

```
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")[2739:2852]
```

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

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 114 entries, 2739 to 2852
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
2739	2739	MATATHWADA	1903	2.6	0.8	0.0	1.7	58.3	104.4	264.2	281.9	173.3	139.9	0.3	5.3	1032.6
2740	2740	MATATHWADA	1904	0.0	0.9	12.1	0.3	7.2	79.2	118.4	57.3	339.0	76.2	0.0	0.0	690.5
2741	2741	MATATHWADA	1905	1.3	2.0	0.0	6.6	4.8	84.6	94.8	137.6	157.8	15.4	0.9	0.0	505.9
2742	2742	MATATHWADA	1906	19.8	0.0	0.1	0.0	0.2	220.6	254.9	156.9	82.1	19.8	3.9	12.4	770.7
2743	2743	MATATHWADA	1907	1.0	6.6	1.1	61.3	0.0	90.4	198.9	155.2	34.7	5.0	0.0	3.7	557.9

```
In [5]: data.shape
```

Out[5]: (114, 20)

```
In [6]: new_data=data.fillna(value=1)
new_data
```

Out[6]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	ANNUAL
2739	2739	MATATHWADA	1903	2.6	0.8	0.0	1.7	58.3	104.4	264.2	281.9	173.3	139.9	0.3	5.3	1032.6
2740	2740	MATATHWADA	1904	0.0	0.9	12.1	0.3	7.2	79.2	118.4	57.3	339.0	76.2	0.0	0.0	690.5
2741	2741	MATATHWADA	1905	1.3	2.0	0.0	6.6	4.8	84.6	94.8	137.6	157.8	15.4	0.9	0.0	505.9
2742	2742	MATATHWADA	1906	19.8	0.0	0.1	0.0	0.2	220.6	254.9	156.9	82.1	19.8	3.9	12.4	770.7
2743	2743	MATATHWADA	1907	1.0	6.6	1.1	61.3	0.0	90.4	198.9	155.2	34.7	5.0	0.0	3.7	557.9
...
2848	2848	MATATHWADA	2012	0.0	0.0	0.0	0.6	2.3	72.2	161.1	101.4	120.0	68.8	0.3	0.0	526.7
2849	2849	MATATHWADA	2013	1.5	9.4	2.6	7.9	6.4	160.9	293.4	136.9	154.1	94.3	7.4	13.1	888.0
2850	2850	MATATHWADA	2014	1.4	13.4	79.0	11.9	7.0	30.4	105.0	178.9	84.5	14.2	19.9	3.3	548.9
2851	2851	MATATHWADA	2015	10.1	1.6	32.0	39.6	12.3	118.3	27.4	112.2	154.3	19.5	4.8	0.0	532.2
2852	2852	VIDARBHA	1901	36.8	39.9	30.9	26.1	7.3	129.7	295.3	368.8	123.4	35.2	0.0	0.0	1093.3

114 rows × 20 columns

```
In [7]: new_data.index
```

Out[7]: RangeIndex(start=2739, stop=2853, step=1)

```
In [8]: new_data.columns
```

Out[8]: 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 [9]: new_data.plot.line()
```

Out[9]: <AxesSubplot:>

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

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

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

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

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

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

```
In [13]: new_data.plot.pie(y='ANNUAL')
```

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



```
In [14]: new_data.plot.scatter(x='YEAR',y='ANNUAL')
```

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



```
In [15]: sns.pairplot(new_data)
```

```
Out[15]: <seaborn.axisgrid.PairGrid at 0x28219643b20>
```

```
In [16]: 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[16]: <AxesSubplot:xlabel='YEAR', ylabel='Density'>
```

```
In [17]: sns.heatmap(new_data.corr())
```

```
Out[17]: <AxesSubplot:>
```