Αναφορά Επενδυτικού Χαρτοφυλακίου

# 1. Επενδυτικό Προφίλ (Persona)

Ο επενδυτής είναι 65 ετών και διαθέτει συνολικό χαρτοφυλάκιο αξίας 1 εκατομμυρίου δολαρίων. Μέχρι σήμερα, έχει επενδύσει αποκλειστικά σε κρατικά και εταιρικά ομόλογα καθώς και σε προθεσμιακές καταθέσεις. Στόχος του είναι να τοποθετήσει το 15% του κεφαλαίου του σε επενδύσεις με μεγαλύτερο ρίσκο (μετοχές και κρυπτονομίσματα), διατηρώντας ωστόσο έναν γενικά συντηρητικό χαρακτήρα.

# 2. Κατανομή Χαρτοφυλακίου

Η προτεινόμενη κατανομή επενδύσεων είναι 90% σε μετοχές και 10% σε κρυπτονομίσματα. Η κατανομή αυτή ισορροπεί μεταξύ σταθερότητας και δυνατοτήτων υπεραπόδοσης. Επιπλέον, επιλέχθηκε ένα μικρό και διαχειρίσιμο χαρτοφυλάκιο, που αποτελείται από 5 μετοχές και 2 κρυπτονομίσματα, ώστε ο επενδυτής να μπορεί να το παρακολουθεί εύκολα και να μην χαθεί μέσα στην πολυπλοκότητα πολλών assets. Η επιλογή αυτή κρίνεται ιδιαίτερα σημαντική, δεδομένου ότι πρόκειται για έναν συντηρητικό επενδυτή που για πρώτη φορά εισέρχεται σε επενδύσεις με υψηλότερο ρίσκο.

## Γιατί 90% Μετοχές και 10% Crypto

Η επιλογή αυτής της κατανομής αντανακλά το συντηρητικό προφίλ του επενδυτή. Οι μετοχές μεγάλων και ώριμων εταιρειών (blue chips) προσφέρουν σταθερότερες αποδόσεις, ισχυρά θεμελιώδη και χαμηλότερη μεταβλητότητα σε σχέση με τα κρυπτονομίσματα. Το 90% της τοποθέτησης σε μετοχές εξασφαλίζει προβλεψιμότητα και διατήρηση αξίας, ενώ το 10% σε crypto επιτρέπει στον επενδυτή να επωφεληθεί από την καινοτομία και τις υψηλές πιθανές αποδόσεις, χωρίς να αλλοιώνεται ο συντηρητικός χαρακτήρας του χαρτοφυλακίου. Επιλέχθηκαν τα δύο πιο ώριμα crypto (Bitcoin και Ethereum) για να περιοριστεί ο κίνδυνος.

# 3. Μετοχές

## Microsoft (MSFT) – 20%

• Έδρα: Redmond, Washington, USA

• Κλάδος: Technology / Software & Cloud Computing

• Market Cap (31/12/2024): ~3,1 τρισ. USD

• Ιδρύθηκε το 1975 από Bill Gates & Paul Allen

Επιλέχθηκε λόγω της ηγετικής θέσης σε software και cloud (Azure), υψηλής κερδοφορίας, σταθερής ανάπτυξης και επενδύσεων σε τεχνητή νοημοσύνη.

## Procter & Gamble (PG) – 15%

• Έδρα: Cincinnati, Ohio, USA

• Κλάδος: Consumer Staples / Household & Personal Products

• Market Cap (31/12/2024): ~400 δισ. USD

• Ιδρύθηκε το 1837

Επιλέχθηκε για τον αμυντικό χαρακτήρα της, τα σταθερά μερίσματα και την ανθεκτικότητα στη διάρκεια υφέσεων, καθώς τα προϊόντα της είναι βασικά αγαθά.

## Johnson & Johnson (JNJ) – 15%

• Έδρα: New Brunswick, New Jersey, USA

• Κλάδος: Healthcare / Pharmaceuticals & Medical Devices

• Market Cap (31/12/2024): ~348 δισ. USD

• Ιδρύθηκε το 1886

Επιλέχθηκε για τη διαφοροποίηση σε φάρμακα, ιατρικές συσκευές και προϊόντα υγείας. Είναι μετοχή με ιστορική συνέπεια στη μερισματική πολιτική και σταθερότητα εσόδων.

## Visa (V) – 20%

• Έδρα: Foster City, California, USA

• Κλάδος: Financial Services / Payments Network

• Market Cap (31/12/2024): ~600 δισ. USD

• Ιδρύθηκε το 1958

Επιλέχθηκε λόγω του ηγετικού ρόλου στις ηλεκτρονικές πληρωμές, της ισχυρής κερδοφορίας και του μεγάλου περιθωρίου ανάπτυξης λόγω της παγκόσμιας μετάβασης σε cashless συναλλαγές.

