Chapter 1: Introduction to the project

1.1 What is data Science

Data Science:

Data Science is a multidisciplinary field that involves using scientific methods, processes, algorithms, and systems to extract knowledge and insights from structured and unstructured data. It combines techniques from statistics, mathematics, computer science, and domain expertise to analyze large datasets and solve complex problems.

Key Components of Data Science:

1. **Data Collection**: Gathering data from various sources (databases, sensors, web scraping, APIs, etc.).
2. **Data Cleaning & Preparation**: Ensuring data quality by handling missing values, duplicates, and inconsistencies.
3. **Exploratory Data Analysis (EDA)**: Understanding data patterns, relationships, and distributions using visualizations and summary statistics.
4. **Modeling & Algorithm Development**: Applying machine learning and statistical models to make predictions or discover patterns.
5. **Interpretation & Communication**: Presenting insights through visualizations, reports, and dashboards to help stakeholders make data-driven decisions.

Applications of Data Science:

- Predictive Analytics (e.g., predicting customer behavior)
- Natural Language Processing (e.g., chatbots, sentiment analysis)
- Image Recognition (e.g., facial recognition, medical imaging)
- Recommendation Systems (e.g., Netflix, Amazon)
- Fraud Detection (e.g., in banking and insurance)

In short, **Data Science helps organizations make smarter, data-driven decisions by transforming raw data into actionable insights**.

This project involves analyzing **crimes against women in India** from **2001 to 2021**, covering various categories such as **rape cases**, **dowry deaths**, **domestic violence**, and **women trafficking** across different states. The analysis includes **trend visualization** over time, **state-wise comparisons**, and identification of key patterns. Missing data is handled to ensure smooth, constant line graphs for each crime category. The project aims to provide actionable insights for policymakers to improve women's safety.

1.2 What is crime against women?

Crimes against women refer to acts of **violence, discrimination, or abuse** specifically directed at women, often rooted in gender inequality and societal norms. These crimes include **physical**, **emotional**, **sexual**, and **economic abuse** that violate women's rights and dignity.

Common examples include:

- **Rape** and **sexual harassment**
- **Domestic violence** and **dowry deaths**
- **Human trafficking** and **forced prostitution**
- **Acid attacks** and **honour killings**

Efforts to address crimes against women involve **legal reforms**, **awareness campaigns**, and **improved support systems** to ensure justice and safety for women.

1.3 Current Scenario of Crimes Against Women in India (as of 2025)

Despite significant legal reforms, increased reporting, and greater public awareness, crimes against women in India remain a major social issue. According to the **National Crime Records Bureau (NCRB)**, the number of reported crimes against women has shown a consistent rise in recent years, with categories like **domestic violence**, **sexual harassment**, and **rape** being the most prevalent.

Key Issues:

1. **Increased Reporting**:

   While the increase in reported cases reflects growing awareness and improved willingness to report crimes, it also highlights the persistent prevalence of gender-based violence.

2. **Domestic Violence**:
   Domestic violence continues to be one of the most frequently reported crimes. The COVID-19 pandemic exacerbated this issue, with many women facing abuse during lockdowns.
3. **Sexual Crimes**:
   Rape and sexual assault cases remain a serious concern. High-profile cases often lead to public outrage, resulting in calls for stricter enforcement of laws and faster judicial processes.
4. **Trafficking and Exploitation**:
   Human trafficking, particularly for forced labor and prostitution, continues to affect vulnerable women, especially in rural and economically weaker areas.
5. **Cybercrime**:
   The rise in internet usage has led to an increase in cybercrimes against women, including online harassment, stalking, and non-consensual sharing of private content.

Government and Social Initiatives:

1. **Legal Framework**:
   - Strict laws like the **Criminal Law (Amendment) Act, 2013** and the **Protection of Women from Domestic Violence Act, 2005** aim to protect women's rights.
   - Fast-track courts have been established to ensure quicker trials in cases of crimes against women.
2. **Helplines and Support Systems**:
   Various helplines and support centres, such as **181 Women’s Helpline** and **One-Stop Crisis Centres**, provide immediate assistance to women in distress.
3. **Awareness Campaigns**:
   Campaigns like **Beti Bachao Beti Padhao** and NGO-led initiatives focus on educating the public about women's rights and gender equality.

Challenges:

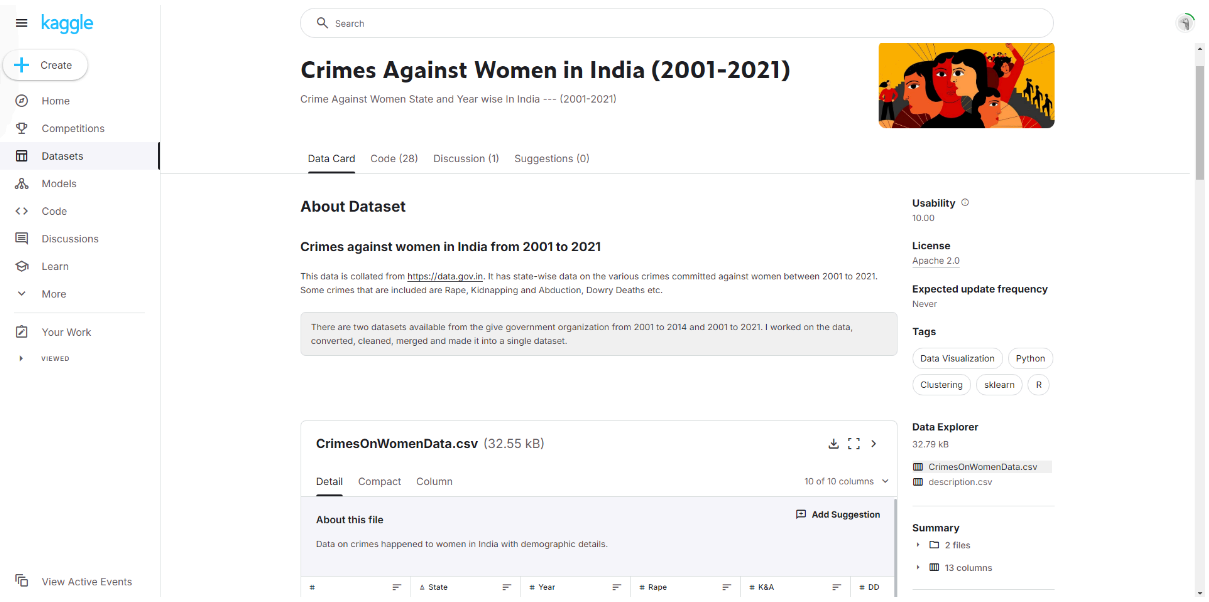
- **Underreporting**: Despite increased reporting, many cases still go unreported due to fear of stigma, lack of trust in law enforcement, and societal pressures.
- **Judicial Delays**: Many cases drag on for years, delaying justice for victims.
- **Social Norms**: Deep-rooted patriarchal mindsets continue to pose a barrier to gender equality and safety.

In conclusion, while India has made progress in addressing crimes against women, much work remains to be done in terms of **changing societal attitudes**, **strengthening law enforcement**, and **providing robust support systems** to ensure the safety and dignity of women.

Data source of the project:

The datasets used for this project are taken from kaggle.com but originally belongs to National Crime Bureau of India. The data refers to State/UT wise crime committed against women categorized by different crime heads during the years .

https://www.kaggle.com/datasets/balajivaraprasad/crimes-against-women-in-india-2001-2021



Crimes Against Women in India (2001-2021)

Crime Against Women State and Year wise In India --- (2001-2021)

About Dataset

Crimes against women in India from 2001 to 2021

This data is collated from https://data.gov.in. It has state-wise data on the various crimes committed against women between 2001 to 2021. Some crimes that are included are Rape, Kidnapping and Abduction, Dowry Deaths etc.

There are two datasets available from the give government organization from 2001 to 2014 and 2001 to 2021. I worked on the data, converted, cleaned, merged and made it into a single dataset.

CrimesOnWomenData.csv (32.55 kB)

10 of 10 columns

About this file

Data on crimes happened to women in India with demographic details.

Usability
10.00

License
Apache 2.0

Expected update frequency
Never

Tags
Data Visualization, Python, Clustering, sklearn, R

Data Explorer
32.79 kB
CrimesOnWomenData.csv
description.csv

Summary
2 files
13 columns

1.4 Software Used in the Project

1. **Python**

   Python is a powerful and versatile programming language widely used in data science and machine learning. In this project, Python is used for data processing, analysis, and visualization due to its rich ecosystem of libraries. Its simple syntax and extensive libraries make it an ideal choice for data-driven projects.

It acts as the backbone, handling data loading, processing, and execution of algorithms.

2. Jupyter Notebook

Jupyter Notebook is an open-source web application that allows users to create and share documents containing live code, equations, visualizations, and narrative text. It is used in this project as the primary environment for writing and executing Python code, performing data analysis, and presenting visualizations in an interactive manner.It provides an interactive environment to develop and document the entire project.

This project strictly follows the 6 phases of the data analytical process namely:

1. The Ask phase

2. The Prepare phase

3. The Process phases

4. The Analysis phase

5. The Share phases

6. The Act phases

This report consists of 7 chapters:

● Chapter1: Introduction

Introduction comprises an overall view of the project which is the main goal of performing this analysis.

● Chapter2: The ask phase

The ask phase briefly discusses the outcomes that are expected from the project.

● Chapter3: Creating a scope of work

Scope of work is a document that forms the outline of the project. It covers every step that we are going to take in the data analytic process of the project.

● Chapter4: The prepare phase

This chapter gives more information about different data sources present and the one opted for the analysis process.

● Chapter5: The process phase

The process of cleaning the dirty dataset and making it ready to go for analyses is discussed in this chapter.

● Chapter6: The analysis phase and the visualise/share phase

This chapter is about analysing the clean data using different techniques of analysis and discusses how the analysed data is put into life by visualising it.

● Chapter7: Conclusion/ suggestions from the analysis

In this chapter the insights provided by the analysis project is discussed. This section summarises the whole project.

Chapter 2: The Ask Phase

2.1 Introduction

The ask phase is all about the clearance of the problem. Considering the needs of the stakeholders and asking them questions about their expectations from the project is all that is covered in the ask phase. This phase is very much important in data analytics because it is the basic building block of the entire process.

Two tasks to be done in this ASK phase are:
● We define the problem and solve it.
● We make sure that we fully understand stakeholder's expectations.

Stakeholder (Definition): A stakeholder is a person or group of people who invest resources in an analytical project and are interested in the outcomes of the project for making data driven decisions.

2.2 Starting the ask phase

In this project, we had no stakeholders, so considering our team as the stakeholder is the best consideration. To get started with the "ask phase" it was needed to consider what would be the outcomes of the projects and what are the insights to be driven from the analysis.

Following are some questions that are to be considered during the ASK phase:

- What would be the insights of this dataset ?
- What could we suggest to overcome the crime rate against women in india?

2.3 The main insights to be driven from analysis

Considering the main objectives of the project, we needed to Analyse the following main points using the dataset:

1. **Overall Crime Trends Over the Years (2001–2021)**

   We can analyze the **year-wise trends** for different crime categories, such as **rape**

cases, **domestic violence**, **dowry deaths**, and **women trafficking**, to understand whether the number of reported cases is increasing, decreasing, or remaining constant over time.

2. **State-Wise Crime Trends**

   We can generate **state-wise line graphs** for each crime category to identify patterns and trends across states. This will help us highlight states with consistently high or low crime rates and detect regional variations in crimes against women.

3. **Comparison Across Different Crime Categories**

   By comparing trends across multiple crime categories, we can identify which types of crimes are rising or falling over time and assess whether specific categories need more attention from policymakers.

4. **Increase in Crimes in Every State Over Time**

   We can calculate and plot the **yearly increase in crimes for each state**, highlighting which states are experiencing rapid increases or decreases in specific crime categories. This will help in identifying regions where intervention is urgently needed.

5. **Handling Missing Data and Ensuring Consistent Visualization**

   We can handle missing data by filling gaps with zeros, ensuring that **every state has a constant line across all years** in the line graphs. This will provide a complete and smooth visualization of crime trends for each state.

6. **Interactive Visualization for Better Exploration**

   We can create **interactive plots** using tools like Plotly, allowing users to explore crime trends dynamically by selecting specific states or crime categories.

Chapter 3: Scope of work (SoW)

Scope of Work (SoW): An agreed upon outline of the work you are going to perform on a project. Scope of work comprises of following things:

Deliverables: A deliverable is a tangible or intangible good or service produced as a result of a project that is intended to be delivered to a customer. A deliverable could be a report, a document, a software product, a server upgrade or any other building block of an overall project.

Timelines: It is the time period that is specified for the analysts to perform the sub tasks or the overall project and submit it to the stakeholder.

Milestones: Milestones are significant tasks you will confirm along your timeline to help everyone know the project is on track.

Reports: Reports are the end results of the sub tasks or the overall project

Statement of Work (SoW): A statement of work is a document that clearly identifies the products and services a vendor or contractor will provide to an organisation. It includes objectives, guidelines, deliverables, schedule, and cost

Data Science Project: Crime against Women in India

Analyst: Akhil Kapoor

Client/Sponsor: None

Purpose: The main purpose of the project is to deliver the suitable audience (i.e people from medical field and research field) a visual platform for analysing the data of different types of crime against women of india.

Scope / Major Project Activities:

Activity	**Description**
Data gathering	In this step we will be gathering the data required for the analysis.
Data cleaning	In this step all the errors in the data will be resolved and the missing values will be completed.
Data transformation	In this step the data will be transformed into the desired format.

Data visualisation	In this step visualisation for the given data will be created.

NOTE: Scope of work is a sub part of the statement of work.

Deliverables

Deliverable	**Description/ Details**
Visualisations of different aspects of analysis	The report includes the visualizations of all the reasons for crime against women that are mentioned in the given data.

Estimated date for completion:

This is my “if all goes well and I have everything I need, this is when I’ll be done” “05/07/24”.

Chapter 4: The Prepare phase

4.1 Introduction

The preparation phase is all about preparing the data for your analysis. It includes the data gathering, storing and checks for the security of the data. In this phase of the analytic process, we look for the data sources that we need to perform the analysis on. The whole prepare phase can be summarised as:

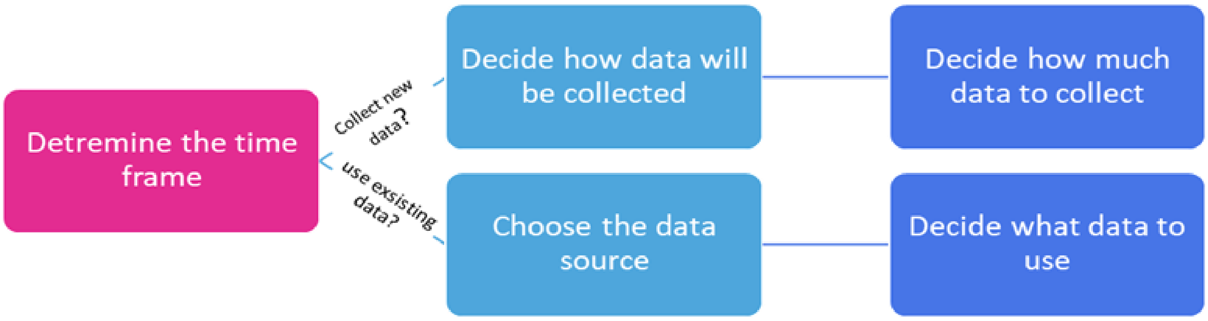
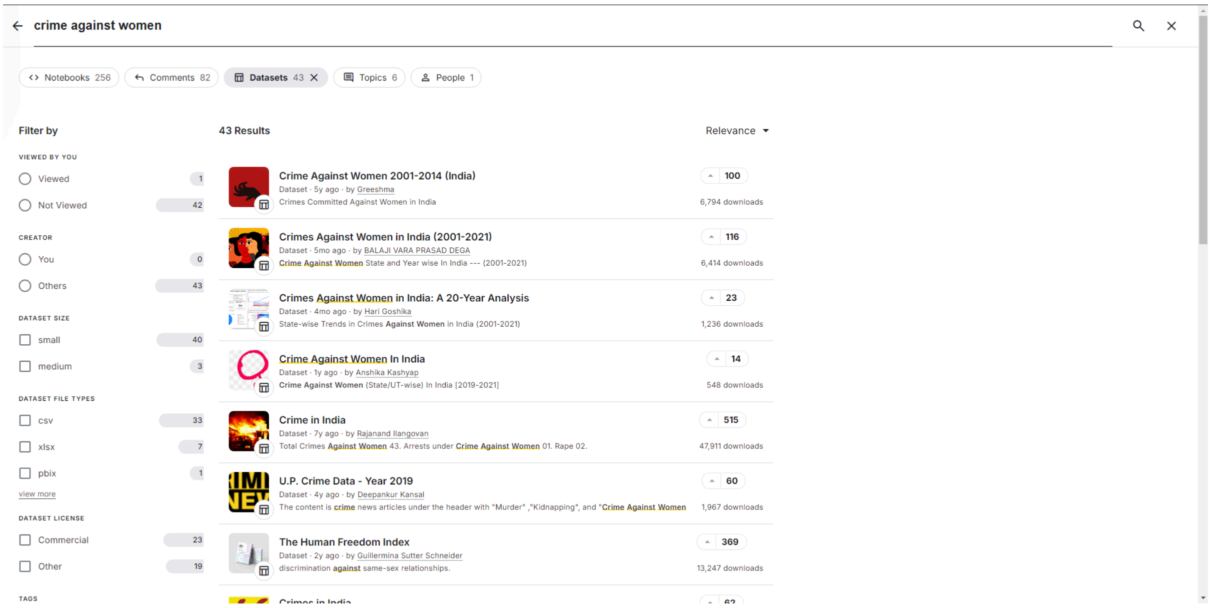


Figure 4.1

Here we choose to collect data from reliable open government data sources (i.e NCRB, refined data from kaggle).



← crime against women

<> Notebooks 256 | ↰ Comments 82 | Datasets 43 ✕ | Topics 6 | People 1

Filter by

VIEWED BY YOU
○ Viewed 1
○ Not Viewed 42

CREATOR
○ You 0
○ Others 43

DATASET SIZE
□ small 40
□ medium 3

DATASET FILE TYPES
□ csv 33
□ xlsx 7
□ pbix 1
view more

DATASET LICENSE
□ Commercial 23
□ Other 19

TAGS

43 Results — Relevance ▾

Crime Against Women 2001-2014 (India)
Dataset · 5y ago · by Greeshma
Crimes Committed Against Women in India
▲ 100 — 6,794 downloads

Crimes Against Women in India (2001-2021)
Dataset · 5mo ago · by BALAJI VARA PRASAD DEGA
Crime Against Women State and Year wise In India --- (2001-2021)
▲ 116 — 6,414 downloads

Crimes Against Women in India: A 20-Year Analysis
Dataset · 4mo ago · by Hari Goshika
State-wise Trends in Crimes Against Women in India (2001-2021)
▲ 23 — 1,236 downloads

Crime Against Women In India
Dataset · 1y ago · by Anshika Kashyap
Crime Against Women (State/UT-wise) In India [2019-2021]
▲ 14 — 548 downloads

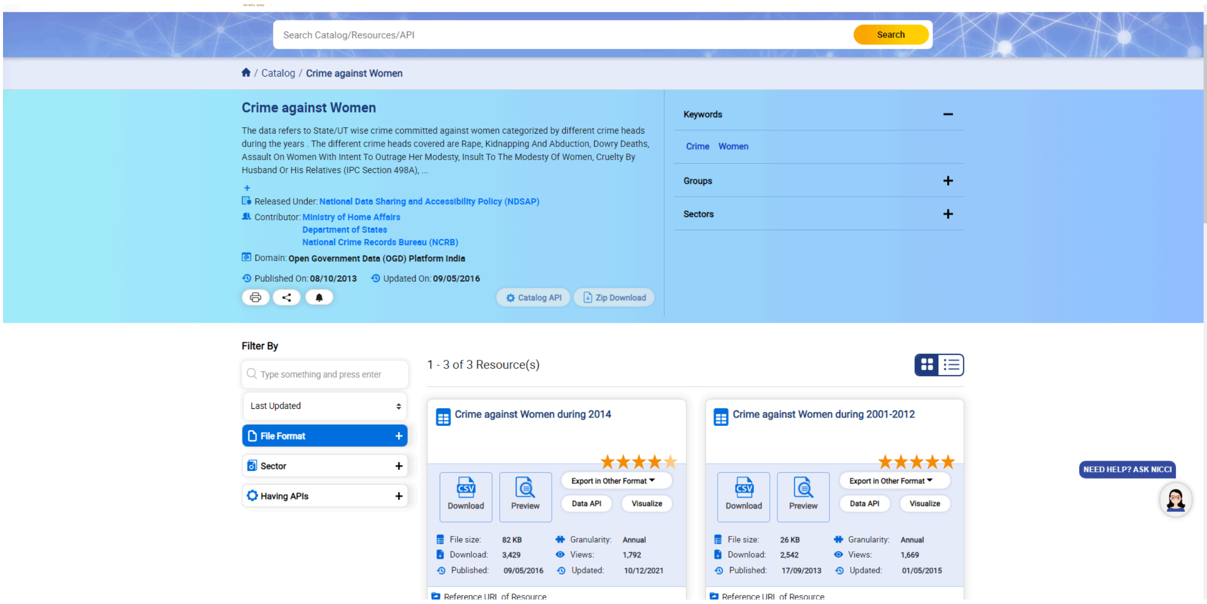
Crime in India
Dataset · 7y ago · by Rajanand Ilangovan
Total Crimes Against Women 43. Arrests under Crime Against Women 01. Rape 02.
▲ 515 — 47,911 downloads

U.P. Crime Data - Year 2019
Dataset · 4y ago · by Deepankur Kansal
The content is crime news articles under the header with "Murder" ,"Kidnapping", and "Crime Against Women
▲ 60 — 1,967 downloads

The Human Freedom Index
Dataset · 2y ago · by Guillermina Sutter Schneider
discrimination against same-sex relationships.
▲ 369 — 13,247 downloads

Crimes in India
▲ 62

There was a variety of data available about crime against women in India on this online open source platform and we selected the data according to the project needs.



Crime against Women

The data refers to State/UT wise crime committed against women categorized by different crime heads during the years . The different crime heads covered are Rape, Kidnapping And Abduction, Dowry Deaths, Assault On Women With Intent To Outrage Her Modesty, Insult To The Modesty Of Women, Cruelty By Husband Or His Relatives (IPC Section 498A), ...

Released Under: National Data Sharing and Accessibility Policy (NDSAP)
Contributor: Ministry of Home Affairs
Department of States
National Crime Records Bureau (NCRB)
Domain: Open Government Data (OGD) Platform India
Published On: 08/10/2013 Updated On: 09/05/2016

1 - 3 of 3 Resource(s)

Crime against Women during 2014

File size:	82 KB	Granularity:	Annual
Download:	3,429	Views:	1,792
Published:	09/05/2016	Updated:	10/12/2021

Crime against Women during 2001-2012

File size:	26 KB	Granularity:	Annual
Download:	2,542	Views:	1,669
Published:	17/09/2013	Updated:	01/05/2015

4.2 Determining the time frame for the collection of data

Time frame required for the collection of data needed for the project depends on the size of the project. For this project we set the milestone of four days for the collection of data.

4.3 Deciding how the data will be collected

The idea is to collect the data from the Internet, websites and web sources like various government websites of India. The preferred source of data for the project data is ncrb.gov.in These government websites are a more reliable source for the data required for the analysis. The data used in this analysis is itself collected by the team members and this type of data is known as first party data. First party data is the most reliable source of the data.

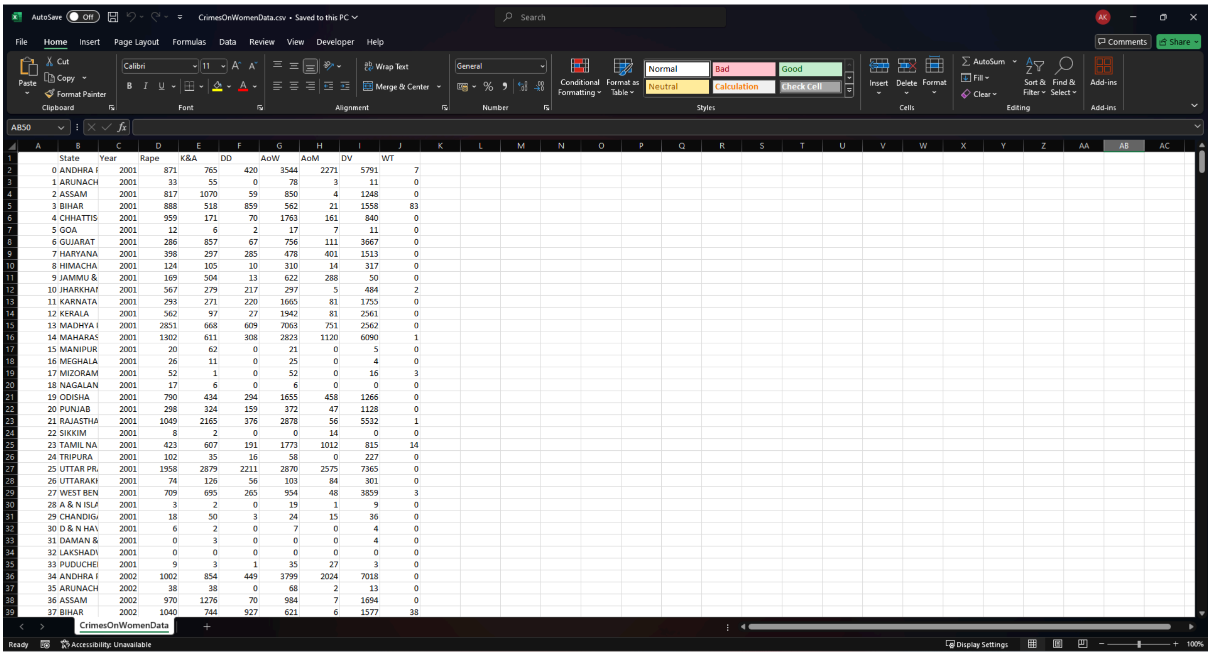
4.4 Deciding how much data to collect

The following points should be considered while deciding the amount of data needed for the analysis:

- The data set should be large enough to avoid any bias.
- The data set should be small enough to prevent overfitting.
- While collecting the data we should keep in mind the business objectives. As "Reliable Data+ Business objectives = Accurate outcomes.

4.5 Dataset glimpse

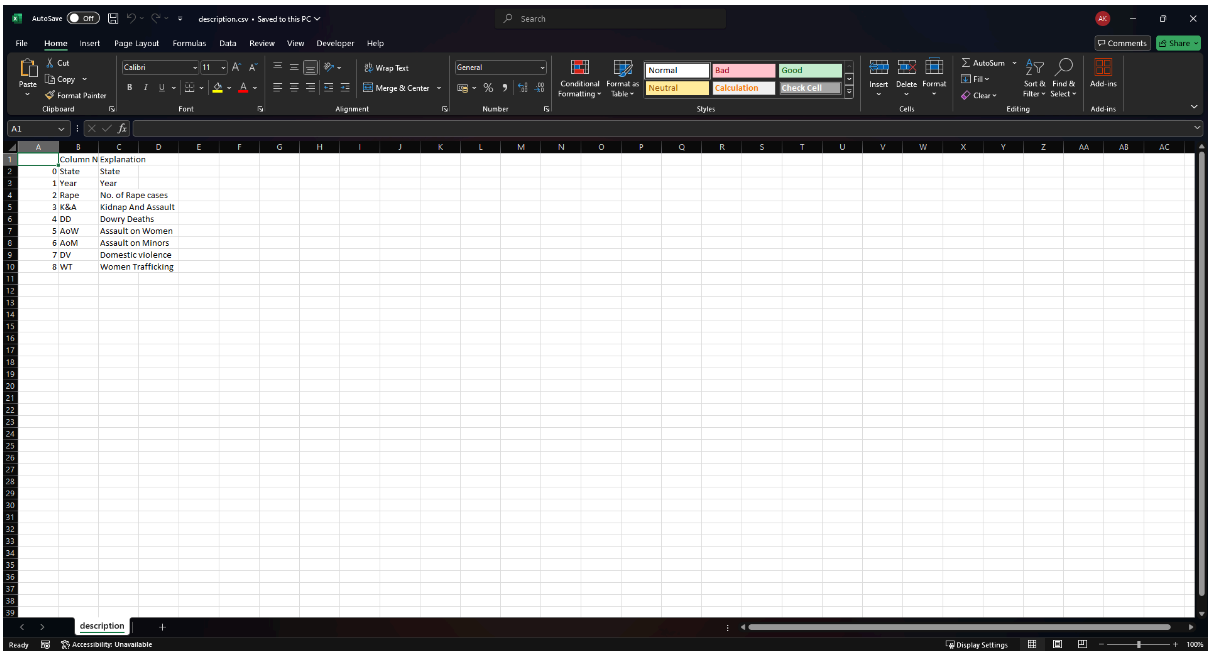
These are all the datasets that have been referred to fulfil this project goal.



