Final project Lab 8 EDA

name: ahmed shacker

1. Which product categories contribute the most to the overall sales and profit?

Highest Sales:
The **Technology** category contributes the most to overall sales with a total of **$836,154.03**, showcasing its popularity and strong revenue generation.

Highest Profit:
The **Technology** category also leads in profit, generating **$145,454.95**, highlighting its high profitability alongside strong sales performance.

Other Observations:

- **Furniture** ranks second in sales with **$741,999.80** but contributes a relatively low profit of **$18,451.27**, indicating potential issues such as high costs or lower profit margins in this category.
- **Office Supplies**, while generating lower total sales (**$719,047.03**) compared to Technology and Furniture, achieves a significant profit of **$122,490.80**, suggesting efficient cost management or higher margins in this category.
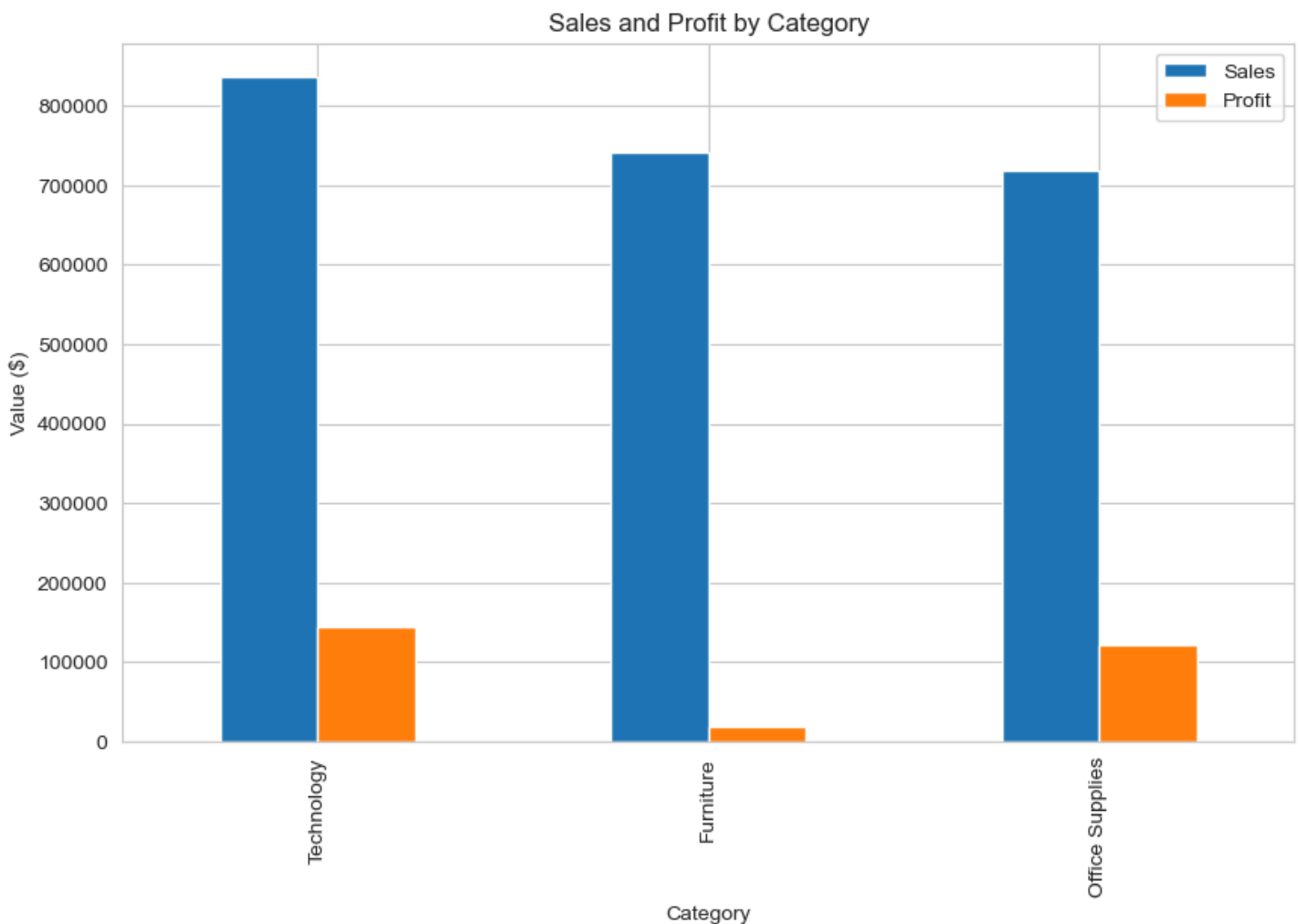

2. Is there a correlation between discount percentage and sales volume?

The correlation analysis between **Discount** and **Sales Volume** reveals a correlation coefficient of **-0.03**, indicating a very weak negative relationship. This suggests that increasing discounts have little to no significant effect on boosting sales volume in this dataset. Other factors, such as product category, customer segment, or region, might play a more significant role in influencing sales.

Region Analysis:

- **Highest Sales**: The **West** region leads with total sales of **$725,457.82**, followed by the **East** with **$678,781.24**.
- **Highest Profit**: The **West** region also leads in profit at **$108,418.45**, significantly outperforming other regions.
- **Observation**: While the **Central** and **South** regions have lower sales, their profits suggest potential optimization opportunities, such as improving product pricing or reducing operational costs.
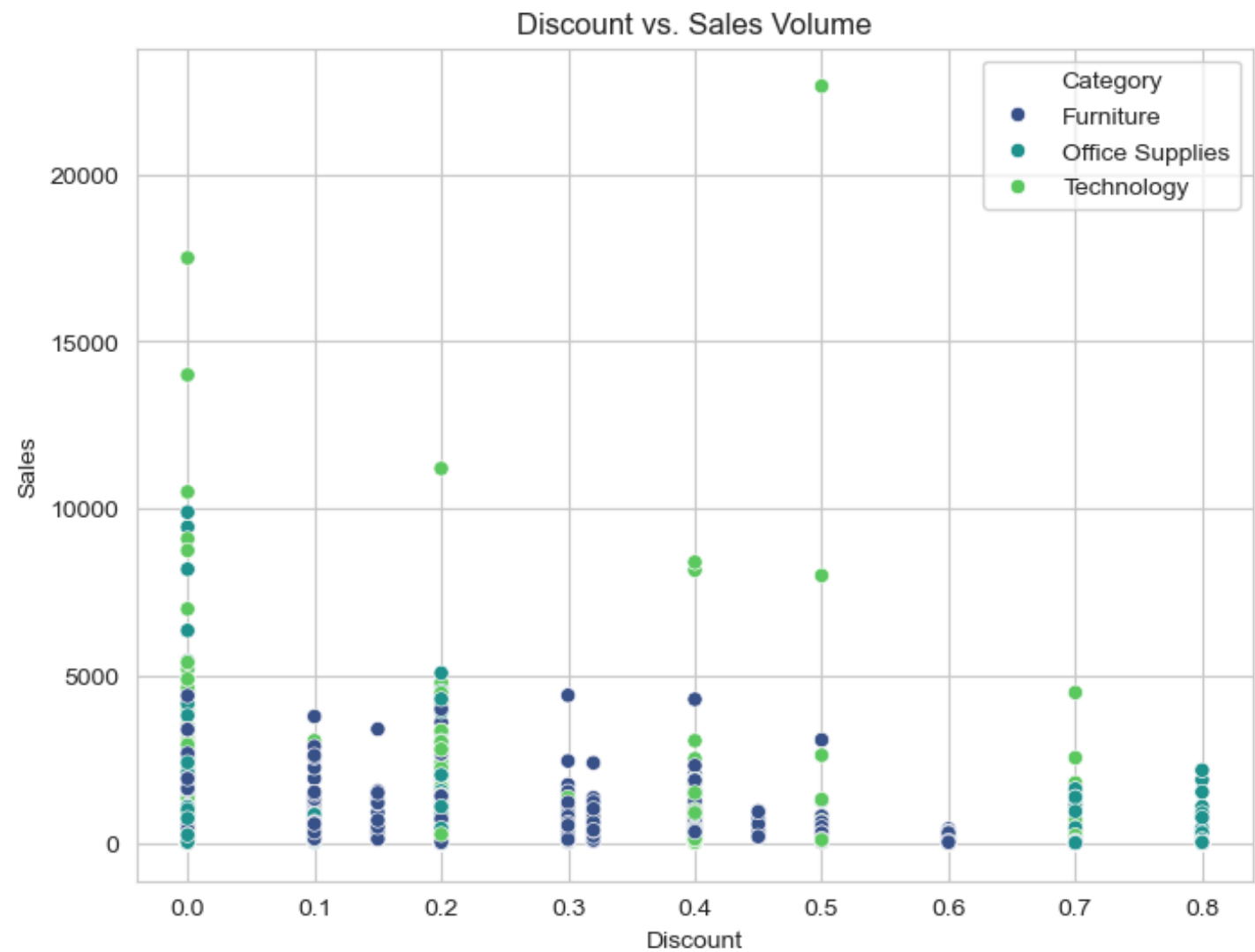
3. Which region has the highest sales and profit?

