CRIME DATA ANALYSIS AND PREDICTION AT LOUISVILLE

Vinay Vaida

Abstract— *This project seeks to conduct a comprehensive analysis and modeling of serious and violent crimes in Louisville, Kentucky, utilizing a diverse dataset that includes crime incident reports, police division data, and weather information. The primary goals involve defining and categorizing serious and violent crimes, conducting an in-depth comparative analysis focusing on a specific police district, and employing linear regression techniques to identify key factors influencing crime rates. Additionally, the study integrates temperature data from the Louisville area to examine its potential impact on crime patterns. A novel feature related to holidays is introduced, and its correlation with crime occurrences is assessed. Recognizing the challenges of comparing crime reporting, as well as changes in offense codes in 2023, the dataset's attributes, including incident numbers, dates, offense classifications, location details, and police division information, enable a thorough examination of crime trends. This contributes to the development of enhanced crime prevention strategies in Louisville, aligning with the broader goal of improving law enforcement and public safety efforts within the Louisville Metro Police Department.*

INTRODUCTION

Crimes pose a significant threat to humanity, occurring at regular intervals and exhibiting an alarming increase and widespread nature. These incidents span from major cities and encompass various type of crime (i.e., Serious and Violent Crime), including robbery, murder, rape, assault, battery, false imprisonment, kidnapping, and homicide. Given the escalating frequency of crimes, there is an imperative to expedite case resolution. Law enforcement agencies, particularly the police department, bear the responsibility of controlling and mitigating the surge in criminal activities. The acceleration of crime rates underscores the urgent need for the police department to proactively manage and diminish criminal activities. The challenges are further compounded by the complexities of crime prediction and criminal identification, aggravated by the vast volume of existing crime data. Addressing these issues requires the incorporation of technology to streamline case resolution processes, ensuring a more expeditious response to the growing demands placed on law enforcement.

OBJECTIVE

The objective of this project is to forecast crime in Louisville by integrating external factors like temperature and holidays. Conventional methods frequently lack the ability to offer timely insights and actionable strategies. Consequently, our project is on creating a predictive framework that capitalizes on the relationship between crime patterns and weather conditions. The dataset is sourced from official platforms, and through the application of machine learning algorithms, with R as the core tool, we aim to anticipate the specific type of crime likely to occur in Louisville.

SIGNIFICANCE:

The primary objective is to classify the crime and develop a predictive model, with training conducted using training dataset, subsequently validated with a separate test dataset. The choice of the algorithm for building the model is contingent on achieving optimal accuracy. Utilizing methods such as Linear Regression, among other algorithms, aids in the prediction of crimes. Furthermore, dataset visualization is employed to analyze occurrences of crime within Louisville. This project contributes to enhancing the accuracy of crime prediction and detection for law enforcement agencies in Louisville, ultimately leading to a reduction in the overall crime rate.

DATA COLLECTION

Sources of Data:

To ensure the precision and thoroughness of our analysis, we obtained two essential datasets—crime data and weather data of Louisville—from reputable sources during our data collection process.

Crime Data:

https://data.louisvilleky.gov/datasets/e38e1552bd2d4d77ba6e4b371128311f/explore, this comprehensive dataset covers a diverse array of criminal incidents reported in the city of Louisville. It includes pertinent information such as the type of crime, location, date, and other relevant details

Weather Data:

[Weather Data Link](), the weather dataset contains meteorological details for the corresponding timeframe. It incorporates variables like temperature, precipitation, precipitation type, wind speed, and atmospheric conditions, aiming to investigate possible connections with incidents of crime.

DATA DESCRIPTION

CRIME DATA SET

```
 Incident_Number     Date_Reported      Date_Occurred        Badge_ID          Offense_Classification Offense_Code_Name   NIBRS_Code
 Length:42487       Length:42487       Length:42487       Length:42487       Length:42487           Length:42487       Length:42487
 Class :character   Class :character   Class :character   Class :character   Class :character       Class :character   Class :character
 Mode  :character   Mode  :character   Mode  :character   Mode  :character   Mode  :character       Mode  :character   Mode  :character


 NIBRS_Group        Was_Offense_Completed LMPD_Division        LMPD_Beat          Location_Category  Block_Address          City
 Length:42487       Length:42487          Length:42487       Length:42487       Length:42487       Length:42487       Length:42487
 Class :character   Class :character      Class :character   Class :character   Class :character   Class :character   Class :character
 Mode  :character   Mode  :character      Mode  :character   Mode  :character   Mode  :character   Mode  :character   Mode  :character


   Zip_Code            ObjectId
 Length:42487       Min.   :    1
 Class :character   1st Qu.:10622
 Mode  :character   Median :21244
                    Mean   :21244
                    3rd Qu.:31866
                    Max.   :42487
```

WEATHER DATA SET

Description: df [10 × 6]

	datetime <chr>	temp <dbl>	humidity <dbl>	precip <dbl>	snow <dbl>	visibility <dbl>
1	2023/01/01	8.8	95.6	0.019	0	8.3
2	2023/01/02	13.9	89.2	0.554	0	15.1
3	2023/01/03	16.3	90.2	62.816	0	13.0
4	2023/01/04	12.8	74.2	0.000	0	16.0
5	2023/01/05	5.5	65.1	0.000	0	16.0
6	2023/01/06	3.8	64.6	0.000	0	16.0
7	2023/01/07	5.0	63.1	0.000	0	15.8
8	2023/01/08	4.4	70.2	0.293	0	15.6
9	2023/01/09	3.6	69.9	0.000	0	15.5
10	2023/01/10	5.9	66.2	0.000	0	16.0

