

A large concrete dam with a walkway and a person standing on it.

【テーマ】

- ・ DL株価データの前処理と描画
 - ・ フリートーク：コードは売れるのか？
-

第7回

BAU STUDY SESSION

DL 株価データの前処理と描画

- ① 株価データをダウンロード
- ② ダウンロードした株価データをDBにインポート
- ③ DBから株価データを取り出しグラフ表示



1. 株価データをダウンロード

1. 株価データをダウンロード

「株価(時系列)取得」

http://wakash.com/cgi-bin/get_kabuka/select_meigara.php

安全ではありません — wakash.com

[Yahoo!ファイナンス](#) から時系列データ（デیلیー）をダウンロードして来て TAB区切りテキストにします。
Yahoo!ファイナンスでは 1983年以降のデータが得られるようです（2011年6月現在）。
対象期間を長くすると、取得までに数分かかることがあります。
このサイトから あなたのパソコンへダウンロードするのは zip形式の圧縮ファイルです。
圧縮ファイルには 1行1日のフォーマットのTAB区切りテキストファイルが1個だけ入っています。
使用例：テキストファイルを Excelへ貼り付ける。

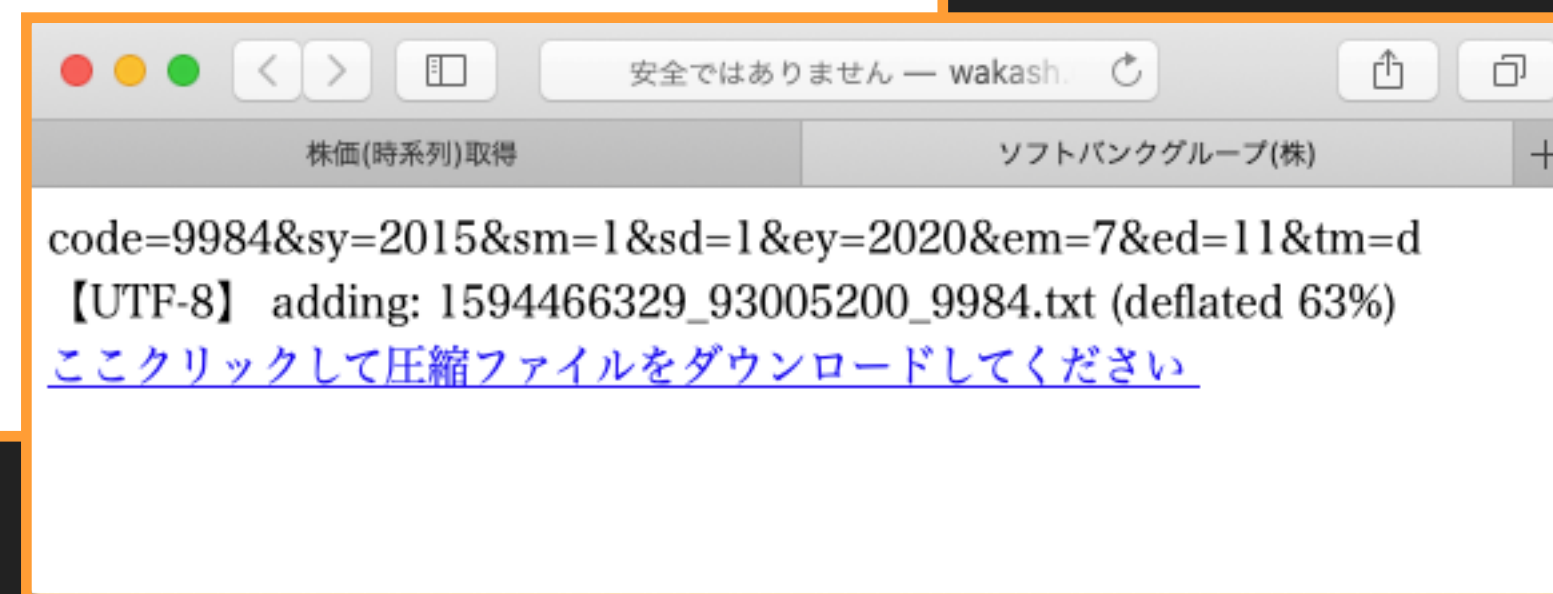
銘柄コード （日経平均:998407 TOPIX:998405）

年 月 日

から
 までのデータ

出力テキストの漢字コード：
☐ Shift-Jis
☐ EUC
☒ UTF-8

- ① 銘柄コード、取得期間、文字コードを入力して [実行] をクリック
- ② 別タブが開いてしばらくすると圧縮ファイルのリンクが表示されるのでクリックする
- ③ 株価データがZIP形式でダウンロードされる



2. ダウンロードした株価データを DBにインポート

2. ダウンロードした株価データをDBにインポート

1. ZIPファイルを解凍

```
import zipfile
with zipfile.ZipFile(zippath) as zf:
    zf.extractall()
```

2. 前処理：数値で始まる行のみを抽出

```
with open( codetsv, mode='w') as wf:
    with open(txtpath) as rf:
        for rline in rf:
            if re.search('^[0-9]', rline):
                wf.write(rline)
```

ソフトバンクグループ(株)

日付 始値 高値 安値 終値 出来高調整後終値*

2020/07/10	6380	6478	6267	6267	30112500	6267
2020/07/09	6191	6380	6184	6291	13366200	06291
2020/07/08	6135	6153	5996	6019	28817500	06019
2020/07/07	6000	6214	5992	6190	36006400	06190
2020/07/06	5839	5926	5805	5918	23873200	05918
2020/07/03	5697	5778	5675	5778	23892400	05778
:						



2020/07/10	6380	6478	6267	6267	30112500	6267
2020/07/09	6191	6380	6184	6291	13366200	06291
2020/07/08	6135	6153	5996	6019	28817500	06019
2020/07/07	6000	6214	5992	6190	36006400	06190
2020/07/06	5839	5926	5805	5918	23873200	05918
2020/07/03	5697	5778	5675	5778	23892400	05778
:						

2. ダウンロードした株価データをDBにインポート

3. 株価ファイルをデータフレームに読み込む

```
df_tsv = pd.read_table(codetsv, index_col=0,  
                      names=('Date', 'Open', 'High', 'Low', 'Close', 'Volume', 'AdjClose'))
```

4. DB接続、インポート

```
conn = sqlite3.connect(dbname)  
df_tsv.to_sql('T' + code, conn, if_exists = 'replace')  
conn.close()
```

3. D B から株価データを取り出し グラフ表示

3. DBから株価データを取り出しグラフ表示

1. DBから株価データを取得

```
sql = F"""
select Date, Open, High, Low, Close, Volume, AdjClose
from {code}
where Date between '{start}' and '{end}'
"""

conn = sqlite3.connect(dbname)
df = pd.read_sql(sql, conn)
conn.close()
```

2. 日付型インデックスの作成

```
df['Dateldx'] = pd.to_datetime(df['Date'], format='%Y/%m/%d')
df = df.set_index('Dateldx')
```

※ 以降の要点は第5回BSSと同様

おまけ. 前処理を手作業で行う

1. grepコマンドで数値で始まる行のみを抽出

```
$ grep "^[0-9]" in-filename.txt > out-filename.txt
```

2. sedコマンドで数値で始まる行のみを抽出

```
$ sed -n "/^[0-9]/p" in-filename.txt > out-filename.txt
```

3. vi でファイルを開いて編集

```
$ vi in-filename.txt
```

(ファイルの内容を表示)

```
:v/^[0-9]/d
```

... 行頭が0-9で始まらない行を削除

```
:wq out-filename.txt
```

... ファイル名を指定して保存して vi を終了

コードは売れるのか？

- ▶ 株価分析プログラムの調査中、株価取得の解説を一通りしながらも最後の部分は有料（500円）というサイトがあった。<https://note.com/tkmngn/n/nf5a0d548b909>
- ▶ その最後の売り文句...
最後のデータ整形・結合部分を実装するための時間と手間を省きたい、という方に限りご購入をおすすめします。仮に読者の労働時給が¥3,000相当だとすると、実装に10分以上かけられるよりは購入いただいた方がお得です。
- ▶ 結局は購入に至らなかったものの、売り文句にはすごく共感したし、ワンコインなら買ってもいいと思った。
- ▶ 作ろうと思えば作れる自分が買ってもいいと思うのだから、作れなかったり、そこまで勉強したくない人がそれを欲しているなら買うだろう。そこに商売の本質が見えた気がした。

コードは売れるのか？

- ▶ Pythonはスクリプト言語なのでコードを隠匿できないし、ネットには技術情報がありふれているが、そのほとんどは我々開発経験者向けの内容で、何も知らない人はそれを見てもどうしたらいいのかわからないと思われる。その溝を埋めることが出来るならコードを売ることもできるのではないか？
- ▶ 商売をしようと思うと、その取り扱う商品についてばかり目が行きがちだが、ペルソナ（サービス・商品の典型的なユーザー像）を具体的・詳細に設定して、どのようにアプローチ出来るかが重要。
- ▶ つまり、商品そのものも大事だが、売り方も同等またはそれ以上に大事。
- ▶ 実際にコードを販売しているサイトが複数散見されて、今でも実際にコードを販売していることがわかったが、パッと見たところ開発者向けの内容であり、今回挙げたようなプログラムを知らないが株価分析だけをしたい人には届かないと思われる。

コードは売れるのか？

- ▶ 環境を作り、コードを組む時間をコストに換算して、それが購入額以上のリターンを得ることが出来るならその人たちは顧客予備軍。
- ▶ コードを組む金銭的成本は人それぞれ。制作時間も違うし、時間的価値も違う。
- ▶ 10人に売れるなら1/10の金額で提供可能。
- ▶ リターンは金銭とは限らない。精神的満足度がコストを上回るならそれも商品になる。
- ▶ 正確には上回るかどうかではなく上回りそうかどうか。確信に近づけば購入に近づく。

コードは売れるのか？

- ▶ この話はしばらく寝かせておいて後日また議論したい。
- ▶ 有志が集まるのなら勉強も兼ねて取り組んでみる。
- ▶ 実際に動くことで見えてくるものもあるはず。
- ▶ 次に動くときはやらなければならないことのピックアップから。

次回のBSS

- ▶ 日程 : 2020年7月15日(水)
- ▶ 司会者 : 有川
- ▶ テーマ : 検討中