

CRIME INCIDENT ANALYSIS

REPORT

Dataset Name: Individual_Incident_2020.csv

Name: Alamaara Kazi

INDEX

Sr. no	Title
1	Introduction
2	Problem Statement
3	Dataset Description
4	Tools and Technologies
5	Exploratory Data Analysis
6	SQL Analysis
7	Excel Analysis
8	Power BI Analysis
9	Observation and Conclusion

1. INTRODUCTION

This project is based on the **Individual_Incident_2020.csv** dataset, which contains crime incident records for the year 2020. Each row in the dataset represents one reported crime incident. The dataset includes important details such as state, incident_number, total_offense, total_victim, total_offender, gun_involvement, and property_value.

Each row in the dataset represents one incident. The data includes information such as state, incident_number, total_offense, total_victim, total_offender, gun_involvement, and property_value.

The main goal of this project is to understand crime patterns, victim distribution, offender details and property loss. The analysis was completed using Python, SQL, Excel and Power BI.

2. PROBLEM STATEMENT

The Individual_Incident_2020 dataset contains crime incident records with 53 columns and multiple data types such as string and float. Since the dataset includes many offense, victim, and offender details, it cannot be directly used for analysis without preparation.

Several columns contain missing values, which can affect the accuracy of results. Some numerical columns required proper formatting and data type conversion before calculations could be performed.

The main objective of this project is to clean the raw dataset, structure it properly, and generate meaningful insights about crime patterns, victim details, offender statistics, and financial impact in 2020.

3. DATASET DESCRIPTION

The Individual_Incident_2020 dataset contains crime incident records reported during the year 2020. Each row in the dataset represents one individual incident. The dataset includes 53 columns that store different types of information related to crime, victims, offenders, and property details.

The dataset contains both categorical and numerical columns.

- **Basic Incident Information**

Some columns provide general details about each incident:

- **State:** shows the state where the incident occurred.
- **ID:** unique record identifier.
- **ORI:** agency reporting code.
- **incident_number:** unique number assigned to each incident.
- **date_HRF and date_SIF:** date-related information.
- **hour:** time of incident.

These columns help in identifying when and where the incident happened.

The table below shows how the dataset is divided into different sections based on the type of information stored in each column.

Category	Important Columns
Basic Info	state, incident_number, date_SIF, hour
Offense Data	total_offense, theft_offense, violence_offense, drug_offense, sex_offense
Victim Data	total_victim, male_victim, female_victim, minor_victim
Offender Data	total_offender, male_offender, female_offender, offender_not_known
Weapon & Drug	gun_involvement, drug_involvement
Property	property_value, stolen_motor

4. TOOLS & TECHNOLOGIES

The following tools were used in the task:

- A. **Python (Pandas, NumPy)**: Data loading and cleaning
- B. **Matplotlib and Seaborn**: Data visualization
- C. **MySQL Workbench**: SQL based analysis
- D. **Microsoft Excel**: Pivot tables and summary charts
- E. **Power BI Desktop**: Interactive dashboard creation

5. EXPLORATORY DATA ANALYSIS (EDA)

The dataset was examined using Python in Jupyter Lab to understand important patterns in crime incidents. Different visual charts were created to study state-wise incidents, victim age groups, and the connection between theft offenses and property loss.

- **Stages Performed during EDA:**

- 1) Imported important libraries such as **Pandas**, **NumPy**, **Matplotlib**, and **Seaborn**.
- 2) Loaded the CSV file into a Pandas DataFrame (df).
- 3) Used df.shape to check the total number of rows and columns.
- 4) Used df.dtypes to check data types of all columns.
- 5) Checked for missing values in each column.
- 6) Verified whether duplicate rows were present in the dataset.

- **Purpose:**

These steps were done to understand:

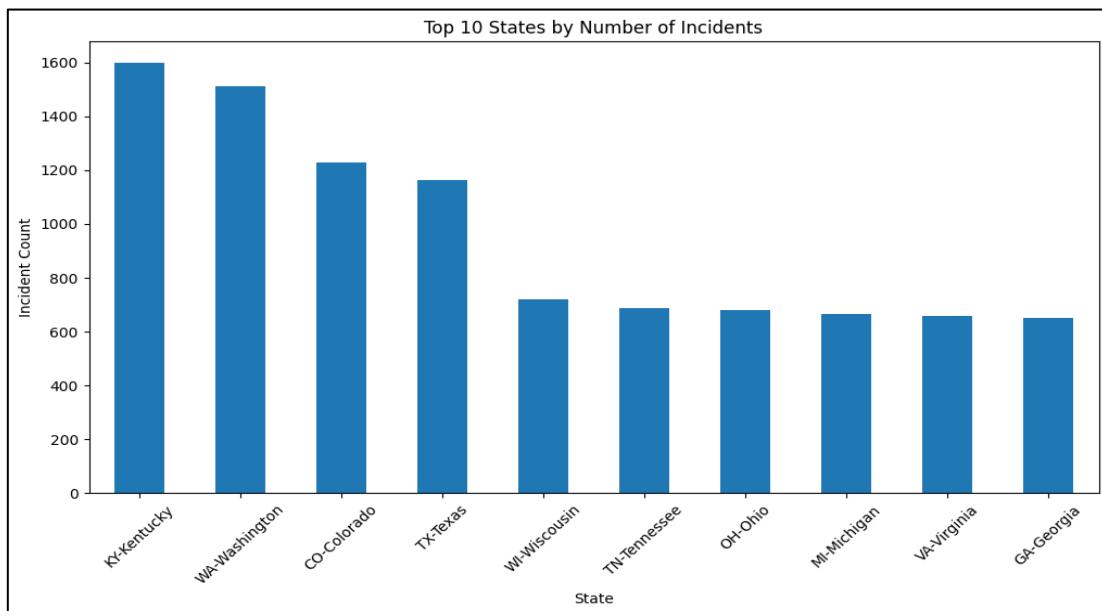
- How large the dataset is
- What type of information is available
- Whether columns are stored correctly (numeric or object type)

Before performing analysis, it is important to know what kind of data we are working with.

5.1 VISUALIZATIONS PERFORMED

After cleaning and preparing the dataset, visual analysis performed using Matplotlib and Seaborn.

I. Top 10 States by Number of Incidents

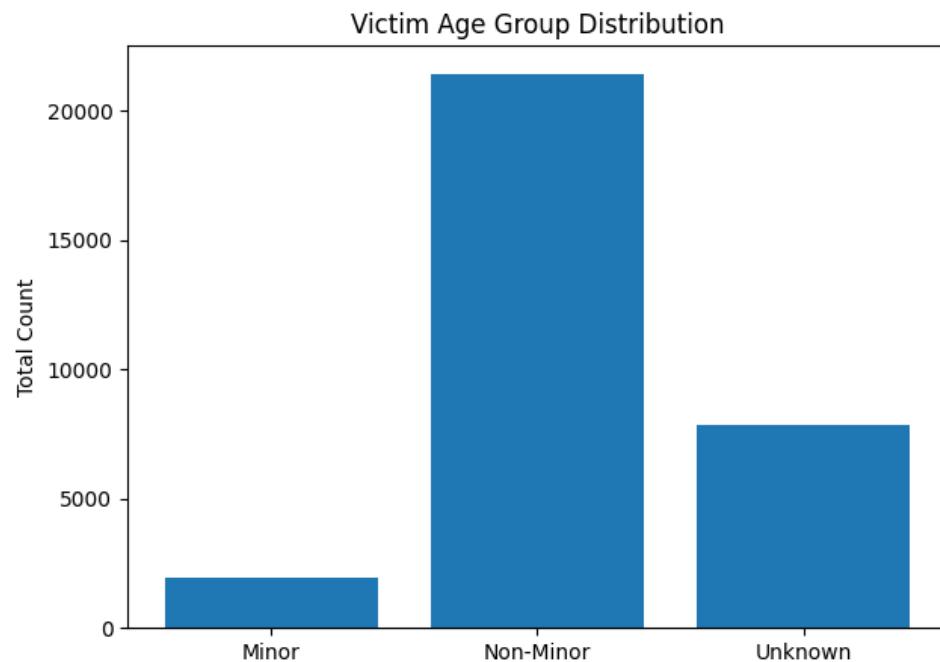


- **A bar chart was used because:**
 - It clearly compares categories.
 - It makes ranking easy to understand
- **What the chart shows:**
 - has the highest number of incidents among the top states.
 - Washington and Colorado also show high incident counts.
 - There is a noticeable gap between the top states and the lower-ranked states.
 - Crime incidents are concentrated in certain states rather than evenly spread.

- **What we understand:**

This chart clearly shows that some states experience more crime cases than others. It helps identify high-incident regions and understand geographic distribution of crime.

II. Victim Age Group Distribution



- **Why this visualization:**

To understand which age group is more affected by crime, three columns were analyzed:

- minor_victim
- non_minor_victim
- unknown_age_victim

A bar chart was selected to compare these groups clearly.

- **What the chart shows:**

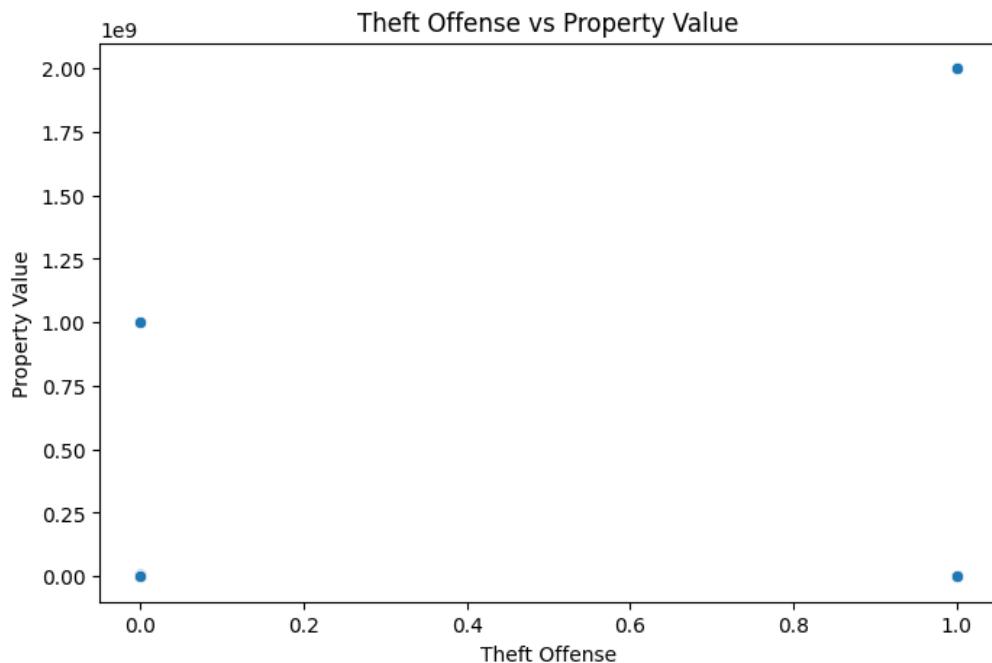
- Non-minor victims have the highest count.
- Minor victims are significantly lower.
- There is a noticeable portion of victims with unknown age.

- **What we understand:**

This shows that adults are more commonly affected compared to minors. It also indicates that some data records do not contain complete age information.

Understanding age distribution is important for social and safety analysis.

III. Theft Offense vs Property Value



- **Why this visualization:**

The column `theft_offense` indicates whether theft was involved in the incident.
The column `property_value` shows the financial loss related to the case.
A scatter plot was created to study the relationship between theft and property loss.

- **What the chart shows:**

- When theft offense is present, property value tends to be higher.
- Some theft cases show extremely large financial damage.
- When theft is not involved, property loss is generally lower.

- **What we understand:**

There is a clear connection between theft and property damage. Theft-related crimes contribute significantly to financial loss.

