

LAB 3. PANDAS (Thời gian thực hiện: (90– 120 phút)

Cho 3 file .csv chứa trong thư mục DATA như sau:

- stocks1.csv : date, symbol, open, high, low, close, volume : chứa thông tin giao dịch chứng khoán các công ty khác nhau
- stocks2.csv : date, symbol, open, high, low, close, volume : chứa thông tin giao dịch chứng khoán các công ty khác nhau
- companies.csv : name, employees, headquarters_city, headquarters_state : chứa thông tin về trụ sở và số lượng nhân viên cho một công ty cụ thể.

LAB 3.1. KHÁM PHÁ DỮ LIỆU. (thời gian thực hiện: 15-20 phút.)

Mục tiêu: Nắm vững cách đọc và khám phá dữ liệu cơ bản từ file CSV.

Nhiệm Vụ:

1. Đọc file stocks1.csv vào DataFrame stocks1.
2. Hiển thị 5 dòng đầu tiên của stocks1.
3. Hiển thị kiểu dữ liệu (dtype) của mỗi cột trong stocks1.
4. Xem thông tin tổng quan (info) của stocks1.

LAB 3.2. XỬ LÝ DỮ LIỆU NULL (thời gian thực hiện: 15-20 phút.)

Mục tiêu: Học cách xác định và xử lý dữ liệu Null.

Nhiệm Vụ:

1. Kiểm tra xem trong stocks1 có dữ liệu Null nào không.
2. Thay thế dữ liệu Null ở cột high bằng giá trị trung bình của cột high.
3. Thay thế dữ liệu Null ở cột low bằng giá trị trung bình của cột low.
4. Hiển thị thông tin tổng quan để xác nhận không còn dữ liệu Null.

LAB 3.3. GỘP DỮ LIỆU VÀ PHÂN TÍCH CƠ BẢN (thời gian thực hiện: 15-20 phút.)

Mục tiêu: Nắm vững cách gộp dữ liệu và thực hiện phân tích cơ bản.

Nhiệm Vụ:

1. Đọc file stocks2.csv vào DataFrame stocks2.
2. Gộp stocks1 và stocks2 thành DataFrame mới tên là stocks.
3. Tính giá trung bình (open, high, low, close) cho mỗi ngày.
4. Hiển thị 5 dòng đầu tiên của kết quả.

LAB 3.4. KẾT HỢP DỮ LIỆU TỪ NHIỀU NGUỒN (thời gian thực hiện: 15-20 phút.)

Mục Tiêu: Nắm được cách kết hợp dữ liệu từ các file khác nhau và thực hiện phân tích cơ bản.

Nhiệm Vụ:

Đọc file companies.csv vào DataFrame companies.

Hiển thị 5 dòng đầu tiên của companies.

Kết hợp stocks (đã tạo từ bài 3) và companies dựa trên cột chung là symbol.

Tính giá đóng cửa (close) trung bình cho mỗi công ty.

Hiển thị kết quả cho 5 công ty đầu tiên.

LAB 3.5: SỬ DỤNG MULTIINDEX VÀ GROUPBY *(thời gian thực hiện: 15-20 phút.)*

Mục Tiêu: Hiểu cách sử dụng MultiIndex và GroupBy để phân tích dữ liệu một cách hiệu quả.

Nhiệm Vụ:

1. Tạo MultiIndex cho DataFrame stocks bằng cách sử dụng cột date và symbol làm chỉ mục.
2. Sử dụng GroupBy để tính giá trung bình (open, high, low, close) và volume trung bình cho mỗi ngày, cho mỗi mã chứng khoán.
3. Sắp xếp dữ liệu theo ngày và mã chứng khoán.
4. Hiển thị kết quả cho 5 ngày đầu tiên.

LAB 3.6: TẠO VÀ PHÂN TÍCH PIVOT TABLE *(thời gian thực hiện: 15-20 phút.)*

Mục Tiêu: Biết cách tạo và sử dụng Pivot Table để phân tích dữ liệu.

Nhiệm Vụ:

1. Tạo Pivot Table từ DataFrame stocks với date làm chỉ mục, symbol làm cột, và giá trị trung bình của close làm giá trị.
2. Thêm một cột tính tổng volume giao dịch cho mỗi mã chứng khoán (symbol).
3. Sắp xếp Pivot Table dựa trên tổng volume giao dịch, từ cao xuống thấp.
4. Hiển thị kết quả cho 5 mã chứng khoán có tổng volume giao dịch cao nhất.