

# 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_madhya maharashtra.csv")
df
```

Out[2]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV
0	2622	MADHYA MAHARASHTRA	1901	18.8	0.6	7.7	36.6	30.4	107.7	215.9	194.1	83.7	68.7	4.4
1	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
2	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
3	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
4	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
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
110	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
111	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
112	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
113	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
114	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

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	2622	MADHYA MAHARASHTRA	1901	18.8	0.6	7.7	36.6	30.4	107.7	215.9	194.1	83.7	68.7	4.4
1	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
2	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
3	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
4	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
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
110	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
111	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
112	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
113	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
114	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

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	2622	MADHYA MAHARASHTRA	1901	18.8	0.6	7.7	36.6	30.4	107.7	215.9	194.1	83.7	68.7	4.4
1	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
2	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

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

115 rows × 20 columns

## Data Cleaning and Data Preprocessing

### describe()

In [5]:

```
df.describe()
```

Out[5]:

	index	YEAR	JAN	FEB	MAR	APR	MAY	JUN
count	115.000000	115.000000	115.000000	115.000000	115.000000	115.000000	115.000000	115.000000
mean	2679.000000	1958.000000	3.054783	1.467826	3.596522	9.146957	22.943478	147.426087
std	33.341666	33.341666	6.666652	2.915282	6.411625	9.402000	22.368097	55.088562
min	2622.000000	1901.000000	0.000000	0.000000	0.000000	0.000000	0.300000	30.700000
25%	2650.500000	1929.500000	0.000000	0.000000	0.200000	3.200000	7.350000	109.650000
50%	2679.000000	1958.000000	0.700000	0.200000	1.500000	6.300000	15.200000	146.400000
75%	2707.500000	1986.500000	2.600000	1.550000	4.200000	12.100000	33.000000	179.900000
max	2736.000000	2015.000000	41.500000	20.000000	41.200000	54.500000	104.200000	293.800000

### 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()
```

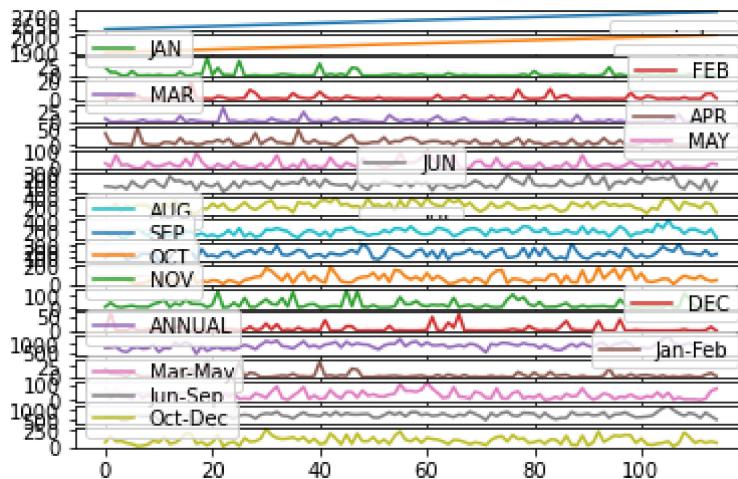
```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 115 entries, 0 to 114
Data columns (total 20 columns):
 #   Column      Non-Null Count  Dtype  
--- 
 0   index        115 non-null    int64  
 1   SUBDIVISION  115 non-null    object  
 2   YEAR         115 non-null    int64  
 3   JAN          115 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)
```

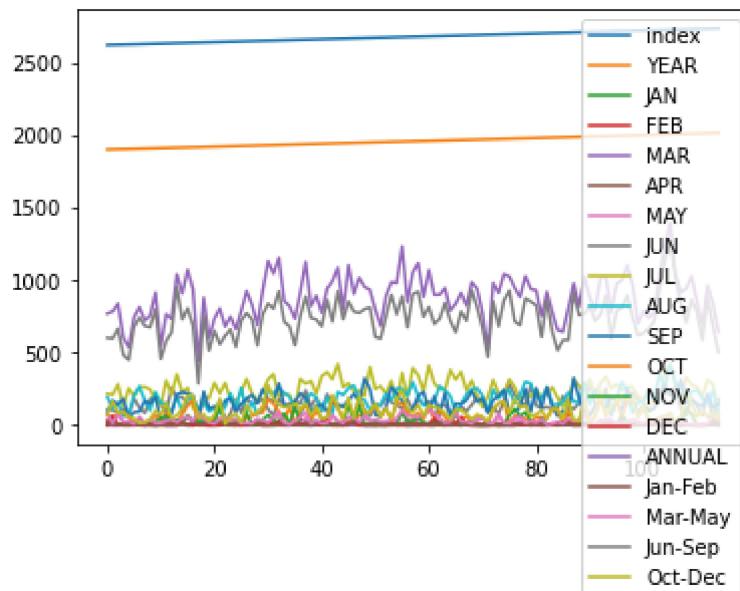
```
Out[11]: array([<AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>], dtype=object)
```



