

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



SOUTH INTERIOR KARNATAKA

```
In [3]: df=data.iloc[3772:3887]
df
```

Out[3]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT
3772	3772	SOUTH INTERIOR KARNATAKA	1901	4.9	31.8	3.0	32.7	109.6	106.0	210.0	109.2	140.8	170.1
3773	3773	SOUTH INTERIOR KARNATAKA	1902	1.9	0.5	6.7	42.6	97.7	91.7	210.0	82.1	138.4	219.1
3774	3774	SOUTH INTERIOR KARNATAKA	1903	0.3	0.0	1.1	11.6	125.1	129.7	284.4	155.7	197.1	154.2
3775	3775	SOUTH INTERIOR KARNATAKA	1904	1.0	0.5	5.2	43.5	144.7	167.9	197.1	73.2	89.6	120.4
3776	3776	SOUTH INTERIOR KARNATAKA	1905	1.7	7.9	14.2	23.6	118.6	95.9	148.4	140.6	43.1	142.8
...
3882	3882	SOUTH INTERIOR KARNATAKA	2011	2.1	12.4	12.4	80.2	83.5	177.1	202.4	199.5	111.2	144.8
3883	3883	SOUTH INTERIOR KARNATAKA	2012	4.6	5.5	8.1	99.0	45.6	81.8	144.7	236.5	100.6	62.8
3884	3884	SOUTH INTERIOR KARNATAKA	2013	0.5	10.1	11.7	34.6	95.6	176.2	307.4	151.7	191.8	103.7
3885	3885	SOUTH INTERIOR KARNATAKA	2014	0.4	2.4	17.7	46.7	130.5	106.8	271.6	254.6	161.6	152.9
3886	3886	SOUTH INTERIOR KARNATAKA	2015	1.7	0.2	24.4	80.5	125.3	218.7	112.0	136.6	164.5	106.1

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
3772	3772	SOUTH INTERIOR KARNATAKA	1901	4.9	31.8	3.0	32.7	109.6	106.0	210.0	109.2	140.8	170.1
3773	3773	SOUTH INTERIOR KARNATAKA	1902	1.9	0.5	6.7	42.6	97.7	91.7	210.0	82.1	138.4	219.1
3774	3774	SOUTH INTERIOR KARNATAKA	1903	0.3	0.0	1.1	11.6	125.1	129.7	284.4	155.7	197.1	154.2
3775	3775	SOUTH INTERIOR KARNATAKA	1904	1.0	0.5	5.2	43.5	144.7	167.9	197.1	73.2	89.6	120.4
3776	3776	SOUTH INTERIOR KARNATAKA	1905	1.7	7.9	14.2	23.6	118.6	95.9	148.4	140.6	43.1	142.8



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

Out[5]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT
3882	3882	SOUTH INTERIOR KARNATAKA	2011	2.1	12.4	12.4	80.2	83.5	177.1	202.4	199.5	111.2	144.8
3883	3883	SOUTH INTERIOR KARNATAKA	2012	4.6	5.5	8.1	99.0	45.6	81.8	144.7	236.5	100.6	62.8
3884	3884	SOUTH INTERIOR KARNATAKA	2013	0.5	10.1	11.7	34.6	95.6	176.2	307.4	151.7	191.8	103.7
3885	3885	SOUTH INTERIOR KARNATAKA	2014	0.4	2.4	17.7	46.7	130.5	106.8	271.6	254.6	161.6	152.9
3886	3886	SOUTH INTERIOR KARNATAKA	2015	1.7	0.2	24.4	80.5	125.3	218.7	112.0	136.6	164.5	106.1



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

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 115 entries, 3772 to 3886
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
3772	3772	SOUTH INTERIOR KARNATAKA	1901	4.9	31.8	3.0	32.7	109.6	106.0	210.0	109.2	140.8	170.1
3773	3773	SOUTH INTERIOR KARNATAKA	1902	1.9	0.5	6.7	42.6	97.7	91.7	210.0	82.1	138.4	219.1
3774	3774	SOUTH INTERIOR KARNATAKA	1903	0.3	0.0	1.1	11.6	125.1	129.7	284.4	155.7	197.1	154.2
3775	3775	SOUTH INTERIOR KARNATAKA	1904	1.0	0.5	5.2	43.5	144.7	167.9	197.1	73.2	89.6	120.4
3776	3776	SOUTH INTERIOR KARNATAKA	1905	1.7	7.9	14.2	23.6	118.6	95.9	148.4	140.6	43.1	142.8
...
3882	3882	SOUTH INTERIOR KARNATAKA	2011	2.1	12.4	12.4	80.2	83.5	177.1	202.4	199.5	111.2	144.8
3883	3883	SOUTH INTERIOR KARNATAKA	2012	4.6	5.5	8.1	99.0	45.6	81.8	144.7	236.5	100.6	62.8
3884	3884	SOUTH INTERIOR KARNATAKA	2013	0.5	10.1	11.7	34.6	95.6	176.2	307.4	151.7	191.8	103.7
3885	3885	SOUTH INTERIOR KARNATAKA	2014	0.4	2.4	17.7	46.7	130.5	106.8	271.6	254.6	161.6	152.9
3886	3886	SOUTH INTERIOR KARNATAKA	2015	1.7	0.2	24.4	80.5	125.3	218.7	112.0	136.6	164.5	106.1

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	115.000000
mean	3829.000000	1958.000000	2.928696	4.163478	9.485217	42.280870	92.100000	141.000000
std	33.341666	33.341666	4.808741	7.560439	13.955998	22.061039	38.219243	46.000000
min	3772.000000	1901.000000	0.000000	0.000000	0.000000	4.800000	9.600000	64.000000
25%	3800.500000	1929.500000	0.100000	0.100000	2.300000	25.450000	63.350000	106.000000
50%	3829.000000	1958.000000	0.800000	1.200000	5.200000	40.000000	90.900000	135.000000
75%	3857.500000	1986.500000	4.100000	4.600000	12.750000	51.400000	114.150000	174.000000
max	3886.000000	2015.000000	24.400000	44.300000	108.900000	127.700000	190.500000	269.000000

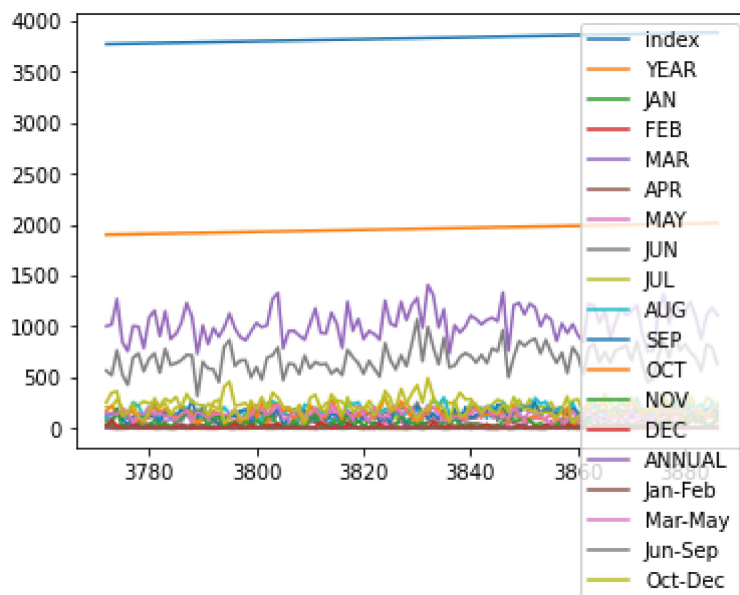
In [9]: `df1.columns`

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

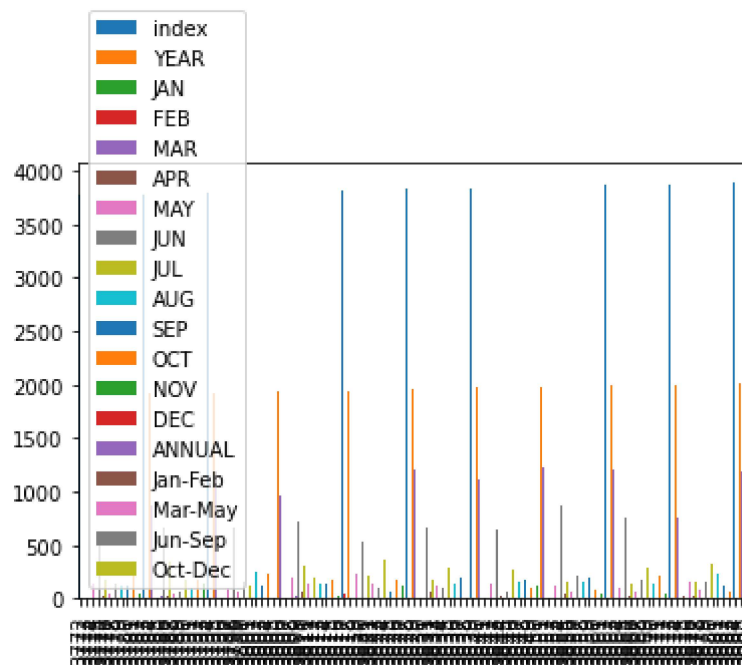
In [10]: `df1.plot.line()`

Out[10]: <AxesSubplot:>



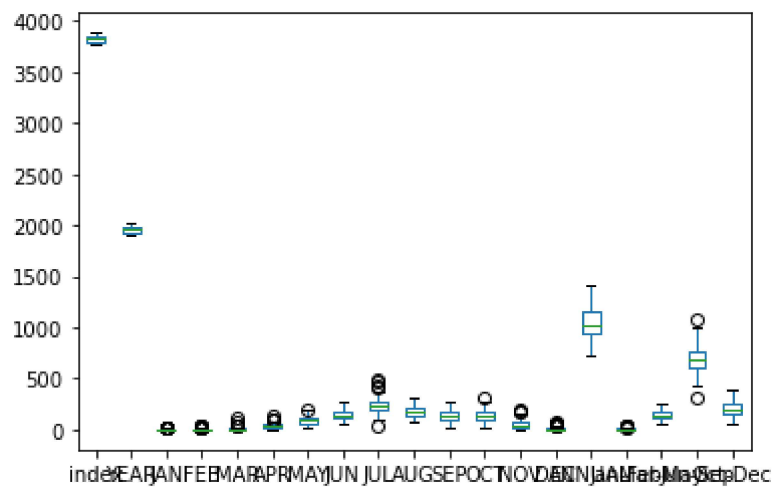
```
In [11]: df1.plot.bar()
```

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



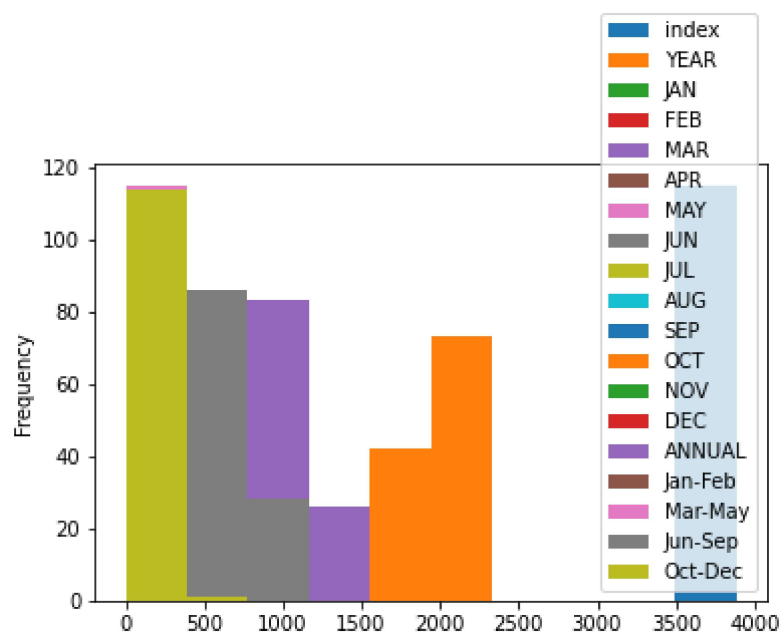
```
In [12]: df1.plot.box()
```

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



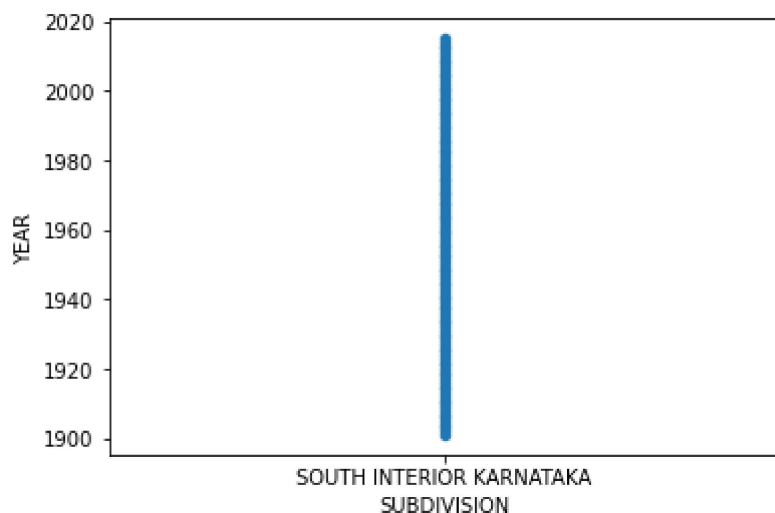
```
In [13]: df1.plot.hist()
```

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



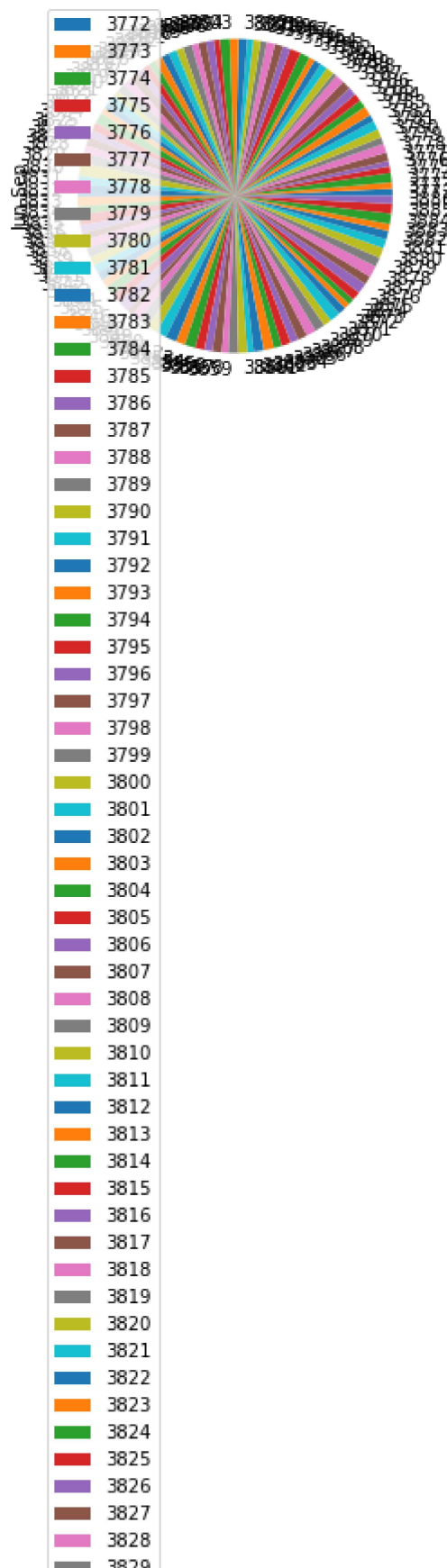
```
In [14]: df1.plot.scatter(x="SUBDIVISION",y="YEAR")
```

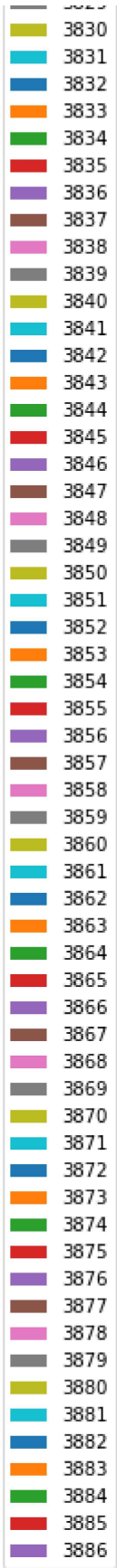
```
Out[14]: <AxesSubplot:xlabel='SUBDIVISION', ylabel='YEAR'>
```




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

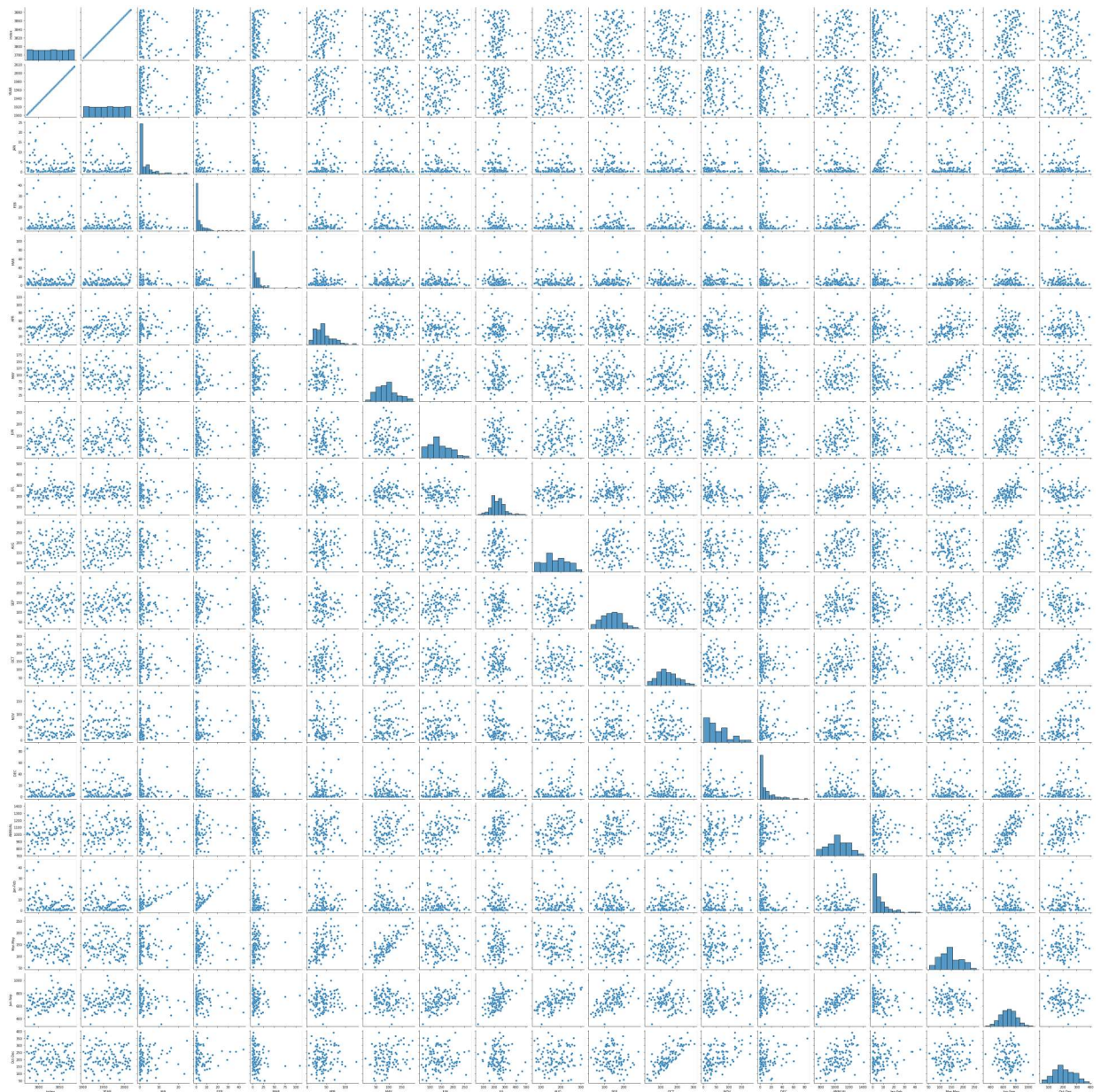
```
Out[15]: array([<AxesSubplot:ylabel='Jun-Sep'>], dtype=object)
```





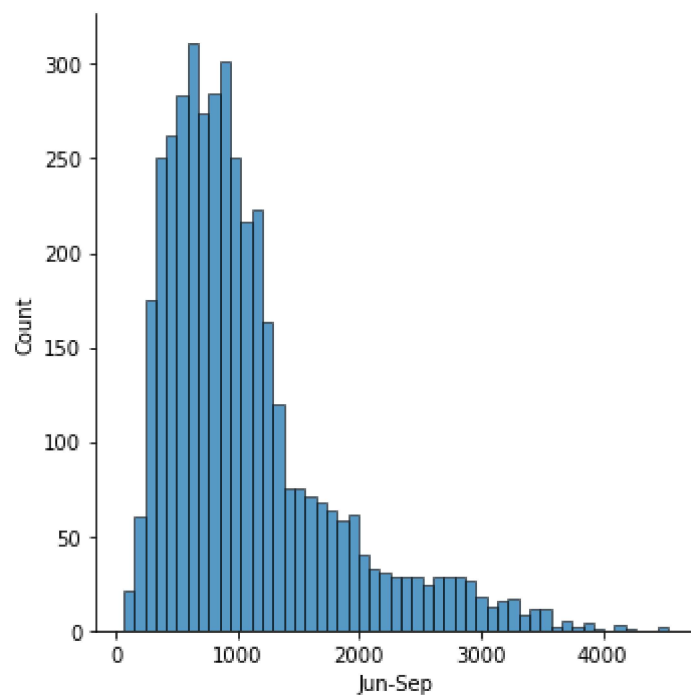
```
In [16]: sns.pairplot(df1)
```

```
Out[16]: <seaborn.axisgrid.PairGrid at 0x1d28fd05340>
```



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

```
Out[17]: <seaborn.axisgrid.FacetGrid at 0x1d29aad7be0>
```



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

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

