Visualizing Crime Hot Spots by Cities on a Map

Abstract—Crimes continue to be a serious social issue, and police forces are tasked with preventing and identifying them. Crimes have an impact on a country's quality of life, economic prosperity, and reputation. In recent years, there has been a significant increase in the crime rate. Around 21000 murders happened in the United States, vigilance should be increased across high crime rate locations to reduce crime. In some places crime rates will be higher and identifying those hot spots will help police to take preventive actions. In this project, we will visualize the crime hot spots using the maps and publicly available data provided by the FBI. We will first process the large datasets and visualized them based on the requirements. We will have an interactive dashboard through which we can identify the hotspots that can be used for crime prediction by police. The dashboard can serve people by showing the crime hotspots and can warn them about the crime hotspots.

Index Terms—crime analysis, hotspot mapping, crime visualization

I. INTRODUCTION

Crime visualization and analytics using big data have become an emerging approach for analyzing crime data. The growing population and unemployment have also been accompanied by an increase in violent crimes and accidents. [4]

Crime statistics in the United States have been tracked since crimes began. They have fluctuated throughout time, with a fast rise around 1900 and a broad bulging peak during the 1970s and early 1990s. After 1992, crime rates began to diminish year by year and have since dropped dramatically. This pattern continued until 2015 when crime rates began to climb modestly. This trend decreased in 2018 and 2019, but violent crime climbed dramatically again in 2020. Despite a spike in violent crime, mainly murders, between 2020 and 2021, total crime remains considerably below the peak witnessed in the United States during the late 1980s and early 1990s, as other crimes such as rape, property crime, and robbery persisted. Crimes are also considered an important factor in the decisionmaking of an individual choice of shifting to another location, or lifestyle approach [9].

The crime rates can be reduced if we can analyze the crime occurrences, locations and situations across the country and take some actions according to it. Most crimes will happen at the same locations and finding the type of crime and crime rate associated with a particular location will help in deploying our resources at the proper time which will eventually result in a decrease in crime.

To carry out the analytical and visual process we have chosen the hotspot mapping technique, this technique is widely adopted in crime-related agencies and this technique will help in taking the best decisions for controlling crime. The hotspot technique is generally a mapping technique used for identifying crime hotspots through point or spatial boundary techniques on maps [5] [8].

It is incredibly difficult to manually analyse and derive significant patterns from the amount, diversity, and frequency of data. Therefore, the following is how we plan to handle these problems:

- To store many types of criminal data, provide a more advanced environment.
- Information on various crime types should be provided, broken down by location and severity. They might concentrate on providing the important regions greater attention with the use of this knowledge.
- Describe the many types of crime that occur throughout a range of time periods, geographic places, crimes against women, crimes involving transportation, etc.
- It provides an environment to store historical data
- Different locations are depicted in maps using various colour schemes to enable comparison.

In this paper, we propose hotspot mapping and different visualization techniques on a map of crime data across the US. The crime data used in the project is downloaded from the official FBI website API, which provides access to large datasets. This data is used to measure crime, assess crime rates and make policy decisions about public safety. Mapbased crime analysis is done by using the big data architecture and AWS intelligence tools like Quicksight, EMR and EC2 for data processing and data Visualization. The investigation is carried out in three important phases, they are data collecting, Data pre-processing and data visualisation.

The major contribution of this paper can be summarized as follows.

- The change in crime rates across the US in the last decade explains the crime data across the US according to crime types.
- We developed a dashboard which helps police administrators, crime researchers and normal people in identifying crime hotspots.

The fundamental aim of this paper is to help crime agencies in analyzing crimes and identify major crime locations and crime types associated with them. And also helps an individual in knowing the crime rates of a particular location in which they want to relocate.

II. LITERATURE REVIEW

Real-time crime forecasting and mass monitoring can drastically cut crime rates. Crime analysis and prediction is a vital activity that may be improved using numerous methodologies and processes. Crimes are widespread societal issues that have an impact on a country's quality of life, economic progress, and reputation. Using Google Maps and numerous R packages, they created a framework for visualising crime networks and evaluating them using various machine

learning methods. However, the programme lacks interaction, and they did not explore a wide range of crime types. We analysed all available data and assessed several crime categories in various regions. [1]

Hotspot mapping's usefulness in identifying criminal hotspots was investigated by Spencer Chainey and his team. Analyzing the locations of crimes is essential to solving the problem of crime. Many professionals in police and crime reduction utilise hotspot mapping to pinpoint geographic patterns of crime. This means that the effectiveness of hotspot maps to forecast geographical patterns of crime and track the investigators was evaluated. In this way, it functions as a fundamental approach for anticipating potential crime hotspots based on the idea that past crime trends might serve as helpful indicators for future trends. This study forced us to specify the investigative techniques depending on the locations and types of crimes. [8]

The approach for visualising geographical and temporal patterns of crime at various degrees of spatial and temporal resolution has been provided in this article. Readers can effectively decode information because to the plots' Trellis-like vertical and horizontal alignment, especially when comparing the longterm levels of crime in each hotspot (as well as similarities between hot spots). Every time a map or graph is made, the creator encodes information that the viewer will decode. In order to effectively decode information, in this article townsley presented "hotspot plot mapping," a methodical, user-friendly method of concurrently exhibiting crime trends in space and time. A method for visualising spatial and temporal patterns of crime at various degrees of resolution, such as a period made up of years, is called "hotspot plot mapping." The study's conclusions provide a better depiction of crime statistics based on time. [5]

Police departments and other law enforcement agencies might benefit from using big data analytics to better understand criminal concerns and get insights that will help them keep track of activities, foresee the possibility of events, and improve decision-making. Big Data Analytics (BDA) has emerged as a method for processing data and obtaining knowledge, according to Mingchen Feng's research from 2019 in "Big Data Analytics and Mining for Effective Visualization and Trends Forecasting of Crime Data." In this research, large data on crime from three US cities was analysed using a variety of cutting-edge big data analytics and visualisation tools. On these datasets, time series models will work effectively to predict future criminal activity. replace [4]

To understand the linkages between criminal data, law enforcement officers must process a lot of it. Police management will be able to discuss existing issues and trends with the aid of analysts using the Spatio-Temporal Visualizer (STV) tool. At TPD, the STV tool has not yet been put into use.

To include several visualisation approaches in our continuing COPLINK project, we have created them. Chen, et al. integrated these methods into a single tool, the Spatio-Temporal Visualizer, to enable simultaneous examination of the same data from three separate perspectives. [7]

Criminal justice organisations turned to Deep Learning, Data Mining, and Machine Learning (ML) to aid in the process of fighting crime by using past crime data to discover and detect crime trends and hotspots, forecast future crime occurrences, and catch offenders and suspects. According to Mokhtar and Xia [15], two genuine crime datasets from the US cities of Chicago and San Francisco underwent state-ofthe-art large crime data processing and visualisation.

III. PROPOSED APPROACH

To visualize the crime data on a map, the following approach is used the approach involves four major steps and are explained in the following sub-sections.

A. Process Work Flow

This investigation involves large datasets, To handle data efficiently spark is used. The whole investigation is carried out by using AWS services and tools like S3, EMR, Quicksight and EC2. The lifecycle of a process flow starts with a problem and then the data required is collected using FBI API [10] and stored in AWS S3 [11]. The data is transformed in the spark cluster running on EMR [12] and after the operations are performed the data is stored in S3. We will access the master data for Quicksight [13] through SPICE.
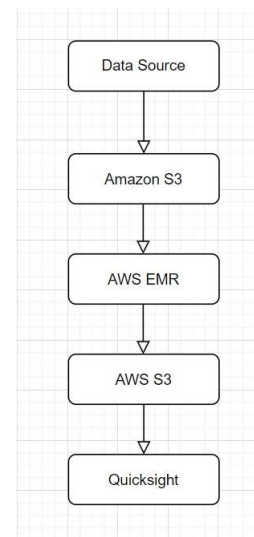

Fig. 1. Process work flow

B. Data Collection

One of three well-known criminal justice reporting systems is commonly used in the United States by professionals to measure crime, evaluate crime statistics, and decide on public

safety policies. The first is the National Incident-Based Reporting System (NIBRS), the second is the Uniform Crime Report (UCR), and the third is the National Crime Victimization Survey (NCVS). Both the UCR and NIBRS datasets are taken into consideration for this inquiry and are accessible through FBI API [10]. Since data was obtained from the FBI's official website, the dataset utilised is authentic and legitimate. The collection includes statistics from 1970 to 2020 on several states and various sorts of crimes.

