

Importing Libraries

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

Importing Datasets

```
In [2]: df=pd.read_csv(r'C:\Users\user\Downloads\Rainfall\NAGA MANI MIZO TRIPURA.csv')
df
```

Out[2]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OC
0	323	NAGA MANI MIZO TRIPURA	1902	4.8	0.5	36.3	297.8	215.5	480.1	392.4	312.8	318.7	102
1	324	NAGA MANI MIZO TRIPURA	1903	6.5	40.5	139.8	45.5	159.9	458.6	300.2	470.6	366.1	166
2	325	NAGA MANI MIZO TRIPURA	1904	2.3	46.9	47.5	290.3	230.5	455.3	423.5	423.6	375.8	128
3	326	NAGA MANI MIZO TRIPURA	1905	9.1	35.3	306.5	161.7	193.6	339.7	450.1	429.9	320.1	246
4	327	NAGA MANI MIZO TRIPURA	1906	7.0	71.5	72.5	99.0	302.7	417.4	475.2	439.2	439.1	142
...
109	432	NAGA MANI MIZO TRIPURA	2011	12.6	3.6	51.4	81.1	334.9	374.2	313.3	367.6	258.3	92
110	433	NAGA MANI MIZO TRIPURA	2012	24.5	10.2	20.3	243.5	163.5	396.2	280.1	342.7	248.7	160
111	434	NAGA MANI MIZO TRIPURA	2013	0.2	5.7	19.7	60.3	348.9	206.6	255.9	291.3	241.4	125
112	435	NAGA MANI MIZO TRIPURA	2014	1.2	21.0	25.4	49.6	192.5	268.3	295.7	372.3	300.9	69
113	436	NAGA MANI MIZO TRIPURA	2015	14.4	14.2	21.6	253.5	198.3	283.9	413.6	334.2	255.9	118

114 rows × 20 columns



Data Cleaning and Data Preprocessing

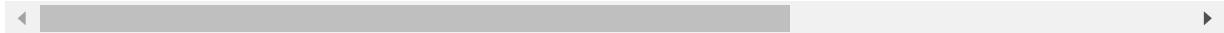
In [3]:

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

Out[3]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OC
0	323	NAGA MANI MIZO TRIPURA	1902	4.8	0.5	36.3	297.8	215.5	480.1	392.4	312.8	318.7	102
1	324	NAGA MANI MIZO TRIPURA	1903	6.5	40.5	139.8	45.5	159.9	458.6	300.2	470.6	366.1	166
2	325	NAGA MANI MIZO TRIPURA	1904	2.3	46.9	47.5	290.3	230.5	455.3	423.5	423.6	375.8	128
3	326	NAGA MANI MIZO TRIPURA	1905	9.1	35.3	306.5	161.7	193.6	339.7	450.1	429.9	320.1	246
4	327	NAGA MANI MIZO TRIPURA	1906	7.0	71.5	72.5	99.0	302.7	417.4	475.2	439.2	439.1	142
...
109	432	NAGA MANI MIZO TRIPURA	2011	12.6	3.6	51.4	81.1	334.9	374.2	313.3	367.6	258.3	92
110	433	NAGA MANI MIZO TRIPURA	2012	24.5	10.2	20.3	243.5	163.5	396.2	280.1	342.7	248.7	160
111	434	NAGA MANI MIZO TRIPURA	2013	0.2	5.7	19.7	60.3	348.9	206.6	255.9	291.3	241.4	125
112	435	NAGA MANI MIZO TRIPURA	2014	1.2	21.0	25.4	49.6	192.5	268.3	295.7	372.3	300.9	69
113	436	NAGA MANI MIZO TRIPURA	2015	14.4	14.2	21.6	253.5	198.3	283.9	413.6	334.2	255.9	118

