

## Importing Libraries

In [1]:

```
1 import numpy as np
2 import pandas as pd
3 import matplotlib.pyplot as plt
4 import seaborn as sns
```

## Importing Datasets

In [2]:

```

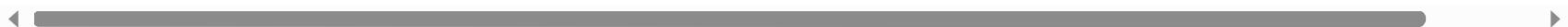
1 df=pd.read_csv('MADHYA MAHARASHTRA.csv')
2 df

```

Out[2]:

		index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	ANNUAL	Jan-Feb	Mar-May	J
0	2623		MADHYA MAHARASHTRA	1902	7.8	0.0	0.1	5.0	9.8	102.6	210.9	114.5	169.5	60.4	40.5	62.9	784.0	7.8	14.9	59
1	2624		MADHYA MAHARASHTRA	1903	7.6	0.0	0.0	3.2	77.2	86.3	281.8	155.5	142.3	74.2	7.6	2.2	837.9	7.6	80.4	66
2	2625		MADHYA MAHARASHTRA	1904	0.4	4.7	1.7	3.0	18.7	114.6	126.5	59.5	183.0	91.1	0.0	0.4	603.5	5.1	23.4	48
3	2626		MADHYA MAHARASHTRA	1905	0.0	1.2	0.0	2.3	23.6	65.0	252.8	79.0	52.6	52.9	8.3	0.0	537.8	1.2	25.9	44
4	2627		MADHYA MAHARASHTRA	1906	10.5	0.8	0.0	0.1	9.3	184.8	199.3	205.0	88.8	19.1	15.5	11.1	744.3	11.3	9.4	67
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
109	2732		MADHYA MAHARASHTRA	2011	0.0	0.3	0.3	5.0	2.9	133.3	261.4	238.1	148.4	62.8	0.0	0.0	852.6	0.3	8.2	78
110	2733		MADHYA MAHARASHTRA	2012	0.0	0.0	0.0	3.0	1.4	67.9	203.0	187.8	129.5	95.2	2.2	0.0	689.8	0.0	4.4	58
111	2734		MADHYA MAHARASHTRA	2013	0.1	5.3	0.8	5.7	6.0	212.4	311.8	147.0	210.3	57.8	4.0	1.3	962.4	5.3	12.4	88
112	2735		MADHYA MAHARASHTRA	2014	3.1	6.2	24.4	7.5	29.8	44.0	277.9	240.3	120.4	38.5	32.8	13.1	838.0	9.3	61.7	68
113	2736		MADHYA MAHARASHTRA	2015	1.4	0.8	41.2	9.6	24.4	177.0	111.7	67.2	146.6	48.3	16.2	0.1	644.5	2.2	75.3	50

114 rows × 20 columns



## Data Cleaning and Data Preprocessing

```
In [3]: 1 df=df.dropna()  
2 df
```

Out[3]:

		index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	ANNUAL	Jan-Feb	Mar-May	J
0	2623		MADHYA MAHARASHTRA	1902	7.8	0.0	0.1	5.0	9.8	102.6	210.9	114.5	169.5	60.4	40.5	62.9	784.0	7.8	14.9	59
1	2624		MADHYA MAHARASHTRA	1903	7.6	0.0	0.0	3.2	77.2	86.3	281.8	155.5	142.3	74.2	7.6	2.2	837.9	7.6	80.4	66
2	2625		MADHYA MAHARASHTRA	1904	0.4	4.7	1.7	3.0	18.7	114.6	126.5	59.5	183.0	91.1	0.0	0.4	603.5	5.1	23.4	48
3	2626		MADHYA MAHARASHTRA	1905	0.0	1.2	0.0	2.3	23.6	65.0	252.8	79.0	52.6	52.9	8.3	0.0	537.8	1.2	25.9	44
4	2627		MADHYA MAHARASHTRA	1906	10.5	0.8	0.0	0.1	9.3	184.8	199.3	205.0	88.8	19.1	15.5	11.1	744.3	11.3	9.4	67
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
109	2732		MADHYA MAHARASHTRA	2011	0.0	0.3	0.3	5.0	2.9	133.3	261.4	238.1	148.4	62.8	0.0	0.0	852.6	0.3	8.2	78
110	2733		MADHYA MAHARASHTRA	2012	0.0	0.0	0.0	3.0	1.4	67.9	203.0	187.8	129.5	95.2	2.2	0.0	689.8	0.0	4.4	58
111	2734		MADHYA MAHARASHTRA	2013	0.1	5.3	0.8	5.7	6.0	212.4	311.8	147.0	210.3	57.8	4.0	1.3	962.4	5.3	12.4	88
112	2735		MADHYA MAHARASHTRA	2014	3.1	6.2	24.4	7.5	29.8	44.0	277.9	240.3	120.4	38.5	32.8	13.1	838.0	9.3	61.7	68
113	2736		MADHYA MAHARASHTRA	2015	1.4	0.8	41.2	9.6	24.4	177.0	111.7	67.2	146.6	48.3	16.2	0.1	644.5	2.2	75.3	50

114 rows × 20 columns



```
In [4]: 1 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')
```

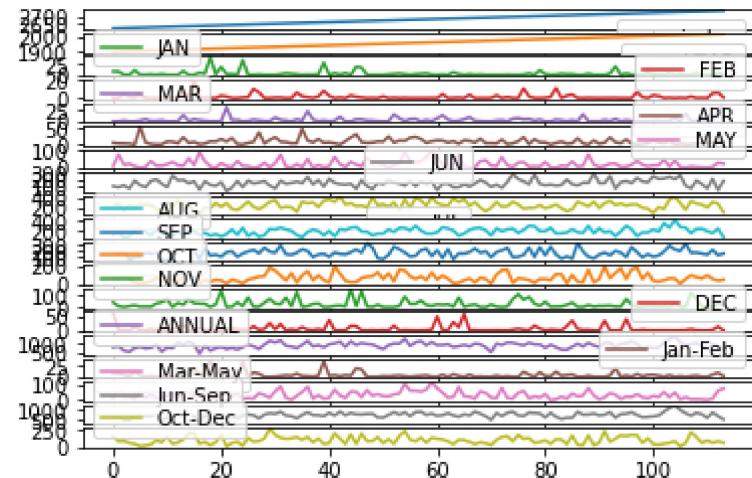
In [5]: 1 df.info()

```
<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

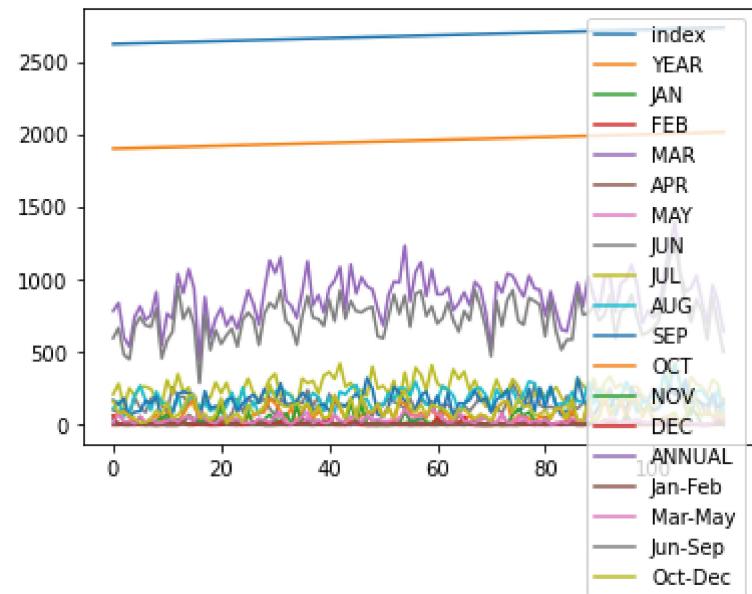
```
In [6]: 1 df.plot.line(subplots=True)
```

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



In [7]: 1 df.plot.line()

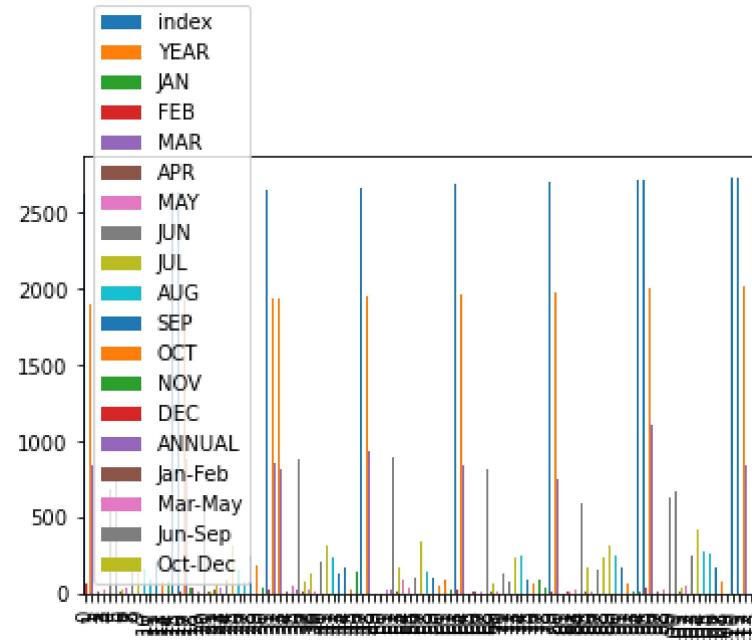
Out[7]: <AxesSubplot:>



## Bar Chart

In [8]: 1 df.plot.bar()

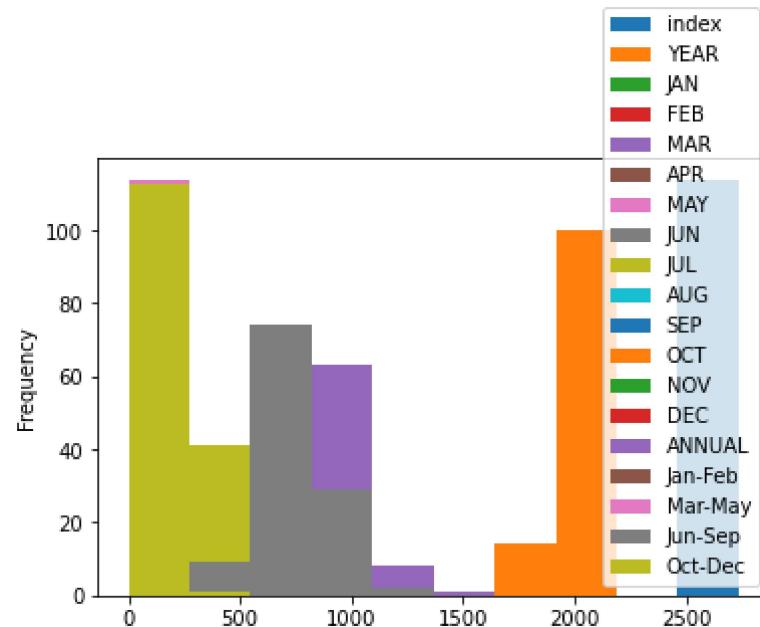
Out[8]: <AxesSubplot:>



## Histogram

In [9]: 1 df.plot.hist()

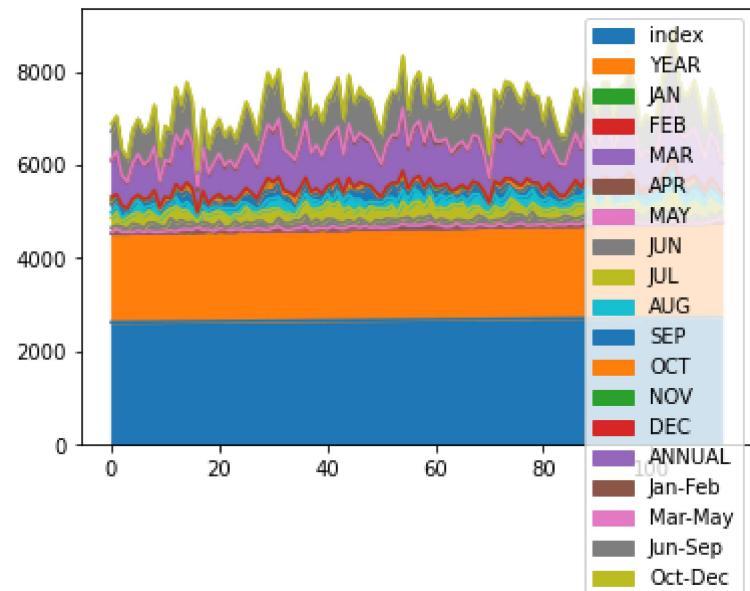
Out[9]: <AxesSubplot:ylabel='Frequency'>



## Area Chart

In [10]: 1 df.plot.area()

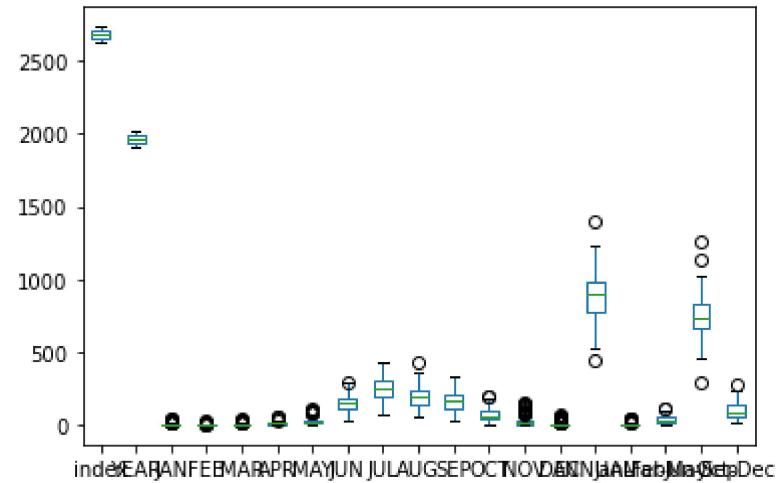
Out[10]: <AxesSubplot:>



## Box Chart

```
In [11]: 1 df.plot.box()
```

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

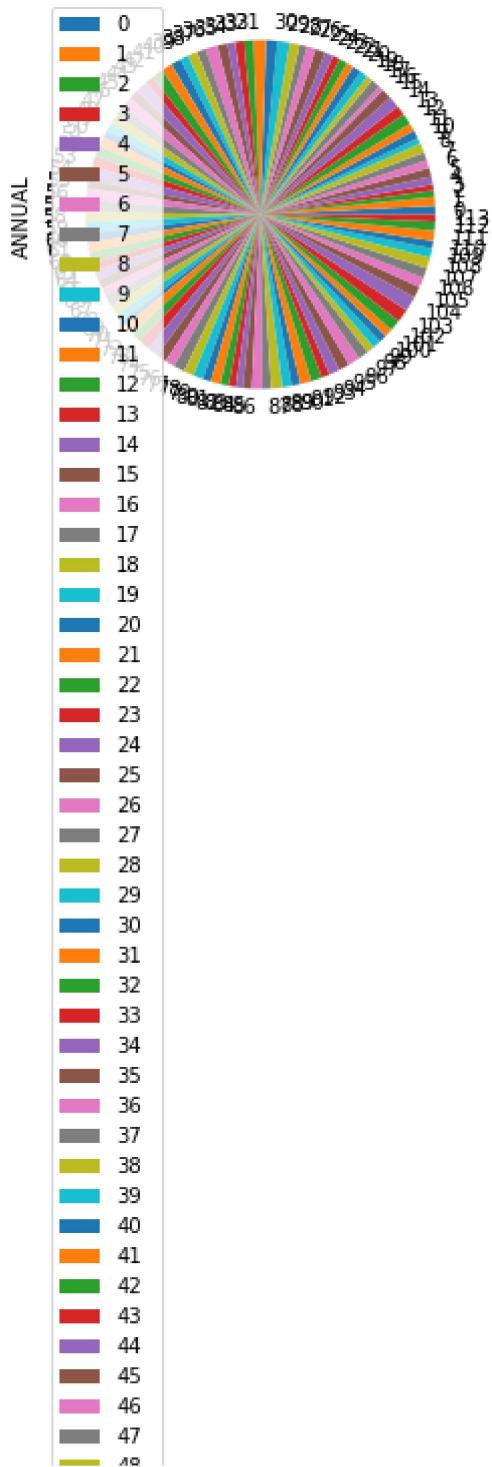


## Pie Chart

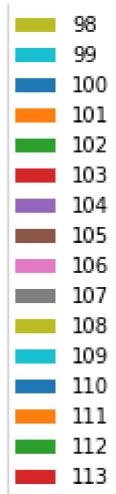
```
In [12]: 1 df.plot.pie(y='ANNUAL')
```

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





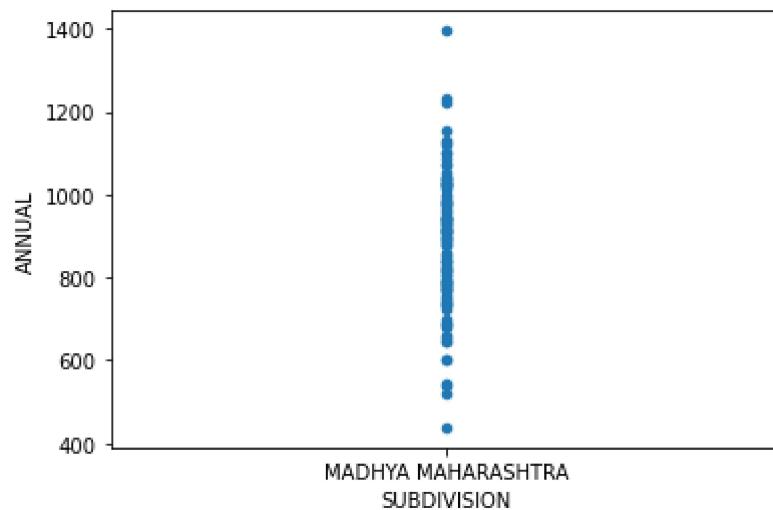
40
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97



## Scatter Plot

```
In [13]: 1 df.plot.scatter(x='SUBDIVISION',y='ANNUAL')
```

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



In [14]: 1 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]: 1 df.describe()
```

Out[15]:

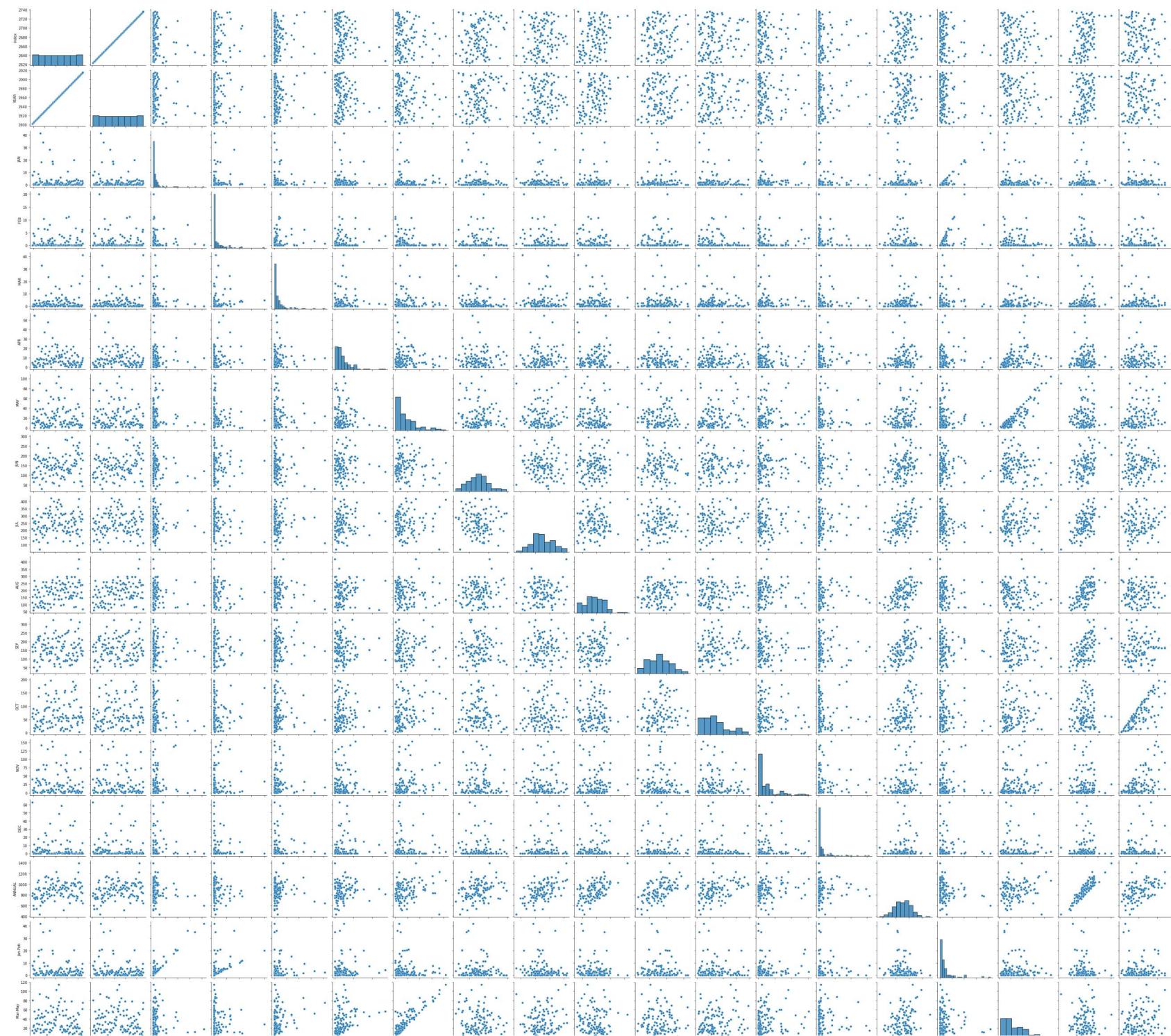
	index	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG
count	114.000000	114.000000	114.000000	114.000000	114.000000	114.000000	114.000000	114.000000	114.000000	114.0
mean	2679.500000	1958.500000	2.916667	1.475439	3.560526	8.906140	22.878070	147.774561	249.270175	184.312281
std	33.052988	33.052988	6.528737	2.927005	6.428251	9.080294	22.455804	55.204323	74.035891	67.789350
min	2623.000000	1902.000000	0.000000	0.000000	0.000000	0.000000	0.300000	30.700000	68.400000	59.500000
25%	2651.250000	1930.250000	0.000000	0.000000	0.200000	3.200000	7.225000	110.025000	198.750000	139.025000
50%	2679.500000	1958.500000	0.650000	0.200000	1.450000	6.250000	15.150000	146.550000	247.950000	186.250000
75%	2707.750000	1986.750000	2.575000	1.625000	4.100000	11.825000	33.050000	180.050000	304.950000	232.050000
max	2736.000000	2015.000000	41.500000	20.000000	41.200000	54.500000	104.200000	293.800000	421.700000	421.700000

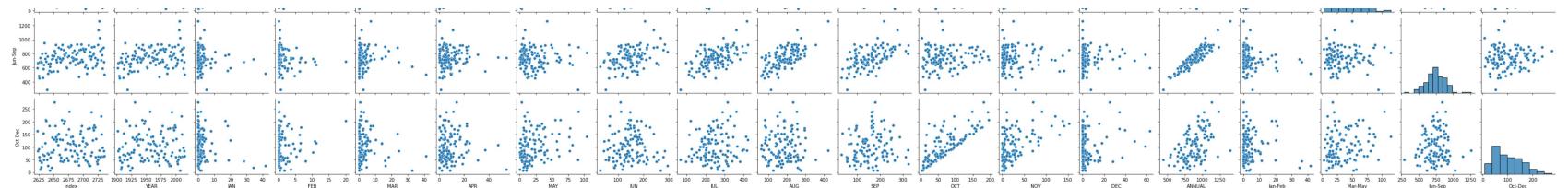
## EDA And Visualization

```
In [16]: 1 sns.pairplot(df)
```

```
Out[16]: <seaborn.axisgrid.PairGrid at 0x210a4123c70>
```



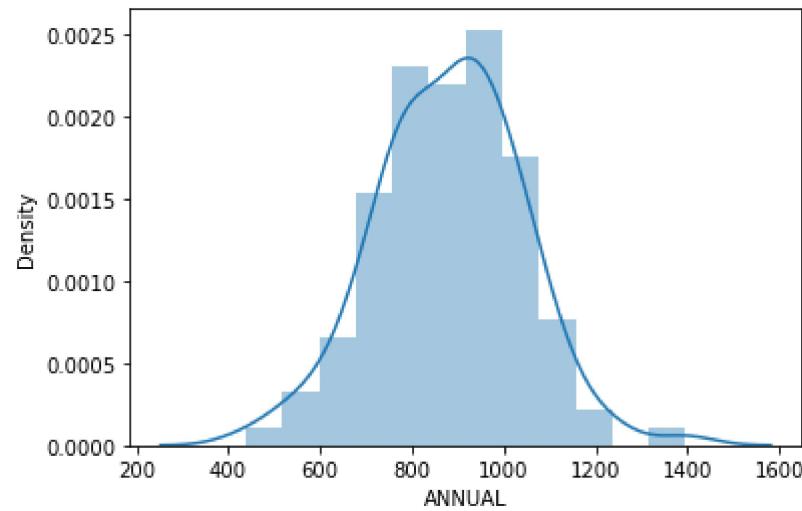




```
In [17]: 1 sns.distplot(df['ANNUAL'])
```

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]: 1 sns.heatmap(df.corr())
```

Out[18]: <AxesSubplot:>

