## **Importing Libraries**

In [1]:

import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

### **Importing Datasets**

In [2]:

df=pd.read\_csv("rainfall in india 1901-2015.csv")
df

Out[2]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	ОСТ
0	0	ANDAMAN & NICOBAR ISLANDS	1901	49.2	87.1	29.2	2.3	528.8	517.5	365.1	481.1	332.6	388.5
1	1	ANDAMAN & NICOBAR ISLANDS	1902	0.0	159.8	12.2	0.0	446.1	537.1	228.9	753.7	666.2	197.2
2	2	ANDAMAN & NICOBAR ISLANDS	1903	12.7	144.0	0.0	1.0	235.1	479.9	728.4	326.7	339.0	181.2
3	3	ANDAMAN & NICOBAR ISLANDS	1904	9.4	14.7	0.0	202.4	304.5	495.1	502.0	160.1	820.4	222.2
4	4	ANDAMAN & NICOBAR ISLANDS	1905	1.3	0.0	3.3	26.9	279.5	628.7	368.7	330.5	297.0	260.7
•••													
4111	4111	LAKSHADWEEP	2011	5.1	2.8	3.1	85.9	107.2	153.6	350.2	254.0	255.2	117.4
4112	4112	LAKSHADWEEP	2012	19.2	0.1	1.6	76.8	21.2	327.0	231.5	381.2	179.8	145.9
4113	4113	LAKSHADWEEP	2013	26.2	34.4	37.5	5.3	88.3	426.2	296.4	154.4	180.0	72.8
4114	4114	LAKSHADWEEP	2014	53.2	16.1	4.4	14.9	57.4	244.1	116.1	466.1	132.2	169.2
4115	4115	LAKSHADWEEP	2015	2.2	0.5	3.7	87.1	133.1	296.6	257.5	146.4	160.4	165.4

4116 rows × 20 columns

#### head

In [3]:

df.head(5)
df

Out[3]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	ОСТ
0	0	ANDAMAN & NICOBAR ISLANDS	1901	49.2	87.1	29.2	2.3	528.8	517.5	365.1	481.1	332.6	388.5
1	1	ANDAMAN & NICOBAR ISLANDS	1902	0.0	159.8	12.2	0.0	446.1	537.1	228.9	753.7	666.2	197.2
2	2	ANDAMAN & NICOBAR ISLANDS	1903	12.7	144.0	0.0	1.0	235.1	479.9	728.4	326.7	339.0	181.2
3	3	ANDAMAN & NICOBAR ISLANDS	1904	9.4	14.7	0.0	202.4	304.5	495.1	502.0	160.1	820.4	222.2
4	4	ANDAMAN & NICOBAR ISLANDS	1905	1.3	0.0	3.3	26.9	279.5	628.7	368.7	330.5	297.0	260.7
•••													
4111	4111	LAKSHADWEEP	2011	5.1	2.8	3.1	85.9	107.2	153.6	350.2	254.0	255.2	117.4
4112	4112	LAKSHADWEEP	2012	19.2	0.1	1.6	76.8	21.2	327.0	231.5	381.2	179.8	145.9
4113	4113	LAKSHADWEEP	2013	26.2	34.4	37.5	5.3	88.3	426.2	296.4	154.4	180.0	72.8
4114	4114	LAKSHADWEEP	2014	53.2	16.1	4.4	14.9	57.4	244.1	116.1	466.1	132.2	169.2
4115	4115	LAKSHADWEEP	2015	2.2	0.5	3.7	87.1	133.1	296.6	257.5	146.4	160.4	165.4

4116 rows × 20 columns

tail

In [4]:

df.tail(5)
df

Out[4]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	ост
0	0	ANDAMAN & NICOBAR ISLANDS	1901	49.2	87.1	29.2	2.3	528.8	517.5	365.1	481.1	332.6	388.5
1	1	ANDAMAN & NICOBAR ISLANDS	1902	0.0	159.8	12.2	0.0	446.1	537.1	228.9	753.7	666.2	197.2
2	2	ANDAMAN & NICOBAR ISLANDS	1903	12.7	144.0	0.0	1.0	235.1	479.9	728.4	326.7	339.0	181.2
3	3	ANDAMAN & NICOBAR ISLANDS	1904	9.4	14.7	0.0	202.4	304.5	495.1	502.0	160.1	820.4	222.2
4	4	ANDAMAN & NICOBAR	1905	1.3	0.0	3.3	26.9	279.5	628.7	368.7	330.5	297.0	260.7

		index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	ОСТ
		ISLANDS												
	•••													
41	11	4111	LAKSHADWEEP	2011	5.1	2.8	3.1	85.9	107.2	153.6	350.2	254.0	255.2	117.4
41	12	4112	LAKSHADWEEP	2012	19.2	0.1	1.6	76.8	21.2	327.0	231.5	381.2	179.8	145.9
41	13	4113	LAKSHADWEEP	2013	26.2	34.4	37.5	5.3	88.3	426.2	296.4	154.4	180.0	72.8
41	14	4114	LAKSHADWEEP	2014	53.2	16.1	4.4	14.9	57.4	244.1	116.1	466.1	132.2	169.2
41	15	4115	LAKSHADWEEP	2015	2.2	0.5	3.7	87.1	133.1	296.6	257.5	146.4	160.4	165.4

4116 rows × 20 columns

# Data Cleaning and Data Preprocessing describe()

[5]:	df.de	df.describe()												
]:		index	YEAR	JAN	FEB	MAR	APR	MAY						
	count	4116.000000	4116.000000	4112.000000	4113.000000	4110.000000	4112.000000	4113.000000	4					
	mean	2057.500000	1958.218659	18.957320	21.805325	27.359197	43.127432	85.745417						
	std	1188.331183	33.140898	33.585371	35.909488	46.959424	67.831168	123.234904						
	min	0.000000	1901.000000	0.000000	0.000000	0.000000	0.000000	0.000000						
	25%	1028.750000	1930.000000	0.600000	0.600000	1.000000	3.000000	8.600000						
	50%	2057.500000	1958.000000	6.000000	6.700000	7.800000	15.700000	36.600000						
	75%	3086.250000	1987.000000	22.200000	26.800000	31.300000	49.950000	97.200000						
	max	4115.000000	2015.000000	583.700000	403.500000	605.600000	595.100000	1168.600000	1					
	4								<b>&gt;</b>					

### shape

```
In [6]: np.shape(df)
Out[6]: (4116, 20)
```

#### size

```
In [7]: np.size(df)
```

```
Out[7]: 82320
```

#### dropna

```
In [8]: df=df.dropna()
```

#### columns

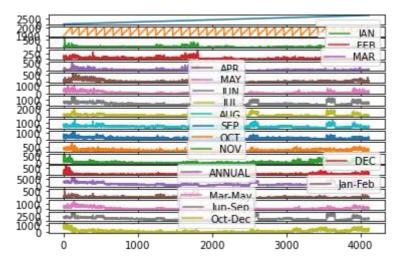
#### info()

```
In [10]:
          df.info()
         <class 'pandas.core.frame.DataFrame'>
         Int64Index: 4090 entries, 0 to 4115
         Data columns (total 20 columns):
                          Non-Null Count Dtype
              Column
          #
              ----
                          -----
          0
              index
                          4090 non-null
                                          int64
              SUBDIVISION 4090 non-null
          1
                                          object
          2
              YEAR
                          4090 non-null
                                          int64
          3
              JAN
                          4090 non-null
                                          float64
          4
              FEB
                          4090 non-null
                                          float64
          5
                          4090 non-null
                                          float64
              MAR
          6
                          4090 non-null
                                          float64
              APR
          7
                          4090 non-null
                                          float64
              MAY
          8
                          4090 non-null
                                          float64
              JUN
          9
                                          float64
              JUL
                          4090 non-null
          10
                          4090 non-null
                                          float64
             AUG
                                          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
          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 [11]: df.plot.line(subplots=True)

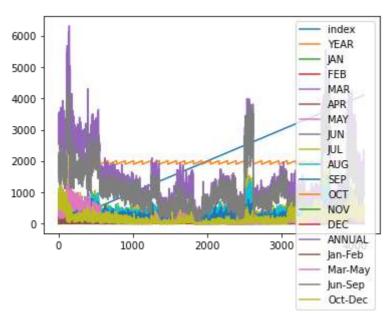
Out[11]: array([<AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>], dtype=object)
```



#### Line chart

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

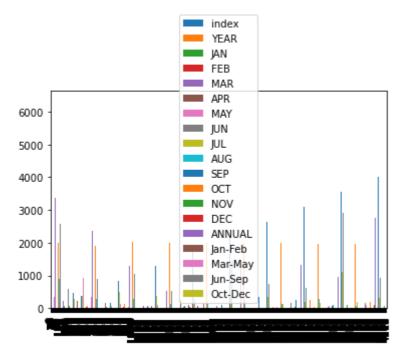
Out[12]: <AxesSubplot:>



#### Bar chart

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

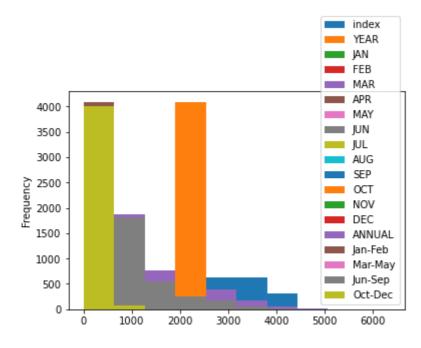
Out[13]: <AxesSubplot:>



## Histogram

```
In [14]: df.plot.hist()
```

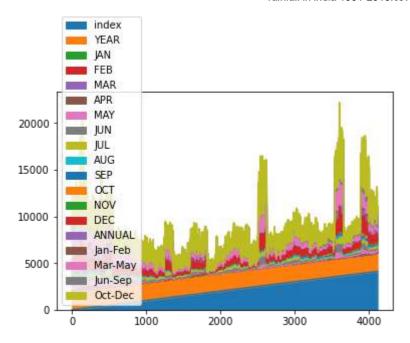
Out[14]: <AxesSubplot:ylabel='Frequency'>



#### Area chart

```
In [15]: df.plot.area()
```

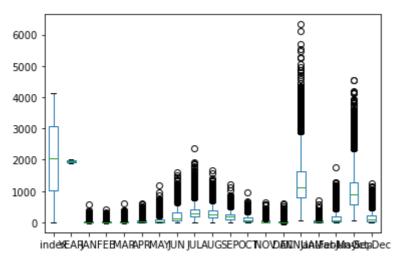
Out[15]: <AxesSubplot:>



#### **Box chart**

```
In [16]: df.plot.box()
```

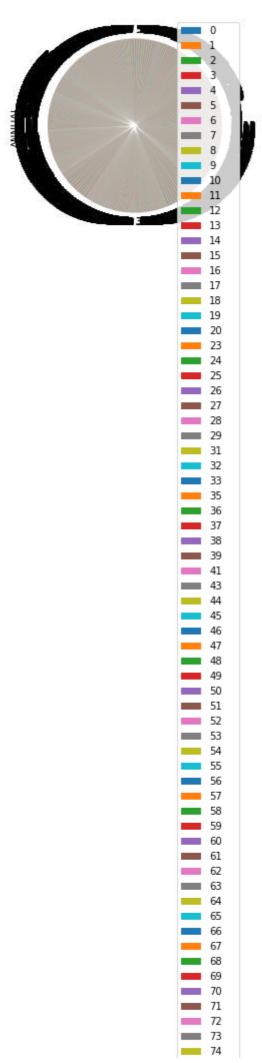
Out[16]: <AxesSubplot:>

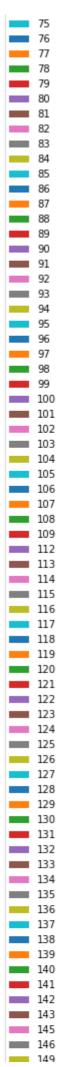


#### Pie chart

```
In [17]: df.plot.pie(y='ANNUAL')
```

Out[17]: <AxesSubplot:ylabel='ANNUAL'>





```
150
         Scatter chart
In [18]:
           df.plot.scatter(x='SUBDIVISION' ,y='ANNUAL')
                                      100
Out[18]: <AxesSubplot:xlabel='SUBDIVISTON', ylabel='ANNUAL'>
                6000
                5000
                4000
             ANNUAL
               3000
               2000
               1000
                  0
           ANSDARAMBIBANIEN
                                         SUBDIVISION
                                        178
In [19]:
           df.info()
          <class 'pandas.core.frame.DataFrame'>
Int64Index: 4090 entries, 0 to 4115
          Data columns (total 20 columns):
                              Non-NuII Count
           #
                Column
                                                Dtype
           0
                index
                              4090 non-ni
                                                int64
                              4090 non-nii
           1
                SUBDIVISION
                                                object
           2
                YEAR
                              4090 non-ni
                                                int64
           3
                JAN
                              4090 non-ni
                                                float64
           4
                FEB
                              4090 non-ni
                                                float64
           5
                MAR
                              4090 non-n
                                                float64
           6
                APR
                              4090 non-n
                                                float64
           7
                MAY
                              4090 non-
                                                float64
           8
                JUN
                              4090 non-
                                                float64
           9
                JUL
                              4090 non-
                                                float64
           10
                              4090 non-
                                                float64
               AUG
           11
                              4090 non
                                                float64
                SEP
           12
               OCT
                              4090 non
                                                float64
           13
               NOV
                              4090 non-
                                                float64
                                                float64
           14
                              4090 non-
               DEC
           15
               ANNUAL
                              4090 non
                                                float64
           16
               Jan-Feb
                              4090 non-
                                                float64
           17
               Mar-May
                              4090 non-
                                                float64
           18
               Jun-Sep
                              4090 non-nu
                                                float64
                              4090 non-nul
           19 Oct-Dec
                                                float64
          dtypes: float64(17), int64(2) object(1)
          memory usage: 671.0+ KB
In [20]:
           df.describe()
                                         213
                                                                        MAR
                                                                                     APR
                                                                                                 MAY
Out[20]:
                       index
                                   YEAR 214
                                                 JAN
                                                             FEB
          count 4090.00000 4090.00000 4090.00000 4090.00000 4090.00000 4090.00000 4090.00000 4090.00000 4
           mean 2057.638875 1958.321271
                                            18.818484
                                                        21.644792
                                                                    27.252494
                                                                                42.714548
                                                                                             84.868044
                                     219
                                         220
```

	index	YEAR	221 <b>JAN</b>	FEB	MAR	APR	MAY	
std	1182.612736	33.148944	33.521719	35.762010	46.829179	67.264863	122.556801	
min	0.000000	1901.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	1034.250000	1930.000000	0.600000	0.600000	1.000000	3.000000	8.500000	
50%	2057.500000	1959.000000	5.900000	6.600000	7.800000	15.500000	36.050000	
75%	3080.750000	1987.000000	21.950000	26.600000	31.100000	49.375000	94.975000	
max	4115.000000	2015.000000	583.700000	403.500000	605.600000	595.100000	1168.600000	1
	min 25% 50% 75%	std         1182.612736           min         0.000000           25%         1034.250000           50%         2057.500000           75%         3080.750000	std         1182.612736         33.148944           min         0.000000         1901.000000           25%         1034.250000         1930.000000           50%         2057.500000         1959.000000           75%         3080.750000         1987.000000	index         YEAR 221         JAN           std         1182.612736         33.148944         33.521719           min         0.000000         1901.000000         0.000000           25%         1034.250000         1930.000000         0.600000           50%         2057.500000         1959.00000         5.900000           75%         3080.750000         1987.000000         21.950000           max         4115.000000         2015.000000         583.700000	std         1182.612736         33.148944         33.521719         35.762010           min         0.000000         1901.000000         0.000000         0.000000           25%         1034.250000         1930.000000         5.90000         6.600000           75%         3080.750000         1987.000000         21.950000         403.500000           max         4115.000000         2015.000000         583.700000         403.5000000	std         1182.612736         33.148944         33.521719         35.762010         46.829179           min         0.000000         1901.000000         0.000000         0.000000         0.000000           25%         1034.250000         1930.000000         5.900000         6.600000         7.800000           50%         2057.500000         1987.000000         21.950000         26.600000         31.100000           max         4115.000000         2015.000000         583.700000         403.500000         605.600000	std         1182.612736         33.148944         33.521719         35.762010         46.829179         67.264863           min         0.000000         1901.000000         0.000000         0.000000         0.000000         0.000000           25%         1034.250000         1930.000000         5.900000         6.600000         7.800000         15.500000           50%         2057.500000         1987.000000         21.950000         26.600000         31.100000         49.375000           max         4115.000000         2015.000000         583.700000         403.500000         605.600000         595.100000	std         1182.612736         33.148944         33.521719         35.762010         46.829179         67.264863         122.556801           min         0.000000         1901.000000         0.000000         0.000000         0.000000         0.000000         0.000000           25%         1034.250000         1930.00000         5.90000         6.600000         7.800000         15.50000         36.050000           75%         3080.750000         1987.000000         583.700000         403.500000         605.600000         595.100000         1168.600000

# EDA AND VISUALIZATION

