Crime prediction

```
import pandas as pd
import numpy as np
df=pd.read_csv("/crime_dataset_india.csv")
```

```
df.shape
```

```
(40160, 14)
```

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 40160 entries, 0 to 40159
Data columns (total 14 columns):
 #   Column              Non-Null Count  Dtype 
---  ------              --------------  ----- 
 0   Report Number       40160 non-null  int64 
 1   Date Reported       40160 non-null  object
 2   Date of Occurrence  40160 non-null  object
 3   Time of Occurrence  40160 non-null  object
 4   City                40160 non-null  object
 5   Crime Code          40160 non-null  int64 
 6   Crime Description   40160 non-null  object
 7   Victim Age          40160 non-null  int64 
 8   Victim Gender       40160 non-null  object
 9   Weapon Used         34370 non-null  object
 10  Crime Domain        40160 non-null  object
 11  Police Deployed     40160 non-null  int64 
 12  Case Closed         40160 non-null  object
 13  Date Case Closed    20062 non-null  object
dtypes: int64(4), object(10)
memory usage: 4.3+ MB
```

Exploratory Data Analysis

```
import matplotlib.pyplot as plt
df.describe()
```

	Report Number	Crime Code	Victim Age	Police Deployed
count	40160.000000	40160.000000	40160.00000	40160.000000
mean	20080.500000	349.360259	44.49126	10.006250
std	11593.337742	144.169205	20.22555	5.467951
min	1.000000	100.000000	10.00000	1.000000
25%	10040.750000	225.000000	27.00000	5.000000
50%	20080.500000	349.000000	44.00000	10.000000
75%	30120.250000	474.000000	62.00000	15.000000
max	40160.000000	599.000000	79.00000	19.000000

```
print(df.isnull().sum)
```

```
<bound method DataFrame.sum of        Report Number  Date Reported  Date of Occurrence
0              False          False               False               False
1              False          False               False               False
2              False          False               False               False
3              False          False               False               False
4              False          False               False               False
...              ...            ...                 ...                 ...
40155          False          False               False               False
40156          False          False               False               False
40157          False          False               False               False
40158          False          False               False               False
40159          False          False               False               False

        City  Crime Code  Crime Description  Victim Age  Victim Gender  \
0      False       False              False       False          False
1      False       False              False       False          False
2      False       False              False       False          False
3      False       False              False       False          False
4      False       False              False       False          False
...      ...         ...                ...         ...            ...
40155  False       False              False       False          False
40156  False       False              False       False          False
40157  False       False              False       False          False
40158  False       False              False       False          False
40159  False       False              False       False          False

       Weapon Used  Crime Domain  Police Deployed  Case Closed  \
0            False         False            False        False
1            False         False            False        False
2            False         False            False        False
3            False         False            False        False
4            False         False            False        False
...            ...           ...              ...          ...
40155        False         False            False        False
40156         True         False            False        False
40157        False         False            False        False
40158        False         False            False        False
40159        False         False            False        False

       Date Case Closed
0                  True
```

```
1                    True
2                    True
3                   False
4                   False
...                   ...
40155                True
40156               False
40157                True
40158                True
40159               False

[40160 rows x 14 columns]>
```

```
print(df.isnull())
```

```
       Report Number  Date Reported  Date of Occurrence  Time of Occurrence  \
0              False          False               False               False   
1              False          False               False               False   
2              False          False               False               False   
3              False          False               False               False   
4              False          False               False               False   
...              ...            ...                 ...                 ...   
40155          False          False               False               False   
40156          False          False               False               False   
40157          False          False               False               False   
40158          False          False               False               False   
40159          False          False               False               False   

        City  Crime Code  Crime Description  Victim Age  Victim Gender  \
0      False       False              False       False          False   
1      False       False              False       False          False   
2      False       False              False       False          False   
3      False       False              False       False          False   
4      False       False              False       False          False   
...      ...         ...                ...         ...            ...   
40155  False       False              False       False          False   
40156  False       False              False       False          False   
40157  False       False              False       False          False   
40158  False       False              False       False          False   
40159  False       False              False       False          False   

       Weapon Used  Crime Domain  Police Deployed  Case Closed  \
0            False         False            False        False   
1            False         False            False        False   
2            False         False            False        False   
3            False         False            False        False   
4            False         False            False        False   
...            ...           ...              ...          ...   
40155        False         False            False        False   
40156         True         False            False        False   
40157        False         False            False        False   
40158        False         False            False        False   
40159        False         False            False        False   

       Date Case Closed  
0                  True  
1                  True  
2                  True  
3                 False  
```

```
4                False
...                ...
40155             True
40156            False
40157             True
40158             True
40159            False

[40160 rows x 14 columns]
```

```
df=df.dropna()
```

```
print(df.isnull())
```

```
       Report Number  Date Reported  Date of Occurrence  Time of Occurrence  \
3              False          False               False               False
4              False          False               False               False
5              False          False               False               False
6              False          False               False               False
10             False          False               False               False
...              ...            ...                 ...                 ...
40144          False          False               False               False
40146          False          False               False               False
40149          False          False               False               False
40153          False          False               False               False
40159          False          False               False               False

        City  Crime Code  Crime Description  Victim Age  Victim Gender  \
3      False       False              False       False          False
4      False       False              False       False          False
5      False       False              False       False          False
6      False       False              False       False          False
10     False       False              False       False          False
...      ...         ...                ...         ...            ...
40144  False       False              False       False          False
40146  False       False              False       False          False
40149  False       False              False       False          False
40153  False       False              False       False          False
40159  False       False              False       False          False

       Weapon Used  Crime Domain  Police Deployed  Case Closed  \
3            False         False            False        False
4            False         False            False        False
5            False         False            False        False
6            False         False            False        False
10           False         False            False        False
...            ...           ...              ...          ...
40144        False         False            False        False
40146        False         False            False        False
40149        False         False            False        False
40153        False         False            False        False
40159        False         False            False        False

       Date Case Closed
3                 False
4                 False
5                 False
6                 False
10                False
...                 ...
```

```
40144             False
40146             False
40149             False
40153             False
40159             False

[17130 rows x 14 columns]
```

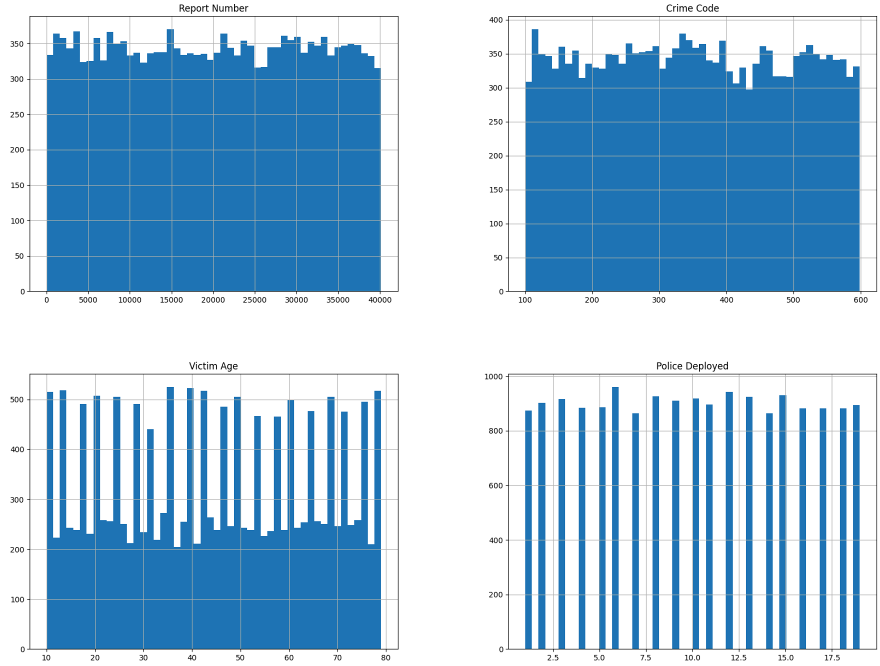
Double-click (or enter) to edit

Histogram

Double-click (or enter) to edit

```
df.hist(bins=50, figsize=(20,15))
```

```
array([[<Axes: title={'center': 'Report Number'}>,
        <Axes: title={'center': 'Crime Code'}>],
       [<Axes: title={'center': 'Victim Age'}>,
        <Axes: title={'center': 'Police Deployed'}>]], dtype=object)
```



