capstone-project-1

April 23, 2023

1 CAPSTONE PROJECT

[1]:

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
from matplotlib  import *
from scipy import stats
import plotly.express as px
import plotly.graph_objects as go
from plotly.subplots import make_subplots


import warnings
warnings.filterwarnings('ignore')
```

1.1 PHASE 1

The population of each state.

Literacy Rate in each state.

Area of each state

1.1.1 The population of each state

I have collected population data from GOV site (https://m.rbi.org.in/Scripts/PublicationsView.aspx?id=21391) i got data of 2001 and 2011.

NOTE : Here population data taken is in thousands (eg : Andaman & nicobar island 2001 population is 356thousand)

[2]:

```
df = pd.read_csv("DataTrained/census.csv")
```

[3]:

```
df.head()
```

[3]:

```
       State/Union Territory   2001    2011
0  Andaman & Nicobar Islands    356     381
1             Andhra Pradesh  76210   84581
```

```
2          Arunachal Pradesh   1098    1384
3                      Assam  26656   31206
4                      Bihar  82999  104099
```

[4]:
```
df['percentage_increase'] = ((df['2011'] - df['2001']) / df['2001']) * 100
```

[5]:
```
df.head()
```

[5]:
```
        State/Union Territory   2001    2011  percentage_increase
0  Andaman & Nicobar Islands    356     381             7.022472
1             Andhra Pradesh  76210   84581            10.984123
2          Arunachal Pradesh   1098    1384            26.047359
3                      Assam  26656   31206            17.069328
4                      Bihar  82999  104099            25.421993
```

Here i have calculated percentage increase in population for each state from 2001 to 2011 and based on percentage increased i have equally filled the data for years 2002 to 2010

[6]:
```
for i, row in df.iterrows():
    start_val = row['2001']
    end_val = row['2011']
    percent_increase = row['percentage_increase']
    for j in range(1, 10):
        new_val = start_val + ((j / 10) * (end_val - start_val))
        df.loc[i, f'part{j}'] = round(new_val, 2)
```

[7]:
```
df.head()
```

[7]:
```
        State/Union Territory   2001    2011  percentage_increase    part1  \
0  Andaman & Nicobar Islands    356     381             7.022472    358.5
1             Andhra Pradesh  76210   84581            10.984123  77047.1
2          Arunachal Pradesh   1098    1384            26.047359   1126.6
3                      Assam  26656   31206            17.069328  27111.0
4                      Bihar  82999  104099            25.421993  85109.0

     part2    part3    part4    part5    part6    part7    part8     part9
0    361.0    363.5    366.0    368.5    371.0    373.5    376.0     378.5
1  77884.2  78721.3  79558.4  80395.5  81232.6  82069.7  82906.8   83743.9
2   1155.2   1183.8   1212.4   1241.0   1269.6   1298.2   1326.8    1355.4
3  27566.0  28021.0  28476.0  28931.0  29386.0  29841.0  30296.0   30751.0
4  87219.0  89329.0  91439.0  93549.0  95659.0  97769.0  99879.0  101989.0
```

Renaming of years

[8]:
```
df = df.rename(columns={'part1': '2002', 'part2': '2003','part3': '2004',␣
 ↪'part4': '2005','part5': '2006', 'part6': '2007',
                       'part7': '2008', 'part8': '2009','part9':'2010'})
```

[9]:

```
df.head()
```

[9]:

```
        State/Union Territory   2001    2011  percentage_increase     2002  \
0  Andaman & Nicobar Islands    356     381             7.022472    358.5
1             Andhra Pradesh  76210   84581            10.984123  77047.1
2          Arunachal Pradesh   1098    1384            26.047359   1126.6
3                      Assam  26656   31206            17.069328  27111.0
4                      Bihar  82999  104099            25.421993  85109.0

      2003     2004     2005     2006     2007     2008     2009      2010
0    361.0    363.5    366.0    368.5    371.0    373.5    376.0     378.5
1  77884.2  78721.3  79558.4  80395.5  81232.6  82069.7  82906.8   83743.9
2   1155.2   1183.8   1212.4   1241.0   1269.6   1298.2   1326.8    1355.4
3  27566.0  28021.0  28476.0  28931.0  29386.0  29841.0  30296.0   30751.0
4  87219.0  89329.0  91439.0  93549.0  95659.0  97769.0  99879.0  101989.0
```

[10]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 35 entries, 0 to 34
Data columns (total 13 columns):
 #   Column                 Non-Null Count  Dtype
---  ------                 --------------  -----
 0   State/Union Territory  35 non-null     object
 1   2001                   35 non-null     int64
 2   2011                   35 non-null     int64
 3   percentage_increase    35 non-null     float64
 4   2002                   35 non-null     float64
 5   2003                   35 non-null     float64
 6   2004                   35 non-null     float64
 7   2005                   35 non-null     float64
 8   2006                   35 non-null     float64
 9   2007                   35 non-null     float64
 10  2008                   35 non-null     float64
 11  2009                   35 non-null     float64
 12  2010                   35 non-null     float64
dtypes: float64(10), int64(2), object(1)
memory usage: 3.7+ KB
```

deleted the percentage increase column and converted the float values to int

[11]:

```
df = df.drop(columns=['percentage_increase'])
columns_to_convert = ['2002', '2003',␣
 ↪'2004','2005','2006','2007','2008','2009','2010']
df[columns_to_convert] = df[columns_to_convert].astype(int)

df.head()
```

```
[11]:        State/Union Territory   2001    2011   2002   2003   2004   2005  \
      0  Andaman & Nicobar Islands    356     381    358    361    363    366
      1             Andhra Pradesh  76210   84581  77047  77884  78721  79558
      2          Arunachal Pradesh   1098    1384   1126   1155   1183   1212
      3                      Assam  26656   31206  27111  27566  28021  28476
      4                      Bihar  82999  104099  85109  87219  89329  91439

          2006   2007   2008   2009    2010
      0    368    371    373    376     378
      1  80395  81232  82069  82906   83743
      2   1241   1269   1298   1326    1355
      3  28931  29386  29841  30296   30751
      4  93549  95659  97769  99879  101989
```

```
[12]: df.shape
```

```
[12]: (35, 12)
```

1.1.2 Literacy Rate in each state

Got this data set from (https://www.kaggle.com/datasets/doncorleone92/govt-of-india-literacy-rate) .

In this data set we have literacy rate in 2001 and 2011 in total and rural and urban separately.

```
[13]: df_l = pd.read_csv("DataTrained/GOI.csv")
```

```
[14]: df_l.head()
```

```
[14]:   Category Country/ States/ Union Territories Name  \
      0  Country                                    INDIA
      1    State                           Andhra Pradesh
      2    State                        Arunachal Pradesh
      3    State                                    Assam
      4    State                                    Bihar

         Literacy Rate (Persons) - Total - 2001  \
      0                                    64.8
      1                                    60.5
      2                                    54.3
      3                                    63.3
      4                                    47.0

         Literacy Rate (Persons) - Total - 2011  \
      0                                    73.0
      1                                    67.0
      2                                    65.4
      3                                    72.2
```

```
4                                  61.8

   Literacy Rate (Persons) - Rural - 2001  \
0                                  58.7
1                                  54.5
2                                  47.8
3                                  59.7
4                                  43.9

   Literacy Rate (Persons) - Rural - 2011  \
0                                  67.8
1                                  60.4
2                                  59.9
3                                  69.3
4                                  59.8

   Literacy Rate (Persons) - Urban - 2001  \
0                                  79.9
1                                  76.1
2                                  78.3
3                                  85.3
4                                  71.9

   Literacy Rate (Persons) - Urban - 2011
0                                  84.1
1                                  80.1
2                                  82.9
3                                  88.5
4                                  76.9
```

[15]:

```
new_row_labels = {
    'A & N Islands': 'Andaman & Nicobar Islands',
    'D & N Haveli': 'Dadra & Nagar Haveli',
    'NCT of Delhi': 'Delhi'
}

df_l['Country/ States/ Union Territories Name'] = df_l['Country/ States/ Union␣
 ↪Territories Name'].replace(new_row_labels)
```

[16]:

```
df_l = df_l.rename(columns={'Literacy Rate (Persons) - Total - 2001':␣
 ↪'Literacy_Rate_2001',
                            'Literacy Rate (Persons) - Total - 2011':␣
 ↪'Literacy_Rate_2011'})
```

[17]:

```
df_l.tail()
```

[17]:

```
           Category Country/ States/ Union Territories Name  \
31  Union Territory                    Dadra & Nagar Haveli
32  Union Territory                             Daman & Diu
33  Union Territory                             Lakshadweep
34  Union Territory                                   Delhi
35  Union Territory                              Puducherry

    Literacy_Rate_2001  Literacy_Rate_2011  \
31                57.6                76.2
32                78.2                87.1
33                86.7                91.8
34                81.7                86.2
35                81.2                85.8

    Literacy Rate (Persons) - Rural - 2001  \
31                                    49.3
32                                    75.8
33                                    85.0
34                                    78.1
35                                    74.0

    Literacy Rate (Persons) - Rural - 2011  \
31                                    64.1
32                                    81.4
33                                    91.6
34                                    81.9
35                                    80.1

    Literacy Rate (Persons) - Urban - 2001  \
31                                    84.4
32                                    82.3
33                                    88.6
34                                    81.9
35                                    84.8

    Literacy Rate (Persons) - Urban - 2011
31                                    89.8
32                                    89.0
33                                    91.9
34                                    86.3
35                                    88.5
```

[18]:

```
df_l = df_l.rename(columns={'Country/ States/ Union Territories Name': 'State/
 ↪Union Territory'})
```

1.1.3 Merging the data frames of population and literacy rate

```
[19]: main_df = pd.merge(df,df_l, on='State/Union Territory')
```

```
[20]: main_df.head()
```

```
[20]:        State/Union Territory   2001    2011   2002   2003   2004   2005  \
      0  Andaman & Nicobar Islands    356     381    358    361    363    366
      1             Andhra Pradesh  76210   84581  77047  77884  78721  79558
      2          Arunachal Pradesh   1098    1384   1126   1155   1183   1212
      3                      Assam  26656   31206  27111  27566  28021  28476
      4                      Bihar  82999  104099  85109  87219  89329  91439

          2006   2007   2008   2009    2010         Category  Literacy_Rate_2001  \
      0    368    371    373    376     378  Union Territory                81.3
      1  80395  81232  82069  82906   83743            State                60.5
      2   1241   1269   1298   1326    1355            State                54.3
      3  28931  29386  29841  30296   30751            State                63.3
      4  93549  95659  97769  99879  101989            State                47.0

         Literacy_Rate_2011  Literacy Rate (Persons) - Rural - 2001  \
      0                86.6                                    78.7
      1                67.0                                    54.5
      2                65.4                                    47.8
      3                72.2                                    59.7
      4                61.8                                    43.9

         Literacy Rate (Persons) - Rural - 2011  \
      0                                    84.5
      1                                    60.4
      2                                    59.9
      3                                    69.3
      4                                    59.8

         Literacy Rate (Persons) - Urban - 2001  \
      0                                    86.6
      1                                    76.1
      2                                    78.3
      3                                    85.3
      4                                    71.9

         Literacy Rate (Persons) - Urban - 2011
      0                                    90.1
      1                                    80.1
      2                                    82.9
      3                                    88.5
      4                                    76.9
```

[21]:
```
main_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 35 entries, 0 to 34
Data columns (total 19 columns):
 #   Column                                Non-Null Count  Dtype
---  ------                                --------------  -----
 0   State/Union Territory                 35 non-null     object
 1   2001                                  35 non-null     int64
 2   2011                                  35 non-null     int64
 3   2002                                  35 non-null     int32
 4   2003                                  35 non-null     int32
 5   2004                                  35 non-null     int32
 6   2005                                  35 non-null     int32
 7   2006                                  35 non-null     int32
 8   2007                                  35 non-null     int32
 9   2008                                  35 non-null     int32
 10  2009                                  35 non-null     int32
 11  2010                                  35 non-null     int32
 12  Category                              35 non-null     object
 13  Literacy_Rate_2001                    35 non-null     float64
 14  Literacy_Rate_2011                    35 non-null     float64
 15  Literacy Rate (Persons) - Rural - 2001  35 non-null     float64
 16  Literacy Rate (Persons) - Rural - 2011  35 non-null     float64
 17  Literacy Rate (Persons) - Urban - 2001  35 non-null     float64
 18  Literacy Rate (Persons) - Urban - 2011  35 non-null     float64
dtypes: float64(6), int32(9), int64(2), object(2)
memory usage: 4.2+ KB
```

1.1.4 Area of each state

I have collected data from https://www.indiastat.com/specimen-tables/geographical-data and prepared a csv file.

[22]:
```
df_a = pd.read_csv("DataTrained/area.csv")
```

[23]:
```
df_a.head()
```

[23]:
```
       State/Union Territory  Area in sq.km
0  Andaman & Nicobar Islands           8249
1             Andhra Pradesh         275045
2          Arunachal Pradesh          83743
3                      Assam          78438
4                      Bihar          94163
```

1.1.5 Merged area of each state with main data frame.

[24]:

```
main_df = pd.merge(main_df, df_a, on='State/Union Territory')
```

[25]:

```
main_df.head()
```

[25]:

```
       State/Union Territory   2001    2011   2002   2003   2004   2005  \
0  Andaman & Nicobar Islands    356     381    358    361    363    366
1             Andhra Pradesh  76210   84581  77047  77884  78721  79558
2          Arunachal Pradesh   1098    1384   1126   1155   1183   1212
3                      Assam  26656   31206  27111  27566  28021  28476
4                      Bihar  82999  104099  85109  87219  89329  91439

    2006   2007   2008   2009    2010         Category  Literacy_Rate_2001  \
0    368    371    373    376     378  Union Territory                81.3
1  80395  81232  82069  82906   83743            State                60.5
2   1241   1269   1298   1326    1355            State                54.3
3  28931  29386  29841  30296   30751            State                63.3
4  93549  95659  97769  99879  101989            State                47.0

   Literacy_Rate_2011  Literacy Rate (Persons) - Rural - 2001  \
0                86.6                                    78.7
1                67.0                                    54.5
2                65.4                                    47.8
3                72.2                                    59.7
4                61.8                                    43.9

   Literacy Rate (Persons) - Rural - 2011  \
0                                    84.5
1                                    60.4
2                                    59.9
3                                    69.3
4                                    59.8

   Literacy Rate (Persons) - Urban - 2001  \
0                                    86.6
1                                    76.1
2                                    78.3
3                                    85.3
4                                    71.9

   Literacy Rate (Persons) - Urban - 2011  Area in sq.km
0                                    90.1           8249
1                                    80.1         275045
2                                    82.9          83743
3                                    88.5          78438
4                                    76.9          94163
```

Using of Melt function.

This function is useful to massage a DataFrame into a format where one or more columns are identifier variables (id_vars), while all other columns, considered measured variables (value_vars), are “unpivoted” to the row axis, leaving just two non-identifier columns, ‘variable’ and ‘value’.(https://pandas.pydata.org/docs/reference/api/pandas.melt.html)

Used this function to convert df which i created into similar format of other df ehich were given.

[26]:
```
main_df_melt = main_df.melt(id_vars=['State/Union␣
 ↪Territory','Category','Literacy_Rate_2001'
                                    ,'Literacy_Rate_2011','Area in sq.
 ↪km','Literacy Rate (Persons) - Rural - 2001',
                                    'Literacy Rate (Persons) - Rural -␣
 ↪2011','Literacy Rate (Persons) - Urban - 2001',
                                    'Literacy Rate (Persons) - Urban - 2011'],␣
 ↪var_name='year', value_name='population')
```

[27]:
```
main_df_melt.head()
```

[27]:
```
       State/Union Territory         Category  Literacy_Rate_2001  \
0  Andaman & Nicobar Islands  Union Territory                81.3
1             Andhra Pradesh            State                60.5
2          Arunachal Pradesh            State                54.3
3                      Assam            State                63.3
4                      Bihar            State                47.0

   Literacy_Rate_2011  Area in sq.km  Literacy Rate (Persons) - Rural - 2001  \
0                86.6           8249                                    78.7
1                67.0         275045                                    54.5
2                65.4          83743                                    47.8
3                72.2          78438                                    59.7
4                61.8          94163                                    43.9

   Literacy Rate (Persons) - Rural - 2011  \
0                                    84.5
1                                    60.4
2                                    59.9
3                                    69.3
4                                    59.8

   Literacy Rate (Persons) - Urban - 2001  \
0                                    86.6
1                                    76.1
2                                    78.3
3                                    85.3
4                                    71.9
```

```
   Literacy Rate (Persons) - Urban - 2011  year  population
0                                    90.1  2001         356
1                                    80.1  2001       76210
2                                    82.9  2001        1098
3                                    88.5  2001       26656
4                                    76.9  2001       82999
```

[28]:

```
main_df_melt
```

[28]:

```
          State/Union Territory         Category  Literacy_Rate_2001  \
0    Andaman & Nicobar Islands  Union Territory                81.3
1               Andhra Pradesh            State                60.5
2            Arunachal Pradesh            State                54.3
3                        Assam            State                63.3
4                        Bihar            State                47.0
..                         …              …                 …
380                 Tamil Nadu            State                73.5
381                    Tripura            State                73.2
382              Uttar Pradesh            State                56.3
383                Uttarakhand            State                71.6
384                West Bengal            State                68.6

     Literacy_Rate_2011  Area in sq.km  \
0                  86.6           8249
1                  67.0         275045
2                  65.4          83743
3                  72.2          78438
4                  61.8          94163
..                  …             …
380                80.1         130060
381                87.2          10486
382                67.7         240928
383                78.8          53483
384                76.3          88752

     Literacy Rate (Persons) - Rural - 2001  \
0                                      78.7
1                                      54.5
2                                      47.8
3                                      59.7
4                                      43.9
..                                      …
380                                    66.2
381                                    69.7
382                                    52.5
383                                    68.1
384                                    63.4
```

```
     Literacy Rate (Persons) - Rural - 2011  \
0                                  84.5
1                                  60.4
2                                  59.9
3                                  69.3
4                                  59.8
..                                  …
380                                73.5
381                                84.9
382                                65.5
383                                76.3
384                                72.1

     Literacy Rate (Persons) - Urban - 2001  \
0                                  86.6
1                                  76.1
2                                  78.3
3                                  85.3
4                                  71.9
..                                  …
380                                82.5
381                                89.2
382                                69.8
383                                81.4
384                                81.2

     Literacy Rate (Persons) - Urban - 2011  year  population
0                                  90.1  2001         356
1                                  80.1  2001       76210
2                                  82.9  2001        1098
3                                  88.5  2001       26656
4                                  76.9  2001       82999
..                                  …    …          …
380                                87.0  2010       71172
381                                93.5  2010        3626
382                                75.1  2010      196450
383                                84.5  2010        9926
384                                84.8  2010       90166

[385 rows x 11 columns]
```

[29]:
```
main_df_melt['year'] = main_df_melt['year'].astype(str)
```

[30]:
```
display(main_df_melt.isnull().any())
```

```
State/Union Territory                     False
```

```
Category                                  False
Literacy_Rate_2001                        False
Literacy_Rate_2011                        False
Area in sq.km                             False
Literacy Rate (Persons) - Rural - 2001    False
Literacy Rate (Persons) - Rural - 2011    False
Literacy Rate (Persons) - Urban - 2001    False
Literacy Rate (Persons) - Urban - 2011    False
year                                      False
population                                False
dtype: bool
```

```
[31]: main_df_melt['State/Union Territory'] = main_df_melt['State/Union Territory'].
       ↪str.upper()
```

1.2 Phase - 2

```
[32]: crime = pd.read_csv("DataTrained/crime.csv")
```

```
[33]: crime.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 385 entries, 0 to 384
Data columns (total 34 columns):
 #   Column                              Non-Null Count  Dtype
---  ------                              --------------  -----
 0   STATE/UT                            385 non-null    object
 1   YEAR                                385 non-null    int64
 2   RESIDENTIAL PREMISES - Dacoity      385 non-null    int64
 3   RESIDENTIAL PREMISES - Robbery      385 non-null    int64
 4   RESIDENTIAL PREMISES - Burglary     385 non-null    int64
 5   RESIDENTIAL PREMISES - Theft        385 non-null    int64
 6   HIGHWAYS - Dacoity                  385 non-null    int64
 7   HIGHWAYS - Robbery                  385 non-null    int64
 8   HIGHWAYS - Burglary                 385 non-null    int64
 9   HIGHWAYS - Theft                    385 non-null    int64
 10  RIVER and SEA - Dacoity             385 non-null    int64
 11  RIVER and SEA - Robbery             385 non-null    int64
 12  RIVER and SEA - Burglary            385 non-null    int64
 13  RIVER and SEA - Theft               385 non-null    int64
 14  RAILWAYS - Dacoity                  385 non-null    int64
 15  RAILWAYS - Robbery                  385 non-null    int64
 16  RAILWAYS - Burglary                 385 non-null    int64
 17  RAILWAYS - Theft                    385 non-null    int64
 18  BANKS - Dacoity                     385 non-null    int64
 19  BANKS - Robbery                     385 non-null    int64
 20  BANKS - Burglary                    385 non-null    int64
 21  BANKS - Theft                          385 non-null    int64
 22  COMMERCIAL ESTABLISHMENTS - Dacoity    385 non-null    int64
 23  COMMERCIAL ESTABLISHMENTS - Robbery    385 non-null    int64
 24  COMMERCIAL ESTABLISHMENTS - Burglary   385 non-null    int64
 25  COMMERCIAL ESTABLISHMENTS - Theft      385 non-null    int64
 26  OTHER PLACES - Dacoity                 385 non-null    int64
 27  OTHER PLACES - Robbery                 385 non-null    int64
 28  OTHER PLACES - Burglary                385 non-null    int64
 29  OTHER PLACES - Theft                   385 non-null    int64
 30  TOTAL - Dacoity                        385 non-null    int64
 31  TOTAL - Robbery                        385 non-null    int64
 32  TOTAL - Burglary                       385 non-null    int64
 33  TOTAL - Theft                          385 non-null    int64
dtypes: int64(33), object(1)
memory usage: 102.4+ KB
```

[34]:
```
crime['YEAR'] = crime['YEAR'].astype(str)
```

[35]:
```
crime = crime.rename(columns={'STATE/UT': 'State/Union Territory'})

crime = crime.rename(columns={'YEAR': 'year'})
```

[36]:
```
new_row_labels = {
    'A & N ISLANDS': 'ANDAMAN & NICOBAR ISLANDS',
    'D & N HAVELI': 'DADRA & NAGAR HAVELI',

}

crime['State/Union Territory'] = crime['State/Union Territory'].
 ↪replace(new_row_labels)
```

[37]:
```
crime
```

[37]:
```
    State/Union Territory  year  RESIDENTIAL PREMISES - Dacoity  \
0          ANDHRA PRADESH  2001                             100
1       ARUNACHAL PRADESH  2001                               9
2                   ASSAM  2001                             381
3                   BIHAR  2001                             818
4            CHHATTISGARH  2001                              54
..                    …     …                             …
380  DADRA & NAGAR HAVELI  2011                               2
381           DAMAN & DIU  2011                               0
382                 DELHI  2011                               9
383           LAKSHADWEEP  2011                               0
384            PUDUCHERRY  2011                               3

     RESIDENTIAL PREMISES - Robbery  RESIDENTIAL PREMISES - Burglary  \
0                               177                             5158
1                                26                               99
2                               191                             1695
3                               326                             2486
4                                42                             3336
..                              …                              …
380                               1                                6
381                               0                               12
382                              85                              944
383                               0                                0
384                               2                               46

     RESIDENTIAL PREMISES - Theft  HIGHWAYS - Dacoity  HIGHWAYS - Robbery  \
0                            4257                  57                 172
1                             131                   0                   0
2                            2901                  46                 136
3                            4741                 162                 826
4                            1417                  10                  38
..                            …                 …                 …
380                            45                   0                   0
381                            14                   0                   0
382                          6018                   2                  26
383                             4                   0                   0
384                            53                   0                   0

     HIGHWAYS - Burglary  HIGHWAYS - Theft  …  \
0                     31                74  …
1                      0                 8  …
2                      7                87  …
3                      0               257  …
4                     12                72  …
..                   …               …  …
380                    0                 0  …
381                    0                 1  …
382                    0               169  …
383                    0                 0  …
384                    0                 0  …

     COMMERCIAL ESTABLISHMENTS - Burglary  COMMERCIAL ESTABLISHMENTS - Theft  \
0                                    1041                               2502
1                                      84                                 54
2                                     442                                967
3                                     231                                686
4                                     370                                299
..                                    …                                …
380                                    16                                  8
381                                    11                                  2
```

```
382                                    189                               2011
383                                      0                                  0
384                                     15                                 36

     OTHER PLACES - Dacoity  OTHER PLACES - Robbery  OTHER PLACES - Burglary  \
0                        37                     232                      862
1                         8                      40                       65
2                        77                     261                      271
3                       210                     880                      505
4                        15                     239                      420
..                      …                     …                      …
380                       0                       0                        0
381                       0                       5                        8
382                      12                     397                      284
383                       0                       0                        0
384                       1                       6                        4

     OTHER PLACES - Theft  TOTAL - Dacoity  TOTAL - Robbery  TOTAL - Burglary  \
0                    8849              214              629              7220
1                     249               22               84               248
2                    1342              532              687              2423
3                    2582             1291             2203              3233
4                    2835               87              338              4144
..                    …              …              …               …
380                    10                7                2                22
381                    30                4                6                31
382                 14618               33              562              1419
383                     0                0                0                 0
384                   618                5               11                65

     TOTAL - Theft
0            16751
1              443
2             5367
3             9701
4             4812
..             …
380             69
381             47
382          22899
383              4
384            707

[385 rows x 34 columns]
```

Merging the main dataset which was created in Phase 1 with crime dataset

[38]:

```
crime = pd.merge(crime, main_df_melt, on=['State/Union Territory','year'])
```