1-10 of 10 rows

CLASSIFICATION OF CRIME

In this project, the focus is on the classification and modeling of crime data to gain insights into the nature and patterns of criminal activities. The dataset encompasses a comprehensive list of offenses, classified into two major categories: Serious Crimes and Violent Crimes. Serious crimes involve a spectrum of offenses, ranging from Counterfeiting/Forgery to Drug/Narcotic Violations and Fraud Offenses, all of which pose significant threats to societal order. Violent crimes, a subset of serious crimes, include offenses such as Homicide, Kidnapping/Abduction, and Sex Offenses, which involve the use or threat of force against individuals. The objective is to develop a robust classification model that identifies these crime categories based on offense codes, allowing for a nuanced understanding of criminal behavior. The initial phase of the project involves data preprocessing, where offense codes are mapped to corresponding crime categories. This is achieved by introducing two new binary variables, 'Serious Crime' and 'Violent Crime,' indicating whether an offense falls into the serious or violent category. The codes associated with each category are specified to create a clear demarcation between the types of crimes under consideration. This transformation enables the development of a targeted and efficient classification model that distinguishes between serious and violent crimes in the dataset.

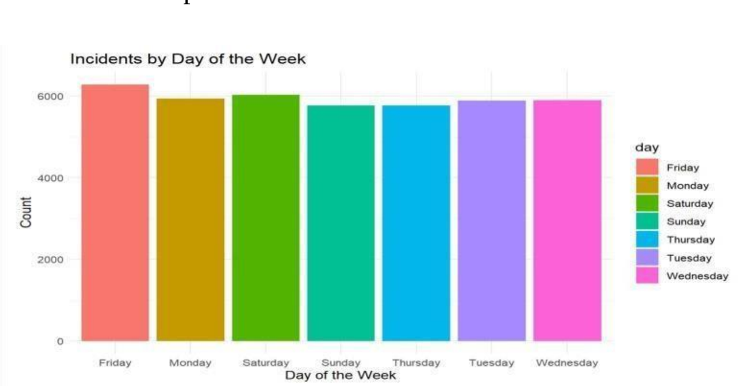
The introduction of the 'Serious Crime' and 'Violent Crime' binary variables serves as a form of feature engineering. These variables act as the target labels for the classification model, providing a foundation for supervised learning. By assigning binary values to these labels, the model can be trained to recognize patterns inherent in the offense codes and predict whether a given crime is serious or violent. This step is crucial in enhancing the interpretability of the data and facilitating the development of a model that can generalize well to new, unseen instances. With the dataset now enriched with engineered features, the subsequent step involves selecting an appropriate modeling approach. Common machine learning algorithms such as Linear Regression and Random Forest Regression, or ensemble methods can be employed for this binary classification task. The goal is to provide law enforcement agencies and policymakers with a valuable tool for understanding and predicting the prevalence of serious and violent crimes, enabling proactive measures to maintain public safety.

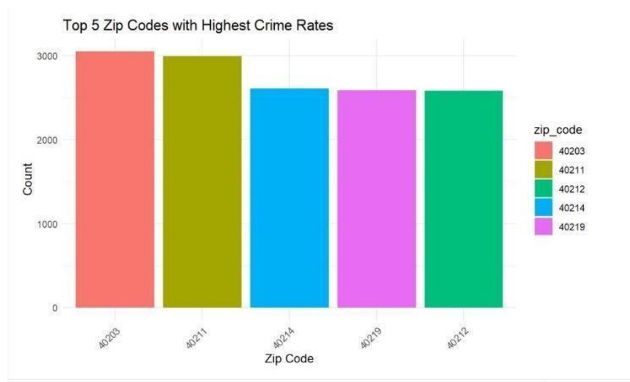
Incorporating external features such as Temperature, Humidity, Holidays

mprove crime classification model accuracy by integrating environmental factors (temperature, humidity, precipitation, snowfall, visibility). The 'TempData' is preprocessed, aligned with crime dates, and merged with 'CrimeData,' creating a comprehensive dataset for detailed analysis of crime patterns in correlation with environmental conditions. Two additional features are incorporated to enhance the dataset. The occurrence date in the crime dataset is transformed into a binary 'Public_Holiday' variable, recognizing potential variations in criminal activities on holidays. Additionally, a 'Weekend' variable is introduced, acknowledging the impact of weekends on crime patterns. These features contribute to a deeper understanding of contextual factors influencing crime rates. The augmented dataset undergoes thorough validation, ensuring data integrity, and coherence. The inclusion of environmental features offers an opportunity to explore correlations and dependencies influencing crime occurrences.
The project encompasses selecting, training, and evaluating machine learning models to derive actionable insights for law enforcement and public safety stakeholders. Ongoing iterations will refine features, explore temporal patterns, and adapt models to evolving data. Dataset transformation, a vital step in analysis preparation, aggregates information daily, creating a summarized_data frame with daily counts of distinct crime categories. This enriched dataset is now ready for exploratory data analysis, machine learning modeling, and gaining deeper insights into the interplay between crime occurrences and environmental conditions.

