PREDICTIVE ANALYTICS

PROJECT REPORT

(Project Semester Sep'25 – Jan'26)

CRIME_REVIEW_FOR_MONTHS_FROM_JAN_TO_SEP

Submitted by

ANIKET SAHA

Registration No: - 12314934

P132 – K23HP

Course Code: - INT234

Under the Guidance of

Dr. Gargi Sharma (UID: 29439)

Discipline of CSE/IT

Lovely School of Computer Science and Engineering

Lovely Professional University, Phagwara

CERTIFICATE

This is to certify that ANIKET SAHA bearing Registration no. 12314934 has completed INT234 project titled, **"CRIME_REVIEW_FOR_MONTHS_FROM_JAN_TO_SEP"** under my guidance and supervision. To the best of my knowledge, the present work is the result of his/her original development, effort and study.

Signature and Name of the Supervisor

Designation of the Supervisor

School of Computer Science and Engineering

Lovely Professional University

Phagwara, Punjab.

Date: 15/12/2025

DECLARATION

I, **ANIKET SAHA**, student of *P132: BTech (Computer Science and Engineering)* under CSE/IT Discipline at, Lovely Professional University, Punjab, hereby declare that all the information furnished in this project report is based on my own intensive work and is genuine.

Date: 15-12-2025 ANIKET SAHA

Registration No: - 12314934 Name of the student

ACKNOWLEDGEMENT

I would like to express my heartfelt gratitude to everyone who has supported and guided me throughout the completion of my project on " ***CRIME_REVIEW_FOR_MONTHS_FROM_JAN_TO_SEP*** "

First and foremost, I extend my sincere thanks to my professor/mentor, Dr. Mrinalini Rana, for their invaluable guidance, encouragement, and expertise, which were instrumental in shaping the direction and scope of this project. Their insights on data analysis techniques and interpretation have greatly enhanced my understanding of the subject.

I am also grateful to my institution Lovely Professional University for providing me with the resources and platform to work on this project. The tools and datasets made available to me were crucial in conducting meaningful analyses.

I would like to acknowledge the contributions of my peers and colleagues, whose constructive feedback and collaboration have helped me refine my work. Their perspectives and suggestions have added depth to my analysis.

Lastly, I owe my deepest gratitude to my family and friends for their unwavering support and motivation throughout the duration of this project. Their encouragement has been a constant source of inspiration.

This project has been a valuable learning experience, and I am deeply thankful to everyone who has contributed to its successful completion.

Sincerely,
ANIKET SAHA

TABLE OF CONTENTS

INTRODUCTION

1. Project Overview

This project focuses on the analysis and modeling of crime data collected from monthly crime review reports (January–September 2025) to identify meaningful patterns, trends, and risk indicators related to public safety. The analysis aims to support law enforcement agencies, policymakers, and governance bodies by providing data-driven insights that can aid in crime prevention and resource planning.

The project involves integrating multiple monthly datasets into a single unified dataset, followed by exploratory data analysis, data cleaning, and the application of supervised machine learning techniques. Key aspects such as crime category distribution, temporal crime trends, severity classification, hotspot detection, and crime forecasting are studied to understand crime behavior over time.

Unlike basic descriptive analysis, this project combines statistical analysis and predictive modeling to move beyond "what happened" and focus on what is likely to happen next. The results are presented in a structured and interpretable manner, making complex crime data easier to understand and actionable for decision-making related to public safety and crime control.

2. Dataset Description

The dataset used in this project consists of monthly crime review records, where each record represents reported crime statistics for specific crime categories within a given month. The data spans nine months (January–September 2025) and is consolidated into a single analytical dataset.

Each dataset includes information related to:

- Broad crime classification (Heads of Crime)
- Major crime categories (Major Heads)
- Sub-classifications of crimes (Minor Heads)
- Crime counts recorded during the current month
- Cumulative crime counts up to the end of the month under review

To support time-based analysis, additional columns such as Month Name, Month Index, and Year are added during data integration. This dataset serves as the foundation for all exploratory analysis, predictive modeling, and public safety insights generated in this project.

3. Objectives

The primary objectives of this project are as follows:

1. Crime Trend Analysis
    - Analyze how different crime categories change over time.
    - Identify crime types that show consistent or rapid growth.
    - Support early detection of emerging crime threats.
2. Crime Severity Classification
    - Classify crime categories into severity levels based on impact and frequency.
    - Assist authorities in prioritizing enforcement and response strategies.
3. High-Crime Risk Identification
    - Distinguish between high-crime and low-crime patterns using supervised classification.
    - Support proactive policing and preventive measures.
4. Crime Forecasting and Prediction
    - Predict future crime levels using historical trends.
    - Aid planning for manpower allocation, patrol deployment, and surveillance.
5. Public Safety Risk Assessment
    - Develop a risk scoring approach to identify crime categories posing the highest threat to society.
    - Enable informed decision-making for governance and public protection.

4. Tools and Techniques

• Data Analysis and Processing

- Integration of multiple monthly crime datasets into a unified structure
- Data cleaning, handling missing values, and standardizing categories
- Feature engineering for time-based and severity-related attributes
- Preparation of datasets for regression and classification models

• Machine Learning and Analytics

- Simple and Multiple Linear Regression for trend and forecasting analysis
- Polynomial Regression and Ordinary Least Squares (OLS) for interpretability
- Logistic Regression and Naive Bayes for classification tasks
- Decision Trees, Support Vector Machines (SVM), and k-Nearest Neighbors (k-NN) for risk modeling

• Evaluation Metrics

- Regression metrics: MAE, MSE, RMSE, R^2
- Classification metrics: Accuracy, Precision, Recall, F1 Score, Confusion Matrix

• Visualization

- Line charts for crime trends
- Bar charts for crime distribution
- Heatmaps for correlation analysis
- Model performance visualizations for interpretability

5. Anticipated Outcomes

The expected outcomes of this project include:

- Identification of crime categories that require immediate attention due to rising trends
- A structured classification of crimes based on severity and social impact
- Predictive insights that support proactive policing and crime prevention
- A data-driven framework for assessing public safety risks
- A scalable analytical approach that can be extended with future crime data

SOURCE OF DATASET

Dataset Source:
Government-published Monthly Crime Review Reports (January–September 2025)

LinkedIn Project Post:

DATA INTRODUCTION

1. Source of Data

The data used in this project is derived from official monthly crime review reports, which provide summarized crime statistics across various crime categories. These reports are commonly used for monitoring law-and-order conditions and assessing crime trends at an administrative level. The data is considered structured and reliable, making it suitable for analytical and predictive modeling purposes.

2. Nature of the Dataset

The dataset is categorical and time-based in nature. Each record represents crime statistics for a specific crime category within a given month. The presence of temporal indicators allows for trend analysis and forecasting.

The dataset contains:

- Crime classification levels (Heads, Major Heads, Minor Heads)
- Monthly crime counts
- Cumulative crime statistics
- Time identifiers (Month and Year)

This structure supports descriptive analysis, comparative evaluation, and supervised learning techniques.

3. Data Storage Format

The datasets are originally stored in CSV format and are merged into a single unified dataset for analysis. CSV format ensures compatibility with data analysis tools and allows efficient preprocessing, transformation, and modeling.

After integration:

- A consistent schema is maintained
- Redundant and empty columns are removed
- Cleaned data is stored for reuse in analysis and modeling
-

4. Data Refresh and Updates

The analytical framework supports easy updates by allowing new monthly crime datasets to be appended to the existing dataset. Upon updating the data source, all analyses and predictive models can be re-run to reflect the latest crime trends.

This design ensures scalability and long-term usability of the project

5. Data Usage in the Analysis

The unified crime dataset is used for:

- Analyzing crime trends over time
- Identifying high-risk and high-severity crime categories
- Building predictive models for crime forecasting
- Supporting public safety planning and policy formulation

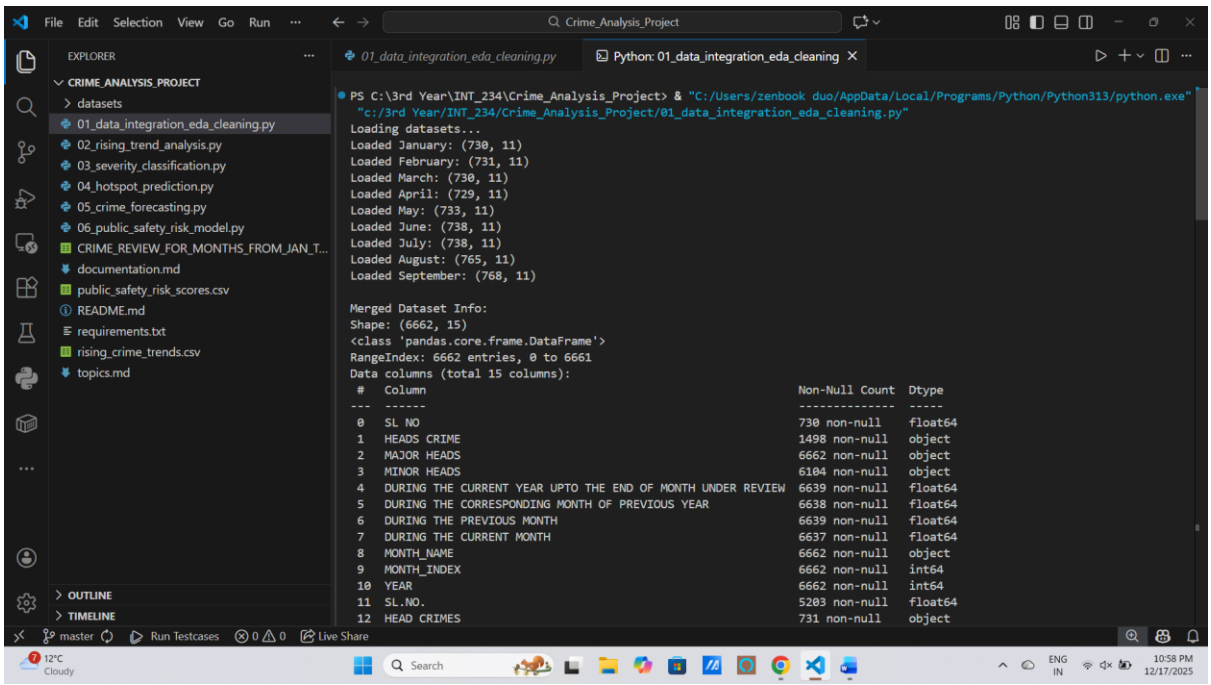
The structured use of this dataset enables clear, insight-driven analysis that aligns with real-world law enforcement and governance needs.

DATA PREPARATION AND PREPROCESSING

Data preparation and preprocessing form a crucial phase of this project, as the quality of analysis and predictive modeling largely depends on the consistency and reliability of the input data. Since the crime data was collected from multiple monthly reports, careful preprocessing was required to transform the raw datasets into a unified and analysis-ready format.

