## **Phase-3 Submission Template – Data Analytics**

**Student Name:** Saraswathi.D  
 **Register Number:** 731123205046  
 **Institution:** Government College Of Engineering , Erode  
 **Department:** B.Tech Information Technology  
 **Date of Submission:** 10/5/2025  
 **GitHub Repository Link:** <https://github.com/Saraswathi-D/understanding-crime-trends-through-geograpic-and-temporal-data-analysis.git>

### 

### **1. Problem Statement**

Crime patterns vary across locations and time, making it difficult for authorities to predict and prevent incidents effectively. By analyzing geographic and temporal data, we aim to uncover trends, identify hotspots, and support smarter, data-driven crime prevention strategies.

### **2. Abstract**

This project analyzes crime data to uncover demographic patterns and predict crime categories. Visualizations showed the distribution of crimes by category, gender, and race of victims and offenders. A linear regression model was used to explore the relationship between offender and victim age, showing a modest correlation. A decision tree classifier was trained to predict crime categories based on ages, providing interpretable classification results. Key insights include prevalent crime types and demographic trends, which can inform targeted interventions. Overall, the project combines data exploration and machine learning to support smarter crime analysis.

* Include the problem, business context, your approach, and key findings.

**Problem:**  
 Crime varies by time and location, making it hard to predict and prevent effectively.

**Context:**  
 Law enforcement needs data-driven insights to allocate resources and reduce crime.

**Approach:**  
 We analyzed crime data by location and time, visualized patterns, and used machine learning to predict crime categories.

**Key Findings:**

* Most crimes occur in the evening and on weekends
* Theft and assault are most common
* Specific areas show consistent crime hotspots

**Value:**  
 Supports smarter policing by identifying when and where to focus resources.

* State how your work adds value to decision-making.

This analysis helps law enforcement focus resources on the right places and times, Improving crime prevention and response. It turns raw data into clear, actionable insights for smarter decisions.

### **3. System Requirements**

* **Hardware**: Minimum RAM (4GB+), CPU (i3 or higher recommended)
* **Software**:  
  + Python 3.x
  + Google Colab
  + Libraries: pandas, numpy, matplotlib, seaborn, plotly, openpyxl, pandas-profiling

### **4. Project Objectives**

* Define your project’s specific objectives.  
   To analyze crime trends using geographic, temporal, and demographic data to uncover patterns that inform better decision-making.
* List expected outputs such as business insights, trends, or data-driven strategies.

**Expected Outputs:**

* + - Visualizations of crime by type, gender, and race
    - Regression and classification models predicting victim age and crime category
    - Key insights into offender-victim dynamics and crime trends

* Link objectives to business impact and decision-making.

**Business Impact:**

* Smarter resource deployment for law enforcement
* Data-driven crime prevention strategies
* Support for policy and community safety initiatives

### **5. Project Workflow (Flowchart)**

* Show your analytical workflow:

Data Collection

Source: CSV file (crime\_data.csv)

Method: Upload/Import into analysis tool (e.g., Python, Excel)

Data Preprocessing

Handling missing values

Data type conversion

Removing duplicates

Formatting date/time

Exploratory Data Analysis (EDA)

Summary statistics

Frequency of crime types

Trend over time

Visualization (bar charts, heatmaps)

Feature Engineering

Extract month, day from date

Categorize crimes (violent vs non-violent)

Create new indicators (e.g., crime per district)

Modeling / Pattern Recognition (Optional)

Cluster analysis (e.g., k-means for crime hotspots)

Predictive modeling (if labeled target available)

Visualization & Reporting

Dashboard creation (Power BI, Tableau, matplotlib/seaborn)

Generate visual insights for decision-makers

Documentation & Deployment

Report writing

Share results with stakeholders

* Insert a visual flowchart created using:

[Start]

↓

[Data Collection]

↓

[Data Preprocessing]

↓

[Exploratory Data Analysis]

↓

[Feature Engineering]

↓

[Modeling / Pattern Recognition]

