



Giảng viên: **Anh Lê Nhật Minh**
Mentor: **Anh Hoàng Minh Khiêm**

E-COMMERCE ANALYSIS



Nguyễn Thanh Phong
Nhữ Văn Tiến
Nguyễn Vũ Hoàng Nguyên



MỤC LỤC



I. BUSINESS DESCRIPTION



II. TỔNG QUAN DATASET



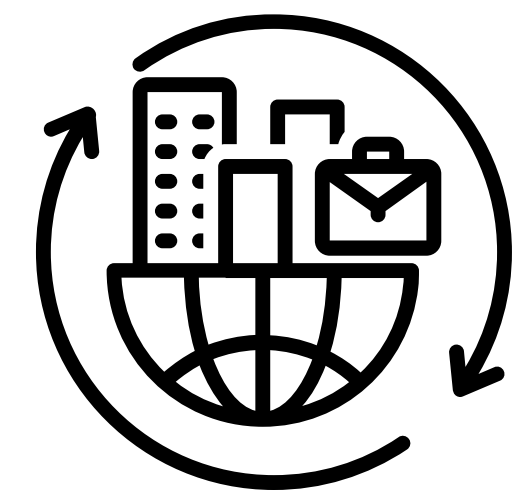
III. PHÂN TÍCH DỮ LIỆU



IV. KẾT LUẬN VÀ ĐỀ XUẤT

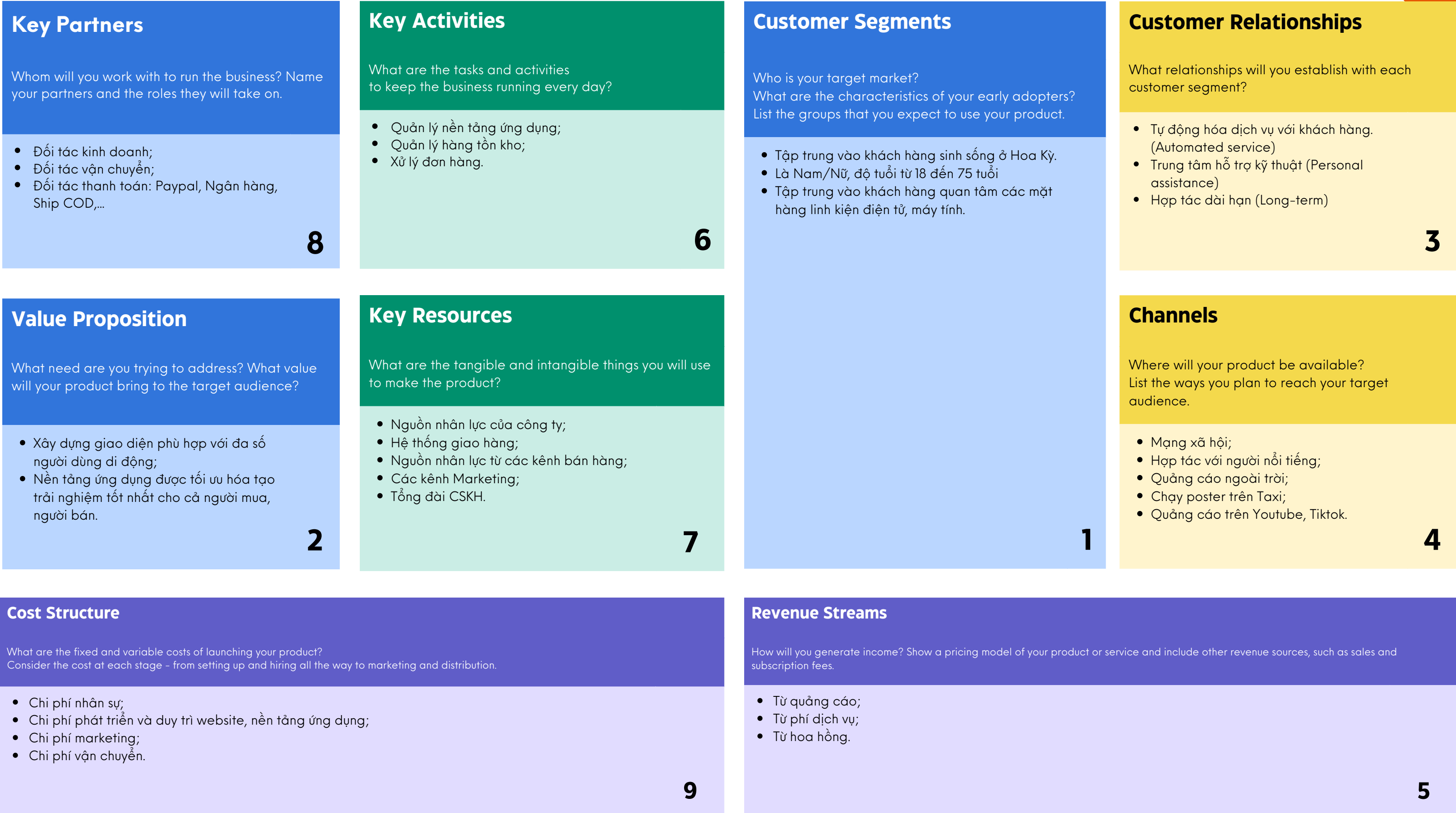


BUSINESS DESCRIPTION



I. Business Description

1. 1. Business Model Canvas



◆ I. Business Description

1. 2. Giới thiệu vấn đề



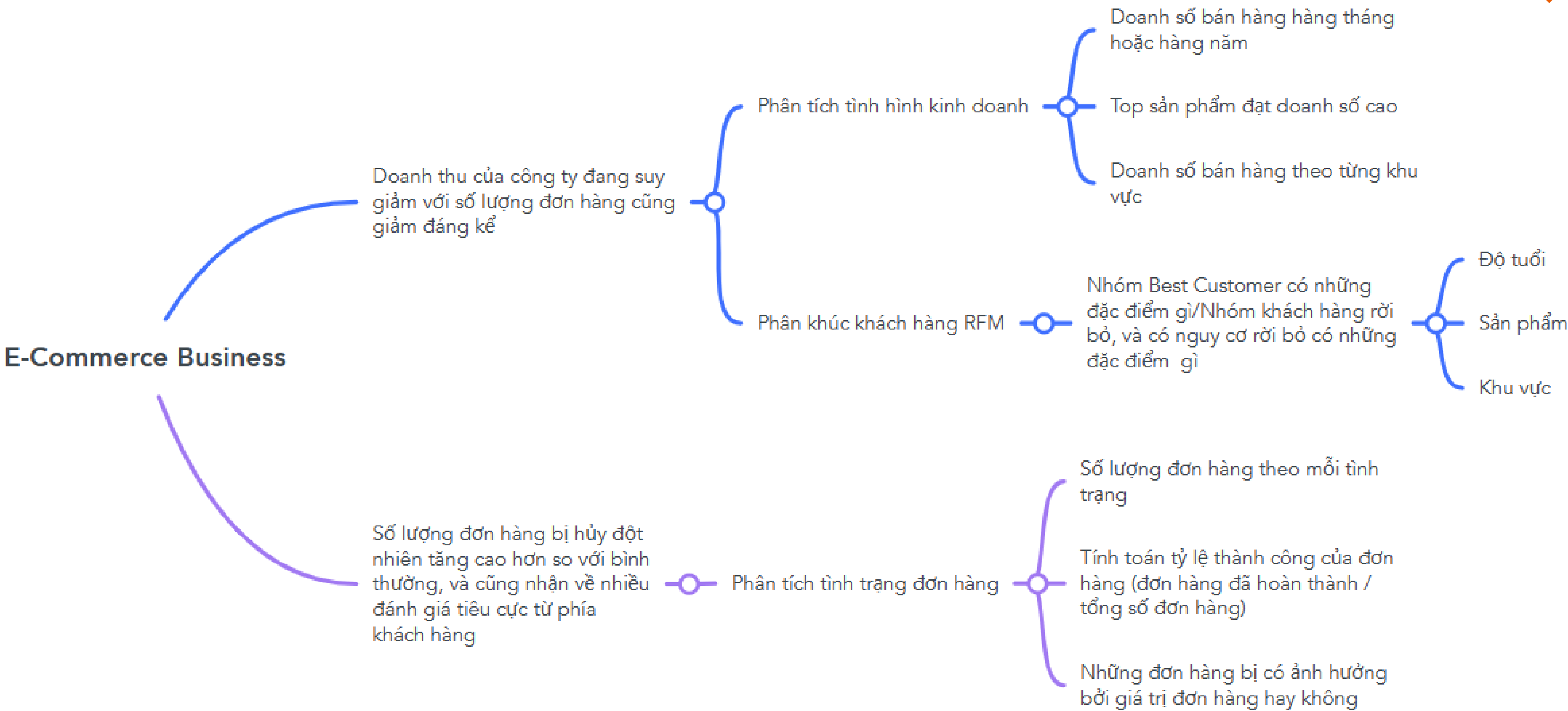
Theo báo cáo của Giám đốc, số lượng đơn hàng và doanh thu của công ty đã giảm trong **Quý 3/2021**.

Mặt khác, số lượng đơn hàng **bị hủy** cao hơn so với bình thường và nhiều **đánh giá tiêu cực** từ khách hàng về quá trình vận chuyển.



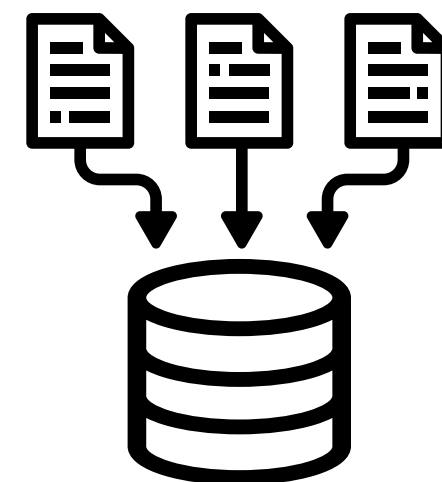
◆ I. Business Description

1. 3. Hướng xử lý





TỔNG QUAN DATASET



◆ II. Tổng quan Dataset

2. 1. Giới thiệu chung



Là bộ dữ liệu chứa các thông tin về bán hàng trực tuyến trên nền tảng ứng dụng di động tại Hoa Kỳ được thu thập từ tháng 10/2020 đến tháng 09/2021. Bộ dữ liệu gồm có **36 cột** và **286.393 dòng** nằm trên cùng 1 bảng.

