

SuperStore Sales Data Analysis

SuperStore Sales Data Analysis

About the dataset

This report summarizes a dataset of sales transactions from a SuperStore, encompassing 5901 rows of data with 23 columns. The dataset includes information such as Row ID, Order ID, Order Date, Ship Date, Ship Mode, Customer ID, Customer Name, Segment, Country, City, State, Region, Product ID, Category, Sub-Category, Product Name, Sales, Quantity, Profit, Returns, and Payment Mode.

The sample data reveals several insights. For instance, the first three orders share the same Order ID ('CA-2019-160304') and Ship Date but vary in Order Date, Product ID, Category, Sub-Category, Product Name, Sales, and Profit, suggesting multiple items were ordered together but recorded on different days. The Ship Mode is predominantly 'Standard Class' in the initial entries. The Returns column mostly contains '#N/A', indicating a lack of return data in the initial entries. The average Sales value is 265.35 with a standard deviation of 474.22, indicating a wide range of sales values. The average Quantity is approximately 3.78 items per order. The Profit column has a mean of 29.70 and a standard deviation of 259.57, with a minimum value of -6599.98 and a maximum of 8399.98, suggesting both substantial profits and losses within the dataset. The dataset covers various geographical locations within the United States, as indicated by the City and State columns.

Relevant Inquiries

Q1.What is the monthly sales trend over the past year? Show the sales growth rate MoM.



Monthly Sales Trend

- **Overall Trend:** Monthly sales show a general upward trend over the analyzed period, particularly in the latter half of 2020.
- **Sales Fluctuations:**
 - Sales started at **\$731.77** in December 2019.
 - A significant jump occurred in **January 2020** to **\$54,763.37**.
 - Sales then fluctuated, with a dip in **February 2020** to **\$52,068.13** and **April 2020** to **\$50,807.54**.
 - Sales began to consistently increase from **July 2020** (\$60,077.42) onwards, reaching a peak of **\$166,185.85** in **December 2020**.

- **Highest Sales:** The highest monthly sales were recorded in **December 2020** at **\$166,185.85**.
- **Lowest Sales (excluding initial month):** The lowest sales after the initial month were in **April 2020** at **\$50,807.54**.

