

A photograph of a modern residential building with a light-colored facade and numerous balconies. The building is set against a clear blue sky with some wispy clouds. The perspective is from a low angle, looking up at the building.

# SUPERSTORE DATA ANALYSIS

GROUP 2

# **TABLE OF CONTENTS**

- Data Instructions
- Consumers Visualization
- Sales Analysis
- RFM Model



# DATA DICTIONARY

Explanation	Attribute
Mã đơn hàng	Order ID
Ngày đặt hàng	Order Date
Ngày giao hàng	Ship Date
Kiểu giao hàng	Ship Mode
Số lượng hàng được đặt	Quantity
Lượng tiền discount	Discount
Lợi nhuận	Profit
Doanh thu	Sales
Mã sản phẩm	Product ID
Phân loại sản phẩm	Category
Tên sản phẩm	Product Name

Explanation	Attribute
Phân loại dựa trên Category	Sub-Category
Mã phân biệt khách hàng	Customer ID
Tên khách hàng	Customer Name
Phân loại khách hàng	Segment
Tuổi của khách hàng	Age
Quốc gia	Country
Thành phố của khách hàng	City
Bang	State
Mã bưu chính	Postal Code
Khu vực	Region
Nhóm tuổi	Age Group

# CLEANING DATA

## Mismatching Data

Xử lí các dữ liệu không khớp ở cột State và City

## Missing value

Xử lí các dữ liệu trống ở cột Customer postal code

## Datatype

Xử lí sự không đồng nhất về định dạng ở cột Order Date, Ship Date

## Outliner

Sau khi kiểm tra với Box Plot thì không có dữ liệu outlier nào nằm ngoài khoảng 99% giá trị của tập dữ liệu

## Duplicates

Loại bỏ các dữ liệu bị trùng lặp

# CUSTOMER VISUALIZATION

- Demographic
- Buying Behavior

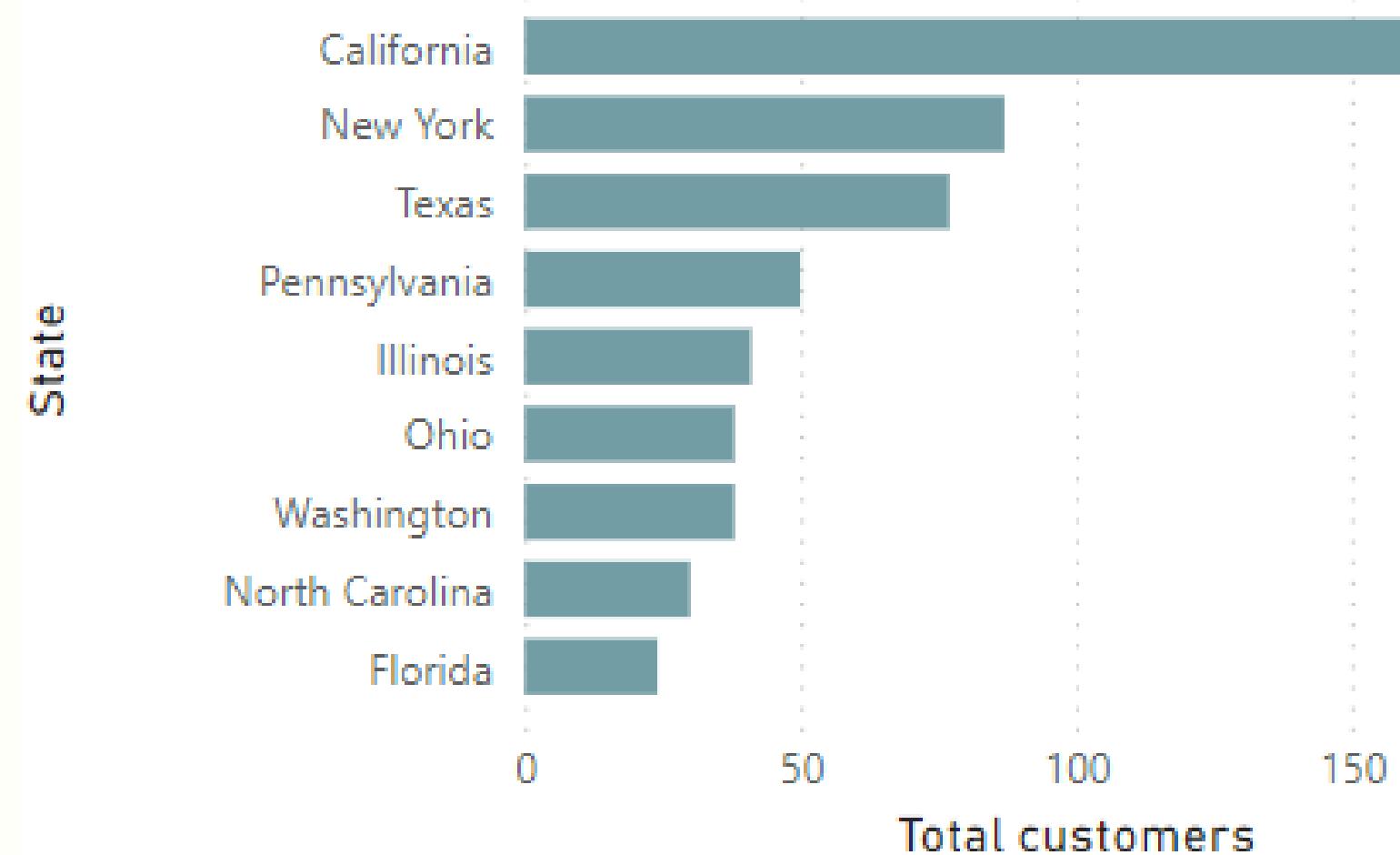


# DEMOGRAPHIC



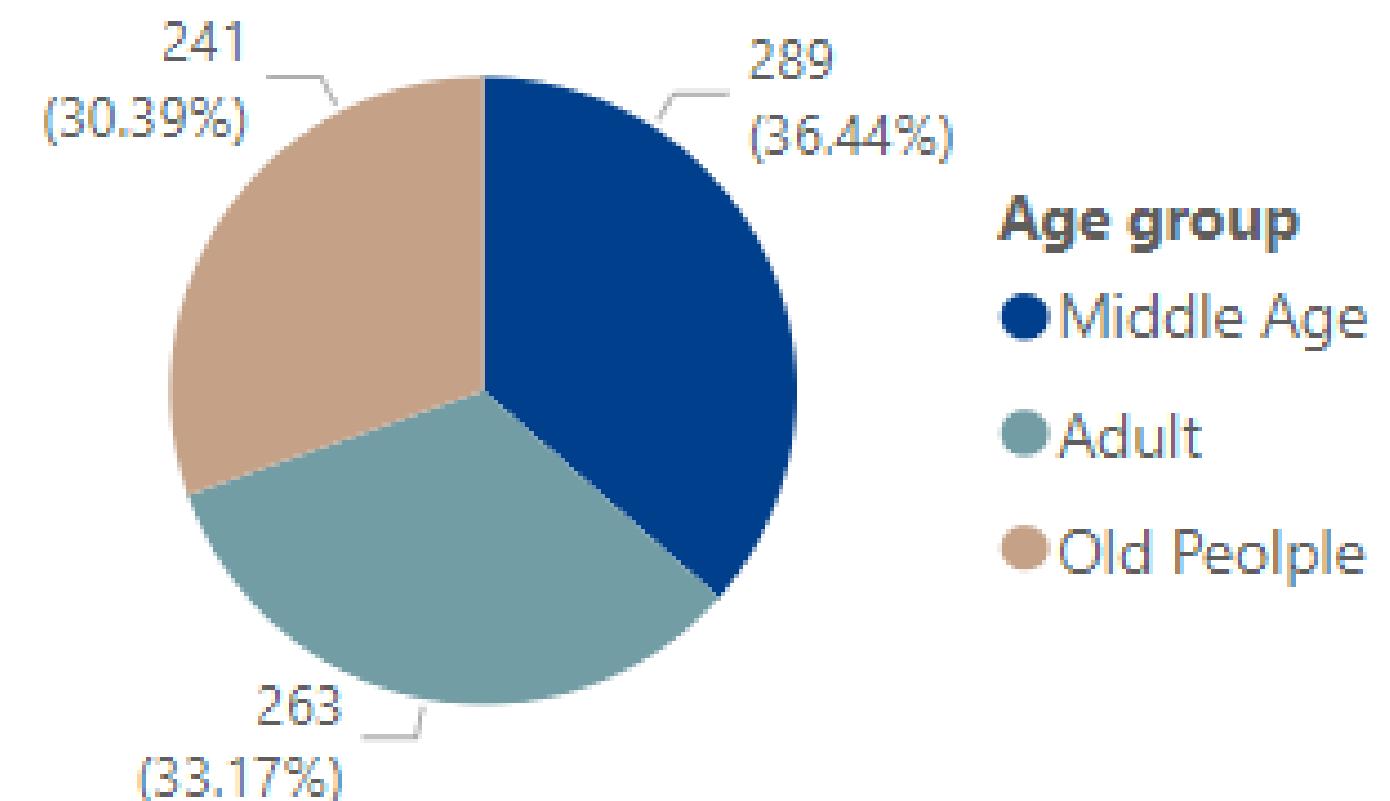
**Khách hàng tập trung chủ yếu ở các tiểu bang lớn, có thói quen mua sắm nhiều. Khoảng 20% đến từ California, hơn 30% đến từ 4 tiểu bang còn lại trong top 5.**

Total customers by State



**Có thể thấy rằng nhóm khách hàng được phân bổ đều theo 3 lứa tuổi, không có sự chênh lệch rõ rệt**

Total customers by Age group

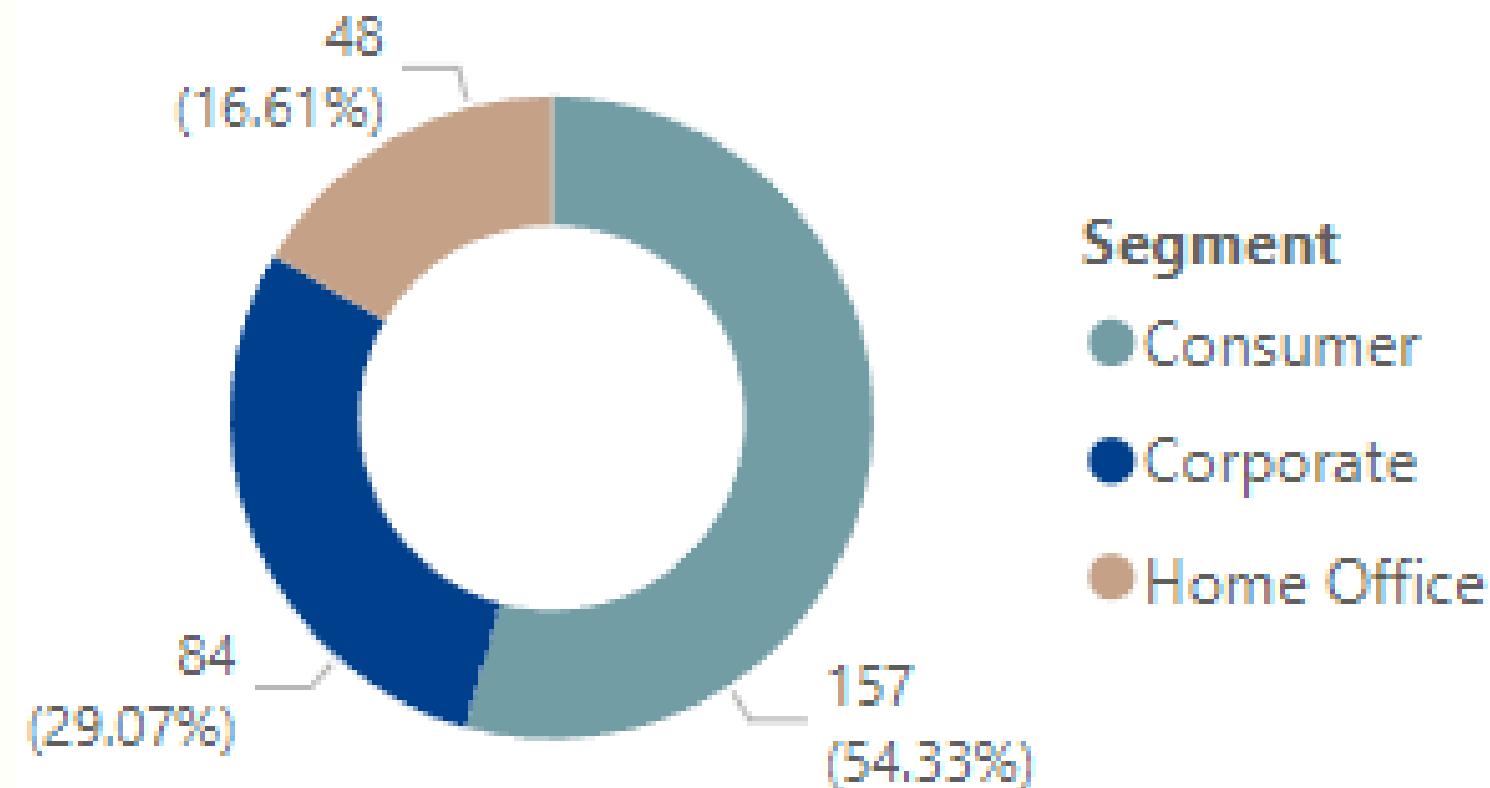


# DEMOGRAPHIC



Nhóm "Người tiêu dùng" chiếm tỷ trọng cao đáng kể. Trong khi đó, các văn phòng tại gia và doanh nghiệp là những nhóm khách hàng công ty cần phải mở rộng thêm với chỉ khoảng 16% và 29%.

