Importing Libraries

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

Importing Datasets

```
In [2]:
    df=pd.read_csv("rainfall in india.csv")
    df
```

Out[2]:		index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	ОСТ
	0	0	ANDAMAN & NICOBAR ISLANDS	1901	49.2	87.1	29.2	2.3	528.8	517.5	365.1	481.1	332.6	388.5
	1	1	ANDAMAN & NICOBAR ISLANDS	1902	0.0	159.8	12.2	0.0	446.1	537.1	228.9	753.7	666.2	197.2
	2	2	ANDAMAN & NICOBAR ISLANDS	1903	12.7	144.0	0.0	1.0	235.1	479.9	728.4	326.7	339.0	181.2
	3	3	ANDAMAN & NICOBAR ISLANDS	1904	9.4	14.7	0.0	202.4	304.5	495.1	502.0	160.1	820.4	222.2
	4	4	ANDAMAN & NICOBAR ISLANDS	1905	1.3	0.0	3.3	26.9	279.5	628.7	368.7	330.5	297.0	260.7
	•••		•••											
	4111	4111	LAKSHADWEEP	2011	5.1	2.8	3.1	85.9	107.2	153.6	350.2	254.0	255.2	117.4
	4112	4112	LAKSHADWEEP	2012	19.2	0.1	1.6	76.8	21.2	327.0	231.5	381.2	179.8	145.9
	4113	4113	LAKSHADWEEP	2013	26.2	34.4	37.5	5.3	88.3	426.2	296.4	154.4	180.0	72.8
	4114	4114	LAKSHADWEEP	2014	53.2	16.1	4.4	14.9	57.4	244.1	116.1	466.1	132.2	169.2
	4115	4115	LAKSHADWEEP	2015	2.2	0.5	3.7	87.1	133.1	296.6	257.5	146.4	160.4	165.4
	4446		00 1											

4116 rows × 20 columns

Data Cleaning and Data Preprocessing

```
In [3]:
    df=df.dropna()
    df
```

Out[3]: SUBDIVISION YEAR JAN index FEB MAR APR MAY JUN JUL AUG SEP OCT ANDAMAN & 0 0 **NICOBAR** 1901 49.2 87.1 29.2 2.3 528.8 517.5 365.1 481.1 332.6 388.5 **ISLANDS** ANDAMAN & **NICOBAR** 0.0 159.8 12.2 0.0 446.1 537.1 228.9 753.7 666.2 197.2 1 1 1902 **ISLANDS** ANDAMAN & 2 2 **NICOBAR** 1903 12.7 144.0 0.0 1.0 235.1 479.9 728.4 326.7 339.0 181.2 **ISLANDS** ANDAMAN & 3 202.4 304.5 495.1 502.0 160.1 820.4 222.2 3 **NICOBAR** 1904 9.4 14.7 0.0 **ISLANDS** ANDAMAN & 4 **NICOBAR** 1905 1.3 0.0 3.3 26.9 279.5 628.7 368.7 330.5 297.0 260.7 **ISLANDS** ... 4111 LAKSHADWEEP 2011 5.1 2.8 3.1 85.9 107.2 153.6 350.2 254.0 255.2 117.4 4111 21.2 327.0 231.5 381.2 179.8 145.9 4112 4112 LAKSHADWEEP 2012 19.2 0.1 1.6 76.8 4113 LAKSHADWEEP 2013 26.2 34.4 37.5 5.3 88.3 426.2 296.4 154.4 180.0 4113 72.8 4114 LAKSHADWEEP 2014 53.2 57.4 244.1 116.1 466.1 132.2 169.2 4114 16.1 4.4 14.9 4115 LAKSHADWEEP 2015 2.2 0.5 3.7 87.1 133.1 296.6 257.5 146.4 160.4 165.4 4090 rows × 20 columns In [4]: df.columns Out[4]: Index(['index', 'SUBDIVISION', 'YEAR', 'JAN', 'FEB', 'MAR', 'APR', 'MAY', 'JUN', 'JUL', 'AUG', 'SEP', 'OCT', 'NOV', 'DEC', 'ANNUAL', 'Jan-Feb', 'Mar-May', 'Jun-Sep', 'Oct-Dec'], dtype='object') In [5]: df.info() <class 'pandas.core.frame.DataFrame'> Int64Index: 4090 entries, 0 to 4115 Data columns (total 20 columns): Non-Null Count Dtype # Column ---0 index 4090 non-null int64 1 SUBDIVISION 4090 non-null object 2 YEAR 4090 non-null int64 3 JAN 4090 non-null float64 4 FEB 4090 non-null float64 5 MAR 4090 non-null float64 6 APR 4090 non-null float64 7 MAY 4090 non-null float64 8 JUN 4090 non-null float64 9 JUL 4090 non-null float64 10 AUG 4090 non-null float64 11 SEP 4090 non-null float64

4090 non-null

float64

OCT

```
NOV
                  4090 non-null
                                  float64
13
14
                  4090 non-null
                                  float64
    DEC
                                  float64
15
    ANNUAL
                  4090 non-null
                                  float64
16
    Jan-Feb
                  4090 non-null
                                  float64
17
    Mar-May
                  4090 non-null
    Jun-Sep
                  4090 non-null
                                  float64
18
19 Oct-Dec
                  4090 non-null
                                  float64
dtypes: float64(17), int64(2), object(1)
memory usage: 671.0+ KB
```

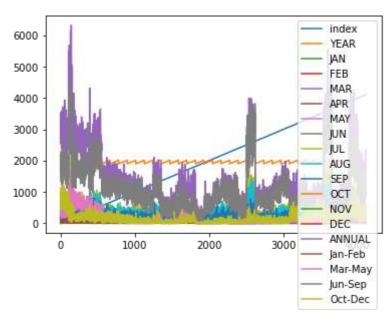
Line chart

```
In [6]:
        df.plot.line(subplots=True)
Out[6]: array([<AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>,
              <AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>,
              <AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>,
              <AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>,
              <AxesSubplot:>, <AxesSubplot:>], dtype=object)
                                                    IAN
                                                   FFR
                                                   MAR
                                  ΔPR
                                  MAY
                                  IUN
                                  IUL
                                  AUG
        1000
                                  SEP
                                  OCT
                                  NOV
                                                   DEC
        50¢
                                 ANNUAL
       5000
500
1000
                                                 Jan-Feb
                                 Mar-May
                                 lun-Sen
                                 Oct-Dec
                      1000
                               2000
                                         3000
                                                  4000
             0
```

Line chart

```
In [7]: df.plot.line()
```

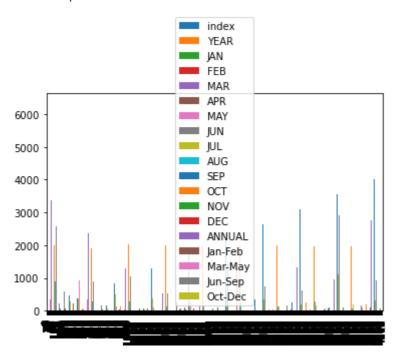
Out[7]: <AxesSubplot:>



Bar chart

```
In [8]: df.plot.bar()
```

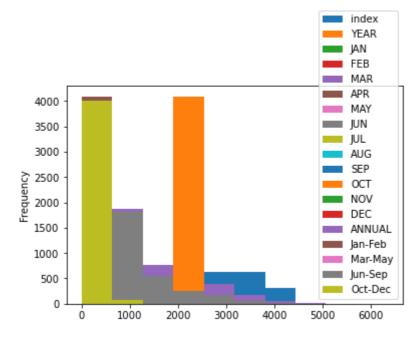
Out[8]: <AxesSubplot:>



Histogram

```
In [9]: df.plot.hist()
```

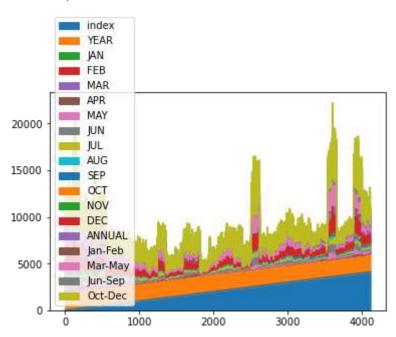
Out[9]: <AxesSubplot:ylabel='Frequency'>



Area chart

```
In [10]: df.plot.area()
```

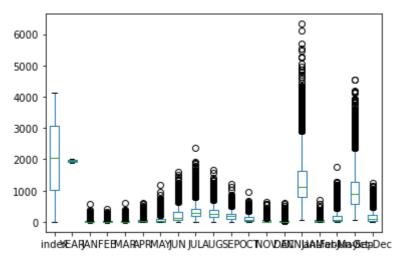
Out[10]: <AxesSubplot:>



Box chart

```
In [11]: df.plot.box()
```

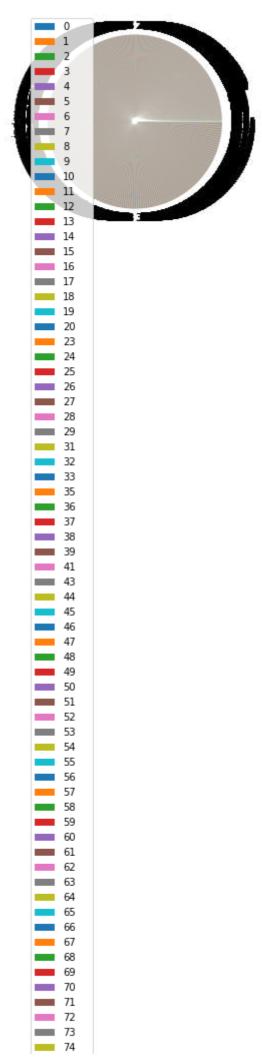
Out[11]: <AxesSubplot:>



Pie chart

```
In [12]: df.plot.pie(y='index')
```

Out[12]: <AxesSubplot:ylabel='index'>



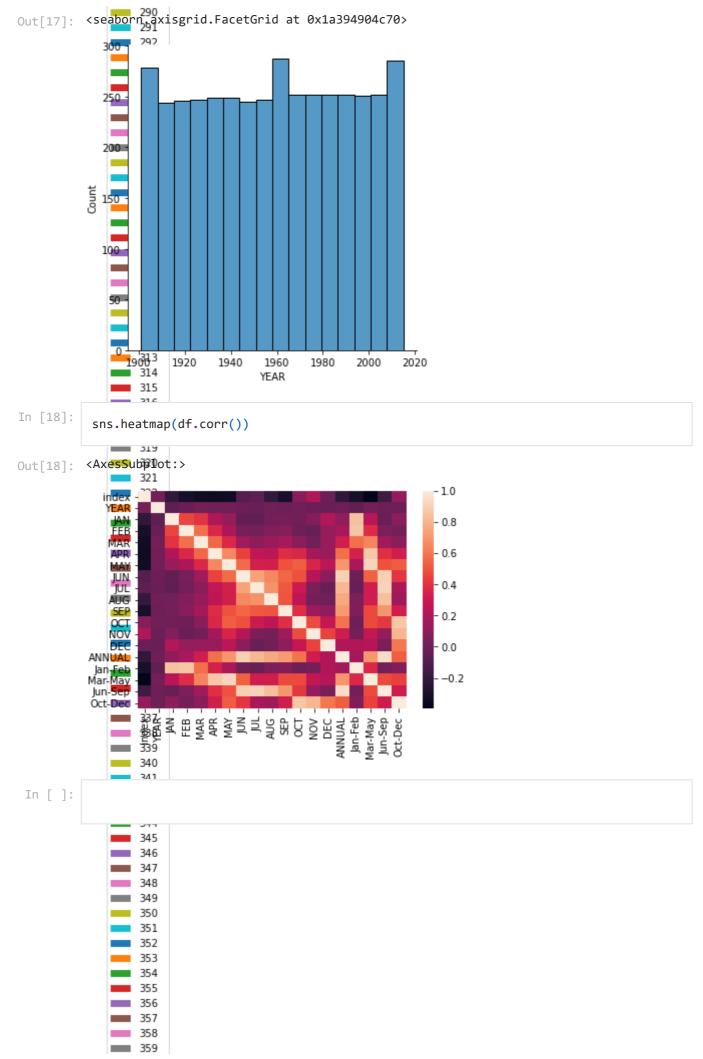
```
150
                   151
In [13]:
           df.plot.scatter(x='YEAR' ,y='index')
          <AxesSubplot:xlabel='YEAR', ylabel='index'>
Out[13]:
             2000
                                   1940
              19706)
                          1920
                                           1960
                                                    1980
                                                             2000
                                                                     2020
              177
                                           YEAR
              178
In [14]:
           df.info()
                  'pandas.core.frame.DataFrame'>
          <class
          Int64Index: 4090 entries, 0 to 4115
          Data columns (total 20 columns):
           #
                               Non-Null Count
               CoTumn
                                                Dtype
                index
SUBDIVISION
           0
                               4090 non-null
                                                 int64
           1
                              4090 non-null
                                                 object
                YEAR 0
           2
                               4090 non-null
                                                 int64
           3
                               4090 non-null
                                                 float64
                   91
           4
                               4090 non-null
                                                 float64
                FEB
                   192
           5
                               4090 non-null
                                                 float64
                   193
           6
                               4090 non-null
                                                 float64
           7
                                                 float64
                MΔ\
                               4090 non-null
                   195
           8
                או זר
                               4090 non-null
                                                 float64
                   196
           9
                               4090 non-null
                                                 float64
                   197
               AUG<sub>198</sub>
           10
                               4090 non-null
                                                 float64
           11
                SEP
                               4090 non-null
                                                 float64
                   199
           12 OCT<sub>200</sub>
13 NOV<sub>201</sub>
                               4090 non-null
                                                 float64
                               4090 non-null
                                                 float64
                   201
           14
               DEC
                               4090 non-null
                                                 float64
           15
              <u>A</u>NNŲĄ́L
                               4090 non-null
                                                 float64
                               4090 non-null
                                                 float64
           16
            17
              Mar<sub>20</sub>6ay
                               4090 non-null
                                                 float64
              Jun-Sep
                               4090 non-null
                                                 float64
           18
           19 Oct-Dec
                               4090 non-null
                                                 float64
          dtypes: 200 oat64(17), int64(2), object(1)
          memory usage: 671.0+ KB
In [15]:
           df.describe()
                  213
                  214 index
                                                  JAN
                                                                                       APR
                                                                                                   MAY
Out[15]:
                                    YEAR
                                                              FEB
                                                                          MAR
          count 4090.000000 4090.000000 4090.000000 4090.000000
                                                                   4090.000000 4090.000000 4090.000000 4
           mean 2057.638875 1958.321271
                                             18.818484
                                                         21.644792
                                                                      27.252494
                                                                                  42.714548
                                                                                               84.868044
              219
                  220
```

		220							
		221 index	YEAR	JAN	FEB	MAR	APR	MAY	
	std	1182.612736	33.148944	33.521719	35.762010	46.829179	67.264863	122.556801	
	min	0.000000	1901.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
2	5%	1034.250000	1930.000000	0.600000	0.600000	1.000000	3.000000	8.500000	
5	0%	2057.500000	1959.000000	5.900000	6.600000	7.800000	15.500000	36.050000	
7	5%	3080.750000	1987.000000	21.950000	26.600000	31.100000	49.375000	94.975000	
n	nax	4115.000000	2015.000000	583.700000	403.500000	605.600000	595.100000	1168.600000	1
		- 22.4							

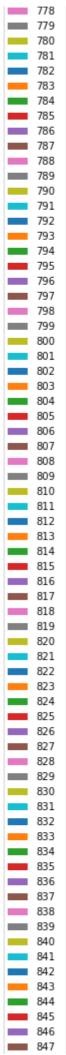
EDA 236 ND VISUALIZATION

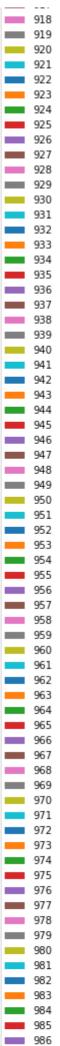
In [16]: sns.pairplot(df) <seaborn.43xisgrid.PairGrid at 0x1a3b26c3ac0> Out[16]: In [17]: sns.displot(df['YEAR'])

local host: 8888/nbc onvert/html/Rainfall-India.ipynb? download=false



Rainfall-India





localhost:8888/nbconvert/html/Rainfall-India.ipynb?download=false

Rainfall-India

2314
2315

232
252

Rainfall-India

Rainfall-India

Rainfall-India

8/4/23, 12:51 PM Rainfall-India