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	ANDAMAN & NICOBAR ISLANDS	1901	49.2	87.1	29.2	2.3	528.8	517.5	365.1	481.1	332.6	;
1	ANDAMAN & NICOBAR ISLANDS	1902	0.0	159.8	12.2	0.0	446.1	537.1	228.9	753.7	666.2	
2	ANDAMAN & NICOBAR ISLANDS	1903	12.7	144.0	0.0	1.0	235.1	479.9	728.4	326.7	339.0	
3	ANDAMAN & NICOBAR ISLANDS	1904	9.4	14.7	0.0	202.4	304.5	495.1	502.0	160.1	820.4	:
4	ANDAMAN & NICOBAR ISLANDS	1905	1.3	0.0	3.3	26.9	279.5	628.7	368.7	330.5	297.0	:
4111	LAKSHADWEEP	2011	5.1	2.8	3.1	85.9	107.2	153.6	350.2	254.0	255.2	
4112	LAKSHADWEEP	2012	19.2	0.1	1.6	76.8	21.2	327.0	231.5	381.2	179.8	
4113	LAKSHADWEEP	2013	26.2	34.4	37.5	5.3	88.3	426.2	296.4	154.4	180.0	
4114	LAKSHADWEEP	2014	53.2	16.1	4.4	14.9	57.4	244.1	116.1	466.1	132.2	
4115	LAKSHADWEEP	2015	2.2	0.5	3.7	87.1	133.1	296.6	257.5	146.4	160.4	
	0 1 2 3 4 4111 4112 4113 4114	ANDAMAN & NICOBAR ISLANDS LAKSHADWEEP 4111 LAKSHADWEEP 4113 LAKSHADWEEP 4114 LAKSHADWEEP	ANDAMAN & 1901 ISLANDS ANDAMAN & 1902 ISLANDS ANDAMAN & 1902 ISLANDS ANDAMAN & 1903 ISLANDS ANDAMAN & 1904 ISLANDS ANDAMAN & 1904 ISLANDS ANDAMAN & 1905 ISLANDS 4111 LAKSHADWEEP 2011 4112 LAKSHADWEEP 2013 4114 LAKSHADWEEP 2014	ANDAMAN & 1901 49.2 ISLANDS ANDAMAN & 1902 0.0 ISLANDS ANDAMAN & 1902 0.0 ISLANDS ANDAMAN & 1903 12.7 ISLANDS ANDAMAN & 1904 9.4 ISLANDS ANDAMAN & 1904 9.4 ISLANDS ANDAMAN & 1905 1.3 ISLANDS	ANDAMAN & 1901 49.2 87.1 SLANDS 1902 0.0 159.8 ISLANDS 1903 12.7 144.0 ISLANDS 1904 9.4 14.7 ISLANDS 1905 1.3 0.0 ISLANDS 1905 1905 1905 1905 1905 1905 1905 1905	ANDAMAN & 1901 49.2 87.1 29.2 SILANDS 1 1 NICOBAR 1902 0.0 159.8 12.2 ANDAMAN & 1903 12.7 144.0 0.0 ISLANDS 1904 9.4 14.7 0.0 SILANDS 1905 1.3 0.0 3.3 SILANDS 1905 1.3 0.0 3.3 SILANDS 1905 1.3 0.0 3.3 SILANDS 1905 1.3 1	ANDAMAN & 1901 49.2 87.1 29.2 2.3 ISLANDS 1902 0.0 159.8 12.2 0.0 ISLANDS 2 NICOBAR 1903 12.7 144.0 0.0 1.0 ISLANDS 3 NICOBAR 1904 9.4 14.7 0.0 202.4 ISLANDS 1SLANDS 1905 1.3 0.0 3.3 26.9 ISLANDS	ANDAMAN & NICOBAR ISLANDS ANDAMAN & 1901 49.2 87.1 29.2 2.3 528.8 11 200 159.8 12.2 0.0 446.1 150.00 159.8 12.2 0.0 446.1 150.00 159.8 12.2 0.0 446.1 150.00 159.8 159.8 150.00 159.00 159.8 150.00 159.00 159.8 150.00 159.8 150.00 159.8 150.00 159.8 150.00 159.8 150.00 159.8 150.00 159.8 150.00 159.8 150.00 159.8 150.00 159.8 150	ANDAMAN & NICOBAR ISLANDS 1901 49.2 87.1 29.2 2.3 528.8 517.5 ANDAMAN & NICOBAR ISLANDS 1902 0.0 159.8 12.2 0.0 446.1 537.1 ANDAMAN & NICOBAR ISLANDS 12.7 144.0 0.0 1.0 235.1 479.9 ANDAMAN & NICOBAR ISLANDS 1904 9.4 14.7 0.0 202.4 304.5 495.1 ANDAMAN & NICOBAR ISLANDS 1.3 0.0 3.3 26.9 279.5 628.7 ANDAMAN & NICOBAR ISLANDS 1.3 0.0 3.3 26.9 279.5 628.7 ANDAMAN & NICOBAR ISLANDS 1.3 2.8 3.1 85.9 107.2 153.6 4111 LAKSHADWEEP 2012 19.2 0.1 1.6 76.8 21.2 327.0 4113 LAKSHADWEEP 2013 26.2 34.4 37.5 5.3 88.3 426.2 4114 LAKSHADWEEP 2014 53.2 16.1 4.4 14.9 57.4 244.1	ANDAMAN & NICOBAR ISLANDS	ANDAMAN & 1901 49.2 87.1 29.2 2.3 528.8 517.5 365.1 481.1 SLANDS 1	ANDAMAN & NICOBAR ISLANDS 1901 49.2 87.1 29.2 2.3 528.8 517.5 365.1 481.1 332.6 ANDAMAN & NICOBAR ISLANDS 1902 0.0 159.8 12.2 0.0 446.1 537.1 228.9 753.7 666.2 ANDAMAN & NICOBAR ISLANDS 1903 12.7 144.0 0.0 1.0 235.1 479.9 728.4 326.7 339.0 ANDAMAN & NICOBAR ISLANDS 1904 9.4 14.7 0.0 202.4 304.5 495.1 502.0 160.1 820.4 ANDAMAN & NICOBAR ISLANDS 1.3 0.0 3.3 26.9 279.5 628.7 368.7 330.5 297.0 ANDAMAN & NICOBAR ISLANDS 1.5 2.8 3.1 85.9 107.2 153.6 350.2 254.0 255.2 A112 LAKSHADWEEP 2012 19.2 0.1 1.6 76.8 21.2 327.0 231.5 381.2 179.8 A113 LAKSHADWEEP 2013 26.2 34.4 37.5 5.3 88.3 426.2 296.4 154.4 180.0 A114 LAKSHADWEEP 2014 53.2 16.1 4.4 14.9 57.4 244.1 116.1 466.1 132.2

4116 rows × 20 columns

ASSAM & MEGHALAYA

In [3]: df=data.iloc[1357:1472]
df

Out[3]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	ОСТ
1357	1357	HARYANA DELHI & CHANDIGARH	1901	35.4	28.9	11.1	0.0	5.1	13.2	126.4	151.5	10.5	2.0
1358	1358	HARYANA DELHI & CHANDIGARH	1902	0.0	0.7	2.9	10.2	15.8	74.6	149.3	97.1	59.8	9.3
1359	1359	HARYANA DELHI & CHANDIGARH	1903	14.7	0.5	2.3	0.5	8.5	8.6	151.6	138.2	97.7	4.0
1360	1360	HARYANA DELHI & CHANDIGARH	1904	7.6	0.7	48.0	0.5	29.3	34.3	109.7	162.9	102.3	1.5
1361	1361	HARYANA DELHI & CHANDIGARH	1905	44.8	20.8	14.0	1.3	7.4	20.1	93.6	23.1	92.6	0.0
1467	1467	HARYANA DELHI & CHANDIGARH	2011	0.7	26.7	6.9	8.9	28.7	94.4	85.0	127.3	133.1	0.0
1468	1468	HARYANA DELHI & CHANDIGARH	2012	8.2	0.2	0.1	11.8	3.8	5.3	68.1	196.6	90.7	2.4
1469	1469	HARYANA DELHI & CHANDIGARH	2013	21.1	52.2	5.3	3.3	1.4	62.1	96.5	161.9	42.8	10.9
1470	1470	HARYANA DELHI & CHANDIGARH	2014	13.0	17.3	26.8	7.5	20.3	25.9	72.3	34.8	67.3	10.5
1471	1471	HARYANA DELHI & CHANDIGARH	2015	12.4	6.6	71.8	34.8	8.4	43.7	130.3	89.2	32.1	3.7

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	ОСТ
1357	1357	HARYANA DELHI & CHANDIGARH	1901	35.4	28.9	11.1	0.0	5.1	13.2	126.4	151.5	10.5	2.0
1358	1358	HARYANA DELHI & CHANDIGARH	1902	0.0	0.7	2.9	10.2	15.8	74.6	149.3	97.1	59.8	9.3
1359	1359	HARYANA DELHI & CHANDIGARH	1903	14.7	0.5	2.3	0.5	8.5	8.6	151.6	138.2	97.7	4.0
1360	1360	HARYANA DELHI & CHANDIGARH	1904	7.6	0.7	48.0	0.5	29.3	34.3	109.7	162.9	102.3	1.5
1361	1361	HARYANA DELHI & CHANDIGARH	1905	44.8	20.8	14.0	1.3	7.4	20.1	93.6	23.1	92.6	0.0

In [5]: df.tail()

Out[5]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	ОСТ
1467	1467	HARYANA DELHI & CHANDIGARH	2011	0.7	26.7	6.9	8.9	28.7	94.4	85.0	127.3	133.1	0.0
1468	1468	HARYANA DELHI & CHANDIGARH	2012	8.2	0.2	0.1	11.8	3.8	5.3	68.1	196.6	90.7	2.4
1469	1469	HARYANA DELHI & CHANDIGARH	2013	21.1	52.2	5.3	3.3	1.4	62.1	96.5	161.9	42.8	10.9
1470	1470	HARYANA DELHI & CHANDIGARH	2014	13.0	17.3	26.8	7.5	20.3	25.9	72.3	34.8	67.3	10.5
1471	1471	HARYANA DELHI & CHANDIGARH	2015	12.4	6.6	71.8	34.8	8.4	43.7	130.3	89.2	32.1	3.7
4													>

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

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 115 entries, 1357 to 1471
Data columns (total 20 columns):

Ducu	COTAMILIS (COC	ar zo coramis,.	
#	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
	67 164/4	->	

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	ОСТ
1357	1357	HARYANA DELHI & CHANDIGARH	1901	35.4	28.9	11.1	0.0	5.1	13.2	126.4	151.5	10.5	2.0
1358	1358	HARYANA DELHI & CHANDIGARH	1902	0.0	0.7	2.9	10.2	15.8	74.6	149.3	97.1	59.8	9.3
1359	1359	HARYANA 1359 DELHI & CHANDIGARH		14.7	0.5	2.3	0.5	8.5	8.6	151.6	138.2	97.7	4.0
1360	1360	HARYANA DELHI & CHANDIGARH	1904	7.6	0.7	48.0	0.5	29.3	34.3	109.7	162.9	102.3	1.5
1361	1361	HARYANA DELHI & CHANDIGARH	1905	44.8	20.8	14.0	1.3	7.4	20.1	93.6	23.1	92.6	0.0
1467	1467	HARYANA DELHI & CHANDIGARH	2011	0.7	26.7	6.9	8.9	28.7	94.4	85.0	127.3	133.1	0.0
1468	1468	HARYANA DELHI & CHANDIGARH	2012	8.2	0.2	0.1	11.8	3.8	5.3	68.1	196.6	90.7	2.4
1469	1469	HARYANA DELHI & CHANDIGARH	2013	21.1	52.2	5.3	3.3	1.4	62.1	96.5	161.9	42.8	10.9
1470	1470	HARYANA DELHI & CHANDIGARH	2014	13.0	17.3	26.8	7.5	20.3	25.9	72.3	34.8	67.3	10.5
1471	1471	HARYANA DELHI & CHANDIGARH	2015	12.4	6.6	71.8	34.8	8.4	43.7	130.3	89.2	32.1	3.7

115 rows × 20 columns

In [8]: df1.describe()

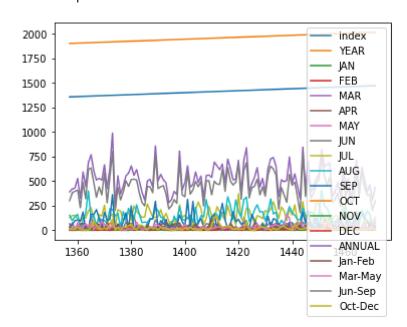
Out[8]:

	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	1414.000000	1958.000000	16.889565	17.433913	12.935652	7.633913	14.533913	48.€
std	33.341666	33.341666	15.514478	18.893422	15.251840	12.847533	15.900347	34.1
min	1357.000000	1901.000000	0.000000	0.000000	0.000000	0.000000	0.000000	2.1
25%	1385.500000	1929.500000	3.550000	2.250000	2.100000	0.800000	3.700000	24.6
50%	1414.000000	1958.000000	14.300000	12.100000	7.200000	2.800000	7.900000	43.7
75%	1442.500000	1986.500000	25.150000	27.850000	17.700000	8.750000	20.700000	66.6
max	1471.000000	2015.000000	66.500000	91.000000	71.800000	82.500000	72.900000	193.5

Data Visulaization

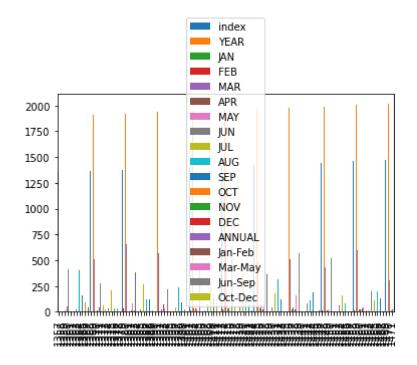
```
In [10]: df1.plot.line()
```

Out[10]: <AxesSubplot:>



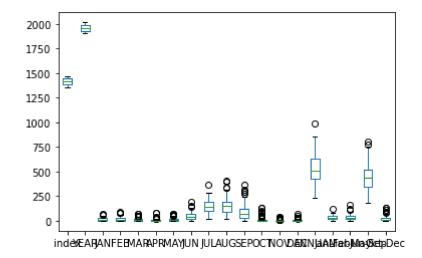
```
In [11]: df1.plot.bar()
```

Out[11]: <AxesSubplot:>



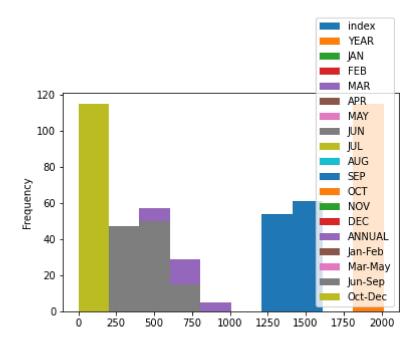
In [12]: df1.plot.box()

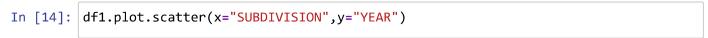
Out[12]: <AxesSubplot:>



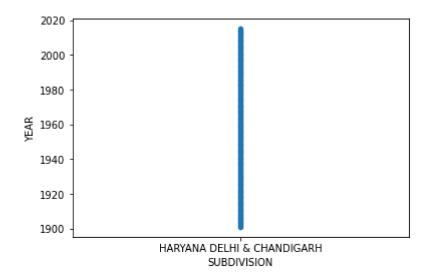
```
In [13]: df1.plot.hist()
```

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



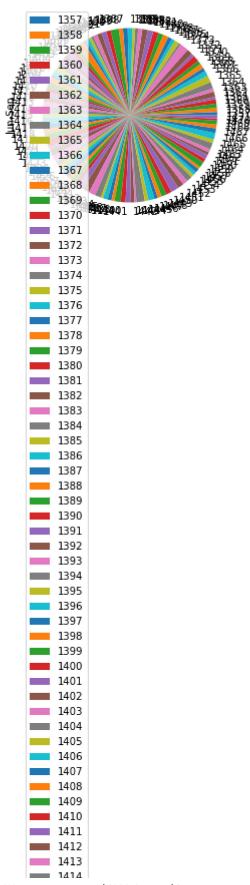


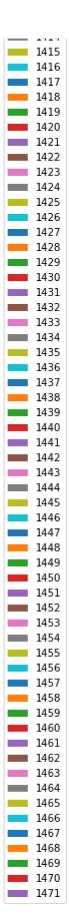
Out[14]: <AxesSubplot:xlabel='SUBDIVISION', ylabel='YEAR'>



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

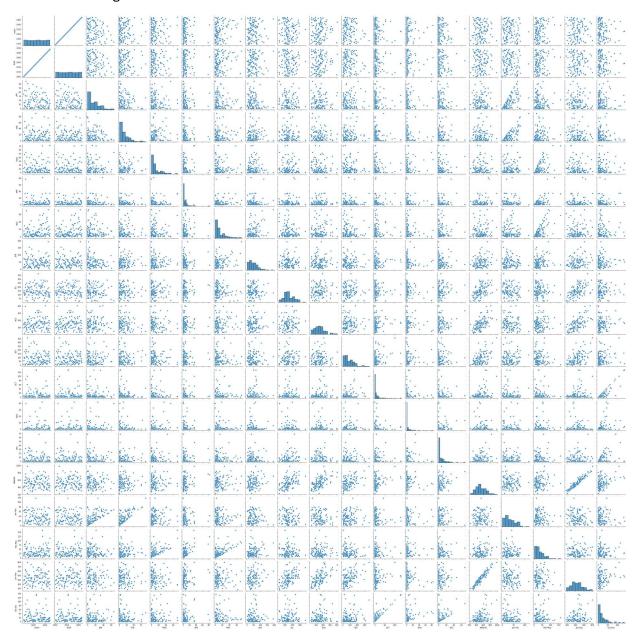
Out[15]: array([<AxesSubplot:ylabel='Jun-Sep'>], dtype=object)





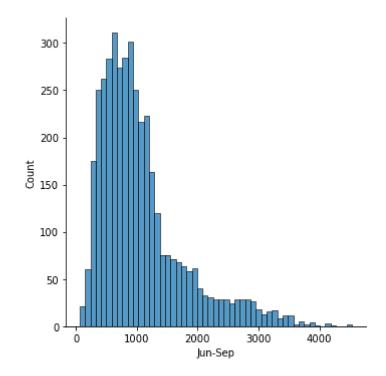
In [16]: sns.pairplot(df1)

Out[16]: <seaborn.axisgrid.PairGrid at 0x20593581310>



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

Out[17]: <seaborn.axisgrid.FacetGrid at 0x2059fb09730>



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

Out[18]: <AxesSubplot:>