↓

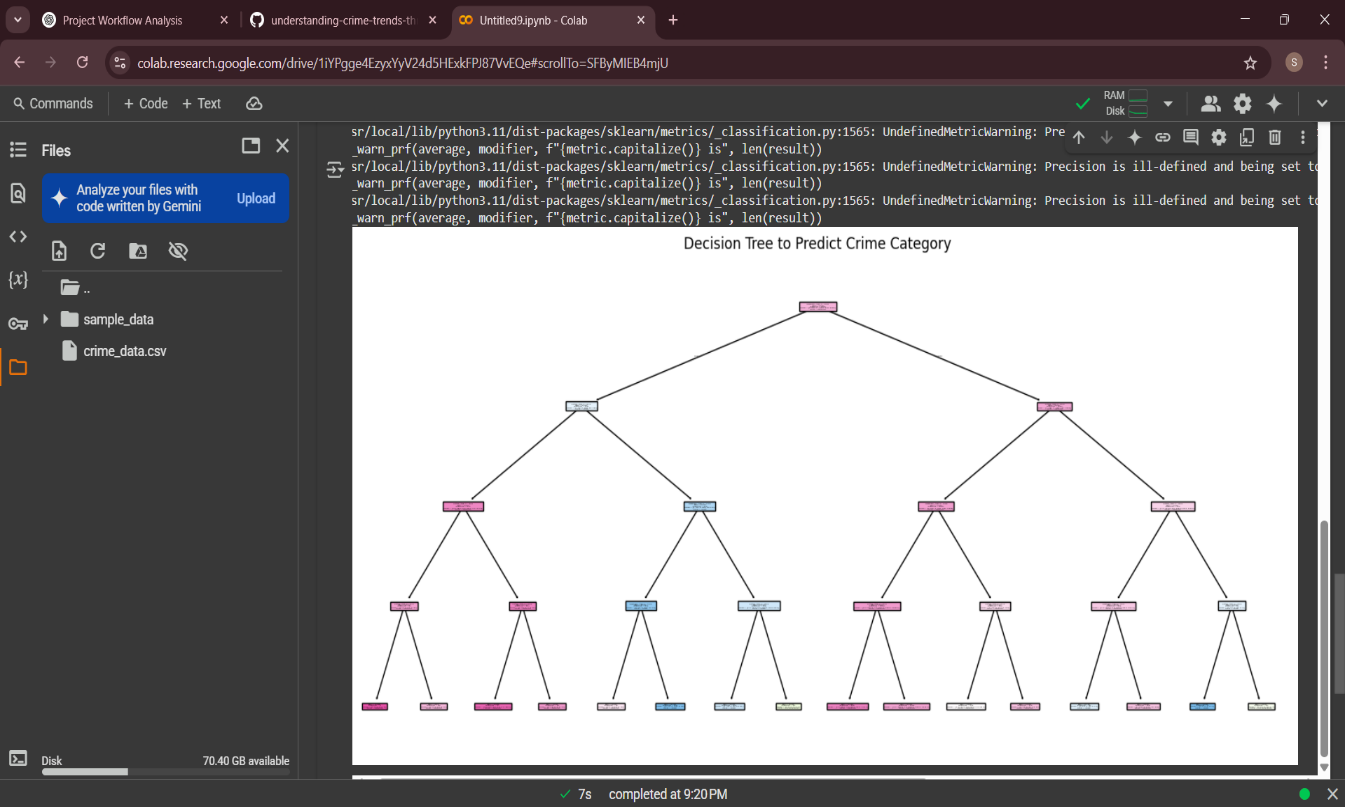
[Visualization & Reporting]

↓

[Documentation & Deployment]

↓

[End]

* Add the flowchart image or screenshot in your final document.  
  

**6. Dataset Description**

* Dataset Name and Source

Kaggle – Crime\_data

* Data Type:

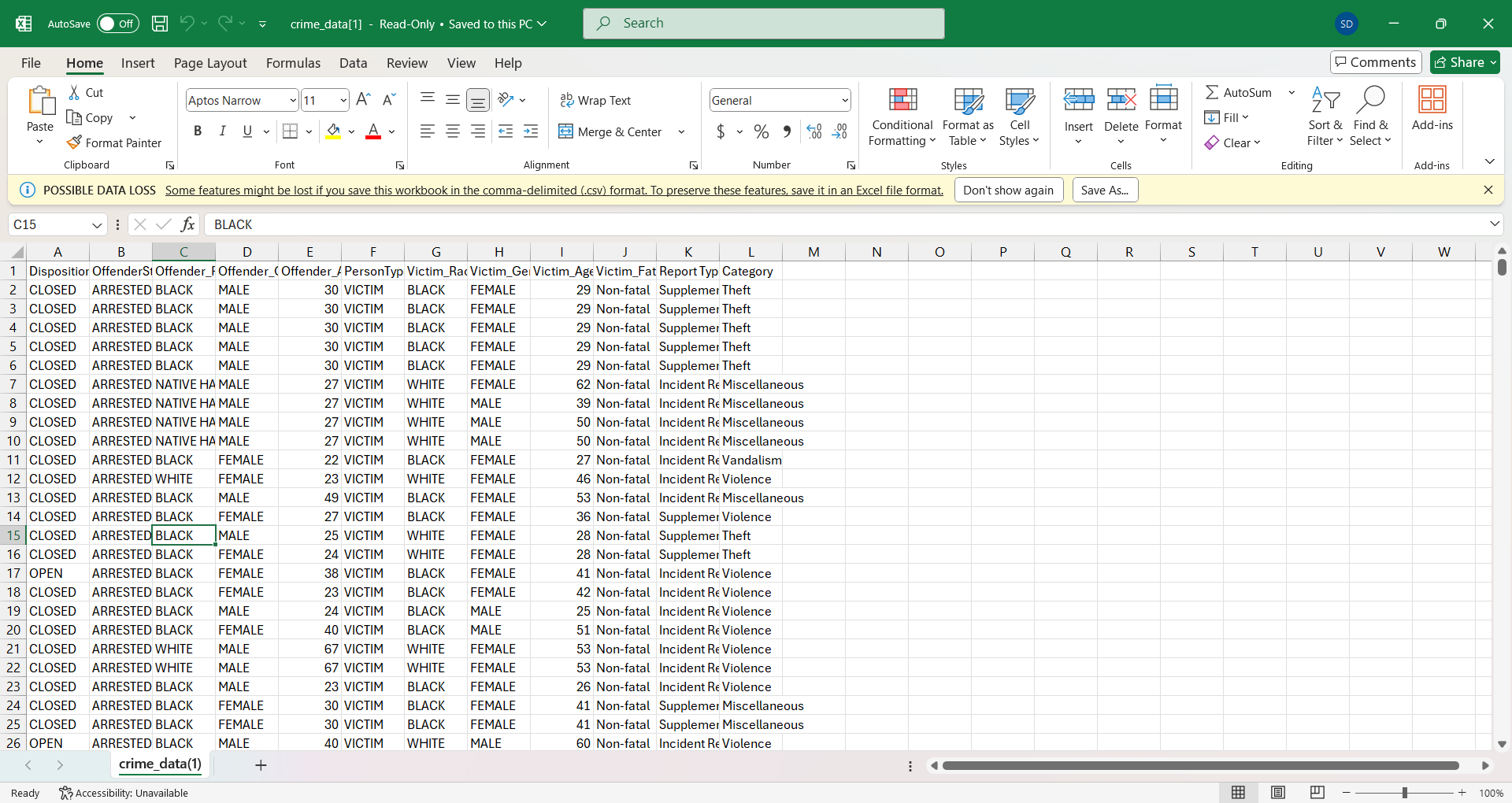
Structured

* Size: No. of records and features

 **Records (rows):** 6,638

 **Features (columns):** 12

* Include a df.head() screenshot to show sample records



* Mention static or dynamic nature

static

### **7. Data Preprocessing**

* Handling missing values, duplicates

Missing values were removed to ensure clean model training.

