Assignment 5

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

FINISHED

Took 0 sec. Last updated by anonymous at February 23 2017, 6:57:59 PM.

```
%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
df.sum()
df.sum(axis=1)
df.mean(axis=1,skipna=False)
df.idxmax()
df.describe()
obj = Series(['a','a','b','c'] * 4)
obj
obj.describe()
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()
returns.tail()
returns.MSFT.corr(returns.IBM)
returns.MSFT.cov(returns.IBM)
returns.corr()
returns.cov()
returns.corrwith(returns.IBM)
returns.corrwith(volume)
```

FINISHED

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

Took 1 sec. Last updated by anonymous at February 23 2017, 7:00:27 PM.

```
%pyspark
```

READY