

# CÁC MÔ HÌNH PHÂN TÍCH DỮ LIỆU

Giảng viên hướng dẫn: TS. Nguyễn Trần Minh Thư

## PHÂN TÍCH DỮ LIỆU BẢN LẺ TRỰC TUYẾN

---

24C12006 – PHẠM PHONG HÀO

24C11060 – TRẦN ĐÌNH SANG

# Table of contents

**01**

---

Bối cảnh & Mục tiêu

**02**

---

Phương pháp tiếp cận

**03**

---

Kết quả

**04**

---

Insights

**05**

---

Kết luận

# **BỐI CẢNH VÀ MỤC TIÊU**

# Business Goal 1

Trong vòng 2 tuần,  
xây dựng báo cáo  
phân tích chuỗi thời  
gian doanh thu theo  
ngày – tuần – tháng  
→ xác định xu hướng  
nền, mùa vụ và biến  
động ngắn hạn.

Mục tiêu:

- ★ hỗ trợ nhà quản lý phân biệt tăng trưởng thực sự và biến động ngẫu nhiên
- ★ làm cơ sở cho điều chỉnh chiến lược kinh doanh trong năm hiện tại.

# Mục tiêu phân tích

Phân tích và làm rõ xu hướng biến động doanh thu theo thời gian trong năm hiện tại

→ giúp doanh nghiệp nhận diện giai đoạn tăng trưởng, suy giảm và thời điểm cần can thiệp chiến lược.

# Bài toán nghiệp vụ

Doanh thu biến động mạnh theo ngày và tuần

→ nhà quản lý khó phân biệt xu hướng thực sự với dao động ngắn hạn

→ ảnh hưởng đến quyết định giá, chương trình bán hàng và phân bổ nguồn lực.

# Câu chuyện dữ liệu

Doanh thu tuần này tăng mạnh so với tuần trước.

- Xu hướng này có lặp lại hay chỉ là biến động ngẫu nhiên?
- Đây là tín hiệu tăng trưởng mới hay hiệu ứng ngắn hạn?

# SMART – Specific & Measurable

- Xác định trend, seasonality và dao động ngắn hạn của doanh thu.
- Đánh giá bằng tăng trưởng tuần, tháng, rolling statistics và kết quả phân rã chuỗi thời gian để phân loại các giai đoạn biến động.



# SMART – Achievable & Relevant

Phân tích khả thi nhờ dữ liệu 12 tháng chi tiết theo ngày.

Sử dụng moving average, thống kê trượt và phân rã chuỗi thời gian.

Kết quả giúp tránh quyết định dựa trên biến động ngẫu nhiên, tối ưu điều chỉnh giá và phân bổ nguồn lực.

# SMART – Time - bound

Thực hiện trong 2 tuần với toàn bộ dữ liệu doanh thu 12 tháng. Kết quả được tổng hợp thành báo cáo hoàn chỉnh và có thể tái sử dụng cho theo dõi định kỳ.

# Business Goal 2

Nâng cao hiệu quả kinh doanh bằng cách xác định và tập trung vào các nhóm khách hàng, dựa trên phân tích xu hướng và đặc điểm từng nhóm khách hàng.

Mục tiêu:

★ Phân tích và dự báo doanh thu theo nhóm khách hàng nhằm xác định các phân khúc khách hàng, ổn định và tiềm năng tăng trưởng trong thời gian 30 ngày tiếp theo.

# Mục tiêu phân tích

***Phân tích và dự báo doanh thu theo nhóm khách hàng*** nhằm xác định các phân khúc khách hàng, ổn định và tiềm năng tăng trưởng trong thời gian 30 ngày tiếp theo.

# Bài toán nghiệp vụ

Doanh thu được tạo ra từ nhiều nhóm khách hàng khác nhau với mức độ đóng góp và tính ổn định không đồng đều. Việc thiếu thông tin về xu hướng và mùa vụ, khả năng duy trì doanh thu theo từng nhóm khách hàng khiến việc tiếp cận chăm sóc từng nhóm khách hàng trở nên khó khăn hơn.

# Câu chuyện dữ liệu

Nghiên cứu này phân chia tập khách hàng theo hai chiều **Tier (Vàng–Bạc–Đồng)** và **BuyerType (Sỉ–Lẻ)**, kết hợp phân tích chuỗi thời gian nhằm tách **xu hướng, mùa vụ và nhiễu**, qua đó làm rõ sự khác biệt về hành vi mua và mức độ đóng góp doanh thu giữa các phân khúc. Kết quả cho thấy **system customer** vẫn đóng vai trò nền tảng nhưng có dấu hiệu chững lại trong ngắn hạn, trong khi **walk-in customer** tuy tỷ trọng thấp hơn lại có mức biến động và seasonality cao, trở thành nguồn động lực tăng trưởng ngắn hạn.

# SMART Goals

- ❑ **Specific:** Phân tích và dự báo doanh thu theo chuỗi thời gian cho các phân khúc khách hàng theo Tier (Gold–Silver–Bronze) và BuyerType (bán sỉ–bán lẻ).
- ❑ **Measurable:** Đo lường mức tăng/giảm doanh thu bằng cách so sánh doanh thu trung bình 30 ngày gần nhất và 30 ngày dự báo cho từng phân khúc.
- ❑ **Achievable:** Sử dụng STL decomposition và Prophet để nắm bắt xu hướng, mùa vụ theo tuần và biến động ngẫu nhiên của doanh thu.
- ❑ **Relevant:** Hỗ trợ nhận diện phân khúc chứng lại và động lực tăng trưởng ngắn hạn, làm cơ sở cho quyết định kinh doanh.
- ❑ **Time-bound:** Dự báo doanh thu 30 ngày tiếp theo phục vụ hoạch định chiến lược ngắn hạn.

## **2. PHƯƠNG PHÁP TIẾP CẬN BÀI TOÁN**



## 2.1. MỤC TIÊU

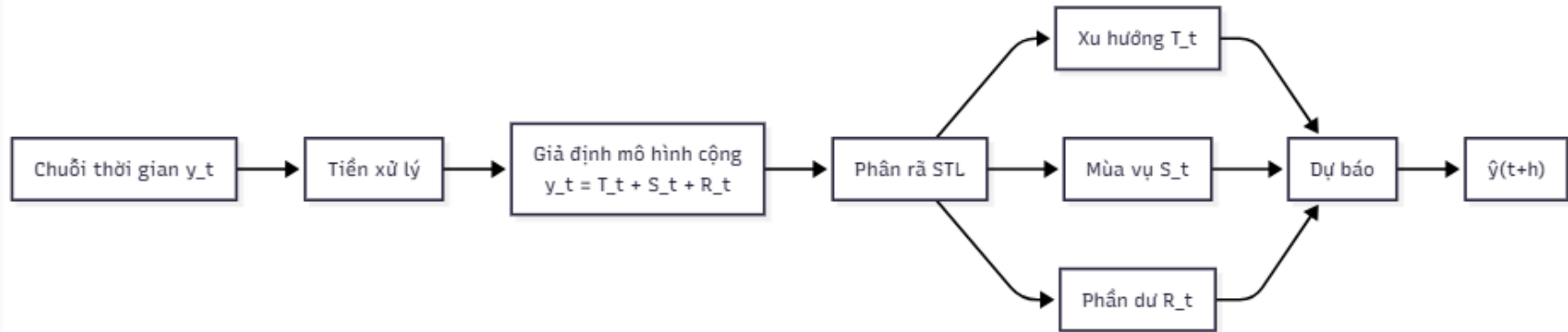
Đánh giá tổng quan kết quả bán hàng theo thời gian  
Phân tích cơ cấu doanh thu theo sản phẩm và khách hàng → nhận diện các xu hướng và rủi ro tiềm ẩn trong hoạt động kinh doanh.

## 2.2. PHẠM VI DỮ LIỆU

Dữ liệu bao gồm hơn 540.000 giao dịch bán lẻ phát sinh trong vòng một năm.