- **order_id**: ID của đơn hàng
- **order_date**: Ngày đặt hàng
- **status**: Trạng thái đơn hàng
- **item_id**: ID của mặt hàng
- **qty_ordered**: Số lượng được đặt hàng
- **price**: Giá của mặt hàng
- **value**: Giá trị đơn hàng
- **discount_amount**: Số tiền giảm giá
- **total**: Tổng giá trị đơn hàng sau khi áp dụng giảm giá
- **category**: Danh mục của mặt hàng
- **payment_method**: Phương thức thanh toán

- **cust_id**: ID của khách hàng
- **gender**: Giới tính
- **age**: Tuổi
- **customer_since**: Thời điểm bắt đầu làm khách hàng
- **E Mail**: Địa chỉ email
- **county**: Quận
- **city**: Thành phố
- **state**: Bang
- **zip**: Mã Zip (mã bưu điện)
- **region**: Vùng miền

Source dataset: <https://www.kaggle.com/datasets/ytgangster/online-sales-in-usa> (Kaggle)

II. Tổng quan Dataset

2. 2. Các vấn đề của bộ dữ liệu

Bộ dữ liệu chưa được chuẩn hóa

0	order_id	286392	non-null	object
1	order_date	286392	non-null	object
2	status	286392	non-null	object
3	item_id	286392	non-null	float64
4	sku	286392	non-null	object
5	qty_ordered	286392	non-null	float64
6	price	286392	non-null	float64
7	value	286392	non-null	float64
8	discount_amount	286392	non-null	float64
9	total	286392	non-null	float64
10	category	286392	non-null	object
11	payment_method	286392	non-null	object
12	bi_st	286392	non-null	object
13	cust_id	286392	non-null	float64
14	year	286392	non-null	int64
15	month	286392	non-null	object
16	ref_num	286392	non-null	int64
17	Name Prefix	286392	non-null	object
18	First Name	286392	non-null	object
19	Middle Initial	286392	non-null	object
20	Last Name	286392	non-null	object
21	Gender	286392	non-null	object
22	age	286392	non-null	float64
23	full_name	286392	non-null	object
24	E Mail	286392	non-null	object
25	Customer Since	286392	non-null	object
26	SSN	286392	non-null	object
27	Phone No.	286392	non-null	object
28	Place Name	286392	non-null	object
29	County	286392	non-null	object
30	City	286392	non-null	object
31	State	286392	non-null	object
32	Zip	286392	non-null	int64
33	Region	286392	non-null	object
34	User Name	286392	non-null	object
35	Discount_Percent	286392	non-null	float64

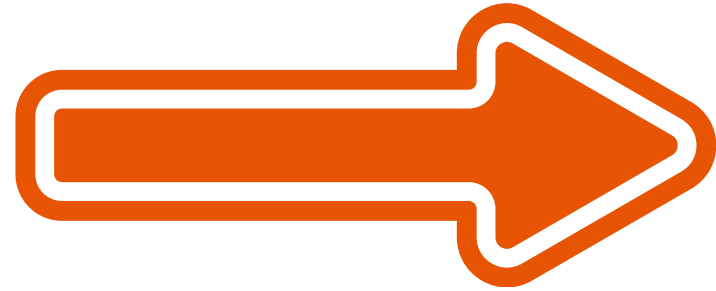
Kiểu dữ liệu nói về ngày tháng chưa đúng

Tên cột chưa được đồng nhất

◆ II. Tổng quan Dataset

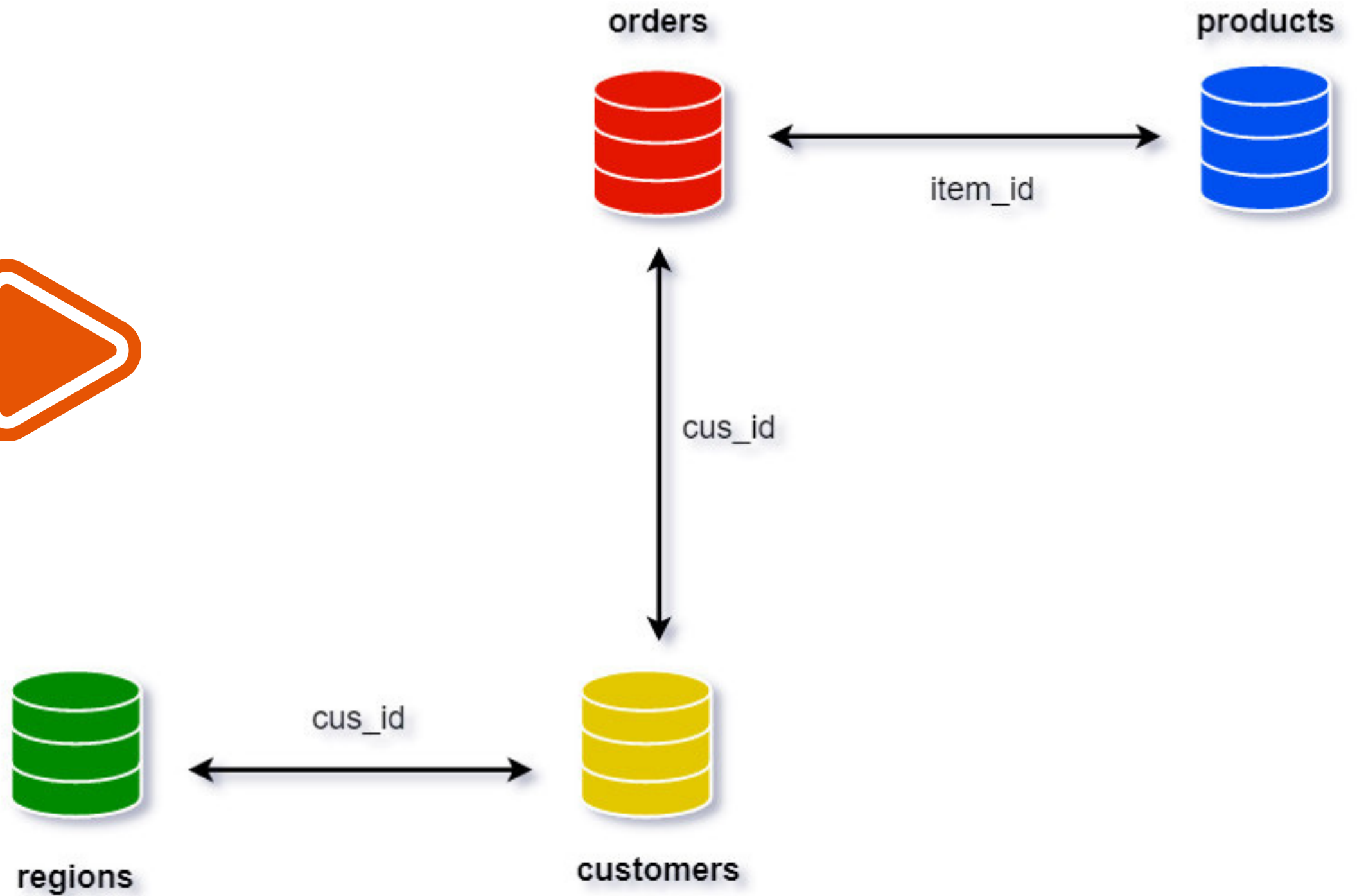
2. 3. Hướng xử lý dữ liệu

Tổ chức dữ liệu thành Schema



Lợi ích của việc chuẩn hóa:

- Tăng sự tổ chức
- Giảm khối lượng dữ liệu
- Tăng tính linh hoạt
- Tối ưu hóa hiệu suất



II. Tổng quan Dataset

2. 4. Các bước xử lý dữ liệu

Các bước xử lý dữ liệu

Kiểm tra và xóa null, duplicate (nếu có)

```
check_null = df.isnull().sum()
print(check_null[check_null>=1])

[3] ✓ 0.2s
... Series([], dtype: int64)

df.drop_duplicates(inplace=True)

[4] ✓ 0.8s
```



Đồng nhất tên các cột

VD: Last Name --> last_name

```
df.columns = df.columns.str.lower()
df.columns = df.columns.str.replace(' ', '_')
```

Đưa các cột ngày về datetime

```
df.order_date = pd.to_datetime(df['order_date'])
df.customer_since = pd.to_datetime(df['customer_since'])
```

	A
206137	100468498
206138	100468498
206139	100468515
206140	100468515
206141	100468499
206142	100468503
206143	100468507
206144	100468510
206145	100468512
206146	100468513
206147	100468513
206148	100468519
206149	100468520
206150	100468520-1
206151	100468521
206152	100468523
206153	100468525
206154	100468525
206155	100479317
206156	100468530

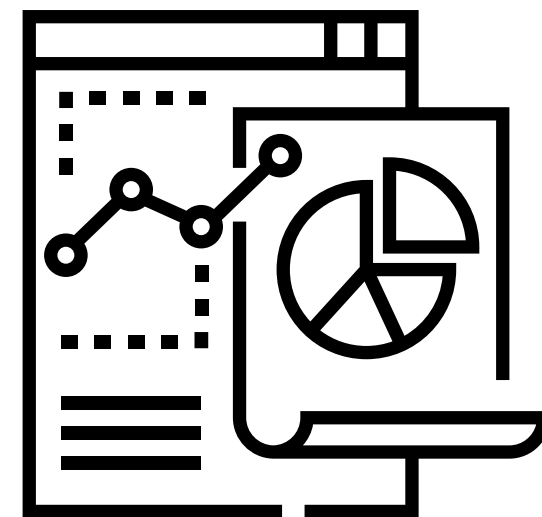
Xóa ký tự "-1" để đồng nhất các id

```
df['order_id'] = df['order_id'].replace('-1$', '', regex=True)

✓ 0.2s
```



