

kcubilecm

August 4, 2023

1 20104169 - SUMESH R

2 Importing Libraries

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

```
[2]: from google.colab import drive
drive.mount('/content/drive')
df=pd.read_csv("/content/drive/MyDrive/mydatasets/rainfall/rainfall_telangana.
↪csv")
df
```

Mounted at /content/drive

```
[2]:
```

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	\
0	3197	TELANGANA	1901	6.9	41.8	7.8	45.2	22.0	123.6	237.8	
1	3198	TELANGANA	1902	0.0	0.0	0.2	10.7	7.3	52.4	146.3	
2	3199	TELANGANA	1903	12.9	4.6	0.0	9.9	40.7	99.2	505.2	
3	3200	TELANGANA	1904	0.0	0.0	10.8	0.8	14.7	104.2	139.5	
4	3201	TELANGANA	1905	0.0	4.3	12.8	27.6	32.2	129.5	82.4	
..	
110	3307	TELANGANA	2011	0.0	11.9	2.6	25.6	9.3	83.9	268.2	
111	3308	TELANGANA	2012	6.7	0.0	0.2	14.0	8.4	124.4	300.3	
112	3309	TELANGANA	2013	2.4	29.0	0.2	24.4	8.5	213.4	453.8	
113	3310	TELANGANA	2014	0.2	2.9	58.3	10.3	73.3	62.3	146.0	
114	3311	TELANGANA	2015	17.5	0.0	43.0	65.7	23.3	266.9	104.4	
		AUG	SEP	OCT	NOV	DEC	ANNUAL	Jan-Feb	Mar-May	Jun-Sep	\
0	177.2	77.7	75.5	12.2	0.0	827.7	48.7	75.0	616.4		
1	142.8	190.5	41.7	31.2	7.3	630.4	0.0	18.2	532.0		
2	246.7	191.9	155.8	15.5	1.1	1283.4	17.5	50.5	1042.9		
3	50.0	162.3	44.4	0.0	0.0	526.7	0.0	26.3	456.0		
4	237.3	179.1	19.6	0.0	0.0	724.9	4.3	72.6	628.4		
..		

110	225.9	107.6	13.9	4.2	0.0	753.1	11.9	37.4	685.6
111	229.9	202.4	83.6	38.7	0.0	1008.6	6.7	22.6	857.0
112	230.6	161.4	205.9	16.4	2.7	1348.7	31.4	33.0	1059.2
113	205.2	146.8	29.6	10.8	0.7	746.4	3.1	141.9	560.2
114	160.5	158.3	15.6	0.3	1.7	857.3	17.5	132.0	690.1

	Oct-Dec
0	87.7
1	80.2
2	172.4
3	44.4
4	19.6
..	...
110	18.1
111	122.3
112	225.0
113	41.0
114	17.6

[115 rows x 20 columns]

3 Data Cleaning and Data Preprocessing

```
[3]: df=df.dropna()
```

```
[4]: df.columns
```

```
[4]: 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')
```

```
[5]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 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
```

```

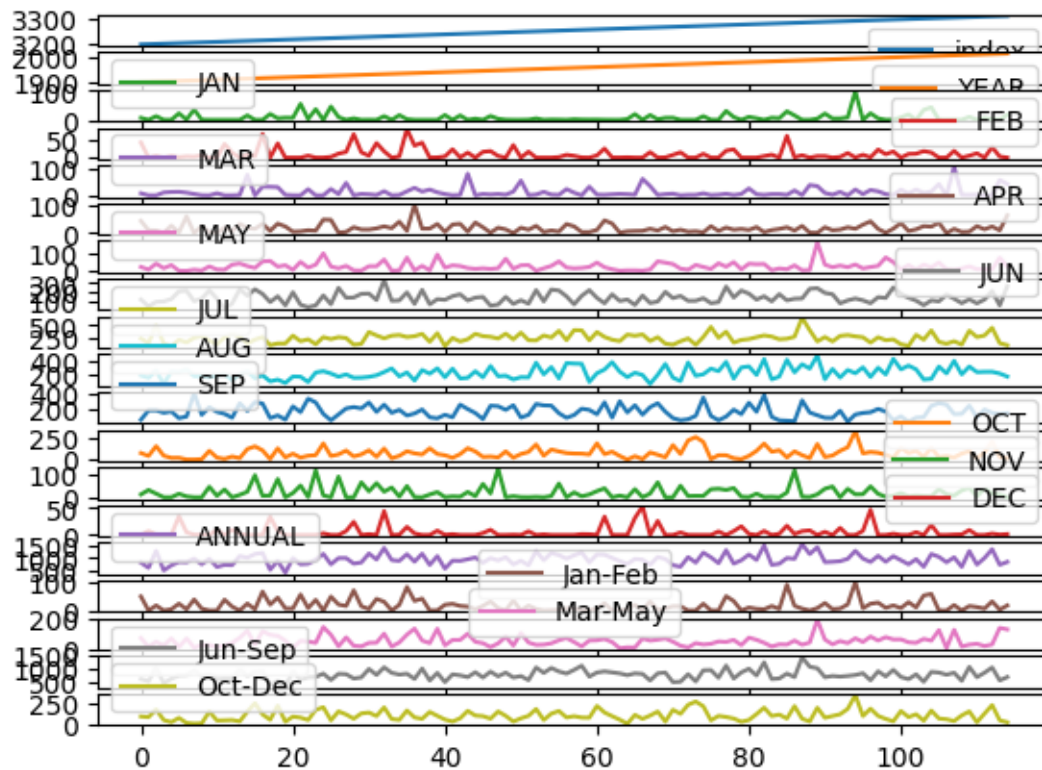
7 MAY 115 non-null float64
8 JUN 115 non-null float64
9 JUL 115 non-null float64
10 AUG 115 non-null float64
11 SEP 115 non-null float64
12 OCT 115 non-null float64
13 NOV 115 non-null float64
14 DEC 115 non-null float64
15 ANNUAL 115 non-null float64
16 Jan-Feb 115 non-null float64
17 Mar-May 115 non-null float64
18 Jun-Sep 115 non-null float64
19 Oct-Dec 115 non-null float64
dtypes: float64(17), int64(2), object(1)
memory usage: 18.1+ KB

```

4 Line chart

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

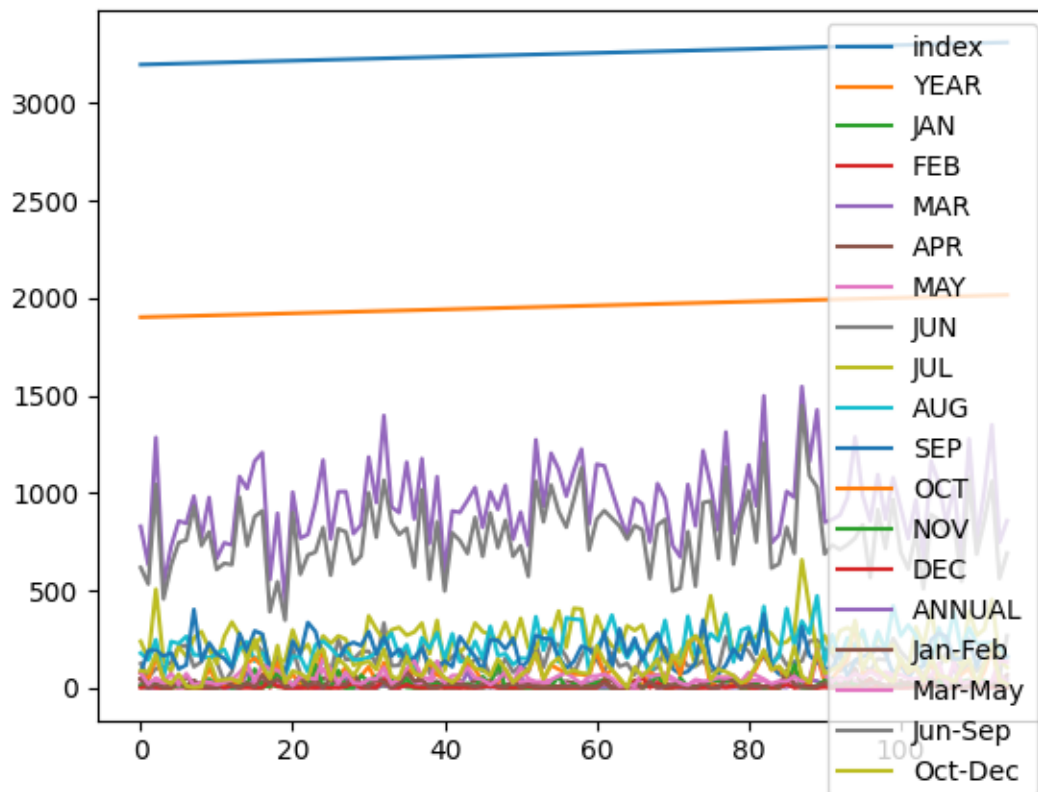
```
[6]: array([<Axes: >, <Axes: >, <Axes: >, <Axes: >, <Axes: >, <Axes: >,
<Axes: >, <Axes: >, <Axes: >, <Axes: >, <Axes: >, <Axes: >,
<Axes: >, <Axes: >, <Axes: >, <Axes: >, <Axes: >, <Axes: >,
<Axes: >], dtype=object)
```



5 Line chart

```
[7]: df.plot.line()
```

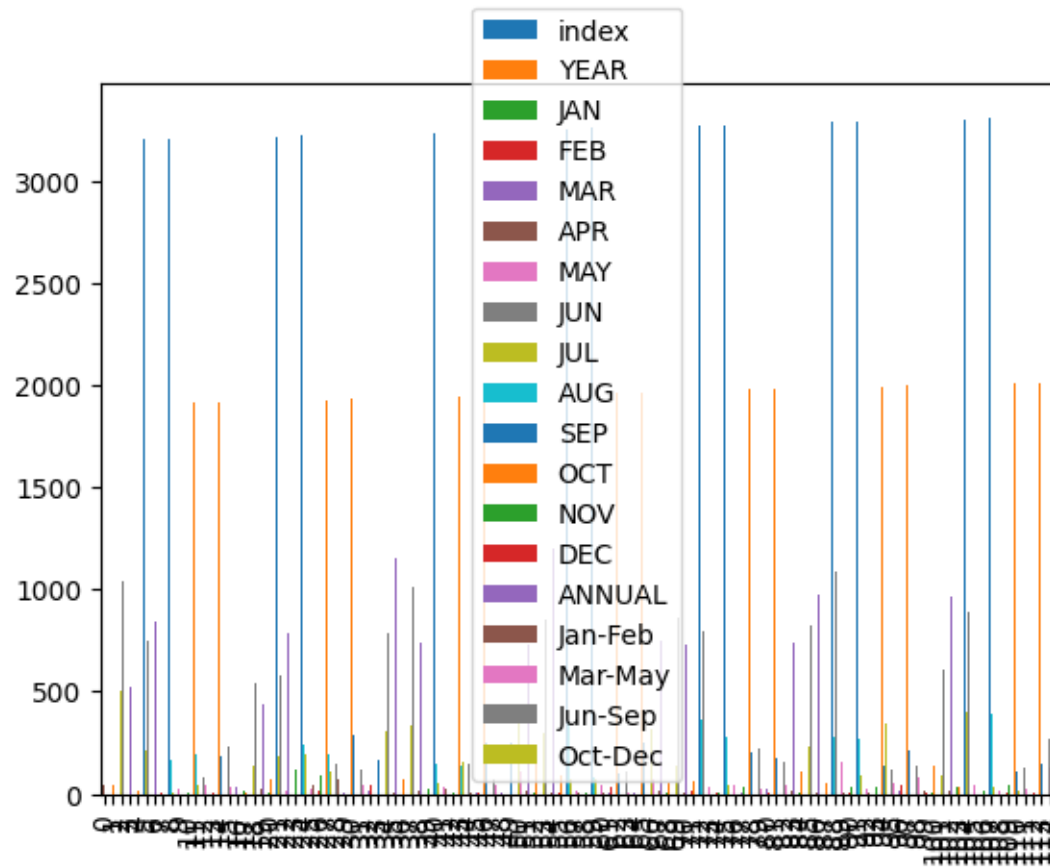
```
[7]: <Axes: >
```



6 Bar chart

```
[8]: df.plot.bar()
```

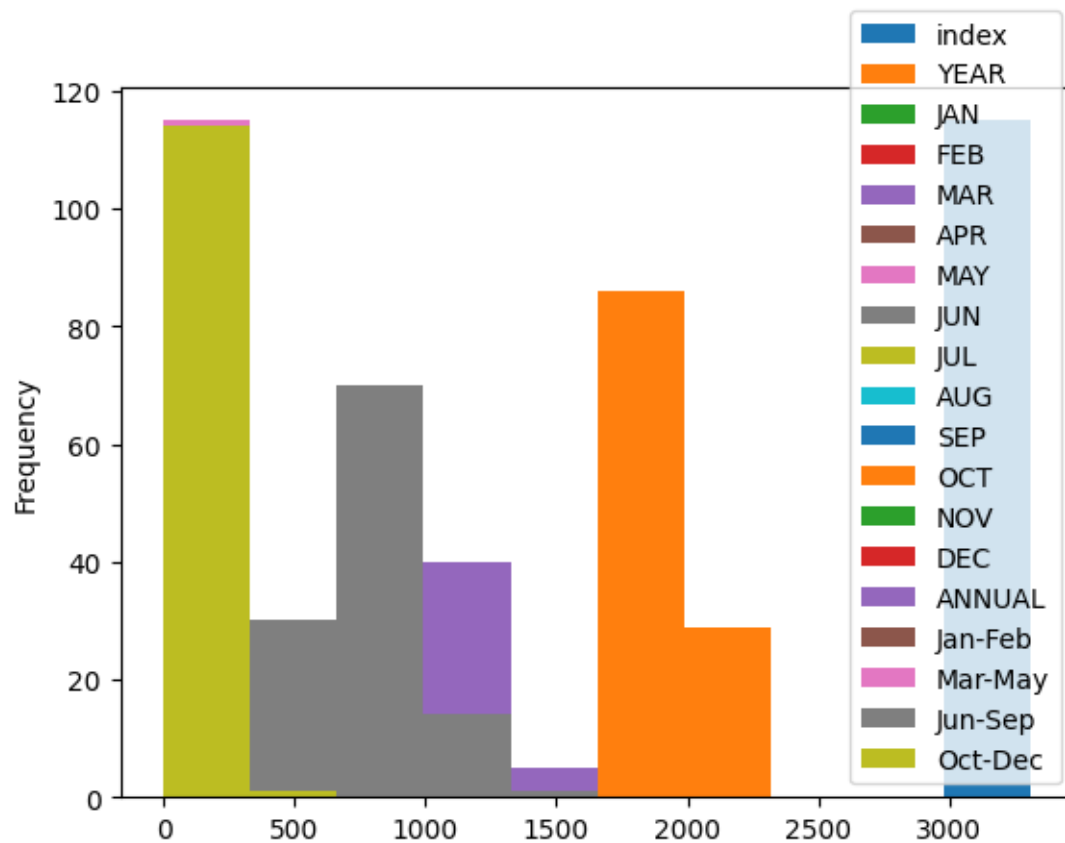
```
[8]: <Axes: >
```



7 Histogram

```
[9]: df.plot.hist()
```

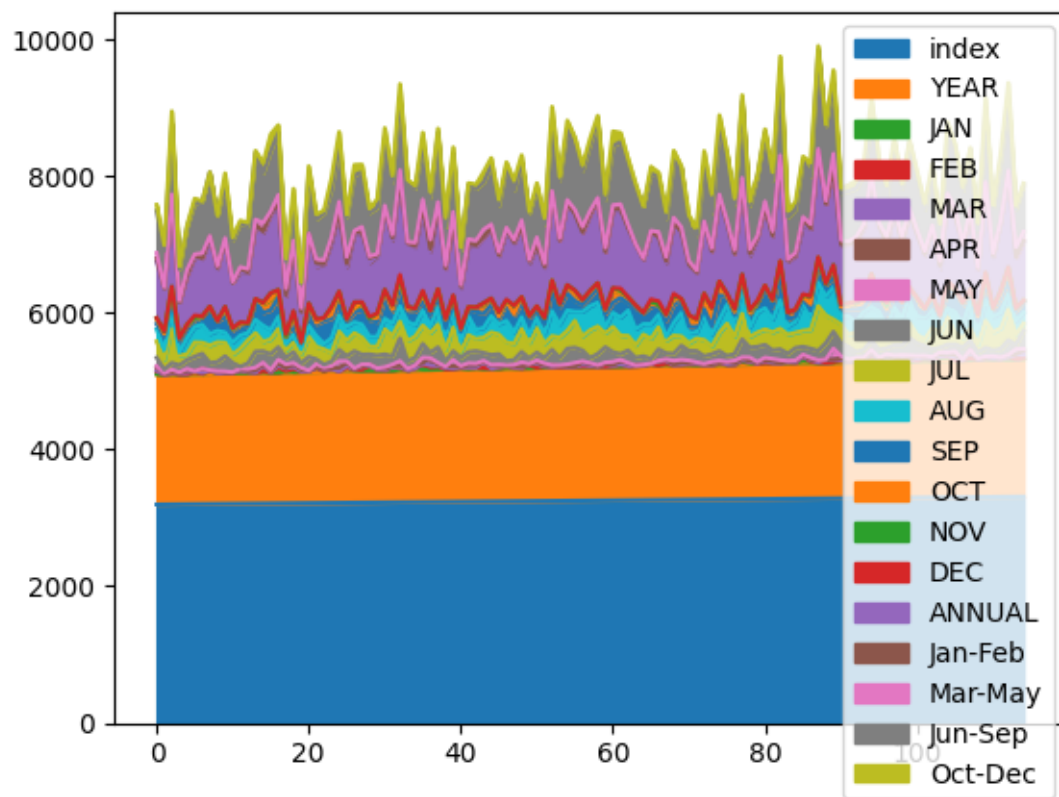
```
[9]: <Axes: ylabel='Frequency'>
```



8 Area chart

```
[10]: df.plot.area()
```

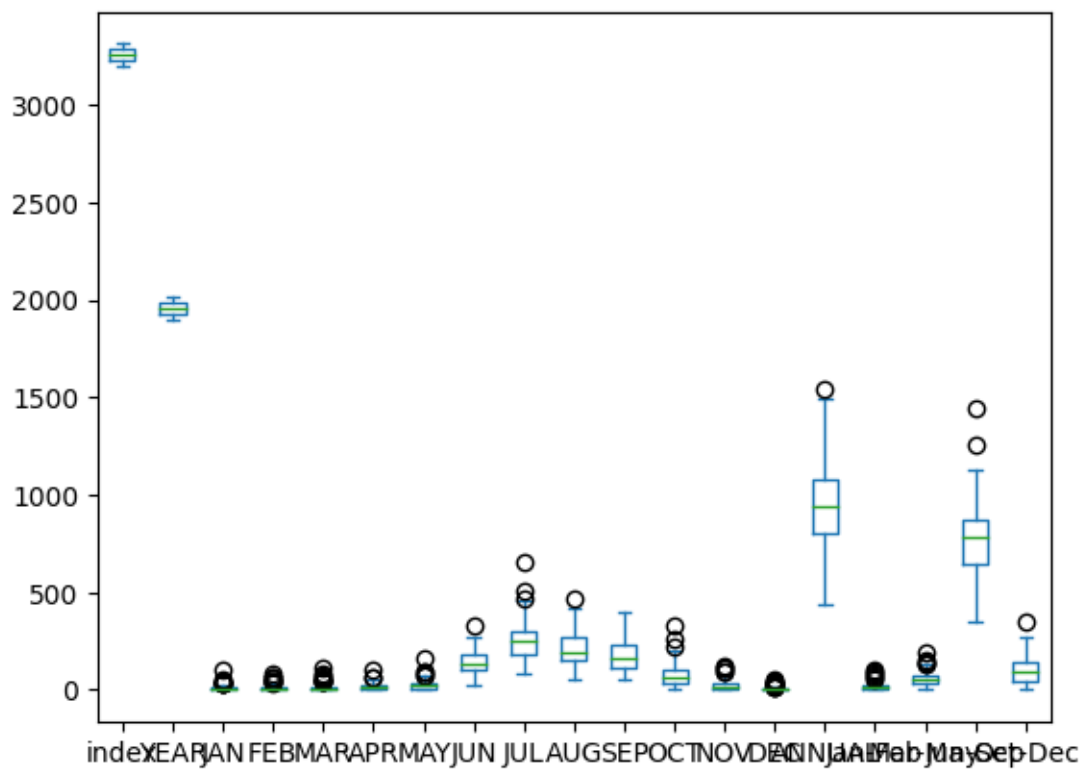
```
[10]: <Axes: >
```



9 Box chart

```
[11]: df.plot.box()
```

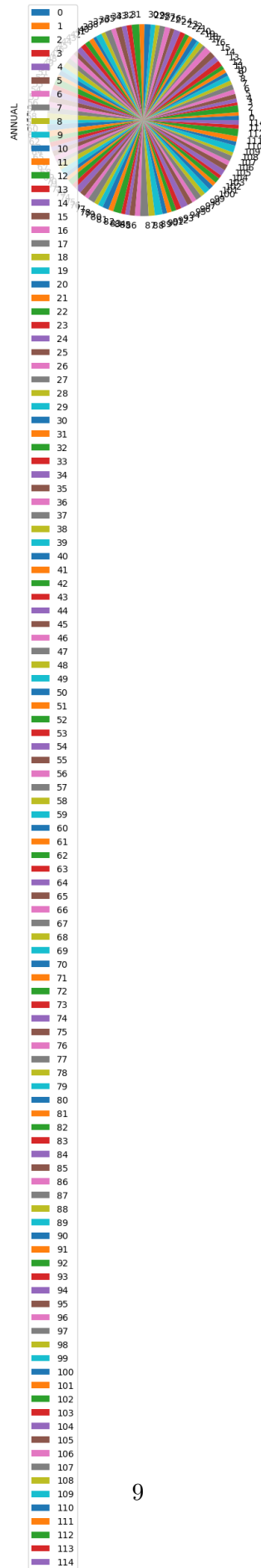
```
[11]: <Axes: >
```



10 Pie chart

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

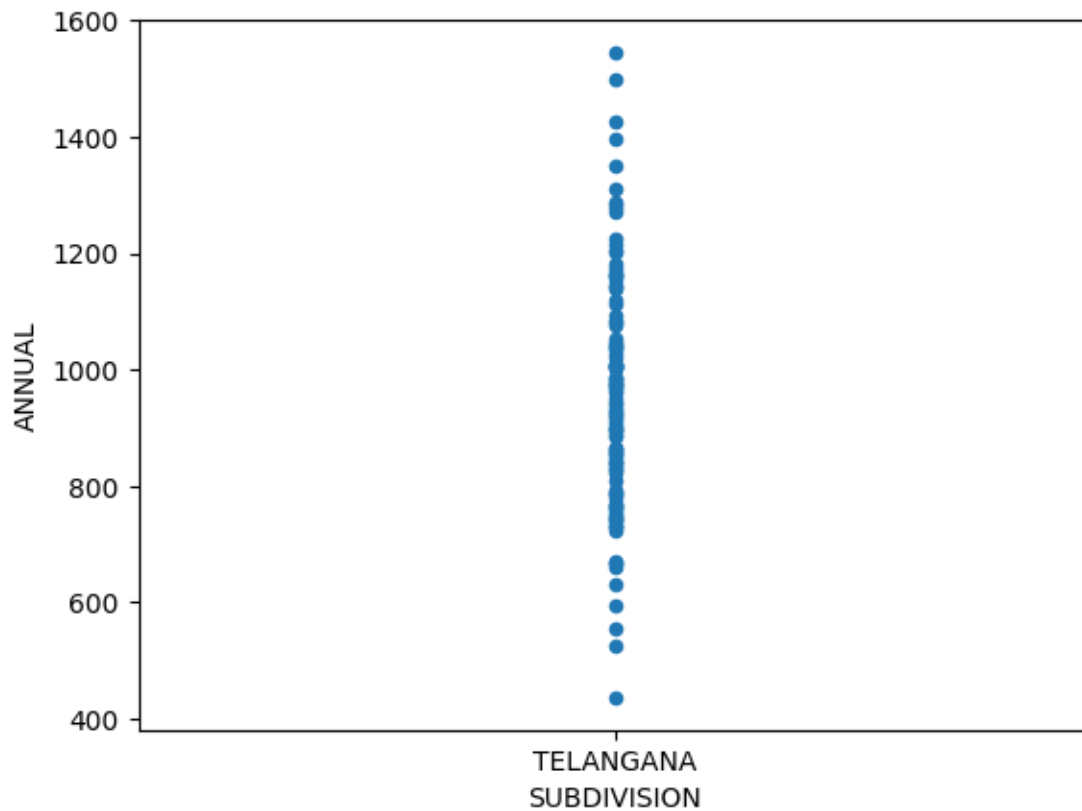
```
[12]: <Axes: ylabel='ANNUAL'>
```

11 Scatter chart

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

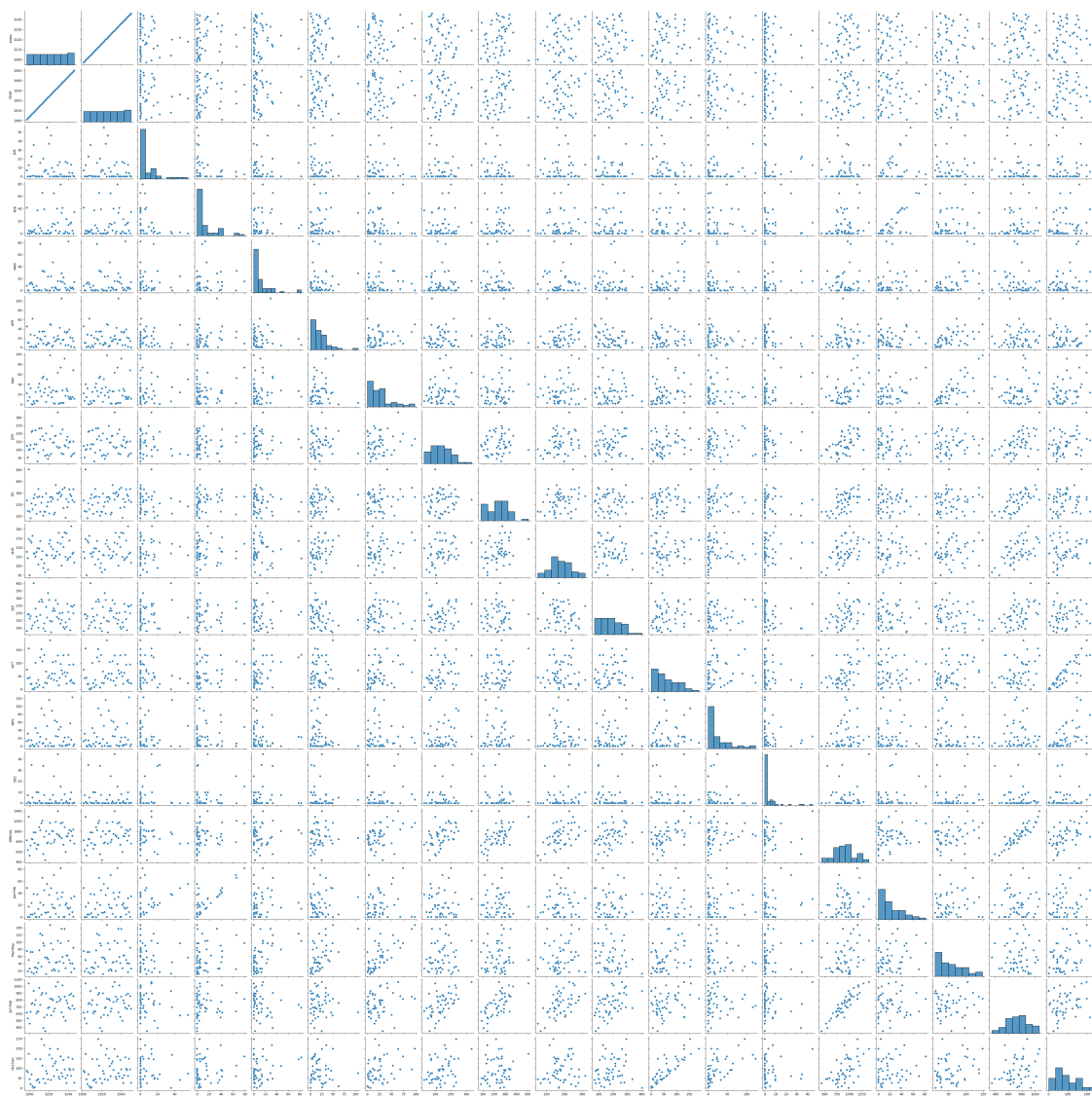
```
[13]: <Axes: xlabel='SUBDIVISION', ylabel='ANNUAL'>
```



12 Seaborn

```
[14]: sns.pairplot(df[0:50])
```

```
[14]: <seaborn.axisgrid.PairGrid at 0x7b9f51458af0>
```



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

<ipython-input-15-5daa97052ca5>:1: UserWarning:

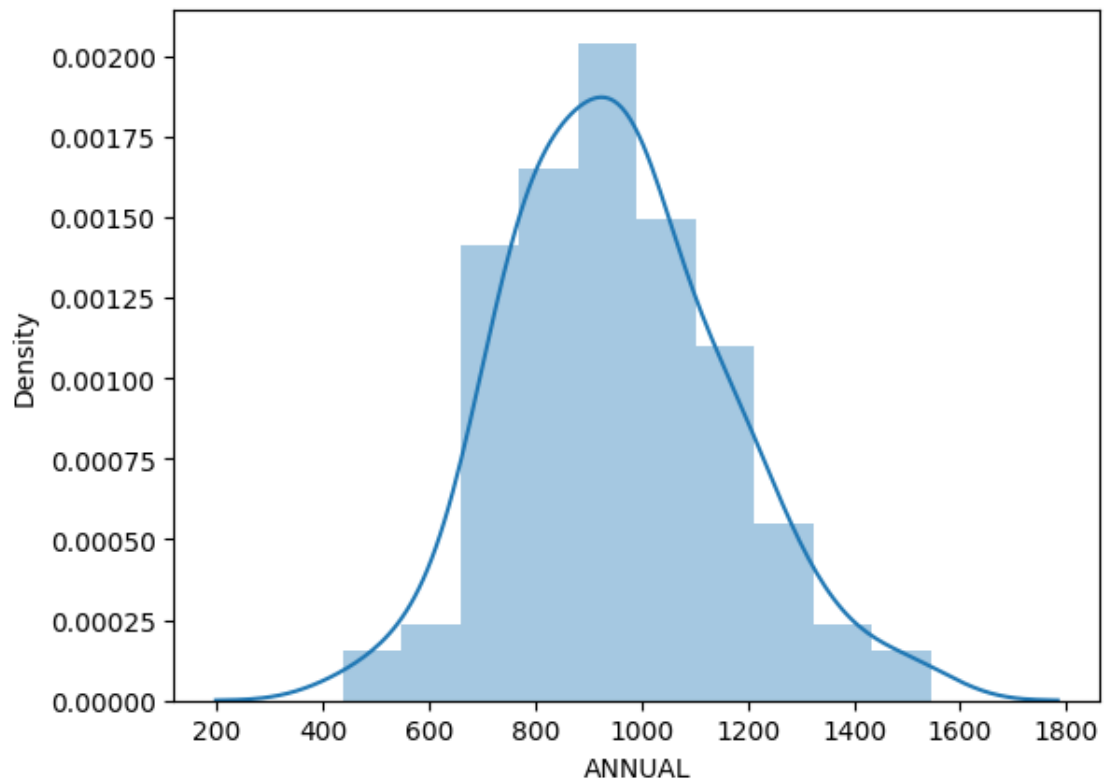
`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see <https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751>

```
sns.distplot(df['ANNUAL'])
```

```
[15]: <Axes: xlabel='ANNUAL', ylabel='Density'>
```



```
[16]: sns.heatmap(df.corr())
```

<ipython-input-16-aa4f4450a243>:1: FutureWarning: The default value of numeric_only in DataFrame.corr is deprecated. In a future version, it will default to False. Select only valid columns or specify the value of numeric_only to silence this warning.

```
sns.heatmap(df.corr())
```

```
[16]: <Axes: >
```

