

ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH  
TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN  
KHOA TOÁN – TIN HỌC

---



BÁO CÁO TRỰC QUAN HÓA DỮ LIỆU  
**BÁO CÁO ĐỒ ÁN CUỐI KỲ**

**Giảng viên:** TS. Tô Đức Khánh

**Sinh viên thực hiện:** Nguyễn Minh Đạt – 22280009

Tp. Hồ Chí Minh, tháng 01/2026

# Mục lục

<b>Danh sách hình vẽ.</b>	<b>iii</b>
<b>Đường dẫn tải xuống Dashboard.</b>	<b>iv</b>
<b>1 Giới thiệu chung</b>	<b>1</b>
1.1 Lý do chọn đề tài.	1
1.2 Giới thiệu bộ dữ liệu.	1
1.3 Mục tiêu đề tài.	2
1.3.1 Mục tiêu tổng quát.	2
1.3.2 Các câu hỏi kinh doanh	3
<b>2 Quy trình xử lý và Ý tưởng thiết kế.</b>	<b>4</b>
2.1 Tiền xử lý dữ liệu.	4
2.1.1 Xử lý dữ liệu bảng Customers.	4
2.1.2 Tối ưu hóa bảng Transactions.	4
2.1.3 Mô hình hóa dữ liệu.	5
2.2 Lựa chọn các chỉ số đo lường chính (KPIs).	5
2.2.1 Nhóm chỉ số Quy mô.	5
2.2.2 Nhóm chỉ số Khách hàng.	5
2.2.3 Nhóm chỉ số Giá trị và Hiệu suất.	6
2.3 Ý tưởng thiết kế dashboard.	6
2.3.1 Sheet 1: Overview.	6
2.3.2 Sheet 2: Customer & Behavior.	7
<b>3 Trình bày Dashboard và Phân tích chi tiết.</b>	<b>9</b>
3.1 Overview.	9
3.1.1 Hệ thống thẻ các chỉ số KPIS.	9
3.1.2 Phân tích xu hướng tăng trưởng.	10
3.1.3 Phân tích cơ cấu và danh mục sản phẩm.	11
3.1.4 Tính tương tác của Dashboard.	13
3.2 Customer & Behavior.	14
3.2.1 Hệ thống thẻ các chỉ số KPIS.	14

3.2.2	Phân tích nhân khẩu học. . . . .	15
3.2.3	Phân tích tương tác marketing và hành vi giá. . . . .	17
3.2.4	Phân tích thị hiếu thẩm mỹ và xu hướng tiêu dùng. . . . .	19
3.2.5	Tính tương tác của Dashboard. . . . .	20
<b>4</b>	<b>Tổng kết và Kiến nghị. . . . .</b>	<b>21</b>
4.1	Tóm tắt các phát hiện quan trọng (Key Insights. . . . .	21
4.2	Đề xuất giải pháp dựa trên dữ liệu. . . . .	22
4.3	Những hạn chế của dữ liệu và phạm vi phân tích . . . . .	23

# Danh sách hình vẽ

1	Overview . . . . .	9
2	Overview's KPIs Cards . . . . .	9
3	Total Revenue & Quantity By Month . . . . .	10
4	Total Revenue By Price Segment . . . . .	11
5	Revenue Share By Product Index . . . . .	12
6	Top 10 Product Type By Total Revenue . . . . .	12
7	Sales Channel Slicer & Page Navigation . . . . .	13
8	Customer & Behavior . . . . .	14
9	Customer & Behavior's KPIs Cards . . . . .	14
10	Total Quantity Sold By Age Group . . . . .	15
11	Distribution Of Customers By Age . . . . .	16
12	Fashion News Frequency . . . . .	17
13	Average Basket Value By Club Member Status . . . . .	18
14	Quantity Sold By Price . . . . .	18
15	Top 10 Graphical Appearance By Quantity Sold . . . . .	19
16	Quantity Sold By Price . . . . .	20

## **Đường dẫn tải xuống Dashboard.**

<https://drive.google.com/file/d/1lYctxd54RNfgIsUe6PWof52lHguDYogk/view?usp=sharing>

# 1. Giới thiệu chung

## 1.1. Lý do chọn đề tài.

Trong bối cảnh ngành thời trang nhanh luôn biến đổi không ngừng, bộ dữ liệu H&M là một lựa chọn điển hình nhờ cấu trúc đa dạng—kết hợp giữa nhân khẩu học khách hàng, chi tiết sản phẩm và lịch sử giao dịch. Thách thức cốt lõi của đề tài không chỉ nằm ở việc xử lý và liên kết các luồng dữ liệu phức tạp, mà còn là chuyển hóa chúng thành những thông tin trực quan, từ đó giải mã quy luật hành vi và các yếu tố quyết định hành vi mua sắm của người tiêu dùng.

## 1.2. Giới thiệu bộ dữ liệu.

Bộ dữ liệu *H&M Personalized Fashion Recommendations* là một trong những tập dữ liệu thực tế lớn nhất trong ngành bán lẻ thời trang, được H&M công bố nhằm thách thức các chuyên gia dữ liệu trong việc xây dựng hệ thống gợi ý (*Recommendation System*) và dự báo hành vi người dùng. Khác với các bộ dữ liệu thương mại điện tử truyền thống chỉ tập trung vào giao dịch, dữ liệu H&M kết hợp đồng thời giữa hành vi mua sắm theo chuỗi thời gian, metadata sản phẩm chuyên sâu và dữ liệu hình ảnh. Với quy mô hơn 31 triệu giao dịch, bộ dữ liệu này yêu cầu khả năng xử lý Big Data và tư duy cá nhân hóa cao độ 1 để tìm ra sự giao thoa giữa xu hướng thời trang nhanh (fast-fashion) và sở thích riêng biệt của từng khách hàng trong tổng số 1.3 triệu người dùng.

**Các đặc điểm nổi bật:**

**File A: articles.csv (Dữ liệu sản phẩm - 105,542 dòng)**

- **article\_id**: Mã định danh duy nhất cho mỗi mặt hàng (khóa chính).
- **product\_code**: Mã dòng sản phẩm.
- **prod\_name**: Tên sản phẩm.
- **product\_type\_no, product\_type\_name**: Mã và tên loại sản phẩm (Váy, Quần, Áo,...).
- **graphical\_appearance\_no, graphical\_appearance\_name**: Kiểu họa tiết (Trơn, Chấm bi, Kẻ caro,...).
- **colour\_group\_code, colour\_group\_name**: Nhóm màu sắc.
- **perceived\_colour\_value\_id, perceived\_colour\_value\_name**: Độ đậm nhạt của màu (Sáng, Tối, Bạc màu,...).
- **index\_code, index\_name**: Phân nhóm cấp cao (Ladieswear, Sport, Baby,...).
- **section\_no, section\_name**: Khu vực trưng bày trong cửa hàng.

- **garment\_group\_no, garment\_group\_name**: Nhóm may mặc (Đồ dệt kim, Đồ lót, ...).
- **detail\_desc**: Mô tả chi tiết sản phẩm bằng văn bản (phục vụ cho phân tích NLP).

**File B: customers.csv (Dữ liệu khách hàng - 1.37 triệu dòng)**

- **customer\_id**: Mã định danh khách hàng (Dạng băm - Hash).
- **FN**: Trạng thái nhận thông báo thời trang (1 nếu có).
- **Active**: Trạng thái hoạt động của khách hàng.
- **club\_member\_status**: Trạng thái thành viên (Active, Pre-create, Left).
- **fashion\_news\_frequency**: Tần suất nhận tin thời trang (Regular, Monthly).
- **age**: Tuổi khách hàng.
- **postal\_code**: Mã bưu điện (Dạng băm).

**File C: transactions\_train.csv (Giao dịch - 31.7 triệu dòng)**

- **t\_dat**: Ngày giao dịch.
- **customer\_id**: Liên kết với bảng khách hàng.
- **article\_id**: Liên kết với bảng sản phẩm.
- **price**: Giá sản phẩm (đã được scale/normalize).
- **sales\_channel\_id**: Kênh bán hàng (1: Online, 2: Tại cửa hàng).

### 1.3. Mục tiêu đề tài.

#### 1.3.1. Mục tiêu tổng quát.

Mục tiêu cốt lõi của Dashboard là cung cấp một bức tranh toàn cảnh về hiệu suất kinh doanh của H&M trong năm 2020, đóng vai trò như một công cụ hỗ trợ ra quyết định chiến lược cho bộ phận Quản lý và Marketing thông qua ba trụ cột chính. Đầu tiên, dự án tập trung phân tích hành vi khách hàng bằng cách phân khúc hóa đối tượng mục tiêu dựa trên nhân khẩu học và hành vi, nhằm xác định các nhóm khách hàng trung thành và tiềm năng. Thứ hai, thông qua việc phân tích các chỉ số kinh doanh để nhận diện những sản phẩm, màu sắc và họa tiết chủ đạo, Dashboard hỗ trợ tối ưu hóa danh mục hàng hóa và bắt kịp các xu hướng theo mùa. Cuối cùng, bằng cách đánh giá sự dịch chuyển hành vi giữa kênh Online và Offline trong bối cảnh thị trường biến động năm 2020, Dashboard cung cấp cơ sở dữ liệu quan trọng để xây dựng chiến lược chuyển đổi kênh phân phối, giúp doanh nghiệp điều phối nguồn lực và ngân sách tiếp thị một cách hiệu quả nhất.

### 1.3.2. Các câu hỏi kinh doanh

Để đạt được các mục tiêu trên, Dashboard tập trung giải quyết 5 nhóm câu hỏi cốt lõi sau:

#### Nhóm 1: Hiệu suất tăng trưởng (Sales Performance)

- Doanh thu và số lượng giao dịch biến động như thế nào theo từng tháng trong năm 2020?
- Những giai đoạn nào được xem là *cao điểm* của H&M?

