

```
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")[3198:3311]
```

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

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 114 entries, 3198 to 3311
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
3198	3198	TELANGANA	1902	0.0	0.0	0.2	10.7	7.3	52.4	146.3	142.8	190.5	41.7	31.2	7.3	630.4
3199	3199	TELANGANA	1903	12.9	4.6	0.0	9.9	40.7	99.2	505.2	246.7	191.9	155.8	15.5	1.1	1283.4
3200	3200	TELANGANA	1904	0.0	0.0	10.8	0.8	14.7	104.2	139.5	50.0	162.3	44.4	0.0	0.0	526.7
3201	3201	TELANGANA	1905	0.0	4.3	12.8	27.6	32.2	129.5	82.4	237.3	179.1	19.6	0.0	0.0	724.9
3202	3202	TELANGANA	1906	22.5	1.2	13.4	2.4	0.7	211.1	210.8	226.7	96.3	20.5	14.9	34.8	855.2

In [5]: data.tail()

Out[5]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	ANNUAL
3307	3307	TELANGANA	2011	0.0	11.9	2.6	25.6	9.3	83.9	268.2	225.9	107.6	13.9	4.2	0.0	753.1
3308	3308	TELANGANA	2012	6.7	0.0	0.2	14.0	8.4	124.4	300.3	229.9	202.4	83.6	38.7	0.0	1008.6
3309	3309	TELANGANA	2013	2.4	29.0	0.2	24.4	8.5	213.4	453.8	230.6	161.4	205.9	16.4	2.7	1348.7
3310	3310	TELANGANA	2014	0.2	2.9	58.3	10.3	73.3	62.3	146.0	205.2	146.8	29.6	10.8	0.7	746.4
3311	3311	TELANGANA	2015	17.5	0.0	43.0	65.7	23.3	266.9	104.4	160.5	158.3	15.6	0.3	1.7	857.3

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
3198	3198	TELANGANA	1902	0.0	0.0	0.2	10.7	7.3	52.4	146.3	142.8	190.5	41.7	31.2	7.3	630.4
3199	3199	TELANGANA	1903	12.9	4.6	0.0	9.9	40.7	99.2	505.2	246.7	191.9	155.8	15.5	1.1	1283.4
3200	3200	TELANGANA	1904	0.0	0.0	10.8	0.8	14.7	104.2	139.5	50.0	162.3	44.4	0.0	0.0	526.7
3201	3201	TELANGANA	1905	0.0	4.3	12.8	27.6	32.2	129.5	82.4	237.3	179.1	19.6	0.0	0.0	724.9
3202	3202	TELANGANA	1906	22.5	1.2	13.4	2.4	0.7	211.1	210.8	226.7	96.3	20.5	14.9	34.8	855.2
...
3307	3307	TELANGANA	2011	0.0	11.9	2.6	25.6	9.3	83.9	268.2	225.9	107.6	13.9	4.2	0.0	753.1
3308	3308	TELANGANA	2012	6.7	0.0	0.2	14.0	8.4	124.4	300.3	229.9	202.4	83.6	38.7	0.0	1008.6
3309	3309	TELANGANA	2013	2.4	29.0	0.2	24.4	8.5	213.4	453.8	230.6	161.4	205.9	16.4	2.7	1348.7
3310	3310	TELANGANA	2014	0.2	2.9	58.3	10.3	73.3	62.3	146.0	205.2	146.8	29.6	10.8	0.7	746.4
3311	3311	TELANGANA	2015	17.5	0.0	43.0	65.7	23.3	266.9	104.4	160.5	158.3	15.6	0.3	1.7	857.3

114 rows × 20 columns

In [8]: new_data.index

Out[8]: RangeIndex(start=3198, stop=3312, 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 0x1f19ffb8ca0>
```

```
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:>
```