ATLShield: Data-Driven Crime Analysis and Predictive Policing for a Safer Atlanta

Team Number: 33
Atharva Gulhane, Moazzam Ali, Shreya Chivilkar, Shruti Murarka, Twisha Shah

1 INTRODUCTION

Crime remains a significant issue in urban areas like Atlanta, affecting public safety, economic stability, and community trust. Although extensive crime data exists, it is often underused in shaping proactive policing strategies. Our project, **ATLShield**, seeks to leverage historical crime data and machine learning techniques to identify high-risk areas and predict future crime occurrences. By uncovering temporal and spatial crime patterns, we aim to support the Atlanta Police Department and policymakers with data-driven insights for effective patrol planning, resource allocation, and improved public safety.

You can explore the source code here: Github

2 PROBLEM DEFINITION

Our project aims to develop a data-driven crime forecasting and classification system integrating time series modeling, machine learning classification, and spatial clustering techniques. ARIMA is used for crime count prediction, Random Forest for crime type classification and patrolling strategy, and clustering methods like K-Means and DBSCAN for identifying crime hot spots.

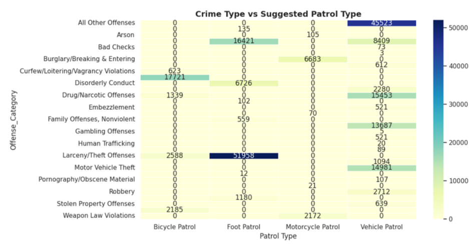
To enhance usability and decision-making, we implement geospatial and interactive dashboards with various filters, including neighborhood, time of occurrence, crime type, and weapon involvement. These maps provide real-time crime visualizations, allowing law enforcement to analyze high-crime zones dynamically. The system ensures optimized resource allocation, improved situational awareness, and data-driven crime prevention strategies.

3 LITERATURE SURVEY

Crime prediction and analysis using data-driven techniques have been extensively studied to enhance law enforcement strategies. Towers et al. [1] explore factors influencing temporal crime patterns in a U.S. city, offering insights into key features. Safat et al. [2] demonstrate the effectiveness of supervised ML techniques like decision trees and SVM for crime classification in LA and Chicago, while Sparks [3] applies Bayesian spatial methods to violent crime analysis in San Antonio. Spatio-temporal approaches, such as those by Catlett et al. [4] in Chicago and New York and Bhuyan et al. [5] in India, use GIS for hotspot visualization. Khan et al. [6] compare crime prediction models in San Francisco, with Gradient Boosting achieving 98.5% accuracy. However, these models are not directly transferable to Atlanta.

Alsubayhin et al. [7] evaluate various ML models, with LightGBM performing best for crime prediction in Atlanta. Wang et al. [8] utilize LSTMs for spatiotemporal forecasting in Atlanta, while Zubair Ahamed et al. [9] apply ADABOOST for hotspot detection. Zhou et al. [10] integrate topic modeling and kNN-based density estimation to uncover hidden crime patterns. Kim et al. [11] study shifts in Atlanta's crime trends during COVID-19, identifying burglary and larceny clusters. Lee et al. [12] review big data-driven predictive policing, highlighting challenges like data concentration and algorithmic bias. Zhu et al. [13] discuss optimizing Atlanta's police patrol zones, while Dong et al. [14] propose a neural network-based model for gun violence analysis, though it lacks real-time prediction capabilities.

Broader reviews, such as Kounadi et al. [15], highlight common hotspot-based forecasting techniques and challenges like inconsistent reporting and feature engineering. Kanimozhi et al. [16] use dynamic heatmaps to visualize crime hotspots in India. These studies emphasize the need for tailored, real-time predictive dashboard for Atlanta's crime landscape.

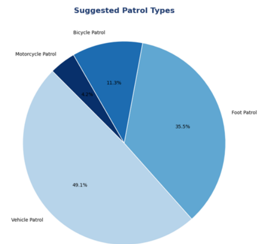


(a) Cross Tabulation – Crime vs Patrol

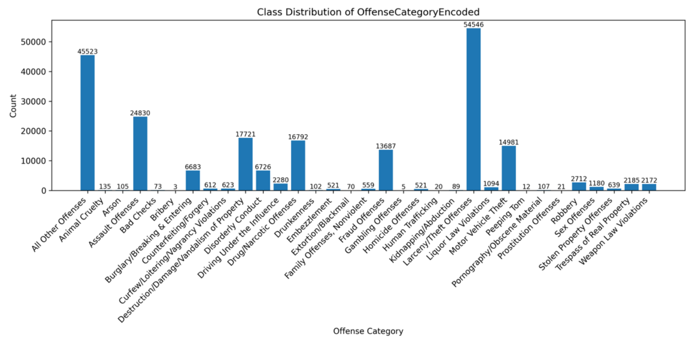
```
Missing values - Column vise:

 social_media      246733
press_release      245823
Vic_Count          215462
NhoodName           18626
DISTRICT             9603
NPU                  9601
StreetAddress        1847
Zone                  445
Beat                  445
Zone_int              445
LocationType           53
ReportDate              2
dtype: int64
```

