



E-Commerce

PROJECT **GLOBAL SUPER STORE**

Presentation By
Huynh Thai Bao

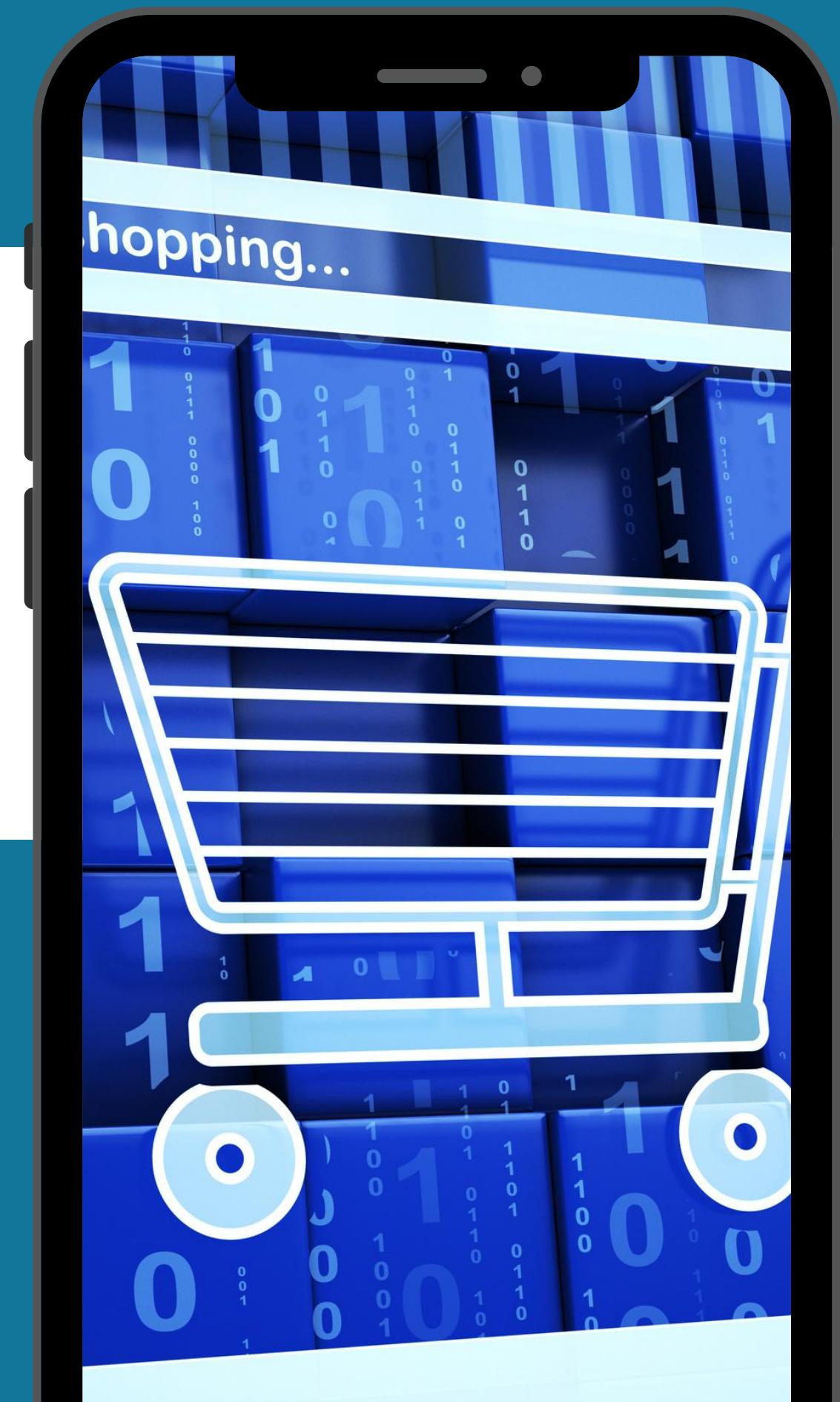


Table Of Content

Introduction

Business Goals – Work Scope

Domain Knowledge - Tools

Data Preparation

Customer Segmentation

Cross Sell

Data Analysis

Conclusion





E-Commerce

01

Introduction



1. Introduction

1.1. About Company

- Global online retail store.
- Headquartered in New York, United States.
- Business areas:
 - Technology
 - Furniture
 - Office Supplies



1. Introduction

1.2. What is the Context?

- In the fiercely competitive landscape of the global online retail market, understanding customer behavior and optimizing business performance have become crucial factors for ensuring sustainable growth and maintaining profitability.
- This project was initiated to address the need for **analyzing** key business metrics such as **Sales, Profit, Customer Characteristics, Transaction Volume, Cart Details, and Growth Rate**. By leveraging historical data, the project aims to deliver valuable and in-depth insights, enabling businesses to better understand their customers, uncover **relationships between products**, and thereby enhance operational efficiency and optimize profitability.





E-Commerce

02

Business Goals

Work Scope



2.1. Business Goals



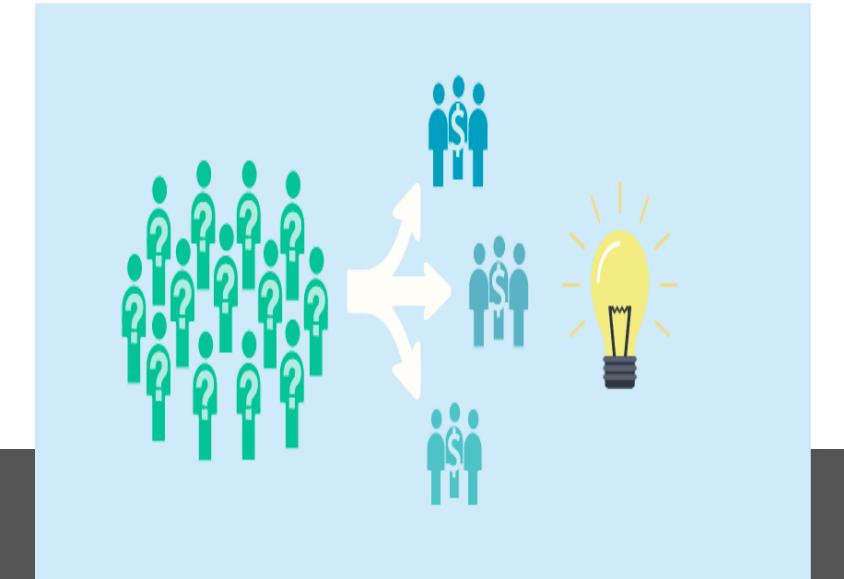
Sales Analysis

Evaluate business performance through key metrics such as sales, profit, costs, and the impact of discount programs, analyzed across dimensions like timeline, product categories, customer segments, and other relevant aspects.



Product Analysis

Identify the key product segments of the business based on sales volume, average price, revenue, and market share, while analyzing the impact of discount programs on customer purchasing trends.



Customer Analysis

Conduct a detailed analysis of customer purchasing trends using the Recency, Frequency, and Monetary metrics, in order to identify the customer segments that contribute most to the business revenue.



2.1. Business Goals



Cross Sell Analysis

Leverage cart data to uncover relationships between products within the business, and then develop cross-selling strategies to enhance the customer's cart value.



Conclusions

Consolidate insights from analyzing historical data, and then draw accurate conclusions that align with the business's current situation.



Recommendations

Based on the assessment of the current situation the business is facing, provide recommendations and necessary actions to be taken in both the short and long term to optimize performance in each analyzed area.



2.2. Work Scope

Data Processing

- Collect data
- Validate data
- Clean data
- Descriptive analysis

Data Modeling

- Create data tables
- Upload data to Power BI
- Set up relationships
- Create necessary tables

Sales Analysis

- Sales, profit, costs
- Growth trends
- Impact of discounts

Product Analysis

- Sales, profit
- Sales trends
- Key product categories
- Impact of discounts

Customer Analysis

- Customer segmentation (RFM)
- Customer profiling

Cross Sell Analysis

- Understand and explain the problem
- Analyze the shopping cart





E-Commerce

03

Domain Knowledge

Tools



3.1. Domain Knowledge



What is E-commerce ?

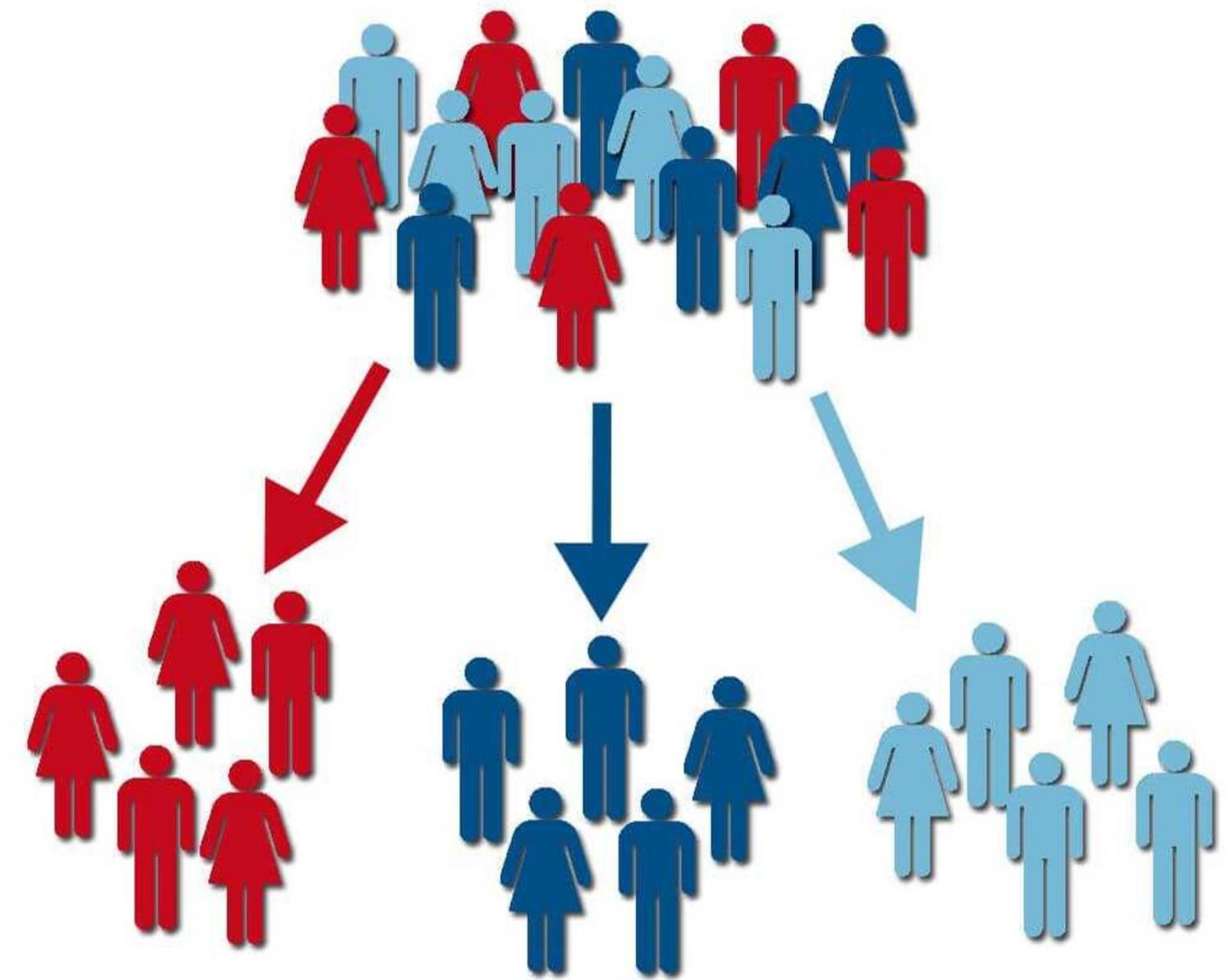
- E-commerce is the process of distributing, selling, purchasing, and marketing products (including goods and services) through telecommunications networks, particularly the internet, television, or other computer networks.
- This activity utilizes digital technology to connect businesses with consumers, making transactions faster, more convenient, and more efficient. E-commerce is not limited to online purchasing but also includes electronic payments, product promotion, order management, and customer support, creating a comprehensive ecosystem for trading and exchanging information.



3.1. Domain Knowledge

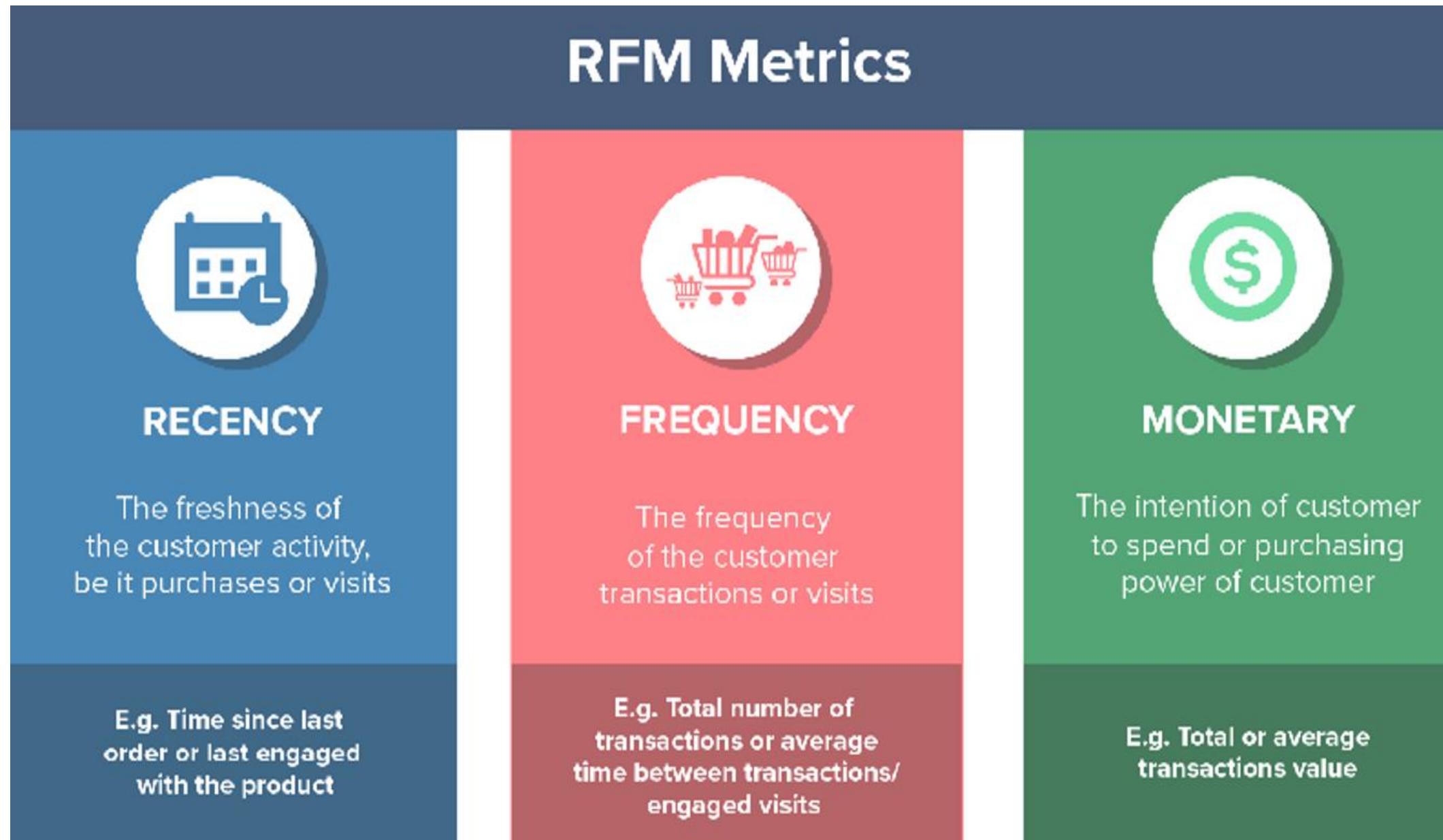
What is Customer Segmentation ?

- Customer segmentation, or market segmentation, is a marketing method that divides customers into smaller groups based on characteristics such as geography, demographics, psychographics, and behavior.
- The goal is to predict the value of each group to the business, and then apply different targeting strategies to maximize value from both high and low-profit customers.



3.1. Domain Knowledge

What is RFM in Customer Segmentation?



RFM stands for Recency, Frequency, and Monetary Value, and it is a technique used in marketing and customer segmentation to analyze and categorize customers based on their transaction behavior. Each of the three components has a specific meaning:

- Recency (R): How recently did the customer make a purchase?
- Frequency (F): How often does the customer make purchases within a specific timeframe?
- Monetary (M): How much money has the customer spent within a specific timeframe?

RFM Customer Segmentation helps businesses better understand their customers, target specific segments with tailored marketing efforts, enhance customer loyalty, and increase profitability through optimized marketing strategies.



3.1. Domain Knowledge

What are Cross Sell and Association Rule?

Cross-selling is a sales technique that involves selling customers a product that relates to or supplements a current or past purchase.

Association Rule is a data mining technique used to discover relationships between items or events in a large dataset. It is commonly applied in market basket analysis to assist with cross-selling strategies. Key Indicators:

- **Support:** The frequency of the pair of products A and B being purchased together.
- **Confidence:** The probability that a customer will buy product B after having purchased product A.
- **Lift:**
 - >1: Buying product A increases the likelihood of buying product B.
 - <1: Buying product A decreases the likelihood of buying product B.
 - =1: The pair of products has no correlation.



3.1. Domain Knowledge

Consider a rule $\{X\} \rightarrow \{Y\}$ where X is the antecedent product list, and Y is the consequent product list.

- **Support:** Represents the frequency of the itemset appearing in the observed data, calculated using the following formula:

$$\text{Support}(\{X\} \rightarrow \{Y\}) = \frac{\text{Total records containing both } X \text{ and } Y}{\text{Total records}}$$

- **Confidence:** Represents the frequency of product Y appearing in the itemset, calculated using the following formula:

$$\text{Confidence}(\{X\} \rightarrow \{Y\}) = \frac{\text{Total records containing both } X \text{ and } Y}{\text{Total records containing } X}$$

- **Lift:** This coefficient compares the likelihood of purchasing Y when X is purchased to the likelihood of purchasing Y without any condition. It measures how purchasing X affects the probability of purchasing Y, calculated as:

$$\text{Lift}(\{X\} \rightarrow \{Y\}) = \frac{\frac{\text{Number of records containing both } X \text{ and } Y}{\text{Number of records containing } X}}{\frac{\text{Number of records containing } Y}{\text{Total number of records}}}$$



3.2. Tools



- Load data from CSV file.
- Clean the data.
- Perform descriptive analysis.
- Create data tables.



- Connect data from SQL Server.
- Build data model.
- Perform customer segmentation.
- Implement cross-selling strategies.
- Build dashboard.



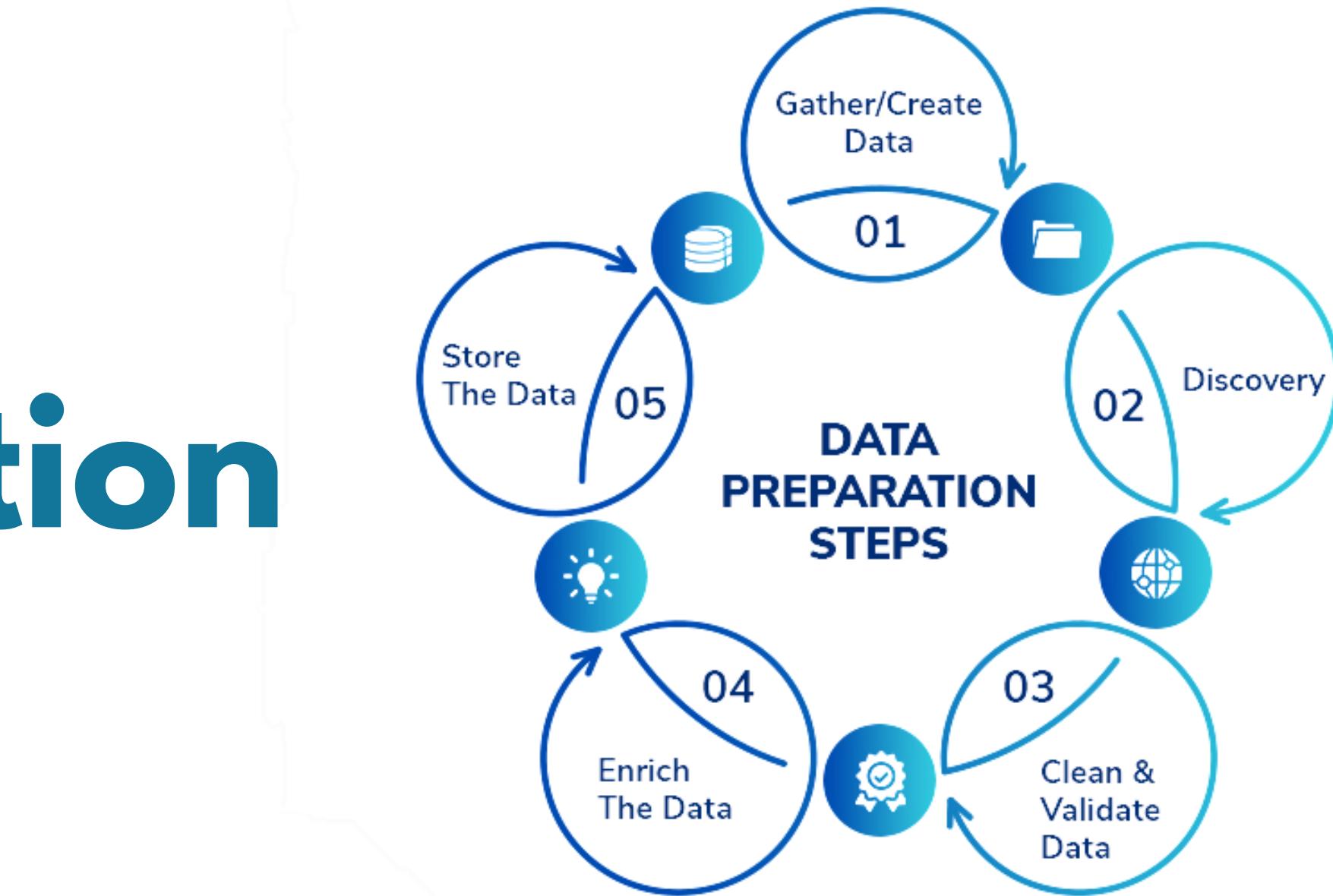
- Introduce the project.
- Summarize the main content.
- Analyze and provide insights.
- Conclusion and recommendations.





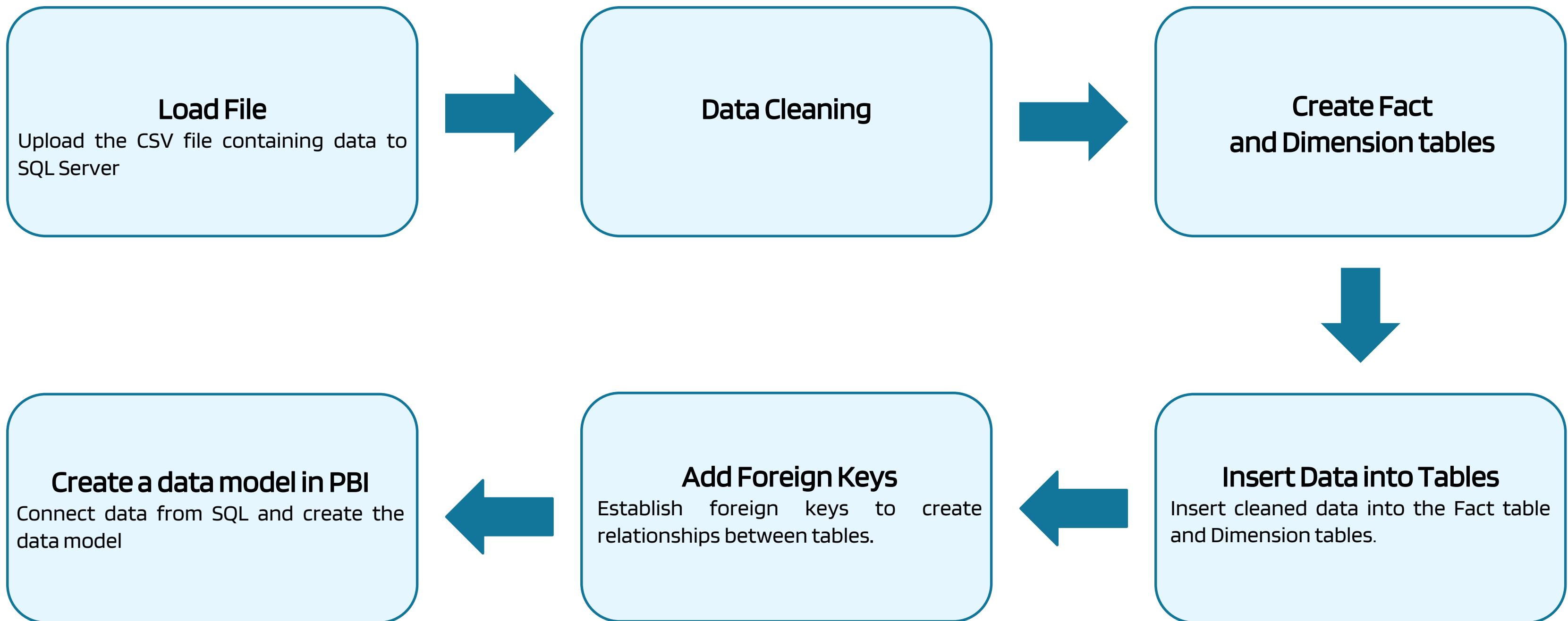
E-Commerce

04 Data Preparation



4. Data Preparation

4.1. Process



4. Data Preparation

4.2. Summary About Dataset

Column Name	Explanation
Row_ID	Unique identifier for the row
Order_ID	Order ID
Order_Date	The date the order was created
Ship_Date	The shipping date for the order
Ship_Mode	Shipping method
Customer_ID	Customer ID
Customer_Name	Full name of the customer
Segment	Customer segment
City	City where the order is delivered
State	State/Province where the order is delivered
Country	Country where the order is delivered
Postal_Code	Postal code of the delivery location

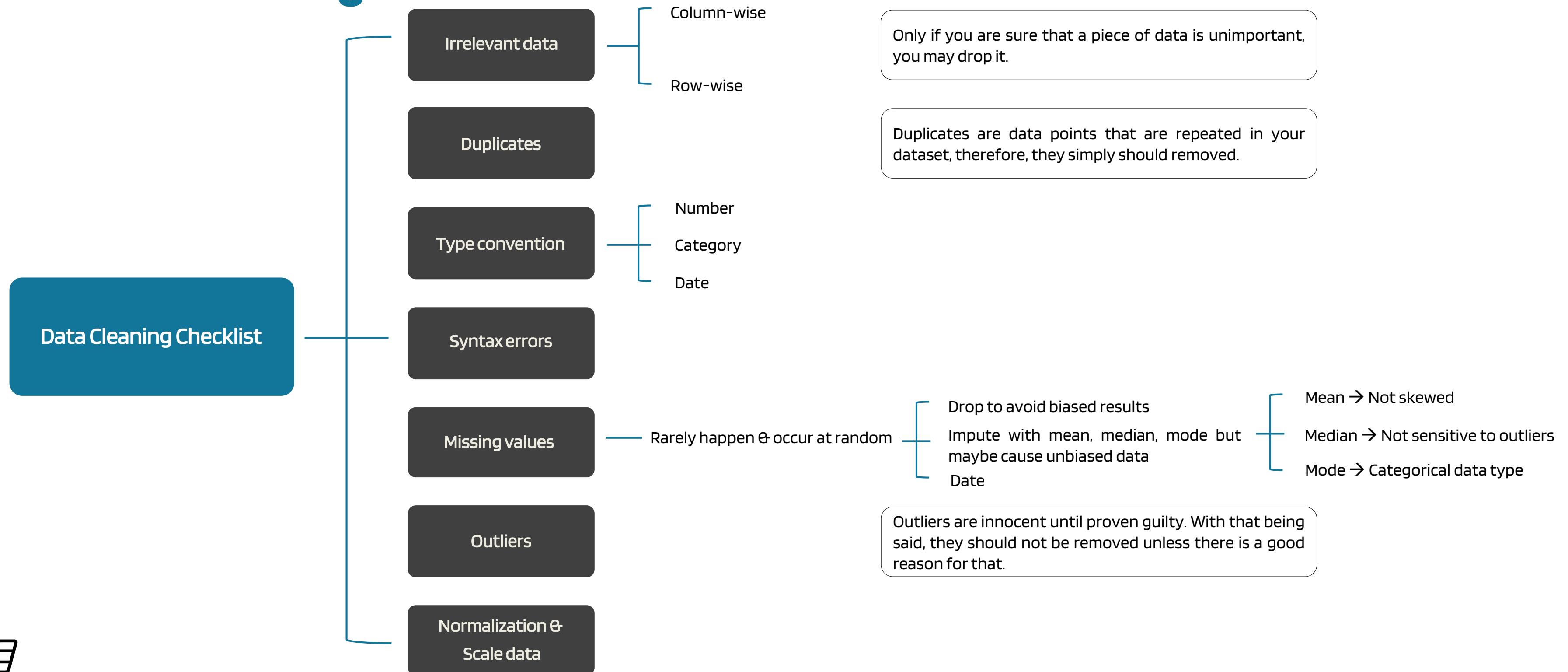
Column Name	Explanation
Market	Market (where the order originates)
Region	Region where the order originates
Product_ID	Product ID
Category	Product category
Sub_Category	Subcategory of the product within its category
Product_Name	Full name of the product
Sales	Total sales value of the order
Quantity	Quantity of the product sold
Discount	Discount percentage for the order
Profit	Total profit from the order
Shipping_Cost	Cost of shipping the order
Order_Priority	Priority level of the order



https://www.kaggle.com/datasets/apoorvaappz/global-super-store-dataset?select=Global_Superstore2.csv

4. Data Preparation

4.3. Data Cleaning



4. Data Preparation

4.4. Create Fact & Dimension Tables

```
--- 1. Create Table Sales
CREATE TABLE Sales (
    OrderID VARCHAR(50),
    OrderDate DATE,
    ShipDate DATE,
    ShipModeID INT,
    CustomerID VARCHAR(50),
    ProductID VARCHAR(50),
    RegionID INT,
    Sales FLOAT,
    UnitPrice FLOAT,
    Quantity INT,
    Discount FLOAT,
    Profit FLOAT,
    ShippingCost FLOAT,
    OrderPriorityID INT
);
```

```
--- 2. Create Table Product
CREATE TABLE Product (
    ProductID VARCHAR(50),
    CategoryID INT,
    SubCategoryID INT,
    ProductName NVARCHAR(255)
);
```

```
--- 3. Create Table Category
CREATE TABLE Category (
    CategoryID INT PRIMARY KEY IDENTITY(1,1),
    CategoryName NVARCHAR(50)
);
```

```
--- 4. Create Table Sub_Category
CREATE TABLE Sub_Category (
    SubCategoryID INT PRIMARY KEY IDENTITY(1,1),
    SubCategoryName NVARCHAR(50)
);
```

```
--- 5. Create Table Region
CREATE TABLE Region (
    RegionID INT PRIMARY KEY IDENTITY(1,1),
    City NVARCHAR(100),
    State NVARCHAR(100),
    CountryID INT,
    MarketID INT
);
```

```
--- 6. Create Table Market
CREATE TABLE Market (
    MarketID INT PRIMARY KEY IDENTITY(1,1),
    MarketName NVARCHAR(100)
);
```

```
--- 7. Create Table Country
CREATE TABLE Country (
    CountryID INT PRIMARY KEY IDENTITY(1,1),
    CountryName NVARCHAR(100)
);
```

```
--- 8. Create Table Customer
CREATE TABLE Customer (
    CustomerID VARCHAR(50) PRIMARY KEY,
    CustomerName NVARCHAR(255),
    SegmentID INT
);
```

```
--- 9. Create Table Cusomter_Segment
CREATE TABLE Customer_Segment (
    SegmentID INT PRIMARY KEY IDENTITY(1,1),
    SegmentName NVARCHAR(100)
);
```

```
--- 10. Create Table Ship_Mode
CREATE TABLE Ship_Mode (
    ShipModeID INT PRIMARY KEY IDENTITY(1,1),
    ShipMode NVARCHAR(50)
);
```

```
--- 11. Create Table Order_Priority
CREATE TABLE Order_Priority (
    OrderPriorityID INT PRIMARY KEY IDENTITY(1,1),
    OrderPriority NVARCHAR(50)
);
```



4. Data Preparation

4.5. Insert Data

--- 1. Insert Data in Category Table

```
INSERT INTO [dbo].[Category] (CategoryName)
SELECT DISTINCT
    Category
FROM [dbo].[Data_Cleaned];
```

--- 2. Insert Data in Sub_Category Table

```
INSERT INTO [dbo].[Sub_Category] (SubCategoryName)
SELECT DISTINCT
    Sub_Category
FROM [dbo].[Data_Cleaned];
```

--- 3. Insert Data in Market Table

```
INSERT INTO [dbo].[Market] (MarketName)
SELECT DISTINCT
    Market
FROM [dbo].[Data_Cleaned];
```

--- 4. Insert Data in Country Table

```
INSERT INTO [dbo].[Country] (CountryName)
SELECT DISTINCT
    Country
FROM [dbo].[Data_Cleaned];
```

--- 5. Insert Data in Customer Segment Table

```
INSERT INTO [dbo].[Customer_Segment] (SegmentName)
SELECT DISTINCT
    Segment
FROM [dbo].[Data_Cleaned];
```

--- 7. Insert Data in Product Table

```
INSERT INTO [dbo].[Product] (ProductID, CategoryID, SubCategoryID, ProductName)
SELECT DISTINCT
    Product_ID,
    (SELECT CategoryID FROM [dbo].[Category] WHERE CategoryName = [dbo].[Data_Cleaned].Category),
    (SELECT SubCategoryID FROM [dbo].[Sub_Category] WHERE SubCategoryName = [dbo].[Data_Cleaned].Sub_Category),
    Product_Name
FROM [dbo].[Data_Cleaned];
```

--- 8. Insert Data in Region Table

```
INSERT INTO [dbo].[Region] (City, State, CountryID, MarketID)
SELECT DISTINCT
    City,
    State,
    (SELECT CountryID FROM Country WHERE CountryName = [dbo].[Data_Cleaned].Country),
    (SELECT MarketID FROM Market WHERE MarketName = [dbo].[Data_Cleaned].Market)
FROM [dbo].[Data_Cleaned];
```



4. Data Preparation

4.5. Insert Data

```
--- 9. Insert Data in Customer Table
INSERT INTO [dbo].[Customer] (CustomerID, CustomerName, SegmentID)
SELECT DISTINCT
    Customer_ID,
    Customer_Name,
    (SELECT SegmentID FROM Customer_Segment WHERE SegmentName = [dbo].[Data_Cleaned].Segment)
FROM [dbo].[Data_Cleaned];
```

```
--- 10. Insert Data in Order Priority Table
INSERT INTO [dbo].[Order_Priority] (OrderPriority)
SELECT DISTINCT
    Order_Priority
FROM [dbo].[Data_Cleaned];
```

```
--- 11. Insert Data in Sales Table
INSERT INTO Sales (OrderID, OrderDate, ShipDate, ShipModeID,
                  CustomerID, ProductID, RegionID, Sales, UnitPrice,
                  Quantity, Discount, Profit, ShippingCost, OrderPriorityID)
SELECT
    dc.OrderID,
    dc.Order_Date,
    dc.Ship_Date,
    sm.ShipModeID,
    dc.Customer_ID,
    dc.Product_ID,
    r.RegionID,
    dc.Sales,
    dc.Unit_Price,
    dc.Quantity,
    dc.Discount,
    dc.Profit,
    dc.Shipping_Cost,
    op.OrderPriorityID
FROM [dbo].[Data_Cleaned] dc
LEFT JOIN [dbo].[Ship_Mode] sm ON dc.Ship_Mode = sm.ShipMode
LEFT JOIN [dbo].[Region] r ON dc.City = r.City
    AND dc.State = r.State
    AND r.CountryID = (SELECT CountryID FROM [dbo].[Country] WHERE CountryName = dc.Country)
    AND r.MarketID = (SELECT MarketID FROM [dbo].[Market] WHERE MarketName = dc.Market)
LEFT JOIN [dbo].[Order_Priority] op ON dc.Order_Priority = op.OrderPriority;
```



4. Data Preparation

4.6. Add Foreign Keys

--- 1. Sales Table

```
ALTER TABLE [dbo].[Sales]
ADD CONSTRAINT FK_OrderPriority FOREIGN KEY (OrderPriorityID)
REFERENCES [dbo].[Order_Priority](OrderPriorityID);
```

```
ALTER TABLE [dbo].[Sales]
ADD CONSTRAINT FK_ShipMode FOREIGN KEY (ShipModeID)
REFERENCES [dbo].[Ship_Mode](ShipModeID);
```

```
ALTER TABLE [dbo].[Sales]
ADD CONSTRAINT FK_Customer FOREIGN KEY (CustomerID)
REFERENCES [dbo].[Customer](CustomerID);
```

```
ALTER TABLE [dbo].[Sales]
ADD CONSTRAINT FK_Region FOREIGN KEY (RegionID)
REFERENCES [dbo].[Region](RegionID);
```

--- 2. Product Table

```
ALTER TABLE [dbo].[Product]
ADD CONSTRAINT FK_Category FOREIGN KEY (CategoryID)
REFERENCES [dbo].[Category](CategoryID);
```

```
ALTER TABLE [dbo].[Product]
ADD CONSTRAINT FK_SubCategory FOREIGN KEY (SubCategoryID)
REFERENCES [dbo].[Sub_Category](SubCategoryID);
```

--- 3. Region Table

```
ALTER TABLE [dbo].[Region]
ADD CONSTRAINT FK_Country FOREIGN KEY (CountryID)
REFERENCES [dbo].[Country](CountryID);
```

```
ALTER TABLE [dbo].[Region]
ADD CONSTRAINT FK_Market FOREIGN KEY (MarketID)
REFERENCES [dbo].[Market](MarketID);
```

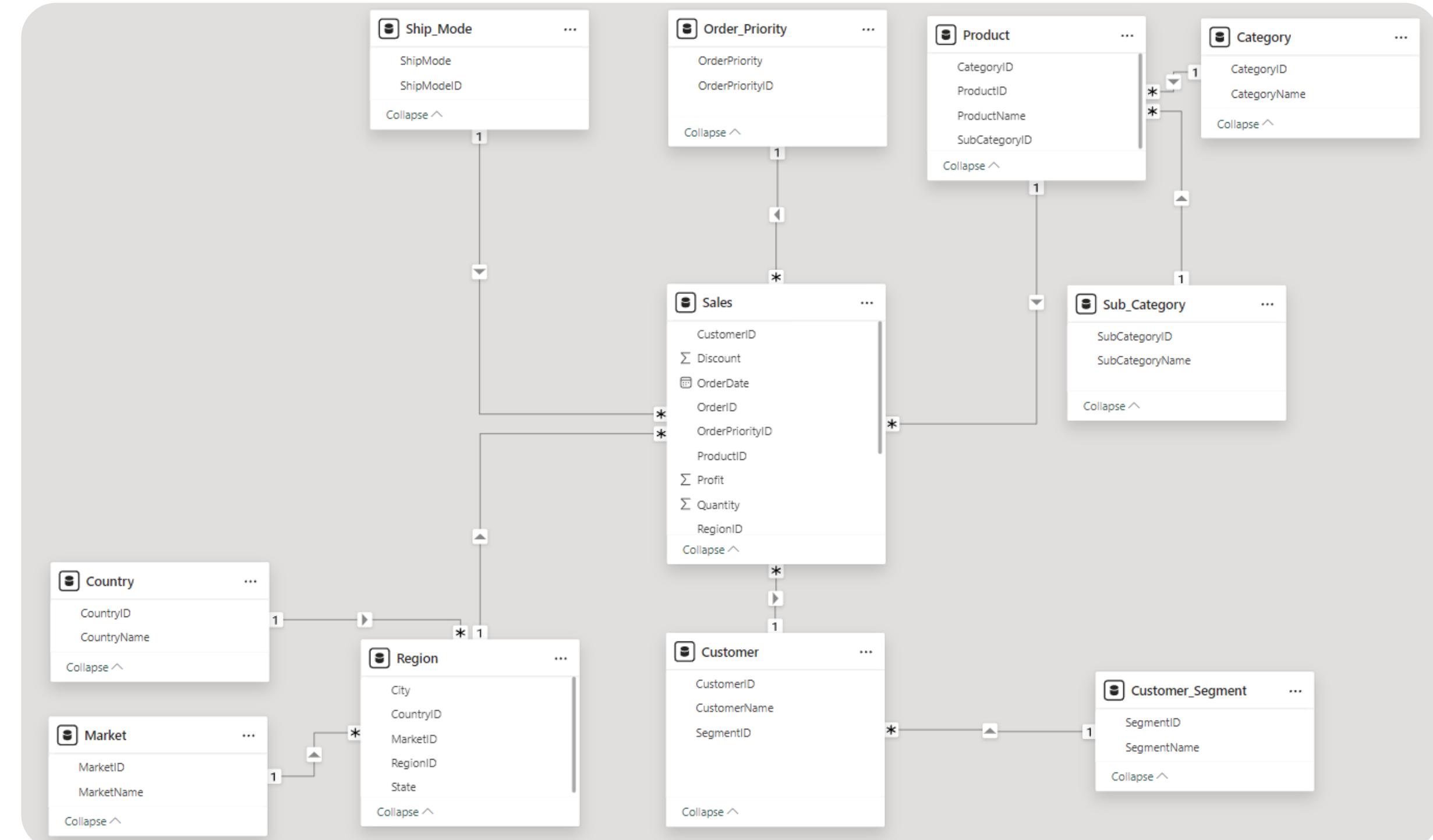
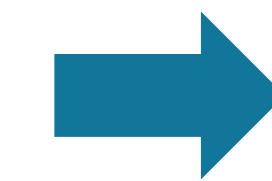
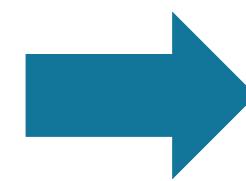
--- 4. Customer Table

```
ALTER TABLE [dbo].[Customer]
ADD CONSTRAINT FK_Segment FOREIGN KEY (SegmentID)
REFERENCES [dbo].[Customer_Segment](SegmentID);
```



4. Data Preparation

4.7. Data Connect & Modeling





E-Commerce

05 Customers Segmentation



5. Customer Segmentation

1

Calculating Recency, Frequency, and Monetary Values

Utilize customer data from the **Dimension Customer** table to calculate the R, F, and M values for each customer by linking it with data from the **Fact Sales** table.

