In [1]:

```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

In [2]:

```
data = pd.read_csv("rape_victims.csv")
```

In [3]:

```
data
```

Out[3]:

State	Cases Reported	Child Victims of Rape (Below 18 Yrs) - Below 6 Years	Child Victims of Rape (Below 18 Yrs) - 6 Years & Above	Child Victims of Rape (Below 18 Yrs) - 12 Years & Above	Child Victims of Rape (Below 18 Yrs) - 16 Years & Above	Child Victims of Rape (Below 18 Yrs) - Total Girl/Child Victims	Women Victims of Rape (Above 18 Yrs) - 18 Years & Above - Below 30 Years	Women Victims of Rape (Above 18 Yrs) - 30 Years & Above - Below 45 Years	Women Victims of Rape (Above 18 Yrs) - 45 Years & Above - Below 60 Years	Women Victims of Rape (Above 18 Yrs) - 60 Years & Above
ndhra adesh	971	16	57	181	251	505	373	76	14	5
nachal adesh	67	4	4	13	11	32	29	9	0	0
ssam	1648	7	24	6	52	89	1043	523	107	5
Bihar	651	0	0	1	3	4	520	111	16	0
tisgarh	2091	41	80	557	541	1219	644	190	42	6
...	...	...	...	...	...	...	...	...	...	...
Haveli	1	0	0	0	0	0	1	0	0	0
ı & Diu	7	0	2	0	0	2	3	2	0	0
elhi UT	1229	0	0	0	0	0	871	337	20	3
dweep	0	0	0	0	0	0	0	0	0	0
ucherry	7	0	0	0	0	0	7	0	0	0

In [4]:

```
data.head()
```

Out[4]:

	Year	Category	State	Cases Reported	Child Victims of Rape (Below 18 Yrs) - Below 6 Years	Child Victims of Rape (Below 18 Yrs) - 6 Years & Above	Child Victims of Rape (Below 18 Yrs) - 12 Years & Above	Child Victims of Rape (Below 18 Yrs) - 16 Years & Above	Child Victims of Rape (Below 18 Yrs) - Total Girl/Child Victims	Women Victims of Rape (Above 18 Yrs) - 18 Years & Above - Below 30 Years
0	2018	State	Andhra Pradesh	971	16	57	181	251	505	373
1	2018	State	Arunachal Pradesh	67	4	4	13	11	32	29
2	2018	State	Assam	1648	7	24	6	52	89	1043
3	2018	State	Bihar	651	0	0	1	3	4	520
4	2018	State	Chhattisgarh	2091	41	80	557	541	1219	644

In [5]:

```
data.tail()
```

Out[5]:

	Year	Category	State	Cases Reported	Child Victims of Rape (Below 18 Yrs) - Below 6 Years	Child Victims of Rape (Below 18 Yrs) - 6 Years & Above	Child Victims of Rape (Below 18 Yrs) - 12 Years & Above	Child Victims of Rape (Below 18 Yrs) - 16 Years & Above	Child Victims of Rape (Below 18 Yrs) - Total Girl/Child Victims	Wor Vict of R (Ab 18 Year Abo Be Ye
103	2016	Union Territory	D&N Haveli	1	0	0	0	0	0	
104	2016	Union Territory	Daman & Diu	7	0	2	0	0	2	
105	2016	Union Territory	Delhi UT	1229	0	0	0	0	0	
106	2016	Union Territory	Lakshadweep	0	0	0	0	0	0	
107	2016	Union Territory	Puducherry	7	0	0	0	0	0	

In [6]:

```
data.shape
```

Out[6]:

```
(108, 15)
```

In [7]:

```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 108 entries, 0 to 107
Data columns (total 15 columns):
 #   Column
Non-Null Count  Dtype
---  ------
--------------  -----
 0   Year
108 non-null    int64
 1   Category
108 non-null    object
 2   State
108 non-null    object
 3   Cases Reported
108 non-null    int64
 4   Child Victims of Rape (Below 18 Yrs) - Below 6 Years
108 non-null    int64
 5   Child Victims of Rape (Below 18 Yrs) - 6 Years & Above
108 non-null    int64
 6   Child Victims of Rape (Below 18 Yrs) - 12 Years & Above
108 non-null    int64
 7   Child Victims of Rape (Below 18 Yrs) - 16 Years & Above
108 non-null    int64
 8   Child Victims of Rape (Below 18 Yrs) - Total Girl/Child Victims
108 non-null    int64
 9   Women Victims of Rape (Above 18 Yrs) - 18 Years & Above - Below 30 Ye
ars  108 non-null    int64
 10  Women Victims of Rape (Above 18 Yrs) - 30 Years & Above - Below 45 Ye
ars  108 non-null    int64
 11  Women Victims of Rape (Above 18 Yrs) - 45 Years & Above - Below 60 Ye
ars  108 non-null    int64
 12  Women Victims of Rape (Above 18 Yrs) - 60 Years & Above
108 non-null    int64
 13  Women Victims of Rape (Above 18 Yrs) - Total Women/Adult Victims
108 non-null    int64
 14  Total Victims
108 non-null    int64
dtypes: int64(13), object(2)
memory usage: 12.8+ KB
```

In [8]:

```
data.describe()
```

Out[8]:

	Year	Cases Reported	Child Victims of Rape (Below 18 Yrs) - Below 6 Years	Child Victims of Rape (Below 18 Yrs) - 6 Years & Above	Child Victims of Rape (Below 18 Yrs) - 12 Years & Above	Child Victims of Rape (Below 18 Yrs) - 16 Years & Above	Ch Victims Rape (Bel 18 Yr To Girl/Ch Victi
count	108.000000	108.000000	108.000000	108.000000	108.000000	108.000000	108.000(
mean	2017.000000	911.796296	8.120370	22.157407	103.092593	143.250000	276.6203
std	0.820303	1268.314753	19.207046	47.079801	242.142714	319.328534	608.8816
min	2016.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000(
25%	2016.000000	57.250000	0.000000	0.000000	0.000000	0.000000	0.750(
50%	2017.000000	477.000000	0.000000	2.000000	6.000000	7.000000	20.000(
75%	2018.000000	1128.000000	5.000000	15.750000	59.000000	68.250000	139.000(
max	2018.000000	5562.000000	105.000000	207.000000	1275.000000	1550.000000	3082.000(

In [9]:

```
data.isnull().sum()
```

Out[9]:

```
Year
0
Category
0
State
0
Cases Reported
0
Child Victims of Rape (Below 18 Yrs) - Below 6 Years
0
Child Victims of Rape (Below 18 Yrs) - 6 Years & Above
0
Child Victims of Rape (Below 18 Yrs) - 12 Years & Above
0
Child Victims of Rape (Below 18 Yrs) - 16 Years & Above
0
Child Victims of Rape (Below 18 Yrs) - Total Girl/Child Victims
0
Women Victims of Rape (Above 18 Yrs) - 18 Years & Above - Below 30 Years
0
Women Victims of Rape (Above 18 Yrs) - 30 Years & Above - Below 45 Years
0
Women Victims of Rape (Above 18 Yrs) - 45 Years & Above - Below 60 Years
0
Women Victims of Rape (Above 18 Yrs) - 60 Years & Above
0
Women Victims of Rape (Above 18 Yrs) - Total Women/Adult Victims
0
Total Victims
0
dtype: int64
```

In [10]:

```
data.columns
```

Out[10]:

```
Index(['Year', 'Category', 'State', 'Cases Reported',
       'Child Victims of Rape (Below 18 Yrs) - Below 6 Years',
       'Child Victims of Rape (Below 18 Yrs) - 6 Years & Above ',
       'Child Victims of Rape (Below 18 Yrs) - 12 Years & Above ',
       'Child Victims of Rape (Below 18 Yrs) - 16 Years & Above ',
       'Child Victims of Rape (Below 18 Yrs) - Total Girl/Child Victims',
       'Women Victims of Rape (Above 18 Yrs) - 18 Years & Above - Below 30
Years',
       'Women Victims of Rape (Above 18 Yrs) - 30 Years & Above - Below 45
Years',
       'Women Victims of Rape (Above 18 Yrs) - 45 Years & Above - Below 60
Years',
       'Women Victims of Rape (Above 18 Yrs) - 60 Years & Above',
       'Women Victims of Rape (Above 18 Yrs) - Total Women/Adult Victims',
       'Total Victims'],
      dtype='object')
```

In [11]:

```
# let's take some general data and plot some simple charts
rape_victims_by_state = data.groupby('State').sum()
rape_victims_by_state.drop('Year', axis = 1, inplace = True)
print('Total Rape Victims = ' ,rape_victims_by_state['Total Victims'].sum())
rape_victims_by_state.sort_values(by = 'Total Victims', ascending = False).head()
```

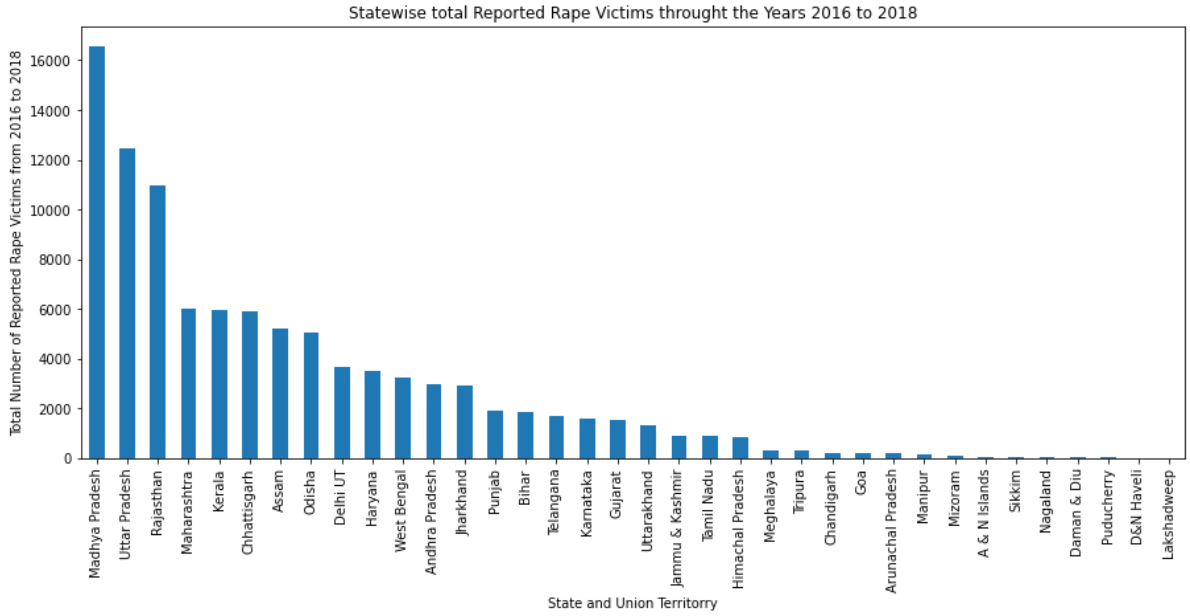
```
Total Rape Victims =  101293
```

Out[11]:

State	Cases Reported	Child Victims of Rape (Below 18 Yrs) - Below 6 Years	Child Victims of Rape (Below 18 Yrs) - 6 Years & Above	Child Victims of Rape (Below 18 Yrs) - 12 Years & Above	Child Victims of Rape (Below 18 Yrs) - 16 Years & Above	Child Victims of Rape (Below 18 Yrs) - Total Girl/Child Victims	Women Victims of Rape (Above 18 Yrs) - 18 Years & Above - Below 30 Years	Women Victims of Rape (Above 18 Yrs) - 30 Years & Above - Below 45 Years	Women Victims of Rape (Above 18 Yrs) - 45 Years & Above - Below 60 Years
Madhya Pradesh	16557	154	556	3693	4602	9005	5100	2157	283
Uttar Pradesh	12438	277	554	1996	1704	4531	6861	2161	88
Rajasthan	10945	31	106	704	1215	2056	6239	2400	271
Kerala	5951	112	359	1078	1777	3326	1655	870	127
Maharashtra	6008	2	4	6	12	24	4380	1460	130

In [12]:

```
plt.subplots(figsize = (15, 6))
cr = rape_victims_by_state['Cases Reported'].sort_values(ascending = False)
ax = cr.plot.bar()
ax.set_xlabel('State and Union Territorry')
ax.set_ylabel('Total Number of Reported Rape Victims from 2016 to 2018')
ax.set_title('Statewise total Reported Rape Victims throught the Years 2016 to 2018')
plt.show()
print(cr)
```




```
State
Madhya Pradesh       16557
Uttar Pradesh        12438
Rajasthan            10945
Maharashtra           6008
Kerala                5951
Chhattisgarh          5907
Assam                 5192
Odisha                5058
Delhi UT              3673
Haryana               3494
West Bengal           3237
Andhra Pradesh        2947
Jharkhand             2918
Punjab                1891
Bihar                 1861
Telangana             1710
Karnataka             1584
Gujarat               1507
Uttarakhand           1309
Jammu & Kashmir        912
Tamil Nadu             897
Himachal Pradesh       842
Meghalaya              325
Tripura                287
Chandigarh             216
Goa                    213
Arunachal Pradesh      185
```

```
Manipur                132
Mizoram                100
A & N Islands           56
Sikkim                  50
Nagaland                30
Daman & Diu             17
Puducherry              14
D&N Haveli               9
Lakshadweep              2
Name: Cases Reported, dtype: int64
```

In [13]:

