

Feb-23



default ▼

```
%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'])

df
```

FINISHED    

	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, 6:54:09 PM.

```
%pyspark
df.sum()
```

FINISHED ▶ 🔍 📖 ⚙️

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

Took 0 sec. Last updated by anonymous at February 23 2017, 6:54:33 PM.

```
%pyspark
df.sum(axis=1)
```

FINISHED ▶ 🔍 📖 ⚙️

```
a    1.40
b    2.60
c     NaN
d   -0.55
dtype: float64
```

Took 0 sec. Last updated by anonymous at February 23 2017, 6:54:58 PM.

```
%pyspark
df.mean(axis=1,skipna=False)
```

FINISHED ▶ 🔍 📖 ⚙️

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

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

```
%pyspark
df.idxmax()
```

FINISHED ▶ ↻ 📖 ⚙️

Page 23

# Zeppelin

~~Took 0 sec. Last updated by anonymous at February 23 2017, 6:55:44 PM.~~

Feb-23






 default ▼

```
%pyspark
df.describe()
```

	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, 6:56:29 PM.

```
%pyspark
obj = Series(['a','a','b','c'] * 4)
obj
obj.describe()
```

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

Took 0 sec. Last updated by anonymous at February 23 2017, 6:56:53 PM.

```
%pyspark
from pandas_datareader import data, wb
all_data = {}
for ticker in ['AAPL','IBM','MSFT','GOOG']:
    all_data[ticker] = web.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()})
returns = price.pct_change()
returns.tail()
```


	AAPL	GOOG	IBM	MSFT
Date				
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, 6:53:12 PM.

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

0.49515377802280924

Took 1 sec. Last updated by anonymous at February 23 2017, 6:53:53 PM.



# Zeppelin

FINISHED
▶
⌵
📖
⚙️

spark
returns.MSFT.csv(returns.IBM)

8-577652963835441e-05
▶
⌵
📖
📝
📁
⬇️
🔖
🗑️
🕒

Feb-23
⌨️
⚙️
🔒
default ▼

Took 0 sec. Last updated by anonymous at February 23 2017, 6:57:17 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, 6:57:34 PM.

```
%pyspark
returns.cov()
```

	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, 6:57:58 PM.

```
%pyspark
returns.corrwith(returns.IBM)
```

AAPL	0.381549
GOOG	0.402872
IBM	1.000000
MSFT	0.495154

dtype: float64

Took 0 sec. Last updated by anonymous at February 23 2017, 6:58:19 PM.

```
%pyspark
returns.corrwith(volume)

AAPL    -0.074323
GOOG    -0.009665
IBM     -0.194432
MSFT    -0.091017

dtype: float64
```

FINISHED ▶ ⌕ 📖 ⚙

Took 0 sec. Last updated by anonymous at February 23 2017, 6:58:37 PM.

READY ▶ 🔍 📖 ⚙️