[39]:

```
crime
```

[39]:

```
    State/Union Territory  year  RESIDENTIAL PREMISES - Dacoity  \
0          ANDHRA PRADESH  2001                             100
1       ARUNACHAL PRADESH  2001                               9
2                   ASSAM  2001                             381
3                   BIHAR  2001                             818
4            CHHATTISGARH  2001                              54
..                    …    …                            …
380  DADRA & NAGAR HAVELI  2011                               2
381           DAMAN & DIU  2011                               0
382                 DELHI  2011                               9
383           LAKSHADWEEP  2011                               0
384            PUDUCHERRY  2011                               3

     RESIDENTIAL PREMISES - Robbery  RESIDENTIAL PREMISES - Burglary  \
0                               177                             5158
1                                26                               99
2                               191                             1695
3                               326                             2486
4                                42                             3336
..                              …                              …
380                               1                                6
381                               0                               12
382                              85                              944
383                               0                                0
384                               2                               46

     RESIDENTIAL PREMISES - Theft  HIGHWAYS - Dacoity  HIGHWAYS - Robbery  \
0                            4257                  57                 172
1                             131                   0                   0
2                            2901                  46                 136
3                            4741                 162                 826
4                            1417                  10                  38
..                            …                 …                 …
380                            45                   0                   0
381                            14                   0                   0
382                          6018                   2                  26
383                             4                   0                   0
384                            53                   0                   0

     HIGHWAYS - Burglary  HIGHWAYS - Theft  …  TOTAL - Theft  \
0                     31                74  …          16751
1                      0                 8  …            443
2                      7                87  …           5367
```

```
3                      0               257  …           9701
4                     12                72  …           4812
..                   …               …  …            …
380                    0                 0  …             69
381                    0                 1  …             47
382                    0               169  …          22899
383                    0                 0  …              4
384                    0                 0  …            707

            Category  Literacy_Rate_2001  Literacy_Rate_2011  Area in sq.km  \
0              State                60.5                67.0         275045
1              State                54.3                65.4          83743
2              State                63.3                72.2          78438
3              State                47.0                61.8          94163
4              State                64.7                70.3         135192
..               …                 …                 …            …
380  Union Territory                57.6                76.2            491
381  Union Territory                78.2                87.1            111
382  Union Territory                81.7                86.2           1483
383  Union Territory                86.7                91.8             30
384  Union Territory                81.2                85.8            490

     Literacy Rate (Persons) - Rural - 2001  \
0                                      54.5
1                                      47.8
2                                      59.7
3                                      43.9
4                                      60.5
..                                      …
380                                    49.3
381                                    75.8
382                                    78.1
383                                    85.0
384                                    74.0

     Literacy Rate (Persons) - Rural - 2011  \
0                                      60.4
1                                      59.9
2                                      69.3
3                                      59.8
4                                      66.0
..                                      …
380                                    64.1
381                                    81.4
382                                    81.9
383                                    91.6
384                                    80.1
```

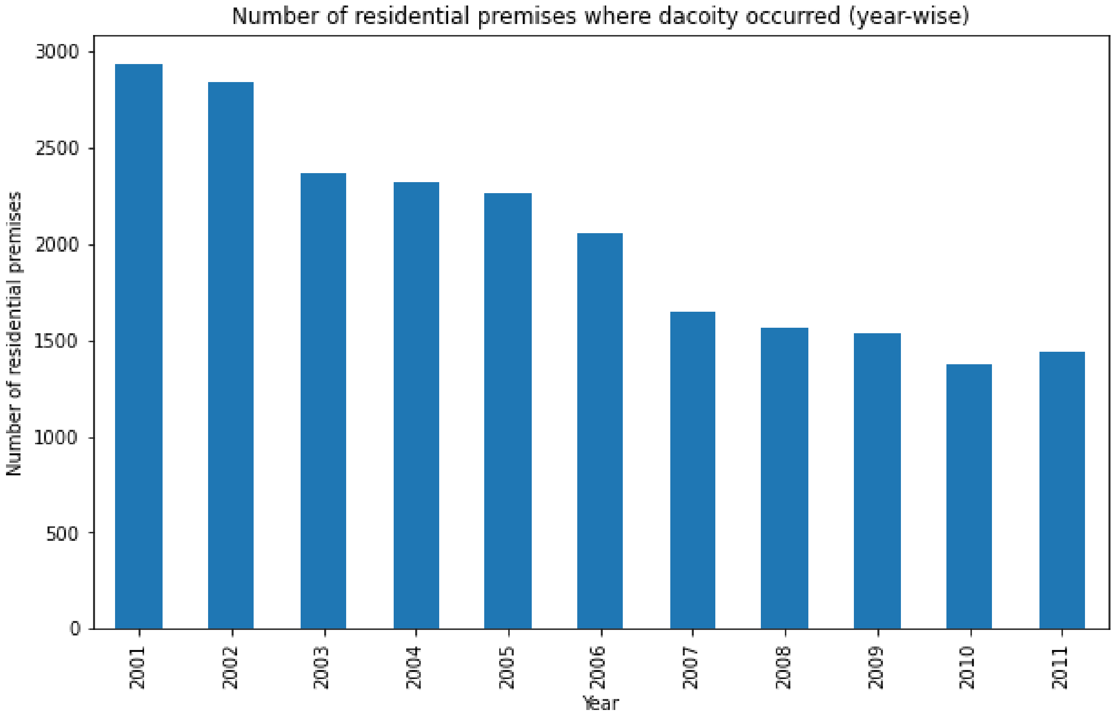
```
     Literacy Rate (Persons) - Urban - 2001  \
0                                      76.1
1                                      78.3
2                                      85.3
3                                      71.9
4                                      80.6
..                                      …
380                                    84.4
381                                    82.3
382                                    81.9
383                                    88.6
384                                    84.8

     Literacy Rate (Persons) - Urban - 2011  population
0                                      80.1       76210
1                                      82.9        1098
2                                      88.5       26656
3                                      76.9       82999
4                                      84.0       20834
..                                      …          …
380                                    89.8         344
381                                    89.0         243
382                                    86.3       16788
383                                    91.9          64
384                                    88.5        1248

[385 rows x 43 columns]
```

[40]:
```
crime['Total - Per. Change'] = (crime.loc[:,'Literacy_Rate_2011'] -
                crime.loc[:,'Literacy_Rate_2001'])/crime.loc[:
 ↪,'Literacy_Rate_2001']
crime['Rural - Per. Change'] = (crime.loc[:,'Literacy Rate (Persons) - Rural -␣
 ↪2011'] -
                crime.loc[:,'Literacy Rate (Persons) - Rural - 2001'])/crime.
 ↪loc[:,'Literacy_Rate_2001']
crime['Urban - Per. Change'] = (crime.loc[:,'Literacy Rate (Persons) - Urban -␣
 ↪2011'] -
                crime.loc[:,'Literacy Rate (Persons) - Urban - 2001'])/crime.
 ↪loc[:,'Literacy_Rate_2001']
```

[41]:
```
crime.sort_values(by='Literacy_Rate_2001', inplace=True)

fig = go.Figure(data = [
    go.Scatter(name='2001', x=crime['State/Union Territory'],␣
 ↪y=crime['Literacy_Rate_2001'], mode='markers'),
    go.Scatter(name='2011', x=crime['State/Union Territory'],␣
  ↪y=crime['Literacy_Rate_2011'], mode='markers')
])

fig.update_layout(barmode='group', title = 'Total Literacy Rate Across Nation :
  ↪')
fig.show()
```

[42]:

```
yearly_data = crime.groupby('year')['RESIDENTIAL PREMISES - Dacoity'].sum()


yearly_data.plot(kind='bar', figsize=(10, 6))
plt.title('Number of residential premises where dacoity occurred (year-wise)')
plt.xlabel('Year')
plt.ylabel('Number of residential premises')
plt.show()
```

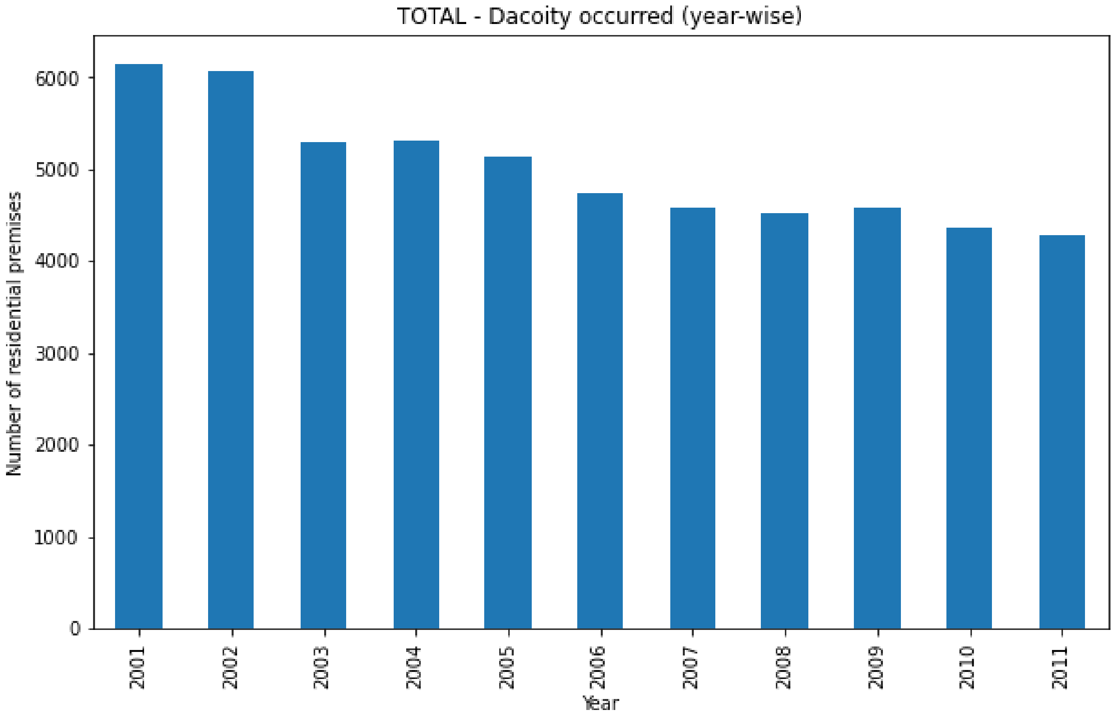


From above graph we can see that from 2001 the residential premises where dacoity occurred decreased geadually

[43]:

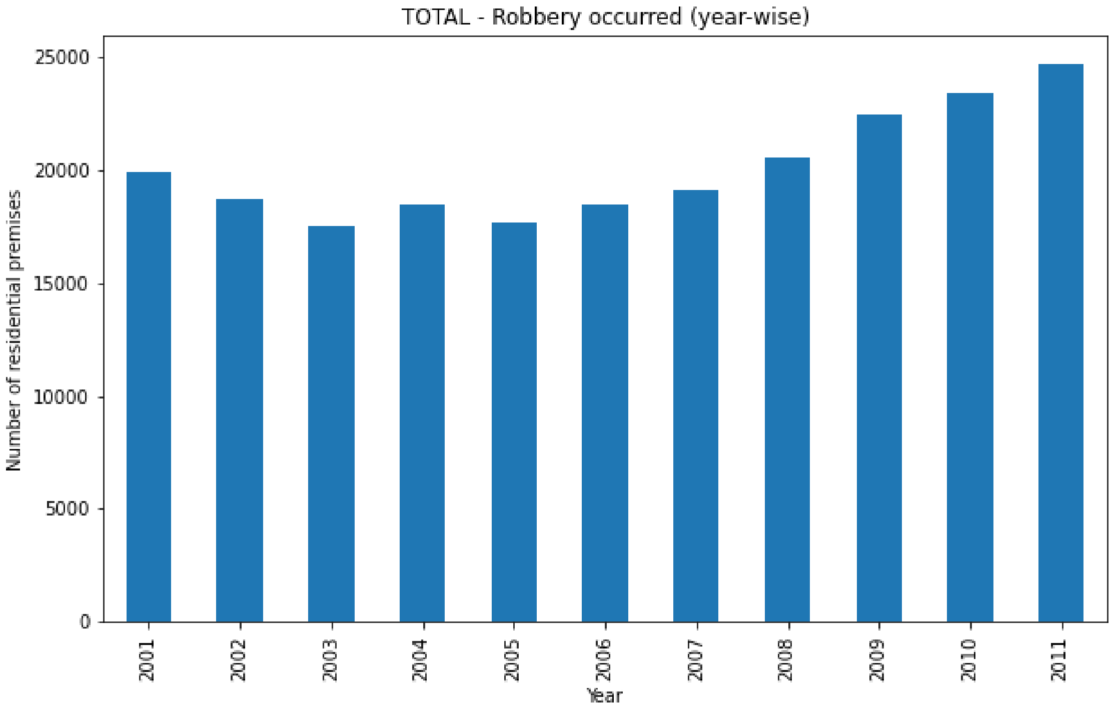
```
yearly_data = crime.groupby('year')['TOTAL - Dacoity'].sum()


yearly_data.plot(kind='bar', figsize=(10, 6))
```

```
plt.title('TOTAL - Dacoity occurred (year-wise)')
plt.xlabel('Year')
plt.ylabel('Number of residential premises')
plt.show()
```



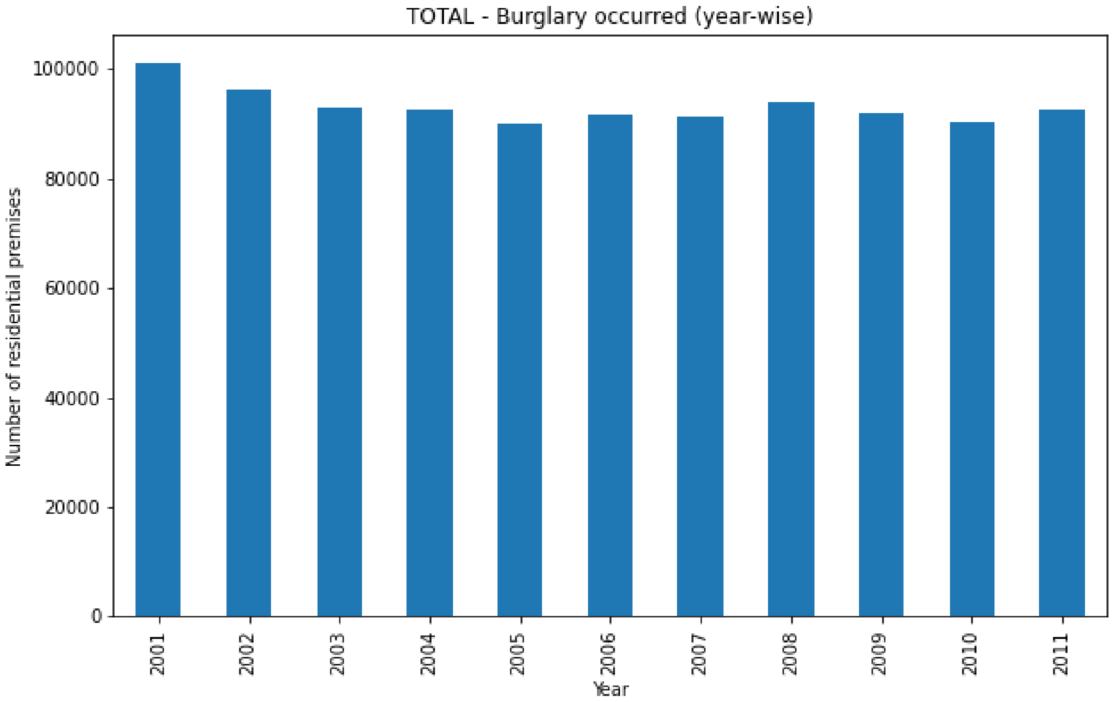
[44]:

```
yearly_data = crime.groupby('year')['TOTAL - Robbery'].sum()


yearly_data.plot(kind='bar', figsize=(10, 6))
plt.title('TOTAL - Robbery occurred (year-wise)')
plt.xlabel('Year')
plt.ylabel('Number of residential premises')
plt.show()
```

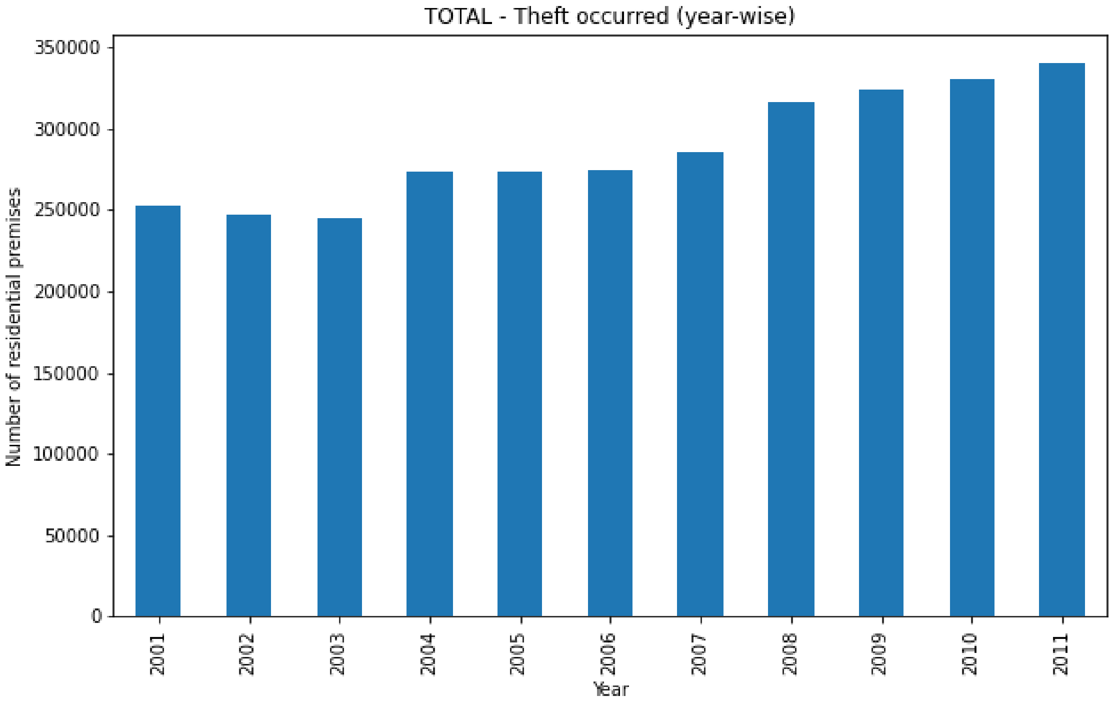


[45]:

```
yearly_data = crime.groupby('year')['TOTAL - Burglary'].sum()


yearly_data.plot(kind='bar', figsize=(10, 6))
plt.title('TOTAL - Burglary occurred (year-wise)')
plt.xlabel('Year')
plt.ylabel('Number of residential premises')
plt.show()
```

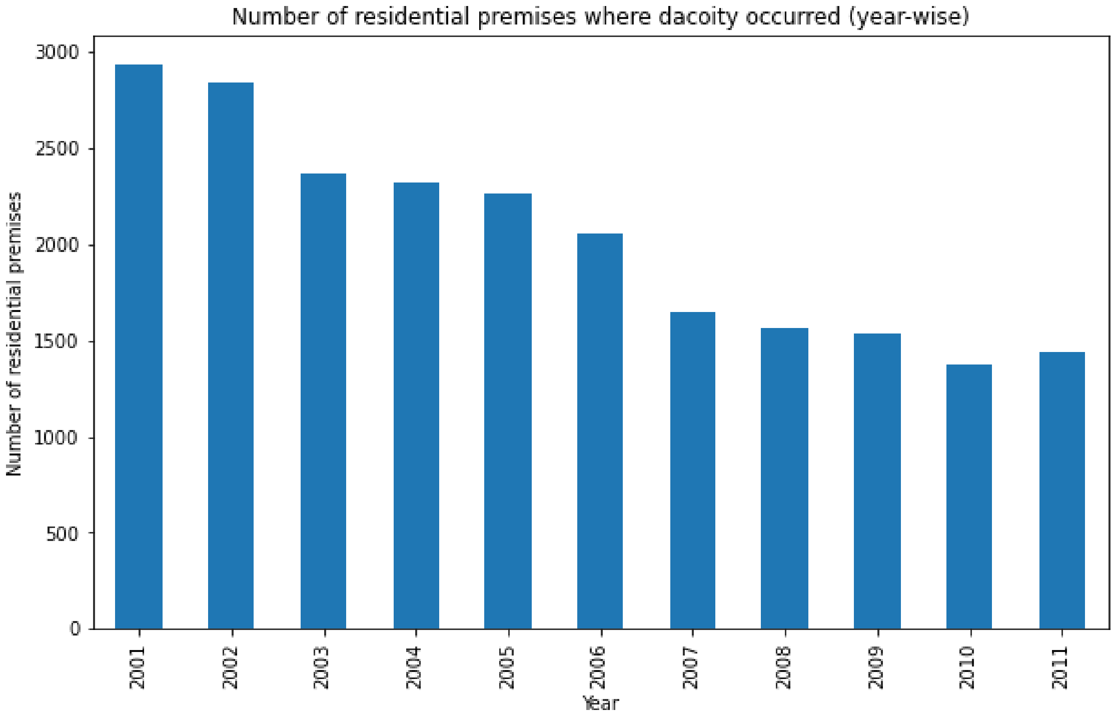


[46]:
```
yearly_data = crime.groupby('year')['TOTAL - Theft'].sum()


yearly_data.plot(kind='bar', figsize=(10, 6))
plt.title('TOTAL - Theft occurred (year-wise)')
plt.xlabel('Year')
plt.ylabel('Number of residential premises')
plt.show()
```



[47]:

```
yearly_data = crime.groupby('year')['RESIDENTIAL PREMISES - Dacoity'].sum()


yearly_data.plot(kind='bar', figsize=(10, 6))
plt.title('Number of residential premises where dacoity occurred (year-wise)')
plt.xlabel('Year')
plt.ylabel('Number of residential premises')
plt.show()
```



[48]:

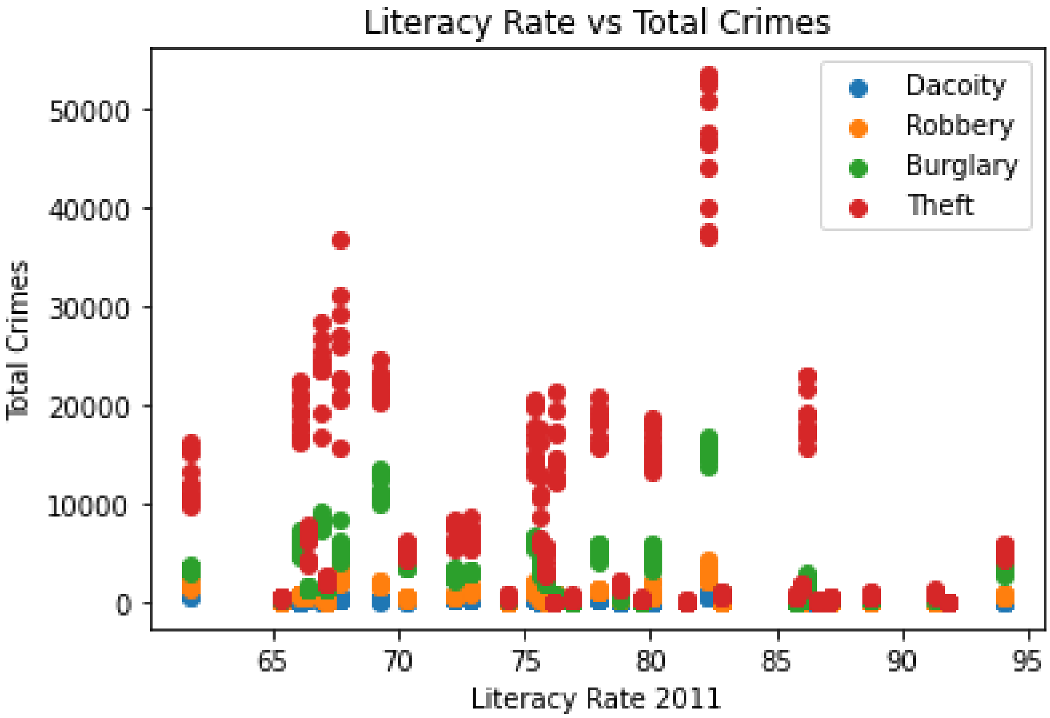
```
req_columns = ['Literacy_Rate_2011', 'TOTAL - Dacoity', 'TOTAL - Robbery',␣
 ↪'TOTAL - Burglary', 'TOTAL - Theft']
crime_1 = crime[req_columns]

crime_1
```

[48]:

```
     Literacy_Rate_2011  TOTAL - Dacoity  TOTAL - Robbery  TOTAL - Burglary  \
107                61.8             1319             2986              3175
38                 61.8             1289             2288              3188
3                  61.8             1291             2203              3233
212                61.8              686             1787              3259
247                61.8              686             1592              3414
..                  …                …                …                 …
221                94.0              121              869              4100
116                94.0              129              639              4580
12                 94.0              176              517              4474
256                94.0               91              816              3882
292                94.0              112              830              3554

     TOTAL - Theft
107          11113
38           10145
3             9701
212          11795
```

```
247          13206
..             …
221           5609
116           5240
12            5441
256           5818
292           5564

[385 rows x 5 columns]
```

```
[49]: # create scatter plot of literacy rate vs total crimes
fig, ax = plt.subplots()
ax.scatter(crime_1['Literacy_Rate_2011'], crime_1['TOTAL - Dacoity'],␣
 ↪label='Dacoity')
ax.scatter(crime_1['Literacy_Rate_2011'], crime_1['TOTAL - Robbery'],␣
 ↪label='Robbery')
ax.scatter(crime_1['Literacy_Rate_2011'], crime_1['TOTAL - Burglary'],␣
 ↪label='Burglary')
ax.scatter(crime_1['Literacy_Rate_2011'], crime_1['TOTAL - Theft'],␣
 ↪label='Theft')
ax.set_xlabel('Literacy Rate 2011')
ax.set_ylabel('Total Crimes')
ax.set_title('Literacy Rate vs Total Crimes')
ax.legend()
plt.show()
```



[50]:

```
avg_lit_rate = crime_1['Literacy_Rate_2011'].mean()
print("Average literacy rate across all states/union territories in 2011 ",␣
 ↪avg_lit_rate)

total_crimes = crime_1.iloc[:, 4:].sum().sum()
print("Total number of crimes across all states/union territories and all␣
 ↪categories: ", total_crimes)
```

```
Average literacy rate across all states/union territories in 2011
77.84545454545425
Total number of crimes across all states/union territories and all categories:
3162902
```

[51]:

```
# creating subset for year 2001 with total crimes.
crime_sub_2001 = crime.loc[crime['year'] == '2001', ['State/Union Territory',␣
 ↪'year', 'TOTAL - Dacoity', 'TOTAL - Robbery',
                                                    'TOTAL - Burglary','TOTAL - Theft',␣
 ↪'Category', 'Literacy_Rate_2001']]
```

