

# EXPLORATORY DATA ANALYSIS AND BUSINESS INSIGHTS

## 1. Introduction

- Brief overview of the dataset consisting of three files Customers.csv, Products.csv, Transactions.csv and to identify trends, patterns, and actionable insights.

## 2. Data Cleaning

- Numerical columns were filled with meaningful default values (mean values were filled for the Price category in Products and transactions datasets).
- Categorical columns were filled with "Unknown" to maintain dataset integrity.
- The datasets were merged correctly using common keys (ProductID and CustomerID).
- After merging, the dataset included transaction details, product details, and customer information, making it ready for comprehensive analysis.

## 3. Exploratory Data Analysis

- Performed summary statistics on the merged dataset to understand key metrics for numerical and categorical columns.
- Grouped by Category, Region and summed TotalValue to identify profitable categories and top-performing regions.
- Counted the number of customers in each region to identify customer distribution.
- Grouped by month to track sales trends and grouped by DayOfWeek to determine which days generate the most sales.



- Calculated average spending by CustomerName to identify high-value customers.
- Counted the frequency of transactions per customer to identify loyal customers.

## 4. Business Insights

### a. Most Profitable Categories

- The **Books** category leads in revenue generation with **\$192,147.47**, followed by **Electronics** and **Clothing**, which generated **\$180,783.50** and **\$166,170.66**, respectively.
- Actionable Insight:** Focus on expanding the inventory and promotional strategies for Books, Electronics, and Clothing to capitalize on their high revenue potential.

### b. Top-Selling Products

- The **ActiveWear Smartwatch** is the best-selling product, generating **\$39,096.97**, followed by **SoundWave Headphones** and **SoundWave Novel** with revenues of **\$25,211.64** and **\$24,507.90**, respectively.
- **Actionable Insight:** Prioritize marketing efforts for ActiveWear and SoundWave products. Bundle deals or discounts on these high-performing products could further boost sales.

#### c. Regional Revenue Analysis

- **South America** drives the highest revenue with **\$219,352.56**, followed by **Europe** and **North America**, which generated **\$166,254.63** and **\$152,313.40**, respectively.
- **Actionable Insight:** Allocate more resources to South America and Europe for market penetration and customer acquisition, as they offer the highest revenue potential.

#### d. Regional Customer Base

- South America also leads in the number of customers (**304**), followed by North America (**244**) and Europe (**234**).
- **Actionable Insight:** Combine revenue and customer base insights to design region-specific strategies, particularly loyalty programs in South America to retain its large customer base.

#### e. Sales Trends Over Time

- Sales peaked in **July 2024 (\$71,366.39)** and **September (\$70,603.75)**, indicating seasonal spikes.
- **Actionable Insight:** Plan for promotional campaigns and stock up inventory ahead of these peak months to maximize revenue.

### 5. Recommendations

- Focus on high-performing regions and products.
- Optimize operations during peak sales times.
- Develop strategies to retain high-value and loyal customers.

### 6. Conclusion

The data cleaning process addressed missing values, ensuring reliable analysis. Sales performance revealed top categories and products, while regional insights highlighted key areas for growth. These insights provide actionable recommendations for optimizing sales and improving customer engagement.