2

Scoring Recency, Frequency, and Monetary

Divide the R, F, and M variables into five segments using the Quantile method, assigning scores on a scale from 1 to 5.

3

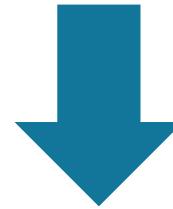
Classifying Customer Segments

Identify customer groups based on the combined scores of the R, F, and M categories, and interpret the significance of each segment.



5. Customer Segmentation

5.1. Calculating Recency, Frequency, and Monetary Values



```
Recency =  
VAR LatestOrderDate =  
CALCULATE(MAX(Sales[OrderDate]),  
| | | ALLEXCEPT(Sales, Customer[CustomerID]) )  
RETURN  
DATEDIFF(LatestOrderDate, DATE(2015,1,1), DAY )
```

```
Frequency =  
CALCULATE(DISTINCTCOUNT(Sales[OrderID]),  
| | | ALLEXCEPT(Sales, Customer[CustomerID]) )
```

```
Monetary = CALCULATE(sum(Sales[Sales]),  
| | | ALLEXCEPT(Sales, Customer[CustomerID]) )
```



5. Customer Segmentation

5.2. Scoring Recency, Frequency, and Monetary

Divide the R, F, and M variables into five segments using the **Quantile method**, assigning scores on a scale from 1 to 5.

Percent	Recency Score
> 80%	1
> 60%	2
> 40%	3
> 20%	4
≤ 20%	5

Percent	Monetary Score
> 80%	5
> 60%	4
> 40%	3
> 20%	2
≤ 20%	1

Percent	Frequency Score
> 80%	5
> 60%	4
> 40%	3
> 20%	2
≤ 20%	1



```
RScore =  
VAR Percent20 = PERCENTILE.EXC(Customer[Recency], 0.2)  
VAR Percent40 = PERCENTILE.EXC(Customer[Recency], 0.4)  
VAR Percent60 = PERCENTILE.EXC(Customer[Recency], 0.6)  
VAR Percent80 = PERCENTILE.EXC(Customer[Recency], 0.8)  
RETURN  
SWITCH(TRUE(),  
    Customer[Recency] >= Percent80, 1,  
    Customer[Recency] >= Percent60, 2,  
    Customer[Recency] >= Percent40, 3,  
    Customer[Recency] >= Percent20, 4, 5)
```



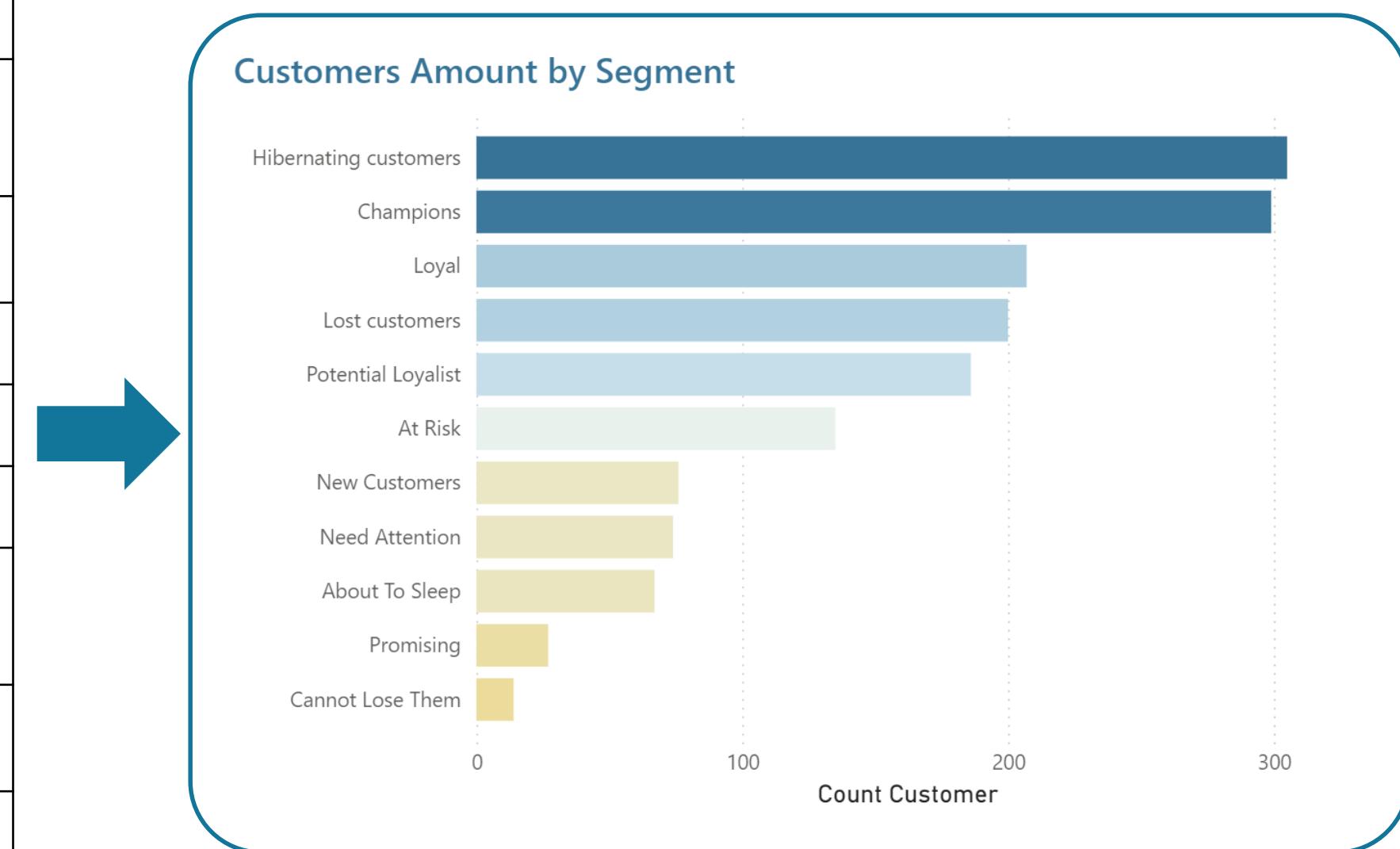
```
FScore =  
VAR Percent20 = PERCENTILE.EXC(Customer[Frequency], 0.2)  
VAR Percent40 = PERCENTILE.EXC(Customer[Frequency], 0.4)  
VAR Percent60 = PERCENTILE.EXC(Customer[Frequency], 0.6)  
VAR Percent80 = PERCENTILE.EXC(Customer[Frequency], 0.8)  
RETURN  
SWITCH(TRUE(),  
    Customer[Frequency] >= Percent80, 5,  
    Customer[Frequency] >= Percent60, 4,  
    Customer[Frequency] >= Percent40, 3,  
    Customer[Frequency] >= Percent20, 2, 1)
```

```
MScore =  
VAR Percent20 = PERCENTILE.EXC(Customer[Monetary], 0.2)  
VAR Percent40 = PERCENTILE.EXC(Customer[Monetary], 0.4)  
VAR Percent60 = PERCENTILE.EXC(Customer[Monetary], 0.6)  
VAR Percent80 = PERCENTILE.EXC(Customer[Monetary], 0.8)  
RETURN  
SWITCH(TRUE(),  
    Customer[Monetary] >= Percent80, 5,  
    Customer[Monetary] >= Percent60, 4,  
    Customer[Monetary] >= Percent40, 3,  
    Customer[Monetary] >= Percent20, 2, 1)
```

5. Customer Segmentation

5.3. Classifying Customer Segments

Segment	RFM Scores	Activity
Champions	555, 554, 544, 545, 454, 455, 445	Recently purchased, orders frequently, and spends the most.
Loyal	543, 444, 435, 355, 354, 345, 344, 335	Regular purchases. Likely to engage in promotional campaigns.
Potential Loyalist	553, 551, 552, 541, 542, 533, 532, 531, 452, 451, 442, 441, 431, 453, 433, 432, 423, 353, 352, 351, 342, 341, 333, 323	Recently purchased and spends well.
Promising	525, 524, 523, 522, 521, 515, 514, 513, 425, 424, 413, 414, 415, 315, 314, 313	Fairly regular customers. Spend often but their last purchase was a few weeks ago.
New Customers	512, 511, 422, 421, 412, 411, 311	New customers.
Need Attention	535, 534, 443, 434, 343, 334, 325, 324	Customers with potential, their last purchase was about a month ago.
About To Sleep	331, 321, 312, 221, 213, 231, 241, 251	Casual customers, haven't purchased in a while.
At Risk	255, 254, 245, 244, 253, 252, 243, 242, 235, 234, 225, 224, 153, 152, 145, 143, 142, 135, 134, 133, 125, 124	Similar to "Cannot lose them but losing," with lower spend and frequency.
Cannot Lose Them	155, 154, 144, 214, 215, 115, 114, 113	Used to place orders most frequently and spend highly but haven't returned in a while.
Hibernating Customers	332, 322, 233, 232, 223, 222, 132, 123, 122, 212, 211	Last purchase was a long time ago, but in the past four weeks, they have visited the website or opened emails.
Lost Customers	111, 112, 121, 131, 141, 151	Their last purchase was the closest among lost customers but showed no activity in the past four weeks.





E-Commerce

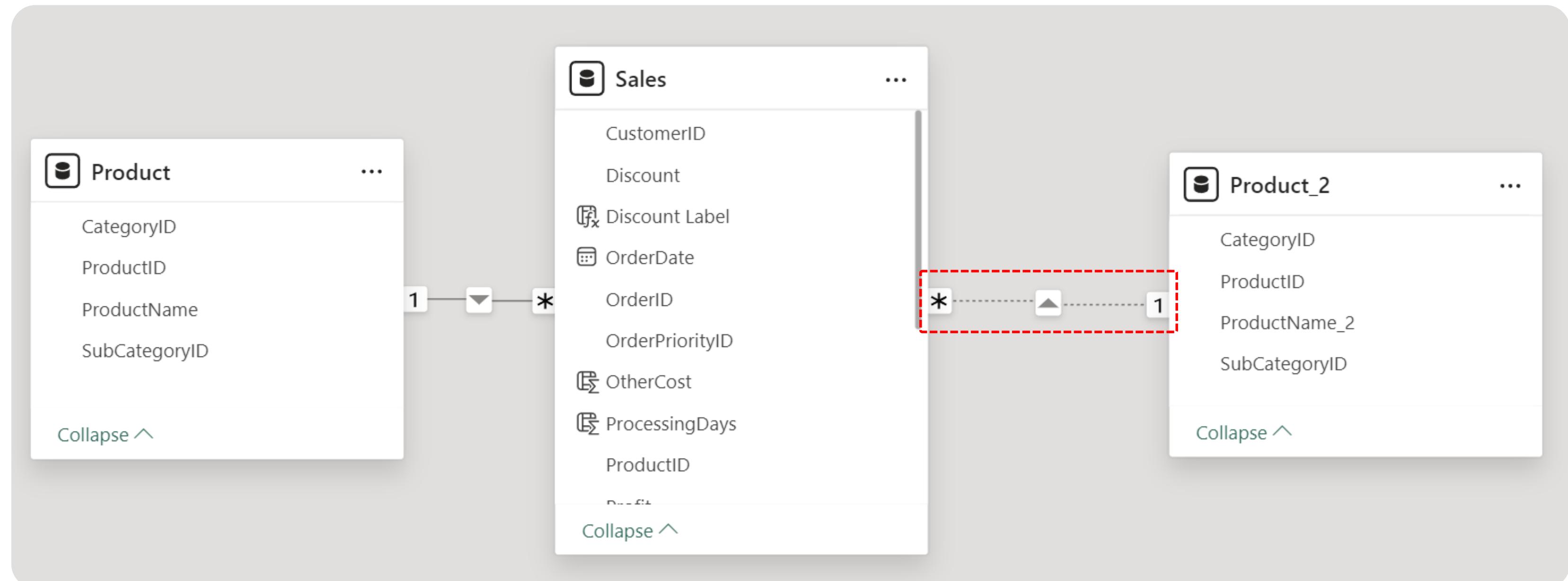
06 Cross Sell



6. Cross Sell

6.1. Expanding Data Modeling for Product Analysis

Use the **Product_2** table created as a Reference in Power Query (linked to Sales via USERELATIONSHIP) to analyze products that are frequently bought together and identify specific product relationships.



6. Cross Sell

6.2. The logic for calculating the metrics

Order Premise: The number of orders that include the first product.

```
Order (Premise) = DISTINCTCOUNT(Sales[OrderID])
```

Order Consequence: The number of orders that include the second product.

```
Order (Consequence) =
CALCULATE(DISTINCTCOUNT(Sales[OrderID]),
    REMOVEFILTERS('Product'),
    USERELATIONSHIP(Sales[ProductID], Product_2[ProductID]))
```

Order Sell Together: The number of orders that include both products.

```
Order (Sell Together) =
VAR SameOrder =
    CALCULATETABLE(
        SUMMARIZE(Sales, Sales[OrderID]),
        REMOVEFILTERS('Product'),
        REMOVEFILTERS(Sales[ProductID]),
        USERELATIONSHIP(Sales[ProductID], Product_2[ProductID]))
VAR Number =
    CALCULATE(DISTINCTCOUNT(Sales[OrderID]), KEEPFILTERS(SameOrder))
RETURN Number
```



6. Cross Sell

6.2. The logic for calculating the metrics

Confidence: is the probability that an order containing the first product (Premise) will also contain the second product (Consequence).

$$\text{Confidence} = [\text{Order Sell Together (Duplicate Filter)}] / [\text{Order (Premise)}]$$

Support: is the ratio of the number of orders containing both products to the total number of orders.

$$\text{Support} = [\text{Order Sell Together (Duplicate Filter)}] / [\text{Order (Total)}]$$

Lift: measures the degree of association between two products. A Lift greater than 1 indicates that the two products have a strong relationship.

$$\text{Lift} = ([\text{Order Sell Together (Duplicate Filter)}] / [\text{Order (Premise)}]) / ([\text{Order (Consequence)}] / [\text{Order (Total)}])$$


6. Cross Sell

6.3. Result

Product Name (Premise)	Order (Premise)	Product Name Consequence	Order (Consequence)	Support	Confidence	Lift
4009 Highlighters		1 Iceberg Mobile Mega Data/Printer Cart	8	0.00%	100.00%	3,219.00
4009 Highlighters		1 Logitech Wireless Marathon Mouse M705	8	0.00%	100.00%	3,219.00
Acco Glide Clips		1 Hand-Finished Solid Wood Document Frame	3	0.00%	100.00%	8,584.00
Ativa V4110MDD Micro-Cut Shredder	2	Staples	222	0.01%	100.00%	116.00
Avaya IP Phone 1140E VoIP phone		1 Global Wood Trimmed Manager's Task Chair, Khaki	14	0.00%	100.00%	1,839.43
Avaya IP Phone 1140E VoIP phone		1 Gould Plastics 18-Pocket Panel Bin, 34w x 5-1/4d x 20-1/2h	10	0.00%	100.00%	2,575.20
Barricks Coffee Table, with Bottom Storage		1 Binney & Smith Canvas, Easy-Erase	29	0.00%	100.00%	888.00
Barricks Coffee Table, with Bottom Storage		1 Brother Personal Copier, Color	18	0.00%	100.00%	1,430.67
Barricks Coffee Table, with Bottom Storage		1 Novimex Round Labels, Alphabetical	19	0.00%	100.00%	1,355.37
Barricks Non-Folding Utility Table with Steel Legs, Laminate Tops		1 Nortel Business Series Terminal T7208 Digital phone	3	0.00%	100.00%	8,584.00
Barricks Non-Folding Utility Table with Steel Legs, Laminate Tops		1 XtraLife ClearVue Slant-D Ring Binders by Cardinal	5	0.00%	100.00%	5,150.40
Barricks Round Table, Rectangular		1 Acme Ruler, High Speed	18	0.00%	100.00%	1,430.67
Barricks Round Table, Rectangular		1 Ibico Binding Machine, Recycled	43	0.00%	100.00%	598.88
Belkin 7 Outlet SurgeMaster Surge Protector with Phone Protection		1 Jabra BIZ 2300 Duo QD Duo Corded Headset	4	0.00%	100.00%	6,438.00
Total	25752		25752			





E-Commerce

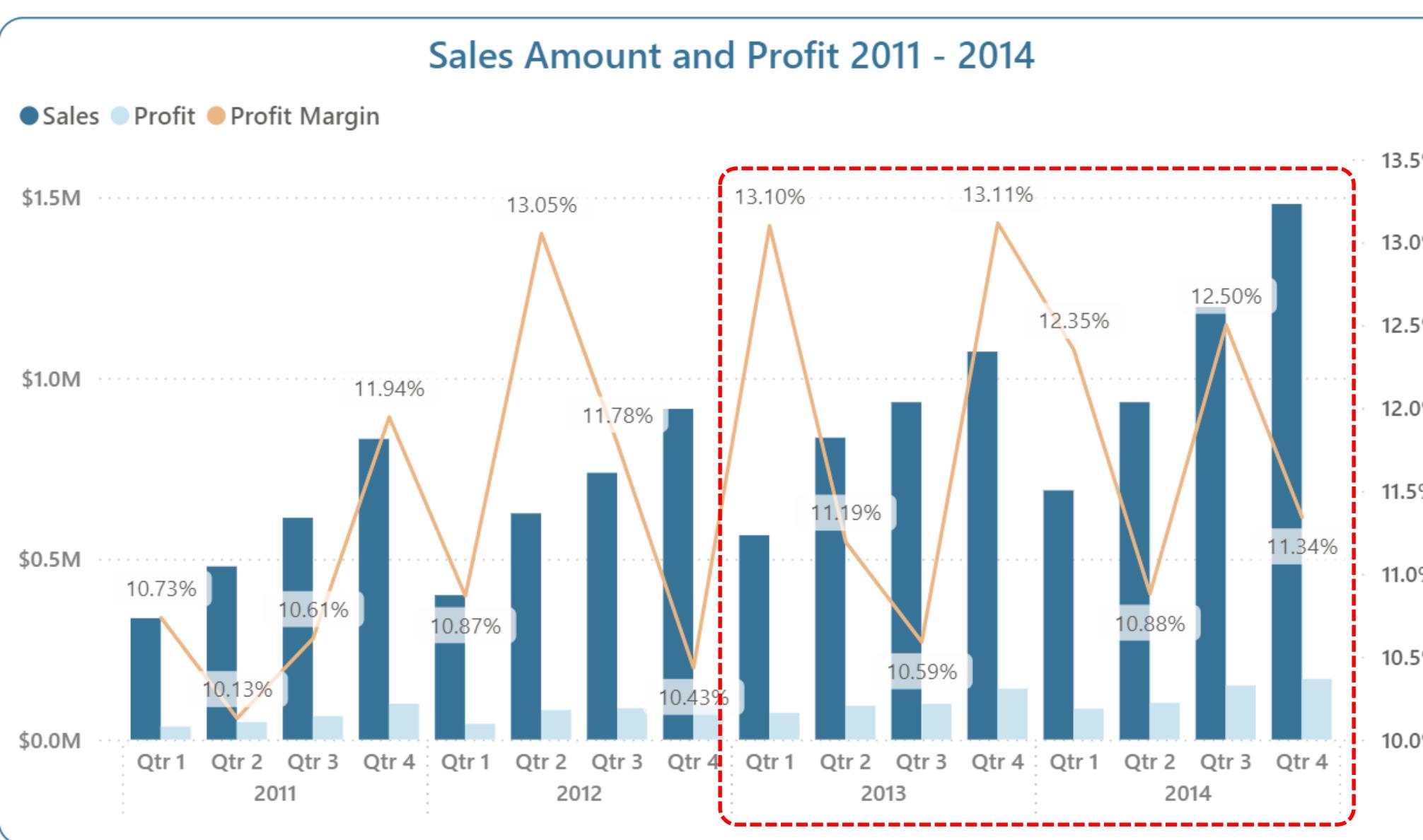
07

Data Analysis



7. Data Analysis

7.1. Business Performance Overview (2011 - 2014)



Sales Growth:

- Sales increased steadily over the years, particularly high in Q4 each year.
- Q4/2014 recorded the highest revenue, nearing the \$1.5 million mark.

Profit Fluctuations:

- Profits fluctuated significantly, with the lowest levels typically in Q1 and Q2 each year.
- The later quarters (Q3 and Q4) generally saw higher profits.

Profit Margin:

- Ranged between 10% and 13.5%, peaking in Q4/2012 (13.11%) and dipping to its lowest in Q2/2011 (10.13%).
- Profit margin was inconsistent, affected by rising costs.

Correlation Between Sales and Profit Margin:

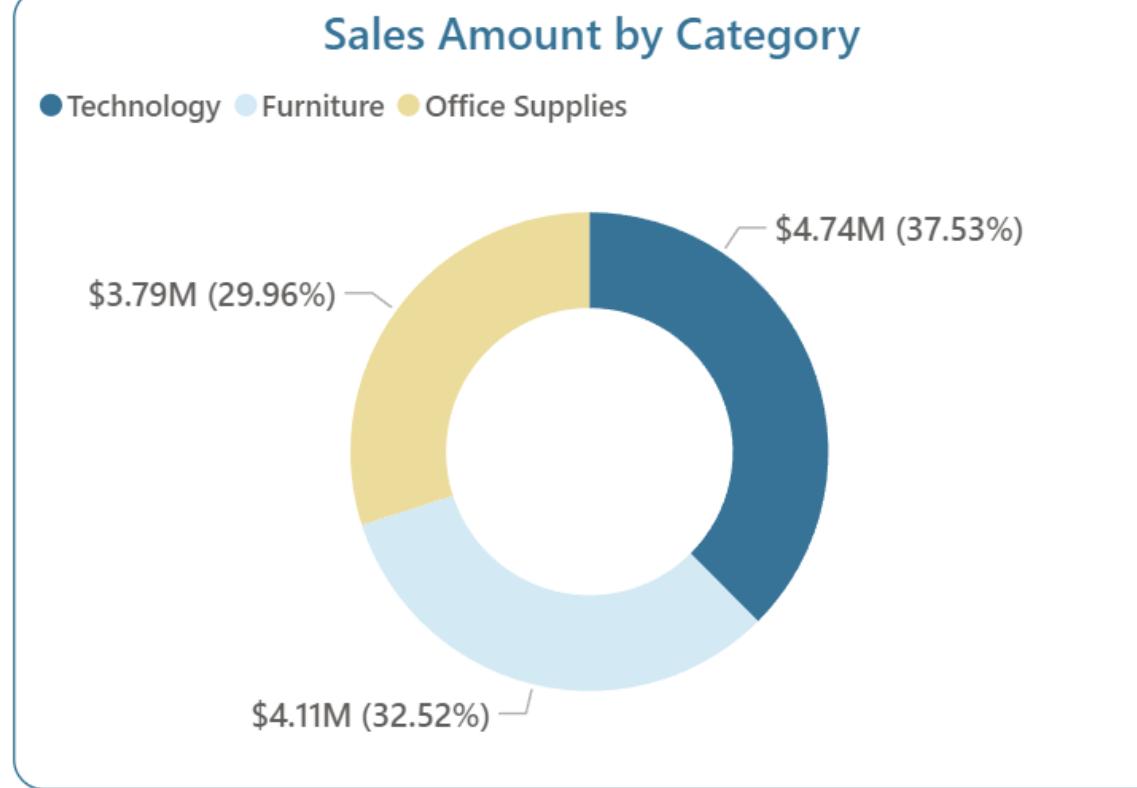
- Sales increased, but profit margins did not grow proportionally, indicating the need for better cost control.
- A downward trend in profit margin was observed from 2013 to 2014.

Outstanding Year-End Performance:

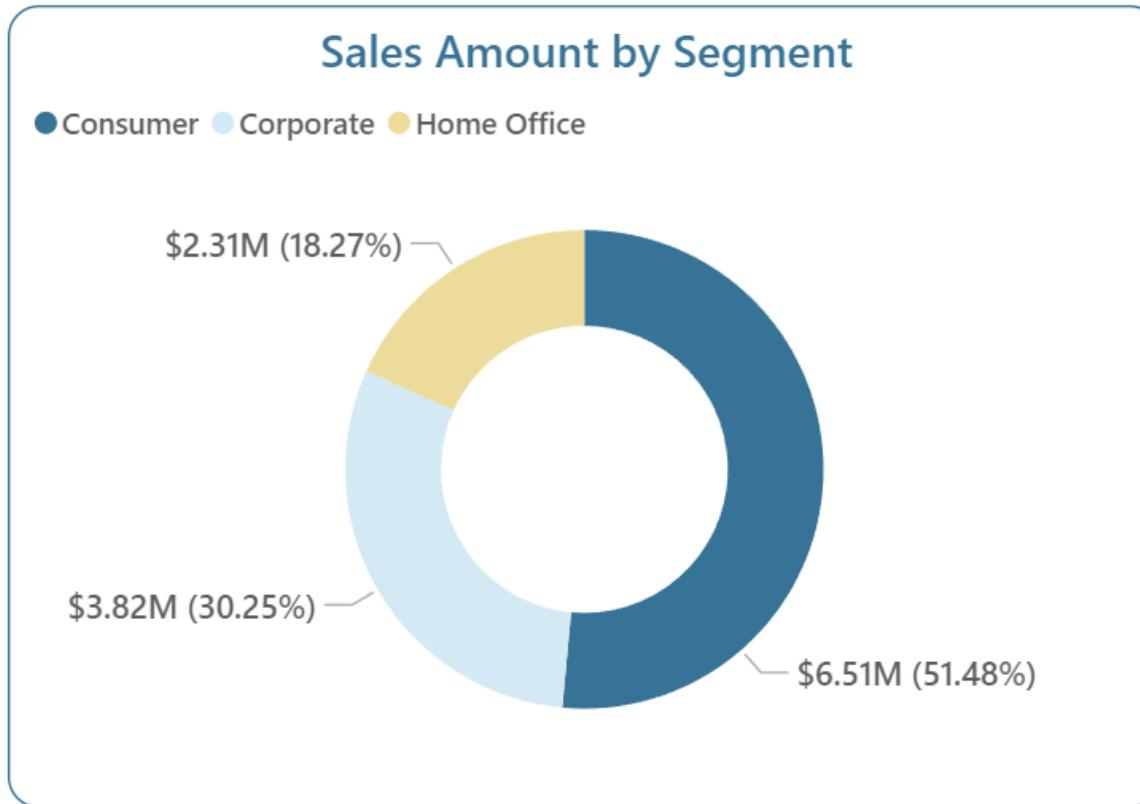
- Q4 each year consistently achieved the highest sales and profit, likely driven by holiday seasons or effective sales campaigns.



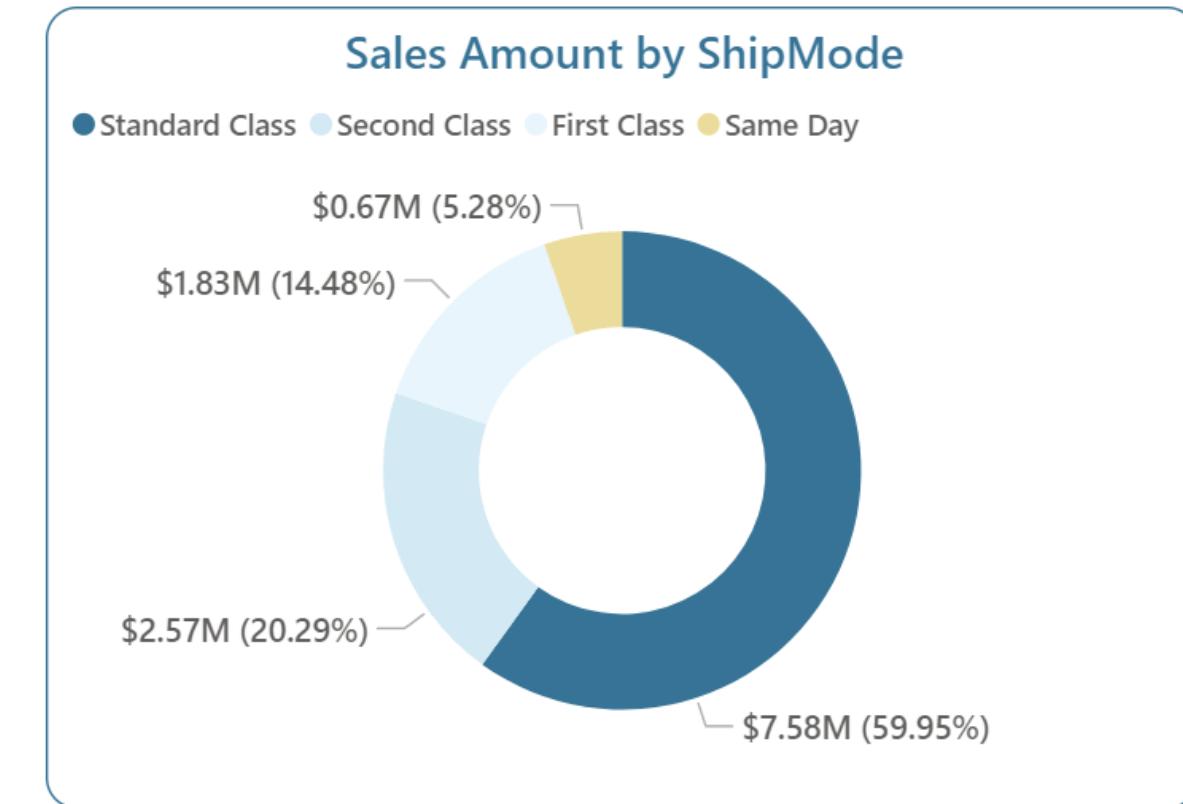
7. Data Analysis



- Sales is relatively evenly distributed across the three product categories.
- Leading the way is **Technology**, accounting for **37.53% of total revenue**, highlighting significant potential and the need for continued focus.
- Although contributing the least, **Office Supplies** still makes up **nearly 30%**, indicating room for growth and the possibility of stronger exploitation in the future.



- The **Consumer Segment** is the primary contributor, accounting for **over 50% of total sales**, underscoring its critical role in business strategy.
- Although **Home Office** has the **smallest share**, it is considered a promising niche market amidst the growing trend of remote work.
- Therefore, it is crucial to focus on maintaining and developing tailored business strategies for each customer segment to optimize effectiveness.