	State	Year	Rape	K&A	DD	AoW	AoM	DV	WT
0	ANDHRA P	2001	871	765	420	3544	2271	5791	7
1	ARUNACH	2001	33	55	0	78	3	11	0
2	ASSAM	2001	817	1070	59	850	4	1248	0
3	BIHAR	2001	888	518	859	562	21	1558	83
4	CHHATTIS	2001	959	171	70	1763	161	840	0
5	GOA	2001	12	6	2	17	7	11	0
6	GUJARAT	2001	286	857	67	756	111	3667	0
7	HARYANA	2001	398	297	285	478	401	1513	0
8	HIMACHA	2001	124	105	10	310	14	317	0
9	JAMMU &	2001	169	504	13	622	288	50	0
10	JHARKHAI	2001	567	279	217	297	5	484	2
11	KARNATA	2001	293	271	220	1665	81	1755	0
12	KERALA	2001	562	97	27	1942	81	2561	0
13	MADHYA P	2001	2851	668	609	7063	751	2562	0
14	MAHARAS	2001	1302	611	308	2823	1120	6090	1
15	MANIPUR	2001	20	62	0	21	0	5	0
16	MEGHALA	2001	26	11	0	25	0	4	0
17	MIZORAM	2001	52	1	0	52	0	16	3
18	NAGALAN	2001	17	6	0	6	0	0	0
19	ODISHA	2001	790	434	294	1655	458	1266	0
20	PUNJAB	2001	298	324	159	372	47	1128	0
21	RAJASTHA	2001	1049	2165	376	2878	56	5532	1
22	SIKKIM	2001	8	2	0	0	14	0	0
23	TAMIL NA	2001	423	607	191	1773	1012	815	14
24	TRIPURA	2001	102	35	16	58	0	227	0
25	UTTAR PR	2001	1958	2879	2211	2870	2575	7365	0
26	UTTARAKH	2001	74	126	56	103	84	301	0
27	WEST BEN	2001	709	695	265	954	48	3859	3
28	A & N ISLA	2001	3	2	0	19	1	9	0
29	CHANDIGA	2001	18	50	3	24	15	36	0
30	D & N HAV	2001	6	2	0	7	0	4	0
31	DAMAN &	2001	0	3	0	0	0	4	0
32	LAKSHADV	2001	0	0	0	0	0	0	0
33	PUDUCHEI	2001	9	3	1	35	27	3	0
34	ANDHRA P	2002	1002	854	449	3799	2024	7018	0
35	ARUNACH	2002	38	38	0	68	2	13	0
36	ASSAM	2002	970	1276	70	984	7	1694	0
37	BIHAR	2002	1040	744	927	621	6	1577	38

CrimesOnWomenData.csv • Saved to this PC
AB50
CrimesOnWomenData
Ready
Accessibility: Unavailable

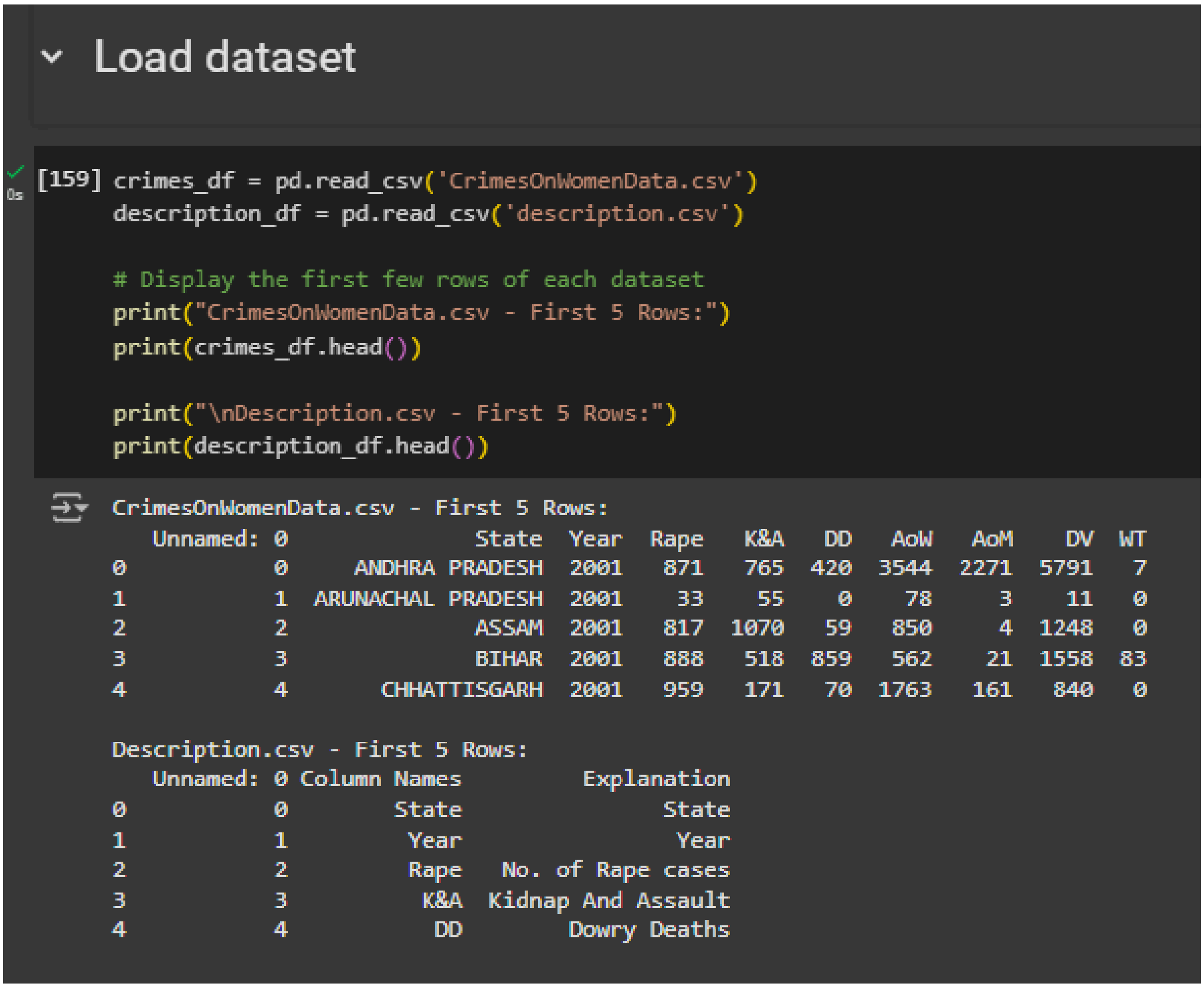
Figure 4.3: Crime Against Women(2001-21)



	Column N	Explanation
0	State	State
1	Year	Year
2	Rape	No. of Rape cases
3	K&A	Kidnap And Assault
4	DD	Dowry Deaths
5	AoW	Assault on Women
6	AoM	Assault on Minors
7	DV	Domestic violence
8	WT	Women Trafficking

description.csv • Saved to this PC
A1
description
Ready
Accessibility: Unavailable

Figure 4.4: Description



Load dataset

```
[159] crimes_df = pd.read_csv('CrimesOnWomenData.csv')
      description_df = pd.read_csv('description.csv')

      # Display the first few rows of each dataset
      print("CrimesOnWomenData.csv - First 5 Rows:")
      print(crimes_df.head())

      print("\nDescription.csv - First 5 Rows:")
      print(description_df.head())
```

```
CrimesOnWomenData.csv - First 5 Rows:
   Unnamed: 0              State  Year  Rape   K&A   DD   AoW   AoM    DV  WT
0           0     ANDHRA PRADESH  2001   871   765  420  3544  2271  5791   7
1           1  ARUNACHAL PRADESH  2001    33    55    0    78     3    11   0
2           2              ASSAM  2001   817  1070   59   850     4  1248   0
3           3              BIHAR  2001   888   518  859   562    21  1558  83
4           4       CHHATTISGARH  2001   959   171   70  1763   161   840   0

Description.csv - First 5 Rows:
   Unnamed: 0 Column Names         Explanation
0           0        State               State
1           1         Year                Year
2           2         Rape   No. of Rape cases
3           3          K&A  Kidnap And Assault
4           4           DD        Dowry Deaths
```

Figure 4.5: Loading the dataset in notebook

Chapter 5: The Process Phase

5.1 Introduction to Process phase

The process phase is all about cleaning the dataset that you have for your analysis.

- A dataset is said to be dirty when the data is incorrect, incomplete, or irrelevant to the problem we are trying to solve.
- Whereas clean data is the data that is complete, correct and relevant to the problem we are trying to solve.

5.2 First step in data cleansing:

Ensuring data integrity

Before we start cleaning out dataset, we need to ensure that all the data present in the dataset is integrated, that it is:

- Complete
- Consistent
- Accurate

In this project, the dataset is complete, accurate and consistent because it is collected from trustworthy and reliable sources. But when data was transformed the errors like missing values and inefficiency in the data was seen. These gaps in the data were covered manually by referring to the original data sources of the government that were downloaded.

5.3 Starting to clean the data

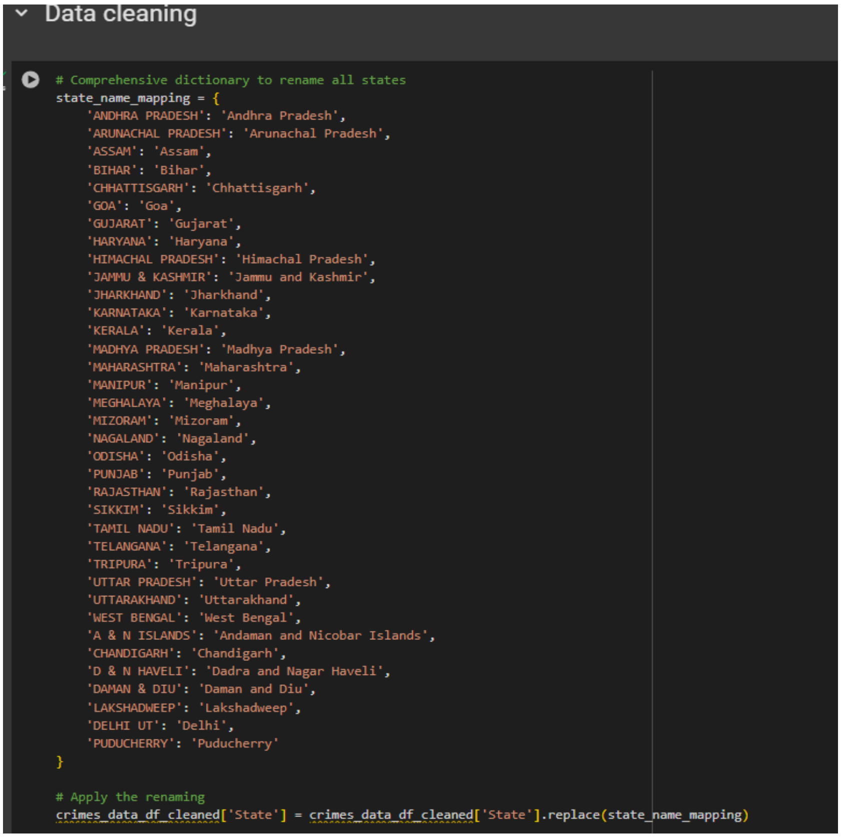
After ensuring the data integrity we can get started with the data cleansing process. To clean our data efficiently we may check for the following flaws in our dataset:

- Duplicate entries
- Typos
- Wrong field entries
- Spelling mistakes
- Missing values or fields etc.

As the dataset is organic that is we collected it, still there is a higher probability of typos and other errors.

Solving typos: Typos usually occur while adding the meta data or abbreviating, We used filtering to solve the typo errors.

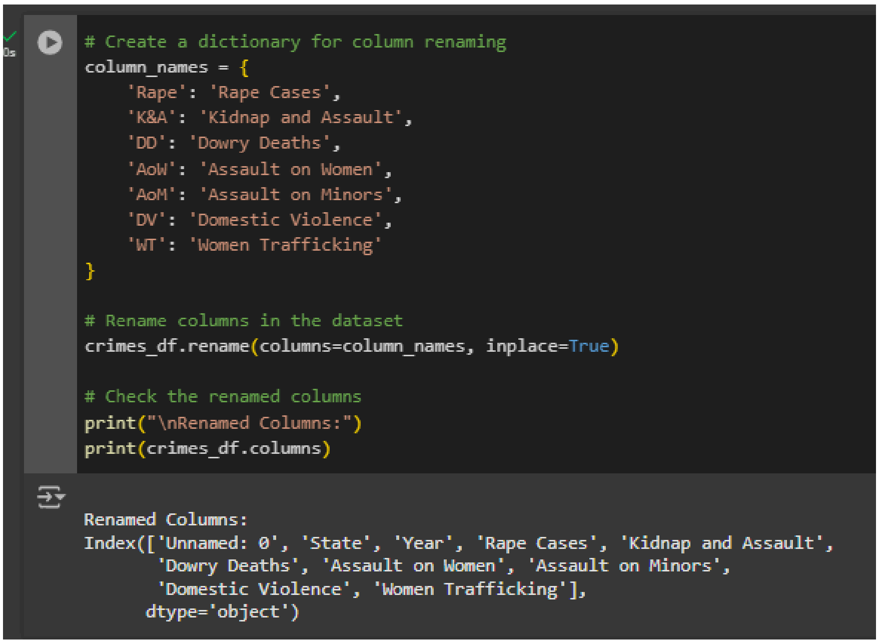
Removing nulls: As the datasets were created organically and we knew the sources, so instead of deleting the nulls we filled them.



Data cleaning

```
# Comprehensive dictionary to rename all states
state_name_mapping = {
    'ANDHRA PRADESH': 'Andhra Pradesh',
    'ARUNACHAL PRADESH': 'Arunachal Pradesh',
    'ASSAM': 'Assam',
    'BIHAR': 'Bihar',
    'CHHATTISGARH': 'Chhattisgarh',
    'GOA': 'Goa',
    'GUJARAT': 'Gujarat',
    'HARYANA': 'Haryana',
    'HIMACHAL PRADESH': 'Himachal Pradesh',
    'JAMMU & KASHMIR': 'Jammu and Kashmir',
    'JHARKHAND': 'Jharkhand',
    'KARNATAKA': 'Karnataka',
    'KERALA': 'Kerala',
    'MADHYA PRADESH': 'Madhya Pradesh',
    'MAHARASHTRA': 'Maharashtra',
    'MANIPUR': 'Manipur',
    'MEGHALAYA': 'Meghalaya',
    'MIZORAM': 'Mizoram',
    'NAGALAND': 'Nagaland',
    'ODISHA': 'Odisha',
    'PUNJAB': 'Punjab',
    'RAJASTHAN': 'Rajasthan',
    'SIKKIM': 'Sikkim',
    'TAMIL NADU': 'Tamil Nadu',
    'TELANGANA': 'Telangana',
    'TRIPURA': 'Tripura',
    'UTTAR PRADESH': 'Uttar Pradesh',
    'UTTARAKHAND': 'Uttarakhand',
    'WEST BENGAL': 'West Bengal',
    'A & N ISLANDS': 'Andaman and Nicobar Islands',
    'CHANDIGARH': 'Chandigarh',
    'D & N HAVELI': 'Dadra and Nagar Haveli',
    'DAMAN & DIU': 'Daman and Diu',
    'LAKSHADWEEP': 'Lakshadweep',
    'DELHI UT': 'Delhi',
    'PUDUCHERRY': 'Puducherry'
}

# Apply the renaming
crimes_data_df_cleaned['State'] = crimes_data_df_cleaned['State'].replace(state_name_mapping)
```

