Lab_feb23

```
%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 38 sec. Last updated by anonymous at February 23 2017, 6:48:13 PM.

```
%pyspark
print(df.sum())
print (df.sum(axis=1))
print(df.mean(axis=1,skipna=False))
print (df.idxmax())
print(df.describe())
obj = Series(['a','a','b','c'] * 4)
obj
obj.describe()
```

FINISHED

```
one    9.25
two   -5.80
dtype: float64
a    1.40
b    2.60
c     NaN
d   -0.55
dtype: float64
a      NaN
b    1.300
c      NaN
d   -0.275
dtype: float64
one    b
two    d
dtype: object
            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
count     16
unique     3
top        a
freq       8
dtype: object
```

Took 0 sec. Last updated by anonymous at February 23 2017, 6:50:33 PM. (outdated)

FINISHED

```
%pyspark
from pandas_datareader import data as  web
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()
```

Took 1 sec. Last updated by anonymous at February 23 2017, 6:52:08 PM.

FINISHED

```
%pyspark
all_data
```

```
{'GOOG':                 Open        High         Low       Close    Volume  \
Date
2010-01-04  626.951088  629.511067  624.241073  626.751061   3927000
2010-01-05  627.181073  627.841071  621.541045  623.991055   6031900
2010-01-06  625.861078  625.861078  606.361042  608.261023   7987100
2010-01-07  609.401025  610.001045  592.651008  594.101005  12876600
2010-01-08  592.000997  603.251034  589.110988  602.021036   9483900
2010-01-11  604.461060  604.461060  594.041028  601.111020  14479800
2010-01-12  597.651011  598.161034  588.001007  590.481036   9742900
2010-01-13  576.490965  588.381027  573.900966  587.090991  13041800
2010-01-14  583.900972  594.200988  582.811024  589.850997   8511900
2010-01-15  593.341025  593.561024  578.041006  580.000965  10909600
2010-01-19  581.201005  590.420997  576.290999  587.620986   8665700
2010-01-20  585.981009  585.981009  575.290986  580.411005   6525700
2010-01-21  583.441002  586.821000  572.251003  582.980970  12662600
2010-01-22  564.500980  570.600979  534.860888  550.010933  13651700
2010-01-25  546.590929  549.880931  535.510900  540.000941   8872800
2010-01-26  537.970896  549.600954  536.290913  542.420931   8743600
```

Took 0 sec. Last updated by anonymous at February 23 2017, 6:54:08 PM.

```
%pyspark
returns.tail()
```

FINISHED

```
                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 0 sec. Last updated by anonymous at February 23 2017, 6:52:21 PM.

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

FINISHED

```
0.495153778023
8.5977652563835441e-05
```

Took 1 sec. Last updated by anonymous at February 23 2017, 6:58:39 PM.

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

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
          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, 6:59:57 PM.

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

FINISHED

```
AAPL    0.381549
GOOG    0.402872
IBM     1.000000
MSFT    0.495154
dtype: float64
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, 7:07:53 PM.

```
%pyspark
```

READY