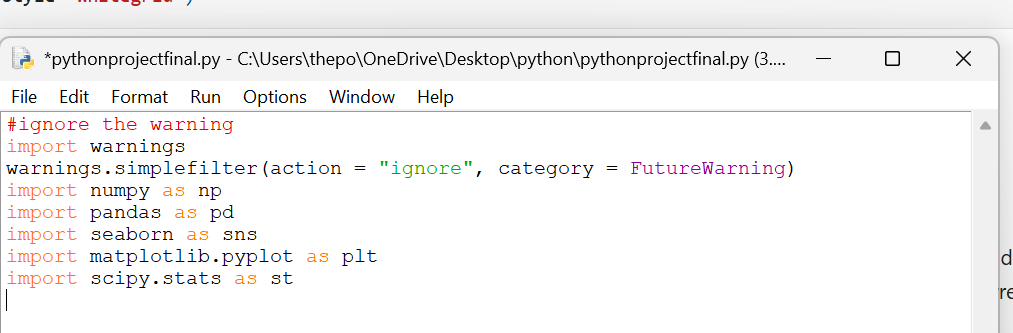
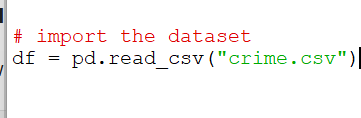
Crime Insights: Data-Driven Crime Analysis

(2020-2025)

* Name – Pratiksha Tripathi
* Reg – no – 12318047
* Roll-no- 30
* Section – K23vm
* In This project I have covered almost every point of python libraries including NumPy pandas mat plot and seaborn
* The Website from which I have taken this dataset is –
* <https://catalog.data.gov/dataset/crime-data-from-2020-to-present>
* This project is based on the crime dataset between the years 2020 to 2025
* 1. Importing the warnings and python libraries in idle python --