[52]:

```
crime_sub_2001
```

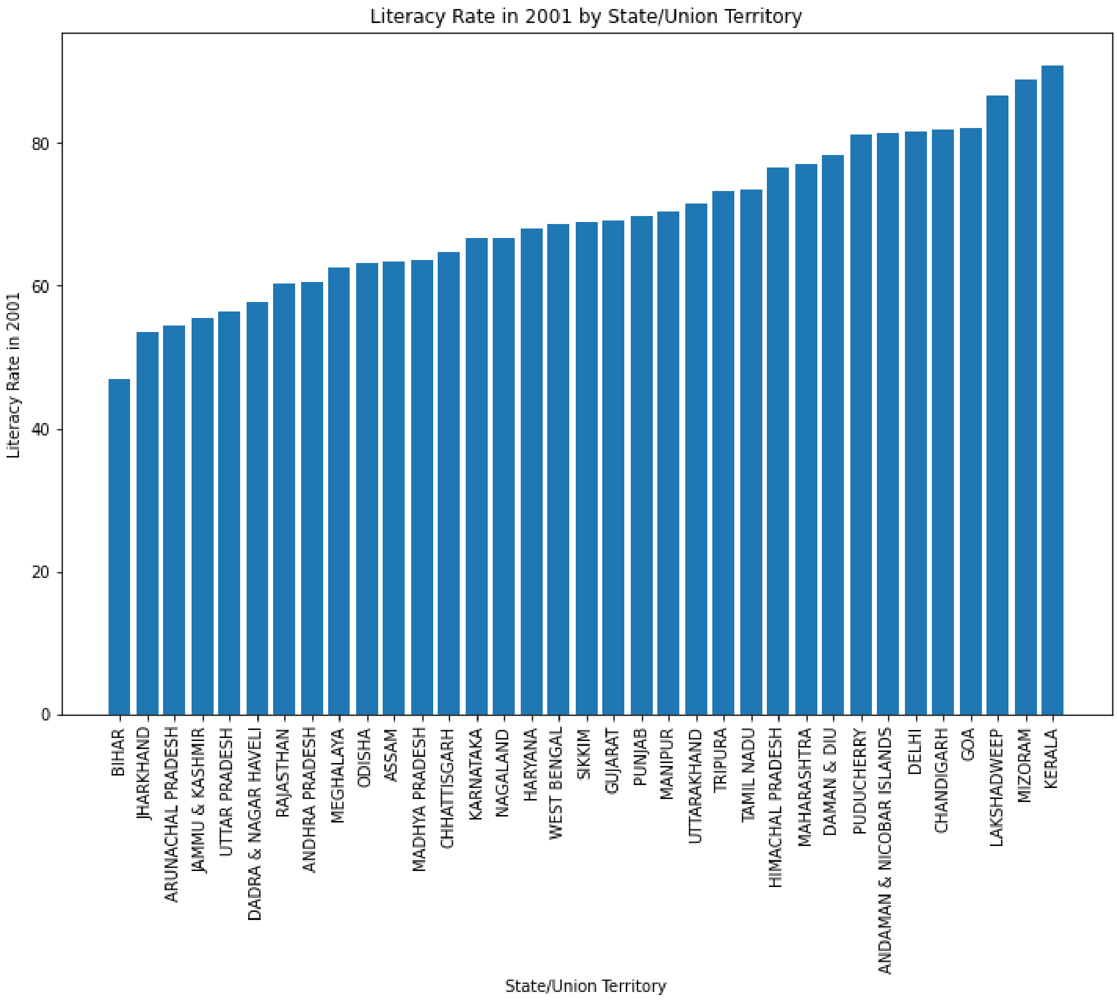
[52]:

```
        State/Union Territory  year  TOTAL - Dacoity  TOTAL - Robbery  \
3                       BIHAR  2001             1291             2203
10                  JHARKHAND  2001              636              647
1           ARUNACHAL PRADESH  2001               22               84
9             JAMMU & KASHMIR  2001               24              161
25              UTTAR PRADESH  2001              905             3825
30       DADRA & NAGAR HAVELI  2001                0                2
21                  RAJASTHAN  2001               60              889
0              ANDHRA PRADESH  2001              214              629
16                  MEGHALAYA  2001               97              125
19                     ODISHA  2001              199              958
2                       ASSAM  2001              532              687
13             MADHYA PRADESH  2001              166             1764
4                CHHATTISGARH  2001               87              338
11                  KARNATAKA  2001              178              847
18                   NAGALAND  2001               11              129
7                     HARYANA  2001               77              397
27                WEST BENGAL  2001              274              660
22                     SIKKIM  2001                0                3
6                     GUJARAT  2001              327              991
20                     PUNJAB  2001               45              131
15                    MANIPUR  2001               20               19
26                UTTARAKHAND  2001               32              191
24                    TRIPURA  2001               26               63
```

```
23                 TAMIL NADU  2001               158              672
8            HIMACHAL PRADESH  2001                 4               28
14                MAHARASHTRA  2001               529             2239
31                DAMAN & DIU  2001                 0                0
34                 PUDUCHERRY  2001                 1                4
28  ANDAMAN & NICOBAR ISLANDS  2001                 0                4
32                      DELHI  2001                48              624
29                 CHANDIGARH  2001                 5               22
5                         GOA  2001                 7               25
33                LAKSHADWEEP  2001                 0                0
17                    MIZORAM  2001                 3               23
12                     KERALA  2001               176              517

    TOTAL - Burglary  TOTAL - Theft          Category  Literacy_Rate_2001
3               3233           9701             State                47.0
10              1266           3827             State                53.6
1                248            443             State                54.3
9               1345           1919             State                55.5
25              8411          27011             State                56.3
30                34             45   Union Territory                57.6
21              7284          16939             State                60.4
0               7220          16751             State                60.5
16               170            271             State                62.6
19              3093           5622             State                63.1
2               2423           5367             State                63.3
13             13549          20263             State                63.7
4               4144           4812             State                64.7
11              6394          12868             State                66.6
18               163            258             State                66.6
7               3109           6117             State                67.9
27               426          14245             State                68.6
22                76             74             State                68.8
6               5142          15834             State                69.1
20              1902           3023             State                69.7
15                75            217             State                70.5
26               533           1419             State                71.6
24               198            259             State                73.2
23              5965          16709             State                73.5
8                844            600             State                76.5
14             15073          39866             State                76.9
31                43             40   Union Territory                78.2
34               111            528   Union Territory                81.2
28                64             65   Union Territory                81.3
32              3029          19276   Union Territory                81.7
29               364           1529   Union Territory                81.9
5                359            576             State                82.0
33                 1             10   Union Territory                86.7
```

```
17                417           878           State               88.8
12               4474          5441           State               90.9
```

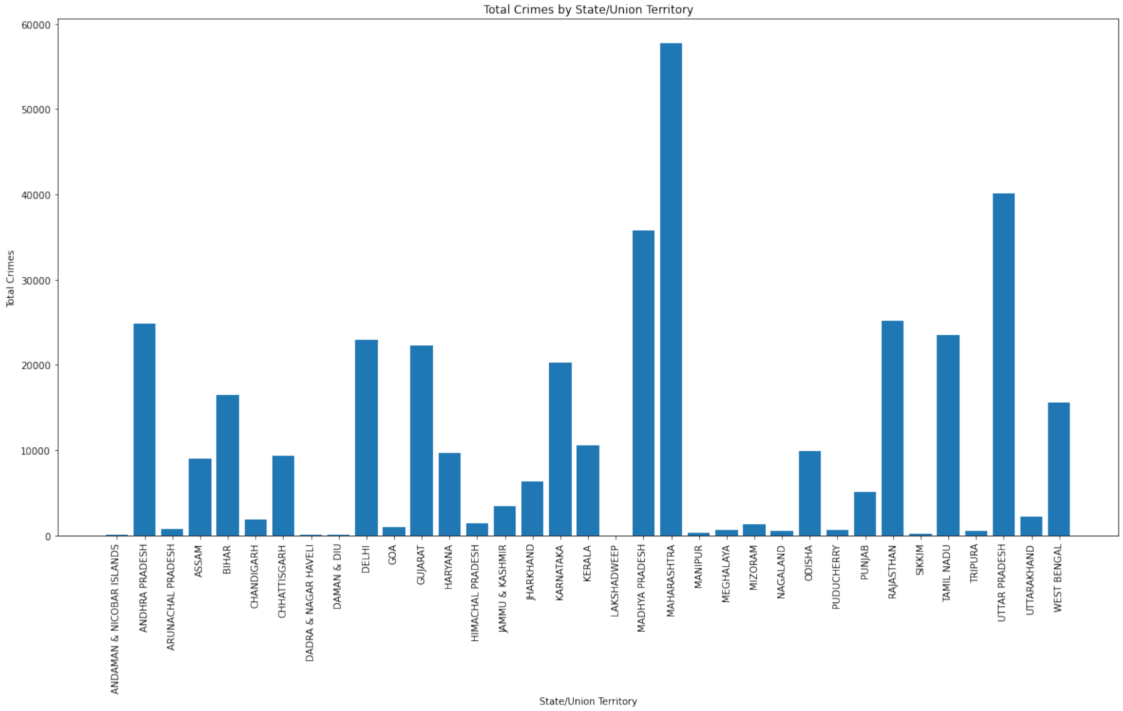
[53]:
```
plt.figure(figsize=(12,8))
plt.bar(crime_sub_2001['State/Union Territory'],␣
  ↪crime_sub_2001['Literacy_Rate_2001'])
plt.xticks(rotation=90)
plt.xlabel('State/Union Territory')
plt.ylabel('Literacy Rate in 2001')
plt.title('Literacy Rate in 2001 by State/Union Territory')
plt.show()
```



From above graph we can see that kerala is having highest literacy rate in the year 2001 and bihar is having the lowest

[54]:

```
total_crimes_2001 = crime_sub_2001.groupby('State/Union Territory')[['TOTAL -␣
↪Dacoity', 'TOTAL - Robbery', 'TOTAL - Burglary', 'TOTAL - Theft']].sum()
total_crimes_2001['total'] = total_crimes_2001.sum(axis=1)

# Create the bar plot
fig, ax = plt.subplots(figsize=(20,10))
ax.bar(total_crimes_2001.index, total_crimes_2001['total'])
ax.set_xlabel('State/Union Territory')
ax.set_ylabel('Total Crimes')
ax.set_title('Total Crimes by State/Union Territory')
plt.xticks(rotation=90)
plt.show()
```

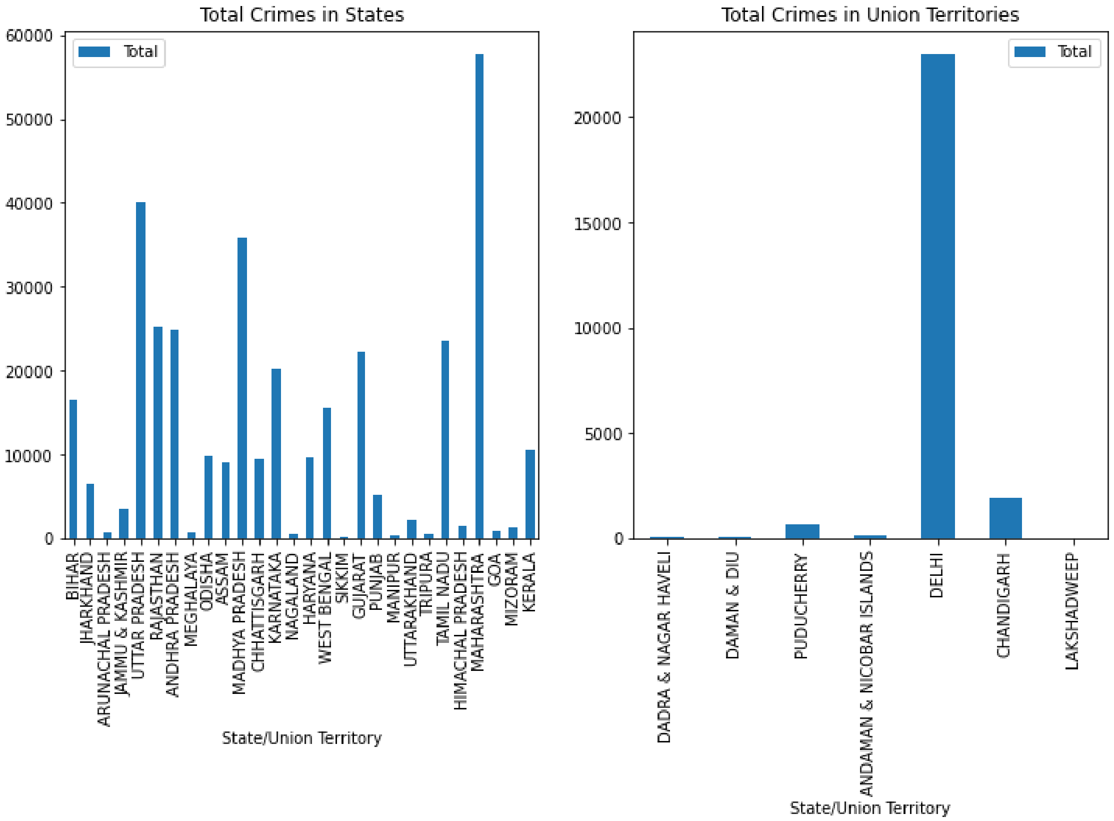


From above graph we can see in year 2001 Maharastra state is having highest crime rate

[55]:

```
state_crime = crime_sub_2001.loc[crime_sub_2001['Category'] == 'State', ['State/
↪Union Territory', 'TOTAL - Dacoity', 'TOTAL - Robbery', 'TOTAL - Burglary',␣
↪'TOTAL - Theft']]
ut_crime = crime_sub_2001.loc[crime_sub_2001['Category'] == 'Union Territory',␣
↪['State/Union Territory', 'TOTAL - Dacoity', 'TOTAL - Robbery', 'TOTAL -␣
↪Burglary', 'TOTAL - Theft']]
state_crime['Total'] = state_crime.iloc[:, 1:].sum(axis=1)
ut_crime['Total'] = ut_crime.iloc[:, 1:].sum(axis=1)
```

```
fig, (ax1, ax2) = plt.subplots(ncols=2, figsize=(12,6))
state_crime.plot(x='State/Union Territory', y='Total', kind='bar', ax=ax1)
ut_crime.plot(x='State/Union Territory', y='Total', kind='bar', ax=ax2)
ax1.set_title('Total Crimes in States')
ax2.set_title('Total Crimes in Union Territories')
plt.show()
```

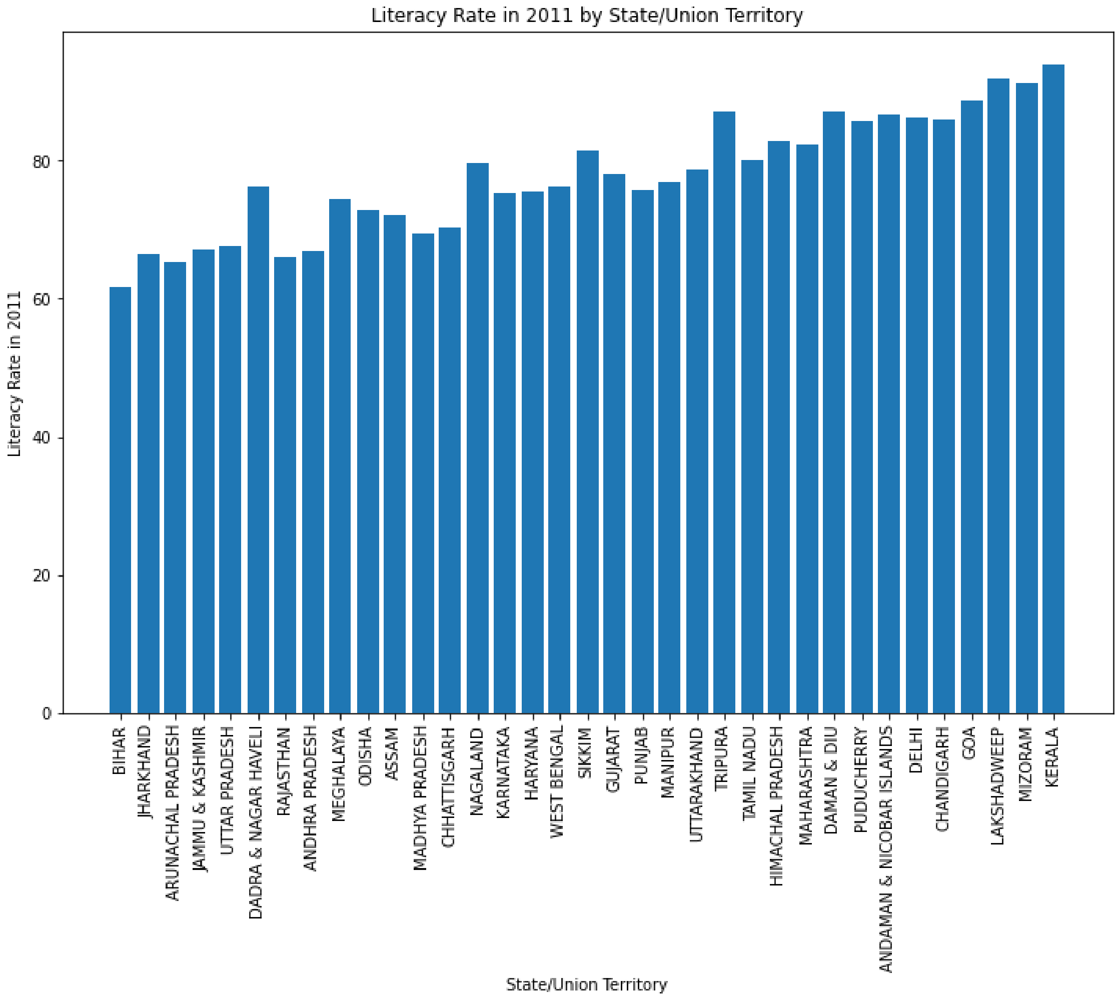


For Union Territory we can see Delhi is having high crime rates

[56]:
```
# creating subset for year 2011 with total crimes.
crime_sub_2011 = crime.loc[crime['year'] == '2011', ['State/Union Territory',␣
 ↪'year', 'TOTAL - Dacoity', 'TOTAL - Robbery',
                                            'TOTAL - Burglary','TOTAL - Theft',␣
 ↪'Category', 'Literacy_Rate_2011']]


total_crimes_2011 = crime_sub_2011.groupby('State/Union Territory')[['TOTAL -␣
 ↪Dacoity', 'TOTAL - Robbery', 'TOTAL - Burglary', 'TOTAL - Theft']].sum()
total_crimes_2011['total'] = total_crimes_2011.sum(axis=1)
```

[57]:

```
plt.figure(figsize=(12,8))
plt.bar(crime_sub_2011['State/Union Territory'],␣
 ↪crime_sub_2011['Literacy_Rate_2011'])
plt.xticks(rotation=90)
plt.xlabel('State/Union Territory')
plt.ylabel('Literacy Rate in 2011')
plt.title('Literacy Rate in 2011 by State/Union Territory')
plt.show()
```



[58]:

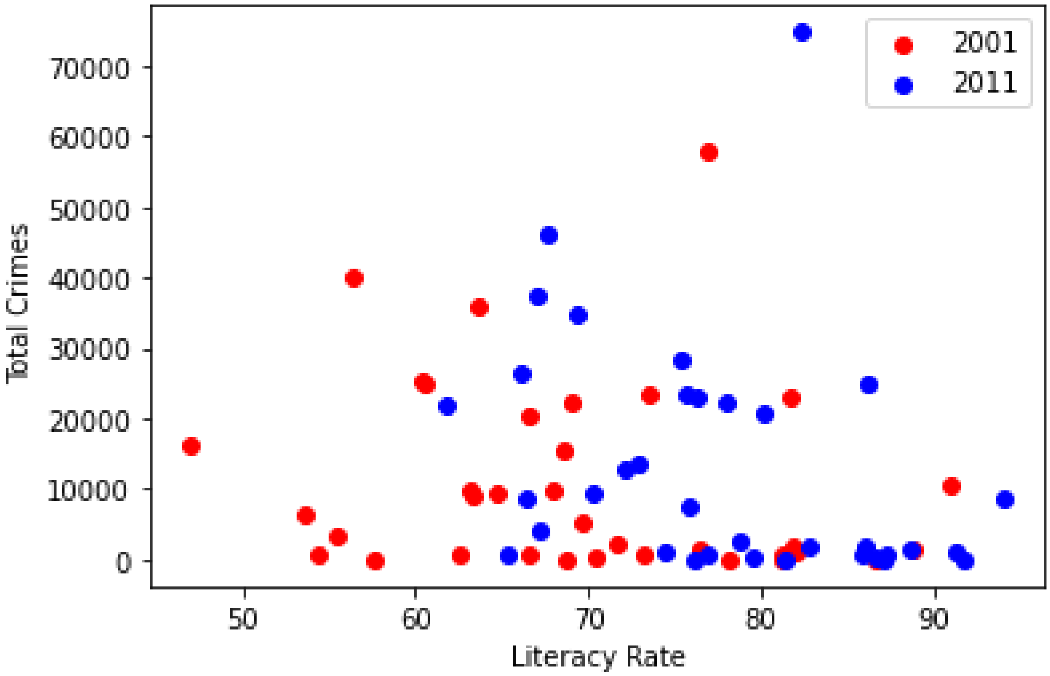
```
# Group by State/Union Territory and calculating the total crimes for each␣
 ↪category
crime_sub_2001 = crime_sub_2001.groupby('State/Union Territory').sum()
crime_sub_2011 = crime_sub_2011.groupby('State/Union Territory').sum()
```

[59]:

```
# Merging the two sub-dataframes based on the State/Union Territory column
```

```
crime_sub = crime_sub_2001.merge(crime_sub_2011, on='State/Union Territory',␣
  ↪suffixes=('_2001', '_2011'))
```

[60]:
```
# Plotting the scatter plot for Total Crimes vs Literacy Rate for both 2001 and␣
  ↪2011
plt.scatter(crime_sub['Literacy_Rate_2001'], total_crimes_2001['total'],␣
  ↪color='red', label='2001')
plt.scatter(crime_sub['Literacy_Rate_2011'], total_crimes_2011['total'],␣
  ↪color='blue', label='2011')
plt.xlabel('Literacy Rate')
plt.ylabel('Total Crimes ')
plt.legend()
plt.show()
```



[61]:
```
crime_sub_2011['Total_Crimes_2011'] = crime_sub_2011.iloc[:, 2:6].sum(axis=1)
```

[62]:
```
crime_sub_2001['Total_Crimes_2001'] = crime_sub_2001.iloc[:, 2:6].sum(axis=1)
```

[63]:
```
print(crime.columns)
```

```
Index(['State/Union Territory', 'year', 'RESIDENTIAL PREMISES - Dacoity',
       'RESIDENTIAL PREMISES - Robbery', 'RESIDENTIAL PREMISES - Burglary',
       'RESIDENTIAL PREMISES - Theft', 'HIGHWAYS - Dacoity',
       'HIGHWAYS - Robbery', 'HIGHWAYS - Burglary', 'HIGHWAYS - Theft',
       'RIVER and SEA - Dacoity', 'RIVER and SEA - Robbery',
       'RIVER and SEA - Burglary', 'RIVER and SEA - Theft',
       'RAILWAYS - Dacoity', 'RAILWAYS - Robbery', 'RAILWAYS - Burglary',
       'RAILWAYS - Theft', 'BANKS - Dacoity', 'BANKS - Robbery',
       'BANKS - Burglary', 'BANKS - Theft',
       'COMMERCIAL ESTABLISHMENTS - Dacoity',
       'COMMERCIAL ESTABLISHMENTS - Robbery',
       'COMMERCIAL ESTABLISHMENTS - Burglary',
       'COMMERCIAL ESTABLISHMENTS - Theft', 'OTHER PLACES - Dacoity',
       'OTHER PLACES - Robbery', 'OTHER PLACES - Burglary',
       'OTHER PLACES - Theft', 'TOTAL - Dacoity', 'TOTAL - Robbery',
       'TOTAL - Burglary', 'TOTAL - Theft', 'Category', 'Literacy_Rate_2001',
       'Literacy_Rate_2011', 'Area in sq.km',
       'Literacy Rate (Persons) - Rural - 2001',
       'Literacy Rate (Persons) - Rural - 2011',
       'Literacy Rate (Persons) - Urban - 2001',
       'Literacy Rate (Persons) - Urban - 2011', 'population',
       'Total - Per. Change', 'Rural - Per. Change', 'Urban - Per. Change'],
      dtype='object')
```

```
[64]: crime_subset = crime[['State/Union Territory', 'year',
                              'RESIDENTIAL PREMISES - Dacoity', 'RESIDENTIAL␣
       ↪PREMISES - Robbery',
                              'RESIDENTIAL PREMISES - Burglary', 'RESIDENTIAL␣
       ↪PREMISES - Theft',
                              'HIGHWAYS - Dacoity', 'HIGHWAYS - Robbery', 'HIGHWAYS␣
       ↪- Burglary',
                              'HIGHWAYS - Theft', 'RIVER and SEA - Dacoity', 'RIVER␣
       ↪and SEA - Robbery',
                              'RIVER and SEA - Burglary', 'RIVER and SEA - Theft',␣
       ↪'RAILWAYS - Dacoity',
                              'RAILWAYS - Robbery', 'RAILWAYS - Burglary', 'RAILWAYS␣
       ↪- Theft',
                              'BANKS - Dacoity', 'BANKS - Robbery', 'BANKS -␣
       ↪Burglary', 'BANKS - Theft',
                              'COMMERCIAL ESTABLISHMENTS - Dacoity', 'COMMERCIAL␣
       ↪ESTABLISHMENTS - Robbery',
                              'COMMERCIAL ESTABLISHMENTS - Burglary', 'COMMERCIAL␣
       ↪ESTABLISHMENTS - Theft',
                              'OTHER PLACES - Dacoity', 'OTHER PLACES - Robbery',␣
       ↪'OTHER PLACES - Burglary',
                              'OTHER PLACES - Theft', 'TOTAL - Dacoity', 'TOTAL -␣
       ↪Robbery', 'TOTAL - Burglary',
                              'TOTAL - Theft', 'Category', 'Literacy_Rate_2001',␣
       ↪'Literacy_Rate_2011',
                              'Area in sq.km', 'Literacy Rate (Persons) - Rural -␣
       ↪2001',
                              'Literacy Rate (Persons) - Rural - 2011', 'Literacy␣
       ↪Rate (Persons) - Urban - 2001',
                              'Literacy Rate (Persons) - Urban - 2011',␣
       ↪'population', 'Total - Per. Change',
                              'Rural - Per. Change', 'Urban - Per. Change']]
```

[65]:

```
# Grouping the data by State/Union Territory, year, and Category (the type of␣
 ↪crime):


grouped = crime_subset.groupby(['State/Union Territory', 'year', 'Category']).
 ↪sum().reset_index()
```

[66]:

```
# Merging the grouped data with the literacy rate data:

literacy_df = crime[['State/Union Territory', 'year', 'Literacy_Rate_2001',␣
 ↪'Literacy_Rate_2011']].drop_duplicates()
merged_df = pd.merge(grouped, literacy_df, on=['State/Union Territory', 'year'])
```

[67]:

```
yearly_total_crime = crime[['year', 'TOTAL - Dacoity', 'TOTAL - Robbery',␣
 ↪'TOTAL - Burglary', 'TOTAL - Theft']].groupby('year').sum()
```

[68]:

```
yearly_total_crime
```

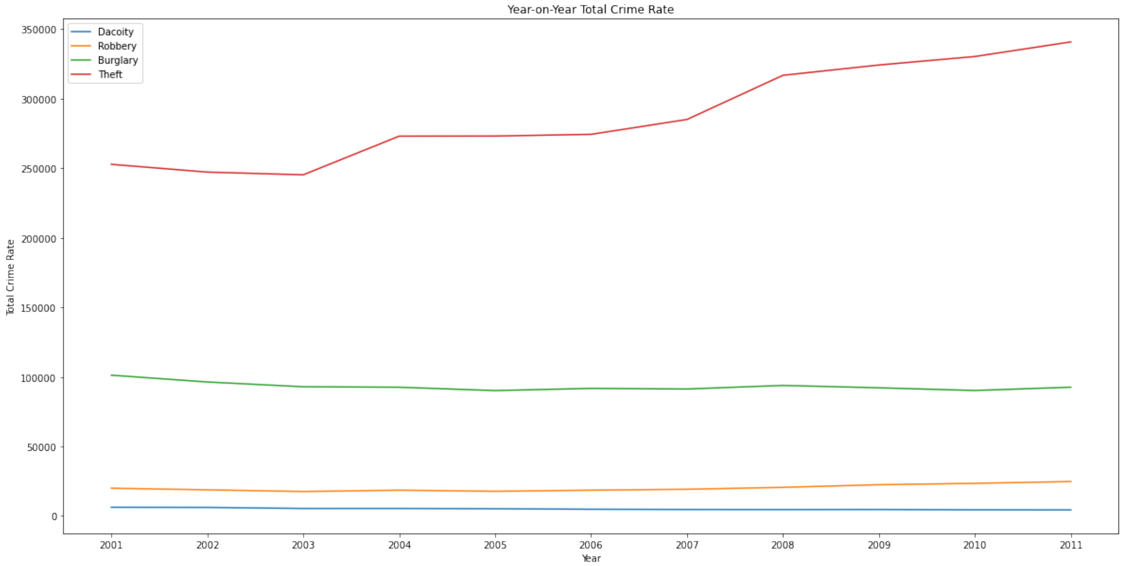
[68]:

```
      TOTAL - Dacoity  TOTAL - Robbery  TOTAL - Burglary  TOTAL - Theft
year
2001             6154            19901            101182         252803
2002             6072            18708             96269         247192
2003             5303            17512             92827         245237
2004             5311            18458             92490         273045
2005             5141            17673             90108         273111
2006             4747            18456             91666         274354
2007             4579            19136             91218         285043
2008             4532            20526             93787         316810
2009             4586            22409             92070         324195
2010             4358            23393             90179         330312
2011             4285            24700             92504         340800
```

[69]:

```
plt.subplots(figsize=(20,10))
plt.plot(yearly_total_crime.index, yearly_total_crime['TOTAL - Dacoity'],␣
 ↪label='Dacoity')
plt.plot(yearly_total_crime.index, yearly_total_crime['TOTAL - Robbery'],␣
 ↪label='Robbery')
plt.plot(yearly_total_crime.index, yearly_total_crime['TOTAL - Burglary'],␣
 ↪label='Burglary')
```

```
plt.plot(yearly_total_crime.index, yearly_total_crime['TOTAL - Theft'],␣
 ↪label='Theft')
plt.xlabel('Year')
plt.ylabel('Total Crime Rate')
plt.title('Year-on-Year Total Crime Rate')
plt.legend()
plt.show()
```



TOTAL - Theft increased year by year for all State/Union Territory. The other crimes like TOTAL - Dacoity,TOTAL - Robbery,TOTAL - Burglary' decreased year by year for State/Union Territory.

[70]:
```
# Calculating total crimes
total_crimes = crime.groupby("State/Union Territory")[["TOTAL - Dacoity",␣
 ↪"TOTAL - Robbery", "TOTAL - Burglary", "TOTAL - Theft"]].sum().sum(axis=1)

total_crimes
```

[70]:
```
State/Union Territory
ANDAMAN & NICOBAR ISLANDS      1956
ANDHRA PRADESH               357402
ARUNACHAL PRADESH              8359
ASSAM                        117472
BIHAR                        205199
CHANDIGARH                    19999
CHHATTISGARH                 104959
DADRA & NAGAR HAVELI           1137
DAMAN & DIU                    1116
DELHI                        237560
```

```
GOA                           10796
GUJARAT                      269821
HARYANA                      161510
HIMACHAL PRADESH              17920
JAMMU & KASHMIR               40813
JHARKHAND                     93945
KARNATAKA                    266498
KERALA                       110832
LAKSHADWEEP                     142
MADHYA PRADESH               389927
MAHARASHTRA                  717452
MANIPUR                        4753
MEGHALAYA                      8876
MIZORAM                       13752
NAGALAND                       5614
ODISHA                       122353
PUDUCHERRY                     7801
PUNJAB                        72051
RAJASTHAN                    279114
SIKKIM                         1778
TAMIL NADU                   232560
TRIPURA                        6612
UTTAR PRADESH                366166
UTTARAKHAND                   25483
WEST BENGAL                  181414
dtype: int64
```

[71]:

```
# Calculating crime rate
area = crime.groupby("State/Union Territory")["Area in sq.km"].mean()
crime_rate = total_crimes/area

crime_rate
```

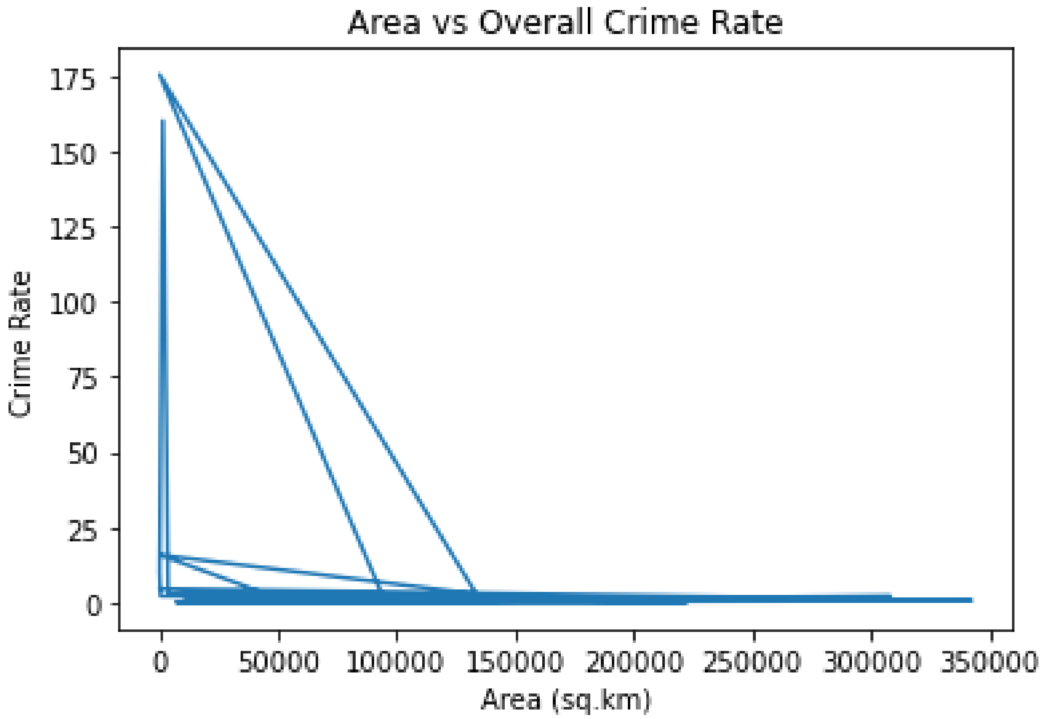
[71]:

```
State/Union Territory
ANDAMAN & NICOBAR ISLANDS      0.237120
ANDHRA PRADESH                 1.299431
ARUNACHAL PRADESH              0.099817
ASSAM                          1.497641
BIHAR                          2.179189
CHANDIGARH                   175.429825
CHHATTISGARH                   0.776370
DADRA & NAGAR HAVELI           2.315682
DAMAN & DIU                   10.054054
DELHI                        160.188806
GOA                            2.916261
GUJARAT                        1.374926
HARYANA                        3.653081
```

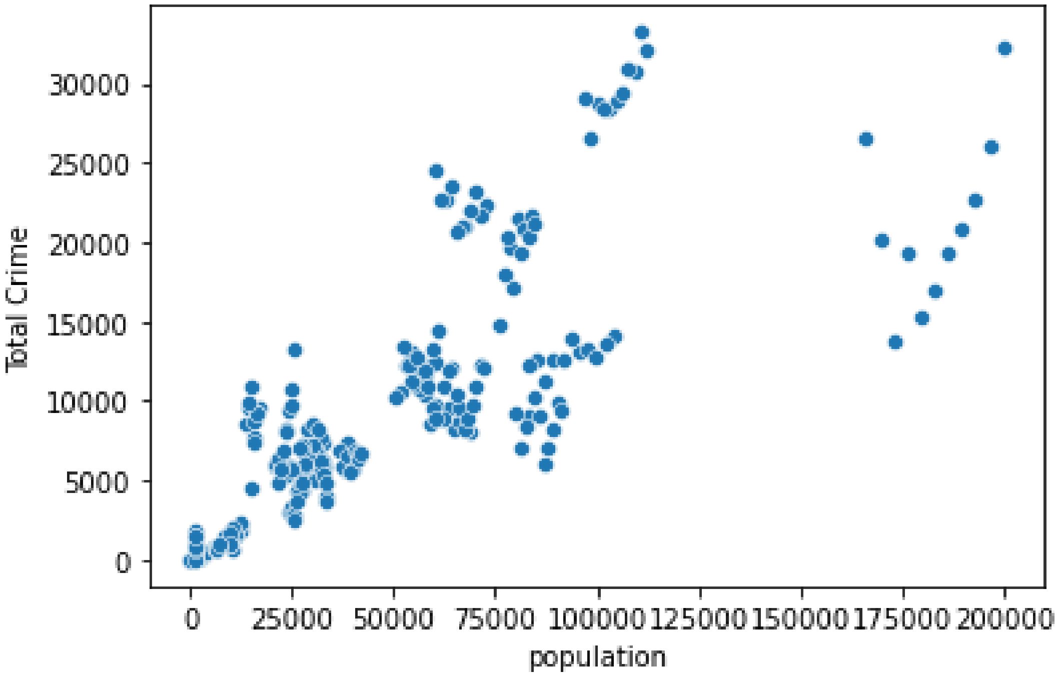
```
HIMACHAL PRADESH               0.321880
JAMMU & KASHMIR                0.183647
JHARKHAND                      1.178496
KARNATAKA                      1.389523
KERALA                         2.852672
LAKSHADWEEP                    4.733333
MADHYA PRADESH                 1.264962
MAHARASHTRA                    2.331562
MANIPUR                        0.212881
MEGHALAYA                      0.395738
MIZORAM                        0.652341
NAGALAND                       0.338621
ODISHA                         0.785790
PUDUCHERRY                    15.920408
PUNJAB                         1.430662
RAJASTHAN                      0.815553
SIKKIM                         0.250564
TAMIL NADU                     1.788098
TRIPURA                        0.630555
UTTAR PRADESH                  1.519815
UTTARAKHAND                    0.476469
WEST BENGAL                    2.044055
dtype: float64
```

[72]:

```
plt.plot(area, crime_rate)
plt.xlabel("Area (sq.km)")
plt.ylabel("Crime Rate")
plt.title("Area vs Overall Crime Rate")
plt.show()
```



[73]:
```
# Calculating total crime count for each row
crime['Total Crime'] = crime.iloc[:, 2:26].sum(axis=1)

# Plotting scatter plot
sns.scatterplot(data=crime, x='population', y='Total Crime')
plt.show()
```



[74]:

```
# Group by state
state_group = crime.groupby('State/Union Territory')

# Calculating total number of crimes for each state
state_crime = state_group.sum()

# Calculating crime rate per capita
state_crime['crime_rate'] = state_crime['TOTAL - Theft'] /␣
 ↪state_crime['population']

# Sorting by crime rate
state_crime.sort_values('crime_rate', ascending=False, inplace=True)
```

1.2.1 Analysis

Crime rate per capita refers to the number of reported crimes in a particular area or jurisdiction, divided by the population of that area. It is a measure of the frequency of crimes in relation to the size of the population

[75]:

```
# Print each state's crime report
for state, row in state_crime.iterrows():
    print(f"State/Union Territory: {state}")
    print(f"Total number of crimes: {row['TOTAL - Theft']}")
    print(f"Crime rate per capita: {row['crime_rate']:.2f}")
    print('\n')
```

```
State/Union Territory: CHANDIGARH
Total number of crimes: 16925.0
Crime rate per capita: 1.57


State/Union Territory: DELHI
Total number of crimes: 209514.0
Crime rate per capita: 1.24


State/Union Territory: MIZORAM
Total number of crimes: 9198.0
Crime rate per capita: 0.84


State/Union Territory: PUDUCHERRY
Total number of crimes: 6708.0
Crime rate per capita: 0.55


State/Union Territory: GOA
Total number of crimes: 7130.0
Crime rate per capita: 0.46


State/Union Territory: HARYANA
Total number of crimes: 113804.0
Crime rate per capita: 0.45


State/Union Territory: MAHARASHTRA
Total number of crimes: 509331.0
Crime rate per capita: 0.44


State/Union Territory: ARUNACHAL PRADESH
Total number of crimes: 4776.0
Crime rate per capita: 0.35


State/Union Territory: MADHYA PRADESH
Total number of crimes: 240678.0
Crime rate per capita: 0.33


State/Union Territory: GUJARAT
Total number of crimes: 197255.0
Crime rate per capita: 0.32
```

```
State/Union Territory: RAJASTHAN
Total number of crimes: 209660.0
Crime rate per capita: 0.30


State/Union Territory: ANDHRA PRADESH
Total number of crimes: 259715.0
Crime rate per capita: 0.29


State/Union Territory: KARNATAKA
Total number of crimes: 181542.0
Crime rate per capita: 0.29


State/Union Territory: ANDAMAN & NICOBAR ISLANDS
Total number of crimes: 1053.0
Crime rate per capita: 0.26


State/Union Territory: ASSAM
Total number of crimes: 76081.0
Crime rate per capita: 0.24


State/Union Territory: TAMIL NADU
Total number of crimes: 173164.0
Crime rate per capita: 0.23


State/Union Territory: DAMAN & DIU
Total number of crimes: 558.0
Crime rate per capita: 0.23


State/Union Territory: DADRA & NAGAR HAVELI
Total number of crimes: 705.0
Crime rate per capita: 0.23


State/Union Territory: CHHATTISGARH
Total number of crimes: 57532.0
Crime rate per capita: 0.23


State/Union Territory: JHARKHAND
```

```
Total number of crimes: 64642.0
Crime rate per capita: 0.20


State/Union Territory: JAMMU & KASHMIR
Total number of crimes: 23745.0
Crime rate per capita: 0.19


State/Union Territory: MEGHALAYA
Total number of crimes: 5461.0
Crime rate per capita: 0.19


State/Union Territory: WEST BENGAL
Total number of crimes: 167477.0
Crime rate per capita: 0.18


State/Union Territory: UTTARAKHAND
Total number of crimes: 18122.0
Crime rate per capita: 0.18


State/Union Territory: ODISHA
Total number of crimes: 73207.0
Crime rate per capita: 0.17


State/Union Territory: KERALA
Total number of crimes: 58275.0
Crime rate per capita: 0.16


State/Union Territory: PUNJAB
Total number of crimes: 45231.0
Crime rate per capita: 0.16


State/Union Territory: NAGALAND
Total number of crimes: 3421.0
Crime rate per capita: 0.16


State/Union Territory: LAKSHADWEEP
Total number of crimes: 106.0
Crime rate per capita: 0.16
```

```
State/Union Territory: SIKKIM
Total number of crimes: 902.0
Crime rate per capita: 0.14


State/Union Territory: MANIPUR
Total number of crimes: 3931.0
Crime rate per capita: 0.14


State/Union Territory: UTTAR PRADESH
Total number of crimes: 274671.0
Crime rate per capita: 0.14


State/Union Territory: BIHAR
Total number of crimes: 135705.0
Crime rate per capita: 0.13


State/Union Territory: HIMACHAL PRADESH
Total number of crimes: 8889.0
Crime rate per capita: 0.12


State/Union Territory: TRIPURA
Total number of crimes: 3788.0
Crime rate per capita: 0.11
```

1.2.2 Phase III

[76]:
```
import sqlite3

# Read the CSV file into a pandas DataFrame
df = pd.read_csv('DataTrained/
 ↪42_District_wise_crimes_committed_against_women_2001_2012.csv')

df.head()
```

[76]:
```
          STATE/UT       DISTRICT  Year  Rape  Kidnapping and Abduction  \
0   ANDHRA PRADESH       ADILABAD  2001    50                        30
1   ANDHRA PRADESH      ANANTAPUR  2001    23                        30
2   ANDHRA PRADESH       CHITTOOR  2001    27                        34
3   ANDHRA PRADESH       CUDDAPAH  2001    20                        20
```

```
4  ANDHRA PRADESH  EAST GODAVARI  2001    23                        26

   Dowry Deaths  Assault on women with intent to outrage her modesty  \
0            16                                                  149
1             7                                                  118
2            14                                                  112
3            17                                                  126
4            12                                                  109

   Insult to modesty of Women  Cruelty by Husband or his Relatives  \
0                          34                                  175
1                          24                                  154
2                          83                                  186
3                          38                                   57
4                          58                                  247

   Importation of Girls
0                     0
1                     0
2                     0
3                     0
4                     0
```

[77]:

```
# Creating a connection to a SQLite database
conn = sqlite3.connect('crime.db')

# Inserting the DataFrame into a SQLite table
df.to_sql('crimes_women', conn, if_exists='replace', index=False)
```

[77]:

```
9017
```

Write SQL query to find the highest number of rapes & Kidnappings that happened in which state, District, and year

[78]:

```
query_1 = """
        SELECT [STATE/UT], DISTRICT, Year, MAX(Rape) as MaxRape,␣
 ↪MAX([Kidnapping and Abduction]) as MaxKidnapping
        FROM crimes_women
        GROUP BY [STATE/UT], DISTRICT, Year
        ORDER BY MaxRape DESC, MaxKidnapping DESC
        LIMIT 1
        """
result_1 = pd.read_sql_query(query_1, conn)
print(result_1)
```

```
         STATE/UT DISTRICT  Year  MaxRape  MaxKidnapping
0  MADHYA PRADESH    TOTAL  2012     3425           1127
```

Write SQL query to find All the lowest number of rapes & Kidnappings that happened in which state, District, and year

```
[79]: query = """
              SELECT [STATE/UT], DISTRICT, Year, MIN(Rape) as MinRape,␣
        ↪MIN([Kidnapping and Abduction]) as MinKidnapping
              FROM crimes_women
              GROUP BY [STATE/UT], DISTRICT, Year
              ORDER BY MinRape ASC, MinKidnapping ASC
              """
      result = pd.read_sql_query(query, conn)
      print(result)
```

```
               STATE/UT DISTRICT  Year  MinRape  MinKidnapping
0        A & N ISLANDS  NICOBAR  2001        0              0
1        A & N ISLANDS  NICOBAR  2003        0              0
2        A & N ISLANDS  NICOBAR  2004        0              0
3        A & N ISLANDS  NICOBAR  2005        0              0
4        A & N ISLANDS  NICOBAR  2006        0              0
…                  …       …    …     …            …
9011  MADHYA PRADESH    TOTAL  2009     2998            841
9012  MADHYA PRADESH    TOTAL  2007     3010            701
9013  MADHYA PRADESH    TOTAL  2010     3135           1030
9014  MADHYA PRADESH    TOTAL  2011     3406           1088
9015  MADHYA PRADESH    TOTAL  2012     3425           1127

[9016 rows x 5 columns]
```

```
[80]: df_1 = pd.read_csv('DataTrained/
        ↪02_District_wise_crimes_committed_against_ST_2001_2012.csv')

      df_1.head()
```

```
[80]:          STATE/UT        DISTRICT  Year  Murder  Rape  Kidnapping Abduction  \
      0  ANDHRA PRADESH        ADILABAD  2001       0     1                     2
      1  ANDHRA PRADESH       ANANTAPUR  2001       0     0                     0
      2  ANDHRA PRADESH        CHITTOOR  2001       0     0                     0
      3  ANDHRA PRADESH        CUDDAPAH  2001       0     0                     0
      4  ANDHRA PRADESH  EAST GODAVARI  2001       0     0                     0

         Dacoity  Robbery  Arson  Hurt  Protection of Civil Rights (PCR) Act  \
      0        0        0      0     2                                     0
      1        0        0      0     7                                     0
      2        0        0      0     2                                     0
      3        0        0      0     2                                     0
      4        0        0      0     0                                     0
```

```
   Prevention of atrocities (POA) Act  Other Crimes Against STs
0                                   0                        13
1                                   1                         6
2                                   0                         0
3                                   2                         0
4                                   0                        14
```

[81]:
```
#conn = sqlite3.connect('crimes.db')

df_1.to_sql('crimes_district', conn, if_exists='replace', index=False)
```

[81]:
```
9018
```

Write SQL query to find the highest number of dacoity/robbery in which district.

[82]:
```
query1 ='''SELECT DISTRICT, MAX(Dacoity + Robbery) AS "[Highest Dacoity/
 ↪Robbery]"
               FROM crimes_district
               GROUP BY DISTRICT
               ORDER BY "[Highest Dacoity/Robbery]" DESC
               LIMIT 1'''

result = pd.read_sql_query(query1, conn)
print(result)
```

```
  DISTRICT  [Highest Dacoity/Robbery]
0    TOTAL                         63
```

[83]:
```
query2 ='''SELECT DISTRICT, MIN(Murder) AS "[Lowest Murders]"
FROM crimes_district
GROUP BY DISTRICT
ORDER BY "[Lowest Murders]" ASC'''

result = pd.read_sql_query(query2, conn)
print(result)
```

```
              DISTRICT  [Lowest Murders]
0     24 PARGANAS NORTH                 0
1     24 PARGANAS SOUTH                 0
2       A and N ISLANDS                 0
3              ADILABAD                 0
4                  AGRA                 0
..                  …               …
808         YAMUNANAGAR                 0
809            YAVATMAL                 0
810           ZUNHEBOTO                 0
811           KONDAGAON                 1
812             MUNGELI                 1
```

```
[813 rows x 2 columns]
```

Write SQL query to find the number of murders in ascending order in district and yearwise

```
[84]: query3 = '''
      SELECT DISTRICT, Year, Murder
      FROM crimes_district
      ORDER BY DISTRICT ASC, Year ASC, Murder ASC'''

      result = pd.read_sql_query(query3, conn)
      print(result)
```

```
                DISTRICT  Year  Murder
0     24 PARGANAS NORTH  2001       0
1     24 PARGANAS NORTH  2002       0
2     24 PARGANAS NORTH  2003       0
3     24 PARGANAS NORTH  2004       0
4     24 PARGANAS NORTH  2005       0
…                   …    …       …
9013          ZUNHEBOTO  2008       0
9014          ZUNHEBOTO  2009       0
9015          ZUNHEBOTO  2010       0
9016          ZUNHEBOTO  2011       0
9017          ZUNHEBOTO  2012       0

[9018 rows x 3 columns]
```

```
[85]: df_2 = pd.read_csv('DataTrained/01_District_wise_crimes_committed_IPC_2001_2012.
       ↪csv')

      df_2.head()
```

```
[85]:         STATE/UT       DISTRICT  YEAR  MURDER  ATTEMPT TO MURDER  \
      0  ANDHRA PRADESH       ADILABAD  2001     101                 60
      1  ANDHRA PRADESH      ANANTAPUR  2001     151                125
      2  ANDHRA PRADESH       CHITTOOR  2001     101                 57
      3  ANDHRA PRADESH       CUDDAPAH  2001      80                 53
      4  ANDHRA PRADESH  EAST GODAVARI  2001      82                 67

         CULPABLE HOMICIDE NOT AMOUNTING TO MURDER  RAPE  CUSTODIAL RAPE  \
      0                                         17    50               0
      1                                          1    23               0
      2                                          2    27               0
      3                                          1    20               0
      4                                          1    23               0

         OTHER RAPE  KIDNAPPING & ABDUCTION  …  ARSON  HURT/GREVIOUS HURT  \
      0          50                       46  …      30                  1131
      1          23                       53  …      69                  1543
      2          27                       59  …      38                  2088
      3          20                       25  …      23                   795
      4          23                       49  …      41                  1244

         DOWRY DEATHS  ASSAULT ON WOMEN WITH INTENT TO OUTRAGE HER MODESTY  \
      0            16                                                  149
      1             7                                                  118
      2            14                                                  112
      3            17                                                  126
      4            12                                                  109

         INSULT TO MODESTY OF WOMEN  CRUELTY BY HUSBAND OR HIS RELATIVES  \
      0                          34                                  175
      1                          24                                  154
      2                          83                                  186
      3                          38                                   57
      4                          58                                  247

         IMPORTATION OF GIRLS FROM FOREIGN COUNTRIES  CAUSING DEATH BY NEGLIGENCE  \
      0                                            0                          181
      1                                            0                          270
      2                                            0                          404
      3                                            0                          233
      4                                            0                          431

         OTHER IPC CRIMES  TOTAL IPC CRIMES
      0              1518              4154
      1               754              4125
      2              1262              5818
      3              1181              3140
      4              2313              6507

      [5 rows x 33 columns]
```

```
[86]: df_2.to_sql('crimes_district_IPC', conn, if_exists='replace', index=False)
```

```
[86]: 9017
```

Write SQL query to find which District in each state/ut has the highest number of murders yearwise. Your output should show STATE/UT, YEAR, DISTRICT, and MURDERS.

```
[87]: query1 = '''SELECT [STATE/UT], YEAR, DISTRICT, MAX(MURDER) AS MURDERS
      FROM crimes_district_IPC
      GROUP BY [STATE/UT], YEAR
      ORDER BY [STATE/UT], YEAR '''
```

```
result = pd.read_sql_query(query1, conn)
print(result)
```

```
          STATE/UT  YEAR DISTRICT  MURDERS
0    A & N ISLANDS  2001  ANDAMAN       13
1    A & N ISLANDS  2002    TOTAL       17
2    A & N ISLANDS  2003    TOTAL       21
3    A & N ISLANDS  2004    TOTAL       15
4    A & N ISLANDS  2005  ANDAMAN       14
..             …   …      …      …
415    WEST BENGAL  2008    TOTAL     1811
416    WEST BENGAL  2009    TOTAL     2068
417    WEST BENGAL  2010    TOTAL     2398
418    WEST BENGAL  2011    TOTAL     2109
419    WEST BENGAL  2012    TOTAL     2252

[420 rows x 4 columns]
```