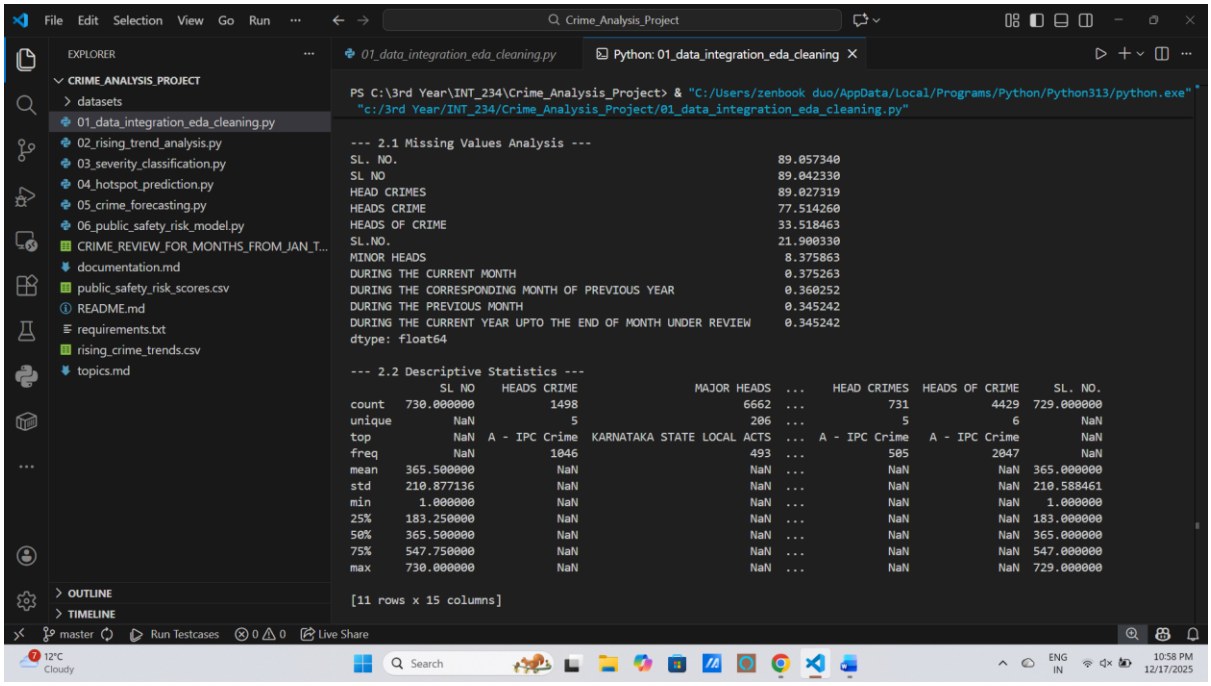
1. Data Integration

The project initially involved **multiple monthly crime datasets** ranging from January to September 2025. Each dataset followed a similar structure but existed as an independent file. To enable longitudinal analysis, all monthly datasets were systematically merged into a **single unified dataset**.
During this process, new identifying attributes such as **Month Name, Month Index, and Year** were added to each record to clearly distinguish data belonging to different months and to support time-based analysis.



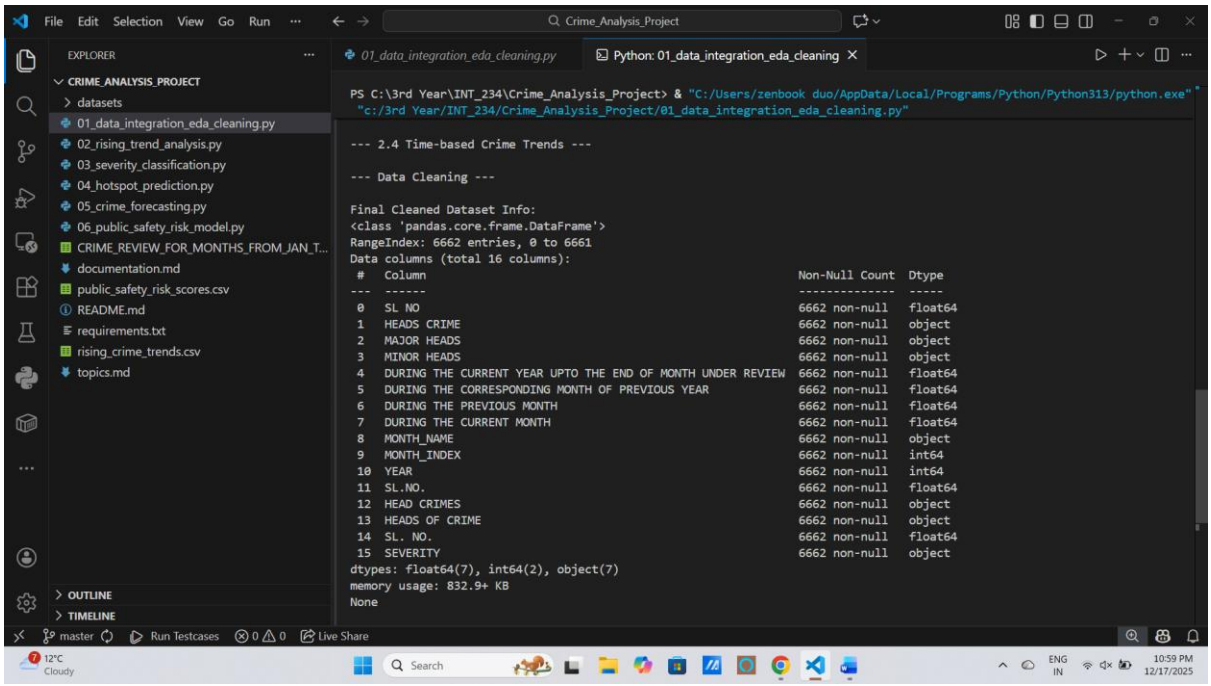
2. Column Standardization and Cleaning

Before merging, all datasets were inspected for column consistency. Columns that were completely empty or redundant (such as unnamed columns generated during data export) were removed. Column names were standardized to ensure uniform naming conventions across all months, preventing schema mismatches during integration.



3. Handling Missing Values

The dataset was examined for missing and null values across all attributes. Columns containing insignificant or non-informative missing values were treated appropriately to avoid bias in analysis. Where required, missing values were either handled through logical replacement or excluded based on their relevance to the analytical objectives. This ensured that the dataset remained both accurate and representative.



4. Data Type Conversion

To maintain analytical correctness, appropriate data types were assigned to each column. Crime classification fields were treated as categorical variables, while crime count fields were converted into numerical formats suitable for statistical analysis and machine learning models. Temporal identifiers such as month index and year were preserved as numeric features to support regression and forecasting tasks.

5. Duplicate and Consistency Checks

The unified dataset was examined for duplicate records and inconsistencies resulting from data integration. Any redundant entries were identified and removed to maintain data integrity. Consistency checks were also performed to ensure that crime categories and labels remained uniform across all months.

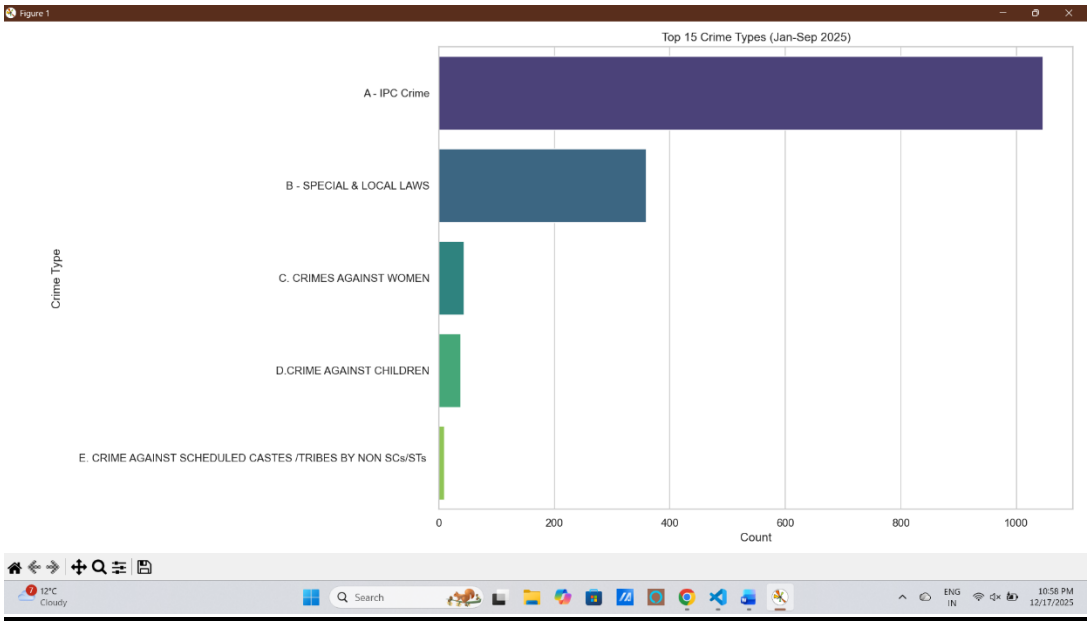
6. Feature Engineering

Additional features were derived to enhance analytical depth and model performance. Time-based attributes enabled trend analysis, while derived indicators such as severity grouping and trend-related measures supported classification and risk assessment objectives. These engineered features played a key role in linking raw crime data to meaningful public safety insights.

7. Final Prepared Dataset

After completing all preprocessing steps, a clean and structured dataset was finalized for further exploratory data analysis and supervised learning. The prepared dataset was validated through summary statistics and sample inspections to confirm its suitability for modeling and interpretation.

Overall, the data preparation and preprocessing stage ensured that the crime data was accurate, consistent, and ready for advanced analysis, forming a reliable foundation for trend detection, classification, forecasting, and public safety risk assessment.



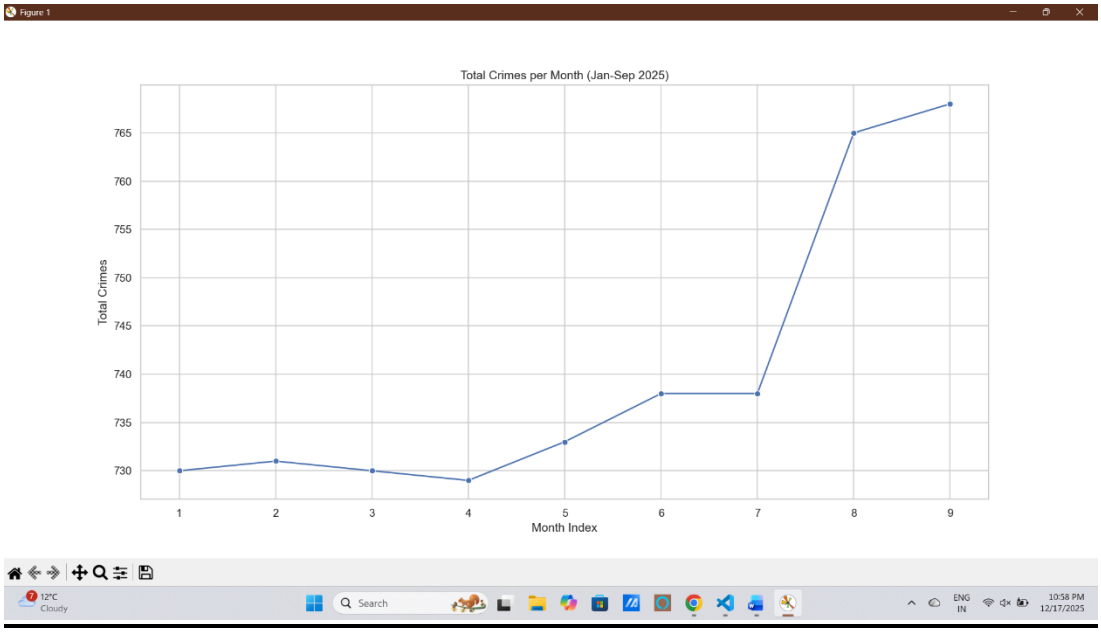
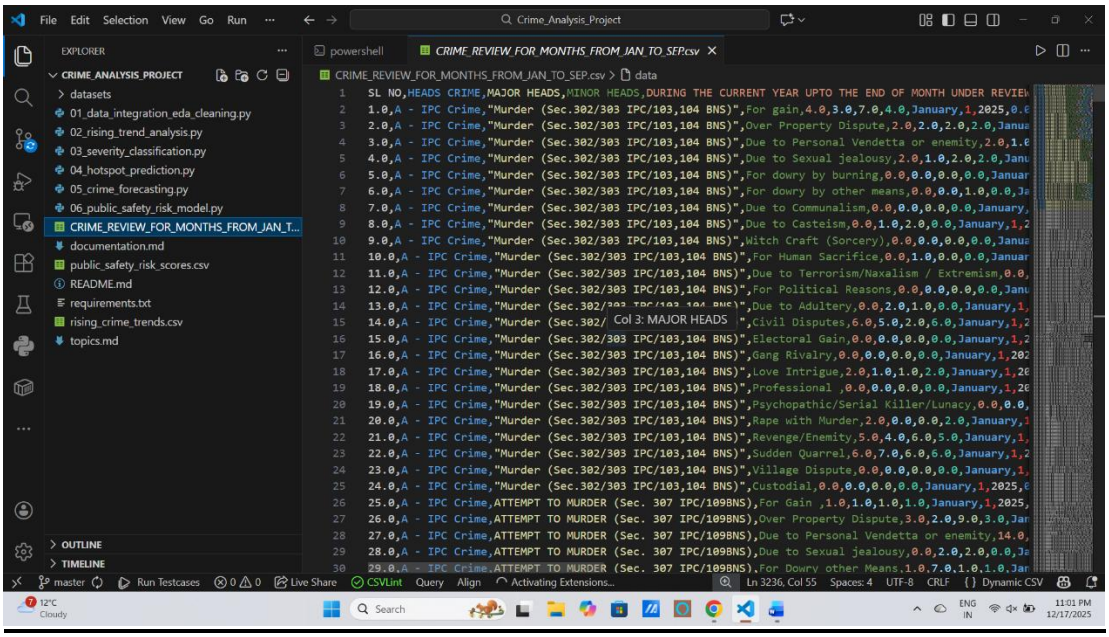