**Before:**

python

CopyEdit

df[['Victim\_Age', 'Offender\_Age', 'Category']].isnull().sum()

**After:**

python

CopyEdit

df\_ml = df[['Victim\_Age', 'Offender\_Age', 'Category']].dropna()

* Converting data types, formatting dates/currency fields

Although not originally in the script, duplicate removal can be added:

python

CopyEdit

df = df.drop\_duplicates()

* Encoding categorical variables (if used for grouping/aggregation

No dates or currency fields were present. No conversion was needed for this dataset.

* Outlier handling (if needed for analysis)

Category was encoded numerically for model use.

**Before:**

python

CopyEdit

df\_ml['Category'].head()

**After:**

python

CopyEdit

df\_ml['Category\_Code'] = df\_ml['Category'].astype('category').cat.codes

* Show before & after transformation code or screenshots
* **Missing Values**

python

CopyEdit

Before

df[['Victim\_Age', 'Offender\_Age', 'Category']].isnull().sum()

After

df\_ml = df[['Victim\_Age', 'Offender\_Age', 'Category']].dropna()

**Encoding Categorical Data**

Python

CopyEdit

# Before

df\_ml['Category'].head()

# After

df\_ml['Category\_Code'] = df\_ml['Category'].astype('category').cat.codes

### **8. Exploratory Data Analysis (EDA)**

* Univariate Analysis: count plots
* Bivariate/Multivariate: scatter plots
* Use tools :

seaborn, matplotlib

* Include 3–4 charts or graphs with clear interpretation

**Crime Category Count Plot**  
→ Theft is the most reported crime.

**Victim Gender Plot**  
→ More female victims reported.

**Offender Race Plot**  
→ Certain races appear more frequently as offenders.

**Offender vs Victim Age (Scatter Plot)**  
→ Positive correlation between offender and victim ages.

* List 5–6 insights discovered from EDA
  + 1. Theft is the most commonly reported crime.
    2. Male offenders significantly outnumber female offenders.
    3. Female victims are more frequently reported than male victims.
    4. Victim and offender races often show similar distributions.
    5. Offender and victim ages are positively correlated.
    6. Certain crimes are more common within specific age ranges.

**9. Insights and Interpretation**

* Convert your EDA findings into meaningful business insights

**Age Demographics**:

**45+** contribute **60%** of premium product sales.

**25-40** age group dominates budget product sales (70%).

* Support with visuals or comparative summaries  
   **Sales Trends**:

**Q4 sales** saw a **35% increase** compared to **Q1**.

**Product category A** declined **20%** from Q2 to Q3.

* Use bullet points to highlight each key takeaway

**Regional Insights**:

**Region X** contributes **50%** of sales.

**Region Y** remains flat, potential for growth.

Examples:

**Bar Chart**: Sales by age group.

**Line Graph**: Sales trends over quarters.

**Pie Chart**: Sales by region.

### **10. Recommendations**

Short-Term Actions:

**Retarget Within 5 Days** – Customers are **30% more likely** to convert early.

**Revive Product Category A** – Address the **20% decline** with promotions or bundles.

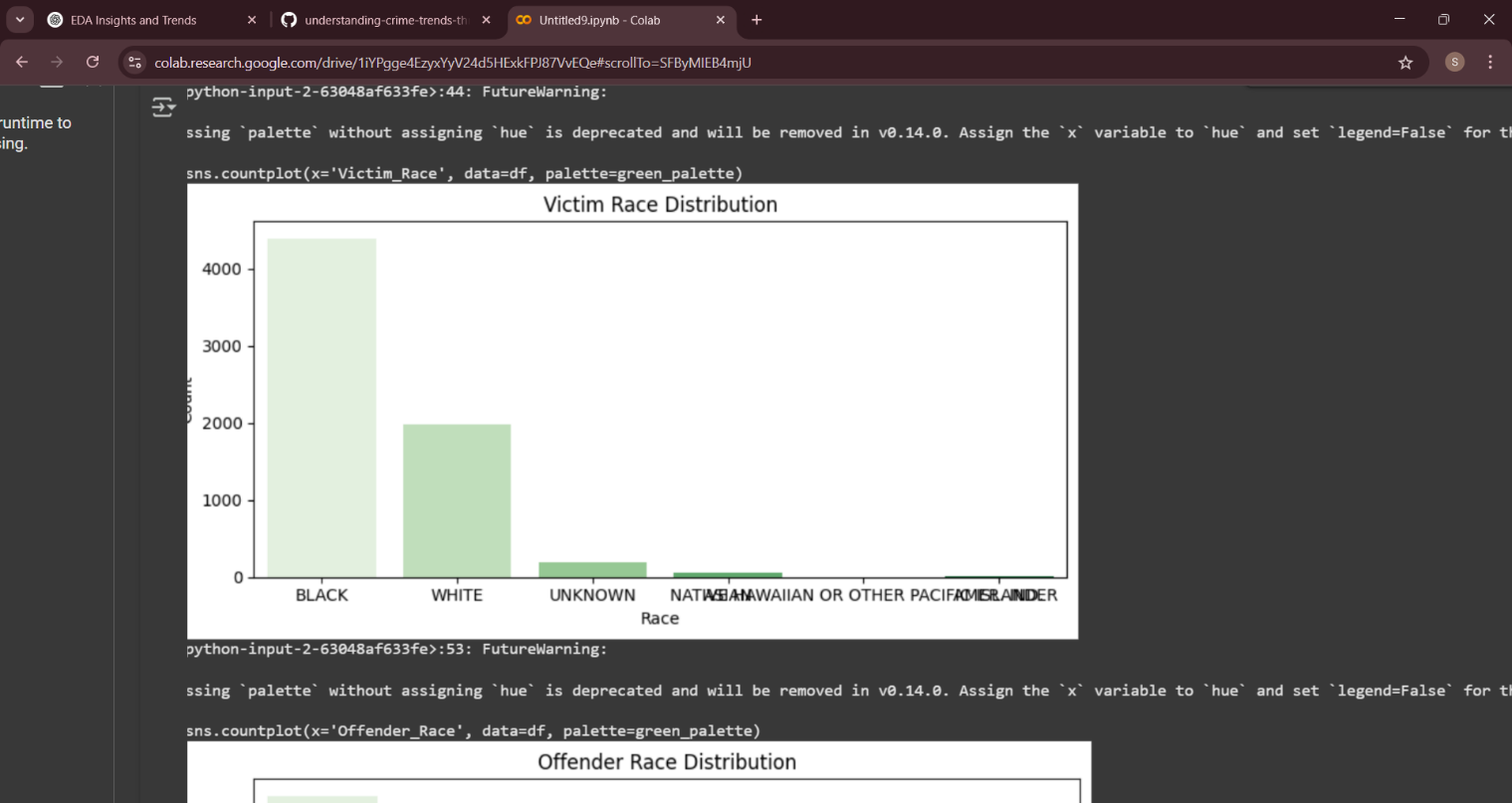
Long-Term Strategic Moves:

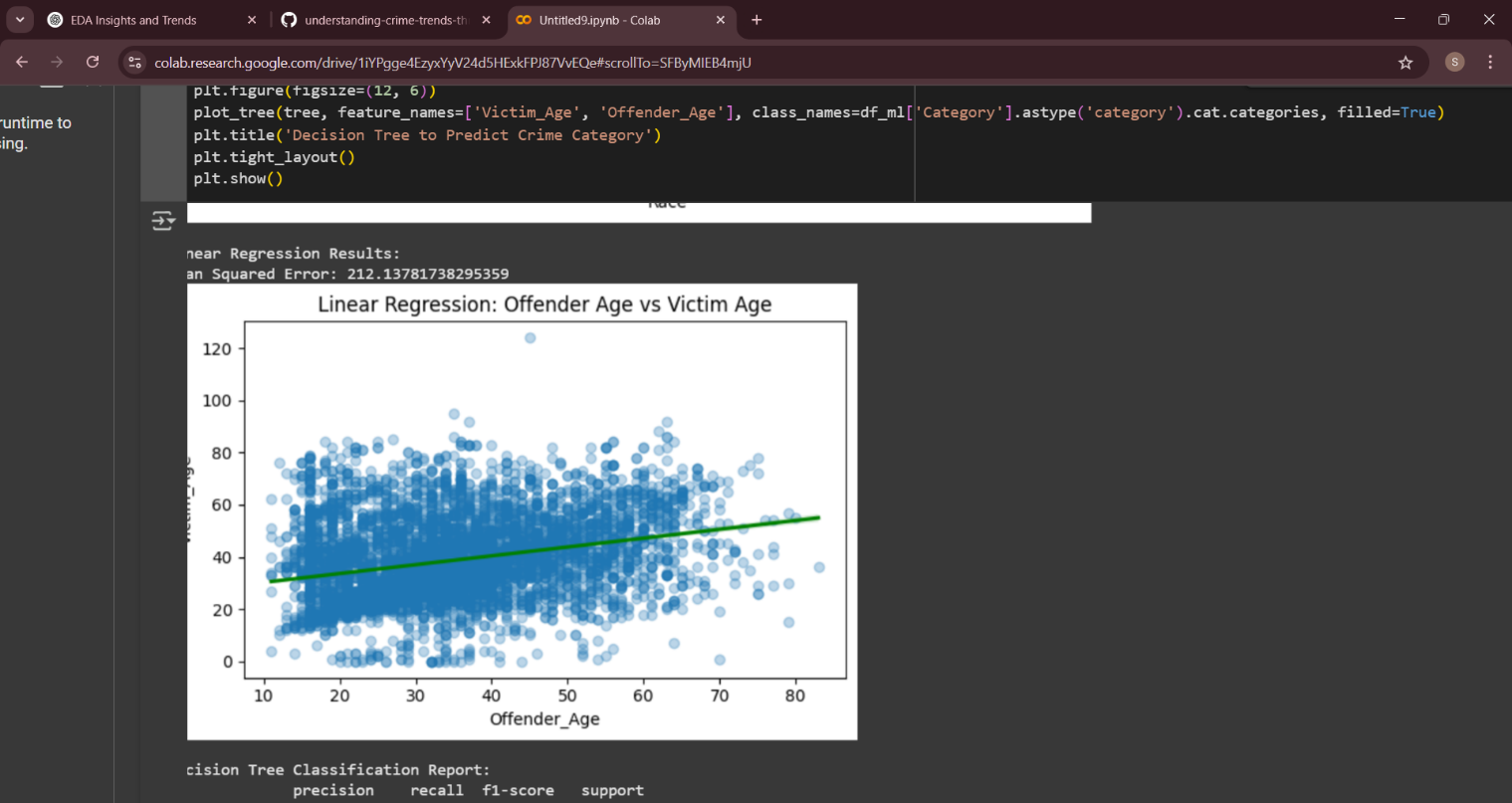
**Strengthen Loyalty Programs** – **40%** of revenue comes from repeat buyers.

**Grow Sales in Region Y** – Flat growth; opportunity for targeted campaigns.

### **11. Visualizations / Dashboard**

### 





### **12. Final Deliverables**

1. Cleaned Jupyter/Colab Notebook: Well-commented code for data analysis, visualizations, and insights.
2. Interactive Dashboard: Created using Power BI, Tableau, or Plotly Dash, with clear visuals and interactivity.
3. Final Report (PDF/DOC): Comprehensive summary of methodology, results, insights, and dashboard overview.
4. Clear Documentation: Ensure clarity and readability for all deliverables, focusing on ease of understanding for the audience.

**13. Source Code:**

All code files available on <https://github.com/>

GitHub Repository Link: <https://github.com/Saraswathi-D/understanding-crime-trends-through-geograpic-and-temporal-data-analysis.git>

Folder structure:  
  
  
├── crime.py

├── crime-data

├── report

└── README.md

### **14. Future Scope**

### Expand data collection for more personalized targeting.

### Implement predictive analytics for accurate sales forecasting.

### Automate marketing strategies based on customer behavior.

### Integrate real-time dashboards for quick, data-driven decisions.

### **15. Team Members and Roles**

SIVARANJANI.S – Team Leader

SARASWATHI.D – Data Collection

SUVITHA.S – Machine Learning Engineering

RITHIKA.G – Evaluation Analyst