PHÂN TÍCH DỮ LIỆU

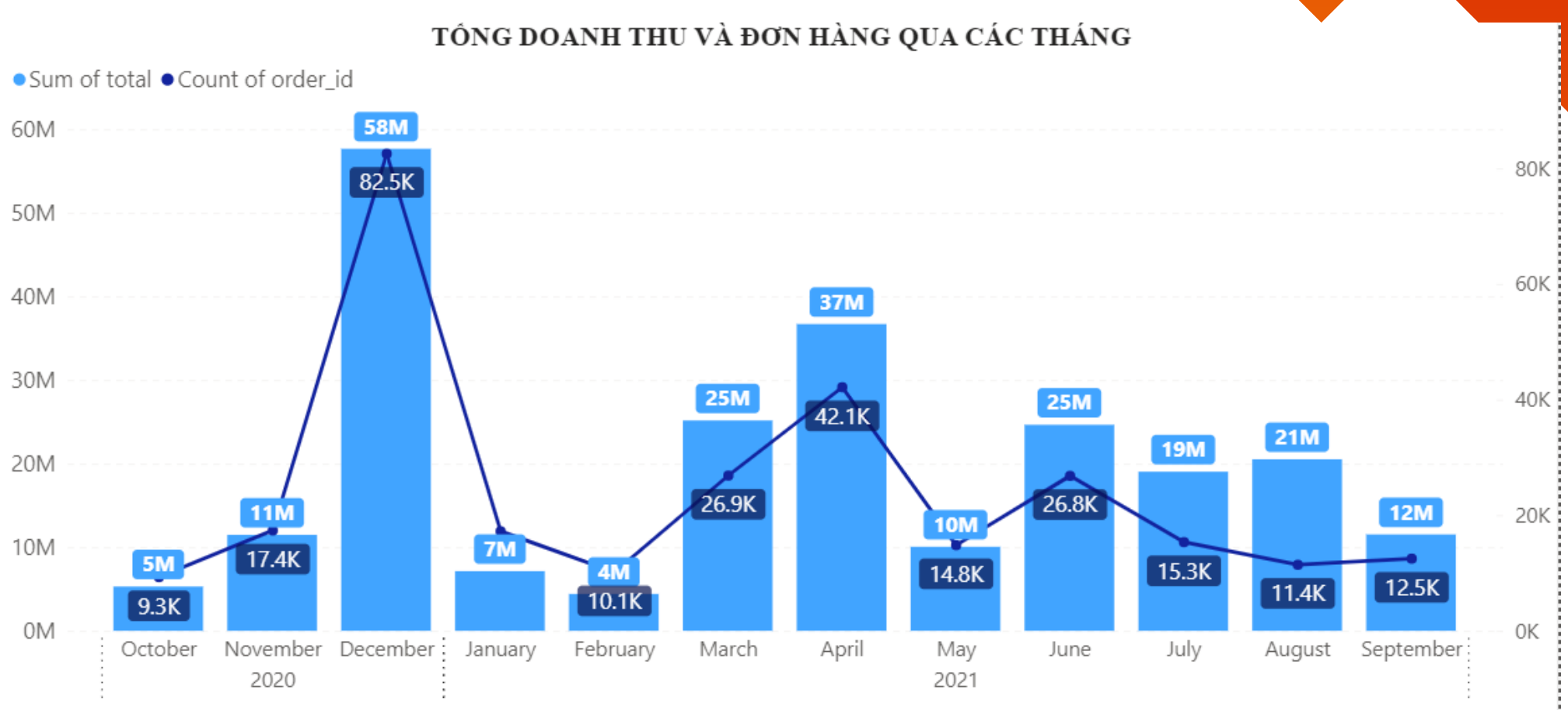
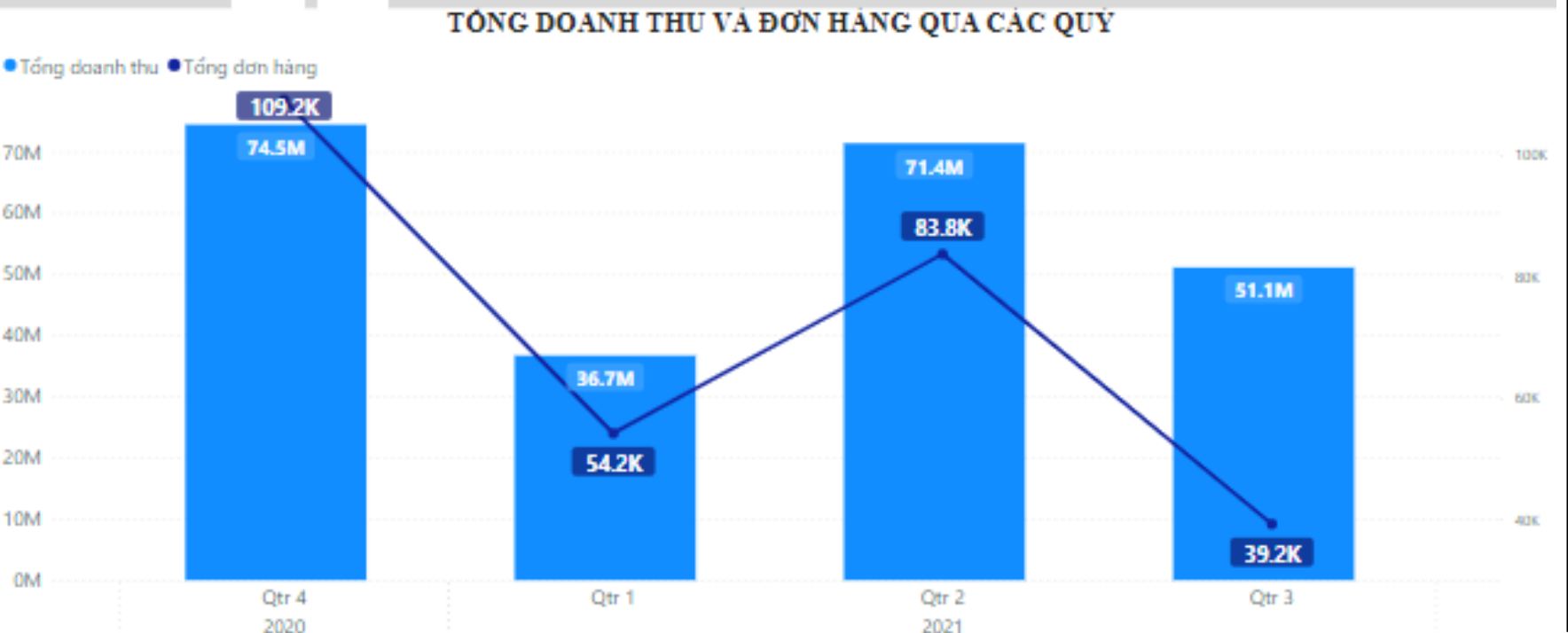


III. PHÂN TÍCH DỮ LIỆU

3. 1. Phân tích tình hình kinh doanh

3. 1. 1. Tình hình kinh doanh qua các năm

TỔNG DOANH THU	SỐ LƯỢNG ĐƠN HÀNG	TỔNG SỐ SẢN PHẨM	SỐ LƯỢNG KHÁCH HÀNG
233.65M	286.39K	15	64.25K



Công ty đã ghi nhận một sự gia tăng vượt bậc trong doanh thu vào Quý 4 năm 2020, khi toàn cầu bị cách ly bởi **đại dịch Covid-19**. Đặc biệt, trong các tháng 10 và 11 năm 2020, doanh thu ghi nhận không quá cao do các chính sách vận chuyển còn hạn chế bởi đại dịch.

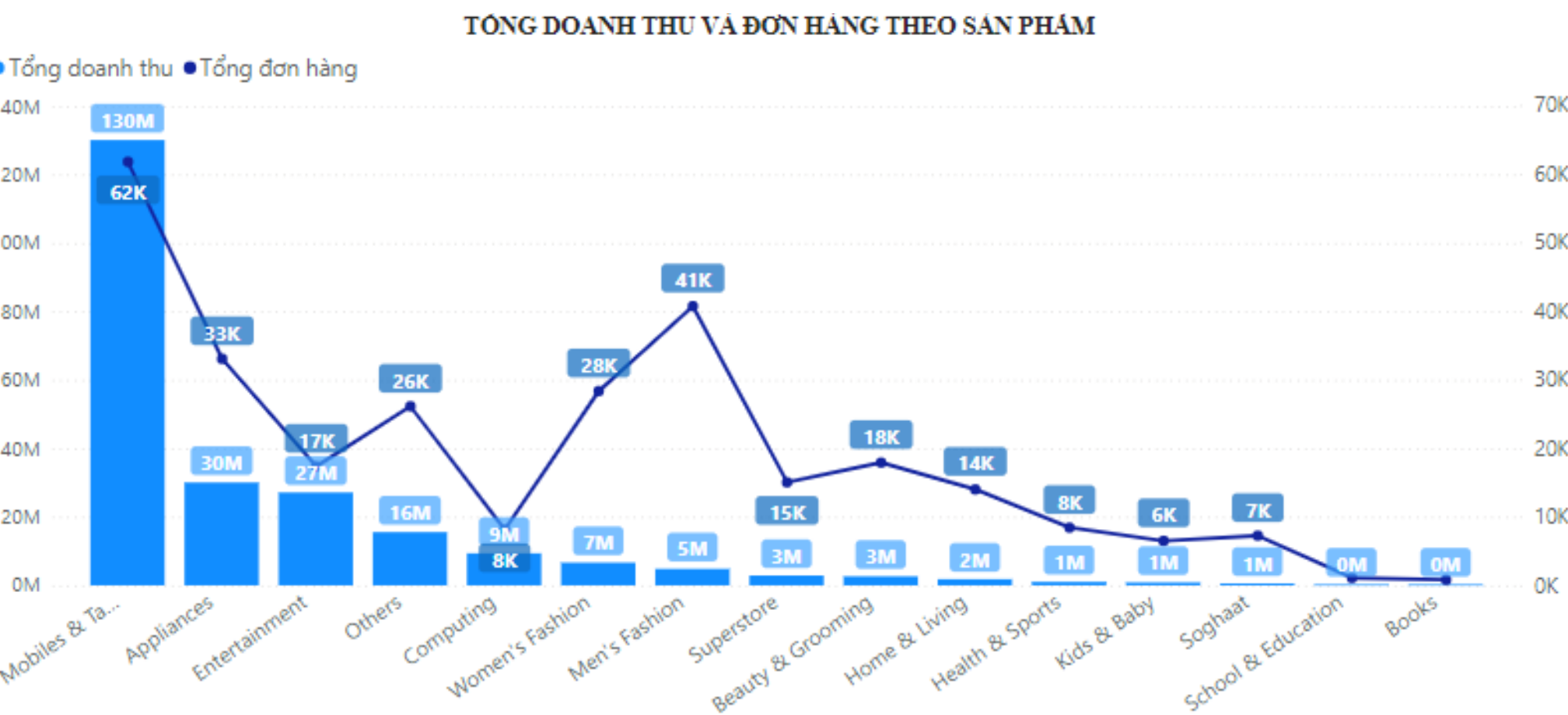
Tuy nhiên, sau khi các chính sách giao hàng an toàn được đưa ra, các chính sách vận chuyển đã được nới lỏng hơn. Năm bắt cơ hội này, công ty đã thực hiện nhiều chiến dịch bán hàng trực tuyến và giảm chi phí vận chuyển, điều này đã tạo ra doanh thu tháng 12/2020 đạt **28,4 triệu USD**, tăng gần **4 lần so với tháng trước**.

Sang năm 2021, doanh thu Quý 3 đã giảm xuống còn 54,1 triệu (**0,3 lần so với quý trước**) do nền kinh tế suy thoái nghiêm trọng bởi đại dịch Covid-19. Mặc dù doanh thu trong tháng 8 vẫn đạt mức tương đối, nhưng số lượng đơn hàng trong hai tháng cuối Quý 3 giảm đáng kể, chỉ còn trung bình khoảng **5,5 nghìn đơn hàng**.

