evelopers-sustainability-hackathon

July 17, 2023

MACHINEHACK GENPACT| GOOGLE FOR DEVELOPERS SUSTAINABILITY HACKATHON

[2]:
```
# importing the necessary libraies
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline
```

FEATURE ENGINEERING

[3]:
```
df1 = pd.read_csv("soilrainfall.csv")
df1.head()
```

[3]:
```
     State Name                         SOIL TYPE PERCENT (Percent)  \
0  Chhattisgarh       LOAMY ALFISOLS - 60% ; USTALF/USTOLLS - 40%
1  Chhattisgarh                                LOAMY ALFISOL - 100%
2  Chhattisgarh  USTALF/USTOLLS - 50% ; LOAMY ALFISOLS - 25% ; …
3  Chhattisgarh                               USTALF/USTOLLS - 100%
4  Chhattisgarh                               USTALF/USTOLLS - 100%

      Year_Rainfall  ANNUAL NORMAL RAINFALL (Millimeters)
0  Average 30 years                                  1277
1  Average 30 years                                  1535
2  Average 30 years                                  1388
3  Average 30 years                                  1327
4  Average 30 years                                  1628
```

[4]:
```
df2 = pd.read_csv("commodityprices.csv")
df2.head()
```

[4]:
```
   Year  Cotton_Price[Dollar/ton]
0  1975               1055.792518
1  1976               1582.035312
2  1977               1399.933700
3  1978               1350.109288
```

```
4  1979                1428.152836
```

[5]:
```
df1.shape
```

[5]:
```
(313, 4)
```

[6]:
```
df2.shape
```

[6]:
```
(48, 2)
```

[7]:
```
df1.describe()
```

[7]:

	ANNUAL NORMAL RAINFALL (Millimeters)
count	313.000000
mean	1204.571885
std	636.098733
min	-1.000000
25%	813.000000
50%	1079.000000
75%	1391.000000
max	3667.000000

[8]:
```
df2.describe()
```

[8]:

	Year	Cotton_Price[Dollar/ton]
count	48.00	48.000000
mean	1998.50	849.508980
std	14.00	751.629995
min	1975.00	0.000000
25%	1986.75	0.000000
50%	1998.50	1193.250575
75%	2010.25	1503.605955
max	2022.00	2048.091980

[9]:
```
df1.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 313 entries, 0 to 312
Data columns (total 4 columns):
 #   Column                                Non-Null Count  Dtype
---  ------                                --------------  -----
 0   State Name                            313 non-null    object
 1   SOIL TYPE PERCENT (Percent)           310 non-null    object
 2   Year_Rainfall                         313 non-null    object
 3   ANNUAL NORMAL RAINFALL (Millimeters)  313 non-null    int64
dtypes: int64(1), object(3)
memory usage: 9.9+ KB
```

[10]:
```
df2.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 48 entries, 0 to 47
Data columns (total 2 columns):
 #   Column                     Non-Null Count  Dtype
---  ------                     --------------  -----
 0   Year                       48 non-null     int64
 1   Cotton_Price[Dollar/ton]   48 non-null     float64
dtypes: float64(1), int64(1)
memory usage: 896.0 bytes
```

ONE HOT ENCODING

[11]:
```
df1['Year_Rainfall']=df1['Year_Rainfall'].replace({'Average 30 years':1,␣
 ↪'other':0})
df1
df1['State Name']=df1['State Name'].replace({'Uttar Pradesh':1,'Madhya Pradesh':
 ↪2,'Rajasthan':3,'Maharashtra':4,'Karnataka':5,'Gujarat':6,'West Bengal':
 ↪7,'Orissa':8,'Punjab':9,'Bihar':10,'Andhra Pradesh':11,'Himachal Pradesh':
 ↪12,'Kerala':13,'Assam':14,'Telangana':15,'Uttarakhand':16,'Haryana':
 ↪17,'Chhattisgarh':18,'Jharkhand':19,'Tamil Nadu':20})
df1
```

[11]:
```
     State Name                          SOIL TYPE PERCENT (Percent)  \
0            18        LOAMY ALFISOLS - 60% ; USTALF/USTOLLS - 40%
1            18                              LOAMY ALFISOL - 100%
2            18  USTALF/USTOLLS - 50% ; LOAMY ALFISOLS - 25% ; …
3            18                             USTALF/USTOLLS - 100%
4            18                             USTALF/USTOLLS - 100%
..          …                                                 …
308          19                             USTALF/USTOLLS - 100%
309          19                             USTALF/USTOLLS - 100%
310          19                             USTALF/USTOLLS - 100%
311          19                             USTALF/USTOLLS - 100%
312          20                                               NaN

     Year_Rainfall  ANNUAL NORMAL RAINFALL (Millimeters)
0                1                                  1277
1                1                                  1535
2                1                                  1388
3                1                                  1327
4                1                                  1628
..             …                                   …
308              1                                  1198
309              1                                  1237
310              1                                  1462
311              1                                  1353
```

```
312              1                                    -1

[313 rows x 4 columns]
```

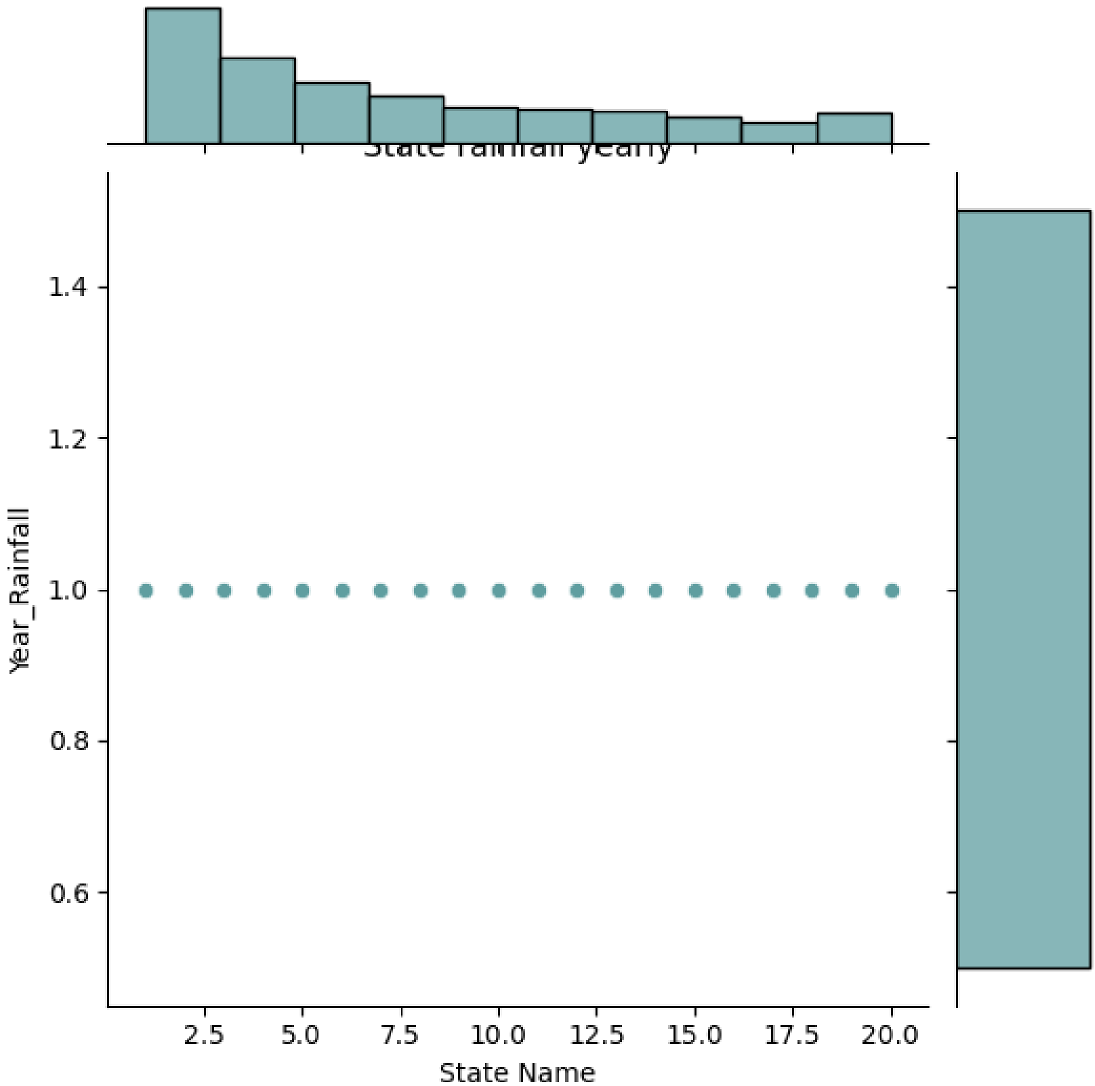
[12]:
```
#displaying the particular number of occurences of different states
df1.value_counts('State Name')
```

[12]:
```
State Name
1     46
2     37
3     26
4     26
5     19
6     18
7     16
20    13
8     13
9     11
10    11
11    11
12    10
13    10
14    10
15     9
16     8
17     7
18     6
19     6
dtype: int64
```

DATA VISUALIZATION AND ANALYSIS

[13]:
```
import seaborn as sns
sns.jointplot(data = df1 , x = 'State Name' , y='Year_Rainfall', color =␣
 ↪'cadetblue')
plt.title("State rainfall yearly")
```

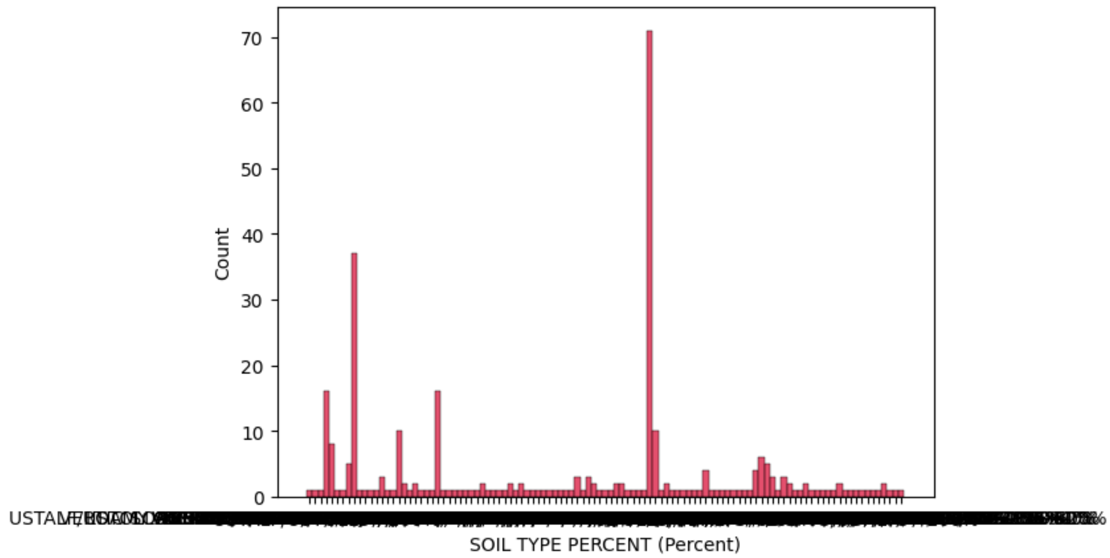
[13]:
```
Text(0.5, 1.0, 'State rainfall yearly')
```



Joint plot above describes the distribution of average rainfall of 30 years over different states which has been constant

[14]:
```
# observing the total percentage of different types of soils
sns.histplot(df1 , x="SOIL TYPE PERCENT (Percent)" , color='crimson')
```

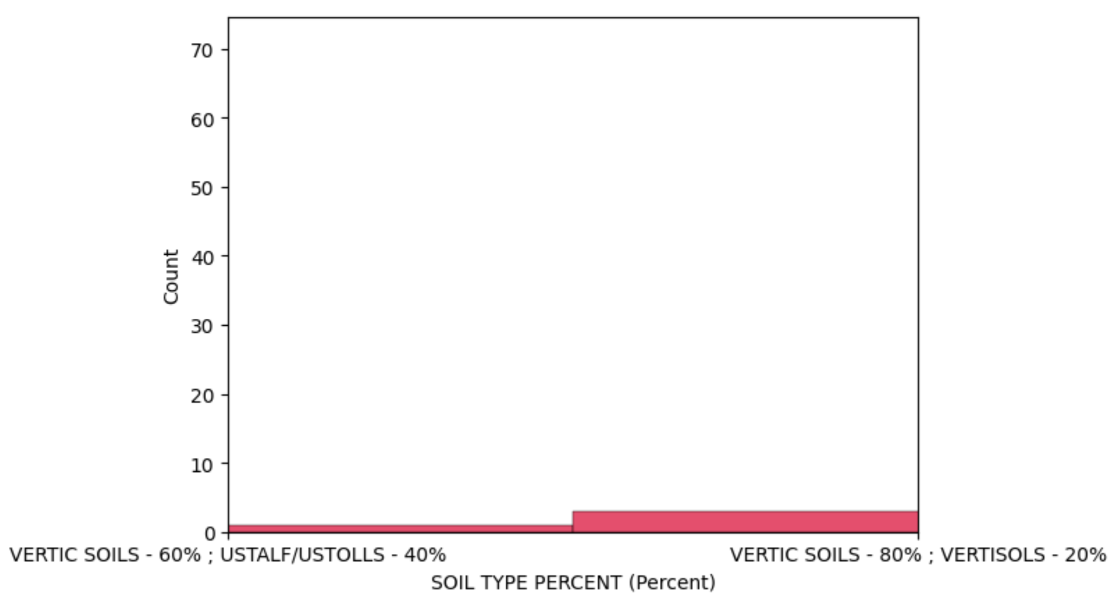
[14]: <Axes: xlabel='SOIL TYPE PERCENT (Percent)', ylabel='Count'>



