Importing Libraries

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

Importing Datasets

In [2]:
```
df=pd.read_csv("rainfall_konkan _ goa.csv")
df
```

Out[2]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV
0	2507	KONKAN & GOA	1901	5.6	0.1	0.4	35.7	19.9	746.1	1075.5	748.0	117.4	38.6	5.4
1	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
2	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
3	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
4	2511	KONKAN & GOA	1905	0.1	0.1	0.0	0.4	8.6	293.7	770.6	305.5	208.3	83.5	12.7
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
110	2617	KONKAN & GOA	2011	0.0	0.0	0.0	3.4	1.1	857.0	1384.1	987.9	468.3	120.3	3.7
111	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
112	2619	KONKAN & GOA	2013	1.8	5.4	0.1	0.1	18.5	1028.3	1478.5	497.6	340.7	149.3	2.7
113	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
114	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

115 rows × 20 columns

head

In [3]:
```
df.head(5)
df
```

Out[3]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV
0	2507	KONKAN & GOA	1901	5.6	0.1	0.4	35.7	19.9	746.1	1075.5	748.0	117.4	38.6	5.4
1	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
2	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
3	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
4	2511	KONKAN & GOA	1905	0.1	0.1	0.0	0.4	8.6	293.7	770.6	305.5	208.3	83.5	12.7
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
110	2617	KONKAN & GOA	2011	0.0	0.0	0.0	3.4	1.1	857.0	1384.1	987.9	468.3	120.3	3.7
111	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
112	2619	KONKAN & GOA	2013	1.8	5.4	0.1	0.1	18.5	1028.3	1478.5	497.6	340.7	149.3	2.7
113	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
114	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

115 rows × 20 columns

tail

In [4]:
```
df.tail(5)
df
```

Out[4]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV
0	2507	KONKAN & GOA	1901	5.6	0.1	0.4	35.7	19.9	746.1	1075.5	748.0	117.4	38.6	5.4
1	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
2	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

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV
3	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.(
4	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
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	.
110	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
111	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
112	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
113	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.(
114	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

115 rows × 20 columns

Data Cleaning and Data Preprocessing

describe()

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

Out[5]:

	index	YEAR	JAN	FEB	MAR	APR	MAY	JUI
count	115.000000	115.000000	115.000000	115.000000	115.000000	115.000000	115.000000	115.00000
mean	2564.000000	1958.000000	1.262609	0.546957	1.374783	4.266087	33.515652	688.56956
std	33.341666	33.341666	3.884234	2.048034	4.749309	9.103141	58.327263	197.23784
min	2507.000000	1901.000000	0.000000	0.000000	0.000000	0.000000	0.000000	238.20000
25%	2535.500000	1929.500000	0.000000	0.000000	0.000000	0.300000	2.900000	552.15000
50%	2564.000000	1958.000000	0.000000	0.000000	0.100000	1.300000	9.700000	706.90000
75%	2592.500000	1986.500000	0.500000	0.100000	0.400000	4.200000	30.300000	804.85000
max	2621.000000	2015.000000	31.800000	18.400000	36.800000	67.300000	345.400000	1118.10000

shape

In [6]:
```
np.shape(df)
```

Out[6]:
```
(115, 20)
```

size

In [7]:
```
np.size(df)
```

Out[7]:
```
2300
```

dropna

In [8]:
```
df=df.dropna()
```

columns

In [9]:
```
df.columns
```

Out[9]:
```
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')
```

info()

In [10]:
```
df.info()
```

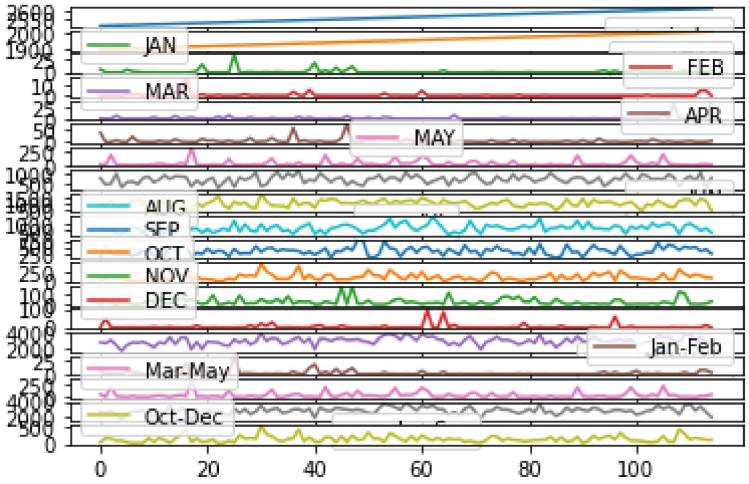
```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 115 entries, 0 to 114
Data columns (total 20 columns):
 #   Column       Non-Null Count  Dtype  
---  ------       --------------  -----  
 0   index        115 non-null    int64  
 1   SUBDIVISION  115 non-null    object 
 2   YEAR         115 non-null    int64  
 3   JAN          115 non-null    float64
 4   FEB          115 non-null    float64
 5   MAR          115 non-null    float64
 6   APR          115 non-null    float64
 7   MAY          115 non-null    float64
 8   JUN          115 non-null    float64
 9   JUL          115 non-null    float64
 10  AUG          115 non-null    float64
 11  SEP          115 non-null    float64
 12  OCT          115 non-null    float64
 13  NOV          115 non-null    float64
 14  DEC          115 non-null    float64
 15  ANNUAL       115 non-null    float64
 16  Jan-Feb      115 non-null    float64
```

```
 17  Mar-May      115 non-null    float64
 18  Jun-Sep      115 non-null    float64
 19  Oct-Dec      115 non-null    float64
dtypes: float64(17), int64(2), object(1)
memory usage: 18.9+ KB
```

Line chart

In [11]:
```
df.plot.line(subplots=True)
```

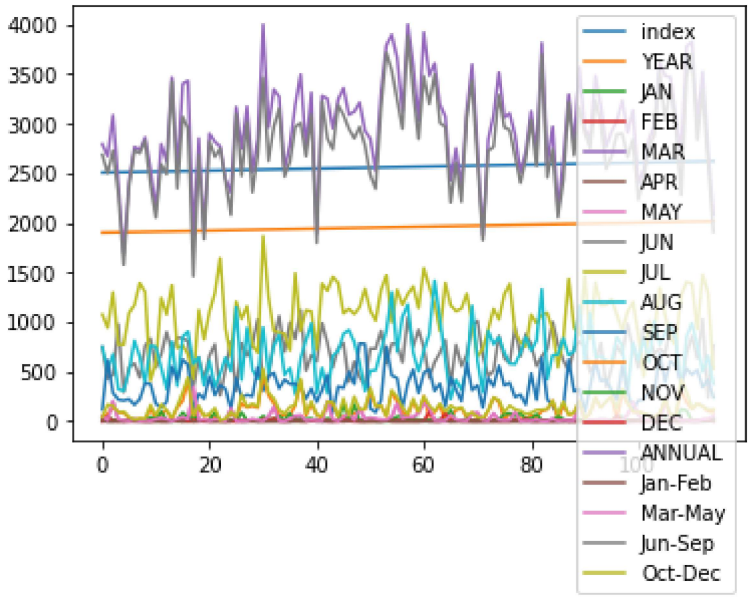
Out[11]:
```
array([<AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>,
       <AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>,
       <AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>,
       <AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>,
       <AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>], dtype=object)
```



Line chart

In [12]:
```
df.plot.line()
```

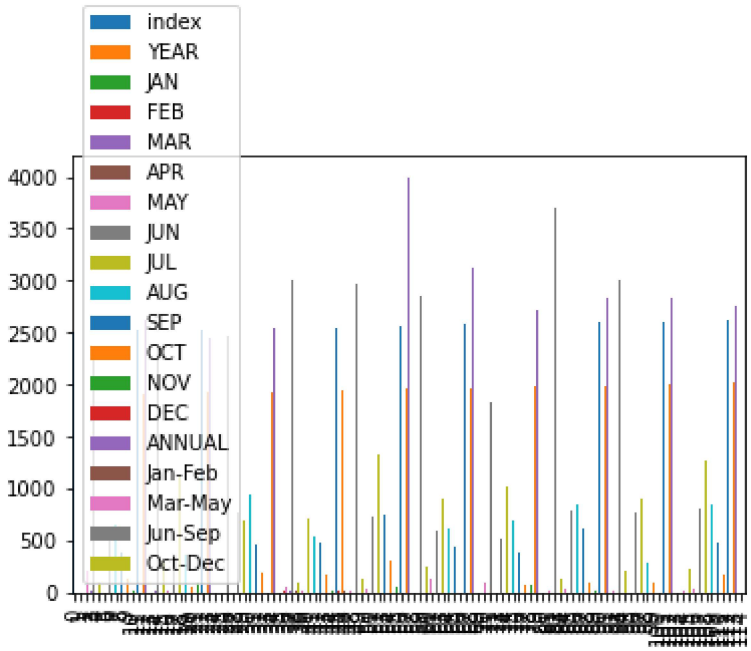
Out[12]:
```
<AxesSubplot:>
```



Bar chart

In [13]:
```
df.plot.bar()
```

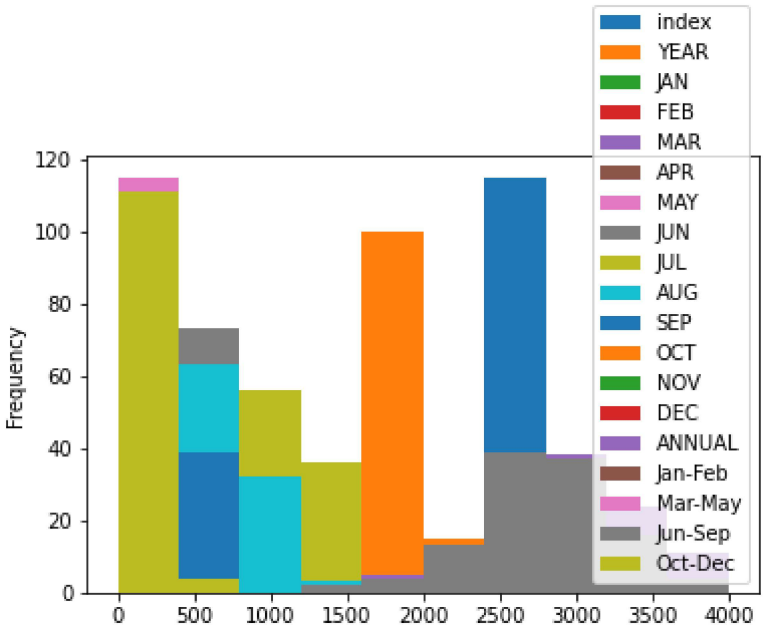
Out[13]: <AxesSubplot:>



Histogram

In [14]:
```
df.plot.hist()
```

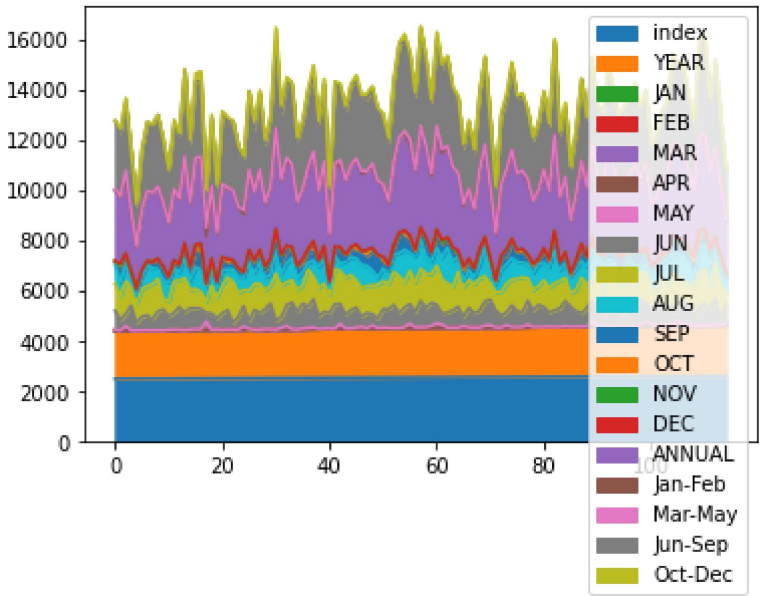
Out[14]: <AxesSubplot:ylabel='Frequency'>



Area chart

In [15]:
```
df.plot.area()
```

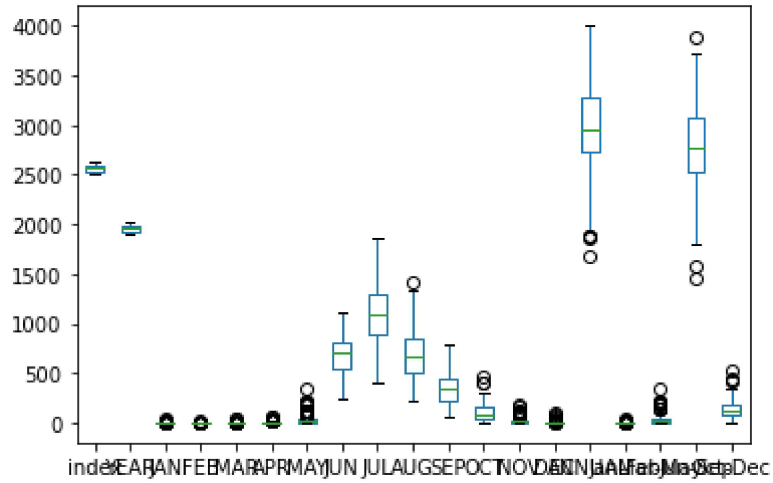
Out[15]: <AxesSubplot:>



Box chart

In [16]:
```
df.plot.box()
```

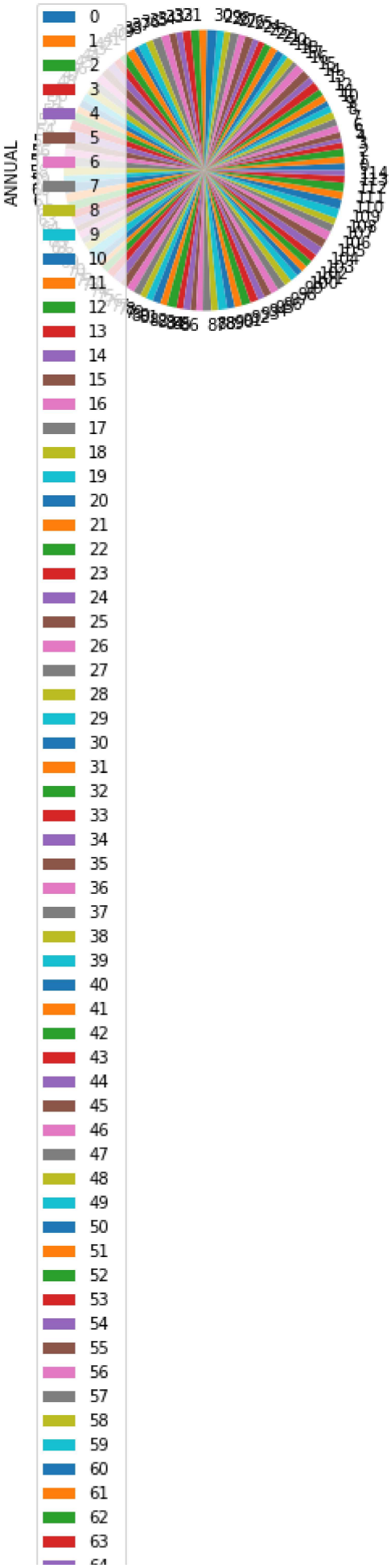
Out[16]: <AxesSubplot:>

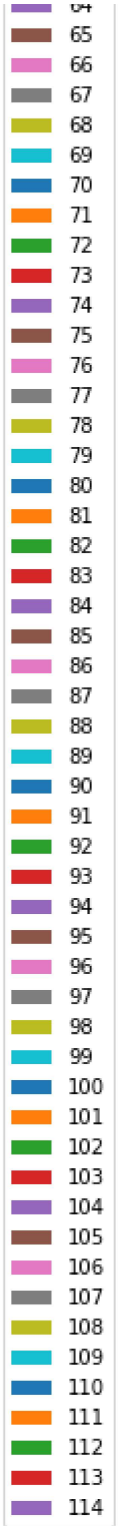


Pie chart

In [17]:
```
df.plot.pie(y='ANNUAL')
```

Out[17]:
```
<AxesSubplot:ylabel='ANNUAL'>
```

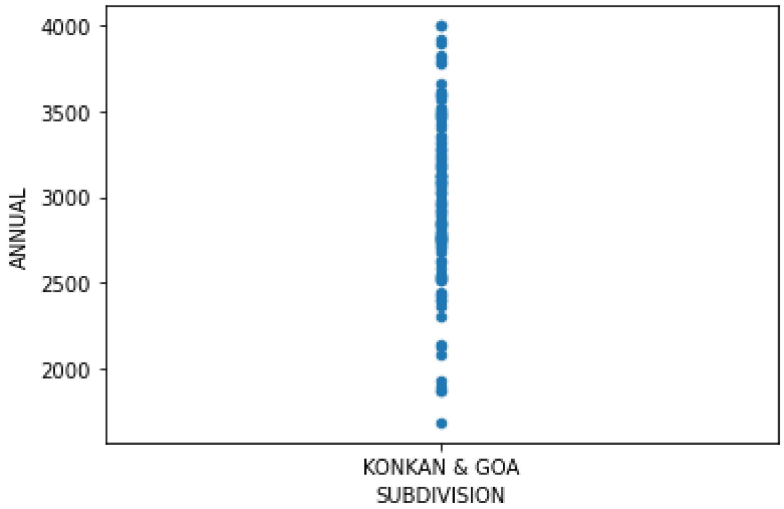

ANNUAL
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63



Scatter chart

In [18]:
```
df.plot.scatter(x='SUBDIVISION' ,y='ANNUAL')
```

Out[18]: <AxesSubplot:xlabel='SUBDIVISION', ylabel='ANNUAL'>



In [19]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 115 entries, 0 to 114
Data columns (total 20 columns):
 #   Column       Non-Null Count  Dtype
---  ------       --------------  -----
 0   index        115 non-null    int64
 1   SUBDIVISION  115 non-null    object
 2   YEAR         115 non-null    int64
 3   JAN          115 non-null    float64
 4   FEB          115 non-null    float64
 5   MAR          115 non-null    float64
 6   APR          115 non-null    float64
 7   MAY          115 non-null    float64
 8   JUN          115 non-null    float64
 9   JUL          115 non-null    float64
 10  AUG          115 non-null    float64
 11  SEP          115 non-null    float64
 12  OCT          115 non-null    float64
 13  NOV          115 non-null    float64
 14  DEC          115 non-null    float64
 15  ANNUAL       115 non-null    float64
 16  Jan-Feb      115 non-null    float64
 17  Mar-May      115 non-null    float64
 18  Jun-Sep      115 non-null    float64
 19  Oct-Dec      115 non-null    float64
dtypes: float64(17), int64(2), object(1)
memory usage: 18.9+ KB
```

In [20]:
```
df.describe()
```

Out[20]:

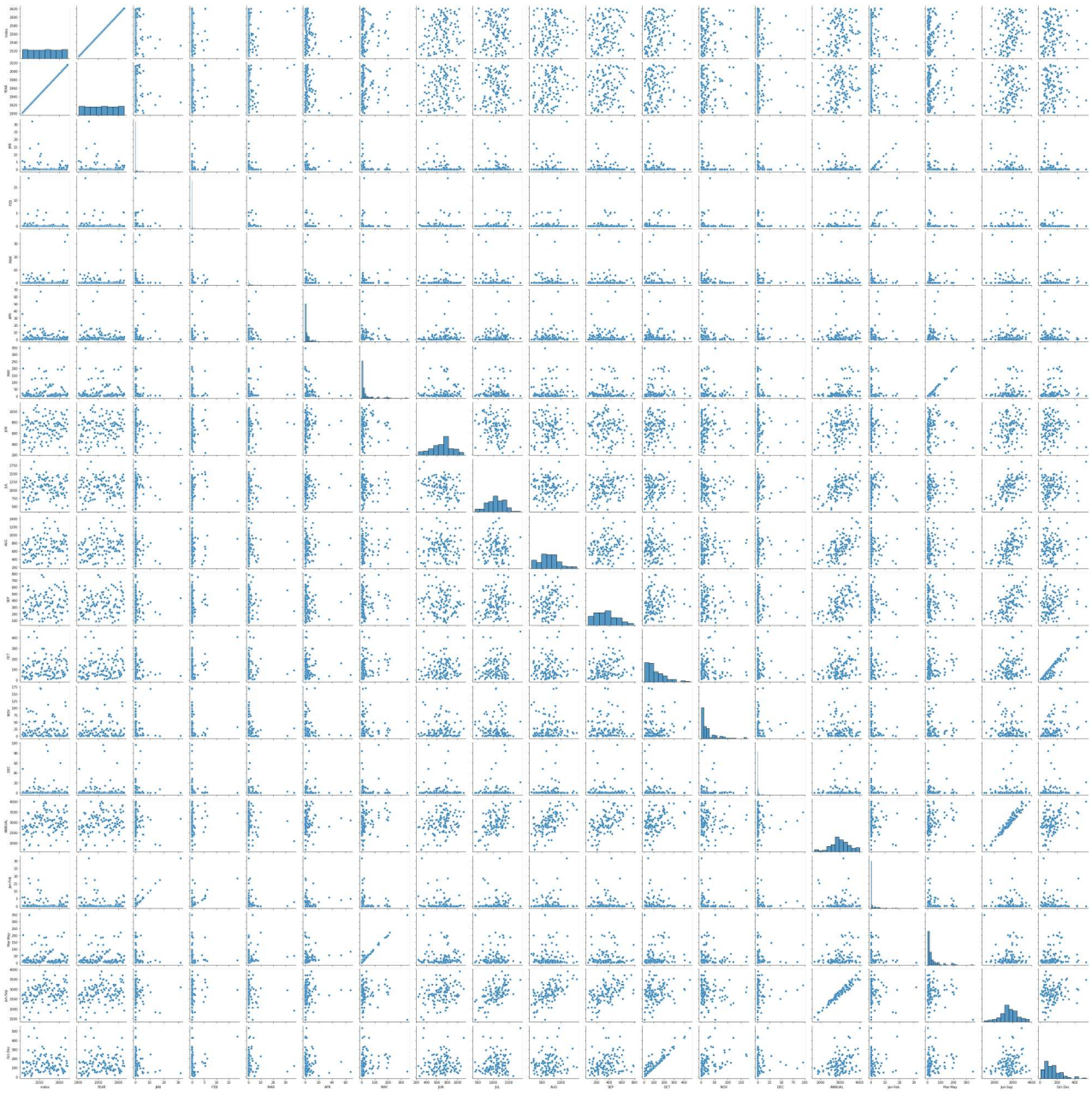
	index	YEAR	JAN	FEB	MAR	APR	MAY	JUI
count	115.000000	115.000000	115.000000	115.000000	115.000000	115.000000	115.000000	115.00000
mean	2564.000000	1958.000000	1.262609	0.546957	1.374783	4.266087	33.515652	688.56956
std	33.341666	33.341666	3.884234	2.048034	4.749309	9.103141	58.327263	197.23784
min	2507.000000	1901.000000	0.000000	0.000000	0.000000	0.000000	0.000000	238.20000
25%	2535.500000	1929.500000	0.000000	0.000000	0.000000	0.300000	2.900000	552.15000

	index	YEAR	JAN	FEB	MAR	APR	MAY	JUI
50%	2564.000000	1958.000000	0.000000	0.000000	0.100000	1.300000	9.700000	706.90000
75%	2592.500000	1986.500000	0.500000	0.100000	0.400000	4.200000	30.300000	804.85000
max	2621.000000	2015.000000	31.800000	18.400000	36.800000	67.300000	345.400000	1118.10000

EDA AND VISUALIZATION

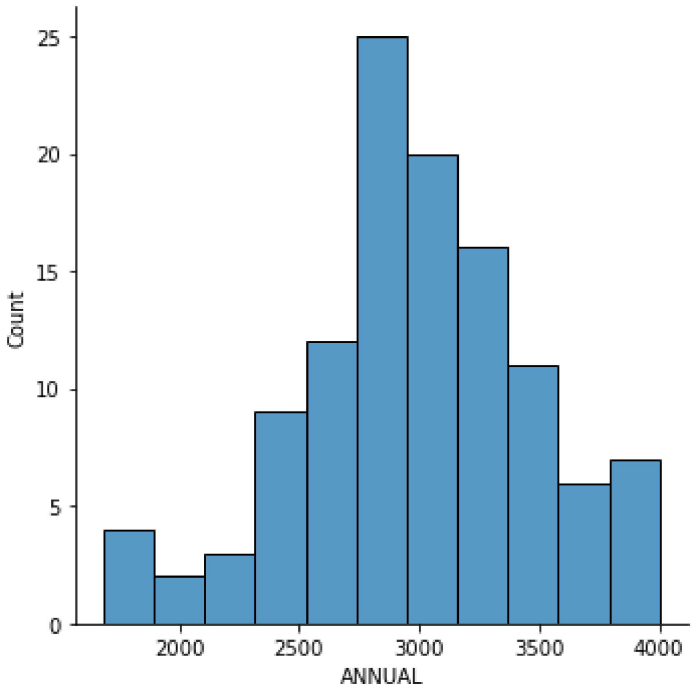
In [21]:
```
sns.pairplot(df)
```

Out[21]: <seaborn.axisgrid.PairGrid at 0x1933b3345e0>



In [22]:
```
sns.displot(df['ANNUAL'])
```

Out[22]: <seaborn.axisgrid.FacetGrid at 0x193461b9430>



In [23]:
```
sns.heatmap(df.corr())
```

Out[23]: <AxesSubplot:>