



- THỰC TẬP ĐỒ ÁN CHUYÊN NGÀNH -  
TÌM HIỂU VỀ PHÂN TÍCH DỮ LIỆU  
(DATA ANALYTICS)

Giáo viên hướng dẫn: TS. Nguyễn Bảo Ân

Sinh viên thực hiện: Cô Nhân Quý - MSSV: 110122150

Tổng quan và Quy trình

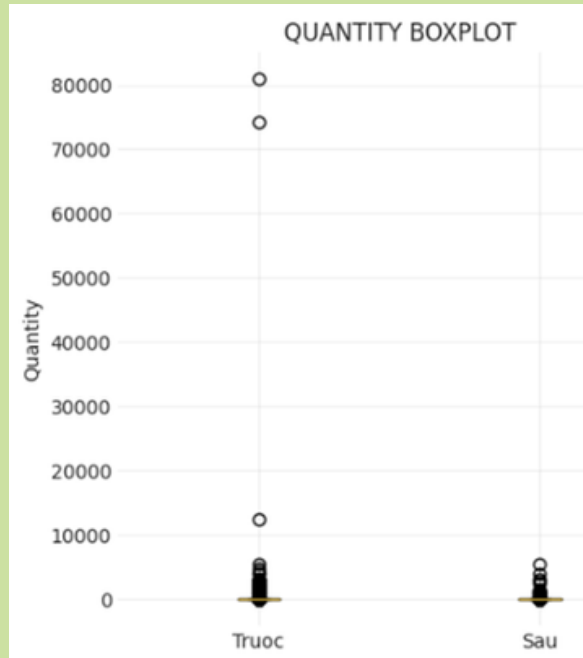
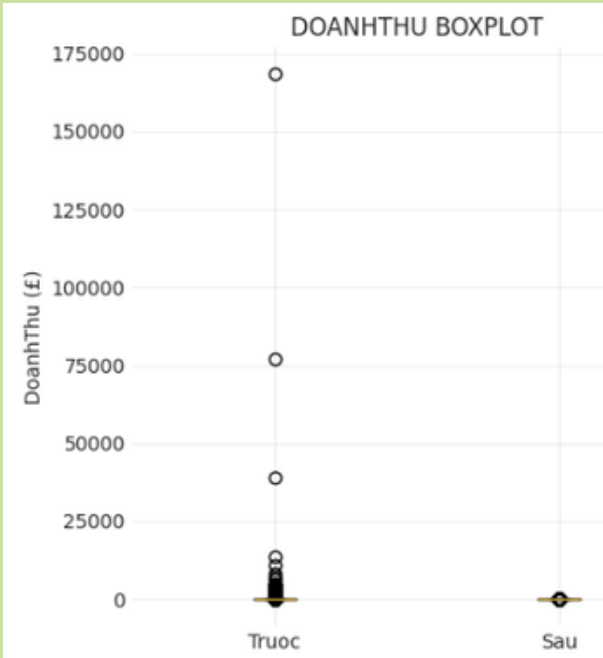
**Mục tiêu:** Áp dụng quy trình phân tích dữ liệu chuẩn trên bộ dữ liệu Online Retail để đánh giá doanh thu và phân khúc khách hàng.

**Công cụ:** Python, Pandas, Matplotlib, Streamlit.

**Quy trình:**

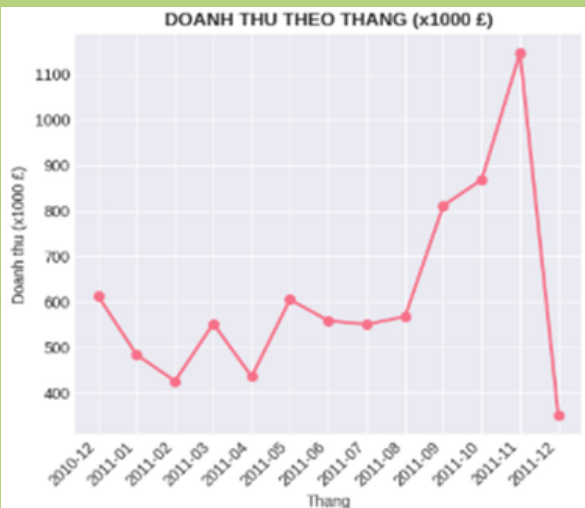
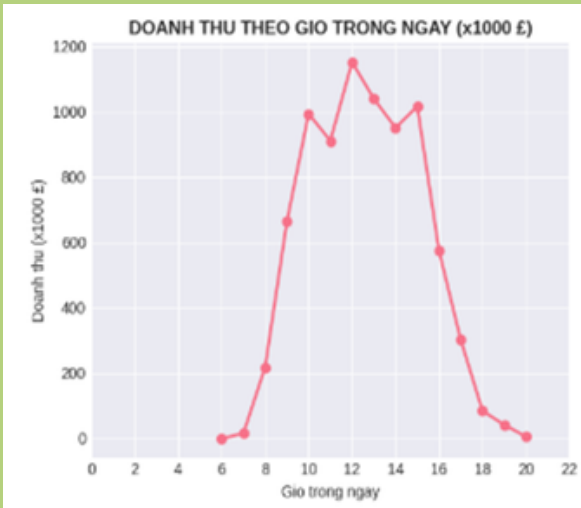
Thu thập dữ liệu - Làm sạch & Tiền xử lý - Phân tích thống kê - Mô hình RFM - Trực quan hóa Dashboard.

Kết quả làm sạch dữ liệu



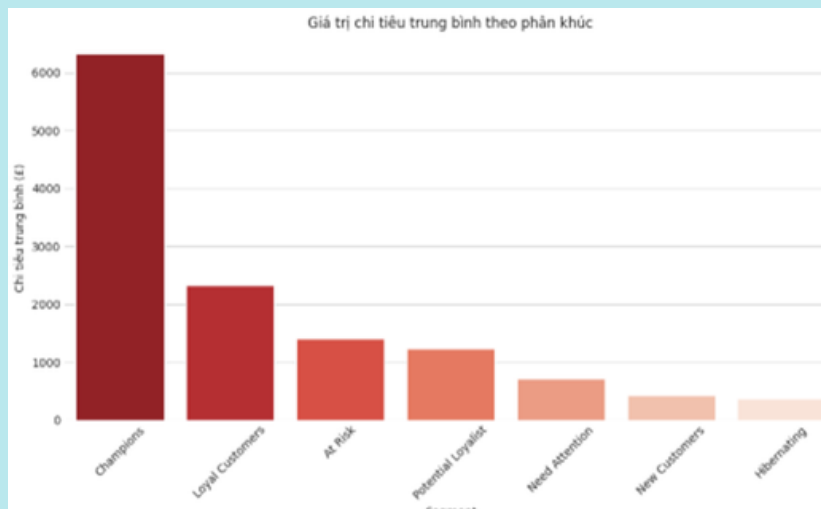
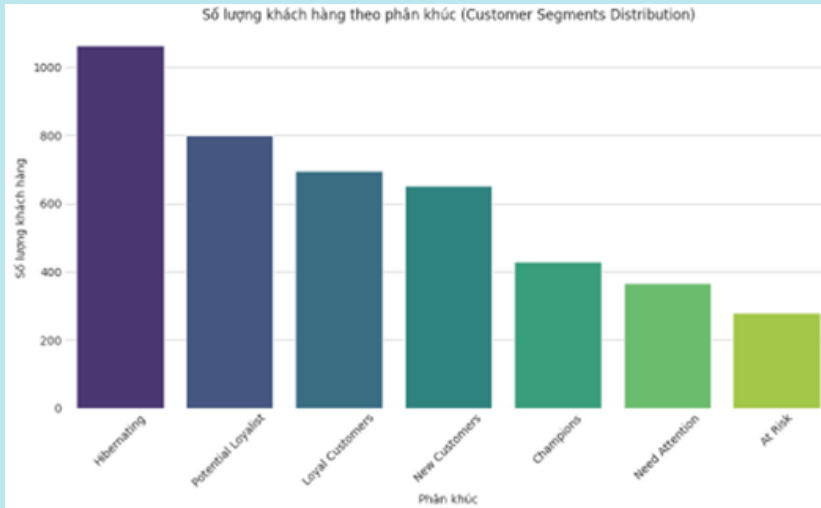
Loại bỏ nhiễu, giảm độ lệch chuẩn >85%, giữ lại 97% dữ liệu sạch

Xu hướng kinh doanh



Đỉnh điểm doanh thu tháng 11, "Khung giờ vàng" mua sắm 10h - 15h.

Phân khúc khách hàng

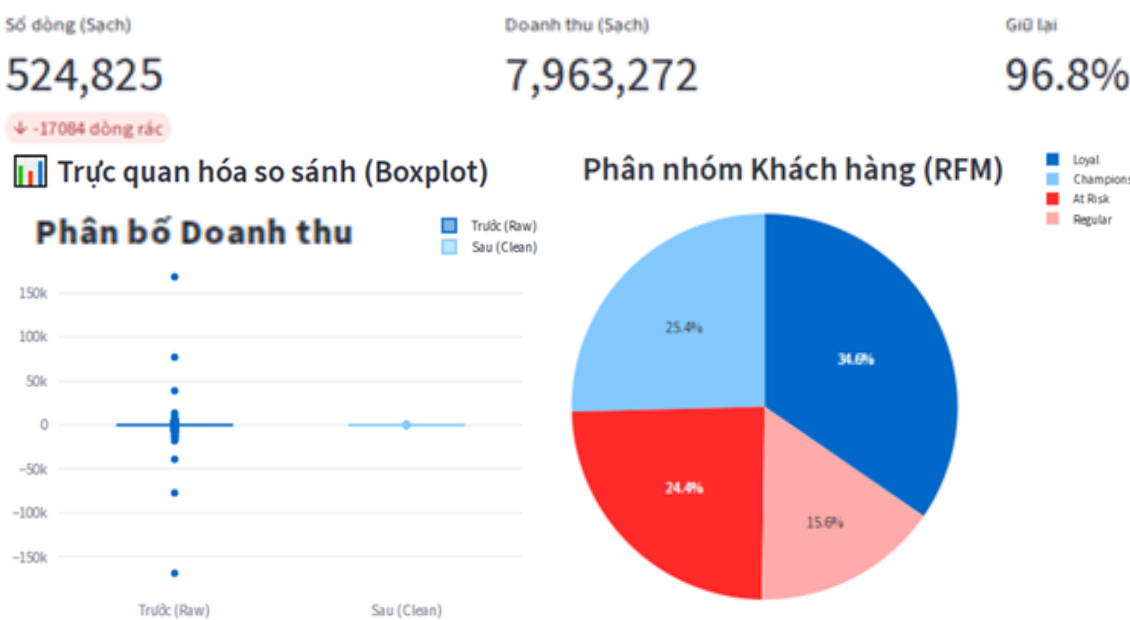


Kết luận

**Kết quả đạt được:**

- Xây dựng thành công quy trình xử lý dữ liệu tự động, giúp giảm độ lệch chuẩn >85% và giữ lại 97% dữ liệu sạch.
- Hoàn thiện ứng dụng Dashboard (Streamlit) trực quan hóa xu hướng doanh thu và phân khúc khách hàng theo thời gian thực.

Hiệu quả làm sạch & So sánh Thống kê



Hướng phát triển

- Nâng cấp công nghệ:** Ứng dụng Machine Learning (K-Means Clustering) để phân nhóm khách hàng tự động và các thuật toán dự báo để dự đoán doanh thu.
- Tối ưu hóa:** Xây dựng hệ thống Gợi ý sản phẩm (Recommendation System) và tự động hóa các kịch bản Marketing theo từng phân khúc RFM.