1. perform EDA on the data that you choose for the project

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

# Load the dataset

data = pd.read\_csv('crime.csv', encoding='iso-8859-1',low\_memory=False) # Replace with your actual dataset path

# Display the first few rows of the dataset

print("Dataset Overview:")

print(data.head())

# Basic statistics of numerical columns

print("\nSummary Statistics:")

print(data.describe())

# Check for missing values

print("\nMissing Values:")

print(data.isnull().sum())

# Distribution of crime types (Offense Code Group)

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

sns.countplot(x='OFFENSE\_CODE\_GROUP', data=data, order=data['OFFENSE\_CODE\_GROUP'].value\_counts().index)

plt.xticks(rotation=90)

plt.title('Distribution of Crime Types')

plt.xlabel('Crime Type')

plt.ylabel('Count')

plt.show()

# Distribution of crimes over the years

plt.figure(figsize=(10, 6))

sns.countplot(x='YEAR', data=data)

plt.title('Distribution of Crimes Over the Years')

plt.xlabel('Year')

plt.ylabel('Count')

plt.show()

# Monthly distribution of crimes

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

sns.countplot(x='MONTH', data=data, hue='YEAR')

plt.title('Monthly Distribution of Crimes')

plt.xlabel('Month')

plt.ylabel('Count')

plt.show()

# Day of the week with most crimes

plt.figure(figsize=(10, 6))

sns.countplot(x='DAY\_OF\_WEEK', data=data, order=['Monday', 'Tuesday', 'Wednesday', 'Thursday', 'Friday', 'Saturday', 'Sunday'])

plt.title('Crimes by Day of the Week')

plt.xlabel('Day of the Week')

plt.ylabel('Count')

plt.show()

**Output:-**

Dataset Overview:

INCIDENT\_NUMBER OFFENSE\_CODE OFFENSE\_CODE\_GROUP \

0 I182080058 2403 Disorderly Conduct

1 I182080053 3201 Property Lost

2 I182080052 2647 Other

3 I182080051 413 Aggravated Assault

4 I182080050 3122 Aircraft

OFFENSE\_DESCRIPTION DISTRICT REPORTING\_AREA SHOOTING \

0 DISTURBING THE PEACE E18 495 NaN

1 PROPERTY - LOST D14 795 NaN

2 THREATS TO DO BODILY HARM B2 329 NaN

3 ASSAULT - AGGRAVATED - BATTERY A1 92 NaN

4 AIRCRAFT INCIDENTS A7 36 NaN

OCCURRED\_ON\_DATE YEAR MONTH DAY\_OF\_WEEK HOUR UCR\_PART \

0 2018-10-03 20:13:00 2018 10 Wednesday 20 Part Two

1 2018-08-30 20:00:00 2018 8 Thursday 20 Part Three

2 2018-10-03 19:20:00 2018 10 Wednesday 19 Part Two

3 2018-10-03 20:00:00 2018 10 Wednesday 20 Part One

4 2018-10-03 20:49:00 2018 10 Wednesday 20 Part Three

STREET Lat Long Location

0 ARLINGTON ST 42.262608 -71.121186 (42.26260773, -71.12118637)

1 ALLSTON ST 42.352111 -71.135311 (42.35211146, -71.13531147)

2 DEVON ST 42.308126 -71.076930 (42.30812619, -71.07692974)

3 CAMBRIDGE ST 42.359454 -71.059648 (42.35945371, -71.05964817)

4 PRESCOTT ST 42.375258 -71.024663 (42.37525782, -71.02466343)

Summary Statistics:

OFFENSE\_CODE YEAR MONTH HOUR \

count 327820.000000 327820.000000 327820.000000 327820.000000

mean 2317.961171 2016.598676 6.672213 13.114840

std 1184.990073 1.009775 3.253984 6.292714

min 111.000000 2015.000000 1.000000 0.000000

25% 1001.000000 2016.000000 4.000000 9.000000

50% 2907.000000 2017.000000 7.000000 14.000000

75% 3201.000000 2017.000000 9.000000 18.000000

max 3831.000000 2018.000000 12.000000 23.000000

Lat Long

count 307188.000000 307188.000000

mean 42.212995 -70.906030

std 2.173496 3.515832

min -1.000000 -71.178674

25% 42.297466 -71.097081

50% 42.325552 -71.077493

75% 42.348624 -71.062482

max 42.395042 -1.000000

Missing Values:

INCIDENT\_NUMBER 0

OFFENSE\_CODE 0

OFFENSE\_CODE\_GROUP 0

OFFENSE\_DESCRIPTION 0

DISTRICT 1774

REPORTING\_AREA 0

SHOOTING 326765

OCCURRED\_ON\_DATE 0

YEAR 0

MONTH 0

DAY\_OF\_WEEK 0

HOUR 0

UCR\_PART 93

STREET 10977

Lat 20632

Long 20632

Location 0

dtype: int64

