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()