```
mp_rape_victims = data[data['State'] == 'Madhya Pradesh']
```

In [14]:

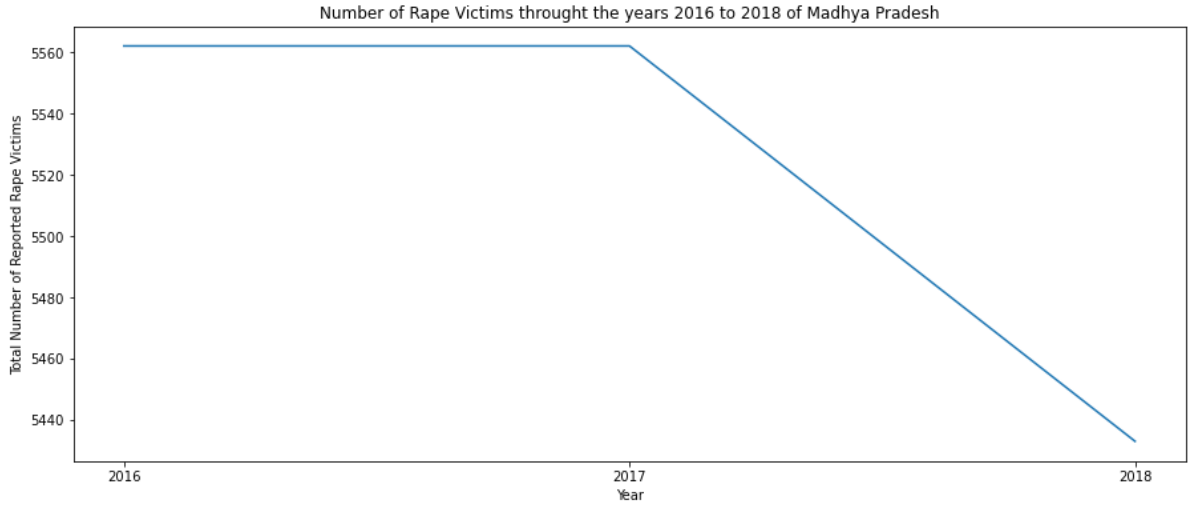
```
mp_rape_victims.head()
```

Out[14]:

	Year	Category	State	Cases Reported	Child Victims of Rape (Below 18 Yrs) - Below 6 Years	Child Victims of Rape (Below 18 Yrs) - 6 Years & Above	Child Victims of Rape (Below 18 Yrs) - 12 Years & Above	Child Victims of Rape (Below 18 Yrs) - 16 Years & Above	Child Victims of Rape (Below 18 Yrs) - Total Girl/Child Victims	Women Victims of Rape (Above 18 Yrs) - 18 Years & Above - Below 30 Years	W V c Y A
13	2018	State	Madhya Pradesh	5433	54	142	1143	1502	2841	1798	
49	2017	State	Madhya Pradesh	5562	50	207	1275	1550	3082	1651	
85	2016	State	Madhya Pradesh	5562	50	207	1275	1550	3082	1651	

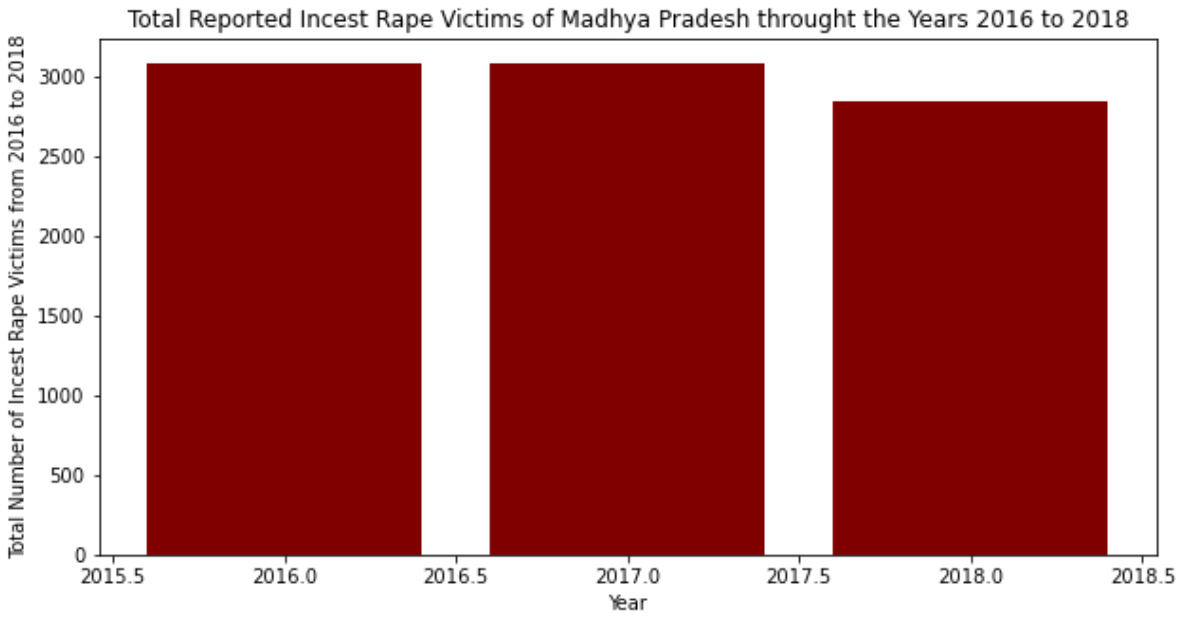
In [15]:

```
# Let's have a look at yearly distribution of number of rape victims in Madhya Pradesh
mp_rape_victims_by_year = mp_rape_victims.groupby('Year').sum()
# plotting the data
plt.subplots(figsize = (15, 6))
ax = mp_rape_victims_by_year['Cases Reported'].plot()
ax.xaxis.set_ticks(np.arange(2016, 2019, 1))
ax.set(xlabel = 'Year', ylabel = 'Total Number of Reported Rape Victims',
       title = 'Number of Rape Victims throught the years 2016 to 2018 of Madhya Pradesh
plt.show()
```



In [16]:

```
mp_incest_rape_cases = mp_rape_victims['Child Victims of Rape (Below 18 Yrs) - Total Gir
year = mp_rape_victims['Year']
fig = plt.figure(figsize = (10, 5))
# creating the bar plot
plt.bar(year, mp_incest_rape_cases , color ='maroon')
plt.xlabel("Year")
plt.ylabel("Total Number of Incest Rape Victims from 2016 to 2018")
plt.title("Total Reported Incest Rape Victims of Madhya Pradesh throught the Years 2016
plt.show()
```



In [20]:

```
x = data.drop(['Cases Reported', 'State', 'Category'], axis = 1)
y = data['Cases Reported']
```

In [21]:

```
x.shape
```

Out[21]:

```
(108, 12)
```

In [22]:

```
y.shape
```

Out[22]:

```
(108,)
```

In [27]:

```
from sklearn.linear_model import LogisticRegression
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(x, y, test_size = 0.2)
```

In [28]:

```
model = LogisticRegression()
model.fit(X_train, y_train)
```

```
c:\python\lib\site-packages\sklearn\linear_model\_logistic.py:814: Converg
enceWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown i
n:
    https://scikit-learn.org/stable/modules/preprocessing.html (https://sc
ikit-learn.org/stable/modules/preprocessing.html)
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-reg
ression (https://scikit-learn.org/stable/modules/linear_model.html#logisti
c-regression)
  n_iter_i = _check_optimize_result(
```

Out[28]:

```
LogisticRegression()
```

In [29]:

```
y_pred = model.predict(X_test)
```

In [30]:

```
print("Training Accuracy :", model.score(X_train, y_train))
print("Testing Accuracy :", model.score(X_test, y_test))
```

```
Training Accuracy : 0.7790697674418605
Testing Accuracy : 0.5
```

In [31]:

```
from sklearn.tree import DecisionTreeRegressor
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(x, y, test_size = 0.2)
```

In [32]:

```
model3 = DecisionTreeRegressor(max_depth=6)
model3.fit(X_train, y_train)
```

Out[32]:

```
DecisionTreeRegressor(max_depth=6)
```

In [33]:

```
y_pred = model3.predict(X_test)
```

In [34]:

```
print("Training Accuracy :", model3.score(X_train, y_train))
print("Testing Accuracy :", model3.score(X_test, y_test))
```

```
Training Accuracy : 0.9998812876120555
Testing Accuracy : 0.7222655420885897
```

In [35]:

```
from sklearn.ensemble import RandomForestRegressor
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(x, y, test_size = 0.2)
```

In [36]:

```
model4 = RandomForestRegressor(n_estimators = 100, random_state = 0)
model4.fit(X_train, y_train)
```

Out[36]:

```
RandomForestRegressor(random_state=0)
```

In [37]:

```
y_pred = model4.predict(X_test)
```

In [38]:

```
print("Training Accuracy :", model4.score(X_train, y_train))
print("Testing Accuracy :", model4.score(X_test, y_test))
```

```
Training Accuracy : 0.997504516754946
Testing Accuracy : 0.99086117195589
```