Above histogram describes the distribution of a large variation of various types of soils due to which there is overlapping on the x-axis

[74]:
```python
sns.histplot(df1 , x="SOIL TYPE PERCENT (Percent)" , color='crimson')
plt.xlim(12,13)
```

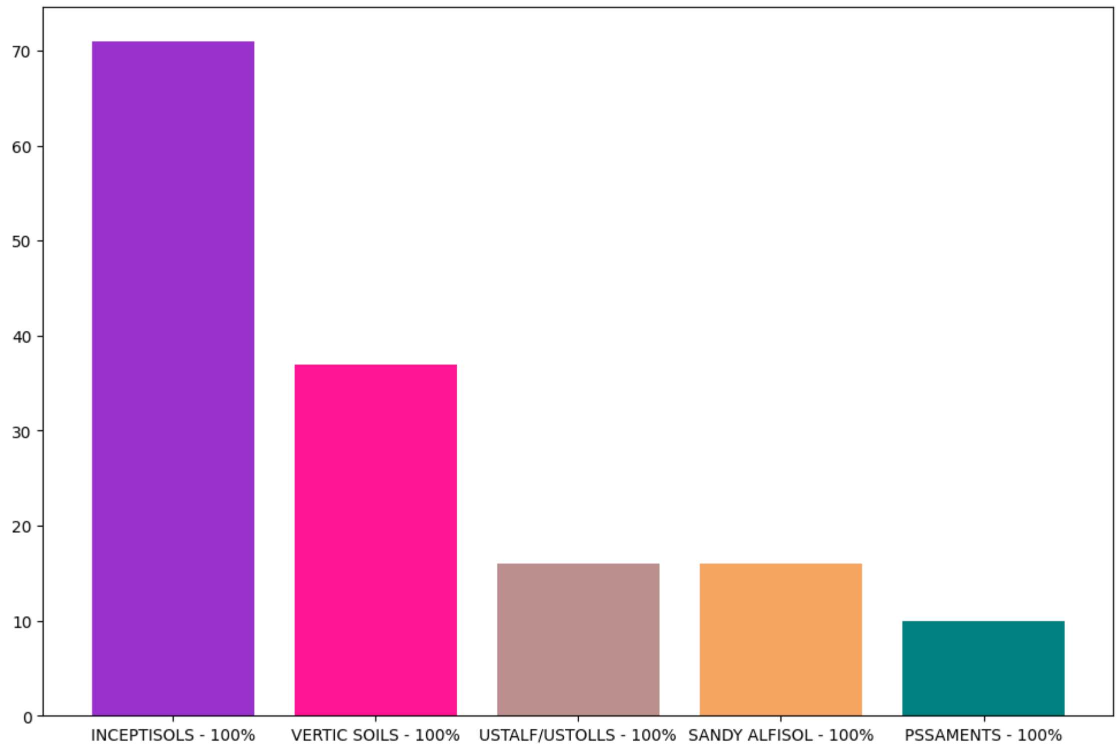
[74]: (12.0, 13.0)



The above histogram describes the variation of various less number of soli types as compared to the previous histogram

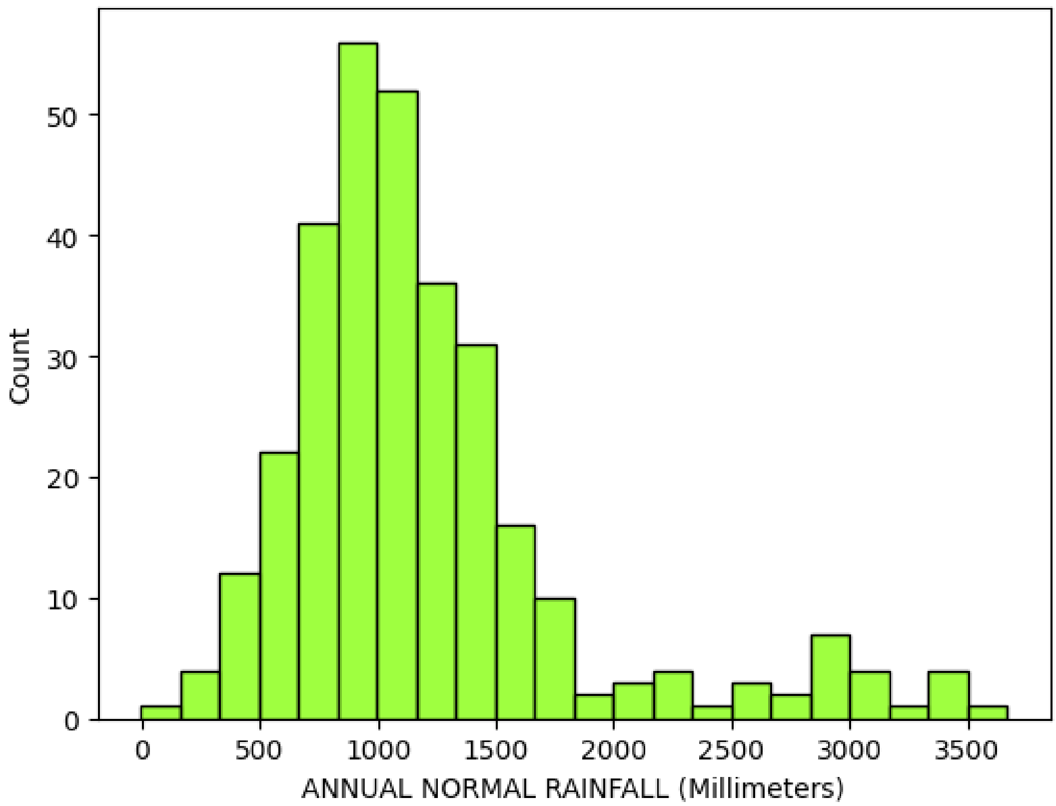
[16]:
```
plt.figure(figsize=(12,8))
plt.bar(list(df1['SOIL TYPE PERCENT (Percent)'].value_counts()[0:5].
 ↪keys()),list(df1['SOIL TYPE PERCENT (Percent)'].value_counts()[0:
 ↪5]),color=['darkorchid','deeppink','rosybrown','sandybrown','teal'])
```

[16]: <BarContainer object of 5 artists>



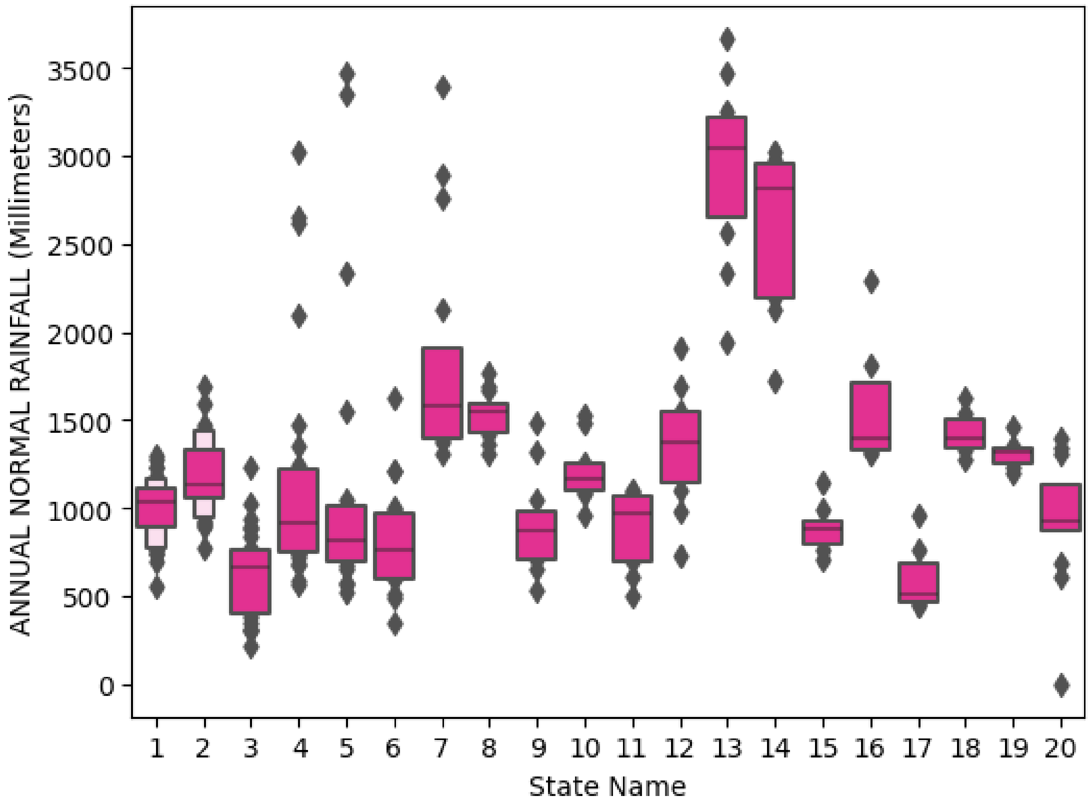
[75]:
```
sns.histplot(df1 , x="ANNUAL NORMAL RAINFALL (Millimeters)" ,␣
 ↪color='chartreuse')
```

[75]: <Axes: xlabel='ANNUAL NORMAL RAINFALL (Millimeters)', ylabel='Count'>



[18]:
```
sns.boxenplot(df1 , x="State Name" , y='ANNUAL NORMAL RAINFALL (Millimeters)',␣
↪color = 'deeppink')
```

[18]:
```
<Axes: xlabel='State Name', ylabel='ANNUAL NORMAL RAINFALL (Millimeters)'>
```



Above histogram and Boxenplot describes the Distribution of annual rainfall in millimeters which shows that a larger portion of the data constitutes to rainfall between 500-1500 mm annually in different states in India.

RESULTS shows us that for soil types such as INCEPTISOLS 100% there is more annual rainfall required in millimeters.

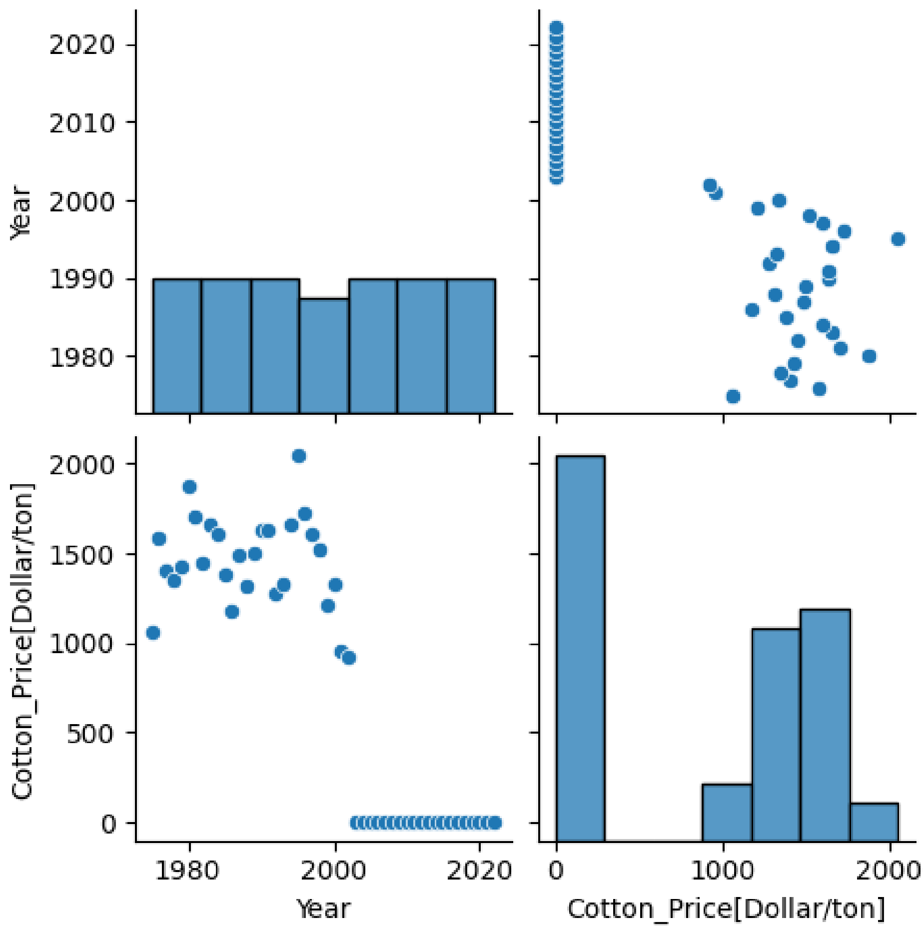
[19]:
```
df2.value_counts()
```

[19]:
```
Year  Cotton_Price[Dollar/ton]
1975  1055.792518                 1
1976  1582.035312                 1
2001  951.734454                  1
2002  918.224230                  1
2003  0.000000                    1
2004  0.000000                    1
2005  0.000000                    1
2006  0.000000                    1
2007  0.000000                    1
2008  0.000000                    1
2009  0.000000                    1
2010  0.000000                    1
```

```
2011   0.000000             1
2012   0.000000             1
2013   0.000000             1
2014   0.000000             1
2015   0.000000             1
2016   0.000000             1
2017   0.000000             1
2018   0.000000             1
2019   0.000000             1
2020   0.000000             1
2021   0.000000             1
2000   1332.472328          1
1999   1211.879614          1
1998   1524.274268          1
1986   1174.621536          1
1977   1399.933700          1
1978   1350.109288          1
1979   1428.152836          1
1980   1869.517760          1
1981   1697.116476          1
1982   1446.892106          1
1983   1655.008234          1
1984   1604.963360          1
1985   1381.855816          1
1987   1485.252494          1
1997   1603.199664          1
1988   1315.055830          1
1989   1496.716518          1
1990   1629.214180          1
1991   1628.993718          1
1992   1276.034056          1
1993   1322.331076          1
1994   1659.858398          1
1995   2048.091980          1
1996   1727.099308          1
2022   0.000000             1
dtype: int64
```

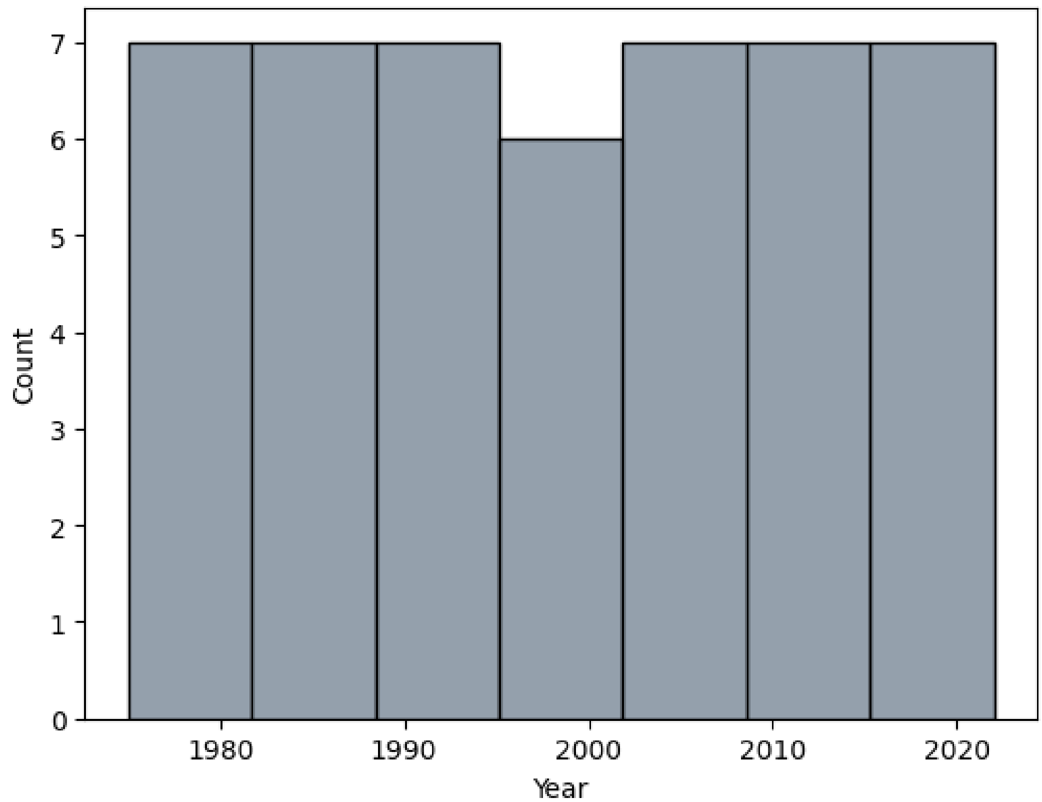
[20]:
```
sns.pairplot(df2)
```

[20]:
```
<seaborn.axisgrid.PairGrid at 0x79a3677234f0>
```



[21]:
```
sns.histplot(df2 , x="Year" , color='slategrey')
```

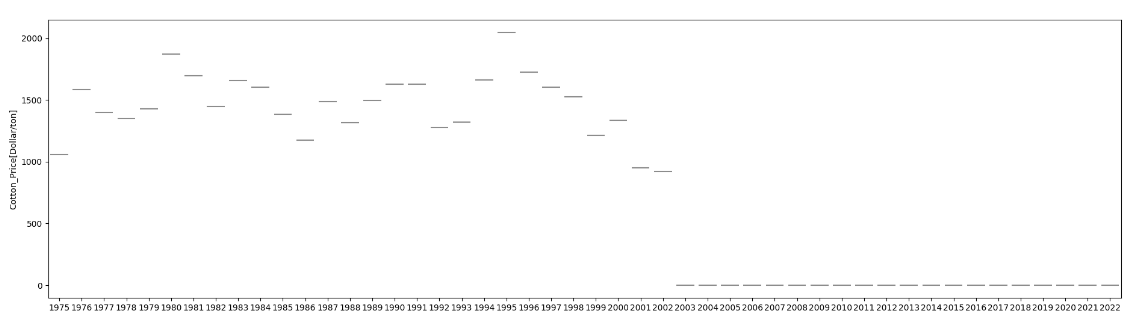
[21]: <Axes: xlabel='Year', ylabel='Count'>



Above histogram describes the years active where the USA cotton commodity was high The plot describes the years from 1980s too 2020 all time high prices except for a downfall in the early 2000s

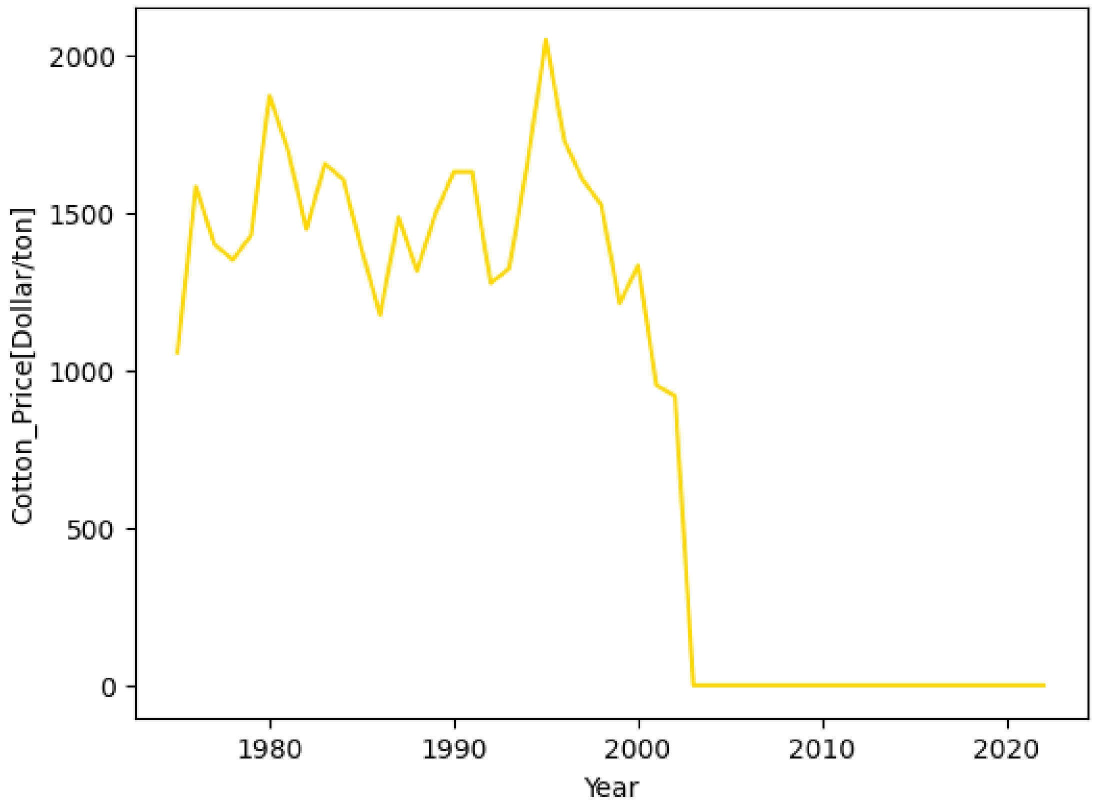
[22]:
```
plt.figure(figsize=(23,6))
sns.boxenplot(df2 , x="Year" , y='Cotton_Price[Dollar/ton]', color = 'black')
#plt.xlim(12,13)
#plt.ylim(1000,2000)
```

[22]: <Axes: xlabel='Year', ylabel='Cotton_Price[Dollar/ton]'>



[23]:
```
sns.lineplot(x="Year", y="Cotton_Price[Dollar/ton]" ,color='gold',data=df2)
```

[23]:
```
<Axes: xlabel='Year', ylabel='Cotton_Price[Dollar/ton]'>
```



Above Boxen plot and line plot describes that the cotton prices in usa were having a variation from the early 1980s till 2003; but after 2003 it has fallen rapidly, declining till almost 0.

MODEL TRAINING

LINEAR REGRESSION ON INDIA SOIL DATA

[24]:
```
# using linear regression algorithm
df3 = pd.read_csv('indiatrain.csv')
df3
```

[24]:
```
     Year  COTTON AREA (1000 ha)  COTTON PRODUCTION (1000 tons)  \
0    1990                    0.0                            0.0
1    1990                    7.0                            3.0
2    1990                   49.0                          238.0
3    1990                   26.0                          120.0
4    1990                  996.0                          289.0
…    …                      …                              …
```

```
1374  1994                 0.0                           0.0
1375  1994                 0.0                           0.0
1376  1994                 0.0                           0.0
1377  1995                11.0                          36.0
1378  1995                24.0                          52.0

      COTTON YIELD (Kg per ha)  TOTAL AREA (1000 ha)  FOREST AREA (1000 ha)  \
0                            0                903.31                 130.36
1                         3333               1451.30                 535.76
2                         4944               1083.84                 327.24
3                         4964                780.54                  85.19
4                         2892                883.69                  68.40
…                         …                   …                     …
1374                         0                375.00                   2.70
1375                         0                343.50                   9.80
1376                         0                617.00                  87.80
1377                      3403                903.00                 131.00
1378                      2551               1451.00                 535.00

      BARREN AND UNCULTIVABLE LAND AREA (1000 ha)  \
0                                          104.08
1                                          202.15
2                                           84.49
3                                           50.32
4                                           66.33
…                                           …
1374                                         0.00
1375                                         4.55
1376                                         9.78
1377                                        98.00
1378                                       192.00

      LAND PUT TO NONAGRICULTURAL USE AREA (1000 ha)  \
0                                             112.64
1                                             126.13
2                                             113.25
3                                              93.54
4                                             114.89
…                                              …
1374                                             NaN
1375                                             NaN
1376                                           94.97
1377                                          121.50
1378                                          128.50

      PERMANENT PASTURES AREA (1000 ha)  OTHER FALLOW AREA (1000 ha)  \
0                                   NaN                         12.75
```

1	7.82	17.86
2	30.71	23.00
3	27.32	43.99
4	22.17	36.84
…	…	…
1374	0.00	NaN
1375	3.04	4.32
1376	0.00	5.43
1377	8.50	12.00
1378	10.50	17.00

	NET CROPPED AREA (1000 ha)	GROSS CROPPED AREA (1000 ha) \
0	498.96	627.26
1	511.69	646.47
2	448.80	698.74
3	447.61	627.21
4	505.26	736.47
…	…	…
1374	290.10	3.00
1375	249.90	0.00
1376	330.00	3.00
1377	511.00	698.00
1378	536.00	716.00

	CROPING INTENSITY (Percent)	NITROGEN CONSUMPTION (tons) \
0	129.12	31003.0
1	127.30	36622.0
2	160.64	NaN
3	143.94	109664.0
4	150.33	88821.0
…	…	…
1374	3.00	24828.0
1375	3.00	20322.0
1376	3.00	16346.0
1377	138.47	39480.0
1378	137.21	32472.0

	PHOSPHATE CONSUMPTION (tons)	POTASH CONSUMPTION (tons) \
0	9306.0	1383.0
1	6835.0	1363.0
2	21088.0	7882.0
3	37972.0	18271.0
4	37910.0	NaN
…	…	…
1374	8530.0	7211.0
1375	7683.0	5301.0
1376	8752.0	4958.0

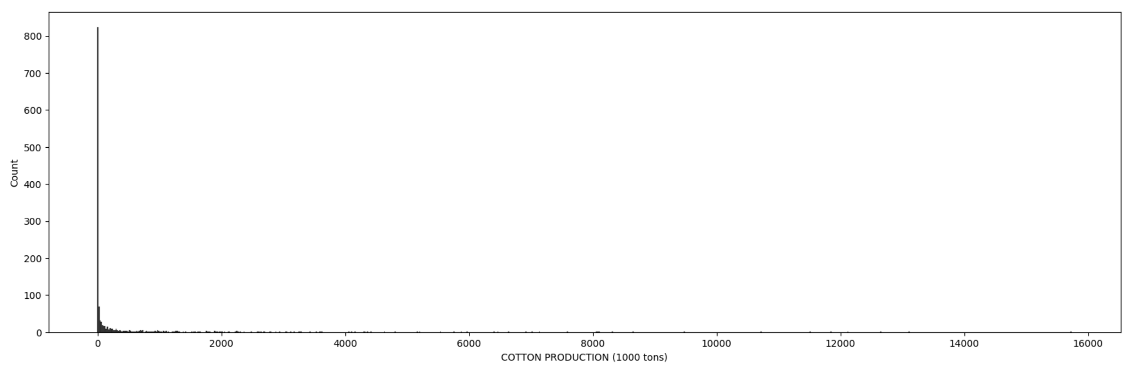
```
1377                            6929.0                          2338.0
1378                            3375.0                          1444.0

      TOTAL CONSUMPTION (tons)  TOTAL PER HA OF NCA (Kg per ha)
0                      41684.0                            85.21
1                      44809.0                            90.08
2                          NaN                           303.24
3                     165898.0                           375.97
4                     139778.0                              NaN
…                         …                           …
1374                   40560.0                           142.76
1375                       NaN                           136.79
1376                   30051.0                            94.33
1377                   48744.0                            96.57
1378                   37287.0                            73.08

[1379 rows x 18 columns]
```

[25]:
```python
plt.figure(figsize=(20,6))
sns.histplot(df3 , x="COTTON PRODUCTION (1000 tons)" , color='black')
```

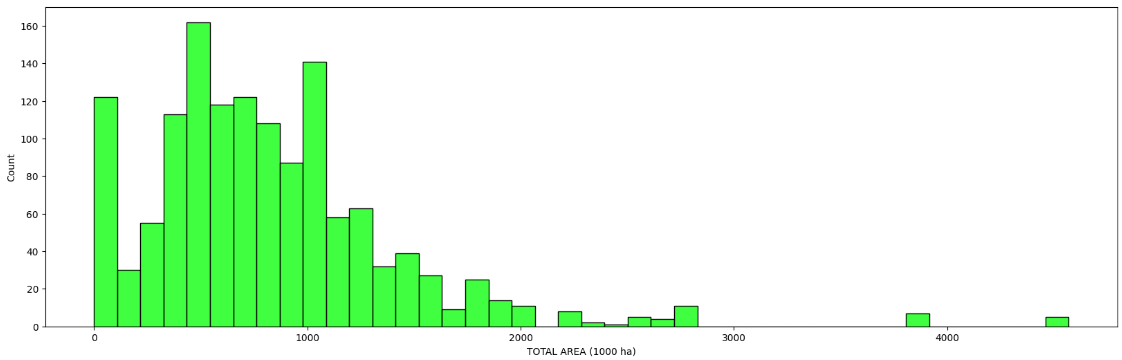
[25]: <Axes: xlabel='COTTON PRODUCTION (1000 tons)', ylabel='Count'>



Cotton Production is denoted high in the early years

[100]:
```python
plt.figure(figsize=(20,6))
sns.histplot(df3 , x="TOTAL AREA (1000 ha)" , color='lime')
```

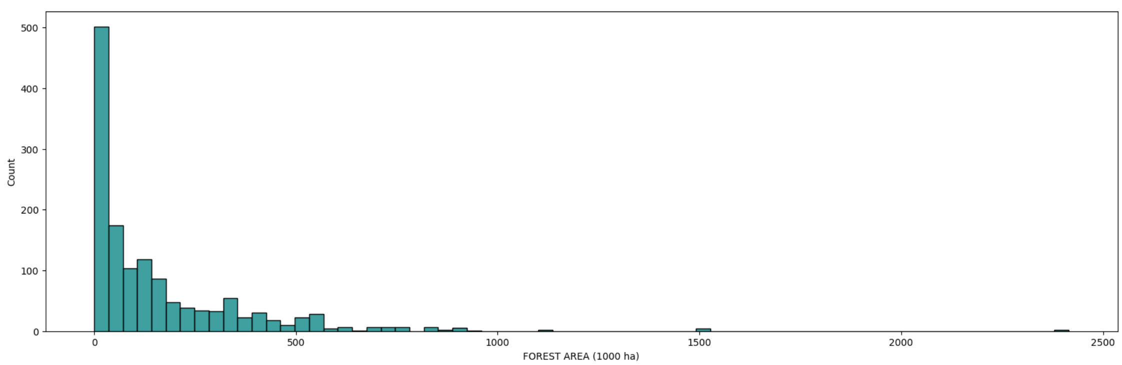
[100]: <Axes: xlabel='TOTAL AREA (1000 ha)', ylabel='Count'>



The total area is also observed to be more in the initial years as compared to latter.

[94]:
```
plt.figure(figsize=(20,6))
sns.histplot(df3 , x="FOREST AREA (1000 ha)" , color='teal')
```

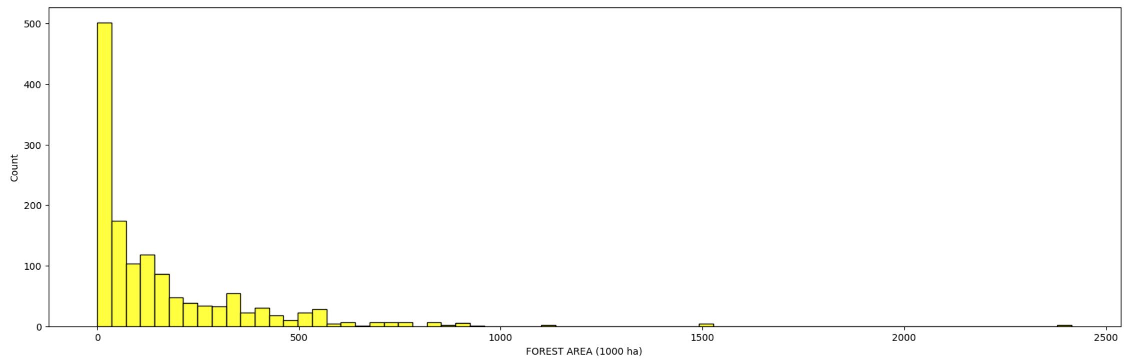
[94]: <Axes: xlabel='FOREST AREA (1000 ha)', ylabel='Count'>



Forest area is also observed to be higher in 1980-1995

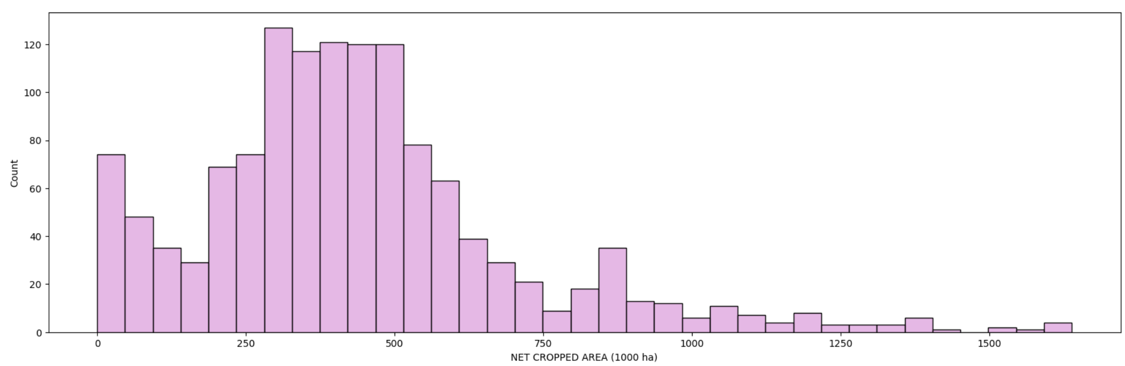
[99]:
```
plt.figure(figsize=(20,6))
sns.histplot(df3 , x="FOREST AREA (1000 ha)" , color='yellow')
```

[99]: <Axes: xlabel='FOREST AREA (1000 ha)', ylabel='Count'>



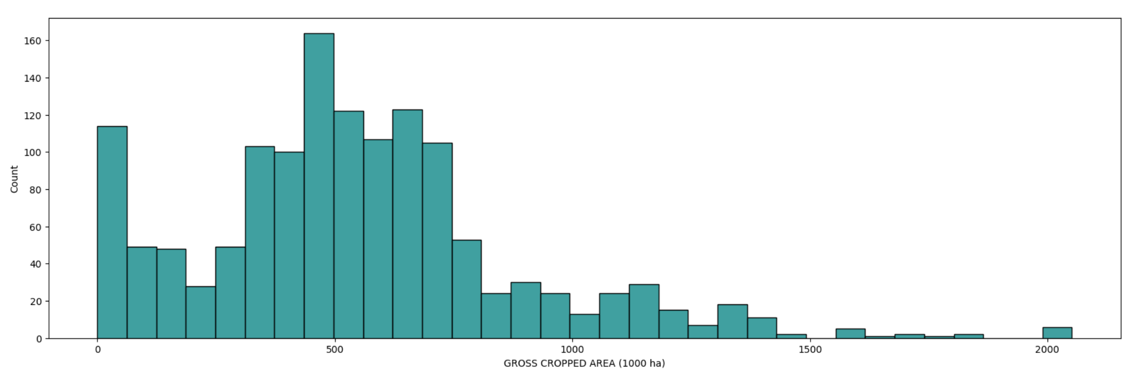
[26]:
```
plt.figure(figsize=(20,6))
sns.histplot(df3 , x="NET CROPPED AREA (1000 ha)" , color='plum')
```

[26]: <Axes: xlabel='NET CROPPED AREA (1000 ha)', ylabel='Count'>



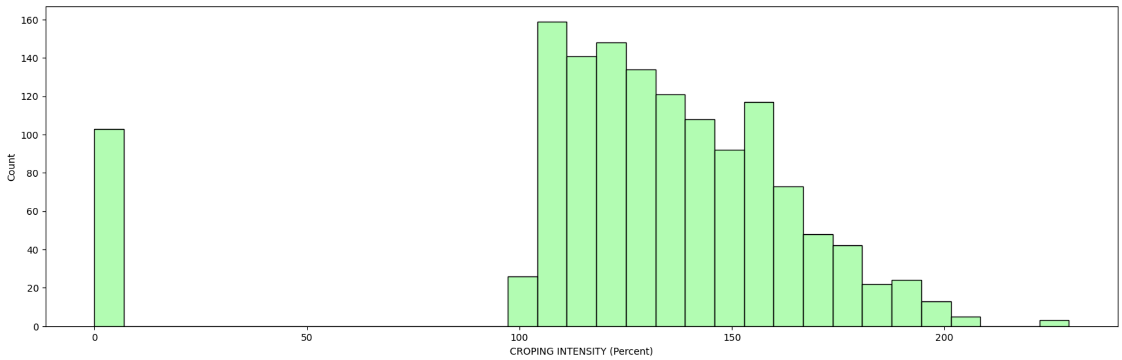
[68]:
```
plt.figure(figsize=(20,6))
sns.histplot(df3 , x="GROSS CROPPED AREA (1000 ha)" , color='teal')
```

[68]: <Axes: xlabel='GROSS CROPPED AREA (1000 ha)', ylabel='Count'>



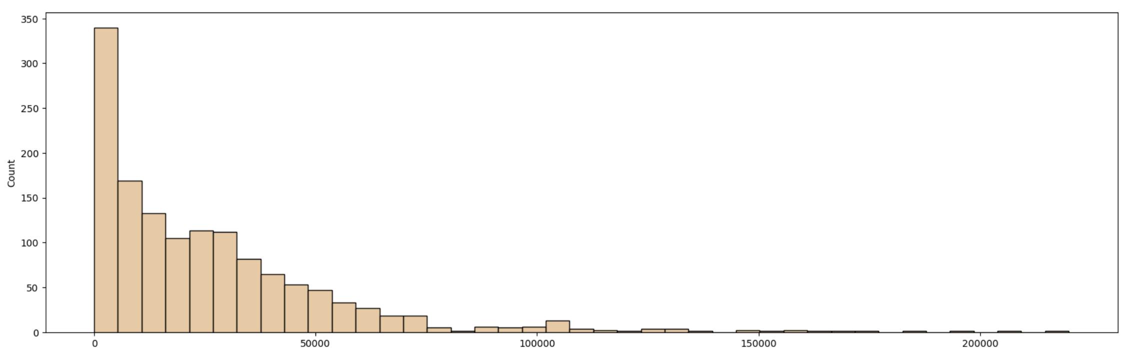
[69]:
```
plt.figure(figsize=(20,6))
sns.histplot(df3 , x="CROPING INTENSITY (Percent)" , color='palegreen')
```

[69]: <Axes: xlabel='CROPING INTENSITY (Percent)', ylabel='Count'>



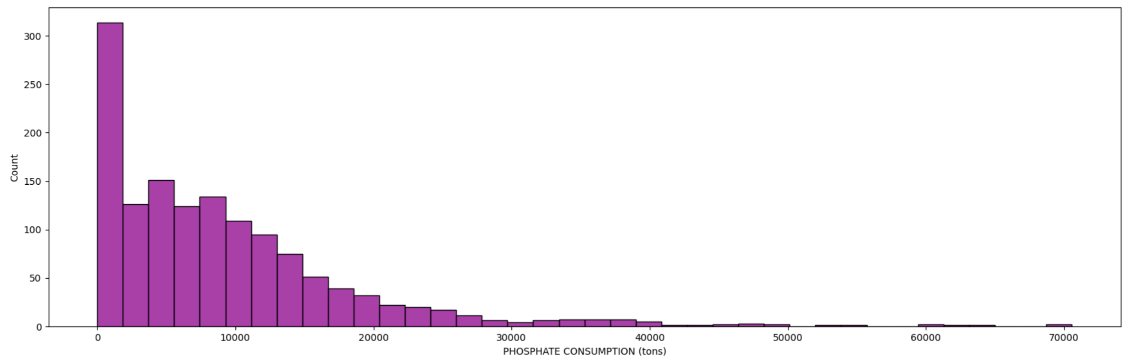
[70]:
```
plt.figure(figsize=(20,6))
sns.histplot(df3 , x="NITROGEN CONSUMPTION (tons)" , color='burlywood')
```

[70]: <Axes: xlabel='NITROGEN CONSUMPTION (tons)', ylabel='Count'>



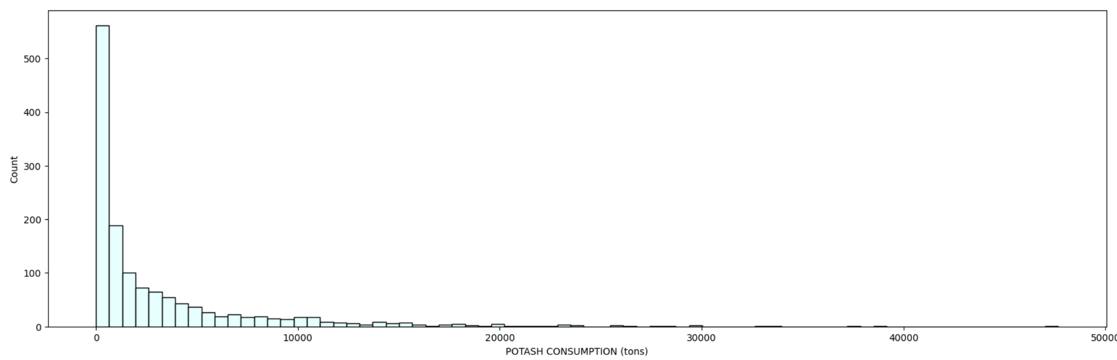
[71]:
```
plt.figure(figsize=(20,6))
sns.histplot(df3 , x="PHOSPHATE CONSUMPTION (tons)" , color='darkmagenta')
```

[71]: <Axes: xlabel='PHOSPHATE CONSUMPTION (tons)', ylabel='Count'>



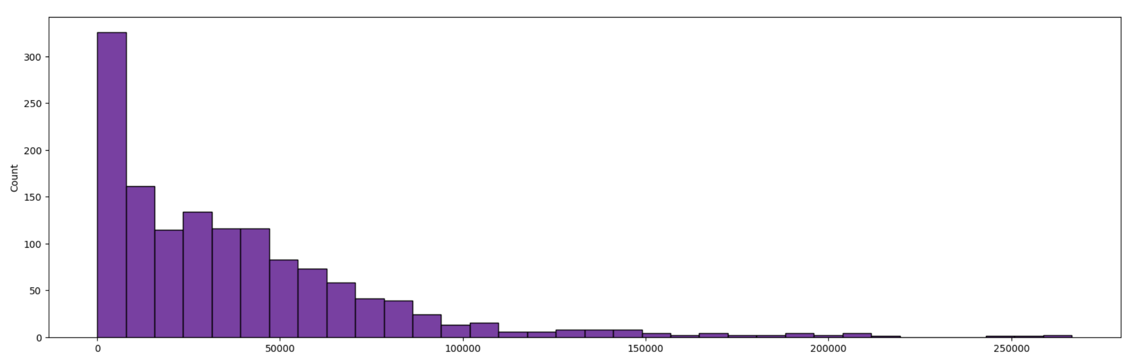

```
[72]: plt.figure(figsize=(20,6))
      sns.histplot(df3 , x="POTASH CONSUMPTION (tons)" , color='lightcyan')
```

```
[72]: <Axes: xlabel='POTASH CONSUMPTION (tons)', ylabel='Count'>
```




```
[73]: plt.figure(figsize=(20,6))
      sns.histplot(df3 , x="TOTAL CONSUMPTION (tons)" , color='indigo')
```

```
[73]: <Axes: xlabel='TOTAL CONSUMPTION (tons)', ylabel='Count'>
```



The net cropped area is also to be higher before 2000s

[42]:
```
X = df3.iloc[:,:-1].values  #independent variable array
y = df3.iloc[:,1].values

df3=df3.replace({'NAN':0})
df3  = df3.fillna(0)   # replacing nan values with zeros
df3
```

[42]:
```
      Year  COTTON AREA (1000 ha)  COTTON PRODUCTION (1000 tons)  \
0     1990                    0.0                            0.0
1     1990                    7.0                            3.0
2     1990                   49.0                          238.0
3     1990                   26.0                          120.0
4     1990                  996.0                          289.0
…     …                     …                            …
1374  1994                    0.0                            0.0
1375  1994                    0.0                            0.0
1376  1994                    0.0                            0.0
1377  1995                   11.0                           36.0
1378  1995                   24.0                           52.0

      COTTON YIELD (Kg per ha)  TOTAL AREA (1000 ha)  FOREST AREA (1000 ha)  \
0                            0                903.31                 130.36
1                         3333               1451.30                 535.76
2                         4944               1083.84                 327.24
3                         4964                780.54                  85.19
4                         2892                883.69                  68.40
…                         …                   …                     …
1374                         0                375.00                   2.70
1375                         0                343.50                   9.80
1376                         0                617.00                  87.80
1377                      3403                903.00                 131.00
1378                      2551               1451.00                 535.00

      BARREN AND UNCULTIVABLE LAND AREA (1000 ha)  \
0                                          104.08
1                                          202.15
2                                           84.49
3                                           50.32
4                                           66.33
…                                           …
1374                                         0.00
1375                                         4.55
1376                                         9.78
```

1377	98.00
1378	192.00

	LAND PUT TO NONAGRICULTURAL USE AREA (1000 ha) \
0	112.64
1	126.13
2	113.25
3	93.54
4	114.89
…	…
1374	0.00
1375	0.00
1376	94.97
1377	121.50
1378	128.50

	PERMANENT PASTURES AREA (1000 ha)	OTHER FALLOW AREA (1000 ha) \
0	0.00	12.75
1	7.82	17.86
2	30.71	23.00
3	27.32	43.99
4	22.17	36.84
…	…	…
1374	0.00	0.00
1375	3.04	4.32
1376	0.00	5.43
1377	8.50	12.00
1378	10.50	17.00

	NET CROPPED AREA (1000 ha)	GROSS CROPPED AREA (1000 ha) \
0	498.96	627.26
1	511.69	646.47
2	448.80	698.74
3	447.61	627.21
4	505.26	736.47
…	…	…
1374	290.10	3.00
1375	249.90	0.00
1376	330.00	3.00
1377	511.00	698.00
1378	536.00	716.00

	CROPING INTENSITY (Percent)	NITROGEN CONSUMPTION (tons) \
0	129.12	31003.0
1	127.30	36622.0
2	160.64	0.0
3	143.94	109664.0

```
4                           150.33                        88821.0
…                              …                              …
1374                          3.00                        24828.0
1375                          3.00                        20322.0
1376                          3.00                        16346.0
1377                        138.47                        39480.0
1378                        137.21                        32472.0

      PHOSPHATE CONSUMPTION (tons)  POTASH CONSUMPTION (tons)  \
0                           9306.0                     1383.0
1                           6835.0                     1363.0
2                          21088.0                     7882.0
3                          37972.0                    18271.0
4                          37910.0                        0.0
…                              …                         …
1374                        8530.0                     7211.0
1375                        7683.0                     5301.0
1376                        8752.0                     4958.0
1377                        6929.0                     2338.0
1378                        3375.0                     1444.0

      TOTAL CONSUMPTION (tons)  TOTAL PER HA OF NCA (Kg per ha)
0                      41684.0                            85.21
1                      44809.0                            90.08
2                          0.0                           303.24
3                     165898.0                           375.97
4                     139778.0                             0.00
…                         …                              …
1374                   40560.0                           142.76
1375                       0.0                           136.79
1376                   30051.0                            94.33
1377                   48744.0                            96.57
1378                   37287.0                            73.08

[1379 rows x 18 columns]
```

[43]:
```
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,random_state=20)
```

[44]:
```
from sklearn.linear_model import LinearRegression
regressor = LinearRegression()
regressor.fit(X_train,y_train)
```

[44]:
```
LinearRegression()
```

[45]:
```
y_pred = regressor.predict(X_test)
y_pred
```

[45]:
```
array([-2.03679101e-13,  5.10293740e-13,  2.00000000e+00,  1.35000000e+02,
       -4.14581778e-12, -1.05683569e-12,  1.73000000e+02,  4.60400000e+03,
       -4.54185036e-12,  2.50000000e+01,  3.80000000e+01,  2.17200000e+03,
       -6.31284460e-13,  2.16500000e+03,  1.70000000e+01,  5.00000000e+00,
       -2.46462026e-13,  3.00000000e+00,  3.05500000e+03,  5.00000000e+00,
        3.90000000e+01,  2.30581668e-13,  1.99000000e+02, -7.41994994e-13,
        3.94000000e+02,  4.00000000e+00,  7.33000000e+02, -2.20953453e-12,
        2.33900358e-13, -3.52400569e-12, -3.97955492e-13,  3.00000000e+00,
        9.02337573e-13,  5.42900000e+03,  1.67717401e-12,  4.00000000e+00,
       -7.30618086e-12, -2.50952901e-12, -2.29054838e-12,  1.20000000e+01,
       -1.31115744e-12,  4.28400000e+03,  5.24000000e+02,  1.99600000e+03,
       -1.97468866e-12,  4.00000000e+00, -5.66452003e-13,  2.00000000e+00,
       -4.37409815e-13, -1.72608588e-12, -8.95753113e-14,  8.52000000e+02,
        9.25000000e+02, -1.44794707e-12,  9.00000000e+00, -4.40783733e-12,
        6.36753272e-13,  2.00000000e+00,  5.50000000e+01,  1.92000000e+02,
        1.09594680e-12,  2.16576793e-13, -1.44655586e-12, -1.28287044e-12,
       -3.33816649e-12, -2.52592440e-12,  7.70700000e+03, -4.12168512e-12,
        2.80400000e+03, -9.68086228e-14, -6.57439568e-13, -3.48168537e-12,
        1.56000000e+02,  2.66200000e+03, -4.79924593e-13,  1.00000000e+01,
       -3.11481731e-12, -1.56005313e-12,  3.54620000e+04, -1.45950704e-12,
       -2.28559896e-12,  1.02000000e+02,  7.36000000e+02, -2.37659514e-12,
       -2.58498210e-12,  4.50000000e+01,  8.75900000e+03, -2.99677807e-12,
        1.77341080e-12,  5.00000000e+00,  3.14000000e+02,  3.21500000e+03,
        7.00000000e+00, -8.05343038e-12, -5.88774146e-12, -5.13342175e-13,
        1.14500000e+03,  5.59095402e-12, -1.51757591e-12,  6.00000000e+00,
       -3.75670544e-12, -2.99653157e-13, -5.86555984e-13,  4.81100000e+03,
       -2.42962481e-12, -3.25050436e-12, -7.40673764e-13, -2.59985687e-12,
        2.60000000e+01, -2.61402271e-12, -1.70718238e-12,  9.79000000e+02,
        1.06000000e+02,  3.00000000e+00, -4.17905197e-13,  1.07000000e+02,
        1.90000000e+01,  1.19700000e+03,  7.00000000e+00,  3.94388799e-13,
        1.98600000e+03,  2.42000000e+02,  4.90000000e+01, -3.41542891e-12,
        3.00000000e+00, -2.02814775e-12,  5.26000000e+02,  6.26800000e+03,
        2.40000000e+01,  2.17232149e-12,  2.00000000e+00, -1.90439105e-12,
       -2.41063990e-12, -1.06781129e-12,  4.01031067e-12,  8.00000000e+00,
        1.38200000e+03,  6.00000000e+00, -1.74980508e-12, -2.94884230e-13,
        4.48192191e-13, -2.14330597e-12, -2.71108446e-12,  2.23000000e+02,
        6.00000000e+00, -5.41780019e-12, -7.20817979e-13,  1.40700000e+03,
        6.80000000e+01, -4.14102726e-12,  2.00000000e+00, -1.71130588e-12,
        1.37900000e+03,  2.10000000e+01, -3.50588097e-12,  6.00000000e+00,
       -5.34040895e-12,  3.70054595e-13, -8.29221914e-13,  5.00000000e+00,
       -1.39585097e-12, -3.28954497e-12, -2.07363747e-12,  1.99901078e-12,
       -2.59586809e-12, -1.67839523e-12,  4.24300000e+03,  1.10000000e+01,
       -1.19155732e-12,  3.08600000e+03,  1.60000000e+01,  1.69887043e-12,
        1.41000000e+02, -5.63522330e-13, -2.22282933e-12, -4.52365587e-12,
```

```
        5.80000000e+01,  9.00000000e+00,  8.00000000e+00,  1.22800000e+03,
        6.65887952e-14,  6.00000000e+02,  5.99000000e+03, -2.96314053e-12,
        3.70000000e+02,  5.00000000e+00,  3.39642957e-13,  3.00000000e+00,
        3.14685866e-12,  1.66324221e-12, -3.62614820e-12,  2.67000000e+02,
       -4.27322123e-13,  7.00000000e+00,  8.07401118e-13,  2.76808274e-13,
        5.70000000e+03, -2.43110157e-12,  5.60000000e+01,  1.09481318e-13,
        9.68325951e-13,  6.00000000e+00,  5.00000000e+00,  6.81000000e+02,
        5.55000000e+02,  2.12000000e+02,  1.23100000e+03, -6.49643887e-12,
        8.00000000e+00,  2.33292041e-12,  8.60000000e+01,  3.81000000e+02,
       -5.19216012e-13,  4.00000000e+00,  2.40000000e+01,  6.82000000e+02,
       -4.09495277e-12,  2.57900000e+03,  2.09900000e+03,  3.30000000e+01,
       -1.85347238e-12, -1.26590060e-12, -1.01258820e-12, -7.26246609e-13,
        6.00000000e+00,  1.24662672e-12,  6.00000000e+00,  1.11000000e+02,
       -1.70782680e-12, -6.60265133e-13,  2.55000000e+02,  1.71891835e-12,
        9.00000000e+00,  2.50000000e+01, -1.85188807e-12,  4.80000000e+01,
        6.60000000e+01,  2.13000000e+02,  1.26635109e-12,  1.49900000e+03,
       -7.92638950e-13,  3.00994393e-12,  5.57200000e+03,  9.33300000e+03,
        2.02000000e+02, -4.86127235e-12, -3.76686919e-12,  1.86000000e+02,
       -5.23514519e-12, -1.41854724e-13, -5.17679726e-14,  3.45776783e-14,
       -2.37678837e-12, -5.62880265e-14, -7.04667802e-13,  9.74965776e-14,
       -2.71768262e-12,  1.30000000e+01, -2.66475088e-12,  6.00000000e+00,
        1.92000000e+02,  1.52000000e+02,  7.00000000e+00,  1.13487282e-12,
       -2.29541343e-12, -9.24000729e-13, -6.96727111e-12, -4.81785678e-13,
        2.30000000e+02,  1.71390000e+04,  1.67000000e+02,  2.50000000e+01,
       -1.58157114e-12,  1.07100000e+03,  9.06000000e+02,  9.64000000e+02])
```

