# **pandas 将“字符类型的日期列”转化成“时间戳索引（DatetimeIndex）”**

[](http://www.jianshu.com/u/c72d7eac4141)

作者 [Tim\_Lee](http://www.jianshu.com/u/c72d7eac4141) 关注

2017.07.04 23:24\* 字数 383 阅读 1118评论 4喜欢 1

假设目前已经引入了 pandas，同时也拥有 pandas 的 DataFrame 类型数据。

import pandas as pd

数据集如下

df.head(3)

date open close high low volume code0 2006-12-18 3.905 3.886 3.943 3.867 171180.67 6000011 2006-12-19 3.886 3.924 3.981 3.867 276799.39 6000012 2006-12-20 3.934 3.934 3.962 3.809 265653.85 600001

查看每一列的类型

df.info()

从结果的第四排可以看见 date 这一列类型是"object"，即字符类型。

<class 'pandas.core.frame.DataFrame'>RangeIndex: 640 entries, 0 to 639

Data columns (total 7 columns):

date 640 non-null object

open 640 non-null float64

close 640 non-null float64

high 640 non-null float64

low 640 non-null float64

volume 640 non-null float64

code 640 non-null object

dtypes: float64(5), object(2)

memory usage: 35.1+ KB

现在的目标是：

* 把 date 这一列用作索引
* 把 date 用作索引时，类型需要是 DatetimeIndex。

# **方法1： .to\_datetime 和 .set\_index**

首先，利用 pandas 的to\_datetime 方法，把 "date" 列的字符类型数据解析成 datetime 对象。

然后，把 "date" 列用作索引。

df['date'] = pd.to\_datetime(df['date'])

df.set\_index("date", inplace=True)

结果：

df.head(3)

open close high low volume code

date 2006-12-18 3.905 3.886 3.943 3.867 171180.67 6000012006-12-19 3.886 3.924 3.981 3.867 276799.39 6000012006-12-20 3.934 3.934 3.962 3.809 265653.85 600001

查看索引是否成为 DatetimeIndex 类型，可以看见确实已经成功转化类型。

df.axes

[DatetimeIndex(['2006-12-18', '2006-12-19', '2006-12-20', '2006-12-21',

'2006-12-22', '2006-12-25', '2006-12-26', '2006-12-27',

'2006-12-28', '2006-12-29',

...

'2009-12-02', '2009-12-03', '2009-12-04', '2009-12-07',

'2009-12-08', '2009-12-09', '2009-12-10', '2009-12-11',

'2009-12-14', '2009-12-15'],

dtype='datetime64[ns]', name='date', length=640, freq=None),

Index(['open', 'close', 'high', 'low', 'volume', 'code'], dtype='object')]

# **方法2： .DatetimeIndex**

首先是原始数据。

df2.head(3)

date open close high low volume code0 2003-08-01 4.997 4.949 5.016 4.949 20709.15 6000021 2003-08-04 4.949 5.045 5.054 4.949 23923.35 6000022 2003-08-05 5.054 5.093 5.131 5.006 35224.00 600002

先把 "date" 列用作索引，然后使用 DatetimeIndex 将字符类型转化成 DateIndex。

df2.set\_index("date", inplace=True)

这个时候索引还是 object 类型，就是字符串类型。

df2.axes

[Index(['2003-08-01', '2003-08-04', '2003-08-05', '2003-08-06', '2003-08-07',

'2003-08-08', '2003-08-11', '2003-08-12', '2003-08-13', '2003-08-14',

...

'2006-03-24', '2006-03-27', '2006-03-28', '2006-03-29', '2006-03-30',

'2006-03-31', '2006-04-03', '2006-04-04', '2006-04-05', '2006-04-06'],

dtype='object', name='date', length=640),

Index(['open', 'close', 'high', 'low', 'volume', 'code'], dtype='object')]

将其转化成 DateIndex 类型。

df2.index = pd.DatetimeIndex(df.index)

再次查看结果

df2.axes

转化成功

[DatetimeIndex(['2006-12-18', '2006-12-19', '2006-12-20', '2006-12-21',

'2006-12-22', '2006-12-25', '2006-12-26', '2006-12-27',

'2006-12-28', '2006-12-29',

...

'2009-12-02', '2009-12-03', '2009-12-04', '2009-12-07',

'2009-12-08', '2009-12-09', '2009-12-10', '2009-12-11',

'2009-12-14', '2009-12-15'],

dtype='datetime64[ns]', name='date', length=640, freq=None),

Index(['open', 'close', 'high', 'low', 'volume', 'code'], dtype='object')]

## **结论：.to\_datetime仅转换格式，.DatetimeIndex还能设置为索引**

两者在转化格式的功能上效果一样，都可以把字符串对象转换成 datetime 对象。

pd.DatetimeIndex 是把某一列进行转换，同时把该列的数据设置为索引 index。  
比如

df2.index = pd.DatetimeIndex(df2["date"])

得到一个以 date 作为索引的结果。

.DatetimeIndex 的问题是原来的 date 列数据仍然存在，形成了重复。

date open close high low volume code

date 2003-08-01 2003-08-01 4.997 4.949 5.016 4.949 20709.15 6000022003-08-04 2003-08-04 4.949 5.045 5.054 4.949 23923.35 6000022003-08-05 2003-08-05 5.054 5.093 5.131 5.006 35224.00 600002

最终还需要把 date 这一列删掉。

del df2["date"]

才能得到正常数据

open close high low volume code

date 2003-08-01 4.997 4.949 5.016 4.949 20709.15 6000022003-08-04 4.949 5.045 5.054 4.949 23923.35 6000022003-08-05 5.054 5.093 5.131 5.006 35224.00 600002