CRIME INCIDENT ANALYSIS REPORT

Database : Individual_Incident_2020

Name-Urfa Ilyas Bhattiwale

Date- 12-02-2026

CONTENTS

INTRODUCTION

1.1 Introduction:

In the modern digital era, data analysis plays an important role in identifying patterns and supporting better decision-making. This project analyzes the Individual_Incident_2020 dataset to study crime trends, offense categories, victim distribution, and weapon involvement across various states

.Different analytical tools such as **Python, SQL, Excel, and Power BI** were used to perform detailed data exploration and visualization. Python was applied for data cleaning and exploratory analysis. SQL was used to organize the dataset in a database and execute advanced queries for deeper insights. Excel helped in summarizing the data using pivot tables and creating interactive dashboards. Power BI was used to design a structured and professional dashboard with key performance indicators and visual comparisons.

By combining data preparation, statistical analysis, and visual reporting, this project converts raw crime data into clear and meaningful insights. The main aim is to identify important crime patterns, compare state-level variations, analyze offense categories, and present the findings in an interactive and easy-to-understand format.

1.2 Objective :

The objective of this project is to analyze the Individual_Incident_2020 crime dataset using Python, SQL, Excel, and Power BI to identify important crime patterns, understand state-wise trends, and present meaningful insights through clear visualizations and interactive dashboards.

1.3 Problem Statement:

Crime incident data contains large amounts of raw and unstructured information, making it difficult to identify meaningful patterns and trends. There is a need to clean, analyze, and visualize the data effectively to understand state-wise crime distribution, offense categories, and victim patterns for better insights and decision-making

DATASET DESCRIPTION

The dataset used in this project is the Individual_Incident_2020 crime dataset. It contains detailed information about crime incidents recorded across different states for the year 2020.

The dataset consists of **2191 records** and **53 columns**, including information related to:

- State name
- Incident number and identifiers
- Date and time of occurrence
- Total offenses
- Total victims and offenders
- Types of offenses (violent, theft, drug, sex, kidnapping, etc.)
- Gun and drug involvement
- Property value associated with incidents
- Victim race distribution
- Stolen property details

The dataset includes both numerical and categorical variables, making it suitable for statistical analysis and visualization. It provides sufficient information to analyze crime trends, state-wise distribution, offense categories, and victim patterns.

TOOLS AND TECHNOLOGIES

The following tools and technologies were used to complete this project:

- **Python (Jupyter Notebook – Anaconda)** – Python was used for data cleaning, preprocessing, and Exploratory Data Analysis (EDA). Jupyter Notebook, installed through Anaconda, was used as the development environment. Libraries such as Pandas, NumPy, Matplotlib, and Seaborn were used for data manipulation and visualization.
- **MySQL (MySQL Workbench)** – Used to store the dataset in a structured database format and perform queries, aggregations, joins, window functions, and advanced analysis.
- **Microsoft Excel** – Used to create Pivot Tables, apply formulas, and design an interactive dashboard for summarizing and presenting insights.
- **Power BI** – Used to develop a professional and interactive dashboard with DAX measures, KPIs, slicers, and dynamic visualizations for better reporting.

These tools collectively helped transform raw crime data into meaningful insights and interactive analytical reports.

PYTHON – EXPLORATORY DATA ANALYSIS (EDA)

Python was used to examine and analyze the Individual_Incident_2020 dataset in order to gain a clear understanding of crime-related information. The purpose of Exploratory Data Analysis (EDA) was to study the dataset's structure, detect patterns, identify irregularities, and explore relationships among different variables**.**

Data Examination

The dataset was first reviewed to understand its overall structure and characteristics.

- Imported the dataset using Pandas.
- Checked the total number of rows and columns.
- Reviewed column names and data types.
- Generated descriptive statistics for numerical variables.
- Identified missing and duplicate records.

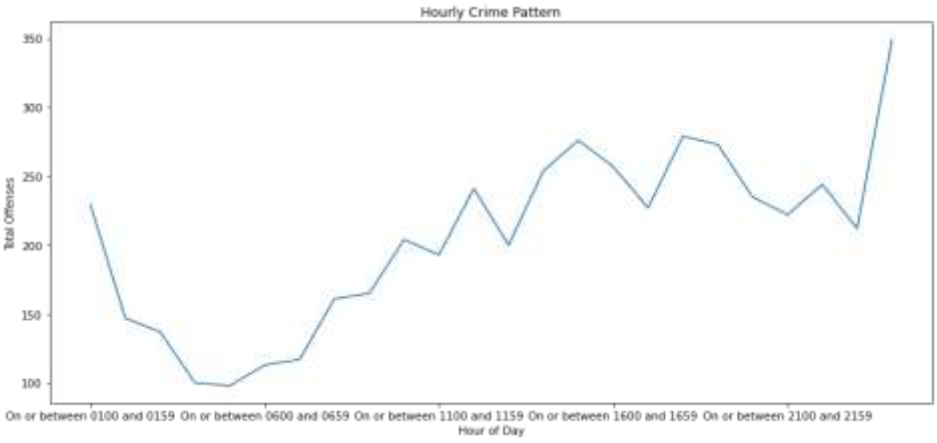
This step helped in understanding the dataset before performing deeper analysis.

Exploratory Data Analysis (EDA)

EDA was performed to examine crime trends, victim distribution, and relationships between key variables in the dataset.

1.Hourly Crime Pattern (Line Graph)

Purpose:
This graph illustrates how the number of crime incidents changes at different hours throughout the day.



Key Observations:

- Lowest crime levels occur between 1 AM and 6 AM.
- Incidents gradually increase during daytime.
- Peak crime activity is observed around 9 PM–10 PM.

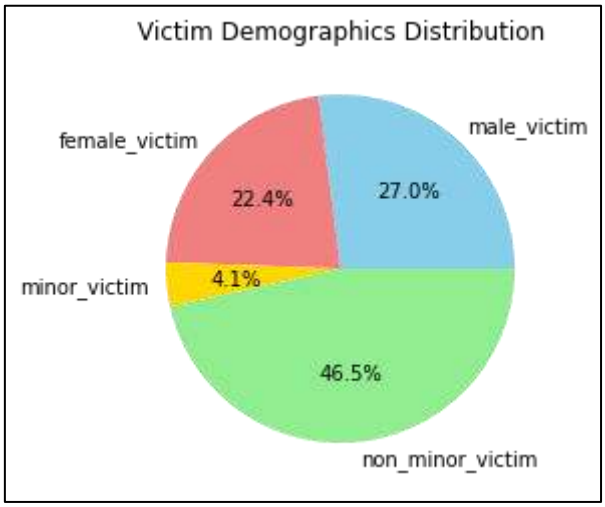
Interpretation:

Evening hours show higher crime frequency, suggesting the need for increased monitoring during this period.

2.Victim Demographics Distribution (Pie Chart)

Purpose:

This chart represents the percentage distribution of victims across different demographic categories.



Distribution:

- Male victims account for 27% of the total.
- Female victims represent 22.4%.
- Minor victims make up 4.1%.
- Non-minor (adult) victims comprise the largest share at 46.5%.

