

feb 23 poornima

```
%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, 7:08:15 PM.

```
%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'])

df
```

FINISHED

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

Took 0 sec. Last updated by anonymous at February 23 2017, 7:08:29 PM.

```
%pyspark
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, web
```

ERROR

Traceback (most recent call last):

File "/var/folders/5n/67rpg35x1w7gxfzv53g2p64pmrz2s/T/zeppelin_pyspark-2206883558564754685.py", line 346, in <module>

raise Exception(traceback.format_exc())

Exception: Traceback (most recent call last):

File "/var/folders/5n/67rpg35x1w7gxfzv53g2p64pmrz2s/T/zeppelin_pyspark-2206883558564754685.py", line 339, in <module>

exec(code)

File "<stdin>", line 9, in <module>

ImportError: cannot import name web

Took 0 sec. Last updated by anonymous at February 23 2017, 7:08:52 PM.

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

FINISHED

Took 0 sec. Last updated by anonymous at February 23 2017, 7:09:47 PM.

```
%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

Took 1 sec. Last updated by anonymous at February 23 2017, 7:11:17 PM.

```
%pyspark
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 0 sec. Last updated by anonymous at February 23 2017, 7:13:32 PM.

```
%pyspark
```

READY