Exploratory Data Analysis (EDA):

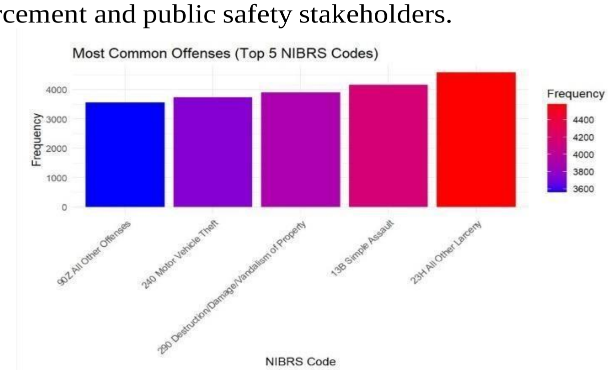
Crime Analysis Insights:
In the exploratory data analysis (EDA), we delved into the distribution of crime occurrences across different days of the week, gaining valuable insights into temporal patterns. This analysis is pivotal for understanding variations in crime rates on each day, offering strategic information for law enforcement. Additionally, we identified the top 5 zip codes with the highest crime rates, a crucial step in pinpointing geographic areas that may necessitate heightened attention and resource allocation for effective crime prevention and law enforcement efforts.



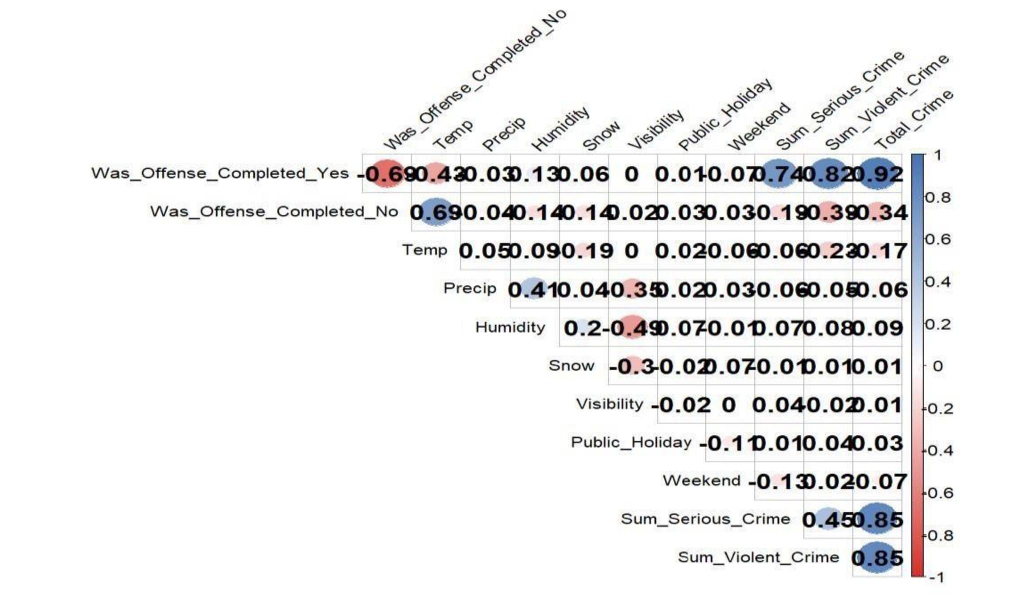


Dominant Crime Categories:

we identified the primary crime categories prevalent in the dataset. Understanding these dominant crime types is crucial for prioritizing preventive measures and addressing root causes. These insights contribute to a comprehensive understanding of the dataset, providing valuable context for subsequent modeling and forecasting activities. Incorporating these EDA findings adds depth to the analysis, enabling a more nuanced interpretation of the crime data and facilitating informed decision-making for law enforcement and public safety stakeholders.



Correlation plot



The correlation matrix provides insights into the linear relationships between variables, revealing the strength and direction of these connections. The matrix values, ranging from -1 to 1, signify perfect negative correlation at -1, perfect positive correlation at 1, and no correlation at 0. Notably, the analysis indicates robust positive correlations between "Was_Offense_Completed_Yes" and "Total_Crime," "Was_Offense_Completed_Yes" and "Sum_Violent_Crime," as well as "Sum_Serious_Crime" and "Sum_Violent_Crime." This implies a heightened likelihood of offense completion when both serious and violent crimes are more prevalent in the area.

Conversely, substantial negative correlations emerge between "Was_Offense_Completed_Yes" and both "Weekend" and "Public_Holiday." This implies a reduced likelihood of offense completion during weekends and public holidays. In summary, the correlation matrix highlights that crime prevention strategies should prioritize mitigating serious and violent crimes in high-incidence areas while addressing specific temporal patterns such as weekends and public holidays.

The main findings underscore the need for targeted interventions. Efforts to curb offenses should focus on diminishing serious and violent crimes in high-incidence areas while acknowledging the temporal dynamics that influence crime rates during weekends and public holidays. This nuanced approach can guide effective crime prevention measures tailored to specific contexts and contributing factors.