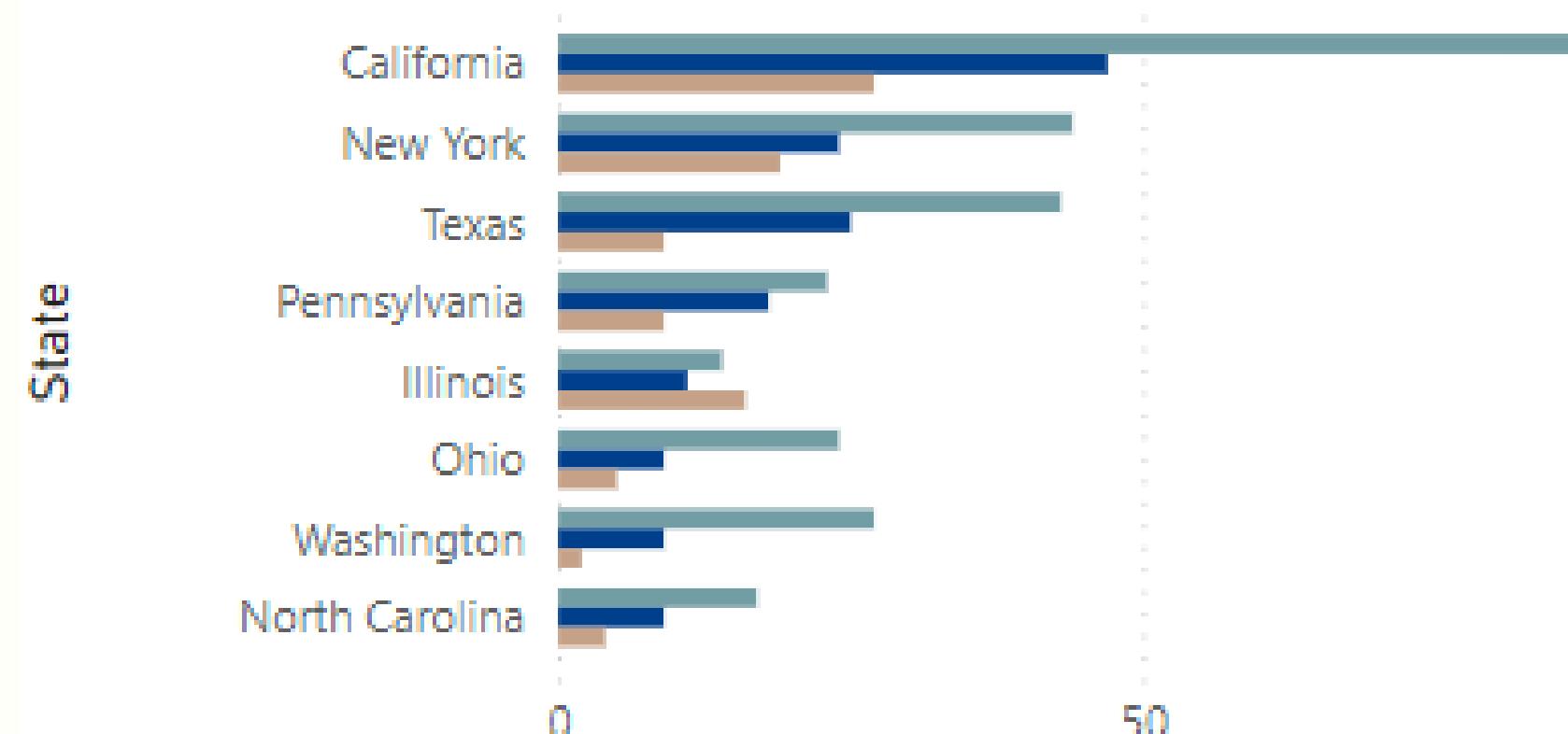
Total customers by Segment



Lượng người trong nhóm khách hàng "Consumers" ở hầu hết các tiểu bang đều chiếm tỷ trọng cao nhất ngoại trừ Illinois

Total customers by State and Segment

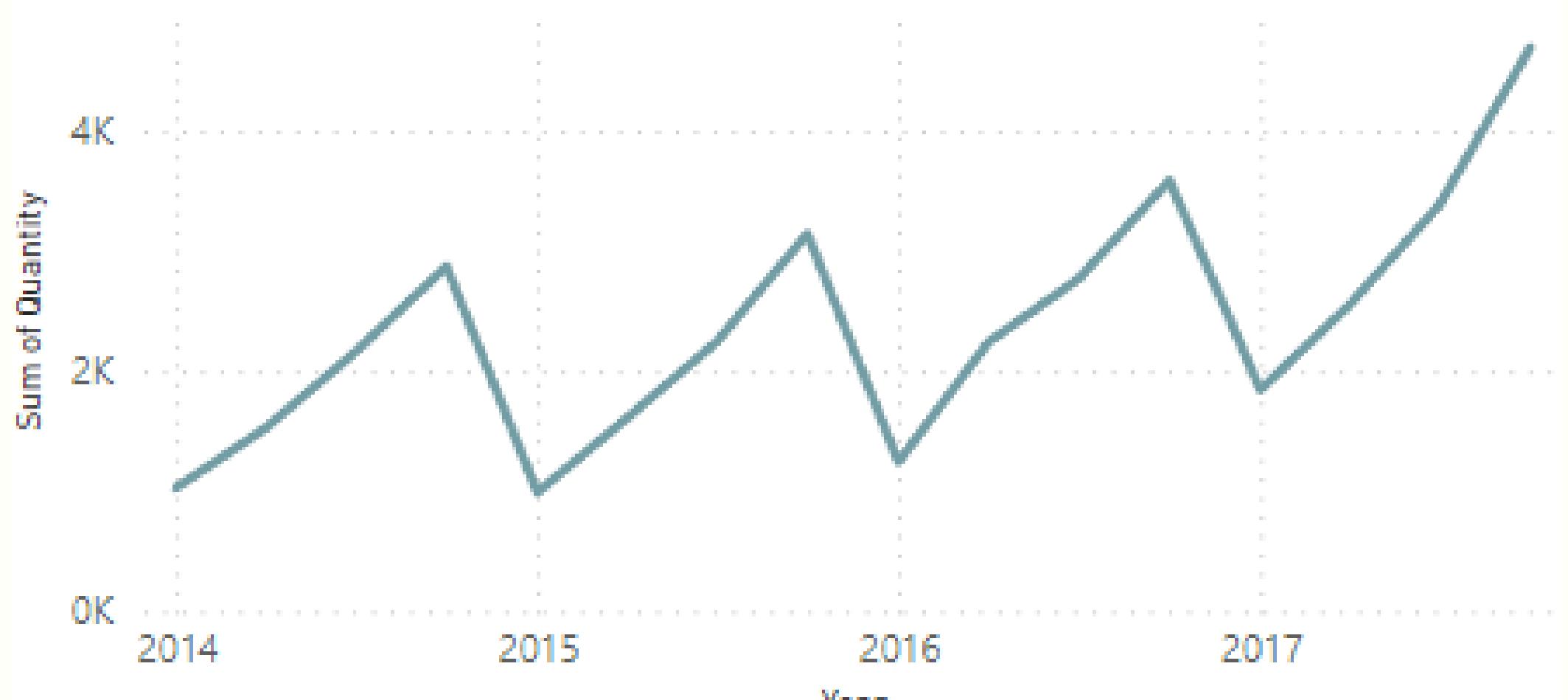
Segment ● Consumer ● Corporate ● Home Office



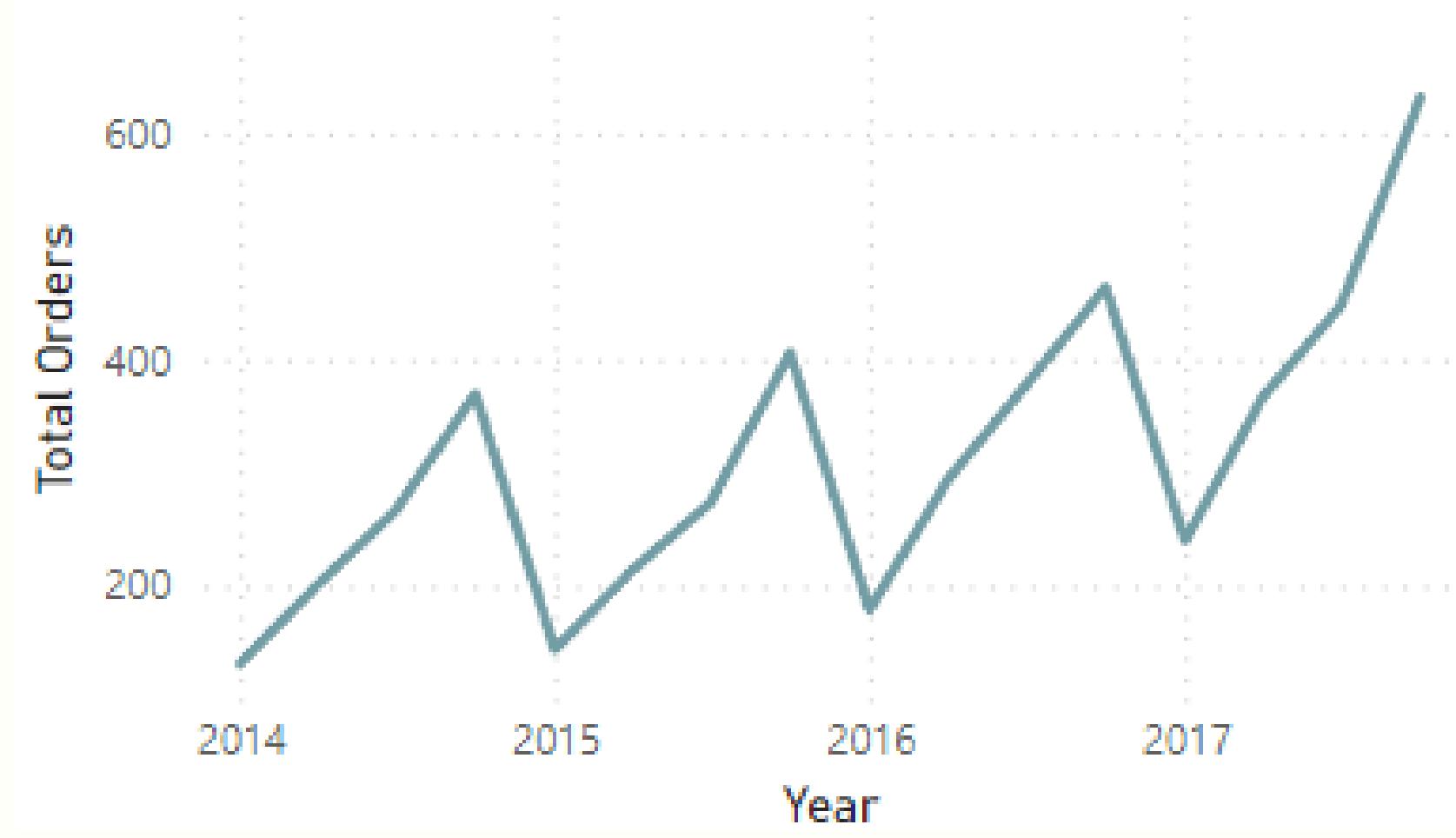
# BUYING BEHAVIOR

Số lượng hàng được đặt có xu hướng tăng dần. Nhu cầu về hàng hóa tại cửa hàng ở thời điểm cuối năm tăng rõ rệt so với đầu năm. Có thể thấy xu hướng này xảy ra do càng gần về cuối năm thì càng có nhiều chương trình giảm giá.

Sum of Quantity by Year and Quarter



Total Orders by Year and Quarter

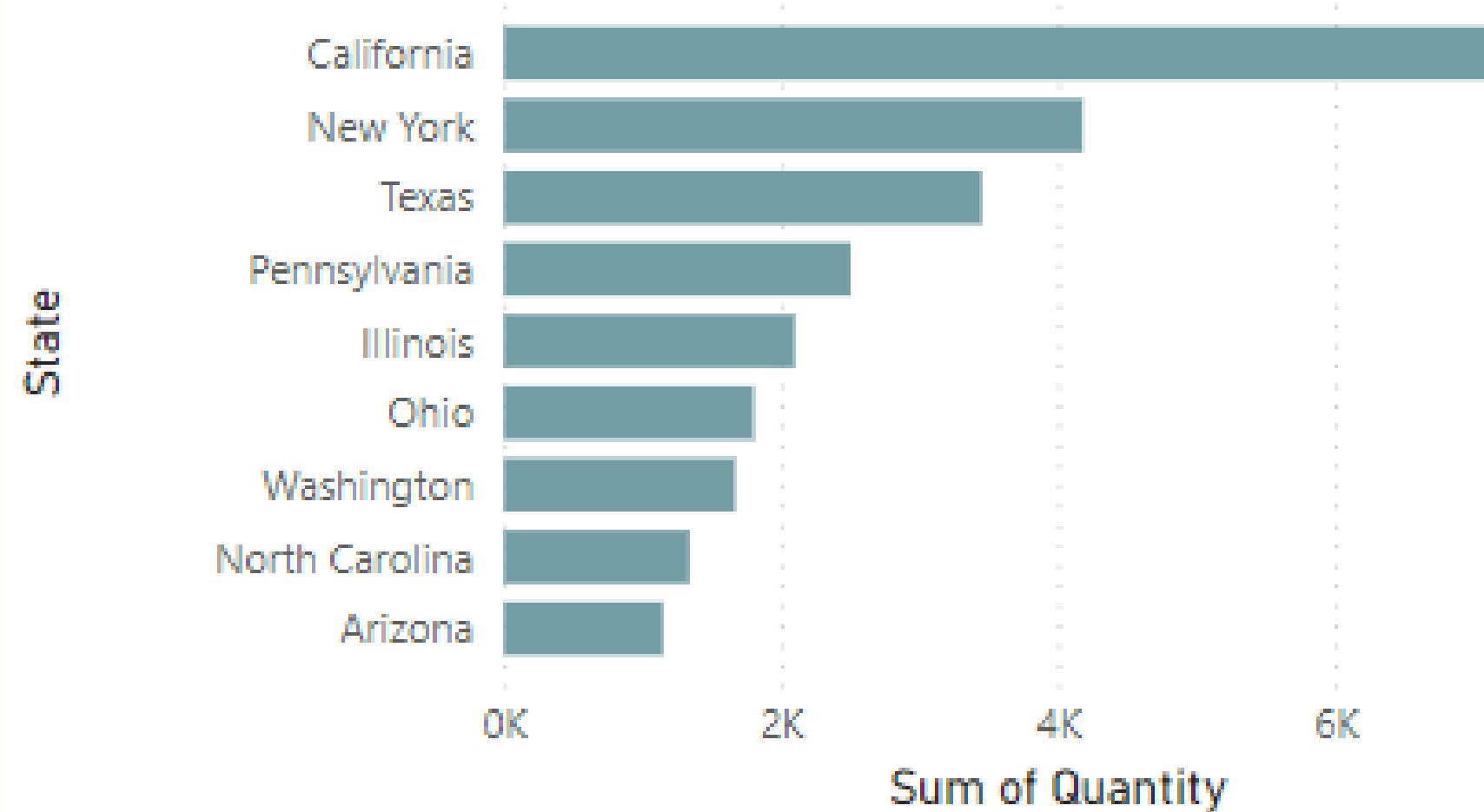


Cùng với số lượng hàng thì số lượng đơn hàng cũng có xu hướng tương tự, tăng dần từ đầu đến cuối năm.

# BUYING BEHAVIOR

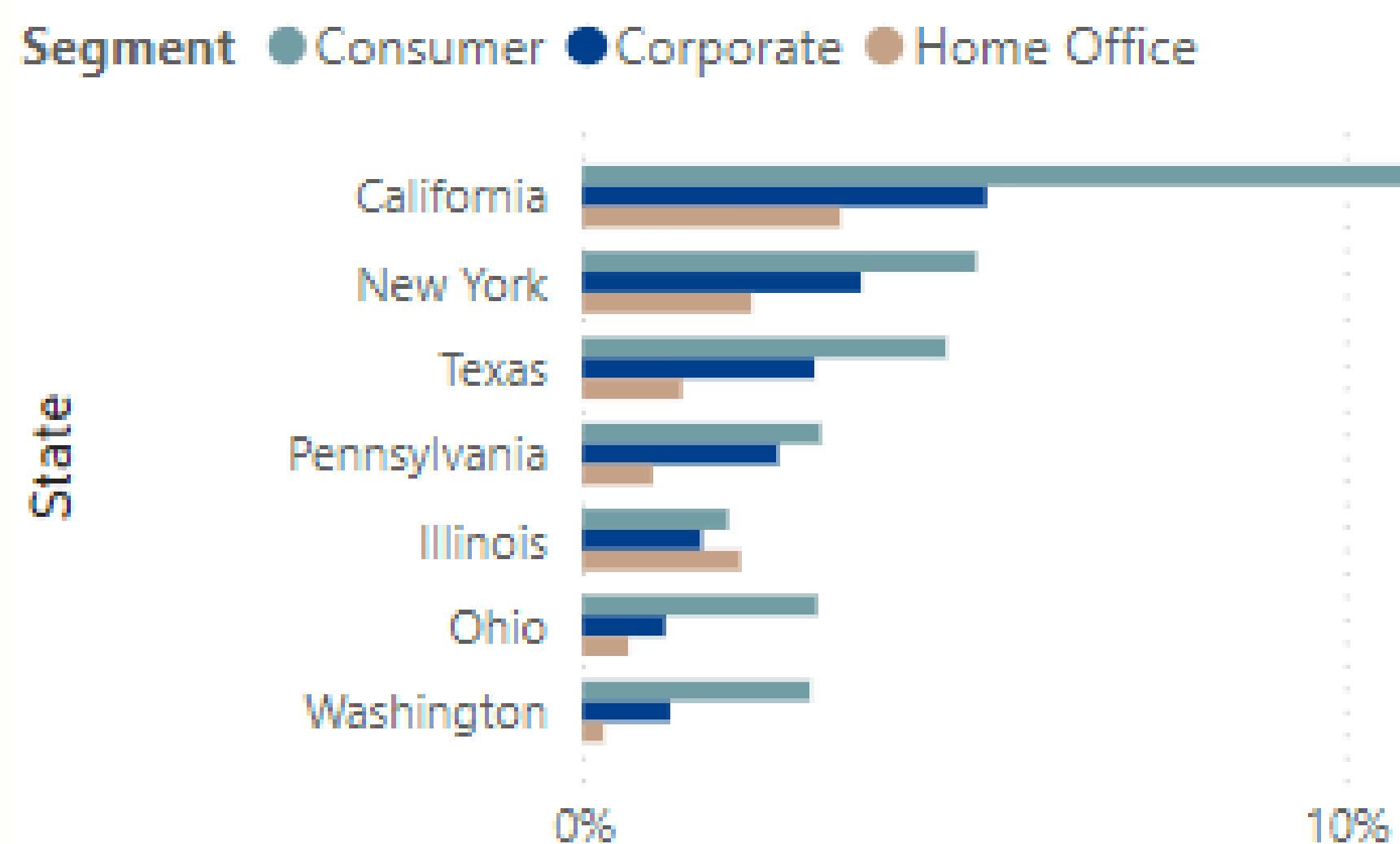
🛒 Số lượng đơn hàng có sự tương đồng rõ rệt đối với số lượng khách hàng.

Sum of Quantity by State



🛒 Nhóm khách hàng "Consumer" tiêu thụ hàng hóa đặc biệt nhiều ở tiểu bang Cali(11% tổng lượng hàng). Trong khi đó ở các tiểu bang khác không có sự chênh lệch quá lớn giữa các nhóm khách hàng

% GT Sum of Quantity by State and Segment

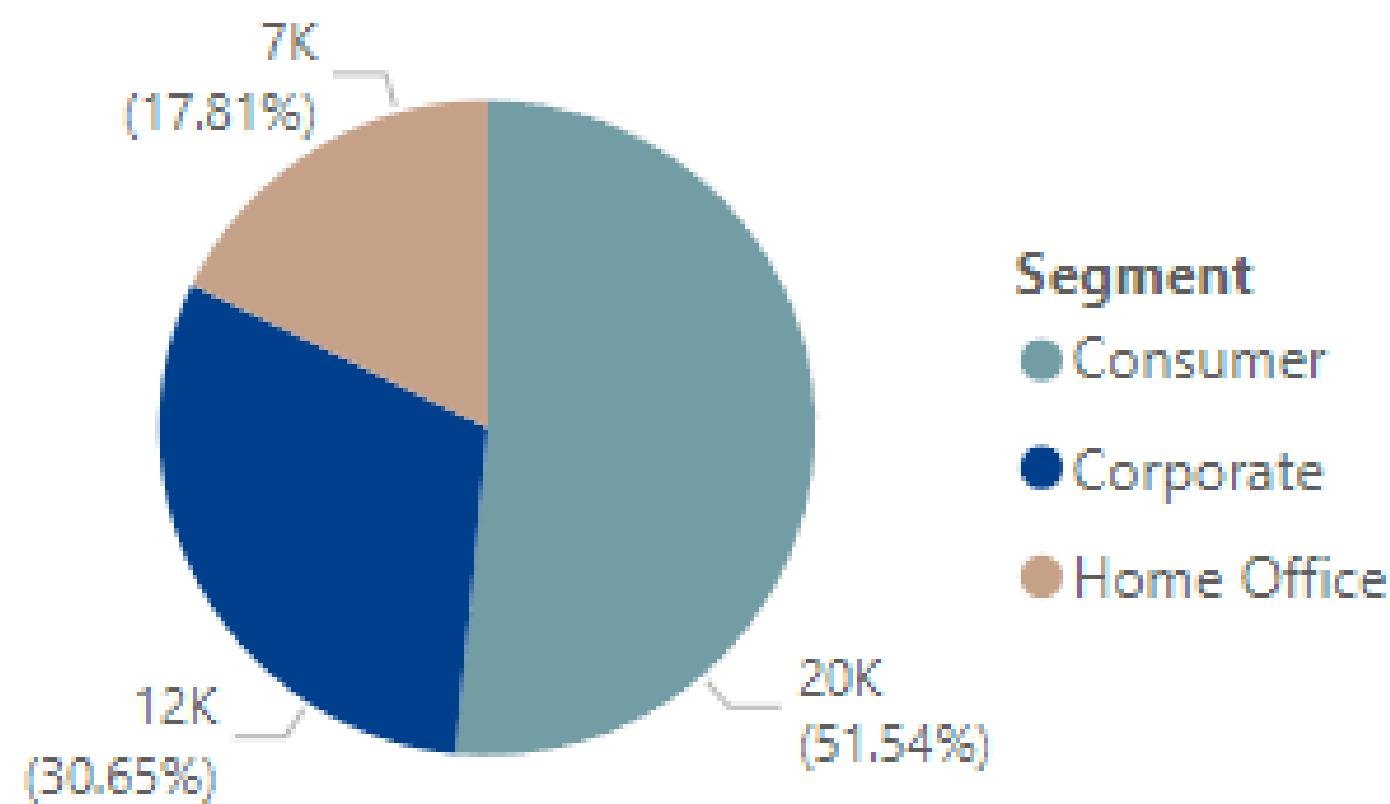


# BUYING BEHAVIOR

**Nhu cầu mua hàng của nhóm khách**

 **hàng "Consumers" chiếm tỷ lệ cao nhất với hơn 50%, trong khi đó 2 nhóm khách hàng còn lại chỉ ở quanh khoảng 20%**

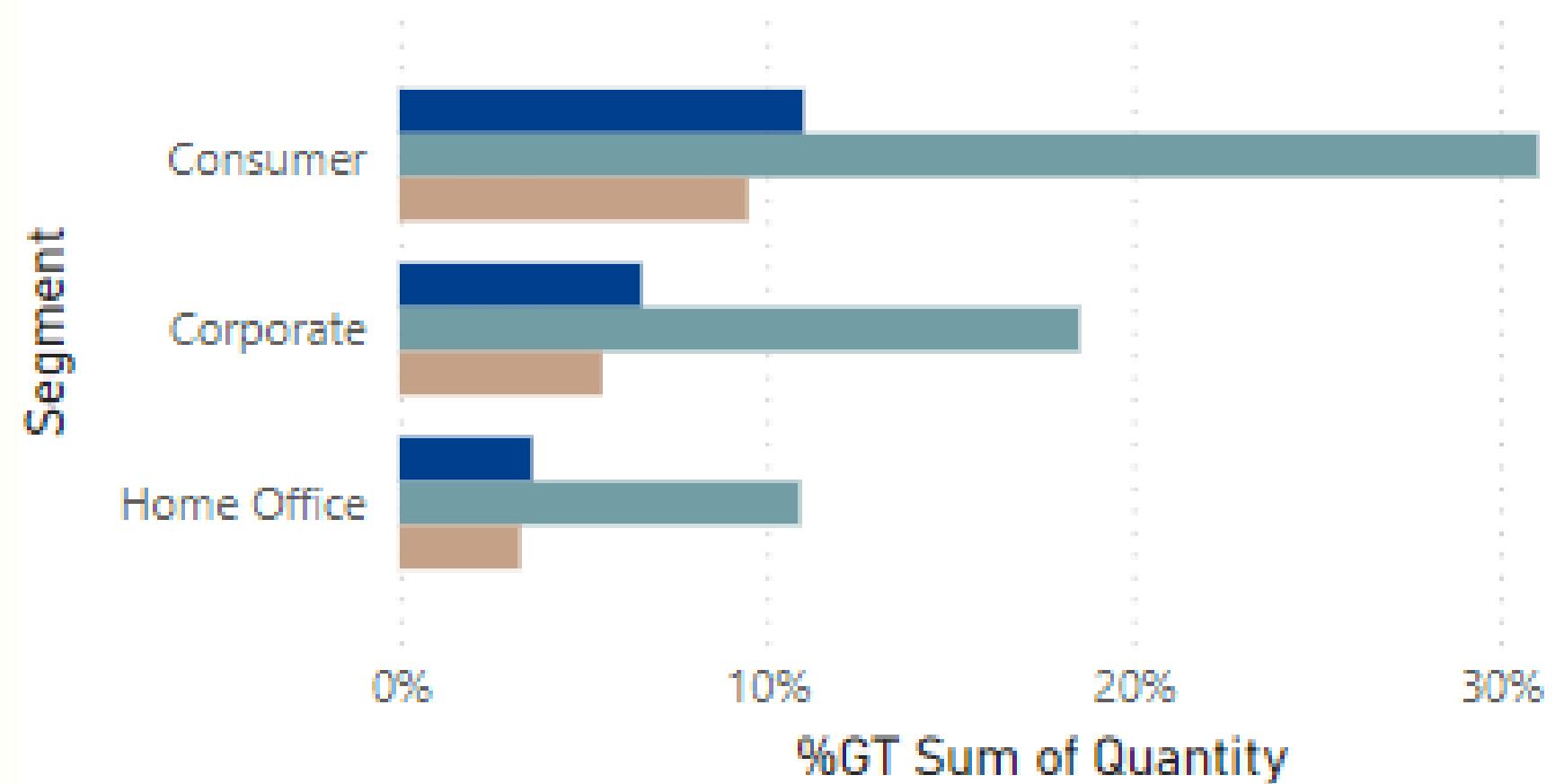
Sum of Quantity by Segment



 **Nhóm khách hàng "Consumers" tiêu thụ văn phòng phẩm nhiều hơn hẳn so với các nhóm còn lại, chiếm hơn 30% tổng lượng hàng tiêu thụ.**

%GT Sum of Quantity by Segment and Category

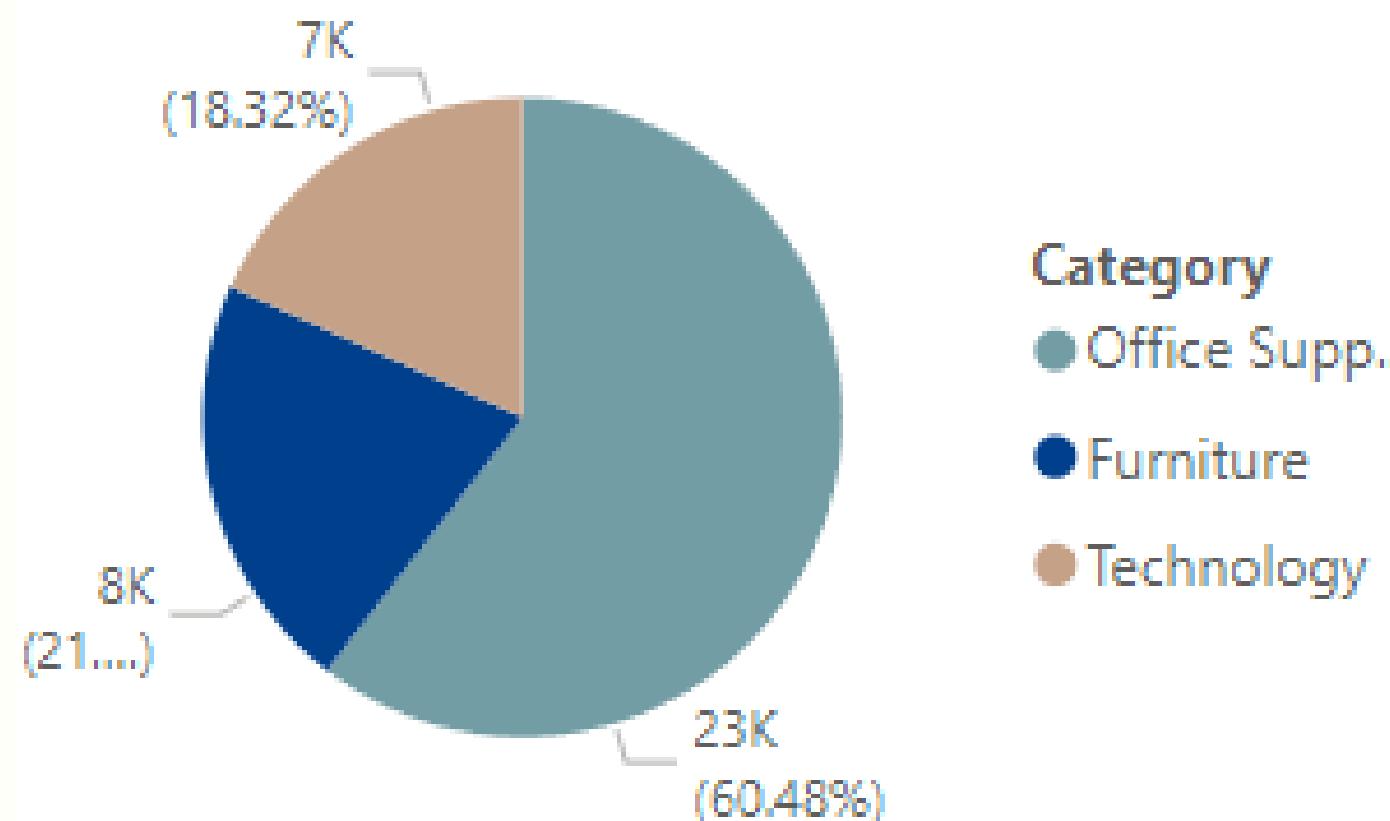
Category ● Furniture ● Office Supplies ● Technology



# BUYING BEHAVIOR

Nhu cầu về Văn phòng phẩm tại cửa hàng là lớn nhất với hơn 60%. Các ngành hàng nội thất và công nghệ có tỷ lệ ở quanh khoảng 20%.

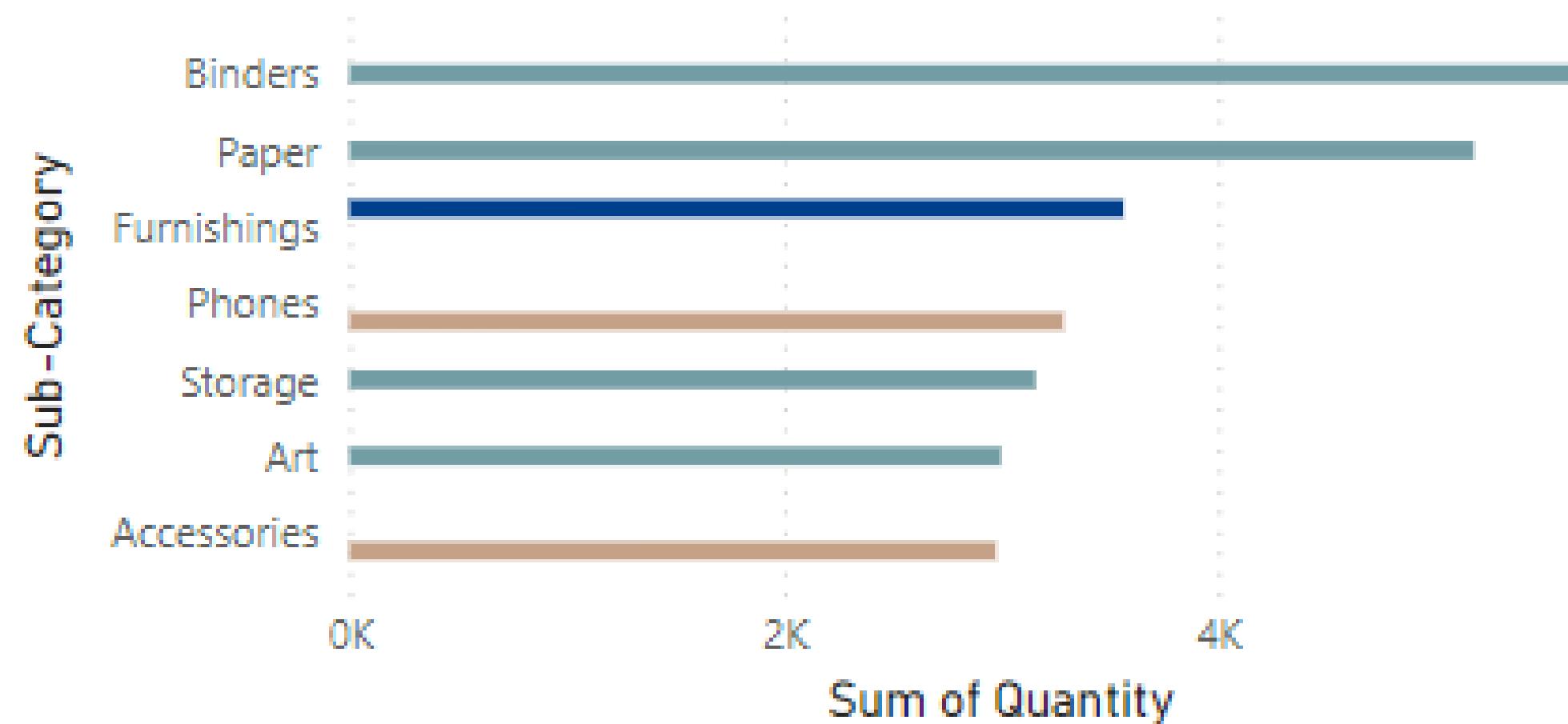
Sum of Quantity by Category



Ngoài các sản phẩm thuộc văn phòng phẩm (Giấy, kẹp,...) thì điện thoại và phụ kiện (thuộc nhóm Technology) cũng khá được ưa chuộng

Sum of Quantity by Sub-Category and Category

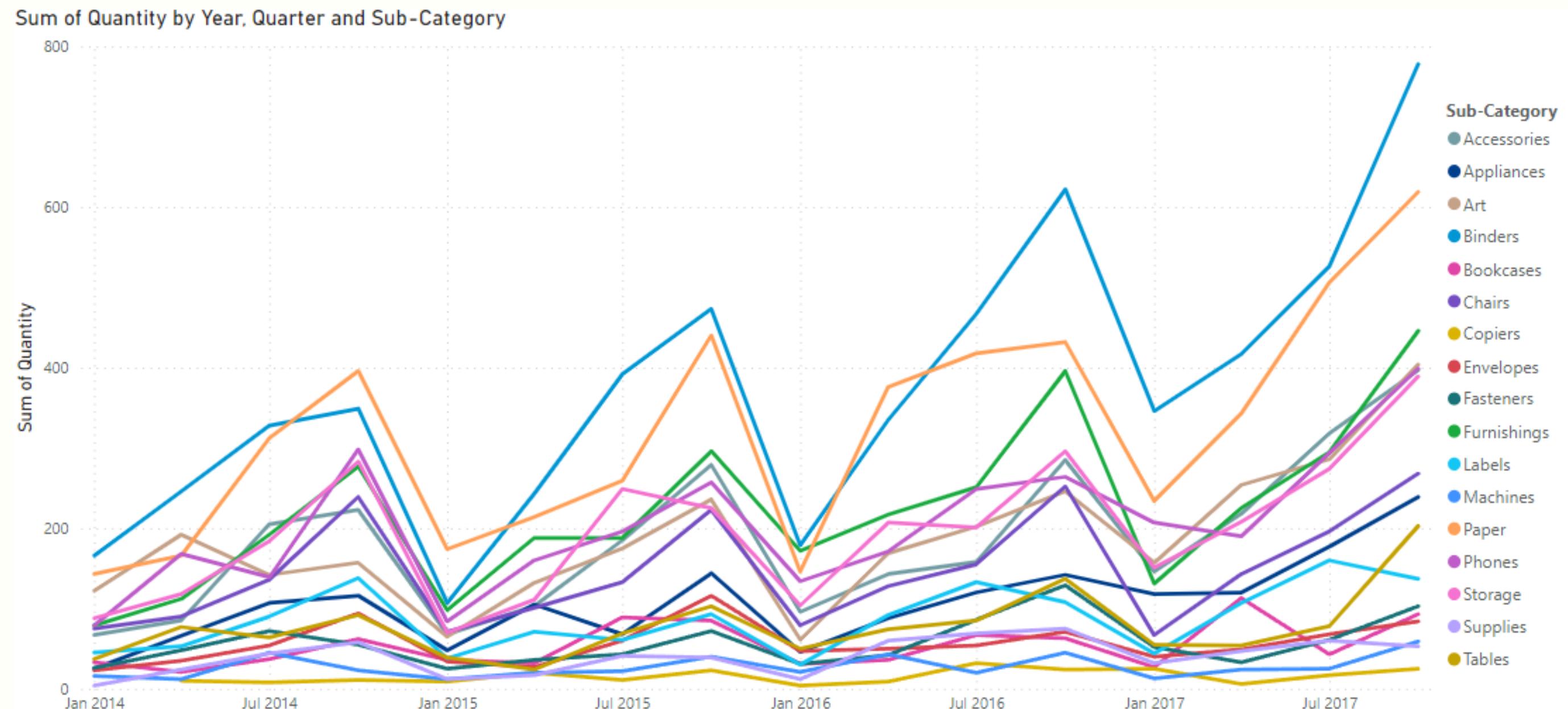
Category ● Furniture ● Office Supplies ● Technology



# BUYING BEHAVIOR



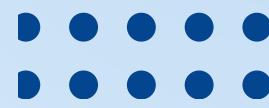
**Nhìn chung thì xu hướng mua với các  
loại hàng hóa là khá tương đồng với  
nhau và giống với xu hướng mua tổng  
thể**



# **SALES ANALYSIS**

**Sales overview - Product -  
Customer - Time and Region of  
sales - RFM analysis**

# Sales Overview



**286.40K**

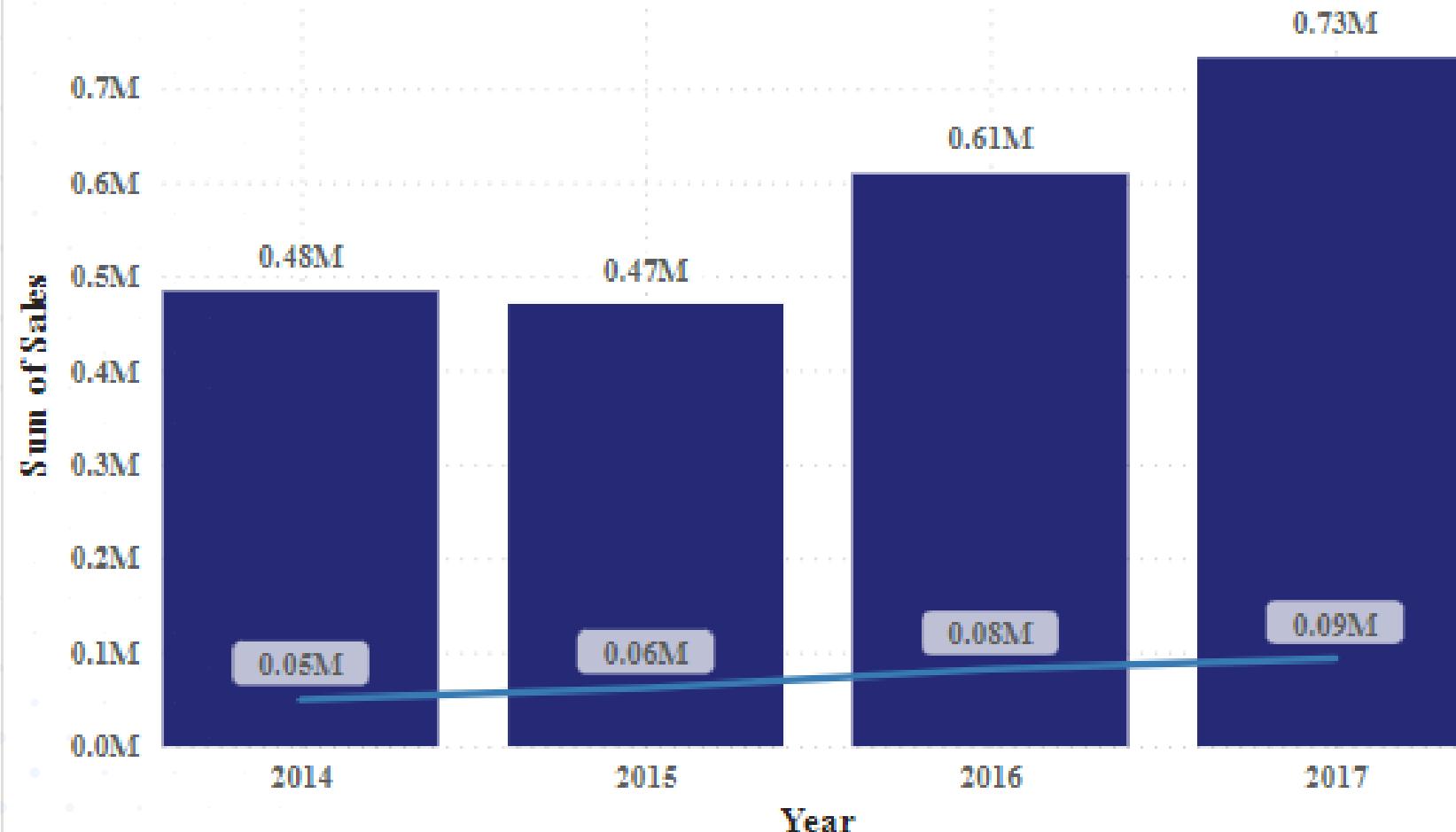
Sum of Profit

**1.56K**

Sum of Discount

Sum of Sales and Sum of Profit by Year

● Sum of Sales ● Sum of Profit



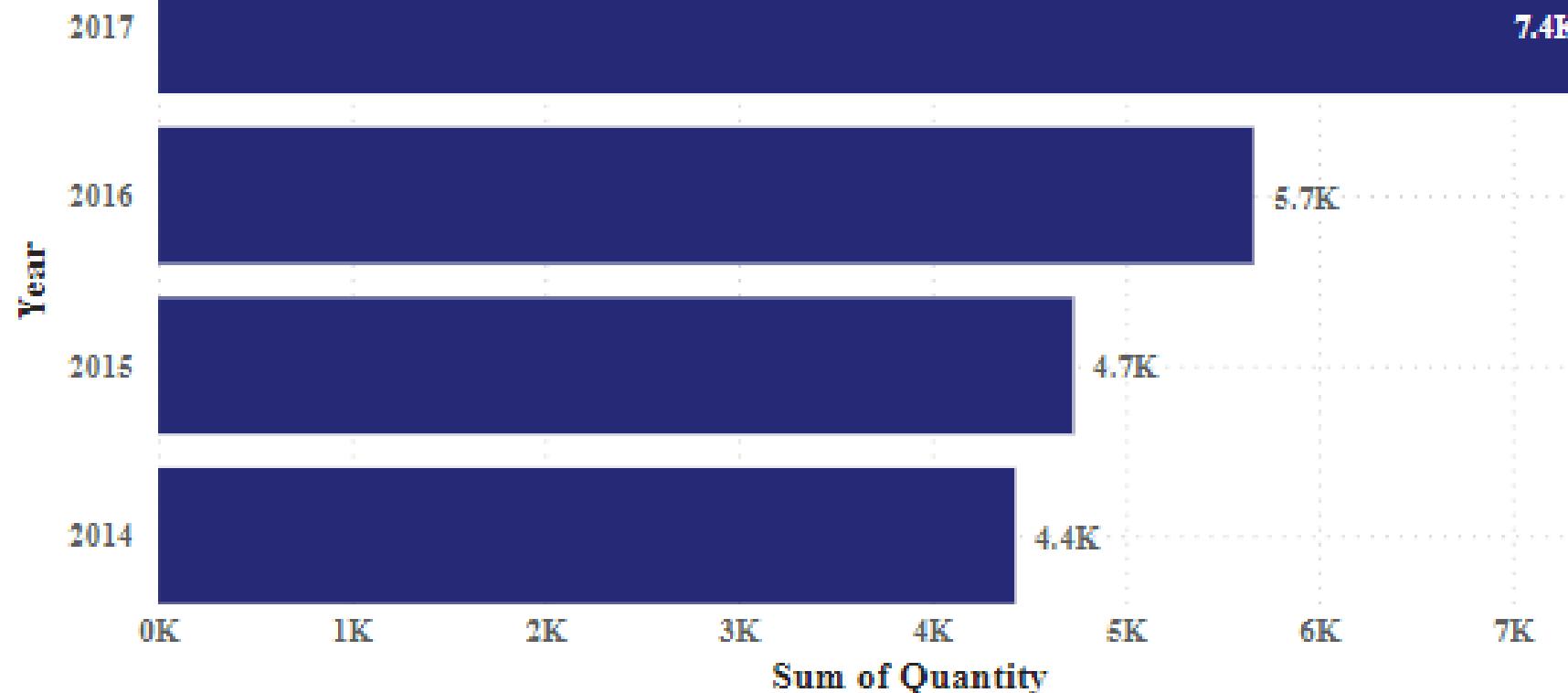
**2.30M**

Sum of Sales

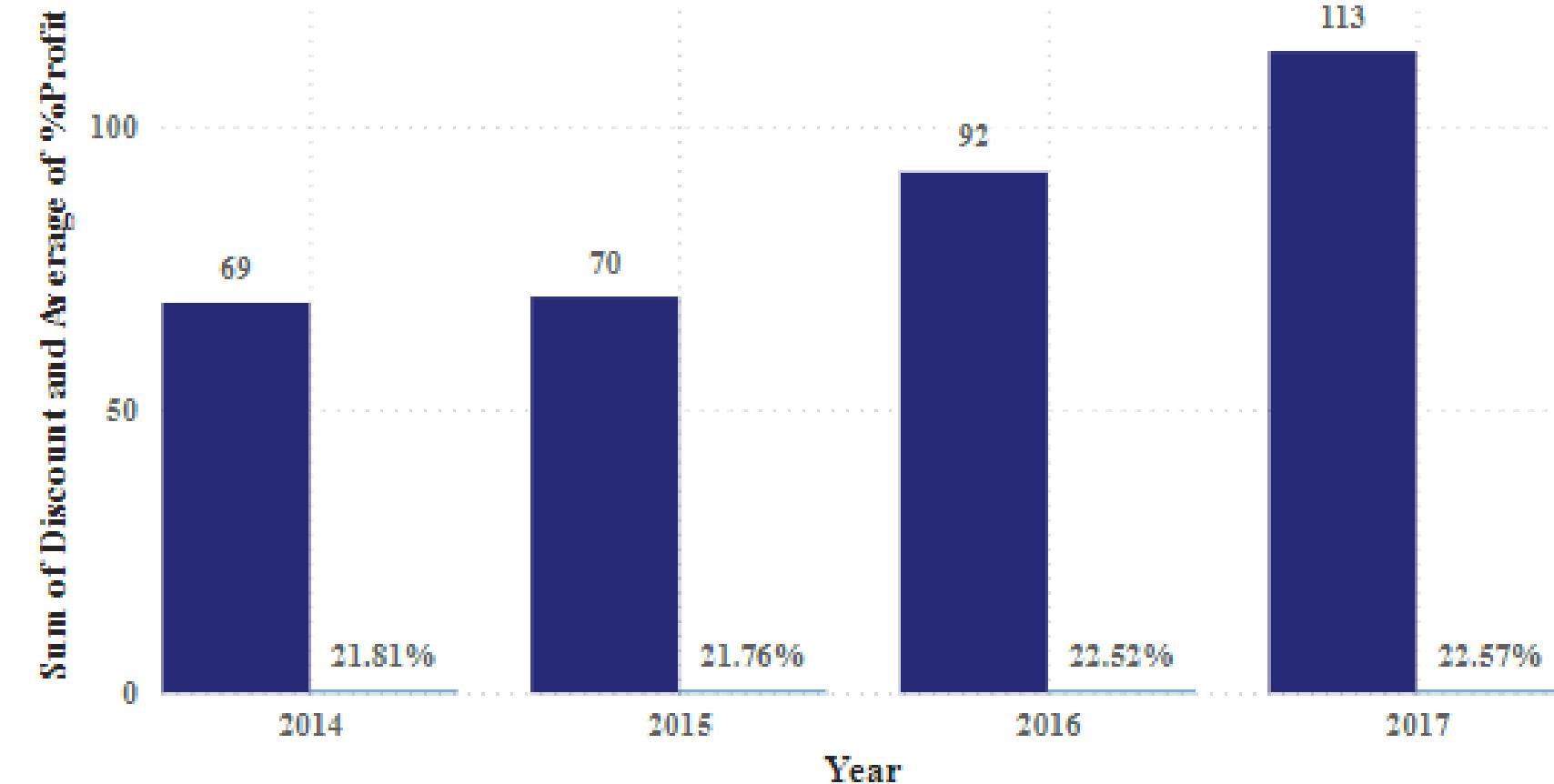
**38K**

Sum of Quantity

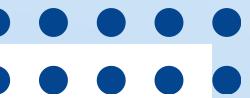
Sum of Quantity by Year



● Sum of Discount ● Average of %Profit

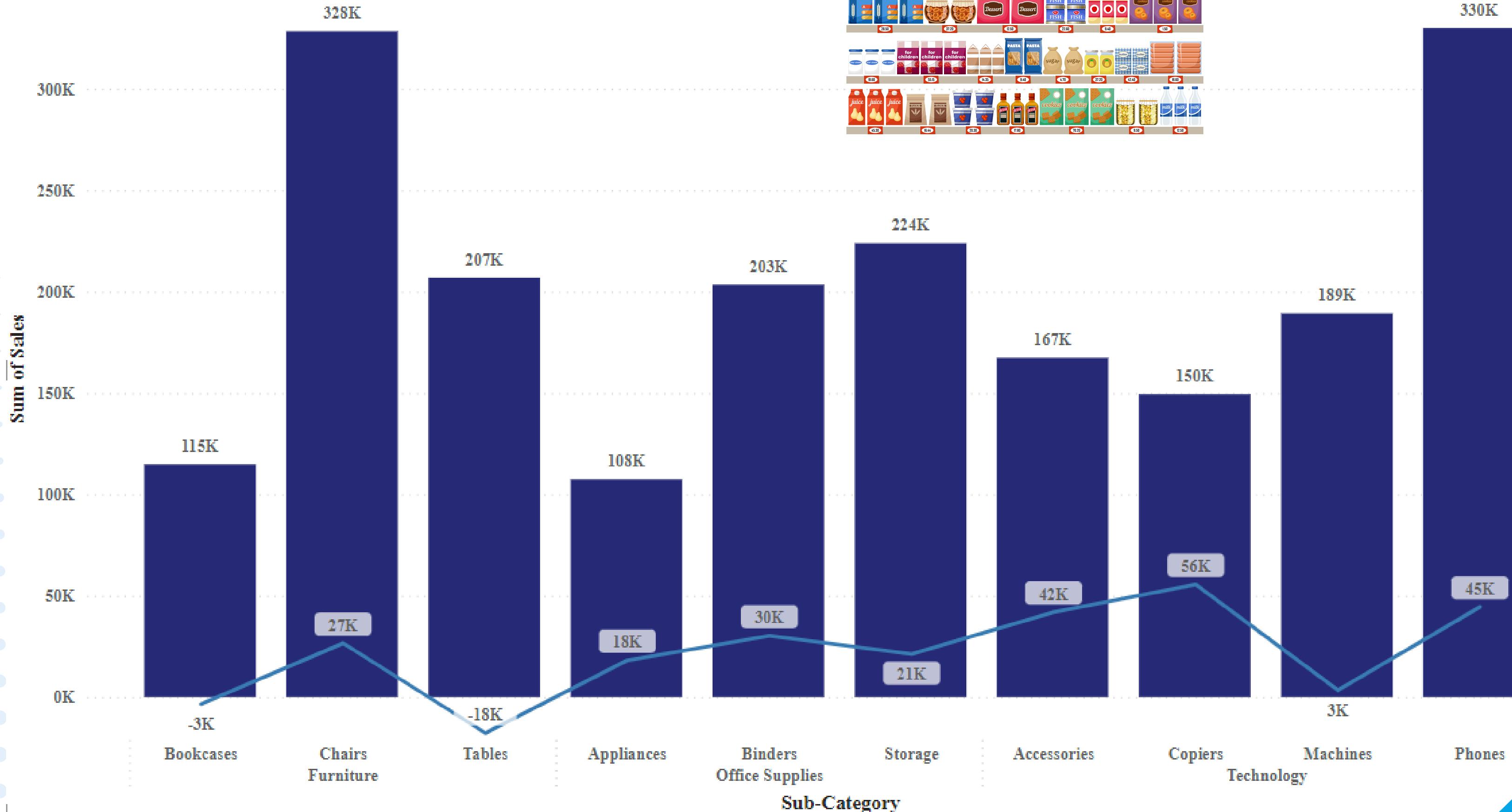


# Product



## Sum of Sales and Sum of Profit by Category and Sub-Category

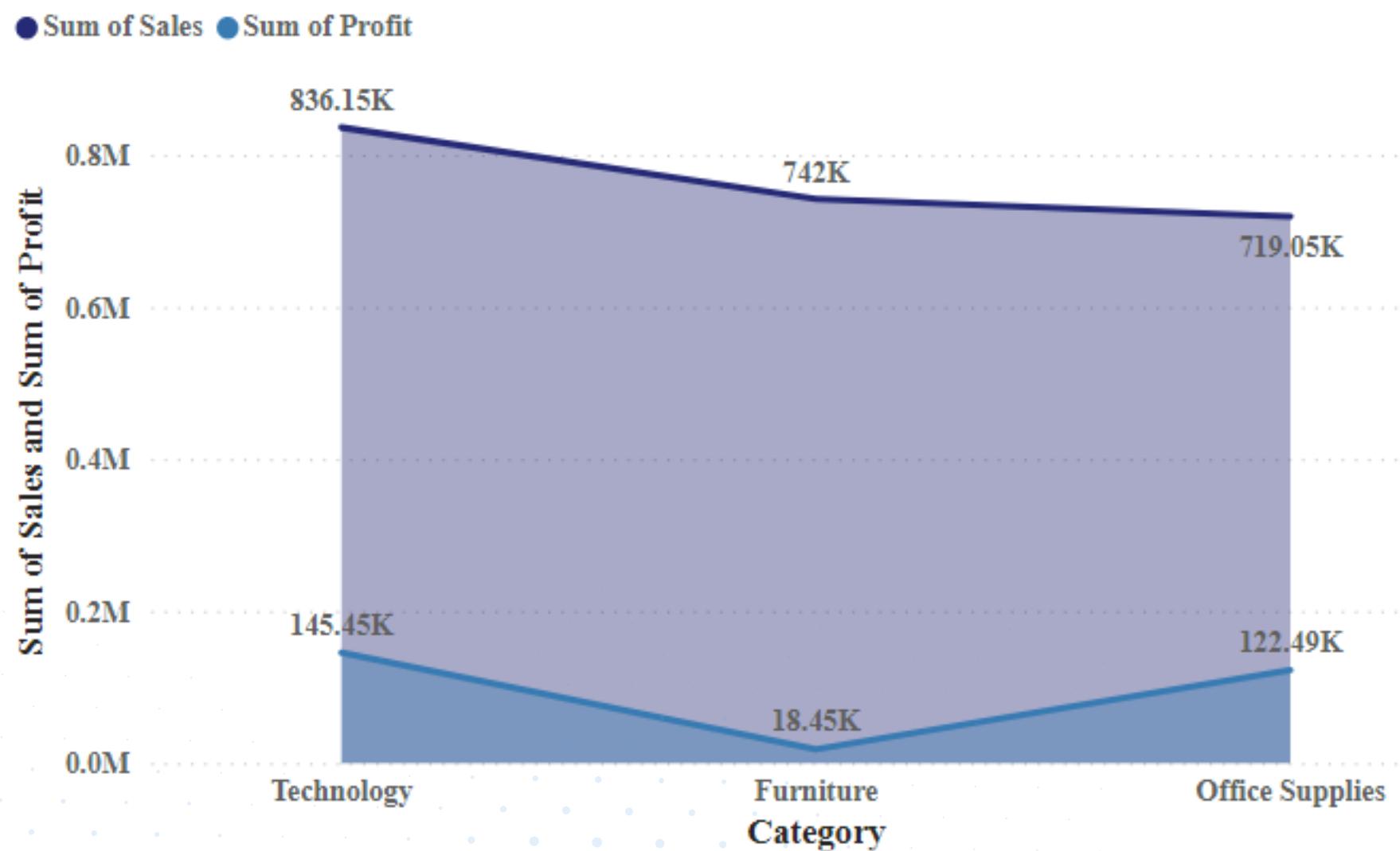
● Sum of Sales ● Sum of Profit



# Product

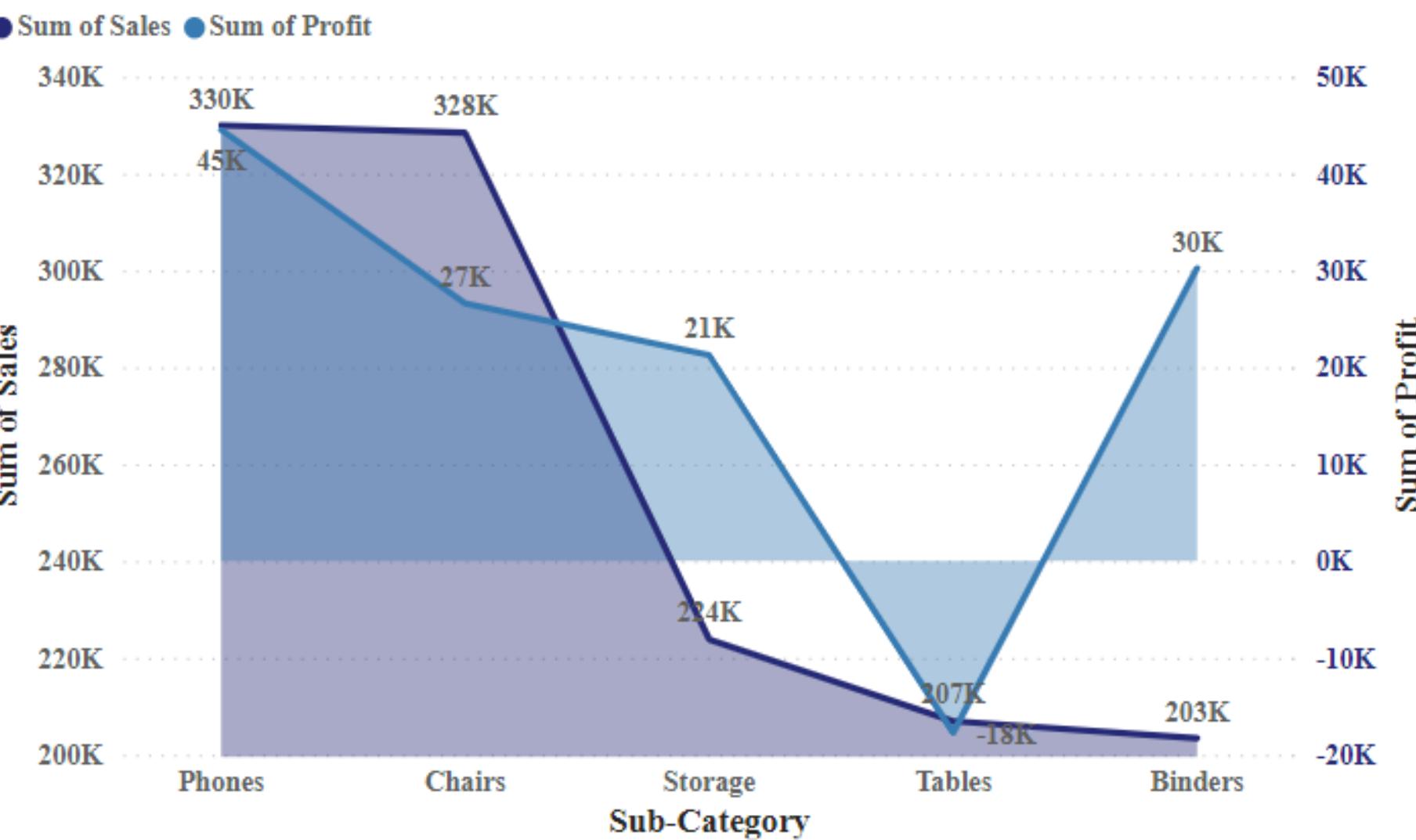
**Doanh thu và lợi nhuận của 3 loại sản phẩm: Furniture, Office Supplys, Technology tăng đều qua các năm từ 2014-2017, doanh thu cao nhất là Technology chiếm 36.34% lợi nhuận 50.7%**

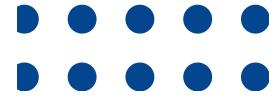
Sum of Sales and Sum of Profit by Category



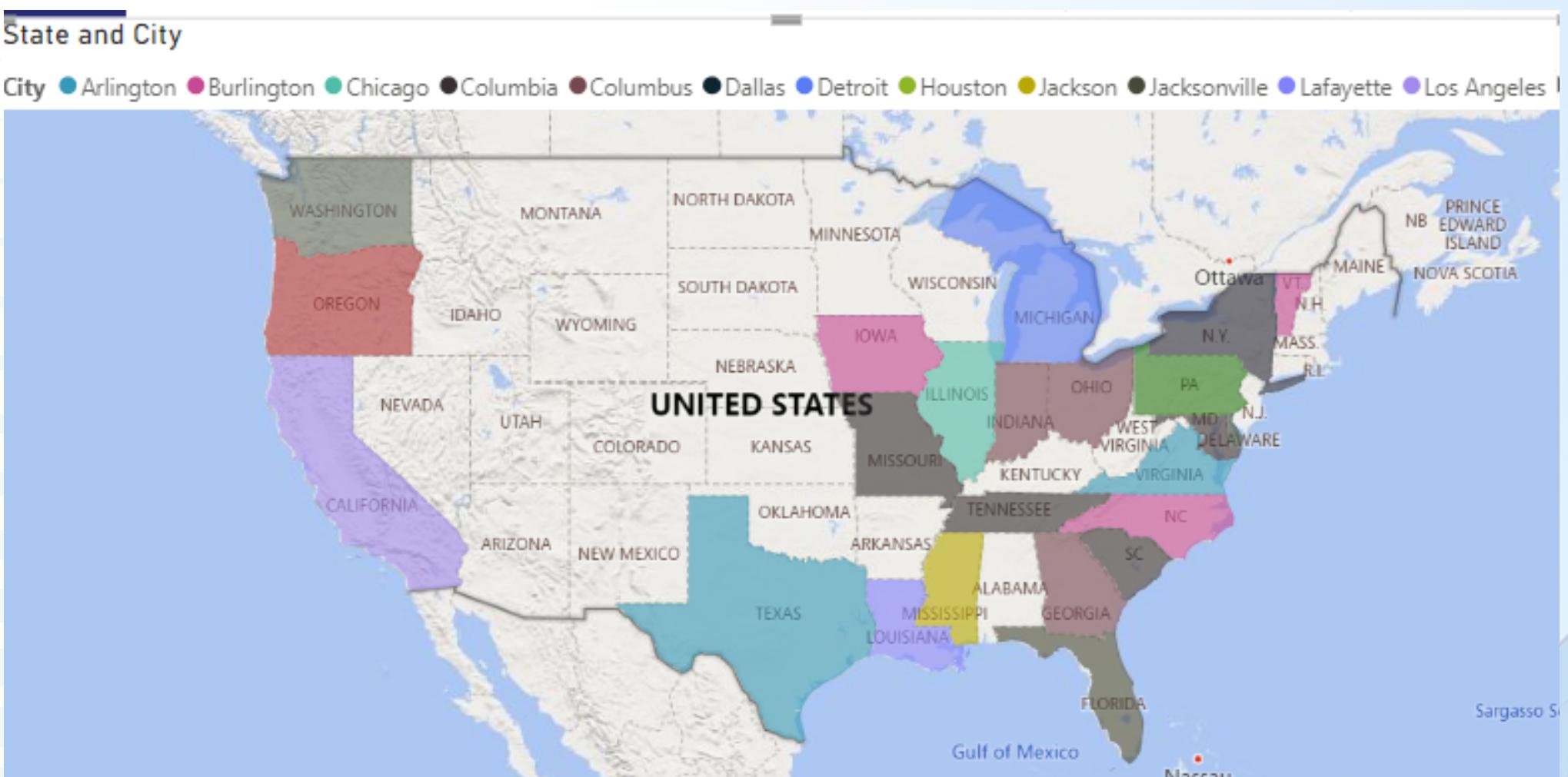
**Doanh thu và lợi nhuận của các sản phẩm: Phones, Chairs, Storage,...tăng đều qua các năm từ 2014-2017, doanh thu cao nhất là phones chiếm 14.35% lợi nhuận 15.8%, tuy nhiên Tables và Bookcares có lợi nhuận âm 18k \$ và 3k \$, doanh số bán ít và nhiều chính sách giảm giá của cửa hàng.**

Sum of Sales and Sum of Profit by Sub-Category

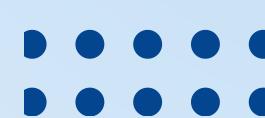




# Time and Region analysis

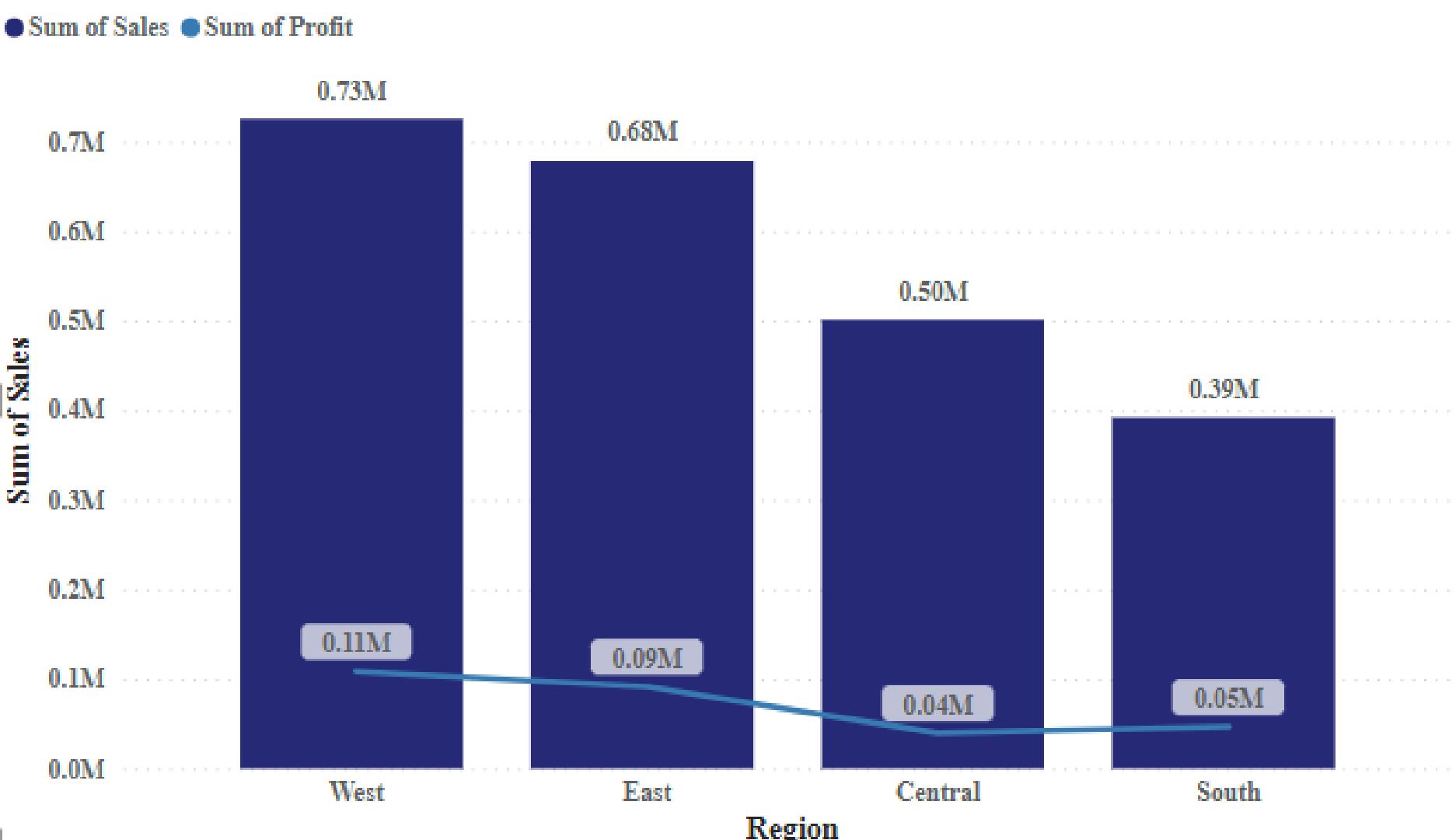


# **REGION - STATE - CITY**

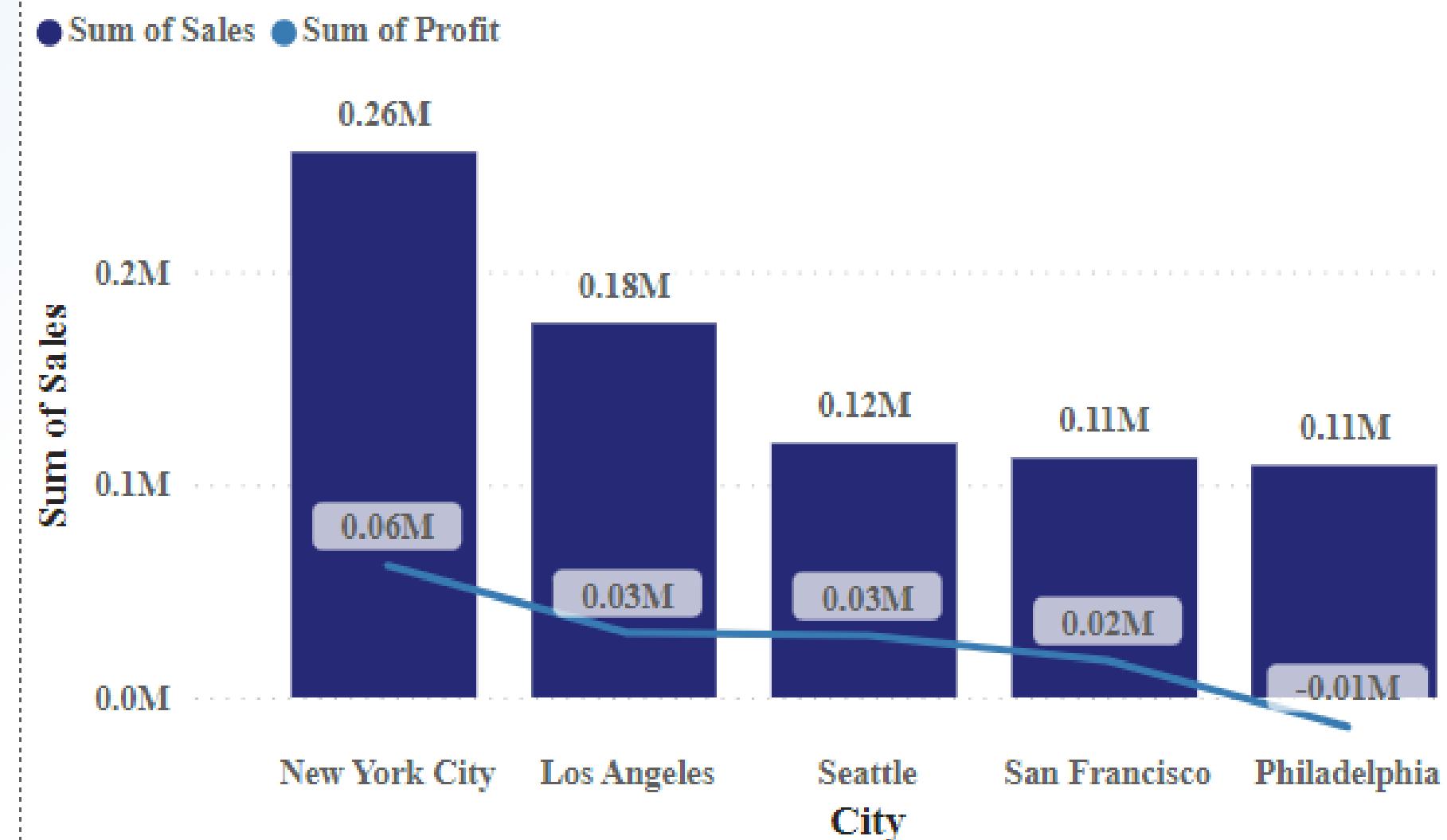


**Doanh thu cao nhất trong 4 vùng miền West, East, Central, South là West với 0.73M \$, chiếm 31.7%, lợi nhuận chiếm 37.93%. New York City là thành phố có doanh thu và lợi nhuận cao nhất, do là nơi giàu nhất, tập trung dân đông và có nhu cầu tiêu dùng lớn nhất, Philadelphia có lợi nhuận âm 0.01M\$ là do là nơi đông dân nhưng nhiều vụ vấn đề về chính trị, thu nhập của người dân mức trung bình thấp so với các thành phố lớn ở Mỹ, và siêu thị đưa ra các chính sách dicount lớn tại đây.**

Sum of Sales and Sum of Profit by Region



Sum of Sales and Sum of Profit by City



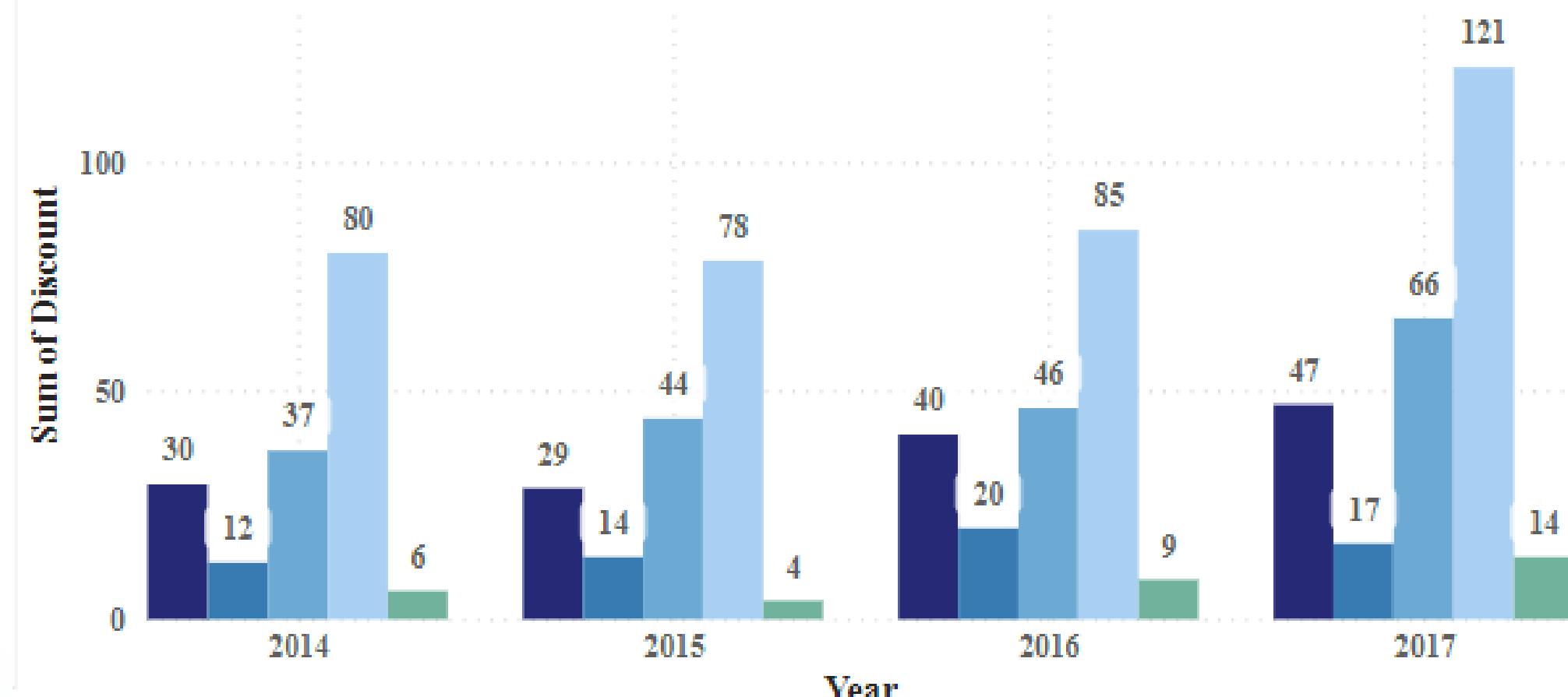
# Time and State

Lợi nhuận và tỷ lệ lợi nhuận của các bang California, New York và Washington tăng đều qua các năm, đây cũng là các bang có doanh số chiếm tỷ trọng doanh thu cao nhất tại Mỹ.

Có 2 bang là Pennsylvania và Texas có lợi nhuận đều âm ở các năm, trung bình âm từ 4K-5K \$/năm, điều này do doanh số 2 bang này khá ít và tỷ lệ discount tại 2 bang này khá lớn, trung bình là 48\$/năm tại Pennsylvania và 86\$/năm tại Texas. Ngoài ra, các mặt hàng ở 2 bang này cũng có nhiều chính sách giảm giá như Tables và Bookcares.

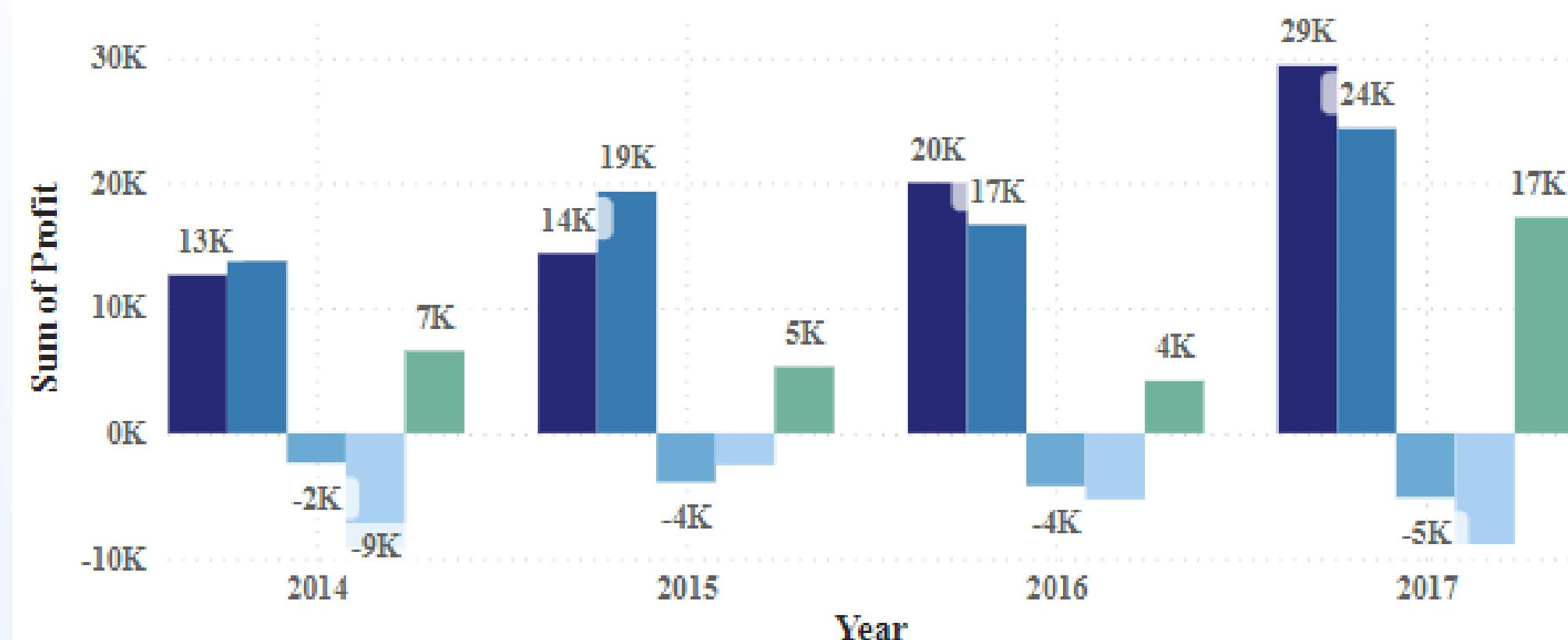
Sum of Discount by Year and State

State ● California ● New York ● Pennsylvania ● Texas ● Washington



Sum of Profit by Year and State

State ● California ● New York ● Pennsylvania ● Texas ● Washington

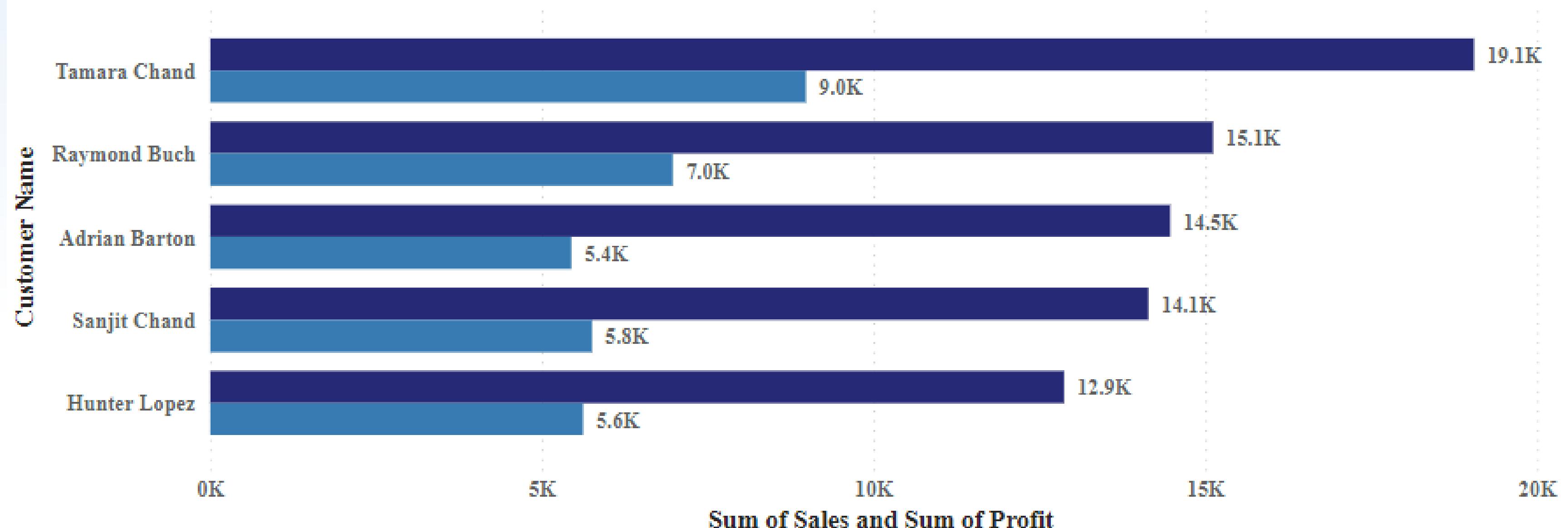


# Customer

5 khách hàng có doanh thu và lợi nhuận cao nhất

Sum of Sales and Sum of Profit by Customer Name

- Sum of Sales
- Sum of Profit



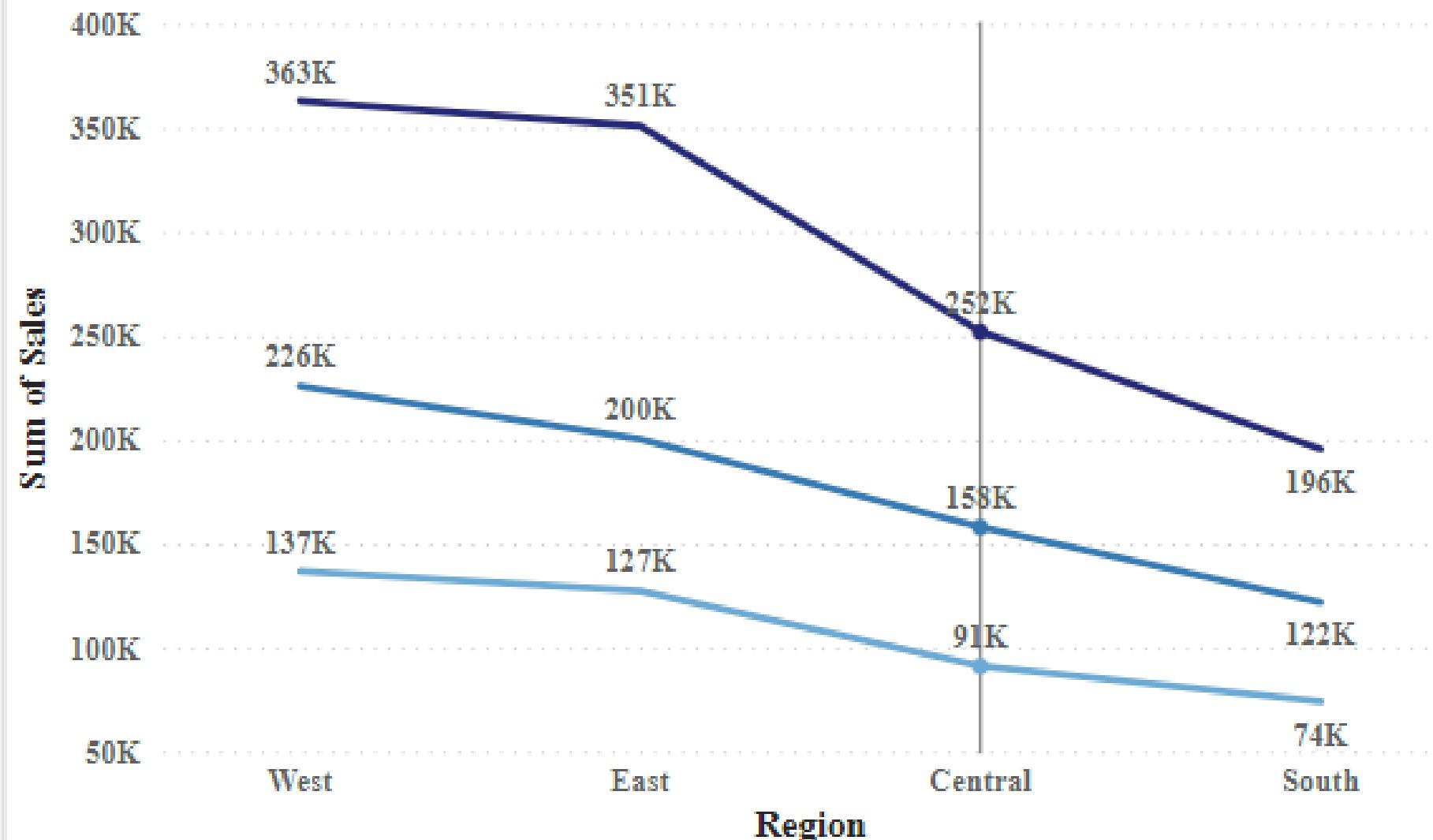
# Customer

Doanh thu của khách hàng tại miền Tây nước Mỹ lớn nhất, và thuộc nhóm "Consumer" với 363K \$, thấp nhất là khu vực Nam Mỹ với nhóm khách hàng thuộc "Home Office" với 74K \$.

Doanh thu của khách hàng sử dụng phương thức vận chuyển "Standard Class", và thuộc nhóm "Consumer" với 0.71M \$, thấp nhất là phương thức "Same Day" với nhóm khách hàng thuộc "Home Office".

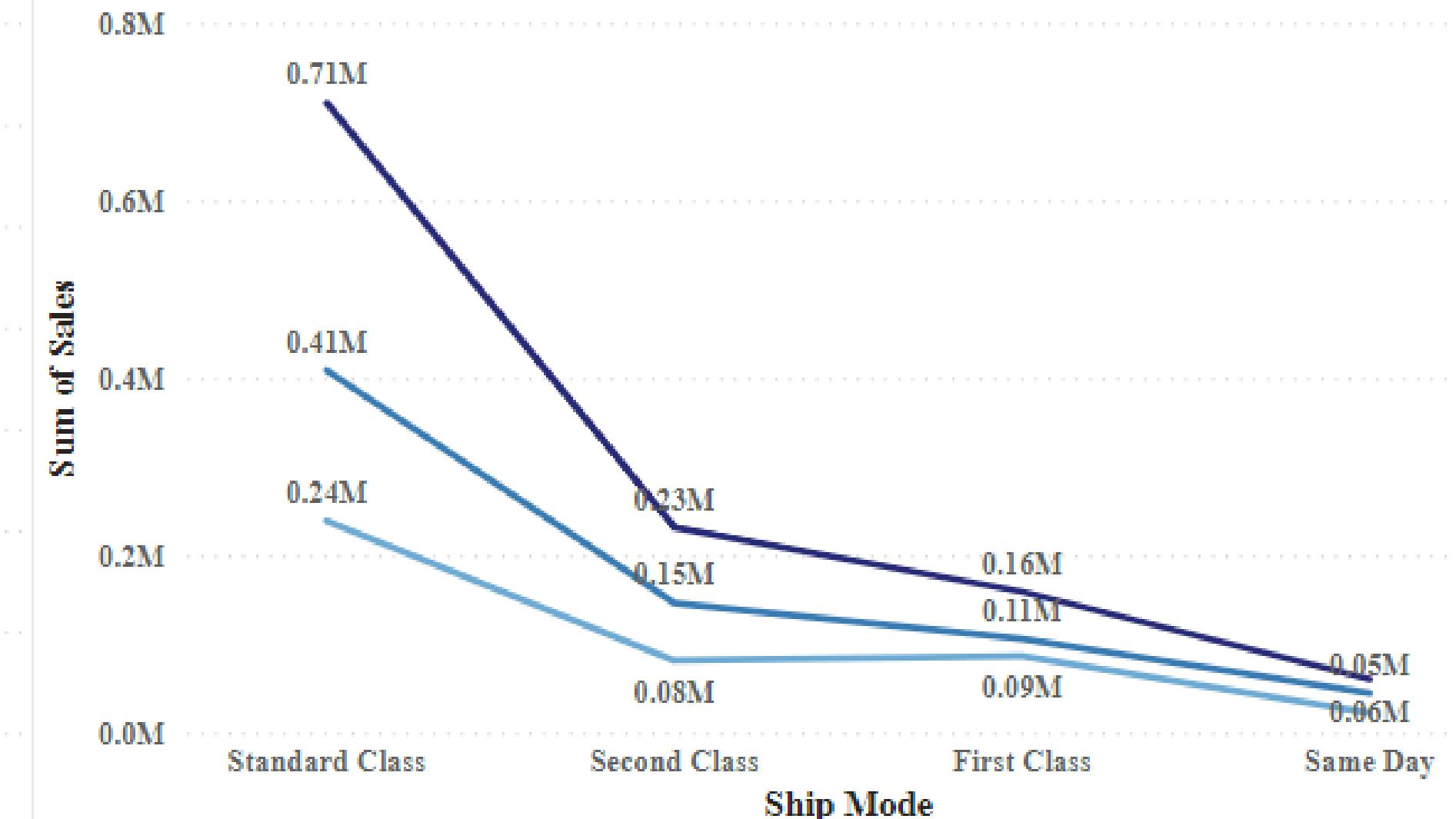
Sum of Sales by Region and Segment

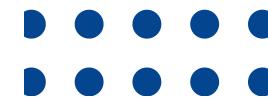
Segment ● Consumer ● Corporate ● Home Office



Sum of Sales by Ship Mode and Segment

Segment ● Consumer ● Corporate ● Home Office





# RFM Analysis



**REGION - STATE - CITY**

# RFM customer segments

STT	Customer Segment	Activity	Actionable Tip	• • • • • • • •
1	Champions	Bought recently, buy often, and spend the most!	Reward them. Can be early adopters for new products. Will promote your brand.	
2	Loyal Customers	Spend good money with us often. Responsive to promotions.	Upsell higher-value products. Ask for reviews. Engage them.	
3	Potential Loyalist	Recent customers, but spent a good amount and bought more than once.	Offer membership/loyalty program, and recommend other products.	
4	Recent Customers	Bought most recently, but not often.	Provide onboarding support, give them early success, and start building relationships.	
5	Promising	Recent shoppers, but haven't spent much.	Create brand awareness, and offer free trials	
6	Customers Needing Attention	Above average recency, frequency, and monetary values. May not have bought it very recently though.	Make limitedtime offers, and recommend based on past purchases. Reactivate them.	
7	About To Sleep	Below average recency, frequency, and monetary values. Will lose them if not reactivated.	Share valuable resources, recommend popular products/renewals at discounts, and reconnect with them.	
8	At Risk	Spent big money and purchased often. But a long time ago. Need to bring them back!	Send personalized emails to reconnect, offer renewals, and provide helpful resources.	
9	Can't Lose Them	Made biggest purchases, and often. But haven't returned for a long time.	Win them back via renewals or newer products, don't lose them to competition, talk to them.	
10	Hibernating	Last purchase was long back, with low spenders and a low number of orders.	Offer other relevant products and special discounts. Recreate brand value.	
11	Lost	Lowest recency, frequency, and monetary scores.	Revive interest with reach-out campaign, ignore otherwise.	

# Creating RFM segments

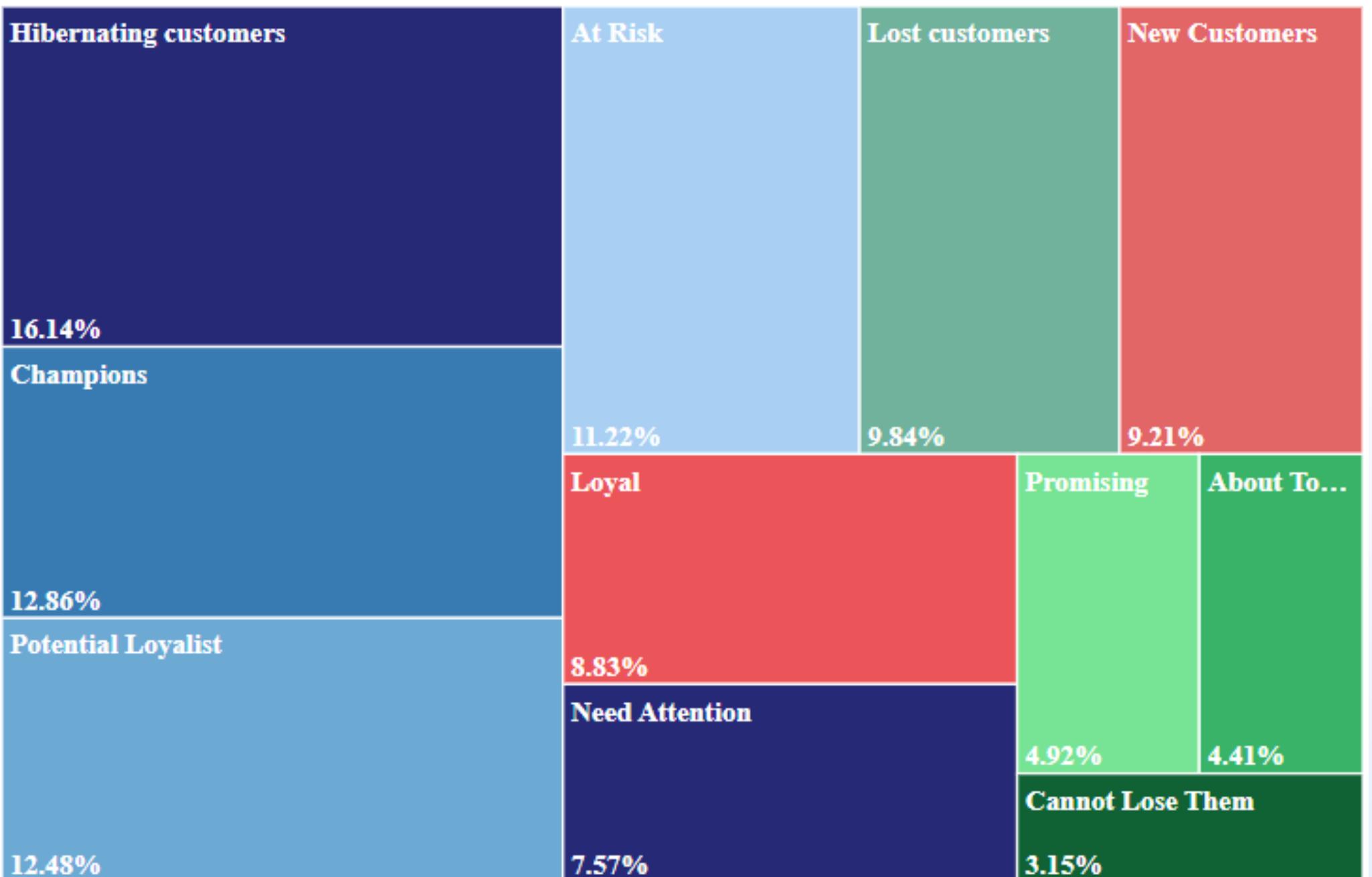


Customer Segment	Recency Score Range	Frequency & Monetary Combined Score Range
<b>Champions</b>	<b>4-5</b>	<b>4-5</b>
<b>Loyal Customers</b>	<b>2-5</b>	<b>3-5</b>
<b>Potential Loyalist</b>	<b>3-5</b>	<b>1-3</b>
<b>Recent Customers</b>	<b>4-5</b>	<b>0-1</b>
<b>Promising</b>	<b>3-4</b>	<b>0-1</b>
<b>Customers Needing Attention</b>	<b>2-3</b>	<b>2-3</b>
<b>About To Sleep</b>	<b>2-3</b>	<b>0-2</b>
<b>At Risk</b>	<b>0-2</b>	<b>2-5</b>
<b>Can't Lose Them</b>	<b>0-1</b>	<b>4-5</b>
<b>Hibernating</b>	<b>1-2</b>	<b>1-2</b>
<b>Lost</b>	<b>0-2</b>	<b>0-2</b>



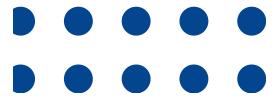
# RFM

%GT Count of RFM-Score by Segment



Hibernating customers là nhóm khách hàng chiếm tỷ lệ cao nhất với 16.14%, ban quản lý cần đưa ra các giải pháp nhằm tái tạo giá trị thương hiệu với đối tượng này, tiếp đến là Champions và Potential Loyalist với 12.86% và 12.48% cho thấy khách hàng tiềm năng chiếm số lượng khá lớn, cần có các chiến lược marketing đến họ sớm nhất, Cung cấp chương trình thành viên/khách hàng thân thiết và giới thiệu các sản phẩm khác,...

Segment	Count of RFM-Score	%GT Count of RFM-Score
About To Sleep	35	4.41%
At Risk	89	11.22%
Cannot Lose Them	25	3.15%
Champions	102	12.86%
Hibernating customers	70	16.14%
Lost customers	78	9.84%
Loyal	70	8.83%
Need Attention	60	7.57%
New Customers	73	9.21%
Potential Loyalist	99	12.48%
Promising	39	4.92%
Total	793	100.00%



# Tổng kết

## Effective Strategies for Sales

### Product

Các loại sản phẩm dòng Technology cần được đẩy mạnh thị phần trong các năm tiếp theo. Các sản phẩm thuộc Furniture, Office Supplies cần có chiến lược quảng bá và nhận diện thương hiệu lớn hơn nhằm tăng doanh thu và lợi nhuận. Thêm sản phẩm hoặc dịch vụ cho đa dạng hơn, chính sách discount phù hợp.

### Customer

Cần có các chính sách khuyến mại, dịch vụ mua hàng vip với từng đối tượng khách hàng, đặc biệt cho các khách đã tiêu dùng nhiều, xây dựng mối quan hệ bền chặt và lâu dài với khách hàng, tập trung cải thiện chất lượng với khách hàng tại miền Tây và cải thiện giá cả với khách hàng khu vực Nam Mỹ.

### Time and Region

Cần gia tăng phạm vi tiếp cận địa lý của siêu thị, hiểu rõ đối tượng mục tiêu, tập trung giải pháp marketing và quảng bá thương hiệu đến các khu vực đông dân và có thu nhập bình quân cao như Central, West, các bang như California, New York.

### RFM

Có các chính sách mua hàng phù hợp với từng đối tượng khách hàng khác nhau, tập trung nhóm khách hàng Champions, Loyal Customers, Potential Loyalist,... xây dựng kế hoạch giảm khách hàng thuộc nhóm Lost, Hibernating với chính sách ưu đãi về giá cả, phương thức vận chuyển, gói khuyến mại, quà tặng, ...



# THANK YOU

