

# Προγραμματιστικά Εργαλεία και Τεχνολογίες για Επιστήμη Δεδομένων

Εργασία στην Python, 14/12/2021, Νίκος Παπασπύρου.

Προθεσμία παράδοσης: Κυριακή 30/1/2022.

## Πρόβλημα “time-travel”

Κάποιος φίλος σας φυσικός ανακάλυψε κάτι πολύ παράξενο: βρήκε τον τρόπο να κάνει αγοραπωλησίες μετοχών στο χρηματιστήριο της Νέας Υόρκης, όχι σήμερα (όπως ο υπόλοιπος κόσμος, πφφφ), αλλά σε οποιαδήποτε ημερομηνία θέλει στο παρελθόν! Συγκεκριμένα, ανακάλυψε έναν άτοκο τραπεζικό λογαριασμό, που την 1/1/1960 περιείχε 1 δολάριο, και κάποιον σχεδόν μαγικό τρόπο να δίνει εντολές για την αγορά ή την πώληση μετοχών όποια μέρα θέλει, χρησιμοποιώντας τα χρήματα που υπάρχουν σε αυτόν το λογαριασμό. Προφανώς, η αγορά μετοχών γίνεται με την αφαίρεση χρημάτων από το λογαριασμό και η πώληση έχει ως αποτέλεσμα την πρόσθεση χρημάτων στο λογαριασμό.

Ο φίλος σας ζητάει τη βοήθειά σας για να βρει κάποια καλή ακολουθία αγορών και πωλήσεων που θα τον κάνει πλούσιο. Αν τον βοηθήσετε, σας προτείνει να μοιραστείτε τα κέρδη. Σκέφτεστε λίγο, ψάχνετε στο Internet και βρίσκετε ότι οι ιστορικές τιμές όλων των μετοχών του χρηματιστηρίου της Νέας Υόρκης είναι διαθέσιμες σε **ένα μεγάλο συμπίεμένο αρχείο**. Συγκεκριμένα, ο κατάλογος Stocks περιέχει για κάθε μετοχή  $s$  ένα αρχείο CSV με όνομα  $s.us.txt$  — για παράδειγμα, για τη μετοχή AAPL της Apple Inc. υπάρχει το αρχείο με όνομα `aapl.us.txt`. Κάθε τέτοιο αρχείο περιέχει, μεταξύ άλλων, για κάθε ημερομηνία  $d$  την τιμή **ανοίγματος** (open) της μετοχής  $s$  τη μέρα  $d$ , τη **μέγιστη** τιμή (high) της μετοχής  $s$  τη μέρα  $d$ , την **ελάχιστη** τιμή (low) της μετοχής  $s$  τη μέρα  $d$ , την τιμή **κλεισίματος** (close) της μετοχής  $s$  τη μέρα  $d$ , και τον **όγκο** (volume) των αγοραπωλησιών της μετοχής  $s$  τη μέρα  $d$  (δηλαδή το πλήθος των μετοχών που άλλαξαν χέρια).

Χρησιμοποιώντας αυτή την πληροφορία, σκέφτεστε ότι θα είναι μάλλον εύκολο να βοηθήσετε το φίλο σας και τον ρωτάτε τι ακριβώς θέλει. Εκείνος σας λέει ότι χρειάζεται μία ακολουθία  $N$  κινήσεων, κάθε μία από τις οποίες θα είναι μία τετράδα της μορφής  $(d_i, m_i, s_i, x_i)$  για  $0 \leq i < N$ , όπου  $d_i$  είναι η μέρα κατά την οποία θα εκτελεστεί,  $m_i$  ο κωδικός της κίνησης (βλ. παρακάτω),  $s_i$  ο κωδικός της μετοχής που αφορά και  $x_i$  το πλήθος των μετοχών είτε προς αγορά είτε προς πώληση.<sup>1</sup>

Τα βασικά είδη κινήσεων που μας συμφέρει να εκτελούμε είναι τα εξής:

- buy-low: αγόρασε στην ελάχιστη τιμή.
- sell-high: πούλησε στη μέγιστη τιμή.

Βάσει των πληροφοριών που έχουμε στη διάθεσή μας, γνωρίζουμε την ελάχιστη και τη μέγιστη τιμή κάθε μετοχής μέσα στη μέρα αλλά δε γνωρίζουμε με ασφάλεια αν η ελάχιστη τιμή προηγήθηκε χρονικά της μέγιστης ή αντίστροφα. Για το λόγο αυτό, θεωρούμε ότι αυτές οι αγορές και οι πωλήσεις γίνονται ανεξάρτητα, δηλαδή πρέπει εκ των προτέρων να έχουμε διαθέσιμα χρήματα για τις μετοχές που θέλουμε να αγοράσουμε και επίσης να έχουμε στην κατοχή μας τις μετοχές που θέλουμε να πουλήσουμε.

