

## 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



## NORTH INTERIOR KARNATAKA

In [3]:

df=data.iloc[3658:3772]  
df

Out[3]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT
3658	3658	NORTH INTERIOR KARNATAKA	1902	0.0	0.0	0.3	22.5	34.4	111.3	83.2	78.1	146.7	118.8
3659	3659	NORTH INTERIOR KARNATAKA	1903	3.5	0.0	0.1	6.9	53.4	102.8	209.4	146.4	189.3	166.4
3660	3660	NORTH INTERIOR KARNATAKA	1904	0.2	0.3	8.5	11.0	46.3	120.6	91.6	48.5	165.1	86.5
3661	3661	NORTH INTERIOR KARNATAKA	1905	0.0	6.0	2.6	16.0	51.2	99.6	60.1	139.2	42.2	85.0
3662	3662	NORTH INTERIOR KARNATAKA	1906	21.3	0.0	0.2	2.6	30.0	142.0	120.3	182.1	116.0	86.2
...	...	...	...	...	...	...	...	...	...	...	...	...	...
3767	3767	NORTH INTERIOR KARNATAKA	2011	0.5	7.2	7.2	41.2	46.8	101.3	150.8	152.0	69.0	73.4
3768	3768	NORTH INTERIOR KARNATAKA	2012	28.5	6.2	0.4	35.4	19.5	60.0	114.5	105.5	79.2	85.2
3769	3769	NORTH INTERIOR KARNATAKA	2013	1.2	6.1	3.0	25.4	47.4	99.4	160.7	73.9	201.0	101.0
3770	3770	NORTH INTERIOR KARNATAKA	2014	0.0	6.1	29.2	26.4	93.0	50.4	136.8	205.2	90.2	80.3
3771	3771	NORTH INTERIOR KARNATAKA	2015	2.4	0.0	27.5	50.8	45.3	89.6	38.5	78.4	150.8	61.2

114 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
<b>3658</b>	3658	NORTH INTERIOR KARNATAKA	1902	0.0	0.0	0.3	22.5	34.4	111.3	83.2	78.1	146.7	118.8
<b>3659</b>	3659	NORTH INTERIOR KARNATAKA	1903	3.5	0.0	0.1	6.9	53.4	102.8	209.4	146.4	189.3	166.4
<b>3660</b>	3660	NORTH INTERIOR KARNATAKA	1904	0.2	0.3	8.5	11.0	46.3	120.6	91.6	48.5	165.1	86.5
<b>3661</b>	3661	NORTH INTERIOR KARNATAKA	1905	0.0	6.0	2.6	16.0	51.2	99.6	60.1	139.2	42.2	85.0
<b>3662</b>	3662	NORTH INTERIOR KARNATAKA	1906	21.3	0.0	0.2	2.6	30.0	142.0	120.3	182.1	116.0	86.2



In [5]: df.tail()

Out[5]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT
<b>3767</b>	3767	NORTH INTERIOR KARNATAKA	2011	0.5	7.2	7.2	41.2	46.8	101.3	150.8	152.0	69.0	73.4
<b>3768</b>	3768	NORTH INTERIOR KARNATAKA	2012	28.5	6.2	0.4	35.4	19.5	60.0	114.5	105.5	79.2	85.2
<b>3769</b>	3769	NORTH INTERIOR KARNATAKA	2013	1.2	6.1	3.0	25.4	47.4	99.4	160.7	73.9	201.0	101.0
<b>3770</b>	3770	NORTH INTERIOR KARNATAKA	2014	0.0	6.1	29.2	26.4	93.0	50.4	136.8	205.2	90.2	80.3
<b>3771</b>	3771	NORTH INTERIOR KARNATAKA	2015	2.4	0.0	27.5	50.8	45.3	89.6	38.5	78.4	150.8	61.2



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

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 114 entries, 3658 to 3771
Data columns (total 20 columns):
#   Column          Non-Null Count  Dtype
---  -
0   index           114 non-null   int64
1   SUBDIVISION     114 non-null   object
2   YEAR            114 non-null   int64
3   JAN             114 non-null   float64
4   FEB             114 non-null   float64
5   MAR             114 non-null   float64
6   APR             114 non-null   float64
7   MAY             114 non-null   float64
8   JUN             114 non-null   float64
9   JUL             114 non-null   float64
10  AUG             114 non-null   float64
11  SEP             114 non-null   float64
12  OCT             114 non-null   float64
13  NOV             114 non-null   float64
14  DEC             114 non-null   float64
15  ANNUAL          114 non-null   float64
16  Jan-Feb         114 non-null   float64
17  Mar-May         114 non-null   float64
18  Jun-Sep         114 non-null   float64
19  Oct-Dec         114 non-null   float64
dtypes: float64(17), int64(2), object(1)
memory usage: 17.9+ 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
3658	3658	NORTH INTERIOR KARNATAKA	1902	0.0	0.0	0.3	22.5	34.4	111.3	83.2	78.1	146.7	118.8
3659	3659	NORTH INTERIOR KARNATAKA	1903	3.5	0.0	0.1	6.9	53.4	102.8	209.4	146.4	189.3	166.4
3660	3660	NORTH INTERIOR KARNATAKA	1904	0.2	0.3	8.5	11.0	46.3	120.6	91.6	48.5	165.1	86.5
3661	3661	NORTH INTERIOR KARNATAKA	1905	0.0	6.0	2.6	16.0	51.2	99.6	60.1	139.2	42.2	85.0
3662	3662	NORTH INTERIOR KARNATAKA	1906	21.3	0.0	0.2	2.6	30.0	142.0	120.3	182.1	116.0	86.2
...	...	...	...	...	...	...	...	...	...	...	...	...	...
3767	3767	NORTH INTERIOR KARNATAKA	2011	0.5	7.2	7.2	41.2	46.8	101.3	150.8	152.0	69.0	73.4
3768	3768	NORTH INTERIOR KARNATAKA	2012	28.5	6.2	0.4	35.4	19.5	60.0	114.5	105.5	79.2	85.2
3769	3769	NORTH INTERIOR KARNATAKA	2013	1.2	6.1	3.0	25.4	47.4	99.4	160.7	73.9	201.0	101.0
3770	3770	NORTH INTERIOR KARNATAKA	2014	0.0	6.1	29.2	26.4	93.0	50.4	136.8	205.2	90.2	80.3
3771	3771	NORTH INTERIOR KARNATAKA	2015	2.4	0.0	27.5	50.8	45.3	89.6	38.5	78.4	150.8	61.2

114 rows × 20 columns



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

Out[8]:

	index	YEAR	JAN	FEB	MAR	APR	MAY	
<b>count</b>	114.000000	114.000000	114.000000	114.000000	114.000000	114.000000	114.000000	114.0
<b>mean</b>	3714.500000	1958.500000	3.008772	3.035088	7.123684	23.924561	46.873684	100.0
<b>std</b>	33.052988	33.052988	6.224851	5.360574	12.727298	15.236731	26.636674	34.0
<b>min</b>	3658.000000	1902.000000	0.000000	0.000000	0.000000	0.200000	3.500000	38.0
<b>25%</b>	3686.250000	1930.250000	0.000000	0.000000	0.500000	12.150000	29.050000	75.0
<b>50%</b>	3714.500000	1958.500000	0.200000	0.300000	3.450000	22.300000	40.600000	99.0
<b>75%</b>	3742.750000	1986.750000	2.400000	3.350000	7.600000	32.100000	62.500000	115.0
<b>max</b>	3771.000000	2015.000000	28.500000	28.400000	109.200000	96.900000	127.300000	235.0

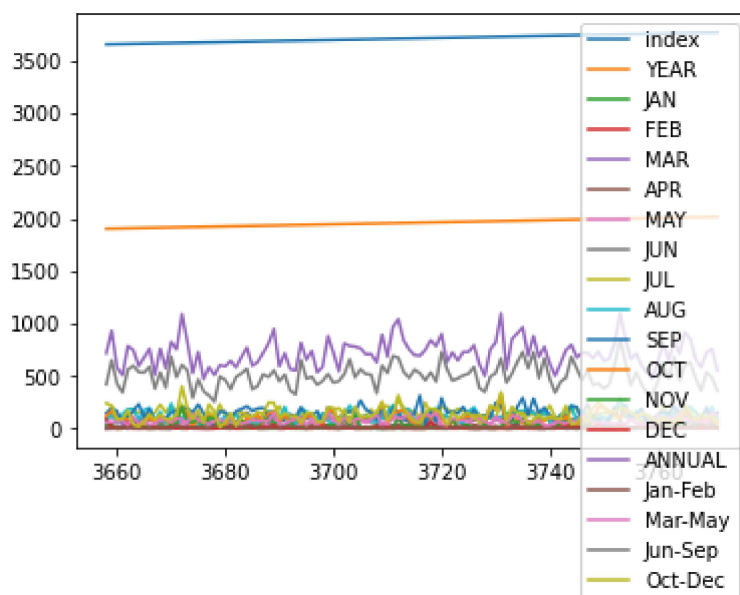
In [9]: `df1.columns`

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

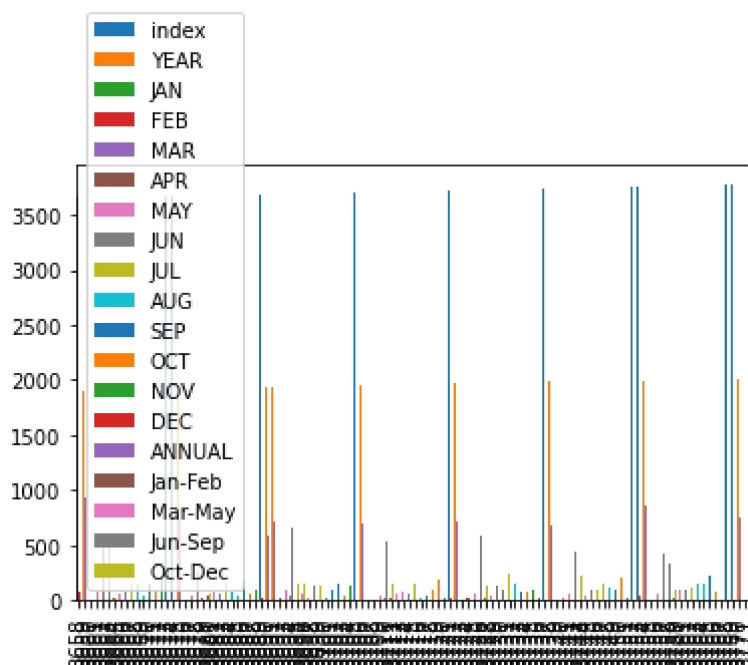
In [10]: `df1.plot.line()`

Out[10]: <AxesSubplot:>



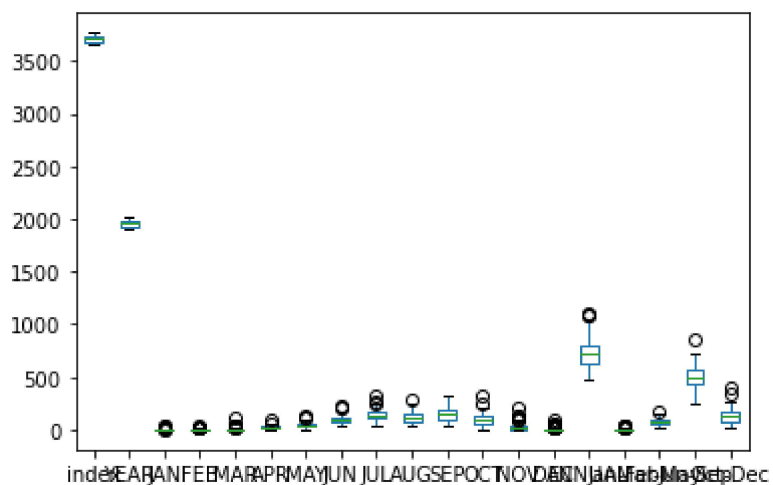
```
In [11]: df1.plot.bar()
```

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



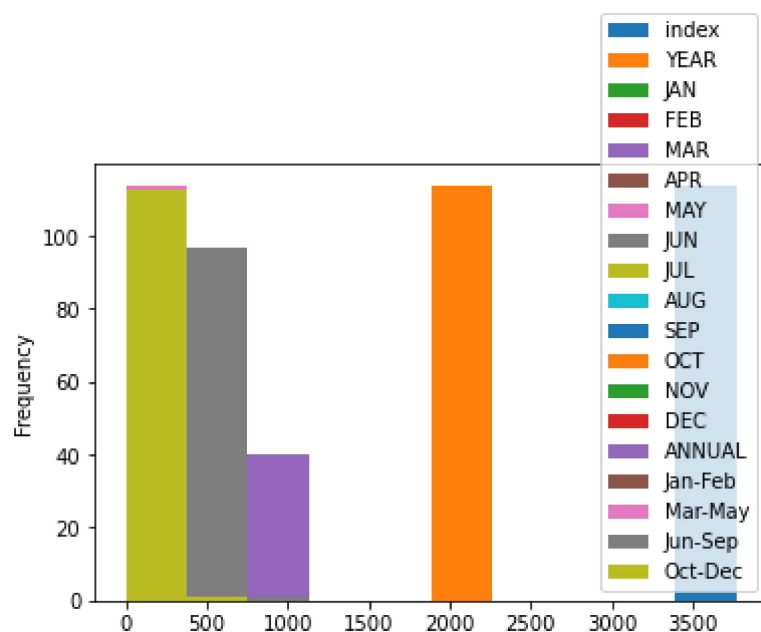
```
In [12]: df1.plot.box()
```

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



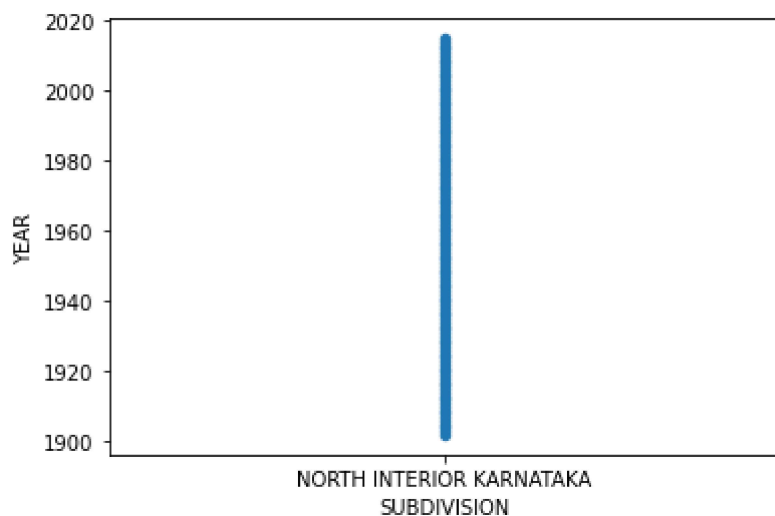
```
In [13]: df1.plot.hist()
```

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



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

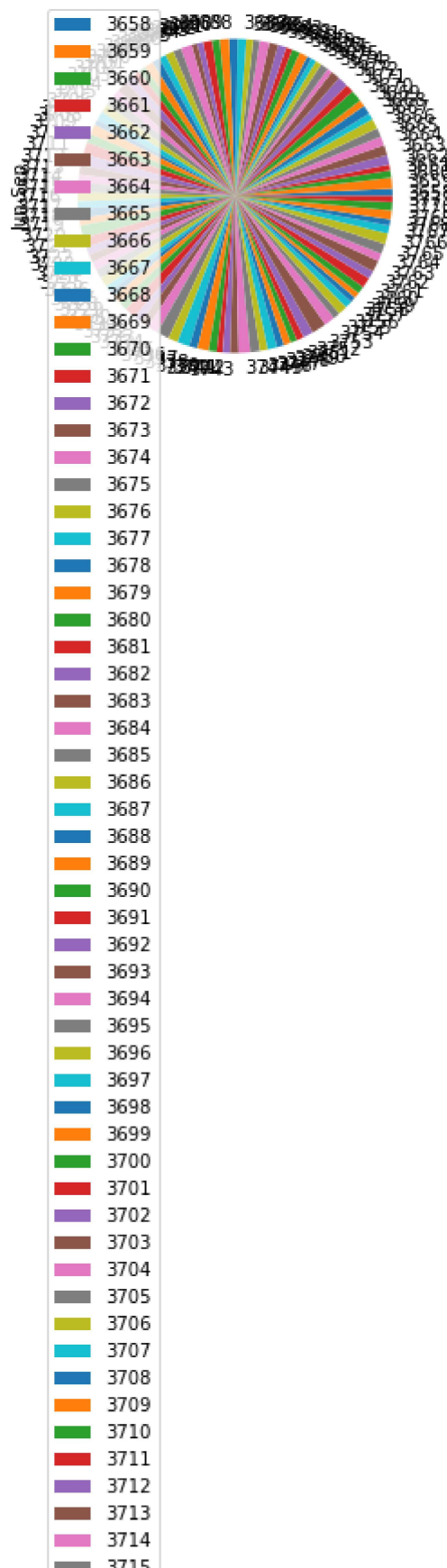
```
Out[14]: <AxesSubplot:xlabel='SUBDIVISION', ylabel='YEAR'>
```

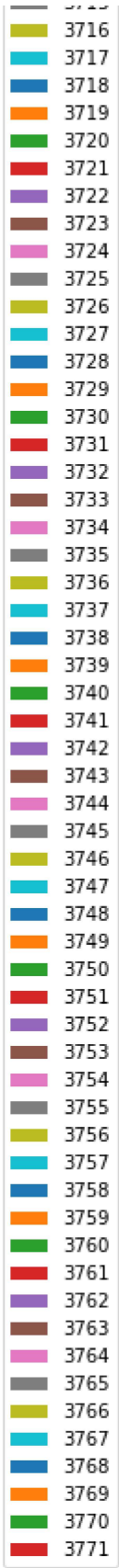




```
In [15]: df2=df1[[ 'Jun-Sep']]
df2.plot.pie(subplots=True)
```

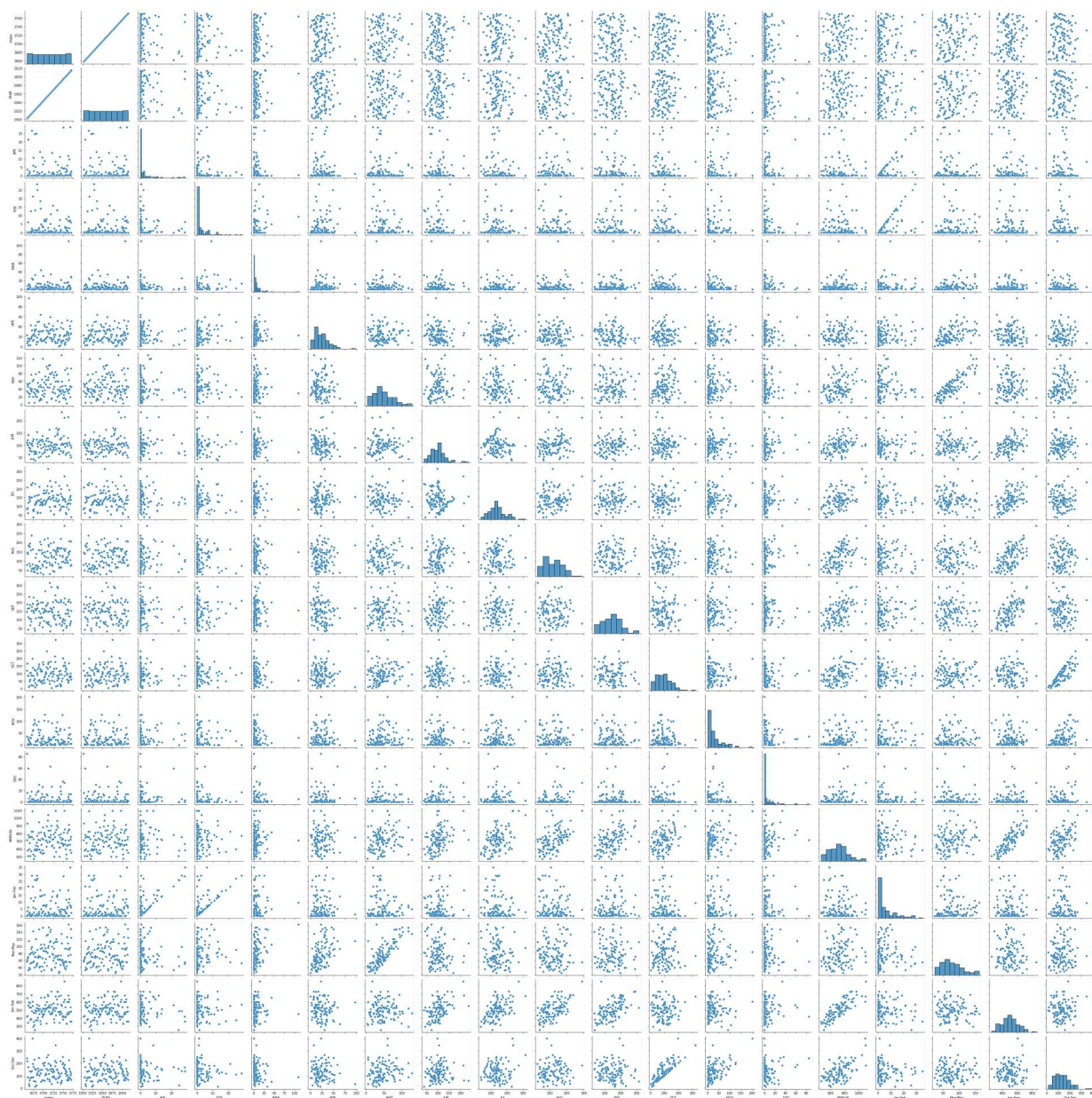
```
Out[15]: array([<AxesSubplot:ylabel='Jun-Sep'>], dtype=object)
```





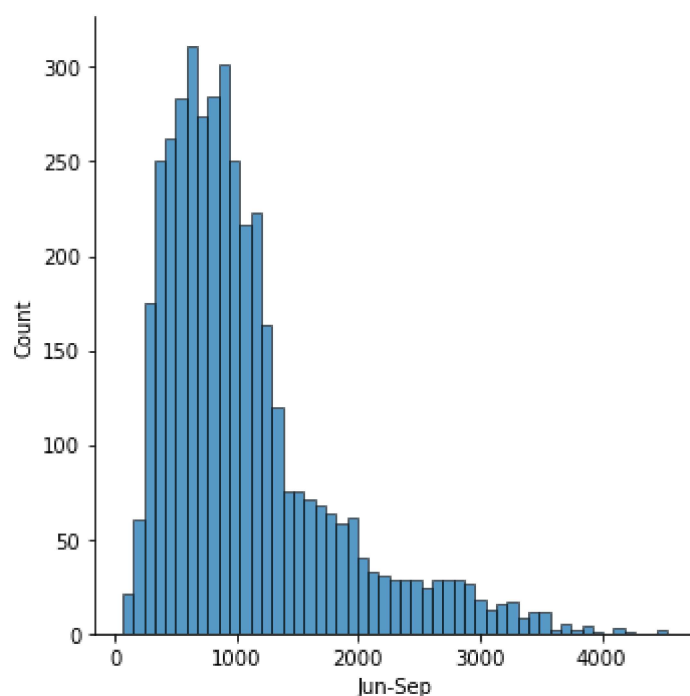
```
In [16]: sns.pairplot(df1)
```

```
Out[16]: <seaborn.axisgrid.PairGrid at 0x27a87d08430>
```



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

```
Out[17]: <seaborn.axisgrid.FacetGrid at 0x27a96243b50>
```



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

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

