

Προγραμματιστικά Εργαλεία
Και Τεχνολογίες για Επιστήμη Δεδομένων

Time Travel



Τρυφωνόπουλος Δημήτριος

ΕΔΕΜ

00370034

2020-2021

Στην εργασία αυτή ασχοληθήκαμε με την μελέτη και αγοραπωλησία 'αμερικάνικων χρηματιστηριακών μετοχών υπό την φανταστική συνθήκη δυνατότητας κινήσεων σε παρελθοντικό χρόνο. Σκοπός ήταν η επιλογή κινήσεων που θα οδηγούσαν σε μέγιστο κέρδος ανάλογα με την στρατηγική που ακολουθήσαμε.

Δεδομένα

Τα δεδομένα αποτελούταν από '14887665' γραμμές οι οποίες περιείχαν τις παρακάτω πληροφορίες:

- Date
- Stock Open-Price: τιμή ανοίγματος της μετοχής για την συγκεκριμένη ημέρα
- Stock High-Price: υψηλότερη τιμή της μετοχής κατά την διάρκεια της ημέρας
- Stock Low-Price: χαμηλότερη τιμή της μετοχής κατά την διάρκεια της ημέρας
- Stock Close-Price: τιμή κλεισίματος της μετοχής για την συγκεκριμένη ημέρα
- Volume:
- Stock Name

Μετά τη φόρτωση των δεδομένων σε μορφή data frame, είχαμε την παρακάτω μορφή

| | Date | Open | High | Low | Close | Volume | OpenInt | Name |
|----------|------------|---------|---------|---------|---------|---------|---------|-------|
| 5675434 | 1962-01-02 | 0.6277 | 0.6362 | 0.6201 | 0.6201 | 2575579 | 0 | GE |
| 6791941 | 1962-01-02 | 6.4130 | 6.4130 | 6.3378 | 6.3378 | 467056 | 0 | IBM |
| 6791942 | 1962-01-03 | 6.3378 | 6.3963 | 6.3378 | 6.3963 | 350294 | 0 | IBM |
| 5675435 | 1962-01-03 | 0.6201 | 0.6201 | 0.6122 | 0.6201 | 1764749 | 0 | GE |
| 5675436 | 1962-01-04 | 0.6201 | 0.6201 | 0.6037 | 0.6122 | 2194010 | 0 | GE |
| ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 4601806 | 2017-11-10 | 37.7000 | 37.7000 | 37.3100 | 37.4100 | 111806 | 0 | EMQQ |
| 12663463 | 2017-11-10 | 23.9300 | 23.9700 | 23.4100 | 23.4100 | 41433 | 0 | SSW_H |
| 12663150 | 2017-11-10 | 24.2300 | 24.3372 | 23.8500 | 23.8900 | 41764 | 0 | SSW_G |
| 12666405 | 2017-11-10 | 1.4700 | 1.4700 | 1.4400 | 1.4400 | 5229 | 0 | SSY |
| 14887664 | 2017-11-10 | 11.6800 | 13.1500 | 11.3043 | 12.4600 | 885587 | 0 | ZYNE |

14887665 rows × 8 columns

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 14887665 entries, 5675434 to 14887664
Data columns (total 8 columns):
#   Column  Dtype
---  -
0    Date   datetime64[ns]
1    Open   float64
2    High   float64
3    Low    float64
4    Close  float64
5    Volume int64
6    OpenInt int64
7    Name   object
dtypes: datetime64[ns](1), float64(4), int64(2), object(1)
memory usage: 1022.3+ MB
```

Επεξεργασία Δεδομένων:

- Αφαιρέσαμε δεδομένα με ελλιπείς εγγραφές για όλα τα Columns
- Τοποθέτηση δεδομένων σε αύξουσα σειρά με βάση την ημερομηνία
- Δημιουργία νέου index για της μετρήσεις
- Δημιουργία καινούργιων κανονικοποιημένων columns για το ποσοστιαίο κέρδος:
 - Profit1: $(High - Open) / Open$
 - Profit2: $(Close - Low) / Low$
 - Best_Profit: Το μεγαλύτερο μεταξύ των (Profit1 & Profit2)

Συνάρτηση BestDailyProfit:

Με την χρήση της παραπάνω συνάρτησης αναζητούμε το Best_Profit ανά μέρα για όλες τις μέρες. Πιο αναλυτικές πληροφορίες επισυνάπτονται ως σχόλια στον κώδικα.

Συνάρτηση BestProfitOrder:

Για όλες τις ημέρες στο dataset, θεωρούμε τις παρακάτω δύο περιπτώσεις:

- Profit1 > Profit2
- Profit2 > Profit1

Στη συνέχεια προσθέτουμε το ανάλογο κόστος αγοράς και πώλησης ανά μετοχή (1%) καθώς και την συνθήκη για την δυνατότητα αγοραπωλησίας όπου θα πρέπει οι υπάρχουσες μετοχές που επιθυμούμε να συναλλάξουμε να μην ξεπερνούν το 10% του συνολικού διαθέσιμου όγκου της εν' λόγω μετοχής.

Η στρατηγική βασισμένη στις 2 παραπάνω κατηγορίες οδηγεί στις εξής 2 κινήσεις:

- Buy-Open → Sell-High
- Buy-Low → Sell-Close

Πιο αναλυτικά σχόλια παρέχονται στον κώδικα.

Small & Large Cases:

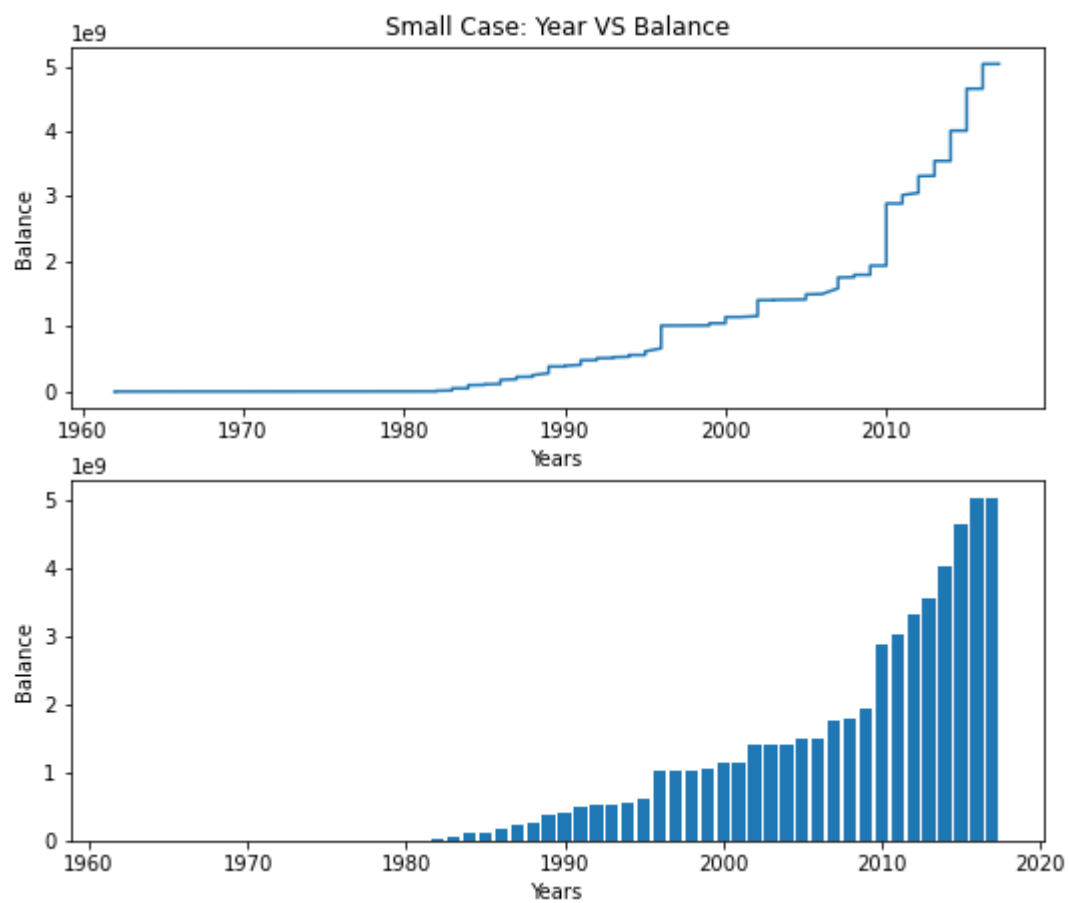
Για την Small Case χρησιμοποιήσαμε τις 9 μετοχές με το καλύτερο Best_Profit για όλο το χρονικό διάστημα.

Για το Large Case χρησιμοποιήσαμε όλες τις διαθέσιμες μετοχές για όλο το χρονικό διάστημα επίσης.

| | Total Transactions | Total Profit |
|------------|--------------------|------------------|
| Small Case | 952 | 5036937366.8293 |
| Large Case | 24494 | 75926513219.3736 |

Γραφήματα

Small Case



Large Case

