

Importing Libraries

In [23]:

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

Importing Datasets

In [25]:

```
df = pd.read_csv(r"C:\Users\user\Downloads\New folder\Arunachal.csv")
df
```

Out[25]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT
0	111	ARUNACHAL PRADESH	1917	21.4	164.5	NaN	269.6	107.9	823.8	909.1	628.4	411.5	191.5
1	112	ARUNACHAL PRADESH	1918	10.4	11.0	191.2	144.6	861.1	1609.9	1303.0	692.6	515.8	121.5
2	113	ARUNACHAL PRADESH	1919	34.5	67.8	28.5	256.9	420.6	973.6	999.0	286.7	628.7	94.5
3	114	ARUNACHAL PRADESH	1920	14.0	196.3	605.6	364.7	173.6	840.6	535.4	896.5	376.7	101.5
4	115	ARUNACHAL PRADESH	1921	78.9	54.3	180.3	358.0	598.0	1233.2	1433.0	885.9	603.4	241.5
...
91	202	ARUNACHAL PRADESH	2011	40.0	51.3	174.5	240.8	219.6	288.4	531.4	277.6	286.7	41.5
92	203	ARUNACHAL PRADESH	2012	57.8	35.8	134.2	403.4	187.4	645.8	638.9	316.0	724.9	241.5
93	204	ARUNACHAL PRADESH	2013	18.5	40.5	115.1	175.1	335.8	290.0	329.6	230.2	316.1	161.5
94	205	ARUNACHAL PRADESH	2014	19.0	101.9	80.3	86.7	299.0	415.8	392.4	599.6	343.0	41.5
95	206	ARUNACHAL PRADESH	2015	30.8	47.5	97.5	287.1	238.9	637.9	329.3	595.5	374.2	61.5

96 rows × 14 columns

Data Cleaning and Data Preprocessing

```
In [26]: df=df.dropna()
```

```
In [27]: df.columns
```

```
Out[27]: 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')
```

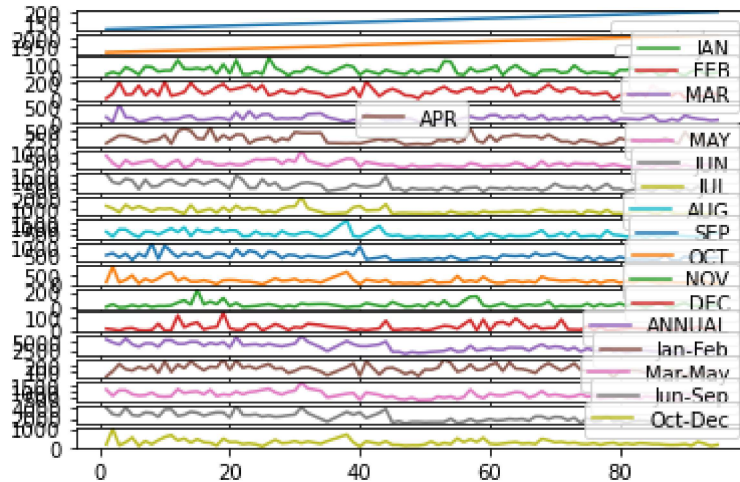
```
In [28]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 91 entries, 1 to 95
Data columns (total 20 columns):
#   Column          Non-Null Count  Dtype
---  -
0   index           91 non-null    int64
1   SUBDIVISION     91 non-null    object
2   YEAR            91 non-null    int64
3   JAN             91 non-null    float64
4   FEB             91 non-null    float64
5   MAR             91 non-null    float64
6   APR             91 non-null    float64
7   MAY             91 non-null    float64
8   JUN             91 non-null    float64
9   JUL             91 non-null    float64
10  AUG             91 non-null    float64
11  SEP             91 non-null    float64
12  OCT             91 non-null    float64
13  NOV             91 non-null    float64
14  DEC             91 non-null    float64
15  ANNUAL          91 non-null    float64
16  Jan-Feb        91 non-null    float64
17  Mar-May        91 non-null    float64
18  Jun-Sep        91 non-null    float64
19  Oct-Dec        91 non-null    float64
dtypes: float64(17), int64(2), object(1)
memory usage: 14.9+ KB
```

Line chart

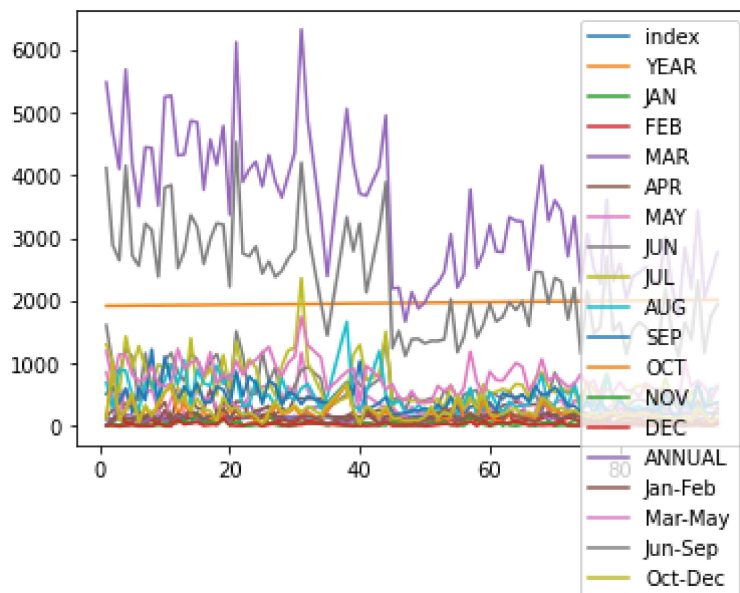
In [29]: `df.plot.line(subplots=True)`

Out[29]: array([<AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>], dtype=object)



In [30]: `df.plot.line()`

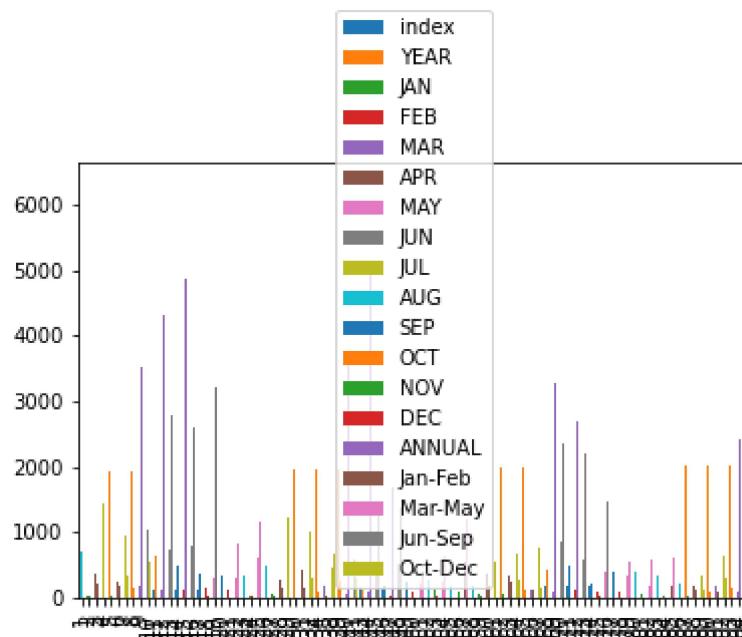
Out[30]: <AxesSubplot:>



Bar chart

```
In [31]: df.plot.bar()
```

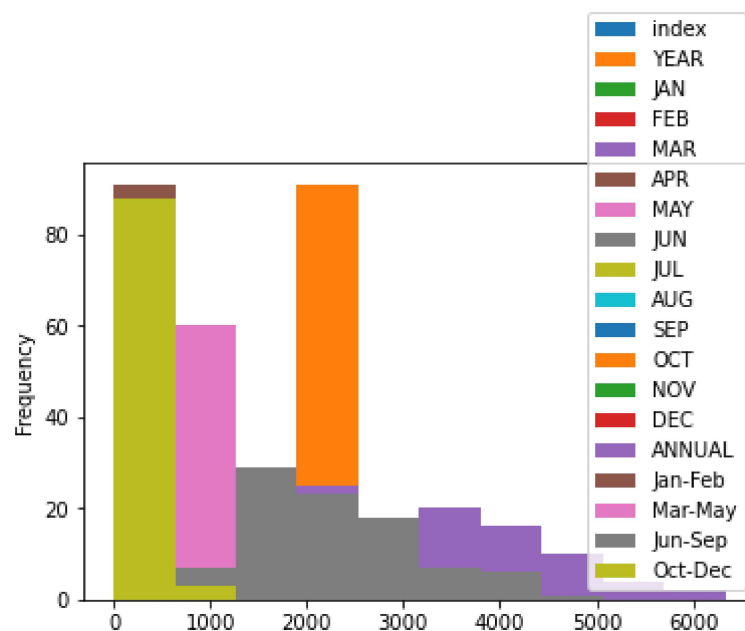
```
Out[31]: <AxesSubplot:>
```



Histogram

```
In [32]: df.plot.hist()
```

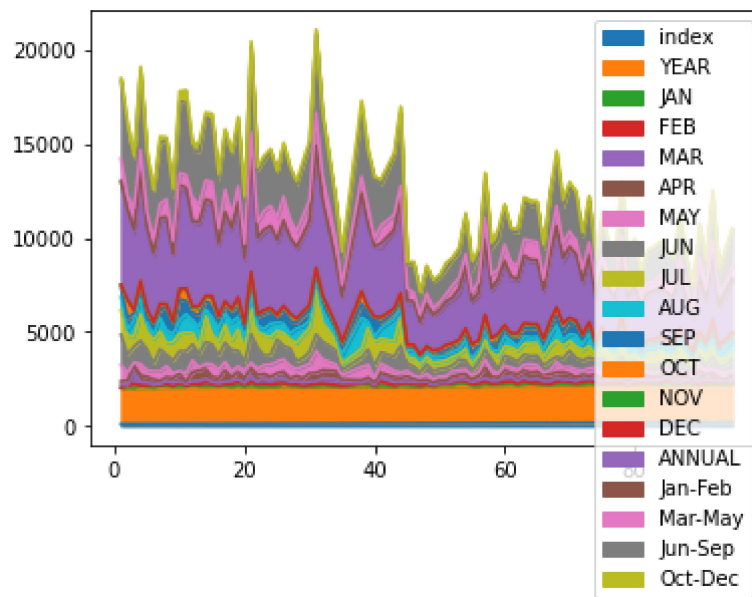
```
Out[32]: <AxesSubplot:ylabel='Frequency'>
```



Area chart

```
In [33]: df.plot.area()
```

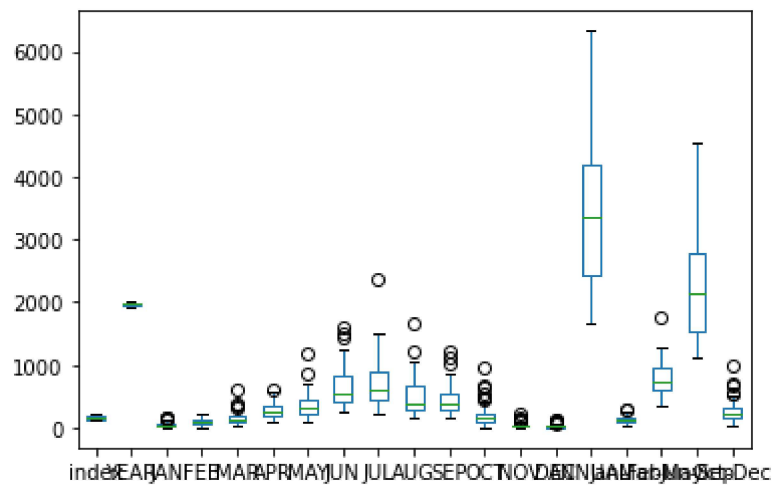
```
Out[33]: <AxesSubplot:>
```



Box plot

```
In [34]: df.plot.box()
```

```
Out[34]: <AxesSubplot:>
```

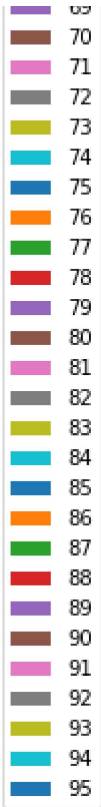


pie chart

```
In [35]: df.plot.pie(y='ANNUAL')
```

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

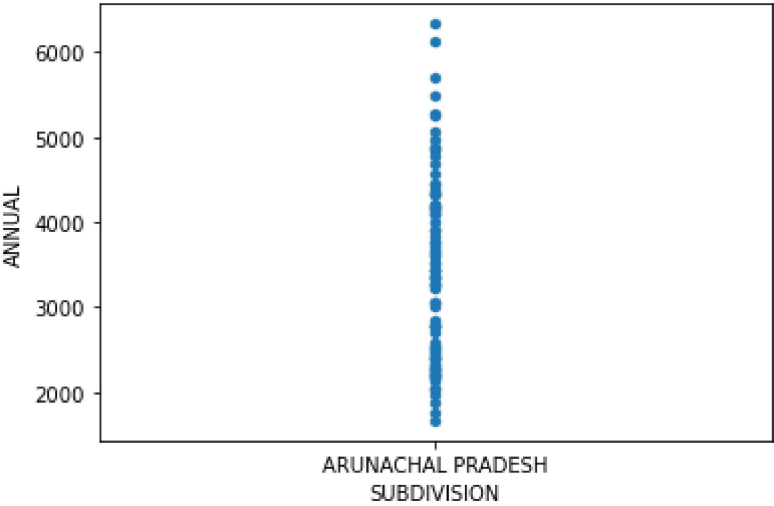


Scatter chart

```
In [36]: df.plot.scatter(x='SUBDIVISION',y='ANNUAL')
```

Out[36]: <AxesSubplot:xlabel='SUBDIVISION', ylabel='ANNUAL'>



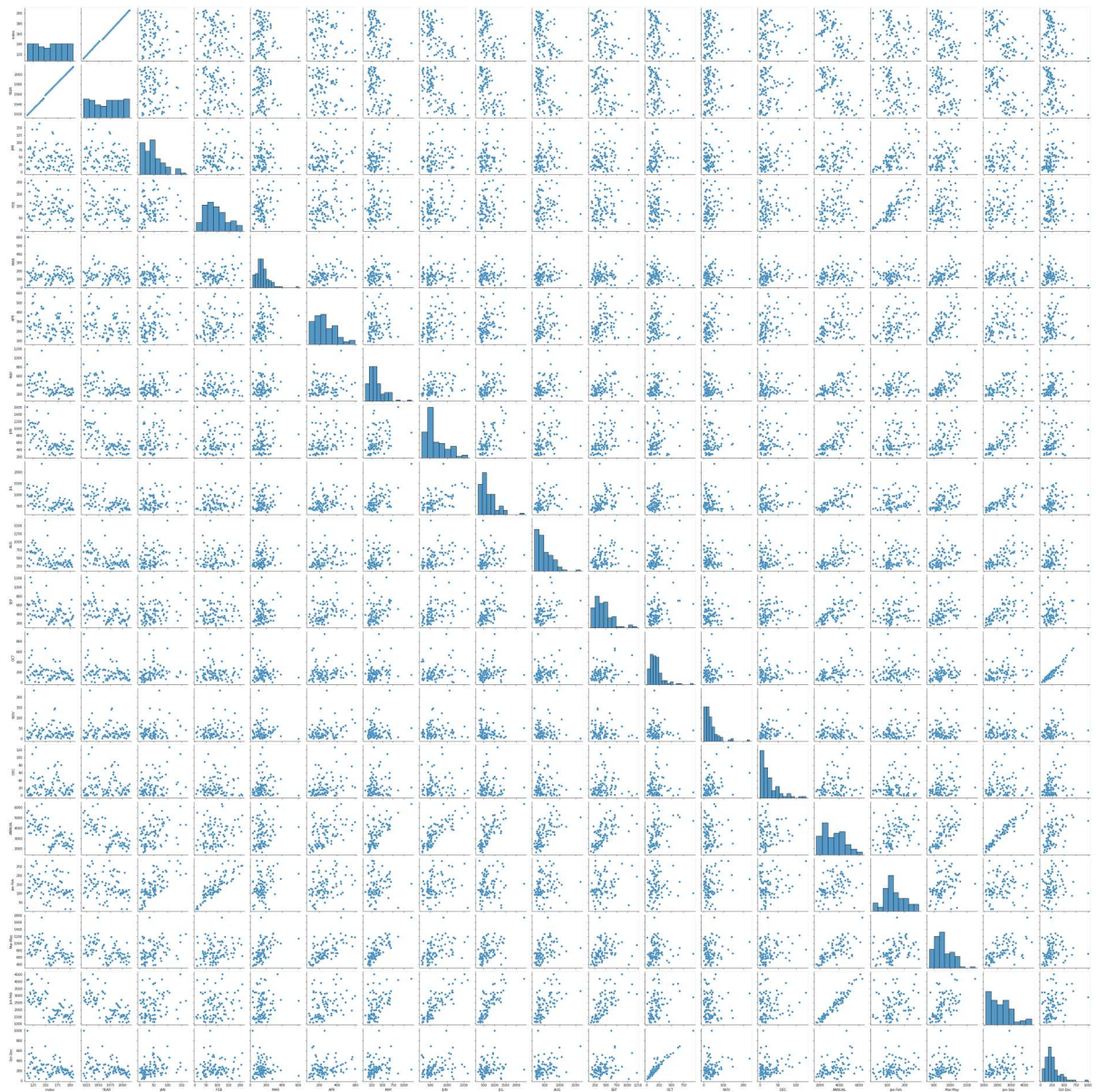
```
In [37]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 91 entries, 1 to 95
Data columns (total 20 columns):
#   Column          Non-Null Count  Dtype
---  -
0   index           91 non-null    int64
1   SUBDIVISION     91 non-null    object
2   YEAR            91 non-null    int64
3   JAN             91 non-null    float64
4   FEB             91 non-null    float64
5   MAR             91 non-null    float64
6   APR             91 non-null    float64
7   MAY             91 non-null    float64
8   JUN             91 non-null    float64
9   JUL             91 non-null    float64
10  AUG             91 non-null    float64
11  SEP             91 non-null    float64
12  OCT             91 non-null    float64
13  NOV             91 non-null    float64
14  DEC             91 non-null    float64
15  ANNUAL          91 non-null    float64
16  Jan-Feb        91 non-null    float64
17  Mar-May        91 non-null    float64
18  Jun-Sep        91 non-null    float64
19  Oct-Dec        91 non-null    float64
dtypes: float64(17), int64(2), object(1)
memory usage: 14.9+ KB
```

EDA AND VISUALIZATION

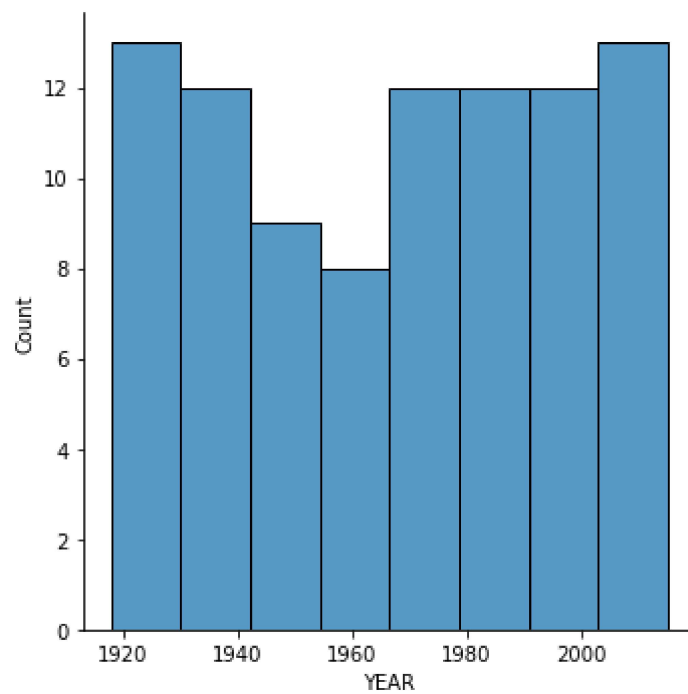
```
In [38]: sns.pairplot(df)
```

```
Out[38]: <seaborn.axisgrid.PairGrid at 0x1f5476aeaf0>
```



```
In [39]: sns.displot(df['YEAR'])
```

```
Out[39]: <seaborn.axisgrid.FacetGrid at 0x1f5476bf700>
```



```
In [40]: sns.heatmap(df.corr())
```

```
Out[40]: <AxesSubplot:>
```

