iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 **Issue:** VI **Month of publication:** June 2021

DOI: https://doi.org/10.22214/ijraset.2021.35623

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VI Jun 2021- Available at www.ijraset.com

Crime Data Analysis in San Francisco, Seattle and India using OpenStreetMap

Tamilaruvi D[1], Manjula K R[1]
[1]*School of Computing, SASTRA Deemed University, Thanjavur 613401, India*

Abstract: ***Crime is rampant in our society for a very expert. San Francisco, Seattle and Indian Police Departments continued to schedule many lawbreaker cases occurring every day and have liberated this information to the community as factor of an uncovered data system. In this project, data interpretation applied to these databases. The focus will be on conducting a detailed interpretation of the prime offences that occur in the town, identifying age trends, and determining how various factors, such as the seasons, contribute to a particular offence. In crime analysis, performing pre-data processing, data visualization on OpenStreetMap, data integration, temporal styles, and correction analysis using R and R studio software. To carry out this procedure using the San Francisco, Seattle and Indian Police Department's Data Databases.***
Keywords: OpenStreetMap, Criminal analysis, temporal trends

I. INTRODUCTION

Offence continues to be a threat to our association. Data officials who work closely with law enforcement officials throughout the United States record hundreds of offences every day. Many towns have signed up for the uncovered Data program, thus making this lawbreaker data, among other types, available to the community. The purpose of this program is to increase taxpayer participation in decision taking through this data to find interesting and useful truths. Scientists and data engineers working in cooperation with the San Francisco Police Department (SFPD), the Seattle Police Department(SPD), the Indian Police Department(IPD) has noted more than 100,000 lawbreaker cases in the police complaints form they have received. With the help of this real data, many designs can found. This will help us to forecast future offences and thus help town police to better defend the people of the town.

A. Motivation

The incentive behind taking up this thesis is that every aware citizen in today's modern world wants to live in a safe environment and neighbourhood. However, it is a known fact that offence in some form, exists in our association. Although we cannot control what goes on around us, we can definitely try to take a little stairs to aid the government and police masteries trying to control it. The SFPD, SPD and IPD has made the Police Complaints data from the year 2003 to 2018 available to the public. Hence, taking innovation from the truths stated above, we resolved to process these data provided and analyse it to identify the trends in offence over the years as well as attempt to forecast the offences in the future.

B. Objectives

1) Visualizing terrestrial trends of criminal task
2) Analysing parts that may influence illegal behaviour

II. OPENSTREETMAP

Created in 2004, OpenStreetMap (OSM) is a easily operated combination project that aims to produce openly reachable worldwide information. Today, numerous geographical qualities are continuously being added to the OSM database, including constructions and their purpose, land use and common transport details. This data allows local governments and groups to boost disaster threat evaluations and crisis drafting and is often used in a variety of disaster threat management applications. As of today, there are more than 5.5 million OSM users and one million propounds who make more than three million changes daily. In the context of natural disasters, the Humanitarian OpenStreetMap Team (HOT) is collaborating on a volunteer mapping project, at first construct after the Haiti earthquake, generating pursuits aimed at enhancing OSM data to support crisis services (https: // wiki. Openstreetmap.org/wiki/ Stats). Often, in the event of a disaster, the need of this essential information, which leads to ploting campaigns, including Mapathons, designed to map the affected areas. OSM data composed in three main ways: (1) using GPS records, which can entered into a database; (2) reliance on orthophotos and satellite imagery with high resolution to trace and digitize digital features; or (3) importing sets of data from external sources such as administrative census data.

Recently, major companies, including Apple, Microsoft, and Facebook have been hiring coordinators to donate to the OSM database. Various tools have growed to keep up OSM mapping attempts, one of which is the Map Swipe app (https://mapswipe.org/), which permits participants to plot and tag attributes of locations on mobile phones based on satellite imagery. Several programs, such as Lost Maps (https://www.missingmaps.org/) allow participants to trace satellite image-based attributes by breaking the map into smaller piece of works, permitting remote participants to work concurrently in the same location (from 2018, there were about a dozen artists. -60,000 contributors to lost maps).

III. LITERATURE REVIEW

Document reviews conducted in detail to understand the present condition in the survey and demonstration of lawbreaker data and a summary of a little selected texts produced below. Hirschfield et al conducted GIS survey of offence data targeted at Merseyside, UK Their project illustrated the pliability of PC ArcView as a structure for combining and targeting many different data sets [6]. Anderson and Joaquin proposed controlling survey chart schemes offence rates using the ARIMA (Autoregressive Integrated Moving Average) time series models and calculation rule schemes to determine whether changes have occurred in the process that caused the watch series [7]. Brown et al proposed ReCAP (Regional Crime Analysis Program), a software program that provides offence analysts the power to take out data to catch lawbreakers. This software produce similar, local, temporary, and mining data methods [8]. Messner conducted an ESDA with murder details of 78 countries in the town and in St. Petersburg. Louis also saw systematized land alter [9]. Murray et al used GIS in the survey of offence data in the town of Brisbane and illustrated the success of geographical survey in forecasting offence occurrence [10]. Ratcliffe discussed future training needs using a simple model to decrease intelligence-led offence. The model proposed that training managers to superior understand the survey presented to them, and how they could use mapping to further prevent and decrease offence, could be as important as increasing the practical skills of offence analysts [11]. Difficult features of performance-based cluster survey to determine tropical observation. This work illustrates the power of statistical methods for surveying local and terrestrial offence data for the town of Ohio [12]. R became pregnant and is still best known for her skills as a mathematical language „. Geographical data survey and observation are an important area for growth within this additional functionality [13]. Kelly et al has been shown „R" as an important tool for performing a variety of geographical survey. They also proposed that calculators and many calculation methods could use „R" in their research [14]. Miller et al suggested RTM as a threat assessment method to help forecast offence by including the underlying causes of offence and placing all of these parts in common worldwide structures in a continuous manner. In fact, the RTM produce a value (weight or non-weighted) that designates the presence, absence or strength of each object in each location in a given geography. A separate merger map (raster) of the same area represented each element. All map layers were integrated into the Geographic Information System (GIS), producing an integrated map - a map of the danger zone [15]. Dyga and Sławinska have proposed histograms and mosaic structures to show offence levels with better clarity, presentation of findings and explanation followed by maps showing offence rates in certain factors of Poland [16]. , use geo-coded data using status files, retrieve hidden information visually, create new variables from restricted data, build an offence predicting engine, and test its accuracy [17].

