

## Importing Libraries

```
In [3]: import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

## Importing Datasets

```
In [4]: df=pd.read_csv("rainfall in india 1901-2015.csv")
df
```

Out[4]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP
0	0	ANDAMAN & NICOBAR ISLANDS	1901	49.2	87.1	29.2	2.3	528.8	517.5	365.1	481.1	332.6
1	1	ANDAMAN & NICOBAR ISLANDS	1902	0.0	159.8	12.2	0.0	446.1	537.1	228.9	753.7	666.2
2	2	ANDAMAN & NICOBAR ISLANDS	1903	12.7	144.0	0.0	1.0	235.1	479.9	728.4	326.7	339.0
3	3	ANDAMAN & NICOBAR ISLANDS	1904	9.4	14.7	0.0	202.4	304.5	495.1	502.0	160.1	820.4
4	4	ANDAMAN & NICOBAR ISLANDS	1905	1.3	0.0	3.3	26.9	279.5	628.7	368.7	330.5	297.0
...	...	...	...	...	...	...	...	...	...	...	...	...
4111	4111	LAKSHADWEEP	2011	5.1	2.8	3.1	85.9	107.2	153.6	350.2	254.0	255.2
4112	4112	LAKSHADWEEP	2012	19.2	0.1	1.6	76.8	21.2	327.0	231.5	381.2	179.8
4113	4113	LAKSHADWEEP	2013	26.2	34.4	37.5	5.3	88.3	426.2	296.4	154.4	180.0
4114	4114	LAKSHADWEEP	2014	53.2	16.1	4.4	14.9	57.4	244.1	116.1	466.1	132.2
4115	4115	LAKSHADWEEP	2015	2.2	0.5	3.7	87.1	133.1	296.6	257.5	146.4	160.4

4116 rows × 20 columns



## Data Cleaning and Data Preprocessing

```
In [5]: df=df.dropna()
```

```
In [6]: df.columns
```

```
Out[6]: 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')
```

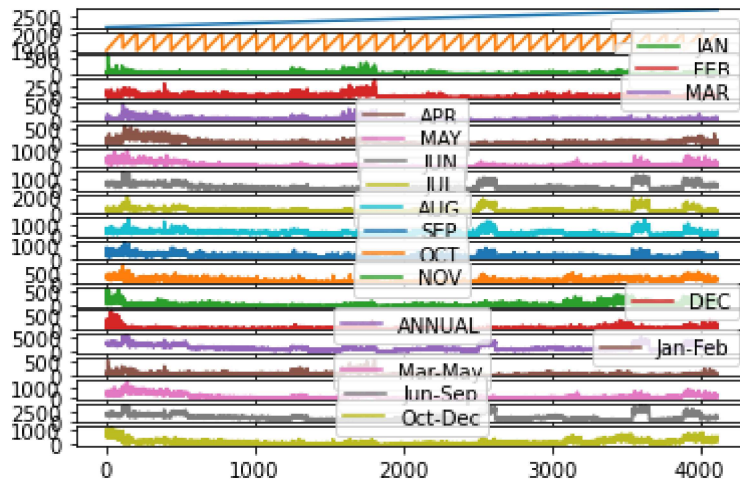
```
In [7]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 4090 entries, 0 to 4115
Data columns (total 20 columns):
#   Column          Non-Null Count  Dtype
---  -
0   index           4090 non-null   int64
1   SUBDIVISION     4090 non-null   object
2   YEAR            4090 non-null   int64
3   JAN             4090 non-null   float64
4   FEB             4090 non-null   float64
5   MAR             4090 non-null   float64
6   APR             4090 non-null   float64
7   MAY             4090 non-null   float64
8   JUN             4090 non-null   float64
9   JUL             4090 non-null   float64
10  AUG             4090 non-null   float64
11  SEP             4090 non-null   float64
12  OCT             4090 non-null   float64
13  NOV             4090 non-null   float64
14  DEC             4090 non-null   float64
15  ANNUAL          4090 non-null   float64
16  Jan-Feb         4090 non-null   float64
17  Mar-May         4090 non-null   float64
18  Jun-Sep         4090 non-null   float64
19  Oct-Dec         4090 non-null   float64
dtypes: float64(17), int64(2), object(1)
memory usage: 671.0+ KB
```

## Line chart

```
In [8]: df.plot.line(subplots=True)
```

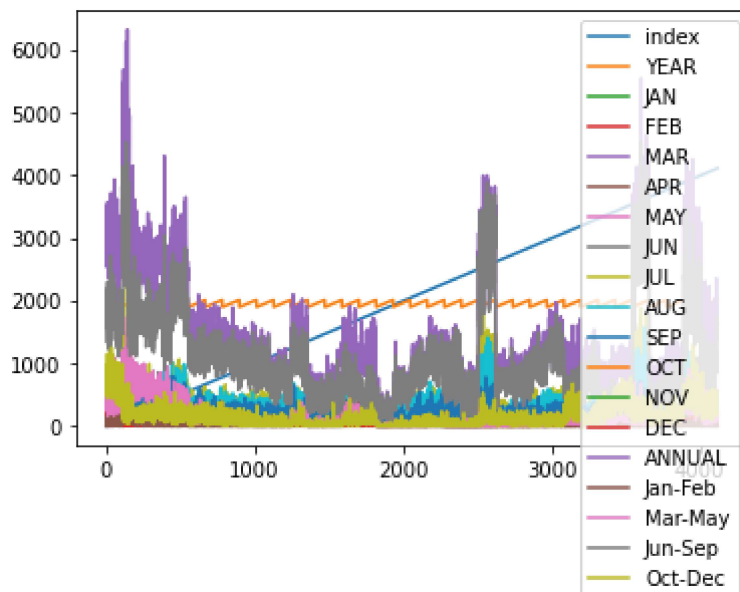
```
Out[8]: array([<AxesSubplot:~>, <AxesSubplot:~>, <AxesSubplot:~>, <AxesSubplot:~>,
<AxesSubplot:~>, <AxesSubplot:~>, <AxesSubplot:~>, <AxesSubplot:~>,
<AxesSubplot:~>, <AxesSubplot:~>, <AxesSubplot:~>, <AxesSubplot:~>,
<AxesSubplot:~>, <AxesSubplot:~>, <AxesSubplot:~>], dtype=object)
```



## Line chart

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

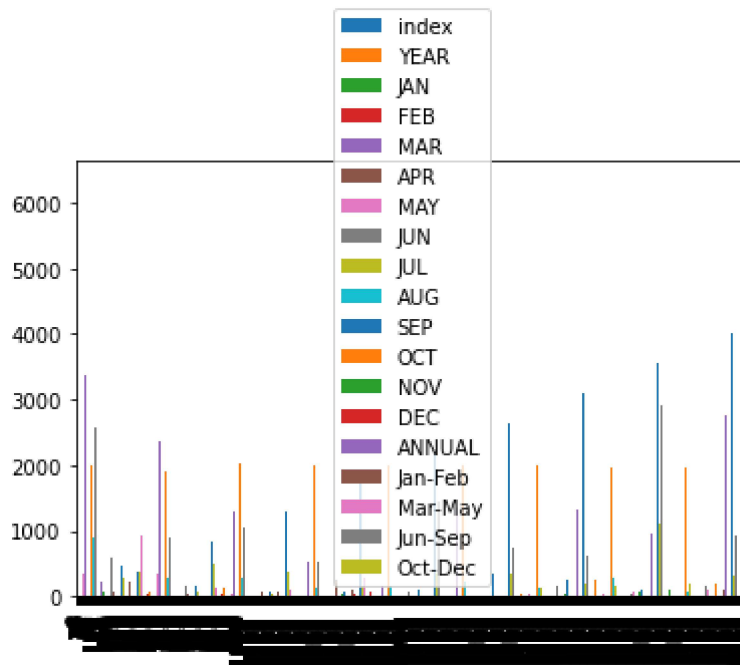
```
Out[9]: <AxesSubplot:~>
```



## Bar chart

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

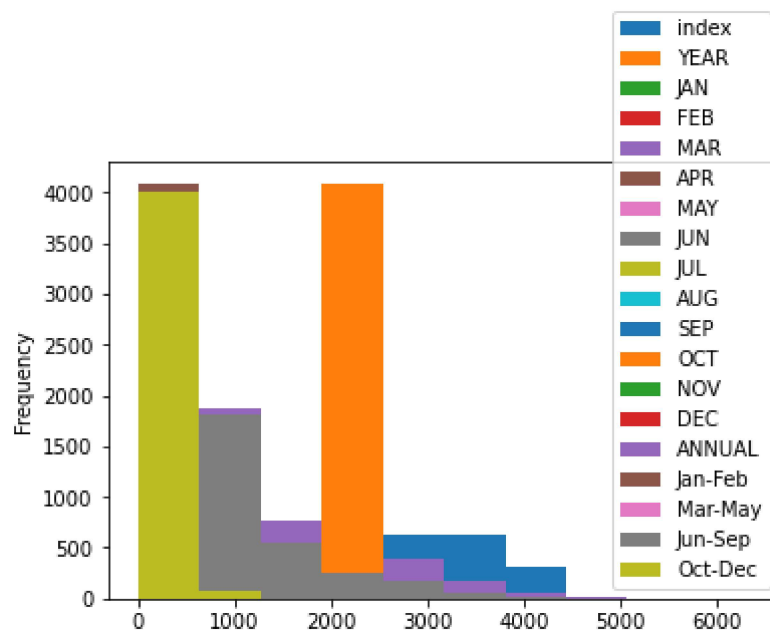
```
Out[10]: <AxesSubplot:>
```



## Histogram

