

## 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\KONKAN GOA.csv')
df
```

Out[2]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	O
0	2508	KONKAN & GOA	1902	0.3	0.0	0.0	0.4	7.6	428.2	943.6	515.1	613.8	7
1	2509	KONKAN & GOA	1903	0.0	0.0	0.1	0.0	201.1	470.5	1298.6	673.9	285.1	14
2	2510	KONKAN & GOA	1904	0.0	0.1	6.6	6.3	4.6	975.8	771.7	321.3	217.0	9
3	2511	KONKAN & GOA	1905	0.1	0.1	0.0	0.4	8.6	293.7	770.6	305.5	208.3	8
4	2512	KONKAN & GOA	1906	5.0	0.9	0.0	0.0	2.9	547.4	1090.9	506.7	222.5	3
...	...	...	...	...	...	...	...	...	...	...	...	...	...
109	2617	KONKAN & GOA	2011	0.0	0.0	0.0	3.4	1.1	857.0	1384.1	987.9	468.3	12
110	2618	KONKAN & GOA	2012	0.0	0.0	0.0	0.6	1.1	633.0	928.5	762.5	515.3	17
111	2619	KONKAN & GOA	2013	1.8	5.4	0.1	0.1	18.5	1028.3	1478.5	497.6	340.7	14
112	2620	KONKAN & GOA	2014	1.3	5.3	1.8	0.7	21.3	238.2	1293.2	658.0	419.5	9
113	2621	KONKAN & GOA	2015	2.7	0.0	36.8	3.6	11.3	764.0	526.5	377.3	240.9	9

114 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	O
0	2508	KONKAN & GOA	1902	0.3	0.0	0.0	0.4	7.6	428.2	943.6	515.1	613.8	7
1	2509	KONKAN & GOA	1903	0.0	0.0	0.1	0.0	201.1	470.5	1298.6	673.9	285.1	14
2	2510	KONKAN & GOA	1904	0.0	0.1	6.6	6.3	4.6	975.8	771.7	321.3	217.0	9
3	2511	KONKAN & GOA	1905	0.1	0.1	0.0	0.4	8.6	293.7	770.6	305.5	208.3	8
4	2512	KONKAN & GOA	1906	5.0	0.9	0.0	0.0	2.9	547.4	1090.9	506.7	222.5	3
...	...	...	...	...	...	...	...	...	...	...	...	...	...
109	2617	KONKAN & GOA	2011	0.0	0.0	0.0	3.4	1.1	857.0	1384.1	987.9	468.3	12
110	2618	KONKAN & GOA	2012	0.0	0.0	0.0	0.6	1.1	633.0	928.5	762.5	515.3	17
111	2619	KONKAN & GOA	2013	1.8	5.4	0.1	0.1	18.5	1028.3	1478.5	497.6	340.7	14
112	2620	KONKAN & GOA	2014	1.3	5.3	1.8	0.7	21.3	238.2	1293.2	658.0	419.5	9
113	2621	KONKAN & GOA	2015	2.7	0.0	36.8	3.6	11.3	764.0	526.5	377.3	240.9	9

114 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')

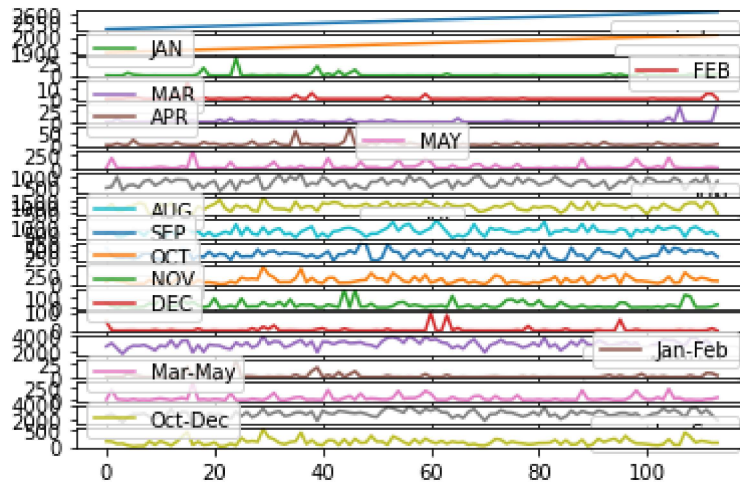
```
In [5]: df.info()
```

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

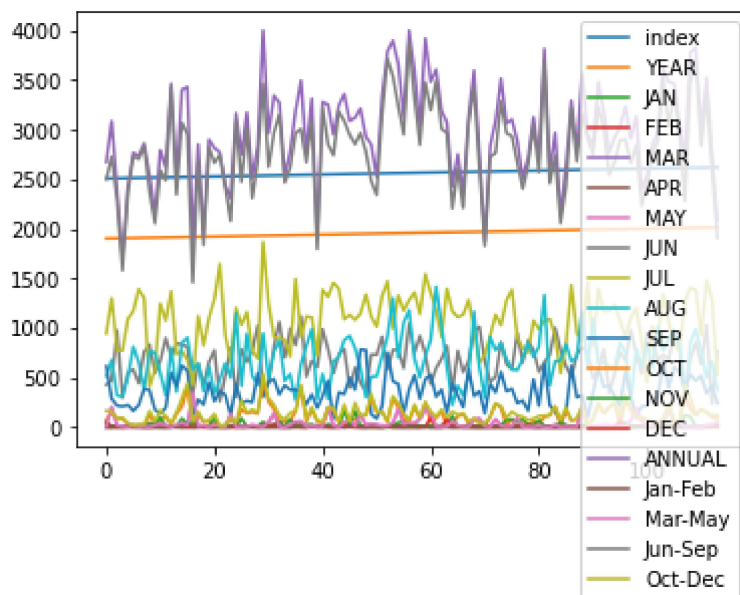
In [6]: `df.plot.line(subplots=True)`

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



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

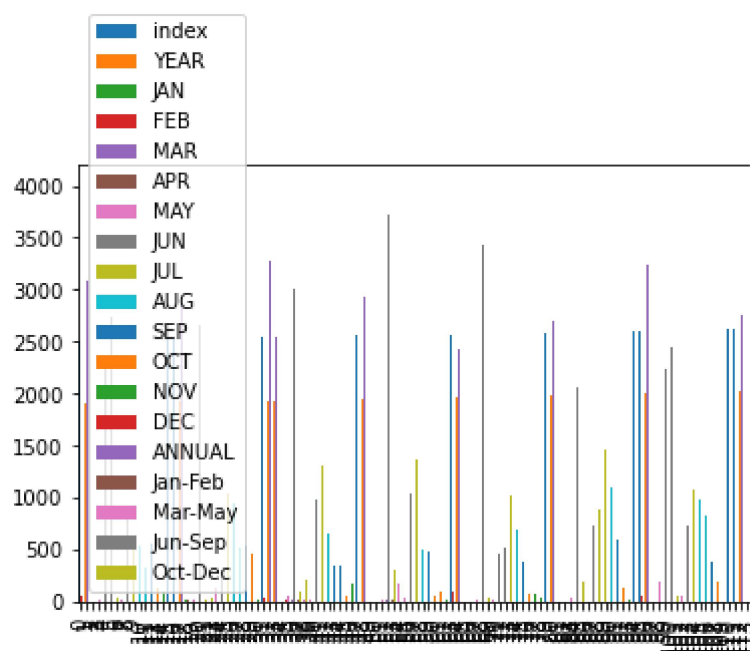
Out[7]: `<AxesSubplot:>`



## Bar Chart

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

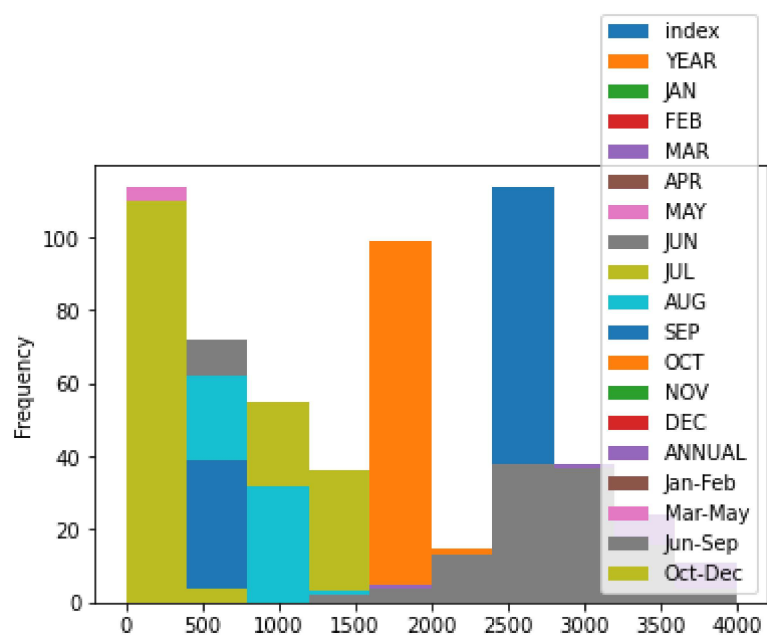
```
Out[8]: <AxesSubplot:>
```



## Histogram

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

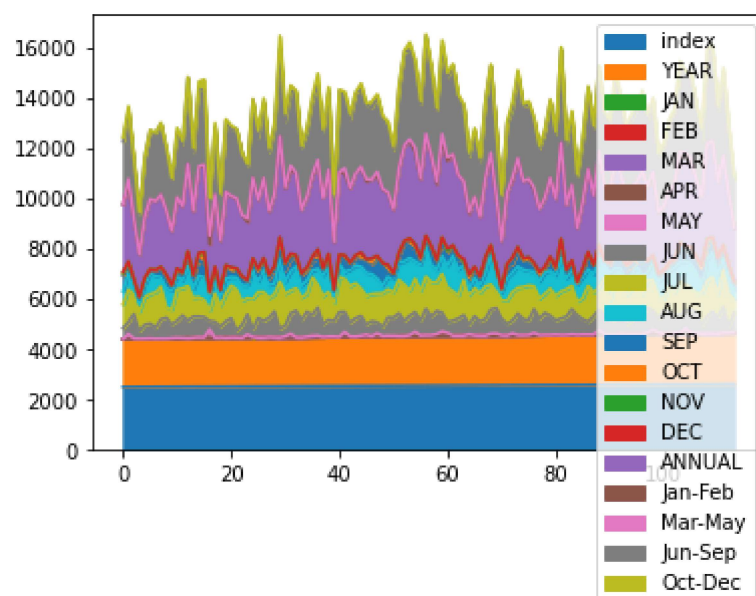
```
Out[9]: <AxesSubplot:ylabel='Frequency'>
```



## Area Chart

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

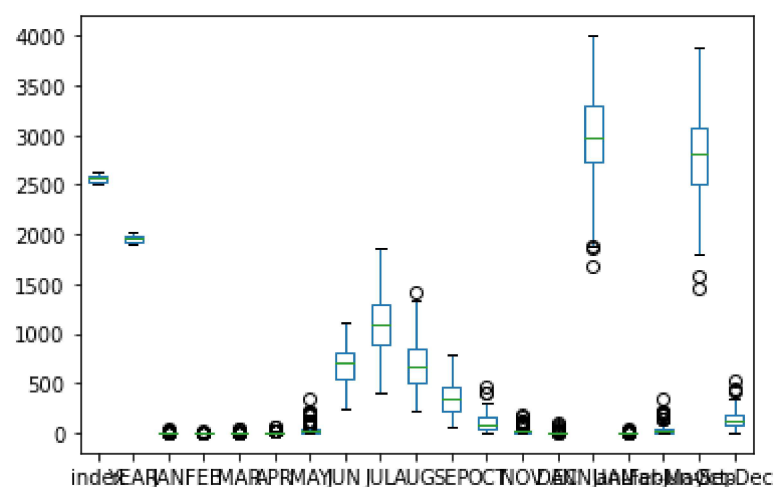
```
Out[10]: <AxesSubplot:>
```



## Box Chart

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

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



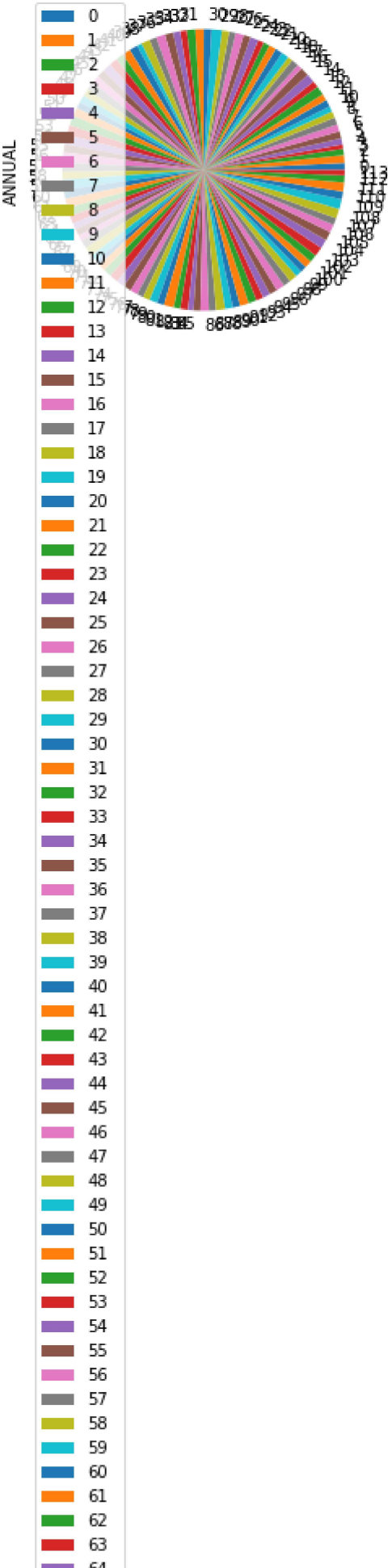
## Pie Chart

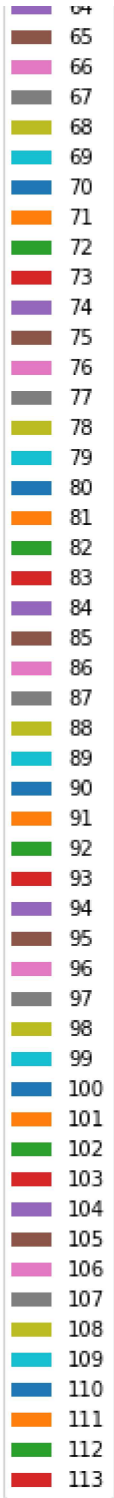
```
In [12]: df.plot.pie(y='ANNUAL')
```

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





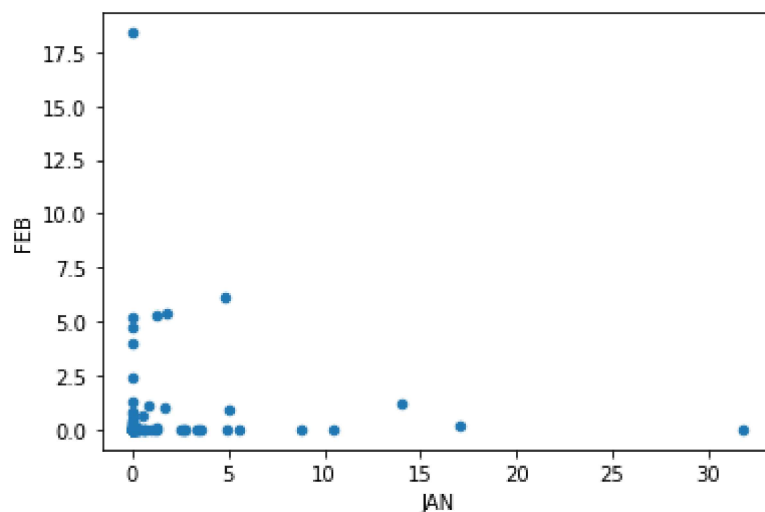




Scatter Plot

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

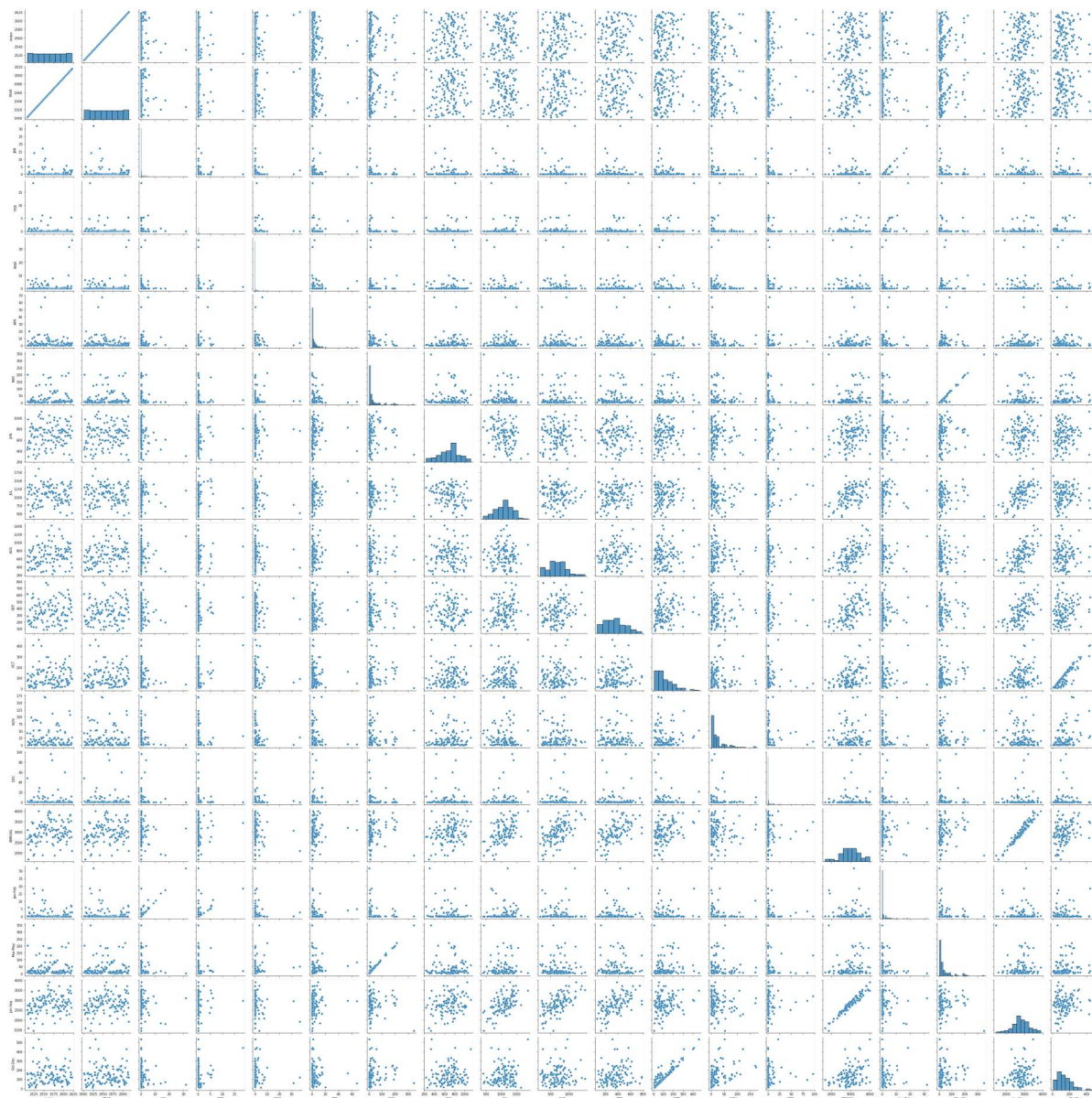
Out[15]:

	index	YEAR	JAN	FEB	MAR	APR	MAY	
count	114.000000	114.000000	114.000000	114.000000	114.000000	114.000000	114.000000	114.000000
mean	2564.500000	1958.500000	1.224561	0.550877	1.383333	3.990351	33.635088	68.000000