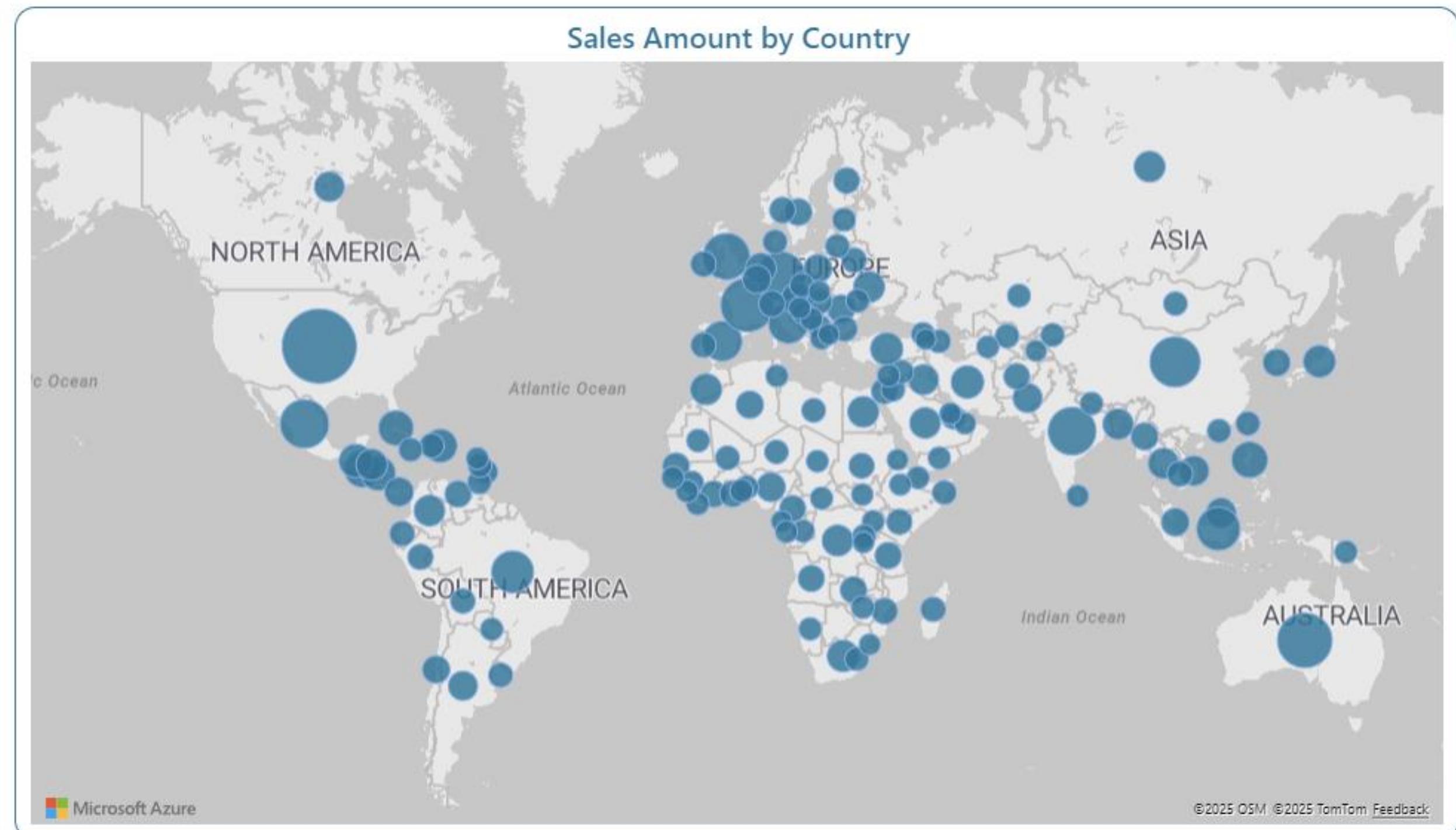


- **Standard Class** is the most preferred shipping method, contributing **nearly 60% of total sales**.
- As shipping speeds increase, customer usage tends to decline, with sales from **Same Day** being the **lowest at only 5.28%**.
- Therefore, it is essential to optimize the operations of Standard Class while enhancing fast shipping services (Same Day, First Class) to better meet diverse customer demands.



7. Data Analysis

- The company's market is continuously on a global scale.
- **The highest sales** comes from developed countries such as the **United States, Germany, China and Japan**, due to their stable economies and strong consumer demand.
- Asia is emerging as a rapidly growing region with significant sales potential and impressive growth rates.
- Revenue from African countries remains modest for now, but the region holds long-term opportunities if investments are made in infrastructure and improvements in customer outreach strategies.



7. Data Analysis

Insights

- Overview of business performance during the period 2011–2014: Promising sales growth accompanied by challenges in profitability.
- During the period 2011–2014, the company experienced impressive sales growth, particularly in the fourth quarter of each year, clearly demonstrating its ability to seize market opportunities and the effectiveness of its sales campaigns.
- However, the profitability picture showed significant fluctuations, not growing in tandem with revenue, highlighting an urgent need to optimize costs and improve operational efficiency. Nevertheless, the company has a solid foundation and significant potential for development in many aspects, provided that appropriate strategic adjustments are implemented.



7. Data Analysis

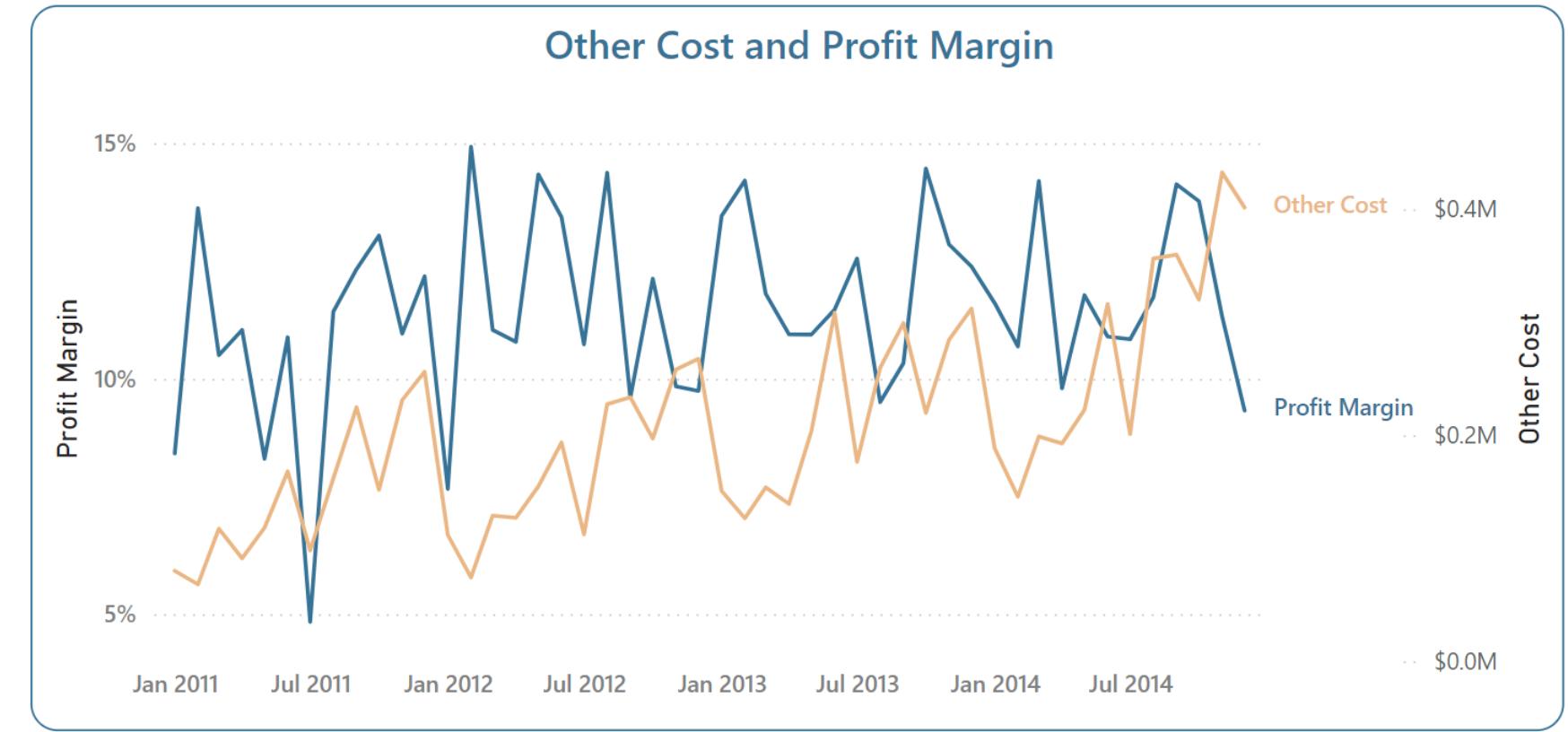
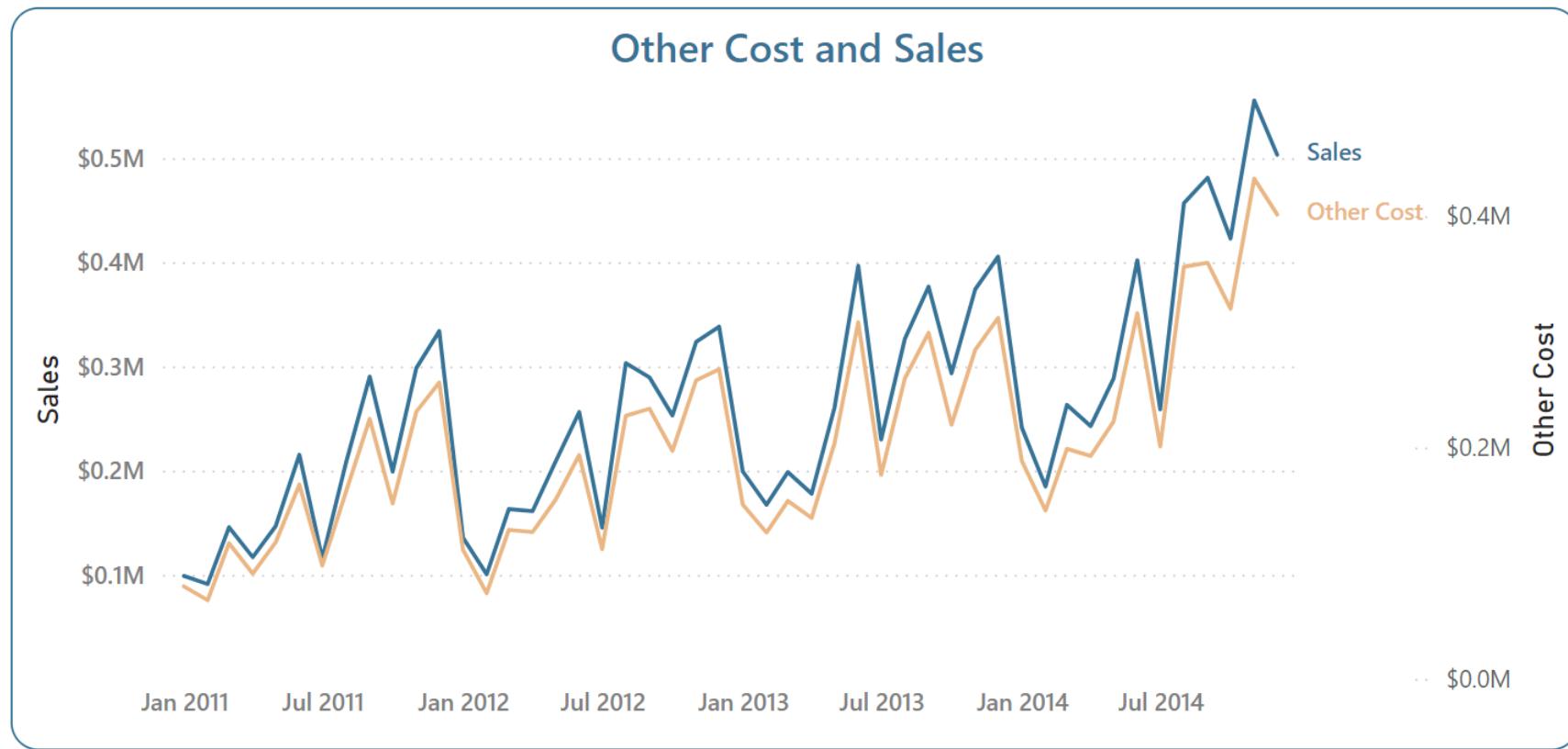
Business Question:

Why did the Profit Margin decrease even as Sales increased?



7. Data Analysis

7.2. Factors affecting Profit Margin



- Sales and Other Costs both increased over time from 2011 to 2014, reaching their highest point in 2014.
- There is a strong correlation between Other Costs and Sales, as reflected in the synchronization of the two lines. When Sales increase, Other Costs also increase, and vice versa.
- The close alignment of Other Costs with Sales is one of the reasons why Profit remains modest despite strong Sales growth.

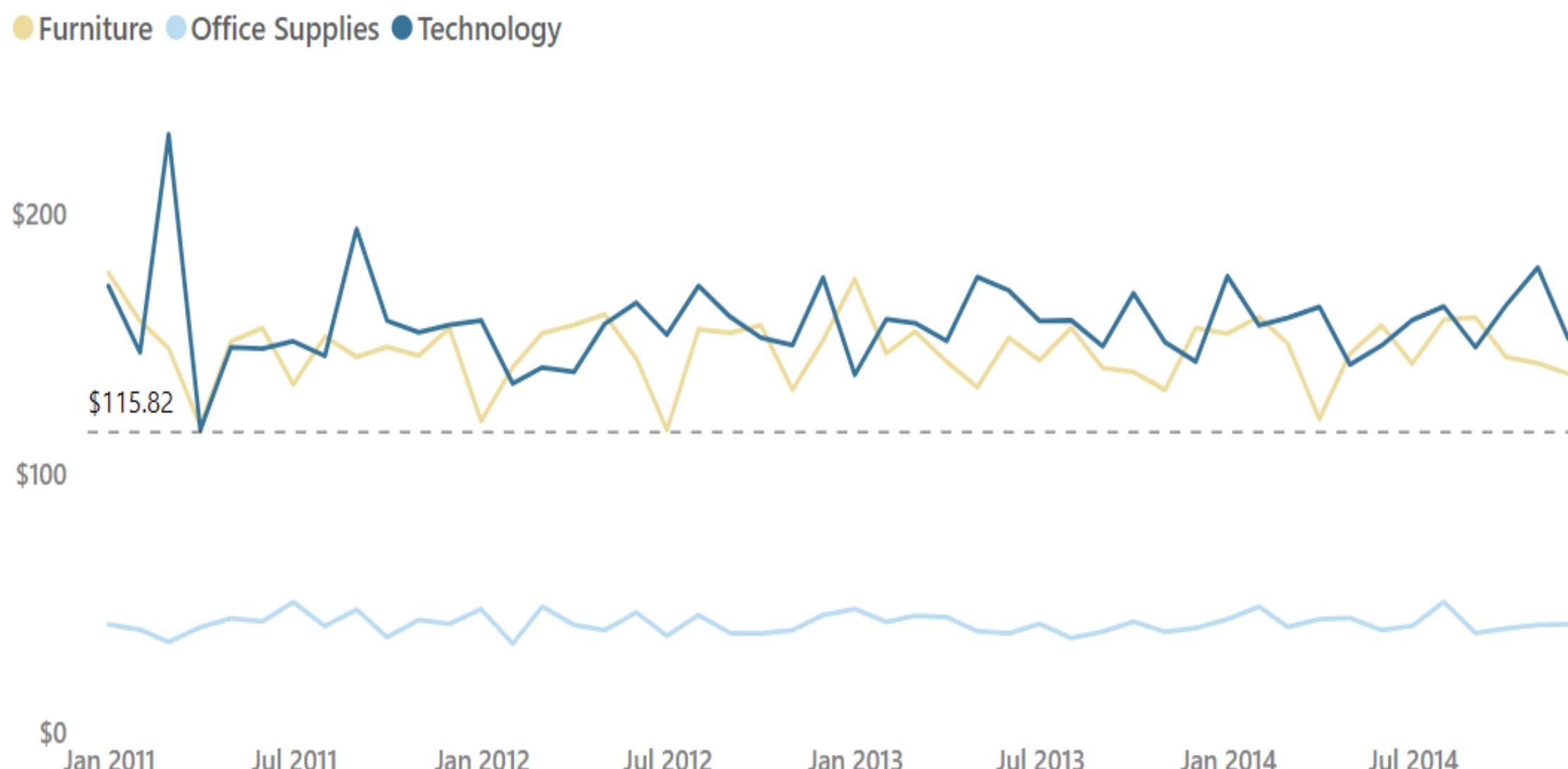
- There is a relatively **inverse correlation** between Other Costs and Profit Margin.
- Other Costs tend to increase over the years, while Profit Margin does not increase steadily but fluctuates significantly.
- When Other Costs increase, it **may impact the profitability** of the business.



OtherCost = Sales[Sales] - Sales[Profit] - Sales[ShippingCost]

7. Data Analysis

Average of UnitPrice by Year-Month and Category



The Impact of Product Pricing?

Product prices in each segment remained relatively stable over time, with no significant fluctuations, despite rising costs.

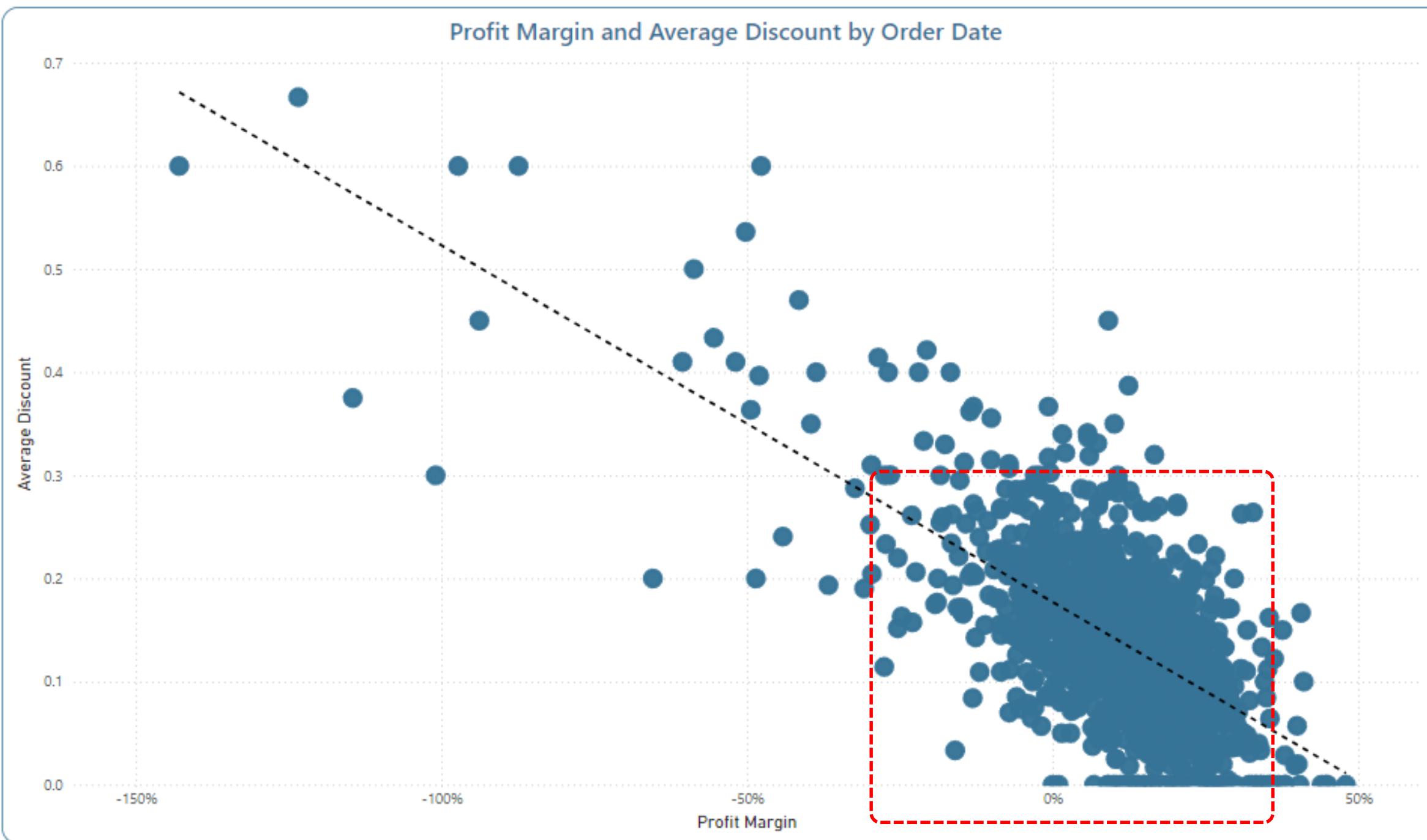
- **High-Price Segment (above average):** Furniture and Technology experienced greater price variability compared to Office Supplies.
- **Low-Price Segment (below average):** Office Supplies showed minimal price fluctuations during the 2011-2014 period.

Positives: The effective pricing strategy helped maintain customer purchasing power.

Negatives: While costs continued to rise, **stable selling prices limited profit growth**, leading to significant fluctuations in Profit Margins.



7. Data Analysis



The Impact of Discount?

Inverse Relationship: The trendline shows a negative correlation between Average Discount and Profit Margin.

Data Distribution:

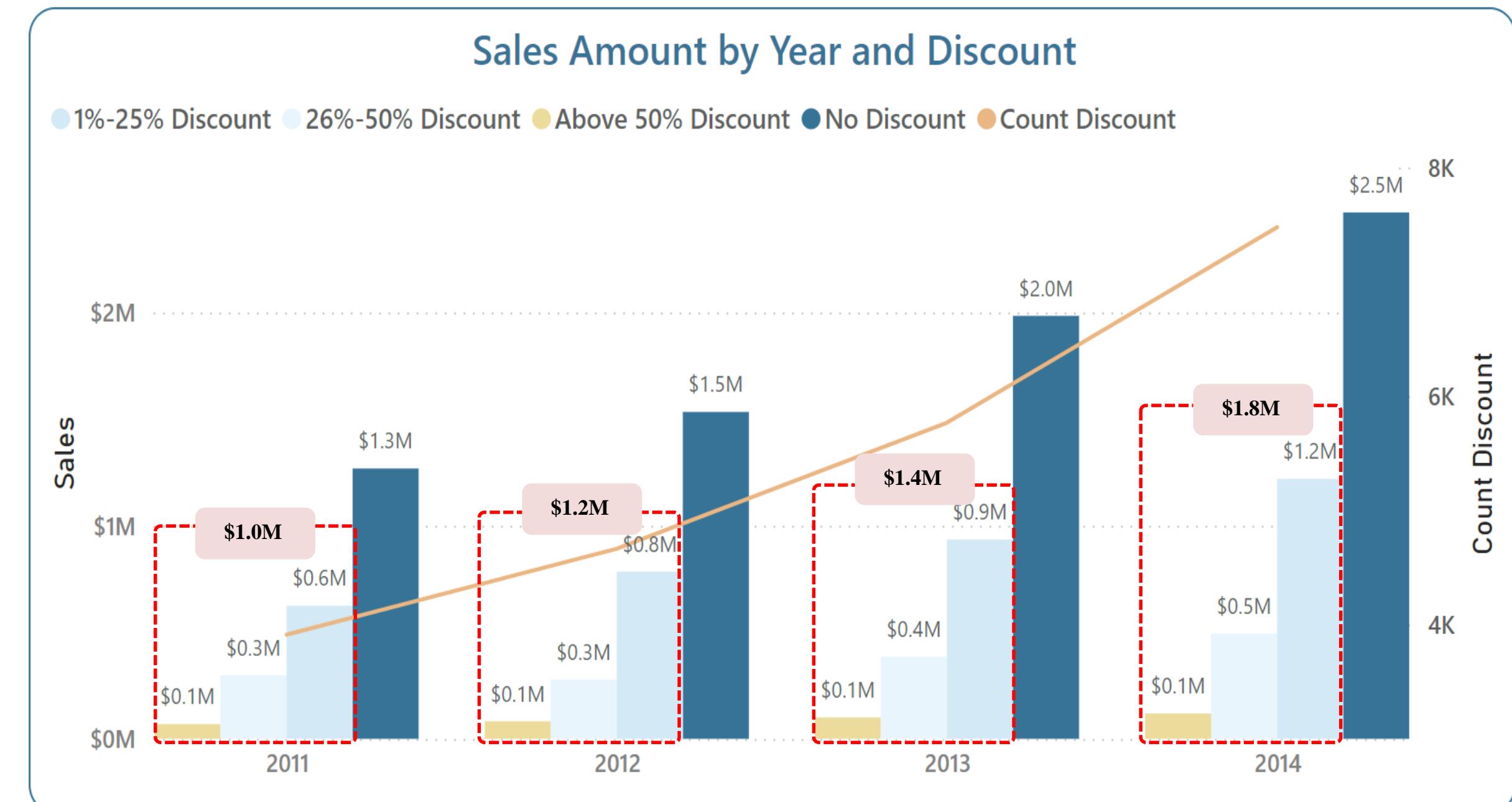
- Most data points are concentrated in the lower-right quadrant, where discounts range from 0% to 30% and profit margins fall between -25% and 25%. This indicates that the majority of orders have moderate discounts, with profit margins fluctuating within a relatively wide range.
- When discounts exceed 30%, the number of data points decreases significantly, and the proportion of negative profit margins increases. This suggests that deep discounts are not only less common but also more likely to lead to losses.



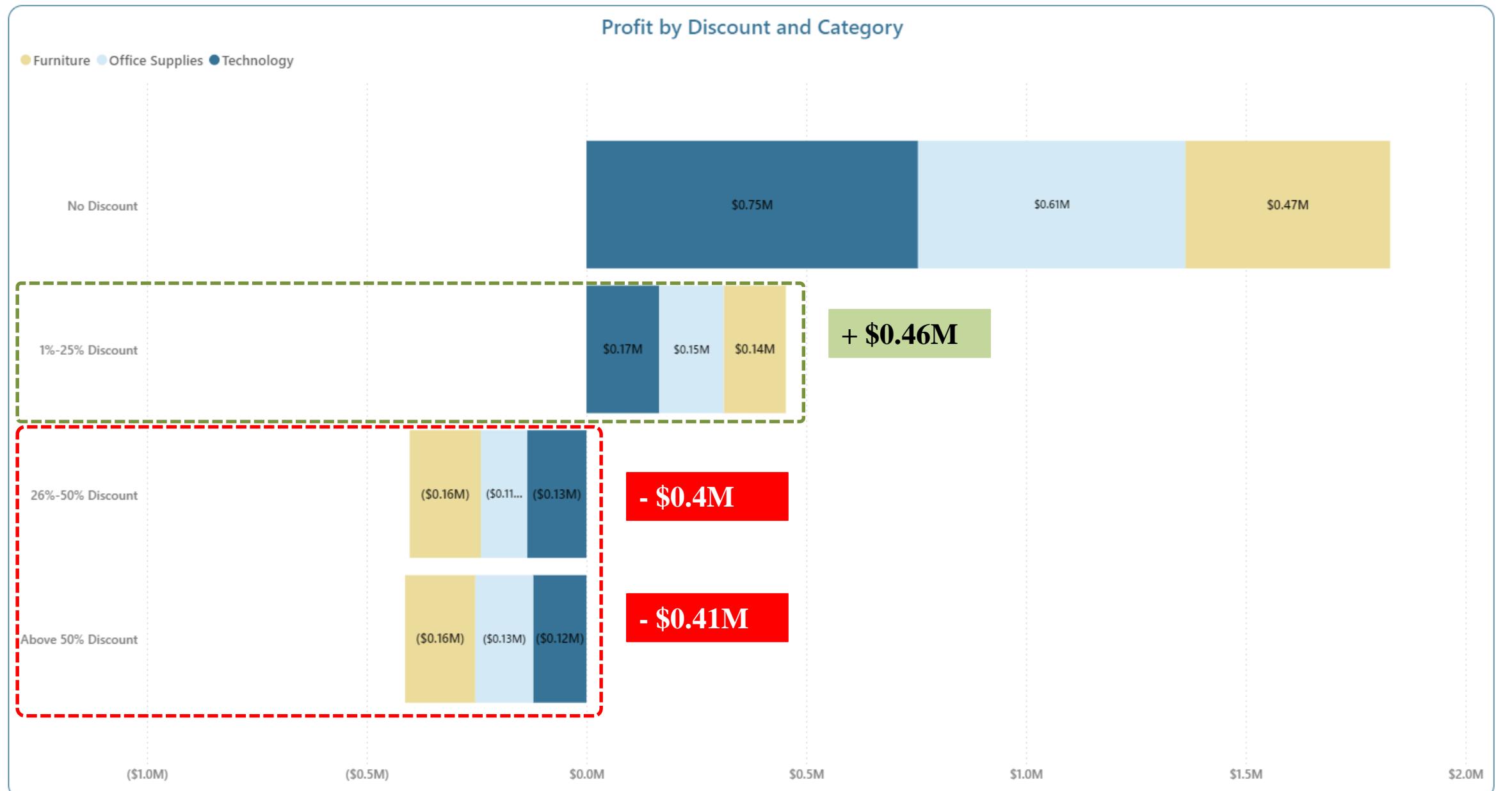
7. Data Analysis

How does Discount affect?

- Overall, **Total Sales** shows a steady upward trend as discount programs are implemented.
- However, the **No Discount** segment remains the primary contributor, accounting for 50% to 70% of annual sales.
- The 1%-25% discount range is the most common after No Discount, generating \$1.2 million in sales in 2014, double the \$0.6 million recorded in 2011.
- In contrast, discounts over 26% contribute insignificantly to sales, indicating that deeper discounts are neither widely applied nor particularly effective.
- Notably, the number of discount programs has steadily increased over the years, highlighting the strategic importance of discount policies in driving sales growth.



7. Data Analysis



Are Discounts truly effective?

Different discount levels yield varying effectiveness:

- **1%-25%:** Generates positive profits across all three segments, indicating that this is an effective and sustainable discount range.
- **26%-50%:** Results in negative profits, highlighting the inefficiency of applying this discount level.
- **Above 50%:** Causes even greater losses to profits, significantly eroding profit margins.
- **No Discount:** Contributions exceptional profits across all three segments, with **Technology** leading the way at \$0.75 million in profit.

High Discount levels are not only **ineffective** but can also lead to significant profit losses, **reducing profit margins** substantially.



7. Data Analysis

Insights

Steady Growth but Fluctuating Profit Margins:

- Sales and profits experienced steady growth from 2011 to 2014, especially with **Q4 consistently achieving the highest levels**, thanks to holiday seasons and promotional programs to stimulate demand.
- However, **profit margins were unstable** due to factors such as costs and discount policies.

Increased Other Costs Significantly Impact Profit Margins:

- Other costs increased in proportion to sales, creating significant pressure on profit margins.
- While **average selling prices remained unchanged**, rising costs have gradually narrowed profit margins over the years.
 - Solution: Implement stringent controls on other costs to optimize profitability.

Suboptimal Discounting Effectiveness :

- **50%-70% of revenue comes from non-discounted orders**, which positively contributes to profit margins.
- Discounts ranging from **1%-25% do not yield significant profits**, while discounts exceeding **26% result in losses**.
 - Solution: Adjust promotional strategies, focusing on **selective discounts** instead of widespread applications.

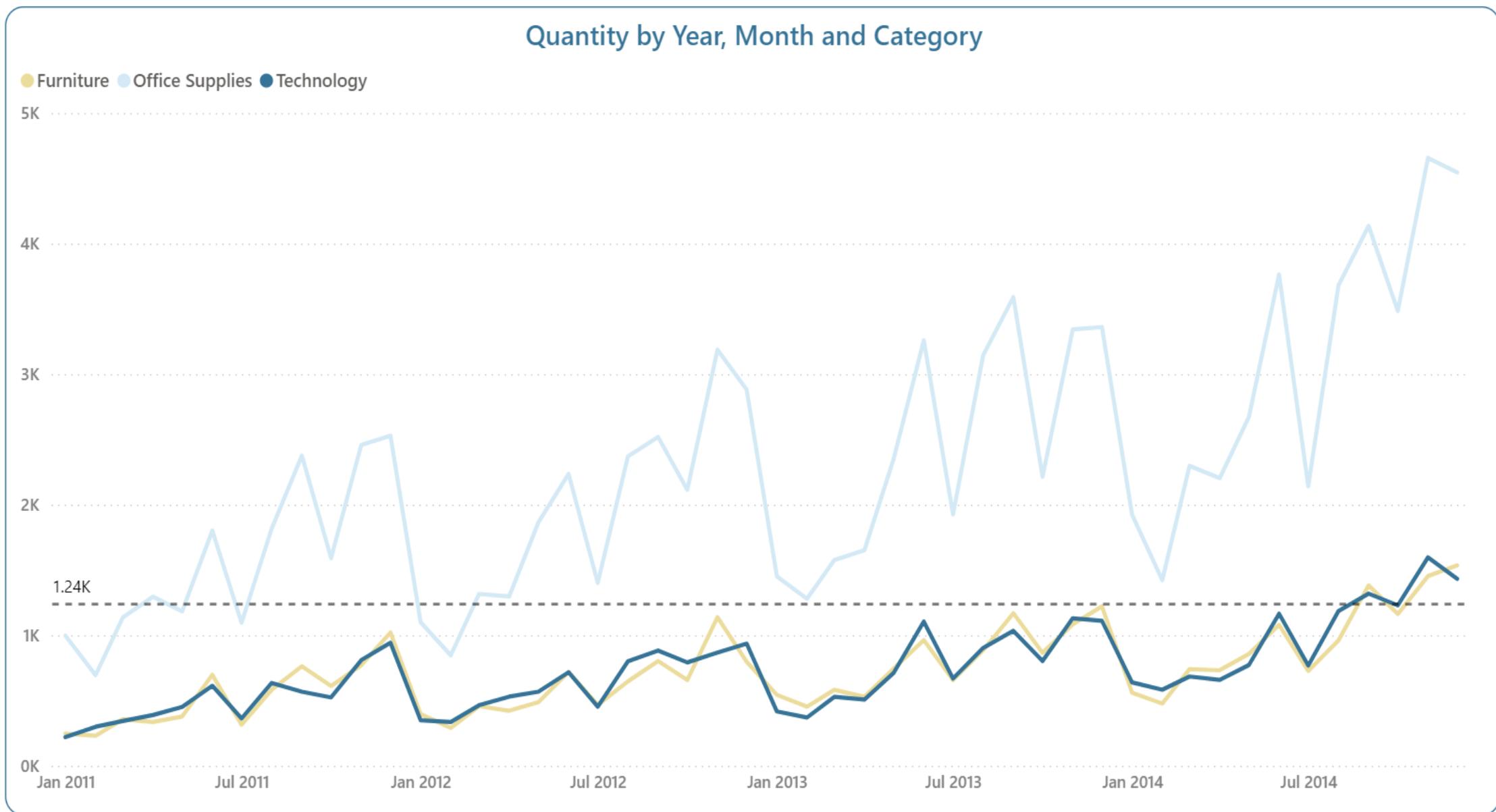
Summary:

- The company has good growth potential but must control costs and optimize its discounting strategy to improve profit margins.
- Develop **flexible promotional programs** and avoid ineffective widespread discounting.
- Control other costs to ensure that profit margins are not continuously eroded over time.



