

In []:

Kaustav Ghosh

Reg 180905188

Roll No 29 C

```
In [1]: import pandas as pd
import numpy as np
```

```
s=pd.Series([1,2,3,6,5,4])
print(s)
```

```
0    1
1    2
2    3
3    6
4    5
5    4
dtype: int64
```

```
In [2]: sum=s.sum()
```

```
In [3]: s.min()
```

```
Out[3]: 1
```

```
In [4]: s.max()
```

```
Out[4]: 6
```

```
In [5]: import pandas as pd
```

```
data=[['Dinesh',10],['Nithya',12],['Raji',13]]
df=pd.DataFrame(data,columns=['name','age'])
df
```

```
Out[5]:
```

	name	age
0	Dinesh	10
1	Nithya	12
2	Raji	13

```
In [6]: data = {'Name':['Kavitha', 'Sudha', 'Raju','Vignesh'],'Age':[28,34,29,42]}
df = pd.DataFrame(data, index=['rank1','rank2','rank3','rank4'])
df
```

```
Out[6]:
```

	Name	Age
rank1	Kavitha	28
rank2	Sudha	34
rank3	Raju	29
rank4	Vignesh	42

```
In [7]: df1=pd.DataFrame({'A':pd.Timestamp('20130102'),'B':np.array([3]*4,dtype='int32'),'C':pd.Categorical(['Male','Female','Male','Female'])})
df1
```

```
Out[7]:
```

	A	B	C
0	2013-01-02	3	Male
1	2013-01-02	3	Female
2	2013-01-02	3	Male
3	2013-01-02	3	Female

```
In [8]: df1.shape
```

```
Out[8]: (4, 3)
```

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

```
Out[9]: A    datetime64[ns]
B              int32
C              category
dtype: object
```

```
In [10]: df1.head()
```

```
Out[10]:
```

	A	B	C
0	2013-01-02	3	Male
1	2013-01-02	3	Female
2	2013-01-02	3	Male
3	2013-01-02	3	Female

```
In [11]: df1.tail()
```

```
Out[11]:
```

	A	B	C
0	2013-01-02	3	Male
1	2013-01-02	3	Female
2	2013-01-02	3	Male
3	2013-01-02	3	Female

```
In [12]: type(df1)
```

```
Out[12]: pandas.core.frame.DataFrame
```

```
In [13]: # df1.summary()
```

```
In [14]: df1.T
```

```
Out[14]:
```

	0	1	2	3
A	2013-01-02 00:00:00	2013-01-02 00:00:00	2013-01-02 00:00:00	2013-01-02 00:00:00
B	3	3	3	3
C	Male	Female	Male	Female

```
In [15]: df1
```

```
Out[15]:
```

	A	B	C
0	2013-01-02	3	Male
1	2013-01-02	3	Female
2	2013-01-02	3	Male
3	2013-01-02	3	Female

```
In [16]: df1.sort_index(axis=0,ascending=False)
```

```
Out[16]:
```

	A	B	C
3	2013-01-02	3	Female
2	2013-01-02	3	Male
1	2013-01-02	3	Female
0	2013-01-02	3	Male

```
In [17]: df1.sort_index(axis=0,ascending=True)
```

```
Out[17]:
```

	A	B	C
0	2013-01-02	3	Male
1	2013-01-02	3	Female
2	2013-01-02	3	Male
3	2013-01-02	3	Female

```
In [18]: df1.sort_index(axis=1,ascending=False)
```

```
Out[18]:
```

	C	B	A
0	Male	3	2013-01-02
1	Female	3	2013-01-02
2	Male	3	2013-01-02
3	Female	3	2013-01-02

```
In [19]: dates=pd.date_range('20200301',periods=100)
         dates
```

```
Out[19]: DatetimeIndex(['2020-03-01', '2020-03-02', '2020-03-03', '2020-03-04',
                        '2020-03-05', '2020-03-06', '2020-03-07', '2020-03-08',
                        '2020-03-09', '2020-03-10', '2020-03-11', '2020-03-12',
                        '2020-03-13', '2020-03-14', '2020-03-15', '2020-03-16',
                        '2020-03-17', '2020-03-18', '2020-03-19', '2020-03-20',
                        '2020-03-21', '2020-03-22', '2020-03-23', '2020-03-24',
                        '2020-03-25', '2020-03-26', '2020-03-27', '2020-03-28',
                        '2020-03-29', '2020-03-30', '2020-03-31', '2020-04-01',
                        '2020-04-02', '2020-04-03', '2020-04-04', '2020-04-05',
                        '2020-04-06', '2020-04-07', '2020-04-08', '2020-04-09',
                        '2020-04-10', '2020-04-11', '2020-04-12', '2020-04-13',
                        '2020-04-14', '2020-04-15', '2020-04-16', '2020-04-17',
                        '2020-04-18', '2020-04-19', '2020-04-20', '2020-04-21',
                        '2020-04-22', '2020-04-23', '2020-04-24', '2020-04-25',
                        '2020-04-26', '2020-04-27', '2020-04-28', '2020-04-29',
                        '2020-04-30', '2020-05-01', '2020-05-02', '2020-05-03',
                        '2020-05-04', '2020-05-05', '2020-05-06', '2020-05-07',
                        '2020-05-08', '2020-05-09', '2020-05-10', '2020-05-11',
                        '2020-05-12', '2020-05-13', '2020-05-14', '2020-05-15',
                        '2020-05-16', '2020-05-17', '2020-05-18', '2020-05-19',
                        '2020-05-20', '2020-05-21', '2020-05-22', '2020-05-23',
                        '2020-05-24', '2020-05-25', '2020-05-26', '2020-05-27',
                        '2020-05-28', '2020-05-29', '2020-05-30', '2020-05-31',
                        '2020-06-01', '2020-06-02', '2020-06-03', '2020-06-04',
                        '2020-06-05', '2020-06-06', '2020-06-07', '2020-06-08'],
                        dtype='datetime64[ns]', freq='D')
```

```
In [20]: df=pd.DataFrame(np.random.randn(100,5),index=dates,columns=list('ABCDE'))
         df
```

Out[20]:

	A	B	C	D	E
2020-03-01	-0.424361	1.324853	0.177166	-2.420857	0.395508
2020-03-02	-0.891346	0.418756	-0.576962	0.789255	0.603036
2020-03-03	-0.565622	0.237257	-1.238142	0.201801	-0.621756
2020-03-04	-0.622072	-0.970570	0.893486	1.205685	-0.061184
2020-03-05	-0.011453	1.207245	-0.249846	-0.263671	-0.827198
...
2020-06-04	0.151475	0.445946	-0.696244	0.029660	0.603972
2020-06-05	0.524293	-0.735005	-0.632562	-1.199961	0.639078
2020-06-06	-1.091008	-0.853599	-0.660759	-0.423268	-0.249291
2020-06-07	0.983981	1.440673	0.066006	-2.699567	0.219518
2020-06-08	0.736044	-0.987208	0.965826	0.027003	0.239819

100 rows × 5 columns

```
In [21]: df.index
```

```
Out[21]: DatetimeIndex(['2020-03-01', '2020-03-02', '2020-03-03', '2020-03-04',
                        '2020-03-05', '2020-03-06', '2020-03-07', '2020-03-08',
                        '2020-03-09', '2020-03-10', '2020-03-11', '2020-03-12',
                        '2020-03-13', '2020-03-14', '2020-03-15', '2020-03-16',
                        '2020-03-17', '2020-03-18', '2020-03-19', '2020-03-20',
                        '2020-03-21', '2020-03-22', '2020-03-23', '2020-03-24',
                        '2020-03-25', '2020-03-26', '2020-03-27', '2020-03-28',
                        '2020-03-29', '2020-03-30', '2020-03-31', '2020-04-01',
                        '2020-04-02', '2020-04-03', '2020-04-04', '2020-04-05',
                        '2020-04-06', '2020-04-07', '2020-04-08', '2020-04-09',
                        '2020-04-10', '2020-04-11', '2020-04-12', '2020-04-13',
                        '2020-04-14', '2020-04-15', '2020-04-16', '2020-04-17',
                        '2020-04-18', '2020-04-19', '2020-04-20', '2020-04-21',
                        '2020-04-22', '2020-04-23', '2020-04-24', '2020-04-25',
                        '2020-04-26', '2020-04-27', '2020-04-28', '2020-04-29',
                        '2020-04-30', '2020-05-01', '2020-05-02', '2020-05-03',
                        '2020-05-04', '2020-05-05', '2020-05-06', '2020-05-07',
                        '2020-05-08', '2020-05-09', '2020-05-10', '2020-05-11',
                        '2020-05-12', '2020-05-13', '2020-05-14', '2020-05-15',
                        '2020-05-16', '2020-05-17', '2020-05-18', '2020-05-19',
                        '2020-05-20', '2020-05-21', '2020-05-22', '2020-05-23',
                        '2020-05-24', '2020-05-25', '2020-05-26', '2020-05-27',
                        '2020-05-28', '2020-05-29', '2020-05-30', '2020-05-31',
                        '2020-06-01', '2020-06-02', '2020-06-03', '2020-06-04',
                        '2020-06-05', '2020-06-06', '2020-06-07', '2020-06-08'],
                        dtype='datetime64[ns]', freq='D')
```

```
In [22]: df.columns
```

```
Out[22]: Index(['A', 'B', 'C', 'D', 'E'], dtype='object')
```

```
In [23]: df.sort_index(axis=0,ascending=True)
```

```
Out[23]:
```

	A	B	C	D	E
2020-03-01	-0.424361	1.324853	0.177166	-2.420857	0.395508
2020-03-02	-0.891346	0.418756	-0.576962	0.789255	0.603036
2020-03-03	-0.565622	0.237257	-1.238142	0.201801	-0.621756
2020-03-04	-0.622072	-0.970570	0.893486	1.205685	-0.061184
2020-03-05	-0.011453	1.207245	-0.249846	-0.263671	-0.827198
...
2020-06-04	0.151475	0.445946	-0.696244	0.029660	0.603972
2020-06-05	0.524293	-0.735005	-0.632562	-1.199961	0.639078
2020-06-06	-1.091008	-0.853599	-0.660759	-0.423268	-0.249291
2020-06-07	0.983981	1.440673	0.066006	-2.699567	0.219518
2020-06-08	0.736044	-0.987208	0.965826	0.027003	0.239819

100 rows × 5 columns

```
In [24]: df[0:3]
```

Out[24]:

	A	B	C	D	E
2020-03-01	-0.424361	1.324853	0.177166	-2.420857	0.395508
2020-03-02	-0.891346	0.418756	-0.576962	0.789255	0.603036
2020-03-03	-0.565622	0.237257	-1.238142	0.201801	-0.621756

```
In [25]: df.iloc[0]
```

Out[25]:

A	-0.424361
B	1.324853
C	0.177166
D	-2.420857
E	0.395508

Name: 2020-03-01 00:00:00, dtype: float64

```
In [26]: df.iloc[0:3]
```

Out[26]:

	A	B	C	D	E
2020-03-01	-0.424361	1.324853	0.177166	-2.420857	0.395508
2020-03-02	-0.891346	0.418756	-0.576962	0.789255	0.603036
2020-03-03	-0.565622	0.237257	-1.238142	0.201801	-0.621756

```
In [27]: # df[0,:2]
```

```
In [28]: df.iloc[0:10,:2]
```

Out[28]:

	A	B
2020-03-01	-0.424361	1.324853
2020-03-02	-0.891346	0.418756
2020-03-03	-0.565622	0.237257
2020-03-04	-0.622072	-0.970570
2020-03-05	-0.011453	1.207245
2020-03-06	-0.501430	1.723194
2020-03-07	-0.900611	0.785145
2020-03-08	-2.039871	2.410184
2020-03-09	-1.141107	-0.461455
2020-03-10	0.804887	-0.828880

```
In [29]: df['A']
```

```
Out[29]: 2020-03-01    -0.424361
          2020-03-02    -0.891346
          2020-03-03    -0.565622
          2020-03-04    -0.622072
          2020-03-05    -0.011453
          ...
          2020-06-04     0.151475
          2020-06-05     0.524293
          2020-06-06    -1.091008
          2020-06-07     0.983981
          2020-06-08     0.736044
          Freq: D, Name: A, Length: 100, dtype: float64
```

```
In [30]: df[['A', 'B']]
```

```
Out[30]:
```

	A	B
2020-03-01	-0.424361	1.324853
2020-03-02	-0.891346	0.418756
2020-03-03	-0.565622	0.237257
2020-03-04	-0.622072	-0.970570
2020-03-05	-0.011453	1.207245
...
2020-06-04	0.151475	0.445946
2020-06-05	0.524293	-0.735005
2020-06-06	-1.091008	-0.853599
2020-06-07	0.983981	1.440673
2020-06-08	0.736044	-0.987208

100 rows × 2 columns

```
In [31]: df[['A', 'B']][0:5]
```

```
Out[31]:
```

	A	B
2020-03-01	-0.424361	1.324853
2020-03-02	-0.891346	0.418756
2020-03-03	-0.565622	0.237257
2020-03-04	-0.622072	-0.970570
2020-03-05	-0.011453	1.207245

In [32]: df[df.A>1.1]

Out[32]:

	A	B	C	D	E
2020-03-11	1.226958	2.768342	1.602426	2.038239	-1.785696
2020-03-31	1.300608	-1.884347	0.384452	-0.687108	-0.536894
2020-04-25	1.513270	-0.689307	0.410927	0.025507	-0.149151
2020-04-30	2.421214	1.584555	0.213267	-1.152837	0.752907
2020-05-01	2.347788	0.904406	-1.302815	0.174765	1.304519
2020-05-07	1.333469	0.748865	0.634736	-3.448722	-0.234078
2020-05-15	1.469526	-1.887684	1.610219	0.459457	-0.055961
2020-05-16	1.373086	-0.017398	-0.789335	0.037940	-1.237515
2020-05-21	2.721519	0.290362	0.298594	-0.537848	-0.517663
2020-05-22	1.908520	0.631277	1.485751	-0.823750	0.161971
2020-06-01	1.180628	1.401385	-0.254090	0.358312	-0.143079
2020-06-02	1.285464	2.452896	-0.360308	1.574432	1.484435

In [32]:

In [33]: df.drop('A',axis=1,inplace=False)

Out[33]:

	B	C	D	E
2020-03-01	1.324853	0.177166	-2.420857	0.395508
2020-03-02	0.418756	-0.576962	0.789255	0.603036
2020-03-03	0.237257	-1.238142	0.201801	-0.621756
2020-03-04	-0.970570	0.893486	1.205685	-0.061184
2020-03-05	1.207245	-0.249846	-0.263671	-0.827198
...
2020-06-04	0.445946	-0.696244	0.029660	0.603972
2020-06-05	-0.735005	-0.632562	-1.199961	0.639078
2020-06-06	-0.853599	-0.660759	-0.423268	-0.249291
2020-06-07	1.440673	0.066006	-2.699567	0.219518
2020-06-08	-0.987208	0.965826	0.027003	0.239819

100 rows × 4 columns


```
In [34]: df
```

```
Out[34]:
```

	A	B	C	D	E
2020-03-01	-0.424361	1.324853	0.177166	-2.420857	0.395508
2020-03-02	-0.891346	0.418756	-0.576962	0.789255	0.603036
2020-03-03	-0.565622	0.237257	-1.238142	0.201801	-0.621756
2020-03-04	-0.622072	-0.970570	0.893486	1.205685	-0.061184
2020-03-05	-0.011453	1.207245	-0.249846	-0.263671	-0.827198
...
2020-06-04	0.151475	0.445946	-0.696244	0.029660	0.603972
2020-06-05	0.524293	-0.735005	-0.632562	-1.199961	0.639078
2020-06-06	-1.091008	-0.853599	-0.660759	-0.423268	-0.249291
2020-06-07	0.983981	1.440673	0.066006	-2.699567	0.219518
2020-06-08	0.736044	-0.987208	0.965826	0.027003	0.239819

100 rows × 5 columns

```
In [35]: series1 = pd.Series([1, 2, 3])
display('series1:', series1)
series2 = pd.Series(['A', 'B', 'C'])
display('series2:', series2)

# concatenating
display('After concatenating:')
display(pd.concat([series1, series2]))
```

```
'series1:'
```

```
0    1
1    2
2    3
dtype: int64
```

```
'series2:'
```

```
0    A
1    B
2    C
dtype: object
```

```
'After concatenating:'
```

```
0    1
1    2
2    3
0    A
1    B
2    C
dtype: object
```

```
In [35]:
```

END

In [35]: