

```
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")[2279:2392]
```

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

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 114 entries, 2279 to 2392
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
2279	2279	GUJARAT REGION	1903	0.3	0.1	1.4	0.0	12.3	30.1	452.9	202.0	183.2	5.4	0.0	0.0	887.6
2280	2280	GUJARAT REGION	1904	0.8	10.6	16.8	0.2	3.9	48.3	194.8	71.8	138.0	6.1	0.1	1.2	492.6
2281	2281	GUJARAT REGION	1905	0.1	0.7	1.1	0.3	0.0	20.1	668.3	37.9	81.3	1.4	0.2	0.1	811.5
2282	2282	GUJARAT REGION	1906	0.0	12.7	0.0	0.0	0.1	177.5	311.5	247.1	134.8	5.9	0.0	0.9	890.6
2283	2283	GUJARAT REGION	1907	0.3	14.6	0.2	2.1	0.4	72.4	325.0	564.4	17.5	0.0	0.0	0.0	997.0

```
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
2279	2279	GUJARAT REGION	1903	0.3	0.1	1.4	0.0	12.3	30.1	452.9	202.0	183.2	5.4	0.0	0.0	887.6
2280	2280	GUJARAT REGION	1904	0.8	10.6	16.8	0.2	3.9	48.3	194.8	71.8	138.0	6.1	0.1	1.2	492.6
2281	2281	GUJARAT REGION	1905	0.1	0.7	1.1	0.3	0.0	20.1	668.3	37.9	81.3	1.4	0.2	0.1	811.5
2282	2282	GUJARAT REGION	1906	0.0	12.7	0.0	0.0	0.1	177.5	311.5	247.1	134.8	5.9	0.0	0.9	890.6
2283	2283	GUJARAT REGION	1907	0.3	14.6	0.2	2.1	0.4	72.4	325.0	564.4	17.5	0.0	0.0	0.0	997.0
...
2388	2388	GUJARAT REGION	2012	0.1	0.0	0.0	0.0	0.0	34.4	178.2	230.3	263.8	7.1	0.0	0.0	714.0
2389	2389	GUJARAT REGION	2013	0.0	0.9	0.1	4.6	0.0	155.7	405.4	211.1	287.3	53.2	0.1	0.0	1118.6
2390	2390	GUJARAT REGION	2014	5.7	0.1	0.2	1.0	1.3	11.6	307.5	138.6	235.1	3.3	1.3	0.0	705.7
2391	2391	GUJARAT REGION	2015	1.8	0.0	6.1	5.5	0.9	120.7	354.7	37.4	93.4	2.2	0.3	0.0	622.9
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

114 rows × 20 columns

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

Out[7]: RangeIndex(start=2279, stop=2393, 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 0x210c1c5e5b0>
```

```
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:>
```