```
import seaborn as sns
corr=df.corr()
sns.heatmap(corr)
```

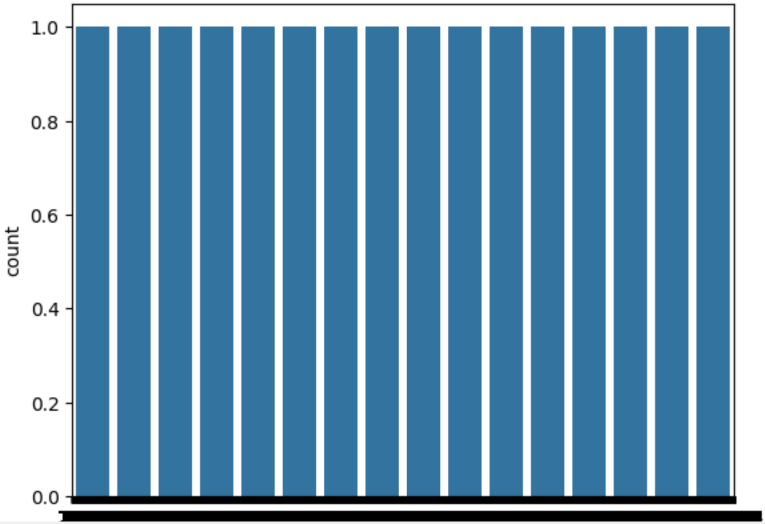
```
---------------------------------------------------------------------------
ValueError                                Traceback (most recent call last)
<ipython-input-75-8a014dbb3e53> in <cell line: 2>()
      1 import seaborn as sns
----> 2 corr=df.corr()
      3 sns.heatmap(corr)

                              3 frames

/usr/local/lib/python3.10/dist-packages/pandas/core/internals/managers.py in
_interleave(self, dtype, na_value)
   1751                 else:
   1752                     arr = blk.get_values(dtype)
-> 1753                 result[rl.indexer] = arr
   1754                 itemmask[rl.indexer] = 1
   1755
```

Next steps: Explain error

Generate randomly select 5 items from a list Close

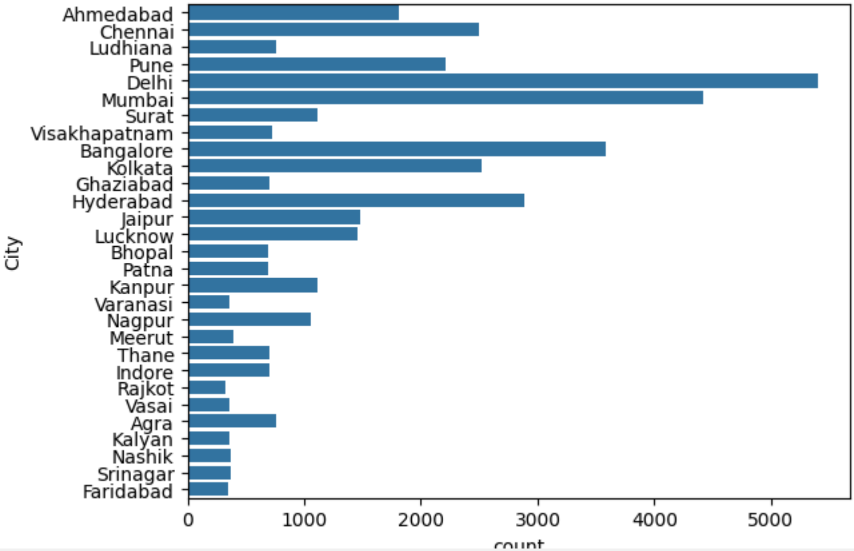
```
sns.countplot(df['Victim Age'])
```