IV. DATASET

The San Francisco crime data created by the San Francisco Police Department on the SF Open Data website, which is part of the open data system. The Seattle crime data created by the Seattle Police Department on Seattle Open Data website. The Indian crime data created by Indian Police Department on Indian Open Data website. These databases contains the following attributes:

1) *IncidntNum:* It is a number field. It means the crime number as recorded on police logs. Same as line number.
2) *Description:* It is a text field. Contains a brief description of the crime. This field provides more information than the Category field but is still limited.
3) *DayOfWeek:* It is a text field. Specifies the day of the week when the crime took place. It takes one of the quotes from Monday, Tuesday, Wednesday, Thursday, Friday, Saturday, and Sunday
4) *Date:* It is the Time field. Specifies the exact date of the crime.
5) *PdDistrict:* It is a text field. It clarifies the police district the crime that took place in San Francisco divided into 10 police districts. It takes one of the values from Southern, Tenderloin, Mission, Central, Northern, Bayview, Richmond, Taraval, Ingleside, Park.
6) *Resolution:* It is a text field. It clarifies the criminal decision. It takes one of these values: Bound, Booked, None · Address: It is a text field. Provides a criminal street address.
7) *X:* It is a country field. It provides long criminal links. · **Y:** It is a country field. Provides latitudinal links for crime.

8) Location: It is a location field. In connection mode, i.e. (X, Y)

9) PdId: It is a number field. It identifies a unique alternative to each registered complaint. Used for data updates or search operations.

10) Category: It is a text field. Specifies the crime category. Initially, there are 39 different values (such as assault, Larceny / Theft, Prostitution, etc.) in this field.

It contains data from 2003 to 2018.

V. SOFTWARE REQUIREMENT

A. R and RStudio

R a free, open source software for statistical analysis, based on the S language. RStudio is a free, open source IDE (integrated development site) for R. (You must enter R before installing RStudio.) Its interface is set so that the user can clearly view graphs, data tables, R code, and output all at once. It also provides an Import-Wizard feature that allows users to import CSV, Excel, SAS (* .sas7bdat), SPSS (* .sav), and Stata (* .dta) files into R without writing code in do so.

VI. IMPLEMENTATION

A. Data Pre-processing

Data processing is a method of data mining that involves converting raw data into an understandable format. In this process, read the data, display the data and process the data using San Francisco, Seattle and Indian police department databases.

B. Data Visualization and Data Aggregation

Data visibility is a clear expression of knowledge and data. By using visual aids such as charts, graphs, and maps, data recognition tools provide an accessible way to identify and understand trends, exports, and patterns in data. In this process, show crime everywhere using OpenStreetMap and edit the graph between Case number and crime year.

Data aggregation is the process of collecting data and presenting it in a concise format. Data can collected from multiple data sources for combining these data sources into a summary of data analysis. In this process, to summarize the details, create a bar-based event structure and create a pie-based chart of events.

C. Temporal Trends

Temporary statistical analysis allows you to evaluate and simulate variable performance on data set over time. In this process, the analysis of theft over time, the map of the time period, the arrests over time.

D. Correlation Analysis

Correlation analysis is a mathematical method used to test the strength of the relationship between two variables. In this process, to perform factor by crime category, factor by police district, factor by month, factor by year.

VII. RESULT

In Data Pre-processing, load the data, display the data using r libraries. In Data display, all text are in capital letters. It is difficult to read. So force the text in the appropriate columns into proper case.

Show 5 entries Search:

	IncidntNum	Category	Descript	DayOfWeek	Date	Time	PdDistrict	Res
1	150060275	NON-CRIMINAL	LOST PROPERTY	Monday	01/19/2015	14:00:00	MISSION	NONI
2	150098210	ROBBERY	ROBBERY, BODILY FORCE	Sunday	02/01/2015	15:45:00	TENDERLOIN	NONI
3	150098210	ASSAULT	AGGRAVATED ASSAULT WITH BODILY FORCE	Sunday	02/01/2015	15:45:00	TENDERLOIN	NONI
4	150098210	SECONDARY CODES	DOMESTIC VIOLENCE	Sunday	02/01/2015	15:45:00	TENDERLOIN	NONI
5	150098226	VANDALISM	MALICIOUS MISCHIEF, VANDALISM OF VEHICLES	Tuesday	01/27/2015	19:00:00	NORTHERN	NONI

Showing 1 to 5 of 100 entries Previous 1 2 3 4 5 ... 20 Next

Fig 7.1 Display the data

Show 5 entries Search:

	IncidntNum	Category	Descript	DayOfWeek	Date	Time	PdDistrict	Resolution
1	150060275	Non-Criminal	Lost Property	Monday	01/19/2015	14:00:00	Mission	None
2	150098210	Robbery	Robbery, Bodily Force	Sunday	02/01/2015	15:45:00	Tenderloin	None
3	150098210	Assault	Aggravated Assault With Bodily Force	Sunday	02/01/2015	15:45:00	Tenderloin	None
4	150098210	Secondary Codes	Domestic Violence	Sunday	02/01/2015	15:45:00	Tenderloin	None
5	150098226	Vandalism	Malicious Mischief, Vandalism Of Vehicles	Tuesday	01/27/2015	19:00:00	Northern	None

Showing 1 to 5 of 100 entries Previous 1 2 3 4 5 ... 20 Next

Fig 7.2 Pre-process the data

In Data Visualization, to display crime incidents, category, locations, address, date, time on the openstreetmap using leaflet library. Marker cluster algorithm is used in this process. It is used to show the incident details on the OpenStreetMap.

