pfjm5zlqa

August 4, 2023

1 20104169 - SUMESH R

2 Importing Libraries

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

[2]:
```
from google.colab import drive
drive.mount('/content/drive')
df=pd.read_csv("/content/drive/MyDrive/mydatasets/rainfall/rainfall_lakshadweep.
  ↪csv")
df
```

```
Mounted at /content/drive
```

[2]:
```
     index  SUBDIVISION  YEAR   JAN   FEB    MAR    APR    MAY    JUN    JUL  \
0     4002  LAKSHADWEEP  1901  22.6  86.4  114.8  263.8   37.3  459.0    0.0
1     4003  LAKSHADWEEP  1902  99.3   9.6   32.6   40.4  179.1  374.2  413.3
2     4004  LAKSHADWEEP  1903  63.5  95.0    0.0   29.5  144.1  212.4  261.8
3     4005  LAKSHADWEEP  1904   0.0   0.0   13.5   13.2  143.3  261.3  256.0
4     4006  LAKSHADWEEP  1905  62.4   0.0    0.0    0.0  166.7  400.7   68.7
..     …          …    …    …    …     …     …     …    …    …
109   4111  LAKSHADWEEP  2011   5.1   2.8    3.1   85.9  107.2  153.6  350.2
110   4112  LAKSHADWEEP  2012  19.2   0.1    1.6   76.8   21.2  327.0  231.5
111   4113  LAKSHADWEEP  2013  26.2  34.4   37.5    5.3   88.3  426.2  296.4
112   4114  LAKSHADWEEP  2014  53.2  16.1    4.4   14.9   57.4  244.1  116.1
113   4115  LAKSHADWEEP  2015   2.2   0.5    3.7   87.1  133.1  296.6  257.5

       AUG    SEP    OCT    NOV   DEC  ANNUAL  Jan-Feb  Mar-May  Jun-Sep  \
0      0.0   46.7  183.7  229.9  15.0  1459.2    109.0    415.9    505.7
1    170.0  214.3  384.2  192.8  49.0  2158.8    108.9    252.1   1171.8
2    202.0  292.1   79.1    NaN   NaN     NaN    158.5    173.6    968.3
3     38.9  219.9  153.6    8.3  68.9  1176.9      0.0    170.0    776.1
4    377.5  107.5  232.1  159.3   0.0  1574.9     62.4    166.7    954.4
..     …     …     …     …     …      …       …       …       …
```

```
109  254.0  255.2  117.4  184.3   14.9  1533.7      7.9     196.2   1013.0
110  381.2  179.8  145.9   12.4    8.8  1405.5     19.3      99.6   1119.5
111  154.4  180.0   72.8   78.1   26.7  1426.3     60.6     131.1   1057.0
112  466.1  132.2  169.2   59.0   62.3  1395.0     69.3      76.7    958.5
113  146.4  160.4  165.4  231.0  159.0  1642.9      2.7     223.9    860.9

     Oct-Dec
0      428.6
1      626.0
2        NaN
3      230.8
4      391.4
..       …
109    316.6
110    167.1
111    177.6
112    290.5
113    555.4

[114 rows x 20 columns]
```

3 Data Cleaning and Data Preprocessing

[3]:

```
df=df.dropna()
```

[4]:

```
df.columns
```

[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')
```

[5]:

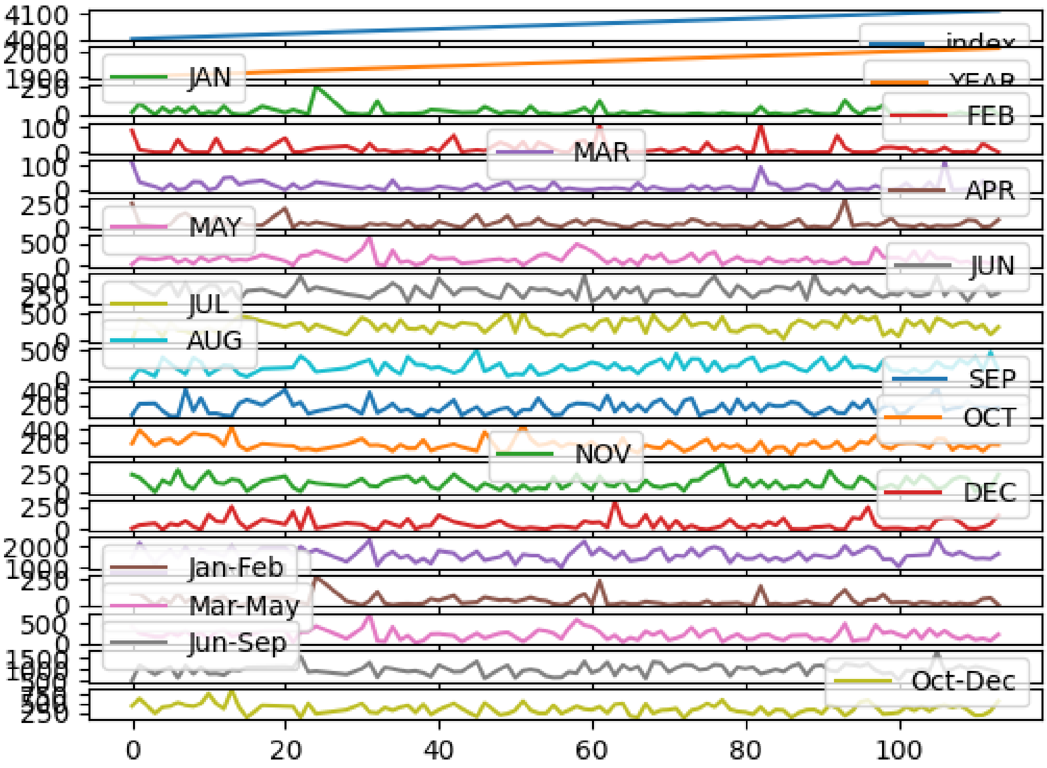
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 103 entries, 0 to 113
Data columns (total 20 columns):
 #   Column       Non-Null Count  Dtype
---  ------       --------------  -----
 0   index        103 non-null    int64
 1   SUBDIVISION  103 non-null    object
 2   YEAR         103 non-null    int64
 3   JAN          103 non-null    float64
 4   FEB          103 non-null    float64
 5   MAR          103 non-null    float64
 6   APR          103 non-null    float64
 7   MAY          103 non-null    float64
 8   JUN          103 non-null    float64
 9   JUL          103 non-null    float64
 10  AUG          103 non-null    float64
 11  SEP          103 non-null    float64
 12  OCT          103 non-null    float64
 13  NOV          103 non-null    float64
 14  DEC          103 non-null    float64
 15  ANNUAL       103 non-null    float64
 16  Jan-Feb      103 non-null    float64
 17  Mar-May      103 non-null    float64
 18  Jun-Sep      103 non-null    float64
 19  Oct-Dec      103 non-null    float64
dtypes: float64(17), int64(2), object(1)
memory usage: 16.9+ KB
```

4 Line chart

```
[6]: df.plot.line(subplots=True)
```

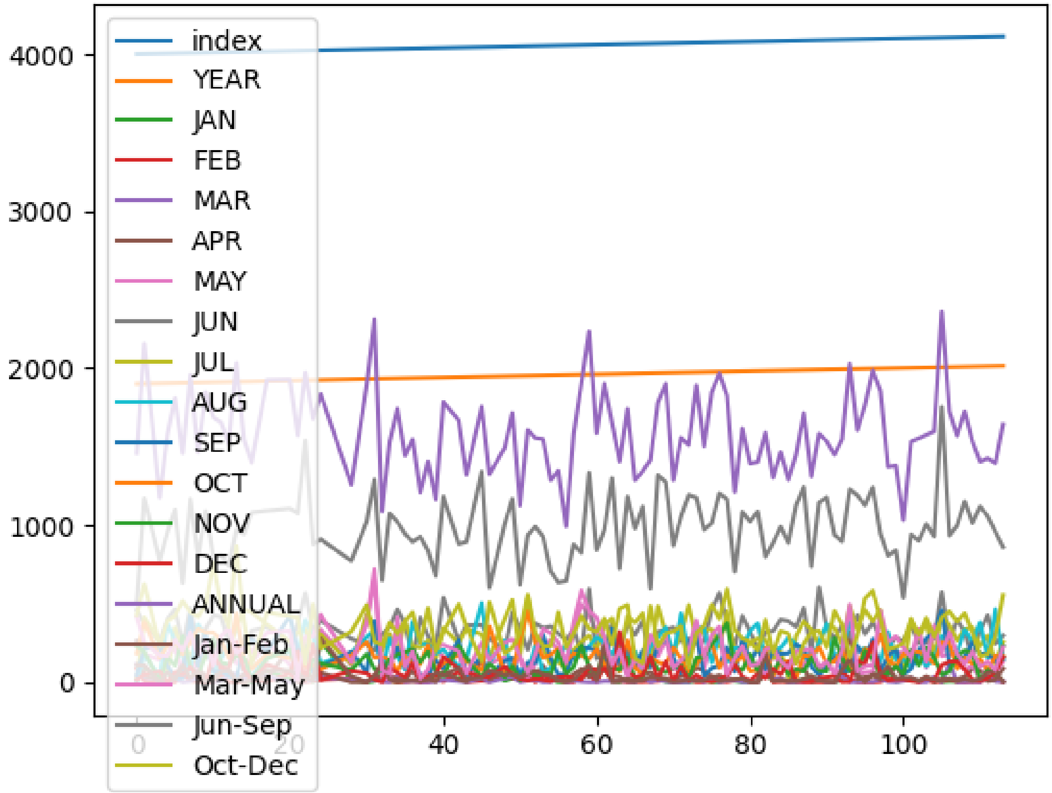
```
[6]: array([<Axes: >, <Axes: >, <Axes: >, <Axes: >, <Axes: >, <Axes: >,
            <Axes: >, <Axes: >, <Axes: >, <Axes: >, <Axes: >, <Axes: >,
            <Axes: >, <Axes: >, <Axes: >, <Axes: >, <Axes: >, <Axes: >,
            <Axes: >], dtype=object)
```



5 Line chart

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

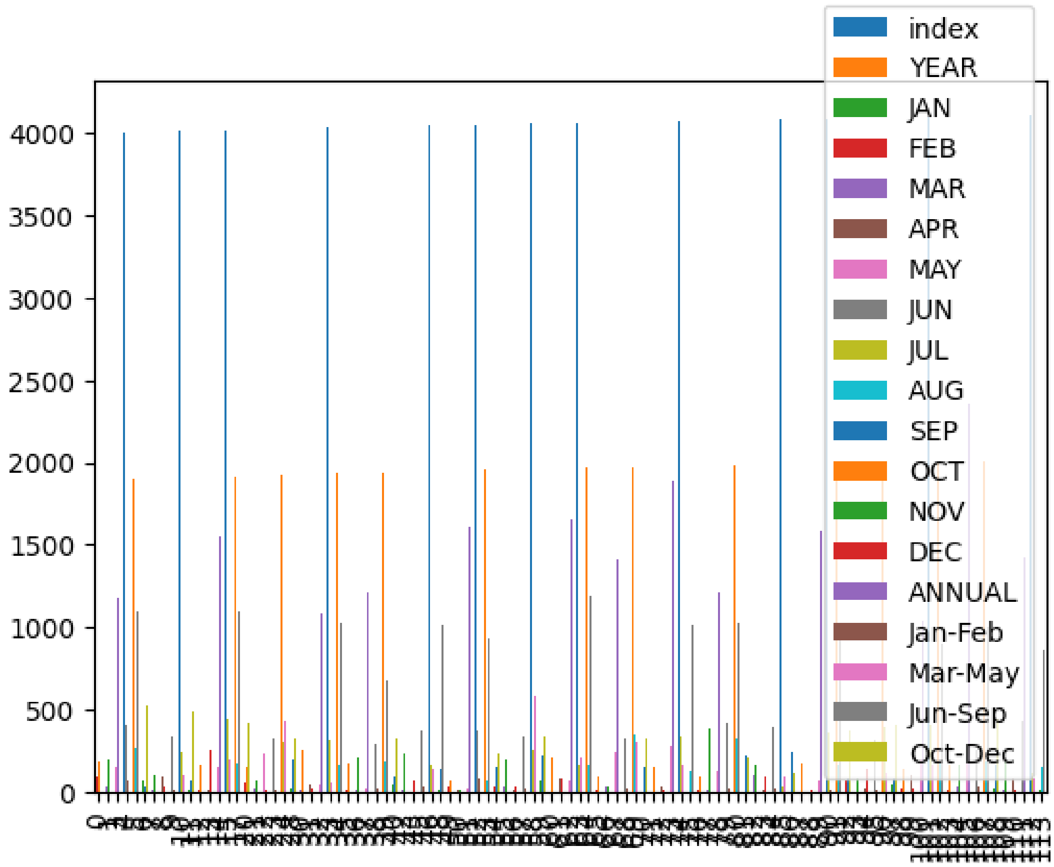
```
[7]: <Axes: >
```



6 Bar chart

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

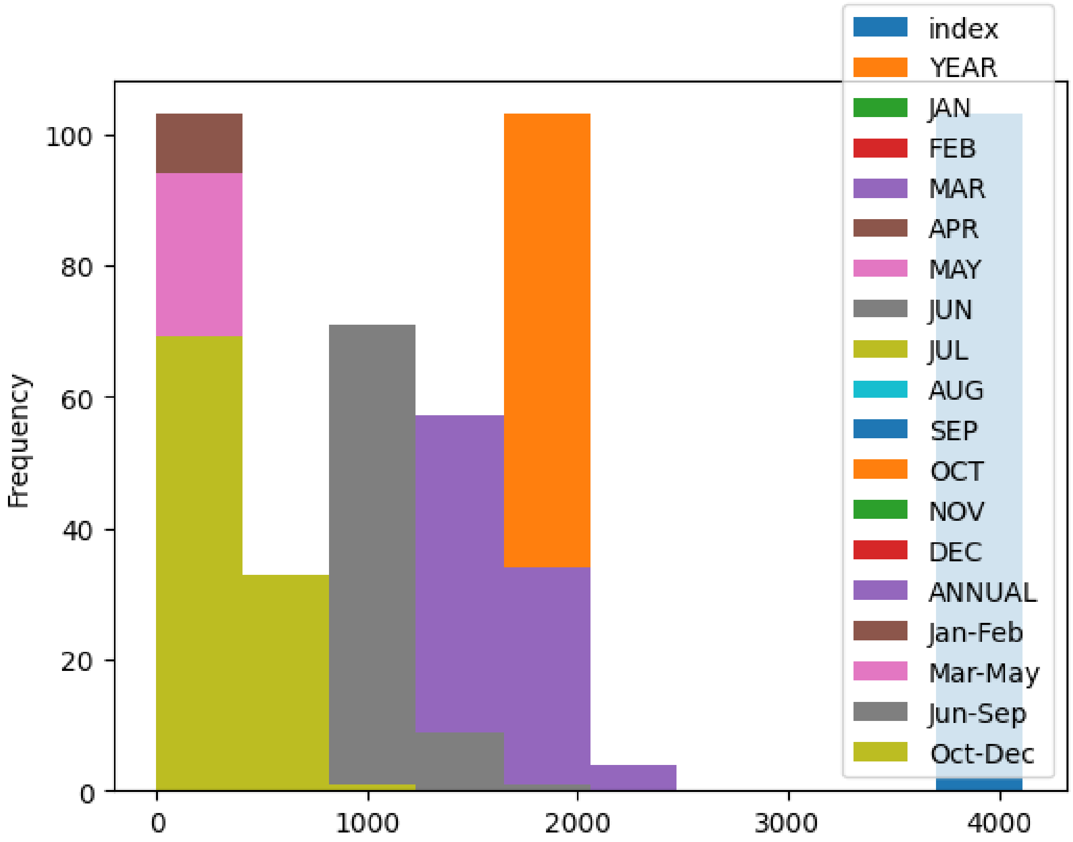
```
[8]: <Axes: >
```



7 Histogram

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

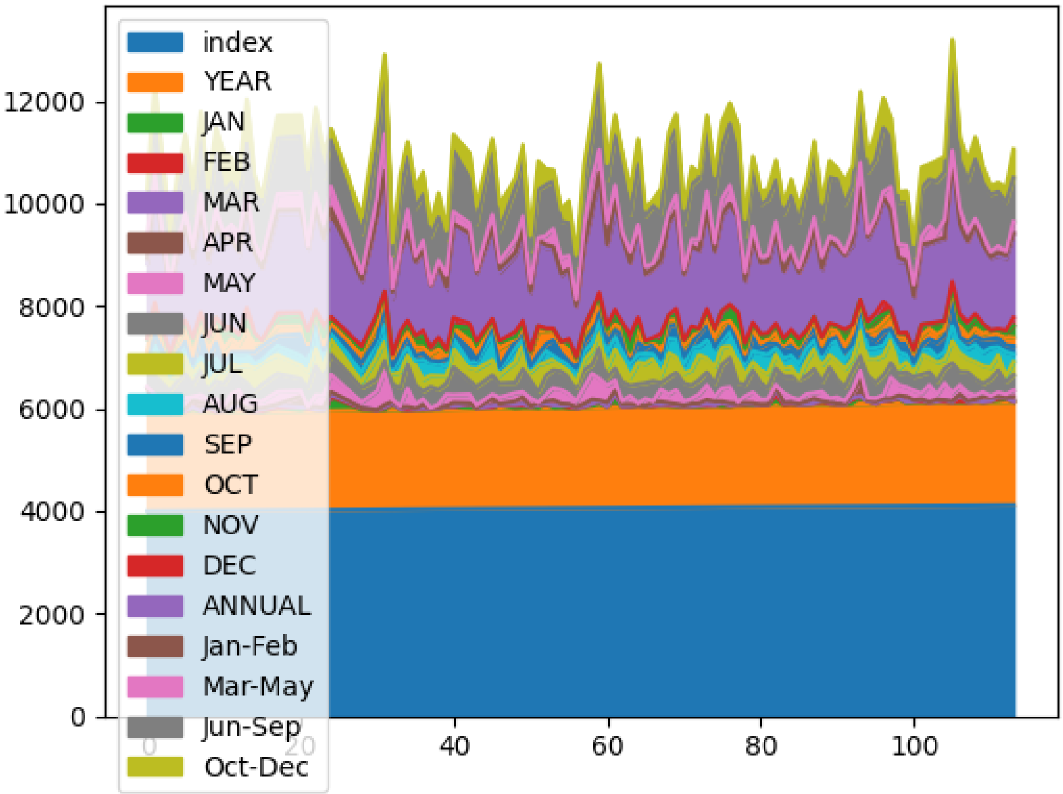
```
[9]: <Axes: ylabel='Frequency'>
```



8 Area chart

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

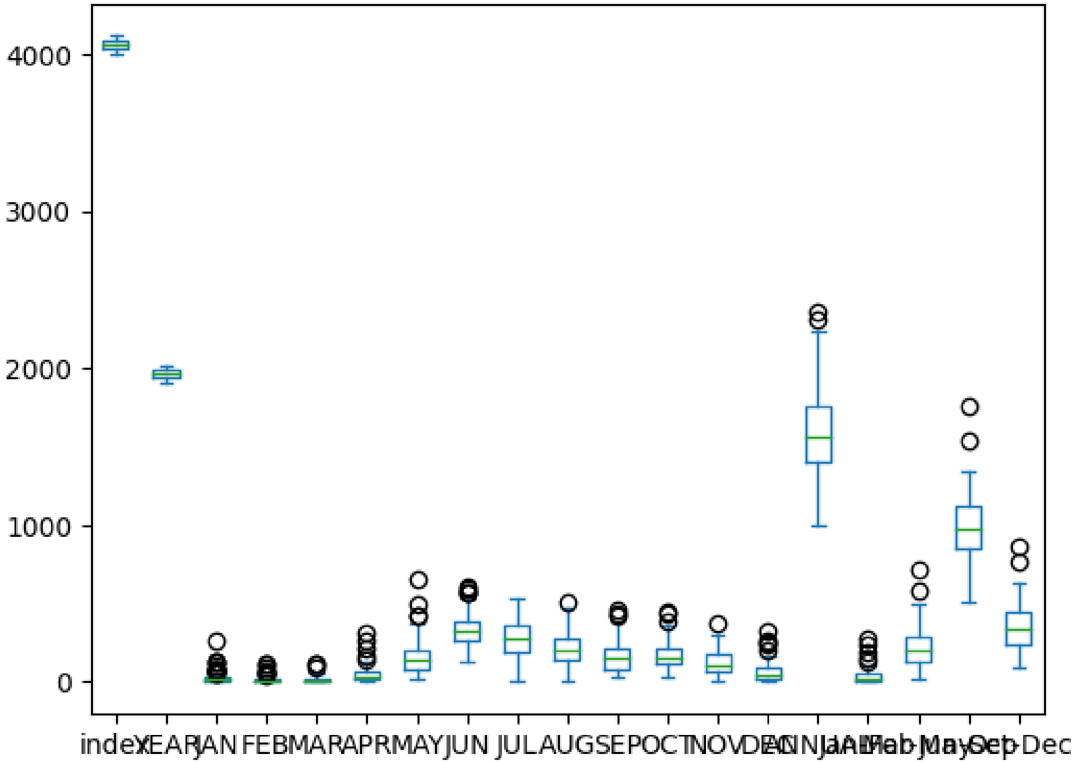
```
[10]: <Axes: >
```



9 Box chart

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

[11]:
```
<Axes: >
```



10 Pie chart

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

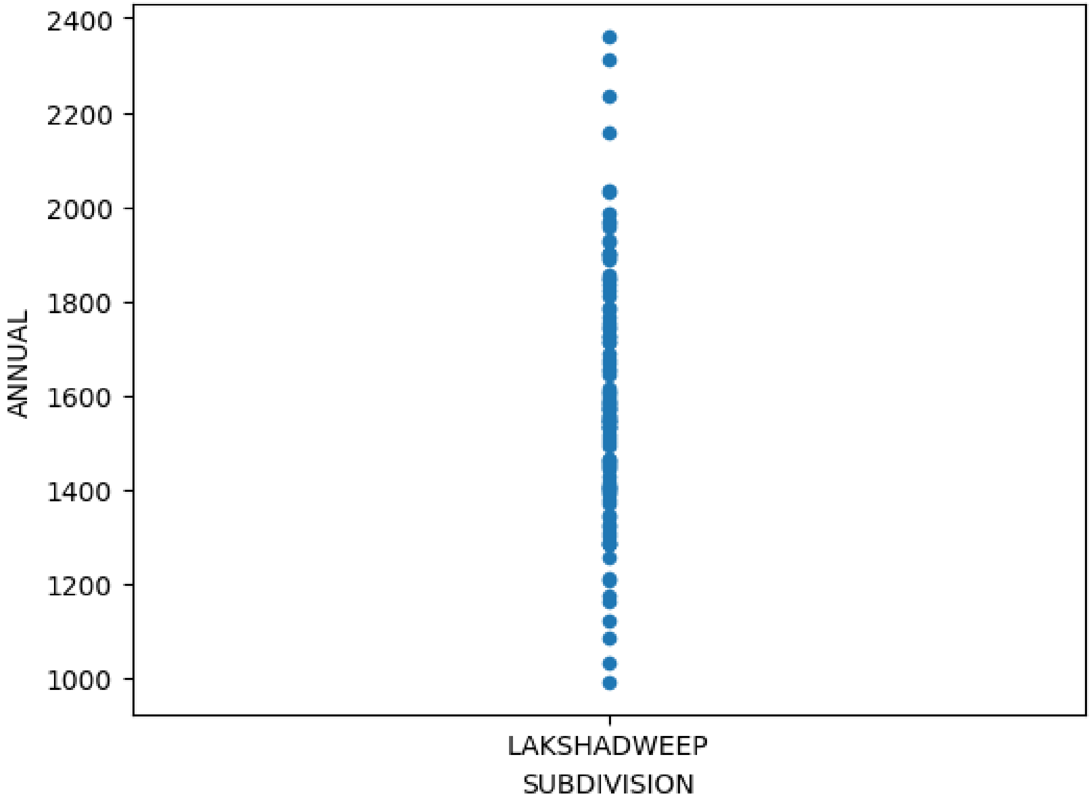
[12]:
```
<Axes: ylabel='ANNUAL'>
```

ANNUAL
0
1
3
4
5
6
7
8
9
10
11
12
13
14
15
17
20
21
22
23
24
28
30
31
32
33
34
35
36
37
38
39
40
42
43
45
46
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113

11 Scatter chart

```
[13]: df.plot.scatter(x='SUBDIVISION' ,y='ANNUAL')
```

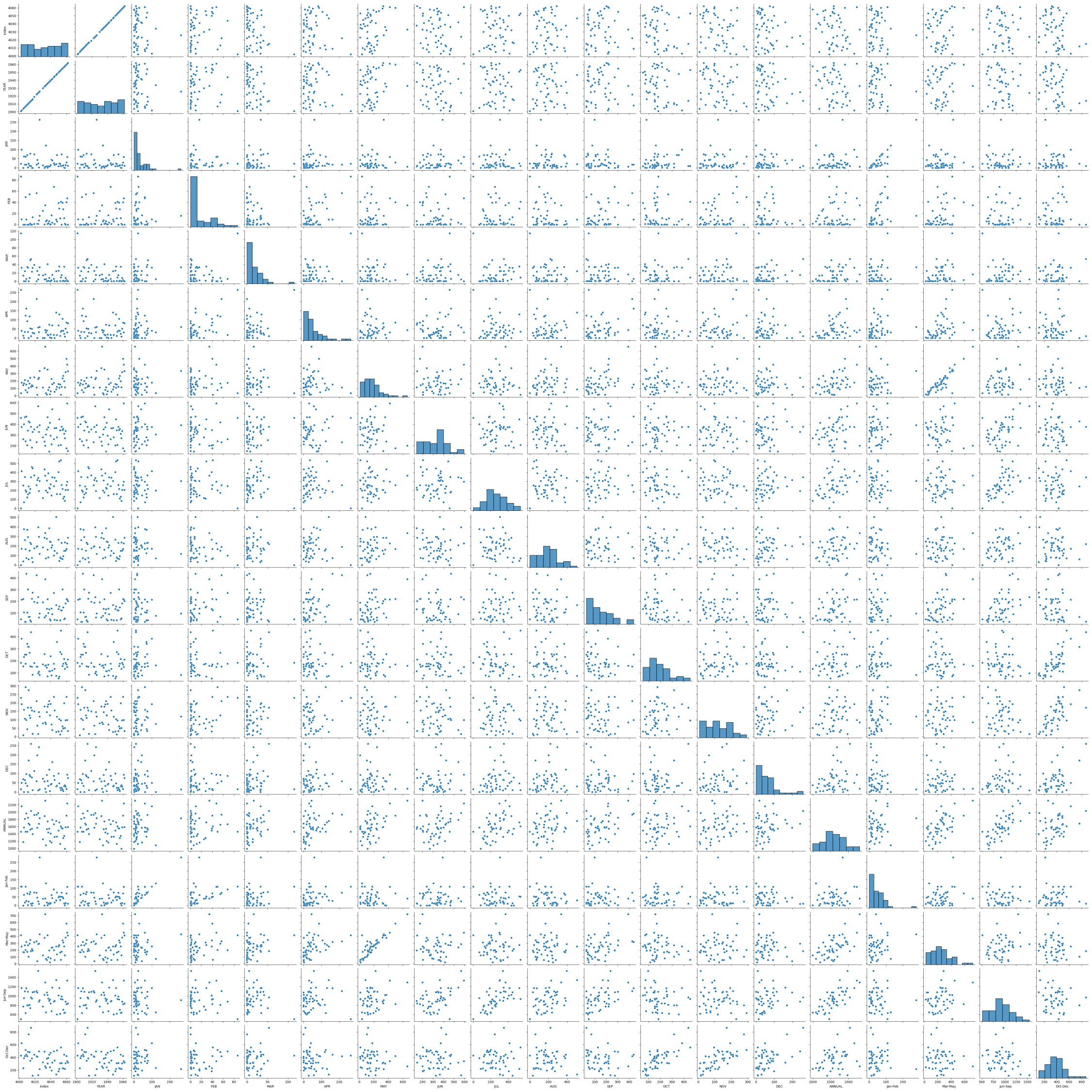
```
[13]: <Axes: xlabel='SUBDIVISION', ylabel='ANNUAL'>
```



12 Seaborn

```
[14]: sns.pairplot(df[0:50])
```

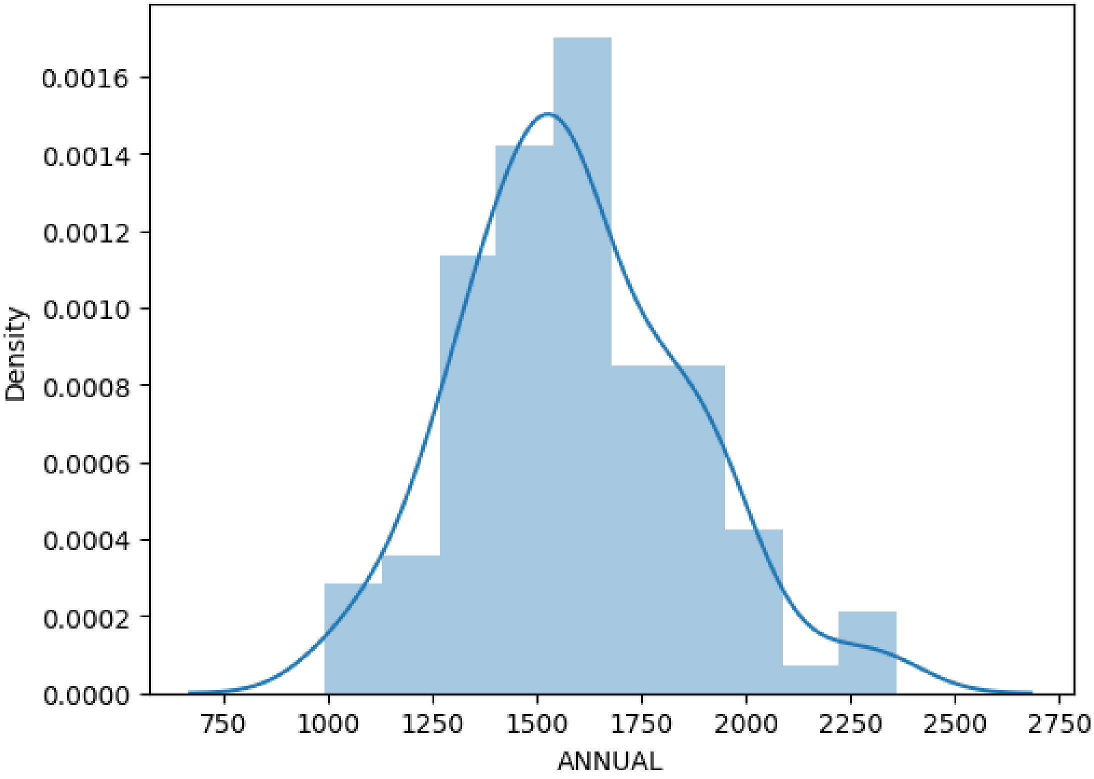
```
[14]: <seaborn.axisgrid.PairGrid at 0x7b5500bf7ca0>
```



[15]:
```
sns.distplot(df['ANNUAL'])
```

```
<ipython-input-15-5daa97052ca5>:1: UserWarning:

`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with
similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751

  sns.distplot(df['ANNUAL'])
```

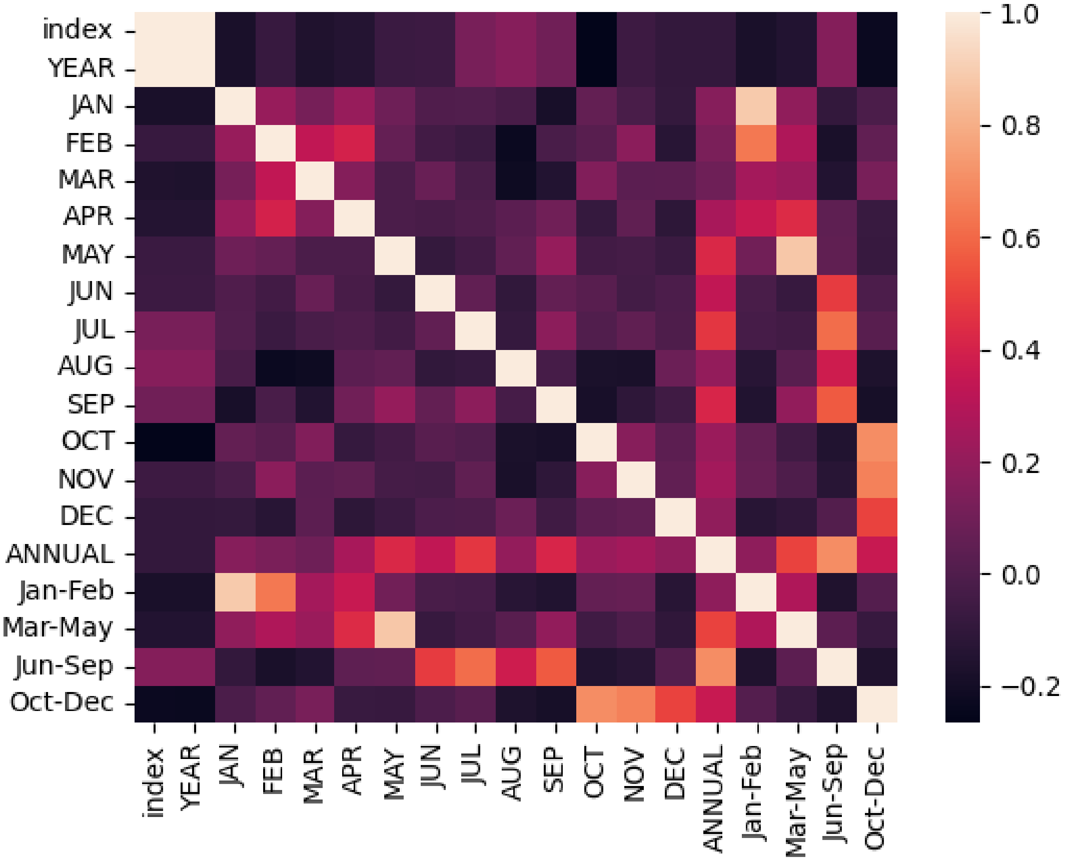
```
[15]: <Axes: xlabel='ANNUAL', ylabel='Density'>
```



```
[16]: sns.heatmap(df.corr())
```

```
<ipython-input-16-aa4f4450a243>:1: FutureWarning: The default value of
numeric_only in DataFrame.corr is deprecated. In a future version, it will
default to False. Select only valid columns or specify the value of numeric_only
to silence this warning.
  sns.heatmap(df.corr())
```

```
[16]: <Axes: >
```


index
YEAR
JAN
FEB
MAR
APR
MAY
JUN
JUL
AUG
SEP
OCT
NOV
DEC
ANNUAL
Jan-Feb
Mar-May
Jun-Sep
Oct-Dec
1.0
0.8
0.6
0.4
0.2
0.0
−0.2