Phase-End Project: US Crime Data Exploration and Analysis

Problem Scenario: An organized quantitative and qualitative investigation is done to find trends in crime and disorder. Information on these patterns helps law enforcement agencies deploy resources more effectively. Crime analysis plays an important role in devising solutions to crime problems and formulating crime prevention strategies.

Q1. Check the dataset for any missing values, and if any are present, delete the row corresponding to that particular cell. Verify once more that 2022 is the only year for which data has been collected.

We To check the dataset for any missing values and delete the rows corresponding to those particular cells, you can use the dropna() method of pandas DataFrame To verify that 2022 is the only year for which data has been collected, I can extract the year from the 'Updated_On' column and check if it contains only the year 2022

We can use Python and libraries such as pandas and matplotlib to achieve this task

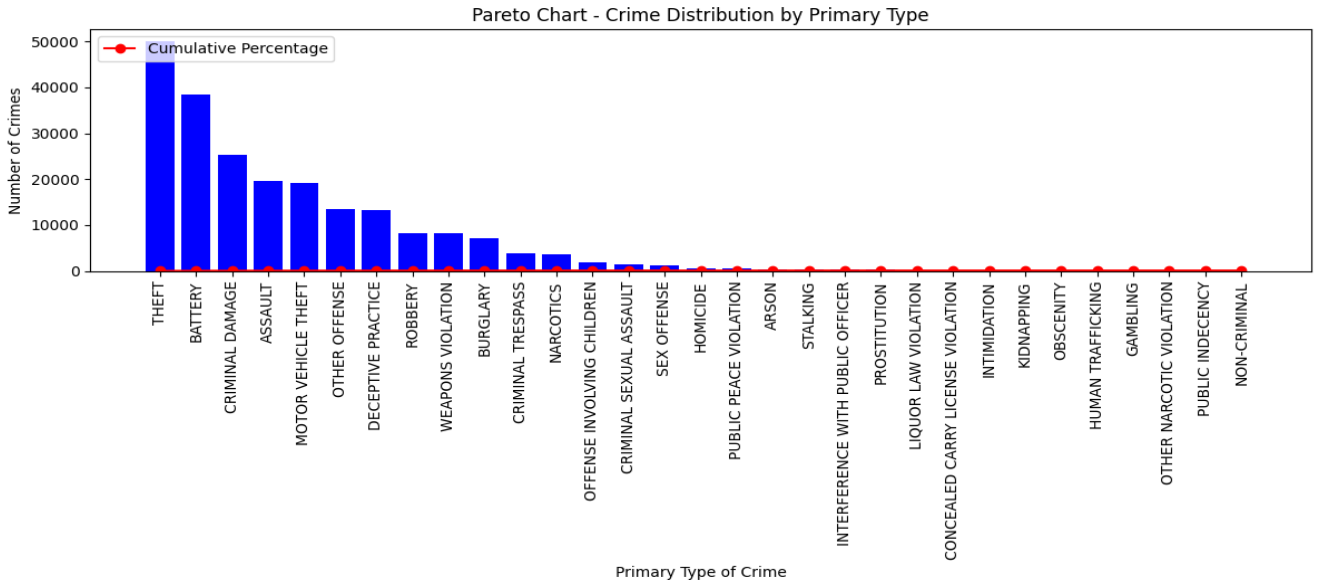
I can use isnull().values.any() to check if the DataFrame contains any missing values. If missing values are present, we use dropna() to remove the rows with missing values. After that, we convert the 'Updated_On' column to a datetime object to extract the year. We then use dt.year.unique() to get the unique years present in the 'Updated_On' column and check if it contains only the year 2022.

import pandas as pd

import matplotlib.pyplot as plt

```
if df.isnull().values.any()
```

Output:-



Q2. Sort the data in the provided dataset based on the number of primary types of crime instances, then create a Pareto chart to visualize the results. Next, sort the data according to the crime's location and description, and then create a histogram to represent the pertinent facts.

To perform the tasks as described, we'll need to use the pandas library for data manipulation and the matplotlib library for visualization. Here's how you can accomplish each step:

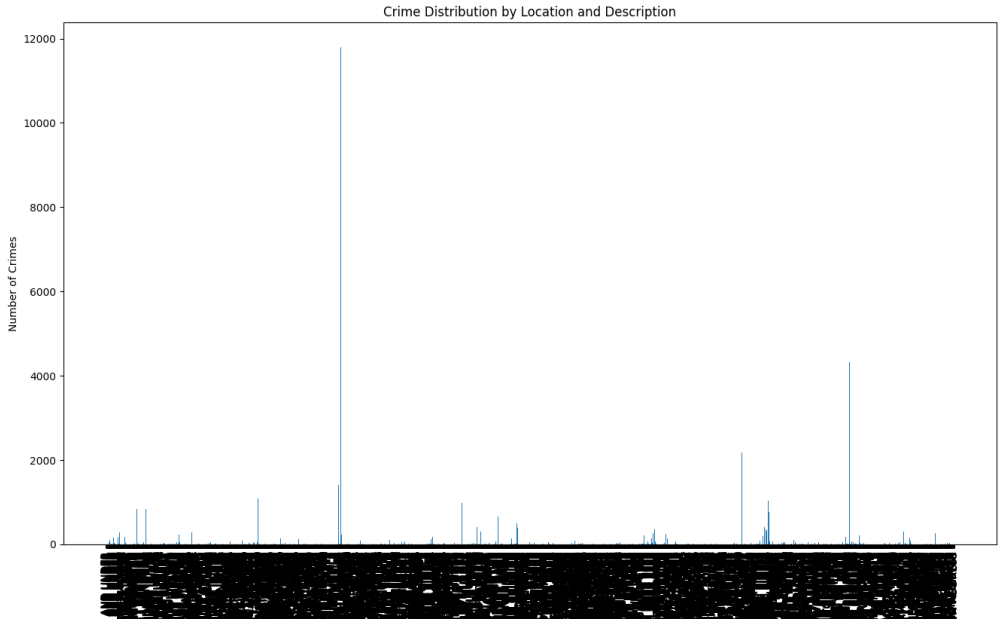
1. Sort the data based on the number of primary types of crime instances and create a Pareto chart:

```
import pandas as pd

import matplotlib.pyplot as plt

# Load the crime data into a DataFrame

df = pd.read_csv('filename')

# Calculate the count of each primary type of crime instances

crime_counts = df['Primary Type'].value_counts()
```