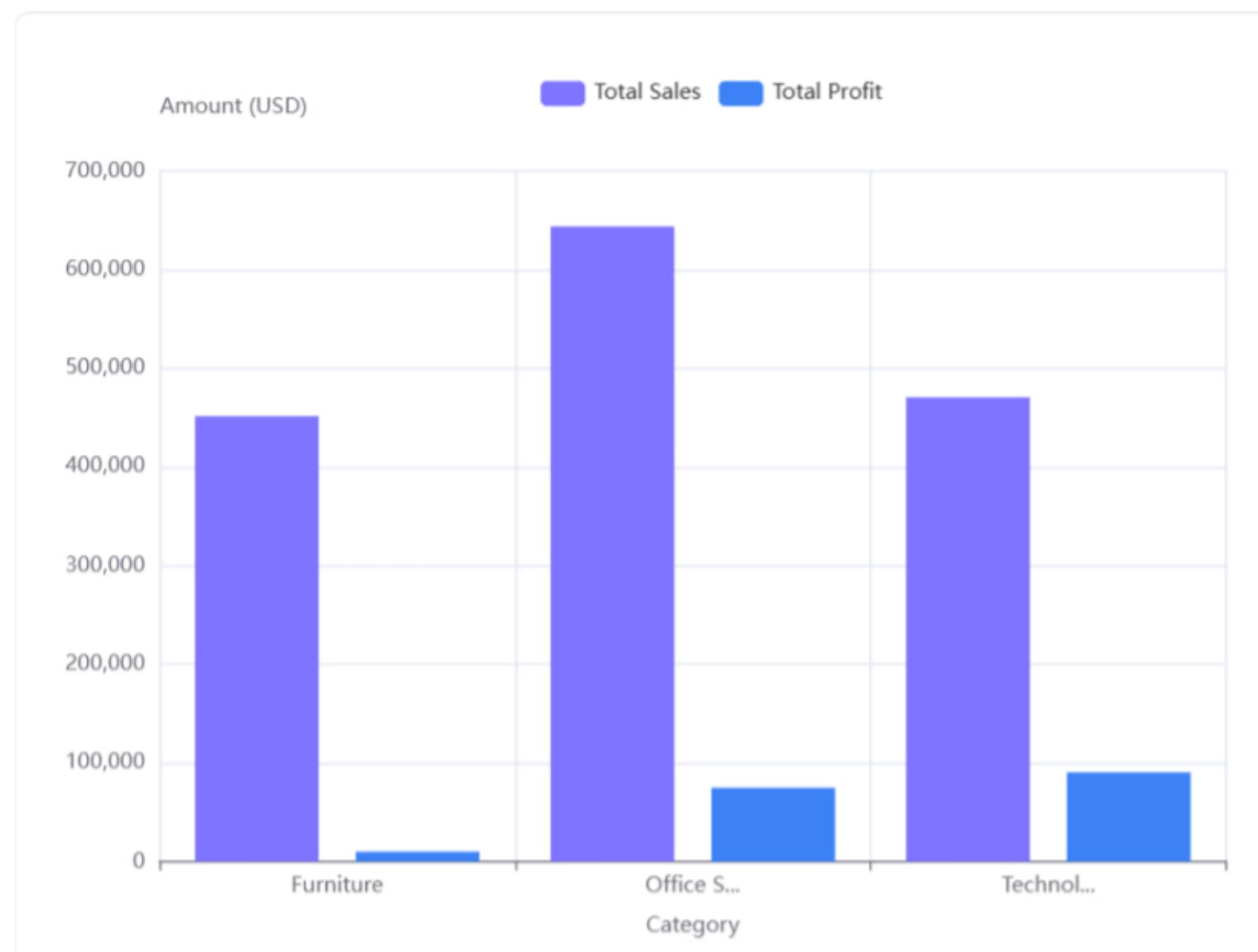
Month-over-Month (MoM) Growth Rate

- **Initial Growth Spike:** The most substantial growth occurred from December 2019 to January 2020, with an extraordinary **7383.71%** MoM growth rate, largely due to the very low base sales in December 2019.
- **Fluctuating Growth:** After the initial spike, the MoM growth rate experienced significant fluctuations throughout 2020.
 - **Negative Growth:** Sales declined in **February 2020 (-4.92%)**, **April 2020 (-20.90%)**, **July 2020 (-13.24%)**, and **October 2020 (-25.80%)**.
 - **Positive Growth:** Notable positive growth rates were observed in **March 2020 (23.36%)**, **May 2020 (31.38%)**, **August 2020 (28.73%)**, **September 2020 (54.91%)**, **November 2020 (47.32%)**, and **December 2020 (26.90%)**.
- **Highest Growth (excluding initial spike):** The highest positive growth rate after the initial spike was **54.91%** in **September 2020**.
- **Largest Decline:** The largest MoM decline was **-25.80%** in **October 2020**.

Conclusion and Insights

- **Strong Year-End Performance:** The sales data indicates a strong performance towards the end of 2020, with sales consistently increasing from August to December and reaching their peak in December. This suggests a positive momentum heading into the new year.
- **Volatility in Growth:** While overall sales show an upward trend, the Month-over-Month growth rate is highly volatile, experiencing both significant gains and declines. This volatility suggests that sales are not growing at a consistent pace and may be influenced by seasonal factors or specific events.
- **Impact of Base Sales:** The extremely high MoM growth rate in January 2020 is an outlier driven by very low sales in the preceding month (December 2019), highlighting the importance of considering the base period when interpreting growth percentages.
- **Areas for Investigation:** The periods of negative growth (e.g., February, April, July, October) warrant further investigation to understand underlying causes such as market conditions, promotional activities, or operational challenges.

Q2. What is the total sales and profit for each product category? Display the profit margin by category.



Furniture

- **Total Sales: 451,508.65 USD**
- **Total Profit: 10,006.61 USD**
- **Profit Margin: 2.22%**

Office Supplies

- **Total Sales: 643,707.69 USD**
- **Total Profit: 74,797.25 USD**
- **Profit Margin: 11.62%**

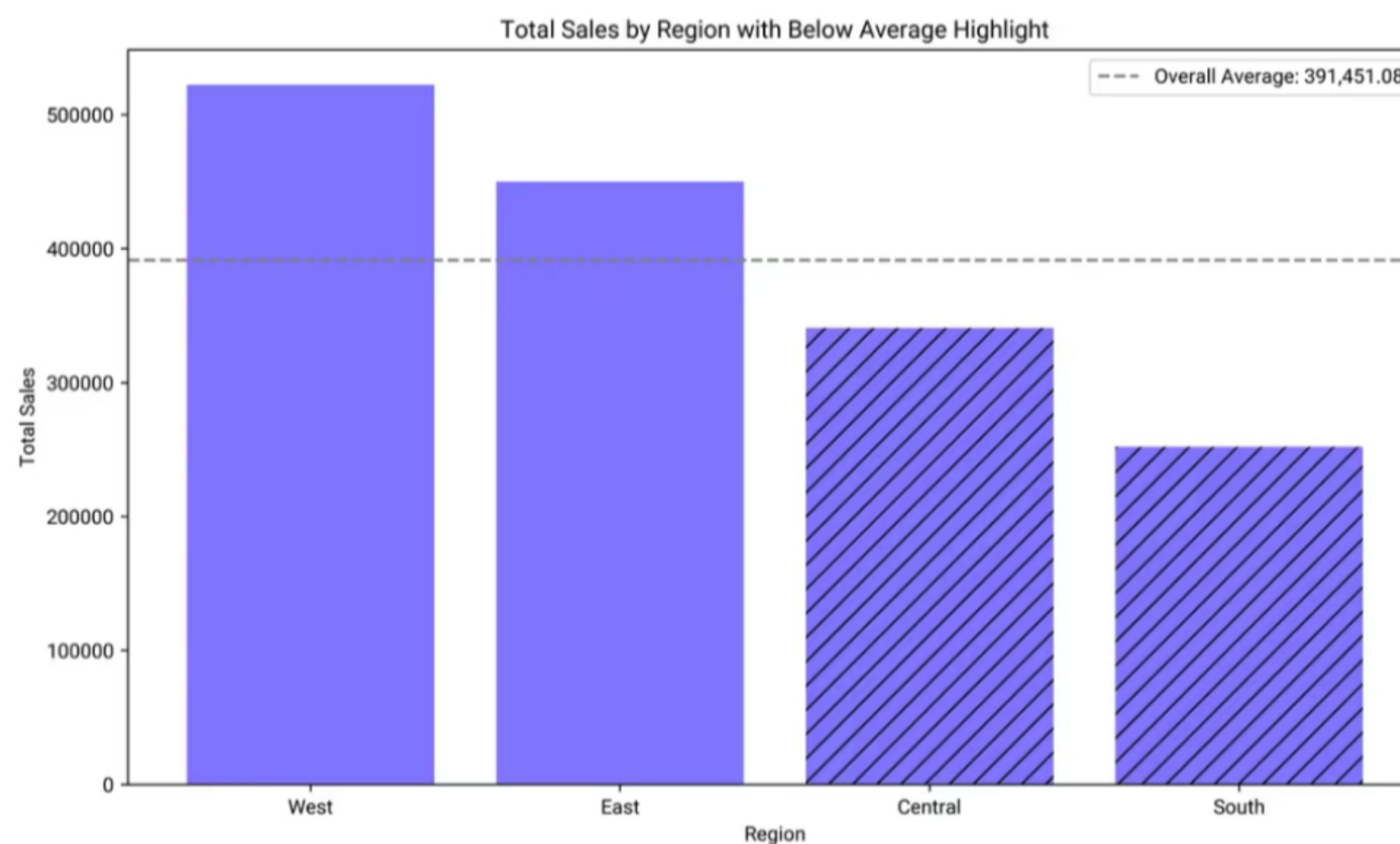
Technology

- Total Sales: 470,587.99 USD
- Total Profit: 90,458.25 USD
- Profit Margin: 19.22%

Conclusion and Insights

- **Profitability by Category:** The **Technology** category demonstrates the highest profitability with a **profit margin of 19.22%**, despite not having the highest total sales.
- **Sales Performance:** **Office Supplies** generated the highest total sales at **643,707.69 USD**, followed by Technology and Furniture.
- **Profit Contribution:** The **Technology** category contributed the most to total profit at **90,458.25 USD**, significantly outperforming Furniture and Office Supplies in terms of profit generated.
- **Underperforming Category:** **Furniture** shows the lowest profit margin at **2.22%** and the lowest total profit at **10,006.61 USD**, indicating it is the least profitable category relative to its sales volume.