This visualization helps in understanding the economic impact of theft offenses.

6. SQL ANALYSIS

The cleaned dataset was imported into MySQL Workbench and stored in the database **Individual_Incident_DB**. The table used for analysis was **incident_table**. SQL was used to summarize data, calculate totals, and identify important crime patterns.

- **SQL Operations Performed**

- Counted total incidents using COUNT(ID) to understand the size of the dataset.
- Grouped data by state to find which states report the highest number of incidents.
- Calculated total_offense, total_victim, and total_offender using SUM() to measure overall crime impact.
- Compared offense categories such as theft_offense, violence_offense, and drug_offense to identify the most frequent crime type.
- Measured gun_involvement and drug_involvement cases to understand weapon and drug-related crimes.
- Calculated total_property_value and stolen_motor cases to analyze financial loss.
- Studied victim distribution by gender and age using male_victim, female_victim, minor_victim, and non_minor_victim.
- Examined offender relationships using offender_wi_family, offender_outside_family, and offender_not_known.
- Analyzed completed and attempted crimes using completed_attempted2.
- Identified peak crime hours using hour column.
- Ranked states based on property loss and gun-related incidents using DENSE_RANK().
- Calculated the percentage of sex_offense out of total_offense.

- **Purpose of SQL Analysis**

The main purpose of SQL analysis was to organize large data efficiently, SQL helped convert raw data into structured and meaningful information for further analysis.

7. EXCEL ANALYSIS

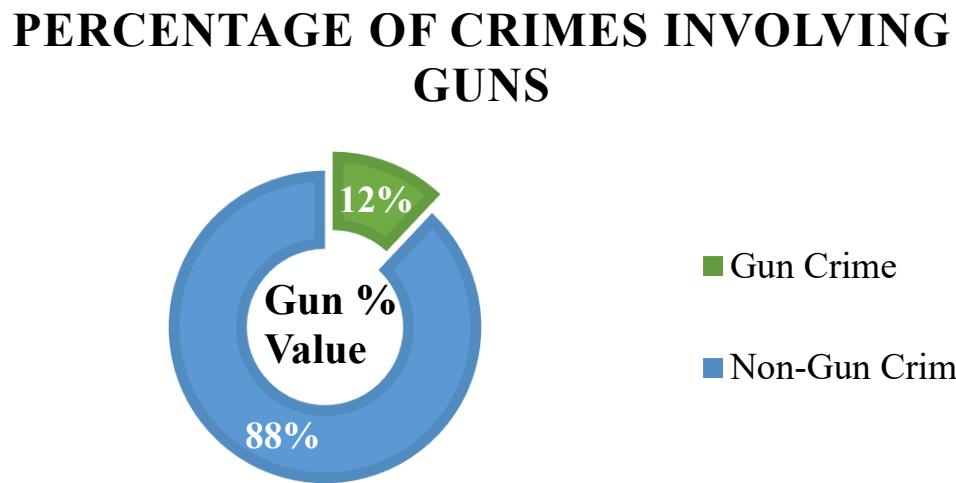
The cleaned dataset was imported into Microsoft Excel for summary and comparison using Pivot Tables and charts. Excel was mainly used to quickly organize large data and identify patterns in crime incidents.

- **Purpose of Using Excel**

- To summarize large data easily using Pivot Tables
- To compare crime categories and percentages
- To calculate totals and percentages
- To create simple visual charts for better understanding.

7.1 Visualization Performed

i. Percentage of Crimes Involving Guns (Donut Chart)



Why this visual was created:

- To understand how many incidents involve gun_involvement.
- To compare gun crimes with non-gun crimes.
- To show crime percentage in a simple visual way.

