1.**INTRODUCTION**

Crime in India is different in different places because of how people live and what they believe. Some areas have more crime because people don't have enough money or jobs. When people feel like they don't have what they need, they might do illegal things to survive. Also, what people think is right or wrong can affect how much crime happens. Some people might not trust the police or follow the rules because of their beliefs. Understanding why crime happens helps us make better plans to stop it. By helping people with money and jobs and making sure everyone feels safe and respected, we can make our communities better and reduce crime. This research paper aims to delve into the nuanced factors influencing crime rates across various regions in India, examining the sociological, economic, and criminological dimensions that underpin the prevalence and variations in criminal activities.

2.**OBJECTIVES**

```
# Importing necessary libraries
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

# Loading the dataset
df = pd.read_csv("/content/crime_data.csv")

# Displaying the first few rows of the dataset
df.head()
```

	States/UTs	District	Year	Murder	Attempt to commit Murder	Culpable Homicide not amounting to Murder	Attempt to commit Culpable Homicide	Rape	Custodial Rape
0	Andhra Pradesh	Anantapur	2014	134	171	8	0	35	0
1	Andhra Pradesh	Chittoor	2014	84	170	2	0	32	0
2	Andhra Pradesh	Cuddapah	2014	80	162	1	0	28	0
3	Andhra Pradesh	East Godavari	2014	64	84	2	0	85	0
4	Andhra Pradesh	Guntakal Railway	2014	14	4	0	0	0	0

5 rows × 91 columns

```
format(df.shape) # Displaying the shape of the DataFrame
```

```
'(838, 91)'
```

```
df.info() # Displaying information about the DataFrame
```

```
 61  Dowry Deaths                                                          838 non-nul
 62  Assault on Women with intent to outrage her Modesty                   838 non-nul
 63  Sexual Harassment                                                     838 non-nul
 64  Assault or use of criminal force to women with intent to Disrobe      838 non-nul
 65  Voyeurism                                                             838 non-nul
 66  Stalking                                                              838 non-nul
 67  Other Assault on Women                                                838 non-nul
 68  Insult to the Modesty of Women                                        838 non-nul
 69  At Office premises                                                    838 non-nul
 70  Other places related to work                                          838 non-nul
 71  In Public Transport system                                            838 non-nul
 72  Places other than 231, 232 & 233                                      838 non-nul
 73  Cruelty by Husband or his Relatives                                   838 non-nul
 74  Importation of Girls from Foreign Country                             838 non-nul
 75  Causing Death by Negligence                                           838 non-nul
 76  Deaths due to negligent driving/act                                   838 non-nul
 77  Deaths due to Other Causes                                            838 non-nul
 78  Offences against State                                                838 non-nul
 79  Sedition                                                              838 non-nul
 80  Other offences against State                                          838 non-nul
 81  Offences promoting enmity between different groups                    838 non-nul
 82  Promoting enmity between different groups                             838 non-nul
 83  Imputation, assertions prejudicial to national integration            838 non-nul
 84  Extortion                                                             838 non-nul
 85  Disclosure of Identity of Victims                                     838 non-nul
 86  Incidence of Rash Driving                                             838 non-nul
 87  HumanTrafficking                                                      838 non-nul
 88  Unnatural Offence                                                     838 non-nul
 89  Other IPC crimes                                                      838 non-nul
```

```
df.describe() # Generating descriptive statistics for the DataFrame
```

	Year	Murder	Attempt to commit Murder	Culpable Homicide not amounting to Murder	Attempt to commit Culpable Homicide	Rape	Custodi Ra
count	838.0	838.000000	838.000000	838.000000	838.000000	838.000000	838.0000
mean	2014.0	81.465394	99.995227	7.992840	10.431981	92.310263	0.4749
std	0.0	302.341738	419.969968	52.272628	71.473497	342.630462	6.7564
min	2014.0	0.000000	0.000000	0.000000	0.000000	0.000000	0.0000
25%	2014.0	16.000000	11.000000	0.000000	0.000000	12.000000	0.0000
50%	2014.0	34.500000	31.000000	2.000000	0.000000	35.000000	0.0000
75%	2014.0	62.750000	65.000000	5.000000	0.000000	70.000000	0.0000
max	2014.0	5150.000000	7355.000000	1412.000000	1433.000000	5076.000000	189.0000

8 rows × 89 columns

```
df.loc[:,:] # Accessing all rows and columns of the DataFrame using the .loc indexer
```

	States/UTs	District	Year	Murder	Attempt to commit Murder	Culpable Homicide not amounting to Murder	Attempt to commit Culpable Homicide	Rape	Cust
0	Andhra Pradesh	Anantapur	2014	134	171	8	0	35	
1	Andhra Pradesh	Chittoor	2014	84	170	2	0	32	
2	Andhra Pradesh	Cuddapah	2014	80	162	1	0	28	
3	Andhra Pradesh	East Godavari	2014	64	84	2	0	85	
4	Andhra Pradesh	Guntakal Railway	2014	14	4	0	0	0	
...	...	...	...	...	...	...	...	...	...
833	Lakshadweep	Lakshadweep	2014	0	0	0	0	1	
834	Lakshadweep	Total	2014	0	0	0	0	1	
835	Puducherry	Karaikal	2014	6	1	0	0	3	
836	Puducherry	Puducherry	2014	19	14	3	0	7	
837	Puducherry	Total	2014	25	15	3	0	10	

838 rows × 91 columns

```
df.columns # Accessing the column names of the DataFrame
```

```
Index(['States/UTs', 'District', 'Year', 'Murder', 'Attempt to commit Murder',
       'Culpable Homicide not amounting to Murder',
       'Attempt to commit Culpable Homicide', 'Rape', 'Custodial Rape',
       'Custodial_Gang Rape', 'Custodial_Other Rape',
       'Rape other than Custodial', 'Rape_Gang Rape', 'Rape_Others',
       'Attempt to commit Rape', 'Kidnapping & Abduction_Total',
       'Kidnapping & Abduction', 'Kidnapping & Abduction in order to Murder',
       'Kidnapping for Ransom',
       'Kidnapping & Abduction of Women to compel her for marriage',
       'Other Kidnapping', 'Dacoity', 'Dacoity with Murder', 'Other Dacoity',
       'Making Preparation and Assembly for committing Dacoity', 'Robbery',
       'Criminal Trespass/Burglary', 'Criminal Trespass or Burglary',
       'House Trespass & House Breaking', 'Theft', 'Auto Theft',
       'Other Thefts', 'Unlawful Assembly', 'Riots', 'Riots_Communal',
       'Riots_Industrial', 'Riots_Political', 'Riots_Caste Conflict',
       'Riots_SC/STs Vs Non-SCs/STs', 'Riots_Other Caste Conflict',
       'Riots_Agrarian', 'Riots_Students', 'Riots_Sectarian', 'Riots_Others',
       'Criminal Breach of Trust', 'Cheating', 'Forgery', 'Counterfeiting',
       'Counterfeit Offences related to Counterfeit Coin',
       'Counterfeiting Government Stamp', 'Counterfeit currency & Bank notes',
```

```
       'Counterfeiting currency notes/Bank notes',
       'Using forged or counterfeiting currency/Bank notes',
       'Possession of forged or counterfeiting currency/Bank notes',
       'Making or Possessing materials for forged currency/Bank notes',
       'Making or Using documents resembling currency', 'Arson',
       'Grievous Hurt', 'Hurt', 'Acid attack', 'Attempt to Acid Attack',
       'Dowry Deaths', 'Assault on Women with intent to outrage her Modesty',
       'Sexual Harassment',
       'Assault or use of criminal force to women with intent to Disrobe',
       'Voyeurism', 'Stalking', 'Other Assault on Women',
       'Insult to the Modesty of Women', 'At Office premises',
       'Other places related to work', 'In Public Transport system',
       'Places other than 231, 232 & 233',
       'Cruelty by Husband or his Relatives',
       'Importation of Girls from Foreign Country',
       'Causing Death by Negligence', 'Deaths due to negligent driving/act',
       'Deaths due to Other Causes', 'Offences against State', 'Sedition',
       'Other offences against State',
       'Offences promoting enmity between different groups',
       'Promoting enmity between different groups',
       'Imputation, assertions prejudicial to national integration',
       'Extortion', 'Disclosure of Identity of Victims',
       'Incidence of Rash Driving', 'HumanTrafficking', 'Unnatural Offence',
       'Other IPC crimes', 'Total Cognizable IPC crimes'],
      dtype='object')
```