## Costco (COST) – 20%

• Έδρα: Issaquah, Washington, USA

• Κλάδος: Retail / Membership-based Wholesale

• Market Cap (31/12/2024): ~410 δισ. USD

• Ιδρύθηκε το 1983

Επιλέχθηκε λόγω του επιτυχημένου συνδρομητικού μοντέλου, της υψηλής πιστότητας πελατών και της ανθεκτικότητας ακόμη και σε περιόδους ύφεσης.

# 4. Κρυπτονομίσματα

## Bitcoin (BTC) – 6%

• Έναρξη: 2009 από τον Satoshi Nakamoto

• Προσφορά: Μέγιστο 21 εκατ. BTC, ~19,9 εκατ. σε κυκλοφορία

• Τιμή (Σεπτέμβριος 2025): ~116.000 USD

Το [Bitcoin](https://coinmarketcap.com/alexandria/article/what-is-bitcoin) είναι ένα ομότιμο διαδικτυακό νόμισμα, που σημαίνει ότι όλες οι [συναλλαγές](https://coinmarketcap.com/alexandria/article/how-long-does-a-bitcoin-transaction-take) πραγματοποιούνται απευθείας μεταξύ ίσων, ανεξάρτητων συμμετεχόντων στο δίκτυο, χωρίς να απαιτείται κάποιος διαμεσολαβητής να επιτρέψει ή να διευκολύνει τις συναλλαγές. 

Επιλέχθηκε ως το παλαιότερο και πιο καθιερωμένο κρυπτονόμισμα, με περιορισμένη προσφορά και ευρεία θεσμική αποδοχή, θεωρείται το 'ψηφιακό χρυσάφι'.

## Ethereum (ETH) – 4%

• Έναρξη: 2015 από τον Vitalik Buterin

• Προσφορά: ~121 εκατ. ETH σε κυκλοφορία

• Τιμή (Σεπτέμβριος 2025): ~4.300–4.700 USD

Επιλέχθηκε ως η μεγαλύτερη πλατφόρμα smart contracts, με κομβικό ρόλο στο οικοσύστημα DeFi και NFTs. Διαθέτει ισχυρή κοινότητα developers και θεσμική στήριξη.

Το Ethereum είναι ένα αποκεντρωμένο σύστημα [blockchain](https://coinmarketcap.com/alexandria/glossary/blockchain) ανοιχτού κώδικα που διαθέτει το δικό του κρυπτονόμισμα, το Ether. Το ETH λειτουργεί ως πλατφόρμα για πολλά άλλα [κρυπτονομίσματα](https://coinmarketcap.com/alexandria/article/what-are-cryptocurrencies), καθώς και για την εκτέλεση αποκεντρωμένων [έξυπνων συμβάσεων](https://coinmarketcap.com/alexandria/glossary/smart-contract).

# 5. Μακροοικονομικοί Δείκτες

Οι ακόλουθοι δείκτες χρησιμοποιούνται για την αξιολόγηση της οικονομικής κατάστασης και επηρεάζουν άμεσα τις αγορές:

• GDP Growth: Μετρά τον ρυθμό ανάπτυξης της οικονομίας. Αν υψηλός → μετοχές ↑. Αν χαμηλός/αρνητικός → φόβος ύφεσης, ενισχύονται αμυντικοί κλάδοι.

• Inflation (CPI): Μετρά τον ρυθμό αύξησης τιμών. Υψηλός πληθωρισμός → υψηλότερα επιτόκια, πιέζονται οι μετοχές, ειδικά τεχνολογίας.

• Unemployment Rate: Μετρά το ποσοστό ανέργων. Υψηλή ανεργία → λιγότερη κατανάλωση, πιέζονται οι αγορές.

• Interest Rate: Βασικό επιτόκιο κεντρικής τράπεζας. Άνοδος → κόστος κεφαλαίου ↑, μετοχές ↓. Πτώση → ενίσχυση risk assets.

• Consumer Confidence: Δείχνει την ψυχολογία νοικοκυριών. Υψηλή → περισσότερες δαπάνες, retail ↑. Χαμηλή → στροφή σε staples.

• Retail Sales: Μετρά τις δαπάνες καταναλωτών. Υψηλές πωλήσεις → στήριξη κερδών, θετικό για μετοχές.

• Industrial Production: Μετρά τη βιομηχανική παραγωγή. Άνοδος → ανάπτυξη, θετικό για cyclical μετοχές.

• Housing Starts: Μετρά νέες οικοδομές. Υψηλές → θετική ένδειξη ζήτησης. Αδύναμες → υποδηλώνουν επιβράδυνση.

• Trade Balance: Μετρά τη διαφορά εξαγωγών–εισαγωγών. Βελτίωση → στηρίζει εξαγωγείς. Επιδείνωση → πιέζει το νόμισμα και εισαγωγείς.

• Money Supply (M2): Μετρά τη ρευστότητα. Αύξηση → στηρίζει μετοχές και crypto. Μείωση → πιέζει αποτιμήσεις.

