



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ  
ΥΠΟΛΟΓΙΣΤΩΝ

ΔΙ-ΙΔΡΥΜΑΤΙΚΟ ΠΜΣ ΕΠΙΣΤΗΜΗ ΔΕΔΟΜΕΝΩΝ ΚΑΙ ΜΗΧΑΝΙΚΗ  
ΜΑΘΗΣΗ

ΜΑΘΗΜΑ: ΠΡΟΓΡΑΜΜΑΤΙΣΤΙΚΑ ΕΡΓΑΛΕΙΑ ΚΑΙ ΤΕΧΝΟΛΟΓΙΕΣ ΓΙΑ ΤΗΝ  
ΕΠΙΣΤΗΜΗ ΔΕΔΟΜΕΝΩΝ

ΤΙΤΛΟΣ ΠΑΡΑΔΟΤΕΟΥ: ΕΞΑΜΗΝΙΑΙΑ ΕΡΓΑΣΙΑ ΣΤΗΝ ΡΥΘΜΟΝ

ΗΜΕΡΟΜΗΝΙΑ ΠΑΡΑΔΟΣΗΣ: 12/01/2025

ΟΝΟΜΑΤΕΠΩΝΥΜΟ: [REDACTED]

ΑΡΙΘΜΟΣ ΜΗΤΡΩΟΥ [REDACTED]

E-MAIL: [REDACTED]



<https://medium.com/magnidata/stock-price-prediction-and-forecasting-via-an-object-oriented-programming-5284fd9f220f>

## **ΕΙΣΑΓΩΓΗ**

Στόχος της παρούσας εργασίας είναι βάσει ενός μεγάλου αρχείου στο οποίο παρατίθενται ιστορικές τιμές μετοχών στο χρηματιστήριο της Νέας Υόρκης (1962-2018) και ξεκινώντας την 1/1/1960 με υπόλοιπο 1 δολαρίου, η δημιουργία 2 ακολουθιών συναλλαγών (αγορά και πώληση μετοχών στο άνοιγμα ή κλείσιμο του χρηματιστηρίου, αγορά μόνο στην χαμηλότερη τιμή και πώληση μόνο στην υψηλότερη τιμή) αγοράς και πώλησης μετοχών η οποία θα αποφέρει μεγάλο κέρδος. Η 1η “μικρή” ακολουθία συναλλαγών θα έχει μήκος  $\leq 1.000$  συναλλαγών ενώ η 2η “μεγάλη” ακολουθία θα έχει μήκος  $\leq 1.000.000$  συναλλαγών. Στα παρακάτω κεφάλαια θα αναλυθεί η διαδικασία προεπεξεργασίας των αρχικών δεδομένων (σειρά τιμών αγοράς/πώλησης για κάθε μετοχή) και παραγωγής των ακολουθιών μαζί με τα αντίστοιχα διαγράμματα αποτίμησης.

## **ΠΡΟΕΠΕΞΕΡΓΑΣΙΑ ΔΕΔΟΜΕΝΩΝ**

Το στάδιο της προεπεξεργασίας δεδομένων έγινε δυνατό μέσω των μεθόδων της κλάσης *StockData* της οποίας ο κώδικας παρατίθεται στο συνημμένο script [StockData.py](#). Πρώτο βήμα προεπεξεργασίας είναι η ανάγνωση των επιμέρους αρχείων κειμένου που περιέχουν τις διαχρονικές σειρές τιμών αγοράς/πώλησης για κάθε μετοχή, ενώ ανάλογα με το μήκος της ακολουθίας των συναλλαγών μέσω της μεθόδου *read\_analyze\_stocks()* επιστρέφονται τα κάτωθι αποτελέσματα:

1. Λίστα με τους πίνακες των σειρών τιμών των μετοχών σαν pandas dataframes για μήκος ακολουθίας συναλλαγών  $\leq 1000$ .
2. Λίστα με τους πίνακες τιμών των μετοχών καθώς και 3 πινάκων που για 3 χρονικές περιόδους (επιλέχθηκαν οι χρονικές περίοδοι 1962-1980, 1980-2000 και 2000-2018) παρουσιάζουν την συνολική ποσοστιαία απόδοση (total return) των μετοχών όπως στον πίνακα παρακάτω.

Stock	Total_Return	Total_Years	Start_Date	End_Date
HD	2307,32	18,27	1981-09-22	1999-12-31
TWX	740,93	7,78	1992-03-19	1999-12-31
CSCO	670,09	9,77	1990-03-26	1999-12-31
MSFT	658,14	13,80	1986-03-13	1999-12-31
AMGN	652,88	15,31	1984-09-07	1999-12-31
WMT	634,45	19,99	1980-01-02	1999-12-31
VIAV	332,75	6,12	1993-11-17	1999-12-31
BRK-A	193,14	19,79	1980-03-17	1999-12-31
QCOM	176,27	8,04	1991-12-16	1999-12-31

Πίνακας 1. Πίνακας αποδόσεων και συνολικών ετών στο χρηματιστήριο μετοχών για τα έτη 1980-2000

Διευκρινίζεται ότι η συνολική απόδοση μιας μετοχής ορίστηκε βάσει του παρακάτω τύπου:

$$Total\ Return = \frac{P_{end} - P_{start}}{P_{start}},$$

όπου  $P_{start}$  η αρχική τιμή και  $P_{end}$  η τελική τιμή της μετοχής για την αρχική και τελική ημερομηνία αντίστοιχα.

Έπειτα, για την “μικρή” ακολουθία  $\leq 1.000$  υλοποιήθηκε και η μέθοδος *filter\_stocks()* βάσει της οποίας διατηρήθηκαν για κάθε χρονική περίοδο μόνο οι αρχικές και οι τελικές τιμές μετοχών βάσει των παρακάτω κριτηρίων:

- Για την χρονική περίοδο 1962-1980 διατηρήθηκαν μόνο οι μετοχές με συνολική απόδοση μεγαλύτερη του 2% που είναι εισηγμένες τουλάχιστον 5 έτη στο χρηματιστήριο.
- Για την χρονική περίοδο 1980-2000 διατηρήθηκαν μόνο οι μετοχές με συνολική απόδοση μεγαλύτερη του 100% που είναι εισηγμένες τουλάχιστον 5 έτη στο χρηματιστήριο.
- Για την χρονική περίοδο 2000-2018 διατηρήθηκαν μόνο οι μετοχές με συνολική απόδοση μεγαλύτερη του 50% που είναι εισηγμένες τουλάχιστον 10 έτη στο χρηματιστήριο.

Τέλος, μέσω της μεθόδου *concat\_stock\_dfs()* συνενώθηκαν οι επιμέρους πίνακες τιμών και για τις 2 περιπτώσεις ακολουθιών. Παράδειγμα για την χρονική περίοδο 1962-1980 παρατίθεται στον παρακάτω πίνακα.

Date	Open	High	Low	Close	Volume	Stock_Name
1970-01-02	0,7587	0,8092	0,7587	0,8092	753088	BA
1970-01-02	0,7383	0,7538	0,7383	0,7538	83787	CVX
1970-01-02	1,0666	1,1097	1,0666	1,0924	62781	HON
1970-01-02	0,28338	0,29161	0,28338	0,29161	3496683	MCD
1970-01-02	0,0219	0,0219	0,0219	0,0219	11519926	MO
1970-01-02	0,32548	0,334	0,32548	0,334	283968	UTX
1972-01-07	0,01592	0,01592	0,01592	0,01592	3787746	INTC
1979-12-31	4,0883	4,1305	4,0801	4,1136	1327214	BA
1979-12-31	2,4343	2,4581	2,4343	2,4343	163504	CVX
1979-12-31	3,38	3,4403	3,3717	3,4148	207878	HON

1979-12-31	0,28077	0,28868	0,28077	0,28077	3519525	INTC
1979-12-31	0,8582	0,8663	0,8582	0,8582	785376	MCD
1979-12-31	0,10243	0,10243	0,10243	0,10243	6951678	MO
1979-12-31	1,1989	1,1989	1,1906	1,1906	564197	UTX

Πίνακας 2. Πίνακας αρχικών και τελικών τιμών μετοχών με υψηλή απόδοση για την χρονική περίοδο 1962-1980

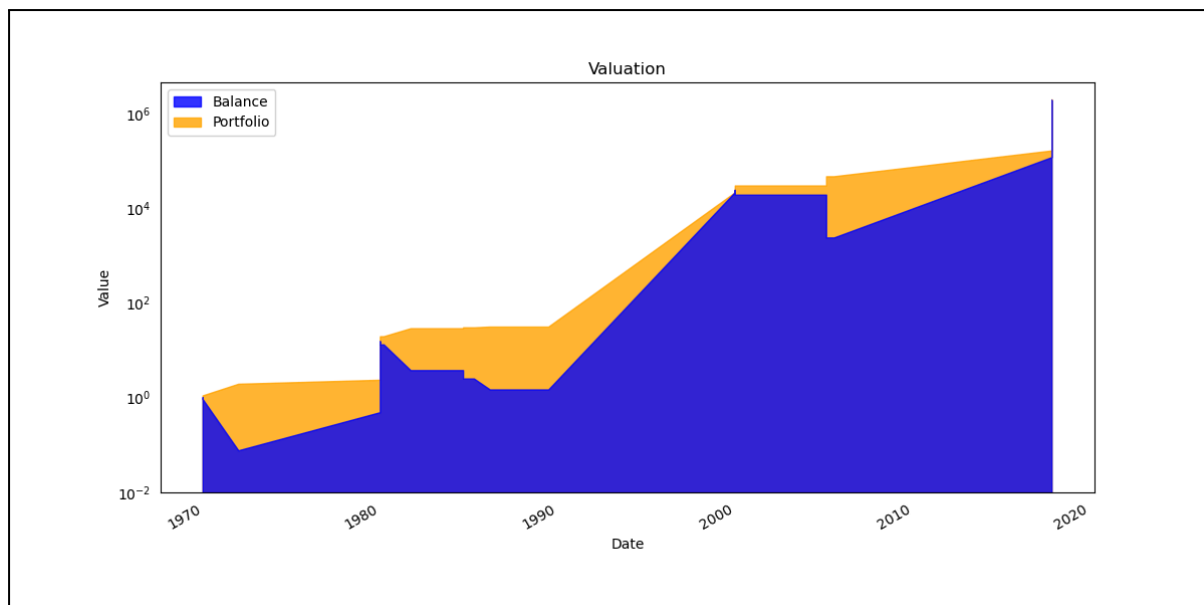
### **ΣΤΡΑΤΗΓΙΚΗ ΔΗΜΙΟΥΡΓΙΑΣ ΜΙΚΡΗΣ ΑΚΟΛΟΥΘΙΑΣ ΣΥΝΑΛΛΑΓΩΝ**

Για την δημιουργία της μικρής ακολουθίας συναλλαγών  $\leq 1.000$  εγγραφών, ξεκινώντας από τον πίνακα με τις αρχικές και τελικές τιμές μετοχών για τις επιμέρους χρονικές στιγμές και τις αντίστοιχες συνολικές αποδόσεις, υπολογίστηκε για κάθε μετοχή ένα βάρος βάσει του οποίου ορίστηκε το ποσοστό επένδυσης σε κάθε διαθέσιμη μετοχή για όγκο μικρότερο του 10% του συνολικού διαθέσιμου. Ο τύπος του βάρους είναι ο παρακάτω, όπου το άθροισμα αναφέρεται στο σύνολο των διαθέσιμων μετοχών για κάθε χρονικό εύρος:

$$Investment\ Weight = \frac{Total\ Return}{\sum_{i=0}^n Total\ Return}$$

Παραδείγματος χάρη, για την μετοχή INTC με συνολική απόδοση ίση με 17,13% στο χρονικό εύρος 1962-1980 υπολογίζεται βάρος ίσο με 0,4759 και επομένως για το ίδιο χρονικό εύρος το 47,59% των επενδύσεων θα γίνουν στην συγκεκριμένη μετοχή. Έπειτα, μέσω της συνάρτησης `Stock_Trader_1000` στο script [Stock\\_Trader\\_1000.py](#) δημιουργείται η ακολουθία των συναλλαγών όπου αγοράζονται οι μετοχές στην αρχή των 3 επιμέρους χρονικών περιόδων βάσει του βάρους που ορίστηκε και πωλούνται στο τέλος της, πάντα τηρώντας τους περιορισμούς της μη υπέρβασης αγοράς ή πώλησης >10% του διαθέσιμου όγκου για μια συγκεκριμένη μετοχή καθώς και της χρέωσης προμήθειας 1% για κάθε μετοχή.

Έπειτα από 24 συναλλαγές, το υπολογιζόμενο κέρδος βρέθηκε ίσο με 1.714.969,49 δολάρια. Επιπλέον, πάλι μέσω της συνάρτησης `Stock_Trader_1000` εξήχθησαν και 2 πίνακες που περιέχουν το κόστος ή το κέρδος από την αγορά και πώληση των μετοχών αντίστοιχα για να υπολογιστεί το υπόλοιπο του λογαριασμού μετά από κάθε κίνηση καθώς και η αποτίμηση, δηλαδή το άθροισμα του υπολοίπου και της αξίας όλων των κατεχόμενων μετοχών στην τιμή κλεισίματος του χρηματιστηρίου. Το διάγραμμα αποτίμησης το οποίο φαίνεται παρακάτω δημιουργήθηκε μέσω της συνάρτησης `Valuation` η οποία βρίσκεται στο script [Valuation.py](#).

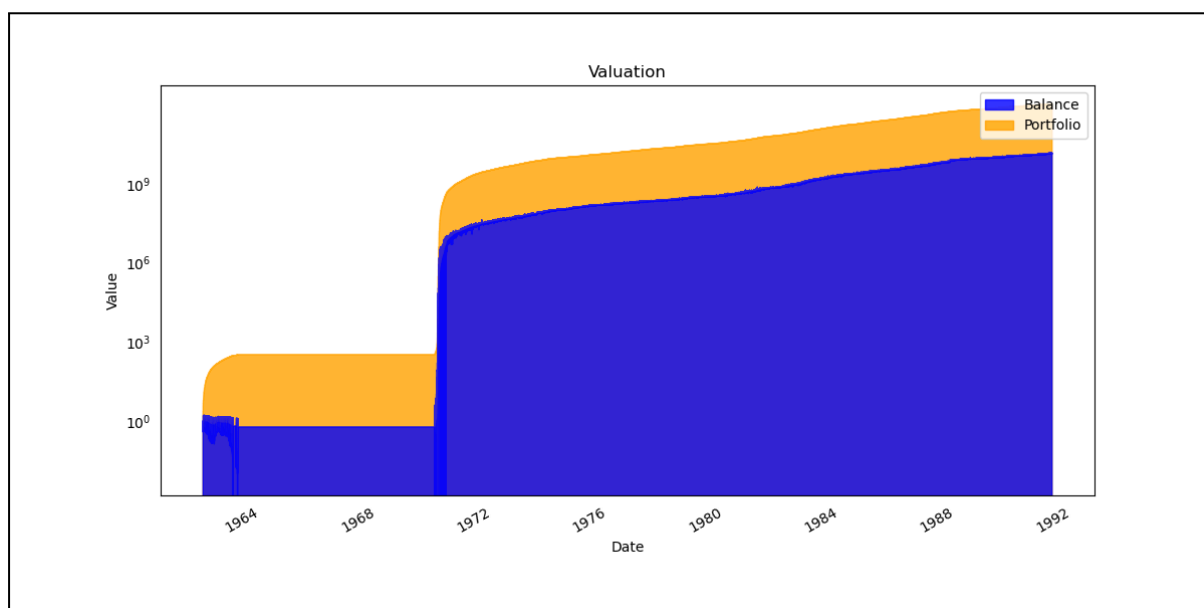


Σχήμα 1. Διάγραμμα αποτίμησης “μικρής” ακολουθίας  $\leq 1.000$  συναλλαγών

Η υλοποίηση της παραπάνω στρατηγικής υλοποιήθηκε μέσω του script [Python\\_APP\\_1k.py](#) με δεδομένα εισόδου ολόκληρο το directory με όλες τις διαθέσιμες μετοχές.

### **ΣΤΡΑΤΗΓΙΚΗ ΔΗΜΙΟΥΡΓΙΑΣ ΜΕΓΑΛΗΣ ΑΚΟΛΟΥΘΙΑΣ ΣΥΝΑΛΛΑΓΩΝ**

Για την δημιουργία της “μεγάλης” ακολουθίας συναλλαγών ακολουθήθηκε μια διαφορετική στρατηγική ξεκινώντας από έναν πίνακα ο οποίος περιέχει όλες τις διαθέσιμες μετοχές από το 1962 έως το 2017. Έπειτα, μέσω της συνάρτησης *StockTrader\_1mil* στο script [StockTrader\\_1mil.py](#) πραγματοποιήθηκαν αγοραπωλησίες μετοχών την ίδια μέρα (intra-day trading) αν υπήρχε δυνατότητα κέρδους. Μέσω της συγκεκριμένης στρατηγικής, μετά από 1.000.000 συναλλαγές που έγιναν στο χρονικό διάστημα από τις 02-01-1962 έως τις 26-04-1991, το κέρδος ήταν ίσο με 14.506.458.292,86 δολάρια. Παρακάτω παρατίθεται και το διάγραμμα αποτίμησης για την συγκεκριμένη ακολουθία.



Σχήμα 2. Διάγραμμα αποτίμησης “μεγάλης” ακολουθίας  $\leq 1.000.000$  συναλλαγών

Κατά αντιστοιχία, η παραπάνω στρατηγική υλοποιήθηκε μέσω του script Python APP 1mil.py με τα ίδια δεδομένα εισόδου.