Zeppelin

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

FINISHED

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

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

FINISHED

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

FINISHED

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

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

FINISHED

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

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()})
```

FINISHED

```
%pyspark
returns = price.pct_change()
returns.tail()
returns.MSFT.corr(returns.IBM)
returns.MSFT.cov(returns.IBM)
```

```
8.5977652563835441e-05
```

FINISHED

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

	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

FINISHED

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

```
AAPL   -0.074323
GOOG   -0.009665
IBM    -0.194432
MSFT   -0.091017
dtype: float64
```

READY