

Phase 2-EDA

I collected the data from different sources for population, Literacy rate and others as well. I found the census data for 2001 then 2011 so I found the Growth rate for population for 10 years and applied that rate (Annually) for finding the population from 2002 to 2012. Here I got columns in my data set are-

- State/UT: Each state present in India.
- Year : 2001 to 2012
- population(total)- total population of state or UT in each year from 2001 to 2012
- Rural : Population of Rural area of that state
- 'Urban' : Population of urban area of that state,
- 'Tot_M' : Total population of Males present in that State/UT
- 'Tot_F' : Total population of Females present in that State/UT,
- 'P_LIT' - Total Literate population of each state,
- 'M_LIT' : Male literacy population of each state
- F_LIT : Female literate population in each state
- Lit_rate : Literacy Rate
- Area- Area in KM square of each state
- Murder- number of murder in each state
- 'ATTEMPT TO MURDER',
- 'CULPABLE HOMICIDE NOT AMOUNTING TO MURDER',
- 'RAPE',
- 'CUSTODIAL RAPE',
- 'OTHER RAPE',
- 'KIDNAPPING & ABDUCTION',
- 'KIDNAPPING AND ABDUCTION OF WOMEN AND GIRLS',
- 'KIDNAPPING AND ABDUCTION OF OTHERS',
- 'DACOITY',
- 'PREPARATION AND ASSEMBLY FOR DACOITY',
- 'ROBBERY',
- 'BURGLARY',
- 'THEFT',
- 'AUTO THEFT',
- 'OTHER THEFT',
- 'RIOTS',
- 'CRIMINAL BREACH OF TRUST',
- 'CHEATING',
- 'COUNTERFEITING',
- 'ARSON',
- 'HURT/GREIVIOUS HURT',
- 'DOWRY DEATHS',
- ASSAULT ON WOMEN WITH INTENT TO OUTRAGE HER MODESTY',
- 'INSULT TO MODESTY OF WOMEN',
- 'CRUELTY BY HUSBAND OR HIS RELATIVES',
- 'IMPORTATION OF GIRLS FROM FOREIGN COUNTRIES',
- 'CAUSING DEATH BY NEGLIGENCE',

- 'OTHER IPC CRIMES',
- 'TOTAL IPC CRIMES',
- 'Total Crimes',
- 'tot_crimes_sc': Total crimes against SC comitted in each year,
- 'total crime against women : Total crimes against Women comitted in each year,
- 'Total crimes against STs': Total crimes against STs comitted in each year,
- 'Total crime against children- Total crimes against Children comitted in each year
- Crime Rate- A crime rate is defined as the total number of crimes performed per a certain number of people in a specified area. This is typically expressed per 100,000 people.

```
In [1]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import plotly.express as px

import plotly
import plotly.graph_objects as go
import plotly.offline as pyo
from plotly.offline import init_notebook_mode

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

```
In [2]: df=pd.read_csv('finaldata.csv')
```

```
In [3]: print("We have {} Rows and {} Columns in our dataframe".format(df.shape[0], df.
df.head())
```

We have 420 Rows and 48 Columns in our dataframe

Out[3]:

	State/UT	Year	population(total)	Rural	Urban	Tot_M	Tot_F	P_LIT	
0	A& N ISLANDS	2001	356152	239954	116198	192972	163180	253135.0	
1	ANDHRA PRADESH	2001	76210007	55401067	20808940	38527413	37682594	39934323.0	23
2	ARUNACHAL PRADESH	2001	1097968	870087	227881	579941	518027	484785.0	
3	ASSAM	2001	26655528	23216288	3439240	13777037	12878491	14015354.0	8
4	BIHAR	2001	82998509	74316709	8681800	43243795	39754714	31109577.0	20

5 rows × 48 columns

In [4]: `df.info()`

```
<class 'pandas.core.frame.DataFrame'>
```

```
RangeIndex: 420 entries, 0 to 419
```

```
Data columns (total 48 columns):
```

#	Column	Non-Null Count	Dty
pe			
---	-----	-----	---
--			
0	State/UT	420 non-null	obj
ect			
1	Year	420 non-null	int
64			
2	population(total)	420 non-null	int
64			
3	Rural	420 non-null	int
64			
4	Urban	420 non-null	int
64			
5	Tot_M	420 non-null	int
64			
6	Tot_F	420 non-null	int
64			
7	P_LIT	420 non-null	flo
at64			
8	M_LIT	420 non-null	flo
at64			
9	F_LIT	420 non-null	int
64			
10	Lit_rate	420 non-null	flo
at64			
11	Area (km2)	420 non-null	int
64			
12	MURDER	420 non-null	int
64			
13	ATTEMPT TO MURDER	420 non-null	int
64			
14	CULPABLE HOMICIDE NOT AMOUNTING TO MURDER	420 non-null	int
64			
15	RAPE	420 non-null	int
64			
16	CUSTODIAL RAPE	420 non-null	int
64			
17	OTHER RAPE	420 non-null	int
64			
18	KIDNAPPING & ABDUCTION	420 non-null	int
64			
19	KIDNAPPING AND ABDUCTION OF WOMEN AND GIRLS	420 non-null	int
64			
20	KIDNAPPING AND ABDUCTION OF OTHERS	420 non-null	int
64			
21	DACOITY	420 non-null	int
64			
22	PREPARATION AND ASSEMBLY FOR DACOITY	420 non-null	int
64			
23	ROBBERY	420 non-null	int
64			
24	BURGLARY	420 non-null	int
64			

25	THEFT	420	non-null	int
64				
26	AUTO THEFT	420	non-null	int
64				
27	OTHER THEFT	420	non-null	int
64				
28	RIOTS	420	non-null	int
64				
29	CRIMINAL BREACH OF TRUST	420	non-null	int
64				
30	CHEATING	420	non-null	int
64				
31	COUNTERFIETING	420	non-null	int
64				
32	ARSON	420	non-null	int
64				
33	HURT/GREVIOUS HURT	420	non-null	int
64				
34	DOWRY DEATHS	420	non-null	int
64				
35	ASSAULT ON WOMEN WITH INTENT TO OUTRAGE HER MODESTY	420	non-null	int
64				
36	INSULT TO MODESTY OF WOMEN	420	non-null	int
64				
37	CRUELTY BY HUSBAND OR HIS RELATIVES	420	non-null	int
64				
38	IMPORTATION OF GIRLS FROM FOREIGN COUNTRIES	420	non-null	int
64				
39	CAUSING DEATH BY NEGLIGENCE	420	non-null	int
64				
40	OTHER IPC CRIMES	420	non-null	int
64				
41	TOTAL IPC CRIMES	420	non-null	int
64				
42	Total Crimes	420	non-null	int
64				
43	tot_crimes_sc	420	non-null	int
64				
44	total crime against women	411	non-null	flo
at64				
45	Total crimes against STs	387	non-null	flo
at64				
46	Total crime against children	418	non-null	flo
at64				
47	Crime Rate	420	non-null	flo
at64				

dtypes: float64(7), int64(40), object(1)
memory usage: 157.6+ KB

Here we can see data tye of each column and memory occupied and information about the Nulls as well

```
In [5]: ## Columns of the dataset
df.columns
```

```
Out[5]: Index(['State/UT', 'Year', 'population(total)', 'Rural', 'Urban', 'Tot_M',
              'Tot_F', 'P_LIT', 'M_LIT', 'F_LIT', 'Lit_rate', 'Area (km2)', 'MURDER',
              'ATTEMPT TO MURDER', 'CULPABLE HOMICIDE NOT AMOUNTING TO MURDER',
              'RAPE', 'CUSTODIAL RAPE', 'OTHER RAPE', 'KIDNAPPING & ABDUCTION',
              'KIDNAPPING AND ABDUCTION OF WOMEN AND GIRLS',
              'KIDNAPPING AND ABDUCTION OF OTHERS', 'DACOITY',
              'PREPARATION AND ASSEMBLY FOR DACOITY', 'ROBBERY', 'BURGLARY', 'THEFT',
              'AUTO THEFT', 'OTHER THEFT', 'RIOTS', 'CRIMINAL BREACH OF TRUST',
              'CHEATING', 'COUNTERFEITING', 'ARSON', 'HURT/GREIVIOUS HURT',
              'DOWRY DEATHS', 'ASSAULT ON WOMEN WITH INTENT TO OUTRAGE HER MODESTY',
              'INSULT TO MODESTY OF WOMEN', 'CRUELTY BY HUSBAND OR HIS RELATIVES',
              'IMPORTATION OF GIRLS FROM FOREIGN COUNTRIES',
              'CAUSING DEATH BY NEGLIGENCE', 'OTHER IPC CRIMES', 'TOTAL IPC CRIMES',
              'Total Crimes', 'tot_crimes_sc', 'total crime against women',
              'Total crimes against STs', 'Total crime against children',
              'Crime Rate'],
              dtype='object')
```

```
In [6]: (df['State/UT'].unique())
```

```
Out[6]: array(['A& N ISLANDS', 'ANDHRA PRADESH', 'ARUNACHAL PRADESH', 'ASSAM',
              'BIHAR', 'CHANDIGARH', 'CHHATTISGARH', 'D & N HAVELI',
              'DAMAN & DIU', 'DELHI', 'GOA', 'GUJARAT', 'HARYANA',
              'HIMACHAL PRADESH', 'JAMMU & KASHMIR', 'JHARKHAND', 'KARNATAKA',
              'KERALA', 'LAKSHADWEEP', 'MADHYA PRADESH', 'MAHARASHTRA',
              'MANIPUR', 'MEGHALAYA', 'MIZORAM', 'NAGALAND', 'ODISHA',
              'PUDUCHERRY', 'PUNJAB', 'RAJASTHAN', 'SIKKIM', 'TAMIL NADU',
              'TRIPURA', 'UTTAR PRADESH', 'UTTARAKHAND', 'WEST BENGAL'],
              dtype=object)
```

```
In [7]: len(df['State/UT'].unique())
```

```
Out[7]: 35
```

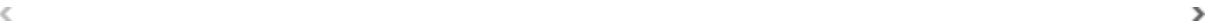
The data has total 35 states and union territories(from 2001 to 2012), but there are 36 as of 2023 (Telangana state is missing in the dataset)

```
In [8]: df.head()
```

Out[8]:

	State/UT	Year	population(total)	Rural	Urban	Tot_M	Tot_F	P_LIT	
0	A& N ISLANDS	2001	356152	239954	116198	192972	163180	253135.0	
1	ANDHRA PRADESH	2001	76210007	55401067	20808940	38527413	37682594	39934323.0	23
2	ARUNACHAL PRADESH	2001	1097968	870087	227881	579941	518027	484785.0	
3	ASSAM	2001	26655528	23216288	3439240	13777037	12878491	14015354.0	8
4	BIHAR	2001	82998509	74316709	8681800	43243795	39754714	31109577.0	20

5 rows × 48 columns



```
In [9]: df.isnull().sum()
```

```
Out[9]: State/UT      0
Year      0
population(total)    0
Rural      0
Urban      0
Tot_M      0
Tot_F      0
P_LIT      0
M_LIT      0
F_LIT      0
Lit_rate    0
Area (km2)    0
MURDER      0
ATTEMPT TO MURDER    0
CULPABLE HOMICIDE NOT AMOUNTING TO MURDER    0
RAPE        0
CUSTODIAL RAPE      0
OTHER RAPE      0
KIDNAPPING & ABDUCTION    0
KIDNAPPING AND ABDUCTION OF WOMEN AND GIRLS    0
KIDNAPPING AND ABDUCTION OF OTHERS    0
DACOITY      0
PREPARATION AND ASSEMBLY FOR DACOITY    0
ROBBERY      0
BURGLARY      0
THEFT        0
AUTO THEFT    0
OTHER THEFT    0
RIOTS        0
CRIMINAL BREACH OF TRUST    0
CHEATING      0
COUNTERFIETING    0
ARSON        0
HURT/GREVIOUS HURT    0
DOWRY DEATHS    0
ASSAULT ON WOMEN WITH INTENT TO OUTRAGE HER MODESTY    0
INSULT TO MODESTY OF WOMEN    0
CRUELTY BY HUSBAND OR HIS RELATIVES    0
IMPORTATION OF GIRLS FROM FOREIGN COUNTRIES    0
CAUSING DEATH BY NEGLIGENCE    0
OTHER IPC CRIMES    0
TOTAL IPC CRIMES    0
Total Crimes    0
tot_crimes_sc    0
total crime against women    9
Total crimes against STs    33
Total crime against children    2
Crime Rate      0
dtype: int64
```

we got few null values in columns

- total crime against women
- Total crimes against STs

- Total crime against children

our dataset

Filling Null Values using mean methods

```
In [10]: df['Total crimes against STs'].describe()
```

```
Out[10]: count      387.000000  
mean       185.077519  
std        393.275773  
min         0.000000  
25%         0.000000  
50%         6.000000  
75%        209.500000  
max       2894.000000  
Name: Total crimes against STs, dtype: float64
```

```
In [11]: df['Total crimes against STs']=df['Total crimes against STs'].fillna(df['Total
```

```
In [12]: df['Total crime against children'].describe()
```

```
Out[12]: count      418.000000  
mean       587.267943  
std       1036.458881  
min         0.000000  
25%         26.000000  
50%        103.000000  
75%        632.250000  
max       6033.000000  
Name: Total crime against children, dtype: float64
```

```
In [13]: df['Total crime against children']=df['Total crime against children'].fillna(df
```

```
In [14]: df['total crime against women']=df['total crime against women'].fillna(df['total
```

__Again Checking for null values.

```
In [15]: df.isnull().sum().sum()
```

```
Out[15]: 0
```

```
In [16]: df['State/UT'].value_counts()
```

```
Out[16]: A& N ISLANDS      12
          PUDUCHERRY      12
          MAHARASHTRA     12
          MANIPUR         12
          MEGHALAYA       12
          MIZORAM         12
          NAGALAND        12
          ODISHA          12
          PUNJAB          12
          LAKSHADWEEP     12
          RAJASTHAN       12
          SIKKIM          12
          TAMIL NADU      12
          TRIPURA        12
          UTTAR PRADESH   12
          UTTARAKHAND     12
          MADHYA PRADESH  12
          KERALA          12
          ANDHRA PRADESH  12
          DAMAN & DIU     12
          ARUNACHAL PRADESH 12
          ASSAM           12
          BIHAR           12
          CHANDIGARH     12
          CHHATTISGARH   12
          D & N HAVELI    12
          DELHI           12
          KARNATAKA       12
          GOA             12
          GUJARAT        12
          HARYANA         12
          HIMACHAL PRADESH 12
          JAMMU & KASHMIR 12
          JHARKHAND       12
          WEST BENGAL     12
          Name: State/UT, dtype: int64
```

Each state is occuring total 12 times as we have our data from 2001 to 2012.

In [17]:

df.describe()

Out[17]:

	Year	population(total)	Rural	Urban	Tot_M	Tot_F	
count	420.000000	4.200000e+02	4.200000e+02	4.200000e+02	4.200000e+02	4.200000e+02	4.
mean	2006.500000	3.210516e+07	2.325755e+07	8.964941e+06	1.659190e+07	1.560647e+07	1.
std	3.456169	4.090185e+07	3.099496e+07	1.152126e+07	2.116180e+07	1.960728e+07	2.
min	2001.000000	6.065000e+04	3.368300e+04	2.696700e+04	3.113100e+04	2.951900e+04	4.
25%	2003.750000	1.302120e+06	7.259828e+05	5.533638e+05	6.866452e+05	6.526302e+05	8.
50%	2006.500000	1.507539e+07	8.597313e+06	3.753114e+06	8.333013e+06	7.219828e+06	1.
75%	2009.250000	5.705401e+07	3.649124e+07	1.504703e+07	2.947709e+07	2.766350e+07	3.
max	2012.000000	2.022266e+08	1.597836e+08	4.826007e+07	1.070925e+08	9.666648e+07	1.

8 rows × 47 columns



The Above describe method showing all the min,max, 25%,50% ,75% statistics of dataset

- Max Lit rate is 94.5 for year 2012 and Min Lit rate was 47 for 2001 year.

Visulaization

- Lets start visulaising the columns , here we starting with the Type of crime like , Murder, Rape, Robbery happend in each year


```

In [18]: #Bar charts of every crime over time from the year 2001 to 2012
fig, axes = plt.subplots(7, 3, figsize=(45, 35))

axes[0,0].set_title("Chart of MURDER cases in India in 2001-2012")
axes[0,0].bar(df['Year'], df['MURDER'], color = 'black');
plt.xlabel('Year') #X-axis
plt.ylabel('Cases of MURDER in India') #Y-axis

axes[0,1].set_title("Chart of ATTEMPT TO MURDER cases in India in 2001-2012")
axes[0,1].bar(df['Year'], df['ATTEMPT TO MURDER'], color = 'violet');
plt.xlabel('Year') #X-axis
plt.ylabel('Cases of ATTEMPT TO MURDER in India') #Y-axis

axes[0,2].set_title("Chart of CULPABLE HOMICIDE NOT AMOUNTING TO MURDER cases in India in 2001-2012")
axes[0,2].bar(df['Year'], df['CULPABLE HOMICIDE NOT AMOUNTING TO MURDER'], color = 'blue');
plt.xlabel('Year') #X-axis
plt.ylabel('Cases of CULPABLE HOMICIDE NOT AMOUNTING TO MURDER in India') #Y-axis

axes[1,0].set_title("Chart of RAPE in 2001-2012")
axes[1,0].bar(df['Year'], df['RAPE'], color = 'cyan');
plt.xlabel('Year') #X-axis
plt.ylabel('Cases of RAPE in India') #Y-axis

axes[1,1].set_title("Chart of CUSTODIAL RAPE cases in India in 2001-2012")
axes[1,1].bar(df['Year'], df['CUSTODIAL RAPE'], color = 'orange');
plt.xlabel('Year') #X-axis
plt.ylabel('Cases of CUSTODIAL RAPE in India') #Y-axis

axes[1,2].set_title("Chart of KIDNAPPING & ABDUCTION in India in 2001-2012")
axes[1,2].bar(df['Year'], df['KIDNAPPING & ABDUCTION'], color = 'red');
plt.xlabel('Year') #X-axis
plt.ylabel('Cases of KIDNAPPING & ABDUCTION in India') #Y-axis

axes[2,0].set_title("Chart of KIDNAPPING AND ABDUCTION OF WOMEN AND GIRLS in India in 2001-2012")
axes[2,0].bar(df['Year'], df['KIDNAPPING AND ABDUCTION OF WOMEN AND GIRLS'], color = 'purple');
plt.xlabel('Year') #X-axis
plt.ylabel('Cases of KIDNAPPING AND ABDUCTION OF WOMEN AND GIRLS in India') #Y-axis

axes[2,1].set_title("Chart of DACOITY in India in 2001-2012")
axes[2,1].bar(df['Year'], df['DACOITY'], color = 'green');
plt.xlabel('Year') #X-axis
plt.ylabel('Cases of DACOITY in India') #Y-axis

axes[2,2].set_title("Chart of ROBBERY in India in 2001-2012")
axes[2,2].bar(df['Year'], df['ROBBERY'], color = 'lightpink');
plt.xlabel('Year') #X-axis
plt.ylabel('Cases of ROBBERY in India') #Y-axis

axes[3,0].set_title("Chart of BURGLARY in India in 2001-2012")
axes[3,0].bar(df['Year'], df['BURGLARY'], color = 'darkgreen');
plt.xlabel('Year') #X-axis
plt.ylabel('Cases of BURGLARY in India') #Y-axis

axes[3,1].set_title("Chart of THEFT in India in 2001-2012")
axes[3,1].bar(df['Year'], df['THEFT'], color = 'darkblue');

```

```
plt.xlabel('Year') #X-axis
plt.ylabel('Cases of THEFT in India') #Y-axis

axes[3,2].set_title("Chart of AUTO THEFT in India in 2001-2012")
axes[3,2].bar(df['Year'], df['AUTO THEFT'], color = 'black');
plt.xlabel('Year') #X-axis
plt.ylabel('Cases of AUTO THEFT in India') #Y-axis

axes[4,0].set_title("Chart of RIOTS in India in 2001-2012")
axes[4,0].bar(df['Year'], df['RIOTS'], color = 'navy');
plt.xlabel('Year') #X-axis
plt.ylabel('Cases of RIOTS') #Y-axis

axes[4,1].set_title("Chart of CRIMINAL BREACH OF TRUST in India in 2001-2012")
axes[4,1].bar(df['Year'], df['CRIMINAL BREACH OF TRUST'], color = 'red');
plt.xlabel('Year') #X-axis
plt.ylabel('Cases of CRIMINAL BREACH OF TRUST in India') #Y-axis

axes[4,2].set_title("Chart of CHEATING in India in 2001-2012")
axes[4,2].bar(df['Year'], df['CHEATING'], color = 'black');
plt.xlabel('Year') #X-axis
plt.ylabel('Cases of CHEATING in India') #Y-axis

axes[5,0].set_title("Chart of COUNTERFIETING in India in 2001-2012")
axes[5,0].bar(df['Year'], df['COUNTERFIETING'], color = 'red');
plt.xlabel('Year') #X-axis
plt.ylabel('Cases of COUNTERFIETING in India') #Y-axis

axes[5,1].set_title("Chart of ARSON in India in 2001-2012")
axes[5,1].bar(df['Year'], df['ARSON'], color = 'cyan');
plt.xlabel('Year') #X-axis
plt.ylabel('Cases of ARSON in India') #Y-axis

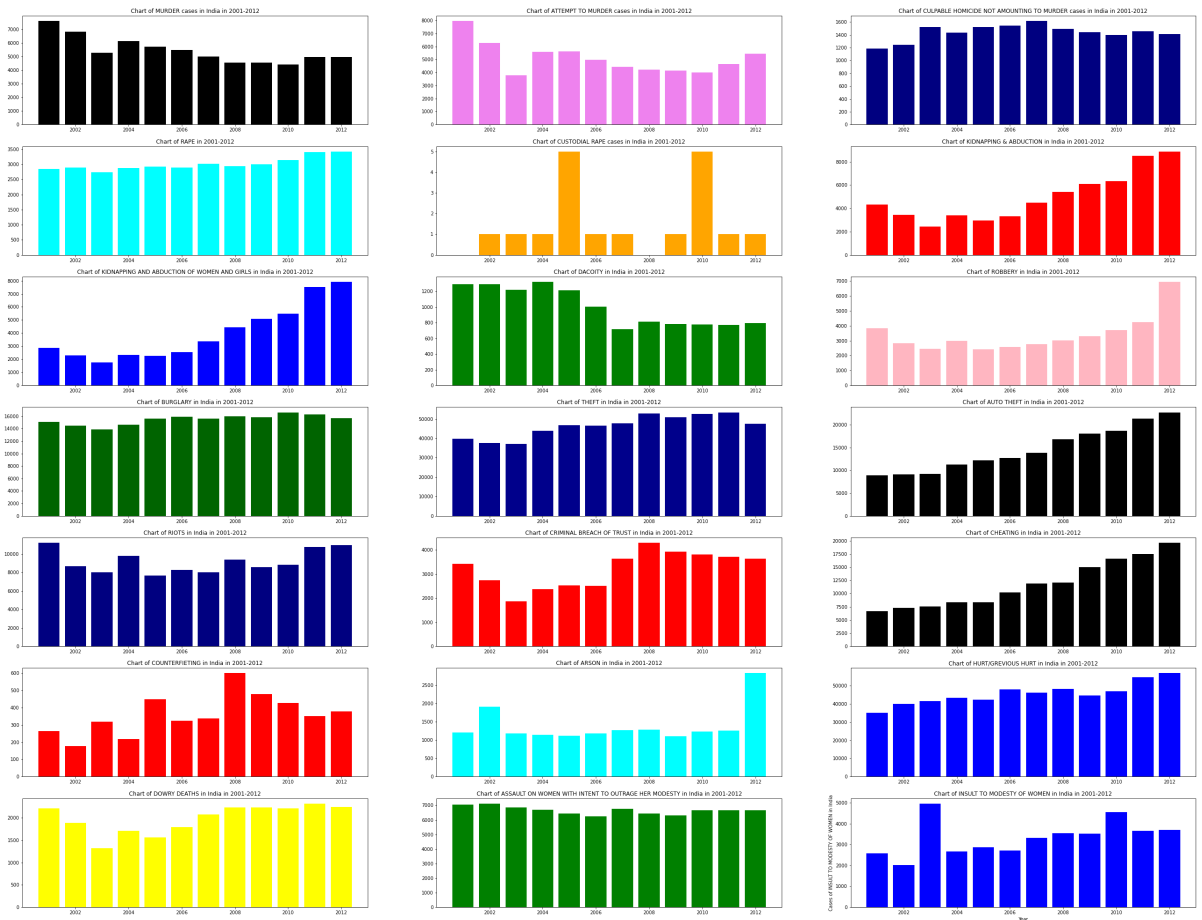
axes[5,2].set_title("Chart of HURT/GREIVIOUS HURT in India in 2001-2012")
axes[5,2].bar(df['Year'], df['HURT/GREIVIOUS HURT'], color = 'blue');
plt.xlabel('Year') #X-axis
plt.ylabel('Cases of HURT/GREIVIOUS HURT in India') #Y-axis

axes[6,0].set_title("Chart of DOWRY DEATHS in India in 2001-2012")
axes[6,0].bar(df['Year'], df['DOWRY DEATHS'], color = 'yellow');
plt.xlabel('Year') #X-axis
plt.ylabel('Cases of DOWRY DEATHS in India') #Y-axis

axes[6,1].set_title("Chart of ASSAULT ON WOMEN WITH INTENT TO OUTRAGE HER MODES")
axes[6,1].bar(df['Year'], df['ASSAULT ON WOMEN WITH INTENT TO OUTRAGE HER MODES'])
plt.xlabel('Year') #X-axis
plt.ylabel('Cases of ASSAULT ON WOMEN WITH INTENT TO OUTRAGE HER MODESTY in India')

axes[6,2].set_title("Chart of INSULT TO MODESTY OF WOMEN in India in 2001-2012")
axes[6,2].bar(df['Year'], df['INSULT TO MODESTY OF WOMEN'], color = 'blue');
plt.xlabel('Year') #X-axis
plt.ylabel('Cases of INSULT TO MODESTY OF WOMEN in India') #Y-axis
```

Out[18]: Text(0, 0.5, 'Cases of INSULT TO MODESTY OF WOMEN in India')



Observations:

- Cases of **Murder** , **Attempt to Murder** and **Daocity** has shown an decreasing trend over the year.
- We can see that the number of cases of **Rapes**, **kidnappings**, **Robbery**, **Auto Theft**, **Cheating**, **Grievous Hurt**, **Arson**, has shown an increasing trend over the ,years.
- **Burglary**, **Dowry Deaths**, **Assault on women wth intent to outrage her modesty**, **Culpable Homicide not amounting to murder** cases have remained the same and do not show any trend (upward or downward)
- **Custodial Rape** shows few peaks like in year 2005 and 2010. Cunterfieting has no fix trend

Plotting Pie chart for different crimes

```
In [19]: tot_murder= df['MURDER'].sum()
tot_rape= df['RAPE'].sum()
tot_dowrydeaths = df['DOWRY DEATHS'].sum()

tot_kidnap= df['KIDNAPPING & ABDUCTION'].sum()
tot_dacoity= df['DACOITY'].sum()
tot_robbery = df['ROBBERY'].sum()
tot_burglary= df['BURGLARY'].sum()
tot_theft= df['THEFT'].sum()
tot_riots = df['DOWRY DEATHS'].sum()

crime_group = ['TOTAL Murder','TOTAL rape','Total Dowry Deaths','Total Kidnapping',
values = [tot_murder,tot_rape,tot_dowrydeaths,tot_kidnap,tot_dacoity,tot_robbery,tot_burglary,tot_theft,tot_riots]

colors = ['crimson','gold','green','yellow','blue','black']

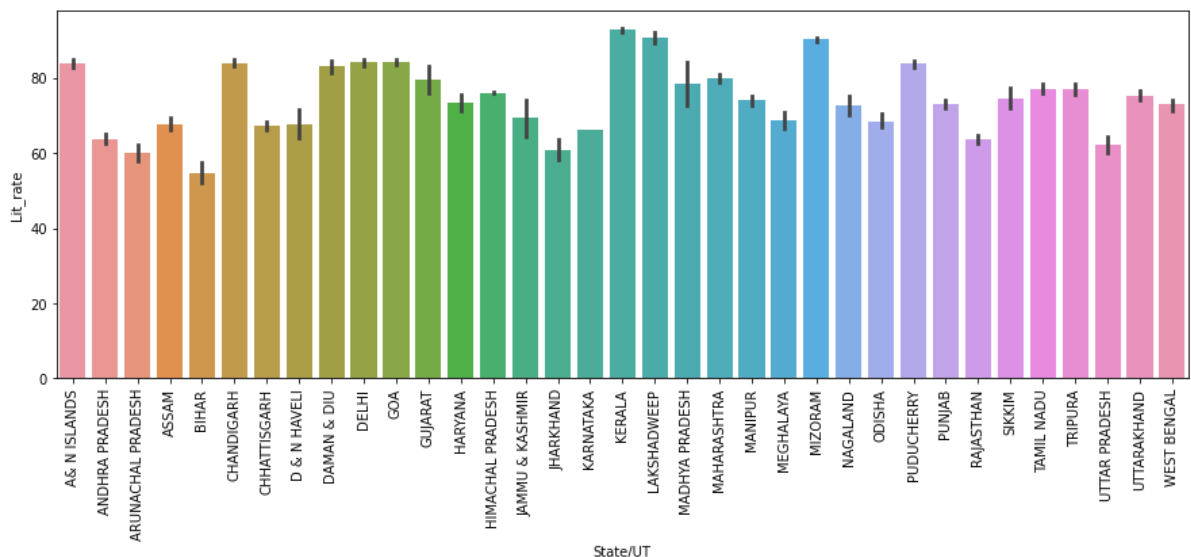
fig = go.Figure(data=[go.Pie(labels=crime_group, values=values,sort=False,
marker=dict(colors=colors),textfont_size=12)])

fig.show(renderer='svg')
```

From above Pie chart we can clearly see that Murder is the most common Crimes that is happening in each state(57%)

State/UT VS Literacy Rate