```
In [11]: df.plot.hist()
```

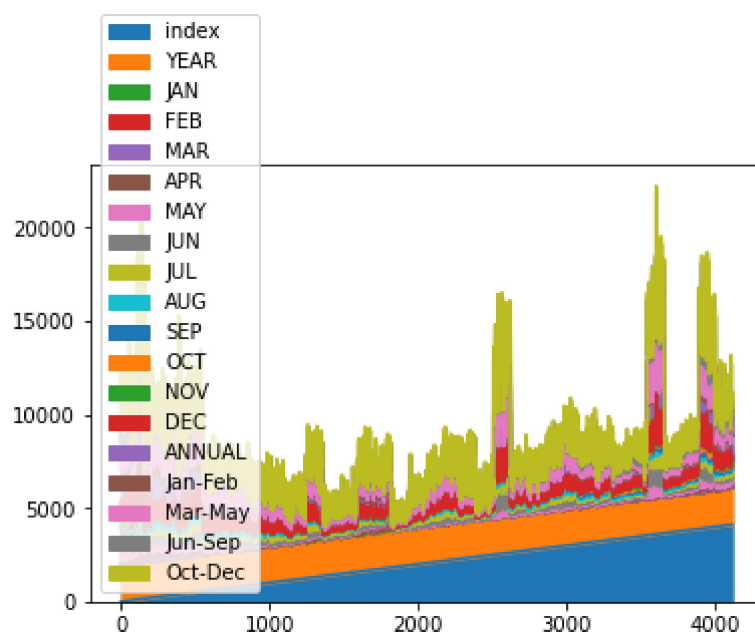
```
Out[11]: <AxesSubplot:ylabel='Frequency'>
```



## Area chart

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

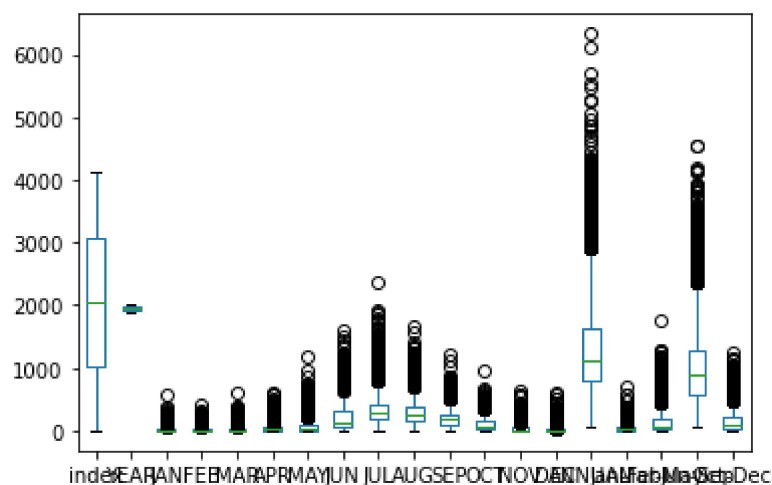
```
Out[12]: <AxesSubplot:>
```



## Box chart

```
In [13]: df.plot.box()
```

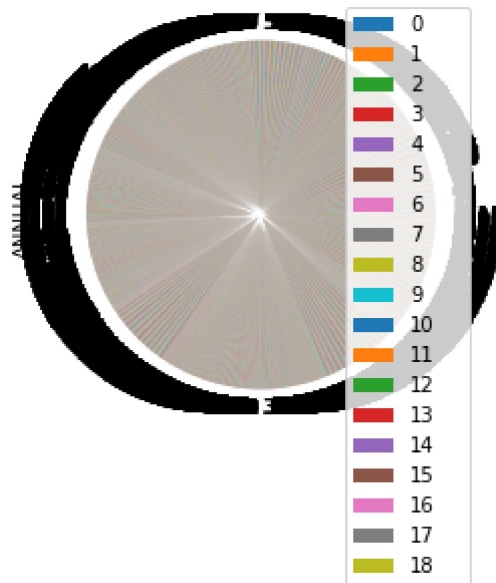
```
Out[13]: <AxesSubplot:>
```



## Pie chart

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

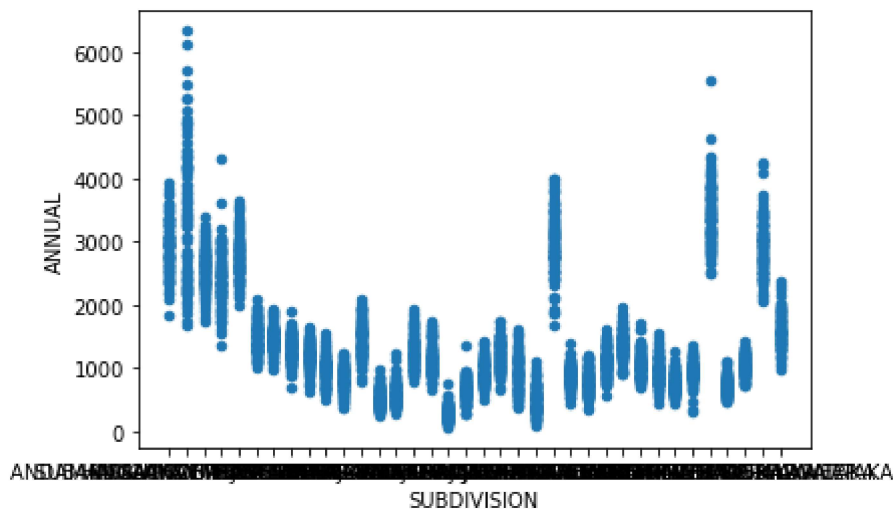
```
Out[14]: <AxesSubplot:ylabel='ANNUAL'>
```



## Scatter chart

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

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

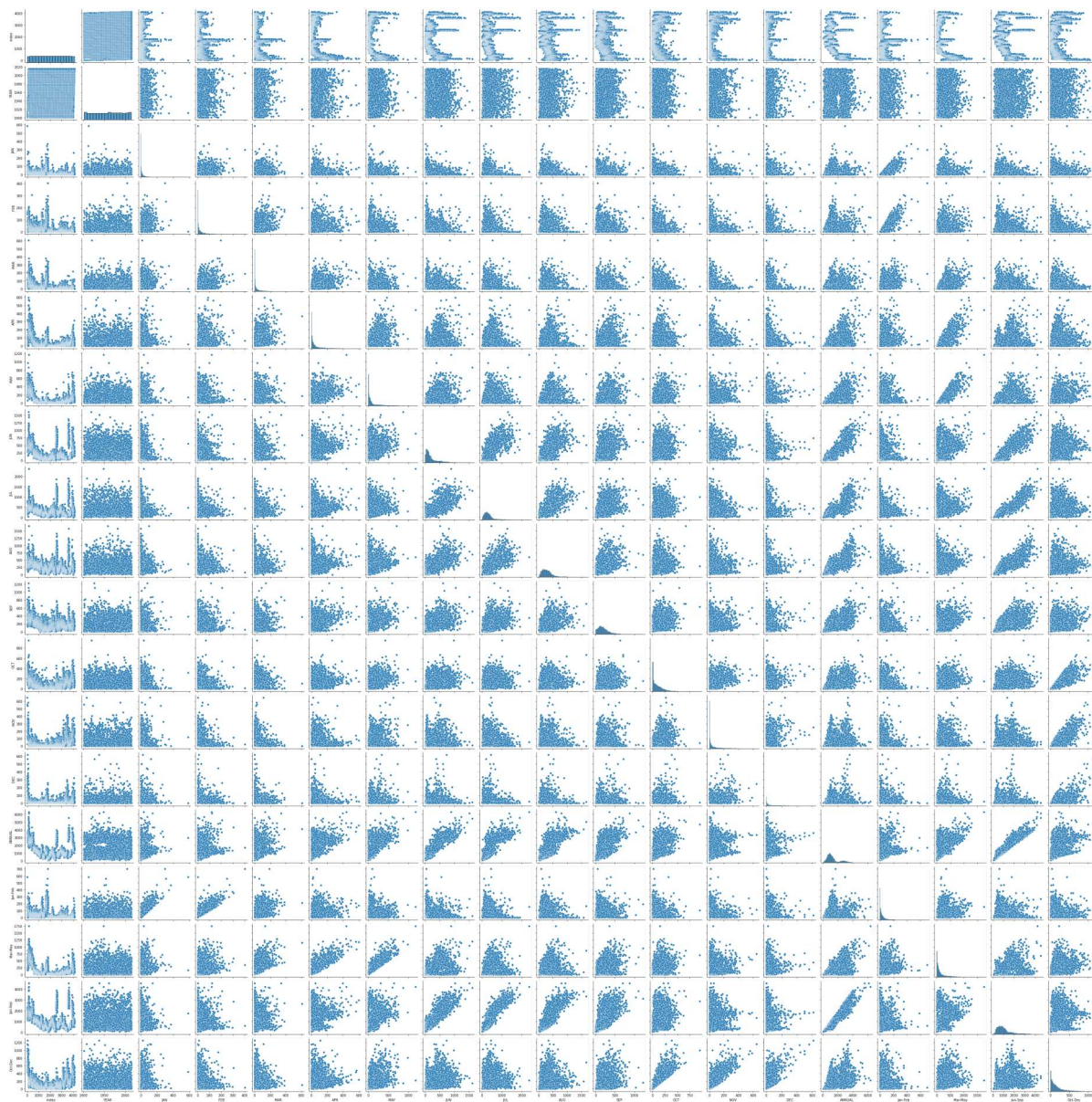


## Seaborn



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

