

Import Libraies

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

In [2]: df=pd.read_csv(r"C:\Users\user\Downloads\FP2_RainFall\rainfall in india 1901-2015.csv")[2392:2506]
df
```

Out[2]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	ANNUAL
2392	2392	SAURASHTRA & KUTCH	1901	1.9	0.0	0.1	0.2	3.2	9.1	87.8	62.5	12.0	3.8	0.0	0.7	181.3
2393	2393	SAURASHTRA & KUTCH	1902	0.1	0.0	0.0	0.5	1.1	14.4	92.9	160.0	123.9	1.5	0.1	6.5	401.1
2394	2394	SAURASHTRA & KUTCH	1903	0.5	0.0	1.7	0.0	3.1	10.5	337.9	96.1	61.9	11.1	0.0	0.0	522.8
2395	2395	SAURASHTRA & KUTCH	1904	1.4	5.8	17.5	0.0	0.0	9.5	111.2	9.4	28.9	0.3	1.7	0.0	185.6
2396	2396	SAURASHTRA & KUTCH	1905	1.5	1.0	0.6	0.4	0.0	6.4	254.5	12.3	12.8	0.4	0.0	0.0	290.0
...
2501	2501	SAURASHTRA & KUTCH	2010	0.1	0.0	0.0	0.0	0.1	71.1	393.8	391.3	187.6	11.6	64.3	0.0	1119.9
2502	2502	SAURASHTRA & KUTCH	2011	0.0	1.4	0.0	0.0	0.0	26.0	212.7	290.9	210.1	1.2	0.1	0.0	742.5
2503	2503	SAURASHTRA & KUTCH	2012	0.0	0.0	0.0	0.2	0.1	22.4	34.7	34.5	228.5	2.4	0.0	1.0	323.8
2504	2504	SAURASHTRA & KUTCH	2013	1.7	0.2	0.1	8.5	0.1	127.7	171.2	83.3	260.2	28.6	0.0	0.0	681.8
2505	2505	SAURASHTRA & KUTCH	2014	0.3	0.0	0.1	0.5	2.1	17.3	137.7	118.8	99.2	5.2	2.7	0.0	383.9

114 rows × 20 columns

Data Cleaning and Preprocessing

In [3]: df.dropna()

Out[3]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	ANNUAL
2392	2392	SAURASHTRA & KUTCH	1901	1.9	0.0	0.1	0.2	3.2	9.1	87.8	62.5	12.0	3.8	0.0	0.7	181.3
2393	2393	SAURASHTRA & KUTCH	1902	0.1	0.0	0.0	0.5	1.1	14.4	92.9	160.0	123.9	1.5	0.1	6.5	401.1
2394	2394	SAURASHTRA & KUTCH	1903	0.5	0.0	1.7	0.0	3.1	10.5	337.9	96.1	61.9	11.1	0.0	0.0	522.8
2395	2395	SAURASHTRA & KUTCH	1904	1.4	5.8	17.5	0.0	0.0	9.5	111.2	9.4	28.9	0.3	1.7	0.0	185.6
2396	2396	SAURASHTRA & KUTCH	1905	1.5	1.0	0.6	0.4	0.0	6.4	254.5	12.3	12.8	0.4	0.0	0.0	290.0
...
2501	2501	SAURASHTRA & KUTCH	2010	0.1	0.0	0.0	0.0	0.1	71.1	393.8	391.3	187.6	11.6	64.3	0.0	1119.9
2502	2502	SAURASHTRA & KUTCH	2011	0.0	1.4	0.0	0.0	0.0	26.0	212.7	290.9	210.1	1.2	0.1	0.0	742.5
2503	2503	SAURASHTRA & KUTCH	2012	0.0	0.0	0.0	0.2	0.1	22.4	34.7	34.5	228.5	2.4	0.0	1.0	323.8
2504	2504	SAURASHTRA & KUTCH	2013	1.7	0.2	0.1	8.5	0.1	127.7	171.2	83.3	260.2	28.6	0.0	0.0	681.8
2505	2505	SAURASHTRA & KUTCH	2014	0.3	0.0	0.1	0.5	2.1	17.3	137.7	118.8	99.2	5.2	2.7	0.0	383.9

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()