```
In [20]: plt.figure(figsize=(15, 5))
sns.barplot(x='State/UT', y='Lit_rate',data=df)
plt.xticks(rotation='90')
plt.savefig("1.jpg")
plt.show()
```

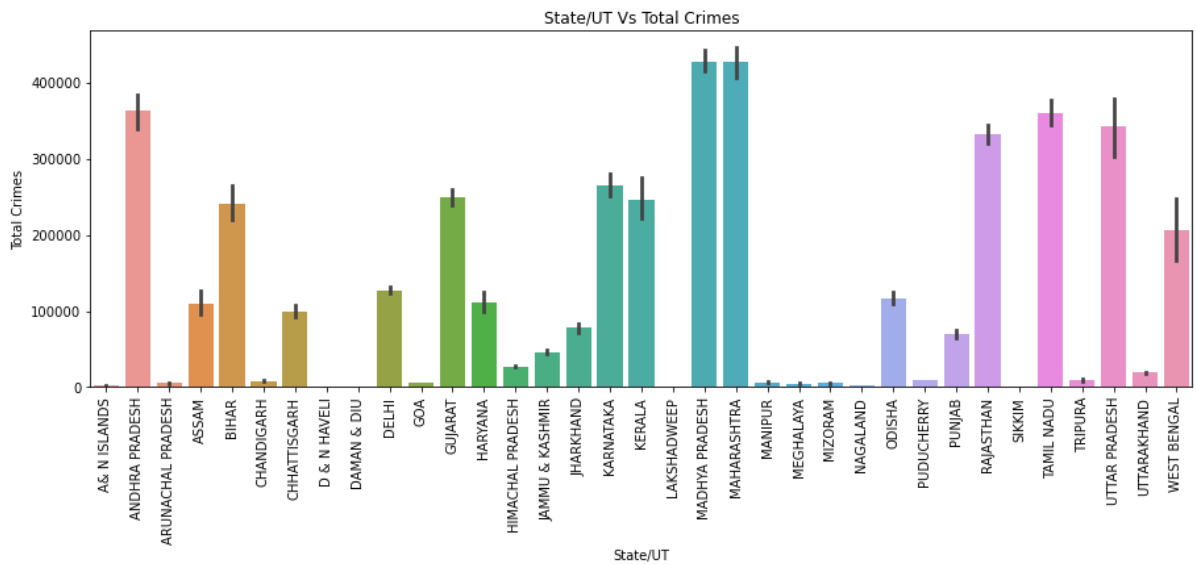


As per this barplot State Kerala has the Highest Literacy_rate and Bihar has the lowest Literacy rate.

State/UT VS Total Crimes

In [21]:

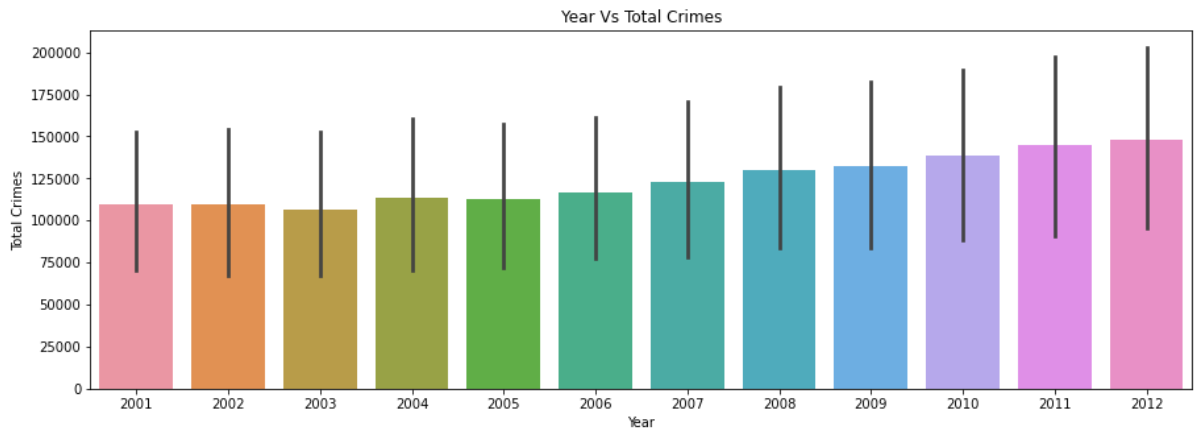
```
plt.figure(figsize=(15, 5))
sns.barplot(x=df['State/UT'], y=df['Total Crimes'])
plt.xticks(rotation='90')
plt.title('State/UT Vs Total Crimes')
plt.savefig("2.jpg")
plt.show()
```



- States Arunachal Pradesh, Chandigarh, D&N Haveli, Goa, Manipur, Meghalaya, Mizoram, Nagaland, Puducherry, Sikkim has least Total crimes as compare to Madhya Pradesh, Maharashtra, Uttar Pradesh has highest Total Crimes

Year Vs Total Crimes

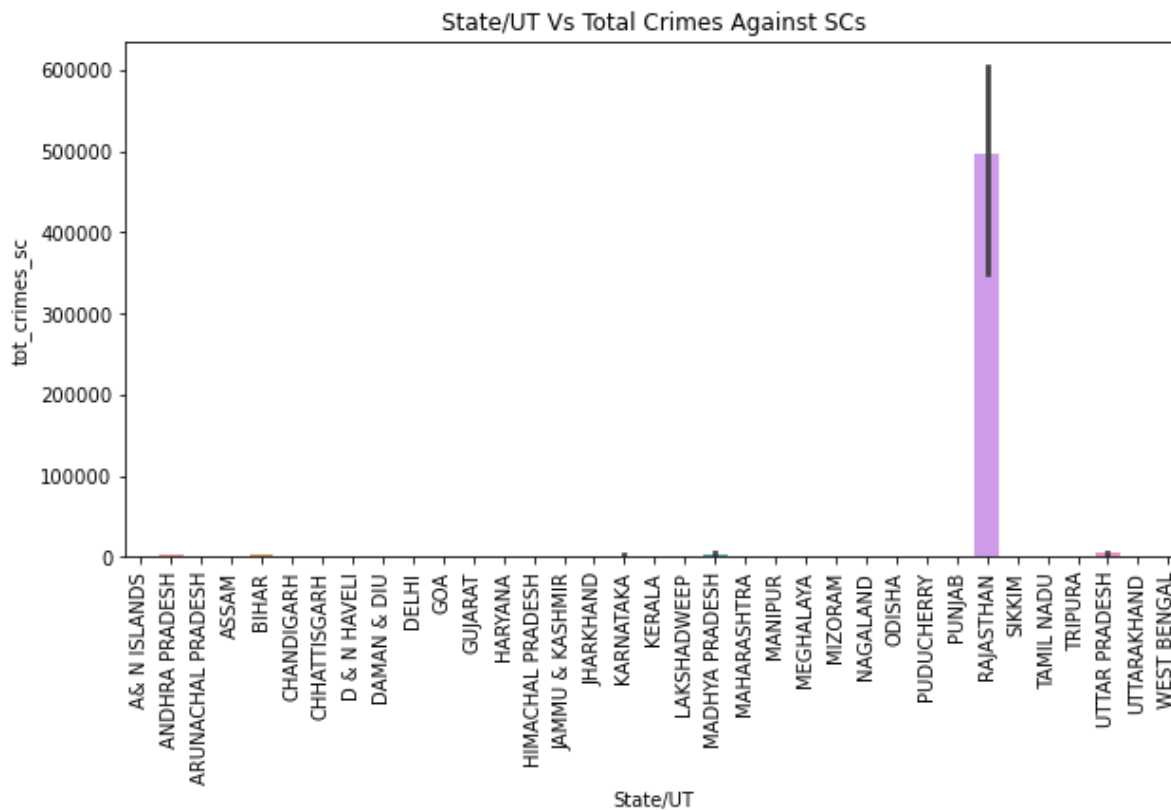
```
In [22]: plt.figure(figsize=(15, 5))
sns.barplot(x=df['Year'], y=df['Total Crimes'])
plt.title('Year Vs Total Crimes')
plt.savefig("3.jpg")
plt.show()
```



- In year 2012 Total Crimes were Highest and 2001 was the lowest.

State/UT VS Total Crimes Against SCs

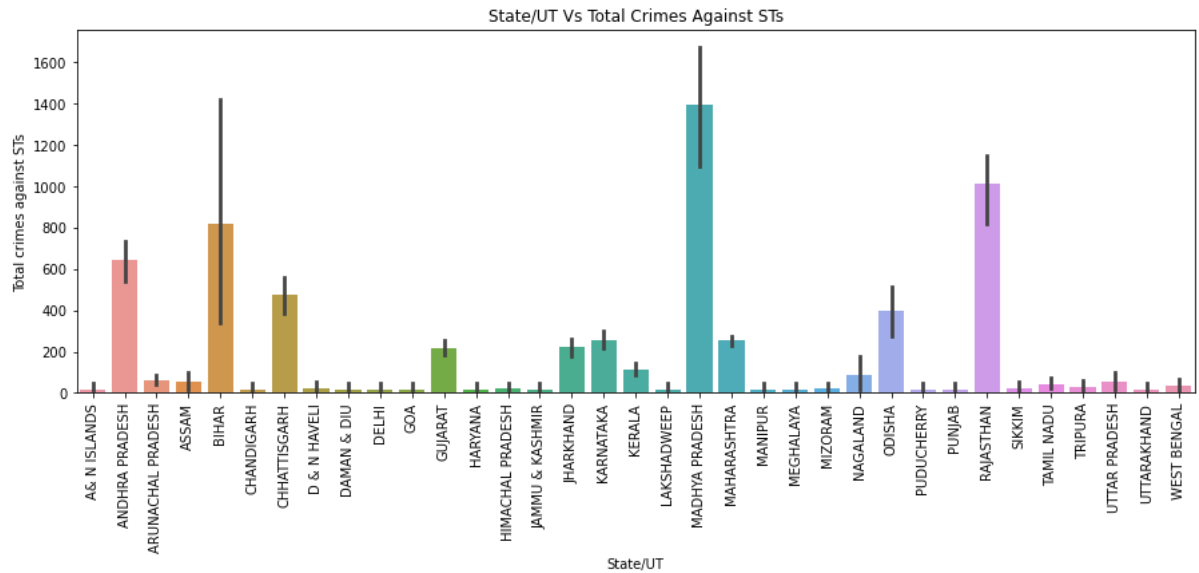
```
In [23]: plt.figure(figsize=(10, 5))
sns.barplot(x=df['State/UT'], y=df['tot_crimes_sc'])
plt.xticks(rotation='90')
plt.title('State/UT Vs Total Crimes Against SCs')
plt.savefig("4.jpg")
plt.show()
```



- States Rajasthan has the highest crimes against SCs , Andhra Pradesh, Bihar ,Karnataka,Madhya Pradesh,Uttar Pradesh has lesser Crimes against SCs

State/UT Vs Total Crimes against STs

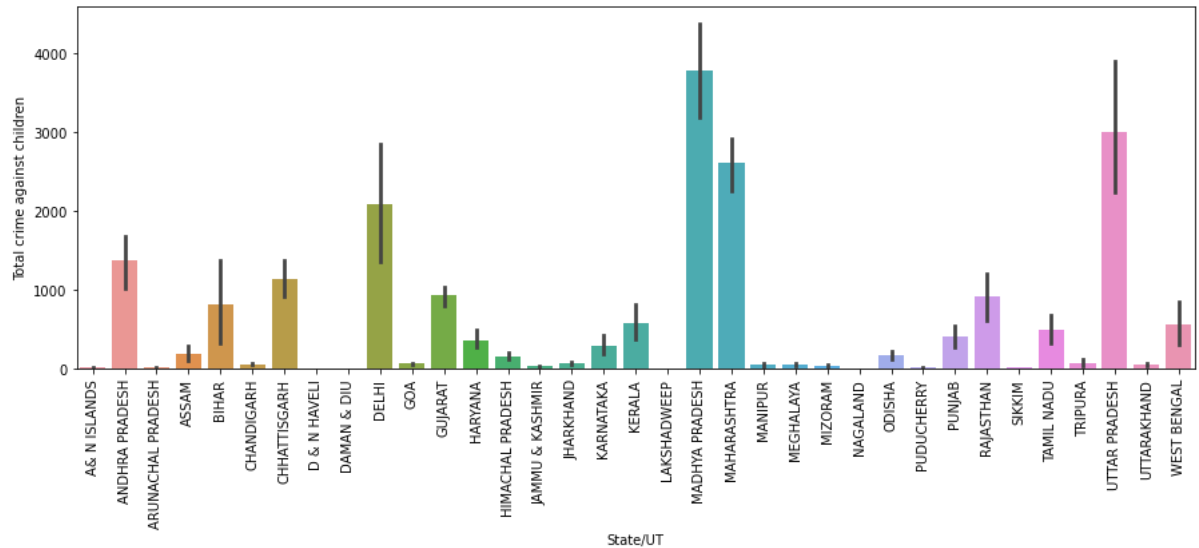
```
In [24]: plt.figure(figsize=(15, 5))
sns.barplot(x=df['State/UT'], y=df['Total crimes against STs'])
plt.xticks(rotation='90')
plt.title('State/UT Vs Total Crimes Against STs')
plt.savefig("5.jpg")
plt.show()
```



- Madhya Pradesh , rajasthan Has higher crimes against STs and Manipur, A & N Island has lowest crimes against STs.

State/UT Vs Total Crimes Against Children

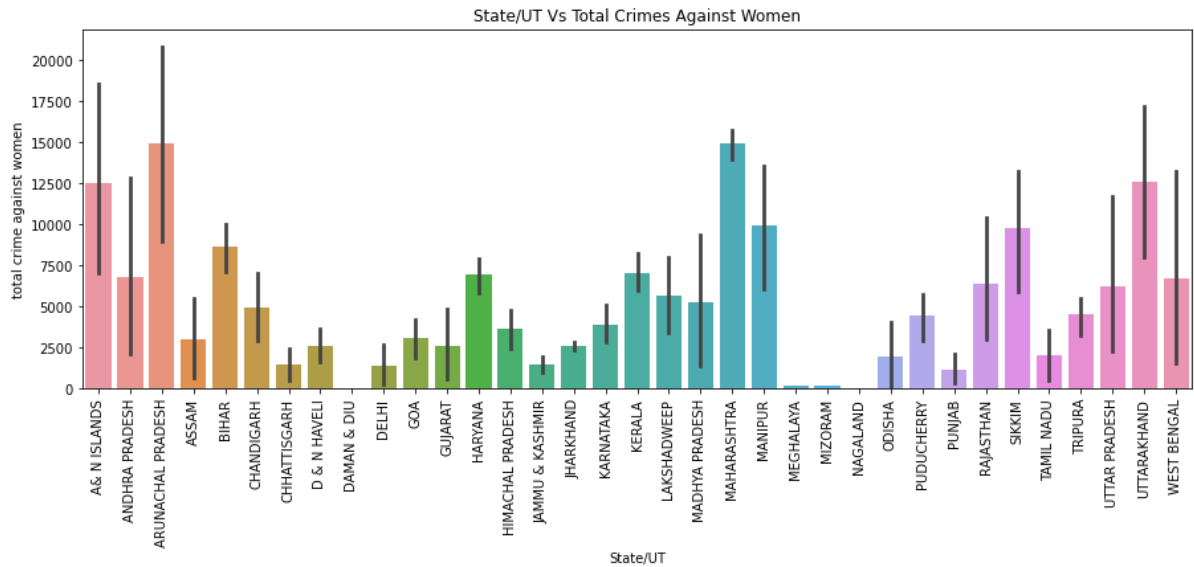
```
In [30]: plt.figure(figsize=(15, 5))
sns.barplot(x=df['State/UT'], y=df['Total crime against children'])
plt.xticks(rotation='90')
plt.title('State/UT Vs Total Crimes Against Children')
plt.savefig("6.jpg")
plt.show()
```



- MP, UP has the Highest Crimes against Children while D & N Haveli, Daman,& Diu has the lowest Crimes against Children

State/UT Vs Total Crime against Women

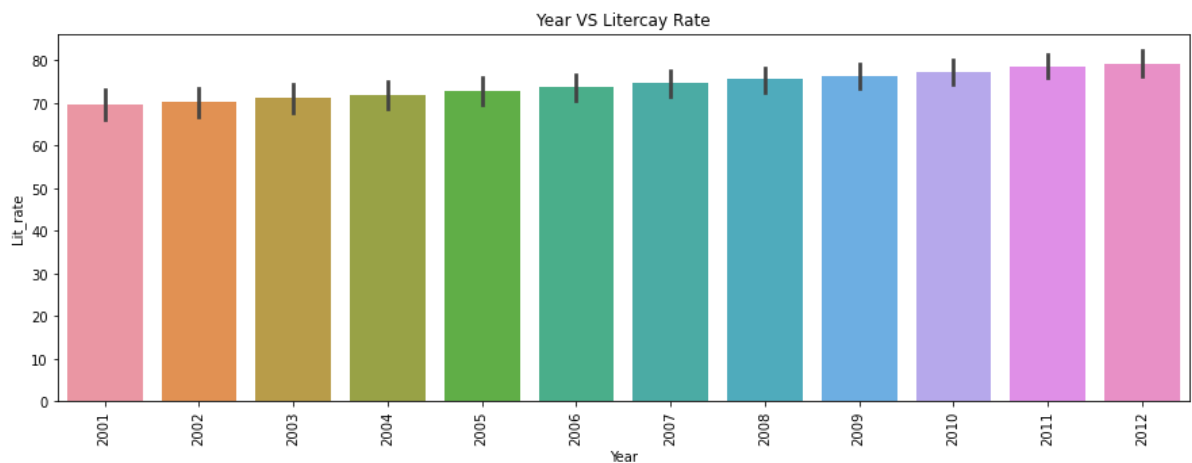
```
In [25]: plt.figure(figsize=(15, 5))
sns.barplot(x=df['State/UT'], y=df['total crime against women'])
plt.xticks(rotation='90')
plt.title('State/UT Vs Total Crimes Against Women')
plt.savefig("7.jpg")
plt.show()
```



- Only Mizoram and Meghalya are the only states which has lesser crime against women and Nagaland has almost 0 crimes against Women.
- Other states have more crimes against Women

Year vs Literacy Rate

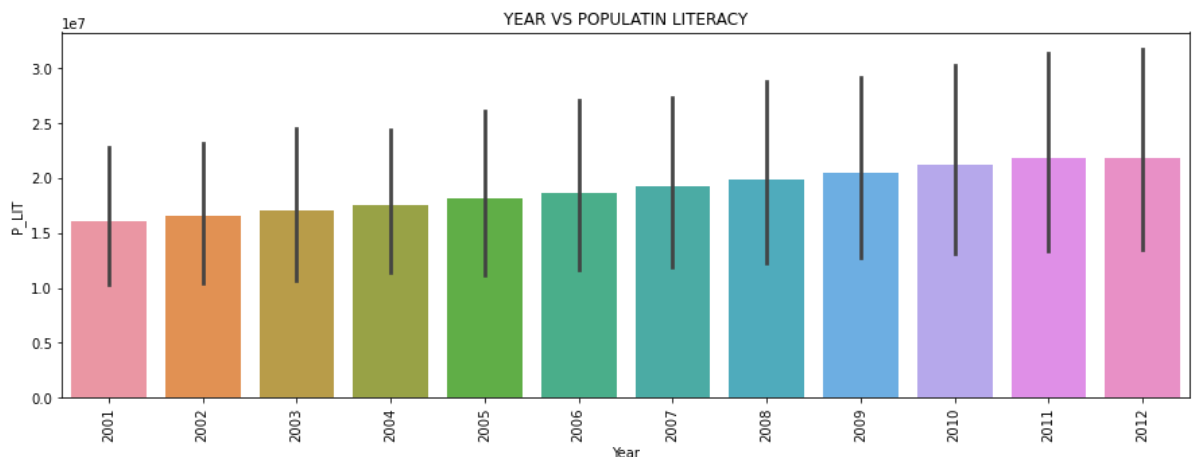
```
In [26]: plt.figure(figsize=(15, 5))
sns.barplot(x=df['Year'], y=df['Lit_rate'])
plt.xticks(rotation='90')
plt.title('Year VS Literacy Rate')
plt.savefig("8.jpg")
plt.show()
```



- Literacy rate for years has been increasing from 2001

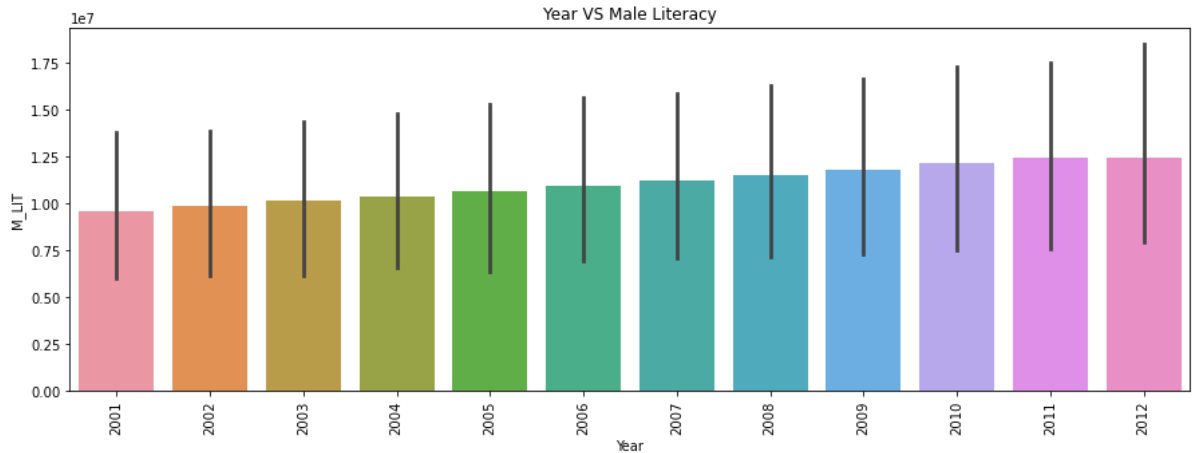
Year Vs Population Literacy

```
In [27]: plt.figure(figsize=(15, 5))
sns.barplot(x=df['Year'], y=df['P_LIT'])
plt.xticks(rotation='90')
plt.title('YEAR VS POPULATIN LITERACY')
plt.savefig("9.jpg")
plt.show()
```



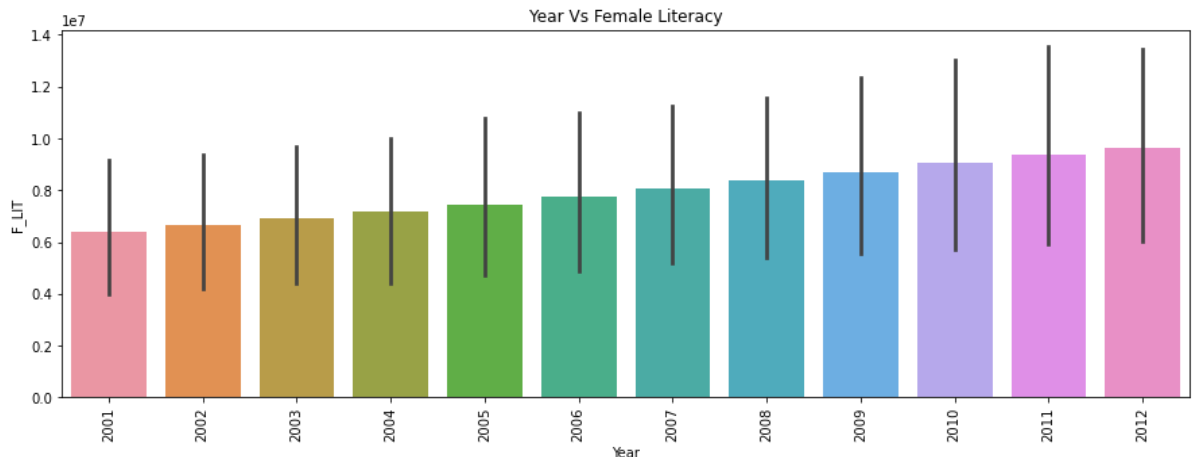
year Vs Male Literacy

```
In [28]: plt.figure(figsize=(15, 5))
sns.barplot(x=df['Year'], y=df['M_LIT'])
plt.xticks(rotation='90')
plt.title('Year VS Male Literacy')
plt.savefig("10.jpg")
plt.show()
```

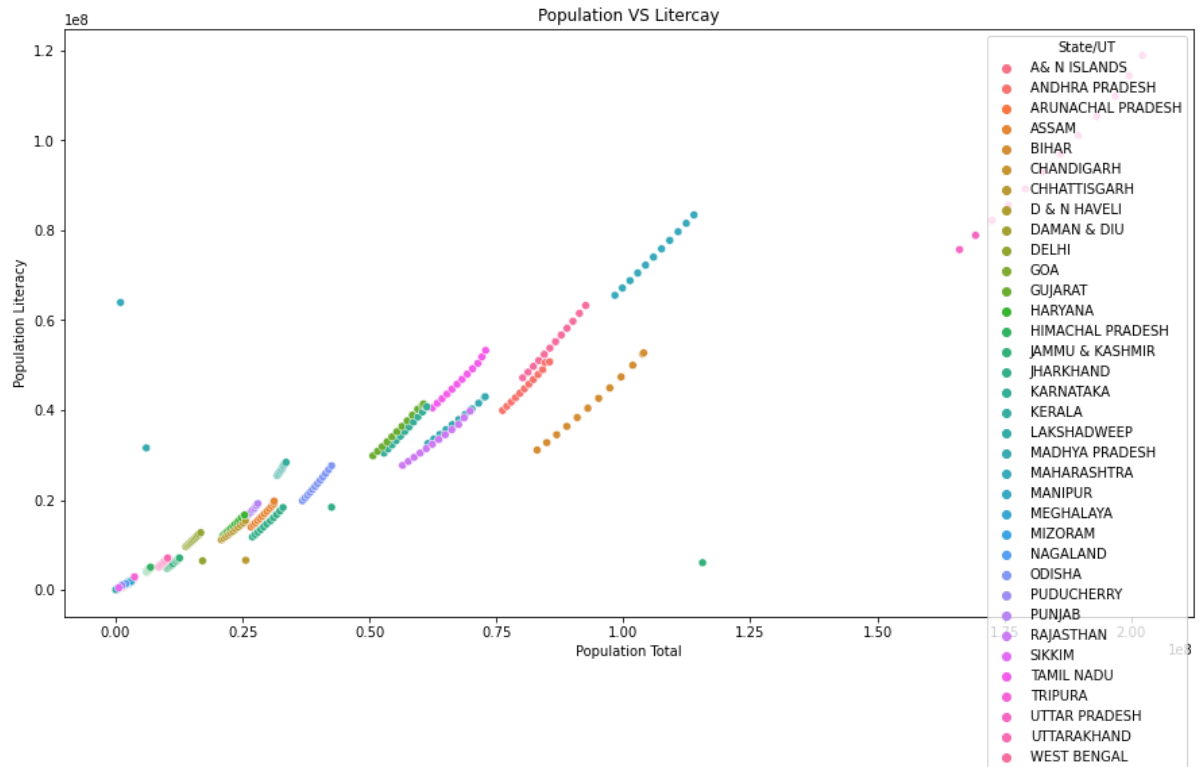


Year VS Female Literacy

```
In [29]: plt.figure(figsize=(15, 5))
sns.barplot(x=df['Year'], y=df['F_LIT'])
plt.xticks(rotation='90')
plt.title('Year Vs Female Literacy')
plt.savefig("11.jpg")
plt.show()
```




```
In [31]: plt.figure(figsize=(15, 8))
sns.scatterplot(x=df['population(total)'], y=df['P_LIT'], hue=df['State/UT'])
plt.xlabel('Population Total')
plt.ylabel('Population Literacy')
plt.title('Population VS Litercay')
plt.savefig("11.jpg")
plt.show()
```



From above scatter its clear for every state/UT , Population is increasing so does Literacy.

Exploratory Data Analysis

2.1 Analysis of Literacy Rate vs Total Crimes.

```
In [32]: new_df=df[['State/UT','Year','Lit_rate', 'Total Crimes', 'tot_crimes_sc', 'total crime against women', 'Total crimes against STs','Total crime against children']]
new_df
```

Out[32]:

	State/UT	Year	Lit_rate	Total Crimes	tot_crimes_sc	total crime against women	Total crimes against STs	Total crime against children
0	A& N ISLANDS	2001	81.30	1386	0	34.0	1.0	0.0
1	ANDHRA PRADESH	2001	60.47	278982	2933	13669.0	512.0	270.0
2	ARUNACHAL PRADESH	2001	54.34	5243	0	180.0	72.0	0.0
3	ASSAM	2001	63.25	81418	6	4048.0	0.0	18.0
4	BIHAR	2001	47.00	189612	1303	4489.0	83.0	83.0
...
415	TAMIL NADU	2012	80.88	415626	1647	6381.0	27.0	1036.0
416	TRIPURA	2012	80.89	13461	76	1559.0	29.0	20.0
417	UTTAR PRADESH	2012	67.98	444403	6202	23033.0	44.0	6033.0
418	UTTARAKHAND	2012	79.22	20122	33	1055.0	3.0	122.0
419	WEST BENGAL	2012	76.88	353008	85	30585.0	91.0	1706.0

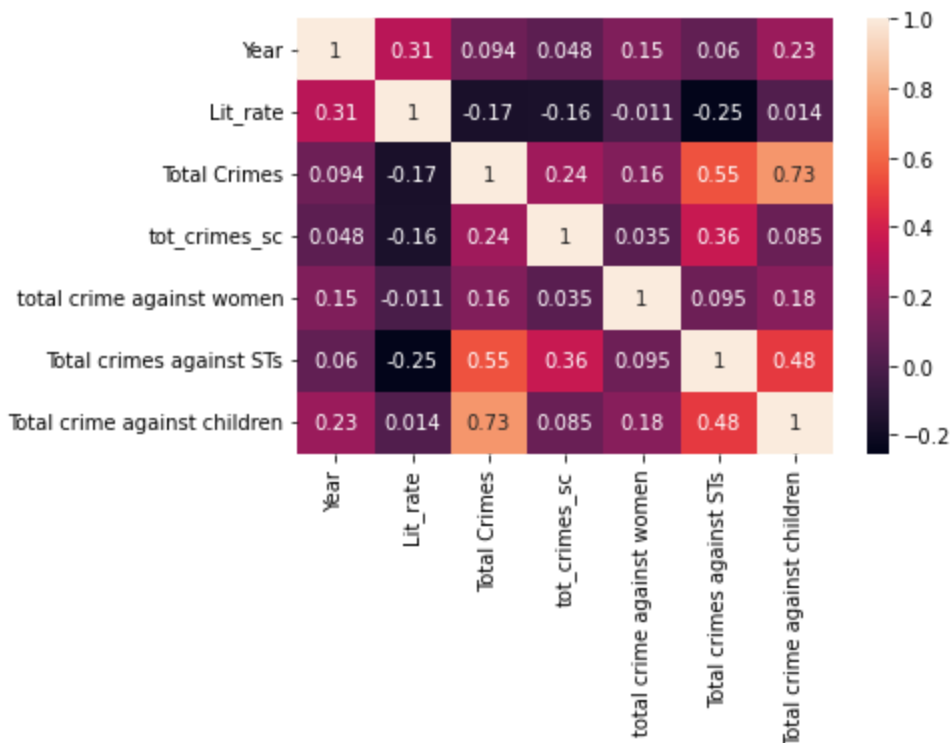
420 rows × 8 columns

