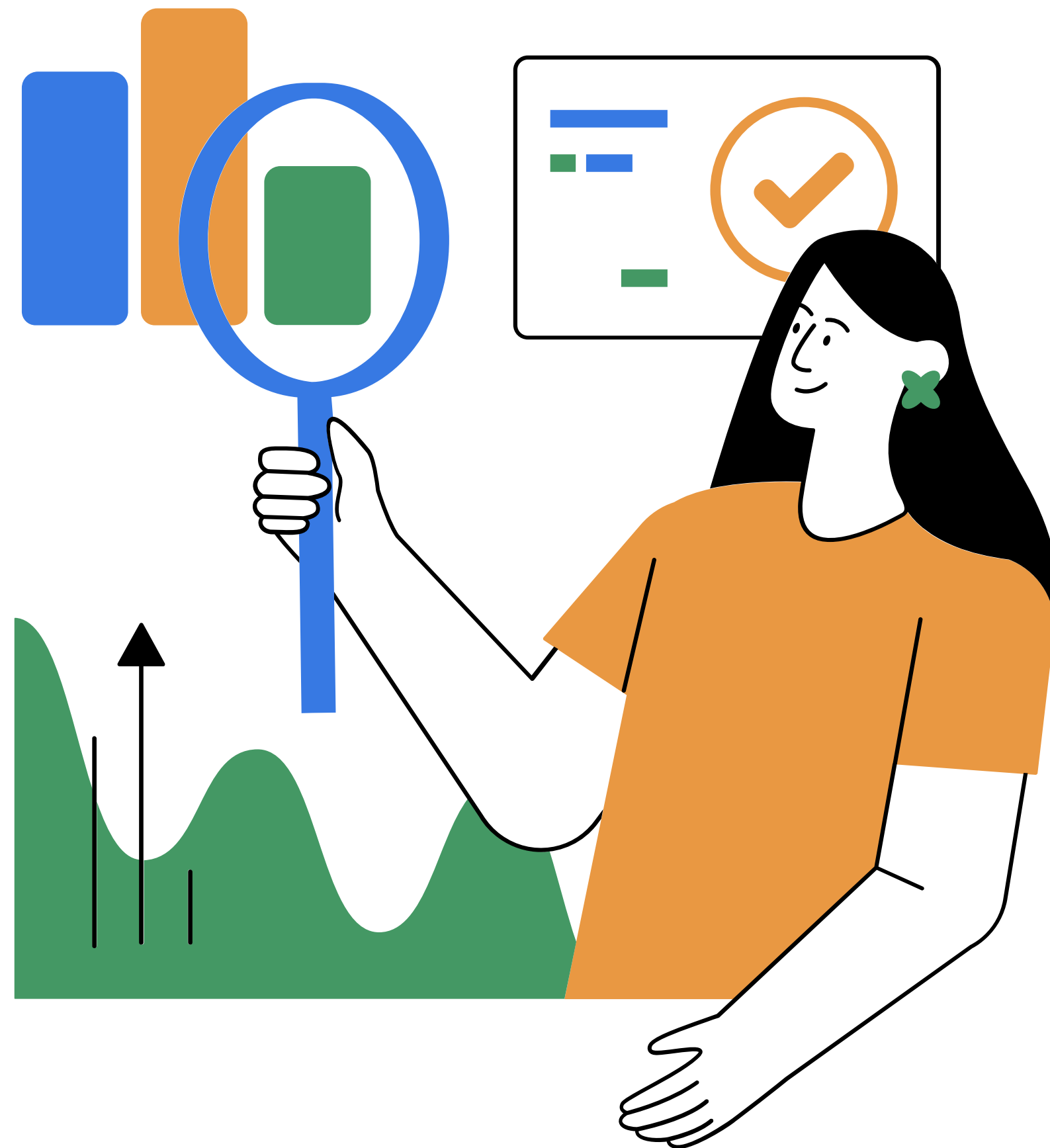
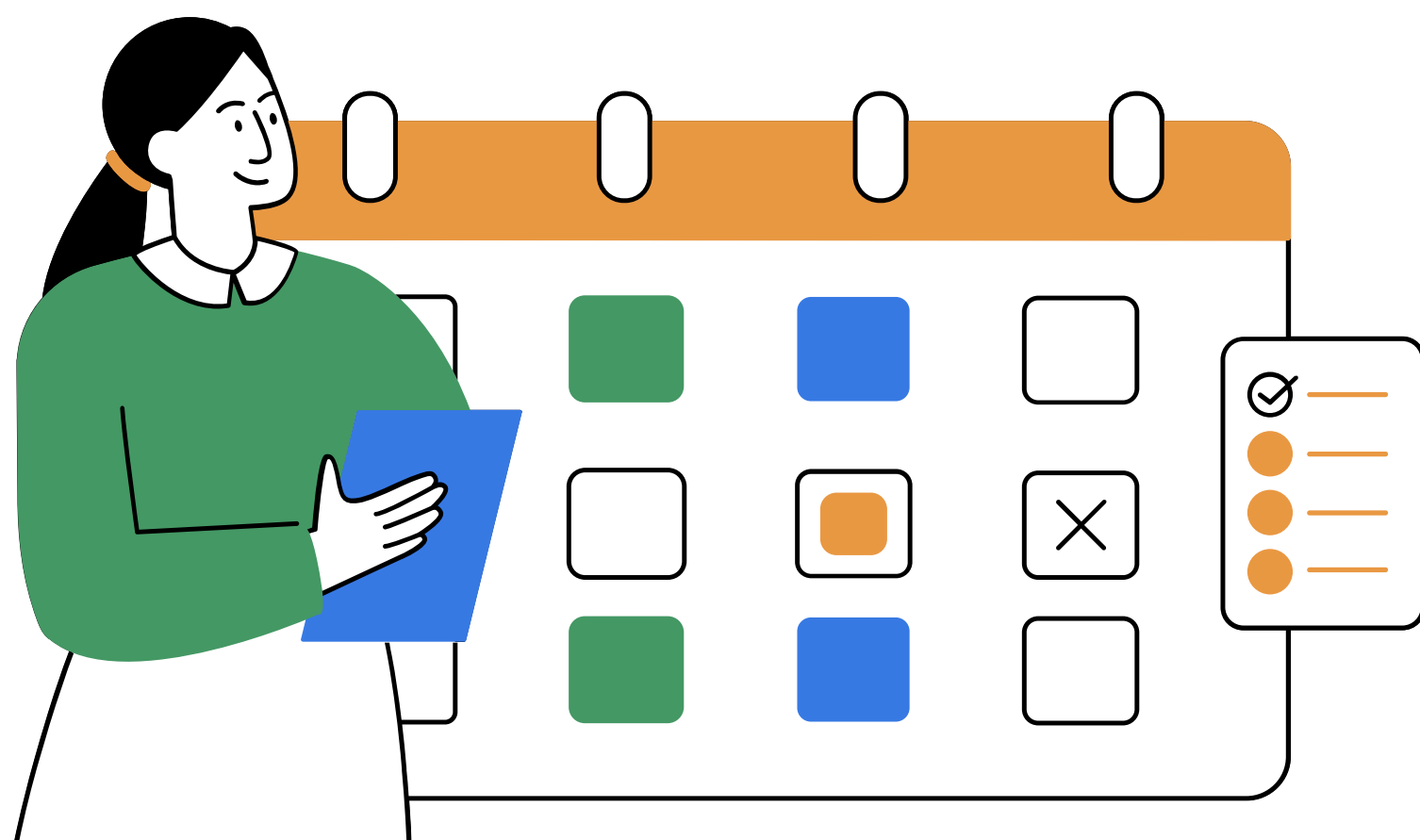


MARKET BASKET ANALYSIS

PHÂN TÍCH GIỎ HÀNG





Mục lục

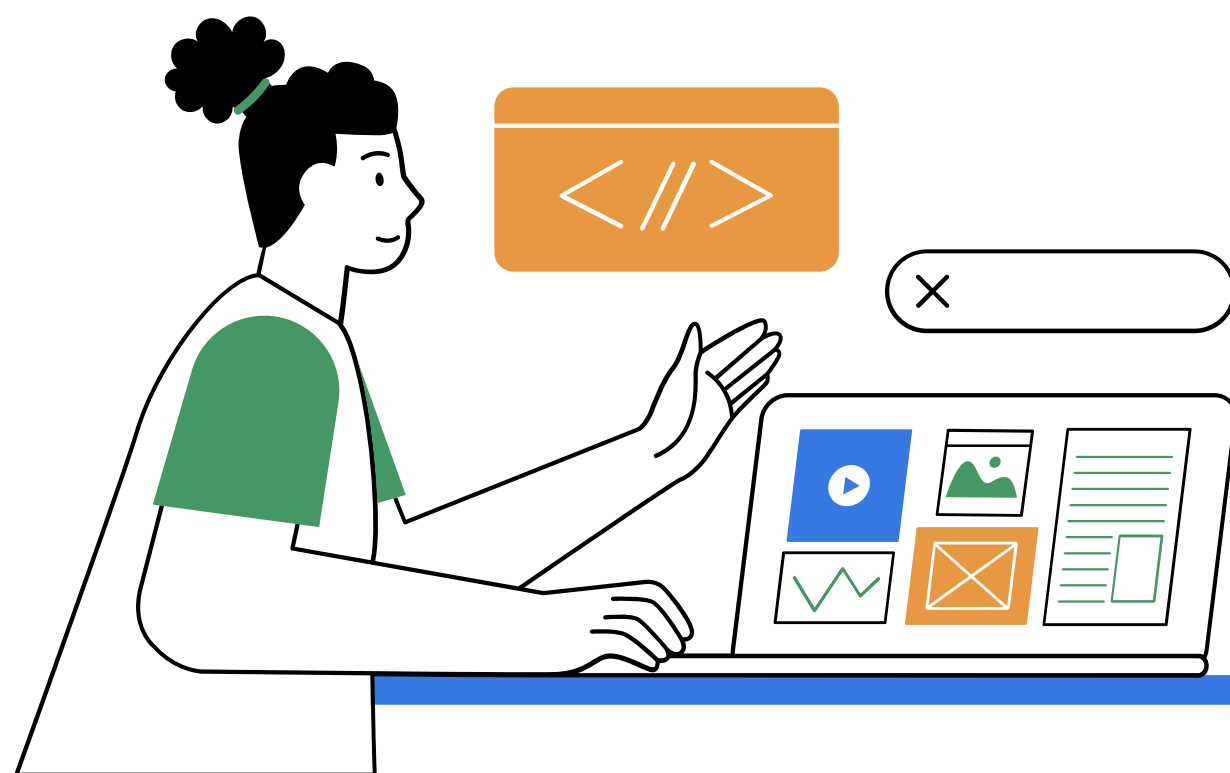
- Tóm tắt điều hành
- Tổng quan dữ liệu
- Phương pháp
- Lựa chọn tham số
- Kết quả
- Insight & Khuyến nghị
- Hạn chế & Rủi ro
- Hướng phát triển



