1. Đề bài

Trong bài tập này, hãy giả sử bạn là 1 chuyên viên phân tích dữ liệu chứng khoán của FiinGroup. Công ty đang rất quan tâm đến mã cổ phiếu MWG trên thị trường. Bạn được giao nhiệm vụ đưa ra những phân tích chuyên sâu về mã cổ phiếu này để đưa ra những tư vấn cho việc liệu có nên đầu tư vào MWG hay không dựa vào tập dữ liệu về giá mở cửa, giá cao nhất, thấp nhất, đóng cửa, khối lượng giao dịch, mua chủ động, bán chủ động, giá trị bán khối ngoại, giá trị mua khối ngoại, giá trị mua/bán ròng. Nhiệm vụ của bạn là trả lời 10 câu hỏi sau cùng với 2 câu hỏi mở về việc đưa ra insight từ các phân tích.

Sau khi bạn hoàn thành bàu thi, hãy chụp và lưu **ảnh chụp màn hình** các kết quả cuối cùng cho bài làm (ví dụ: biểu đồ, bảng, kết quả tính toán, v.v.). Chúng sẽ cần được chia sẻ trong phần Đánh giá ngang hàng sau đây của mô-đun Dự án cuối cùng.

1. Về bộ dữ liệu

|  |  |  |
| --- | --- | --- |
| Tên thuộc tính | Mô tả | Kiểu dữ liệu |
| ticker | Tên mã | str |
| timestamp | Thời gian giao dịch | str |
| open | Giá mở cửa | float |
| low | Giá thấp nhất | float |
| high | Giá cao nhất | float |
| close | Giá đóng cửa | float |
| volume | Khối lượng giao dịch | int |
| bu | Khối lượng mua chủ động | int |
| sd | Khối lượng bán chủ động | int |
| fb | Giá trị mua khối ngoại | int |
| fs | Giá trị bán khối ngoại | int |
| fn | Giá trị mua/bán ròng khối ngoại | int |

1. Câu hỏi
2. Phân tích và trực quan dữ liệu cơ bản (3 điểm)

Câu 1: Khám phá dữ liệu cơ bản (1 điểm)

Nhiệm vụ:

Tải data theo đường link đính kèm và hiển thị:

* Bảng biểu của tập dữ liệu
* Các ngày có trong tập dữ liệu (sắp xếp theo thứ tự tăng dần)
* Tìm ngày có biên độ dao động giá (High - Low) lớn nhất

Câu 2: So sánh hành vi phiên đầu tuần và cuối tuần (1 điểm)

Nhiệm vụ:

* Tính % thay đổi giá đóng cửa vào ngày thứ Hai và thứ Sáu mỗi tuần.
* So sánh phân phối biến động giữa hai ngày này (dùng boxplot hoặc histogram).

Gợi ý: sử dụng datetime.weekday() để lọc.

Câu 3: Phân tích khoảng gap mở cửa (1 điểm)

Nhiệm vụ:

* Tính khoảng chênh lệch giữa giá mở cửa hôm nay và giá đóng cửa hôm qua (gọi là gap).
* Phân loại gap: “Gap Up”, “Gap Down”, “No Gap” dựa trên ngưỡng ±1%.
* Thống kê tỷ lệ từng loại gap.

1. Phân tích nâng cao (4 điểm)

Câu 4: Tính biến động theo ATR (1 điểm)

Nhiệm vụ:

* Tính chỉ báo Average True Range (ATR) với cửa sổ 14 ngày.
* Vẽ biểu đồ ATR và giải thích ngắn gọn ý nghĩa.
* Highlight các giai đoạn ATR tăng vọt (gợi ý: ATR > percentile 90).

Câu 5: Xác định các phiên breakout mạnh (1 điểm)

Nhiệm vụ:

* Phiên breakout mạnh là phiên có giá đóng cửa vượt 10% so với đỉnh gần nhất trong 30 phiên trước đó.
* Tìm và hiển thị các phiên breakout đó kèm giá trị Volume tương ứng.
* Trực quan bằng biểu đồ đánh dấu breakout.

Câu 6: Chiến lược theo vùng giá tích lũy (1 điểm)

Nhiệm vụ:

* Xác định vùng giá tích lũy: biên độ dao động trong ±3% kéo dài tối thiểu 10 phiên.
* Mua khi giá vượt ra khỏi vùng tích lũy với khối lượng tăng ít nhất 50%.
* Mua khi giá vượt ra khỏi vùng tích lũy với khối lượng tăng ít nhất 50%.

Câu 7: Chỉ báo MFI (Money Flow Index) (1 điểm)

Nhiệm vụ:

* Mua khi giá vượt ra khỏi vùng tích lũy với khối lượng tăng ít nhất 50%.
* Vẽ MFI cùng biểu đồ giá.
* Đánh dấu các vùng MFI > 80 và < 20.

1. Phân tích kỹ thuật chuyên sâu (2 điểm)

Câu 8: Phân tích ảnh hưởng của dòng tiền tự doanh tới xu hướng giá (1 điểm)

Nhiệm vụ:

* Tính hệ số tương quan giữa giá trị mua/bán ròng tự doanh với giá đóng cửa và % thay đổi giá.
* Vẽ scatter plot giữa giá trị mua ròng tự doanh và giá đóng cửa hôm sau.
* Có thể dùng shift() để kiểm tra hiệu ứng trễ.

Câu 9: Đánh giá hiệu suất của chiến lược đơn giản theo khối lượng (0.5 điểm)

Nhiệm vụ:

* Chiến lược: Mua khi khối lượng tăng > 2 lần so với trung bình 10 phiên, giá đóng cửa tăng > 2%.
* Bán sau 3 phiên.
* Tính tổng lợi nhuận, số lệnh, tỷ lệ thắng/lỗ.

Câu 10: Phát hiện điểm dị biệt bằng KMeans (0.5 điểm)

Nhiệm vụ:

* Áp dụng KMeans clustering với 3 cụm lên tập dữ liệu (sử dụng các đặc trưng như % thay đổi giá, khối lượng, giá trị giao dịch...).
* Xác định cụm “dị biệt” dựa trên khoảng cách trung tâm cụm.
* Vẽ biểu đồ phân cụm theo thời gian và nêu nhận định.

1. Đưa ra insight (1 điểm)

Câu 11: Đánh giá chiến lược breakout trong thời điểm có tin tức gây sốc (1 điểm)

* Lấy giai đoạn có sự kiện tin tức lớn (VD: cổ phiếu bị loại khỏi ETF, có giao dịch nội gián, kết quả lỗ quý bất ngờ...).
* Áp dụng lại chiến lược breakout ở câu 5 trong giai đoạn đó.
* Đánh giá ưu và nhược điểm khi chiến lược hoạt động trong môi trường biến động cực mạnh.