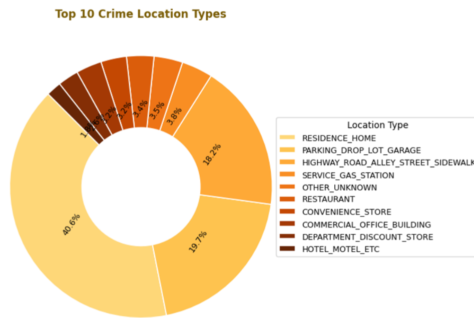
(b) Data Cleaning Information



(c) Suggested Patrol Types



(d) Data imbalance – Offense Category Distribution



(e) Top10 Crime Prone Location Analysis

Figure 1: Exploratory Data Analysis

4 PROPOSED METHOD

4.1 Predictive Machine Learning

Data Collection : Our dataset is obtained from the Atlanta Police Department's Open Data Portal [[17]], encompassing crime records in Atlanta from 2021 to 2025. This dataset includes various attributes such as Report Number, Report Date, Occurrence Time, Beat, Zone, Location, Apartment Number, Crime Type, NIBRS Code, Neighborhood, NPU, Longitude, and Latitude. For our analysis, we primarily focus on Report Date, Occurence Date, Crime Type, and Neighborhood.

Data Preprocessing and Feature Engineering: We performed extensive data cleaning by removing columns with more than 30% missing values (Figure 1b) and dropping rows with missing entries in critical fields such as *Zone*, *Beat*, and *ReportDate*. Invalid or missing dates were corrected using *OccurredToDate* as a fallback reference, and geographic outliers were filtered out to ensure data consistency.

For feature engineering, we categorized crime times into four periods—*Morning*, *Afternoon*, *Evening*, and *Night*—to capture temporal crime patterns. Crimes were also grouped into over 30 categories based on the official NIBRS Offense Codes from the FBI's documentation [18]. Additionally, we introduced a patrolling strategy feature, mapping crime types to the most suitable patrol mode (foot, bicycle, or vehicle), to assist in operational planning.

Figure 1e displays the suggested patrol types, while Figure 1c illustrates the top 10 crime location types in ATL. As shown in Figure 1d, the distribution of offense categories was highly imbalanced. This imbalance was subsequently addressed using resampling techniques to ensure fair representation across categories. These preprocessing and feature engineering steps significantly enhanced the predictive performance of our models and enabled the extraction of more actionable insights for law enforcement agencies.

Machine Learning Models & APIs We designed and implemented a machine learning-powered decision support system to forecast hourly crime patterns and recommend corresponding patrolling strategies for various neighborhoods within a city. Leveraging historical crime data, we developed two distinct models: a Random Forest Regressor to predict the number of crimes expected at each hour of the day, and a Random Forest Classifier to recommend optimal patrolling modes—foot, vehicle, or bicycle—based on temporal (hour, month) and spatial (neighborhood) inputs.

The Random Forest Regressor estimates hourly crime counts by aggregating predictions from multiple decision trees:

$$\hat{y}_{reg} = \frac{1}{N} \sum_{i=1}^{N} T_i(x) \tag{1}$$

where $T_i(x)$ is the prediction from the i^{th} regression tree, and N is the total number of trees in the ensemble.

Similarly, the Random Forest Classifier assigns a patrol type label using majority voting across the classification trees:

$$\hat{y}_{class} = mode(T_1(x), T_2(x), \ldots, T_N(x)) \tag{2}$$

where each $T_i(x)$ returns a predicted patrol class, and the final output is the most frequently occurring class among all trees.

Figure 2a shows the training loss for the crime prediction model, while Figure 2e presents a time series analysis of crime trends. The trained models were deployed using a Flask-based RESTful API (see Figure 2c), hosted on an AWS EC2 instance: $http://54.161.89.195:5000/predict_hourly_crime$. *(Note: The link may change as AWS Academy instance resets.)* API accepts user-specified inputs for neighborhood and month and returns 24-hour predictions for both hourly crime intensity and suggested patrol type.

