

# Customer Shopping Behavior Analysis

## 1. Giới thiệu dự án (Project Overview)

Dự án tập trung phân tích hành vi mua sắm của khách hàng thông qua bộ dữ liệu gồm 3.900 giao dịch của một công ty bán lẻ. Mục tiêu là hiểu rõ các yếu tố ảnh hưởng đến doanh thu, tần suất mua, phân khúc khách hàng, mức độ sử dụng ưu đãi và xu hướng theo mùa.

Câu hỏi kinh doanh chính cần trả lời:

**“Doanh nghiệp có thể tận dụng dữ liệu hành vi mua sắm của khách hàng như thế nào để nhận diện xu hướng, cải thiện tương tác khách hàng và tối ưu chiến lược marketing – sản phẩm?”**

Hai dashboard trực quan đã được xây dựng để hỗ trợ phân tích toàn diện.

Cụ thể:

- Xác định phân khúc khách hàng mang lại doanh thu cao nhất.
- Kiểm tra mức độ phụ thuộc vào khuyến mãi.
- Phân tích hiệu quả danh mục sản phẩm.
- Xác định xu hướng theo mùa và độ tuổi để hỗ trợ marketing.
- Tìm ra các vấn đề rủi ro như khách hàng mới thấp, churn cao,...

## 2. Tổng quan dataset

- **Số dòng:** 3.900
- **Số cột:** 18
- **Nhóm thông tin chính:**
  - **Nhân khẩu học:** Tuổi, Giới tính, Khu vực, Tình trạng đăng ký
  - **Thông tin giao dịch:** Sản phẩm, Danh mục, Giá trị đơn hàng, Mùa, Kích cỡ, Màu sắc
  - **Hành vi mua sắm:** Giảm giá, Promo Code, Số lần mua trước, Tần suất mua, Đánh giá, Hình thức vận chuyển
- **Dữ liệu thiếu:** 37 giá trị thiếu ở cột Review Rating → đã xử lý bằng median theo từng danh mục sản phẩm.

## Phương pháp sử dụng

- Descriptive Analytics: mô tả dữ liệu hiện tại (doanh thu, đơn hàng, số khách).
- Diagnostic Analytics: tìm nguyên nhân phía sau sự khác biệt (vì sao Loyal mua nhiều? vì sao New thấp?).
- Segment Analysis / Customer Profiling (RFM hoặc phân nhóm loyalty).
- Category Analysis: phân tích hiệu quả danh mục (Clothing, Accessories...).
- Time-series exploration: phân mùa.
- Discount sensitivity analysis: kiểm tra mức độ ảnh hưởng khuyến mãi.

### **Chiều phân tích (dimensions)**

- Khách hàng: Loyalty level, Age group, Gender.
- Sản phẩm: Category, Rating.
- Đơn hàng: Quantity, Price, Discount, Payment method.
- Thời gian: Season

## **3. Chuẩn bị & làm sạch dữ liệu (Python)**

### **3.1 Khám phá ban đầu**

- Dùng df.info(), df.describe() để xem cấu trúc, kiểu dữ liệu, phân phối.

### **3.2 Xử lý dữ liệu thiếu**

- Điền giá trị thiếu ở Rating bằng median theo Category.

### **3.3 Chuẩn hóa dữ liệu**

- Đổi tên cột sang dạng snake\_case
- Chuẩn hóa các giá trị dạng text (ví dụ: Male ↔ male)

### **3.4 Feature Engineering**

- Tạo cột **age\_group** theo độ tuổi
- Tạo cột **purchase\_frequency\_days** từ timestamp
- Loại bỏ **promo\_code\_used** vì trùng chức năng với **discount\_applied**

### **3.5 Tích hợp vào SQL Server**

- Kết nối Python → SQL Server

- Đẩy dữ liệu sạch vào database để dùng cho SQL và dashboard

#### 4. DashBoard



Figure 1: Page 1



Figure 2: Page 2

## **5. Insights và hành động đề xuất**

**Insight:** Khách hàng thân thiết (Loyal) đóng góp gần như toàn bộ doanh thu chứng tỏ nếu nhóm này giảm thì doanh thu sẽ giảm theo

**Quyết định:** triển khai các chương trình duy trì dành cho các khách hàng trung thành (như free ship, giảm giá, tích điểm và đổi quà)

**Insight:** Số lượng khách hàng mới rất thấp dẫn đến khó tăng trưởng.

**Quyết định:** đưa ra các ưu đãi mua lần đầu, hoặc chương trình giới thiệu bạn bè và người nhà đến mua được nhận thêm ưu đãi.

**Insight:** Clothing là nguồn doanh thu & giao dịch chính. Đây là core product nên phải ưu tiên danh mục này

**Quyết định:** Khi chạy các chương trình marketing thì nên ưu tiên danh mục này

**Insight:** Quá nhiều khách hàng phụ thuộc vào khuyến mãi, khuyến mãi có thể thúc đẩy người dùng mua thường xuyên hơn nhưng có thể làm giảm doanh số

**Quyết định:** Thay vì tập trung vào giảm giá thì có thể tặng các dịch vụ khác như miễn phí trả hàng hoặc có thể phân phát mã giảm giá dựa theo mức độ trung thành cả khách hàng thay vì phân phát cho tất cả.

**Insight:** nhóm young Adults và mùa spring/ fall có cơ hội tăng trưởng ngắn hạn. Young adult có doanh thu và số đơn cao nhất, và mùa sản phẩm được tiêu thụ mạnh là vào spring và fall.

**Quyết định:** Có thể chạy các chương trình quảng cáo tập trung vào nhóm Young Adults vào mùa Spring và Fall. Ngoài ra có thể đưa ra các chiến dịch thúc đẩy mua sắm hơn (Như khuyến mãi hoặc giảm giá) vào 2 mùa chủ chốt này