

请根据以下说明，完成之后的练习。

A，一共四道题目，请提供 py 后缀的 python 源代码，如有必要，可以写一个 report。

B，请参考 numpy 和 pandas 相应文档，选取适当的方法完成任务。

C，按照代码完成度（20%），代码正确率（40%），是否熟练使用数据分析相关库（30%），代码风格（10%）来进行考核。

E1 你可以看到一些 csv 文件，每个 csv 文件是股票的日频数据，以 s\_info\_windcode 命名，

注意：a，以下所说的股票代码是指股票在交易所的代码，即 s\_info\_windcode 的前 6 位。  
SZ 表示深圳，SH 表示上海。

b，以下用 symbol 代表股票代码，symbols 表示股票代码列表。

c，以下用 date 表示 trade\_dt，turnover 代表 s\_dq\_amout，其余对应去掉 s\_dq。

#### 4.28 中国 A 股日行情-AShareEODPrices

字段	字段类型	字段名称	备注
OBJECT_ID	VARCHAR2(100)	对象 ID	
s_info_windcode	VARCHAR2(40)	Wind 代码	
trade_dt	VARCHAR2(8)	交易日期	
crncy_code	VARCHAR2(10)	货币代码	
s_dq_preclose	NUMBER(20,4)	昨收盘价(元)	数据保留 2 位小数
s_dq_open	NUMBER(20,4)	开盘价(元)	数据保留 2 位小数
s_dq_high	NUMBER(20,4)	最高价(元)	数据保留 2 位小数
s_dq_low	NUMBER(20,4)	最低价(元)	数据保留 2 位小数
s_dq_close	NUMBER(20,4)	收盘价(元)	数据保留 2 位小数
s_dq_change	NUMBER(20,4)	涨跌(元)	数据保留 2 位小数

s_dq_pctchange	NUMBER(20,4)	涨跌幅(%)	
s_dq_volume	NUMBER(20,4)	成交量(手)	
s_dq_amount	NUMBER(20,4)	成交金额(千元)	
s_dq_adjpreclose	NUMBER(20,4)	复权昨收盘价(元)	昨收盘价*复权因子
s_dq_adjopen	NUMBER(20,4)	复权开盘价(元)	开盘价*复权因子
s_dq_adjhigh	NUMBER(20,4)	复权最高价(元)	最高价*复权因子
s_dq_adjlow	NUMBER(20,4)	复权最低价(元)	最低价*复权因子
s_dq_adjclose	NUMBER(20,4)	复权收盘价(元)	收盘价*复权因子
s_dq_adjfactor	NUMBER(20,6)	复权因子	初始值为 1；当日复权因子=前一交易日收盘价/当日昨收盘价*前一交易日复权因子。
S_dq_avgprice	NUMBER(20,4)	均价(VWAP)	成交金额/成交量
s_dq_tradestatus	VARCHAR2(10)	交易状态	

说明：本表的价格字段，都是保留 2 位小数。

设计 StockData 类，

E1.1 StockData.\_\_init\_\_(path)

初始化函数，path 为 csv 数据文件所在路径

E1.2 StockData.read(symbols)

symbols 是股票代码列表，包括 1~n 个股票代码，函数从对应 symbol 的文件中读取数据并选取合适结构保存在类中

E1.3 StockData.get\_data\_by\_symbol(symbol, start\_date, end\_date)

获取某个 symbol 从 start\_date 到 end\_date 之间的所有日频数据

E1.4 StockData.get\_data\_by\_date(adata, symbols):

获取某一天中，对应 symbols 的所有日频数据

E1.5 StockData.get\_data\_by\_field(field, symbols):

获取 symbols 在某个 field 上所有的所有交易日的数据

其中 E1.3, E1.4, E1.5 返回 pandas.DataFrame 结构，E1.6 返回 pandas.Series 结构

E1.3 dataframe 格式如下：

date	open	high	low	close
20170320				
20170321				
20170322				
20170323				
20170324				

E1.4 dataframe 格式如下：

symbols	open	high	low	close
600000				
600036				
000001				
000002				
000004				

E1.5 dataframe 格式如下：

date	600000	600036	000001	000004
20170320				
20170321				
20170322				
20170323				
20170324				

注：停牌或者缺失数据可以不处理，日期范围最大选取近十年的。

E2, csv 数据均为日频数据, 你需要做一些采样和转换

E2.1 StockData.format\_date(symbol)

symbol 的数据中 date 在读入时, 可能是 int 或者 string 类型, 请将这一列中的每个 date

转化为 datetime 类型 (或者 pandas.Timestamp 类型)

E2.2 StockData.plot(symbol, field)

做出 symbol 关于 field 的走势图, field 取 open, high, low, close, vwap 中的某一个。

走势图为折线图, 取 volume 或者 turnover 中的某一个时, 走势图为柱状图。

简单画一下就行, 类似于一般股票网站上价量的图。

E2.3 StockData.adjust\_data(symbol)

对代表价格 (open, high, low, close) 进行 (从后向前) 复权。计算方法如下:

## ■ Wind 复权算法

基于“分红再投资”的假设, 对价格进行调整, 弥补因权益事件造成的除权或除息缺口, 方便投资收益率的计算。

### ● 算法:

- 1、上市首日的复权因子 AF 为 1
- 2、 $T$  日复权因子  $AF = T-1$  日收盘价 /  $T$  日昨收盘价  $\times T-1$  日 AF
- 3、 $T$  日复权价(后复权) =  $T$  日收盘价  $\times T$  日 AF
- 4、 $T$  日复权价(前复权) =  $T$  日收盘价  $\times T$  日 AF / 最新交易日 AF

对于每一天的价格均要复权。

E2.4 StockData.resample(symbol, freq)

将 symbol 的日频数据进行 resample (重采样), freq 表示天数, 采样后的时间为区间的左端时间

采样只需计算完成 open, high, low, close, volume, turnover, vwap 这些 field 即可

例如: StockData data

df = data.resample("600000", 5)

表示对 600000 的日频数据进行 5 天的采样，其中  
open, close 表示这 5 天第一日开盘价和最后一日的收盘价  
high, low 表示这 5 天的最高价和最低价  
volume, turnover 表示这 5 天的总成交量，总成交额  
vwap 表示这 5 天的成交量加权平均价格

**注意：**

a, E2.1 和 E2.3 的函数直接改变了 StockData 存储的数据。对于每个 symbol 而言，就是改变了存储这个 symbol 日频数据的 dataframe 中相应 field 的值。

B, E2.4 返回 dataframe

E3, 你需要根据 3.10 的 excel 中的说明，做一些计算，所用到的数据，是在 E2.3 中前复权过的数据，缺失或者停牌数据不处理。

E3.1 StockData.moving\_average(symbol, field, window)

计算 symbol 对应 field 的移动平均， window 表示滑动窗口。

E3.2 StockData.ema(symbol, params)

计算 symbol 日频数据的 ema 指标

E3.3 StockData.atr(symbol, params)

计算 symbol 日频数据的 atr 指标

E3.4 StockData.rsi(symbol, params)

计算 symbol 日频数据的 rsi 指标

E3.5 StockData.macd(symbol, params)

计算 symbol 日频数据的 macd 指标

**注意：**

a E3.1 ~ E3.5 均返回一个 pandas.Series 对象，index 是交易日 (date)

b E3.2 ~ E3.5 中的 params 表示计算用的参数，有几个参数，分别表示什么意思，自己定义。这里只是一个参数说明。

c E3.1 ~ E3.5 其中涉及到滑动窗口（或叫频率，周期）的参数，请分别测试当窗口为 5，20，60 的时候的计算结果，可以用 E2.2 中的 plot 函数画图看下结果。

d 指标计算方法在 excel 中如果没有或者和你知道不一致，请网上自行查找，或者按照自己的想法解答，在注释中写清楚计算方法即可。

e 很多指标由于涉及滑动窗口，需要做一些 rolling 的计算，请参考 pandas 相关文档。

E4, 根据 E2.3 前复权后收盘价 (close)，做如下计算：

a, symbol 表示股票代码，freq 表示频率

b, freq 参数分别取字母 m, q, h, y 对应每月，每季度，每半年，每年

E4.1 `StockData.calc_return(symbol, freq)`

计算股票每月，每季度，每半年，每年的回报。

E4.2 `StockData.calc_sharpe_ratio(symbol, freq)`

计算股票每月，每季度，每半年，每年的回报。

E4.3 `StockData.calc_max_drawdown_ratio(symbol):`

计算股票 close 的最大回撤率