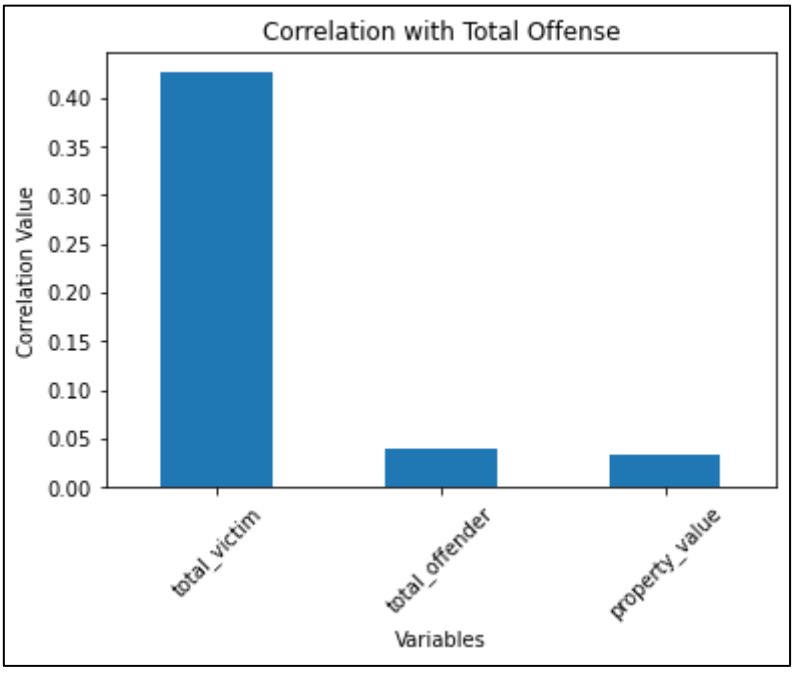
Interpretation:

Adults form the largest victim group, while minors represent a smaller portion. Preventive strategies may focus more on adult populations.

3.Correlation Analysis (Bar Chart):

Purpose:

The correlation matrix was created to analyze the relationship between key numerical variables, including total_offense, total_victim, total_offender, and property_value. A bar chart was used to visually compare the strength of these relationships.



Key Observations:

- Total offense & total victim: Moderate positive correlation (0.43).
- Total offense & total offender: Very weak correlation (0.04).
- Total offense & property value: Very weak correlation (0.03).

Interpretation:

Victim count shows the strongest relationship with total offenses, making it more relevant for understanding crime trends.

SQL ANALYSIS

SQL was used to organize and examine the Individual_Incident_2020 dataset within a relational database system. The goal was to structure the data efficiently and extract useful insights through query-based analysis.

Database Preparation

- Created a database and imported the dataset into a table.
- Verified column names and data types.
- Ensured data consistency before performing analysis.

Data Retrieval and Filtering

- Used SELECT statements to retrieve required fields.
- Applied WHERE conditions to filter records based on state and offense type.
- Sorted results using ORDER BY.
- Limited output using LIMIT for quick review.

Summary and Aggregation

- Calculated totals using SUM() for offenses, victims, and property value.
- Counted records using COUNT().
- Found average values using AVG().
- Performed state-wise analysis using GROUP BY.
- Filtered grouped results using HAVING.

Advanced Analysis

- Used CASE expressions to categorize crime data.
- Applied subqueries and CTEs for structured comparisons.
- Implemented window functions like RANK() and LAG() to analyze rankings and trends.
- Calculated percentages and ratios using arithmetic operations.

