FINISHED

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

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

Took 26 sec. Last updated by anonymous at February 16 2017, 9:17:49 PM.

FINISHED

```
%pyspark
print(obj.values)
obj.index
obj2 = Series([4, 7, -5, 3], index=['d', 'b', 'a', 'c'])
obj2
```

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

Took 0 sec. Last updated by anonymous at February 16 2017, 9:17:53 PM.

FINISHED

```
%pyspark
print(obj2.index)
obj2['a']
obj2['d'] = 6
obj2[['c', 'a', 'd']]
obj2
```

```
Index([u'd', u'b', u'a', u'c'], dtype='object')
d    6
b    7
a   -5
c    3
dtype: int64
```

Took 0 sec. Last updated by anonymous at February 16 2017, 9:20:06 PM.

FINISHED

```
%pyspark
print(obj2[obj2 > 0])
obj2 * 2
```

```
d    6
b    7
c    3
```

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

Took 0 sec. Last updated by anonymous at February 16 2017, 9:18:15 PM.

FINISHED

```
%pyspark
import numpy as np
np.exp(obj2)
print('b' in obj2)
'e' in obj2
```

```
True
False
```

Took 0 sec. Last updated by anonymous at February 16 2017, 9:18:38 PM.

FINISHED

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

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

Took 0 sec. Last updated by anonymous at February 16 2017, 9:19:48 PM.

FINISHED

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

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

Took 0 sec. Last updated by anonymous at February 16 2017, 9:21:48 PM.

FINISHED

```
%pyspark
pd.isnull(obj4)
```

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

Took 0 sec. Last updated by anonymous at February 16 2017, 9:22:36 PM.

FINISHED

```
%pyspark
pd.notnull(obj4)
```

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

Took 0 sec. Last updated by anonymous at February 16 2017, 9:22:42 PM.

FINISHED

```
%pyspark
obj4.isnull()
```

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

Took 0 sec. Last updated by anonymous at February 16 2017, 9:22:49 PM.

FINISHED

```
%pyspark
print(obj3)
obj4
```

```
Ohio      35000
Oregon    16000
Texas     71000
Utah       5000
dtype: int64
California        NaN
Ohio          35000.0
Oregon        16000.0
Texas         71000.0
dtype: float64
```

Took 0 sec. Last updated by anonymous at February 16 2017, 9:23:00 PM.

FINISHED

```
%pyspark
obj3 + obj4
obj4.name = 'population'
obj4.index.name = 'state'
obj4
```

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

Took 0 sec. Last updated by anonymous at February 16 2017, 9:24:00 PM.

FINISHED

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

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

Took 0 sec. Last updated by anonymous at February 16 2017, 9:24:43 PM.

FINISHED

```
%pyspark
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
```

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

Took 0 sec. Last updated by anonymous at February 16 2017, 9:25:02 PM.

FINISHED


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

Took 0 sec. Last updated by anonymous at February 16 2017, 9:27:07 PM.

FINISHED

```
%pyspark
frame2.columns
frame2['state']
frame2.year
frame2.ix['three']
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

Took 0 sec. Last updated by anonymous at February 16 2017, 9:28:22 PM.

FINISHED

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

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

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

Took 0 sec. Last updated by anonymous at February 16 2017, 9:28:36 PM.

FINISHED

```
%pyspark
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   NaN    True
four   2001  Nevada  2.4  -1.5   False
five   2002  Nevada  2.9  -1.7   False
```

Took 1 sec. Last updated by anonymous at February 16 2017, 9:29:50 PM.

Lab3Feb16 Untitled Untitled Untitled Untitled Untitled Untitled Untitled Untitled Untitled Untitled Untitled Unti

FINISHED

Zeppelin

Lab3Feb16

default

```
%pyspark
...e2['eastern'] = ...ame2.state == 'Ohio'
...e2
```

```
       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   NaN    True
four   2001  Nevada  2.4  -1.5   False
five   2002  Nevada  2.9  -1.7   False
```

Took 0 sec. Last updated by anonymous at February 16 2017, 9:30:49 PM.

FINISHED

```
%pyspark
del frame2['eastern']
frame2.columns
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
```

Took 0 sec. Last updated by anonymous at February 16 2017, 9:32:32 PM.

FINISHED

```
%pyspark
frame3.T
```

```
        2000  2001  2002
Nevada   NaN   2.4   2.9
Ohio     1.5   1.7   3.6
```

Took 0 sec. Last updated by anonymous at February 16 2017, 9:32:43 PM.

FINISHED

```
%pyspark
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, 9:33:21 PM.

READY

Lab3Feb16 Untitled Untitled Untitled Untitled Untitled Untitled Untitled Untitled Untitled Untitled Untitled Unti

Zeppelin

Lab3Feb16 default