```
In [33]: new_df.corr()
```

```
Out[33]:
```

	Year	Lit_rate	Total Crimes	tot_crimes_sc	total crime against women	Total crimes against STs	Total crime against children
Year	1.000000	0.314890	0.094007	0.048076	0.150851	0.060426	0.229476
Lit_rate	0.314890	1.000000	-0.167573	-0.163993	-0.011096	-0.253828	0.014062
Total Crimes	0.094007	-0.167573	1.000000	0.242294	0.155126	0.547101	0.728111
tot_crimes_sc	0.048076	-0.163993	0.242294	1.000000	0.035324	0.359817	0.084627
total crime against women	0.150851	-0.011096	0.155126	0.035324	1.000000	0.095342	0.176379
Total crimes against STs	0.060426	-0.253828	0.547101	0.359817	0.095342	1.000000	0.484496
Total crime against children	0.229476	0.014062	0.728111	0.084627	0.176379	0.484496	1.000000

```
In [35]: plt.figure(figsize=(6,4))
sns.heatmap(new_df.corr(),annot=True)
plt.savefig('heatmap1.jpg')
plt.show()
```



I see that the above heatmap shows the correlation matrix data wherein there are positive as well as negative correlations between the target label and other feature columns. A zero correlation indicates that there is no relationship between the variables. Looking at the above representation I see that quality column is positively correlated with alcohol and it is negatively

correlated with the volatile acidity. The quality column is least correlated with residual sugar showing a coefficient value of 0.014 that close to 0. Similarly we can bifurcate all the other positively and negatively correlated feature columns with respect to the target label.

Also there are some highly positive and negative correlated feature columns that can pose the concern for multicollinearity. If the correlation coefficient, assuming it to be the variable 'r', is exactly +1 or -1, then it is called perfect multicollinearity. But even if this 'r' is close to -1 or +1 then one of the features should be removed from the model if at all possible.

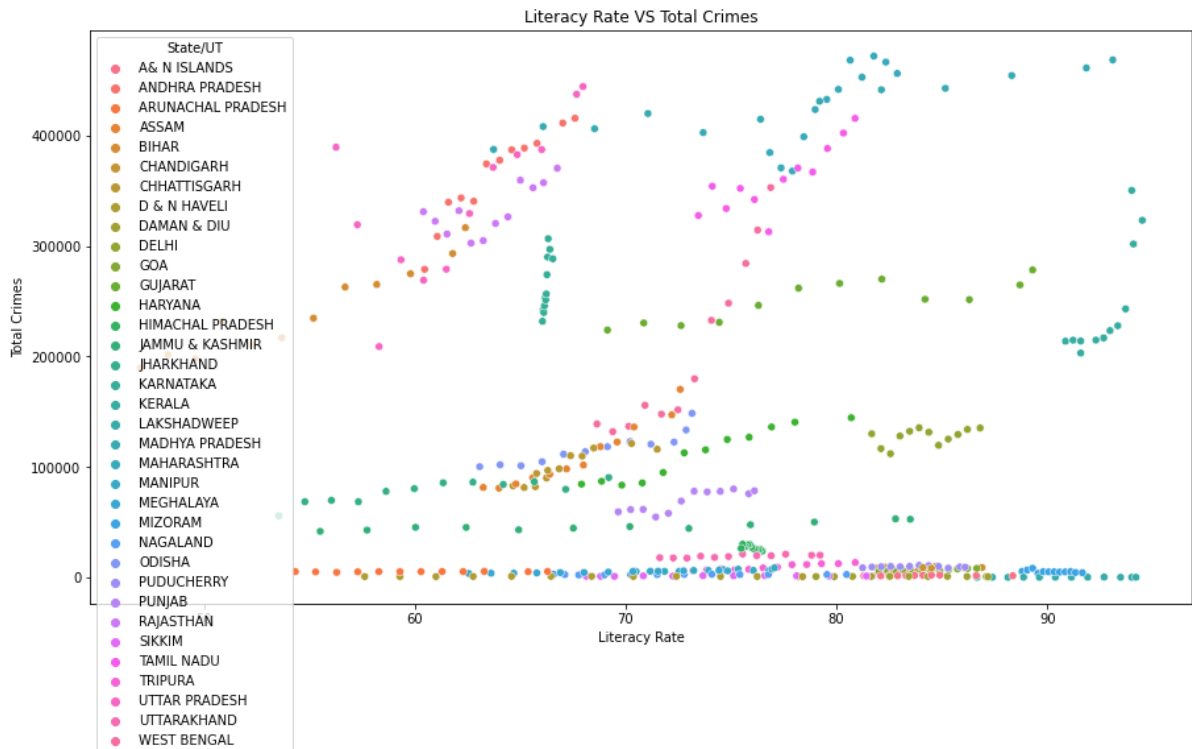
- **Lit_rate is Negatively corelated with the Tootal Crimes,, 'tot_crimes_sc', 'total crime against women','Total crimes against STs', 'Total crime against children'**

```
In [36]: import plotly.io as pio
pio.renderers.default = 'svg'

import plotly.offline as pyo
pyo.init_notebook_mode(connected=True)
```

Literacy rate VS Total Crimes

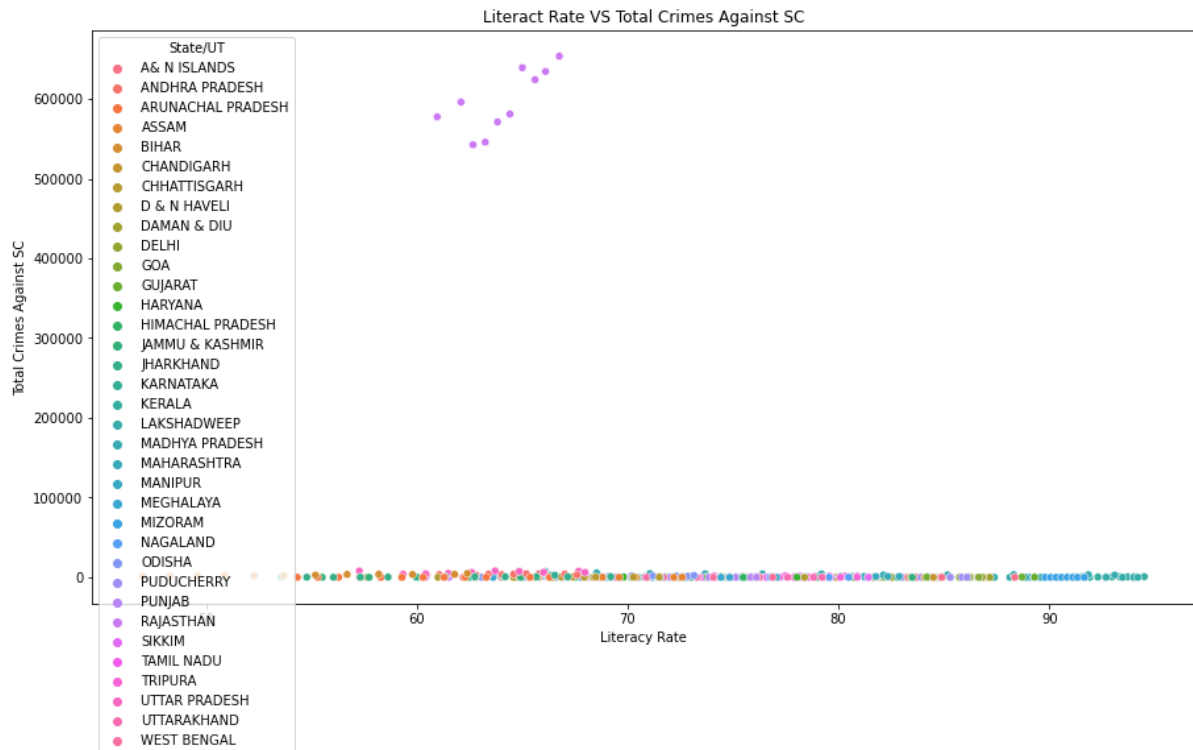
```
In [44]: #plotting scatter plot
plt.figure(figsize=(15, 8))
sns.scatterplot(x=df['Lit_rate'], y=df['Total Crimes'], hue=df['State/UT'])
plt.xlabel('Literacy Rate')
plt.ylabel('Total Crimes')
plt.title('Literacy Rate VS Total Crimes')
plt.savefig("12.jpg")
plt.show()
```



__We can see Lit_rate VS Total Crimes for evver state here

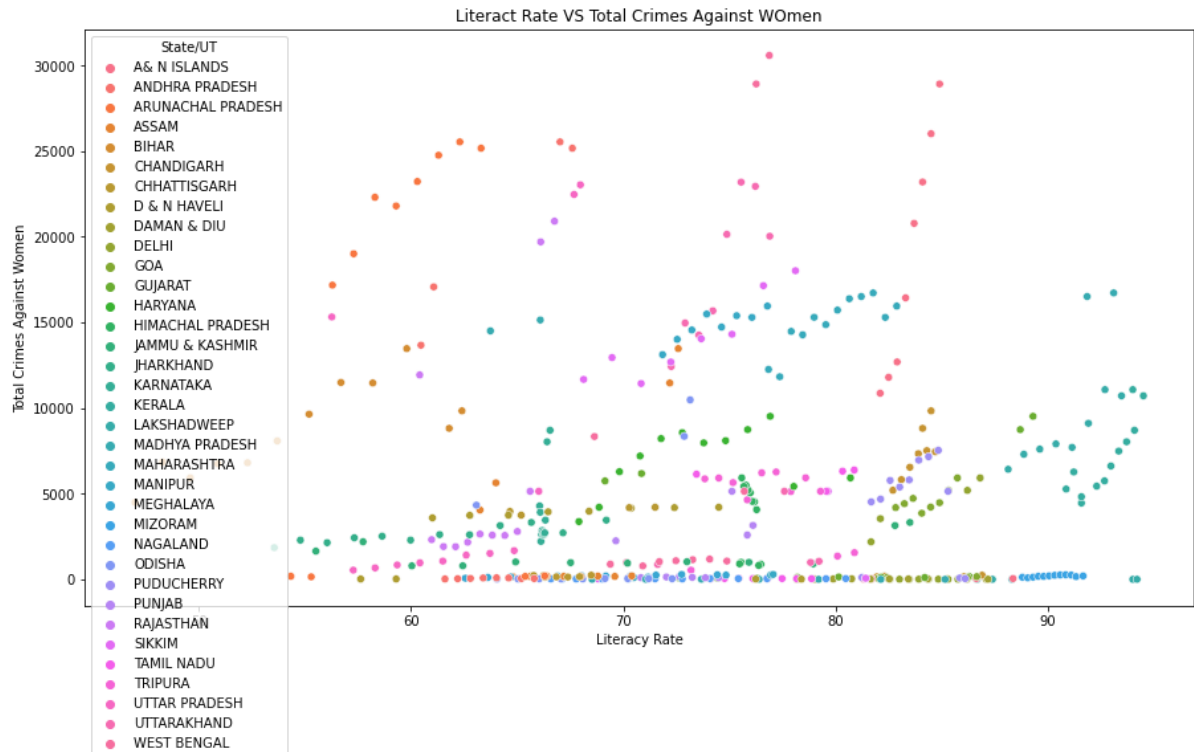
From the above scatter plot its clear that Higher the Literacy rate lower the Total Crimes would be.

```
In [45]: ## Literacy Rate VS Tot Crimes SC
plt.figure(figsize=(15, 8))
sns.scatterplot(x=df['Lit_rate'], y=df['tot_crimes_sc'], hue=df['State/UT'])
plt.xlabel('Literacy Rate')
plt.ylabel('Total Crimes Against SC')
plt.title('Literact Rate VS Total Crimes Against SC')
plt.savefig("13.jpg")
plt.show()
```



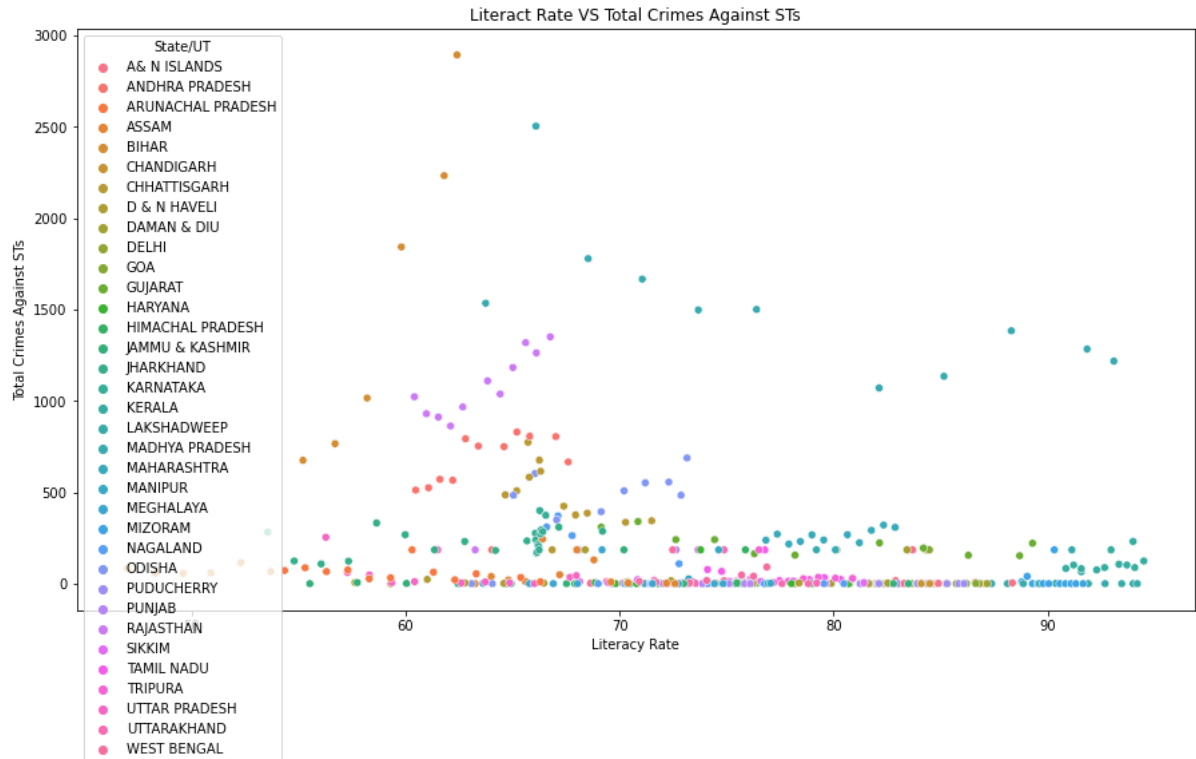
from this above scatterplot we can clearly see if Litercay Rate keep on increasing, there is decline in the tot_crimes_sc