```
<Axes: ylabel='count'>
```



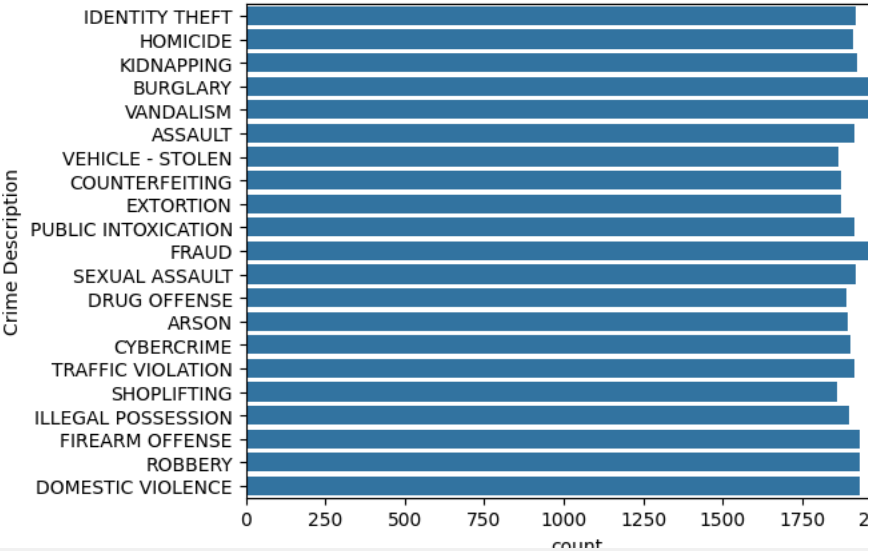
```
sns.countplot(df['City'])
```

```
<Axes: xlabel='count', ylabel='City'>
```



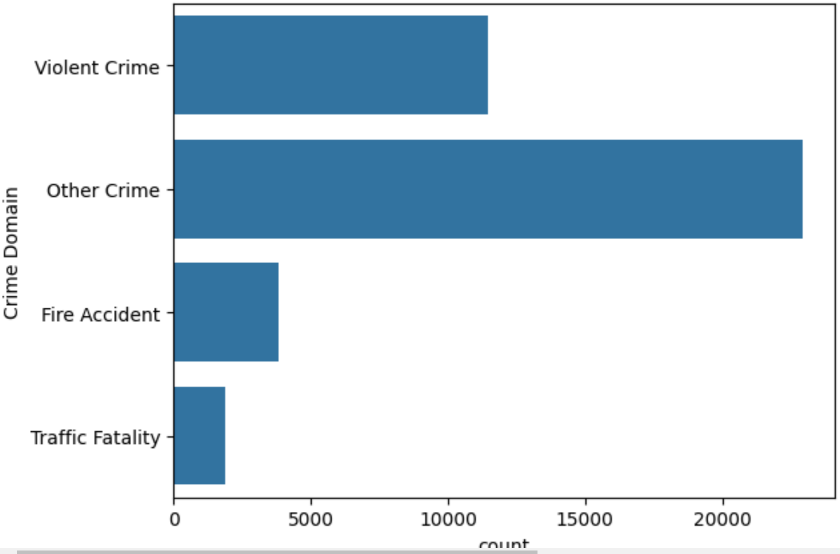
```
sns.countplot(df['Crime Description'])
```

```
<Axes: xlabel='count', ylabel='Crime Description'>
```



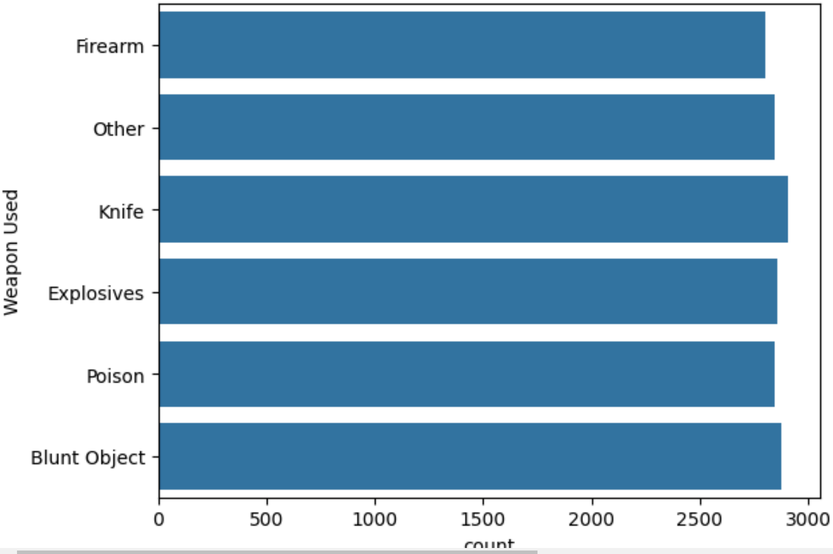
```
sns.countplot(df['Crime Domain'])
```

```
<Axes: xlabel='count', ylabel='Crime Domain'>
```



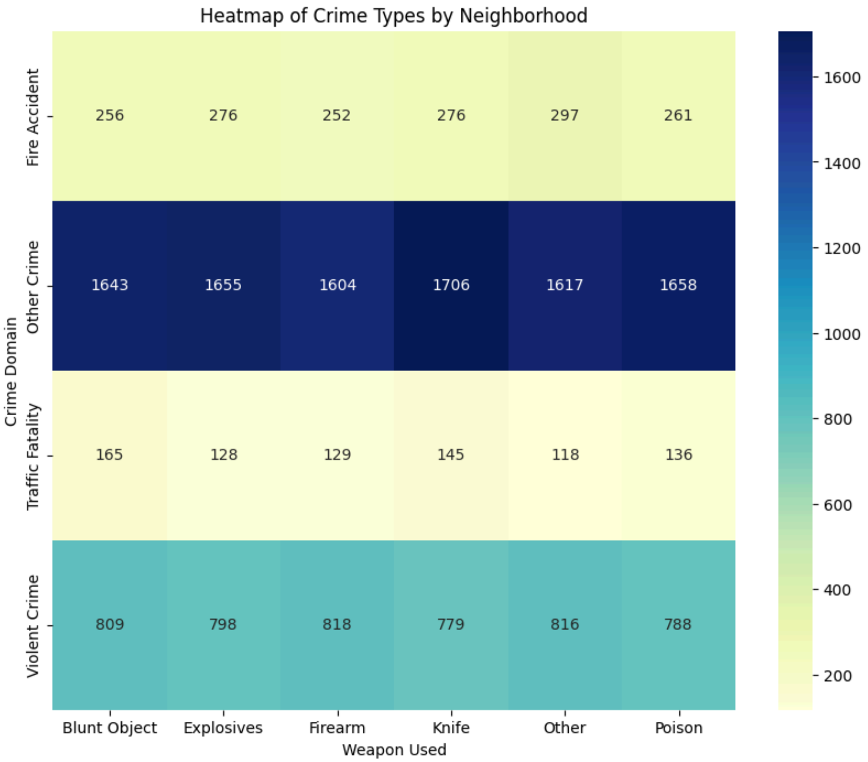
```
sns.countplot(df['Weapon Used'])
```

```
<Axes: xlabel='count', ylabel='Weapon Used'>
```



```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

```
# Create a cross-tabulation between two categorical variables
cross_tab = pd.crosstab(df['Crime Domain'], df['Weapon Used'])

# Plot the cross-tab as a heatmap
plt.figure(figsize=(10, 8))
sns.heatmap(cross_tab, cmap="YlGnBu", annot=True, fmt="d")
plt.title("Heatmap of Crime Types by Neighborhood")
plt.show()
```



Generate 10 random numbers using numpy Close

```
from sklearn.preprocessing import LabelEncoder

# Load your dataset
df = pd.read_csv('/crime_dataset_india.csv') # Replace with your file path

# Convert all columns with string values to numerical using Label Encoding
# Convert all columns with string values to numerical using Label Encoding
label_encoders = {}
for column in df.columns:
    if df[column].dtype == 'object': # Check if column is of string type
        le = LabelEncoder()
        df[column] = le.fit_transform(df[column].astype(str)) # Encode the string data
        label_encoders[column] = le # Save the encoder if you need to decode later

# Calculate the correlation matrix
correlation_matrix = df.corr()

# Set up the matplotlib figure
plt.figure(figsize=(10, 8))

# Create a heatmap
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', fmt='.2f', linewidths=0.5)

# Add labels and title
plt.title("Heatmap of Encoded Crime Prediction Dataset")
plt.xlabel("Features")
plt.ylabel("Features")

# Display the heatmap
plt.show()
```