std	33.052988	33.052988	3.879799	2.056642	4.769388	8.647522	58.570653	19.000000
min	2508.000000	1902.000000	0.000000	0.000000	0.000000	0.000000	0.000000	25.000000
25%	2536.250000	1930.250000	0.000000	0.000000	0.000000	0.300000	2.900000	54.000000
50%	2564.500000	1958.500000	0.000000	0.000000	0.050000	1.250000	9.650000	70.000000
75%	2592.750000	1986.750000	0.450000	0.100000	0.375000	4.075000	30.450000	80.000000
max	2621.000000	2015.000000	31.800000	18.400000	36.800000	67.300000	345.400000	114.000000

# EDA And Visualization

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

```
Out[16]: <seaborn.axisgrid.PairGrid at 0x1be5ac0ab50>
```

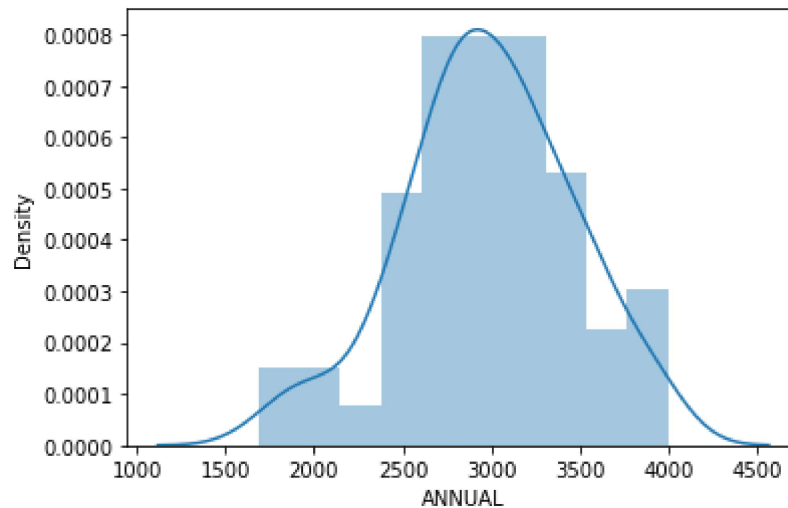


```
In [17]: sns.distplot(df['ANNUAL'])
```

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:>
```

