

# 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\UTTARAKHAND.csv')  
df
```

Out[2]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	C
0	1242	UTTARAKHAND	1901	134.5	81.4	44.5	5.9	60.8	33.6	381.1	612.3	167.1	1
1	1243	UTTARAKHAND	1902	0.0	17.0	52.2	63.7	52.1	113.1	444.1	327.5	220.4	3
2	1244	UTTARAKHAND	1903	68.0	7.9	87.6	10.3	37.5	83.0	251.6	442.7	249.3	5
3	1245	UTTARAKHAND	1904	40.0	5.2	78.3	13.6	61.1	180.1	449.6	417.2	174.1	
4	1246	UTTARAKHAND	1905	115.4	80.7	99.8	26.1	70.3	111.5	299.9	349.5	129.5	
...	...	...	...	...	...	...	...	...	...	...	...	...	...
110	1352	UTTARAKHAND	2011	30.9	65.2	18.0	30.9	84.2	223.1	433.3	523.7	148.4	
111	1353	UTTARAKHAND	2012	38.8	11.9	28.1	39.2	9.1	46.0	387.1	419.5	220.6	
112	1354	UTTARAKHAND	2013	73.0	188.3	22.0	24.7	18.2	488.9	413.4	359.4	111.3	2
113	1355	UTTARAKHAND	2014	45.9	99.9	68.4	37.6	52.9	62.9	462.7	264.2	107.9	4
