

Importing Libraries

In [19]:

```
1 import numpy as np
2 import pandas as pd
3 import matplotlib.pyplot as plt
4 import seaborn as sns
```

Importing Datasets

In [20]:

```

1 df=pd.read_csv('PUNJAB.csv')
2 df

```

Out[20]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	ANNUAL	Jan-Feb	Mar-May	Ju-S
0	1472	PUNJAB	1901	55.7	50.1	25.2	2.1	25.2	10.4	178.2	145.0	24.4	3.7	0.0	3.3	523.5	105.9	52.5	351
1	1473	PUNJAB	1902	0.0	0.8	9.9	10.9	29.6	49.9	125.6	94.9	67.2	9.0	0.0	0.1	398.0	0.8	50.4	331
2	1474	PUNJAB	1903	29.5	0.5	45.0	1.3	9.2	5.2	212.2	119.1	132.5	6.9	0.0	9.5	571.0	29.9	55.5	461
3	1475	PUNJAB	1904	24.2	1.7	87.8	1.2	13.8	22.0	59.9	124.0	73.8	7.4	9.8	25.9	451.5	25.9	102.9	271
4	1476	PUNJAB	1905	53.0	40.3	24.3	0.5	2.2	19.2	122.6	50.3	111.1	1.2	0.0	9.4	434.3	93.3	27.0	301
...	
110	1582	PUNJAB	2011	3.5	35.6	8.2	17.8	18.9	162.9	120.9	193.5	140.2	0.0	1.0	2.6	705.2	39.2	44.9	611
111	1583	PUNJAB	2012	62.6	3.2	1.9	31.1	1.6	11.9	120.2	135.1	112.3	2.2	0.4	11.0	493.6	65.8	34.7	371
112	1584	PUNJAB	2013	9.3	50.1	11.6	3.4	3.6	120.3	117.9	217.1	24.4	16.2	6.1	6.6	586.6	59.4	18.6	471
113	1585	PUNJAB	2014	21.8	20.1	30.3	24.5	20.8	20.6	76.3	41.9	105.8	6.0	0.7	14.1	382.7	41.9	75.5	241
114	1586	PUNJAB	2015	17.7	31.3	68.5	29.8	16.7	48.3	130.2	88.6	69.2	9.0	0.8	0.7	510.8	49.0	115.0	331

115 rows × 20 columns



Out[20]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	ANNUAL	Jan-Feb	Mar-May	Jt S
0	1472	PUNJAB	1901	55.7	50.1	25.2	2.1	25.2	10.4	178.2	145.0	24.4	3.7	0.0	3.3	523.5	105.9	52.5	351
1	1473	PUNJAB	1902	0.0	0.8	9.9	10.9	29.6	49.9	125.6	94.9	67.2	9.0	0.0	0.1	398.0	0.8	50.4	331
2	1474	PUNJAB	1903	29.5	0.5	45.0	1.3	9.2	5.2	212.2	119.1	132.5	6.9	0.0	9.5	571.0	29.9	55.5	461
3	1475	PUNJAB	1904	24.2	1.7	87.8	1.2	13.8	22.0	59.9	124.0	73.8	7.4	9.8	25.9	451.5	25.9	102.9	271
4	1476	PUNJAB	1905	53.0	40.3	24.3	0.5	2.2	19.2	122.6	50.3	111.1	1.2	0.0	9.4	434.3	93.3	27.0	301
...	
110	1582	PUNJAB	2011	3.5	35.6	8.2	17.8	18.9	162.9	120.9	193.5	140.2	0.0	1.0	2.6	705.2	39.2	44.9	611
111	1583	PUNJAB	2012	62.6	3.2	1.9	31.1	1.6	11.9	120.2	135.1	112.3	2.2	0.4	11.0	493.6	65.8	34.7	371
112	1584	PUNJAB	2013	9.3	50.1	11.6	3.4	3.6	120.3	117.9	217.1	24.4	16.2	6.1	6.6	586.6	59.4	18.6	471
113	1585	PUNJAB	2014	21.8	20.1	30.3	24.5	20.8	20.6	76.3	41.9	105.8	6.0	0.7	14.1	382.7	41.9	75.5	241
114	1586	PUNJAB	2015	17.7	31.3	68.5	29.8	16.7	48.3	130.2	88.6	69.2	9.0	0.8	0.7	510.8	49.0	115.0	331

115 rows × 20 columns



Data Cleaning and Data Preprocessing

In [21]:

```

1 df=df.dropna()
2 df

```

Out[21]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	ANNUAL	Jan-Feb	Mar-May	Ju-S
0	1472	PUNJAB	1901	55.7	50.1	25.2	2.1	25.2	10.4	178.2	145.0	24.4	3.7	0.0	3.3	523.5	105.9	52.5	351
1	1473	PUNJAB	1902	0.0	0.8	9.9	10.9	29.6	49.9	125.6	94.9	67.2	9.0	0.0	0.1	398.0	0.8	50.4	331
2	1474	PUNJAB	1903	29.5	0.5	45.0	1.3	9.2	5.2	212.2	119.1	132.5	6.9	0.0	9.5	571.0	29.9	55.5	461
3	1475	PUNJAB	1904	24.2	1.7	87.8	1.2	13.8	22.0	59.9	124.0	73.8	7.4	9.8	25.9	451.5	25.9	102.9	271
4	1476	PUNJAB	1905	53.0	40.3	24.3	0.5	2.2	19.2	122.6	50.3	111.1	1.2	0.0	9.4	434.3	93.3	27.0	301
...	
110	1582	PUNJAB	2011	3.5	35.6	8.2	17.8	18.9	162.9	120.9	193.5	140.2	0.0	1.0	2.6	705.2	39.2	44.9	611
111	1583	PUNJAB	2012	62.6	3.2	1.9	31.1	1.6	11.9	120.2	135.1	112.3	2.2	0.4	11.0	493.6	65.8	34.7	371
112	1584	PUNJAB	2013	9.3	50.1	11.6	3.4	3.6	120.3	117.9	217.1	24.4	16.2	6.1	6.6	586.6	59.4	18.6	471
113	1585	PUNJAB	2014	21.8	20.1	30.3	24.5	20.8	20.6	76.3	41.9	105.8	6.0	0.7	14.1	382.7	41.9	75.5	241
114	1586	PUNJAB	2015	17.7	31.3	68.5	29.8	16.7	48.3	130.2	88.6	69.2	9.0	0.8	0.7	510.8	49.0	115.0	331

115 rows × 20 columns



Out[21]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	ANNUAL	Jan-Feb	Mar-May	Jt S
0	1472	PUNJAB	1901	55.7	50.1	25.2	2.1	25.2	10.4	178.2	145.0	24.4	3.7	0.0	3.3	523.5	105.9	52.5	351
1	1473	PUNJAB	1902	0.0	0.8	9.9	10.9	29.6	49.9	125.6	94.9	67.2	9.0	0.0	0.1	398.0	0.8	50.4	331
2	1474	PUNJAB	1903	29.5	0.5	45.0	1.3	9.2	5.2	212.2	119.1	132.5	6.9	0.0	9.5	571.0	29.9	55.5	461
3	1475	PUNJAB	1904	24.2	1.7	87.8	1.2	13.8	22.0	59.9	124.0	73.8	7.4	9.8	25.9	451.5	25.9	102.9	271
4	1476	PUNJAB	1905	53.0	40.3	24.3	0.5	2.2	19.2	122.6	50.3	111.1	1.2	0.0	9.4	434.3	93.3	27.0	301
...	
110	1582	PUNJAB	2011	3.5	35.6	8.2	17.8	18.9	162.9	120.9	193.5	140.2	0.0	1.0	2.6	705.2	39.2	44.9	611
111	1583	PUNJAB	2012	62.6	3.2	1.9	31.1	1.6	11.9	120.2	135.1	112.3	2.2	0.4	11.0	493.6	65.8	34.7	371
112	1584	PUNJAB	2013	9.3	50.1	11.6	3.4	3.6	120.3	117.9	217.1	24.4	16.2	6.1	6.6	586.6	59.4	18.6	471
113	1585	PUNJAB	2014	21.8	20.1	30.3	24.5	20.8	20.6	76.3	41.9	105.8	6.0	0.7	14.1	382.7	41.9	75.5	241
114	1586	PUNJAB	2015	17.7	31.3	68.5	29.8	16.7	48.3	130.2	88.6	69.2	9.0	0.8	0.7	510.8	49.0	115.0	331

115 rows × 20 columns



In [22]: 1 df.columns

Out[22]: 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')

Out[22]: 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 [23]: 1 df.info()

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 115 entries, 0 to 114
Data columns (total 20 columns):
 #   Column      Non-Null Count  Dtype  
--- 
 0   index       115 non-null    int64  
 1   SUBDIVISION 115 non-null    object  
 2   YEAR        115 non-null    int64  
 3   JAN         115 non-null    float64 
 4   FEB         115 non-null    float64 
 5   MAR         115 non-null    float64 
 6   APR         115 non-null    float64 
 7   MAY         115 non-null    float64 
 8   JUN         115 non-null    float64 
 9   JUL         115 non-null    float64 
 10  AUG         115 non-null    float64 
 11  SEP         115 non-null    float64 
 12  OCT         115 non-null    float64 
 13  NOV         115 non-null    float64 
 14  DEC         115 non-null    float64 
 15  ANNUAL      115 non-null    float64 
 16  Jan-Feb     115 non-null    float64 
 17  Mar-May     115 non-null    float64 
 18  Jun-Sep     115 non-null    float64 
 19  Oct-Dec     115 non-null    float64 
dtypes: float64(17), int64(2), object(1)
memory usage: 18.9+ KB
<class 'pandas.core.frame.DataFrame'>
Int64Index: 115 entries, 0 to 114
Data columns (total 20 columns):
 #   Column      Non-Null Count  Dtype  
--- 
 0   index       115 non-null    int64  
 1   SUBDIVISION 115 non-null    object  
 2   YEAR        115 non-null    int64  
 3   JAN         115 non-null    float64 
 4   FEB         115 non-null    float64 
 5   MAR         115 non-null    float64 
 6   APR         115 non-null    float64 
 7   MAY         115 non-null    float64 
 8   JUN         115 non-null    float64 
 9   JUL         115 non-null    float64 
 10  AUG         115 non-null    float64
```

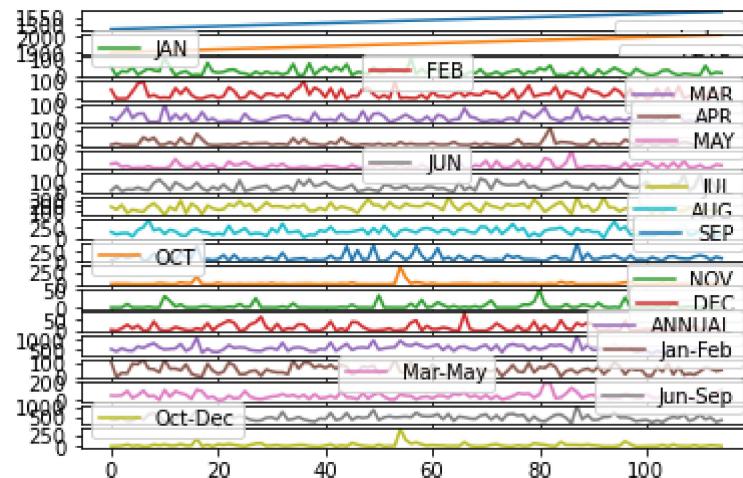
```
11 SEP           115 non-null   float64
12 OCT           115 non-null   float64
13 NOV           115 non-null   float64
14 DEC           115 non-null   float64
15 ANNUAL        115 non-null   float64
16 Jan-Feb       115 non-null   float64
17 Mar-May       115 non-null   float64
18 Jun-Sep       115 non-null   float64
19 Oct-Dec       115 non-null   float64
dtypes: float64(17), int64(2), object(1)
memory usage: 18.9+ KB
```

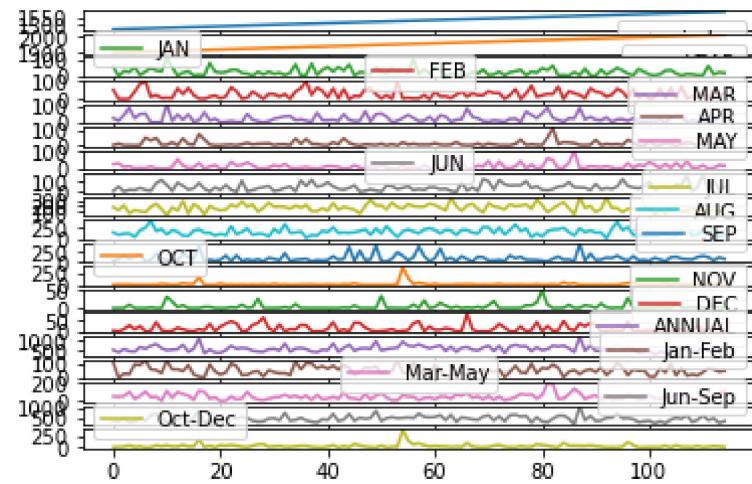
Line Chart

```
In [24]: 1 df.plot.line(subplots=True)
```

```
Out[24]: array([<AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>], dtype=object)
```

```
Out[24]: array([<AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>], dtype=object)
```

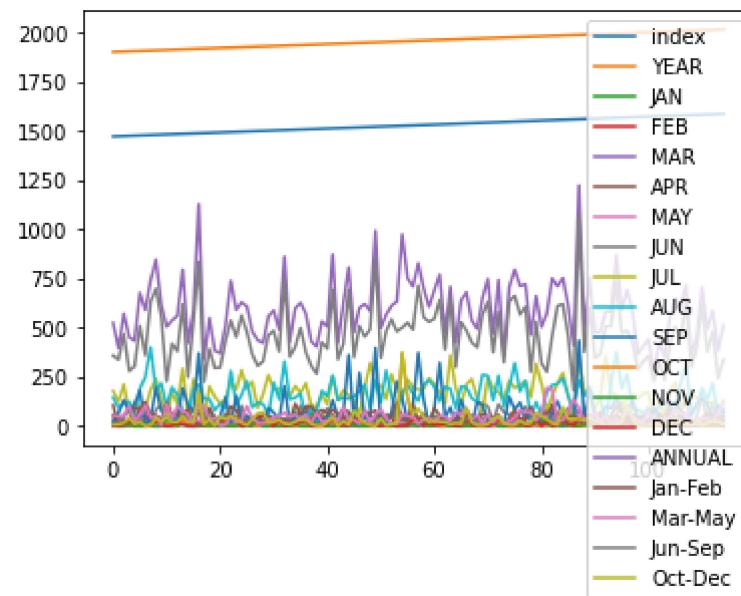
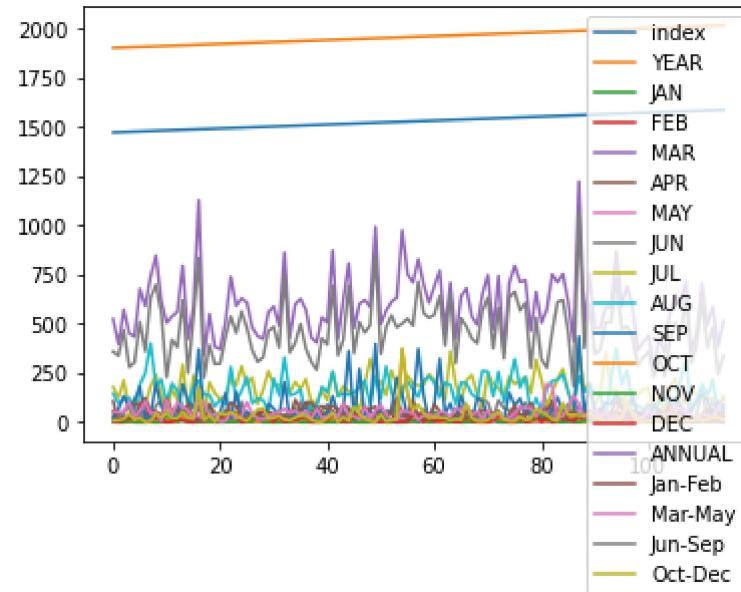




```
In [25]: 1 df.plot.line()
```

```
Out[25]: <AxesSubplot:>
```

```
Out[25]: <AxesSubplot:>
```

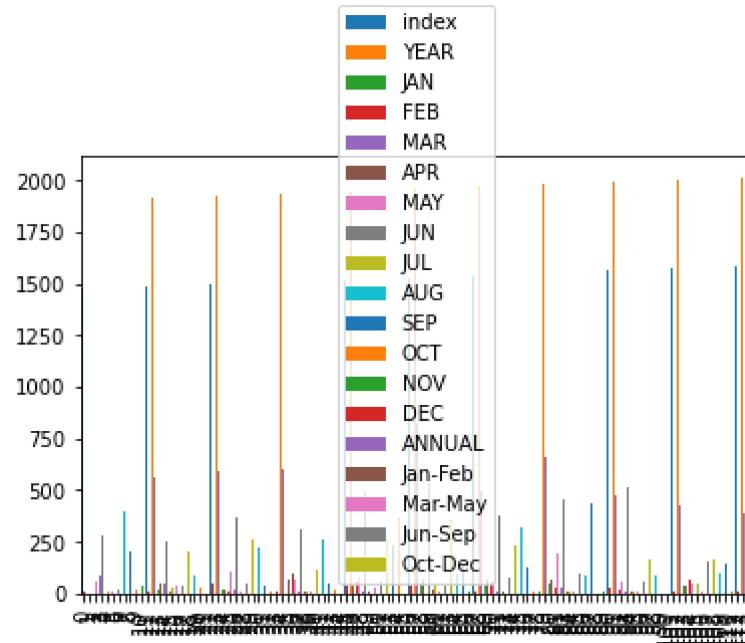


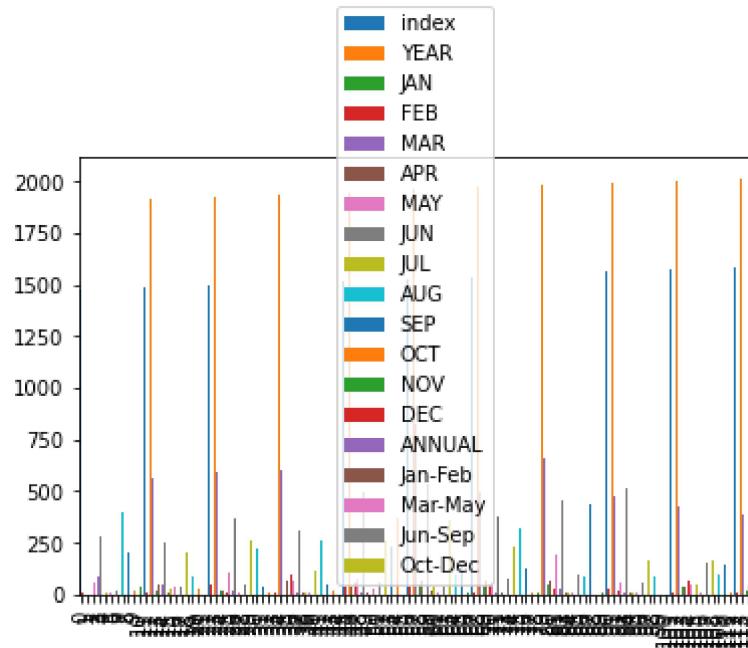
Bar Chart

```
In [26]: 1 df.plot.bar()
```

```
Out[26]: <AxesSubplot:>
```

```
Out[26]: <AxesSubplot:>
```



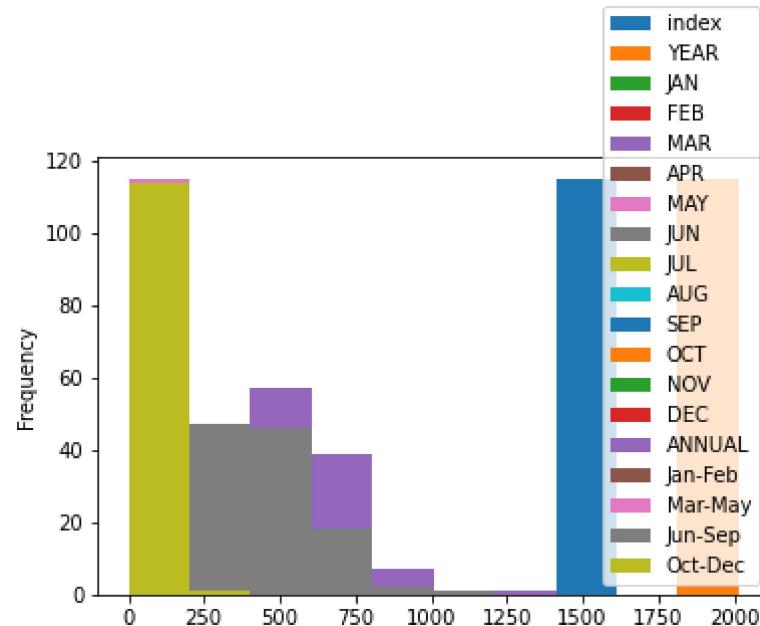


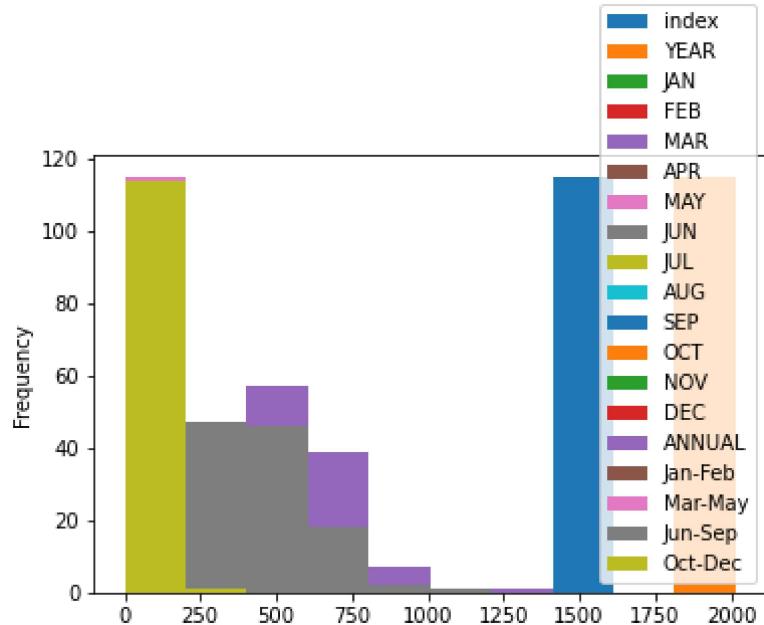
Histogram

```
In [27]: 1 df.plot.hist()
```

```
Out[27]: <AxesSubplot:ylabel='Frequency'>
```

```
Out[27]: <AxesSubplot:ylabel='Frequency'>
```



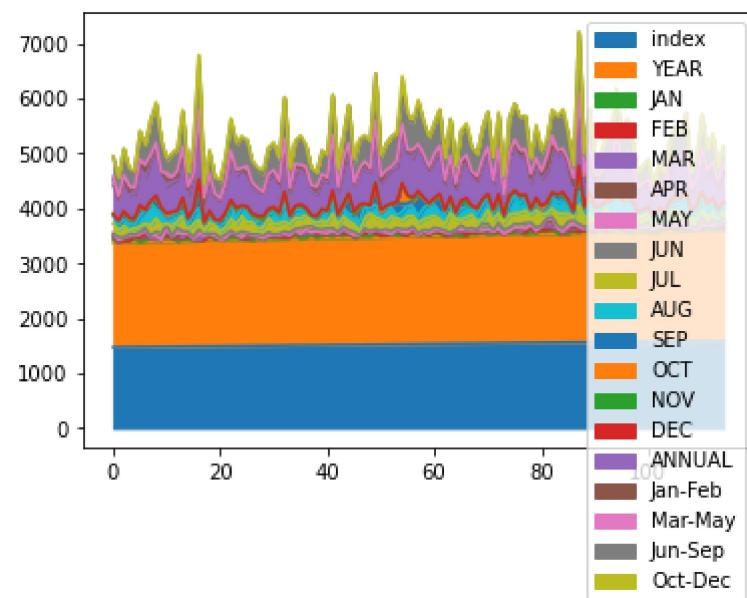
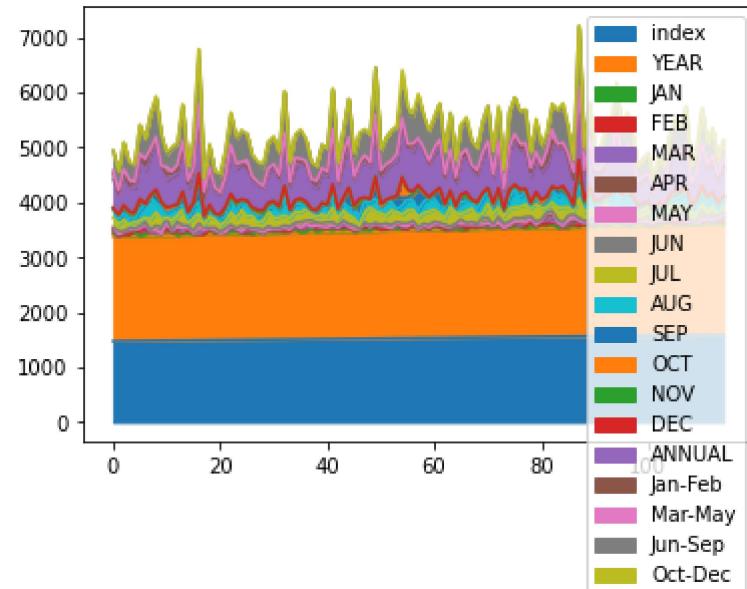


Area Chart

```
In [28]: 1 df.plot.area()
```

```
Out[28]: <AxesSubplot:>
```

```
Out[28]: <AxesSubplot:>
```

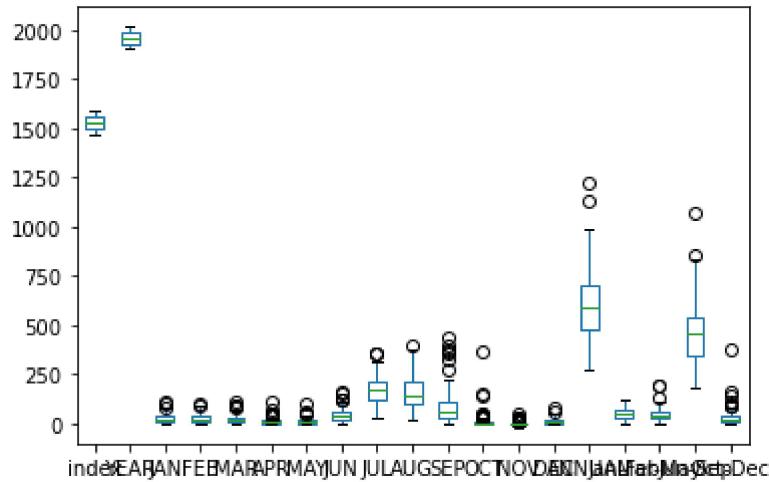
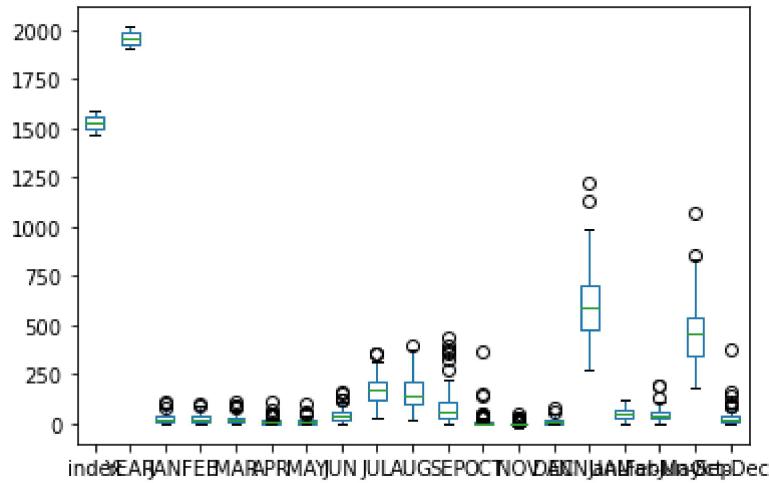


Box Chart

```
In [29]: 1 df.plot.box()
```

```
Out[29]: <AxesSubplot:>
```

```
Out[29]: <AxesSubplot:>
```

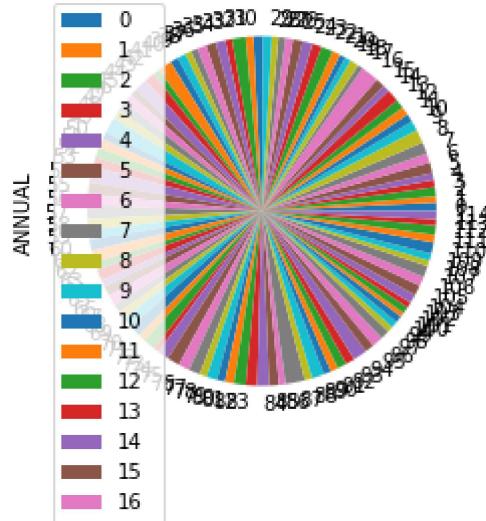


Pie Chart

```
In [30]: 1 df.plot.pie(y='ANNUAL')
```

```
Out[30]: <AxesSubplot:ylabel='ANNUAL'>
```

```
Out[30]: <AxesSubplot:ylabel='ANNUAL'>
```

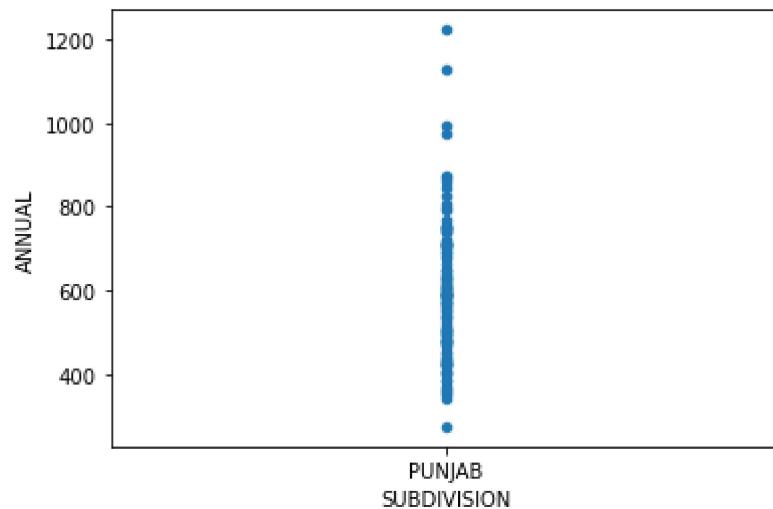
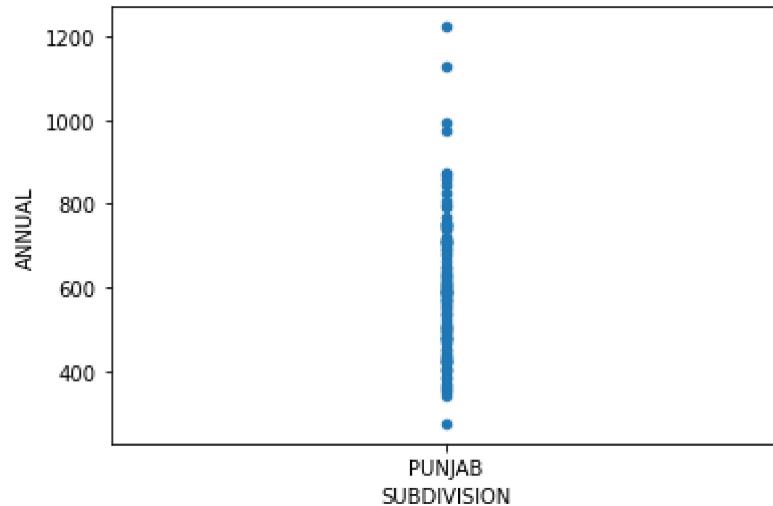


Scatter Plot

```
In [31]: 1 df.plot.scatter(x='SUBDIVISION',y='ANNUAL')
```

```
Out[31]: <AxesSubplot:xlabel='SUBDIVISION', ylabel='ANNUAL'>
```

```
Out[31]: <AxesSubplot:xlabel='SUBDIVISION', ylabel='ANNUAL'>
```



In [32]: 1 df.info()

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 115 entries, 0 to 114
Data columns (total 20 columns):
 #   Column      Non-Null Count  Dtype  
--- 
 0   index       115 non-null    int64  
 1   SUBDIVISION 115 non-null    object  
 2   YEAR        115 non-null    int64  
 3   JAN         115 non-null    float64 
 4   FEB         115 non-null    float64 
 5   MAR         115 non-null    float64 
 6   APR         115 non-null    float64 
 7   MAY         115 non-null    float64 
 8   JUN         115 non-null    float64 
 9   JUL         115 non-null    float64 
 10  AUG         115 non-null    float64 
 11  SEP         115 non-null    float64 
 12  OCT         115 non-null    float64 
 13  NOV         115 non-null    float64 
 14  DEC         115 non-null    float64 
 15  ANNUAL      115 non-null    float64 
 16  Jan-Feb     115 non-null    float64 
 17  Mar-May     115 non-null    float64 
 18  Jun-Sep     115 non-null    float64 
 19  Oct-Dec     115 non-null    float64 
dtypes: float64(17), int64(2), object(1)
memory usage: 18.9+ KB
<class 'pandas.core.frame.DataFrame'>
Int64Index: 115 entries, 0 to 114
Data columns (total 20 columns):
 #   Column      Non-Null Count  Dtype  
--- 
 0   index       115 non-null    int64  
 1   SUBDIVISION 115 non-null    object  
 2   YEAR        115 non-null    int64  
 3   JAN         115 non-null    float64 
 4   FEB         115 non-null    float64 
 5   MAR         115 non-null    float64 
 6   APR         115 non-null    float64 
 7   MAY         115 non-null    float64 
 8   JUN         115 non-null    float64 
 9   JUL         115 non-null    float64 
 10  AUG         115 non-null    float64
```

```
11 SEP          115 non-null    float64
12 OCT          115 non-null    float64
13 NOV          115 non-null    float64
14 DEC          115 non-null    float64
15 ANNUAL       115 non-null    float64
16 Jan-Feb      115 non-null    float64
17 Mar-May      115 non-null    float64
18 Jun-Sep      115 non-null    float64
19 Oct-Dec      115 non-null    float64
dtypes: float64(17), int64(2), object(1)
memory usage: 18.9+ KB
```

In [33]: 1 df.describe()

Out[33]:

	index	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG
count	115.000000	115.000000	115.000000	115.000000	115.000000	115.000000	115.000000	115.000000	115.000000	115.0
mean	1529.000000	1958.000000	25.246087	26.786957	23.651304	12.660000	14.136522	46.466957	168.963478	158.167826
std	33.341666	33.341666	22.306656	23.473612	22.890109	16.751778	15.185232	33.349257	67.729623	73.659290
min	1472.000000	1901.000000	0.000000	0.000000	0.000000	0.000000	0.100000	1.600000	26.400000	17.400000
25%	1500.500000	1929.500000	7.250000	5.650000	6.900000	2.550000	3.350000	21.600000	124.100000	96.000000
50%	1529.000000	1958.000000	21.600000	21.300000	15.800000	6.700000	9.200000	40.700000	167.400000	146.500000
75%	1557.500000	1986.500000	36.100000	40.600000	33.650000	15.700000	19.700000	60.150000	208.650000	208.400000
max	1586.000000	2015.000000	112.100000	96.000000	108.500000	113.200000	98.300000	162.900000	359.300000	399.100000

Out[33]:

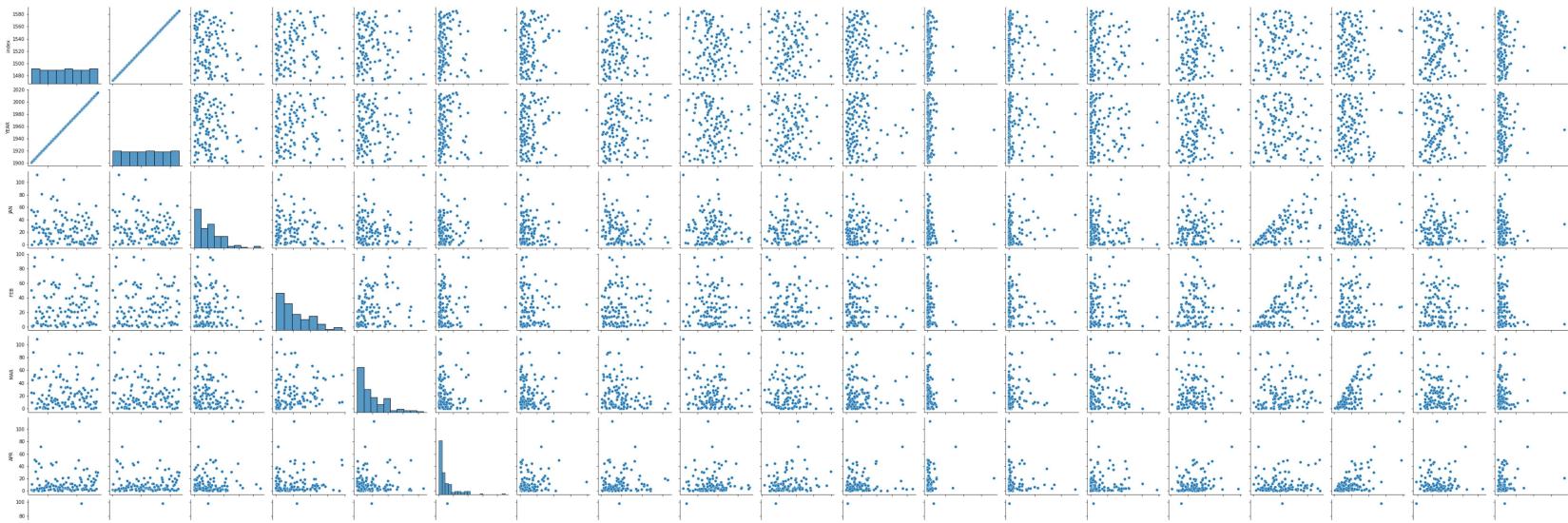
	index	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG
count	115.000000	115.000000	115.000000	115.000000	115.000000	115.000000	115.000000	115.000000	115.000000	115.0
mean	1529.000000	1958.000000	25.246087	26.786957	23.651304	12.660000	14.136522	46.466957	168.963478	158.167826
std	33.341666	33.341666	22.306656	23.473612	22.890109	16.751778	15.185232	33.349257	67.729623	73.659290
min	1472.000000	1901.000000	0.000000	0.000000	0.000000	0.000000	0.100000	1.600000	26.400000	17.400000
25%	1500.500000	1929.500000	7.250000	5.650000	6.900000	2.550000	3.350000	21.600000	124.100000	96.000000
50%	1529.000000	1958.000000	21.600000	21.300000	15.800000	6.700000	9.200000	40.700000	167.400000	146.500000
75%	1557.500000	1986.500000	36.100000	40.600000	33.650000	15.700000	19.700000	60.150000	208.650000	208.400000
max	1586.000000	2015.000000	112.100000	96.000000	108.500000	113.200000	98.300000	162.900000	359.300000	399.100000

EDA And Visualization

```
In [34]: 1 sns.pairplot(df)
```

```
Out[34]: <seaborn.axisgrid.PairGrid at 0x2450ce1ca60>
```

```
Out[34]: <seaborn.axisgrid.PairGrid at 0x2450ce1ca60>
```

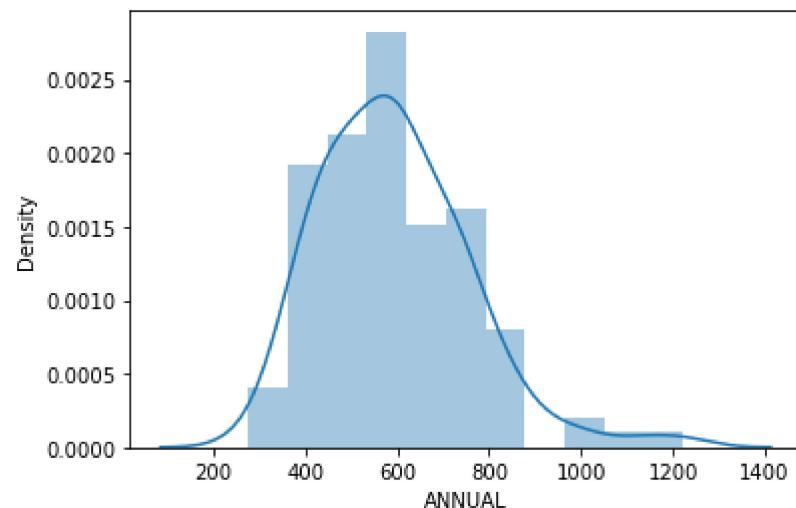


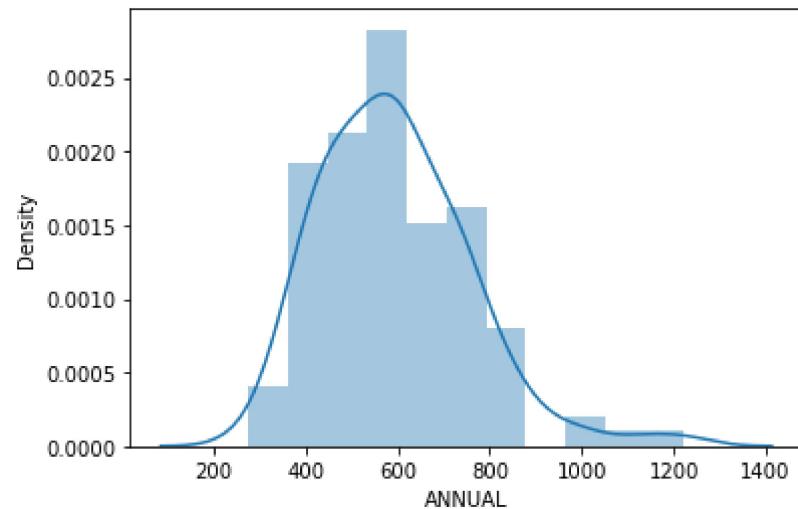
```
In [35]: 1 sns.distplot(df['ANNUAL'])
```

```
C:\ProgramData\Anaconda3\lib\site-packages\seaborn\distributions.py:2557: FutureWarning: `distplot` is a dep  
recated 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)  
C:\ProgramData\Anaconda3\lib\site-packages\seaborn\distributions.py:2557: FutureWarning: `distplot` is a dep  
recated 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[35]: <AxesSubplot:xlabel='ANNUAL', ylabel='Density'>
```

```
Out[35]: <AxesSubplot:xlabel='ANNUAL', ylabel='Density'>
```





```
In [36]: 1 sns.heatmap(df.corr())
```

Out[36]: <AxesSubplot:>

Out[36]: <AxesSubplot:>

