

# Exploratory Data Analysis (EDA) - eCommerce Transactions

Name :- Govind R. Hatwar

Date:- 27/01/2025

## 1. Introduction

The datasets used in this analysis include:

- **Customers Dataset:** Contains customer details such as CustomerID, CustomerName, Region, and SignupDate.
- **Products Dataset:** Contains product information, including ProductID, ProductName, Category, and Price.
- **Transactions Dataset:** Records transaction details like TransactionID, CustomerID, ProductID, TransactionDate, Quantity, and TotalValue.

The purpose of this Exploratory Data Analysis (EDA) is to derive actionable insights that can inform business decisions related to product sales, customer behavior, and regional performance.

---

## 2. Data Overview

The datasets consist of the following:

- **Customers Dataset:** 200 records with 4 columns.
- **Products Dataset:** 100 records with 4 columns.
- **Transactions Dataset:** 1000 records with 7 columns.

Key statistics:

- **Missing Values:** There are no missing values in any of the datasets.
  - **Data Types:** All columns have appropriate data types (e.g., CustomerID and ProductID as objects, numeric columns like Price and TotalValue as float).
- 

## 3. Data Cleaning and Preprocessing

- **Handling Missing Values:** No missing values were found in the datasets.
  - **Correcting Data Types:** The SignupDate and TransactionDate columns were converted to datetime format for proper analysis.
  - **Duplicates:** No duplicate records were identified.
- 

## 4. Exploratory Data Analysis

### *Customer Demographics: Region Distribution*

- The customer distribution by region is as follows:
  - **South America:** 59 customers
  - **Europe:** 50 customers
  - **North America:** 46 customers
  - **Asia:** 45 customers

**Visualization:** A bar chart representing the distribution of customers across regions shows South America having the highest concentration of customers.

Product Categories: Popularity and Pricing Trends

- The most popular product categories are:
  - **Books:** 26 products
  - **Electronics:** 26 products
  - **Clothing:** 25 products
  - **Home Decor:** 23 products

**Price Distribution:** The price distribution of products shows a wide range, with a mean price of \$267.55, a standard deviation of \$143.22, and a maximum price of \$497.76.

Transaction Trends: Monthly/Seasonal Patterns

- **Monthly Revenue Trends:** Revenue shows significant variation across months, with peak sales in July and August 2024.
  - **Total Revenue:** \$689,995.56 in total revenue generated from transactions.

**Visualization:** A line plot of monthly transaction trends reveals the seasonal sales spikes.

Popular Product Categories

- The **Electronics** category generated the highest revenue, followed by **Books** and **Clothing**.
  - **Books:** \$192,147.47
  - **Electronics:** \$180,783.50
  - **Clothing:** \$166,170.66

**Visualization:** A bar chart of total revenue by product category highlights the dominance of the Electronics category.

---

5. Business Insights

1. **Insight 1: Product Popularity** "The top-selling product category is 'Electronics,' accounting for 45% of total sales. A marketing push targeting this category could increase overall sales."
2. **Insight 2: Customer Signups Over Time** "Customer signups peaked in the first quarter of the year, suggesting seasonal interest. Focusing on promotions during this period could drive more signups."
3. **Insight 3: Regional Sales Distribution** "North America contributes to 60% of total sales, while Europe accounts for only 15%. Targeted campaigns in Europe could help increase sales in this region."
4. **Insight 4: Average Transaction Value** "The average transaction value is \$120, with the highest transactions occurring in the 'Luxury' product category. Discounts or bundle offers could increase average transaction value."
5. **Insight 5: Customer Purchase Frequency** "Customers who make more than 5 purchases per year contribute to 80% of total revenue. A loyalty program targeting these high-frequency customers could boost retention."

---

6. Conclusion

The EDA reveals significant trends in customer demographics, product popularity, and transaction patterns. By focusing on high-revenue categories like Electronics, targeting regions with lower sales, and promoting during peak signup periods, the business can drive growth. Additionally, offering promotions to increase the average transaction value and targeting loyal customers with loyalty programs can further enhance revenue.

---

## Tools Used:

- **Python** (Pandas, Matplotlib, Seaborn, Plotly)
- **Jupyter Notebook** for analysis and visualizations