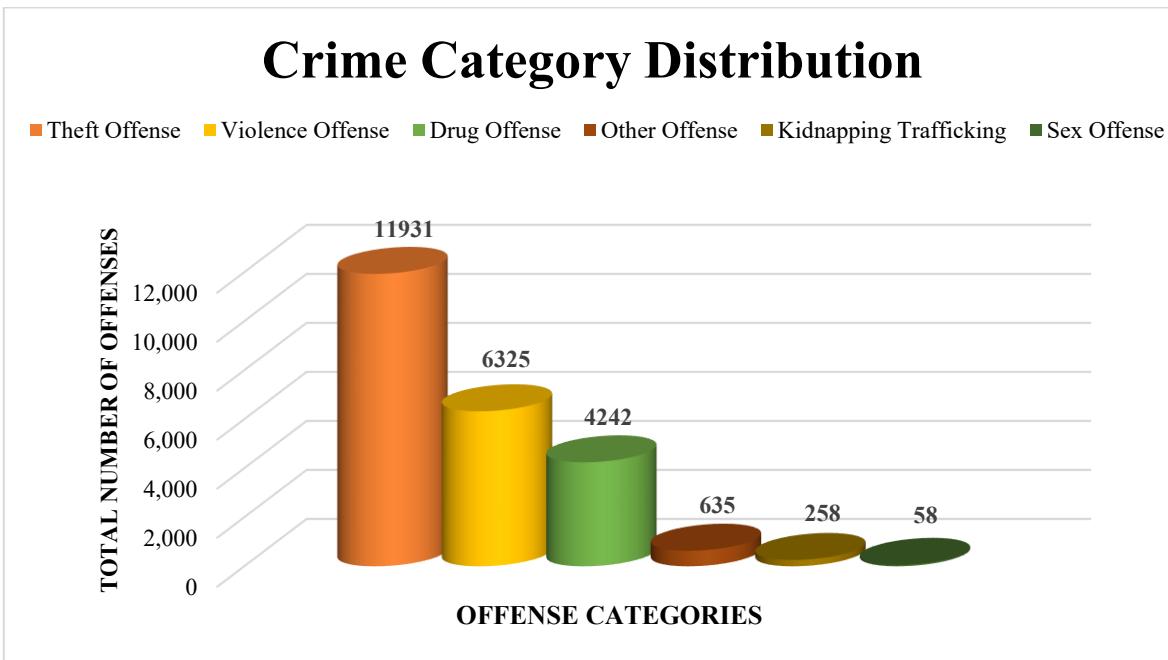
What this visual shows:

- Around 12% of incidents involve guns.
- Around 88% of incidents do not involve guns.

What we understand:

- Gun crimes are smaller in number compared to non-gun crimes.
- Even though percentage is low, gun crimes may be more serious.
- Most crime cases are not weapon-based.

ii. Crime Category Distribution (Column Chart)



Why this visual was created:

- To compare different offense categories.
- To identify which offense type is highest.
- To understand overall crime pattern.

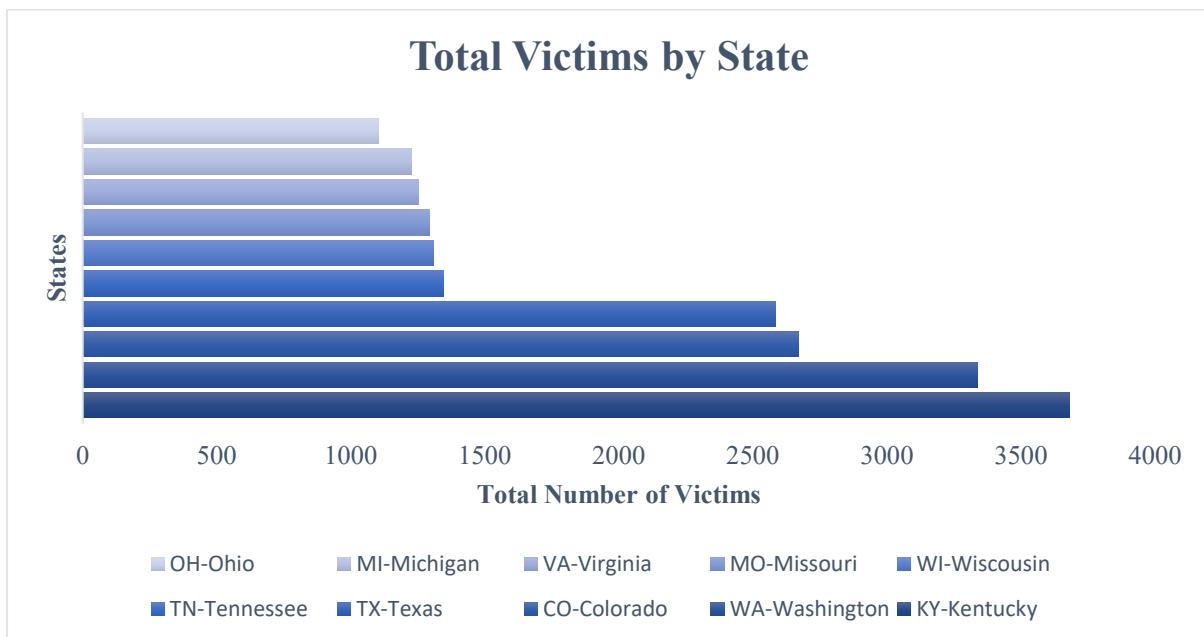
What this visual shows:

- theft_offense has the highest number of cases.
- violence_offense is second highest.
- drug_offense comes after that.
- kidnapping_trafficking and sex_offense are much lower.

What we understand:

- Property-related crimes are more common.
- Serious crimes exist but are fewer in number.
- Crime distribution is not equal across categories.

iii. Total Victims by State (Horizontal Bar Chart)



Why this visual was created:

- To compare total_victim numbers across states.
- To identify which states have more victims.
- To study geographical impact of crime.

What this visual shows:

- Some states have higher victim counts.
- Kentucky and Washington are among the top states.
- Other states have moderate or lower counts.

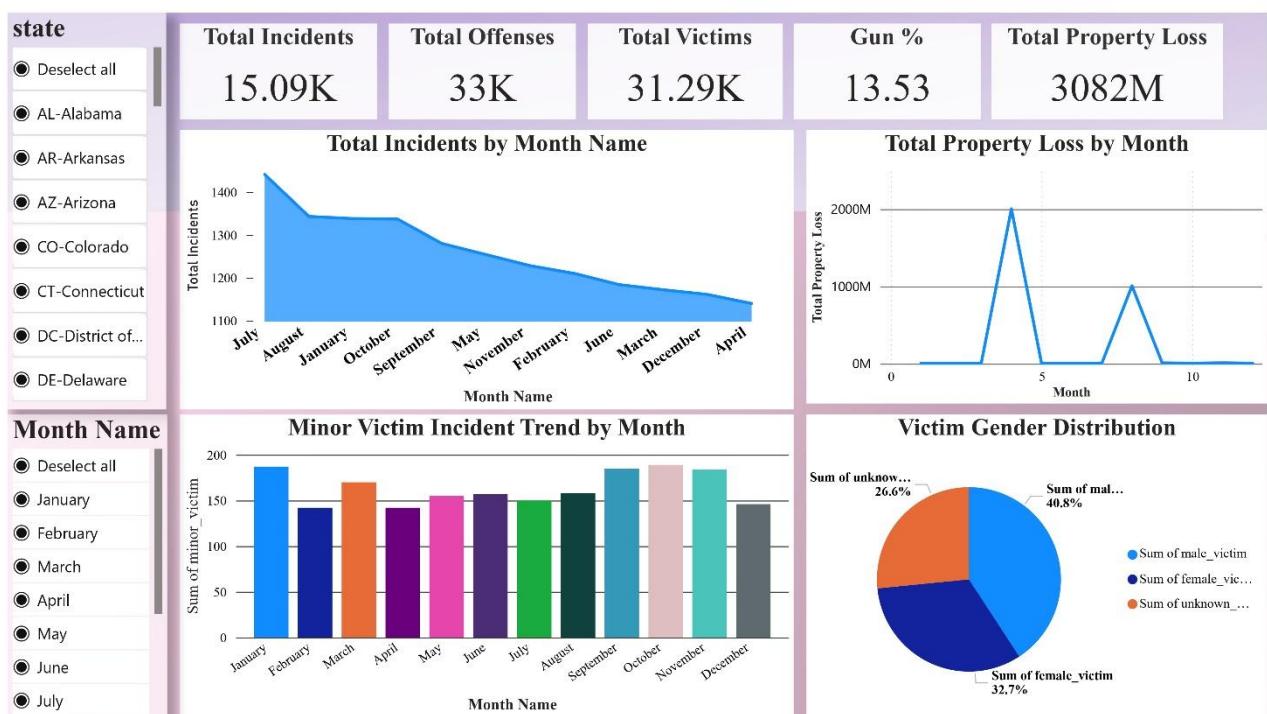
What we understand:

- Crime impact differs by state.
- Certain states need more attention and control measures.
- Victim distribution is not uniform across regions.

8. POWER BI ANALYSIS

The cleaned dataset was imported into Power BI Desktop to create an interactive dashboard. The purpose of using Power BI was to present crime data in a clear and visual format so that trends and patterns can be easily understood.

The dashboard includes summary cards, charts, and filters to analyse incidents, victims, property loss, and crime distribution.



1) Overall Crime Summary:

This section shows the main numbers of the dataset in card format:

- Total Incidents
- Total Offenses
- Total Victims
- Gun Percentage
- Total Property Loss

Explanation:

These summary cards give a quick overview of the crime situation. It helps understand the total scale of incidents, victim impact, gun involvement rate, and financial loss at one glance

2) Total Incident by Month Name

This is a line/area chart that shows total incidents for each month.

Explanation:

It shows how crime cases change throughout the year. Some months have higher incidents while others have lower counts. This helps identify seasonal patterns in crime activity.

3) Total Property Loss by Month Name

This chart displays total property_value lost in each month.

Explanation:

It helps understand how financial damage varies month by month. Certain months show very high property loss, which may be linked to theft-related incidents.

4) Minor Victim Trend by Month

This is a column chart showing minor_victim cases for each month.

Explanation:

It helps analyse how incidents involving minors change across months. It shows whether certain months have higher impact on children and teenagers.

5) Victim Gender Distribution

This is a pie chart showing:

- male victim
- female victim
- unknown sex victim

Explanation:

It shows the percentage distribution of victims by gender. It helps understand which gender is more affected and whether there are missing gender records.

6) State and Month Filters (Slicers)

These are interactive filters that allow selection of:

- State
- Month

Explanation:

These filters help users analyse specific states or months. The dashboard updates automatically based on selected values, making the analysis more interactive.

The dashboard combines summary numbers, trends, and comparisons in one place.

It helps understand:

- Total crime impact
- Monthly patterns
- Financial loss
- Minor involvement
- Gender distribution

All charts together give a complete overview of crime data in a simple visual format.

9. OBSERVATION AND CONCLUSION

A Overall Observation

After analysing the dataset using Python, SQL, Excel, and Power BI, several important patterns were identified:

The total number of incidents and victims is high, showing a significant crime impact.

Theft offenses are the most common crime category compared to other offenses.

Around 13% of incidents involve guns, while the majority are non-gun crimes.

Property loss is substantial, indicating strong financial impact of crime.

Crime incidents and property damage vary across different months.

Some states report higher incidents and victim counts than others.

Male victims are slightly higher compared to female victims.

Non-minor (adult) victims are more common than minor victims.

Overall, crime distribution is not equal across states or months, and certain categories such as theft contribute more to total incidents.

B Conclusion

This project demonstrates a complete data analysis process using multiple tools.

The dataset was cleaned, structured, and analysed to convert raw data into meaningful insights.

The analysis shows clear patterns in crime types, victim distribution, gun involvement, and property loss.

Using Python helped in data cleaning and visualization. SQL helped in summarizing and ranking data efficiently. Excel helped in quick comparisons using pivot tables. Power BI combined all findings into an interactive dashboard.

The overall study provides a clear understanding of crime patterns and their impact. These insights can help in better planning, monitoring, and decision-making related to crime control.