```
In [46]: plt.figure(figsize=(15, 8))
sns.scatterplot(x=df['Lit_rate'], y=df['total crime against women'], hue=df['State'])
plt.xlabel('Literacy Rate')
plt.ylabel('Total Crimes Against Women')
plt.title('Literact Rate VS Total Crimes Against WOMen')
plt.savefig("14.jpg")
plt.show()
```



___Total crime against women kept decreasing with the increament in Literacy rate for each state

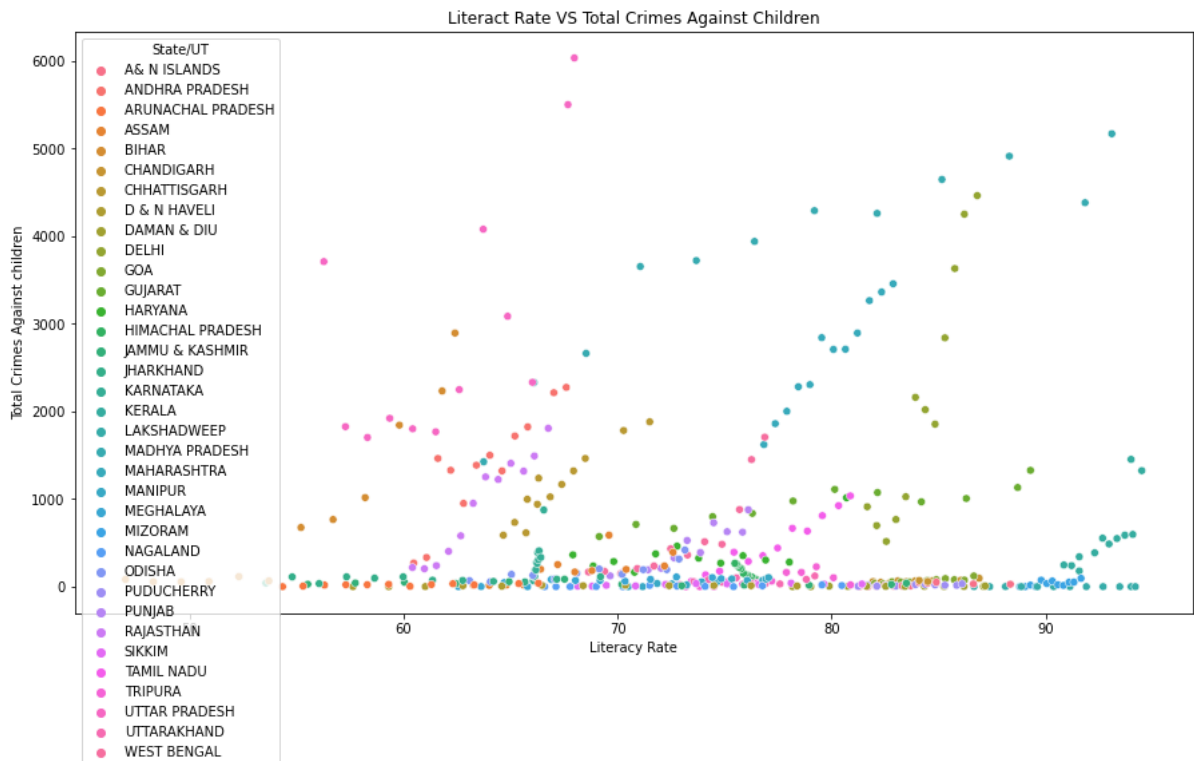
```
In [47]: plt.figure(figsize=(15, 8))
sns.scatterplot(x=df['Lit_rate'], y=df['Total crimes against STs'], hue=df['State'])
plt.xlabel('Literacy Rate')
plt.ylabel('Total Crimes Against STs')
plt.title('Literacy Rate VS Total Crimes Against STs')
plt.savefig("15.jpg")
plt.show()
```



Lit_rate is negatively correlated with Total crimes against STs.


```
In [48]: plt.figure(figsize=(15, 8))
sns.scatterplot(x=df['Lit_rate'], y=df['Total crime against children'], hue=df['State/UT'])
plt.xlabel('Literacy Rate')
plt.ylabel('Total Crimes Against children')
plt.title('Literact Rate VS Total Crimes Against Children')
plt.savefig("16.jpg")

plt.show()
```



Lit_rate is negatively correlated with Total crimes against children.

2.2 Analysis of the type of crime vs each state vs Literacy rate.

State Vs MURDER Vs Litercay Rate

```
In [58]: fig = px.scatter(df, x="Lit_rate", y="MURDER", color="State/UT", marginal_y="vi
          marginal_x="box", trendline='ols', template="simple_white")
fig.show(renderer='svg')
```

__Here we can see that as soon as State ahs higher Litercay rate rate the count of MURDER cases decreasing

EACH STATE VS LIT RATE VS ATTEMPT TO MURDER CRIME

```
In [60]: fig = px.scatter(df, x="Lit_rate", y="ATTEMPT TO MURDER", color="State/UT", marginal_x="box", trendline='ols', template="simple_white")
fig.show(renderer='svg')
```

__We can select /Deselect the State and observe the pattern here for any state/UT.

- **UTTAR PRADESH has the highest number f ATTEMPT TO MURDER CASES(7964 with LIT_rate 56.2)**
- **Lakshdweep State has 0 ATTEMPT TO MURDER CASES with Literacy 94.2 which is th highest for mentioned states**

EACH State VS RAPE VS Lit_rate

```
In [62]: fig = px.scatter(df, x="Lit_rate", y="RAPE", color="State/UT", marginal_y="viol", marginal_x="box", template="simple_white")
fig.show(renderer='svg')
```

__State MP has Highest Number of Rape cases while Lakshdweep has Lowest Rape Cases 0.

EACH State VS CULPABLE HOMICIDE NOT AMOUNTING TO MURDER VS Lit_rate

```
In [64]: fig = px.scatter(df, x="Lit_rate", y="CULPABLE HOMICIDE NOT AMOUNTING TO MURDER", marginal_x="box", template="simple_white")
fig.show(renderer='svg')
```

- **Only UP has the CULPABLE HOMICIDE NOT AMOUNTING TO MURDER cases around 1418 and states likes AP, Tripura,SIKKIM,Lakshdeep has minim CULPABLE HOMICIDE NOT AMOUNTING TO MURDER like 3,6,0 etc**

Each state VS Lit rate VS Type of CRIME(CUSTODIAL RAPE)

```
In [66]: fig = px.scatter(df, x="Lit_rate", y="CUSTODIAL RAPE", color="State/UT", marginal_x="box", template="simple_white")
fig.show(renderer='svg')
```

- **__ Kerala with 94 Litracay rate has 0 Custodial rape__**
- **__ Assam with 70 Litracay Rate has 5 Custodial Rape which was MAXimum in the category of Custodial Rape__**

Custodial rapes are those which are committed by the public servant on the premises. We know that prosecuting the public servant is very hard in our nation. In the country where Government stops from lodging the report to suppress the crime, in that nation it's hard to report the case. That's why these are very less reported hence showing less numbers.

Each State VS Lit Rate VS Type of crime(KIDNAPPING & ABDUCTION)

```
In [68]: fig = px.scatter(df, x="Lit_rate", y="KIDNAPPING & ABDUCTION", color="State/UT",
                        marginal_x="box", template="simple_white")
fig.show(renderer='svg')
```

- Top 3 states for KIDNAPPING & ABDUCTION cases are Uttar Pradesh, West Bengal and BIHAR
- Top 2 states with lowest number of auto theft cases Sikkim, Mizoram.

Each state vs Lit rate VS type of Crime(KIDNAPPING AND ABDUCTION OF WOMEN AND GIRLS)

```
In [70]: fig = px.scatter(df, x="Lit_rate", y="KIDNAPPING AND ABDUCTION OF WOMEN AND GIRLS",
                        marginal_x="box", template="simple_white")
fig.show(renderer='svg')
```

- Uttar Pradesh has highest number of KIDNAPPING AND ABDUCTION OF WOMEN AND GIRLS cases 7910 and Lakshdeep has 0 KIDNAPPING AND ABDUCTION OF WOMEN AND GIRLS cases

Each state vs Lit rate VS type of crime(DACOITY)

```
In [73]: fig = px.scatter(df, x="Lit_rate", y="DACOITY", color="State/UT", marginal_y="box",
                        marginal_x="box", template="simple_white")
fig.show(renderer='svg')
```

- Bihar, UTTAR PRADESH, and MAharashtra are the top 3 states with highest Daocity cases.
- Tripura, Sikkim, MANipur has the lowest daocity cases..

Each state VS Lit rate VS Type of Crime(ROBBERY)

```
In [75]: fig = px.scatter(df, x="Lit_rate", y="ROBBERY", color="State/UT", marginal_y="box",
                        marginal_x="box", template="simple_white")
fig.show(renderer='svg')
```

- Maharashtra state has 6949 cases and Lakshdweep has 0 Robbery cases

Each state VS Lit Rate VS Type of Crime(BURGLARY)

```
In [77]: fig = px.scatter(df, x="Lit_rate", y="BURGLARY", color="State/UT", marginal_y='vic
          marginal_x="box", template="simple_white")
fig.show(renderer='svg')
```

- Maharashtra has around 15000 cases of BURGLARY and Lakshdweep has 3 BURGLARY cases

Each State VS Lit rate VS Type of Crime(THEFT)

```
In [79]: fig = px.scatter(df, x="Lit_rate", y="THEFT", color="State/UT", marginal_y="vic
          marginal_x="box", template="simple_white")
fig.show(renderer='svg')
```

- __ Maharashtra has around 50K (highest)cases of THEFT and Lakshdweep has 7(Lowest) cases of THEFT.

Each State VSLit rate VS Type of Crime(AUTO THEFT)

```
In [81]: fig = px.scatter(df, x="Lit_rate", y="AUTO THEFT", color="State/UT", marginal_y
          marginal_x="box", template="simple_white")
fig.show(renderer='svg')
```

- __ Uttar Pradesh has 22000(highest)AUTO THEFT cases where else Lakshdweep has 1 AUTO THEFT caseswhich is the Lowest

Each state VS Lit rate VS Type of Crime(RIOTS)

```
In [83]: fig = px.scatter(df, x="Lit_rate", y="RIOTS", color="State/UT", marginal_y="vic
          marginal_x="box", template="simple_white")
fig.show(renderer='svg')
```

-__ Mizoram(91.33 Literacy rate) has 0 (Lowest) Riots cases while the Rajasthan(Lit rate 60.1) has 11.2K(highest) Cases of Riots which is the Highest

Each State VS Literacy Rate VS Type of Crime(CRIMINAL BREACH OF TRUST)

```
In [85]: fig = px.scatter(df, x="Lit_rate", y="CRIMINAL BREACH OF TRUST", color="State/UT",
                        marginal_x="box", template="simple_white")
fig.show(renderer='svg')
```

- UP has CRIMINAL BREACH OF TRUST cases around 4296(highest) and Lakshdweep has 0 CRIMINAL BREACH OF TRUST

Each State VS Lit rate VS Type of Crime(CHEATING)

```
In [87]: fig = px.scatter(df, x="Lit_rate", y="CHEATING", color="State/UT", marginal_y='vic',
                        marginal_x="box", template="simple_white")
fig.show(renderer='svg')
```

- __Rajasthan has highest cases of Cheating 19.46K and Lakshdweep has lowest cases 0

Each state VS Lit rate vs Type of Crime(COUNTERFEITING)

```
In [89]: fig = px.scatter(df, x="Lit_rate", y="COUNTERFEITING", color="State/UT", marginal_y='vic',
                        marginal_x="box", template="simple_white")
fig.show(renderer='svg')
```

- State Tamil Nadu has highest numbers of COUNTERFEITING cases 599 and Lakshdweep has 0 COUNTERFEITING cases

Each state VS Lit rate vs Type of Crime(ARSON)

```
In [91]: fig = px.scatter(df, x="Lit_rate", y="ARSON", color="State/UT", marginal_y="vic",
                        marginal_x="box", template="simple_white")
fig.show(renderer='svg')
```

- State Assam has highest numbers of ARSON cases 2830 and Lkshdweep has 3 ARSON cases

Each state VS Lit rate vs Type of Crime(HURT/GREIVIOUS HURT)

```
In [93]: fig = px.scatter(df, x="Lit_rate", y="HURT/GREIVIOUS HURT", color="State/UT", marginal_x="box", template="simple_white")
fig.show(renderer='svg')
```

- state Tamil Nadu has highest numbers of HURT/GREIVIOUS HURT cases 56.7K and Lakshdweep has 1 HURT/GREIVIOUS HURT cases

Each state VS Lit rate vs Type of Crime(DOWRY DEATHS)

```
In [95]: fig = px.scatter(df, x="Lit_rate", y="DOWRY DEATHS", color="State/UT", marginal_x="box", template="simple_white")
fig.show(renderer='svg')
```

- State Assam has highest numbers of Dowry Death 2322 cases 2830 and Lkshdweep has 0 Dowry Death cases

Each state VS Lit rate vs Type of Crime(INSULT TO MODESTY OF WOMEN)

```
In [97]: fig = px.scatter(df, x="Lit_rate", y="INSULT TO MODESTY OF WOMEN", color="State/UT", marginal_x="box", template="simple_white")
fig.show(renderer='svg')
```

- State Uttar Pradesh has highest numbers of Insult to Modest of women 4970 cases 2830 and Lkshdweep has 0 Insult to Modest of women cases

Each state VS Lit rate vs Type of Crime(CRUELTY BY HUSBAND OR HIS RELATIVES)

```
In [99]: fig = px.scatter(df, x="Lit_rate", y="CRUELTY BY HUSBAND OR HIS RELATIVES", color="State/UT", marginal_x="box", template="simple_white")
fig.show(renderer='svg')
```

- State West Bengal has highest numbers of CRUELTY BY HUSBAND OR HIS RELATIVES 19.36K cases and Lakshdweep has 0 CRUELTY BY HUSBAND OR HIS RELATIVES

Each state VS Lit rate vs Type of Crime(IMPORTATION OF GIRLS FROM FOREIGN COUNTRIES)

```
In [101]: fig = px.scatter(df, x="Lit_rate", y="IMPORTATION OF GIRLS FROM FOREIGN COUNTRIES",
                        marginal_x="box", template="simple_white")
fig.show(renderer='svg')
```

- State Bihar has 83 (highest) numbers of IMPORTATION OF GIRLS FROM FOREIGN COUNTRIES cases 2830 and Lakshdweep,Kerala,Mizoram have 0 IMPORTATION OF GIRLS FROM FOREIGN COUNTRIES cases

Each state VS Lit rate vs Type of Crime(CAUSING DEATH BY NEGLIGENCE)

```
In [103]: fig = px.scatter(df, x="Lit_rate", y="CAUSING DEATH BY NEGLIGENCE", color="State/UT",
                        marginal_x="box", template="simple_white")
fig.show(renderer='svg')
```

- State Tamil Nadu has 16.07K (highest) numbers of CAUSING DEATH BY NEGLIGENCE cases and Lakshdweep have 0 CAUSING DEATH BY NEGLIGENCE cases

Each state VS Lit rate vs Type of Crime(OTHER IPC CRIMES)

```
In [105]: fig = px.scatter(df, x="Lit_rate", y="OTHER IPC CRIMES", color="State/UT", marginal_x="box",
                        template="simple_white")
fig.show(renderer='svg')
```

-State Tamil Nadu has 124.17K (highest) numbers of OTHER IPC Crimes cases and Lakshdweep have 5 OTHER IPC CRIMES cases

From Above analysis of each type of crime and Literacy rate and State/UT it is some how clear that Lakshdweep is the only UT which has Lowest crimes and Literacy rate is Quite good like its less than Kerla but Higher than most of the States/UT and has amost 0 crime rate .

2.3 Analysis of year-on-year total crime rate.

- Crime rate= e following formula is used to calculate a crime rate
- $CR = \frac{TC}{P} \times 100,000$
- Where CR is the rate of crimes per 100,000 people
- TC is the total number of crimes committed in an area

- P is the total population of the area

```
In [22]: #selecting d/f states/UT for analysis
df_states=df[['State/UT','Year','Crime Rate']]
```

```
In [23]: # Making data set for different state
#Selecting A& N Island only
select_andman = df_states.loc[df['State/UT'] == 'A& N ISLANDS']
select_andhra = df_states.loc[df['State/UT'] == 'ANDHRA PRADESH']
select_arunachal = df_states.loc[df['State/UT'] == 'ARUNACHAL PRADESH']
select_assam4 = df_states.loc[df['State/UT'] == 'ASSAM']
select_bihar = df_states.loc[df['State/UT'] == 'BIHAR']
select_chandigarh = df_states.loc[df['State/UT'] == 'CHANDIGARH']
select_chhatisgarh = df_states.loc[df['State/UT'] == 'CHHATTISGARH']
select_haveli = df_states.loc[df['State/UT'] == 'D & N HAVELI']
select_daman = df_states.loc[df['State/UT'] == 'DAMAN & DIU']
select_delhi = df_states.loc[df['State/UT'] == 'DELHI']
select_goa = df_states.loc[df['State/UT'] == 'GOA']
select_gujarat = df_states.loc[df['State/UT'] == 'GUJARAT']
select_haryana = df_states.loc[df['State/UT'] == 'HARYANA']
select_himachal = df_states.loc[df['State/UT'] == 'HIMACHAL PRADESH']
select_jk = df_states.loc[df['State/UT'] == 'JAMMU & KASHMIR']
select_jharkhand = df_states.loc[df['State/UT'] == 'JHARKHAND']
select_karnataka = df_states.loc[df['State/UT'] == 'KARNATAKA']
select_kerala= df_states.loc[df['State/UT'] == 'KERALA']
select_lakshdweep= df_states.loc[df['State/UT'] == 'LAKSHADWEEP']
select_mp= df_states.loc[df['State/UT'] == 'MADHYA PRADESH']
select_maharashtra = df_states.loc[df['State/UT'] == 'MAHARASHTRA']
select_manipur= df_states.loc[df['State/UT'] == 'MANIPUR']
select_meghalaya = df_states.loc[df['State/UT'] == 'MEGHALAYA']
select_mizoram = df_states.loc[df['State/UT'] == 'MIZORAM']
select_nagaland= df_states.loc[df['State/UT'] == 'NAGALAND']
select_odisha= df_states.loc[df['State/UT'] == 'ODISHA']
select_puducherry = df_states.loc[df['State/UT'] == 'PUDUCHERRY']
select_punjab = df_states.loc[df['State/UT'] == 'PUNJAB']
select_rajasthan = df_states.loc[df['State/UT'] == 'RAJASTHAN']
select_sikkim = df_states.loc[df['State/UT'] == 'SIKKIM']
select_tamilnadu = df_states.loc[df['State/UT'] == 'TAMIL NADU']
select_tripura= df_states.loc[df['State/UT'] == 'TRIPURA']
select_up = df_states.loc[df['State/UT'] == 'UTTAR PRADESH']
select_uk= df_states.loc[df['State/UT'] == 'UTTARAKHAND']
select_wb= df_states.loc[df['State/UT'] == 'WEST BENGAL']
```

Crime Rate Analysis of A & Islands year wise

```
In [25]: fig = px.line(select_andman, x='Year', y='Crime Rate', text='Crime Rate', markers=True)
fig.show(renderer='svg')
```

- As per above line plot crime rate was lowest in 2002 and highest was in 2010 for A & N Islands

Crime Rate Analysis of Andhra Pradesh year wise

```
In [26]: fig = px.line(select_andhra, x='Year', y='Crime Rate', text='Crime Rate', marker=fig.show(renderer='svg'))
```

- For Andhra Pradesh Crime rate was lowest in 2001 and highest was in 2012

Crime Rate Analysis of Arunachal Pradesh year wise

```
In [27]: fig = px.line(select_arunachal, x='Year', y='Crime Rate', text='Crime Rate', marker=fig.show(renderer='svg'))
```

As per above Line plot Arunachal Pradesh has Lowest crime rate in 2012 and highest in 2001 which is a good thing that means Crime is decreasing in this state.

Crime Rate Analysis of Assam year wise

```
In [28]: fig = px.line(select_assam4, x='Year', y='Crime Rate', text='Crime Rate', marker=fig.show(renderer='svg'))
```

As per above Line plot Assam has Hishest crime rate in 2012 and Lowest in 2001

Crime Rate Analysis of Bihar year wise

```
In [29]: fig = px.line(select_bihar, x='Year', y='Crime Rate', text='Crime Rate', marker=fig.show(renderer='svg'))
```

It is clear from above plot that Bihar has Increasing trend for Crime rate that means Crime rate is keep on increasing from 2001

```
In [30]: fig = px.line(select_chandigarh, x='Year', y='Crime Rate', text='Crime Rate', marker=fig.show(renderer='svg'))
```

Chandigarh Crime kept fluctuating from 2001 , Crime rate was highest in 2002 and lowest in 2003

Crime Rate Analysis of Chhatisgarh year wise

```
In [31]: fig = px.line(select_chhatisgarh, x='Year', y='Crime Rate', text='Crime Rate',
fig.show(renderer='svg')
```

Chhatisgarh has Lowest crime rate recorded in 2003 and highest was in 2011

Crime Rate Analysis of D & N Haveli year wise

```
In [32]: fig = px.line(select_haveli, x='Year', y='Crime Rate', text='Crime Rate', marker
fig.show(renderer='svg')
```

Highest Crime rate was recorded in 2005 and lowest recorded in 2012

Crime Rate Analysis of Daman & Diu year wise

```
In [33]: fig = px.line(select_daman, x='Year', y='Crime Rate', text='Crime Rate', marker
fig.show(renderer='svg')
```

Daman & Diu Crime rate was highest in 2005 since then its decreasing till 2012.

Crime Rate Analysis of Delhi year wise

```
In [34]: fig = px.line(select_delhi, x='Year', y='Crime Rate', text='Crime Rate', marker
fig.show(renderer='svg')
```

Delhi Crime rate was at peak in 2001 and its fluctuated till 2008 and it got decreased in 2010

Crime Rate Analysis of GOA year wise

```
In [35]: fig = px.line(select_goa, x='Year', y='Crime Rate', text='Crime Rate', markers=
fig.show(renderer='svg')
```

Goa Crime rate was lowest in 2004 and Highest recorded in 2012

Crime Rate Analysis of GUJARAT year wise

```
In [36]: fig = px.line(select_gujarat, x='Year', y='Crime Rate', text='Crime Rate', markers=True)
fig.show(renderer='svg')
```

Gujarat has highest crime rate was in 2007 and lowest crime rate was in 2010

Crime Rate Analysis of HARYANA year wise

```
In [37]: fig = px.line(select_haryana, x='Year', y='Crime Rate', text='Crime Rate', markers=True)
fig.show(renderer='svg')
```

Crime rate of Haryana keep on increasing since 2004.

Crime Rate Analysis of HIMACHAL PRADESH year wise

```
In [38]: fig = px.line(select_himachal, x='Year', y='Crime Rate', text='Crime Rate', markers=True)
fig.show(renderer='svg')
```

Lowest Crime rate For Himachal Pradesh was recorded in 2012 and highest was in 2007

Crime Rate Analysis of JAMMU & KASHMIR year wise

```
In [39]: fig = px.line(select_jk, x='Year', y='Crime Rate', text='Crime Rate', markers=True)
fig.show(renderer='svg')
```

For Jammu & Kashmir 2007 was only year when there were less crime

Crime Rate Analysis of JHARKHAND year wise

```
In [40]: fig = px.line(select_jharkhand, x='Year', y='Crime Rate', text='Crime Rate', markers=True)
fig.show(renderer='svg')
```

Lowest Crime rate was in 2001 and Highest was in 2007.

Crime Rate Analysis of Karnataka year wise

```
In [41]: fig = px.line(select_karnataka, x='Year', y='Crime Rate', text='Crime Rate', markers=True)
fig.show(renderer='svg')
```

For Karnataka 2001 had less crime rate and year 2010 highest crime rate.

Crime Rate Analysis of Kerala year wise

```
In [42]: fig = px.line(select_kerala, x='Year', y='Crime Rate', text='Crime Rate', markers=True)
fig.show(renderer='svg')
```

Crime rate was lower till 2009 and then it got increased and highest recorded in 2011

Lakshdweep Crime rate yearly analysis

```
In [43]: fig = px.line(select_lakshdweep, x='Year', y='Crime Rate', text='Crime Rate', markers=True)
fig.show(renderer='svg')
```

Lakshdweep was less Crime prone in 2001 and was highest in 2007

For MadhyaPradesh Crime Rate Analysis Yearly

```
In [44]: fig = px.line(select_mp, x='Year', y='Crime Rate', text='Crime Rate', markers=True)
fig.show(renderer='svg')
```

MP State has highest crime rate in 2001 and from 2002 it decreased.

Maharashtra Crime Rate Analysis Year wise

```
In [45]: fig = px.line(select_maharashtra, x='Year', y='Crime Rate', text='Crime Rate', markers=True)
fig.show(renderer='svg')
```

Maharashtra Had lowest crime rate since 2002 till 2012

Manipur Crime rate Analysis year wise

```
In [46]: fig = px.line(select_manipur, x='Year', y='Crime Rate', text='Crime Rate', markers=True)
fig.show(renderer='svg')
```

Manipur has lowest Crime rate in 2010 and highest in 2008

Meghalya Crime rate Analysis Year wise

```
In [47]: fig = px.line(select_meghalaya, x='Year', y='Crime Rate', text='Crime Rate', markers=True)
fig.show(renderer='svg')
```

Meghayala state had lowest crime rate in 2003 and highest was in 2011

Mizoram Crime Rate Analysis Year wise

```
In [48]: fig = px.line(select_mizoram, x='Year', y='Crime Rate', text='Crime Rate', markers=True)
fig.show(renderer='svg')
```

Mizoram lowest crime rate was in 2012 and highest was in 2003

Nagaland Crime Rate Analysis Year wise

```
In [49]: fig = px.line(select_nagaland, x='Year', y='Crime Rate', text='Crime Rate', markers=True)
fig.show(renderer='svg')
```

Nagaland crime rate was lowest in 2003 and highest was in 2001

Odisha Crime rate Analysis year wise

```
In [51]: fig = px.line(select_odisha, x='Year', y='Crime Rate', text='Crime Rate', markers=True)
fig.show(renderer='svg')
```

__Odisha crime rate was highest in 2012 and lowest was in 2003

PuduCherry Crime rate Analysis

```
In [52]: fig = px.line(select_puducherry, x='Year', y='Crime Rate', text='Crime Rate', markers=True)
fig.show(renderer='svg')
```

__Puducherry had lowest crime rate was in 2010 and highest was in 2007

Punjab Crime rate Analysis Year wise

```
In [53]: fig = px.line(select_punjab, x='Year', y='Crime Rate', text='Crime Rate', markers=True)
fig.show(renderer='svg')
```

__Crime rate of Punjab was lowest in 2004 and highest was in 2007

Rajasthan Crime rate Analysis Year wise

```
In [54]: fig = px.line(select_rajasthan, x='Year', y='Crime Rate', text='Crime Rate', markers=True)
fig.show(renderer='svg')
```

__Crime rate of Rajasthan was in 2005 and highest was in 2001

Sikkim Crime Rate Analysis Year wise

```
In [55]: fig = px.line(select_sikkim, x='Year', y='Crime Rate', text='Crime Rate', markers=True)
fig.show(renderer='svg')
```

__Crime rate of Sikkim was in 2003 and highest was in 2008.

Tamil Nadu Crime Rate Analyss year wise

```
In [56]: fig = px.line(select_tamilnadu, x='Year', y='Crime Rate', text='Crime Rate', markers=True)
fig.show(renderer='svg')
```

__Tamil Nadu crime rate was highest in 2012 and lowest was in 2006

Tripura Crime rate analysis Yearwise

```
In [57]: fig = px.line(select_tripura, x='Year', y='Crime Rate', text='Crime Rate', markers=True)
fig.show(renderer='svg')
```

Crime rate of Tripura was in 2001 lowest and highest was in 2012

Uttar Pradesh Crime rate Analysis Year wise

```
In [58]: fig = px.line(select_up, x='Year', y='Crime Rate', text='Crime Rate', markers=True)
fig.show(renderer='svg')
```

__Uttar Pradesh had lowest crime rate was in 2003 and highest was in 2001

Uttarakhand Crime Rate Analysis Year wise

```
In [59]: fig = px.line(select_uk, x='Year', y='Crime Rate', text='Crime Rate', markers=True)
fig.show(renderer='svg')
```

__Uttarakhand state had lowest crime rate was in 2012 and highest was in 2007

West Bengal Crime Rtae Analysis Year wise

```
In [60]: fig = px.line(select_wb, x='Year', y='Crime Rate', text='Crime Rate', markers=True)
fig.show(renderer='svg')
```

Crime rate of West Bengal keep on increasing since 2006 and 2002 was the year when crime rate was lowest

2.4 Analysis of area vs overall crime

In [65]:

df_area=df.groupby('Area (km2)').min()

In [66]:

dfarea=pd.pivot_table(df, values ='Area (km2)', index =['State/UT', 'Year', 'Total Crimes'],aggfunc = max)
dfarea

Out[66]:

			Area (km2)
State/UT	Year	Total Crimes	
A& N ISLANDS	2001	1386	8249
	2002	1269	8249
	2003	1352	8249
	2004	1612	8249
	2005	1462	8249
...
WEST BENGAL	2008	232767	88752
	2009	248291	88752
	2010	284277	88752
	2011	314488	88752
	2012	353008	88752

420 rows × 1 columns

```
In [67]: df_area=df[['State/UT', 'Year', 'Total Crimes', 'Area (km2)']]
df_area
```

```
Out[67]:
```

	State/UT	Year	Total Crimes	Area (km2)
0	A& N ISLANDS	2001	1386	8249
1	ANDHRA PRADESH	2001	278982	275045
2	ARUNACHAL PRADESH	2001	5243	83743
3	ASSAM	2001	81418	78438
4	BIHAR	2001	189612	94163
...
415	TAMIL NADU	2012	415626	130058
416	TRIPURA	2012	13461	10486
417	UTTAR PRADESH	2012	444403	240928
418	UTTARAKHAND	2012	20122	53483
419	WEST BENGAL	2012	353008	88752

420 rows × 4 columns

```
In [68]: fig = px.scatter(df_area, x="Area (km2)", y="Total Crimes", size="Area (km2)",
                        size_max=50)
fig.show('svg')
```


Above Scatter plot is showing us the Area VS Total Crimes. Since the Area will remain the same so the size of the circle are equal for every state Area Correspondingly. Different Color showing the different State/UT and Total Crimes on y-axis.

Observations

- From above representaiton its is clear that smaller the Area of the state lesser Total Crimes would be. For example Lakshdweep Area is 72 KM2 and Total crimes are only 127 which is showing afetr hovering over the Lakshdweep circle
- Similarly with Puducherry state Area-479km2 and Total Crimes 9252.
- The Area of Rajasthan state is 342239(Largest) Km2 and Total Crime are like 370.502K
- While the Uttar Pradesh has 240928 km2 area and Total Crimes are 444.403K Total Crimes which is the Highest

2.5 Analysis of Population vs overall Crime

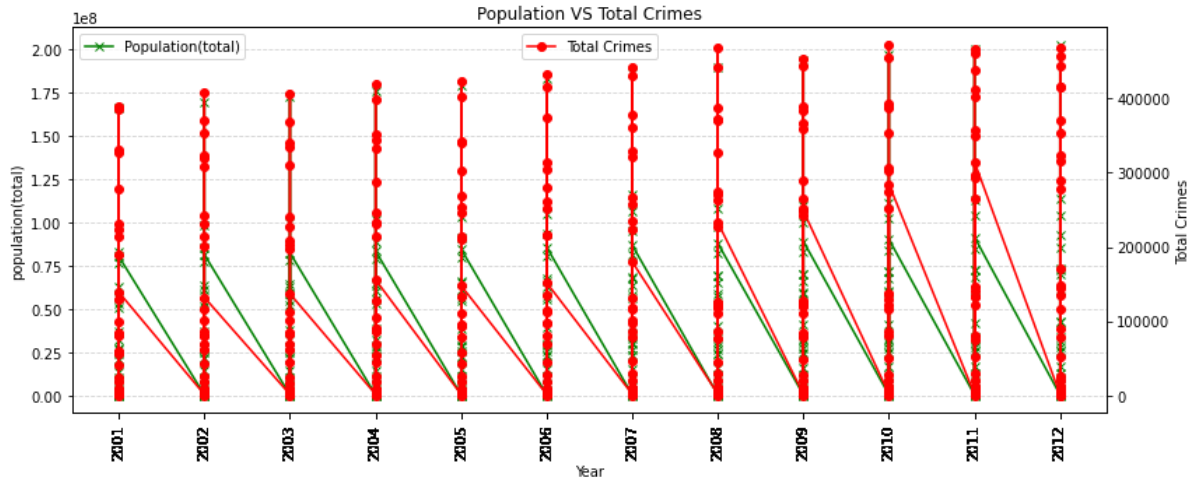
```
In [106]: fig, ax = plt.subplots(figsize=(12,5))
ax2 = ax.twinx()
ax.set_title('Population VS Total Crimes')
ax.set_xlabel('Year')

ax.plot(df['Year'],df['population(total)'], color='green', marker='x')
ax2.plot(df['Year'],df['Total Crimes'], color='red', marker='o')

ax.set_ylabel('population(total)')
ax2.set_ylabel('Total Crimes')
ax.legend(['Population(total)'])
ax2.legend(['Total Crimes'], loc='upper center')
ax.set_xticks(df['Year'])
ax.set_xticklabels(df['Year'], rotation=90)
ax.yaxis.grid(color='lightgray', linestyle='dashed')

plt.savefig('18.jpg')

plt.tight_layout()
plt.show()
```



In above line plot we can Clearly see that More the population chances of Crimes would be higher. x axis showing population increament year wise correspondigly increament in Total Crimes. y-axis on lefy side showing Population data and Y axis on right side shwoing Total crimes

- Population is one of the important factors influencing incidence of crime. A number of socio-economic factors, besides population, could influence the crime situation at a particular place.