Tóm tắt điều hành

Mục tiêu dự án là khai thác tập phổ biến và luật kết hợp từ dữ liệu bán lẻ để xác định nhóm sản phẩm thường được mua cùng, phục vụ đề xuất cross-sell/upsell, thiết kế combo và tối ưu bố trí kệ hàng. Do tỷ lệ giao dịch chứa ≥ 2 sản phẩm thấp (4,070/294,188), mô hình sử dụng ngưỡng support và confidence thấp, đánh giá hiệu quả qua chỉ số lift.

Tổng quan dữ liệu



- Số dòng: 302,010 | Số cột: 30
- Cột giao dịch (Transaction):
Transaction_ID
- Cột mặt hàng (Item):
Product_Category
- Tổng số giao dịch: 294,188
- Giao dịch có ≥ 2 mặt hàng: 4,070
- Số nhóm hàng (unique): 5

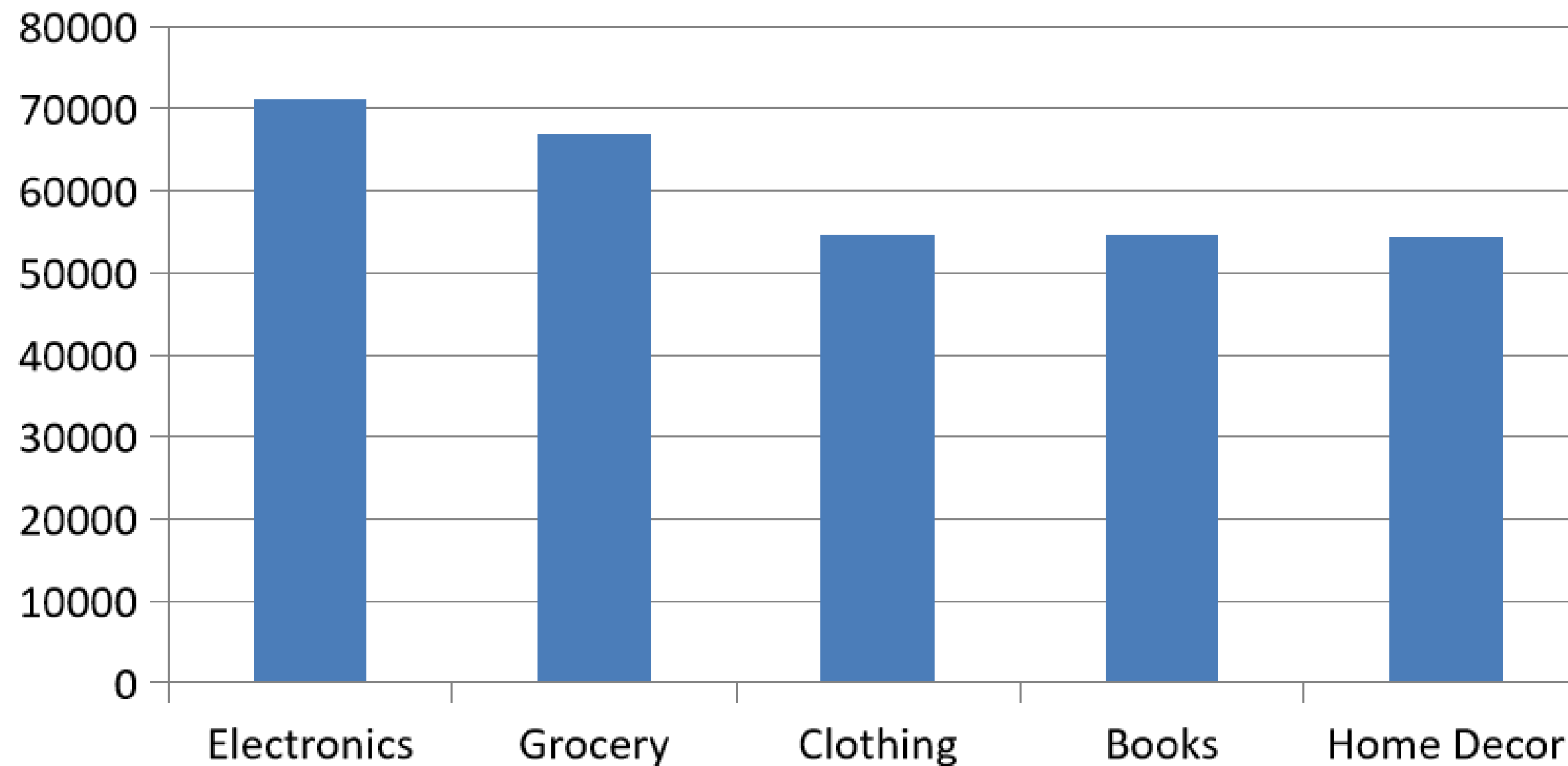
- Danh sách cột (rút gọn):
- Transaction_ID
- Customer_ID
- Name
- Email
- Phone
- Address
- City
- State
- +....

DATASET:

<https://www.kaggle.com/datasets/bhavikjikadara/retail-transactional-dataset/data>

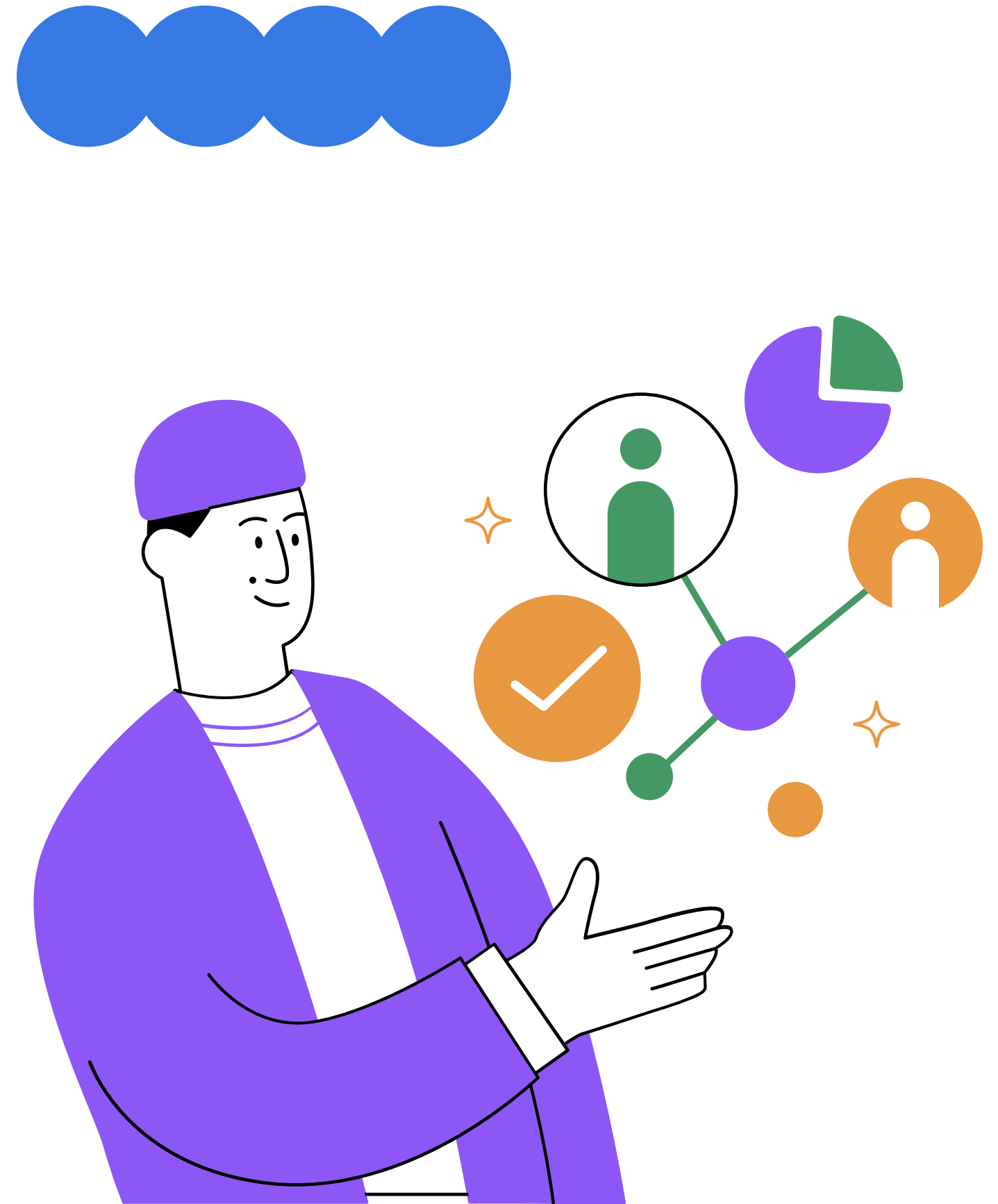


Top nhóm hàng (Product_Category)

