

Python Assingment



default ▾

FINISHED ▶ ⌵ 📖 ⚙

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

df = 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'])
```

Took 0 sec. Last updated by anonymous at February 23 2017, 6:36:26 PM.

FINISHED ▶ ⌵ 📖 ⚙

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

df

```
   one  two
a  1.40 NaN
b  7.10 -4.5
c   NaN NaN
d  0.75 -1.3
```

Took 0 sec. Last updated by anonymous at February 23 2017, 6:35:01 PM.

FINISHED ▶ ⌵ 📖 ⚙

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

df.sum()

```
one    9.25
two   -5.80
dtype: float64
```

Took 0 sec. Last updated by anonymous at February 23 2017, 6:34:12 PM.

FINISHED ▶ ⌵ 📖 ⚙

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

df.sum(axis=1)

```
a    1.40
b    2.60
c     NaN
d   -0.55
```

Python Assingment



FINISHED 

default

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

df.mean(axis=1,skipna=False)

a      NaN
b      1.300
c      NaN
d     -0.275
dtype: float64
```

Took 0 sec. Last updated by anonymous at February 23 2017, 6:34:38 PM.

FINISHED 

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

df.idxmax()

one     b
two     d
dtype: object
```

Took 0 sec. Last updated by anonymous at February 23 2017, 6:35:55 PM.

FINISHED 

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

df.describe()

count      one      two
count  3.000000  2.000000
mean    3.083333 -2.900000
std      3.493685  2.262742
min      0.750000 -4.500000
25%      1.075000 -3.700000
50%      1.400000 -2.900000
75%      4.250000 -2.100000
max       7.100000 -1.300000
```

Took 0 sec. Last updated by anonymous at February 23 2017, 6:36:15 PM.

FINISHED 

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

obj = Series(['a','a','b','c'] * 4)
obj

0      a
1      a
2      b
3      c
```



Zeppelin

Python Assingment



default ▾

```
4 a
5 a
6 a
7 c
8 a
9 a
10 b
11 c
12 a
13 a
14 b
15 c
```

dtype: object

Took 0 sec. Last updated by anonymous at February 23 2017, 6:37:15 PM.

FINISHED ▶ ⌵ 📖 ⚙

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

obj.describe()
```

```
count      16
unique       3
top         a
freq        8
dtype: object
```

Took 0 sec. Last updated by anonymous at February 23 2017, 6:38:04 PM.

FINISHED ▶ ⌵ 📖 ⚙

```
%pyspark

from pandas_datareader import data, wb
all_data = {}
```

Took 0 sec. Last updated by anonymous at February 23 2017, 7:05:20 PM.

FINISHED ▶ ⌵ 📖 ⚙

```
%pyspark
import pandas_datareader as wb

for ticker in ['AAPL','IBM','MSFT','GOOG']:
    all_data[ticker] = wb.get_data_yahoo(ticker)
price = DataFrame({tic: data['Adj Close']
                    for tic, data in all_data.items()})
volume = DataFrame({tic: data['Volume']
                     for tic, data in all_data.items()})
```

Took 1 sec. Last updated by anonymous at February 23 2017, 7:10:10 PM.

FINISHED ▶ ⌵ 📖 ⚙

```
%pyspark
```

Python Assignment Zeppelin

Python Assignment

	AAPL	GOOG	IBM	MSFT
2017-02-15	0.003629	-0.001792	0.008605	-0.000619
2017-02-16	-0.001181	0.006325	-0.001376	-0.000155
2017-02-17	0.002734	0.004744	-0.004189	0.001550
2017-02-21	0.007221	0.004335	-0.002269	-0.002012
2017-02-22	0.002999	-0.001082	0.004937	-0.002016

Took 1 sec. Last updated by anonymous at February 23 2017, 7:10:36 PM.

```
%pyspark

returns.MSFT.corr(returns.IBM)
```

FINISHED ▶ ⌵ 📖 ⚙

0.49515377802280924

Took 0 sec. Last updated by anonymous at February 23 2017, 7:16:03 PM.

```
%pyspark

returns.MSFT.cov(returns.IBM)
```

FINISHED ▶ ⌵ 📖 ⚙

8.5977652563835441e-05

Took 0 sec. Last updated by anonymous at February 23 2017, 7:16:35 PM.

```
%pyspark

returns.corr()
```

FINISHED ▶ ⌵ 📖 ⚙

	AAPL	GOOG	IBM	MSFT
AAPL	1.000000	0.409541	0.381549	0.388972
GOOG	0.409541	1.000000	0.402872	0.470820
IBM	0.381549	0.402872	1.000000	0.495154
MSFT	0.388972	0.470820	0.495154	1.000000

Took 0 sec. Last updated by anonymous at February 23 2017, 7:17:29 PM.

```
%pyspark

returns.cov()
```

FINISHED ▶ ⌵ 📖 ⚙

	AAPL	GOOG	IBM	MSFT
AAPL	0.000270	0.000105	0.000075	0.000093
GOOG	0.000105	0.000244	0.000075	0.000107
IBM	0.000075	0.000075	0.000144	0.000086
MSFT	0.000093	0.000107	0.000086	0.000210

Took 0 sec. Last updated by anonymous at February 23 2017, 7:17:45 PM.

```
%pyspark
```

FINISHED ▶ ⌵ 📖 ⚙



returns.corrwith(returns.IBM)

Zeppelin

Python Assingment

```
AAPL    0.381549
GOOG    0.403872
IBM      1.000000
MSFT    0.495154
dtype: float64
```



default ▾

Took 0 sec. Last updated by anonymous at February 23 2017, 7:18:05 PM.

```
%pyspark
```

FINISHED ▶ ⌵ ⌵ ⚙

```
returns.corrwith(volume)
```

```
AAPL    -0.074323
GOOG    -0.009670
IBM      -0.194432
MSFT    -0.091017
dtype: float64
```

Took 0 sec. Last updated by anonymous at February 23 2017, 7:18:22 PM.

READY ▶ ⌵ ⌵ ⚙