#### Nhóm 2: Phân tích Kênh và Sản phẩm (Channel & Product Insights)

- Tỷ lệ đóng góp doanh thu giữa kênh *Online* và *Cửa hàng* có sự chênh lệch như thế nào?
- Nhóm sản phẩm nào mang lại doanh thu cao nhất? Top 10 sản phẩm cụ thể nào được khách hàng mua nhiều nhất?

#### Nhóm 3: Chân dung và Hành vi khách hàng (Customer Demographics)

- Phân khúc độ tuổi nào là nhóm khách hàng có mức chi tiêu cao nhất?
- Trạng thái thành viên và việc nhận tin tức thời trang có ảnh hưởng đáng kể đến giá trị đơn hàng trung bình hay không?

#### Nhóm 4: Xu hướng thiết kế (Design Trends)

- Những tổ hợp giữa màu sắc và họa tiết nào đang chiếm ưu thế trên thị trường?
- Khách hàng có xu hướng ưu tiên các sản phẩm cơ bản hay các thiết kế có họa tiết phức tạp hơn?

#### Nhóm 5: Chiến lược giá (Pricing Strategy)

- Mối quan hệ giữa mức giá sản phẩm và số lượng bán ra thể hiện xu hướng như thế nào?
- Hệ số co giãn của cầu theo giá: các mặt hàng có giá cao có duy trì được lượng giao dịch ổn định hay nhu cầu chủ yếu tập trung vào phân khúc giá thấp?



## 2. Quy trình xử lý và Ý tưởng thiết kế.

### 2.1. Tiền xử lý dữ liệu.

Do đặc thù bộ dữ liệu H&M có quy mô cực lớn với hàng chục triệu bản ghi, quy trình tiền xử lý dữ liệu đóng vai trò quyết định để đảm bảo tính ổn định và hiệu suất của Dashboard trên Power BI.

#### 2.1.1. Xử lý dữ liệu bảng Customers.

**Chuẩn hóa trạng thái hoạt động:** Nhằm đưa dữ liệu về dạng có thể định lượng và phục vụ cho các phép toán thống kê, các giá trị “null” trong hai cột FN và Active được thay thế bằng giá trị 0.

**Xử lý dữ liệu thành viên:** Đối với cột club\_member\_status, các giá trị trống được thay thế bằng nhãn “Unknown”. Song song đó, cột fashion\_news\_frequency được đồng bộ hóa bằng cách chuyển tất cả các giá trị “NONE” về một chuẩn duy nhất là “None”, đồng thời các ô trống (blank) cũng được lấp đầy bằng giá trị này để tránh sự sai lệch giữa các biến định tính tương đương.

**Phân nhóm đối tượng khách hàng:** Biến định lượng age được chuyển đổi thành biến định tính age\_group thông qua logic điều kiện. Cụ thể, các khách hàng có độ tuổi bị trống được gán vào nhóm “Unknown”. Những khách hàng có độ tuổi nhỏ hơn 25 được phân loại vào nhóm “Young”, trong khi nhóm “Adult” bao gồm các khách hàng trong khoảng từ 25 đến dưới 54 tuổi. Các khách hàng từ 55 tuổi trở lên được xếp vào nhóm “Senior”.

#### 2.1.2. Tối ưu hóa bảng Transactions.

Với hơn 31 triệu giao dịch, việc tải toàn bộ dữ liệu vào Power BI có thể gây quá tải và làm giảm hiệu suất xử lý. Do đó, áp dụng kỹ thuật lọc dữ liệu tại nguồn trong Power Query bằng cách thiết lập các điều kiện lọc chặt chẽ trong tính năng Date Filters, dữ liệu được giới hạn chính xác trong năm tài chính 2020 ( $01/01/2020 \leq t \leq 31/12/2020$ ).

Nhằm sâu sắc hóa việc phân tích chiến lược giá và hành vi chi tiêu của khách hàng, một thuộc tính mới mang tên Price Segment đã được thiết lập. Cụ thể, các giao dịch được phân loại vào ba phân khúc chiến lược bao gồm: phân khúc giá rẻ (Budget) cho các sản phẩm có mức giá  $price \leq 0.02$ , phân khúc phổ thông (Regular) khi mức giá nằm trong khoảng  $0.02 < price \leq 0.05$ , và phân khúc cao cấp (Premium) đối với các sản phẩm có giá trị  $price > 0.05$ .

### 2.1.3. Mô hình hóa dữ liệu.

Sau khi làm sạch, các bảng được kết nối với nhau theo mô hình Star Schema để tối ưu hóa việc truy vấn, trong đó bảng Transactions đóng vai trò là bảng Trung tâm còn các bảng Customers và Articles đóng vai trò là các bảng Xung quanh, liên kết qua các khóa ngoại `customer_id` và `article_id`.

## 2.2. Lựa chọn các chỉ số đo lường chính (KPIs).

Để đánh giá hiệu quả hoạt động kinh doanh và hành vi khách hàng của H&M trong năm 2020, Dashboard tập trung vào 5 chỉ số đo lường cốt lõi (KPIs). Các chỉ số này được tính toán bằng ngôn ngữ DAX để đảm bảo tính linh hoạt khi lọc dữ liệu.

### 2.2.1. Nhóm chỉ số Quy mô.

Nhóm chỉ số này được sử dụng nhằm đánh giá quy mô thị trường, mức độ phủ sóng thương hiệu và hiệu quả tăng trưởng tổng thể của doanh nghiệp.

**Total Revenue (Tổng doanh thu):** được tính bằng tổng giá trị các giao dịch phát sinh trong tập dữ liệu:

$$\text{Total Revenue} = \text{SUM}(\text{transactions\_train}[\text{price}])$$

Chỉ số này phản ánh tổng giá trị giao dịch sau khi đã được chuẩn hóa và là thước đo trực tiếp nhất để đánh giá tăng trưởng doanh thu cũng như hiệu quả tài chính tổng thể của doanh nghiệp trong giai đoạn phân tích.

**Total Quantity Sold (Tổng sản lượng bán ra):** được xác định thông qua số lượng các dòng giao dịch:

$$\text{Total Quantity Sold} = \text{COUNTROWS}(\text{transactions\_train})$$

Chỉ số này đo lường số lượng sản phẩm đã được tiêu thụ và đặc biệt quan trọng đối với ngành thời trang nhanh, nơi vòng đời sản phẩm ngắn và tốc độ quay vòng tồn kho đóng vai trò then chốt trong chiến lược vận hành.

### 2.2.2. Nhóm chỉ số Khách hàng.

Nhóm chỉ số này tập trung đánh giá khả năng thu hút và duy trì khách hàng của doanh nghiệp.

**Total Customers (Tổng số khách hàng):** được xác định bằng số lượng khách hàng duy nhất có phát sinh giao dịch trong năm 2020:

$$\text{Total Customers} = \text{DISTINCTCOUNT}(\text{transactions\_train}[\text{customer\_id}])$$

Chỉ số này giúp phân biệt giữa "tăng trưởng nhờ khách hàng mới" hay "tăng trưởng nhờ khách hàng cũ mua nhiều lần", từ đó hỗ trợ đánh giá chiến lược giữ chân khách hàng.

### 2.2.3. Nhóm chỉ số Giá trị và Hiệu suất.

Nhóm chỉ số này đi sâu vào chất lượng đơn hàng và hành vi chi tiêu của khách hàng.

**Average Order Value - AOV (Giá trị đơn hàng trung bình):** được xác định bằng tỷ lệ giữa tổng doanh thu và tổng số giao dịch:

$$\text{Average Order Value} = \text{DIVIDE}([\text{TotalRevenue}], [\text{TotalQuantitySold}])$$

Chỉ số này cho biết trung bình mỗi lần mua sắm, khách hàng chi tiêu bao nhiêu. Đây là chỉ số then chốt để đánh giá hiệu quả của các chiến lược *Upselling* và *Cross-selling*, cũng như mức độ thành công của các chiến dịch tối ưu hóa giỏ hàng.

**Quantity Sold (Sản lượng theo mặt hàng):** Sản lượng bán ra theo từng mặt hàng được tính bằng số lần xuất hiện của mỗi `article_id` trong bảng giao dịch:

$$\text{Quantity Sold} = \text{COUNT}(\text{transactions\_train}[\text{article\_id}])$$

Chỉ số này được sử dụng để xác định các sản phẩm *Best-seller*, đồng thời cung cấp thông tin quan trọng cho bộ phận thu mua và thiết kế trong việc nhận diện xu hướng về mẫu mã, màu sắc và chất liệu cho các mùa kinh doanh tiếp theo.

## 2.3. Ý tưởng thiết kế dashboard.

Để hiện thực hóa các mục tiêu phân tích, Dashboard được thiết kế theo cấu trúc 2 trang riêng biệt nhưng có tính liên kết chặt chẽ, đi từ kết quả kinh doanh tổng thể đến chi tiết hành vi khách hàng.

### 2.3.1. Sheet 1: Overview.

Mục tiêu của trang phân tích này là cung cấp một cái nhìn vĩ mô về hiệu suất bán hàng, từ đó hỗ trợ nhà quản lý trả lời câu hỏi trọng tâm: “*Doanh nghiệp đang vận hành ra sao và đâu là các dòng sản phẩm chủ lực?*”. Trang tổng quan đóng vai trò như một bảng điều khiển chiến lược, cho phép đánh giá nhanh tình trạng hoạt động trước khi đi sâu vào các phân tích chi tiết hơn.

**Hệ thống thẻ số KPIs:** Các chỉ số cốt lõi được trình bày dưới dạng thẻ số (*KPI Cards*) và bố trí tại khu vực phía trên của dashboard. Cách sắp xếp này giúp người xem nhanh chóng nắm bắt các chỉ số mang tính sống còn, bao gồm Tổng doanh thu, Tổng sản lượng bán ra, Tổng số khách hàng và Giá trị đơn hàng trung bình. Nhờ đó, tình hình kinh doanh tổng thể được truyền tải một cách trực quan và tức thời.

