

Import Libraies

```
In [1]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

In [2]: df=pd.read_csv(r"C:\Users\user\Downloads\FP2_RainFall\rainfall in india 1901-2015.csv")[2852:2965]
df
```

Out[2]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	ANNUAL
2852	2852	VIDARBHA	1901	36.8	39.9	30.9	26.1	7.3	129.7	295.3	368.8	123.4	35.2	0.0	0.0	1093.5
2853	2853	VIDARBHA	1902	1.6	0.1	0.0	6.5	4.1	38.0	270.7	204.7	150.9	29.6	16.1	26.7	748.5
2854	2854	VIDARBHA	1903	5.2	4.0	0.1	2.5	37.8	121.2	475.5	325.5	154.8	100.8	2.0	0.0	1229.4
2855	2855	VIDARBHA	1904	4.3	2.4	12.9	0.2	14.8	148.9	158.3	151.8	196.9	61.7	0.0	0.9	753.2
2856	2856	VIDARBHA	1905	7.3	12.7	12.4	16.2	14.0	81.0	254.5	216.3	321.3	6.0	0.2	0.0	941.8
...
2961	2961	VIDARBHA	2010	14.0	1.9	7.5	0.3	1.9	144.8	454.5	384.0	238.0	51.0	37.0	3.2	1338.7
2962	2962	VIDARBHA	2011	0.0	1.2	0.1	7.7	0.6	137.9	247.1	302.8	191.0	4.7	0.0	0.0	893.2
2963	2963	VIDARBHA	2012	3.1	0.1	0.0	0.6	0.2	125.5	370.5	316.2	249.4	34.9	7.3	0.0	1107.7
2964	2964	VIDARBHA	2013	6.6	13.0	3.8	2.8	0.5	366.7	535.5	326.1	131.7	133.5	0.0	0.0	1520.0
2965	2965	VIDARBHA	2014	1.2	18.3	49.6	2.6	4.0	63.3	337.6	191.7	224.9	17.3	6.2	2.3	919.7

114 rows × 20 columns

Data Cleaning and Preprocessing

In [3]: df.dropna()

Out[3]:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	ANNUAL
2852	2852	VIDARBHA	1901	36.8	39.9	30.9	26.1	7.3	129.7	295.3	368.8	123.4	35.2	0.0	0.0	1093.5
2853	2853	VIDARBHA	1902	1.6	0.1	0.0	6.5	4.1	38.0	270.7	204.7	150.9	29.6	16.1	26.7	748.5
2854	2854	VIDARBHA	1903	5.2	4.0	0.1	2.5	37.8	121.2	475.5	325.5	154.8	100.8	2.0	0.0	1229.4
2855	2855	VIDARBHA	1904	4.3	2.4	12.9	0.2	14.8	148.9	158.3	151.8	196.9	61.7	0.0	0.9	753.2
2856	2856	VIDARBHA	1905	7.3	12.7	12.4	16.2	14.0	81.0	254.5	216.3	321.3	6.0	0.2	0.0	941.8
...
2961	2961	VIDARBHA	2010	14.0	1.9	7.5	0.3	1.9	144.8	454.5	384.0	238.0	51.0	37.0	3.2	1338.7
2962	2962	VIDARBHA	2011	0.0	1.2	0.1	7.7	0.6	137.9	247.1	302.8	191.0	4.7	0.0	0.0	893.2
2963	2963	VIDARBHA	2012	3.1	0.1	0.0	0.6	0.2	125.5	370.5	316.2	249.4	34.9	7.3	0.0	1107.7
2964	2964	VIDARBHA	2013	6.6	13.0	3.8	2.8	0.5	366.7	535.5	326.1	131.7	133.5	0.0	0.0	1520.0
2965	2965	VIDARBHA	2014	1.2	18.3	49.6	2.6	4.0	63.3	337.6	191.7	224.9	17.3	6.2	2.3	919.7

114 rows × 20 columns



In [4]: df.columns

Out[4]: 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')

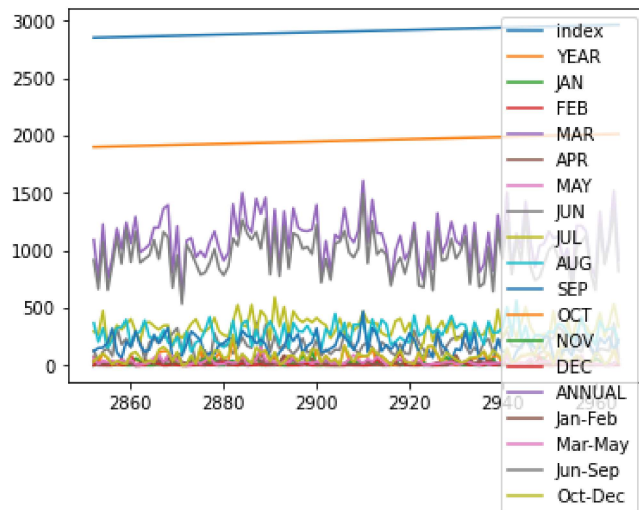
In [5]: df.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 114 entries, 2852 to 2965
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
```

Line chart

```
In [6]: df.plot.line()
```

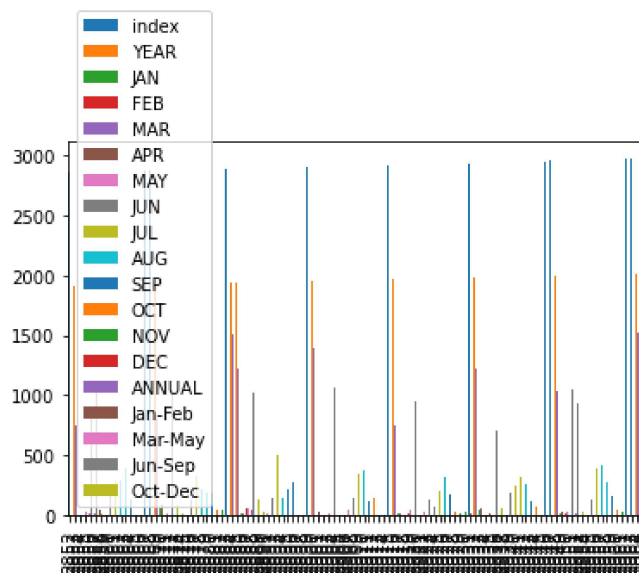
```
Out[6]: <AxesSubplot:>
```



Bar chart

```
In [7]: df.plot.bar()
```

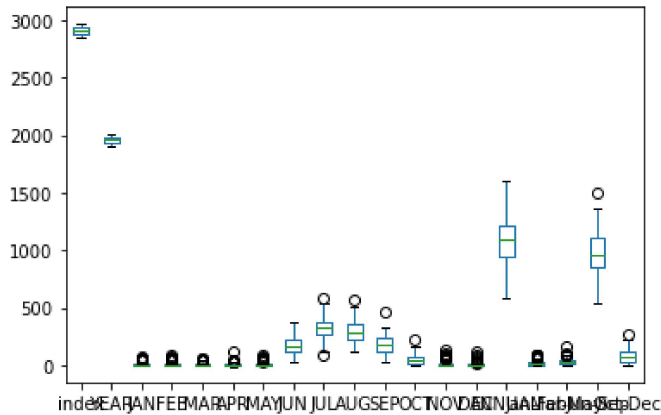
```
Out[7]: <AxesSubplot:>
```



Box chart

```
In [8]: df.plot.box()
```

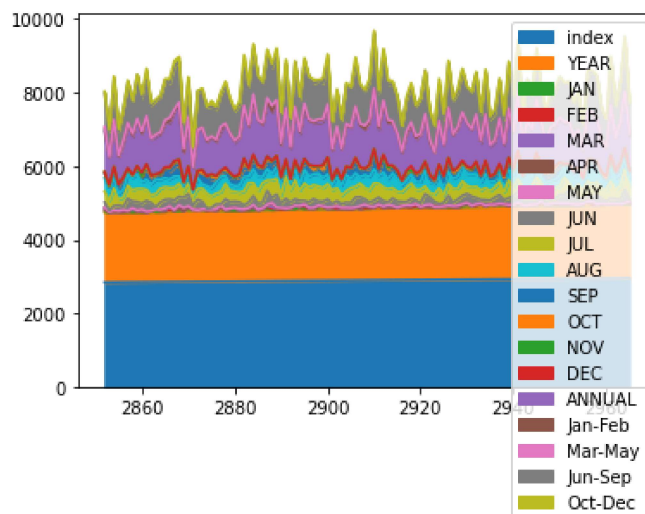
```
Out[8]: <AxesSubplot:>
```



Area chart

```
In [9]: df.plot.area()
```

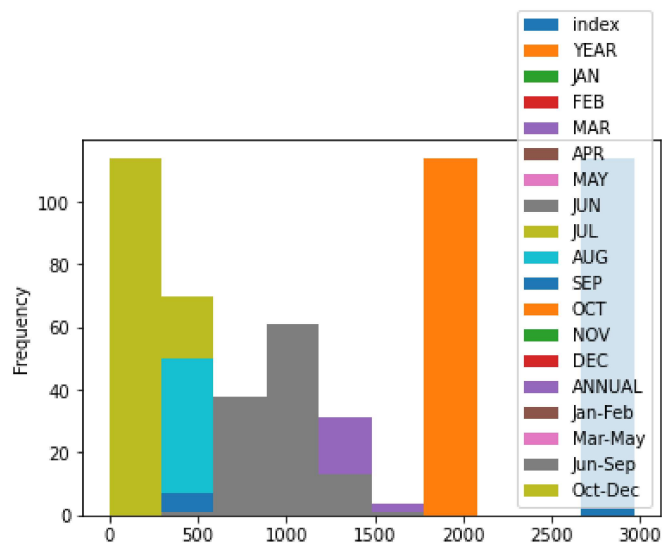
```
Out[9]: <AxesSubplot:>
```



Histogram

```
In [10]: df.plot.hist()
```

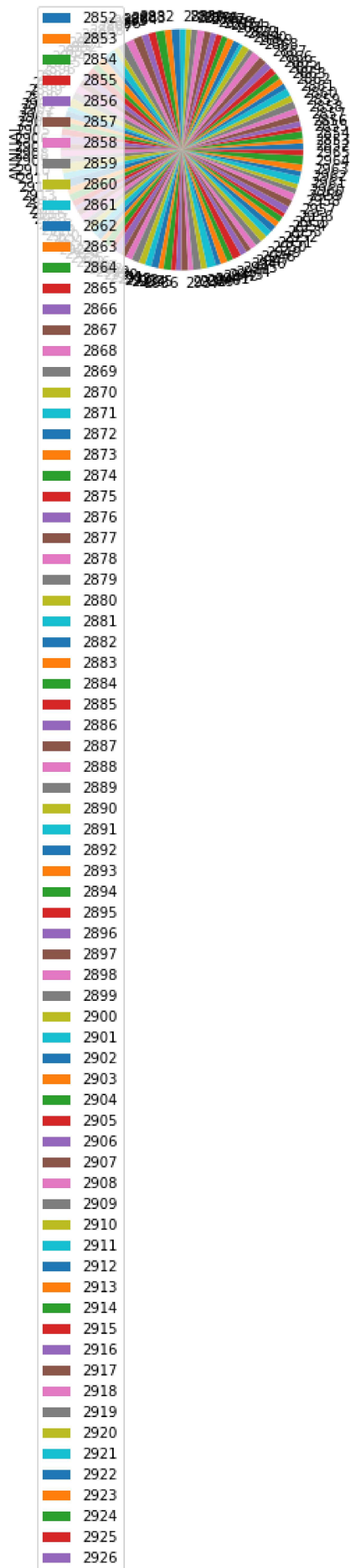
```
Out[10]: <AxesSubplot:ylabel='Frequency'>
```

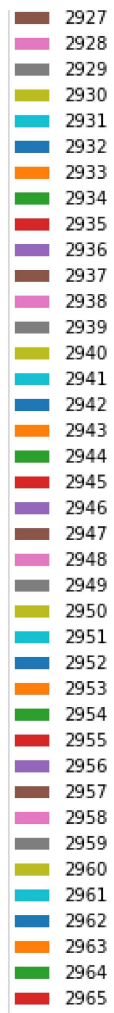


pie chart

```
In [11]: df.plot.pie(y="ANNUAL")
```

```
Out[11]: <AxesSubplot:ylabel='ANNUAL'>
```

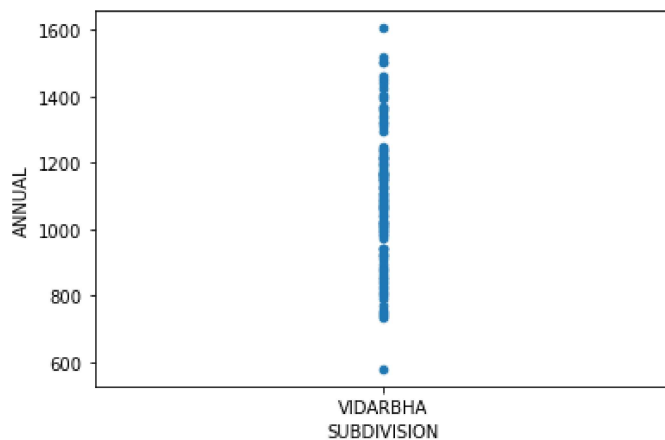





Scatter chart

```
In [12]: df.plot.scatter(y='ANNUAL',x='SUBDIVISION')
```

```
Out[12]: <AxesSubplot:xlabel='SUBDIVISION', ylabel='ANNUAL'>
```



```
In [13]: df.describe()
```

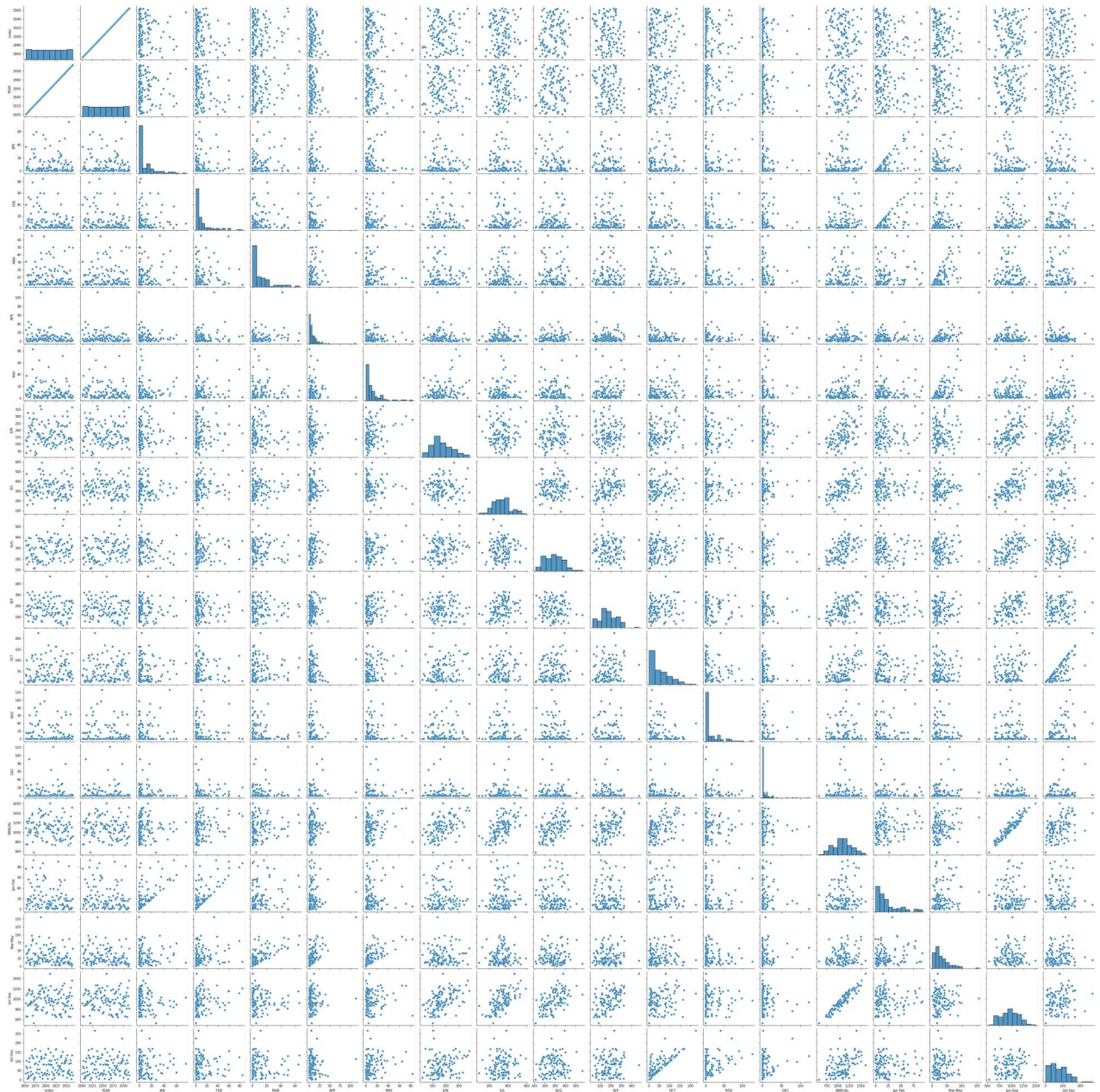
Out[13]:

	index	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL
count	114.000000	114.000000	114.000000	114.000000	114.000000	114.000000	114.000000	114.000000	114.000000
mean	2908.500000	1957.500000	10.425439	12.046491	11.394737	9.271930	11.540351	172.867544	331.114912
std	33.052988	33.052988	15.099416	17.351581	14.737129	13.117728	14.668368	77.482313	91.238694
min	2852.000000	1901.000000	0.000000	0.000000	0.000000	0.000000	0.000000	21.400000	88.100000
25%	2880.250000	1929.250000	0.350000	1.100000	0.925000	2.625000	2.225000	116.325000	264.125000
50%	2908.500000	1957.500000	3.750000	4.850000	4.850000	5.600000	6.150000	156.250000	335.450000
75%	2936.750000	1985.750000	14.575000	14.350000	15.725000	11.800000	14.950000	231.200000	380.075000
max	2965.000000	2014.000000	74.900000	84.900000	65.900000	112.700000	83.100000	377.500000	591.000000

EDA AND VISUALIZATION

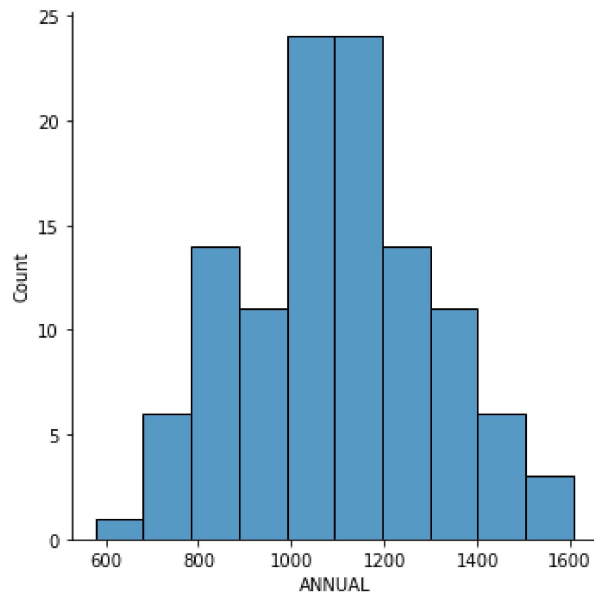
```
In [14]: sns.pairplot(df)
```

```
Out[14]: <seaborn.axisgrid.PairGrid at 0x1f97850fdf0>
```



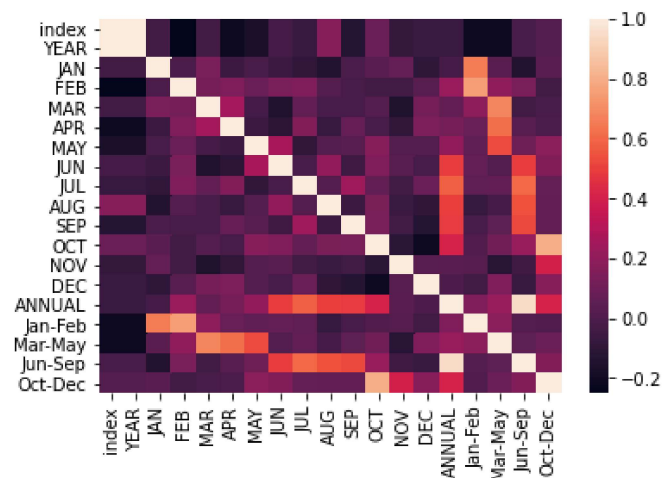
```
In [15]: sns.displot(df['ANNUAL'])
```

```
Out[15]: <seaborn.axisgrid.FacetGrid at 0x1f905cceb50>
```



```
In [16]: sns.heatmap(df.corr())
```

```
Out[16]: <AxesSubplot:>
```



```
In [ ]:
```