

1. Introduction

This report summarizes an exploratory data analysis (EDA) of an eCommerce Transactions dataset. The dataset includes three tables: **Customers**, **Products**, and **Transactions**. After merging and cleaning the data, various analyses were conducted to understand customer distribution, product performance, and revenue trends.

2. Exploratory Data Analysis (EDA)

2.1 Data Loading and Cleaning

- **Libraries Used:** `pandas`, `numpy`, `matplotlib`, `seaborn`
- **DataFrames:**
 - `customers_df` (CustomerID, CustomerName, Region, SignupDate)
 - `products_df` (ProductID, ProductName, Category, Price)
 - `transactions_df` (TransactionID, CustomerID, ProductID, TransactionDate, Quantity, TotalValue, Price)

Cleaning Steps:

1. Converted date columns (`TransactionDate`, `SignupDate`) to datetime format.
2. Ensured numeric columns (`Quantity`, `Price`, `TotalValue`) were properly typed.
3. Dropped duplicates if found in `CustomerID`, `ProductID`, and `TransactionID`.

2.2 Merging Datasets

1. Merged `transactions_df` with `products_df` on `ProductID`.
2. Merged the resulting dataset with `customers_df` on `CustomerID`, creating a unified `full_df`.

2.3 High-Level Statistics

- **Unique Customers:** 199
- **Unique Products:** 100
- **Total Transactions:** 1000
- **Total Revenue:** \$689,995.56
- **Average Order Value (AOV):** \$689.99

Descriptive Stats (Quantity, TotalValue, Price)

Metric	Quantity	TotalValue	Price_trans
--------	----------	------------	-------------

count	1000	1000	1000
mean	2.54	689.99	272.55
std	1.12	493.14	140.74
min	1.00	16.08	16.08
25%	2.00	295.29	147.95
50%	3.00	588.88	299.93
75%	4.00	1011.66	404.40
max	4.00	1991.04	497.76

2.4 Visual Explorations

1. Customers by Region

- Pie chart shows the regional distribution:
 - South America: 29.5%
 - North America: 23.0%
 - Europe: 25.0%
 - Asia: 22.5%

2. Monthly Transaction Count (2024-01 to 2024-11)

- Initial spike in January and February (over 80–100 transactions).
- Moderate fluctuations the rest of the year with a dip around September–October.

3. Total Sales by Product Category

- **Books:** Highest total sales (~\$200k)
- **Electronics:** Second (~\$170k)
- **Clothing:** Third (~\$150k)
- **Home Decor:** Fourth (~\$120k)

3. Business Insights

1. Books Are the Top Revenue-Generating Category

Books category leads with nearly \$200k in revenue, representing the single largest share among product lines. This indicates a consistent or large-scale demand for reading materials. A deeper look into titles or sub-categories (e.g., fiction vs. non-fiction) could reveal further opportunities to expand the most popular segments

and run promotional campaigns to upsell related items.

2. **Moderate Seasonal Peaks in Early 2024**

Transaction counts surge in January and February 2024 but then stabilize. This could relate to New Year promotions or holiday gift cards. Planning targeted marketing during this early-year peak may optimize campaign budgets. Monitoring subsequent months helps refine re-stocking plans and align marketing with slow periods to drive traffic.

3. **High Average Order Value (AOV) Indicates Bundling Opportunities**

With an AOV of \$689.99, each transaction is relatively high-value. Cross-selling and bundling strategies could be effective, particularly pairing top categories like Books or Electronics with complementary items. This approach can nudge shoppers toward larger purchases per transaction.

4. **Geographic Opportunities in South America**

South America accounts for nearly 30% of customers, slightly higher than other regions. This is a significant customer base deserving of localized promotions or expanded shipping options. Exploring region-specific product preferences could further bolster sales and brand loyalty within this market.

5. **Potential Growth in Underrepresented Categories**

While Books and Electronics dominate, categories like Clothing and Home Decor still yield substantial sales (\$150k+ and \$120k+). Introducing new products or styles within these categories could capture customers who want more variety. Monitoring category-level trends (e.g., seasonal clothing or home décor during holiday seasons) can guide restocking cycles and marketing pitches.

4. Conclusion

From the EDA, we see strong performance in the Books category, a substantial presence in South America, and an early-year surge in transaction volume. These findings suggest opportunities for **targeted marketing**, **product expansion**, and **strategic inventory planning**. By optimizing around these insights, the business can improve customer retention and boost overall revenue.