Figure 1
Total Crimes per Month (Jan-Sep 2025)
Total Crimes
765
760
755
750
745
740
735
730
1
2
3
4
5
6
7
8
9
Month Index
12°C
Cloudy
Search
ENG
IN
10:58 PM
12/17/2025


File Edit Selection View Go Run
Crime_Analysis_Project
EXPLORER
CRIME_ANALYSIS_PROJECT
datasets
01_data_integration_eda_cleaning.py
02_rising_trend_analysis.py
03_severity_classification.py
04_hotspot_prediction.py
05_crime_forecasting.py
06_public_safety_risk_model.py
CRIME_REVIEW_FOR_MONTHS_FROM_JAN_T...
documentation.md
public_safety_risk_scores.csv
README.md
requirements.txt
rising_crime_trends.csv
topics.md
OUTLINE
TIMELINE
powershell
CRIME_REVIEW_FOR_MONTHS_FROM_JAN_TO_SEP.csv
CRIME_REVIEW_FOR_MONTHS_FROM_JAN_TO_SEP.csv > data
SL NO,HEADS CRIME,MAJOR HEADS,MINOR HEADS,DURING THE CURRENT YEAR UPTO THE END OF MONTH UNDER REVIEW
1.0,A - IPC Crime,"Murder (Sec.302/303 IPC/103,104 BNS)",For gain,4.0,3.0,7.0,4.0,January,1,2025,0.0
2.0,A - IPC Crime,"Murder (Sec.302/303 IPC/103,104 BNS)",Over Property Dispute,2.0,2.0,2.0,2.0,Janua
3.0,A - IPC Crime,"Murder (Sec.302/303 IPC/103,104 BNS)",Due to Personal Vendetta or enemity,2.0,1.0
4.0,A - IPC Crime,"Murder (Sec.302/303 IPC/103,104 BNS)",Due to Sexual jealousy,2.0,1.0,2.0,2.0,Janu
5.0,A - IPC Crime,"Murder (Sec.302/303 IPC/103,104 BNS)",For dowry by burning,0.0,0.0,0.0,0.0,Januar
6.0,A - IPC Crime,"Murder (Sec.302/303 IPC/103,104 BNS)",For dowry by other means,0.0,0.0,1.0,0.0,Ja
7.0,A - IPC Crime,"Murder (Sec.302/303 IPC/103,104 BNS)",Due to Communalism,0.0,0.0,0.0,0.0,January,
8.0,A - IPC Crime,"Murder (Sec.302/303 IPC/103,104 BNS)",Due to Casteism,0.0,1.0,2.0,0.0,January,1,2
9.0,A - IPC Crime,"Murder (Sec.302/303 IPC/103,104 BNS)",Witch Craft (Sorcery),0.0,0.0,0.0,0.0,Janua
10.0,A - IPC Crime,"Murder (Sec.302/303 IPC/103,104 BNS)",For Human Sacrifice,0.0,1.0,0.0,0.0,Januar
11.0,A - IPC Crime,"Murder (Sec.302/303 IPC/103,104 BNS)",Due to Terrorism/Naxalism / Extremism,0.0,
12.0,A - IPC Crime,"Murder (Sec.302/303 IPC/103,104 BNS)",For Political Reasons,0.0,0.0,0.0,0.0,Janu
13.0,A - IPC Crime,"Murder (Sec.302/303 IPC/103,104 BNS)",Due to Adultery,0.0,2.0,1.0,0.0,January,1,
14.0,A - IPC Crime,"Murder (Sec.302/ Col 3: MAJOR HEADS ",Civil Disputes,6.0,5.0,2.0,6.0,January,1,2
15.0,A - IPC Crime,"Murder (Sec.302/303 IPC/103,104 BNS)",Electoral Gain,0.0,0.0,0.0,0.0,January,1,2
16.0,A - IPC Crime,"Murder (Sec.302/303 IPC/103,104 BNS)",Gang Rivalry,0.0,0.0,0.0,0.0,January,1,202
17.0,A - IPC Crime,"Murder (Sec.302/303 IPC/103,104 BNS)",Love Intrigue,2.0,1.0,1.0,2.0,January,1,20
18.0,A - IPC Crime,"Murder (Sec.302/303 IPC/103,104 BNS)",Professional ,0.0,0.0,0.0,0.0,January,1,20
19.0,A - IPC Crime,"Murder (Sec.302/303 IPC/103,104 BNS)",Psychopathic/Serial Killer/Lunacy,0.0,0.0,
20.0,A - IPC Crime,"Murder (Sec.302/303 IPC/103,104 BNS)",Rape with Murder,2.0,0.0,0.0,2.0,January,1
21.0,A - IPC Crime,"Murder (Sec.302/303 IPC/103,104 BNS)",Revenge/Enemity,5.0,4.0,6.0,5.0,January,1,
22.0,A - IPC Crime,"Murder (Sec.302/303 IPC/103,104 BNS)",Sudden Quarrel,6.0,7.0,6.0,6.0,January,1,2
23.0,A - IPC Crime,"Murder (Sec.302/303 IPC/103,104 BNS)",Village Dispute,0.0,0.0,0.0,0.0,January,1,
24.0,A - IPC Crime,"Murder (Sec.302/303 IPC/103,104 BNS)",Custodial,0.0,0.0,0.0,0.0,January,1,2025,0
25.0,A - IPC Crime,ATTEMPT TO MURDER (Sec. 307 IPC/109BNS),For Gain ,1.0,1.0,1.0,1.0,January,1,2025,
26.0,A - IPC Crime,ATTEMPT TO MURDER (Sec. 307 IPC/109BNS),Over Property Dispute,3.0,2.0,9.0,3.0,Jan
27.0,A - IPC Crime,ATTEMPT TO MURDER (Sec. 307 IPC/109BNS),Due to Personal Vendetta or enemity,14.0,
28.0,A - IPC Crime,ATTEMPT TO MURDER (Sec. 307 IPC/109BNS),Due to Sexual jealousy,0.0,2.0,2.0,0.0,Ja
29.0,A - IPC Crime,ATTEMPT TO MURDER (Sec. 307 IPC/109BNS),For Dowry other Means,1.0,7.0,1.0,1.0,Jan
master
Run Testcases
Live Share
CSVLint
Query
Align
Activating Extensions...
Ln 3236, Col 55
Spaces: 4
UTF-8
CRLF
{ } Dynamic CSV
12°C
Cloudy
Search
ENG
IN
11:01 PM
12/17/2025

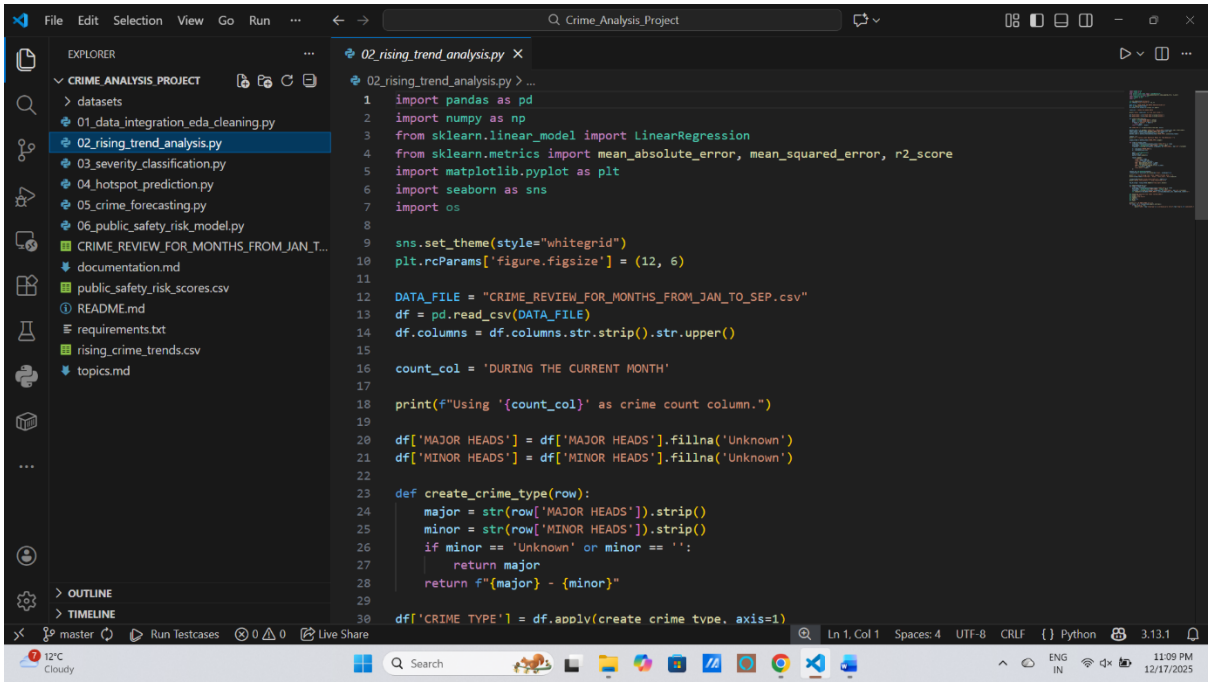
DATA ANALYSIS

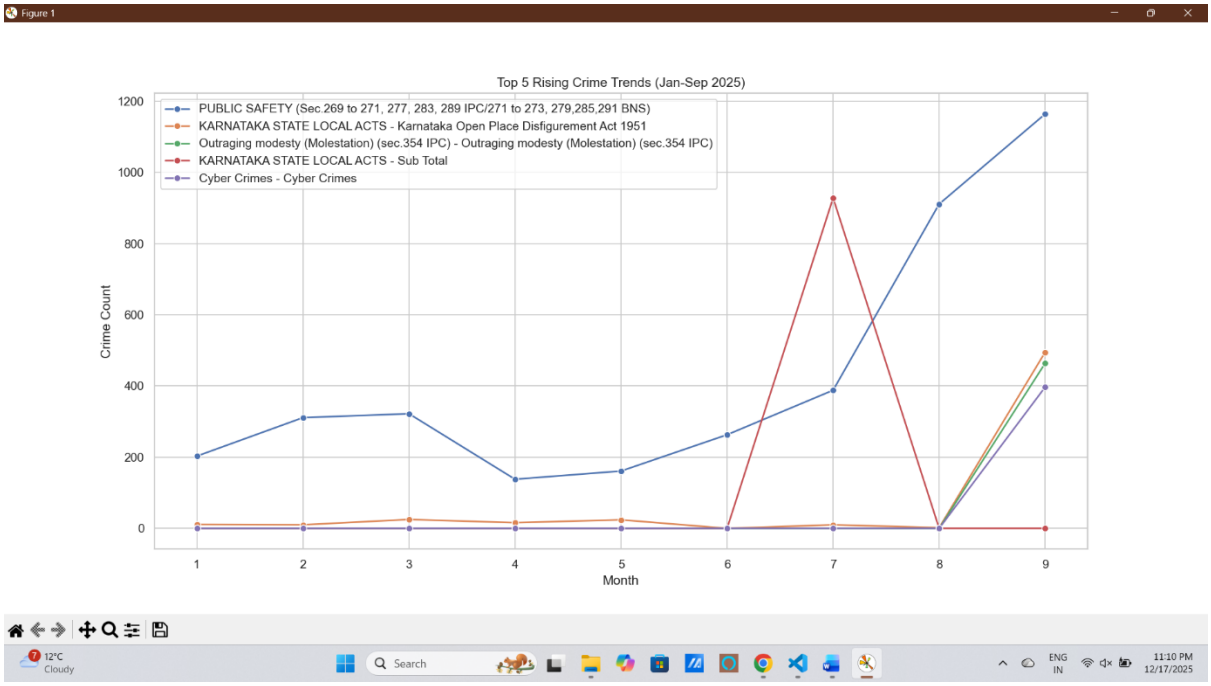
The data analysis phase focuses on extracting meaningful insights from the prepared crime dataset and translating them into information that can support public safety and governance decisions. Using the unified dataset covering January to September 2025, multiple analytical techniques were applied to understand crime behavior, identify emerging patterns, and support predictive modeling.

