Business Report: eCommerce Transactions Dataset Analysis

1. Executive Summary

The eCommerce Transactions dataset was analyzed to extract actionable insights, focusing on customer behavior, product sales, and overall transaction patterns. This report outlines key business insights derived from the dataset and recommendations to improve customer engagement, product performance, and revenue generation.

2. Data Overview

- **Customers.csv:** Contains customer demographics and sign-up details.
- **Products.csv:** Lists product details, including category and price.
- **Transactions.csv:** Documents transactional information, including products purchased, quantity, and total value.

3. Key Insights after EDA

A. Data Quality Check
Upon inspecting the dataset, no missing values or duplicate entries were identified across all three files—Customers.csv, Products.csv, and Transactions.csv. This ensures the dataset is clean and ready for analysis without requiring additional preprocessing for data quality.

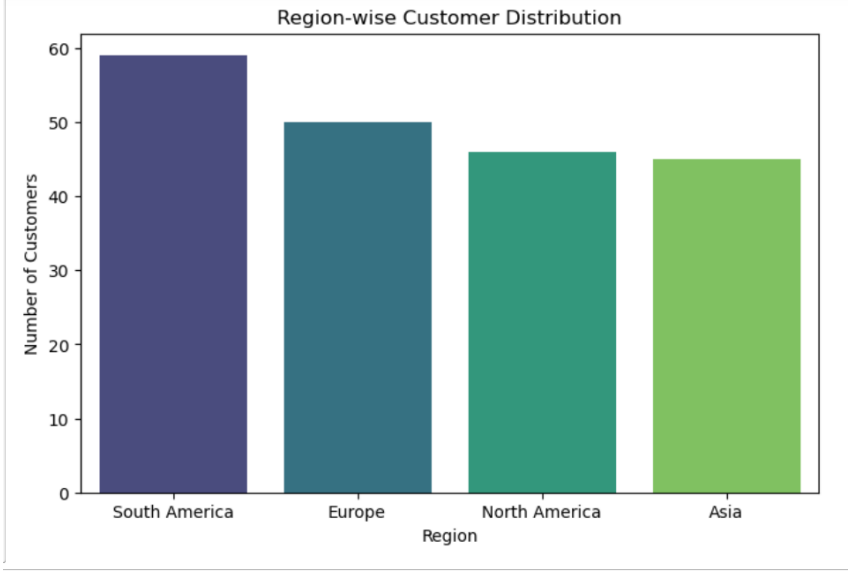
B. Data Preparation

1. **Date Conversion**: The SignupDate column from the **Customers.csv** file and the TransactionDate column from the **Transactions.csv** file were successfully converted to datetime format for accurate time-based analysis.
2. **Data Merging**:
    - The **Transactions.csv** file was merged with **Customers.csv** using the CustomerID column with a left join to combine transactional data with customer information.
    - The resulting dataset was then merged with **Products.csv** on the ProductID column using a left join, creating a comprehensive dataset that includes details on customers, products, and transactions.

C. Business Insights:

1. Region-Wise Customer Distribution
A bar plot was created to show the distribution of customers across different regions. This helps identify the most

significant regions in terms of customer base, which can guide targeted marketing and regional sales strategies.



Specific Observations:

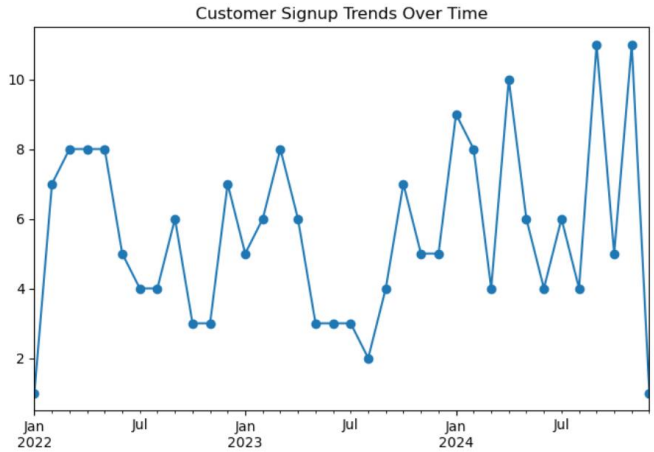
- **South America:** Has the highest number of customers with approximately 60.
- **Europe:** Follows with around 50 customers.
- **North America:** Has approximately 45 customers.
- **Asia:** Has the lowest number of customers with about 45

Inferences:

- The company has a significant customer base in South America.
- Europe and North America also have a considerable number of customers.
- Asia has the smallest customer base among the regions shown

2. Customer Signup Trends Over Time

A line plot was created to analyze the monthly trends of customer signups over time. This visualization highlights the fluctuations in signups, providing insights into periods of high customer acquisition, which can guide future marketing and customer engagement strategies.



Overall:

The graph shows how the number of customer signups has changed over time, specifically from January 2022 to July 2024. The x-axis represents the time period, and the y-axis indicates the number of signups.

Key Observations:

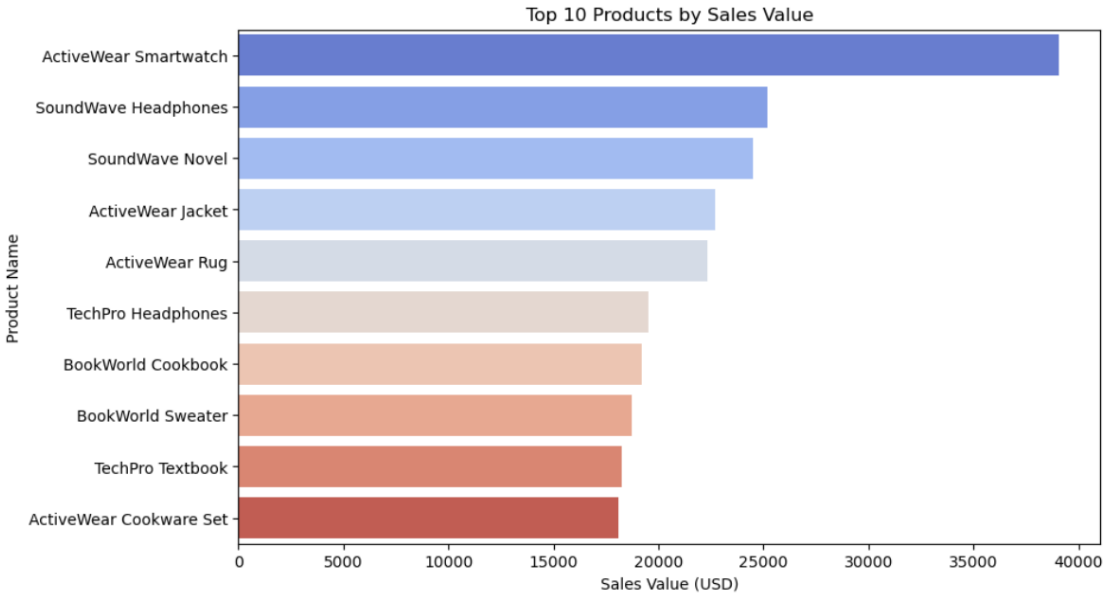
1. **Fluctuating Signups:** The number of signups hasn't been consistent. There are periods of high signups followed by periods of low signups.
2. **High Points:** There are several peaks where the number of signups significantly increases. These peaks occur around:
    - January 2023
    - July 2023
    - January 2024
    - July 2024
3. **Low Points:** Conversely, there are also periods with lower signups, notably around:
    - July 2022
    - January 2023

Possible Inferences:

- **Seasonality:** The fluctuations might suggest seasonal trends in customer signups. For example, there could be increased interest or activity during certain months or times of the year.
- **Marketing Campaigns:** The peaks could coincide with successful marketing campaigns or promotions.
- **External Factors:** External events or economic conditions might also influence customer signups.

3. Top 10 Products by Sales Value

A bar plot was created to display the top 10 products based on their total sales value. This analysis identifies the most revenue-generating products, which is essential for inventory planning, marketing focus, and maximizing profitability.



Key Observations:

- **ActiveWear Smartwatch is the Top Seller:** It has the highest sales value, exceeding $40,000.

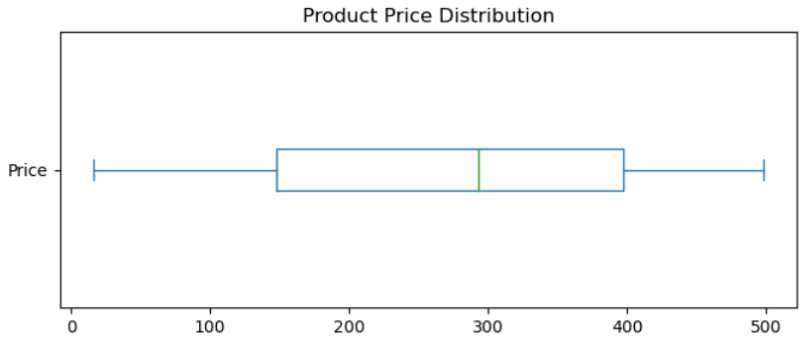
- **SoundWave Headphones and Novel:** These products follow closely behind the smartwatch, with sales values around $25,000.
- **Consistent Sales:** There's a relatively consistent gap between the top 3 products and the rest of the list.
- **ActiveWear Cookware Set:** This product has the lowest sales value among the top 10.

Possible Inferences:

- **Strong Demand for ActiveWear:** Products under the "ActiveWear" brand seem to be popular, with multiple items appearing in the top 10.
- **SoundWave Brand Appeal:** "SoundWave" products are also well-received, with both headphones and a novel ranking high.
- **Cookware Set Performance:** The lower sales value of the ActiveWear Cookware Set might indicate lower demand or a different pricing strategy.

4. Product Price Distribution

A box plot was used to visualize the distribution of product prices. This helps identify the price range, median price, and potential outliers, providing valuable insights for pricing strategies and market competitiveness.

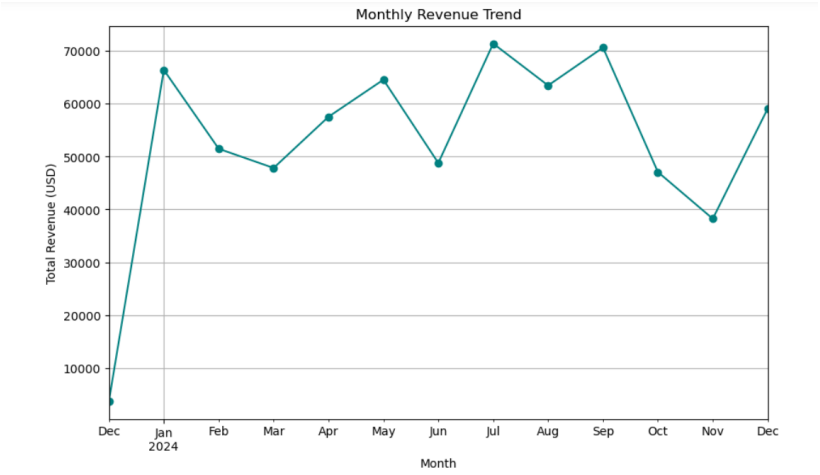


Key Observations:

- **Median Price:** The median price appears to be around 300. This means that half of the products have a price below 300, and half have a price above 300.
- **Quartiles:** The box represents the interquartile range (IQR), which spans from approximately 200 to 400. This indicates that 50% of the products fall within this price range.
- **Whiskers:** The whiskers extend from the box to the minimum and maximum values within 1.5 times the IQR. In this case, the whiskers reach up to around 500 on the right side

5. Monthly Revenue Trend

A line plot was created to analyze the monthly revenue trend over time. This visualization tracks total revenue per month, helping to identify peak sales periods and assess the overall financial performance, which can inform future sales strategies and forecasting.



Key Observations:

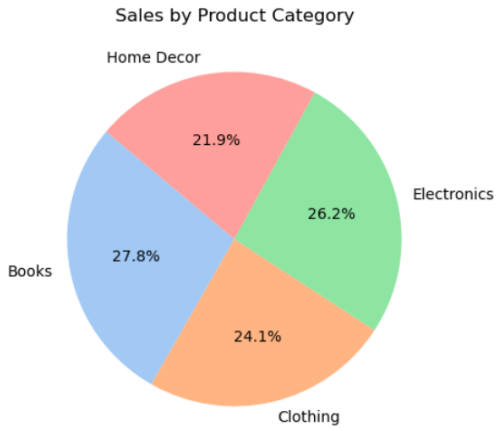
- **Fluctuating Revenue:** The revenue has not been consistent throughout the year. There are periods of high revenue followed by periods of lower revenue.
- **High Points:** The highest revenue points occur in:
    - January 2024
    - July 2024
    - December 2024
- **Low Points:** The lowest revenue points occur in:
    - December 2023
    - November 2024

Possible Inferences:

- **Seasonality:** The fluctuations might suggest seasonal trends in revenue. For example, there could be increased sales during certain months or times of the year.
- **Marketing Campaigns:** The peaks could coincide with successful marketing campaigns or promotions.
- **External Factors:** External events or economic conditions might also influence revenue.

6. Sales Distribution by Product Category

A pie chart was used to visualize the distribution of sales across different product categories. This chart provides a clear picture of which categories contribute the most to total revenue, aiding in decision-making for inventory management and targeted marketing campaigns.



The pie chart titled "Sales by Product Category" shows the distribution of sales across four different product categories.

Here are the key takeaways:

1. Books Lead the Sales: The largest slice of the pie belongs to "Books," accounting for 27.8% of total sales.

2. Electronics Close Second: "Electronics" comes in second with 26.2% of the sales.

3. Home Decor and Clothing: "Home Decor" and "Clothing" have similar shares with 21.9% and 24.1% of sales respectively.

Inferences:

- The company seems to have a strong presence in the book and electronics markets.
- Home Decor and Clothing categories could potentially benefit from increased marketing or product development efforts to boost their sales share.

7. High-Value Transactions
To identify high-value transactions, the dataset was filtered to include only the top 5% of transactions based on revenue. This subset highlights the transactions with the highest total value, providing insights into customer spending behavior and potential areas for personalized marketing or loyalty programs targeting high-spending customers.

1.**High-value transactions contribute significantly** to total revenue, with amounts ranging from $1669.48 to $1991.04.

2. **A small percentage of customers** are responsible for a large share of sales, highlighting the importance of targeting them.

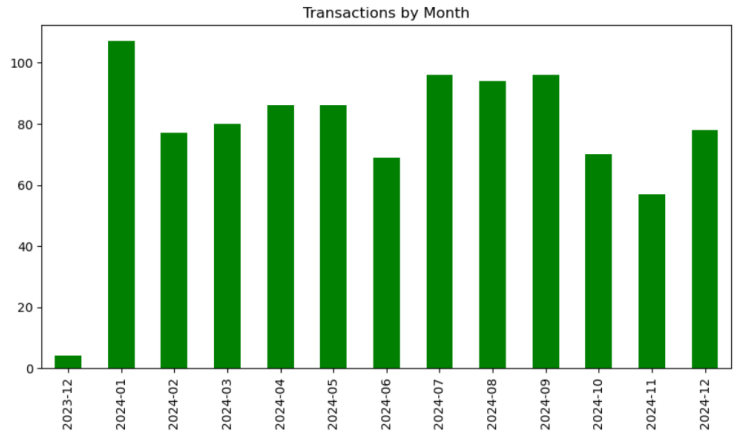
3.**Frequent high-value buyers** indicate the presence of loyal, high-spending customers.

4.**Potential for upselling or cross-selling** by analyzing the products linked to these high-value transactions.

5. **Revenue optimization** can be achieved by focusing on this segment with tailored marketing and exclusive offers.

6. **Customer retention strategies** should be a priority for the high-value transaction group to boost lifetime value.

8. Monthly Transaction Trends
The dataset was analyzed by counting the number of transactions per month. This monthly breakdown reveals seasonal trends, highlighting months with higher or lower transaction volumes. By visualizing these trends, businesses can identify peak sales periods, plan marketing campaigns accordingly, and adjust inventory management.



This graph displays the number of transactions for each month from December 2023 to December 2024. The x-axis represents the months, and the y-axis indicates the number of transactions.

Key Observations:

- **December 2023:** The number of transactions was very low in December 2023.
- **January 2024:** A significant increase in transactions occurred in January 2024.
- **Fluctuating Transactions:** The number of transactions fluctuates throughout the year, with some months showing higher activity than others.
- **Highest Transactions:** The highest number of transactions seems to be in January 2024, followed by September and August 2024.
- **Lowest Transactions:** The lowest number of transactions occurred in December 2023 and November 2024.

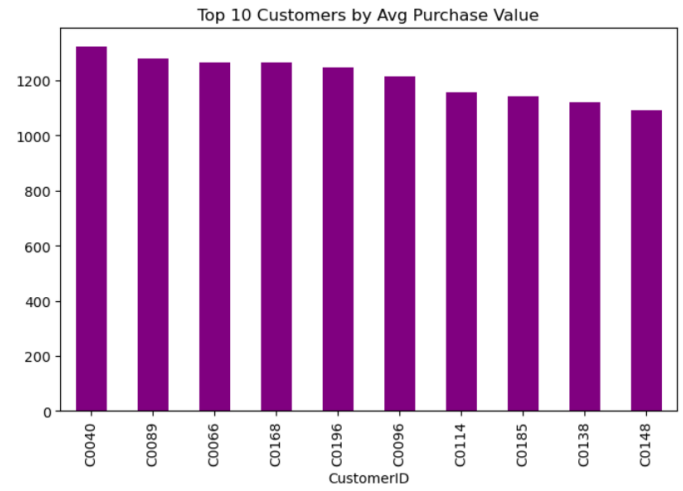
Possible Inferences:

- **Seasonality:** The fluctuations might suggest seasonal trends in transactions. There could be increased activity during certain months or times of the year.
- **Marketing Campaigns:** The peaks could coincide with successful marketing campaigns or promotions.
- **External Factors:** External events or economic conditions might also influence transaction numbers.

9. Repeat Customers
This analysis identifies customers who made more than one purchase by counting the number of transactions for each customer. The dataset filters for customers with more than one purchase, revealing repeat customers.

-> Number of Repeat Customers: 187

10. Top 10 Customers by Average Purchase Value
This analysis identifies the top 10 customers based on their average purchase value. By calculating the mean transaction value for each customer, we can pinpoint high-value customers who spend more per transaction. These customers can be targeted for premium offers or exclusive promotions to further increase their spending and loyalty.



The bar graph titled "Top 10 Customers by Avg Purchase Value" shows the average purchase value for the top 10 customers.

Here are some key observations:

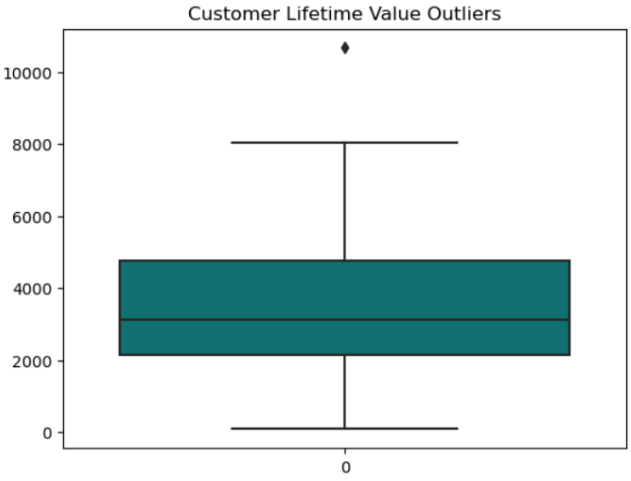
- **Customer C0040:** This customer has the highest average purchase value, with a value exceeding 1200.
- **High Average Purchase Values:** The top 10 customers all have high average purchase values, ranging from around 1000 to 1200.
- **Consistent Values:** There is a relatively consistent gap between the average purchase values of the top customers.

Inferences:

- The company has a group of high-value customers who consistently make large purchases.
- These customers are likely important to the company's revenue and profitability.

11. Identifying Outliers in Customer Lifetime Value (CLV)

This analysis visualizes outliers in customer lifetime value by plotting the total revenue per customer using a box plot. The outliers indicate customers with exceptionally high total revenue, which could highlight valuable, high-spending customers who may warrant targeted marketing efforts or special attention to maintain their loyalty.



This box plot provides a visual summary of the distribution of customer lifetime values, highlighting potential outliers.

Key Observations:

- **Median:** The median customer lifetime value appears to be around 3000.
- **Interquartile Range (IQR):** The IQR, represented by the box, spans from approximately 2000 to 4000. This means that 50% of the customers have lifetime values within this range.
- **Whiskers:** The whiskers extend from the box to the minimum and maximum values within 1.5 times the IQR. In this case, the upper whisker extends to around 8000.
- **Outlier:** There is one data point above the upper whisker, marked with a diamond. This point represents a customer with an exceptionally high lifetime value, considered an outlier.

12. Customer Retention Rate

The customer retention rate is calculated as the percentage of repeat customers relative to the total unique customers. It provides insight into how many of the customers are returning to make additional purchases. A higher retention rate indicates customer loyalty and the effectiveness of engagement strategies.

Average Order Value (AOV)

The average order value is the mean of the total transaction values across all orders. It reflects the average amount spent by customers per transaction and can help assess the pricing and sales strategies**.**

Top 10 Customers by Lifetime Value

The top 10 customers are identified based on the total revenue generated from each customer. These high-value customers are key to the business's success, and focusing on retaining them can yield substantial long-term revenue growth.

```
Customer Retention Rate: 93.50%
Average Order Value: $690.00
Top 10 Customers by Lifetime Value:
CustomerID
C0141    10673.87
C0054     8040.39
C0065     7663.70
C0156     7634.45
C0082     7572.91
C0188     7111.32
C0059     7073.28
C0028     6819.57
C0099     6715.72
C0165     6708.10
```

13. Correlation Analysis of Key Variables

The correlation matrix analyzes the relationships between key numerical columns: TransactionPrice, ProductPrice, Quantity, and TotalValue. By calculating these correlations, we can understand how these variables are interrelated, which can help in optimizing pricing and sales strategies.

Heatmap Visualization

A heatmap is used to visualize the correlation matrix, where the intensity of the color indicates the strength of the correlation. This allows us to quickly identify which variables are closely related, such as whether higher product prices result in higher total transaction values or if quantity impacts total sales.

Color Scale:

- Red indicates a strong positive correlation.
- Blue indicates a strong negative correlation.
- White indicates no correlation.

Values in the Cells:

- The numbers in each cell represent the correlation coefficient between the two variables. A value of 1.00 means a perfect positive correlation, -1.00 means a perfect negative correlation, and 0 means no correlation.

Observations:

- **TransactionPrice and ProductPrice:** These two variables have a strong positive correlation (0.72). This makes sense because, in general, higher product prices would lead to higher transaction prices.
- **Quantity and TotalValue:** There's also a strong positive correlation (0.61) between quantity and total value. More items purchased typically result in a higher total value.
- **Quantity and TransactionPrice:** The correlation between quantity and transaction price is very weak (-0.01). This suggests that the number of items purchased doesn't have a significant impact on the price of each item.
- **TransactionPrice and TotalValue:** There's a strong positive correlation (0.72) between transaction price and total value. This is expected, as higher transaction prices generally lead to a higher total value.

14. **Total Revenue Calculation**
This code calculates the total revenue by multiplying the TransactionPrice and Quantity for each transaction. By summing these values, we get the overall revenue from all transactions. This helps in understanding the total financial performance and provides a benchmark for analyzing sales and setting targets.

-> Total Revenue: $689,995.56