* 2. Importing the data set



# Using Pandas

* 3. Overview of the data set

. check the dimension of the data set for that we have use shape attribute



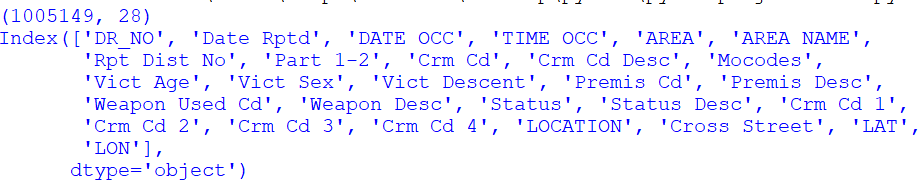
Ans – The output of the code is the



. check the columns of the dataset for that I used attribute



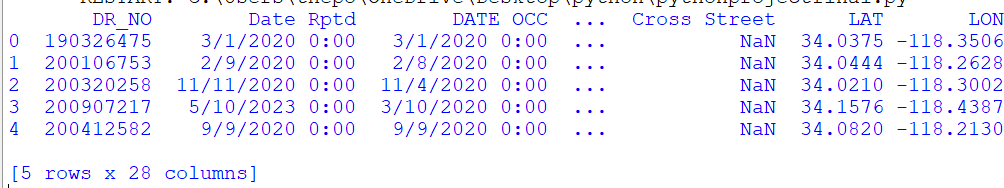
Ans – The output of the code is the



. check the top 5 row of the dataset

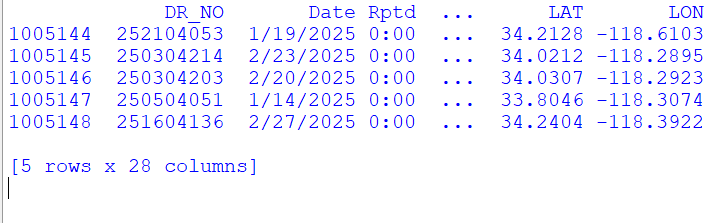


Ans – The output of the code is



. check the list 5 rows of the dataset

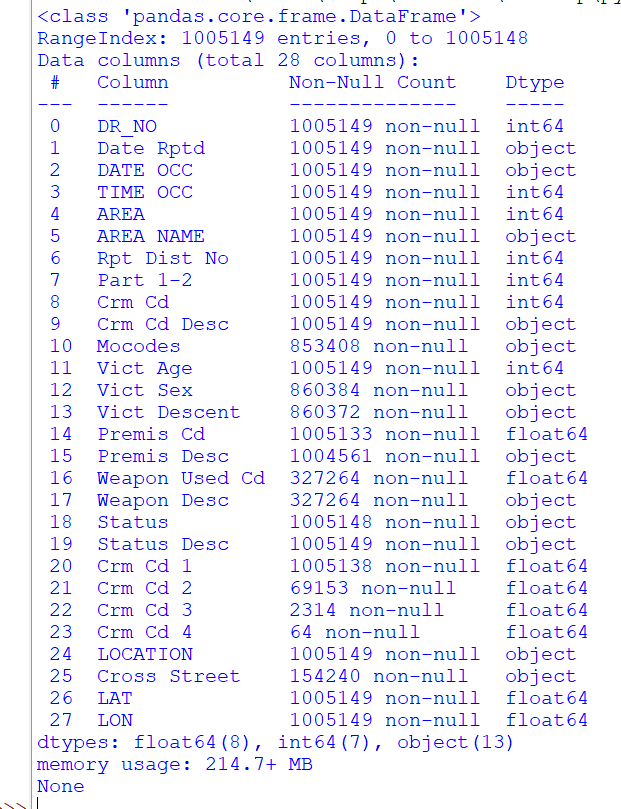


Ans – The output of the code is

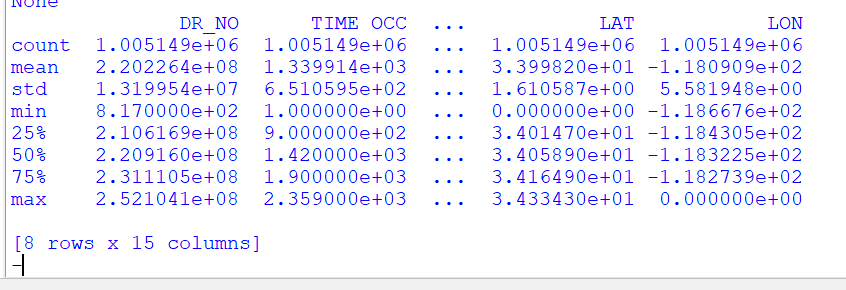
. checking all the information of the dataset and details then we use info function



Ans – The output of the code is



. checking for the describe method it will give you the summary of the invention

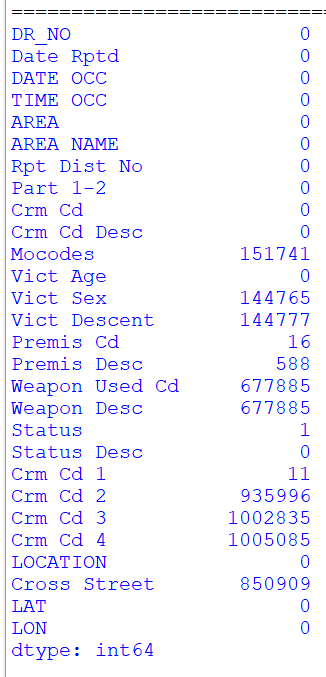
Ans -

* 4. **Check for anomalies in the dataset**

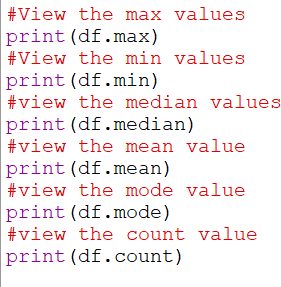
. check for missing numeric values Check for the missing number in the dataset and their sum



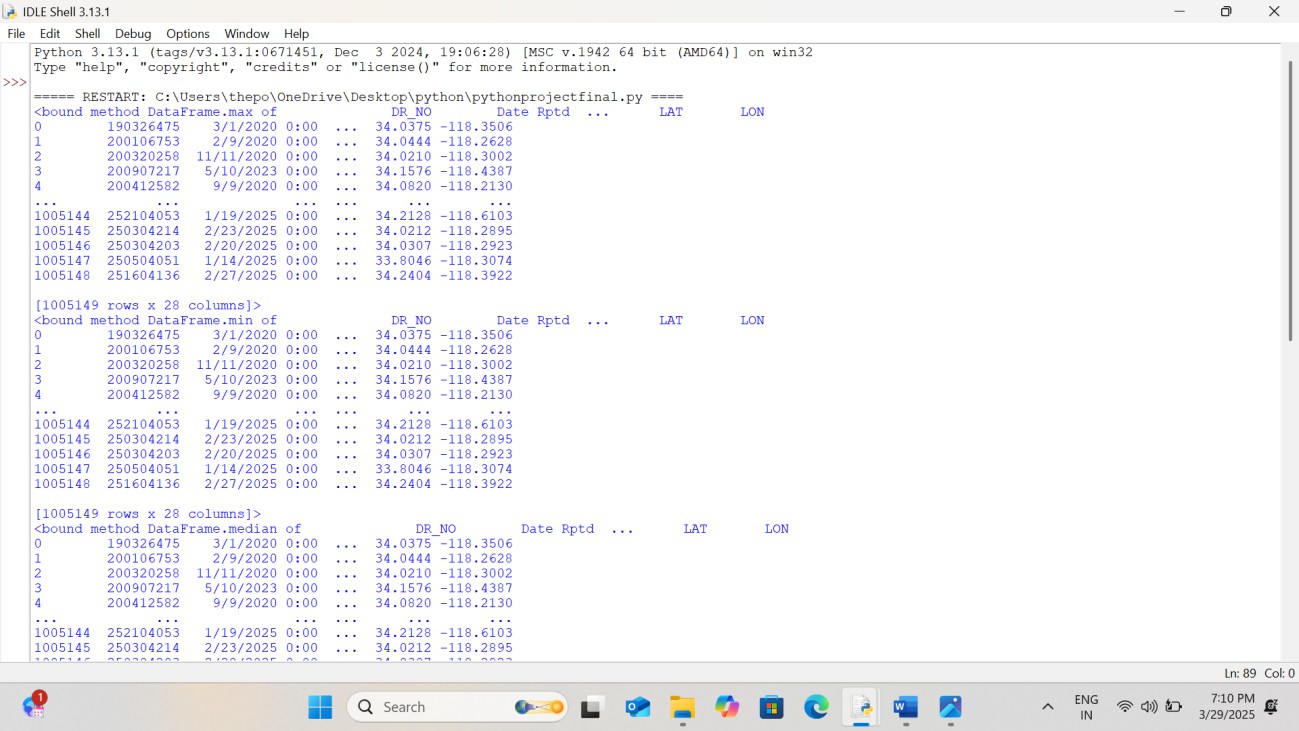
Ans – The output is

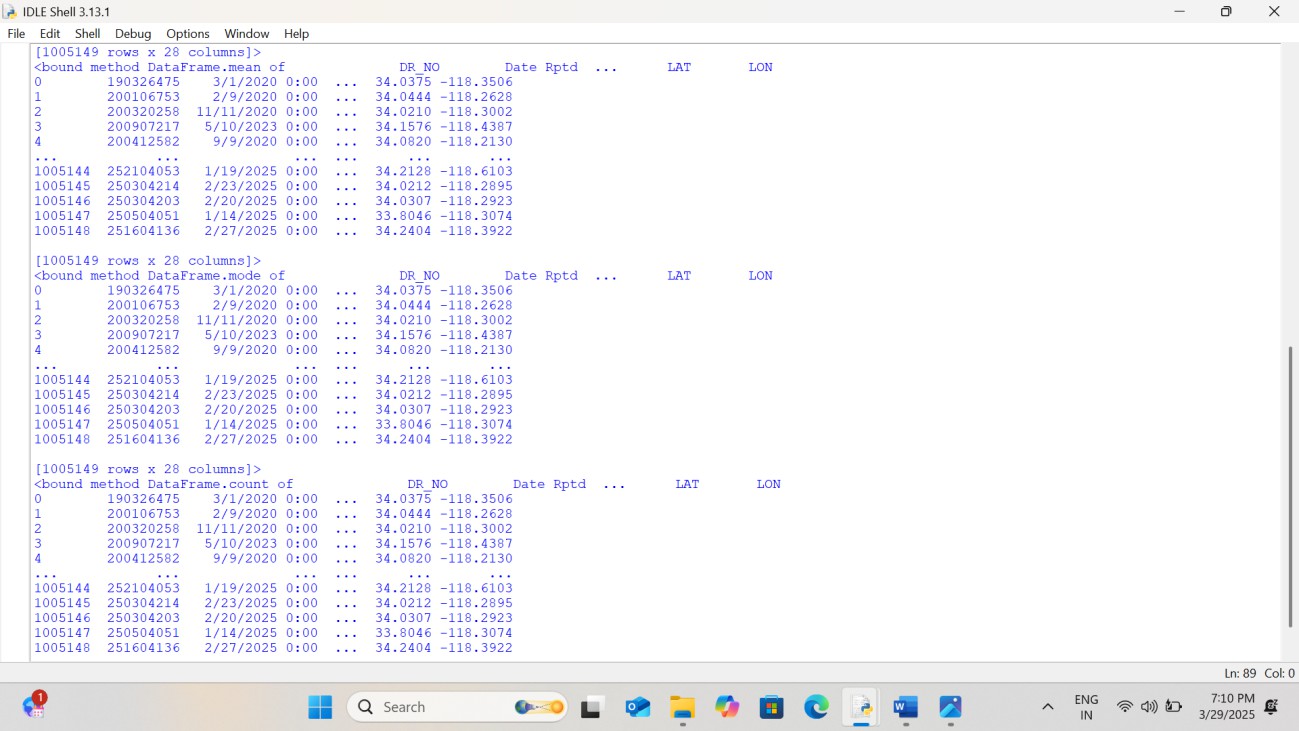


* 5. Checking for the max value min values median mode count and sum in one pic



Ans - The output of the code is



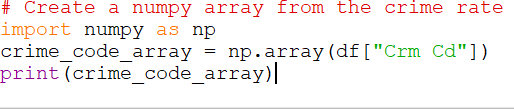


* A close up of a word  AI-generated content may be incorrect.6. Checking for the cleaning of the dataset