114	1356	UTTARAKHAND	2015	54.5	62.6	127.3	57.3	38.0	186.6	337.0	305.3	52.6	1

115 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	C
0	1242	UTTARAKHAND	1901	134.5	81.4	44.5	5.9	60.8	33.6	381.1	612.3	167.1	1
1	1243	UTTARAKHAND	1902	0.0	17.0	52.2	63.7	52.1	113.1	444.1	327.5	220.4	3
2	1244	UTTARAKHAND	1903	68.0	7.9	87.6	10.3	37.5	83.0	251.6	442.7	249.3	5
3	1245	UTTARAKHAND	1904	40.0	5.2	78.3	13.6	61.1	180.1	449.6	417.2	174.1	
4	1246	UTTARAKHAND	1905	115.4	80.7	99.8	26.1	70.3	111.5	299.9	349.5	129.5	
...	...	...	...	...	...	...	...	...	...	...	...	...	...
110	1352	UTTARAKHAND	2011	30.9	65.2	18.0	30.9	84.2	223.1	433.3	523.7	148.4	
111	1353	UTTARAKHAND	2012	38.8	11.9	28.1	39.2	9.1	46.0	387.1	419.5	220.6	
112	1354	UTTARAKHAND	2013	73.0	188.3	22.0	24.7	18.2	488.9	413.4	359.4	111.3	2
113	1355	UTTARAKHAND	2014	45.9	99.9	68.4	37.6	52.9	62.9	462.7	264.2	107.9	4