3.**Correlation Analysis**

In our research, we used correlation analysis to understand how different factors relate to crime rates in India. To ensure accuracy, we focused on numerical data and identified highly correlated variables. Removing these redundant variables helped us avoid problems like overfitting and made our analysis clearer and more reliable. This approach allowed us to provide valuable insights into the complex dynamics of crime in India, informing better policymaking and contributing to the field of criminology.

```
import numpy as np  # Import NumPy library for numerical computations

# Select only numeric columns for correlation calculation
numeric_df = df.select_dtypes(include=['number'])  # Filter the DataFrame to include only

# Calculate correlation matrix
corr_matrix = numeric_df.corr().abs()  # Compute the absolute correlation matrix for nume

# Drop highly correlated columns
upper = corr_matrix.where(np.triu(np.ones(corr_matrix.shape), k=1).astype(bool))  # Creat
to_drop = [column for column in upper.columns if any(upper[column] > 0.95)]  # Identify c
df.drop(to_drop, axis=1, inplace=True)  # Drop highly correlated columns from the DataFra



corr_matrix # Calculate correlation matrix
```

	Year	Murder	Attempt to commit Murder	Culpable Homicide not amounting to Murder	Attempt to commit Culpable Homicide	Rape	Custodial Rape	C
Year	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
Murder	NaN	1.000000	0.884060	0.790658	0.376151	0.832772	0.563463	
Attempt to commit Murder	NaN	0.884060	1.000000	0.687222	0.429030	0.777617	0.410184	
Culpable Homicide not amounting to Murder	NaN	0.790658	0.687222	1.000000	0.379721	0.578272	0.908445	
Attempt to commit Culpable Homicide	NaN	0.376151	0.429030	0.379721	1.000000	0.441808	0.231547	
...	...	...	...	...	...	...	...	
Incidence of Rash Driving	NaN	0.439894	0.411060	0.254836	0.631262	0.505905	0.098792	
HumanTrafficking	NaN	0.547694	0.463260	0.203829	0.146832	0.567222	0.003721	
Unnatural Offence	NaN	0.572030	0.461110	0.501062	0.643885	0.721444	0.353102	
Other IPC crimes	NaN	0.842423	0.777138	0.505239	0.438218	0.871093	0.254742	
Total Cognizable IPC crimes	NaN	0.874939	0.809774	0.582874	0.565206	0.908808	0.325657	

89 rows × 89 columns

**Result *

"Murder" has a strong positive correlation with "Total Cognizable IPC crimes" (0.874939), indicating a strong relationship between the number of murder cases and the overall number of cognizable IPC crimes reported.

"Rape" also shows a strong positive correlation with "Total Cognizable IPC crimes" (0.908808), suggesting a strong relationship between the number of rape cases and the overall number of cognizable IPC crimes reported.

"Incidence of Rash Driving" exhibits a moderate positive correlation with "Total Cognizable IPC crimes" (0.725232), indicating a moderate relationship between the incidence of rash driving and the overall number of cognizable IPC crimes reported.

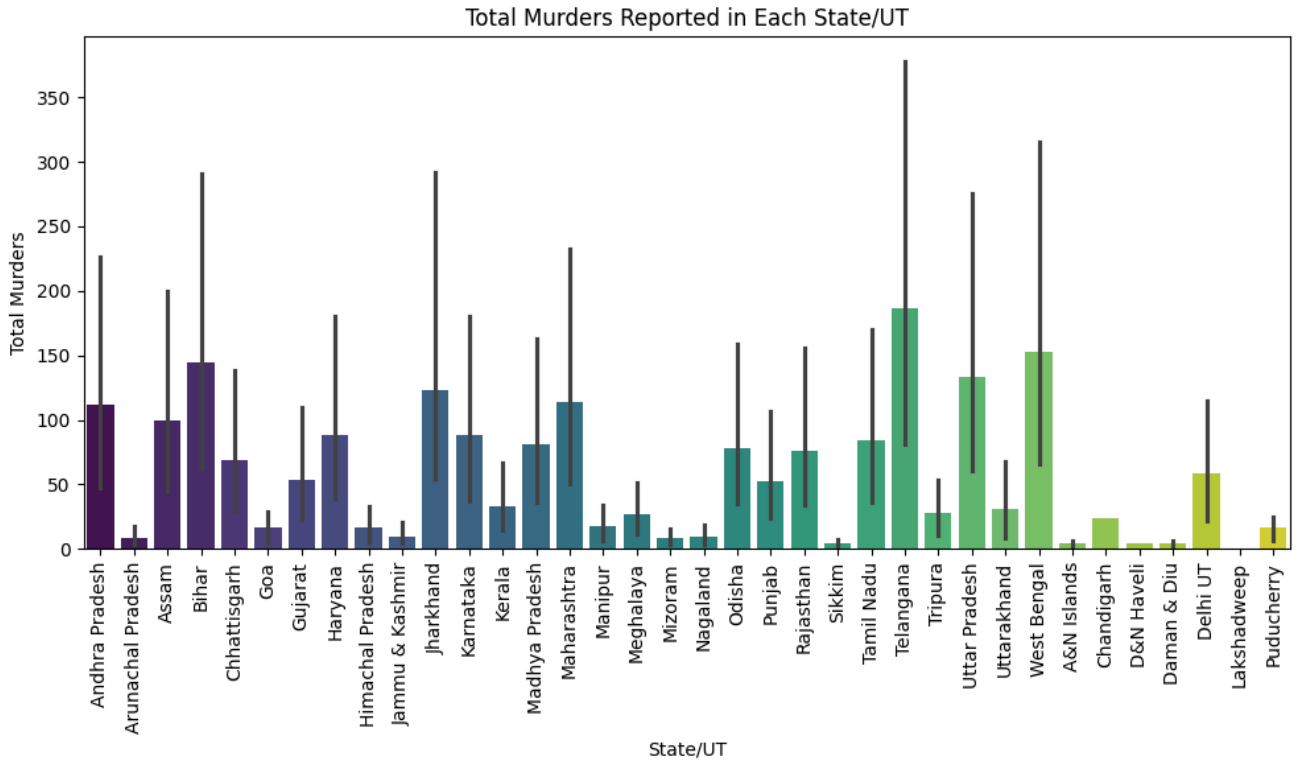
"HumanTrafficking" has a moderate positive correlation with "Total Cognizable IPC crimes" (0.511246), implying a moderate relationship between the occurrence of human trafficking and the overall number of cognizable IPC crimes reported.

4.**DATA VISUALIZATION**

Total No.of Murders

```
import pandas as pd  # Import the pandas library for data manipulation and analysis

df = pd.read_csv("/content/crime_data.csv")  # Read the CSV file containing crime data in

data = df[['States/UTs', 'Murder']]  # Extract only the columns 'States/UTs' and 'Murder'

data.dropna(inplace=True)  # Remove any rows with missing values in the selected columns

plt.figure(figsize=(10, 6))  # Set the figure size for the plot

# Create a bar plot showing the total number of murders reported for each state/UT
sns.barplot(data=data, x='States/UTs', y='Murder', palette='viridis')  # Use Seaborn to c
plt.title('Total Murders Reported in Each State/UT')  # Add a title to the plot
plt.xlabel('State/UT')  # Add a label for the x-axis
plt.ylabel('Total Murders')  # Add a label for the y-axis
plt.xticks(rotation=90)  # Rotate x-axis labels for better readability
plt.tight_layout()  # Adjust layout to prevent overlapping labels
plt.show()  # Display the plot
```

```
<ipython-input-10-5945d3cfff85>:7: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/us
  data.dropna(inplace=True)  # Remove any rows with missing values in the selected co
<ipython-input-10-5945d3cfff85>:12: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.

  sns.barplot(data=data, x='States/UTs', y='Murder', palette='viridis')  # Use Seabor
```