C. Data Preprocessing

The datasets that were gathered contain many records, many of which are empty records, necessitating preprocessing and data cleansing. In order to handle huge volumes of data, pyspark and AWS EMR are used.

1) Pyspark setup on AWS EMR: A managed cluster platform called Amazon EMR (formerly known as Amazon Elastic MapReduce) makes it easier to run big data frameworks like Apache Hadoop and Apache Spark on AWS to process and analyse enormous volumes of data. It processes data for analytics purposes and business intelligence tasks using these frameworks and associated open-source projects. It also transform and transport huge volumes of data into and out of other AWS data storage like Amazon Simple Storage (AWS S3) Service using Amazon EMR. [12] To create an

Fig. 2. Conifguration of software requirements on EMR

EMR cluster on the EC2 instance, we need to access amazon EMR through a console and need to create a cluster suitable for our requirements. After selecting create a cluster, select Advanced options and configure the software requirements needed. The Fig.2. shows the selected software configurations for this project, we have selected one master and one core node instance of m5.xlarge and used them for the investigation and the remaining options were set to default and the cluster is created. The EMR provides us with the on-cluster application user interface through which we can access JupyterHub. [12] The Fig.3. shows the applications that can be accessed through EMR. We can run several instances of a single
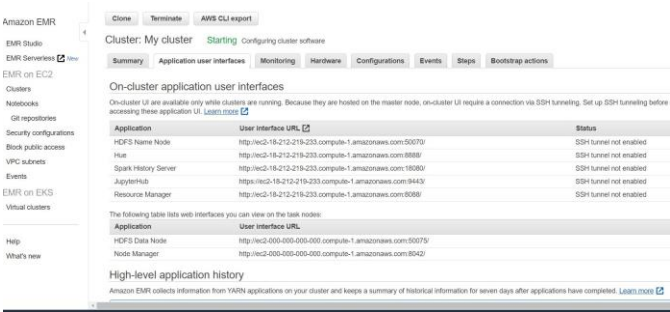

Fig. 3. Application Program Interface on EMR

Jupyter notebook server on JupyterHub. Amazon EMR creates a Docker container on the cluster's master node when you build a cluster using JupyterHub. Every component needed for Jupyter, including JupyterHub, runs inside a container. When you build a cluster using JupyterHub, Spark and Apache Livy is immediately installed. There is a default Pyspark 3 kernel for Jupyter that has to be chosen. [12]

The master dataset is built using the Jupyter notebook and kept in S3 storage services, and it comprises crime information relating to state, crime type, number of crimes, etc. The main dataset is utilised for analysis. [11] [12] [13].

- year
- state-abbr
- state-name
- population
- violent-crime
- homicide
- rape-legacy
- rape-revised
- robbery
- aggravated-assault
- property-crime
- burglary
- larcency
- motor-vehicle-theft

D. Data Visualization

Both art and science go into data visualisation. It is a type of visual language. It entails the development and research of data visualisation. Data communication through statistical visuals and plots is the main objective of data visualisation. Effective visualisation enhances our ability to evaluate and make sense of facts and evidence. The project offers the creation of crime density maps, which aid in the analysis of crime trends by criminal analysts. In order to investigate and prevent crimes, law enforcement and intelligence organisations need to understand the patterns of criminal activity. Analyzing crimes through location and maps as they happen in a place greatly aids comprehension. This study offers a revolutionary method for mapping past crime data and forecasting potential crimes in the future.

The AWS platform's interactive and visual tool aids in the visualisation of the dataset across states. The police and law enforcement investigators can study the local crime kinds with the use of crime hotspot maps. With the help of this tool, users will be able to visually filter the dataset so they can make judgments.

The AWS tool Quicksight is used for the following investigations.

