```
from pandas import Series, DataFrame
import pandas as pd
import numpy as np
```

Series

- Series는 index를 공유함에 있어 list와 다르다.

```
# 버전 확인
print("pandas 버전 ", pd.__version__)
```

```
pandas 버전  0.24.2
```

```
obj = Series([4,7,-5,3])
obj
```

```
0    4
1    7
2   -5
3    3
dtype: int64
```

```
obj.index    # 인덱스 확인
```

```
RangeIndex(start=0, stop=4, step=1)
```

```
obj.values   # 내장 값 확인
```

```
array([ 4,  7, -5,  3])
```

```
obj2 = Series([4,7,-5,3],index = ['d','b','a','c'])  # 생성시 인덱스 지정
obj2
```

```
d    4
b    7
a   -5
c    3
dtype: int64
```

```
obj2.index
```

```
Index(['d', 'b', 'a', 'c'], dtype='object')
```

```
obj2['a']
```

```
-5
```

```
obj2['d'] = 6
obj2[['c','a','d']]
```

```
c    3
a   -5
d    6
dtype: int64
```

```
obj2[obj2>0]
```

```
d    6
b    7
c    3
dtype: int64
```

```
obj2*2
```

```
d    12
b    14
a   -10
c     6
dtype: int64
```

```
np.exp(obj2)
```

```
d     403.428793
b    1096.633158
a       0.006738
c      20.085537
dtype: float64
```

```
# 간단 예제
## 홍길동 팀 대항 게임 5일간 점수
score = Series([1000, 14000, 3000, 3000, 1000])

## 날짜를 인덱스로 시리즈 만들기
gildong = Series([1000, 14000, 3000, 3000, 1000],
                 index = [20191005, 20191006, 20191007, 20191008, 20191009])
```

```
# 데이터프레임
A = pd.DataFrame({'date' : [20191005, 20191006, 20191007, 20191008, 20191009],
                  'gildong' : [1000, 14000, 3000, 3000, 1000],
                  'toto' : [2000, 12000, 5600, 3000, 1200]})
```

```
A = A.set_index('date')   # 인덱스 지정
A
```

date	gildong	toto
20191005	1000	2000
20191006	14000	12000
20191007	3000	5600
20191008	3000	3000
20191009	1000	1200

```
'b' in obj2
```

True

```
'e' in obj2
```

False

```
sdata = {'Ohio' : 35000, 'Texas' : 71000, 'Oregon': 16000, 'Utah' : 5000}
obj3 = Series(sdata)
obj3
```

```
Ohio      35000
Texas     71000
Oregon    16000
Utah       5000
dtype: int64
```

```
states = ['California','Ohio','Oregon','Texas']
obj4 = Series(sdata, index = states)
obj4
```

```
California        NaN
Ohio          35000.0
Oregon        16000.0
Texas         71000.0
dtype: float64
```

```
pd.isnull(obj4)
```

```
California     True
Ohio          False
Oregon        False
Texas         False
dtype: bool
```

```
pd.notnull(obj4)
```

```
California    False
Ohio           True
Oregon         True
Texas          True
dtype: bool
```

```
obj4.isnull()
```

```
California     True
Ohio          False
Oregon        False
Texas         False
dtype: bool
```

```
obj3 + obj4
```

```
California         NaN
Ohio           70000.0
Oregon         32000.0
Texas         142000.0
Utah               NaN
dtype: float64
```

```
obj4.name = 'population'      # 시리즈 이름
obj4.index.name = 'state'     # 인덱스 이름
obj4
```

```
state
California        NaN
Ohio          35000.0
Oregon        16000.0
Texas         71000.0
Name: population, dtype: float64
```

```
obj
```

```
0    4
1    7
2   -5
3    3
dtype: int64
```

```
obj.index = ['Bob','Steve','Jeff','Ryan']
obj
```

```
Bob      4
Steve    7
Jeff    -5
Ryan     3
dtype: int64
```

DataFrame

```
data = {'state' : ['Ohio','Ohio','Ohio','Nevada','Nevada'],
        'year' : [2000,2001,2002,2001,2002],
        'pop' : [1.5, 1.7, 3.6, 2.4, 2.9]}
frame = DataFrame(data)
```

```
frame
```

	state	year	pop
0	Ohio	2000	1.5
1	Ohio	2001	1.7
2	Ohio	2002	3.6
3	Nevada	2001	2.4
4	Nevada	2002	2.9

```
DataFrame(data, columns = ['year','state','pop'])
```

	year	State	pop
0	2000	Ohio	1.5
1	2001	Ohio	1.7
2	2002	Ohio	3.6
3	2001	Nevada	2.4
4	2002	Nevada	2.9

```
frame2 = DataFrame(data, columns = ['year','state','pop','debt'],
                   index = ['one','two','three','four','five'])
frame2
```

	year	state	pop	debt
one	2000	Ohio	1.5	NaN
two	2001	Ohio	1.7	NaN
three	2002	Ohio	3.6	NaN
four	2001	Nevada	2.4	NaN
five	2002	Nevada	2.9	NaN

```
frame2.columns
```

```
frame2['state']
```

```
one        Ohio
two        Ohio
three      Ohio
four     Nevada
five     Nevada
Name: state, dtype: object
```

```
frame.year
```

```
0    2000
1    2001
2    2002
3    2001
4    2002
Name: year, dtype: int64
```

```
frame2.ix['three']
```

```
C:₩Python₩Anaconda3₩lib₩site-packages₩ipykernel_launcher.py:1: DeprecationWarning:
.ix is deprecated. Please use
.loc for label based indexing or
.iloc for positional indexing

See the documentation here:
http://pandas.pydata.org/pandas-docs/stable/indexing.html#ix-indexer-is-deprecated
  """Entry point for launching an IPython kernel.
year     2002
state    Ohio
pop       3.6
debt      NaN
Name: three, dtype: object
```

```
frame2.loc['three']
```

```
year     2002
state    Ohio
pop       3.6
debt      NaN
Name: three, dtype: object
```

```
frame2['debt'] = 16.5
frame2
```

	year	state	pop	debt
one	2000	Ohio	1.5	16.5
two	2001	Ohio	1.7	16.5
three	2002	Ohio	3.6	16.5
four	2001	Nevada	2.4	16.5
five	2002	Nevada	2.9	16.5

```
frame2.debt = 15
frame2
```

	year	state	pop	debt
one	2000	Ohio	1.5	15
two	2001	Ohio	1.7	15
three	2002	Ohio	3.6	15
four	2001	Nevada	2.4	15
five	2002	Nevada	2.9	15

```
frame2['debt'] = np.arange(5.)
frame2
```

	year	state	pop	debt
one	2000	Ohio	1.5	0.0
two	2001	Ohio	1.7	1.0
three	2002	Ohio	3.6	2.0
four	2001	Nevada	2.4	3.0
five	2002	Nevada	2.9	4.0

```
val = Series([-1.2,-1.5,-1.7],index = ['two','three','five'])
frame2.debt = val
frame2
```

```
frame2['eastern'] = frame2.state == 'Ohio'
frame2
```

	year	state	pop	debt	eastern
one	2000	Ohio	1.5	NaN	True
two	2001	Ohio	1.7	-1.2	True
three	2002	Ohio	3.6	-1.5	True
four	2001	Nevada	2.4	NaN	False
five	2002	Nevada	2.9	-1.7	False

```
# 칼럼 삭제
del frame2['eastern']
frame2
```

	year	state	pop	debt
one	2000	Ohio	1.5	NaN
two	2001	Ohio	1.7	-1.2
three	2002	Ohio	3.6	-1.5
four	2001	Nevada	2.4	NaN
five	2002	Nevada	2.9	-1.7

```
# 중첩된 dict
# 바깥 사전의 키값은 칼럼이 되고, 그 안의 키값은 로우가 된다.

pop = { 'Nevada' : {2001: 2.4, 2002: 2.9},
        'Ohio' : {2000: 1.5, 2001: 1.7, 2002 : 3.6}}
frame3 = DataFrame(pop)
```

```
frame3
```

	Nevada	Ohio
2000	NaN	1.5
2001	2.4	1.7
2002	2.9	3.6

```
frame3.T
```

```
DataFrame(pop, index = [2001,2002,2003])
```

	Nevada	Ohio
2001	2.4	1.7
2002	2.9	3.6
2003	NaN	NaN

```
frame3
```

	Nevada	Ohio
2000	NaN	1.5
2001	2.4	1.7
2002	2.9	3.6

```
pdata = {'Ohio' : frame3['Ohio'][:-1],
        'Nevada' : frame3['Nevada'][:2]}
DataFrame(pdata)
```

	Ohio	Nevada
2000	1.5	NaN
2001	1.7	2.4

```
frame3.index.name = 'year' ; frame3.columns.name = 'state'
frame3
```

state	Nevada	Ohio
year		
2000	NaN	1.5
2001	2.4	1.7
2002	2.9	3.6

```
frame3.values
```

```
array([[nan, 1.5],
       [2.4, 1.7],
       [2.9, 3.6]])
```

```
frame2.values
```

```
array([[2000, 'Ohio', 1.5, nan],
       [2001, 'Ohio', 1.7, -1.2],
       [2002, 'Ohio', 3.6, -1.5],
       [2001, 'Nevada', 2.4, nan],
       [2002, 'Nevada', 2.9, -1.7]], dtype=object)
```

```
frame2
```

	year	state	pop	debt
one	2000	Ohio	1.5	NaN
two	2001	Ohio	1.7	-1.2
three	2002	Ohio	3.6	-1.5
four	2001	Nevada	2.4	NaN
five	2002	Nevada	2.9	-1.7

```
obj = Series(range(3), index=['a','b','c'])
index = obj.index
index
```

```
Index(['a', 'b', 'c'], dtype='object')
```

```
index[1:]
```

```
Index(['b', 'c'], dtype='object')
```

```
#색인 객체 변경 시 TypeError 발생
```

```
index = pd.Index(np.arange(3))
obj2 = Series([1.5,-2.5,0],index=index)
obj2.index is index
```

```
True
```

```
frame3
```

state	Nevada	Ohio
year		
2000	NaN	1.5
2001	2.4	1.7
2002	2.9	3.6

```
'Ohio' in frame3.columns
```

True

```
2003 in frame3.index
```

False