114 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')
```

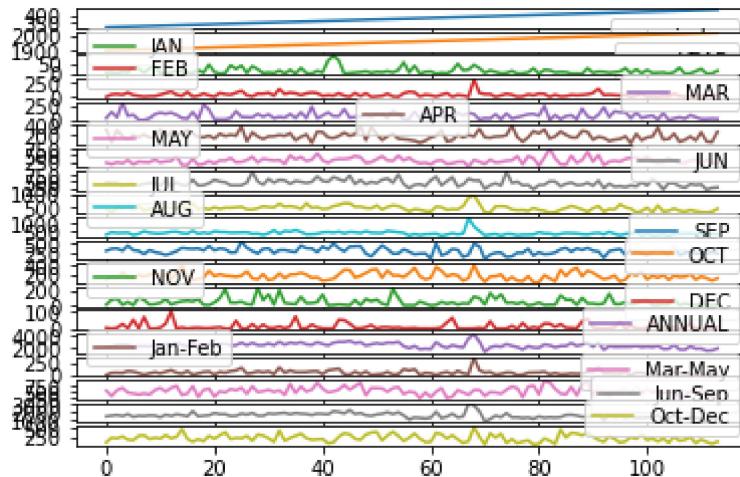
In [5]: df.info()

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 114 entries, 0 to 113
Data columns (total 20 columns):
 #   Column      Non-Null Count  Dtype  
--- 
 0   index       114 non-null    int64  
 1   SUBDIVISION 114 non-null    object  
 2   YEAR        114 non-null    int64  
 3   JAN         114 non-null    float64 
 4   FEB         114 non-null    float64 
 5   MAR         114 non-null    float64 
 6   APR         114 non-null    float64 
 7   MAY         114 non-null    float64 
 8   JUN         114 non-null    float64 
 9   JUL         114 non-null    float64 
 10  AUG         114 non-null    float64 
 11  SEP         114 non-null    float64 
 12  OCT         114 non-null    float64 
 13  NOV         114 non-null    float64 
 14  DEC         114 non-null    float64 
 15  ANNUAL      114 non-null    float64 
 16  Jan-Feb     114 non-null    float64 
 17  Mar-May     114 non-null    float64 
 18  Jun-Sep     114 non-null    float64 
 19  Oct-Dec     114 non-null    float64 
dtypes: float64(17), int64(2), object(1)
memory usage: 18.7+ KB
```

Line Chart

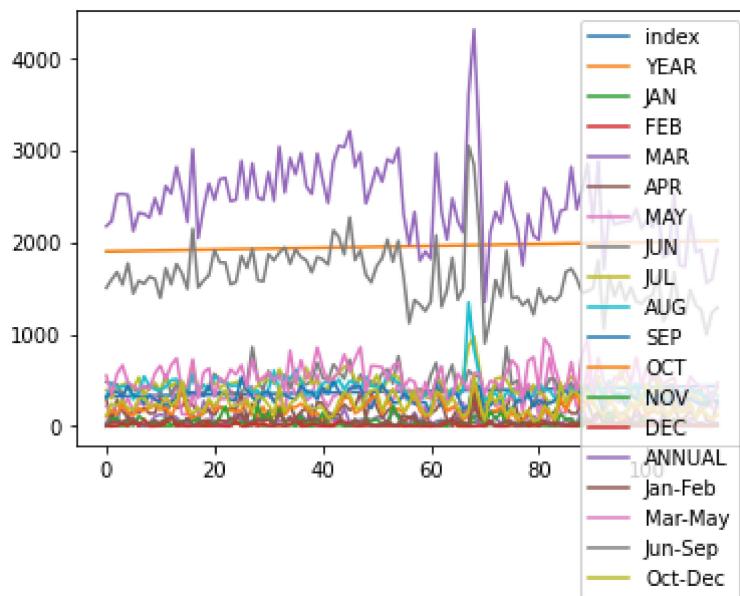
In [6]: `df.plot.line(subplots=True)`

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



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

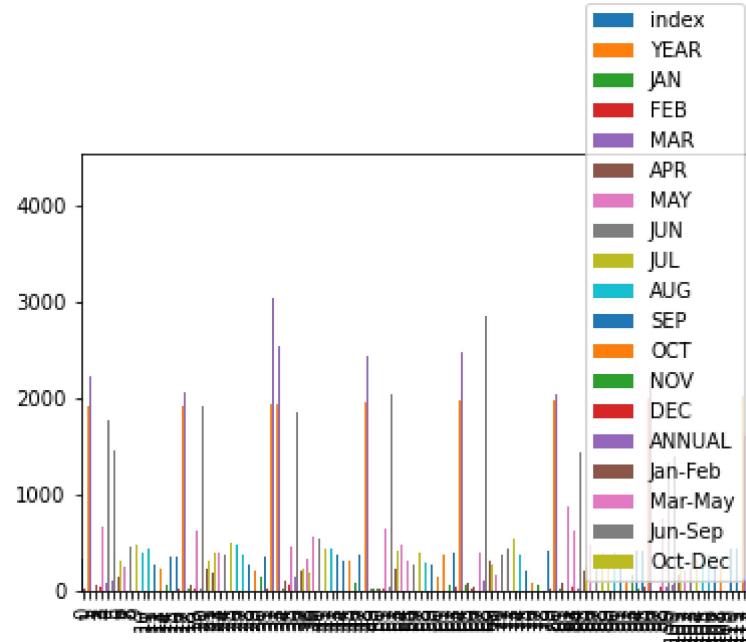
Out[7]: `<AxesSubplot:>`



Bar Chart

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

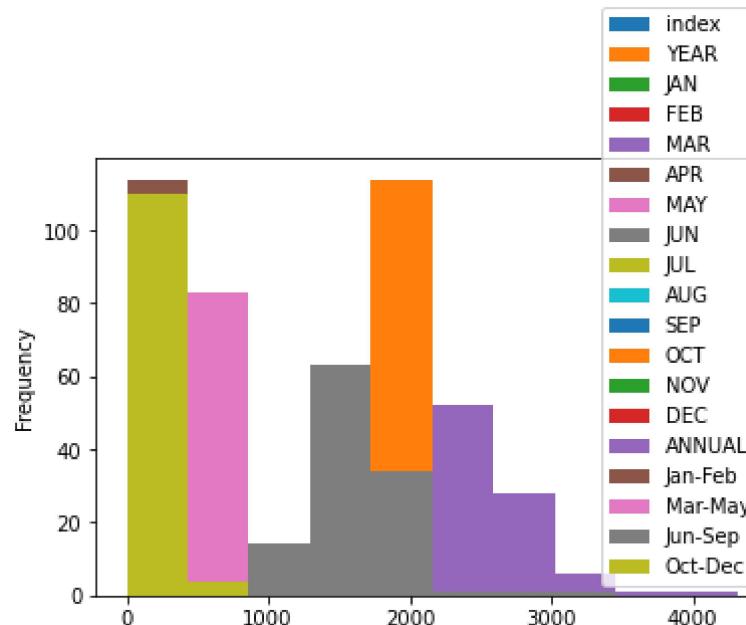
Out[8]: <AxesSubplot:>



Histogram

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

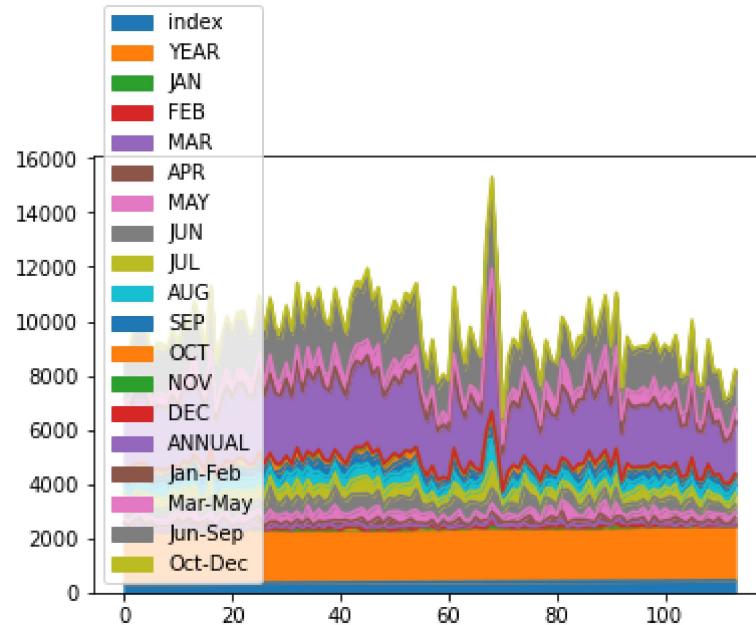
Out[9]: <AxesSubplot:ylabel='Frequency'>



Area Chart

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

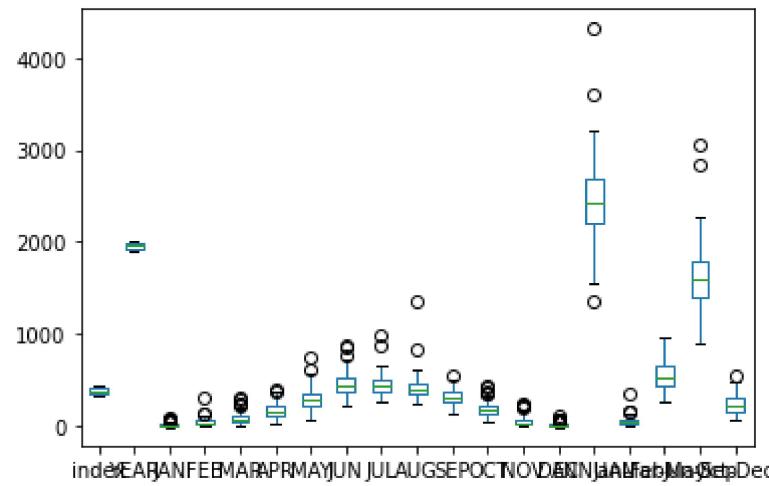
Out[10]: <AxesSubplot:>



Box Chart

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

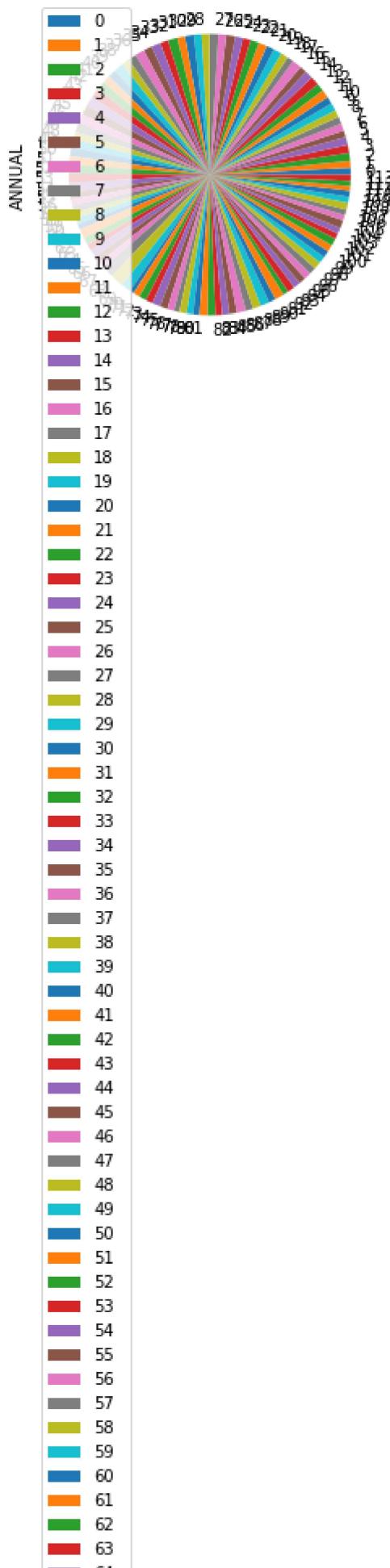
Out[11]: <AxesSubplot:>



Pie Chart

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

```
Out[12]: <AxesSubplot:ylabel='ANNUAL'>
```

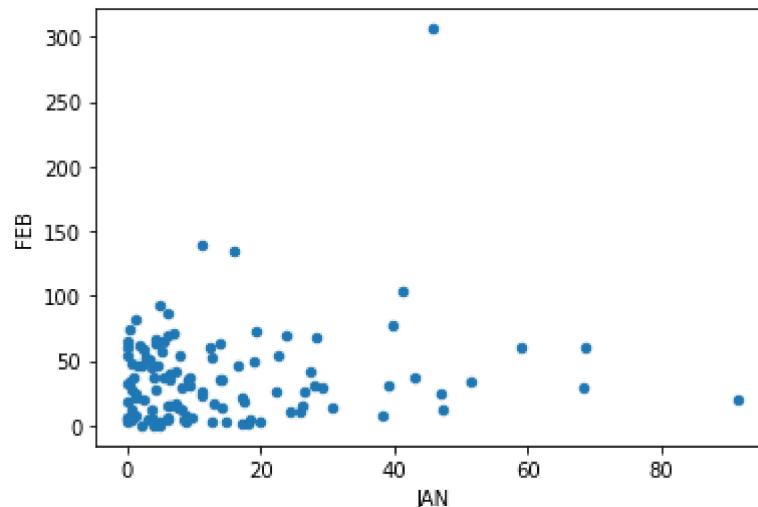



64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113

Scatter Plot

```
In [13]: df.plot.scatter(x='JAN',y='FEB')
```

```
Out[13]: <AxesSubplot:xlabel='JAN', ylabel='FEB'>
```



```
In [14]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 114 entries, 0 to 113
Data columns (total 20 columns):
 #   Column      Non-Null Count  Dtype  
--- 
 0   index       114 non-null    int64  
 1   SUBDIVISION 114 non-null    object  
 2   YEAR        114 non-null    int64  
 3   JAN         114 non-null    float64 
 4   FEB         114 non-null    float64 
 5   MAR         114 non-null    float64 
 6   APR         114 non-null    float64 
 7   MAY         114 non-null    float64 
 8   JUN         114 non-null    float64 
 9   JUL         114 non-null    float64 
 10  AUG         114 non-null    float64 
 11  SEP         114 non-null    float64 
 12  OCT         114 non-null    float64 
 13  NOV         114 non-null    float64 
 14  DEC         114 non-null    float64 
 15  ANNUAL      114 non-null    float64 
 16  Jan-Feb     114 non-null    float64 
 17  Mar-May     114 non-null    float64 
 18  Jun-Sep     114 non-null    float64 
 19  Oct-Dec     114 non-null    float64 
dtypes: float64(17), int64(2), object(1)
memory usage: 18.7+ KB
```

In [15]: df.describe()

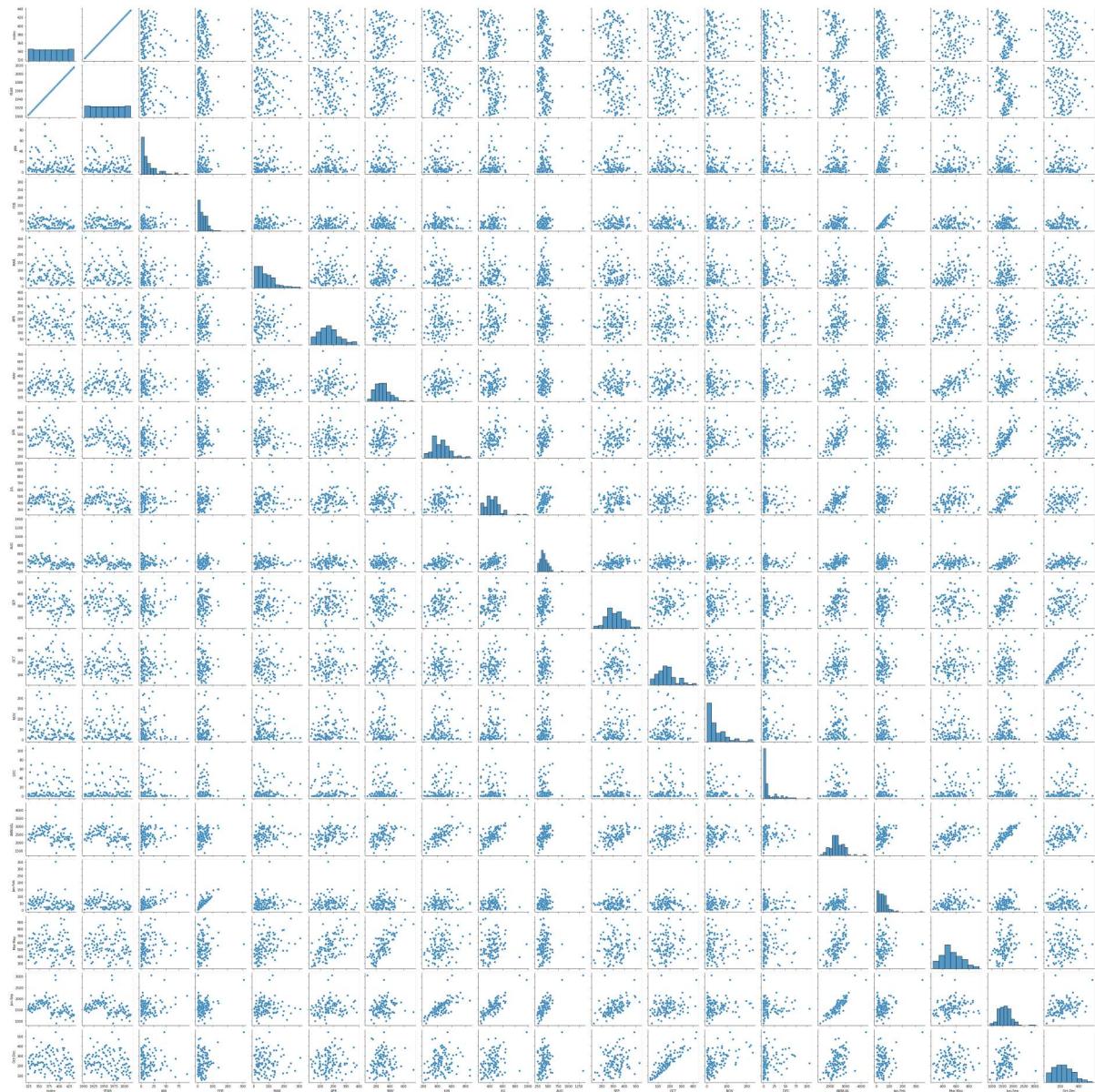
Out[15]:

	index	YEAR	JAN	FEB	MAR	APR	MAY	
count	114.000000	114.000000	114.000000	114.000000	114.000000	114.000000	114.000000	114
mean	379.500000	1958.500000	14.045614	36.814912	77.618421	170.421930	292.302632	445
std	33.052988	33.052988	16.699717	37.903151	60.756869	78.834683	112.069767	124
min	323.000000	1902.000000	0.000000	0.000000	3.100000	26.300000	73.500000	206
25%	351.250000	1930.250000	3.075000	11.600000	32.975000	112.875000	211.275000	361
50%	379.500000	1958.500000	7.800000	30.050000	63.400000	161.600000	279.550000	440
75%	407.750000	1986.750000	18.525000	53.450000	105.275000	214.050000	352.350000	512
max	436.000000	2015.000000	91.400000	306.300000	306.500000	383.800000	743.000000	861

EDA And Visualization

```
In [16]: sns.pairplot(df)
```

```
Out[16]: <seaborn.axisgrid.PairGrid at 0x15a39ad6d60>
```

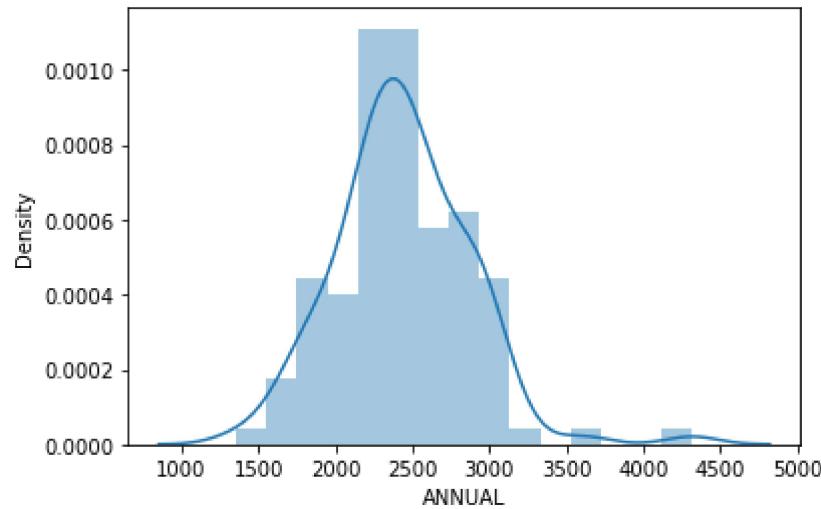


In [17]: `sns.distplot(df['ANNUAL'])`

C:\ProgramData\Anaconda3\lib\site-packages\seaborn\distributions.py:2557: FutureWarning: `distplot` is a deprecated function and will be removed in a future version. Please adapt your code to use either `displot` (a figure-level function with similar flexibility) or `histplot` (an axes-level function for histograms).

`warnings.warn(msg, FutureWarning)`

Out[17]: <AxesSubplot:xlabel='ANNUAL', ylabel='Density'>



In [18]: `sns.heatmap(df.corr())`

Out[18]: <AxesSubplot:>