## Line chart

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

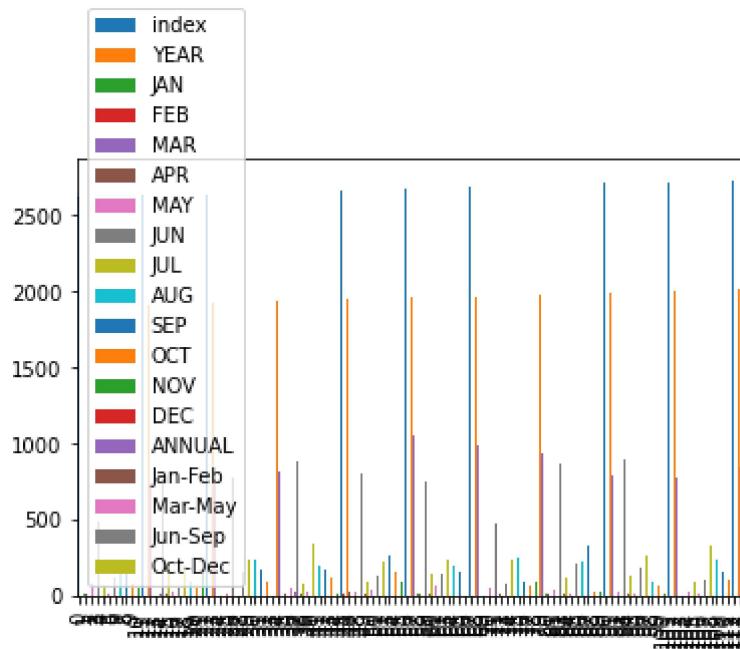
```
Out[12]: <AxesSubplot:>
```



## Bar chart

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

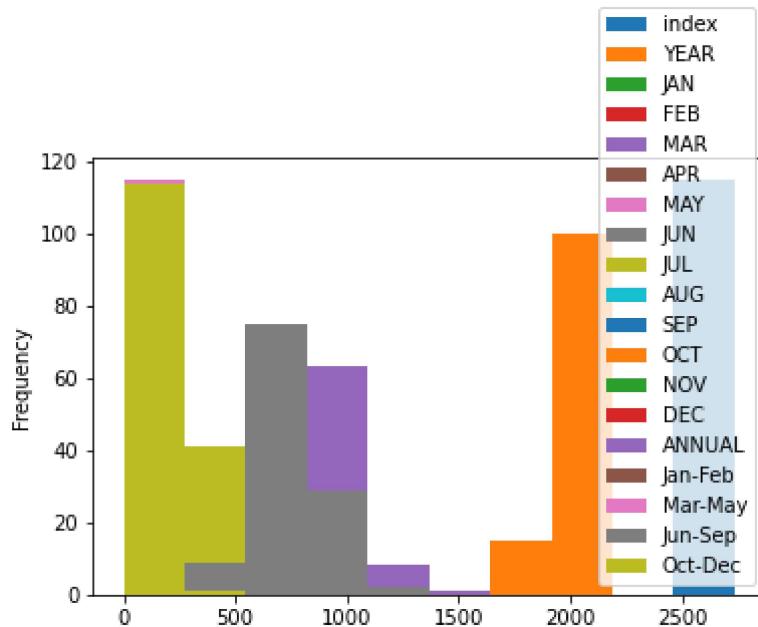
Out[13]: <AxesSubplot:>



## Histogram

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

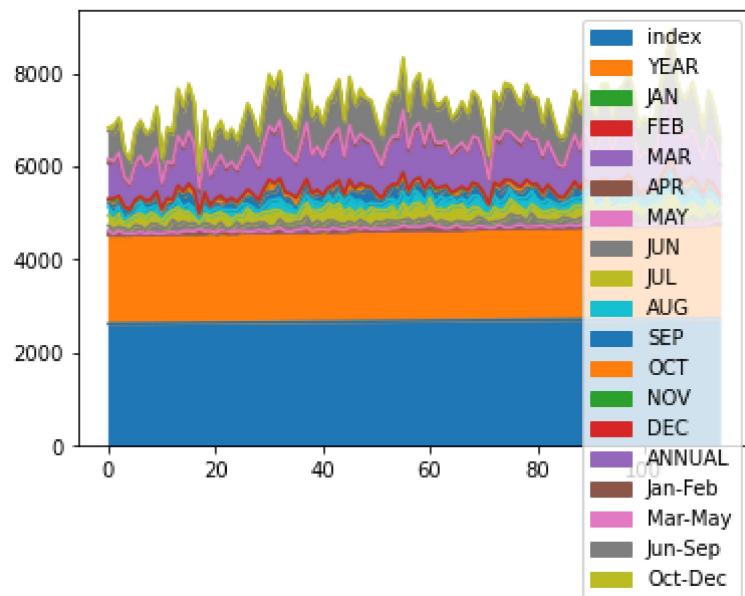
Out[14]: <AxesSubplot:ylabel='Frequency'>



## Area chart

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

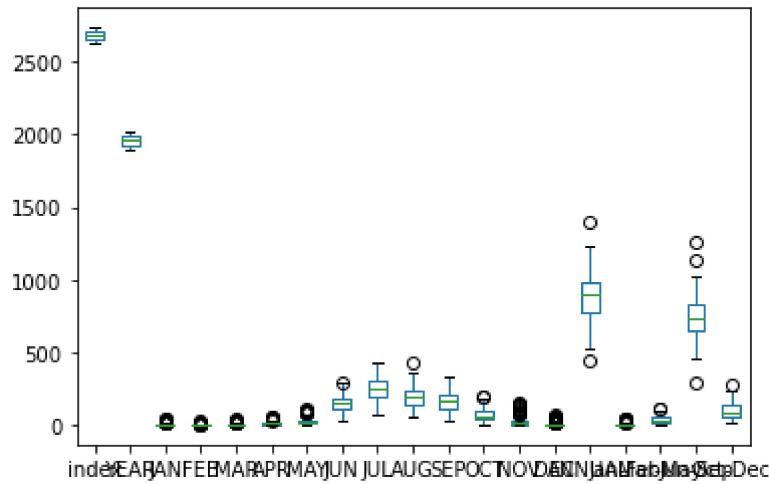
Out[15]: <AxesSubplot:>



## Box chart

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

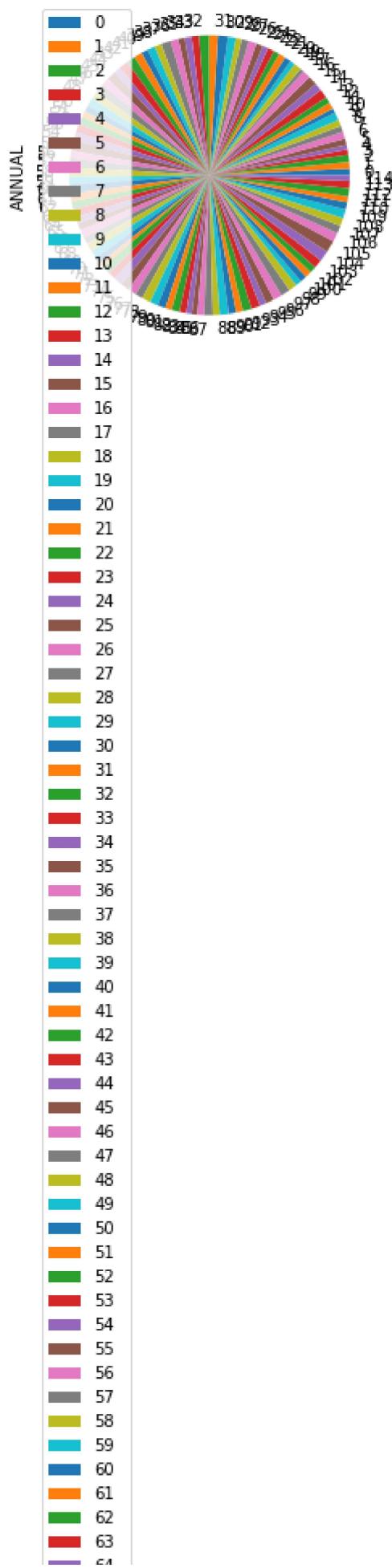
Out[16]: <AxesSubplot:>

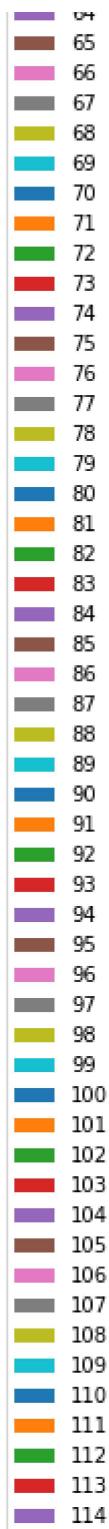


## Pie chart

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

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

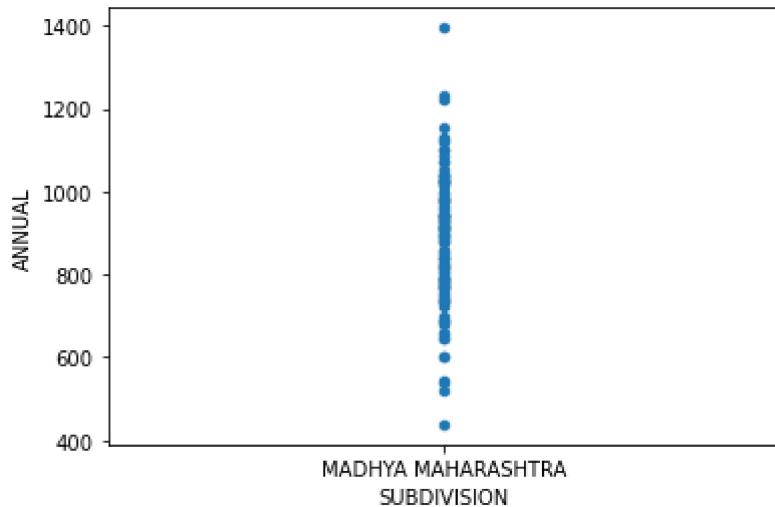




## Scatter chart

```
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]:

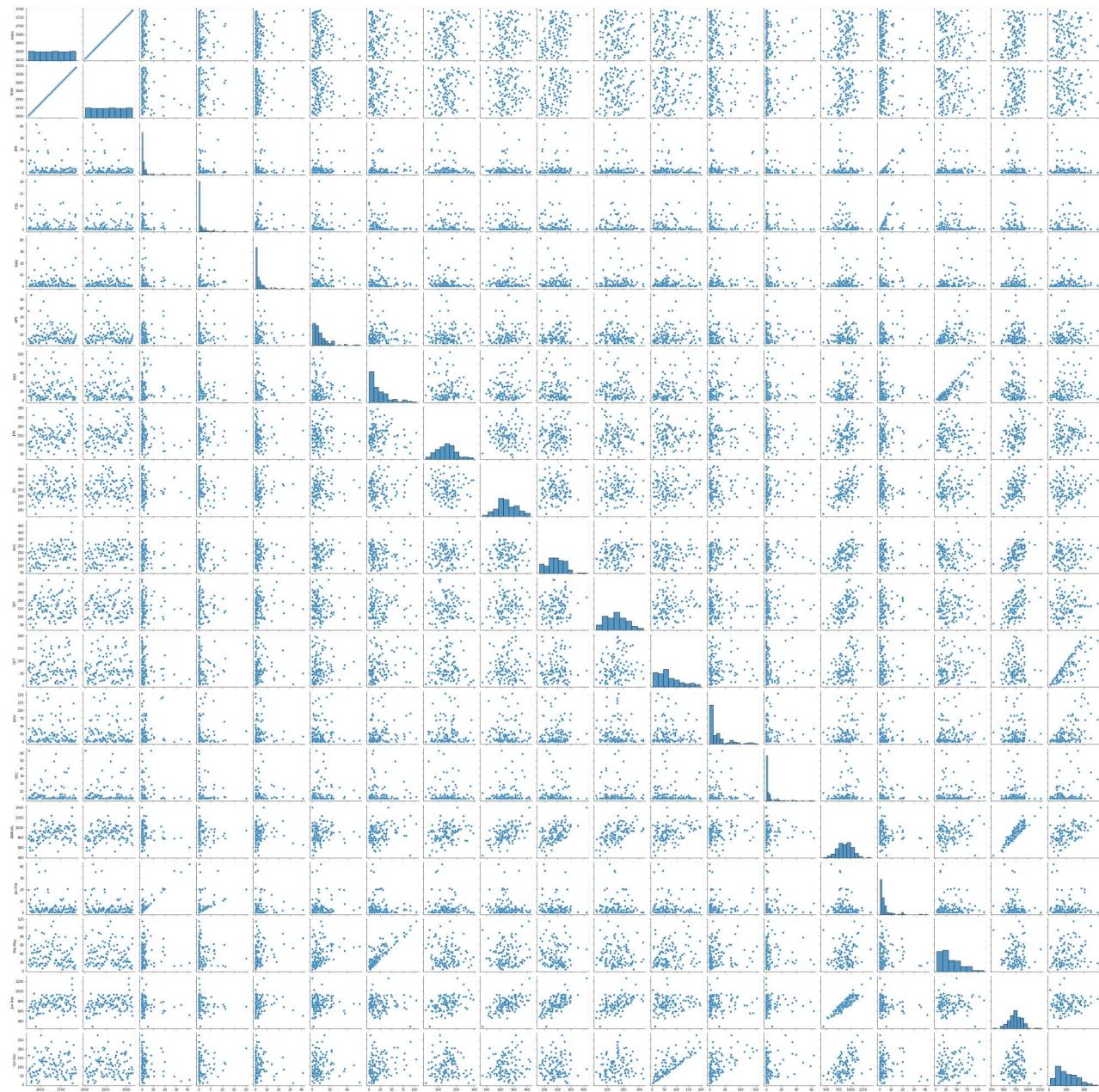
	index	YEAR	JAN	FEB	MAR	APR	MAY	JUN
<b>count</b>	115.000000	115.000000	115.000000	115.000000	115.000000	115.000000	115.000000	115.000000
<b>mean</b>	2679.000000	1958.000000	3.054783	1.467826	3.596522	9.146957	22.943478	147.426087
<b>std</b>	33.341666	33.341666	6.666652	2.915282	6.411625	9.402000	22.368097	55.088562
<b>min</b>	2622.000000	1901.000000	0.000000	0.000000	0.000000	0.000000	0.300000	30.700000
<b>25%</b>	2650.500000	1929.500000	0.000000	0.000000	0.200000	3.200000	7.350000	109.650000

	index	YEAR	JAN	FEB	MAR	APR	MAY	JUN
<b>50%</b>	2679.000000	1958.000000	0.700000	0.200000	1.500000	6.300000	15.200000	146.400000
<b>75%</b>	2707.500000	1986.500000	2.600000	1.550000	4.200000	12.100000	33.000000	179.900000
<b>max</b>	2736.000000	2015.000000	41.500000	20.000000	41.200000	54.500000	104.200000	293.800000

## EDA AND VISUALIZATION

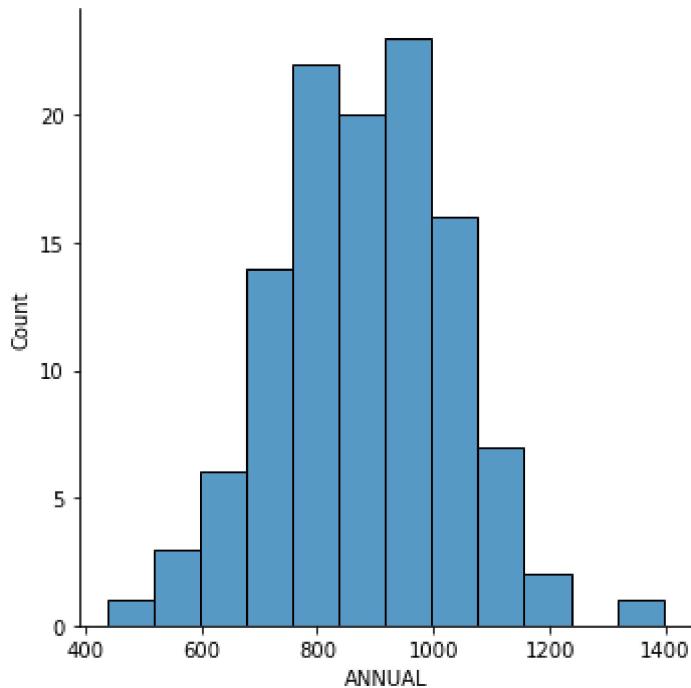
In [21]: `sns.pairplot(df)`

Out[21]: <seaborn.axisgrid.PairGrid at 0x186f198ffa0>



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

Out[22]: &lt;seaborn.axisgrid.FacetGrid at 0x186fafbdb80&gt;

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

Out[23]: &lt;AxesSubplot:&gt;

