

# 20104016

## DEENA

### 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_telangana.csv")
df
```

Out[2]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT
0	3197	TELANGANA	1901	6.9	41.8	7.8	45.2	22.0	123.6	237.8	177.2	77.7	75.5
1	3198	TELANGANA	1902	0.0	0.0	0.2	10.7	7.3	52.4	146.3	142.8	190.5	41.7
2	3199	TELANGANA	1903	12.9	4.6	0.0	9.9	40.7	99.2	505.2	246.7	191.9	155.8
3	3200	TELANGANA	1904	0.0	0.0	10.8	0.8	14.7	104.2	139.5	50.0	162.3	44.4
4	3201	TELANGANA	1905	0.0	4.3	12.8	27.6	32.2	129.5	82.4	237.3	179.1	19.6
...	...	...	...	...	...	...	...	...	...	...	...	...	...
110	3307	TELANGANA	2011	0.0	11.9	2.6	25.6	9.3	83.9	268.2	225.9	107.6	13.9
111	3308	TELANGANA	2012	6.7	0.0	0.2	14.0	8.4	124.4	300.3	229.9	202.4	83.6
112	3309	TELANGANA	2013	2.4	29.0	0.2	24.4	8.5	213.4	453.8	230.6	161.4	205.9
113	3310	TELANGANA	2014	0.2	2.9	58.3	10.3	73.3	62.3	146.0	205.2	146.8	29.6
114	3311	TELANGANA	2015	17.5	0.0	43.0	65.7	23.3	266.9	104.4	160.5	158.3	15.6

115 rows × 14 columns

### Data Cleaning and Data Preprocessing

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

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

```
Out[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')
```

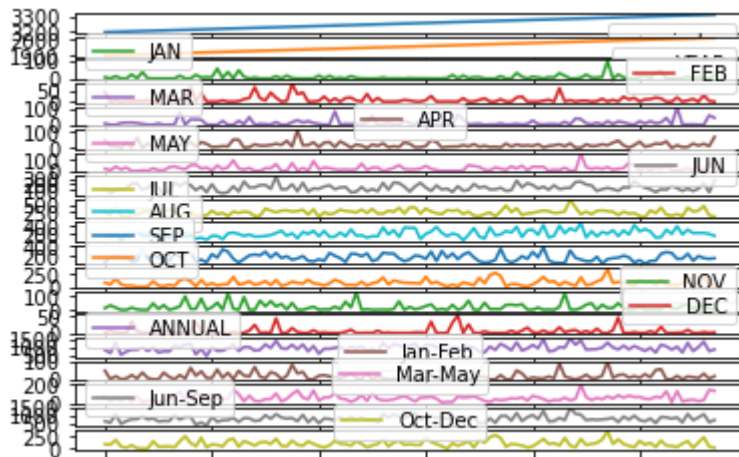
```
In [5]: df.info()
```

```
<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
```

## Line chart

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

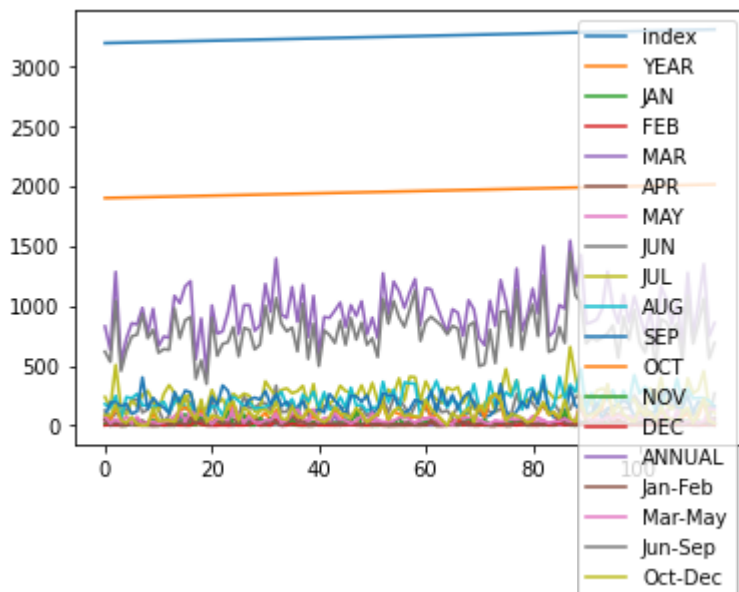
Out[6]: array([<AxesSubplot:~>, <AxesSubplot:~>, <AxesSubplot:~>, <AxesSubplot:~>, <AxesSubplot:~>, <AxesSubplot:~>, <AxesSubplot:~>, <AxesSubplot:~>, <AxesSubplot:~>, <AxesSubplot:~>, <AxesSubplot:~>, <AxesSubplot:~>, <AxesSubplot:~>, <AxesSubplot:~>, <AxesSubplot:~>], dtype=object)



## Line chart

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

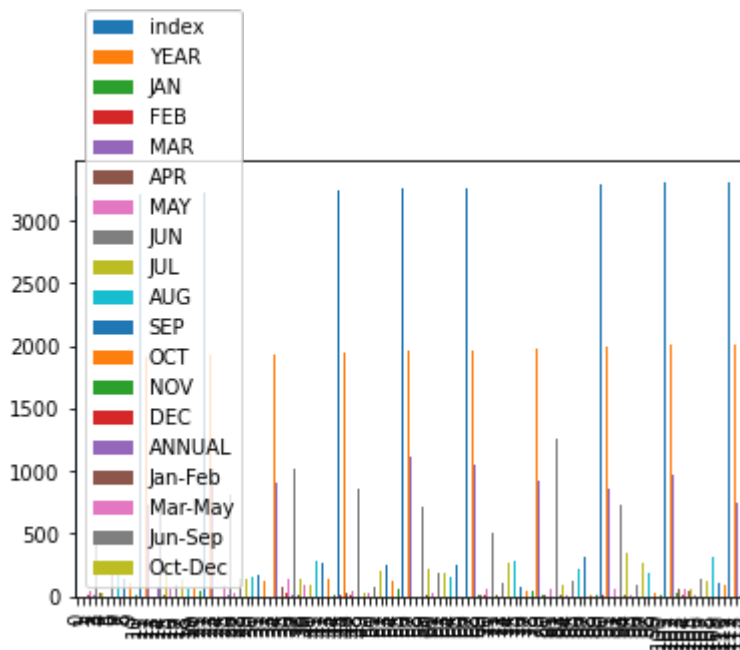
Out[7]: <AxesSubplot:~>



## Bar chart

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

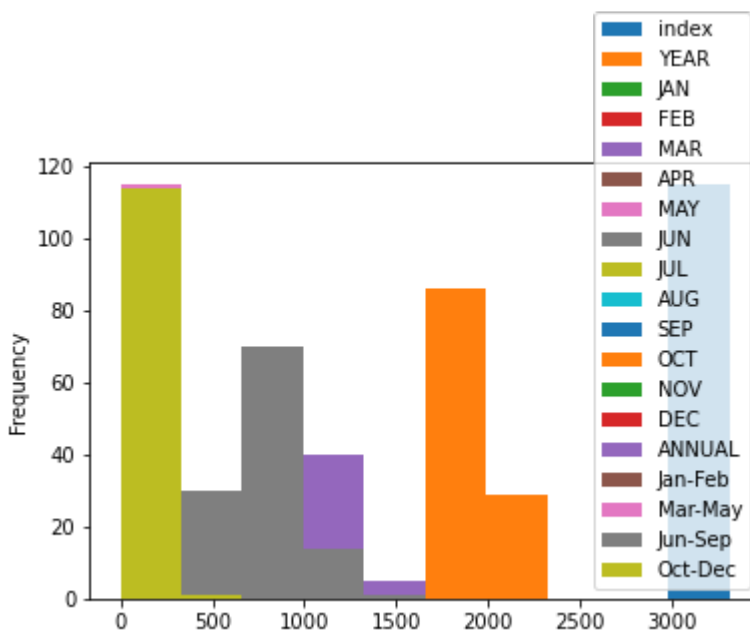
Out[8]: `<AxesSubplot:>`



## Histogram

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

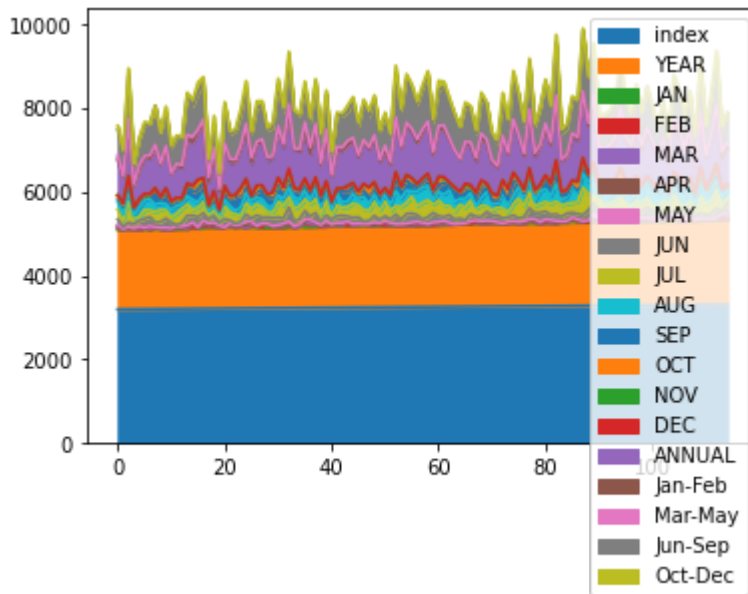
Out[9]: `<AxesSubplot:ylabel='Frequency'>`



## Area chart

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

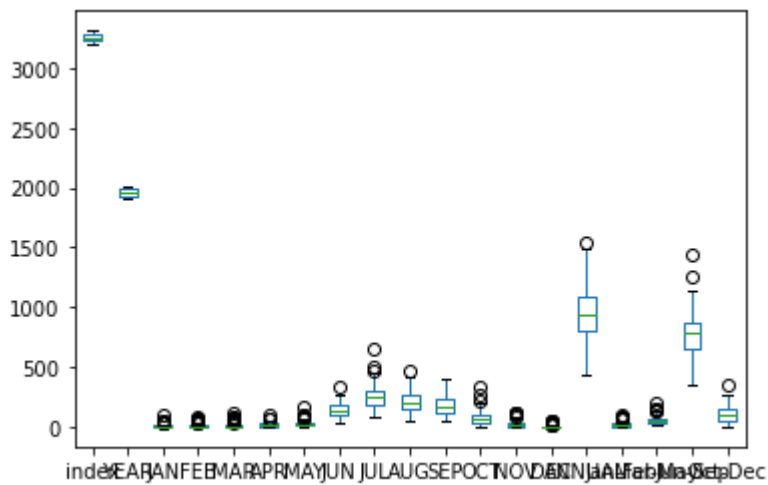
Out[10]: `<AxesSubplot:>`



## Box chart

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

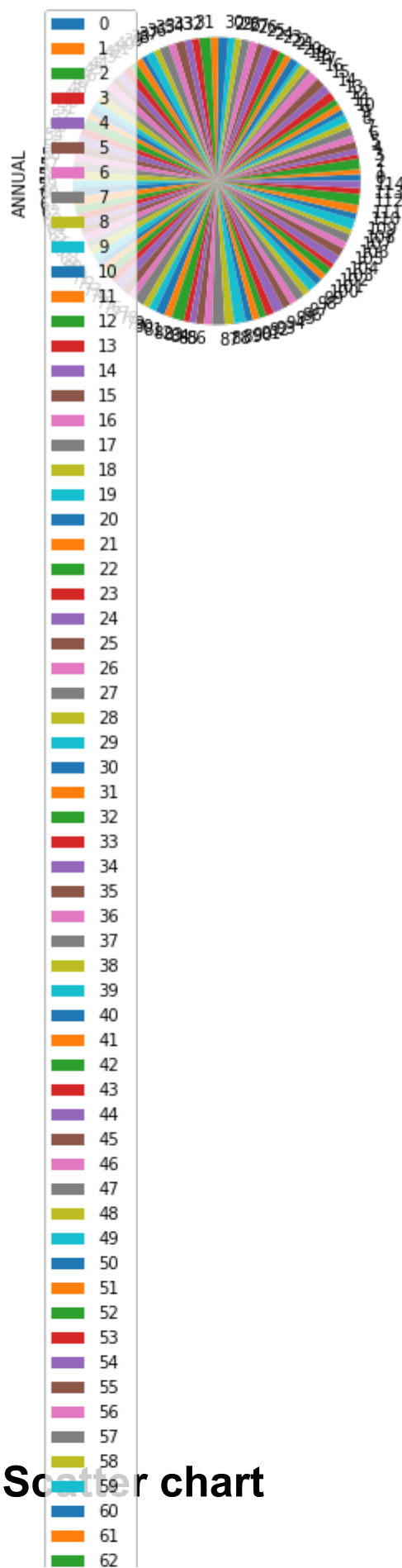
Out[11]: `<AxesSubplot:>`



## Pie chart

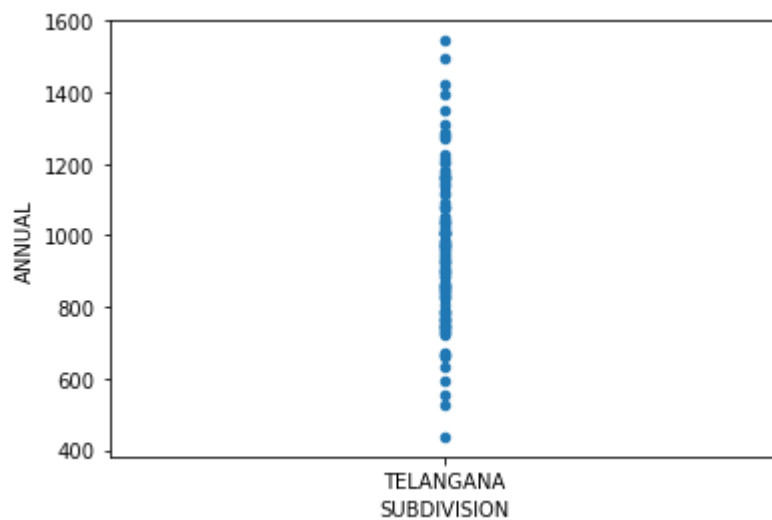
In [12]: `df.plot(figsize=(10, 5), color='r', label='ANNUAL')`

Out[12]: `<AxesSubplot:ylabel='ANNUAL'>`



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

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



```
In [14]: 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
```



In [15]: `df.describe()`

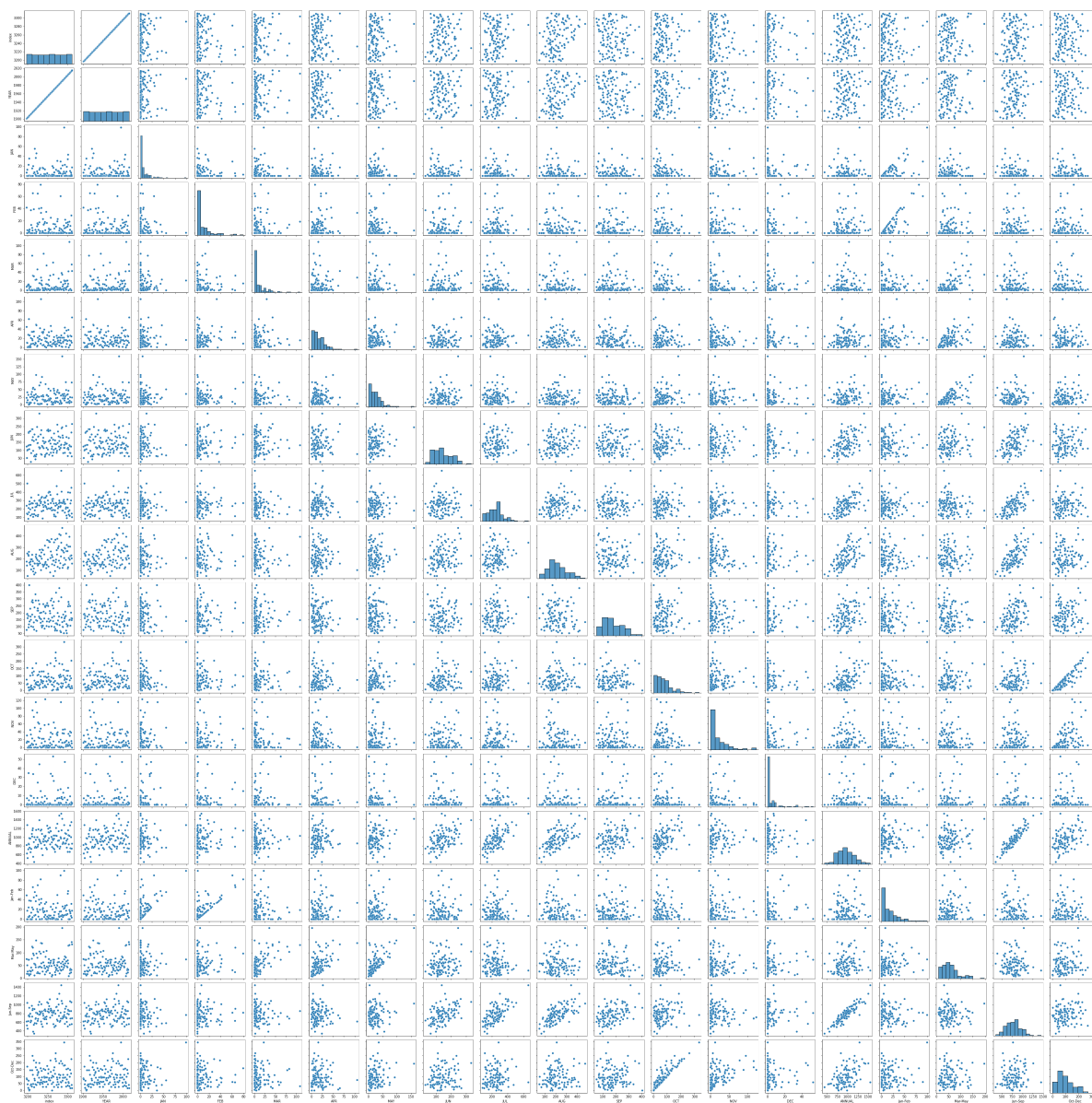
Out[15]:

	index	YEAR	JAN	FEB	MAR	APR	MAY	
<b>count</b>	115.000000	115.000000	115.000000	115.000000	115.000000	115.000000	115.000000	11
<b>mean</b>	3254.000000	1958.000000	7.702609	9.688696	12.614783	18.185217	25.373913	14
<b>std</b>	33.341666	33.341666	13.881222	15.223808	18.781819	15.931307	23.642777	5
<b>min</b>	3197.000000	1901.000000	0.000000	0.000000	0.000000	0.100000	0.200000	2
<b>25%</b>	3225.500000	1929.500000	0.000000	0.000000	1.400000	6.950000	8.450000	10
<b>50%</b>	3254.000000	1958.000000	1.000000	3.400000	4.700000	14.000000	20.600000	13
<b>75%</b>	3282.500000	1986.500000	9.700000	13.900000	15.250000	24.850000	34.500000	18
<b>max</b>	3311.000000	2015.000000	98.700000	79.100000	108.600000	105.600000	159.800000	33

## EDA AND VISUALIZATION

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

Out[16]: `<seaborn.axisgrid.PairGrid at 0x21b5eddaa00>`

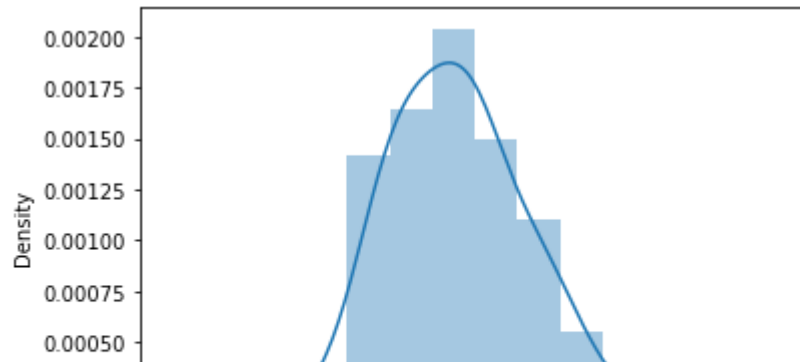


In [17]: `sns.distplot(df['ANNUAL'])`

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:>`