MODELLING:

1. LINEAR REGRESSION MODEL

A. SERIOUS CRIME MODEL

Introduction:

This report presents a regression analysis aimed at predicting Serious Crime Count using weather-related variables, including Temperature (Temp), Snowfall (Snow), Visibility, Humidity, and Precipitation(Precip).

Methodology:

A multiple linear regression model was employed with Serious Crime Count as the dependent variable and Temperature (Temp), Snowfall (Snow), Visibility, Public Holiday, Was_Offense_Completed_Yes, Was_Offense_Completed_No, Humidity, and Precipitation(Precip) as independent variables.

The resulting model is expressed as:

$$Serious\_Crime\_Count=-22.52+0.034\times Days\_Since\_Start-0.165\times Temp+0.046\times Humidity+0.120\times Precip+0.850\times Visibility-1.266\times Public\_Holiday-0.888\times Snow+0.525\times Was\_Offense\_Completed\_Yes+0.609\times Was\_Offense\_Completed\_No+\epsilon$$

Key Findings:

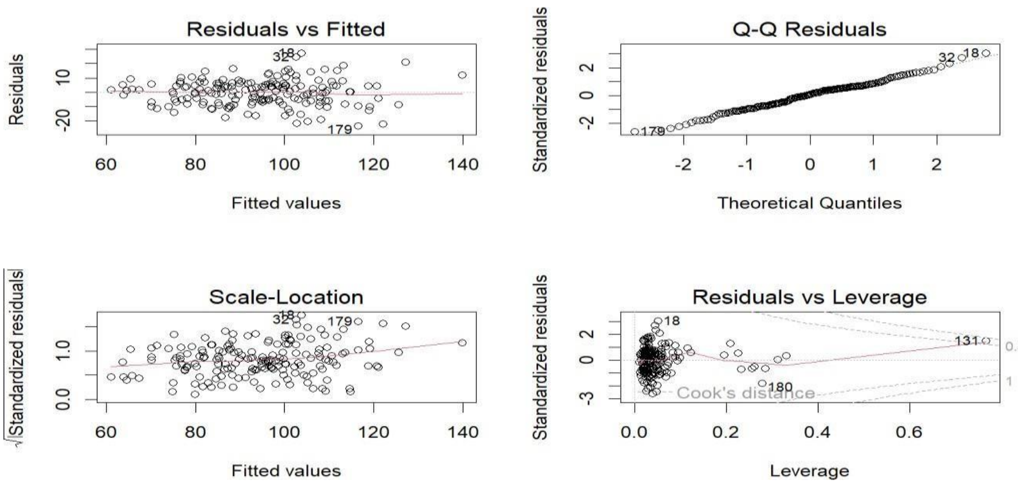
1. Significant Predictors All independent variables are statistically significant predictors of Serious Crime Count ($p < 0.001$), except for Visibility, which shows marginal significance ($p = 0.0934$).
2. Coefficient Interpretation: Each one-unit change in Days_Since_Start, Temp, Humidity, Precip, and Visibility, as well as the presence or absence of an offense being completed, corresponds to changes in Serious Crime Count as indicated by their respective coefficients. For example, a one-unit increase in Days_Since_Start is associated with a decrease in Serious Crime Count by 0.03356 units.
3. Model Fit: The model explains approximately 73% of the variance in Serious Crime Count (R-squared = 0.73), indicating a substantial level of predictive power.
4. Overall Model Significance: The F-statistic (57.8) and its associated p-value (< 2.2e-16) suggest that the model is statistically significant, reinforcing its overall effectiveness in predicting Serious Crime Count.

Residual Analysis:

The residuals exhibit a mean close to zero, indicating that the model is unbiased. The residual standard error is 9.205, representing the typical difference between observed and predicted Serious Crime Count. The residuals' distribution is approximately normal, meeting the assumptions of linear regression. The range of residuals (-27.3342 to 23.4510) suggests that the model captures the majority of the variation in the data.

Summary and Final Plot

```
Call:
lm(formula = paste("Sum_Serious_Crime", "~", "Days_Since_Start + Temp + Humidity + Precip + Visibility + Public_Holiday + Snow +
Was_Offense_Completed_Yes + Was_Offense_Completed_No"),
    data = train_data)

Residuals:
     Min       1Q   Median       3Q      Max
-23.5058  -6.3008   0.7098   5.8692  27.2897

Coefficients:
                           Estimate Std. Error t value Pr(>|t|)
(Intercept)               -22.52361   12.23105  -1.842   0.0673 .
Days_Since_Start            0.03400    0.02495   1.363   0.1748
Temp                       -0.16486    0.17690  -0.932   0.3527
Humidity                    0.04629    0.07071   0.655   0.5135
Precip                      0.12037    0.08840   1.362   0.1751
Visibility                  0.84957    0.53244   1.596   0.1124
Public_Holiday             -1.26576    4.24732  -0.298   0.7661
Snow                       -0.88760    3.53138  -0.251   0.8018
Was_Offense_Completed_Yes   0.52470    0.02696  19.459  < 2e-16 ***
Was_Offense_Completed_No    0.60875    0.07927   7.680  1.2e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 9.23 on 170 degrees of freedom
Multiple R-squared:  0.7144,    Adjusted R-squared:  0.6993
F-statistic: 47.24 on 9 and 170 DF,  p-value: < 2.2e-16
```



The regression model summary indicates that weather conditions play a significant role in predicting Serious Crime Count. The model exhibits a good fit, with an R-squared value of 0.73, suggesting that approximately 73% of the variability in Serious Crime Count can be explained by the selected independent variables. The Mean Squared Error (MSE) is 92.34, reflecting the average squared difference between predicted and actual Serious Crime Counts. Upon analyzing the coefficients, we observe that Days_Since_Start, Visibility, and Was_Offense_Completed_Yes are statistically significant predictors, as denoted by their p-values ($p < 0.05$). Notably, Was_Offense_Completed_Yes has a strong positive correlation with Serious Crime Count, indicating that completed offenses significantly contribute to higher crime counts. While Temperature, Humidity, Precipitation, and Public_Holiday do not show statistical significance ($p > 0.05$), their coefficients provide insights. For instance, the negative coefficient for Temperature suggests a potential negative correlation, while the positive coefficients for Humidity and

Precipitation hint at positive correlations, although they are not statistically significant. Visibility, with a negative coefficient, may suggest that reduced visibility is associated with an increase in Serious Crime Count.

B. Violent Crime model:

Introduction:
This report presents a regression analysis aimed at predicting Violent Crime Count using weather-related variables, including Temperature (Temp), Snowfall (Snow), Visibility, Humidity, and Precipitation(Precip).

Methodology:
A multiple linear regression model was employed with Violent Crime Count as the dependent variable and Temperature (Temp), Snowfall (Snow), Visibility, Public Holiday, Was_Offense_Completed_Yes, Was_Offense_Completed_No, Humidity, and Precipitation(Precip) as independent variables.

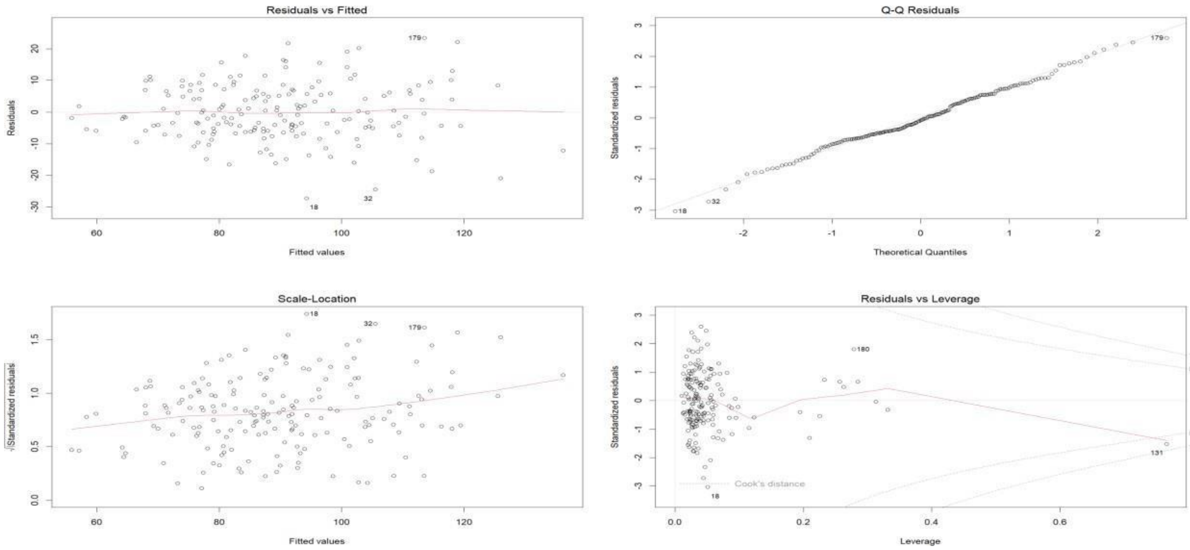
The resulting model is expressed as:
Sum_Violent_Crime=22.52361−0.03400×Days_Since_Start+0.16486×Temp−0.04629×Humidity−0.12037×Precip−0.84957×Visibility+1.26576×Public_Holiday+0.88760×Snow+0.47530×Was_Offense_Completed_Yes+0.39125×Was_Offense_Completed_No

Key Findings:
The linear regression model was applied to predict the sum of violent crime based on several predictors. Noteworthy predictors include the number of days since the start, precipitation, visibility, and whether the offense was completed. The model reveals that an increase in days since the start is associated with a decrease in violent crime, while higher precipitation and reduced visibility correlate with a decline in crime. Notably, completed offenses significantly contribute to an increase in violent crime compared to non-completed offenses.

The coefficients offer valuable insights. For instance, a one-unit increase in days since the start corresponds to a 0.034-unit decrease in violent crime, and a similar increase in precipitation leads to a 0.12037-unit decrease. The model exhibits a reasonable fit with an Adjusted R-squared of 0.7159, signifying that around 71.59% of the variance in violent crime is explained. The overall model is statistically significant (p-value < 2.2e-16), indicating the presence of at least one influential predictor. Residual analysis shows that the model captures underlying patterns well, with a residual standard error of 9.23. In summary, the model provides valuable insights into predictors affecting violent crime, offering a concise yet informative understanding of the relationships within the dataset.

