

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

```
In [2]: a=np.array([1,2,3,4])  
a
```

```
Out[2]: array([1, 2, 3, 4])
```

```
In [3]: li=np.array([[9,78,65],[5,4,9]])  
print(li)  
  
[[ 9 78 65]  
 [ 5  4  9]]
```

```
In [4]: li=np.array((1,2,3,4))  
li
```

```
Out[4]: array([1, 2, 3, 4])
```

```
In [5]: a=np.arange(100,20,-4)  
a
```

```
Out[5]: array([100, 96, 92, 88, 84, 80, 76, 72, 68, 64, 60, 56, 52,  
              48, 44, 40, 36, 32, 28, 24])
```

```
In [6]: x=np.full((5,5),5)  
x
```

```
Out[6]: array([[5, 5, 5, 5, 5],  
              [5, 5, 5, 5, 5],  
              [5, 5, 5, 5, 5],  
              [5, 5, 5, 5, 5],  
              [5, 5, 5, 5, 5]])
```

```
In [7]: r2=np.arange(100,20,-4).reshape(4,5)  
r2
```

```
Out[7]: array([[100, 96, 92, 88, 84],  
              [ 80, 76, 72, 68, 64],  
              [ 60, 56, 52, 48, 44],  
              [ 40, 36, 32, 28, 24]])
```

```
In [8]: r2=np.arange(100,20,-4).reshape(4,4)
r2
```

```
-----
ValueError                                Traceback (most recent call last)
<ipython-input-8-a801fbd557fe> in <module>
----> 1 r2=np.arange(100,20,-4).reshape(4,4)
      2 r2

ValueError: cannot reshape array of size 20 into shape (4,4)
```

```
In [9]: r2=np.arange(100,20,-1).reshape(4,5)
r2
```

```
-----
ValueError                                Traceback (most recent call last)
<ipython-input-9-e9b3847f209d> in <module>
----> 1 r2=np.arange(100,20,-1).reshape(4,5)
      2 r2

ValueError: cannot reshape array of size 80 into shape (4,5)
```

```
In [10]: r2=np.arange(100,20).reshape(4,4)
r2
```

```
-----
ValueError                                Traceback (most recent call last)
<ipython-input-10-6cb4390c2327> in <module>
----> 1 r2=np.arange(100,20).reshape(4,4)
      2 r2

ValueError: cannot reshape array of size 0 into shape (4,4)
```

```
In [12]: r2=np.arange(100,20,-4).reshape(4,5)
r2
```

```
Out[12]: array([[100,  96,  92,  88,  84],
                [ 80,  76,  72,  68,  64],
                [ 60,  56,  52,  48,  44],
                [ 40,  36,  32,  28,  24]])
```

Pandas

- It is one of the most important module of data science
- used to work with data in the form of frames/tables
- stands for Panel data
- Data Analysis, Manipulation and cleaning -Data Analysis means analysing the data
 - statistics of the data, volume of the data etc
 - Data Manipulation means transferring the data from one format to another

- Merging,concatnating etc
- Data Cleaning means removing the null ddata
 - NUll data means NaN(Not a Number)

- 2 datastructures in pandas 1).Series
 - Sequential data 2).DataFrame
 - Data in tabular form ie,rows ad columns
- *installation**
- pip istall pandas
- import pandas

In []:

In []:

In []:

In [13]: `import pandas as pd`In [14]: `pd.__version`

```

-----
AttributeError                                Traceback (most recent call last)
<ipython-input-14-29895f600985> in <module>
----> 1 pd.__version

C:\ProgramData\Anaconda3\lib\site-packages\pandas\__init__.py in __getattr__(name)
    242         return _SparseArray
    243
--> 244     raise AttributeError(f"module 'pandas' has no attribute '{name}'")
    245
    246

AttributeError: module 'pandas' has no attribute '__version'

```

In [15]: `pd.__version__`

Out[15]: '1.2.4'

we can create series using str,tuple,list and dict

```
In [16]: # with string
st='I am mounav'
pd.Series(st)
```

```
Out[16]: 0    I am mounav
dtype: object
```

```
In [17]: #input.split()
li=input("I am inevitable").split()
pd.Series(li)
```

I am inevitable but I am iron man

```
Out[17]: 0    but
1         I
2         am
3        iron
4         man
dtype: object
```

```
In [18]: #using tuple/list
t=(6,7,8,2,10)
pd.Series(t)
```

```
Out[18]: 0     6
1     7
2     8
3     2
4    10
dtype: int64
```

```
In [19]: #using range()
pd.Series(range(10,30,6))
```

```
Out[19]: 0    10
1    16
2    22
3    28
dtype: int64
```

```
In [21]: #using dictionary
dic={'34':'mounav', '12':'charan', 12:91, 'tuple':(9,8, 'hi'), "list":[9,6,4]}
s=pd.Series(dic)
s
```

```
Out[21]: 34          mounav
12          charan
12           91
tuple    (9, 8, hi)
list     [9, 6, 4]
dtype: object
```

```
In [22]: s.index      #getting index values
```

```
Out[22]: Index(['34', '12', 12, 'tuple', 'list'], dtype='object')
```

```
In [26]: #creating series using numpy array  
import numpy as np  
ar=np.arange(10,40,5)#array  
ass=pd.Series(ar)  
ass
```

```
Out[26]: 0    10  
        1    15  
        2    20  
        3    25  
        4    30  
        5    35  
        dtype: int32
```

```
In [27]: #modifying index values  
ass[0]
```

```
Out[27]: 10
```

```
In [28]: ass[4]
```

```
Out[28]: 30
```

```
In [29]: ass[2]
```

```
Out[29]: 20
```

```
In [30]: ass[1]
```

```
Out[30]: 15
```

```
In [31]: ass.index=[4,5,6,7,9,1]      #index changing  
ass
```

```
Out[31]: 4    10  
        5    15  
        6    20  
        7    25  
        9    30  
        1    35  
        dtype: int32
```

```
In [32]: ass
```

```
Out[32]: 4    10
          5    15
          6    20
          7    25
          9    30
          1    35
          dtype: int32
```

```
In [37]: li2=[9,3,4,5,6,'hi']
          inx=[num for num in range(5,11)]
          pd.Series(li2,index=inx)
```

```
Out[37]: 5      9
          6      3
          7      4
          8      5
          9      6
          10     hi
          dtype: object
```

```
In [38]: inx
```

```
Out[38]: [5, 6, 7, 8, 9, 10]
```

```
In [43]: #pass the user index to series created using dict
          #len(dic)
          ix=[num for num in range(10,14)]
          pd.Series(dic,index=ix)
```

```
Out[43]: 10     NaN
          11     NaN
          12     91
          13     NaN
          dtype: object
```

```
In [76]: s
```

```
Out[76]: 34      mounav
          12      charan
          12      91
          tuple  (9, 8, hi)
          list   [9, 6, 4]
          dtype: object
```

In [77]: dic

Out[77]: {'34': 'mounav',
 '12': 'charan',
 12: 91,
 'tuple': (9, 8, 'hi'),
 'list': [9, 6, 4]}

In [78]: df=pd.DataFrame(dic)
 print(df)

| | 34 | 12 | 12 | tuple | list |
|---|--------|--------|----|-------|------|
| 0 | mounav | charan | 91 | 9 | 9 |
| 1 | mounav | charan | 91 | 8 | 6 |
| 2 | mounav | charan | 91 | hi | 4 |

In [79]: dic

Out[79]: {'34': 'mounav',
 '12': 'charan',
 12: 91,
 'tuple': (9, 8, 'hi'),
 'list': [9, 6, 4]}

In [80]: df=pd.DataFrame(dic)
 print(df)

| | 34 | 12 | 12 | tuple | list |
|---|--------|--------|----|-------|------|
| 0 | mounav | charan | 91 | 9 | 9 |
| 1 | mounav | charan | 91 | 8 | 6 |
| 2 | mounav | charan | 91 | hi | 4 |

In [81]: data={'int':90,'float':90.67,'str':"Mounav"}
 pd.DataFrame(data,index=[1,2,3])

Out[81]:

| | int | float | str |
|---|-----|-------|--------|
| 1 | 90 | 90.67 | Mounav |
| 2 | 90 | 90.67 | Mounav |
| 3 | 90 | 90.67 | Mounav |

In [82]: pd.concat([df,df2]) *#concatnating two dictionary*

...

In [87]: d2={"mobile":'samsung',12:91,"tu":(3,4,'hello')}
 d2

Out[87]: {'mobile': 'samsung', 12: 91, 'tu': (3, 4, 'hello')}

In [84]: df

Out[84]:

| | 34 | 12 | 12 | tuple | list |
|---|--------|--------|----|-------|------|
| 0 | mounav | charan | 91 | 9 | 9 |
| 1 | mounav | charan | 91 | 8 | 6 |
| 2 | mounav | charan | 91 | hi | 4 |

In [89]: pd.concat([df,d2]) *#concatnating two dictionary*

...

In [3]: `import pandas as pd`
`df3=pd.DataFrame({'Int_rate':[2,1,2,3], 'Int_gdp':[50,45,45,67]},index=[2041,2002,`
`df3`

Out[3]:

| | Int_rate | Int_gdp |
|------|----------|---------|
| 2041 | 2 | 50 |
| 2002 | 1 | 45 |
| 2003 | 2 | 45 |
| 2004 | 3 | 67 |

In [4]: *#joining two dataframes*
`df4=pd.DataFrame({'low_tier_hpi':[80,90,70,60], 'unemployment':[1,3,5,6]},index=[`
`df4`

Out[4]:

| | low_tier_hpi | unemployment |
|------|--------------|--------------|
| 2001 | 80 | 1 |
| 2002 | 90 | 3 |
| 2003 | 70 | 5 |
| 2004 | 60 | 6 |

In [10]: df3.join(df4)

Out[10]:

| | Int_rate | Int_gdp | low_tier_hpi | unemployment |
|------|----------|---------|--------------|--------------|
| 2041 | 2 | 50 | NaN | NaN |
| 2002 | 1 | 45 | 90.0 | 3.0 |
| 2003 | 2 | 45 | 70.0 | 5.0 |
| 2004 | 3 | 67 | 60.0 | 6.0 |


```
In [11]: #merging the dataframes having similar column value
m=pd.DataFrame({'a':[4,5,6,7], 'b':[30,50,60,70], 'c':['a','b','c','d']},index=['first', 'second', 'third', 'fourth'])
```

Out[11]:

| | a | b | c |
|--------|---|----|---|
| first | 4 | 30 | a |
| second | 5 | 50 | b |
| third | 6 | 60 | c |
| fourth | 7 | 70 | d |

```
In [12]: n=pd.DataFrame({'a':[4,9,6,7], 'b':[30,80,60,70], 'c':['a','b','c','d']},index=['first', 'second', 'third', 'fourth'])
```

Out[12]:

| | a | b | c |
|--------|---|----|---|
| first | 4 | 30 | a |
| second | 9 | 80 | b |
| third | 6 | 60 | c |
| fourth | 7 | 70 | d |

```
In [40]: m.merge(n)
```

Out[40]:

| | a | b | c |
|---|---|----|---|
| 0 | 4 | 30 | a |
| 1 | 6 | 60 | c |
| 2 | 7 | 70 | d |

```
In [42]: new=m.merge(n)
new.columns=['f','s','th']#renaming the col values
new.index=[5,6,7]#updating row values
new
```

Out[42]:

| | f | s | th |
|---|---|----|----|
| 5 | 4 | 30 | a |
| 6 | 6 | 60 | c |
| 7 | 7 | 70 | d |

In [43]: `m.merge(n,on='c')` *#merging is based on particular col*

Out[43]:

| | a_x | b_x | c | a_y | b_y |
|---|-----|-----|---|-----|-----|
| 0 | 4 | 30 | a | 4 | 30 |
| 1 | 5 | 50 | b | 9 | 80 |
| 2 | 6 | 60 | c | 6 | 60 |
| 3 | 7 | 70 | d | 7 | 70 |

In [44]: `n.merge(m,on='a')`

Out[44]:

| | a | b_x | c_x | b_y | c_y |
|---|---|-----|-----|-----|-----|
| 0 | 4 | 30 | a | 30 | a |
| 1 | 6 | 60 | c | 60 | c |
| 2 | 7 | 70 | d | 70 | d |

In [48]: *#data range*
`dates=pd.date_range('2022-01-01','2002-01-10')`
`pd.Series(dates)`

Out[48]: Series([], dtype: datetime64[ns])

In [49]: `dt=pd.date_range('2022-01-10',periods=5)`
`pd.Series(dt)`

Out[49]:

| | |
|---|------------|
| 0 | 2022-01-10 |
| 1 | 2022-01-11 |
| 2 | 2022-01-12 |
| 3 | 2022-01-13 |
| 4 | 2022-01-14 |

dtype: datetime64[ns]

In [50]: `temp=[30,32,35,40,31]`
`pd.Series(temp,index=dt)`

Out[50]:

| | |
|------------|----|
| 2022-01-10 | 30 |
| 2022-01-11 | 32 |
| 2022-01-12 | 35 |
| 2022-01-13 | 40 |
| 2022-01-14 | 31 |

Freq: D, dtype: int64

```
In [51]: new
```

```
Out[51]:
```

| | f | s | th |
|---|---|----|----|
| 5 | 4 | 30 | a |
| 6 | 6 | 60 | c |
| 7 | 7 | 70 | d |

```
In [52]: new.index[1]#getting index values
```

```
Out[52]: 6
```

```
In [53]: new
```

```
Out[53]:
```

| | f | s | th |
|---|---|----|----|
| 5 | 4 | 30 | a |
| 6 | 6 | 60 | c |
| 7 | 7 | 70 | d |

Working with .csv file

- reading .csv file located at same location
- `pd.read_csv('filename')`

```
In [54]: dir(pd)
```

...

```
In [58]: marks=pd.read_csv('marks.csv')
marks
```

Out[58]:

| | Roll | name | Maths | Discrete | OS | CO | DS |
|----|------|------|-------|----------|----|----|----|
| 0 | A0 | X | 60 | 70 | 30 | 70 | 90 |
| 1 | 91 | X | 60 | 70 | 30 | 70 | 90 |
| 2 | 53 | X | 60 | 70 | 30 | 70 | 90 |
| 3 | 54 | X | 60 | 70 | 30 | 70 | 90 |
| 4 | 55 | X | 60 | 70 | 30 | 70 | 90 |
| 5 | 56 | X | 60 | 70 | 30 | 70 | 90 |
| 6 | 57 | X | 60 | 70 | 30 | 70 | 90 |
| 7 | 58 | X | 60 | 70 | 30 | 70 | 90 |
| 8 | 59 | X | 60 | 70 | 30 | 70 | 90 |
| 9 | 60 | X | 60 | 70 | 30 | 70 | 90 |
| 10 | 61 | X | 60 | 70 | 30 | 70 | 90 |
| 11 | 62 | X | 60 | 70 | 30 | 70 | 90 |
| 12 | 69 | X | 60 | 70 | 30 | 70 | 90 |
| 13 | 70 | X | 60 | 70 | 30 | 70 | 90 |

```
In [59]: marks=pd.read_csv('marks.csv')
marks
```

Out[59]:

| | Roll | name | Maths | Discrete | OS | CO | DS |
|----|------|------|-------|----------|----|----|----|
| 0 | A0 | X | 60 | 70 | 30 | 70 | 90 |
| 1 | 91 | X | 60 | 70 | 30 | 70 | 90 |
| 2 | 53 | X | 60 | 70 | 30 | 70 | 90 |
| 3 | 54 | X | 60 | 70 | 30 | 70 | 90 |
| 4 | 55 | X | 60 | 70 | 30 | 70 | 90 |
| 5 | 56 | X | 60 | 70 | 30 | 70 | 90 |
| 6 | 57 | X | 60 | 70 | 30 | 70 | 90 |
| 7 | 58 | X | 60 | 70 | 30 | 70 | 90 |
| 8 | 59 | X | 60 | 70 | 30 | 70 | 90 |
| 9 | 60 | X | 60 | 70 | 30 | 70 | 90 |
| 10 | 61 | X | 60 | 70 | 30 | 70 | 90 |
| 11 | 62 | X | 60 | 70 | 30 | 70 | 90 |
| 12 | 69 | X | 60 | 70 | 30 | 70 | 90 |
| 13 | 70 | X | 60 | 70 | 30 | 70 | 90 |

```
In [60]: marks=pd.read_csv('marks.csv')
marks
```

Out[60]:

| | Roll | name | Maths | Discrete | OS | CO | DS |
|----|------|------|-------|----------|----|----|----|
| 0 | A0 | X | 60 | 70 | 30 | 70 | 90 |
| 1 | 91 | X | 60 | 70 | 30 | 70 | 90 |
| 2 | 53 | X | 60 | 70 | 30 | 70 | 90 |
| 3 | 54 | X | 60 | 70 | 30 | 70 | 90 |
| 4 | 55 | X | 60 | 70 | 30 | 70 | 90 |
| 5 | 56 | X | 60 | 70 | 30 | 70 | 90 |
| 6 | 57 | X | 60 | 70 | 30 | 70 | 90 |
| 7 | 58 | X | 60 | 70 | 30 | 70 | 90 |
| 8 | 59 | X | 60 | 70 | 30 | 70 | 90 |
| 9 | 60 | X | 60 | 70 | 30 | 70 | 90 |
| 10 | 61 | X | 60 | 70 | 30 | 70 | 90 |
| 11 | 62 | X | 60 | 70 | 30 | 70 | 90 |
| 12 | 69 | X | 60 | 70 | 30 | 70 | 90 |
| 13 | 70 | X | 60 | 70 | 30 | 70 | 90 |

In [61]: marks

Out[61]:

| | Roll | name | Maths | Discrete | OS | CO | DS |
|----|------|------|-------|----------|----|----|----|
| 0 | A0 | X | 60 | 70 | 30 | 70 | 90 |
| 1 | 91 | X | 60 | 70 | 30 | 70 | 90 |
| 2 | 53 | X | 60 | 70 | 30 | 70 | 90 |
| 3 | 54 | X | 60 | 70 | 30 | 70 | 90 |
| 4 | 55 | X | 60 | 70 | 30 | 70 | 90 |
| 5 | 56 | X | 60 | 70 | 30 | 70 | 90 |
| 6 | 57 | X | 60 | 70 | 30 | 70 | 90 |
| 7 | 58 | X | 60 | 70 | 30 | 70 | 90 |
| 8 | 59 | X | 60 | 70 | 30 | 70 | 90 |
| 9 | 60 | X | 60 | 70 | 30 | 70 | 90 |
| 10 | 61 | X | 60 | 70 | 30 | 70 | 90 |
| 11 | 62 | X | 60 | 70 | 30 | 70 | 90 |
| 12 | 69 | X | 60 | 70 | 30 | 70 | 90 |
| 13 | 70 | X | 60 | 70 | 30 | 70 | 90 |

In [62]: *#Data Processig*
#select 5 random samples
marks.sample() *#random samples in our data*

Out[62]:

| | Roll | name | Maths | Discrete | OS | CO | DS |
|---|------|------|-------|----------|----|----|----|
| 5 | 56 | X | 60 | 70 | 30 | 70 | 90 |

In [63]: marks.sample(3) *#3 random samples*

Out[63]:

| | Roll | name | Maths | Discrete | OS | CO | DS |
|----|------|------|-------|----------|----|----|----|
| 11 | 62 | X | 60 | 70 | 30 | 70 | 90 |
| 12 | 69 | X | 60 | 70 | 30 | 70 | 90 |
| 0 | A0 | X | 60 | 70 | 30 | 70 | 90 |

```
In [64]: marks.head() #first five samples
```

Out[64]:

| | Roll | name | Maths | Discrete | OS | CO | DS |
|---|------|------|-------|----------|----|----|----|
| 0 | A0 | X | 60 | 70 | 30 | 70 | 90 |
| 1 | 91 | X | 60 | 70 | 30 | 70 | 90 |
| 2 | 53 | X | 60 | 70 | 30 | 70 | 90 |
| 3 | 54 | X | 60 | 70 | 30 | 70 | 90 |
| 4 | 55 | X | 60 | 70 | 30 | 70 | 90 |

```
In [65]: marks.head(3)
```

Out[65]:

| | Roll | name | Maths | Discrete | OS | CO | DS |
|---|------|------|-------|----------|----|----|----|
| 0 | A0 | X | 60 | 70 | 30 | 70 | 90 |
| 1 | 91 | X | 60 | 70 | 30 | 70 | 90 |
| 2 | 53 | X | 60 | 70 | 30 | 70 | 90 |

```
In [66]: marks.head(1)
```

Out[66]:

| | Roll | name | Maths | Discrete | OS | CO | DS |
|---|------|------|-------|----------|----|----|----|
| 0 | A0 | X | 60 | 70 | 30 | 70 | 90 |

```
In [67]: marks.tail() #last 5 samples
```

Out[67]:

| | Roll | name | Maths | Discrete | OS | CO | DS |
|----|------|------|-------|----------|----|----|----|
| 9 | 60 | X | 60 | 70 | 30 | 70 | 90 |
| 10 | 61 | X | 60 | 70 | 30 | 70 | 90 |
| 11 | 62 | X | 60 | 70 | 30 | 70 | 90 |
| 12 | 69 | X | 60 | 70 | 30 | 70 | 90 |
| 13 | 70 | X | 60 | 70 | 30 | 70 | 90 |

In [68]: `marks.tail(4)`

Out[68]:

| | Roll | name | Maths | Discrete | OS | CO | DS |
|----|------|------|-------|----------|----|----|----|
| 10 | 61 | X | 60 | 70 | 30 | 70 | 90 |
| 11 | 62 | X | 60 | 70 | 30 | 70 | 90 |
| 12 | 69 | X | 60 | 70 | 30 | 70 | 90 |
| 13 | 70 | X | 60 | 70 | 30 | 70 | 90 |

In [69]: `marks.tail(1)`

Out[69]:

| | Roll | name | Maths | Discrete | OS | CO | DS |
|----|------|------|-------|----------|----|----|----|
| 13 | 70 | X | 60 | 70 | 30 | 70 | 90 |

In [70]: `marks['Roll']`

Out[70]:

```
0    A0
1    91
2    53
3    54
4    55
5    56
6    57
7    58
8    59
9    60
10   61
11   62
12   69
13   70
Name: Roll, dtype: object
```

In [71]: `marks['Roll']`

Out[71]:

```
0    A0
1    91
2    53
3    54
4    55
5    56
6    57
7    58
8    59
9    60
10   61
11   62
12   69
13   70
Name: Roll, dtype: object
```



```
In [77]: new=pd.read_csv('marks.csv',usecols=['name','OS'])  
new
```

Out[77]:

| | name | OS |
|----|------|----|
| 0 | X | 30 |
| 1 | X | 30 |
| 2 | X | 30 |
| 3 | X | 30 |
| 4 | X | 30 |
| 5 | X | 30 |
| 6 | X | 30 |
| 7 | X | 30 |
| 8 | X | 30 |
| 9 | X | 30 |
| 10 | X | 30 |
| 11 | X | 30 |
| 12 | X | 30 |
| 13 | X | 30 |

```
In [74]: marks
```

Out[74]:

| | Roll | name | Maths | Discrete | OS | CO | DS |
|----|------|------|-------|----------|----|----|----|
| 0 | A0 | X | 60 | 70 | 30 | 70 | 90 |
| 1 | 91 | X | 60 | 70 | 30 | 70 | 90 |
| 2 | 53 | X | 60 | 70 | 30 | 70 | 90 |
| 3 | 54 | X | 60 | 70 | 30 | 70 | 90 |
| 4 | 55 | X | 60 | 70 | 30 | 70 | 90 |
| 5 | 56 | X | 60 | 70 | 30 | 70 | 90 |
| 6 | 57 | X | 60 | 70 | 30 | 70 | 90 |
| 7 | 58 | X | 60 | 70 | 30 | 70 | 90 |
| 8 | 59 | X | 60 | 70 | 30 | 70 | 90 |
| 9 | 60 | X | 60 | 70 | 30 | 70 | 90 |
| 10 | 61 | X | 60 | 70 | 30 | 70 | 90 |
| 11 | 62 | X | 60 | 70 | 30 | 70 | 90 |
| 12 | 69 | X | 60 | 70 | 30 | 70 | 90 |
| 13 | 70 | X | 60 | 70 | 30 | 70 | 90 |

```
In [76]: marks[['Roll', 'name', 'CO']]
```

Out[76]:

| | Roll | name | CO |
|----|------|------|----|
| 0 | A0 | X | 70 |
| 1 | 91 | X | 70 |
| 2 | 53 | X | 70 |
| 3 | 54 | X | 70 |
| 4 | 55 | X | 70 |
| 5 | 56 | X | 70 |
| 6 | 57 | X | 70 |
| 7 | 58 | X | 70 |
| 8 | 59 | X | 70 |
| 9 | 60 | X | 70 |
| 10 | 61 | X | 70 |
| 11 | 62 | X | 70 |
| 12 | 69 | X | 70 |
| 13 | 70 | X | 70 |

```
In [78]: marks.dtypes
```

#data type of each column

Out[78]:

| | |
|----------|--------|
| Roll | object |
| name | object |
| Maths | int64 |
| Discrete | int64 |
| OS | int64 |
| CO | int64 |
| DS | int64 |

dtype: object

```
In [79]: marks.index
```

Out[79]: RangeIndex(start=0, stop=14, step=1)

In [80]: marks.describe

Out[80]: <bound method NDFrame.describe of

| | Roll | name | Maths | Discrete | OS | CO | DS |
|----|------|------|-------|----------|----|----|-----|
| 0 | A0 | X | 60 | 70 | 30 | 70 | 90 |
| 1 | 91 | X | 60 | 70 | 30 | 70 | 90 |
| 2 | 53 | X | 60 | 70 | 30 | 70 | 90 |
| 3 | 54 | X | 60 | 70 | 30 | 70 | 90 |
| 4 | 55 | X | 60 | 70 | 30 | 70 | 90 |
| 5 | 56 | X | 60 | 70 | 30 | 70 | 90 |
| 6 | 57 | X | 60 | 70 | 30 | 70 | 90 |
| 7 | 58 | X | 60 | 70 | 30 | 70 | 90 |
| 8 | 59 | X | 60 | 70 | 30 | 70 | 90 |
| 9 | 60 | X | 60 | 70 | 30 | 70 | 90 |
| 10 | 61 | X | 60 | 70 | 30 | 70 | 90 |
| 11 | 62 | X | 60 | 70 | 30 | 70 | 90 |
| 12 | 69 | X | 60 | 70 | 30 | 70 | 90 |
| 13 | 70 | X | 60 | 70 | 30 | 70 | 90> |

In [81]: marks.info

Out[81]: <bound method DataFrame.info of

| | Roll | name | Maths | Discrete | OS | CO | DS |
|----|------|------|-------|----------|----|----|-----|
| 0 | A0 | X | 60 | 70 | 30 | 70 | 90 |
| 1 | 91 | X | 60 | 70 | 30 | 70 | 90 |
| 2 | 53 | X | 60 | 70 | 30 | 70 | 90 |
| 3 | 54 | X | 60 | 70 | 30 | 70 | 90 |
| 4 | 55 | X | 60 | 70 | 30 | 70 | 90 |
| 5 | 56 | X | 60 | 70 | 30 | 70 | 90 |
| 6 | 57 | X | 60 | 70 | 30 | 70 | 90 |
| 7 | 58 | X | 60 | 70 | 30 | 70 | 90 |
| 8 | 59 | X | 60 | 70 | 30 | 70 | 90 |
| 9 | 60 | X | 60 | 70 | 30 | 70 | 90 |
| 10 | 61 | X | 60 | 70 | 30 | 70 | 90 |
| 11 | 62 | X | 60 | 70 | 30 | 70 | 90 |
| 12 | 69 | X | 60 | 70 | 30 | 70 | 90 |
| 13 | 70 | X | 60 | 70 | 30 | 70 | 90> |

In [82]: marks.columns

Out[82]: Index(['Roll', 'name', 'Maths', 'Discrete', 'OS', 'CO', 'DS'], dtype='object')

In [83]: marks

Out[83]:

| | Roll | name | Maths | Discrete | OS | CO | DS |
|----|------|------|-------|----------|----|----|----|
| 0 | A0 | X | 60 | 70 | 30 | 70 | 90 |
| 1 | 91 | X | 60 | 70 | 30 | 70 | 90 |
| 2 | 53 | X | 60 | 70 | 30 | 70 | 90 |
| 3 | 54 | X | 60 | 70 | 30 | 70 | 90 |
| 4 | 55 | X | 60 | 70 | 30 | 70 | 90 |
| 5 | 56 | X | 60 | 70 | 30 | 70 | 90 |
| 6 | 57 | X | 60 | 70 | 30 | 70 | 90 |
| 7 | 58 | X | 60 | 70 | 30 | 70 | 90 |
| 8 | 59 | X | 60 | 70 | 30 | 70 | 90 |
| 9 | 60 | X | 60 | 70 | 30 | 70 | 90 |
| 10 | 61 | X | 60 | 70 | 30 | 70 | 90 |
| 11 | 62 | X | 60 | 70 | 30 | 70 | 90 |
| 12 | 69 | X | 60 | 70 | 30 | 70 | 90 |
| 13 | 70 | X | 60 | 70 | 30 | 70 | 90 |

In [89]: marks.describe()

Out[89]:

| | Maths | Discrete | OS | CO | DS |
|-------|-------|----------|------|------|------|
| count | 14.0 | 14.0 | 14.0 | 14.0 | 14.0 |
| mean | 60.0 | 70.0 | 30.0 | 70.0 | 90.0 |
| std | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| min | 60.0 | 70.0 | 30.0 | 70.0 | 90.0 |
| 25% | 60.0 | 70.0 | 30.0 | 70.0 | 90.0 |
| 50% | 60.0 | 70.0 | 30.0 | 70.0 | 90.0 |
| 75% | 60.0 | 70.0 | 30.0 | 70.0 | 90.0 |
| max | 60.0 | 70.0 | 30.0 | 70.0 | 90.0 |

In [88]: marks.Maths.value_counts()

Out[88]: 60 14
Name: Maths, dtype: int64

```
In [90]: marks.Maths.value_counts(1)
```

```
Out[90]: 60    1.0  
         Name: Maths, dtype: float64
```

```
In [91]: marks.Maths.value_counts(3)
```

```
Out[91]: 60    1.0  
         Name: Maths, dtype: float64
```

```
In [92]: marks.sort_values('Roll',ascending=False)
```

```
Out[92]:
```

| | Roll | name | Maths | Discrete | OS | CO | DS |
|----|------|------|-------|----------|----|----|----|
| 0 | A0 | X | 60 | 70 | 30 | 70 | 90 |
| 1 | 91 | X | 60 | 70 | 30 | 70 | 90 |
| 13 | 70 | X | 60 | 70 | 30 | 70 | 90 |
| 12 | 69 | X | 60 | 70 | 30 | 70 | 90 |
| 11 | 62 | X | 60 | 70 | 30 | 70 | 90 |
| 10 | 61 | X | 60 | 70 | 30 | 70 | 90 |
| 9 | 60 | X | 60 | 70 | 30 | 70 | 90 |
| 8 | 59 | X | 60 | 70 | 30 | 70 | 90 |
| 7 | 58 | X | 60 | 70 | 30 | 70 | 90 |
| 6 | 57 | X | 60 | 70 | 30 | 70 | 90 |
| 5 | 56 | X | 60 | 70 | 30 | 70 | 90 |
| 4 | 55 | X | 60 | 70 | 30 | 70 | 90 |
| 3 | 54 | X | 60 | 70 | 30 | 70 | 90 |
| 2 | 53 | X | 60 | 70 | 30 | 70 | 90 |

```
In [94]: marks['total']=100
```

In [95]: marks

Out[95]:

| | Roll | name | Maths | Discrete | OS | CO | DS | total |
|----|------|------|-------|----------|----|----|----|-------|
| 0 | A0 | X | 60 | 70 | 30 | 70 | 90 | 100 |
| 1 | 91 | X | 60 | 70 | 30 | 70 | 90 | 100 |
| 2 | 53 | X | 60 | 70 | 30 | 70 | 90 | 100 |
| 3 | 54 | X | 60 | 70 | 30 | 70 | 90 | 100 |
| 4 | 55 | X | 60 | 70 | 30 | 70 | 90 | 100 |
| 5 | 56 | X | 60 | 70 | 30 | 70 | 90 | 100 |
| 6 | 57 | X | 60 | 70 | 30 | 70 | 90 | 100 |
| 7 | 58 | X | 60 | 70 | 30 | 70 | 90 | 100 |
| 8 | 59 | X | 60 | 70 | 30 | 70 | 90 | 100 |
| 9 | 60 | X | 60 | 70 | 30 | 70 | 90 | 100 |
| 10 | 61 | X | 60 | 70 | 30 | 70 | 90 | 100 |
| 11 | 62 | X | 60 | 70 | 30 | 70 | 90 | 100 |
| 12 | 69 | X | 60 | 70 | 30 | 70 | 90 | 100 |
| 13 | 70 | X | 60 | 70 | 30 | 70 | 90 | 100 |

In [96]: marks.Maths.sum() *#summation of one col data*

Out[96]: 840

In [97]: marks.sum()

Out[97]: Roll A091535455565758596061626970
 name XXXXXXXXXXXXXXXXX
 Maths 840
 Discrete 980
 OS 420
 CO 980
 DS 1260
 total 1400
 dtype: object

In [98]: `marks.isnull()` *#checks for null values*

Out[98]:

| | Roll | name | Maths | Discrete | OS | CO | DS | total |
|----|-------|-------|-------|----------|-------|-------|-------|-------|
| 0 | False | False | False | False | False | False | False | False |
| 1 | False | False | False | False | False | False | False | False |
| 2 | False | False | False | False | False | False | False | False |
| 3 | False | False | False | False | False | False | False | False |
| 4 | False | False | False | False | False | False | False | False |
| 5 | False | False | False | False | False | False | False | False |
| 6 | False | False | False | False | False | False | False | False |
| 7 | False | False | False | False | False | False | False | False |
| 8 | False | False | False | False | False | False | False | False |
| 9 | False | False | False | False | False | False | False | False |
| 10 | False | False | False | False | False | False | False | False |
| 11 | False | False | False | False | False | False | False | False |
| 12 | False | False | False | False | False | False | False | False |
| 13 | False | False | False | False | False | False | False | False |

In [102]: *#finding the sum/count of null values in our data*
`marks.isnull().sum()`

Out[102]:

| | |
|----------|---|
| Roll | 0 |
| name | 0 |
| Maths | 0 |
| Discrete | 0 |
| OS | 0 |
| CO | 0 |
| DS | 0 |
| total | 0 |

dtype: int64

In [103]: `marks.nunique()`

Out[103]:

| | |
|----------|----|
| Roll | 14 |
| name | 1 |
| Maths | 1 |
| Discrete | 1 |
| OS | 1 |
| CO | 1 |
| DS | 1 |
| total | 1 |

dtype: int64

```
In [105]: clean=marks.dropna()      #removing null characters
```

```
In [106]: clean
```

Out[106]:

| | Roll | name | Maths | Discrete | OS | CO | DS | total |
|----|------|------|-------|----------|----|----|----|-------|
| 0 | A0 | X | 60 | 70 | 30 | 70 | 90 | 100 |
| 1 | 91 | X | 60 | 70 | 30 | 70 | 90 | 100 |
| 2 | 53 | X | 60 | 70 | 30 | 70 | 90 | 100 |
| 3 | 54 | X | 60 | 70 | 30 | 70 | 90 | 100 |
| 4 | 55 | X | 60 | 70 | 30 | 70 | 90 | 100 |
| 5 | 56 | X | 60 | 70 | 30 | 70 | 90 | 100 |
| 6 | 57 | X | 60 | 70 | 30 | 70 | 90 | 100 |
| 7 | 58 | X | 60 | 70 | 30 | 70 | 90 | 100 |
| 8 | 59 | X | 60 | 70 | 30 | 70 | 90 | 100 |
| 9 | 60 | X | 60 | 70 | 30 | 70 | 90 | 100 |
| 10 | 61 | X | 60 | 70 | 30 | 70 | 90 | 100 |
| 11 | 62 | X | 60 | 70 | 30 | 70 | 90 | 100 |
| 12 | 69 | X | 60 | 70 | 30 | 70 | 90 | 100 |
| 13 | 70 | X | 60 | 70 | 30 | 70 | 90 | 100 |

```
In [107]: marks      #original
```

Out[107]:

| | Roll | name | Maths | Discrete | OS | CO | DS | total |
|----|------|------|-------|----------|----|----|----|-------|
| 0 | A0 | X | 60 | 70 | 30 | 70 | 90 | 100 |
| 1 | 91 | X | 60 | 70 | 30 | 70 | 90 | 100 |
| 2 | 53 | X | 60 | 70 | 30 | 70 | 90 | 100 |
| 3 | 54 | X | 60 | 70 | 30 | 70 | 90 | 100 |
| 4 | 55 | X | 60 | 70 | 30 | 70 | 90 | 100 |
| 5 | 56 | X | 60 | 70 | 30 | 70 | 90 | 100 |
| 6 | 57 | X | 60 | 70 | 30 | 70 | 90 | 100 |
| 7 | 58 | X | 60 | 70 | 30 | 70 | 90 | 100 |
| 8 | 59 | X | 60 | 70 | 30 | 70 | 90 | 100 |
| 9 | 60 | X | 60 | 70 | 30 | 70 | 90 | 100 |
| 10 | 61 | X | 60 | 70 | 30 | 70 | 90 | 100 |
| 11 | 62 | X | 60 | 70 | 30 | 70 | 90 | 100 |
| 12 | 69 | X | 60 | 70 | 30 | 70 | 90 | 100 |
| 13 | 70 | X | 60 | 70 | 30 | 70 | 90 | 100 |

