

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

In [3]:

df=data.iloc[1817:1932]
df

Out[3]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT
1817	1817	WEST RAJASTHAN	1901	6.7	0.0	1.1	0.0	6.1	3.0	79.0	59.2	1.0	2.1
1818	1818	WEST RAJASTHAN	1902	0.0	0.0	0.0	0.5	4.0	49.1	27.0	71.3	41.8	1.8
1819	1819	WEST RAJASTHAN	1903	1.7	1.3	5.5	0.0	4.2	2.7	154.8	87.1	49.3	0.1
1820	1820	WEST RAJASTHAN	1904	3.8	2.9	16.3	0.7	11.4	14.6	39.8	45.6	21.4	1.4
1821	1821	WEST RAJASTHAN	1905	6.3	4.8	0.7	1.3	0.3	4.9	30.1	0.6	64.5	0.0
...
1927	1927	WEST RAJASTHAN	2011	0.0	11.8	1.5	1.5	7.8	24.4	88.5	166.8	116.3	0.1
1928	1928	WEST RAJASTHAN	2012	0.5	0.0	0.0	9.5	10.4	5.3	40.4	166.7	92.0	1.9
1929	1929	WEST RAJASTHAN	2013	8.6	21.8	4.2	3.1	1.7	37.6	104.5	138.2	58.7	10.1
1930	1930	WEST RAJASTHAN	2014	0.8	2.2	4.7	8.4	23.0	13.8	94.3	69.6	84.9	0.5
1931	1931	WEST RAJASTHAN	2015	1.4	0.9	30.3	25.2	15.5	53.2	234.6	60.5	35.7	1.1

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	N
1817	1817	WEST RAJASTHAN	1901	6.7	0.0	1.1	0.0	6.1	3.0	79.0	59.2	1.0	2.1	
1818	1818	WEST RAJASTHAN	1902	0.0	0.0	0.0	0.5	4.0	49.1	27.0	71.3	41.8	1.8	
1819	1819	WEST RAJASTHAN	1903	1.7	1.3	5.5	0.0	4.2	2.7	154.8	87.1	49.3	0.1	
1820	1820	WEST RAJASTHAN	1904	3.8	2.9	16.3	0.7	11.4	14.6	39.8	45.6	21.4	1.4	
1821	1821	WEST RAJASTHAN	1905	6.3	4.8	0.7	1.3	0.3	4.9	30.1	0.6	64.5	0.0	

In [5]:

df.tail()

Out[5]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	
1927	1927	WEST RAJASTHAN	2011	0.0	11.8	1.5	1.5	7.8	24.4	88.5	166.8	116.3	0.1	
1928	1928	WEST RAJASTHAN	2012	0.5	0.0	0.0	9.5	10.4	5.3	40.4	166.7	92.0	1.9	
1929	1929	WEST RAJASTHAN	2013	8.6	21.8	4.2	3.1	1.7	37.6	104.5	138.2	58.7	10.1	
1930	1930	WEST RAJASTHAN	2014	0.8	2.2	4.7	8.4	23.0	13.8	94.3	69.6	84.9	0.5	
1931	1931	WEST RAJASTHAN	2015	1.4	0.9	30.3	25.2	15.5	53.2	234.6	60.5	35.7	1.1	

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

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 115 entries, 1817 to 1931
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
1817	1817	WEST RAJASTHAN	1901	6.7	0.0	1.1	0.0	6.1	3.0	79.0	59.2	1.0	2.1
1818	1818	WEST RAJASTHAN	1902	0.0	0.0	0.0	0.5	4.0	49.1	27.0	71.3	41.8	1.8
1819	1819	WEST RAJASTHAN	1903	1.7	1.3	5.5	0.0	4.2	2.7	154.8	87.1	49.3	0.1
1820	1820	WEST RAJASTHAN	1904	3.8	2.9	16.3	0.7	11.4	14.6	39.8	45.6	21.4	1.4
1821	1821	WEST RAJASTHAN	1905	6.3	4.8	0.7	1.3	0.3	4.9	30.1	0.6	64.5	0.0
...
1927	1927	WEST RAJASTHAN	2011	0.0	11.8	1.5	1.5	7.8	24.4	88.5	166.8	116.3	0.1
1928	1928	WEST RAJASTHAN	2012	0.5	0.0	0.0	9.5	10.4	5.3	40.4	166.7	92.0	1.9
1929	1929	WEST RAJASTHAN	2013	8.6	21.8	4.2	3.1	1.7	37.6	104.5	138.2	58.7	10.1
1930	1930	WEST RAJASTHAN	2014	0.8	2.2	4.7	8.4	23.0	13.8	94.3	69.6	84.9	0.5
1931	1931	WEST RAJASTHAN	2015	1.4	0.9	30.3	25.2	15.5	53.2	234.6	60.5	35.7	1.1

115 rows × 20 columns

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

Out[8]:

	index	YEAR	JAN	FEB	MAR	APR	MAY	
count	115.000000	115.000000	115.000000	115.000000	115.000000	115.000000	115.000000	115.0
mean	1874.000000	1958.000000	3.327826	4.930435	3.986087	3.571304	9.443478	28.6
std	33.341666	33.341666	4.551914	7.858800	7.813965	5.916803	10.853168	22.8
min	1817.000000	1901.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.4
25%	1845.500000	1929.500000	0.350000	0.200000	0.200000	0.400000	1.950000	13.3
50%	1874.000000	1958.000000	1.600000	1.300000	1.100000	1.400000	6.100000	21.9
75%	1902.500000	1986.500000	4.000000	5.950000	5.200000	3.750000	12.150000	39.6
max	1931.000000	2015.000000	21.400000	39.100000	59.000000	36.100000	56.800000	143.2

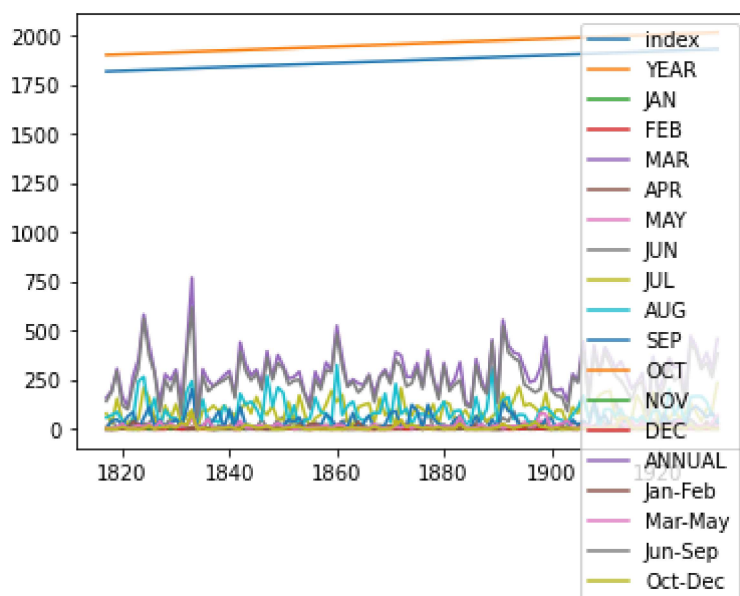
In [9]: `df1.columns`

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 Visualaization

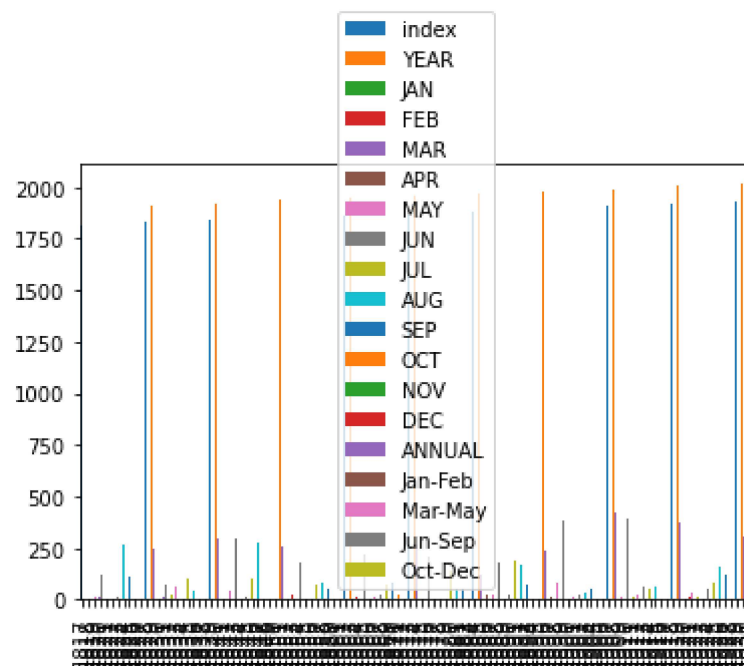
In [10]: `df1.plot.line()`

Out[10]: <AxesSubplot:>



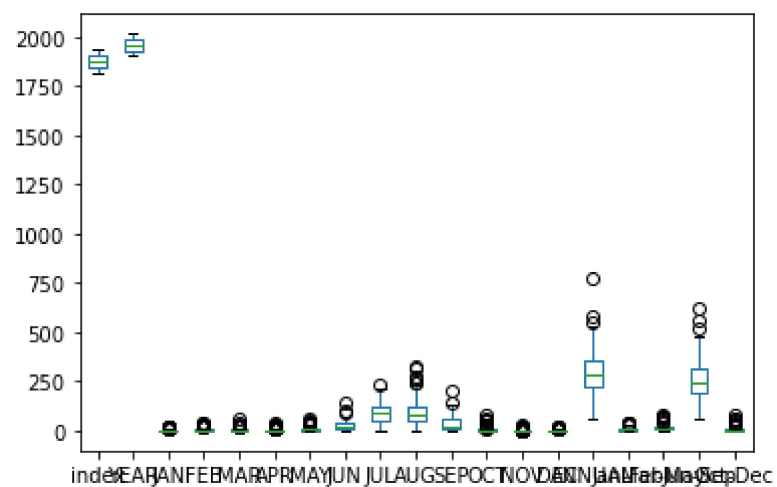
```
In [11]: df1.plot.bar()
```

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



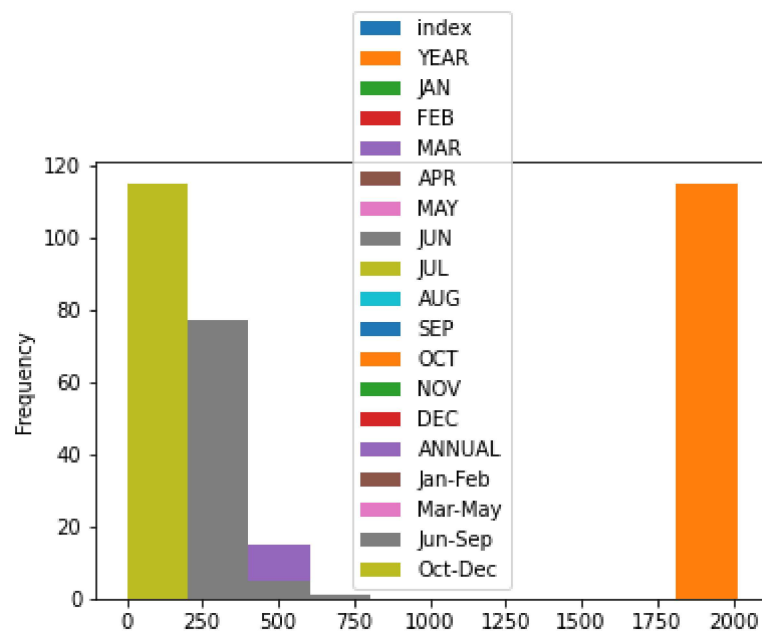
```
In [12]: df1.plot.box()
```

```
Out[12]: <AxesSubplot:>
```



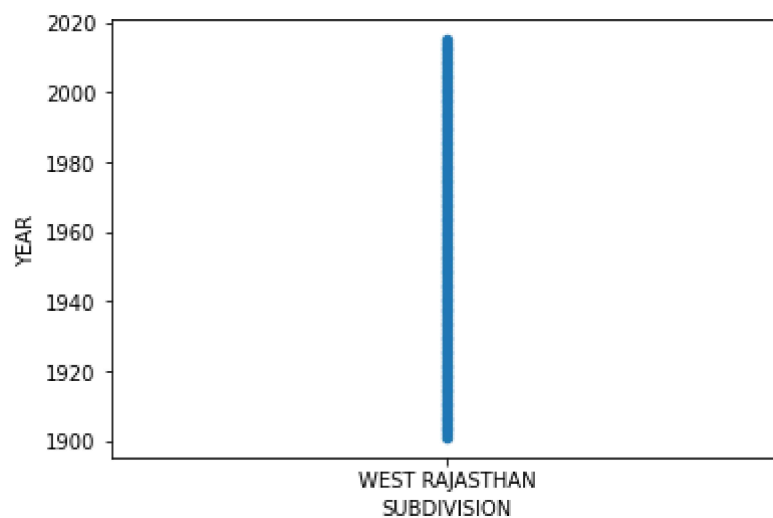
```
In [13]: df1.plot.hist()
```

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



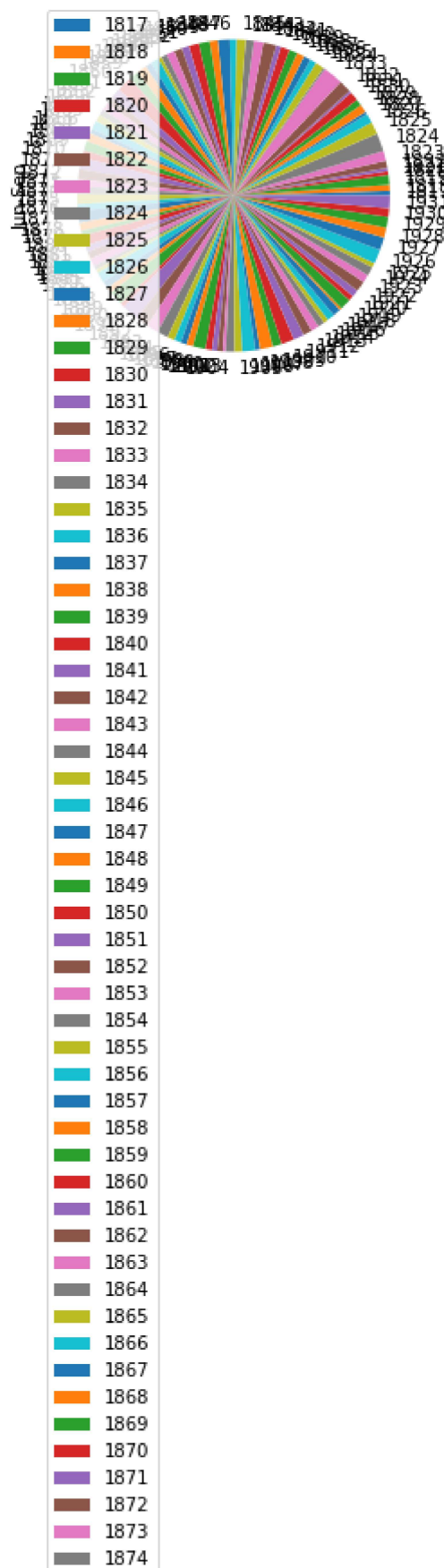
```
In [14]: df1.plot.scatter(x="SUBDIVISION",y="YEAR")
```

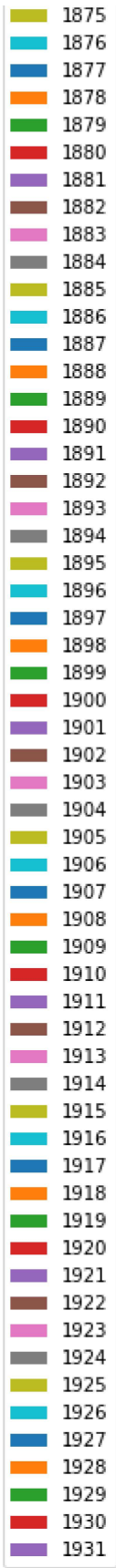
```
Out[14]: <AxesSubplot:xlabel='SUBDIVISION', ylabel='YEAR'>
```




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

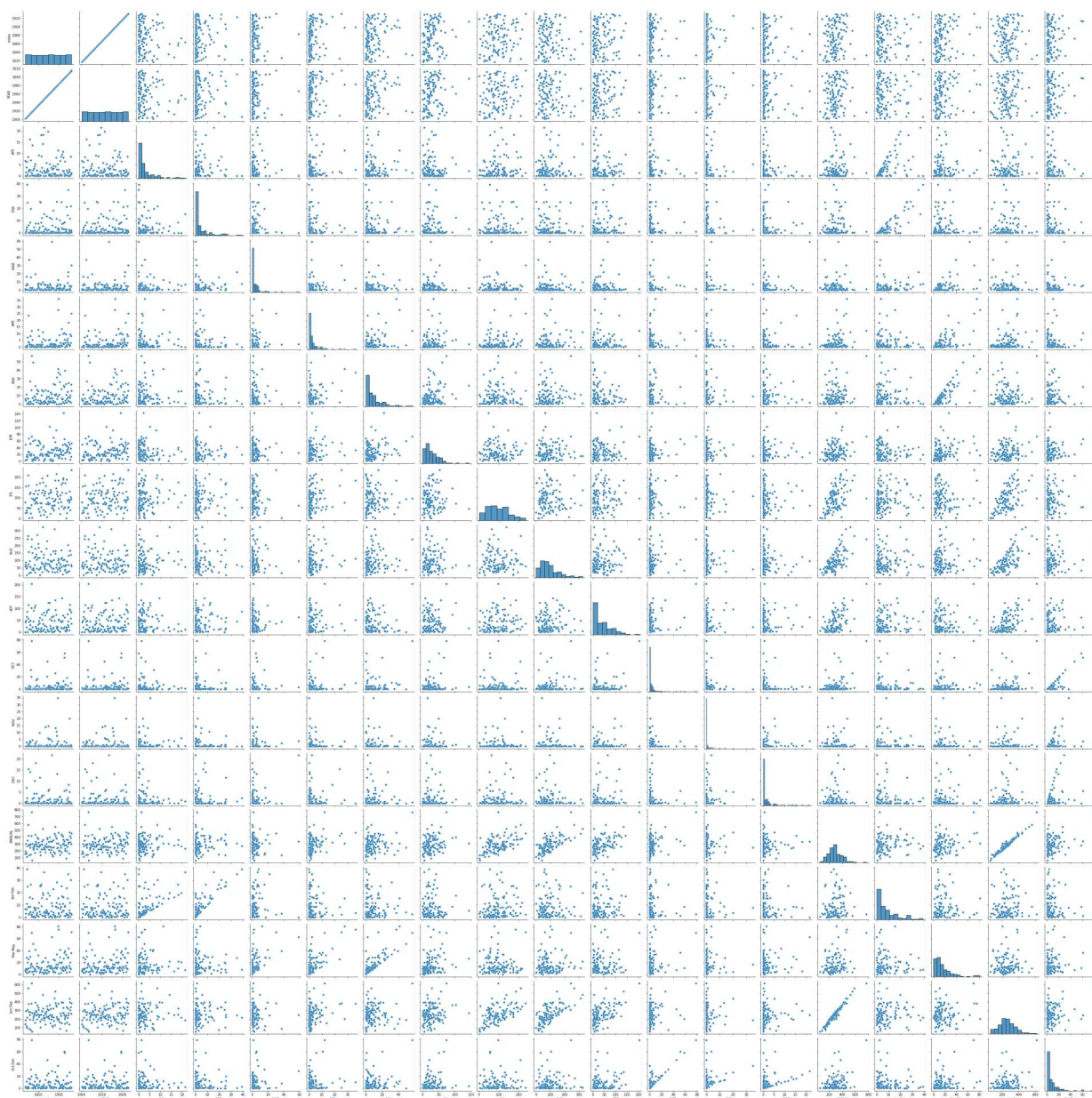
```
Out[15]: array([<AxesSubplot:ylabel='Jun-Sep'>], dtype=object)
```





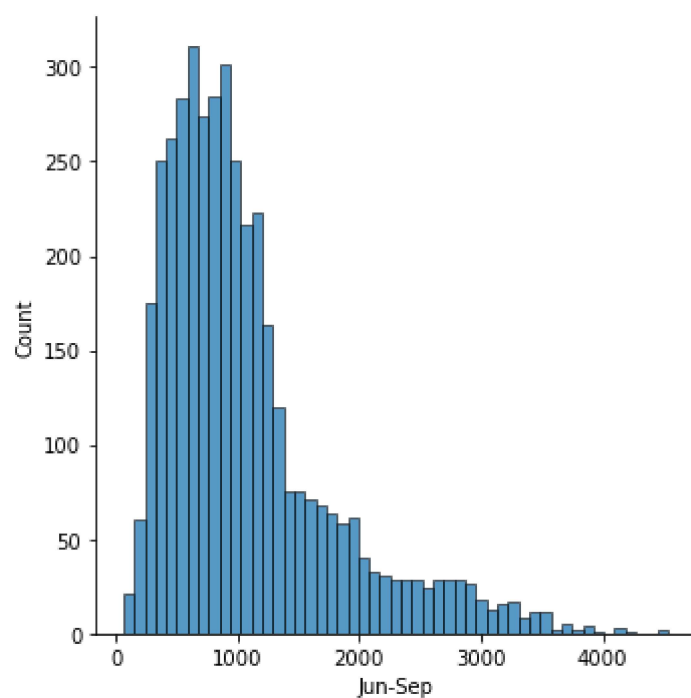
```
In [16]: sns.pairplot(df1)
```

```
Out[16]: <seaborn.axisgrid.PairGrid at 0x29879b75520>
```



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

```
Out[17]: <seaborn.axisgrid.FacetGrid at 0x298031b71c0>
```



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

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