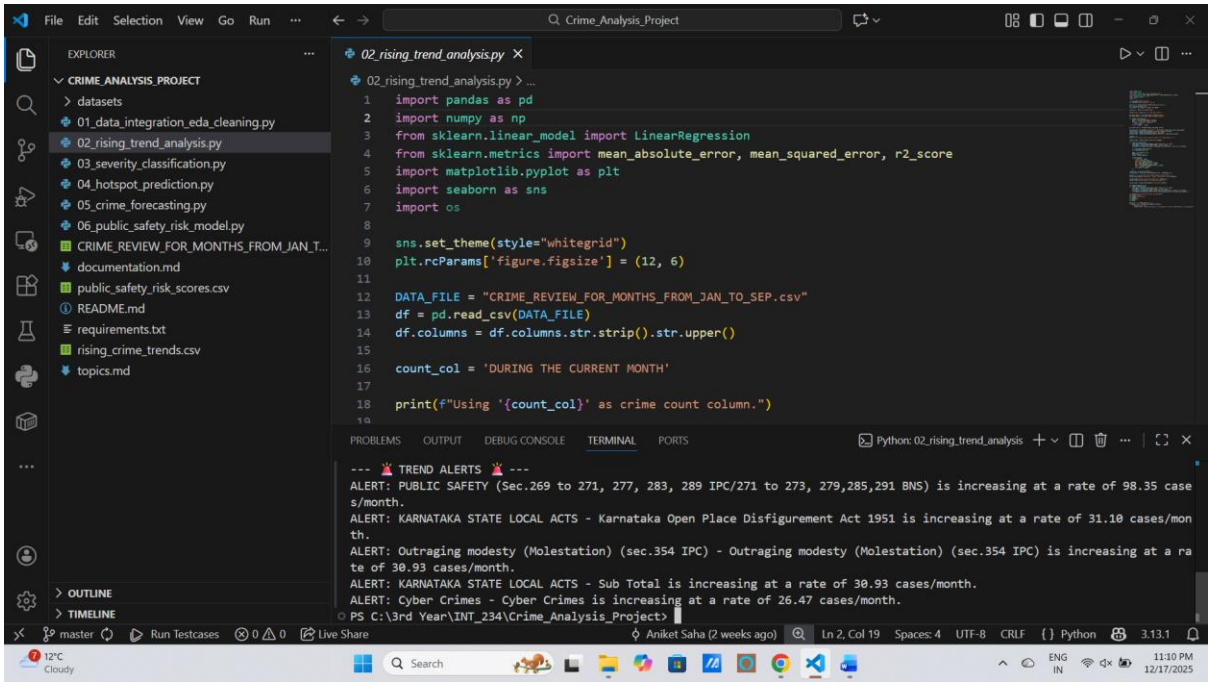
1. Exploratory Analysis of Crime Distribution

An initial exploratory analysis was conducted to understand how crime incidents are distributed across different crime categories. The analysis revealed that certain **major crime categories consistently account for a large proportion of total reported cases**, while several minor categories contribute relatively fewer incidents. This uneven distribution highlights the need for targeted intervention strategies rather than uniform policing approaches.

Crime counts were also examined across months, revealing fluctuations in reporting patterns. Some categories showed stable behavior, while others exhibited noticeable increases or decreases over time, indicating potential seasonal or behavioral influences.







2. Temporal Trend Analysis

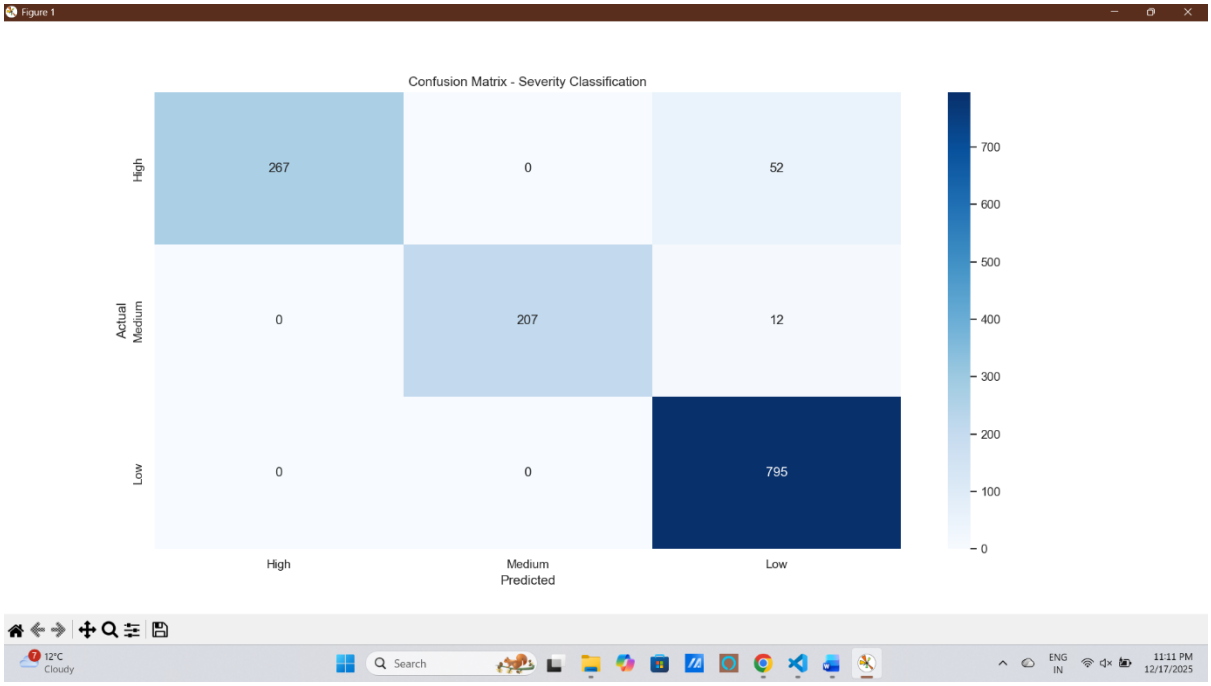
Time-based analysis was performed using the month identifiers introduced during preprocessing. Monthly crime trends were analyzed for major crime categories to detect growth patterns and emerging risks. The results indicated that **specific crime types demonstrate a steady upward trend over the observed period**, suggesting the need for early preventive measures.

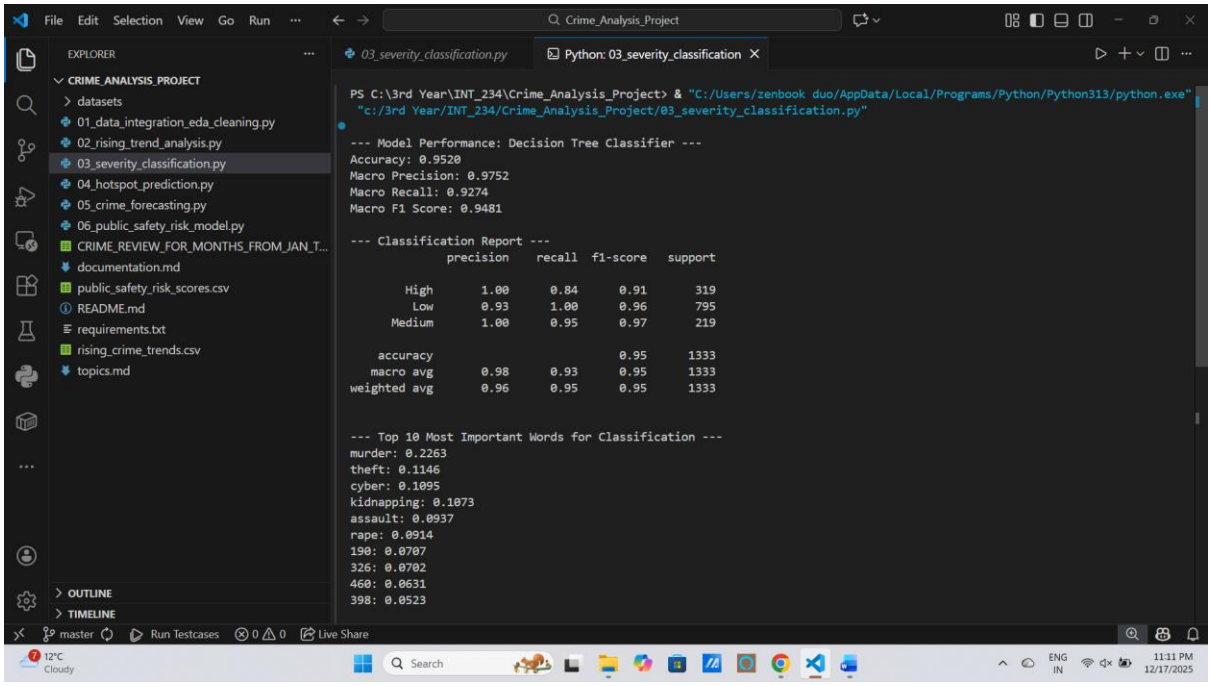
Simple linear regression was used to quantify these trends and assess whether changes in crime levels were statistically meaningful. Crimes with positive trend slopes were identified as potential areas of concern for law enforcement planning.

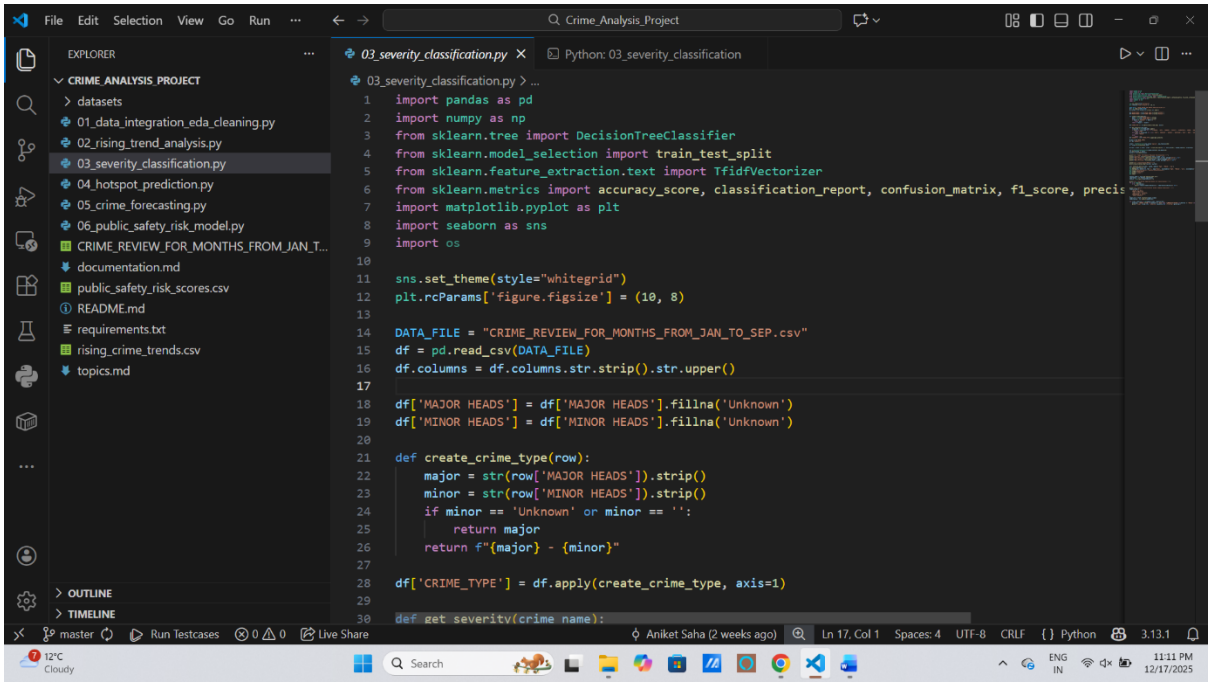
3. Crime Severity and Impact Analysis

Crime categories were further analyzed based on their severity and societal impact. By examining frequency and classification levels, crimes were grouped into severity tiers. This analysis provided a structured understanding of which crime types pose greater threats to public safety and require rapid response mechanisms.

The severity-based analysis supports strategic decision-making by allowing authorities to allocate resources proportionately, focusing on crimes that have both high frequency and high social impact.



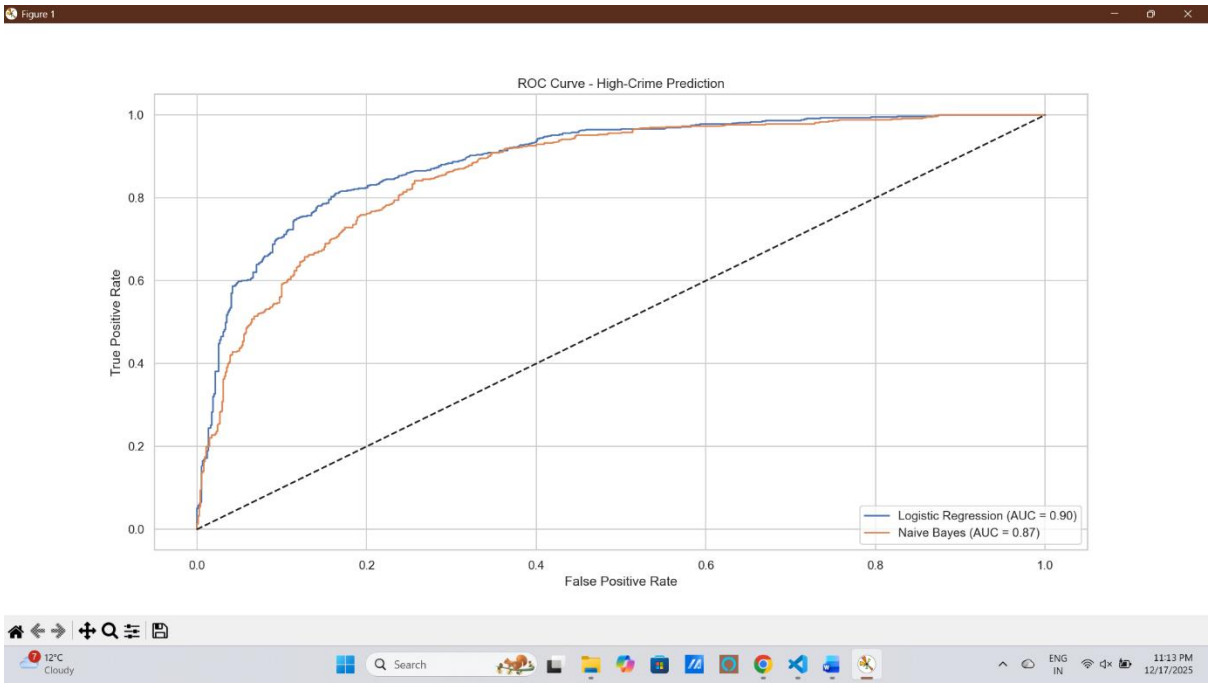


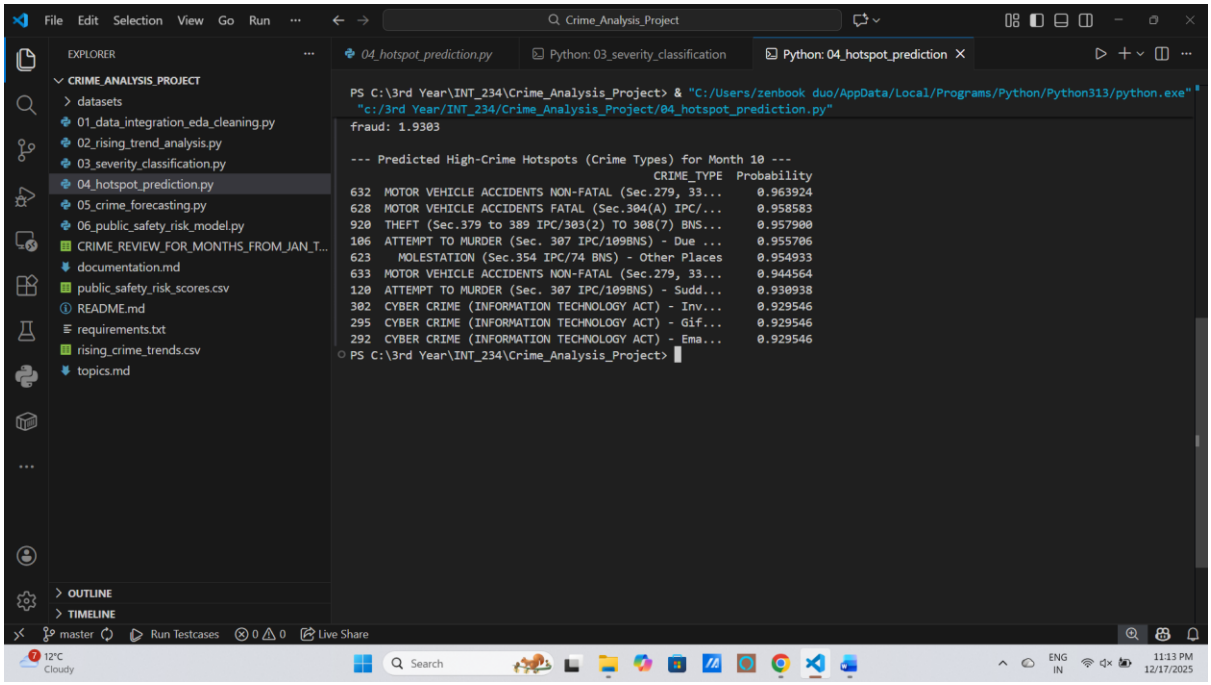


4. High-Crime Risk Identification

Classification-based analysis was carried out to distinguish between high-crime and low-crime patterns within the dataset. By applying supervised learning techniques, crime categories with consistently higher incident counts were identified as high-risk. The results demonstrated that historical crime frequency and category characteristics play a significant role in determining risk levels.

This analysis highlights the importance of data-driven early warning systems that can assist law enforcement in identifying crime-prone areas or categories before situations escalate.