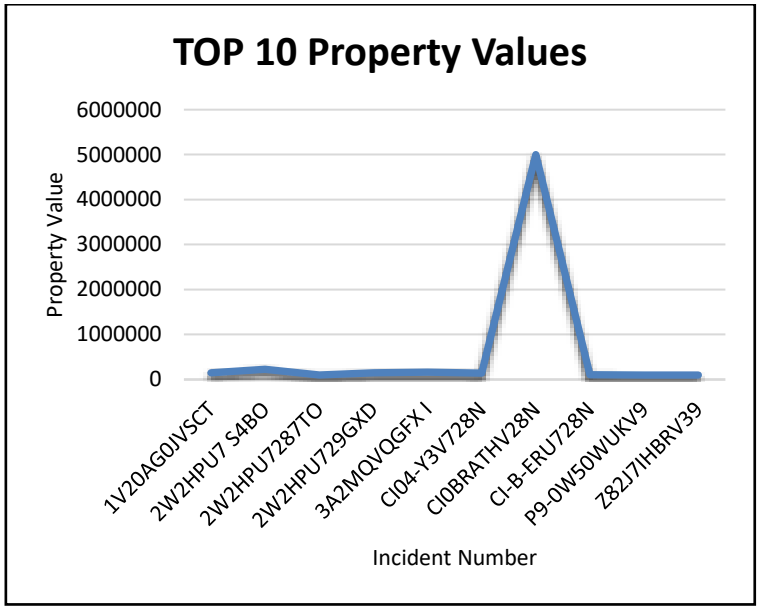
EXCEL ANALYSIS

The Pivot Table charts created in Excel provide insights into property impact, victim distribution, and crime concentration across states. Each visualization highlights a different aspect of the dataset.

1. Top 10 Property Values (Line Chart)

Purpose:

To analyze property values associated with the top 10 recorded incidents.



Key Observations:

- Most incidents report property values below 1,000,000.
- One incident (CI0BRATHY28N) shows a significantly higher value, close to 5,000,000.

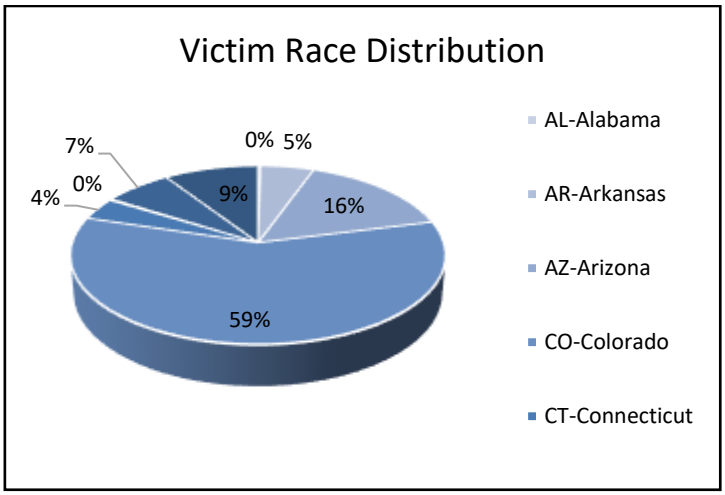
Interpretation:

The unusually high value represents a potential outlier. It may indicate a major financial loss event or require further validation to rule out data inconsistencies.

2. Victim Race Distribution (3D Pie Chart)

Purpose:

To examine the distribution of victims across different states.



Key Observations:

- Arizona accounts for the largest share (59%).
- Arkansas (16%) and Delaware (9%) contribute notable portions.
- Some states such as Connecticut and Georgia show minimal or no recorded values.

Interpretation:

The distribution appears concentrated in Arizona. This pattern may reflect regional demographics, reporting differences, or localized crime patterns.

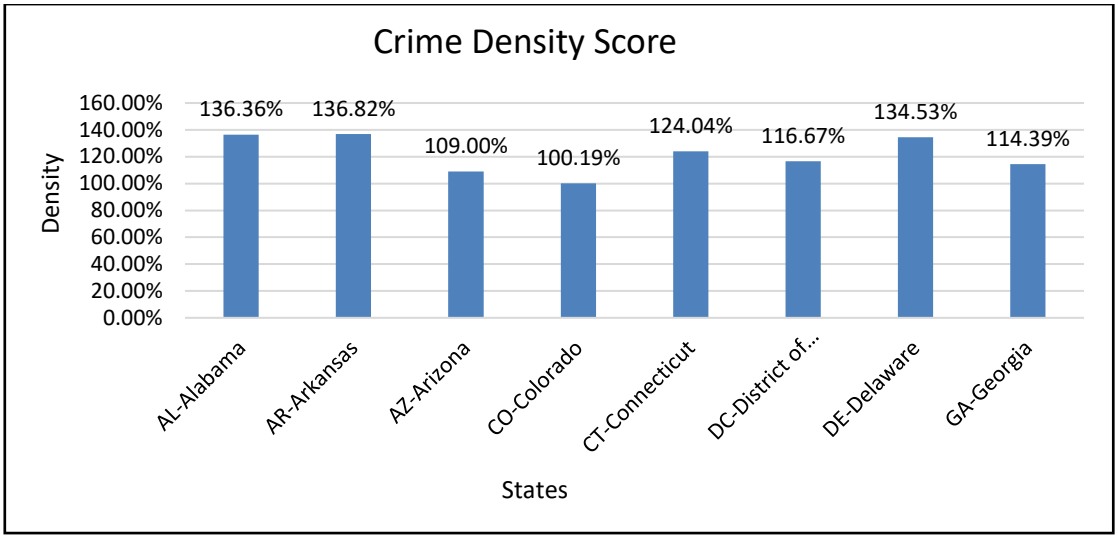
3. Crime Density Score (Bar Chart)

