feb16 poornima

```
%pyspark
from pandas import Series, DataFrame
import pandas as pd
```

FINISHED

Took 0 sec. Last updated by anonymous at February 16 2017, 7:22:28 PM.

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

FINISHED

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

Took 0 sec. Last updated by anonymous at February 16 2017, 7:24:47 PM.

```
%pyspark
obj.values
obj.index
obj2 = Series([4, 7, -5, 3], index=['d', 'b', 'a', 'c'])
obj2
obj2.index
obj2['a']
obj2['d'] = 6
```

FINISHED

Took 0 sec. Last updated by anonymous at February 16 2017, 7:25:02 PM.

```
%pyspark
obj2[['c', 'a', 'd']]
obj2
obj2[obj2 > 0]
obj2 * 2
```

FINISHED

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

Took 0 sec. Last updated by anonymous at February 16 2017, 7:25:15 PM.

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

FINISHED

```
sdata = {'Ohio': 35000, 'Texas': 71000, 'Oregon': 16000, 'Utah': 5000}
obj3 = Series(sdata)
obj3
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, 7:25:32 PM.

```
%pyspark
pd.isnull(obj4)
pd.notnull(obj4)
obj4.isnull()
obj3
obj4
obj3 + obj4
obj4.name = 'population'
obj4.index.name = 'state'
obj4
obj.index = ['Bob', 'Steve', 'Jeff', 'Ryan']
obj
```

FINISHED

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

Took 0 sec. Last updated by anonymous at February 16 2017, 7:25:53 PM.

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

FINISHED

Took 0 sec. Last updated by anonymous at February 16 2017, 7:26:09 PM.

```
%pyspark
frame2
frame2.columns
frame2['state']
frame2.year
frame2.ix['three']
frame2['debt'] = 16.5
frame2
frame2['debt'] = np.arange(5.)
```

FINISHED

```
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
```

Took 0 sec. Last updated by anonymous at February 16 2017, 7:26:21 PM.

```
%pyspark
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
```

FINISHED

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

Took 0 sec. Last updated by anonymous at February 16 2017, 7:26:37 PM.

```
%pyspark
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)
```

FINISHED

```
      Nevada  Ohio
2000     NaN   1.5
2001     2.4   1.7
```

Took 1 sec. Last updated by anonymous at February 16 2017, 7:26:50 PM.

```
%pyspark
frame3.index.name = 'year'; frame3.columns.name = 'state'
frame3
frame3.values
frame2.values
```

FINISHED

```
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, 7:27:01 PM.

```
%pyspark
```

READY