1) *Visualization of Crime Hotspots across the US:* The crime data is derived from the dataset and is based on the state and total crimes that occured. The display is based on the hotspot mapping approach. The crime is depicted across the US in Fig. 4 using point maps. The amount of offences will vary depending on the size and intricacy of the colour. It also gives a summary of the information. The data can also be represented based on the year and the data frame can also be adjusted based on the requirement and identify the crime trends across states over the years. The finest representation of this visualisation is at the national level, and it will provide administrators all the information they need to improve services and reduce crime.

2) *Visualization of crime rates across states :* According to FBI standards [10], it is not appropriate to compare crime statistics from different states. Therefore, based on the state
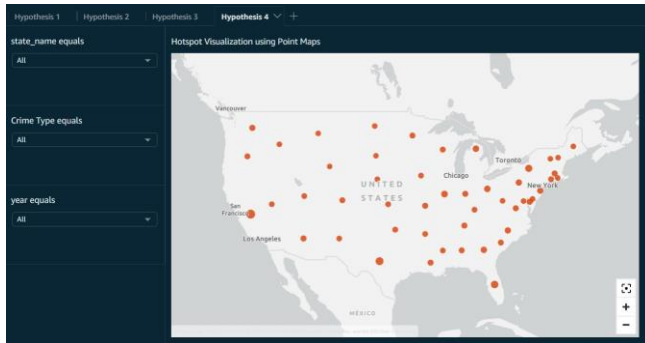

Fig. 4. Visualization of hot spots across US Quicksight Dashboard

abbreviation, state name, crime kinds, and crime count, the dataset's crime data is chosen. This visualisation offers an interactive visual dashboard as well as a page for selecting different states. In Fig. 5, a state's pie chart is displayed, with the different types of crimes indicated according to the overall number of crimes.
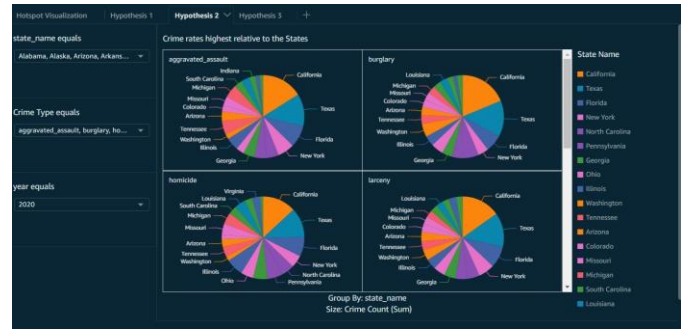

Fig. 5. Visualization of crime rates across US Quicksight Dashboard

3) *Visualization of crime types highest relative to states:* Based on the crime kinds, state name, and year trends, the dataset is used to produce the data needed for the graphics. The analysis of the sorts of crimes that commonly occur in a given location and the improvement of security measures based on those crimes are both aided by this type of depiction of crimes based on crime categories across different states.

4) *Visualizations and hotspot mapping techniques uses:* A vital and important tool for policing is the creation of maps that show hotspots. Researchers and analysts utilise these to investigate the incidence of hotspots in certain places and the reasons behind them, which helps them develop ideas. This enables academics to explain why crime happens in some locations while not occurring in others. These may be used by crime analysts to improve judgments, focus resources, develop strategies, and support law enforcement.

IV. RESULTS AND DISCUSSIONS

• The visual tool is built using big data architecture and the required investigation is done on the hypothesis using the proposed method and the following are the results of the investigation.

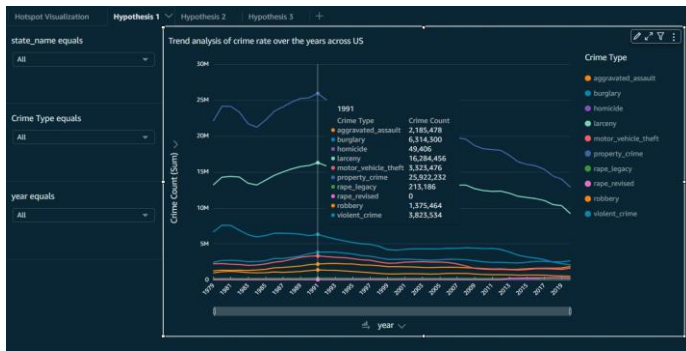Fig. 6.　　Visualization of crime types highest relative to states Quicksight Dashboard

Fig. 7. Trend analysis by crime type Quicksight Dashboard

- The live visual rich dashboard is available using link Quicksight Dashboard
- The tool is interactive in nature and users will have the flexibility and control over the data.
- Visualization of crime hot spots across the US is developed. Fig.4 explains about different crime types in point maps and the size of the bubble represents the number of crimes.
- Visualization of crime rates across the US is developed. The FBI has strictly mentioned a rule to not compare states with data provided by them. So, we have developed a pie chart Fig.5 where different states and crime types filter is present and according to the user requirement, they can filter out the data needed.
- Visualization of crime types highest relative to states across the US is developed. Fig.6 represents the visual table along with sorting filter, this filter will make user able to locate the crime types highest relative to states.
- The current visualizations will help law enforcement officers, policemen to identify the hotspots of different crimes and will support analysis.
- Trend analysis of crime visual is developed and can be used across states. The trend analysis varies from state to crime type.
- The dashboard can serve many purposes and they can be customised as per the requirement.

V. Conclusions

The solution we created is built on a big data architecture combining Spark and AWS services to carry out our studies on crime data visualisation methodologies. Through an interactive visual dashboard, this system aids public, law enforcement, and crime researchers in analysing crime throughout the US. The analysis and discovery of crime trends across states and crime categories will be made easier with the use of these interactive and visual feature applications. It is clear that law enforcement organisations may be quite effective in their efforts to prevent crime and combat it.

This study demonstrates a hotspot mapping-based visualisation approach that law enforcement organisations may use to forecast both statewide and local levels of crime. We'll work to create an automated data flow in the future for the intelligence tool. When crime reporting agencies update their data, the intelligence tool likewise updates its data and incorporates machine learning to provide some predictive skills.

References

[1] H. ToppiReddy, B. Saini, G. Mahajan. "Crime Prediction Monitoring Framework Based on Spatial Analysis". Procedia Computer Science 132 (Jan. 2018), pp. 696–705.

[2] S. Bayoumi, H. Alshabib. "A Review of Crime Analysis and Visualization. Case Study: Maryland State, USA". Apr. 2018.

[3] agenauer, M. Helbich, M. Leitner. "Visualization of Crime Trajectories with Self Organizing Maps: A Case Study on Evaluating the Impact of Hurricanes on Spatiotemporal Crime Hotspots". Jan. 2011.

[4] M. Feng et al., "Big Data Analytics and Mining for Effective Visualization and Trends Forecasting of Crime Data," in IEEE Access, vol. 7, pp. 106111-106123, 2019, doi: 10.1109/ACCESS.2019.2930410.

[5] Townsley, Michael. (2008). Visualising Space Time Patterns in Crime: The Hotspot Plot. Crime Patterns and Analysis. 1.

[6] Eck, J. E., Chainey, S., Cameron, J. G., Leitner, M., Wilson, R. E. (2000). Mapping Crime: Understanding Hot Spots. PsycEXTRA Dataset, 1. https://doi.org/10.1037/E536552006-001

[7] Chen, Hsiu-chin Atabakhsh, Homa Petersen, Tim Schroeder, Jennifer Buetow, Ty Chaboya, Luis O'Toole, Christopher Chau, Michael Cushna, Tom Casey, Dan Huang, Zan. (2003). COPLINK: Visualization for Crime Analysis..

[8] Chainey, Spencer Tompson, Lisa Uhlig, Sebastian. (2008). The Utility of Hotspot Mapping for Predicting Spatial Patterns of Crime. Security Journal. 21. 4-28. 10.1057/palgrave.sj.8350066.

[9] Wikimedia Foundation. (2022, October 5). Crime in the United States. Wikipedia. Retrieved November 12, 2022, from https://en.wikipedia.org/wiki/CrimeintheUnitedStates.

[10] Crime data Available at: https://crime-dataexplorer.fr.cloud.gov/pages/docApi (Accessed: November 15, 2022).

[11] S3 (2002) Amazon. Strand Street Press. Available at: https://docs.aws.amazon.com/s3/index.html (Accessed: November 11, 2022).
https://doi.org/10.1145/3523089.3523094