

DATACO SMART SUPPLY CHAIN

Group 2 - Mentor Vi Thai

1. Nguyễn Ngọc Bảo Châu
2. Phạm Nguyên Chương
3. Nguyễn Thị Thu Hà
4. Trần Tiến Sang



MỤC LỤC

1. TỔNG QUAN

- 1.1 Giới thiệu Dataset
- 1.2 Các hướng phân tích

2. PHÂN TÍCH KHÁM PHÁ DỮ LIỆU

- 2.1 Công cụ xử lý dữ liệu
- 2.2 Tiền xử lý dữ liệu
- 2.3 Khám phá dữ liệu

3. PHÂN TÍCH DỮ LIỆU

- 3.1 Tổng quan tình hình kinh doanh của công ty
- 3.2 Phân tích RFM, tỷ lệ giao hàng chậm trễ, các loại hình giao hàng,..

4. KẾT LUẬN VÀ ĐỀ XUẤT



01

TỔNG QUAN



1. TỔNG QUAN

1.1 Giới thiệu về dataset

Dataset dựa trên số liệu của công ty DataCo Smart được đặt tại Mỹ, gồm 53 cột và 180519 dòng. Công ty kinh doanh đa dạng mặt hàng với hơn 50 danh mục hàng hóa như đồ thể thao, đồ điện tử, quần áo, đồng hồ, sách,... Công ty kinh doanh đa quốc gia và có hơn 150 mạng lưới vận chuyển trên toàn thế giới.

Thời gian dữ liệu từ 01/01/2015 - 28/02/2018

Days_for_shipping_(real):

Tổng số ngày theo thực tế

Days_for_shipment_(scheduled):

Tổng số ngày theo dự kiến

Delivery_Status:

Trạng thái đơn hàng

Late_delivery_risk

Đúng hẹn/ Trễ hẹn

Category_Id:

Mã danh mục sản phẩm

Category_Name:

Tên danh mục sản phẩm

Customer_City:

Thành phố nơi khách hàng

Customer_Fname:

Tên khách hàng

Customer_Id:

Mã khách hàng

Customer_Segment:

Phân khúc khách hàng

Customer_State:

Tiểu bang của khách hàng

1. TỔNG QUAN

1.1 Giới thiệu về dataset

Order_City

Thành phố đơn hàng gửi đến

Order_Country

Quốc gia đơn hàng gửi đến

Order_date_(DateOrders)

Ngày đặt hàng

Order_Id

Mã đơn hàng

Order_Item_Discount_Rate

Tỷ lệ chiết khấu từng sản phẩm

Order_Item_Discount

Tổng giá trị chiết khấu

Order_Item_Quantity

Số lượng sản phẩm

Sales

Doanh thu bán hàng

Order_Item_Total

Doanh thu sau chiết khấu

Market

Thị trường

Order_Item_Id

Mã sản phẩm

Order_Item_Product_Price

Đơn giá từng sản phẩm

Product_Name

Tên sản phẩm

shipping_date_(DateOrders)

Ngày nhận hàng

Shipping_Mode

Phương thức vận chuyển

Type

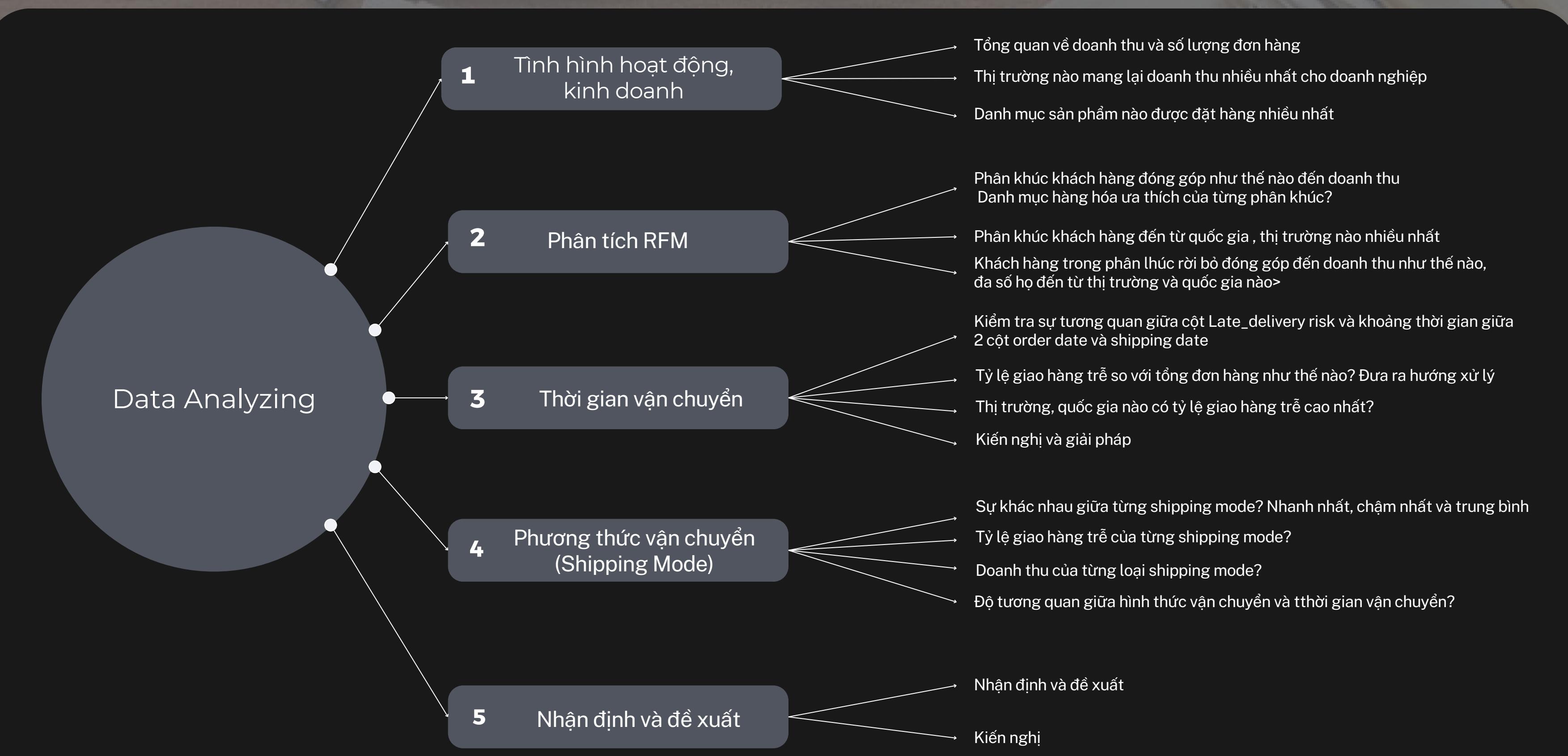
Phương thức thanh toán

Logistics



1. TỔNG QUAN

1.2 Các hướng phân tích



[02]

PHÂN TÍCH KHÁM PHÁ DỮ LIỆU

2. PHÂN TÍCH KHÁM PHÁ DỮ LIỆU

2.1 Công cụ xử lý data



2.2 Tiền xử lý dữ liệu

2.2.1 EDA data

df.info()

0 Type	180519	non-null	object	180519	non-null	object	180519	non-null	object	
1 Days for shipping (real)	180519	non-null	int64	27 Order Customer Id	180519	non-null	int64	180519	non-null	int64
2 Days for shipment (scheduled)	180519	non-null	int64	28 order date (DateOrders)	180519	non-null	object	180519	non-null	object
3 Benefit per order	180519	non-null	float64	29 Order Id	180519	non-null	int64	180519	non-null	int64
4 Sales per customer	180519	non-null	float64	30 Order Item Cardprod Id	180519	non-null	int64	180519	non-null	float64
5 Delivery Status	180519	non-null	object	31 Order Item Discount	180519	non-null	float64	180519	non-null	float64
6 Late_delivery_risk	180519	non-null	int64	32 Order Item Discount Rate	180519	non-null	float64	180519	non-null	float64
7 Category Id	180519	non-null	int64	33 Order Item Id	180519	non-null	int64	180519	non-null	int64
8 Category Name	180519	non-null	object	34 Order Item Product Price	180519	non-null	float64	180519	non-null	float64
9 Customer City	180519	non-null	object	35 Order Item Profit Ratio	180519	non-null	float64	180519	non-null	float64
10 Customer Country	180519	non-null	object	36 Order Item Quantity	180519	non-null	int64	180519	non-null	int64
11 Customer Email	180519	non-null	object	37 Sales	180519	non-null	float64	180519	non-null	float64
12 Customer Fname	180519	non-null	object	38 Order Item Total	180519	non-null	float64	180519	non-null	float64
13 Customer Id	180519	non-null	int64	39 Order Profit Per Order	180519	non-null	float64	180519	non-null	float64
14 Customer Lname	180511	non-null	object	40 Order Region	180519	non-null	object	180519	non-null	object
15 Customer Password	180519	non-null	object	41 Order State	180519	non-null	object	180519	non-null	object
16 Customer Segment	180519	non-null	object	42 Order Status	180519	non-null	object	180519	non-null	object
17 Customer State	180519	non-null	object	43 Order Zipcode	24840	non-null	float64	24840	non-null	float64
18 Customer Street	180519	non-null	object	44 Product Card Id	180519	non-null	int64	180519	non-null	int64
19 Customer Zipcode	180516	non-null	float64	45 Product Category Id	180519	non-null	int64	180519	non-null	int64
20 Department Id	180519	non-null	int64	46 Product Description	0	non-null	float64	0	non-null	float64
21 Department Name	180519	non-null	object	47 Product Image	180519	non-null	object	180519	non-null	object
22 Latitude	180519	non-null	float64	48 Product Name	180519	non-null	object	180519	non-null	float64
23 Longitude	180519	non-null	float64	49 Product Price	180519	non-null	float64	180519	non-null	int64
24 Market	180519	non-null	object	50 Product Status	180519	non-null	int64	180519	non-null	object
25 Order City	180519	non-null	object	51 shipping date (DateOrders)	180519	non-null	object	180519	non-null	object
				52 Shipping Mode	180519	non-null	object			

- Kiểm tra dữ liệu

Customer Country
['Puerto Rico' 'EE. UU.']
xxxx-----xxxx-----xxxx
Customer Email
['XXXXXXXXXX']
xxxx-----xxxx-----xxxx

Customer Password
['XXXXXXXXXX']
Product Description
[nan]

Kiểm tra các cột dữ liệu
không đóng góp vào
bảng dữ liệu.
XÓA

2.2 Tiết xử lý dữ liệu

2.2.1 EDA data

- Định dạng các cột order date (DateOrders) và shipping date (DateOrders)

```
## Convert all date columns to datetime type
df['order date (DateOrders)'] = df['order date (DateOrders)'].str.split(" ", expand=True)[0]
df['shipping date (DateOrders)'] = df['shipping date (DateOrders)'].str.split(" ", expand=True)[0]
```

- Xóa các cột Product Description, Customer Email, Customer Password, Customer Country, Order Zipcode

```
## Delete columns that consist of no values Product Description, Customer Email, Customer Password, Customer Country
df = df.drop(columns=['Product Description', 'Customer Email', 'Customer Password', 'Customer Country', 'Order Zipcode'])
```

- Kiểm tra và điền các dòng còn thiếu dữ liệu

```
## Fill in Customer Lname, Customer Zipcode
df['Customer Lname'] = df['Customer Lname'].fillna(value = "No Lname")

## Check the Customer City to fillna in Customer Zipcode
df[df['Customer Zipcode'].isnull()] ['Customer City'] ## Kết quả trả ra toàn bộ zipcode N/A là CA
## Fillna in Customer Zipcode
df ['Customer Zipcode'] = df ['Customer Zipcode'] .fillna(value='95758')

## Change the number in Customer State to right location
df['Customer State'] = df['Customer State'].replace(['95758', '91732'], 'CA')
# df['Customer Zipcode'] = df['Customer Zipcode'].replace({'95758': 'CA', '91732': 'CA'})
```

2.2 Tiên xử lý dữ liệu

2.2.1 EDA data

- Đổi khoảng cách (space) các cột thành dấu gạch dưới

```
## Change all space in columns to underscore (_)
df.columns = df.columns.str.replace(' ', '_')
```

- Đổi giá trị “Same day” trong cột Shipping mode thành giá trị “Priority” do có thời gian giao hàng trong này

```
## Due to days in delivery = 0, as always shipping mode is same day, so we change the shipping mode
df['Shipping_Mode'] = df['Shipping_Mode'].str.replace("Same Day","Priority")
```

- Xóa các dòng dữ liệu có chứa giá trị Shipping cancel trong cột Delivery Status để phân tích chính xác về lượng khách hàng, doanh thu

```
## Delete rows where the value in the "Delivery_Status" column is "Shipping canceled",
## and the payment type column has the signal of 'SUSPECTED_FRAUD' or 'CANCELED'
df = df.drop(df[df['Delivery_Status'] == 'Shipping canceled'].index)
```

2.2 Tiên xử lý dữ liệu

2.2.1 EDA data

- Các tên quốc gia trong cột Order Country là Tiếng Tây Ban Nha, đổi toàn bộ dữ liệu Quốc gia sang Tiếng Anh.

```
## Change the Order Country in Spanish to English
translation_dict = {
    'Indonesia': 'Indonesia', 'India': 'India', 'Australia': 'Australia', 'China': 'China',
    'Corea del Sur': 'South Korea', 'Singapur': 'Singapore', 'Turquía': 'Turkey',
    'Mongolia': 'Mongolia', 'Estados Unidos': 'United States', 'Nigeria': 'Nigeria',
    'República Democrática del Congo': 'Democratic Republic of the Congo', 'Senegal': 'Senegal',
    'Marruecos': 'Morocco', 'Alemania': 'Germany', 'Francia': 'France', 'Países Bajos': 'Netherlands',
    'Reino Unido': 'United Kingdom', 'Guatemala': 'Guatemala', 'El Salvador': 'El Salvador',
    'Panamá': 'Panama', 'República Dominicana': 'Dominican Republic', 'Venezuela': 'Venezuela',
    'Colombia': 'Colombia', 'Honduras': 'Honduras', 'Brasil': 'Brazil', 'México': 'Mexico',
    'Uruguay': 'Uruguay', 'Argentina': 'Argentina', 'Cuba': 'Cuba', 'Perú': 'Peru',
    'Nicaragua': 'Nicaragua', 'Ecuador': 'Ecuador', 'Angola': 'Angola', 'Sudán': 'Sudan',
    'Somalia': 'Somalia', 'Costa de Marfil': 'Ivory Coast', 'Egipto': 'Egypt', 'Italia': 'Italy',
    'España': 'Spain', 'Suecia': 'Sweden', 'Austria': 'Austria', 'Zambia': 'Zambia', 'Níger': 'Niger',
    'SudAfrica': 'South Africa', 'Mozambique': 'Mozambique', 'Tanzania': 'Tanzania',
    'Ruanda': 'Rwanda', 'Israel': 'Israel', 'Nueva Zelanda': 'New Zealand', 'Bangladés': 'Bangladesh',
    'Tailandia': 'Thailand', 'Irak': 'Iraq', 'Arabia Saudí': 'Saudi Arabia', 'Filipinas': 'Philippines',
    'Kazajistán': 'Kazakhstan', 'Irán': 'Iran', 'Myanmar (Birmania)': 'Myanmar (Burma)',
    'Uzbekistán': 'Uzbekistan', 'Benín': 'Benin', 'Camerún': 'Cameroon', 'Kenia': 'Kenya',
    'Qatar': 'Qatar', 'Etiopia': 'Ethiopia', 'Mali': 'Mali', 'Albania': 'Albania',
    'República del Congo': 'Republic of the Congo', 'Bosnia y Herzegovina': 'Bosnia and Herzegovina',
    'Omán': 'Oman', 'Túnez': 'Tunisia', 'Sierra Leona': 'Sierra Leone', 'Yibuti': 'Djibouti',
    'Burundi': 'Burundi', 'Montenegro': 'Montenegro', 'Gabón': 'Gabon', 'Sudán del Sur': 'South Sudan',
    'Luxemburgo': 'Luxembourg', 'Liberia': 'Liberia', 'Namibia': 'Namibia', 'Mauritania': 'Mauritania',
    'Grecia': 'Greece', 'Suazilandia': 'Eswatini', 'Guyana': 'Guyana', 'Guayana Francesa': 'French Guiana',
    'República Centroafricana': 'Central African Republic', 'Taiwán': 'Taiwan', 'Estonia': 'Estonia',
    'Líbano': 'Lebanon', 'Chipre': 'Cyprus', 'Guinea-Bissau': 'Guinea-Bissau', 'Surinam': 'Suriname',
    'Belice': 'Belize', 'Eslovenia': 'Slovenia', 'República de Gambia': 'Republic of the Gambia',
    'Botswana': 'Botswana', 'Armenia': 'Armenia', 'Guinea Ecuatorial': 'Equatorial Guinea',
    'Kuwait': 'Kuwait', 'Bután': 'Bhutan', 'Chad': 'Chad', 'Serbia': 'Serbia', 'Sáhara Occidental': 'Western Sahara'}
df['Order_Country'] = df['Order_Country'].map(translation_dict)
```

