

In [9]:

```
import pandas as pd #import the pandas and aliasing as pd
import numpy as np
data = np.array(['a','b','c','d'])
s1 = pd.Series(data)
print(s1)
data = np.array([1,2,3,4])
s2 = pd.Series(data)
print(s2)
data = np.array(['a','b','c','d'])
s = pd.Series(data,index=[100,101,102,103])
print(s)
s = pd.Series([1,2,3,4,5],index = ['a','b','c','d','e'])
print(s[0],s[1],s[2])
print(s['a'])
print(s[['a','c','d']])
```

```
0    a
1    b
2    c
3    d
dtype: object
0    1
1    2
2    3
3    4
dtype: int64
100   a
101   b
102   c
103   d
dtype: object
1 2 3
1
a    1
c    3
d    4
dtype: int64
```

In [18]:

```
import pandas as pd
df = pd.DataFrame()
print(df)
data = [1,2,3,4,5]
df1 = pd.DataFrame(data)
print(df1)
data = [['Alex',10],['Bob',12],['larke',13]]
df2 = pd.DataFrame(data,columns=['Name','Age'])
print(df2)
data = {'Name':['Tom', 'Jack', 'Steve', 'Ricky'],
        'Age':[28,34,29,42]}
df = pd.DataFrame(data)
print(df)
d = {'one' : pd.Series([1, 2, 3], index=['a', 'b', 'c']),
     'two' : pd.Series([1, 2, 3, 4], index=['a', 'b', 'c', 'd'])}
df = pd.DataFrame(d)
print(df)
```

Empty DataFrame

Columns: []

Index: []

```
0
0 1
1 2
2 3
3 4
4 5
   Name  Age
0  Alex   10
1   Bob   12
2  larke   13
   Name  Age
0   Tom   28
1  Jack   34
2  Steve   29
3  Ricky   42
   one  two
a  1.0    1
b  2.0    2
c  3.0    3
d  NaN    4
```

In [58]:

```

import pandas as pd
d = {'Math' : pd.Series([90,10,34,23,43],index=[0,1,2,3,4]),
      'C-programming' : pd.Series([70,20,25,32,73],index=[0,1,2,3,4]),
      'Java': pd.Series([50,41,67,34,23],index=[0,1,2,3,4]),
      'Web-programming':pd.Series([55,66,23,76,67],index=[0,1,2,3,4])}
Studentdf = pd.DataFrame(d)
print(Studentdf)
print('-----')
print(Studentdf[2:4]) #Slice Rows
print('-----')
print(Studentdf.loc[0])
print('-----')
print(Studentdf.loc[2])
print('-----')
print(Studentdf.loc[0:2])
print('-----')
df = pd.DataFrame([[1, 2], [3, 4]], columns = ['a','b'])
df2 = pd.DataFrame([[5, 6], [7, 8]], columns = ['a','b'])
df = df.append(df2)
print(df)
print('-----')
extend = {"Math": pd.Series([50,41],index=[5,6]),
          'C-programming' : pd.Series([37,34],index=[5,6]),
          "Java":pd.Series([84,73],index=[5,6]),
          "Web-programming":pd.Series([82,55],index=[5,6])}
studentExtend = pd.DataFrame(extend)
Studentdf = Studentdf.append(studentExtend)
print(Studentdf)
print('-----')
Studentdf = Studentdf.drop(0)
print(Studentdf)
print('-----')
print(Studentdf.describe())
print('-----')
Studentdf = Studentdf.reindex([0,1,2,3,4,5,6,7])
print(Studentdf)
print('-----')
print(Studentdf['Math'].isnull())
print(Studentdf['C-programming'].isnull())
print(Studentdf['Java'].isnull())
print(Studentdf['Web-programming'].isnull())
print('-----')
print(Studentdf['Math'].notnull())
print(Studentdf['C-programming'].notnull())
print(Studentdf['Java'].notnull())
print(Studentdf['Web-programming'].notnull())
print('-----')
s1 = Studentdf
s2 = Studentdf
s1 = s1.fillna(0)
s2 = s2.fillna(method='pad')
print(s1)
print('-----')
print(s2)
print('-----')
Studentdf = Studentdf.drop(3)
print(Studentdf)

```

	Math	C-programming	Java	Web-programming
0	90	70	50	55
1	10	20	41	66
2	34	25	67	23
3	23	32	34	76
4	43	73	23	67

	Math	C-programming	Java	Web-programming
2	34	25	67	23
3	23	32	34	76

```

Math          90
C-programming 70
Java          50
Web-programming 55
Name: 0, dtype: int64

```

```

Math          34
C-programming 25
Java          67
Web-programming 23
Name: 2, dtype: int64

```

	Math	C-programming	Java	Web-programming
0	90	70	50	55
1	10	20	41	66
2	34	25	67	23

	a	b
0	1	2
1	3	4
0	5	6
1	7	8

	Math	C-programming	Java	Web-programming
0	90	70	50	55
1	10	20	41	66
2	34	25	67	23
3	23	32	34	76
4	43	73	23	67
5	50	37	84	82
6	41	34	73	55

	Math	C-programming	Java	Web-programming
1	10	20	41	66
2	34	25	67	23
3	23	32	34	76
4	43	73	23	67
5	50	37	84	82
6	41	34	73	55

	Math	C-programming	Java	Web-programming
count	6.000000	6.000000	6.000000	6.000000
mean	33.500000	36.833333	53.666667	61.500000
std	14.707141	18.776759	24.328310	21.002381
min	10.000000	20.000000	23.000000	23.000000
25%	25.750000	26.750000	35.750000	57.750000
50%	37.500000	33.000000	54.000000	66.500000
75%	42.500000	36.250000	71.500000	73.750000
max	50.000000	73.000000	84.000000	82.000000

	Math	C-programming	Java	Web-programming
0	NaN	NaN	NaN	NaN
1	10.0	20.0	41.0	66.0
2	34.0	25.0	67.0	23.0
3	23.0	32.0	34.0	76.0
4	43.0	73.0	23.0	67.0
5	50.0	37.0	84.0	82.0
6	41.0	34.0	73.0	55.0
7	NaN	NaN	NaN	NaN

-----

```

0      True
1     False
2     False
3     False
4     False
5     False
6     False
7      True

```

Name: Math, dtype: bool

```

0      True
1     False
2     False
3     False
4     False
5     False
6     False
7      True

```

Name: C-programming, dtype: bool

```

0      True
1     False
2     False
3     False
4     False
5     False
6     False
7      True

```

Name: Java, dtype: bool

```

0      True
1     False
2     False
3     False
4     False
5     False
6     False
7      True

```

Name: Web-programming, dtype: bool

-----

```

0     False
1      True
2      True
3      True
4      True
5      True
6      True
7     False

```

Name: Math, dtype: bool

```

0     False
1      True
2      True
3      True
4      True

```

5 True

6 True

7 False

Name: C-programming, dtype: bool

0 False

1 True

2 True

3 True

4 True

5 True

6 True

7 False

Name: Java, dtype: bool

0 False

1 True

2 True

3 True

4 True

5 True

6 True

7 False

Name: Web-programming, dtype: bool

-----

	Math	C-programming	Java	Web-programming
0	0.0	0.0	0.0	0.0
1	10.0	20.0	41.0	66.0
2	34.0	25.0	67.0	23.0
3	23.0	32.0	34.0	76.0
4	43.0	73.0	23.0	67.0
5	50.0	37.0	84.0	82.0
6	41.0	34.0	73.0	55.0
7	0.0	0.0	0.0	0.0

-----

	Math	C-programming	Java	Web-programming
0	NaN	NaN	NaN	NaN
1	10.0	20.0	41.0	66.0
2	34.0	25.0	67.0	23.0
3	23.0	32.0	34.0	76.0
4	43.0	73.0	23.0	67.0
5	50.0	37.0	84.0	82.0
6	41.0	34.0	73.0	55.0
7	41.0	34.0	73.0	55.0

-----

	Math	C-programming	Java	Web-programming
0	NaN	NaN	NaN	NaN
1	10.0	20.0	41.0	66.0
2	34.0	25.0	67.0	23.0
4	43.0	73.0	23.0	67.0
5	50.0	37.0	84.0	82.0
6	41.0	34.0	73.0	55.0
7	NaN	NaN	NaN	NaN

In [59]:

```

import pandas as pd
data = {'Name':['Tom', 'Jack', 'Steve', 'Ricky'],
        'Age':[28,34,29,42]}
df = pd.DataFrame(data)
df.sum()
print(Studentdf.count())
print('-----')
print(Studentdf.sum(axis = 0))
print('-----')
Studentdf.sum(axis = 1)

```

```

Math          5
C-programming 5
Java          5
Web-programming 5
dtype: int64
-----
Math          178.0
C-programming 189.0
Java          288.0
Web-programming 293.0
dtype: float64
-----

```

Out[59]:

```

0         0.0
1       137.0
2       149.0
4       206.0
5       253.0
6       203.0
7         0.0
dtype: float64

```

In [33]:

```
import pandas as pd
import numpy as np
#Create a Dictionary of series
d = {'Name':pd.Series(['Tom','James','Ricky','Vin','Steve','Smith','Jack',
'Lee','David','Gasper','Betina','Andres']),
'Age':pd.Series([25,26,25,23,30,29,23,34,40,30,51,46]),
'Rating':pd.Series
([4.23,3.24,3.98,2.56,3.20,4.6,3.8,3.78,2.98,4.80,4.10,3.65])
}
#Create a DataFrame
df = pd.DataFrame(d)
df.describe()
```

Out[33]:

	Age	Rating
count	12.000000	12.000000
mean	31.833333	3.743333
std	9.232682	0.661628
min	23.000000	2.560000
25%	25.000000	3.230000
50%	29.500000	3.790000
75%	35.500000	4.132500
max	51.000000	4.800000



In [36]:

```
import pandas as pd
df = pd.DataFrame(np.random.randn(5, 3), index=['a', 'c', 'e', 'f', 'h'], columns=['one', 'two', 'three'])
df = df.reindex(['a', 'b', 'c', 'd', 'e', 'f', 'g', 'h'])
print(df['one'].isnull())
print(df['two'].notnull())
```

```
a    False
b     True
c    False
d     True
e    False
f    False
g     True
h    False
Name: one, dtype: bool
a     True
b    False
c     True
d    False
e     True
f     True
g    False
h     True
Name: two, dtype: bool
```

In [43]:

```
left = pd.DataFrame({
'id':[1,2,3,4,5],
'Name': ['Alex', 'Amy', 'Allen', 'Alice', 'Ayoung'],
'subject_id':['sub1','sub2','sub4','sub6','sub5']})
right = pd.DataFrame(
{'id':[1,2,3,4,5],
'subject_name':['Math','English',
'Social Science','Programming','Physics']})
print(left)
print(right)
```

```
id    Name subject_id
0     1   Alex      sub1
1     2   Amy      sub2
2     3  Allen      sub4
3     4  Alice      sub6
4     5 Ayoung      sub5
id    subject_name
0     1         Math
1     2       English
2     3  Social Science
3     4   Programming
4     5       Physics
```

In [45]:

```
merged_table = pd.merge(left,right,on='id')
print(merged_table)
```

	id	Name	subject_id	subject_name
0	1	Alex	sub1	Math
1	2	Amy	sub2	English
2	3	Allen	sub4	Social Science
3	4	Alice	sub6	Programming
4	5	Ayoung	sub5	Physics

In [60]:

```
section_left = pd.DataFrame({
'id':[0,1,2,3,4,5],
'section':['one','one','one','two','two','two']})
section_right = pd.DataFrame(
{'id':[0,1,2,3,4,5],
'Name':["Mat","Mike","John","Peecha","Prabu","Bunchu"],
'Math':[34,12,34,12,65,23],
'English':[23,13,51,62,12,42]})
score = pd.merge(section_left,section_right,on='id')
print(score)
score_sec = score.groupby('section')
score_sec.mean()
```

	id	section	Name	Math	English
0	0	one	Mat	34	23
1	1	one	Mike	12	13
2	2	one	John	34	51
3	3	two	Peecha	12	62
4	4	two	Prabu	65	12
5	5	two	Bunchu	23	42

Out[60]:

	id	Math	English
section			
one	1.0	26.666667	29.000000
two	4.0	33.333333	38.666667