

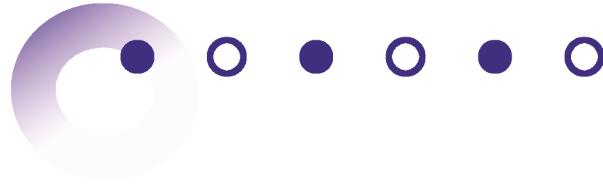


Báo cáo dự án

Winner GR Analytics

Le Van Duy





CONTENTS

01 Mở đầu dự án

02 Bối cảnh dữ liệu

03 Kiến trúc dự án

04 Khám phá dữ liệu

05 Vấn đề & Cơ hội





01

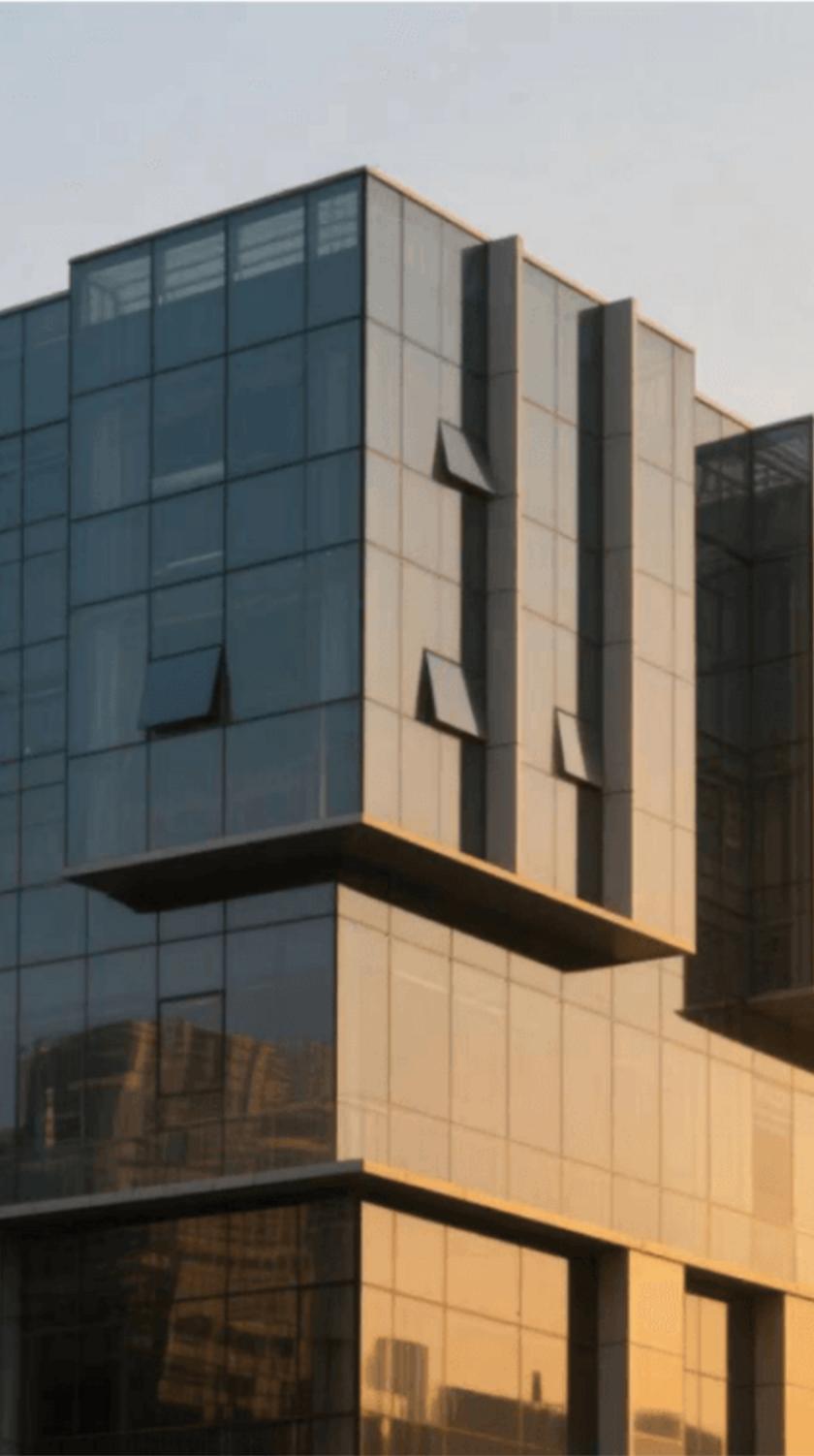
Mở đầu dự án



Winner Group Analytics là gì?