Các giao dịch bất thường như đơn hàng bị hủy hoặc giá trị không hợp lệ đã được loại bỏ nhằm đảm bảo tính chính xác và nhất quán của kết quả

## 2.3. Quy trình chung



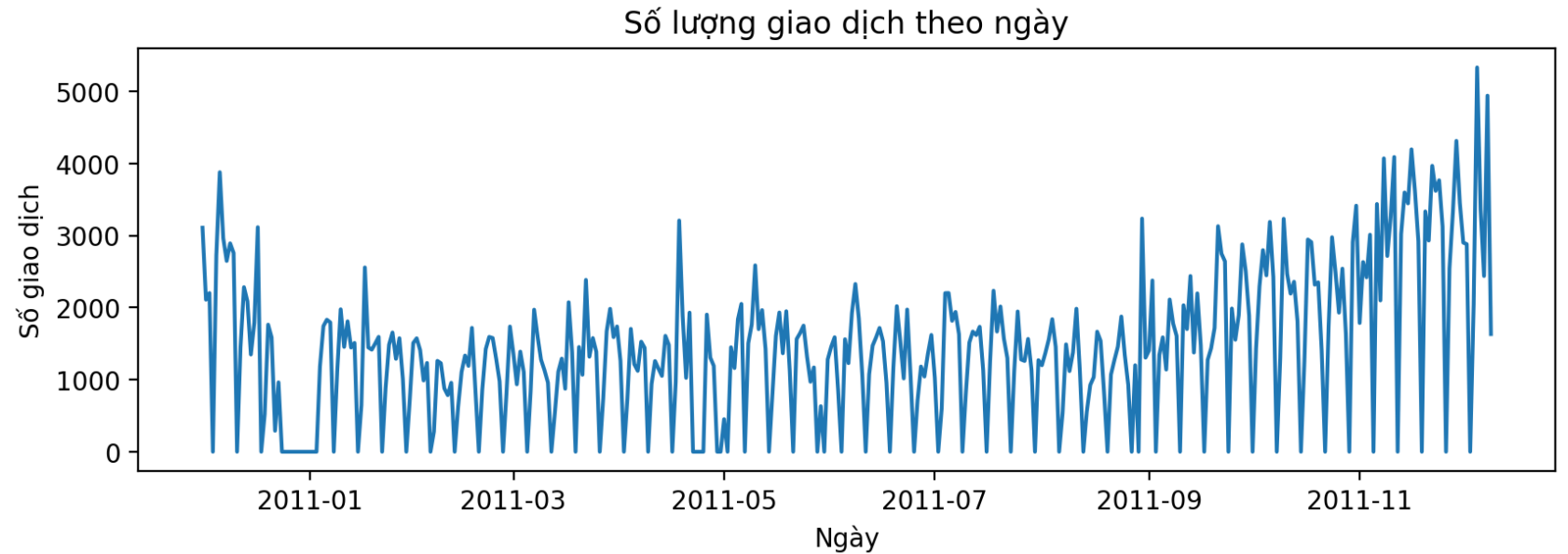
### **3. KẾT QUẢ**

# 3. 1. Tổng quan dữ liệu

Mỗi record chứa thông tin:

- Thời điểm mua hàng,
- Mã và mô tả sản phẩm, số lượng,
- Đơn giá
- Tổng giá trị giao dịch
- Định danh khách hàng
- Quốc gia phát sinh giao dịch

# 3. 1. Tổng quan dữ liệu



Hình 1. Số lượng giao dịch theo ngày trong bộ dữ liệu Online Retail, cho thấy mật độ dữ liệu theo thời gian và các khoảng biến động đáng chú ý.

# Xu hướng doanh thu theo thời gian

Doanh thu có xu hướng tăng trưởng theo thời gian, đặc biệt rõ rệt vào các giai đoạn cuối năm.

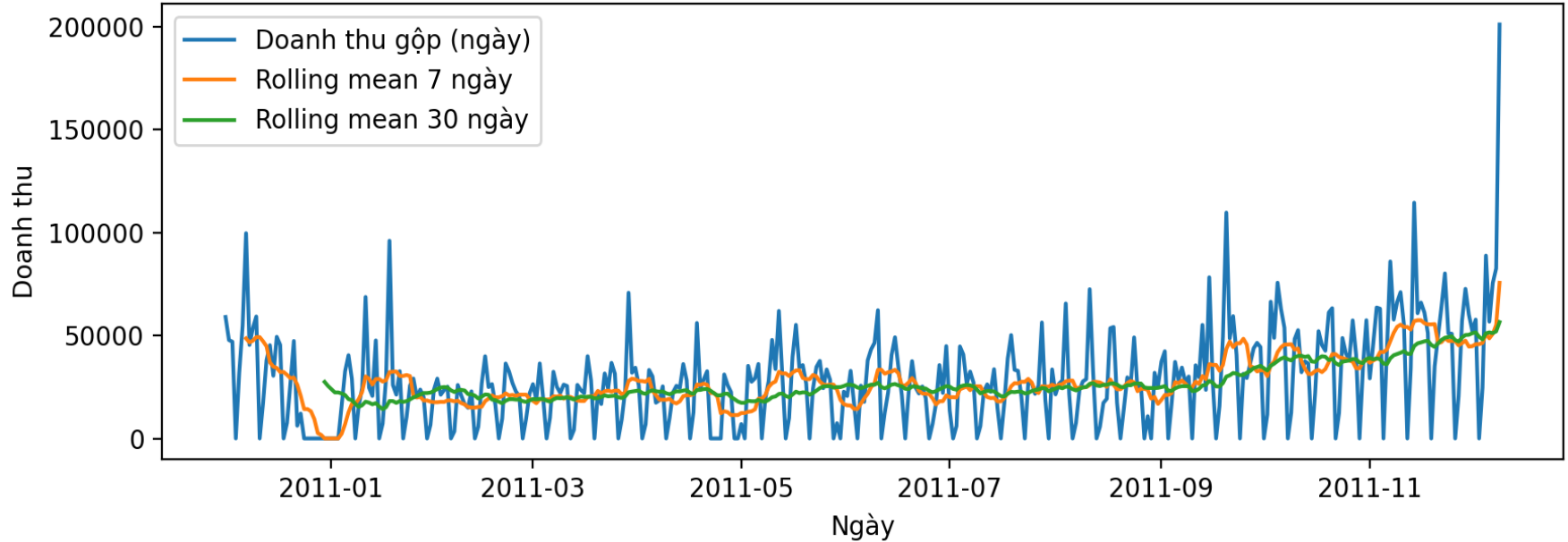
Tuy nhiên, chuỗi doanh thu tồn tại nhiều biến động do ảnh hưởng của yếu tố mùa vụ và các sự kiện ngắn hạn

# Làm trơn & biến động doanh thu

Đường làm trơn (rolling 7 ngày và 30 ngày) giúp làm rõ xu hướng tổng thể và giảm nhiễu ngắn hạn. Phân tích hệ số biến thiên cho thấy mức độ ổn định của doanh thu thay đổi theo từng giai đoạn trong năm.

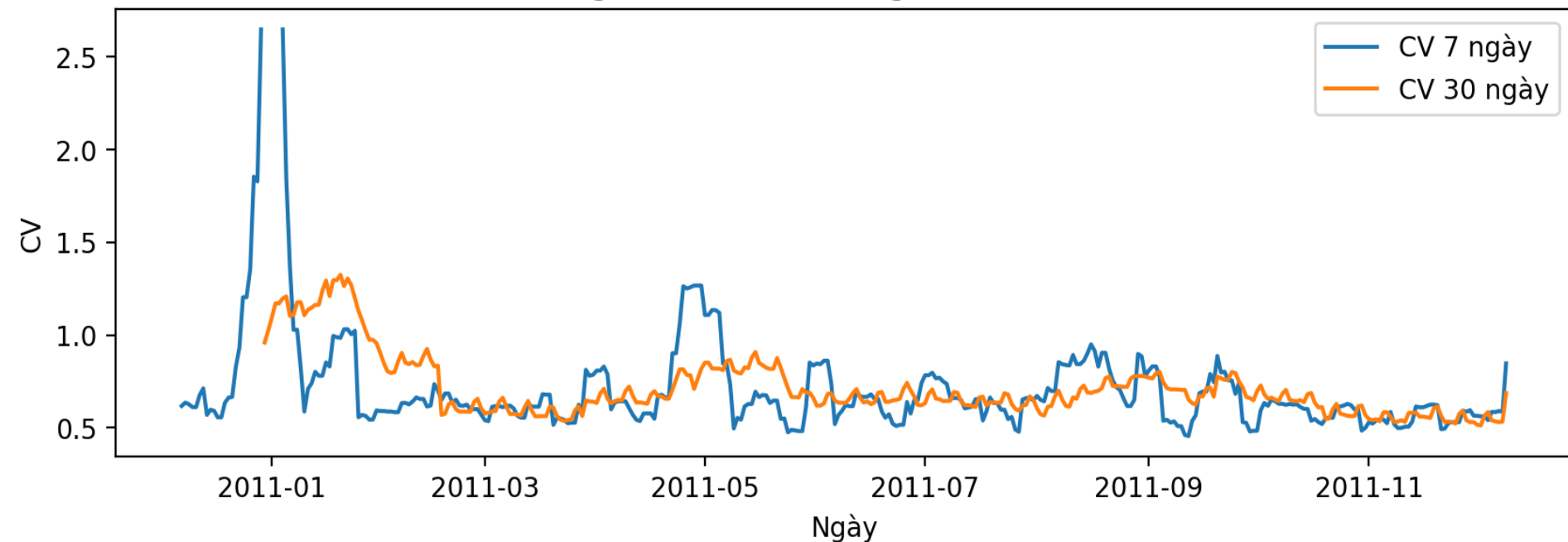


Doanh thu gộp theo ngày và đường làm trơn



Hình 2. Doanh thu gộp theo ngày và đường làm trơn (7 ngày, 30 ngày). Đường làm trơn giúp làm rõ xu hướng tổng thể và giảm nhiễu ngắn hạn trong chuỗi doanh thu.

Độ biến động doanh thu (Rolling Coefficient of Variation)

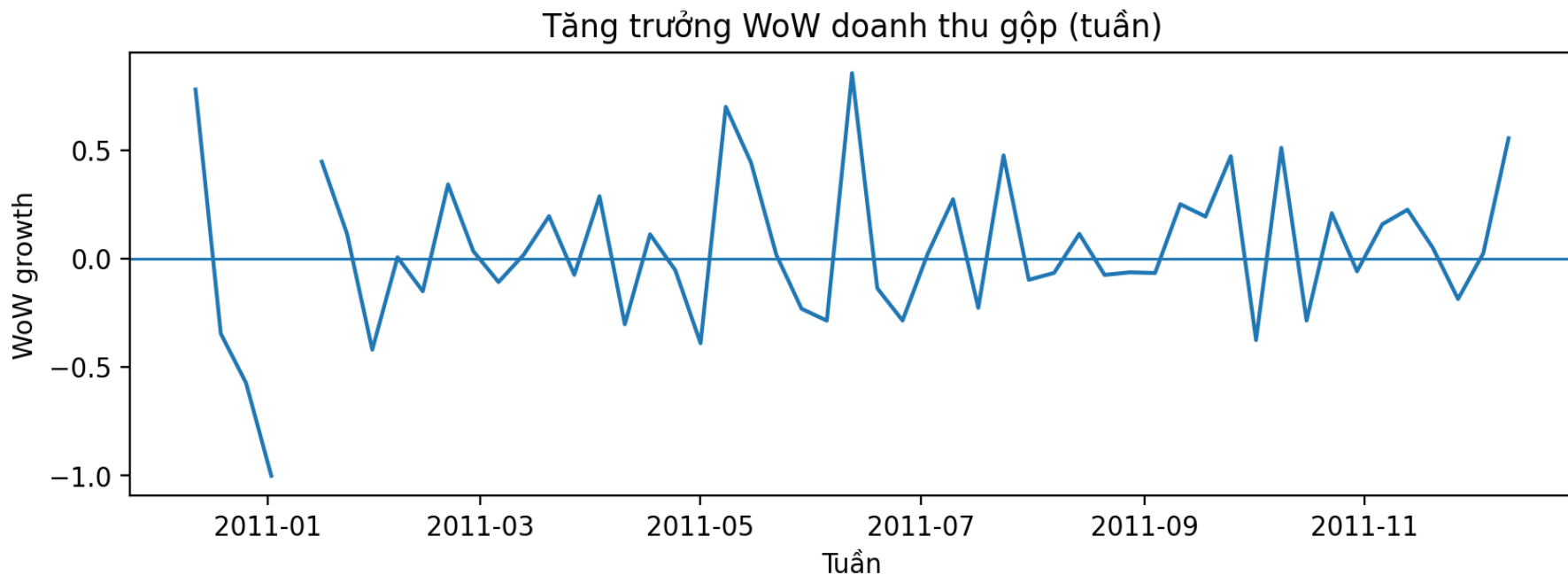


Hình 3. Độ biến động doanh thu theo thời gian thông qua hệ số biến thiên (CV) rolling 7 ngày và 30 ngày, phản ánh mức độ ổn định/dao động của chuỗi doanh thu

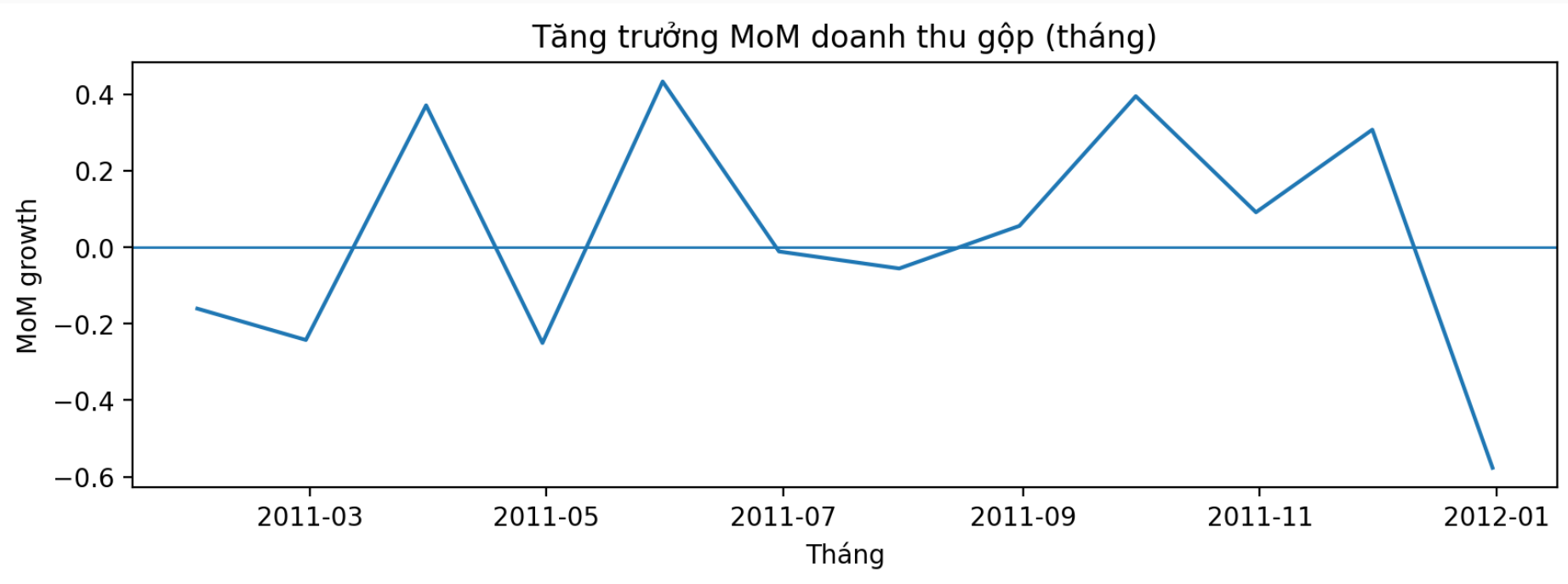
# Tăng trưởng WoW & MoM

Tăng trưởng doanh thu theo tuần (WoW) phản ánh biến động ngắn hạn và hiệu ứng mùa vụ

Tăng trưởng theo tháng (MoM) cung cấp góc nhìn vĩ mô hơn về các giai đoạn cao điểm và thấp điểm.



Hình 4. Tăng trưởng doanh thu gộp theo tuần (WoW). Các pha tăng/giảm mạnh phản ánh biến động theo chu kỳ ngắn hạn và hiệu ứng mùa vụ.



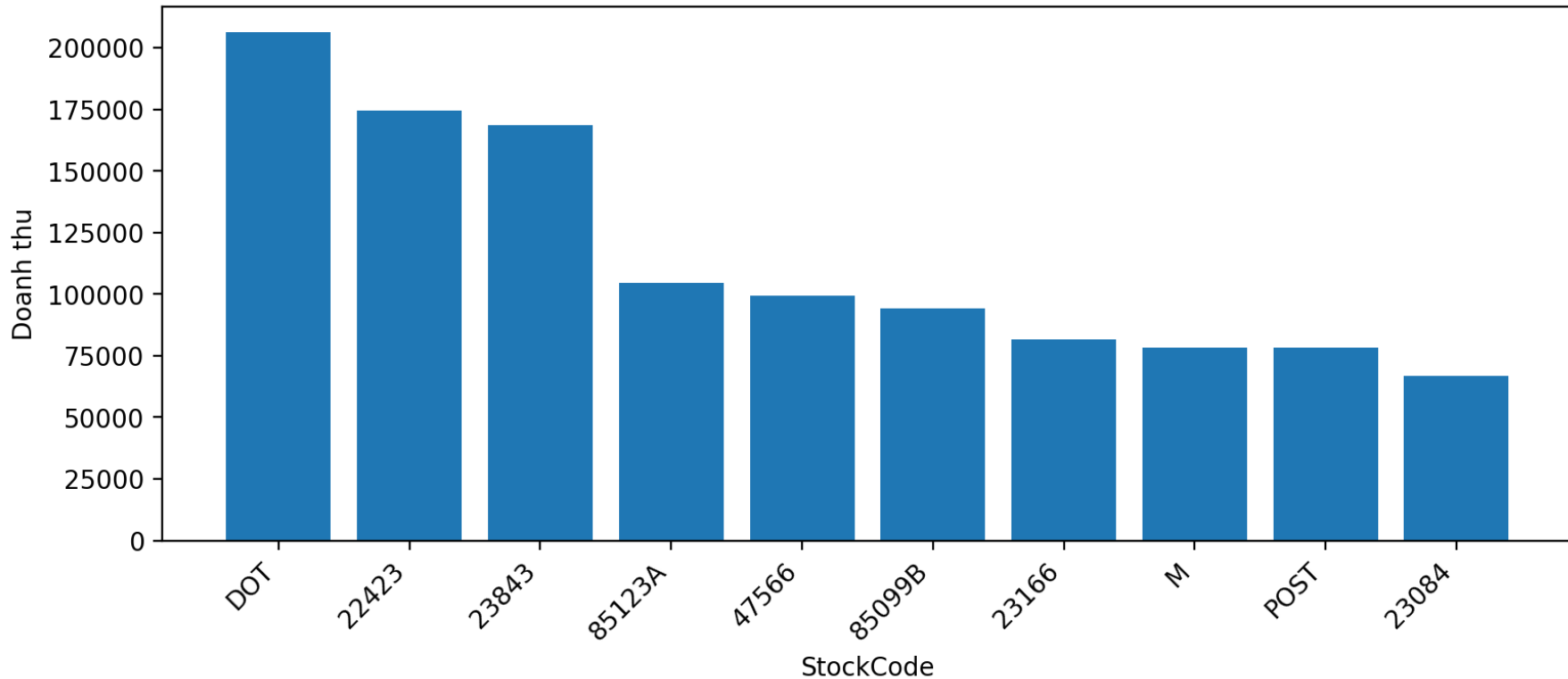
Hình 5. Tăng trưởng doanh thu gộp theo tháng (MoM), cung cấp góc nhìn vĩ mô hơn về xu hướng và các giai đoạn cao điểm/thấp điểm trong năm.

# Phân tích theo sản phẩm

Doanh thu phân bổ không đồng đều giữa các sản phẩm. Một nhóm nhỏ sản phẩm tạo ra phần lớn doanh thu, trong khi nhiều sản phẩm khác đóng góp rất thấp.

➔ Kết quả gợi ý chiến lược tập trung vào các sản phẩm chủ lực.

Top 10 sản phẩm theo doanh thu gộp



Hình 6. Top 10 sản phẩm theo doanh thu gộp (chỉ tính giao dịch bán). Kết quả cho thấy doanh thu tập trung vào một nhóm sản phẩm chủ lực.

# Phân tích theo khách hàng

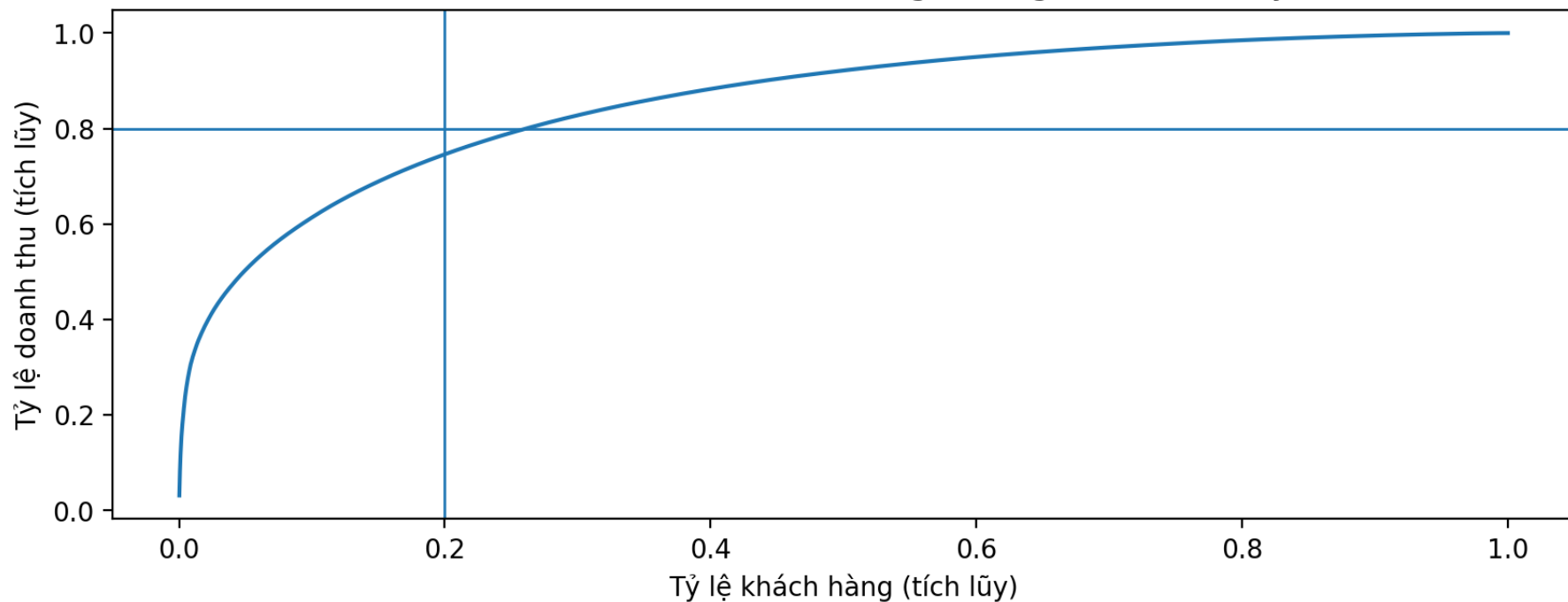
Doanh thu tập trung mạnh vào một tỷ lệ nhỏ khách hàng, phù hợp với nguyên lý Pareto.

→ Nhóm khách hàng giá trị cao đóng vai trò quyết định nhưng cũng tiềm ẩn rủi ro nếu không có chiến lược duy trì và mở rộng phù hợp

→ Cần quan tâm đến việc phát triển nhóm khách hàng trung bình, nơi tạo ra nguồn doanh thu ổn định hơn trong dài hạn.



