp. 319–342, 2004.

[10] C. A. Piña-García and L. Ramírez-Ramírez, "Exploring crime patterns in mexico city," *Journal of Big Data*, vol. 6, pp. 1–21, Jul 2019.

[11] X. Wang and D. E. Brown, "The spatio-temporal generalized additive model for criminal incidents," in *Proceedings of 2011 IEEE International Conference on Intelligence and Security Informatics*, pp. 42–47, July 2011.

[12] H. Liu and D. E. Brown, "Criminal incident prediction using a point-pattern-based density model," *International Journal of Forecasting*, vol. 19, no. 4, pp. 603 – 622, 2003.

[13] Y. Xue and D. E. Brown, "Spatial analysis with preference specification of latent decision makers for criminal event prediction," *Decision Support Systems*, vol. 41, no. 3, pp. 560 – 573, 2006.

[14] "NYPD Crime Data Historic Dataset." `https://data.cityofnewyork.us/Public-Safety/NYPD-Complaint-Data-Historic/qgea-i56i`.

[15] D. Yang, T. Heaney, A. Tonon, L. Wang, and P. Cudré-Mauroux, "Crimetelescope: crime hotspot prediction based on urban and social media data fusion," *World Wide Web*, vol. 21, pp. 1323–1347, Sep 2018.

[16] G. Kelling, J. Wilson, and M. Sadr Touhid-Khaneh, *Broken Windows: The Police and Neighborhood Safety (Persian Version)*, pp. 179–204. 08 2003.

[17] M. Smith and D. Brown, "Application of discrete choice analysis to attack point patterns," *Information Systems and e-Business Management*, vol. 5, no. 3, pp. 255–274, 2007.

[18] S. Chainey, L. Tompson, and S. Uhlig, "The utility of hotspot mapping for predicting spatial patterns of crime," *Security Journal*, vol. 21, pp. 4–28, Feb 2008.

[19] M. S. Gerber, "Predicting crime using twitter and kernel density estimation," *Decision Support Systems*, vol. 61, pp. 115 – 125, 2014.

[20] "NYC Neighborhood and ZIP Code Dataset." `https://www.health.ny.gov/statistics/cancer/registry/appendix/neighborhoods.htm`.

[21] "NYC ZIP Code Dataset." `https://public.opendatasoft.com/explore/dataset/us-zip-code-latitude-and-longitude/table/?refine.state=NY&q=New+York`.

[22] D. Wroth and A. Han, "205 the ul safety index: quantifying safety around the world," *Injury Prevention*, vol. 22, pp. A75.1–A75, 09 2016.