Mục đích của dự án

Winner Group Analytics được xây dựng để chuyển đổi dữ liệu phân tán từ Pancake POS thành những insights hành động, hỗ trợ quyết định kinh doanh dựa trên dữ liệu và tăng hiệu quả vận hành cũng như doanh thu cho doanh nghiệp thời trang online tại Hà Nội.



Vai trò của dự án

Dự án này không chỉ giúp doanh nghiệp hiểu rõ tình hình kinh doanh mà còn tạo nền tảng cho phân tích nâng cao như Machine Learning và dự báo xu hướng, từ đó mang lại lợi thế cạnh tranh bền vững.

Vấn đề cần giải quyết

Dữ liệu phân tán

Dữ liệu 4 năm của doanh nghiệp chưa được khai thác và quản lý tập trung, khiến việc phân tích và ra quyết định trở nên khó khăn.

Thiếu hệ thống phân tích

Doanh nghiệp không có hệ thống phân tích tập trung, dẫn đến việc không thể phát hiện các pattern và đo lường KPI một cách hệ thống.

Trải nghiệm khách hàng kém

Với tỷ lệ hoàn thành đơn hàng thấp và tỷ lệ trả hàng cao, trải nghiệm khách hàng kém, ảnh hưởng đến doanh thu và uy tín của doanh nghiệp.



02

Bối cảnh dữ liệu





Quy mô dữ liệu 4 năm

Tổng quan dữ liệu

Từ 2021 đến 2025, doanh nghiệp ghi nhận 40.236 đơn hàng, 46.611 sản phẩm đã bán, phục vụ 36.090 khách hàng với 37 mã hàng, tổng doanh thu đạt khoảng 18 tỷ đồng.

Chất lượng dữ liệu

Dữ liệu được lưu trữ qua Pancake POS API dạng JSON, với tỷ lệ missing values dưới 2%, đảm bảo độ hoàn chỉnh cao, sẵn sàng cho quá trình làm sạch và phân tích.

Cơ hội khai thác

Với lượng dữ liệu lớn và chất lượng tốt, doanh nghiệp có cơ sở để khai thác sâu, phát hiện xu hướng và insight, từ đó đưa ra chiến lược kinh doanh hiệu quả.

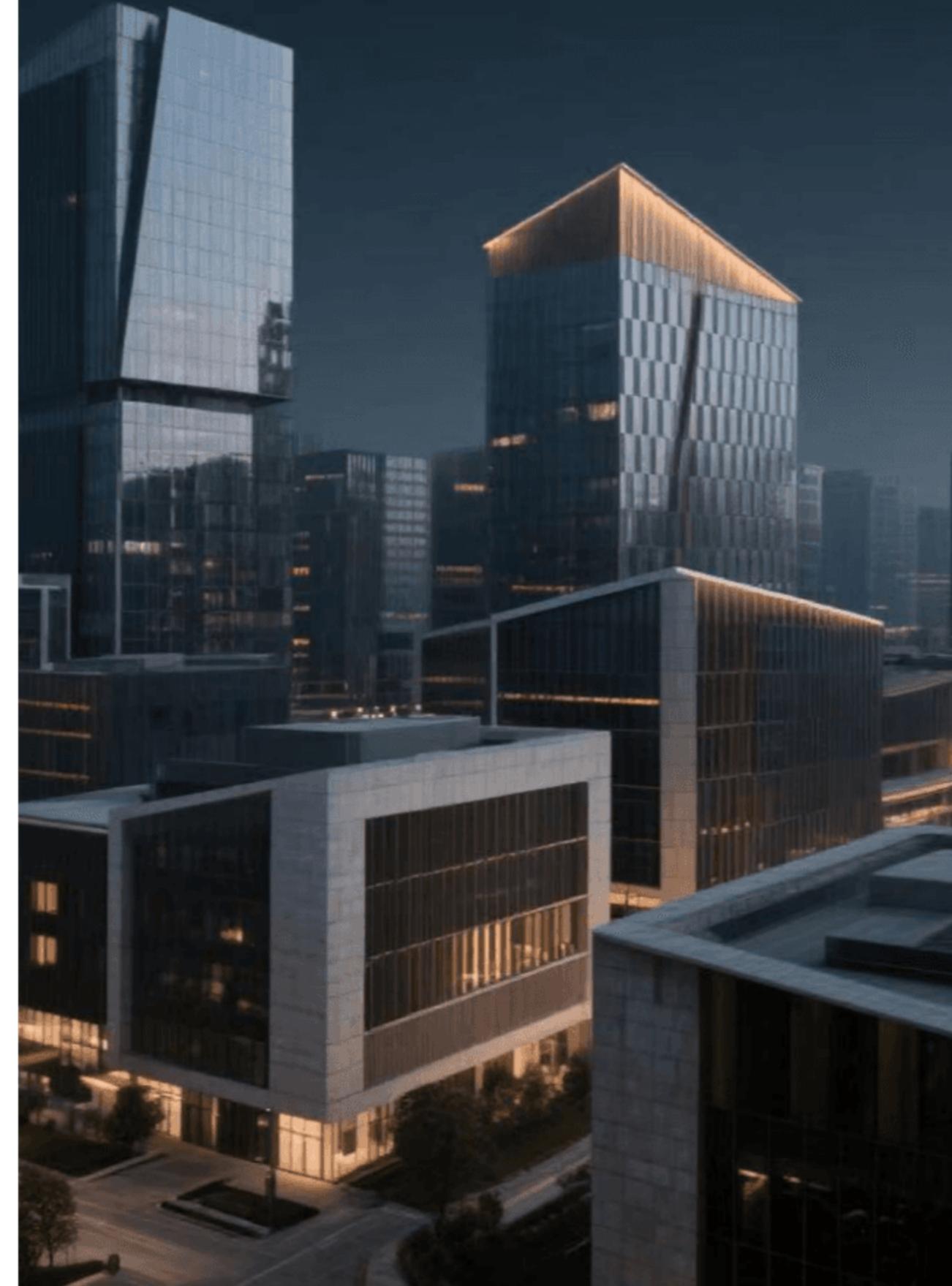
Cấu trúc dữ liệu gốc

Các bảng dữ liệu gốc

Dữ liệu gốc gồm các bảng: orders, customers, products, payments, shipping, warehouses, provinces, districts. Mỗi bảng chứa thông tin chi tiết về đơn hàng, khách hàng, sản phẩm, thanh toán, vận chuyển và kho hàng.

Ý nghĩa dữ liệu gốc

Các bảng dữ liệu gốc là nền tảng để xây dựng mô hình phân tích sau này, giúp doanh nghiệp hiểu rõ hoạt động kinh doanh, từ đó đưa ra các quyết định chiến lược và chiến thuật phù hợp.



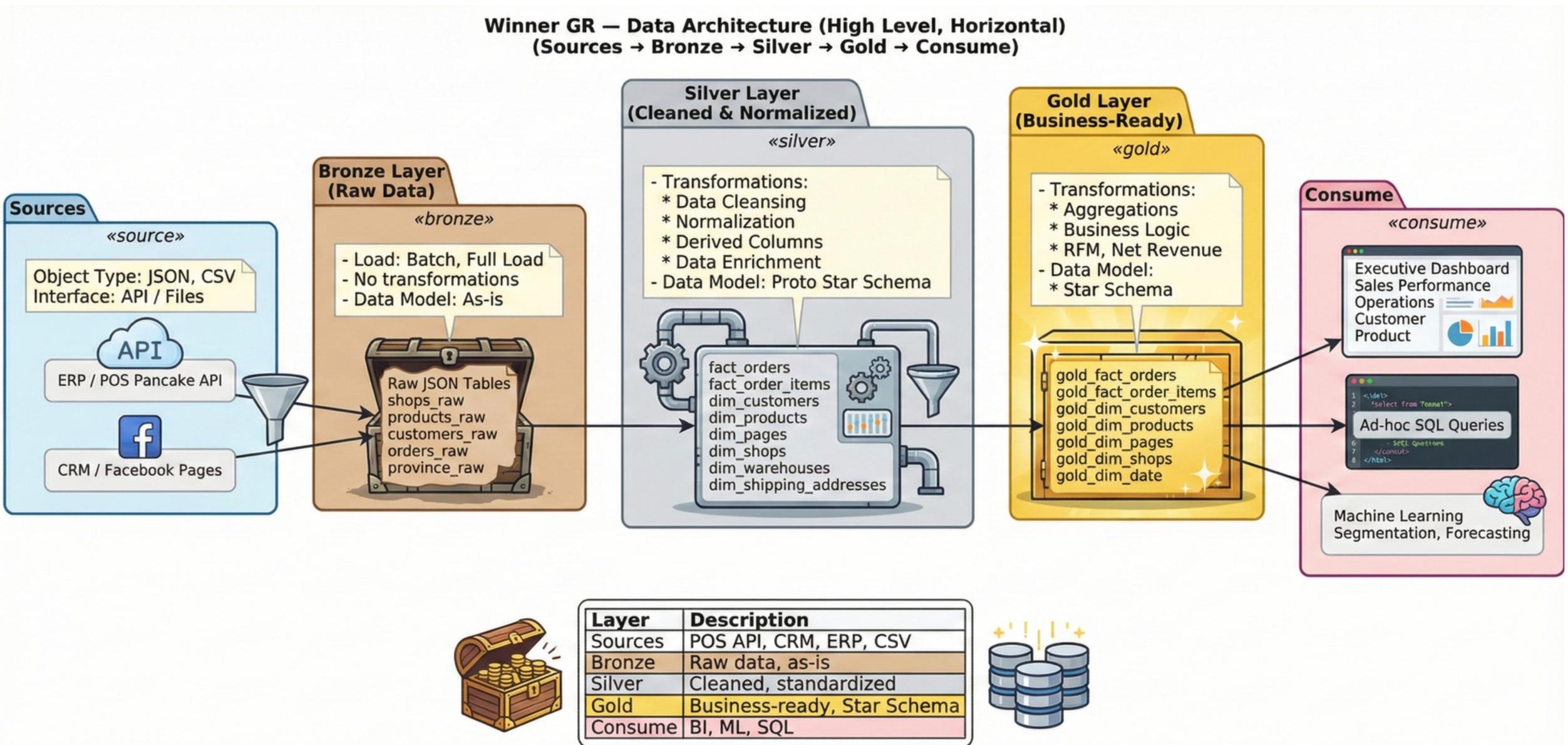


03

Kiến trúc dự án



Kiến trúc dự án



Mô hình

Dự án áp dụng mô hình gồm ba tầng: Bronze lưu trữ dữ liệu thô từ API, Silver làm sạch và chuẩn hóa, Gold xây dựng Star Schema tối ưu cho phân tích.

Lợi ích của Kiến trúc 3 tầng

Mô hình Medallion đảm bảo dữ liệu gốc được bảo toàn, dễ truy vết và mở rộng, phù hợp với quy mô và nhu cầu của doanh nghiệp.

Quy trình làm việc

Dữ liệu từ Pancake POS API được đưa vào tầng Bronze, sau đó qua tầng Silver để làm sạch và chuẩn hóa, cuối cùng đến tầng Gold để phục vụ phân tích và báo cáo.



Star Schema tầng Gold

Cấu trúc Star Schema

Tầng Gold gồm 2 fact tables (orders, order_items) và 8 dimension tables (customers, products, shop, date, pages, warehouses, shipping, payments).

Lợi ích của Star Schema

Thiết kế Star Schema tối ưu cho truy vấn nhanh, hỗ trợ drill-down, roll-up và là chuẩn cho data warehouse, sẵn sàng phục vụ dashboard và báo cáo phân tích.



04

Khám phá dữ liệu



Tổng quan doanh thu và đơn hàng

Doanh thu và đơn hàng

Tổng doanh thu đạt 17,9 tỷ đồng, sau giảm giá 16,5 tỷ, AOV 411.218đ, tỷ lệ giảm giá 7,99% hợp lý. Có 40.236 đơn, nhưng chỉ 42,9% hoàn thành, 33,9% bị trả, 1,11 đơn/khách cho thấy hiệu suất vận hành và trải nghiệm khách hàng cần cải thiện mạnh mẽ.

Ý nghĩa của số liệu

Các số liệu này phản ánh hiệu quả kinh doanh hiện tại của doanh nghiệp, đồng thời chỉ ra những điểm cần cải thiện như tỷ lệ hoàn thành đơn hàng và tỷ lệ trả hàng.

Hướng cải thiện

Doanh nghiệp cần tập trung cải thiện tỷ lệ hoàn thành đơn hàng và giảm tỷ lệ trả hàng để tăng hiệu quả vận hành và trải nghiệm khách hàng.

Phân tích khách hàng

Đặc điểm khách hàng

36.090 khách, nữ chiếm 54%, retention chỉ 9,68%, 90,3% mua một lần. Phân khúc RFM cho thấy lượng VIP nhỏ, lớp khách At-Risk và Lost lớn, CLV trung bình thấp.

Ý nghĩa của phân tích

Phân tích này giúp doanh nghiệp hiểu rõ hơn về khách hàng, từ đó đưa ra chiến lược marketing và chăm sóc khách hàng hiệu quả hơn.

Hướng khai thác

Doanh nghiệp có thể tập trung vào phân khúc nữ, tăng retention rate và CLV, đồng thời cải thiện trải nghiệm khách hàng để giảm tỷ lệ mất khách.



