# **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")

aat

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

Andaman&Nicobar Islands ¶

In [3]: df=data.iloc[0:110]
df

Out[3]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	О
0	0	ANDAMAN & NICOBAR ISLANDS	1901	49.2	87.1	29.2	2.3	528.8	517.5	365.1	481.1	332.6	38
1	1	ANDAMAN & NICOBAR ISLANDS	1902	0.0	159.8	12.2	0.0	446.1	537.1	228.9	753.7	666.2	19
2	2	ANDAMAN & NICOBAR ISLANDS	1903	12.7	144.0	0.0	1.0	235.1	479.9	728.4	326.7	339.0	18
3	3	ANDAMAN & NICOBAR ISLANDS	1904	9.4	14.7	0.0	202.4	304.5	495.1	502.0	160.1	820.4	22
4	4	ANDAMAN & NICOBAR ISLANDS	1905	1.3	0.0	3.3	26.9	279.5	628.7	368.7	330.5	297.0	26
105	105	ANDAMAN & NICOBAR ISLANDS	2011	265.9	84.8	272.8	111.4	326.5	383.2	583.2	441.5	757.1	21
106	106	ANDAMAN & NICOBAR ISLANDS	2012	119.9	45.6	30.9	55.8	533.9	458.2	317.3	369.6	868.9	20
107	107	ANDAMAN & NICOBAR ISLANDS	2013	67.1	37.6	43.0	46.3	509.3	777.0	564.8	336.7	473.6	45
108	108	ANDAMAN & NICOBAR ISLANDS	2014	41.9	8.6	0.0	11.1	238.0	416.6	467.6	321.6	412.9	40
109	109	ANDAMAN & NICOBAR ISLANDS	2015	126.8	7.6	3.1	138.2	331.9	346.4	328.9	480.0	523.3	25

110 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
0	0	ANDAMAN & NICOBAR ISLANDS	1901	49.2	87.1	29.2	2.3	528.8	517.5	365.1	481.1	332.6	388.5
1	1	ANDAMAN & NICOBAR ISLANDS	1902	0.0	159.8	12.2	0.0	446.1	537.1	228.9	753.7	666.2	197.2
2	2	ANDAMAN & NICOBAR ISLANDS	1903	12.7	144.0	0.0	1.0	235.1	479.9	728.4	326.7	339.0	181.2
3	3	ANDAMAN & NICOBAR ISLANDS	1904	9.4	14.7	0.0	202.4	304.5	495.1	502.0	160.1	820.4	222.2
4	4	ANDAMAN & NICOBAR ISLANDS	1905	1.3	0.0	3.3	26.9	279.5	628.7	368.7	330.5	297.0	260.7

In [5]: df.tail()

### Out[5]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OC
105	105	ANDAMAN & NICOBAR ISLANDS	2011	265.9	84.8	272.8	111.4	326.5	383.2	583.2	441.5	757.1	212
106	106	ANDAMAN & NICOBAR ISLANDS	2012	119.9	45.6	30.9	55.8	533.9	458.2	317.3	369.6	868.9	209
107	107	ANDAMAN & NICOBAR ISLANDS	2013	67.1	37.6	43.0	46.3	509.3	777.0	564.8	336.7	473.6	455
108	108	ANDAMAN & NICOBAR ISLANDS	2014	41.9	8.6	0.0	11.1	238.0	416.6	467.6	321.6	412.9	402
109	109	ANDAMAN & NICOBAR ISLANDS	2015	126.8	7.6	3.1	138.2	331.9	346.4	328.9	480.0	523.3	252
4													•

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

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 110 entries, 0 to 109
Data columns (total 20 columns):

#	Column	Non-Null Count	Dtype
0	index	110 non-null	int64
1	SUBDIVISION	110 non-null	object
2	YEAR	110 non-null	int64
3	JAN	110 non-null	float64
4	FEB	110 non-null	float64
5	MAR	108 non-null	float64
6	APR	108 non-null	float64
7	MAY	109 non-null	float64
8	JUN	108 non-null	float64
9	JUL	108 non-null	float64
10	AUG	108 non-null	float64
11	SEP	107 non-null	float64
12	OCT	108 non-null	float64
13	NOV	108 non-null	float64
14	DEC	107 non-null	float64
15	ANNUAL	104 non-null	float64
16	Jan-Feb	110 non-null	float64
17	Mar-May	107 non-null	float64
18	Jun-Sep	107 non-null	float64
<b>1</b> 9	Oct-Dec	107 non-null	float64
dtyp	es: float64(1	7), int64(2), ob	oject(1)

memory usage: 17.3+ KB

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

# Out[7]:

	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 [8]: df1.describe()

#### Out[8]:

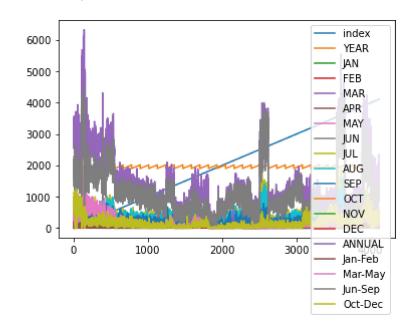
	index	YEAR	JAN	FEB	MAR	APR	MAY
cour	t 4116.000000	4116.000000	4116.000000	4116.000000	4116.000000	4116.000000	4116.000000
mea	n 2057.500000	1958.218659	18.938897	21.789431	27.319315	43.085520	85.682920
st	d 1188.331183	33.140898	33.574242	35.901220	46.936787	67.811512	123.211711
mi	n 0.000000	1901.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	6 1028.750000	1930.000000	0.600000	0.600000	1.000000	3.000000	8.600000
50%	<b>6</b> 2057.500000	1958.000000	6.000000	6.700000	7.800000	15.600000	36.400000
75%	<b>3</b> 086.250000	1987.000000	22.125000	26.800000	31.225000	49.825000	96.825000
ma	<b>x</b> 4115.000000	2015.000000	583.700000	403.500000	605.600000	595.100000	1168.600000

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

# **Data Visulaization**

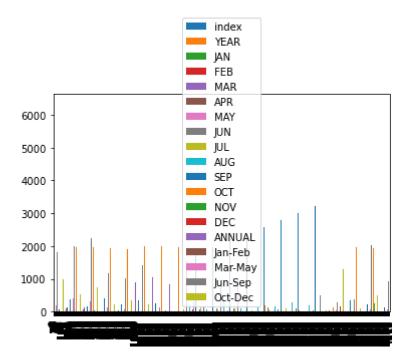
```
In [10]: df1.plot.line()
```

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



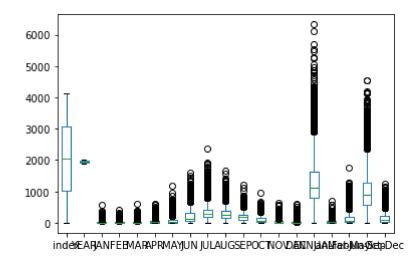
In [11]: df1.plot.bar()

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



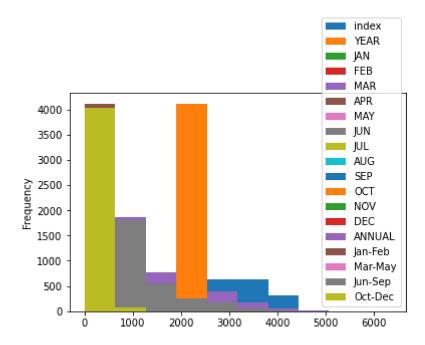


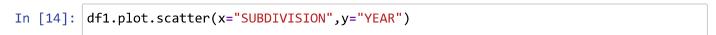
## Out[12]: <AxesSubplot:>



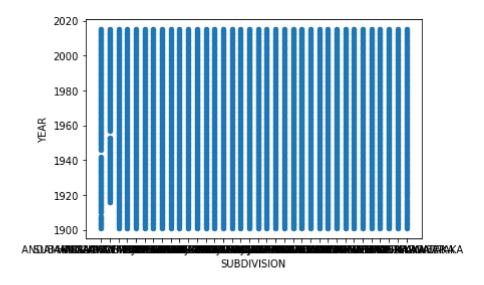
```
In [13]: df1.plot.hist()
```

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





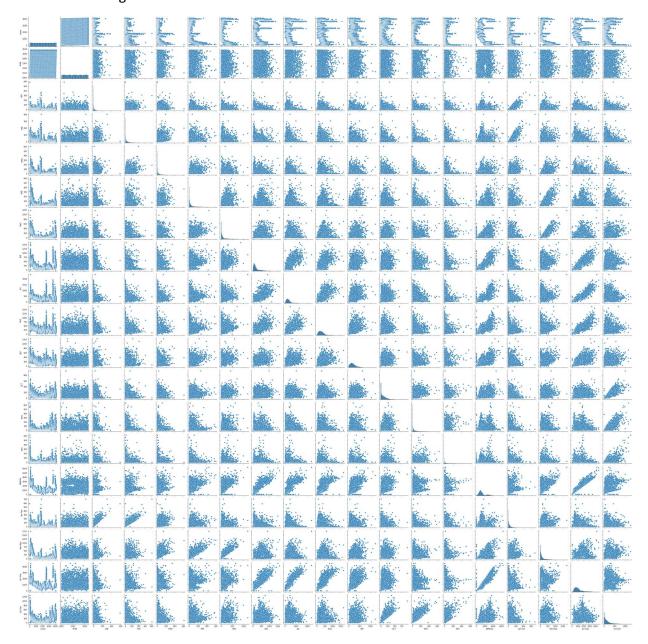
Out[14]: <AxesSubplot:xlabel='SUBDIVISION', ylabel='YEAR'>





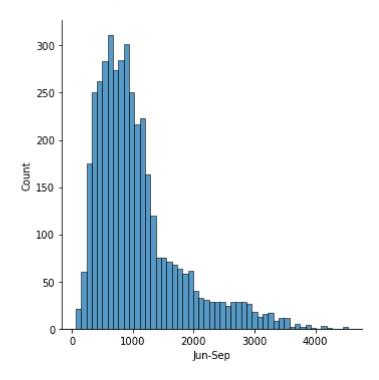
In [16]: sns.pairplot(df1)

Out[16]: <seaborn.axisgrid.PairGrid at 0x17170ee1e20>



```
In [18]: sns.displot(data["Jun-Sep"])
```

Out[18]: <seaborn.axisgrid.FacetGrid at 0x17170dfffd0>



In [19]: sns.heatmap(df1.corr())

### Out[19]: <AxesSubplot:>