```
Out[16]: <seaborn.axisgrid.PairGrid at 0x182380c6d60>
```

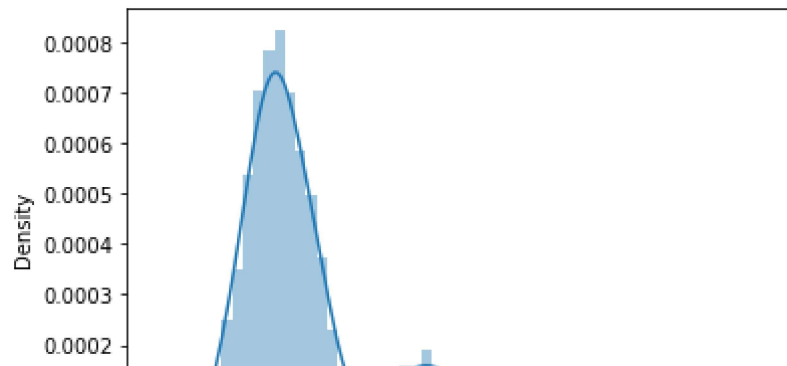


In [17]: `sns.distplot(df['ANNUAL'])`

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='ANNUAL', ylabel='Density'>`



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

Out[18]: `<AxesSubplot:>`

