BASEMAP VISUALISATION IN PYTHON FOR CYBER CRIMES IN 2018

AIM:

To Visualize the Cyber Crimes in 2018 csv file using the Base map, Pandas, Shapefile and Matplotlib python libraries with the help of Jupyter Notebook and Python 3.10.6.

PROCEDURE:

Download the csv file from the https://ncrb.gov.in/en/crime-in-india-table-additional-table-and-chapter-contents?page=27 and use the Pandas library to load this CSV file, and convert it into the dataframe. read_csv () method is used to read CSV files.

INTRODUCTION ABOUT THE PYTHON LIBRARIES:

PANDAS:

- PANDAS – Python Data Analysis
- Pandas library was created by Wes McKinney in 2008.
- Pandas is an open-source library, that is made mainly for working with the labeled data.
- Pandas is fast and it has high performance.
- Pandas provides various data structures and operations for manipulating numerical data.

Pandas functions and methods used in the project:

drop () - This method is used to remove the unwanted rows and columns in the datasets.

read_csv () - This method is used to import the csv file into the Jupyter Notebook or other python applications.

head () – This method will always show the first five rows and columns, if we don't give any value to the head method. Suppose you give any value to this method it shows the given number of rows and columns.

MATPLOTLIB**:**

- Matplotlib was created by John D. Hunter in 2002.
- Matplotlib is one of the most popular Python packages used for data visualization.
- It is a cross-platform library for making 2D plots as well as 3D plots from data in arrays.
- You can generate plots, histograms, bar charts, scatter plots and other types of charts with just a few lines of code.
- It's often used in web application servers, shells and python scripts.
- Pyplot is a Matplotlib module that provides simple functions for adding plot elements such as lines, images, text, etc. to the axes in the current figure.

Matplotlib functions and methods used in my project:

scatter () - Scatter plots are used to plot data points on horizontal and vertical axis in the attempt to show how much one variable is affected by another. Each row in the data table is represented by a marker the position depends on its values in the columns set on the X and Y axes. A third variable can be set to correspond to the color or size of the markers, thus adding yet another dimension to the plot.

text () - Add text at an arbitrary location of the Axes.

BASEMAP:

- Basemap and mpl_toolkits are necessary to show the map.
- Basemap cannot be directly import with the help of the mpl_toolkits the Basemap can be imported. In the same way you not need to install the mpl_toolkits, if you install the matplotlib library it automatically installed.
- The map is created using the Basemap class, which has many options. Without passing any option, the map has the Plate Carrée projection centered at longitude and latitude = 0.
- With the help of the shapefile, you imported the India map.
- After setting the map, we can draw what we want. In this case, the coast lines layer, which comes already with the library, using the method drawcoastlines ().

These are about the introduction to the python libraries used here.

ALGORITHM FOR PLOTTING THE GRAPH:

Step 1: Start the program.

Step 2: Import the pandas and matplotlib libraries with the alias name np and pd respectively.

Step 3: Define the empty list un1 and un4.

Step 4: With the help of the read_csv () method to open and read the data in the csv file using the alias name pd.

Step 5: Remove the unwanted data using the drop () method and get the wanted data using the head () method in pandas.

Step 6: Using the for loop to store the elements from the csv file to un1 and un4 lists.

Step 7: With the help of the datum in the lists to plot the scatter and line plot in the single graph.

Step 8: And give the labels and title.

Step 9: Display the graph

Step 10: Stop the program.

STEP BY STEP PROGRAM EXECUTION:

Given inputs

```
In [32]: # first import the necessary libraries
         import pandas as pd
         import matplotlib.pyplot as plt
```

Given input

```
In [20]: # define the empty list
         un1=[]
         un4=[]
```

Given input

```
In [8]: # read the csv file using read_csv() method
        df=pd.read_csv(r"C:\Users\OM MURUGA\Desktop\cybercrime.csv",encoding='ISO-8859-1')
        print(df)
```

Output for given input.

Out[9]:

	TABLE 9A.1	Unnamed: 1	Unnamed: 2	Unnamed: 3	Unnamed: 4	Unnamed: 5	Unnamed: 6	Unnamed: 7
0	Cyber Crimes (State/UT-wise) □ 2016-2018	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1	S. No	State/UT	2016.0	2017.0	2018.0	Percentage Share of State/UT (2018)	Mid-Year Projected Population (in Lakhs) (2018)+	Rate of Total Cyber Crimes (2018)++
2	1	2	3.0	4.0	5.0	6	7	8
3	STATES:	NaN	NaN	NaN	NaN	NaN	NaN	NaN
4	1	Andhra Pradesh	616.0	931.0	1207.0	4.4	520.3	2.3
5	2	Arunachal Pradesh	4.0	1.0	7.0	0.0	14.9	0.5
6	3	Assam	696.0	1120.0	2022.0	7.4	340.4	5.9
7	4	Bihar	309.0	433.0	374.0	1.4	1183.3	0.3
8	5	Chhattisgarh	90.0	171.0	139.0	0.5	284.7	0.5
9	6	Goa	31.0	13.0	29.0	0.1	15.3	1.9
10	7	Gujarat	362.0	458.0	702.0	2.6	673.2	1.0
11	8	Haryana	401.0	504.0	418.0	1.5	284.0	1.5
12	9	Himachal Pradesh	31.0	56.0	69.0	0.3	72.7	0.9
13	10	Jammu & Kashmir	28.0	63.0	73.0	0.3	134.3	0.5
14	11	Jharkhand	259.0	720.0	930.0	3.4	370.5	2.5
15	12	Karnataka	1101.0	3174.0	5839.0	21.4	654.5	8.9
16	13	Kerala	283.0	320.0	340.0	1.2	350.0	1.0
17	14	Madhya Pradesh	258.0	490.0	740.0	2.7	814.7	0.9
18	15	Maharashtra	2380.0	3604.0	3511.0	12.9	1213.9	2.9
19	16	Manipur	11.0	74.0	29.0	0.1	30.8	0.9
20	17	Meghalaya	39.0	39.0	74.0	0.3	32.0	2.3
21	18	Mizoram	1.0	10.0	6.0	0.0	11.8	0.5
22	19	Nagaland	2.0	0.0	2.0	0.0	21.3	0.1
23	20	Odisha	317.0	824.0	843.0	3.1	435.5	1.9
24	21	Punjab	102.0	176.0	239.0	0.9	297.0	0.8
25	22	Rajasthan	941.0	1304.0	1104.0	4.1	765.9	1.4
26	23	Sikkim	1.0	1.0	1.0	0.0	6.6	0.2
27	24	Tamil Nadu	144.0	228.0	295.0	1.1	754.6	0.4
28	25	Telangana	593.0	1209.0	1205.0	4.4	370.3	3.3
29	26	Tripura	8.0	7.0	20.0	0.1	39.6	0.5
30	27	Uttar Pradesh	2639.0	4971.0	6280.0	23.0	2230.0	2.8
31	28	Uttarakhand	62.0	124.0	171.0	0.6	110.6	1.5
32	29	West Bengal	478.0	568.0	335.0	1.2	965.0	0.3
33	NaN	TOTAL STATE(S)	12187.0	21593.0	27004.0	99.1	12997.9	2.1
34	UNION TERRITORIES:	NaN	NaN	NaN	NaN	NaN	NaN	NaN
35	30	A & N Islands	3.0	3.0	7.0	0.0	4.0	1.8
36	31	Chandigarh	26.0	32.0	30.0	0.1	11.7	2.6
37	32	D&N Haveli	1.0	1.0	0.0	0.0	5.3	0.0
38	33	Daman & Diu	0.0	0.0	0.0	0.0	4.0	0.0
39	34	Delhi UT	98.0	162.0	189.0	0.7	195.6	1.0
40	35	Lakshadweep	0.0	0.0	4.0	0.0	0.7	6.0
41	36	Puducherry	2.0	5.0	14.0	0.1	14.8	0.9
42	NaN	TOTAL UT(S)	130.0	203.0	244.0	0.9	236.0	1.0
43	NaN	TOTAL (ALL INDIA)	12317.0	21796.0	27248.0	100.0	13233.8	2.1
44	Note : i) '++' Crime Rate is calculated as Cri...	NaN	NaN	NaN	NaN	NaN	NaN	TABLE 9A.1 Page 1 of 1
45	NaN	ii) '+ Population Source: Technical group on ...	NaN	NaN	NaN	NaN	NaN	NaN
46	NaN	III) As per data provided by States/UTs	NaN	NaN	NaN	NaN	NaN	NaN
47	NaN	iv) Clarifications are pending from West Benga...	NaN	NaN	NaN	NaN	NaN	NaN
48	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

Given input

In [18]:
```
# remove the unwanted rows and columns using the drop() method and get the wanted rows and colums using the head() method
df.drop(columns ='Unnamed: 2')
df.head(32)
new=df.drop(labels=['Unnamed: 2','Unnamed: 3','Unnamed: 5','Unnamed: 6','Unnamed: 7'],axis=1)
new.head(32)
b=new.drop(0)
dfi=b[["Unnamed: 1","Unnamed: 4"]]
a=dfi.head(32)
print(a)
b=a['Unnamed: 4']
#print(b[3:32])
```

Output for given input.

	Unnamed: 1	Unnamed: 4
1	State/UT	2018.0
2	2	5.0
3	NaN	NaN
4	Andhra Pradesh	1207.0
5	Arunachal Pradesh	7.0
6	Assam	2022.0
7	Bihar	374.0
8	Chhattisgarh	139.0
9	Goa	29.0
10	Gujarat	702.0
11	Haryana	418.0
12	Himachal Pradesh	69.0
13	Jammu & Kashmir	73.0
14	Jharkhand	930.0
15	Karnataka	5839.0
16	Kerala	340.0
17	Madhya Pradesh	740.0
18	Maharashtra	3511.0
19	Manipur	29.0
20	Meghalaya	74.0
21	Mizoram	6.0
22	Nagaland	2.0
23	Odisha	843.0
24	Punjab	239.0
25	Rajasthan	1104.0
26	Sikkim	1.0
27	Tamil Nadu	295.0
28	Telangana	1205.0
29	Tripura	20.0
30	Uttar Pradesh	6280.0
31	Uttarakhand	171.0
32	West Bengal	335.0

Given input

In [21]:
```
# using the for loop to store the elements in the list
for i in range(4,33,1):
    un4.append(b[i])
print(un4)
c=a['Unnamed: 1']
for i in range(4,33,1):
    un1.append(c[i])
print(un1)
```

ALGORITHM FOR SHOWING THE MAP AND PLOTTING DATA IN THE MAP:

Step 1: Start the program.

Step 2: Import the Basemap from mpl_toolkits and import the matplotlib, shapefile and pandas.

Step 3: Create the empty list.

Step 4: With the help of the read_csv () method to open and read the data in the csv file using the alias name pd.

Step 5: Remove the unwanted data using the drop () method and get the wanted data using the head () method in pandas.

Step 6: Using the for loop to append the data in the list un4.

Step 7: With the help of the Basemap import the India map using the latitude and longitude and draw the coastlines and boundaries.

Step 8: Adjust the image size using the plt.figure () method.

Step 9: Download the India shape file and import using the shapefile library into the program and show each state in India.

Step 10: Using the python 3.10.6 find the latitude and longitude for each state in India and stored in the particular list un1.

Step 11: Using the two lists un1 and un4 plot the scatter plot in India map.

Step 12: Display the Plotted map.

Step 13: Stop the program.

STEP BY STEP PROGRAM EXECUTION:

Given input

In [34]:
```
# import the necessary libraries
from mpl_toolkits.basemap import Basemap
import matplotlib.pyplot as plt
import shapefile as shp
import pandas as pd
```

In [36]:
```
# create the empty list and read data from csv file
un4=[]
df=pd.read_csv(r"C:\Users\OM MURUGA\Desktop\cybercrime.csv",encoding='ISO-8859-1')
```

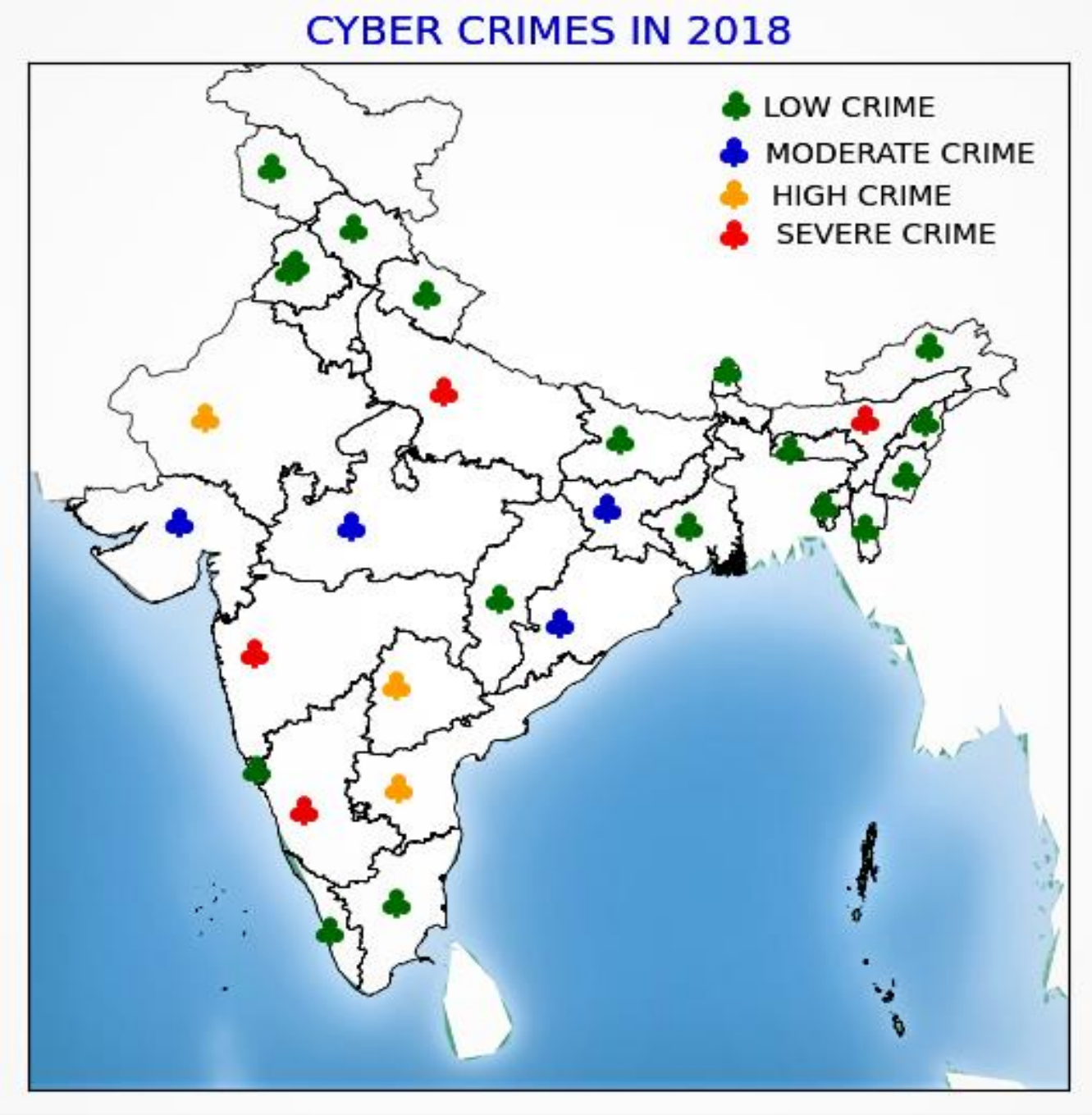
Given input

In [39]:
```
# remove the unwanted data using the drop() method and get the wanted data using the head() method and store in un4
df.drop(columns ='Unnamed: 2')
df.head(32)
new=df.drop(labels=['Unnamed: 2','Unnamed: 3','Unnamed: 5','Unnamed: 6','Unnamed: 7'],axis=1)
new.head(31)
b=new.drop(0)
dfi=b[["Unnamed: 1","Unnamed: 4"]]
a=dfi.head(31)
b=a['Unnamed: 4']
for i in range(2,31,1):
    un4.append(b[i])
```

Given input

In [40]:
```
# with the help of the basemap and matplotlib show the map plot the data
m=Basemap(llcrnrlon=67,llcrnrlat=5,urcrnrlon=99,urcrnrlat=37,projection='mill',resolution='c')
fig=plt.figure(figsize=(20,15))
m.readshapefile(r"C:\Users\OM MURUGA\Desktop\India_State_Shapefile\India_State_Shapefile\India_State_Boundary",name='states',draw
m.shadedrelief()
m.fillcontinents(color='white')
plt.title("CYBER CRIMES IN 2018",fontsize=14,color='blue')
un1=[[1.256e+06,1.104e+06],[3.073e+06,2.714e+06],[2.852e+06,2.449e+06],[2.014e+06,2.376e+06],[1.602e+06,1.795e+06],[7.71e+05,1.1
for i in range(0,29,1):
    if(un4[i]<=500):
        m.scatter(un1[i][0],un1[i][1],marker=r'$\clubsuit$',s=100,color='g',zorder=5)
    elif(un4[i]>500 and un4[i]<=1000):
        m.scatter(un1[i][0],un1[i][1],marker=r'$\clubsuit$',s=100,color='b',zorder=5)
    elif(un4[i]>1000 and un4[i]<=2000):
        m.scatter(un1[i][0],un1[i][1],marker=r'$\clubsuit$',s=100,color='orange',zorder=5)
    else:
        m.scatter(un1[i][0],un1[i][1],marker=r'$\clubsuit$',s=100,color='r',zorder=5)
m.scatter(2.411e+06,3.596e+06,marker=r'$\clubsuit$',s=100,color='g')
plt.text(2.514e+06,3.552e+06,'LOW CRIME')
m.scatter(2.404e+06,3.427e+06,marker=r'$\clubsuit$',s=100,color='b')
plt.text(2.521e+06,3.383e+06,'MODERATE CRIME')
m.scatter(2.404e+06,3.273e+06,marker=r'$\clubsuit$',s=100,color='orange')
plt.text(2.543e+06,3.229e+06,'HIGH CRIME')
m.scatter(2.404e+06,3.126e+06,marker=r'$\clubsuit$',s=100,color='r')
plt.text(2.558e+06,3.089e+06,'SEVERE CRIME')
plt.show()
```

Output for given input.



RESULT :

Thus, the program was verified and executed successfully with the help of the python libraries.