**Phân tích xu hướng theo thời gian:** Để phân tích sự biến động doanh thu theo thời gian, biểu đồ cột kết hợp với đường (*Line and Stacked Column Chart*) được sử dụng với trục thời gian theo tháng trong năm 2020. Cách tiếp cận này cho phép nhận diện các đặc điểm mang tính chu kỳ của ngành thời trang nhanh, đồng thời phát hiện các điểm đột biến bất thường về doanh thu, từ đó hỗ trợ việc đánh giá tác động của yếu tố mùa vụ và các chiến dịch kinh doanh.

**Cơ cấu danh mục sản phẩm:** Cơ cấu danh mục sản phẩm được phân tích thông qua sự kết hợp giữa biểu đồ cột ngang (*Bar Chart*) và biểu đồ Donut. Các biểu đồ này phản ánh tỷ trọng đóng góp doanh thu của từng nhóm hàng, mặt hàng cụ thể cũng như phân khúc giá. Nhờ đó, doanh nghiệp có thể phân loại sản phẩm theo ma trận doanh thu và tốc độ bán lẻ, hỗ trợ việc xác định danh mục sản phẩm chủ lực và định hướng chiến lược kinh doanh trong các giai đoạn tiếp theo.

### 2.3.2. Sheet 2: Customer & Behavior.

Mục tiêu của trang phân tích này là chuyển trọng tâm từ “*Sản phẩm*” sang “*Con người*”, qua đó hỗ trợ các quyết định liên quan đến Marketing, cá nhân hóa trải nghiệm khách hàng và chiến lược định giá. Trang dashboard này giúp doanh nghiệp hiểu rõ ai là người tạo ra doanh thu, họ tương tác với thương hiệu như thế nào và yếu tố nào ảnh hưởng đến hành vi chi tiêu.

**Phân khúc nhân khẩu học:** Doanh thu được trực quan hóa theo các nhóm tuổi nhằm xác định “*phân khúc vàng*” — nhóm khách hàng đóng góp lợi nhuận lớn nhất cho doanh nghiệp. Kết quả phân tích này đóng vai trò định hướng cho việc xây dựng thông điệp truyền thông và ngôn ngữ thiết kế phù hợp trong các chiến dịch quảng cáo và tiếp thị.

**Hiệu quả kênh Marketing và mức độ tương tác:** Để đánh giá tác động của các kênh tiếp thị, biểu đồ so sánh theo biến *fashion\_news\_frequency* được sử dụng nhằm phân tích sự khác biệt trong hành vi mua sắm giữa các nhóm khách hàng có và không thường xuyên nhận tin tức thời trang. Đây là cơ sở quan trọng để kiểm chứng hiệu quả của kênh Email Marketing đối với việc thúc đẩy doanh thu và tần suất mua hàng.

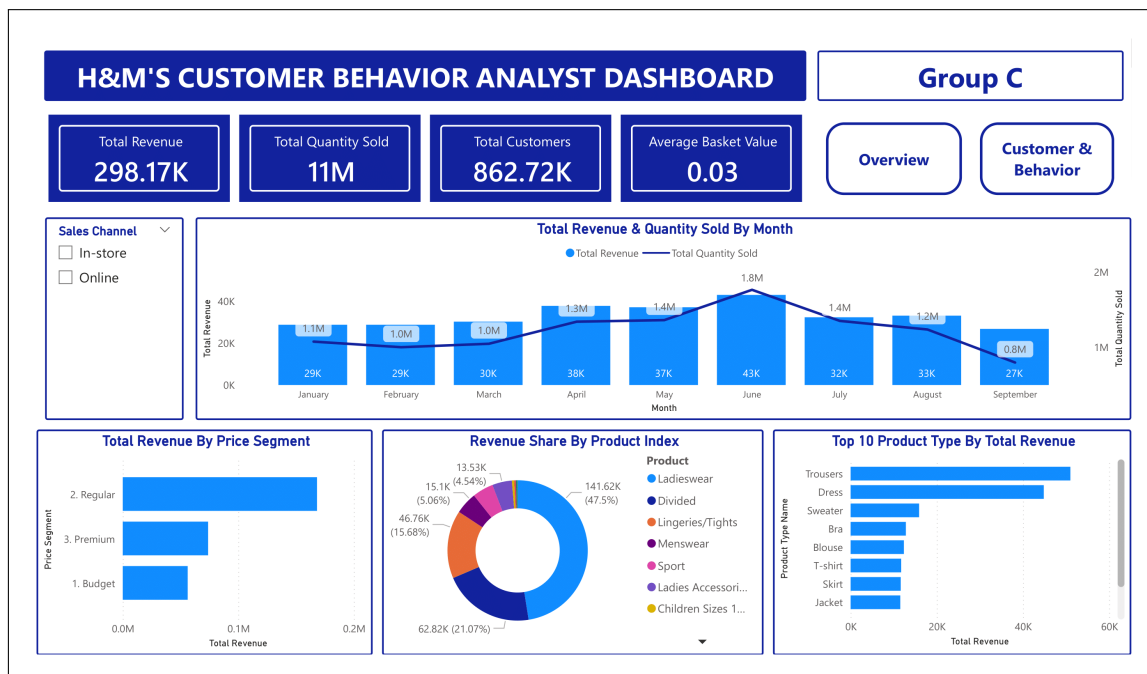
**Lòng trung thành và Giá trị khách hàng:** Mức chi tiêu trung bình được phân tích theo trạng thái thành viên nhằm đánh giá vai trò của các chính sách khách hàng thân thiết. Phân tích này giúp xác định liệu các chương trình ưu đãi dành cho hội viên có thực sự gia tăng giá trị đơn hàng trung bình hay chỉ đơn thuần tạo ra chi phí marketing mà không mang lại hiệu quả tương xứng.

**Phân tích độ nhạy giá:** Mối quan hệ giữa đơn giá sản phẩm và sản lượng bán ra được trực quan hóa nhằm phân tích độ nhạy cảm của khách hàng đối với giá. Thông qua đó, doanh nghiệp có thể xác định “*điểm ngọt*” của mức giá — nơi mà khách hàng sẵn sàng chấp nhận chi trả cao hơn mà không làm suy giảm đáng kể nhu cầu mua sắm.

**Phân tích thị hiếu thẩm mỹ:** Thị hiếu thẩm mỹ của khách hàng được phân tích thông qua việc xếp hạng top 10 họa tiết và màu sắc theo tổng doanh thu. Những kết quả này cung cấp insight trực tiếp cho bộ phận thiết kế và phát triển sản phẩm, giúp nắm bắt xu hướng thẩm mỹ của thị trường và định hướng chiến lược thiết kế cho các mùa kinh doanh tiếp theo.

### 3. Trình bày Dashboard và Phân tích chi tiết.

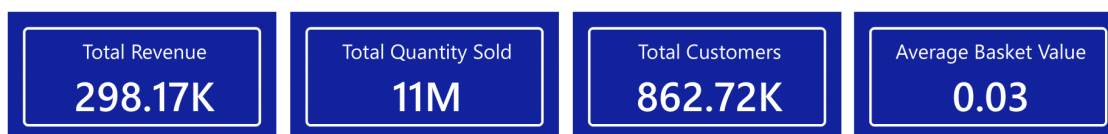
#### 3.1. Overview.



Hình 1: Overview

Trang Overview được thiết kế để cung cấp cái nhìn toàn diện về hiệu suất của H&M trong năm 2020. Dashboard sử dụng bố cục phân tầng thông tin rõ rệt từ trên xuống dưới.

##### 3.1.1. Hệ thống thẻ các chỉ số KPIS.



Hình 2: Overview's KPIs Cards

Được bố trí tại vị trí trên cùng dashboard, hệ thống thẻ các chỉ số KPIS cung cấp cái nhìn tức thì về quy mô vận hành và hiệu quả tổng thể của hoạt động kinh doanh.

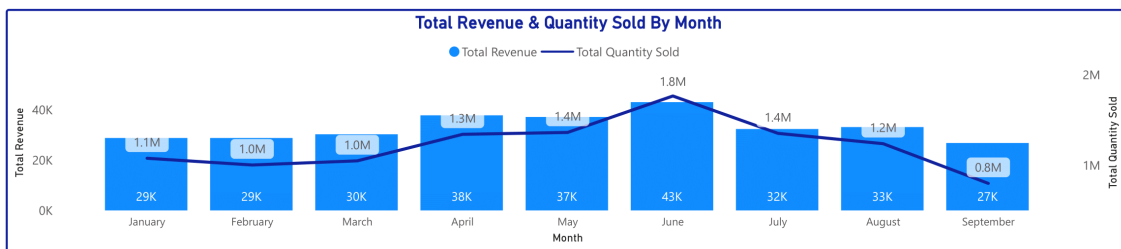
Chỉ số *Total Revenue* đạt mức 298.17K cùng với *Total Quantity Sold* lên tới 11 triệu đơn vị phản ánh tổng giá trị doanh thu và sản lượng hàng hóa đã được tiêu thụ trong giai đoạn phân tích. Quy mô bán ra ở mức hàng chục triệu sản phẩm cho thấy áp lực rất lớn đặt lên hệ thống logistics, chuỗi cung ứng và công tác quản lý

tồn kho, vốn là những yếu tố sống còn trong ngành thời trang nhanh.

Chỉ số *Total Customers* ghi nhận 862.72K khách hàng duy nhất có phát sinh giao dịch, cho thấy doanh nghiệp đang sở hữu một tệp khách hàng có quy mô rất lớn. Đây được xem là tài sản chiến lược quan trọng, tạo nền tảng cho các hoạt động phân tích hành vi, cá nhân hóa trải nghiệm và triển khai các chiến dịch CRM trong dài hạn.

Đối với *Average Basket Value (AOV)*, giá trị trung bình đạt mức 0.03. Cần lưu ý rằng dữ liệu giá trong bộ dữ liệu H&M đã được chuẩn hóa từ nguồn gốc, do đó chỉ số này không phản ánh giá trị tiền tệ tuyệt đối mà thể hiện giá trị giỏ hàng trung bình trên đơn vị sản phẩm. Trong bối cảnh này, AOV đóng vai trò như một chỉ số tương đối, hữu ích cho việc so sánh hiệu quả chi tiêu giữa các phân khúc khách hàng và nhóm sản phẩm thay vì dùng để đánh giá doanh thu tuyệt đối.

### 3.1.2. Phân tích xu hướng tăng trưởng.



Hình 3: Total Revenue & Quantity By Month

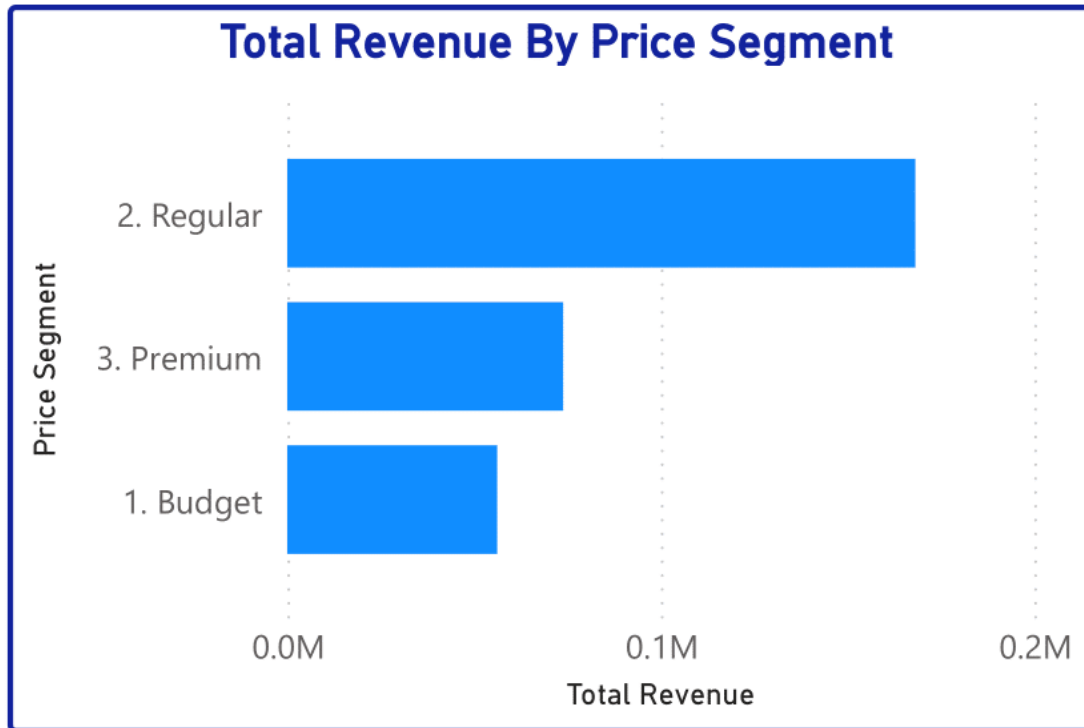
Biểu đồ kết hợp *Total Revenue* và *Quantity Sold by Month* được đặt ở vị trí trung tâm của trang báo cáo. Trục hoành của biểu đồ thể hiện dòng thời gian theo tháng, kéo dài từ tháng 1 đến tháng 9 năm 2020. Trong khi đó, doanh thu được biểu diễn bằng biểu đồ cột (Bar), còn sản lượng bán ra được thể hiện bằng biểu đồ đường (Line), cho phép quan sát đồng thời mối quan hệ giữa giá trị và khối lượng tiêu thụ theo thời gian.

Từ góc độ phân tích, biểu đồ phản ánh rõ rệt tính mùa vụ trong hoạt động kinh doanh. Doanh thu và sản lượng tăng trưởng tương đối ổn định trong nửa đầu năm và đạt đỉnh vào tháng 6, với mức doanh thu xấp xỉ 43 nghìn. Sự chuyển động gần như đồng pha giữa đường sản lượng và cột doanh thu cho thấy giá bán trung bình không có biến động lớn theo thời gian, hàm ý rằng tăng trưởng doanh thu chủ yếu được dẫn dắt bởi khối lượng tiêu thụ thay vì thay đổi về giá.

Tuy nhiên, xu hướng sụt giảm mạnh vào tháng 9, khi doanh thu giảm xuống còn khoảng 27 nghìn, là một tín hiệu đáng chú ý từ góc độ quản trị. Diễn biến này gợi ý nhu cầu triển khai các chương trình kích cầu trong ngắn hạn, đồng thời chuẩn bị sớm cho các bộ sưu tập Thu-Đông nhằm khôi phục đà tăng trưởng và hạn chế rủi ro suy giảm doanh thu trong các tháng tiếp theo.

### 3.1.3. Phân tích cơ cấu và danh mục sản phẩm.

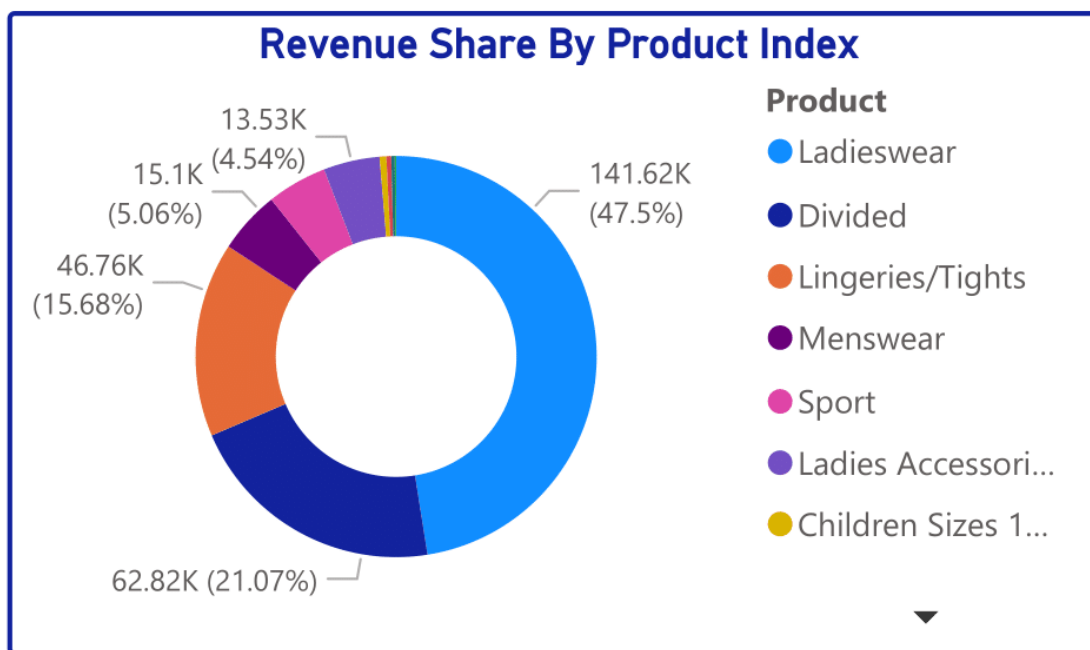
Nhóm biểu đồ trong mục này nhằm giải mã “ADN” doanh thu của H&M thông qua ba góc độ phân tích chính, bao gồm phân khúc giá, tỷ trọng nhóm hàng và danh mục sản phẩm chủ lực.



**Hình 4:** Total Revenue By Price Segment

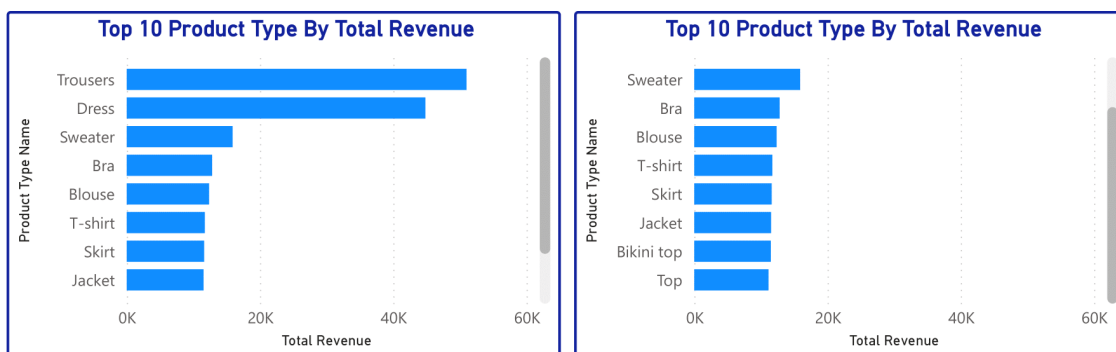
Xét theo phân khúc giá, nhóm sản phẩm thuộc phân khúc *Regular* chiếm ưu thế tuyệt đối trong cơ cấu doanh thu. Kết quả này cho thấy chiến lược định vị “thời trang đại chúng” của H&M là hoàn toàn phù hợp, khi thương hiệu tập trung phục vụ nhóm khách hàng đại trà với mức chi trả trung bình và tần suất mua sắm cao, từ đó tạo ra nguồn doanh thu ổn định và bền vững.





**Hình 5:** Revenue Share By Product Index

Ở góc độ cơ cấu nhóm hàng, phân tích tỷ trọng doanh thu theo *Index* cho thấy nhóm *Ladieswear* đóng góp tới 47.5% tổng doanh thu, gần gấp đôi so với nhóm *Divided* với tỷ trọng 21.07%. Điều này khẳng định phân khúc khách hàng nữ là nhóm khách hàng cốt lõi, đóng vai trò như “nguồn sống” của thương hiệu. Do đó, các chiến dịch marketing, phát triển sản phẩm và phân bổ ngân sách truyền thông cần ưu tiên mạnh mẽ cho nhóm hàng này nhằm tối đa hóa hiệu quả kinh doanh.