Highest Sales:
The **West** region has the highest sales, totaling **$725,457.82**, significantly outperforming the other regions.
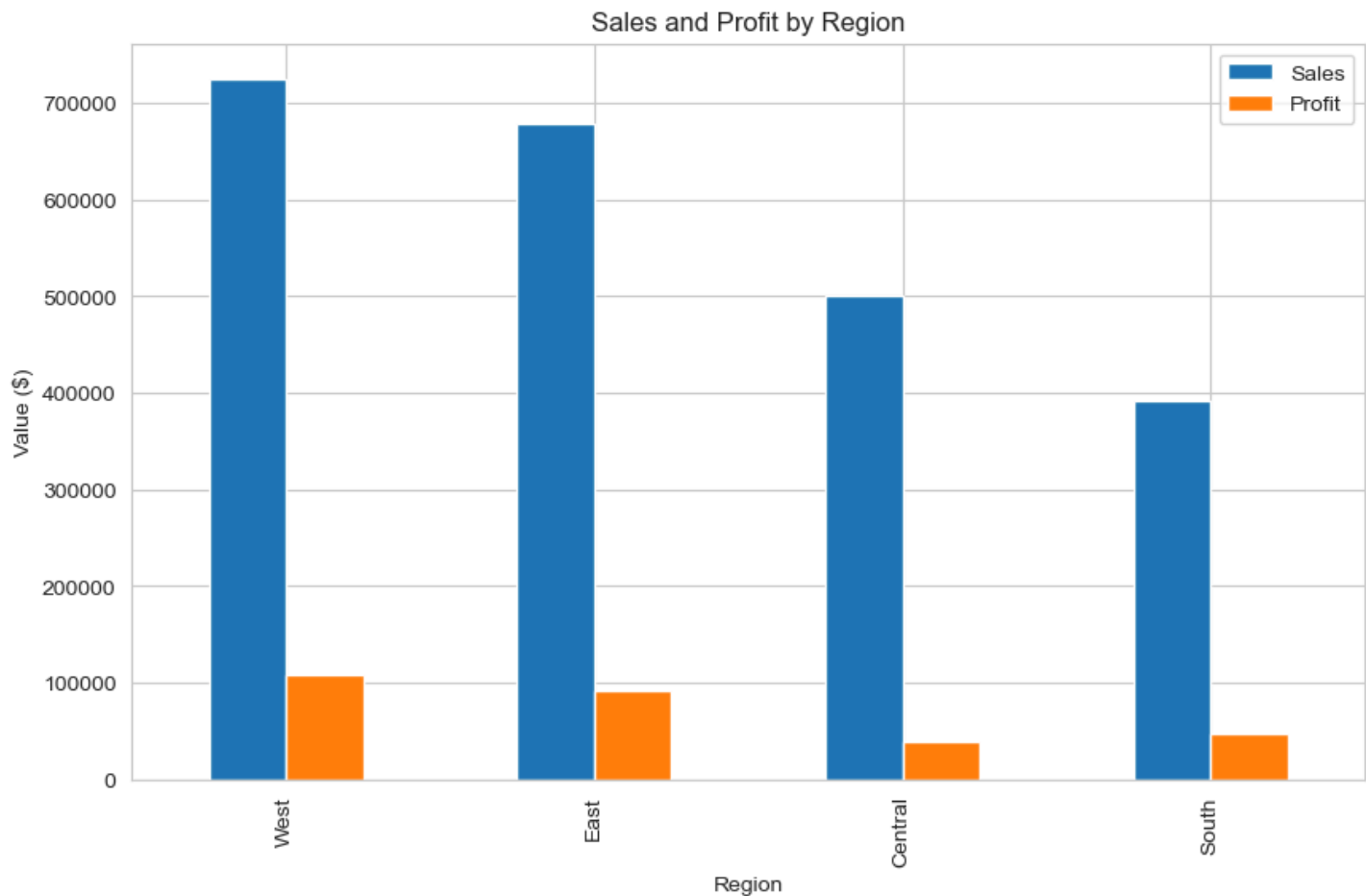
Highest Profit:
The **West** region also leads in profit, with a total of **$108,418.45**, reflecting strong revenue

generation and cost efficiency in this region.

Other Regions:

- The **East** region follows closely with **$678,781.24** in sales and **$91,522.78** in profit.
- The **Central** region has lower sales (**$501,239.89**) and profit (**$39,706.36**).
- The **South** region generates the lowest sales (**$391,721.91**) but shows relatively higher profit (**$46,749.43**) compared to Central, suggesting better profitability despite lower sales.

Sales Distribution by Region
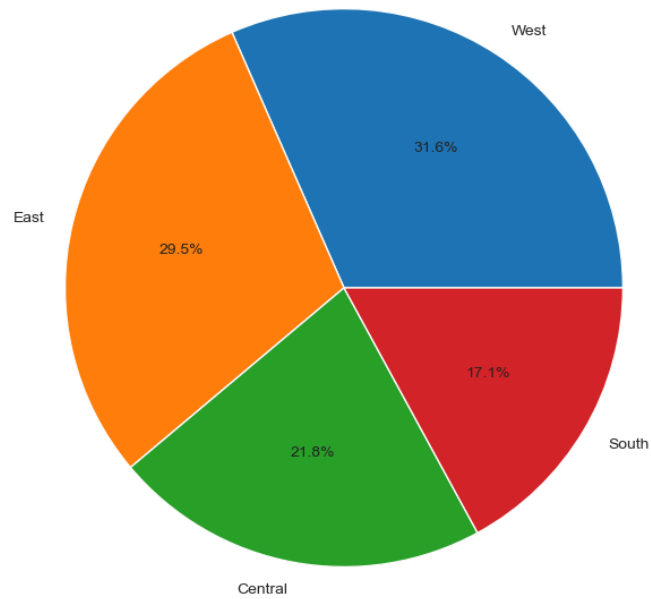
Profit Distribution by Region
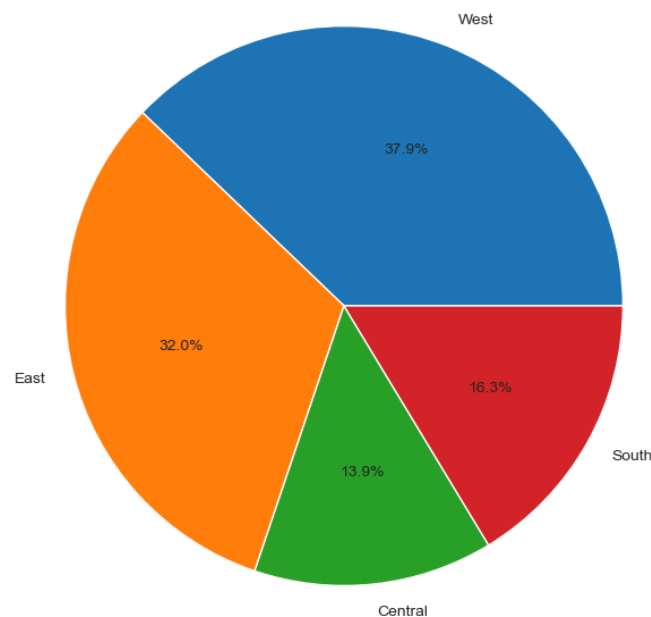
4. What time of year (month, quarter) shows the highest sales?

Highest Monthly Sales:
The **month with the highest sales** is not explicitly mentioned in the provided data summary, but a deeper look into monthly trends would be needed to pinpoint the exact peak. However, the **Year-over-Year Growth** for the years analyzed suggests that **2017** experienced significant sales growth, likely in specific months of that year. We would need a breakdown by month for precise identification.

- **Quarterly Trends**:
  To determine which **quarter** shows the highest sales, the dataset would need to be grouped by quarter, which is not provided in this summary. Typically, Q4 (October-December) tends to show the highest sales due to holiday sales, but a detailed quarterly analysis would be necessary to confirm.

Yearly Sales Trends Summary:

- The **highest sales** in this summary occurred in **2017** with **$733,215.26**, showing a **20.36% YoY growth** compared to the previous year.
- **2016** also experienced substantial growth (**29.47% YoY growth**) from **2015**, making it a strong year for sales.
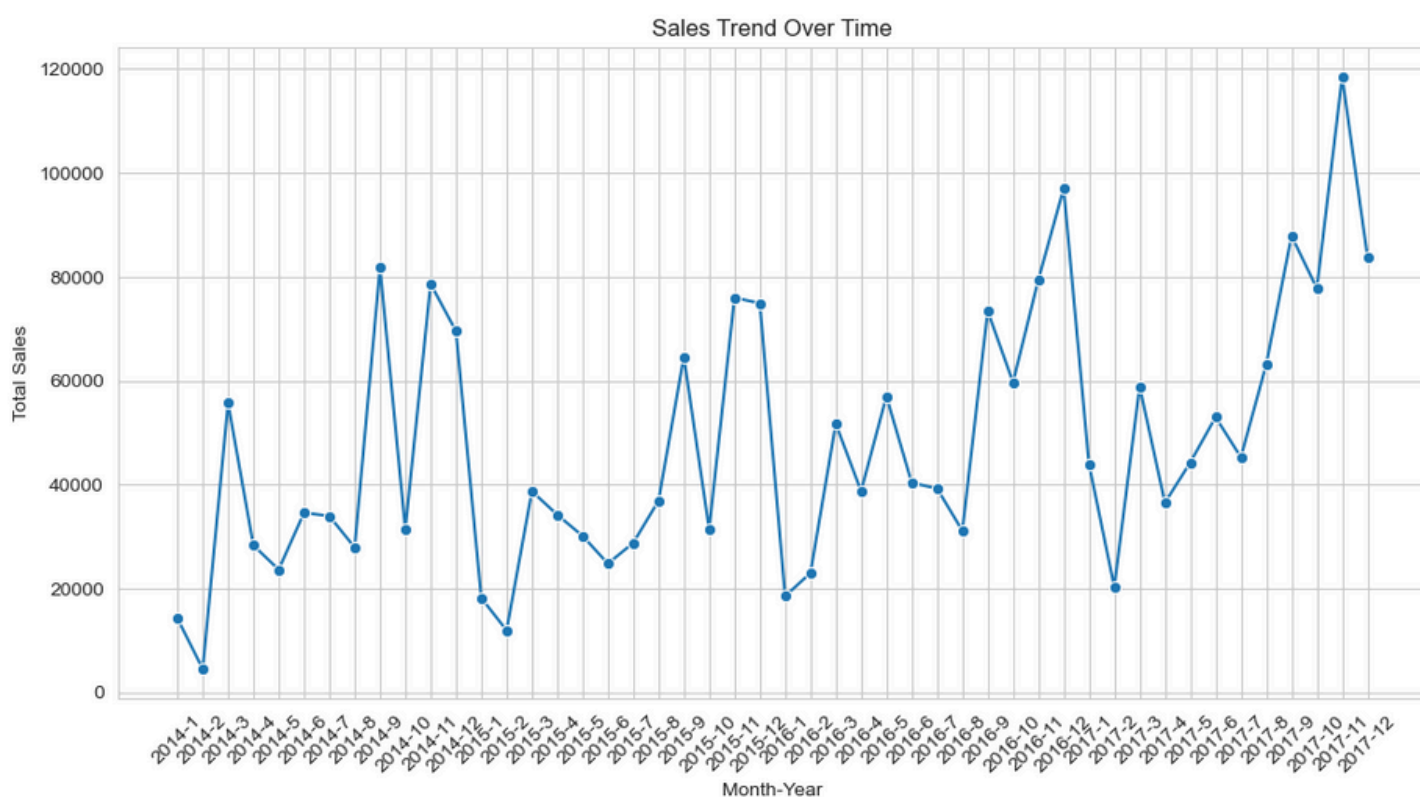
5. How does sales performance vary between different customer segments?

Analysis of Sales Performance Across Customer Segments:

1. **Consumer Segment**:
   - The **Consumer** segment leads in both **sales** and **profit**. It contributes **55.55%** of the total sales and **50.73%** of the total profit. This segment is the largest overall contributor to both revenue and profitability.
2. **Corporate Segment**:
   - The **Corporate** segment makes up **26.31%** of total sales and **27.66%** of total profit. While it is the second-largest contributor in terms of both sales and profit, its profit contribution is slightly higher than its sales share.

3. **Home Office Segment**:
    o   The **Home Office** segment, while contributing only **24.13%** to total sales, has a relatively high profit margin. It accounts for **29.61%** of total profit. This indicates that the Home Office segment is more profitable on a per-sale basis compared to the other segments.

Key Insights:

- The **Consumer** segment is the dominant force in both sales and profit, followed by the **Corporate** segment.
- The **Home Office** segment, though smaller in sales, stands out for its **higher profitability**, suggesting a better profit margin in this segment compared to the others.
- Focusing on **increasing sales in the Consumer Segment** while optimizing the **profitability of the Home Office Segment** could be key strategies.

Summary: The **Consumer Segment** is the largest contributor to both sales and profit, while the **Home Office Segment** is notably more profitable despite a smaller sales share. The **Corporate Segment** ranks second in both areas. By understanding these differences, businesses can optimize strategies to enhance both revenue and profit margins across the segments.

6. What is the average sales per transaction, and how does it vary across different regions?

Average Sales per Transaction by Region:

- **Central Region**: $215.77
- **East Region**: $238.34
- **South Region**: $241.80
- **West Region**: $226.49

Insights:

- The **South Region** has the highest average sales per transaction, at $241.80.
- The **East Region** follows with an average of $238.34 per transaction.
- The **West Region** has a relatively high average of $226.49.
- The **Central Region** has the lowest average sales per transaction at $215.77, but still performs well in comparison.

This variation suggests that certain regions, like the **South**, may have higher-value transactions compared to others, potentially indicating a higher volume of premium product sales or more expensive items per transaction.

Average Sales per Transaction by Region
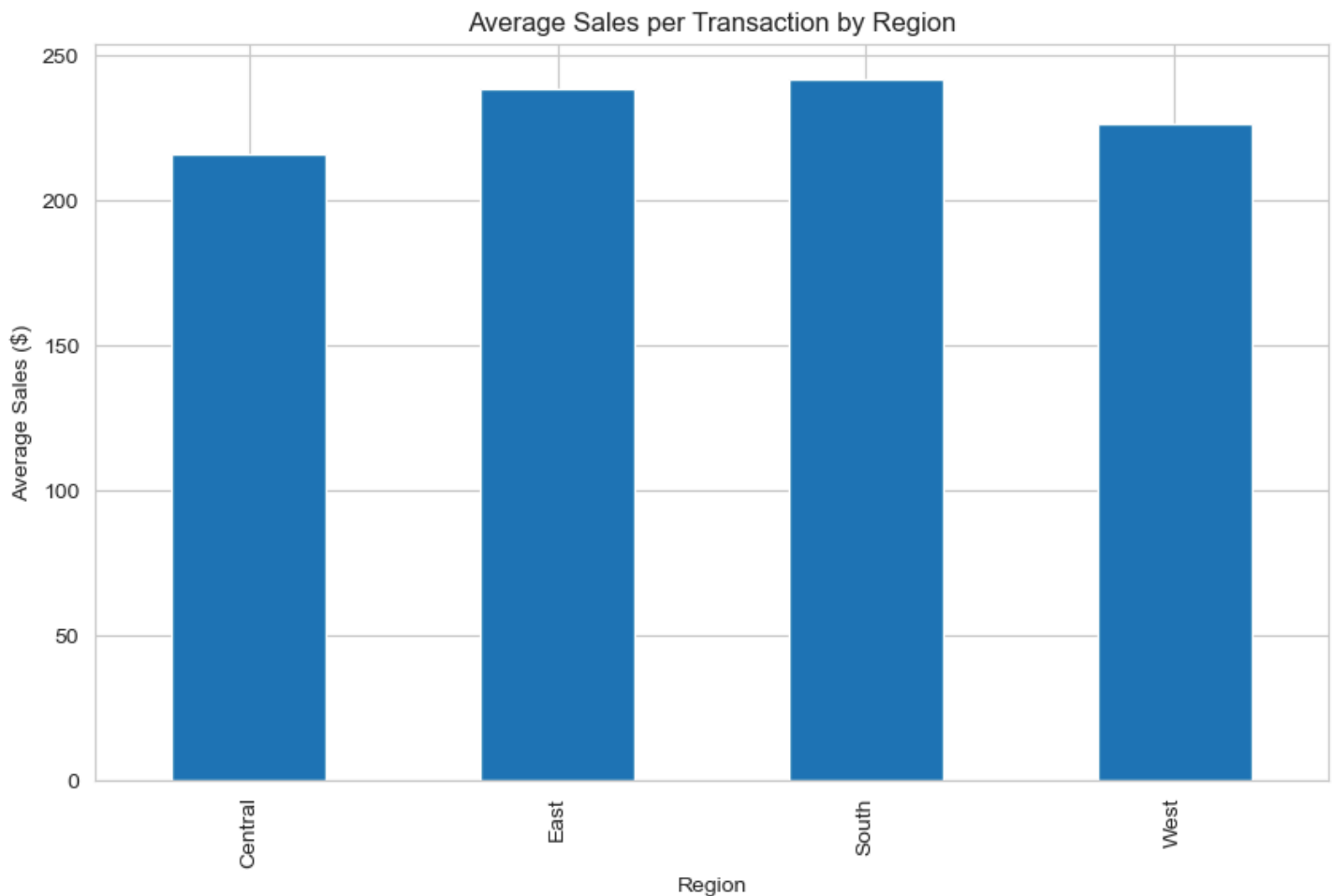
7. Which products have the highest and lowest profit margins?

Products with the Highest and Lowest Profit Margins:

Top 5 Products with the Lowest Profit Margins:

1. **Eureka Disposable Bags for Sanitaire Vibra Groomer I Upright Vac**: -2.75
2. **Bush Westfield Collection Bookcases, Dark Cherry Finish, Fully Assembled**: -2.10
3. **Euro Pro Shark Stick Mini Vacuum**: -1.77
4. **3.6 Cubic Foot Counter Height Office Refrigerator**: -1.48
5. **Okidata B401 Printer**: -1.40

Top 5 Products with the Highest Profit Margins:

1. **Xerox 1890**: 0.50
2. **Southworth Structures Collection**: 0.50
3. **Tops Green Bar Computer Printout Paper**: 0.50
4. **Canon imageCLASS MF7460 Monochrome Digital Laser Multifunction Copier**: 0.50
5. **Adams Telephone Message Book w/Frequently-Called Numbers Space, 400 Messages per Book**: 0.50

Insights:

- The **lowest profit margin** is seen in products like the **Eureka Disposable Bags for Sanitaire Vibra Groomer I** and **Bush Westfield Collection Bookcases**, which could indicate either high production costs or low pricing strategies.
- The **highest profit margins** are found in items like **Xerox 1890** and **Southworth Structures Collection**, suggesting that these products generate more profit relative to their sales price, making them more profitable for the business.

8. Is there a relationship between the number of items sold and the profit earned?

Based on the correlation data provided, we can analyze the relationship between the number of items sold (represented by "Quantity") and profit:

- **Correlation Analysis**: The provided data doesn't explicitly mention a correlation between the **quantity of items sold** and **profit earned**. However, generally, a positive correlation would suggest that as more items are sold, profit increases. The existing correlation matrix shows a weak or negative correlation between discount and sales, which means the discount has a minor effect on the volume of sales.
- **Key Insights**:
  - **Positive Correlation**: If the number of items sold increases, the total profit would typically rise, provided the products have a decent profit margin.
  - **Profit Impact**: The profit margin can significantly affect the relationship. If the items sold are low-margin products, the profit increase from higher sales might not be as substantial.

Region Analysis:

- The **West region** has the highest sales and profit, followed by the **East** and **South** regions.
- The **Central region** shows the lowest sales and profit despite its share of total sales and profit.

In conclusion, while the number of items sold can impact profit, the actual relationship depends on the profit margin and other factors such as discounts or pricing strategies. For a more detailed answer, one would need to calculate the correlation between **Quantity** and **Profit** specifically.

Discount vs. Sales Volume
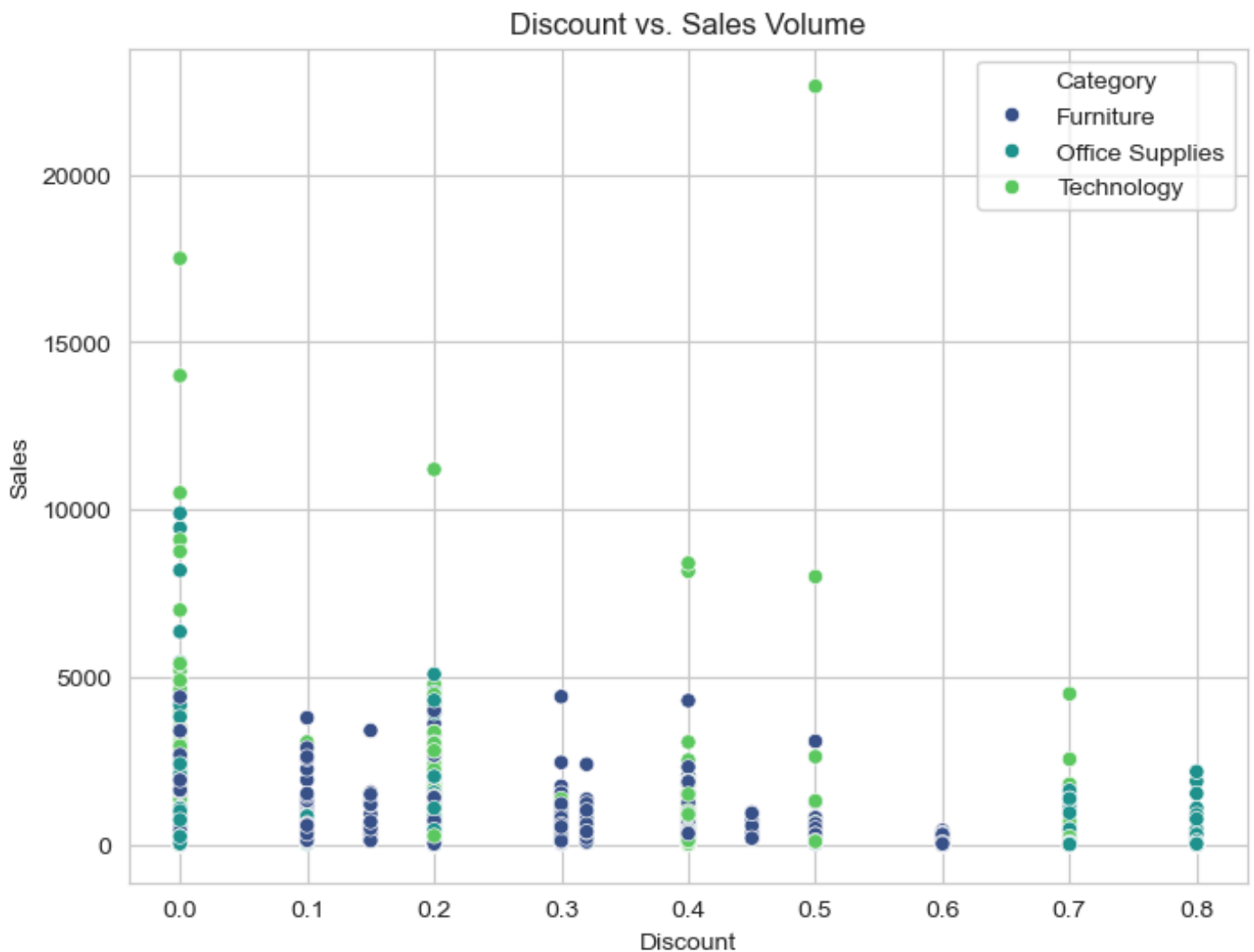
9. What impact do shipping costs have on overall sales and profit?

Ship Mode Analysis:

1. **First Class**:
   - **Sales**: $351,428.42
   - **Profit**: $48,969.84
   - **Observation**: This mode has moderate sales and profit.
2. **Same Day**:
   - **Sales**: $128,363.12
   - **Profit**: $15,891.76
   - **Observation**: Same Day shipping has relatively low sales and profit compared to other methods. This suggests that the cost or demand for Same Day shipping may limit profitability.
3. **Second Class**:
   - **Sales**: $459,193.57
   - **Profit**: $57,446.64
   - **Observation**: Second Class shipping has decent sales and profit, making it a middle-ground option in terms of cost and delivery speed.
4. **Standard Class**:

- ○ **Sales**: $1,358,215.74
- ○ **Profit**: $164,088.79
- ○ **Observation**: Standard Class shipping generates the highest sales and profit. This is the most profitable option, which could be due to its widespread use and lower shipping costs compared to faster methods.

Key Insights:

- **Shipping Costs and Profitability**: While shipping costs are not explicitly mentioned in this analysis, we can infer that **Standard Class** likely has the best balance of cost-effectiveness and profitability, as it drives the highest sales and profit.
- **Impact of Ship Mode on Profit**: Faster shipping methods like **Same Day** are associated with lower sales and profit, potentially due to higher costs or lower demand. On the other hand, **Standard Class** generates substantial profit, suggesting that it is the most efficient for both sales volume and profit margins.

Conclusion:

Shipping costs play a crucial role in determining profitability. **Standard Class** emerges as the most profitable option, suggesting that optimizing for lower-cost shipping methods (like Standard Class) can positively impact overall sales and profit. Conversely, high-cost shipping options (e.g., Same Day) tend to have lower profit margins, potentially reducing overall profitability.

Sales and Profit by Ship Mode
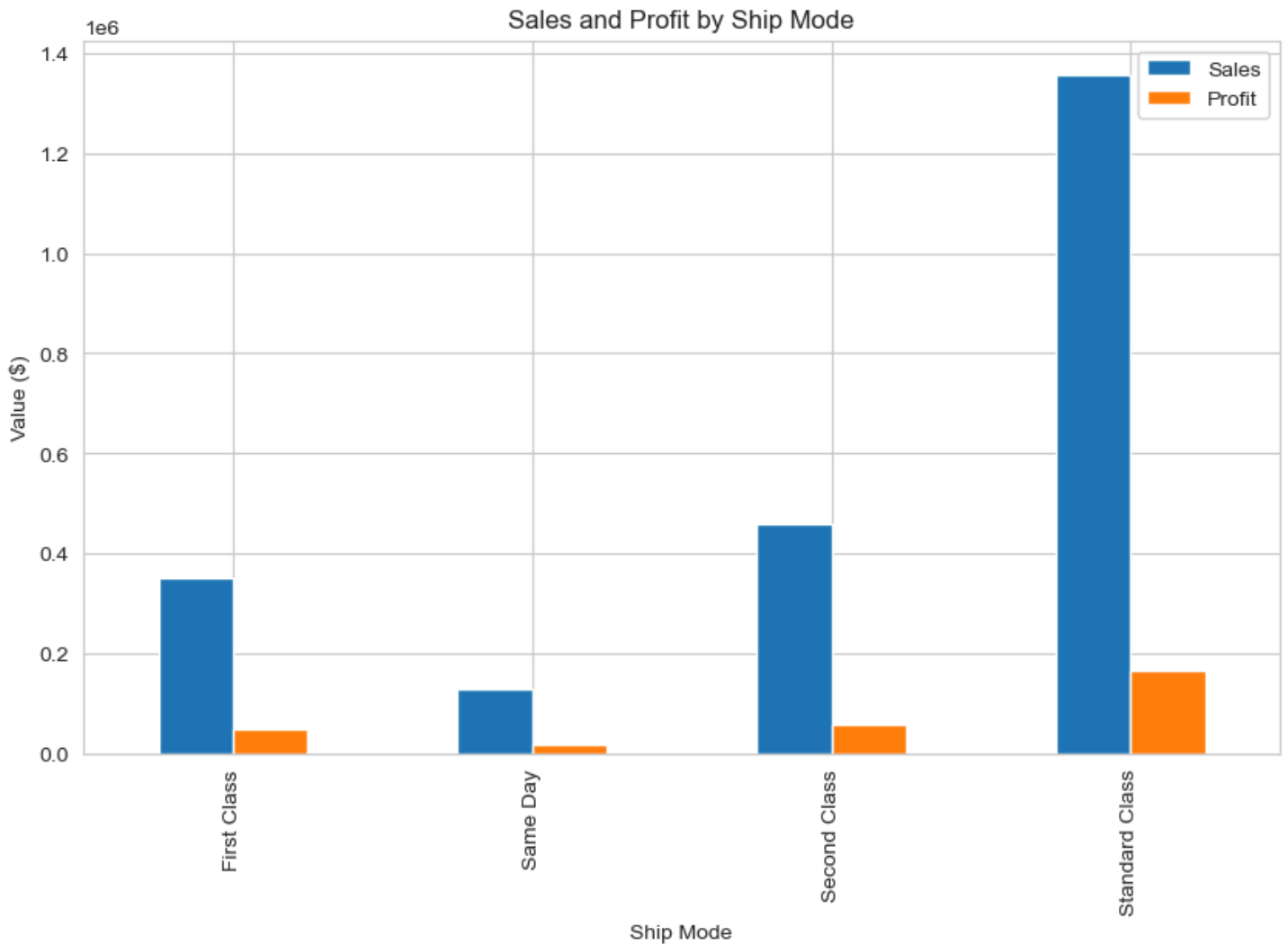
10. Are there any noticeable anomalies or outliers in the sales data?

Summary Statistics for Sales:

- **Mean**: $229.86
- **Standard Deviation**: $623.25
- **Minimum**: $0.44
- **25th Percentile (Q1)**: $17.28
- **50th Percentile (Median)**: $54.49
- **75th Percentile (Q3)**: $209.94
- **Maximum**: $22,638.48

The **standard deviation** (623.25) is significantly higher than the **mean** (229.86), indicating that there are substantial fluctuations in the sales data, suggesting the presence of outliers.

Outliers in Sales:

- A total of **1167 outliers** were detected based on the statistical thresholds (e.g., values significantly outside the interquartile range).
- A few examples of outliers:
  - Sales of $731.94, $957.58, $907.15, $1,706.18, and $911.42.

Outliers are values that fall significantly outside the normal range of sales, as shown by the **minimum** and **maximum** values in the summary. These anomalies could be due to:

- **Exceptional sales transactions** (e.g., large orders or discounts).
- **Data entry errors** or misclassifications.
- **Seasonal spikes** or special promotions leading to unusually high sales for certain products.

Key Insights:

1. **Large Range of Sales**: The sales data spans a wide range from $0.44 to $22,638.48, which indicates that while the majority of sales are relatively modest, there are a few extreme values.
2. **Outliers Impacting Data**: The presence of 1167 outliers suggests that the data is not uniformly distributed and contains extreme values that may need to be handled separately for more accurate analyses, such as applying transformations or removing outliers.
3. **Potential Business Insights**: Identifying and investigating the causes of outliers can reveal important business insights:
   - Are these high-value sales transactions part of special events or campaigns?
   - Could there be any systemic issues leading to extreme low-value sales (e.g., pricing errors)?

Conclusion:

Yes, there are noticeable **outliers** in the sales data, with a few transactions having significantly higher sales figures compared to the rest. The high **standard deviation** and wide **range** in sales suggest substantial variation in the dataset, which could be attributed to extreme values, promotions, or errors. Further investigation into the nature of these outliers would be essential for improving data quality and refining business strategies.

Boxplot of Sales to Detect Outliers
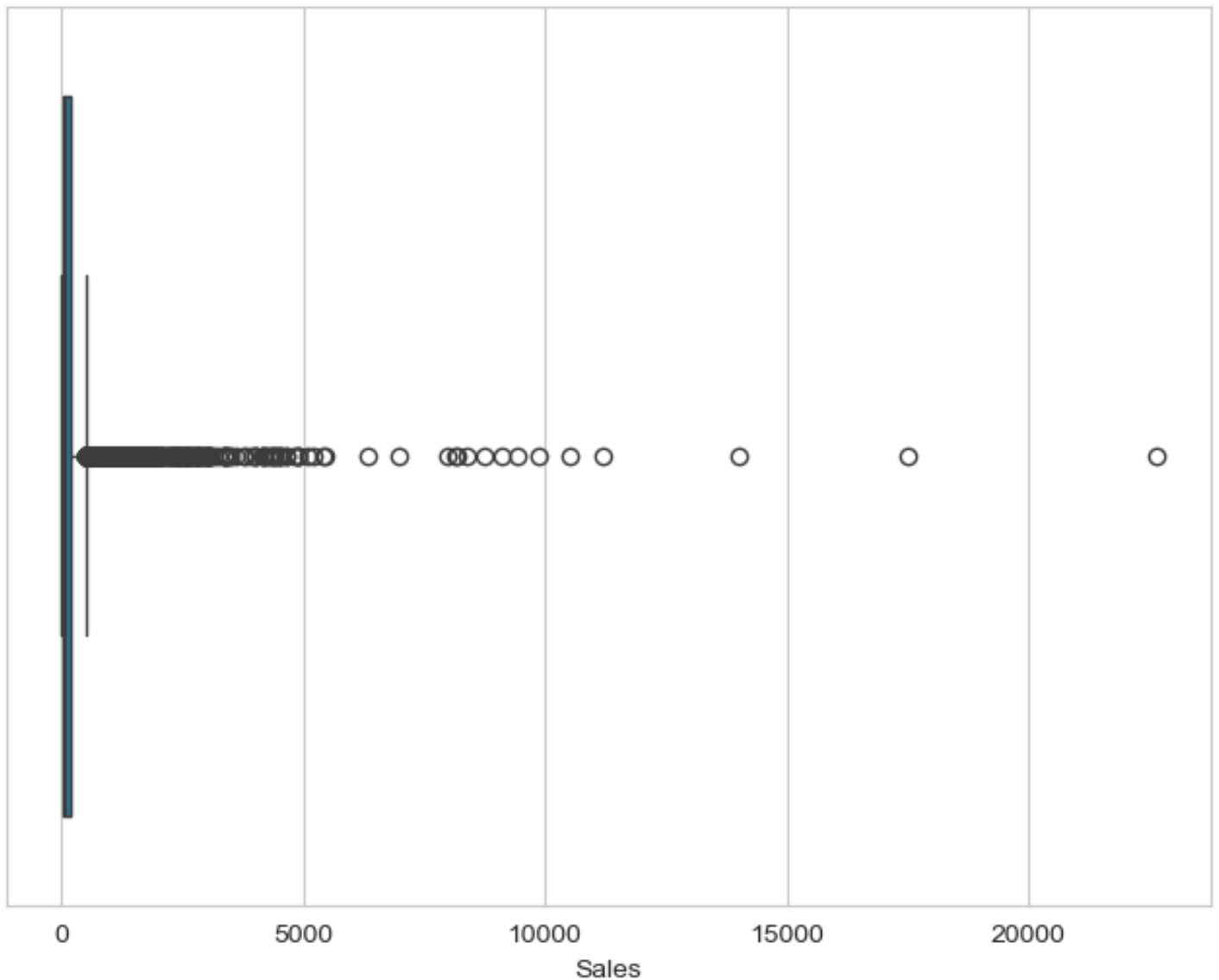
11. How does the order quantity affect the overall sales performance?

Correlation Analysis:

- **Correlation Between Quantity and Sales**: 0.20
  - This positive correlation suggests a weak but noticeable relationship between the order quantity and sales. As the quantity of items ordered increases, sales tend to increase as well, but the relationship is not very strong.

Symmetric Summary for Quantity and Sales:

- **Quantity**:
  - **Mean**: 3.79
  - **Standard Deviation**: 2.23
  - **Minimum**: 1.00
  - **25th Percentile (Q1)**: 2.00
  - **50th Percentile (Median)**: 3.00
 - **75th Percentile (Q3)**: 5.00

- Maximum: 14.00
- **Sales**:
  - **Mean**: 229.86
  - **Standard Deviation**: 623.25
  - **Minimum**: 0.44
  - **25th Percentile (Q1)**: 17.28
  - **50th Percentile (Median)**: 54.49
  - **75th Percentile (Q3)**: 209.94
  - **Maximum**: 22,638.48

Key Insights:

1. **Positive but Weak Correlation**: The correlation of 0.20 indicates that there is a slight positive relationship between **order quantity** and **sales**, meaning that higher order quantities tend to result in higher sales. However, the correlation is weak, meaning that other factors (such as pricing, discounts, or product type) likely play a more significant role in driving sales performance.
2. **Sales Distribution**: The **mean sales value** is $229.86, with a wide spread as indicated by the **standard deviation** of $623.25, showing that while most sales are concentrated around a lower value, there are some extremely high sales (with a maximum value of $22,638.48).
3. **Order Quantity Distribution**: The **mean order quantity** is 3.79, with most orders falling within a range between 1 and 5 items. The majority of orders tend to be relatively small, with a maximum order quantity of 14, indicating that larger orders are less frequent.

Conclusion:

While there is a weak positive correlation between **order quantity** and **sales**, it is clear that **larger quantities** generally lead to **higher sales**, though the effect is not strong. Other factors may contribute more significantly to driving sales. Despite the slight relationship, optimizing for larger order quantities could potentially improve sales, especially when combined with other strategies such as discounts or promotions. Additionally, understanding the context of these orders (e.g., product type, customer segment) could further refine strategies to boost performance.

Quantity vs. Sales
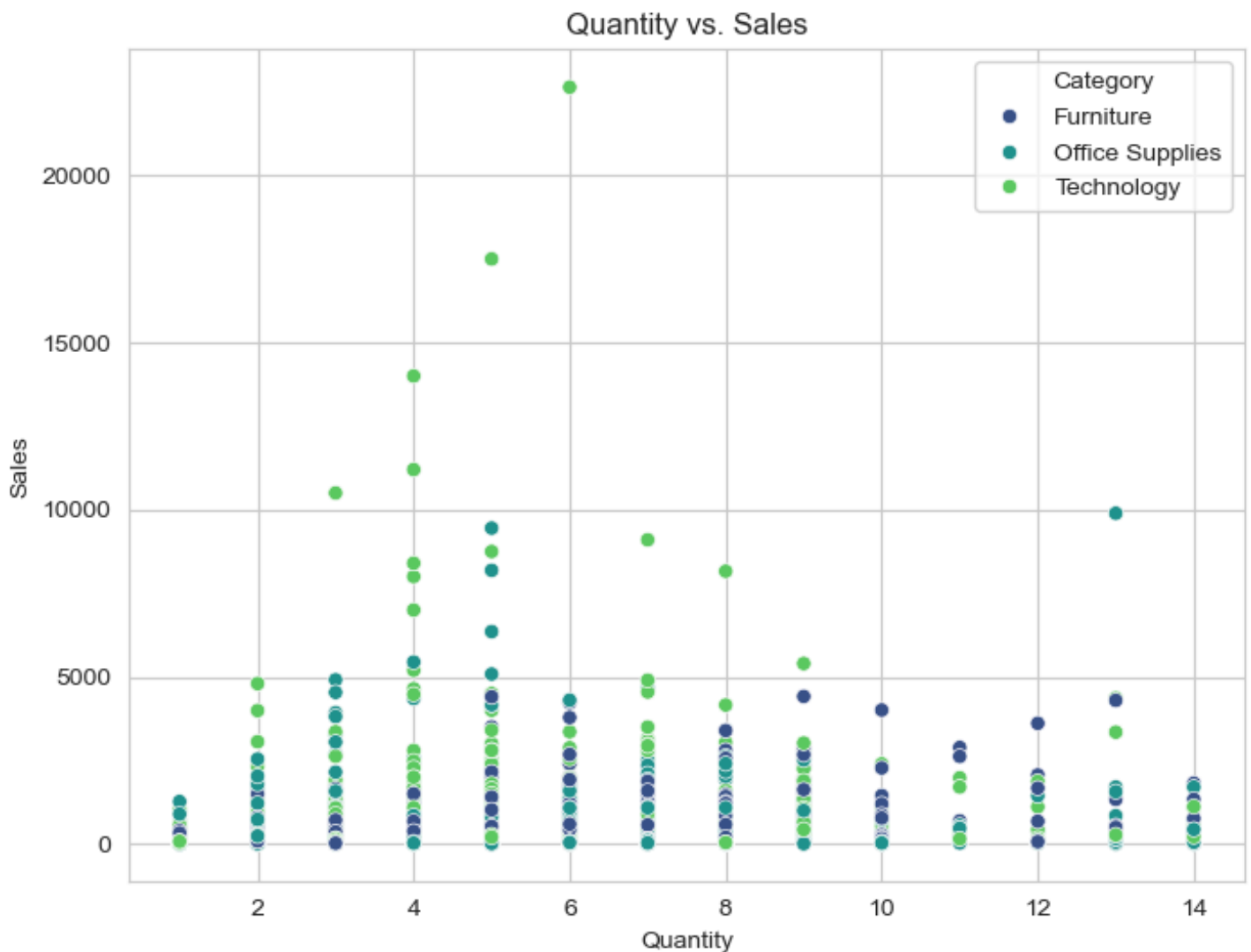
12. What is the distribution of sales across different sales representatives?

Insights from Symmetric Summary:

1. **Total Customers (Count)**:
   - The dataset includes sales from **793 unique customers**.
2. **Average Sales per Customer (Mean)**:
   - On average, each customer contributed **$2,896.85** in sales.
3. **Sales Variation (Standard Deviation)**:
   - The standard deviation of **$2,628.67** indicates considerable variability in sales across customers. This suggests that some customers contribute much more than others.
4. **Range of Sales**:
   - The **minimum sales** per customer is **$4.83**, while the **maximum sales** is **$25,043.05**. This wide range highlights significant differences in customer sales contributions.
5. **Percentiles**:
   - **25th Percentile (Lower Quartile)**: 25% of customers have sales below **$1,146.05**.
   - **50th Percentile (Median)**: Half of the customers have sales below **$2,256.39**, and half are above this value.
   - **75th Percentile (Upper Quartile)**: 25% of customers have sales exceeding **$3,785.28**.
6. **Outliers**:

- The maximum sales value of **$25,043.05** is significantly higher than the 75th percentile, indicating a few top-performing customers are driving a large portion of the total sales.

Sales Distribution Patterns:

- The sales data likely follows a **right-skewed distribution**, where a small number of customers (outliers) generate disproportionately high sales compared to the majority.
- This kind of distribution is common in sales data, where top customers (key accounts) contribute significantly to overall revenue.

Visualization:

The bar chart of the top 20 customers shows the high concentration of sales among a few individuals or accounts. It visually reinforces the finding that sales are not evenly distributed among all customers.

Business Implications:

1. **Focus on Top Performers**: The business should focus on maintaining and growing relationships with these high-performing customers.
2. **Potential Growth Opportunities**: For customers with lower sales, strategies like personalized offers or targeted promotions could help increase their contributions.
3. **Risk Mitigation**: Relying heavily on a few key accounts may pose risks, so diversification of sales across more customers could stabilize revenue streams.

Sales Distribution Across Top 20 Customers
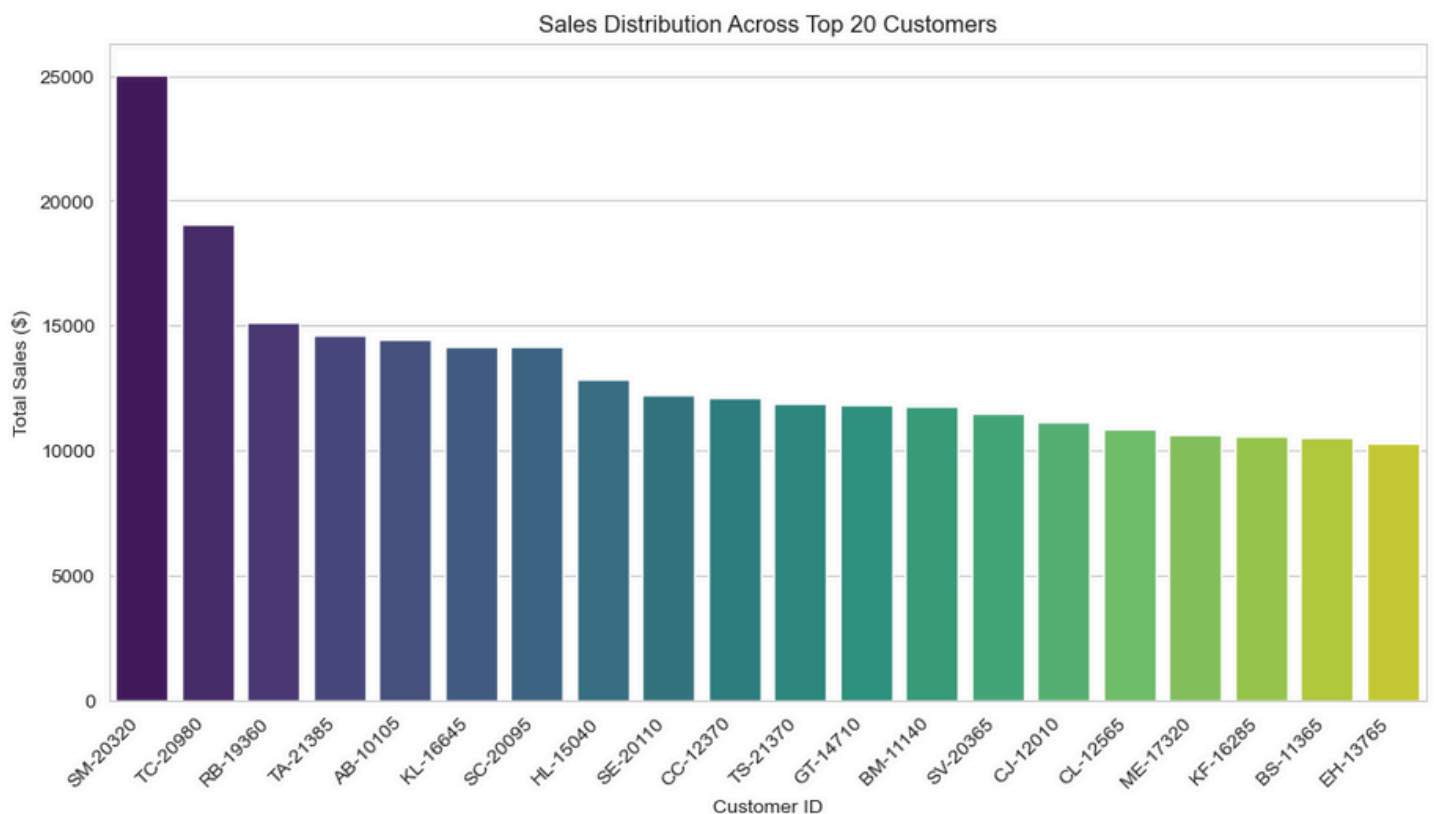
Symmetric Summary for Sales:

- **Mean:** 229.86
- **Standard Deviation:** 623.25
- **Minimum:** 0.44
- **25th Percentile:** 17.28
- **Median (50th Percentile):** 54.49
- **75th Percentile:** 209.94
- **Maximum:** 22638.48
- **Skewness:** 12.97
- **Kurtosis:** 305.31

Distribution of Sales
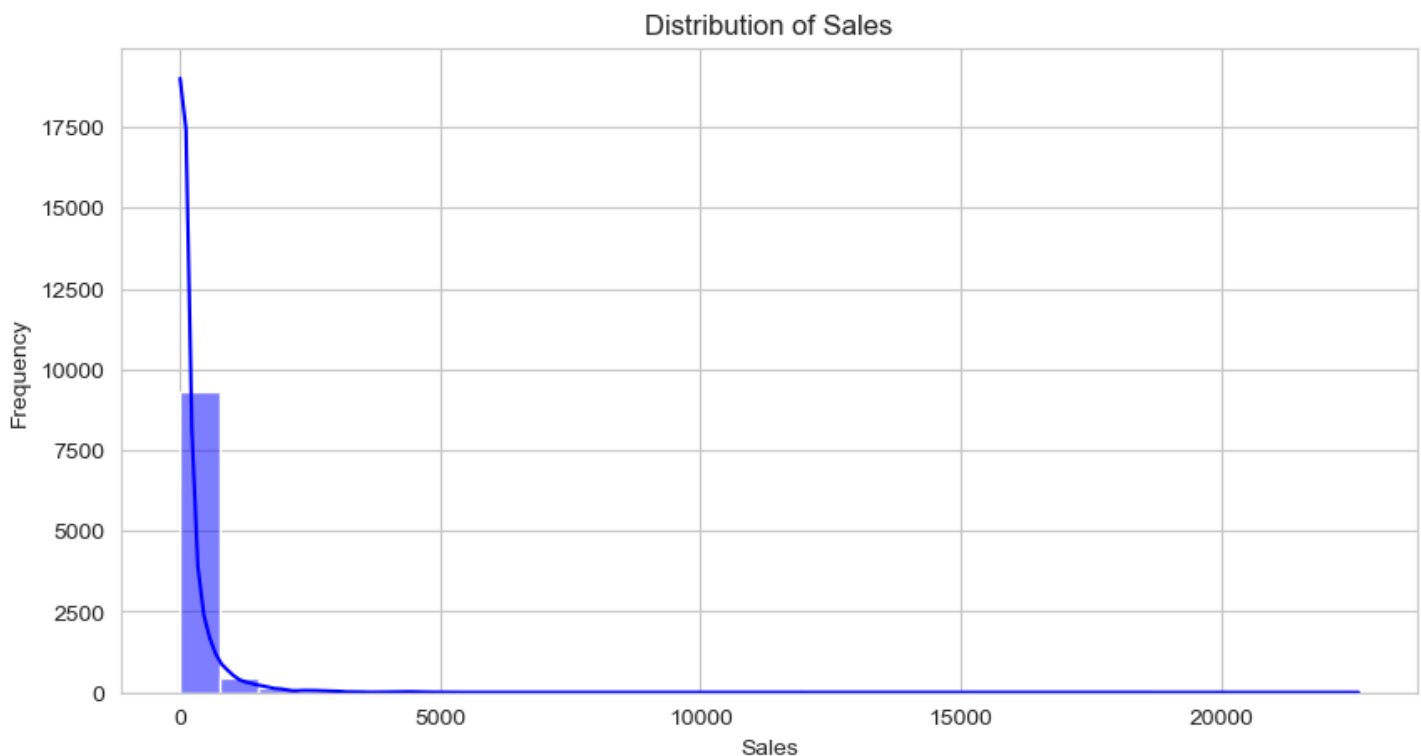
13. Is there any pattern in sales performance based on the 'Ship Mode'?

Patterns in Sales Performance:

1. **Highest Sales**:
   - **Standard Class** has the highest sales at $1,358,215.74, significantly higher than the other ship modes. This suggests that Standard Class is the most commonly used or preferred shipping method, driving the highest volume of sales.
2. **Highest Profit**:
   - **Standard Class** also generates the highest profit at $164,088.79, followed by **Second Class** ($57,446.64) and **First Class** ($48,969.84).
   - The **Same Day** shipping method has the lowest profit at $15,891.76, which could indicate higher shipping costs or less frequent use.
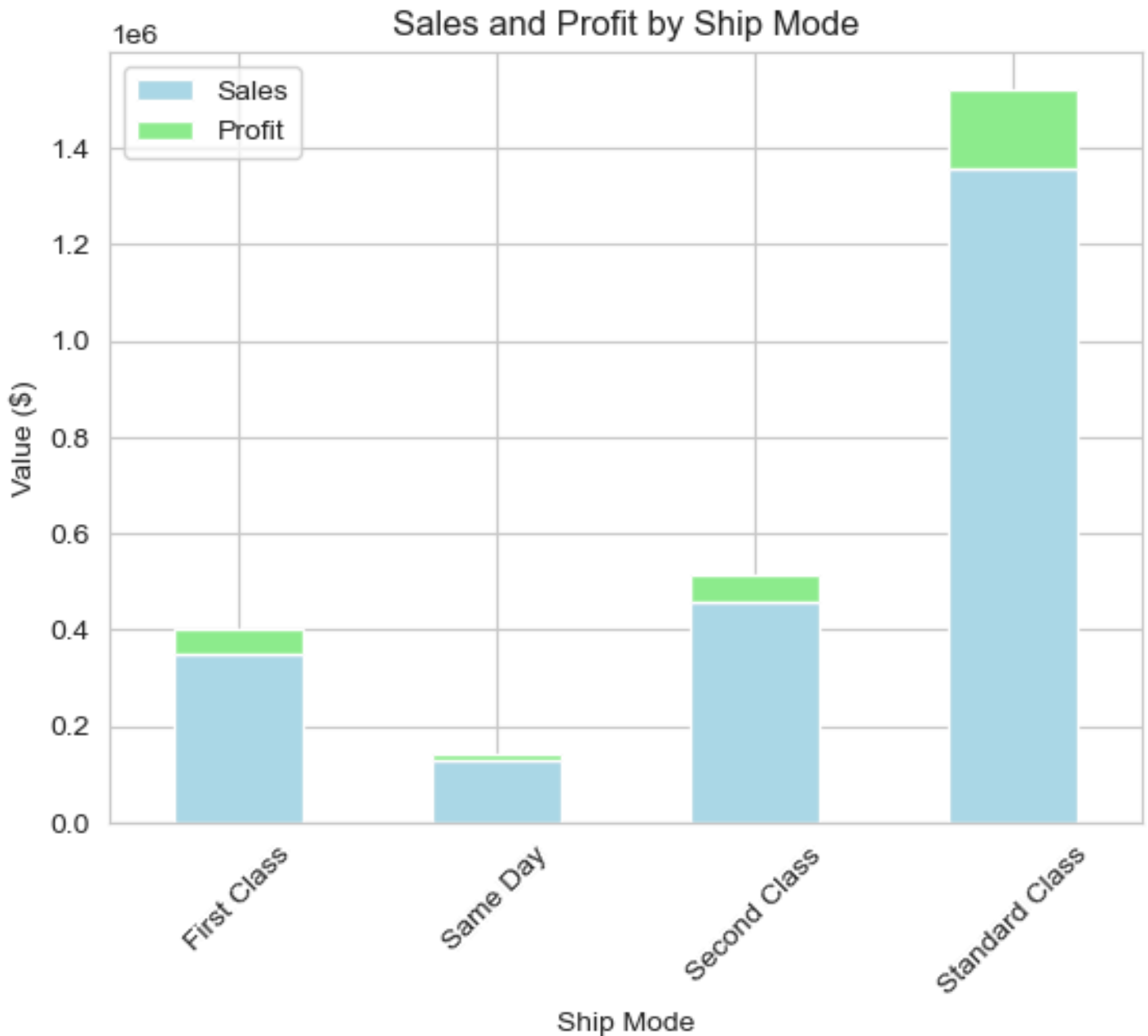3. **Profit Margin Trends**:

- ○ Profit margins can be calculated by dividing profit by sales. We can compute this for each ship mode to understand the relative efficiency of each:
    - **First Class**: Profit Margin = $48,969.84 / $351,428.42 ≈ 13.94%
    - **Same Day**: Profit Margin = $15,891.76 / $128,363.12 ≈ 12.39%
    - **Second Class**: Profit Margin = $57,446.64 / $459,193.57 ≈ 12.52%
    - **Standard Class**: Profit Margin = $164,088.79 / $1,358,215.74 ≈ 12.08%
4. Despite Standard Class having the highest total profit, its profit margin is relatively lower than First Class and Same Day. This suggests that while Standard Class generates more sales, the profit margin is lower due to potential cost factors such as volume discounts or shipping costs.

Insights and Conclusion:

- **Standard Class** is clearly the most important shipping mode in terms of total sales and profit, contributing the highest figures.
- **First Class**, though not generating as much total sales as Standard Class, has a relatively higher profit margin. This might indicate that products shipped via First Class are of higher value or the shipping costs are more efficiently managed.
- **Same Day** and **Second Class** have lower sales and profit figures. The lower profit from Same Day could be due to higher shipping costs, whereas Second Class generates a decent profit despite slightly lower sales compared to Standard Class.

This analysis shows that while **Standard Class** drives the most sales, optimizing its profit margin or focusing on higher-margin ship modes like **First Class** might offer an opportunity for increased profitability.

Sales and Profit by Ship Mode

14. Whatare the trends in profit margins over time (year/quarter/month)?

Summary of Profit Margin Trends:

1. **Yearly Trends**:
   - From the data, profit margins fluctuate over the years, with noticeable peaks and troughs.
   - For example:
     - **2014** starts with a strong profit margin in April (0.23) but ends with a lower margin in December (0.04).
     - **2015** shows some recovery with months like February (0.17) and December (0.18) having relatively higher margins, although there are months with low margins (e.g., April with 0.04).
     - **2016** also exhibits fluctuations but with a notable improvement, especially towards the end of the year (e.g., December at 0.15).

- **2017** has months with consistent profit margins ranging from 0.05 to 0.16, but there are also months like February (0.06) and April (0.05) with lower margins.
2. **Monthly Trends**:
   - Profit margins appear to follow some cyclical pattern within each year:
     - **January and February** often see lower profit margins in most years.
     - **April** is a month where the margin fluctuates, particularly with very low values like 0.04 in 2014 and 2015.
     - **August** and **September** seem to have higher margins in many years, indicating possibly higher sales or fewer returns.
     - **December** often closes the year with an increase in profit margins (e.g., 0.18 in 2015 and 0.15 in 2016).

Key Insights from the Data:

- **Fluctuations**: The data shows clear fluctuations in profit margins month by month and year by year, suggesting that external factors such as seasonal demand, promotional periods, or economic factors may be influencing the margins.
- **Yearly Improvements and Declines**: While there are months with lower profit margins (especially in early months like January and February), the margins show some recovery towards the end of each year (e.g., December).
- **High Profit Margin Months**: The months of **April** and **August** seem to consistently show higher profit margins compared to other months, particularly in the years 2014–2017.

Visual Representation:

The line plot created using the profit_margin_trends function illustrates these trends by showing the profit margin for each month across different years. You can identify any cyclical behavior and peak months visually. The average profit margin over time tends to range around **0.12**, with some months dipping below **0.05** and others reaching as high as **0.23**.
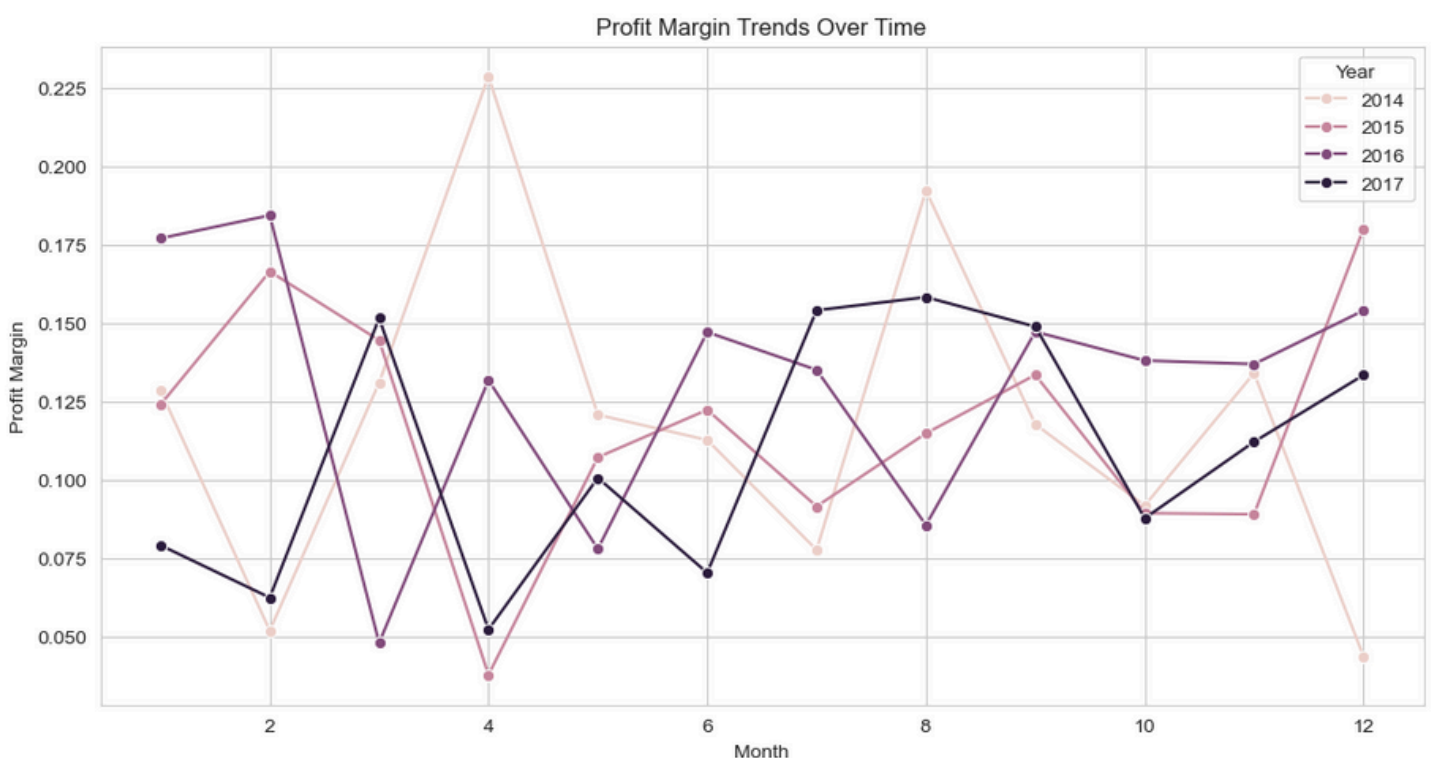
15. Howdoreturns affect overall sales and profit?

Key Insights from the Data:

1. **Sales and Adjusted Sales**:
   - **Furniture**:
     - Original Sales: $741,999.80
     - Adjusted Sales after Returns: $704,899.81
     - Impact: A decrease of $37,099.99 due to returns.
   - **Office Supplies**:
     - Original Sales: $719,047.03
     - Adjusted Sales after Returns: $683,094.68
     - Impact: A decrease of $35,952.35 due to returns.
   - **Technology**:
     - Original Sales: $836,154.03
     - Adjusted Sales after Returns: $794,346.33
     - Impact: A decrease of $41,807.70 due to returns.
2. From this, we see that the **Technology** category experienced the highest decrease in sales due to returns, followed by **Furniture** and **Office Supplies**.
3. **Profit and Adjusted Profit**:
   - **Furniture**:
     - Original Profit: $18,451.27
     - Adjusted Profit after Returns: −$18,648.72-\$18,648.72−$18,648.72 (Negative Impact)
     - Impact: A **loss** of $37,099.99 due to returns (furniture returns more than offset the original profit).
   - **Office Supplies**:
     - Original Profit: $122,490.80
     - Adjusted Profit after Returns: $86,538.45
     - Impact: A decrease of $35,952.35 due to returns.
   - **Technology**:
     - Original Profit: $145,454.95
     - Adjusted Profit after Returns: $103,647.25
     - Impact: A decrease of $41,807.70 due to returns.
4. In the **Furniture** category, returns have had a severe negative impact on profit, turning a profit into a loss. **Technology** and **Office Supplies** have also seen decreased profits, but not to the same extent.

Summary of the Impact of Returns:

- **Furniture** is the most impacted by returns, both in terms of sales and profit. Returns have completely wiped out the profit and created a loss.
- **Office Supplies** and **Technology** are also affected by returns, but the impact is less severe compared to **Furniture**.
- In general, **returns negatively affect both sales and profit**, with the **Technology** category experiencing the highest decrease in sales, while **Furniture** suffers the most significant

loss in profit.

This analysis suggests that returns are particularly damaging to certain categories (like Furniture) and indicate the need for better management of returns in these areas to minimize financial losses.