2. Sort the data based on the crime's location and description, and create a histogram:

```
# Sort the data based on the crime's location and description

sorted_location_df = df.sort_values(by=['Location', 'Description'], inplace=False)

# Create a histogram to represent the pertinent facts
```

The first code snippet creates a Pareto chart for the primary types of crime instances, which visually represents the most significant types of crimes in descending order. The second code snippet creates a histogram to show the frequency distribution of crime locations. Both visualizations will help you gain insights into the dataset and the distribution of crime instances.



Q3. Use conditional formatting in excel to generate data bars for the top occurring crime based on its primary type. Use red color to identify the crime where theft is over $500 in residence and arrest has not been done.

To apply conditional formatting in Excel using Python and generate data bars for the top occurring crime based on its primary type, as well as use red color to identify the crime where theft is over $500 in residence and arrest has not been done, we can use the openpyxl library. The openpyxl library allows us to interact with Excel files and apply conditional formatting rules programmatically.

```
import openpyxl

from openpyxl.styles import PatternFill, NamedStyle

from openpyxl.formatting.rule import CellIsRule

workbook = openpyxl.load_workbook('crime_data.xlsx')
```

Q4. Generate a frequency distribution chart of the data for the community area and note any skewness

To generate a frequency distribution chart of the data for the community area and analyze any skewness, you can use Python with pandas and matplotlib libraries.

```
import pandas as pd

import matplotlib.pyplot as plt
```

- Import the necessary libraries:
- Load the crime data into a DataFrame:

- import matplotlib.pyplot as pltCreate a frequency distribution of the data for the community area:
- Plot the frequency distribution chart:
  plt.figure(figsize=(12, 6))
  plt.bar(community_area_counts.index, community_area_counts.values)
- To analyze the skewness, you can calculate the skewness measure using the pandas
  skewness = community_area_counts.skew()
  print("Skewness:", skewness)
- The skewness measure indicates the direction and extent of skewness in the data. If skewness > 0, the distribution is positively skewed (tail on the right), if skewness < 0, the distribution is negatively skewed (tail on the left), and if skewness = 0, the distribution is approximately symmetric

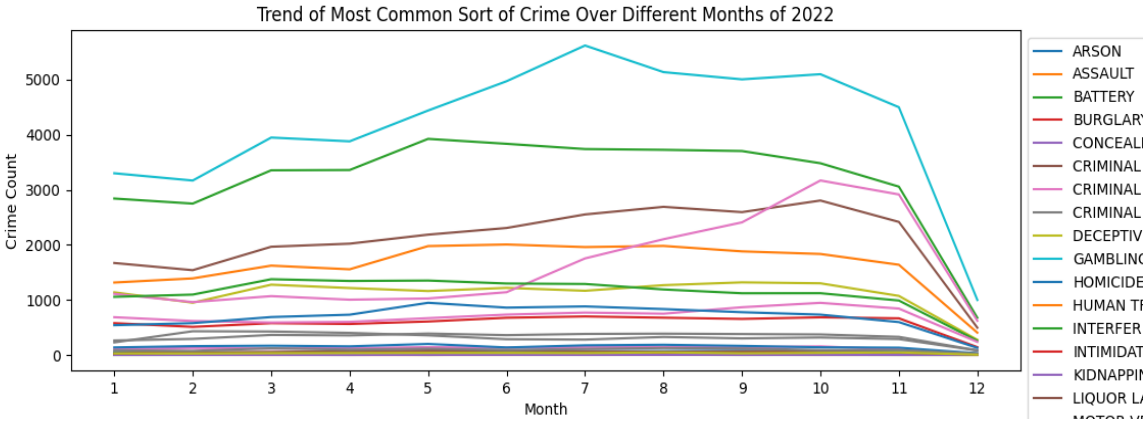
Q5. To find trends and patterns for the most common sort of crime over the different months of 2022, create a heatmap and sparklines.

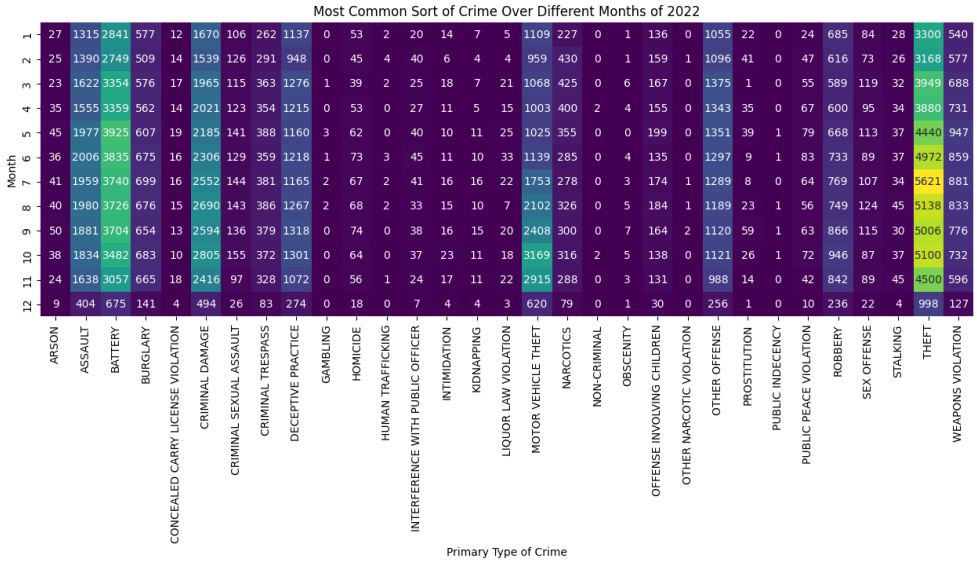
To find trends and patterns for the most common sort of crime over the different months of 2022, we can create a heatmap and sparklines using Python with pandas, seaborn, and matplotlib libraries. Here's how you can do it

- Import the necessary libraries:

```
import pandas as pd

import seaborn as sns

import matplotlib.pyplot as plt
```

- Load the crime data into a DataFrame:
- Convert the date column to a pandas datetime object:
- Extract the month from the date column
- Calculate the most common sort of crime over different months of 2022:
- Create a pivot table to prepare the data for the heatmap:
- Create a heatmap to visualize the most common sort of crime over different months of 2022
  plt.figure(figsize=(12, 8))
  sns.heatmap(crime_pivot, annot=True, fmt='', cmap='viridis', cbar=False)
  plt.xlabel('Primary Type of Crime')
  plt.ylabel('Month')
  plt.title('Most Common Sort of Crime Over Different Months of 2022')
  plt.show()
- Create sparklines to show the trend of the most common sort of crime:
- The heatmap will display the most common sort of crime over different months of 2022, and the sparklines will show the trend of each crime type over the months..

- 



Q6. To predict the future number of total crimes across all categories, create a trendline for the frequency of each key crime type occurring in 2022

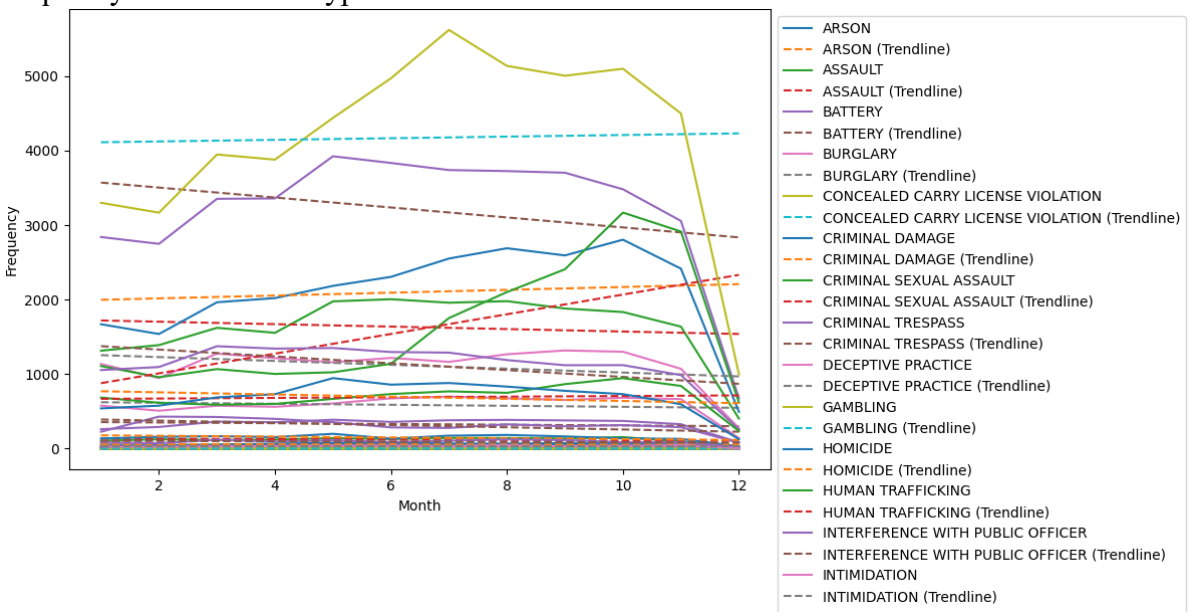
To predict the future number of total crimes across all categories and create trendlines for the frequency of each key crime type occurring in 2022, we can use Python with pandas and matplotlib libraries. We will fit a linear regression model to the crime data for each crime type and then use the model to predict future values.