**Hình 6:** Top 10 Product Type By Total Revenue

Cuối cùng, xét theo danh mục sản phẩm chủ lực, các mặt hàng cơ bản như *Trousers* và *Dress* dẫn đầu danh sách mười nhóm sản phẩm có doanh thu cao nhất. Việc duy trì doanh số lớn ở các mặt hàng thiết yếu này giúp H&M đảm bảo dòng tiền ổn định và giảm thiểu rủi ro so với các nhóm sản phẩm mang tính xu hướng ngắn hạn như *Jacket* hay *Accessory*, vốn có mức độ biến động cao theo mùa và thị hiếu.

### 3.1.4. Tính tương tác của Dashboard.

Một trong những điểm mạnh nổi bật của Dashboard này nằm ở khả năng tương tác linh hoạt, cho phép người dùng không chỉ quan sát dữ liệu dưới dạng tĩnh mà còn chủ động khai thác thông tin theo các nhu cầu phân tích cụ thể.



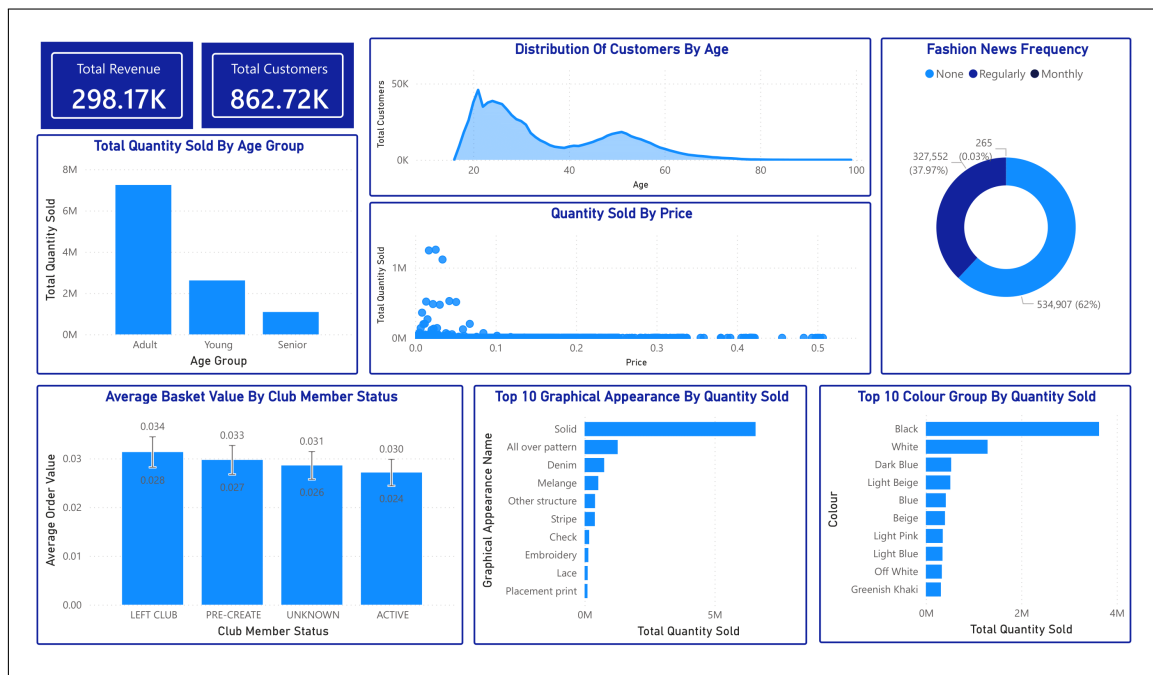
**Hình 7:** Sales Channel Slicer & Page Navigation

Trước hết, bộ lọc kênh bán hàng được bố trí ở phía bên trái trang báo cáo, cho phép người dùng chuyển đổi nhanh chóng giữa dữ liệu bán hàng Online và In-store. Việc tích hợp Slicer giúp nhà quản lý dễ dàng so sánh hiệu suất kinh doanh giữa các kênh phân phối, từ đó đưa ra các điều chỉnh chiến lược phù hợp cho từng nền tảng. Điều này đặc biệt có ý nghĩa trong bối cảnh năm 2020, khi hành vi mua sắm của khách hàng có sự dịch chuyển mạnh mẽ sang các kênh số.

Bên cạnh đó, Dashboard được thiết kế với hệ thống điều hướng trang chuyên nghiệp, thay thế cho thanh tab mặc định ở phía dưới của Power BI. Các nút điều hướng được bố trí ở góc trên bên phải, giúp người dùng di chuyển giữa các trang báo cáo một cách trực quan và nhất quán. Cụ thể, nút *Customer & Behavior* cho phép chuyển nhanh sang trang phân tích chi tiết về chân dung khách hàng và hành vi mua sắm, qua đó hỗ trợ đào sâu insight mà không làm gián đoạn mạch tư duy phân tích tổng thể.

Cuối cùng, toàn bộ các biểu đồ trên trang *Overview* được liên kết với nhau thông qua cơ chế tương tác chéo. Khi người dùng lựa chọn một trường ở một biểu đồ cụ thể, chẳng hạn như *Ladieswear* trên biểu đồ Tỷ trọng doanh thu các nhóm hàng, tất cả các chỉ số KPI, biểu đồ xu hướng theo thời gian và danh sách sản phẩm chủ lực sẽ tự động được lọc để chỉ phản ánh dữ liệu của nhóm hàng đó. Cơ chế này giúp người dùng nhanh chóng chuyển từ góc nhìn tổng quan sang phân tích chuyên sâu mà không cần thao tác lọc thủ công, nâng cao đáng kể hiệu quả khai thác dữ liệu.

### 3.2. Customer & Behavior.



Hình 8: Customer & Behavior

Trong khi trang Overview tập trung vào kết quả kinh doanh, trang Customer & Behavior được thiết kế để giải mã các yếu tố thúc đẩy đằng sau những con số đó. Trang này tập trung trả lời câu hỏi chiến lược: "Ai là người đang thực sự nuôi sống doanh nghiệp và thói quen tiêu dùng của họ là gì?"

#### 3.2.1. Hệ thống thẻ các chỉ số KPIS.



Hình 9: Customer & Behavior's KPIs Cards

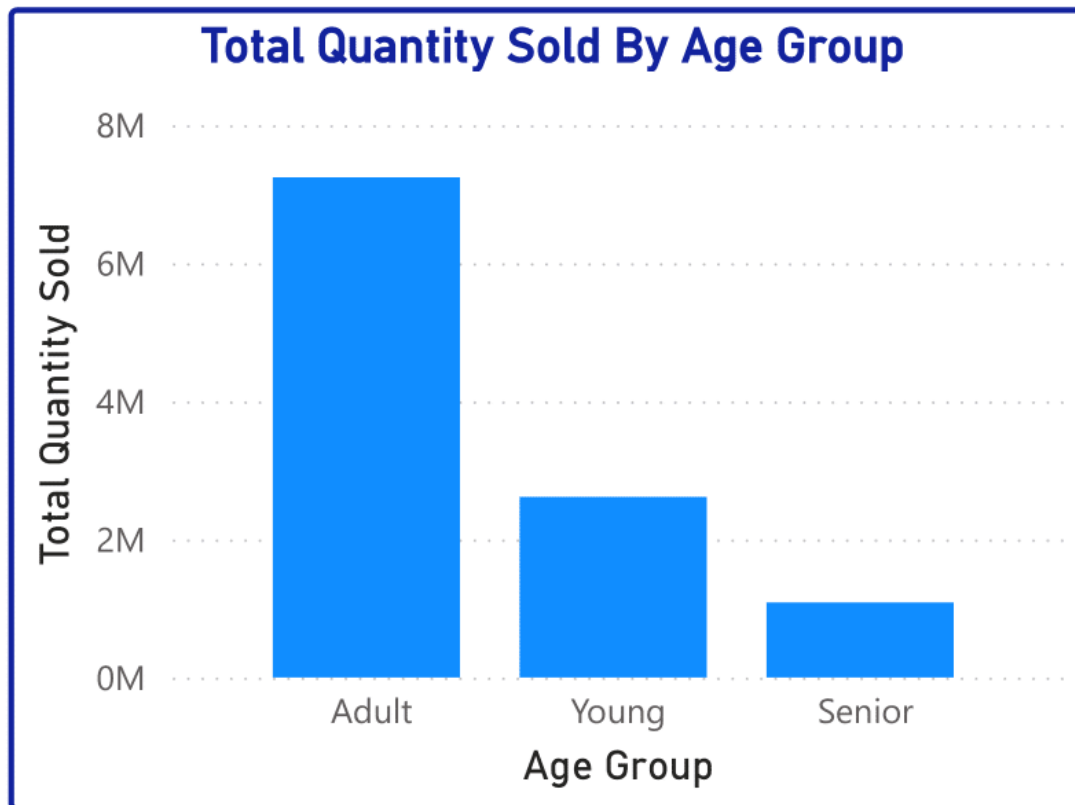
Dashboard duy trì hai chỉ số cốt lõi là *Total Revenue* (298.17K) và *Total Customers* (862.72K) tại vị trí góc trái của trang báo cáo. Việc giữ cố định các chỉ số này đóng vai trò như một "mỏ neo" tham chiếu, giúp người xem luôn nhận thức rõ quy mô

giá trị kinh tế và quy mô tệp khách hàng trong suốt quá trình phân tích.

Trong bối cảnh trang báo cáo tập trung mạnh vào các biểu đồ liên quan đến sản lượng bán ra (*Quantity*), hệ thống chỉ số tham chiếu này giúp tránh sự nhầm lẫn giữa số lượng giao dịch lớn và mức đóng góp doanh thu thực tế. Nhờ đó, nhà quản lý có thể đánh giá chính xác liệu sự gia tăng về sản lượng có thực sự tạo ra giá trị kinh tế tương xứng, hay chỉ phản ánh các mặt hàng có đơn giá thấp, từ đó hỗ trợ việc ra quyết định về chiến lược giá và danh mục sản phẩm.

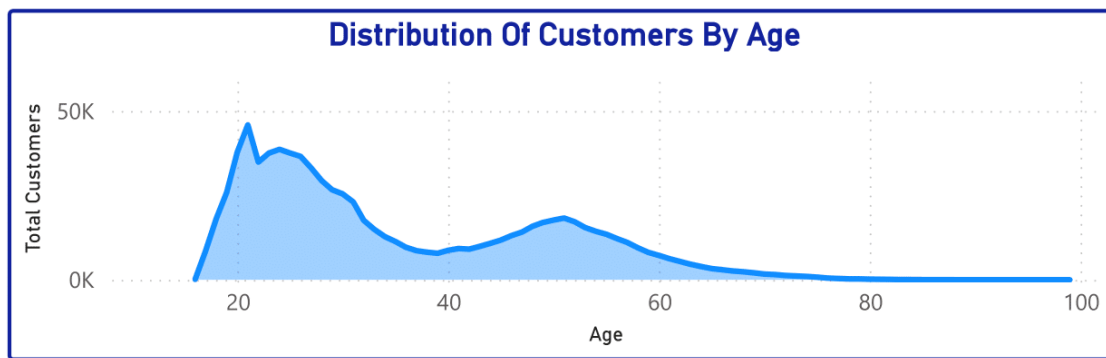
### 3.2.2. Phân tích nhân khẩu học.

Phân tích nhân khẩu học làm rõ đặc điểm độ tuổi của khách hàng H&M thông qua việc kết hợp giữa quy mô tiêu thụ và mật độ phân bố khách hàng. Cách tiếp cận này cho phép đánh giá đồng thời *ai là nhóm khách hàng đông nhất* và *ai là nhóm tạo ra sản lượng lớn nhất*.



**Hình 10:** Total Quantity Sold By Age Group

Xét theo sản lượng tiêu thụ, biểu đồ *Quantity Sold by Age Group* cho thấy nhóm *Adult* (25–54 tuổi) chiếm ưu thế tuyệt đối với hơn 7 triệu sản phẩm được bán ra, cao gấp gần ba lần so với nhóm *Young*. Kết quả này phản ánh vai trò của nhóm khách hàng trưởng thành như một “động cơ chính” thúc đẩy doanh số nhờ mức thu nhập ổn định và nhu cầu mua sắm thường xuyên.



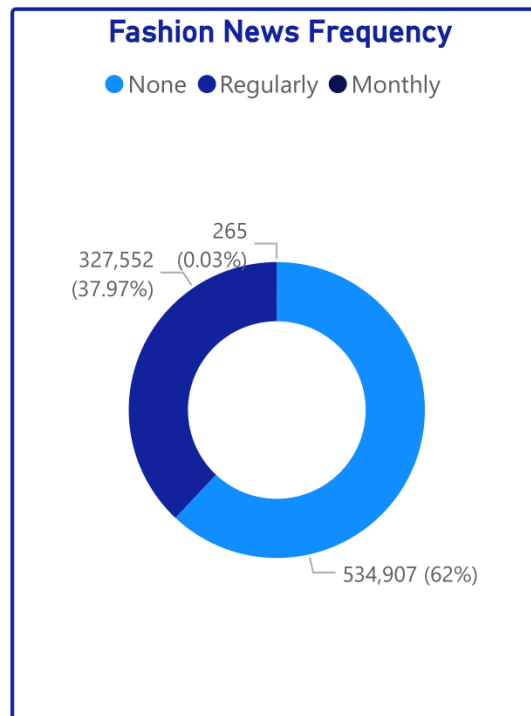
**Hình 11:** Distribution Of Customers By Age

Ở góc độ phân bố khách hàng, biểu đồ vùng thể hiện mật độ khách hàng theo độ tuổi cho thấy đỉnh tập trung rõ rệt trong khoảng 20–25 tuổi. Điều này cho thấy nhóm khách hàng trẻ chiếm tỷ trọng lớn về mặt số lượng cá thể tham gia mua sắm, dù mức độ tiêu thụ trên mỗi khách hàng không cao bằng nhóm trưởng thành.

Sự đối lập giữa hai góc nhìn này mang lại một insight quan trọng: trong khi nhóm *Young* tạo nên độ phủ thương hiệu và lưu lượng khách hàng lớn, thì nhóm *Adult* mới là nhóm đóng góp chủ yếu vào sản lượng bán ra. Đây là cơ sở để H&M xây dựng chiến lược cân bằng giữa nội dung marketing mang tính trẻ trung, định hướng thương hiệu, và danh mục sản phẩm tập trung vào các nhu cầu thực tế của nhóm khách hàng trưởng thành.

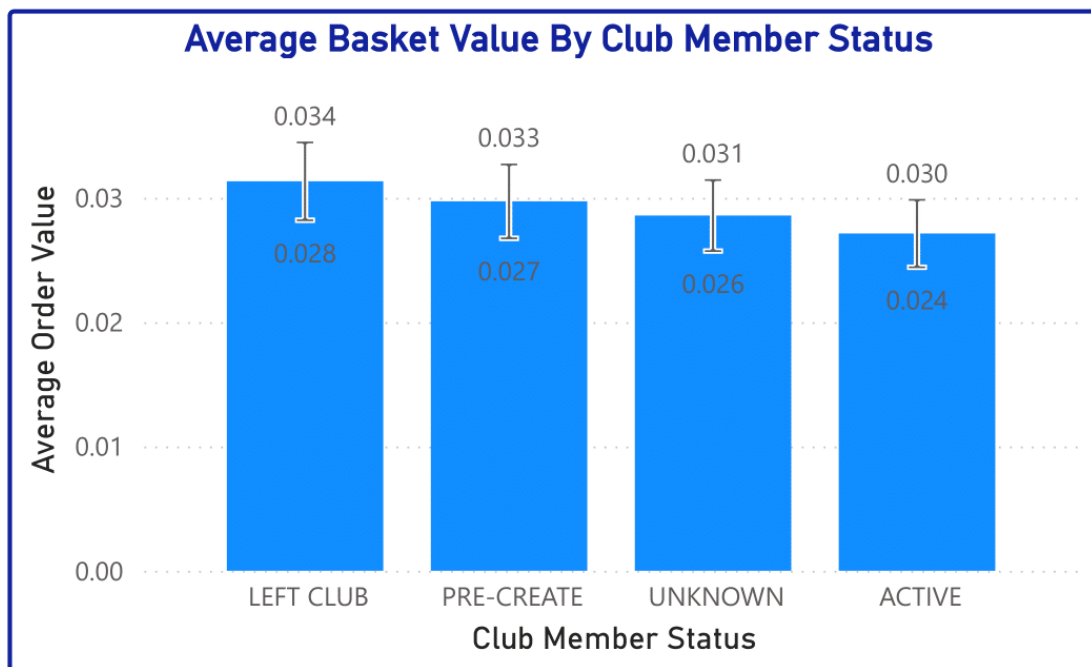
### 3.2.3. Phân tích tương tác marketing và hành vi giá.

Khu vực phân tích này tập trung đánh giá mức độ tương tác truyền thông của khách hàng cũng như hành vi phản ứng với giá bán, từ đó làm rõ hiệu quả của các kênh marketing và chiến lược định giá của H&M.



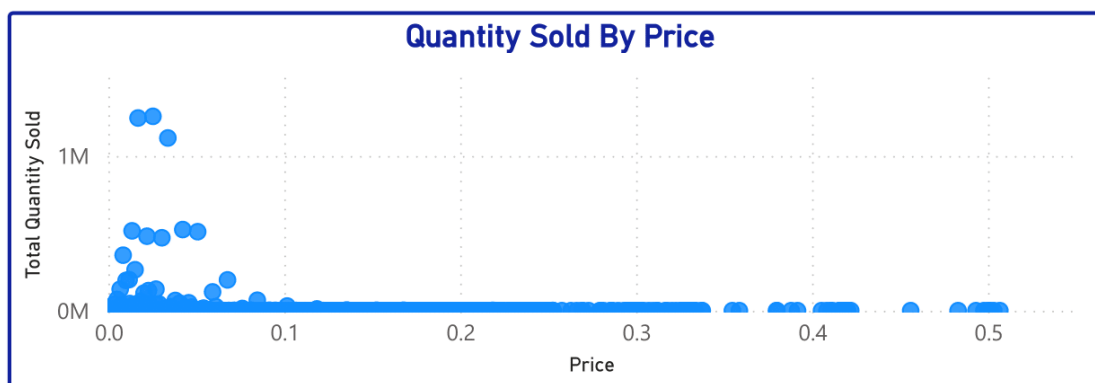
**Hình 12:** Fashion News Frequency

Xét về hiệu quả truyền thông, biểu đồ Donut thể hiện tần suất nhận tin thời trang cho thấy một khoảng trống lớn trong hoạt động kết nối khách hàng. Cụ thể, có tới 62% khách hàng không đăng ký nhận bất kỳ thông tin thời trang nào, trong khi chỉ 38% duy trì kết nối thường xuyên. Kết quả này phản ánh dư địa rất lớn cho các chiến dịch remarketing, đặc biệt là các chương trình email marketing hoặc cá nhân hóa nội dung nhằm chuyển đổi nhóm khách hàng “thụ động” thành nhóm có mức độ tương tác và trung thành cao hơn.



**Hình 13:** Average Basket Value By Club Member Status

Ở góc độ lòng trung thành và giá trị đơn hàng, biểu đồ *Average Basket Value by Member Status* cho thấy một kết quả đáng chú ý: nhóm *Active* lại có giá trị giỏ hàng trung bình thấp nhất, xấp xỉ 0.027, trong khi nhóm đã rời câu lạc bộ đạt mức cao hơn, khoảng 0.034. Điều này cho thấy các thành viên hoạt động có xu hướng mua sắm thường xuyên với giá trị mỗi đơn hàng nhỏ, phản ánh hành vi mua nhiều lần, mỗi lần một ít, trong khi các khách hàng ít gắn bó hơn lại có xu hướng dồn nhu cầu mua sắm vào những đơn hàng lớn hơn nhưng ít lần hơn.