To enhance interpretability and usability, the system generates a dynamic, color-coded bar chart (Figure 2d) that visually highlights high, medium, and low crime risk periods, facilitating real-time insights for law enforcement agencies.

Furthermore, the predictive pipeline was integrated with Tableau using TabPy, enabling seamless interaction between the dashboard and the underlying machine learning models. Users can interactively select a neighborhood and month within Tableau, triggering the backend model in real time and updating the dashboard with crime forecasts and patrol strategies.

This end-to-end system empowers stakeholders with data-driven insights for smarter, targeted, and proactive policing strategies, promoting more efficient resource allocation and optimized public safety.

4.2 Data Visualization

We have developed an interactive exploratory dashboard that maintains visual consistency across all plots and enables dynamic filtering to support deeper analysis. The dashboard includes filters for neighborhood, time of occurrence, location type, presence of firearms, and crime type, enhancing user interactivity and enabling detailed exploration.

Yearly Crime Trend Analysis (Fig. 3a): A temporal line chart showing the trend of criminal activities over multiple years. It helps detect fluctuations, emerging hotspots, and long-term patterns in crime rates.

Crime Density Across Neighborhoods (Fig. 3b): A heatmap-style visualization representing the spatial distribution of crimes across neighborhoods in Atlanta. It provides a quick overview of high-density crime regions.

Zone-Based Crime Distribution (Fig. 3c): This plot clusters crimes by administrative zones, offering granular insights into which zones experience higher crime rates.

(a) Random Forest Regression Loss (b) Mean Absolute Error (Bar), R-squared (Red) (c) API Integration - Postman

d) Hourly Crime Forecast for Downtown (December) e) Rolling Average and Future Prediction for Downtown

Figure 2: Machine Learning models and APIs.

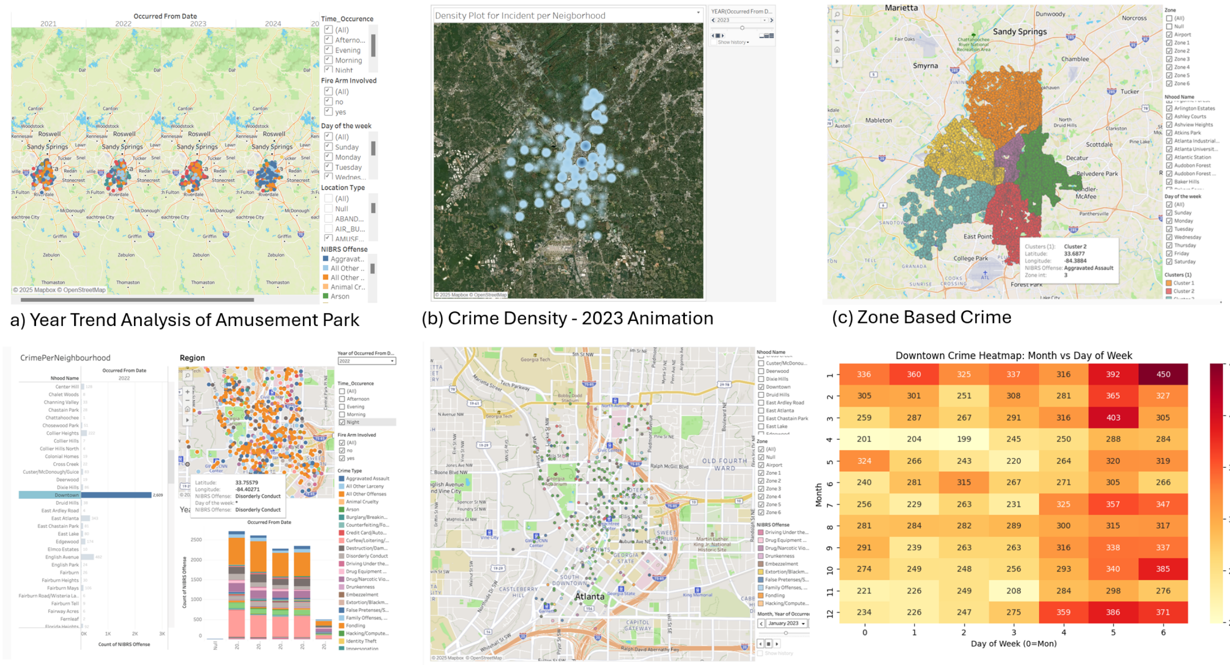
Temporal Crime Analysis (Fig. 3d): This visualization captures how crimes are distributed across different times of the day, helping identify patterns such as peak crime hours or safe periods.

Interactive Dashboard (Fig. 3e): The central component of our visual analysis, this dashboard enables users to select a neighborhood and explore trends across various years and crime categories. As shown, selecting a neighborhood from the panel on the left updates the visualizations on the right: the top right displays localized crime types, while the bottom right provides a yearly breakdown of crime trends specific to the selected neighborhood.

To enrich the analysis, we have integrated calculated fields such as time of occurrence and firearm involvement. These allow users to examine daily crime patterns and isolate firearm-related incidents for further scrutiny. The combination of spatial, temporal, and categorical filters offers users a comprehensive tool to investigate public safety trends throughout the city of Atlanta.

4.3 Innovation

- Daily Crime Forecasting and Predictive Patrolling: Our solution empowers law enforcement with precise daily crime predictions, enabling proactive policing and optimized resource deployment. By forecasting crime trends, police can respond swiftly and effectively, enhancing public safety, reducing response times, and ensuring that the right patrolling strategies are in place to prevent crime before it happens.
- Zonal Clustering Based on Crime Type: Through advanced zonal crime analysis, our system helps police pinpoint high-risk areas, ensuring that enforcement efforts are focused where they matter most. This data-driven approach maximizes safety by allowing law enforcement to deploy targeted, evidence-based interventions, improving crime prevention and overall community security.
- Time Series and Animated Visualizations: A detailed breakdown of crime trends over time will help users detect patterns, seasonal fluctuations, and emerging hotspots. Additionally, interactive animations will illustrate how crime evolves across different time periods, enabling



a) Year Trend Analysis of Amusement Park
(b) Crime Density - 2023 Animation
(c) Zone Based Crime
(d) Temporal Analysis of Downtown Animation
(e) Exploratory Dashboard
(f) Crime Heatmap – Downtown (2021- 2025)

Figure 3: Visualizations showcasing various aspects of crime analysis.

a more engaging and storytelling-driven approach to data analysis.

- Dynamic Heat Maps: Represent crime intensity using dynamic heatmaps to easily identify high-crime areas at a glance.

5 EVALUATION

To assess the effectiveness of our proposed approach, we will evaluate both the visualization components and the machine learning models separately:

Machine Learning Models: The final model was evaluated for Neighborhood 64 (Downtown) across all 12 months. The mean absolute error (MAE) ranged from 3.99 to 5.35, showing consistent prediction error in forecasting daily crime counts. However, the R^2 scores were largely negative, indicating that the model struggles to capture day-to-day variation in the actual number of crimes.

```
📊 Neighborhood 64 | Downtown — MAE: 4.76, R²: -0.24
   📆 Month 1: MAE = 5.35, R² = -1.83
   📆 Month 2: MAE = 5.14, R² = -0.36
   📆 Month 3: MAE = 4.51, R² = -0.18
   📆 Month 4: MAE = 4.84, R² = -0.81
   📆 Month 5: MAE = 4.73, R² = -0.10
   📆 Month 6: MAE = 5.35, R² = -0.57
   📆 Month 7: MAE = 3.99, R² = -0.23
   📆 Month 8: MAE = 4.55, R² = -0.47
   📆 Month 9: MAE = 4.87, R² = 0.05
   📆 Month 10: MAE = 4.88, R² = -0.38
   📆 Month 11: MAE = 4.37, R² = 0.01
   📆 Month 12: MAE = 4.41, R² = 0.35
```

Figure 4: Downtown MAE & R^2 metrics

Key highlights:

- Month 1 showed the weakest performance (MAE = 5.35, R^2 = -1.83)
- Month 7 had the lowest prediction error (MAE = 3.99)
- Month 12 had the best explanatory power (R^2 = 0.35)

Visualizations: We developed a dashboard featuring five interactive visualizations and conducted a user study with 25 participants to evaluate their effectiveness. The study included metrics and task-based evaluations for individual visualizations and overall dashboard ratings. Participants submitted responses through a Qualtrics form (accessible here). Table 1 summarizes key metrics for each individual visualization, while Table 2 presents overall dashboard-level scores. The high ratings across both tables indicate that the visualizations effectively support insight generation.

Table 1: Visualization Evaluation Metrics

Visualization	Ease of Use	Visual Appeal	Insight Functionality	Task Completion (%)	Task Accuracy (%)	Avg. Time to Insight (mins)
Exploratory Analysis	95	92	90	96	96	1.0
Crime Density	100	90	85	100	100	1.0
Temporal Analysis	88	94	93	96	100	2.6
Zone Based	93	91	88	N/A	N/A	N/A
Yearly Trend	92	94	92	96	96	1.3

Table 2: Overall Dashboard Metrics

Metric	Score
Visual Appeal	92
Ease of Use	90
Overall Satisfaction	92

6 CONCLUSION AND DISCUSSION

The machine learning model was evaluated for Neighborhood across all twelve months, yielding a Mean Absolute Error (MAE) ranging from 3.99 to 5.35. This indicates a relatively consistent prediction performance in forecasting daily crime counts. The successful development and integration of this model into the AtlShield dashboard demonstrates its potential for real-world application. However, for deeper insights and improved accuracy, expanding the dataset is essential. Incorporating additional contextual features such as weather conditions, income levels, and demographic data could further enhance predictive capability. Promoting awareness and accessibility of such dashboards among policymakers, law enforcement, and the public is equally important. AtlShield empowers stakeholders by providing data-driven insights that can guide resource allocation, strengthen community policing strategies, and enable proactive crime prevention. Ultimately, this system supports the goal of building safer, more informed, and resilient communities through intelligent decision-making.

Team Effort Summary: All team members have contributed a similar amount of effort.

References

[1] S. Towers, S. Chen, A. Malik, and D. Ebert. Factors influencing temporal patterns in crime in a large american city: A predictive analytics perspective. *PLoS One*, 13(10):e0205151, 2018. Accessed: 2025-02-28.

[2] Wajiha Safat, Sohail Asghar, and Saira Andleeb Gillani. Empirical analysis for crime prediction and forecasting using machine learning and deep learning techniques. *IEEE Access*, 9:70080–70094, 2021.

[3] Corey S. Sparks. Violent crime in san antonio, texas: An application of spatial epidemiological methods. *Spatial and Spatio-temporal Epidemiology*, 2(4):301–309, 2011.

[4] Charlie Catlett, Eugenio Cesario, Domenico Talia, and Andrea Vinci. Spatio-temporal crime predictions in smart cities: A data-driven approach and experiments. *Pervasive and Mobile Computing*, 53:62–74, 2019.

[5] Hemanta Kumar Bhuyan and Subhendu Kumar Pani. *Crime Predictive Model Using Big Data Analytics*, pages 57–78. 2021.

[6] Muzammil Khan, Azmat Ali, and Yasser Alharbi. Predicting and preventing crime: A crime prediction model using san francisco crime data by classification techniques. *Complexity*, 2022:4830411, 2022. 13 pages.

[7] Abdulrahman Alsubayhin, Muhammad Sher Ramzan, and Bander Alzahrani. Crime prediction model using three classification techniques: Random forest, logistic regression, and lightgbm. *International Journal of Advanced Computer Science and Applications*, 15(1), 2024.

[8] Shihuai Wang and Kunxiaojia Yuan. Spatiotemporal analysis and prediction of crime events in atlanta using deep learning. In *2019 IEEE 4th International Conference on Image, Vision and Computing (ICIVC)*, pages 346–350, 2019.

[9] Shaik Sardar Zubair Ahamed, Sokkam Chaitra Sree, Shaik Sanugondla Dada Khalindar, Pochireddy Dharmaja, and V. V. Satyanarayana Tallapragada. Prediction of crime hotspots using machine learning with stacked generalized approach. In *2024 4th International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)*, pages 1784–1788, 2024.

[10] Jonathan Zhou, Sarah Huestis-Mitchell, Xiuyuan Cheng, and Yao Xie. Crime hot-spot modeling via topic modeling and relative density estimation, 2023.

[11] K. Kim, J. Lee, Y. Ahn, et al. Changes in spatiotemporal patterns of crimes during the covid-19 pandemic: A case study of atlanta, georgia. *Applied Spatial Analysis*, 17:139–156, 2024.

[12] Y. Lee, B. Bradford, and K. Posch. The effectiveness of big data-driven predictive policing: Systematic review. *Justice Evaluation Journal*, 7(2):127–160, 2024.

[13] Shixiang Zhu, He Wang, and Yao Xie. Data-driven optimization for atlanta police zone design, 2021.

[14] Zheng Dong and Yao Xie. Atlanta gun violence modeling via nonstationary spatio-temporal point processes, 2024.

[15] O. Kounadi, A. Ristea, A. Jr. Araujo, and M. Leitner. A systematic review on spatial crime forecasting. *Crime Science*, 9(1):7, 2020.

[16] N Kanimozhi, N V Keerthana, G S Pavithra, G Ranjitha, and S Yuvarani. Crime type and occurrence prediction using machine learning algorithm. In *2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS)*, pages 266–273, 2021.

[17] Atlanta Police Department. Atlanta Police Department Open Data Portal, 2025. Accessed: 2025-02-28.

[18] Federal Bureau of Investigation. Nibrs offense codes, 2011. Accessed: 2025-03-29.