- Load the crime data into a DataFrame:
- Convert the date column to a pandas datetime object:
- Extract the month from the date column

```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.linear_model import LinearRegression
```

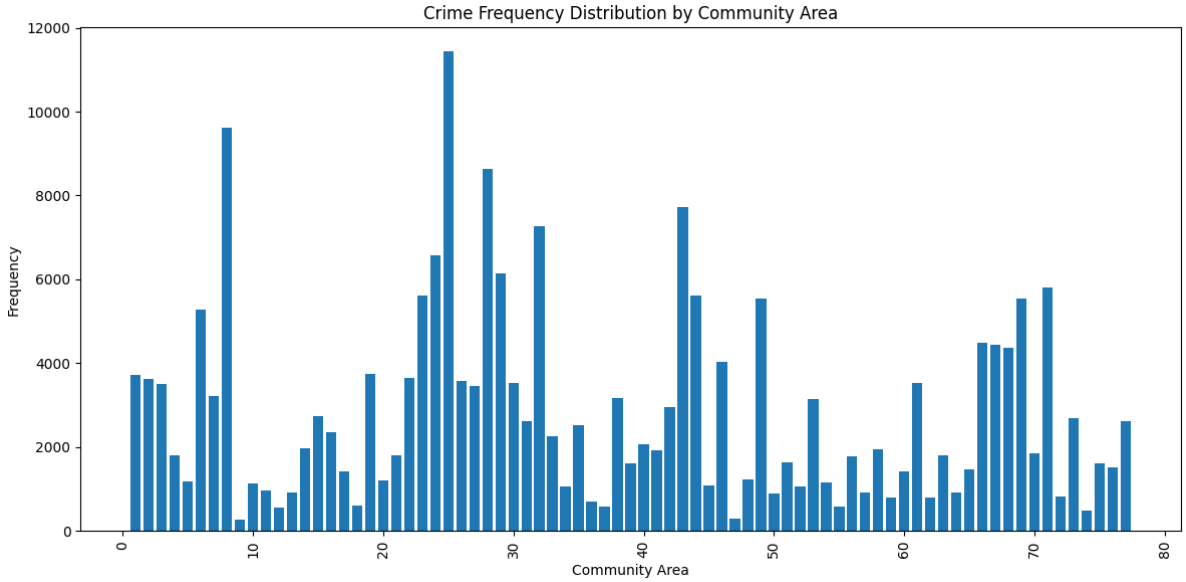
- Convert the date column to a pandas datetime object (if not already done):
- Filter the data for the year 2022:

- Create a pivot table to get the frequency of each key crime type for each month in 2022:
- Fit a linear regression model for each crime type to predict future values:
- Create a trendline plot for each key crime type:
- we used a linear regression model to fit a trendline for each key crime type. The model predicts the future frequency of each crime type for the next 12 months (you can change this to any desired period). The trendlines are plotted alongside the actual frequency of each crime type in 2022

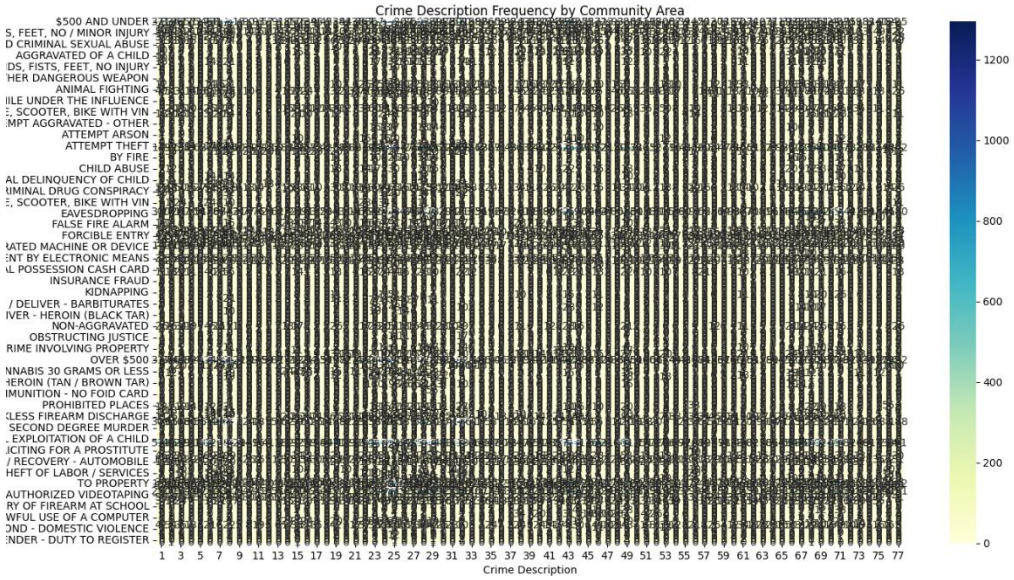


Q7. Create a cross-pivot table for the community area and crime description, then analyze the results

To create a cross-pivot table for the community area and crime description and analyze the results, we can use Python with pandas. A cross-pivot table, also known as a contingency table or a crosstab, will give us insights into the relationship between



community areas and crime descriptions.