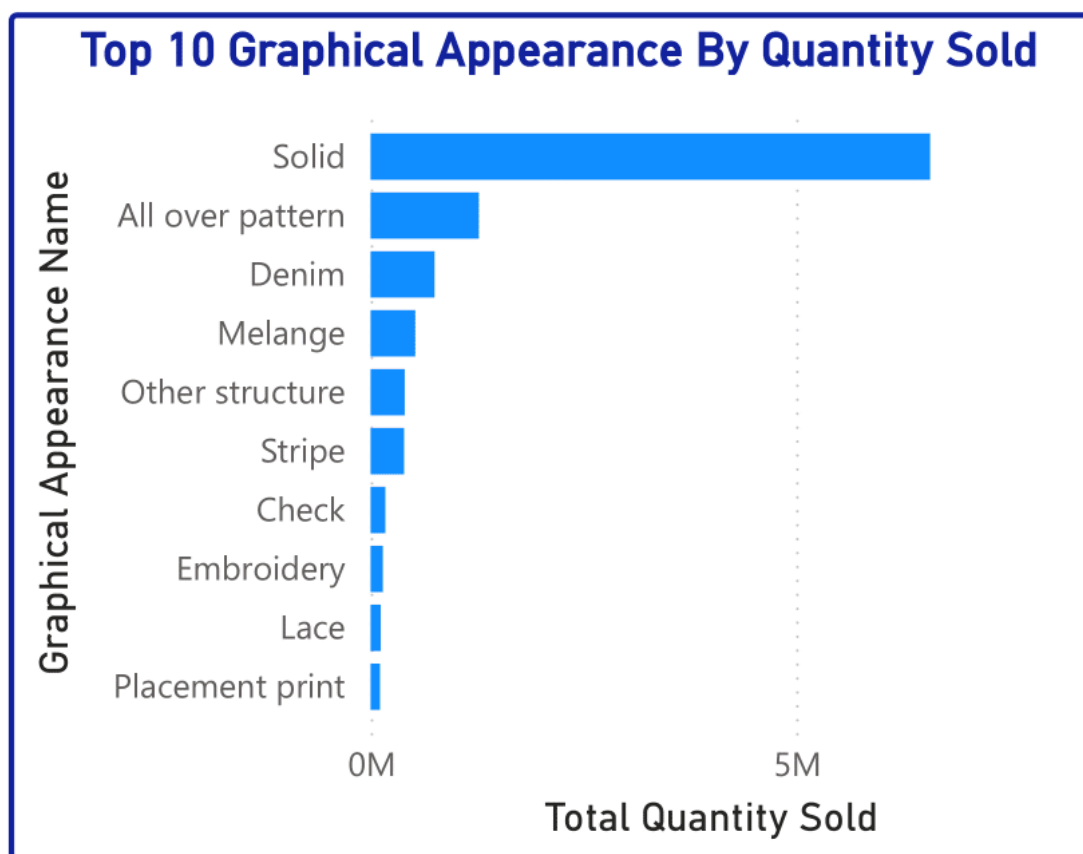
**Hình 14:** Quantity Sold By Price

Cuối cùng, hành vi nhạy cảm với giá được thể hiện rõ qua biểu đồ phân tán *Quantity Sold by Price*. Các điểm dữ liệu tập trung dày đặc trong khoảng giá từ 0.0 đến 0.05, cho thấy phần lớn sản lượng tiêu thụ đến từ các sản phẩm có mức giá rất thấp. Điều này khẳng định mức độ nhạy cảm cao về giá của khách hàng H&M, đồng

thời cũng cố lợi thế cạnh tranh cốt lõi của thương hiệu dựa trên chiến lược “giá rẻ nhưng vẫn hợp xu hướng”, đặc biệt trong bối cảnh thị trường thời trang nhanh có tính cạnh tranh gay gắt.

#### 3.2.4. Phân tích thị hiếu thẩm mỹ và xu hướng tiêu dùng.

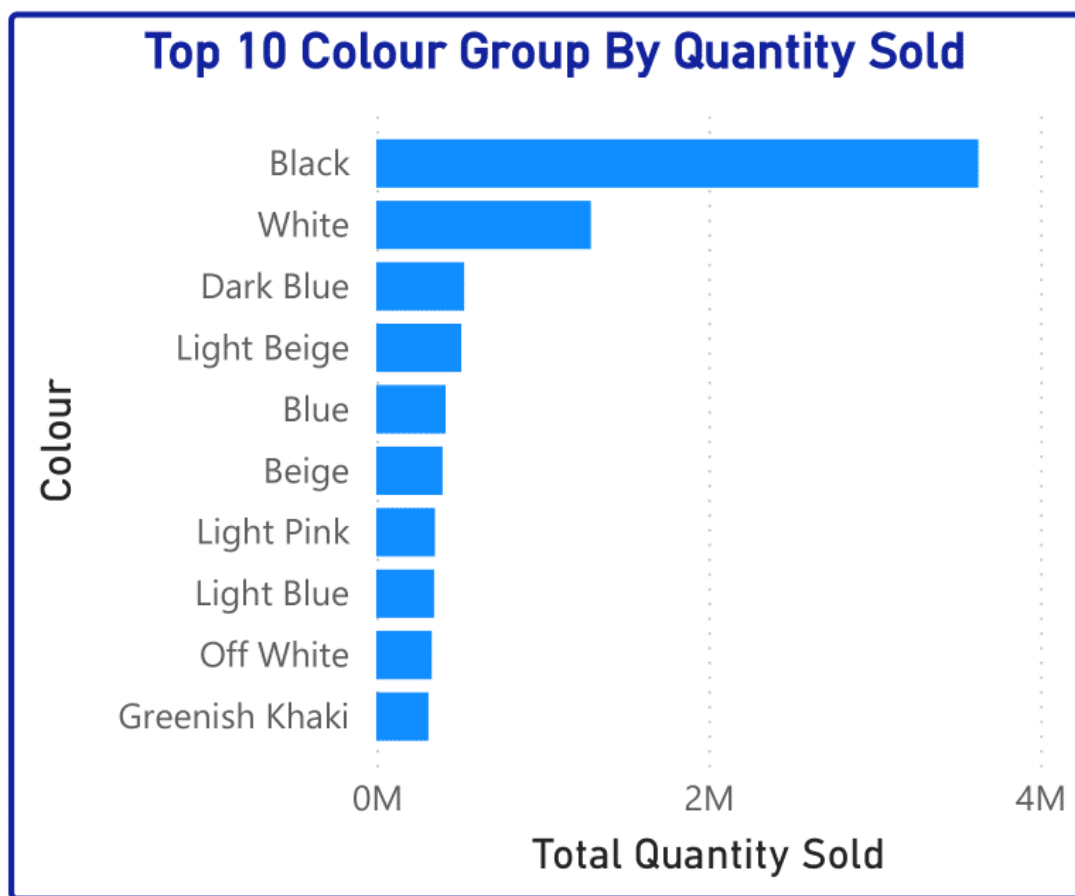
Khu vực phân tích này cung cấp các insight trực tiếp phục vụ cho bộ phận thiết kế và lập kế hoạch sản xuất, thông qua việc làm rõ những yếu tố thẩm mỹ chi phối hành vi lựa chọn sản phẩm của khách hàng H&M.



Hình 15: Top 10 Graphical Appearance By Quantity Sold

Xét trên khía cạnh họa tiết (*Graphical Appearance*), nhóm sản phẩm mang họa tiết *Solid* chiếm ưu thế tuyệt đối với hơn 6 triệu sản phẩm được bán ra. Kết quả này cho thấy xu hướng tối giản vẫn giữ vai trò là giá trị cốt lõi và bền vững trong tiêu dùng thời trang. Các thiết kế đơn sắc không chỉ dễ dàng phối hợp trong nhiều hoàn cảnh sử dụng mà còn giúp giảm thiểu rủi ro lỗi mốt, qua đó duy trì sức mua ổn định theo thời gian.





**Hình 16:** Quantity Sold By Price

Ở chiều cạnh màu sắc (*Colour Group*), các gam màu trung tính chiếm lĩnh hoàn toàn cơ cấu doanh số, nổi bật nhất là Black (đen), White (trắng), Dark Blue (xanh đậm) và Beige (màu be). Điều này phản ánh một nghịch lý thú vị của thị trường thời trang nhanh: mặc dù tốc độ ra mắt sản phẩm cao và mang tính xu hướng, khách hàng của H&M vẫn ưu tiên những lựa chọn an toàn, linh hoạt và có tính ứng dụng cao trong sinh hoạt hàng ngày. Xu hướng này củng cố vai trò của các tông màu cơ bản như một “trụ cột doanh thu”, đồng thời cho phép thương hiệu cân bằng giữa yếu tố thời trang và tính thực dụng trong chiến lược thiết kế.

### 3.2.5. Tính tương tác của Dashboard.

Sheet Customer & Behavior được thiết lập chế độ tương tác đa chiều, cho phép người dùng thực hiện các "truy vấn thị giác" chỉ bằng một cú click chuột. Ví dụ: Khi người dùng nhấn vào nhóm tuổi "Adult" trên biểu đồ Total Quantity Sold By Age Group, toàn bộ các biểu đồ bao gồm cả các biểu đồ về Hòa tiết và Màu sắc sẽ tự động lọc để hiển thị riêng thị hiếu của phân khúc này. Điều này có thể giúp nhà phân tích trả lời ngay lập tức câu hỏi: "Nhóm khách hàng trưởng thành đang ưu tiên màu sắc và họa tiết nào nhất?"

## 4. Tổng kết và Kiến nghị.

### 4.1. Tóm tắt các phát hiện quan trọng (Key Insights).

Tổng hợp từ các phân tích trên hai Dashboard, nghiên cứu rút ra một số phát hiện chiến lược cốt lõi phản ánh rõ nét mô hình vận hành, hành vi tiêu dùng và định vị thị trường của H&M trong năm 2020.

Về hiệu suất kinh doanh, H&M đang vận hành trên một quy mô rất lớn với hơn 862 nghìn khách hàng phát sinh giao dịch trong năm. Doanh thu thể hiện tính chu kỳ rõ rệt, tăng trưởng ổn định từ đầu năm, đạt đỉnh vào tháng 6 (cao điểm mùa hè) và bắt đầu suy giảm vào tháng 9. Cấu trúc doanh thu cho thấy phân khúc giá phổ thông cùng với nhóm hàng Ladieswear — chiếm tới 47.5% tổng doanh thu — đóng vai trò là hai trụ cột chiến lược, duy trì dòng tiền và sự ổn định cho toàn bộ hệ thống kinh doanh.

Xét về hành vi khách hàng, dữ liệu cho thấy sự phân hóa rõ ràng giữa quy mô số lượng và giá trị đóng góp. Mặc dù nhóm khách hàng trẻ (độ tuổi 20–25) chiếm tỷ trọng lớn nhất về mặt số lượng cá thể, nhóm khách hàng trưởng thành mới là đối tượng tiêu thụ nhiều sản phẩm nhất. Điều này phản ánh vai trò trung tâm của nhóm khách hàng có thu nhập ổn định trong việc thúc đẩy doanh số thực tế, đồng thời gợi ý sự cần thiết phải phân biệt giữa chiến lược thu hút và chiến lược khai thác giá trị khách hàng.

Về phân tích kênh truyền thông chỉ ra một khoảng trống đáng kể trong chiến lược tiếp cận khách hàng trực tiếp. Tỷ lệ khách hàng không đăng ký nhận tin tức thời trang lên tới 62%, cho thấy dư địa rất lớn để H&M mở rộng các hoạt động CRM và Marketing cá nhân hóa. Việc khai thác hiệu quả nhóm khách hàng “thụ động” này có thể trở thành đòn bẩy quan trọng nhằm gia tăng mức độ gắn kết và giá trị vòng đời khách hàng trong tương lai.

Phân tích về lòng trung thành mang lại một phát hiện mang tính nghịch lý. Giá trị giỏ hàng trung bình (AOV) của nhóm thành viên đang hoạt động lại thấp nhất so với các nhóm còn lại, chỉ vào khoảng 0.027. Điều này cho thấy nhóm khách hàng trung thành của H&M có xu hướng mua sắm thường xuyên với quy mô nhỏ cho mỗi giao dịch, trong khi nhóm khách hàng ít gắn kết hơn lại có xu hướng tích lũy nhu cầu để thực hiện các đơn hàng lớn. Phát hiện này gợi mở rằng giá trị của nhóm khách hàng trung thành không nằm ở kích thước đơn hàng, mà ở tần suất mua và sự ổn định dài hạn.

Về độ nhạy cảm với giá, hành vi tiêu dùng của khách hàng H&M cho thấy mức độ nhạy cảm rất cao. Phần lớn sản lượng tiêu thụ tập trung dày đặc ở vùng giá cực thấp, từ 0.0 đến 0.05, và giảm mạnh khi giá tăng. Kết quả này tái khẳng định lợi thế cạnh tranh cốt lõi của H&M nằm ở chiến lược “thời trang giá rẻ”, trong đó khối lượng giao dịch lớn đóng vai trò bù đắp cho biên lợi nhuận thấp trên từng sản phẩm đơn lẻ.

Cuối cùng, về thị hiếu sản phẩm, khách hàng của H&M thể hiện xu hướng ưu tiên các lựa chọn mang tính an toàn và ứng dụng cao. Các dòng sản phẩm cơ bản với họa tiết trơn và bảng màu trung tính như đen, trắng và beige liên tục dẫn đầu về sản lượng bán ra. Kết quả này cho thấy dù hoạt động trong lĩnh vực thời trang nhanh, nhu cầu cốt lõi của khách hàng vẫn xoay quanh tính linh hoạt, dễ phối đồ và ít rủi ro lỗi mốt.

## 4.2. Đề xuất giải pháp dựa trên dữ liệu.

Dựa trên các phát hiện từ phân tích Dashboard, nghiên cứu đề xuất một số định hướng giải pháp nhằm tối ưu hóa hiệu quả kinh doanh, nâng cao mức độ tương tác của khách hàng và củng cố lợi thế cạnh tranh dài hạn của H&M.

Trước hết, về chiến lược Marketing theo phân khúc khách hàng, dữ liệu cho thấy sự khác biệt rõ rệt giữa các nhóm độ tuổi. Đối với nhóm khách hàng Trẻ, các chiến dịch nên tập trung vào việc xây dựng hình ảnh thương hiệu, nhấn mạnh yếu tố xu hướng và tính thời trang nhằm duy trì sự gắn kết của một tệp khách hàng đông đảo nhưng có mức chi tiêu chưa cao. Ngược lại, với nhóm khách hàng Trưởng thành, vốn là động lực chính của sản lượng tiêu thụ, doanh nghiệp nên ưu tiên các chương trình khuyến mãi theo gói hoặc cơ chế ưu đãi dành cho khách hàng thân thiết. Cách tiếp cận này giúp khai thác tối đa sức mua của nhóm khách hàng có thu nhập ổn định và nhu cầu tiêu dùng bền vững.

Về cải thiện mức độ tương tác và kết nối với khách hàng, tỷ lệ lớn khách hàng không đăng ký nhận tin tức thời trang cho thấy tiềm năng chưa được khai thác hiệu quả. Nhóm đề xuất triển khai các chiến dịch khuyến khích đăng ký nhận tin đi kèm ưu đãi tức thời nhằm chuyển đổi nhóm khách hàng thụ động thành nhóm có tương tác thường xuyên. Song song đó, việc tận dụng dữ liệu hành vi để cá nhân hóa nội dung truyền thông, chẳng hạn như thông báo các bộ sưu tập mới phù hợp với màu sắc hoặc họa tiết mà khách hàng từng mua, sẽ giúp tăng xác suất chuyển đổi và nâng cao trải nghiệm mua sắm.

Trong công tác quản trị danh mục sản phẩm và kho vận, kết quả phân tích cho thấy các mặt hàng cơ bản như quần dài và váy đơn sắc đóng vai trò trụ cột doanh thu nhờ tính ứng dụng cao và rủi ro lỗi mốt thấp. Do đó, doanh nghiệp cần ưu tiên duy trì mức tồn kho ổn định cho nhóm sản phẩm này nhằm đảm bảo dòng tiền liên tục. Bên cạnh đó, giai đoạn doanh thu cao điểm vào giữa năm, đặc biệt là tháng 6, có thể được tận dụng để triển khai các chương trình xả kho giữa mùa, giúp giải phóng hàng tồn từ đầu năm và tối ưu hóa chi phí lưu kho.

Cuối cùng, về chiến lược giá, hành vi tiêu dùng cho thấy khách hàng H&M có độ nhạy cảm cao với giá và tập trung mạnh vào phân khúc phổ thông. Vì vậy, nguồn lực nghiên cứu và phát triển sản phẩm nên được ưu tiên cho phân khúc *Regular*, nơi có khả năng tạo ra khối lượng giao dịch lớn và doanh thu ổn định. Việc mở rộng quá sâu sang các phân khúc giá cao cấp không chỉ tiềm ẩn rủi ro về tồn kho mà còn không phù hợp với đặc điểm chi tiêu của tệp khách hàng hiện tại.

### 4.3. Những hạn chế của dữ liệu và phạm vi phân tích

Mặc dù bộ dữ liệu H&M cung cấp quy mô lớn và mức độ chi tiết cao, quá trình xử lý và phân tích vẫn tồn tại một số hạn chế khách quan có thể ảnh hưởng đến độ chính xác tuyệt đối của các chỉ số được tính toán.

Thứ nhất, dữ liệu giao dịch được ghi nhận ở cấp độ sản phẩm (*item-level*) và không có trường định danh đơn hàng (*order\_id*). Việc thiếu vắng khóa định danh này khiến các chỉ số liên quan trực tiếp đến đơn hàng, chẳng hạn như tổng số đơn hàng hay số lượng sản phẩm trung bình trên mỗi đơn, không thể được tính toán một cách chính xác. Trong báo cáo này, các chỉ số trên chỉ mang tính xấp xỉ, được suy luận dựa trên sự kết hợp giữa *customer\_id* và ngày giao dịch, do đó cần được diễn giải một cách thận trọng.

Thứ hai, dữ liệu giá bán đã được chuẩn hóa và ẩn danh, dẫn đến việc mất đi ngữ cảnh tài chính tuyệt đối. Nhóm nghiên cứu không thể xác định đơn vị tiền tệ gốc cũng như quy đổi doanh thu về giá trị thị trường thực tế. Vì vậy, các phân tích liên quan đến doanh thu trong báo cáo chủ yếu mang ý nghĩa tương đối, phục vụ cho việc so sánh xu hướng, tỷ trọng đóng góp và hành vi chi tiêu giữa các phân khúc, thay vì phản ánh quy mô tài chính tuyệt đối của doanh nghiệp.

Cuối cùng, phạm vi thời gian của nghiên cứu chỉ giới hạn trong năm 2020, một giai đoạn đặc biệt chịu ảnh hưởng mạnh từ bối cảnh dịch bệnh toàn cầu. Do đó, các kết luận về tính mùa vụ, chu kỳ kinh doanh và hành vi tiêu dùng có thể mang tính đặc thù theo thời điểm. Để có cái nhìn toàn diện và dài hạn hơn, các phân tích trong tương lai nên được mở rộng sang nhiều năm liên tiếp nhằm kiểm chứng mức độ ổn định của các xu hướng được quan sát trong nghiên cứu này.