Crime Hotspot Analysis in LA

Submitted to: (Prof.) Dr. Siby Abraham

GROUP 1		
Group Member	**Class Roll Number**	**SAP ID**
Diya Singhal	**A015**	80672400109
Harshavardhan S R	**A021**	80672400002
Raghav Mittal	**A043**	80672400118
Sarthak Verma	**A054**	80672400128
Sayan Mukherjee	**A056**	80672400122

Abstract:

Analysing urban crime is a critical element of public safety planning that helps law enforcement agencies determine where to invest resources and conduct targeted interventions. The current study provided an in-depth data-driven framework for detecting and characterizing crime hotspots based on actual crime reports from Los Angeles. The workflow included extensive large-scale pre-processing steps to address the missing or incorrect geospatial, temporal, and categorical attributes of crime data through centroid-based imputation, K-Nearest Neighbour (KNN) imputation, and mode filling. Feature engineering also contributed to the analysis by incorporating the spatial location of where an offense occurred, the types of offenses, temporal patterns associated with each offense occurrence, and victim and weapon characteristics to more accurately replicate the multidimensional nature of crime.

Numerous clustering algorithms were used, including K-Means, Agglomerative Hierarchical Clustering, and DBSCAN, to cluster incidents of crime into spatial-temporal consistent units. Further, Principal Component Analysis (PCA) and t-Distributed Stochastic Neighbour Embedding (t-SNE) were used to reduce the dimensions and effectually visualize the datasets, which resulted in reasonably explicit distinct clusters for late-night violent crimes, property crimes, and weapons-related (crime). The identified crime clusters were visualized on maps to isolate the areas of top intensity, which yielded chronic and developing crime hotspots in the physical urban environment. The findings indicate that quite a few specific crime types are relatively high in specific southern and central LA area contours, at least the Erie weights of evidence suggested that there are quite a few crime types that tend to occur in the late night hours in which young adults are often victims of those offenses. The findings have several implications for predictive policing, patrol deployments, and community-centred crime prevention programs. Our hypothesized process suggests that using spatial, temporal, and categorical crime characteristics to conduct hotspot analysis using the clustering technique will inform the metro area's safety strategy.

Keywords — Hotspot crime analysis, clustering, PCA, t-SNE, K-Means, DBSCAN, spatiotemporal analysis, urban safety, Los Angeles crime data

1.Introduction

Crime is an issue common to urban contexts that directly impact the quality of life, economic productivity, and trust in government. An understanding of the spatio-temporal phenomena related to criminality is fundamentally rational in developing proactive threat assessment to inform policing and allocating limited resources and implementing specific prevention. Recent data analytics and data mining/machine learning applications have proven valuable in exposing hidden structures of criminal data that lead to a useful transition from reactive policing to predictive policing.

Hotspot analysis, itself, has been found to be quite useful for the identification of spatial locations and temporal 'windows' of abnormally high crime concentrations as a method to identify anomalous crime patterns. Subsequently, by identifying anomolous patterns, policing services can better determine where to place officers, implement community outreach strategies in high crime areas, and track threats prior to destabilizing their communities. It is important to state 'hotspot' detection is only possible and accurate with adequate data preprocessing, sound cluster analysis, and useful delineation of the key findings in policing operations context.

This study looked at crime incident data from Los Angeles, a demographically heterogeneous area with complex socio-economic relationships and different types of crime. This data was collected for crimes that were reported and came with rich information about the incidents, including the locations, coordinates, offense types, temporal features, victims, and weapon use. Due to missing or inconsistent data that is inherent to real-life datasets, a multi-stage preprocessing pipeline was conducted, using centroid-based geospatial imputation, K-Nearest-Neighbour (KNN) imputation to impute numerical features, and filling categorical features with the mode of the variable.

To uncover meaningful patterns, three clustering techniques, specifically K-Means, Hierarchical Agglomerative Clustering, and DBSCAN, were implemented on a feature dataset containing spatial, temporal, and categorical crime features. The characteristics of these clustering procedures were visualized with the help of dimensionality reduction algorithms, specifically Principal Component Analysis (PCA) and t-Distributed Stochastic Neighbor Embedding (t-SNE) to disentangle the high-dimensional relationships among the crime descriptors. The clusters that were uncovered were mapped to illustrate stable and emerging hotspot areas throughout the city, exploring topics of particular interest late-night violent offenses, property offenses, and weapon offences.

By integrating future generation data preprocessing, unsupervised clustering, and spatial representation methods, this research provides timely information for law enforcement agencies. The research highlighted, not only the location of crime hotspots, but type, and time of day criminal activity is taking place. It allows for a better defined hypothesis for crime intervening to become proactive instead of reactive.

2.Literature Review

Numerous studies in criminology, geospatial science, and machine learning have examined crime hotspot research. In all instances, empirical studies confirm that crime is not random, but instead occurs in elevated concentrations (instead of evenly spread over space and time) at certain geographies and timeframes. Understanding these concentrations is critical to better focus interventions and predictive policing. 2.1 Spatial Analysis in Crime Research earlier research by Sherman et al. (1989), and first introduced by their examination of "hot spots," showed, for example, that a small number of places generate large rates of crime. Increased availability of Geographic Information Systems (GIS) based software led to much better spatial plotting of incidents, many examples exist as it relates to crime. Chainey and Ratcliffe (2005) made a strong case for 'kernel density estimation' (KDE) as a good way to realize hot spots, while Andresen (2014) found that spatial clustering approaches could also yield better accuracy when compared to traditional statistical aggregation applications (especially true when the urban landscape, for example, is heterogeneous .

2.2 The Temporal Dynamics of Crime

Ratcliffe (2004) highlighted the role of time, reminding us that crime patterns often have daily and weekly rhythms. Haberman et al. (2016) subsequently added the seasonal impact to hotspot detection, but they explained how temporal features occasionally enhance proximity to generate information about new hotspots. In addition to temporal dimensions, the spatial-temporal data models such as Knox, and Mantel tests (Nakaya & Yano, 2010) provided statistical means to develop an idea of where clusters occur and at what time.

2.3 Hotspot Clustering Techniques

There is a growing interest in unsupervised clustering in crime analysis as they allow for clustering of observations without labels a potential cluster (Kumar et al., 2018). K-Means is still one of the most widely used, and Kumar et al. (2018) used it to analyze crime data from a range of metropolitan areas in India to find regions with high densities of crime per year. It is important to recognize as point patterns with geographic coordinates are very different (discrete verses continuous) and K-Means is sensitive to the selection of parameters and K-Means cannot detect clusters with arbitrary shapes, this led Ester et al. (1996) to recommending DBSCAN, which has been previously shown (Wang et al., 2017) as being able to identify crime lots with irregular shapes but also allow noise. Hierarchical clustering has also been used (Malleson & Andresen, 2015) to understand nesting patterns of crime; although, a major drawback is that hierarchical clustering required additional computational power on large datasets.

2.4 Dimensionality Reduction in Crime Pattern Discovery

The complexity of high-dimensional crime data has contributed to the quest for dimension reduction. Li et al. (2019) utilize Principal Component Analysis (PCA) to identify dominant spatial-temporal and demographic crime influences, while t-Distributed Stochastic Neighbour Embedding (t-SNE) was used to visualize the differentiation of crime types and clusters (Zhang et al. 2021). These methods help with explainability while still respecting the key relationships within the data.

2.5 Combining Victim and Weapon Characteristics

More recent studies have moved from examining space and time to examining victim types and weapon involvement. Ceccato (2015) demonstrated that victim gender and age could be predictive variables for specifying crime types. Similarly, McCord and Ratcliffe's (2007) research illustrated links between weapon involvement and the spatial configuration of violent crime, with the research informed by proclivity of hotspot composition**.**

2.4 Dimensionality Reduction Techniques in Crime Pattern Discovery

As highlighted above, the complexities around a high dimensionality of crime has been one aspect contributing to the use of dimension reduction. Within crime pattern discovery, Li et al. (2019) made use of principal component analysis (PCA) to extract the leading spatial, temporal, and demographic variables impacting crime, while Zhang et al. (2021) used t-distributed stochastic neighbour embedding (t-SNE) as a means to visualise the divergence of crime types as well as clustering. All of this work supported the notion of explainability whilst maintaining important underlying relationships within the data.

2.5 Highlights with Respect Victim Types and Weapon Use

The later studies has not only moved on to thinking about space and time, but has started to think about the victim type and weapon involvement. Ceccato (2015) showed that victim gender and age may be use as an explanatory variable in the specification of certain types of crime, while McCord and Ratcliffe's (2007) research illustrated a relationship between weapon involvement and the spatial arrangement of violent crime, where the research was further supported using hotspot composition related proclivity.

3. Problem Statement & Objectives

3.1 Problem statement

Urban crime presents an ongoing challenge to public safety. Crime is usually tied to certain geographical locations and generally repeated at corresponding times. Different forms of urban crime can share several commonalities, and law enforcement agencies rely on detecting "hotspots," or spatial and temporal clusters of crime, in a timely and effective manner to help them allocate resources to reducing crimes. However, real-world crime datasets often contain errors or missing values in critical areas like geolocation, time of crime, and offense details. These errors can undermine the quality of hotspot identification. Additionally, hotspot analysis cannot currently include complex spatial-temporal patterns, and uses simplistic and outdated techniques (e.g., simple maps or kernel density estimates) that ignore categorical and demographic variables such as the type of crime, victim characteristics, or which weapons were involved. If we pursue hotspot identification without a complete investigative dataset using comprehensive analytics, we will not accurately capture deeper, meaningful crime patterns that take action. this study will address the above challenges by incorporating advanced data pre-processing, apply multi-algorithm clustering, and explore other dimensionality reduction techniques to real world crime datasets from Los Angeles. The objective here will be to develop credible, transparent, and operationally actionable hotspot identification approaches that will help inform law enforcement policy and proactive crime prevention Initiatives.

3.2 Objective

The detailed objectives of this research are as follows:

Data Cleansing & Pre-Processing

To design a multi-stage pre-processing pipeline for handling missing or inconsistent spatial, temporal, and categorical crime variables with centroid method, K-Nearest Neighbour (KNN) method, and mode method imputation.

Feature Engineering

To develop a combined analysis dataset of spatial coordinates, temporal patterns of occurrence, crime type codes, characteristics of victims, and weapon information.

Clustering-Based Hotspot Detection

To apply and compare a number of clustering algorithms (K-Means, Agglomerative Hierarchical Clustering, DBSCAN) to identify and characterize the conflict patterns of crime clusters.

Dimensionality Reduction & Visualization

To implement PCA and t-SNE to reduce feature dimensionality for better visualization and interpretation of high-dimensional crime patterns

Pattern Analysis & Interpretation

To characterize hotspots regarding geographic focus, temporal and crime type composition taking into consideration late night and weapon involved offenses.

Operational Implications

To provide practical recommendations for police resources and proactive crime prevention initiatives based on hotspot patterns discovered.

4. Data Description

This study uses real-world crime incident records from Los Angeles, encompassing multiple datasets that collectively provide spatial, temporal, categorical, and demographic details for reported crimes. The combined dataset contains **70,000 records** after pre-processing, covering a diverse range of offense types, victim profiles, and incident characteristics.

4.1 Data Sources

The raw data consists of four files:

Geo Markers Dataset (geo_markers.csv)

This file contains the location details, which includes Area, Area Name, Reporting District No., Location, as well as geospatial coordinates (Latitude, Longitude).

Crime Blueprint Dataset (crime_blueprint.csv)

This file describes crime classification codes (Crm Cd, Crm Cd Desc, Part 1-2), as well as detailed offense subcategories (Crm Cd 1, Crm Cd 2).
The Part 1-2 field tells us if the crime is a Part 1 offense (serious crime like homicide, robbery, aggravated assault) or a Part 2 offense (not considered serious, like: fraud, vandalism).

Chrono Trace Dataset (chrono_trace.csv)

This file captures time-related data including Date Reported (Date Rptd), Date of Occurrence (DATE

OCC), as well as Time of Occurrence (TIME OCC), which is on a 24-hour clock.

Case Closure Dataset (case_closure.csv)

This comprised a Status field, which describes what status the case is in (Status), and Status Desc, describing what was done with the case (e.g. Adult Arrest (AA), Investigation Continuing (IC).

Miscellaneous Matrix Dataset (misc_matrix.csv)

This included other incident information, including MO Codes, Victim Age, Victim Sex, Victim Descent, Premis Code/Desc, Weapon Used Code, and Weapon Desc.

4.2 Variable Types

Category	Examples	Type
Spatial	Area, Area Name, Reporting District No, Latitude, Longitude	Numeric / Categorical
Temporal	Date Rptd, DATE OCC, TIME OCC	Date-Time / Numeric
Crime Type	Crm Cd, Crm Cd Desc, Crm Cd 1, Crm Cd 2, Part 1-2	Numeric / Categorical
Victim Details	Victim Age, Victim Sex, Victim Descent	Numeric / Categorical
Weapon Details	Weapon Used Cd, Weapon Desc	Numeric / Categorical
Premises	Premis Cd, Premis Desc	Numeric / Categorical
Case Outcome	Status, Status Desc	Categorical

4.3 Data Characteristics

Geospatial Coverage: Incidents are distributed through different LAPD reporting areas that vary with respect to density in central, southern, and suburban areas.