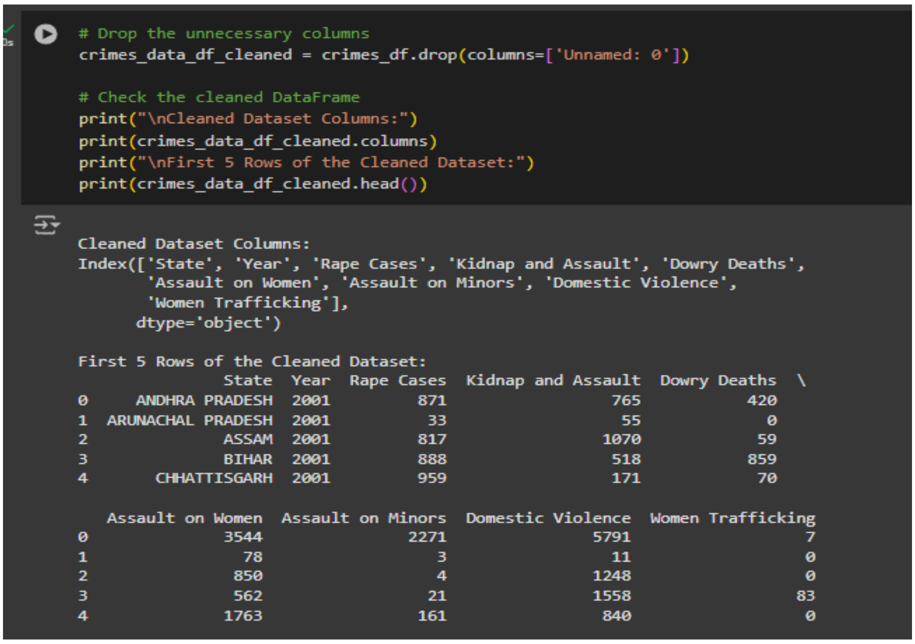
Figure 5.1: Mapping to Reduce Duplicity



```
# Create a dictionary for column renaming
column_names = {
    'Rape': 'Rape Cases',
    'K&A': 'Kidnap and Assault',
    'DD': 'Dowry Deaths',
    'AoW': 'Assault on Women',
    'AoM': 'Assault on Minors',
    'DV': 'Domestic Violence',
    'WT': 'Women Trafficking'
}

# Rename columns in the dataset
crimes_df.rename(columns=column_names, inplace=True)

# Check the renamed columns
print("\nRenamed Columns:")
print(crimes_df.columns)
```

```
Renamed Columns:
Index(['Unnamed: 0', 'State', 'Year', 'Rape Cases', 'Kidnap and Assault',
       'Dowry Deaths', 'Assault on Women', 'Assault on Minors',
       'Domestic Violence', 'Women Trafficking'],
      dtype='object')
```

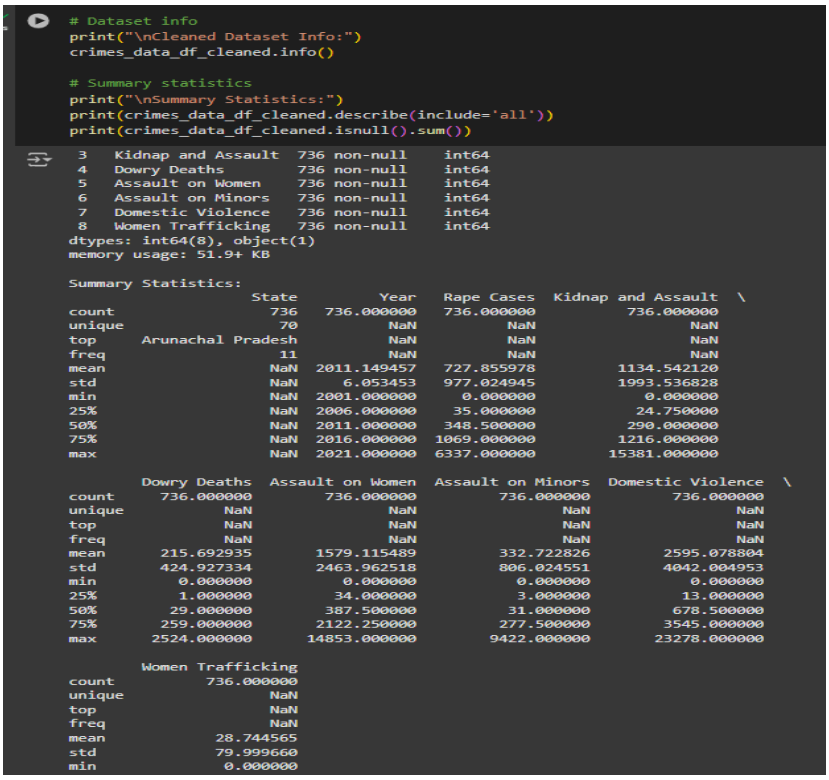
Figure 5.2: Renaming columns



```
# Drop the unnecessary columns
crimes_data_df_cleaned = crimes_df.drop(columns=['Unnamed: 0'])

# Check the cleaned DataFrame
print("\nCleaned Dataset Columns:")
print(crimes_data_df_cleaned.columns)
print("\nFirst 5 Rows of the Cleaned Dataset:")
print(crimes_data_df_cleaned.head())
```

```
Cleaned Dataset Columns:
Index(['State', 'Year', 'Rape Cases', 'Kidnap and Assault', 'Dowry Deaths',
       'Assault on Women', 'Assault on Minors', 'Domestic Violence',
       'Women Trafficking'],
      dtype='object')

First 5 Rows of the Cleaned Dataset:
               State  Year  Rape Cases  Kidnap and Assault  Dowry Deaths  \
0     ANDHRA PRADESH  2001         871                 765           420
1  ARUNACHAL PRADESH  2001          33                  55             0
2              ASSAM  2001         817                1070            59
3              BIHAR  2001         888                 518           859
4       CHHATTISGARH  2001         959                 171            70

   Assault on Women  Assault on Minors  Domestic Violence  Women Trafficking
0              3544               2271               5791                  7
1                78                  3                 11                  0
2               850                  4               1248                  0
3               562                 21               1558                 83
4              1763                161                840                  0
```

Figure 5.3: dropping unnecessary columns



```
# Dataset info
print("\nCleaned Dataset Info:")
crimes_data_df_cleaned.info()

# Summary statistics
print("\nSummary Statistics:")
print(crimes_data_df_cleaned.describe(include='all'))
print(crimes_data_df_cleaned.isnull().sum())
```

```
 3   Kidnap and Assault  736 non-null    int64
 4   Dowry Deaths        736 non-null    int64
 5   Assault on Women    736 non-null    int64
 6   Assault on Minors   736 non-null    int64
 7   Domestic Violence   736 non-null    int64
 8   Women Trafficking   736 non-null    int64
dtypes: int64(8), object(1)
memory usage: 51.9+ KB

Summary Statistics:
                    State         Year   Rape Cases  Kidnap and Assault  \
count                 736   736.000000   736.000000          736.000000
unique                 70          NaN          NaN                 NaN
top     Arunachal Pradesh          NaN          NaN                 NaN
freq                   11          NaN          NaN                 NaN
mean                  NaN  2011.149457   727.855978         1134.542120
std                   NaN     6.053453   977.024945         1993.536828
min                   NaN  2001.000000     0.000000            0.000000
25%                   NaN  2006.000000    35.000000           24.750000
50%                   NaN  2011.000000   348.500000          290.000000
75%                   NaN  2016.000000  1069.000000         1216.000000
max                   NaN  2021.000000  6337.000000        15381.000000

        Dowry Deaths  Assault on Women  Assault on Minors  Domestic Violence  \
count     736.000000        736.000000         736.000000         736.000000
unique           NaN               NaN                NaN                NaN
top              NaN               NaN                NaN                NaN
freq             NaN               NaN                NaN                NaN
mean      215.692935       1579.115489         332.722826        2595.078804
std       424.927334       2463.962518         806.024551        4042.004953
min         0.000000          0.000000           0.000000           0.000000
25%         1.000000         34.000000           3.000000          13.000000
50%        29.000000        387.500000          31.000000         678.500000
75%       259.000000       2122.250000         277.500000        3545.000000
max      2524.000000      14853.000000        9422.000000       23278.000000

        Women Trafficking
count          736.000000
unique                NaN
top                   NaN
freq                  NaN
mean            28.744565
std             79.999660
min              0.000000
```

Figure 5.4: checking for null values

Chapter 6: The Analyze Phase and Share Phase

After getting the data from dirty to clean, the next step in the data analytic process is to analyse the data. To get started with the analysis of the data we need to refer to the Scope of Work again. SoW will give us an idea of what aspects of the data we need to explore.

Analysis phase: Analysing the collected data involves using tools to transform and organise that information so that useful conclusions, predictions, and data drive informed decisions could be drawn.

Share phase: In this phase data analysts interpret results and share them with the stakeholders in order to make effective data driven decisions. Data is being shared in the visual form in order to make more clarity about the data to the entire team.

6.1 Noting down the analysis activities

Referring to the scope of work, the following is the list of analysis activities that we need to perform on the project data: Creating visualisation for crime against women in different categories with respect to place and year is the main goal of this project.

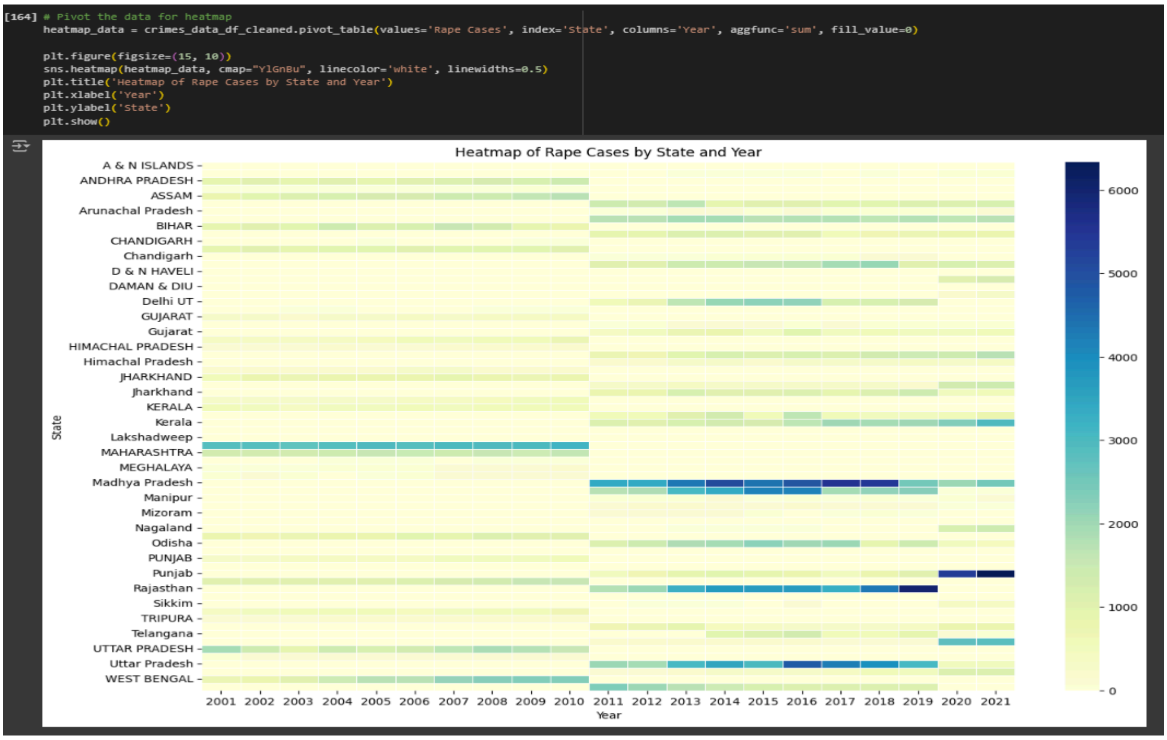
6.2 Creating models for identifying the underlying trends

In this project, I used Linear Regression, Random Forest Regressor, Gradient Boosting Regressor, and XGBoost Regressor to predict crime trends over time. These models were trained using historical crime data, with the year as a key feature and the number of cases as the target variable. By evaluating the models' performance using metrics like Mean Absolute Error (MAE) and R^2 Score, I compared their accuracy in forecasting future crime rates. This predictive analysis helps in identifying potential future increases in specific crimes and regions, aiding in better decision-making and policy formulation.

6.3 Creating Visualisations

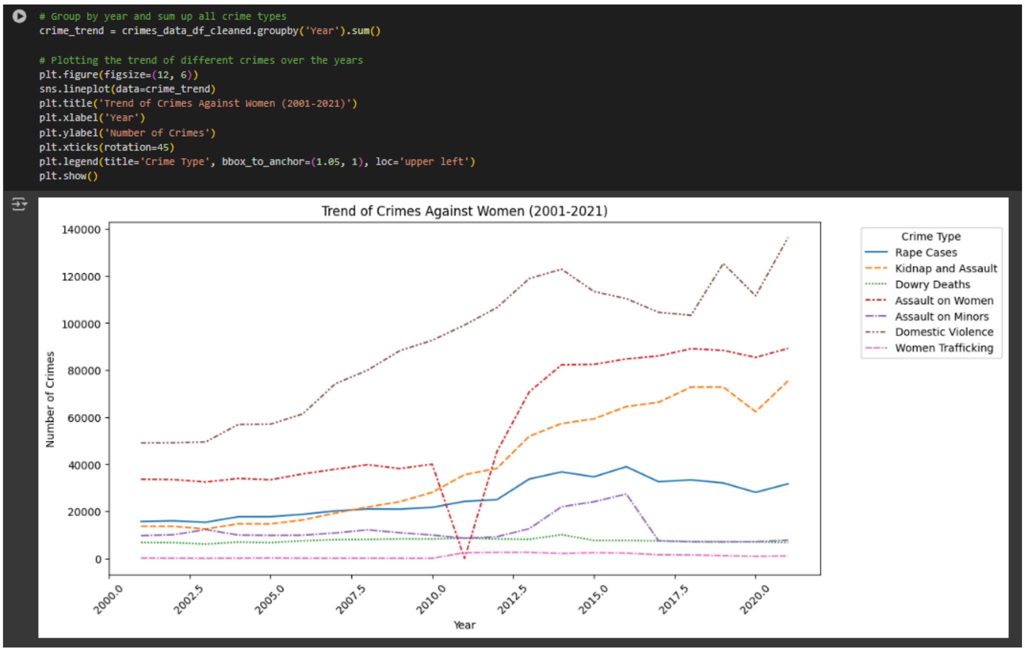
To show different aspects of the data and visually represent all the activities performed in the analysis, are shown through the images below :

- Plotting the datasets.



