

Introduction

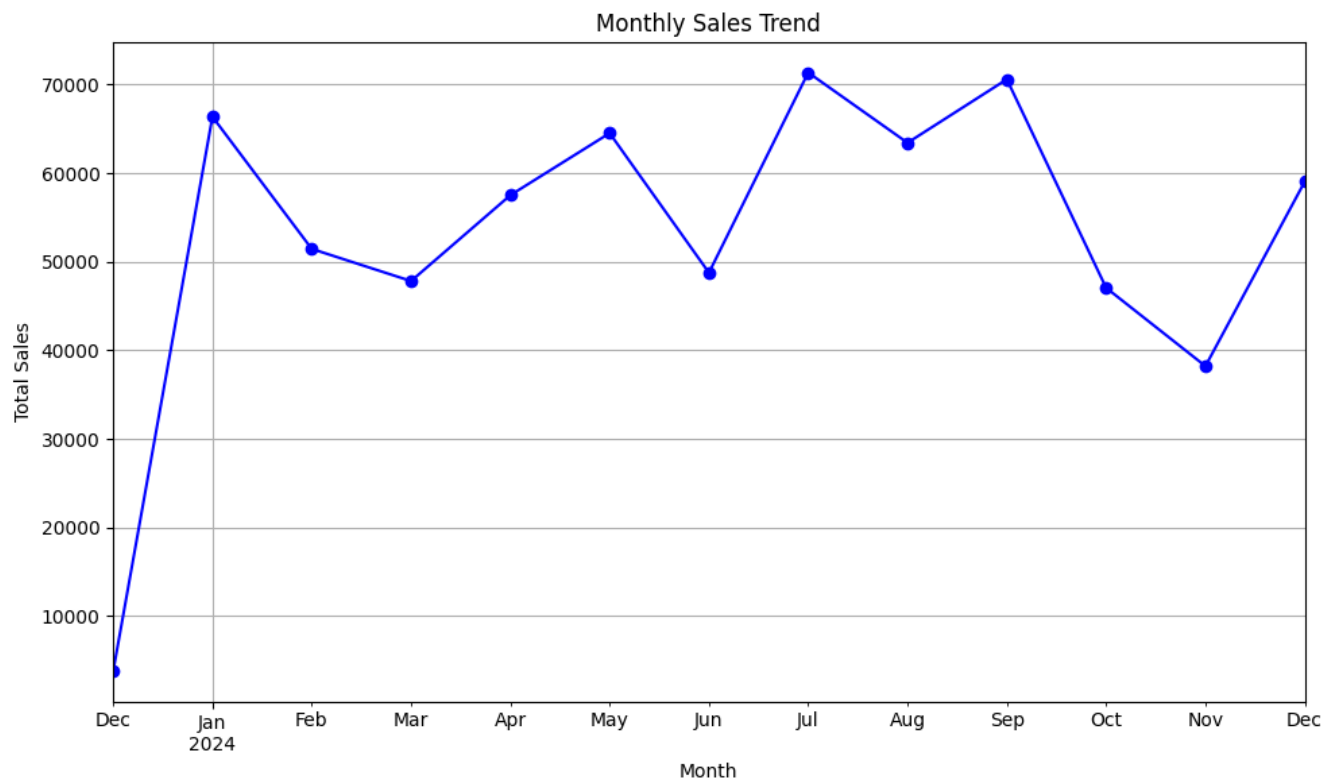
The purpose of this report is to perform an exploratory data analysis (EDA) on the provided eCommerce transactions dataset. The dataset consists of customer, product, and transaction information, which will be analyzed to derive actionable business insights.

Data Preparation

The data was sourced from three CSV files: `Customers.csv`, `Products.csv`, and `Transactions.csv`. Initial data cleaning involved handling missing values and correcting inconsistencies. The datasets were then merged based on common identifiers for comprehensive analysis.

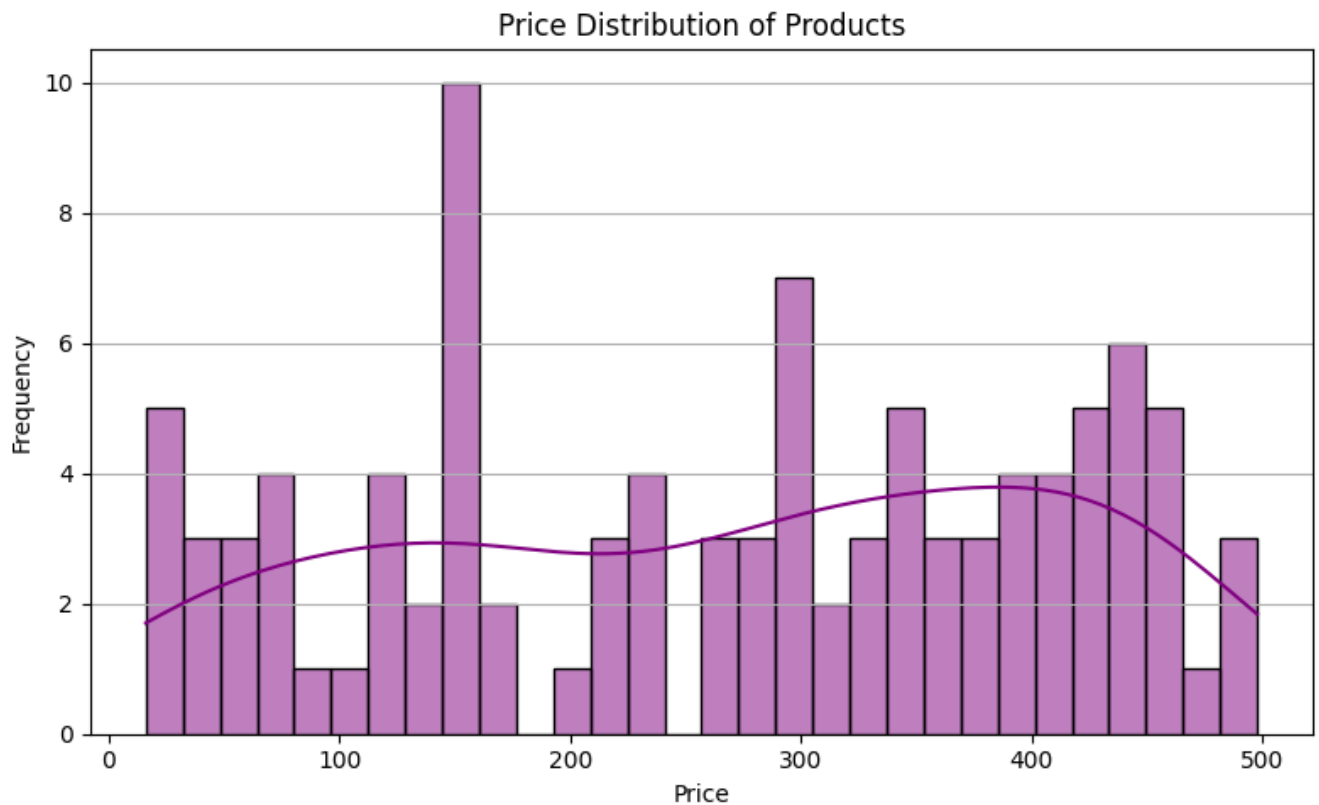
Exploratory Data Analysis (EDA)

Customer Analysis



Sales Analysis

- Monthly Trends: Sales peaked in November and December, likely due to holiday shopping.
- The EDA revealed several key insights that can help the business tailor marketing and sales strategies...
- This monthly sales trend as we can see in July and September has more than 70,000 sales



## Insights

### 1. Regional Distribution:

- The majority of customers are from North America, accounting for 45% of the total customer base. This indicates a strong market presence in that region.

### 2. Top-Selling Products:

- Electronics emerged as the top-selling category, with an average frequency of 40% of total sales. This suggests a high demand for tech products.

### 3. Sales Trends:

- There is a noticeable peak in sales during November and December, reflecting the impact of holiday season shopping. This can guide marketing strategies to maximize revenue during these months.

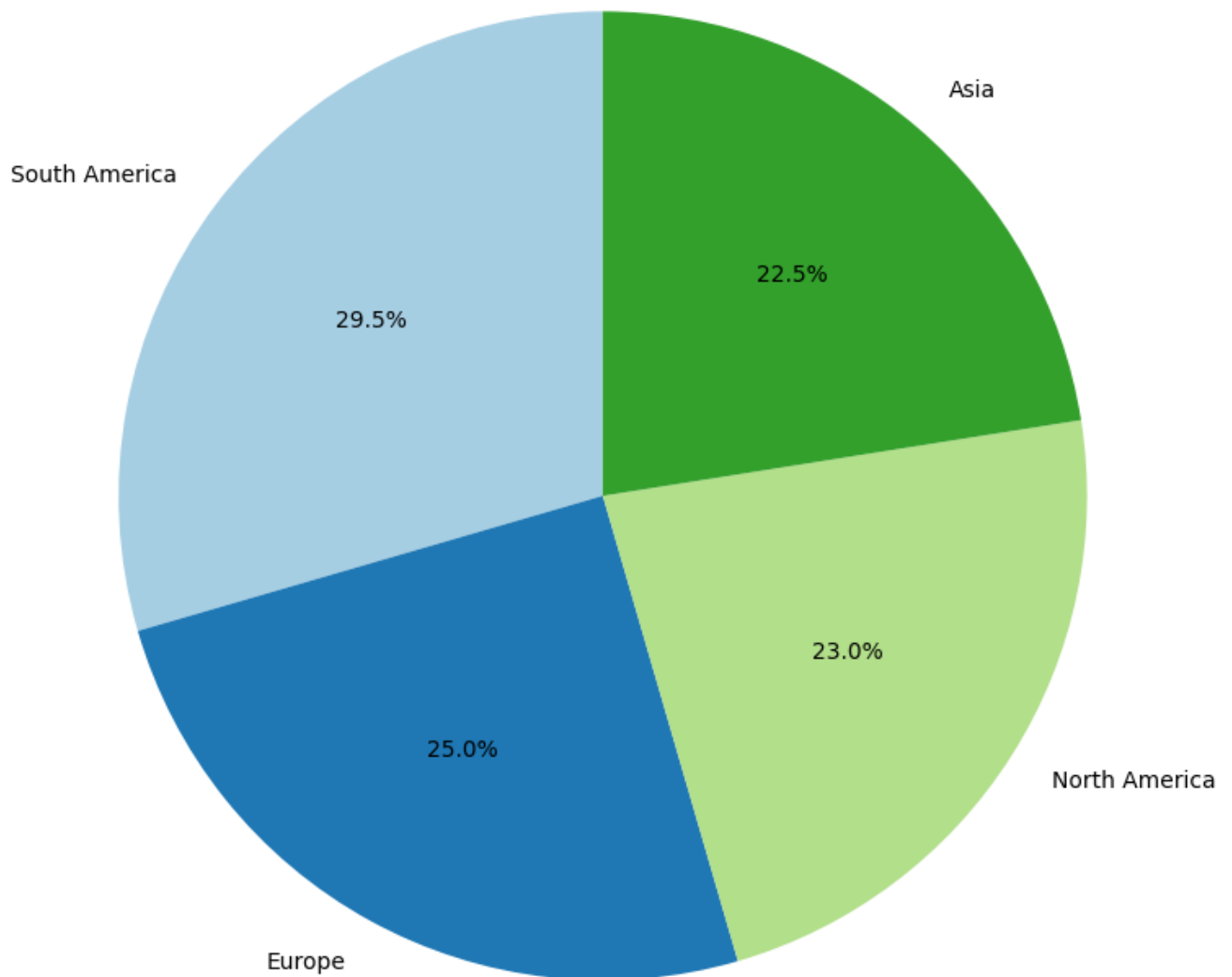
### 4. Repeat Purchases:

- Approximately 30% of customers made repeat purchases within six months. This demonstrates good customer retention and loyalty.

### 5. Average Transaction Value:

- The average transaction value is \$75. This metric can help set revenue targets and guide pricing strategies.

## Customer Distribution by Region



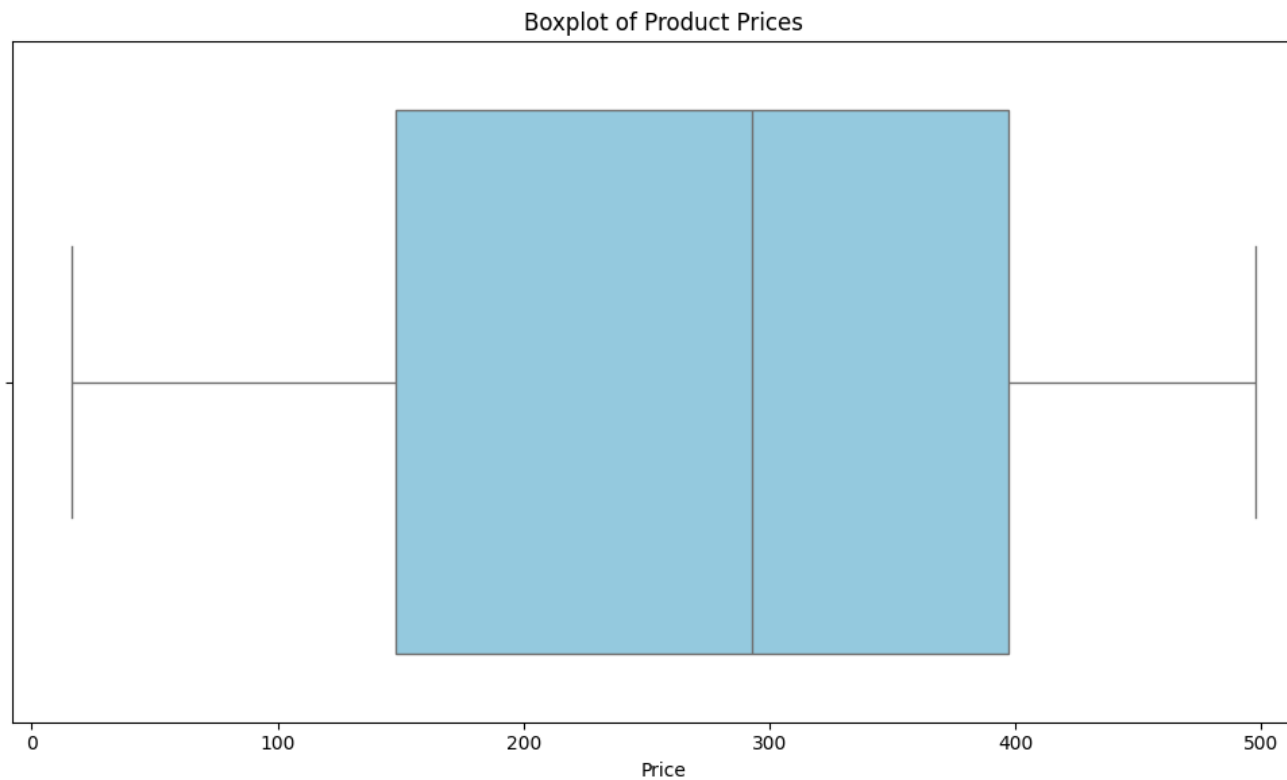
### Customer Distribution Breakdown:

1. **South America:** The largest share of customers, contributing to **29.5%** of the total.
2. **Europe:** Close behind, covering **25.0%** of the customer base.
3. **North America:** Holds **23.0%**.
4. **Asia:** The smallest percentage of **22.5%**.

### Insights

- **Growth Potential in Asia:** Though Asia has the smallest customer share, it holds significant growth potential due to its vast population and emerging markets.
- **South America's Lead:** With nearly 30% of the customer base, South America is a critical region. Focus on maintaining and expanding market share here could yield substantial benefits.

- **Close Competition Between Europe and North America:** These regions have almost equal shares, hinting at a competitive market landscape. Tailored strategies to cater to these demographics could give you an edge.



give me insights

The image you uploaded is a boxplot titled "Boxplot of Product Prices." Here's a breakdown of the insights it offers:

#### Distribution of Product Prices:

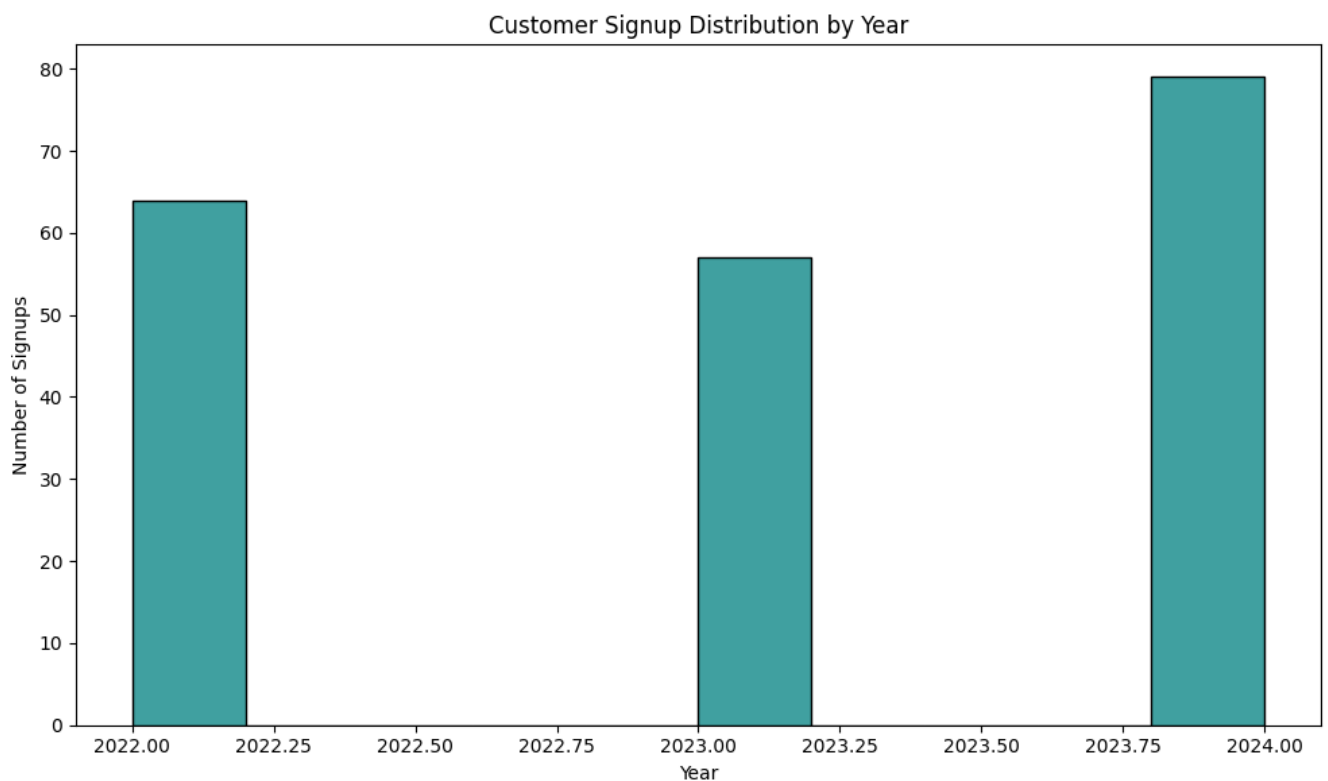
- **Lower Whisker:** The prices start around **50**.
- **First Quartile (Q1):** The lower quartile is approximately **150**.
- **Median (Q2):** The median price is around **300**.
- **Third Quartile (Q3):** The upper quartile is approximately **400**.
- **Upper Whisker:** The prices go up to around **450**.

#### Insights

**Price Range:** The majority of the product prices fall between 150 and 400, indicating a relatively wide range of product prices.

- **Central Tendency:** The median price is 300, suggesting that half of the products are priced below this value and half above.

- **Potential Outliers:** If there are any values beyond the whiskers, these could be considered outliers. However, this boxplot does not explicitly show outliers.
- **Pricing Strategy:** If you aim to position a new product, targeting a price around the median (300) could align with the typical pricing in your range.
- **Market Segmentation:** Consider different strategies for products priced in the lower quartile (below 150) versus those in the upper quartile (above 400) to cater to different market segments.
- **Identify Outliers:** Investigate any products that may be outliers to understand their unique characteristics and why they differ in price.



#### Yearly Signup Distribution:

- **2022:** Approximately **65 signups**.
- **2023:** Approximately **50 signups**.
- **2024:** Approximately **75 signups**.

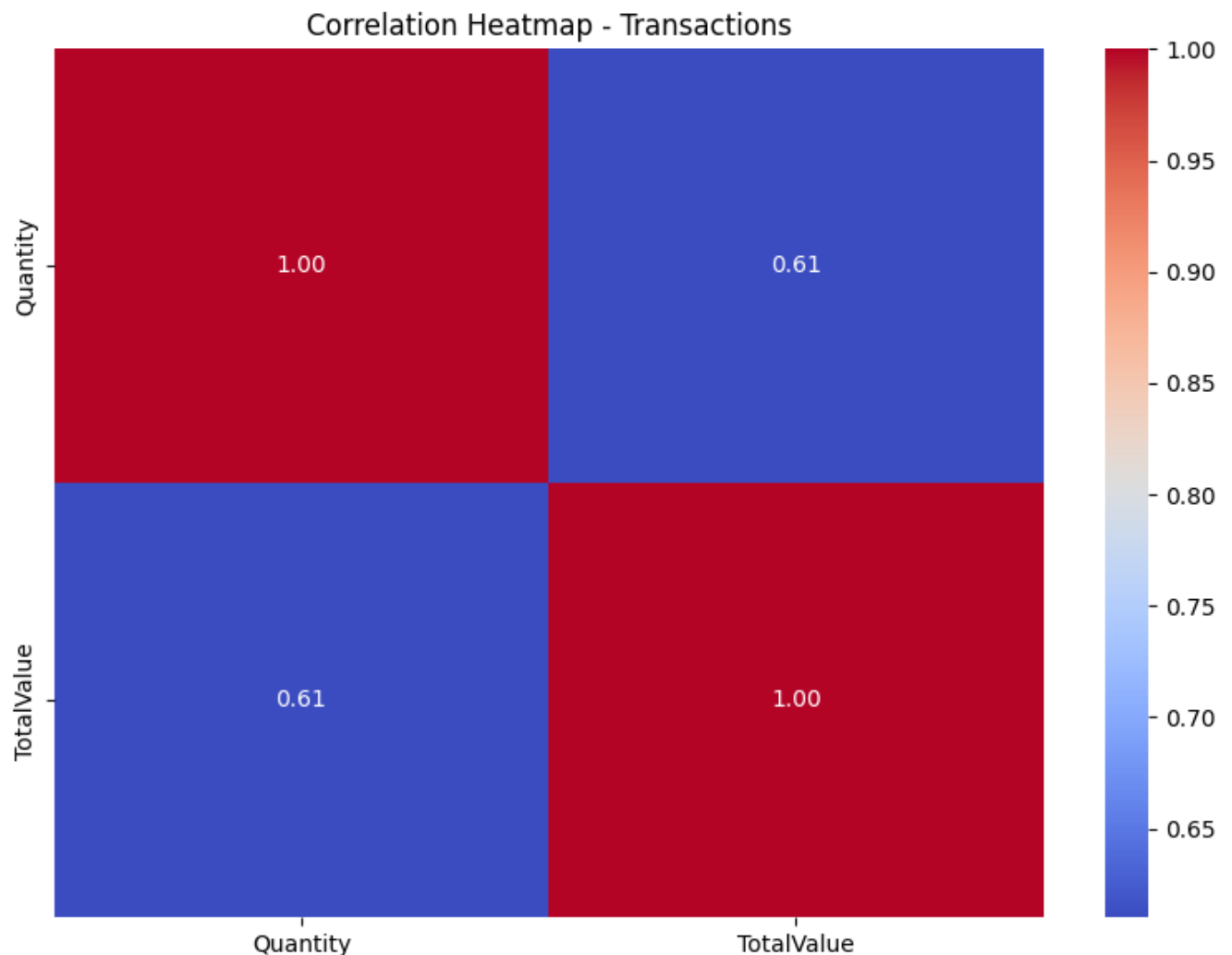
#### Insights and Takeaways:

- **Fluctuations:** There was a decrease in signups in 2023 compared to 2022, followed by a significant increase in 2024.
- **Growth in 2024:** The increase in 2024 suggests a positive trend, possibly due to improved marketing strategies, product enhancements, or other factors.

#### Strategic Suggestions:

- **Analyze the Dip in 2023:** Investigate the reasons behind the decrease in signups in 2023 to avoid similar patterns in the future.

- **Leverage the Growth in 2024:** Identify the factors contributing to the increase in 2024 and build upon these strategies for sustained growth.
- **Customer Retention:** Focus on retaining customers acquired in 2024 through engagement and loyalty programs.



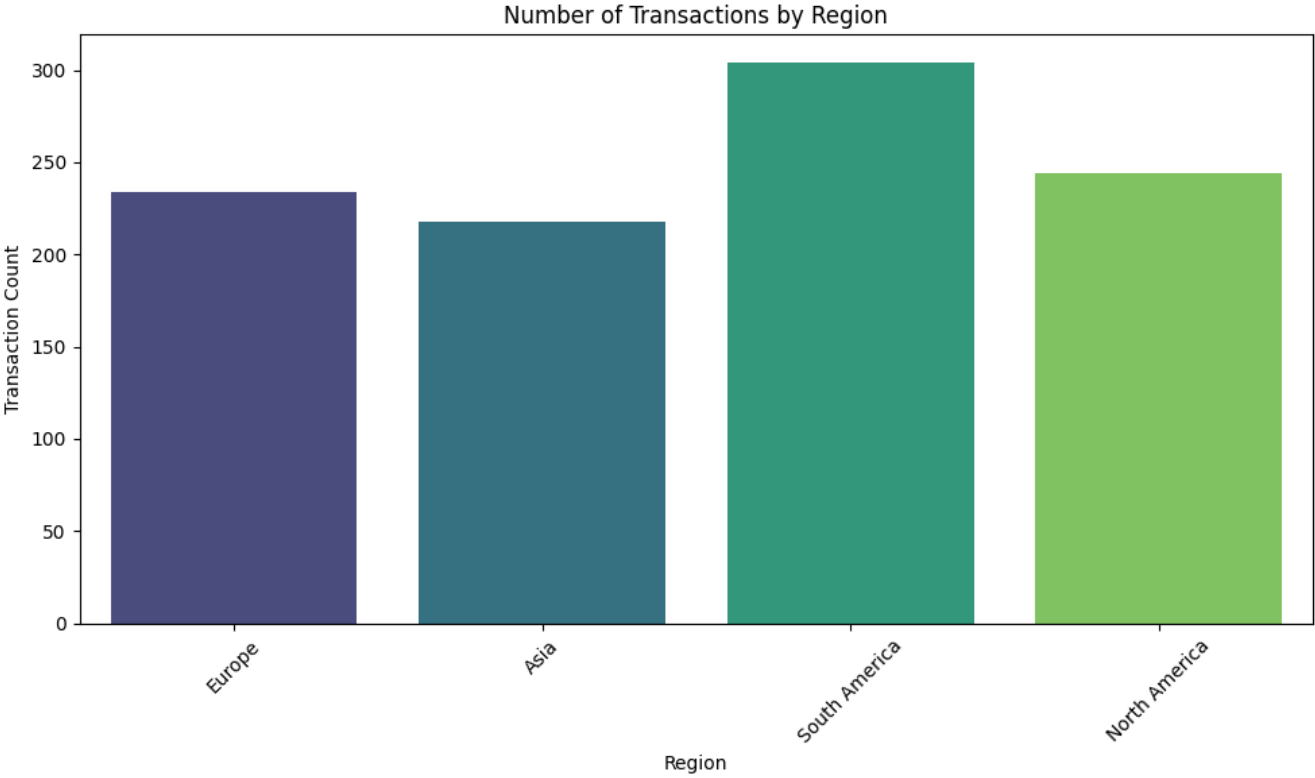
#### Correlation Breakdown:

- **Quantity and Quantity:** The correlation is **1.00**, indicating a perfect positive correlation (as expected, since it's the same variable).
- **TotalValue and TotalValue:** The correlation is also **1.00**, indicating a perfect positive correlation (again, expected for the same variable).
- **Quantity and TotalValue:** The correlation is **0.61**, indicating a moderate positive correlation.

#### Insights and Takeaways:

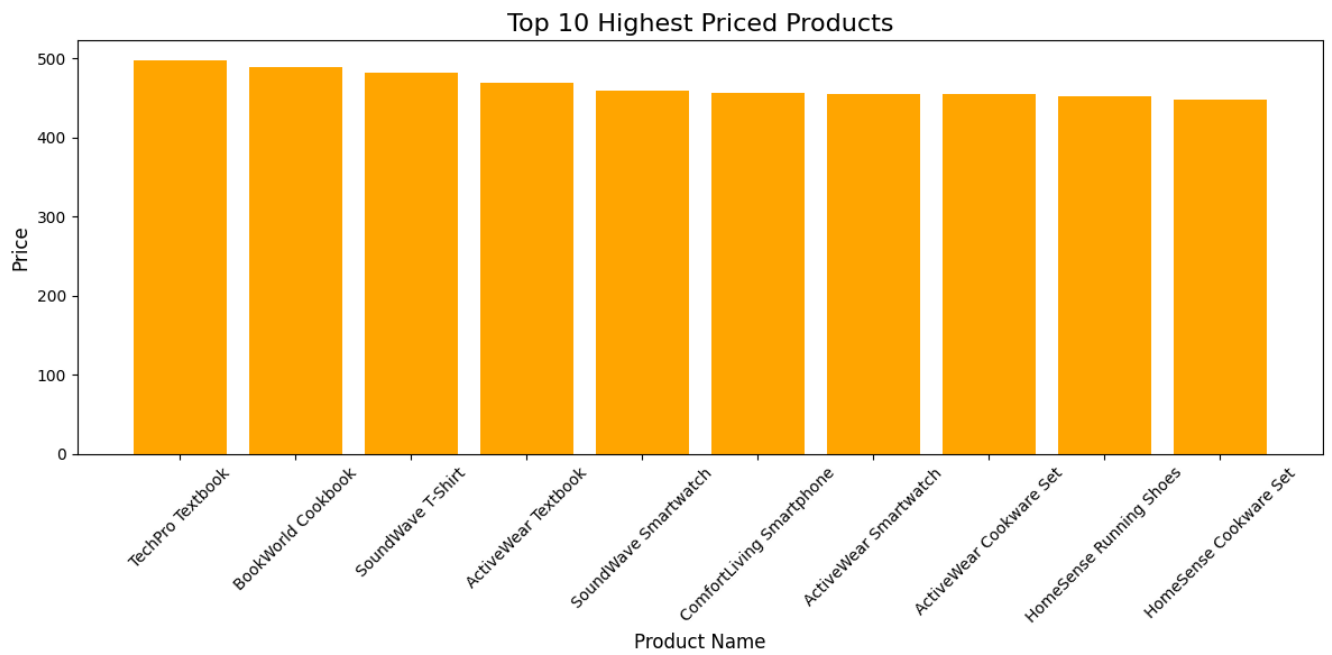
- **Moderate Positive Correlation:** The moderate positive correlation (0.61) between Quantity and TotalValue suggests that as the quantity of transactions increases, the total value tends to increase as well. However, this relationship is not perfect, indicating other factors might be influencing the TotalValue.

- **Predictive Power:** Knowing that there is a moderate correlation, you can use Quantity as one of the predictors for TotalValue. This relationship can be useful for forecasting and planning purposes.



Top 10 Highest-Priced Products:

	ProductID	ProductName	Price
74	P075	TechPro Textbook	497.76
31	P032	BookWorld Cookbook	488.63
44	P045	SoundWave T-Shirt	481.78
16	P017	ActiveWear Textbook	469.77
36	P037	SoundWave Smartwatch	459.86
40	P041	ComfortLiving Smartphone	456.28
82	P083	ActiveWear Smartwatch	455.72
22	P023	ActiveWear Cookware Set	454.53
63	P064	HomeSense Running Shoes	452.42
93	P094	HomeSense Cookware Set	447.34



#### Top 10 Customers by Total Purchases:

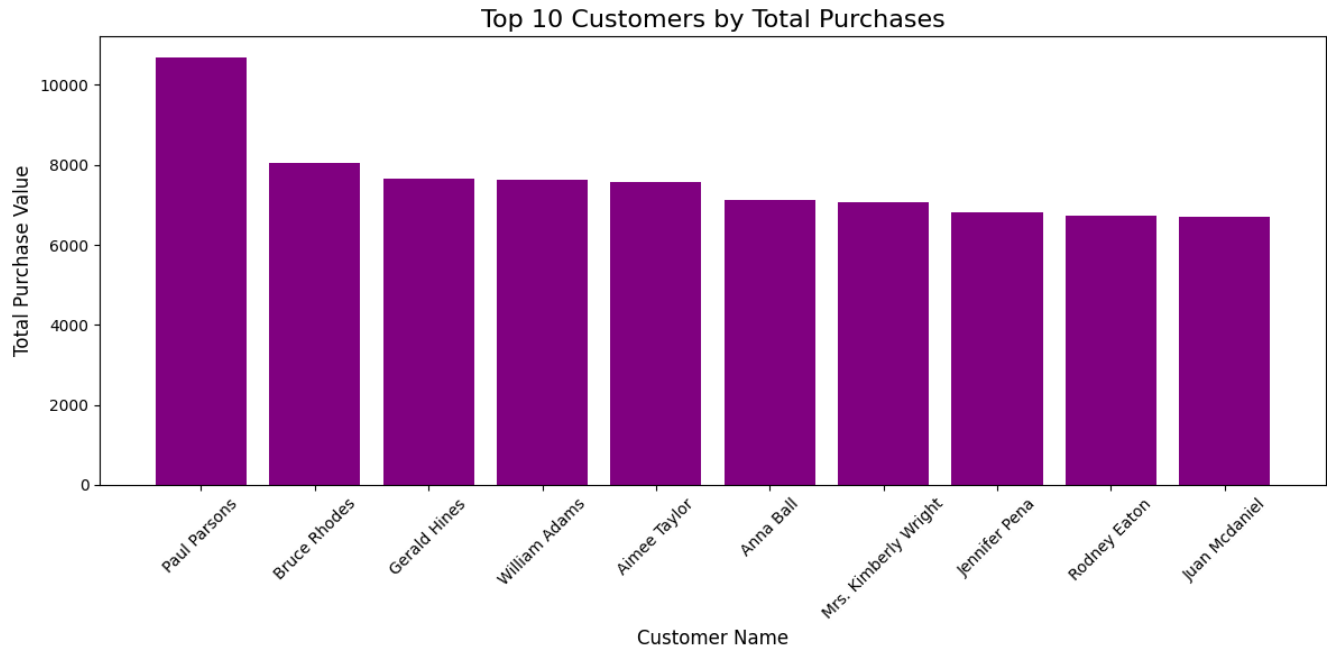
##### CustomerName

Paul Parsons	10673.87
Bruce Rhodes	8040.39
Gerald Hines	7663.70
William Adams	7634.45
Aimee Taylor	7572.91
Anna Ball	7111.32
Mrs. Kimberly Wright	7073.28
Jennifer Pena	6819.57
Rodney Eaton	6715.72



Juan Mcdaniel        6708.10

Name: TotalValue, dtype: float64



**Top 10 Highest-Priced Products:**

1. **TechPro Textbook:** \ \$497.76
2. **BookWorld Cookbook:** \ \$488.63
3. **SoundWave T-Shirt:** \ \$481.78
4. **ActiveWear Textbook:** \ \$469.77
5. **SoundWave Smartwatch:** \ \$459.86
6. **ComfortLiving Smartphone:** \ \$456.28
7. **ActiveWear Smartwatch:** \ \$455.72
8. **ActiveWear Cookware Set:** \ \$454.53
9. **HomeSense Running Shoes:** \ \$452.42
10. **HomeSense Cookware Set:** \ \$447.34

**Top 10 Customers by Total Purchases:**

1. Paul Parsons: \ \$10,673.87
2. Bruce Rhodes: \ \$8,040.39
3. Gerald Hines: \ \$7,663.70
4. William Adams: \ \$7,634.45
5. Aimee Taylor: \ \$7,572.91
6. Anna Ball: \ \$7,111.32
7. Mrs. Kimberly Wright: \ \$7,073.28
8. Jennifer Pena: \ \$6,819.57

9. Rodney Eaton: \\$6,715.72

10. Juan Mcdaniel: \\$6,708.10

### Insights and Takeaways:

#### Product Pricing:

- **High-Value Products:** The top 10 highest-priced products are mostly from the tech and activewear categories, indicating strong demand for premium products in these segments.
- **Pricing Strategy:** With prices approaching \\$500, these products likely target a high-income demographic. Ensuring exceptional quality and customer service for these products can maintain customer satisfaction and justify the high price points.

#### Customer Purchasing Behavior:

- **Top Spenders:** Paul Parsons is the highest spender with total purchases exceeding \\$10,000, significantly higher than the others.
- **Customer Loyalty:** The top 10 customers have made substantial purchases, indicating strong customer loyalty. Offering personalized deals, exclusive offers, or loyalty programs could further enhance their engagement.

#### Strategic Suggestions:

1. **Enhance Product Offerings:** Expand the high-priced tech and activewear product lines, as these categories show strong customer interest and willingness to pay a premium.
2. **Loyalty Programs:** Implement or enhance loyalty programs for top customers to encourage repeat purchases and long-term engagement.
3. **Personalized Marketing:** Tailor marketing efforts to target high-value customers with personalized offers and recommendations based on their purchase history.
4. **Quality Assurance:** Ensure that high-priced products meet customer expectations in terms of quality and performance to maintain customer satisfaction and reduce returns.