Purpose:

To compare crime intensity levels across states.

Key Observations:

- Arkansas (136.82%) and Alabama (136.36%) show the highest crime density scores.
- Colorado (100.19%) and Arizona (109%) are closer to the baseline.



Interpretation:

Higher scores indicate stronger crime concentration relative to the average. Arkansas and Alabama demonstrate comparatively higher crime intensity, while Colorado remains near the standard level.

POWER BI DASHBOARD

Power BI was utilized to develop an interactive dashboard for analyzing the Individual_Incident_2020 dataset. The objective was to convert raw data into clear, visual insights that support understanding of crime patterns and trends.

Key Dashboard Components

1. Key Performance Indicators (KPIs)

- Displayed overall totals such as incidents, offenses, victims, and offenders.
- Provided a quick snapshot of the dataset's main statistics.

2. State-Level Crime Comparison

- Visualized total offenses across states using comparative charts.
- Helped identify states with higher levels of recorded crime.

3. Victim and Offender Analysis

- Compared total victims and offenders to examine crime impact.
- Highlighted patterns in crime involvement.

4. Offense Type Breakdown

- Categorized crimes into groups such as violent, theft, and drug-related offenses.
- Identified the most common types of offenses.

5. Interactive Features

- Included slicers for filtering by state, month, and offense category.
- Enabled dynamic interaction between visuals for deeper exploration.

Overall Impact

The Power BI dashboard enhanced data interpretation by presenting crime information in a structured and interactive format. It simplified complex data, supported comparison across different categories, and improved the clarity of insights for reporting and decision-making purposes.

OVERALL OBSERVATION

The analysis carried out using Python (EDA), SQL, Excel, and Power BI revealed clear and consistent crime trends within the dataset. Through Python-based exploratory analysis, key patterns were identified, including increased crime activity during late evening hours, a higher proportion of adult victims, and a moderate positive relationship between total offenses and total victims. SQL was used to organize the data efficiently and perform structured queries for state-level comparisons, offense category summaries, and aggregated insights.

Excel pivot tables helped uncover property value outliers, demographic distributions, and differences in crime density across states. Power BI integrated these findings into interactive dashboards with dynamic filters for state, month, and offense type, making the analysis more accessible and visually impactful.

Overall, the combined use of these tools converted raw crime data into well-structured, meaningful insights. The study highlighted time-based patterns, demographic trends, regional variations, and key variable relationships, supporting better understanding and data-driven decision-making.

CONCLUSION

This project examined the Individual_Incident_2020 dataset using Python, SQL, Excel, and Power BI to uncover meaningful crime trends and patterns. Data preprocessing and exploratory analysis in Python enhanced data accuracy and highlighted key relationships within the dataset. SQL was used to organize the data into a structured format and perform detailed analytical queries. Excel supported the creation of pivot tables and dashboard summaries, while Power BI presented the results through interactive and visually engaging reports.

The analysis identified clear patterns in hourly crime activity, victim demographics, crime concentration across states, and the association between total offenses and victim counts. Overall, the project demonstrates how integrating multiple analytical tools can transform raw crime data into clear, structured insights that assist in informed decision-making and professional reporting.