bonus_multiindex

February 29, 2024

1 Bonus video: How do I use the MultiIndex in pandas?

Full course: pandas in 30 days

[1]:
```python
import pandas as pd
```

[2]:
```python
stocks = pd.read_csv('http://bit.ly/smallstocks')
stocks
```

[2]:
```
         Date   Close    Volume Symbol
0  2016-10-03   31.50  14070500   CSCO
1  2016-10-03  112.52  21701800   AAPL
2  2016-10-03   57.42  19189500   MSFT
3  2016-10-04  113.00  29736800   AAPL
4  2016-10-04   57.24  20085900   MSFT
5  2016-10-04   31.35  18460400   CSCO
6  2016-10-05   57.64  16726400   MSFT
7  2016-10-05   31.59  11808600   CSCO
8  2016-10-05  113.05  21453100   AAPL
```

[3]:
```python
stocks.index
```

[3]:
```
RangeIndex(start=0, stop=9, step=1)
```

[4]:
```python
stocks.groupby('Symbol').Close.mean()
```

[4]:
```
Symbol
AAPL    112.856667
CSCO     31.480000
MSFT     57.433333
Name: Close, dtype: float64
```

1.1 Series with MultiIndex

[5]:
```python
ser = stocks.groupby(['Symbol', 'Date']).Close.mean()
ser
```

[5]:
```
Symbol  Date
AAPL    2016-10-03    112.52
        2016-10-04    113.00
        2016-10-05    113.05
CSCO    2016-10-03     31.50
        2016-10-04     31.35
        2016-10-05     31.59
MSFT    2016-10-03     57.42
        2016-10-04     57.24
        2016-10-05     57.64
Name: Close, dtype: float64
```

[6]:
```python
ser.index
```

[6]:
```
MultiIndex([('AAPL', '2016-10-03'),
            ('AAPL', '2016-10-04'),
            ('AAPL', '2016-10-05'),
            ('CSCO', '2016-10-03'),
            ('CSCO', '2016-10-04'),
            ('CSCO', '2016-10-05'),
            ('MSFT', '2016-10-03'),
            ('MSFT', '2016-10-04'),
            ('MSFT', '2016-10-05')],
           names=['Symbol', 'Date'])
```

[7]:
```python
ser.unstack()
```

[7]:
```
Date    2016-10-03  2016-10-04  2016-10-05
Symbol
AAPL        112.52      113.00      113.05
CSCO         31.50       31.35       31.59
MSFT         57.42       57.24       57.64
```

[8]:
```python
df = stocks.pivot_table(values='Close', index='Symbol', columns='Date')
df
```

[8]:
```
Date    2016-10-03  2016-10-04  2016-10-05
Symbol
AAPL        112.52      113.00      113.05
CSCO         31.50       31.35       31.59
MSFT         57.42       57.24       57.64
```

1.2 Selection from Series with MultiIndex

[9]:
```python
ser
```

[9]:
```
Symbol  Date
AAPL    2016-10-03    112.52
        2016-10-04    113.00
        2016-10-05    113.05
CSCO    2016-10-03     31.50
        2016-10-04     31.35
        2016-10-05     31.59
MSFT    2016-10-03     57.42
        2016-10-04     57.24
        2016-10-05     57.64
Name: Close, dtype: float64
```

[10]:
```python
ser.loc['AAPL']
```

[10]:
```
Date
2016-10-03    112.52
2016-10-04    113.00
2016-10-05    113.05
Name: Close, dtype: float64
```

[11]:
```python
ser.loc['AAPL', '2016-10-03']
```

[11]:
```
112.52
```

[12]:
```python
ser.loc[:, '2016-10-03']
```

[12]:
```
Symbol
AAPL    112.52
CSCO     31.50
MSFT     57.42
Name: Close, dtype: float64
```

[13]:
```python
df
```

[13]:
```
Date    2016-10-03  2016-10-04  2016-10-05
Symbol
AAPL        112.52      113.00      113.05
CSCO         31.50       31.35       31.59
MSFT         57.42       57.24       57.64
```

[14]:
```python
df.loc['AAPL']
```

[14]:
```
Date
2016-10-03    112.52
```

```
2016-10-04    113.00
2016-10-05    113.05
Name: AAPL, dtype: float64
```

[15]:
```
df.loc['AAPL', '2016-10-03']
```

[15]:
```
112.52
```

[16]:
```
df.loc[:, '2016-10-03']
```

[16]:
```
Symbol
AAPL    112.52
CSCO     31.50
MSFT     57.42
Name: 2016-10-03, dtype: float64
```

1.3 DataFrame with MultiIndex

[17]:
```
stocks.set_index(['Symbol', 'Date'], inplace=True)
stocks
```

[17]:
```
                    Close    Volume
Symbol Date
CSCO   2016-10-03   31.50  14070500
AAPL   2016-10-03  112.52  21701800
MSFT   2016-10-03   57.42  19189500
AAPL   2016-10-04  113.00  29736800
MSFT   2016-10-04   57.24  20085900
CSCO   2016-10-04   31.35  18460400
MSFT   2016-10-05   57.64  16726400
CSCO   2016-10-05   31.59  11808600
AAPL   2016-10-05  113.05  21453100
```

[18]:
```
stocks.index
```

[18]:
```
MultiIndex([('CSCO', '2016-10-03'),
            ('AAPL', '2016-10-03'),
            ('MSFT', '2016-10-03'),
            ('AAPL', '2016-10-04'),
            ('MSFT', '2016-10-04'),
            ('CSCO', '2016-10-04'),
            ('MSFT', '2016-10-05'),
            ('CSCO', '2016-10-05'),
            ('AAPL', '2016-10-05')],
           names=['Symbol', 'Date'])
```

[19]:
```
stocks.sort_index(inplace=True)
stocks
```

[19]:
```
                    Close    Volume
Symbol Date
AAPL   2016-10-03  112.52  21701800
       2016-10-04  113.00  29736800
       2016-10-05  113.05  21453100
CSCO   2016-10-03   31.50  14070500
       2016-10-04   31.35  18460400
       2016-10-05   31.59  11808600
MSFT   2016-10-03   57.42  19189500
       2016-10-04   57.24  20085900
       2016-10-05   57.64  16726400
```

1.4 Selection from DataFrame with MultiIndex

[20]:
```
stocks.loc['AAPL']
```

[20]:
```
             Close    Volume
Date
2016-10-03  112.52  21701800
2016-10-04  113.00  29736800
2016-10-05  113.05  21453100
```

[21]:
```
stocks.loc[('AAPL', '2016-10-03'), :]
```

[21]:
```
Close          112.52
Volume    21701800.00
Name: (AAPL, 2016-10-03), dtype: float64
```

[22]:
```
stocks.loc[('AAPL', '2016-10-03'), 'Close']
```

[22]:
```
112.52
```

[23]:
```
stocks.loc[['AAPL', 'MSFT'], :]
```

[23]:
```
                    Close    Volume
Symbol Date
AAPL   2016-10-03  112.52  21701800
       2016-10-04  113.00  29736800
       2016-10-05  113.05  21453100
MSFT   2016-10-03   57.42  19189500
       2016-10-04   57.24  20085900
       2016-10-05   57.64  16726400
```

[24]:
```
stocks.loc[(['AAPL', 'MSFT'], '2016-10-03'), :]
```

[24]:
```
                    Close    Volume
Symbol Date
AAPL   2016-10-03  112.52  21701800
```

```
MSFT   2016-10-03   57.42  19189500
```

[25]:
```
stocks.loc[(['AAPL', 'MSFT'], '2016-10-03'), 'Close']
```

[25]:
```
Symbol  Date
AAPL    2016-10-03    112.52
MSFT    2016-10-03     57.42
Name: Close, dtype: float64
```

[26]:
```
stocks.loc[('AAPL', ['2016-10-03', '2016-10-04']), 'Close']
```

[26]:
```
Symbol  Date
AAPL    2016-10-03    112.52
        2016-10-04    113.00
Name: Close, dtype: float64
```

[27]:
```
stocks.loc[(slice(None), ['2016-10-03', '2016-10-04']), :]
```

[27]:
```
                    Close    Volume
Symbol Date
AAPL   2016-10-03  112.52  21701800
       2016-10-04  113.00  29736800
CSCO   2016-10-03   31.50  14070500
       2016-10-04   31.35  18460400
MSFT   2016-10-03   57.42  19189500
       2016-10-04   57.24  20085900
```

1.5 Merging DataFrames with MultiIndexes

[28]:
```
close = pd.read_csv('http://bit.ly/smallstocks', usecols=[0, 1, 3],␣
 ↪index_col=['Symbol', 'Date']).sort_index()
close
```

[28]:
```
                    Close
Symbol Date
AAPL   2016-10-03  112.52
       2016-10-04  113.00
       2016-10-05  113.05
CSCO   2016-10-03   31.50
       2016-10-04   31.35
       2016-10-05   31.59
MSFT   2016-10-03   57.42
       2016-10-04   57.24
       2016-10-05   57.64
```

[29]:
```
volume = pd.read_csv('http://bit.ly/smallstocks', usecols=[0, 2, 3],␣
 ↪index_col=['Symbol', 'Date']).sort_index()
```

```
volume
```

[29]:

```
                   Volume
Symbol Date
AAPL   2016-10-03  21701800
       2016-10-04  29736800
       2016-10-05  21453100
CSCO   2016-10-03  14070500
       2016-10-04  18460400
       2016-10-05  11808600
MSFT   2016-10-03  19189500
       2016-10-04  20085900
       2016-10-05  16726400
```

[30]:

```
both = pd.merge(close, volume, left_index=True, right_index=True)
both
```

[30]:

```
                    Close    Volume
Symbol Date
AAPL   2016-10-03  112.52  21701800
       2016-10-04  113.00  29736800
       2016-10-05  113.05  21453100
CSCO   2016-10-03   31.50  14070500
       2016-10-04   31.35  18460400
       2016-10-05   31.59  11808600
MSFT   2016-10-03   57.42  19189500
       2016-10-04   57.24  20085900
       2016-10-05   57.64  16726400
```

[31]:

```
both.reset_index()
```

[31]:

```
  Symbol        Date   Close    Volume
0   AAPL  2016-10-03  112.52  21701800
1   AAPL  2016-10-04  113.00  29736800
2   AAPL  2016-10-05  113.05  21453100
3   CSCO  2016-10-03   31.50  14070500
4   CSCO  2016-10-04   31.35  18460400
5   CSCO  2016-10-05   31.59  11808600
6   MSFT  2016-10-03   57.42  19189500
7   MSFT  2016-10-04   57.24  20085900
8   MSFT  2016-10-05   57.64  16726400
```