

EDA

In [48]:

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

In [49]:

```
a=pd.read_csv(r"C:\Users\user\Downloads\5_Instagram data - 5_Instagram data.csv")
a
```

Out[49]:

	Impressions	From Home	From Hashtags	From Explore	From Other	Saves	Comments	Shares	Likes	Profile Visits	Follows
0	3920	2586	1028	619	56	98	9	5	162	35	2
1	5394	2727	1838	1174	78	194	7	14	224	48	10
2	4021	2085	1188	0	533	41	11	1	131	62	12
3	4528	2700	621	932	73	172	10	7	213	23	8
4	2518	1704	255	279	37	96	5	4	123	8	0
...
114	13700	5185	3041	5352	77	573	2	38	373	73	80
115	5731	1923	1368	2266	65	135	4	1	148	20	18

	Impressions	From Home	From Hashtags	From Explore	From Other	Saves	Comments	Shares	Likes	Profile Visits	Follows
116	4139	1133	1538	1367	33	36	0	1	92	34	10
117	32695	11815	3147	17414	170	1095	2	75	549	148	214
118	36919	13473	4176	16444	2547	653	5	26	443	611	228

119 rows × 13 columns

In [50]:

```
print(a.isnull())
```

	Impressions	From Home	From Hashtags	From Explore	From Other	Saves	Comments	Shares	Likes	Profile Visits	Follows
0	False	False	False	False	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False	False	False	False
..
114	False	False	False	False	False	False	False	False	False	False	False
115	False	False	False	False	False	False	False	False	False	False	False
116	False	False	False	False	False	False	False	False	False	False	False
117	False	False	False	False	False	False	False	False	False	False	False
118	False	False	False	False	False	False	False	False	False	False	False
	Comments	Shares	Likes	Profile Visits	Follows	Caption	Hashtags				
0	False	False	False	False	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False	False	False	False
..
114	False	False	False	False	False	False	False	False	False	False	False
115	False	False	False	False	False	False	False	False	False	False	False
116	False	False	False	False	False	False	False	False	False	False	False
117	False	False	False	False	False	False	False	False	False	False	False
118	False	False	False	False	False	False	False	False	False	False	False

[119 rows × 13 columns]

In [51]:

```
print(a.fillna(value=0))
```

	Impressions	From Home	From Hashtags	From Explore	From Other	Saves	\
0	3920	2586	1028	619	56	98	
1	5394	2727	1838	1174	78	194	
2	4021	2085	1188	0	533	41	
3	4528	2700	621	932	73	172	
4	2518	1704	255	279	37	96	
..	
114	13700	5185	3041	5352	77	573	
115	5731	1923	1368	2266	65	135	
116	4139	1133	1538	1367	33	36	
117	32695	11815	3147	17414	170	1095	
118	36919	13473	4176	16444	2547	653	

	Comments	Shares	Likes	Profile Visits	Follows	\
0	9	5	162	35	2	
1	7	14	224	48	10	
2	11	1	131	62	12	
3	10	7	213	23	8	
4	5	4	123	8	0	
..	
114	2	38	373	73	80	
115	4	1	148	20	18	
116	0	1	92	34	10	
117	2	75	549	148	214	
118	5	26	443	611	228	

Caption \

0	Here are some of the most important data visua...
1	Here are some of the best data science project...
2	Learn how to train a machine learning model an...
3	Here's how you can write a Python program to d...
4	Plotting annotations while visualizing your da...
..	...
114	Here are some of the best data science certifi...
115	Clustering is a machine learning technique use...
116	Clustering music genres is a task of grouping ...
117	Here are some of the best data science certifi...
118	175 Python Projects with Source Code solved an...

Hashtags

0	#finance #money #business #investing #investme...
1	#healthcare #health #covid #data #datascience ...
2	#data #datascience #dataanalysis #dataanalytic...
3	#python #pythonprogramming #pythonprojects #py...
4	#datavizualization #datascience #data #dataana...
..	...
114	#datascience #datasciencejobs #datascientrai...
115	#machinelearning #machinelearningalgorithms #d...
116	#machinelearning #machinelearningalgorithms #d...
117	#datascience #datasciencejobs #datascientrai...
118	#python #pythonprogramming #pythonprojects #py...

[119 rows x 13 columns]

In [52]:

```
a.dropna()
```

Out[52]:

	Impressions	From Home	From Hashtags	From Explore	From Other	Saves	Comments	Shares	Likes	Profile Visits	Follows
0	3920	2586	1028	619	56	98	9	5	162	35	2

	Impressions	From Home	From Hashtags	From Explore	From Other	Saves	Comments	Shares	Likes	Profile Visits	Follows
1	5394	2727	1838	1174	78	194	7	14	224	48	10
2	4021	2085	1188	0	533	41	11	1	131	62	12
3	4528	2700	621	932	73	172	10	7	213	23	8
4	2518	1704	255	279	37	96	5	4	123	8	0
...
114	13700	5185	3041	5352	77	573	2	38	373	73	80
115	5731	1923	1368	2266	65	135	4	1	148	20	18
116	4139	1133	1538	1367	33	36	0	1	92	34	10
117	32695	11815	3147	17414	170	1095	2	75	549	148	214
118	36919	13473	4176	16444	2547	653	5	26	443	611	228

Impressions	From Home	From Hashtags	From Explore	From Other	Saves	Comments	Shares	Likes	Profile Visits	Follows
-------------	-----------	---------------	--------------	------------	-------	----------	--------	-------	----------------	---------

119 rows × 13 columns

In [53]:

```
print("Head\n\n",a.head())
```

Head

	Impressions	From Home	From Hashtags	From Explore	From Other	Saves	\
0	3920	2586	1028	619	56	98	
1	5394	2727	1838	1174	78	194	
2	4021	2085	1188	0	533	41	
3	4528	2700	621	932	73	172	
4	2518	1704	255	279	37	96	
Comments	Shares	Likes	Profile Visits	Follows			\
0	9	5	162	35	2		
1	7	14	224	48	10		
2	11	1	131	62	12		
3	10	7	213	23	8		
4	5	4	123	8	0		

Caption \

- 0 Here are some of the most important data visua...
- 1 Here are some of the best data science project...
- 2 Learn how to train a machine learning model an...
- 3 Here's how you can write a Python program to d...
- 4 Plotting annotations while visualizing your da...

Hashtags

- 0 #finance #money #business #investing #investme...
- 1 #healthcare #health #covid #data #datascience ...
- 2 #data #datascience #dataanalysis #dataanalytic...
- 3 #python #pythonprogramming #pythonprojects #py...
- 4 #datavisualization #datascience #data #dataana...

In [54]:

```
print("Tail\n\n",a.tail())
```

Tail

	Impressions	From Home	From Hashtags	From Explore	From Other	Saves	\
114	13700	5185	3041	5352	77	573	
115	5731	1923	1368	2266	65	135	
116	4139	1133	1538	1367	33	36	
117	32695	11815	3147	17414	170	1095	
118	36919	13473	4176	16444	2547	653	
Comments	Shares	Likes	Profile Visits	Follows			\
114	2	38	373	73	80		
115	4	1	148	20	18		
116	0	1	92	34	10		
117	2	75	549	148	214		
118	5	26	443	611	228		

Caption \

- 114 Here are some of the best data science certifi...

```
115 Clustering is a machine learning technique use...
116 Clustering music genres is a task of grouping ...
117 Here are some of the best data science certifi...
118 175 Python Projects with Source Code solved an...
```

Hashtags

```
114 #datascience #datasciencejobs #datascientrai...
115 #machinelearning #machinelearningalgorithms #d...
116 #machinelearning #machinelearningalgorithms #d...
117 #datascience #datasciencejobs #datascientrai...
118 #python #pythonprogramming #pythonprojects #py...
```

In [55]:

```
a.describe()
```

Out[55]:

	Impressions	From Home	From Hashtags	From Explore	From Other	Saves	Comments	Profile Visits	Follows	Likes	Shares	Time
count	119.000000	119.000000	119.000000	119.000000	119.000000	119.000000	119.000000	119.000000	119.000000	119.000000	119.000000	119.000000
mean	5703.991597	2475.789916	1887.512605	1078.100840	171.092437	153.310924	6.663866	6.663866	6.663866	6.663866	6.663866	6.663866
std	4843.780105	1489.386348	1884.361443	2613.026132	289.431031	156.317731	3.544576	3.544576	3.544576	3.544576	3.544576	3.544576
min	1941.000000	1133.000000	116.000000	0.000000	9.000000	22.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	3467.000000	1945.000000	726.000000	157.500000	38.000000	65.000000	4.000000	4.000000	4.000000	4.000000	4.000000	4.000000
50%	4289.000000	2207.000000	1278.000000	326.000000	74.000000	109.000000	6.000000	6.000000	6.000000	6.000000	6.000000	6.000000
75%	6138.000000	2602.500000	2363.500000	689.500000	196.000000	169.000000	8.000000	8.000000	8.000000	8.000000	8.000000	8.000000
max	36919.000000	13473.000000	11817.000000	17414.000000	2547.000000	1095.000000	19.000000	19.000000	19.000000	19.000000	19.000000	19.000000

```
◀ ━━━━━━ ▶
```

In [56]:

```
np.shape(a)
```

Out[56]: (119, 13)

In [57]:

```
np.size(a)
```

Out[57]: 1547

In [58]:

```
np.ndim(a)
```

Out[58]: 2

In [69]:

```
a.iloc[0:3]
```

Out[69]:

	Impressions	From Home	From Hashtags	From Explore	From Other	Saves	Comments	Shares	Likes	Profile Visits	Follows	Time
0	3920	2586	1028	619	56	98	9	5	162	35	2	15

im

	Impressions	From Home	From Hashtags	From Explore	From Other	Saves	Comments	Shares	Likes	Profile Visits	Follows	Created
1	5394	2727	1838	1174	78	194	7	14	224	48	10	2023-01-01T00:00:00Z
2	4021	2085	1188	0	533	41	11	1	131	62	12	2023-01-01T00:00:00Z
3	4528	2700	621	932	73	172	10	7	213	23	8	2023-01-01T00:00:00Z

In [68]:

```
a.loc[0:3]
```

Out[68]:

	Impressions	From Home	From Hashtags	From Explore	From Other	Saves	Comments	Shares	Likes	Profile Visits	Follows	Created
0	3920	2586	1028	619	56	98	9	5	162	35	2	2023-01-01T00:00:00Z
1	5394	2727	1838	1174	78	194	7	14	224	48	10	2023-01-01T00:00:00Z
2	4021	2085	1188	0	533	41	11	1	131	62	12	2023-01-01T00:00:00Z
3	4528	2700	621	932	73	172	10	7	213	23	8	2023-01-01T00:00:00Z



In [102...]:

```
a.columns
```

```
Out[102]: Index(['Impressions', 'From Home', 'From Hashtags', 'From Explore',
   'From Other', 'Saves', 'Comments', 'Shares', 'Likes', 'Profile Visits',
   'Follows', 'Caption', 'Hashtags'],
  dtype='object')
```

```
In [103]: a.index
```

```
Out[103]: RangeIndex(start=0, stop=119, step=1)
```

```
In [89]: a1=a[['Comments', "Shares"]]
a1
a1.mean()
```

```
Out[89]: Comments    6.663866
Shares      9.361345
dtype: float64
```

```
In [90]: a1.median()
```

```
Out[90]: Comments    6.0
Shares      6.0
dtype: float64
```

```
In [91]: a1.mode()
```

```
Out[91]:    Comments  Shares
0            6       3
```

```
In [92]: print(a1.sum())
```

```
Comments    793
Shares     1114
dtype: int64
```

```
In [93]: print(a1.cumsum())
```

	Comments	Shares
0	9	5
1	16	19
2	27	20
3	37	27
4	42	31
..
114	782	1011
115	786	1012
116	786	1013
117	788	1088
118	793	1114

```
[119 rows x 2 columns]
```

```
In [94]: print(a1.count())
```

```
Comments      119  
Shares       119  
dtype: int64
```

```
In [95]: print(a1.min())
```

```
Comments      0  
Shares       0  
dtype: int64
```

```
In [96]: print(a1.max())
```

```
Comments     19  
Shares      75  
dtype: int64
```

```
In [98]: from numpy import cov  
from scipy.stats import pearsonr  
from scipy.stats import spearmanr
```

```
In [99]: a1=a["Comments"].values  
a2=a["Saves"].values  
print(np.cov(a1,a2))
```

```
[[ 1.25640222e+01 -1.49115511e+01]  
 [-1.49115511e+01  2.44352330e+04]]
```

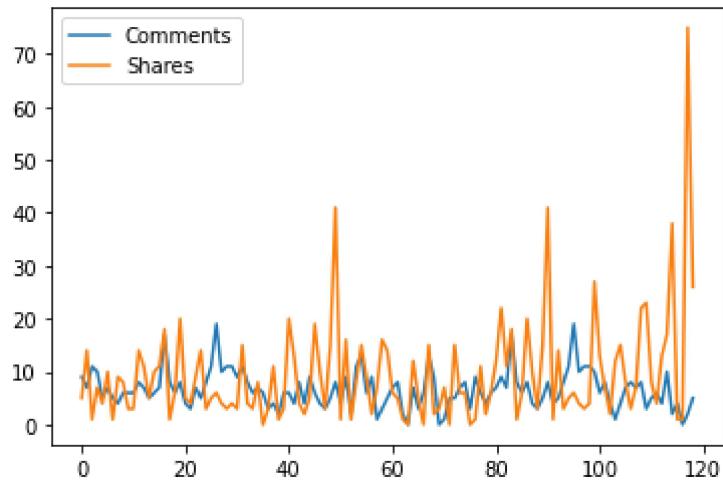
```
In [100...  
a1=a["Comments"].values  
a2=a["Saves"].values  
spearman =pd.Series(a1).corr(pd.Series(a2),method='spearman')  
pearson =pd.Series(a1).corr(pd.Series(a2),method='pearson')  
print(spearman)  
print(pearson)
```

```
0.18289066665208123  
-0.026912263707561033
```

Visualization

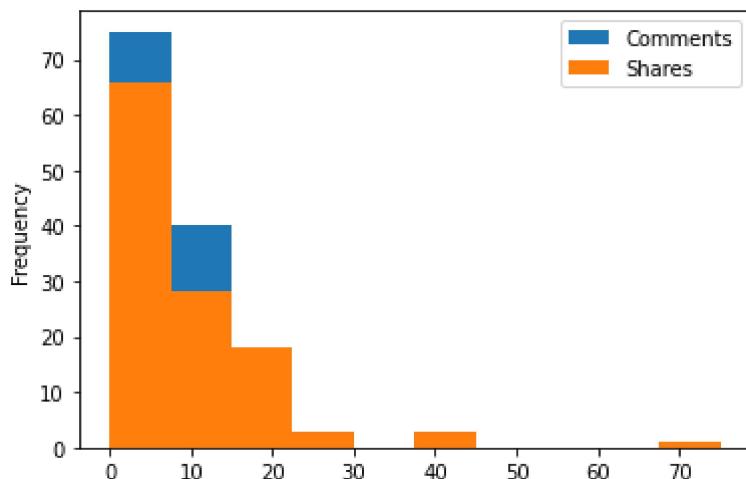
```
In [76]: a1.plot.line()
```

```
Out[76]: <AxesSubplot:>
```



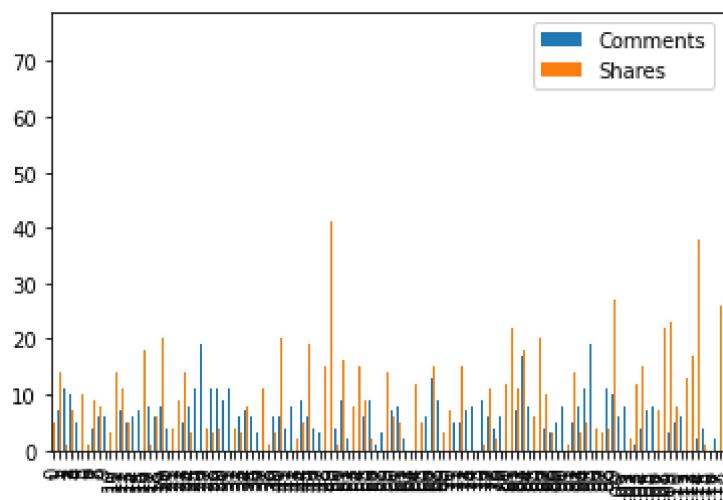
```
In [78]: a1.plot.hist()
```

```
Out[78]: <AxesSubplot:ylabel='Frequency'>
```



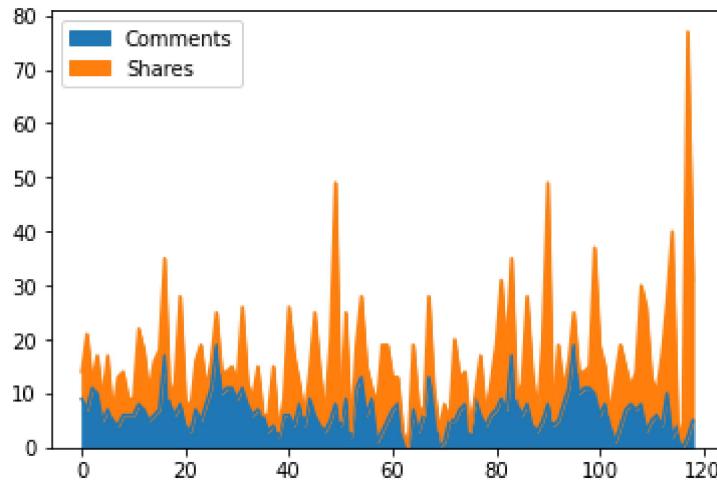
```
In [79]: a1.plot.bar()
```

```
Out[79]: <AxesSubplot:>
```



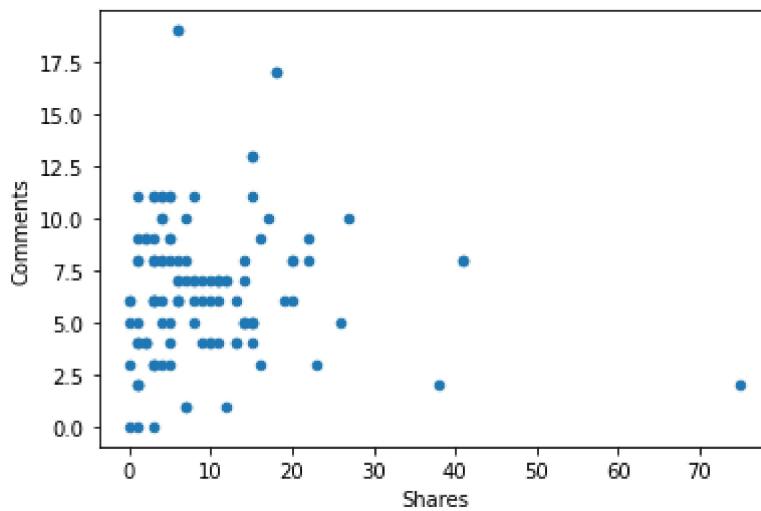
```
In [80]: a1.plot.area()
```

```
Out[80]: <AxesSubplot:>
```



```
In [81]: a1.plot.scatter(x="Shares",y="Comments")
```

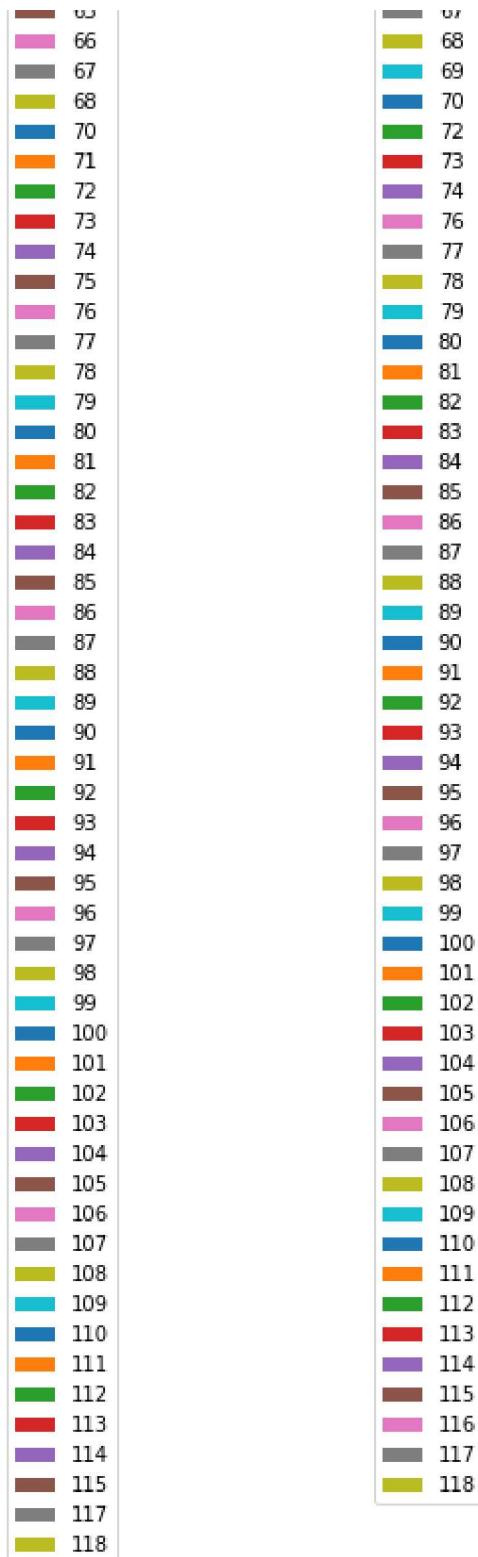
```
Out[81]: <AxesSubplot:xlabel='Shares', ylabel='Comments'>
```



```
In [82]: a1.plot.pie(subplots=True)
```

```
Out[82]: array([<AxesSubplot:ylabel='Comments'>, <AxesSubplot:ylabel='Shares'>],  
dtype=object)
```





```
In [83]: a1.boxplot()
```

```
Out[83]: <AxesSubplot:>
```

