

```
In [1]: import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
import warnings
warnings.filterwarnings('ignore')
```

```
In [3]: df = pd.read_csv('Capstone project phase 2.csv')
df
```

Out[3]:

	State	Year	Population of Each state	Litracy rate	Area in Sq Km	Type of Crime	Total Crimes	MURDER	ATTE MUF
0	Andhra Pradesh	2001	75728400	66.40	1,62,975	MURDER	130089	1555	
1	Arunachal Pradesh	2001	1098328	66.95	83,743	ATTEMPT TO MURDER	2342	53	
2	Assam	2001	26638600	73.18	78,438	CULPABLE HOMICIDE NOT AMOUNTING TO MURDER	36877	481	
3	Bihar	2001	82879910	69.82	94,163	RAPE	88432	3419	
4	Chhattisgarh	2001	20834530	71.04	1,35,192	CUSTODIAL RAPE	38460	529	
...	
319	Tamil Nadu	2012	635963102	81.33	1,30,058	DOWRY DEATHS	200474	2954	
320	Tripura	2012	32659810	88.75	1,12,077	ASSAULT ON WOMEN WITH INTENT TO OUTRAGE HER MO...	6264	79	
321	Uttar Pradesh	2012	179673604	69.78	2,40,928	INSULT TO MODESTY OF WOMEN	198093	4811	
322	Uttarakhand	2012	89449107	79.64	53,483	CRUELTY BY HUSBAND OR HIS RELATIVES	8882	161	
323	West Bengal	2012	86571309	78.08	88,752	IMPORTATION OF GIRLS FROM FOREIGN COUNTRIES	161427	2854	

324 rows × 37 columns

```
In [4]: # Display the first few rows of the dataset  
print(df.head())
```

	State	Year	Population of Each state	Litracy rate \
0	Andhra Pradesh	2001	75728400	66.40
1	Arunachal Pradesh	2001	1098328	66.95
2	Assam	2001	26638600	73.18
3	Bihar	2001	82879910	69.82
4	Chhattisgarh	2001	20834530	71.04

	Area in Sq Km	Type of Crime	Total Crimes
0	1,62,975	MURDER	130089
1	83,743	ATTEMPT TO MURDER	2342
2	78,438	CULPABLE HOMICIDE NOT AMOUNTING TO MURDER	36877
3	94,163	RAPE	88432
4	1,35,192	CUSTODIAL RAPE	38460

	MURDER	ATTEMPT TO MURDER	CULPABLE HOMICIDE NOT AMOUNTING TO MURDER
...	\		
0	1555	1555	136
...			
1	53	53	3
...			
2	481	481	40
...			
3	3419	3419	250
...			
4	529	529	45
...			

	ARSON	HURT/GREVIOUS HURT	DOWRY DEATHS \
0	872	34947	420
1	13	466	0
2	441	5805	59
3	502	7544	859
4	215	5477	70

	ASSAULT ON WOMEN WITH INTENT TO OUTRAGE HER MODESTY \
0	3544
1	78
2	850
3	562
4	1763

	INSULT TO MODESTY OF WOMEN	CRUELTY BY HUSBAND OR HIS RELATIVES \
0	2271	5791
1	3	11
2	4	1248
3	21	1558
4	161	840

	IMPORTATION OF GIRLS FROM FOREIGN COUNTRIES	CAUSING DEATH BY NEGLIGENCE \
0	7	740
0		
1	0	
0		
2	0	201
0		
3	83	240
6		
4	0	68
9		

	OTHER IPC CRIMES	TOTAL IPC CRIMES
0	34344	130089
1	618	2342
2	9315	36877
3	36667	88432
4	15790	38460

[5 rows x 37 columns]

```
In [5]: # Check for missing values
print(df.isnull().sum())
```

```
State      0
Year      0
Population of Each state  0
Litracys rate  0
Area in Sq Km  0
Type of Crime  0
Total Crimes  0
MURDER      0
ATTEMPT TO MURDER  0
CULPABLE HOMICIDE NOT AMOUNTING TO MURDER  0
RAPE        0
CUSTODIAL RAPE  0
OTHER RAPE   0
KIDNAPPING & ABDUCTION  0
KIDNAPPING AND ABDUCTION OF WOMEN AND GIRLS  0
KIDNAPPING AND ABDUCTION OF OTHERS  0
DACOITY      0
PREPARATION AND ASSEMBLY FOR DACOITY  0
ROBBERY      0
BURGLARY     0
THEFT        0
AUTO THEFT   0
OTHER THEFT  0
RIOTS        0
CRIMINAL BREACH OF TRUST  0
CHEATING     0
COUNTERFIETING  0
ARSON        0
HURT/GREVIOUS HURT  0
DOWRY DEATHS  0
ASSAULT ON WOMEN WITH INTENT TO OUTRAGE HER MODESTY  0
INSULT TO MODESTY OF WOMEN  0
CRUELTY BY HUSBAND OR HIS RELATIVES  0
IMPORTATION OF GIRLS FROM FOREIGN COUNTRIES  0
CAUSING DEATH BY NEGLIGENCE  0
OTHER IPC CRIMES  0
TOTAL IPC CRIMES  0
dtype: int64
```

```
In [6]: # Check data types  
print(df.dtypes)
```

State	object
Year	int64
Population of Each state	int64
Litracy rate	float64
Area in Sq Km	object
Type of Crime	object
Total Crimes	int64
MURDER	int64
ATTEMPT TO MURDER	int64
CULPABLE HOMICIDE NOT AMOUNTING TO MURDER	int64
RAPE	int64
CUSTODIAL RAPE	int64
OTHER RAPE	int64
KIDNAPPING & ABDUCTION	int64
KIDNAPPING AND ABDUCTION OF WOMEN AND GIRLS	int64
KIDNAPPING AND ABDUCTION OF OTHERS	int64
DACOITY	int64
PREPARATION AND ASSEMBLY FOR DACOITY	int64
ROBBERY	int64
BURGLARY	int64
THEFT	int64
AUTO THEFT	int64
OTHER THEFT	int64
RIOTS	int64
CRIMINAL BREACH OF TRUST	int64
CHEATING	int64
COUNTERFIETING	int64
ARSON	int64
HURT/GREVIOUS HURT	int64
DOWRY DEATHS	int64
ASSAULT ON WOMEN WITH INTENT TO OUTRAGE HER MODESTY	int64
INSULT TO MODESTY OF WOMEN	int64
CRUELTY BY HUSBAND OR HIS RELATIVES	int64
IMPORTATION OF GIRLS FROM FOREIGN COUNTRIES	int64
CAUSING DEATH BY NEGLIGENCE	int64
OTHER IPC CRIMES	int64
TOTAL IPC CRIMES	int64
dtype:	object

```
In [8]: # Handle missing values (if any)
df = df.dropna()
df
```

Out[8]:

MURDER	ATTEMPT TO MURDER	CULPABLE HOMICIDE NOT AMOUNTING TO MURDER	...	ARSON	HURT/GREVIOUS HURT	DOWRY DEATHS	ASSAULT ON WOMEN WITH INTENT TO OUTRAGE HER MODESTY	INS MODE WOI
1555	1555	136	...	872	34947	420	3544	:
53	53	3	...	13	466	0	78	
481	481	40	...	441	5805	59	850	
3419	3419	250	...	502	7544	859	562	
529	529	45	...	215	5477	70	1763	
...	
2954	2954	44	...	726	22100	110	1494	
79	79	2	...	59	1504	37	314	
4811	4811	1410	...	327	12290	2244	3247	
161	161	44	...	17	1022	71	139	
2854	2854	522	...	396	15923	593	3345	

In [9]:

```
State                object
Year                int64
Population of Each state  int64
Literacy rate        float64
Area in Sq Km        object
Type of Crime        object
Total Crimes         int64
MURDER               int64
ATTEMPT TO MURDER     int64
CULPABLE HOMICIDE NOT AMOUNTING TO MURDER int64
RAPE                 int64
CUSTODIAL RAPE       int64
OTHER RAPE           int64
KIDNAPPING & ABDUCTION int64
KIDNAPPING AND ABDUCTION OF WOMEN AND GIRLS int64
KIDNAPPING AND ABDUCTION OF OTHERS         int64
DACOITY              int64
PREPARATION AND ASSEMBLY FOR DACOITY        int64
ROBBERY              int64
BURGLARY             int64
THEFT                int64
AUTO THEFT           int64
OTHER THEFT          int64
RIOTS               int64
CRIMINAL BREACH OF TRUST int64
CHEATING             int64
COUNTERFIETING       int64
ARSON                int64
HURT/GREVIOUS HURT   int64
DOWRY DEATHS         int64
ASSAULT ON WOMEN WITH INTENT TO OUTRAGE HER MODESTY int64
INSULT TO MODESTY OF WOMEN int64
CRUELTY BY HUSBAND OR HIS RELATIVES         int64
IMPORTATION OF GIRLS FROM FOREIGN COUNTRIES int64
CAUSING DEATH BY NEGLIGENCE int64
OTHER IPC CRIMES     int64
TOTAL IPC CRIMES     int64
dtype: object
```

```
In [33]: df['State'] = pd.to_numeric(df['State'], errors='coerce')
df = df.dropna(axis=1)
df
```

Out[33]:

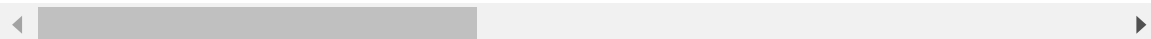
	State	Year	Population of Each state	Litracy rate	Area in Sq Km	Type of Crime	Total Crimes	MURDER	ATTEMPT TO MURDER
0	-1.0	2001	75728400	66.40	1,62,975	MURDER	130089	1555	1555
1	-1.0	2001	1098328	66.95	83,743	ATTEMPT TO MURDER	2342	53	53
2	-1.0	2001	26638600	73.18	78,438	CULPABLE HOMICIDE NOT AMOUNTING TO MURDER	36877	481	481
3	-1.0	2001	82879910	69.82	94,163	RAPE	88432	3419	3419
4	-1.0	2001	20834530	71.04	1,35,192	CUSTODIAL RAPE	38460	529	529
...
319	-1.0	2012	635963102	81.33	1,30,058	DOWRY DEATHS	200474	2954	2954
320	-1.0	2012	32659810	88.75	1,12,077	ASSAULT ON WOMEN WITH INTENT TO OUTRAGE HER MO...	6264	79	79
321	-1.0	2012	179673604	69.78	2,40,928	INSULT TO MODESTY OF WOMEN	198093	4811	4811
322	-1.0	2012	89449107	79.64	53,483	CRUELTY BY HUSBAND OR HIS RELATIVES	8882	161	161
323	-1.0	2012	86571309	78.08	88,752	IMPORTATION OF GIRLS FROM FOREIGN COUNTRIES	161427	2854	2854
324 rows × 37 columns									


```
In [34]: df['State'].fillna(df['State'].mean(), inplace=True)
df['State'].fillna(df['State'].median(), inplace=True)
df = df.ffill()
df = df.bfill()
df['State'] = df['State'].interpolate(method='linear')
df.fillna(-1, inplace=True)
df
```

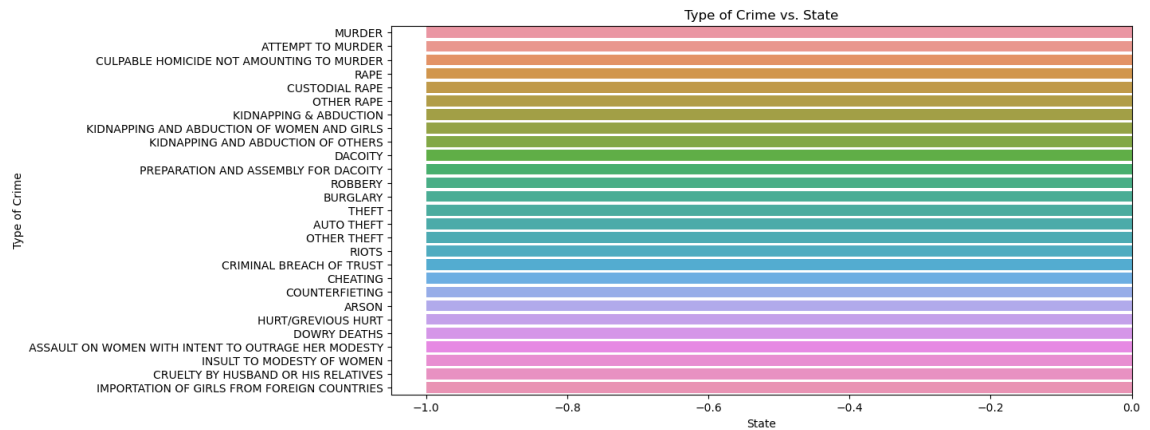
Out[34]:

	State	Year	Population of Each state	Litracy rate	Area in Sq Km	Type of Crime	Total Crimes	MURDER	ATTEMPT TO MURDER
0	-1.0	2001	75728400	66.40	1,62,975	MURDER	130089	1555	1555
1	-1.0	2001	1098328	66.95	83,743	ATTEMPT TO MURDER	2342	53	53
2	-1.0	2001	26638600	73.18	78,438	CULPABLE HOMICIDE NOT AMOUNTING TO MURDER	36877	481	481
3	-1.0	2001	82879910	69.82	94,163	RAPE	88432	3419	3419
4	-1.0	2001	20834530	71.04	1,35,192	CUSTODIAL RAPE	38460	529	529
...
319	-1.0	2012	635963102	81.33	1,30,058	DOWRY DEATHS	200474	2954	2954
320	-1.0	2012	32659810	88.75	1,12,077	ASSAULT ON WOMEN WITH INTENT TO OUTRAGE HER MO...	6264	79	79
321	-1.0	2012	179673604	69.78	2,40,928	INSULT TO MODESTY OF WOMEN	198093	4811	4811
322	-1.0	2012	89449107	79.64	53,483	CRUELTY BY HUSBAND OR HIS RELATIVES	8882	161	161
323	-1.0	2012	86571309	78.08	88,752	IMPORTATION OF GIRLS FROM FOREIGN COUNTRIES	161427	2854	2854

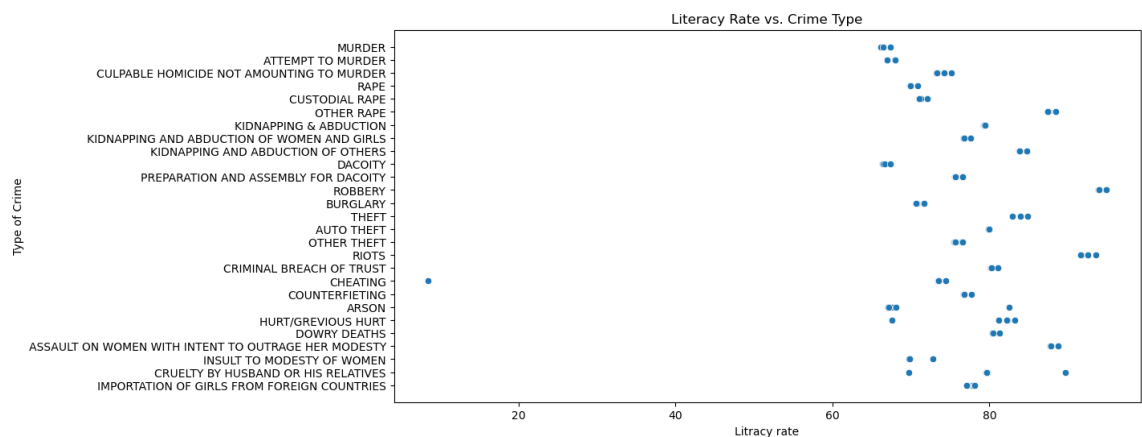
324 rows × 37 columns



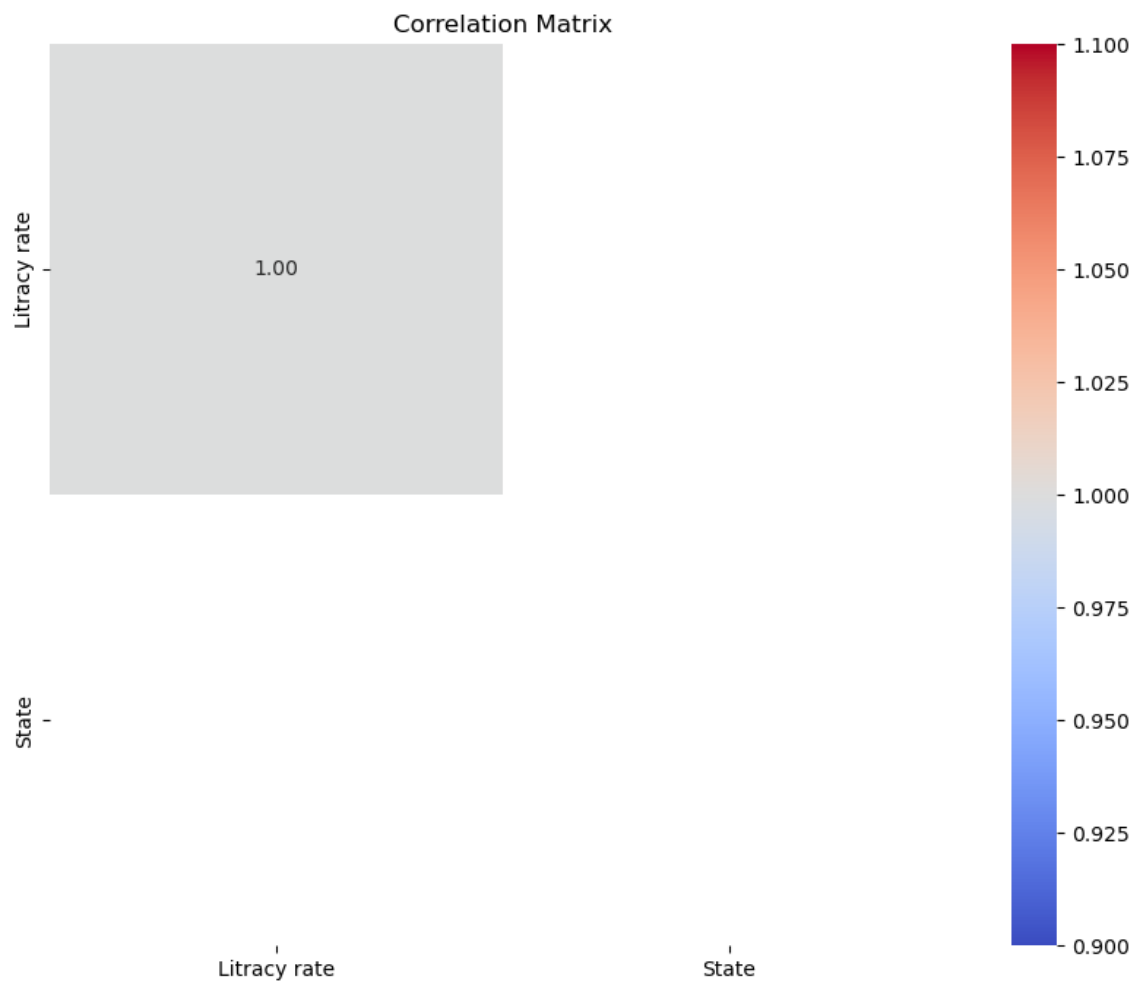
```
In [35]: # Create a bar plot of crime types vs. state
plt.figure(figsize=(12, 6))
sns.barplot(x='State', y='Type of Crime', data=df)
plt.title('Type of Crime vs. State')
plt.show()
```



```
In [36]: # Create a scatter plot of literacy rate vs. crime type
plt.figure(figsize=(12, 6))
sns.scatterplot(x='Literacy rate', y='Type of Crime', data=df)
plt.title('Literacy Rate vs. Crime Type')
plt.show()
```



```
In [37]: # Create a heatmap to visualize the correlation between variables
correlation_matrix = df[['Litracy rate', 'State', 'Type of Crime']].corr()
plt.figure(figsize=(10, 8))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', fmt='.2f')
plt.title('Correlation Matrix')
plt.show()
```



In []: