

CASE 1 — PHÂN TÍCH BIẾN ĐỘNG GIÁ VÀNG GIỮA TP.HCM VÀ HÀ NỘI

(chỉ sử dụng thư viện NUMPY)

I. Đặt vấn đề

Anh/Chị là nhà phân tích dữ liệu tại một công ty tài chính. Công ty muốn theo dõi biến động giá vàng giữa hai khu vực TP.HCM và Hà Nội trong suốt 6 tháng để nhận diện mức chênh lệch giá theo thời gian. Dữ liệu gồm hai tệp CSV: `gold_price_hcm.csv` và `gold_price_hanoi.csv`, mỗi tệp có dạng: YYYY-MM-DD, price.

II. Mục tiêu

1. Vận dụng NumPy để xử lý dữ liệu dạng mảng 1 chiều hoặc 2 chiều.
2. Thực hành các thao tác nền tảng: slicing, indexing, reshaping, stacking.
3. Áp dụng boolean mask để lọc dữ liệu theo điều kiện.
4. Áp dụng fancy indexing để trích xuất dữ liệu theo danh sách vị trí.
5. Sử dụng broadcasting để chuẩn hoá hoặc tính toán giữa các mảng có kích thước khác nhau.
6. Tính toán các thống kê cơ bản bằng NumPy: mean, std, max, min, argmax, argmin.
7. Sử dụng `np.diff()` để phân tích biến động theo thời gian.
8. Tạo mảng tổng hợp kết quả và lưu ra file CSV bằng `np.savetxt()`.

III. Nhiệm vụ bắt buộc

1. Đọc `gold_price_hcm.csv` và `gold_price_hn.csv` bằng `np.loadtxt` hoặc `np.genfromtxt`.
2. Tách ngày và giá thành các mảng riêng.
3. Chuyển cột ngày sang `datetime64[D]`.
4. Kiểm tra dtype, shape, in 5 giá trị đầu và cuối.
5. Slicing: lấy 7 ngày đầu, 7 ngày giữa và 7 ngày cuối.
6. Boolean mask: tìm ngày TP.HCM cao hơn Hà Nội.
7. Fancy indexing: tìm top-5 ngày giá cao nhất mỗi khu vực.
8. Broadcasting: tính chênh lệch giá theo từng ngày.
9. Dùng `np.diff` để phân tích biến động ngày liền kề.
10. Tính mean, std, max, min, argmax, argmin.
11. Tạo bảng tổng hợp và lưu `ket_qua_case_1.csv` bằng `np.savetxt`.
12. (Tuỳ chọn) Vẽ biểu đồ line hoặc bar minh hoạ.

IV. Gợi ý kỹ thuật NumPy

Xem lại TLHT và nội dung các bài đã học và tìm hiểu mở rộng thêm trên internet, có thể sử dụng một số công cụ (hàm/kỹ thuật):

`np.loadtxt`, `np.genfromtxt`, `np.mean`, `np.std`, `np.diff`, `np.argmax`, `np.argmin`, structured array, slicing, boolean indexing, broadcast, axis=0/1.

V. Tập dữ liệu sử dụng

- `gold_price_hcm.csv`
- `gold_price_hanoi.csv`

VI. Sản phẩm nộp bài

- Báo cáo nhóm file .docx hoặc .pdf, ghi rõ nhật ký đóng góp nhóm (5 người)
- File code .ipynb
- Bộ CSV
- Các file được up lên repository: **Lớp_Nhóm_N_case1** trên github theo cấu trúc:

```
NhomN_Case_n/  
├── code_case_n.ipynb  
├── ket_qua_case_n.csv  
└── bao_cao_NhomX_case0n.pdf/docx
```

Trong đó:

- N: là số thứ tự của case study do nhóm được phân công

Lưu ý: *Trường hợp các bài sao chép hoặc sử dụng hoàn toàn bằng Ai sẽ bị trừ 50% điểm.*

VII. Thời gian hoàn thành

Nhóm 5 người — 05 ngày

Up bài lên Github và invite giảng viên trước **0 giờ ngày 26/11/2025**.

VIII. Thang điểm

4.0 — Xử lý NumPy

3.0 — Phân tích & thống kê

1.5 — Báo cáo

0.5 — Vẽ biểu đồ đơn giản

1.0 — Phỏng vấn nhóm

IX. Checklist

1. Đọc 2 file thành công
2. Tách đúng cột
3. Convert datetime
4. Slicing đúng 3 đoạn
5. Mask đúng điều kiện
6. Fancy-index đúng top-5
7. Broadcasting đúng
8. diff đúng
9. Thống kê đầy đủ
10. Lưu CSV thành công