Zeppelin

2017-02-16-Class-Lab

default

FINISHED

```
%pyspark
from pandas import Series, DataFrame
import pandas as pd
obj = Series([4, 7, -5, 3])
obj
obj.values
obj.index
```

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

Took 0 sec. Last updated by anonymous at February 16 2017, 6:44:16 AM.

FINISHED

```
%pyspark
from pandas import Series, DataFrame
import pandas as pd
obj2 = Series([4, 7, -5, 3], index=['d', 'b', 'a', 'c'])
obj2
obj2.index
obj2['a']
obj2['d'] = 6
obj2[['c', 'a', 'd']]
obj2
obj2[obj2 > 0]
obj2 * 2
```

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

Took 0 sec. Last updated by anonymous at February 16 2017, 6:44:07 AM.

FINISHED

```
%pyspark
```

2017-02-16-Class-Lab

2017-02-16-Class-Lab

```
import numpy as np
'b' in obj2
'e' in obj2
```

Took 0 sec. Last updated by anonymous at February 16 2017, 6:45:33 AM.

FINISHED

```
%pyspark
import numpy as np
obj3 = Series(sdata)
obj3
states = ['California', 'Ohio', 'Oregon', 'Texas']
```

Took 0 sec. Last updated by anonymous at February 16 2017, 6:45:49 AM.

FINISHED

```
%pyspark
import numpy as np
obj4 = Series(sdata, index=states)
obj4
pd.isnull(obj4)
pd.notnull(obj4)
obj4.isnull()
```

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

2017-02-16-Class-LabUntitled Untitled Untitled Untitled Untitled Untitled Untitled Untitled Untitled Untitled

Took 0 sec. Last updated by anonymous at February 16 2017, 6:46:08 AM.

FINISHED

```
%pyspark
import numpy as np
obj3
obj4
obj3 + obj4
obj4.name = 'population'
obj4.index.name = 'state'
obj4
obj.index = ['Bob', 'Steve', 'Jeff', 'Ryan']
obj
```

Steve 4
Jeff -5
Ryan -3
dtype: int64

Took 0 sec. Last updated by anonymous at February 16 2017, 6:47:03 AM.

Zeppelin

2017-02-16-Class-Lab

default

FINISHED

```
%pyspark
import numpy as np
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
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

Took 0 sec. Last updated by anonymous at February 16 2017, 6:51:37 AM.

FINISHED

```
%pyspark
import numpy as np
frame2 = DataFrame(data, columns=['year', 'state', 'pop', 'debt'],
                   index=['one', 'two', 'three', 'four', 'five'])
frame2
frame2.columns
frame2['state']
frame2.year
frame2.ix['three']
frame2['debt'] = 16.5
frame2
frame2['debt'] = np.arange(5.)
frame2
val = Series([-1.2, -1.5, -1.7], index=['two', 'four', 'five'])
frame2['debt'] = val
frame2
```

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

```
Index([u'year', u'state', u'pop', u'debt'], dtype='object')
```

default

Took 0 sec. Last updated by anonymous at February 16 2017, 6:52:09 AM.

FINISHED

```
%pyspark
import numpy as np
pop = {'Nevada': {2001: 2.4, 2002: 2.9},
       'Ohio': {2000: 1.5, 2001: 1.7, 2002: 3.6}}
frame3 = DataFrame(pop)
frame3
frame3.T
pdata = {'Ohio': frame3['Ohio'][:-1],
        'Nevada': frame3['Nevada'][:2]}
DataFrame(pdata)
frame3.index.name = 'year'; frame3.columns.name = 'state'
frame3
frame3.values
frame2.values
```

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

Took 0 sec. Last updated by anonymous at February 16 2017, 6:52:49 AM.

READY