Ans – The output of the code is the

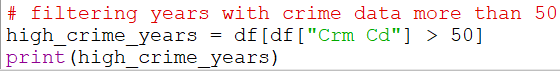


# CREATION OF NUMPY ARRAY 1.



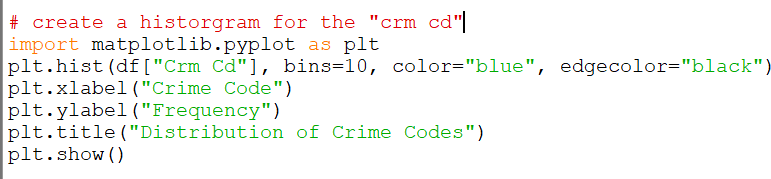
Ans – The output of the code is



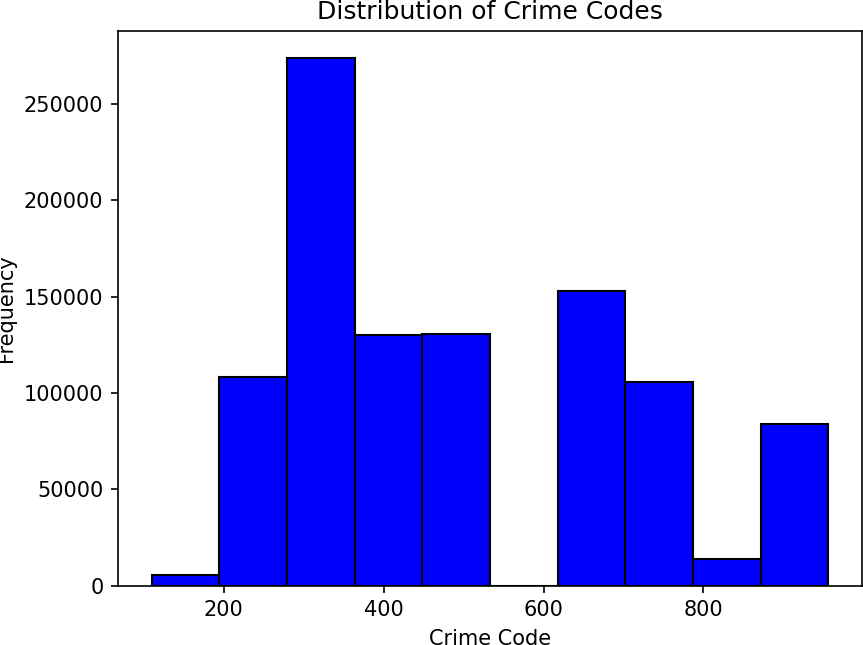
2. filtering years with crime data more than 50

# HISTOGRAM

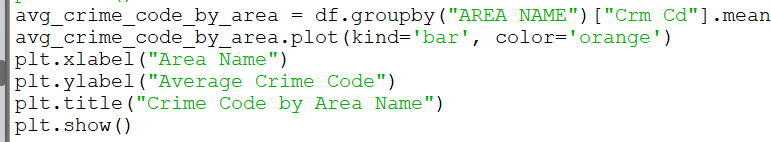
1. creating a histogram based on crime data



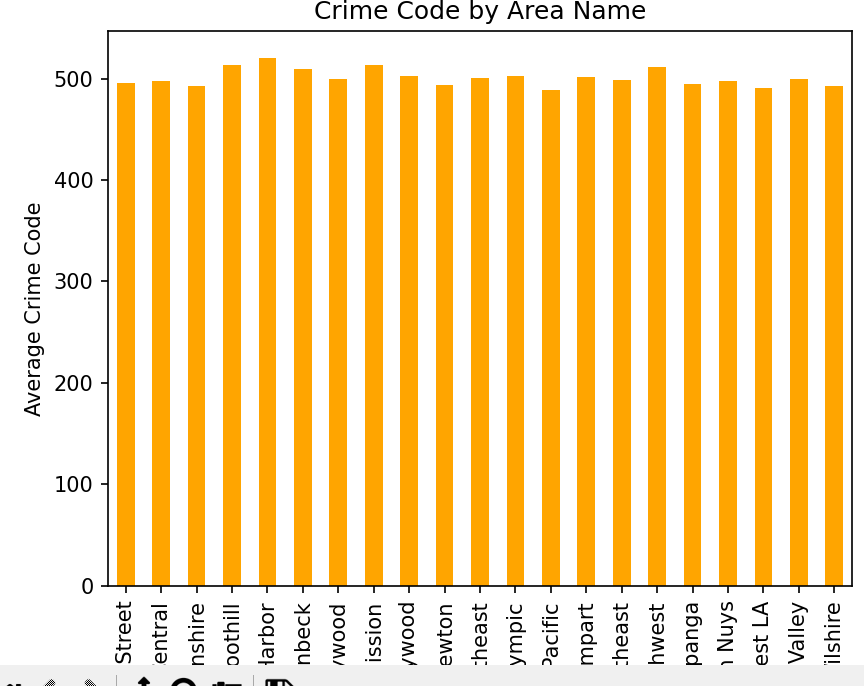
Ans – The output of the code is



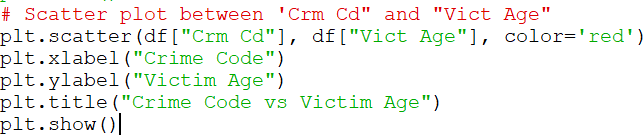
1. creating a Bar chart in which Bar chart comparing “ Crm Cd” across “Area Name”



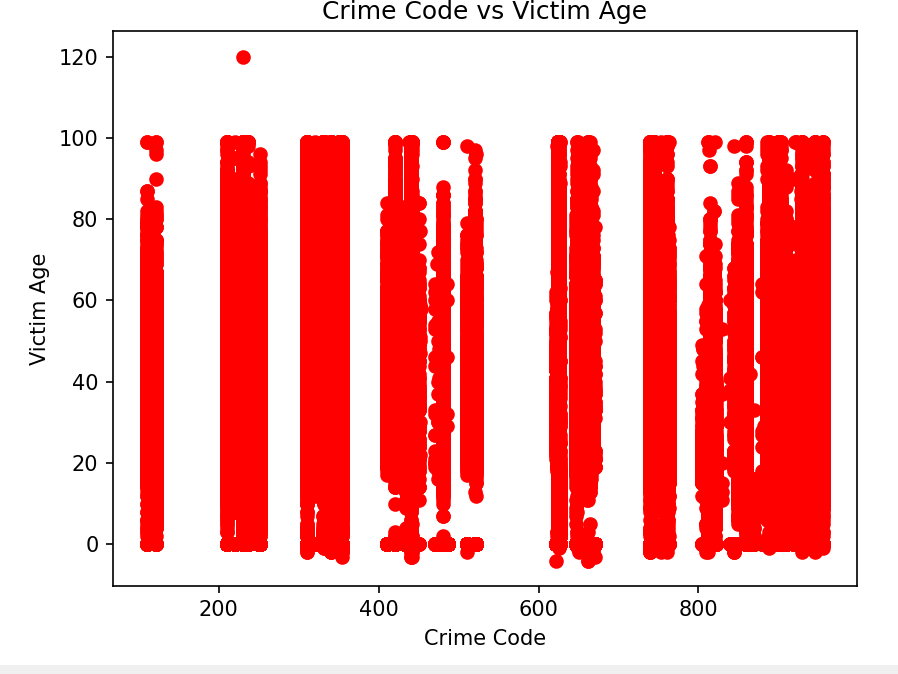
Ans – The output of the code is



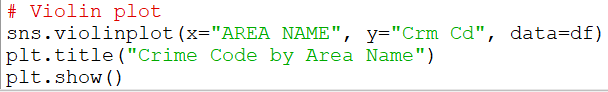
1. creating a scatter plot between “CRM Cd” and “Vict Age”:\



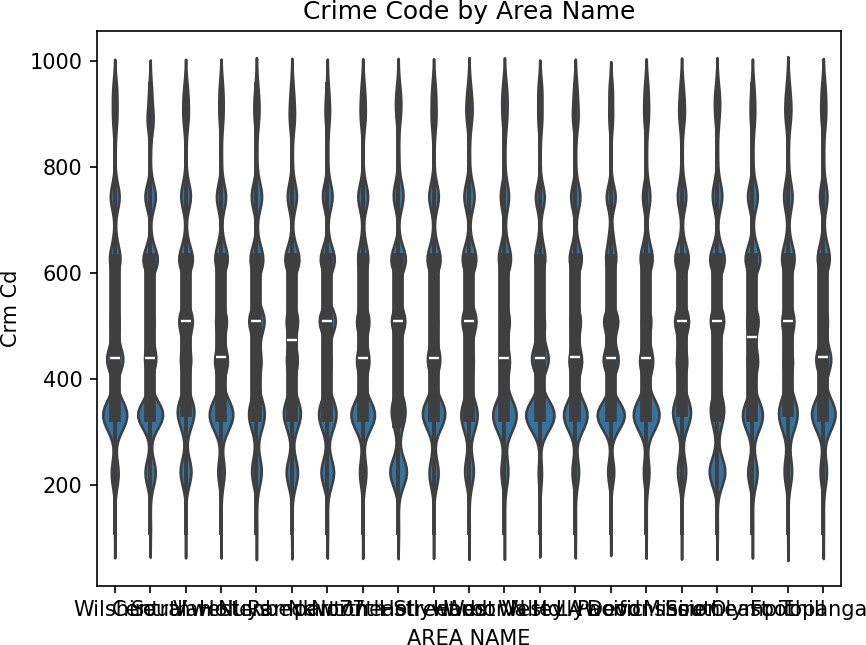
Ans – The output of the code is



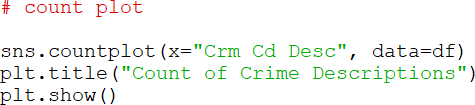
# Seaborn python libraries

1. Creating a Violin plot for “CRM Cd” grouped by “AREA NAME”:

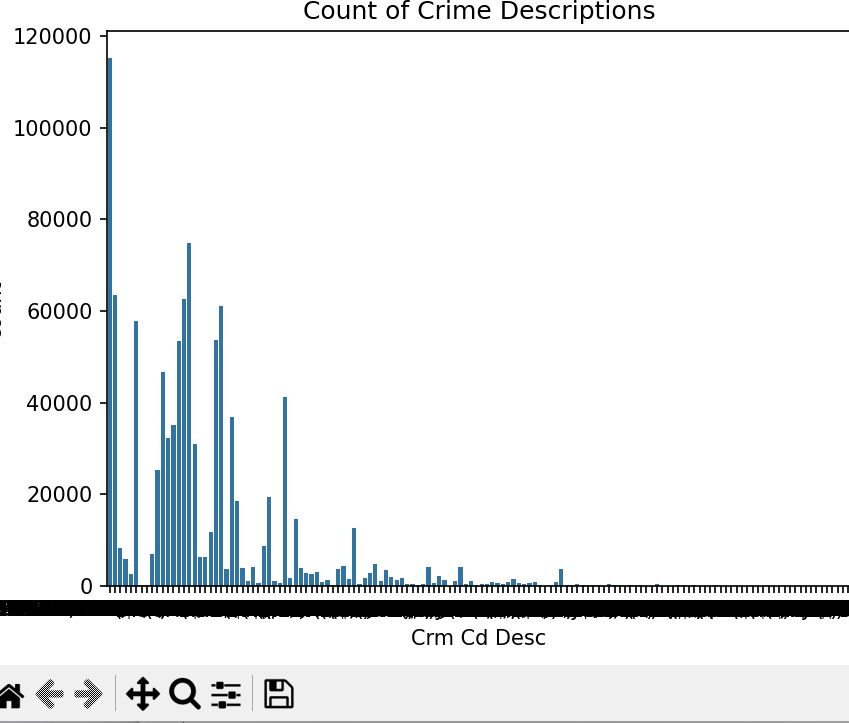
Ans – The output of the code is

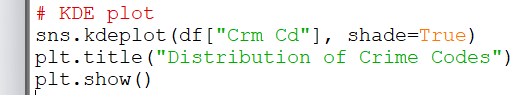


1. Creating a count plot

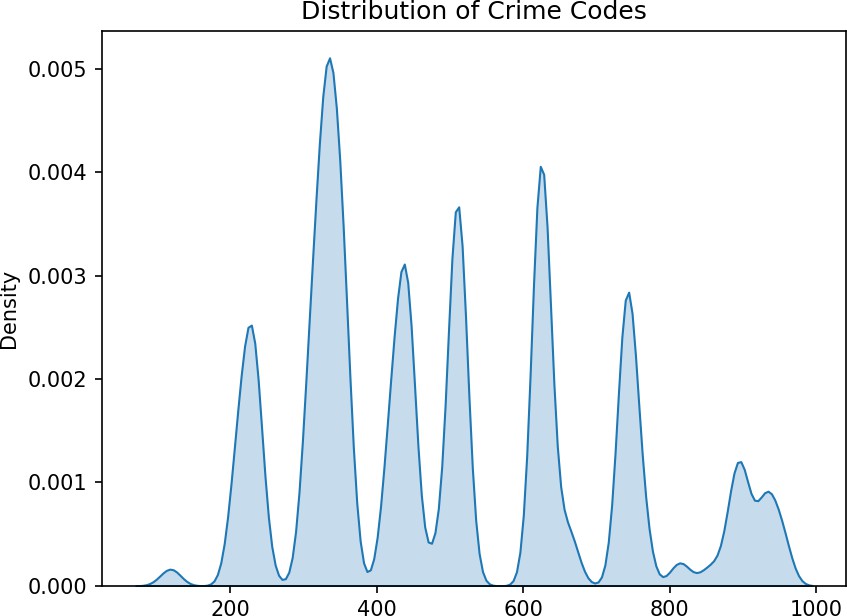


Ans – The output of the code is



1. Creating a KDE Graph

Ans – The output of the code is the



BELOW I HAVE WRITTEN ALL THE CODE WHICH I HAVE WRITE IN IDLE PYTHON

import warnings

warnings.simplefilter(action="ignore", category=FutureWarning)

import numpy as np import pandas as pd import seaborn as sns

import matplotlib.pyplot as plt import scipy.stats as st

# Importing the dataset

df = pd.read\_csv("crime.csv")

# Checking the dimensions of the dataset # print(df.shape)

# Listing the columns of the dataset

# print(df.columns)

# Viewing the first five rows of the dataset # print(df.head())

# Viewing the last five rows of the dataset # print(df.tail())

# Viewing all the information about the dataset # print(df.info())

# Descriptive statistics of the dataset # print(df.describe())

# Checking for missing values in the dataset and their total count # print(df.isnull().sum())

# Viewing the maximum values in the dataset # print(df.max())

# Viewing the minimum values in the dataset # print(df.min())

# Viewing the median values in the dataset # print(df.median())

# Viewing the mean values in the dataset # print(df.mean())

# Viewing the mode values in the dataset # print(df.mode())

# Counting non-null values in each column # print(df.count())

# Cleaning the dataset by dropping rows with missing values # print(df.dropna(inplace=True))

# Creating a numpy array from the crime rate # crime\_code\_array = np.array(df["Crm Cd"])

# print(crime\_code\_array)

# Filtering years with crime data greater than 50 high\_crime\_years = df[df["Crm Cd"] > 50]

# print(high\_crime\_years)

# Creating a histogram for the "Crm Cd" column

# plt.hist(df["Crm Cd"], bins=10, color="blue", edgecolor="black") # plt.xlabel("Crime Code")

# plt.ylabel("Frequency")

# plt.title("Distribution of Crime Codes") # plt.show()

# Creating a bar chart to show the average crime code by area

# avg\_crime\_code\_by\_area = df.groupby("AREA NAME")["Crm Cd"].mean() # avg\_crime\_code\_by\_area.plot(kind='bar', color='orange')

# plt.xlabel("Area Name")

# plt.ylabel("Average Crime Code")

# plt.title("Crime Code by Area Name") # plt.show()

# Creating a line graph to show the trend of crime code across dates # plt.plot(df["DATE OCC"], df["Crm Cd"], marker='o')

# plt.xlabel("Date of Occurrence") # plt.ylabel("Crime Code")

# plt.title("Trend of Crime Code Across Dates") # plt.show()

# Scatter plot between 'Crm Cd' and 'Vict Age'

# plt.scatter(df["Crm Cd"], df["Vict Age"], color='red') # plt.xlabel("Crime Code")

# plt.ylabel("Victim Age")

# plt.title("Crime Code vs Victim Age") # plt.show()

# Boxplot for "Date Rptd" distribution by year # import seaborn as sns

# sns.boxplot(x="Year", y="Date Rtpd", data=df) # plt.title("Crime Code Distribution by Year") # plt.show()

# Creating a heatmap to visualize the correlation between features # sns.heatmap(df.corr(), annot=True, cmap="coolwarm")

# plt.title("Feature Correlation Heatmap") # plt.show()

# Violin plot for crime code distribution by area

# sns.violinplot(x="AREA NAME", y="Crm Cd", data=df) # plt.title("Crime Code by Area Name")

# plt.show()

# Count plot for crime descriptions

# sns.countplot(x="Crm Cd Desc", data=df) # plt.title("Count of Crime Descriptions")

# plt.show()

# KDE plot for the distribution of crime codes # sns.kdeplot(df["Crm Cd"], shade=True)

# plt.title("Distribution of Crime Codes") # plt.show()