

TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN
KHOA CÔNG NGHỆ THÔNG TIN
Cao học khóa 34
Các mô hình phân tích dữ liệu



BÁO CÁO ĐỒ ÁN MÔN HỌC
Phân tích dữ liệu bán lẻ trực tuyến

Giảng viên hướng dẫn:	TS. Nguyễn Trần Minh Thư
Nhóm thực hiện:	Nhóm 04 Phạm Phong Hào - 24C12006 Trần Đình Sang - 24C11060
Ngành:	Hệ thống thông tin

TP. Hồ Chí Minh, Tháng 1/2026

Mục lục

1	Bối cảnh và Mục tiêu	1
1.1	Bối cảnh chung	1
1.2	Các mục tiêu nghiệp vụ	1
1.2.1	Phân tích doanh thu trên dòng thời gian chung	1
1.2.2	Phân tích doanh thu theo từng phân khúc của khách hàng	3
2	Dữ liệu nghiên cứu	5
2.1	Tổng quan về tập dữ liệu	5
2.2	Các nhận định trên dữ liệu	6
3	Xử lý dữ liệu	7
3.1	Nhận định về tính chất time-series của tập dữ liệu	7
3.2	Xử lý dữ liệu thiếu	8
3.2.1	Thuật toán điền giá trị thiếu bằng trung bình trượt (Moving Average)	8
3.2.2	Kết quả điền thiếu	9
4	Phương pháp tiếp cận	10
4.1	Phạm vi phân tích	10
4.2	Phân rã STL	10
4.3	Các mô hình dự đoán	11
4.3.1	SARIMA	11
4.3.2	Prophet	12
5	Phân tích tính chất doanh thu chung	13
5.1	Tổng quan dữ liệu	13
5.2	Xu hướng doanh thu theo thời gian	13
5.3	Phân tích theo sản phẩm	15
5.4	Phân tích theo khách hàng	16
5.5	Phân tích theo quốc gia	17
6	Phân tích tính chất doanh thu theo phân khúc khách hàng	18
6.1	Phân khúc khách hàng	18
6.2	Phân tích doanh thu theo chiều Tier	19
6.2.1	Tỷ trọng các nhóm khách hàng theo Tier	19
6.2.2	Phân rã STL theo từng Tier	20
6.3	Phân tích doanh thu theo chiều BuyerType	24
6.3.1	Tỷ trọng các nhóm khách hàng theo BuyerType	24
6.3.2	Phân rã STL theo từng BuyerType	25
6.3.3	Dự đoán doanh thu 30 ngày theo từng BuyerType	27

7	Insights và Quyết định	31
7.1	Insights: Phân tích doanh thu xu hướng chung	31
7.1.1	Nhận định và Rủi ro	31
7.1.2	Phân tích và đánh giá kết quả dự báo bằng mô hình SARIMA . . .	31
7.2	Insights: Phân tích doanh thu theo phân khúc khách hàng	34
7.2.1	Đặc điểm nhóm Tier	34
7.2.2	Đặc điểm nhóm bán sỉ/lẻ	35
7.2.3	Đánh giá hiệu suất dự báo và Lựa chọn mô hình	36
8	Thảo luận	37
8.1	Thảo luận về đặc điểm doanh thu chung	37
8.1.1	Thảo luận về kết quả phân tích	37
8.1.2	Thảo luận về kết quả dự báo bằng SARIMA	37
8.1.3	Hàm ý tổng hợp	37
8.2	Thảo luận về doanh thu theo nhóm khách hàng	38
9	Kết luận và Hướng phát triển	39
9.1	Kết luận	39
9.2	Hướng phát triển	39
10	Phụ lục	40

Chương 1

Bối cảnh và Mục tiêu

1.1 Bối cảnh chung

Trong môi trường bán lẻ trực tuyến, mỗi giao dịch không chỉ đơn thuần là một lần mua bán mà còn là một tín hiệu phản ánh nhu cầu thị trường, hành vi khách hàng và mức độ biến động của doanh thu theo thời gian. Toàn bộ những tín hiệu này được ghi lại trong bộ dữ liệu *Online Retail*, bao gồm hơn 500,000 dòng giao dịch phát sinh trong vòng một năm hoạt động của doanh nghiệp.

Tuy nhiên, dữ liệu chỉ thực sự có ý nghĩa khi nó giúp trả lời các câu hỏi nghiệp vụ cốt lõi. Đối với doanh nghiệp này, ba vấn đề mang tính sống còn được đặt ra: (1) làm thế nào để dự đoán nhu cầu trong ngắn hạn nhằm chuẩn bị hàng tồn kho phù hợp; (2) làm thế nào để nhận diện xu hướng và mức độ biến động của doanh thu theo thời gian; và (3) nhóm khách hàng nào thực sự tạo ra giá trị bền vững và xứng đáng được ưu tiên đầu tư. Việc thiếu các phân tích dựa trên dữ liệu có thể khiến doanh nghiệp rơi vào trạng thái bị động, dẫn đến tồn kho dư thừa, mất cơ hội doanh thu do thiếu hàng, hoặc phân bổ nguồn lực marketing kém hiệu quả.

Xuất phát từ bối cảnh đó, đề án này hướng tới mục tiêu chuyển hóa dữ liệu giao dịch thô thành các insight có khả năng hỗ trợ ra quyết định trong quản lý chuỗi cung ứng và hoạt động kinh doanh. Thay vì dừng lại ở việc mô tả dữ liệu, nghiên cứu tập trung xây dựng các mục tiêu kinh doanh rõ ràng cùng một lộ trình phân tích có cấu trúc.

Chúng tôi nghiên cứu và phân tích về doanh thu trên dòng thời gian 1 năm. Mục tiêu chính là **Đánh giá hiệu quả doanh thu trong 1 năm, xác định các giai đoạn cao-thấp điểm, các yếu tố đóng góp và các nguyên nhân làm suy giảm doanh thu để hỗ trợ tối ưu hoạt động bán hàng trong ngắn hạn**. Chúng tôi xác định có hai mục tiêu con phân tích về tính chất của doanh thu trên dòng thời gian chung và theo phân khúc của khách hàng.

1.2 Các mục tiêu nghiệp vụ

1.2.1 Phân tích doanh thu trên dòng thời gian chung

Business Goal 1: Trong vòng 2 tuần phân tích, xây dựng một báo cáo phân tích chuỗi thời gian doanh thu theo ngày–tuần–tháng nhằm xác định rõ trend, seasonality và mức độ biến động ngắn hạn. Từ đó hỗ trợ nhà quản lý phân biệt được tăng trưởng thực sự và biến động ngẫu nhiên, làm cơ sở cho các quyết định điều chỉnh chiến lược kinh doanh trong năm hiện tại.

Mục tiêu: Phân tích nhằm làm rõ cách doanh thu thay đổi theo thời gian, không chỉ ở mức độ quan sát biến động mà còn ở mức độ nhận diện bản chất của các thay đổi đó. Cụ thể, mục tiêu là xác định các giai đoạn tăng trưởng và suy giảm có tính hệ thống, các chu kỳ lặp lại theo tuần hoặc theo tháng, cũng như đánh giá mức độ biến động ngắn hạn xung quanh xu hướng chung. Qua đó, phân tích giúp chuyển từ việc “nhìn thấy doanh thu tăng hoặc giảm” sang việc “hiểu vì sao doanh thu thay đổi và mức độ ổn định của sự thay đổi đó”.

Bài toán nghiệp vụ: Doanh thu biến động đáng kể theo ngày và theo tuần, gây khó khăn trong việc đánh giá liệu sự tăng/giảm doanh thu là xu hướng thực sự hay chỉ là dao động ngắn hạn, từ đó ảnh hưởng đến các quyết định điều chỉnh giá, chương trình bán hàng và phân bổ nguồn lực.

Câu chuyện dữ liệu: Khi quan sát thấy doanh thu tăng mạnh trong một hoặc vài tuần liên tiếp, liệu đây có phải là dấu hiệu khởi đầu của một giai đoạn tăng trưởng mới hay chỉ là một biến động ngắn hạn do các yếu tố tạm thời. Tương tự, những giai đoạn doanh thu suy giảm đặt ra nghi vấn đó là xu hướng suy giảm cần can thiệp chiến lược hay chỉ là dao động thông thường quanh một xu hướng ổn định. Phân tích dữ liệu chuỗi thời gian được sử dụng để trả lời những câu hỏi này một cách có căn cứ, thay vì dựa trên cảm nhận hoặc so sánh ngắn hạn.

SMART Goals:

- *Specific:* Phân tích là làm rõ cấu trúc biến động doanh thu theo thời gian bằng cách xác định trend, seasonal và các dao động ngắn hạn. Phân biệt được đâu là xu hướng tăng trưởng hoặc suy giảm mang tính hệ thống, đâu là biến động ngẫu nhiên xảy ra trong ngắn hạn. Từ đó giúp nhà quản lý hiểu đúng bản chất nhịp độ doanh thu theo ngày, tuần và tháng.
- *Measurable:* Đánh giá thông qua các chỉ số định lượng, bao gồm tốc độ tăng trưởng doanh thu tuần và tháng, mức độ biến động doanh thu được đo lường bằng các thống kê mô tả và rolling statistics, cũng như kết quả phân rã chuỗi thời gian. Dựa trên các chỉ số này, doanh thu được phân loại thành các giai đoạn tăng trưởng bền vững, suy giảm có hệ thống hoặc biến động ngẫu nhiên, cho phép kiểm chứng và so sánh kết quả phân tích một cách khách quan.
- *Achievable:* Phân tích có thể thực hiện được vì dữ liệu doanh thu theo thời gian đã được ghi nhận đầy đủ trong suốt 12 tháng và có độ chi tiết theo ngày. Các phương pháp được sử dụng đều là những kỹ thuật phân tích chuỗi thời gian mô tả phổ biến và phù hợp với quy mô dữ liệu, bao gồm làm mượt chuỗi bằng moving average, phân tích thống kê trượt và phân rã chuỗi thời gian nội năm, không đòi hỏi các mô hình dự báo phức tạp hay nguồn lực tính toán lớn.
- *Relevant:* Kết quả phân tích có ý nghĩa trực tiếp đối với hoạt động quản trị kinh doanh, giúp nhà quản lý tránh đưa ra các quyết định dựa trên những biến động ngắn hạn mang tính ngẫu nhiên. Việc hiểu rõ xu hướng và mùa vụ doanh thu giúp cho việc ra quyết định điều chỉnh giá, triển khai chương trình bán hàng, cũng như phân bổ nguồn lực phù hợp theo từng giai đoạn trong năm, qua đó nâng cao hiệu quả theo dõi và điều hành kinh doanh trong ngắn hạn.
- *Time-bound:* Phân tích được thực hiện trong 2 tuần, sử dụng toàn bộ dữ liệu doanh thu của 12 tháng trong năm hiện tại. Kết quả được tổng hợp thành một báo cáo hoàn chỉnh, bao gồm đánh giá tổng quan xu hướng cả năm và phân tích chi tiết

theo từng tháng, đồng thời có thể được tái sử dụng như công cụ theo dõi định kỳ để hỗ trợ các quyết định quản lý trong các chu kỳ tiếp theo.

1.2.2 Phân tích doanh thu theo từng phân khúc của khách hàng

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.

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: “Hiện tại, tập khách hàng hệ thống (system customer) vẫn chưa được phân tích đầy đủ ở góc độ cấu trúc và hành vi, do các phân tích doanh thu trước đây chủ yếu dừng ở mức tổng hợp, chưa phản ánh được đặc điểm bên trong của tập khách hàng. Do đó, nghiên cứu này tiến hành phân chia khách hàng theo hai chiều Tier (Vàng – Bạc – Đồng) và BuyerType (Sỉ – Lẻ), nhằm đa dạng hóa góc nhìn và khai thác sâu hơn tiềm năng của từng phân khúc. Việc phân nhóm theo hai chiều này giúp làm rõ sự khác biệt về hành vi mua và mức độ đóng góp doanh thu. Trên từng nhóm khách hàng sau phân loại, phân tích chuỗi thời gian được thực hiện để tách các thành phần trend, seasonality và residual. Kết quả cho thấy nhóm system customer ở hầu hết các Tier vẫn đóng góp phần lớn doanh thu và đóng vai trò nền tảng, tuy nhiên xu hướng doanh thu của nhóm này có dấu hiệu chững lại trong các tháng gần đây, với dự báo ngắn hạn thiên về đi ngang hoặc giảm nhẹ. Ngược lại, nhóm walk-in customer tuy chiếm tỷ trọng doanh thu thấp hơn nhưng có mức biến động và seasonality cao, đóng góp đáng kể vào các nhịp tăng trưởng ngắn hạn ở một số Tier. Điều này cho thấy động lực tăng trưởng doanh thu đang có xu hướng dịch chuyển từ nhóm khách hàng nền tảng sang các phân khúc linh hoạt hơn trong ngắn hạn.”

SMART Goals.

- *Specific:* Phân tích và dự báo doanh thu theo chuỗi thời gian cho từng phân nhóm khách hàng được xác định theo hai chiều *Tier* (Vàng, Bạc, Đồng) và *BuyerType* (system customer, walk-in customer), nhằm làm rõ vai trò nền tảng và động lực tăng trưởng của từng nhóm.
- *Measurable:* Đánh giá độ chính xác dự báo trên từng chuỗi doanh thu theo phân nhóm khách hàng, với chỉ số sai số MAPE không vượt quá 15% ở giai đoạn kiểm định.
- *Achievable:* Áp dụng các phương pháp phân tích chuỗi thời gian phù hợp với dữ liệu quy mô trung bình, bao gồm phân rã STL kết hợp với các mô hình dự báo đơn biến (Holt–Winters, SARIMA), nhằm nắm bắt xu hướng, tính mùa vụ và biến động ngẫu nhiên của doanh thu.
- *Relevant:* Kết quả phân tích và dự báo cung cấp cơ sở định lượng để doanh nghiệp đánh giá mức độ chững lại của nhóm *system customer* và tiềm năng tăng trưởng

ngắn hạn của nhóm *walk-in customer* theo từng Tier, từ đó hỗ trợ hoạch định chiến lược chăm sóc khách hàng và phân bổ nguồn lực kinh doanh.

- *Time-bound*: Huấn luyện mô hình trên dữ liệu lịch sử, dự đoán tương lai, kế hoạch dự trù phát triển khoảng 1 tuần.

Chương 2

Dữ liệu nghiên cứu

2.1 Tổng quan về tập dữ liệu

Bảng 2.1: Bảng phân tích thuộc tính dữ liệu (Data Info)

Cột	Non-null	Kiểu dữ liệu	Nhận xét
InvoiceNo	541,909	object	Đúng chuẩn: dạng string, gồm số và ký tự 'C' cho đơn hủy
StockCode	541,909	object	Hợp lệ
Description	540,455	object	Có ~1.454 dòng thiếu mô tả
Quantity	541,909	int	Có thể âm (đơn trả hàng)
InvoiceDate	541,909	object → datetime	Không có lỗi ngày tháng
UnitPrice	541,909	float	Có thể gồm giá bất thường
CustomerID	406,829	float	Thiếu 135.080 giá trị (khoảng 25%)
Country	541,909	object	Hợp lệ

Mô tả chi tiết các cột dữ liệu:

- **InvoiceNo:** Mã hóa đơn định danh duy nhất cho mỗi giao dịch, trong đó các mã bắt đầu bằng ký tự "C" biểu thị giao dịch bị hủy.
- **StockCode:** Mã sản phẩm định danh duy nhất cho từng mặt hàng được bán trong hệ thống.
- **Description:** Tên hoặc mô tả văn bản của sản phẩm tương ứng với từng dòng giao dịch.
- **Quantity:** Số lượng đơn vị sản phẩm được giao dịch trong mỗi hóa đơn.
- **InvoiceDate:** Thời điểm (ngày và giờ) mà giao dịch được phát sinh và ghi nhận trong hệ thống.
- **UnitPrice:** Giá bán của một đơn vị sản phẩm, được tính bằng bảng Anh (GBP).

- **CustomerID:** Mã định danh khách hàng, được sử dụng để phân biệt các khách hàng khác nhau và có thể tồn tại giá trị thiếu.
- **Country:** Tên quốc gia nơi khách hàng cư trú tại thời điểm thực hiện giao dịch.

2.2 Các nhận định trên dữ liệu

Dựa trên kết quả tiền xử lý, các thông số thống kê cơ bản được tóm tắt như sau:

- **Số dòng trùng lặp:** 5.268
- **Số dòng lỗi ngày tháng:** 0
- **Số mã sản phẩm bất thường:** 71

Nhìn chung, dữ liệu E-Commerce này có chất lượng **khá tốt** nhờ các đặc điểm:

- **Định dạng:** Không phát hiện lỗi định dạng nghiêm trọng.
- **Cấu trúc:** Dữ liệu đồng nhất, không bị vỡ cấu trúc giữa các cột.
- **Tính toàn vẹn:** Không có cột nào bị hỏng (corrupted) hoàn toàn.
- **Độ nhiễu:** Số lượng dòng trùng lặp chiếm tỷ lệ nhỏ so với tổng thể.

Lưu ý: Mặc dù trường CustomerID thiếu khá nhiều giá trị ($\sim 25\%$), nhưng điều này có thể **chấp nhận được** trong ngữ cảnh thương mại điện tử, vì không phải giao dịch nào người dùng cũng đăng nhập hoặc cung cấp thông tin định danh khách hàng.

Chương 3

Xử lý dữ liệu

3.1 Nhận định về tính chất time-series của tập dữ liệu

Đánh giá sơ bộ:

- Cột liên quan đến thời gian: InvoiceDate, datatype là timestamp.
- Mỗi dòng dữ liệu đang được định danh bởi 2 cột: InvoiceNo, StockCode.
- Đối tượng nghiên cứu mục tiêu: Khách hàng (customer), nên cần phải đưa về dạng mỗi khách hàng sẽ có một dòng thời gian mua hàng riêng.

Đánh giá chi tiết về bản chất time-series:

- Vai trò của biến thời gian: Biến InvoiceDate không đóng vai trò là time index (chỉ mục loại thời gian).
- Đơn vị quan sát (unit): Mỗi record là một sản phẩm đã mua theo giao dịch.
- Cấu trúc lấy mẫu theo thời gian: không đều, dữ liệu 1 năm.
- Quá trình sinh dữ liệu: Đây là dữ liệu ghi nhận thực tế.

Kết luận của việc đánh giá bản chất chuỗi thời gian của dữ liệu:

- Dữ liệu ban đầu không là dạng chuỗi thời gian, chỉ ở dạng transaction (giao dịch) thông thường.
- Khả năng chuyển đổi thành time-series: Có. Nếu xác định rõ đối tượng cần nghiên cứu (đối tượng phải tồn tại trong dữ liệu) thì sẽ biến đổi được về dạng time-series.

Xét về cơ bản dữ liệu này không phải dạng time-series, chúng tôi liệt kê những điều kiện để định nghĩa một tập dữ liệu gọi là time-series như sau:

Bảng 3.1: Đối chiếu điều kiện dữ liệu chuỗi thời gian

Điều kiện	Có hay Không?
Có cột timestamp	Có
Có time index	Không
Đơn vị quan sát theo thời gian	Không, mỗi record quan sát bằng InvoiceNo, StockCode
Thứ tự thời gian có ý nghĩa	Không
Tần suất xác định	Không. Việc tính tần suất trung bình không có ý nghĩa.
Phụ thuộc theo thời gian	Không
Là một stochastic process	Không

3.2 Xử lý dữ liệu thiếu

3.2.1 Thuật toán điền giá trị thiếu bằng trung bình trượt (Moving Average)

Để xử lý các khoảng trống dữ liệu trong chuỗi thời gian, hàm `fill_missing_by_moving_average` được triển khai nhằm mục đích làm mịn và điền đầy các giá trị dựa trên phương pháp trung bình trượt. Các bước thực hiện cụ thể như sau:

- **Bước 1: Chuẩn hóa dữ liệu thời gian:** Chuyển đổi cột thời gian sang định dạng `datetime` của Pandas. Đồng thời, loại bỏ các dòng dữ liệu bị lỗi thời gian (NaN) để đảm bảo tính nhất quán cho chỉ mục thời gian.
- **Bước 2: Sắp xếp và Gom nhóm theo tần suất:** Dữ liệu được sắp xếp theo trình tự thời gian tăng dần. Nếu tần suất yêu cầu là theo ngày (D), thuật toán thực hiện làm tròn thời gian về mốc 00:00:00 và gom nhóm (`groupby`) để tính tổng giá trị (ví dụ: tổng doanh thu) phát sinh trong ngày đó.
- **Bước 3: Tái chỉ mục (Reindexing) toàn bộ dải thời gian:** Tạo ra một dải thời gian liên tục từ ngày bắt đầu đến ngày kết thúc dựa trên tần suất `freq`. Những ngày không có giao dịch trong dữ liệu gốc sẽ được chèn vào dưới dạng giá trị thiếu (NaN). Một mặt nạ (mask) được tạo ra để ghi nhớ chính xác vị trí của các ngày bị thiếu này.
- **Bước 4: Tính toán trung bình trượt (Moving Average):** Sử dụng kỹ thuật của sổ trượt (rolling window) với kích thước `window` tùy chỉnh. Thuật toán hỗ trợ cả cửa sổ trượt trung tâm (`center=True`) hoặc cửa sổ trượt lùi. Tham số `min_periods` được thiết lập để đảm bảo vẫn tính được trung bình ngay cả khi cửa sổ chứa một số giá trị thiếu.
- **Bước 5: Điền giá trị (Filling):** Dựa trên cấu hình, thuật toán sẽ thực hiện:
 - Nếu `only_fill_missing_days=True`: Chỉ điền các giá trị vào những ngày bị thiếu do quá trình tái chỉ mục tạo ra.

– Nếu **False**: Điền vào tất cả các vị trí NaN hiện có trong cột giá trị.

- **Bước 6: Xử lý biên (Edge case handling)**: Trong trường hợp ở các đầu mút của chuỗi thời gian không đủ dữ liệu để tính trung bình trượt, thuật toán sử dụng phương pháp *Forward Fill* và *Backward Fill* để đảm bảo chuỗi thời gian hoàn toàn không còn giá trị thiếu.
- **Bước 7: Định dạng đầu ra**: Kết quả được trả về dưới dạng DataFrame với cấu trúc cột ban đầu hoặc được thiết lập làm chỉ mục thời gian tùy theo yêu cầu của người dùng.

Algorithm 1 Điền giá trị thiếu bằng Trung bình trượt (Moving Average)

Require: df (Dữ liệu gốc), $window$ (Kích thước cửa sổ), $freq$ (Tần suất)

Ensure: df_filled (Dữ liệu đã được làm đầy và liên tục theo thời gian)

```
1: Bắt đầu
2:  $out \leftarrow$  Chuyển đổi cột thời gian sang định dạng chuẩn và loại bỏ dòng rỗng
3: Sắp xếp  $out$  tăng dần theo thời gian
4:                                     ▷ Bước 1: Gom nhóm dữ liệu theo tần suất
5: if  $freq$  là theo Ngày then
6:    $out\_daily \leftarrow$  Gom nhóm theo ngày và tính tổng giá trị  $value\_col$ 
7: else
8:    $out\_daily \leftarrow$  Gom nhóm theo thời gian gốc và tính tổng
9: end if
10:                                     ▷ Bước 2: Tạo dải thời gian liên tục
11:  $full\_range \leftarrow$  Tạo danh sách thời gian liên tục từ  $Min(Date)$  đến  $Max(Date)$  với
    bước nhảy  $freq$ 
12:  $reindexed \leftarrow$  Chèn  $full\_range$  vào dữ liệu. Những ngày mới chèn sẽ có giá trị là
     $NaN$ 
13:  $mask\_missing \leftarrow$  Đánh dấu vị trí các ngày vừa được chèn thêm
14:                                     ▷ Bước 3: Tính toán giá trị trung bình trượt
15:  $rolling\_mean \leftarrow$  Tính trung bình trong cửa sổ  $window$  (có thể dùng center window)
16:                                     ▷ Bước 4: Điền đầy giá trị
17: for mỗi dòng  $i$  trong  $reindexed$  do
18:   if  $i$  nằm trong  $mask\_missing$  then
19:      $reindexed[i] \leftarrow rolling\_mean[i]$ 
20:   end if
21: end for
22:                                     ▷ Bước 5: Xử lý các giá trị biên còn sót lại
23:  $filled \leftarrow$  Điền các giá trị  $NaN$  còn lại bằng giá trị liền trước ( $ffill$ ) hoặc liền sau
    ( $bfill$ )
24: return  $result$  (Dữ liệu đã xử lý xong định dạng)
25: Kết thúc
```

3.2.2 Kết quả điền thiếu

Vị trí: Xem notebook `code/problem-02/CustomerAndRevenue.ipynb`

Hàm: `fill_missing_by_moving_average`, phần **Phân tích nhóm mua lẻ**.

Chương 4

Phương pháp tiếp cận

4.1 Phạm vi phân tích

Nghiên cứu được thực hiện với mục tiêu đánh giá tổng quan kết quả bán hàng theo thời gian, đồng thời phân tích cơ cấu doanh thu theo sản phẩm và khách hàng nhằm nhận diện các xu hướng và rủi ro tiềm ẩn trong hoạt động kinh doanh. Thay vì tập trung vào dự báo dài hạn, ưu tiên phân tích mô tả và khám phá dữ liệu nhằm hỗ trợ ra quyết định trong ngắn hạn, phù hợp với bối cảnh quản trị của doanh nghiệp bán lẻ trực tuyến.

Dữ liệu bao gồm hơn 540.000 giao dịch bán lẻ phát sinh trong vòng một năm. Dữ liệu được phân tích theo nhiều chiều, bao gồm thời gian, sản phẩm, khách hàng và quốc gia. Trước khi tiến hành phân tích, 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

4.2 Phân rã STL

Phân tích chuỗi thời gian được thực hiện nhằm tách và đánh giá các thành phần cấu trúc cơ bản của chuỗi dữ liệu, bao gồm xu hướng dài hạn, mùa vụ và nhiễu ngẫu nhiên. Quy trình phân tích được xây dựng dựa trên mô hình cộng, trong đó mỗi quan sát được biểu diễn như tổng của các thành phần riêng biệt.

Mô hình chuỗi thời gian cộng Chuỗi thời gian y_t tại thời điểm t được giả định tuân theo mô hình cộng:

$$y_t = T_t + S_t + R_t,$$

trong đó T_t là thành phần xu hướng (trend) phản ánh biến động dài hạn, S_t là thành phần mùa vụ (seasonality) mang tính chu kỳ, và R_t là phần dư (residual) đại diện cho nhiễu ngẫu nhiên và các biến động không giải thích được.

Tiền xử lý chuỗi thời gian Dữ liệu được sắp xếp theo thứ tự thời gian và chuẩn hóa về tần suất cố định. Các giá trị bị thiếu được xử lý thông qua nội suy theo thời gian nhằm đảm bảo tính liên tục của chuỗi, điều kiện cần cho các phương pháp phân rã và mô hình hóa tiếp theo.

Phân rã chuỗi thời gian bằng STL Phương pháp STL (Seasonal-Trend decomposition using Loess) được áp dụng để ước lượng riêng rẽ các thành phần T_t , S_t và R_t . STL cho phép xử lý mùa vụ với chu kỳ xác định trước và có khả năng chống nhiễu tốt nhờ cơ chế

hồi quy cục bộ và tùy chọn robust. Kết quả phân rã cung cấp cái nhìn rõ ràng về cấu trúc nội tại của chuỗi thời gian.

Phân tích thành phần xu hướng Thành phần xu hướng T_t được sử dụng để đánh giá hướng biến động dài hạn của chuỗi. Tốc độ tăng hoặc giảm được ước lượng thông qua hồi quy tuyến tính:

$$T_t \approx \alpha + \beta t,$$

trong đó hệ số β biểu diễn độ dốc xu hướng theo thời gian. Giá trị này được dùng để so sánh mức độ tăng trưởng hoặc suy giảm giữa các chuỗi khác nhau.

Phân tích mùa vụ Thành phần mùa vụ S_t phản ánh các mẫu lặp lại theo chu kỳ. Mức độ ảnh hưởng của mùa vụ được đánh giá thông qua biên độ dao động, thường được đo bằng chênh lệch giữa giá trị cực đại và cực tiểu trong một chu kỳ:

$$\text{Amplitude}(S) = \max(S_t) - \min(S_t).$$

Phân tích phần dư Phần dư R_t đại diện cho các biến động ngẫu nhiên không thuộc xu hướng hay mùa vụ. Đặc điểm của R_t được đánh giá thông qua các thống kê mô tả như trung bình và độ lệch chuẩn:

$$\sigma_R = \sqrt{\frac{1}{n} \sum_{t=1}^n (R_t - \bar{R})^2}.$$

Thành phần này được kỳ vọng có trung bình gần bằng 0 và không chứa cấu trúc thời gian rõ rệt.

Tái cấu trúc và dự báo chuỗi thời gian Sau khi phân rã, các thành phần được sử dụng để tái cấu trúc và dự báo chuỗi trong tương lai. Dự báo được thực hiện bằng cách ngoại suy xu hướng, lặp lại mẫu mùa vụ và mô hình hóa phần dư:

$$\hat{y}_{t+h} = \hat{T}_{t+h} + \hat{S}_{t+h} + \hat{R}_{t+h}.$$

Cách tiếp cận này cho phép tách biệt vai trò của từng thành phần trong quá trình dự báo và hỗ trợ diễn giải kết quả một cách minh bạch.

4.3 Các mô hình dự đoán

4.3.1 SARIMA

(1) Mô hình SARIMA

SARIMA (*Seasonal AutoRegressive Integrated Moving Average*) là phần mở rộng của mô hình ARIMA, cho phép mô hình hóa đồng thời tính tự hồi quy, sai phân và trung bình trượt ở cả mức độ thường và mùa vụ. Mô hình SARIMA được ký hiệu là:

$$\text{SARIMA}(p, d, q)(P, D, Q)_s$$

Trong đó, dạng tổng quát của mô hình được biểu diễn như sau:

$$\Phi_P(B^s)\phi_p(B)(1-B)^d(1-B^s)^D y_t = \Theta_Q(B^s)\theta_q(B)\varepsilon_t$$

với B là toán tử trễ, s là chu kỳ mùa vụ, ϕ_p và θ_q lần lượt là các đa thức tự hồi quy và trung bình trượt, còn ε_t là nhiễu trắng. SARIMA đặc biệt phù hợp với các chuỗi doanh thu có tính mùa vụ rõ ràng và cấu trúc tuyến tính ổn định, giúp nắm bắt tốt động lực ngắn hạn trong dữ liệu lịch sử.

4.3.2 Prophet

Mô hình Prophet

Prophet là mô hình dự báo chuỗi thời gian dạng cộng, được thiết kế để xử lý tốt dữ liệu kinh doanh có xu hướng thay đổi, mùa vụ phức tạp và nhiễu lớn. Chuỗi thời gian được biểu diễn dưới dạng:

$$y(t) = g(t) + s(t) + h(t) + \varepsilon_t$$

Trong đó:

- $g(t)$ là thành phần xu hướng, được mô hình hóa bằng hàm tuyến tính hoặc logistic có các điểm thay đổi (*change points*);
- $s(t)$ là thành phần mùa vụ (tuần, năm), được biểu diễn bằng chuỗi Fourier;
- $h(t)$ là tác động của các sự kiện đặc biệt (nếu có);
- ε_t là nhiễu ngẫu nhiên.

Prophet cho phép kiểm soát mức độ linh hoạt của xu hướng thông qua tham số *changepoint prior scale*, đồng thời tự động xử lý các khoảng trống dữ liệu và ngoại lệ, phù hợp với bối cảnh dữ liệu bán lẻ thực tế.

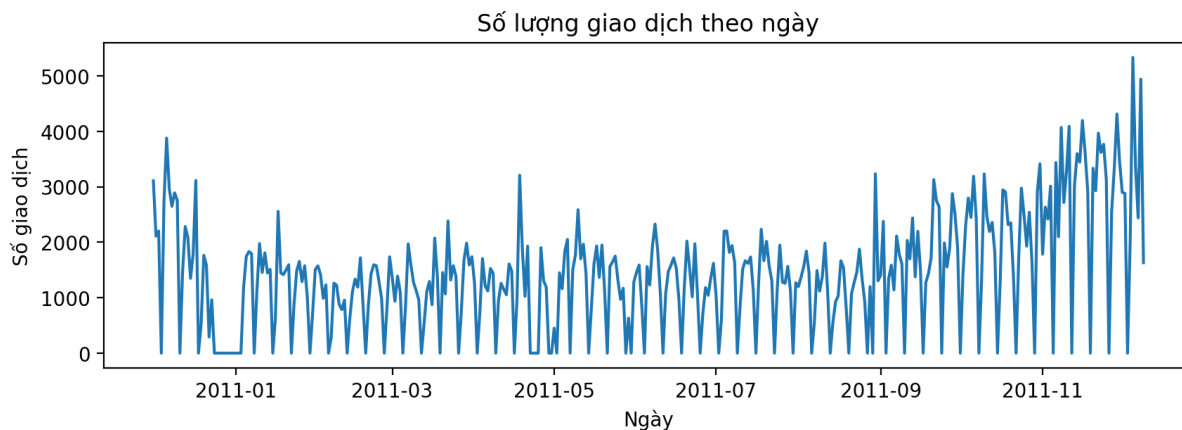
Chương 5

Phân tích tính chất doanh thu chung

5.1 Tổng quan dữ liệu

Dữ liệu giao dịch phản ánh chi tiết quá trình mua bán của khách hàng trong môi trường bán lẻ trực tuyến. Mỗi record chứa thông tin về thời điểm mua hàng, mã và mô tả sản phẩm, số lượng, đơn giá, tổng giá trị giao dịch, định danh khách hàng và quốc gia phát sinh giao dịch. Sau quá trình làm sạch, tập dữ liệu cho thấy mức độ đầy đủ và nhất quán cao, đủ điều kiện để tiến hành các phân tích đa chiều.

Từ góc độ phân tích, cấu trúc dữ liệu cho phép liên kết hành vi mua hàng của từng khách hàng với các sản phẩm cụ thể và theo dõi sự biến động doanh thu theo thời gian. Điều này tạo tiền đề cho việc khám phá các đặc điểm mùa vụ, mức độ tập trung doanh thu cũng như sự khác biệt trong hành vi tiêu dùng giữa các nhóm khách hàng.

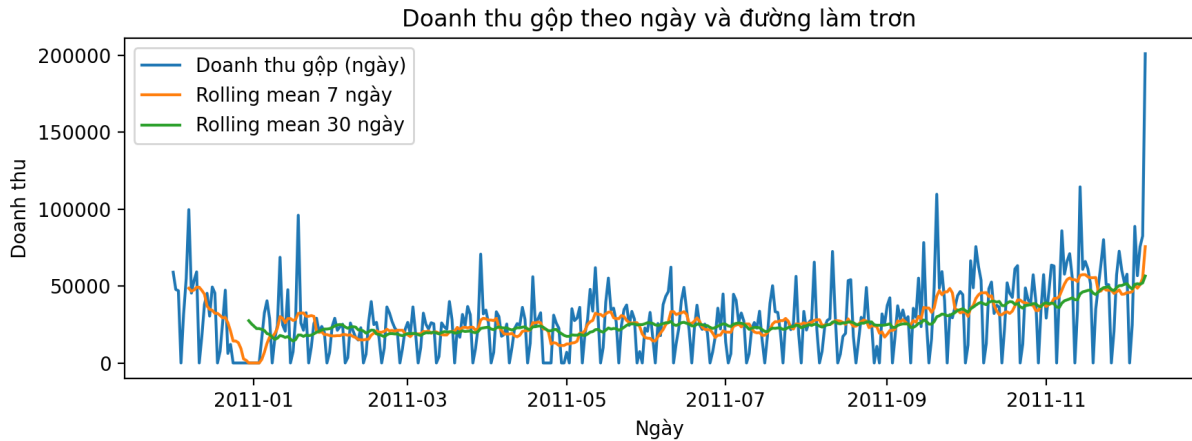


Hình 5.1: Số lượng giao dịch theo ngày trong bộ dữ liệu Online Retail, cho thấy mật độ dữ liệu theo thời gian và các khoảng biến động đáng chú ý.

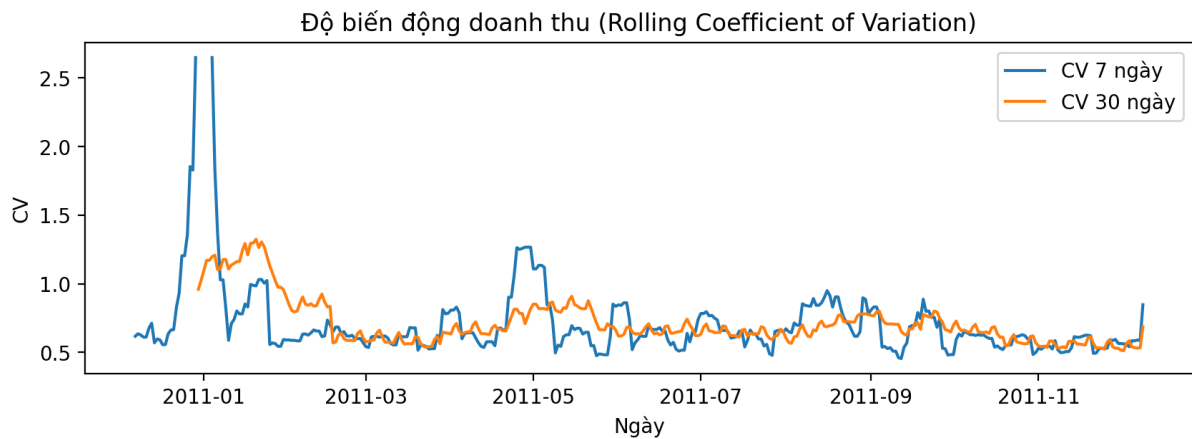
5.2 Xu hướng doanh thu theo thời gian

Kết quả phân tích cho thấy doanh thu có xu hướng tăng trưởng theo thời gian, với mức tăng rõ rệt trong các giai đoạn cuối năm. Tuy nhiên, giữa các thời điểm quan sát tồn tại những biến động đáng kể, phản ánh ảnh hưởng mạnh mẽ của yếu tố mùa vụ và các sự kiện ngắn hạn. Sự kết hợp giữa xu hướng dài hạn và dao động ngắn hạn cho thấy doanh thu không tăng theo một đường tuyến tính đơn giản mà chịu tác động của nhiều

yếu tố đồng thời.

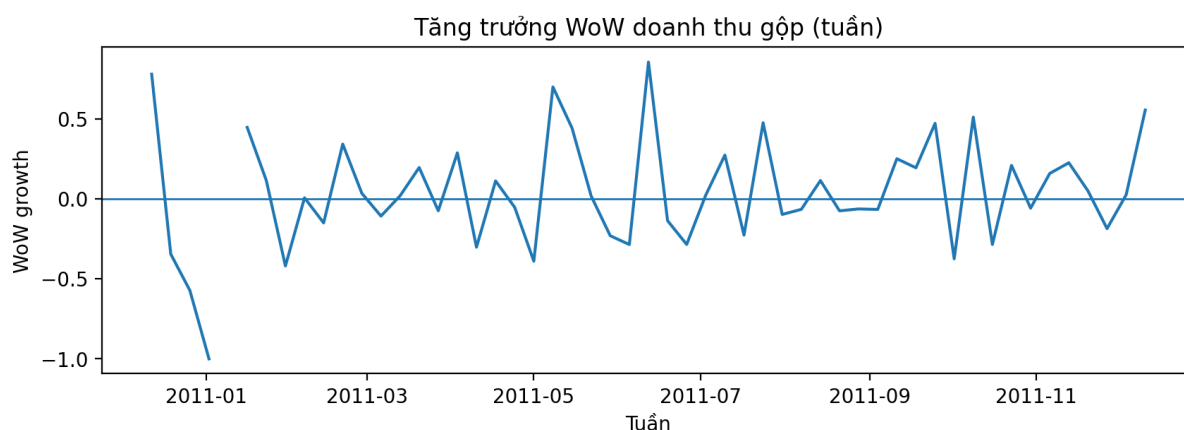


Hình 5.2: Doanh thu gộp theo ngày và đường làm trơn (7 ngày, 30 ngày). Đường làm trơn giúp làm rõ xu hướng tổng thể và giảm nhiễu ngắn hạn trong chuỗi doanh thu.

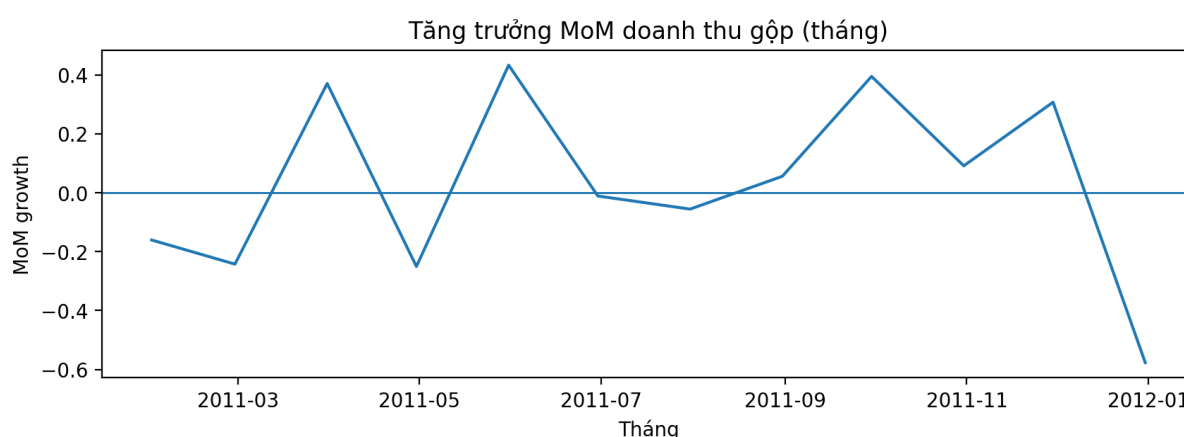


Hình 5.3: Độ biến động doanh thu theo thời gian thông qua hệ số biến thiên (CV) rolling 7 ngày và 30 ngày, phản ánh mức độ ổn định/dao động của chuỗi doanh thu.

Phát hiện này nhấn mạnh tầm quan trọng của việc xem xét yếu tố thời vụ khi đánh giá hiệu quả kinh doanh. Nếu chỉ dựa trên so sánh doanh thu giữa các giai đoạn ngắn mà không điều chỉnh theo chu kỳ mùa vụ, chúng ta có thể đưa ra các kết luận sai lệch và các quyết định điều chỉnh chưa phù hợp.



Hình 5.4: 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.

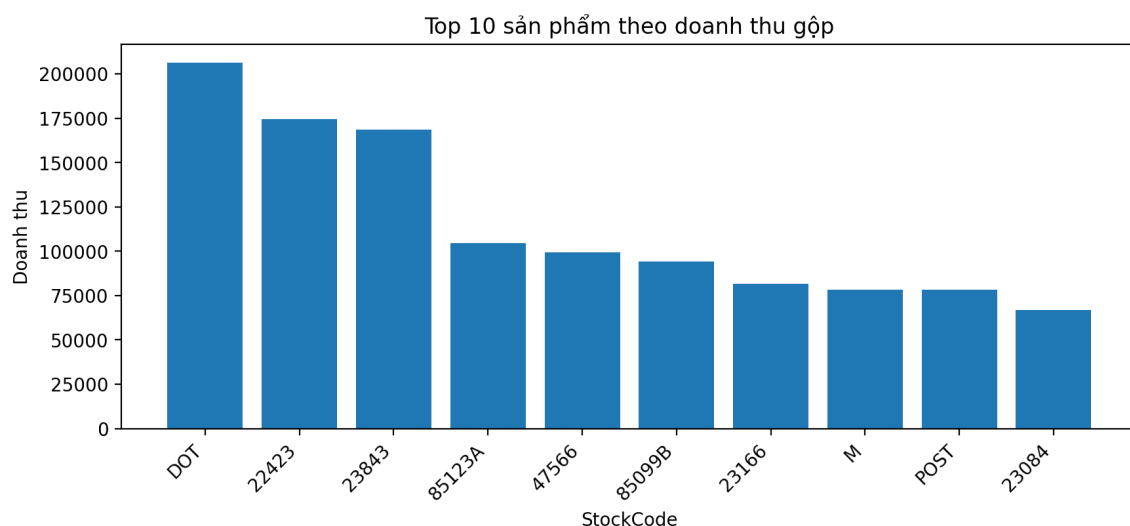


Hình 5.5: Tăng trưởng doanh thu gộp theo tháng (MoM), cung cấp góc nhìn vĩ mô hơn về xu hướng và các giai đoạn cao điểm/thấp điểm trong năm.

5.3 Phân tích theo sản phẩm

Phân tích cơ cấu doanh thu theo sản phẩm cho thấy sự phân bổ không đồng đều. Một nhóm nhỏ sản phẩm tạo ra phần lớn tổng doanh thu, trong khi nhiều sản phẩm khác chỉ đóng góp ở mức rất thấp. Điều này phản ánh đặc điểm phổ biến trong bán lẻ trực tuyến là một số sản phẩm đóng vai trò chủ lực trong tổng doanh thu.

Thông qua đó có thể thấy rằng doanh thu sẽ đạt được hiệu quả cao hơn nếu tập trung nguồn lực vào nhóm sản phẩm chủ lực, đồng thời rà soát và tinh gọn danh mục các sản phẩm kém hiệu quả. Việc duy trì quá nhiều sản phẩm có doanh thu thấp không chỉ làm tăng chi phí quản lý mà còn có thể làm phân tán sự tập trung trong chiến lược kinh doanh.

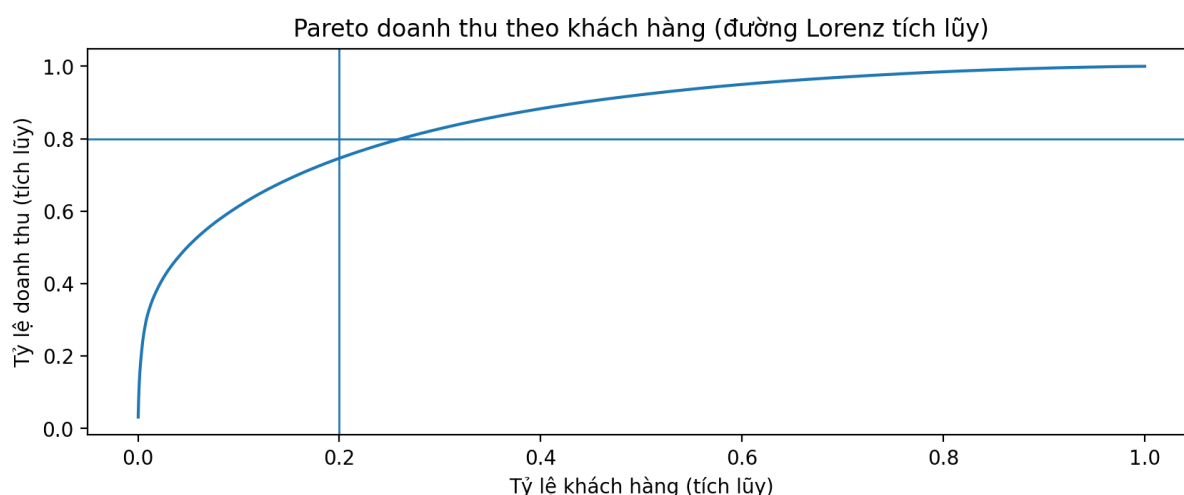


Hình 5.6: Top 10 sản phẩm theo doanh thu gộp (chỉ tính giao dịch bán). Kết quả cho thấy doanh thu tập trung vào một nhóm sản phẩm chủ lực.

5.4 Phân tích theo khách hàng

Kết quả phân tích cho thấy doanh thu tập trung mạnh vào một tỷ lệ nhỏ khách hàng, phù hợp với nguyên lý Pareto trong kinh doanh. Nhóm khách hàng này đóng góp phần lớn doanh thu và có vai trò quyết định đối với hiệu quả tài chính của doanh nghiệp.

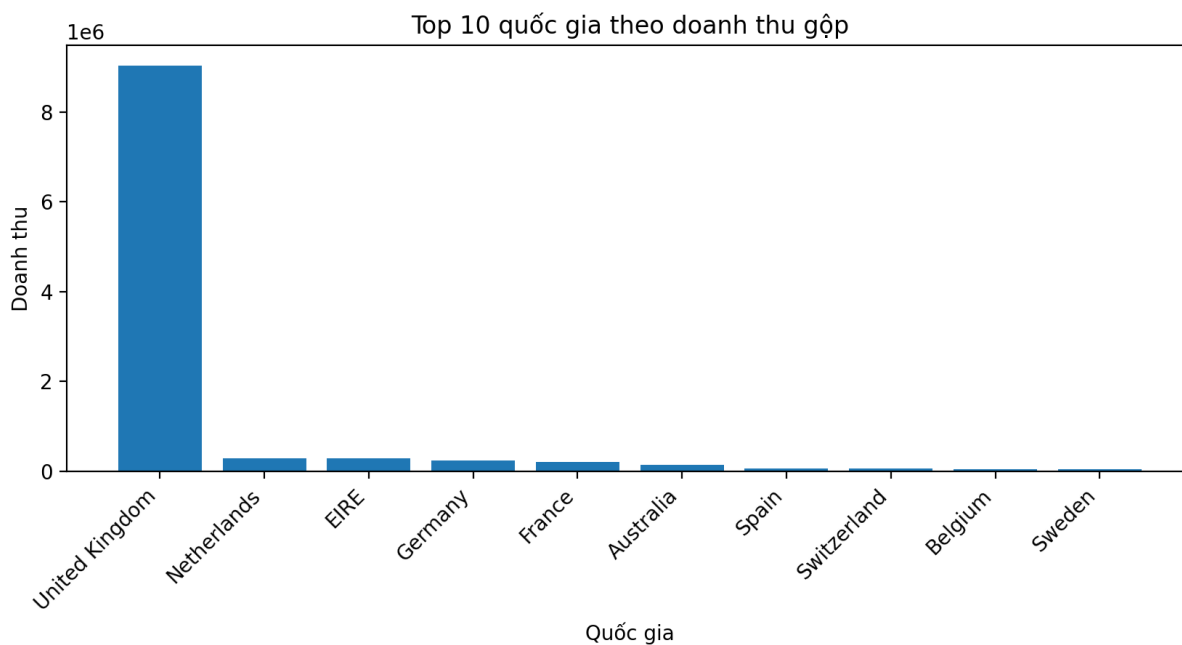
Tuy nhiên, mức độ tập trung cao cũng chứa nhiều rủi ro. Nếu chúng ta không có chiến lược duy trì và mở rộng nhóm khách hàng giá trị cao, sự thay đổi hành vi của một số ít khách hàng có thể dẫn đến biến động lớn về doanh thu. Do đó, bên cạnh việc chăm sóc nhóm khách hàng chủ chốt, cần quan tâm đến việc phát triển nhóm khách hàng trung bình nhằm tạo ra nguồn doanh thu ổn định hơn trong dài hạn.



Hình 5.7: 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.

5.5 Phân tích theo quốc gia

Xét theo quốc gia, doanh thu chủ yếu tập trung tại thị trường United Kingdom, trong khi các thị trường khác chỉ đóng góp một tỷ trọng nhỏ. Sự tập trung này mang lại lợi thế về mặt vận hành và hiểu biết thị trường, nhưng đồng thời cũng làm gia tăng rủi ro nếu nhu cầu tại thị trường chính suy giảm.



Hình 5.8: 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.

Chương 6

Phân tích tính chất doanh thu theo phân khúc khách hàng

6.1 Phân khúc khách hàng

Quy trình phân chia tập khách hàng được xây dựng nhằm phản ánh sự khác biệt về mức độ đóng góp doanh thu và hành vi mua hàng thông qua hai tiêu chí độc lập: Tier khách hàng và Buyer Type. Hai tiêu chí này được xác định dựa trên các đặc trưng định lượng rút ra từ dữ liệu giao dịch tổng hợp theo từng khách hàng.

Chuẩn hóa dữ liệu theo khách hàng Từ dữ liệu giao dịch ban đầu, các chỉ tiêu được tổng hợp ở cấp độ khách hàng, bao gồm tổng doanh thu, tổng số lượng mua, số hóa đơn và giá trị đơn hàng trung bình (AOV). Với mỗi khách hàng i , các đại lượng cơ bản được xác định như sau:

$$\begin{aligned} \text{Revenue}_i &= \sum_t \text{Revenue}_{i,t}, & \text{Quantity}_i &= \sum_t \text{Quantity}_{i,t}, \\ \text{AOV}_i &= \frac{\text{Revenue}_i}{\text{Bills}_i}, & \text{QtyPerBill}_i &= \frac{\text{Quantity}_i}{\text{Bills}_i}. \end{aligned}$$

Các giá trị không hợp lệ hoặc không dương được loại bỏ nhằm đảm bảo tính nhất quán của dữ liệu đầu vào.

Phân hạng Tier khách hàng dựa trên phân phối doanh thu Tier khách hàng được xác định dựa trên phân phối doanh thu tổng hợp theo khách hàng. Cụ thể, các ngưỡng phân vị được sử dụng để phân chia khách hàng thành ba nhóm:

- *Gold*: nhóm khách hàng có doanh thu nằm trong top $p_1\%$ của phân phối.
- *Silver*: nhóm khách hàng có doanh thu nằm trong khoảng $(p_1, p_2]\%$.
- *Bronze*: các khách hàng còn lại.

Việc xác định ngưỡng được thực hiện thông qua các phân vị:

$$\tau_{\text{Gold}} = Q_{1-p_1}(\text{Revenue}), \quad \tau_{\text{Silver}} = Q_{1-p_2}(\text{Revenue}),$$

trong đó $Q_q(\cdot)$ là hàm phân vị tại mức q . Để xử lý trường hợp nhiều khách hàng có doanh thu trùng với giá trị ngưỡng, hai chiến lược cắt được xem xét: (i) cắt nghiêm ngặt theo thứ hạng, hoặc (ii) gộp toàn bộ các khách hàng có giá trị bằng ngưỡng vào nhóm trên nhằm tránh tách rời các quan sát đồng hạng.

Trích xuất đặc trưng hành vi mua hàng Để phân biệt hành vi mua sỉ và mua lẻ, các đặc trưng hành vi được xây dựng từ dữ liệu tổng hợp, bao gồm tổng số lượng mua, AOV và số lượng trung bình trên mỗi hóa đơn. Các đặc trưng này phản ánh đồng thời quy mô giao dịch và cấu trúc đơn hàng của từng khách hàng.

Xác định ngưỡng hành vi bằng phương pháp phân vị Các ngưỡng hành vi được xác định dựa trên phân vị của phân phối đặc trưng:

$$\begin{aligned}\tau_Q &= Q_q(\text{Quantity}), \\ \tau_{AOV} &= Q_q(\text{AOV}), \\ \tau_{QPB} &= Q_q(\text{QtyPerBill}).\end{aligned}$$

với q thường được lựa chọn trong khoảng $[0.90, 0.95]$ để tập trung vào nhóm khách hàng có hành vi nổi bật. Việc sử dụng phân vị giúp mô hình hóa hành vi tương đối thay vì phụ thuộc vào ngưỡng tuyệt đối.

Phân loại Buyer Type theo quy tắc kết hợp Buyer Type được xác định thông qua quy tắc kết hợp các ngưỡng hành vi. Một khách hàng được gán nhãn *Wholesale* nếu đồng thời thỏa mãn:

$$\text{Quantity}_i \geq \tau_Q \quad \text{và} \quad \text{QtyPerBill}_i \geq \tau_{QPB},$$

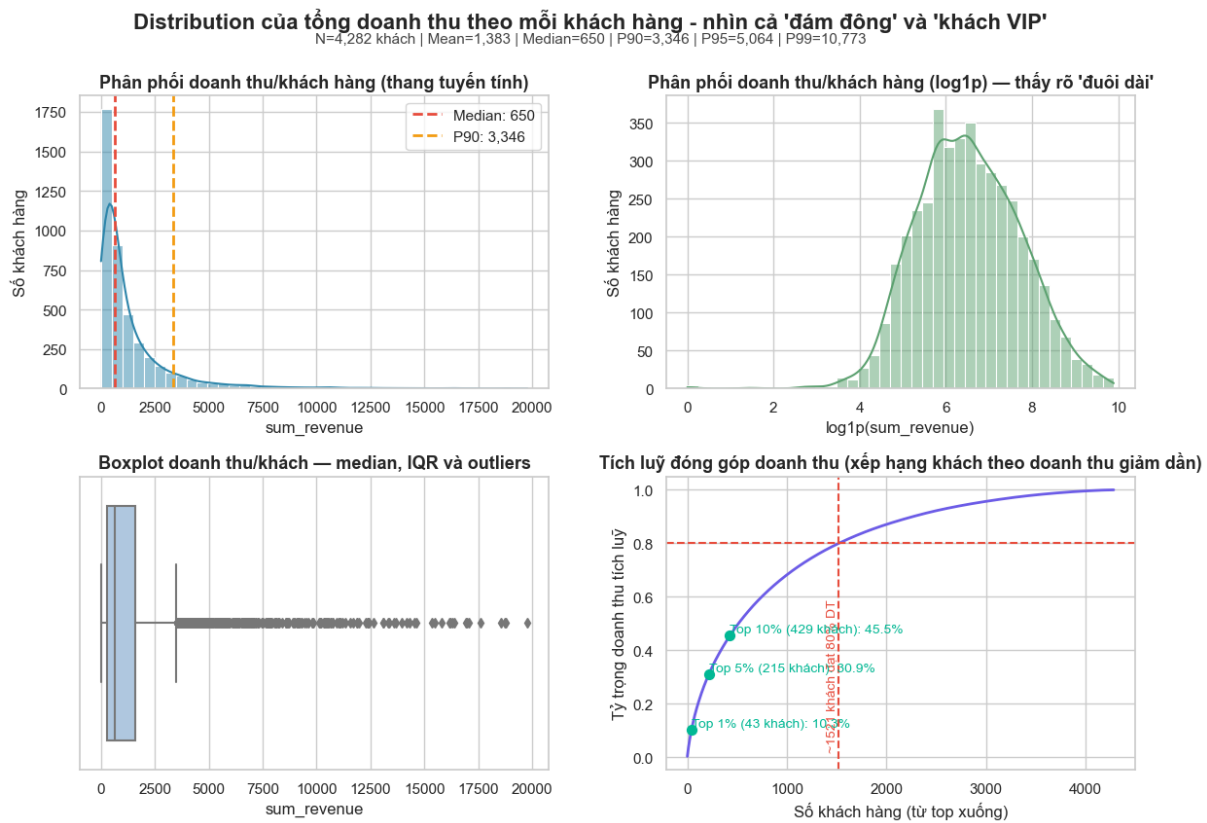
trong khi các trường hợp còn lại được gán nhãn *Retail*. Quy tắc này đảm bảo rằng nhóm Wholesale không chỉ có tổng khối lượng mua lớn mà còn thể hiện hành vi mua theo lô ở cấp độ từng giao dịch.

Đánh giá phân bố và tính nhất quán Sau khi phân chia, các nhóm khách hàng được đánh giá thông qua các chỉ số mô tả như số lượng khách hàng, tổng doanh thu, tỷ trọng doanh thu và các thống kê phân phối. Bước này nhằm đảm bảo tính hợp lý của việc phân nhóm trước khi sử dụng các nhóm khách hàng như các đơn vị phân tích độc lập trong các bước tiếp theo.

6.2 Phân tích doanh thu theo chiều Tier

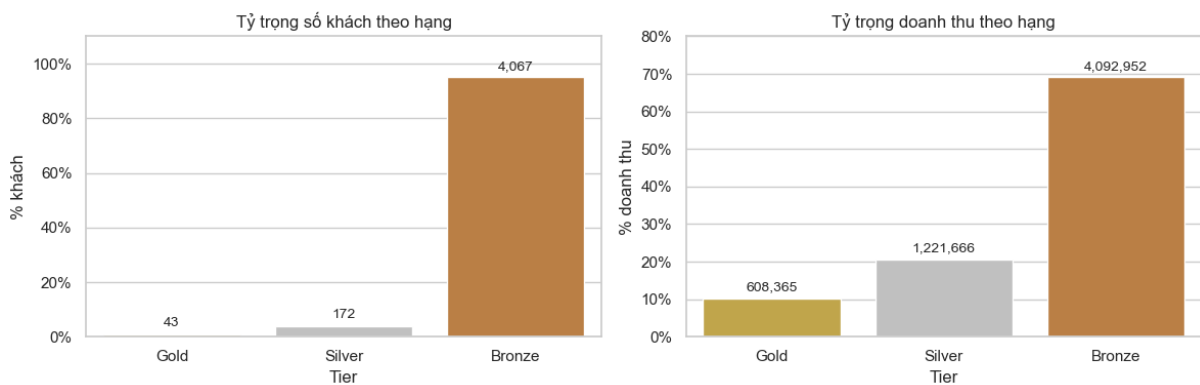
6.2.1 Tỷ trọng các nhóm khách hàng theo Tier

Phân chia tập khách hàng hệ thống theo "nhóm đám đông" và "khách VIP". trong đó nhóm Gold chiếm quantile 1-5%, Nhóm Silver từ 5-10%, Nhóm Bronze là phần còn lại. Các thông tin được ghi nhận thông qua hình sau:



Hình 6.1: Phân phối của tổng doanh thu theo mỗi khách hàng - nhóm 'đám đông' và 'khách VIP'

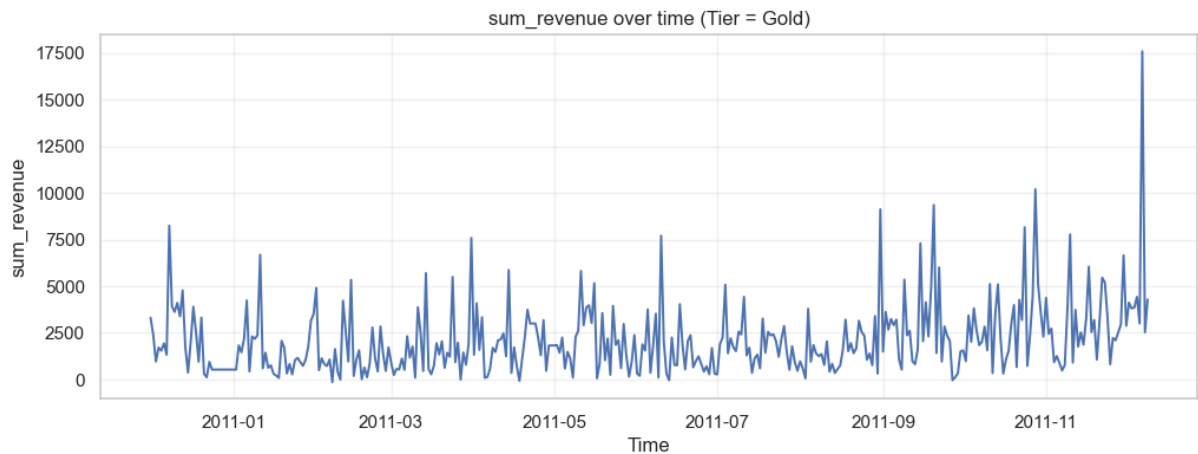
Sau khi định nghĩa và phân chia, tập khách hàng hệ thống được chia theo Tier, như sau:



Hình 6.2: Tỷ trọng theo số lượng khách và doanh thu của các nhóm Tier

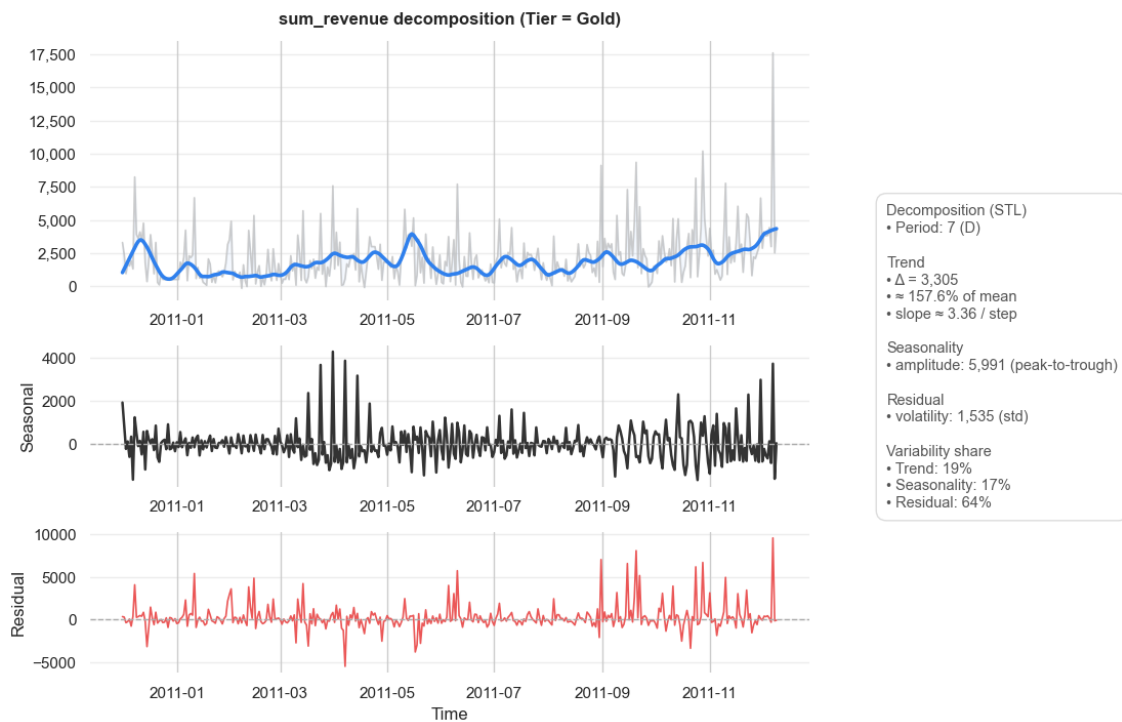
6.2.2 Phân rã STL theo từng Tier

Nhóm Vàng (Gold) Có mẫu được quan sát và phân tích:



Hình 6.3: Doanh thu theo ngày nhóm Tier = Gold

Kết quả phân tích:

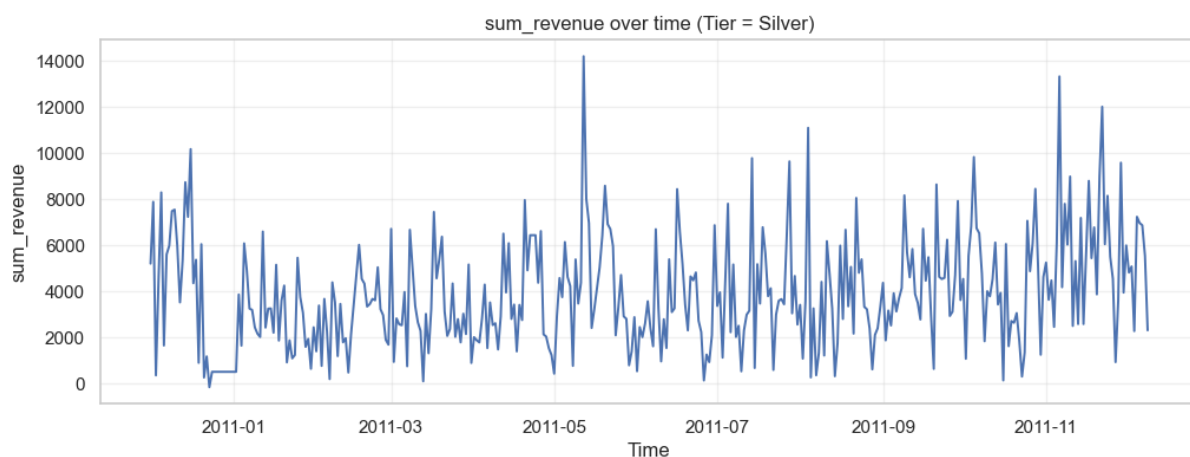


Hình 6.4: Phân tích nhóm Tier = Gold

Thu được:

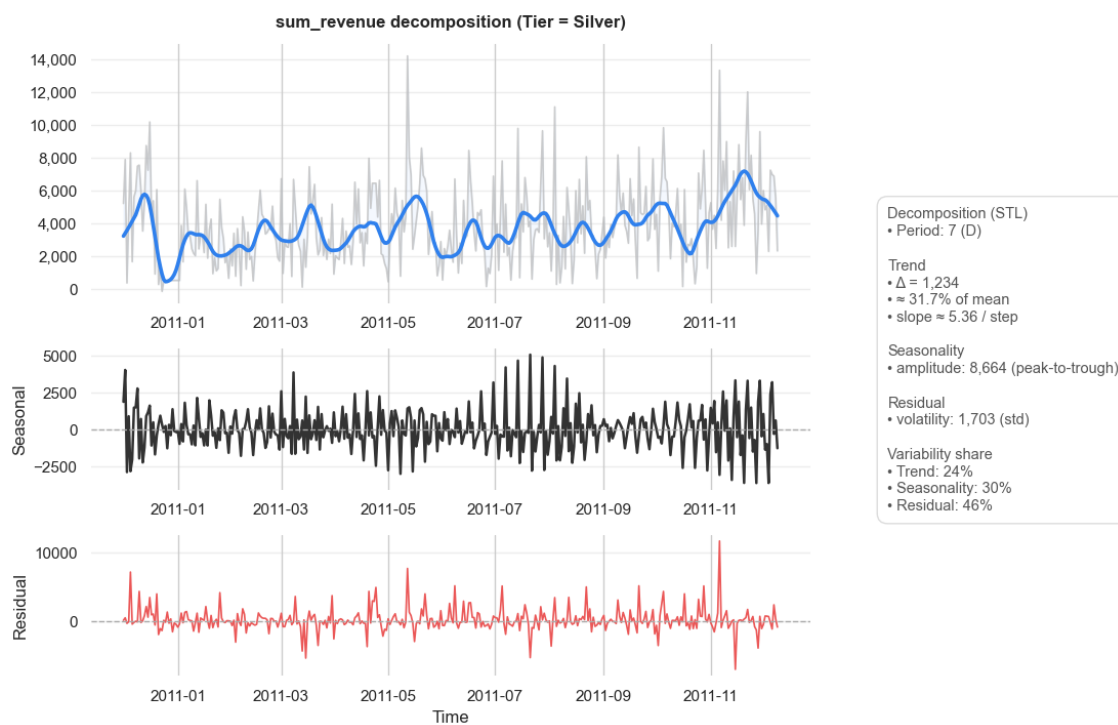
Method : STL, freq = D , period = 7
Trend (Δ) : 3,305 (157.6%)
Seasonality : 5,991 (peak-to-trough)
Residual : 1,535 (std)

Nhóm Bạc (Silver) Có mẫu được quan sát và phân tích:



Hình 6.5: Doanh thu theo ngày nhóm Tier = Silver

Kết quả phân tích:

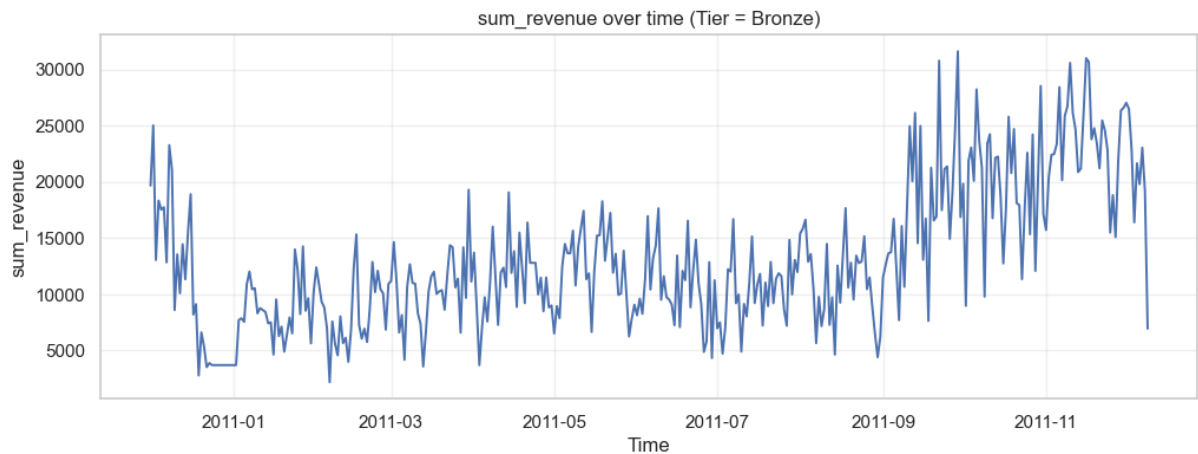


Hình 6.6: Phân tích nhóm Tier = Silver

Thu được:

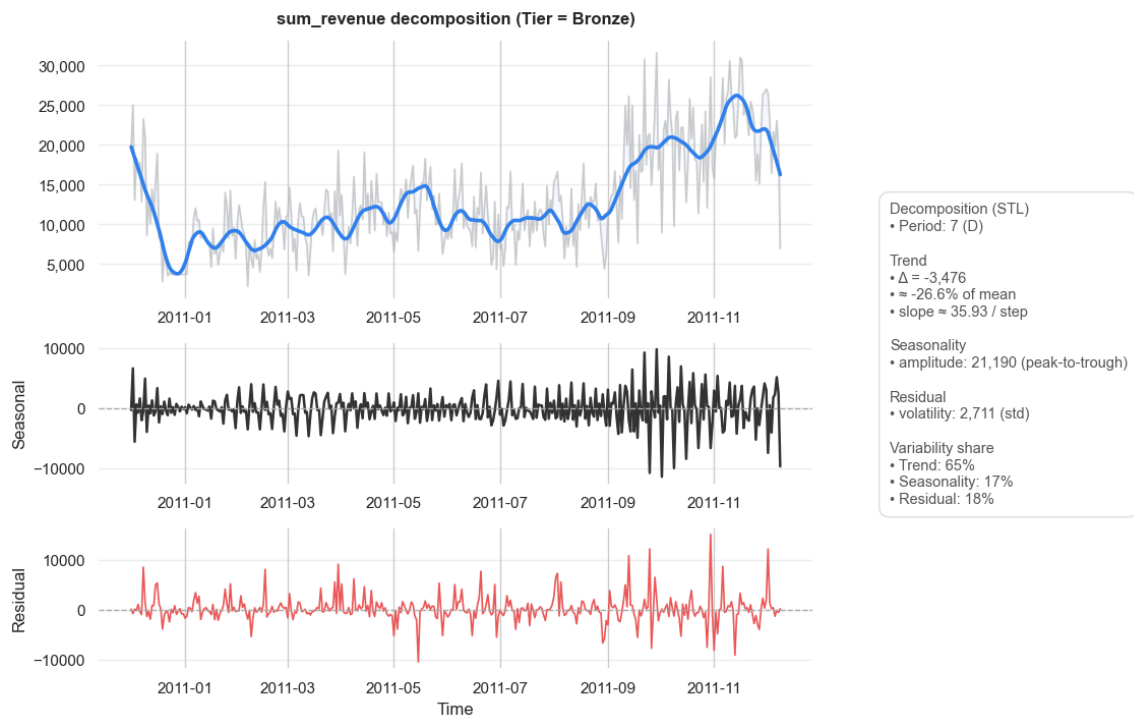
Method : STL, freq = D , period = 7
Trend (Δ) : 1,234 (31.7%)
Seasonality : 8,664 (peak-to-trough)
Residual : 1,703 (std)

Nhóm Đồng (Bronze) Có mẫu được quan sát và phân tích:



Hình 6.7: Doanh thu theo ngày nhóm Tier = Bronze

Kết quả phân tích:



Hình 6.8: Phân tích nhóm Tier = Bronze

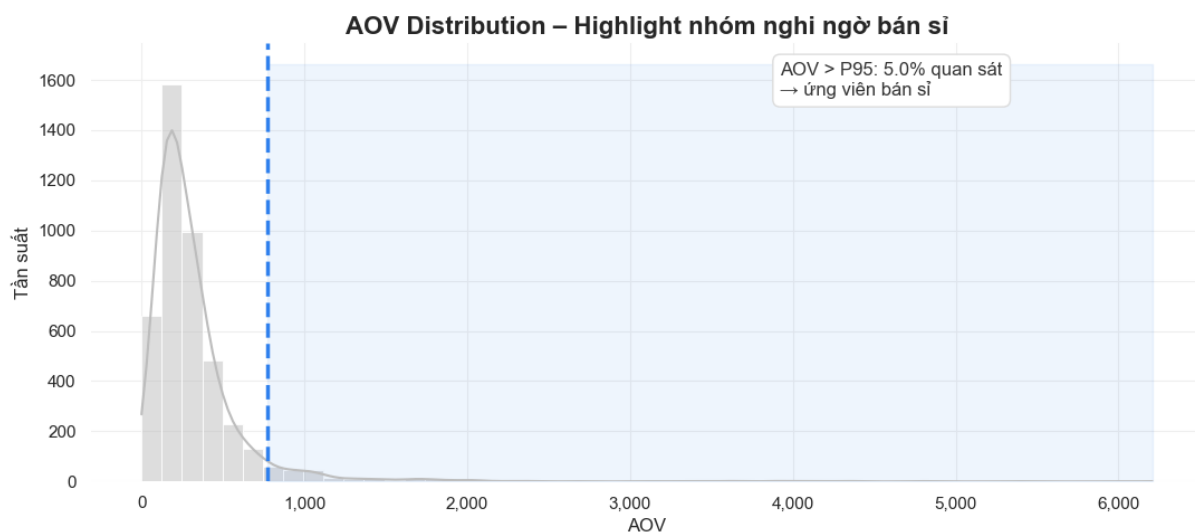
Thu được:

Method : STL, freq = D , period = 7
Trend (Δ) : $-3,476$ (-26.6%)
Seasonality : 21,190 (peak-to-trough)
Residual : 2,711 (std)

6.3 Phân tích doanh thu theo chiều BuyerType

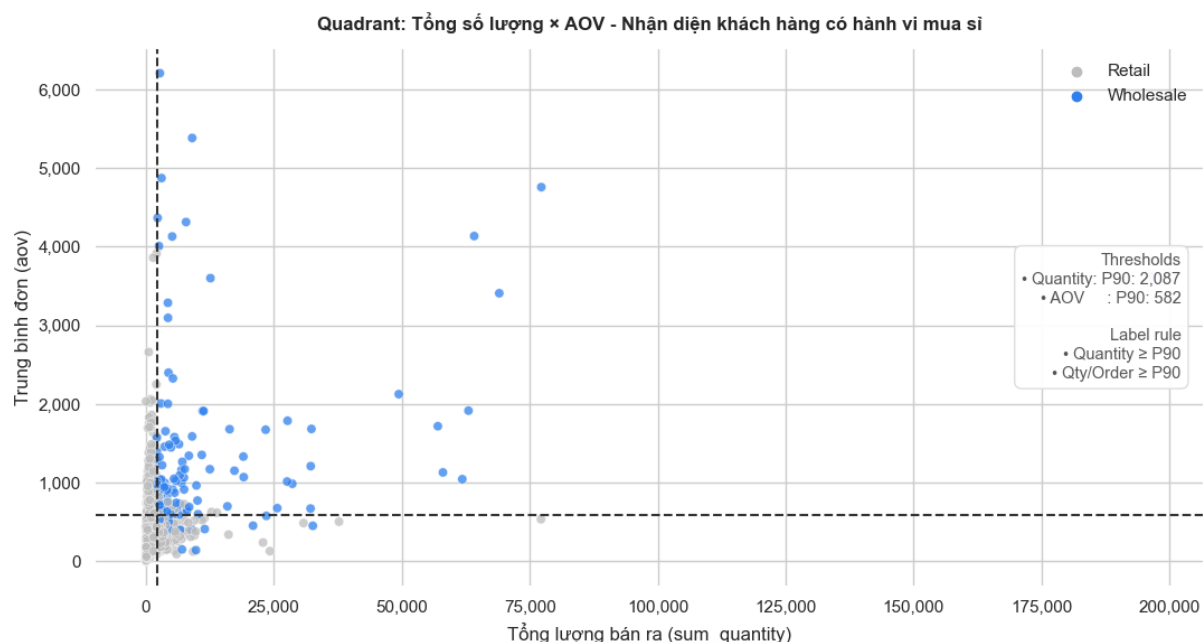
6.3.1 Tỷ trọng các nhóm khách hàng theo BuyerType

Xét về việc phân chia theo tiêu chí bản chất mua bán (Bán sỉ, lẻ) của khách hàng hệ thống, chúng tôi khảo sát được kết quả như hình sau:



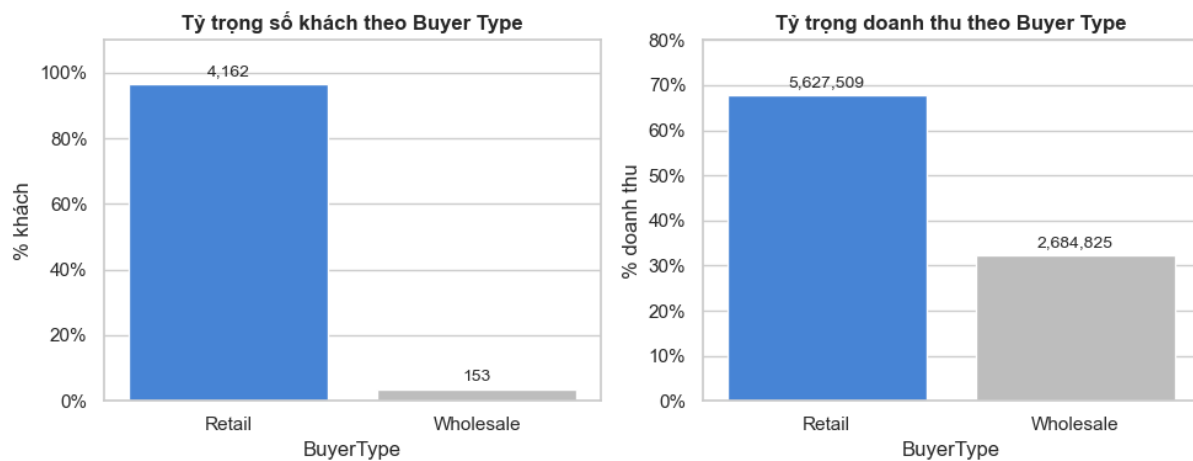
Hình 6.9: Nghi ngờ nhóm bán sỉ

Sau khi nghi ngờ, chúng tôi tiến hành định nghĩa và chia như sau:



Hình 6.10: Ngưỡng phân chia bán sỉ và lẻ

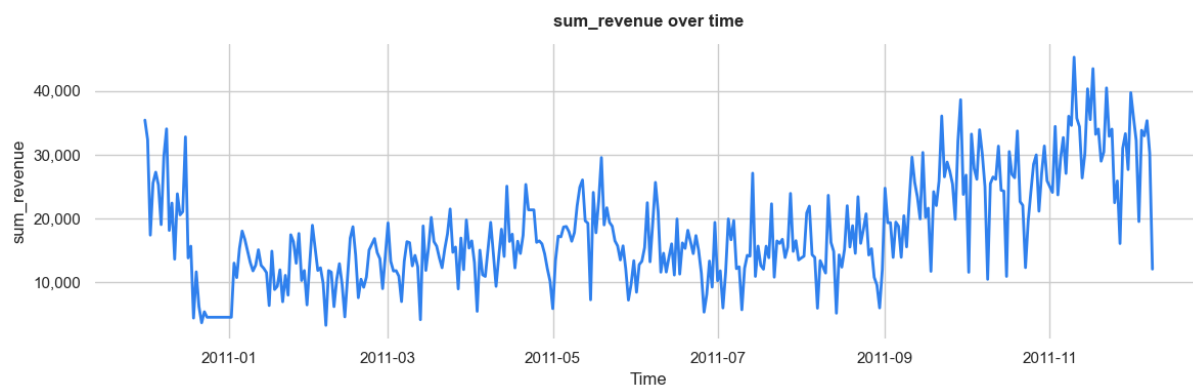
Kết thúc phân chia, chúng tôi thu được tỷ trọng các nhóm phân chia là:



Hình 6.11: Tỷ trọng nhóm bán sỉ và lẻ

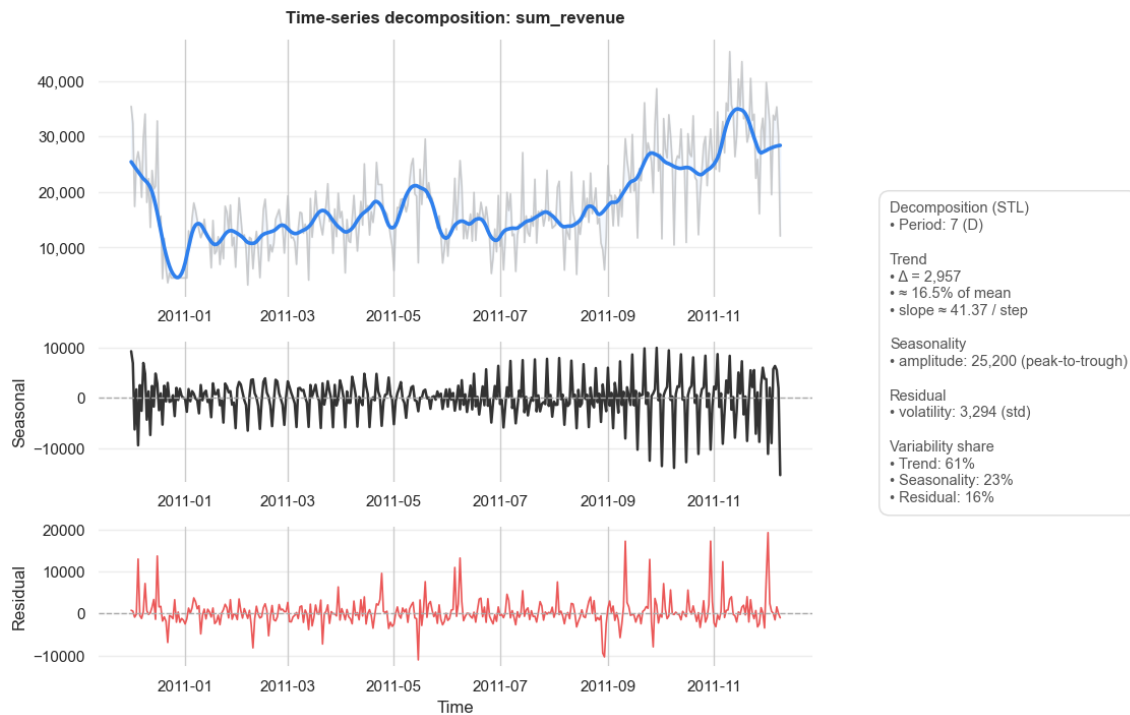
6.3.2 Phân rã STL theo từng BuyerType

Nhóm bán lẻ (retail) Có mẫu được quan sát và phân tích:



Hình 6.12: Doanh thu theo ngày nhóm Bán lẻ

Kết quả phân tích:

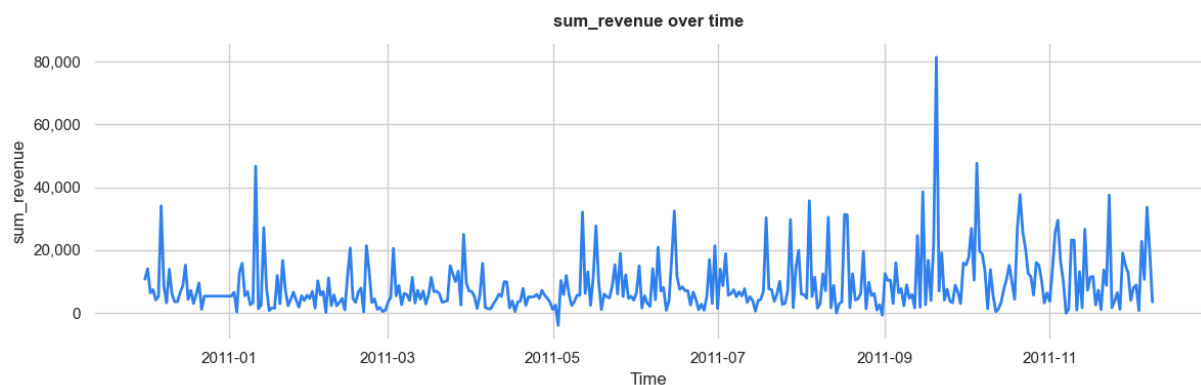


Hình 6.13: Phân tích nhóm Bán lẻ

Thu được:

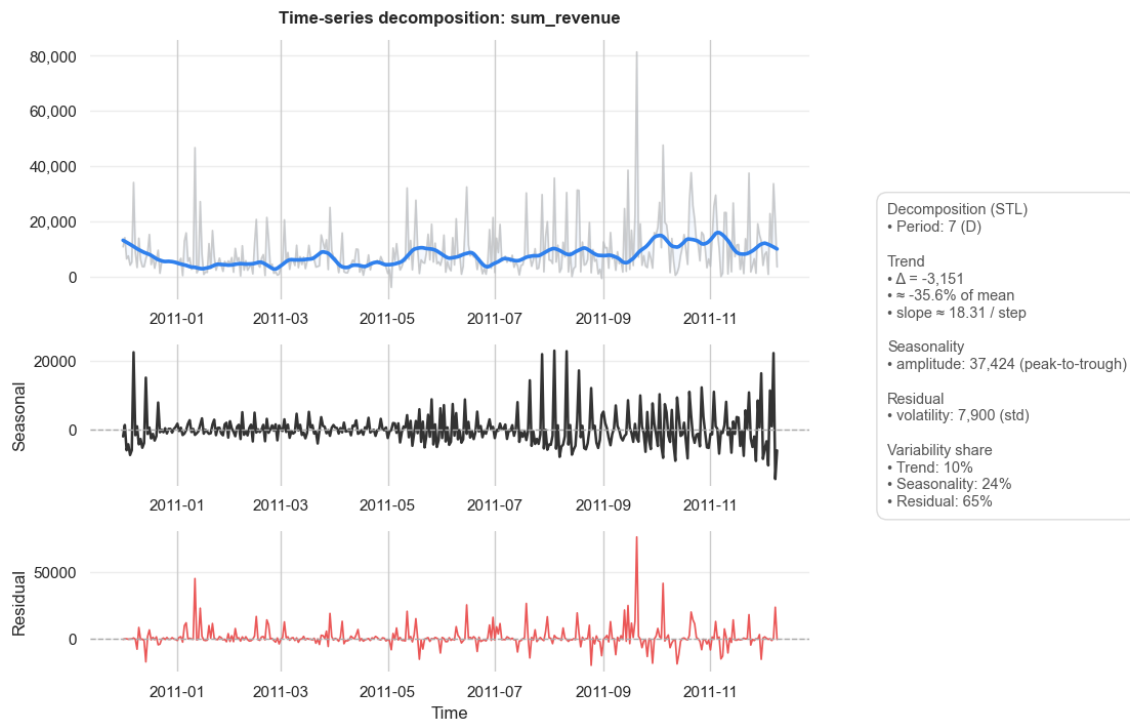
Method : STL, freq = D , period = 7
Trend (Δ) : 2,957 (16.5%)
Seasonality : 25,200 (peak-to-trough)
Residual : 3,294 (std)

Nhóm bán sỉ (wholesale) Có mẫu được quan sát và phân tích:



Hình 6.14: Doanh thu theo ngày nhóm Bán sỉ

Kết quả phân tích:



Hình 6.15: Phân tích nhóm Bán sỉ

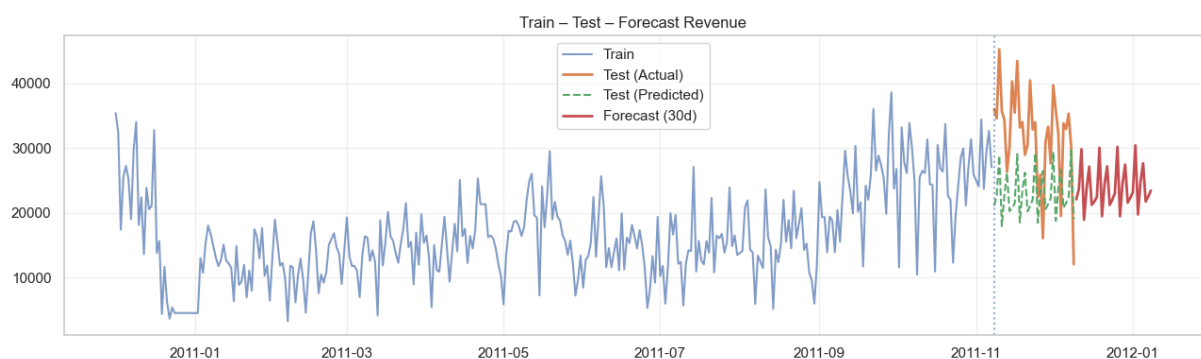
Thu được:

Method : STL, freq = D , period = 7
Trend (Δ) : $-3,151$ (-35.6%)
Seasonality : 37,424 (peak-to-trough)
Residual : 7,900 (std)

6.3.3 Dự đoán doanh thu 30 ngày theo từng BuyerType

Nhóm bán lẻ

Kết quả dự đoán khi dùng SARIMA dự đoán doanh thu nhóm bán lẻ:



Hình 6.16: Dự đoán doanh thu 30 ngày tới của nhóm Bán lẻ

Đánh giá tăng trưởng

TB doanh thu 30 ngày qua : 31.783,72

TB dự báo 30 ngày tới : 23.921,70

Mức độ tăng trưởng : $-7.862,02$ ($-24,74\%$)

Nhận định : Giảm đáng kể

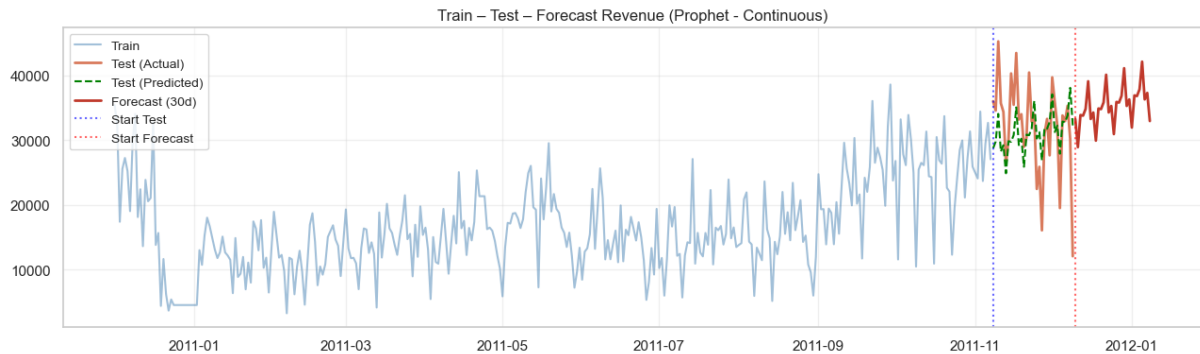
Thông tin mô hình

Luồng xử lý (Pipeline) : STL \rightarrow SARIMAX(residual)

Chu kỳ mùa vụ : 7 ngày

Tỷ lệ train-test : 11:1

Kết quả dự đoán khi dùng prophet dự đoán doanh thu nhóm bán lẻ:



Hình 6.17: Dự đoán doanh thu 30 ngày tới của nhóm Bán lẻ

Đánh giá tăng trưởng

TB doanh thu 30 ngày qua : 31.783,72

TB dự báo 30 ngày tới : 35.383,54

Mức độ tăng trưởng : $+3.599,82$ ($+11,33\%$)

Nhận định : Tăng rõ rệt

Thông tin mô hình

Mô hình : Prophet

Mùa vụ : Weekly seasonality

Tỷ lệ train-test : 11:1

Kết luận : Chọn mô hình Prophet

Lý do chính : Xử lý nhiễu tốt, phản ánh đúng đà tăng trưởng thực tế.

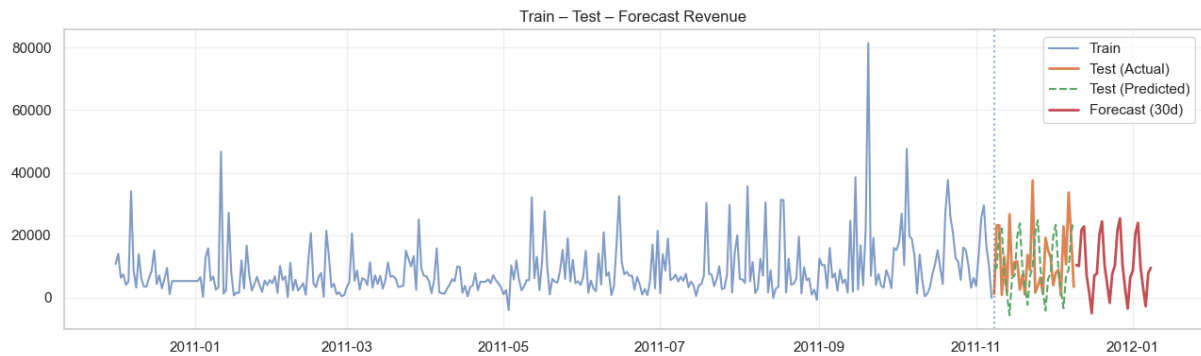
So sánh kịch bản : Prophet ($+11,33\%$) vs SARIMA ($-24,74\%$)

Tính khả thi : Tránh các dự báo bị quan quá mức từ nhiễu ngắn hạn.

Mô hình	Dự báo TB	Tăng trưởng	Đề xuất
SARIMA (STL)	23.921,70	-24,74%	Loại bỏ (Nhiều)
Prophet	35.383,54	+11,33%	Lựa chọn

Nhóm bán sỉ

Kết quả dự đoán khi dùng SARIMA dự đoán doanh thu nhóm bán sỉ:



Hình 6.18: Dự đoán doanh thu 30 ngày tới của nhóm Bán sỉ

Kết quả dự báo doanh thu 30 ngày tiếp theo cho nhóm khách hàng bán sỉ được thực hiện bằng mô hình SARIMA.

Đánh giá tăng trưởng

TB doanh thu 30 ngày qua : 11.289,52

TB dự báo 30 ngày tới : 10.119,08

Mức độ tăng trưởng : $-1.170,44$ ($-10,37\%$)

Nhận định : Giảm đáng kể

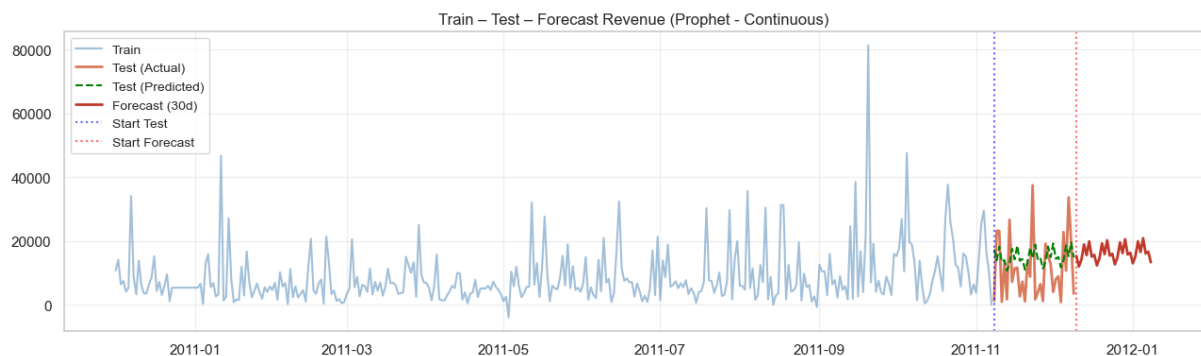
Thông tin mô hình

Luồng xử lý (Pipeline) : STL \rightarrow SARIMAX(residual)

Chu kỳ mùa vụ : 7 ngày

Tỷ lệ train-test : 11:1

Kết quả dự đoán khi dùng prophet dự đoán doanh thu nhóm bán sỉ:



Hình 6.19: Dự đoán doanh thu 30 ngày tới của nhóm Bán sỉ

Đánh giá tăng trưởng

TB doanh thu 30 ngày qua : 11.289,52

TB dự báo 30 ngày tới : 16.244,72

Mức độ tăng trưởng : +4.955,20 (+43,89%)

Nhận định : Tăng rõ rệt

Thông tin mô hình

Mô hình : Prophet

Mùa vụ : Weekly seasonality

Tỷ lệ train-test : 11:1

Kết luận : Chọn mô hình Prophet

Lý do chính : Tách biệt tốt xu hướng và mùa vụ, ít nhạy cảm với nhiễu cuối kỳ.

So sánh kịch bản : Prophet (+43,89%) vs SARIMA (-10,37%)

Tính khả thi : Phản ánh đúng đà tăng trưởng; tránh dự báo bị quan sai lệch.

Mô hình	Dự báo TB	Tăng trưởng	Đề xuất
SARIMA (STL)	10.119,08	-10,37%	Loại bỏ (Bị nhiễu)
Prophet	16.244,72	+43,89%	Lựa chọn

Chương 7

Insights và Quyết định

7.1 Insights: Phân tích doanh thu xu hướng chung

7.1.1 Nhận định và Rủi ro

Nhận định chính

Tổng hợp các kết quả phân tích cho thấy doanh thu của doanh nghiệp có xu hướng tăng trưởng tích cực nhưng phụ thuộc đáng kể vào một số sản phẩm và khách hàng chủ chốt. Bên cạnh đó, yếu tố mùa vụ đóng vai trò quan trọng trong việc định hình biến động doanh thu theo thời gian. Cơ cấu doanh thu chưa thực sự cân bằng giữa các nhóm sản phẩm và khách hàng, cho thấy dư địa cải thiện trong chiến lược kinh doanh.

Rủi ro tiềm ẩn

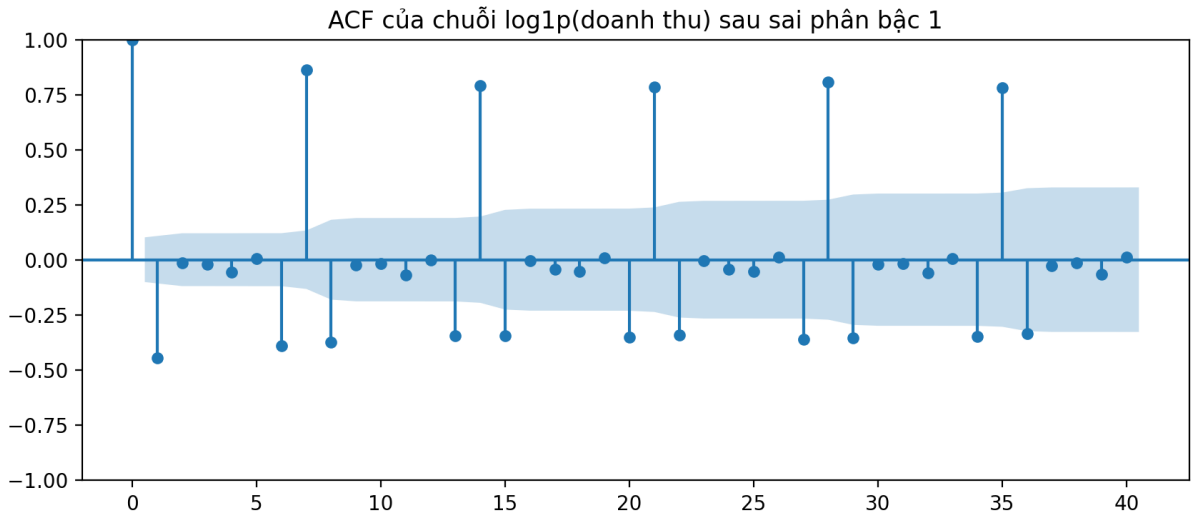
Mức độ tập trung doanh thu cao khiến chúng ta phải đối mặt với rủi ro nếu mất đi các khách hàng lớn hoặc khi nhu cầu đối với sản phẩm chủ lực suy giảm. Ngoài ra, danh mục sản phẩm phân tán với nhiều sản phẩm có đóng góp thấp có thể làm giảm hiệu quả vận hành. Cuối cùng, việc không xem xét đầy đủ yếu tố mùa vụ trong phân tích và dự báo có thể dẫn đến sai lệch trong kế hoạch bán hàng và tồn kho.

7.1.2 Phân tích và đánh giá kết quả dự báo bằng mô hình SARIMA

Động cơ lựa chọn mô hình SARIMA

Trong bối cảnh dữ liệu doanh thu thể hiện rõ xu hướng theo thời gian và tồn tại yếu tố mùa vụ nội năm, các mô hình chuỗi thời gian có khả năng xử lý đồng thời xu hướng, tính dừng và mùa vụ là cần thiết. Mô hình SARIMA là mô hình hóa cấu trúc tự tương quan của chuỗi doanh thu, đồng thời tách biệt thành phần mùa vụ và phi mùa vụ trong cùng một khuôn khổ thống nhất.

Trước khi xây dựng mô hình, chuỗi doanh thu đã được kiểm tra tính dừng và được sai phân ở các bậc phù hợp nhằm đảm bảo điều kiện áp dụng của mô hình ARIMA. Việc xác định các tham số của SARIMA được thực hiện dựa trên phân tích hàm tự tương quan (ACF), hàm tự tương quan riêng phần (PACF) và so sánh các tiêu chí thông tin như AIC và BIC.



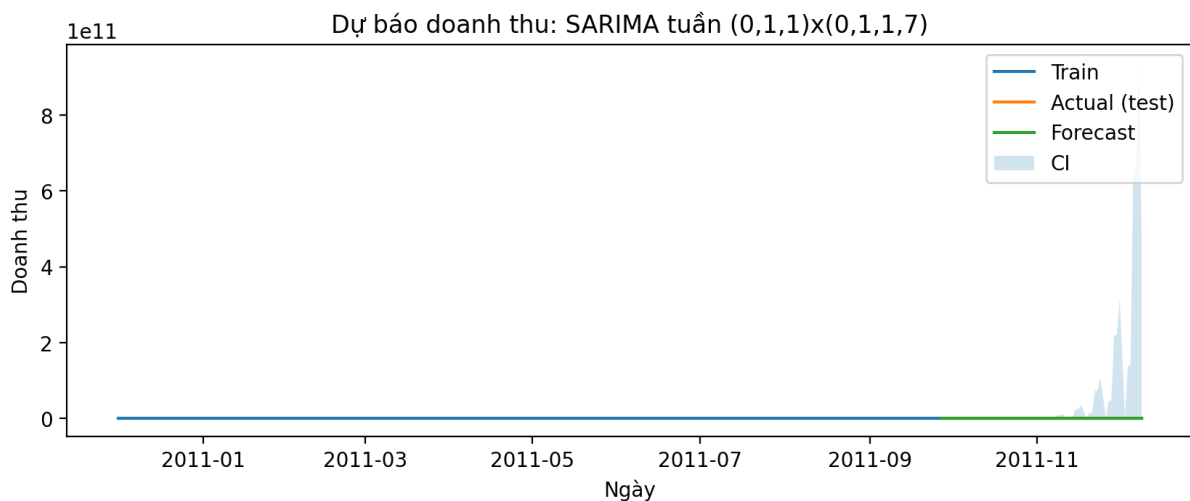
Hình 7.1: Hàm tự tương quan (ACF) và tự tương quan riêng phần (PACF) của chuỗi doanh thu sau khi sai phân

Mô hình SARIMA thứ nhất: Mô hình cơ sở

Mô hình SARIMA thứ nhất được xây dựng với mục tiêu đóng vai trò mô hình cơ sở, phản ánh cấu trúc tự tương quan chính của chuỗi doanh thu mà không quá phức tạp về mặt tham số. Mô hình này tập trung nắm bắt xu hướng chung và thành phần mùa vụ cơ bản của dữ liệu.

Kết quả ước lượng cho thấy các tham số tự hồi quy và trung bình trượt ở cả phần phi mùa vụ và mùa vụ đều có ý nghĩa thống kê, cho thấy mô hình đã khai thác được các mối quan hệ phụ thuộc theo thời gian trong chuỗi doanh thu. Phần dư của mô hình sau khi ước lượng có đặc điểm gần với nhiễu trắng, cho thấy mô hình đã giải thích được phần lớn cấu trúc hệ thống của dữ liệu.

Tuy nhiên, khi đánh giá hiệu quả dự báo trên tập kiểm tra, mô hình SARIMA thứ nhất vẫn còn tồn tại sai lệch nhất định trong các giai đoạn biến động mạnh. Điều này cho thấy mô hình cơ sở phù hợp để mô tả xu hướng tổng thể nhưng chưa phản ứng đủ linh hoạt trước các thay đổi đột ngột trong nhu cầu.



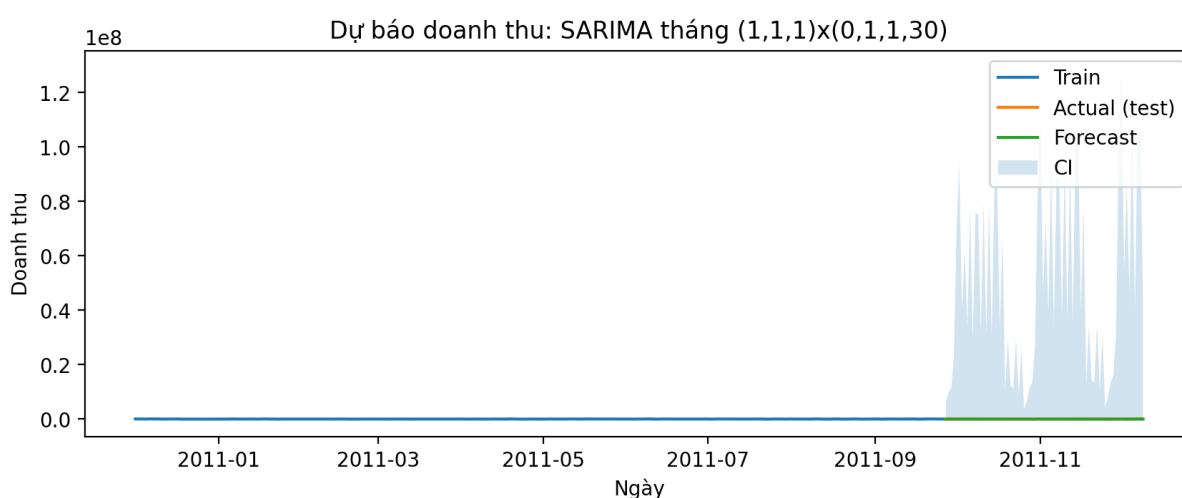
Hình 7.2: Kết quả dự báo doanh thu ngắn hạn bằng mô hình SARIMA thứ nhất

Mô hình SARIMA thứ hai: Mô hình cải tiến

Nhằm cải thiện khả năng dự báo, mô hình SARIMA thứ hai được xây dựng với các tham số khác nhằm nắm bắt tốt hơn các dao động ngắn hạn và thành phần mùa vụ phức tạp hơn. Việc bổ sung hoặc điều chỉnh các bậc tự hồi quy và trung bình trượt giúp mô hình phản ánh chi tiết hơn hành vi của chuỗi doanh thu.

Kết quả cho thấy mô hình SARIMA thứ hai đạt giá trị AIC và BIC thấp hơn so với mô hình cơ sở, phản ánh mức độ phù hợp cao hơn với dữ liệu. Phân tích phần dư tiếp tục cho thấy không còn cấu trúc tự tương quan đáng kể, củng cố tính hợp lệ của mô hình.

Quan trọng hơn, khi so sánh trên tập kiểm tra, mô hình SARIMA thứ hai cho sai số dự báo thấp hơn một cách nhất quán so với mô hình thứ nhất. Điều này đặc biệt rõ rệt trong các giai đoạn doanh thu biến động mạnh, cho thấy mô hình cải tiến có khả năng thích ứng tốt hơn với những thay đổi trong hành vi mua hàng theo thời gian.

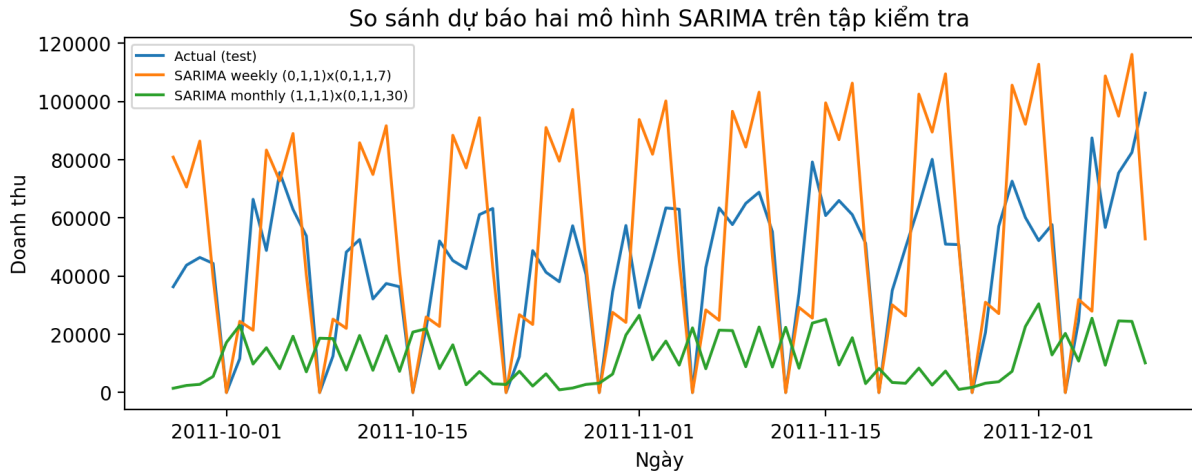


Hình 7.3: Kết quả dự báo doanh thu ngắn hạn bằng mô hình SARIMA thứ hai với độ chính xác cải thiện

So sánh và đánh giá hai mô hình SARIMA

Việc so sánh hai mô hình SARIMA cho thấy sự đánh đổi điển hình giữa tính đơn giản và độ chính xác dự báo. Mô hình SARIMA thứ nhất có cấu trúc gọn nhẹ, dễ diễn giải và phù hợp để theo dõi xu hướng dài hạn của doanh thu. Trong khi đó, mô hình SARIMA thứ hai có độ phức tạp cao hơn nhưng mang lại cải thiện rõ rệt về hiệu quả dự báo ngắn hạn.

Từ góc độ ứng dụng thực tiễn, mô hình SARIMA thứ nhất có thể được sử dụng như một công cụ giám sát xu hướng doanh thu tổng thể, giúp nhận diện sớm các thay đổi mang tính cấu trúc. Ngược lại, mô hình SARIMA thứ hai phù hợp hơn cho các bài toán dự báo ngắn hạn phục vụ lập kế hoạch tồn kho, phân bổ nhân sự và tối ưu hóa chiến dịch bán hàng theo thời gian.



Hình 7.4: 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

Kết quả dự báo từ hai mô hình SARIMA cung cấp căn cứ định lượng quan trọng cho việc ra quyết định trong quản trị bán lẻ trực tuyến. Thứ nhất, việc dự báo doanh thu theo chu kỳ cho phép chúng ta chủ động chuẩn bị tồn kho và nguồn lực trong các giai đoạn cao điểm, đồng thời hạn chế rủi ro tồn kho dư thừa trong các giai đoạn thấp điểm.

Thứ hai, sự khác biệt về hiệu quả giữa hai mô hình nhấn mạnh tầm quan trọng của việc lựa chọn mô hình phù hợp với mục tiêu quản trị. Trong bối cảnh cần phản ứng nhanh với biến động thị trường, việc ưu tiên mô hình có khả năng dự báo chính xác hơn trong ngắn hạn là cần thiết, ngay cả khi mô hình đó phức tạp hơn về mặt kỹ thuật.

Cuối cùng, việc tích hợp kết quả dự báo từ mô hình SARIMA vào quy trình ra quyết định giúp doanh nghiệp chuyển từ cách quản lý dựa trên kinh nghiệm sang quản lý dựa trên dữ liệu, qua đó nâng cao tính chủ động và hiệu quả trong hoạt động kinh doanh.

Hạn chế và hướng phát triển tiếp theo

Mặc dù hai mô hình SARIMA đã cho thấy hiệu quả trong việc dự báo doanh thu ngắn hạn, nghiên cứu vẫn tồn tại một số hạn chế. Mô hình hiện tại chỉ khai thác thông tin nội tại của chuỗi doanh thu mà chưa xem xét các biến ngoại sinh như chương trình khuyến mãi, yếu tố kinh tế vĩ mô hoặc sự kiện đặc biệt. Trong các nghiên cứu tiếp theo, việc mở rộng sang các mô hình SARIMAX hoặc kết hợp với các phương pháp học máy có thể giúp cải thiện hơn nữa độ chính xác dự báo.

7.2 Insights: Phân tích doanh thu theo phân khúc khách hàng

7.2.1 Đặc điểm nhóm Tier

Dựa vào phân rã STL, chúng tôi phát biểu một số đặc điểm nổi bật như sau:

Nhóm Gold đóng vai trò trụ cột doanh thu và thể hiện xu hướng tăng trưởng rõ rệt. Kết quả STL cho thấy thành phần xu hướng (*trend*) có mức tăng tuyệt đối lớn và

tốc độ tăng cao so với giá trị trung bình, phản ánh sự mở rộng ổn định của nhóm khách hàng giá trị cao. Biên độ mùa vụ (*seasonality*) tương đối thấp, cho thấy nhu cầu ít bị chi phối bởi các yếu tố thời vụ. Thành phần nhiễu (*residual*) nhỏ cho thấy hành vi mua có tính dự đoán cao, phù hợp cho các chiến lược hoạch định cung ứng dài hạn.

Nhóm Silver ghi nhận tăng trưởng dương nhưng với mức độ vừa phải và độ nhạy mùa vụ cao hơn. Xu hướng doanh thu tăng nhẹ, cho thấy tiềm năng mở rộng nhưng chưa ổn định như nhóm Gold. Biên độ mùa vụ lớn hơn phản ánh sự phụ thuộc đáng kể vào các chu kỳ khuyến mãi hoặc thời điểm tiêu dùng. Điều này cho thấy nhóm Silver phù hợp với các chiến lược kích cầu theo mùa và các chính sách giá linh hoạt.

Nhóm Bronze thể hiện xu hướng suy giảm doanh thu và mức độ biến động cao. Thành phần xu hướng âm cho thấy sự sụt giảm dài hạn, trong khi biên độ mùa vụ lớn phản ánh hành vi mua không ổn định và phụ thuộc mạnh vào các yếu tố ngắn hạn. Residual tương đối cao cho thấy mức độ nhiễu lớn, làm giảm khả năng dự báo chính xác. Nhóm này chủ yếu đóng vai trò tạo doanh thu thời điểm, ít phù hợp cho hoạch định chuỗi cung ứng mang tính dài hạn.

Tổng hợp Kết quả phân tích cho thấy sự phân hóa rõ rệt về động lực doanh thu giữa các Tier. Trong khi nhóm Gold mang tính nền tảng và ổn định cho hệ thống bán lẻ, nhóm Silver và Bronze phản ánh các nguồn biến động và cơ hội tăng trưởng ngắn hạn. Việc kết hợp các chiến lược cung ứng và marketing theo đặc trưng từng Tier là cần thiết để tối ưu hóa hiệu quả toàn chuỗi.

7.2.2 Đặc điểm nhóm bán sỉ/lẻ

Dựa vào phân rã STL, chúng tôi phát biểu một số đặc điểm của nhóm bán sỉ và bán lẻ:

Nhóm Bán lẻ (Retail) thể hiện xu hướng tăng trưởng dương và đóng vai trò ổn định trong doanh thu ngắn hạn. Kết quả phân rã STL cho thấy thành phần xu hướng tăng nhẹ, phản ánh sự mở rộng đều của nhu cầu tiêu dùng cuối. Biên độ mùa vụ tương đối lớn cho thấy doanh thu bán lẻ chịu ảnh hưởng đáng kể từ các chu kỳ tuần, chương trình khuyến mãi và hành vi mua sắm theo thời điểm. Thành phần nhiễu ở mức trung bình cho thấy mức độ biến động chấp nhận được, phù hợp cho các mô hình dự báo ngắn hạn và hoạch định tồn kho linh hoạt.

Nhóm Bán sỉ (Wholesale) ghi nhận xu hướng suy giảm rõ rệt và mức độ biến động cao. Thành phần xu hướng âm với độ lớn đáng kể phản ánh sự co hẹp trong nhu cầu đặt hàng dài hạn từ các khách hàng sỉ. Biên độ mùa vụ rất lớn cho thấy doanh thu phụ thuộc mạnh vào các đơn hàng theo đợt, mang tính không đều và khó dự đoán. Residual cao phản ánh mức độ nhiễu lớn, làm gia tăng rủi ro sai lệch trong dự báo và đòi hỏi các chiến lược quản trị cung ứng thận trọng hơn.

So sánh và hàm ý quản trị Phân tích cho thấy sự dịch chuyển động lực doanh thu từ kênh bán sỉ sang bán lẻ trong ngắn hạn. Trong khi bán lẻ mang lại dòng doanh thu ổn định và có xu hướng tăng, bán sỉ đóng vai trò nguồn doanh thu lớn nhưng

kém ổn định và đang suy giảm. Điều này gợi ý doanh nghiệp cần tái cân bằng chiến lược chuỗi cung ứng, ưu tiên khả năng phản ứng nhanh cho kênh bán lẻ, đồng thời rà soát lại chính sách hợp tác và dự báo nhu cầu đối với khách hàng bán sỉ.

7.2.3 Đánh giá hiệu suất dự báo và Lựa chọn mô hình

Dựa trên các đặc điểm về xu hướng và mùa vụ đã phân tích qua STL, chúng tôi tiến hành thực nghiệm dự báo 30 ngày tới và ghi nhận sự phân hóa rõ rệt giữa hai phương pháp tiếp cận:

Sự nhạy cảm với nhiễu của SARIMA Mô hình SARIMA (kết hợp STL) có xu hướng đưa ra các dự báo mang tính thận trọng cực đoan. Kết quả thực nghiệm cho thấy SARIMA phản ứng mạnh với các biến động tiêu cực ngắn hạn ở cuối chu kỳ dữ liệu (residuals), dẫn đến các dự báo sụt giảm từ **-10,37%** đến **-24,74%**. Điều này cho thấy mô hình dễ bị "đánh lừa" bởi nhiễu cục bộ, gây ra rủi ro dự báo bị quan sai lệch so với thực tế vận hành bền vững của các nhóm khách hàng chiến lược như Tier Gold.

Tính ổn định và Khả năng bám sát xu hướng của Prophet Mô hình Prophet thể hiện ưu thế vượt trội trong việc duy trì đà tăng trưởng tự nhiên của doanh nghiệp. Bằng cách tách biệt thành phần xu hướng (*trend*) và điểm chuyển đổi (*change points*) linh hoạt, Prophet đưa ra kịch bản tăng trưởng tích cực từ **+11,33%** đến **+43,89%**. Kết quả này phù hợp hơn với quan sát về sự mở rộng của nhóm Retail và nhóm Gold, vốn là những phân khúc ít bị chi phối bởi nhiễu ngắn hạn.

Kết luận và Hàm ý quản trị Chúng tôi đề xuất lựa chọn mô hình Prophet làm cơ sở hoạch định chuỗi cung ứng. Việc sử dụng mức dự báo trung bình **16.244,72** (tăng 43,89%) từ Prophet giúp doanh nghiệp chuẩn bị đủ nguồn lực tồn kho để đáp ứng nhu cầu thị trường, tránh việc thiếu hụt hàng hóa do các nhận định tiêu cực từ SARIMA.

Mô hình	Dự báo TB	Tăng trưởng	Độ tin cậy	Đề xuất
SARIMA (STL)	10.119,08	-10,37%	Thấp	Loại bỏ (Bị nhiễu)
Prophet	16.244,72	+43,89%	Cao	Lựa chọn

Chương 8

Thảo luận

8.1 Thảo luận về đặc điểm doanh thu chung

8.1.1 Thảo luận về kết quả phân tích

Kết quả phân tích cho thấy doanh thu có xu hướng tăng trưởng theo thời gian nhưng chịu ảnh hưởng mạnh của yếu tố mùa vụ và biến động ngắn hạn. Các đường làm trơn và chỉ số biến động (CV) cho thấy doanh thu không tăng tuyến tính mà dao động theo chu kỳ tuần và tháng. Điều này hàm ý rằng việc đánh giá hiệu quả kinh doanh cần được thực hiện trong bối cảnh chu kỳ mùa vụ, tránh so sánh trực tiếp giữa các giai đoạn ngắn hạn.

Doanh thu tập trung cao vào một nhóm nhỏ sản phẩm và khách hàng. Hiện tượng này phù hợp với nguyên lý Pareto trong bán lẻ, trong đó một tỷ lệ nhỏ sản phẩm và khách hàng tạo ra phần lớn doanh thu. Mặc dù giúp tối ưu hiệu quả ngắn hạn, mức độ tập trung cao cũng làm gia tăng rủi ro nếu nhu cầu đối với các đối tượng chủ lực suy giảm.

Về mặt thị trường, sự phụ thuộc lớn vào United Kingdom vừa là lợi thế vừa là rủi ro. Doanh nghiệp hưởng lợi từ sự am hiểu thị trường và hiệu quả vận hành, nhưng đồng thời dễ bị tổn thương trước các cú sốc cầu mang tính khu vực.

8.1.2 Thảo luận về kết quả dự báo bằng SARIMA

Hai mô hình SARIMA cho thấy khả năng nắm bắt tốt xu hướng và mùa vụ của chuỗi doanh thu. Mô hình cơ sở phù hợp cho giám sát xu hướng dài hạn, trong khi mô hình cải tiến đạt độ chính xác cao hơn trong dự báo ngắn hạn, đặc biệt ở các giai đoạn biến động mạnh.

Sự đánh đổi giữa tính đơn giản và độ chính xác dự báo là rõ ràng. Kết quả này gợi ý rằng việc lựa chọn mô hình cần gắn với mục tiêu quản trị cụ thể: theo dõi xu hướng tổng thể hay phục vụ lập kế hoạch tác nghiệp ngắn hạn.

8.1.3 Hàm ý tổng hợp

Tổng hợp các kết quả cho thấy doanh thu tăng trưởng nhưng chưa bền vững do phụ thuộc vào mùa vụ, sản phẩm và khách hàng chủ lực. Việc kết hợp phân tích mô tả và dự báo chuỗi thời gian giúp doanh nghiệp không chỉ hiểu cấu trúc doanh thu hiện tại mà còn chủ động hơn trong hoạch định tồn kho, phân bổ nguồn lực và xây dựng chiến lược khách hàng theo từng phân khúc.

8.2 Thảo luận về doanh thu theo nhóm khách hàng

Phân tích theo Tier cho thấy sự phân hóa rõ rệt về động lực doanh thu. Nhóm Gold đóng vai trò trụ cột với xu hướng tăng trưởng ổn định và mức nhiễu thấp, cho thấy hành vi mua có tính dự đoán cao. Ngược lại, nhóm Silver và đặc biệt là Bronze thể hiện mức độ biến động lớn và xu hướng kém bền vững hơn, phản ánh vai trò chủ yếu trong việc tạo dao động ngắn hạn thay vì giá trị dài hạn.

Sự khác biệt giữa kênh bán lẻ và bán sỉ phản ánh sự dịch chuyển cấu trúc doanh thu. Doanh thu bán lẻ tăng trưởng dương và ổn định hơn, trong khi doanh thu bán sỉ suy giảm và có mức biến động cao. Điều này cho thấy mô hình kinh doanh đang dịch chuyển dần sang phục vụ tiêu dùng cuối, với yêu cầu cao hơn về khả năng phản ứng nhanh và quản trị tồn kho linh hoạt.

Chương 9

Kết luận và Hướng phát triển

9.1 Kết luận

Kết quả phân tích cho thấy doanh thu có xu hướng tăng trưởng theo thời gian nhưng chịu ảnh hưởng rõ rệt của yếu tố mùa vụ và biến động ngắn hạn. Doanh thu không phân bổ đồng đều mà tập trung cao vào một số sản phẩm và nhóm khách hàng chủ chốt, phản ánh đặc trưng điển hình của hoạt động bán lẻ.

Phân tích theo nhóm khách hàng cho thấy sự khác biệt rõ rệt về động lực doanh thu. **Nhóm khách hàng giá trị cao và kênh bán lẻ đóng vai trò nền tảng**, với mức độ ổn định tương đối cao và khả năng dự báo tốt. Ngược lại, các nhóm còn lại thể hiện mức độ biến động lớn hơn và tiềm ẩn nhiều rủi ro trong hoạch định.

Kết quả dự báo bằng các mô hình SARIMA cho thấy mô hình có khả năng nắm bắt tốt xu hướng và mùa vụ của chuỗi doanh thu trong ngắn hạn. **Điều này cung cấp cơ sở định lượng quan trọng cho các quyết định về tồn kho, phân bổ nguồn lực và lập kế hoạch bán hàng**, góp phần nâng cao hiệu quả quản trị chuỗi cung ứng dựa trên dữ liệu.

9.2 Hướng phát triển

Trong giai đoạn tiếp theo, nghiên cứu có thể được mở rộng theo hai hướng chính. Thứ nhất, phát triển hệ thống *Dashboard* trực quan nhằm hỗ trợ nhà quản lý theo dõi doanh thu, xu hướng bán hàng và kết quả dự báo theo thời gian thực, qua đó nâng cao khả năng giám sát và ra quyết định dựa trên dữ liệu.

Thứ hai, mở rộng phạm vi phân tích sang các bài toán nghiệp vụ chuyên sâu hơn, bao gồm phân tích hành vi khách vãng lai, xây dựng hệ thống gợi ý mặt hàng phù hợp cho từng nhóm khách hàng, định nghĩa và theo dõi vòng đời (*customer lifecycle*) của khách hàng nhằm nắm bắt đặc điểm và giá trị theo thời gian. Bên cạnh đó, việc tích hợp phân tích quản lý tồn kho để nhận diện sớm các rủi ro *stockout* và *overstock* sẽ góp phần tối ưu hóa chuỗi cung ứng và hiệu quả vận hành tổng thể.

Chương 10

Phụ lục

Tài nguyên và mã nguồn

Toàn bộ mã nguồn, dữ liệu và các kết quả trung gian được sử dụng trong nghiên cứu này được công bố tại các liên kết sau:

- **GitHub Repository:**
https://github.com/WindPham/24C12006_24C11060_DA_models
- **Google Drive (Dataset & Figures):**
<https://drive.google.com/drive/folders/1Geb1TOChDHtS45UEFd7N0ngPhLwGeQpE?usp=sharing>

Các tài nguyên trên bao gồm mã tiền xử lý dữ liệu, mô hình phân tích chuỗi thời gian, kết quả trực quan hóa và các tệp hỗ trợ phục vụ cho quá trình tái lập và mở rộng nghiên cứu.