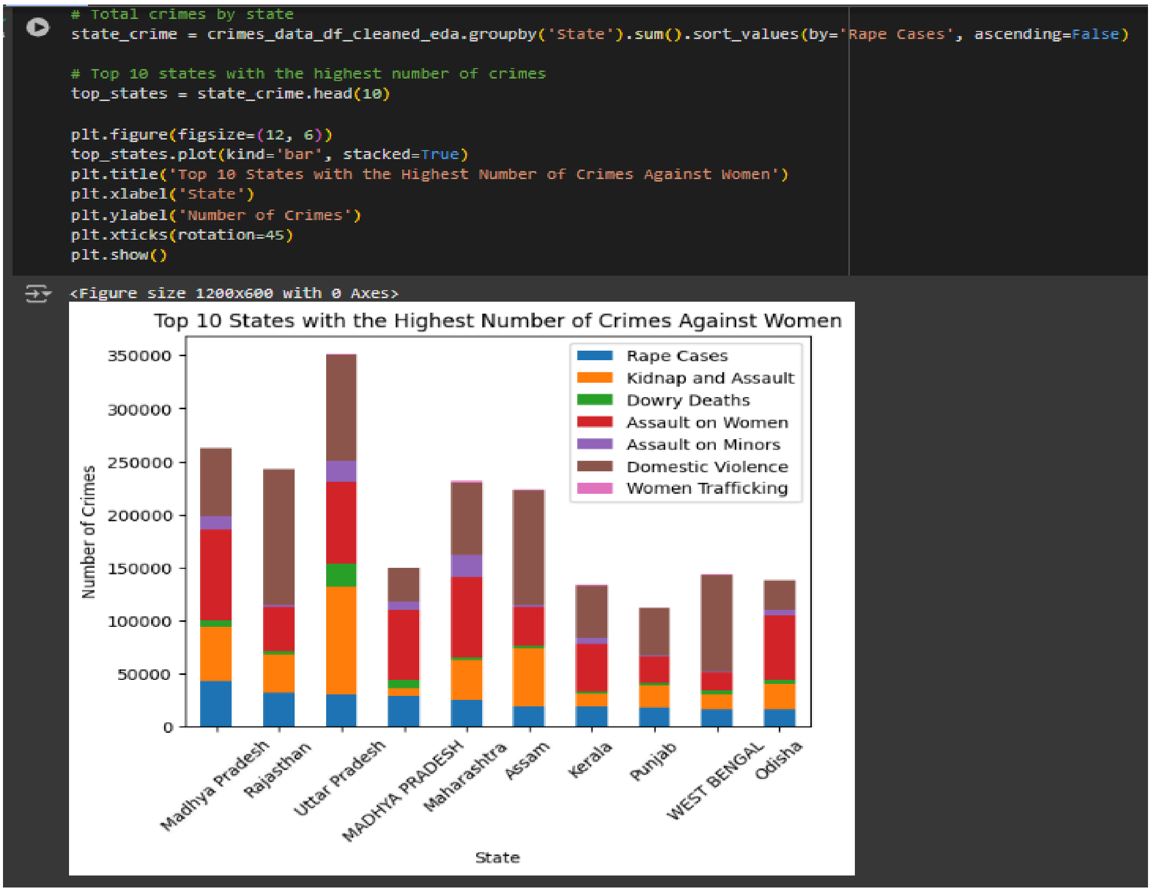
```
[164] # Pivot the data for heatmap
      heatmap_data = crimes_data_df_cleaned.pivot_table(values='Rape Cases', index='State', columns='Year', aggfunc='sum', fill_value=0)

      plt.figure(figsize=(15, 10))
      sns.heatmap(heatmap_data, cmap="YlGnBu", linecolor='white', linewidths=0.5)
      plt.title('Heatmap of Rape Cases by State and Year')
      plt.xlabel('Year')
      plt.ylabel('State')
      plt.show()
```

Figure 6.1: Heatmap of Rape Cases in India by State and Year (2001–2021)



```
# Group by year and sum up all crime types
crime_trend = crimes_data_df_cleaned.groupby('Year').sum()

# Plotting the trend of different crimes over the years
plt.figure(figsize=(12, 6))
sns.lineplot(data=crime_trend)
plt.title('Trend of Crimes Against Women (2001-2021)')
plt.xlabel('Year')
plt.ylabel('Number of Crimes')
plt.xticks(rotation=45)
plt.legend(title='Crime Type', bbox_to_anchor=(1.05, 1), loc='upper left')
plt.show()
```

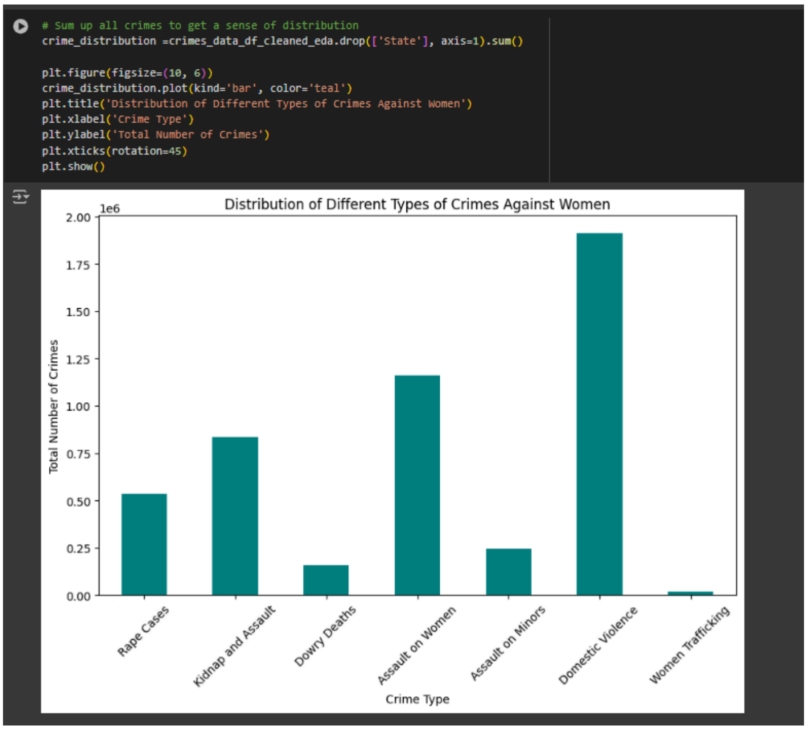
Figure 6.2: Trend of Crimes Against Women in India by Category (2001–2021)



```
# Total crimes by state
state_crime = crimes_data_df_cleaned_eda.groupby('State').sum().sort_values(by='Rape Cases', ascending=False)

# Top 10 states with the highest number of crimes
top_states = state_crime.head(10)

plt.figure(figsize=(12, 6))
top_states.plot(kind='bar', stacked=True)
plt.title('Top 10 States with the Highest Number of Crimes Against Women')
plt.xlabel('State')
plt.ylabel('Number of Crimes')
plt.xticks(rotation=45)
plt.show()
```

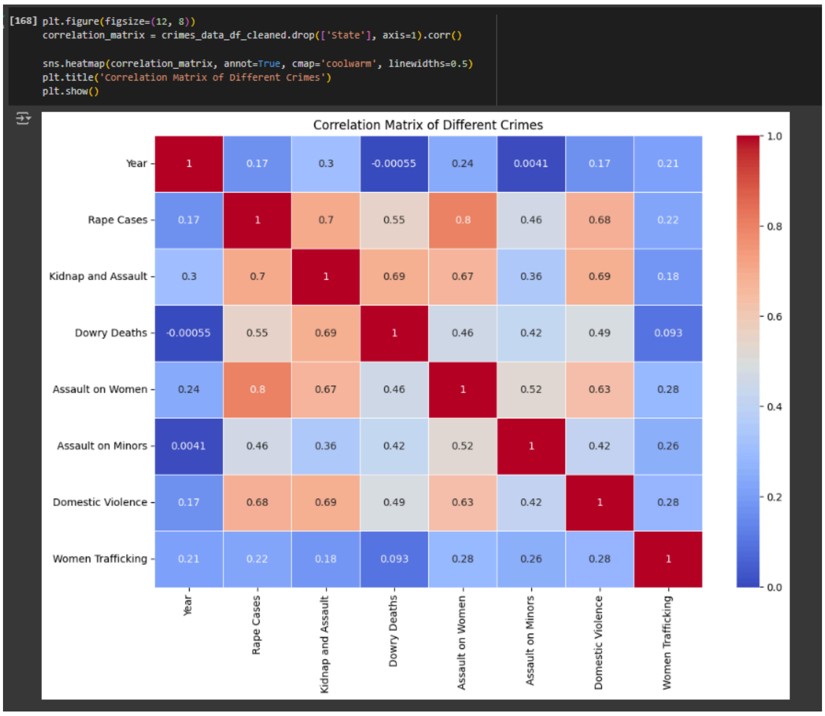
```
<Figure size 1200x600 with 0 Axes>
```

Figure 6.3: Top 10 States with the Highest Number of Crimes Against Women (2001–2021)



```
# Sum up all crimes to get a sense of distribution
crime_distribution =crimes_data_df_cleaned_eda.drop(['State'], axis=1).sum()

plt.figure(figsize=(10, 6))
crime_distribution.plot(kind='bar', color='teal')
plt.title('Distribution of Different Types of Crimes Against Women')
plt.xlabel('Crime Type')
plt.ylabel('Total Number of Crimes')
plt.xticks(rotation=45)
plt.show()
```

Figure 6.4: Distribution of Different Types of Crimes Against Women in India (2001–2021)

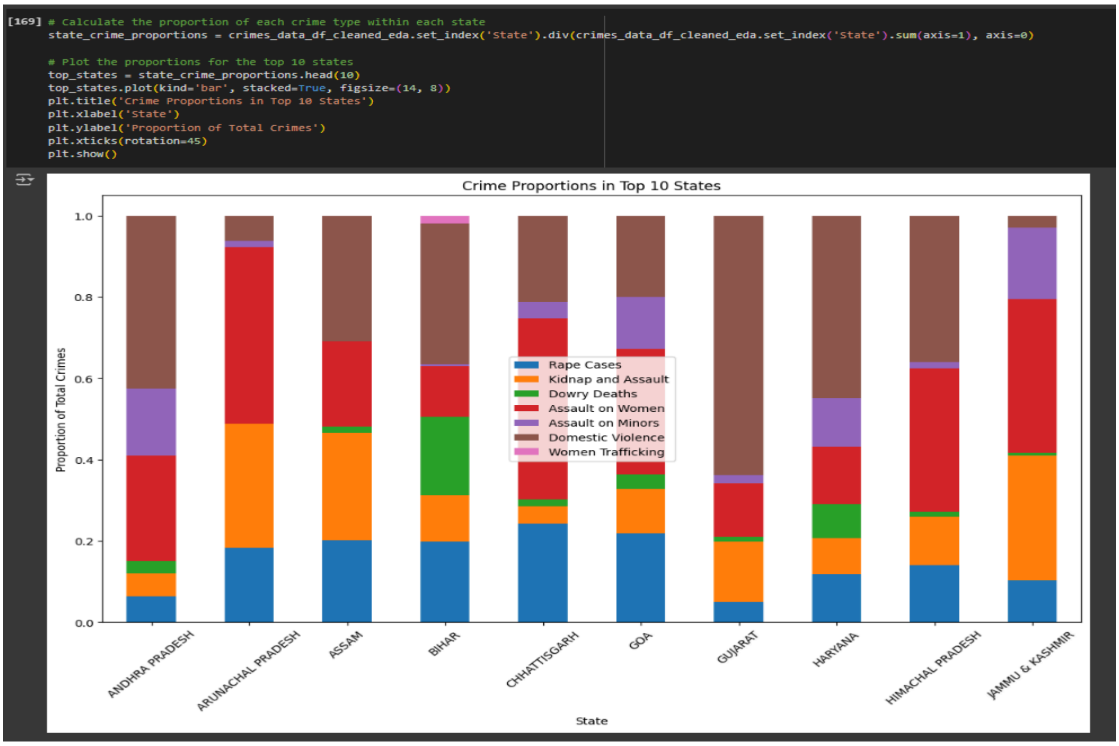


```
[168] plt.figure(figsize=(12, 8))
      correlation_matrix = crimes_data_df_cleaned.drop(['State'], axis=1).corr()

      sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', linewidths=0.5)
      plt.title('Correlation Matrix of Different Crimes')
      plt.show()
```

Correlation Matrix of Different Crimes

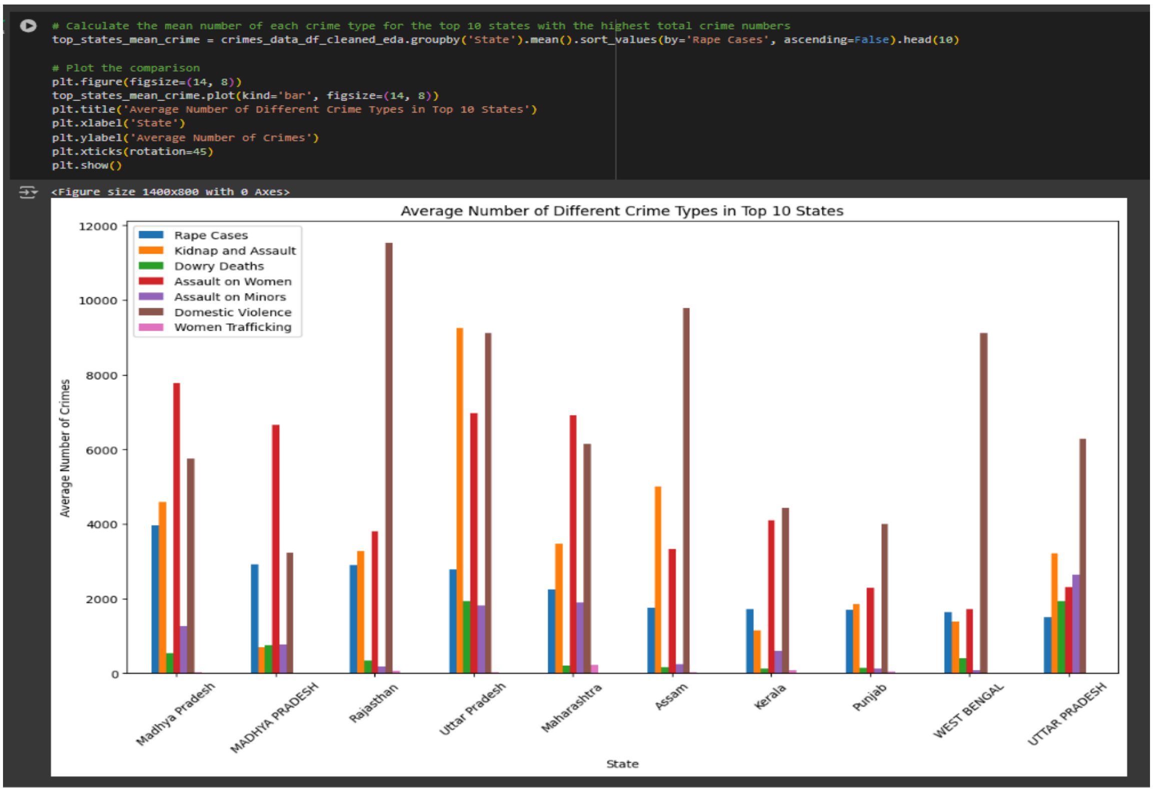
	Year	Rape Cases	Kidnap and Assault	Dowry Deaths	Assault on Women	Assault on Minors	Domestic Violence	Women Trafficking
Year	1	0.17	0.3	-0.00055	0.24	0.0041	0.17	0.21
Rape Cases	0.17	1	0.7	0.55	0.8	0.46	0.68	0.22
Kidnap and Assault	0.3	0.7	1	0.69	0.67	0.36	0.69	0.18
Dowry Deaths	-0.00055	0.55	0.69	1	0.46	0.42	0.49	0.093
Assault on Women	0.24	0.8	0.67	0.46	1	0.52	0.63	0.28
Assault on Minors	0.0041	0.46	0.36	0.42	0.52	1	0.42	0.26
Domestic Violence	0.17	0.68	0.69	0.49	0.63	0.42	1	0.28
Women Trafficking	0.21	0.22	0.18	0.093	0.28	0.26	0.28	1

Figure 6.5: Correlation Matrix of Different Types of Crimes Against Women



```
[169] # Calculate the proportion of each crime type within each state
      state_crime_proportions = crimes_data_df_cleaned_eda.set_index('State').div(crimes_data_df_cleaned_eda.set_index('State').sum(axis=1), axis=0)

      # Plot the proportions for the top 10 states
      top_states = state_crime_proportions.head(10)
      top_states.plot(kind='bar', stacked=True, figsize=(14, 8))
      plt.title('Crime Proportions in Top 10 States')
      plt.xlabel('State')
      plt.ylabel('Proportion of Total Crimes')
      plt.xticks(rotation=45)
      plt.show()
```