7. Data Analysis

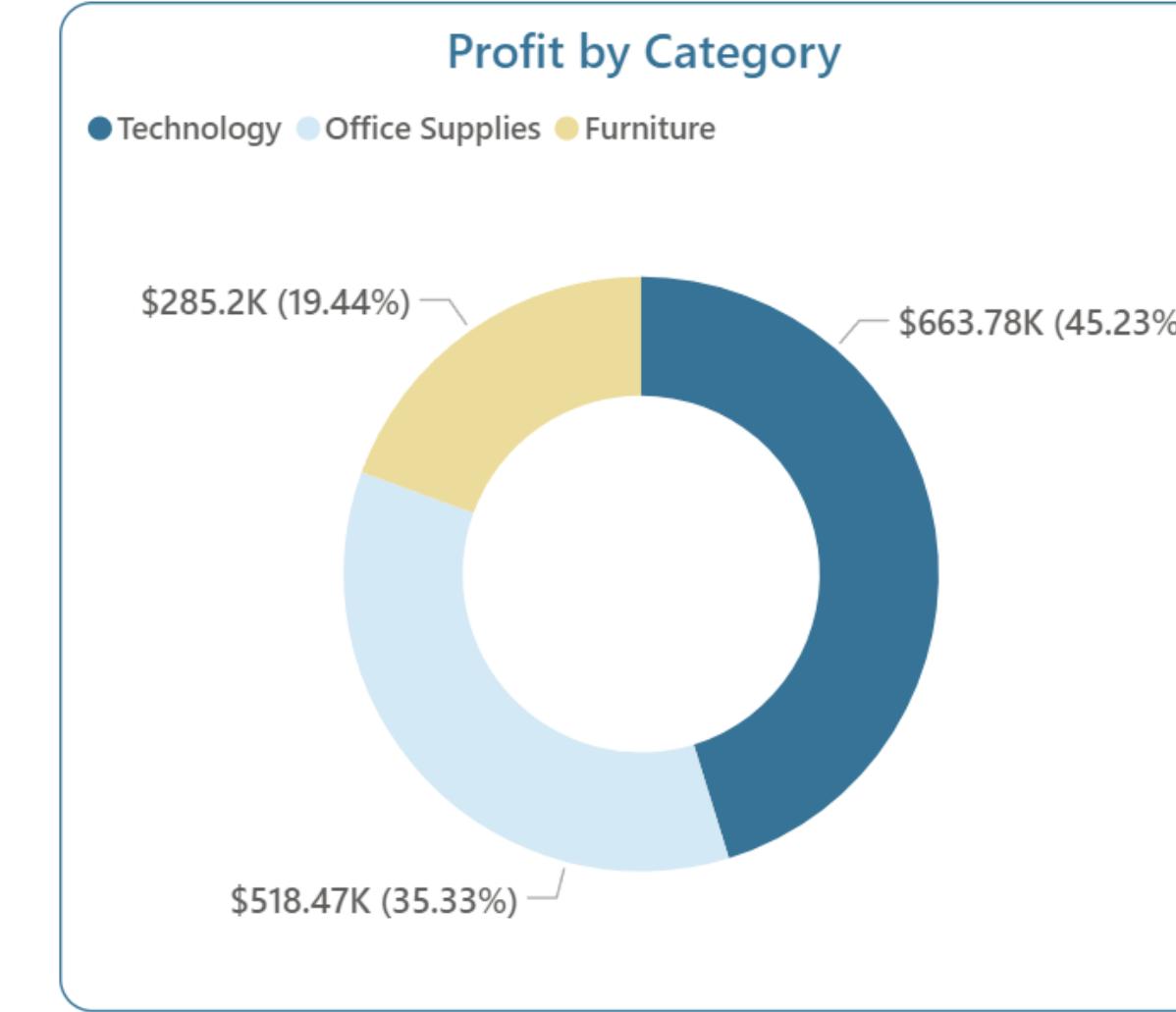
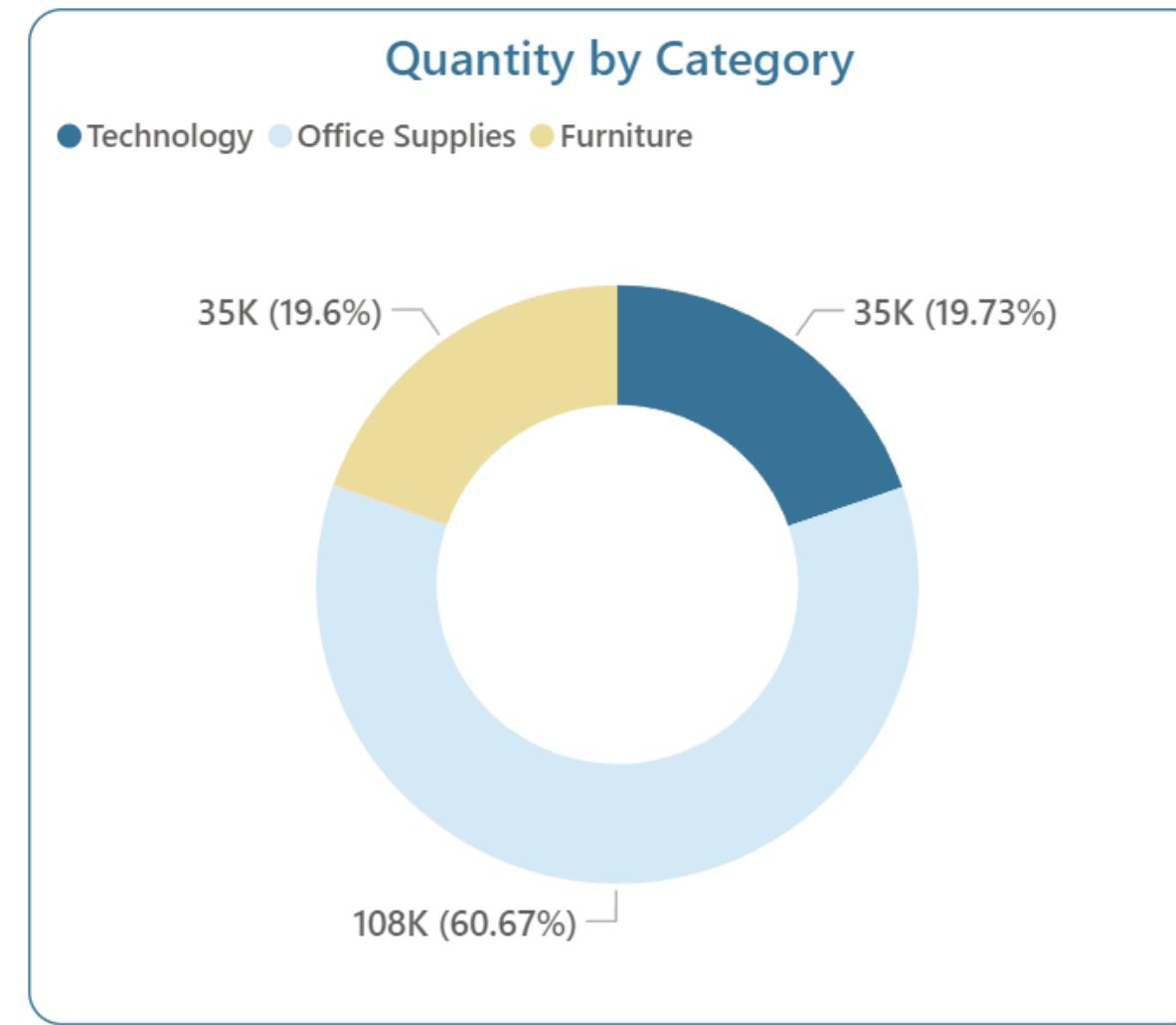
7.3. Summary Category



- From 2011 to 2014, the total quantity of products sold exhibited a steady growth trend, aligning with the increase in revenue.
- Although **Technology** initially led in revenue generation, **Office Supplies** truly stood out by surpassing the average sales volume in June 2011 and maintaining consistent growth until the end of 2014, solidifying its role as the primary category.
- Meanwhile, both **Furniture** and **Technology** showed more gradual growth. Most of their sales remained below the average level until September 2014, when they began to experience a notable breakthrough.



7. Data Analysis

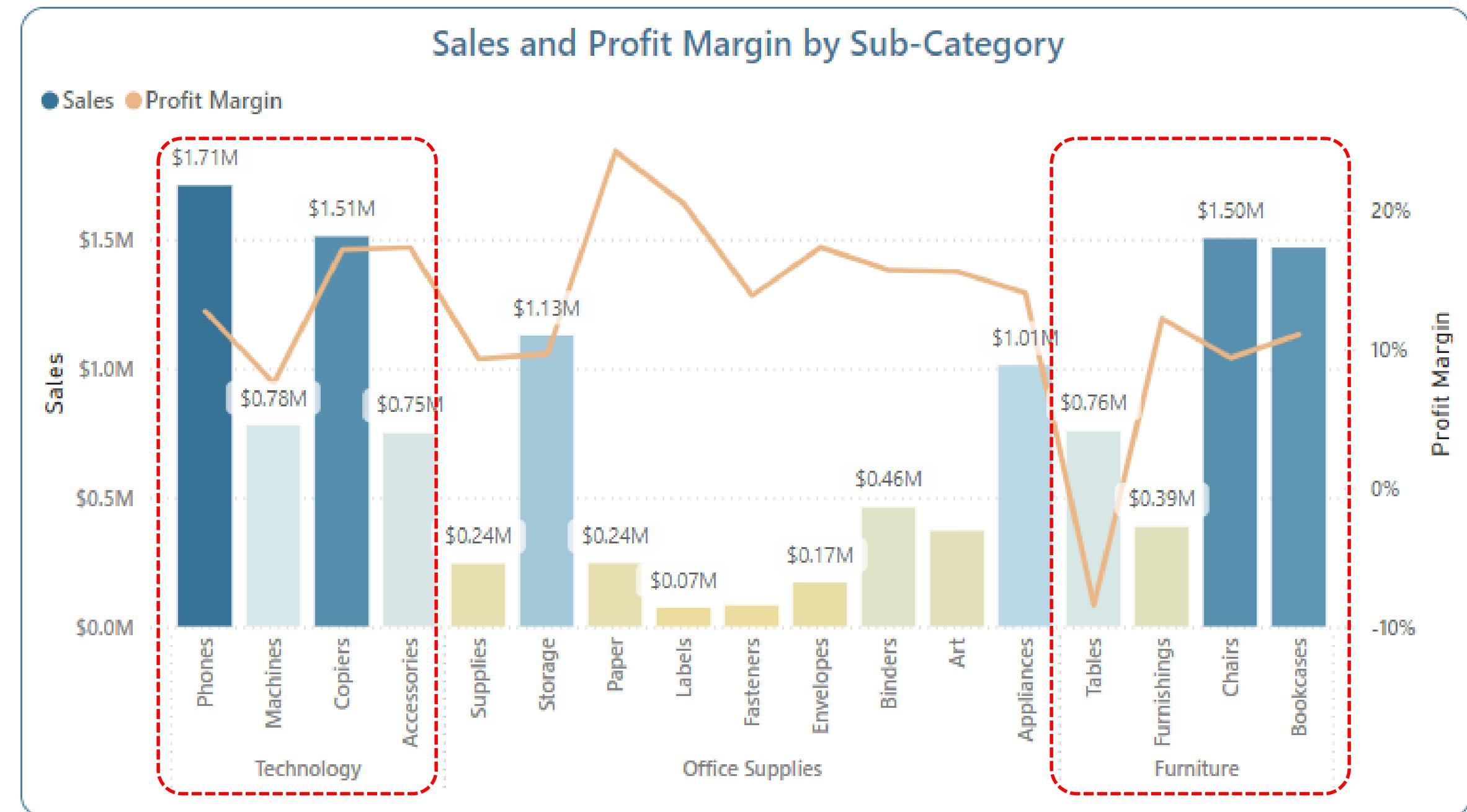


- Office Supplies account for 60.67% of total sales (108K units), but due to low pricing, they contribute only 35.33% of total profit.
- Technology is the highest-value category, making up just 19.73% of sales volume but contributing 45.23% of total profit (\$663.78K), highlighting strong growth potential.
- Furniture has high selling prices but low sales volume (19.6% of total sales) and the lowest profit contribution (19.44% of total profit).



7. Data Analysis

- Phones & Copiers (Technology) and Chairs and Bookcases (Furniture) are the top four best-selling categories.
- Technology maintains a stable profit margin, with Copiers achieving the highest margin.
- Furniture shows clear differentiation, as Chairs and Bookcases sustain stable profitability, while Furnishings have a negative profit margin.
- Office Supplies have lower sales volume but the highest profit margin, indicating strong potential for efficient exploitation.



7. Data Analysis

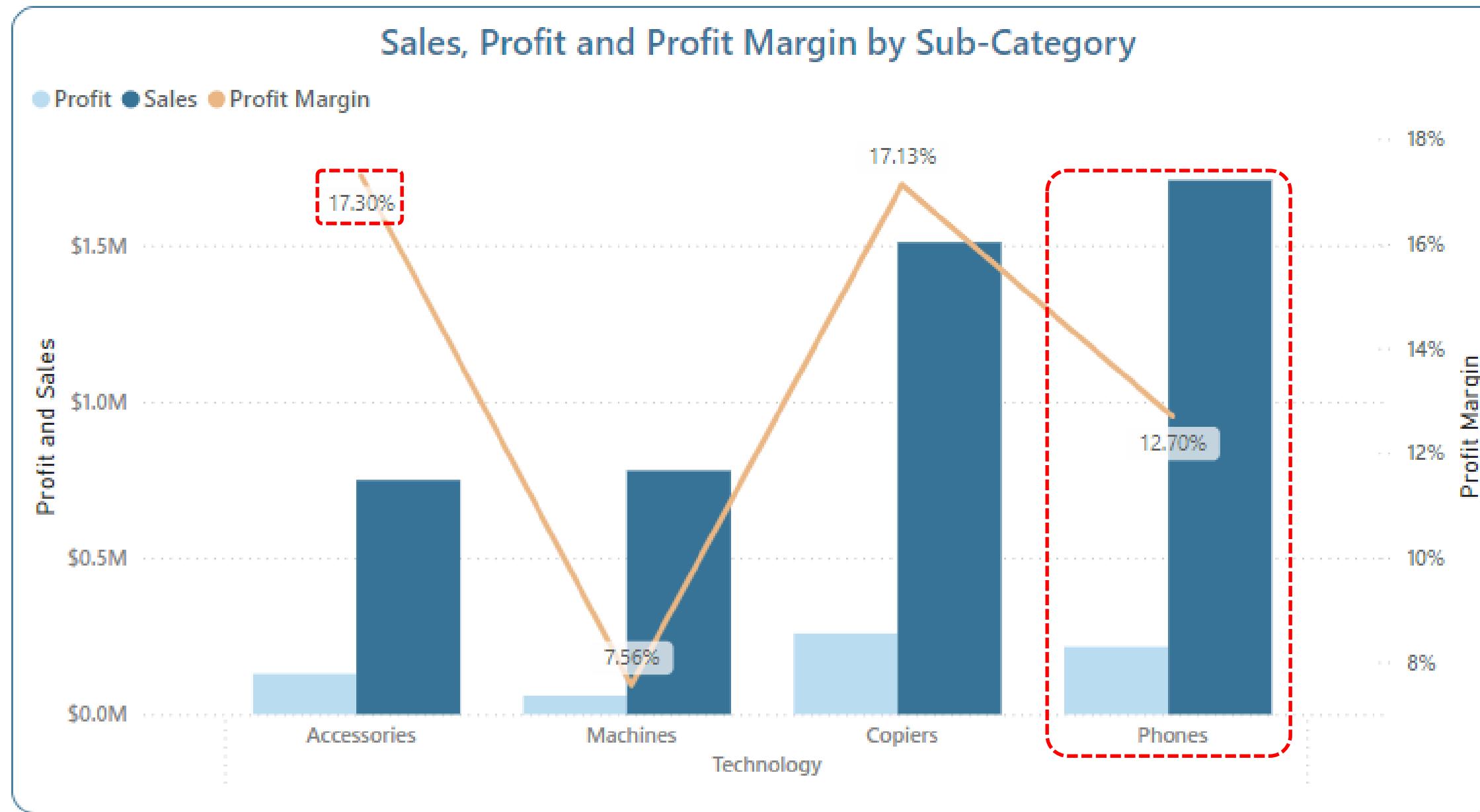
Business Question:

What factors are affecting the profitability of each Category?



7. Data Analysis

Technology

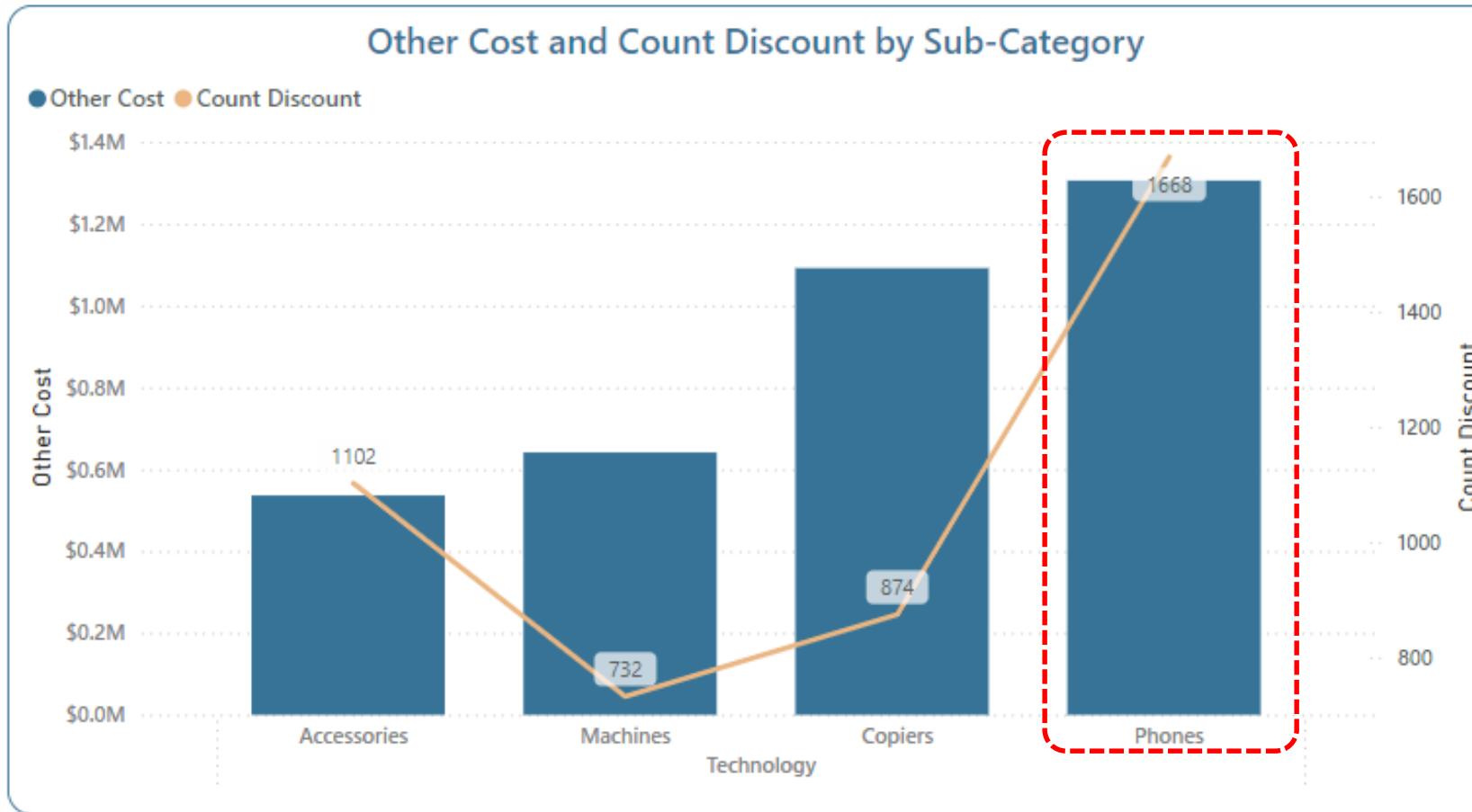


- Phones have the highest sales in the Technology category, making a significant contribution to total revenue. However, the profit margin is only 12.70%, an average level, indicating potential for improving business efficiency.
- Copiers have a high profit margin (17.13%) despite lower sales than Phones, demonstrating strong profitability.
- Accessories have the lowest sales but achieve the highest profit margin (17.30%), showing high profitability despite not being a core product.
- Overall trend: The Technology category is performing well, with three out of four products exceeding a 10% profit margin.



7. Data Analysis

Technology

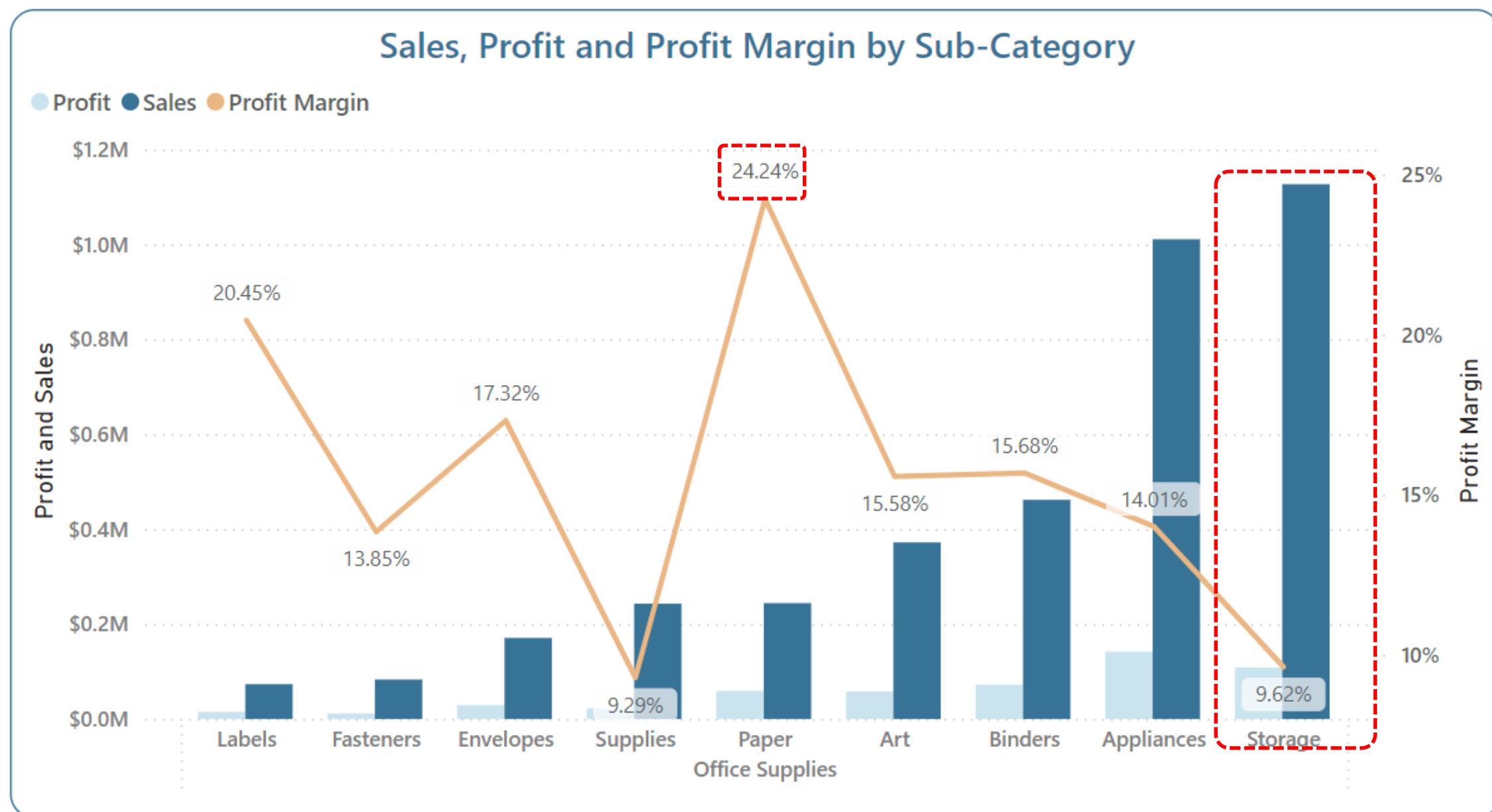


- **Phones:** Low profit margin due to high discount and large Other Cost.
- **Accessories:** Achieved the highest profit margin, with medium discount, low Other Cost and high revenue (10.9K products sold).
- **Machines:** Has the lowest quantity sold (4.9K products sold), despite low discount and Other Cost not being high.
- Other Cost, Discount and Quantity Sold are 3 important factors that strongly influence the profit margin of the Sub-Categories within the Technology group.



7. Data Analysis

Office Supplies

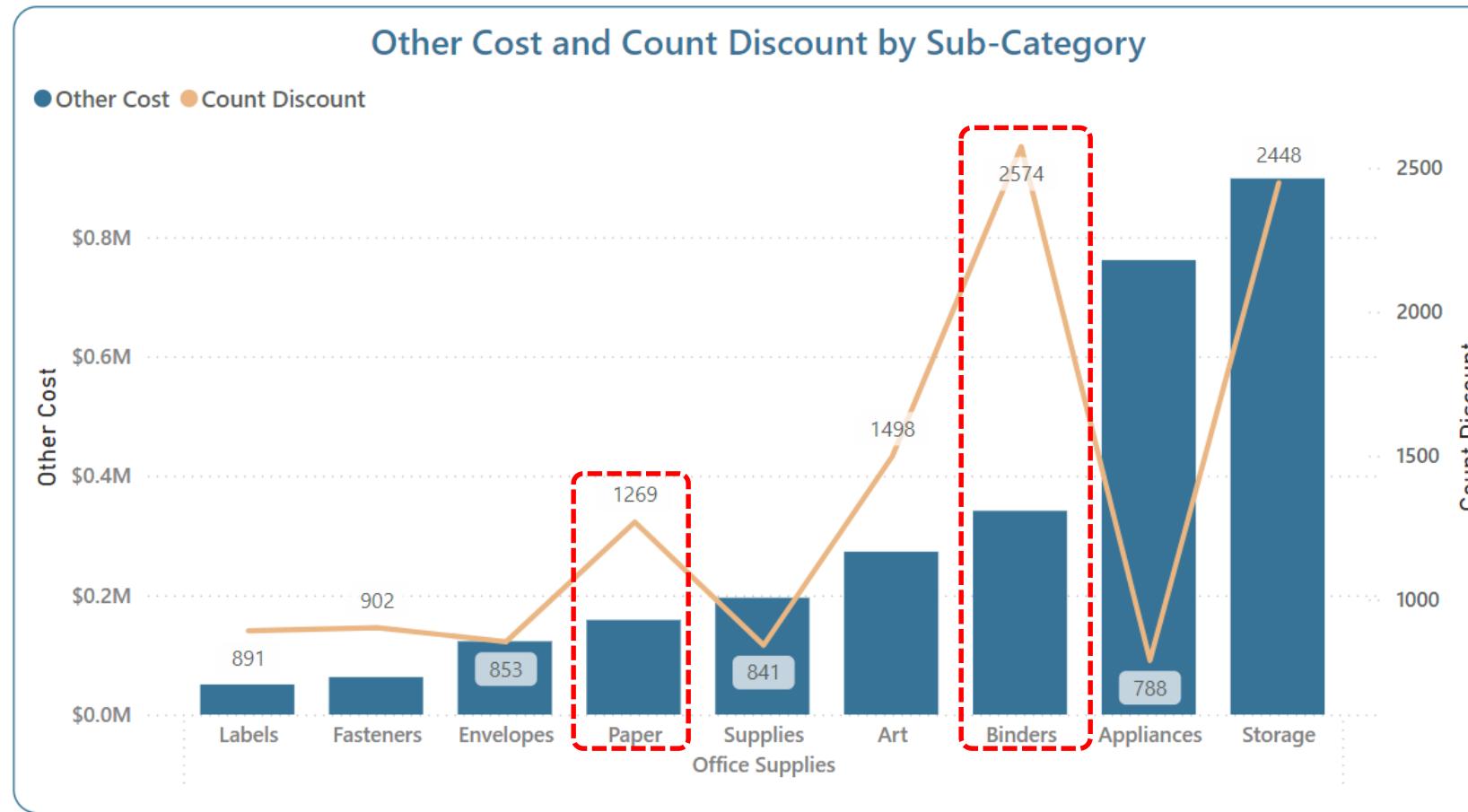


- Storage leads in sales and almost takes the top position in profit, but the profit margin belongs to the lowest group (9.62%), indicating many issues need to be improved.
- Paper sales is more than 5 times lower than Storage, but the profit margin is the highest (24.24%), showing high efficiency.
- General Trend: Product groups with high sales usually have lower profit margins and vice versa.



7. Data Analysis

Office Supplies

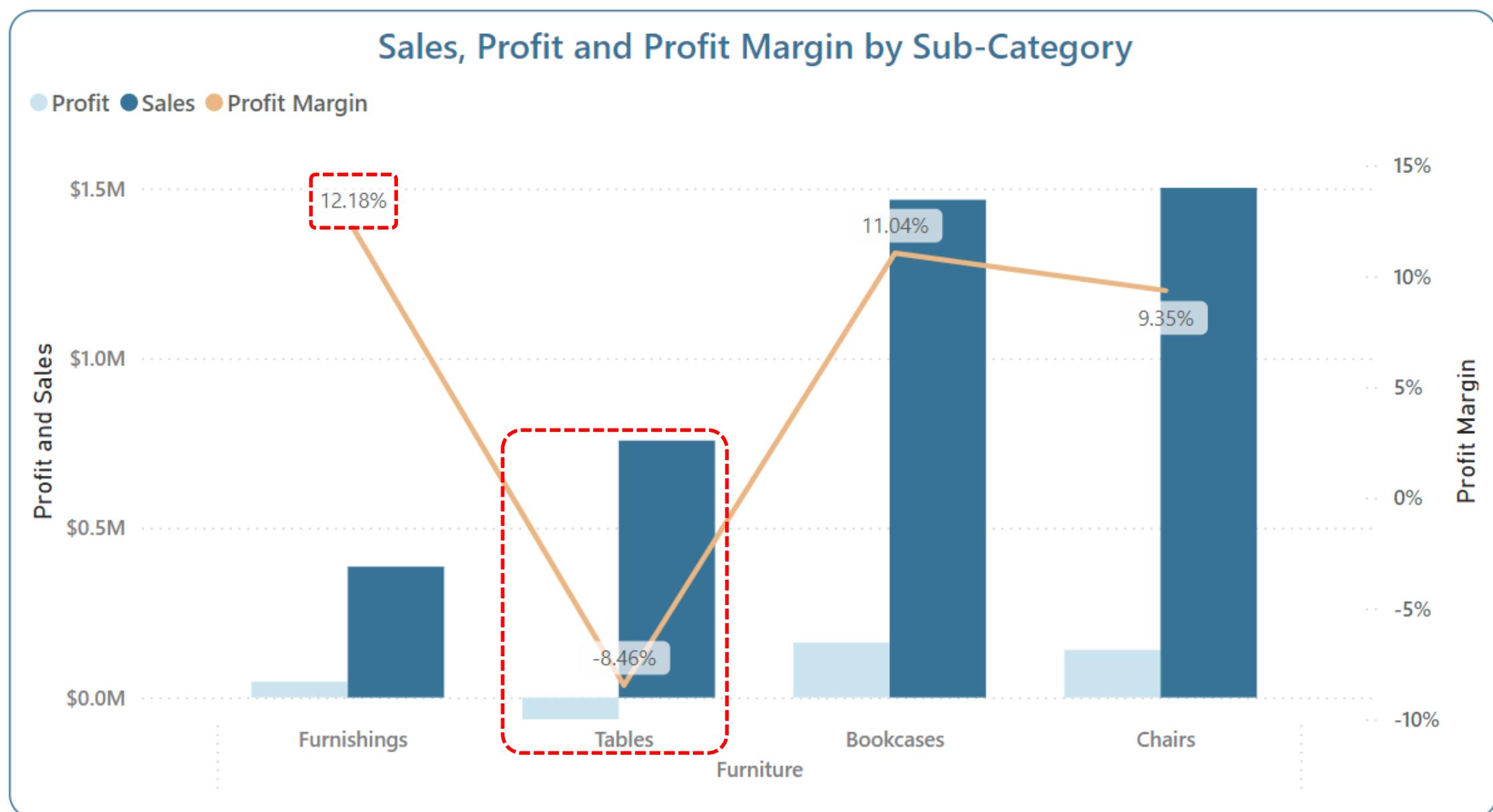


- **Storage:** Poor profit margin due to high Other Cost and large Discount amount, reducing business efficiency.
- **Paper:** Although Other Cost is low (under 0.2 million USD), it still achieves high sales (top 4 with 12.8K products sold), showing good profitability.
- **Binders:** Has low Other Cost, but highest Discount and leading product sold (21.4K products sold), proving efficiency when applying discounts.
- Applying Discount to the right target audience, controlling Other Cost reasonably will promote purchasing power and positive profit margin results.



7. Data Analysis

Furniture

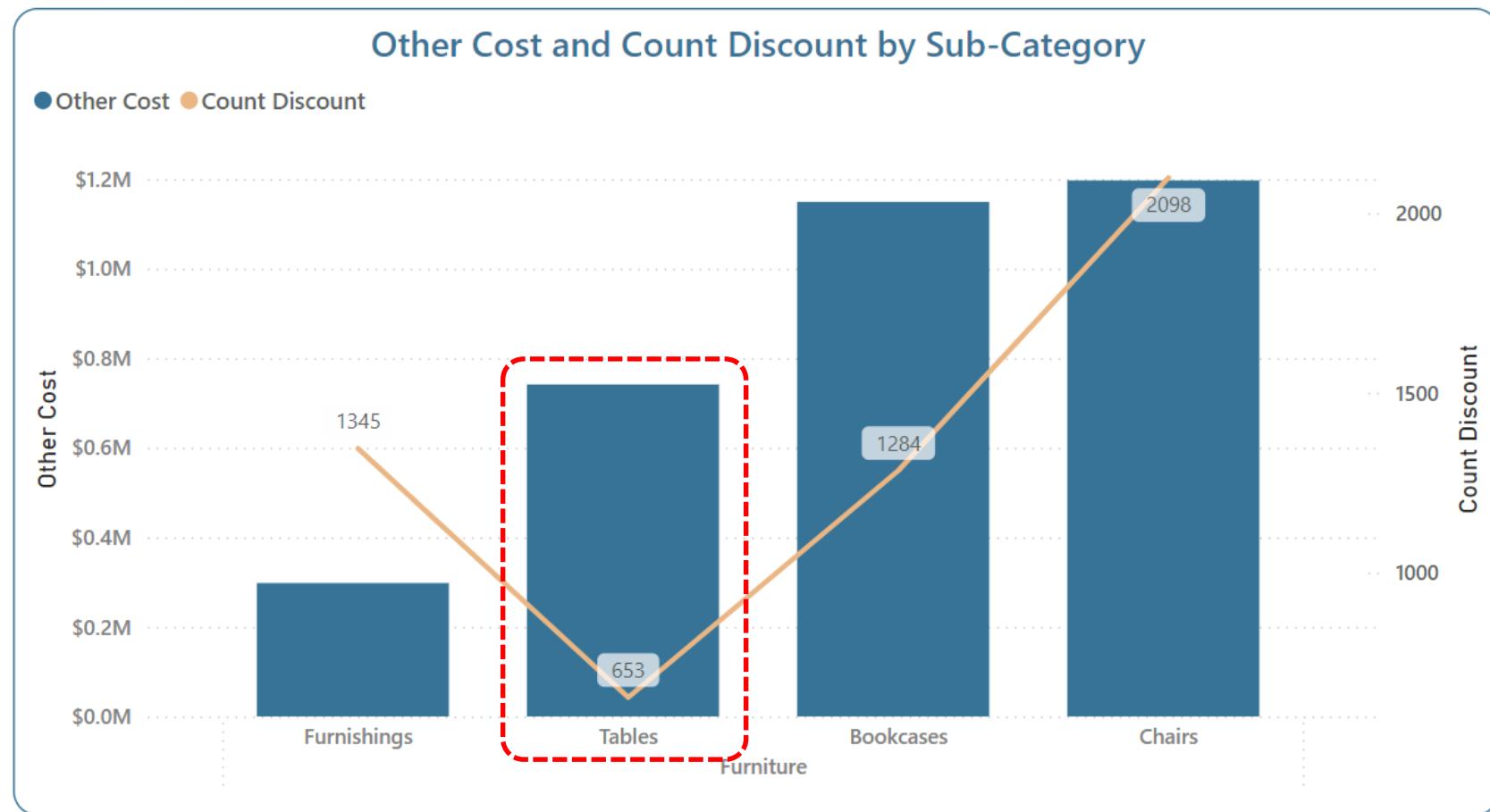


- Chairs and Bookcases have high sales and stable profit margins (9.35% and 11.04%), playing an important role in the Furniture category.
- Furnishings have the highest profit margin (12.18%) but low revenue.
- However, the company needs to review the business strategy for Tables, because this category is currently losing money.



7. Data Analysis

Furniture



- Furnishings have the lowest Other Cost and effective discount application, helping to rank second in quantity sold (3.1K products sold) and lead in profit margin within the Furniture group.
- Bookcases & Chairs, although having high costs and large discount levels, still achieve high sales (8.3K & 12.3K products sold), affirming their role as core categories.
- Tables are facing serious problems:
 - Costs are not too high, discounts are low, but revenue is the lowest in the group (3.1K products sold).
 - Is the only category incurring losses in profit margin and having negative growth.



7. Data Analysis

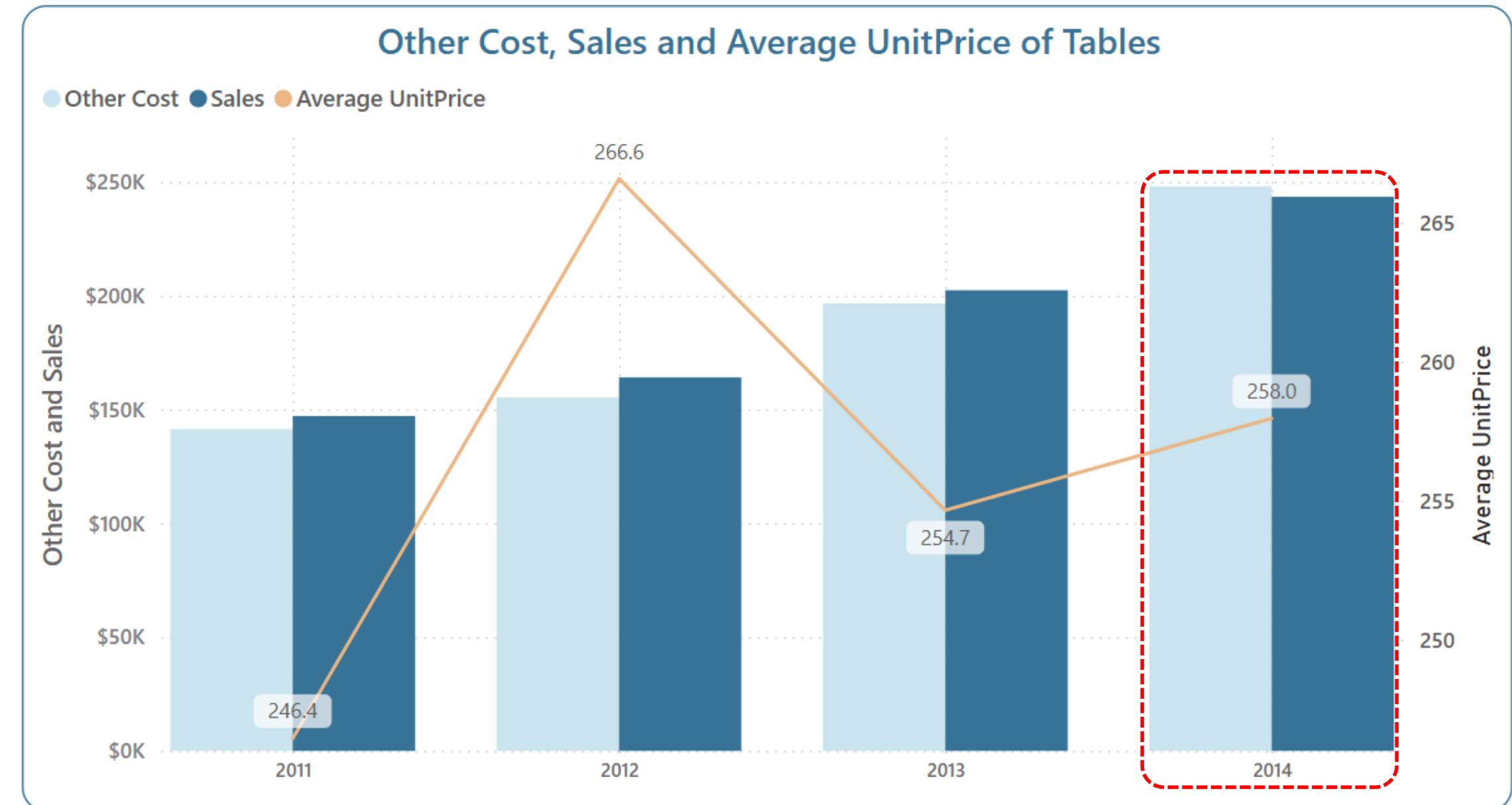
Business Question:

Why are Tables not profitable?



