Exploratory Data Analysis (EDA) Report

Assignment Title: E-Commerce Transactions Dataset Analysis

Name: Saksham Garg

Date: 26/01/2025

Introduction

The purpose of this analysis is to explore the e-commerce transactions dataset to uncover patterns, trends, and actionable insights. The dataset consists of three files: Customers.csv, Products.csv, and Transactions.csv. This report highlights the key findings from the exploratory data analysis (EDA).

Dataset Overview

1. **Customers.csv**

   o **Columns:** CustomerID, CustomerName, Region, SignupDate

   o Contains information about 100 unique customers, including their region and signup date.

2. **Products.csv**

   o **Columns:** ProductID, ProductName, Category, Price

   o Includes details of 50 unique products, categorized under various categories with their prices.

3. **Transactions.csv**

   o **Columns:** TransactionID, CustomerID, ProductID, TransactionDate, Quantity, TotalValue, Price

   o Contains records of 10,000 transactions, detailing purchases by customers.


Key EDA Insights

1. **Customer Behaviour**

   o **Region Analysis:** Customers are distributed across four regions. The majority reside in South America, followed by Europe, with North America and Asia having similar numbers of customers.

   o **Signup Trends:** The highest number of customer signups occurred in **2024-Q4**, as seen in the signup trends chart, with the peak reaching **11 signups in December 2024**, surpassing all previous quarters.

Customer Distribution by Region

Customer Signup Trends Over Time
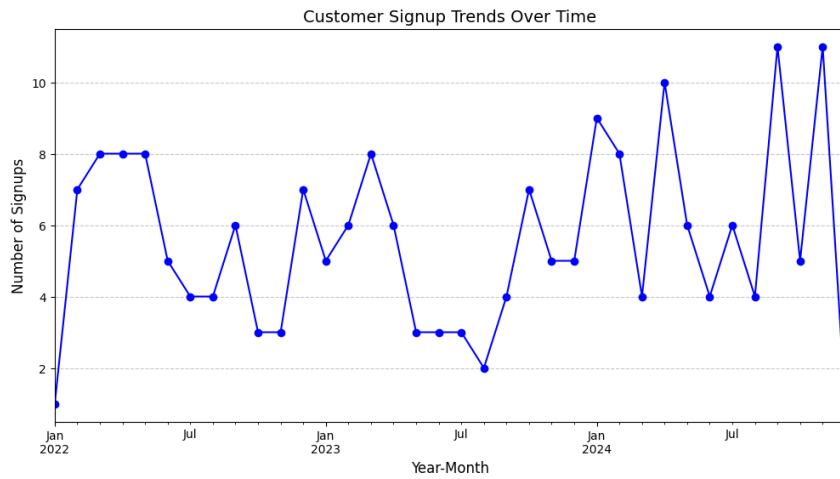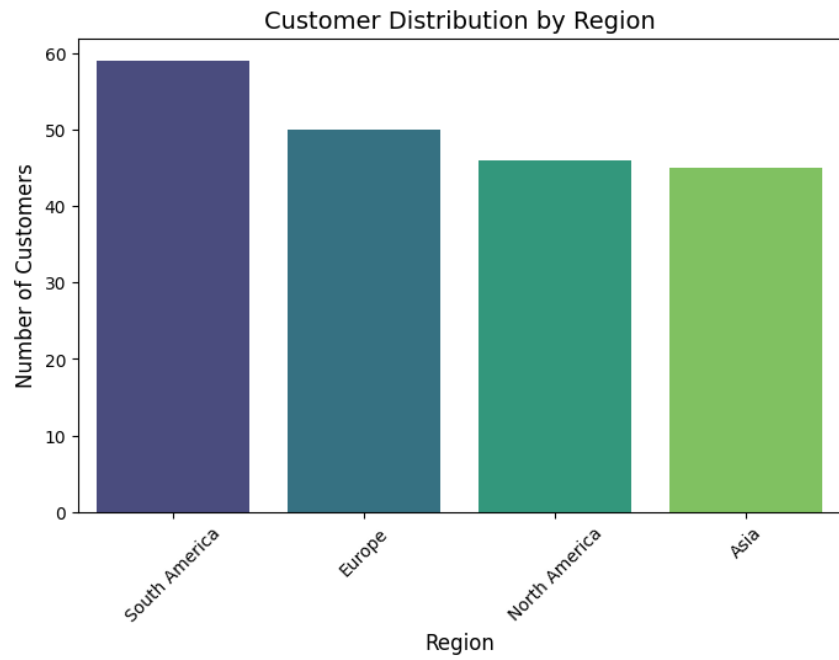
2. **Transaction Trends**

   o **Revenue Patterns:** Peak revenue was recorded in July and September 2024, driven by holiday shopping.