```

In [108]: fig, ax = plt.subplots(figsize=(12,5))
          ax2 = ax.twinx()
          ax.set_title('Population VS CRIME RATE')
          ax.set_xlabel('Year')

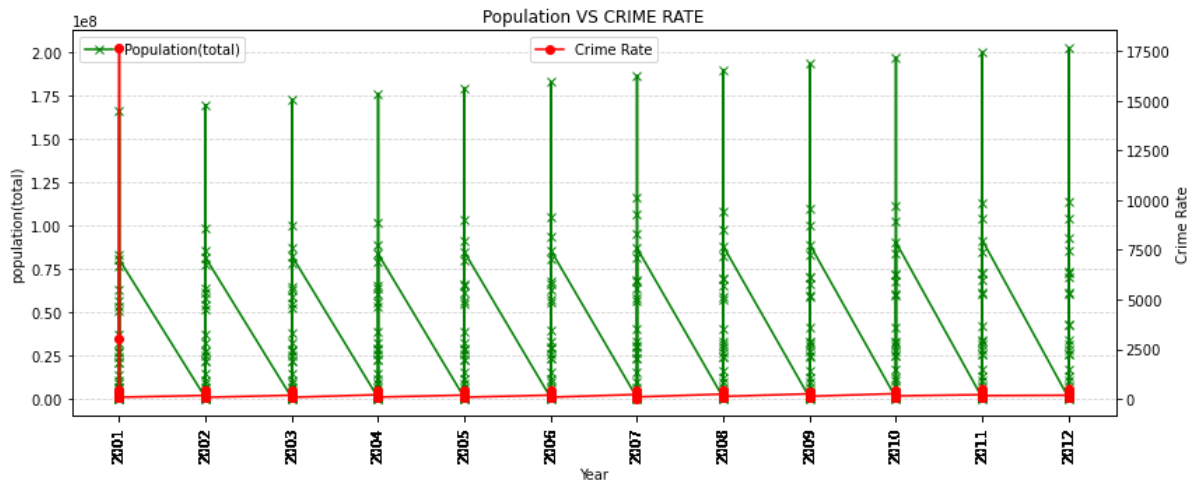
          plt.savefig('19.jpg')

          ax.plot(df['Year'],df['population(total)'], color='green', marker='x')
          ax2.plot(df['Year'],df['Crime Rate'], color='red', marker='o')

          ax.set_ylabel('population(total)')
          ax2.set_ylabel('Crime Rate')
          ax.legend(['Population(total)'])
          ax2.legend(['Crime Rate'], loc='upper center')
          ax.set_xticks(df['Year'])
          ax.set_xticklabels(df['Year'], rotation=90)
          ax.yaxis.grid(color='lightgray', linestyle='dashed')
          plt.savefig('20.jpg')

          plt.tight_layout()
          plt.show()

```



2.6 Each state crime report. There is no fixed format to write a report, you can write a report inside the notebook itself based on what you have analyzed in the above points.

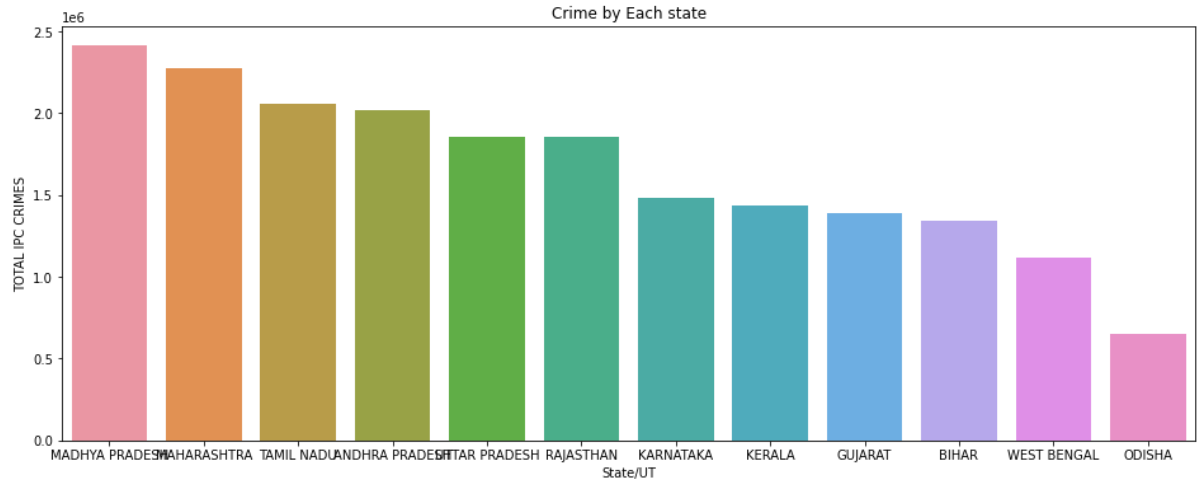
```
In [110]: tc= df.groupby("State/UT")["TOTAL IPC CRIMES"].sum().sort_values(ascending = False)
print("Crime by Each state:\n",tc)
```

Crime by Each state:

State/UT	TOTAL IPC CRIMES
MADHYA PRADESH	2413770
MAHARASHTRA	2273436
TAMIL NADU	2060176
ANDHRA PRADESH	2018981
UTTAR PRADESH	1858074
RAJASTHAN	1855916
KARNATAKA	1481063
KERALA	1437459
GUJARAT	1385775
BIHAR	1346293
WEST BENGAL	1119304
ODISHA	647946
DELHI	633174
ASSAM	597764
HARYANA	595303
CHHATTISGARH	561027
JHARKHAND	422351
PUNJAB	384131
JAMMU & KASHMIR	259155
HIMACHAL PRADESH	154948
UTTARAKHAND	103204
PUDUCHERRY	54116
TRIPURA	52734
CHANDIGARH	40807
MANIPUR	35072
GOA	32051
ARUNACHAL PRADESH	27652
MIZORAM	26146
MEGHALAYA	25249
NAGALAND	13133
A& N ISLANDS	9102
SIKKIM	7000
D & N HAVELI	4651
DAMAN & DIU	4572
LAKSHADWEEP	743

Name: TOTAL IPC CRIMES, dtype: int64

```
In [111]: plt.figure(figsize=(16,6))
sns.barplot(x = tc.index[:12], y = tc[:12]).set(title = "Crime by Each state")
plt.savefig('crimebyeachstate.jpg')
plt.show()
```



In []:

From ABOve Bar plot we can see Top 12 states which have higher IPC Crimes - TOP 3 States are Madhya Pradesh, Maharashtra and Tamil Nadu

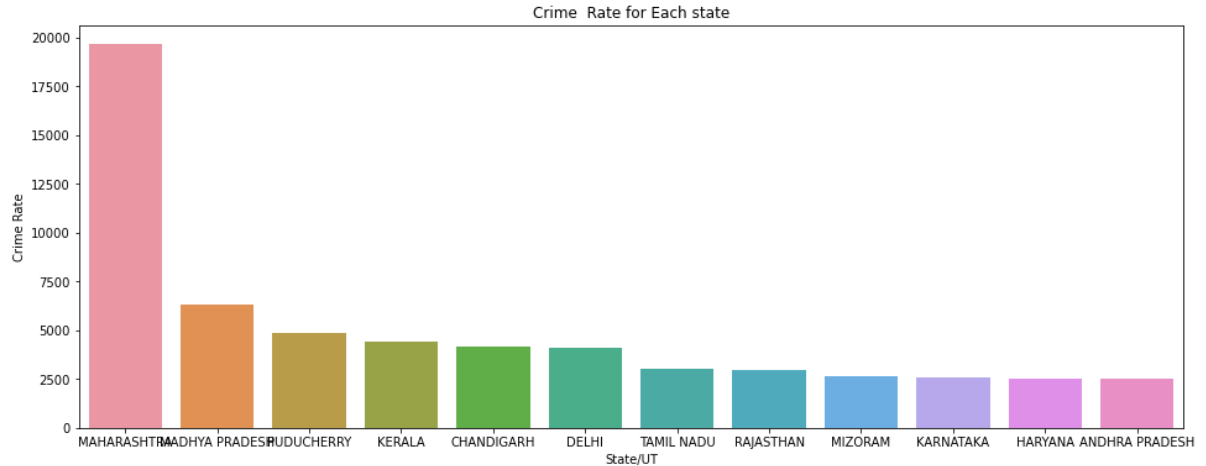
```
In [112]: tcr= df.groupby("State/UT")["Crime Rate"].sum().sort_values(ascending = False)
print("Crime Rate of Each state:\n",tcr)
```

Crime Rate of Each state:

State/UT	
MAHARASHTRA	19655.285754
MADHYA PRADESH	6316.673756
PUDUCHERRY	4869.180671
KERALA	4385.598206
CHANDIGARH	4155.725611
DELHI	4126.855153
TAMIL NADU	3037.821579
RAJASTHAN	2946.820293
MIZORAM	2661.527940
KARNATAKA	2578.819811
HARYANA	2532.205598
ANDHRA PRADESH	2488.914059
GUJARAT	2478.217328
A& N ISLANDS	2457.210401
CHHATTISGARH	2389.783310
HIMACHAL PRADESH	2386.476742
DAMAN & DIU	2324.749802
GOA	2276.257698
ARUNACHAL PRADESH	2227.536009
JAMMU & KASHMIR	2100.831570
ASSAM	2035.844909
D & N HAVELI	1675.369068
ODISHA	1633.077477
TRIPURA	1509.964138
PUNJAB	1463.824145
MANIPUR	1453.769068
BIHAR	1422.610111
JHARKHAND	1370.455004
WEST BENGAL	1281.098158
SIKKIM	1209.923472
LAKSHADWEEP	1182.225532
UTTARAKHAND	1102.138263
UTTAR PRADESH	1003.792457
MEGHALAYA	940.503341
NAGALAND	661.155333

Name: Crime Rate, dtype: float64

```
In [113]: plt.figure(figsize=(16,6))
sns.barplot(x = tcr.index[:12], y = tcr[:12]).set(title = "Crime Rate for Each
plt.savefig('crimeratestate.jpg')
plt.show()
```



- Maharashtra State is the one which is having the highest crime rate and Nagaland is the state which has lowest crime rate

In []:

In []:

- A& N ISLANDS- In 2001 Andaman had 184.75 Crime Rate which got declined in 2012 as population got increased and it was highest in 2009 (250.04 Crime Rate). Its Area is 8249KM²
- ANDHRA PRADESH- Total IPC crimes were 2018981 in 2012 in which that population were highest too
- ARUNACHAL PRADESH- with count of Total IPC Crimes 27652 – on population of 1407014 which make s crime rate of 171.
- ASSAM- 597764 with total IPC crimes Assam state comes at 14th position
- BIHAR- 1346293 Total IPC crimes comes at 10th position
- CHANDIGARH- 40807 total ipc crimes is less populated and crimes are less here as compare to other state /UTs
- CHHATTISGARH-with total ipc crimes 561027 Chhatisgarh comes at 15th position
- D & N HAVELI- Being an UT with an area of 491 km square has only 4651 total IPC crimes in the span of 11 years.
- DAMAN & DIU- Its also an UT which has less area and crimes are 4651 , with an area of 12 km square it means lesser the area , less populated and less crimes
- DELHI – Being the Capital of India , Delhi has 633174 crime count in the span of 2001 to 2012. Delhi is vast in area wise 16579 km² which is large in size.
- GOA- with an area of 3702 km² , Goa is the smallest state of India also it has 32051 count of crime , which has happened in the span of 11 years.(2001 to 2012)
- GUJARAT – is the hub for textile business , and it has crimes 1385775 and it comes at 8th position.

- HARYANA – 595303 total ipc crimes happened in the duration of 11 year it is 44212 km² spread and it has 245 crime rate per capita.
- HIMACHAL PRADESH – 154948 crimes were happened in the duration of 11 years (2001-2012)
- JAMMU & KASHMIR – 259155 crimes were reported in the duration of 11 years (2001-2012)
- JHARKHAND – 422351 crimes were reported in the span of 11 years (2001 to 2012)
- KARNATAKA -1481063 crimes were happened in the duration of 11 years with the population count of 61290592 that means it has crime rate of 214 (avg)
- KERALA – Being the no.1 state in literacy – Kerala still has 1437459 total crime count in the span of 11 years
- LAKSHADWEEP – 743 the lowest crime count for any UT also its in small in area as well only 32 km square.
- MADHYA PRADESH – is in the Top 3 states with higher crimes 2413770, also its large in area and Population wise as well.
- MAHARASHTRA – 2nd State with highest crime with largest population and area covered
- MANIPUR- Small state and less populated as well
- MEGHALAYA – Small state and less populated
- MIZORAM – 1110132 count of Population and total Crimes are 35072.
- NAGALAND – 13133 crimes were happened in the duration of 2001 to 2012
- ODISHA – 647946 crimes were happened in the duration of 11 years.
- PUDUCHERRY-54116 total crimes were happened in the span of 1 years
- PUNJAB - 384131 crimes were happened in the duration of 11 years.
- RAJASTHAN- In terms of area, the largest state in India is Rajasthan.It covers an area of over 342,000 square kilometers
- SIKKIM- 7000 crimes were reported in the span of 11 years.
- TAMIL NADU – 3rd State with highest crimes in india 2060176 crime count.
- TRIPURA – 52734 total ipc counts in the duration of 11 years.
- UTTAR PRADESH- Most Populated state and at 5th position in total crimes against
- UTTARAKHAND – 103204 total counts IPC crimes in the duration of 11 years.
- WEST BENGAL – 1110304 crimes were recorded in the duration of 2001 to 2012

In []:

In []: