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 ANDAMAN & LAKSHADWEEP ANDAMAN & LAKSHADWEEP ANDAMAN & LAKSHADWEEP ANDAMAN & NICOBAR ISLANDS	ANDAMAN & 1901 ISLANDS ANDAMAN & 1902 ISLANDS ANDAMAN & 1902 ISLANDS ANDAMAN & 1903 ISLANDS ANDAMAN & 1903 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 ISLANDS 1902 0.0 159.8 ISLANDS 2 NICOBAR ISLANDS 1903 12.7 144.0 ISLANDS 1904 9.4 14.7 ISLANDS 1905 1.3 0.0 ISLANDS 1905 1.3 0.0 ISLANDS 1905 1.3 0.0 ISLANDS 1905 1.3 0.0 ISLANDS 1905 1.3 2.8 4111 LAKSHADWEEP 2011 5.1 2.8 4112 LAKSHADWEEP 2012 19.2 0.1 4113 LAKSHADWEEP 2014 53.2 16.1	ANDAMAN & NICOBAR ISLANDS 1901 49.2 87.1 29.2 SISLANDS 1902 0.0 159.8 12.2 ANDAMAN & NICOBAR ISLANDS 1903 12.7 144.0 0.0 ISLANDS 1904 9.4 14.7 0.0 ANDAMAN & NICOBAR ISLANDS 1904 9.4 14.7 0.0 SISLANDS 1905 1.3 0.0 3.3 ISLANDS 1905 1.3 0.0 3.3 ISLANDS 1905 1.3 0.0 3.3 ISLANDS 1905 1.3 1905 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 4 NICOBAR 1905 1.3 0.0 3.3 26.9 ISLANDS 3 NICOBAR 1905 1.3 0.0 3.3 26.9 ISLANDS 4 NICOBAR 1905 1.3 2.8 3.1 85.9 4112 LAKSHADWEEP 2012 19.2 0.1 1.6 76.8 4113 LAKSHADWEEP 2014 53.2 16.1 4.4 14.9	ANDAMAN & ISLANDS ISLA	ANDAMAN & NICOBAR ISLANDS 1901 49.2 87.1 29.2 2.3 528.8 517.5 15.4 1	ANDAMAN & NICOBAR ISLANDS	ANDAMAN & 1901 49.2 87.1 29.2 2.3 528.8 517.5 365.1 481.1 SLANDS 1902 0.0 159.8 12.2 0.0 446.1 537.1 228.9 753.7 ISLANDS 1903 12.7 144.0 0.0 1.0 235.1 479.9 728.4 326.7 ISLANDS 1904 9.4 14.7 0.0 202.4 304.5 495.1 502.0 160.1 1914 1915 1915 1915 1915 1915 1915 191	ANDAMAN & NICOBAR 1901 49.2 87.1 29.2 2.3 528.8 517.5 365.1 481.1 332.6 ANDAMAN & NICOBAR 1902 0.0 159.8 12.2 0.0 446.1 537.1 228.9 753.7 666.2 ISLANDS NICOBAR 1903 12.7 144.0 0.0 1.0 235.1 479.9 728.4 326.7 339.0 ISLANDS NICOBAR 1904 9.4 14.7 0.0 202.4 304.5 495.1 502.0 160.1 820.4 ISLANDS NICOBAR 1905 1.3 0.0 3.3 26.9 279.5 628.7 368.7 330.5 297.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1

4116 rows × 20 columns

UTTARAKHAND

In [3]: df=data.iloc[1242:1357]
df

Out[3]:

index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	C
1242	UTTARAKHAND	1901	134.5	81.4	44.5	5.9	60.8	33.6	381.1	612.3	167.1	
1243	UTTARAKHAND	1902	0.0	17.0	52.2	63.7	52.1	113.1	444.1	327.5	220.4	3
1244	UTTARAKHAND	1903	68.0	7.9	87.6	10.3	37.5	83.0	251.6	442.7	249.3	Ę
1245	UTTARAKHAND	1904	40.0	5.2	78.3	13.6	61.1	180.1	449.6	417.2	174.1	
1246	UTTARAKHAND	1905	115.4	80.7	99.8	26.1	70.3	111.5	299.9	349.5	129.5	
1352	UTTARAKHAND	2011	30.9	65.2	18.0	30.9	84.2	223.1	433.3	523.7	148.4	
1353	UTTARAKHAND	2012	38.8	11.9	28.1	39.2	9.1	46.0	387.1	419.5	220.6	
1354	UTTARAKHAND	2013	73.0	188.3	22.0	24.7	18.2	488.9	413.4	359.4	111.3	2
1355	UTTARAKHAND	2014	45.9	99.9	68.4	37.6	52.9	62.9	462.7	264.2	107.9	۷
1356	UTTARAKHAND	2015	54.5	62.6	127.3	57.3	38.0	186.6	337.0	305.3	52.6	,
	1242 1243 1244 1245 1246 1352 1353 1354	1242 UTTARAKHAND 1243 UTTARAKHAND 1244 UTTARAKHAND 1245 UTTARAKHAND 1246 UTTARAKHAND 1352 UTTARAKHAND 1353 UTTARAKHAND 1354 UTTARAKHAND 1355 UTTARAKHAND	1242 UTTARAKHAND 1901 1243 UTTARAKHAND 1902 1244 UTTARAKHAND 1903 1245 UTTARAKHAND 1904 1246 UTTARAKHAND 1905 1352 UTTARAKHAND 2011 1353 UTTARAKHAND 2012 1354 UTTARAKHAND 2013 1355 UTTARAKHAND 2014	1242 UTTARAKHAND 1901 134.5 1243 UTTARAKHAND 1902 0.0 1244 UTTARAKHAND 1903 68.0 1245 UTTARAKHAND 1904 40.0 1246 UTTARAKHAND 1905 115.4 1352 UTTARAKHAND 2011 30.9 1353 UTTARAKHAND 2012 38.8 1354 UTTARAKHAND 2013 73.0 1355 UTTARAKHAND 2014 45.9	1242 UTTARAKHAND 1901 134.5 81.4 1243 UTTARAKHAND 1902 0.0 17.0 1244 UTTARAKHAND 1903 68.0 7.9 1245 UTTARAKHAND 1904 40.0 5.2 1246 UTTARAKHAND 1905 115.4 80.7 1352 UTTARAKHAND 2011 30.9 65.2 1353 UTTARAKHAND 2012 38.8 11.9 1354 UTTARAKHAND 2013 73.0 188.3 1355 UTTARAKHAND 2014 45.9 99.9	1242 UTTARAKHAND 1901 134.5 81.4 44.5 1243 UTTARAKHAND 1902 0.0 17.0 52.2 1244 UTTARAKHAND 1903 68.0 7.9 87.6 1245 UTTARAKHAND 1904 40.0 5.2 78.3 1246 UTTARAKHAND 1905 115.4 80.7 99.8 1352 UTTARAKHAND 2011 30.9 65.2 18.0 1353 UTTARAKHAND 2012 38.8 11.9 28.1 1354 UTTARAKHAND 2013 73.0 188.3 22.0 1355 UTTARAKHAND 2014 45.9 99.9 68.4	1242 UTTARAKHAND 1901 134.5 81.4 44.5 5.9 1243 UTTARAKHAND 1902 0.0 17.0 52.2 63.7 1244 UTTARAKHAND 1903 68.0 7.9 87.6 10.3 1245 UTTARAKHAND 1904 40.0 5.2 78.3 13.6 1246 UTTARAKHAND 1905 115.4 80.7 99.8 26.1 1352 UTTARAKHAND 2011 30.9 65.2 18.0 30.9 1353 UTTARAKHAND 2012 38.8 11.9 28.1 39.2 1354 UTTARAKHAND 2013 73.0 188.3 22.0 24.7 1355 UTTARAKHAND 2014 45.9 99.9 68.4 37.6	1242 UTTARAKHAND 1901 134.5 81.4 44.5 5.9 60.8 1243 UTTARAKHAND 1902 0.0 17.0 52.2 63.7 52.1 1244 UTTARAKHAND 1903 68.0 7.9 87.6 10.3 37.5 1245 UTTARAKHAND 1904 40.0 5.2 78.3 13.6 61.1 1246 UTTARAKHAND 1905 115.4 80.7 99.8 26.1 70.3 1352 UTTARAKHAND 2011 30.9 65.2 18.0 30.9 84.2 1353 UTTARAKHAND 2012 38.8 11.9 28.1 39.2 9.1 1354 UTTARAKHAND 2013 73.0 188.3 22.0 24.7 18.2 1355 UTTARAKHAND 2014 45.9 99.9 68.4 37.6 52.9	1242 UTTARAKHAND 1901 134.5 81.4 44.5 5.9 60.8 33.6 1243 UTTARAKHAND 1902 0.0 17.0 52.2 63.7 52.1 113.1 1244 UTTARAKHAND 1903 68.0 7.9 87.6 10.3 37.5 83.0 1245 UTTARAKHAND 1904 40.0 5.2 78.3 13.6 61.1 180.1 1246 UTTARAKHAND 1905 115.4 80.7 99.8 26.1 70.3 111.5	1242 UTTARAKHAND 1901 134.5 81.4 44.5 5.9 60.8 33.6 381.1 1243 UTTARAKHAND 1902 0.0 17.0 52.2 63.7 52.1 113.1 444.1 1244 UTTARAKHAND 1903 68.0 7.9 87.6 10.3 37.5 83.0 251.6 1245 UTTARAKHAND 1904 40.0 5.2 78.3 13.6 61.1 180.1 449.6 1246 UTTARAKHAND 1905 115.4 80.7 99.8 26.1 70.3 111.5 299.9 </th <th>1242 UTTARAKHAND 1901 134.5 81.4 44.5 5.9 60.8 33.6 381.1 612.3 1243 UTTARAKHAND 1902 0.0 17.0 52.2 63.7 52.1 113.1 444.1 327.5 1244 UTTARAKHAND 1903 68.0 7.9 87.6 10.3 37.5 83.0 251.6 442.7 1245 UTTARAKHAND 1904 40.0 5.2 78.3 13.6 61.1 180.1 449.6 417.2 1246 UTTARAKHAND 1905 115.4 80.7 99.8 26.1 70.3 111.5 299.9 349.5 </th> <th>1242 UTTARAKHAND 1901 134.5 81.4 44.5 5.9 60.8 33.6 381.1 612.3 167.1 1243 UTTARAKHAND 1902 0.0 17.0 52.2 63.7 52.1 113.1 444.1 327.5 220.4 1244 UTTARAKHAND 1903 68.0 7.9 87.6 10.3 37.5 83.0 251.6 442.7 249.3 1245 UTTARAKHAND 1904 40.0 5.2 78.3 13.6 61.1 180.1 449.6 417.2 174.1 1246 UTTARAKHAND 1905 115.4 80.7 99.8 26.1 70.3 111.5 299.9 349.5 129.5 </th>	1242 UTTARAKHAND 1901 134.5 81.4 44.5 5.9 60.8 33.6 381.1 612.3 1243 UTTARAKHAND 1902 0.0 17.0 52.2 63.7 52.1 113.1 444.1 327.5 1244 UTTARAKHAND 1903 68.0 7.9 87.6 10.3 37.5 83.0 251.6 442.7 1245 UTTARAKHAND 1904 40.0 5.2 78.3 13.6 61.1 180.1 449.6 417.2 1246 UTTARAKHAND 1905 115.4 80.7 99.8 26.1 70.3 111.5 299.9 349.5	1242 UTTARAKHAND 1901 134.5 81.4 44.5 5.9 60.8 33.6 381.1 612.3 167.1 1243 UTTARAKHAND 1902 0.0 17.0 52.2 63.7 52.1 113.1 444.1 327.5 220.4 1244 UTTARAKHAND 1903 68.0 7.9 87.6 10.3 37.5 83.0 251.6 442.7 249.3 1245 UTTARAKHAND 1904 40.0 5.2 78.3 13.6 61.1 180.1 449.6 417.2 174.1 1246 UTTARAKHAND 1905 115.4 80.7 99.8 26.1 70.3 111.5 299.9 349.5 129.5

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	OC
1242	1242	UTTARAKHAND	1901	134.5	81.4	44.5	5.9	60.8	33.6	381.1	612.3	167.1	16
1243	1243	UTTARAKHAND	1902	0.0	17.0	52.2	63.7	52.1	113.1	444.1	327.5	220.4	31
1244	1244	UTTARAKHAND	1903	68.0	7.9	87.6	10.3	37.5	83.0	251.6	442.7	249.3	57
1245	1245	UTTARAKHAND	1904	40.0	5.2	78.3	13.6	61.1	180.1	449.6	417.2	174.1	6
1246	1246	UTTARAKHAND	1905	115.4	80.7	99.8	26.1	70.3	111.5	299.9	349.5	129.5	0
4													•

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

Out[5]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	01
1352	1352	UTTARAKHAND	2011	30.9	65.2	18.0	30.9	84.2	223.1	433.3	523.7	148.4	(
1353	1353	UTTARAKHAND	2012	38.8	11.9	28.1	39.2	9.1	46.0	387.1	419.5	220.6	4
1354	1354	UTTARAKHAND	2013	73.0	188.3	22.0	24.7	18.2	488.9	413.4	359.4	111.3	29
1355	1355	UTTARAKHAND	2014	45.9	99.9	68.4	37.6	52.9	62.9	462.7	264.2	107.9	4(
1356	1356	UTTARAKHAND	2015	54.5	62.6	127.3	57.3	38.0	186.6	337.0	305.3	52.6	16

In [6]: df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 115 entries, 1242 to 1356
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
44	C1+C4/4	7) : -+ < 4 / 2 \ - 1	

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	C
1242	1242	UTTARAKHAND	1901	134.5	81.4	44.5	5.9	60.8	33.6	381.1	612.3	167.1	
1243	1243	UTTARAKHAND	1902	0.0	17.0	52.2	63.7	52.1	113.1	444.1	327.5	220.4	3
1244	1244	UTTARAKHAND	1903	68.0	7.9	87.6	10.3	37.5	83.0	251.6	442.7	249.3	Ę
1245	1245	UTTARAKHAND	1904	40.0	5.2	78.3	13.6	61.1	180.1	449.6	417.2	174.1	
1246	1246	UTTARAKHAND	1905	115.4	80.7	99.8	26.1	70.3	111.5	299.9	349.5	129.5	
1352	1352	UTTARAKHAND	2011	30.9	65.2	18.0	30.9	84.2	223.1	433.3	523.7	148.4	
1353	1353	UTTARAKHAND	2012	38.8	11.9	28.1	39.2	9.1	46.0	387.1	419.5	220.6	
1354	1354	UTTARAKHAND	2013	73.0	188.3	22.0	24.7	18.2	488.9	413.4	359.4	111.3	2
1355	1355	UTTARAKHAND	2014	45.9	99.9	68.4	37.6	52.9	62.9	462.7	264.2	107.9	2
1356	1356	UTTARAKHAND	2015	54.5	62.6	127.3	57.3	38.0	186.6	337.0	305.3	52.6	•

115 rows × 20 columns

In [8]: df1.describe()

Out[8]:

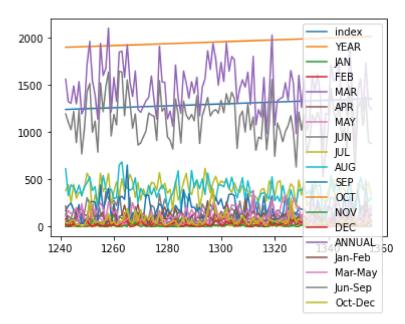
	index	YEAR	JAN	FEB	MAR	APR	MAY	
count	115.000000	115.000000	115.000000	115.000000	115.000000	115.000000	115.000000	115.
mean	1299.000000	1958.000000	53.797391	63.452174	57.272174	35.166087	55.338261	162.
std	33.341666	33.341666	40.887384	44.040532	42.438752	24.116540	36.597919	86.
min	1242.000000	1901.000000	0.000000	0.000000	0.000000	1.100000	3.600000	33.
25%	1270.500000	1929.500000	21.400000	27.950000	22.850000	18.250000	28.050000	104.
50%	1299.000000	1958.000000	49.700000	60.100000	47.700000	30.700000	50.500000	139.
75%	1327.500000	1986.500000	76.200000	88.100000	80.600000	51.200000	71.450000	211.
max	1356.000000	2015.000000	211.400000	188.300000	190.300000	132.900000	270.200000	488.
4								

In [9]: | df1.columns

Data Visulaization

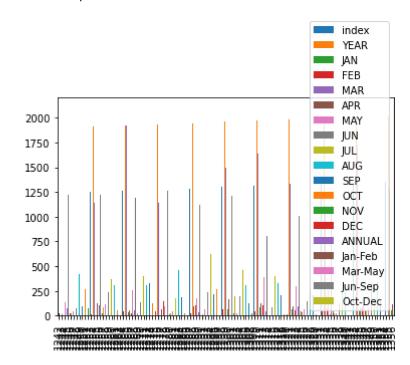
```
In [10]: df1.plot.line()
```

Out[10]: <AxesSubplot:>



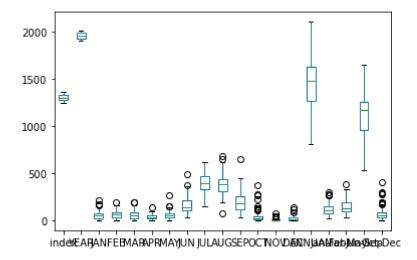
In [11]: df1.plot.bar()

Out[11]: <AxesSubplot:>



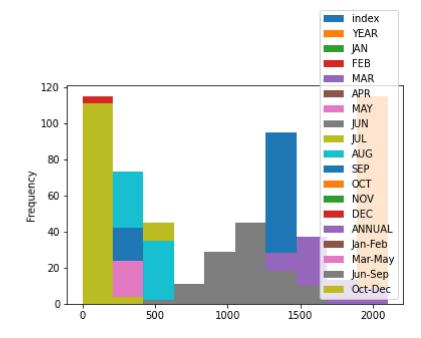
In [12]: df1.plot.box()

Out[12]: <AxesSubplot:>



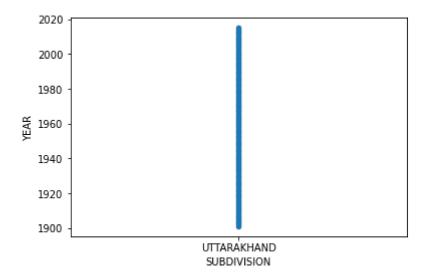
In [13]: df1.plot.hist()

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



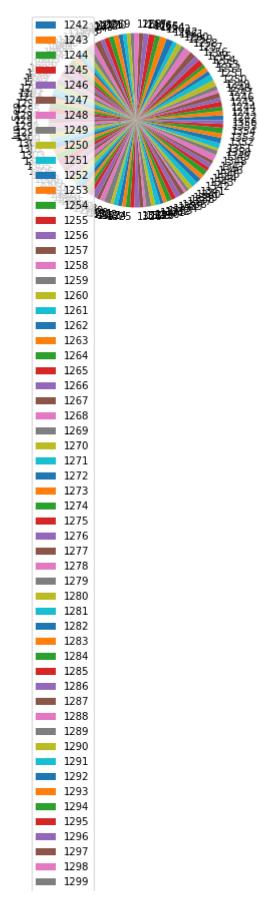
In [14]: df1.plot.scatter(x="SUBDIVISION",y="YEAR")

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)



```
1300
 1301
  1302
 1303
1304
 1305
1306
1307
1308
1309
1310
 1311
 1312
1313
1314
 1315
1316
1317
1318
1319
 1320
  1321
1322
 1323
1324
1325
1326
1327
1328
1329
1330
 1331
1332
1333
 1334
 1335
1336
1337
1338
1339
1340
 1341
 1342
1343
1344
 1345
1346
1347
1348
1349
1350
  1351
1352
 1353
 1354
1355
1356
```

```
In [ ]: sns.pairplot(df1)
In [ ]: sns.displot(data["Jun-Sep"])
```

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