IT5006 Milestone 2 — Group 1

HE YIHAN (A0309390A) — JIN ZIQI (A0290561E) — LI XINYUE (A0291291A)
LI XU (A0304365J) — PAN XIAOYU (A0304475E)
https://github.com/Lemonziz/IT5006Project

---

0 Problem Statement

The predictive task is to forecast the monthly crime count by region. This will assist police departments in optimizing resource allocation and personnel deployment.

1 Literature Review

1.1 Summary of existing systems

Predictive policing systems, a statistical approach to help predict criminal activity and policing decision making [1], have been widely adopted in various regions to enhance law enforcement efficiency. Below are some notable systems and their effectiveness.
PredPol utilizes an ETAS(Epidemic-type aftershock) model integrated with machine learning algorithms, leveraging historical crime records as input features. The model's effectiveness is assessed based on the crime reduction rate and the Prediction Accuracy Index (PAI) [2, 3]. In experiments conducted in Los Angeles, PredPol resulted in a 7.4% decrease in crime [4]. Taking a different approach, RTMDx uses RTM model [5] and focuses on built-environment data, such as streetlight density and land use, in addition to crime records. This system used RRI (Recapture Rate Index) as its evaluation metric [6], and has led to a 35% reduction in gun violence in Newark and a 33% decrease in motor vehicle theft in Colorado Springs [7].
PreCobs [8] (Pre Crime Observation System) is based on Near Repeat Theory, focuses on historical burglary patterns and residential typology data to generates grid maps highlighting high-risk zones based on time, location of past incidents. The system reduced burglaries by 30%-40% in several European cities [9].
Unlike the previous systems, HunchLab [10] integrated Multivariate Regression Models, RTM(Risk Terrain Modeling) with machine learning algorithms. It incorporates a more diverse range of inputs compared with PredPol, including jurisdictional boundaries, crime records, geographical data (such as points of interest), and temporal factors like weather and holidays. It has shown a 31% reduction in property crime in Philadelphia and a notable decrease in violent crime in Chicago.
While all these systems rely on historical crime records as their primary input, they differ in other key aspects such as feature engineer techniques and effectiveness.

1.2 Review of Model methods

From the above systems, it can be seen that the currently used models include four main categories: Statistical, Spatial Analysis and Machine Learning Models.

Statical analyses are based on historical data and statistical methods, suitable for analyzing crime trends and hotspots, including regression analysis and KDE [**?**, 11] (Kernel Density Estimation). While spatial analysis models focus on the geographical distribution characteristics of crime, including RTM [12](Risk Terrain Modeling) and Near Repeat Theory. RTM [13] evaluates environmental factors contributing to crime risks by analyzing spatial correlations between locations and physical or social conditions. As Near Repeat Theory is based on criminological theories that crimes tend to cluster in time and space, these models predict repeat offenses near prior incidents. Machine learning model are suitable for big data environments, including random forest, SVM [14] and neural networks. They are used for classification tasks like predicting crime types or high-risk areas.

1.3 Feature engineer techniques

Feature engineering is crucial for machine learning and data analysis of crime data, as it contributes to modifying the data's features (including spatial, temporal etc.) to better represent the nature of the problem and improve the model accuracy [15], including calculation approach, representation and feature importance. For example, when engineering crime hotspots in urban environment, Borges et al. [16] deconstruct "street network" as dead-end density, major road length and so on, and decided to calculate major road length by summing up all roads of this type. For building types, they use binary representation to present if certain buildings are within the area. Finally, they adopt a random forest classifier to determine the importance of features.
In feature engineering, analyze and deconstruct the features, select proper representations and importance are necessary techniques to improve the model.

1.4 Evaluation metrics

To test and adjust the efficiency and efficacy of models, evaluation metrics are introduced to measure the abovementioned systems and models.
Predictive Accuracy Index (PAI) is commonly used in the field of criminal geography. It measures the concentration of crimes captured within predicted high-risk areas relative to their size. Based on it, Levine [17] proposed using RRI (Recapture Rate Index) together with PAI to measure accuracy.
Accuracy and precision are commonly used in classification tasks.
Accuracy represents the proportion of correctly predicted units among all units, which is greatly influenced by the crime level and spatial clustering of incidents in the study area. Precision refers to the proportion of units where crimes actually occurred among the predicted crime occurrence units.
Fairness and Transparency are also qualitative metrics, which are implied in the above systems. Fairness means to evaluate algorithmic bias to ensure equitable policing across different demographic groups. Transparency refers to assess whether predictions are explainable to law enforcement officers and the public.

2 Dataset and Preprocessing

In this data analysis project, we will use the open dataset from the Chicago Police Department's Citizen Law Enforcement Analysis and Reporting system, including crime data from 2001 to present (excluding the most recent 7 days) in Chicago. The dataset contains 22 columns, providing information about temporal aspects (e.g., date, year, updated_on), spatial aspects (e.g., block, location, ward, community area, beat), categorical features (e.g., FBI code, domestic, primary type), case handling (e.g., arrest, domestic), and detailed descriptions of the incidents.

2.1 Data Loading and Preprocessing

The analysis begins by loading the raw crime dataset from the `chicago_crime.csv` file, containing crime records from 2017 to 2024. After converting the date column to datetime format, the code filters complete time period data from December 2017 to 2024. During preprocessing, unnecessary fields like community areas and coordinates are removed, and missing values are checked. Notably, geographical data is obfuscated to protect privacy while maintaining block-level accuracy - a key characteristic of this dataset.

2.2 Temporal Feature Engineering

The code extracts multi-dimensional temporal features, including basic time dimensions like year, month, day, and hour. It further creates rich temporal derivatives:

1. Dividing hours into four periods: Night (0-6), Morning (6-12), Afternoon (12-18), and Evening (18-24)
2. Categorizing months into Winter (Dec-Feb), Spring (Mar-May), Summer (Jun-Aug), and Fall (Sep-Nov)
3. Flagging weekdays vs weekends
4. Identifying holidays using the U.S. federal holiday calendar

These temporal features help capture periodic patterns and seasonal variations in criminal behavior.

2.3 Spatial Feature Standardization

For crime location information, the code implements standardized processing of location descriptions. Free-text location descriptions in raw data are mapped to a 98-category standard classification system including residences, commercial areas, transportation hubs, etc. This standardization not only resolves inconsistencies in original location descriptions but also enables systematic spatial analysis. Each location is assigned a unique ID and standardized name, establishing the foundation for subsequent spatial pattern analysis.

2.4 Data Aggregation and Feature Derivation

The code aggregates raw event-level data to the "location-year-month" dimension, calculating various statistical features:

1. Absolute values: total crimes, arrest counts, domestic violence incidents
2. Ratio features: temporal distribution of crimes, arrest rates, domestic violence rates
3. Trigonometric transformations: sine/cosine of month for better seasonal pattern capture
4. Time ID creation for temporal sorting and calculation

This aggregation transforms raw event data into structured data suitable for time series analysis.

