Day 6

Instagram Dataset

In [1]:

```
import numpy as np
import pandas as pd
```

In [2]:

```
d=pd.read_csv(r"c:\Users\user\Downloads\5_instagram data.csv")
d
```

Out[2]:

	Impressions	From Home	From Hashtags	From Explore	From Other	Saves	Comments	Shares	Likes	Profile Visits
0	3920	2586	1028	619	56	98	9	5	162	35
1	5394	2727	1838	1174	78	194	7	14	224	48
2	4021	2085	1188	0	533	41	11	1	131	62
3	4528	2700	621	932	73	172	10	7	213	23
4	2518	1704	255	279	37	96	5	4	123	8
...	...	...	...	...	...	...	...	...	...	...
114	13700	5185	3041	5352	77	573	2	38	373	73
115	5731	1923	1368	2266	65	135	4	1	148	20
116	4139	1133	1538	1367	33	36	0	1	92	34
117	32695	11815	3147	17414	170	1095	2	75	549	148
118	36919	13473	4176	16444	2547	653	5	26	443	611

119 rows × 13 columns

Mean,median,mode,describe

In [3]:

```
data=pd.DataFrame(d[['Saves','Shares']][0:500])
data
```

Out[3]:

	Saves	Shares
0	98	5
1	194	14
2	41	1
3	172	7
4	96	4
...	...	...
114	573	38
115	135	1
116	36	1
117	1095	75
118	653	26

119 rows × 2 columns

In [4]:

```
print(data.mean())
```

```
Saves     153.310924
Shares      9.361345
dtype: float64
```

In [5]:

```
print(data.median())
```

```
Saves     109.0
Shares      6.0
dtype: float64
```

In [6]:

```
data.fillna(value=1)
```

Out[6]:

	Saves	Shares
0	98	5
1	194	14
2	41	1
3	172	7
4	96	4
...	...	...
114	573	38
115	135	1
116	36	1
117	1095	75
118	653	26

119 rows × 2 columns

In [7]:

```
print(data.mode())
```

```
   Saves  Shares
0     40     3.0
1    135     NaN
2    144     NaN
```

In [8]:

```
print(data.describe())
```

```
            Saves      Shares
count   119.000000  119.000000
mean    153.310924    9.361345
std     156.317731   10.089205
min      22.000000    0.000000
25%      65.000000    3.000000
50%     109.000000    6.000000
75%     169.000000   13.500000
max    1095.000000   75.000000
```

Sum,cumsum,count,min,max

In [9]:

```
print(data.sum())
```

```
Saves     18244
Shares     1114
dtype: int64
```

In [10]:

```
print(data.cumsum())
```

```
     Saves  Shares
0       98       5
1      292      19
2      333      20
3      505      27
4      601      31
..     ...     ...
114  16325    1011
115  16460    1012
116  16496    1013
117  17591    1088
118  18244    1114

[119 rows x 2 columns]
```

In [11]:

```
print(data.count())
```

```
Saves     119
Shares    119
dtype: int64
```

In [12]:

```
print(data.min())
```

```
Saves     22
Shares     0
dtype: int64
```

In [13]:

```
print(data.max())
```

```
Saves     1095
Shares      75
dtype: int64
```

covariance and correlation (spearman and pearsons)

In [14]:

```
data1=data['Saves'][0:10]
data1
```

Out[14]:

```
0     98
1    194
2     41
3    172
4     96
5     74
6     22
7    135
8    155
9    122
Name: Saves, dtype: int64
```

In [15]:

```
data2=data['Shares'][0:10]
data2
```

Out[15]:

```
0     5
1    14
2     1
3     7
4     4
5    10
6     1
7     9
8     8
9     3
Name: Shares, dtype: int64
```

In [16]:

```
from numpy import cov
print(cov(data1,data2))
```

```
[[3091.87777778  171.35555556]
 [ 171.35555556   17.51111111]]
```

In [17]:

```
from scipy.stats import pearsonr
print(pearsonr(data1,data2))
```

```
(0.7364278065804685, 0.015141913073821655)
```

In [18]:

```
from scipy.stats import spearmanr
print(spearmanr(data1,data2))
```

```
SpearmanrResult(correlation=0.6565379871744699, pvalue=0.039204386332556754)
```

In [ ]:

```

```