```

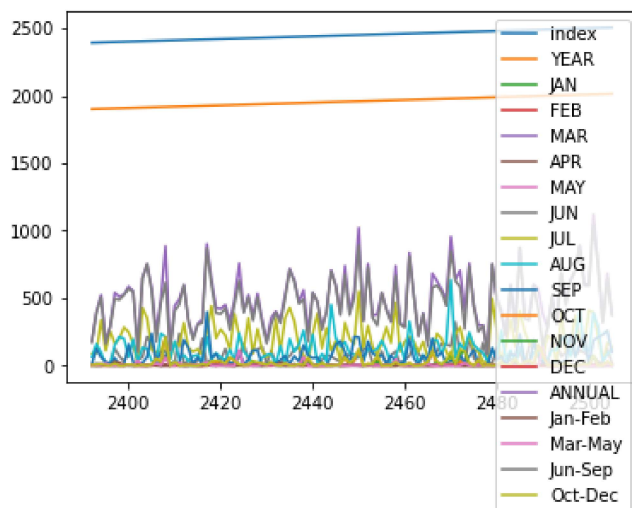
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 114 entries, 2392 to 2505
Data columns (total 20 columns):
#   Column          Non-Null Count  Dtype
---  -
0   index           114 non-null    int64
1   SUBDIVISION     114 non-null    object
2   YEAR            114 non-null    int64
3   JAN             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: 17.9+ KB

```

Line chart

In [6]: df.plot.line()

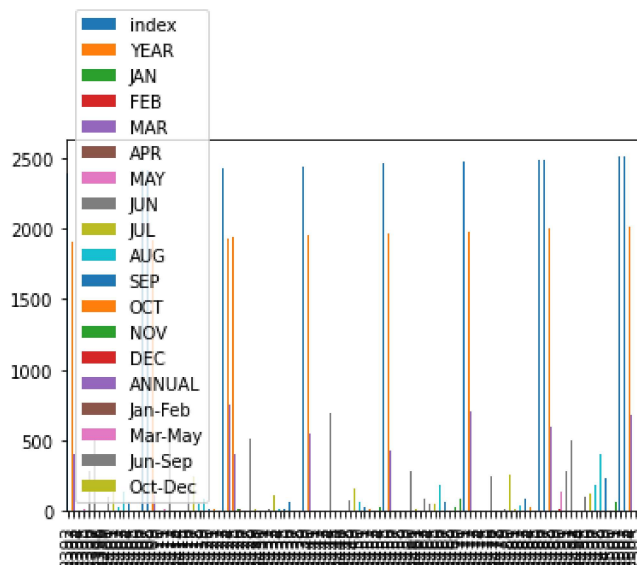
Out[6]: <AxesSubplot:>



Bar chart

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

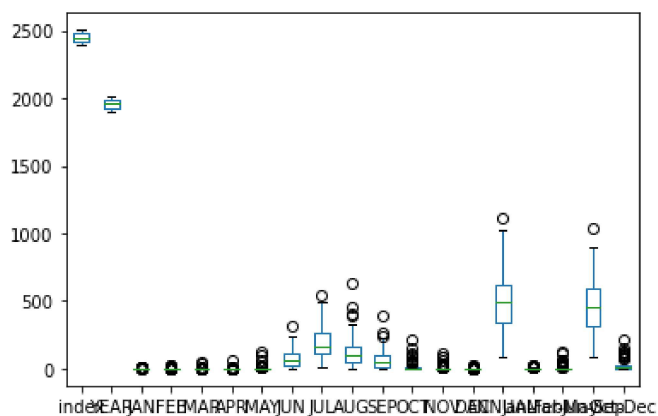
```
Out[7]: <AxesSubplot:>
```



Box chart

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

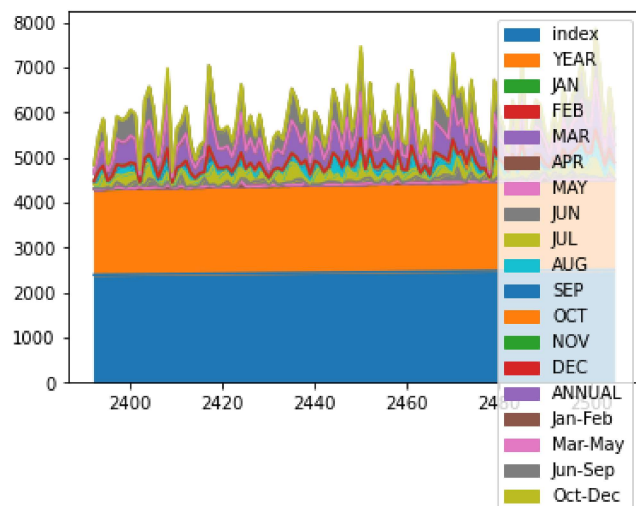
```
Out[8]: <AxesSubplot:>
```



Area chart

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

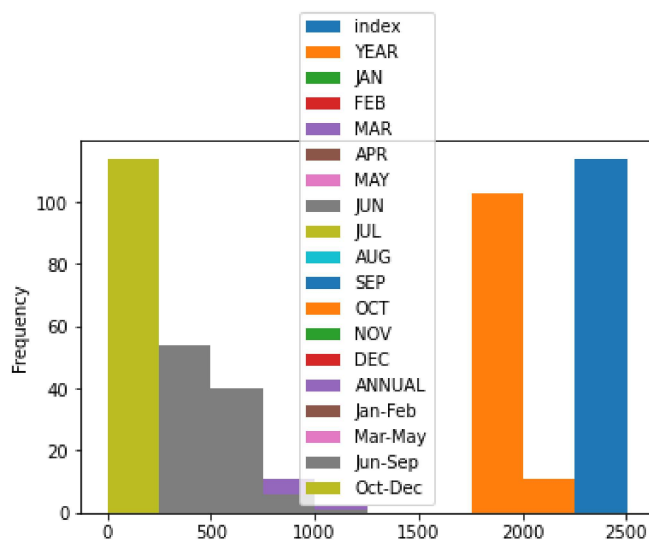
```
Out[9]: <AxesSubplot:>
```



Histogram

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

```
Out[10]: <AxesSubplot:ylabel='Frequency'>
```

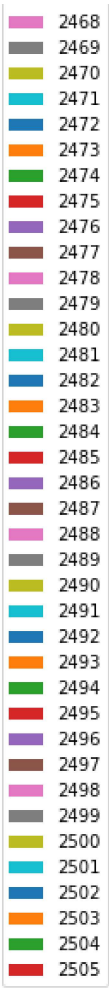


pie chart

```
In [11]: df.plot.pie(y="ANNUAL")
```

```
Out[11]: <AxesSubplot:ylabel='ANNUAL'>
```

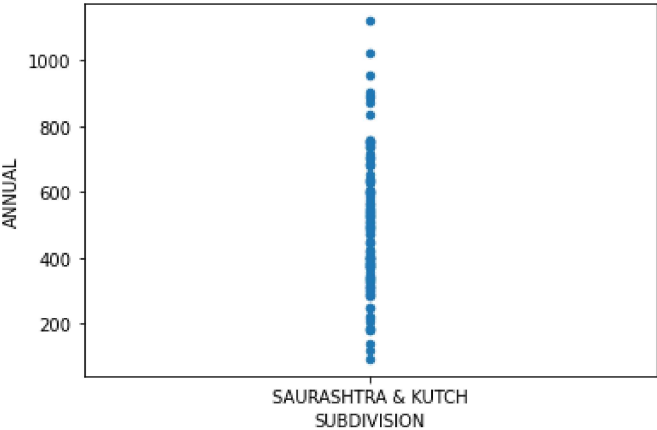


Scatter chart

```
In [12]: df.plot.scatter(y='ANNUAL',x='SUBDIVISION')
```

Out[12]: <AxesSubplot:xlabel='SUBDIVISION', ylabel='ANNUAL'>



```
In [13]: df.describe()
```

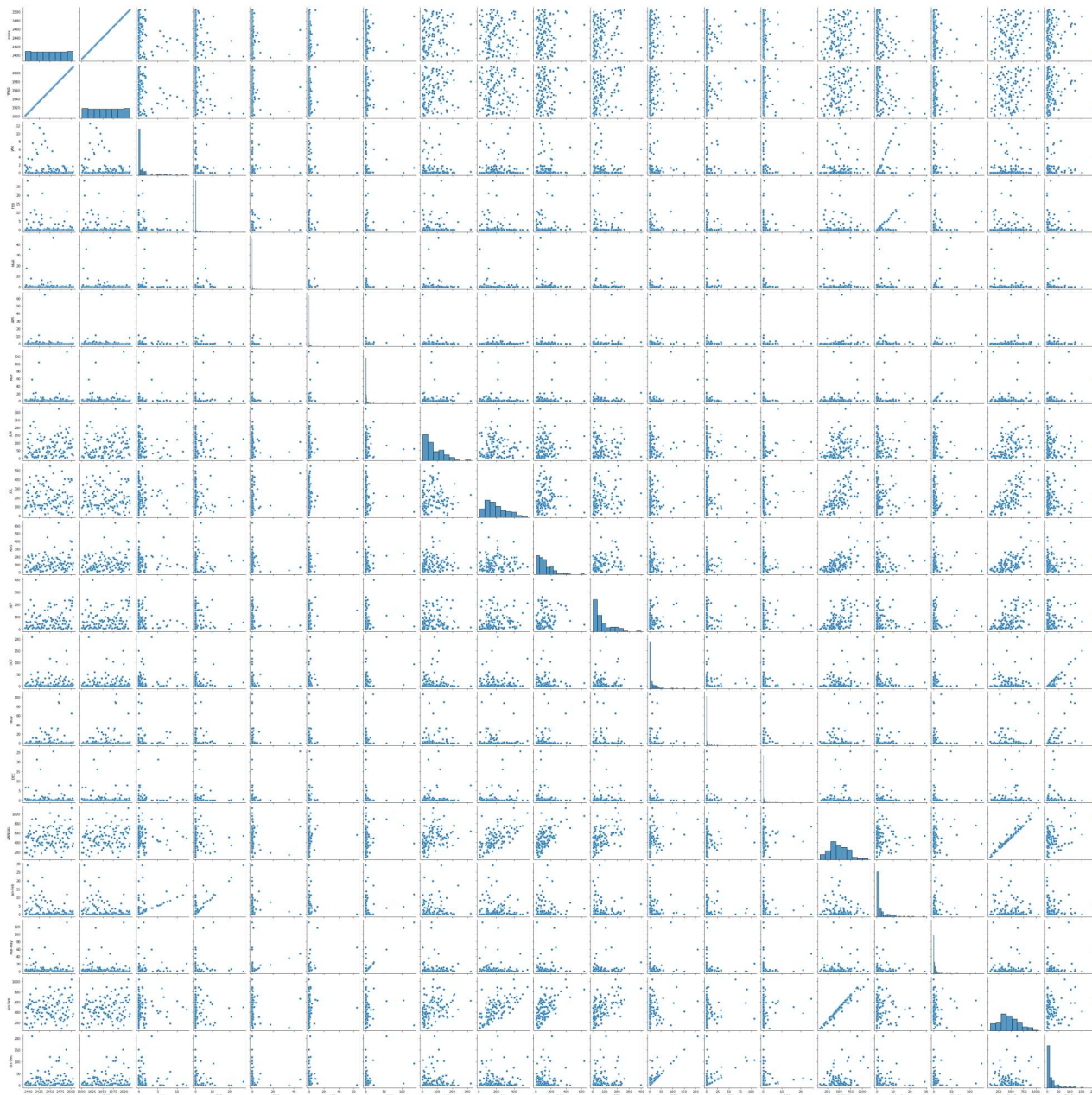
Out[13]:

	index	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	
count	114.000000	114.000000	114.000000	114.000000	114.000000	114.000000	114.000000	114.000000	114.000000	114.000000
mean	2448.500000	1957.500000	1.141228	1.629825	1.269298	1.175439	4.696491	74.035965	194.692105	194.692105
std	33.052988	33.052988	2.385087	4.286714	5.709147	6.185432	16.656466	63.238094	120.227961	120.227961
min	2392.000000	1901.000000	0.000000	0.000000	0.000000	0.000000	0.000000	1.200000	8.000000	8.000000
25%	2420.250000	1929.250000	0.000000	0.000000	0.000000	0.000000	0.000000	18.225000	107.075000	107.075000
50%	2448.500000	1957.500000	0.150000	0.000000	0.000000	0.000000	0.500000	61.350000	167.850000	167.850000
75%	2476.750000	1985.750000	1.000000	0.575000	0.375000	0.500000	2.700000	114.325000	266.075000	266.075000
max	2505.000000	2014.000000	12.500000	28.200000	46.200000	64.400000	131.900000	321.800000	549.100000	549.100000

EDA AND VISUALIZATION

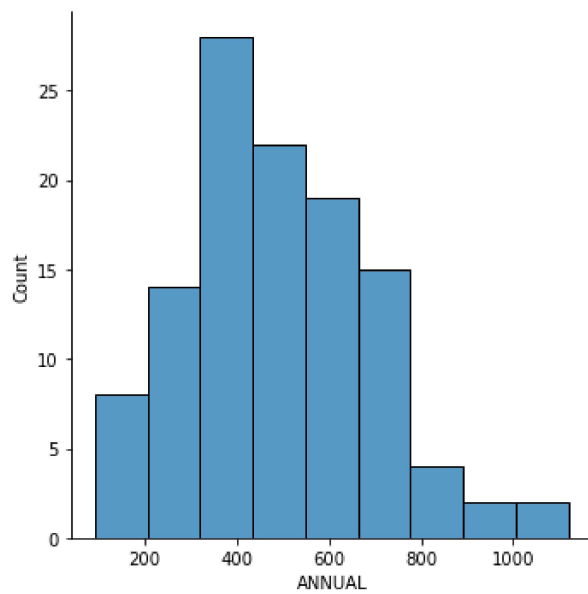
```
In [14]: sns.pairplot(df)
```

```
Out[14]: <seaborn.axisgrid.PairGrid at 0x20179cbcac0>
```



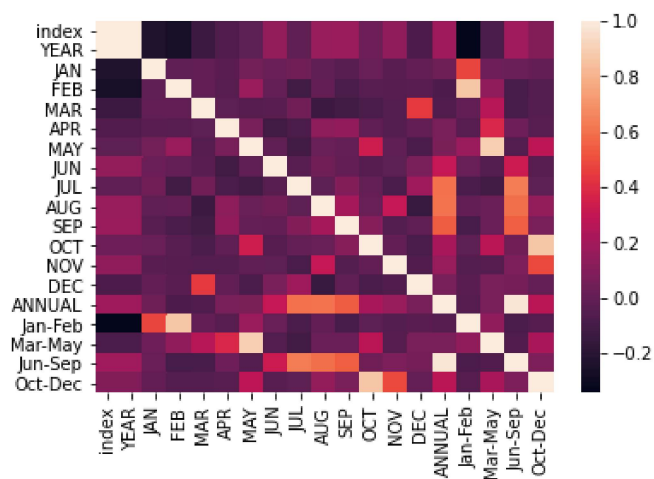
```
In [15]: sns.displot(df['ANNUAL'])
```

```
Out[15]: <seaborn.axisgrid.FacetGrid at 0x20108fb1a30>
```



```
In [16]: sns.heatmap(df.corr())
```

```
Out[16]: <AxesSubplot:>
```



```
In [ ]:
```