Pareto doanh thu theo khách hàng (đường Lorenz tích lũy)

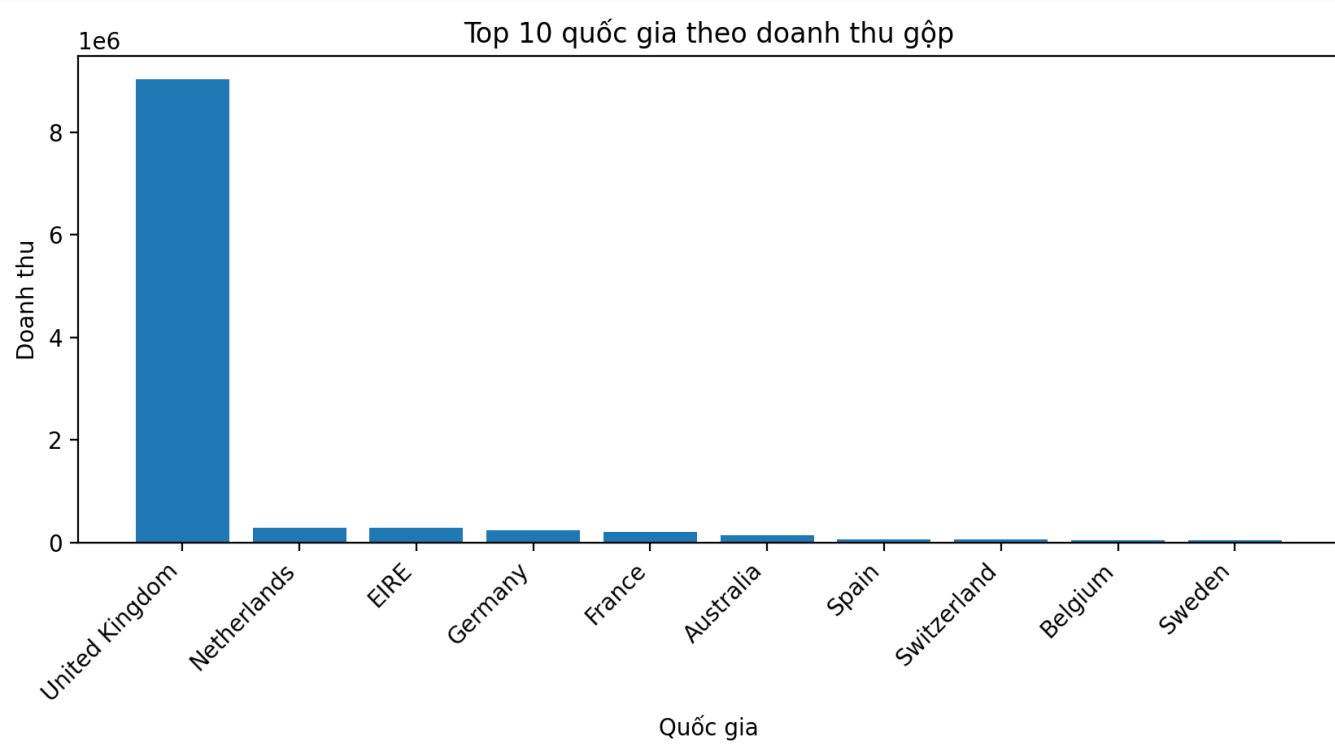


Hình 7. Đường tích lũy doanh thu theo khách hàng (Pareto/Lorenz) trên các giao dịch bán. Đường cong cho thấy mức độ tập trung doanh thu và gợi ý chiến lược ưu tiên nhóm khách hàng giá trị cao.

# Phân tích theo quốc gia

Doanh thu

- Chủ yếu tập trung tại thị trường United Kingdom
  - Các thị trường khác đóng góp tỷ trọng nhỏ
- Sự tập trung này mang lại lợi thế vận hành nhưng đồng thời làm gia tăng rủi ro phụ thuộc thị trường chính.

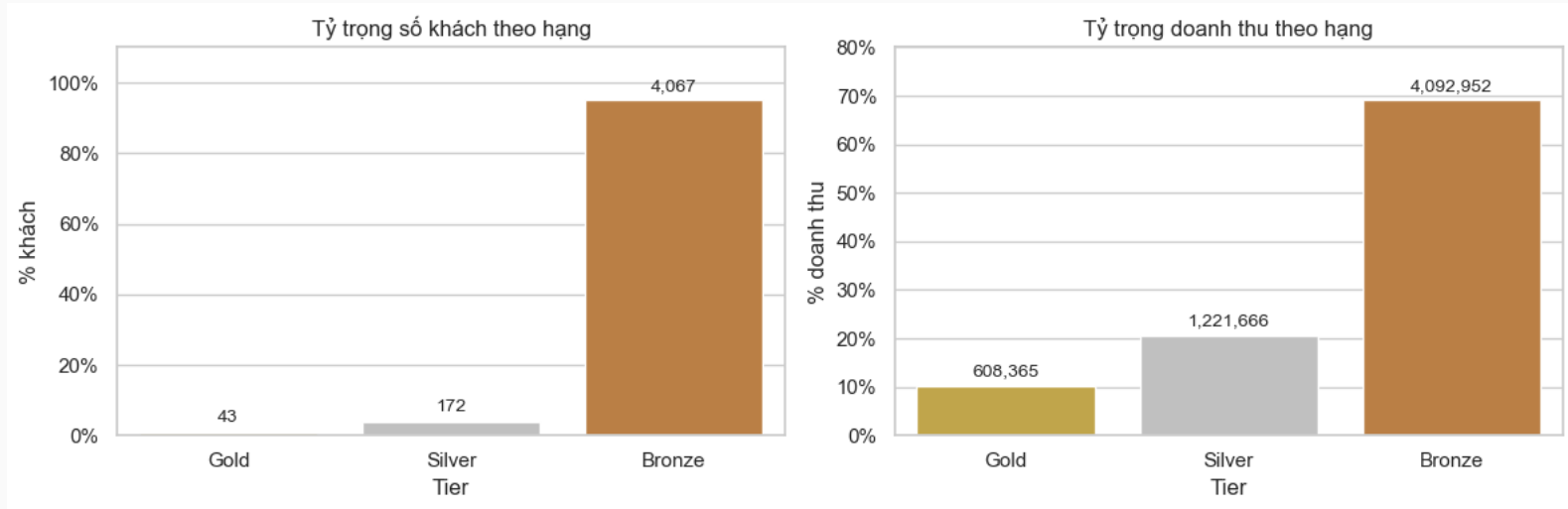


Hình 8.

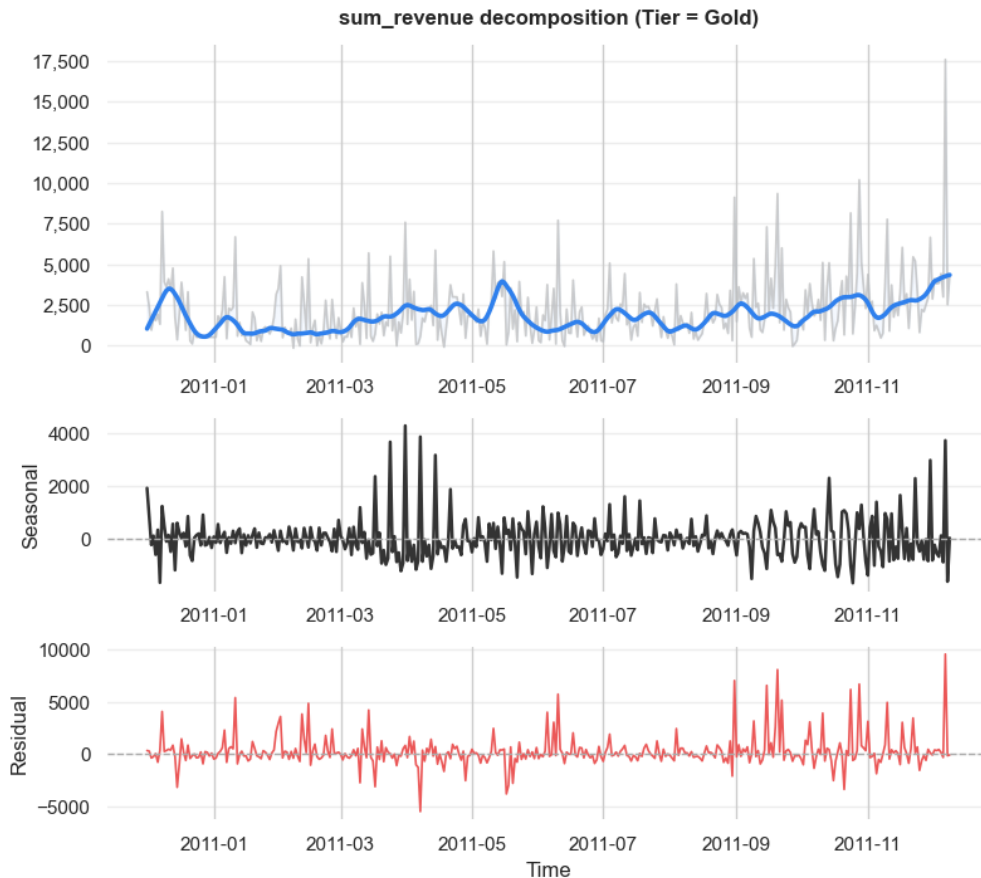
Top 10 quốc gia theo doanh thu gộp. Kết quả phản ánh mức độ tập trung thị trường và là cơ sở để thảo luận rủi ro phụ thuộc thị trường chính.

# Phân tích theo Tier

Chia nhóm Tier thành Gold – Silver - Bronze



# Phân tích theo Tier



## Decomposition (STL)

- Period: 7 (D)

## Trend

- $\Delta = 3,305$
- $\approx 157.6\%$  of mean
- slope  $\approx 3.36$  / step

## Seasonality

- amplitude: 5,991 (peak-to-trough)

## Residual

- volatility: 1,535 (std)

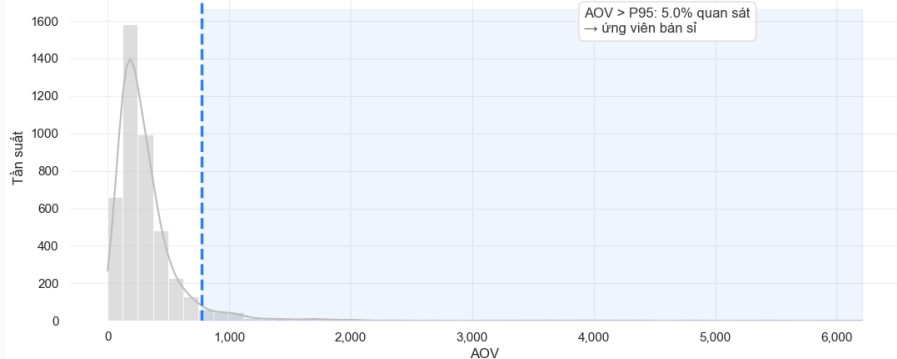
## Variability share

- Trend: 19%
- Seasonality: 17%
- Residual: 64%

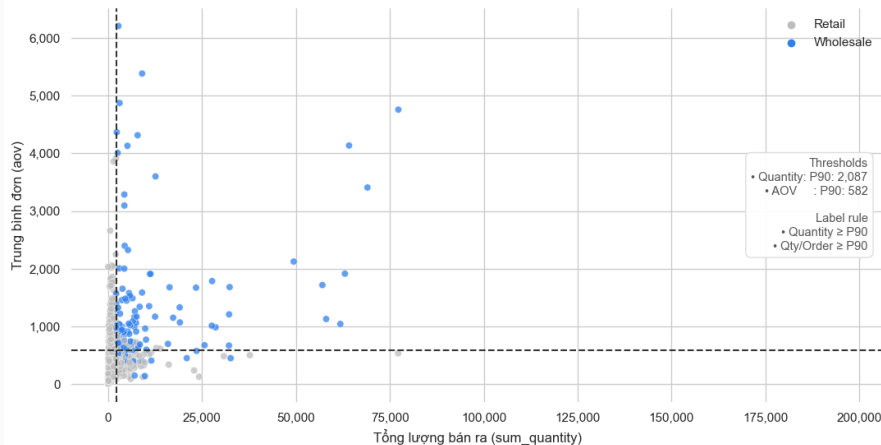
# Phân tích theo BuyerType

Nghi ngờ nhóm 5% là bán sỉ

AOV Distribution – Highlight nhóm nghi ngờ bán sỉ

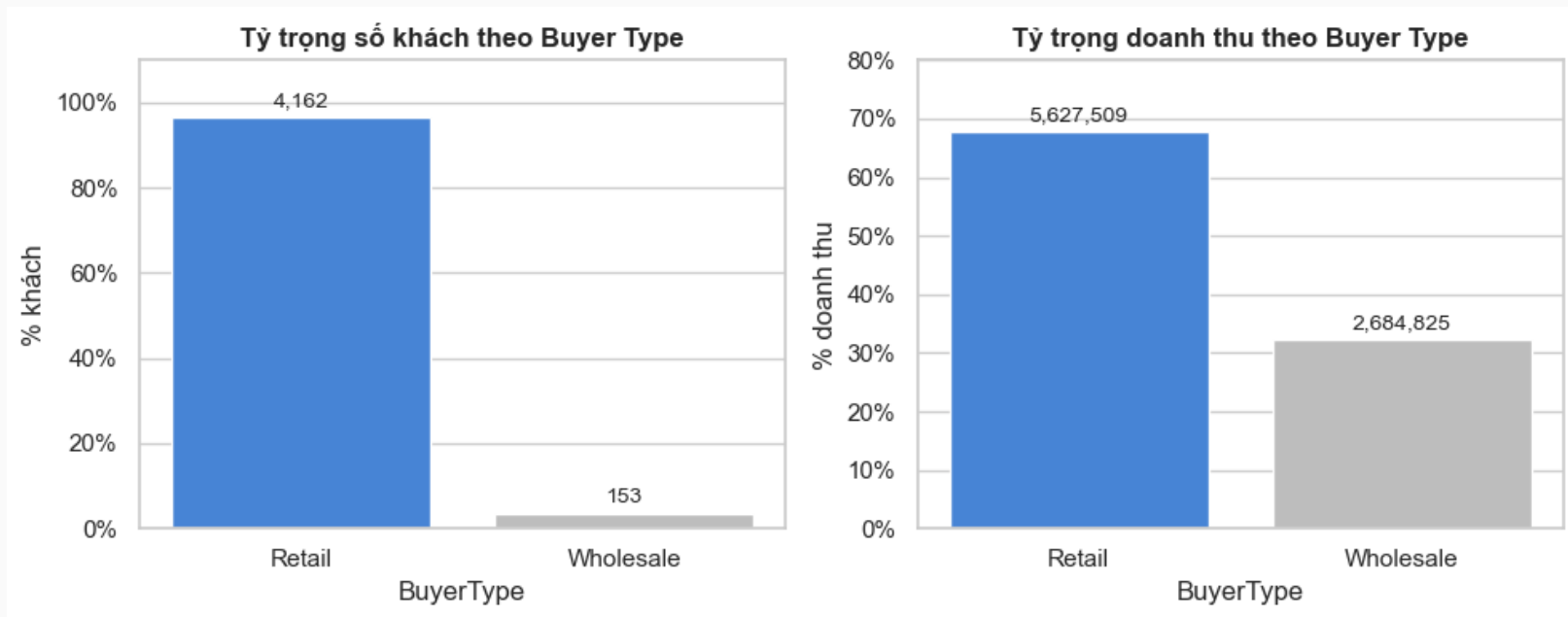


Quadrant: Tổng số lượng × AOV - Nhận diện khách hàng có hành vi mua sỉ



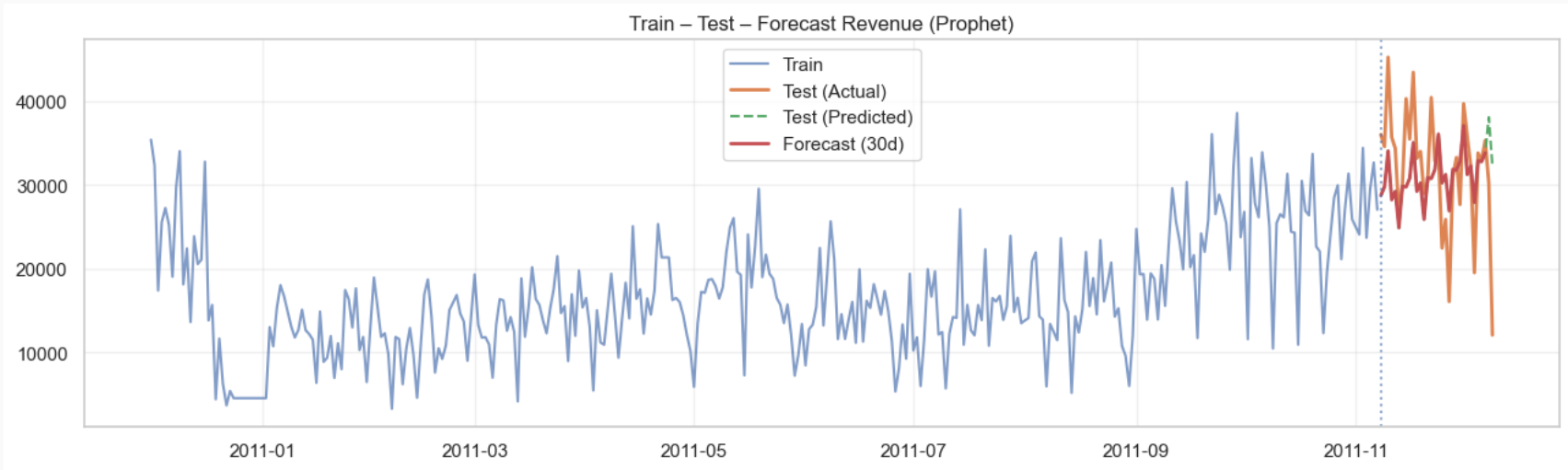
# Phân tích theo BuyerType

Chia nhóm Bán lẻ và nhóm Bán sỉ



# Phân tích theo BuyerType

Dự đoán nhóm doanh thu của tập bán lẻ





## **4. INSIGHTS**

## 4.1. NHẬN ĐỊNH VÀ RỦI RỎ

# Nhận định chính

Doanh thu có xu hướng tăng trưởng tích cực nhưng phụ thuộc vào một số sản phẩm và khách hàng chủ chốt.

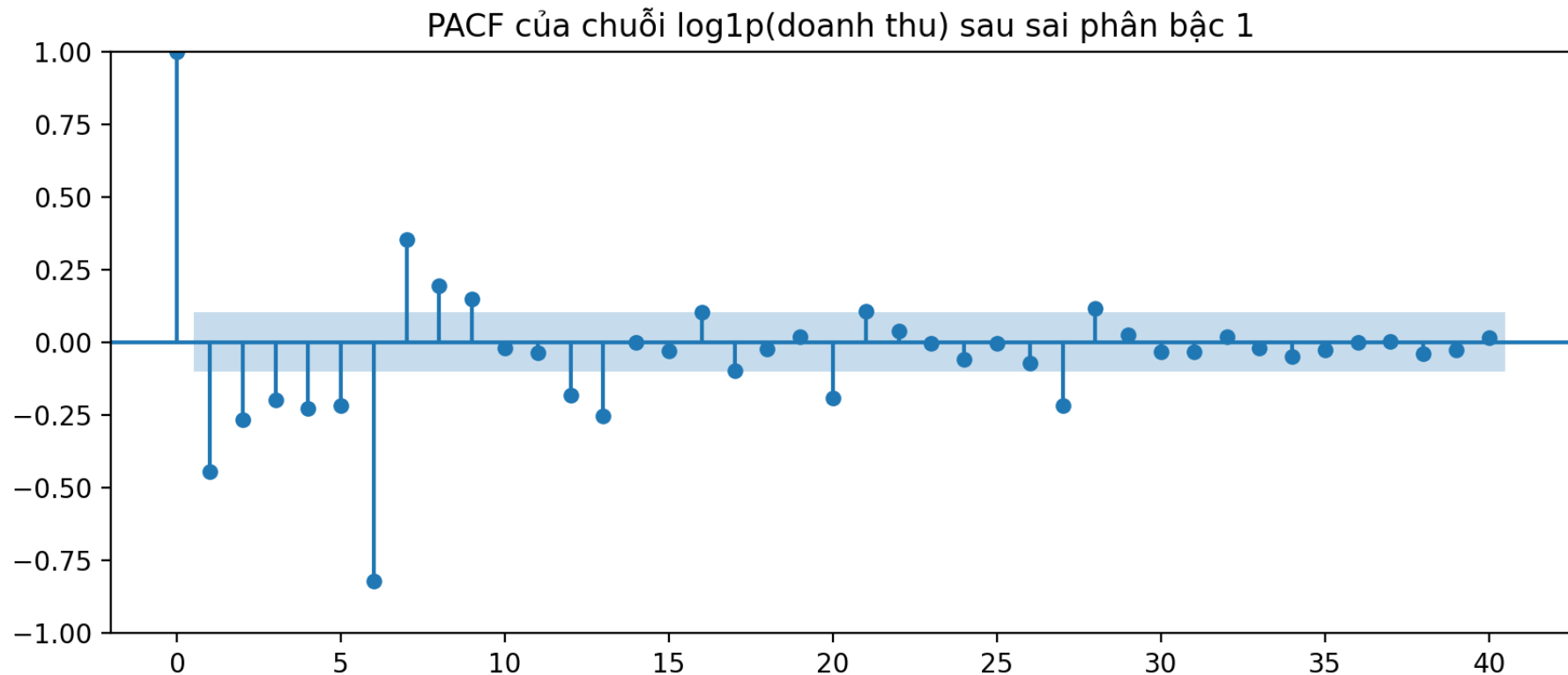
Yếu tố mùa vụ đóng vai trò quan trọng trong việc hình thành các pha tăng – giảm doanh thu trong năm, cho thấy cơ cấu doanh thu chưa thực sự cân bằng.

# Rủi ro tiềm ẩn

Mức độ tập trung doanh thu cao làm gia tăng rủi ro khi mất khách hàng lớn hoặc khi nhu cầu sản phẩm chủ lực suy giảm.

Danh mục sản phẩm phân tán với nhiều sản phẩm hiệu quả thấp có thể làm giảm hiệu quả vận hành.

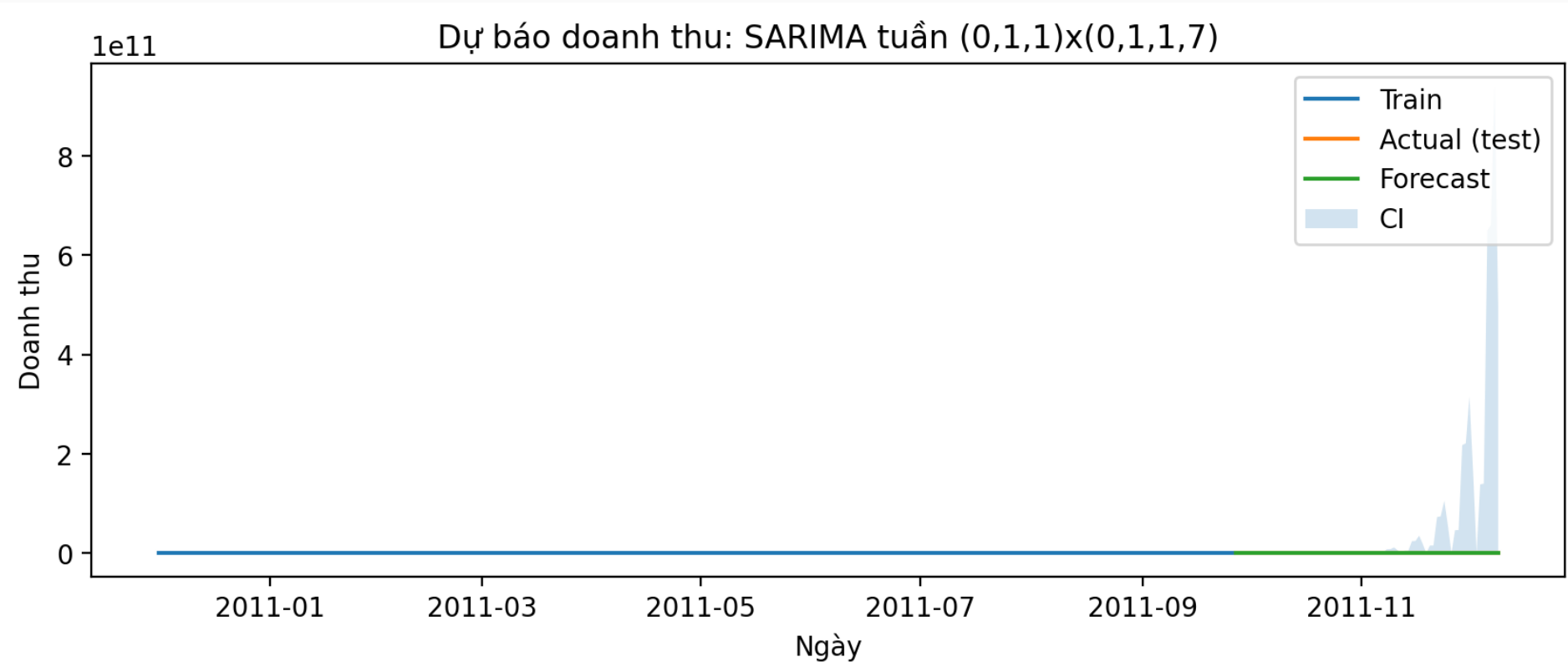
## 4.2. PHÂN TÍCH VÀ ĐÁNH GIÁ KẾT QUẢ DỰ BÁO BẰNG MÔ HÌNH SARIMA



Hình 9. Hàm tự tương quan (ACF) và tự tương quan riêng phần (PACF) của chuỗi doanh thu sau khi sai phân

## **B. SARIMA – Mô hình thứ nhất (Cơ sở)**

Mô hình cơ sở có cấu trúc gọn nhẹ, tập trung nắm bắt xu hướng chung và mùa vụ cơ bản. Phù hợp để mô tả tổng thể nhưng còn hạn chế trong các giai đoạn biến động mạnh.



Hình 10. Kết quả dự báo doanh thu ngắn hạn bằng mô hình SARIMA thứ nhất



# SARIMA – Mô hình thứ hai (Cải tiến)

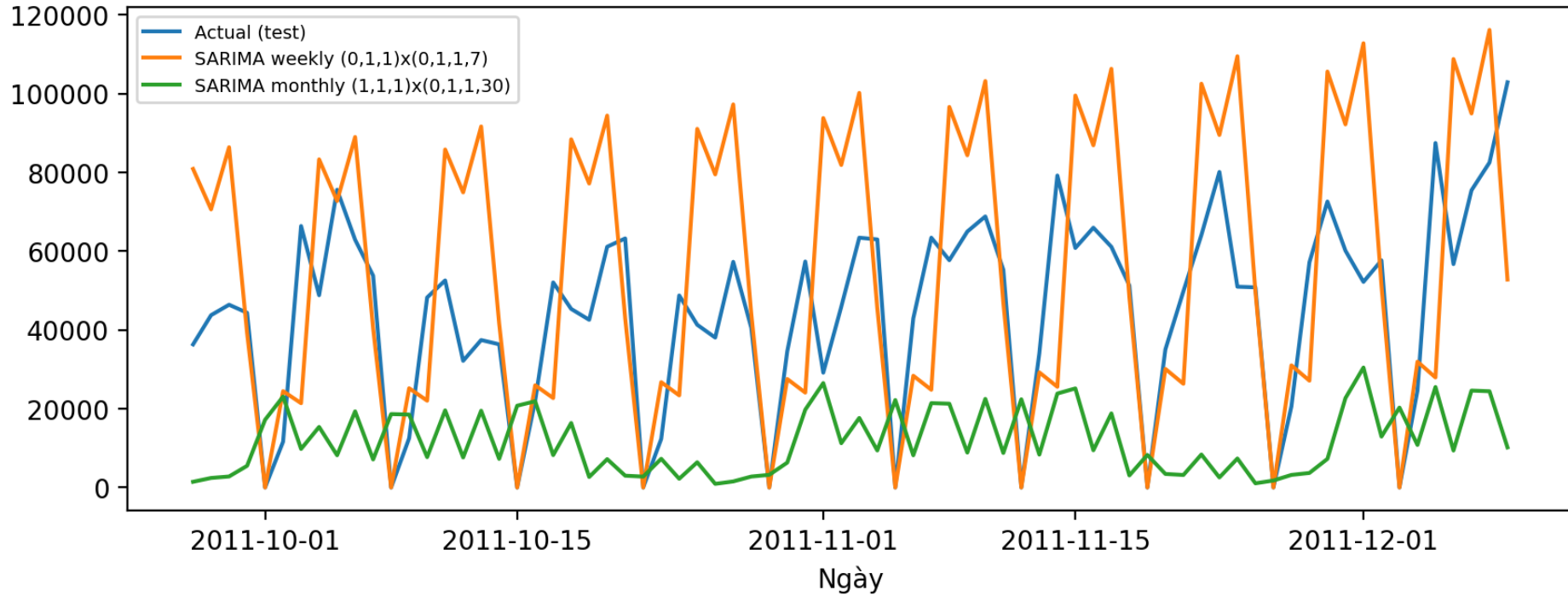
Mô hình cải tiến điều chỉnh tham số để phản ánh tốt hơn dao động ngắn hạn và mùa vụ phức tạp. Kết quả cho thấy AIC/BIC thấp hơn và sai số dự báo nhỏ hơn.

# So sánh hai mô hình SARIMA

SARIMA 1 phù hợp giám sát xu hướng dài hạn nhờ tính đơn giản và khả năng diễn giải tốt

SARIMA 2 phù hợp hơn cho dự báo ngắn hạn trong bối cảnh biến động cao

So sánh dự báo hai mô hình SARIMA trên tập kiểm tra



Hình 12. So sánh kết quả dự báo của hai mô hình SARIMA với doanh thu thực tế

# Hàm ý quản trị từ kết quả dự báo

- ★ Dự báo giúp doanh nghiệp chủ động chuẩn bị tồn kho và nguồn lực ở các giai đoạn cao điểm, đồng thời giảm rủi ro tồn kho dư thừa ở giai đoạn thấp điểm.
- ★ Nhấn mạnh tầm quan trọng của việc lựa chọn mô hình phù hợp
- ★ Chuyển từ cách quản lý dựa trên kinh nghiệm sang quản lý dựa trên dữ liệu → nâng cao tính chủ động và hiệu quả

# Hạn chế & Hướng phát triển

Mô hình hiện tại chưa xét biến ngoại sinh như khuyến mãi hay yếu tố vĩ mô hay chương trình đặc biệt.  
Hướng tiếp theo là mở rộng SARIMAX hoặc kết hợp các phương pháp học máy.

## **4.3. ĐẶC ĐIỂM CÁC CHIỀU**

# Đặc điểm doanh thu theo Tier khách hàng

## Gold

- **Trụ cột doanh thu**, xu hướng tăng mạnh và ổn định
- Mùa vụ thấp → **dễ dự báo**, phù hợp **hoạch định dài hạn**

## Silver

- **Tăng trưởng dương nhưng vừa phải**
- Mùa vụ rõ → nhạy với **khuyến mãi & chu kỳ tiêu dùng**

## Bronze

- **Xu hướng suy giảm**, biến động cao
- Nhiều lớn → **khó dự báo**, chủ yếu tạo doanh thu ngắn hạn

## Hàm ý quản trị

- Gold: nền tảng cung ứng ổn định
- Silver & Bronze: nguồn biến động và **động lực tăng trưởng ngắn hạn**, cần chiến lược linh hoạt

# Đặc điểm doanh thu theo Buyer Type

## Bán lẻ (Retail)

- Xu hướng tăng nhẹ, doanh thu ổn định
- Mùa vụ theo tuần rõ → phù hợp dự báo ngắn hạn & quản trị tồn kho linh hoạt

## Bán sỉ (Wholesale)

- Xu hướng suy giảm, biến động rất cao
- Phụ thuộc đơn hàng theo đợt → rủi ro dự báo lớn

## Hàm ý quản trị

- Động lực doanh thu dịch chuyển từ bán sỉ sang bán lẻ trong ngắn hạn
- Cần ưu tiên khả năng phản ứng nhanh cho kênh bán lẻ, đồng thời rà soát chiến lược khách hàng sỉ



## **5. KẾT LUẬN**

# Kết luận

- ❑ **Doanh thu có xu hướng tăng theo thời gian** nhưng chịu ảnh hưởng rõ rệt của **mùa vụ và biến động ngắn hạn**, đồng thời tập trung vào một số **sản phẩm và nhóm khách hàng chủ chốt**, phản ánh đặc trưng điển hình của bán lẻ.
- ❑ **Động lực doanh thu khác biệt rõ rệt giữa các phân khúc**: nhóm khách hàng giá trị cao và kênh bán lẻ đóng vai trò nền tảng với mức độ ổn định và khả năng dự báo tốt hơn, trong khi các nhóm còn lại biến động mạnh và tiềm ẩn rủi ro hoạch định.
- ❑ **Mô hình SARIMA cho kết quả dự báo ngắn hạn hiệu quả**, nắm bắt tốt xu hướng và mùa vụ, cung cấp cơ sở định lượng cho các quyết định về tồn kho, phân bổ nguồn lực và lập kế hoạch bán hàng trong quản trị chuỗi cung ứng.

CẢM ƠN