III. PHÂN TÍCH DỮ LIỆU

3. 1. Phân tích tình hình kinh doanh

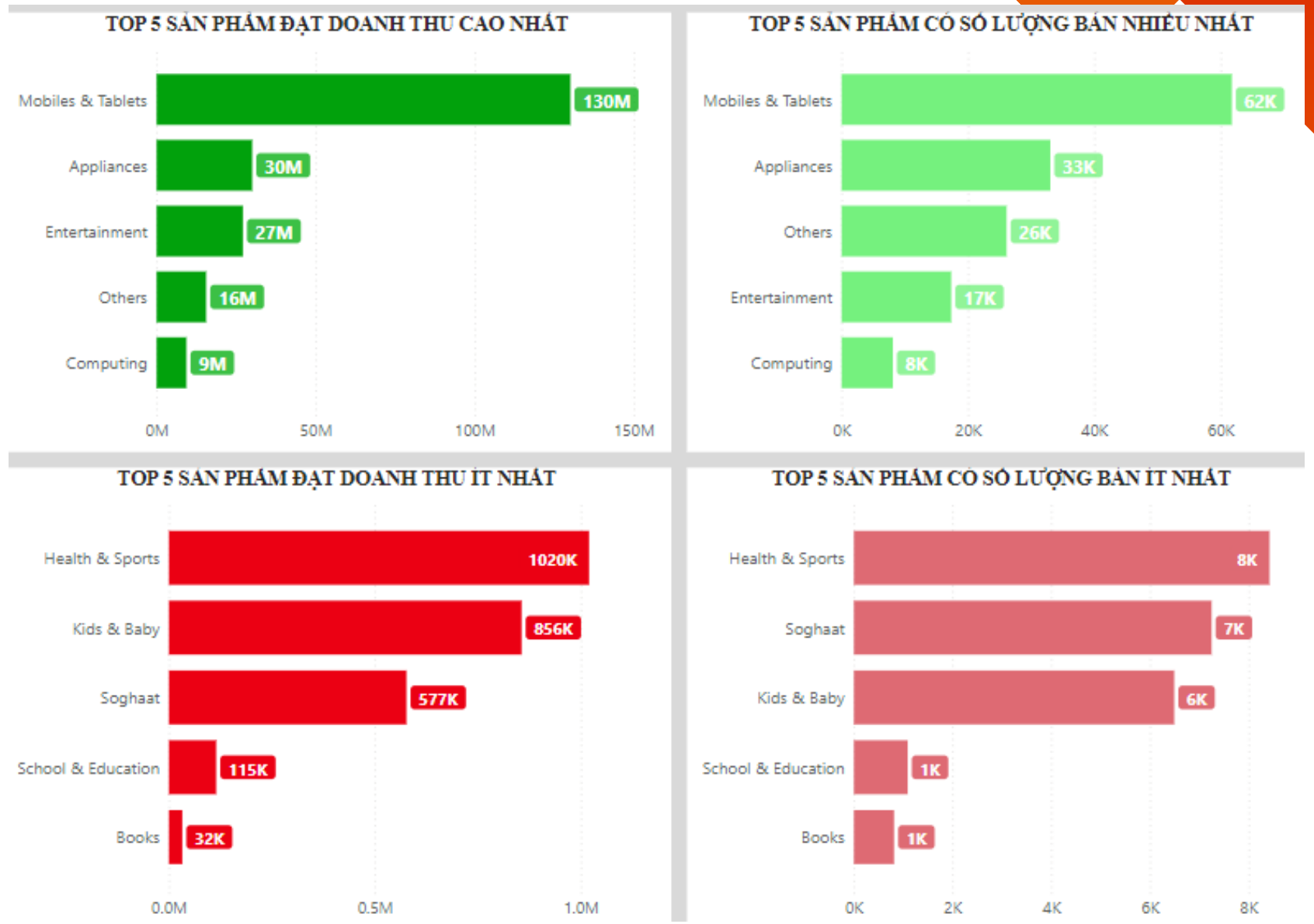
3. 1. 2. Tình hình kinh doanh ở các sản phẩm



Phần lớn các mặt hàng chính bán chạy nhất ở công ty chủ yếu là các sản phẩm liên quan đến **thiết bị và các đồ điện tử, máy tính, điện thoại**.

Mặc dù sản phẩm thời trang nam nữ không có doanh thu cao, nhưng số lượng đơn hàng của chúng vẫn cao hơn so với các sản phẩm có doanh thu cao khác.

! Các mặt hàng có doanh thu thấp thì số lượng bán ra cũng không đạt kỳ vọng như mong muốn.



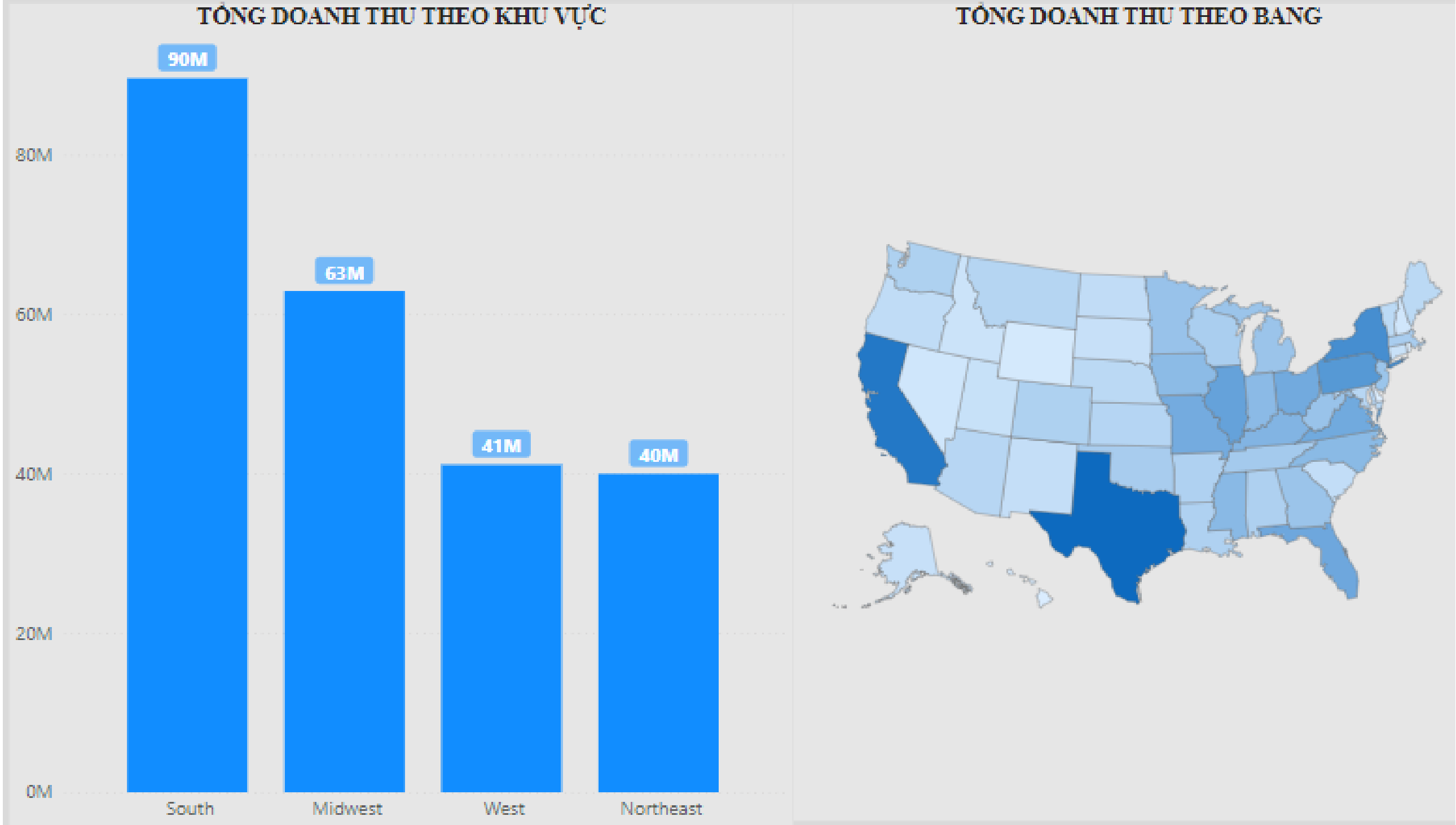
Mặt hàng Mobiles & Tablets đạt doanh số và số lượng đơn hàng cao nhất là **130 triệu và 62 nghìn đơn**.

Bốn mặt hàng còn lại tuy có doanh thu trong top 5 nhưng số lượng sản phẩm bán ra lại không quá nhiều.

III. PHÂN TÍCH DỮ LIỆU

3. 1. Phân tích tình hình kinh doanh

3. 1. 3. Tình hình kinh doanh ở các khu vực



Nam Hoa Kỳ và Trung Tây Hoa Kỳ là 2 khu vực có doanh thu cao nhất lần lượt là 90 triệu và 63 triệu

Texas, California và New York là 3 tiểu bang có doanh thu cao nhất tại Hoa Kỳ

III. PHÂN TÍCH DỮ LIỆU

3. 1. Phân tích tình hình kinh doanh

3. 1. 4. Đề xuất



Nghiên cứu và phân tích thị trường

- Đảm bảo rằng doanh nghiệp hiểu rõ nhu cầu và mong muốn của người tiêu dùng đối với các sản phẩm và dịch vụ cụ thể.
- Nghiên cứu cạnh tranh và phân tích thị trường để hiểu các yếu tố ảnh hưởng đến doanh thu.



Chạy các chương trình khuyến mãi và quảng cáo

- Mở rộng hoạt động quảng cáo và tiếp thị để nâng cao hiệu suất quảng cáo và thu hút nhiều người xem hơn.
- Các chiến dịch quảng cáo có thể được thực hiện thông qua các kênh truyền thông xã hội, quảng cáo trực tuyến hoặc quảng cáo truyền thống.



Tối ưu hóa giá cả và chiến lược giá

- Xem xét việc điều chỉnh giá cả để tăng doanh số bán hàng.
- Thiết lập giảm giá tạm thời, chiến lược giá cả linh hoạt, hoặc việc tạo ra các gói sản phẩm giá trị để thu hút khách hàng.



Nâng cao chất lượng dịch vụ

- Tăng cường dịch vụ và hỗ trợ khách hàng để tạo ra trải nghiệm khách hàng tích cực.
- Thông qua sự chia sẻ tích cực từ phía khách hàng hiện tại, việc này có thể giúp tăng cường lòng trung thành của khách hàng và tạo ra nguồn cung khách hàng tiềm năng mới.

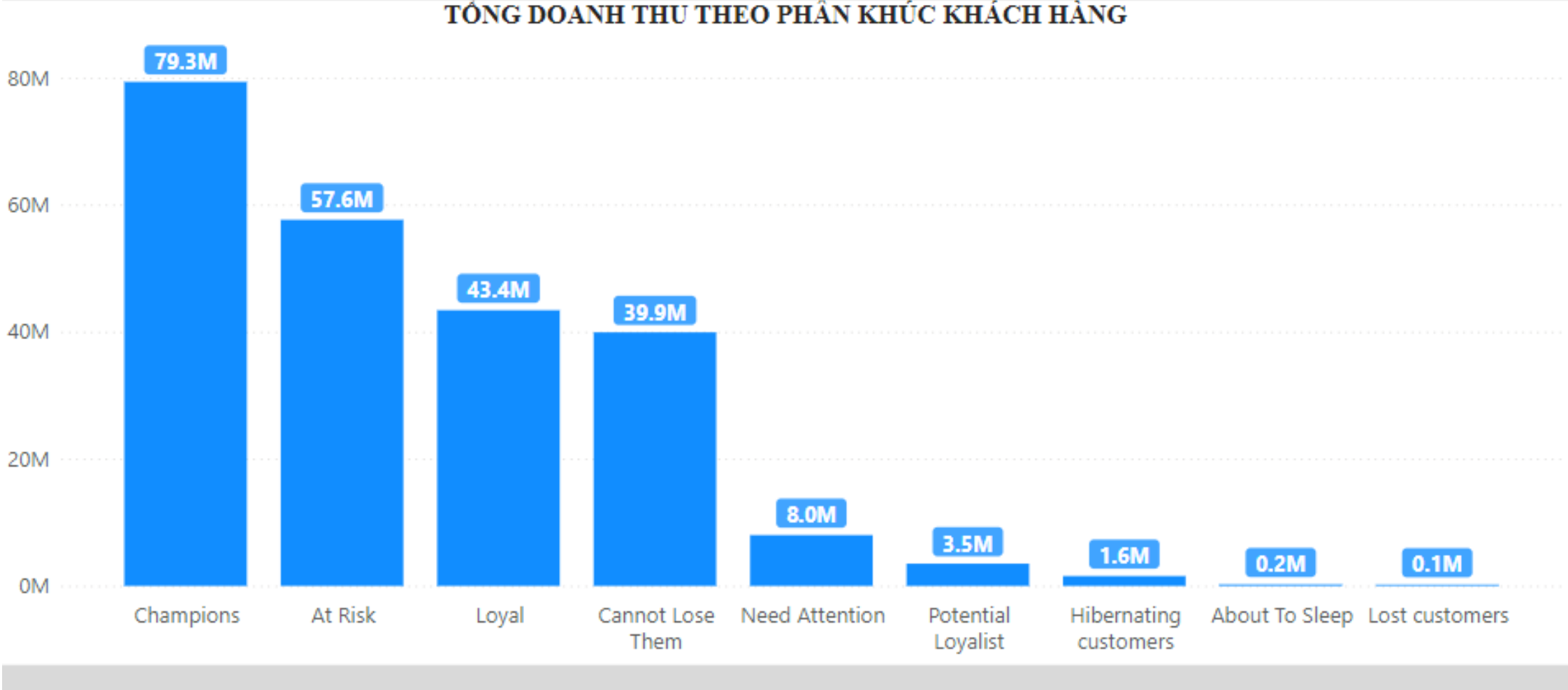
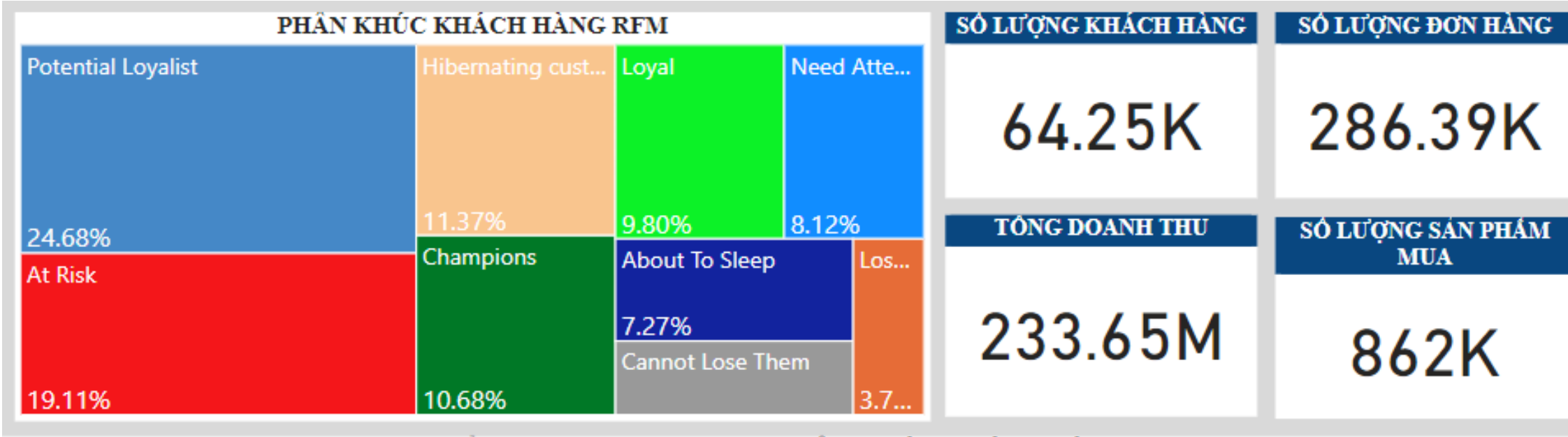


Tích hợp công nghệ: Sử dụng công nghệ để cải thiện trải nghiệm khách hàng và quá trình kinh doanh.

III. PHÂN TÍCH DỮ LIỆU

3. 2. Phân tích phân khúc khách hàng RFM

3. 2. 1. Nhóm khách hàng hiện tại



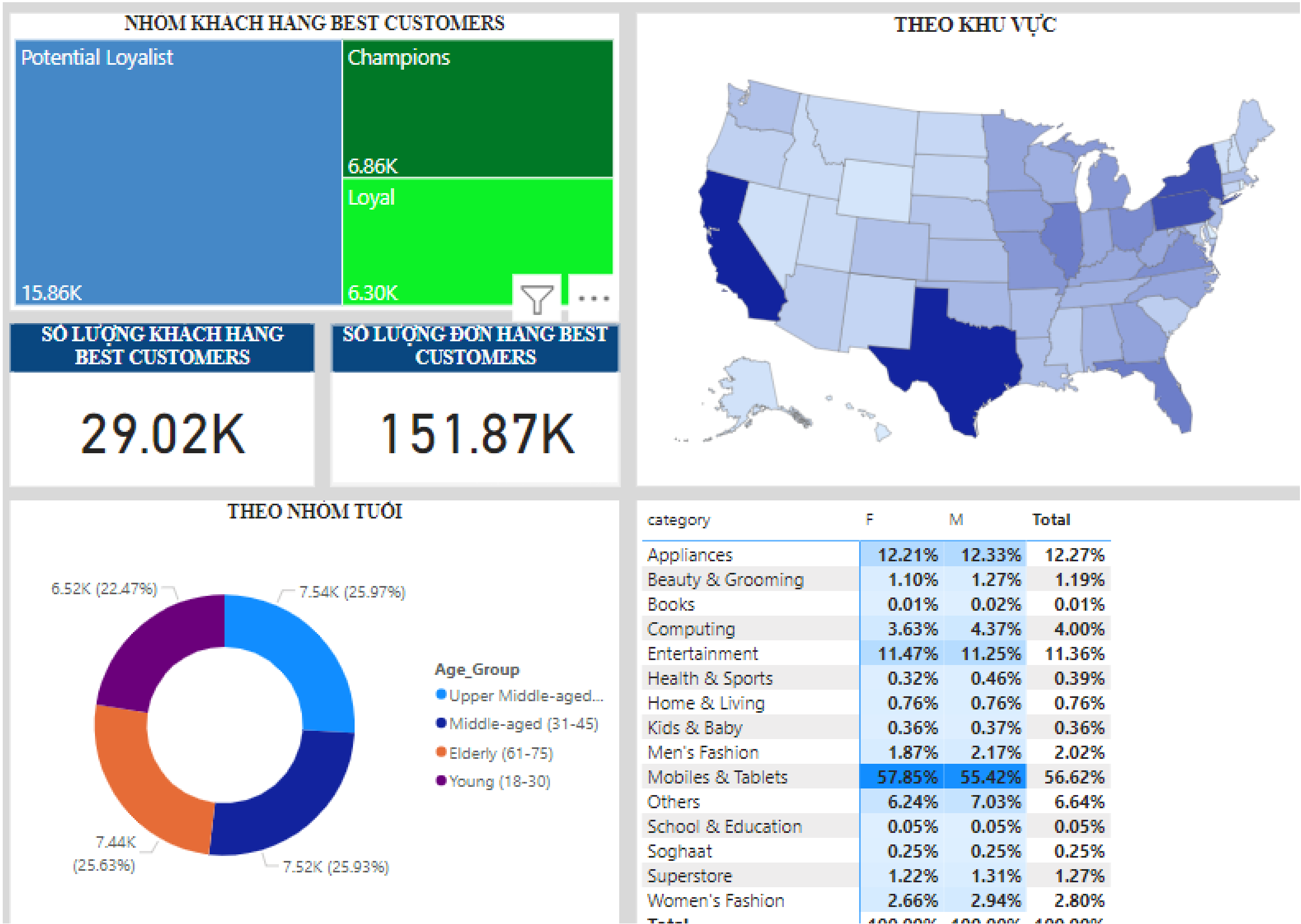
- Nhóm khách hàng **Potential Loyalist** chiếm tỷ trọng **cao nhất** (~25% tổng số).
 - Nhưng trái với kỳ vọng khi kinh doanh thì nhóm khách hàng **Champions** chiếm tỷ trọng không quá cao, chỉ xếp thứ 4 (~10.7%)
- => Điều này có thể cho thấy rằng công ty đang **thiếu hiệu quả** trong việc chuyển đổi khách hàng tiềm năng.

Bổ sung thêm cho luận điểm trên, nhóm khách hàng **At Risk** và nhóm khách hàng **Hibernating Customers** chiếm tỷ trọng lớn thứ 2 và thứ 3 với tỷ trọng lần lượt là **19.1%** và **11,37%**.

III. PHÂN TÍCH DỮ LIỆU

3. 2. Phân tích phân khúc khách hàng RFM

3. 2. 2. Phát triển từ nhóm khách hàng tiềm năng



III. PHÂN TÍCH DỮ LIỆU

3. 2. Phân tích phân khúc khách hàng RFM

3. 2. 2. Phát triển từ nhóm khách hàng tiềm năng

ĐỀ XUẤT

Duy trì doanh thu từ nhóm Champions



Chăm sóc khách hàng tốt hơn



Cung cấp sản phẩm chất lượng hơn



Tạo giá trị cho khách hàng



Tích cực tương tác với khách hàng

Chuyển khách hàng tiềm năng thành khách hàng trung thành (Loyal)



Lập kế hoạch cho nhóm khách hàng tiềm năng

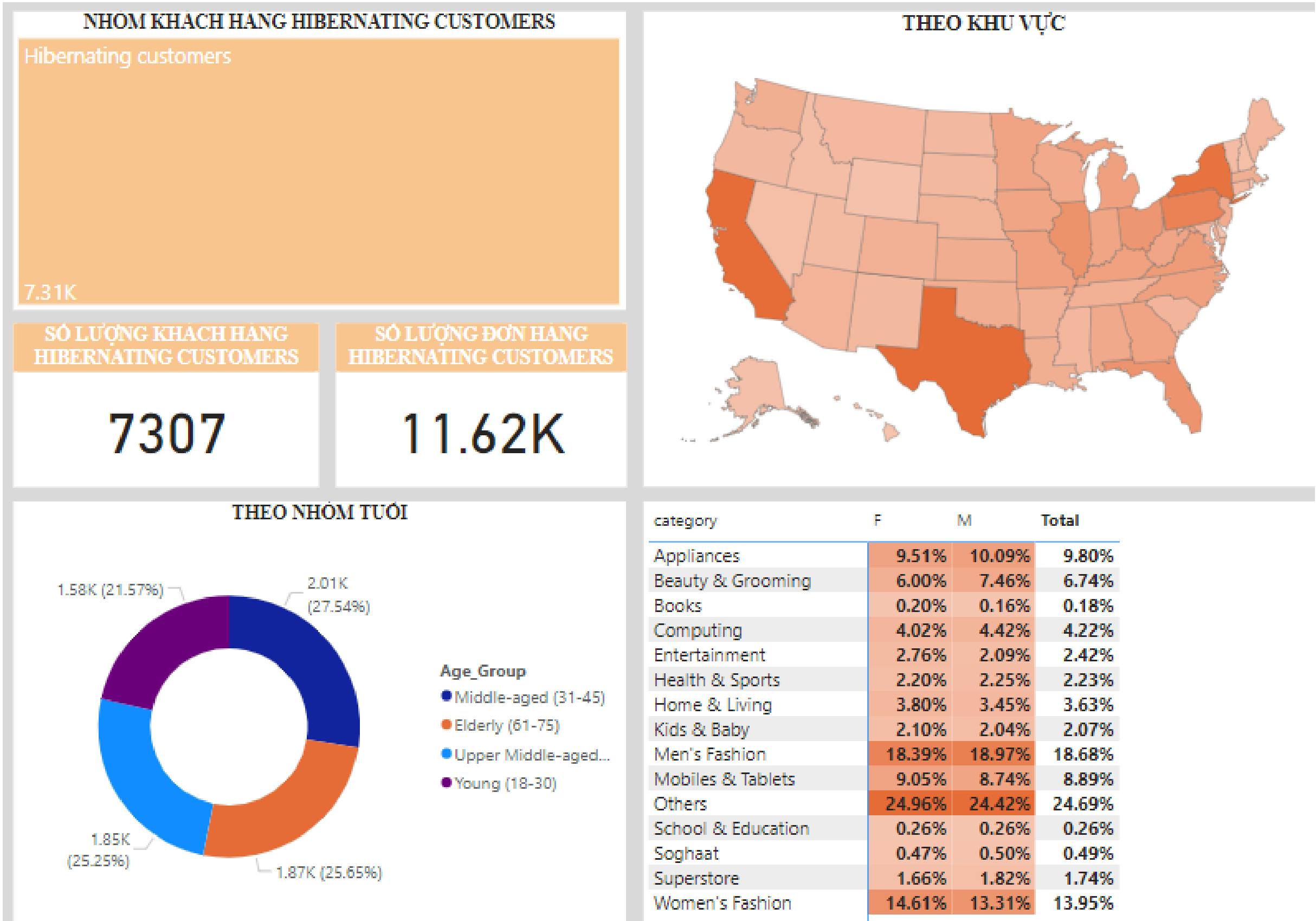


Đảm bảo theo sát kế hoạch đã được đề ra

III. PHÂN TÍCH DỮ LIỆU

3. 2. Phân tích phân khúc khách hàng RFM

3. 2. 2. Phát triển từ nhóm khách hàng tiềm năng



HIBERNATING CUSTOMERS

Phần lớn từ Texas, California và New York

Chủ yếu là Men and Women's Fashion.

Theo đánh giá của các chuyên gia, việc tìm kiếm một khách hàng mới thường tốn từ **5 đến 25 lần** chi phí so với việc giữ chân một khách hàng hiện có
=> Việc giữ chân nhóm khách hàng này là rất quan trọng để tạo niềm tin, lòng trung thành cũng tăng doanh thu cho công ty.

◆ III. PHÂN TÍCH DỮ LIỆU

3. 2. Phân tích phân khúc khách hàng RFM

3. 2. 2. Phát triển từ nhóm khách hàng tiềm năng

ĐỀ XUẤT



- Tìm ra lý do tại sao những người này không tiếp tục mua hàng
- Tìm hiểu tại sao họ thường mua hàng và có mối quan hệ tích cực với công ty
- Cải thiện dịch vụ chăm sóc khách hàng

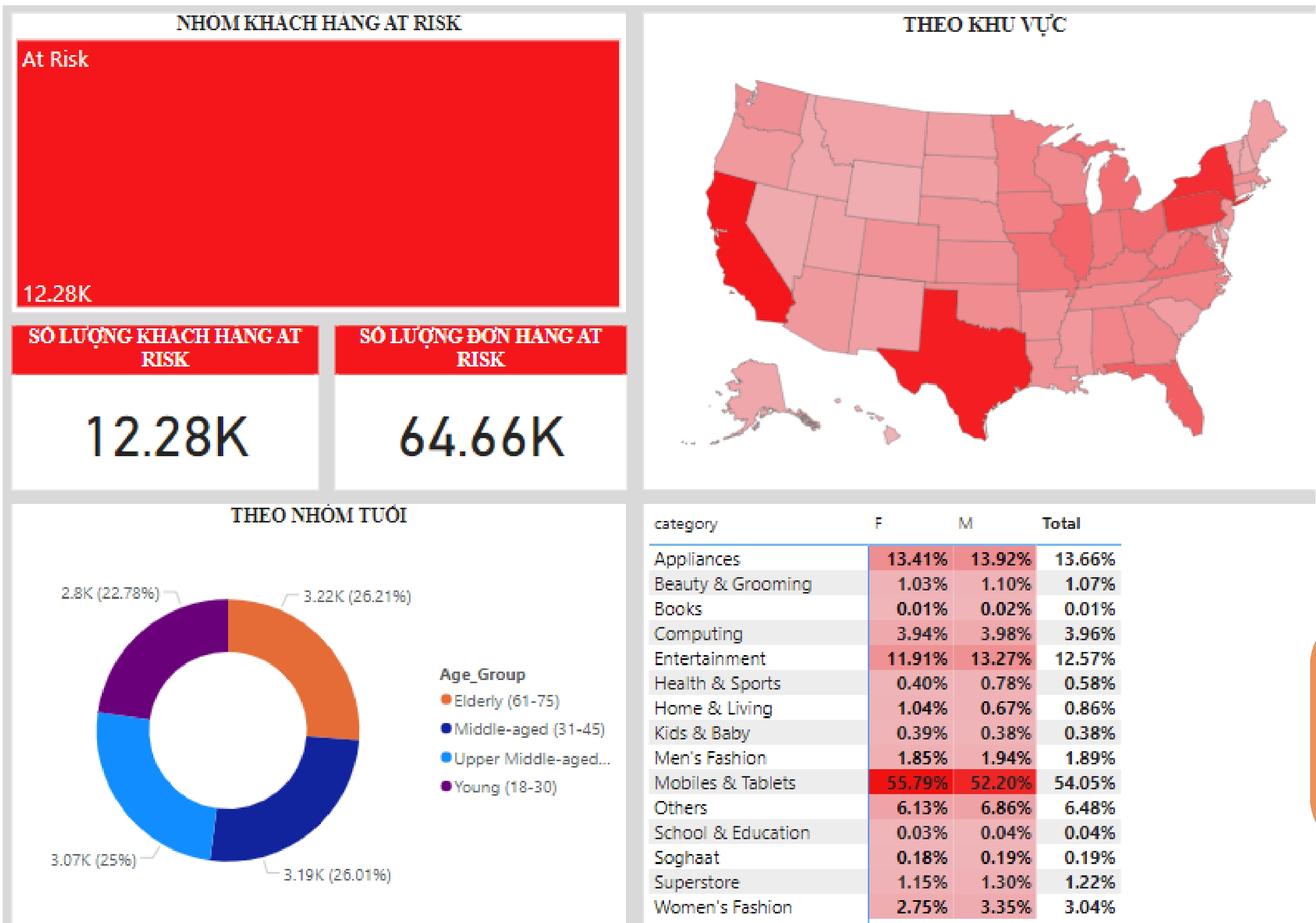


- Thường mua các mặt hàng liên quan đến quần áo thời trang
- Đẩy mạnh các chiến lược giảm giá
- Phát triển nhiều sản phẩm mới
- Tạo nhiều cơ hội hơn cho khách hàng tiếp cận các mặt hàng còn lại.

III. PHÂN TÍCH DỮ LIỆU

3. 2. Phân tích phân khúc khách hàng RFM

3. 2. 3. Nhóm khách hàng có nguy cơ rời bỏ At Risk



AT RISK

AGE+ Phần lớn là nhóm Người già (khoảng 26%)

Mobiles & Tablets Chủ yếu là Mobiles & Tablets

Theo lịch sử mua hàng của nhóm này, họ có tần suất mua hàng **thấp**, giá trị hàng hòa **thấp** và đã mua **khá lâu**.
=> Khả năng khách hàng rời bỏ công ty rất cao nếu không có các chiến dịch thu hút họ quay trở lại.

III. PHÂN TÍCH DỮ LIỆU

3. 2. Phân tích phân khúc khách hàng RFM

3. 2. 3. Nhóm khách hàng có nguy cơ rời bỏ At Risk

ĐỀ XUẤT



Phân tích và xem xét hành vi

Sử dụng dữ liệu mua hàng của khách hàng để phân tích và đánh giá hành vi



Tùy chỉnh giao tiếp

Giao tiếp một cách cá nhân hóa với khách hàng để hiểu và giải quyết các vấn đề mà họ đang gặp phải.



Cải thiện dịch vụ

Tăng cường chất lượng dịch vụ và hỗ trợ khách hàng để giải quyết các vấn đề và nhu cầu của họ một cách nhanh chóng và hiệu quả.



Chương trình khuyến mãi

Tạo ra các chương trình khuyến mãi và ưu đãi đặc biệt cho khách hàng có nguy cơ rời bỏ để thu hút họ quay trở lại.



Xây dựng mối quan hệ

Tập trung vào việc xây dựng mối quan hệ lâu dài và chặt chẽ với khách hàng.



Giải quyết vấn đề nhanh chóng

Xử lý các vấn đề và khiếu nại từ khách hàng một cách nhanh chóng và chuyên nghiệp.



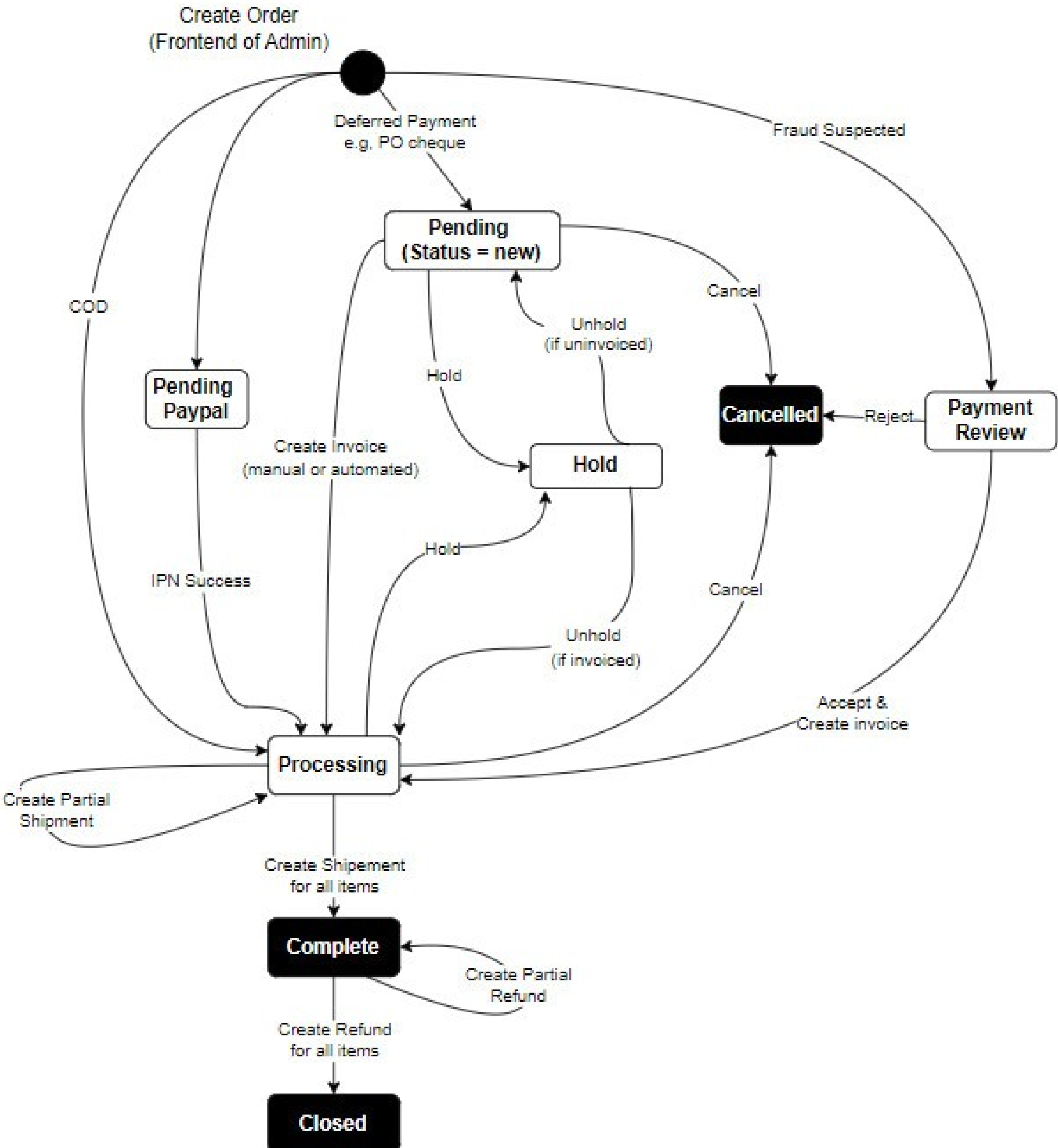
Theo dõi và đánh giá

Theo dõi và đánh giá thường xuyên việc tiến triển của mối quan hệ với khách hàng có nguy cơ rời bỏ.

◆ **III. PHÂN TÍCH DỮ LIỆU**

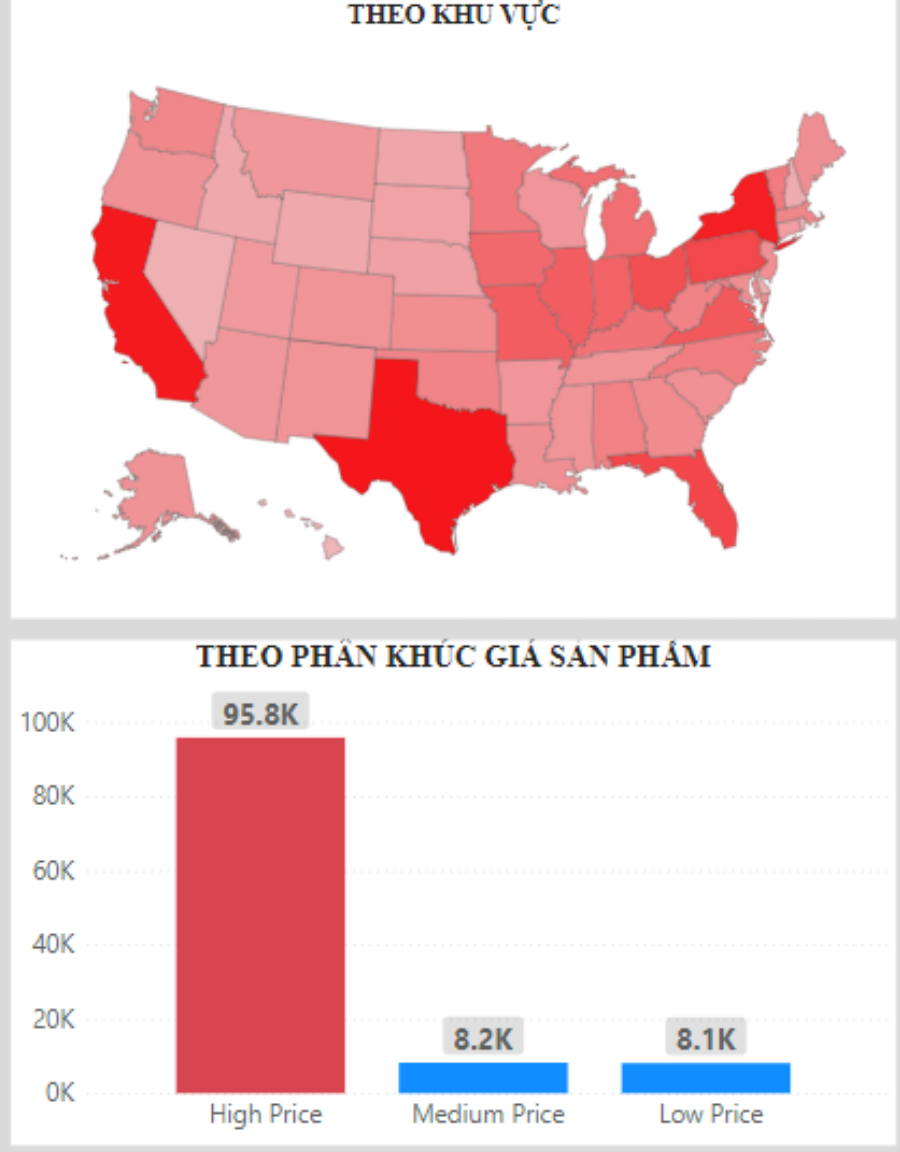
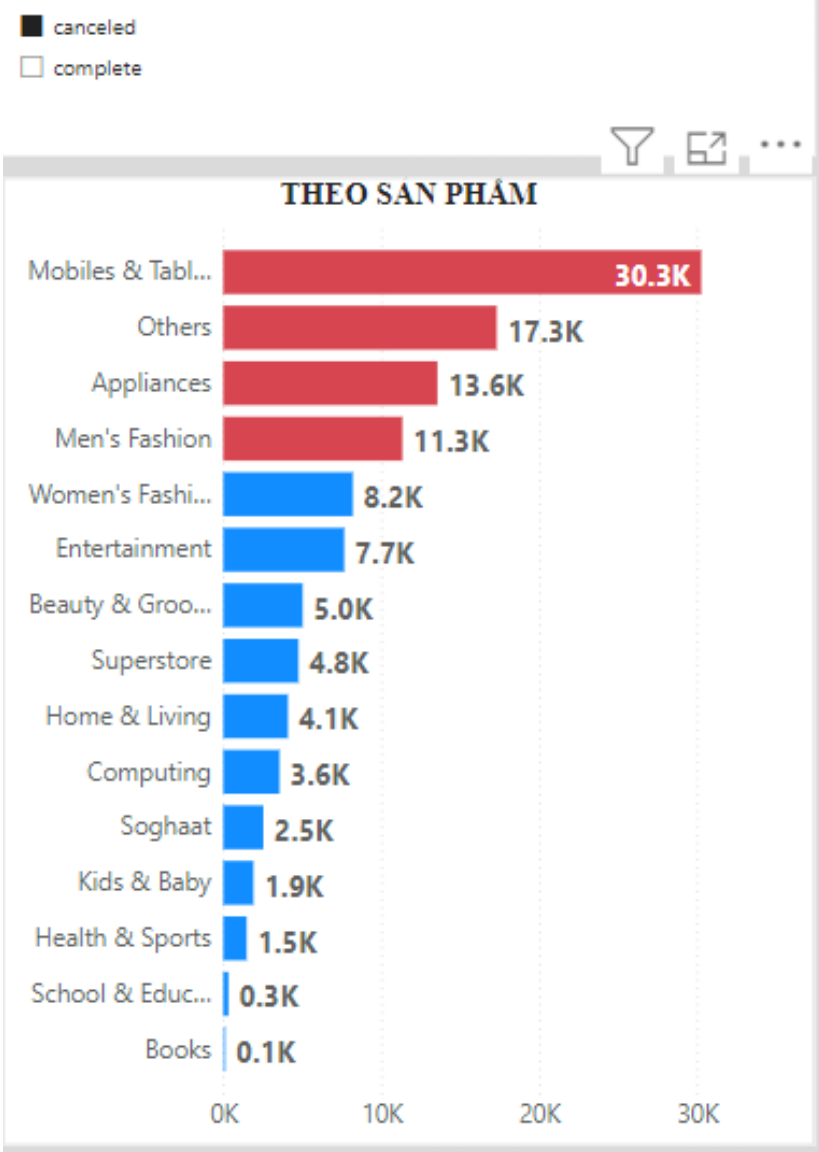
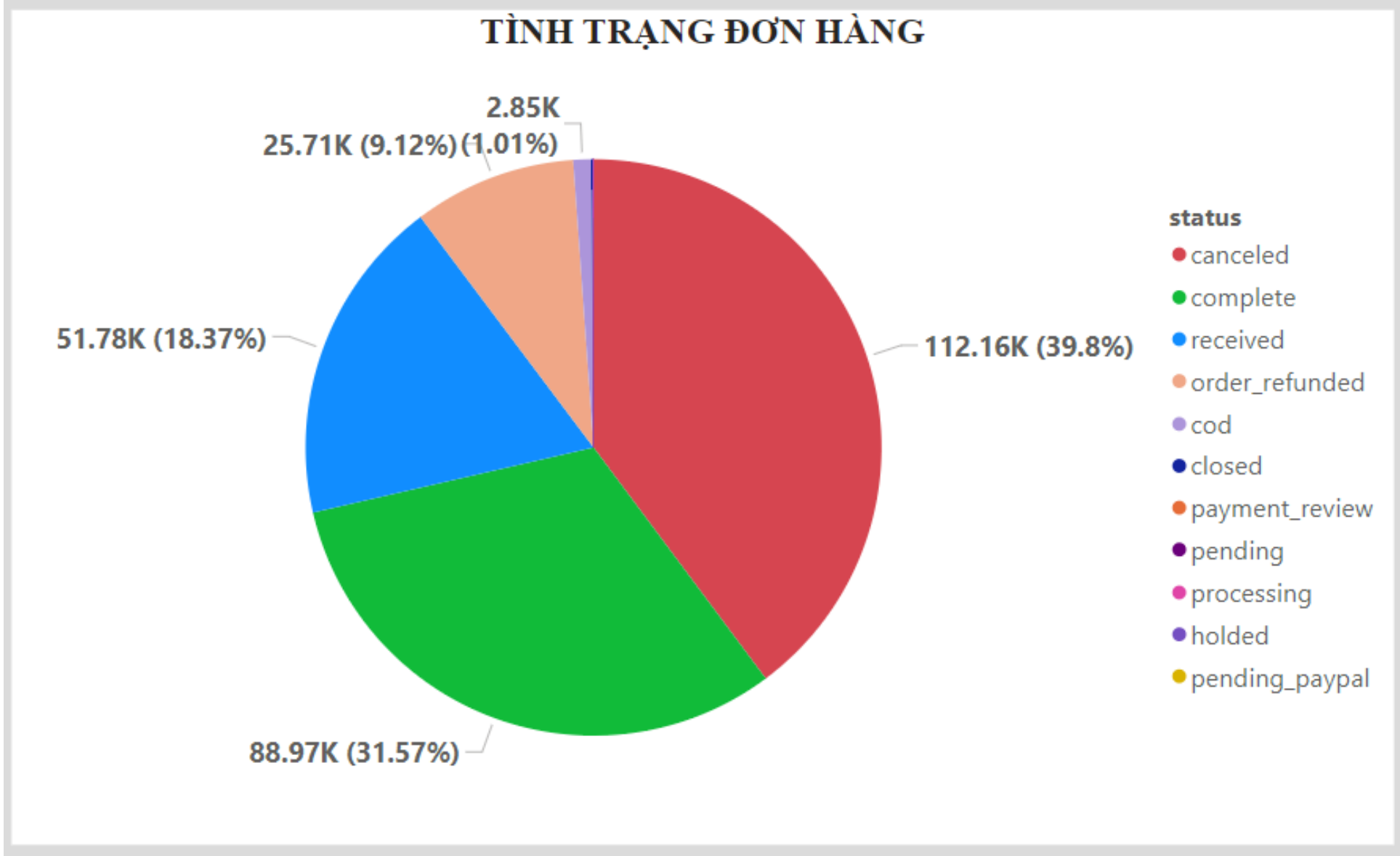
3. 3. Phân tích tình trạng đơn hàng

**QUY TRÌNH XỬ LÝ
ĐƠN HÀNG**



III. PHÂN TÍCH DỮ LIỆU

3. 3. Phân tích tình trạng đơn hàng



 Tình trạng đơn hàng bị hủy trong 9 tháng đầu năm 2021 đột ngột tăng cao lên đến gần 113 nghìn đơn hàng (~40%)

 Chủ yếu là Mobiles & Tablets, Appliances và các mặt hàng quần áo.

 Chủ yếu từ Texas, California và New York

III. PHÂN TÍCH DỮ LIỆU

3. 3. Phân tích tình trạng đơn hàng

ĐỀ XUẤT



Tìm hiểu sâu các nguyên nhân khiến khách hàng hủy đơn và giải quyết các vấn đề tồn tại.



Thúc đẩy sự tương tác và phản hồi của khách hàng để hiểu rõ hơn về nhu cầu và mong muốn của họ.



Xem xét cơ chế thanh toán qua các kênh có gây cản trở hay không.

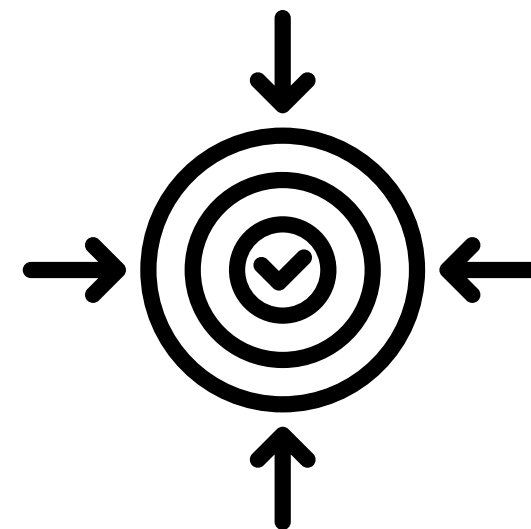


Khi nhận được hàng thành công, có thể tạo ra các xu hướng và chương trình cho khách hàng.



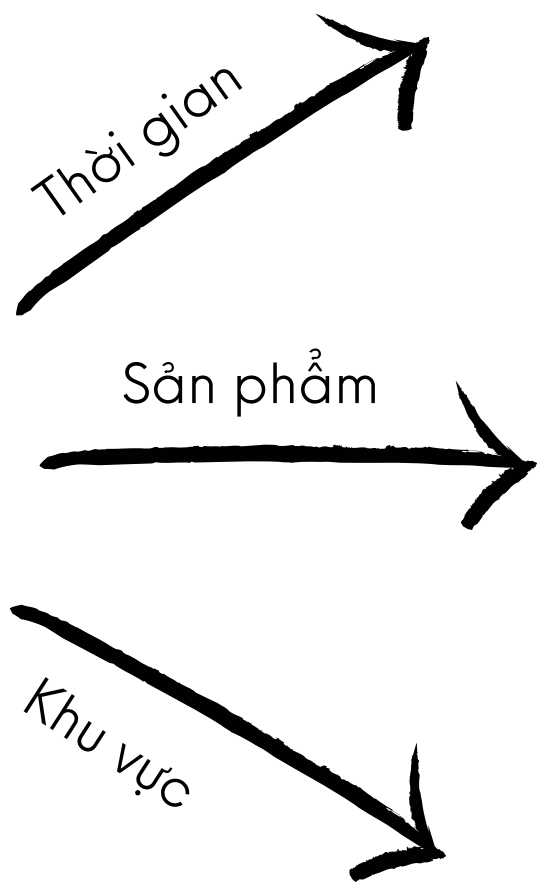


KẾT LUẬN VÀ ĐỀ XUẤT



◆ IV. KẾT LUẬN VÀ ĐỀ XUẤT

Tình hình kinh doanh



- Doanh thu Quý 3 giảm
- Số lượng đơn hàng giảm mạnh

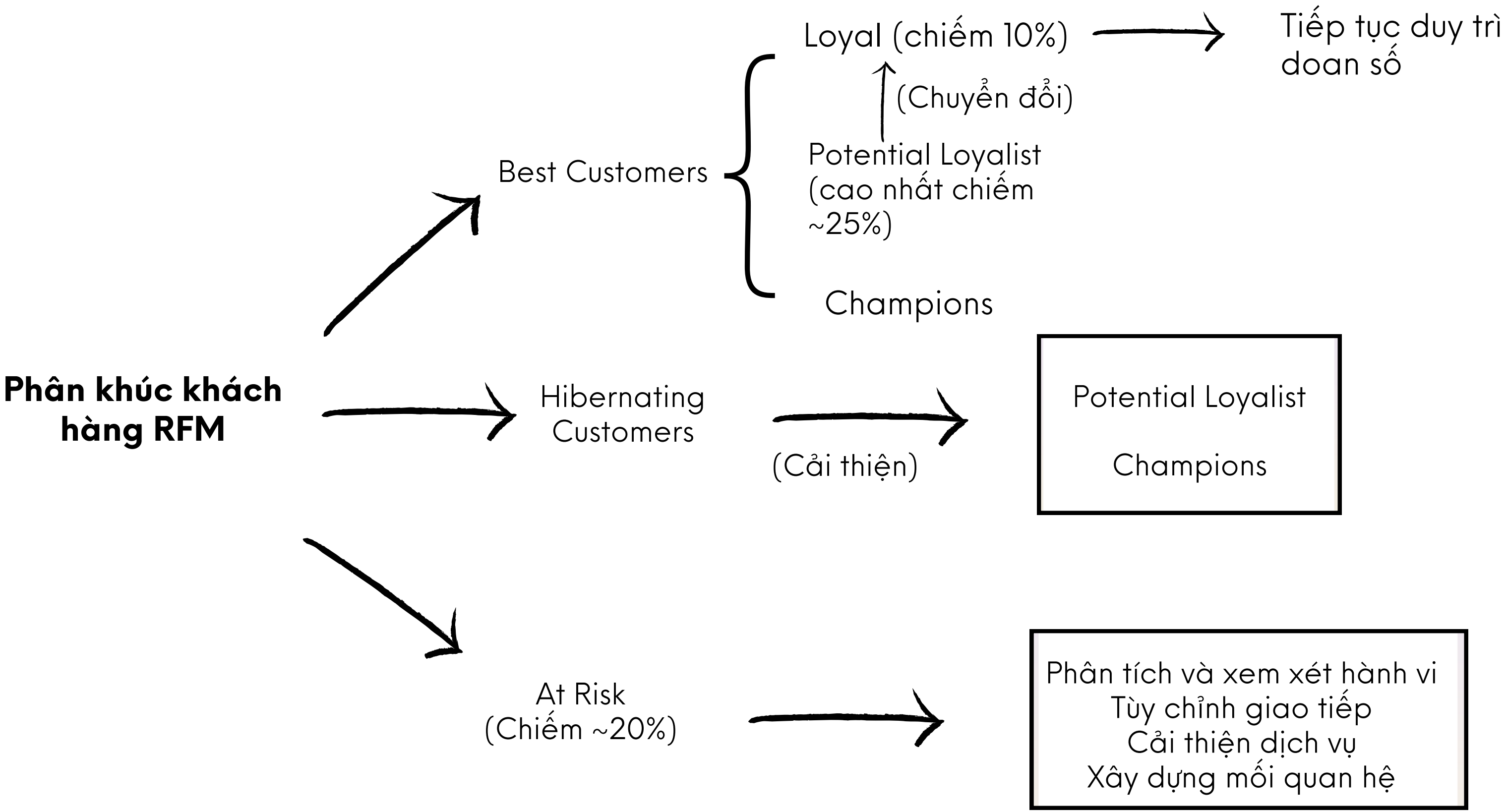
- Doanh thu chính là các sản phẩm thiết bị, công nghệ
- Số lượng đơn hàng quần áo nam nữ tăng cao

- South và Midwest đạt doanh thu cao nhất
- 3 tiểu bang Texas, California và New York là 3 nơi có doanh thu cao nhất



- Nghiên cứu và phân tích thị trường
- Chạy các chương trình khuyến mãi và quảng cáo
- Tối ưu hóa giá cả và chiến lược giá
- Nâng cao chất lượng dịch vụ
- Tích hợp công nghệ

◆ IV. KẾT LUẬN VÀ ĐỀ XUẤT

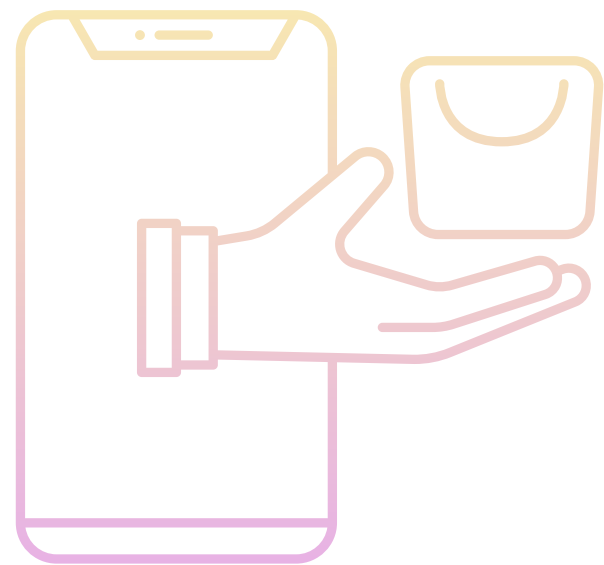


◆ IV. KẾT LUẬN VÀ ĐỀ XUẤT





**Thank
For
Your
Watching!**



NHÓM 1

E-Commerce Analysis