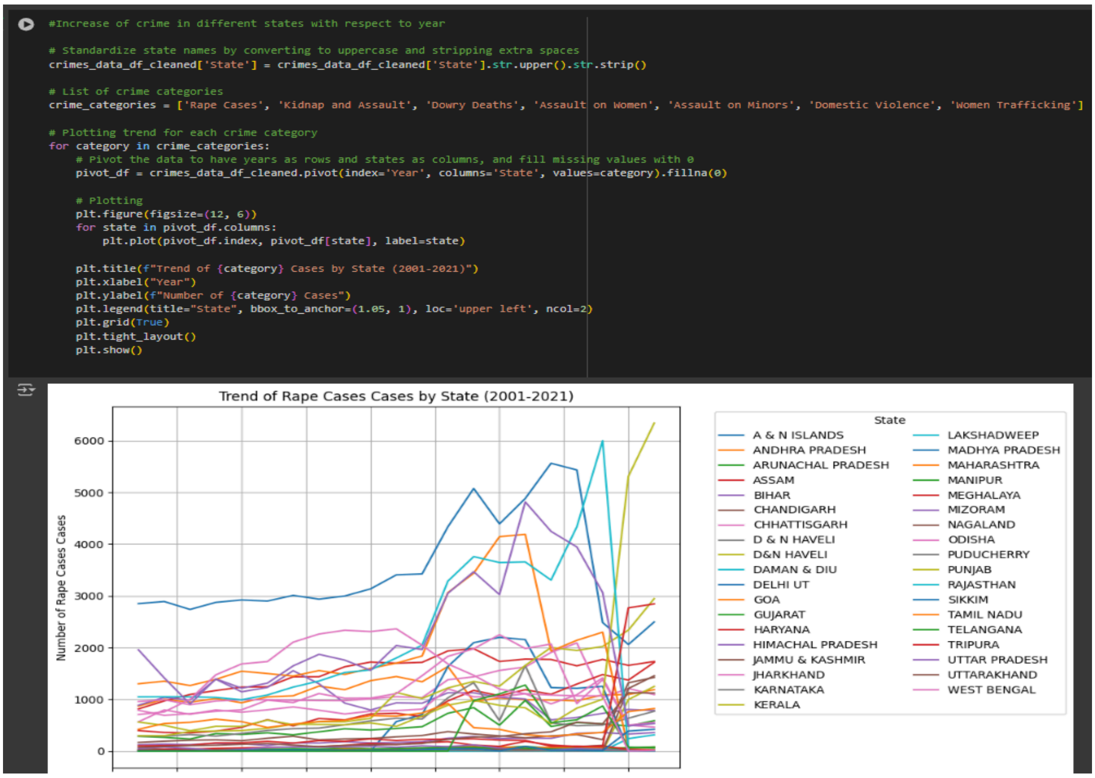
Figure 6.6: Proportion of Different Types of Crimes Against Women in Top 10 States



```
# Calculate the mean number of each crime type for the top 10 states with the highest total crime numbers
top_states_mean_crime = crimes_data_df_cleaned_eda.groupby('State').mean().sort_values(by='Rape Cases', ascending=False).head(10)

# Plot the comparison
plt.figure(figsize=(14, 8))
top_states_mean_crime.plot(kind='bar', figsize=(14, 8))
plt.title('Average Number of Different Crime Types in Top 10 States')
plt.xlabel('State')
plt.ylabel('Average Number of Crimes')
plt.xticks(rotation=45)
plt.show()
```

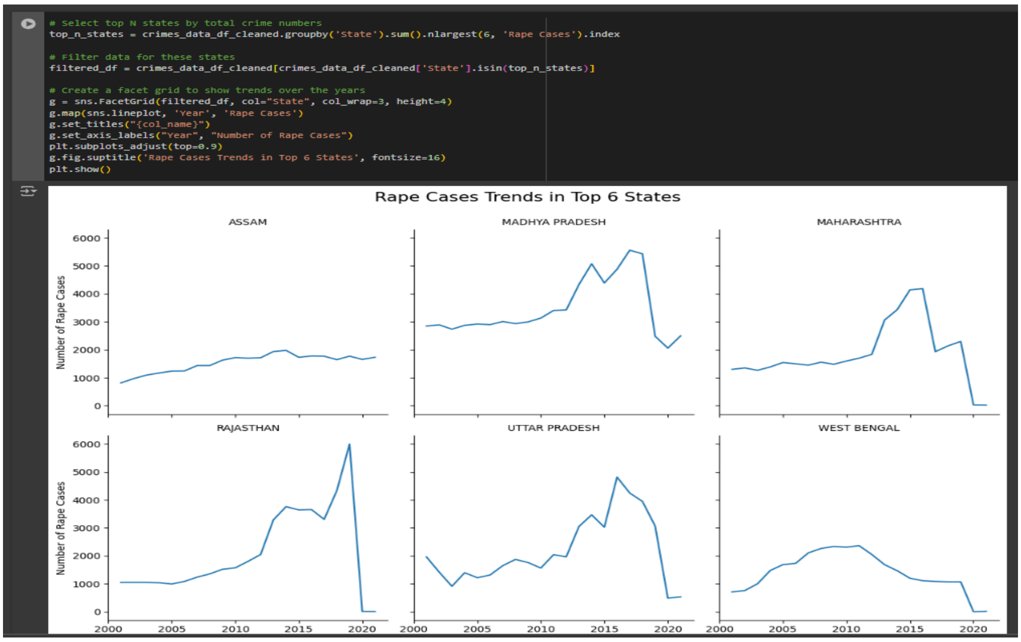
```
<Figure size 1400x800 with 0 Axes>
```

Figure 6.7: Average Number of Different Crime Types in Top 10 States (2001–2021)



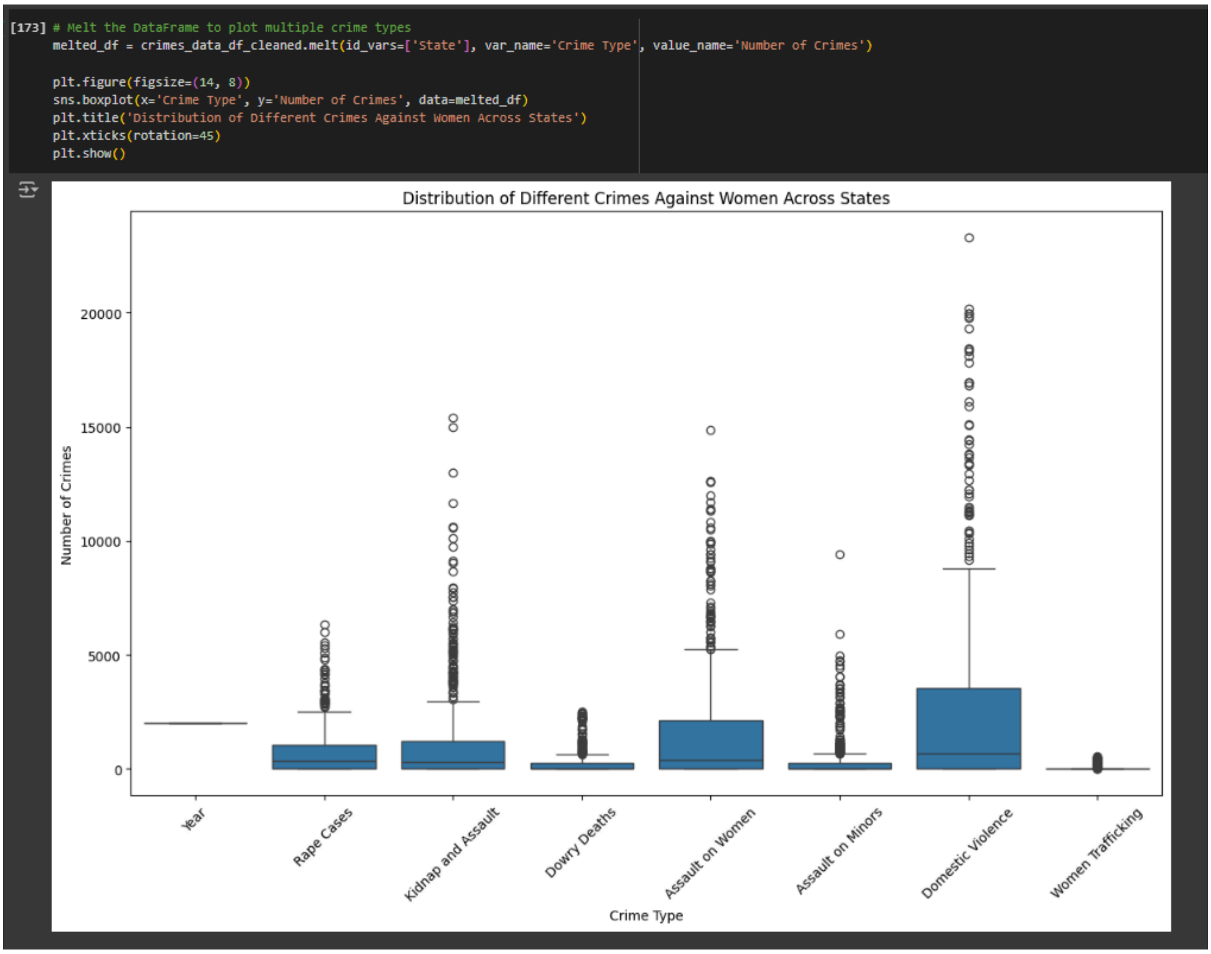
```
#Increase of crime in different states with respect to year

# Standardize state names by converting to uppercase and stripping extra spaces
crimes_data_df_cleaned['State'] = crimes_data_df_cleaned['State'].str.upper().str.strip()

# List of crime categories
crime_categories = ['Rape Cases', 'Kidnap and Assault', 'Dowry Deaths', 'Assault on Women', 'Assault on Minors', 'Domestic Violence', 'Women Trafficking']

# Plotting trend for each crime category
for category in crime_categories:
    # Pivot the data to have years as rows and states as columns, and fill missing values with 0
    pivot_df = crimes_data_df_cleaned.pivot(index='Year', columns='State', values=category).fillna(0)

    # Plotting
    plt.figure(figsize=(12, 6))
    for state in pivot_df.columns:
        plt.plot(pivot_df.index, pivot_df[state], label=state)

    plt.title(f"Trend of {category} Cases by State (2001-2021)")
    plt.xlabel("Year")
    plt.ylabel(f"Number of {category} Cases")
    plt.legend(title="State", bbox_to_anchor=(1.05, 1), loc='upper left', ncol=2)
    plt.grid(True)
    plt.tight_layout()
    plt.show()
```

Figure 6.8: Trend of Different Cases by State (2001–2021)



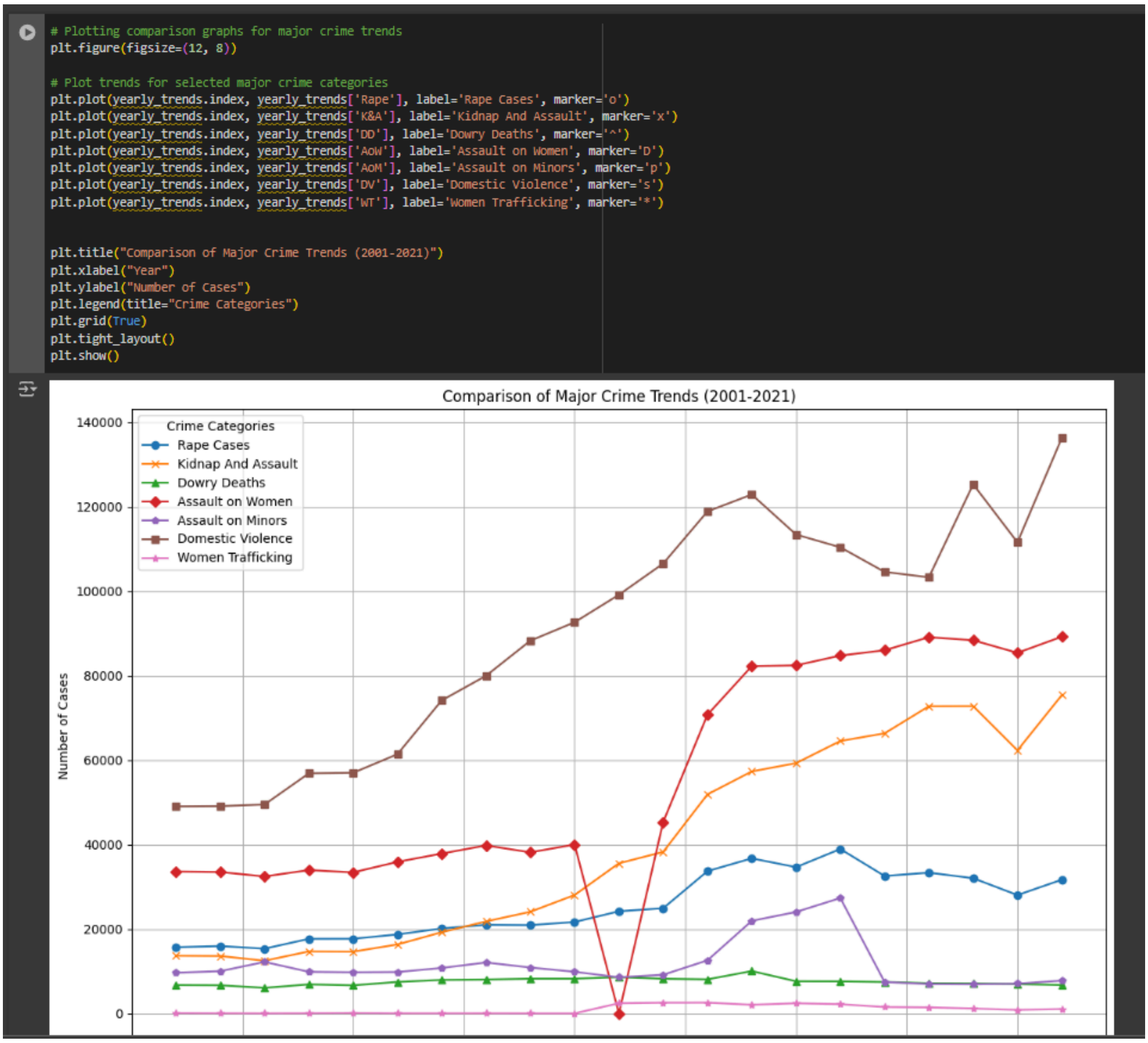
```
# Select top N states by total crime numbers
top_n_states = crimes_data_df_cleaned.groupby('State').sum().nlargest(6, 'Rape Cases').index

# Filter data for these states
filtered_df = crimes_data_df_cleaned[crimes_data_df_cleaned['State'].isin(top_n_states)]

# Create a facet grid to show trends over the years
g = sns.FacetGrid(filtered_df, col="State", col_wrap=3, height=4)
g.map(sns.lineplot, 'Year', 'Rape Cases')
g.set_titles("{col_name}")
g.set_axis_labels("Year", "Number of Rape Cases")
plt.subplots_adjust(top=0.9)
g.fig.suptitle('Rape Cases Trends in Top 6 States', fontsize=16)
plt.show()
```

Figure 6.9: Rape Cases Trends in Top 6 States (2001–2021)



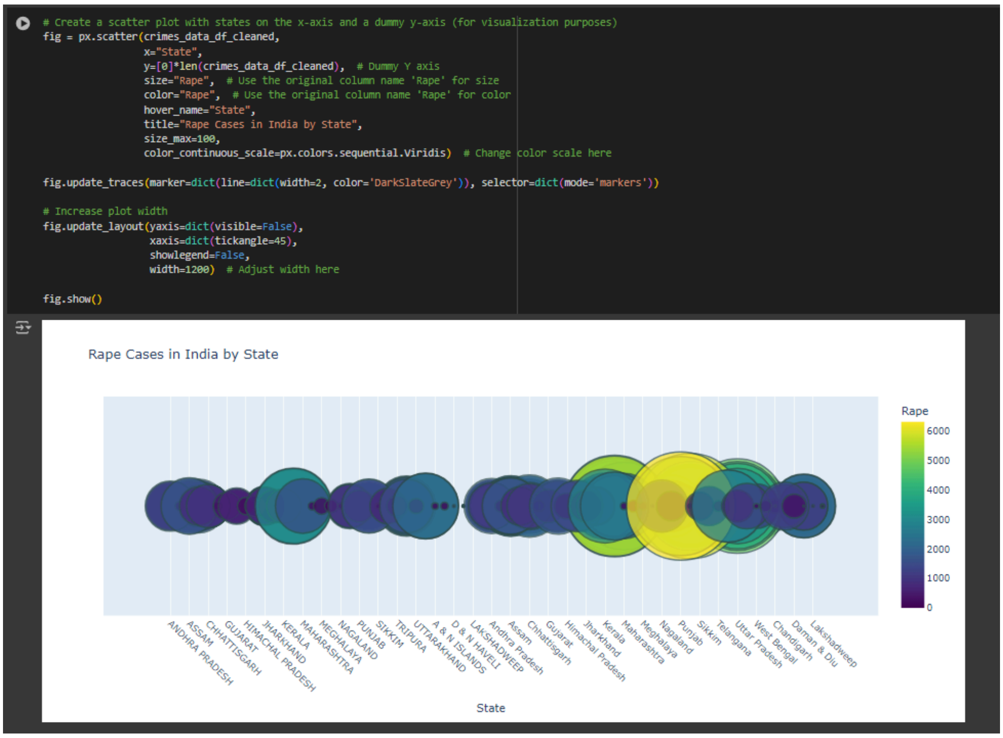
```
[173] # Melt the DataFrame to plot multiple crime types
      melted_df = crimes_data_df_cleaned.melt(id_vars=['State'], var_name='Crime Type', value_name='Number of Crimes')

      plt.figure(figsize=(14, 8))
      sns.boxplot(x='Crime Type', y='Number of Crimes', data=melted_df)
      plt.title('Distribution of Different Crimes Against Women Across States')
      plt.xticks(rotation=45)
      plt.show()
```

Figure 6.10: Distribution of Different Types of Crimes Against Women Across States (2001–2021)



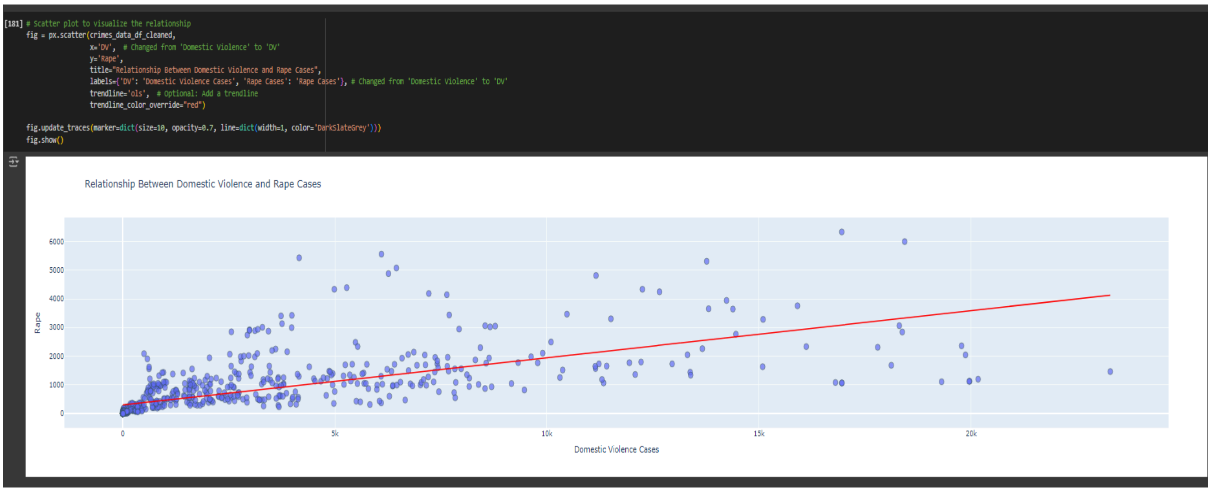
```
# Plotting comparison graphs for major crime trends
plt.figure(figsize=(12, 8))

# Plot trends for selected major crime categories
plt.plot(yearly_trends.index, yearly_trends['Rape'], label='Rape Cases', marker='o')
plt.plot(yearly_trends.index, yearly_trends['K&A'], label='Kidnap And Assault', marker='x')
plt.plot(yearly_trends.index, yearly_trends['DD'], label='Dowry Deaths', marker='^')
plt.plot(yearly_trends.index, yearly_trends['AoW'], label='Assault on Women', marker='D')
plt.plot(yearly_trends.index, yearly_trends['AoM'], label='Assault on Minors', marker='p')
plt.plot(yearly_trends.index, yearly_trends['DV'], label='Domestic Violence', marker='s')
plt.plot(yearly_trends.index, yearly_trends['WT'], label='Women Trafficking', marker='*')


plt.title("Comparison of Major Crime Trends (2001-2021)")
plt.xlabel("Year")
plt.ylabel("Number of Cases")
plt.legend(title="Crime Categories")
plt.grid(True)
plt.tight_layout()
plt.show()
```

Figure 6.11: Comparison of Major Crime Trends Against Women (2001–2021)



```
# Create a scatter plot with states on the x-axis and a dummy y-axis (for visualization purposes)
fig = px.scatter(crimes_data_df_cleaned,
                 x="State",
                 y=[0]*len(crimes_data_df_cleaned),  # Dummy Y axis
                 size="Rape",  # Use the original column name 'Rape' for size
                 color="Rape",  # Use the original column name 'Rape' for color
                 hover_name="State",
                 title="Rape Cases in India by State",
                 size_max=100,
                 color_continuous_scale=px.colors.sequential.Viridis)  # Change color scale here

fig.update_traces(marker=dict(line=dict(width=2, color='DarkSlateGrey')), selector=dict(mode='markers'))

# Increase plot width
fig.update_layout(yaxis=dict(visible=False),
                  xaxis=dict(tickangle=45),
                  showlegend=False,
                  width=1200)  # Adjust width here

fig.show()
```

Figure 6.12: Rape Cases in India by State (2001–2021)

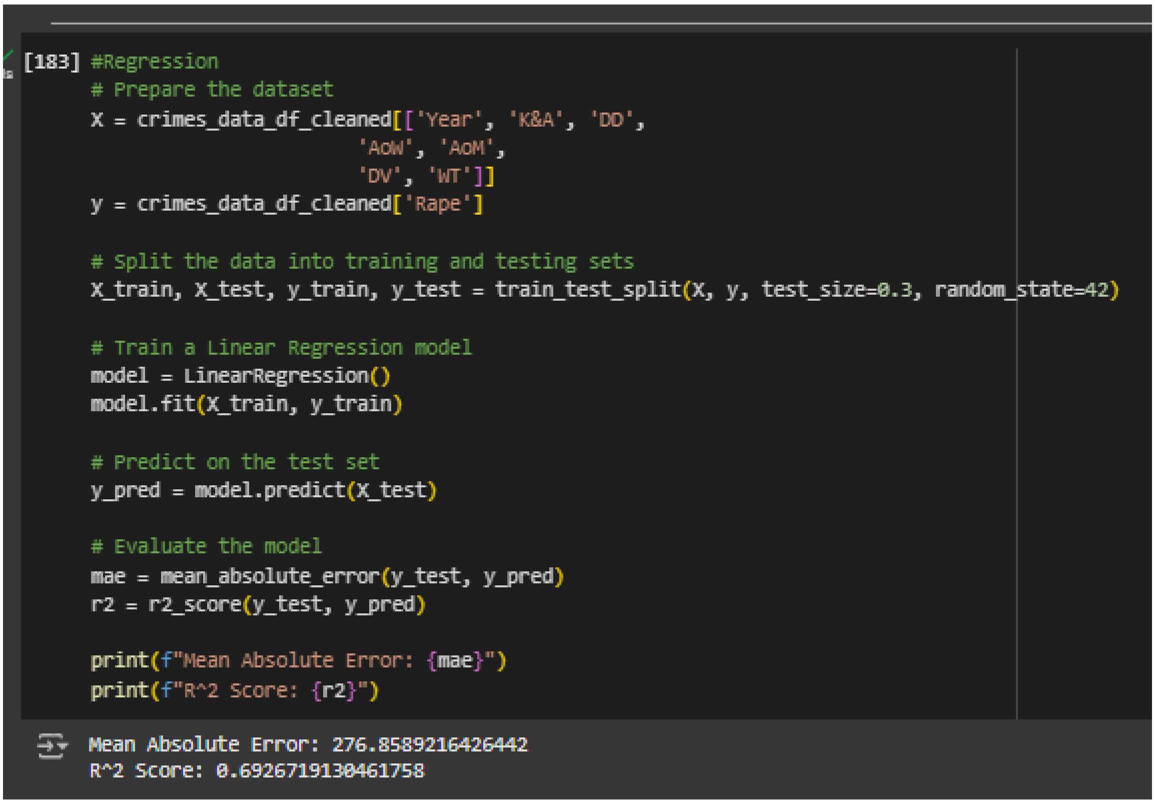


```
[181] # Scatter plot to visualize the relationship
      fig = px.scatter(crimes_data_df_cleaned,
                       x='DV',  # Changed from 'Domestic Violence' to 'DV'
                       y='Rape',
                       title="Relationship Between Domestic Violence and Rape Cases",
                       labels={'DV': 'Domestic Violence Cases', 'Rape Cases': 'Rape Cases'}, # Changed from 'Domestic Violence' to 'DV'
                       trendline='ols',  # Optional: Add a trendline
                       trendline_color_override="red")

      fig.update_traces(marker=dict(size=10, opacity=0.7, line=dict(width=1, color='DarkSlateGrey')))
      fig.show()
```

Figure 6.13: Relationship Between Domestic Violence and Rape Cases (2001–2021)

6.3 Creating Models

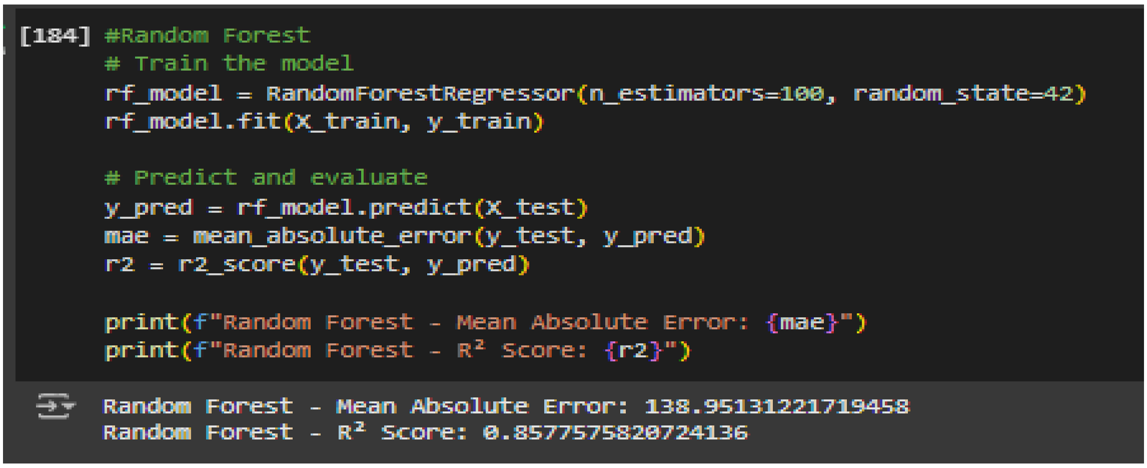
To demonstrate the various aspects of the data analysis and visually represent the predictive modeling activities performed in the project, several models were utilized, and their outputs are presented through the images below. These models, including Linear Regression, Random Forest Regressor, Gradient Boosting Regressor, and XGBoost Regressor, were applied to forecast future crime trends and evaluate the relationships between different types of crimes. The figures showcase the accuracy of these models and help highlight key insights derived from the analysis.



```
[183] #Regression
      # Prepare the dataset
      X = crimes_data_df_cleaned[['Year', 'K&A', 'DD',
                                  'AoW', 'AoM',
                                  'DV', 'WT']]
      y = crimes_data_df_cleaned['Rape']

      # Split the data into training and testing sets
      X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=42)

      # Train a Linear Regression model
      model = LinearRegression()
      model.fit(X_train, y_train)

      # Predict on the test set
      y_pred = model.predict(X_test)

      # Evaluate the model
      mae = mean_absolute_error(y_test, y_pred)
      r2 = r2_score(y_test, y_pred)

      print(f"Mean Absolute Error: {mae}")
      print(f"R^2 Score: {r2}")
```

```
Mean Absolute Error: 276.8589216426442
R^2 Score: 0.6926719130461758
```

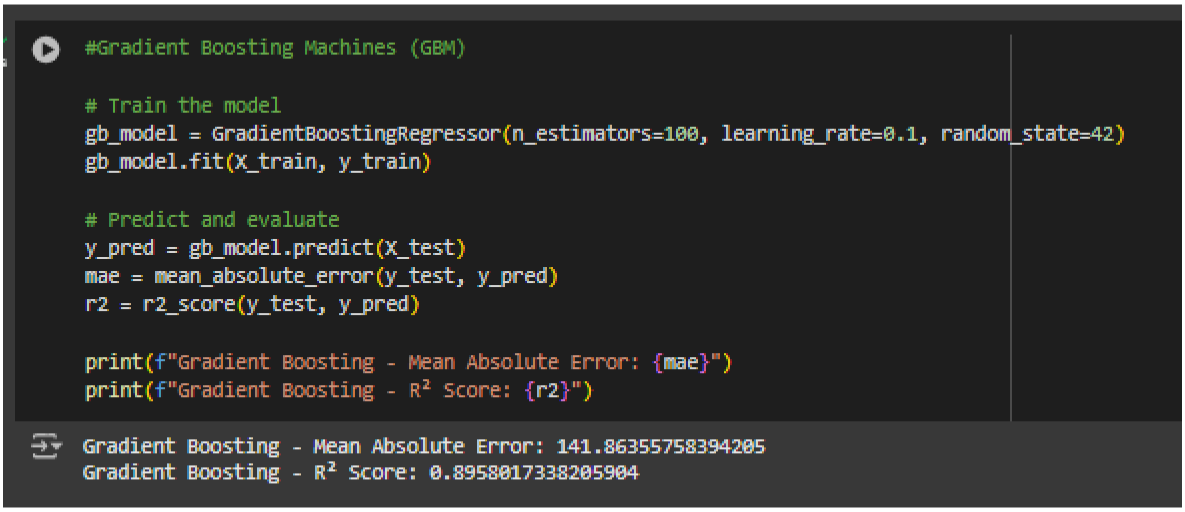
Figure 6.14: Linear Regression Model to Predict Rape Cases Based on Other Crime Categories



```
[184] #Random Forest
      # Train the model
      rf_model = RandomForestRegressor(n_estimators=100, random_state=42)
      rf_model.fit(X_train, y_train)

      # Predict and evaluate
      y_pred = rf_model.predict(X_test)
      mae = mean_absolute_error(y_test, y_pred)
      r2 = r2_score(y_test, y_pred)

      print(f"Random Forest - Mean Absolute Error: {mae}")
      print(f"Random Forest - R² Score: {r2}")
```

```
Random Forest - Mean Absolute Error: 138.95131221719458
Random Forest - R² Score: 0.8577575820724136
```

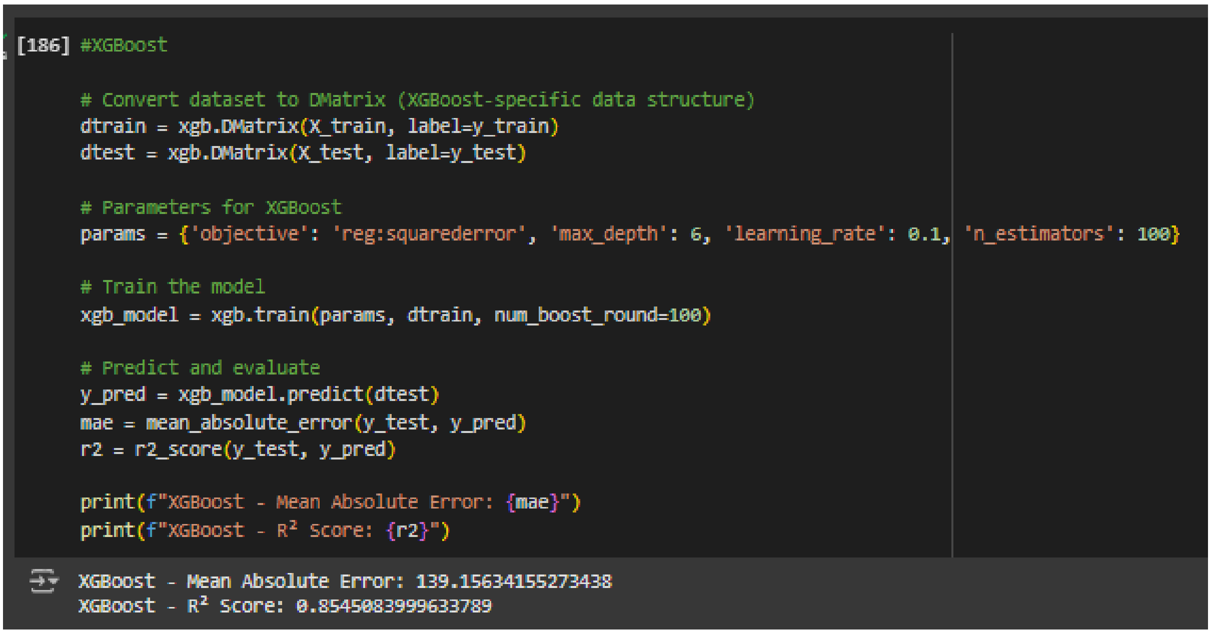
Figure 6.15: Random Forest Regressor Model to Predict Rape Cases and Evaluate Performance



```
#Gradient Boosting Machines (GBM)

# Train the model
gb_model = GradientBoostingRegressor(n_estimators=100, learning_rate=0.1, random_state=42)
gb_model.fit(X_train, y_train)

# Predict and evaluate
y_pred = gb_model.predict(X_test)
mae = mean_absolute_error(y_test, y_pred)
r2 = r2_score(y_test, y_pred)

print(f"Gradient Boosting - Mean Absolute Error: {mae}")
print(f"Gradient Boosting - R² Score: {r2}")
```

```
Gradient Boosting - Mean Absolute Error: 141.86355758394205
Gradient Boosting - R² Score: 0.8958017338205904
```

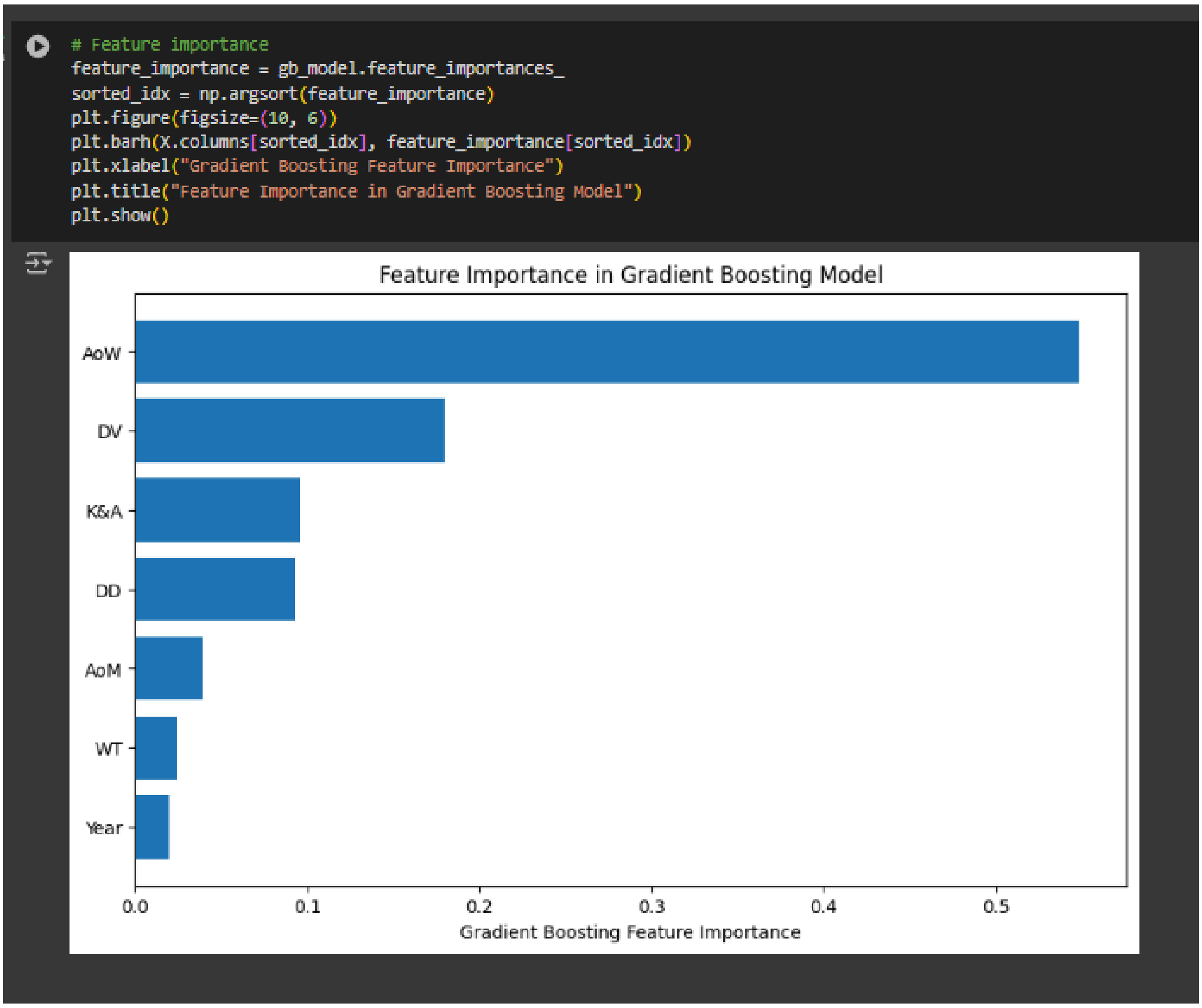
Figure 6.16: Gradient Boosting Regressor Model to Predict Rape Cases and Evaluate Accuracy



```
[186] #XGBoost

# Convert dataset to DMatrix (XGBoost-specific data structure)
dtrain = xgb.DMatrix(X_train, label=y_train)
dtest = xgb.DMatrix(X_test, label=y_test)

# Parameters for XGBoost
params = {'objective': 'reg:squarederror', 'max_depth': 6, 'learning_rate': 0.1, 'n_estimators': 100}

# Train the model
xgb_model = xgb.train(params, dtrain, num_boost_round=100)

# Predict and evaluate
y_pred = xgb_model.predict(dtest)
mae = mean_absolute_error(y_test, y_pred)
r2 = r2_score(y_test, y_pred)

print(f"XGBoost - Mean Absolute Error: {mae}")
print(f"XGBoost - R² Score: {r2}")
```

```
XGBoost - Mean Absolute Error: 139.15634155273438
XGBoost - R² Score: 0.8545083999633789
```

Figure 6.17: XGBoost Regressor Model to Predict Rape Cases and Evaluate Performance



```
# Feature importance
feature_importance = gb_model.feature_importances_
sorted_idx = np.argsort(feature_importance)
plt.figure(figsize=(10, 6))
plt.barh(X.columns[sorted_idx], feature_importance[sorted_idx])
plt.xlabel("Gradient Boosting Feature Importance")
plt.title("Feature Importance in Gradient Boosting Model")
plt.show()
```

Feature Importance in Gradient Boosting Model
AoW
DV
K&A
DD
AoM
WT
Year
0.0
0.1
0.2
0.3
0.4
0.5
Gradient Boosting Feature Importance

Figure 6.18: Feature Importance in Gradient Boosting Model for Predicting Rape Cases

Chapter 7 Conclusion and Key Insights

This project focused on analyzing crimes against women in India over a period of two decades (2001–2021), utilizing data-driven methods to explore various crime categories, trends, and patterns. By employing advanced data analysis techniques and machine learning models, this study provides valuable insights into the nature and extent of crimes against women across different states and timeframes.

The analysis was conducted using a comprehensive dataset that included multiple crime categories such as **rape cases**, **domestic violence**, **dowry deaths**, **kidnap and assault**, **assault on women and minors**, and **women trafficking**. Several visualizations were created to represent the data effectively, including **heatmaps**, **line graphs**, **scatter plots**, and **box plots**, highlighting key aspects of the data. Furthermore, predictive models such as **Linear Regression**, **Random Forest Regressor**, **Gradient Boosting Regressor**, and **XGBoost Regressor** were applied to forecast crime trends and assess the relationships between different crime categories.

Key Insights from the Analysis

1. **Overall Increase in Crimes Against Women**: The data indicates a general upward trend in crimes against women across India from 2001 to 2021. Notably, categories such as **domestic violence** and **assault on women** have shown significant increases over time. This suggests that either the prevalence of these crimes has increased, or there has been an improvement in reporting mechanisms, leading to more cases being officially recorded.
2. **State-Wise Crime Trends**: The state-wise analysis revealed that certain states, such as **Madhya Pradesh**, **Rajasthan**, **Uttar Pradesh**, **Maharashtra**, and **West Bengal**, consistently reported high numbers of crimes across multiple categories. In contrast, smaller states and union territories like **Lakshadweep**, **Nagaland**, and **Sikkim** reported relatively lower numbers of crimes. This regional disparity highlights the need for targeted interventions and policy measures in high-crime states.
3. **Crime-Specific Patterns**:
    - **Rape Cases**: States like **Madhya Pradesh**, **Rajasthan**, and **Uttar Pradesh** reported the highest number of rape cases. The trend analysis indicates a gradual increase in rape cases over the years, peaking around 2015.
    - **Domestic Violence**: Domestic violence cases showed the highest overall numbers compared to other crime categories. This emphasizes the urgent need for effective domestic violence prevention programs and support systems.
    - **Dowry Deaths**: Although dowry deaths have not shown a sharp increase over time, they remain a significant concern in states like **Uttar Pradesh** and **Bihar**.
    - **Women Trafficking**: Trafficking cases, while fewer in number compared to other categories, were concentrated in certain regions, particularly in northeastern and border states like **West Bengal** and **Assam**.

4. **Correlation Between Crime Categories**: The correlation analysis revealed strong relationships between certain crime types. For instance, **domestic violence** and **rape cases** showed a positive correlation, indicating that regions with high domestic violence cases also tend to report more rape cases. This insight can help policymakers design integrated intervention strategies addressing multiple forms of violence against women.
5. **Feature Importance in Predictive Models**: The feature importance analysis using the **Gradient Boosting Regressor** highlighted that **assault on women (AoW)** and **domestic violence (DV)** were the most significant predictors of rape cases. This suggests that addressing these specific crimes could potentially lead to a reduction in rape cases as well.
6. **Model Performance and Prediction**: Among the predictive models used, **Gradient Boosting Regressor** and **Random Forest Regressor** provided the best performance, with **R^2 scores** of 0.89 and 0.86, respectively. These models can be used to forecast future crime trends and aid in proactive policymaking by identifying regions and crime categories likely to experience a surge in reported cases.

Recommendations

Based on the insights derived from this analysis, several recommendations can be made:

1. **Targeted Interventions in High-Crime States**: States such as **Madhya Pradesh**, **Rajasthan**, and **Uttar Pradesh** require targeted interventions to address the high prevalence of crimes against women. This could include stricter law enforcement, improved support services, and awareness campaigns.
2. **Strengthening Domestic Violence Support Systems**: Given the high incidence of domestic violence, there is a need to strengthen support systems for victims, including expanding helplines, shelters, and legal aid services. Special attention should be given to ensuring that these services are accessible in rural and remote areas.
3. **Integrated Policy Approaches**: The positive correlation between different crime categories suggests that an integrated approach to policy formulation is essential. For example, programs aimed at reducing domestic violence should also address sexual violence and dowry-related crimes.
4. **Enhancing Reporting Mechanisms**: Efforts should be made to improve crime reporting mechanisms, particularly in states with historically low reporting rates. This includes creating a safer environment for victims to come forward and ensuring that law enforcement agencies handle cases sensitively.
5. **Use of Predictive Models for Crime Prevention**: The predictive models developed in this project can be further refined and deployed by law enforcement agencies to anticipate future crime trends. This will allow for timely interventions and better resource allocation.

Conclusion

This project demonstrates the power of data-driven approaches in understanding and addressing social issues such as crimes against women. By leveraging statistical analysis, machine learning models, and visualizations, we have uncovered critical insights that can inform future policies and interventions. However, it is important to note that while data analysis can guide decision-making, effective implementation requires collaboration between government agencies, law enforcement, civil society, and the community.

Future work could focus on incorporating additional datasets, such as socio-economic factors, literacy rates, and employment statistics, to gain a more holistic understanding of the factors contributing to crimes against women. Additionally, real-time crime data could be used to build dynamic models capable of providing continuous updates and predictions.

Ultimately, the goal is to create a safer environment for women in India by combining data-driven insights with effective policy implementation and social change.

References

1. National Crime Records Bureau (NCRB) - Official Reports on Crimes Against Women in India: NCRB Website
2. Python Data Science Libraries Documentation: Pandas Documentation, NumPy Documentation
3. Machine Learning Models - Scikit-Learn Documentation: Scikit-Learn Documentation
4. Gradient Boosting and XGBoost Algorithms: Gradient Boosting Documentation, XGBoost Documentation
5. Seaborn and Matplotlib for Data Visualization: Seaborn Documentation, Matplotlib Documentation
6. Plotly for Interactive Visualizations: Plotly Documentation