



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 BỘ DATASET



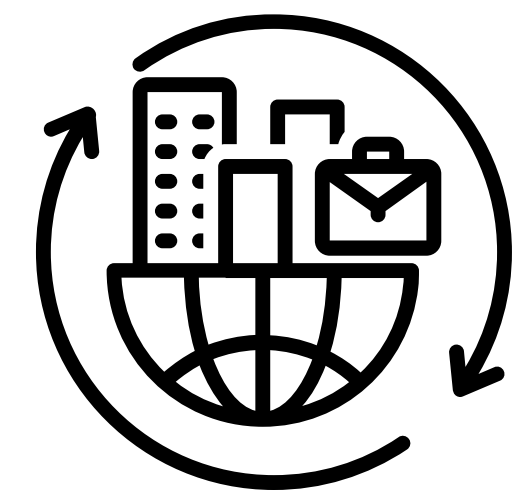
III. PHÂN TÍCH DỮ LIỆU



IV. KẾT LUẬN VÀ ĐỀ XUẤT

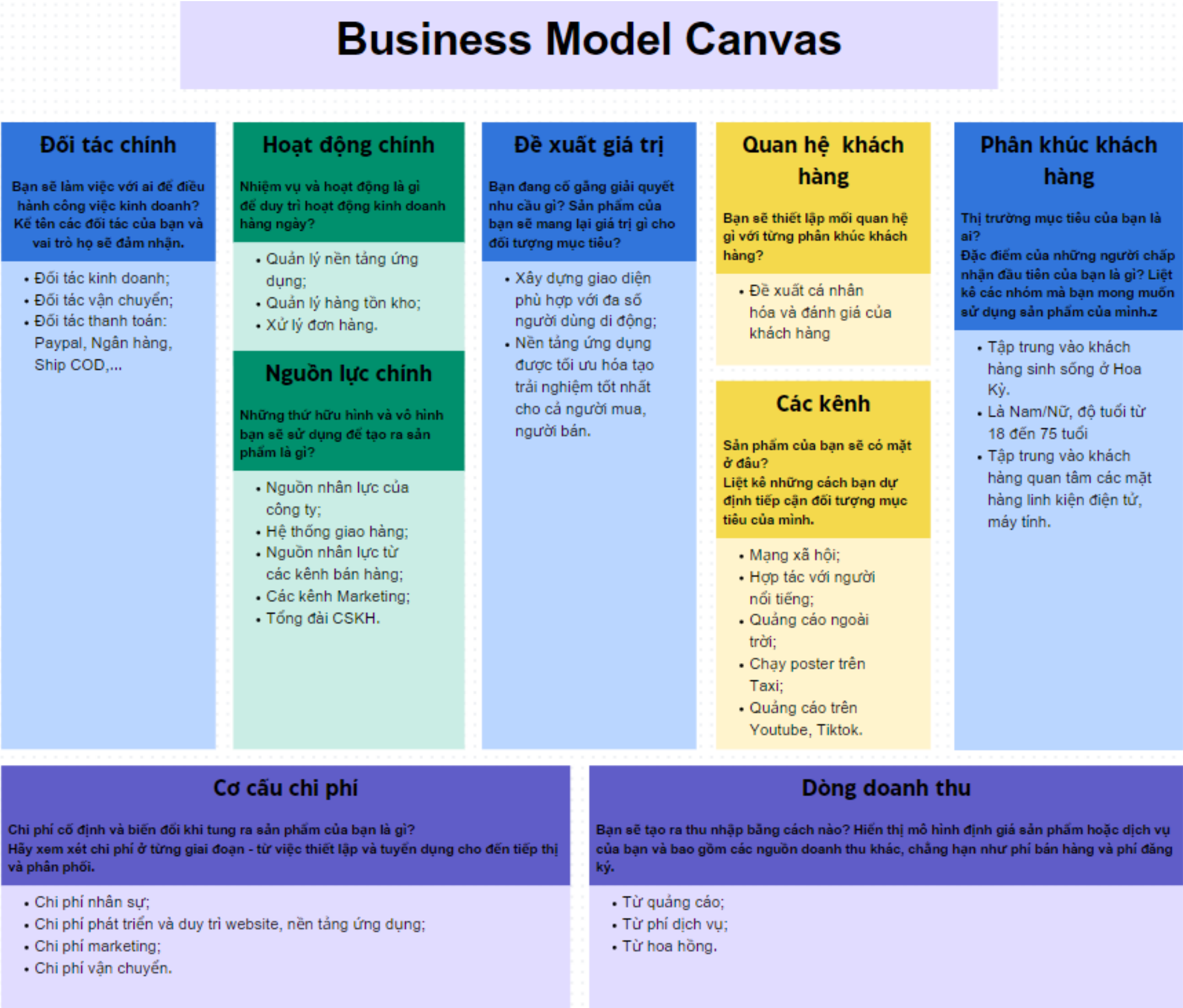


BUSINESS DESCRIPTION



◆ I. Business Description

1. 1. Business Model Canvas



◆ I. Business Description

1. 2. Giới thiệu vấn đề



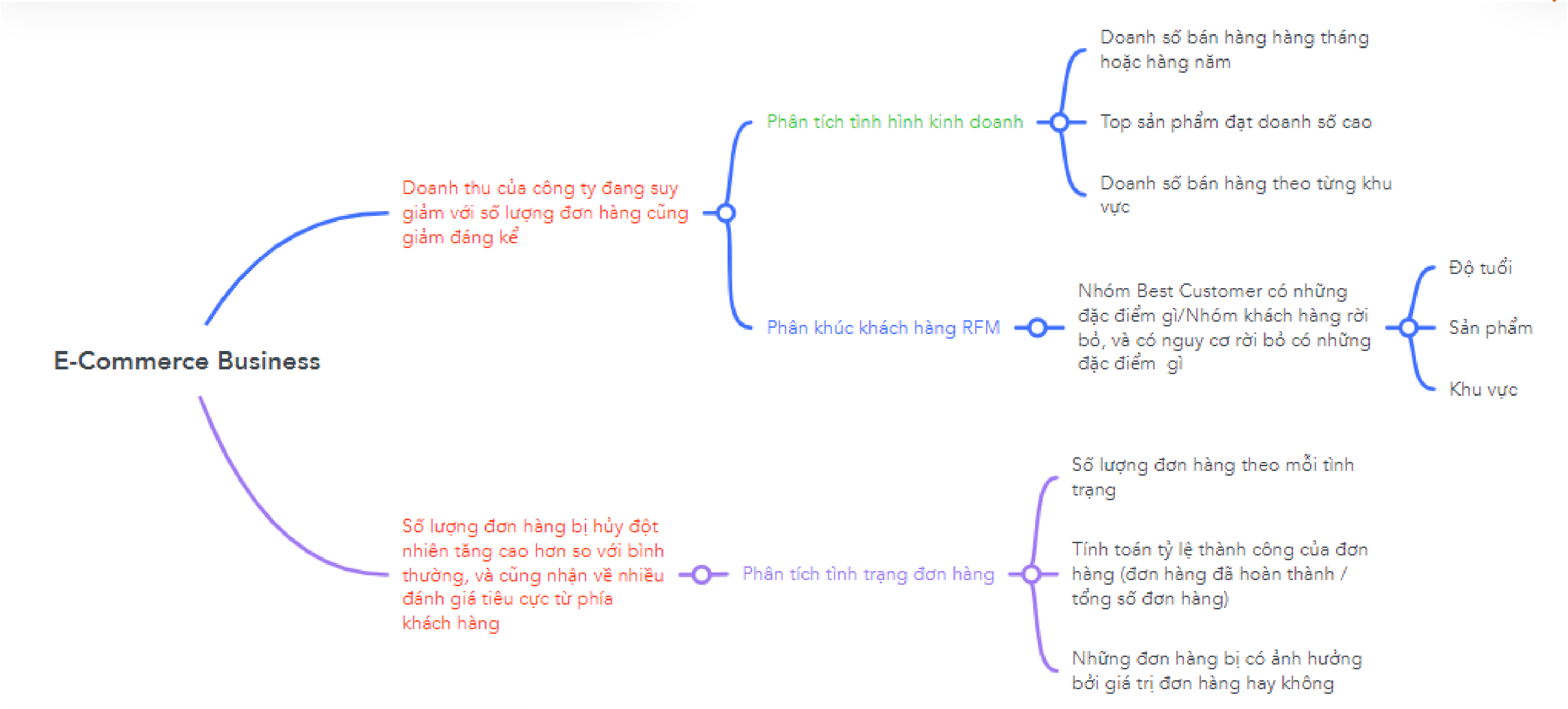
Theo báo cáo của Giám đốc, số lượng đơn hàng và doanh thu của công ty đã giảm trong **Quý 3/2021**.

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.

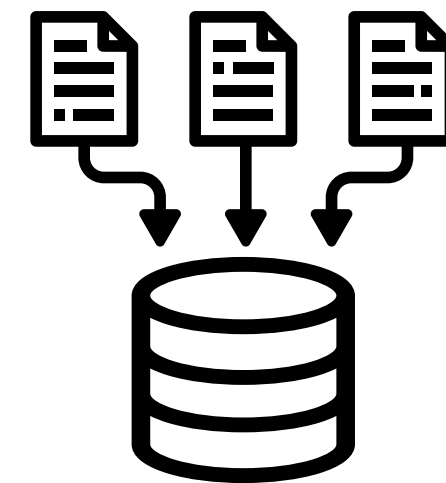


◆ I. Business Description

1. 3. Hướng xử lý



TỔNG QUAN BỘ 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

(*) Chỉ lấy những cột có giá trị sử dụng

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

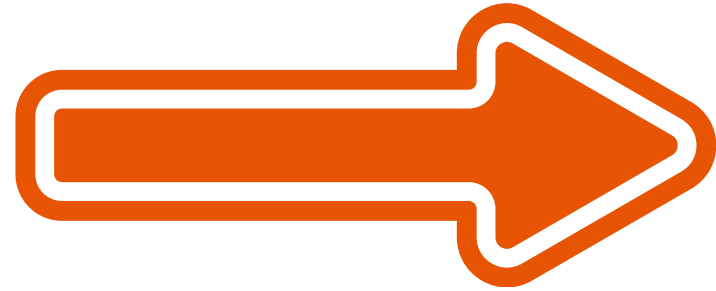
Kiểu dữ liệu nói về ngày tháng chưa đúng

Tên cột chưa được đồng nhất

◆ II. Tổng quan Dataset

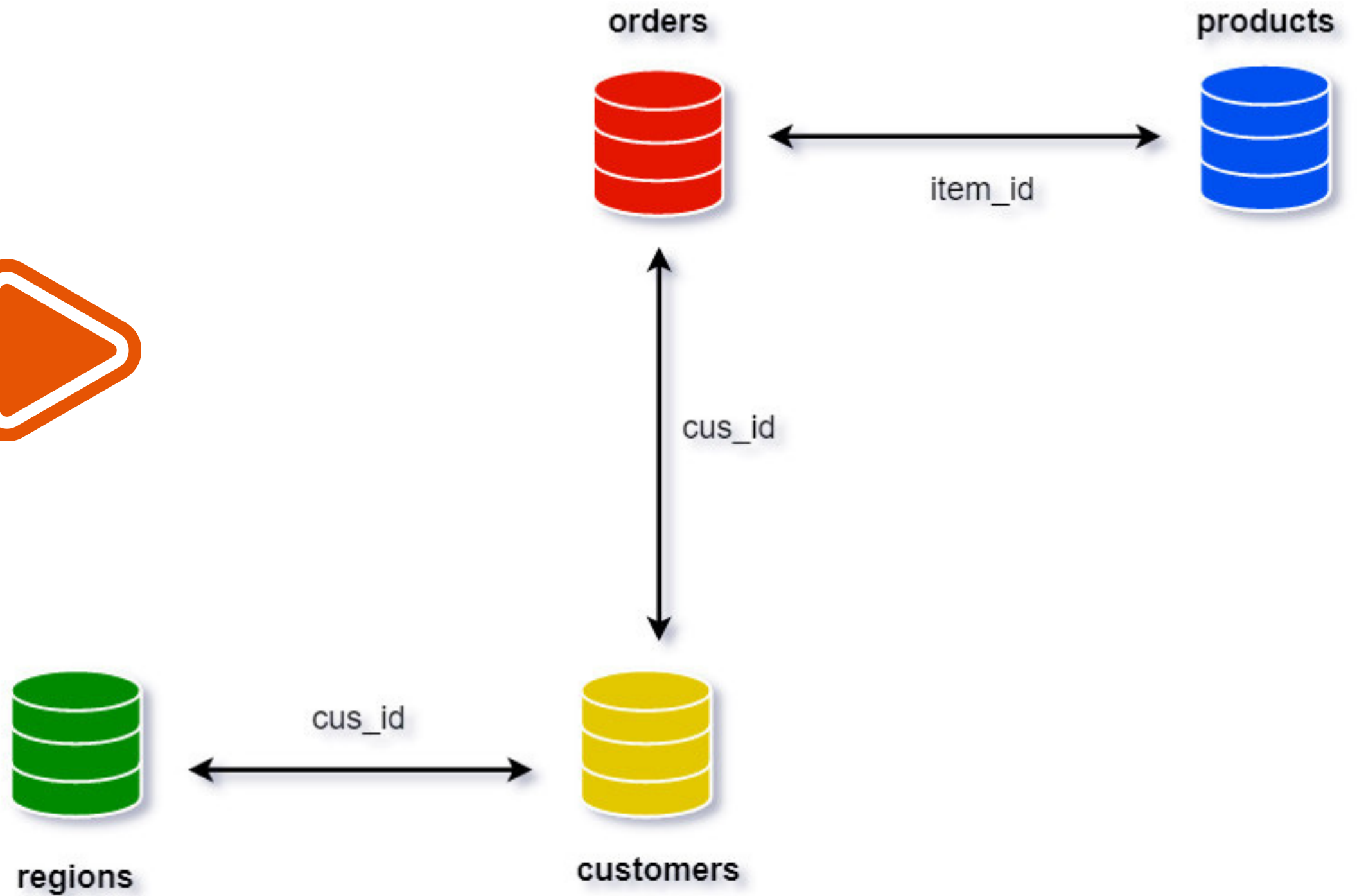
2. 3. Hướng xử lý dữ liệu

Tổ chức dữ liệu thành Schema



Lợi ích của việc chuẩn hóa:

- 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)

DataFrame rỗng --> không có giá trị null

```
check_duplicate = df.duplicated().sum()
print(check_duplicate)
```

✓ 0.5s

0

Không có dòng nào bị duplicate



Đồng nhất tên các cột

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

Kết quả sau khi sửa

```
df.columns

Index(['order_id', 'order_date', 'status', 'item_id', 'sku', 'qty_ordered',
      'price', 'value', 'discount_amount', 'total', 'category',
      'payment_method', 'bi_st', 'cust_id', 'year', 'month', 'ref_num',
      'name_prefix', 'first_name', 'middle_initial', 'last_name', 'gender',
      'age', 'full_name', 'e_mail', 'customer_since', 'ssn', 'phone_no',
      'place_name', 'county', 'city', 'state', 'zip', 'region', 'user_name',
      'discount_percent'],
      dtype='object')
```

II. Tổng quan Dataset

2. 4. Các bước xử lý dữ liệu

Các bước xử lý dữ liệu

Đưa các cột ngày về datetime

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

Kết quả sau khi đổi

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 286392 entries, 0 to 286391
Data columns (total 2 columns):
#   Column          Non-Null Count  Dtype
---  -
0   order_date      286392 non-null datetime64[ns]
1   customer_since  286392 non-null datetime64[ns]
dtypes: datetime64[ns](2)
memory usage: 4.4 MB
```



	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
```

Kiểm tra lại

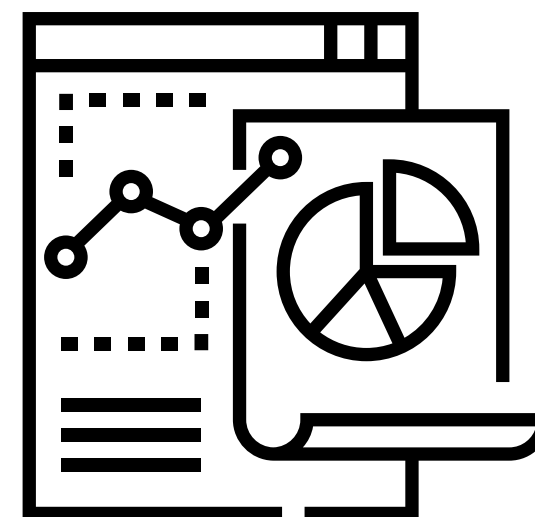
```
df[df['order_id']=='100468520-1']
✓ 0.1s
```

order_id	order_date	status	item_id	sku	qty_ordered	price	value	discount_amount	total	...	SSN	Phone No.	Place Name	County	City
----------	------------	--------	---------	-----	-------------	-------	-------	-----------------	-------	-----	-----	-----------	------------	--------	------

0 rows x 36 columns



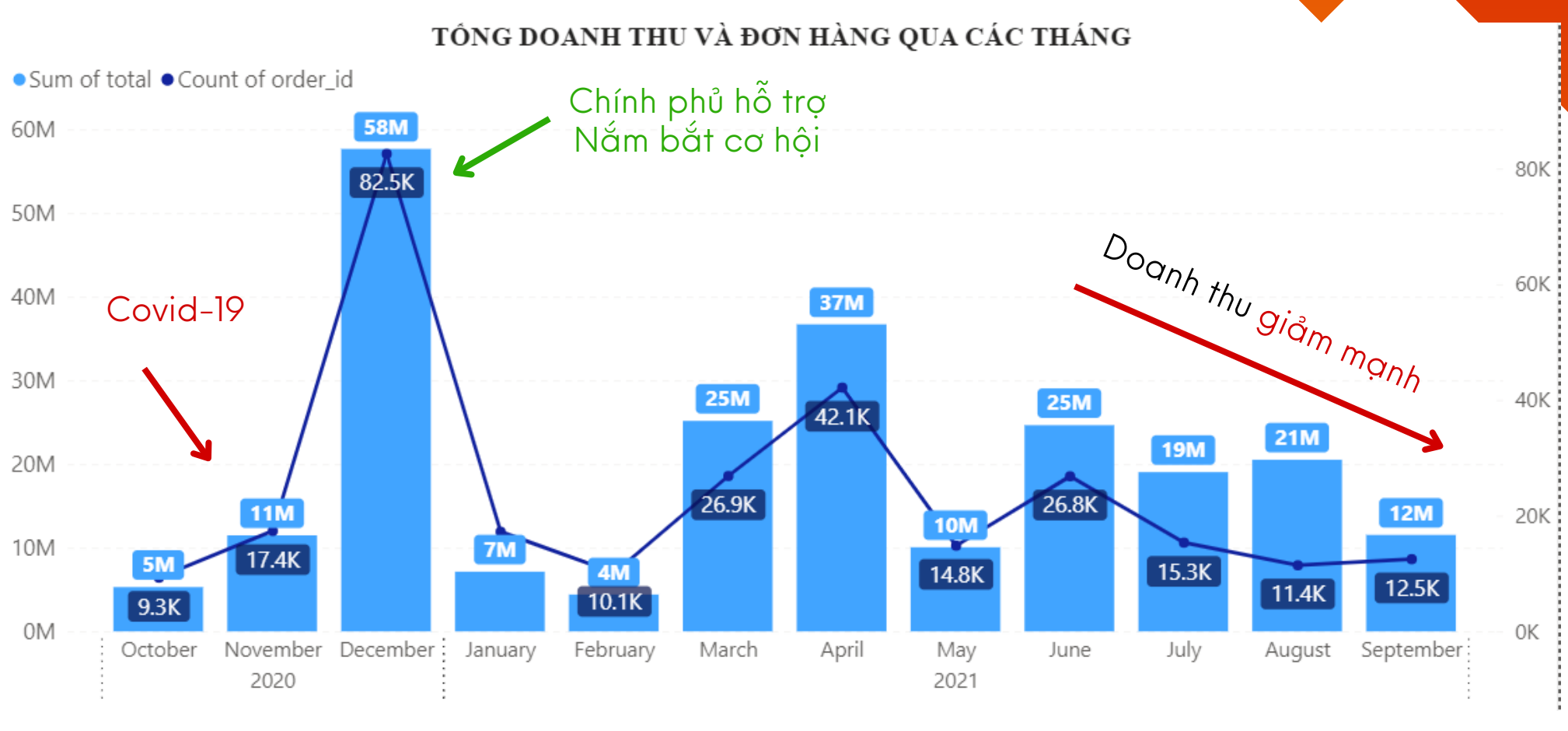
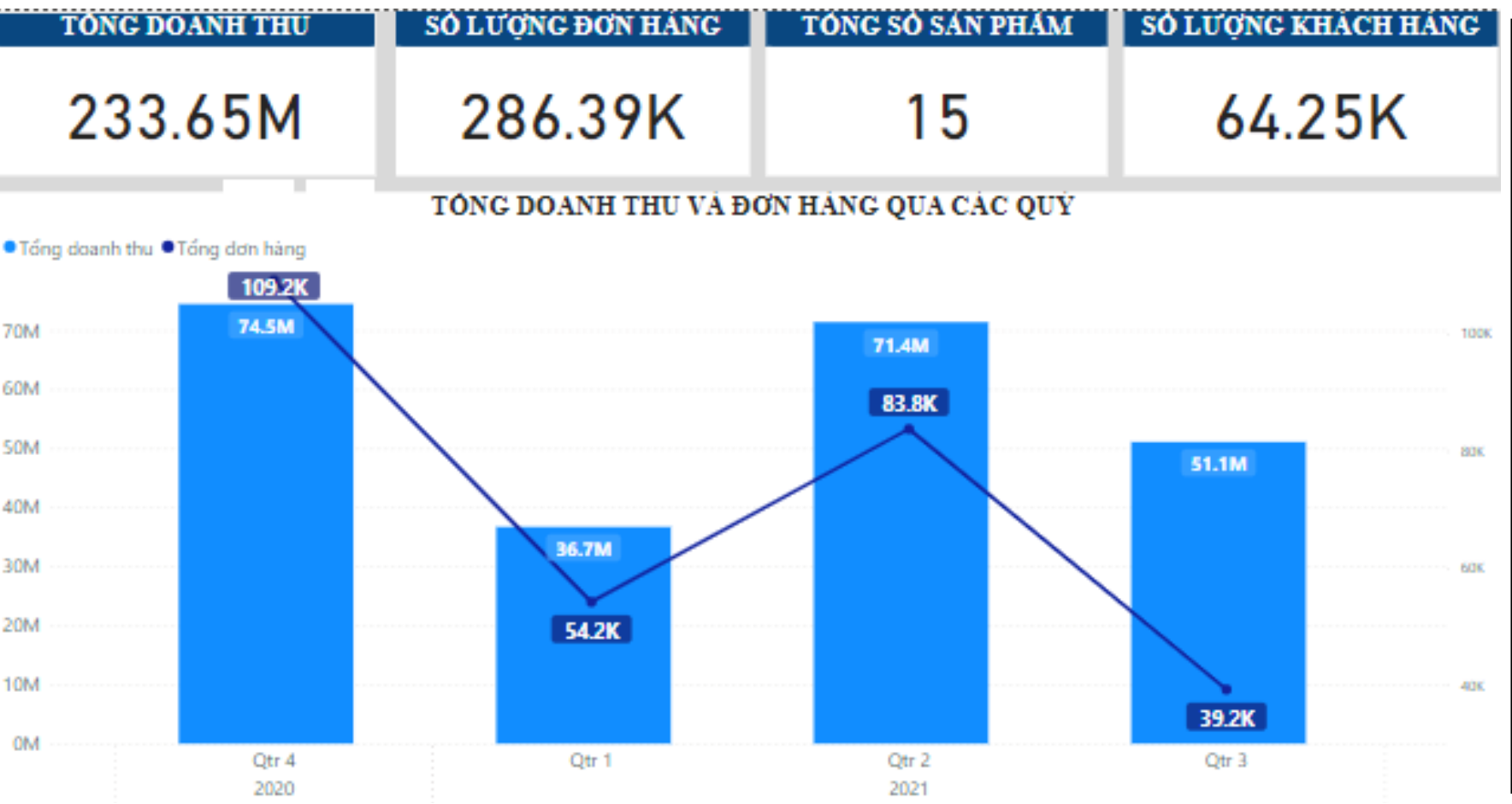
PHÂN TÍCH DỮ LIỆU



III. PHÂN TÍCH DỮ LIỆU

3. 1. Phân tích tình hình kinh doanh

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



Tuy nhiên, việc hạn chế vận chuyển đã làm **giảm** doanh thu trong tháng 10 và 11/2020.

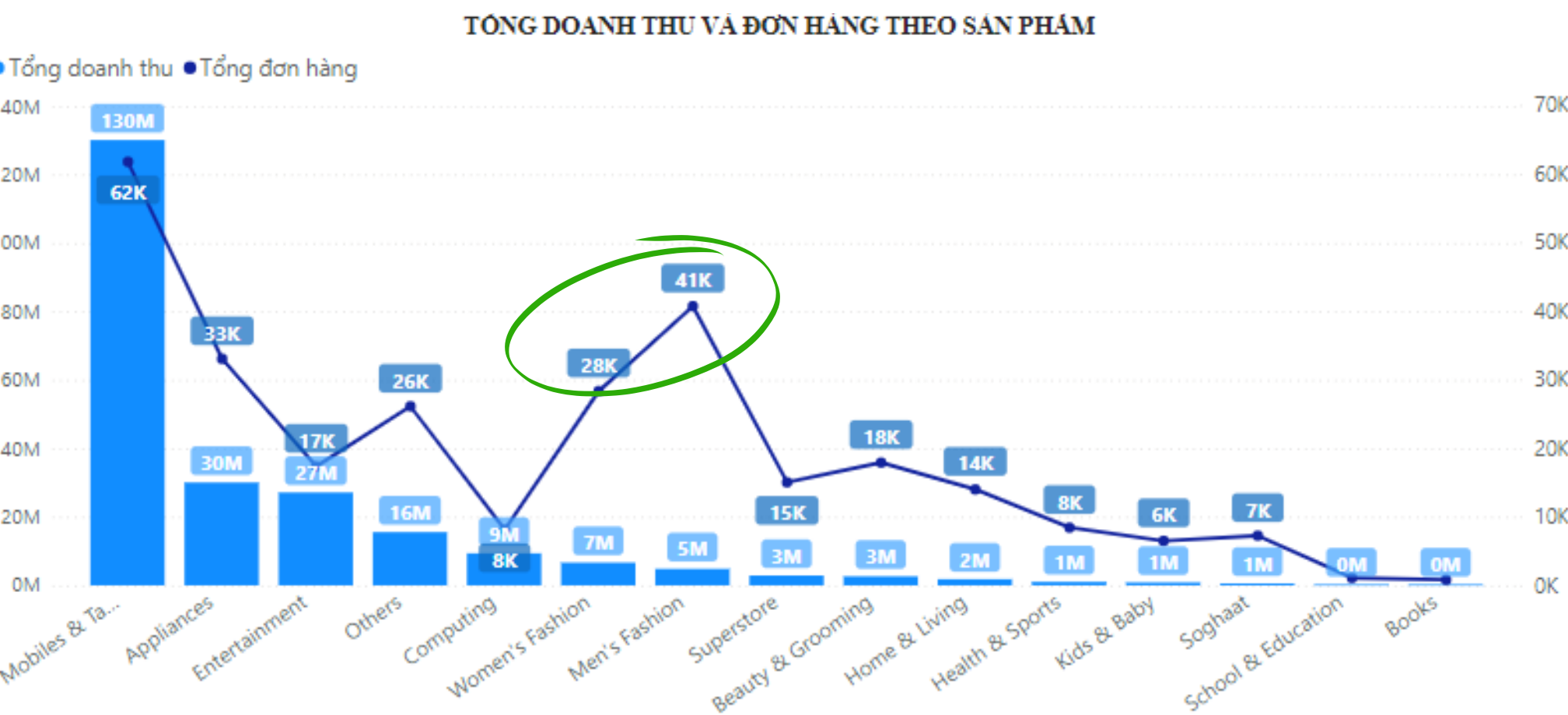
Doanh thu tháng 12/2020 đạt **58,4 triệu USD**, **tăng** gần **4 lần** so với tháng trước.

Năm 2021, doanh thu Quý 3 **giảm** còn **51,1 triệu USD (0,3 lần so với quý trước)**. Số lượng đơn hàng cuối Quý 3 **giảm** còn trung bình khoảng **5,5 nghìn đơn hàng**.

III. PHÂN TÍCH DỮ LIỆU

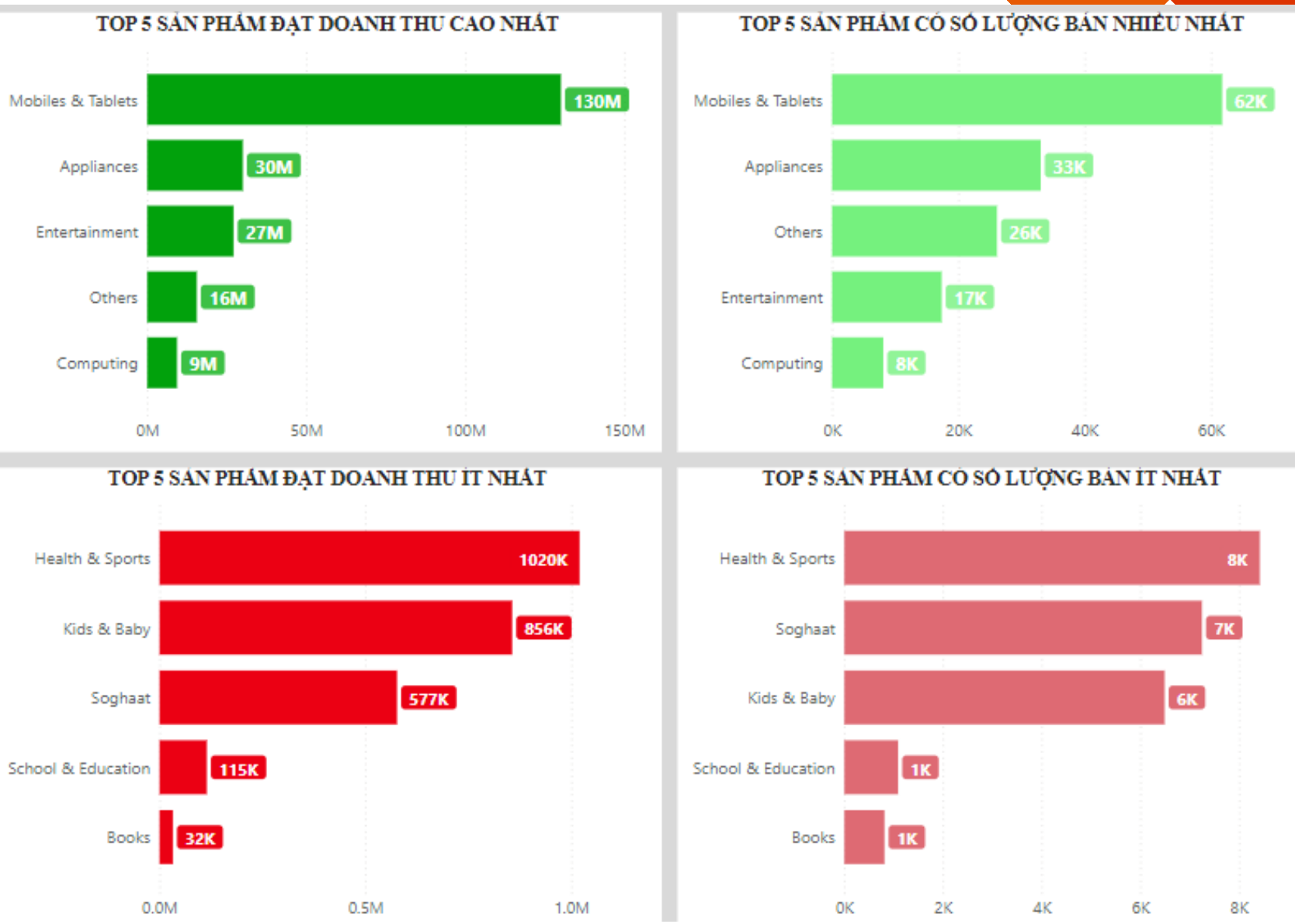
3. 1. Phân tích tình hình kinh doanh

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



Chủ yếu là các sản phẩm liên quan đến **thiết bị và các đồ điện tử, máy tính, điện thoại**.

Sản phẩm thời trang nam nữ **không có doanh thu cao**, nhưng số lượng đơn hàng ở mức **cao**



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

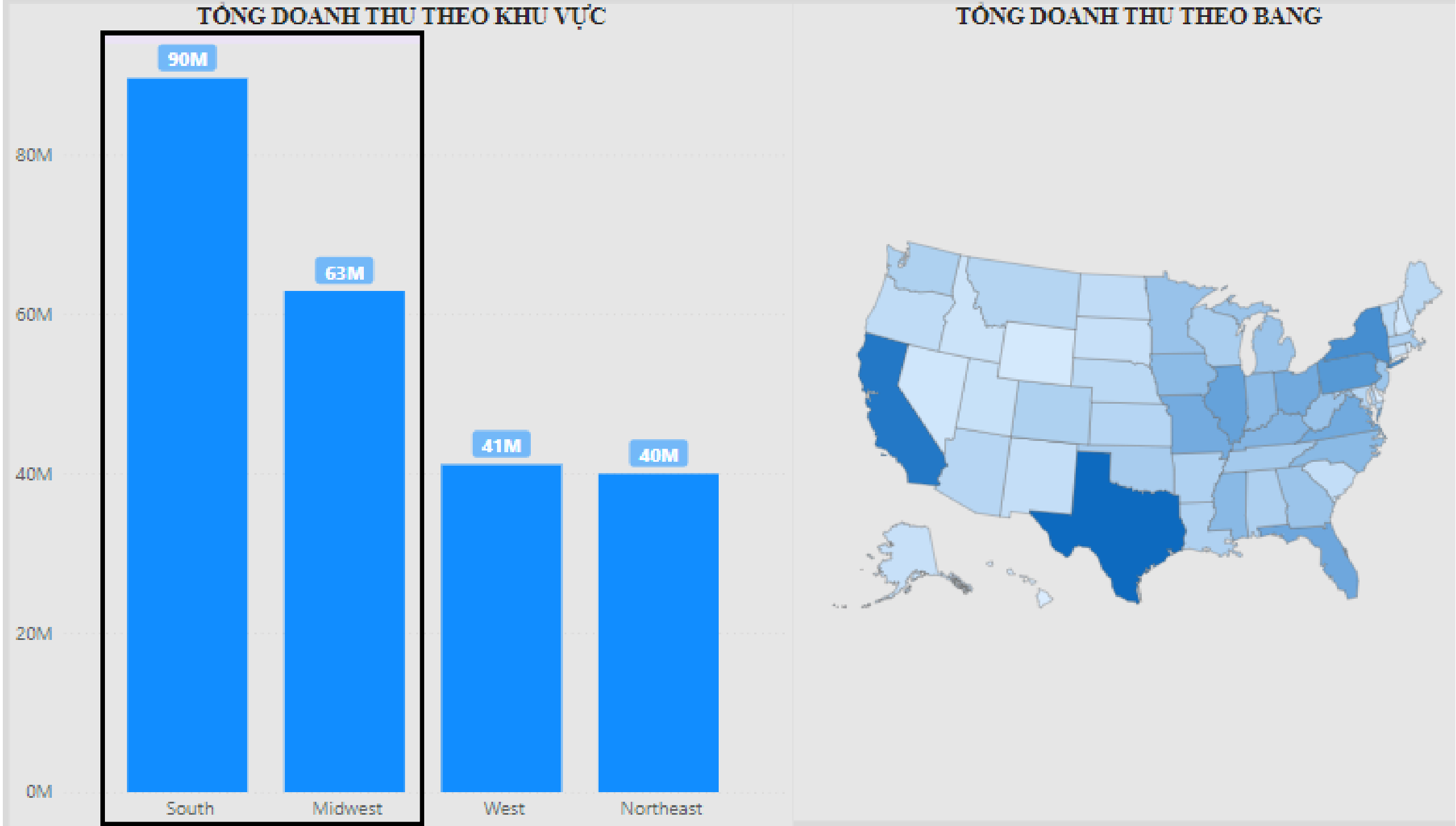
Bốn mặt hàng còn lại tuy có doanh thu trong top 5 nhưng số lượng sản phẩm bán ra lại **không quá nhiều**.

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

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** (chiếm ~65% tổng doanh thu)

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

- Hiểu rõ nhu cầu và mong muốn của thị trường
- Nghiên cứu đối thủ.



Chạy các chương trình khuyến mãi và quảng cáo

- Mở rộng hoạt động quảng cáo, tiếp thị sản phẩm
- Chiến lược giảm giá, gói ưu đãi



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 chất lượng đội CSKH
- Xây dựng tập khách hàng trung thành, tiềm năng

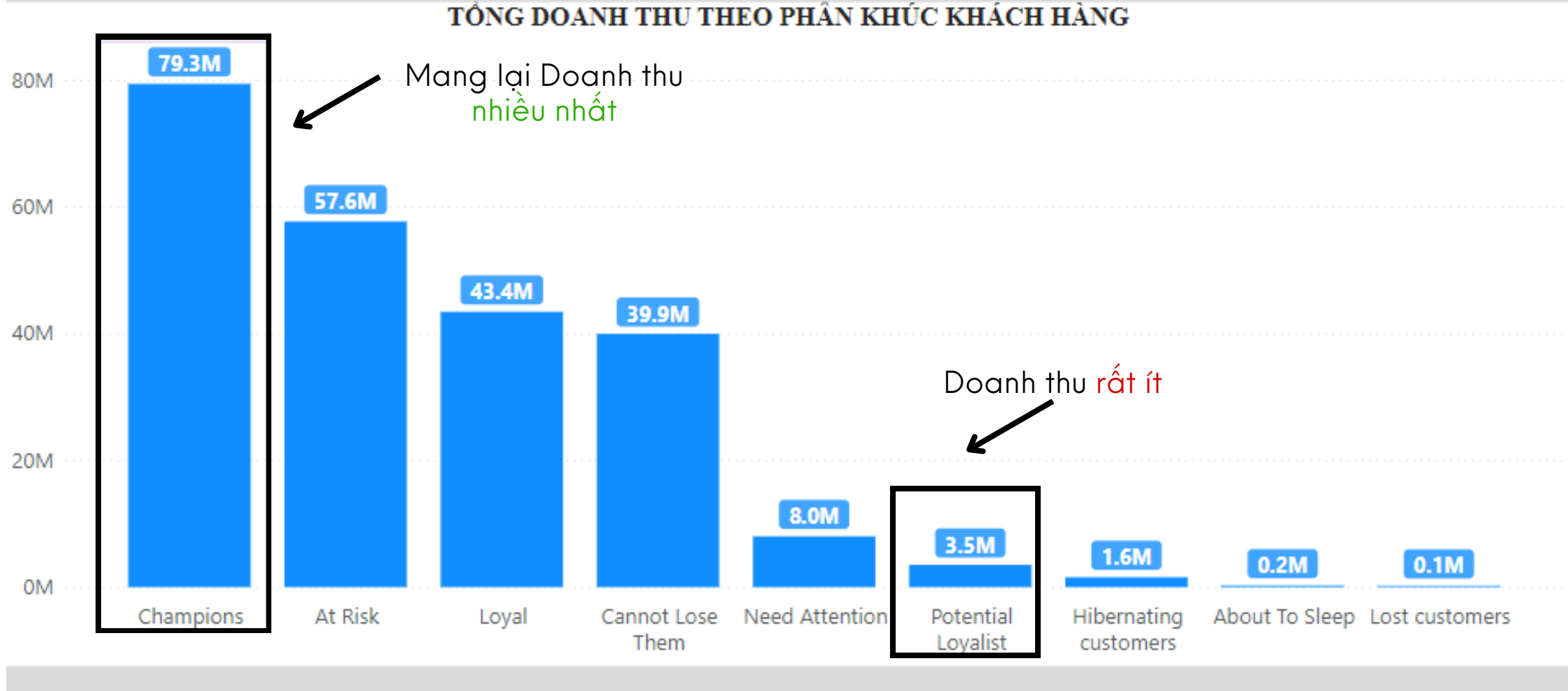
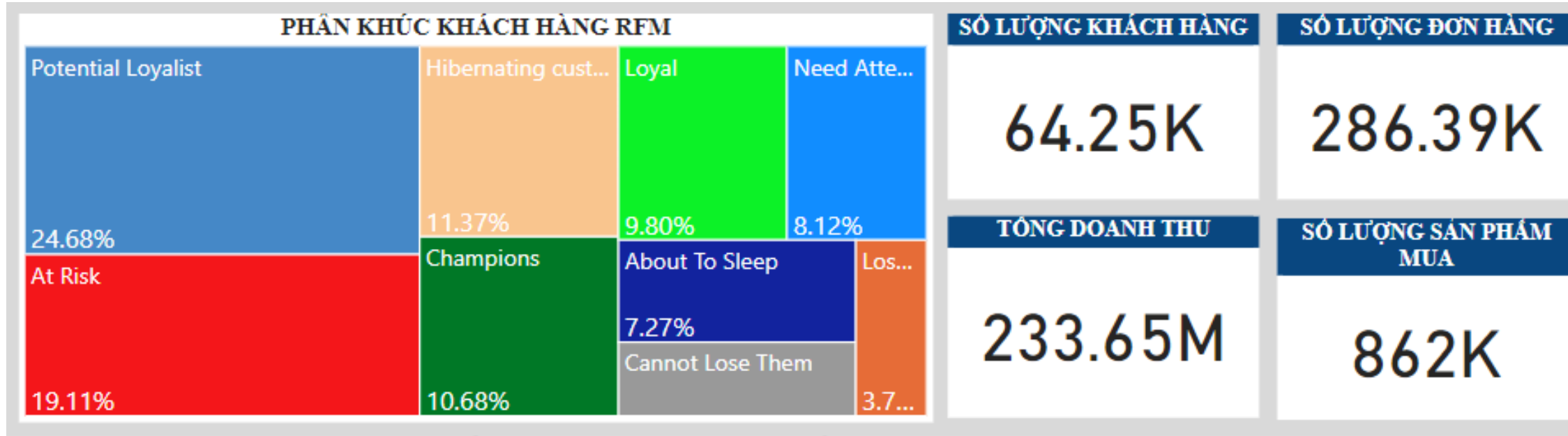


Tích hợp công nghệ: Sử dụng công nghệ để cải thiện trải nghiệm khách hàng và quá trình kinh doanh.

III. PHÂN TÍCH DỮ LIỆU

3. 2. Phân tích phân khúc khách hàng RFM

3. 2. 1. Nhóm khách hàng hiện tại



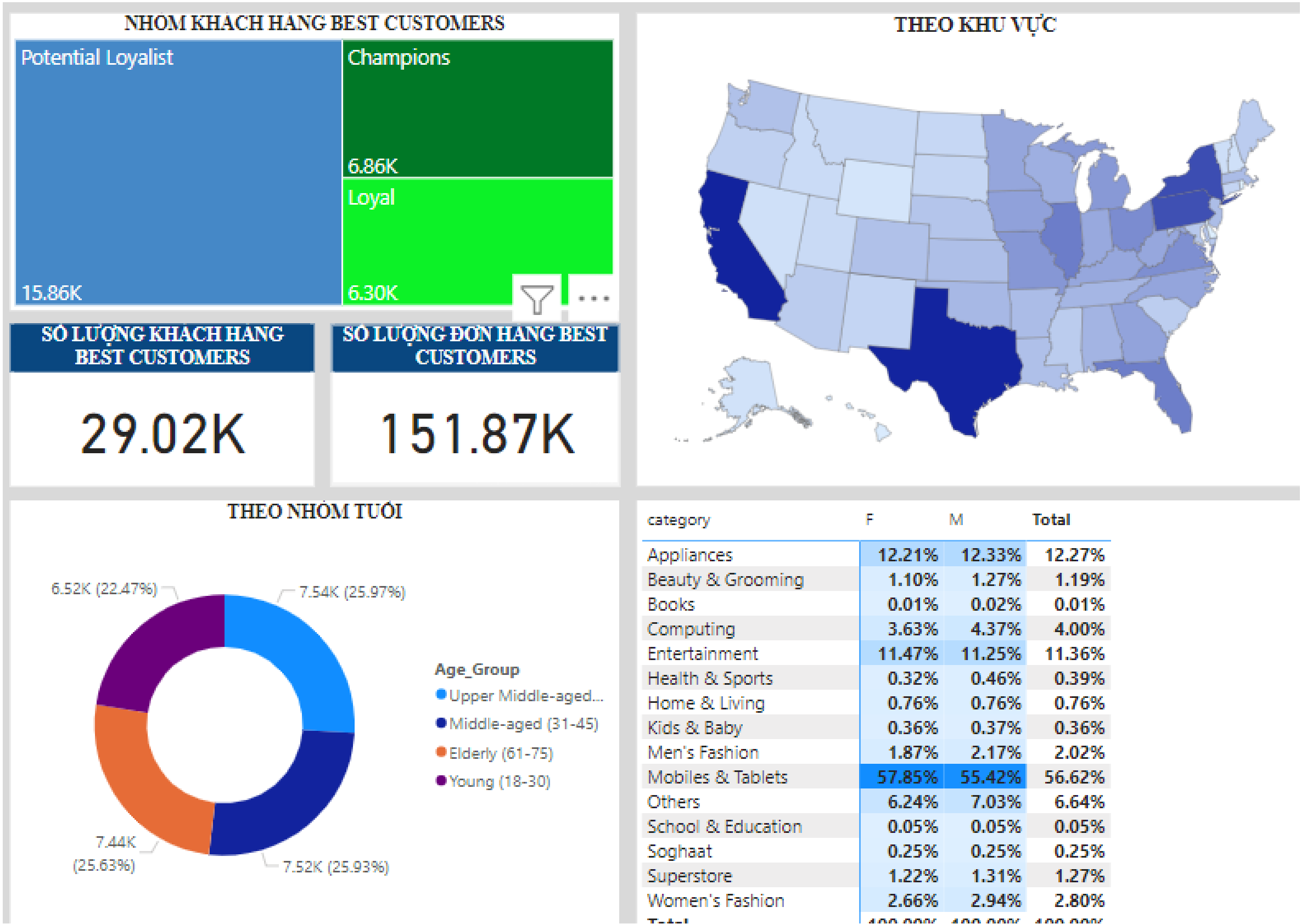
- Nhóm khách hàng **Potential Loyalist** chiếm tỷ trọng **cao nhất** (~25% tổng số).
 - Champions** chiếm tỷ trọng **không quá cao**, chỉ xếp thứ 4 (~10.7%)
- => Công ty đang **thiếu hiệu quả** trong việc chuyển đổi khách hàng tiềm năng.

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



BEST CUSTOMERS

29,02 nghìn khách hàng

152 nghìn đơn hàng

Phần lớn từ **Texas, California** và **New York**

Chủ yếu là **Mobiles & Tablets**

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

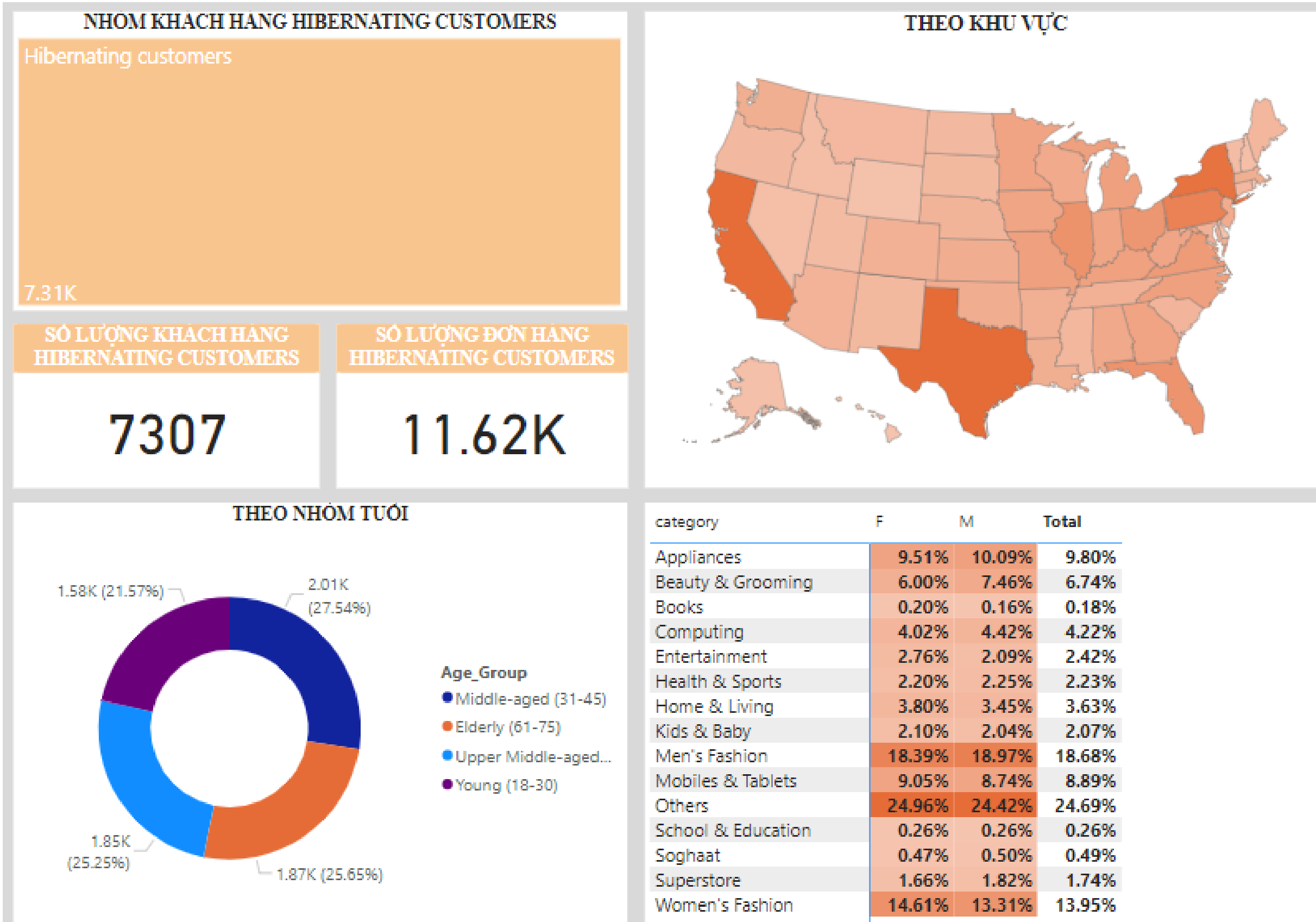
Chuyển khách hàng tiềm năng thành khách hàng trung thành (Loyal)

-  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

Texas, California và New York

Men and Women's Fashion.

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

◆ 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 khách hàng 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

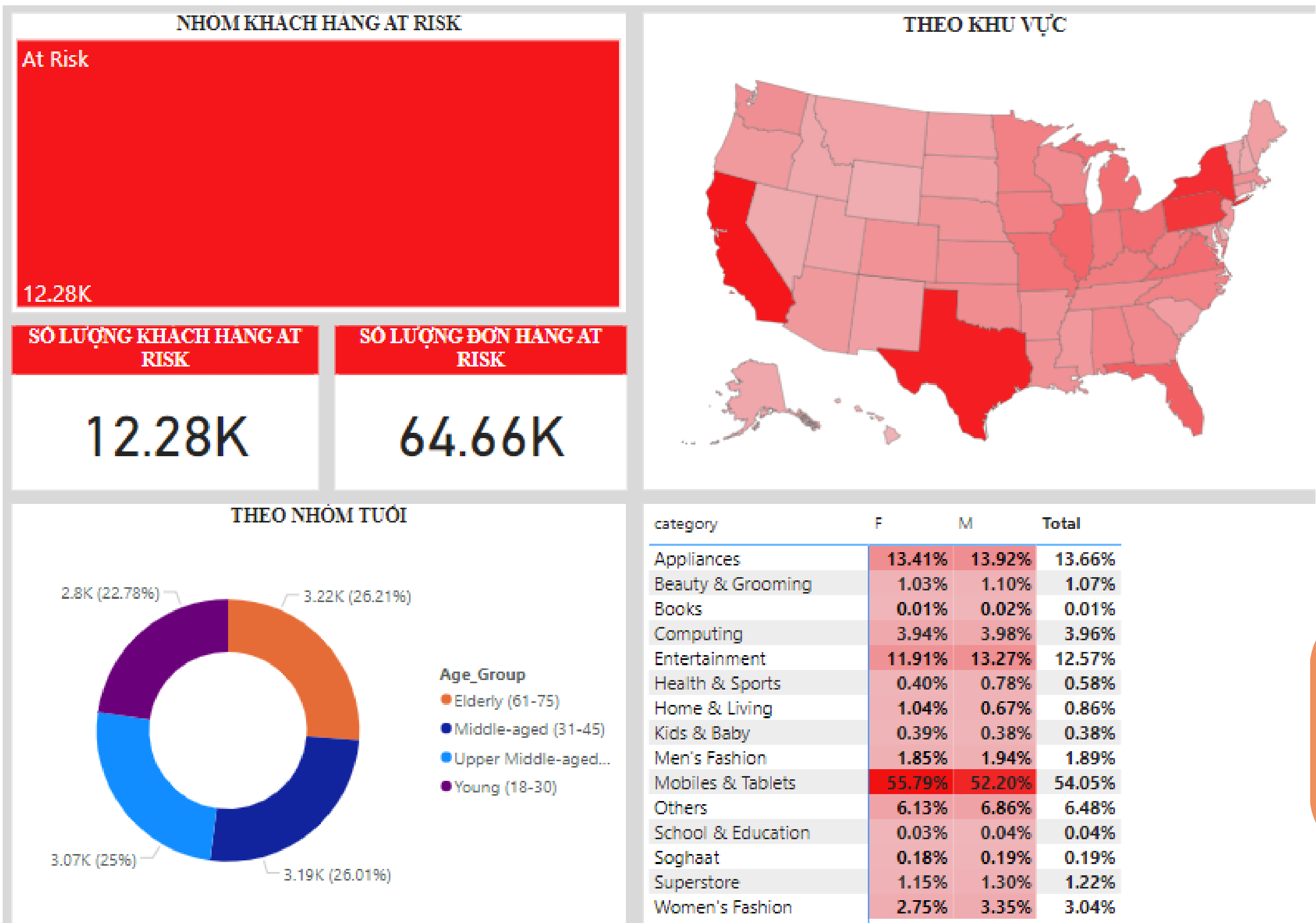


- Thường mua các mặt hàng liên quan đến quần áo thời trang
- Đẩy mạnh các chiến lược giảm giá
- Phát triển nhiều sản phẩm mới
- Tạo nhiều cơ hội hơn cho khách hàng tiếp cận các mặt hàng còn lại.

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

Phần lớn là nhóm **Người già** (khoảng **26%**)

Chủ yếu là **Mobiles & Tablets**

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



Tùy chỉnh
giao tiếp



Cải thiện dịch vụ



Chương trình
khuyến mãi



Xây dựng
mối quan hệ



Giải quyết
vấn đề nhanh chóng

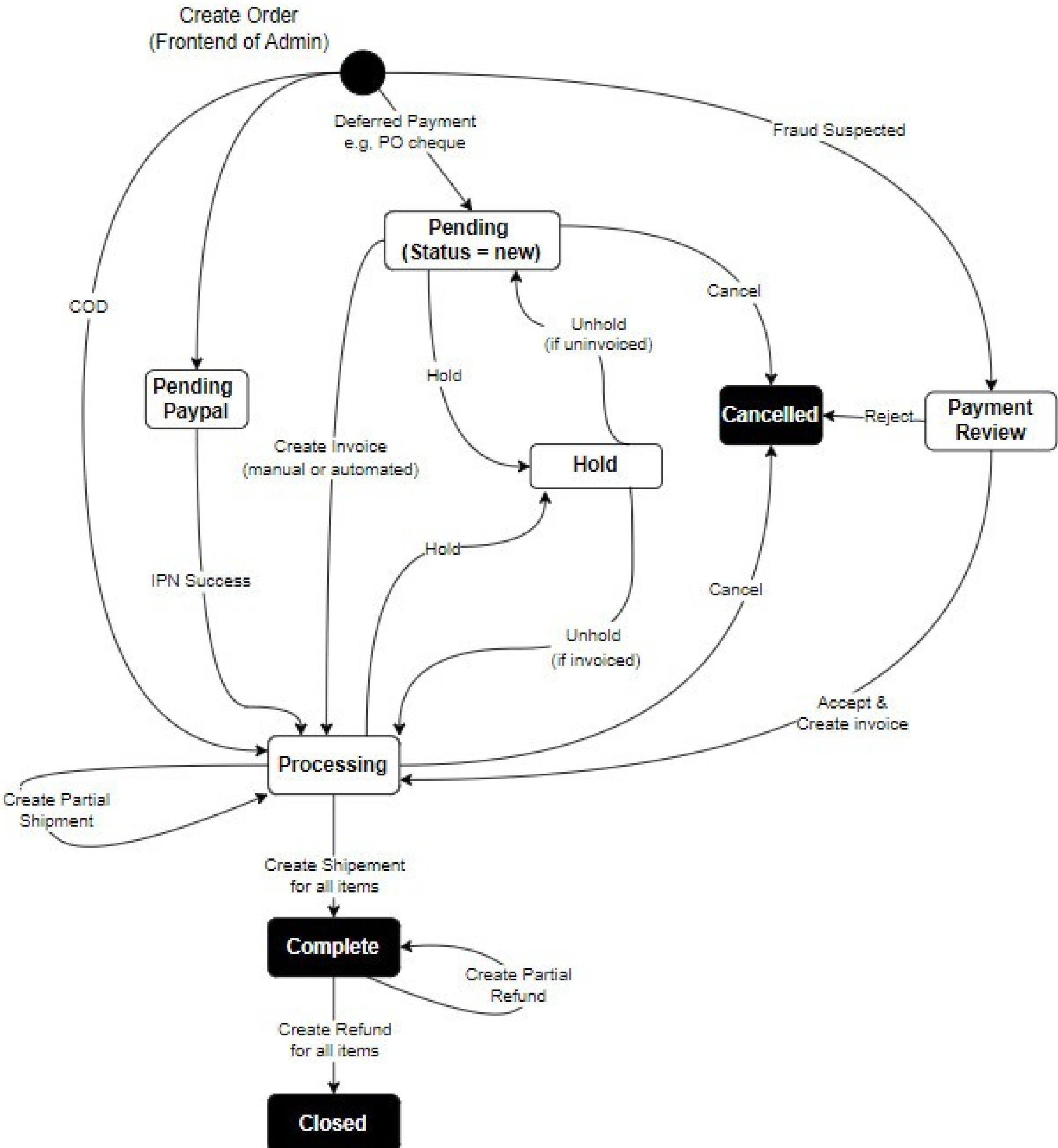


Theo dõi
và đánh giá

◆ **III. PHÂN TÍCH DỮ LIỆU**

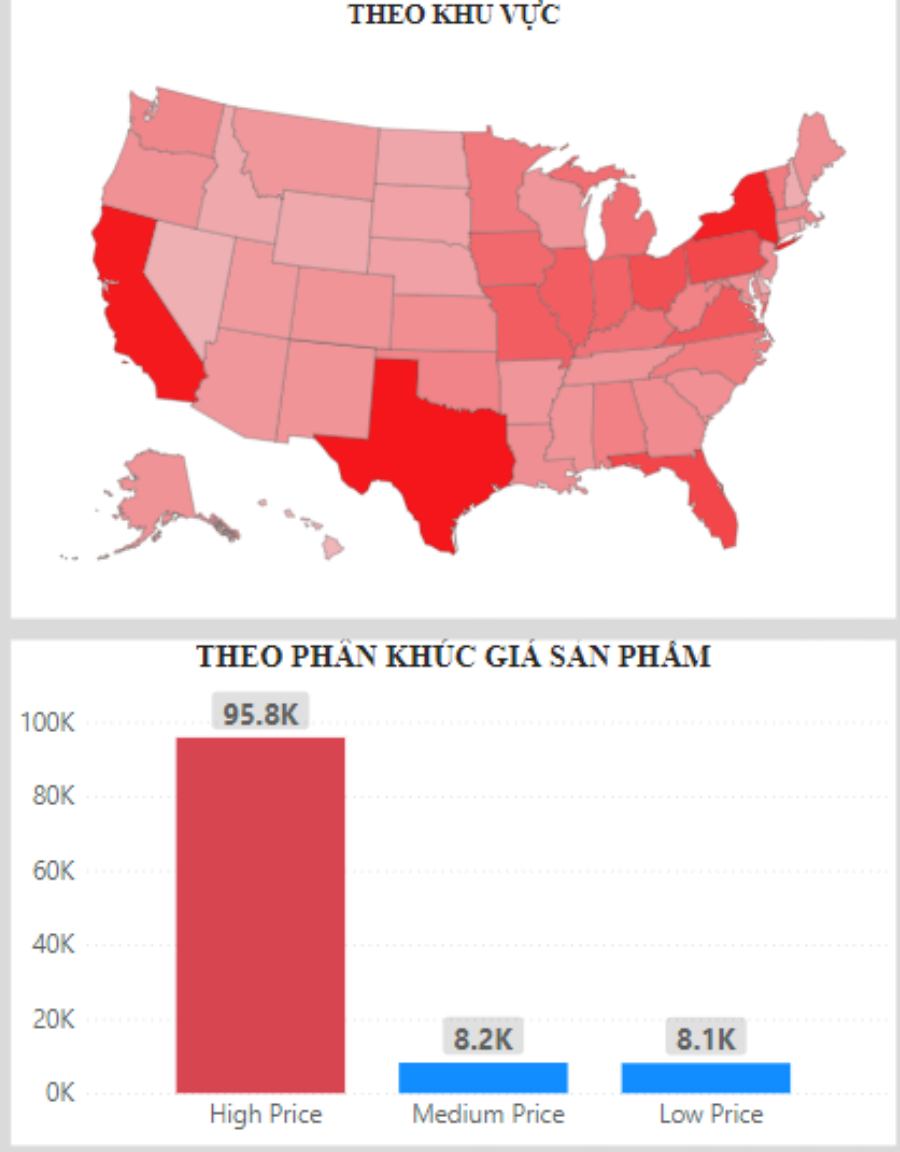
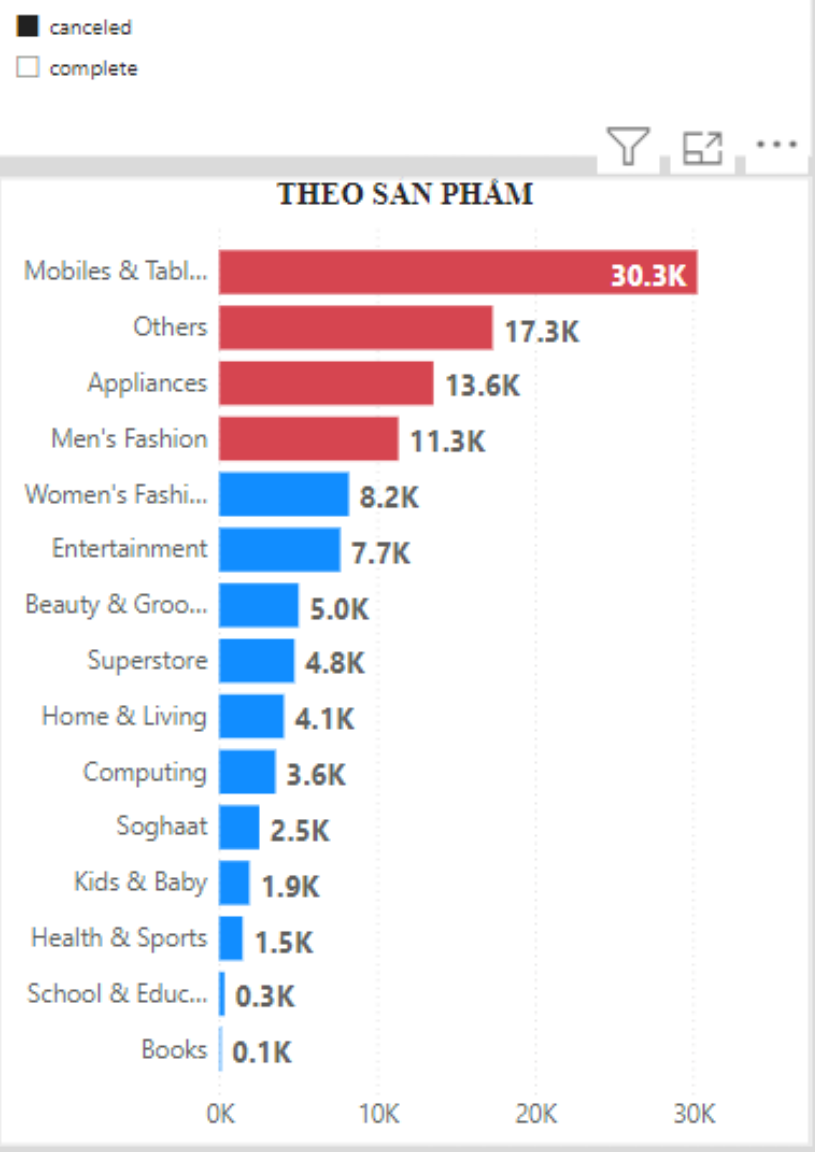
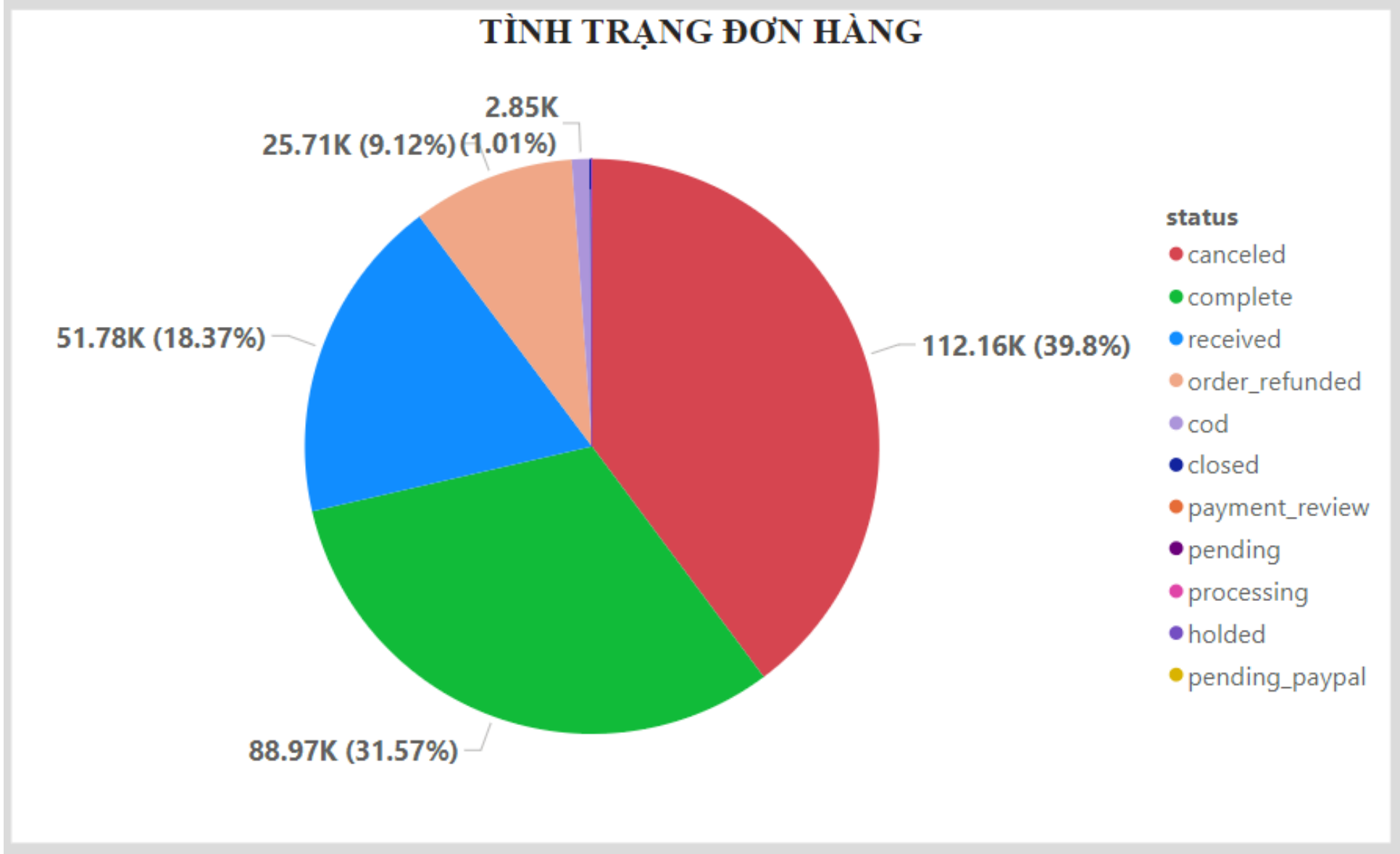
3. 3. Phân tích tình trạng đơn hàng

**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%)

Mobiles & Tablets, Appliances và **các mặt hàng quần áo**.

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



Thúc đẩy sự tương tác và phản hồi của khách hàng



Xem xét cơ chế thanh toán qua các kênh

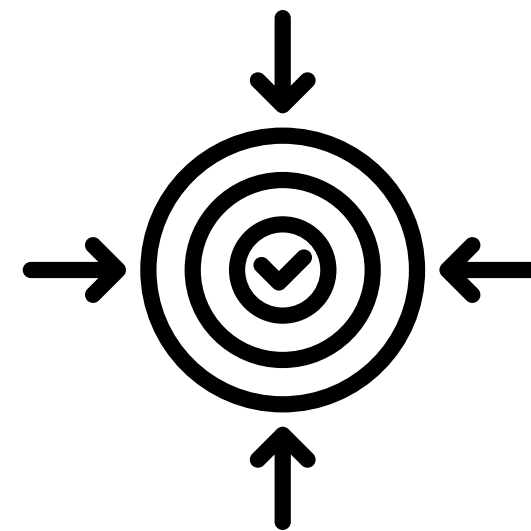


Tạo ra các chương trình cho khách hàng khi hoàn tất nhận hàng.





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

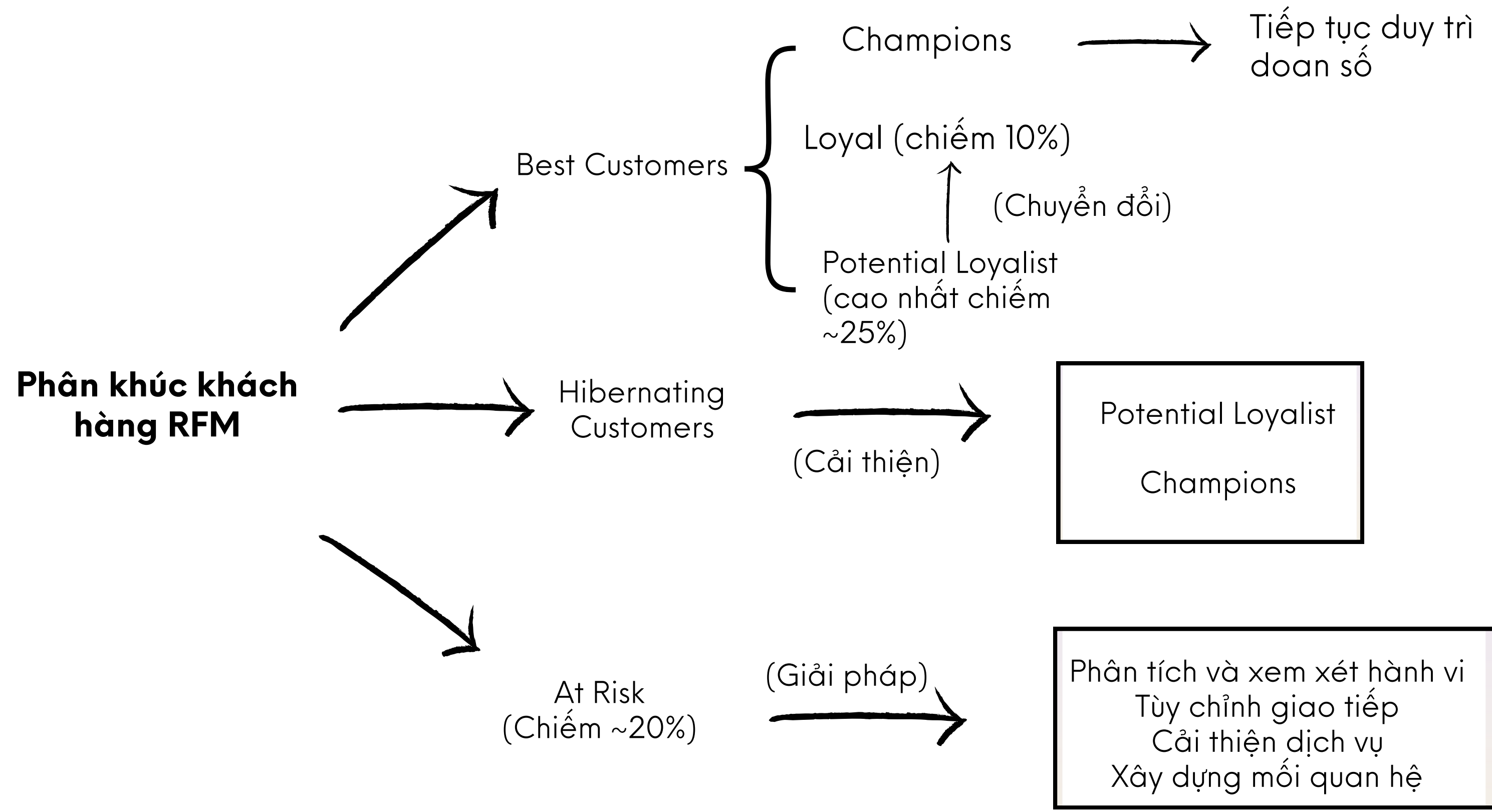
- Doanh thu chính là các sản phẩm thiết bị, công nghệ
- Số lượng đơn hàng quần áo nam nữ tăng cao

- 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

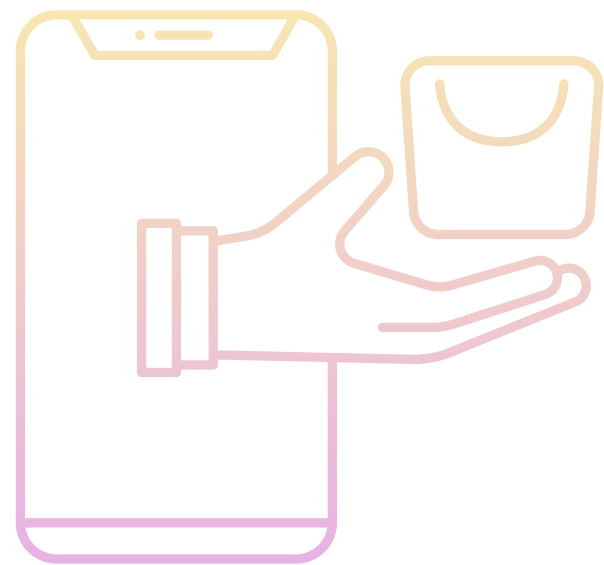


◆ IV. KẾT LUẬN VÀ ĐỀ XUẤT





**Thank
For
Your
Watching!**



NHÓM 1

E-Commerce Analysis