2.2 Tiên xử lý dữ liệu

2.2.2 Final Dataset

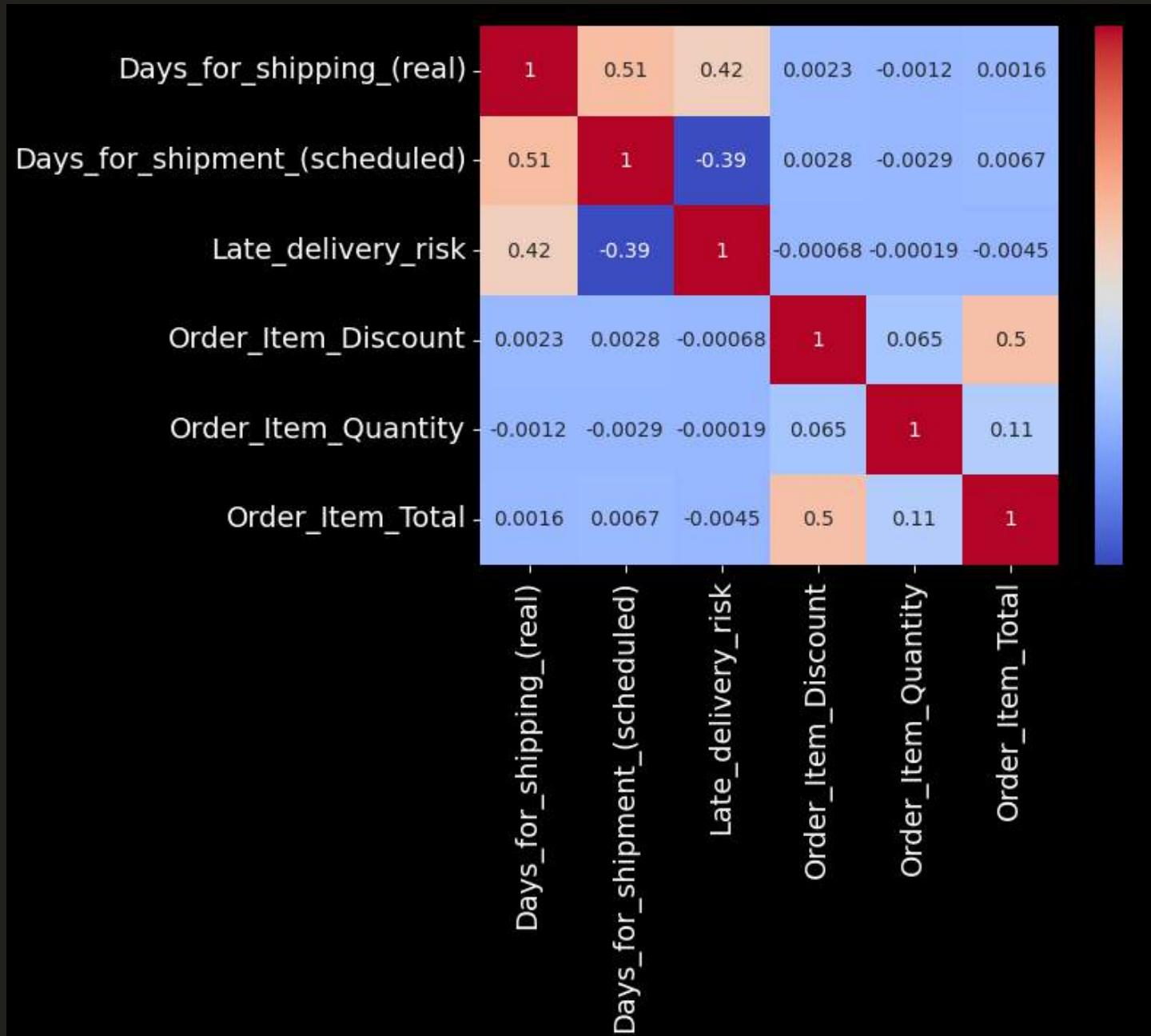
Bảng dữ liệu còn 48 cột và 172765 dòng

0	Type	172765	non-null	object
1	Days_for_shipping_(real)	172765	non-null	int64
2	Days_for_shipment_(scheduled)	172765	non-null	int64
3	Benefit_per_order	172765	non-null	float64
4	Sales_per_customer	172765	non-null	float64
5	Delivery_Status	172765	non-null	object
6	Late_delivery_risk	172765	non-null	int64
7	Category_Id	172765	non-null	int32
8	Category_Name	172765	non-null	object
9	Customer_City	172765	non-null	object
10	Customer_Fname	172765	non-null	object
11	Customer_Id	172765	non-null	int32
12	Customer_Lname	172765	non-null	object
13	Customer_Segment	172765	non-null	object
14	Customer_State	172765	non-null	object
15	Customer_Street	172765	non-null	object
16	Customer_Zipcode	172765	non-null	int32
17	Department_Id	172765	non-null	int32
18	Department_Name	172765	non-null	object
19	Latitude	172765	non-null	float64
20	Longitude	172765	non-null	float64
21	Market	172765	non-null	object
22	Order_City	172765	non-null	object
23	Order_Country	172765	non-null	object
24	Order_Customer_Id	172765	non-null	int32
25	order_date_(DateOrders)	172765	non-null	object
26	Order_Id	172765	non-null	int32
27	Order_Item_Cardprod_Id	172765	non-null	int32
28	Order_Item_Discount	172765	non-null	float64
29	Order_Item_Discount_Rate	172765	non-null	float64
30	Order_Item_Id	172765	non-null	int64
31	Order_Item_Product_Price	172765	non-null	float64
32	Order_Item_Profit_Ratio	172765	non-null	float64
33	Order_Item_Quantity	172765	non-null	int64
34	Sales	172765	non-null	float64
35	Order_Item_Total	172765	non-null	float64
36	Order_Profit_Per_Order	172765	non-null	float64
37	Order_Region	172765	non-null	object
38	Order_State	172765	non-null	object
39	Order_Status	172765	non-null	object
40	Product_Card_Id	172765	non-null	int32
41	Product_Category_Id	172765	non-null	int32
42	Product_Image	172765	non-null	object
43	Product_Name	172765	non-null	object
44	Product_Price	172765	non-null	float64
45	Product_Status	172765	non-null	int64
46	shipping_date_(DateOrders)	172765	non-null	object
47	Shipping_Mode	172765	non-null	object

2.3 Khám phá dữ liệu

2.3.2 Kiểm tra tính tương quan giữa các cột có giá trị số

```
## Check the correlation between delivery days, late risk, total quantity, discount rate
correlation = df[['Days_for_shipping_(real)', 'Days_for_shipment_(scheduled)', 'Late_delivery_risk', 'Order_Item_Discount', 'Order_Item_Quantity', 'Order_Item_Total']]
sns.heatmap(correlation,cmap='coolwarm', annot=True)
plt.gca().tick_params(axis='x', colors='white', labelsize=14)
plt.gca().tick_params(axis='y', colors='white', labelsize=14)
plt.gca().yaxis.label.set_color('white')
plt.gca().xaxis.label.set_color('white')
plt.gcf().patch.set_facecolor('black')
```



- **Độ trễ giao hàng (-) thời gian giao hàng dự kiến**

Thời gian giao hàng dự kiến càng tăng thì độ giao hàng trễ sẽ giảm.

- **Độ giao hàng trễ (+) thời gian giao hàng thực tế.**

Thời gian giao hàng thực tăng đồng nghĩa với việc góp phần làm tăng tỷ lệ giao hàng trễ.

- **Tỷ lệ giảm giá (+) tổng doanh thu bán hàng**

Việc tăng tỷ lệ giảm giá hoặc chương trình khuyến mại giúp làm tăng doanh số bán hàng.

[03]

PHÂN TÍCH DỮ LIỆU

3. PHÂN TÍCH DỮ LIỆU

3.1 Phân tích tình hình hoạt động kinh doanh



Destination Order
164

Pháp, Đức, UK, Ý, Mexico, Brazil, Úc, Trung Quốc, Ấn Độ, Hoa Kỳ, Canada, Nigeria, Morocco,...

Market
5

Châu Âu, châu Á - Thái Bình Dương, Châu Phi, Hoa Kỳ và Canada (USCA) và Mỹ Latinh

Category Number
50

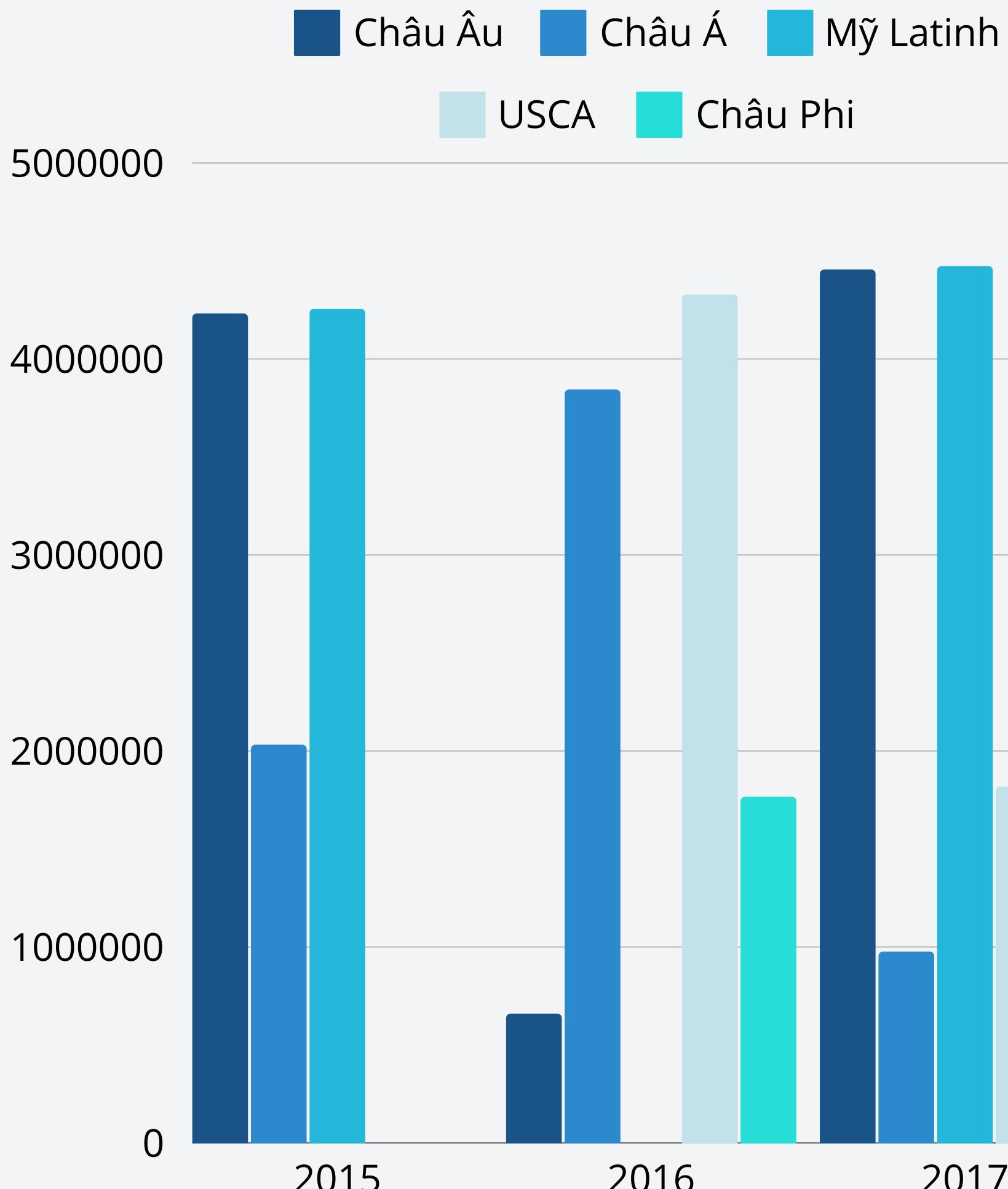
Fishing, Cleats, Camping & Hiking, Cardio Equipment,...

Customer Segment
3

Corporate, Home, Consumer

• USCA: Hoa Kỳ và Canada

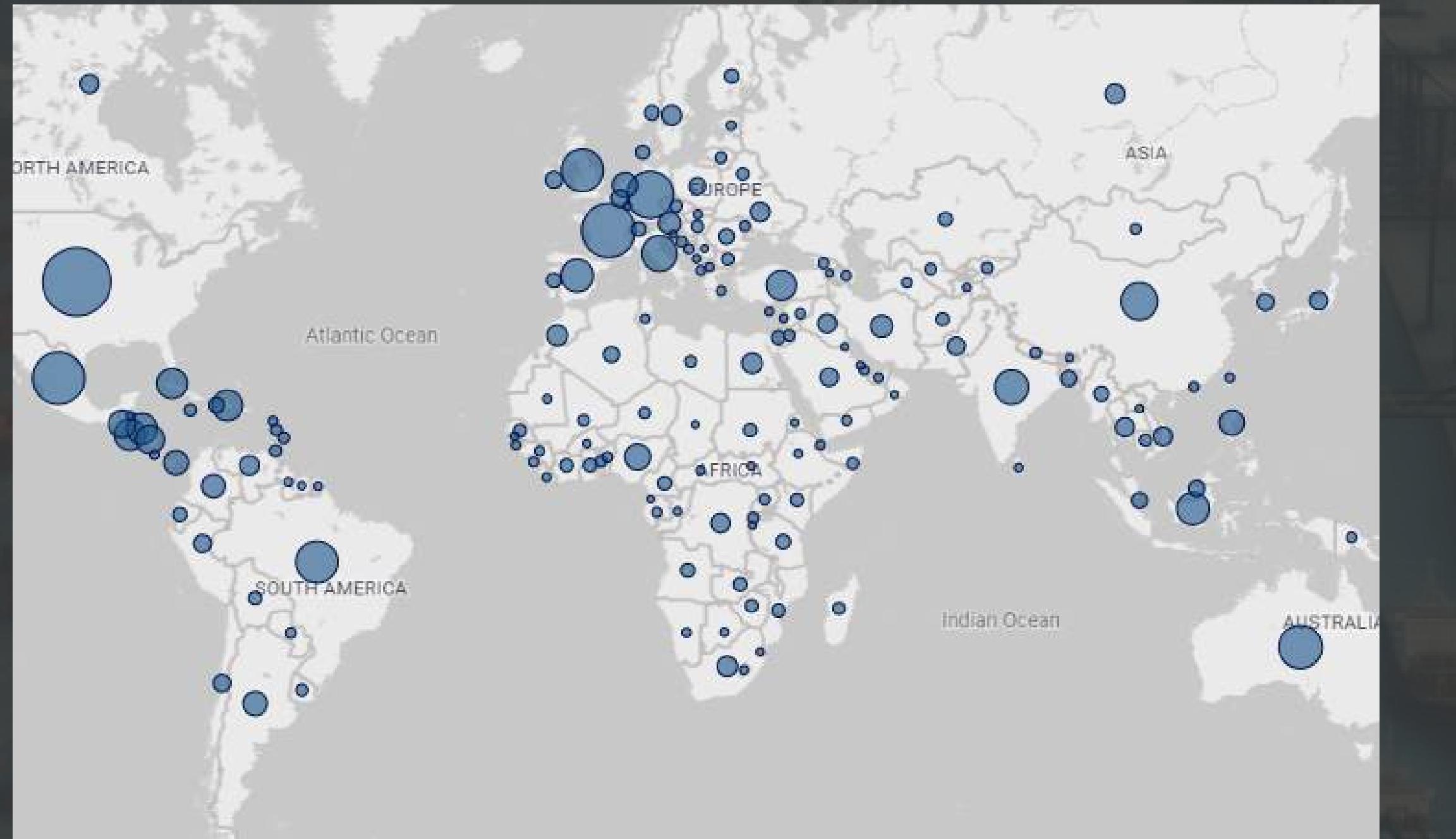
Doanh thu theo từng thị trường (Market)



3. PHÂN TÍCH DỮ LIỆU

3.1 Phân tích tình hình hoạt động kinh doanh

Doanh thu thường đến từ các quốc gia nào?



Hoạt động kinh doanh của công ty có mạng lưới phân phối trên toàn thế giới , trong đó có Việt Nam.



Đa số doanh thu tập trung nhiều ở Mỹ Latinh và châu Âu, vì vậy nên tập trung đẩy mạnh, xây dựng, hợp tác với các hàng hàng không để rút ngắn thời gian giao hàng.

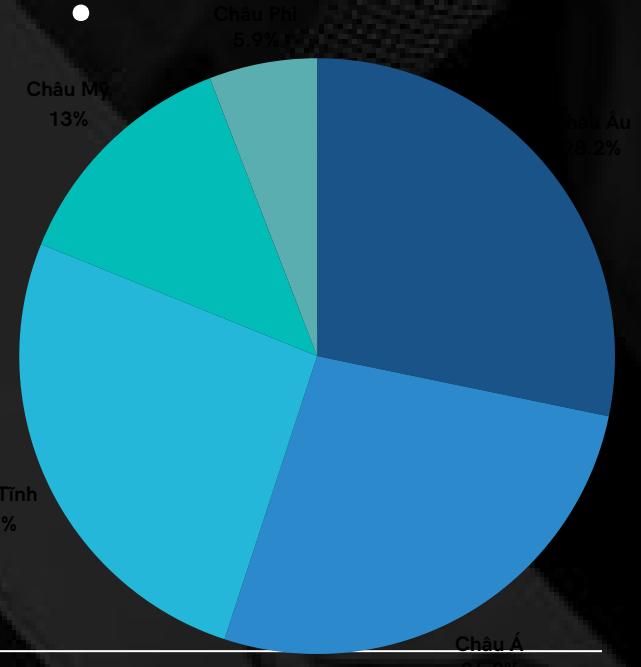


Châu Á - Thái Bình Dương là một thị trường năng động và đầy tiềm năng. Tuy nhiên, điểm giao hàng nằm rải rác ở nhiều nơi.

3.1 Phân tích tình hình hoạt động kinh doanh

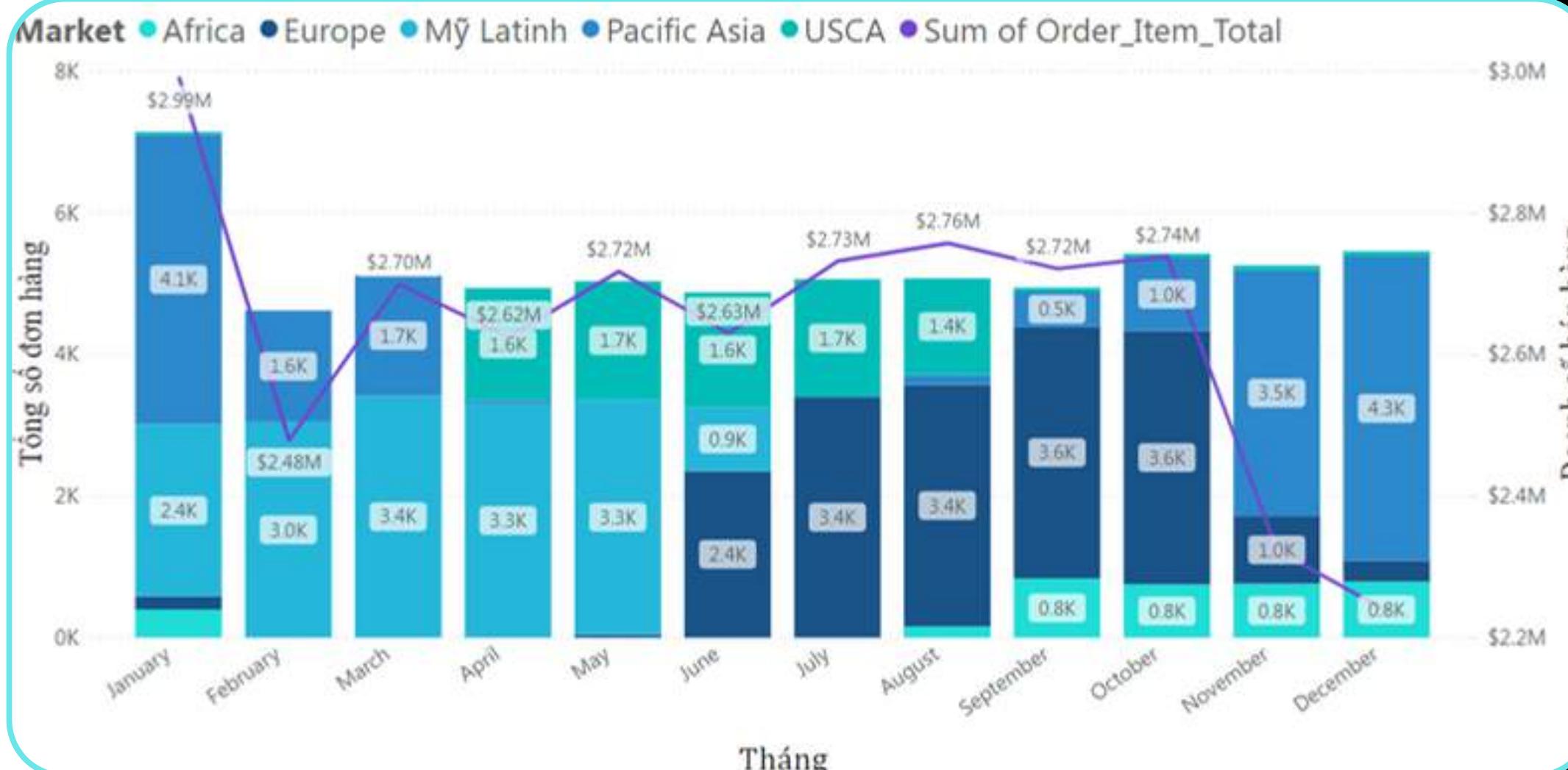
DOANH THU VÀ TỔNG SỐ ĐƠN HÀNG THEO THỊ TRƯỜNG

Total Customer Quantity in 3 years	Revenue in 3 years	Total Order Quantity in 3 years
20,261	31.36M	60,860



- USCA: Hoa Kỳ và Canada

Tổng đơn hàng theo từng tháng và thị trường



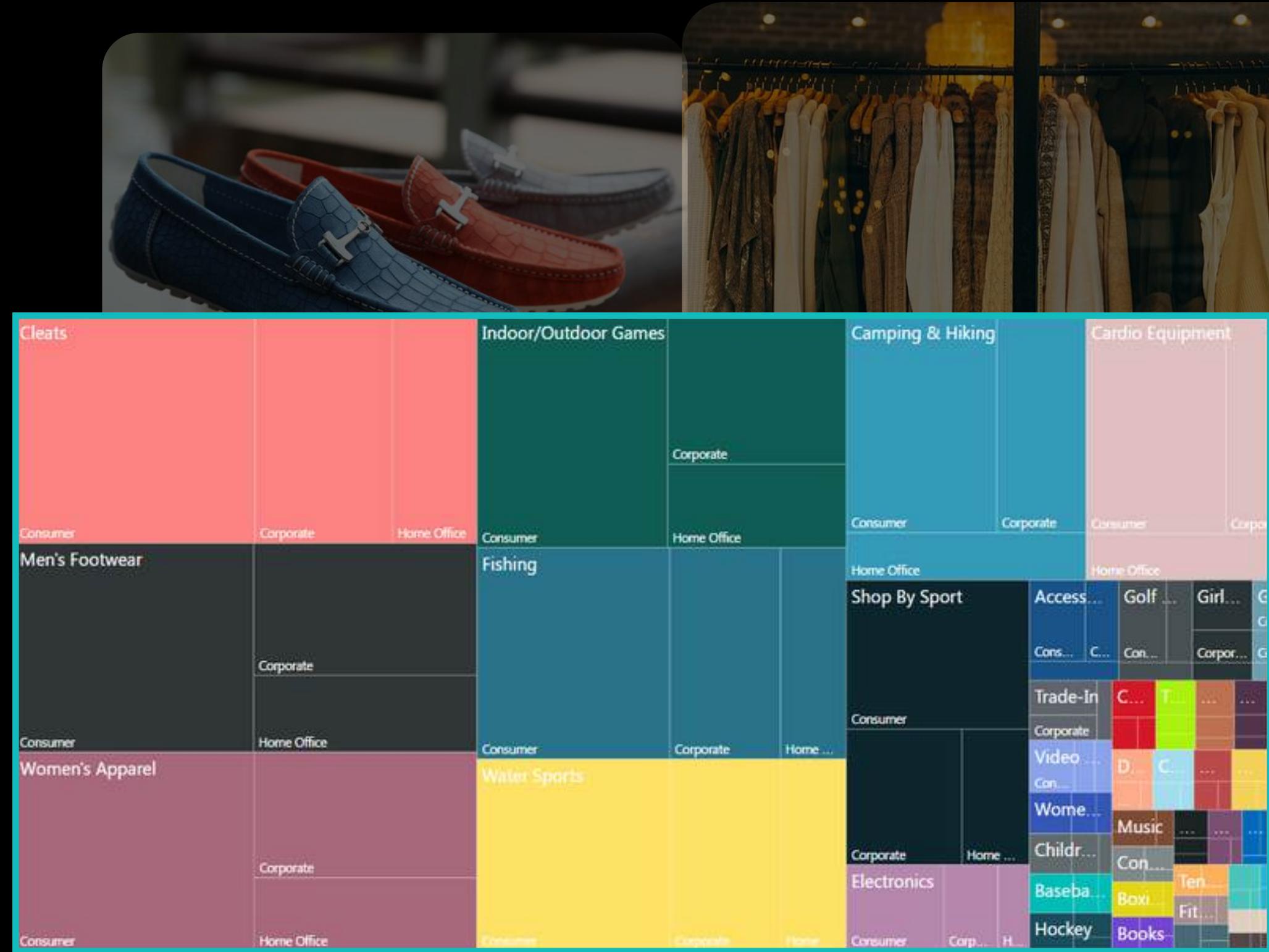
Các đơn hàng, vận chuyển được đặt theo từng thời điểm cụ thể như sau:

- **Châu Phi:** Từ tháng 9 đến tháng 1 năm sau.
- **Châu Âu:** Từ tháng 6 đến tháng 11 (**Cao điểm từ tháng 7 đến tháng 10**).
- **Mỹ Latinh:** Từ tháng 1 đến tháng 6 (**Cao điểm vào tháng 2 đến tháng 6**).
- **Châu Á:** Từ tháng 9 đến tháng 3 năm sau (**Cao điểm vào tháng 11 đến tháng 1**).

- **Hoa Kỳ và Canada:** Từ tháng 4 đến tháng 8 , các đơn hàng đồng đều qua các tháng.

Các danh mục hàng hóa được đặt nhiều nhất

- Cleats
- Men's Footwear
- Women's Apparel
- Indoor/Outdoor Games
- Fishing
- Water Sport
- Camping & Hiking



Bảng phân bố các hạng mục được đặt nhiều nhất

KẾT LUẬN

- Công ty đang phát triển tốt trong linh giai đoạn 2015-2017.
- Công ty có doanh thu và số lượng đơn hàng ổn định qua từng tháng trong 3 năm.
- Châu Á, Châu Âu và Mỹ Latinh là 3 châu lục có tổng đơn hàng và doanh thu **lớn nhất 80%** giá trị kinh doanh của công ty.
- Tháng 1 là tháng có đơn hàng cao nhất trong 3 năm
- Mỗi thị trường đều có thời gian đặt hàng nhất định. Tại từng thời điểm, công ty có thể đưa ra nhiều chính sách ưu đãi để thu hút khách hàng từ các khu vực.
- In-stock các mặt hàng được đặt nhiều nhất để kịp thời giao hàng đến khách hàng trong thời gian sớm.

ĐỀ XUẤT



Xây dựng chiến lược và phân bổ nhân sự theo từng quý để tối ưu và tập trung thời gian vào xử lý các đơn hàng thị trường trọng điểm theo bảng đề xuất dưới đây:

Quý 1	Quý 2	Quý 3	Quý 4
Châu Á	Mỹ Latinh	Châu Âu	Châu Á
Mỹ Latinh	USCA	USCA	Châu Phi, Châu Âu



Xây dựng hợp tác với các đối tác tại Châu Á - Thái Bình Dương để thành lập kho trung gian, rút ngắn thời gian, chi phí vận chuyển đến các khu vực xung quanh Châu Á.



In-stock các mặt hàng được đặt nhiều nhất để kịp thời giao hàng đến khách hàng trong thời gian sớm.



3. PHÂN TÍCH DỮ LIỆU

3.2 Phân tích RFM, tỷ lệ giao hàng chậm trễ, các loại hình giao hàng,..

PHÂN TÍCH RFM

PHÂN KHÚC KHÁCH HÀNG THEO RFM

Best Customer

- Champions
- Loyal
- Potential Loyalist

Need Attention

- Promising
- New Customer
- Need Attention

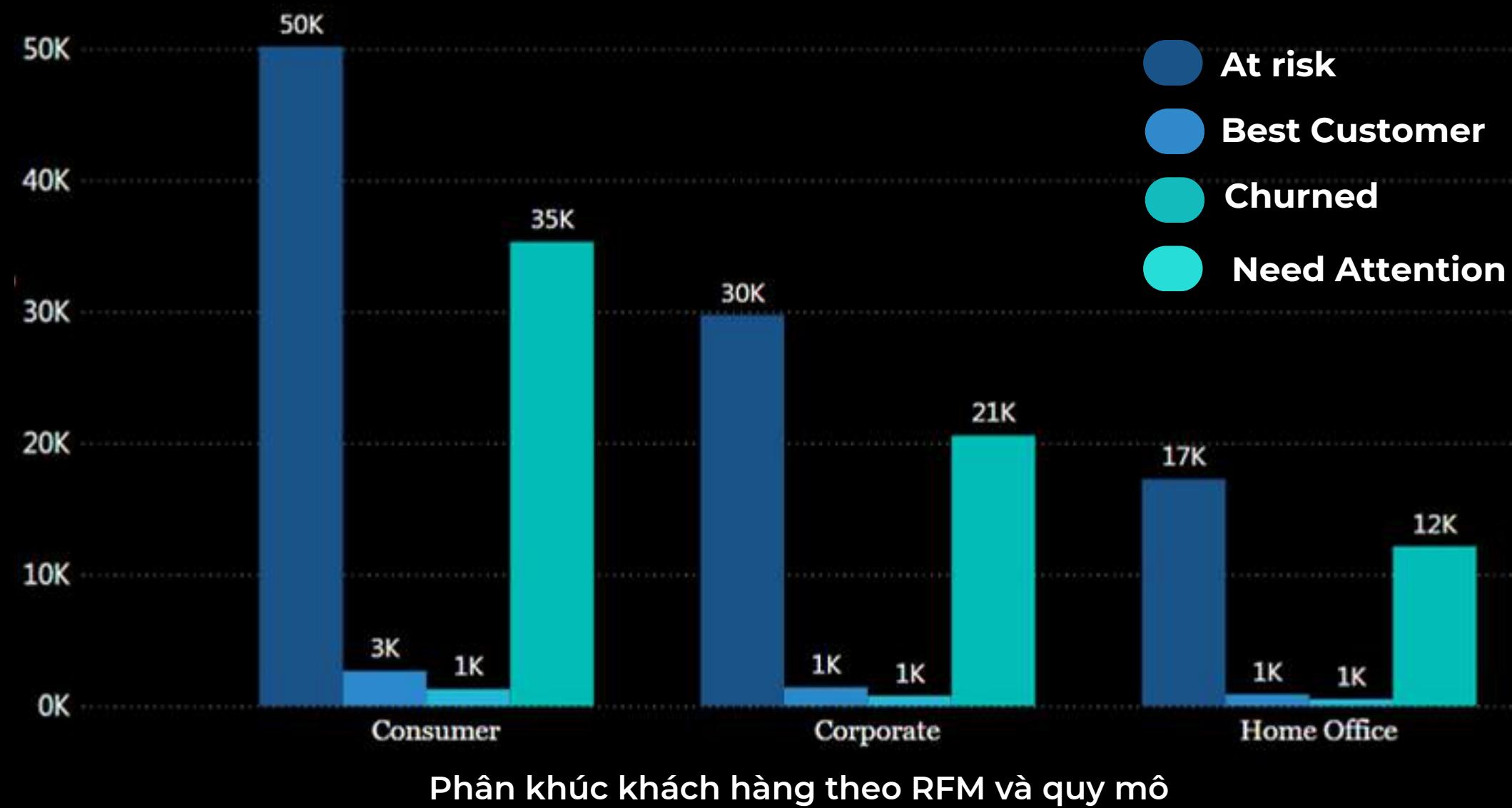
At Risk

- About to Sleep
- At risk
- Cannot Lose Them

Churned

- Hibernating customers
- Lost customers

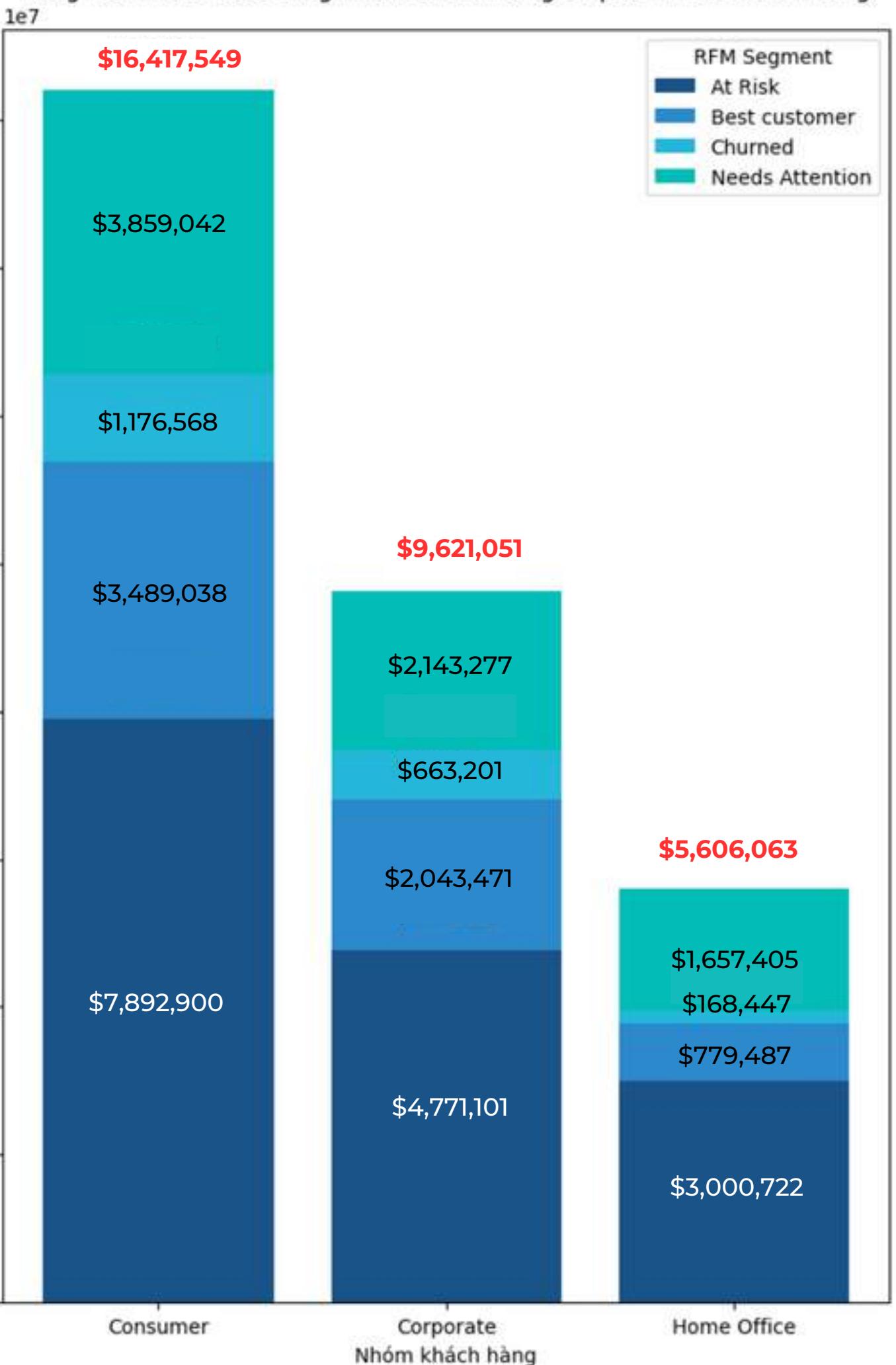
PHÂN NHÓM KHÁCH HÀNG THEO RFM VÀ QUY MÔ



Nhóm khách hàng **Consumer chiếm hơn 50% trên tổng số 20261 khách hàng**, với nhóm Corporate và Home Office đứng sau. Điều này đề xuất việc cần cải thiện mối quan hệ hợp tác với hai nhóm khách hàng này từ phía công ty.

Ngoài ra, sự xuất hiện của các phân khúc "needs attention" và "at risk" ở mức báo động là dấu hiệu rằng có nguy cơ cao về việc khách hàng rời bỏ, và do đó, cần phải xem xét các biện pháp khắc phục.

Tổng doanh thu theo từng nhóm khách hàng và phân khúc khách hàng



TỔNG QUAN VỀ DOANH THU THEO RFM

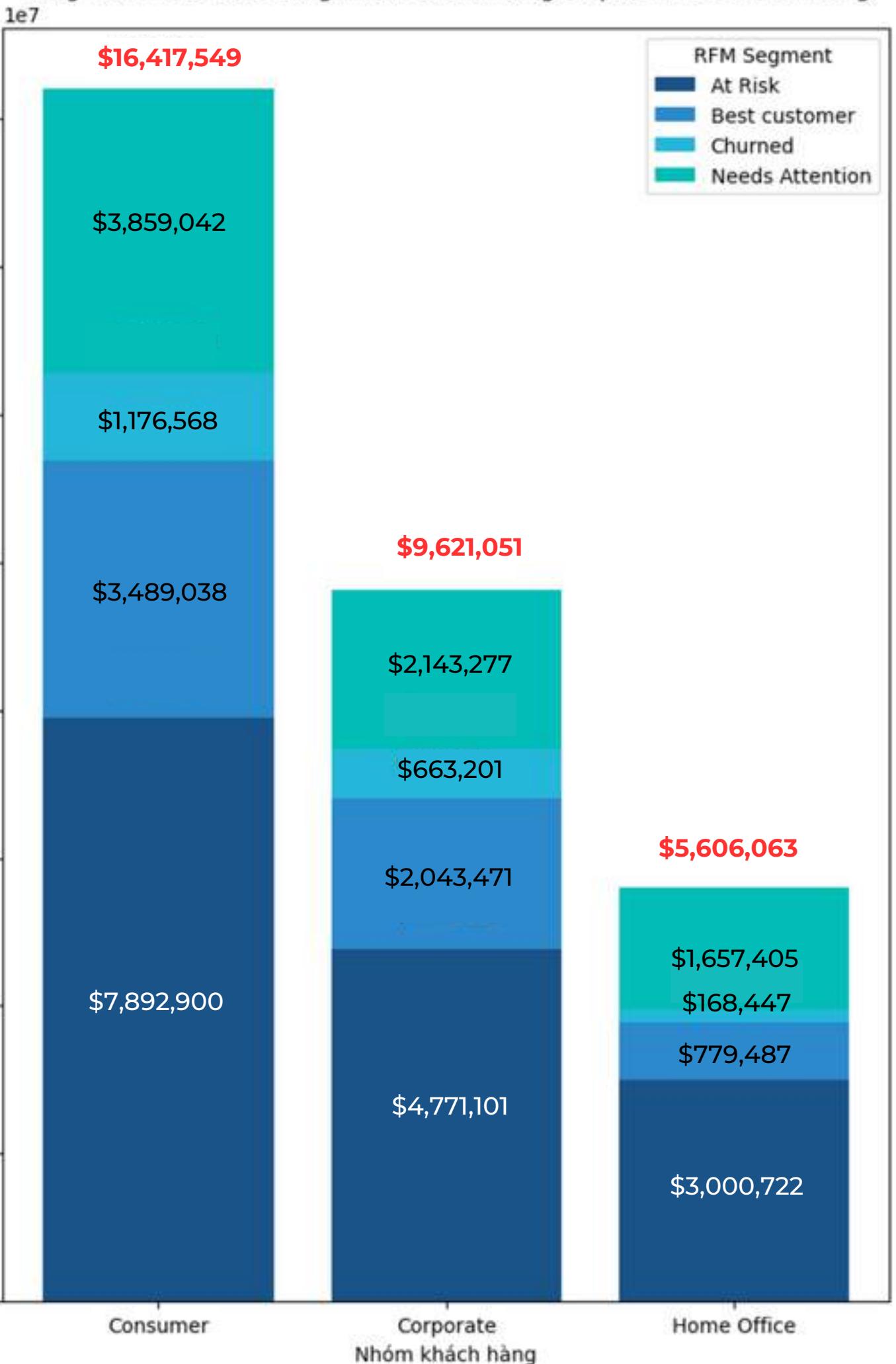
Nhận xét:

Consumer tạo ra doanh thu lớn nhất, đóng góp \$16,417,549. Tuy nhiên, có **\$11,381,938 (khoảng 69%)** đến từ phân khúc "**At Risk**" và "**Churned**". Điều này có thể ảnh hưởng đến sự tăng trưởng doanh thu trong tương lai.

Corporate tạo ra doanh thu lên đến **\$9,621,051**. Tuy nhiên, sự phụ thuộc quá nhiều vào khách hàng ở phân khúc "**At Risk**" là một **dấu hiệu đáng báo động (chiếm 50%)**.

Home Office tạo ra doanh thu **\$5,606,063**. Tuy nhiên, chỉ có \$779,487 (khoảng 17%) đến từ phân khúc "Best customer", trong khi **\$3,000,722 (khoảng 56%)** đến từ phân khúc "**At Risk**".





TỔNG QUAN VỀ DOANH THU THEO RFM

Đề xuất:

Xây Dựng Hệ Thống CRM Hiệu Quả:

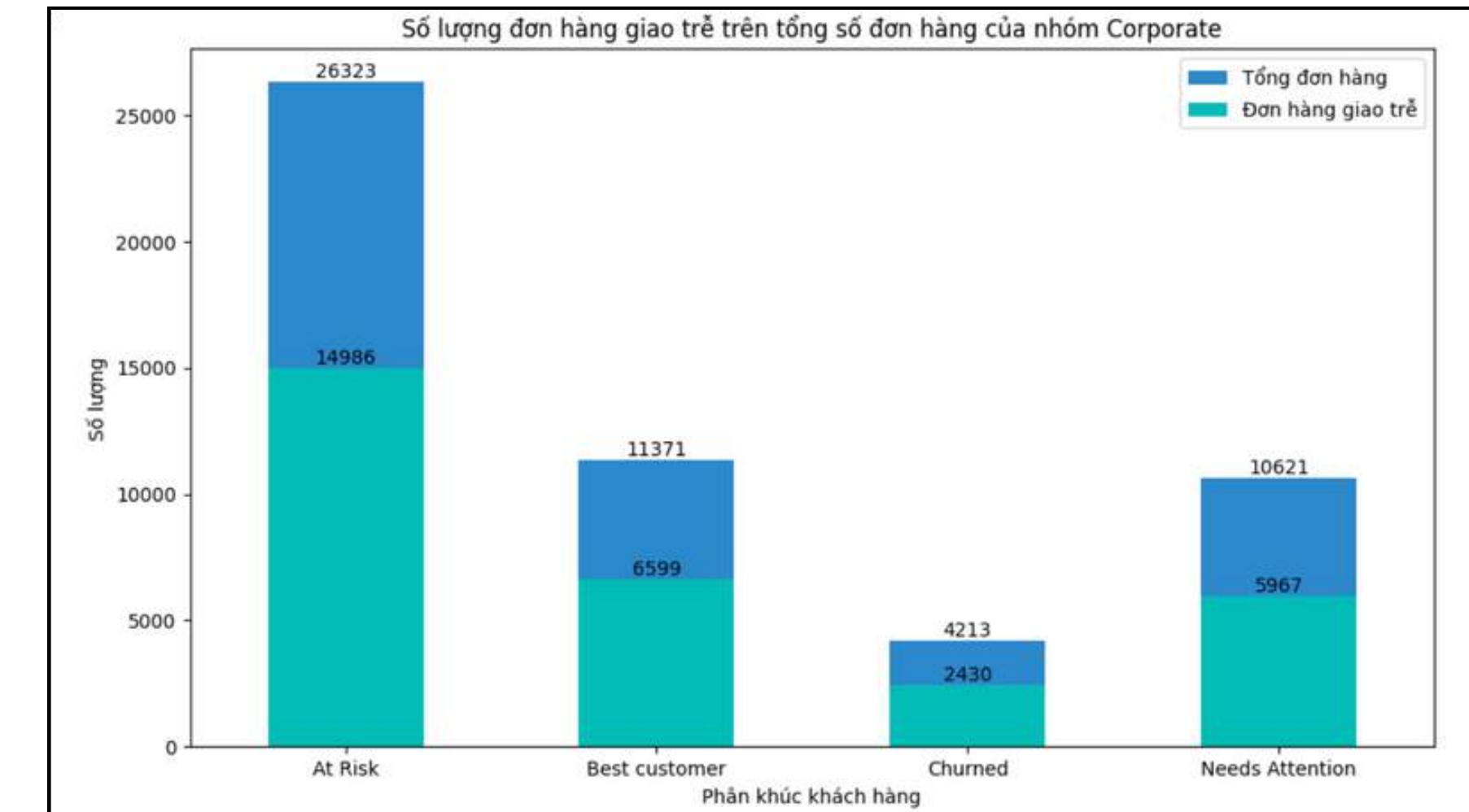
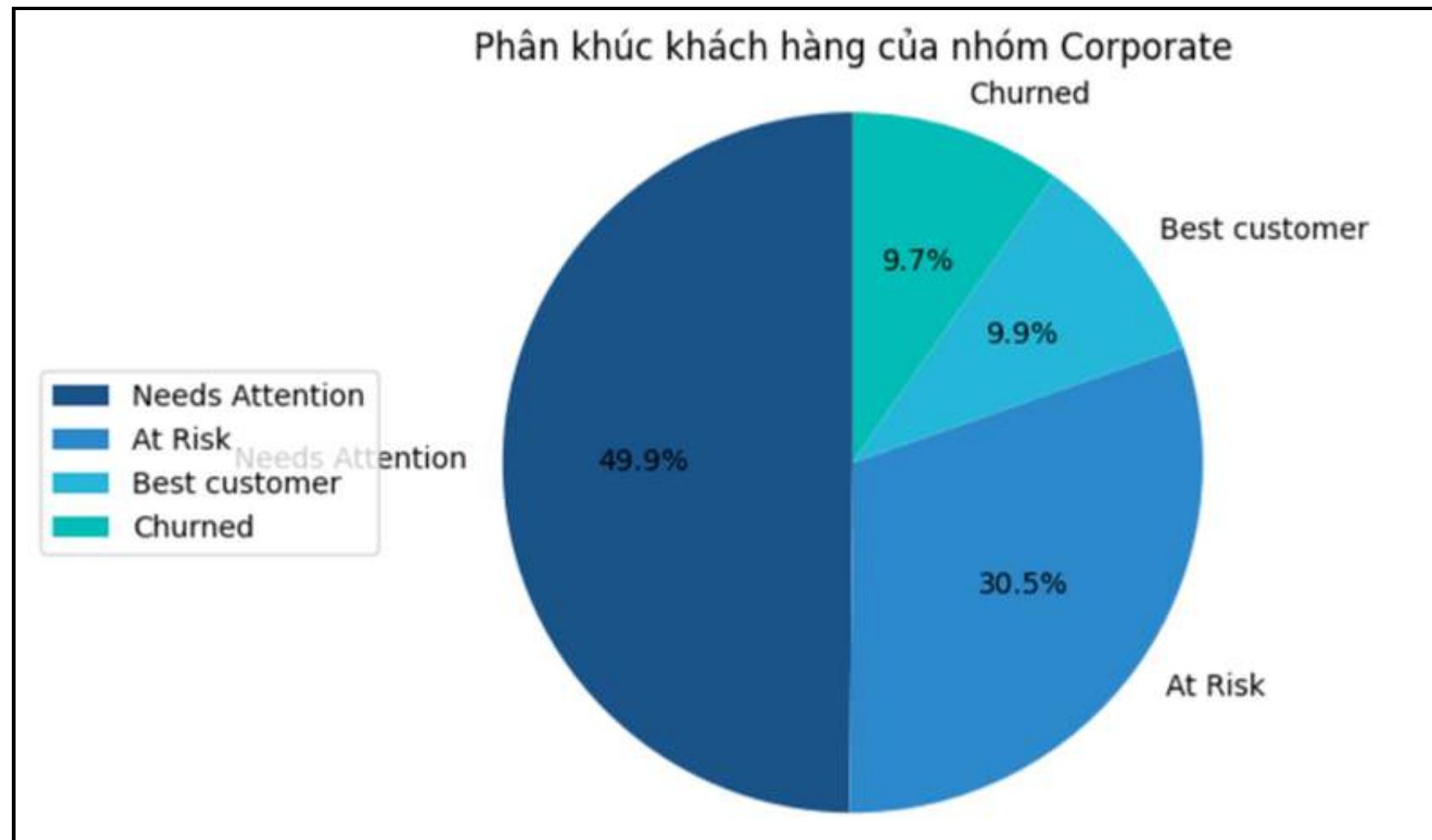
Triển khai hệ thống quản lý mối quan hệ khách hàng (CRM) tiên tiến để thu thập, phân tích và sử dụng dữ liệu khách hàng một cách hiệu quả. Từ đó, tăng cường khả năng hiểu biết về hành vi và nhu cầu của từng phân khúc khách hàng và **cá nhân hóa chiến lược tiếp thị và chăm sóc khách hàng**.

Tăng cường tương tác với khách hàng:

Phát triển chương trình khuyến mãi và thưởng đặc biệt dành riêng cho khách hàng ở phân khúc "At Risk" và "Churned" để khuyến khích họ tiếp tục mua hàng và trở thành khách hàng trung thành.



PHÂN TÍCH RFM NHÓM KHÁCH HÀNG CORPORATE



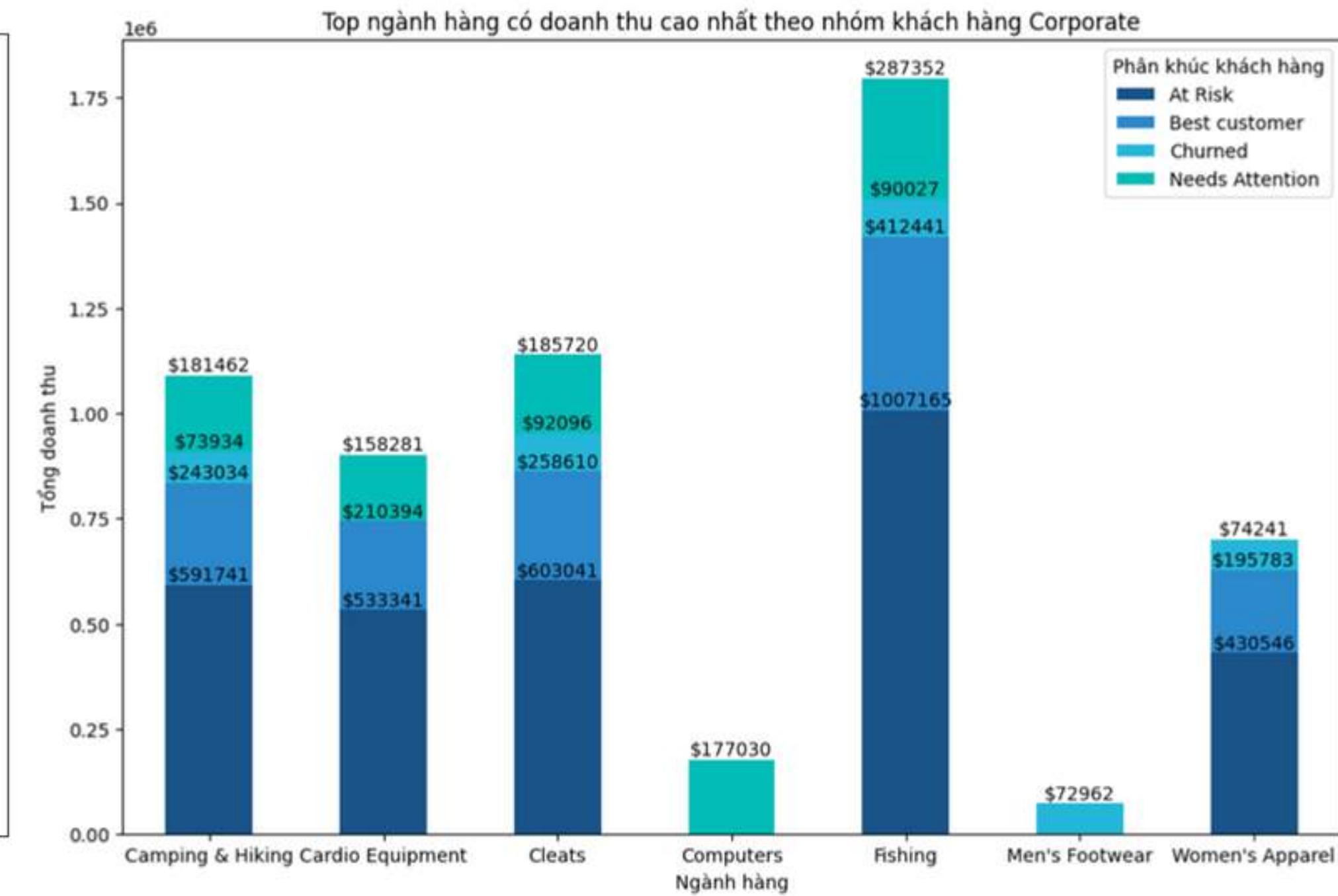
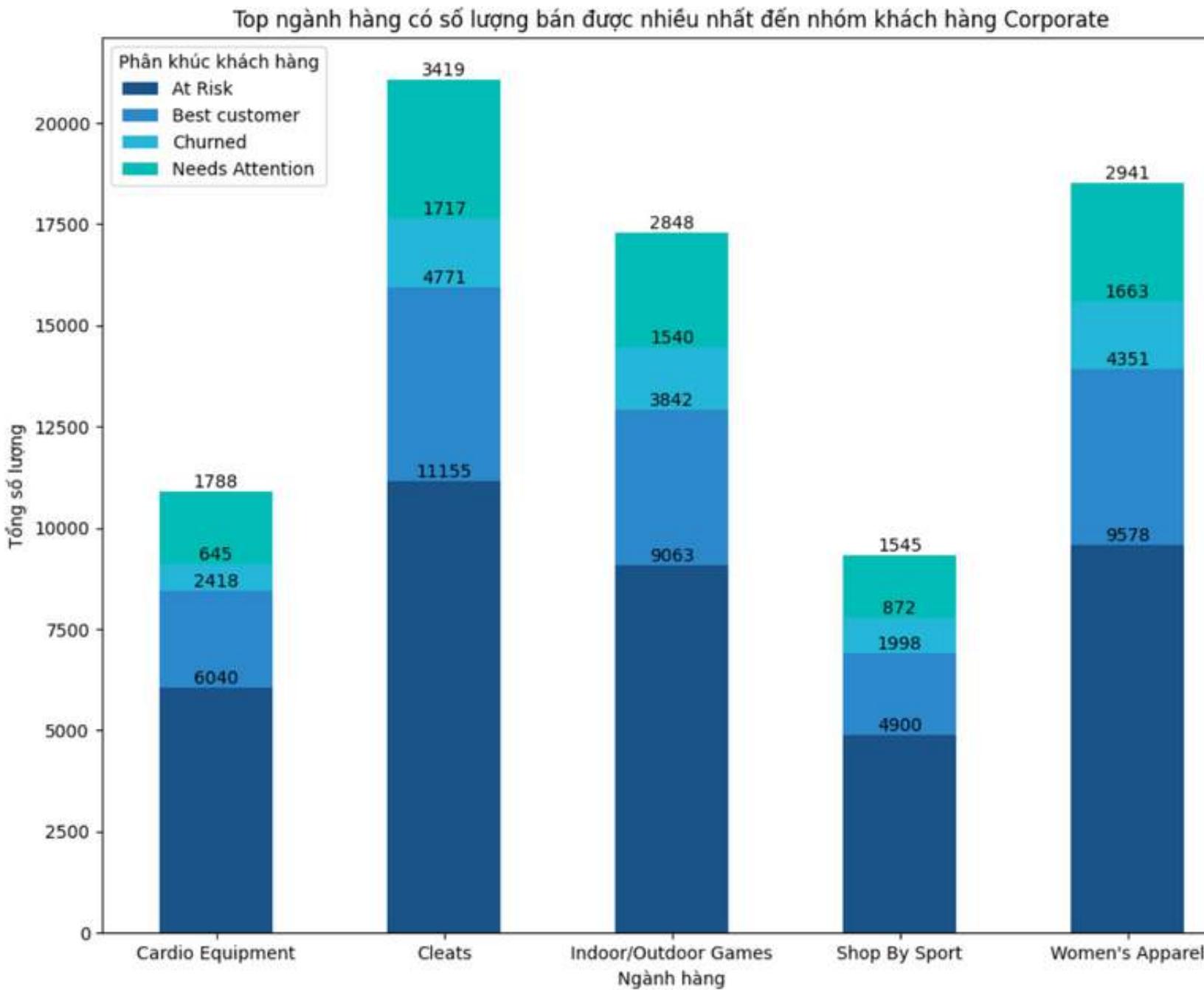
Sơ đồ phân phối khách hàng Corporate:

Tình hình phân bổ khách hàng trong nhóm Corporate cho thấy sự tập trung lớn vào **phân khúc Needs Attention và At Risk, chiếm hơn 80% tổng số khách hàng**. Điều này cho thấy có thể cần tập trung nhiều hơn vào việc duy trì và thu hút khách hàng mới trong nhóm này.

Tổng số đơn hàng và đơn hàng bị trễ:

- Tổng số **đơn hàng bị trễ chiếm 57%** tổng số đơn hàng.
- Tỷ lệ đơn hàng trễ khá cao, đặc biệt là trong phân khúc **At Risk**, với hơn **49% tỷ lệ đơn hàng bị trễ**. Điều này có thể gây ra sự không hài lòng và tăng nguy cơ mất mát khách hàng.

Phân tích RFM nhóm khách hàng Corporate



Cleats và Women's Apparel là hai ngành hàng được bán nhiều nhất cho nhóm khách hàng Corporate. Tuy nhiên, ngành hàng **Fishing mang lại doanh thu cao nhất**, đạt tổng cộng **\$1,796,985**. Điều này có thể chỉ ra rằng mặc dù Cleats được bán nhiều, nhưng sản phẩm trong ngành Fishing có giá trị cao hơn, dẫn đến doanh thu cao hơn. Mặt khác, riêng **nhóm khách hàng Churned** lại tập trung mua **ngành hàng Men's Footwear** nhiều nhất, cần nghiên cứu về lý do tại sao nhóm khách hàng này chọn rời bỏ và tìm kiếm phản hồi từ họ để cải thiện trải nghiệm mua sắm và giữ chân họ lại.

ĐỀ XUẤT HÀNH ĐỘNG ÁP DỤNG VỚI NHÓM KHÁCH HÀNG CORPORATE



Cải thiện quy trình và thời gian giao hàng:

Tối ưu hóa quy trình và hệ thống giao hàng để giảm thiểu thời gian giao hàng và giải quyết các vấn đề liên quan đến đơn hàng bị trễ.



Chiến lược tiếp thị cho ngành hàng Fishing:

Tạo chiến dịch quảng cáo và tiếp thị tập trung vào sản phẩm Fishing để thu hút khách hàng và khuyến khích mua hàng từ khách hàng At Risk và Churned.



Giảm số lượng khách hàng rời bỏ:

Tiến hành cuộc khảo sát khách hàng nhóm Churned về sản phẩm Men's Footwear.

Cải thiện chất lượng và dịch vụ liên quan đến Men's Footwear dựa trên phản hồi từ khách hàng để giảm tỷ lệ churn trong tương lai.



Tối ưu hóa bán hàng cho ngành hàng Women's Apparel:

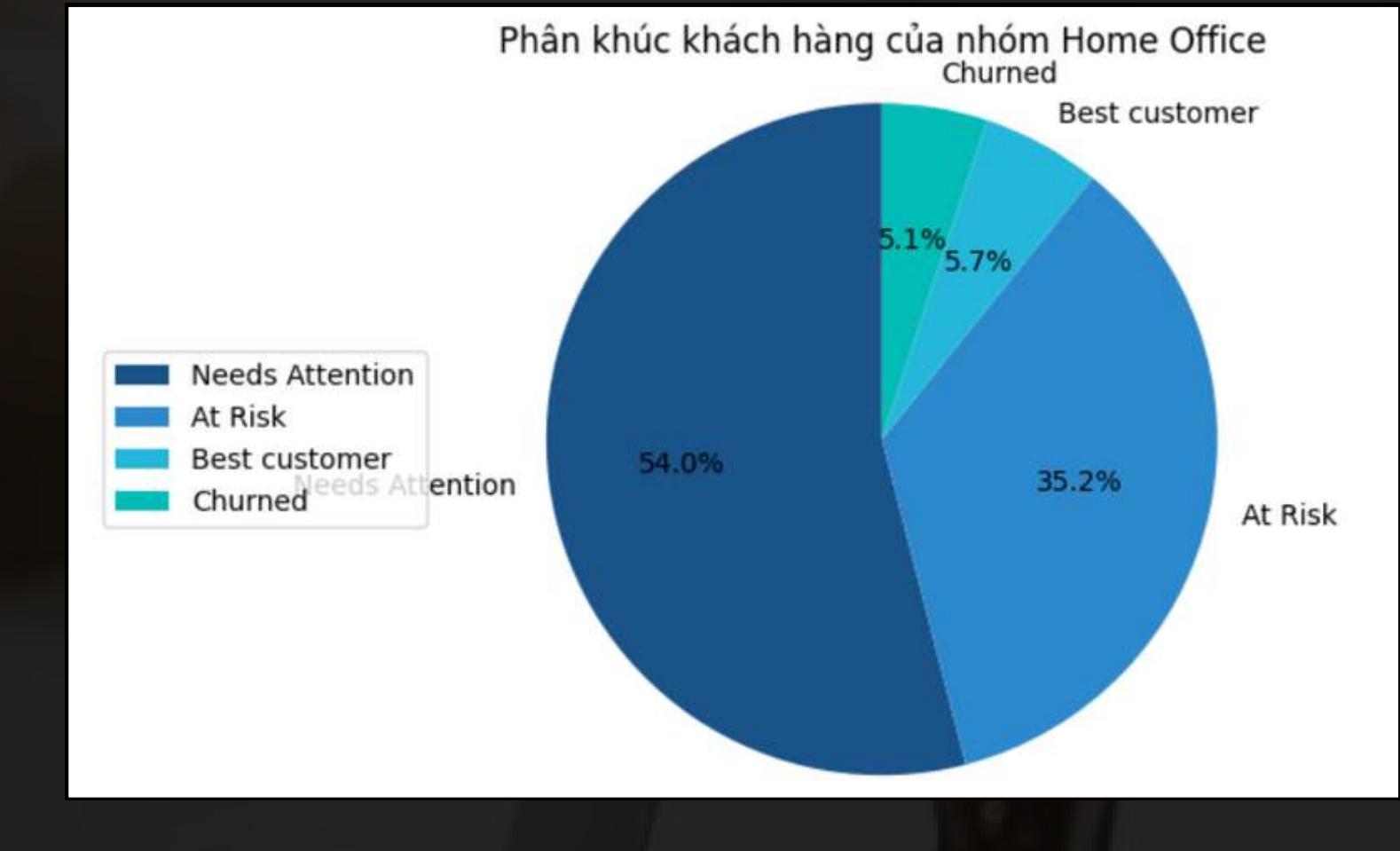
Xây dựng chương trình khách hàng trung thành và ưu đãi để tăng cường mua sắm lặp lại.

Tạo chiến dịch quảng cáo và tiếp thị đặc biệt nhắm vào khách hàng Best customer và Needs Attention để tăng doanh thu từ Women's Apparel.

PHÂN TÍCH RFM NHÓM KHÁCH HÀNG HOME OFFICE

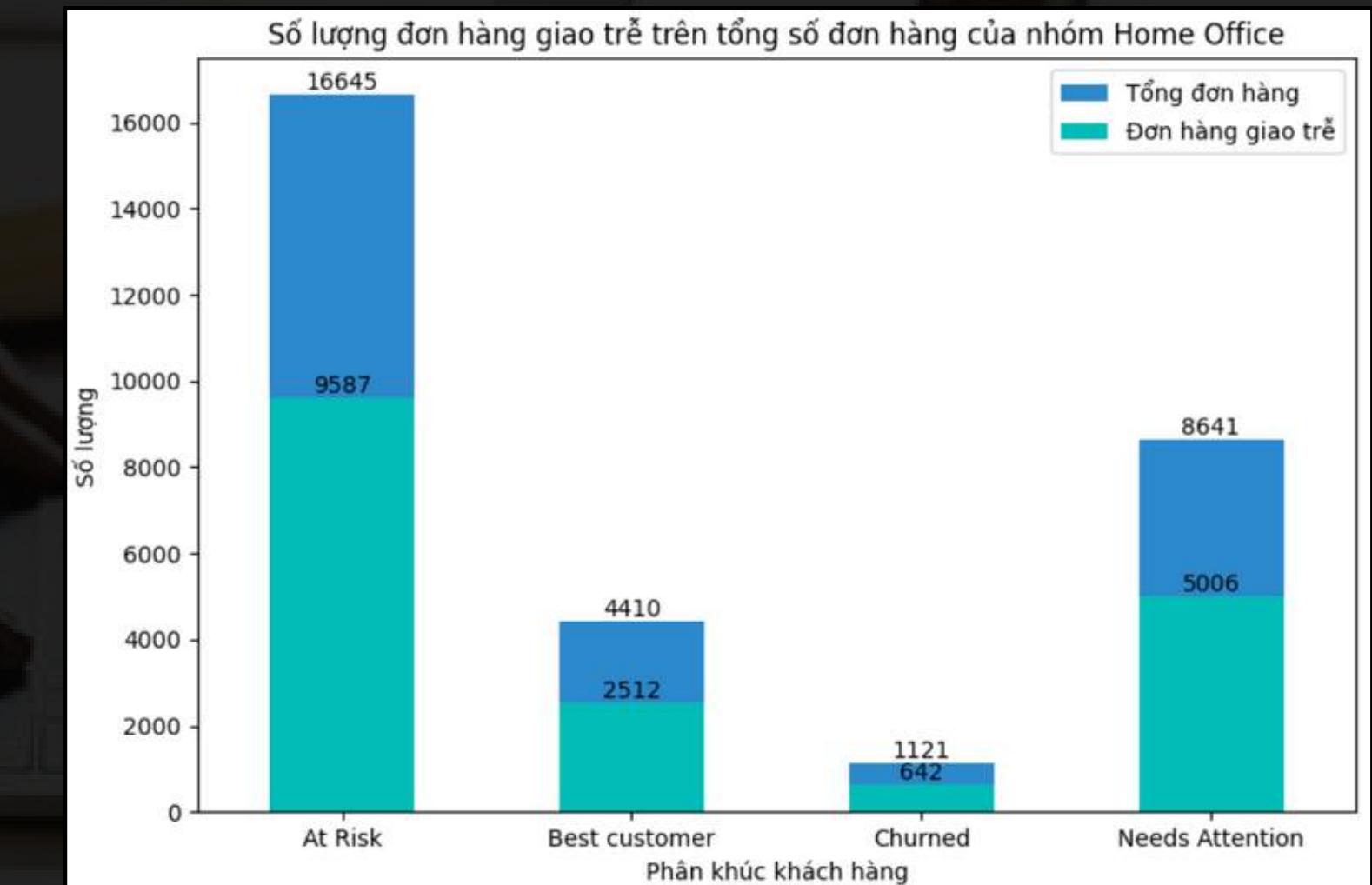
Phân bố khách hàng Home Office:

Trong tổng số 3,654 khách hàng Home Office, phần lớn (**89.2%**) **thuộc vào các phân khúc Needs Attention và At Risk**. Chỉ có một số nhỏ (5.7%) được xem là Best customers. Điều này chỉ ra rằng đa số khách hàng Home Office đang đối diện với các vấn đề và cần sự chú ý đặc biệt.

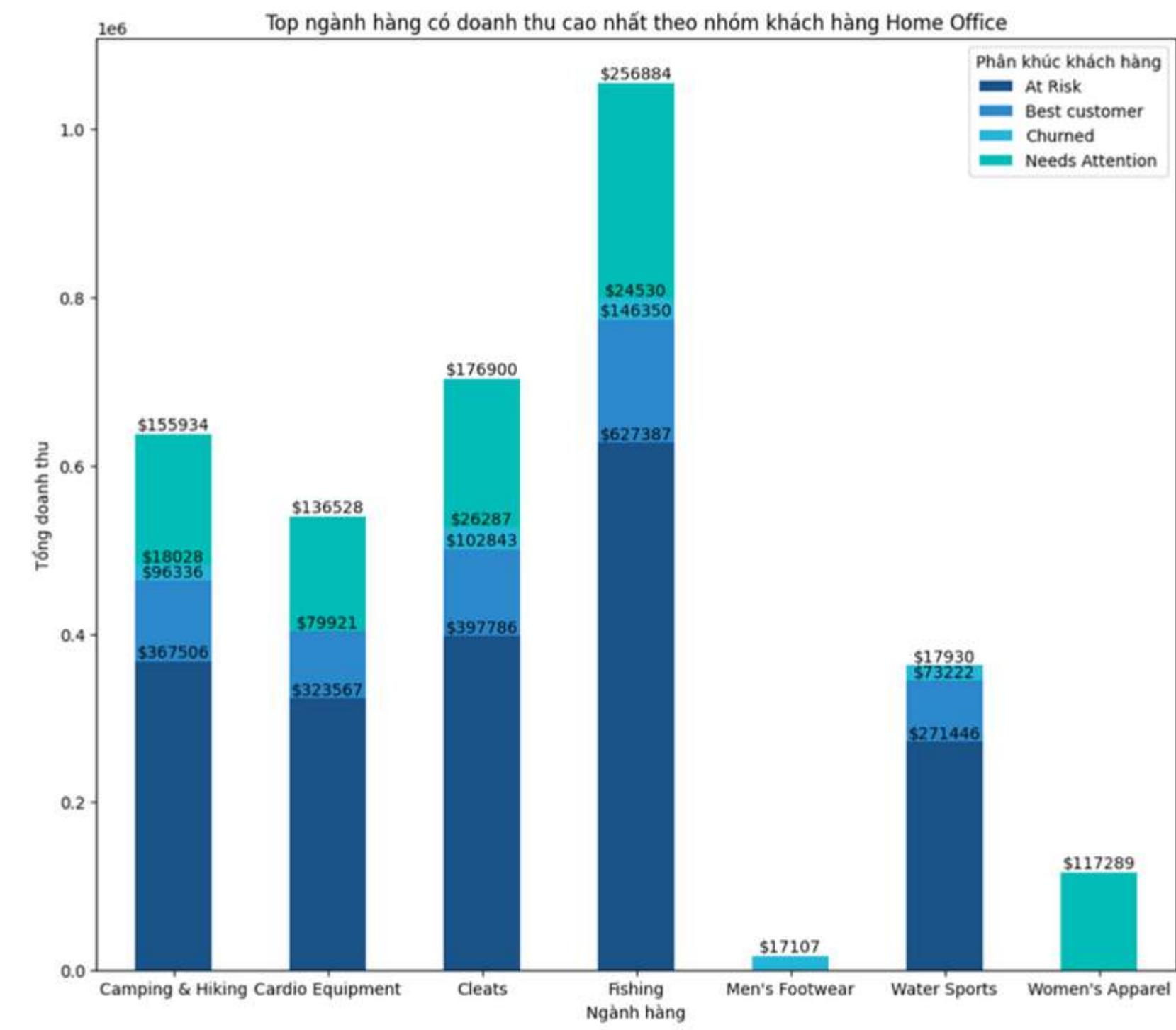
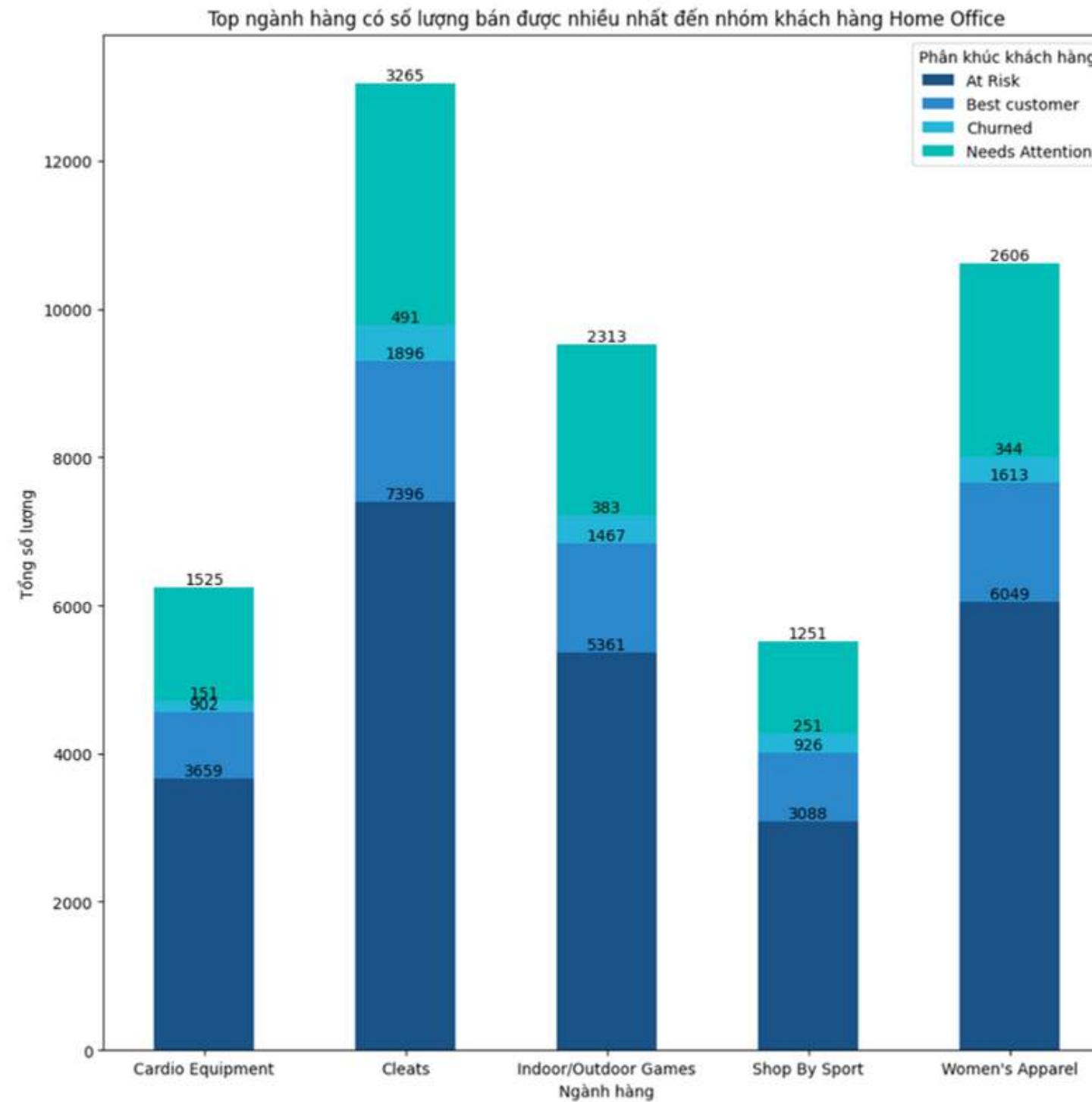


Tổng đơn hàng và số đơn hàng bị giao trễ:

Tổng **số lượng đơn hàng bị giao trễ** là 17,747, chiếm tỷ lệ lớn (**57.5%**) trong tổng số 30,817 đơn hàng. Số lượng đơn hàng bị trễ tăng theo từng phân khúc, đặc biệt là ở **phân khúc At Risk**, với tổng cộng **9,587 đơn hàng giao trễ**. Điều này gợi ý rằng việc đơn hàng bị giao trễ có thể ảnh hưởng đến mối quan hệ với khách hàng, đặc biệt là ở nhóm khách hàng có nguy cơ cao như At Risk.



Phân tích RFM nhóm khách hàng Home Office



Cleats và Women's Apparel là hai trong số những ngành hàng được bán nhiều nhất, với Cleats dẫn đầu về số lượng bán. Tuy nhiên, khi xét doanh thu, Fishing và Cleats là hai ngành hàng mang lại doanh thu cao nhất. Sự tương quan này cho thấy rằng, mặc dù Cleats và Women's Apparel có số lượng bán cao, nhưng giá trị của các đơn hàng từ Fishing và Cleats cao hơn, dẫn đến doanh thu cao hơn.

Đề xuất hành động đối với nhóm khách hàng Home Office



CẢI THIỆN DỊCH VỤ GIAO HÀNG:

Thực hiện đánh giá lại quy trình giao hàng và xem xét việc sử dụng các đối tác vận chuyển khác nhau để giảm thiểu thời gian giao hàng và số lượng đơn hàng bị trễ.



TỐI ƯU HOÁ CHIẾN LƯỢC BÁN HÀNG:

Phát triển các chiến lược tiếp thị đặc biệt nhằm vào phân khúc Best customers và Needs Attention trong nhóm Home Office như cung cấp ưu đãi đặc biệt và khuyến mãi để tăng cường sự trung thành và giữ chân khách hàng.



ĐỔI CHIẾN LƯỢC TIẾP THỊ:

Chú trọng vào việc tối ưu hóa chiến lược tiếp thị và bán hàng cho các ngành hàng mang lại giá trị cao như Fishing và Cleats, thay vì tập trung nhiều vào các ngành hàng mang lại giá trị thấp.

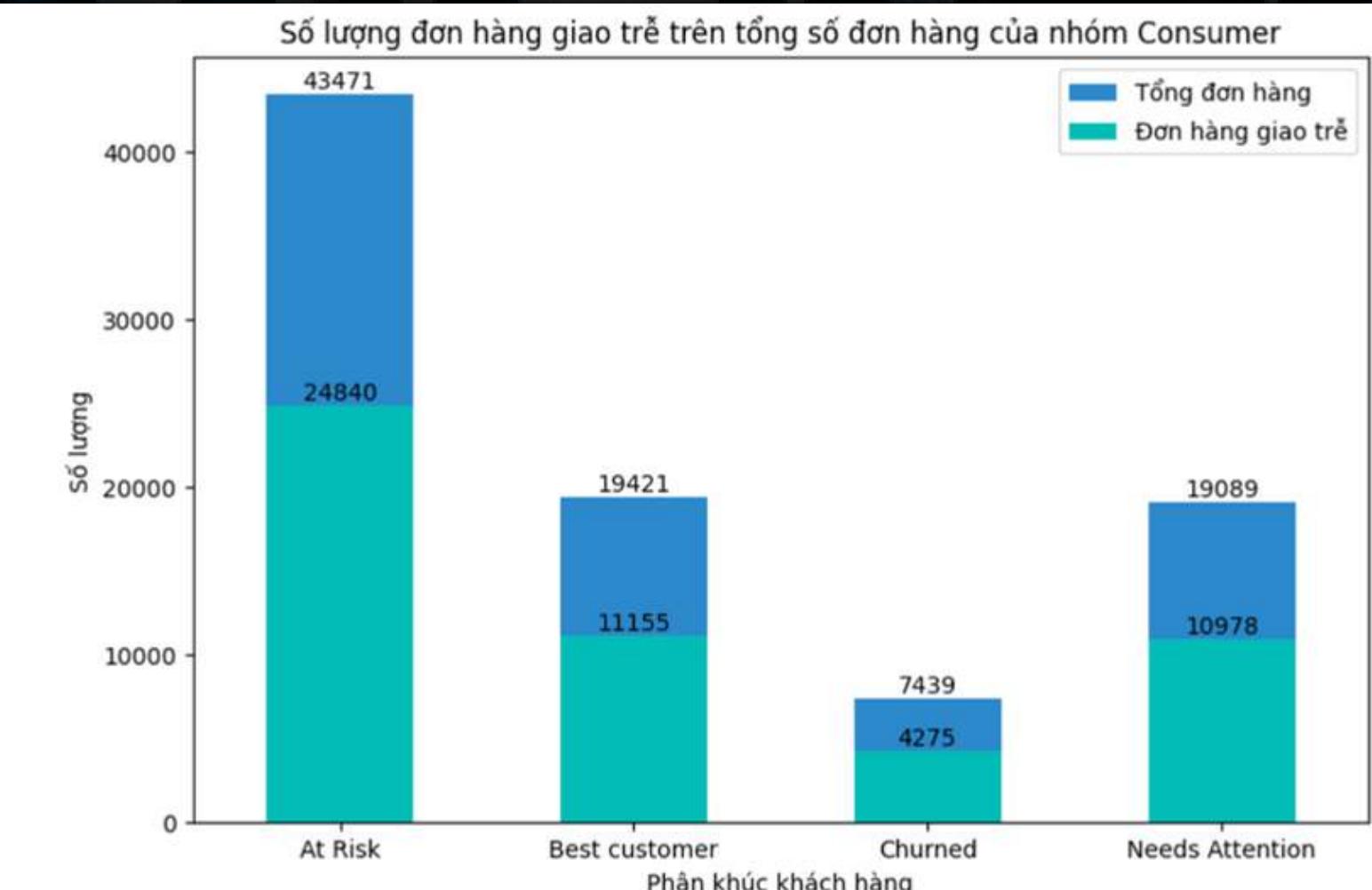
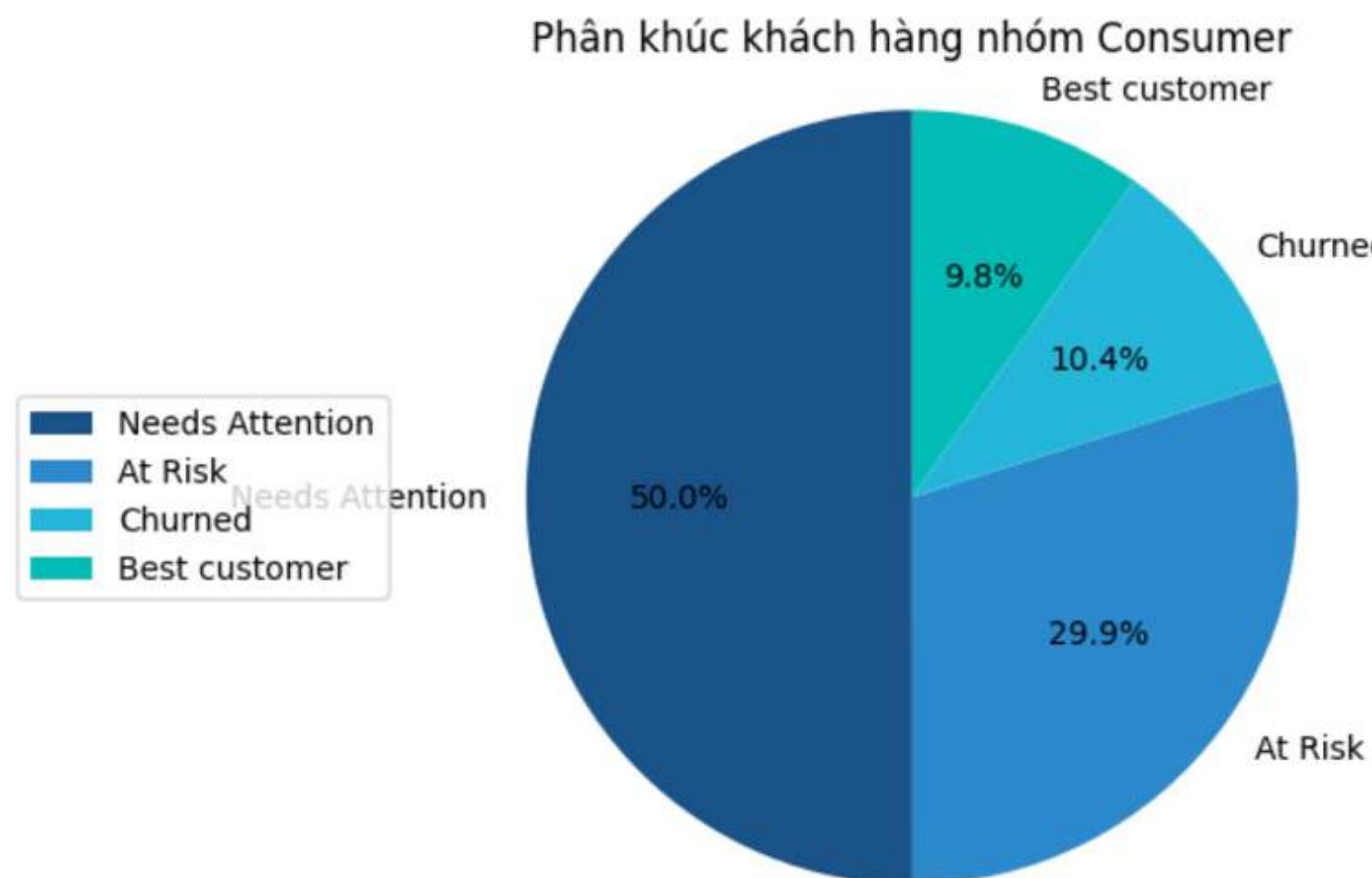
Phân tích RFM nhóm khách hàng Consumer

Sơ đồ phân phối khách hàng Corporate:

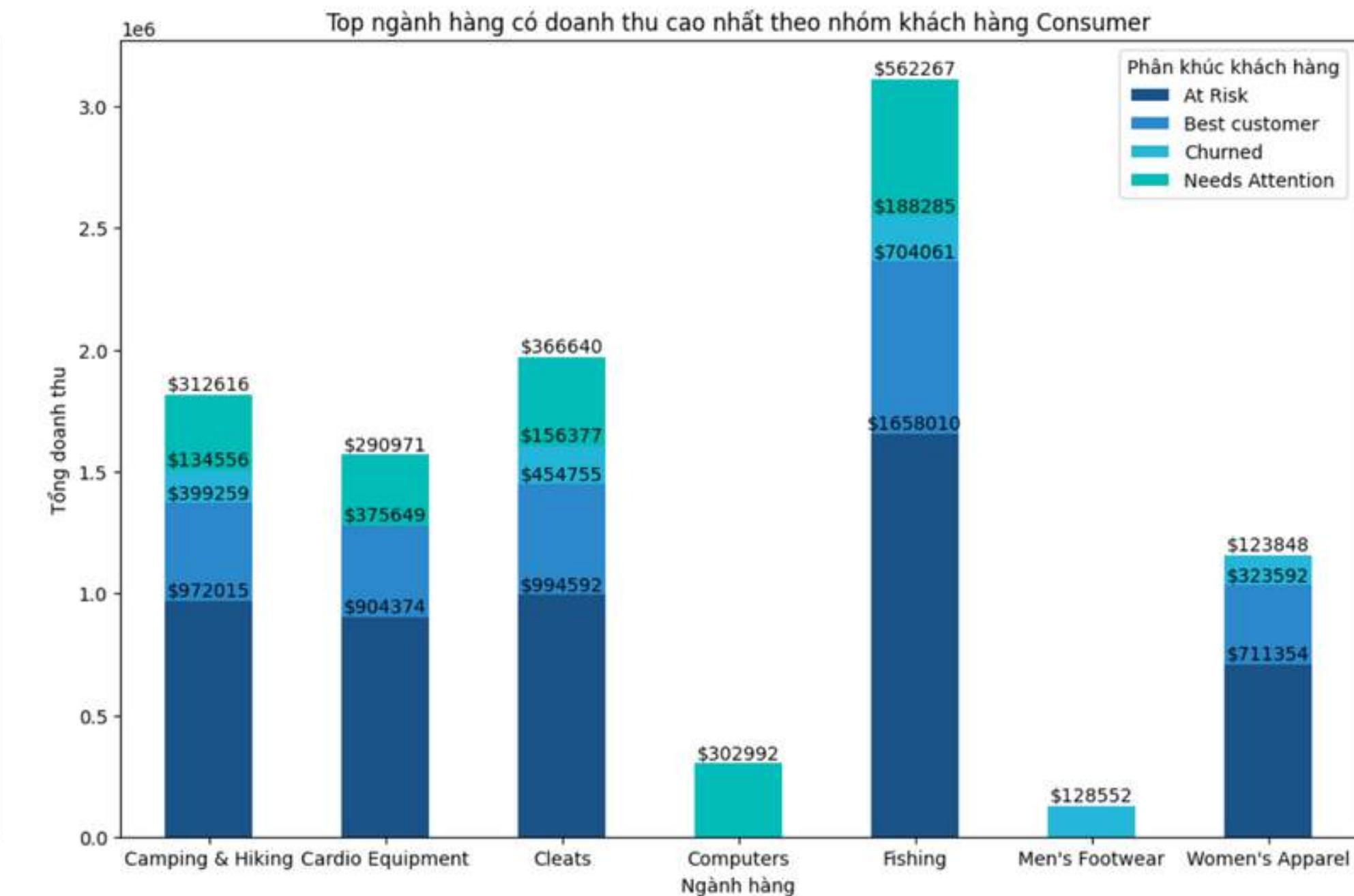
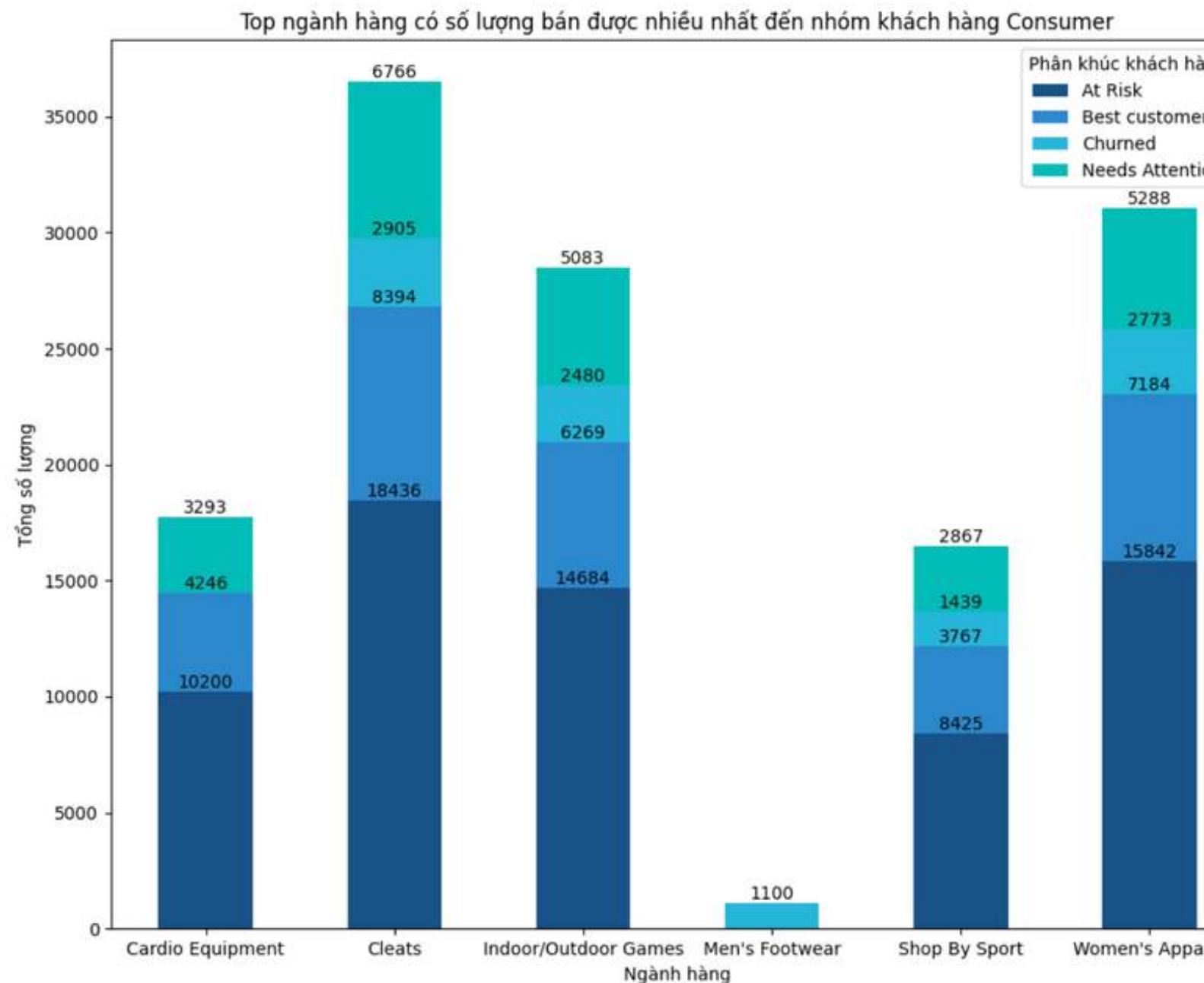
Số lượng khách hàng Consumer có sự phân bố tương đối đồng đều giữa các phân khúc. Tuy nhiên, tỷ lệ **At risk đang ở mức cao**, chiếm khoảng **30%** tổng số khách hàng.

Đơn hàng giao trễ:

Số **đơn hàng bị giao trễ** trong nhóm Consumer rất đáng chú ý khi chiếm **gần 60%** tổng số đơn hàng. Phân khúc At Risk có tỷ lệ đơn hàng bị trễ cao nhất, lên đến gần 50% tổng số đơn hàng.



Phân tích RFM nhóm khách hàng Consumer



Tương tự với nhóm khách hàng Corporate và Home Office, Cleats và Women's Apparel là hai ngành hàng có số lượng bán cao nhất đối với nhóm khách hàng Consumer, với Cleats đứng đầu với tổng số lượng bán là 36,501. Có thể thấy rằng các sản phẩm trong ngành hàng Fishing và Cleats không chỉ được bán nhiều mà còn mang lại doanh thu cao, đặc biệt là từ phân khúc khách hàng At Risk.

Đề xuất hành động đối với nhóm khách hàng Consumer

Tăng cường chuyển đổi phân khúc At Risk

Phân khúc khách hàng "At Risk" chiếm tỷ lệ lớn trong nhóm khách hàng Consumer, công ty đối mặt với nguy cơ họ có thể chuyển sang phân khúc "Churned".

Cần tổ chức các chiến dịch tiếp thị đa kênh (email marketing, quảng cáo trực tuyến) tập trung vào sản phẩm Cleats và Women's Apparel.

Tạo ra các gói ưu đãi và khuyến mãi đặc biệt dành riêng cho phân khúc này, như giảm giá, quà tặng kèm theo, hoặc vận chuyển miễn phí.

Giảm thiểu số lượng khách hàng rời bỏ

Khảo sát và nghiên cứu lý do khiến khách hàng "Churned" rời bỏ sau khi mua sản phẩm Men's Footwear.

Cải thiện chất lượng sản phẩm, dịch vụ hậu mãi hoặc chính sách đổi trả để giải quyết những vấn đề được phản ánh.

Tổ chức các chương trình giới thiệu sản phẩm mới hoặc thử nghiệm sản phẩm miễn phí để thu hút lại khách hàng "Churned" và tạo cơ hội tái kích thích mua sắm.

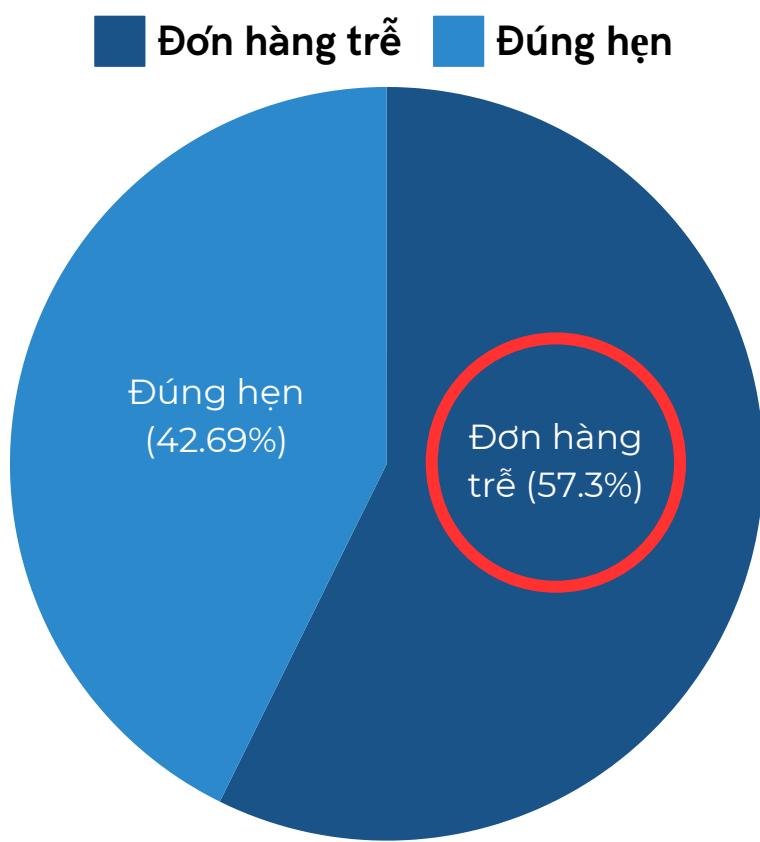


PHÂN TÍCH

THỜI GIAN VẬN CHUYỂN

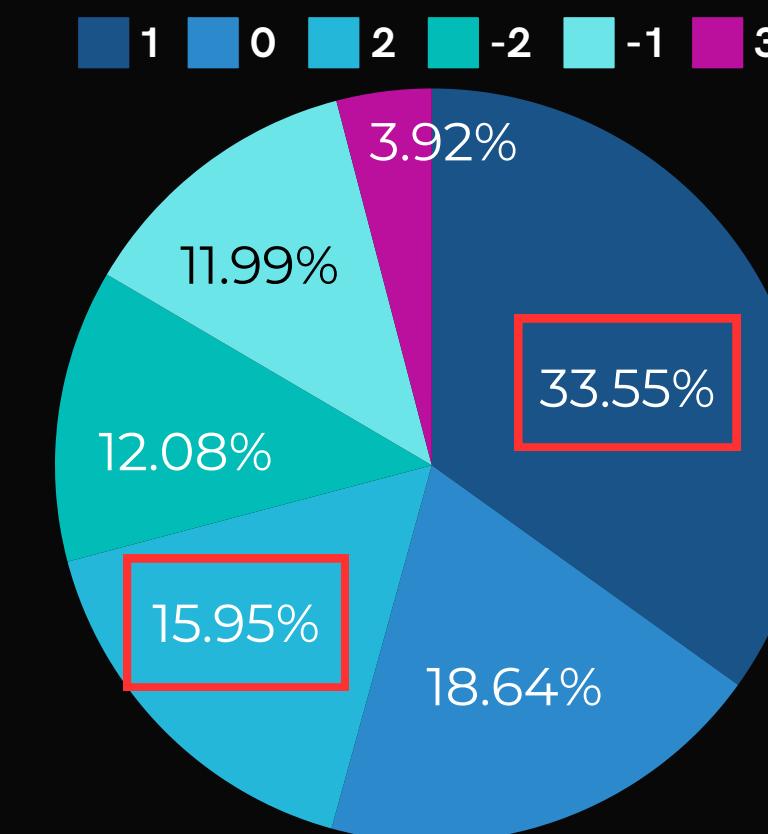
THỜI GIAN VẬN CHUYỂN

Tỷ trọng trạng thái đơn hàng



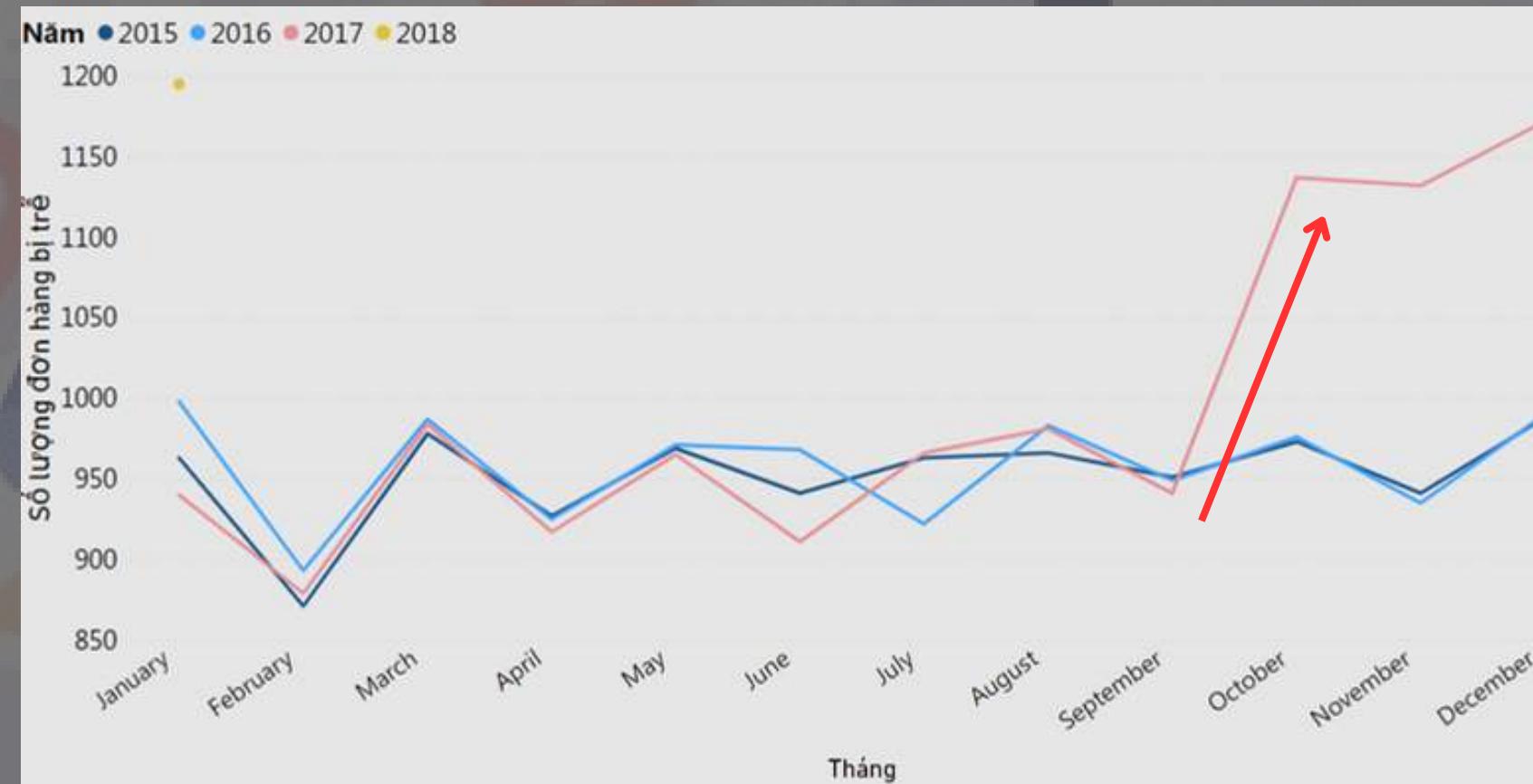
→ Tỷ lệ đơn giao hàng bị trễ lên đến 57,29% so với tổng số đơn hàng được giao thành công.

Tỷ trọng trạng thái đơn hàng



→ Phản ánh các đơn hàng thường giao trễ hơn từ 1 đến 2 ngày so với dự kiến. Cụ thể: các đơn hàng giao trễ hơn 1 ngày chiếm khoảng 21,13 nghìn đơn hàng đơn hàng (~33,55%), đơn hàng giao trễ hơn 2 ngày chiếm 10,01 nghìn đơn hàng (~15,95%).

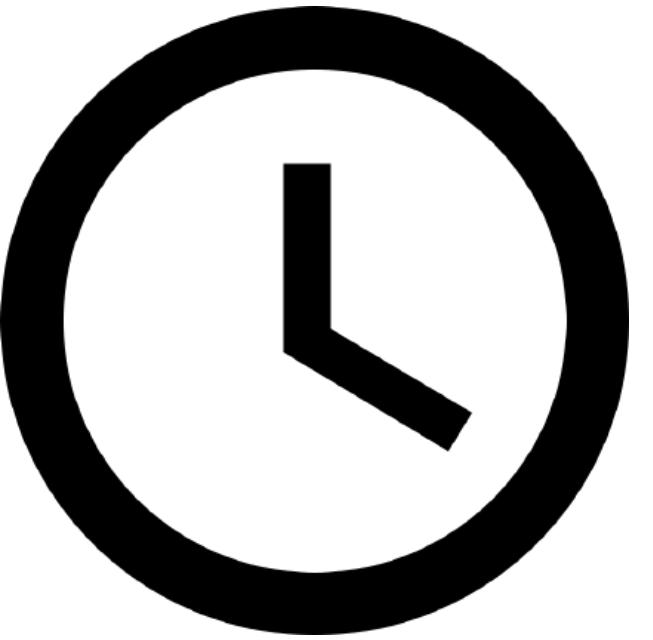
THỜI GIAN VẬN CHUYỂN



- Trong giai đoạn từ 2015 đến 2017, tỷ lệ đơn hàng trễ tăng cao vào các tháng như: **tháng 6, tháng 8, tháng 9 và tháng 12.**
- Từ tháng 9 năm 2017 đến tháng 1 năm 2018, **có sự tăng đột biến** trong số lượng đơn hàng giao trễ do nhu cầu tăng cao so với các năm trước.
- Nguyên nhân là do:
 - + **Các tháng cuối năm:** Sự tăng trong các tháng cuối năm là do **nhu cầu cao trong mùa lễ hội**, dẫn đến quá tải cho hệ thống logistics. **Cuối năm tại phương Tây là thời điểm các dịp lễ lớn như Halloween, Lễ Giáng Sinh và Tết Dương Lịch.** Doanh nghiệp thường có khó khăn trong việc duy trì hiệu suất chuỗi cung ứng trong thời gian này, dẫn đến tình trạng ùn ứ đơn hàng.
 - + **Giai đoạn tháng 6 đến tháng 8:** Đây là thời điểm bước vào mùa hè, là **thời gian phổ biến cho kỳ nghỉ và du lịch**, dẫn đến gia tăng nhu cầu mua hàng hóa.

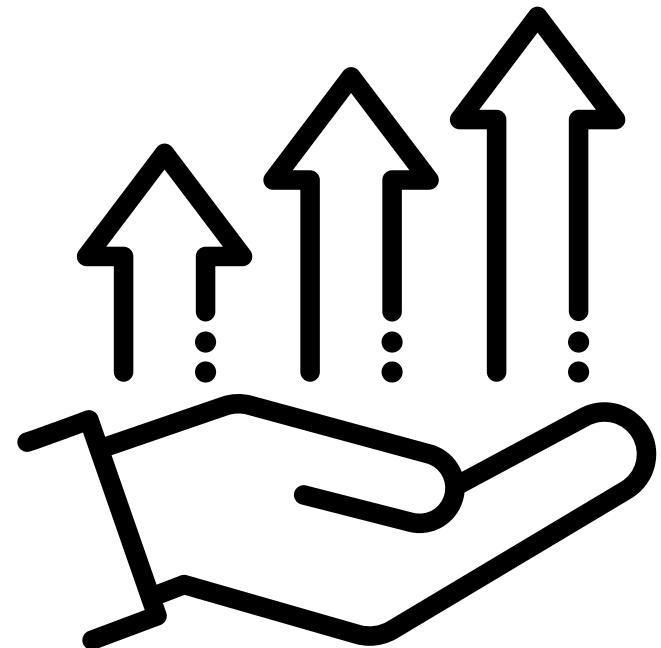


GIẢI PHÁP



ĐIỀU CHỈNH THỜI GIAN GIAO HÀNG DỰ KIẾN

Cần có sự điều chỉnh trong việc dự báo nhu cầu và quản lý chuỗi cung ứng, đặc biệt là trong các giai đoạn cao điểm như mùa hè và các tháng cuối năm. Trên cơ sở đó đưa ra thời gian giao hàng dự kiến hợp lý. Hạn chế giao hàng trễ dẫn đến tình trạng đánh mất khách hàng.



NÂNG CAO QUY TRÌNH

Cần tăng cường nguồn lực logistic, cải thiện quy trình xử lý đơn hàng, và tăng cường trao đổi với khách hàng về thời gian giao hàng dự kiến

THỜI GIAN GIAO HÀNG TRÊN THEO VỊ TRÍ ĐỊA LÝ



THEO THỊ TRƯỜNG



THEO KHU VỰC

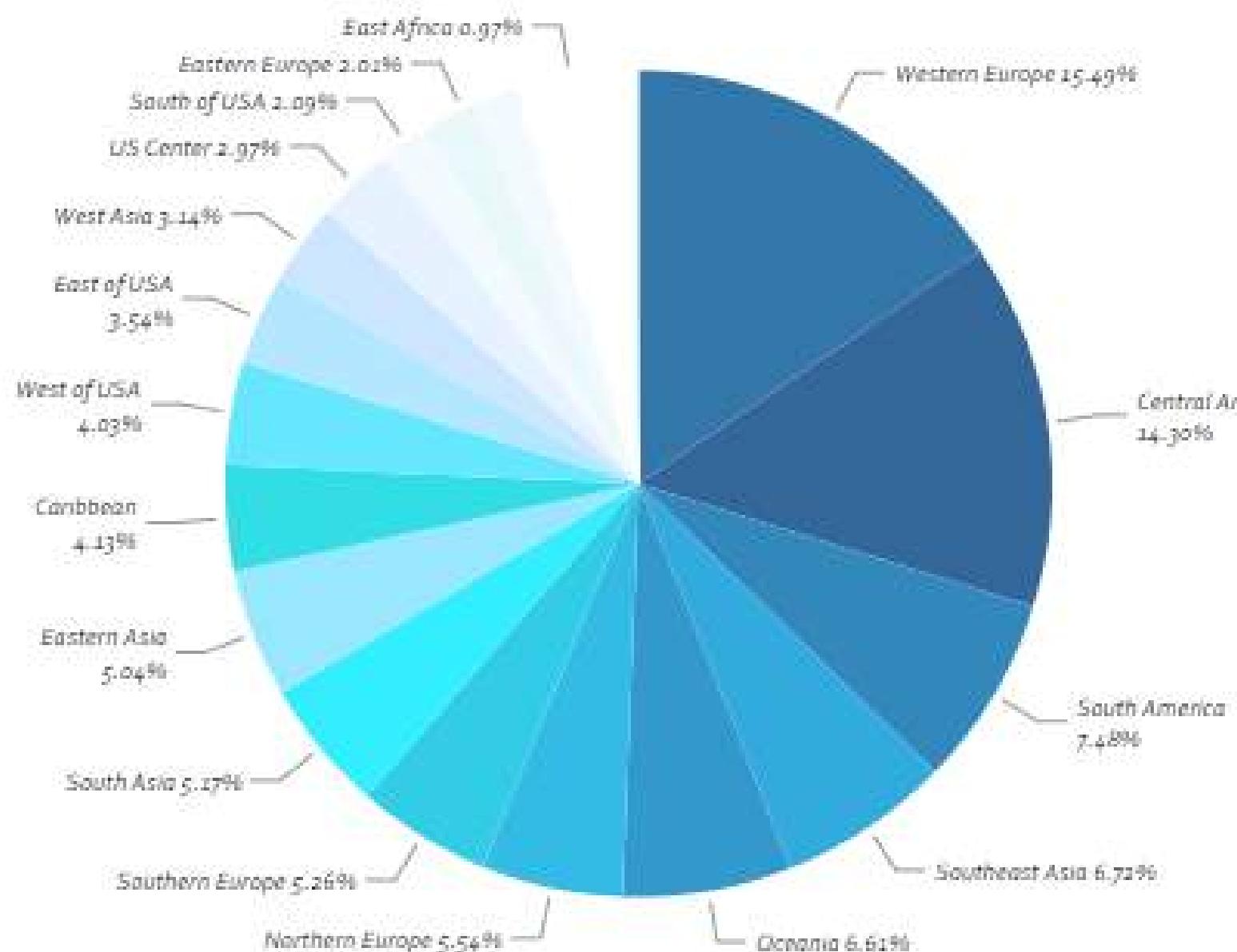
PHÂN BỐ THEO THỊ TRƯỜNG

- Tỷ lệ giao hàng trễ của các thị trường tương đương nhau và ở mức cao (>50%)
- **Thị trường Châu Á - Thái Bình Dương (Pacific Asia), Châu Âu (Europe) và Mỹ La Tinh (LATAM) là 3 thị trường có đơn hàng được giao nhiều nhất** trong số 5 thị trường hoạt động, đây là ba thị trường tiềm năng và phát triển, có tiềm năng khai thác cao.
- Tuy nhiên, **thị trường Mỹ La Tinh lại có tỷ lệ giao hàng trễ thấp** hơn so với hai thị trường còn lại - Châu Âu và Châu Á - Thái Bình Dương
- Đáng chú ý, tại **thị trường USCA (Canada và Hoa Kỳ)** có khối lượng đơn hàng thấp nhưng tỷ lệ giao hàng trễ khá cao (~57,52%).

Số lượng đơn hàng và tỷ lệ đơn hàng giao trễ phân bố theo thị trường



Tỷ trọng đơn hàng giao trả theo khu vực



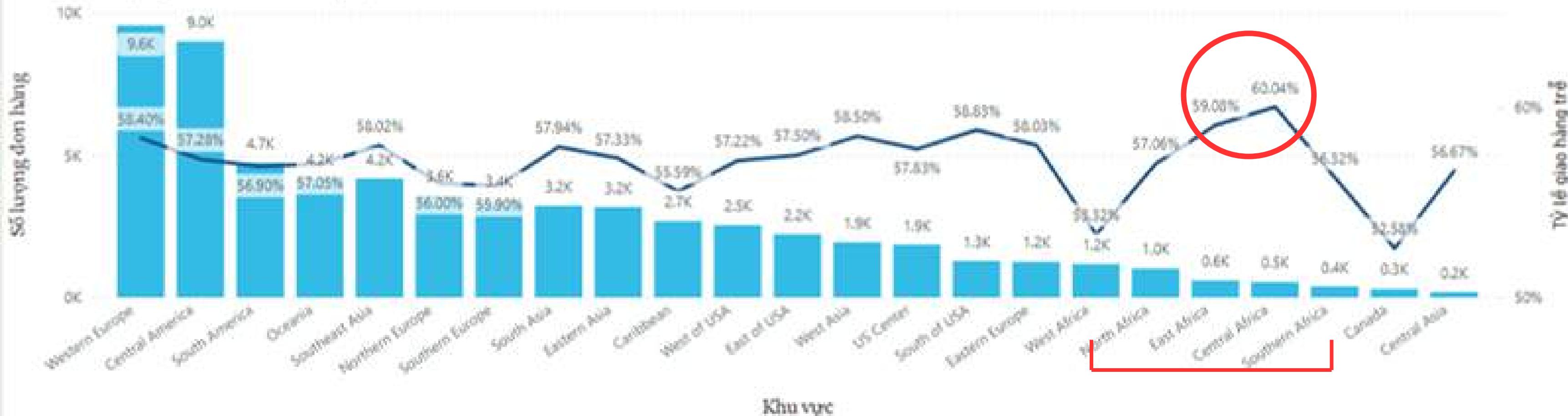
- Khu vực
- Western Europe
 - Central America
 - South America
 - Southeast Asia
 - Oceania
 - Northern Europe
 - Southern Europe
 - South Asia
 - Eastern Asia
 - Caribbean
 - West of USA
 - East of USA
 - West Asia
 - US Center
 - South of USA
 - Eastern Europe
 - West Africa
 - North Africa
 - East Africa
 - Central Africa
 - Southern Africa
 - Canada
 - Central Asia

PHÂN BỐ THEO KHU VỰC VẬN CHUYỂN

- **Khu vực trung tâm, khu vực phía Nam của Mỹ và Bắc Âu** là 3 khu vực có tỷ lệ đơn hàng giao trả trên nhiều nhất, lần lượt chiếm tỷ trọng là 15,68%, 15,3% và 8,19%. Đồng thời 3 khu vực này cũng có khối lượng đơn hàng cao nhất.
- Nguyên nhân là do các khu vực này có lượng đơn hàng lớn đồng nghĩa với áp lực lớn cho hệ thống vận chuyển và giao nhận.

Số lượng đơn hàng và tỷ lệ đơn hàng giao trễ phân bố theo khu vực

● Total Orders by Region ● Late Order Rate by Region



PHÂN BỐ THEO KHU VỰC

- **Khu vực châu Phi có khá ít đơn hàng, tuy nhiên tỷ lệ đơn hàng giao trễ lại cao.** Cao nhất là khu vực trung tâm châu Phi. Nguyên nhân là do hệ thống logistic và cơ sở hạ tầng chưa phát triển tốt, gây ra khó khăn trong quá trình giao hàng.
 - **Khu vực Nam Âu:** Với tỷ lệ giao hàng trễ là 55.90%, khu vực này có hệ thống vận chuyển và quản lý chuỗi cung ứng hiệu quả, hạ tầng vận chuyển phát triển và quy trình vận hành được tối ưu hóa.
 - **Khu vực Caribbean:** Với tỷ lệ giao hàng trễ là 55.59%, khu vực này cũng có một hệ thống vận chuyển và quản lý chuỗi cung ứng khá tốt, quy trình vận hành được điều chỉnh và tối ưu hóa cho phù hợp với điều kiện địa phương.

KIẾN NGHỊ - GIẢI PHÁP

Đề xuất phương án làm giảm tỷ lệ giao hàng trễ tại các thị trường và khu vực:

Tăng cường nghiên cứu

Cần tiến hành nghiên cứu chi tiết về các yếu tố cụ thể ảnh hưởng đến hiệu suất giao hàng trong từng thị trường và khu vực

Tối ưu hóa quy trình

Cần tối ưu hóa quy trình vận chuyển và quản lý chuỗi cung ứng tại các thị trường và khu vực có tỷ lệ đơn hàng giao trễ cao, nhằm tăng cường hiệu suất và giảm thiểu sự cố

Tăng cường hợp tác

Cần tăng cường hợp tác với các đối tác logistic và nhà cung cấp địa phương tại các khu vực có tỷ lệ giao hàng trễ cao, nhằm cải thiện quá trình giao hàng và giảm thiểu thời gian giao hàng.

PHÂN TÍCH **SHIPPING MODE**

(Phương thức vận chuyển)

SHIPPING MODE

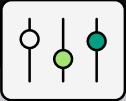
Thời gian giao hàng dự kiến, thực tế cho từng shipping mode

Shipping_Mode	Max of Days_for_shipment _(scheduled)	Min of Days_for_shipping _(real)	Max of Days_for_shipping_ (real)	% late_shipment_per_shipping _mode
Priority	0	0	1	48.37%
First Class	1	2	2	100.00%
Second Class	2	2	6	79.99%
Standard Class	4	2	6	39.85%
Tổng cộng	4	0	6	57.31%

Tỷ lệ trễ hàng tất cả các hạng dịch vụ đều cao. Điều này gây ảnh hưởng rất nghiêm trọng đến uy tín của doanh nghiệp và mất lòng tin từ khách hàng.

SHIPPING MODE

Tính độ tương quan giữa Shipping Mode và thời gian giao hàng



Calculate the relationship between Shipping Mode và Late_delivery_risk

```
df['Shipping_Mode'] = df['Shipping_Mode'].map({'Priority': 4, 'First Class': 3, 'Second Class': 2, 'Standard Class': 1})  
df[['Shipping_Mode', 'Product_Price', 'Late_delivery_risk', 'Days_for_shipment_(scheduled)']].corr()
```

Shipping mode <=> Độ trễ giao hàng

TƯƠNG QUAN DƯƠNG

Việc nâng cấp hạng dịch vụ vận chuyển giao hàng không làm tăng thời gian đúng hẹn.

**PHẢI ĐIỀU CHỈNH THỜI GIAN GIAO HÀNG
DỰ KIẾN ĐỂ GIÚP LÀM GIẢM TỶ LỆ TRỄ
GIAO HÀNG VÀ CUNG CẤP THÔNG TIN
CHÍNH XÁC CHO KHÁCH HÀNG.**



	Shipping_Mode	Product_Price	Late_delivery_risk	Days_for_shipment_(scheduled)
Shipping_Mode	1.000000	-0.007403	0.350452	-0.984968
Product_Price	-0.007403	1.000000	-0.002666	0.007125
Late_delivery_risk	0.350452	-0.002666	1.000000	-0.389626
Days_for_shipment_(scheduled)	-0.984968	0.007125	-0.389626	1.000000

**Shipping mode <=> Thời gian giao hàng
dự kiến**

TƯƠNG QUAN ÂM

**VIỆC TĂNG THỜI GIAN GIAO HÀNG DỰ
KIẾN SẼ LÀM GIẢM TỶ LỆ CHẬM GIAO
HÀNG**

- **First Class:** Với tỷ lệ trễ hàng là 100%, dịch vụ này cần được ưu tiên xem xét và cải thiện ngay lập tức. Trễ hàng ở mức 100% có thể gây mất lòng tin từ khách hàng và ảnh hưởng đến uy tín của doanh nghiệp.
- **Priority:** Mặc dù tỷ lệ trễ hàng không cao nhất nhưng vẫn có thể gây ra sự bất tiện và không hài lòng từ phía khách hàng vì đây là loại phương thức vận chuyển cao nhất mà khách hàng không nhận lại được chất lượng tương ứng.
- **Second Class:** Dịch vụ này cũng đang gặp vấn đề nghiêm trọng. Cần phải thực hiện các biện pháp cải thiện như tăng cường quản lý quy trình, đào tạo nhân viên, hoặc nâng cấp hạ tầng để giảm tỷ lệ trễ hàng.
- **Standard class:** Đây là hạng dịch vụ thấp nhất với thời gian giao hàng dài nhất, do đó tỷ lệ này thấp hơn so với các lớp dịch vụ khác, nhưng vẫn còn động lực để nâng cao chất lượng dịch vụ và giảm thiểu trễ hàng.

**ĐÁNH GIÁ ĐỘ TƯƠNG QUAN GIỮA SHIPPING
MODE VÀ THỜI GIAN GIAO HÀNG**

SHIPPING MODE



ĐỀ XUẤT

Điều chỉnh lại thời gian giao hàng dự kiến để phù hợp với từng loại Shipping mode



Shipping mode	Thời gian giao hàng dự kiến
Priority	1 ngày (24 tiếng)
First Class	2 ngày (48 tiếng)
Second Class	3-4 ngày
Standard	5-6 ngày



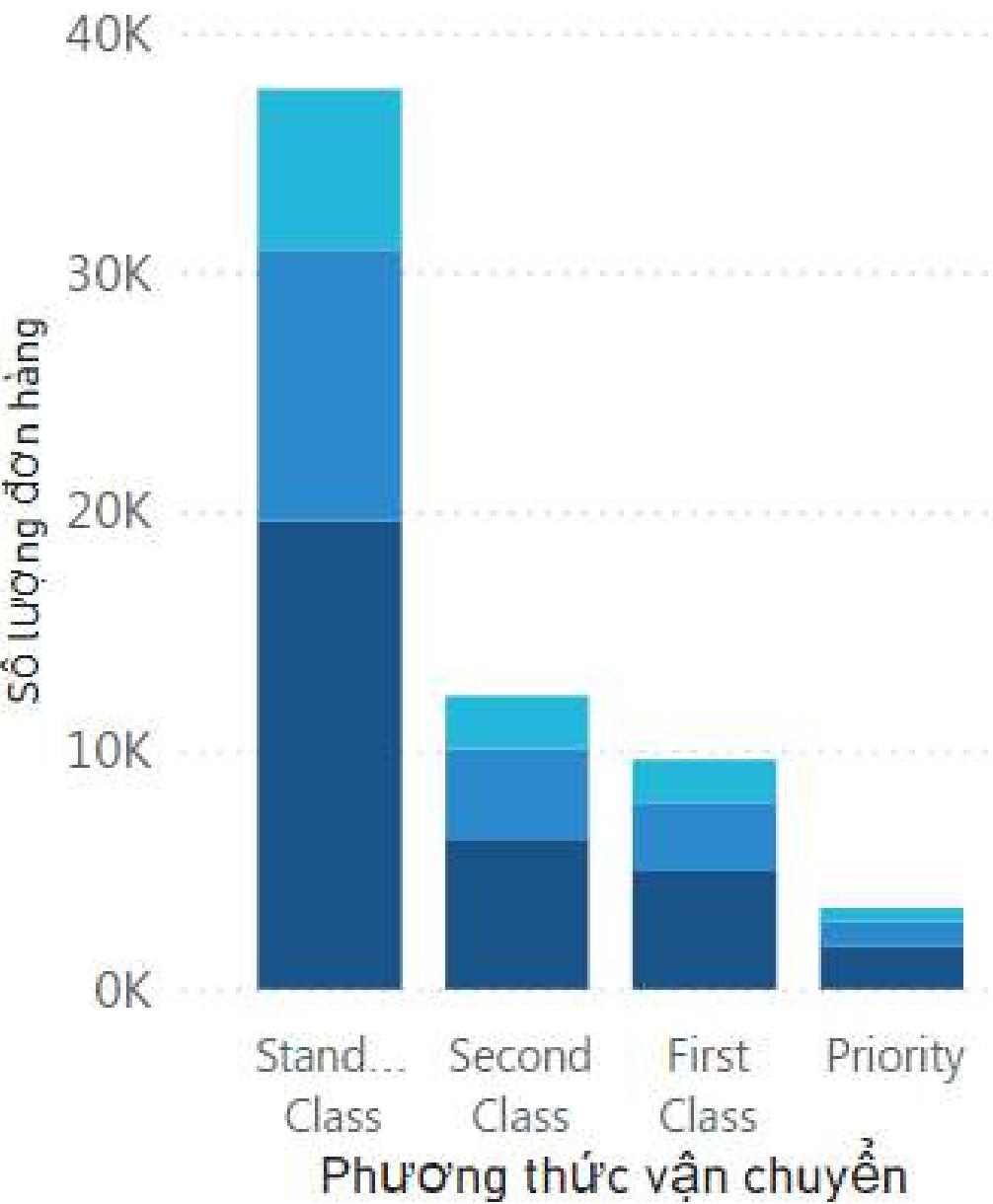
Kết quả sau khi điều chỉnh lại thời gian giao hàng dự kiến

Shipping_Mode	Max of Days_for_shipping_(scheduled)	Min of Days_for_shipping_(real)	Max of Days_for_shipping_(real)	% late_shipment_per_shipping_mode
Priority	1	0	1	1
First Class	2	2	2	2
Second Class	4	2	6	40.05%
Standard Class	6	2	6	6
Tổng cộng	6	0	6	7.80%

- Việc tăng thời gian dự kiến nhằm giảm tỷ lệ trễ hàng đối với công ty. Ngoài ra, khách hàng có thông tin chính xác thời gian nhận hàng để cân đối lịch cá nhân, công việc.
- Tăng phiếu giảm giá (voucher) cho khách hàng trong trường hợp bị chậm đơn.

SHIPPING MODE

Quy mô ● Consu... ● Corpor... ● Home ...



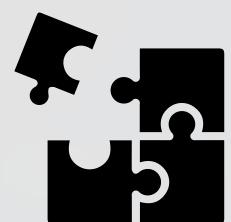
FAVORITE

Loại hình vận chuyển Standard Class được đặt nhiều nhất, vì vậy công ty phải tối ưu được thời gian vận chuyển thực tế nhằm mang lại sự tin tưởng, hài lòng của khách hàng.

Priority



First Class



Second Class



Standard Class



04

KẾT LUẬN VÀ ĐỀ XUẤT

4.1 KẾT LUẬN



Tình hình kinh doanh

Công ty có chiến lược kinh doanh tốt, quản lý và nhận biết rủi ro tốt. Tại từng thời điểm trong năm nên phân bổ nhân sự đồng đều xử lý đơn hàng nhanh nhất.



Lưu trữ các mặt hàng có số lượng bán hàng nhiều nhất.

Đối với các mặt hàng có số lượng bán ra nhiều, công ty nên dự trữ hàng trước hoặc cải tiến sản phẩm, đóng gói với thiết kế bắt mắt để thu hút khách hàng.



Kiểm soát quy trình vận chuyển hàng

Quy trình và thủ tục vận chuyển hàng chưa hiệu quả với độ trễ hàng cao.



Tập trung nâng cao chất lượng dịch vụ, đào tạo nội bộ

Luôn luôn đào tạo nội bộ, xây dựng quy trình vận chuyển để có thể đáp ứng nhanh với lượng đơn hàng lớn.

4.2 KIẾN NGHỊ

01.

Hợp tác với các nhà phân phối

Xây dựng mối quan hệ hợp tác với các nhà phân phối khác, có các kho trung gian tại Châu Á - Thái Bình Dương.

02.

Phân khúc hóa chiến lược tiếp thị

Dựa vào sở thích và nhu cầu của từng nhóm/ phân khúc khách hàng để tập trung tiếp thị các ngành hàng phù hợp với sở thích và nhu cầu của họ.

03.

Điều chỉnh thời gian giao hàng dự kiến

Điều chỉnh lại thời gian giao hàng dự kiến để đảm bảo rằng khách hàng nắm rõ về thời gian giao hàng chính xác cho đơn hàng.

04.

Áp dụng công nghệ tiên tiến

Sử dụng công nghệ để theo dõi và quản lý, cập nhật trạng thái đơn hàng.

05.

Cải thiện quy trình vận chuyển

Cải thiện chất lượng dịch vụ và giảm thiểu tỷ lệ trễ hàng bằng cách đánh giá và tối ưu hóa các quy trình và hạ tầng.

06.

Bồi thường cho khách hàng

Đối với đơn hàng giao chậm hoặc chất lượng hàng hóa không đúng, áp dụng các chính sách bồi thường cho khách hàng như tăng phiếu giảm giá hoặc miễn phí giao hàng cho đợt sau.

**“SUPPLY CHAIN:
GETTING THE RIGHT PRODUCT,
TO THE RIGHT PLACE,
IN THE RIGHT QUANTITY,
WITH THE RIGHT QUALITY,
AT THE RIGHT TOTAL COST,
AT THE BEST TIME.”**

Collected



BI55 – FINAL PROJECT

THANK YOU FOR YOUR ATTENTION

Nguyễn Ngọc Bảo Châu

Phạm Nguyên Chương

Nguyễn Thị Thu Hà

Trần Tiến Sang