Q3.What is the total sales by region? Highlight regions with sales below the average.



Total Sales by Region

- **West:** The West region recorded the **highest total sales** at **522,441.05**.
- **East:** The East region followed with total sales of **450,234.67**.
- **Central:** The Central region had total sales of **341,007.52**.
- **South:** The South region reported the **lowest total sales** at **252,121.08**.

Overall Average Sales and Below Average Regions

- **Overall Average Sales:** The overall average sales across all regions was **391,451.08**.
- **Regions Below Average:**
 - **Central:** Total sales of 341,007.52 are **below the overall average**.
 - **South:** Total sales of 252,121.08 are **below the overall average**.
- **Regions Above Average:**
 - **West:** Total sales of 522,441.05 are **above the overall average**.
 - **East:** Total sales of 450,234.67 are **above the overall average**.

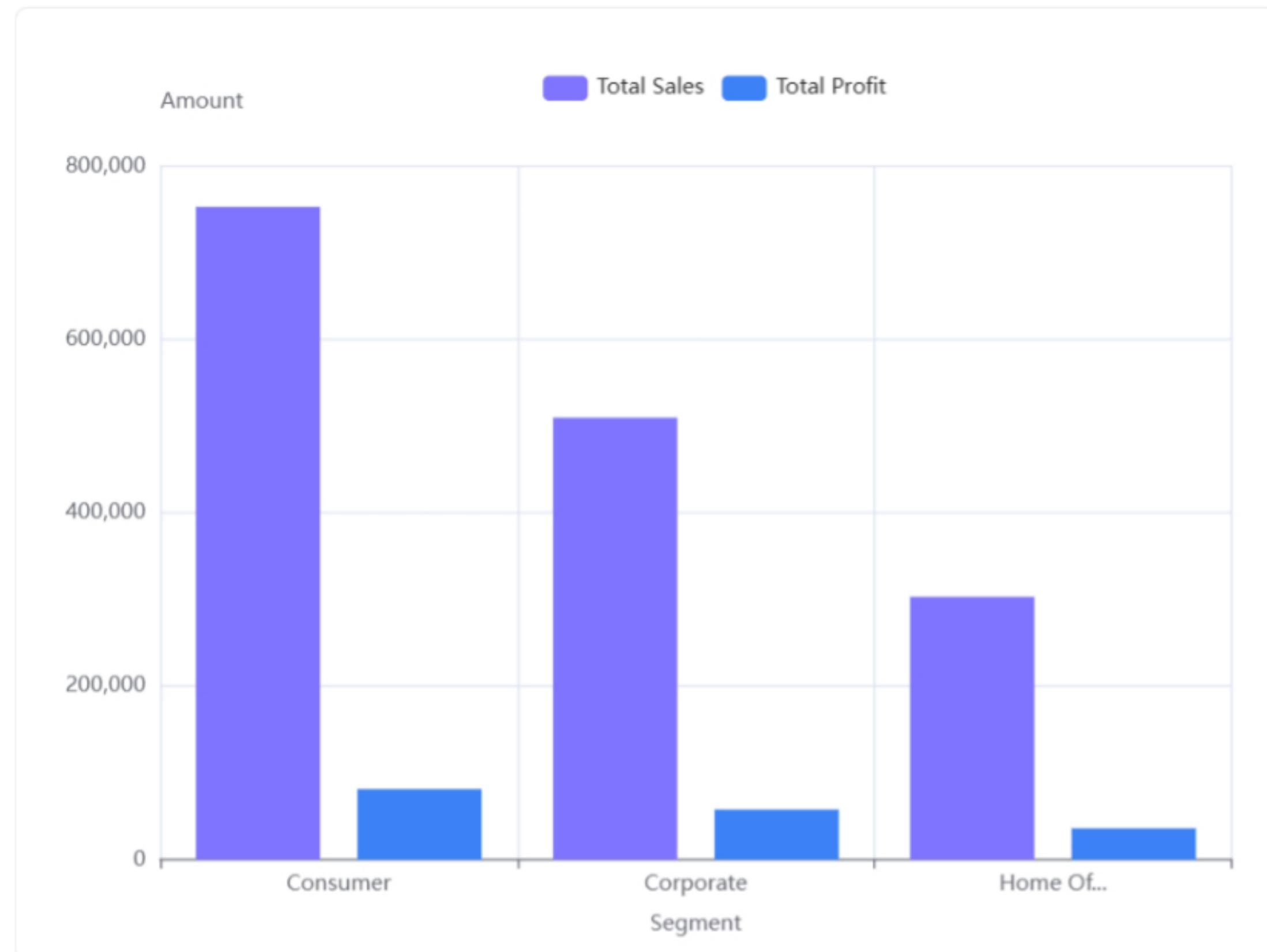
Visual Representation of Sales Performance

- **Bar Chart Visualization:** A bar chart effectively displays the total sales for each region.
- **Highlighting Below Average Regions:** Regions with sales below the overall average (Central and South) are visually highlighted with a **striped pattern** on their bars, while regions above average (West and East) have solid bars.
- **Average Sales Line:** A dashed grey line across the chart indicates the **overall average sales** value (391,451.08), providing a clear benchmark for comparison.

Conclusion and Insights

- **Sales Disparity:** There is a significant disparity in sales performance across regions, with the West and East regions performing considerably better than the Central and South regions.
- **Underperforming Regions:** The **Central and South regions** are identified as underperforming, as their total sales fall below the overall average. This suggests potential areas for strategic focus to improve sales.
- **Performance Benchmark:** The overall average sales serve as a crucial benchmark, allowing for quick identification of regions that are excelling or lagging in sales performance.