Ένας άλλος φίλος οικονομολόγος όμως μας πρότεινε ότι μπορούμε να έχουμε μεγαλύτερο κέρδος αν κάνουμε και *intra-day trading*, δηλαδή αν αγοράζουμε και πουλάμε *την ίδια μέρα*, εκμεταλλευόμενοι τη διακύμανση των μετοχών. Για το λόγο αυτό, προσθέτουμε και τα εξής είδη κινήσεων:

- buy-open: αγόρασε στην τιμή ανοίγματος.
- sell-open: πούλησε στην τιμή ανοίγματος.
- buy-close: αγόρασε στην τιμή κλεισίματος.
- sell-close: πούλησε στην τιμή κλεισίματος.

Οι κινήσεις εκτελούνται με χρονολογική σειρά. Μέσα στην ίδια μέρα, οι κινήσεις εκτελούνται με την εξής σειρά:

$\{ \text{buy-open}, \text{sell-open} \} \gg \{ \text{buy-low}, \text{sell-high} \} \gg \{ \text{buy-close}, \text{sell-close} \}$

δηλαδή σε τρεις ομάδες: κατά το άνοιγμα, κατά τη διάρκεια της μέρας και κατά το κλείσιμο. Πριν να εκτελεστούν οι εντολές κάθε ομάδας, πρέπει να υπάρχουν στο λογαριασμό αρκετά χρήματα για όλες τις εντολές αγορών και να έχουμε στην κατοχή μας όλες τις μετοχές προς πώληση. Η σειρά με την οποία εκτελούνται οι εντολές κάθε ομάδας δεν έχει σημασία.

Ορίζουμε ως **κέρδος** (σε δολάρια) μίας έγκυρης ακολουθίας κινήσεων το υπόλοιπο του λογαριασμού αφού εκτελεστούν όλες οι κινήσεις, όπως περιγράφηκε παραπάνω, έχοντας ξεκινήσει με 1 δολάριο στο λογαριασμό την 1/1/1960.

Δεδομένης μίας έγκυρης ακολουθίας, ορίζουμε ως **αποτίμηση** (σε δολάρια) για μία συγκεκριμένη ημέρα το άθροισμα του υπολοίπου του λογαριασμού και της αξίας όλων των μετοχών που βρίσκονται στην κατοχή μας στο κλείσιμο εκείνης της ημέρας (δηλαδή το άθροισμα των γινομένων των μετοχών επί των αντίστοιχων τιμών κλεισίματος).

**Προσοχή!** Προκειμένου να μην αλλοιωθεί το συνεχές του χωροχρόνου και να μην εξαφανιστούν η Νέα Υόρκη και το χρηματιστήριό της από πρόσωπο γης,<sup>2</sup> πρέπει να πληρούνται οι εξής προϋποθέσεις:

<sup>1</sup> Για λόγους απλότητας, στην εργασία αυτή κάνουμε την αφελή παραδοχή ότι οι αγοραπωλησίες μετοχών δεν έχουν προμήθεια.

<sup>2</sup> Μεγάλη μπαρούφα, λες και δεν του αλλάξαμε ήδη τα φώτα του συνεχούς, αλλά όλες οι σχετικές ταινίες sci-fi έτσι λένε.

1. Το συνολικό πλήθος των μετοχών  $s$  που αγοράζουμε την ημέρα  $d$  δεν πρέπει να ξεπερνά το 10% του όγκου της μετοχής  $s$  την ημέρα  $d$ .
2. Ομοίως, το συνολικό πλήθος των μετοχών  $s$  που πουλάμε την ημέρα  $d$  δεν πρέπει να ξεπερνά το 10% του όγκου της μετοχής  $s$  την ημέρα  $d$ .
3. Αν πριν το άνοιγμα της ημέρας  $d$  έχουμε στην κατοχή μας  $n \geq 0$  μετοχές  $s$ , τότε κατά τη διάρκεια της ημέρας  $d$  δεν μπορούμε να αγοράσουμε συνολικά περισσότερες από  $n + 1$  μετοχές  $s$ .

### Τι πρέπει να παραδώσετε:

Επισυνάψτε σε e-mail που θα στείλετε στη διεύθυνση [progds@courses.softlab.ntua.gr](mailto:progds@courses.softlab.ntua.gr) μέχρι την καταληκτική ημερομηνία παράδοσης τα ακόλουθα:

1. Μία έγκυρη ακολουθία  $N \leq 1000$  κινήσεων με μεγάλο κέρδος.

Θα αποθηκεύσετε την ακολουθία σε ένα αρχείο κειμένου `small.txt` με την εξής γραμμογράφηση. Η πρώτη γραμμή θα πρέπει να περιέχει το πλήθος  $N$  των κινήσεων. Κάθε μία από τις επόμενες  $N$  γραμμές θα περιγράφει μία κίνηση, σύμφωνα με το παράδειγμα που ακολουθεί (το παράδειγμα οδηγεί σε κέρδος 401.405, 23731 δολαρίων):

```
71
1962-06-25 buy-low GE 1
1962-06-27 buy-low GE 1
1965-12-31 sell-high GE 2
1970-05-25 buy-low GE 1
1970-05-26 buy-low GE 1
1970-05-27 buy-low GE 2
1971-12-20 sell-high GE 2
1972-02-24 buy-low GE 2
1973-01-11 sell-high GE 4
1974-09-11 buy-low GE 1
1974-09-12 buy-low GE 2
1974-09-13 buy-low GE 3
1974-09-16 buy-low GE 4
1984-09-14 sell-high GE 10
1984-10-09 buy-low AAPL 1
1984-10-10 buy-low AAPL 2
1984-10-11 buy-low AAPL 4
1984-10-12 buy-low AAPL 8
1984-11-16 buy-low AAPL 12
1984-11-19 buy-low AAPL 27
1985-01-15 sell-high AAPL 54
1985-08-02 buy-low AAPL 1
1985-08-05 buy-low AAPL 2
1985-08-06 buy-low AAPL 4
1985-08-07 buy-low AAPL 8
1985-08-12 buy-low AAPL 16
1985-08-15 buy-low AAPL 32
1985-08-16 buy-low AAPL 9
1985-09-26 buy-low GE 1
1985-10-04 buy-low GE 2
1985-10-07 buy-low GE 2
1987-10-19 sell-open AAPL 72
1987-10-19 buy-low AAPL 73
1987-10-19 sell-high GE 5
1987-10-19 buy-close GE 6
1997-12-16 sell-high GE 6
1997-12-29 buy-low AAPL 74
1997-12-30 buy-low AAPL 148
1998-01-06 buy-low AAPL 116
1998-01-06 sell-high AAPL 71
2000-01-20 sell-high AAPL 340
2001-09-05 buy-low AMZN 1
2001-09-17 buy-low AMZN 2
```

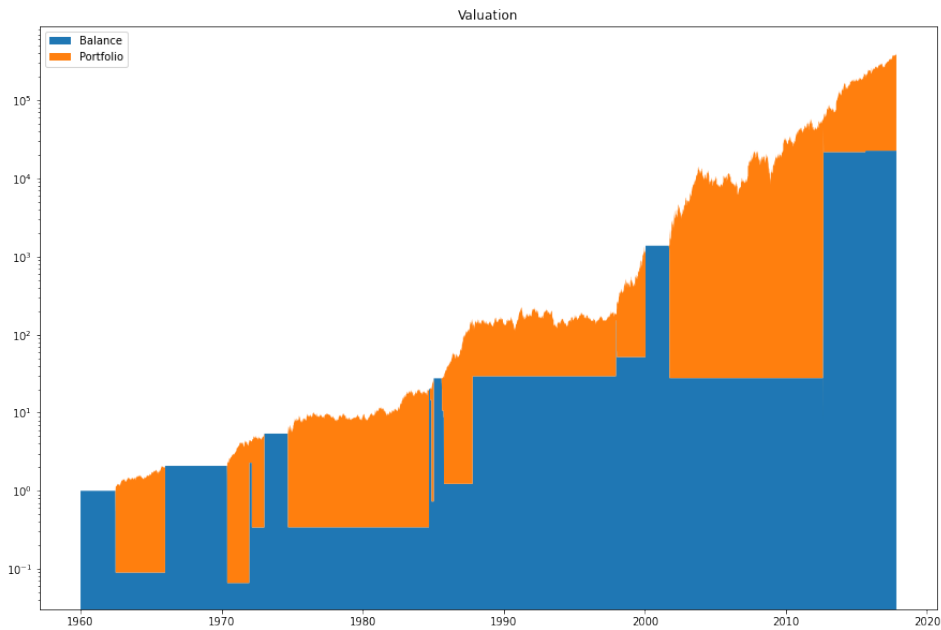
2001-09-19 buy-low AMZN 4  
 2001-09-21 buy-low AMZN 8  
 2001-09-26 buy-low AMZN 16  
 2001-09-27 buy-low AMZN 32  
 2001-09-28 buy-low AMZN 64  
 2001-10-01 buy-low AMZN 108  
 2012-08-16 buy-low FB 1  
 2012-08-17 sell-high AMZN 3  
 2012-08-17 buy-close FB 2  
 2012-08-20 buy-low FB 4  
 2012-08-21 buy-low FB 8  
 2012-08-22 buy-low FB 16  
 2012-08-22 sell-high AMZN 2  
 2012-08-24 buy-low FB 32  
 2012-08-27 sell-open AMZN 5  
 2012-08-27 buy-low FB 64  
 2012-08-27 sell-high AMZN 10  
 2012-08-28 buy-low FB 128  
 2012-08-29 sell-high AMZN 20  
 2012-08-30 buy-low FB 256  
 2012-08-30 sell-high AMZN 195  
 2012-08-31 buy-low FB 512  
 2012-09-04 buy-low FB 1024  
 2015-08-24 buy-open FB 250  
 2015-08-24 buy-low FB 34  
 2015-08-24 sell-high FB 250  
 2015-08-24 sell-close FB 10  
 2017-11-01 sell-high FB 2071

2. Μία έγκυρη ακολουθία  $N \leq 1.000.000$  κινήσεων με μεγάλο κέρδος, αποθηκευμένη σε ένα αρχείο κειμένου `large.txt` με τον ίδιο τρόπο όπως παραπάνω.
3. Μία έκθεση `report.pdf` που να περιγράφει τη μέθοδο που ακολουθήσατε για να βρείτε τις δύο ακολουθίες. Για κάθε ακολουθία, φροντίστε να συμπεριλάβετε στην έκθεσή σας ένα διάγραμμα της αποτίμησής σας συναρτήσει του χρόνου, όπως το Σχήμα 1 (που αντιστοιχεί στο δοθέν παράδειγμα).
4. Ένα συμπιεσμένο αρχείο `scripts.zip` (ή αντίστοιχο) που να περιέχει τυχόν `scripts` (ή `notebooks`) που χρησιμοποιήσατε για να διαβάσετε και να επεξεργαστείτε τα ιστορικά δεδομένα, να βρείτε τις ακολουθίες κινήσεων, να σχεδιάσετε τα διαγράμματα, κ.λπ.

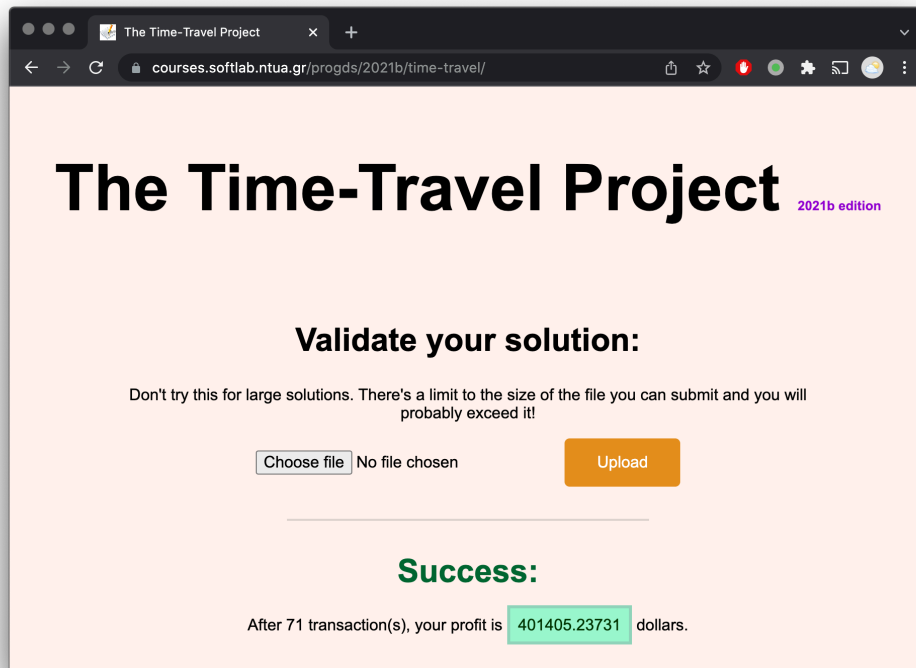
Μπορείτε να δοκιμάσετε τις ακολουθίες που παράγετε υποβάλλοντας αυτές στην παρακάτω ιστοσελίδα (βλ. Σχήμα 2).

<https://courses.softlab.ntua.gr/progds/2021b/time-travel/>

**Προσοχή:** Το να βρείτε το μέγιστο δυνατό κέρδος είναι ένα πολύ δύσκολο αλγοριθμικό πρόβλημα. Προτείνω να μην το επιχειρήσετε. Επιστρατεύστε τη διαίσθηση και τη φαντασία σας και κάνετε ό,τι καλύτερο μπορείτε. Αν ποτέ βρεθείτε στην ευχάριστη θέση να μπορείτε να ταξιδέψετε πίσω στο χρόνο και η εργασία αυτή σας φανεί χρήσιμη, ο διδάσκων θα δέχεται δωρεές ύψους 1% των κερδών σας. ☺



Σχήμα 1: Διάγραμμα αποτίμησης.



Σχήμα 2: Ιστοσελίδα επικύρωσης ακολουθιών κινήσεων.