# 6. Μεθοδολογία Δεδομένων

Στην παρούσα εργασία αξιοποιούμε τα παρεχόμενα αρχεία δεδομένων και επιλέγουμε στοχευμένα ένα υποσύνολο τίτλων μετοχών και κρυπτονομισμάτων (ενδεικτικά MSFT, PG, JNJ, V, COST και BTC, ETH), προκειμένου να υλοποιήσουμε την ανάλυση που απαιτεί η άσκηση.

Τα δεδομένα των κρυπτονομισμάτων μετασχηματίστηκαν σε ημερήσιες παρατηρήσεις μέσω ομαδοποίησης κατά ημερομηνία και σύμβολο. Ως τιμή ανοίγματος ορίστηκε η παρατήρηση της 00:00 κάθε ημέρας, ως τιμή κλεισίματος η παρατήρηση της 18:00 της ίδιας ημέρας, ως υψηλή τιμή το μέγιστο ενδοημερήσιο επίπεδο, ως χαμηλή τιμή το ελάχιστο ενδοημερήσιο επίπεδο, ενώ ο ημερήσιος όγκος υπολογίστηκε ως το άθροισμα όλων των ενδοημερήσιων όγκων του 24ώρου. Η συγκεκριμένη διαδικασία επιλέχθηκε διότι τα κρυπτονομίσματα διαπραγματεύονται σε καθεστώς συνεχούς αγοράς, σε αντίθεση με τις μετοχές που ακολουθούν συγκεκριμένο ωράριο· συνεπώς απαιτείται τυποποίηση των ενδοημερήσιων παρατηρήσεων σε ημερήσια OHLCV, ώστε να διασφαλίζεται η συγκρισιμότητα των σειρών και να αποφεύγονται μεροληψίες που απορρέουν από διαφορετικά ημερολόγια συναλλαγών. Επιπλέον, για λόγους συμβατότητας με το ημερολόγιο διαπραγμάτευσης των μετοχών, η ανάλυση εστιάζει σε εργάσιμες ημέρες (Δευτέρα–Παρασκευή). Η υλοποίηση πραγματοποιήθηκε σε Python με σαφείς κανόνες ομαδοποίησης και συναθροίσεων, ώστε τα παραγόμενα ημερήσια σύνολα OHLCV να είναι μεθοδολογικά συνεπή, αναπαραγώγιμα και άμεσα συγκρίσιμα με τα αντίστοιχα δεδομένα των μετοχών.

# 7. Κώδικας

A screen shot of a computer program

AI-generated content may be incorrect.

**Τι κάνει κάθε γραμμή/μπλοκ.**

564 — Μετατρέπει τη στήλη Timestamp του test\_crypto σε datetime για να μπορούμε να πάρουμε ημερομηνία/ώρα.

566–571 — Δημιουργεί ενδιάμεσο DataFrame: προσθέτει στήλες Date (μόνο ημερομηνία) και Hour (μόνο ώρα) και φιλτράρει εργάσιμες ημέρες (Δευτέρα–Παρασκευή).

573–583 — Ομαδοποίηση ανά Date, Symbol και κατασκευή ημερήσιων bars με τους κανόνες της άσκησης:  
— Open = τιμή στην 00:00   
— High = μέγιστη τιμή της ημέρας  
— Low = ελάχιστη τιμή της ημέρας  
— Close = τιμή στην 18:00   
— Volume = άθροισμα όλου του ημερήσιου όγκου  
Έπειτα απορρίπτει μέρες χωρίς Open/Close και μηδενίζει το index (reset).

585–587 — Θέτει τη σειρά εμφάνισης των συμβόλων (π.χ. BTC πριν από ETH) και ταξινομεί ανά Symbol, Date για σταθερή/ευανάγνωστη έξοδο.

589–592 — Διασφαλίζει ότι οι στήλες Date είναι datetime τόσο στο daily\_df (crypto) όσο και στο test\_stocks (μετοχές), για συμβατότητα.

595 — Συνενώνει κατακόρυφα τα ημερήσια crypto (daily\_df) με τα ημερήσια των μετοχών (test\_stocks) σε ένα ενιαίο merged\_df.

599–600 — Τελική ταξινόμηση του merged\_df ανά Symbol, Date και ανανέωση index.