Hiệu suất sản phẩm

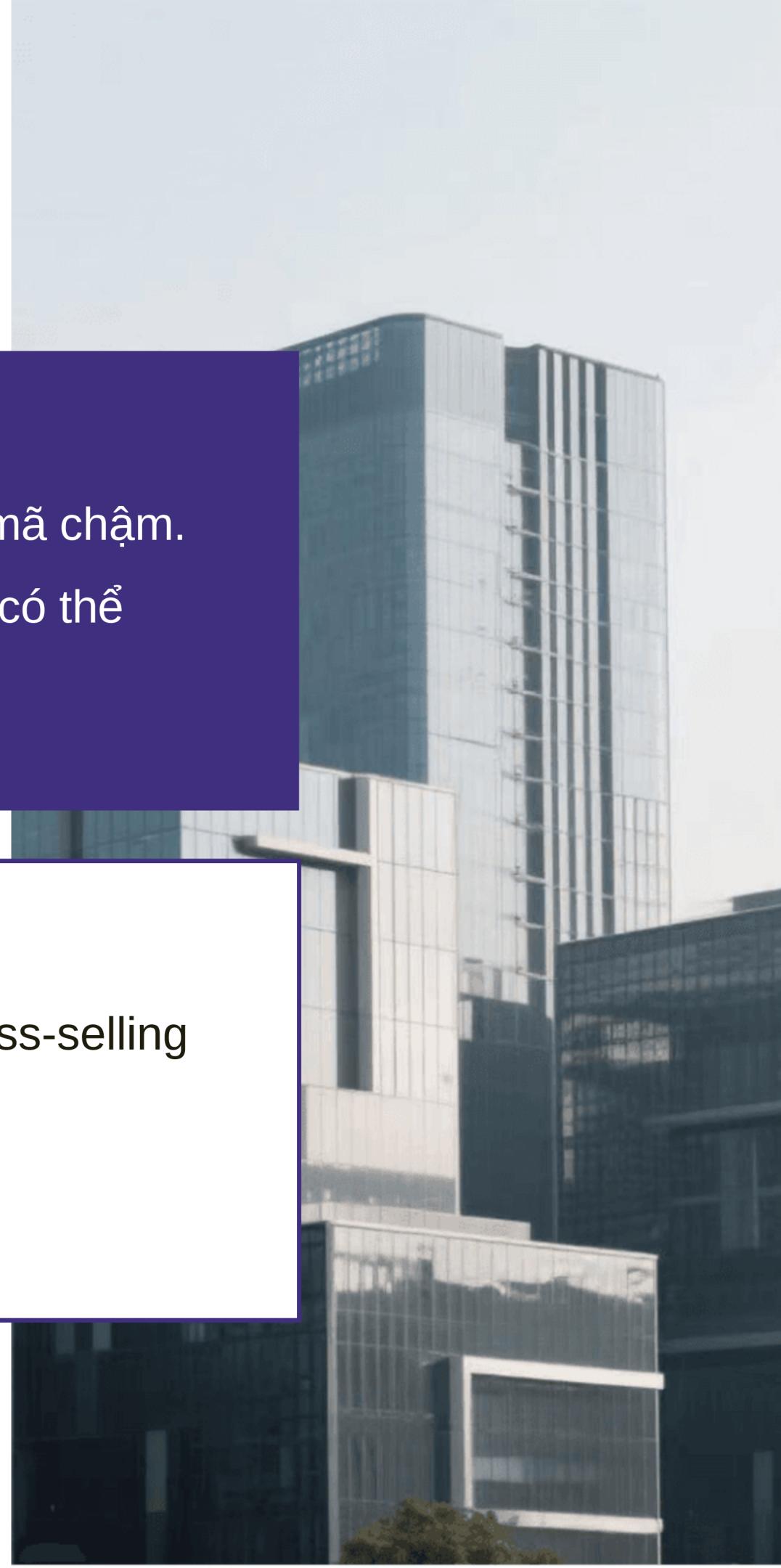
Tổng quan sản phẩm

37 mã hàng, 46.611 items bán, top 10 sản phẩm chiếm phần lớn doanh thu, nhiều mã chậm.

Phân tích lifecycle cho thấy sản phẩm cần được tái cơ cấu, bundle và cross-selling có thể tăng AOV.

Hướng tối ưu

Doanh nghiệp cần tái cơ cấu sản phẩm, loại bỏ những mã chậm và tăng cường cross-selling để nâng cao hiệu quả kinh doanh và tăng doanh thu.





05

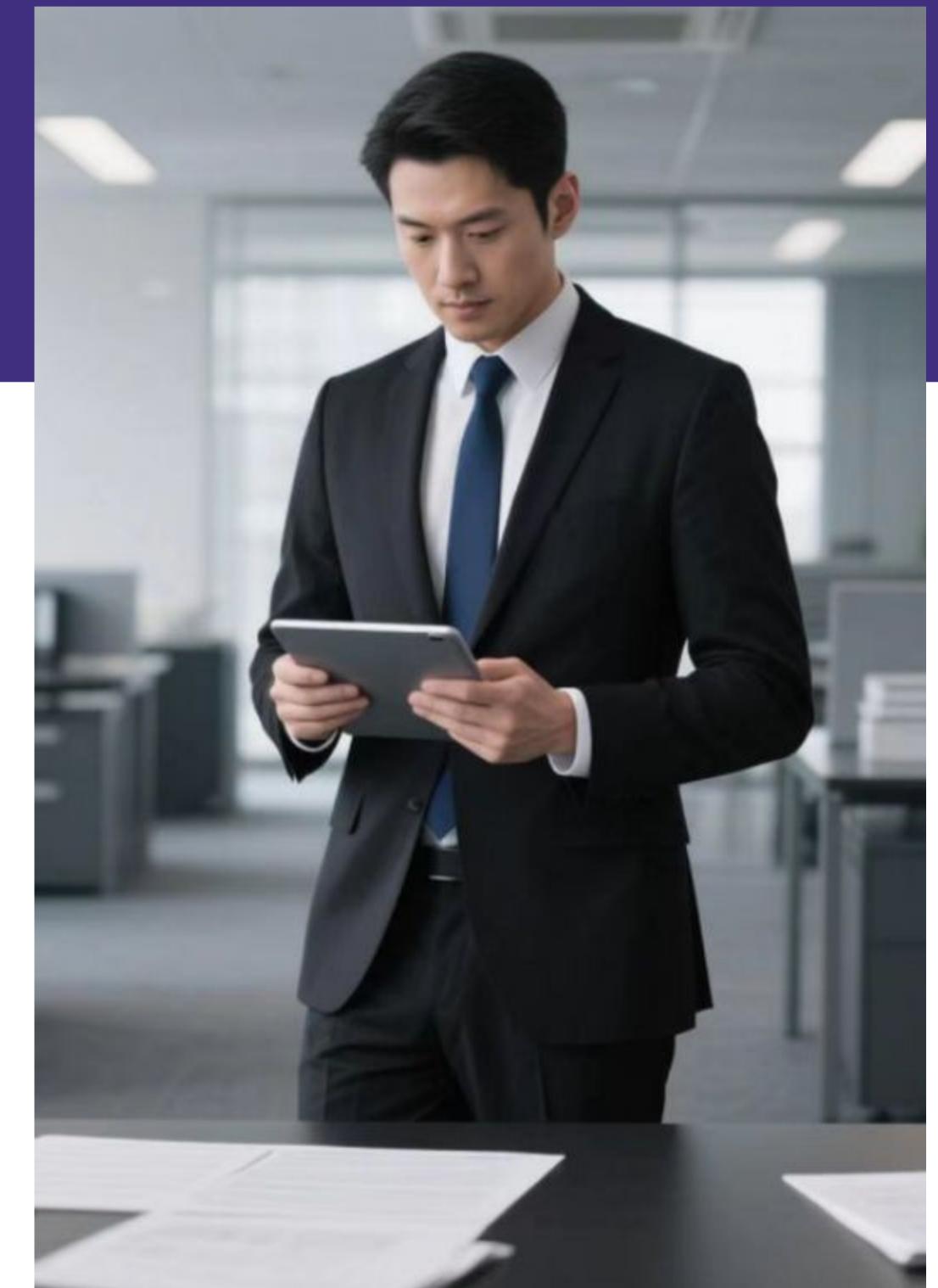
Vấn đề & Cơ hội



Ba vấn đề nghiêm trọng

Tỷ lệ giữ chân khách hàng thấp

Retention rate chỉ 9,68% cực thấp làm mất doanh thu lặp lại và tăng chi phí, cần được ưu tiên cải thiện trong thời gian ngắn nhất.



Tỷ lệ hoàn hàng cao

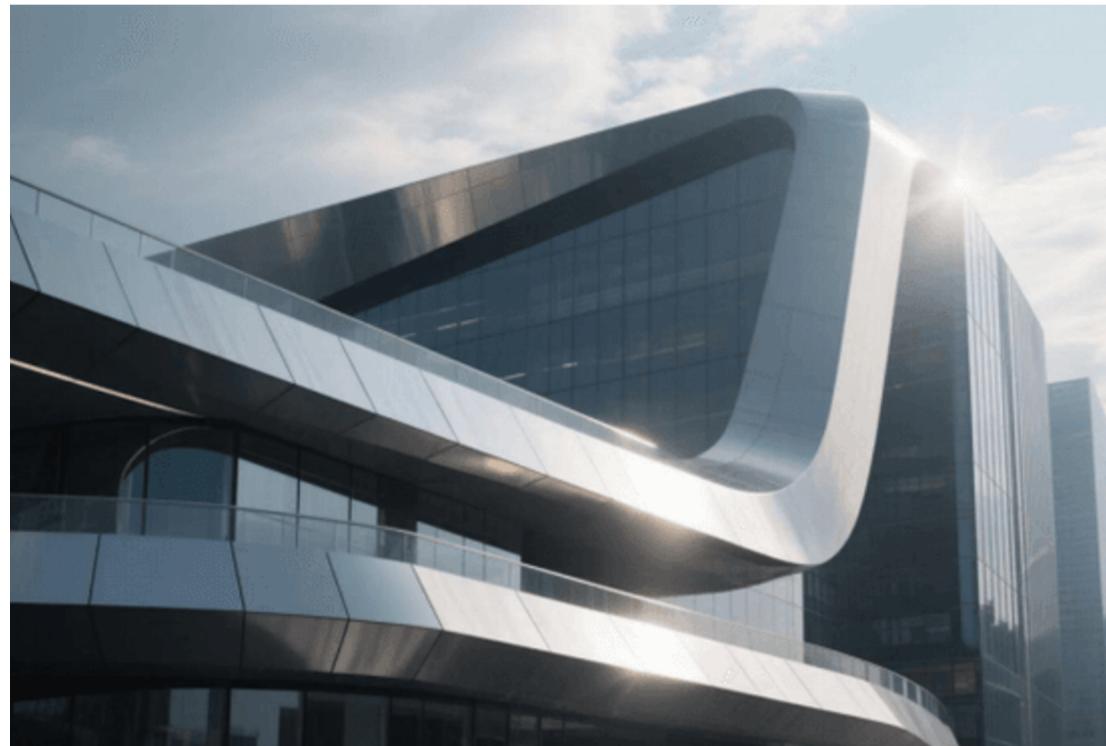
Return rate 33,9% cao gây tốn logistics và giảm lợi nhuận, ảnh hưởng lớn đến hiệu quả kinh doanh của doanh nghiệp.

Tỷ lệ hoàn thành đơn thấp

Fulfillment rate 42,9% thấp khiến trải nghiệm khách kém và doanh thu thực thu thiệt hại, cần được nâng cao gấp.



Cơ hội tăng trưởng



Tăng retention rate

Nếu retention tăng từ 9,68% lên 20% doanh thu có thể tăng gần gấp đôi, mang lại lợi ích kinh doanh to lớn cho doanh nghiệp.



Giảm return rate

Giảm return xuống 15% tiết kiệm chi phí lớn, giúp doanh nghiệp giảm thiểu tổn thất và tăng lợi nhuận.



Tập trung vào nữ

Tập trung nữ 54% và cross-selling nâng AOV, tận dụng lợi thế thị trường nữ để tăng doanh thu và hiệu quả kinh doanh.



06

Hành động cụ thể





Ngắn hạn 1-3 tháng

Giảm return rate

Giảm return về <20% bằng khảo sát lý do trả, cải thiện mô tả và chất lượng sản phẩm, từ đó nâng cao trải nghiệm khách hàng và giảm chi phí trả hàng.

Tăng retention rate

Tăng retention lên >15% qua loyalty program, email remarketing, faster shipping và personalized offer, giúp doanh nghiệp tăng doanh thu lặp lại.

Nâng fulfillment rate

Nâng fulfillment >60% bằng tối ưu quy trình đơn, inventory và communication, từ đó cải thiện trải nghiệm khách hàng và tăng doanh thu thực thu.



Trung hạn 3-6 tháng

Triển khai phân khúc RFM và personalization

Triển khai phân khúc RFM và personalization toàn diện, marketing theo segment, recommendation theo hành vi, giúp doanh nghiệp hiểu rõ hơn về khách hàng và tăng hiệu quả marketing.



Tối ưu tồn kho

Tối ưu tồn kho dựa trên velocity, loại bỏ chậm, tăng bestseller, giúp doanh nghiệp quản lý kho hàng hiệu quả và giảm chi phí tồn kho.



07

Kết quả & Đường hướng



Sản phẩm kỹ thuật đạt được

Pipeline hoàn chỉnh

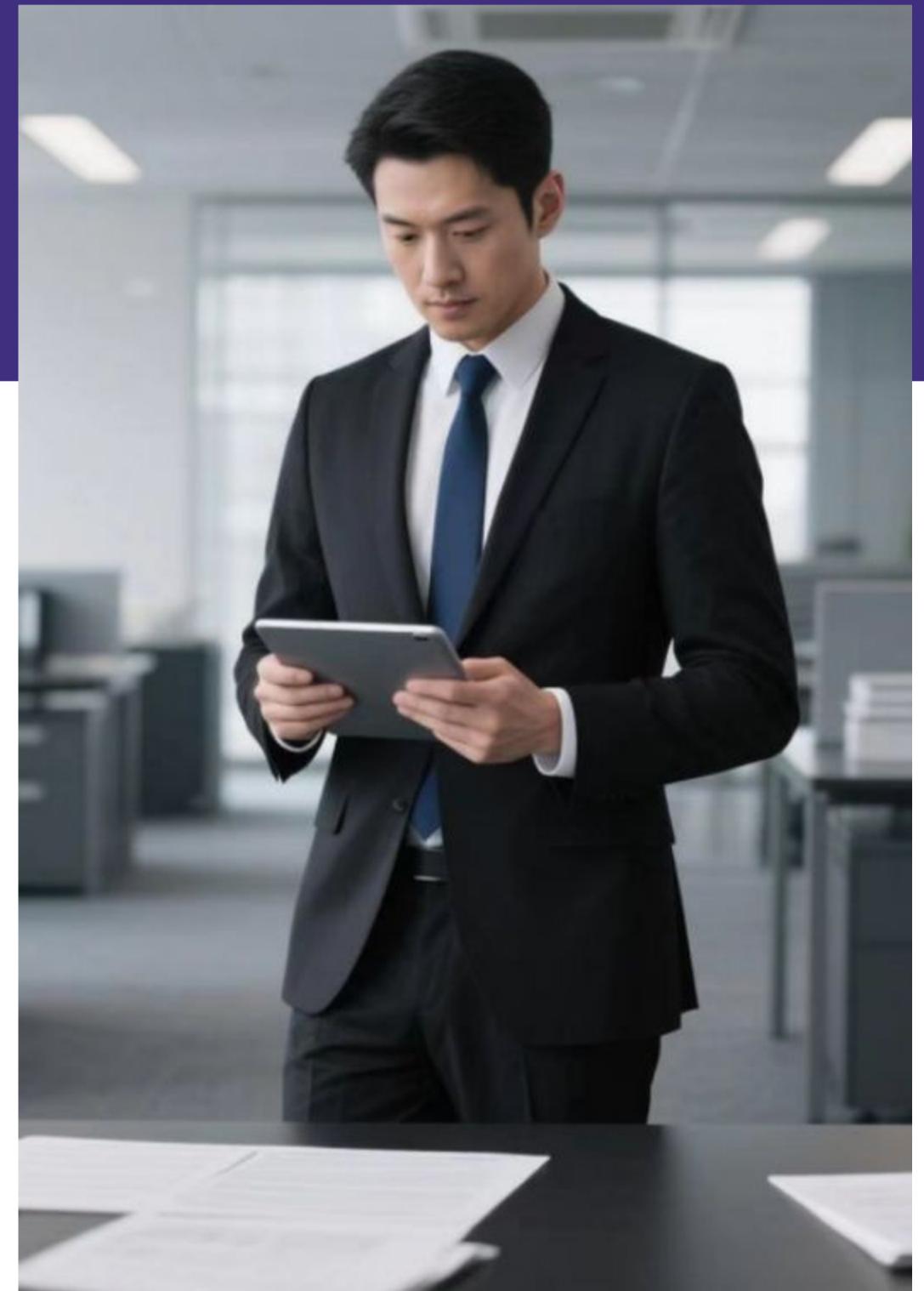
Hoàn thành pipeline Bronze-Silver-Gold với 20+ notebook, Star Schema 2 fact 8 dim, dictionary đầy đủ, đảm bảo dữ liệu được quản lý và khai thác hiệu quả.

EDA và PowerBI

EDA 4 notebook, 20+ visualization, PowerBI 5 trang tương tác, drill-down, export, giúp doanh nghiệp có insights rõ ràng và báo cáo chuyên nghiệp.

SQL và tài liệu

SQL RBAC, tài liệu architecture, hướng dẫn vận hành, đảm bảo quy trình và kiến trúc được ghi nhận và có thể tái sử dụng.



Giá trị kinh doanh mang lại

Hiểu rõ toàn cảnh

Hiểu rõ toàn cảnh 4 năm, phát hiện critical issues, 15+ khuyến nghị ưu tiên, giúp doanh nghiệp có hướng đi rõ ràng và chiến lược phù hợp.

Tăng trưởng và lợi nhuận

Tiềm năng tăng retention gấp đôi doanh thu, giảm return tiết kiệm chi phí, xây dựng văn hóa data-driven, ROI cao trong 6-12 tháng tới.



THANK YOU

