# 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_coastal andhra pradesh.csv")
df
```

Out[2]:

| | index | SUBDIVISION | YEAR | JAN | FEB | MAR | APR | MAY | JUN | JUL | AUG | SEP | OCT |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 3082 | COASTAL ANDHRA PRADESH | 1901 | 18.8 | 80.9 | 7.2 | 28.7 | 68.7 | 77.7 | 113.0 | 133.7 | 125.3 | 173.4 |
| **1** | 3083 | COASTAL ANDHRA PRADESH | 1902 | 2.0 | 0.0 | 2.8 | 23.9 | 37.6 | 72.6 | 144.5 | 236.1 | 204.5 | 262.( |
| **2** | 3084 | COASTAL ANDHRA PRADESH | 1903 | 0.8 | 13.3 | 0.2 | 6.2 | 73.4 | 154.0 | 248.6 | 258.0 | 216.5 | 159.7 |
| **3** | 3085 | COASTAL ANDHRA PRADESH | 1904 | 1.3 | 0.0 | 5.4 | 3.0 | 136.3 | 107.8 | 120.2 | 117.7 | 116.8 | 240.9 |
| **4** | 3086 | COASTAL ANDHRA PRADESH | 1905 | 1.1 | 16.7 | 68.0 | 37.0 | 68.8 | 84.4 | 64.6 | 210.8 | 170.2 | 66.( |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| **110** | 3192 | COASTAL ANDHRA PRADESH | 2011 | 0.0 | 17.9 | 0.9 | 62.3 | 67.9 | 86.8 | 196.0 | 215.8 | 129.7 | 74.( |
| **111** | 3193 | COASTAL ANDHRA PRADESH | 2012 | 37.6 | 0.0 | 2.7 | 24.0 | 39.3 | 95.4 | 221.9 | 221.2 | 246.5 | 140.( |
| **112** | 3194 | COASTAL ANDHRA PRADESH | 2013 | 2.0 | 29.6 | 0.2 | 48.0 | 28.2 | 127.5 | 162.4 | 123.1 | 132.0 | 411.5 |
| **113** | 3195 | COASTAL ANDHRA PRADESH | 2014 | 0.4 | 1.2 | 9.1 | 6.0 | 112.9 | 45.7 | 151.8 | 177.8 | 144.5 | 195.( |
| **114** | 3196 | COASTAL ANDHRA PRADESH | 2015 | 2.0 | 0.6 | 5.5 | 32.3 | 34.1 | 283.8 | 116.0 | 192.0 | 201.8 | 59.7 |

115 rows × 20 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')
```

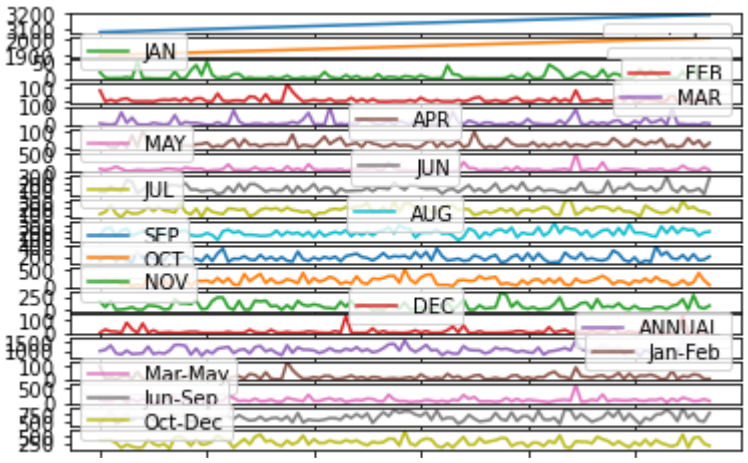
In [5]:
```
df.info()
```

```
<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)
```

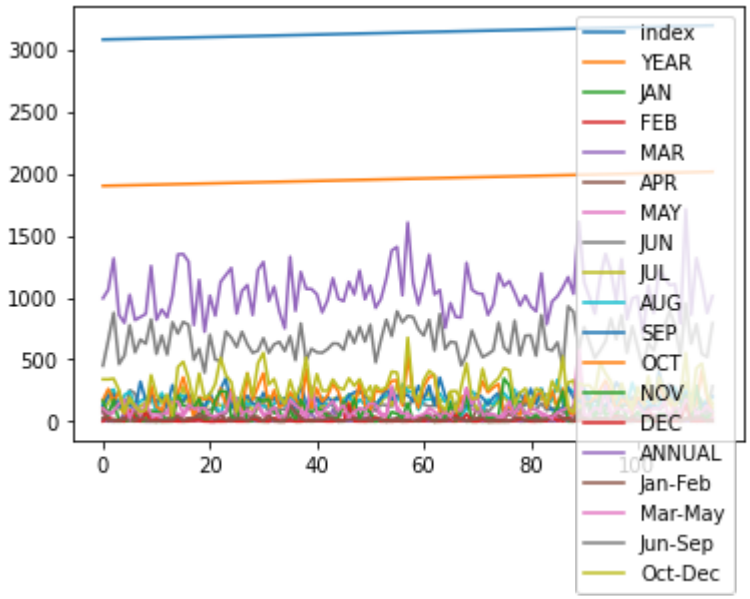
Out[6]:
```
array([<AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>,
       <AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>,
       <AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>,
       <AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>,
    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



# Line chart

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

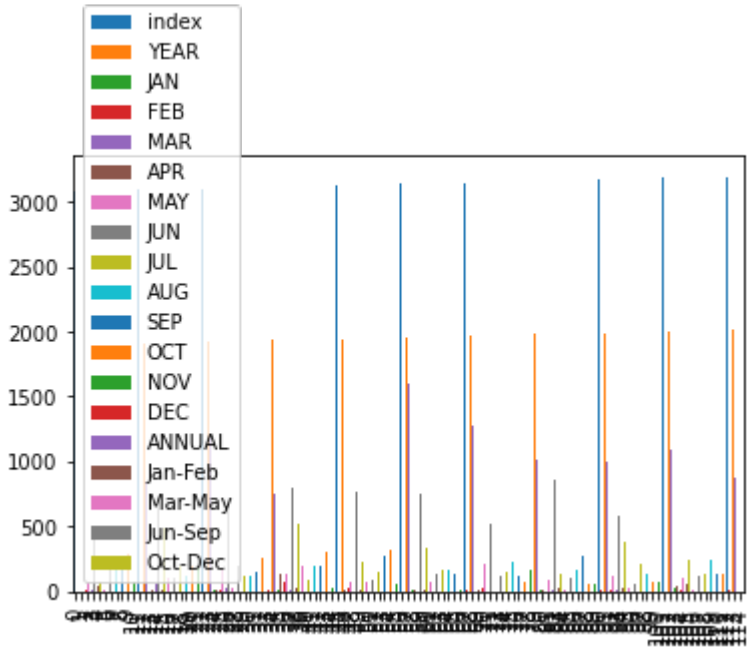
Out[7]: <AxesSubplot:>




# Bar chart

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

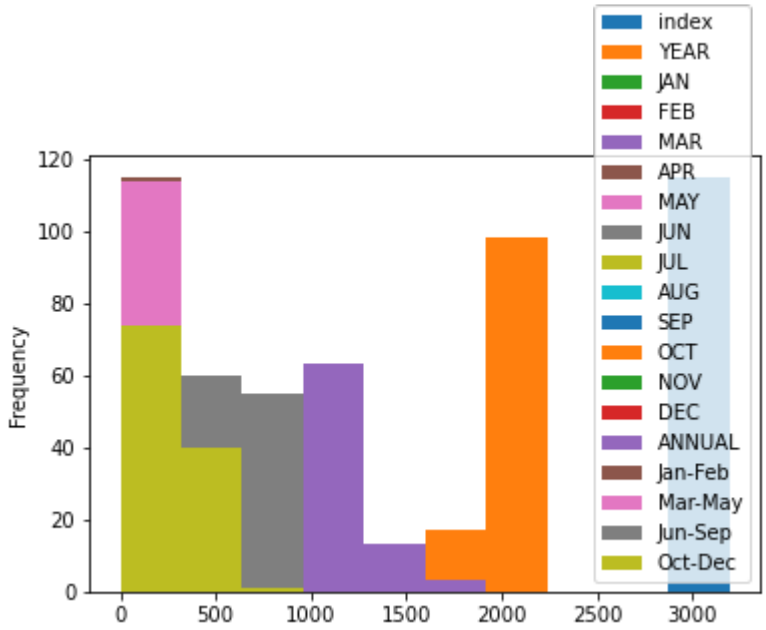
Out[8]: <AxesSubplot:>



# Histogram

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

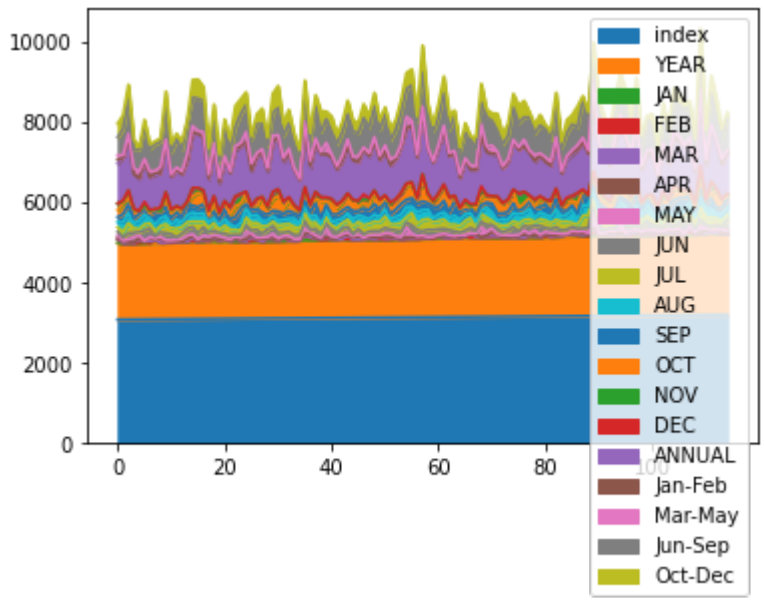
Out[9]: <AxesSubplot:ylabel='Frequency'>




# Area chart

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

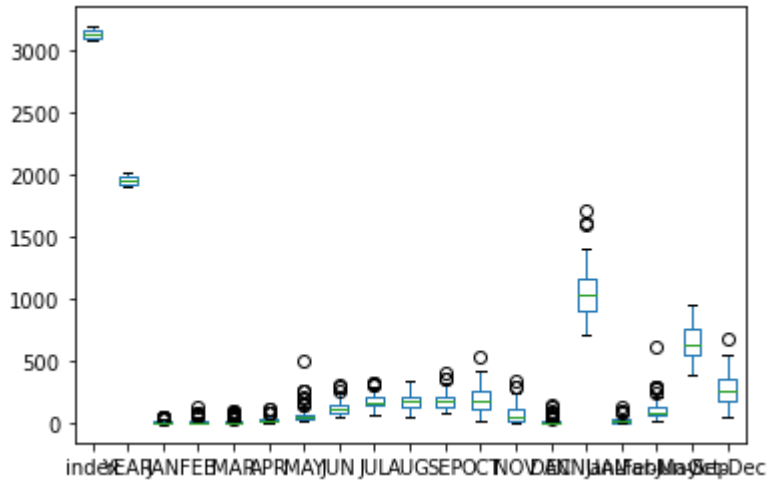
Out[10]: <AxesSubplot:>



# Box chart

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

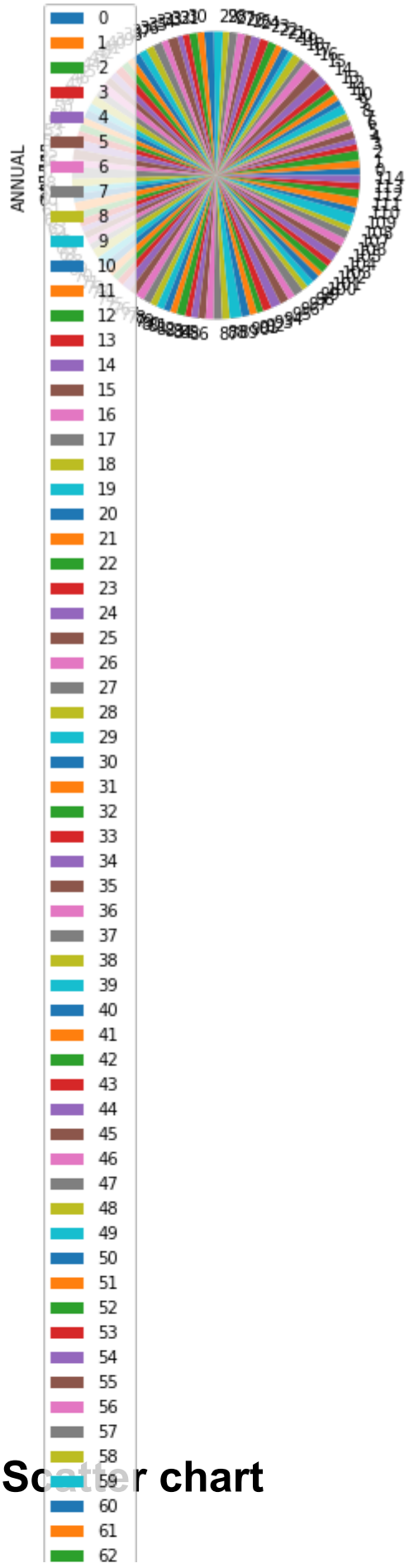
Out[11]: <AxesSubplot:>




# Pie chart

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

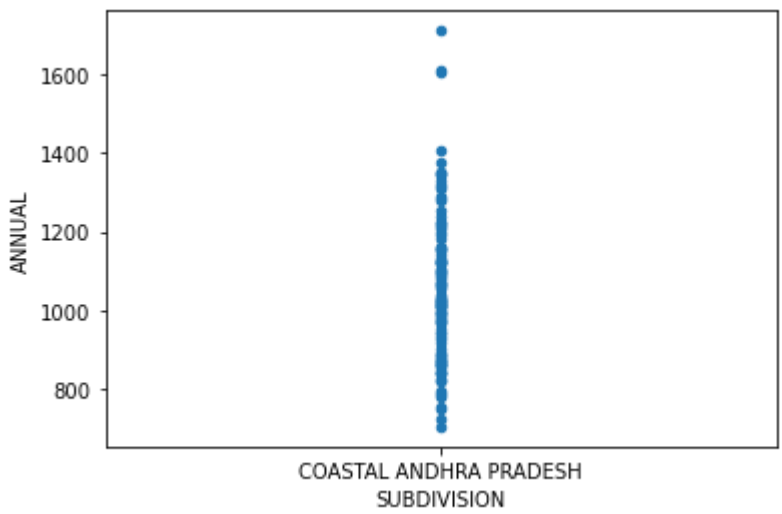
Out[12]:
```
<AxesSubplot:ylabel='ANNUAL'>
```




## Scatter chart

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()
```

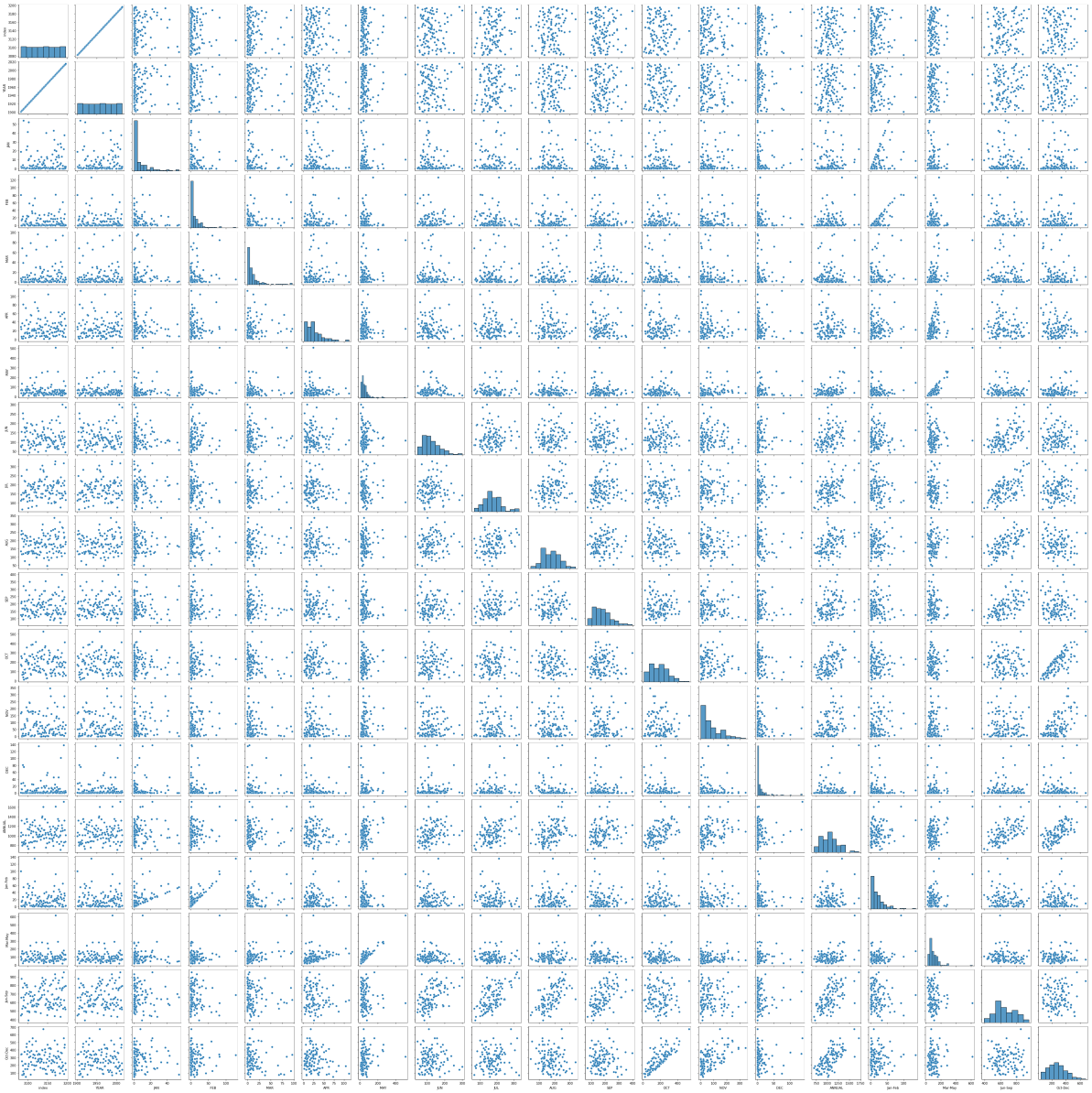
Out[15]:

| | index | YEAR | JAN | FEB | MAR | APR | MAY | |
|---|---|---|---|---|---|---|---|---|
| **count** | 115.000000 | 115.000000 | 115.000000 | 115.000000 | 115.000000 | 115.000000 | 115.000000 | 11 |
| **mean** | 3139.000000 | 1958.000000 | 7.483478 | 12.923478 | 13.221739 | 26.740870 | 62.549565 | 12 |
| **std** | 33.341666 | 33.341666 | 11.524748 | 19.997058 | 20.036216 | 21.310873 | 63.719734 | 5 |
| **min** | 3082.000000 | 1901.000000 | 0.000000 | 0.000000 | 0.000000 | 1.100000 | 10.500000 | 4 |
| **25%** | 3110.500000 | 1929.500000 | 0.200000 | 0.450000 | 1.550000 | 12.950000 | 31.350000 | 8 |
| **50%** | 3139.000000 | 1958.000000 | 2.000000 | 5.100000 | 5.700000 | 21.500000 | 44.400000 | 11 |
| **75%** | 3167.500000 | 1986.500000 | 10.300000 | 17.400000 | 14.250000 | 35.050000 | 69.800000 | 15 |
| **max** | 3196.000000 | 2015.000000 | 54.100000 | 127.100000 | 96.600000 | 112.200000 | 507.700000 | 30 |

# EDA AND VISUALIZATION

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

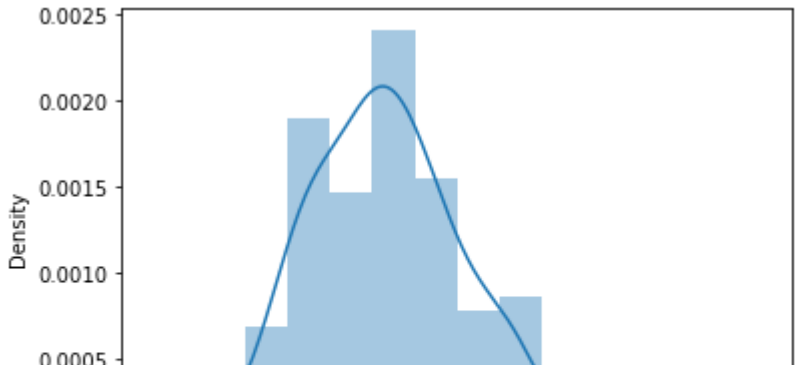
Out[16]: <seaborn.axisgrid.PairGrid at 0x1d80b3bf880>



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

```
C:\ProgramData\Anaconda3\lib\site-packages\seaborn\distributions.py:2557: Fut
ureWarning: `distplot` is a deprecated function and will be removed in a futu
re version. Please adapt your code to use either `displot` (a figure-level fu
nction with similar flexibility) or `histplot` (an axes-level function for hi
stograms).
  warnings.warn(msg, FutureWarning)
```

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




In [18]:
```
sns.heatmap(df.corr())
```

Out[18]: <AxesSubplot:>