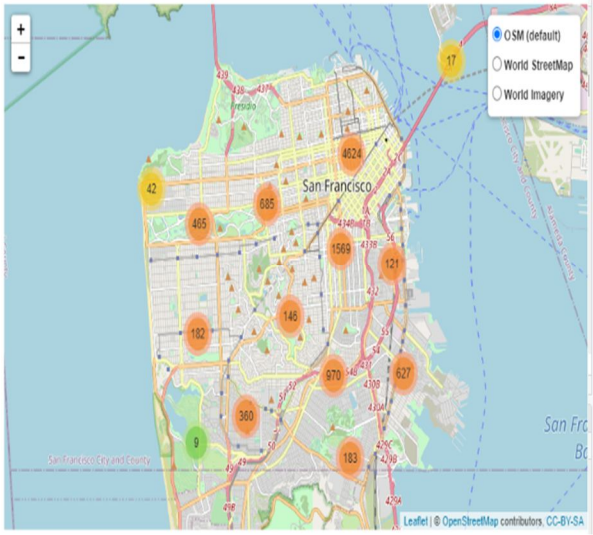


Fig 7.3 Crime across space in San Francisco

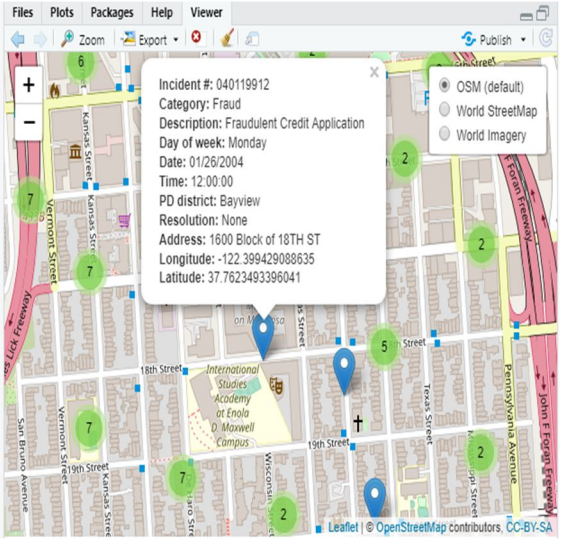


Fig 7.4 San Francisco incident details

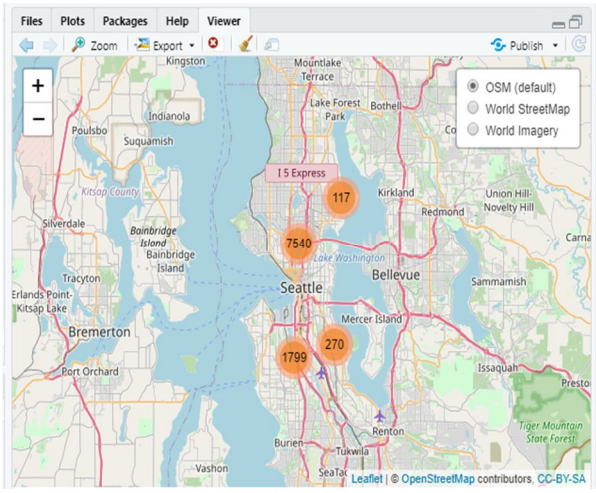


Fig 7.5 Crime across space in Seattle

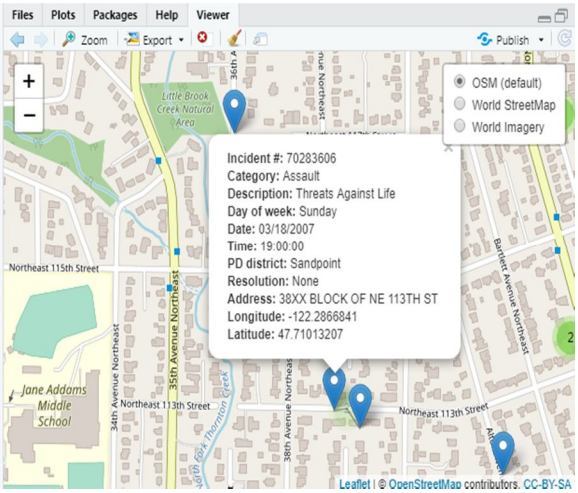


Fig 7.6 Seattle incident details

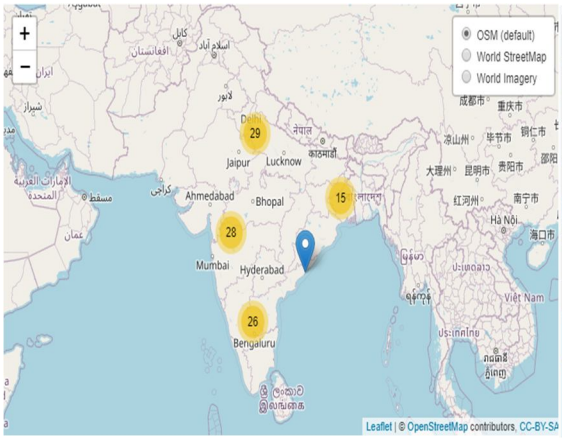


Fig 7.7 Crime across space in India

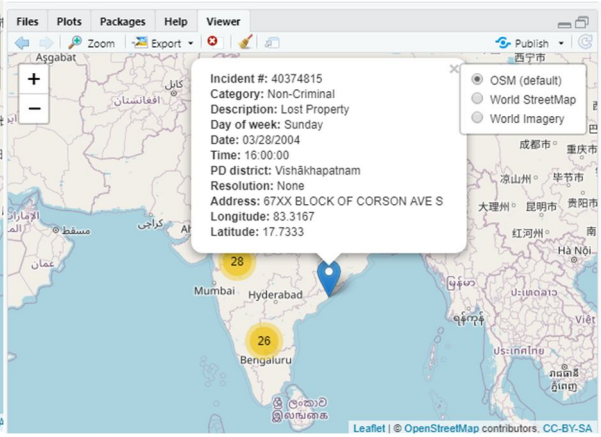


Fig 7.8 Indian incident details

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VI Jun 2021- Available at www.ijraset.com

In Data Aggregation, to recap the data by incident type, generate a bar plot by incident type and generate a pie chart by incident type.

Show 10 entries

Search:

	Category	Frequency	Percentage
1	Larceny/Theft	477975	0.227255113864095
2	Other Offenses	303027	0.144075391785962
3	Non-Criminal	236937	0.112652638555609
4	Assault	191384	0.0909942836168545
5	Vehicle Theft	126228	0.0600156043994707
6	Drug/Narcotic	117875	0.056044137343439
7	Vandalism	114718	0.0545431291432842
8	Warrants	99821	0.0474602912726144
9	Burglary	91067	0.0432981671724705
10	Suspicious Occ	79740	0.0379126999937716

Showing 1 to 10 of 20 entries

Previous 1 2 Next

Fig 7.9 Summarized data

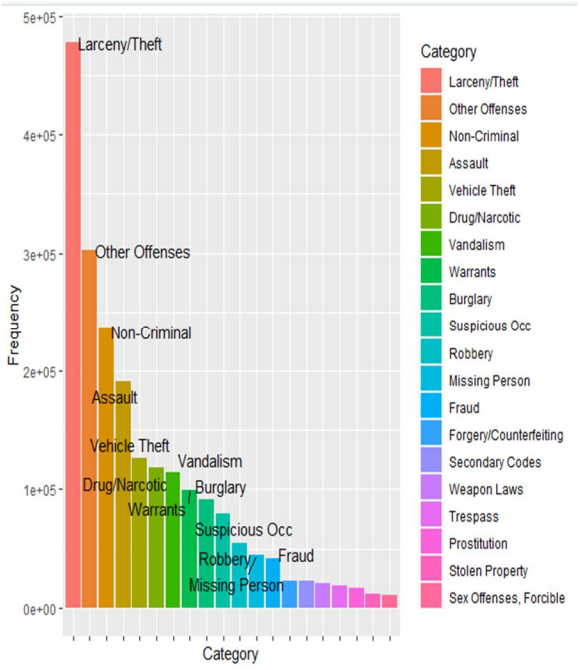


Fig 7.10 Bar plot by incident details

In summarized format,
Percentage = Frequency / Sum of Frquency

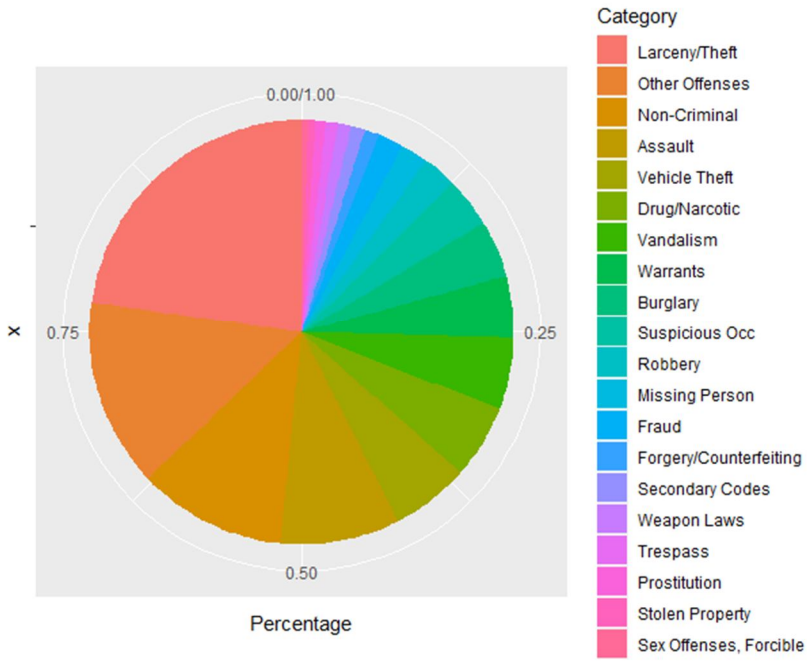


Fig 7.11 Pie chart by incident details

In Temporal trends, Cluster total of robberies by and Time to create heat map. To generate a graph of arrests over hour.

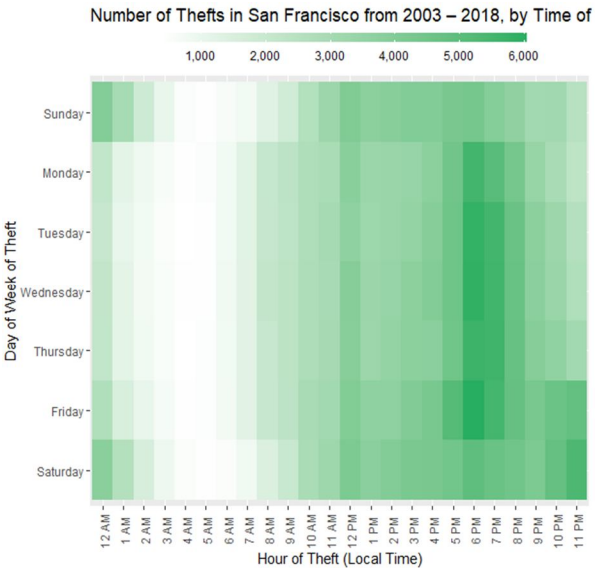


Fig 7.12 Theft Time Heatmap in San Francisco

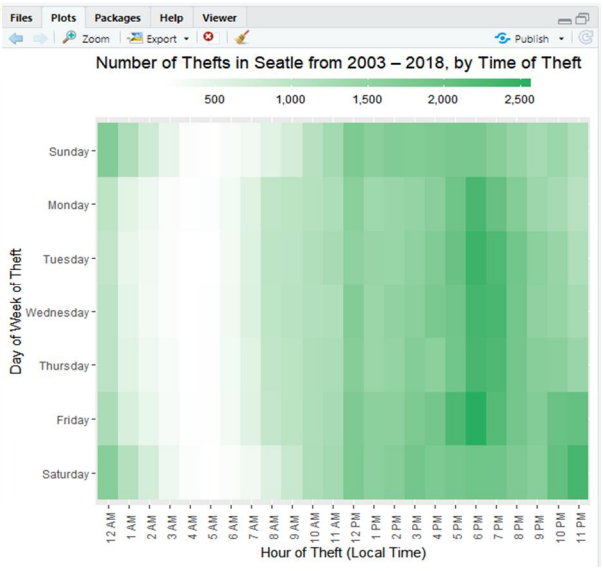


Fig 7.13 Theft Time Heatmap in Seattle

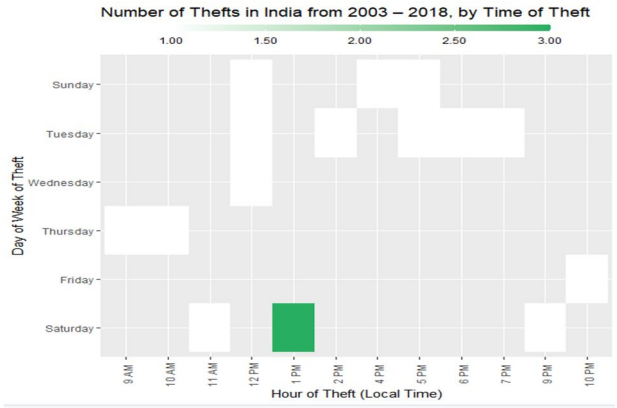


Fig 7.14 Theft Time Heatmap in india

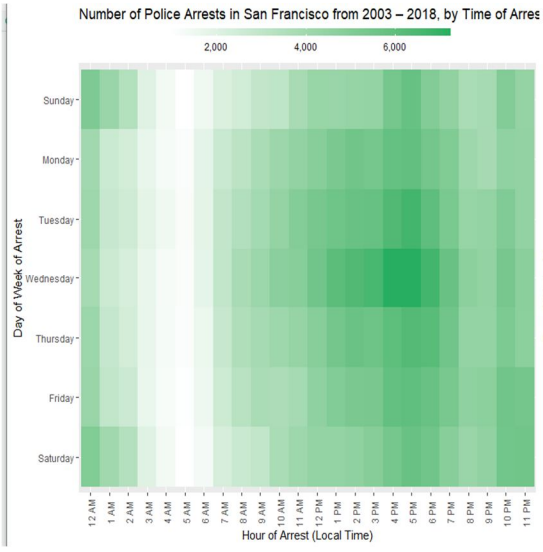


Fig 7.15 Arrest over time in San Francisco

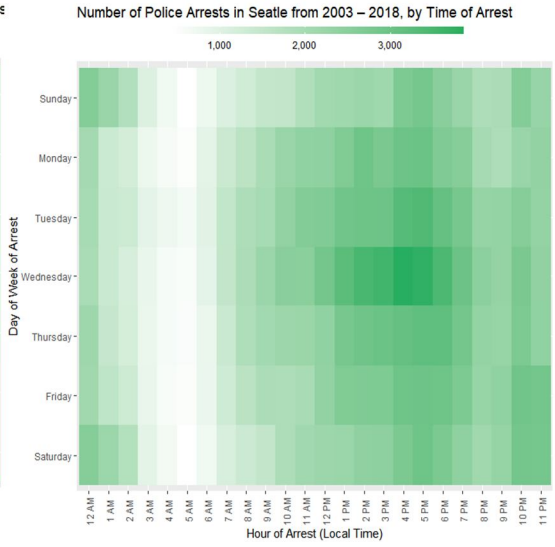


Fig 7.16 Arrest over time in Seattle

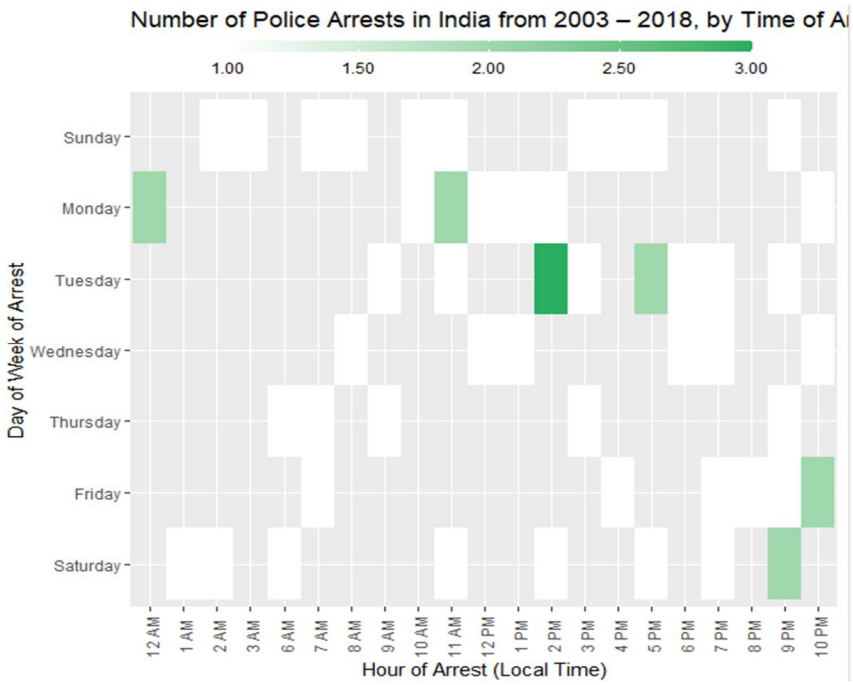


Fig 7.17 Arrest over time in India

In Correlation Analysis, plot the graph between Day of Cop and Hour of Cop by number of police arrests based on crime category and time of arrest.

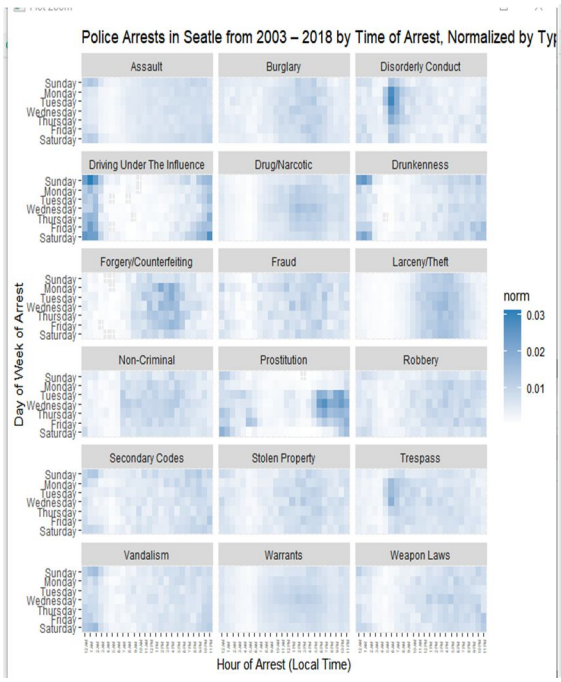


Fig 7.18 Factor by Crime category in San Francisco

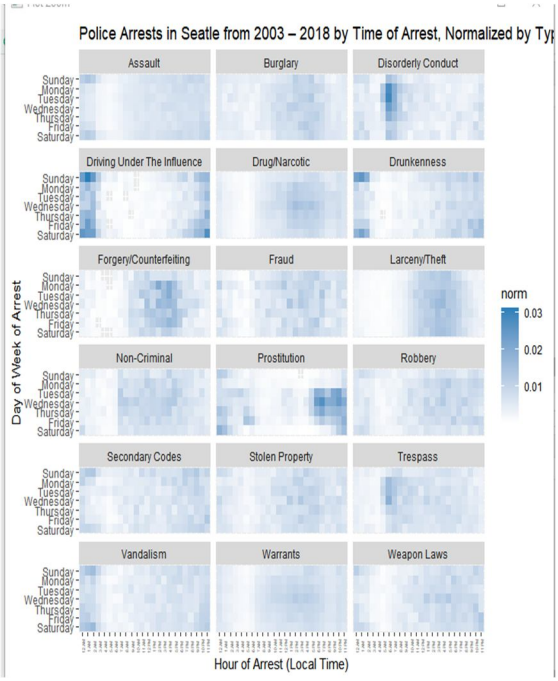


Fig 7.19 Factor by Crime Category in Seattle

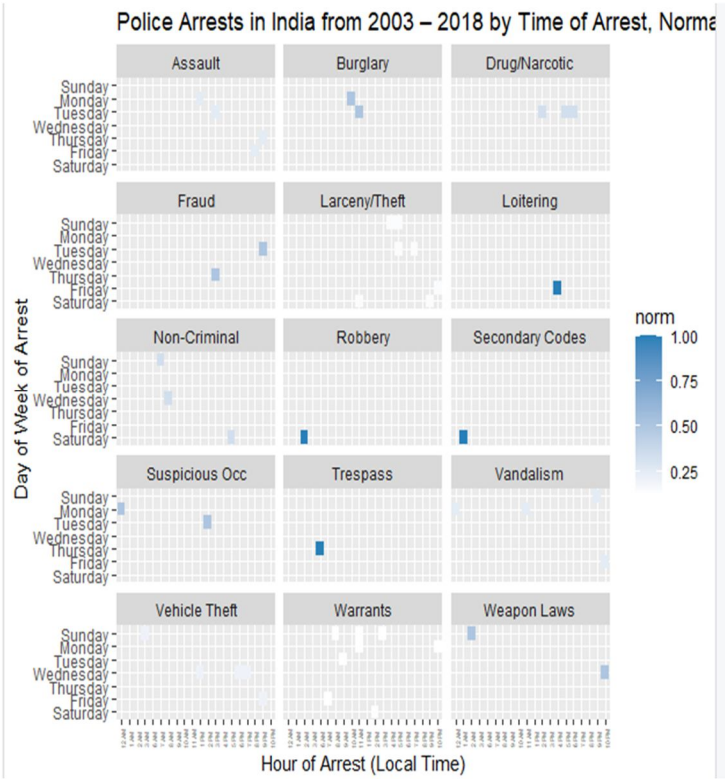


Fig 7.20 Factor by Crime category in India

In Factor by Police District, plot the graph between Day of Cop and Hour of Cop based on police arrests based on time of arrest and district.

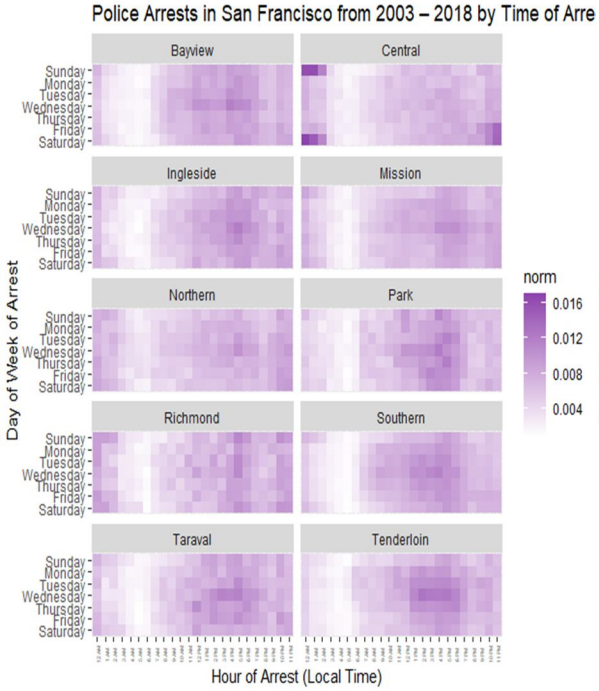


Fig 7.21 Factor by Police district in San Francisco

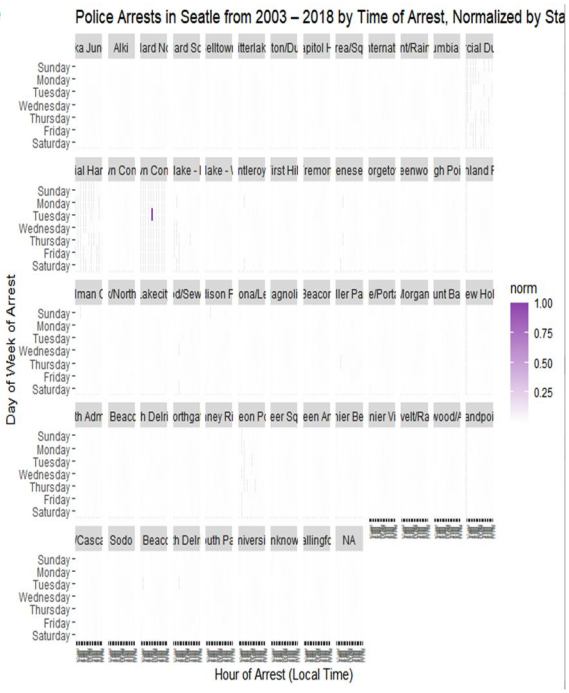


Fig 7.22 Factor by Police district in Seattle

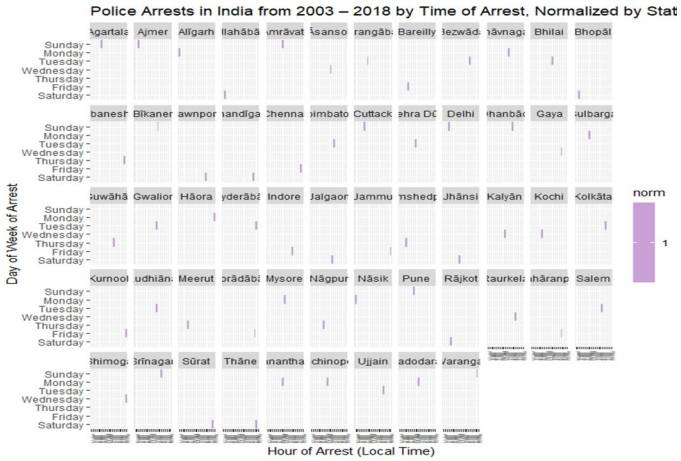


Fig 7.23 Factor by Police district in India

In Factor by Month, plot the graph between Day of Cop and Hour of Cop based on time of arrest and month.

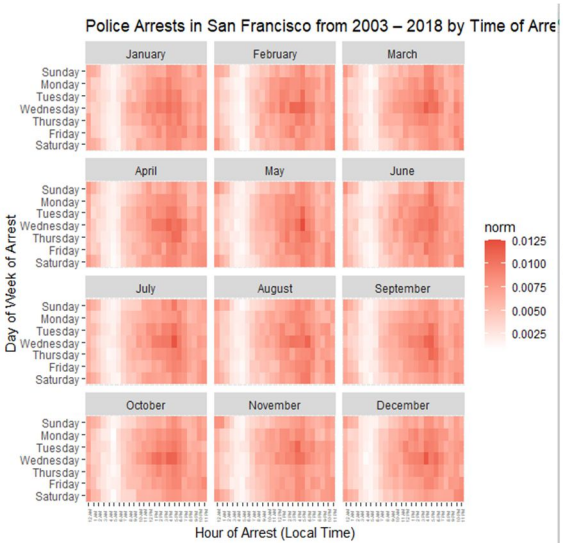


Fig 7.24 Factor by Month in San Francisco

Fig 7.25 Factor by Month in Seattle

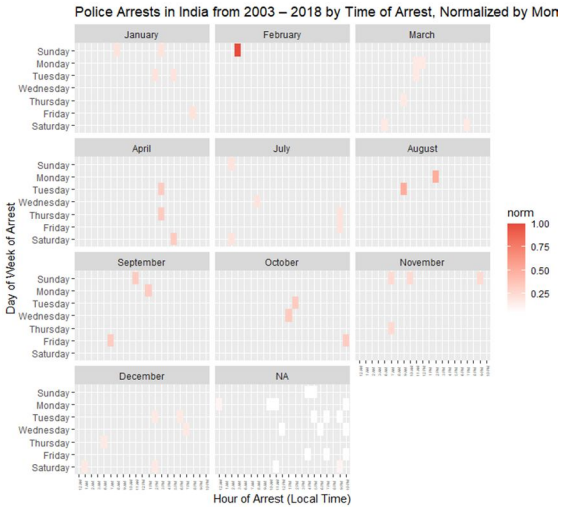


Fig 7.26 Factor by Month in India

In Factor by year, plot the graph between Day of Cop and Hour of Cop based on time of arrest and year.

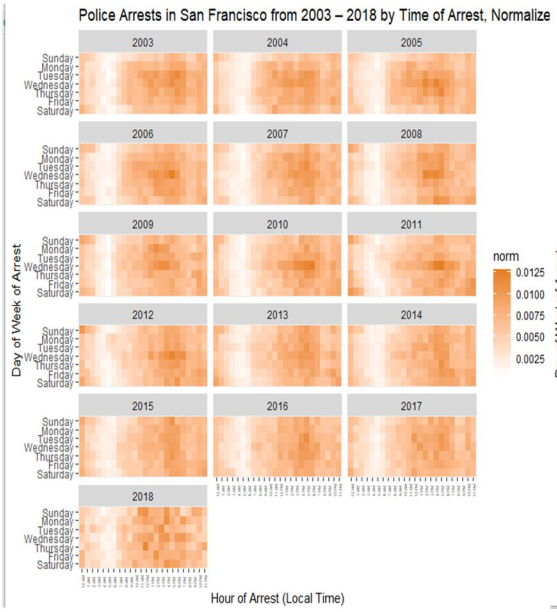


Fig 7.27 Factor by Year in San Francisco

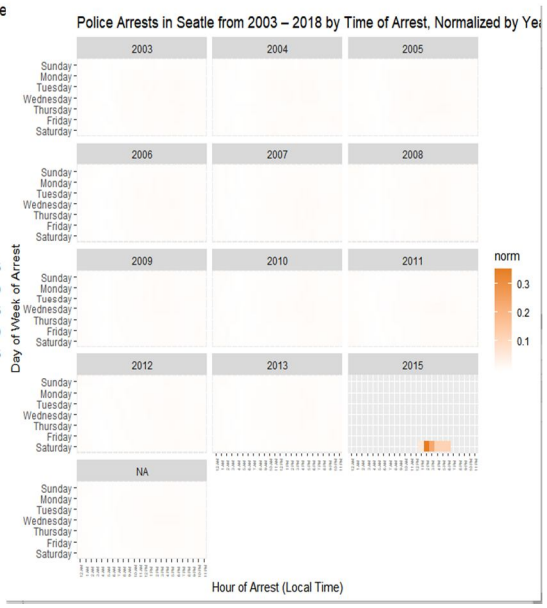


Fig 7.28 Factor by Year in Seattle

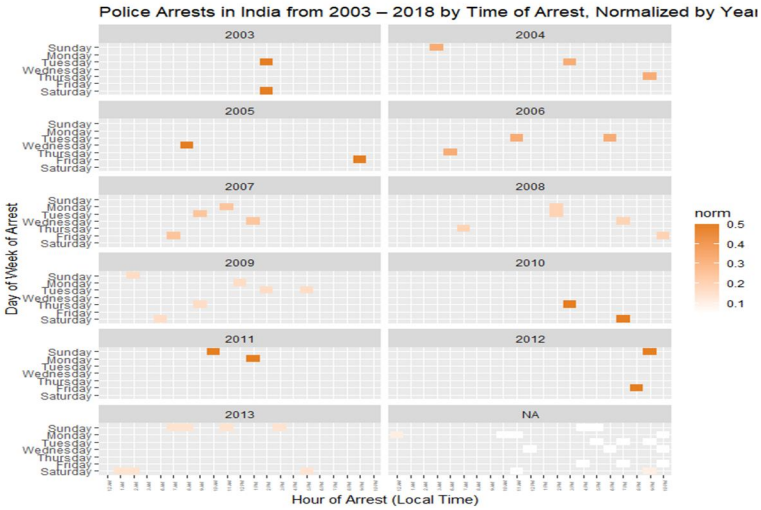


Fig 7.29 Factor by Year in India

VIII. CONLUSION

In this project, a detailed analysis of the various types of cases conducted in San Francisco, Seattle and India. In addition, this same model has tested on other crime data sets such as Chicago's crime data, which is one of the known databases. It would be interesting to see how crime in other cities compared to these crime datasets.

REFERENCES

[1] Garima, A. and Alaiad, A., 2019, June. Crime analysis in Chicago city. In 2019 10th International Conference on Information and Communication Systems (ICICS) (pp. 166-172). IEEE.

