Zeppelin

```
%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'])
```

FINISHED

```
%pyspark
df
```

FINISHED

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

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

FINISHED

```
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
```

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

FINISHED

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

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

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

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

```
one    b
two    d
dtype: object
```

```
%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
```

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

```
%pyspark
obj
```

```
0     a
1     a
2     b
3     c
4     a
5     a
6     b
7     c
8     a
9     a
10    b
11    c
12    a
13    a
14    b
15    c
dtype: object
```

FINISHED ▷ ⅀⅀ 📖 ⚙

```
%pyspark
obj.describe()
```

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

FINISHED ▷ ⅀⅀ 📖 ⚙

```
%pyspark
from pandas_datareader import data as wb
all_data = {}
for ticker in ['AAPL','IBM','MSFT','GOOG','INTC','HPQ']:
    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	HPQ	IBM	INTC	MSFT
Date						
2017-02-15	0.003629	-0.001792	-0.004981	0.008605	0.003340	-0.000619
2017-02-16	-0.001181	0.006325	-0.003755	-0.001376	0.009986	-0.000155
2017-02-17	0.002734	0.004744	0.001884	-0.004189	0.001923	0.001550
2017-02-21	0.007221	0.004335	0.011285	-0.002269	0.001096	-0.002012
2017-02-22	0.002999	-0.001082	0.004340	0.004937	-0.012322	-0.002016

FINISHED ▷ ⅀⅀ 📖 ⚙

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

```
0.49515377802280924
```

```
%pyspark
returns.HPQ.corr(returns.INTC)
```

```
0.46280121957842657
```

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

```
8.5977652563835441e-05
```

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

```
0.00014178582732356362
```

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

```
          AAPL      GOOG       HPQ       IBM      INTC      MSFT
AAPL  1.000000  0.409541  0.317160  0.381549  0.388781  0.388972
GOOG  0.409541  1.000000  0.277883  0.402872  0.374907  0.470820
HPQ   0.317160  0.277883  1.000000  0.449908  0.462801  0.385346
IBM   0.381549  0.402872  0.449908  1.000000  0.503128  0.495154
INTC  0.388781  0.374907  0.462801  0.503128  1.000000  0.559970
MSFT  0.388972  0.470820  0.385346  0.495154  0.559970  1.000000
```

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

```
          AAPL      GOOG       HPQ       IBM      INTC      MSFT
AAPL  0.000270  0.000105  0.000109  0.000075  0.000094  0.000093
GOOG  0.000105  0.000244  0.000091  0.000075  0.000086  0.000107
HPQ   0.000109  0.000091  0.000435  0.000113  0.000142  0.000116
IBM   0.000075  0.000075  0.000113  0.000144  0.000089  0.000086
INTC  0.000094  0.000086  0.000142  0.000089  0.000216  0.000119
MSFT  0.000093  0.000107  0.000116  0.000086  0.000119  0.000210
```

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

```
AAPL    0.381549
GOOG    0.402872
HPQ     0.449908
IBM     1.000000
INTC    0.503128
MSFT    0.495154
dtype: float64
```

FINISHED ▷ ⠶ 📖 ⚙

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

```
AAPL    0.388781
GOOG    0.374907
HPQ     0.462801
IBM     0.503128
INTC    1.000000
MSFT    0.559970
dtype: float64
```

FINISHED ▷ ⠶ 📖 ⚙

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

```
AAPL    -0.074323
GOOG    -0.009670
HPQ     -0.123092
IBM     -0.194432
INTC    -0.030195
MSFT    -0.091017
dtype: float64
```

FINISHED ▷ ⠶ 📖 ⚙