

1. Thành viên nhóm

- Nguyễn Thị Thu Ngân
- Phạm Thị Hồng Như
- Trần Thu Nguyên
- Đinh Ngọc Dương
- Bùi Việt Phương

2. Phương pháp xử lý và phân tích dữ liệu

Thực hiện theo 5 nhiệm vụ được giao trong Case Study:

- Nhiệm vụ 1: Làm sạch dữ liệu từ 3 file CSV gốc (`stock_price.csv`, `stock_info.csv`, `transactions.csv`).
- Nhiệm vụ 2: Truy vấn và thống kê mô tả dữ liệu sạch.
- Nhiệm vụ 3: Phân nhóm và tổng hợp dữ liệu bằng `groupby`.
- Nhiệm vụ 4: Merge dữ liệu để tạo file hoàn chỉnh.
- Nhiệm vụ 5: Tạo bảng pivot và phân tích đa chiều.

3. Các bước thực hiện chính

Nhiệm vụ 1 – Làm sạch dữ liệu

- Chuẩn hóa mã cổ phiếu: viết hoa, xóa khoảng trắng.
- Xử lý cột `date`: chuyển về kiểu datetime, xử lý lỗi định dạng.
- Xử lý cột `close_price` và `volume`: loại bỏ ký tự chữ, chuyển về số.
- Chuẩn hóa cột `exchange` và `sector`: viết hoa, sửa lỗi chính tả.
- Xuất 3 file sạch: `stock_price_clean.csv`, `stock_info_clean.csv`, `transactions_clean.csv`.

KẾT QUẢ CUỐI CÙNG: Đã tạo 3 file sạch chứa 10 mã cổ phiếu chuẩn hóa, giá và khối lượng đã xử lý số, ngày tháng hợp lệ

Nhiệm vụ 2 – Truy vấn và thống kê

- Lọc mã cổ phiếu theo sàn HOSE.
- Đếm số lượng mã theo ngành.
- Tìm mã có giá đóng cửa bị thiếu.
- Tính khối lượng giao dịch trung bình theo mã.
- Tìm mã có trong `stock_info` nhưng không có trong `stock_price`.

KẾT QUẢ CUỐI CÙNG: Phát hiện 7/10 mã thiếu dữ liệu: VIC, VHM, FPT, MWG, VCB, BID, SSI không có trong stock_price

Nhiệm vụ 3 – GroupBy và tổng hợp

- Tính giá đóng cửa trung bình theo ngành.
- Tính khối lượng giao dịch trung bình theo sàn.
- Tìm mã có biến động giá mạnh nhất.
- Tổng hợp giá theo tháng.
- Xác định ngành có thanh khoản thấp nhất.

KẾT QUẢ CUỐI CÙNG: Ngành Bán Lẻ có thanh khoản thấp nhất với khối lượng TB 8,753

Nhiệm vụ 4 – Merge dữ liệu

- Merge `stock_info` với `stock_price` theo `ticker`.

- Merge thêm `transactions` để có tổng khối lượng giao dịch.
- Phát hiện các trường hợp mismatch (mã có giá nhưng không có thông tin).
- Xuất file `merged_full.csv`.

KẾT QUẢ CUỐI CÙNG: File hoàn chỉnh với 400 dòng, 6 cột, trong đó 280 dòng không có thông tin ngành/sàn

Nhiệm vụ 5 – Pivot Table và Stack/Unstack

- Tạo pivot table giá đóng cửa theo ngành và tháng.
- Tạo pivot table khối lượng giao dịch theo sàn và ngành.
- Thực hành chuyển đổi `stack()` và `unstack()`.
- Phân tích ngành tăng trưởng mạnh nhất

KẾT QUẢ CUỐI CÙNG: Ngành Bán Lẻ có tăng trưởng giá tốt nhất trong giai đoạn phân tích

4. Kết quả chính

- Dữ liệu đã được làm sạch và chuẩn hóa, sẵn sàng cho phân tích.
- Các bảng thống kê, tổng hợp đã được tạo ra đầy đủ.
- File dữ liệu hợp nhất `merged_full.csv` đã xuất thành công.
- Báo cáo này tổng hợp quá trình làm việc của nhóm.

5. Nhật ký nhóm

- Nguyễn Thị Thu Ngân:Nhiệm Vụ 4
- Phạm Thị Hồng Như:Nhiệm vụ 3
- Trần Thu Uyên:Nhiệm vụ 1
- Đinh Ngọc Dương:Nhiệm vụ 2
- Bùi Việt Phương:Nhiệm vụ 5

6. Kết luận

- Nhóm đã hoàn thành đầy đủ 5 nhiệm vụ của Case Study 02.
- Dữ liệu đã được làm sạch, tổng hợp và phân tích theo yêu cầu.
- Các kết quả đã được lưu trữ và báo cáo đầy đủ.
- Quá trình làm việc nhóm được thực hiện có phân công rõ ràng và đảm bảo tiến độ.