Exploratory Data Analysis (EDA)

1. Overview of the Datasets

Customers Dataset:

- Contains 200 records with columns: CustomerID, CustomerName, Region, and SignupDate.
- Data types are all objects except for the SignupDate column after conversion to datetime.
- No missing values were detected.

Products Dataset:

- Includes 100 products with details: ProductID, ProductName, Category, and Price.
- The Price column is numerical, while others are categorical.
- No missing values were found.

Transactions Dataset:

- Contains 1,000 transaction records with columns such as TransactionID, CustomerID, ProductID, TransactionDate, Quantity, TotalValue, and Price.
- No missing values were detected.
- Numerical columns like Quantity and TotalValue provide insights into purchase behavior

Merged Dataset:

- Datasets were merged on CustomerID and ProductID to create a unified view.
- Final dataset includes 1,000 entries with 13 columns.

2. Customer Analysis

- **Total Customers:**

  - The dataset contains 200 unique customers.

- **Customer Distribution by Region:**

  - South America has the highest customer count (59), followed by Europe (50), North America (46), and Asia (45).

- **Customer Signup Trends:**

  - Customer signups increased in recent years, indicating growing engagement and acquisition efforts.

3. Product Analysis

- **Top-Selling Products:**

  - The top product is **ActiveWear Smartwatch**, with 100 units sold.

- o Other popular products include **SoundWave Headphones** (97 units) and **HomeSense Desk Lamp** (81 units).

- **Revenue by Product Category:**

  - o The Books category generated the highest revenue ($192,147.47), followed by Electronics ($180,783.50), Clothing ($166,170.66), and Home Decor ($150,893.93).

4. Transaction Analysis

- **Average Transaction Value:**

  - o The average transaction value is **$690**, reflecting a premium customer base.

- **Transaction Trends**:

  - o Transaction volume and quantity show seasonal peaks, which could correlate with promotional campaigns or holidays.

- **High-Value Transactions**:

  - o Transactions exceeding the 95th percentile in total value mostly involve expensive electronics and books.

5. Regional Insights

- **Revenue by Region**:

  - o South America is the most significant contributor to revenue, generating **$219,352.56** (35.6% of total revenue).
  - o Europe, North America, and Asia contribute 26.9%, 24.7%, and 24.7%, respectively.

- **Regional Trends**:

  - o Each region shows unique preferences for product categories, enabling region-specific marketing strategies.

6. Visualization Insights

- **Bar Charts**:

  - o Used to visualize customer distribution by region and top-selling products.

- **Line Charts**:

  - o Highlight trends in customer signups and transactions over time.

- **Pie Charts**:

  - o Show regional revenue distribution.

Exploratory Data Analysis (EDA)

7. Business Implications

- **Market Focus**:

  - South America, being the highest revenue generator, should be the focus for marketing and sales strategies.

- **Product Expansion**:

  - High-performing categories like Books and Electronics should be expanded.

- **Customer Engagement**:

  - Premium customers (high transaction values) can be targeted with loyalty programs and personalized recommendations.

- **Inventory Management**:

  - Popular products like **ActiveWear Smartwatch** need to be prioritized in stock planning.

8. Limitations

- **Data Completeness**:

  - Although there were no missing values, the dataset size limits deeper segmentation insights.

- **Time Frame**:

  - The dataset might not capture long-term trends due to its limited timeframe.