

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('KONKAN GOA.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
0	2508		KONKAN & GOA	1902	0.3	0.0	0.0	0.4	7.6	428.2	943.6	515.1	613.8	74.3	42.7	48.0	2673.9	0.3	8.0
1	2509		KONKAN & GOA	1903	0.0	0.0	0.1	0.0	201.1	470.5	1298.6	673.9	285.1	140.8	12.4	1.7	3084.3	0.0	201.2
2	2510		KONKAN & GOA	1904	0.0	0.1	6.6	6.3	4.6	975.8	771.7	321.3	217.0	90.3	0.0	0.0	2393.7	0.1	17.5
3	2511		KONKAN & GOA	1905	0.1	0.1	0.0	0.4	8.6	293.7	770.6	305.5	208.3	83.5	12.1	0.0	1682.8	0.2	9.0
4	2512		KONKAN & GOA	1906	5.0	0.9	0.0	0.0	2.9	547.4	1090.9	506.7	222.5	31.6	10.2	0.7	2418.9	5.9	2.9
...	
109	2617		KONKAN & GOA	2011	0.0	0.0	0.0	3.4	1.1	857.0	1384.1	987.9	468.3	120.3	3.1	0.0	3825.2	0.0	4.5
110	2618		KONKAN & GOA	2012	0.0	0.0	0.0	0.6	1.1	633.0	928.5	762.5	515.3	175.1	2.3	0.0	3018.4	0.0	1.7
111	2619		KONKAN & GOA	2013	1.8	5.4	0.1	0.1	18.5	1028.3	1478.5	497.6	340.7	149.3	2.1	1.5	3524.0	7.3	18.7
112	2620		KONKAN & GOA	2014	1.3	5.3	1.8	0.7	21.3	238.2	1293.2	658.0	419.5	98.7	8.0	11.7	2757.5	6.6	23.8
113	2621		KONKAN & GOA	2015	2.7	0.0	36.8	3.6	11.3	764.0	526.5	377.3	240.9	91.4	27.3	0.0	2082.0	2.7	51.7

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
0	2508		KONKAN & GOA	1902	0.3	0.0	0.0	0.4	7.6	428.2	943.6	515.1	613.8	74.3	42.7	48.0	2673.9	0.3	8.0
1	2509		KONKAN & GOA	1903	0.0	0.0	0.1	0.0	201.1	470.5	1298.6	673.9	285.1	140.8	12.4	1.7	3084.3	0.0	201.2
2	2510		KONKAN & GOA	1904	0.0	0.1	6.6	6.3	4.6	975.8	771.7	321.3	217.0	90.3	0.0	0.0	2393.7	0.1	17.5
3	2511		KONKAN & GOA	1905	0.1	0.1	0.0	0.4	8.6	293.7	770.6	305.5	208.3	83.5	12.1	0.0	1682.8	0.2	9.0
4	2512		KONKAN & GOA	1906	5.0	0.9	0.0	0.0	2.9	547.4	1090.9	506.7	222.5	31.6	10.2	0.7	2418.9	5.9	2.9
...	
109	2617		KONKAN & GOA	2011	0.0	0.0	0.0	3.4	1.1	857.0	1384.1	987.9	468.3	120.3	3.1	0.0	3825.2	0.0	4.5
110	2618		KONKAN & GOA	2012	0.0	0.0	0.0	0.6	1.1	633.0	928.5	762.5	515.3	175.1	2.3	0.0	3018.4	0.0	1.7
111	2619		KONKAN & GOA	2013	1.8	5.4	0.1	0.1	18.5	1028.3	1478.5	497.6	340.7	149.3	2.1	1.5	3524.0	7.3	18.7
112	2620		KONKAN & GOA	2014	1.3	5.3	1.8	0.7	21.3	238.2	1293.2	658.0	419.5	98.7	8.0	11.7	2757.5	6.6	23.8
113	2621		KONKAN & GOA	2015	2.7	0.0	36.8	3.6	11.3	764.0	526.5	377.3	240.9	91.4	27.3	0.0	2082.0	2.7	51.7

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')
```

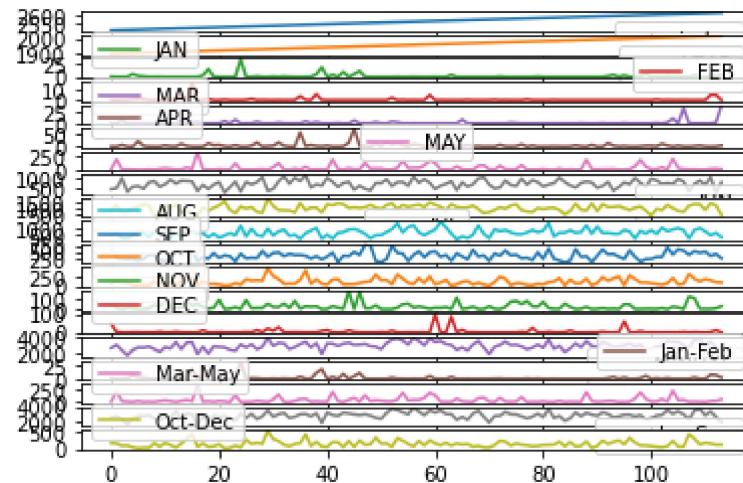
In [5]: 1 df.info()

```
<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

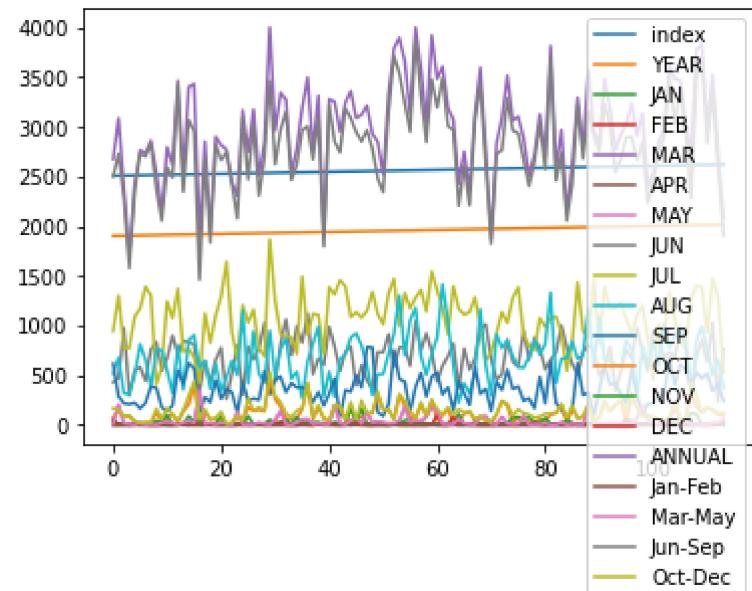
```
In [6]: 1 df.plot.line(subplots=True)
```

```
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()

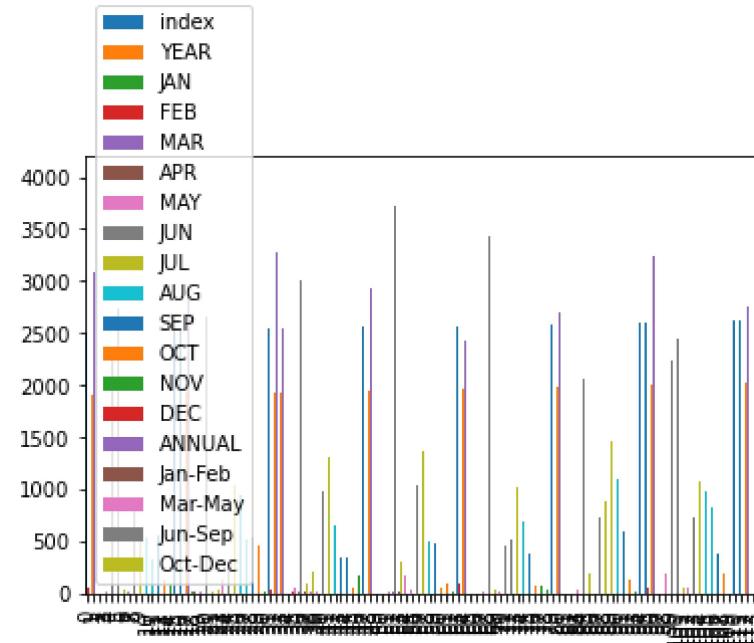
Out[7]: <AxesSubplot:>



Bar Chart

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

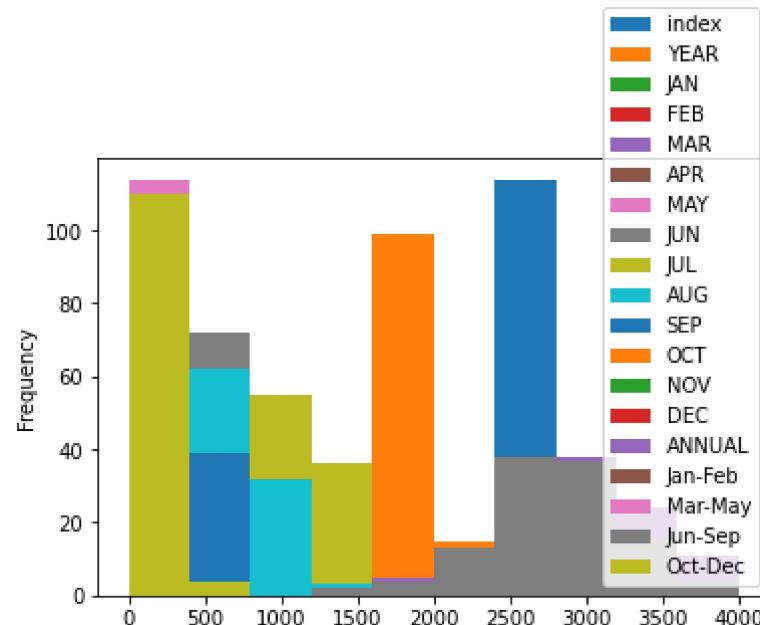
Out[8]: <AxesSubplot:>



Histogram

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

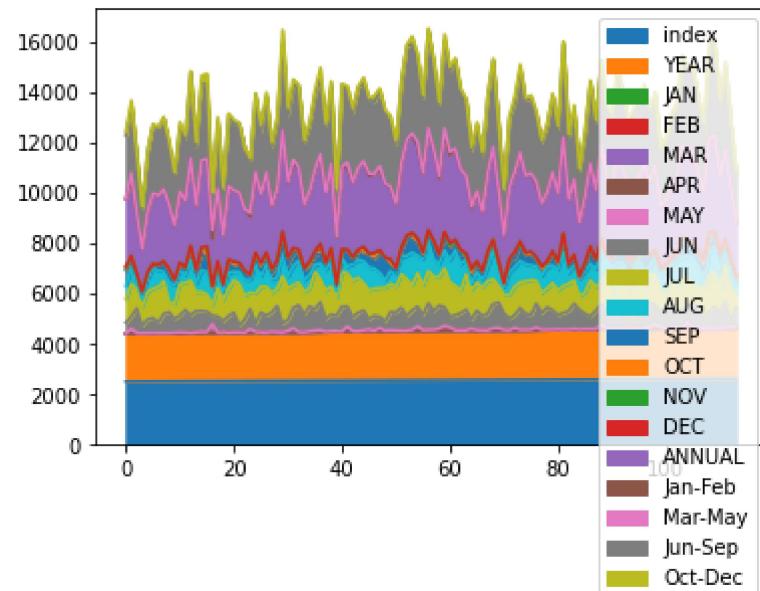
Out[9]: <AxesSubplot:ylabel='Frequency'>



Area Chart

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

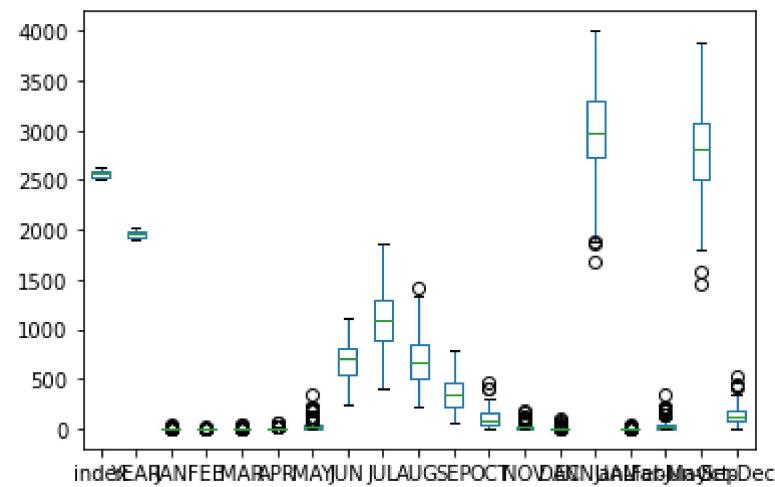
Out[10]: <AxesSubplot:>



Box Chart

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

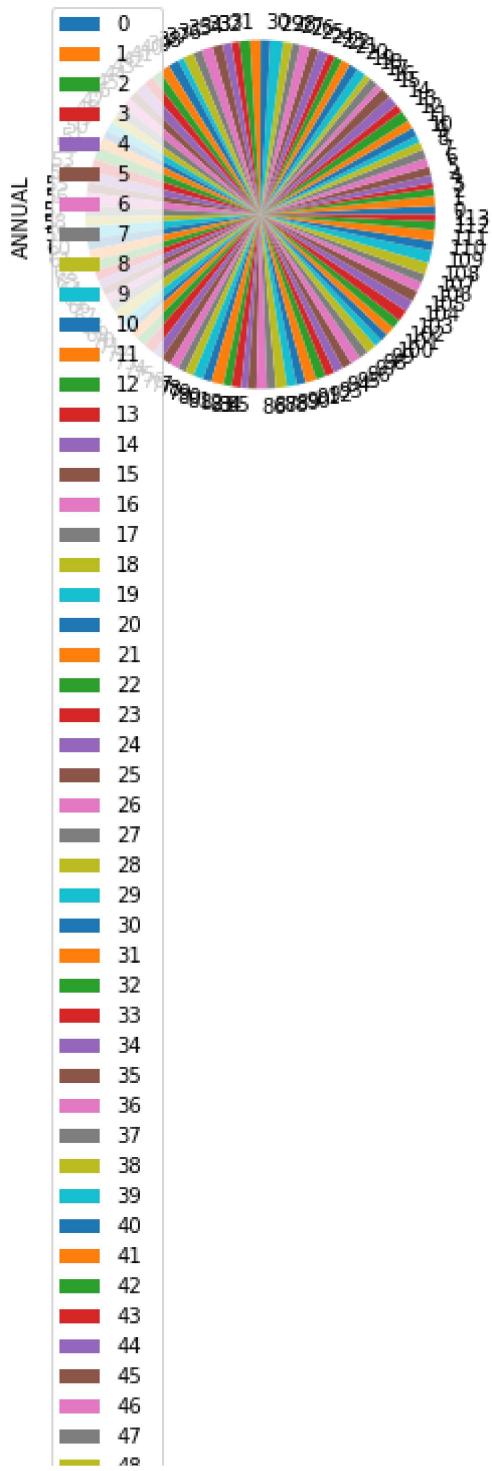
```
Out[11]: <AxesSubplot:>
```



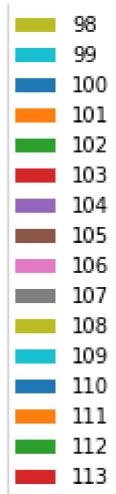
Pie Chart

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

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

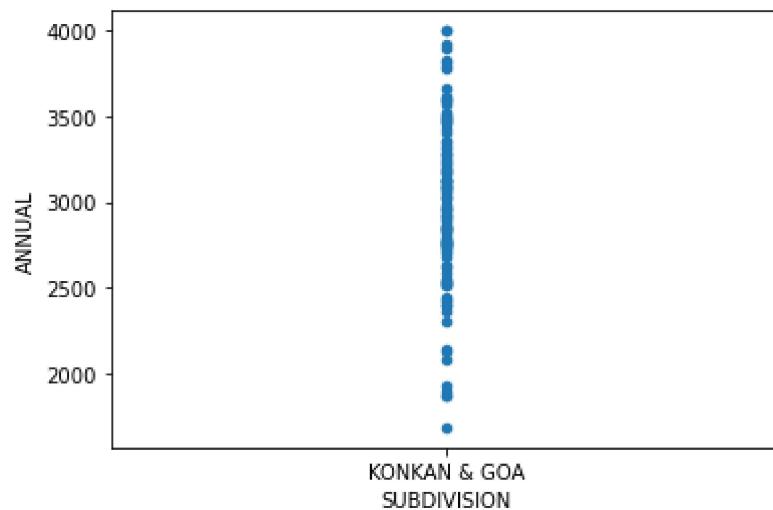
40
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97



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()
```

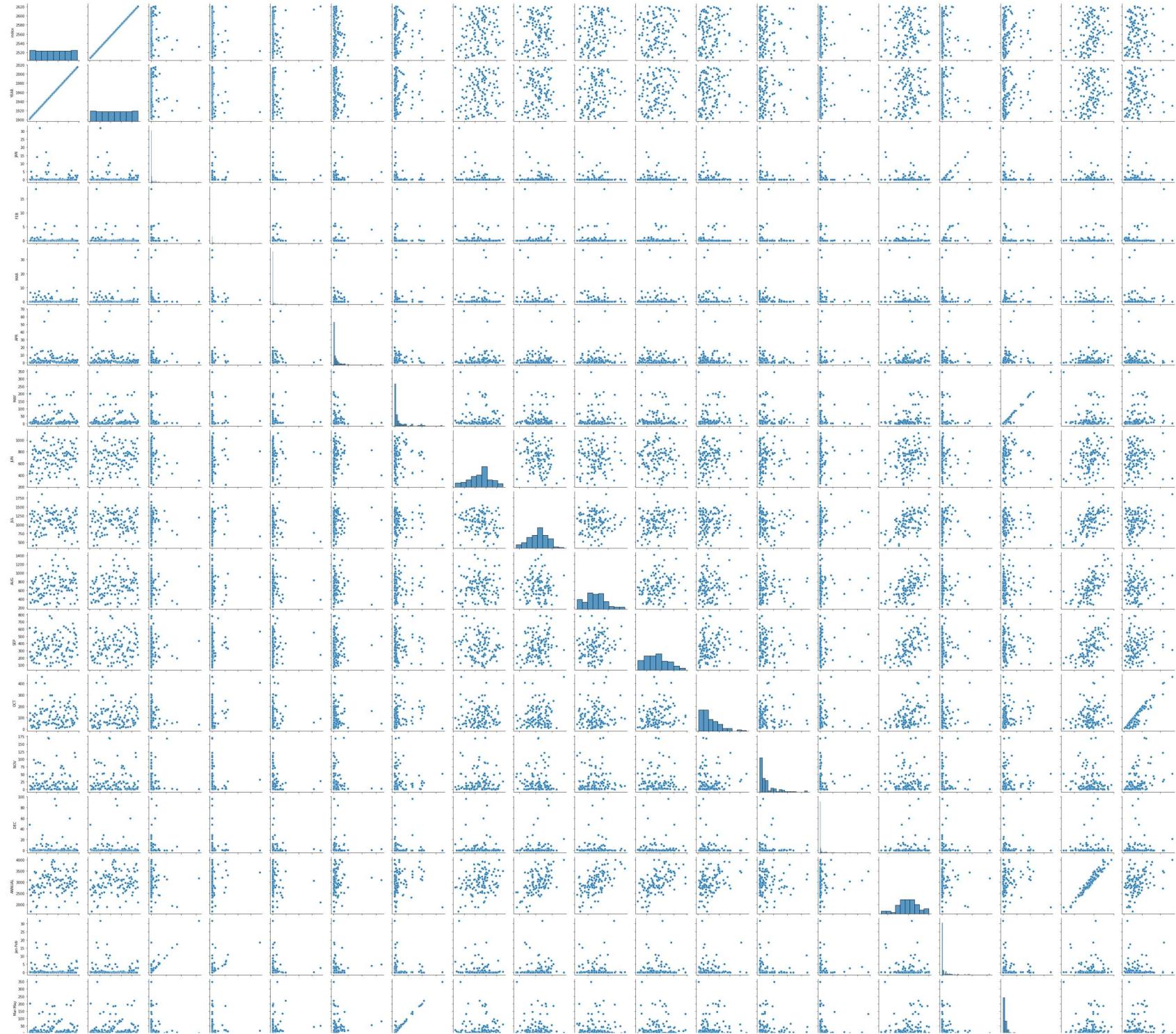
Out[15]:

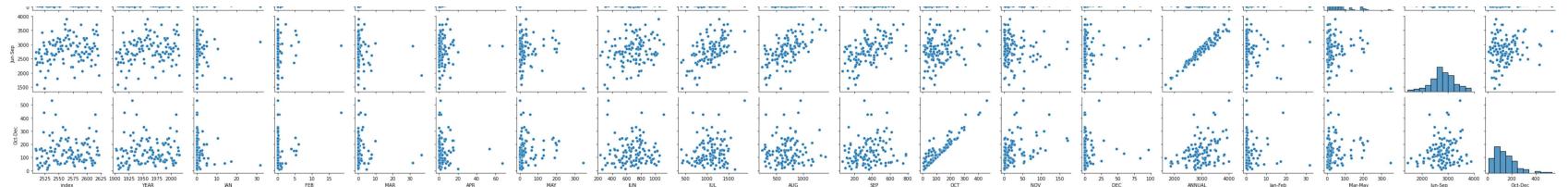
	index	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	
count	114.000000	114.000000	114.000000	114.000000	114.000000	114.000000	114.000000	114.000000	114.000000	114.000000	11
mean	2564.500000	1958.500000	1.224561	0.550877	1.383333	3.990351	33.635088	688.064912	1073.008772	682.184211	35
std	33.052988	33.052988	3.879799	2.056642	4.769388	8.647522	58.570653	198.034071	276.321871	250.888801	16
min	2508.000000	1902.000000	0.000000	0.000000	0.000000	0.000000	0.000000	238.200000	407.800000	224.100000	6
25%	2536.250000	1930.250000	0.000000	0.000000	0.000000	0.300000	2.900000	549.775000	895.325000	508.775000	22
50%	2564.500000	1958.500000	0.000000	0.000000	0.050000	1.250000	9.650000	703.600000	1096.700000	663.750000	33
75%	2592.750000	1986.750000	0.450000	0.100000	0.375000	4.075000	30.450000	806.225000	1288.500000	844.600000	45
max	2621.000000	2015.000000	31.800000	18.400000	36.800000	67.300000	345.400000	1118.100000	1867.200000	1414.700000	78

EDA And Visualization

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

```
Out[16]: <seaborn.axisgrid.PairGrid at 0x219ad7ca790>
```

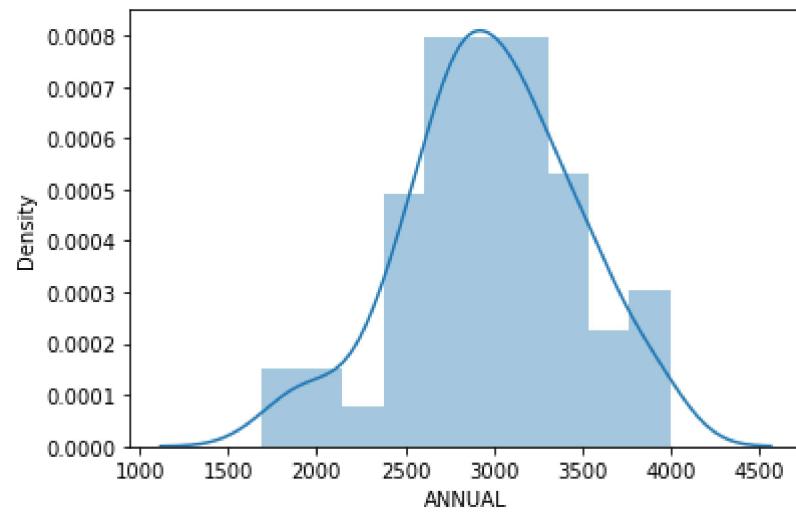





```
In [17]: 1 sns.distplot(df['ANNUAL'])
```

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:>
```