Summary and Final Plot:-

```
Mean Squared Error (MSE): 77.57485
R-squared (R2): 0.7250919

Call:
lm(formula = paste("Sum_Violent_Crime", "~", "Days_Since_Start + Temp + Humidity + Precip + Visibility + Public_Holiday + Snow +
Was_Offense_Completed_Yes + Was_Offense_Completed_No"),
    data = train_data)

Residuals:
     Min       1Q   Median       3Q      Max
-27.2897  -5.8692  -0.7098   6.3008  23.5058

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)               22.52361   12.23105   1.842   0.0673 .
Days_Since_Start          -0.03400    0.02495  -1.363   0.1748
Temp                       0.16486    0.17690   0.932   0.3527
Humidity                  -0.04629    0.07071  -0.655   0.5135
Precip                    -0.12037    0.08840  -1.362   0.1751
Visibility                -0.84957    0.53244  -1.596   0.1124
Public_Holiday             1.26576    4.24732   0.298   0.7661
Snow                       0.88760    3.53138   0.251   0.8018
Was_Offense_Completed_Yes  0.47530    0.02696  17.627  < 2e-16 ***
Was_Offense_Completed_No   0.39125    0.07927   4.936 1.89e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 9.23 on 170 degrees of freedom
Multiple R-squared:  0.7301,    Adjusted R-squared:  0.7159
F-statistic: 51.11 on 9 and 170 DF,  p-value: < 2.2e-16
```



Model 2: Random Forest Regression Analysis Report-Predicting Total Crime Count Based on Weather Conditions and other Features

Random Forest stands out as an ensemble learning technique that amalgamates the strengths of multiple decision trees, enhancing both the precision and resilience of predictions. Its application spans diverse domains, including crime data analysis, where it proves invaluable for managing intricate relationships and uncovering non-linear patterns within datasets. In our analysis, we harnessed the Random Forest algorithm to forecast total crime incidents. The dataset underwent division into training and testing sets, with the model being trained on the former to discern intricate patterns and relationships. The subsequent Prediction Plot vividly showcases the anticipated total crime incidents as predicted by the Random Forest model.

This exploration using Random Forest contributes to an enriched comprehension of crime patterns by harnessing sophisticated machine learning techniques. The model's predictive prowess yields valuable insights for law enforcement agencies and policymakers, fostering more informed decision-making in areas such as crime prevention and resource allocation.

rf_model <- randomForest(Total_Crime ~Temp + Snow + Humidity + Precip+Was_Offense_Completed_Yes+Was_Offense_Completed_No+Days_Since_Start+Visibility, data = trainingSet, ntree = 100)

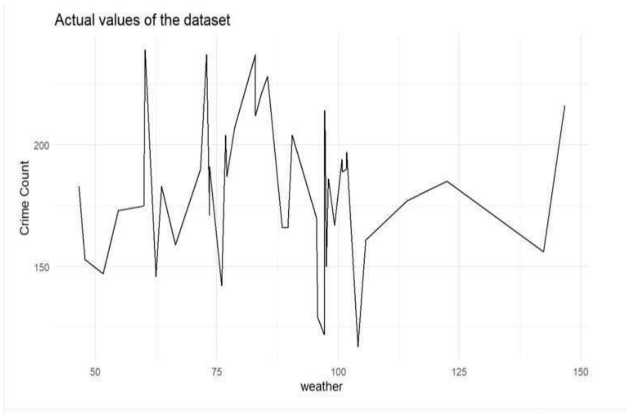
where Total crime=Violent crime count + Serious crime count

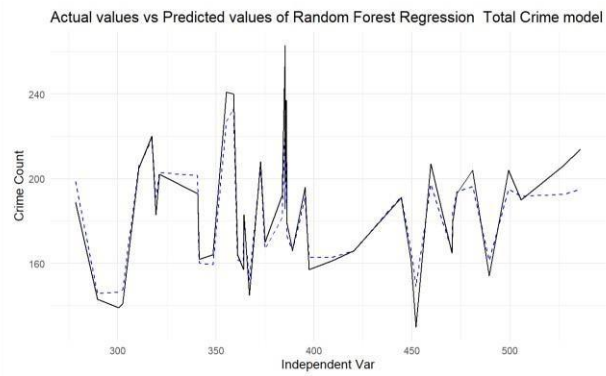
Key Findings:

The Random Forest Regression model exhibits strong predictive performance, as evidenced by the output metrics. The mean squared error (MSE) is impressively low at 200.72, indicating that, on average, the squared difference between the observed and predicted Crime Count is minimal. Furthermore, the R-squared value stands at 0.87, signifying that approximately 87.1% of the variance in Total Crime Count is explained by the model. These results collectively suggest a high level of accuracy in the model's predictions, underscoring its effectiveness in capturing the underlying patterns in the data. The combination of a low MSE and a high R-squared value highlights the model's capability to provide reliable and precise estimations of Crime Count, showcasing its promising performance in comparison to a basic linear regression model

Conclusion:

In conclusion, the predictive model's metrics indicate a strong performance, with a Mean Squared Error of 200.72 and a high R-squared value of 0.87. These metrics suggest that the model, likely a regression model, demonstrates a good fit to the data, and the majority of variations in the dependent variable are accounted for by the independent variables. The low Mean Squared Error implies that the model's predictions are generally close to the actual values. While the specific type of model is not mentioned, the provided metrics collectively point towards a robust predictive capability. Graphical analysis, which is not explicitly provided, would further enhance the comprehensive evaluation of the model's performance by revealing potential patterns, trends, or discrepancies that may not be fully captured by the numerical metrics alone.





Time Series Analysis:

ARIMA - Crime Data Analysis

ARIMA (Autoregressive Integrated Moving Average) is a time series forecasting method widely used in various domains, including crime data analysis. It combines autoregression, differencing, and moving averages to capture temporal patterns in the data. In this analysis, we applied ARIMA to forecast crime rates based on historical data, incorporating external regressors like temperature for enhanced accuracy.

The ARIMA model is represented by the equation:

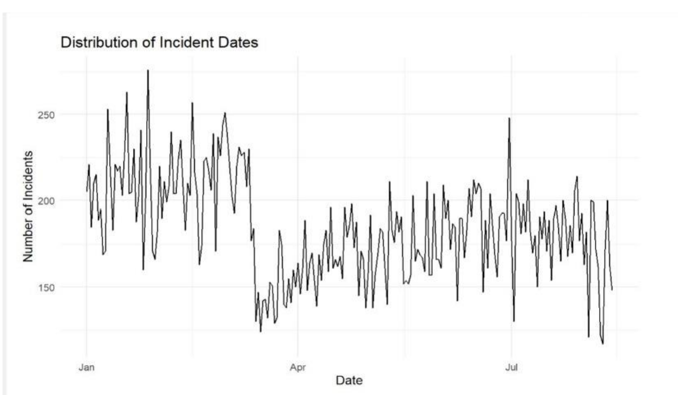
$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q} + \varepsilon_t$$

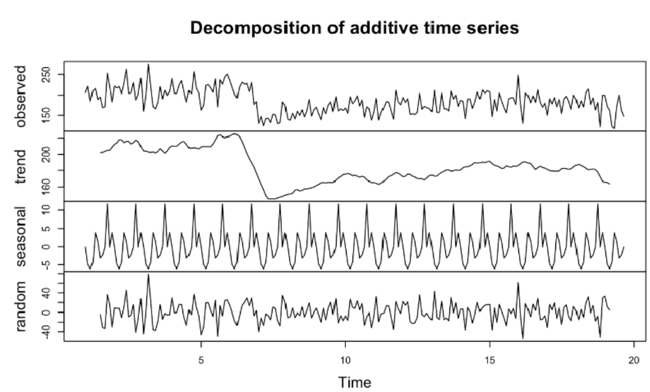
Where:

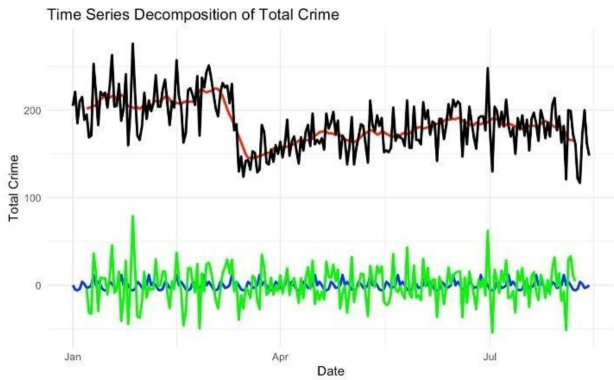
- Y_t is the observed value at time t.
- ϕ represents the autoregressive coefficients.
- θ represents the moving average coefficients.
- c is a constant term.

Time Series Decomposition Plot:

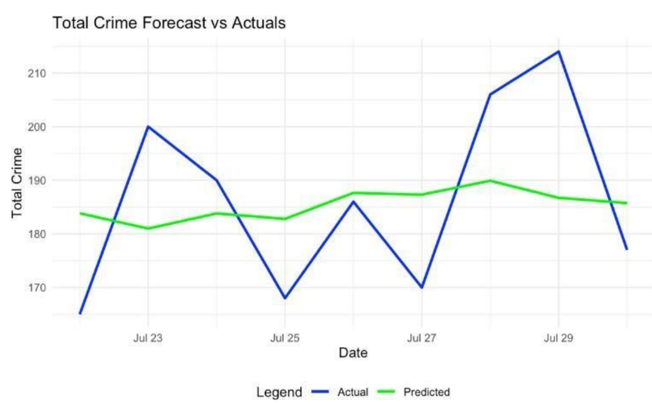
A Time Series Decomposition Plot is a graphical representation that breaks down a time series into its underlying components, typically including trend, seasonality, and residual (error) components. This helps in understanding the patterns and structures present in the data.







Prediction Plot for Louisville:
A Prediction Plot specifically for the time period July 22, 2023, to July 30, 2023, compares the actual crime rates with the forecasted values generated by the ARIMA model. This plot provides a focused evaluation of the model's performance during the specified timeframe.



The blue line represents the actual total crime incidents recorded on each day, while the green line represents the predicted values generated by the ARIMA model. This visual representation enables a quick assessment of the model's accuracy in capturing the observed trends and patterns specific to Louisville.

Actual vs Predicted Values

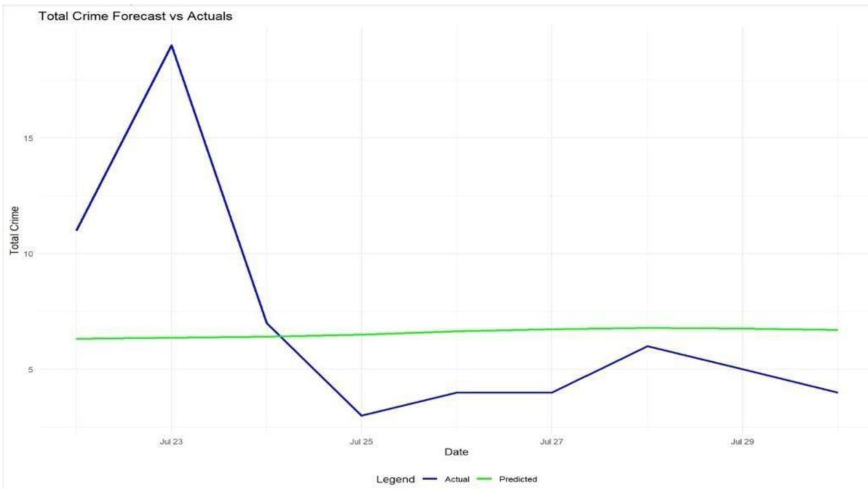
Date	Actual	Predicted
2023-07-22	165	183.8315
2023-07-23	200	180.9831
2023-07-24	190	183.8069
2023-07-25	168	182.7872
2023-07-26	186	187.6331
2023-07-27	170	187.3087
2023-07-28	206	189.9036
2023-07-29	214	186.7181
2023-07-30	177	185.7425

```
ARIMA Model Summary:
Series: train_data$Total_Crime
Regression with ARIMA(1,1,2) errors

Coefficients:
          ar1     ma1      ma2    xreg
      -0.9138  0.1824  -0.7882  1.0588
s.e.   0.0408  0.0478   0.0443  0.4198

sigma^2 = 549.3:  log likelihood = -918.01
AIC=1846.03   AICc=1846.34   BIC=1862.55

Training set error measures:
                    ME     RMSE      MAE       MPE     MAPE      MASE       ACF1
Training set -0.913373 23.1456 18.00149 -1.963036 10.02432 0.7711636 -0.0318418
```

The ARIMA model was employed for forecasting, utilizing key variables such as temperature. The model parameters were determined as ARIMA(1,1,2), where 1 represents the order of auto ARIMA (P), 1 signifies the order of integration (d), and 2 denotes the order of moving averages (q). Further details regarding error metrics can be found in the model summary presented below.

Prediction Plot for Louisville, Division-4, Beat 1

The prediction plot below illustrates the forecasted total crime incidents in Louisville, Division-4, Beat 1, from July 22, 2023, to July 30, 2023. The plot compares the actual crime occurrences against the predictions made by the ARIMA model, providing a visual representation of the model's performance.



The blue line represents the actual total crime incidents recorded on each day, while the green line represents the predicted values generated by the ARIMA model. This visual representation enables a quick assessment of the model's accuracy in capturing the observed trends and patterns specific to Louisville, Division-4, Beat 1.

Actual vs Predicted Values

Date	Actual	Predicted
2023-07-22	11	6.324266
2023-07-23	19	6.364702
2023-07-24	7	6.417672
2023-07-25	3	6.502986
2023-07-26	4	6.640441
2023-07-27	4	6.732283
2023-07-28	6	6.794361
2023-07-29	5	6.764156
2023-07-30	4	6.700589

```
ARIMA Model Summary:
Series: train_data$Total_Crime
Regression with ARIMA(1,1,2) errors

Coefficients:
         ar1      ma1     ma2    xreg
      0.9293  -1.8208  0.8209  0.0392
s.e.  0.0648   0.0944  0.0938  0.0515

sigma^2 = 11.65:  log likelihood = -523.9
AIC=1057.8   AICc=1058.11   BIC=1074.24

Training set error measures:
                     ME     RMSE      MAE      MPE     MAPE      MASE       ACF1
Training set -0.1918451 3.370498 2.743232 -42.3081 65.64167 0.7715341 0.01172366
```

Conclusion:

In conclusion, the ARIMA model, with its robust forecasting capabilities, emerges as a pivotal tool for informed decision-making in law enforcement and public safety. By delving into time series analysis, as exemplified by the ARIMA model, we unlock a powerful means of comprehending and predicting sequential data patterns. This not only provides invaluable insights for proactive crime prevention but also facilitates strategic resource allocation. The multifaceted analysis, incorporating diverse modeling techniques such as linear regression and random forest regression, culminates in a comprehensive preventive and alerting system. Beyond aiding the public in avoiding high-crime areas, the model extends its utility by offering guidance on safe travel practices during specific weather conditions. This holistic approach harnesses predictive analytics to augment public safety and awareness, mitigating potential risks associated with both crime and adverse weather. It stands as a testament to the transformative power of data-driven strategies in addressing intricate urban challenges and fostering a safer environment for the community. This comprehensive strategy encapsulates the resilience and adaptability required to meet the challenges of urban environments, providing a foundation for a safer and more secure community.