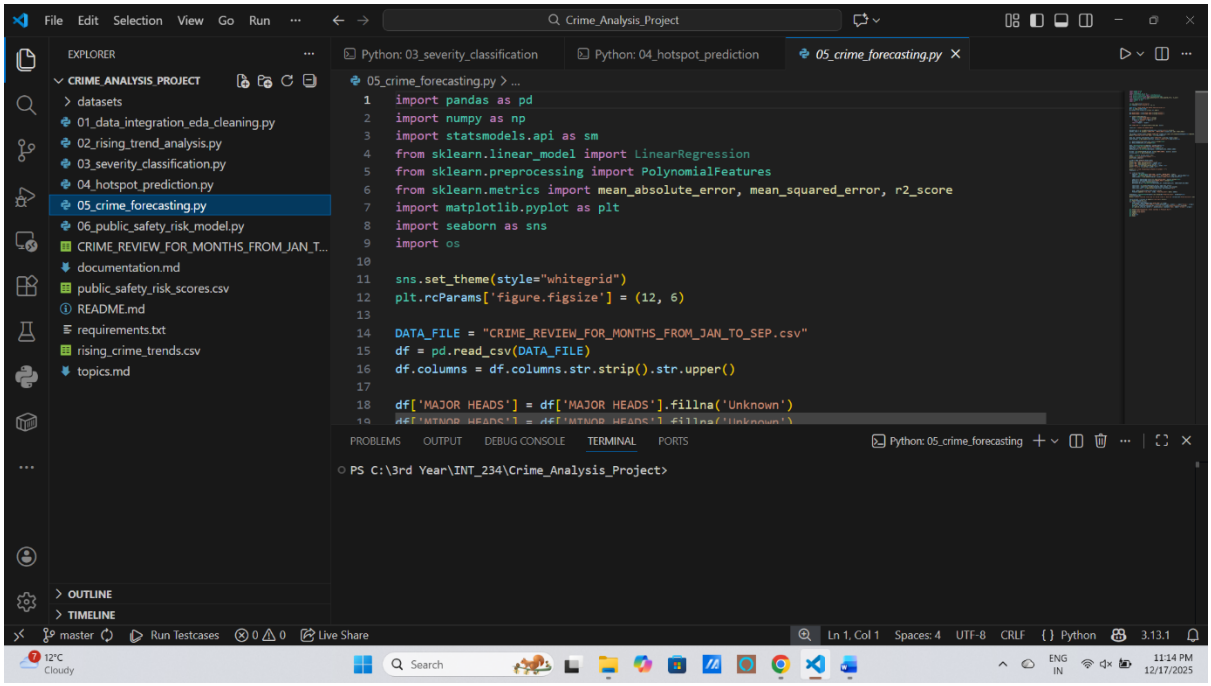


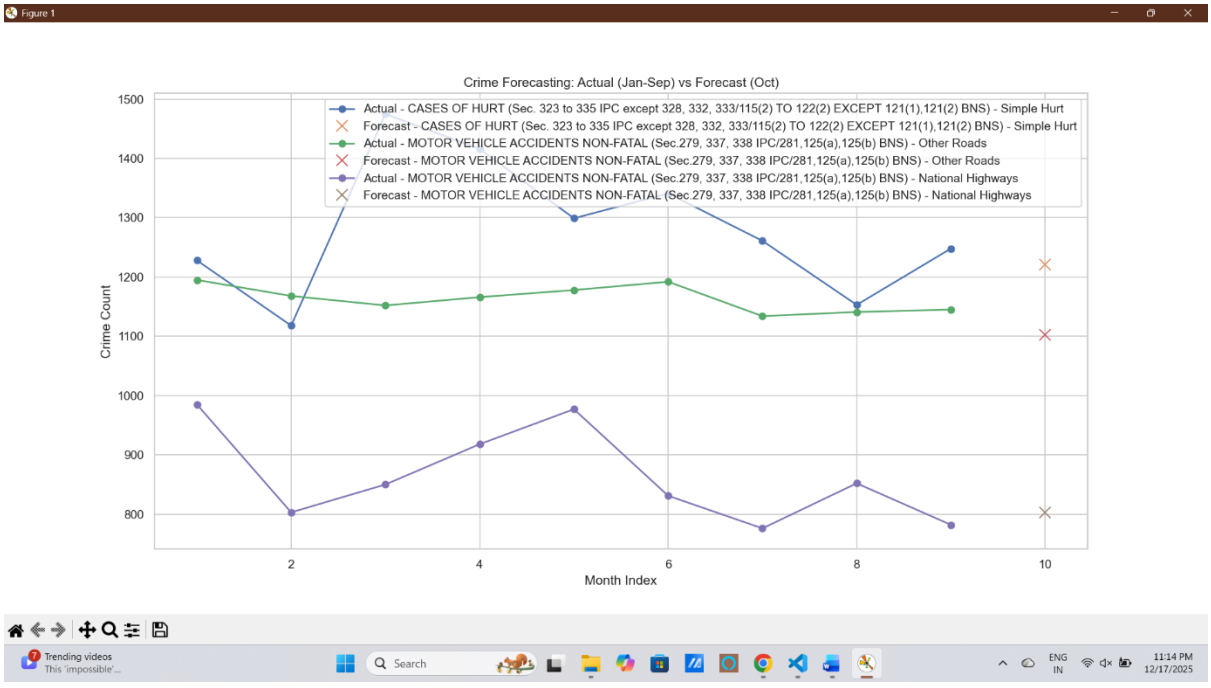
5. Predictive and Forecasting Analysis

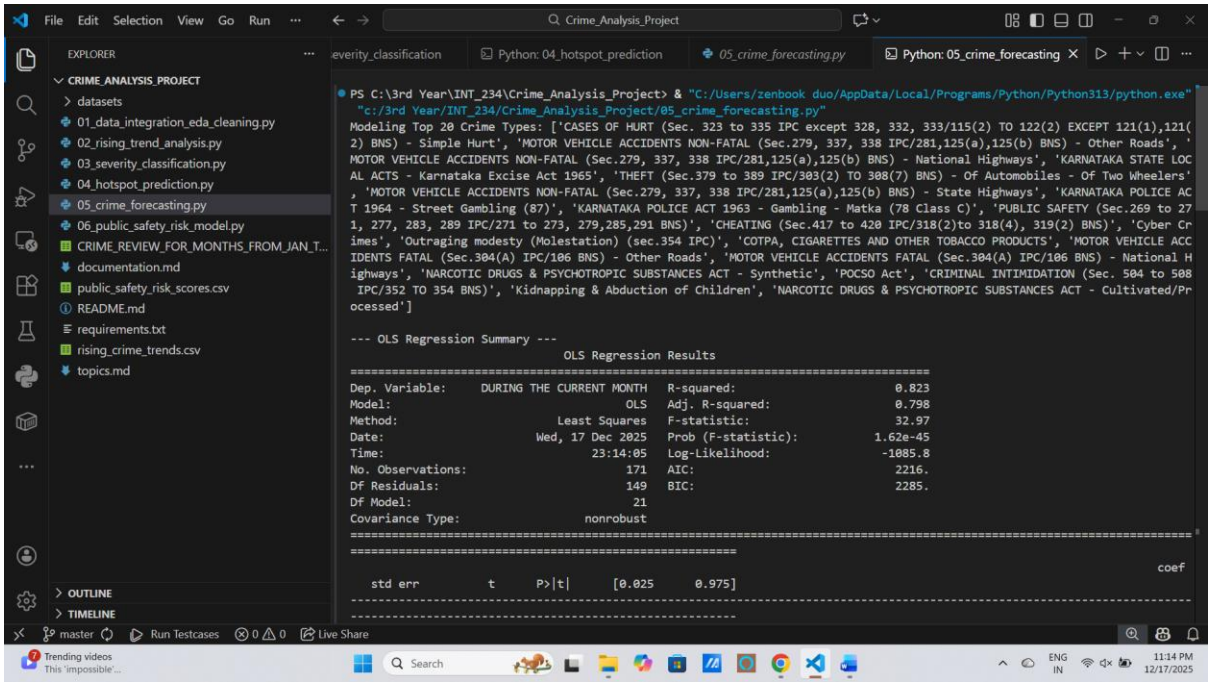
Predictive modeling techniques were applied to estimate future crime behavior based on historical trends. Regression-based models were used to forecast crime counts, providing insights into potential future demand on law enforcement resources. These predictions are

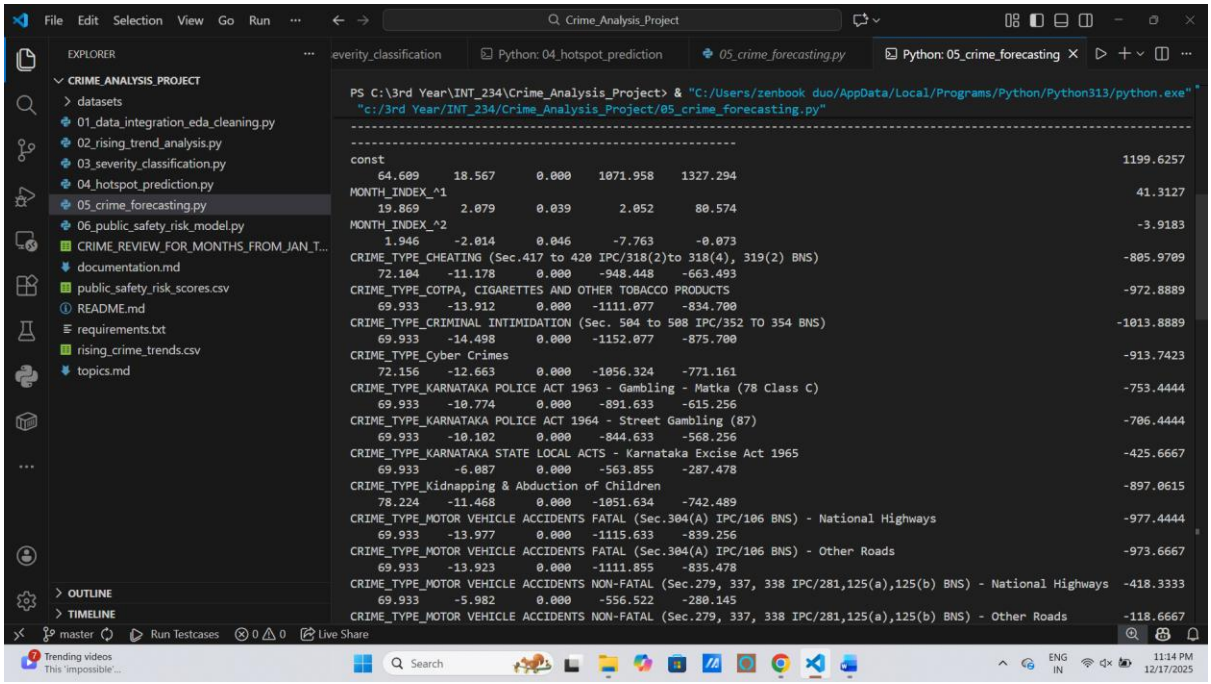
particularly useful for planning patrol schedules, manpower deployment, and preventive surveillance strategies.

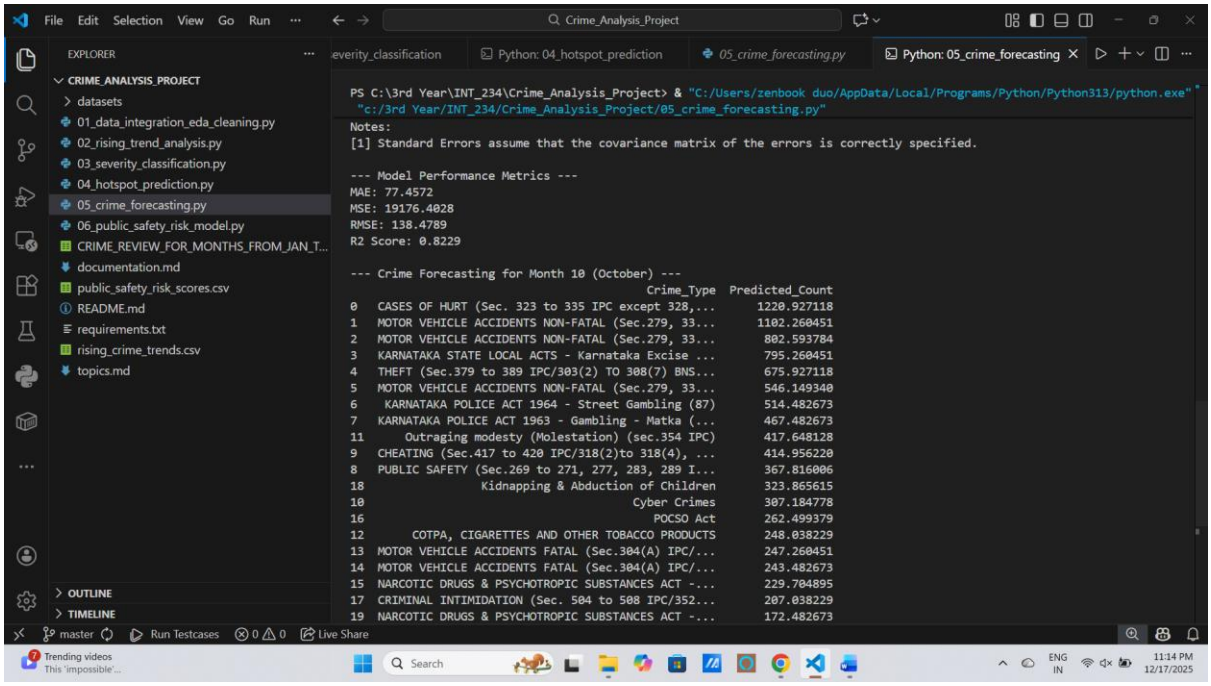
The forecasting analysis showed that trend-aware models can reasonably approximate short-term crime behavior, reinforcing the value of predictive analytics in crime prevention initiatives.






File Edit Selection View Go Run
Crime_Analysis_Project
EXPLORER
CRIME_ANALYSIS_PROJECT
datasets
01_data_integration_eda_cleaning.py
02_rising_trend_analysis.py
03_severity_classification.py
04_hotspot_prediction.py
05_crime_forecasting.py
06_public_safety_risk_model.py
CRIME_REVIEW_FOR_MONTHS_FROM_JAN_T...
documentation.md
public_safety_risk_scores.csv
README.md
requirements.txt
rising_crime_trends.csv
topics.md
OUTLINE
TIMELINE
Python: 04_hotspot_prediction
05_crime_forecasting.py
Python: 05_crime_forecasting
PS C:\3rd Year\INT_234\Crime_Analysis_Project> & "C:/Users/zenbook duo/AppData/Local/Programs/Python/Python313/python.exe" "c:/3rd Year/INT_234/Crime_Analysis_Project/05_crime_forecasting.py"
Modeling Top 20 Crime Types: ['CASES OF HURT (Sec. 323 to 335 IPC except 328, 332, 333/115(2) TO 122(2) EXCEPT 121(1),121(2) BNS) - Simple Hurt', 'MOTOR VEHICLE ACCIDENTS NON-FATAL (Sec.279, 337, 338 IPC/281,125(a),125(b) BNS) - Other Roads', 'MOTOR VEHICLE ACCIDENTS NON-FATAL (Sec.279, 337, 338 IPC/281,125(a),125(b) BNS) - National Highways', 'KARNATAKA STATE LOCAL ACTS - Karnataka Excise Act 1965', 'THEFT (Sec.379 to 389 IPC/303(2) TO 308(7) BNS) - Of Automobiles - Of Two Wheelers', 'MOTOR VEHICLE ACCIDENTS NON-FATAL (Sec.279, 337, 338 IPC/281,125(a),125(b) BNS) - State Highways', 'KARNATAKA POLICE ACT 1964 - Street Gambling (87)', 'KARNATAKA POLICE ACT 1963 - Gambling - Matka (78 Class C)', 'PUBLIC SAFETY (Sec.269 to 271, 277, 283, 289 IPC/271 to 273, 279,285,291 BNS)', 'CHEATING (Sec.417 to 420 IPC/318(2)to 318(4), 319(2) BNS)', 'Cyber Crimes', 'Outraging modesty (Molestation) (sec.354 IPC)', 'COTPA, CIGARETTES AND OTHER TOBACCO PRODUCTS', 'MOTOR VEHICLE ACCIDENTS FATAL (Sec.304(A) IPC/106 BNS) - Other Roads', 'MOTOR VEHICLE ACCIDENTS FATAL (Sec.304(A) IPC/106 BNS) - National Highways', 'NARCOTIC DRUGS & PSYCHOTROPIC SUBSTANCES ACT - Synthetic', 'POCSO Act', 'CRIMINAL INTIMIDATION (Sec. 504 to 508 IPC/352 TO 354 BNS)', 'Kidnapping & Abduction of Children', 'NARCOTIC DRUGS & PSYCHOTROPIC SUBSTANCES ACT - Cultivated/Processed']
--- OLS Regression Summary ---
OLS Regression Results
Dep. Variable: DURING THE CURRENT MONTH R-squared: 0.823
Model: OLS Adj. R-squared: 0.798
Method: Least Squares F-statistic: 32.97
Date: Wed, 17 Dec 2025 Prob (F-statistic): 1.62e-45
Time: 23:14:05 Log-Likelihood: -1085.8
No. Observations: 171 AIC: 2216.
Df Residuals: 149 BIC: 2285.
Df Model: 21
Covariance Type: nonrobust
coef
std err t P>|t| [0.025 0.975]
master
Run Testcases
Live Share
Trending videos
This 'impossible'...
Search
ENG
IN
11:14 PM
12/17/2025

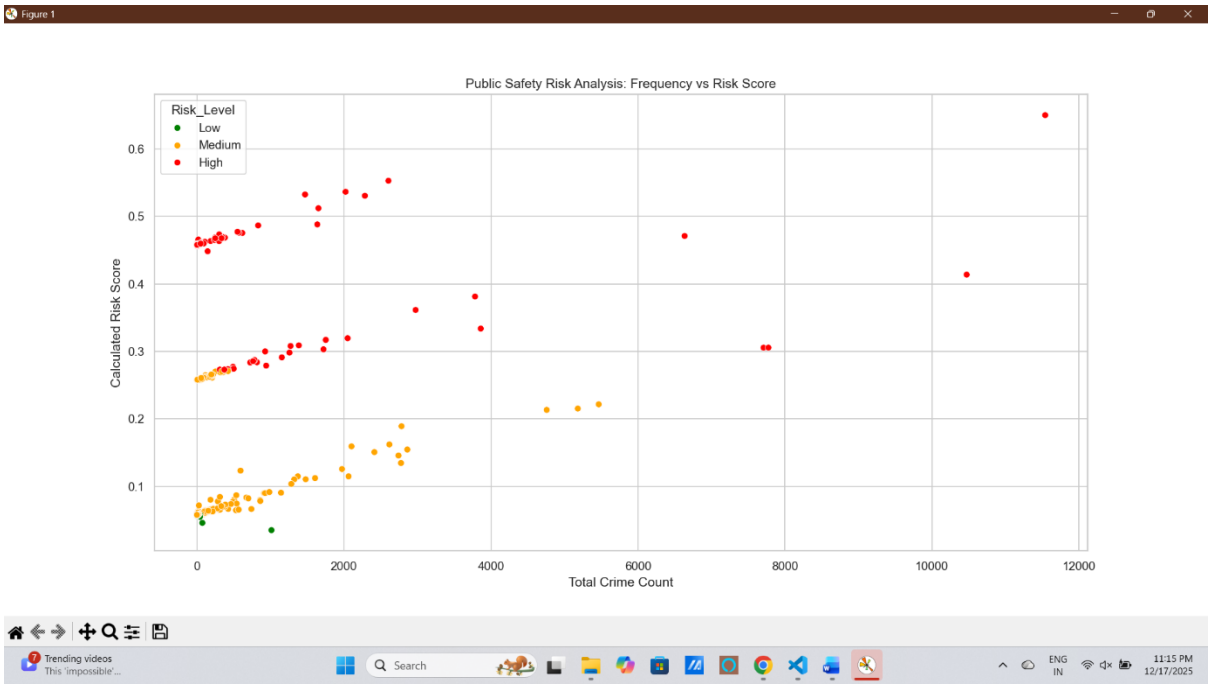

File Edit Selection View Go Run
Crime_Analysis_Project
EXPLORER
CRIME_ANALYSIS_PROJECT
datasets
01_data_integration_eda_cleaning.py
02_rising_trend_analysis.py
03_severity_classification.py
04_hotspot_prediction.py
05_crime_forecasting.py
06_public_safety_risk_model.py
CRIME_REVIEW_FOR_MONTHS_FROM_JAN_T...
documentation.md
public_safety_risk_scores.csv
README.md
requirements.txt
rising_crime_trends.csv
topics.md
OUTLINE
TIMELINE
Python: 04_hotspot_prediction
05_crime_forecasting.py
Python: 05_crime_forecasting
PS C:\3rd Year\INT_234\Crime_Analysis_Project> & "C:/Users/zenbook duo/AppData/Local/Programs/Python/Python313/python.exe" "c:/3rd Year/INT_234/Crime_Analysis_Project/05_crime_forecasting.py"
const 1199.6257
64.609 18.567 0.000 1071.958 1327.294
MONTH_INDEX_^1 41.3127
19.869 2.079 0.039 2.052 80.574
MONTH_INDEX_^2 -3.9183
1.946 -2.014 0.046 -7.763 -0.073
CRIME_TYPE_CHEATING (Sec.417 to 420 IPC/318(2)to 318(4), 319(2) BNS) -805.9709
72.104 -11.178 0.000 -948.448 -663.493
CRIME_TYPE_COTPA, CIGARETTES AND OTHER TOBACCO PRODUCTS -972.8889
69.933 -13.912 0.000 -1111.077 -834.700
CRIME_TYPE_CRIMINAL INTIMIDATION (Sec. 504 to 508 IPC/352 TO 354 BNS) -1013.8889
69.933 -14.498 0.000 -1152.077 -875.700
CRIME_TYPE_Cyber Crimes -913.7423
72.156 -12.663 0.000 -1056.324 -771.161
CRIME_TYPE_KARNATAKA POLICE ACT 1963 - Gambling - Matka (78 Class C) -753.4444
69.933 -10.774 0.000 -891.633 -615.256
CRIME_TYPE_KARNATAKA POLICE ACT 1964 - Street Gambling (87) -706.4444
69.933 -10.102 0.000 -844.633 -568.256
CRIME_TYPE_KARNATAKA STATE LOCAL ACTS - Karnataka Excise Act 1965 -425.6667
69.933 -6.087 0.000 -563.855 -287.478
CRIME_TYPE_Kidnapping & Abduction of Children -897.0615
78.224 -11.468 0.000 -1051.634 -742.489
CRIME_TYPE_MOTOR VEHICLE ACCIDENTS FATAL (Sec.304(A) IPC/106 BNS) - National Highways -977.4444
69.933 -13.977 0.000 -1115.633 -839.256
CRIME_TYPE_MOTOR VEHICLE ACCIDENTS FATAL (Sec.304(A) IPC/106 BNS) - Other Roads -973.6667
69.933 -13.923 0.000 -1111.855 -835.478
CRIME_TYPE_MOTOR VEHICLE ACCIDENTS NON-FATAL (Sec.279, 337, 338 IPC/281,125(a),125(b) BNS) - National Highways -418.3333
69.933 -5.982 0.000 -556.522 -280.145
CRIME_TYPE_MOTOR VEHICLE ACCIDENTS NON-FATAL (Sec.279, 337, 338 IPC/281,125(a),125(b) BNS) - Other Roads -118.6667
master
Run Testcases
Live Share
Trending videos
This 'impossible'...
Search
ENG
IN
11:14 PM
12/17/2025

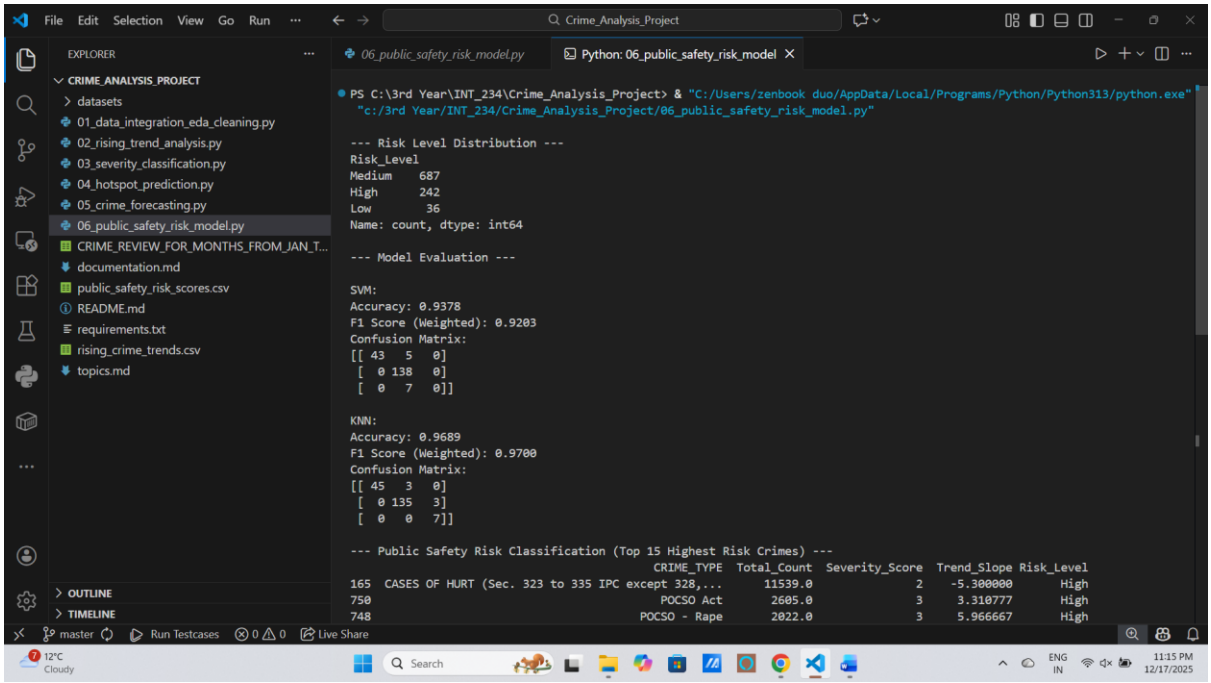


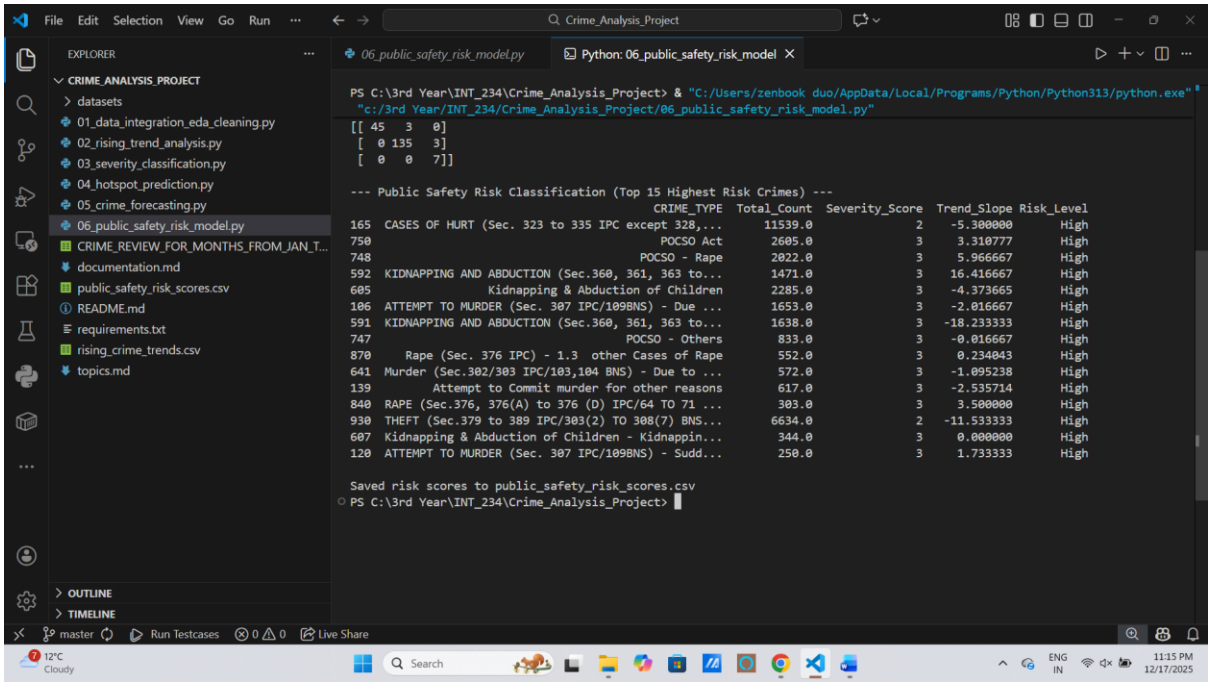
6. Public Safety Risk Assessment

A composite analysis combining crime frequency, trend behavior, and severity levels was used to assess overall public safety risk. Crime categories were ranked based on their potential threat to society, enabling a clearer prioritization framework for policy intervention.

