3.1 - MultiIndex

April 11, 2017

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

In [2]: # /usr/lib/python3/dist-packages/pandas/tools/tests/data/quotes2.csv
        df = pd.read_csv('quotes2.csv')

In [3]: df.dtypes
```

```
time       object
ticker     object
bid       float64
ask       float64
dtype: object
```

```
In [4]: df.drop_duplicates(inplace=True)

In [5]: idf = df.set_index('ticker')

In [6]: idf = df.set_index(['ticker', 'time'])

In [7]: idf.loc['AAPL']
```

```
                         bid    ask
time
20160525 13:30:00.075  98.55  98.56
20160525 13:30:00.076  98.55  98.56
20160525 13:30:00.080  98.55  98.56
20160525 13:30:00.084  98.55  98.56
20160525 13:30:00.086  98.55  98.63
20160525 13:30:00.088  98.65  98.63
20160525 13:30:00.089  98.63  98.63
20160525 13:30:00.104  98.63  98.63
20160525 13:30:00.104  98.62  98.63
20160525 13:30:00.105  98.62  98.63
20160525 13:30:00.107  98.62  98.63
20160525 13:30:00.115  98.62  98.63
20160525 13:30:00.118  98.62  98.63
```

```
20160525 13:30:00.128  98.62  98.63
20160525 13:30:00.129  98.62  98.63
20160525 13:30:00.129  98.61  98.63
20160525 13:30:00.130  98.61  98.63
20160525 13:30:00.131  98.61  98.62
20160525 13:30:00.135  98.61  98.62
20160525 13:30:00.136  98.61  98.62
20160525 13:30:00.144  98.61  98.62
20160525 13:30:00.145  98.61  98.62
20160525 13:30:00.145  98.61  98.63
20160525 13:30:00.145  98.60  98.63
```

```
In [8]: idf.loc['AAPL']
```

```
                         bid    ask
time
20160525 13:30:00.075  98.55  98.56
20160525 13:30:00.076  98.55  98.56
20160525 13:30:00.080  98.55  98.56
20160525 13:30:00.084  98.55  98.56
20160525 13:30:00.086  98.55  98.63
20160525 13:30:00.088  98.65  98.63
20160525 13:30:00.089  98.63  98.63
20160525 13:30:00.104  98.63  98.63
20160525 13:30:00.104  98.62  98.63
20160525 13:30:00.105  98.62  98.63
20160525 13:30:00.107  98.62  98.63
20160525 13:30:00.115  98.62  98.63
20160525 13:30:00.118  98.62  98.63
20160525 13:30:00.128  98.62  98.63
20160525 13:30:00.129  98.62  98.63
20160525 13:30:00.129  98.61  98.63
20160525 13:30:00.130  98.61  98.63
20160525 13:30:00.131  98.61  98.62
20160525 13:30:00.135  98.61  98.62
20160525 13:30:00.136  98.61  98.62
20160525 13:30:00.144  98.61  98.62
20160525 13:30:00.145  98.61  98.62
20160525 13:30:00.145  98.61  98.63
20160525 13:30:00.145  98.60  98.63
```

```
In [9]: idf.sort_index(inplace=True)

In [10]: # Prima selezione parziale: è ambigua! (sarebbe idf.loc[('AAPL', slice(Non
         idf.loc['AAPL', 'bid']
```

```
time
20160525 13:30:00.075    98.55
20160525 13:30:00.076    98.55
20160525 13:30:00.080    98.55
20160525 13:30:00.084    98.55
20160525 13:30:00.086    98.55
20160525 13:30:00.088    98.65
20160525 13:30:00.089    98.63
20160525 13:30:00.104    98.63
20160525 13:30:00.104    98.62
20160525 13:30:00.105    98.62
20160525 13:30:00.107    98.62
20160525 13:30:00.115    98.62
20160525 13:30:00.118    98.62
20160525 13:30:00.128    98.62
20160525 13:30:00.129    98.62
20160525 13:30:00.129    98.61
20160525 13:30:00.130    98.61
20160525 13:30:00.131    98.61
20160525 13:30:00.135    98.61
20160525 13:30:00.136    98.61
20160525 13:30:00.144    98.61
20160525 13:30:00.145    98.61
20160525 13:30:00.145    98.61
20160525 13:30:00.145    98.60
Name: bid, dtype: float64
```

```
In [11]: # Cosa fa pandas (ma non ambigua)
         idf.loc[('AAPL',slice(None)), 'bid']
```

```
ticker  time
AAPL    20160525 13:30:00.075    98.55
        20160525 13:30:00.076    98.55
        20160525 13:30:00.080    98.55
        20160525 13:30:00.084    98.55
        20160525 13:30:00.086    98.55
        20160525 13:30:00.088    98.65
        20160525 13:30:00.089    98.63
        20160525 13:30:00.104    98.63
        20160525 13:30:00.104    98.62
        20160525 13:30:00.105    98.62
        20160525 13:30:00.107    98.62
        20160525 13:30:00.115    98.62
        20160525 13:30:00.118    98.62
        20160525 13:30:00.128    98.62
```

```
         20160525 13:30:00.129    98.62
         20160525 13:30:00.129    98.61
         20160525 13:30:00.130    98.61
         20160525 13:30:00.131    98.61
         20160525 13:30:00.135    98.61
         20160525 13:30:00.136    98.61
         20160525 13:30:00.144    98.61
         20160525 13:30:00.145    98.61
         20160525 13:30:00.145    98.61
         20160525 13:30:00.145    98.60
Name: bid, dtype: float64
```

```
In [12]: idf = idf[~idf.index.duplicated()]
```

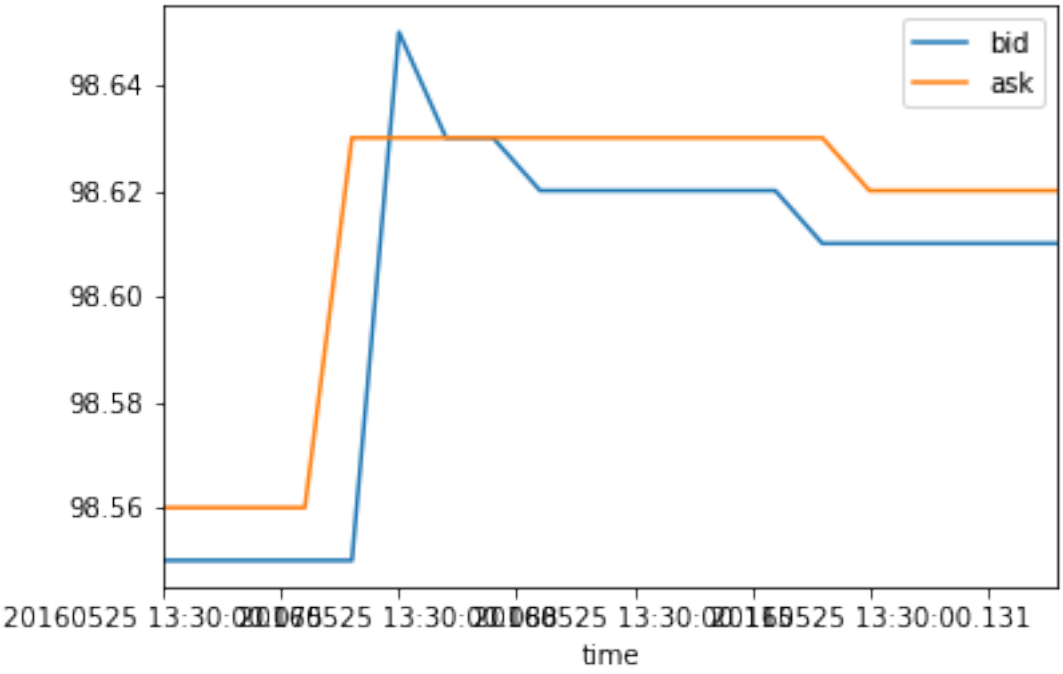
0.1 Manipolazione

```
In [13]: idf.head()
```

```
                                  bid    ask
ticker time
AAPL   20160525 13:30:00.075  98.55  98.56
       20160525 13:30:00.076  98.55  98.56
       20160525 13:30:00.080  98.55  98.56
       20160525 13:30:00.084  98.55  98.56
       20160525 13:30:00.086  98.55  98.63
```

```
In [14]: %matplotlib inline
         idf.loc['AAPL'].plot()

Out[14]: <matplotlib.axes._subplots.AxesSubplot at 0x7fbb4dd110b8>
```


bid
ask
98.64
98.62
98.60
98.58
98.56
20160525 13:30:00.131
time