

```
In [2]: import pandas as pd
```

```
In [2]: pd.__version__
```

```
Out[2]: '1.4.4'
```

```
In [3]: list = [1,2,3,4,5,6]
print(list)
```

```
[1, 2, 3, 4, 5, 6]
```

```
In [4]: series = pd.Series(list)
print(series)
print(type(series))
```

```
0    1
1    2
2    3
3    4
4    5
5    6
dtype: int64
<class 'pandas.core.series.Series'>
```

```
In [5]: empty = pd.Series([])
empty
```

```
C:\Users\harsh\AppData\Local\Temp\ipykernel_2436\3069909210.py:1: FutureWarning: T
he default dtype for empty Series will be 'object' instead of 'float64' in a futur
e version. Specify a dtype explicitly to silence this warning.
```

```
empty = pd.Series([])
```

```
Out[5]: Series([], dtype: float64)
```

```
In [6]: a = pd.Series(['p','q','r','s','t'], index = [10,11,12,13,14])
a
```

```
Out[6]: 10    p
11    q
12    r
13    s
14    t
dtype: object
```

```
In [7]: b = pd.Series(['p','q','r','s','t'], index = [10,11,12,13,14], name = 'alphabets')
b
```

```
Out[7]: 10    p
11    q
12    r
13    s
14    t
Name: alphabets, dtype: object
```

```
In [8]: scalar_series = pd.Series(0.5)
scalar_series
```

```
Out[8]: 0    0.5
dtype: float64
```

```
In [9]: scalar_series = pd.Series(0.5, index = [1,2,3])
scalar_series
```

```
Out[9]: 1    0.5
         2    0.5
         3    0.5
        dtype: float64
```

```
In [11]: dict_series = pd.Series({'p':1, 'q':2, 'r':3,'s':4, 't':5})
dict_series
```

```
Out[11]: p    1
          q    2
          r    3
          s    4
          t    5
         dtype: int64
```

```
In [12]: dict_series[0]
```

```
Out[12]: 1
```

```
In [13]: dict_series[1:3]
```

```
Out[13]: q    2
          r    3
         dtype: int64
```

```
In [14]: max(dict_series)
```

```
Out[14]: 5
```

```
In [15]: dict_series = pd.Series({'p':[1,5,6], 'q':[2,6,7], 'r':[3,9,0],'s':[4,4,5], 't':[5,1,2]})
dict_series
```

```
Out[15]: p    [1, 5, 6]
          q    [2, 6, 7]
          r    [3, 9, 0]
          s    [4, 4, 5]
          t    [5, 1, 2]
         dtype: object
```

```
In [3]: df = pd.DataFrame()
print(df)
```

```
Empty DataFrame
Columns: []
Index: []
```

```
In [4]: lst = [1,2,3,4,5]
df = pd.DataFrame(lst)
df
```

```
Out[4]: 0
      1
      2
      3
      4
      5
```

```
In [5]: lst = [[1,2,3,4,5], [11,12,13,14,15]]
df = pd.DataFrame(lst)
```

df

Out[5]:	0	1	2	3	4
0	1	2	3	4	5
1	11	12	13	14	15

```
In [6]: a = [{"a":5, "b":7, "c":9, "d":2},
            {"a":4, "b":8, "c":19, "d":12}]
df = pd.DataFrame(a)
df
```

Out[6]:	a	b	c	d
0	5	7	9	2
1	4	8	19	12

```
In [7]: b = {'RollNo.':pd.Series([1,2,3,4,5]),
            'Maths':pd.Series([45,65,14,98,87]),
            'Physics':pd.Series([65,65,34,22,66])}
df = pd.DataFrame(b)
df
```

Out[7]:	RollNo.	Maths	Physics
0	1	45	65
1	2	65	65
2	3	14	34
3	4	98	22
4	5	87	66

```
In [12]: df = pd.read_csv(r'C:\Users\harsh\Desktop\Python - Data Analysis\Pandas\Salary_Data.csv')
df
```

Out[12]:

	YearsExperience	Salary
0	1.1	39343.0
1	1.3	46205.0
2	1.5	37731.0
3	2.0	43525.0
4	2.2	39891.0
5	2.9	56642.0
6	3.0	60150.0
7	3.2	54445.0
8	3.2	64445.0
9	3.7	57189.0
10	3.9	63218.0
11	4.0	55794.0
12	4.0	56957.0
13	4.1	57081.0
14	4.5	61111.0
15	4.9	67938.0
16	5.1	66029.0
17	5.3	83088.0
18	5.9	81363.0
19	6.0	93940.0
20	6.8	91738.0
21	7.1	98273.0
22	7.9	101302.0
23	8.2	113812.0
24	8.7	109431.0
25	9.0	105582.0
26	9.5	116969.0
27	9.6	112635.0
28	10.3	122391.0
29	10.5	121872.0

In [13]:

type(df)

Out[13]:

pandas.core.frame.DataFrame

In [14]:

df.columns

Out[14]:

Index(['YearsExperience', 'Salary'], dtype='object')

```
In [15]: df.shape
```

```
Out[15]: (30, 2)
```

```
In [16]: df.size
```

```
Out[16]: 60
```

```
In [17]: df.head()
```

```
Out[17]:
```

	YearsExperience	Salary
0	1.1	39343.0
1	1.3	46205.0
2	1.5	37731.0
3	2.0	43525.0
4	2.2	39891.0

```
In [18]: df.head(2)
```

```
Out[18]:
```

	YearsExperience	Salary
0	1.1	39343.0
1	1.3	46205.0

```
In [19]: df.tail()
```

```
Out[19]:
```

	YearsExperience	Salary
25	9.0	105582.0
26	9.5	116969.0
27	9.6	112635.0
28	10.3	122391.0
29	10.5	121872.0

```
In [20]: df.tail(8)
```

```
Out[20]:
```

	YearsExperience	Salary
22	7.9	101302.0
23	8.2	113812.0
24	8.7	109431.0
25	9.0	105582.0
26	9.5	116969.0
27	9.6	112635.0
28	10.3	122391.0
29	10.5	121872.0

In [21]: `df.describe()`

	YearsExperience	Salary
count	30.000000	30.000000
mean	5.313333	76003.000000
std	2.837888	27414.429785
min	1.100000	37731.000000
25%	3.200000	56720.750000
50%	4.700000	65237.000000
75%	7.700000	100544.750000
max	10.500000	122391.000000

In [22]: `df.info()`

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 30 entries, 0 to 29
Data columns (total 2 columns):
 #   Column           Non-Null Count  Dtype  
---  -- 
 0   YearsExperience  30 non-null    float64
 1   Salary           30 non-null    float64
dtypes: float64(2)
memory usage: 608.0 bytes
```

In [25]: `df2 = pd.read_csv(r'C:\Users\harsh\Desktop\Python - Data Analysis\Pandas\Restaurant.csv')`
`df2.head()`

	Rank	Restaurant	Content	Sales	YOY_Sales	Units	YOY_Units	Headquarters	Segment_Cat
0	1	McDonald's		NaN	40412	4.9%	13846	-0.5%	NaN
1	2	Starbucks		NaN	21380	8.6%	15049	3.0%	NaN
2	3	Chick-fil-A	While Popeyes got a lot of the chicken buzz in...	11320	13.0%	2470	5.0%	NaN	Quick Service
3	4	Taco Bell	NaN	11293	9.0%	6766	2.7%	NaN	Quick Service
4	5	Burger King	NaN	10204	2.7%	7346	0.2%	NaN	Quick Service

In [26]: `df2.shape`

Out[26]: (250, 9)

In [27]: `df2.info()`

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 250 entries, 0 to 249
Data columns (total 9 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   Rank              250 non-null    int64  
 1   Restaurant        250 non-null    object  
 2   Content            33 non-null    object  
 3   Sales              250 non-null    int64  
 4   YOY_Sales          250 non-null    object  
 5   Units              250 non-null    int64  
 6   YOY_Units           250 non-null    object  
 7   Headquarters       52 non-null    object  
 8   Segment_Category   250 non-null    object  
dtypes: int64(3), object(6)
memory usage: 17.7+ KB
```

In [28]: `df2.describe()`

	Rank	Sales	Units
count	250.000000	250.000000	250.000000
mean	125.500000	1242.740000	850.076000
std	72.312977	3365.22882	2296.151659
min	1.000000	126.000000	13.000000
25%	63.250000	181.000000	85.000000
50%	125.500000	330.000000	207.000000
75%	187.750000	724.750000	555.250000
max	250.000000	40412.000000	23801.000000

In [35]: `df = pd.read_csv(r'C:\Users\harsh\Desktop\Python - Data Analysis\Pandas\sample.csv')`

	Roll No.	Physics	Chemistry	Maths	Computer
0	1	56.0	57.0	58.0	59.0
1	2	23.0	24.0	25.0	26.0
2	3	89.0	25.0	26.0	27.0
3	4	45.0	26.0	27.0	28.0
4	5	23.0	27.0	28.0	29.0

In [36]: `df.isnull()`

Out[36]:

	Roll No.	Physics	Chemistry	Maths	Computer
0	False	False	False	False	False
1	False	False	False	False	False
2	False	False	False	False	False
3	False	False	False	False	False
4	False	False	False	False	False
5	False	False	True	False	False
6	False	False	False	False	False
7	False	False	False	False	False
8	False	True	False	False	False
9	False	False	False	False	False
10	False	True	False	False	False
11	False	False	True	False	False
12	False	False	False	False	False
13	False	False	True	True	False
14	False	True	False	False	False
15	False	False	False	False	False
16	False	False	False	False	False
17	False	False	False	False	False
18	False	False	False	False	False
19	False	False	False	False	False
20	False	False	False	False	False
21	False	False	False	False	False
22	False	False	False	False	False
23	False	False	False	False	False
24	False	False	True	True	False
25	False	False	False	False	False
26	False	False	False	False	False
27	False	False	False	False	True
28	False	False	False	False	False
29	False	False	False	False	False

In [37]:

`df.isnull().sum()`

Out[37]:

Roll No.	0
Physics	3
Chemistry	4
Maths	2
Computer	1
dtype:	int64

```
In [38]: df.isnull().sum().sum()
```

```
Out[38]: 10
```

```
In [39]: df.shape
```

```
Out[39]: (30, 5)
```

```
In [40]: df2 = df.dropna()
```

```
In [42]: df2.shape
```

```
Out[42]: (22, 5)
```

```
In [43]: df3 = df.dropna(axis = 1)
```

```
In [45]: df3.shape
```

```
Out[45]: (30, 1)
```

```
In [46]: df.dropna(how = 'any') #if any row value is null then remove that row
```

Out[46]:

	Roll No.	Physics	Chemistry	Maths	Computer
0	1	56.0	57.0	58.0	59.0
1	2	23.0	24.0	25.0	26.0
2	3	89.0	25.0	26.0	27.0
3	4	45.0	26.0	27.0	28.0
4	5	23.0	27.0	28.0	29.0
6	7	12.0	13.0	14.0	15.0
7	8	78.0	14.0	15.0	16.0
9	10	45.0	16.0	17.0	18.0
12	13	22.0	23.0	24.0	25.0
15	16	44.0	44.0	44.0	44.0
16	17	45.0	45.0	45.0	45.0
17	18	46.0	46.0	46.0	46.0
18	19	47.0	47.0	47.0	47.0
19	20	48.0	48.0	48.0	48.0
20	21	49.0	49.0	49.0	49.0
21	22	50.0	50.0	50.0	50.0
22	23	51.0	51.0	51.0	51.0
23	24	52.0	52.0	52.0	52.0
25	26	54.0	33.0	33.0	54.0
26	27	55.0	34.0	34.0	55.0
28	29	57.0	36.0	36.0	66.0
29	30	58.0	37.0	37.0	43.0

In [47]: `df.dropna(how = 'all') #if all row values is null then remove that row`

Out[47]:

	Roll No.	Physics	Chemistry	Maths	Computer
0	1	56.0	57.0	58.0	59.0
1	2	23.0	24.0	25.0	26.0
2	3	89.0	25.0	26.0	27.0
3	4	45.0	26.0	27.0	28.0
4	5	23.0	27.0	28.0	29.0
5	6	90.0	NaN	29.0	30.0
6	7	12.0	13.0	14.0	15.0
7	8	78.0	14.0	15.0	16.0
8	9	NaN	15.0	16.0	17.0
9	10	45.0	16.0	17.0	18.0
10	11	NaN	17.0	18.0	19.0
11	12	88.0	NaN	19.0	20.0
12	13	22.0	23.0	24.0	25.0
13	14	90.0	NaN	NaN	42.0
14	15	NaN	43.0	43.0	43.0
15	16	44.0	44.0	44.0	44.0
16	17	45.0	45.0	45.0	45.0
17	18	46.0	46.0	46.0	46.0
18	19	47.0	47.0	47.0	47.0
19	20	48.0	48.0	48.0	48.0
20	21	49.0	49.0	49.0	49.0
21	22	50.0	50.0	50.0	50.0
22	23	51.0	51.0	51.0	51.0
23	24	52.0	52.0	52.0	52.0
24	25	53.0	NaN	NaN	53.0
25	26	54.0	33.0	33.0	54.0
26	27	55.0	34.0	34.0	55.0
27	28	56.0	35.0	35.0	NaN
28	29	57.0	36.0	36.0	66.0
29	30	58.0	37.0	37.0	43.0

In [48]: `df.dropna(inplace = True)`In [49]: `df.shape`

Out[49]: (22, 5)

In [51]: `df.fillna(0)`

Out[51]:

	Roll No.	Physics	Chemistry	Maths	Computer
0	1	56.0	57.0	58.0	59.0
1	2	23.0	24.0	25.0	26.0
2	3	89.0	25.0	26.0	27.0
3	4	45.0	26.0	27.0	28.0
4	5	23.0	27.0	28.0	29.0
6	7	12.0	13.0	14.0	15.0
7	8	78.0	14.0	15.0	16.0
9	10	45.0	16.0	17.0	18.0
12	13	22.0	23.0	24.0	25.0
15	16	44.0	44.0	44.0	44.0
16	17	45.0	45.0	45.0	45.0
17	18	46.0	46.0	46.0	46.0
18	19	47.0	47.0	47.0	47.0
19	20	48.0	48.0	48.0	48.0
20	21	49.0	49.0	49.0	49.0
21	22	50.0	50.0	50.0	50.0
22	23	51.0	51.0	51.0	51.0
23	24	52.0	52.0	52.0	52.0
25	26	54.0	33.0	33.0	54.0
26	27	55.0	34.0	34.0	55.0
28	29	57.0	36.0	36.0	66.0
29	30	58.0	37.0	37.0	43.0

In [56]:

```
df = pd.read_csv(r'C:\Users\harsh\Desktop\Python - Data Analysis\Pandas\sample.csv')
df.head()
```

Out[56]:

	Roll No.	Physics	Chemistry	Maths	Computer
0	1	56.0	57.0	58.0	59.0
1	2	23.0	24.0	25.0	26.0
2	3	89.0	25.0	26.0	27.0
3	4	45.0	26.0	27.0	28.0
4	5	23.0	27.0	28.0	29.0

In [57]:

```
df.isnull().sum()
```

Out[57]:

Roll No.	0
Physics	3
Chemistry	4
Maths	2
Computer	1
dtype: int64	

In [58]: `df.fillna(0)`

	Roll No.	Physics	Chemistry	Maths	Computer
0	1	56.0	57.0	58.0	59.0
1	2	23.0	24.0	25.0	26.0
2	3	89.0	25.0	26.0	27.0
3	4	45.0	26.0	27.0	28.0
4	5	23.0	27.0	28.0	29.0
5	6	90.0	0.0	29.0	30.0
6	7	12.0	13.0	14.0	15.0
7	8	78.0	14.0	15.0	16.0
8	9	0.0	15.0	16.0	17.0
9	10	45.0	16.0	17.0	18.0
10	11	0.0	17.0	18.0	19.0
11	12	88.0	0.0	19.0	20.0
12	13	22.0	23.0	24.0	25.0
13	14	90.0	0.0	0.0	42.0
14	15	0.0	43.0	43.0	43.0
15	16	44.0	44.0	44.0	44.0
16	17	45.0	45.0	45.0	45.0
17	18	46.0	46.0	46.0	46.0
18	19	47.0	47.0	47.0	47.0
19	20	48.0	48.0	48.0	48.0
20	21	49.0	49.0	49.0	49.0
21	22	50.0	50.0	50.0	50.0
22	23	51.0	51.0	51.0	51.0
23	24	52.0	52.0	52.0	52.0
24	25	53.0	0.0	0.0	53.0
25	26	54.0	33.0	33.0	54.0
26	27	55.0	34.0	34.0	55.0
27	28	56.0	35.0	35.0	0.0
28	29	57.0	36.0	36.0	66.0
29	30	58.0	37.0	37.0	43.0

In [59]: `df.fillna(2)`

Out[59]:

	Roll No.	Physics	Chemistry	Maths	Computer
0	1	56.0	57.0	58.0	59.0
1	2	23.0	24.0	25.0	26.0
2	3	89.0	25.0	26.0	27.0
3	4	45.0	26.0	27.0	28.0
4	5	23.0	27.0	28.0	29.0
5	6	90.0	2.0	29.0	30.0
6	7	12.0	13.0	14.0	15.0
7	8	78.0	14.0	15.0	16.0
8	9	2.0	15.0	16.0	17.0
9	10	45.0	16.0	17.0	18.0
10	11	2.0	17.0	18.0	19.0
11	12	88.0	2.0	19.0	20.0
12	13	22.0	23.0	24.0	25.0
13	14	90.0	2.0	2.0	42.0
14	15	2.0	43.0	43.0	43.0
15	16	44.0	44.0	44.0	44.0
16	17	45.0	45.0	45.0	45.0
17	18	46.0	46.0	46.0	46.0
18	19	47.0	47.0	47.0	47.0
19	20	48.0	48.0	48.0	48.0
20	21	49.0	49.0	49.0	49.0
21	22	50.0	50.0	50.0	50.0
22	23	51.0	51.0	51.0	51.0
23	24	52.0	52.0	52.0	52.0
24	25	53.0	2.0	2.0	53.0
25	26	54.0	33.0	33.0	54.0
26	27	55.0	34.0	34.0	55.0
27	28	56.0	35.0	35.0	2.0
28	29	57.0	36.0	36.0	66.0
29	30	58.0	37.0	37.0	43.0

In [60]: `df.fillna({'Physics': 'none', 'Chemistry': 0, 'Maths': 30})`

Out[60]:

	Roll No.	Physics	Chemistry	Maths	Computer
0	1	56.0	57.0	58.0	59.0
1	2	23.0	24.0	25.0	26.0
2	3	89.0	25.0	26.0	27.0
3	4	45.0	26.0	27.0	28.0
4	5	23.0	27.0	28.0	29.0
5	6	90.0	0.0	29.0	30.0
6	7	12.0	13.0	14.0	15.0
7	8	78.0	14.0	15.0	16.0
8	9	none	15.0	16.0	17.0
9	10	45.0	16.0	17.0	18.0
10	11	none	17.0	18.0	19.0
11	12	88.0	0.0	19.0	20.0
12	13	22.0	23.0	24.0	25.0
13	14	90.0	0.0	30.0	42.0
14	15	none	43.0	43.0	43.0
15	16	44.0	44.0	44.0	44.0
16	17	45.0	45.0	45.0	45.0
17	18	46.0	46.0	46.0	46.0
18	19	47.0	47.0	47.0	47.0
19	20	48.0	48.0	48.0	48.0
20	21	49.0	49.0	49.0	49.0
21	22	50.0	50.0	50.0	50.0
22	23	51.0	51.0	51.0	51.0
23	24	52.0	52.0	52.0	52.0
24	25	53.0	0.0	30.0	53.0
25	26	54.0	33.0	33.0	54.0
26	27	55.0	34.0	34.0	55.0
27	28	56.0	35.0	35.0	NaN
28	29	57.0	36.0	36.0	66.0
29	30	58.0	37.0	37.0	43.0

In [61]: df.fillna(method = 'ffill') #axis = 0

Out[61]:

	Roll No.	Physics	Chemistry	Maths	Computer
0	1	56.0	57.0	58.0	59.0
1	2	23.0	24.0	25.0	26.0
2	3	89.0	25.0	26.0	27.0
3	4	45.0	26.0	27.0	28.0
4	5	23.0	27.0	28.0	29.0
5	6	90.0	27.0	29.0	30.0
6	7	12.0	13.0	14.0	15.0
7	8	78.0	14.0	15.0	16.0
8	9	78.0	15.0	16.0	17.0
9	10	45.0	16.0	17.0	18.0
10	11	45.0	17.0	18.0	19.0
11	12	88.0	17.0	19.0	20.0
12	13	22.0	23.0	24.0	25.0
13	14	90.0	23.0	24.0	42.0
14	15	90.0	43.0	43.0	43.0
15	16	44.0	44.0	44.0	44.0
16	17	45.0	45.0	45.0	45.0
17	18	46.0	46.0	46.0	46.0
18	19	47.0	47.0	47.0	47.0
19	20	48.0	48.0	48.0	48.0
20	21	49.0	49.0	49.0	49.0
21	22	50.0	50.0	50.0	50.0
22	23	51.0	51.0	51.0	51.0
23	24	52.0	52.0	52.0	52.0
24	25	53.0	52.0	52.0	53.0
25	26	54.0	33.0	33.0	54.0
26	27	55.0	34.0	34.0	55.0
27	28	56.0	35.0	35.0	55.0
28	29	57.0	36.0	36.0	66.0
29	30	58.0	37.0	37.0	43.0

In [63]:

`df.fillna(method = 'bfill')`

Out[63]:

	Roll No.	Physics	Chemistry	Maths	Computer
0	1	56.0	57.0	58.0	59.0
1	2	23.0	24.0	25.0	26.0
2	3	89.0	25.0	26.0	27.0
3	4	45.0	26.0	27.0	28.0
4	5	23.0	27.0	28.0	29.0
5	6	90.0	13.0	29.0	30.0
6	7	12.0	13.0	14.0	15.0
7	8	78.0	14.0	15.0	16.0
8	9	45.0	15.0	16.0	17.0
9	10	45.0	16.0	17.0	18.0
10	11	88.0	17.0	18.0	19.0
11	12	88.0	23.0	19.0	20.0
12	13	22.0	23.0	24.0	25.0
13	14	90.0	43.0	43.0	42.0
14	15	44.0	43.0	43.0	43.0
15	16	44.0	44.0	44.0	44.0
16	17	45.0	45.0	45.0	45.0
17	18	46.0	46.0	46.0	46.0
18	19	47.0	47.0	47.0	47.0
19	20	48.0	48.0	48.0	48.0
20	21	49.0	49.0	49.0	49.0
21	22	50.0	50.0	50.0	50.0
22	23	51.0	51.0	51.0	51.0
23	24	52.0	52.0	52.0	52.0
24	25	53.0	33.0	33.0	53.0
25	26	54.0	33.0	33.0	54.0
26	27	55.0	34.0	34.0	55.0
27	28	56.0	35.0	35.0	66.0
28	29	57.0	36.0	36.0	66.0
29	30	58.0	37.0	37.0	43.0

In [62]:

```
df['Physics'].fillna(value=df['Physics'].mean())
```

```
Out[62]: 0    56.000000
         1    23.000000
         2    89.000000
         3    45.000000
         4    23.000000
         5    90.000000
         6    12.000000
         7    78.000000
         8    52.814815
         9    45.000000
        10   52.814815
        11   88.000000
        12   22.000000
        13   90.000000
        14   52.814815
        15   44.000000
        16   45.000000
        17   46.000000
        18   47.000000
        19   48.000000
        20   49.000000
        21   50.000000
        22   51.000000
        23   52.000000
        24   53.000000
        25   54.000000
        26   55.000000
        27   56.000000
        28   57.000000
        29   58.000000
Name: Physics, dtype: float64
```

```
In [64]: df.fillna(method = 'bfill', inplace = True)
df
```

Out[64]:

	Roll No.	Physics	Chemistry	Maths	Computer
0	1	56.0	57.0	58.0	59.0
1	2	23.0	24.0	25.0	26.0
2	3	89.0	25.0	26.0	27.0
3	4	45.0	26.0	27.0	28.0
4	5	23.0	27.0	28.0	29.0
5	6	90.0	13.0	29.0	30.0
6	7	12.0	13.0	14.0	15.0
7	8	78.0	14.0	15.0	16.0
8	9	45.0	15.0	16.0	17.0
9	10	45.0	16.0	17.0	18.0
10	11	88.0	17.0	18.0	19.0
11	12	88.0	23.0	19.0	20.0
12	13	22.0	23.0	24.0	25.0
13	14	90.0	43.0	43.0	42.0
14	15	44.0	43.0	43.0	43.0
15	16	44.0	44.0	44.0	44.0
16	17	45.0	45.0	45.0	45.0
17	18	46.0	46.0	46.0	46.0
18	19	47.0	47.0	47.0	47.0
19	20	48.0	48.0	48.0	48.0
20	21	49.0	49.0	49.0	49.0
21	22	50.0	50.0	50.0	50.0
22	23	51.0	51.0	51.0	51.0
23	24	52.0	52.0	52.0	52.0
24	25	53.0	33.0	33.0	53.0
25	26	54.0	33.0	33.0	54.0
26	27	55.0	34.0	34.0	55.0
27	28	56.0	35.0	35.0	66.0
28	29	57.0	36.0	36.0	66.0
29	30	58.0	37.0	37.0	43.0

In [65]: `df.replace(to_replace=26, value = 30)`

Out[65]:

	Roll No.	Physics	Chemistry	Maths	Computer
0	1	56.0	57.0	58.0	59.0
1	2	23.0	24.0	25.0	30.0
2	3	89.0	25.0	30.0	27.0
3	4	45.0	30.0	27.0	28.0
4	5	23.0	27.0	28.0	29.0
5	6	90.0	13.0	29.0	30.0
6	7	12.0	13.0	14.0	15.0
7	8	78.0	14.0	15.0	16.0
8	9	45.0	15.0	16.0	17.0
9	10	45.0	16.0	17.0	18.0
10	11	88.0	17.0	18.0	19.0
11	12	88.0	23.0	19.0	20.0
12	13	22.0	23.0	24.0	25.0
13	14	90.0	43.0	43.0	42.0
14	15	44.0	43.0	43.0	43.0
15	16	44.0	44.0	44.0	44.0
16	17	45.0	45.0	45.0	45.0
17	18	46.0	46.0	46.0	46.0
18	19	47.0	47.0	47.0	47.0
19	20	48.0	48.0	48.0	48.0
20	21	49.0	49.0	49.0	49.0
21	22	50.0	50.0	50.0	50.0
22	23	51.0	51.0	51.0	51.0
23	24	52.0	52.0	52.0	52.0
24	25	53.0	33.0	33.0	53.0
25	30	54.0	33.0	33.0	54.0
26	27	55.0	34.0	34.0	55.0
27	28	56.0	35.0	35.0	66.0
28	29	57.0	36.0	36.0	66.0
29	30	58.0	37.0	37.0	43.0

In [66]:

`df.replace(34, 10000)`

Out[66]:

	Roll No.	Physics	Chemistry	Maths	Computer
0	1	56.0	57.0	58.0	59.0
1	2	23.0	24.0	25.0	26.0
2	3	89.0	25.0	26.0	27.0
3	4	45.0	26.0	27.0	28.0
4	5	23.0	27.0	28.0	29.0
5	6	90.0	13.0	29.0	30.0
6	7	12.0	13.0	14.0	15.0
7	8	78.0	14.0	15.0	16.0
8	9	45.0	15.0	16.0	17.0
9	10	45.0	16.0	17.0	18.0
10	11	88.0	17.0	18.0	19.0
11	12	88.0	23.0	19.0	20.0
12	13	22.0	23.0	24.0	25.0
13	14	90.0	43.0	43.0	42.0
14	15	44.0	43.0	43.0	43.0
15	16	44.0	44.0	44.0	44.0
16	17	45.0	45.0	45.0	45.0
17	18	46.0	46.0	46.0	46.0
18	19	47.0	47.0	47.0	47.0
19	20	48.0	48.0	48.0	48.0
20	21	49.0	49.0	49.0	49.0
21	22	50.0	50.0	50.0	50.0
22	23	51.0	51.0	51.0	51.0
23	24	52.0	52.0	52.0	52.0
24	25	53.0	33.0	33.0	53.0
25	26	54.0	33.0	33.0	54.0
26	27	55.0	10000.0	10000.0	55.0
27	28	56.0	35.0	35.0	66.0
28	29	57.0	36.0	36.0	66.0
29	30	58.0	37.0	37.0	43.0

In [67]:

```
df.replace(to_replace=[50,51,52,53,54,55,56,57,58,59], value = 'A')
```

Out[67]:

	Roll No.	Physics	Chemistry	Maths	Computer
0	1	A	A	A	A
1	2	23.0	24.0	25.0	26.0
2	3	89.0	25.0	26.0	27.0
3	4	45.0	26.0	27.0	28.0
4	5	23.0	27.0	28.0	29.0
5	6	90.0	13.0	29.0	30.0
6	7	12.0	13.0	14.0	15.0
7	8	78.0	14.0	15.0	16.0
8	9	45.0	15.0	16.0	17.0
9	10	45.0	16.0	17.0	18.0
10	11	88.0	17.0	18.0	19.0
11	12	88.0	23.0	19.0	20.0
12	13	22.0	23.0	24.0	25.0
13	14	90.0	43.0	43.0	42.0
14	15	44.0	43.0	43.0	43.0
15	16	44.0	44.0	44.0	44.0
16	17	45.0	45.0	45.0	45.0
17	18	46.0	46.0	46.0	46.0
18	19	47.0	47.0	47.0	47.0
19	20	48.0	48.0	48.0	48.0
20	21	49.0	49.0	49.0	49.0
21	22	A	A	A	A
22	23	A	A	A	A
23	24	A	A	A	A
24	25	A	33.0	33.0	A
25	26	A	33.0	33.0	A
26	27	A	34.0	34.0	A
27	28	A	35.0	35.0	66.0
28	29	A	36.0	36.0	66.0
29	30	A	37.0	37.0	43.0

In [68]: df.replace(to_replace=[50,51,52,53], value = ['A', 'B', 'C', 'D'])

Out[68]:

	Roll No.	Physics	Chemistry	Maths	Computer
0	1	56.0	57.0	58.0	59.0
1	2	23.0	24.0	25.0	26.0
2	3	89.0	25.0	26.0	27.0
3	4	45.0	26.0	27.0	28.0
4	5	23.0	27.0	28.0	29.0
5	6	90.0	13.0	29.0	30.0
6	7	12.0	13.0	14.0	15.0
7	8	78.0	14.0	15.0	16.0
8	9	45.0	15.0	16.0	17.0
9	10	45.0	16.0	17.0	18.0
10	11	88.0	17.0	18.0	19.0
11	12	88.0	23.0	19.0	20.0
12	13	22.0	23.0	24.0	25.0
13	14	90.0	43.0	43.0	42.0
14	15	44.0	43.0	43.0	43.0
15	16	44.0	44.0	44.0	44.0
16	17	45.0	45.0	45.0	45.0
17	18	46.0	46.0	46.0	46.0
18	19	47.0	47.0	47.0	47.0
19	20	48.0	48.0	48.0	48.0
20	21	49.0	49.0	49.0	49.0
21	22	A	A	A	A
22	23	B	B	B	B
23	24	C	C	C	C
24	25	D	33.0	33.0	D
25	26	54.0	33.0	33.0	54.0
26	27	55.0	34.0	34.0	55.0
27	28	56.0	35.0	35.0	66.0
28	29	57.0	36.0	36.0	66.0
29	30	58.0	37.0	37.0	43.0

In [70]: df['Physics'].replace(to_replace=[50,51,52,53], value = ['A', 'B', 'C', 'D'], inplace=True)

Out[70]:

	Roll No.	Physics	Chemistry	Maths	Computer
0	1	56.0	57.0	58.0	59.0
1	2	23.0	24.0	25.0	26.0
2	3	89.0	25.0	26.0	27.0
3	4	45.0	26.0	27.0	28.0
4	5	23.0	27.0	28.0	29.0
5	6	90.0	13.0	29.0	30.0
6	7	12.0	13.0	14.0	15.0
7	8	78.0	14.0	15.0	16.0
8	9	45.0	15.0	16.0	17.0
9	10	45.0	16.0	17.0	18.0
10	11	88.0	17.0	18.0	19.0
11	12	88.0	23.0	19.0	20.0
12	13	22.0	23.0	24.0	25.0
13	14	90.0	43.0	43.0	42.0
14	15	44.0	43.0	43.0	43.0
15	16	44.0	44.0	44.0	44.0
16	17	45.0	45.0	45.0	45.0
17	18	46.0	46.0	46.0	46.0
18	19	47.0	47.0	47.0	47.0
19	20	48.0	48.0	48.0	48.0
20	21	49.0	49.0	49.0	49.0
21	22	A	50.0	50.0	50.0
22	23	B	51.0	51.0	51.0
23	24	C	52.0	52.0	52.0
24	25	D	33.0	33.0	53.0
25	26	54.0	33.0	33.0	54.0
26	27	55.0	34.0	34.0	55.0
27	28	56.0	35.0	35.0	66.0
28	29	57.0	36.0	36.0	66.0
29	30	58.0	37.0	37.0	43.0

In [72]: df.replace('[A-Za-z]', 0, regex = True)

Out[72]:

	Roll No.	Physics	Chemistry	Maths	Computer
0	1	56.0	57.0	58.0	59.0
1	2	23.0	24.0	25.0	26.0
2	3	89.0	25.0	26.0	27.0
3	4	45.0	26.0	27.0	28.0
4	5	23.0	27.0	28.0	29.0
5	6	90.0	13.0	29.0	30.0
6	7	12.0	13.0	14.0	15.0
7	8	78.0	14.0	15.0	16.0
8	9	45.0	15.0	16.0	17.0
9	10	45.0	16.0	17.0	18.0
10	11	88.0	17.0	18.0	19.0
11	12	88.0	23.0	19.0	20.0
12	13	22.0	23.0	24.0	25.0
13	14	90.0	43.0	43.0	42.0
14	15	44.0	43.0	43.0	43.0
15	16	44.0	44.0	44.0	44.0
16	17	45.0	45.0	45.0	45.0
17	18	46.0	46.0	46.0	46.0
18	19	47.0	47.0	47.0	47.0
19	20	48.0	48.0	48.0	48.0
20	21	49.0	49.0	49.0	49.0
21	22	0.0	50.0	50.0	50.0
22	23	0.0	51.0	51.0	51.0
23	24	0.0	52.0	52.0	52.0
24	25	0.0	33.0	33.0	53.0
25	26	54.0	33.0	33.0	54.0
26	27	55.0	34.0	34.0	55.0
27	28	56.0	35.0	35.0	66.0
28	29	57.0	36.0	36.0	66.0
29	30	58.0	37.0	37.0	43.0

In [73]: `df.replace(to_replace = 15, method = 'ffill')`

Out[73]:

	Roll No.	Physics	Chemistry	Maths	Computer
0	1	56.0	57.0	58.0	59.0
1	2	23.0	24.0	25.0	26.0
2	3	89.0	25.0	26.0	27.0
3	4	45.0	26.0	27.0	28.0
4	5	23.0	27.0	28.0	29.0
5	6	90.0	13.0	29.0	30.0
6	7	12.0	13.0	14.0	30.0
7	8	78.0	14.0	14.0	16.0
8	9	45.0	14.0	16.0	17.0
9	10	45.0	16.0	17.0	18.0
10	11	88.0	17.0	18.0	19.0
11	12	88.0	23.0	19.0	20.0
12	13	22.0	23.0	24.0	25.0
13	14	90.0	43.0	43.0	42.0
14	14	44.0	43.0	43.0	43.0
15	16	44.0	44.0	44.0	44.0
16	17	45.0	45.0	45.0	45.0
17	18	46.0	46.0	46.0	46.0
18	19	47.0	47.0	47.0	47.0
19	20	48.0	48.0	48.0	48.0
20	21	49.0	49.0	49.0	49.0
21	22	A	50.0	50.0	50.0
22	23	B	51.0	51.0	51.0
23	24	C	52.0	52.0	52.0
24	25	D	33.0	33.0	53.0
25	26	54.0	33.0	33.0	54.0
26	27	55.0	34.0	34.0	55.0
27	28	56.0	35.0	35.0	66.0
28	29	57.0	36.0	36.0	66.0
29	30	58.0	37.0	37.0	43.0

In [74]:

`df.replace(to_replace = 15, method = 'bfill')`

Out[74]:

	Roll No.	Physics	Chemistry	Maths	Computer
0	1	56.0	57.0	58.0	59.0
1	2	23.0	24.0	25.0	26.0
2	3	89.0	25.0	26.0	27.0
3	4	45.0	26.0	27.0	28.0
4	5	23.0	27.0	28.0	29.0
5	6	90.0	13.0	29.0	30.0
6	7	12.0	13.0	14.0	16.0
7	8	78.0	14.0	16.0	16.0
8	9	45.0	16.0	16.0	17.0
9	10	45.0	16.0	17.0	18.0
10	11	88.0	17.0	18.0	19.0
11	12	88.0	23.0	19.0	20.0
12	13	22.0	23.0	24.0	25.0
13	14	90.0	43.0	43.0	42.0
14	16	44.0	43.0	43.0	43.0
15	16	44.0	44.0	44.0	44.0
16	17	45.0	45.0	45.0	45.0
17	18	46.0	46.0	46.0	46.0
18	19	47.0	47.0	47.0	47.0
19	20	48.0	48.0	48.0	48.0
20	21	49.0	49.0	49.0	49.0
21	22	A	50.0	50.0	50.0
22	23	B	51.0	51.0	51.0
23	24	C	52.0	52.0	52.0
24	25	D	33.0	33.0	53.0
25	26	54.0	33.0	33.0	54.0
26	27	55.0	34.0	34.0	55.0
27	28	56.0	35.0	35.0	66.0
28	29	57.0	36.0	36.0	66.0
29	30	58.0	37.0	37.0	43.0

In [81]: df = pd.read_csv(r'C:\Users\harsh\Desktop\Python - Data Analysis\Pandas\sample2.csv')
df.head()

Out[81]:

	Section	Branch	Physics	Chemistry	Maths	Computer	DOB
Roll No.							
1	A	CS	56.0	57.0	58.0	59.0	01-01-2001
2	A	ECE	23.0	24.0	25.0	26.0	02-01-2001
3	B	MECH	89.0	25.0	26.0	27.0	03-01-2001
4	C	MECH	45.0	26.0	27.0	28.0	04-01-2001
5	A	CS	23.0	27.0	28.0	29.0	05-01-2001

In [82]:

`df.loc[1]`

Out[82]:

```
Section          A
Branch          CS
Physics         56.0
Chemistry       57.0
Maths           58.0
Computer        59.0
DOB             01-01-2001
Name: 1, dtype: object
```

In [85]:

`df.loc[5]`

Out[85]:

```
Section          A
Branch          CS
Physics         23.0
Chemistry       27.0
Maths           28.0
Computer        29.0
DOB             05-01-2001
Name: 5, dtype: object
```

In [86]:

`df.loc[[5,6,7,8]]`

Out[86]:

	Section	Branch	Physics	Chemistry	Maths	Computer	DOB
Roll No.							

5	A	CS	23.0	27.0	28.0	29.0	05-01-2001
6	A	ECE	90.0	NaN	29.0	30.0	06-01-2001
7	B	CS	12.0	13.0	14.0	15.0	07-01-2001
8	C	NaN	78.0	14.0	15.0	16.0	08-01-2001

In [87]:

`df.loc[5, 'Physics']`

Out[87]:

23.0

In [88]:

`df.loc[5:15, 'Chemistry']`

```
Out[88]: Roll No.
5      27.0
6      NaN
7      13.0
8      14.0
9      15.0
10     16.0
11     17.0
12     NaN
13     23.0
14     NaN
15     43.0
Name: Chemistry, dtype: float64
```

```
In [89]: df.loc[df['Physics'] < 50]
```

Roll No.	Section	Branch	Physics	Chemistry	Maths	Computer	DOB
2	A	ECE	23.0	24.0	25.0	26.0	02-01-2001
4	C	MECH	45.0	26.0	27.0	28.0	04-01-2001
5	A	CS	23.0	27.0	28.0	29.0	05-01-2001
7	B	CS	12.0	13.0	14.0	15.0	07-01-2001
10	A	CS	45.0	16.0	17.0	18.0	10-01-2001
13	A	CS	22.0	23.0	24.0	25.0	13-01-2001
16	C	Nan	44.0	44.0	44.0	44.0	16-01-2001
17	A	MECH	45.0	45.0	45.0	45.0	17-01-2001
18	A	MECH	46.0	46.0	46.0	46.0	18-01-2001
19	B	ECE	47.0	47.0	47.0	47.0	19-01-2001
20	C	MECH	48.0	48.0	48.0	48.0	20-01-2001
21	A	MECH	49.0	49.0	49.0	49.0	21-01-2001

```
In [90]: df.loc[df['Physics'] > 80]
```

Roll No.	Section	Branch	Physics	Chemistry	Maths	Computer	DOB
3	B	MECH	89.0	25.0	26.0	27.0	03-01-2001
6	A	ECE	90.0	Nan	29.0	30.0	06-01-2001
12	C	CS	88.0	Nan	19.0	20.0	12-01-2001
14	A	CS	90.0	Nan	Nan	42.0	14-01-2001

```
In [92]: df.loc[df['Physics'] > 80, ['Maths']]
```

Out[92]:

Maths

Roll No.	
3	26.0
6	29.0
12	19.0
14	NaN

In [93]:

`df.iloc[0]`

Out[93]:

```
Section      A
Branch       CS
Physics     56.0
Chemistry   57.0
Maths       58.0
Computer    59.0
DOB         01-01-2001
Name: 1, dtype: object
```

In [95]:

`df.iloc[[0,1,2]]`

Out[95]:

	Section	Branch	Physics	Chemistry	Maths	Computer	DOB
--	---------	--------	---------	-----------	-------	----------	-----

Roll No.							
1	A	CS	56.0	57.0	58.0	59.0	01-01-2001
2	A	ECE	23.0	24.0	25.0	26.0	02-01-2001
3	B	MECH	89.0	25.0	26.0	27.0	03-01-2001

In [98]:

`df.iloc[:,0]`

```
Out[98]: Roll No.  
1     A  
2     A  
3     B  
4     C  
5     A  
6     A  
7     B  
8     C  
9     A  
10    A  
11    B  
12    C  
13    A  
14    A  
15    B  
16    C  
17    A  
18    A  
19    B  
20    C  
21    A  
22    A  
23    B  
24    C  
25    A  
26    A  
27    B  
28    C  
29    A  
30    A  
Name: Section, dtype: object
```

In [100...]

df.iloc[:,1]

```
Out[100]: Roll No.
1      CS
2      ECE
3      MECH
4      MECH
5      CS
6      ECE
7      CS
8      NaN
9      ECE
10     CS
11     ECE
12     CS
13     CS
14     CS
15     ECE
16     NaN
17     MECH
18     MECH
19     ECE
20     MECH
21     MECH
22     MECH
23     ECE
24     MECH
25     MECH
26     ECE
27     CS
28     CS
29     CS
30     CS
Name: Branch, dtype: object
```

In [101... df.iloc[0:5, 1]

```
Out[101]: Roll No.
1      CS
2      ECE
3      MECH
4      MECH
5      CS
Name: Branch, dtype: object
```

In [102... df.iloc[0:5, 1:4]

Out[102]: **Branch Physics Chemistry**

Roll No.			
1	CS	56.0	57.0
2	ECE	23.0	24.0
3	MECH	89.0	25.0
4	MECH	45.0	26.0
5	CS	23.0	27.0

In [103... df = pd.read_csv(r'C:\Users\harsh\Desktop\Python - Data Analysis\Pandas\sample2.csv')
df.head()

	Roll No.	Section	Branch	Physics	Chemistry	Maths	Computer	DOB
0	1	A	CS	56.0	57.0	58.0	59.0	01-01-2001
1	2	A	ECE	23.0	24.0	25.0	26.0	02-01-2001
2	3	B	MECH	89.0	25.0	26.0	27.0	03-01-2001
3	4	C	MECH	45.0	26.0	27.0	28.0	04-01-2001
4	5	A	CS	23.0	27.0	28.0	29.0	05-01-2001

In [104...]: `branch_group = df.groupby(by = 'Branch')`
`branch_group`

Out[104]: <pandas.core.groupby.generic.DataFrameGroupBy object at 0x0000020AA5081760>

In [105...]: `branch_group.groups`

Out[105]: `{'CS': [0, 4, 6, 9, 11, 12, 13, 26, 27, 28, 29], 'ECE': [1, 5, 8, 10, 14, 18, 22, 25], 'MECH': [2, 3, 16, 17, 19, 20, 21, 23, 24]}`

In [108...]: `a = df.groupby(by = ['Branch', 'Section'])`
`a.groups`

Out[108]: `{('CS', 'A'): [0, 4, 9, 12, 13, 28, 29], ('CS', 'B'): [6, 26], ('CS', 'C'): [11, 27], ('ECE', 'A'): [1, 5, 8, 25], ('ECE', 'B'): [10, 14, 18, 22], ('MECH', 'A'): [16, 17, 20, 21, 24], ('MECH', 'B'): [2], ('MECH', 'C'): [3, 19, 23], (nan, 'C'): [7, 15]}`

In [111...]: `for group, data_frame in a:`
 `print(a)`
 `print(data_frame)`

```
<pandas.core.groupby.generic.DataFrameGroupBy object at 0x0000020AA508AE50>
    Roll No. Section Branch Physics Chemistry Maths Computer DOB
0          1      A    CS     56.0      57.0    58.0    59.0 01-01-2001
4          5      A    CS     23.0      27.0    28.0    29.0 05-01-2001
9         10      A    CS     45.0      16.0    17.0    18.0 10-01-2001
12        13      A    CS     22.0      23.0    24.0    25.0 13-01-2001
13        14      A    CS     90.0      NaN     NaN    42.0 14-01-2001
28        29      A    CS     57.0      36.0    36.0    66.0 29-01-2001
29        30      A    CS     58.0      37.0    37.0    43.0 30-01-2001
<pandas.core.groupby.generic.DataFrameGroupBy object at 0x0000020AA508AE50>
    Roll No. Section Branch Physics Chemistry Maths Computer DOB
6          7      B    CS     12.0      13.0    14.0    15.0 07-01-2001
26        27      B    CS     55.0      34.0    34.0    55.0 27-01-2001
<pandas.core.groupby.generic.DataFrameGroupBy object at 0x0000020AA508AE50>
    Roll No. Section Branch Physics Chemistry Maths Computer DOB
11        12      C    CS     88.0      NaN     19.0    20.0 12-01-2001
27        28      C    CS     56.0      35.0    35.0     NaN 28-01-2001
<pandas.core.groupby.generic.DataFrameGroupBy object at 0x0000020AA508AE50>
    Roll No. Section Branch Physics Chemistry Maths Computer DOB
1          2      A    ECE     23.0      24.0    25.0    26.0 02-01-2001
5          6      A    ECE     90.0      NaN     29.0    30.0 06-01-2001
8          9      A    ECE     NaN      15.0    16.0    17.0 09-01-2001
25        26      A    ECE     54.0      33.0    33.0    54.0 26-01-2001
<pandas.core.groupby.generic.DataFrameGroupBy object at 0x0000020AA508AE50>
    Roll No. Section Branch Physics Chemistry Maths Computer DOB
10        11      B    ECE     NaN      17.0    18.0    19.0 11-01-2001
14        15      B    ECE     NaN      43.0    43.0    43.0 15-01-2001
18        19      B    ECE     47.0      47.0    47.0    47.0 19-01-2001
22        23      B    ECE     51.0      51.0    51.0    51.0 23-01-2001
<pandas.core.groupby.generic.DataFrameGroupBy object at 0x0000020AA508AE50>
    Roll No. Section Branch Physics Chemistry Maths Computer DOB
16        17      A    MECH     45.0      45.0    45.0    45.0 17-01-2001
17        18      A    MECH     46.0      46.0    46.0    46.0 18-01-2001
20        21      A    MECH     49.0      49.0    49.0    49.0 21-01-2001
21        22      A    MECH     50.0      50.0    50.0    50.0 22-01-2001
24        25      A    MECH     53.0      NaN     NaN    53.0 25-01-2001
<pandas.core.groupby.generic.DataFrameGroupBy object at 0x0000020AA508AE50>
    Roll No. Section Branch Physics Chemistry Maths Computer DOB
2          3      B    MECH     89.0      25.0    26.0    27.0 03-01-2001
<pandas.core.groupby.generic.DataFrameGroupBy object at 0x0000020AA508AE50>
    Roll No. Section Branch Physics Chemistry Maths Computer DOB
3          4      C    MECH     45.0      26.0    27.0    28.0 04-01-2001
19        20      C    MECH     48.0      48.0    48.0    48.0 20-01-2001
23        24      C    MECH     52.0      52.0    52.0    52.0 24-01-2001
```

In [112]:
`df1 = pd.DataFrame({'Roll No.':[1,2,3,4,5], 'Physics':[34,54,87,52,54]})
df1`

Out[112]: **Roll No. Physics**

0	1	34
1	2	54
2	3	87
3	4	52
4	5	54

In [113]:
`df2 = pd.DataFrame({'Roll No.':[1,2,3,4,5], 'Chemistry':[65,74,98,34,21]})
df2`

Out[113]: **Roll No. Chemistry**

0	1	65
1	2	74
2	3	98
3	4	34
4	5	21

In [114... pd.merge(df1, df2, on = 'Roll No.')

Out[114]: **Roll No. Physics Chemistry**

0	1	34	65
1	2	54	74
2	3	87	98
3	4	52	34
4	5	54	21

In [115... pd.merge(df2, df1, on = 'Roll No.')

Out[115]: **Roll No. Chemistry Physics**

0	1	65	34
1	2	74	54
2	3	98	87
3	4	34	52
4	5	21	54

In [116... pd.merge(df1, df2)

Out[116]: **Roll No. Physics Chemistry**

0	1	34	65
1	2	54	74
2	3	87	98
3	4	52	34
4	5	54	21

In [117... df3 = pd.DataFrame({'Roll No.':[1,2,3,6,7], 'Physics':[34,67,34,89,12]})
df4 = pd.DataFrame({'Roll No.':[1,2,3,8,9], 'Chemistry':[34,67,34,89,12]})

In [119... pd.merge(df3, df4)

Out[119]:

	Roll No.	Physics	Chemistry
0	1	34	34
1	2	67	67
2	3	34	34

In [120...]:

```
pd.merge(df3, df4, how = 'left')
```

Out[120]:

	Roll No.	Physics	Chemistry
0	1	34	34.0
1	2	67	67.0
2	3	34	34.0
3	6	89	NaN
4	7	12	NaN

In [121...]:

```
pd.merge(df3, df4, how = 'right')
```

Out[121]:

	Roll No.	Physics	Chemistry
0	1	34.0	34
1	2	67.0	67
2	3	34.0	34
3	8	NaN	89
4	9	NaN	12

In [122...]:

```
pd.merge(df3, df4, how = 'outer')
```

Out[122]:

	Roll No.	Physics	Chemistry
0	1	34.0	34.0
1	2	67.0	67.0
2	3	34.0	34.0
3	6	89.0	NaN
4	7	12.0	NaN
5	8	NaN	89.0
6	9	NaN	12.0

In [123...]:

```
df1 = pd.DataFrame({'Roll No.':[1,2,3,4,5], 'Maths':[45,78,45,90,66], 'Physics':[34,67,34,89,12])
df2 = pd.DataFrame({'Roll No.':[6,7,8,9,10], 'Maths':[78,73,45,90,69], 'Physics':[89,67,45,90,12])
print(df1)
print(df2)
```

```
Roll No. Maths Physics
0      1    45     33
1      2    78     67
2      3    45     12
3      4    90     90
4      5    66     44
Roll No. Maths Physics
0      6    78     23
1      7    73     67
2      8    45     88
3      9    90      0
4     10    69     98
```

In [127...]: `df1.append(df2)`

C:\Users\harsh\AppData\Local\Temp\ipykernel_16160\3062608662.py:1: FutureWarning:
The frame.append method is deprecated and will be removed from pandas in a future
version. Use pandas.concat instead.
df1.append(df2)

Out[127]:

	Roll No.	Maths	Physics
0	1	45	33
1	2	78	67
2	3	45	12
3	4	90	90
4	5	66	44
0	6	78	23
1	7	73	67
2	8	45	88
3	9	90	0
4	10	69	98

In [129...]: `df2.append(df1)`

C:\Users\harsh\AppData\Local\Temp\ipykernel_16160\2914556596.py:1: FutureWarning:
The frame.append method is deprecated and will be removed from pandas in a future
version. Use pandas.concat instead.
df2.append(df1)

Out[129]:

	Roll No.	Maths	Physics
0	6	78	23
1	7	73	67
2	8	45	88
3	9	90	0
4	10	69	98
0	1	45	33
1	2	78	67
2	3	45	12
3	4	90	90
4	5	66	44

In [130...]

```
df1.append(df2, ignore_index = True)
```

C:\Users\harsh\AppData\Local\Temp\ipykernel_16160\2363792934.py:1: FutureWarning:
The frame.append method is deprecated and will be removed from pandas in a future
version. Use pandas.concat instead.
df1.append(df2, ignore_index = True)

Out[130]:

	Roll No.	Maths	Physics
0	1	45	33
1	2	78	67
2	3	45	12
3	4	90	90
4	5	66	44
5	6	78	23
6	7	73	67
7	8	45	88
8	9	90	0
9	10	69	98

In [131...]

```
df1.append(df2, ignore_index = True, sort = True)
```

C:\Users\harsh\AppData\Local\Temp\ipykernel_16160\1627638081.py:1: FutureWarning:
The frame.append method is deprecated and will be removed from pandas in a future
version. Use pandas.concat instead.
df1.append(df2, ignore_index = True, sort = True)

Out[131]:

	Maths	Physics	Roll No.
0	45	33	1
1	78	67	2
2	45	12	3
3	90	90	4
4	66	44	5
5	78	23	6
6	73	67	7
7	45	88	8
8	90	0	9
9	69	98	10

In [132...]

```
df1 = pd.DataFrame({'Roll No.':[1,2,3,4,5], 'Chemistry':[45,78,45,90,66], 'Physics':[33,67,12,90,44]})
df2 = pd.DataFrame({'Roll No.':[6,7,8,9,10], 'Maths':[78,73,45,90,69], 'Physics':[23,67,88,0,98]})
print(df1)
print(df2)
```

	Roll No.	Chemistry	Physics
0	1	45	33
1	2	78	67
2	3	45	12
3	4	90	90
4	5	66	44

	Roll No.	Maths	Physics
0	6	78	23
1	7	73	67
2	8	45	88
3	9	90	0
4	10	69	98

In [134...]

```
df1.append(df2, ignore_index = True, sort = True)
```

C:\Users\harsh\AppData\Local\Temp\ipykernel_16160\1627638081.py:1: FutureWarning:
The frame.append method is deprecated and will be removed from pandas in a future
version. Use pandas.concat instead.

```
df1.append(df2, ignore_index = True, sort = True)
```

Out[134]:

	Chemistry	Maths	Physics	Roll No.
0	45.0	NaN	33	1
1	78.0	NaN	67	2
2	45.0	NaN	12	3
3	90.0	NaN	90	4
4	66.0	NaN	44	5
5	NaN	78.0	23	6
6	NaN	73.0	67	7
7	NaN	45.0	88	8
8	NaN	90.0	0	9
9	NaN	69.0	98	10

In [135...]

```
df1 = pd.DataFrame({'Roll No.':[1,2,3,4,5], 'Maths':[45,78,45,90,66], 'Physics':[33,67,12,90,44], 'Chemistry':[56,89,33,12,89]})  
df2 = pd.DataFrame({'Roll No.':[6,7,8,9,10], 'Maths':[78,73,45,90,69], 'Physics':[23,67,88,0,98], 'Chemistry':[None]*5})  
print(df1)  
print(df2)
```

	Roll No.	Maths	Physics	Chemistry
0	1	45	33	56
1	2	78	67	89
2	3	45	12	33
3	4	90	90	12
4	5	66	44	89

	Roll No.	Maths	Physics
0	6	78	23
1	7	73	67
2	8	45	88
3	9	90	0
4	10	69	98

In [136...]

```
df1.append(df2, ignore_index = True, sort = True)
```

C:\Users\harsh\AppData\Local\Temp\ipykernel_16160\1627638081.py:1: FutureWarning:
The frame.append method is deprecated and will be removed from pandas in a future
version. Use pandas.concat instead.

```
df1.append(df2, ignore_index = True, sort = True)
```

Out[136]:

	Chemistry	Maths	Physics	Roll No.
0	56.0	45	33	1
1	89.0	78	67	2
2	33.0	45	12	3
3	12.0	90	90	4
4	89.0	66	44	5
5	NaN	78	23	6
6	NaN	73	67	7
7	NaN	45	88	8
8	NaN	90	0	9
9	NaN	69	98	10

In [139...]

```
df = pd.read_csv(r'C:\Users\harsh\Desktop\Python - Data Analysis\Pandas\sample2.csv')
df.head()
```

Out[139]:

	Roll No.	Section	Branch	Physics	Chemistry	Maths	Computer	DOB
0	1	A	CS	56.0	57.0	58.0	59.0	01-01-2001
1	2	A	ECE	23.0	24.0	25.0	26.0	02-01-2001
2	3	B	MECH	89.0	25.0	26.0	27.0	03-01-2001
3	4	C	MECH	45.0	26.0	27.0	28.0	04-01-2001
4	5	A	CS	23.0	27.0	28.0	29.0	05-01-2001

In [140...]

```
pd.pivot_table(df, index='Branch')
```

Out[140]:

	Chemistry	Computer	Maths	Physics	Roll No.
Branch					
CS	30.888889	37.200000	30.200	51.090909	16.000000
ECE	32.857143	35.875000	32.750	53.000000	13.875000
MECH	42.625000	44.222222	42.875	53.000000	17.111111

In [142...]

```
pd.pivot_table(df, index = 'Branch', aggfunc='sum')
```

Out[142]:

	Chemistry	Computer	Maths	Physics	Roll No.
Branch					
CS	278.0	372.0	302.0	562.0	176
ECE	230.0	287.0	262.0	265.0	111
MECH	341.0	398.0	343.0	477.0	154

In [143...]

```
pd.pivot_table(df, index = 'Branch', aggfunc='count')
```

Out[143]:

	Chemistry	Computer	DOB	Maths	Physics	Roll No.	Section
Branch							
CS	9	10	11	10	11	11	11
ECE	7	8	8	8	5	8	8
MECH	8	9	9	8	9	9	9

In [144...]

```
pd.pivot_table(df, index = 'Branch', aggfunc='max')
```

Out[144]:

	Chemistry	Computer	DOB	Maths	Physics	Roll No.	Section
Branch							
CS	57.0	66.0	30-01-2001	58.0	90.0	30	C
ECE	51.0	54.0	26-01-2001	51.0	90.0	26	B
MECH	52.0	53.0	25-01-2001	52.0	89.0	25	C

In [145...]

```
pd.pivot_table(df, index = 'Branch', columns = 'Section')
```

Out[145]:

	Chemistry			Computer			Maths			
Section	A	B	C	A	B	C	A	B	C	A
Branch										
CS	32.666667	23.5	35.0	40.285714	35.0	20.000000	33.333333	24.00	27.000000	50.142857
ECE	24.000000	39.5	NaN	31.750000	40.0	NaN	25.750000	39.75	NaN	55.666667
MECH	47.500000	25.0	42.0	48.600000	27.0	42.666667	47.500000	26.00	42.333333	48.600000