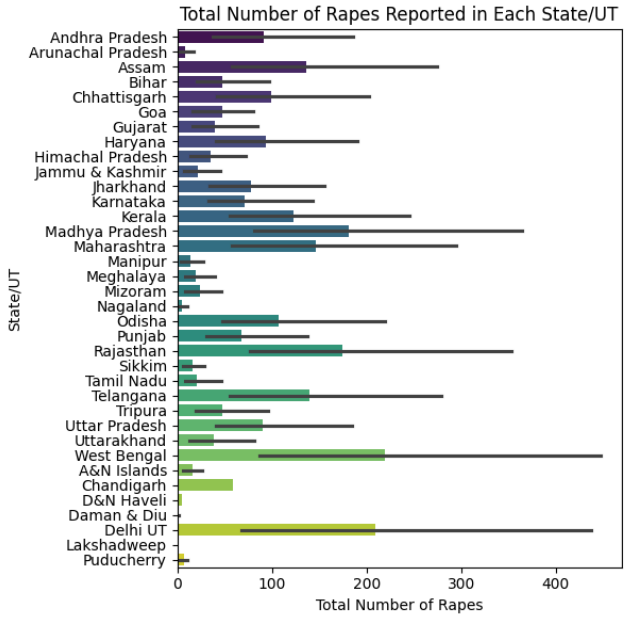
The plot displays the total number of murders reported in each state/union territory (State/UT) in India. A bar plot was chosen for this visualization because it effectively compares the total murder counts across different categories (states/UTs) by representing each category as a separate bar, making it easy to identify variations and trends.

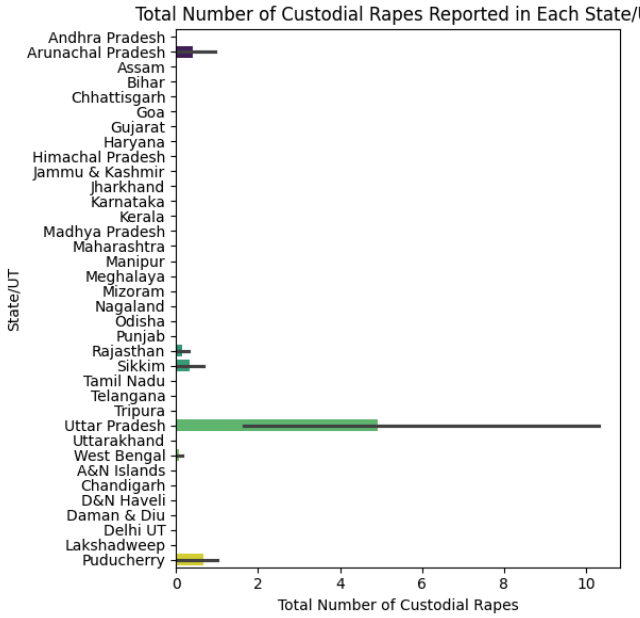
RESULT : Based on the figure displaying the crime rate for the year 2019, it is evident that Telegana had the highest crime rate among all states/union territories.

Total no. of Rapes reported in each state compared to Total number of custodial rapes reported in each state

```
import pandas as pd  # Import pandas for data manipulation
import seaborn as sns  # Import seaborn for data visualization
import matplotlib.pyplot as plt  # Import matplotlib for plotting

# Load the dataset
df = pd.read_csv("/content/crime_data.csv")  # Read the crime dataset into a DataFrame

# Select relevant columns (State/UT, Rape, and Custodial Rape)
data = df[['States/UTs', 'Rape', 'Custodial Rape']]  # Extract columns related to Rape an

# Remove rows with missing values
data.dropna(inplace=True)  # Drop rows with missing values in any of the selected columns

# Plotting
plt.figure(figsize=(12, 6))  # Set the figure size

# Plot the total number of rapes for each state/UT
plt.subplot(1, 2, 1)  # Create subplot 1 (left side)
sns.barplot(data=data, x='Rape', y='States/UTs', palette='viridis')  # Create bar plot fo
plt.title('Total Number of Rapes Reported in Each State/UT')  # Add title
plt.xlabel('Total Number of Rapes')  # Add label for x-axis
plt.ylabel('State/UT')  # Add label for y-axis

# Plot the total number of custodial rapes for each state/UT
plt.subplot(1, 2, 2)  # Create subplot 2 (right side)
sns.barplot(data=data, x='Custodial Rape', y='States/UTs', palette='viridis')  # Create b
plt.title('Total Number of Custodial Rapes Reported in Each State/UT')  # Add title
plt.xlabel('Total Number of Custodial Rapes')  # Add label for x-axis
plt.ylabel('State/UT')  # Add label for y-axis

plt.tight_layout()  # Adjust layout to prevent overlapping labels
plt.show()  # Display the plot
```

```
<ipython-input-11-d514c4999f74>:12: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/us
  data.dropna(inplace=True)  # Drop rows with missing values in any of the selected c
<ipython-input-11-d514c4999f74>:19: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.

  sns.barplot(data=data, x='Rape', y='States/UTs', palette='viridis')  # Create bar p
<ipython-input-11-d514c4999f74>:26: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.

  sns.barplot(data=data, x='Custodial Rape', y='States/UTs', palette='viridis')  # Cr
```





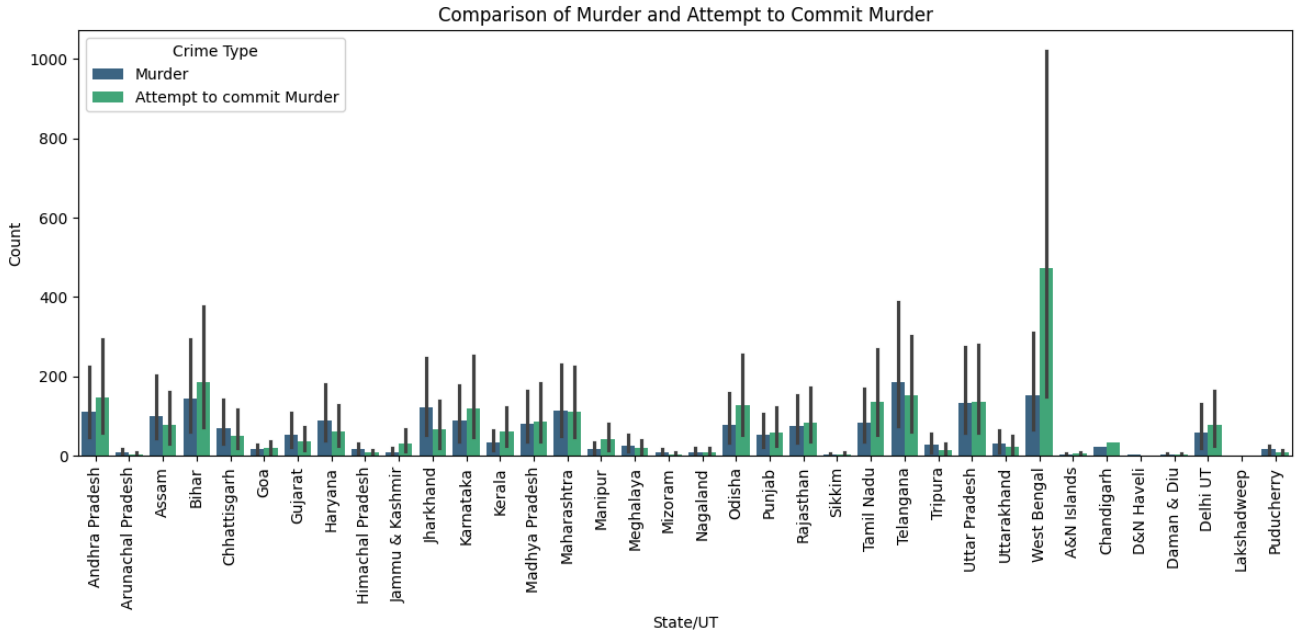
The plot consists of two side-by-side bar charts. The left chart shows the total number of rapes reported in each state/union territory (State/UT), while the right chart displays the total number of custodial rapes reported in each State/UT. I used this plot because it allows for a direct comparison of the total number of reported rapes and custodial rapes across different states/UTs. The use of bars makes it easy to visually compare the magnitudes of these two types of crimes for each State/UT.

RESULT :Based on the provided graph, it is evident that West Bengal has the highest number of rapes and Uttar Pradesh has the highest number of custodial rapes among all states/union territories during the year 2019.

Comparison of Murder and Attempt to commit Murder

```
import pandas as pd  # Import pandas for data manipulation
import seaborn as sns  # Import seaborn for data visualization
import matplotlib.pyplot as plt  # Import matplotlib for plotting

# Load the dataset
df = pd.read_csv("/content/crime_data.csv")  # Read the crime dataset into a DataFrame

# Select relevant columns (State/UT, Murder, and Attempt to commit Murder)
data = df[['States/UTs', 'Murder', 'Attempt to commit Murder']]  # Extract columns relate

# Remove rows with missing values
data.dropna(inplace=True)  # Drop rows with missing values in any of the selected columns

# Melt the DataFrame to reshape it for visualization
data_melted = pd.melt(data, id_vars=['States/UTs'], var_name='Crime Type', value_name='Co

# Plotting
plt.figure(figsize=(12, 6))  # Set the figure size

# Create a grouped bar plot
sns.barplot(data=data_melted, x='States/UTs', y='Count', hue='Crime Type', palette='virid
plt.title('Comparison of Murder and Attempt to Commit Murder')  # Add title to the plot
plt.xlabel('State/UT')  # Add label for x-axis
plt.ylabel('Count')  # Add label for y-axis
plt.xticks(rotation=90)  # Rotate x-axis labels for better readability
plt.legend(title='Crime Type')  # Add legend with title
plt.tight_layout()  # Adjust layout to prevent overlapping labels
plt.show()  # Display the plot
```

```
<ipython-input-12-308c03c8175d>:12: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/us
  data.dropna(inplace=True)  # Drop rows with missing values in any of the selected c
```



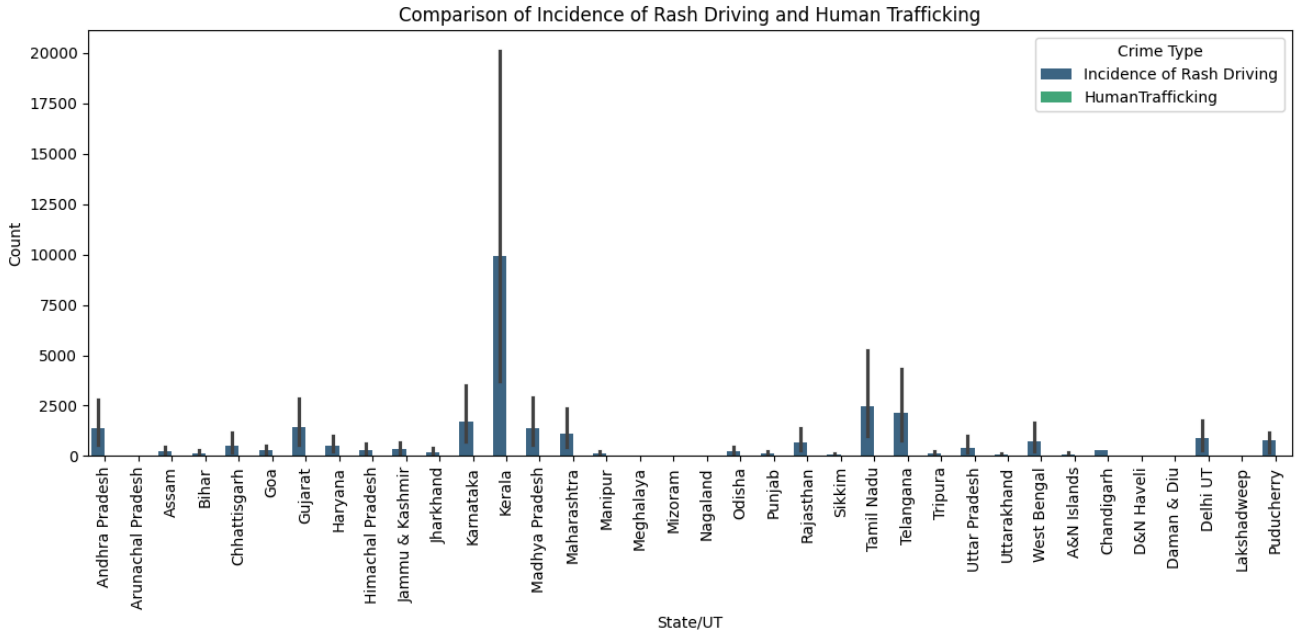
The plot is a grouped bar plot comparing the counts of "Murder" and "Attempt to commit Murder" across different states/union territories (States/UTs) in India. This plot was chosen because it allows for a direct visual comparison between the two types of crimes within each state/UT. By grouping the bars by crime type, we can easily observe the relative magnitudes of "Murder" and "Attempt to commit Murder" within each region.

RESULT : Based on the graph presented, it is evident that Telangana recorded the highest number of Murder and West Bengal recorded the highest Attempt to commit Murder cases among all states/union territories.

COMPARISON OF INCIDENCE OF RASH DRIVING AND HUMAN TRAFFICKING

```
import pandas as pd  # Import pandas for data manipulation
import seaborn as sns  # Import seaborn for data visualization
import matplotlib.pyplot as plt  # Import matplotlib for plotting

# Load the dataset
df = pd.read_csv("/content/crime_data.csv")

# Select relevant columns (State/UT, Incidence of Rash Driving, and Human Trafficking)
data = df[['States/UTs', 'Incidence of Rash Driving', 'HumanTrafficking']]

# Remove rows with missing values
data.dropna(inplace=True)

# Reshape the DataFrame using pd.melt() to treat 'Incidence of Rash Driving' and 'Human T
data_melted = pd.melt(data, id_vars=['States/UTs'], var_name='Crime Type', value_name='Co

# Plotting
plt.figure(figsize=(12, 6))

# Create a side-by-side bar plot comparing Incidence of Rash Driving and Human Traffickin
sns.barplot(data=data_melted, x='States/UTs', y='Count', hue='Crime Type', palette='virid
plt.title('Comparison of Incidence of Rash Driving and Human Trafficking')  # Add title t
plt.xlabel('State/UT')  # Add label for x-axis
plt.ylabel('Count')  # Add label for y-axis
plt.xticks(rotation=90)  # Rotate x-axis labels for better readability
plt.legend(title='Crime Type')  # Add legend with title
plt.tight_layout()  # Adjust layout to prevent overlapping labels
plt.show()  # Display the plot
```

```
<ipython-input-17-11c04c4d58fb>:12: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/us
  data.dropna(inplace=True)
```



This code compares the counts of "Incidence of Rash Driving" and "Human Trafficking" across different states/union territories (States/UTs) in India using a side-by-side bar plot.

Result : Based on the graph presented, it is evident that Kerala recorded the highest number of Rash Driving and there was no incident of Human Trafficking.

Total Cognizable IPC crimes in each State

```
import matplotlib.pyplot as plt  # Import matplotlib for plotting

# Group the data by 'States/UTs' and calculate the total cognizable IPC crimes for each s
total_ipc_crimes_by_state = df.groupby('States/UTs')['Total Cognizable IPC crimes'].sum()

# Plotting
plt.figure(figsize=(10, 10))  # Set the figure size
plt.pie(total_ipc_crimes_by_state, labels=total_ipc_crimes_by_state.index, autopct='%1.1f
plt.title('Proportion of Total Cognizable IPC Crimes by State')  # Add title to the plot
plt.axis('equal')  # Set the aspect ratio to be equal to make it a perfect circle
plt.show()  # Display the plot
```

The pie chart illustrates the proportion of total cognizable IPC crimes attributed to each state/union territory (States/UTs) in India. It visually represent the distribution of total cognizable IPC crimes across different regions. Each slice of the pie corresponds to a specific state/UT, and its size depicts the relative contribution of that region to the overall total of cognizable IPC crimes.

Result : Based on the pie chart above , we can see that Madhya Pradesh has the larger number of Cognizable IPC crimes i.e 12.9 % comapared to other states.

Regression Plotting:

Rape ~ Custodial Rape


```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.linear_model import LinearRegression

# Extract the relevant columns
data = df[['Custodial Rape', 'Rape']]

# Drop rows with missing values
data.dropna(inplace=True)

# Define the independent variable (X) and the dependent variable (y)
X = data[['Rape']]
y = data['Custodial Rape']

# Initialize and fit the linear regression model
model = LinearRegression()
model.fit(X, y)
```