Q4.What is the sales distribution across different customer segments? Show the profit contribution by each segment.



Total Sales per Segment

- **Consumer Segment:** The Consumer segment recorded the **highest total sales** at **753,002.13**.
- **Corporate Segment:** The Corporate segment had total sales of **509,743.13**.
- **Home Office Segment:** The Home Office segment had the **lowest total sales** at **303,059.07**.

Profit Contribution by Segment

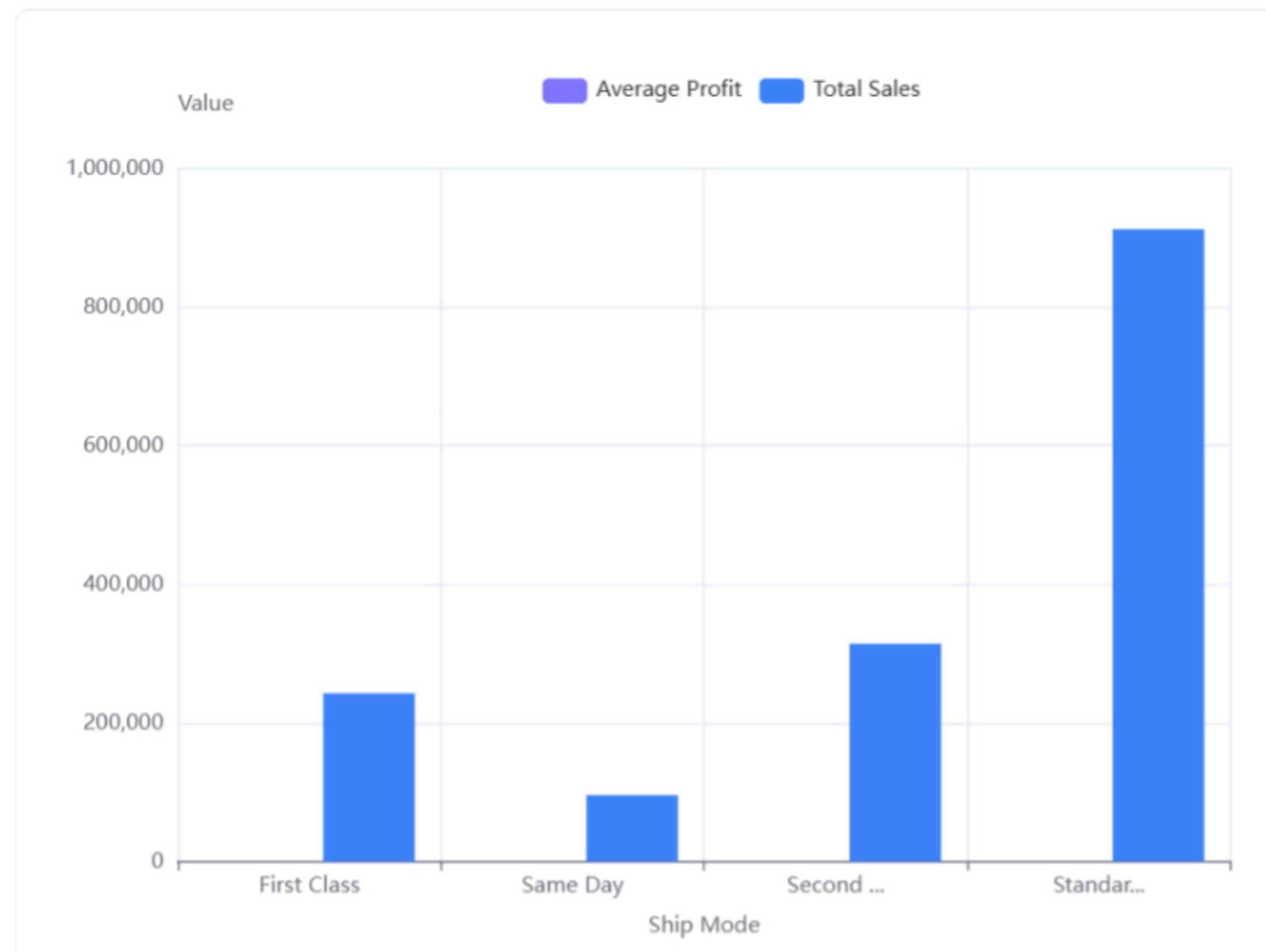
Total Profit per Segment

- **Consumer Segment:** The Consumer segment generated the **highest total profit** of **81,338.59**.
- **Corporate Segment:** The Corporate segment contributed **57,805.80** in total profit.
- **Home Office Segment:** The Home Office segment yielded the **lowest total profit** at **36,117.72**.

Conclusion and Insights

- **Dominance of Consumer Segment:** The **Consumer segment** is the **leading contributor** to both sales and profit, significantly outperforming the Corporate and Home Office segments in both metrics.
- **Sales and Profit Correlation:** There is a **direct correlation** between sales volume and profit generation across segments; segments with higher sales generally yield higher profits.
- **Segment Performance Hierarchy:** The segments can be ranked in descending order of both sales and profit as **Consumer > Corporate > Home Office**, indicating a consistent performance trend across the customer base.

Q5.What is the average profit per order for each shipping mode? Compare the sales volume for each shipping mode.



Average Profit per Order by Shipping Mode

- **Second Class:** This mode has the **highest average profit per order** at **\$32.20**.
- **First Class:** The average profit per order for this mode is **\$31.02**.
- **Standard Class:** This mode has an average profit per order of **\$28.91**.
- **Same Day:** This mode shows the **lowest average profit per order** at **\$25.61**.

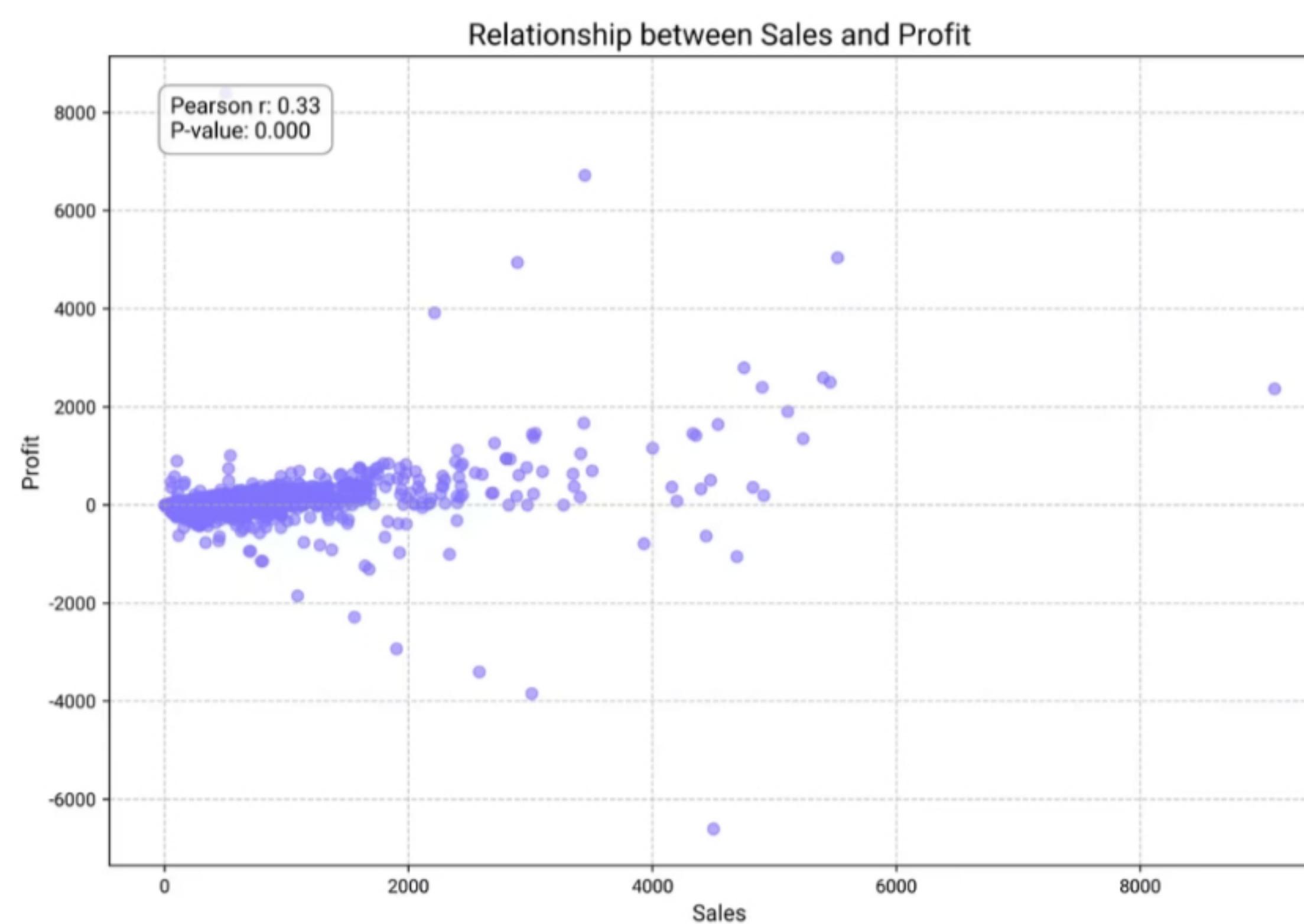
Sales Volume Comparison by Shipping Mode

- **Standard Class:** This mode accounts for the **overwhelming majority of total sales**, with a volume of **\$912,401.04**. This is significantly higher than any other shipping mode.
- **Second Class:** This mode has the second-highest total sales volume at **\$314,508.06**.
- **First Class:** The total sales for this mode are **\$242,936.72**.
- **Same Day:** This mode has the **lowest total sales volume** at **\$95,958.50**.

Conclusion and Insights

- **Dominance of Standard Class:** While "Standard Class" has a moderate average profit per order, its **exceptionally high total sales volume** makes it the most significant contributor to overall revenue. This suggests that a large number of orders are shipped via this mode, despite it not being the most profitable per order.
- **Profitability vs. Volume Trade-off:** "Second Class" shipping offers the **highest average profit per order**, indicating its efficiency in generating profit on a per-order basis. However, its total sales volume is considerably lower than "Standard Class," suggesting fewer orders utilize this mode.
- **Lowest Performers:** "Same Day" shipping consistently performs the lowest in both metrics, having the **lowest average profit per order** and the **lowest total sales volume**. This indicates it is neither highly profitable nor widely used.
- **Strategic Implications:** Businesses might consider strategies to encourage the use of "Second Class" shipping to boost overall profitability, while also understanding the critical role "Standard Class" plays in driving sales volume. Analyzing the reasons for the high volume in "Standard Class" and low volume in "Second Class" could provide further actionable insights.

Q6.What is the correlation coefficient between sales and profit? Is the correlation statistically significant?



Pearson Correlation Coefficient

- **Correlation Coefficient:** The Pearson correlation coefficient between Sales and Profit is **0.326**. This indicates a **weak positive linear relationship** between Sales and Profit.

Statistical Significance

- **P-value:** The calculated p-value is **4.537e-146**, which is extremely small.
- **Significance Level:** With an alpha level of 0.05, the correlation is determined to be **Statistically Significant**. This is because the p-value (4.537e-146) is much less than the alpha level (0.05).

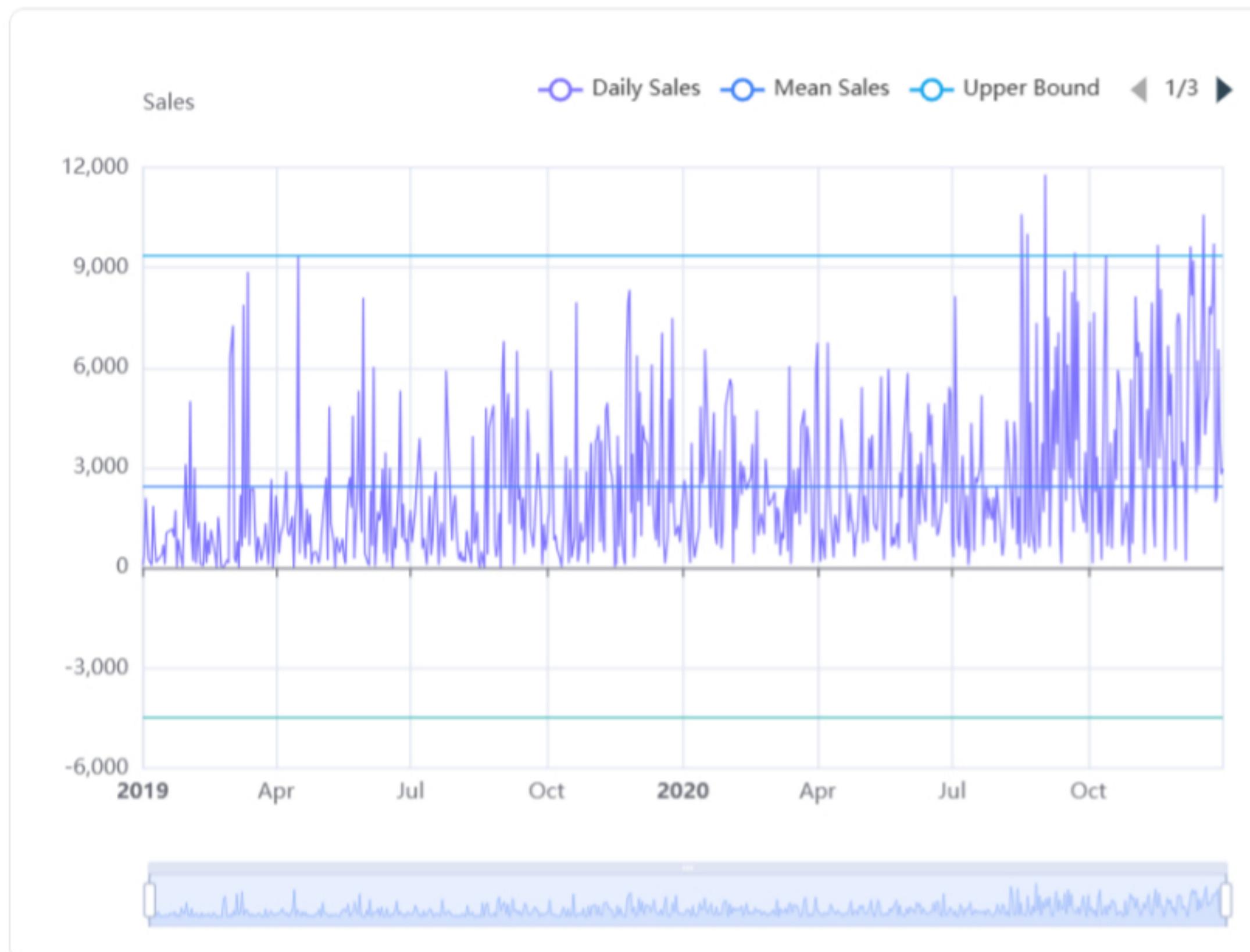
Visual Representation

- **Scatter Plot Annotation:** The scatter plot visually confirms the calculated Pearson r of 0.33 and a P-value of 0.000 (rounded from the extremely small actual value), reinforcing the statistical findings.
- **Data Distribution:** The plot shows a general trend where higher sales tend to be associated with higher profits, although there is significant scatter, especially at lower sales values, and instances of negative profit even with considerable sales.

Conclusion and Insights

- **Correlation Strength:** There is a **weak positive correlation** between Sales and Profit, meaning that as sales increase, profit tends to increase, but not in a strong or perfectly linear fashion.
- **Statistical Certainty:** Despite the correlation being weak, it is **highly statistically significant**. This implies that the observed positive relationship is unlikely to have occurred by random chance, suggesting a genuine underlying connection between sales and profit in the dataset.
- **Practical Implication:** While statistically significant, the weak correlation coefficient (0.326) suggests that sales alone are not a strong predictor of profit. Other factors likely play a substantial role in determining profitability, and a high volume of sales does not guarantee high profits, as evidenced by instances of negative profit even with high sales values.

Q7.Are there any outlier sales values? Identify anomalies in the daily sales trend using a 3-sigma rule.



Anomaly Detection Parameters

- **Mean Daily Sales:** The average daily sales calculated across all recorded days is approximately **2,435.15**.
- **Standard Deviation of Daily Sales:** The variability in daily sales is significant, with a standard deviation of approximately **2,305.86**.
- **Upper Anomaly Bound (Mean + 3*Std):** Sales exceeding **9,352.75** are considered unusually high.
- **Lower Anomaly Bound (Mean - 3*Std):** Sales falling below **-4,482.44** are considered unusually low. Given that sales values cannot be negative, this lower bound effectively means any positive sales value is above the lower anomaly threshold.

Identified Anomalies

- **Presence of Anomalies:** Yes, there are outlier sales values identified as anomalies.
- **Nature of Anomalies:** All identified anomalies are **high sales events**, meaning daily sales significantly exceeded the calculated upper bound. There are no anomalies detected below the lower bound, as the lower bound is negative and all sales are positive.
- **Visual Confirmation:** The "Daily Sales Trend with Anomaly Detection" chart clearly shows several **red diamond markers** indicating these anomalous daily sales points, all of which are positioned above the orange "Upper Bound" line. These anomalous sales values appear to be concentrated at the higher end of the sales spectrum, some even exceeding 11,000.

Conclusion and Insights

- **Outlier Sales Values Exist:** Based on the 3-sigma rule, there are indeed **outlier sales values** in the daily sales trend. These outliers represent days with exceptionally high sales figures.
- **High Sales Anomalies:** All detected anomalies are instances of **unusually high daily sales**, specifically those exceeding the calculated upper bound of approximately **9,352.75**. The maximum daily sales observed was **11,778.85**, which is well above this threshold.
- **No Low Sales Anomalies:** Due to the nature of sales data (non-negative values) and the calculated negative lower bound of **-4,482.44**, no days were flagged for unusually low sales.
- **Significance of Anomalies:** These high sales anomalies could indicate significant events such as successful promotional campaigns, seasonal peaks, or large individual orders, warranting further investigation to understand their root causes and potential for replication.

Q8.What is the return rate for each product category? Which category has the highest return rate?



Return Rate Details

- **Furniture:** The return rate for the Furniture category is **0.0%**.
- **Office Supplies:** The return rate for the Office Supplies category is **0.0%**.
- **Technology:** The return rate for the Technology category is **0.0%**.

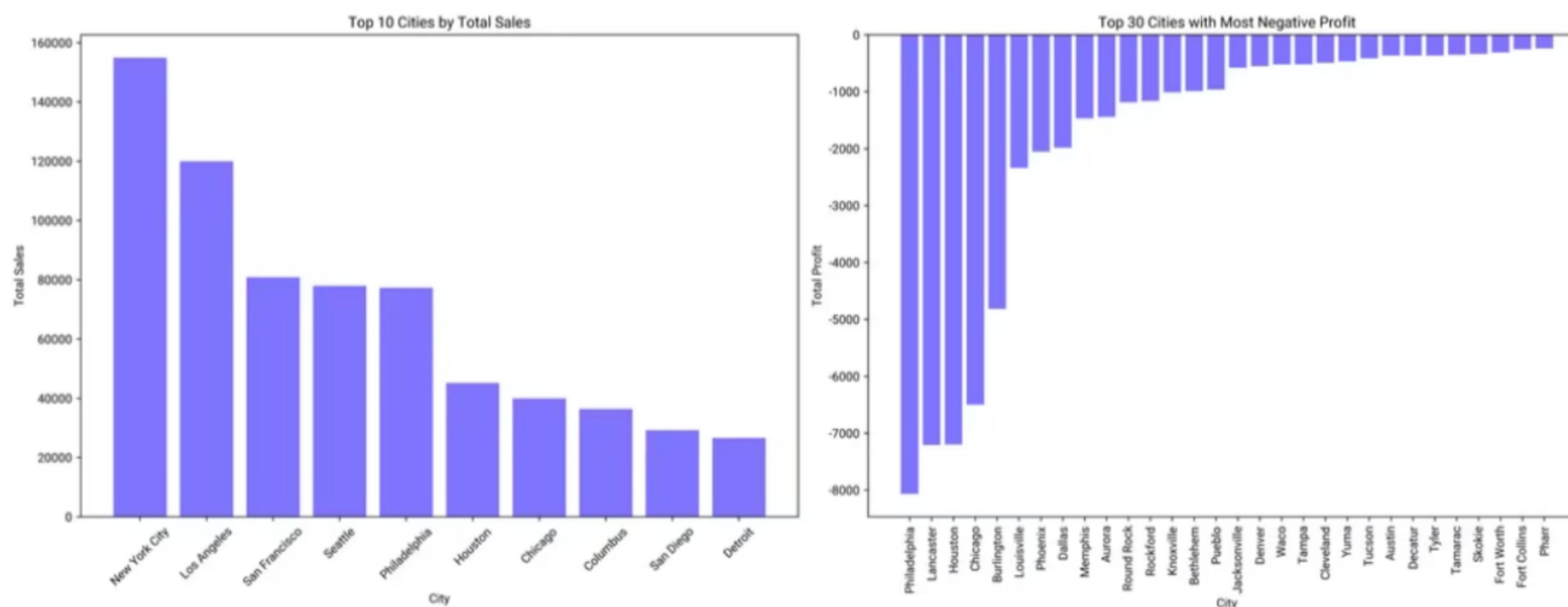
Visualization of Return Rates

- The bar chart visually confirms that all product categories (Furniture, Office Supplies, and Technology) have a **0% return rate**.

Conclusion and Insights

- Uniform Return Rates:** All product categories, namely **Furniture, Office Supplies, and Technology**, exhibit a **0.0% return rate**.
- Highest Return Rate:** Since all categories have a 0.0% return rate, **no single category has the highest return rate**; they all share the same, lowest possible return rate.
- Data Anomaly/Implication:** The consistent 0% return rate across all categories suggests either an **exceptionally low return volume** for the analyzed period or a potential **data anomaly** where return information might not be fully captured or correctly interpreted (e.g., the 'Returns' column might not accurately reflect returns or there were no returns in the dataset). Further investigation into the raw data or the definition of 'Returns' might be necessary if this result is unexpected.

Q9.What are the top 10 cities by sales? Identify cities with negative profit margins.



Top 10 Cities by Total Sales

- New York City** : Generated the highest total sales at **\$154,919.70**.
- Los Angeles** : Ranked second with total sales of **\$120,017.44**.
- San Francisco** : Followed with total sales of **\$80,891.42**.
- Seattle** : Recorded total sales of **\$77,924.54**.
- Philadelphia** : Had total sales of **\$77,305.44**.
- Houston** : Total sales amounted to **\$45,159.51**.
- Chicago** : Generated total sales of **\$39,993.16**.
- Columbus** : Total sales were **\$36,398.89**.
- San Diego** : Recorded total sales of **\$29,269.55**.
- Detroit** : The tenth city on the list, with total sales of **\$26,631.21**.

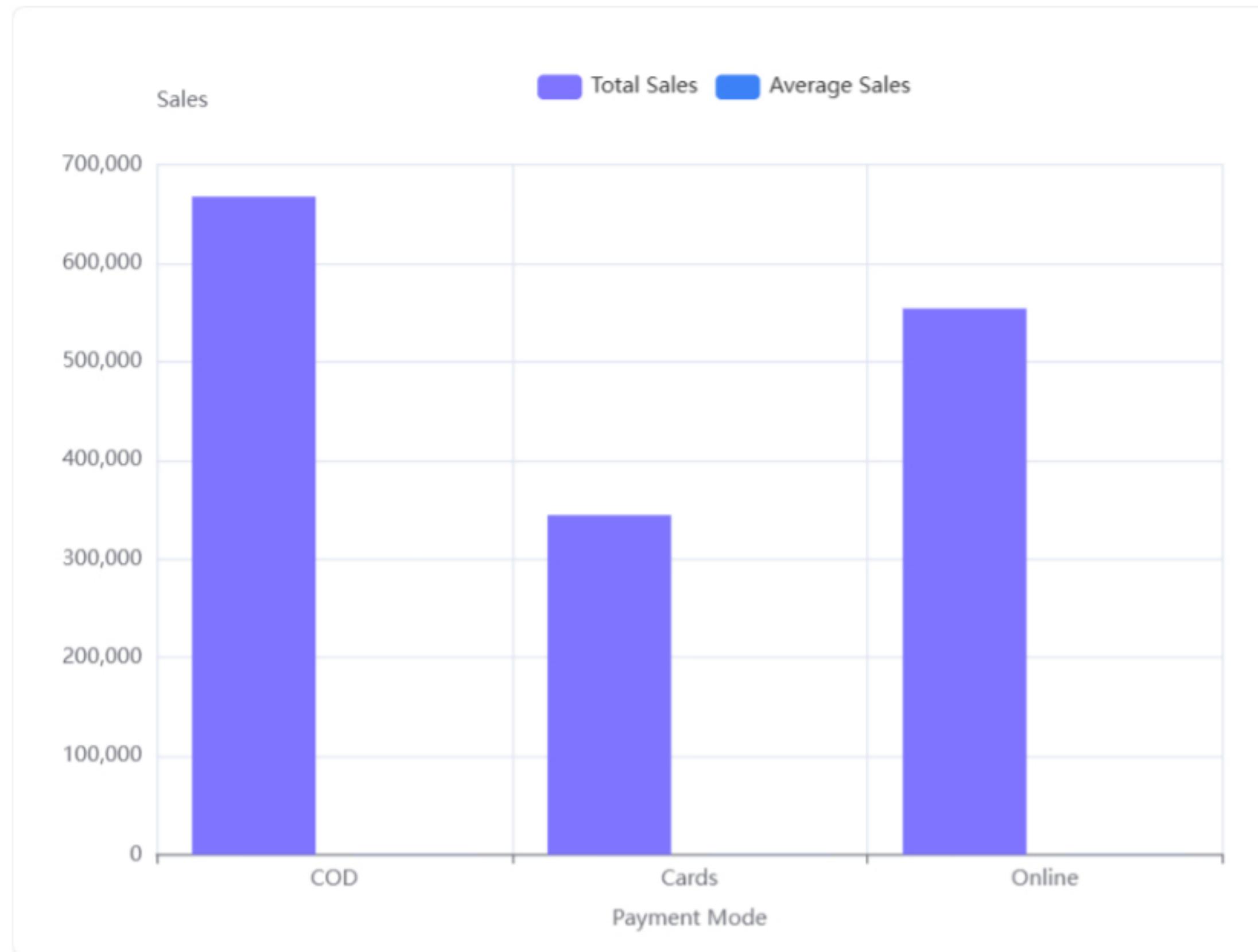
Cities with Negative Profit Margins

- Total Count:** A significant number of cities, **93 in total**, experienced negative profit margins.
- Most Significant Losses:**
 - Philadelphia:** Incurred the largest loss with a total profit of **-\$8,065.97**.
 - Lancaster:** Followed with a substantial loss of **-\$7,207.38**.
 - Houston:** Also experienced a considerable loss of **-\$7,197.72**.
 - Chicago:** Recorded a loss of **-\$6,497.36**.
 - Burlington:** Showed a negative profit of **-\$4,814.80**.
- Range of Losses:** The negative profit values ranged from a minimum loss of **-\$0.24** (Highland Park) to a maximum loss of **-\$8,065.97** (Philadelphia).

Conclusion and Insights

- Sales Concentration:** The top 10 cities by sales demonstrate a strong concentration of revenue, with New York City and Los Angeles significantly leading the sales figures. This indicates that a large portion of the company's sales are generated from a few key urban centers.
- Profitability Challenges:** Despite strong sales in some areas, a substantial number of cities (93) are operating at a loss. Notably, **Philadelphia, Houston, and Chicago** appear in both the top 10 sales list and the list of cities with the most significant negative profits. This highlights a critical issue where high sales volume does not necessarily translate to profitability, suggesting potential problems with cost management, pricing strategies, or product mix in these specific high-volume, low-profit cities.
- Strategic Focus:** To improve overall profitability, it is crucial to investigate the underlying reasons for negative profits in the identified cities, especially those with high sales volume. This could involve analyzing operational costs, product pricing, and competitive landscapes in these markets.

Q10.What is the distribution of sales by payment mode? What is the average order value for each payment mode?



Sales Distribution by Payment Mode

- **COD (Cash on Delivery):** Generated the **highest total sales** at **667,417.75**.
- **Online:** Accounted for the **second highest total sales** with **553,993.46**.
- **Cards:** Had the **lowest total sales** among the payment modes, amounting to **344,393.11**.

Average Order Value by Payment Mode

- **COD (Cash on Delivery):** Showed the **highest average sales value** per transaction at **272.08**.
- **Cards:** Had an average sales value of **268.22**.
- **Online:** Recorded the **lowest average sales value** per transaction at **256.00**.

Conclusion and Insights

- **Dominance of COD in Total Sales:** COD is the **leading payment mode in terms of total sales**, significantly outperforming both Online and Cards. This suggests a strong customer preference or reliance on cash payments for larger overall transaction volumes.
- **Highest Average Order Value for COD:** Not only does COD generate the most total sales, but it also has the **highest average order value**, indicating that customers using COD tend to make larger individual purchases.
- **Online vs. Cards Performance:** While Online payments contribute more to total sales than Cards, the average order value for Online transactions is the lowest, suggesting a higher volume of smaller transactions. Conversely, Cards, despite lower total sales, have a slightly higher average order value than Online, implying fewer but potentially larger transactions compared to online payments.