**Week 2**

**Problem 1**

Στόχος μας είναι η ημερολογιακή ευθυγράμμιση του ενιαίου συνόλου μετοχών και κρυπτονομισμάτων, δεδομένου ότι τα crypto διαπραγματεύονται 24/7 ενώ οι μετοχές μόνο τις εργάσιμες. Χρησιμοποιήσαμε ημερήσια στοιχεία μετοχών για την περίοδο 2015–2024 και προσθέσαμε συνθετικές τιμές αποκλειστικά στα Σαββατοκύριακα των μετοχών, χωρίς καμία παρέμβαση στα δεδομένα των crypto. Για κάθε σύμβολο διατηρήσαμε ανέπαφες τις εργάσιμες παρατηρήσεις και, για όλες τις ημερομηνίες Σαββάτου–Κυριακής, τοποθετήσαμε μία σταθερή τιμή ίση με τη διάμεσο των εργάσιμων τιμών κλεισίματος στο ίδιο δείγμα (μία διάμεσος ανά σύμβολο). Η διάμεσος επιλέχθηκε ως μέθοδος συμπλήρωσης διότι είναι ανθεκτική σε ακραίες τιμές και περιορίζει την επίδραση μεμονωμένων αιχμών, παρέχοντας σταθερή και αναπαραγώγιμη τιμή αναφοράς για τις μη διαπραγματευόμενες ημέρες. Επιπλέον, δεν «επινοεί» ενδιάμεση κίνηση μέσα στο Σαββατοκύριακο (όπως θα έκανε μια γραμμική παρεμβολή μεταξύ Παρασκευής και Δευτέρας), αποφεύγοντας την τεχνητή εισαγωγή κατεύθυνσης στην τάση. Με αυτόν τον απλό και ουδέτερο κανόνα επιτυγχάνεται η ημερολογιακή ενοποίηση με ελάχιστες υποθέσεις και χωρίς αλλοίωση των εργάσιμων τιμών, με την παραδοχή ότι η σταθερή διάμεσος υποεκτιμά τη διακύμανση του Σ/Κ — αποδεκτός συμβιβασμός για τον σκοπό της άσκησης. Ως προς τα αποτελέσματα, ο μέσος όρος «συγκλίνει» προς τη διάμεσο: η προσθήκη πολλών τιμών ίσων με τη διάμεσο αυξάνει το βάρος στο κέντρο της κατανομής, οπότε όταν ο αρχικός μέσος ήταν πάνω από τη διάμεσο (τυπικό σε ανοδικές σειρές), ο νέος μέσος μειώνεται (π.χ. JNJ, MSFT, PG, V), ενώ όπου ο αρχικός μέσος ήταν χαμηλότερος, αυξάνεται (π.χ. COST). Παράλληλα, η διασπορά μειώνεται συστηματικά, επειδή οι ίδιες τιμές των Σ/Κ «συμπιέζουν» τις ουρές και πνίγουν την επίδραση ακραίων εργάσιμων τιμών, με εντονότερη πτώση της τυπικής απόκλισης σε πιο ευμετάβλητες σειρές.

**Problem 2**

Στο πλαίσιο του προβλήματος δοκιμάστηκαν διαφορετικά μεγέθη παραθύρου (3, 7, 14, 20, 30 και 60 ημερών) για την εμφύτευση ελλιπών τιμών με rolling median. Η αξιολόγηση έγινε με τις μετρικές MAE και RMSE προκειμένου να εκτιμηθεί ποιο παράθυρο αποδίδει καλύτερα.

Τα αποτελέσματα έδειξαν διαφοροποίηση μεταξύ των δύο μετρικών: το παράθυρο 30 ημερών πέτυχε τη χαμηλότερη τιμή MAE, ενώ το παράθυρο 60 ημερών εμφάνισε τη χαμηλότερη τιμή RMSE. Η διαφορά αυτή ερμηνεύεται από τη φύση των μετρικών. Το MAE υπολογίζει τον μέσο όρο της απόλυτης απόκλισης, δίνοντας ισότιμο βάρος σε όλα τα σφάλματα, ενώ το RMSE δίνει μεγαλύτερη βαρύτητα στα μεγάλα σφάλματα λόγω του τετραγωνισμού.

Για το συγκεκριμένο dataset, επιλέξαμε ως βέλτιστη λύση το παράθυρο των 30 ημερών, καθώς η μετρική MAE θεωρείται πιο ερμηνεύσιμη και σταθερή σε σχέση με την επίδραση μεμονωμένων μεγάλων αποκλίσεων. Παρόλα αυτά, η ανάλυση έδειξε ότι το παράθυρο των 60 ημερών μειώνει σημαντικά τις ακραίες αποκλίσεις, γεγονός που μπορεί να είναι χρήσιμο σε εφαρμογές όπου οι μεγάλες διαφορές θεωρούνται κρίσιμες.

A graph with a line and a blue line

AI-generated content may be incorrect.

A graph with a line going up

AI-generated content may be incorrect.

**PROBLEM 3**

Στο παρόν σκέλος αξιολογήσαμε τρεις απλές μεθόδους συμπλήρωσης κενών σε ημερήσιες χρονοσειρές τιμών: Forward fill , Backward fill και μια Combined παραλλαγή (διαδοχικό forward–backward για τυχόν εναπομείναντα κενά). Αντίστοιχα υπολογίσαμε τις μετρικές  MAE, RMSE και MAPE για κάθε μια μέθοδο. Τα αποτελέσματα συνοψίζονται ως εξής:

A close up of a card

AI-generated content may be incorrect.

Η MAPE ανατινάσσεται (∞) επειδή ο παρονομαστής λαμβάνει τιμές μηδέν στο δείγμα, άρα είναι μη κατάλληλη για την παρούσα εφαρμογή. Συνεπώς, η σύγκριση πρέπει να στηριχθεί στις MAE και RMSE, που μετρούν σφάλματα σε απόλυτες μονάδες και παραμένουν σταθερές ως προς μηδενικές τιμές.

Με κριτήριο τις MAE και RMSE, η μέθοδος Backward υπερέχει οριακά (MAE: 323.74 vs 323.76, RMSE: 362.62 vs 362.69). Η μικρή, αλλά σταθερή υπεροχή της μεθόδου Backward fill οφείλεται στο ότι χρησιμοποιεί την επόμενη διαθέσιμη τιμή για να γεμίσει τα κενά, η οποία συνήθως είναι πιο κοντά στην «πραγματική» τιμή. Αντίθετα, το Forward fill κρατά την προηγούμενη τιμή και μπορεί να «παγώσει» παλιότερες πληροφορίες.

Στην παρούσα εφαρμογή, η πιο χρήσιμη μετρική είναι η **RMSE**. Ο λόγος είναι ότι, σε αντίθεση με τη MAE που μετρά μόνο το μέσο απόλυτο σφάλμα, η RMSE «τιμωρεί» περισσότερο τα μεγάλα λάθη επειδή υψώνει τα σφάλματα στο τετράγωνο. Έτσι, παρέχει μια πιο αυστηρή εικόνα για το πόσο καλά ή άσχημα αποδίδει η μέθοδος σε σημεία όπου τα κενά εμφανίζονται κοντά σε μεγάλες μεταβολές ή «σοκ» της χρονοσειράς. Η MAE παραμένει χρήσιμη λόγω της απλότητας και της άμεσης ερμηνείας της (σε μονάδες της χρονοσειράς), ωστόσο η RMSE είναι πιο ευαίσθητη και πρακτική όταν απαιτείται αξιολόγηση της σταθερότητας μιας μεθόδου σε πιο δύσκολες περιπτώσεις.

Συμπερασματικά, η καλύτερη μέθοδος ανάμεσα στις δοκιμασθείσες είναι η Backward fill, βάσει χαμηλότερων τιμών MAE και RMSE. Παρ’ όλα αυτά, το μέγεθος των σφαλμάτων και η αλλοίωση της τάσης υποδεικνύουν ότι για παραγωγική χρήση θα χρειαστεί πλουσιότερη και πιο εξελιγμένη μέθοδος.

**PROBLEM 4**

Υλοποιήθηκε και δοκιμάστηκε ένα σύνολο τριών προσαρμοσμένων μεθόδων παρεμβολής για την κάλυψη κενών σε ημερήσιες τιμές μετοχών:

1. Γραμμική παρεμβολή, η οποία εκτιμά ενδιάμεσες τιμές ως ευθύγραμμη σύνδεση μεταξύ των άκρων ενός κενού
2. Πολυωνυμική παρεμβολή χαμηλού βαθμού (2), που εισάγει ελεγχόμενη καμπυλότητα ώστε να αποτυπώνει μη γραμμικές μεταβολές χωρίς υπερβολικές ταλαντώσεις·
3. Απλή εκθετική εξομάλυνση , όπου η εκτιμημένη τιμή ορίζεται αναδρομικά, δίνοντας μεγαλύτερο βάρος στις πιο πρόσφατες παρατηρήσεις.

Η ποιότητα εκτίμησης μετρήθηκε ως εξής:

A white background with black text

AI-generated content may be incorrect.

Η MAPE κατέστη μη χρήσιμη (∞), επειδή λαμβάνει μηδενικές τίμες ο παρανομιστής συνεπώς η σύγκριση βασίστηκε σε MAE/RMSE. Επιπλέον, ελέγχθηκε η διατήρηση τάσης (slope) πριν/μετά την παρεμβολή, με μικρές αλλά υπαρκτές αποκλίσεις που υποδηλώνουν ήπια αλλοίωση του σήματος.

Η πολυωνυμική επιτυγχάνει το ελάχιστο MAE, άρα την καλύτερη μέση απόλυτη ακρίβεια σε όλο το δείγμα, ενώ η εκθετική εξομάλυνση πετυχαίνει το ελάχιστο RMSE, άρα ελαφρώς μικρότερα μεγάλα σφάλματα (καθώς το RMSE ποινικοποιεί δυσανάλογα τις μεγάλες αποκλίσεις). Δεδομένης της συνήθους προτεραιότητας στις χρηματοοικονομικές εφαρμογές να περιορίζονται τα μεγάλα λάθη, η βέλτιστη επιλογή είναι η Εκθετική Εξομάλυνση βάσει RMSE. Αν ο στόχος είναι ομοιόμορφη μέση ακρίβεια ανεξαρτήτως outliers, η Πολυωνυμική υπερέχει οριακά σε MAE. Σε κάθε περίπτωση, οι διαφορές είναι μικρές, η Γραμμική παραμένει στιβαρό baseline, και συνιστάται επιπλέον διερεύνηση της παραμέτρου α και του βαθμού του πολυωνύμου για βελτιστοποίηση σε παραγωγικό περιβάλλον.

**PROBLEM 5**

Στην ανάλυση που πραγματοποιήθηκε εμφανίστηκαν ορισμένα προβλήματα τα οποία σχετίζονται με τους δείκτες αξιολόγησης. Το πρώτο ζήτημα αφορά τον δείκτη MAPE, ο οποίος σε όλα τα αποτελέσματα εμφάνισε τιμή άπειρο. Αυτό συμβαίνει επειδή ο τύπος του MAPE περιλαμβάνει διαίρεση με τις πραγματικές τιμές και όταν αυτές είναι μηδενικές, η διαίρεση οδηγεί σε άπειρο. Για να αντιμετωπιστεί το πρόβλημα αυτό αντικαταστάθηκε ο MAPE με τον δείκτη sMAPE (Symmetric MAPE), ο οποίος χρησιμοποιεί στο παρονομαστή το άθροισμα των απόλυτων τιμών πραγματικών και εκτιμημένων δεδομένων. Με τον τρόπο αυτό αποφεύγεται η διαίρεση με το μηδέν και προκύπτουν σταθερές και ρεαλιστικές τιμές.

Ένα δεύτερο πρόβλημα που εντοπίστηκε ήταν η εμφάνιση NaN στον συντελεστή συσχέτισης. Η αιτία είναι ότι η αρχική χρονοσειρά παρουσιάζει μηδενική διακύμανση, καθώς η τάση της είναι επίπεδη. Σε αυτή την περίπτωση η τυπική απόκλιση είναι μηδέν και κατά συνέπεια ο συντελεστής συσχέτισης δεν μπορεί να υπολογιστεί, με αποτέλεσμα να προκύπτει NaN. Για να λυθεί το ζήτημα εφαρμόστηκε έλεγχος πριν τον υπολογισμό, έτσι ώστε όταν δεν υπάρχει διακύμανση να επιστρέφεται τιμή μη ορισμένη ή μηδέν αντί για NaN.

Τέλος, παρατηρήθηκε ότι ορισμένες μετρικές, όπως το correlation, τείνουν να δίνουν πολύ μεγάλες ή παραπλανητικές τιμές σε περιπτώσεις φτωχής διακύμανσης των δεδομένων. Αυτό καθιστά τους δείκτες αυτούς μη αξιόπιστους για την αξιολόγηση της ποιότητας της παρεμβολής. Για τον λόγο αυτό δόθηκε έμφαση σε πιο σταθερούς δείκτες, όπως ο MAE και ο sMAPE, οι οποίοι αποδείχθηκαν πιο ανθεκτικοί σε ακραίες περιπτώσεις.

Με τις παραπάνω διορθώσεις οι δείκτες αξιολόγησης σταθεροποιήθηκαν και τα αποτελέσματα είναι πλέον πιο αξιόπιστα και ερμηνεύσιμα.  
  
  
**PROBLEM 6**

Στο πρόβλημα 6 εφαρμόστηκε μέθοδος παρεμβολής με χρήση KNN (K-Nearest Neighbors) μέσω της κλάσης KNNImputer της βιβλιοθήκης scikit-learn. Η διαδικασία έγινε σε δύο εκδοχές: η πρώτη ήταν η απλή εκδοχή (basic KNN), όπου εφαρμόζεται απευθείας η μέθοδος χωρίς καμία κανονικοποίηση, ενώ η δεύτερη ήταν η εκδοχή με κλιμάκωση (scaled KNN), στην οποία εφαρμόστηκε πρώτα τυποποίηση με τον αλγόριθμο StandardScaler, ώστε να εξομαλυνθούν οι αποστάσεις μεταξύ των χαρακτηριστικών. Παράλληλα, έγινε και σύγκριση με μια απλή μέθοδο υποκατάστασης που βασίζεται στη διάμεσο (median imputation).

Συνολικά δοκιμάστηκαν διαφορετικές τιμές για τον αριθμό γειτόνων K, συγκεκριμένα K = 5, 10, 15 και 20. Σε όλες τις περιπτώσεις, τόσο η βασική όσο και η κλιμακωμένη εκδοχή του KNN κατάφεραν να συμπληρώσουν επιτυχώς όλες τις ελλείπουσες τιμές χωρίς να παραμείνει κανένα κενό. Ωστόσο, οι μετρικές αξιολόγησης έδειξαν ότι δεν υπήρχε ουσιαστική διαφοροποίηση μεταξύ των διαφορετικών τιμών του K, καθώς τα αποτελέσματα σε MAE και RMSE παρέμειναν σταθερά ανεξάρτητα από τον αριθμό των γειτόνων.

Πιο συγκεκριμένα, τα αποτελέσματα για τον δείκτη MAE έδειξαν ότι τόσο το basic όσο και το scaled KNN κατέληξαν στην ίδια μέση απόλυτη απόκλιση περίπου 92.5, ανεξαρτήτως K. Αντίθετα, η μέθοδος median imputation έδωσε χαμηλότερο MAE περίπου 77.4, επομένως φάνηκε να υπερτερεί. Αντίστοιχη εικόνα προέκυψε και από τον δείκτη RMSE, όπου το KNN εμφάνισε τιμές γύρω στις 185.1, ενώ η διάμεσος κατέγραψε μικρότερο σφάλμα περίπου 169.7. Η κλιμάκωση δεν οδήγησε σε βελτίωση, αφού τα αποτελέσματα scaled KNN ήταν ταυτόσημα με εκείνα του basic KNN.pasted-image.tiff

Από τα παραπάνω προκύπτει ότι, για το συγκεκριμένο dataset, το KNN imputation δεν παρουσίασε ουσιαστική βελτίωση σε σχέση με την απλούστερη μέθοδο της διάμεσου. Παρά το γεγονός ότι θεωρητικά η μέθοδος KNN μπορεί να εκμεταλλευτεί καλύτερα τη δομή των δεδομένων και να προσφέρει πιο αξιόπιστη συμπλήρωση, στην πράξη τα αποτελέσματα έδειξαν ότι η διάμεσος αποτελεί πιο αποδοτική λύση. Συνεπώς, η προτεινόμενη μέθοδος για την περίπτωση αυτή είναι η median imputation, καθώς προσφέρει μικρότερα σφάλματα και μεγαλύτερη αξιοπιστία.

**PROBLEM 7**

Οι συναρτήσεις χρησιμότητας αποτελούν βασικό εργαλείο στην οικονομική θεωρία, καθώς μοντελοποιούν τη στάση των επενδυτών απέναντι στον κίνδυνο και τον πλούτο. Στα κρυπτονομίσματα, όπου η μεταβλητότητα είναι ιδιαίτερα υψηλή, οι συναρτήσεις χρησιμότητας βοηθούν στην κατανόηση των διαφορετικών συμπεριφορών επενδυτών.

Τρεις απλές μορφές είναι:

* **Ριψοκίνδυνος επενδυτής:** U(W)=W2U(W)=W2. Κυνηγά υψηλές αποδόσεις ακόμη και με μεγάλο ρίσκο.
* **Αποστροφή στον κίνδυνο:** U(W)=WU(W)=W. Προτιμά σταθερότητα και αποφυγή απωλειών.
* **Ουδέτερος στον κίνδυνο:** U(W)=WU(W)=W. Επικεντρώνεται μόνο στην αναμενόμενη αξία, αδιαφορώντας για τη διακύμανση.

Η διαφοροποίηση είναι έντονη όταν εξετάσουμε τα χαρακτηριστικά των επενδυτών:

* **Νεότεροι** έχουν χρόνο να ανακάμψουν, άρα τείνουν σε πιο ριψοκίνδυνη στάση.
* **Ηλικιωμένοι** ενδιαφέρονται κυρίως για διατήρηση κεφαλαίου, επομένως η χρησιμότητά τους μοιάζει με U(W)=WU(W)=W.
* **Πλούσιοι** μπορούν να αναλάβουν μεγαλύτερο ρίσκο, καθώς οι απώλειες δεν απειλούν σοβαρά την ασφάλειά τους.
* **Φτωχότεροι** συχνά είναι πιο συντηρητικοί, αφού μια μεγάλη ζημιά μπορεί να έχει δραματικές συνέπειες.

Αυτό επηρεάζει και τον χειρισμό των **ελλιπών δεδομένων (missing data)**. Ένας επενδυτής με αποστροφή στον κίνδυνο θα αντιμετωπίσει τα κενά με απαισιόδοξες εκτιμήσεις (π.χ. χαμηλότερες αποδόσεις ή αυξημένη μεταβλητότητα), για να μην υποτιμήσει τον κίνδυνο. Αντίθετα, ένας ριψοκίνδυνος μπορεί να αγνοήσει τα κενά ή να τα συμπληρώσει με πιο θετικές υποθέσεις, θεωρώντας ότι η αγορά θα κινηθεί ανοδικά.

**Συγκεκριμένο παράδειγμα:** Ένας 28χρονος με μικρό κεφάλαιο επενδύει σε Bitcoin. Ακόμη και αν χάσει, θεωρεί ότι έχει χρόνο να ανακάμψει, άρα χρησιμοποιεί U(W)=W2U(W)=W2. Αν λείπουν δεδομένα για αποδόσεις συγκεκριμένων μηνών, μπορεί να τα αντικαταστήσει με μέσους όρους αγοράς, διατηρώντας αισιόδοξη στάση. Αντίθετα, μια 65χρονη επενδύτρια που επενδύει για τη σύνταξή της θα χρησιμοποιήσει U(W)=WU(W)=W. Αν λείπουν δεδομένα, πιθανόν να τα θεωρήσει ένδειξη αστάθειας και να αποφύγει εντελώς την τοποθέτηση.

Συμπερασματικά, οι συναρτήσεις χρησιμότητας έχουν νόημα στις επενδύσεις σε κρυπτονομίσματα, καθώς αποτυπώνουν πώς διαφορετικοί επενδυτές αξιολογούν το ρίσκο και πώς αντιμετωπίζουν την αβεβαιότητα.

**PROBLEM 8**

Στο όγδοο πρόβλημα πραγματοποιήθηκε η ενοποίηση δύο διαφορετικών συνόλων δεδομένων: το πρώτο περιείχε ημερήσια στοιχεία για κρυπτονομίσματα και το δεύτερο ημερήσια στοιχεία για μετοχές. Για να διασφαλιστεί η συγκρισιμότητα, επιλέχθηκε κοινό χρονικό εύρος, ώστε όλα τα assets να έχουν τις ίδιες ημερομηνίες παρατήρησης.

Κατά την ενοποίηση των δύο πινάκων εμφανίστηκαν ορισμένες ελλείπουσες τιμές, οι οποίες αντιμετωπίστηκαν με τη μέθοδο της διάμεσου (simple median imputation). Η συγκεκριμένη μέθοδος είχε ήδη αποδειχθεί η πιο αξιόπιστη στα προηγούμενα πειράματα, καθώς παρείχε τα χαμηλότερα σφάλματα σε σύγκριση με πιο σύνθετες τεχνικές. Έτσι διασφαλίστηκε η πληρότητα του τελικού πίνακα χωρίς να εισαχθούν σημαντικές στρεβλώσεις στα δεδομένα.

**PROBLEM 9**

Ο πίνακας εμφανίζει ενδεικτικά τις πρώτες γραμμές των αποδόσεων με στήλες Simple και Log για κάθε τίτλο∙ φαίνεται ότι στις μικρές ημερήσιες μεταβολές οι log αποδόσεις είναι σχεδόν ίσες με τις simple, ενώ σε μεγαλύτερα άλματα αποκλίνουν ελαφρά.

A screenshot of a computer code

AI-generated content may be incorrect.

Δημιουργήθηκε ο πίνακας που συνοψίζει μέση τιμή, τυπική απόκλιση και ασυμμετρία για αποδόσεις simple και log.  
Mean: Οι μέσες simple αποδόσεις εμφανίζονται υψηλότερες από τις αντίστοιχες log για τα ίδια περιουσιακά στοιχεία. Αυτό είναι αναμενόμενο, επειδή οι log αποδόσεις ενσωματώνουν τη σύνθετη κεφαλαιοποίηση και τείνουν να είναι ελαφρώς μικρότερες—η διαφορά φαίνεται περισσότερο όταν η μεταβλητότητα ή ο ορίζοντας είναι μεγάλοι.  
Std: Η μεταβλητότητα είναι γενικά παρόμοια μεταξύ simple και log για τα περισσότερα assets. Σε τίτλους με πολύ υψηλή μεταβλητότητα, οι simple αποδόσεις δίνουν ελαφρώς μεγαλύτερο std, καθώς μεγάλες ημερήσιες κινήσεις «φουσκώνουν» τη διασπορά.  
Skewness: Στις simple αποδόσεις ορισμένες μετοχές εμφανίζουν ισχυρή θετική ασυμμετρία (δεξιές ουρές). Οι log αποδόσεις είναι πιο συμμετρικές και «κοντά» στην κανονικότητα, με εξαίρεση το BTC που παρουσιάζει αρνητική ασυμμετρία (βαριές πτωτικές ουρές). Αυτό αποτελεί βασικό πλεονέκτημα των log returns για στατιστική μοντελοποίηση.  
Υπολογίστηκαν ξεχωριστοί πίνακες συσχέτισης για simple και log.  
Μετοχές μεταξύ τους: Καταγράφεται πολύ υψηλή θετική συσχέτιση, υποδηλώνοντας ισχυρό κοινό συστημικό παράγοντα και περιορισμένη διαφοροποίηση εντός του equity basket.  
Crypto vs Μετοχές: Τα BTC/ETH εμφανίζουν πολύ χαμηλές έως μηδενικές συσχετίσεις με τις μετοχές, άρα προσφέρουν δυνητική διαφοροποίηση σε μικτά χαρτοφυλάκια.  
Μεταξύ crypto: Η συσχέτιση BTC–ETH εμφανίζεται ασθενώς αρνητική στο συγκεκριμένο δείγμα· πρόκειται για ασυνήθιστο εύρημα που πιθανόν εξηγείται από το μέγεθος ή την ιδιαίτερη περίοδο του δείγματος.