2.5 Feature Importance Analysis

Four complementary methods comprehensively evaluate feature importance:

1. Pearson correlation analysis measures linear relationships
2. Random Forest captures non-linear feature importance
3. F-regression assesses linear relationship strength
4. Lasso regression performs feature selection via regularization

Visual comparison of results from these methods forms a comprehensive feature importance evaluation. The analysis pays special attention to multicollinearity issues, identifying and handling highly correlated feature pairs to provide scientific basis for feature selection in subsequent modeling.

2.6 Time Series Feature Engineering

To prepare for time series modeling, the code creates crucial temporal features:

1. Monthly percentage change in crime count as prediction target
2. 1-12 month lags to capture temporal dependencies
3. 3-6 month moving averages to identify trends

These feature engineering steps enable models to learn temporal dynamics of crime patterns. After ensuring data completeness by handling missing values, the final output is a time series dataset rich in spatiotemporal features, providing ideal input for time series prediction models like RNN and LSTM.

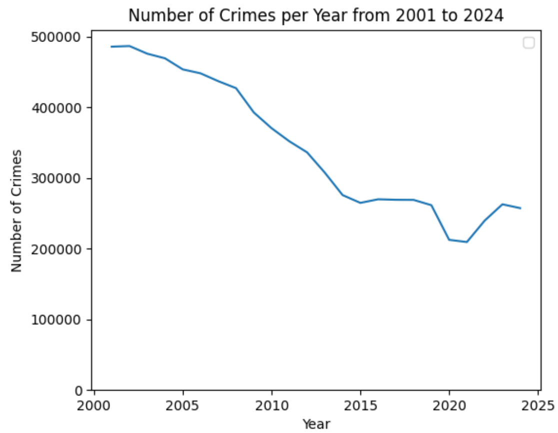


Figure 3.1: Crime count per year from 2001 to 2024.

3 Explotary Data Analysis

3.1 initial analysis

At the beginning of the analysis, we didn't notice that the data from 2015 showed a significantly different pattern compared to other years, so we didn't exclude the year and do the following analysis.

3.1.1 Temporal Distribution Study

As shown in Figure 3.1, from 2001 to 2015, the number of crimes per year decreases almost linearly. From 2015 to 2018, the number of crimes per year are relatively stable. From 2018 to 2024, the annual crime number varies a lot, decreasing rapidly and then increasing rapidly.

Figure 3.2(A) shows the total number of crimes between 2001 and 2024 in different months. February has the lowest crime count, and July has the highest crime count. The total crime count increases from February to July and decreases from July to December in general.

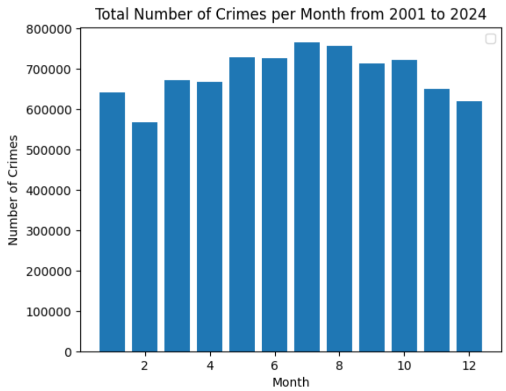
As displayed in Figure 3.2(B), in general, the crime count increases from February to August and decreases from August to December. The trend coincides with Figure 3.2(A), indicating that the trend of crime counts versus month is stable over time.

3.1.2 Spatial Distribution Study

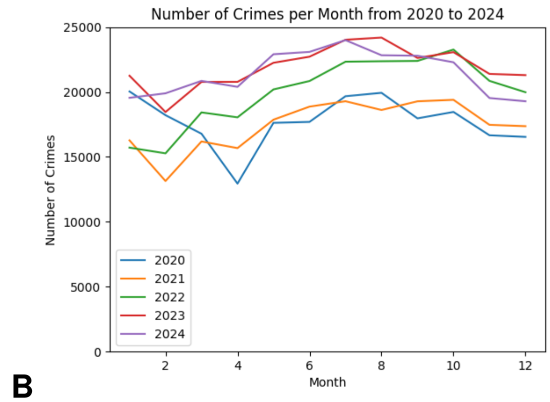
As demonstrated in Figure 3.3(A), between 2001 and 2024, the central and south central area of Chicago had the highest crimes. The results of spatial distribution of crimes of 2022, 2023 and 2024 in Figure 3.3(B), (C) and (D), also exhibit similar patterns. This indicates that the spatial distribution of crimes is relatively consistent over time.

3.1.3 Crime Correlation Analysis

Figure 3.4 reveals the relationship between various factors. The district of crimes and Beat have strong positive linear relationship. Ward has similar positive linear relationship with District and



(a) Total Number of Crimes per Month from 2001 to 2024.



(b) Number of Crimes per Month from 2020 to 2024.

Figure 3.2: (A) Count of all crimes between 2001 and 2024 in different months. (B) Crime count per month from 2020 to 2024.

Beat, attributed to the strong correlation relationship between District and Beat.

Furthermore, X Coordinate has fully linear relationship with Longitude, due to their same geographical meaning. The same for Y Coordinate and Latitude. In the following discussion, X Coordinate plays the same role with Longitude. So does for Y Coordinate and Latitude.

Community Area has relatively strong linear relationship with Latitude, and relatively weak linear relationship with Longitude. In addition, it has similar correlation relationship with Beat, District and Ward, explained by their close relationship.

Additionally, Longitude has similar linear relationship with Beat, District, Ward and Latitude, induced by their close relationship. Latitude also has similar linear relationship with Beat, District, and Ward, for the same interpration.

The rest of features almost have no linear relationship.

In summary, the results reveal the complex interaction between multiple features of the Chicago crime data.

3.2 Further analysis

After realizing that the data from 2015 was problematic, we decided to focus our analysis only on the data from 2018 onwards and base the analysis on the preprocessed data.

3.2.1 Distribution of crime types

Figure 3.5a illustrates the 15 top crime types and their frequencies. We can see that:

- Theft is the most prevalent crime by a significant margin (40,000 cases), dominating all other crime types.
- Violent crimes (e.g., assault, robbery) and property crimes (e.g., criminal damage, motor vehicle theft) are major categories, occupying the top five positions.

