

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

```
In [2]: data=pd.read_csv(r"C:\Users\user\Desktop\vicky\rainfall\rainfall in india 1901-2015.csv")[2968:3081]
```

```
In [3]: data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 114 entries, 2968 to 3081
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
```

```
In [4]: data.head()
```

Out[4]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	ANNUAL
2968	2968	CHHATTISGARH	1902	0.6	6.5	0.4	13.9	10.3	37.2	403.8	236.6	198.1	4.7	8.1	3.7	923.6
2969	2969	CHHATTISGARH	1903	6.2	13.9	0.4	6.8	51.1	110.7	365.9	396.0	212.0	168.0	0.1	0.0	1331.6
2970	2970	CHHATTISGARH	1904	0.0	8.6	32.3	0.2	77.5	369.5	303.6	483.6	86.8	129.3	1.0	0.0	1492.6
2971	2971	CHHATTISGARH	1905	50.3	22.6	19.0	24.6	31.8	40.4	443.7	270.8	338.8	8.9	0.0	0.0	1251.6
2972	2972	CHHATTISGARH	1906	25.0	91.0	52.5	0.0	4.1	210.1	445.2	258.3	242.3	41.2	10.2	14.7	1394.6

In [5]:

data.tail()

Out[5]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	ANNUAL
3077	3077	CHHATTISGARH	2011	0.3	11.5	2.6	35.0	16.8	183.5	272.6	379.8	382.2	15.5	0.0	2.8	1302.7
3078	3078	CHHATTISGARH	2012	36.6	4.8	1.1	14.9	9.4	147.3	430.6	442.2	245.3	19.8	20.4	5.0	1377.4
3079	3079	CHHATTISGARH	2013	2.8	19.7	4.9	45.8	5.7	263.6	418.8	336.6	140.9	180.9	0.3	0.0	1420.0
3080	3080	CHHATTISGARH	2014	2.3	29.0	21.4	17.3	25.0	104.9	416.7	327.7	252.7	77.9	2.6	1.1	1278.7
3081	3081	CHHATTISGARH	2015	15.8	1.2	21.2	37.0	13.0	257.6	248.6	286.6	216.9	17.7	0.6	1.5	1117.0

In [6]:

data.shape

Out[6]:

(114, 20)

In [7]:

new_data=data.fillna(value=1)
new_data

Out[7]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	ANNUAL
2968	2968	CHHATTISGARH	1902	0.6	6.5	0.4	13.9	10.3	37.2	403.8	236.6	198.1	4.7	8.1	3.7	923.0
2969	2969	CHHATTISGARH	1903	6.2	13.9	0.4	6.8	51.1	110.7	365.9	396.0	212.0	168.0	0.1	0.0	1331.0
2970	2970	CHHATTISGARH	1904	0.0	8.6	32.3	0.2	77.5	369.5	303.6	483.6	86.8	129.3	1.0	0.0	1492.4
2971	2971	CHHATTISGARH	1905	50.3	22.6	19.0	24.6	31.8	40.4	443.7	270.8	338.8	8.9	0.0	0.0	1251.7
2972	2972	CHHATTISGARH	1906	25.0	91.0	52.5	0.0	4.1	210.1	445.2	258.3	242.3	41.2	10.2	14.7	1394.0
...
3077	3077	CHHATTISGARH	2011	0.3	11.5	2.6	35.0	16.8	183.5	272.6	379.8	382.2	15.5	0.0	2.8	1302.7
3078	3078	CHHATTISGARH	2012	36.6	4.8	1.1	14.9	9.4	147.3	430.6	442.2	245.3	19.8	20.4	5.0	1377.4
3079	3079	CHHATTISGARH	2013	2.8	19.7	4.9	45.8	5.7	263.6	418.8	336.6	140.9	180.9	0.3	0.0	1420.0
3080	3080	CHHATTISGARH	2014	2.3	29.0	21.4	17.3	25.0	104.9	416.7	327.7	252.7	77.9	2.6	1.1	1278.7
3081	3081	CHHATTISGARH	2015	15.8	1.2	21.2	37.0	13.0	257.6	248.6	286.6	216.9	17.7	0.6	1.5	1117.0

114 rows × 20 columns

In [8]:

new_data.index

Out[8]:

RangeIndex(start=2968, stop=3082, step=1)

In [9]:

new_data.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')

```
In [10]: new_data.plot.line()
```

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

```
In [11]: new_data.plot.bar()
```

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

```
In [12]: new_data.plot.area()
```

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

```
In [13]: new_data.plot.hist()
```

```
Out[13]: <AxesSubplot:ylabel='Frequency'>
```

```
In [14]: new_data.plot.pie(y='ANNUAL')
```

```
Out[14]: <AxesSubplot:ylabel='ANNUAL'>
```



```
In [15]: new_data.plot.scatter(x='YEAR',y='ANNUAL')
```

```
Out[15]: <AxesSubplot:xlabel='YEAR', ylabel='ANNUAL'>
```



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

```
Out[16]: <seaborn.axisgrid.PairGrid at 0x23e0c726b80>
```

```
In [17]: sns.distplot(data['YEAR'])
```

```
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='YEAR', ylabel='Density'>
```

```
In [18]: sns.heatmap(new_data.corr())
```

```
Out[18]: <AxesSubplot:>
```