This risk-oriented analysis moves beyond descriptive statistics and emphasizes actionable insights, allowing governance bodies to focus on crime types that pose immediate and long-term risks to public safety.




File Edit Selection View Go Run
Crime_Analysis_Project
EXPLORER
CRIME_ANALYSIS_PROJECT
datasets
01_data_integration_eda_cleaning.py
02_rising_trend_analysis.py
03_severity_classification.py
04_hotspot_prediction.py
05_crime_forecasting.py
06_public_safety_risk_model.py
CRIME_REVIEW_FOR_MONTHS_FROM_JAN_T...
documentation.md
public_safety_risk_scores.csv
README.md
requirements.txt
rising_crime_trends.csv
topics.md
06_public_safety_risk_model.py
Python: 06_public_safety_risk_model
PS C:\3rd Year\INT_234\Crime_Analysis_Project> & "C:/Users/zenbook duo/AppData/Local/Programs/Python/Python313/python.exe" "c:/3rd Year/INT_234/Crime_Analysis_Project/06_public_safety_risk_model.py"
--- Risk Level Distribution ---
Risk_Level
Medium 687
High 242
Low 36
Name: count, dtype: int64
--- Model Evaluation ---
SVM:
Accuracy: 0.9378
F1 Score (Weighted): 0.9203
Confusion Matrix:
[[43 5 0]
[0 138 0]
[0 7 0]]
KNN:
Accuracy: 0.9689
F1 Score (Weighted): 0.9700
Confusion Matrix:
[[45 3 0]
[0 135 3]
[0 0 7]]
--- Public Safety Risk Classification (Top 15 Highest Risk Crimes) ---
CRIME_TYPE Total_Count Severity_Score Trend_Slope Risk_Level
165 CASES OF HURT (Sec. 323 to 335 IPC except 328,... 11539.0 2 -5.300000 High
750 POCSO Act 2605.0 3 3.310777 High
748 POCSO - Rape 2022.0 3 5.966667 High
OUTLINE
TIMELINE
master
Run Testcases
0 0
Live Share
12°C
Cloudy
Search
ENG
IN
11:15 PM
12/17/2025


File Edit Selection View Go Run
Crime_Analysis_Project
EXPLORER
CRIME_ANALYSIS_PROJECT
datasets
01_data_integration_eda_cleaning.py
02_rising_trend_analysis.py
03_severity_classification.py
04_hotspot_prediction.py
05_crime_forecasting.py
06_public_safety_risk_model.py
CRIME_REVIEW_FOR_MONTHS_FROM_JAN_T...
documentation.md
public_safety_risk_scores.csv
README.md
requirements.txt
rising_crime_trends.csv
topics.md
06_public_safety_risk_model.py
Python: 06_public_safety_risk_model
PS C:\3rd Year\INT_234\Crime_Analysis_Project> & "C:/Users/zenbook duo/AppData/Local/Programs/Python/Python313/python.exe" "c:/3rd Year/INT_234/Crime_Analysis_Project/06_public_safety_risk_model.py"
[[45 3 0]
[0 135 3]
[0 0 7]]
--- Public Safety Risk Classification (Top 15 Highest Risk Crimes) ---
CRIME_TYPE Total_Count Severity_Score Trend_Slope Risk_Level
165 CASES OF HURT (Sec. 323 to 335 IPC except 328,... 11539.0 2 -5.300000 High
750 POCSO Act 2605.0 3 3.310777 High
748 POCSO - Rape 2022.0 3 5.966667 High
592 KIDNAPPING AND ABDUCTION (Sec.360, 361, 363 to... 1471.0 3 16.416667 High
605 Kidnapping & Abduction of Children 2285.0 3 -4.373665 High
106 ATTEMPT TO MURDER (Sec. 307 IPC/109BNS) - Due ... 1653.0 3 -2.016667 High
591 KIDNAPPING AND ABDUCTION (Sec.360, 361, 363 to... 1638.0 3 -18.233333 High
747 POCSO - Others 833.0 3 -0.016667 High
870 Rape (Sec. 376 IPC) - 1.3 other Cases of Rape 552.0 3 0.234043 High
641 Murder (Sec.302/303 IPC/103,104 BNS) - Due to ... 572.0 3 -1.095238 High
139 Attempt to Commit murder for other reasons 617.0 3 -2.535714 High
840 RAPE (Sec.376, 376(A) to 376 (D) IPC/64 TO 71 ... 303.0 3 3.500000 High
930 THEFT (Sec.379 to 389 IPC/303(2) TO 308(7) BNS... 6634.0 2 -11.533333 High
607 Kidnapping & Abduction of Children - Kidnappin... 344.0 3 0.000000 High
120 ATTEMPT TO MURDER (Sec. 307 IPC/109BNS) - Sudd... 250.0 3 1.733333 High
Saved risk scores to public_safety_risk_scores.csv
PS C:\3rd Year\INT_234\Crime_Analysis_Project>
OUTLINE
TIMELINE
master
Run Testcases
0 0
Live Share
12°C
Cloudy
Search
ENG
IN
11:15 PM
12/17/2025

PROJECT OBJECTIVES AND ANALYSIS

This project was designed with the primary goal of using data analytics and machine learning to support **public safety and informed decision-making**. Each objective focuses on a practical problem related to crime prevention and governance, and the analysis carried out for each objective is summarized below.

Objective 1: Identification of Rising Crime Trends

Objective:
To identify crime categories that show increasing trends over time and may pose emerging threats to public safety.

Analysis:
Monthly crime data from January to September 2025 was analyzed using time-based techniques to observe changes across different crime categories. Trend analysis was conducted to identify categories with consistent growth patterns. Simple linear regression was applied to quantify these trends and measure the rate of increase or decrease. Crime types exhibiting positive and sustained growth were flagged as high-priority areas requiring early intervention. This analysis helps authorities anticipate potential crime escalation and take preventive measures in advance.

Objective 2: Crime Severity Classification

Objective:
To classify crime categories based on their severity and social impact in order to prioritize enforcement and response strategies.

Analysis:
Crime categories were evaluated based on frequency and classification hierarchy to determine their relative severity. A supervised classification approach was used to group crimes into severity levels such as high, medium, and low impact. Decision tree models were employed due to their interpretability, allowing clear understanding of how crime attributes influence severity classification. The resulting severity framework supports structured decision-making for resource allocation and rapid response planning.

Objective 3: High-Crime Risk Identification

Objective:
To distinguish between high-crime and low-crime patterns using historical data and supervised learning techniques.

Analysis:
A binary classification approach was adopted to label crime categories as high-risk or low-

risk based on their incident frequency. Models such as logistic regression and Naive Bayes were applied to learn patterns associated with elevated crime levels. Performance metrics including accuracy, precision, recall, and F1-score were used to evaluate model effectiveness. This analysis enables early identification of crime-prone patterns, supporting proactive policing and crime prevention strategies.

Objective 4: Crime Forecasting and Trend Prediction

Objective:
To predict future crime levels using historical data to assist in planning and resource management.

Analysis:
Regression-based models were used to forecast crime counts based on time-related features and crime category attributes. Multiple linear regression and polynomial regression techniques were applied to capture both linear and non-linear trends. Model performance was evaluated using error-based metrics such as MAE, MSE, RMSE, and R^2. The forecasting results provide valuable insights for manpower planning, patrol scheduling, and surveillance deployment.

Objective 5: Public Safety Risk Assessment

Objective:
To develop a risk assessment framework that identifies crime categories posing the highest threat to public safety.

Analysis:
A composite risk assessment approach was developed by combining crime frequency, trend behavior, and severity levels. Classification techniques such as Support Vector Machines or k-Nearest Neighbors were used to categorize crimes into different risk levels. The resulting risk scores enable policymakers to prioritize interventions and focus on crimes that have the greatest potential impact on society. This objective strengthens the link between analytical findings and actionable governance decisions.

CONCLUSION

This project successfully demonstrates the application of **data analytics and machine learning techniques** for analyzing crime data in a structured and meaningful manner. By transforming raw monthly crime review reports into a unified and analysis-ready dataset, the project provides clear insights into crime trends, severity patterns, risk levels, and potential future behavior related to public safety.

The analysis highlights noticeable variations across crime categories over the observed period, revealing crime types that exhibit increasing trends and require early attention. Severity-based classification and risk assessment further enhance the understanding of which crimes pose greater threats to society, enabling more informed prioritization of law enforcement resources. Predictive modeling and forecasting contribute additional value by supporting proactive planning and preventive policing strategies.

The analytical approach adopted in this project allows complex crime data to be interpreted systematically, combining exploratory analysis with supervised learning techniques. The integration of trend analysis, classification, and forecasting ensures both descriptive and predictive insights, bridging the gap between historical data and future-oriented decision-making.

Overall, this project illustrates how data-driven methodologies can support public safety initiatives and governance by enabling evidence-based decision-making. The analytical framework developed is scalable and can be extended with additional crime data or advanced modeling techniques, making it valuable for academic evaluation as well as practical applications in crime prevention and policy planning.

FUTURE SCOPE

While the current project provides meaningful insights into crime patterns and public safety risks using monthly crime review data, there are several opportunities to further enhance and extend the scope of the analysis in the future.

One potential improvement is the integration of **real-time or regularly updated crime datasets**. By connecting the analytical framework to continuously updated data sources, the system could reflect the most recent crime trends, enabling near real-time monitoring and quicker response from law enforcement agencies.

The project can also be extended by incorporating **advanced predictive and forecasting techniques**. More sophisticated time-series models and deep learning approaches could be applied to improve the accuracy of crime trend predictions and long-term forecasting, helping authorities plan resources more effectively.

Another important area of enhancement is the inclusion of **geographical and spatial data**. By integrating location-based information such as districts, police stations, or zones, the analysis could be expanded to identify geographical crime hotspots and region-specific risk patterns, significantly strengthening preventive policing efforts.

The analytical framework can further benefit from the integration of **external contextual data**, such as population density, socioeconomic indicators, weather conditions, or major events. Combining crime data with these external factors may provide deeper insights into the underlying causes of crime and improve model interpretability.

Additionally, the system can be enhanced by developing **interactive dashboards and decision-support tools** for stakeholders. Features such as drill-down analysis, dynamic filtering, and alert-based reporting would allow decision-makers to explore crime patterns in greater detail and respond more effectively.

Overall, the project establishes a scalable and extensible foundation for crime analytics. With the inclusion of richer datasets, advanced analytical models, and interactive visualization layers, it has the potential to evolve into a comprehensive decision-support system for law enforcement and public safety governance in the future.

REFERENCES

1. Government Crime Review Reports (January–September 2025).
   Available from: Official government or law enforcement public data portals.

2. Kaggle. Public Datasets Repository.
   Available from: https://www.kaggle.com/
   Used for understanding data analysis workflows and reference datasets.

3. Scikit-learn Documentation.
   Available from: https://scikit-learn.org/
   Referenced for supervised machine learning algorithms, model evaluation metrics, and implementation guidelines.

4. Pandas Documentation.
   Available from: https://pandas.pydata.org/
   Used for data manipulation, preprocessing, and dataset integration techniques.

5. Data Analytics and Machine Learning Best Practices.
   Referenced for model selection, evaluation strategies, and ethical use of data in public safety applications.