



CHAPTER

05

Pandas 라이브러리

Pandas Library

컴퓨터소프트웨어공학과
김대영

파이썬 실습 2

```
import pandas as pd
import numpy as np

def main():

    obj = pd.Series(np.arange(4.), index=
['a','b','c','d'])
    print(obj)
    print(obj['b'])
    print(obj[2:4])
    print(obj[[1,3]])
    print(obj['b':'c'])

    obj = pd.Series(np.arange(4.), index=
['a','b','c','d'])
    print(obj['b':'c'])
    obj['b':'c'] = 5
    print(obj)

    data = pd.DataFrame(np.arange(16).reshape((4,4)),
index=['Ohio', 'Colorado','Utah','New York'],
columns=['one','two','three','four'])
    print(data)
    print(data['two'])
    print(data[:2])

    data = pd.DataFrame(np.arange(16).reshape((4,4)),
index=['Ohio', 'Colorado','Utah','New York'],
columns=['one','two','three','four'])
    d = data.loc['Colorado',['two','three']]
    print(d)
    d = data.iloc[2,[3,0,1]]
    print(d)
```

```
s1 = pd.Series([7.3,-2.5,3.4,1.5], index=
['a','c','d','e'])
s2 = pd.Series([-2.1,3.6,-1.5,4,3.1], index=
['a','c','e','f','g'])
s = s1 + s2
print(s)

df1 = pd.DataFrame(np.arange(9).reshape((3,3)),
columns=list('bcd'), index=['Ohio','Texas','Colorado'])
df2 = pd.DataFrame(np.arange(12).reshape((4,3)),
columns=list('bde'), index=
['Utah','Ohio','Texas','Oregon'])
df = df1 + df2
print(df1)
print(df2)
print(df)

df1 = pd.DataFrame(np.arange(12).reshape((3,4)),
columns=list('abcd'))
df2 = pd.DataFrame(np.arange(20).reshape((4,5)),
columns=list('abcde'))
df2.loc[1,'b'] = np.nan
print(df1)
print(df2)

df = df1.add(df2, fill_value=0)
print(df)

df = 1/df1
print(df)

df = df1.rdiv(1)
print(df)
```

파이썬 실습 2

```
frame = pd.DataFrame(np.arange(12).reshape((4,3)),
columns=list('bde'),
index=['Utah','Ohio','Texas','Oregon'])
series = frame.iloc[0]
print(frame)
print(series)
val = frame - series
print(val)

frame = pd.DataFrame(np.arange(12).reshape((4,3)),
columns=list('bde'),
index=['Utah','Ohio','Texas','Oregon'])
series3 = frame['d']
print(frame)
print(series3)
val = frame.sub(series3, axis='index')
print(val)

frame = pd.DataFrame(np.random.randn(4,3),
columns=list('bde'), index=['Utah','Ohio','Texas',
'Oregon'])
print(frame)
# print(np.abs(frame))
f = lambda x : x.max() - x.min()
# v = frame.apply(f)
# print(v)

v= frame.apply(f, axis='columns')
print(v)
def f(x):
    return pd.Series([x.min(), x.max()],
index=['min','max'])
v = frame.apply(f)
print(v)
```

```
format = lambda x: '%.2f' % x
v = frame.applymap(format)
print(v)

obj = pd.Series(range(4), index = ['d','a','b','c'])
print(obj.sort_index())
frame = pd.DataFrame(np.arange(8).reshape((2,4)),
index=['three','one'], columns=['d','a','b','c'])
print(frame.sort_index())
print(frame.sort_index(axis=1))

obj = pd.Series([7,-5,7,4,2,0,4])
print(obj.rank())
print(obj.rank(method='first'))
obj = pd.Series(range(5), index=
['a','a','b','b','c'])
print(obj)
print(obj.index.is_unique)
print(obj['a'])

df = pd.DataFrame([[1.4, np.nan], [7.1, -4.5],
[np.nan, np.nan], [0.75, -1.3]], index=['a','b','c','d'],
columns=['one','two'])
print(df)
print(df.sum())
print(df.sum(axis='columns'))
print(df.mean(axis='columns', skipna=False))
print(df.idxmax())
print(df.cumsum())
print(df.describe())

if __name__ == '__main__': # main()
    main()
```