Spatiotemporal analysis in LA Crime

Vinita Kumari Jhakra

Abstract— This paper focuses on analysing the crime data owned and reported by the Los Angeles Police Department (LAPD). We analyse the data to uncover the patterns and insights of the crime data. We have taken the crime data for the last three years, 2021–2023. We used the data mining techniques K-Mean, MDS, and visualisation analysis like temporal analysis, spatio-temporal analysis, and demographic analysis using Python and Tableau. With the effective collaboration of visualisation and KDD, we would find the answers we aim to investigate in this study..

1 Problem Statement

The Los Angeles Police Department (LAPD), known as the City of Los Angeles Police Department maintains the daily crimes reported in all the locations in LA. LAPD is dedicated to enhancing public safety to an area encompassing 468 square miles and 21 community areas, representing approximately over 4 million residents as of 2023 [1]. Due to high rate of crime without any information beforehand, LAPD had developed predictive policing with PredPol and Operation Laser, the department's flagship data-driven programs. Data analytics play a vital role in identifying the hot shot of crime that include to analyse and find the various factors which contribute to Violent crime. The analysis of past historical crime data uncovers the significant insight on the crime factors.

With this study, we aim to investigate the temporal and spatial patterns in LA Crime dataset for last three years i.e. 2021-2023. Temporal patterns potentially provide the insights about the crimes when they are most likely to occur. There are some trends which need to be identified:

- Which area has which type of crime occurred most? Which area is considered the least safe?
- When does different crime occurred based on time, week and months?
- How is crime distributed based on victim's demography?
- Is there any pattern of criminal's characteristic?

We will apply visual analytics on the dataset to illustrate the insights. The dataset is favourable for both type of analysis i.e. location (Lat, Long) and temporal (time, date) across last three years.

2 State of the Art

Shyam Varan Nath explains about how data mining technique specifically clustering can help in finding geo-spatial pattern of crime [2]. He took data from Sherrif's office, under non-disclosure agreements from the crime reporting system. He visualised geo-spatial plot of crime on the map, which shows the cluster and hotspot of the crimes where they are closely populated. We have used density and choropleth map and hue based on density instead of scatter plot map.

He used k-means clustering for which features were selected based on the importance of crime type on different attribute, e.g., homicide is more related to age than theft. These features were selected with the help of a domain expert and mainly used for suspect findings. He obtained different clusters on different pair of attributes, which were visualized and interpreted on map.

We have used k-means clustering based crime types on different areas and also used K-mode clustering to find the mode of combination of attributes. We have used word cloud for MO codes which explains patterns of the criminal used in crime. We have focussed mainly on Violent Crime.

Zhanhong Wang et al. collected from Shanghai Police Cases Online Information System and focused mainly on theft and robbery.

The monthly hotspots of thefts and robberies in Shanghai in 2009 are analysed and mapped by using the hotspot analysis tool of ArcGIS 9.3. They have calculated Getis-Ord G∗i to obtain clustering. GiZScore and P value were calculated for theft and robbery and based on that hot and cold zones were analysed. Cold zones were taken where p value is greater than 0.05.PCA is adopted to 18 indicators (e.g. resident population density and floating population density) to find the main related factors of the crime hot spots [3].

We have used heatmap of hours vs week to analyse the temporal crime density. Also, we have focussed on violent and property crime so plotted on choropleth map using Tableau.

Sahar Bayoumi et al. explains how nowadays, crime mapping and analysis upgraded from using pins to computer software such as GIS technology which has a significant influence on crime analysis. It can be visualized on a map to analyse where, how and why crime occurs. They collected the data for crimes on Maryland State, US in 2016 with over 71000 records. They focused on mainly three types of crimes namely -crime against property, a crime against society, a crime against person. They have used Tableau for visualisation. They have visualised scatter plot of three types of crimes on map and used bar charts for Victim rates in diff areas and explain how crime rates vary on different time [4].

We have used Choropleth map to distinguished on crime type (violent and Property) and then heatmap of all categories of time to show the temporal pattern. And performed MDS for crime pattern in different areas in LA. And performed MDS for Crime types based on time.

3 Properties of the Data

Data is owned by LAPD OpenData and provided by Los Angeles Police Department (LAPD) [1], where overall 561,270 number of crimes have been occurred in Los Angeles over the last 3 years 2021-2023. LAPD has 4 major bureaus –

Central, Valley, South and West Bureau and under these bureaus there are 21 areas all together. This dataset reflects incidents of crime occurred in the City of Los Angeles from 2021-2023 Fig. 1. The information presented originates from authentic crime reports recorded on paper, implying the possibility of inherent inaccuracies in the transcribed data due to the manual transcription process [1]. Dataset has 28 features where each row indicates crime incident. ‘DATE OCC’, ‘Date Rptd’ and ‘TIME OCC’ are the time related attributes which would help in temporal analysis to drive the crime patterns in hours, week and months. AREA is Geographic Area with in the LAPD which further joined with shape file of LAPD for plotting choropleth map. Attribute ‘Crm Cd Desc’ describes the crime occurred in LA with 137 crime types. Fig. 2 illustrate the top 10 crimes out of 137 crime types. This attribute is further enriched with classifications and sub-classifications using the UCR Compstat file and UCR handbook provided as reference documents [1]. This data enrichment would help to identify the crime pattern on higher classifications in order to conclude on Violent crime and Property crimes. Mocodes means “mode of operation” is another attribute which, post the enrichment using the MO Code numeric reference file, would provide the insights to gain the knowledge on which mode of operation is followed most by the criminals we have created the Word cloud for Violent crime. LAT and LON would be used to visualise the crime on the map which can then interpret the pattern of the crime. Victim age, Sex and Descent are the attribute related to victim’s demography.

We have started with checking the data quality of the data and ensuring the data accuracy which further would allow drawing the meaningful visualisation. As part of DQ checks, we have investigated the missing values in the dataset and using the “isnull()” function in python. Then we performed the duplicate record presence verification using the python function “duplicated()” which then dropped from the dataset. We also dropped not relevant attributes from the dataset which won’t provide the insights. Data occurred and Date reported are converted to Date type. Additional data transformation is applied to get the Hours, months and week for the crime using the Data occurred column. “Vict Sex” attribute has multiple values defined 'M', 'F', 'X', 'H', '-' along with missing and any value except the ‘M’ and ‘F’ are transformed to ‘X’ as an Unknown victim sex. This could be due some of the crimes don’t demand any victim involved. “Vict Descent” has been transformed into the meaningful description using the code and value pair provided in the source [1].

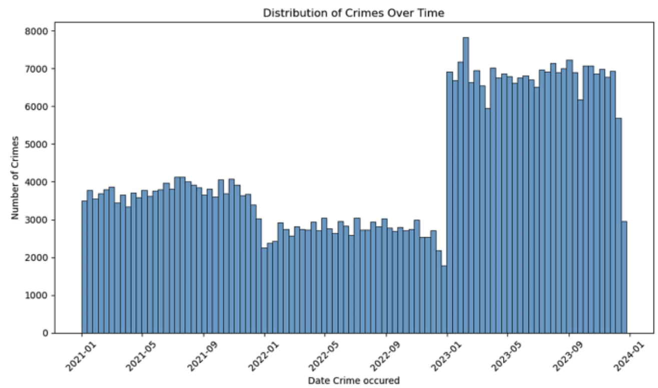


Fig. 1. Histogram of crime occurred in 2021-23

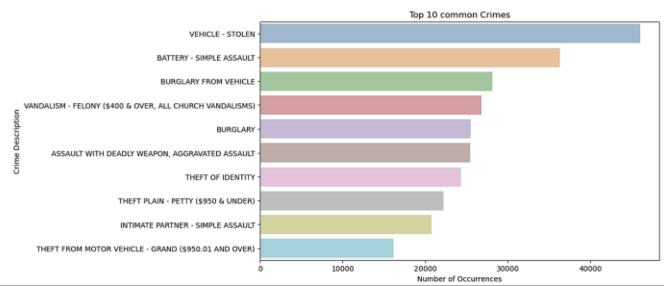


Fig. 2. Bar graph 10 most common crime

4 Analysis

4.1 Approach

Both Knowledge discovery and data mining (KDD) and visualisation methods have their know limitation when leveraged standalone. But Visual analytics brings both of these together in order to steer the analysis to get the insight and describe the underlying phenomena described in data in meaningful ways [5]. We will be using Python to perform data quality analysis, transformations, clustering and visualisation. We will also leverage Tableau for visualisations with its advantage of user-friendly GUI and easy to perform actions like aggregation and creating various insightful graphs.

This section presents the approach followed to solve the problem with effective use of Visual analytics. Fig. 3 describes the very first step is data preparation which comprises of the data quality handling (i.e. removing the duplicates, missing data handling, Outliers) and data mapping with combining the Primary Dataset file of LA Crime with reference data files for MO codes and crime classification from UCR report. The key challenge of visual analytics for Crime data is how to effectively analyse the high volume and complex dataset with time pressure and decision making, thus aggregation and elimination become necessary to deal with overcluttered data [6]. In data transformation, dimension reduction algorithm MDS is performed to illuminate the trend of 137 crime types are aggregated in 21 area and also in 24 hrs a day. We have used Python for this activity.

We performed the data mining comprises of various graphical analysis techniques Geo-Spatial analysis &

Temporal analysis and clustering algorithms i.e. K-mean, K-mode.

As part of K-mean clustering, elbow method is applied to find the appropriate no. of clusters. Human cognitive is needed here to examine to interpret the cluster against crime based on the centroids. K-mode clustering is also performed to deep dive into the further granular combination of crime vs area, ethnicity, age etc. of the victim. Spatiotemporal and Demographic analysis visualisation are also performed to find the patterns of the crime.

As part of Geo-Spatial analysis, MDS is perform whose result is further collectively analysed with Choropleth map. Human reasoning is needed to examine the visual pattern of distance between different 21 Areas in LA on the basis of Crime types.

In Temporal analysis, the pattern of crime against the seasons, days and time are analysed with the help of MDS and heatmap. MDS projects the crime types over distance of the hours and Heatmap provides the trend of crime hours of the day over the week. This requires human reasoning to identify the major hours of the crime on particular day based on color density. Line chart is plotted to illuminate the monthly trend of overall crime over the last 3 years.

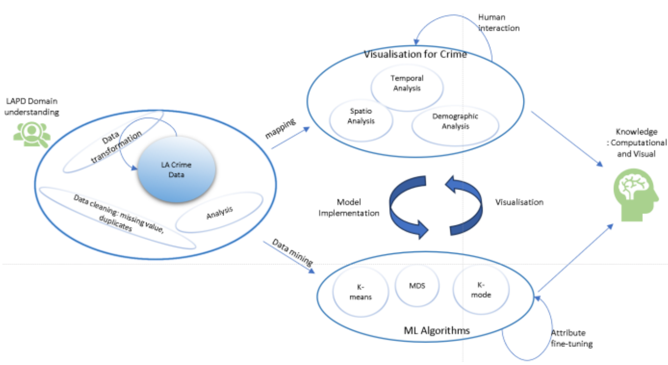


Fig. 3. Analysis workflow plan diagram. Adapted from [5]

4.2 Process

In this section, we have further described each process in the analysis. Initially, we started with visualising all the crimes on the Geo-spatial plot of the crimes. But large number of crimes made it difficult to get the insights of the pattern. So, we performed K-means clustering on aggregating the 137 crimes type over the 21 areas in LA where appropriate number of clusters i.e. 2 clusters were obtained by applying the Elbow method. We had then taken a mean of each cluster and observed that cluster 0 have got higher mean count of majority of the crime types than cluster 1.

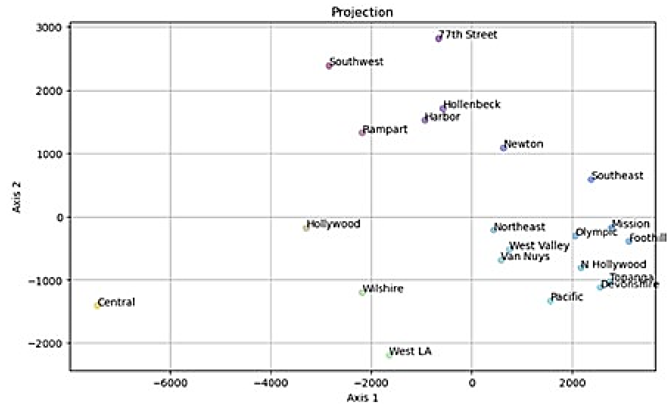


Fig. 4. 2D projection of 137 crime types in 21 Areas of LA

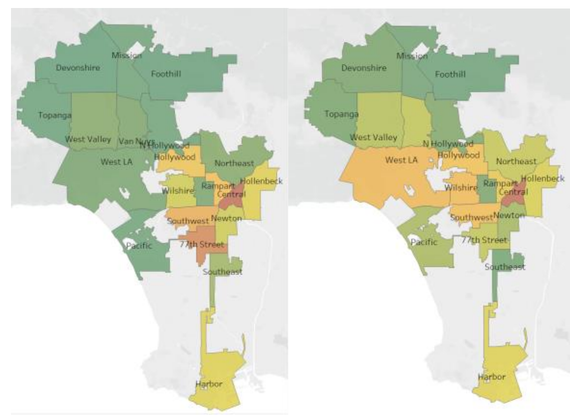


Fig. 5(a). Choropleth map-violent crime (b) Property crime

We obtained silhouette score of 57%. We had plotted those clusters against areas in Choropleth map which gave the insights that a greater number of crimes of different type of crimes are observed in cluster 0 than cluster 1. K-mean clustering only provided the insights of crime by clustering them into 2 and that is not sufficient enough to conclude the crime pattern in different areas. Then, we started with visualising all the 137 crimes types in 21 areas using bar chart and that provided quite a complex visualisation nearly impossible to interpret by human as data is cluttered. Hence, we performed dimension reduction (MDS) which is best suited to non-linear data. As also mention above, data is aggregated based on crime type of all the 21 areas. MDS helped to find the similarity of crime types using the distance between the areas in LA. Fig. 4 illustrates the projection of areas based on the crime type. All the areas with similar type of crimes are neighbour to each other where areas ‘77th Street’ and ‘Central’ shown as outlier. To understand further the pattern of the crime in these areas, we have considered the enriched crime classifications to bin the crimes in major categories. We mainly focused on two types of crimes i.e. Violent and Property crime. Color palette used in choropleth maps in Fig. 5(a) and Fig. 5(b) is “Temperature Diverging” which indicate the Green as low, Yellow as moderate and Red as High crimes in particular. These graphs showed that Central area has most crimes occurred which justified it as the outlier in Fig. 4. Also, ’77th Street’ has the 2nd largest crime occurred in Violent Crime classification whereas Property crimes are relatively less thus an outlier in Fig. 4. Area cluster on the most right shows in Fig. 4 has very less crime occurred when relate it with choropleth maps in Fig. 5(a) and Fig. 5(b) shown as green. Area "West LA" has less violent crime (Green) but moderate Property crime (Yellow). Rest other areas are following the moderate crime theme.

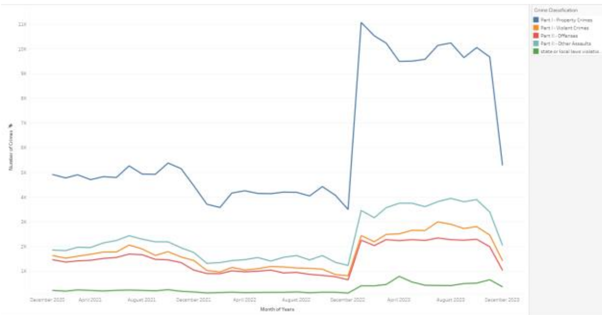
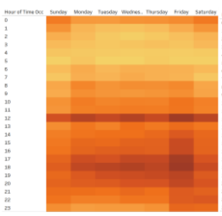
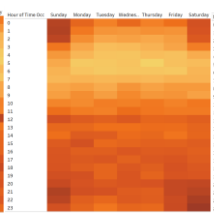


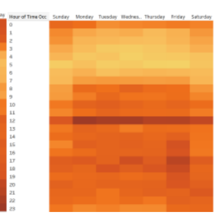
Fig. 6. Line graph- trend of crimes in major 5 classification over the 3 years on monthly basis



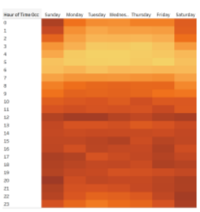
a) Part I - Property crime



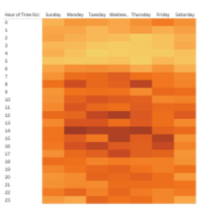
b) Part I-violent crime



c) Part II – Offences



d) Part II – Other assaults



e) State or local laws violations

Fig. 7. Heat map of major 5 categories of the crime with respect to hours in the week

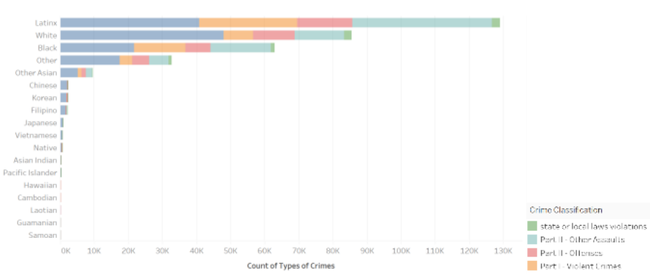


Fig. 8. Bar graph Crime type of based on victim's ethnicity

Temporal analysis: As already mentioned in last section, Line graph Fig. 6 is plotted to demonstrate the trend of crimes in major 5 classification over the 3 years on monthly basis. It gives the insights that crimes took sudden rise in 2023 where it was low in 2022 compared to 2021. We have also observed that every year, crimes went down in the month of December and again take a rise in January every year. This could be due to festive season during December and people stay at home with families. It is also noticed that property crime is always superior to others and it has taken sudden steep highest rise in year 2023.

To visualise when all the type of crime occurred on hourly basis on a day it demands the dimension reduction (MDS) by aggregating each type of crime against the hourly time on the day. We observed that most of the crimes occurred at the same time (graph not included in this report). To visualise it further, we have created Heatmap Fig. 7 of major 5 categories of the crime with respect to hours in the week. It was observed that how different types of the crime vary during the week. Part1-Property crime and Part-II offences & Other assaults more likely to occurred at 1200. Part I- violent crime occurred most during mid night of Saturdays and Sundays. Part1-Property crime also occurred during every evening but most on Friday's evening. Part-II Other assaults crimes followed the pattern of rise post afternoon. And last but not the least, State or local laws violations occurred mostly during the 0800 to 1700 on working days.

It has been observed in data that not all the crimes are reported on the day when crime occurred. Hence, we have calculated the difference between the date occurred and the date reported in no. of days. Then these are binned into the categories for e.g. 'Same day', '1 Day', '2 Day' etc. against the major 5 classification. This is then demonstrated in Fig. 9 that Part I – Violent crime and Part II – Other Assaults are mostly reported on the same day they occurred. Part II – Offenses crime has 50% are reported on the same day and rest following the '1 Day' '2 Day' and so on. It is observed that around 30% of the Property crimes are reported on the same day. It is very important to practice the prompt reporting culture so that law enforcement effectively optimise the efforts required and take preventive action as soon as possible. This would help to adjust the petrol schedules by knowing actual hours of crime during the day.

Demographic visualisation: Using the latitude and longitude provided for the crimes, we have plotted the scattered coloured with Ethnicity on Tableau so that we can see the pattern of crime against ethnicity over whole LA. It is observed that Black, White and Latin X are the ethnicity which are prime victims in LA. Black victims are spread all over across the LA but majority targeted in south side of LA. White and Latin X victims are targeted all over the LA equally. Further we have visualised using the bar plot which crime the particular ethnicity is most targeted with. Fig. 8 illustrates that Latinx ethnicity has suffered with most crimes then White followed by Black. Even though Latin X had suffered with highest no. of crimes but White had most Property crime victims. Latin X had topped in Part II – Other Assaults crimes followed by Part I - Violent crimes. As Latin X, White and Black suffered the most crime comparing to others so it may be possible that these are the ethnicity had contributed to most population as well in LA. We have also examined the pattern of the age of the victims in major 5 classifications. We have interpreted through KDE plot (figure not included) that Part I – Violent crime and Part II – Other Assaults are the major crimes classification spread across all the ages of victims that can be distinguish with hue. We have also observed that "assault with deadly weapon" and sexual abuse are the two crime classifications where Victims are under 18 ages. There is no gender biased pattern has been observed. As mentioned in earlier section that we have MO code in dataset typically stands for "Modus Operandi" that means "mode of operation" and refers to a criminals' characteristic method of committing the crime of Activities associated with the suspect in commission of the crime. It is observed that each crime has more than one MO code associated to it. So, in order to gain the knowledge which mode of operation is followed most by the criminals we have created the Word cloud for Violent crime Fig. 10 which indicates the top three mode of operations followed in crime "Hit with weapon", "Brandishes weapon" and "Victim knew Suspect". Knowing the mode of operation of the criminals, law enforcement can take preventive measures and put the course of action for tackling such criminals effectively.

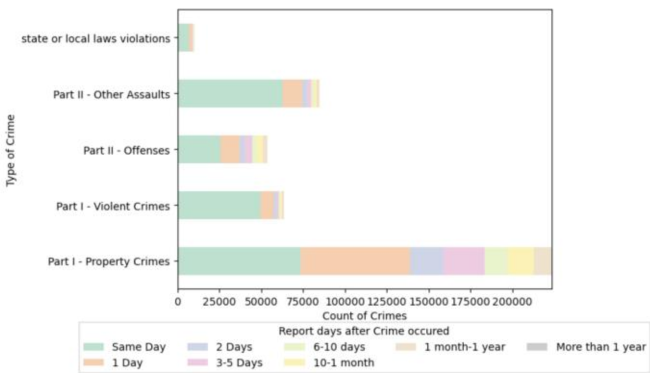


Fig. 9. Crimes classification spread across all the ages of victims.

Fig. 10. Word cloud of Modus operandi

4.3 Results

Choropleth map of K-Means cluster in Fig. 12 illustrates the dense area of all the crime types.

Also, we have identified the hot spots (of two major crime classification) marked as red and cold spots marked as green and rest crime are moderate in Fig. 5. Fig. 11 shows the density graph of all crimes, accurately identified that area "Central" is the main spot of highest number of crimes in all the 5 major crimes classification. It is also proven that central has victims belong to all the ethnicity exist in LA. Each of the 5 major crime classifications have their set pattern of crime occurred with regards to time in hours, week and month, this has been illustrated in Heatmap and line chart in Fig. 7 and Fig. 6 respectively. Fig. 8 concludes that the highest number of crimes along with highest Part II – Other Assaults crimes are occurred in Latin X descent. And White ethnicity victims are in top place in Part II - Property crime with overall 2nd place on highest number of crimes occurred. With enrichment of MO description taken by criminal, it has been illustrated in Word Cloud of violent crime in Fig. 10.