7. Data Analysis

- Sales of Tables have grown steadily over the years, indicating that the product still has certain demand in the market.
- Other Costs increased over time, peaking in 2014, even exceeding sales, putting pressure on profits.
- The average selling price fluctuated sharply, peaking in 2012 (266.6 USD), but decreased in 2013 (254.7 USD) before slightly recovering in 2014 (258.0 USD).
- Negative impact on profit and profit margin, due to rising costs while selling prices are unstable, making Tables an ineffective category.



7. Data Analysis

7.4. Cross Sell

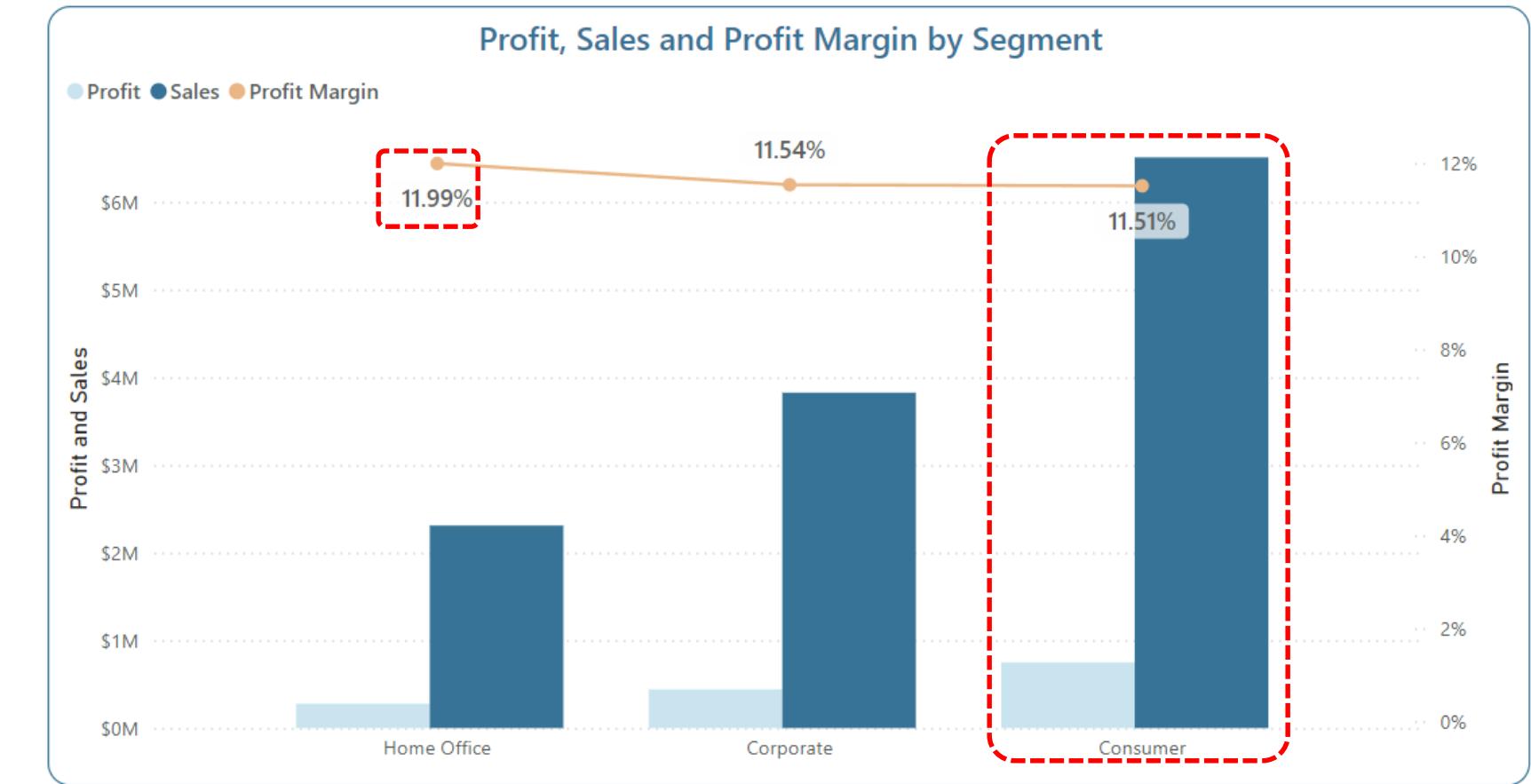
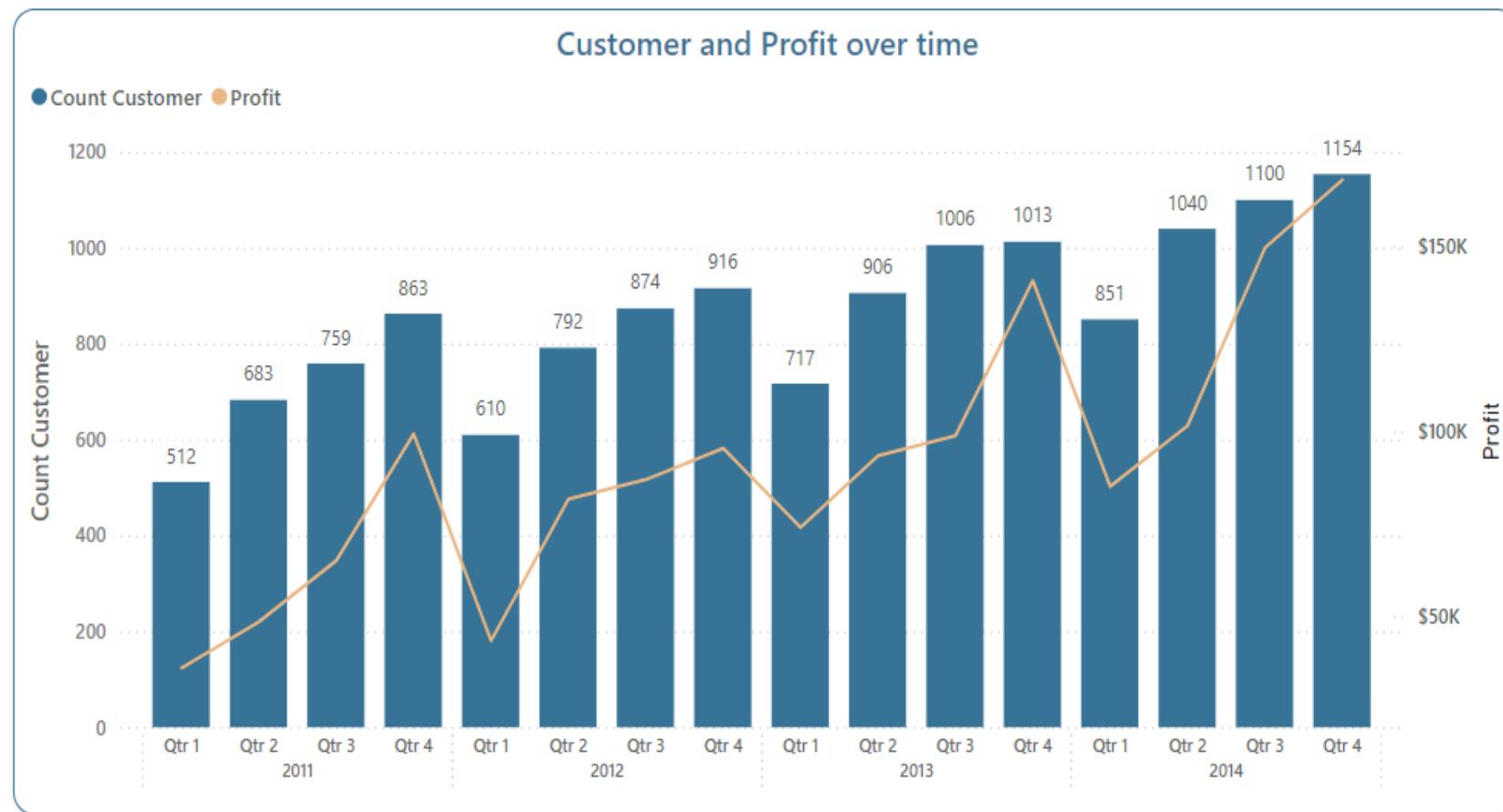
Product Name (Premise)	Order (Premise)	Product Name Consequence	Order (Consequence)	Support	Confidence	Lift
4009 Highlighters		1 Iceberg Mobile Mega Data/Printer Cart	8	0.00%	100.00%	3,219.00
4009 Highlighters		1 Logitech Wireless Marathon Mouse M705	8	0.00%	100.00%	3,219.00
Acco Glide Clips		1 Hand-Finished Solid Wood Document Frame	3	0.00%	100.00%	8,584.00
Ativa V4110MDD Micro-Cut Shredder	2	Staples	222	0.01%	100.00%	116.00
Avaya IP Phone 1140E VoIP phone		1 Global Wood Trimmed Manager's Task Chair, Khaki	14	0.00%	100.00%	1,839.43
Avaya IP Phone 1140E VoIP phone		1 Gould Plastics 18-Pocket Panel Bin, 34w x 5-1/4d x 20-1/2h	10	0.00%	100.00%	2,575.20
Barricks Coffee Table, with Bottom Storage		1 Binney & Smith Canvas, Easy-Erase	29	0.00%	100.00%	888.00
Barricks Coffee Table, with Bottom Storage		1 Brother Personal Copier, Color	18	0.00%	100.00%	1,430.67
Barricks Coffee Table, with Bottom Storage		1 Novimex Round Labels, Alphabetical	19	0.00%	100.00%	1,355.37
Barricks Non-Folding Utility Table with Steel Legs, Laminate Tops		1 Nortel Business Series Terminal T7208 Digital phone	3	0.00%	100.00%	8,584.00
Barricks Non-Folding Utility Table with Steel Legs, Laminate Tops		1 XtraLife ClearVue Slant-D Ring Binders by Cardinal	5	0.00%	100.00%	5,150.40
Barricks Round Table, Rectangular		1 Acme Ruler, High Speed	18	0.00%	100.00%	1,430.67
Barricks Round Table, Rectangular		1 Ibico Binding Machine, Recycled	43	0.00%	100.00%	598.88
Belkin 7 Outlet SurgeMaster Surge Protector with Phone Protection		1 Jabra BIZ 2300 Duo QD Duo Corded Headset	4	0.00%	100.00%	6,438.00
Total	25752		25752			

- A very low Support indicates that the data contains a wide variety of products and orders, leading to a decreased frequency of specific product pairs.
- Meanwhile, a very high Confidence suggests that when one product (Premise) appears, the likelihood of the other product (Consequence) appearing is very high.
- On the other hand, a highly fluctuating Lift coefficient indicates that product pairs have a very strong relationship, with a much higher likelihood of being purchased together compared to random chance.



7. Data Analysis

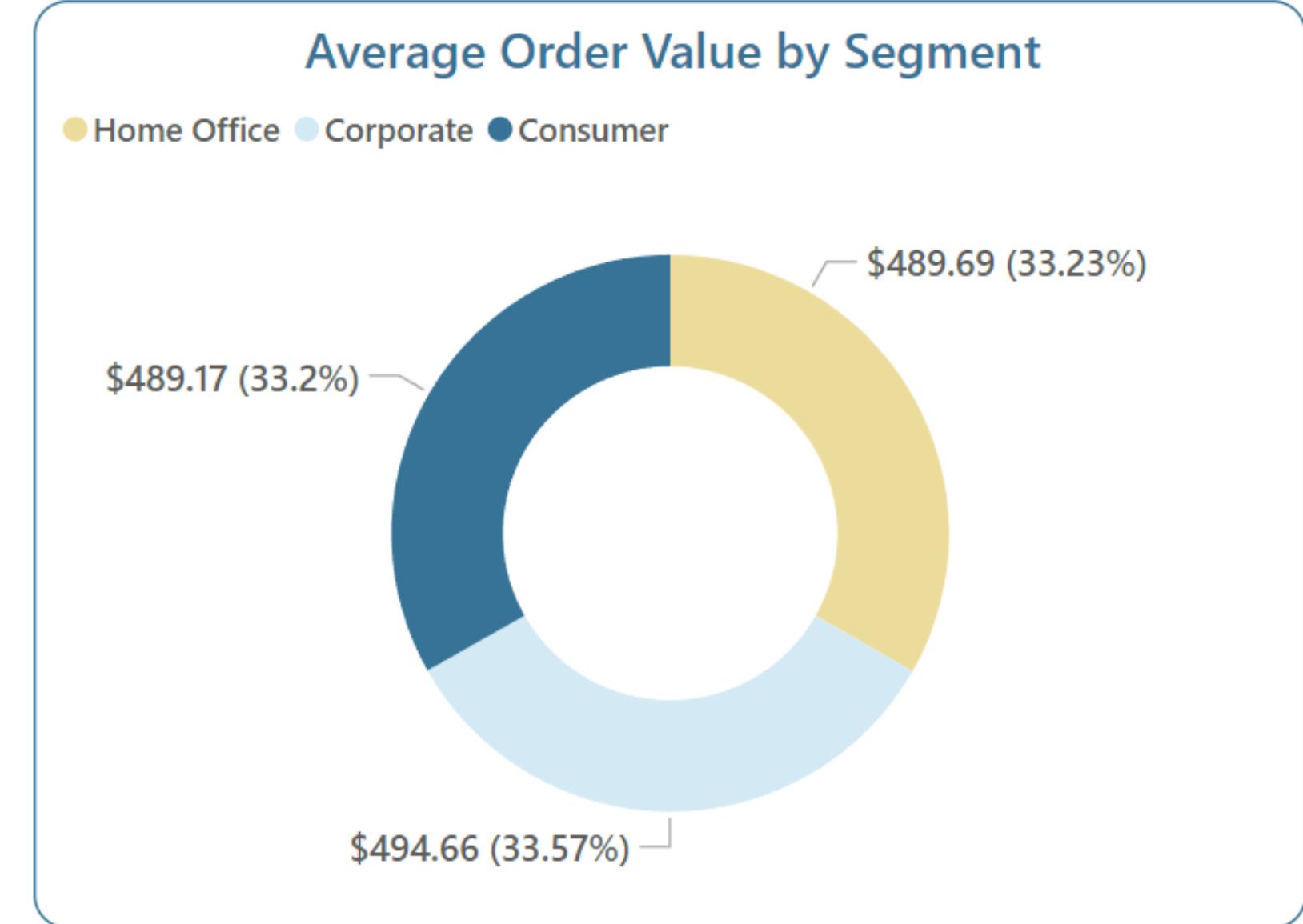
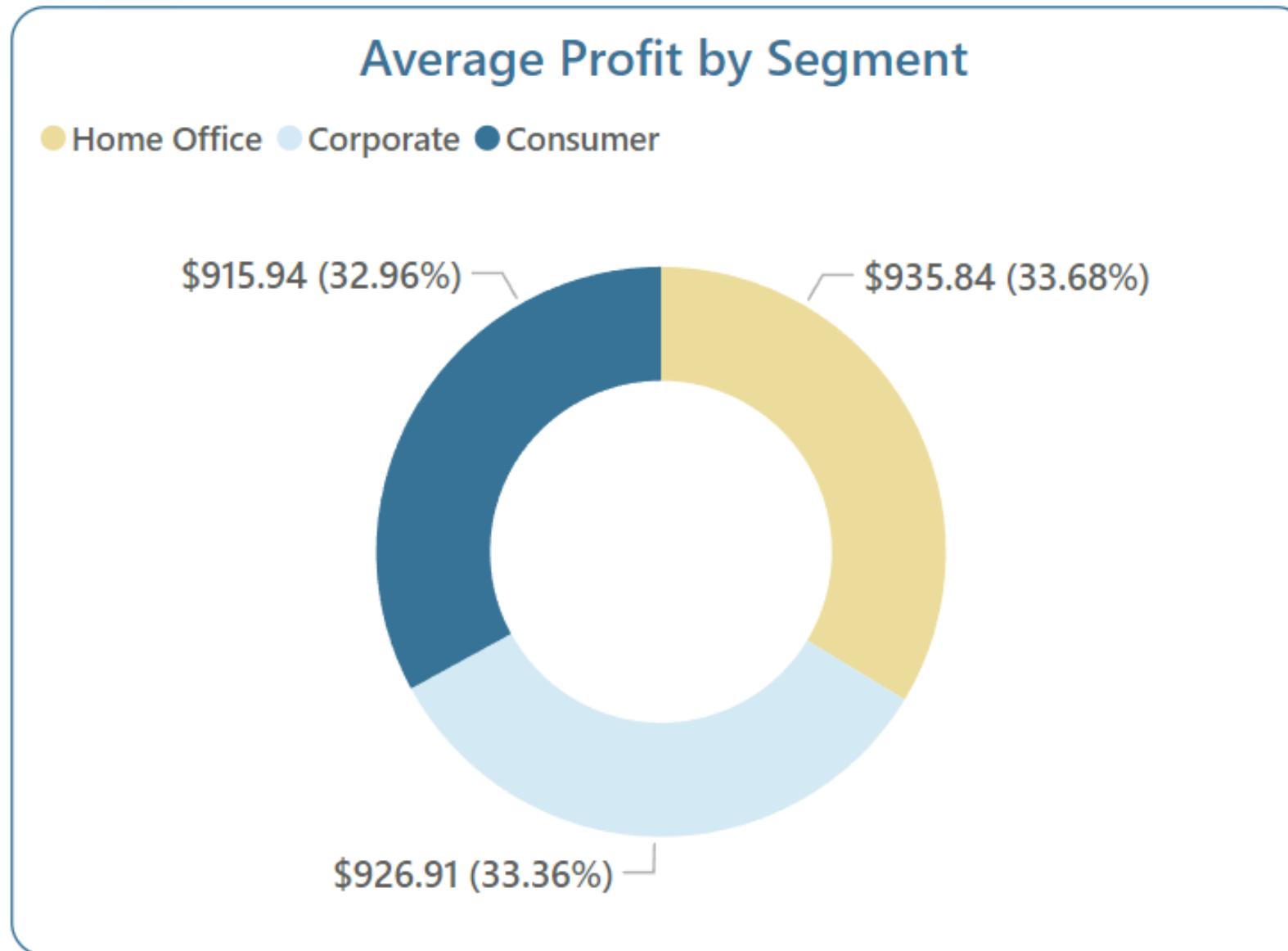
7.5. Summary Customer



- As of the end of 2014, the business had a total of 1,590 customers.
- The number of customers and profit both gradually increased over time, showing stable expansion.
- The Consumer group contributes the most in terms of both sales and profit.
- Although having a low number of customers, Home Office has the highest profit margin, showing better profitability compared to other groups.



7. Data Analysis



It can be seen that all three customer groups (Home Office, Corporate, Consumer) have a relatively even distribution in terms of both average profit and average order value (fluctuating around 33% for each group)



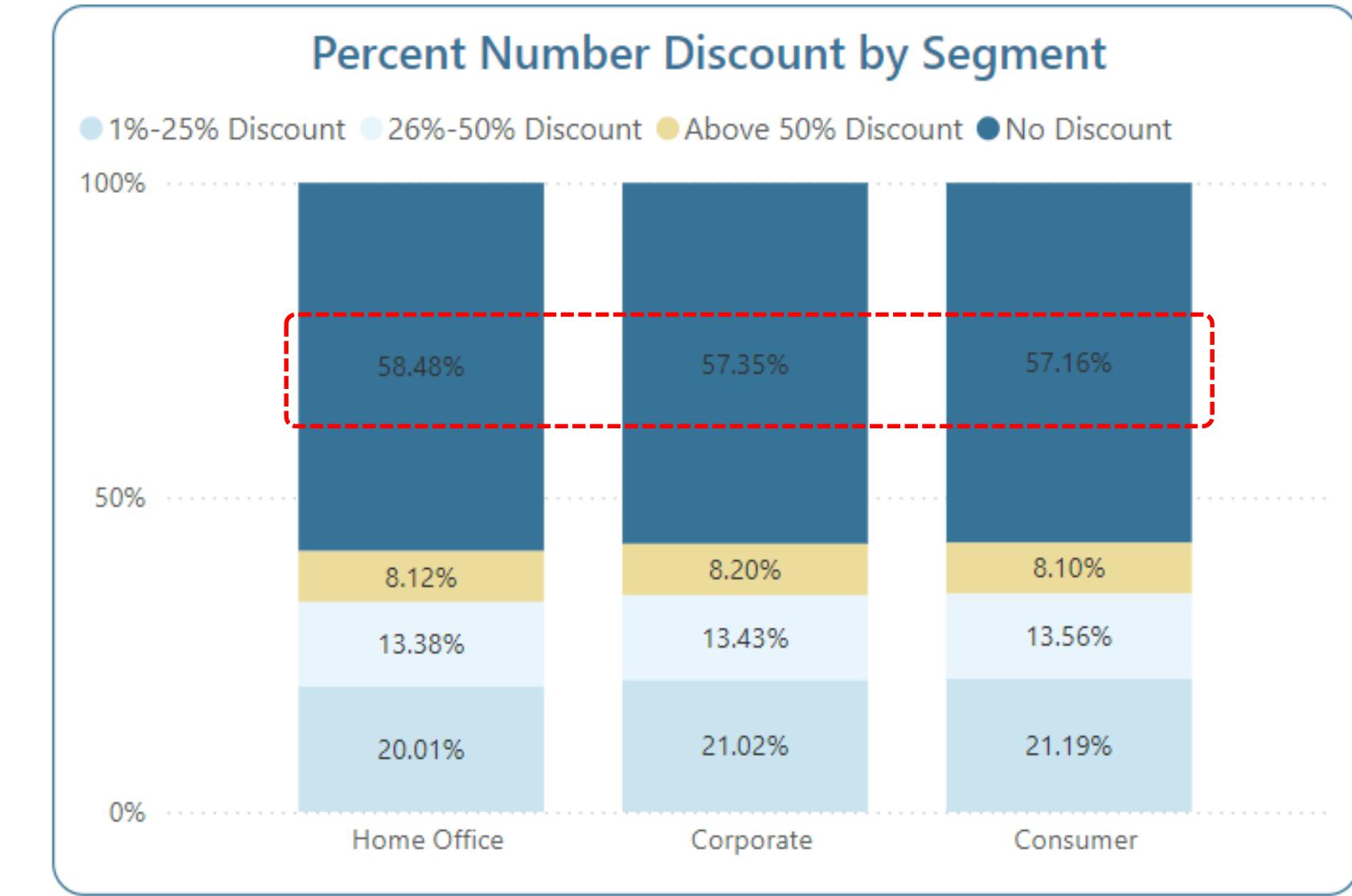
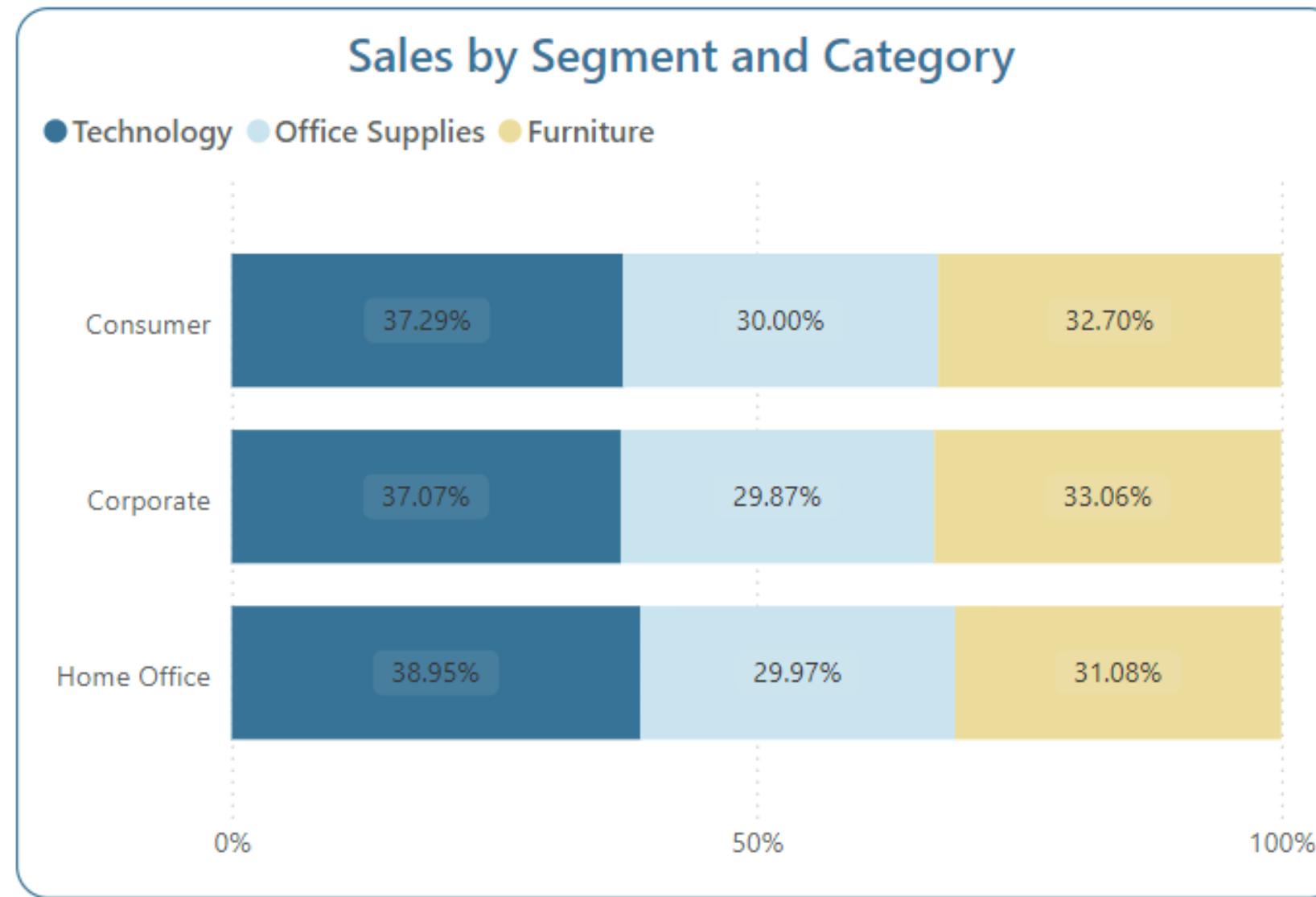
7. Data Analysis

Business Question:

Why is Home Office's Profit Margin higher than the other two segments?



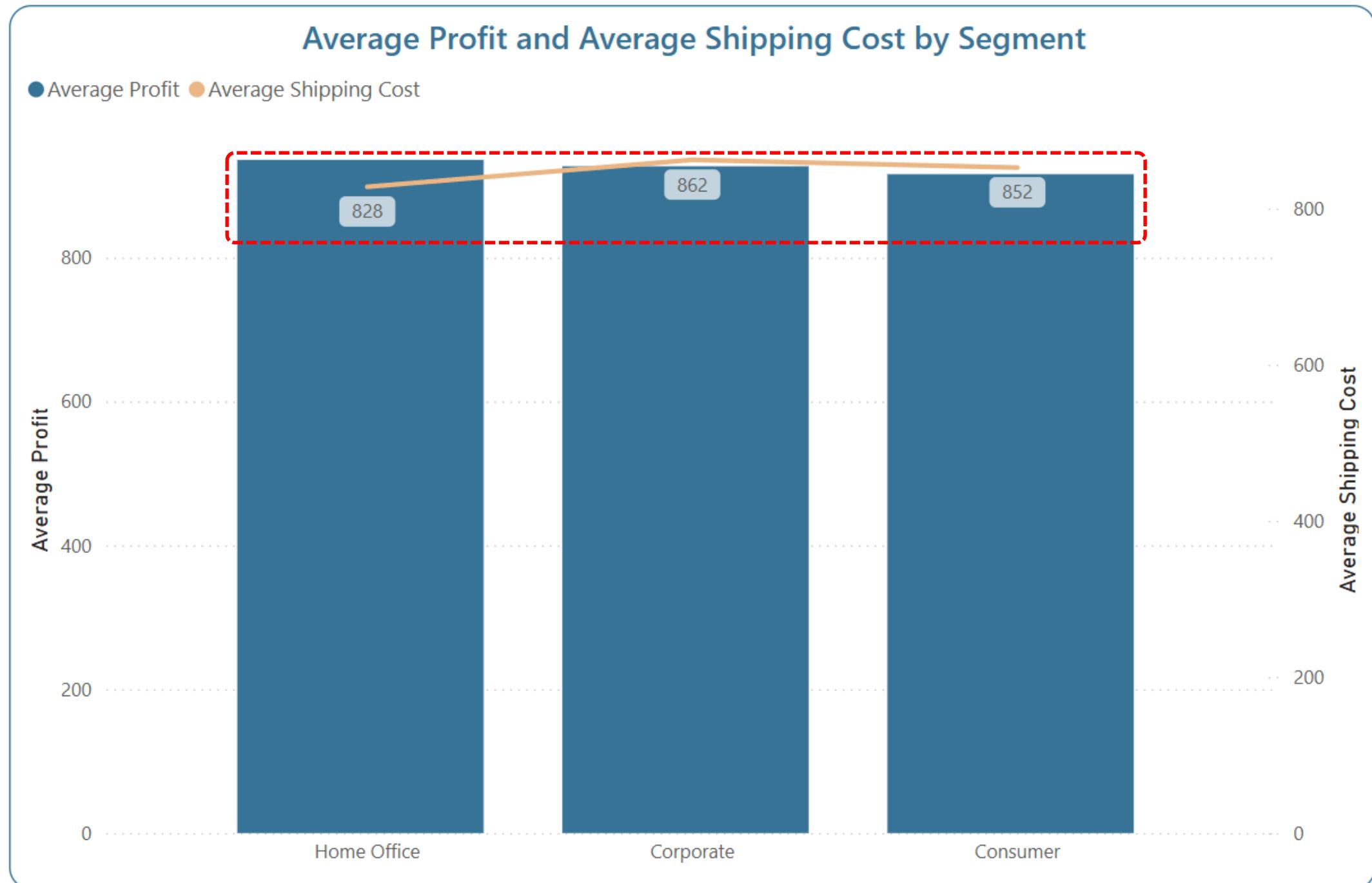
7. Data Analysis



- Similarity between segments: All three segments (Home Office, Corporate, Consumer) have similar purchasing behavior, with revenue ratios from categories like Technology (~37-38%), Office Supplies (~30%), Furniture (~30-32%) remaining almost unchanged.
- The percentage of non-discounted orders (No Discount) for Home Office is the highest (20.01%), which helps retain more profit.
- The discount level among segments does not have a significant difference, but the fact that there is less discount has helped Home Office optimize profits.



7. Data Analysis



Shipping costs affecting profit:

- Home Office has lower shipping costs compared to the other two segments, although the percentage is insignificant.
- This helps to maintain a better profit margin.



7. Data Analysis

Insight

Customer segments have similar purchasing behavior, leading to no significant difference in profit margins. Some factors that could influence this include:

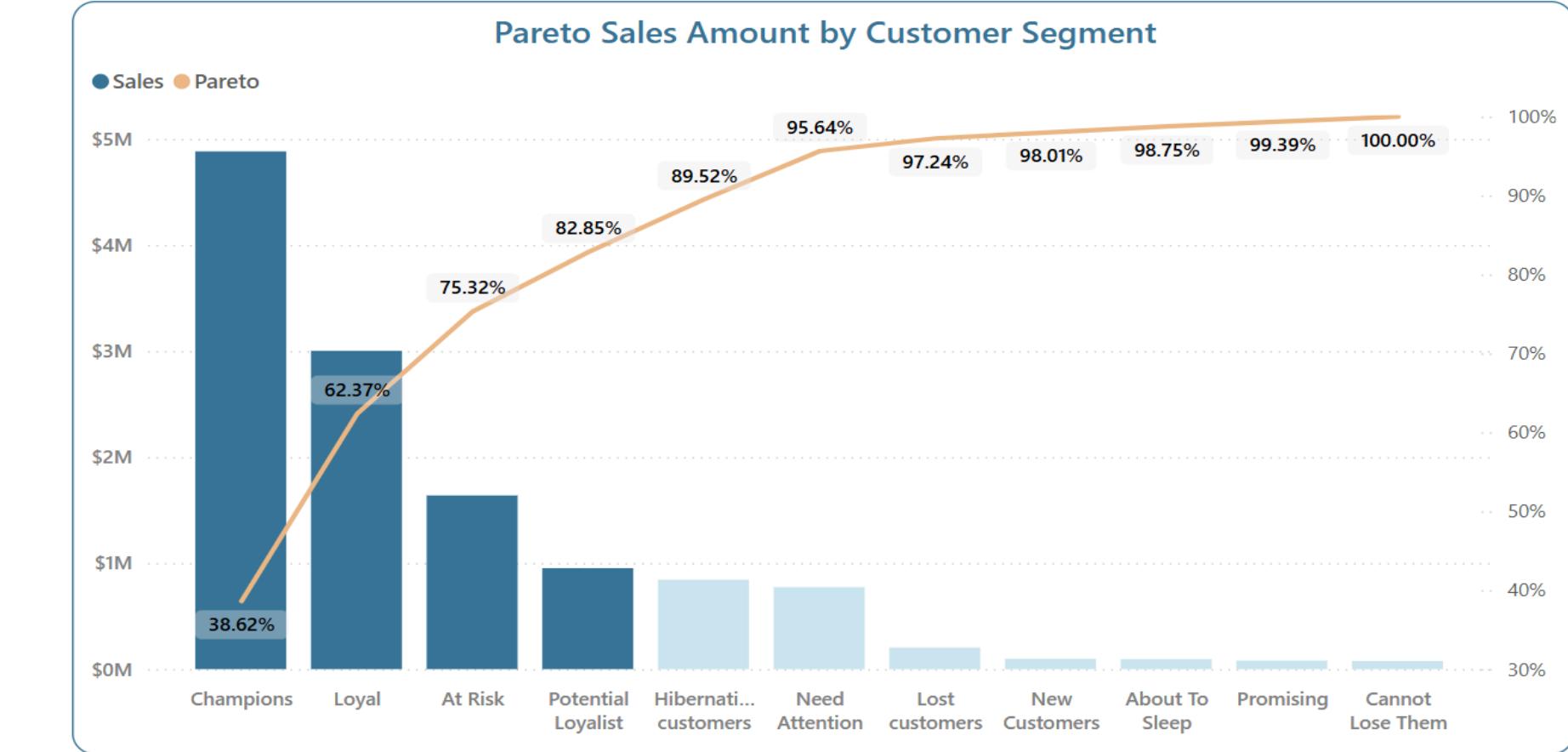
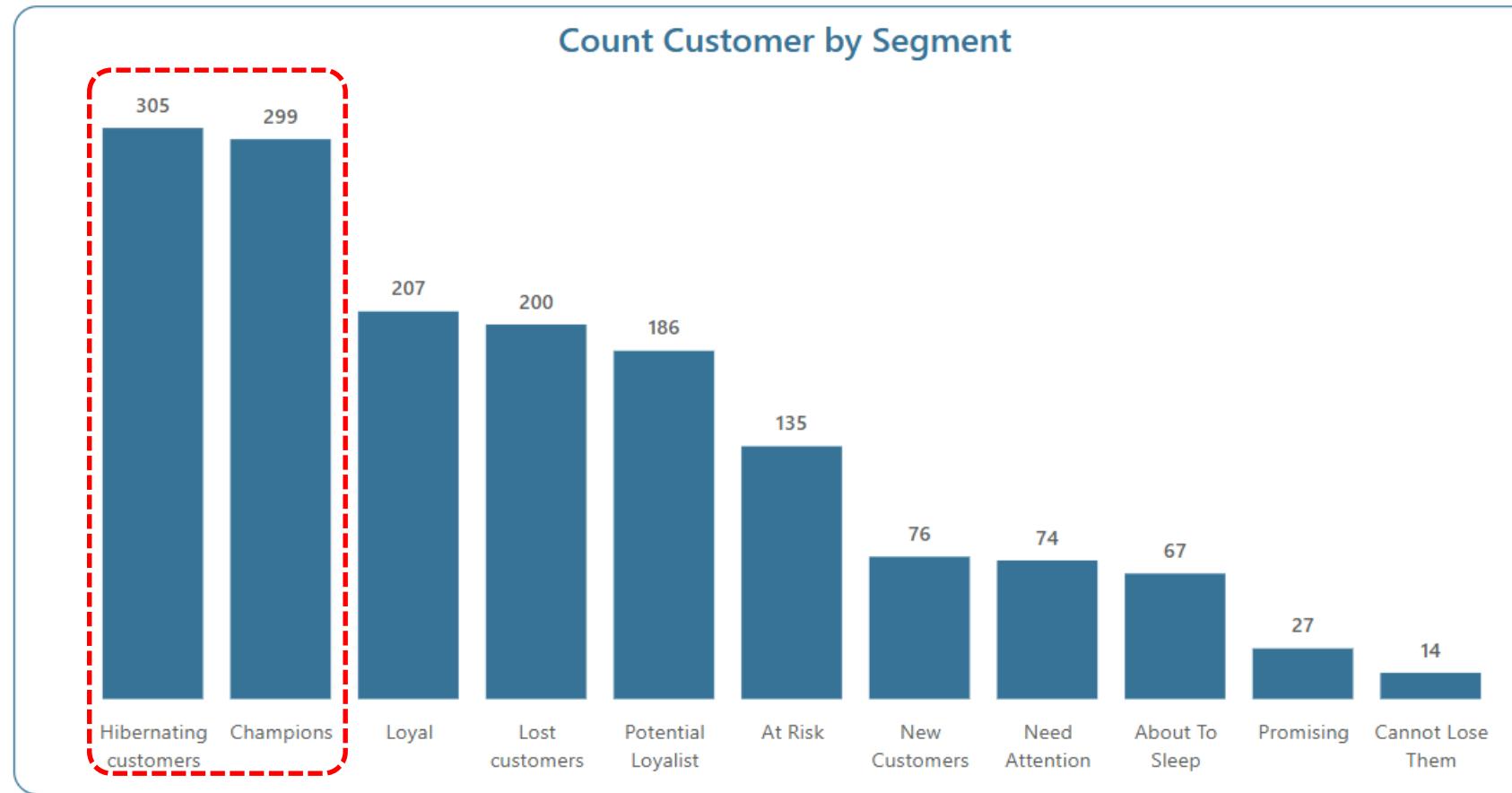
- The Home Office segment receives fewer promotional programs compared to Consumer and Corporate.
- Average shipping costs for Home Office are lower than the other two segments, contributing to profit retention. The similarity in purchasing behavior among segments creates challenges in personalizing business strategies, making it difficult to implement customer care programs, promotions, or growth drivers.

Recommendation: Implement Customer Segmentation based on the purchasing behavior of customer groups in more detail. This will help optimize targeted business strategies, improve decision-making efficiency, and promote sustainable growth towards each customer segment instead of customer groups as it is currently.



7. Data Analysis

7.6. Customer Segmentation and Analysis



Number of customers by segment:

- After implementing Customer Segmentation, the company categorized 1590 customers into 11 different groups.
- Hibernating** (305 customers) and **Champions** (299 customers) are the **most populated** groups, indicating the company has a large proportion of effective customers.
- Lost Customers** (200 customers) are a **matter of concern**, and there needs to be an outreach strategy to recover this group.

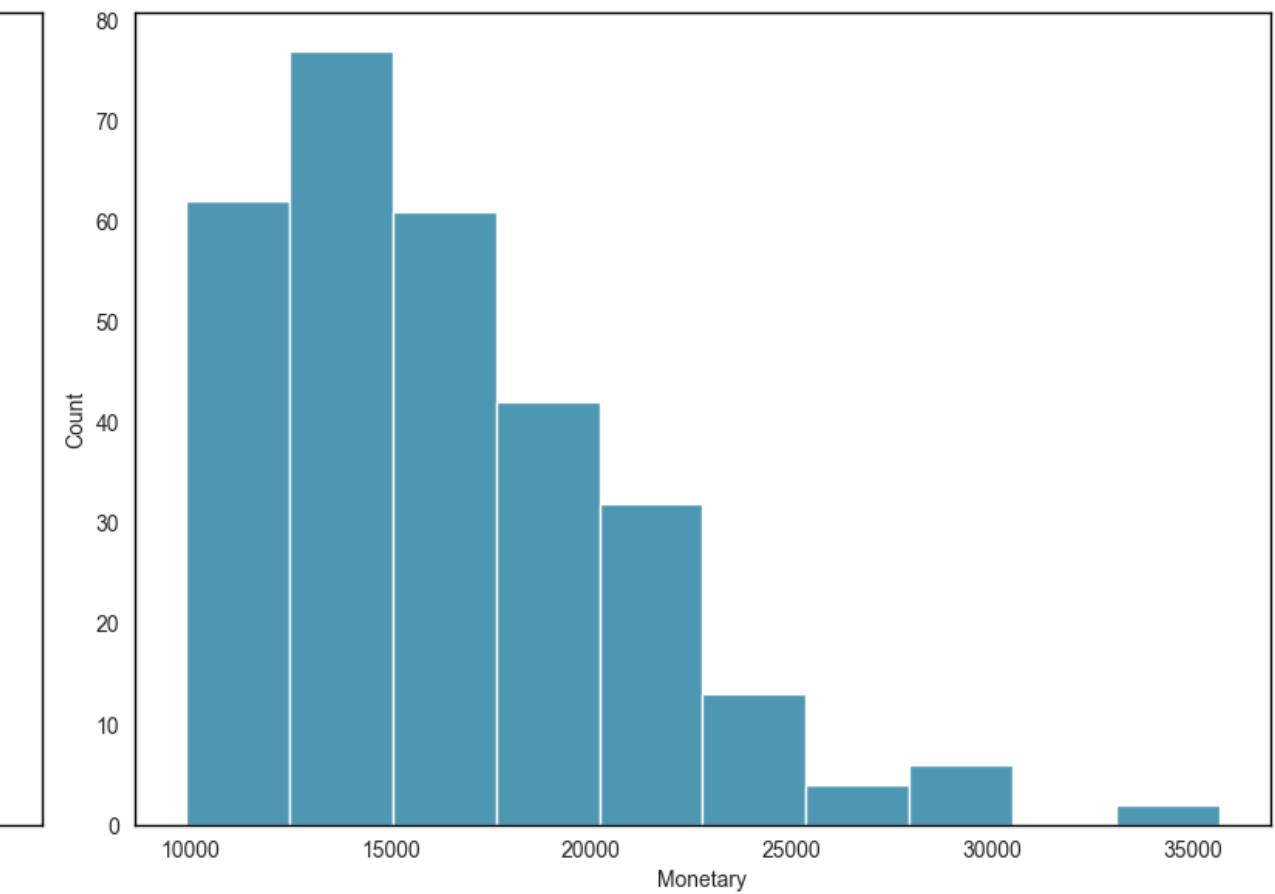
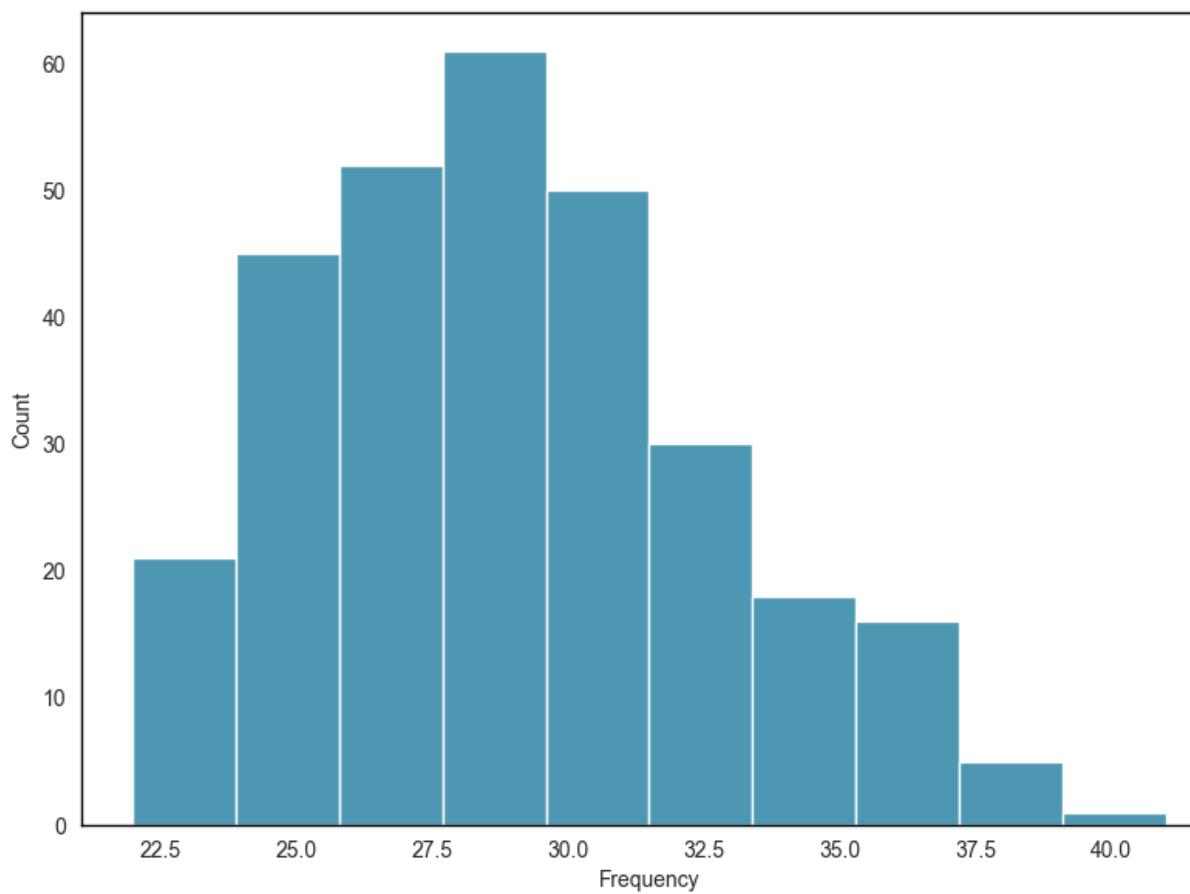
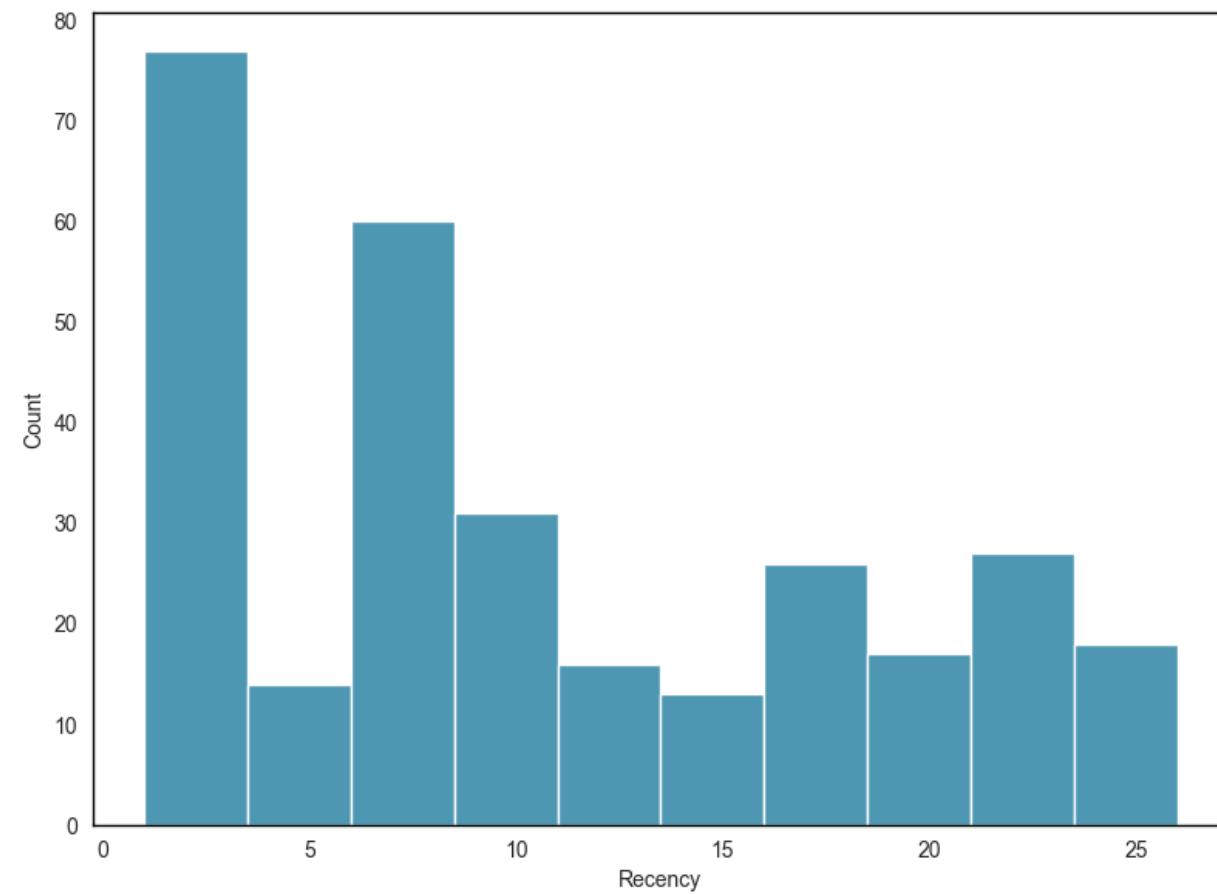
Sales by customer segment (Pareto Chart):

- 80% of revenue comes from 4 main groups:** Champions, Loyal, At Risk, and Potential Loyalist.
- Champions** contribute the most (38.62%), followed by **Loyal** (62.37%), showing that these **are two important customer groups that need focused care**.
- The At Risk and Potential Loyalist groups still have high sales contribution potential, so there should be customer retention and purchase encouragement programs.



7. Data Analysis

Champions



Observation: These are your best customers. They buy recently, frequently, and spend the most.

Analysis:

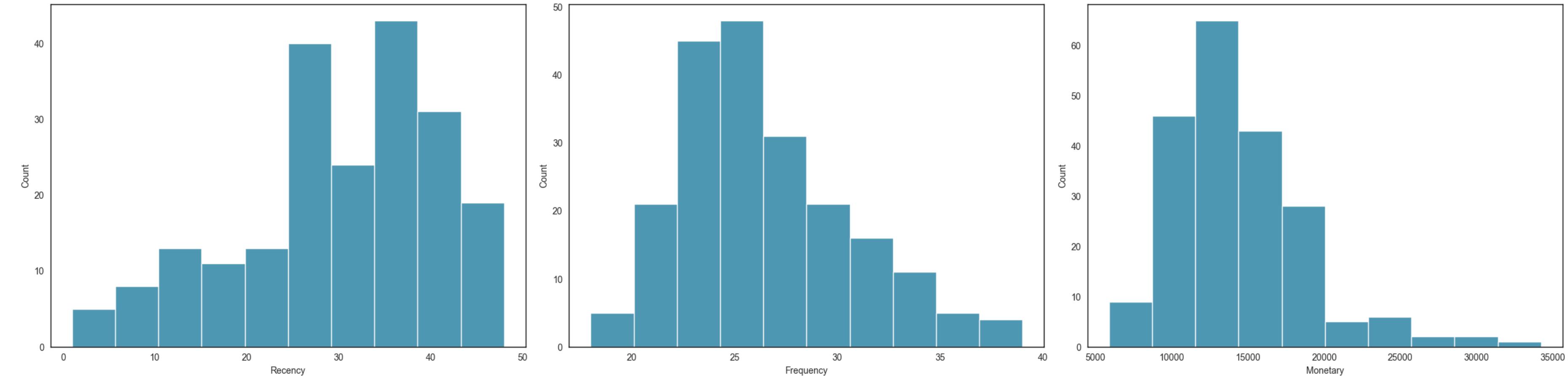
- **Recency:** Bought very recently (Low Recency).
- **Frequency:** Number of times buying frequently (High Frequency).
- **Monetary:** High purchase value (High Monetary).

Actionable Tip: Upsell, offer personalized policies. Also, ask for reviews.



7. Data Analysis

Loyal



Observation: Customers place orders frequently and interact with promotional programs.

Analysis:

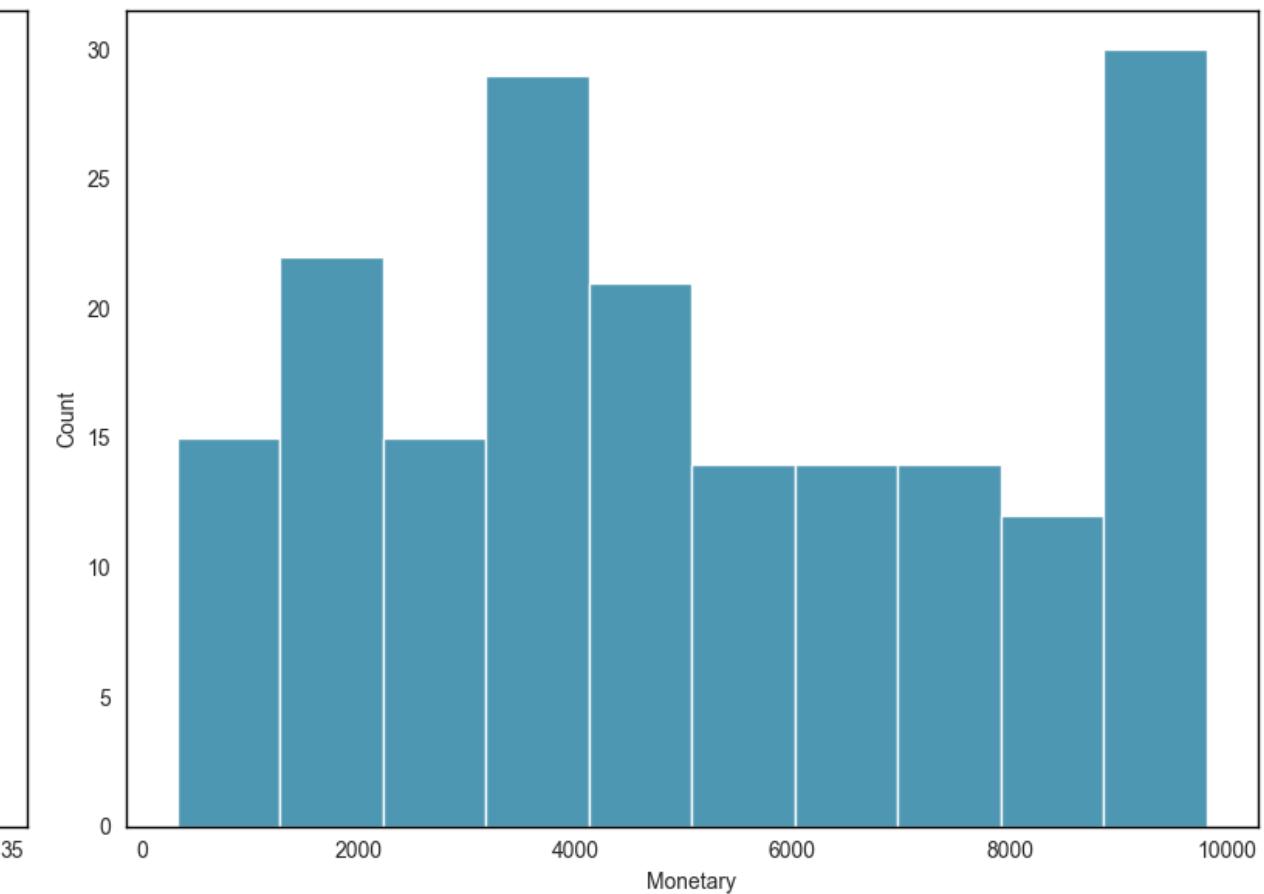
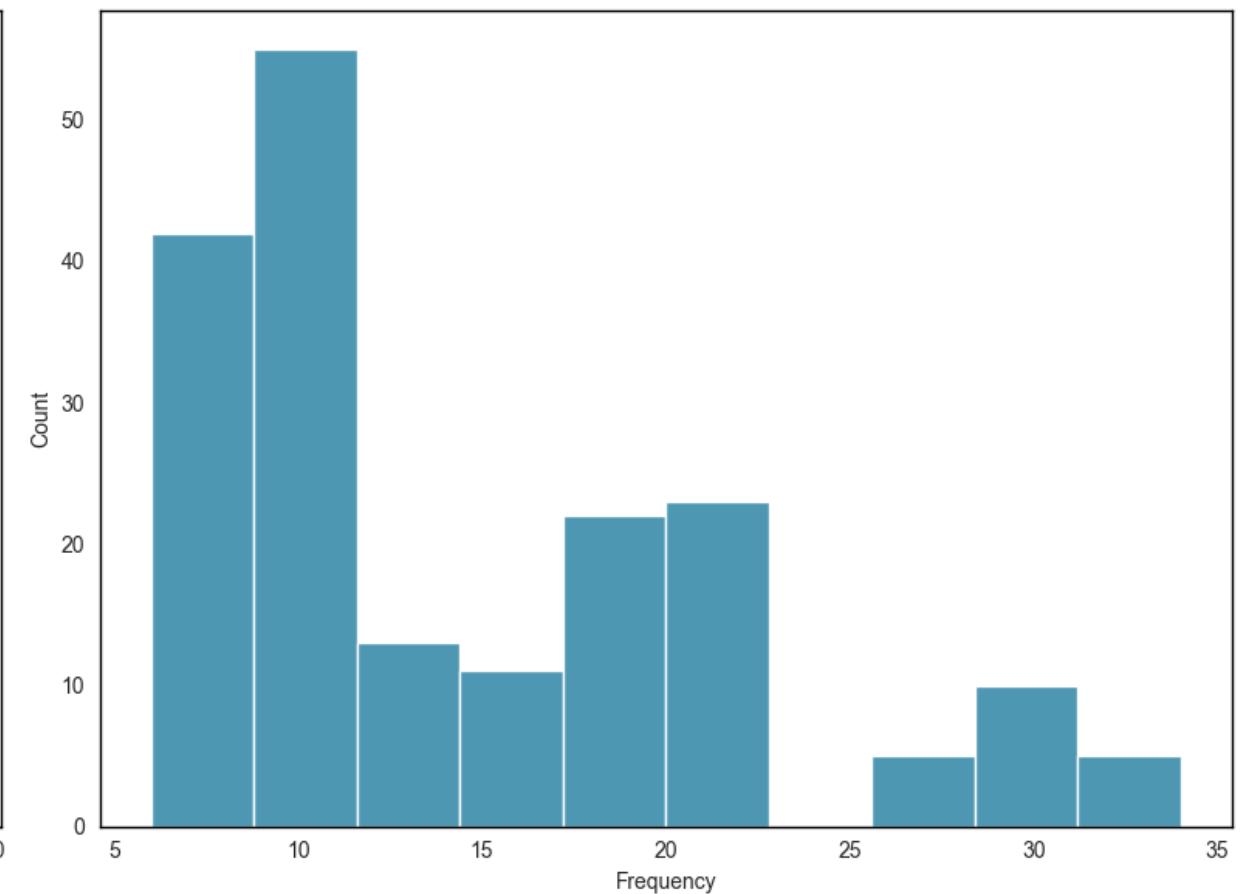
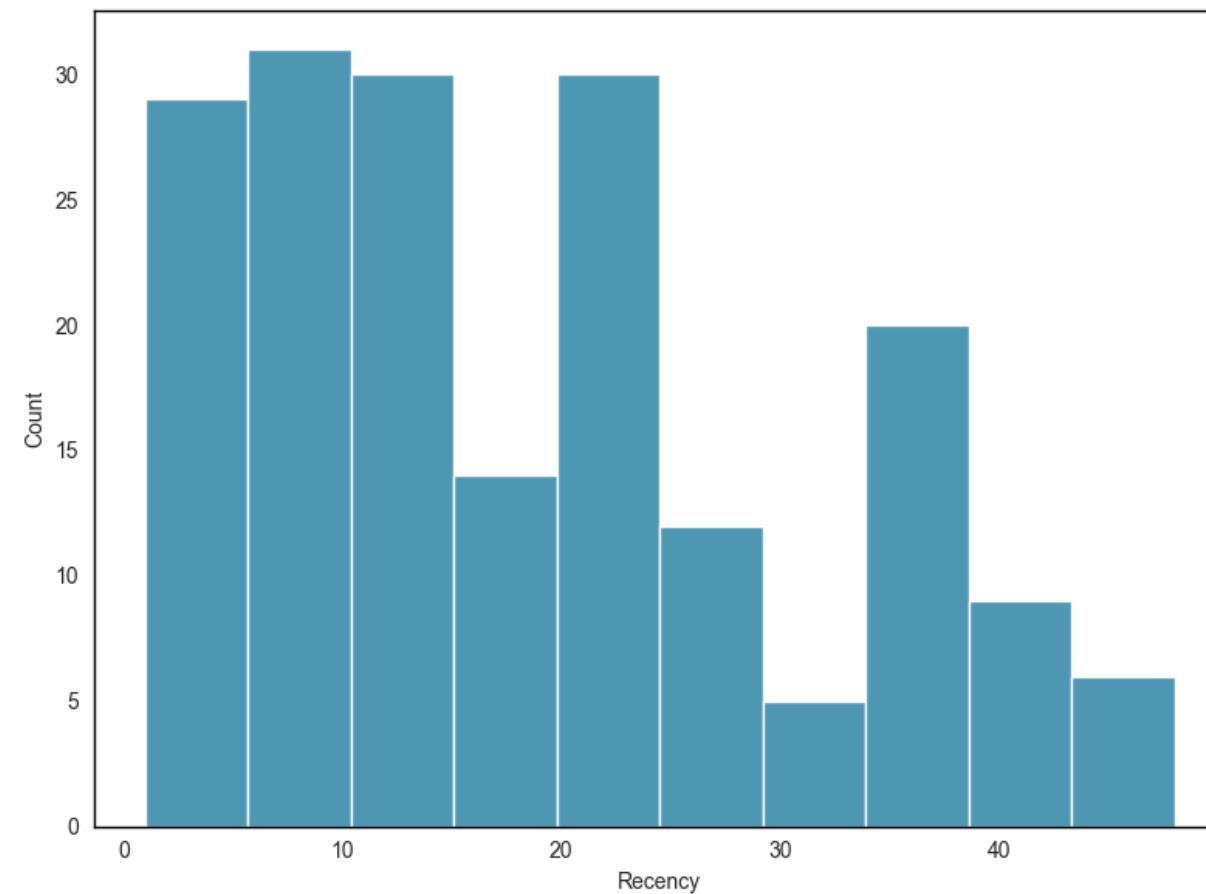
- **Recency:** Not very high
- **Frequency:** Frequent
- **Monetary:** Quite good

Actionable Tip: Reward them. They could be early adopters of new products. They are very likely to become a very good referral channel.



7. Data Analysis

Potential Loyalist



Observation: Bought recently, spends well.

Analysis:

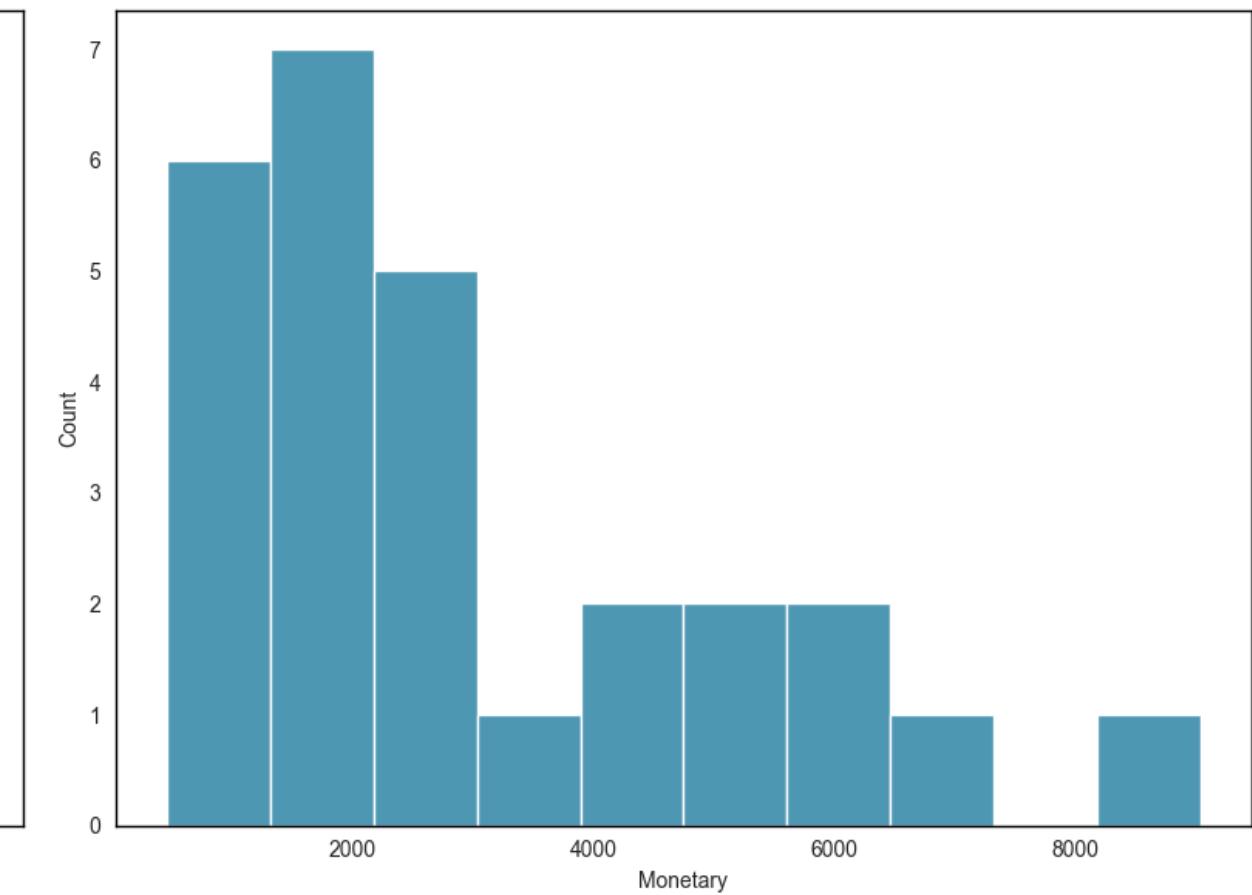
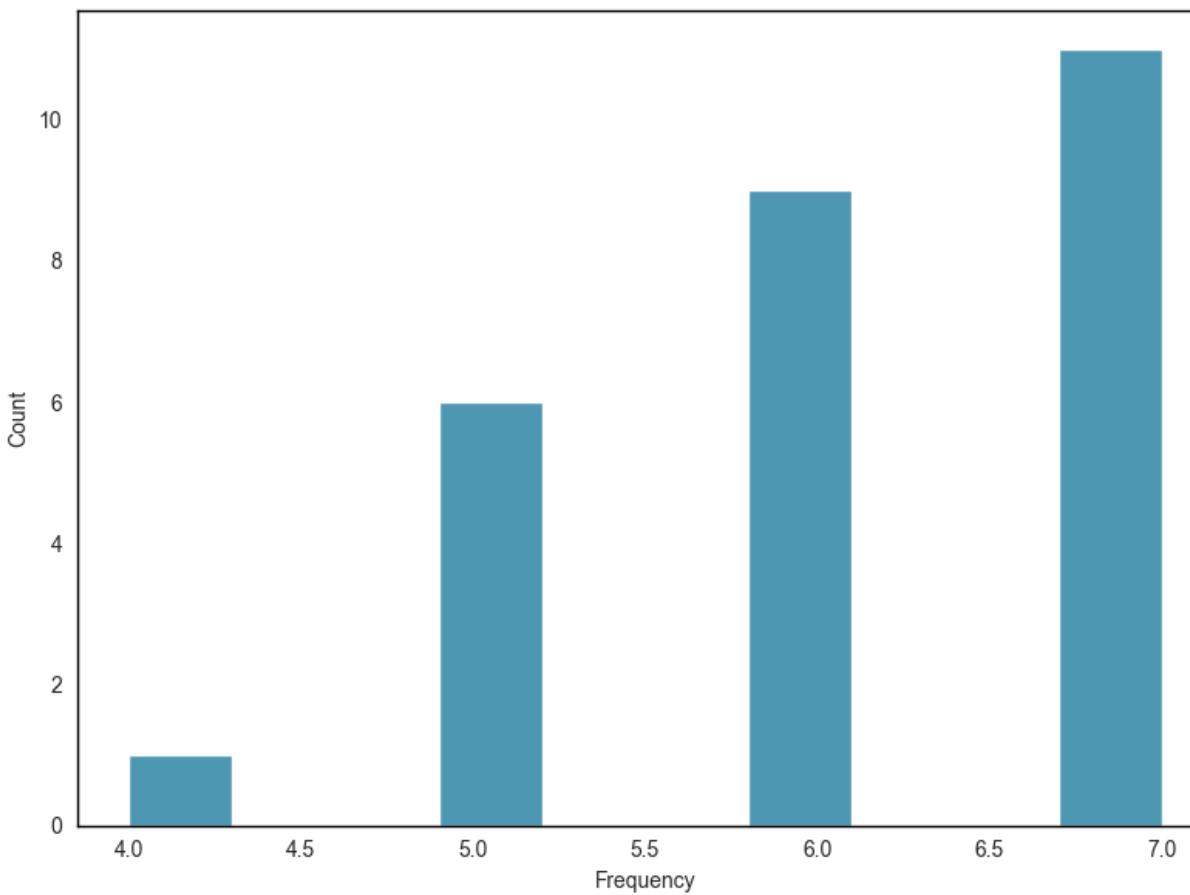
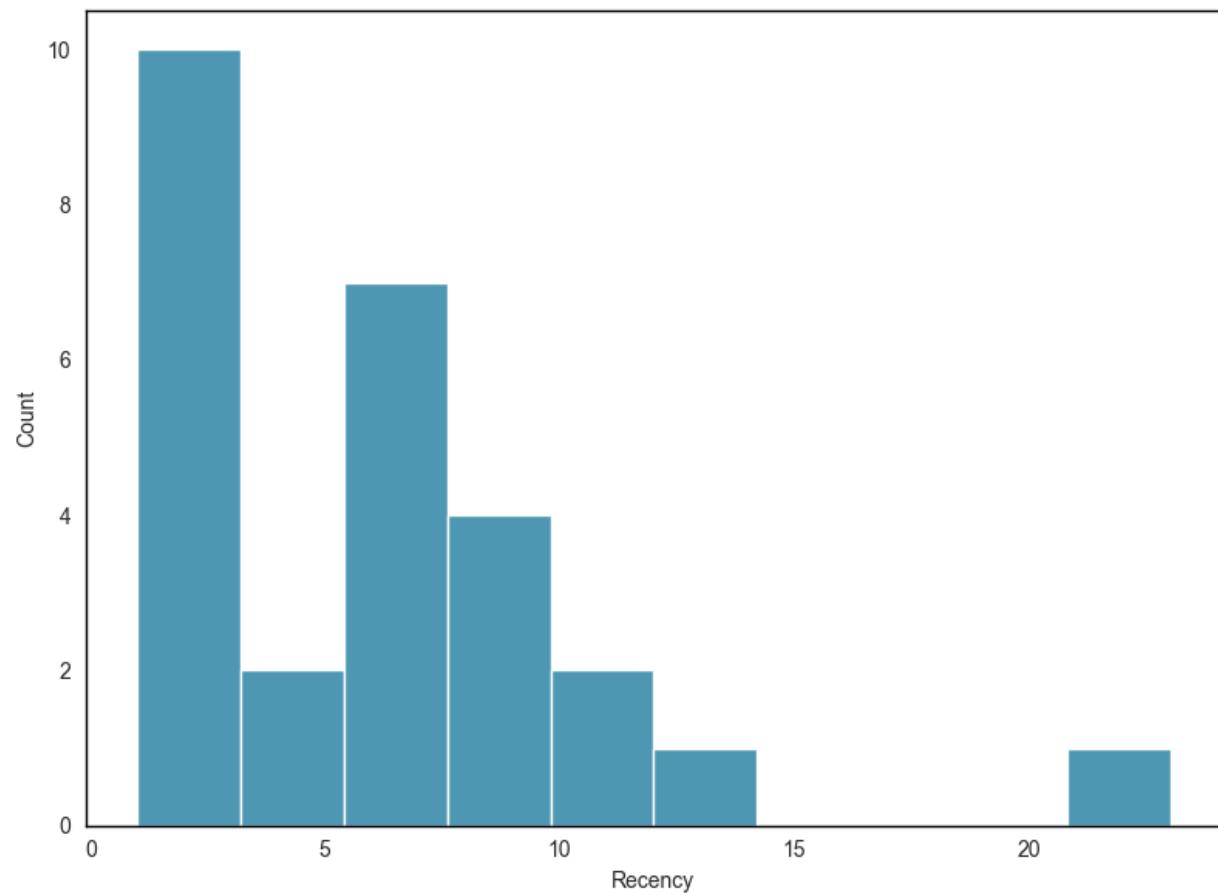
- **Recency:** Bought recently
- **Frequency:** Not frequent
- **Monetary:** Quite good

Actionable Tip: Offer a membership program / loyalty program. Keep them engaged. Offer personalized policies.



7. Data Analysis

Promising



Observation: Overall are loyal customers. They spend frequently and quite a lot of money. However, their last purchase was a few weeks ago.

Analysis:

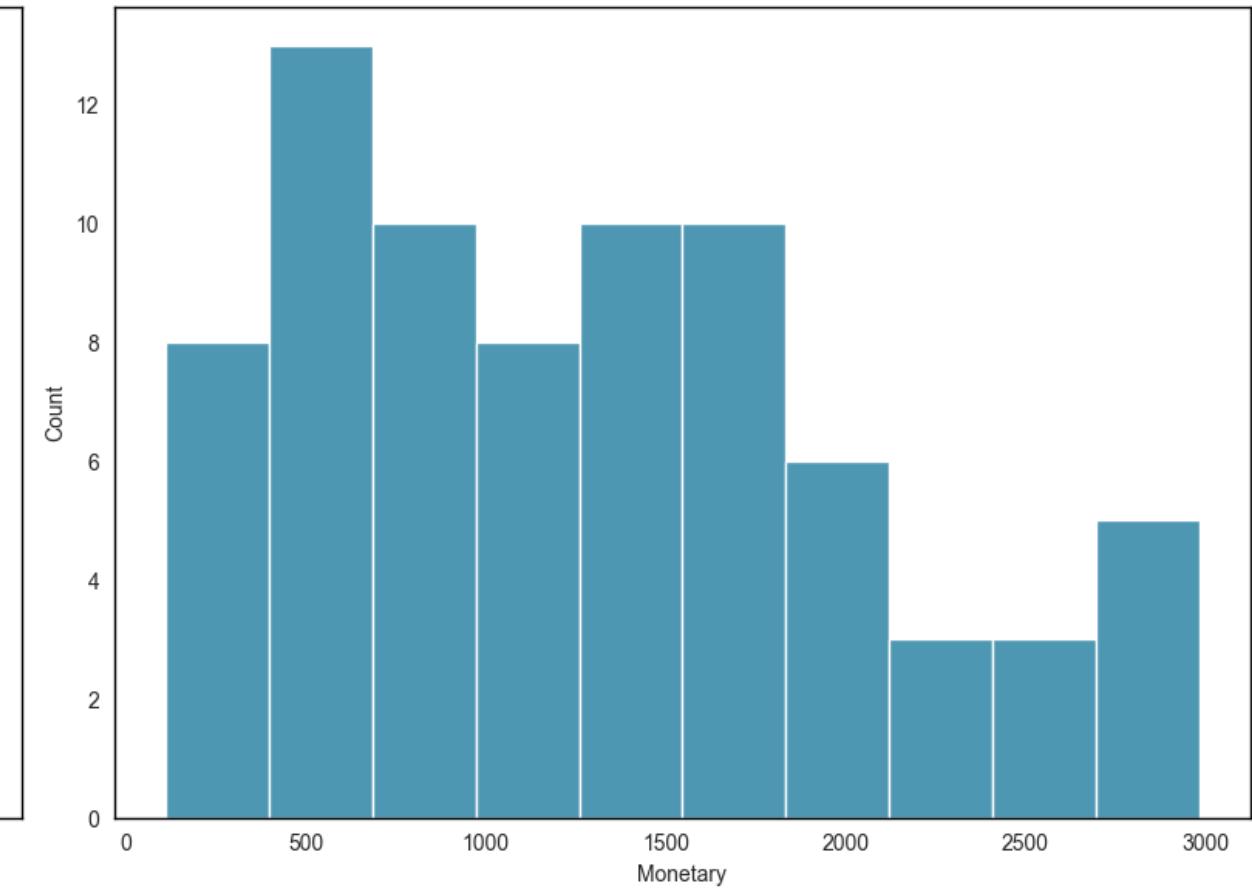
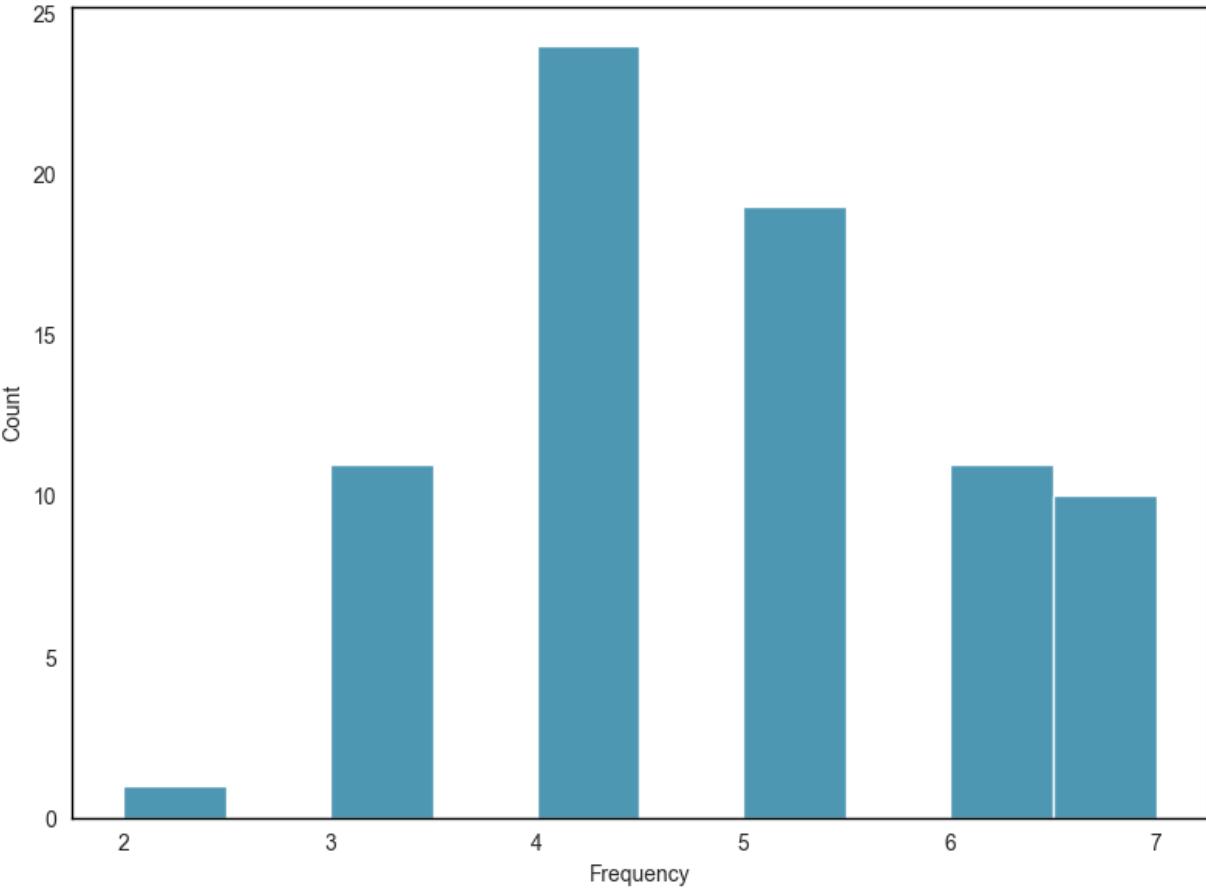
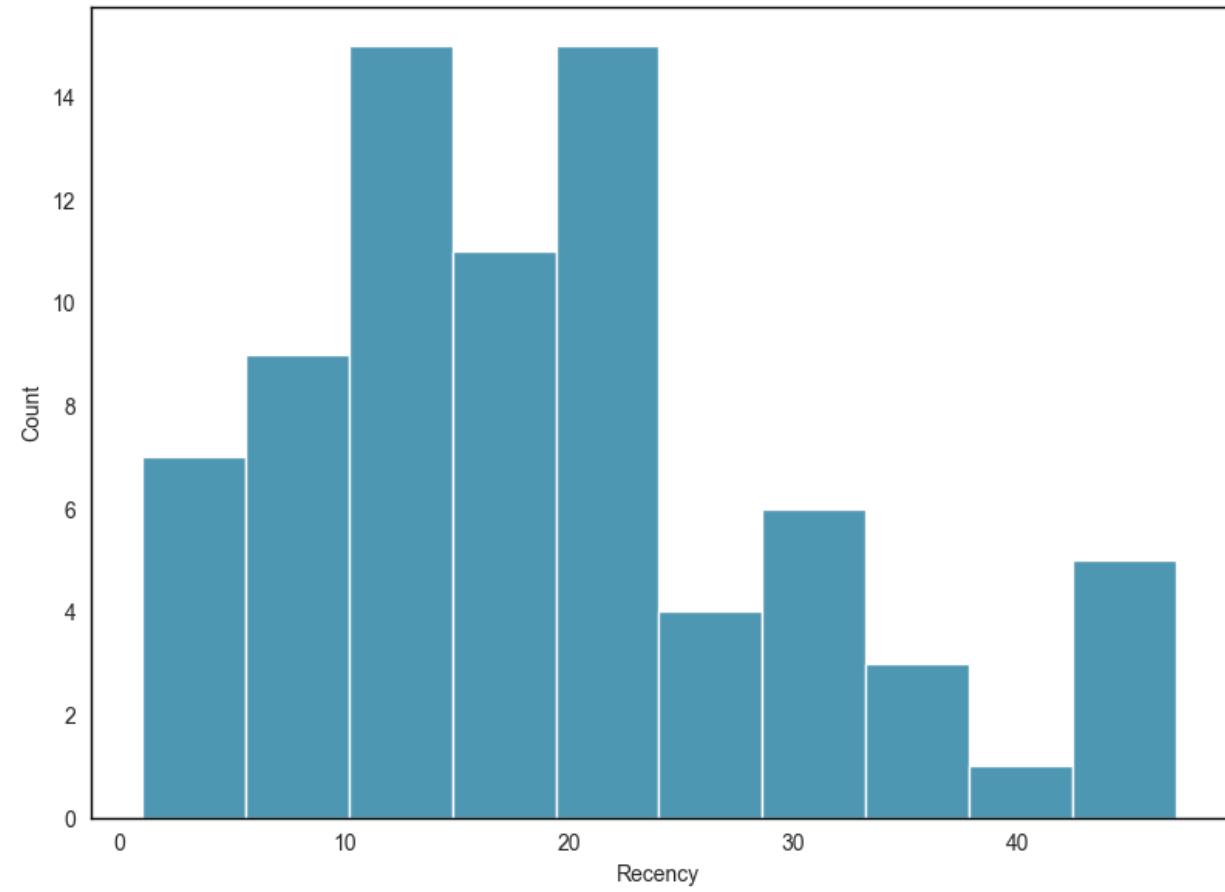
- **Recency:** Not high
- **Frequency:** Average
- **Monetary:** Average

Actionable Tip: Offer discounts. Bring them back to the platform and keep them engaged. Offer personalized recommendations.



7. Data Analysis

New



Observation: New customer

Analysis:

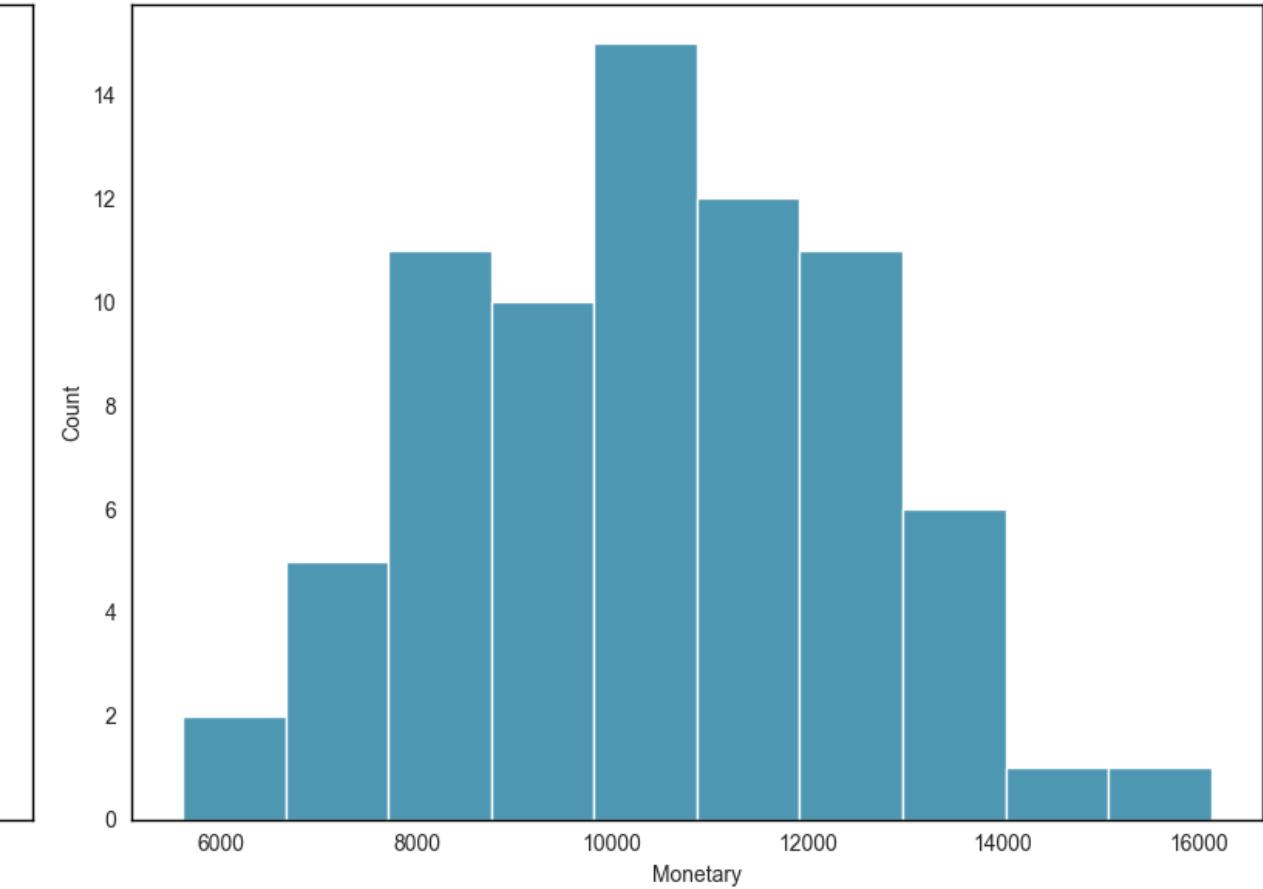
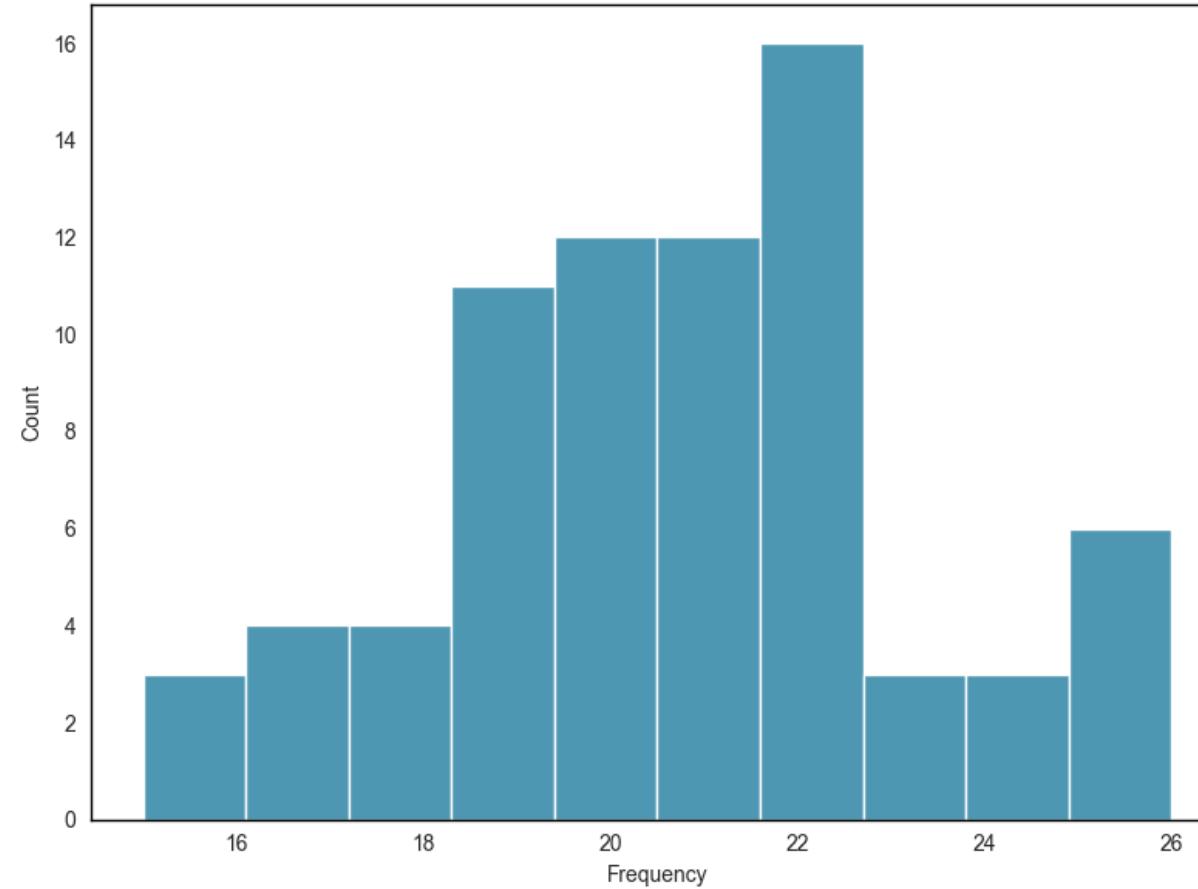
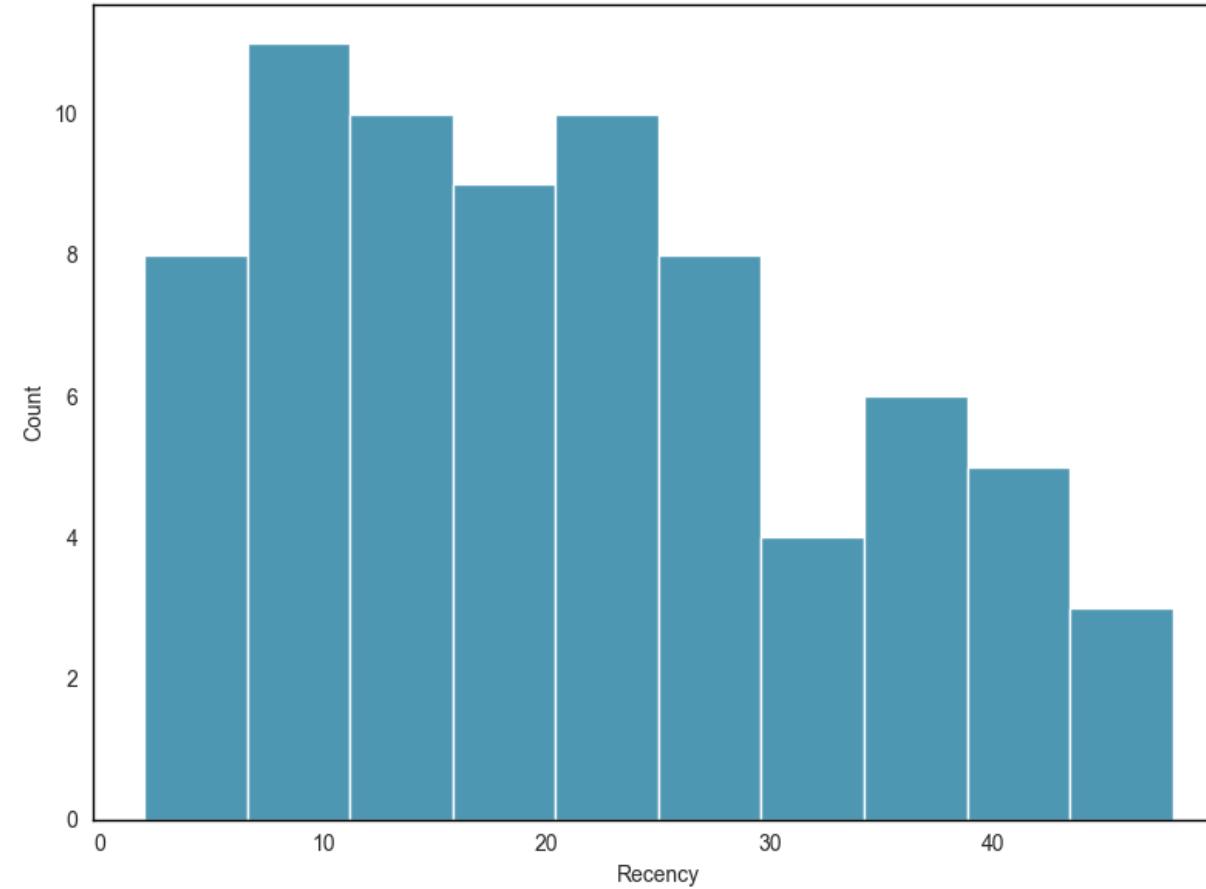
- **Recency:** New
- **Frequency:** Low
- **Monetary:** Low

Actionable Tip: Include in onboarding campaigns, start building a relationship.



7. Data Analysis

Need Attention



Observation: Core loyal customers whose last purchase was more than a month ago.

Analysis:

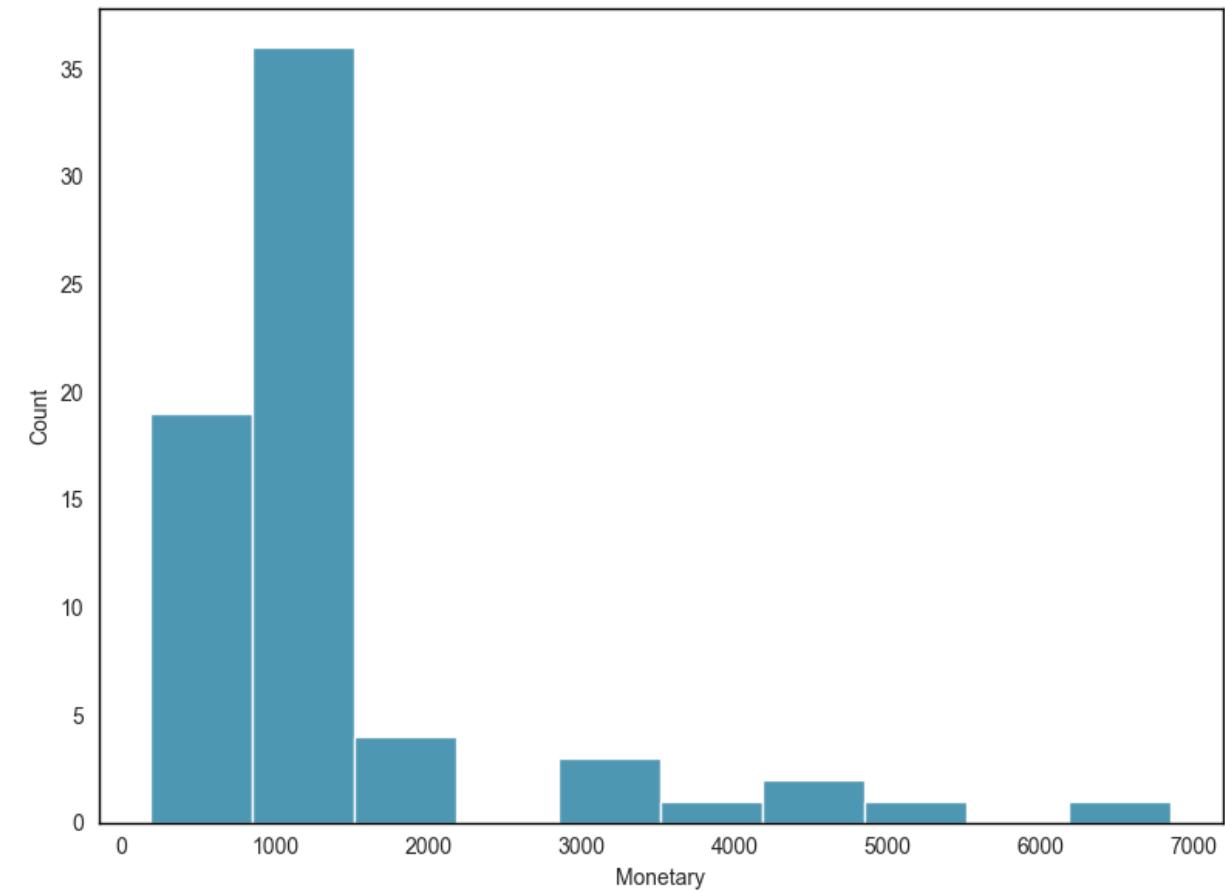
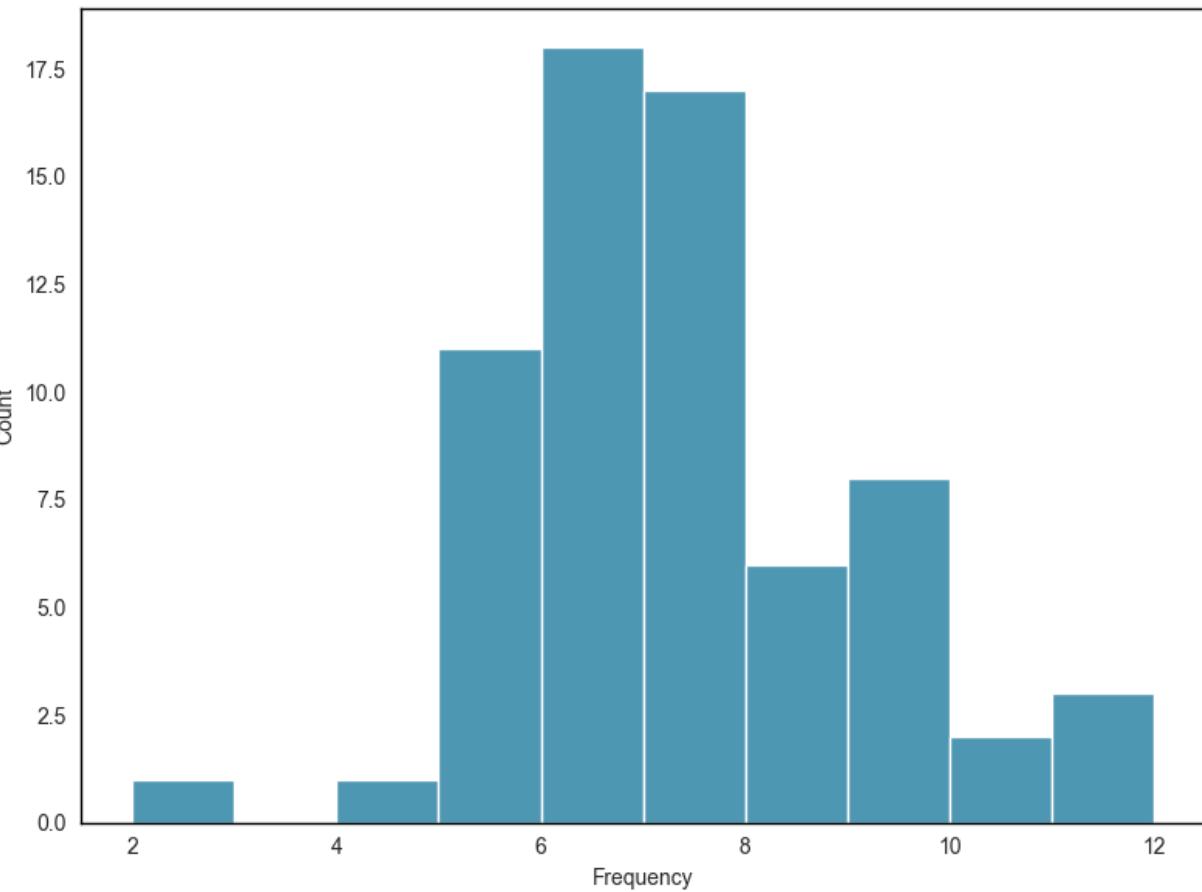
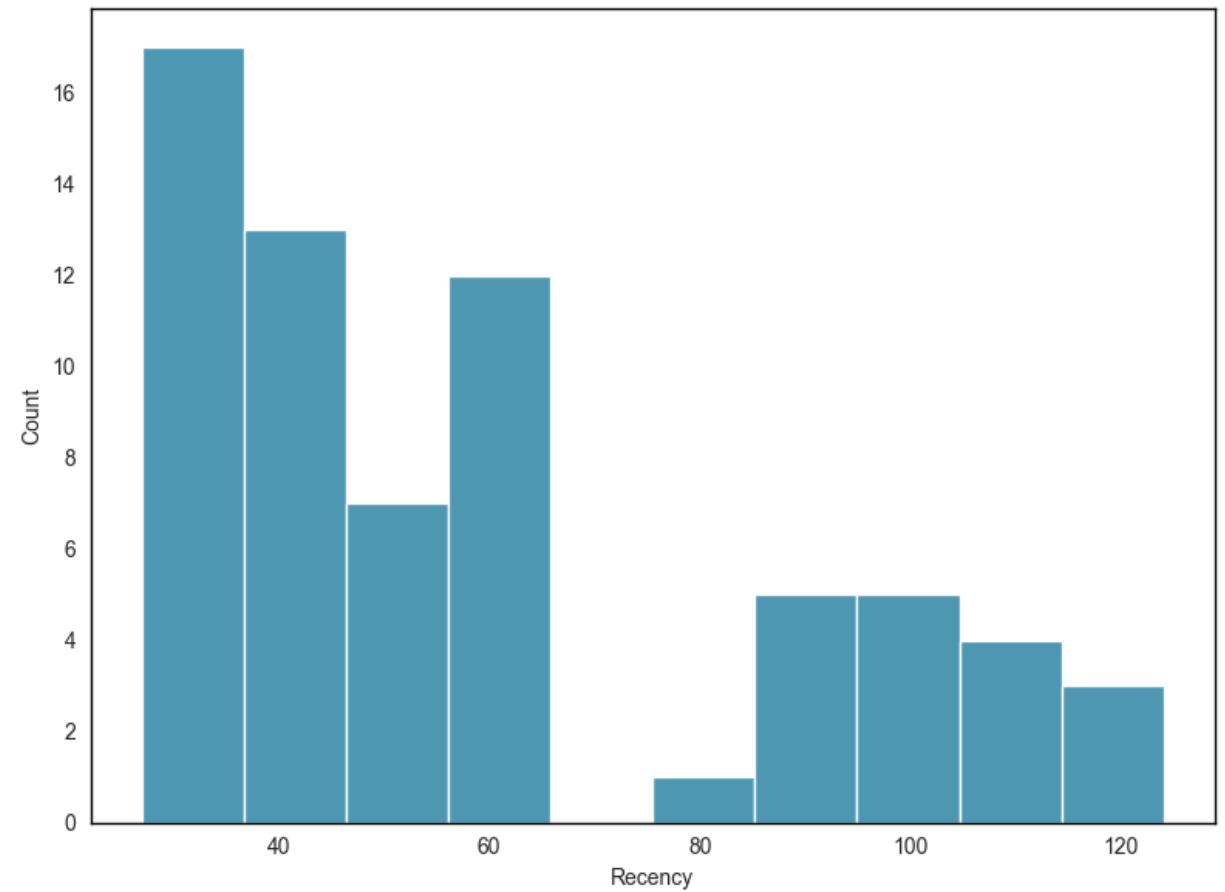
- **Recency:** Average
- **Frequency:** Average
- **Monetary:** Average

Actionable Tip: Offer time-limited promotions. Offer personalized policies.



7. Data Analysis

About To Sleep



Observation: Popular customers, bought recently.

Analysis:

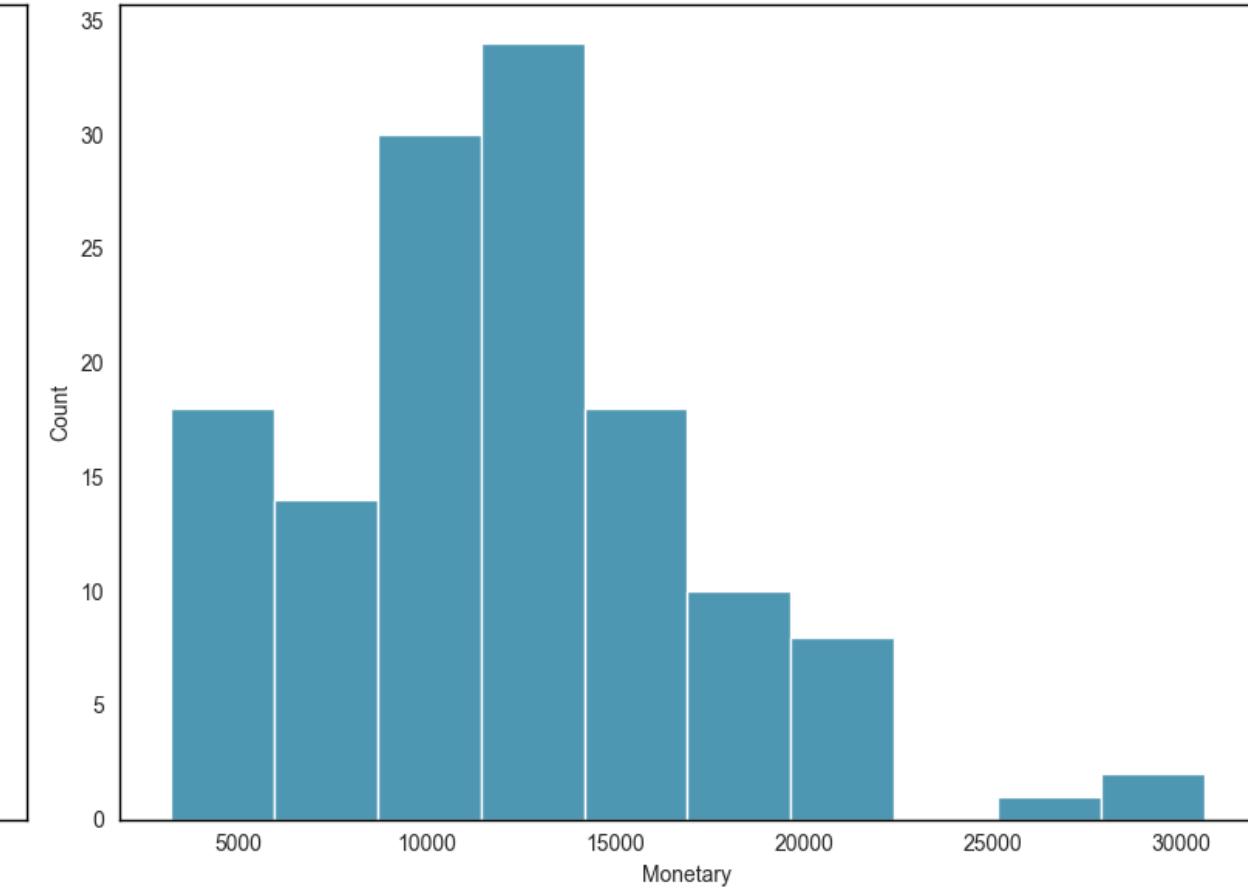
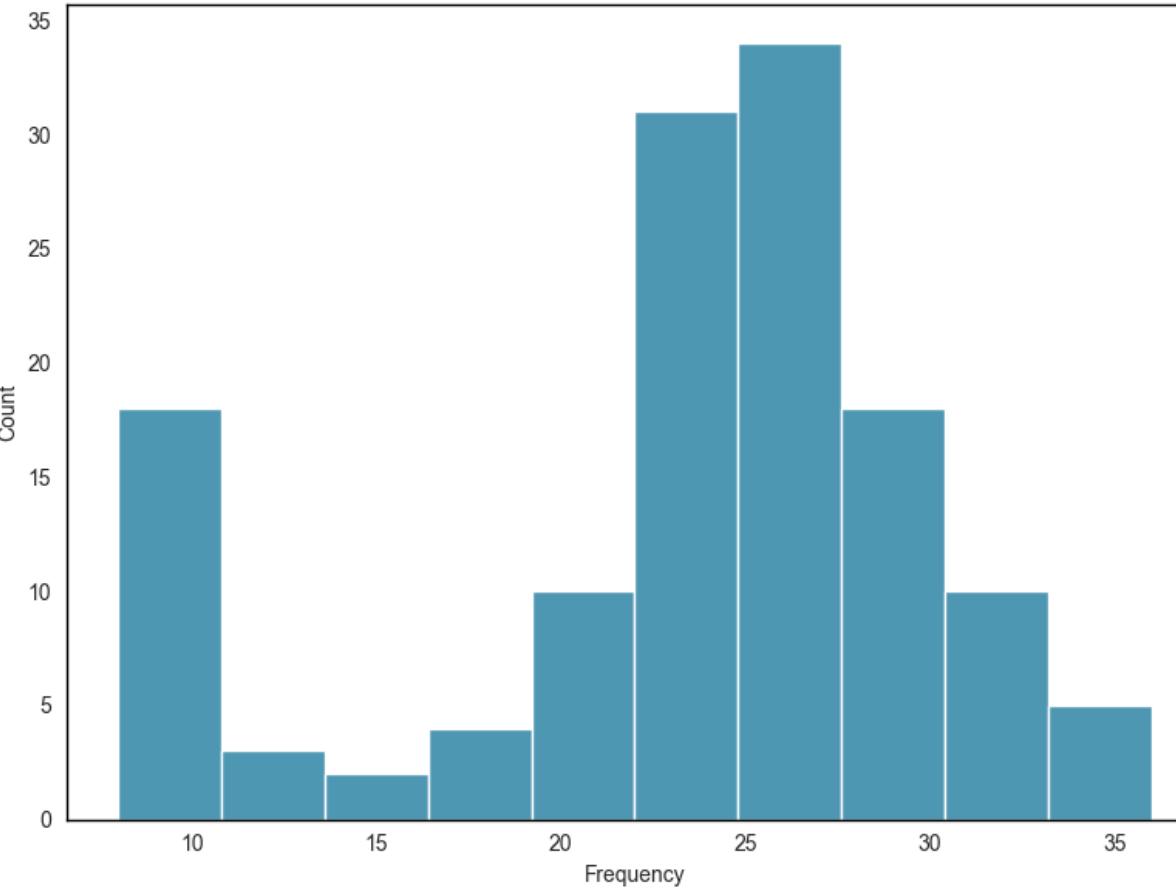
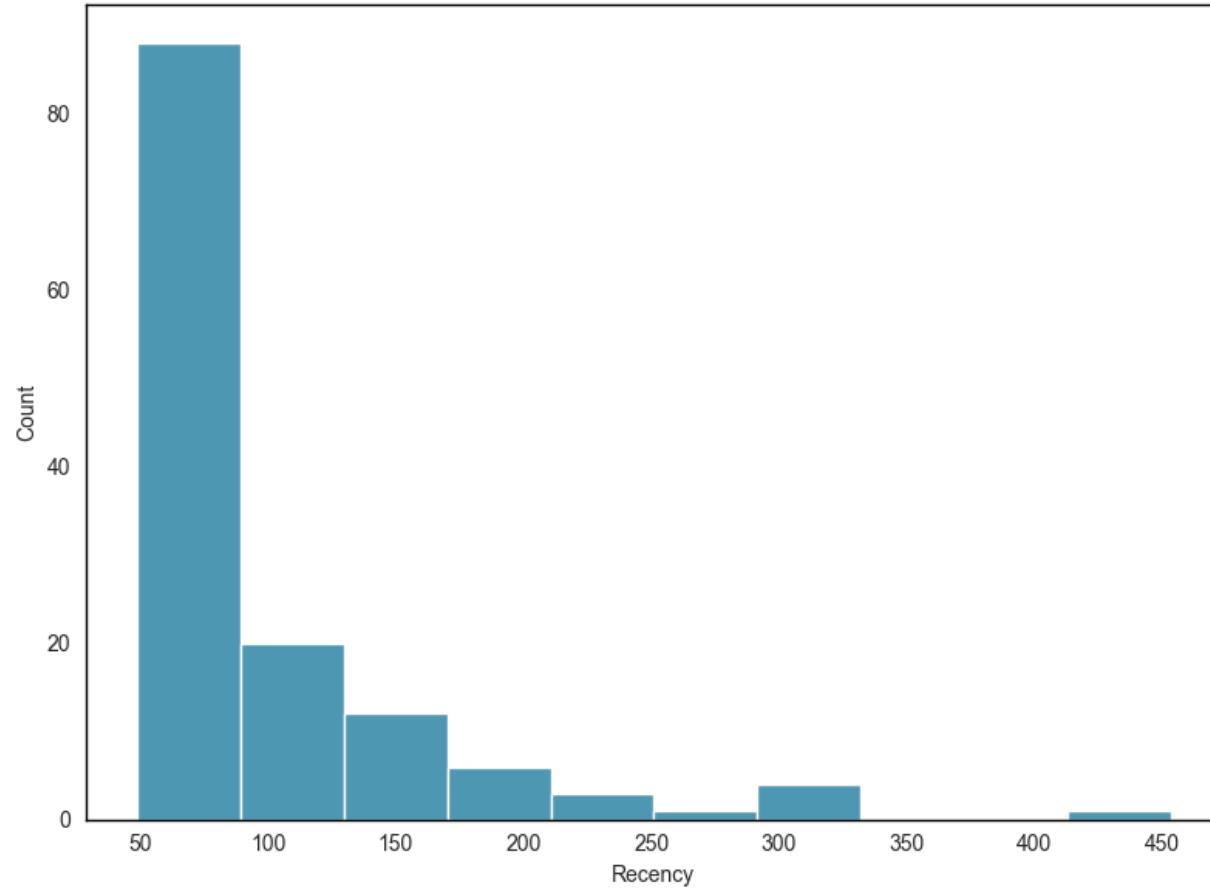
- **Recency:** Longer ago
- **Frequency:** Infrequent
- **Monetary:** Low

Actionable Tip: Offer time-limited promotions.



7. Data Analysis

At Risk



Observation: Similar to "Can't lose them but losing" but lower average order value and frequency.

Analysis:

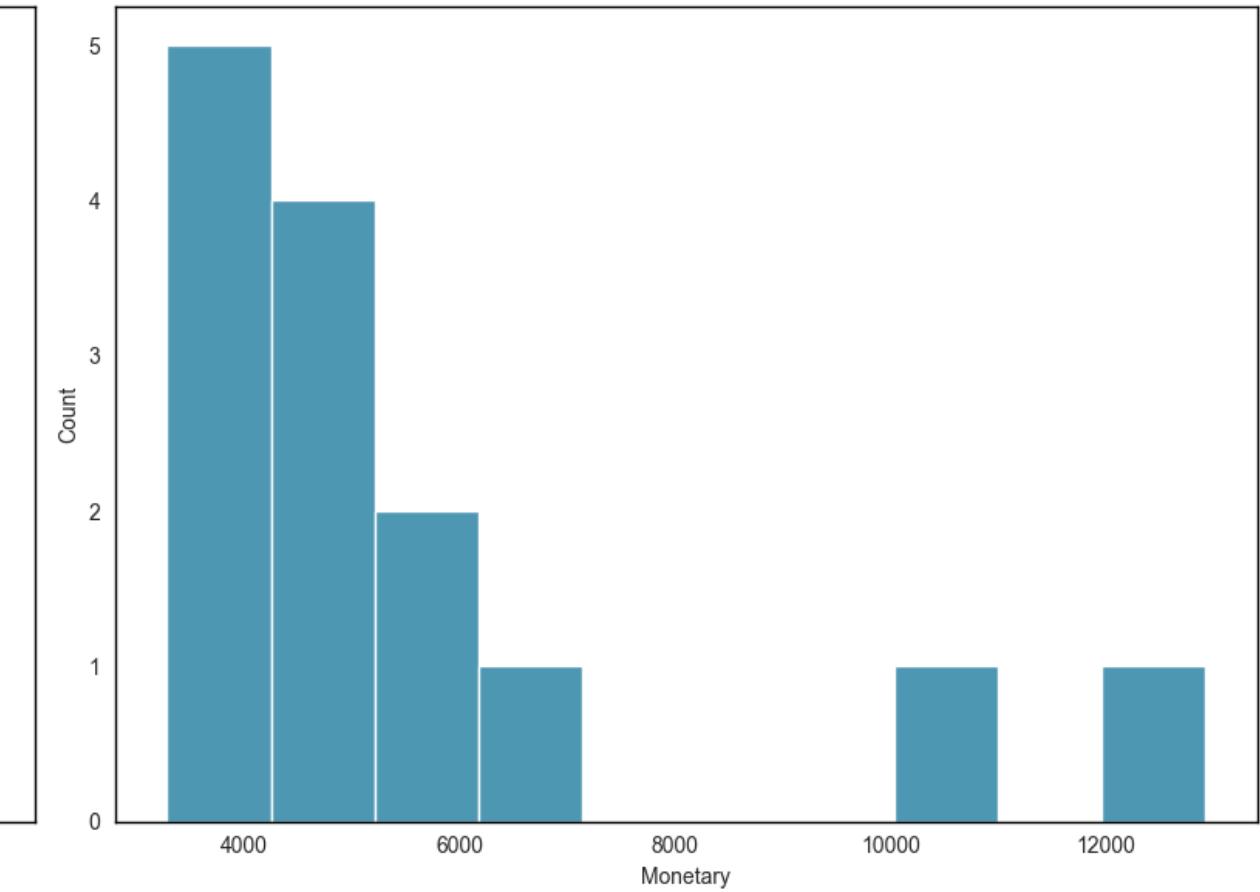
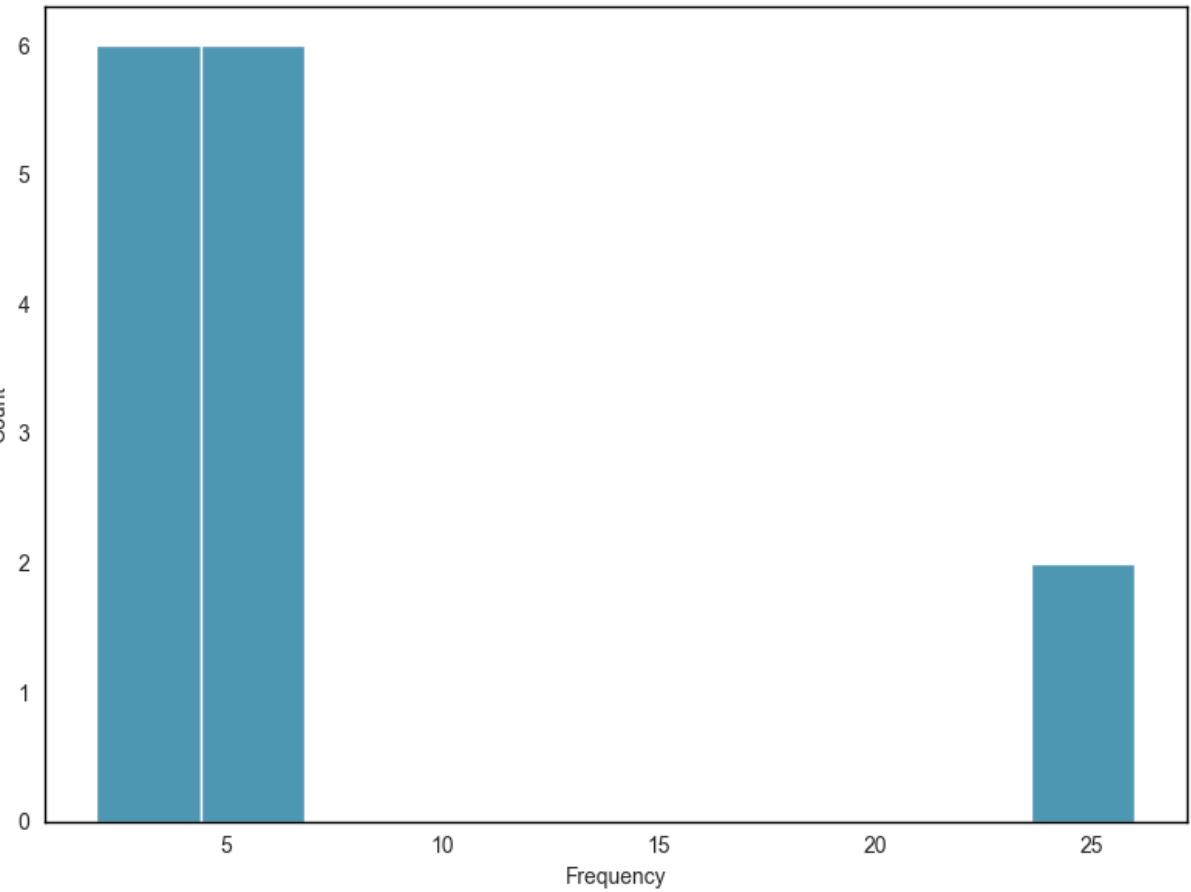
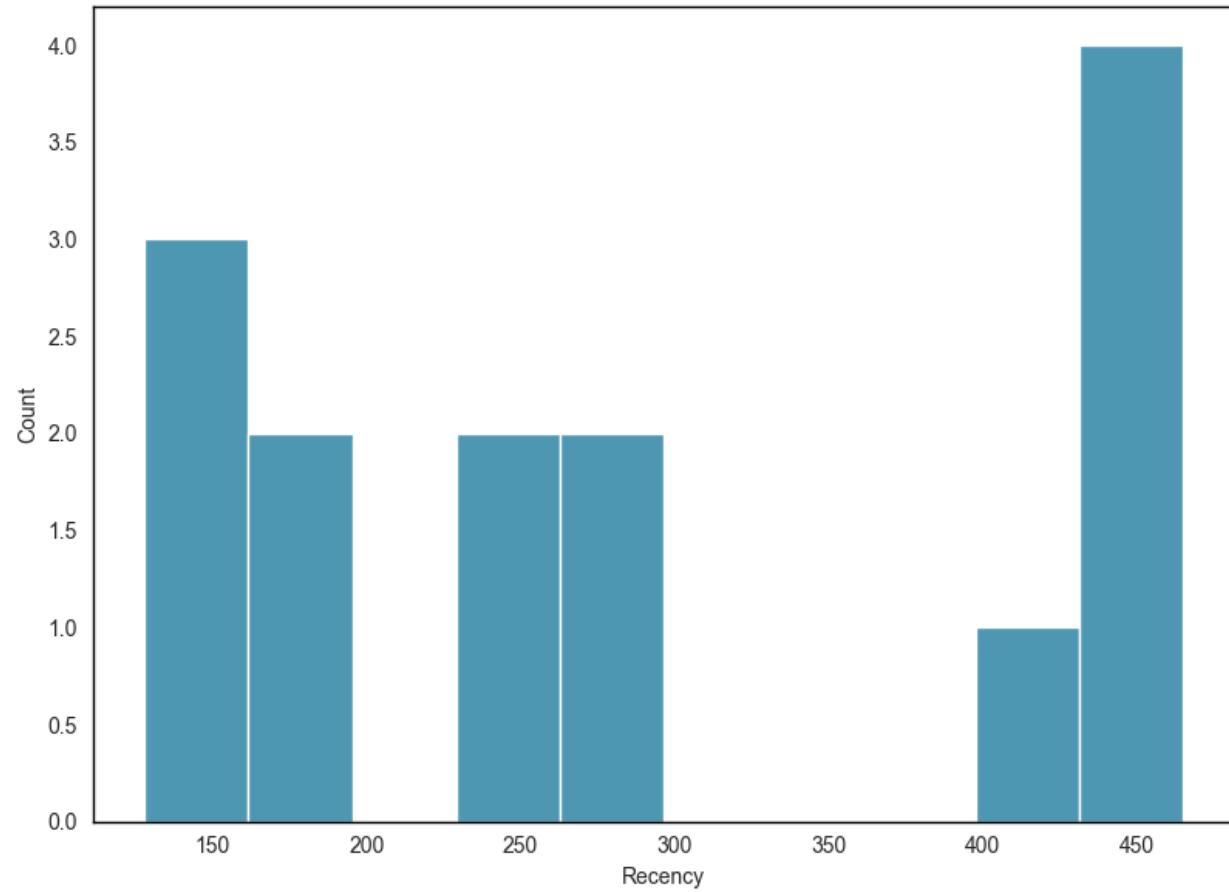
- **Recency:** Quite long ago
- **Frequency:** Low
- **Monetary:** Quite good

Actionable Tip: Send personalized emails. Offer promotional policies.



7. Data Analysis

Can't Lose Them



Observation: Make the largest and most frequent orders. But haven't returned for a long time.

Analysis:

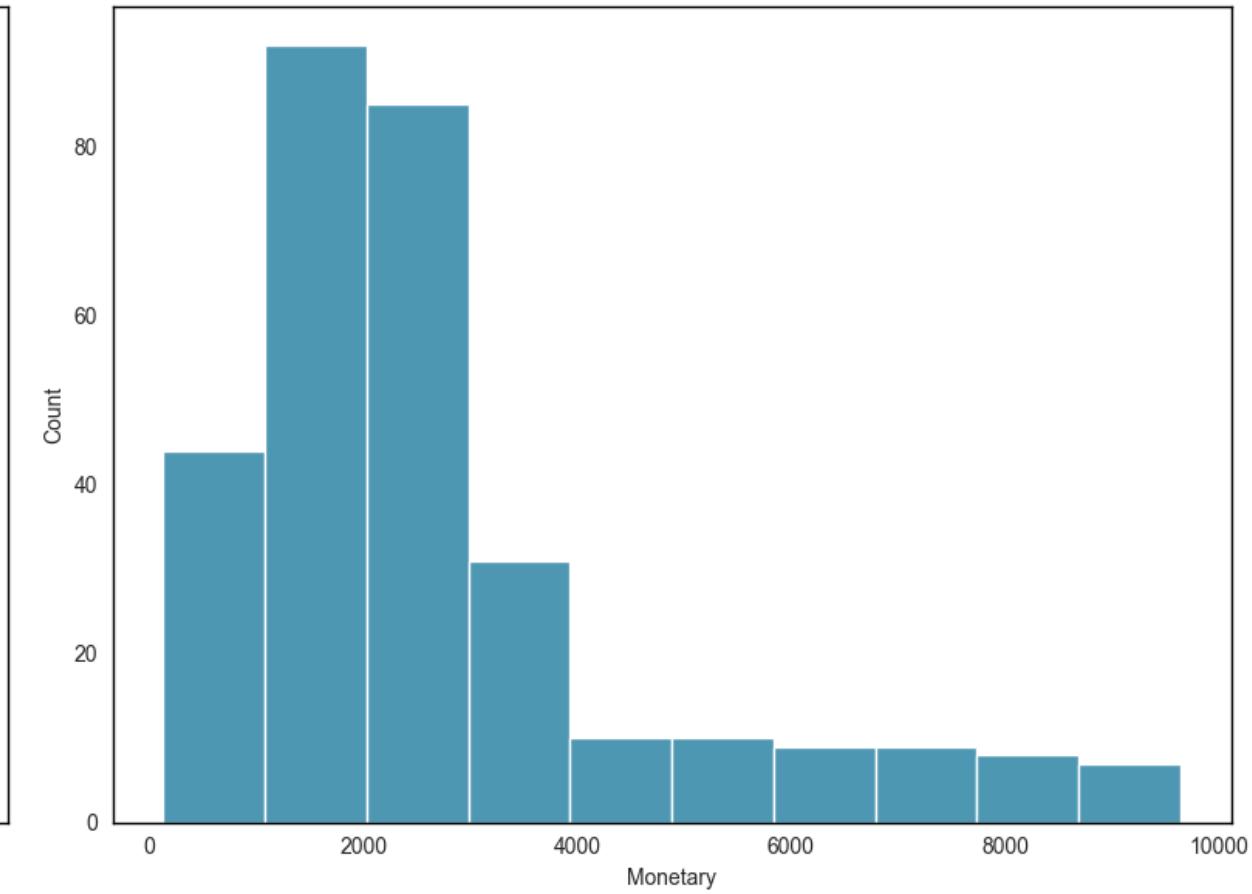
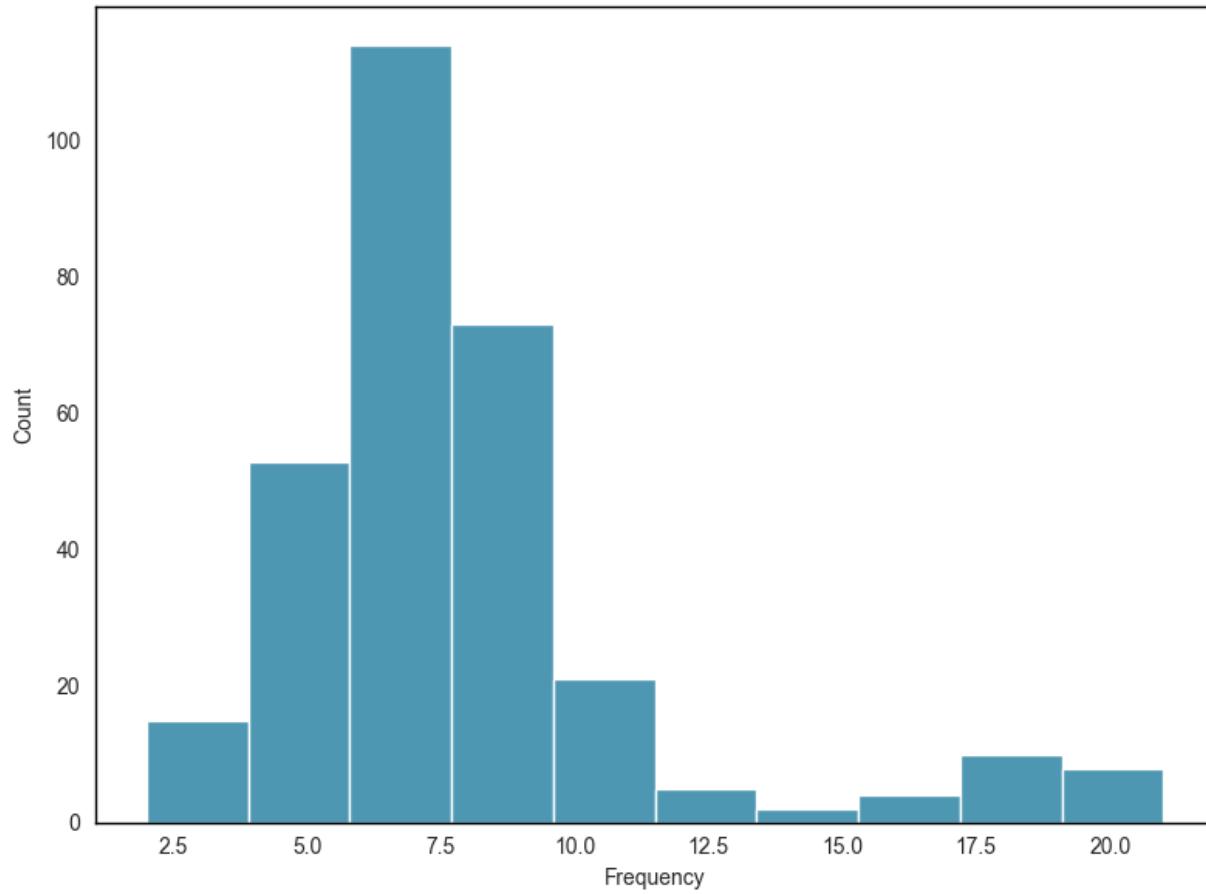
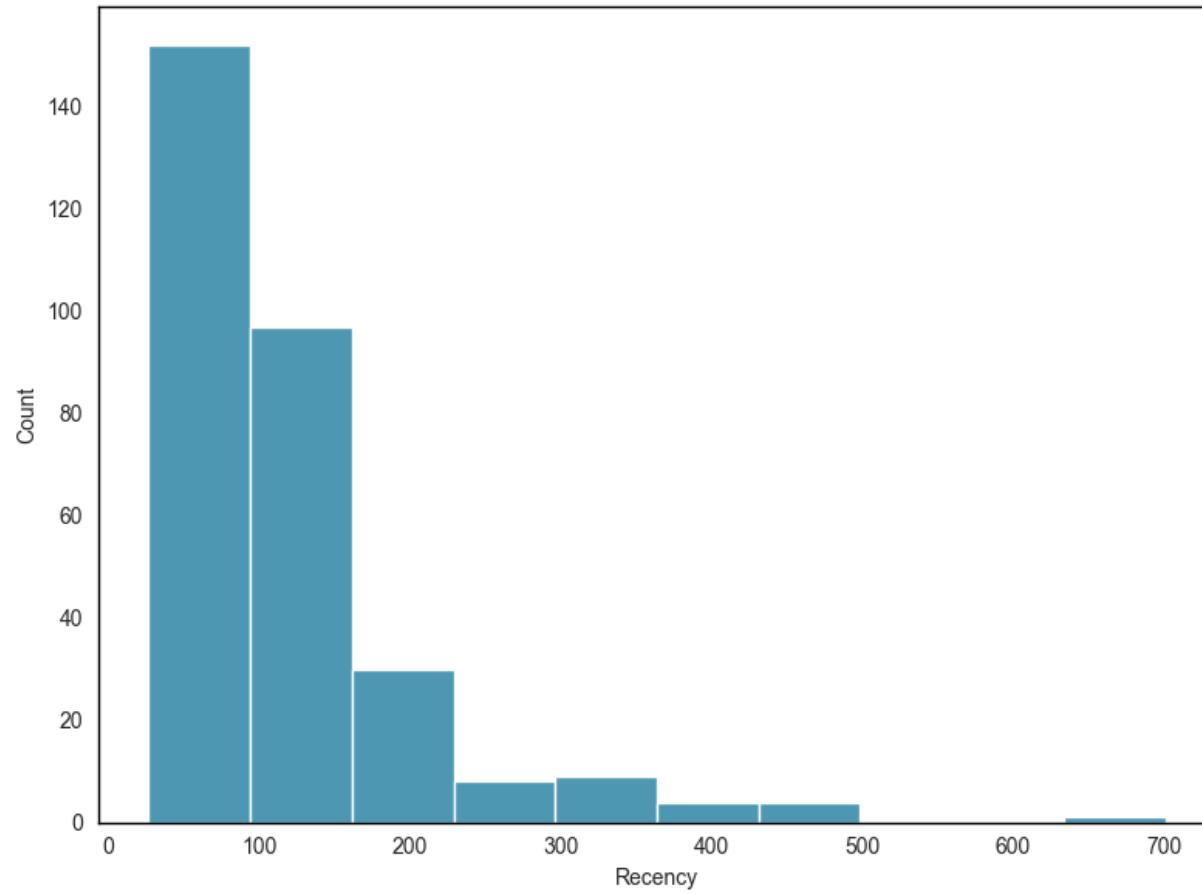
- **Recency:** Very long ago
- **Frequency:** Very low
- **Monetary:** High

Actionable Tip: Win them back through renewals or newer products, don't lose them to competitors. Talk to them if necessary. Spend time personalizing as much as possible.



7. Data Analysis

Hibernating



Observation: Last purchase was quite a long time ago, but they have visited the website or opened emails.

Analysis:

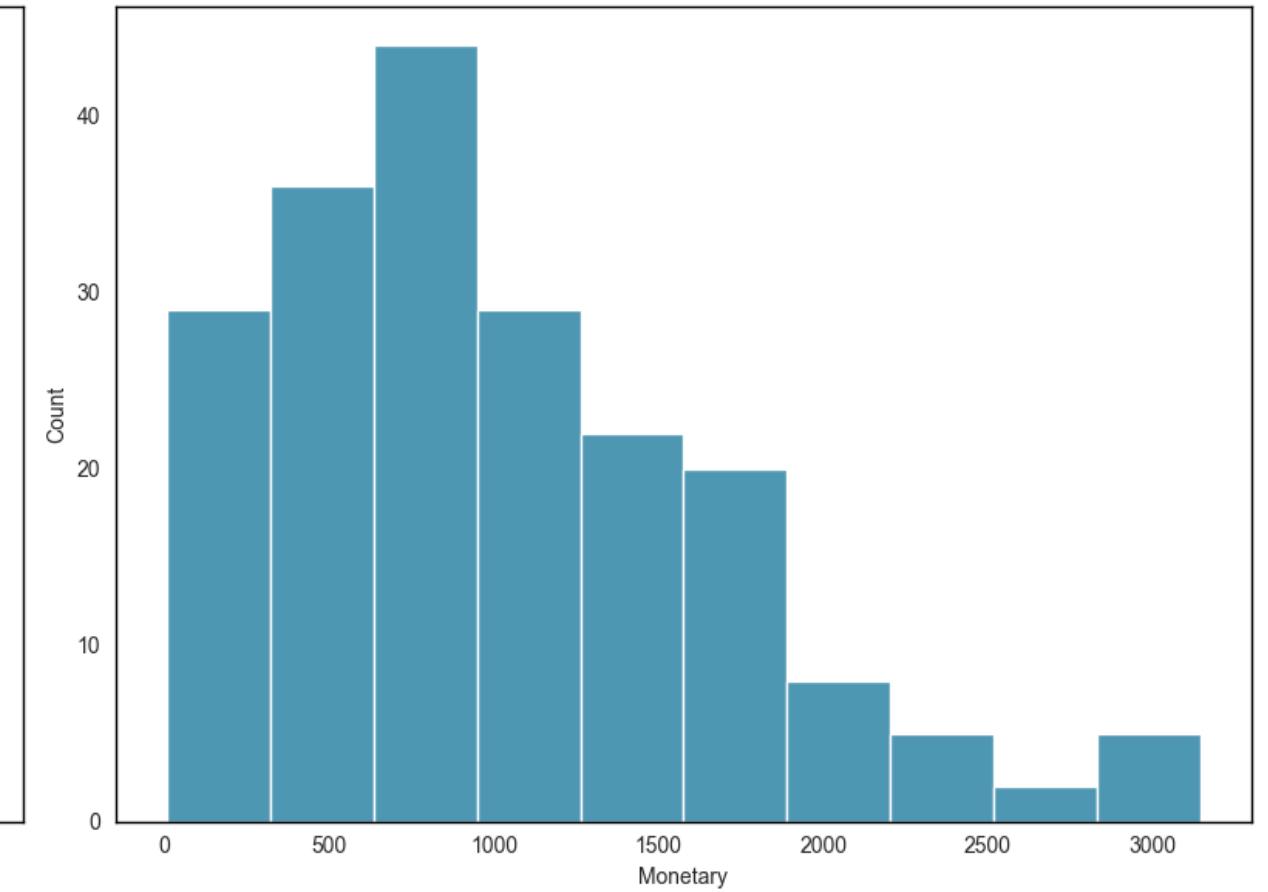
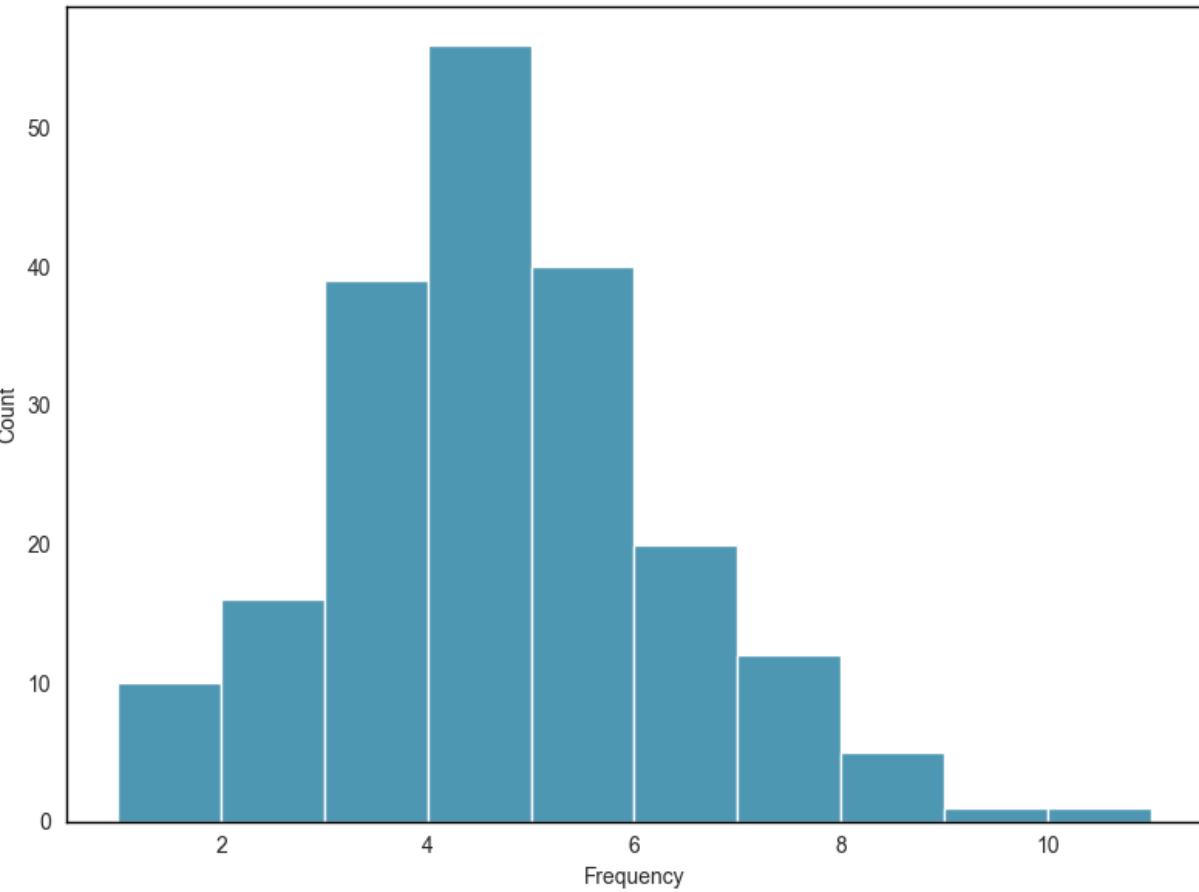
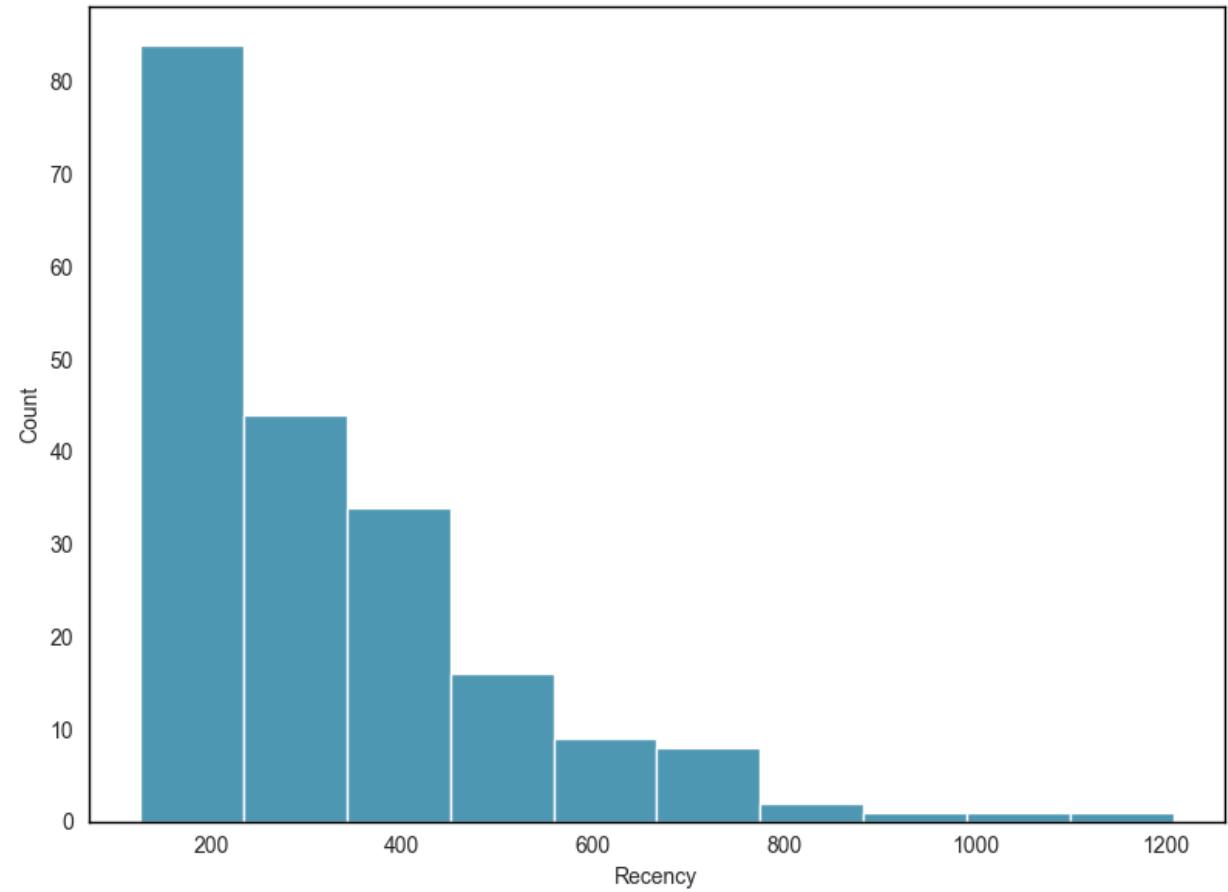
- **Recency:** Very long ago
- **Frequency:** Low
- **Monetary:** Very low

Actionable Tip: Personalize email subject lines... Attract their attention by offering specific discounts for a specific product.



7. Data Analysis

Lost



Observation: Last purchase was quite a long time ago and completely no interaction in the past 4 weeks.

Analysis:

- **Recency:** Very long ago
- **Frequency:** Low
- **Monetary:** Very low

Actionable Tip: Skip in outreach campaigns.





E-Commerce

08 Conclusion



8. Conclusion

Insights

- Purchasing trends are seasonal, decreasing at the beginning of the year and increasing significantly towards the end of the year.
- Sales growth is stable; however, operating costs (Other Cost) also increase linearly with profit, leading to a narrowing profit margin.
- High discount rates do not help increase profits; on the contrary, they reduce profit margins.
- The number of products sold increases steadily over time, but unchanged prices cause profit margins to decline.
- The product portfolio is diverse; however, the frequency of product pairings is very low despite having strong correlations.
- Four key customer groups contribute 80% of the revenue: Champions, Loyal, At Risk, and Potential Loyalist. However, the Lost group has the highest number, with more than 200 customers.

Solution

- Limit discounts to a reasonable level:
 - Apply more discounts in the 1-25% range.
 - Restrict or eliminate discounts of 26-50% and above 50%.
 - Replace discounts with promotional programs such as free gifts with purchases to stimulate demand throughout the year.
- Reduce unnecessary costs to improve profit margins. Adjust pricing strategies appropriately to optimize profitability.
- Increase the recommendation rate of product pairings with high correlation to enhance cross-selling efficiency and increase cart value.
- Maximize engagement with loyal and potential customers. Implement specific actions to attract customers who have left or are at risk of leaving, encouraging them to return.





Thank you