

Chương 8

Ví dụ về một Prompt Toàn Diện theo ACTOR để Làm Sạch Dữ Liệu

Hãy cùng xem cách áp dụng khung ACTOR vào tình huống làm sạch tệp dữ liệu giao dịch khách hàng (`DuLieu_Loi_KhaoSat.csv`) đã nêu ở các phần trước.

A - Actor (Vai trò): "Bạn hãy đóng vai một chuyên gia phân tích dữ liệu, có kinh nghiệm sâu sắc trong việc làm sạch và chuẩn bị **dữ liệu kinh doanh** cho phần mềm SPSS. Bạn rất thành thạo các kỹ thuật trong Excel và có khả năng giải thích các giải pháp một cách rõ ràng, dễ hiểu cho người mới bắt đầu."

C - Context (Bối cảnh): "Tôi đang làm sạch một tệp dữ liệu từ Excel có tên là `DuLieu_Loi_KhaoSat.csv`. Đây là dữ liệu giao dịch của một cửa hàng trực tuyến với **N=150**. Sau khi kiểm tra sơ bộ, tôi đã phát hiện các vấn đề chính sau:

1. **Cột `HinhThucThanhToan` (biến định danh):** Mã hóa không nhất quán. Có các giá trị như 'The Tin Dung', 'the', 'COD', 'Tiền mặt', 'Vi Dien Tu', 'Momo'. Tôi muốn chuẩn hóa tất cả về 3 loại chính là 'COD', 'Thẻ tín dụng', và 'Ví điện tử'.
2. **Cột `TongTien` (biến tỷ lệ):** Có các giá trị ngoại lai bất hợp lý, bao gồm một giá trị âm (-250000) và một giá trị rất lớn so với phần còn lại (50000000).
3. **Cột `NgayMua` (biến ngày tháng):** Bị lỗi định dạng. Định dạng chuẩn cần là DD/MM/YYYY nhưng trong tệp lại có cả định dạng YYYY-MM-DD và dạng văn bản như 5 Tháng 8 2025.
4. **Cột `ThoiGianGiaoHang` (biến tỷ lệ):** Có một số ô bị bỏ trống (missing data)."

T - Task (Nhiệm vụ):

"Dựa trên bối cảnh trên, tôi yêu cầu bạn hỗ trợ tôi thực hiện tuần tự các tác vụ sau:

1. Với cột `HinhThucThanhToan`, hãy gợi ý một phương pháp hiệu quả trong Excel để chuẩn hóa các giá trị không nhất quán về các mã chuẩn đã định trước.
2. Với cột `TongTien`, hãy giải thích cách tiếp cận để xử lý cả hai loại giá trị ngoại lai (âm và rất lớn). Hãy đề xuất phương án xử lý tốt nhất cho từng trường hợp và giải thích lý do.

3. Với cột **NgayMua**, hãy cung cấp các bước cụ thể trong Excel để tôi có thể xác định và chuyển đổi tất cả các định dạng khác nhau về một định dạng duy nhất là DD/MM/YYYY.
4. Với cột **ThoiGianGiaoHang**, hãy nêu ra các phương án để xử lý các ô trống (ví dụ: xóa bỏ dòng, thay thế bằng giá trị trung bình/trung vị) và cho biết phương án nào có thể là tốt nhất trong trường hợp này."

O - Objective (Mục tiêu):

"Mục tiêu cuối cùng của tôi là có một bộ dữ liệu đã được làm sạch, nhất quán và đáng tin cậy, sẵn sàng để nhập vào phần mềm SPSS mà không gặp lỗi. Đồng thời, tôi cũng muốn hiểu rõ cơ sở lý luận đằng sau các quyết định làm sạch dữ liệu mà mình đã thực hiện."

R - Requirements (Yêu cầu cụ thể về đầu ra):

"1. Vui lòng trình bày giải pháp cho từng vấn đề một cách riêng biệt, có đánh số thứ tự tương ứng với các nhiệm vụ ở phần Task. 2. Đối với các gợi ý sử dụng công thức hoặc tính năng của Excel, hãy giải thích ngắn gọn ý nghĩa của chúng. 3. Ngôn ngữ giải thích cần rõ ràng, mạch lạc, tập trung vào tính ứng dụng và dễ hiểu cho người không có nền tảng kỹ thuật sâu. 4. Câu trả lời của bạn nên tập trung vào các giải pháp thực hiện được trong môi trường Excel, vì đây là bước chuẩn bị trước khi đưa dữ liệu vào SPSS."