

# FINAL ASSESSMENT 2

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

```
In [2]: #importing dataset
data=pd.read_csv(r"C:\Users\user\Downloads\rainfall in india 1901-2015.csv")
data
```

Out[2]:

|      | index | SUBDIVISION               | YEAR | JAN  | FEB   | MAR  | APR   | MAY   | JUN   | JUL   | AUG   | SEP   |   |
|------|-------|---------------------------|------|------|-------|------|-------|-------|-------|-------|-------|-------|---|
| 0    | 0     | ANDAMAN & NICOBAR ISLANDS | 1901 | 49.2 | 87.1  | 29.2 | 2.3   | 528.8 | 517.5 | 365.1 | 481.1 | 332.6 | : |
| 1    | 1     | ANDAMAN & NICOBAR ISLANDS | 1902 | 0.0  | 159.8 | 12.2 | 0.0   | 446.1 | 537.1 | 228.9 | 753.7 | 666.2 | : |
| 2    | 2     | ANDAMAN & NICOBAR ISLANDS | 1903 | 12.7 | 144.0 | 0.0  | 1.0   | 235.1 | 479.9 | 728.4 | 326.7 | 339.0 | : |
| 3    | 3     | ANDAMAN & NICOBAR ISLANDS | 1904 | 9.4  | 14.7  | 0.0  | 202.4 | 304.5 | 495.1 | 502.0 | 160.1 | 820.4 | : |
| 4    | 4     | ANDAMAN & NICOBAR ISLANDS | 1905 | 1.3  | 0.0   | 3.3  | 26.9  | 279.5 | 628.7 | 368.7 | 330.5 | 297.0 | : |
| ...  | ...   | ...                       | ...  | ...  | ...   | ...  | ...   | ...   | ...   | ...   | ...   | ...   |   |
| 4111 | 4111  | LAKSHADWEEP               | 2011 | 5.1  | 2.8   | 3.1  | 85.9  | 107.2 | 153.6 | 350.2 | 254.0 | 255.2 |   |
| 4112 | 4112  | LAKSHADWEEP               | 2012 | 19.2 | 0.1   | 1.6  | 76.8  | 21.2  | 327.0 | 231.5 | 381.2 | 179.8 | : |
| 4113 | 4113  | LAKSHADWEEP               | 2013 | 26.2 | 34.4  | 37.5 | 5.3   | 88.3  | 426.2 | 296.4 | 154.4 | 180.0 |   |
| 4114 | 4114  | LAKSHADWEEP               | 2014 | 53.2 | 16.1  | 4.4  | 14.9  | 57.4  | 244.1 | 116.1 | 466.1 | 132.2 | : |
| 4115 | 4115  | LAKSHADWEEP               | 2015 | 2.2  | 0.5   | 3.7  | 87.1  | 133.1 | 296.6 | 257.5 | 146.4 | 160.4 | : |

4116 rows × 20 columns

# PUNJAB

```
In [3]: df=data.iloc[1472:1587]
df
```

Out[3]:

|      | index | SUBDIVISION | YEAR | JAN  | FEB  | MAR  | APR  | MAY  | JUN   | JUL   | AUG   | SEP   | OCT  |
|------|-------|-------------|------|------|------|------|------|------|-------|-------|-------|-------|------|
| 1472 | 1472  | PUNJAB      | 1901 | 55.7 | 50.1 | 25.2 | 2.1  | 25.2 | 10.4  | 178.2 | 145.0 | 24.4  | 3.7  |
| 1473 | 1473  | PUNJAB      | 1902 | 0.0  | 0.8  | 9.9  | 10.9 | 29.6 | 49.9  | 125.6 | 94.9  | 67.2  | 9.0  |
| 1474 | 1474  | PUNJAB      | 1903 | 29.5 | 0.5  | 45.0 | 1.3  | 9.2  | 5.2   | 212.2 | 119.1 | 132.5 | 6.9  |
| 1475 | 1475  | PUNJAB      | 1904 | 24.2 | 1.7  | 87.8 | 1.2  | 13.8 | 22.0  | 59.9  | 124.0 | 73.8  | 7.4  |
| 1476 | 1476  | PUNJAB      | 1905 | 53.0 | 40.3 | 24.3 | 0.5  | 2.2  | 19.2  | 122.6 | 50.3  | 111.1 | 1.2  |
| ...  | ...   | ...         | ...  | ...  | ...  | ...  | ...  | ...  | ...   | ...   | ...   | ...   | ...  |
| 1582 | 1582  | PUNJAB      | 2011 | 3.5  | 35.6 | 8.2  | 17.8 | 18.9 | 162.9 | 120.9 | 193.5 | 140.2 | 0.0  |
| 1583 | 1583  | PUNJAB      | 2012 | 62.6 | 3.2  | 1.9  | 31.1 | 1.6  | 11.9  | 120.2 | 135.1 | 112.3 | 2.2  |
| 1584 | 1584  | PUNJAB      | 2013 | 9.3  | 50.1 | 11.6 | 3.4  | 3.6  | 120.3 | 117.9 | 217.1 | 24.4  | 16.2 |
| 1585 | 1585  | PUNJAB      | 2014 | 21.8 | 20.1 | 30.3 | 24.5 | 20.8 | 20.6  | 76.3  | 41.9  | 105.8 | 6.0  |
| 1586 | 1586  | PUNJAB      | 2015 | 17.7 | 31.3 | 68.5 | 29.8 | 16.7 | 48.3  | 130.2 | 88.6  | 69.2  | 9.0  |

115 rows × 20 columns

# Data Cleaning and Preprocessing

```
In [4]: df.head()
```

Out[4]:

|      | index | SUBDIVISION | YEAR | JAN  | FEB  | MAR  | APR  | MAY  | JUN  | JUL   | AUG   | SEP   | OCT |
|------|-------|-------------|------|------|------|------|------|------|------|-------|-------|-------|-----|
| 1472 | 1472  | PUNJAB      | 1901 | 55.7 | 50.1 | 25.2 | 2.1  | 25.2 | 10.4 | 178.2 | 145.0 | 24.4  | 3.7 |
| 1473 | 1473  | PUNJAB      | 1902 | 0.0  | 0.8  | 9.9  | 10.9 | 29.6 | 49.9 | 125.6 | 94.9  | 67.2  | 9.0 |
| 1474 | 1474  | PUNJAB      | 1903 | 29.5 | 0.5  | 45.0 | 1.3  | 9.2  | 5.2  | 212.2 | 119.1 | 132.5 | 6.9 |
| 1475 | 1475  | PUNJAB      | 1904 | 24.2 | 1.7  | 87.8 | 1.2  | 13.8 | 22.0 | 59.9  | 124.0 | 73.8  | 7.4 |
| 1476 | 1476  | PUNJAB      | 1905 | 53.0 | 40.3 | 24.3 | 0.5  | 2.2  | 19.2 | 122.6 | 50.3  | 111.1 | 1.2 |

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

Out[5]:

|             | index | SUBDIVISION | YEAR | JAN  | FEB  | MAR  | APR  | MAY  | JUN   | JUL   | AUG   | SEP   | OCT  |
|-------------|-------|-------------|------|------|------|------|------|------|-------|-------|-------|-------|------|
| <b>1582</b> | 1582  | PUNJAB      | 2011 | 3.5  | 35.6 | 8.2  | 17.8 | 18.9 | 162.9 | 120.9 | 193.5 | 140.2 | 0.0  |
| <b>1583</b> | 1583  | PUNJAB      | 2012 | 62.6 | 3.2  | 1.9  | 31.1 | 1.6  | 11.9  | 120.2 | 135.1 | 112.3 | 2.2  |
| <b>1584</b> | 1584  | PUNJAB      | 2013 | 9.3  | 50.1 | 11.6 | 3.4  | 3.6  | 120.3 | 117.9 | 217.1 | 24.4  | 16.2 |
| <b>1585</b> | 1585  | PUNJAB      | 2014 | 21.8 | 20.1 | 30.3 | 24.5 | 20.8 | 20.6  | 76.3  | 41.9  | 105.8 | 6.0  |
| <b>1586</b> | 1586  | PUNJAB      | 2015 | 17.7 | 31.3 | 68.5 | 29.8 | 16.7 | 48.3  | 130.2 | 88.6  | 69.2  | 9.0  |

In [6]: `df.info()`

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 115 entries, 1472 to 1586
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
```

```
In [7]: #filling null values
df1=df.fillna(0)
df1
```

Out[7]:

|             | index | SUBDIVISION | YEAR | JAN  | FEB  | MAR  | APR  | MAY  | JUN   | JUL   | AUG   | SEP   | OCT  |
|-------------|-------|-------------|------|------|------|------|------|------|-------|-------|-------|-------|------|
| <b>1472</b> | 1472  | PUNJAB      | 1901 | 55.7 | 50.1 | 25.2 | 2.1  | 25.2 | 10.4  | 178.2 | 145.0 | 24.4  | 3.7  |
| <b>1473</b> | 1473  | PUNJAB      | 1902 | 0.0  | 0.8  | 9.9  | 10.9 | 29.6 | 49.9  | 125.6 | 94.9  | 67.2  | 9.0  |
| <b>1474</b> | 1474  | PUNJAB      | 1903 | 29.5 | 0.5  | 45.0 | 1.3  | 9.2  | 5.2   | 212.2 | 119.1 | 132.5 | 6.9  |
| <b>1475</b> | 1475  | PUNJAB      | 1904 | 24.2 | 1.7  | 87.8 | 1.2  | 13.8 | 22.0  | 59.9  | 124.0 | 73.8  | 7.4  |
| <b>1476</b> | 1476  | PUNJAB      | 1905 | 53.0 | 40.3 | 24.3 | 0.5  | 2.2  | 19.2  | 122.6 | 50.3  | 111.1 | 1.2  |
| ...         | ...   | ...         | ...  | ...  | ...  | ...  | ...  | ...  | ...   | ...   | ...   | ...   | ...  |
| <b>1582</b> | 1582  | PUNJAB      | 2011 | 3.5  | 35.6 | 8.2  | 17.8 | 18.9 | 162.9 | 120.9 | 193.5 | 140.2 | 0.0  |
| <b>1583</b> | 1583  | PUNJAB      | 2012 | 62.6 | 3.2  | 1.9  | 31.1 | 1.6  | 11.9  | 120.2 | 135.1 | 112.3 | 2.2  |
| <b>1584</b> | 1584  | PUNJAB      | 2013 | 9.3  | 50.1 | 11.6 | 3.4  | 3.6  | 120.3 | 117.9 | 217.1 | 24.4  | 16.2 |
| <b>1585</b> | 1585  | PUNJAB      | 2014 | 21.8 | 20.1 | 30.3 | 24.5 | 20.8 | 20.6  | 76.3  | 41.9  | 105.8 | 6.0  |
| <b>1586</b> | 1586  | PUNJAB      | 2015 | 17.7 | 31.3 | 68.5 | 29.8 | 16.7 | 48.3  | 130.2 | 88.6  | 69.2  | 9.0  |

115 rows × 20 columns

```
In [8]: df1.describe()
```

Out[8]:

|              | index       | YEAR        | JAN        | FEB        | MAR        | APR        | MAY        |
|--------------|-------------|-------------|------------|------------|------------|------------|------------|
| <b>count</b> | 115.000000  | 115.000000  | 115.000000 | 115.000000 | 115.000000 | 115.000000 | 115.000000 |
| <b>mean</b>  | 1529.000000 | 1958.000000 | 25.246087  | 26.786957  | 23.651304  | 12.660000  | 14.136522  |
| <b>std</b>   | 33.341666   | 33.341666   | 22.306656  | 23.473612  | 22.890109  | 16.751778  | 15.185232  |
| <b>min</b>   | 1472.000000 | 1901.000000 | 0.000000   | 0.000000   | 0.000000   | 0.000000   | 0.100000   |
| <b>25%</b>   | 1500.500000 | 1929.500000 | 7.250000   | 5.650000   | 6.900000   | 2.550000   | 3.350000   |
| <b>50%</b>   | 1529.000000 | 1958.000000 | 21.600000  | 21.300000  | 15.800000  | 6.700000   | 9.200000   |
| <b>75%</b>   | 1557.500000 | 1986.500000 | 36.100000  | 40.600000  | 33.650000  | 15.700000  | 19.700000  |
| <b>max</b>   | 1586.000000 | 2015.000000 | 112.100000 | 96.000000  | 108.500000 | 113.200000 | 98.300000  |

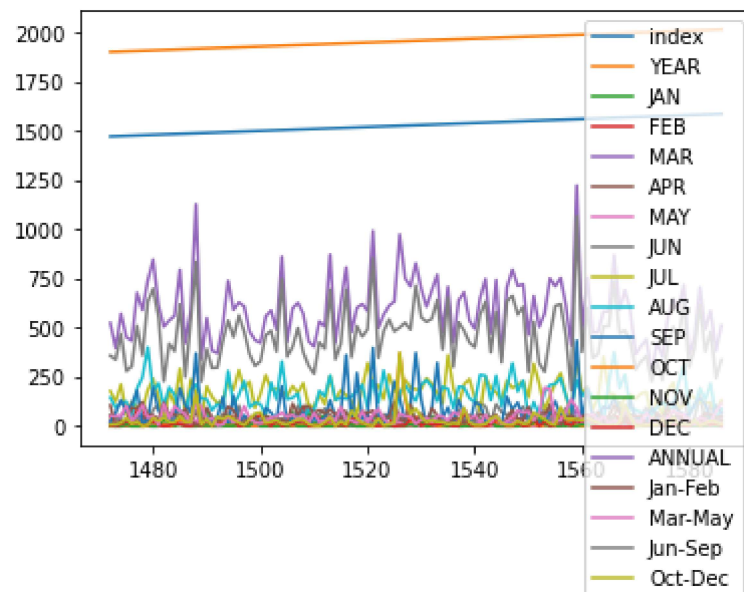
```
In [9]: df1.columns
```

Out[9]: 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')

# Data Visulaization

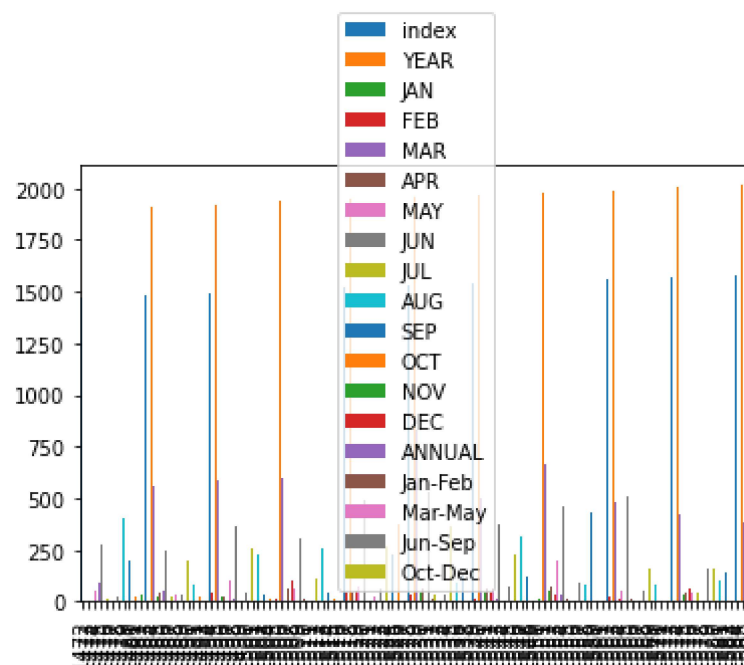
```
In [10]: df1.plot.line()
```

```
Out[10]: <AxesSubplot:>
```



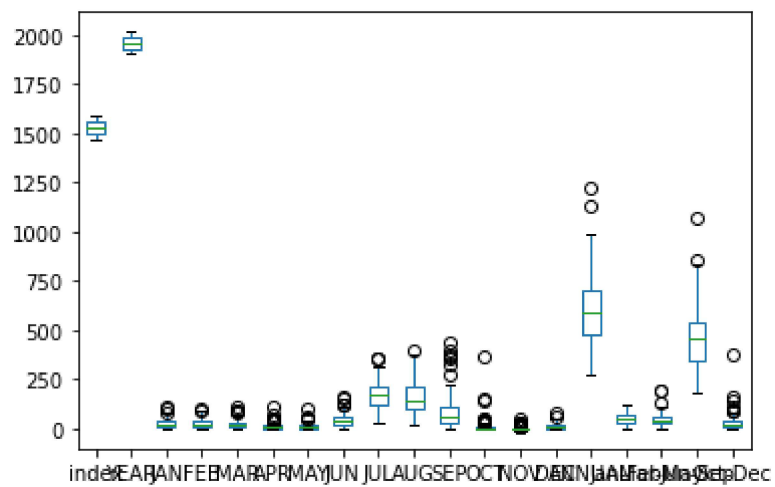
```
In [11]: df1.plot.bar()
```

```
Out[11]: <AxesSubplot:>
```



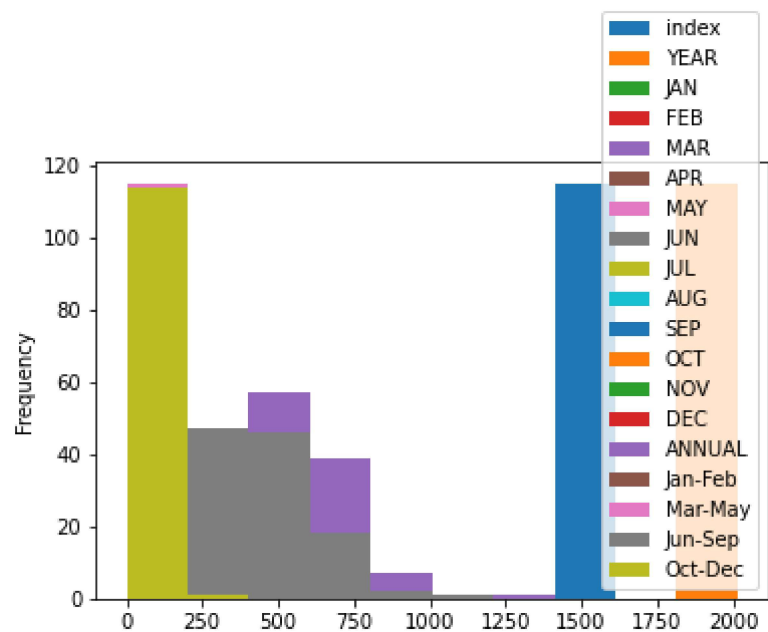
In [12]: `df1.plot.box()`

Out[12]: `<AxesSubplot:>`



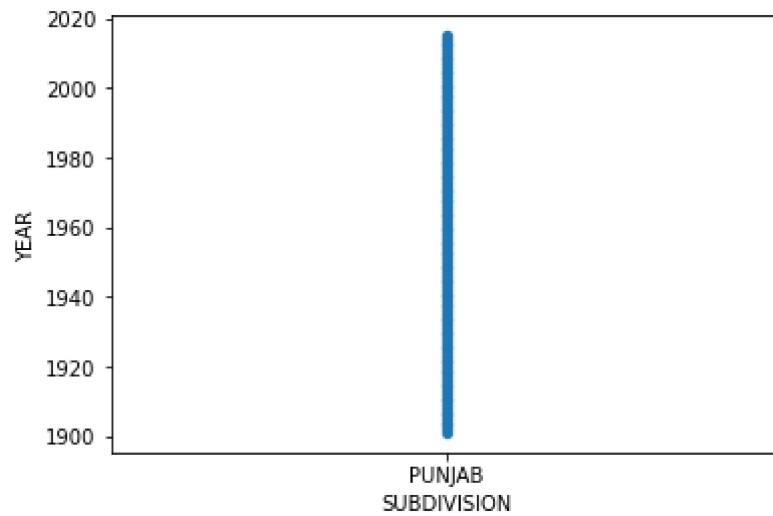
In [13]: `df1.plot.hist()`

Out[13]: `<AxesSubplot:ylabel='Frequency'>`



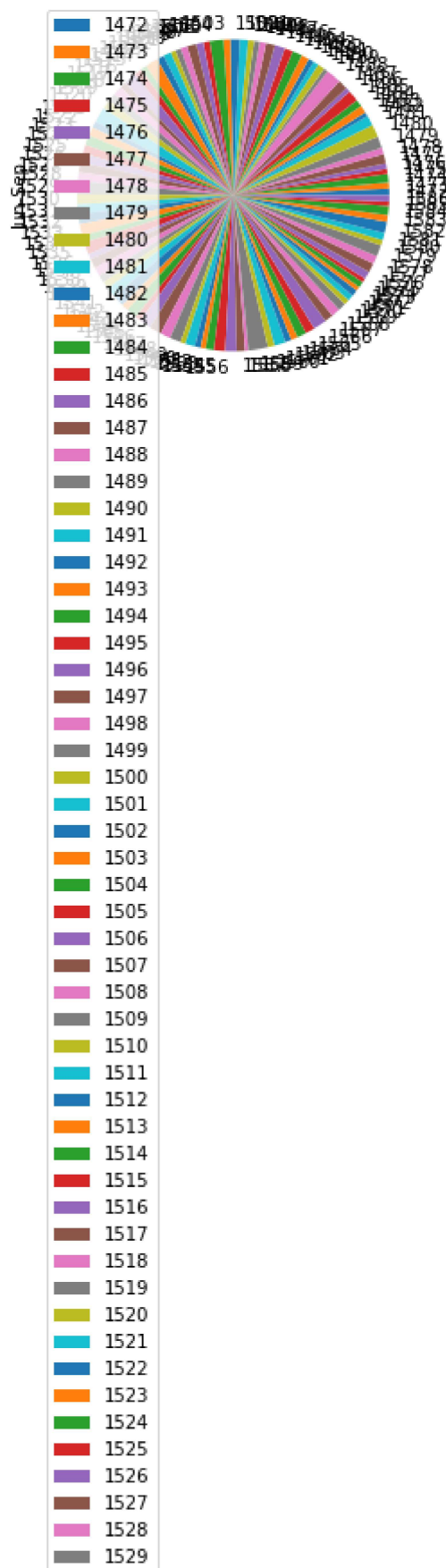
```
In [14]: df1.plot.scatter(x="SUBDIVISION",y="YEAR")
```

```
Out[14]: <AxesSubplot:xlabel='SUBDIVISION', ylabel='YEAR'>
```



```
In [15]: df2=df1[[ 'Jun-Sep']]  
df2.plot.pie(subplots=True)
```

```
Out[15]: array([<AxesSubplot:ylabel='Jun-Sep'>], dtype=object)
```

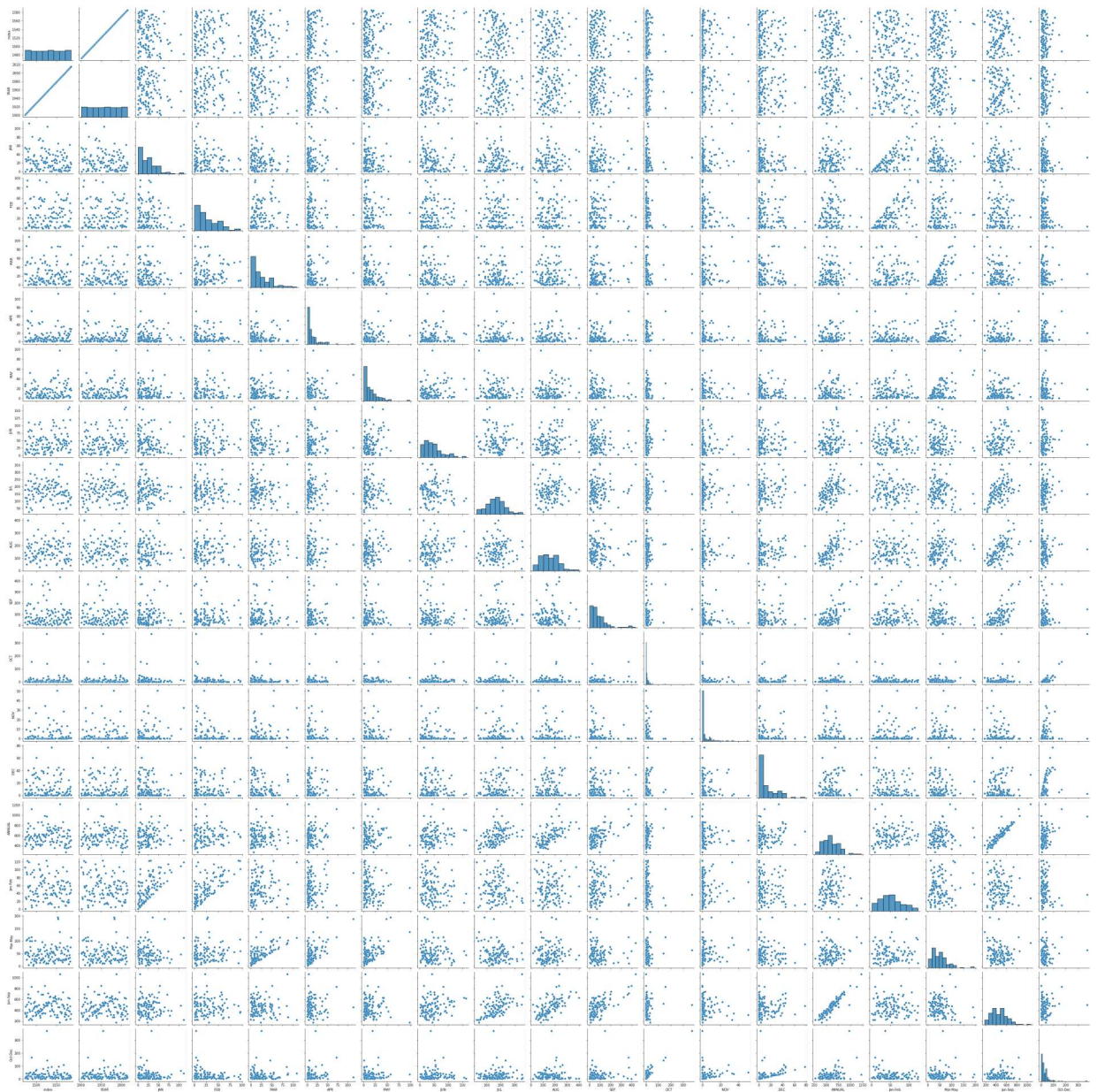




|  |      |
|--|------|
|  | 1530 |
|  | 1531 |
|  | 1532 |
|  | 1533 |
|  | 1534 |
|  | 1535 |
|  | 1536 |
|  | 1537 |
|  | 1538 |
|  | 1539 |
|  | 1540 |
|  | 1541 |
|  | 1542 |
|  | 1543 |
|  | 1544 |
|  | 1545 |
|  | 1546 |
|  | 1547 |
|  | 1548 |
|  | 1549 |
|  | 1550 |
|  | 1551 |
|  | 1552 |
|  | 1553 |
|  | 1554 |
|  | 1555 |
|  | 1556 |
|  | 1557 |
|  | 1558 |
|  | 1559 |
|  | 1560 |
|  | 1561 |
|  | 1562 |
|  | 1563 |
|  | 1564 |
|  | 1565 |
|  | 1566 |
|  | 1567 |
|  | 1568 |
|  | 1569 |
|  | 1570 |
|  | 1571 |
|  | 1572 |
|  | 1573 |
|  | 1574 |
|  | 1575 |
|  | 1576 |
|  | 1577 |
|  | 1578 |
|  | 1579 |
|  | 1580 |
|  | 1581 |
|  | 1582 |
|  | 1583 |
|  | 1584 |
|  | 1585 |
|  | 1586 |

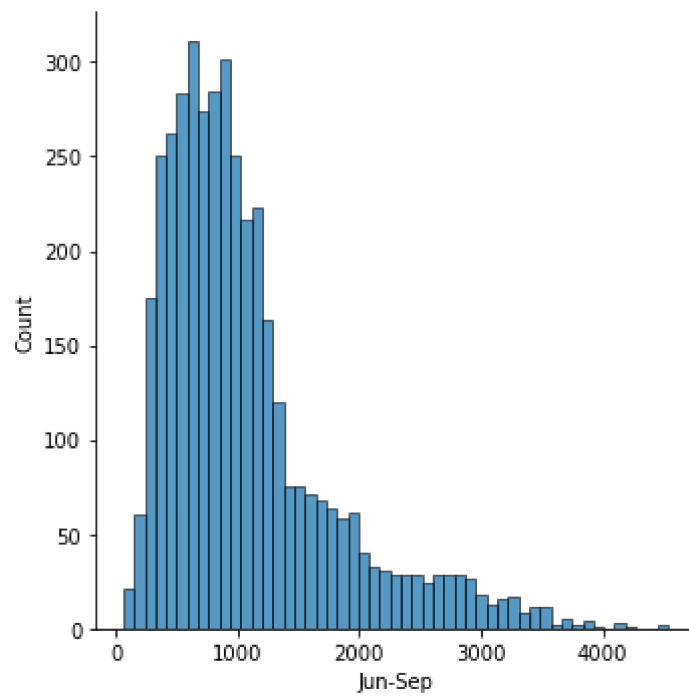
```
In [16]: sns.pairplot(df1)
```

```
Out[16]: <seaborn.axisgrid.PairGrid at 0x1fd98655760>
```



```
In [17]: sns.displot(data["Jun-Sep"])
```

```
Out[17]: <seaborn.axisgrid.FacetGrid at 0x1fd95282a00>
```



```
In [18]: sns.heatmap(df1.corr())
```

```
Out[18]: <AxesSubplot:>
```

