

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\TAMIL NADU.csv')
df
```

Out[2]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT
0	3428	TAMIL NADU	1902	67.2	9.8	25.1	21.9	84.7	39.3	55.1	113.8	98.6	282.2
1	3429	TAMIL NADU	1903	19.3	7.8	1.7	18.2	128.5	58.5	72.6	115.0	210.4	128.1
2	3430	TAMIL NADU	1904	35.2	0.1	0.7	19.5	121.9	34.9	89.0	40.4	85.7	163.2
3	3431	TAMIL NADU	1905	6.5	7.5	17.2	64.8	83.7	49.8	39.0	101.8	73.5	250.4
4	3432	TAMIL NADU	1906	52.4	12.9	17.0	8.5	39.6	43.6	76.0	195.2	65.3	162.8
...
109	3537	TAMIL NADU	2011	4.3	11.2	8.0	91.5	33.4	56.0	45.5	128.9	76.0	200.4
110	3538	TAMIL NADU	2012	3.0	0.1	2.5	35.5	41.9	30.1	46.5	98.0	84.9	235.2
111	3539	TAMIL NADU	2013	3.9	30.9	30.0	20.3	42.0	54.6	42.7	110.7	113.5	127.9
112	3540	TAMIL NADU	2014	7.4	6.1	8.1	8.3	139.1	47.8	50.6	117.7	98.9	252.2
113	3541	TAMIL NADU	2015	8.3	2.3	21.7	108.8	112.4	62.4	43.5	81.6	98.4	132.6

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	OCT
0	3428	TAMIL NADU	1902	67.2	9.8	25.1	21.9	84.7	39.3	55.1	113.8	98.6	282.2
1	3429	TAMIL NADU	1903	19.3	7.8	1.7	18.2	128.5	58.5	72.6	115.0	210.4	128.1
2	3430	TAMIL NADU	1904	35.2	0.1	0.7	19.5	121.9	34.9	89.0	40.4	85.7	163.2
3	3431	TAMIL NADU	1905	6.5	7.5	17.2	64.8	83.7	49.8	39.0	101.8	73.5	250.4
4	3432	TAMIL NADU	1906	52.4	12.9	17.0	8.5	39.6	43.6	76.0	195.2	65.3	162.8
...
109	3537	TAMIL NADU	2011	4.3	11.2	8.0	91.5	33.4	56.0	45.5	128.9	76.0	200.4
110	3538	TAMIL NADU	2012	3.0	0.1	2.5	35.5	41.9	30.1	46.5	98.0	84.9	235.2
111	3539	TAMIL NADU	2013	3.9	30.9	30.0	20.3	42.0	54.6	42.7	110.7	113.5	127.9
112	3540	TAMIL NADU	2014	7.4	6.1	8.1	8.3	139.1	47.8	50.6	117.7	98.9	252.2
113	3541	TAMIL NADU	2015	8.3	2.3	21.7	108.8	112.4	62.4	43.5	81.6	98.4	132.6

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')

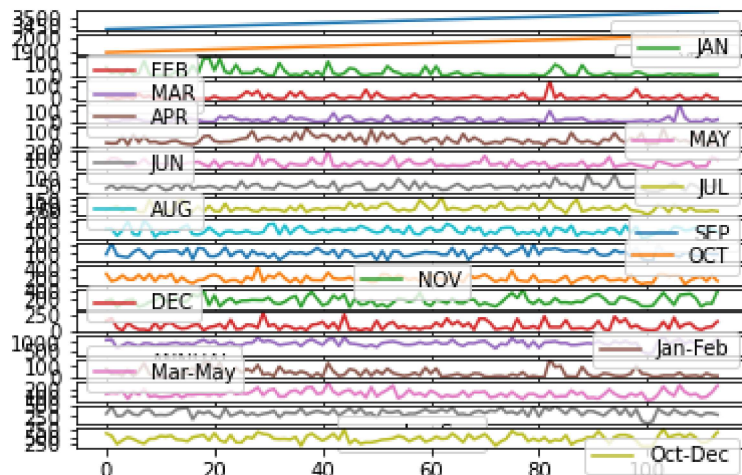
```
In [5]: df.info()
```

```
<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

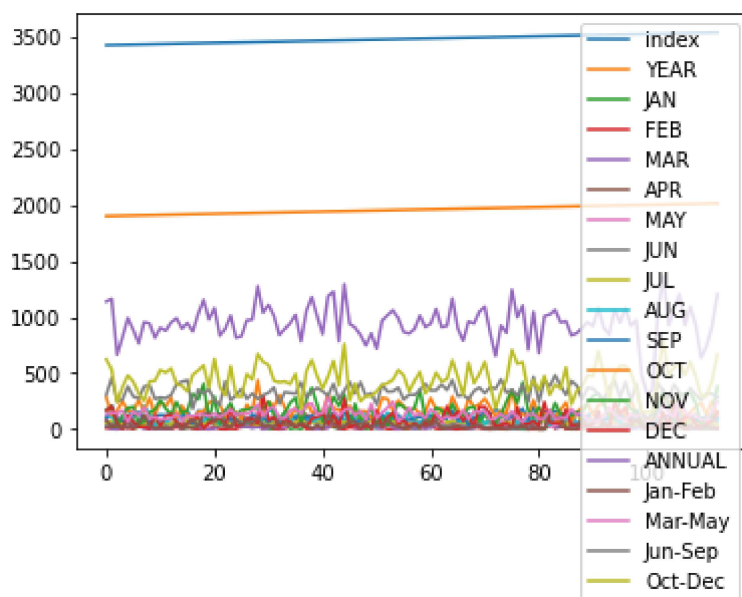
```
In [6]: df.plot.line(subplots=True)
```

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



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

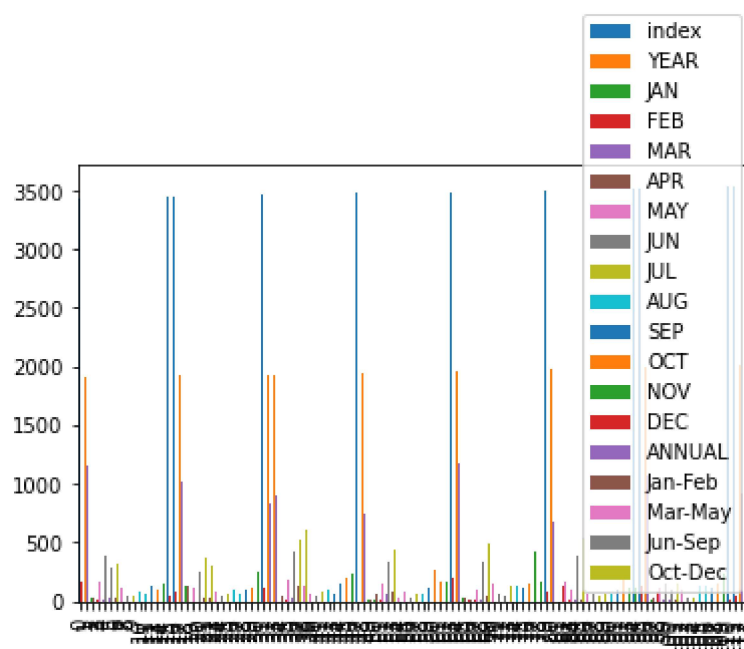
```
Out[7]: <AxesSubplot:~>
```



Bar Chart

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

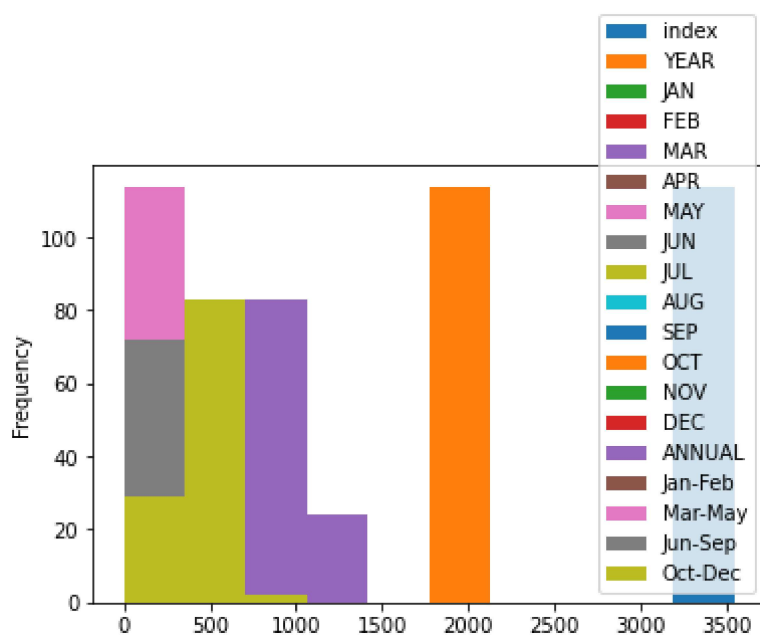
```
Out[8]: <AxesSubplot:>
```



Histogram

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

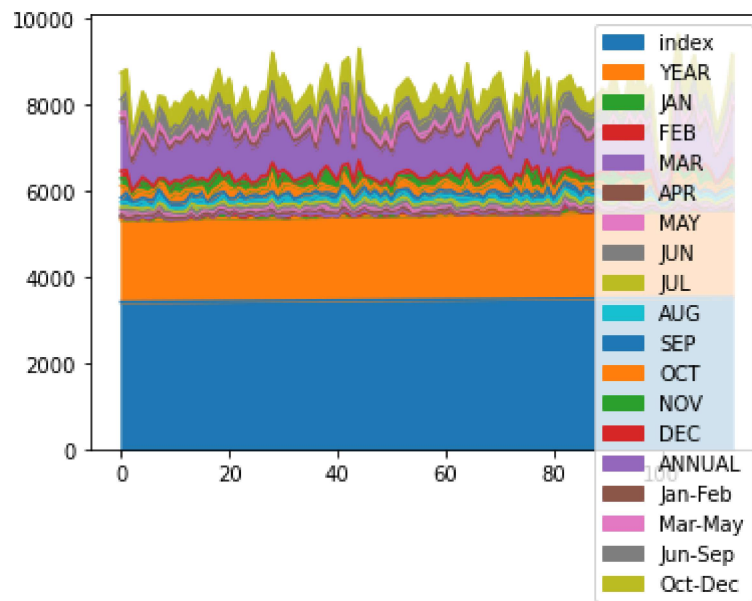
```
Out[9]: <AxesSubplot:ylabel='Frequency'>
```



Area Chart

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

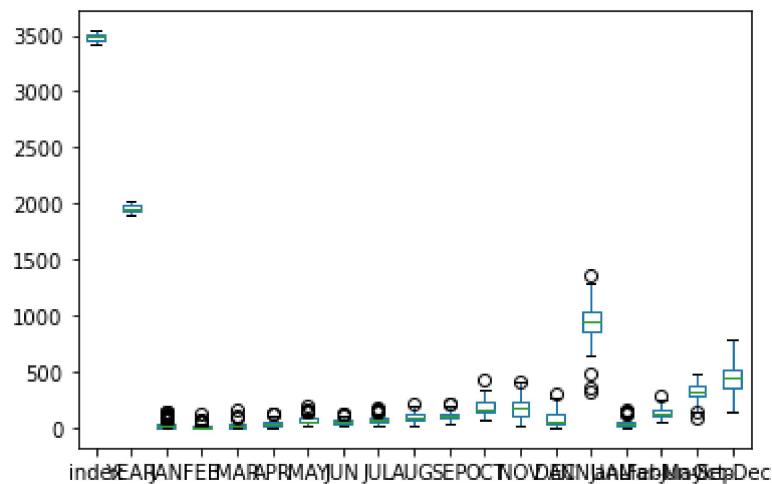
```
Out[10]: <AxesSubplot:>
```



Box Chart

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

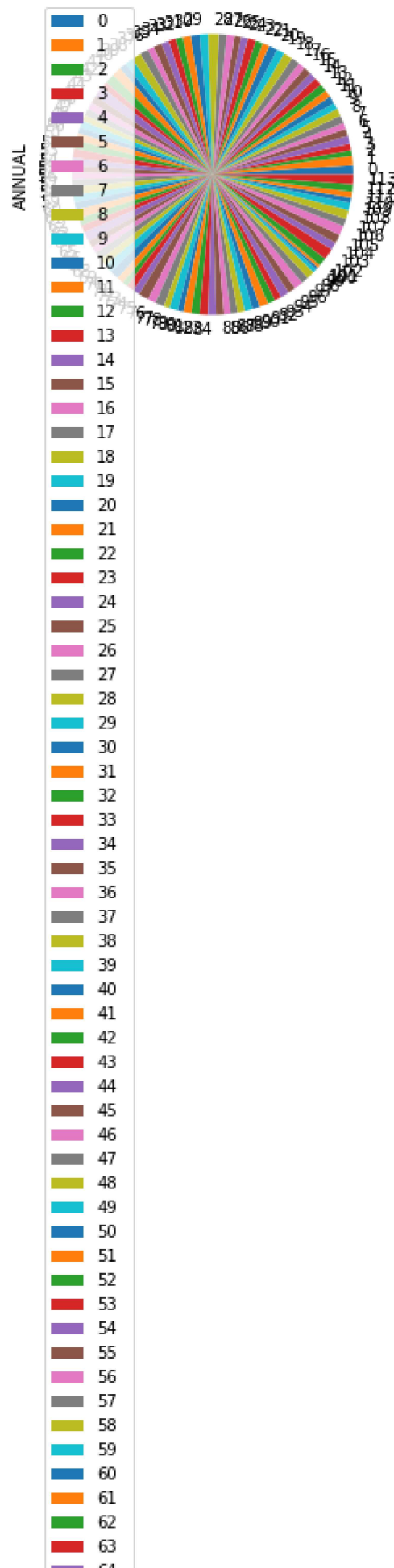
```
Out[11]: <AxesSubplot:>
```

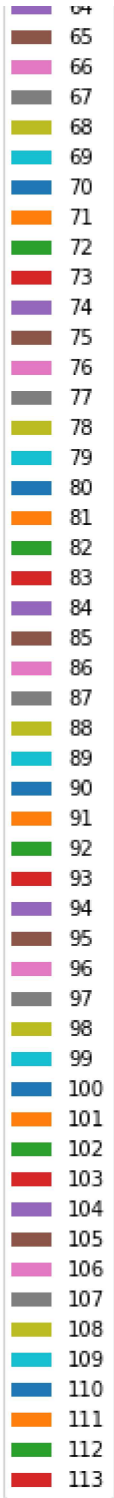


Pie Chart

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

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

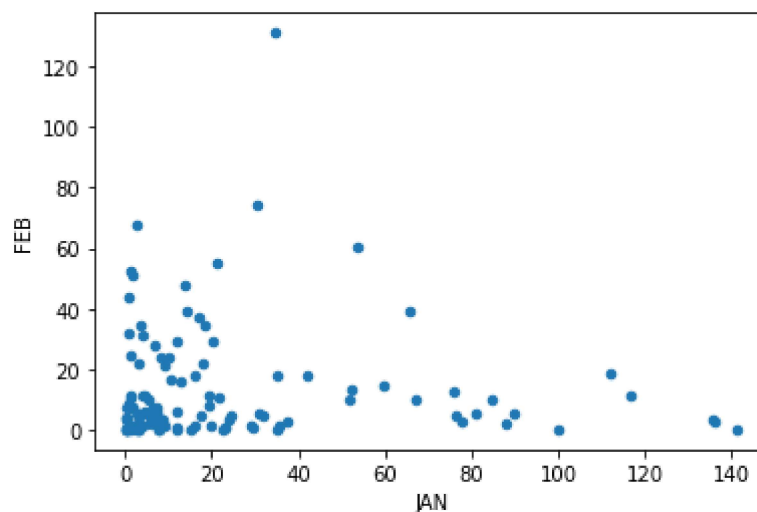





Scatter Plot

```
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()
```

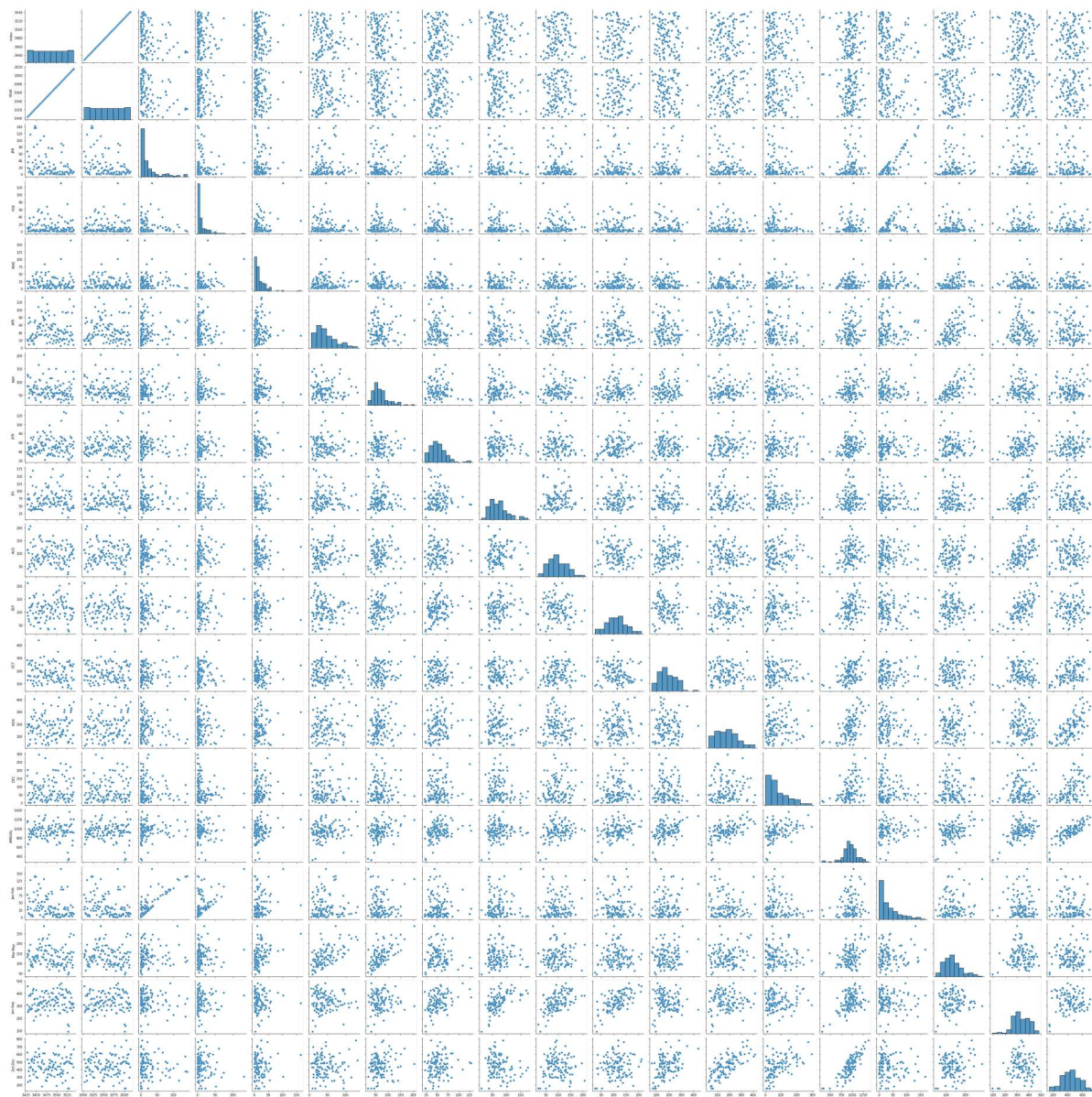
Out[15]:

	index	YEAR	JAN	FEB	MAR	APR	MAY	
count	114.000000	114.000000	114.000000	114.000000	114.000000	114.000000	114.000000	11
mean	3484.500000	1958.500000	23.813158	13.197368	19.456140	45.074561	69.885088	5
std	33.052988	33.052988	32.396065	19.436488	22.487869	28.030608	31.913229	1
min	3428.000000	1902.000000	0.100000	0.000000	0.000000	5.500000	19.800000	2
25%	3456.250000	1930.250000	2.925000	1.200000	5.075000	23.625000	49.825000	3
50%	3484.500000	1958.500000	9.600000	5.450000	11.900000	37.200000	61.000000	4
75%	3512.750000	1986.750000	29.500000	17.675000	26.700000	59.650000	82.700000	6
max	3541.000000	2015.000000	141.200000	131.300000	164.700000	132.100000	204.400000	12

EDA And Visualization

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

```
Out[16]: <seaborn.axisgrid.PairGrid at 0x1b2782e1460>
```

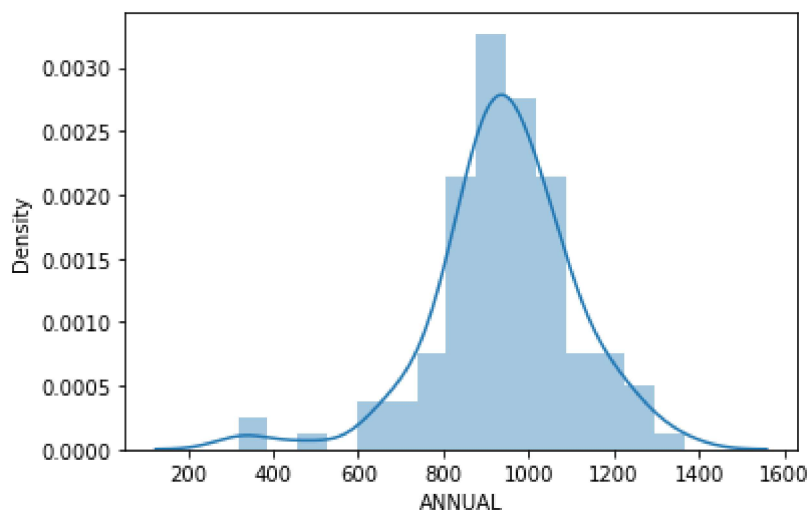


```
In [17]: sns.distplot(df['ANNUAL'])
```

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:>
```