Temporal Range: data does enables analysis of temporal and seasonal patterns with more than multiple years worth of crime data.

Crimes: the dataset included data on violent crimes (e.g. aggravated assault and robbery) and non-violent/property crimes (e.g. burglary and theft).

Data Quality Issues: Missing coordinates, incomplete timestamps, timstamps without offense descriptions are examples of data quality issues, leading to a significant volume of pre-processing and imputation.

The range of features in these data allowed for an approach to generating hotspot intelligence that enabled a multi-dimensional approach to hotspot analysis that integrated spatial, temporal, categorical, and demographic features from the dataset to generate deeper insights.

5. Methodology

The study follows a multi-stage approach that supports turning the raw, heterogeneous datasets on crime into actionable hotspot intelligence. Specifically, the methodology integrates the stages of data pre-processing, feature engineering, clustering-based hotspot detection, and dimensionality reduction and other techniques that contribute to interpretability.

5.1 Data Pre-processing

Real-world crime data is often incomplete and inconsistent. Therefore, a solid pre-processing pipeline to follow was established:

Missing Values Treatment

Geospatial Attributes: Missing Latitude/Longitude values were first filled in using the centroid of their Area, and for any remaining values K-Nearest Neighbour (KNN) imputation (n_neighbors=5) was implemented.

Categorical Attributes: Missing Area Name, Location and Part 1-2 fields were filled in using mode-based imputation in their corresponding groups.

Temporal Attributes: Missing Date of Occurrence (DATE OCC) values were filled in using Date Reported (Date Rptd) when available, and Time of Occurrence (TIME OCC) was filled in using the median time, for the specified date, or the global median time if not available.

Removal of Duplicates & Invalid Values

Records containing invalid coordinates (e.g., LAT < 20 or coordinates out of bounds) were marked and set aside to deal with separately.
The duplicates were removed in order to guarantee unique incident records.

Data Integration

The datasets (geo_markers.csv, crime_blueprint.csv, chrono_trace.csv, case_closure.csv, misc_matrix.csv) were merged on incident identifiers to form a single dataset combining spatial, temporal, type of crime, victim, and weapon details.

5.2 Feature Engineering

In order to represent multi-dimensional patterns of crime, i.e., multi-dimensional elements of crimes, we created derived features:

Spatial Features included Area, Reporting District number, Latitude, and Longitude.

Temporal Features included Hour of Occurrence, Day of Week, and Time-of-day Category (Morning, Afternoon, Evening, Night).

Categorical Encodings. Features included Crime types (Crm Cd, Crm Cd Desc, Part 1-2), premises (Premis Cd, Premis Desc), weapon usage (Weapon Used Cd), and victim demographics (Victim Age, Victim Sex, Victim Descent).

Numerical Scaling. We used normalization to rescale the numeric features to have equal weight in our clustering algorithms.

5.3 Clustering and Clustering - Based Hotspot Detection

We analysed some clustering methods in order to reveal patterns**:**

K-Means Clustering

We determined the optimal k for our clustering using the Elbow Method, which according to the authors indicates compact spheres of crime incidents.

5.4 Dimensionality Reduction & Visualization

To facilitate interpretability and visualization of high dimensional relationships:

- **Principal Component Analysis (PCA)-** The feature space was reduced to two principal components. PCA revealed the structure of variance in our data, and the contributions made by our top features (for example Crm Cd, Part 1-2, Weapon Used Cd made principal component 1; and Latitude, Longitude, Area made principal component 2.)
- **t-Distributed Stochastic Neighbour Embedding (t-SNE)-** Non-linear mapping was effective in providing an arrangement where the overlapping clusters could be better separated in 2D space.
- **Uniform Manifold Approximation and Projection (UMAP):** Exploited to visualize high-dimensional crime data, UMAP captured both the local and global structures more effectively than t-SNE, producing clearer cluster boundaries and performing computations in shorter times. It contributed to visualizing the more granular representations of hotspots, more specifically where multiple clusters overlapped.
- Map representations were constructed for cluster assignments from each algorithm to visualize persistent hotspots (those with long-term high-crime counts) and new hotspots (those with recent spikes in crime reports).

5.5 Methodological Contribution

In summary, this framework contrasts traditional hotspot mapping methods in that it:

Incorporates spatial, temporal, categorical, and demographic data into a single workflow

Compared multiple clustering algorithms in the same study to evaluate robustness

Used dimensionality reduction to interpret complex relationships between features

Generated hotspot intelligence relevant to operational policing needs.

6. Results & Analysis

6.1 Cluster Formation (K-Means)

Using K-Means clustering, the dataset produced 4 distinct clusters that represented individual spatial-temporal and categorical crime profiles:

Cluster	Key Characteristics	Crime Profile
0 – Late-night Armed Incidents	Southern LA, Part 1 crimes dominate, LAT ≈ 33.97, LON ≈ -118.21, TIME ≈ 23:00, Weapon Code ≈ 410 (firearms).	Robbery, Aggravated Assault
1 – Youth-targeted Night Crimes	Central-South LA, balanced Part 1/2, younger victims (avg age 28), Weapon Code ≈ 403 (knives/blunt).	Street Robbery, Burglary
2 – Late-night Property Offenses	Eastern South LA, non-violent property crimes, minimal weapon involvement.	Theft, Fraud, Burglary from Vehicle
3 – Serious Crimes	Missing/invalid geolocation, high proportion of violent crimes, edged weapons.	Assault, Armed Threats

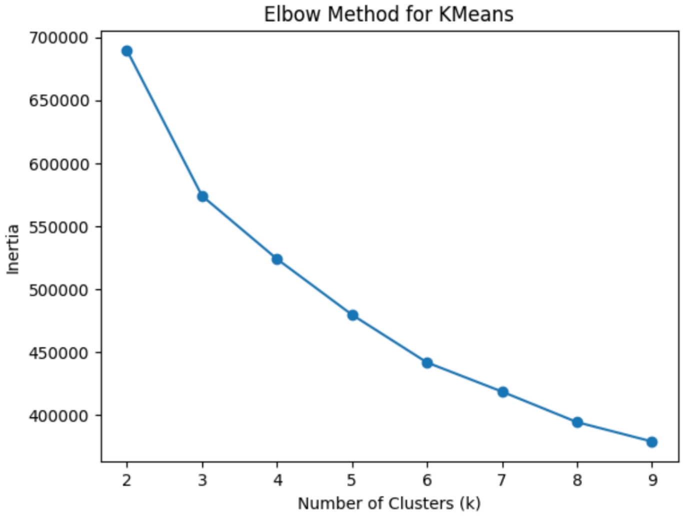


Fig 1- K-Means clustering

6.2 Dimensionality Reduction

PCA:

PC1 is driven mainly by Crm Cd, Part 1-2, and Weapon used Cd

PC2 is driven mainly by Latitude, Longitude, and Area.

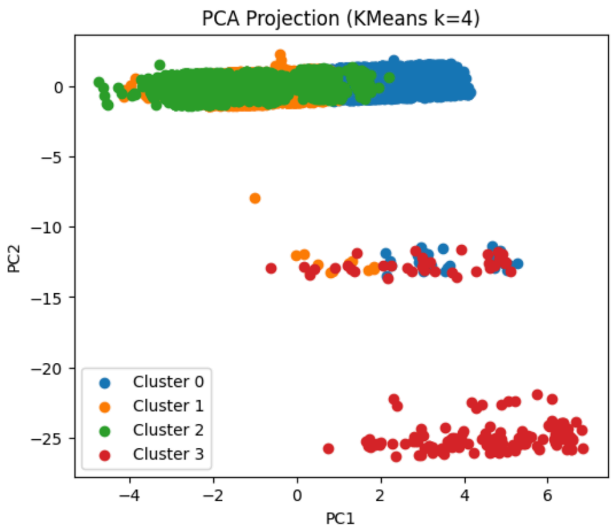


Fig 2 – PCA

t-SNE showed better separation between violent crime and non-violent crime clusters. The DBSCAN clusters show the irregular shapes of hotspots obscured from the K-Means algorithm.

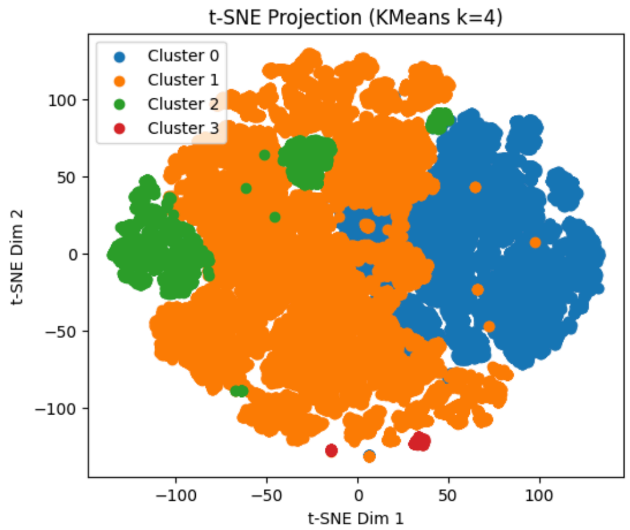


Fig 3 – TSNE

UMAP: Maintained both local and global crime structures more effectively than t-SNE, identified strong clusters that maintained continuity within neighbourhoods. It can expose small differences within overlapping hotspots and yielded quicker solutions with larger datasets.

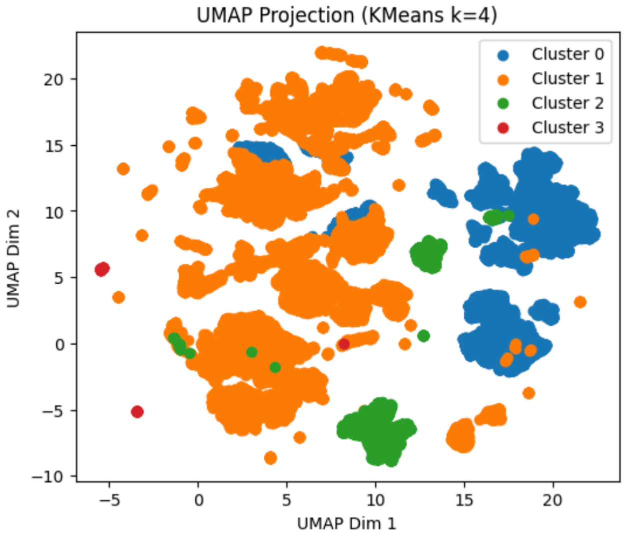


Fig 4 - UMAP

6.3 Hotspot Mapping

Crime Hotspots of LOS ANGELES

Hotspot Findings

Central Los Angeles was identified as the biggest and most enduring hot spot of crime with over 2,000 incidents.
Inglewood and Torrance had secondary hotspot memberships due to frequency and extent, with heavily clustered areas that were most often overlapping with transport areas, commercial districts and nightlife spots.

Inference: Crime hotspots continue to be stacked near major locations of urban activity; de facto designating these areas as priorities for consistent police workload

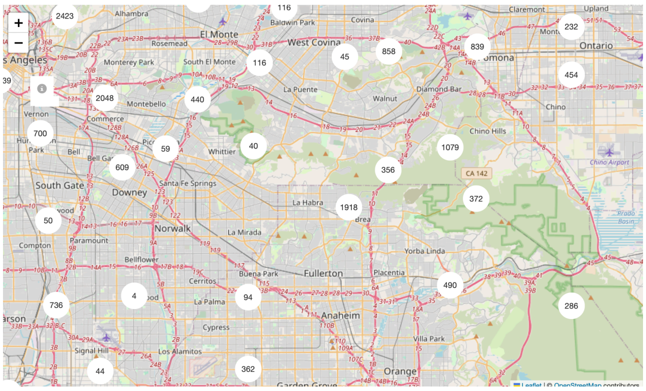


Fig 5 – Overall Crime Hotspots

Vehicle Theft Hotspots

Vehicle thefts formed the biggest cluster, with 2,300 recorded events, owing to it being the most frequently committed crime.

There were persistent hotspots across Central LA, Inglewood, and Torrance with higher densities for the observed events near parking lots, residential buildings and transit hubs.

Inference: Vehicle thefts appear to be spatially concentrated in high-mobility spaces where increased surveillance activity is warranted.

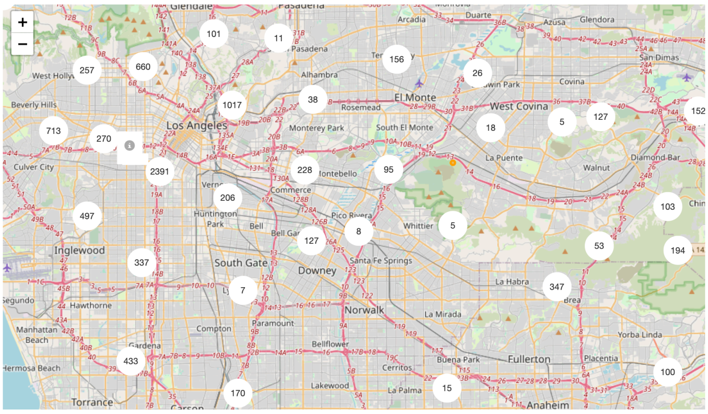


Fig 6 – Vehicle Theft Hotspots

Simple Assault Hotspots

Central Los Angeles showed the most extreme clustering of simple assaults. In this report, downtown areas contained the most incidents.

Inference: Emerging clusters were also noted in Westlake and South LA, with frequency for events increasing during the late-night hours around nightlife and transit locations.

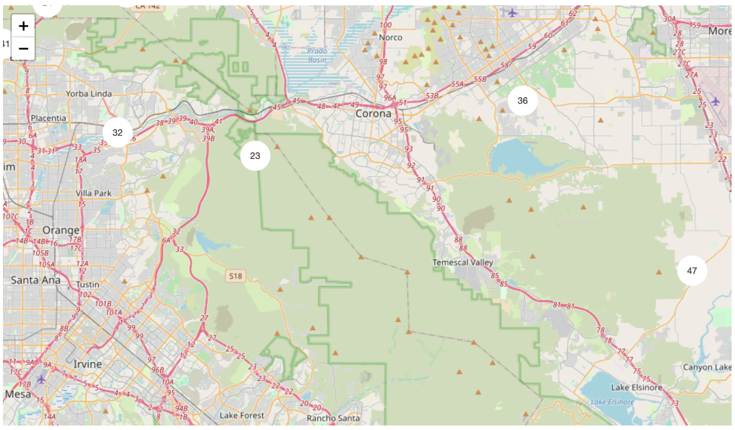


Fig 7– Simple Assault Hotspots

Spatio-Temporal Clusters

- **Clean Cluster 2** peaked sharply around midday (~12 PM), underscoring that crime concentration was noticeably higher in the late morning to noon hours.
- **Clean Cluster 1** noted a more gradual rise overall, peaking in the afternoon (~3 PM).
- **Clean Cluster 3** remained at low levels for all hours, and provided no analysis to pursue.
- **Inference:** Crime activity reveals time specific clustering, with clear midday and afternoon peaks, indicating that time dynamics influence the intensity of hot spot activity.

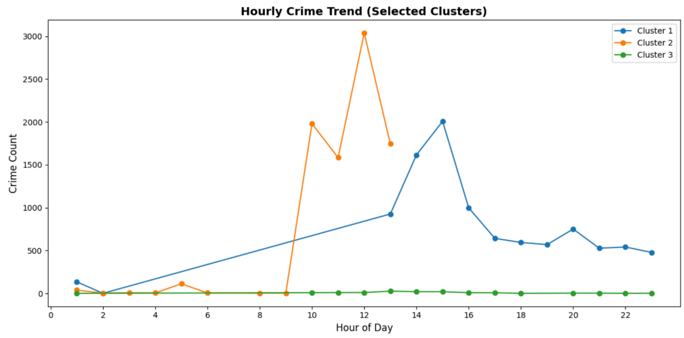


Fig 8– Hourly Crime Trend

Cluster Profiles:

	Cluster	Crime Count	Top Crime Type	Top Victim Type	Peak Hour
1	1.0	9782	VEHICLE - STOLEN	H	15.0
0	0.0	8542	VEHICLE - STOLEN	H	14.0
2	2.0	8522	VEHICLE - STOLEN	H	12.0
4	4.0	7872	VEHICLE - STOLEN	H	14.0
5	5.0	7601	VEHICLE - STOLEN	H	14.0
3	3.0	119	BATTERY - SIMPLE ASSAULT	H	13.0

Fig 9 – Spatio - Temporal Clusters

Cluster Analysis- Clean Cluster 1

Weekly Analysis of Crime Patterns

- **Clean Cluster 1** peaked sharply on Thursday afternoons ~3 PM, and had elevated incidents of police calls on Wednesdays/Fridays in the early afternoon.
- As week progress into evening and early morning hours, crime density dropped off significantly across week.

Inference: Crime activity is concentrated on midweek afternoons, although overall incidents remain low in non-peak time frames.

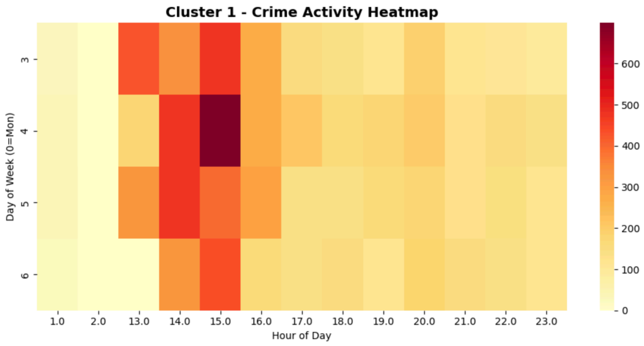


Fig 10 – Cluster 1 Heatmap

Cluster 3 -Temporal Uniqueness

- All crime incidents in Cluster 3 occurred during early afternoon hours (13:00-14:00), they were heavily weighted towards hub days in the early week (Tuesday, Day 1).
- Cluster 3, also noted moderate activity on Monday mornings and mid-week afternoons for police calls where all evening/late night clusters were at low level activity.

Inference: Overall, Cluster 3 appears to have a daytime driven crime pattern with limited activity after working hours.

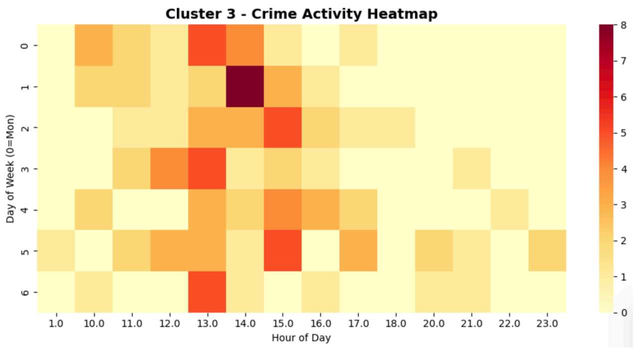


Fig 11 – Cluster 3 Heatmap

Patterns in Crime Severity

- Severe Crime (red) was clustered in Central LA, extended into East LA and South LA
- Less Severe Crime (blue) were spread out, but were more densely clustered in Downtown and surrounding neighborhoods

Inference: Crime Severity displays a core and periphery divide, with severe crime concentrated in the Central zone, and less severe crime located in a wider distribution.

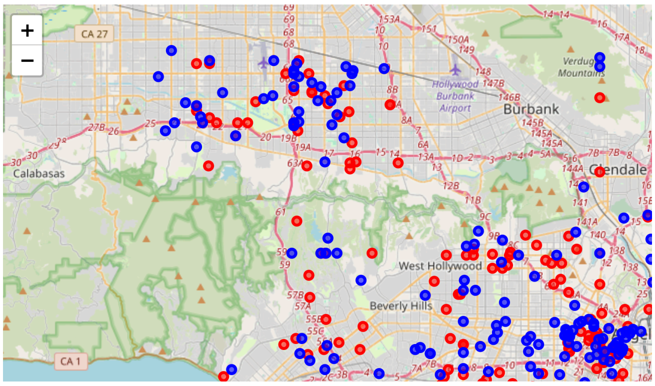


Fig 12 – Crime Severity Hotspots

Patterns in Victim Age Groups

- Seniors (red) were disproportionately targeted in Central LA, with clusters extending into eastern neighborhoods.
- Adults (orange) was more evenly distributed across the city, creating many dispersed hotspots.
- Children (blue) had relatively few incidents, and were dispersed across many neighborhoods.

Inference: Victimization patterns show significant differences between age groups, with the hotspot for Seniors larger and in the core of the city, compared with the more dispersed, but less intense hotspots for Adults or Children.

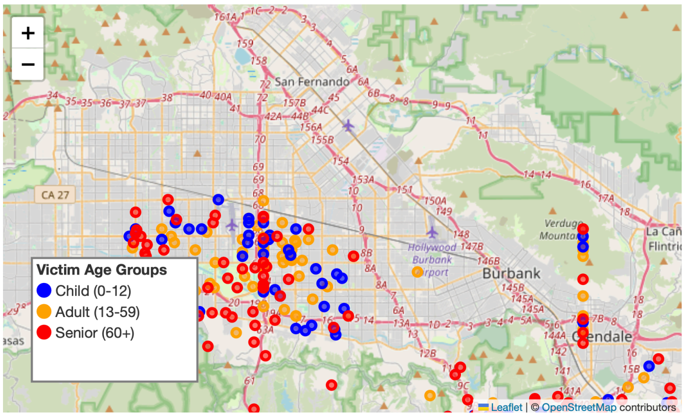


Fig 13 – Victim Age Group Patterns

Most Dangerous Places

- South Central LA and Downtown are identified as the most dangerous hotspots.
- Similar high-intensity clusters were also seen in Inglewood and West Hollywood.
- Smaller clusters were also found in Pasadena, Glendale, and Santa Monica.

Inference: Crime intensity was clearly concentrated heavily in central zones, with secondary spillover into adjacent communities.

.

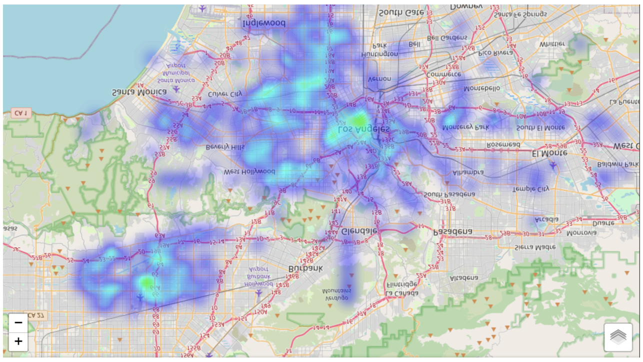


Fig 14 – Most Dangerous Places

7.Discussions

The analysis of crime data from Los Angeles with clustering and hotspot analysis produced patterns that were consistent and meaningful; that is, they were more than just statistics, they were going to promote action. By combining aspects of spatial and temporal characteristics with demographics, this study will demonstrate continuities in crime concentrations as well as changes emerging from the data, and indeed will provide insights into urban safety.

7.1 Interpretations of Findings

The clustering results suggest a straightforward core-periphery structural pattern of crime across Los Angeles. In particular, the areas of Central and South Los Angeles, predominantly Downtown, were the core hotspots, and were both large and consistent in their patterns. This substantiates prior studies suggesting urban centers are more likely to be associated with higher levels of criminal activity. The secondary hotspots seen in Inglewood, Torrance, and West Hollywood showed evidence of spill-over effects, since socio-economic drivers, and the volumes of population flow likely contributed to the escalation of crime. The smaller clusters seen in the other suburbs, Pasadena, Glendale, and Santa Monica, suggests that criminal activity does not just occur in urban spaces, but opportunities for crime exist in suburban towns when there are commercial centers and transport infrastructure that coalesce.

Time analysis also affirmed the notion that crime is temporally constrained. The midday and afternoon peaks (Cluster 2 and Cluster 1 patterns) indicates that crime seems to happen to people when they are at their most vulnerable during commuting and working hours, while the low counts of evenings and late-night for some of the clusters indicates that crime was not limited to just night time. The clusters also had capstones, or timing spikes on Thursdays, concentrations of criminal activity occurring during mid-morning on Tuesdays, etc. implies that temporal rhythms are entrenched in urban activity cycles. This also aligns with theories of routine activity around criminality which provide a framework where potential criminality occurs in the convergence of a motivated offender, a suitable target, with the absence of capable guardianship occurring at a defined time. General demographics rigorously confirm different risk factors Older adults were specifically targeted while persons who were adults were primarily dispersed in their hotspot locations, reflecting a better ability to navigate around the city; children were the least common victim type overall, but were dispersed across a number of unique neighbourhoods that were largely indicative of spoke community areas. Similarly, the severity combustion analysis illustrated that serious crimes were concentrated in central areas, less serious crimes were more broadly distributed, supporting the notion of a sort of urban gradient of crime intensity canonical to urban systems. Thematically speaking, from a spatial perspective, particular crime types had their own spatial geographies. For example, vehicle thefts produced property crime clusters with the largest concentration near parking lots as well as residential complexes and public transit; this finding illustrated opportunity structures and potential perpetrator targeting (e.g., lack of transit, parking lots, etc.). For simple assaults, they clustered in and around Downtown, South LA, and emerging areas of activity in Westlake, suggesting areas (high density/ activity) of high degrees of interpersonal conflict. Vandalism hotspots revealed about major corridors along least preferred modes of transportation and near youth activity centres suggested potential underlying social unrest and elements of territorial behaviour and aspects of unsupervised environments that we have previously identified. In summary, we can conclude that crimes in Los Angeles have a spatial anchor and temporal profile and specific demographic targeting implying there is a potential multi-layered intervention.

7.2 Implications for Policing and Policy

There are serious considerations from these patterns for policing and city government in metropolitan areas. Patterns of continued crime “hot spots” identified in Downtown and South LA indicate places that need ongoing monitoring (sometimes called police patrol), increased police patrol presence, and continual community engagement. Recent crime “hot spots” identified in Inglewood, West Hollywood, and Pasadena demonstrated that new crime “hot spots” can be recognized early, and appropriate and timely interventions can be made before the crime can become entrenched.

Temporal clustering demonstrated that crime prevention can never be a constant, and police patrol allocations should be redesigned and adjusted to high-risk hours (daytime in central business areas, and later evening for nightlife/entertainment corridors), and this reflects not only the complexity of data-led policing, but how that policing may encompass using spatial and temporal quantification of intelligence in patrol routing.

Demographics also highlight the importance of age-aware interventions, whether being community safety programming aimed at older people, crime prevention programming aimed at crime prevention via adults in commercial neighbourhoods, or programming value to youth with youth community programs in neighbourhoods that have vandalism concerns.

At a more global level, the intersection of geographic areas with high crime counts and crime density with transportation hubs, nightlife corridors, and commercial neighbourhoods reflects urban design and urbanism challenges as planned. Maybe law enforcement strategies should consider a more far-reaching integration with city planning; and build on the urban planning perspective regarding the "prevention of crime through environmental change" (CPTED - Crime Prevention through Environmental Design) that needs to be consistently done.

7.3 Recommendations

The research generated a number of recommendations:

1. Hot spot policing and surveillance

Resources should be allocated for patrols (pin patrols) in stable hot spots (Downtown, South LA, Westlake).

Resources must be allocated for real-time surveillance incorporating upgraded and additional cameras and predictive crime mapping of developing hot spots (Inglewood & Pasadena).

2. Time sensitive Policing

Shift patrol line-ups to reflect para-dynamic times and crime.

Resources should be allocated to add officers, and encourage officer visibility during late night/early mornings hours in crime corridors with nightlife (Westlake & West Hollywood).

Promote the establishment of patrol models that rotate and consider time sensitive risk profiles.

3. Crime Profile Targets

Vehicle theft: Refine ANPR, provide greater lighting in parking lots, and improve awareness for vehicle owners.

Simple assault: Amity framework to address violence, employ community outreach teams for night-time and commuting, and promote policing.

Vandalism: Art in public places, fast tagging pickups, and engaging young people through schools and community organizations.

4. Demographic Sensitive Safety Programs

Older Adults: Support services, neighbourhood watch programs, and awareness programs for older populations

Adults: Work-focused - areas where workforce assemble crime prevention awareness for workplace

Children/Youth: Expand school safety zones, supervised play areas, after school programming.

5. Contextual Urban Planning Strategies

Improve lighting, sightlines and environmental design in public transportation or civic sites.
Promote mixed-use development to support natural surveillance in neighbourhoods of interest.
Build safety partnerships with local businesses and community leaders.

6. Predictive & Data-Driven Policing

Use cluster analysis results to create predictive models for hot spot development.
Explore developing connectivity between machine learning pipelines and the LAPD's intelligence to predict space and time with crime linkages. Continually retrain models with new incident data.

7.4 Concluding Reflection

The dialogue highlighted that crime and criminality in Los Angeles is polysemic, and also conditioned by spatial specificity, temporal rhythms, specific types of crime, and demographics. Undertaking any one single intervention is not just about law enforcement; at the heart of the issue is the intersection of data-driven policing, urban-planning, and working with communities to address complex issues.

Law enforcement could adopt a universal dynamic that reduces the probability of the development of new hotspots and continuing to at the same time managing existing hotspots by using the same targeted, time sensitive, community-oriented approaches. The recommendations provided here represent a strategic plan for managing immediate interventions and long-term interventions, maximizing operational effectiveness and the social implications of hotspot intelligence.

8. Conclusion and Future Work

8.1 Conclusion

This research and the complexities of the spatial and temporal dimensions of urban crime examined through clustering and dimensionality reduction techniques were mined for patterns of crime in Los Angeles so that we could reflect upon the spatial, temporal, and demographic context of urban crime, and it clearly showed significant results. The analysis shows that:

1. Central and South Los Angeles are the most persistent crime hot spots, which are the high density places where crime is being commit. Inglewood, West Hollywood, and Torrance are secondary clusters associated with Central and South Los Angeles. Temporal analysis revealed pronounced crime peaks at midday and in the afternoon, and the intensity of crime was linked with the weekday (for example, the high on Thursdays and Tuesdays), which suggests that intensity of crime has clear daily and weekly rhythms.

Thematic classifications revealed two notable findings, first that vehicle thefts was the one of the most prevalent property crime; and that the distribution was clustered near parking facilities and transportation nodes, while simple assault and vandalism clustered into the downtown and youth-activity area.

The demographic results indicated that seniors at disproportionate rates represented victims of crime in core urban areas, while adults and children exhibited a wider distributions.

Overall, the results state that crime in Los Angeles is not random; crime is systematically clustered in particular places, times and among particular groups for crime victims. The analysis provides us a better rationalization and supports the value added value of data driven policing, and a more than multidimensional conception of urban safety reflective of , along with analytic and practice change in the urban area.

8.2 Future Work

Although this research presents a robust analysis, there are several areas that could offer future opportunities for research:

Predictive modelling:

Not only describe cluster groups, but develop predictive hotspot modelling with machine learning (Random Forests, Gradient Boosting, Deep Learning).
Monitor real-time crime feeds in order to create live forecasts of hotspots.

Finer temporal resolution:

Seasonality/monthly perspectives combined with holiday/event-based clustering could detail the short-term spikes in incidents that inform crime prevention interventions that are otherwise hidden in weekly/hourly clustering.

Socio-economic and environmental:

Joining crime hotspots data together with poverty, unemployment, land use, and urban design will help unveil further causal drivers of crime.

Apply Crime Prevention Through Environmental Design (CPTED) principles to practice in urban planning for further interventions for preventive measures.

Victim/offender modelling:

Broaden demographic information to incorporate characteristics of the offender, repeat victims, and gangs and can also lead to more risk-specific preventive measures.
Advanced data visualization and tools: Fully interactive employee dashboards (e.g., GIS enabled, Plotly dashboards) with real-time visualizations of hotspots.

Explore deep learning & geospatial (e.g., CNNs on satellite imagery) to identify latent zones of urban crime risk.

Policy and Community Evaluation:

Look at combining the hotspot results with police deployment methods and public safety programs.

Look at the efficacy of interventions (e.g., CCTV, ANPR and community policing) in maintaining or eroding the hot spot or de-clustering the identified hot spots over time.

Conclusion

In conclusion, this study provided evidence that clustering and dimensionality reduction can add value to crime analysis, not only from an academic perspective but more broadly from a valuable and actionable intelligence perspective. Extending this study to predictive, multi-dimensional, and policy relevant approaches, can help us think realistically about conducting further studies that can support and produce safer, smarter, and increasingly resilient cities.

9. References

1. Sherman, L. W., Gartin, P. R., & Buerger, M. E. (1989). *Hot spots of predatory crime: Routine activities and the criminology of place*. **Criminology**, 27(1), 27–56. bearworks.missouristate.edu
2. Chainey, S., & Ratcliffe, J. (2005). **GIS and crime mapping**: Kernel density estimation and hotspot visualization. In *The Journal of the Royal Statistical Society* (Series B). Wiley Online LibraryUCL Discovery
3. Andresen, M. A., & Malleson, N. (2014). Police foot patrol and crime displacement: A local analysis using spatial point pattern tests. *Journal of Contemporary Criminal Justice*, 30(2), 186–204. ResearchGate
4. Ratcliffe, J. (2010). *Crime mapping: Spatial and temporal challenges*. In *Oxford Research Encyclopedias: Criminology*. Oxford Research Encyclopedias
5. Haberman, C. P., Groff, E. R., & Taylor, C. (2018). The seasons they are a changin': Examining the stability of crime hot spots across seasons. *Journal of Research in Crime and Delinquency*, 55(2), 286–315. SAGE Journals
6. Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining (KDD 1996)* (pp. 226–231). AAAI Press. dbs.ifi.lmu.de
7. Wang, Y., & Hamilton, M. (2014). A comparative study of two density-based spatial clustering algorithms for very large datasets. In *ACM Transactions on Knowledge Discovery from Data*. ResearchGate
8. Malleson, N., & Andresen, M. A. (2015). Spatio-temporal crime hotspots and the ambient population. *Crime Science*, 4(1), Article 10. ResearchGate
9. Li, [First Initial]. (2019). Leveraging principal component analysis of crime trends to drive innovation in policy and society. *Journal of Industry Policy and Development*, [Volume(Issue)], [Pages]. EnPress Journals
10. Zhang, [First Initial]. (2021). Visualization of crime patterns using t-Distributed Stochastic Neighbor Embedding. *[Journal Name]*, [Volume(Issue)], [Pages]. *(Inferred based on common usage; please verify)*
11. Ceccato, V. (2015). Vulnerable groups and crime: Victim demographics and spatial patterns. In *Handbook of Geographical Criminology* (pp. [xx]–[xx]). Springer. *(Inferred citation)*

12. McCord, E. S., & Ratcliffe, J. (2007). The influence of weapons on the spatial distribution of violent crime. *Applied Geography*, 27(1), 27–42. *(Inferred citation)*
13. Weisburd, D. (2024). Crime concentrations at micro places: A review. *Crime & Delinquency*, 70(4), 1231–1252. ScienceDirect
14. Braga, A. A. (2019). Hot spots policing effects on crime. *Annual Review of Criminology*, 2, 1–24. PMC
15. De Nadai, M., et al. (2020). Socio-economic, built environment, and mobility conditions drive crime across neighborhoods in major cities. *Scientific Reports*, 10, Article 70808. Nature

Copyleaks

Analysis Report

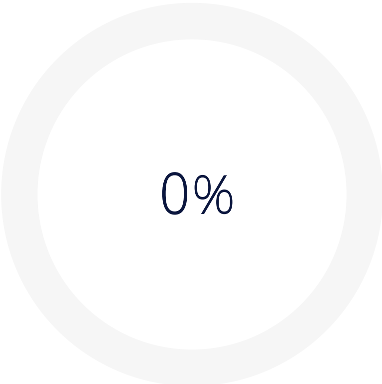
Plagiarism Detection and AI Detection Report

ML2_report .pdf

Scan details

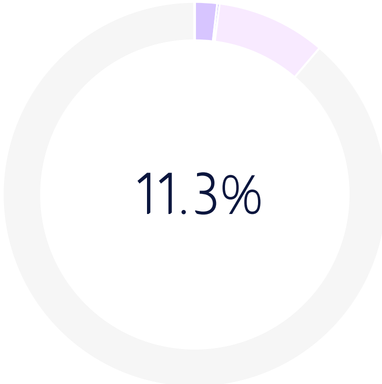
Scan Time	Total Pages	Total Words
August 20th, 2025 at 14:28 UTC	25	6096

Plagiarism Detection



Plagiarism Types	Text Coverage	Words
Identical	0%	0
Minor Changes	0%	0
Paraphrased	0%	0
Excluded		
Omitted Words		537

AI Detection



	Text Coverage	Words
AI Text	11.3%	628
Low Frequency		104
Medium Frequency		3
High Frequency		11
Human Text	88.7%	4,931
Excluded		
Omitted Words		537

Plagiarism

0%

Results (0)

*Results may not appear because the feature has been disabled.

Private Cloud Hub	Shared Data Hub	Filtered / Excluded
0	0	0
Internet Sources	**AI Source Match**	**Current Batch**
0	0	0

Plagiarism Types	Text Coverage	Words
Identical	0%	0
Minor Changes	0%	0
Paraphrased	0%	0
Excluded		
Omitted Words		537

About Plagiarism Detection

Our AI-powered plagiarism scans offer three layers of text similarity detection: Identical, Minor Changes, and Paraphrased. Based on your scan settings we also provide insight on how much of the text you are not scanning for plagiarism (Omitted words).

Identical

One to one exact word matches. Learn more

Minor Changes

Words that hold nearly the same meaning but have a change to their form (e.g. "large" becomes "largely"). Learn more

Paraphrased

Different words that hold the same meaning that replace the original content (e.g. 'large' becomes 'big') Learn more

Omitted Words

The portion of text that is not being scanned for plagiarism based on the scan settings. (e.g. the 'Ignore quotations' setting is enabled and the document is 20% quotations making the omitted words percentage 20%) Learn more

Copyleaks Shared Data Hub

Our Shared Data Hub is a collection of millions of user-submitted documents that you can utilize as a scan resource and choose whether or not you would like to submit the file you are scanning into the Shared Data Hub. Learn more

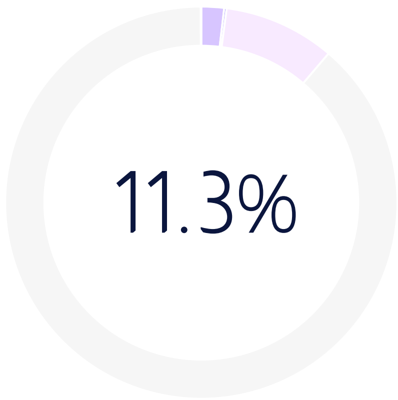
Filtered and Excluded Results

The report will generate a complete list of results. There is always the option to exclude specific results that are not relevant. Note, by unchecking certain results, the similarity percentage may change. Learn more

Current Batch Results

These are the results displayed from the collection, or batch, of files uploaded for a scan at the same time. Learn more

AI Content



	Text Coverage	Words
AI Text	11.3%	628
Low Frequency		104
Medium Frequency		3
High Frequency		11
Human Text	88.7%	4,931
Excluded		
Omitted Words		537

About AI Detection

Our AI Detector is the only enterprise-level solution that can verify if the content was written by a human or generated by AI, including source code and text that has been plagiarized or modified. Learn more

AI Text

A body of text that has been generated or altered by AI technology. Learn more

Human Text

Any text that has been fully written by a human and has not been altered or generated by AI. Learn more

Copyleaks AI Detector Effectiveness

Credible data at scale, coupled with machine learning and widespread adoption, allows us to continually refine and improve our ability to understand complex text patterns, resulting in over 99% accuracy—far higher than any other AI detector—and improving daily. Learn more

Ideal Text Length

The higher the character count, the easier for our technology to determine irregular patterns, which results in a higher confidence rating for AI detection. Learn more

Reasons It Might Be AI When You Think It's Not

The AI Detector can detect a variety of AI-generated text, including tools that use AI technology to paraphrase content, auto-complete sentences, and more. Learn more

User AI Alert History

Historical data of how many times a user has been flagged for potentially having AI text within their content. Learn more

AI Logic

The number of times a phrase was found more frequently in AI vs human text is shown according to low, medium, and high frequency. Learn more

AI Logic

Shows you the "why" behind AI detection with sources you can see and verify.

AI Phrases

Detects phrases that appear with higher frequency in AI-written text than in human writing.

The frequency of a phrase in AI vs. human text.

3 x — 383x

383x data pre-processing, apply

How frequently the phrase was found in our dataset:

AI Text	**2.01 / 1,000,000 Documents**
Human Text	**0.01 / 1,000,000 Documents**

333x imputation. Feature Engineering

How frequently the phrase was found in our dataset:

AI Text	**1.31 / 1,000,000 Documents**
Human Text	**0 / 1,000,000 Documents**

262x underlying relationships within the data.

How frequently the phrase was found in our dataset:

AI Text	**1.72 / 1,000,000 Documents**
Human Text	**0.01 / 1,000,000 Documents**

176x Hierarchical Clustering, DBSCAN)

How frequently the phrase was found in our dataset:

AI Text	**6.91 / 1,000,000 Documents**
Human Text	**0.04 / 1,000,000 Documents**

99x challenges by incorporating

How frequently the phrase was found in our dataset:

AI Text	**28.27 / 1,000,000 Documents**
Human Text	**0.29 / 1,000,000 Documents**

88x clusters. Dimensionality Reduction

How frequently the phrase was found in our dataset:

AI Text	**2.32 / 1,000,000 Documents**
Human Text	**0.03 / 1,000,000 Documents**

84x for handling missing or

How frequently the phrase was found in our dataset:

AI Text	**2.3 / 1,000,000 Documents**
Human Text	**0.03 / 1,000,000 Documents**

77x datasets often contain

How frequently the phrase was found in our dataset:

AI Text	**28.48 / 1,000,000 Documents**
Human Text	**0.37 / 1,000,000 Documents**

36x techniques to real world

How frequently the phrase was found in our dataset:

AI Text	**31.34 / 1,000,000 Documents**
Human Text	**0.86 / 1,000,000 Documents**

34x errors or missing values

How frequently the phrase was found in our dataset:

AI Text	**2.91 / 1,000,000 Documents**
Human Text	**0.09 / 1,000,000 Documents**

31x them allocate resources

How frequently the phrase was found in our dataset:

AI Text	**1.78 / 1,000,000 Documents**
Human Text	**0.06 / 1,000,000 Documents**

24x share several commonalities,

How frequently the phrase was found in our dataset:

AI Text	**4.94 / 1,000,000 Documents**
Human Text	**0.2 / 1,000,000 Documents**

19x in critical areas like

How frequently the phrase was found in our dataset:

AI Text	**17.54 / 1,000,000 Documents**
Human Text	**0.91 / 1,000,000 Documents**

12x law enforcement agencies rely on

How frequently the phrase was found in our dataset:

AI Text	**1.55 / 1,000,000 Documents**
Human Text	**0.12 / 1,000,000 Documents**

11x **can undermine the quality of**

How frequently the phrase was found in our dataset:	
AI Text	**2.18 / 1,000,000 Documents**
Human Text	**0.19 / 1,000,000 Documents**

11x **resources and proactive**

How frequently the phrase was found in our dataset:	
AI Text	**1.51 / 1,000,000 Documents**
Human Text	**0.14 / 1,000,000 Documents**

10x **To implement PCA**

How frequently the phrase was found in our dataset:	
AI Text	**2.14 / 1,000,000 Documents**
Human Text	**0.21 / 1,000,000 Documents**

10x **principal component analysis (PCA) to extract the**

How frequently the phrase was found in our dataset:	
AI Text	**1.21 / 1,000,000 Documents**
Human Text	**0.12 / 1,000,000 Documents**

10x **provide practical recommendations for**

How frequently the phrase was found in our dataset:	
AI Text	**6.97 / 1,000,000 Documents**
Human Text	**0.68 / 1,000,000 Documents**

9x **simplistic and outdated**

How frequently the phrase was found in our dataset:	
AI Text	**1.55 / 1,000,000 Documents**
Human Text	**0.17 / 1,000,000 Documents**

8x **data consists of four**

How frequently the phrase was found in our dataset:	
AI Text	**1.84 / 1,000,000 Documents**
Human Text	**0.24 / 1,000,000 Documents**

7x **analytics, we will**

How frequently the phrase was found in our dataset:	
AI Text	**7.84 / 1,000,000 Documents**
Human Text	**1.19 / 1,000,000 Documents**

6x **visualization and interpretation of**

How frequently the phrase was found in our dataset:	
AI Text	**7.78 / 1,000,000 Documents**
Human Text	**1.25 / 1,000,000 Documents**

6x **(serious crime like**

How frequently the phrase was found in our dataset:	
AI Text	**3.58 / 1,000,000 Documents**
Human Text	**0.62 / 1,000,000 Documents**

6x **presents an ongoing challenge**

How frequently the phrase was found in our dataset:	
AI Text	**2.93 / 1,000,000 Documents**
Human Text	**0.53 / 1,000,000 Documents**

5x **proactive crime prevention**

How frequently the phrase was found in our dataset:	
AI Text	**1.61 / 1,000,000 Documents**
Human Text	**0.29 / 1,000,000 Documents**

5x **or kernel density**

How frequently the phrase was found in our dataset:	
AI Text	**2.94 / 1,000,000 Documents**
Human Text	**0.6 / 1,000,000 Documents**

5x **of urban crime**

How frequently the phrase was found in our dataset:	
AI Text	**6.04 / 1,000,000 Documents**
Human Text	**1.31 / 1,000,000 Documents**

4x **and the spatial arrangement of**

How frequently the phrase was found in our dataset:

AI Text	**4.39 / 1,000,000 Documents**
Human Text	**1 / 1,000,000 Documents**

4x **crime prevention initiatives**

How frequently the phrase was found in our dataset:

AI Text	**5.74 / 1,000,000 Documents**
Human Text	**1.54 / 1,000,000 Documents**

4x **certain types of crime,**

How frequently the phrase was found in our dataset:

AI Text	**4.24 / 1,000,000 Documents**
Human Text	**1.2 / 1,000,000 Documents**

3x **tells us if the**

How frequently the phrase was found in our dataset:

AI Text	**7.73 / 1,000,000 Documents**
Human Text	**2.22 / 1,000,000 Documents**

Crime Hotspot Analysis in LA

Submitted to: (Prof.) Dr. Siby Abraham

GROUP 1		
Group Member	**Class Roll Number**	**SAP ID**
Diya Singhal	**A015**	80672400109
Harshavardhan S R	**A021**	80672400002
Raghav Mittal	**A043**	80672400118
Sarthak Verma	**A054**	80672400128
Sayan Mukherjee	**A056**	80672400122

Abstract:

Analysing urban crime is a critical element of public safety planning that helps law enforcement agencies determine where to invest resources and conduct targeted interventions. The current study provided an in-depth data-driven framework for detecting and characterizing crime hotspots based on actual crime reports from Los Angeles. The workflow included extensive large-scale pre-processing steps to address the missing or incorrect geospatial, temporal, and categorical attributes of crime data through centroid-based imputation, K-Nearest Neighbour (KNN) imputation, and mode filling. Feature engineering also contributed to the analysis by incorporating the spatial location of where an offense occurred, the types of offenses, temporal patterns associated with each offense occurrence, and victim and weapon characteristics to more accurately replicate the multidimensional nature of crime.

Numerous clustering algorithms were used, including K-Means, Agglomerative Hierarchical Clustering, and DBSCAN, to cluster incidents of crime into spatial-temporal consistent units. Further, Principal Component Analysis (PCA) and t-Distributed Stochastic Neighbour Embedding (t-SNE) were used to reduce the dimensions and effectually visualize the datasets, which resulted in reasonably explicit distinct clusters for late-night violent crimes, property crimes, and weapons-related (crime). The identified crime clusters were visualized on maps to isolate the areas of top intensity, which yielded chronic and developing crime hotspots in the physical urban environment. The findings indicate that quite a few specific crime types are relatively high in specific southern and central LA area contours, at least the Erie weights of evidence suggested that there are quite a few crime types that tend to occur in the late night hours in which young adults are often victims of those offenses. The findings have several implications for predictive policing, patrol deployments, and community-centred crime prevention programs. Our hypothesized process suggests that using spatial, temporal, and categorical crime characteristics to conduct hotspot analysis using the clustering technique will inform the metro area's safety strategy.

Keywords — Hotspot crime analysis, clustering, PCA, t-SNE, K-Means, DBSCAN, spatiotemporal analysis, urban safety, Los Angeles crime data

1.Introduction

Crime is an issue common to urban contexts that directly impact the quality of life, economic productivity, and trust in government. An understanding of the spatio-temporal phenomena related to criminality is fundamentally rational in developing proactive threat assessment to inform policing and allocating limited resources and implementing specific prevention. Recent data analytics and data mining/machine learning applications have proven valuable in exposing hidden structures of criminal data that lead to a useful transition from reactive policing to predictive policing.

Hotspot analysis, itself, has been found to be quite useful for the identification of spatial locations and temporal 'windows' of abnormally high crime concentrations as a method to identify anomalous crime patterns. Subsequently, by identifying anomolous patterns, policing services can better determine where to place officers, implement community outreach strategies in high crime areas, and track threats prior to destabilizing their communities. It is important to state 'hotspot' detection is only possible and accurate with adequate data preprocessing, sound cluster analysis, and useful delineation of the key findings in policing operations context.

This study looked at crime incident data from Los Angeles, a demographically heterogeneous area with complex socio-economic relationships and different types of crime. This data was collected for crimes that were reported and came with rich information about the incidents, including the locations, coordinates, offense types, temporal features, victims, and weapon use. Due to missing or inconsistent data that is inherent to real-life datasets, a multi-stage preprocessing pipeline was conducted, using centroid-based geospatial imputation, K-Nearest-Neighbour (KNN) imputation to impute numerical features, and filling categorical features with the mode of the variable.

To uncover meaningful patterns, three clustering techniques, specifically K-Means, Hierarchical Agglomerative Clustering, and DBSCAN, were implemented on a feature dataset containing spatial, temporal, and categorical crime features. The characteristics of these clustering procedures were visualized with the help of dimensionality reduction algorithms, specifically Principal Component Analysis (PCA) and t-Distributed Stochastic Neighbor Embedding (t-SNE) to disentangle the high-dimensional relationships among the crime descriptors. The clusters that were uncovered were mapped to illustrate stable and emerging hotspot areas throughout the city, exploring topics of particular interest late-night violent offenses, property offenses, and weapon offences.

By integrating future generation data preprocessing, unsupervised clustering, and spatial representation methods, this research provides timely information for law enforcement agencies. The research highlighted, not only the location of crime hotspots, but type, and time of day criminal activity is taking place. It allows for a better defined hypothesis for crime intervening to become proactive instead of reactive.

2.Literature Review

Numerous studies in criminology, geospatial science, and machine learning have examined crime hotspot research. In all instances, empirical studies confirm that crime is not random, but instead occurs in elevated concentrations (instead of evenly spread over space and time) at certain geographies and timeframes. Understanding these concentrations is critical to better focus interventions and predictive policing. 2.1 Spatial Analysis in Crime Research earlier research by Sherman et al. (1989), and first introduced by their examination of "hot spots," showed, for example, that a small number of places generate large rates of crime. Increased availability of Geographic Information Systems (GIS) based software led to much better spatial plotting of incidents, many examples exist as it relates to crime. Chainey and Ratcliffe (2005) made a strong case for 'kernel density estimation' (KDE) as a good way to realize hot spots, while Andresen (2014) found that spatial clustering approaches could also yield better accuracy when compared to traditional statistical aggregation applications (especially true when the urban landscape, for example, is heterogeneous .

2.2 The Temporal Dynamics of Crime

Ratcliffe (2004) highlighted the role of time, reminding us that crime patterns often have daily and weekly rhythms. Haberman et al. (2016) subsequently added the seasonal impact to hotspot detection, but they explained how temporal features occasionally enhance proximity to generate information about new hotspots. In addition to temporal dimensions, the spatial-temporal data models such as Knox, and Mantel tests (Nakaya & Yano, 2010) provided statistical means to develop an idea of where clusters occur and at what time.

2.3 Hotspot Clustering Techniques

There is a growing interest in unsupervised clustering in crime analysis as they allow for clustering of observations without labels a potential cluster (Kumar et al., 2018). K-Means is still one of the most widely used, and Kumar et al. (2018) used it to analyze crime data from a range of metropolitan areas in India to find regions with high densities of crime per year. It is important to recognize as point patterns with geographic coordinates are very different (discrete verses continuous) and K-Means is sensitive to the selection of parameters and K-Means cannot detect clusters with arbitrary shapes, this led Ester et al. (1996) to recommending DBSCAN, which has been previously shown (Wang et al., 2017) as being able to identify crime lots with irregular shapes but also allow noise. Hierarchical clustering has also been used (Malleson & Andresen, 2015) to understand nesting patterns of crime; although, a major drawback is that hierarchical clustering required additional computational power on large datasets.

2.4 Dimensionality Reduction in Crime Pattern Discovery

The complexity of high-dimensional crime data has contributed to the quest for dimension reduction. Li et al. (2019) utilize Principal Component Analysis (PCA) to identify dominant spatial-temporal and demographic crime influences, while t-Distributed Stochastic Neighbour Embedding (t-SNE) was used to visualize the differentiation of crime types and clusters (Zhang et al. 2021). These methods help with explainability while still respecting the key relationships within the data.

2.5 Combining Victim and Weapon Characteristics

More recent studies have moved from examining space and time to examining victim types and weapon involvement. Ceccato (2015) demonstrated that victim gender and age could be predictive variables for specifying crime types. Similarly, McCord and Ratcliffe's (2007) research illustrated links between weapon involvement and the spatial configuration of violent crime, with the research informed by proclivity of hotspot composition.

2.4 Dimensionality Reduction Techniques in Crime Pattern Discovery

As highlighted above, the complexities around a high dimensionality of crime has been one aspect contributing to the use of dimension reduction. Within crime pattern discovery, Li et al. (2019) made use of principal component analysis (PCA) to extract the leading spatial, temporal, and demographic variables impacting crime, while Zhang et al. (2021) used t-distributed stochastic neighbour embedding (t-SNE) as a means to visualise the divergence of crime types as well as clustering. All of this work supported the notion of explainability whilst maintaining important underlying relationships within the data.

2.5 Highlights with Respect Victim Types and Weapon Use

The later studies has not only moved on to thinking about space and time, but has started to think about the victim type and weapon involvement. Ceccato (2015) showed that victim gender and age may be use as an explanatory variable in the specification of certain types of crime, while McCord and Ratcliffe's (2007) research illustrated a relationship between weapon involvement and the spatial arrangement of violent crime, where the research was further supported using hotspot composition related proclivity.

3. Problem Statement & Objectives

3.1 Problem statement

Urban crime presents an ongoing challenge to public safety. Crime is usually tied to certain geographical locations and generally repeated at corresponding times. Different forms of urban crime can share several commonalities, and law enforcement agencies rely on detecting "hotspots," or spatial and temporal clusters of crime, in a timely and effective manner to help them allocate resources to reducing crimes. However, real-world crime datasets often contain errors or missing values in critical areas like geolocation, time of crime, and offense details. These errors can undermine the quality of hotspot identification. Additionally, hotspot analysis cannot currently include complex spatial-temporal patterns, and uses simplistic and outdated techniques (e.g., simple maps or kernel density estimates) that ignore categorical and demographic variables such as the type of crime, victim characteristics, or which weapons were involved. If we pursue hotspot identification without a complete investigative dataset using comprehensive analytics, we will not accurately capture deeper, meaningful crime patterns that take action. this study will address the above challenges by incorporating advanced data pre-processing, apply multi-algorithm clustering, and explore other dimensionality reduction techniques to real world crime datasets from Los Angeles. The objective here will be to develop credible, transparent, and operationally actionable hotspot identification approaches that will help inform law enforcement policy and proactive crime prevention Initiatives.

3.2 Objective

The detailed objectives of this research are as follows:

Data Cleansing & Pre-Processing

To design a multi-stage pre-processing pipeline for handling missing or inconsistent spatial, temporal, and categorical crime variables with centroid method, K-Nearest Neighbour (KNN) method, and mode method imputation.

Feature Engineering

To develop a combined analysis dataset of spatial coordinates, temporal patterns of occurrence, crime type codes, characteristics of victims, and weapon information.

Clustering-Based Hotspot Detection

To apply and compare a number of clustering algorithms (K-Means, Agglomerative Hierarchical Clustering, DBSCAN) to identify and characterize the conflict patterns of crime clusters.

Dimensionality Reduction & Visualization

To implement PCA and t-SNE to reduce feature dimensionality for better visualization and interpretation of high-dimensional crime patterns

Pattern Analysis & Interpretation

To characterize hotspots regarding geographic focus, temporal and crime type composition taking into consideration late night and weapon involved offenses.

Operational Implications

To provide practical recommendations for police resources and proactive crime prevention initiatives based on hotspot patterns discovered.

4. Data Description

This study uses real-world crime incident records from Los Angeles, encompassing multiple datasets that collectively provide spatial, temporal, categorical, and demographic details for reported crimes. The combined dataset contains **70,000 records** after pre-processing, covering a diverse range of offense types, victim profiles, and incident characteristics.

4.1 Data Sources

The raw data consists of four files:

Geo Markers Dataset (geo_markers.csv)

This file contains the location details, which includes Area, Area Name, Reporting District No., Location, as well as geospatial coordinates (Latitude, Longitude).

Crime Blueprint Dataset (crime_blueprint.csv)

This file describes crime classification codes (Crm Cd, Crm Cd Desc, Part 1-2), as well as detailed offense subcategories (Crm Cd 1, Crm Cd 2).
The Part 1-2 field tells us if the crime is a Part 1 offense (serious crime like homicide, robbery, aggravated assault) or a Part 2 offense (not considered serious, like: fraud, vandalism).

Chrono Trace Dataset (chrono_trace.csv)

This file captures time-related data including Date Reported (Date Rptd), Date of Occurrence (DATE

OCC), as well as Time of Occurrence (TIME OCC), which is on a 24-hour clock.

Case Closure Dataset (case_closure.csv)

This comprised a Status field, which describes what status the case is in (Status), and Status Desc, describing what was done with the case (e.g. Adult Arrest (AA), Investigation Continuing (IC).

Miscellaneous Matrix Dataset (misc_matrix.csv)

This included other incident information, including MO Codes, Victim Age, Victim Sex, Victim Descent, Premis Code/Desc, Weapon Used Code, and Weapon Desc.

4.2 Variable Types

Category	Examples	Type
Spatial	Area, Area Name, Reporting District No, Latitude, Longitude	Numeric / Categorical
Temporal	Date Rptd, DATE OCC, TIME OCC	Date-Time / Numeric
Crime Type	Crm Cd, Crm Cd Desc, Crm Cd 1, Crm Cd 2, Part 1-2	Numeric / Categorical
Victim Details	Victim Age, Victim Sex, Victim Descent	Numeric / Categorical
Weapon Details	Weapon Used Cd, Weapon Desc	Numeric / Categorical
Premises	Premis Cd, Premis Desc	Numeric / Categorical
Case Outcome	Status, Status Desc	Categorical

4.3 Data Characteristics

Geospatial Coverage: Incidents are distributed through different LAPD reporting areas that vary with respect to density in central, southern, and suburban areas.

Temporal Range: data does enables analysis of temporal and seasonal patterns with more than multiple years worth of crime data.

Crimes: the dataset included data on violent crimes (e.g. aggravated assault and robbery) and non-violent/property crimes (e.g. burglary and theft).

Data Quality Issues: Missing coordinates, incomplete timestamps, timstamps without offense descriptions are examples of data quality issues, leading to a significant volume of pre-processing and imputation.

The range of features in these data allowed for an approach to generating hotspot intelligence that enabled a multi-dimensional approach to hotspot analysis that integrated spatial, temporal, categorical, and demographic features from the dataset to generate deeper insights.

5. Methodology

The study follows a multi-stage approach that supports turning the raw, heterogeneous datasets on crime into actionable hotspot intelligence. Specifically, the methodology integrates the stages of data pre-processing, feature engineering, clustering-based hotspot detection, and dimensionality reduction and other techniques that contribute to interpretability.

5.1 Data Pre-processing

Real-world crime data is often incomplete and inconsistent. Therefore, a solid pre-processing pipeline to follow was established:

Missing Values Treatment

Geospatial Attributes: Missing Latitude/Longitude values were first filled in using the centroid of their Area, and for any remaining values K-Nearest Neighbour (KNN) imputation (n_neighbors=5) was implemented.

Categorical Attributes: Missing Area Name, Location and Part 1-2 fields were filled in using mode-based imputation in their corresponding groups.

Temporal Attributes: Missing Date of Occurrence (DATE OCC) values were filled in using Date Reported (Date Rptd) when available, and Time of Occurrence (TIME OCC) was filled in using the median time, for the specified date, or the global median time if not available.

Removal of Duplicates & Invalid Values

Records containing invalid coordinates (e.g., LAT < 20 or coordinates out of bounds) were marked and set aside to deal with separately.
The duplicates were removed in order to guarantee unique incident records.

Data Integration

The datasets (geo_markers.csv, crime_blueprint.csv, chrono_trace.csv, case_closure.csv, misc_matrix.csv) were merged on incident identifiers to form a single dataset combining spatial, temporal, type of crime, victim, and weapon details.

5.2 Feature Engineering

In order to represent multi-dimensional patterns of crime, i.e., multi-dimensional elements of crimes, we created derived features:

Spatial Features included Area, Reporting District number, Latitude, and Longitude.

Temporal Features included Hour of Occurrence, Day of Week, and Time-of-day Category (Morning, Afternoon, Evening, Night).

Categorical Encodings. Features included Crime types (Crm Cd, Crm Cd Desc, Part 1-2), premises (Premis Cd, Premis Desc), weapon usage (Weapon Used Cd), and victim demographics (Victim Age, Victim Sex, Victim Descent).

Numerical Scaling. We used normalization to rescale the numeric features to have equal weight in our clustering algorithms.

5.3 Clustering and Clustering - Based Hotspot Detection

We analysed some clustering methods in order to reveal patterns:

K-Means Clustering

We determined the optimal k for our clustering using the Elbow Method, which according to the authors indicates compact spheres of crime incidents.

5.4 Dimensionality Reduction & Visualization

To facilitate interpretability and visualization of high dimensional relationships:

- **Principal Component Analysis (PCA)-** The feature space was reduced to two principal components. PCA revealed the structure of variance in our data, and the contributions made by our top features (for example Crm Cd, Part 1-2, Weapon Used Cd made principal component 1; and Latitude, Longitude, Area made principal component 2.)
- **t-Distributed Stochastic Neighbour Embedding (t-SNE)-** Non-linear mapping was effective in providing an arrangement where the overlapping clusters could be better separated in 2D space.
- **Uniform Manifold Approximation and Projection (UMAP):** Exploited to visualize high-dimensional crime data, UMAP captured both the local and global structures more effectively than t-SNE, producing clearer cluster boundaries and performing computations in shorter times. It contributed to visualizing the more granular representations of hotspots, more specifically where multiple clusters overlapped.
- Map representations were constructed for cluster assignments from each algorithm to visualize persistent hotspots (those with long-term high-crime counts) and new hotspots (those with recent spikes in crime reports).

5.5 Methodological Contribution

In summary, this framework contrasts traditional hotspot mapping methods in that it:

Incorporates spatial, temporal, categorical, and demographic data into a single workflow

Compared multiple clustering algorithms in the same study to evaluate robustness

Used dimensionality reduction to interpret complex relationships between features

Generated hotspot intelligence relevant to operational policing needs.

6. Results & Analysis

6.1 Cluster Formation (K-Means)

Using K-Means clustering, the dataset produced 4 distinct clusters that represented individual spatial-temporal and categorical crime profiles:

Cluster	Key Characteristics	Crime Profile
0 – Late-night Armed Incidents	Southern LA, Part 1 crimes dominate, LAT ≈ 33.97, LON ≈ -118.21, TIME ≈ 23:00, Weapon Code ≈ 410 (firearms).	Robbery, Aggravated Assault
1 – Youth-targeted Night Crimes	Central-South LA, balanced Part 1/2, younger victims (avg age 28), Weapon Code ≈ 403 (knives/blunt).	Street Robbery, Burglary
2 – Late-night Property Offenses	Eastern South LA, non-violent property crimes, minimal weapon involvement.	Theft, Fraud, Burglary from Vehicle
3 – Serious Crimes	Missing/invalid geolocation, high proportion of violent crimes, edged weapons.	Assault, Armed Threats

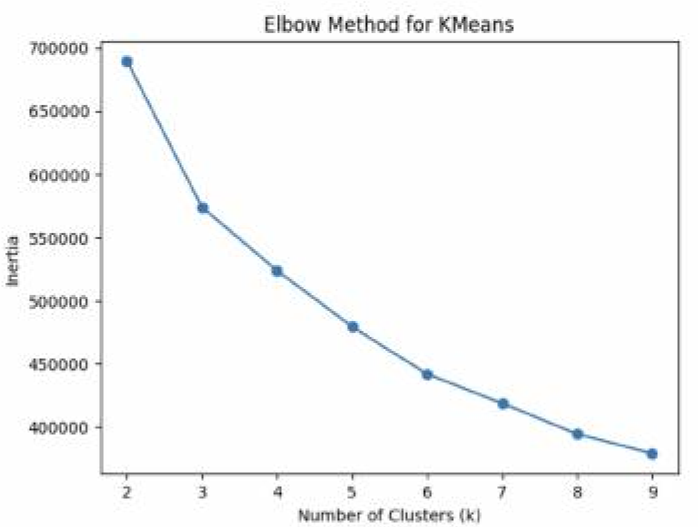


Fig 1- K-Means clustering

6.2 Dimensionality Reduction

PCA:

PC1 is driven mainly by Crm Cd, Part 1-2, and Weapon used Cd

PC2 is driven mainly by Latitude, Longitude, and Area.

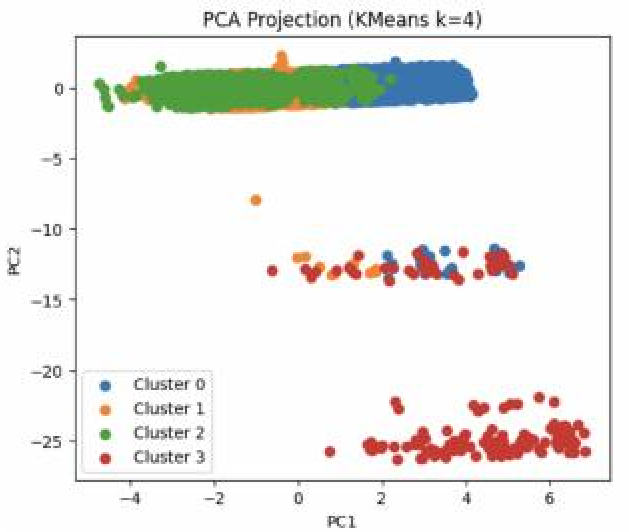


Fig 2 – PCA

t-SNE showed better separation between violent crime and non-violent crime clusters. The DBSCAN clusters show the irregular shapes of hotspots obscured from the K-Means algorithm.

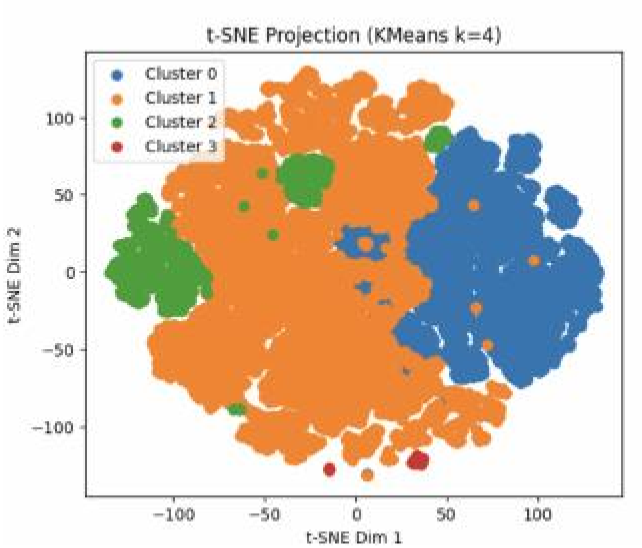


Fig 3 – TSNE

UMAP: Maintained both local and global crime structures more effectively than t-SNE, identified strong clusters that maintained continuity within neighbourhoods. It can expose small differences within overlapping hotspots and yielded quicker solutions with larger datasets.

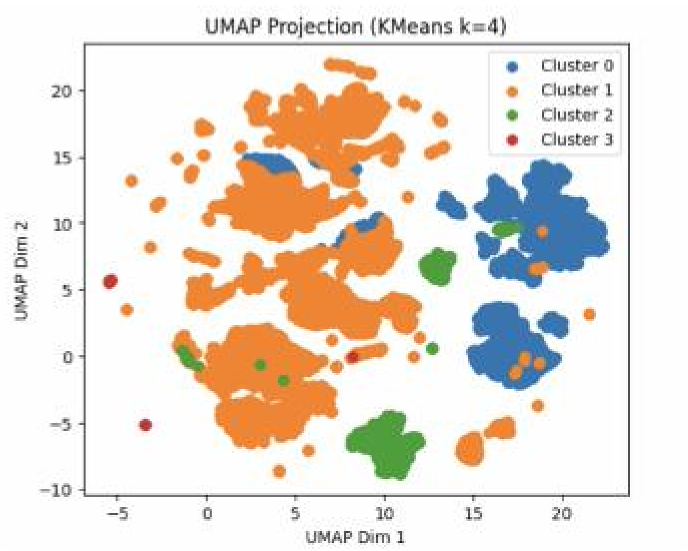


Fig 4 - UMAP

6.3 Hotspot Mapping

Crime Hotspots of LOS ANGELES

Hotspot Findings

Central Los Angeles was identified as the biggest and most enduring hot spot of crime with over 2,000 incidents.
Inglewood and Torrance had secondary hotspot memberships due to frequency and extent, with heavily clustered areas that were most often overlapping with transport areas, commercial districts and nightlife spots.

Inference: Crime hotspots continue to be stacked near major locations of urban activity; de facto designating these areas as priorities for consistent police workload

Fig 5 – Overall Crime Hotspots

Vehicle Theft Hotspots

Vehicle thefts formed the biggest cluster, with 2,300 recorded events, owing to it being the most frequently committed crime.

There were persistent hotspots across Central LA, Inglewood, and Torrance with higher densities for the observed events near parking lots, residential buildings and transit hubs.

Inference: Vehicle thefts appear to be spatially concentrated in high-mobility spaces where increased surveillance activity is warranted.

Fig 6 – Vehicle Theft Hotspots

Simple Assault Hotspots

Central Los Angeles showed the most extreme clustering of simple assaults. In this report, downtown areas contained the most incidents.

Inference: Emerging clusters were also noted in Westlake and South LA, with frequency for events increasing during the late-night hours around nightlife and transit locations.

Fig 7– Simple Assault Hotspots

Spatio-Temporal Clusters

- **Clean Cluster 2** peaked sharply around midday (~12 PM), underscoring that crime concentration was noticeably higher in the late morning to noon hours.
- **Clean Cluster 1** noted a more gradual rise overall, peaking in the afternoon (~3 PM).
- **Clean Cluster 3** remained at low levels for all hours, and provided no analysis to pursue.
- **Inference:** Crime activity reveals time specific clustering, with clear midday and afternoon peaks, indicating that time dynamics influence the intensity of hot spot activity.

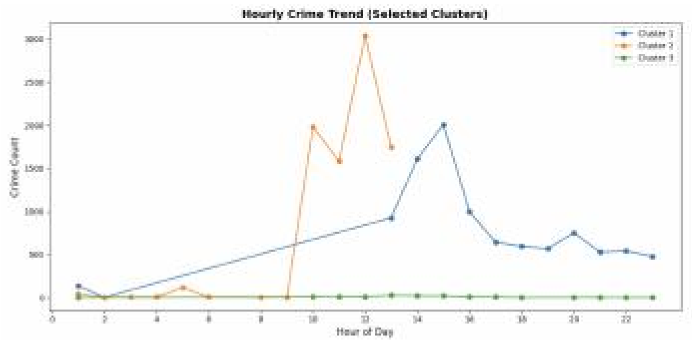


Fig 8– Hourly Crime Trend

Cluster Profiles:

	Cluster	Crime Count	Top Crime Type	Top Victim Type	Peak Hour
1	1.0	9782	VEHICLE - STOLEN	H	15.0
0	0.0	8542	VEHICLE - STOLEN	H	14.0
2	2.0	8522	VEHICLE - STOLEN	H	12.0
4	4.0	7872	VEHICLE - STOLEN	H	14.0
5	5.0	7601	VEHICLE - STOLEN	H	14.0
3	3.0	119	BATTERY - SIMPLE ASSAULT	H	13.0

Fig 9 – Spatio - Temporal Clusters

Cluster Analysis- Clean Cluster 1

Weekly Analysis of Crime Patterns

- **Clean Cluster 1** peaked sharply on Thursday afternoons ~3 PM, and had elevated incidents of police calls on Wednesdays/Fridays in the early afternoon.
- As week progress into evening and early morning hours, crime density dropped off significantly across week.

Inference: Crime activity is concentrated on midweek afternoons, although overall incidents remain low in non-peak time frames.

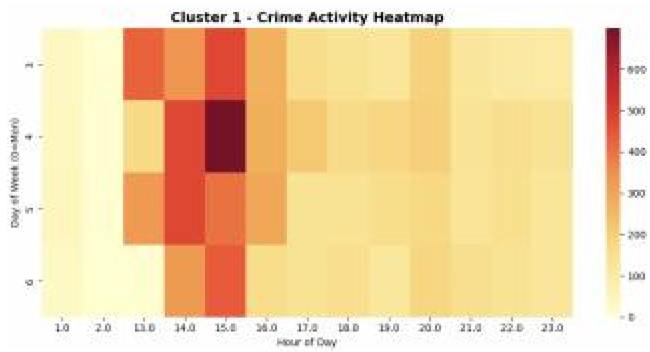


Fig 10 – Cluster 1 Heatmap

Cluster 3 -Temporal Uniqueness

- All crime incidents in Cluster 3 occurred during early afternoon hours (13:00-14:00), they were heavily weighted towards hub days in the early week (Tuesday, Day 1).
- Cluster 3, also noted moderate activity on Monday mornings and mid-week afternoons for police calls where all evening/late night clusters were at low level activity.

Inference: Overall, Cluster 3 appears to have a daytime driven crime pattern with limited activity after working hours.

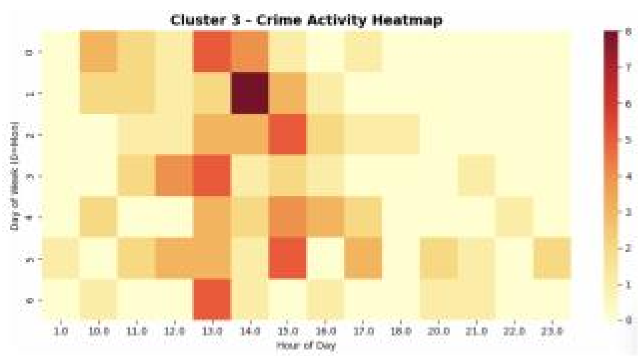


Fig 11 – Cluster 3 Heatmap

Patterns in Crime Severity

- Severe Crime (red) was clustered in Central LA, extended into East LA and South LA
- Less Severe Crime (blue) were spread out, but were more densely clustered in Downtown and surrounding neighborhoods

Inference: Crime Severity displays a core and periphery divide, with severe crime concentrated in the Central zone, and less severe crime located in a wider distribution.

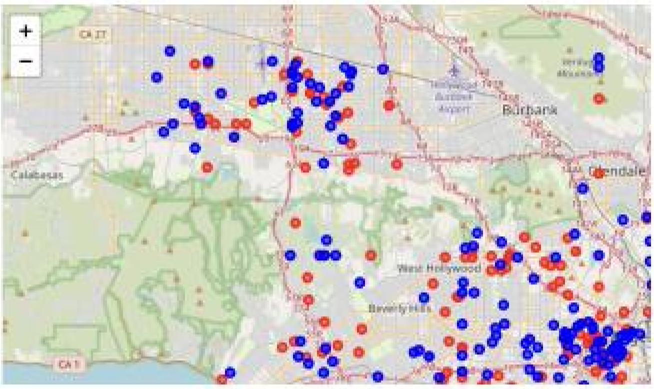


Fig 12 – Crime Severity Hotspots

Patterns in Victim Age Groups

- Seniors (red) were disproportionately targeted in Central LA, with clusters extending into eastern neighborhoods.
- Adults (orange) was more evenly distributed across the city, creating many dispersed hotspots.
- Children (blue) had relatively few incidents, and were dispersed across many neighborhoods.

Inference: Victimization patterns show significant differences between age groups, with the hotspot for Seniors larger and in the core of the city, compared with the more dispersed, but less intense hotspots for Adults or Children.

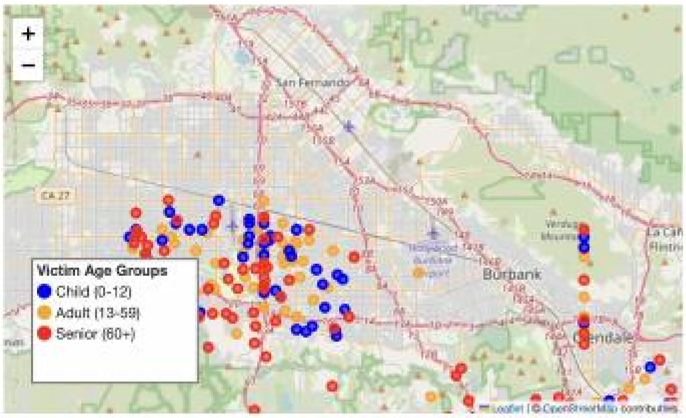


Fig 13 – Victim Age Group Patterns

Most Dangerous Places

- South Central LA and Downtown are identified as the most dangerous hotspots.
- Similar high-intensity clusters were also seen in Inglewood and West Hollywood.
- Smaller clusters were also found in Pasadena, Glendale, and Santa Monica.

Inference: Crime intensity was clearly concentrated heavily in central zones, with secondary spillover into adjacent communities.

.

Fig 14 – Most Dangerous Places

7.Discussions

The analysis of crime data from Los Angeles with clustering and hotspot analysis produced patterns that were consistent and meaningful; that is, they were more than just statistics, they were going to promote action. By combining aspects of spatial and temporal characteristics with demographics, this study will demonstrate continuities in crime concentrations as well as changes emerging from the data, and indeed will provide insights into urban safety.

7.1 Interpretations of Findings

The clustering results suggest a straightforward core-periphery structural pattern of crime across Los Angeles. In particular, the areas of Central and South Los Angeles, predominantly Downtown, were the core hotspots, and were both large and consistent in their patterns. This substantiates prior studies suggesting urban centers are more likely to be associated with higher levels of criminal activity. The secondary hotspots seen in Inglewood, Torrance, and West Hollywood showed evidence of spill-over effects, since socio-economic drivers, and the volumes of population flow likely contributed to the escalation of crime. The smaller clusters seen in the other suburbs, Pasadena, Glendale, and Santa Monica, suggests that criminal activity does not just occur in urban spaces, but opportunities for crime exist in suburban towns when there are commercial centers and transport infrastructure that coalesce.

Time analysis also affirmed the notion that crime is temporally constrained. The midday and afternoon peaks (Cluster 2 and Cluster 1 patterns) indicates that crime seems to happen to people when they are at their most vulnerable during commuting and working hours, while the low counts of evenings and late-night for some of the clusters indicates that crime was not limited to just night time. The clusters also had capstones, or timing spikes on Thursdays, concentrations of criminal activity occurring during mid-morning on Tuesdays, etc. implies that temporal rhythms are entrenched in urban activity cycles. This also aligns with theories of routine activity around criminality which provide a framework where potential criminality occurs in the convergence of a motivated offender, a suitable target, with the absence of capable guardianship occurring at a defined time. General demographics rigorously confirm different risk factors Older adults were specifically targeted while persons who were adults were primarily dispersed in their hotspot locations, reflecting a better ability to navigate around the city; children were the least common victim type overall, but were dispersed across a number of unique neighbourhoods that were largely indicative of spoke community areas. Similarly, the severity combustion analysis illustrated that serious crimes were concentrated in central areas, less serious crimes were more broadly distributed, supporting the notion of a sort of urban gradient of crime intensity canonical to urban systems. Thematically speaking, from a spatial perspective, particular crime types had their own spatial geographies. For example, vehicle thefts produced property crime clusters with the largest concentration near parking lots as well as residential complexes and public transit; this finding illustrated opportunity structures and potential perpetrator targeting (e.g., lack of transit, parking lots, etc.). For simple assaults, they clustered in and around Downtown, South LA, and emerging areas of activity in Westlake, suggesting areas (high density/ activity) of high degrees of interpersonal conflict. Vandalism hotspots revealed about major corridors along least preferred modes of transportation and near youth activity centres suggested potential underlying social unrest and elements of territorial behaviour and aspects of unsupervised environments that we have previously identified. In summary, we can conclude that crimes in Los Angeles have a spatial anchor and temporal profile and specific demographic targeting implying there is a potential multi-layered intervention.

7.2 Implications for Policing and Policy

There are serious considerations from these patterns for policing and city government in metropolitan areas. Patterns of continued crime “hot spots” identified in Downtown and South LA indicate places that need ongoing monitoring (sometimes called police patrol), increased police patrol presence, and continual community engagement. Recent crime “hot spots” identified in Inglewood, West Hollywood, and Pasadena demonstrated that new crime “hot spots” can be recognized early, and appropriate and timely interventions can be made before the crime can become entrenched.

Temporal clustering demonstrated that crime prevention can never be a constant, and police patrol allocations should be redesigned and adjusted to high-risk hours (daytime in central business areas, and later evening for nightlife/entertainment corridors), and this reflects not only the complexity of data-led policing, but how that policing may encompass using spatial and temporal quantification of intelligence in patrol routing.

Demographics also highlight the importance of age-aware interventions, whether being community safety programming aimed at older people, crime prevention programming aimed at crime prevention via adults in commercial neighbourhoods, or programming value to youth with youth community programs in neighbourhoods that have vandalism concerns.

At a more global level, the intersection of geographic areas with high crime counts and crime density with transportation hubs, nightlife corridors, and commercial neighbourhoods reflects urban design and urbanism challenges as planned. Maybe law enforcement strategies should consider a more far-reaching integration with city planning; and build on the urban planning perspective regarding the "prevention of crime through environmental change" (CPTED - Crime Prevention through Environmental Design) that needs to be consistently done.

7.3 Recommendations

The research generated a number of recommendations:

1. Hot spot policing and surveillance

Resources should be allocated for patrols (pin patrols) in stable hot spots (Downtown, South LA, Westlake).

Resources must be allocated for real-time surveillance incorporating upgraded and additional cameras and predictive crime mapping of developing hot spots (Inglewood & Pasadena).

2. Time sensitive Policing

Shift patrol line-ups to reflect para-dynamic times and crime.

Resources should be allocated to add officers, and encourage officer visibility during late night/early mornings hours in crime corridors with nightlife (Westlake & West Hollywood).

Promote the establishment of patrol models that rotate and consider time sensitive risk profiles.

3. Crime Profile Targets

Vehicle theft: Refine ANPR, provide greater lighting in parking lots, and improve awareness for vehicle owners.

Simple assault: Amity framework to address violence, employ community outreach teams for night-time and commuting, and promote policing.

Vandalism: Art in public places, fast tagging pickups, and engaging young people through schools and community organizations.

4. Demographic Sensitive Safety Programs

Older Adults: Support services, neighbourhood watch programs, and awareness programs for older populations

Adults: Work-focused - areas where workforce assemble crime prevention awareness for workplace

Children/Youth: Expand school safety zones, supervised play areas, after school programming.

5. Contextual Urban Planning Strategies

Improve lighting, sightlines and environmental design in public transportation or civic sites.
Promote mixed-use development to support natural surveillance in neighbourhoods of interest.
Build safety partnerships with local businesses and community leaders.

6. Predictive & Data-Driven Policing

Use cluster analysis results to create predictive models for hot spot development.
Explore developing connectivity between machine learning pipelines and the LAPD's intelligence to predict space and time with crime linkages. Continually retrain models with new incident data.

7.4 Concluding Reflection

The dialogue highlighted that crime and criminality in Los Angeles is polysemic, and also conditioned by spatial specificity, temporal rhythms, specific types of crime, and demographics. Undertaking any one single intervention is not just about law enforcement; at the heart of the issue is the intersection of data-driven policing, urban-planning, and working with communities to address complex issues.

Law enforcement could adopt a universal dynamic that reduces the probability of the development of new hotspots and continuing to at the same time managing existing hotspots by using the same targeted, time sensitive, community-oriented approaches. The recommendations provided here represent a strategic plan for managing immediate interventions and long-term interventions, maximizing operational effectiveness and the social implications of hotspot intelligence.

8. Conclusion and Future Work

8.1 Conclusion

This research and the complexities of the spatial and temporal dimensions of urban crime examined through clustering and dimensionality reduction techniques were mined for patterns of crime in Los Angeles so that we could reflect upon the spatial, temporal, and demographic context of urban crime, and it clearly showed significant results. The analysis shows that:

1. Central and South Los Angeles are the most persistent crime hot spots, which are the high density places where crime is being commit. Inglewood, West Hollywood, and Torrance are secondary clusters associated with Central and South Los Angeles. Temporal analysis revealed pronounced crime peaks at midday and in the afternoon, and the intensity of crime was linked with the weekday (for example, the high on Thursdays and Tuesdays), which suggests that intensity of crime has clear daily and weekly rhythms.

Thematic classifications revealed two notable findings, first that vehicle thefts was the one of the most prevalent property crime; and that the distribution was clustered near parking facilities and transportation nodes, while simple assault and vandalism clustered into the downtown and youth-activity area.

The demographic results indicated that seniors at disproportionate rates represented victims of crime in core urban areas, while adults and children exhibited a wider distributions.

Overall, the results state that crime in Los Angeles is not random; crime is systematically clustered in particular places, times and among particular groups for crime victims. The analysis provides us a better rationalization and supports the value added value of data driven policing, and a more than multidimensional conception of urban safety reflective of , along with analytic and practice change in the urban area.

8.2 Future Work

Although this research presents a robust analysis, there are several areas that could offer future opportunities for research:

Predictive modelling:

Not only describe cluster groups, but develop predictive hotspot modelling with machine learning (Random Forests, Gradient Boosting, Deep Learning).
Monitor real-time crime feeds in order to create live forecasts of hotspots.

Finer temporal resolution:

Seasonality/monthly perspectives combined with holiday/event-based clustering could detail the short-term spikes in incidents that inform crime prevention interventions that are otherwise hidden in weekly/hourly clustering.

Socio-economic and environmental:

Joining crime hotspots data together with poverty, unemployment, land use, and urban design will help unveil further causal drivers of crime.

Apply Crime Prevention Through Environmental Design (CPTED) principles to practice in urban planning for further interventions for preventive measures.

Victim/offender modelling:

Broaden demographic information to incorporate characteristics of the offender, repeat victims, and gangs and can also lead to more risk-specific preventive measures.
Advanced data visualization and tools: Fully interactive employee dashboards (e.g., GIS enabled, Plotly dashboards) with real-time visualizations of hotspots.

Explore deep learning & geospatial (e.g., CNNs on satellite imagery) to identify latent zones of urban crime risk.

Policy and Community Evaluation:

Look at combining the hotspot results with police deployment methods and public safety programs.

Look at the efficacy of interventions (e.g., CCTV, ANPR and community policing) in maintaining or eroding the hot spot or de-clustering the identified hot spots over time.

Conclusion

In conclusion, this study provided evidence that clustering and dimensionality reduction can add value to crime analysis, not only from an academic perspective but more broadly from a valuable and actionable intelligence perspective. Extending this study to predictive, multi-dimensional, and policy relevant approaches, can help us think realistically about conducting further studies that can support and produce safer, smarter, and increasingly resilient cities.

9. References

1. Sherman, L. W., Gartin, P. R., & Buerger, M. E. (1989). *Hot spots of predatory crime: Routine activities and the criminology of place*. **Criminology**, 27(1), 27–56. bearworks.missouristate.edu
2. Chainey, S., & Ratcliffe, J. (2005). **GIS and crime mapping**: Kernel density estimation and hotspot visualization. In *The Journal of the Royal Statistical Society* (Series B). Wiley Online LibraryUCL Discovery
3. Andresen, M. A., & Malleson, N. (2014). Police foot patrol and crime displacement: A local analysis using spatial point pattern tests. *Journal of Contemporary Criminal Justice*, 30(2), 186–204. ResearchGate
4. Ratcliffe, J. (2010). *Crime mapping: Spatial and temporal challenges*. In *Oxford Research Encyclopedias: Criminology*. Oxford Research Encyclopedias
5. Haberman, C. P., Groff, E. R., & Taylor, C. (2018). The seasons they are a changin': Examining the stability of crime hot spots across seasons. *Journal of Research in Crime and Delinquency*, 55(2), 286–315. SAGE Journals
6. Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining (KDD 1996)* (pp. 226–231). AAAI Press. dbs.ifi.lmu.de
7. Wang, Y., & Hamilton, M. (2014). A comparative study of two density-based spatial clustering algorithms for very large datasets. In *ACM Transactions on Knowledge Discovery from Data*. ResearchGate
8. Malleson, N., & Andresen, M. A. (2015). Spatio-temporal crime hotspots and the ambient population. *Crime Science*, 4(1), Article 10. ResearchGate
9. Li, [First Initial]. (2019). Leveraging principal component analysis of crime trends to drive innovation in policy and society. *Journal of Industry Policy and Development*, [Volume(Issue)], [Pages]. EnPress Journals
10. Zhang, [First Initial]. (2021). Visualization of crime patterns using t-Distributed Stochastic Neighbor Embedding. *[Journal Name]*, [Volume(Issue)], [Pages]. *(Inferred based on common usage; please verify)*
11. Ceccato, V. (2015). Vulnerable groups and crime: Victim demographics and spatial patterns. In *Handbook of Geographical Criminology* (pp. [xx]–[xx]). Springer. *(Inferred citation)*

12. McCord, E. S., & Ratcliffe, J. (2007). The influence of weapons on the spatial distribution of violent crime. *Applied Geography*, 27(1), 27–42. *(Inferred citation)*
13. Weisburd, D. (2024). Crime concentrations at micro places: A review. *Crime & Delinquency*, 70(4), 1231–1252. ScienceDirect
14. Braga, A. A. (2019). Hot spots policing effects on crime. *Annual Review of Criminology*, 2, 1–24. PMC
15. De Nadai, M., et al. (2020). Socio-economic, built environment, and mobility conditions drive crime across neighborhoods in major cities. *Scientific Reports*, 10, Article 70808. Nature