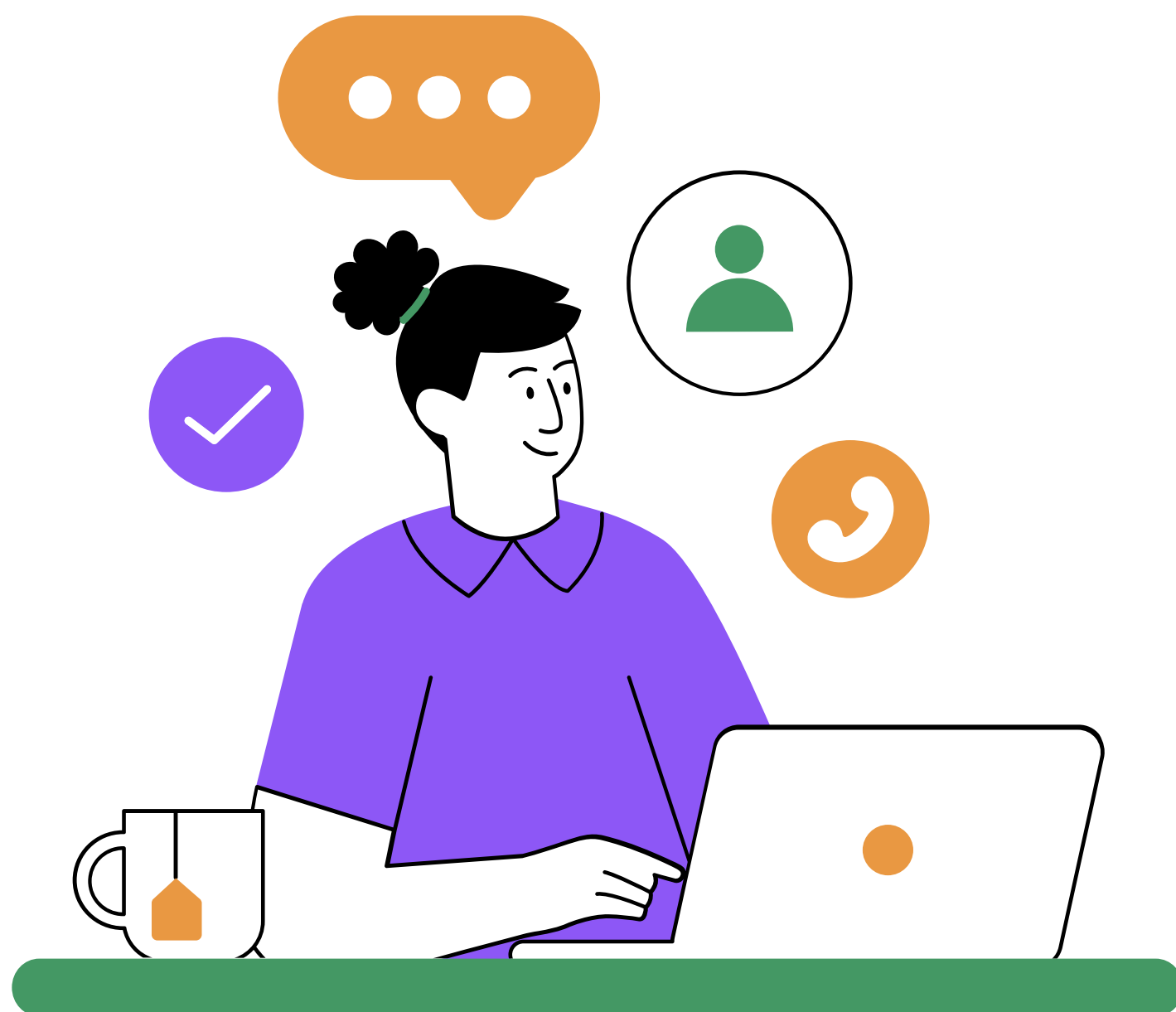


Phương pháp

- Tạo giỏ hàng: nhóm theo Transaction_ID → tập các Product_Category duy nhất/hoá đơn.
- Mã hoá one-hot bằng TransactionEncoder
- Apriori (ID lẻ): duyệt theo mức, sinh ứng viên + cắt tỉa theo tính chất kháng đơn điệu.
- FP-Growth (ID chẵn): nén dữ liệu bằng FP-tree, khai thác không cần sinh ứng viên.
- FP-Max (ID chẵn): tìm tập phổ biến tối đại (không có siêu tập nào cũng phổ biến).
- Association Rules: tính support, confidence, lift (thêm leverage, conviction).



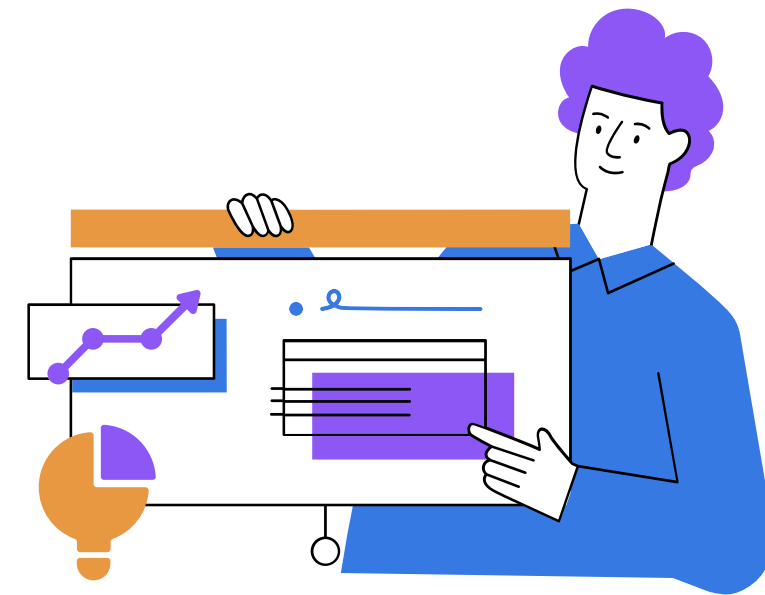
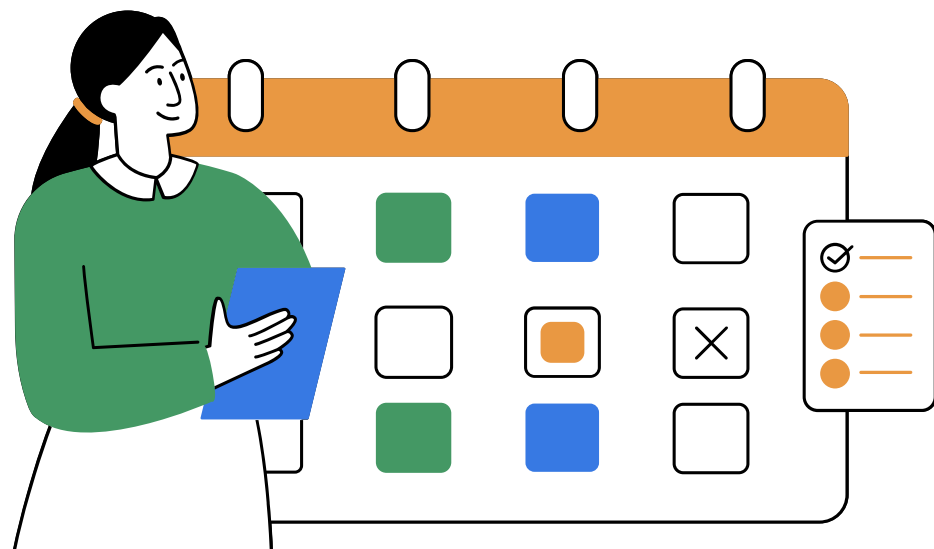
Lựa chọn tham số



- min_support: $k/Nk/N$, với $N=294,188N = 294,188$ hoặc $N=4,070$ (giỏ ≥ 2 món).
- Confidence: Quét $0.8 \rightarrow 0.1$ (có thể xuống 0.01 nếu dữ liệu thừa).
- Lọc: Lift ≥ 1.0 để ưu tiên luật có sức nâng
- Mục tiêu 50–300 luật
- Ưu tiên: Cấu hình sát biên dưới để luật gọn, dễ diễn giải.

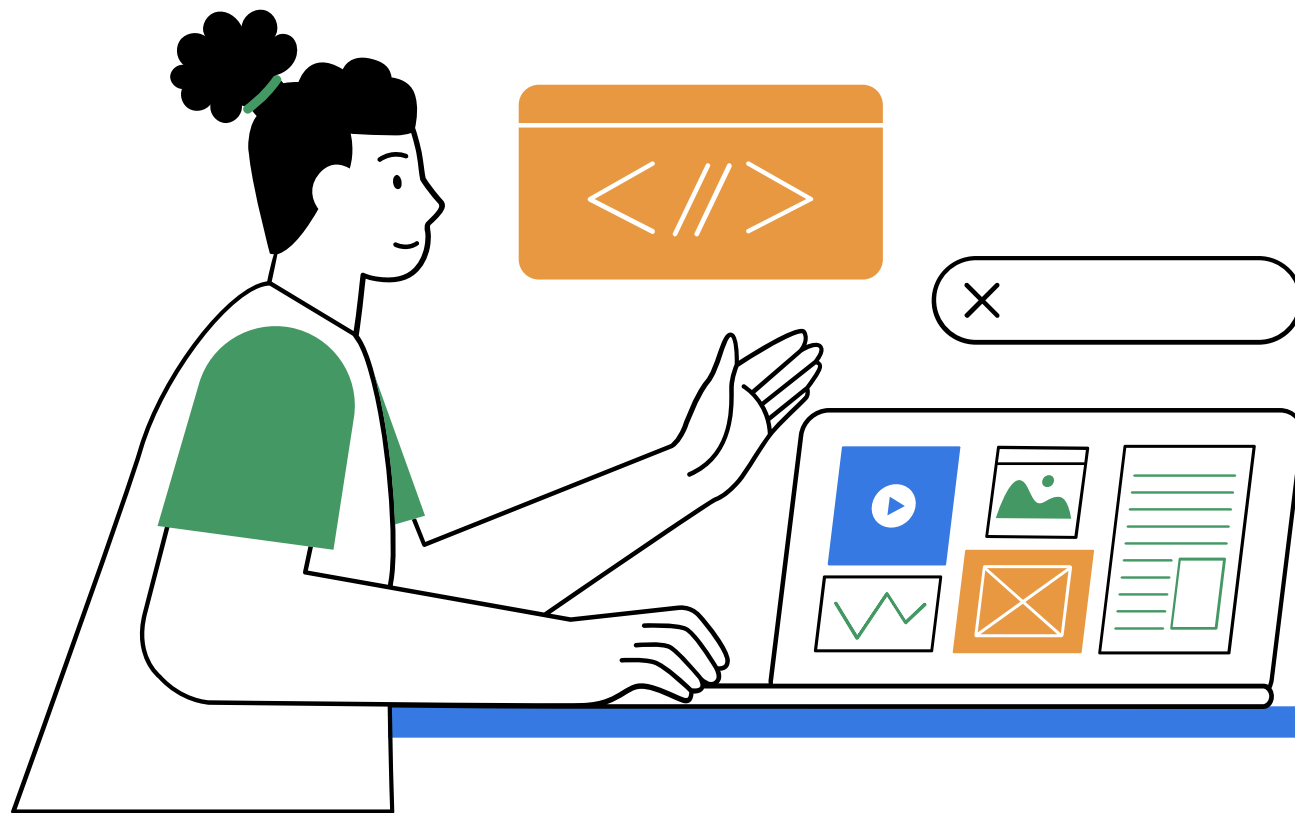
Kết quả thí nghiệm (CSV)

- frequent_itemsets.csv → thống kê số lượng theo độ dài (1, 2, 3...).
- association_rules.csv → sắp xếp theo lift, confidence; chọn Top-10 luật.
- (ID chẵn)
frequent_itemsets_maximal.csv → các tập tối đại (FP-Max).



- Mẫu bảng để paste:
antecedents → consequents | support |
confidence | lift
VD: {PARTY, PAPER CUP} → BALLOON |
0.012 | 0.38 | 1.26

Insight & khuyến nghị

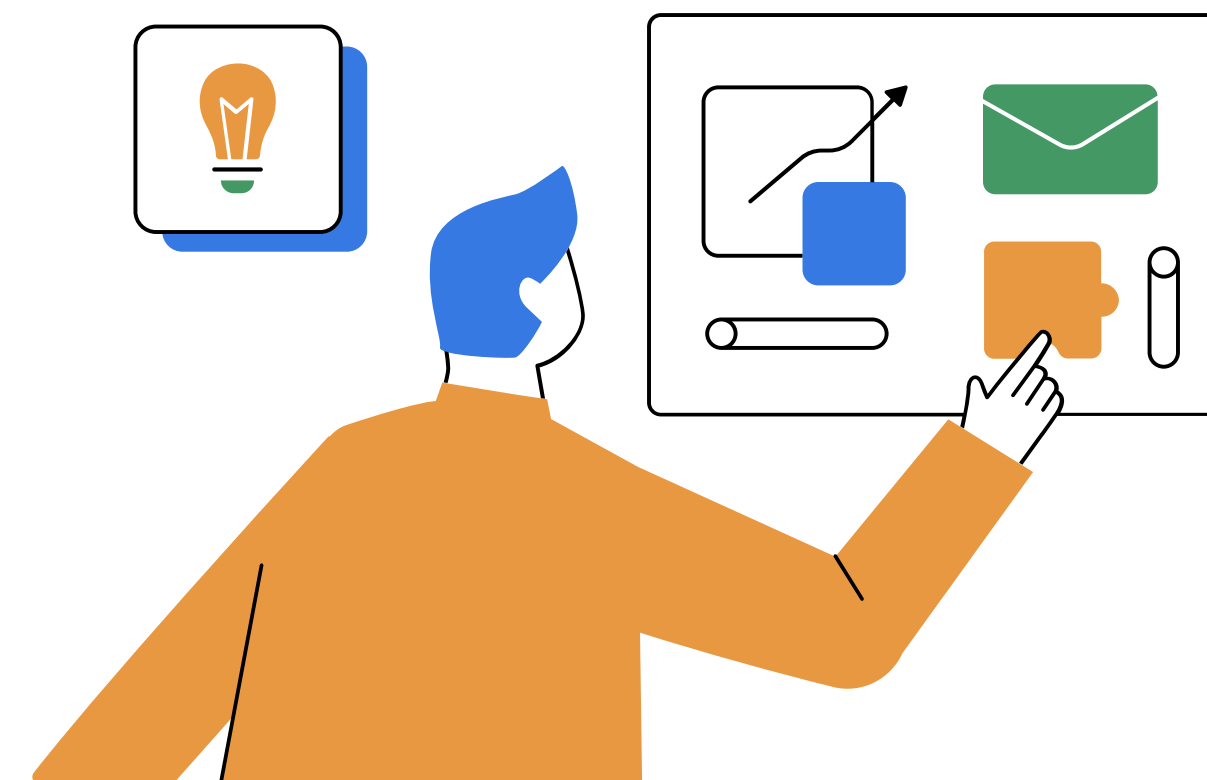


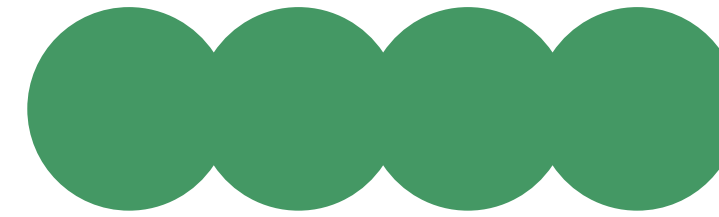
- Combo & Cross-sell: Dựa trên luật lift cao để đề xuất tại trang checkout.
- Bố trí kệ: Đặt gần nhau các nhóm hàng thường mua cùng.
- Khuyến mại: Tạo gói 2–3 nhóm hàng liên kết mạnh.



Hạn chế & Rủi ro

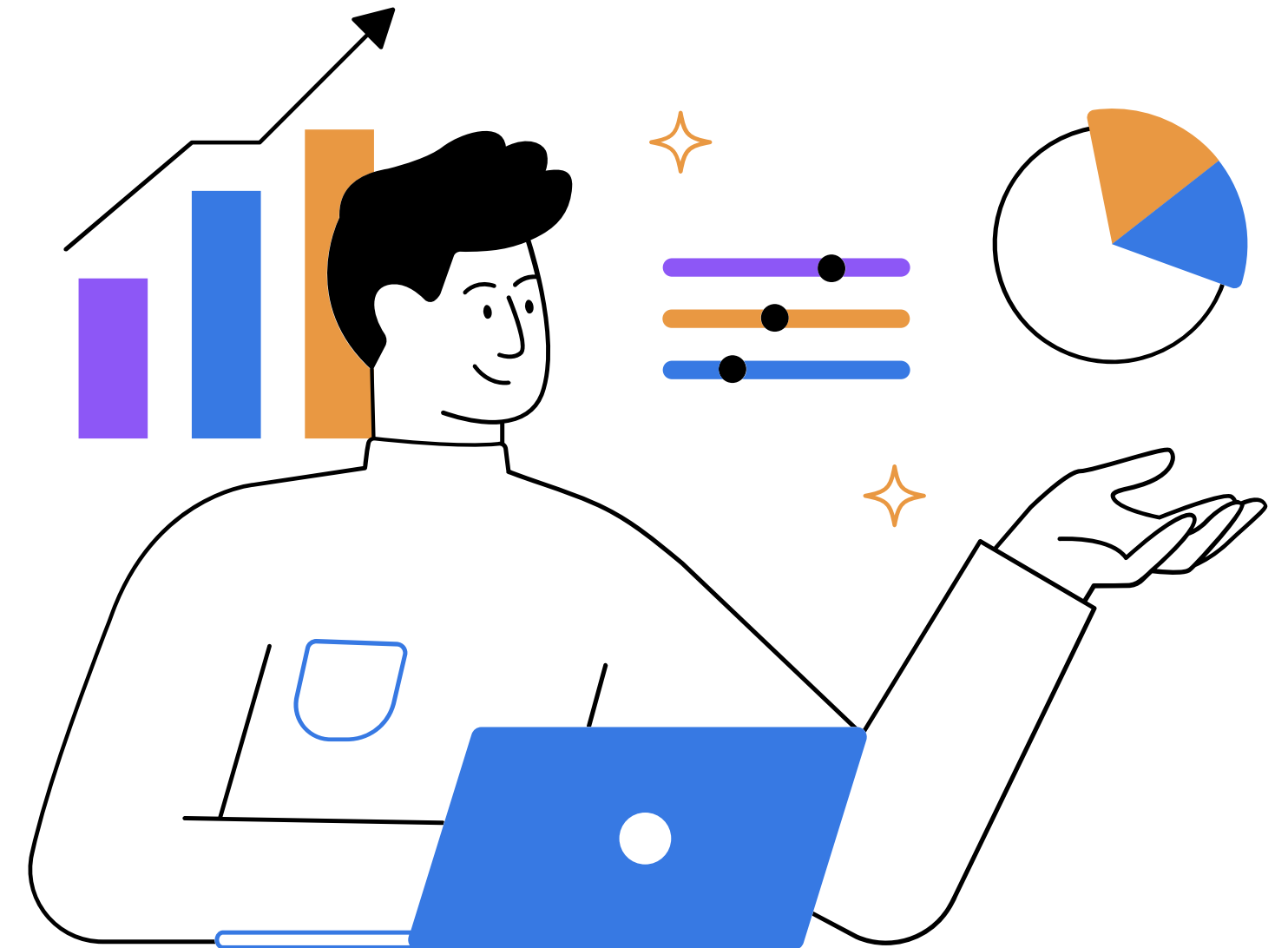
- Dữ liệu thưa: chỉ 4,070/294,188 giao dịch có ≥ 2 món; chỉ 5 nhóm hàng.
- Chưa xét mùa vụ/khuyến mãi/địa lý; rule mining không thể hiện nhân-quả.
- Cần kiểm tra độ ổn định luật theo thời gian / theo phân khúc khách hàng.





Hướng phát triển?

- Gom giỏ theo Customer + Date để tăng số item/giỏ.
- So sánh Apriori vs FP-Growth về thời gian & chất lượng luật.
- Thử thêm thuộc tính Brand/Type vào item thay vì chỉ Category.
- Triển khai thử nghiệm A/B cho gợi ý combo.



XIN CẢM ƠN THẦY/CÔ VÀ CÁC BẠN ĐÃ LẮNG NGHE!

Hỏi & Đáp (Q&A)