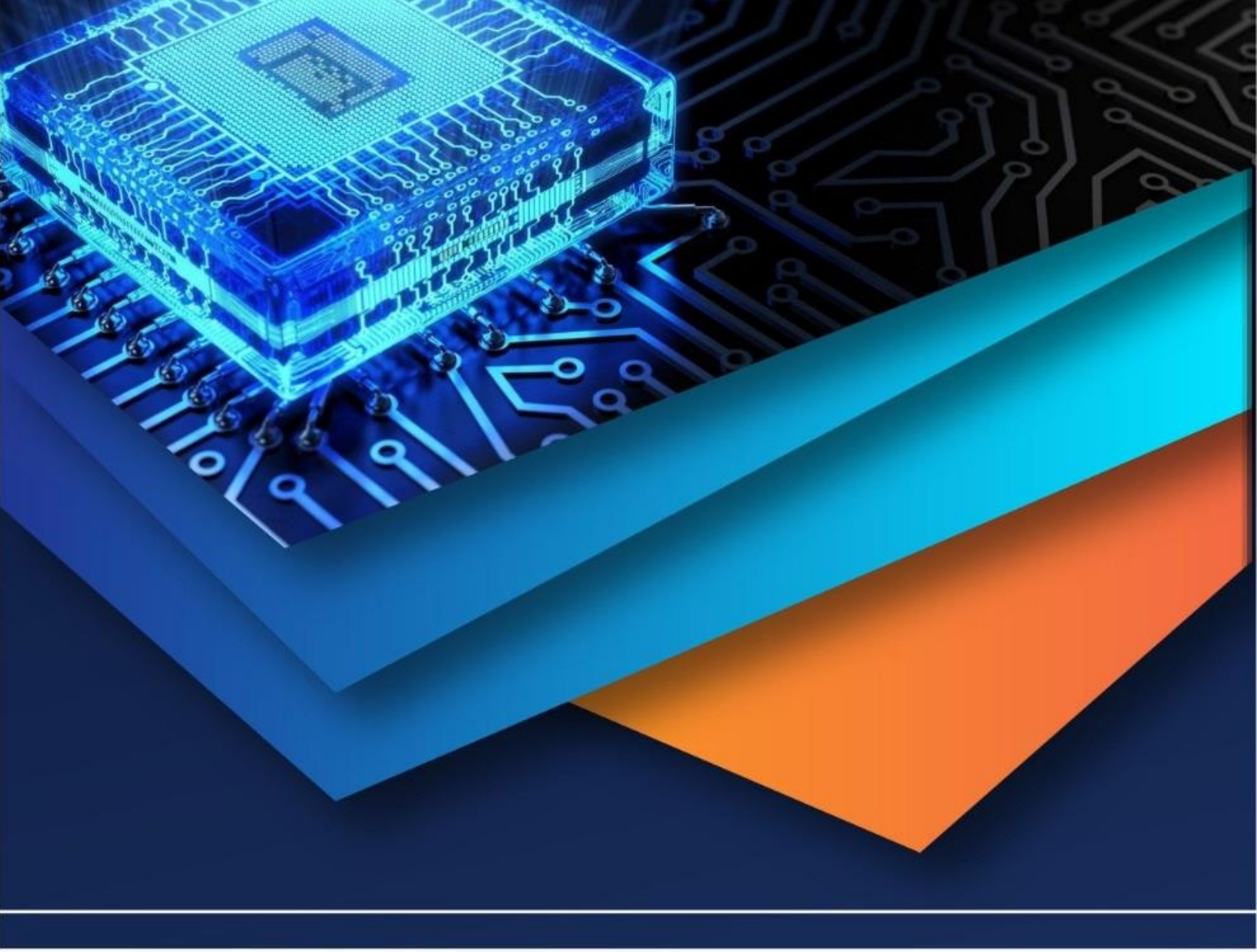
[2] Krishnendu, S.G., Lakshmi, P.P. and Nitha, L., 2020, March. Crime Analysis and Prediction using Optimized K-Means Algorithm. In 2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC) (pp. 915-918). IEEE.

[3] Bappee, F.K., Junior, A.S. and Matwin, S., 2018, May. Predicting crime using spatial features. In Canadian Conference on Artificial Intelligence (pp. 367-373). Springer, Cham.

[4] Kumar, R. and Nagpal, B., 2019. Analysis and prediction of crime patterns using big data. International Journal of Information Technology, 11(4), pp.799-805.

[5] Feng, M., Zheng, J., Han, Y., Ren, J. and Liu, Q., 2018, July. Big data analytics and mining for crime data analysis, visualization and prediction. In International Conference on Brain Inspired Cognitive Systems (pp. 605-614). Springer, Cham.

[6] Hirschfield, P. Brown and P. Todd "GIS and the analysis of spatially-referenced crime data: Experiences in Merseyside, U.K.," International Journal of Geographical Information Systems, 9(2), 1994, 191-210.
[7] E. Anderson and D. Joaquin, "Using process control chart techniques to analyse crime rates in Houston, Texas," Journal of the Operational Research Society, 47(7), 1996, 871-881.
[8] D.E. Brown, "The Regional Crime Analysis Program (RECAP): A Framework for to Catch Criminals," Proceedings of the IEEE International Conference on Systems, Man and Cybernetics, 3, 1998, 2848-2853.
[9] S.F Messner, L. Anselin, R. D. Baller, D.F. Hawkins, G. Deane, and S.E. Tolnay, "The Spatial Patterning of County Homicide Rates: An Application of Exploratory Spatial Data Analysis," Journal of Quantitative Criminology, 15(4), 1999, 423-424.
[10] A.T. Murray, M. Ingrid, S.W. John and P. Mullins, "Exploratory Spatial Data Analysis Techniques for Examining Urban Crime: Implications for Evaluating Treatment," The British Journal of Criminology,41(2), 2001, 309-329.
[11] J. Ratcliffe, "Crime Mapping and the Training Needs of Law Enforcement," European Journal on Criminal Policy and Research, 10, Kluwer Academic Publishers, Netherlands, 2004, 65-83.
[12] H. Grubesic, T. Murray, and Alan "Detecting Hot spots using cluster analysis and GIS," Proceedings from the Fifth Annual International Crime Mapping Research Conference, 26, 2001.
[13] T. Laurent, A.R. Gazen, and C.T.Agnan, "GeoXp: an R package for interactive exploratory spatial data analysis," Journal of Statistical Software, 47(2), 2012.
[14] K. Kelly, K. Lai, and P.J. Wu, "Using R for Data Analysis-A Best Practice for Research," Book of Best Advanced Practices in Quantitative methods, Sage publications, 34, 2008, 535-570.
[15] J. Miller, L. W. Kennedy and J. M. Caplan, "Risk terrain modeling: brokering criminological theory and GIS methods for crime forecasting," Journal of Justice quarterly, 28(2), 2011, 360-381.
[16] A. Dyga, and M. Sławioska "Application of R in crime data analysis", Kielce University of Technology, Poland, 2015.
[17] Y. Zhao, and Y. Cen, "Crime Analyses Using R by A. Sengupta, M. Kumar and S. Upadhyay," Data Mining Applications with R, Academic Press, ch.13, 2014, 367-395.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY