

Lab 5 022317



```
%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 ▶ ⌵ 📖 ⚙

df

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

FINISHED ▶ ⌵ 📖 ⚙

```
print(df.sum())
```

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

FINISHED ▶ ⌵ 📖 ⚙

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

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

FINISHED ▶ ⌵ 📖 ⚙

```
df.mean(axis=1)
```

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

FINISHED ▶ ⌵ 📖 ⚙

```
df.idxmax()
```

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

FINISHED ▶ ⌵ 📖 ⚙

```
df.cumsum()
```

```
   one  two
a  1.40 NaN
b  8.50 -4.5
c   NaN NaN
d  9.25 -5.8
```

FINISHED ▶ ⌵ 📖 ⚙

```
df.describe()
```

FINISHED ▶ ⌵ 📖 ⚙

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

FINISHED ▶ ⌵ 📖 ⚙

```
obj.describe()
```

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

```

one 1
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
     .200000

```

```

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

```

```

%pyspark
from pandas_datareader import data as web
all_data = {}
for ticker in ['AAPL','IBM','MSFT','GOOG','PEG']:
    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()

```

FINISHED ▷ ⌵ 📖 ⚙️

	AAPL	GOOG	IBM	MSFT	PEG
Date					
2017-02-15	0.003629	-0.001792	0.008605	-0.000619	-0.005066
2017-02-16	-0.001181	0.006325	-0.001376	-0.000155	0.010414
2017-02-17	0.002734	0.004744	-0.004189	0.001550	-0.003894
2017-02-21	0.007221	0.004335	-0.002269	-0.002012	0.019315
2017-02-22	0.002999	-0.001082	0.004937	-0.002016	-0.001354

%pyspark FINISHED ▷ ⌵ 📖 ⚙️

```

returns.MSFT.corr
(returns.IBM)
returns.MSFT.cov(returns
.IBM)

```

8.5977652563835441e-05

%pyspark FINISHED ▷ ⌵ 📖 ⚙️

```

returns.corr()

```

%pyspark FINISHED ▷ ⌵ 📖 ⚙️

```

returns.cov()

```

	AAPL	GOOG
IBM	MSFT	PEG
AAPL	1.000000	0.409541
0.381549	0.388972	0.214978
GOOG	0.409541	1.000000
0.402872	0.470820	0.253036
IBM	0.381549	0.402872
1.000000	0.495154	0.357348
MSFT	0.388972	0.470820
0.495154	1.000000	0.339156
PEG	0.214978	0.253036
0.357348	0.339156	1.000000

	AAPL	GOOG
IBM	MSFT	PEG
AAPL	0.000270	0.000105
0.000075	0.000093	0.000042
GOOG	0.000105	0.000244
0.000075	0.000107	0.000047
IBM	0.000075	0.000075
0.000144	0.000086	0.000051
MSFT	0.000093	0.000107
0.000086	0.000210	0.000058
PEG	0.000042	0.000047
0.000051	0.000058	0.000142

%pyspark

FINISHED ▷ ⌵ 📖 ⚙

returns.corrwith(returns.IBM)

```
AAPL    0.381549
GOOG    0.402872
IBM      1.000000
MSFT    0.495154
PEG      0.357348
dtype: float64
```

%pyspark

FINISHED ▷ ⌵ 📖 ⚙

returns.corrwith(volume)

```
AAPL    -0.074323
GOOG    -0.009670
IBM      -0.194432
MSFT    -0.091017
PEG      -0.029628
dtype: float64
```

READY ▷ ⌵ 📖 ⚙