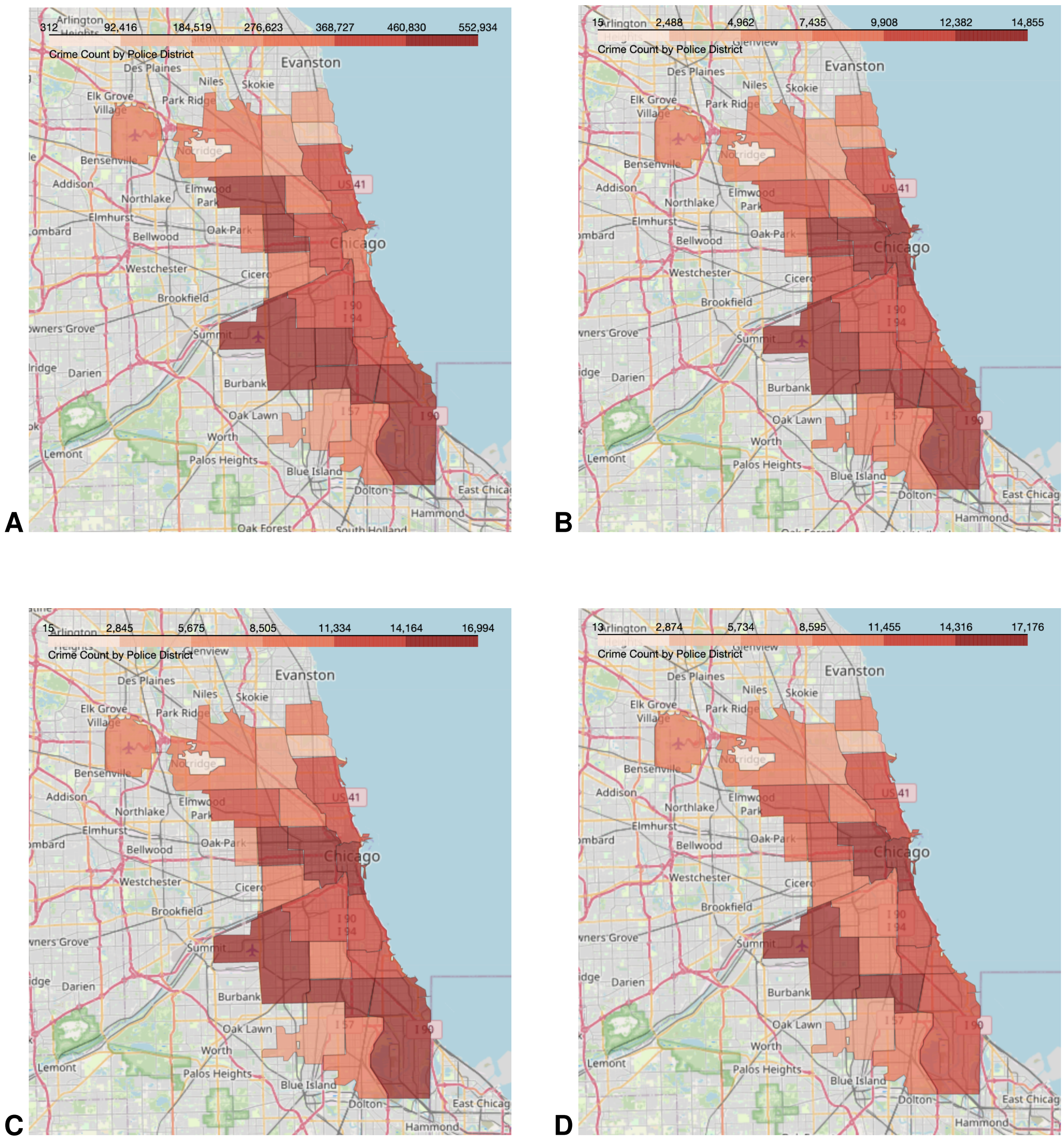


Figure 3.3: (A) Spatial distribution of all crimes between 2001 and 2024. (B) Crime spatial distribution in 2022. (C) Crime spatial distribution in 2023. (D) Crime spatial distribution in 2024.

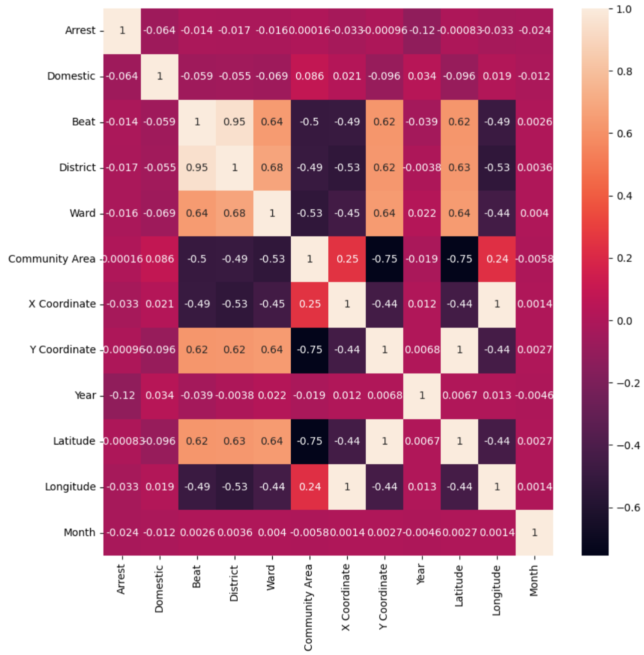


Figure 3.4: Correlation coefficients between various features.

- Sex-related crimes (sex offence, criminal sexual assault) and crimes involving children (offense involving children) have the lowest reported rates, which may reflect underreporting or data collection limitations.

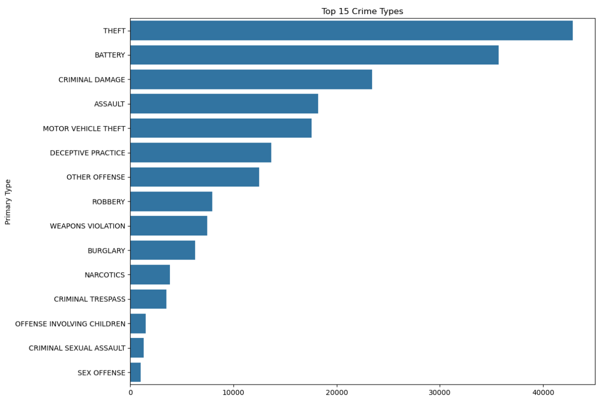
3.2.2 Time pattern analysis

Figure 3.5b and Figure 3.5c illustrates the comparative weekly and hourly distribution.

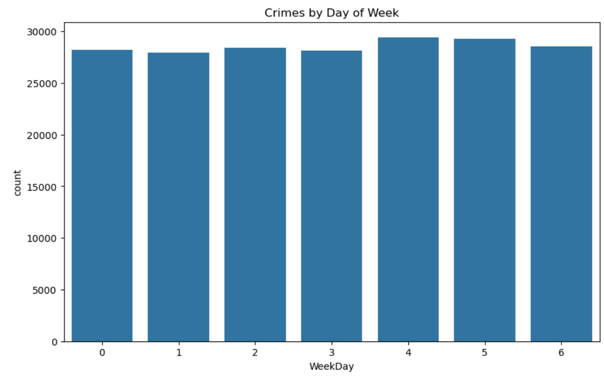
Based on the two charts, we can observe different temporal patterns in crime occurrences by day of the week and hour of the day. While crimes are evenly spread across different days, they are not evenly distributed by time of day. There is a sharp peak at midnight, a low point during early morning hours, and another sustained high level of activity during the afternoon to early evening.

- Crimes by Day of Week: The number of crimes remains relatively stable throughout the week, with only slight fluctuations. Weekdays 4 and 5 (typically Thursday and Friday) show a slightly higher count compared to other days, while days 1 and 3 (likely Monday and Wednesday) appear to have slightly fewer incidents. This suggests that criminal activity is distributed fairly evenly across the week, with only minor increases toward the end of the workweek.

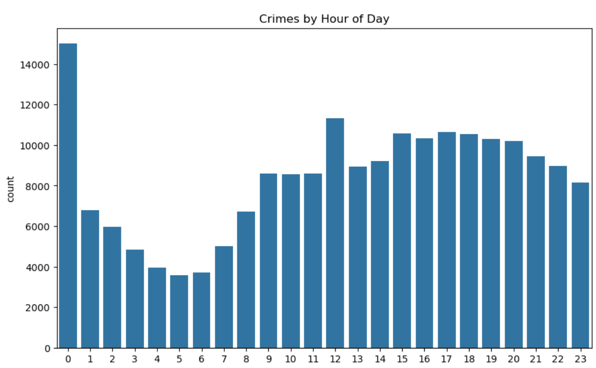
- Crimes by Hour of Day: The variation by hour is much more significant. There is a noticeable peak at midnight (hour 0), which is the highest among all hours. This may reflect crimes happening late at night or just after evening social activities. After midnight, the number of crimes drops sharply and remains low during the early morning hours (around 3 AM to 6



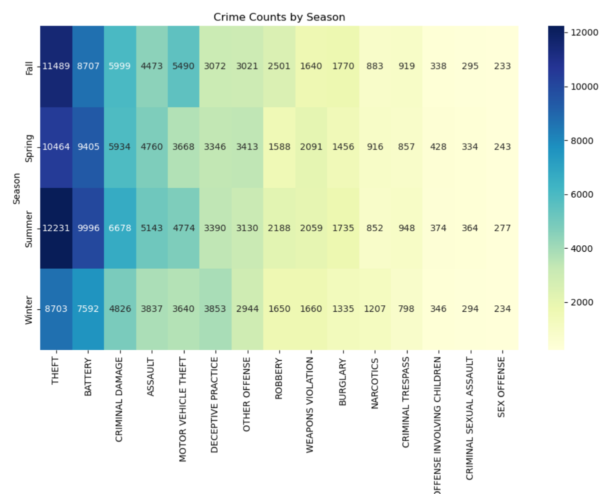
(a) 15 Top Crime Types



(b) Crimes By Day Of Week



(c) Crime By Hour Of Day



(d) Crime Counts By Season

Figure 3.5: Various Crime Analysis Figures.

AM). Crime counts begin to rise again from around 8 AM and remain consistently high from late morning (11 AM) through the afternoon and evening, with a second peak around 12 PM to 5 PM. This indicates that daytime and early evening hours are also active periods for criminal activity.

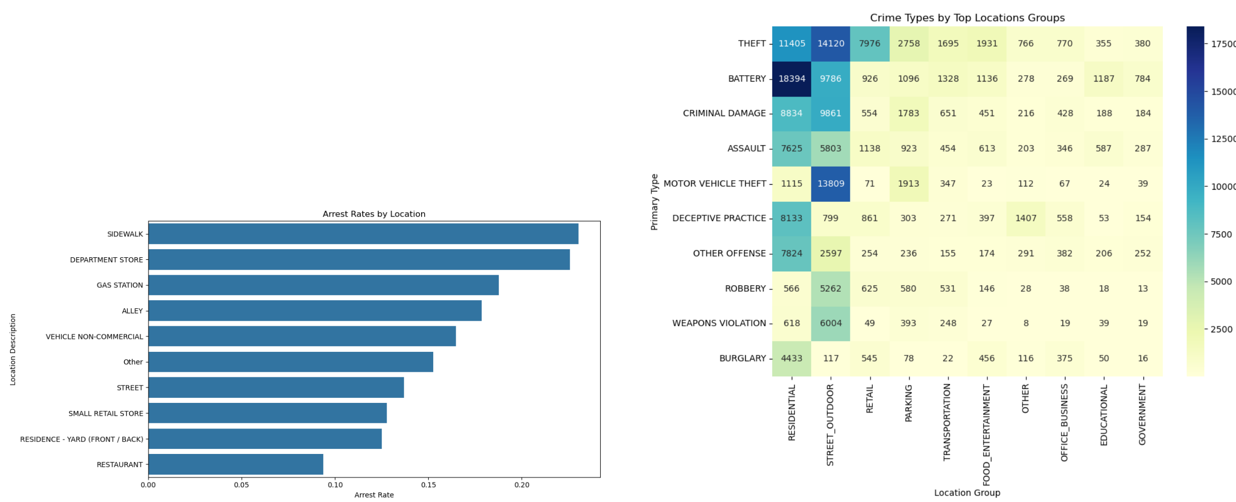
Besides,Figure 3.5d shows Seasonal Patterns:

- Summer Surge: Higher temperatures and increased outdoor activity may explain spikes in THEFT, BATTERY, and ASSAULT. Tourism/nightlife in summer could contribute (e.g., ROBBERY peaks at 2,188 in summer vs. 1,650 in winter).

- Winter Decline: Cold weather reduces outdoor presence, leading to fewer opportunities for theft/violence. MOTOR VEHICLE THEFT drops by 30% from summer to winter (5,143 vs. 3,837).

3.2.3 Relationship between crime and location

Figure 3.6a and figure 3.6b shows that Locations with the highest arrest rates (e.g., sidewalks, department stores) also exhibit the highest crime frequencies (e.g., theft, battery).



(a) Arrest Rates By Location

(b) Crime Types By Top Location

Figure 3.6: Comparison of Arrest Rates and Crime Types by Location.

Apart from that, we can see that:

- Street/Outdoor Areas (STREET, SIDEWALK, ALLEY) have highest arrest rates (Fig. 1) and peak crime frequencies for Violent crimes (BATTERY, ASSAULT), property crimes (THEFT, MOTOR VEHICLE THEFT). That may be caused by high foot traffic, surveillance blind spots, and elevated conflict potential.

- Retail/Commercial Areas's high arrest rates is linked to frequent theft (THEFT: 14,120 incidents),deceptive practices (DECEPTIVE PRACTICE: 8,133 incidents). This may be caused by Concentration of goods, cash transactions, and fraud opportunities.

4 Model Development

In this section, we implement multiple predictive models to forecast monthly crime counts in specific locations for 2024 based on historical data from 2018-2023. We develop four distinct models: *Linear Regression* (as our baseline), *XGBoost* [18], *LSTM* [19], and *CNN* [20].

4.1 Model Architectures

For our predictive modeling approaches, we implemented the following:

4.1.1 *Linear Regression*

We employed Linear Regression as our baseline model due to its interpretability and computational efficiency. The model utilized the default parameters from the scikit-learn implementation [21], using standard Ordinary Least Squares (OLS) without regularization. This approach provides a simple yet effective foundation against which to compare more complex models.

4.1.2 *XGBoost*

We applied *XGBoost* (Extreme Gradient Boosting), an implementation of gradient boosted decision trees known for its performance in structured data problems. Our `XGBRegressor` was configured with 200 estimators, a learning rate of 0.1, maximum depth of 5, minimum child weight of 7, gamma value of 0.1, and a fixed random state (42). These hyperparameters were selected to balance model complexity and generalization performance, with regularization parameters helping to prevent overfitting.

4.1.3 *LSTM* (Long Short-Term Memory)

For capturing complex temporal dynamics, we implemented a custom *LSTM* neural network architecture using PyTorch [22]. The network architecture consists of:

- An embedding layer (dimension 16) to encode categorical location features
- A stacked *LSTM* with 3 layers and hidden size of 128
- A fully connected layer reducing dimensions to 64 with ReLU activation
- A dropout layer (rate 0.5) for regularization
- A final output layer projecting to the target dimension

The model was trained using Adam [23] optimizer with a learning rate of 0.001 and batch size of 32 for 100 epochs. This architecture was specifically designed to capture sequential patterns and long-term dependencies in the crime data time series.

4.1.4 *CNN*

Our methodology employed convolutional neural networks *CNN* [20] for time series forecasting, leveraging a sliding window mechanism to capture localized temporal features. The hierarchical convolutional layers were designed to inherently detect short-term patterns within sequential data, thereby circumventing the computational overhead required for modeling long-range temporal dependencies characteristic of recurrent architectures.

Table 1: Structure of CNN

Layer Name	Input Shape	Output Shape	Parameters
fc1 (Linear)	(346, 91)	(346, 128)	11,776
ReLU	(346, 128)	(346, 128)	0
Dropout	(346, 128)	(346, 128)	0
fc2 (Linear)	(346, 128)	(346, 64)	8,256
ReLU	(346, 64)	(346, 64)	0
Dropout	(346, 64)	(346, 64)	0
fc3 (Linear)	(346, 64)	(346, 32)	2,080
ReLU	(346, 32)	(346, 32)	0
fc4 (Linear)	(346, 32)	(346, 1)	33

5 Model Evaluation

We conducted a comprehensive evaluation of our predictive models using a multi-metric approach to assess their efficacy in forecasting crime patterns across both spatial (location-based) and temporal dimensions.

5.1 Overall Spatial Performance Metrics

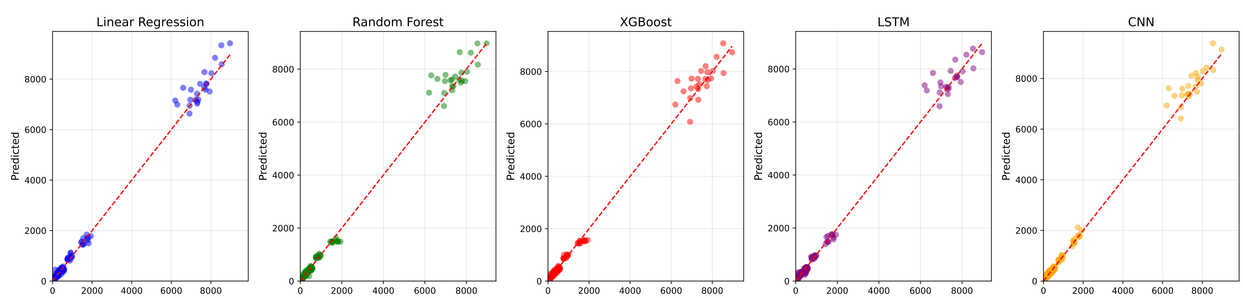


Figure 5.1: Scatter plots comparing predicted versus actual monthly crime counts across different location categories for all four models. The red dashed line represents perfect prediction (y=x).

Figure 5.1 presents a visual comparison of predicted versus actual values across all four models. The scatter plots reveal several important insights:

- The distribution of points reveals distinct clusters in the lower range (0-2000) and upper range (6000-9000), representing different location categories with varying crime volumes.

- The consistent alignment with the reference line across all models suggests that even the simpler Linear Regression approach effectively captures the primary patterns in location-specific crime prediction.

To evaluate the predictive accuracy of our models, we employ several complementary metrics. *Root Mean Square Error* (RMSE) [24] quantifies the standard deviation of prediction errors, with lower values indicating better performance. The *coefficient of determination* (R^2) [25] measures the proportion of variance in the dependent variable explained by the model, with values closer to 1 representing superior fit. *Mean Absolute Error* (MAE) [24] measures the average magnitude of errors without considering direction, providing an intuitive understanding of prediction accuracy in the same units as the target variable. Additionally, we calculate *Mean Absolute Percentage Error* (MAPE) [26], defined as $\text{MAPE} = \frac{\text{MAE}}{\bar{y}_{\text{test}}} \times 100\%$ where $\bar{y}_{\text{test}}$ is the mean of actual test values, to contextualize error magnitude relative to actual crime volumes, enabling more intuitive cross-model comparison.

Table 2: Overall Performance Comparison of Models for Location-Specific Crime Prediction

Model	**RMSE**	**R^2**	**MAE**	**MAPE**
Linear Regression	132.99	0.99	56.13	7.55%
XGBoost	139.71	0.99	50.88	6.84%
LSTM	137.48	0.99	47.70	6.41%
CNN	133.69	0.99	50.83	6.83%

Table 2 presents a comparative analysis of these performance metrics across all implemented models when predicting monthly crime counts across different location categories.

The location-specific performance analysis reveals several noteworthy findings:

- Despite comparable R^2 values (0.99) across all models, significant variations in error metrics highlight the nuanced differences in predictive capabilities.

- *LSTM* demonstrates superior performance in terms of Mean Absolute Error (MAE = 47.70) and MAPE (6.41%), suggesting its effectiveness in capturing location-specific crime patterns.

- *CNN* shows strong spatial crime prediction capabilities, achieving competitive error metrics (RMSE=133.69, MAE=50.83) comparable to top performers like Linear Regression. Its exceptional explanatory power (R^2=0.99) and low percentage error (MAPE=6.83) confirm effective pattern recognition in localized crime distribution, though temporal trend modeling requires architectural enhancements for holistic spatiotemporal analysis.

- *XGBoost* follows closely with an MAE of 50.88 (MAPE: 6.84%), outperforming Linear Regression models.

- Interestingly, while *LSTM* demonstrates the lowest MAE and MAPE values among all models, indicating superior performance in terms of absolute error metrics.

- The Linear Regression model, despite its simplicity, maintains competitive performance with an RMSE of 132.99, suggesting that certain crime patterns follow linear relationships with the predictive features.

Figure 5.2 presents a detailed visualization of actual versus predicted crime counts across the eight highest-crime locations in 2024 with *LSTM* and *XGBoost*. For clearer visualization, we include only these two methods, as they demonstrated superior performance with the lowest MAE and MAPE metrics as shown in Table 2. Notable findings include significant volatility in actual crime data across all locations, with both *XGBoost* and *LSTM* models generally capturing overall trends but struggling to predict extreme peaks and sudden drops. For instance, both models missed the sharp mid-year spike in Restaurant crimes and the severe October decline in Commercial/Office Building incidents. The visualizations demonstrate that public spaces experience higher crime volumes overall, while enclosed locations like Department/Discount Stores show more pronounced seasonal variations. While both predictive models perform reasonably well in tracking general crime patterns, they tend to smooth out the real-world volatility, with *LSTM* generally providing more consistent predictions compared to *XGBoost*'s occasionally higher variability.

5.2 Temporal Performance Analysis

Beyond spatial analysis, we evaluated model robustness across the temporal dimension by aggregating monthly crime counts across all locations to assess the models' capabilities in capturing broader temporal crime trends as shown in Table 3.

The temporal analysis yields contrasting insights compared to the location-specific evaluation:

- Linear Regression demonstrates superior performance in temporal prediction, achieving the lowest RMSE (959.44), highest R^2 (0.65), lowest MAE (832.50), and best MAPE (3.88%).

- CNN shows a relatively high RMSE of 1205.87 and the lowest R^2 value (0.45) among all models. Furthermore, the MAE of 995.16 and MAPE of 4.64 indicate that while the CNN model can capture some crime trends, it fails to provide accurate predictions for broader temporal patterns.

- The notable disparity in R^2 values between location-specific predictions (0.99) and aggregated monthly forecasts (0.50-0.65) suggests potential differences in the underlying predictability of these tasks. This phenomenon may reflect changes in variance structure, signal composition, or the relative influence of predictors across different levels of aggregation, rather than necessarily indicating increased prediction difficulty. Further analysis is warranted to determine the precise mechanisms driving this statistical divergence.

- The MAPE values remain relatively low (3.88%-4.30%), which translates to absolute MAE values of 832.50-922.92 against total monthly crime volumes of approximately 13,000-24,000. This level of prediction error—representing a deviation of less than 4.5% from actual values—demonstrates reasonable accuracy for operational planning purposes, while still leaving room for refinement in precise volume forecasting.

Figure 5.3 illustrates the comparative performance of all models in predicting aggregated monthly crime counts. During the training period (2019-2023), all models closely track actual

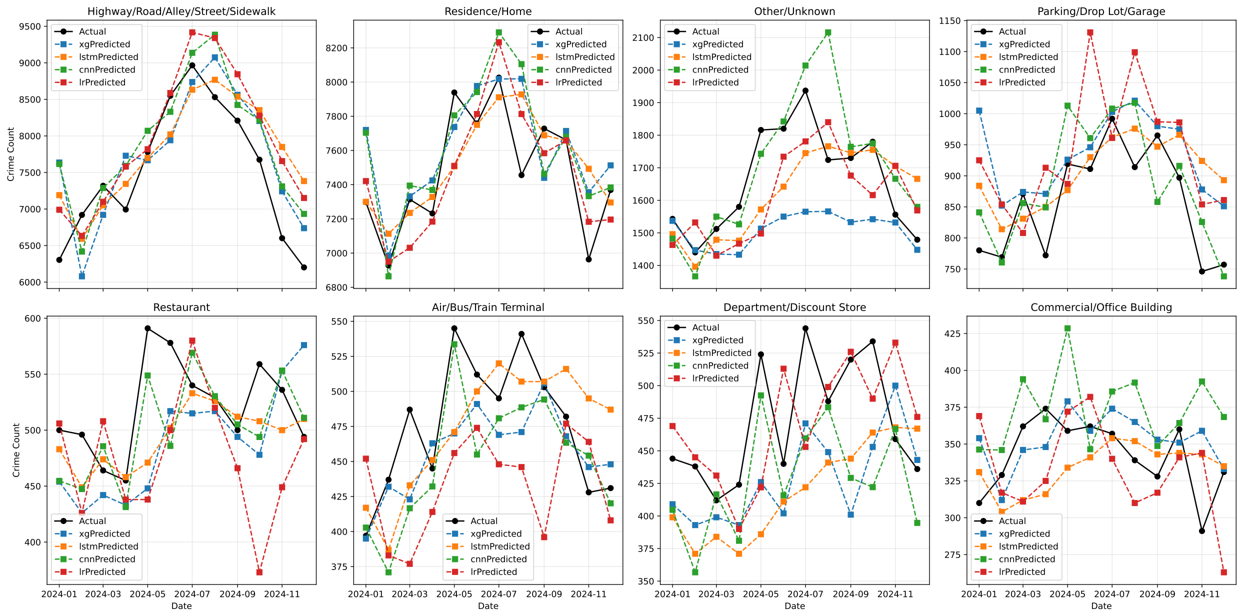
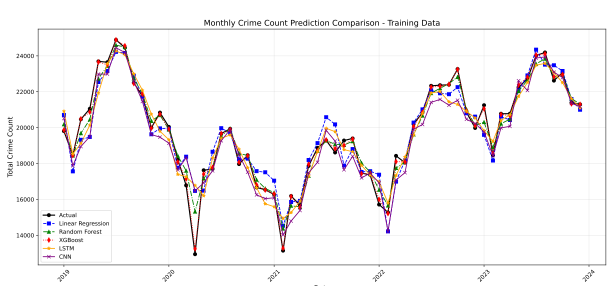


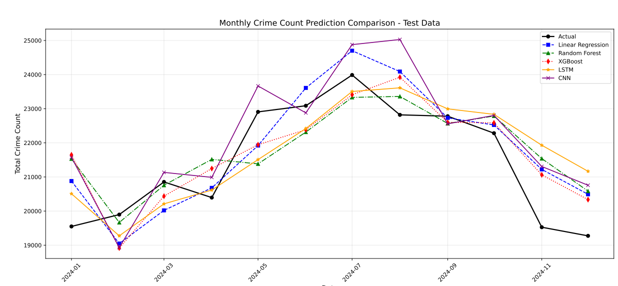
Figure 5.2: Monthly crime trends in 2024 for the eight locations with highest crime counts, comparing actual data with *XGBoost* and *LSTM* model predictions.

Table 3: Temporal Performance Metrics for Aggregated Monthly Crime Prediction

Model	**RMSE**	$\mathbf{R^2}$	**MAE**	**MAPE**
Linear Regression	959.44	0.65	832.50	3.88%
XGBoost	1034.53	0.59	897.75	4.19%
LSTM	1086.20	0.55	913.25	4.26%
CNN	1205.87	0.45	995.16	4.64%



(a) Monthly Crime Count Prediction on Training Data (2019-2023)



(b) Monthly Crime Count Prediction on Test Data (2024)

Figure 5.3: Comparison of model performance in predicting aggregated monthly crime counts across all locations during the training period (a) versus the test period (b).

crime counts, capturing seasonal patterns and fluctuations with high fidelity. However, the 2024 test period reveals notable challenges, with all models failing to capture the sharp decline in November-December. Linear Regression shows the most pronounced overestimation during mid-year, while *LSTM* better adapts to early-year fluctuations. This divergence between training and test performance highlights the models' limitations in identifying novel patterns not represented in historical data, explaining the lower R^2 values observed in temporal analysis.

5.3 Summary of Model Performance

Our evaluation reveals key insights for crime forecasting applications:

- *LSTM* excels in location-specific prediction (MAE: 47.70, MAPE: 6.41%), while Linear Regression performs best for temporal forecasting (RMSE: 959.44, R^2: 0.65)
- *CNN* is effective in capturing location-specific crime predictions, with an MAE of 50.83. However, its ability to capture broader temporal crime trends is weaker, as the R^2 value drops from 0.99 to 0.45. Improving the model's ability to predict temporal trends remains an area for further research.
- A substantial performance gap exists between spatial (R^2: 0.99) and temporal (R^2: 0.50-0.65) prediction tasks
- All models effectively capture general trends but fail to predict extreme events and sudden changes
- Practical implementation would benefit from a hybrid approach that selects models based on whether spatial precision or temporal accuracy is prioritized

5.4 Ethical Analysis

5.4.1 Privacy-Preserving Data Handling

To safeguard individual privacy while maintaining analytical utility, we implement the following protective measures:

Temporal Aggregation. All crime data is aggregated into monthly time windows as the minimum temporal unit. This prevents tracking of individual incidents or identifying patterns at sub-monthly resolution

Geographic Anonymization. Analysis uses location types (e.g., "Residence/Home", "Commercial Building") rather than specific addresses or street-level data. Enforces minimum area sizes by grouping all instances of the same venue category within a city district

5.4.2 Geographic Distribution Analysis

The test set is used instead of the training data to provide an objective evaluation of model performance on data not seen during training. And since the number of samples varies significantly across different locations, the geographical performance comparison is conducted using Mean Absolute Percentage Error (MAPE) instead of Mean Absolute Error (MAE) to account for scale differences and allow for fairer cross-location evaluation.

Figure 5.2 has shown the crime prediction trends in different kinds of locations. And along with table 4, we can see that:

On one hand, the four models (Linear Regression, XGBoost, LSTM, and CNN) are generally able to capture the overall trend of crime rates in areas such as transportation hubs (Air/Bus/Train Terminal), roadways (Highway/Road/Alley/Street/Sidewalk), restaurants, and residential areas (Residence/Home). While the magnitude of the prediction error varies, the overall patterns predicted by the models align reasonably well with the actual trends. This suggests that crime activities in these areas may follow stronger temporal or behavioral patterns, making them easier for the models to learn.

On the other hand, for locations like department/discount stores and parking lots/garages, the prediction accuracy is notably lower, with significantly higher mean absolute errors across all models. Several factors may contribute to this outcome:

- Crimes in these areas may be more sporadic and volatile, lacking clear temporal sequences or spatial regularities that models rely on.
- The data might be skewed or imbalanced, affecting the models' ability to generalize.
- External factors not captured by the models — such as holidays, promotions, or weather — may heavily influence criminal activity in these areas.

Table 4: Model Performance Across Locations (MAPE)

Location	**Linear Reg.**	**XGBoost**	**LSTM**	**CNN**
Air/Bus/Train Terminal	11.89	**5.55**	7.72	6.16
Commercial/Office Building	9.98	6.97	7.44	10.71
Department/Discount Store	9.13	10.50	12.21	**9.68**
Highway/Road/Alley	**7.39**	8.03	7.66	7.29
Parking Lot/Garage	12.21	9.22	8.48	**6.31**
Residence/Home	**2.29**	2.88	2.30	2.89
Restaurant	11.14	9.17	**6.04**	6.40

To be specific, in case of transport Hubs, XGBoost has the best prediction ability(5.55 MAPE). While in residential areas, we should use LSTM to predict(2.30 MAPE). And CNN shows advantage on roads crime prediction(7.29 MAPE).

However, despite the good performance in other kinds of location, the CNN model tends to over-predict crime counts in commercial buildings, with a MAPE of 10.71, while in department stores, the LSTM model underperforms, exhibiting a higher MAPE of 12.21.

5.4.3 Seasonal Performance by Model

Figure 5.4 and Table 5 show that during spring, all models exhibit relatively poor performance, with mean absolute errors (MAE) exceeding 1000. That may be caused by

- Unstable Human Activity: Spring includes holidays such as Spring Break and Easter, during which outdoor activities and social interactions increase. This leads to a seasonal rise in crime, but the high variability makes it difficult for models to capture consistent patterns.

- Dramatic Weather Transitions: The shift from cold to warm weather brings changes in crime types (e.g., theft, street altercations), making it harder for models to generalize.

In contrast, summer sees the best performance from CNN, with its MAE dropping to approximately 500, indicating improved predictive capability during this season. More regular behavior patterns may play the role. With stable weather and increased outdoor activity, crime trends become more spatially consistent and frequent, benefiting models like CNN that excel at extracting such localized patterns.

In autumn, the LSTM model proves to be the most stable, also maintaining an MAE close to 500. This may be caused by:

- Return to Routine: With schools back in session and fewer holidays, social rhythms stabilize, and crime patterns become more predictable, aligning well with LSTM's strength in modeling sequential data.

- Consistent Behavioral Trends: As people resume structured work and school routines, behavioral patterns are less volatile, making it easier for LSTM to model long-term dependencies.

Finally, in winter, the Linear Regression model achieves the lowest error among all models, with an MAE of 400, suggesting better alignment with seasonal crime trends during this period. This may be caused by:

- cold Weather Suppresses Certain Crimes. Outdoor crimes often decline during winter due to low temperatures, resulting in a more linear trend in crime rates that linear regression can model effectively.

- Lower Volatility in Data. Overall crime volumes tend to be lower and less volatile in winter, favoring simpler models like linear regression.

Table 5: Seasonal Performance by Model (MAE)

Season	**Linear Reg.**	**XGBoost**	**LSTM**	**CNN**
Spring	1000	1380.67	1200	**1492.48**
Summer	700	800	600	**500**
Autumn	550	650	**500**	700
Winter	**400**	600	1000	800

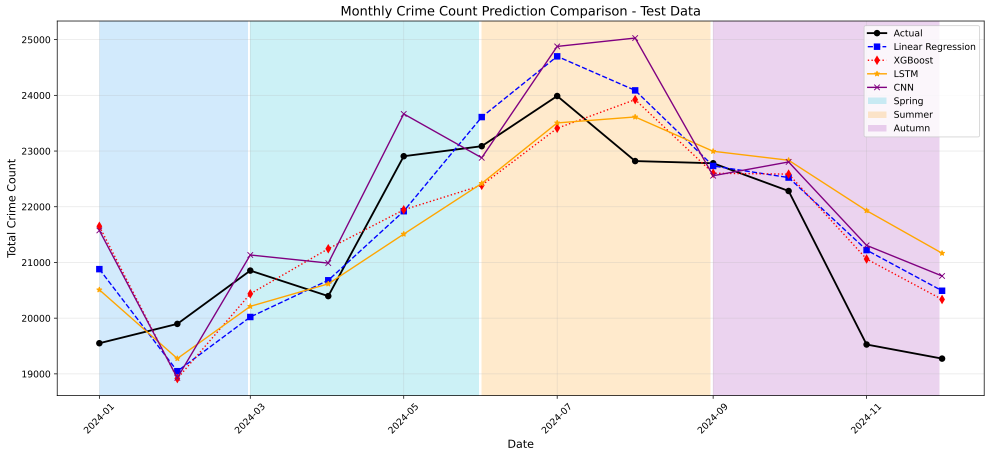


Figure 5.4: Seasonal Absolute Error by Model (MAE)

References

[1] A. Meijer and M. Wessels, "Predictive policing: Review of benefits and drawbacks," *International Journal of Public Administration*, vol. 42, no. 12, pp. 1031–1039, 2019.

[2] A. G. Ferguson, "Policing predictive policing," *Washington University Law Review*, vol. 94, no. 5, pp. 1115–1194, 2017.

[3] G. O. Mohler, M. B. Short, and S. M. et al., "Randomized controlled field trials of predictive policing," *Journal of the American Statistical Association*, vol. 110, no. 512, pp. 1399–1411, 2015.

[4] G. Turner, J. Brantingham, and G. Mohler, "Technology talk: Predictive policing in action in atlanta, georgia," *Police Chief Magazine*, Jul. 30 2022. [Online]. Available: https://www.policechiefmagazine.org/predictive-policing-in-action-in-atlanta-georgia/

[5] L. W. Kennedy, J. M. Caplan, and E. L. P. et al., "Vulnerability and exposure to crime: Applying risk terrain modeling to the study of assault in chicago," *Applied Spatial Analysis and Policy*, vol. 9, no. 4, pp. 529–548, 2016.

[6] J. M. Caplan, L. W. Kennedy, and J. D. B. et al., "Crime in context: Utilizing risk terrain modeling and conjunctive analysis of case configurations," *Journal of Contemporary Criminal Justice*, vol. 33, no. 2, pp. 133–151, 2017.

[7] A. G. Ferguson, "Predictive policing theory," in *The Cambridge Handbook of Policing in the United States*. Cambridge: Cambridge University Press, 2019, pp. 491–510.

[8] A. Tieu, "Precobs app predicts crimes before they happen," Jun. 10 2015. [Online]. Available: https://futurism.com/precobs-app-predicts-crimes-before-they-happen

[9] J. H. Ratcliffe, R. B. Taylor, and A. P. A. et al., "The philadelphia predictive policing experiment," *Journal of Experimental Criminology*, vol. 17, no. 1, pp. 15–41, 2021.

[10] Y. Hu, F. Wang, and C. G. et al., “A spatio-temporal kernel density estimation framework for predictive crime hot spot mapping,” *Applied Geography*, vol. 99, pp. 89–97, 2018.

[11] A. Malik, R. Maciejewski, and S. T. et al., “Proactive spatio-temporal resource allocation for community policing,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 20, no. 12, pp. 1863–1872, 2014.

[12] J. M. Caplan, L. W. Kennedy, and E. L. P. et al., “Using vulnerability and exposure to improve robbery prediction,” *Applied Spatial Analysis and Policy*, vol. 13, no. 1, pp. 113–136, 2020.

[13] A. Bilen and A. B. Ozer, “Cyber-attack method prediction using machine learning,” *PeerJ Computer Science*, vol. 7, p. e475, 2021.

[14] W. Hardyns and A. Rummens, “Predictive policing as a new tool for law enforcement?” *European Journal on Criminal Policy and Research*, vol. 24, no. 3, pp. 201–218, 2018.

[15] P. Duboue, *The Art of Feature Engineering: Essentials for Machine Learning*. Cambridge University Press, 2020.

[16] J. Becker and A. Becker, “Predictive accuracy index in evaluating dataset shift,” *Procedia Computer Science*, vol. 225, pp. 3342–3351, 2023.

[17] D. M. W. Powers, “Evaluation: From precision, recall to roc and correlation,” *arXiv*, 2020. [Online]. Available: https://arxiv.org/abs/2010.16061

[18] T. Chen and C. Guestrin, “Xgboost: A scalable tree boosting system,” in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD ’16. ACM, Aug. 2016, p. 785–794. [Online]. Available: http://dx.doi.org/10.1145/2939672.2939785

[19] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural Comput.*, vol. 9, no. 8, p. 1735–1780, Nov. 1997. [Online]. Available: https://doi.org/10.1162/neco.1997.9.8.1735

[20] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[21] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, “Scikit-learn: Machine learning in Python,” *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[22] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Köpf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, “Pytorch: An imperative style, high-performance deep learning library,” 2019. [Online]. Available: https://arxiv.org/abs/1912.01703

[23] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” 2017. [Online]. Available: https://arxiv.org/abs/1412.6980

[24] T. O. Hodson, "Root-mean-square error (rmse) or mean absolute error (mae): when to use them or not," *Geoscientific Model Development*, vol. 15, no. 14, pp. 5481–5487, 2022. [Online]. Available: https://gmd.copernicus.org/articles/15/5481/2022/

[25] N. Draper and H. Smith, *Applied regression analysis*, ser. Wiley series in probability and mathematical statistics. New York [u.a.]: Wiley, 1966. [Online]. Available: http://gso.gbv.de/DB=2.1/CMD?ACT=SRCHA&SRT=YOP&IKT=1016&TRM=ppn+022791892&sourceid=fbw_bibsonomy

[26] A. de Myttenaere, B. Golden, B. Le Grand, and F. Rossi, "Mean absolute percentage error for regression models," *Neurocomputing*, vol. 192, p. 38–48, Jun. 2016. [Online]. Available: http://dx.doi.org/10.1016/j.neucom.2015.12.114