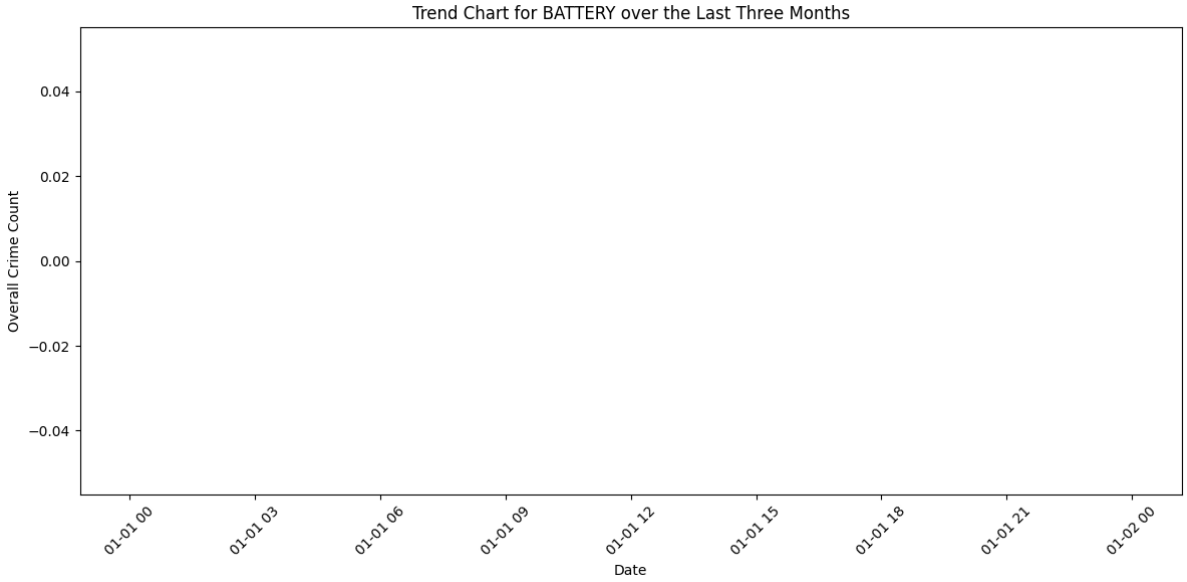
Q8. Create an excel dashboard that displays trend charts over the last three months for the several key categories of crimes based on the overall number of crimes that have occurred. (To make a dynamic chart, use form controls.)

To create an Excel dashboard with trend charts over the last three months for key categories of crimes based on the overall number of crimes that have occurred and make it dynamic using form controls, we'll use the openpyxl library in Python. We'll create an Excel template with form controls (data validation drop-downs) to select the key crime categories and then insert trend charts based on the selected categories.

- Install the required libraries:

  pip install openpyxl

- Prepare the Excel template with form controls:
- Create an Excel template file (e.g., crime_dashboard_template.xlsx) and set up the
- Data: This sheet will contain the crime data, including columns for 'Date', 'Primary Type', and other relevant information.
- Dashboard: This sheet will be the dashboard where we'll insert the form controls and trend charts.
- On the Dashboard sheet, insert a data validation drop-down cell for selecting key crime categories. Follow these steps manually in Excel:
- Select a cell where you want the drop-down to appear.
- Go to "Data" > "Data Validation" on the Excel ribbon.
- In the "Data Validation" dialog box, select "List" from the "Allow" drop-down.
- In the "Source" field, enter the range of key crime categories in your data, for example, Data!$B$2:$B$10.
- Click "OK" to apply the data validation drop-down.



Q9. Create a SQL database containing data related to the case number, primary crime category, crime description, crime location, and arrest status using the dataset.

```
/****** Object:  Table [dbo].[uscrime$]    Script Date:
7/16/2023 1:46:46 PM ******/
SET ANSI_NULLS ON
GO

SET QUOTED_IDENTIFIER ON
GO

CREATE TABLE [dbo].[uscrime$](
        [id] int NOT NULL,
      [case_number] [nvarchar](max) NULL,
      [primary_type] [nvarchar](max) NULL,
```

```
	[description] [nvarchar](max) NULL,
	[location_description] [nvarchar](max) NULL,
	[arrest] [bit] NULL,
)

GO
```

Q10. Make a database in SQL where theft costs more than $500.

```
select * from uscrime$ where description = 'OVER $500'
select count(*) from uscrime$ where description = 'OVER $500'
select count(*) from uscrime$
select * from uscrime$ where id = '1'
```

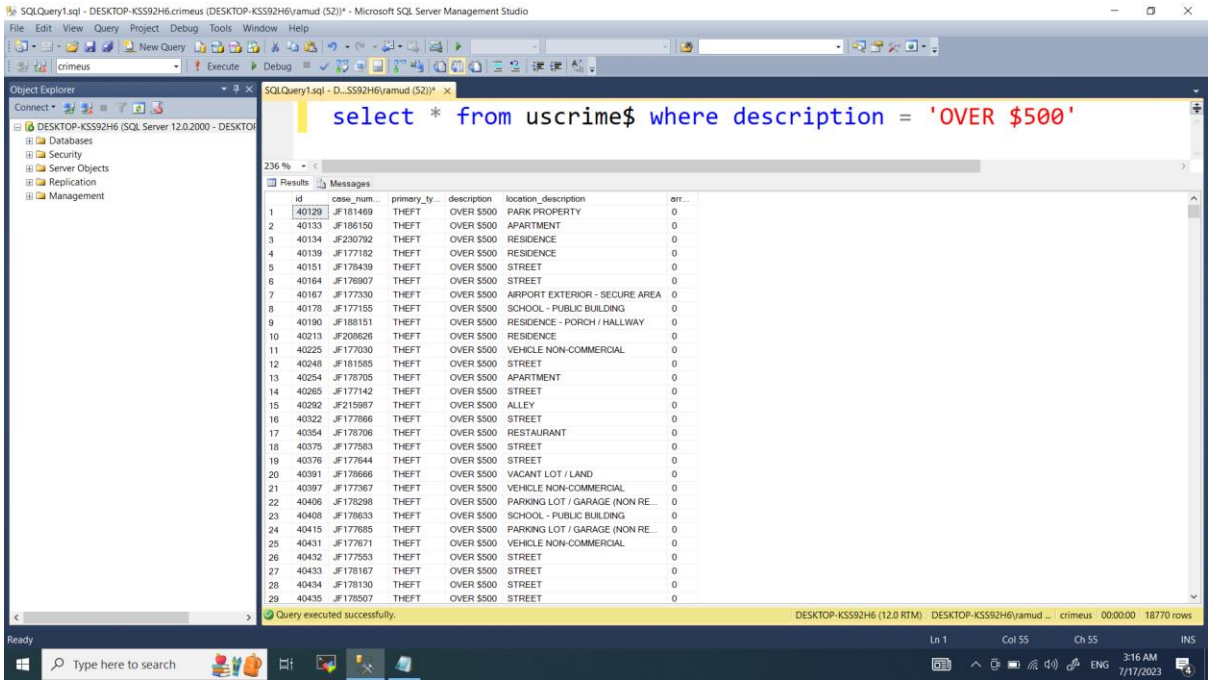
Q11. Determine the overall number of cases for each major category of crime.

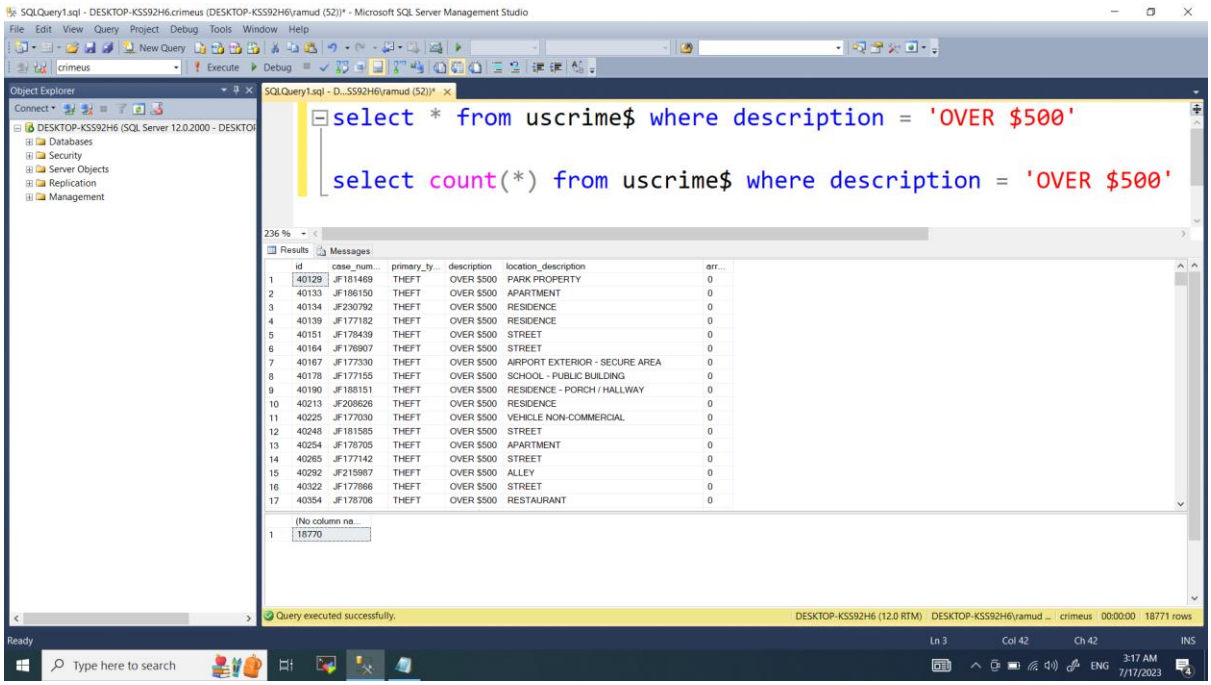
```
SELECT primary_type, COUNT(*) AS number_of_incedent FROM
uscrime$ WHERE location_description = 'APARTMENT' GROUP BY
primary_type ORDER BY COUNT(*) DESC
```

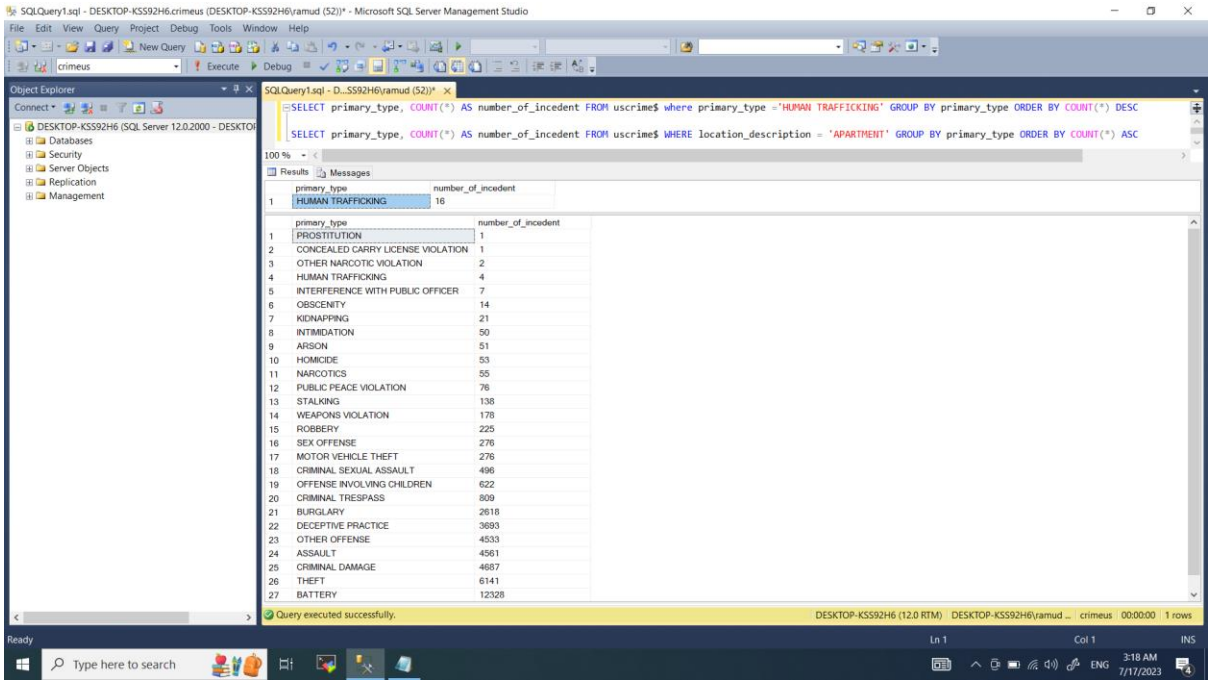
Q12. Apply 1NF normalization to the dataset provided.

```
SELECT primary_type, COUNT(*) AS number_of_incedent FROM
uscrime$ where primary_type ='HUMAN TRAFFICKING' GROUP BY
primary_type ORDER BY COUNT(*) DESC

SELECT primary_type, COUNT(*) AS number_of_incedent FROM
uscrime$ WHERE location_description = 'APARTMENT' GROUP BY
primary_type ORDER BY COUNT(*) ASC


SELECT COUNT(*) FROM uscrime$ where primary_type =
'PROSTITUTION'
```

Sql output:-

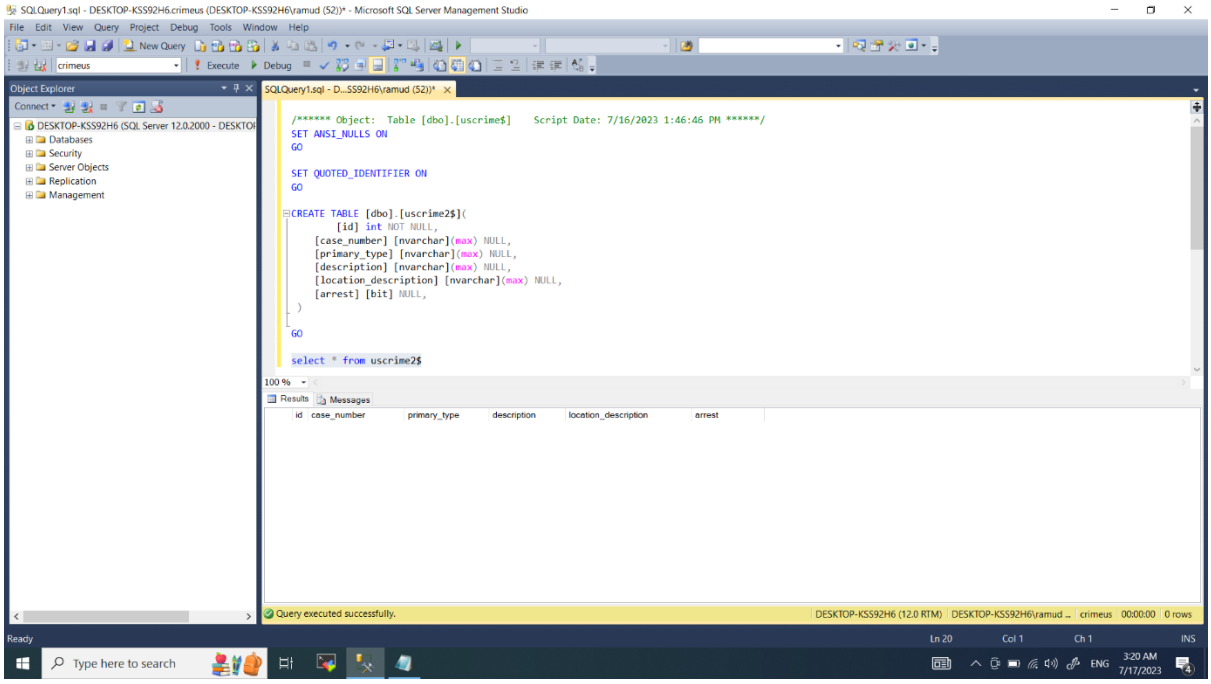

SQLQuery1.sql - DESKTOP-KSS92H6.crimeus (DESKTOP-KSS92H6\ramud (52))* - Microsoft SQL Server Management Studio
File Edit View Query Project Debug Tools Window Help
New Query
crimeus
Execute
Debug
Object Explorer
Connect
DESKTOP-KSS92H6 (SQL Server 12.0.2000 - DESKTOP
Databases
Security
Server Objects
Replication
Management
SQLQuery1.sql - D...SS92H6\ramud (52))*
select * from uscrime$ where description = 'OVER $500'
236 %
Results
Messages
id case_num... primary_ty... description location_description arr...
1 40129 JF181469 THEFT OVER $500 PARK PROPERTY 0
2 40133 JF186150 THEFT OVER $500 APARTMENT 0
3 40134 JF230792 THEFT OVER $500 RESIDENCE 0
4 40139 JF177182 THEFT OVER $500 RESIDENCE 0
5 40151 JF178439 THEFT OVER $500 STREET 0
6 40164 JF176907 THEFT OVER $500 STREET 0
7 40167 JF177330 THEFT OVER $500 AIRPORT EXTERIOR - SECURE AREA 0
8 40178 JF177155 THEFT OVER $500 SCHOOL - PUBLIC BUILDING 0
9 40190 JF188151 THEFT OVER $500 RESIDENCE - PORCH / HALLWAY 0
10 40213 JF208626 THEFT OVER $500 RESIDENCE 0
11 40225 JF177030 THEFT OVER $500 VEHICLE NON-COMMERCIAL 0
12 40248 JF181585 THEFT OVER $500 STREET 0
13 40254 JF178705 THEFT OVER $500 APARTMENT 0
14 40265 JF177142 THEFT OVER $500 STREET 0
15 40292 JF215987 THEFT OVER $500 ALLEY 0
16 40322 JF177866 THEFT OVER $500 STREET 0
17 40354 JF178706 THEFT OVER $500 RESTAURANT 0
18 40375 JF177583 THEFT OVER $500 STREET 0
19 40376 JF177644 THEFT OVER $500 STREET 0
20 40391 JF178666 THEFT OVER $500 VACANT LOT / LAND 0
21 40397 JF177367 THEFT OVER $500 VEHICLE NON-COMMERCIAL 0
22 40406 JF178298 THEFT OVER $500 PARKING LOT / GARAGE (NON RE... 0
23 40408 JF178633 THEFT OVER $500 SCHOOL - PUBLIC BUILDING 0
24 40415 JF177685 THEFT OVER $500 PARKING LOT / GARAGE (NON RE... 0
25 40431 JF177671 THEFT OVER $500 VEHICLE NON-COMMERCIAL 0
26 40432 JF177553 THEFT OVER $500 STREET 0
27 40433 JF178167 THEFT OVER $500 STREET 0
28 40434 JF178130 THEFT OVER $500 STREET 0
29 40435 JF178507 THEFT OVER $500 STREET 0
Query executed successfully.
DESKTOP-KSS92H6 (12.0 RTM) DESKTOP-KSS92H6\ramud ... crimeus 00:00:00 18770 rows
Ready
Ln 1 Col 55 Ch 55 INS
Type here to search
ENG
3:16 AM
7/17/2023


SQLQuery1.sql - DESKTOP-KSS92H6.crimeus (DESKTOP-KSS92H6\ramud (52))* - Microsoft SQL Server Management Studio
File Edit View Query Project Debug Tools Window Help
New Query
crimeus
Execute
Debug
Object Explorer
Connect
DESKTOP-KSS92H6 (SQL Server 12.0.2000 - DESKTOP
Databases
Security
Server Objects
Replication
Management
SQLQuery1.sql - D...SS92H6\ramud (52))*
select * from uscrime$ where description = 'OVER $500'
select count(*) from uscrime$ where description = 'OVER $500'
236 %
Results
Messages
id case_num... primary_ty... description location_description arr...
1 40129 JF181469 THEFT OVER $500 PARK PROPERTY 0
2 40133 JF186150 THEFT OVER $500 APARTMENT 0
3 40134 JF230792 THEFT OVER $500 RESIDENCE 0
4 40139 JF177182 THEFT OVER $500 RESIDENCE 0
5 40151 JF178439 THEFT OVER $500 STREET 0
6 40164 JF176907 THEFT OVER $500 STREET 0
7 40167 JF177330 THEFT OVER $500 AIRPORT EXTERIOR - SECURE AREA 0
8 40178 JF177155 THEFT OVER $500 SCHOOL - PUBLIC BUILDING 0
9 40190 JF188151 THEFT OVER $500 RESIDENCE - PORCH / HALLWAY 0
10 40213 JF208626 THEFT OVER $500 RESIDENCE 0
11 40225 JF177030 THEFT OVER $500 VEHICLE NON-COMMERCIAL 0
12 40248 JF181585 THEFT OVER $500 STREET 0
13 40254 JF178705 THEFT OVER $500 APARTMENT 0
14 40265 JF177142 THEFT OVER $500 STREET 0
15 40292 JF215987 THEFT OVER $500 ALLEY 0
16 40322 JF177866 THEFT OVER $500 STREET 0
17 40354 JF178706 THEFT OVER $500 RESTAURANT 0
(No column na...
1 18770
Query executed successfully.
DESKTOP-KSS92H6 (12.0 RTM) DESKTOP-KSS92H6\ramud ... crimeus 00:00:00 18771 rows
Ready
Ln 3 Col 42 Ch 42 INS
Type here to search
ENG
3:17 AM
7/17/2023



	primary_type	number_of_incedent
1	HUMAN TRAFFICKING	16

	primary_type	number_of_incedent
1	PROSTITUTION	1
2	CONCEALED CARRY LICENSE VIOLATION	1
3	OTHER NARCOTIC VIOLATION	2
4	HUMAN TRAFFICKING	4
5	INTERFERENCE WITH PUBLIC OFFICER	7
6	OBSCENITY	14
7	KIDNAPPING	21
8	INTIMIDATION	50
9	ARSON	51
10	HOMICIDE	53
11	NARCOTICS	55
12	PUBLIC PEACE VIOLATION	76
13	STALKING	138
14	WEAPONS VIOLATION	178
15	ROBBERY	225
16	SEX OFFENSE	276
17	MOTOR VEHICLE THEFT	276
18	CRIMINAL SEXUAL ASSAULT	496
19	OFFENSE INVOLVING CHILDREN	622
20	CRIMINAL TRESPASS	809
21	BURGLARY	2618
22	DECEPTIVE PRACTICE	3693
23	OTHER OFFENSE	4533
24	ASSAULT	4561
25	CRIMINAL DAMAGE	4687
26	THEFT	6141
27	BATTERY	12328



id	case_number	primary_type	description	location_description	arrest