[46]:
```
y_test
```

[46]:
```
array([0.0000e+00, 0.0000e+00, 2.0000e+00, 1.3500e+02, 0.0000e+00,
       0.0000e+00, 1.7300e+02, 4.6040e+03, 0.0000e+00, 2.5000e+01,
       3.8000e+01, 2.1720e+03, 0.0000e+00, 2.1650e+03, 1.7000e+01,
       5.0000e+00, 0.0000e+00, 3.0000e+00, 3.0550e+03, 5.0000e+00,
       3.9000e+01, 0.0000e+00, 1.9900e+02, 0.0000e+00, 3.9400e+02,
       4.0000e+00, 7.3300e+02, 0.0000e+00, 0.0000e+00, 0.0000e+00,
       0.0000e+00, 3.0000e+00, 0.0000e+00, 5.4290e+03, 0.0000e+00,
       4.0000e+00, 0.0000e+00, 0.0000e+00, 0.0000e+00, 1.2000e+01,
       0.0000e+00, 4.2840e+03, 5.2400e+02, 1.9960e+03, 0.0000e+00,
       4.0000e+00, 0.0000e+00, 2.0000e+00, 0.0000e+00, 0.0000e+00,
       0.0000e+00, 8.5200e+02, 9.2500e+02, 0.0000e+00, 9.0000e+00,
       0.0000e+00, 0.0000e+00, 2.0000e+00, 5.5000e+01, 1.9200e+02,
       0.0000e+00, 0.0000e+00, 0.0000e+00, 0.0000e+00, 0.0000e+00,
       0.0000e+00, 7.7070e+03, 0.0000e+00, 2.8040e+03, 0.0000e+00,
       0.0000e+00, 0.0000e+00, 1.5600e+02, 2.6620e+03, 0.0000e+00,
       1.0000e+01, 0.0000e+00, 0.0000e+00, 3.5462e+04, 0.0000e+00,
       0.0000e+00, 1.0200e+02, 7.3600e+02, 0.0000e+00, 0.0000e+00,
       4.5000e+01, 8.7590e+03, 0.0000e+00, 0.0000e+00, 5.0000e+00,
       3.1400e+02, 3.2150e+03, 7.0000e+00, 0.0000e+00, 0.0000e+00,
```

```
0.0000e+00, 1.1450e+03, 0.0000e+00, 0.0000e+00, 6.0000e+00,
0.0000e+00, 0.0000e+00, 0.0000e+00, 4.8110e+03, 0.0000e+00,
0.0000e+00, 0.0000e+00, 0.0000e+00, 2.6000e+01, 0.0000e+00,
0.0000e+00, 9.7900e+02, 1.0600e+02, 3.0000e+00, 0.0000e+00,
1.0700e+02, 1.9000e+01, 1.1970e+03, 7.0000e+00, 0.0000e+00,
1.9860e+03, 2.4200e+02, 4.9000e+01, 0.0000e+00, 3.0000e+00,
0.0000e+00, 5.2600e+02, 6.2680e+03, 2.4000e+01, 0.0000e+00,
2.0000e+00, 0.0000e+00, 0.0000e+00, 0.0000e+00, 0.0000e+00,
8.0000e+00, 1.3820e+03, 6.0000e+00, 0.0000e+00, 0.0000e+00,
0.0000e+00, 0.0000e+00, 0.0000e+00, 2.2300e+02, 6.0000e+00,
0.0000e+00, 0.0000e+00, 1.4070e+03, 6.8000e+01, 0.0000e+00,
2.0000e+00, 0.0000e+00, 1.3790e+03, 2.1000e+01, 0.0000e+00,
6.0000e+00, 0.0000e+00, 0.0000e+00, 0.0000e+00, 5.0000e+00,
0.0000e+00, 0.0000e+00, 0.0000e+00, 0.0000e+00, 0.0000e+00,
0.0000e+00, 4.2430e+03, 1.1000e+01, 0.0000e+00, 3.0860e+03,
1.6000e+01, 0.0000e+00, 1.4100e+02, 0.0000e+00, 0.0000e+00,
0.0000e+00, 5.8000e+01, 9.0000e+00, 8.0000e+00, 1.2280e+03,
0.0000e+00, 6.0000e+02, 5.9900e+03, 0.0000e+00, 3.7000e+02,
5.0000e+00, 0.0000e+00, 3.0000e+00, 0.0000e+00, 0.0000e+00,
0.0000e+00, 2.6700e+02, 0.0000e+00, 7.0000e+00, 0.0000e+00,
0.0000e+00, 5.7000e+03, 0.0000e+00, 5.6000e+01, 0.0000e+00,
0.0000e+00, 6.0000e+00, 5.0000e+00, 6.8100e+02, 5.5500e+02,
2.1200e+02, 1.2310e+03, 0.0000e+00, 8.0000e+00, 0.0000e+00,
8.6000e+01, 3.8100e+02, 0.0000e+00, 4.0000e+00, 2.4000e+01,
6.8200e+02, 0.0000e+00, 2.5790e+03, 2.0990e+03, 3.3000e+01,
0.0000e+00, 0.0000e+00, 0.0000e+00, 0.0000e+00, 6.0000e+00,
0.0000e+00, 6.0000e+00, 1.1100e+02, 0.0000e+00, 0.0000e+00,
2.5500e+02, 0.0000e+00, 9.0000e+00, 2.5000e+01, 0.0000e+00,
4.8000e+01, 6.6000e+01, 2.1300e+02, 0.0000e+00, 1.4990e+03,
0.0000e+00, 0.0000e+00, 5.5720e+03, 9.3330e+03, 2.0200e+02,
0.0000e+00, 0.0000e+00, 1.8600e+02, 0.0000e+00, 0.0000e+00,
0.0000e+00, 0.0000e+00, 0.0000e+00, 0.0000e+00, 0.0000e+00,
0.0000e+00, 0.0000e+00, 1.3000e+01, 0.0000e+00, 6.0000e+00,
1.9200e+02, 1.5200e+02, 7.0000e+00, 0.0000e+00, 0.0000e+00,
0.0000e+00, 0.0000e+00, 0.0000e+00, 2.3000e+02, 1.7139e+04,
1.6700e+02, 2.5000e+01, 0.0000e+00, 1.0710e+03, 9.0600e+02,
9.6400e+02])
```

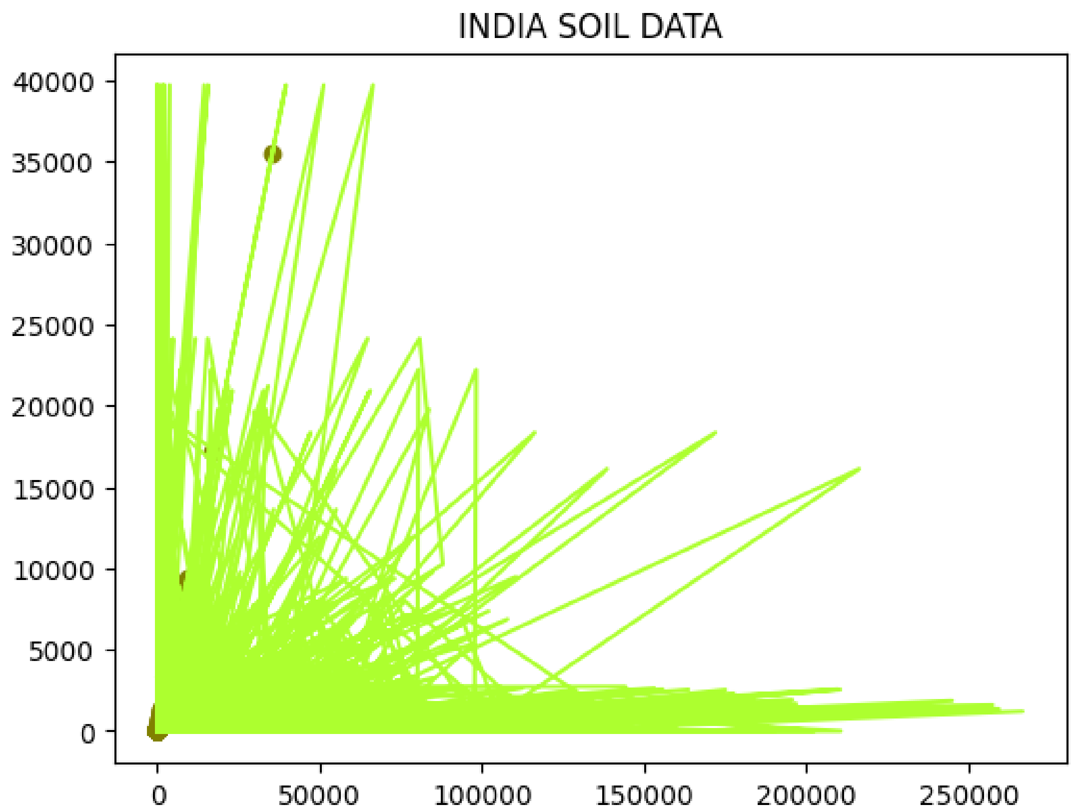
HERE y_test is the actual data array of india soil train data and y_pred is the predicted data array

[76]:
```
plt.scatter(y_pred, y_test, color='olive') # plotting the observation line

plt.plot(X_train, regressor.predict(X_train), color='greenyellow') # plotting␣
↪the regression line

plt.title("INDIA SOIL DATA")
```

```
#plt.xlabel("Years of experience")
#plt.ylabel("Salaries")
plt.show()
```



Above observation describes how due to difference in forest area for crop production the comsumption, soils nutritional content and profit of cotton has declined over the years

MODEL EVALUATION

[51]:
```
# predicting the mean absolute error
from sklearn.metrics import mean_absolute_error
print("MAE",mean_absolute_error(y_test,y_pred))
```

```
MAE 2.4755320017743695e-12
```

[52]:
```
# predicting the mean squared error
from sklearn.metrics import mean_squared_error
print("MSE",mean_squared_error(y_test,y_pred))
```

```
MSE 1.4327796219398725e-23
```

[53]:
```
# predicting the root mean squared error
print("RMSE",np.sqrt(mean_squared_error(y_test,y_pred)))
```

```
RMSE 3.7852075530145935e-12
```

[ ]:
```
# predicting the root mean squared log error
print("RMSE",np.log(np.sqrt(mean_squared_error(y_test,y_pred))))
```

[54]:
```
from sklearn.metrics import r2_score
r2 = r2_score(y_test,y_pred)
print(r2)
```

```
1.0
```

Since the r2 score is 1 it indicates us that the linear regression model is improved.

LINEAR REGRESSION ON USA COTTON PRODUCTION DATA

[40]:
```
# using linear regression algorithm
df4 = pd.read_csv('USA_train.csv')
df4
```

[40]:
```
     Year Planted (1000 Acres) Harvested (1000 Acres)  \
0    1975                  385                    370
1    1975                  700                    680
2    1975                  NaN                    268
3    1975                  900                    875
4    1975                    4                      4
..    …                     …                      …
555  2002                  200                    180
556  2002                  290                    200
557  2002                  565                    530
558  2002                5,600                  4,500
559  2002                  100                     98

     Yield (Pounds/ Harvested Area)  Average Temperature Value  \
0                               406                  66.825000
1                               486                  63.875000
2                              1028                  61.891667
3                              1074                        NaN
4                               347                  73.208333
..                               …                          …
555                             560                  60.416667
556                             316                  65.700000
557                             743                  61.725000
558                             541                  67.025000
559                             469                  60.558333

     Average Temperature Anomaly  Maximum Temperature Value  \
```

```
0                     1.566667                    75.066667
1                     2.225000                    73.808333
2                    -0.733333                    74.483333
3                    -0.383333                    72.700000
4                     4.133333                    83.983333
..                         …                          …
555                   2.433333                    75.133333
556                   1.683333                    76.058333
557                   1.683333                    72.566667
558                   2.716667                    80.941667
559                   1.975000                    71.800000

     Maximum Temperature Anomaly  Minimum Temperature Value  \
0                     2.266667                    54.566667
1                     1.208333                    52.975000
2                    -0.291667                    45.283333
3                    -0.466667                    46.825000
4                     2.308333                    64.408333
..                         …                          …
555                   1.116667                    50.708333
556                   2.625000                    53.325000
557                   4.408333                    48.841667
558                   0.333333                    57.100000
559                   3.233333                    49.350000

     Minimum Temperature Anomaly  …  Heating Degree Days Value  \
0                     2.866667  …                 224.916667
1                     3.258333  …                 299.250000
2                     0.850000  …                 211.166667
3                     1.691667  …                 303.083333
4                     2.941667  …                  49.916667
..                         …   …                       …
555                   2.791667  …                 316.333333
556                   2.741667  …                 217.916667
557                   4.941667  …                 318.166667
558                   3.116667  …                 170.000000
559                   3.758333  …                 347.833333

     Heating Degree Days Anomaly  Palmer Drought Severity Index (PDSI) Value  \
0                    -1.083333                                    8.161667
1                     8.750000                                    6.674167
2                    43.416667                                    0.584167
3                          NaN                                    2.650000
4                    -6.666667                                    1.191667
..                         …                                          …
555                  11.916667                                    4.065000
556                  -1.666667                                         NaN
```

```
557                -2.500000                                       3.680000
558                 8.750000                                       4.174167
559               -16.000000                                            NaN

     Palmer Drought Severity Index (PDSI) Anomaly  \
0                                        6.840833
1                                        6.331667
2                                        0.403333
3                                        4.479167
4                                        1.801667
..                                            …
555                                      0.377500
556                                      0.236667
557                                      0.880833
558                                      1.235833
559                                      0.539167

     Palmer Hydrological Drought Index (PHDI) Value  \
0                                                NaN
1                                           5.414167
2                                           1.584167
3                                           1.613333
4                                           4.085000
..                                               …
555                                         1.434167
556                                        -1.291667
557                                         2.127500
558                                         0.911667
559                                         1.013333

     Palmer Hydrological Drought Index (PHDI) Anomaly  \
0                                             4.684167
1                                             4.894167
2                                             0.375000
3                                             4.324167
4                                             1.762500
..                                                 …
555                                           1.620833
556                                          -0.115833
557                                           3.180000
558                                           1.660833
559                                           0.632500

     Palmer Modified Drought Index (PMDI) Value  \
0                                      6.132500
1                                      5.098333
2                                      1.140833
```

```
3                                              1.405000
4                                              2.155000
..                                                  ...
555                                            1.524167
556                                           -0.941667
557                                            2.747500
558                                            1.031667
559                                            0.455833

     Palmer Modified Drought Index (PMDI) Anomaly  Palmer Z-Index Value  \
0                                        5.821667              4.857500
1                                        3.680000              1.537500
2                                        2.805000              3.648333
3                                        2.100833              3.068333
4                                        1.778333              3.239167
..                                            ...                   ...
555                                           NaN              3.279167
556                                     -0.740000              2.412500
557                                      3.960000              4.494167
558                                      2.886667              4.301667
559                                     -1.858333              0.540833

     Palmer Z-Index Anomaly
0                  2.718333
1                  3.382500
2                  0.893333
3                  2.322500
4                  3.455833
..                      ...
555                4.037500
556                1.468333
557                3.199167
558                3.251667
559                3.409167

[560 rows x 24 columns]
```

[41]:
```python
X = df4.iloc[:,:-1].values  #independent variable array
y = df4.iloc[:,1].values

df4=df4.replace({'NAN':0})
df4  = df4.fillna(0)   # replacing nan values with zeros
df4
```

[41]:
```
     Year Planted (1000 Acres) Harvested (1000 Acres)  \
0    1975                  385                    370
1    1975                  700                    680
```

```
2    1975                       0                      268
3    1975                     900                      875
4    1975                       4                        4
..    …                        …                        …
555  2002                     200                      180
556  2002                     290                      200
557  2002                     565                      530
558  2002                   5,600                    4,500
559  2002                     100                       98

     Yield (Pounds/ Harvested Area)  Average Temperature Value  \
0                               406                  66.825000
1                               486                  63.875000
2                              1028                  61.891667
3                              1074                   0.000000
4                               347                  73.208333
..                                …                          …
555                             560                  60.416667
556                             316                  65.700000
557                             743                  61.725000
558                             541                  67.025000
559                             469                  60.558333

     Average Temperature Anomaly  Maximum Temperature Value  \
0                       1.566667                  75.066667
1                       2.225000                  73.808333
2                      -0.733333                  74.483333
3                      -0.383333                  72.700000
4                       4.133333                  83.983333
..                             …                          …
555                     2.433333                  75.133333
556                     1.683333                  76.058333
557                     1.683333                  72.566667
558                     2.716667                  80.941667
559                     1.975000                  71.800000

     Maximum Temperature Anomaly  Minimum Temperature Value  \
0                       2.266667                  54.566667
1                       1.208333                  52.975000
2                      -0.291667                  45.283333
3                      -0.466667                  46.825000
4                       2.308333                  64.408333
..                             …                          …
555                     1.116667                  50.708333
556                     2.625000                  53.325000
557                     4.408333                  48.841667
558                     0.333333                  57.100000
```

```
559                   3.233333                 49.350000

     Minimum Temperature Anomaly  …  Heating Degree Days Value  \
0                       2.866667  …                 224.916667
1                       3.258333  …                 299.250000
2                       0.850000  …                 211.166667
3                       1.691667  …                 303.083333
4                       2.941667  …                  49.916667
..                           …  …                          …
555                     2.791667  …                 316.333333
556                     2.741667  …                 217.916667
557                     4.941667  …                 318.166667
558                     3.116667  …                 170.000000
559                     3.758333  …                 347.833333

     Heating Degree Days Anomaly  Palmer Drought Severity Index (PDSI) Value  \
0                      -1.083333                                    8.161667
1                       8.750000                                    6.674167
2                      43.416667                                    0.584167
3                       0.000000                                    2.650000
4                      -6.666667                                    1.191667
..                           …                                          …
555                    11.916667                                    4.065000
556                    -1.666667                                    0.000000
557                    -2.500000                                    3.680000
558                     8.750000                                    4.174167
559                   -16.000000                                    0.000000

     Palmer Drought Severity Index (PDSI) Anomaly  \
0                                        6.840833
1                                        6.331667
2                                        0.403333
3                                        4.479167
4                                        1.801667
..                                            …
555                                      0.377500
556                                      0.236667
557                                      0.880833
558                                      1.235833
559                                      0.539167

     Palmer Hydrological Drought Index (PHDI) Value  \
0                                          0.000000
1                                          5.414167
2                                          1.584167
3                                          1.613333
4                                          4.085000
```

```
..                                                …
555                                        1.434167
556                                       -1.291667
557                                        2.127500
558                                        0.911667
559                                        1.013333

     Palmer Hydrological Drought Index (PHDI) Anomaly  \
0                                            4.684167
1                                            4.894167
2                                            0.375000
3                                            4.324167
4                                            1.762500
..                                                  …
555                                          1.620833
556                                         -0.115833
557                                          3.180000
558                                          1.660833
559                                          0.632500

     Palmer Modified Drought Index (PMDI) Value  \
0                                      6.132500
1                                      5.098333
2                                      1.140833
3                                      1.405000
4                                      2.155000
..                                            …
555                                    1.524167
556                                   -0.941667
557                                    2.747500
558                                    1.031667
559                                    0.455833

     Palmer Modified Drought Index (PMDI) Anomaly  Palmer Z-Index Value  \
0                                        5.821667              4.857500
1                                        3.680000              1.537500
2                                        2.805000              3.648333
3                                        2.100833              3.068333
4                                        1.778333              3.239167
..                                              …                     …
555                                      0.000000              3.279167
556                                     -0.740000              2.412500
557                                      3.960000              4.494167
558                                      2.886667              4.301667
559                                     -1.858333              0.540833

     Palmer Z-Index Anomaly
```

```
0                  2.718333
1                  3.382500
2                  0.893333
3                  2.322500
4                  3.455833
..                      …
555                4.037500
556                1.468333
557                3.199167
558                3.251667
559                3.409167

[560 rows x 24 columns]
```

[55]:
```python
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,random_state=20)

from sklearn.linear_model import LinearRegression
regressor = LinearRegression()
regressor.fit(X_train,y_train)
```

[55]:
```
LinearRegression()
```

[56]:
```python
y_pred = regressor.predict(X_test)
y_pred
```

[56]:
```
array([-2.03679101e-13,  5.10293740e-13,  2.00000000e+00,  1.35000000e+02,
       -4.14581778e-12, -1.05683569e-12,  1.73000000e+02,  4.60400000e+03,
       -4.54185036e-12,  2.50000000e+01,  3.80000000e+01,  2.17200000e+03,
       -6.31284460e-13,  2.16500000e+03,  1.70000000e+01,  5.00000000e+00,
       -2.46462026e-13,  3.00000000e+00,  3.05500000e+03,  5.00000000e+00,
        3.90000000e+01,  2.30581668e-13,  1.99000000e+02, -7.41994994e-13,
        3.94000000e+02,  4.00000000e+00,  7.33000000e+02, -2.20953453e-12,
        2.33900358e-13, -3.52400569e-12, -3.97955492e-13,  3.00000000e+00,
        9.02337573e-13,  5.42900000e+03,  1.67717401e-12,  4.00000000e+00,
       -7.30618086e-12, -2.50952901e-12, -2.29054838e-12,  1.20000000e+01,
       -1.31115744e-12,  4.28400000e+03,  5.24000000e+02,  1.99600000e+03,
       -1.97468866e-12,  4.00000000e+00, -5.66452003e-13,  2.00000000e+00,
       -4.37409815e-13, -1.72608588e-12, -8.95753113e-14,  8.52000000e+02,
        9.25000000e+02, -1.44794707e-12,  9.00000000e+00, -4.40783733e-12,
        6.36753272e-13,  2.00000000e+00,  5.50000000e+01,  1.92000000e+02,
        1.09594680e-12,  2.16576793e-13, -1.44655586e-12, -1.28287044e-12,
       -3.33816649e-12, -2.52592440e-12,  7.70700000e+03, -4.12168512e-12,
        2.80400000e+03, -9.68086228e-14, -6.57439568e-13, -3.48168537e-12,
        1.56000000e+02,  2.66200000e+03, -4.79924593e-13,  1.00000000e+01,
       -3.11481731e-12, -1.56005313e-12,  3.54620000e+04, -1.45950704e-12,
```

```
-2.28559896e-12,  1.02000000e+02,  7.36000000e+02, -2.37659514e-12,
-2.58498210e-12,  4.50000000e+01,  8.75900000e+03, -2.99677807e-12,
 1.77341080e-12,  5.00000000e+00,  3.14000000e+02,  3.21500000e+03,
 7.00000000e+00, -8.05343038e-12, -5.88774146e-12, -5.13342175e-13,
 1.14500000e+03,  5.59095402e-12, -1.51757591e-12,  6.00000000e+00,
-3.75670544e-12, -2.99653157e-13, -5.86555984e-13,  4.81100000e+03,
-2.42962481e-12, -3.25050436e-12, -7.40673764e-13, -2.59985687e-12,
 2.60000000e+01, -2.61402271e-12, -1.70718238e-12,  9.79000000e+02,
 1.06000000e+02,  3.00000000e+00, -4.17905197e-13,  1.07000000e+02,
 1.90000000e+01,  1.19700000e+03,  7.00000000e+00,  3.94388799e-13,
 1.98600000e+03,  2.42000000e+02,  4.90000000e+01, -3.41542891e-12,
 3.00000000e+00, -2.02814775e-12,  5.26000000e+02,  6.26800000e+03,
 2.40000000e+01,  2.17232149e-12,  2.00000000e+00, -1.90439105e-12,
-2.41063990e-12, -1.06781129e-12,  4.01031067e-12,  8.00000000e+00,
 1.38200000e+03,  6.00000000e+00, -1.74980508e-12, -2.94884230e-13,
 4.48192191e-13, -2.14330597e-12, -2.71108446e-12,  2.23000000e+02,
 6.00000000e+00, -5.41780019e-12, -7.20817979e-13,  1.40700000e+03,
 6.80000000e+01, -4.14102726e-12,  2.00000000e+00, -1.71130588e-12,
 1.37900000e+03,  2.10000000e+01, -3.50588097e-12,  6.00000000e+00,
-5.34040895e-12,  3.70054595e-13, -8.29221914e-13,  5.00000000e+00,
-1.39585097e-12, -3.28954497e-12, -2.07363747e-12,  1.99901078e-12,
-2.59586809e-12, -1.67839523e-12,  4.24300000e+03,  1.10000000e+01,
-1.19155732e-12,  3.08600000e+03,  1.60000000e+01,  1.69887043e-12,
 1.41000000e+02, -5.63522330e-13, -2.22282933e-12, -4.52365587e-12,
 5.80000000e+01,  9.00000000e+00,  8.00000000e+00,  1.22800000e+03,
 6.65887952e-14,  6.00000000e+02,  5.99000000e+03, -2.96314053e-12,
 3.70000000e+02,  5.00000000e+00,  3.39642957e-13,  3.00000000e+00,
 3.14685866e-12,  1.66324221e-12, -3.62614820e-12,  2.67000000e+02,
-4.27322123e-13,  7.00000000e+00,  8.07401118e-13,  2.76808274e-13,
 5.70000000e+03, -2.43110157e-12,  5.60000000e+01,  1.09481318e-13,
 9.68325951e-13,  6.00000000e+00,  5.00000000e+00,  6.81000000e+02,
 5.55000000e+02,  2.12000000e+02,  1.23100000e+03, -6.49643887e-12,
 8.00000000e+00,  2.33292041e-12,  8.60000000e+01,  3.81000000e+02,
-5.19216012e-13,  4.00000000e+00,  2.40000000e+01,  6.82000000e+02,
-4.09495277e-12,  2.57900000e+03,  2.09900000e+03,  3.30000000e+01,
-1.85347238e-12, -1.26590060e-12, -1.01258820e-12, -7.26246609e-13,
 6.00000000e+00,  1.24662672e-12,  6.00000000e+00,  1.11000000e+02,
-1.70782680e-12, -6.60265133e-13,  2.55000000e+02,  1.71891835e-12,
 9.00000000e+00,  2.50000000e+01, -1.85188807e-12,  4.80000000e+01,
 6.60000000e+01,  2.13000000e+02,  1.26635109e-12,  1.49900000e+03,
-7.92638950e-13,  3.00994393e-12,  5.57200000e+03,  9.33300000e+03,
 2.02000000e+02, -4.86127235e-12, -3.76686919e-12,  1.86000000e+02,
-5.23514519e-12, -1.41854724e-13, -5.17679726e-14,  3.45776783e-14,
-2.37678837e-12, -5.62880265e-14, -7.04667802e-13,  9.74965776e-14,
-2.71768262e-12,  1.30000000e+01, -2.66475088e-12,  6.00000000e+00,
 1.92000000e+02,  1.52000000e+02,  7.00000000e+00,  1.13487282e-12,
-2.29541343e-12, -9.24000729e-13, -6.96727111e-12, -4.81785678e-13,
```

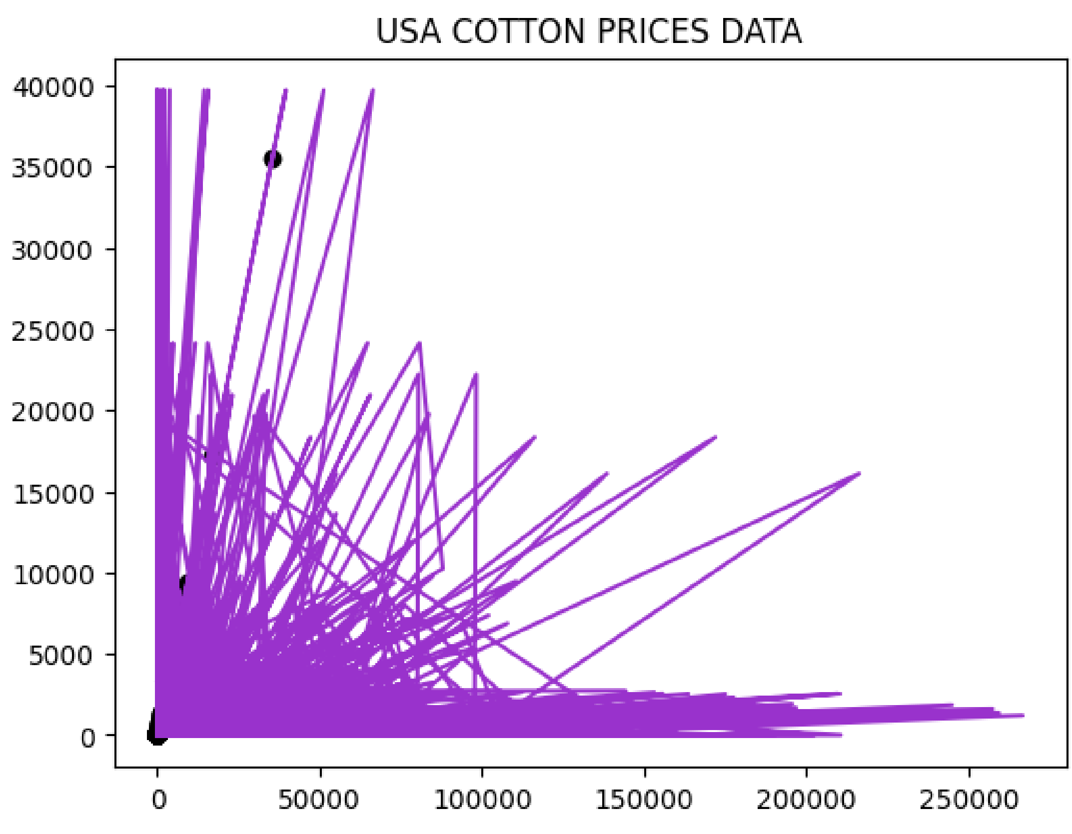
```
 2.30000000e+02,  1.71390000e+04,  1.67000000e+02,  2.50000000e+01,
-1.58157114e-12,  1.07100000e+03,  9.06000000e+02,  9.64000000e+02])
```

[57]:
```
y_test
```

[57]:
```
array([0.0000e+00, 0.0000e+00, 2.0000e+00, 1.3500e+02, 0.0000e+00,
       0.0000e+00, 1.7300e+02, 4.6040e+03, 0.0000e+00, 2.5000e+01,
       3.8000e+01, 2.1720e+03, 0.0000e+00, 2.1650e+03, 1.7000e+01,
       5.0000e+00, 0.0000e+00, 3.0000e+00, 3.0550e+03, 5.0000e+00,
       3.9000e+01, 0.0000e+00, 1.9900e+02, 0.0000e+00, 3.9400e+02,
       4.0000e+00, 7.3300e+02, 0.0000e+00, 0.0000e+00, 0.0000e+00,
       0.0000e+00, 3.0000e+00, 0.0000e+00, 5.4290e+03, 0.0000e+00,
       4.0000e+00, 0.0000e+00, 0.0000e+00, 0.0000e+00, 1.2000e+01,
       0.0000e+00, 4.2840e+03, 5.2400e+02, 1.9960e+03, 0.0000e+00,
       4.0000e+00, 0.0000e+00, 2.0000e+00, 0.0000e+00, 0.0000e+00,
       0.0000e+00, 8.5200e+02, 9.2500e+02, 0.0000e+00, 9.0000e+00,
       0.0000e+00, 0.0000e+00, 2.0000e+00, 5.5000e+01, 1.9200e+02,
       0.0000e+00, 0.0000e+00, 0.0000e+00, 0.0000e+00, 0.0000e+00,
       0.0000e+00, 7.7070e+03, 0.0000e+00, 2.8040e+03, 0.0000e+00,
       0.0000e+00, 0.0000e+00, 1.5600e+02, 2.6620e+03, 0.0000e+00,
       1.0000e+01, 0.0000e+00, 0.0000e+00, 3.5462e+04, 0.0000e+00,
       0.0000e+00, 1.0200e+02, 7.3600e+02, 0.0000e+00, 0.0000e+00,
       4.5000e+01, 8.7590e+03, 0.0000e+00, 0.0000e+00, 5.0000e+00,
       3.1400e+02, 3.2150e+03, 7.0000e+00, 0.0000e+00, 0.0000e+00,
       0.0000e+00, 1.1450e+03, 0.0000e+00, 0.0000e+00, 6.0000e+00,
       0.0000e+00, 0.0000e+00, 0.0000e+00, 4.8110e+03, 0.0000e+00,
       0.0000e+00, 0.0000e+00, 0.0000e+00, 2.6000e+01, 0.0000e+00,
       0.0000e+00, 9.7900e+02, 1.0600e+02, 3.0000e+00, 0.0000e+00,
       1.0700e+02, 1.9000e+01, 1.1970e+03, 7.0000e+00, 0.0000e+00,
       1.9860e+03, 2.4200e+02, 4.9000e+01, 0.0000e+00, 3.0000e+00,
       0.0000e+00, 5.2600e+02, 6.2680e+03, 2.4000e+01, 0.0000e+00,
       2.0000e+00, 0.0000e+00, 0.0000e+00, 0.0000e+00, 0.0000e+00,
       8.0000e+00, 1.3820e+03, 6.0000e+00, 0.0000e+00, 0.0000e+00,
       0.0000e+00, 0.0000e+00, 0.0000e+00, 2.2300e+02, 6.0000e+00,
       0.0000e+00, 0.0000e+00, 1.4070e+03, 6.8000e+01, 0.0000e+00,
       2.0000e+00, 0.0000e+00, 1.3790e+03, 2.1000e+01, 0.0000e+00,
       6.0000e+00, 0.0000e+00, 0.0000e+00, 0.0000e+00, 5.0000e+00,
       0.0000e+00, 0.0000e+00, 0.0000e+00, 0.0000e+00, 0.0000e+00,
       0.0000e+00, 4.2430e+03, 1.1000e+01, 0.0000e+00, 3.0860e+03,
       1.6000e+01, 0.0000e+00, 1.4100e+02, 0.0000e+00, 0.0000e+00,
       0.0000e+00, 5.8000e+01, 9.0000e+00, 8.0000e+00, 1.2280e+03,
       0.0000e+00, 6.0000e+02, 5.9900e+03, 0.0000e+00, 3.7000e+02,
       5.0000e+00, 0.0000e+00, 3.0000e+00, 0.0000e+00, 0.0000e+00,
       0.0000e+00, 2.6700e+02, 0.0000e+00, 7.0000e+00, 0.0000e+00,
       0.0000e+00, 5.7000e+03, 0.0000e+00, 5.6000e+01, 0.0000e+00,
       0.0000e+00, 6.0000e+00, 5.0000e+00, 6.8100e+02, 5.5500e+02,
       2.1200e+02, 1.2310e+03, 0.0000e+00, 8.0000e+00, 0.0000e+00,
```

```
8.6000e+01, 3.8100e+02, 0.0000e+00, 4.0000e+00, 2.4000e+01,
6.8200e+02, 0.0000e+00, 2.5790e+03, 2.0990e+03, 3.3000e+01,
0.0000e+00, 0.0000e+00, 0.0000e+00, 0.0000e+00, 6.0000e+00,
0.0000e+00, 6.0000e+00, 1.1100e+02, 0.0000e+00, 0.0000e+00,
2.5500e+02, 0.0000e+00, 9.0000e+00, 2.5000e+01, 0.0000e+00,
4.8000e+01, 6.6000e+01, 2.1300e+02, 0.0000e+00, 1.4990e+03,
0.0000e+00, 0.0000e+00, 5.5720e+03, 9.3330e+03, 2.0200e+02,
0.0000e+00, 0.0000e+00, 1.8600e+02, 0.0000e+00, 0.0000e+00,
0.0000e+00, 0.0000e+00, 0.0000e+00, 0.0000e+00, 0.0000e+00,
0.0000e+00, 0.0000e+00, 1.3000e+01, 0.0000e+00, 6.0000e+00,
1.9200e+02, 1.5200e+02, 7.0000e+00, 0.0000e+00, 0.0000e+00,
0.0000e+00, 0.0000e+00, 0.0000e+00, 2.3000e+02, 1.7139e+04,
1.6700e+02, 2.5000e+01, 0.0000e+00, 1.0710e+03, 9.0600e+02,
9.6400e+02])
```

[62]:
```
plt.scatter(y_pred, y_test, color='black') # plotting the observation line
plt.plot(X_train, regressor.predict(X_train), color='darkorchid') # plotting␣
 ↪the regression line
plt.title("USA COTTON PRICES DATA")
plt.show()
```



[63]:
```
# predicting the mean absolute error
from sklearn.metrics import mean_absolute_error
print("MAE",mean_absolute_error(y_test,y_pred))
```

```
MAE 2.4755320017743695e-12
```

[64]:
```
# predicting the mean squared error
from sklearn.metrics import mean_squared_error
print("MSE",mean_squared_error(y_test,y_pred))
```

```
MSE 1.4327796219398725e-23
```

[65]:
```
# predicting the root mean squared error
print("RMSE",np.sqrt(mean_squared_error(y_test,y_pred)))
```

```
RMSE 3.7852075530145935e-12
```

[66]:
```
# predicting the root mean squared log error
print("RMSE",np.log(np.sqrt(mean_squared_error(y_test,y_pred))))
```

```
RMSE -26.299920394871595
```

[67]:
```
from sklearn.metrics import r2_score
r2 = r2_score(y_test,y_pred)
print(r2)
```

```
1.0
```

Results indicate that the soil variation in India has been deteriorated by climatic changes since the early 2000s. As compared to the late 1975s there was a presense of rich soil nutrients in the soil which were producing maximum cotton production in and profitting USA with maximum prices. But, due to the climatic conditions globally the soil quality has been degraded which has vastly impacted cotton production due to which cotton prices have dropped along with way decreasing at the early peak of the 21st century. If the conditions of climate change are kept constant then there is high possibility of global warning. Hence, to combat such issues sustainability measurements are implemented.