114	1356	UTTARAKHAND	2015	54.5	62.6	127.3	57.3	38.0	186.6	337.0	305.3	52.6	1

115 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')

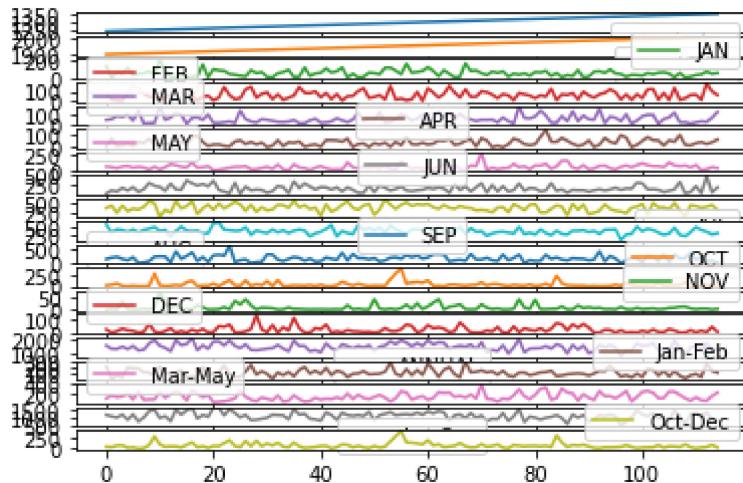
In [5]: df.info()

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 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

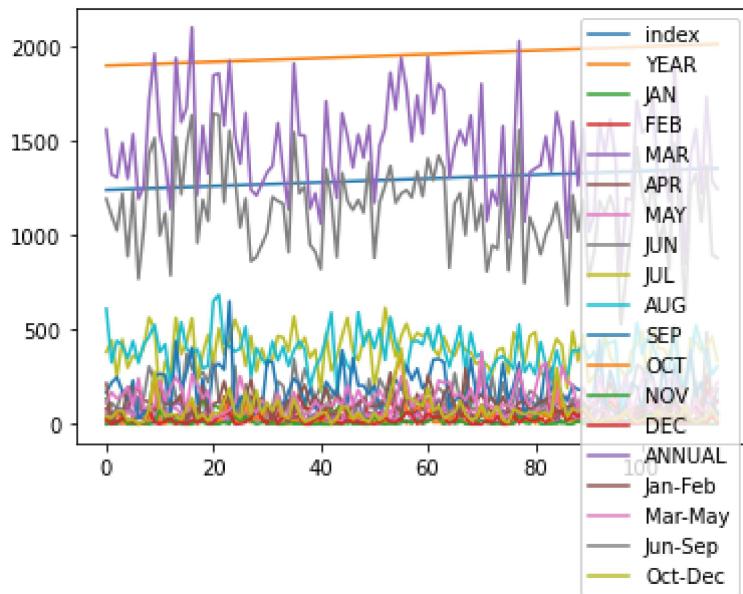
```
In [6]: df.plot.line(subplots=True)
```

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



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

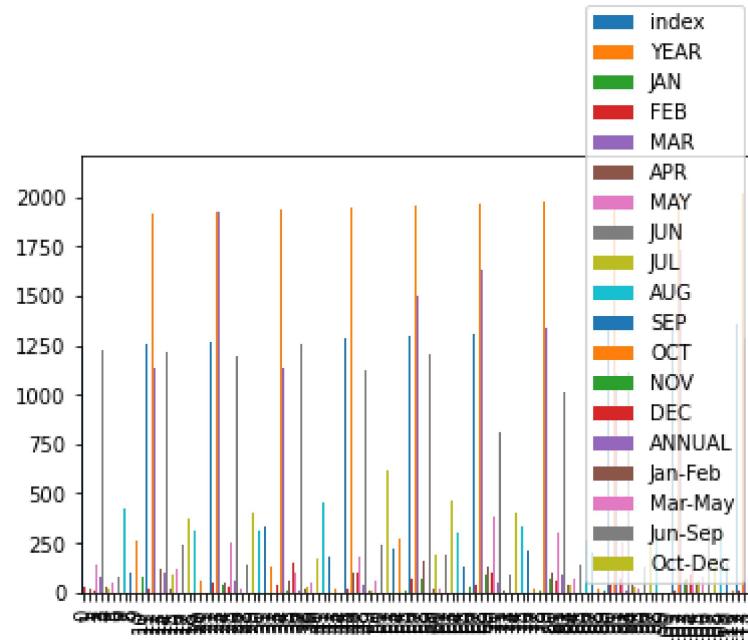
```
Out[7]: <AxesSubplot:>
```



## Bar Chart

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

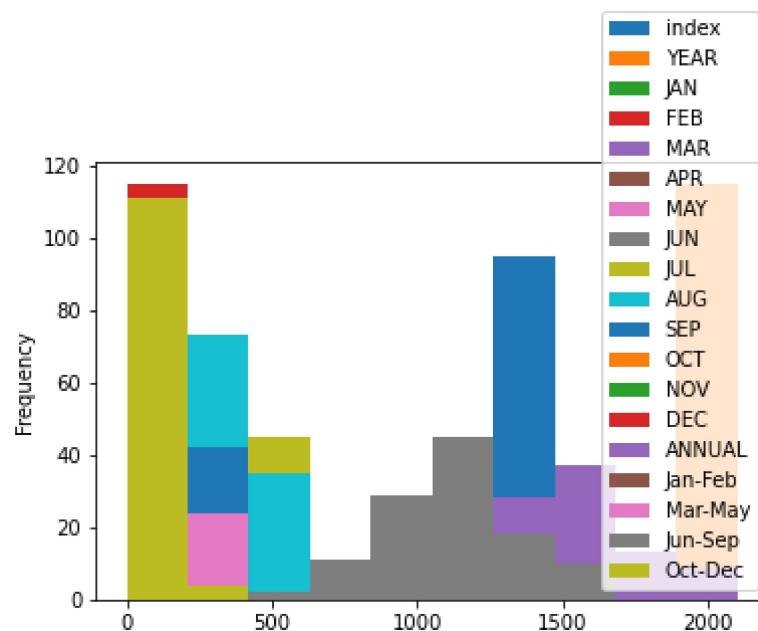
Out[8]: <AxesSubplot:>



## Histogram

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

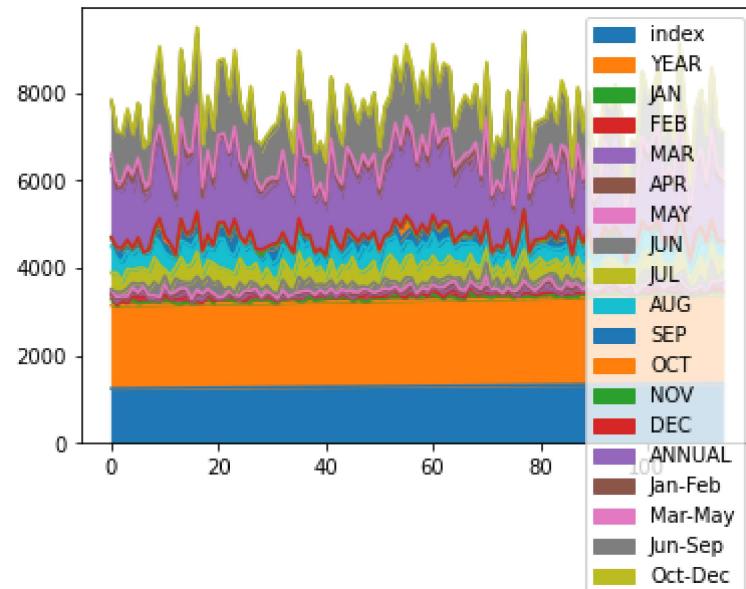
Out[9]: <AxesSubplot:ylabel='Frequency'>



## Area Chart

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

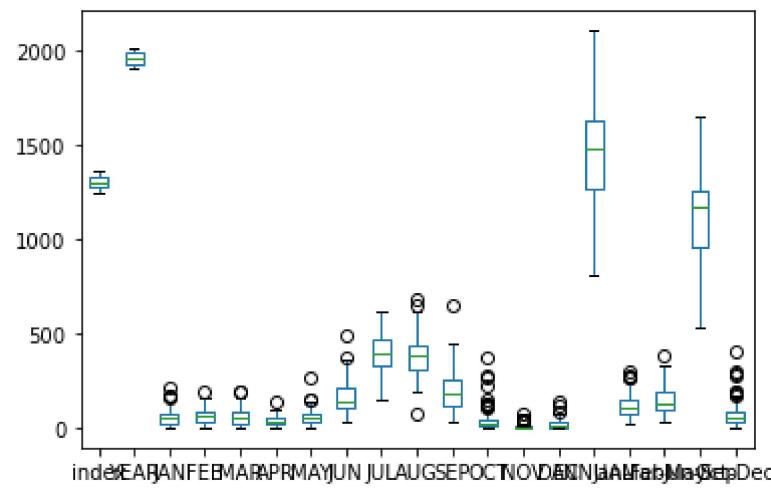
```
Out[10]: <AxesSubplot:>
```



## Box Chart

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

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

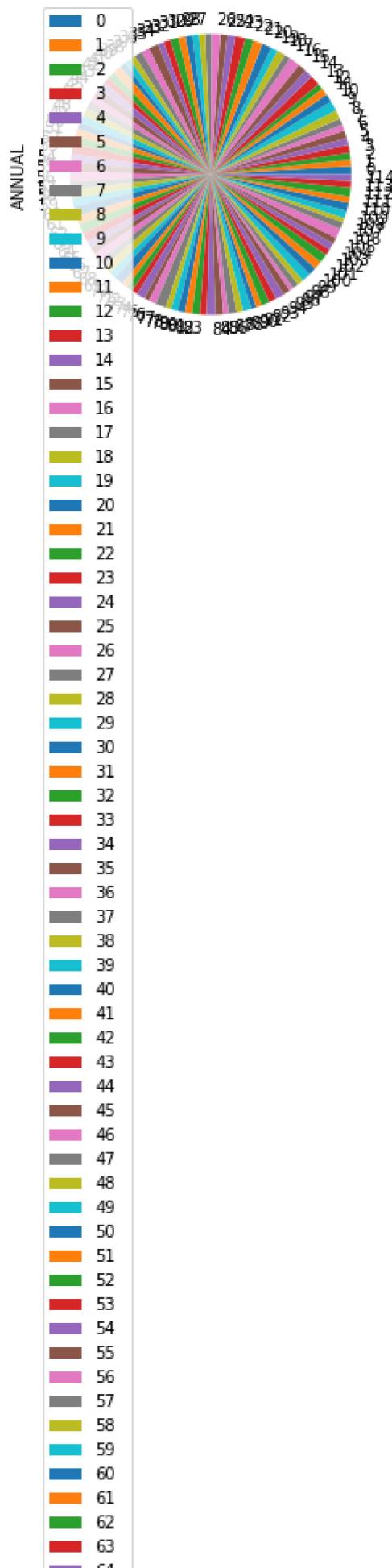


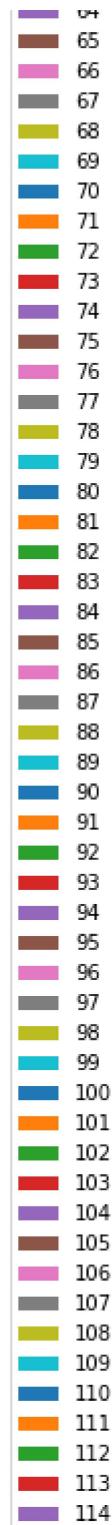
## Pie Chart

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

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



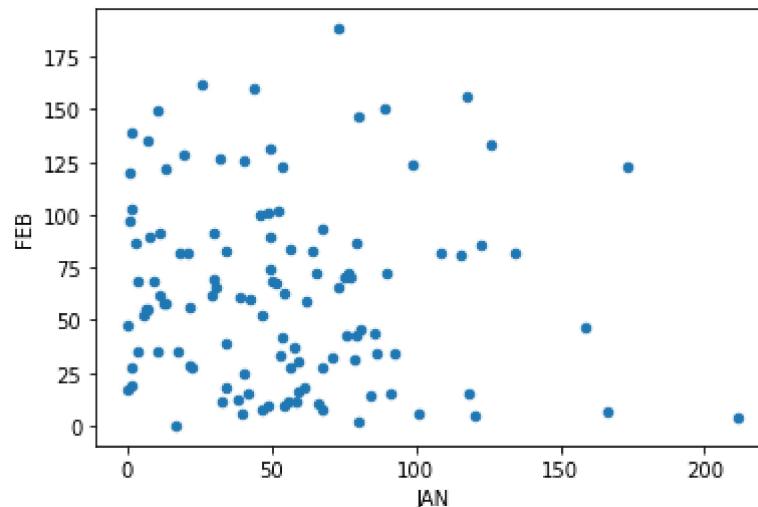




# Scatter Plot

```
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: 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 [15]: df.describe()

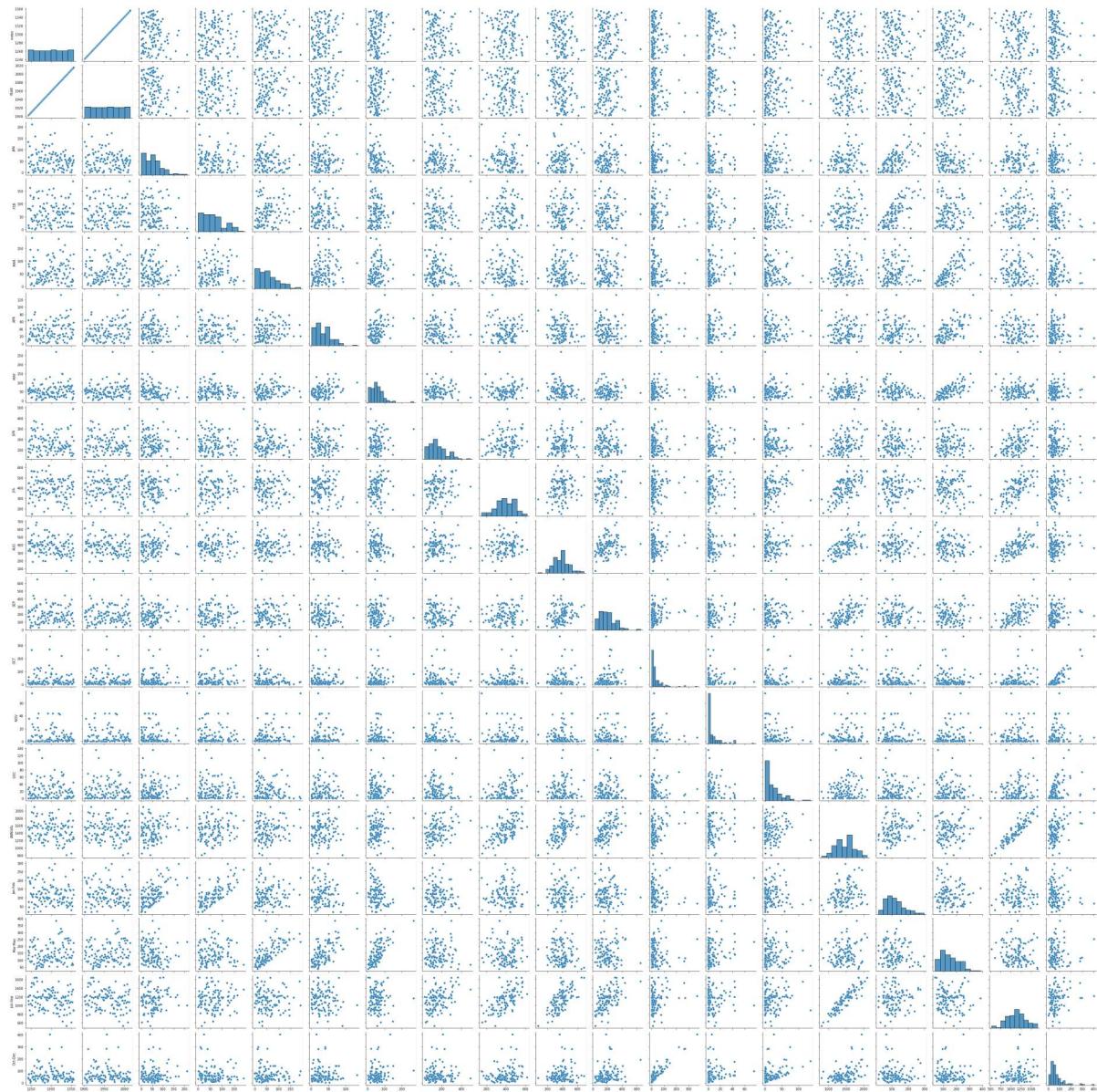
Out[15]:

	index	YEAR	JAN	FEB	MAR	APR	MAY	
<b>count</b>	115.000000	115.000000	115.000000	115.000000	115.000000	115.000000	115.000000	11
<b>mean</b>	1299.000000	1958.000000	53.797391	63.452174	57.272174	35.166087	55.338261	16
<b>std</b>	33.341666	33.341666	40.887384	44.040532	42.438752	24.116540	36.597919	8
<b>min</b>	1242.000000	1901.000000	0.000000	0.000000	0.000000	1.100000	3.600000	3
<b>25%</b>	1270.500000	1929.500000	21.400000	27.950000	22.850000	18.250000	28.050000	10
<b>50%</b>	1299.000000	1958.000000	49.700000	60.100000	47.700000	30.700000	50.500000	13
<b>75%</b>	1327.500000	1986.500000	76.200000	88.100000	80.600000	51.200000	71.450000	21
<b>max</b>	1356.000000	2015.000000	211.400000	188.300000	190.300000	132.900000	270.200000	48

## EDA And Visualization

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

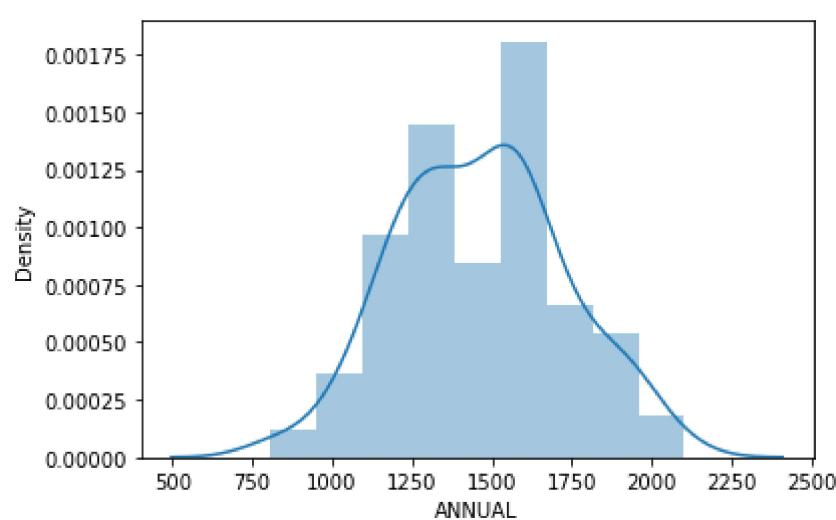
```
Out[16]: <seaborn.axisgrid.PairGrid at 0x21be5bf4640>
```



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:>