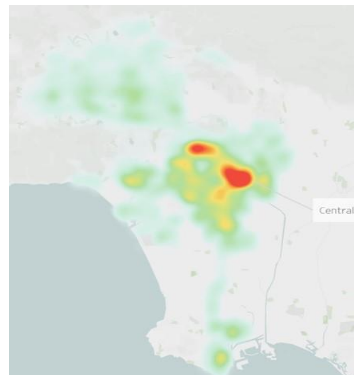


Fig. 11. Density map of overall crimes

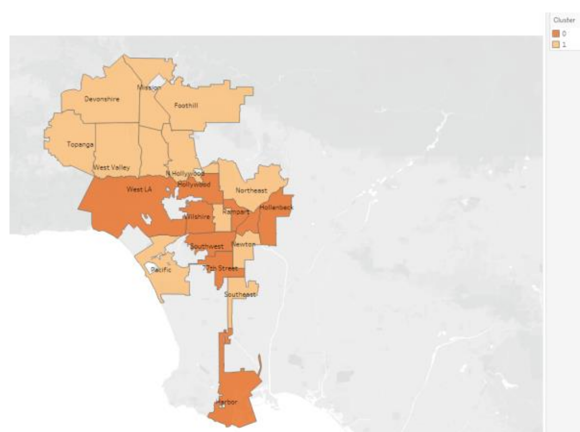


Fig. 12. Choropleth map of K-Mean clusters

5 Critical Reflection

As my data is categorical it is best suited to use the K-mode clustering algorithm. I have applied K-mode clustering by selecting the features Vict Age, Descent, Crime sub-classification and Weapon Desc. But the data is very diverse in terms of these attributes finding mode was the difficult task. So, we proceeded with other method like K-mean clustering, dimension reduction method (MDS), binning the crimes into 5 major crime classification which, together with other visual graphics, gave the better insights to my research questions. I have found my research answers at satisfactory level but this could have been improved by leveraging the other attributes like unemployment rate, inflection rate. Now day, the crimes are increasing dramatically and it is very important to focus on why part rather limiting to the how, where and when, thus my findings are partially resolving the problem of crime. This can be taken up further to next level.

We have used tableau and python for this study and it supported well with my analysis. Literature review shows that jigsaw system and ArchGIS software could interpret better results in crime analysis. SOM (Self organising map) is a unsupervised learning neural network algorithm specifically for non-linear data for dimension reduction. It is an effective algorithm for Spatio-temporal modelling for finding crime hotspots [5]. Lesson learns during this study is that crime data needs a lot of manipulation and aggregation as the data is huge and finding pattern straightaway is almost impossible and complex. Also, it is very important to understand both data and domain before performing any task of visual analytics effectively. And I have understood the equal importance of the human and computational analysis in visual analytics.

Table of word counts

Section	Expected word count	Actual word count
Problem statement	250	243
State of the art	500	485
Properties of the data	500	482
Analysis: Approach	500	441
Analysis: Process	1500	1380
Analysis: Results	200	199
Critical reflection	500	281

REFERENCES

The list below provides examples of formatting references.

[1] LAPD Online. “LAPD Organization Chart.”.

[2] Nath, Shyam Varan. “Crime Pattern Detection Using Data Mining.” In 2006 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology Workshops, 41–44, 2006

[3] Wang, Zhanhong, Jianping Wu, and Bailang Yu. “Analyzing Spatio-Temporal Distribution of Crime Hot-Spots and Their Related Factors in Shanghai, China.” In 2011 19th International Conference on Geoinformatics, 1–6, 2011

[4] Bayoumi, Sahar, Sarah AlDakhil, Eman AlNakhilan, Ebtehal Al Taleb, and Hana AlShabib. “A Review of Crime Analysis and Visualization. Case Study: Maryland State, USA.” In 2018 21st Saudi Computer Society National Computer Conference (NCC), 1–6, 2018.

[5] Keim, D., Kohlhammer, J., Ellis, G., Mansmann, F. (eds.): Mastering the Information Age : Solving Problems with Visual Analytics. Goslar : Eurographics Association (2010). DOI 10.2312/14803.

[6] Ku, Chih-Hao, Alicia Iriberri, and Goutam Jena. “Visual Analytics for Crime Analysis and Decision Support,” 2016. https://doi.org/10.4018/978-1-5225-0463-4.