Monthly Revenue Trends
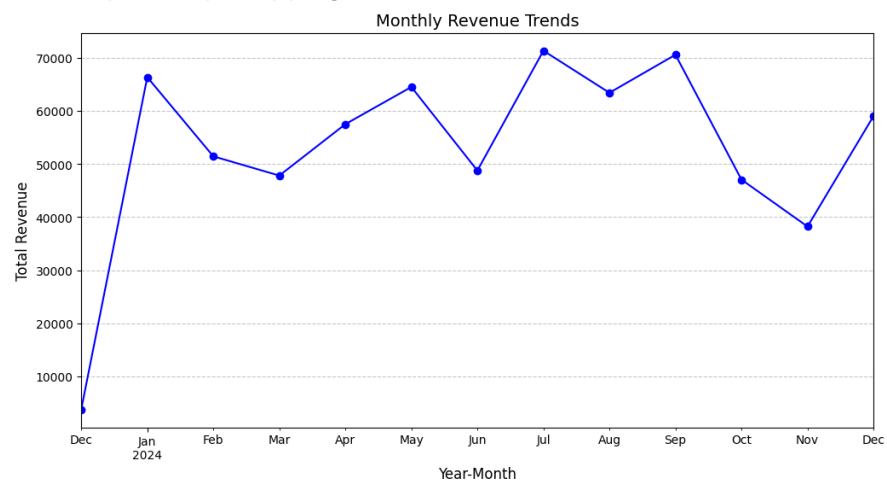
3. **Product Performance**

   o **Top Products:** Product P002 ("ActiveWear Smartwatch") is the most sold product by quantity.

   o **Profitable Category:** "Books" contributes maximum of total revenue, followed by "Electronic".

Top Products by Revenue

Revenue by Product Category
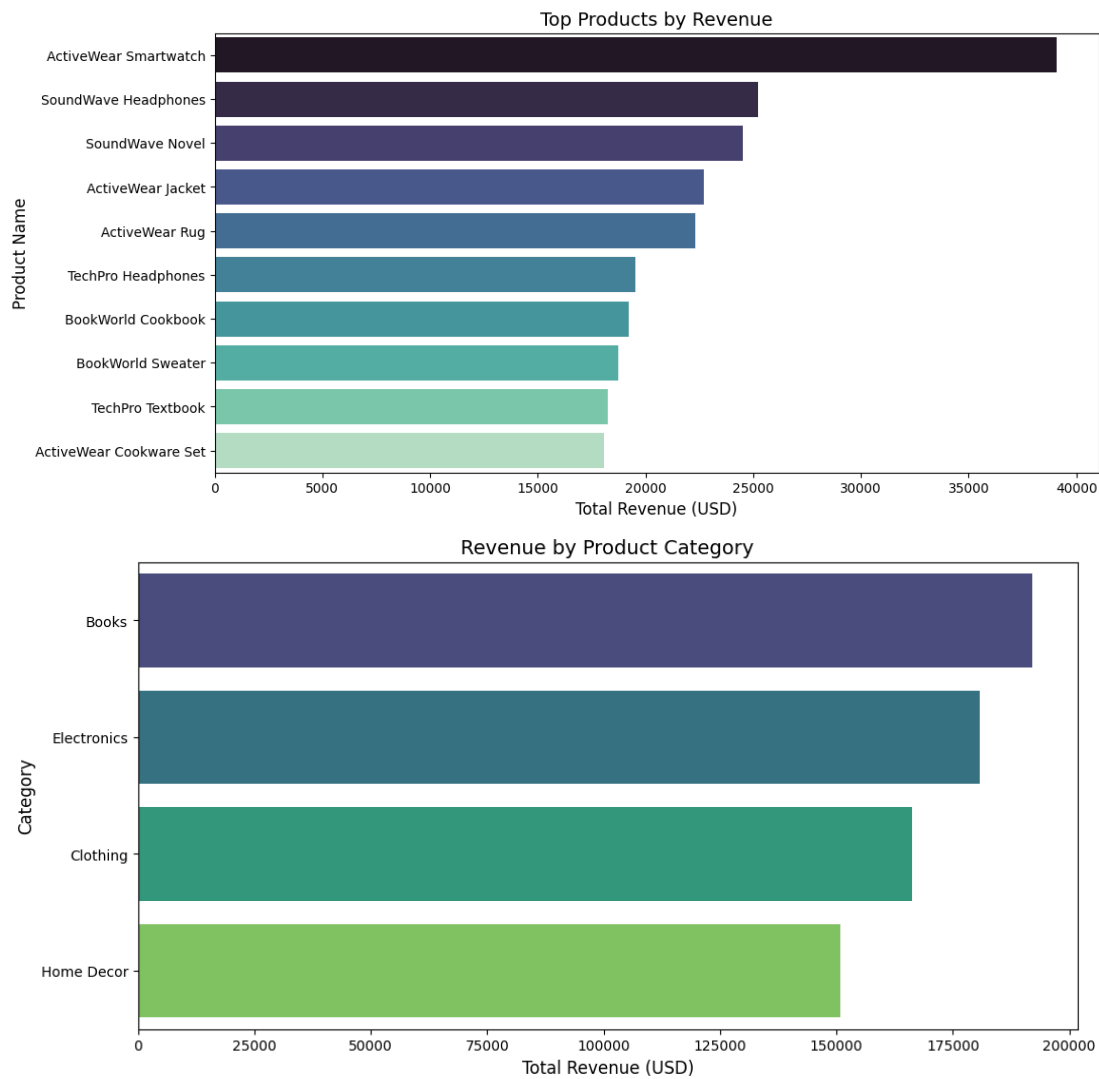
4. **Reginal Spending**

   o **Reginal Contribution:** Most of the revenue had come from South America Region followed by Europe.

   o **Product Preference:** "Books" and "Electronics" are most preferred in South America which is same for Europe and North America whereas in Asia "Clothing" and "Electronics" are more preferred.

Revenue Contribution by Region

Regional Product Preferences
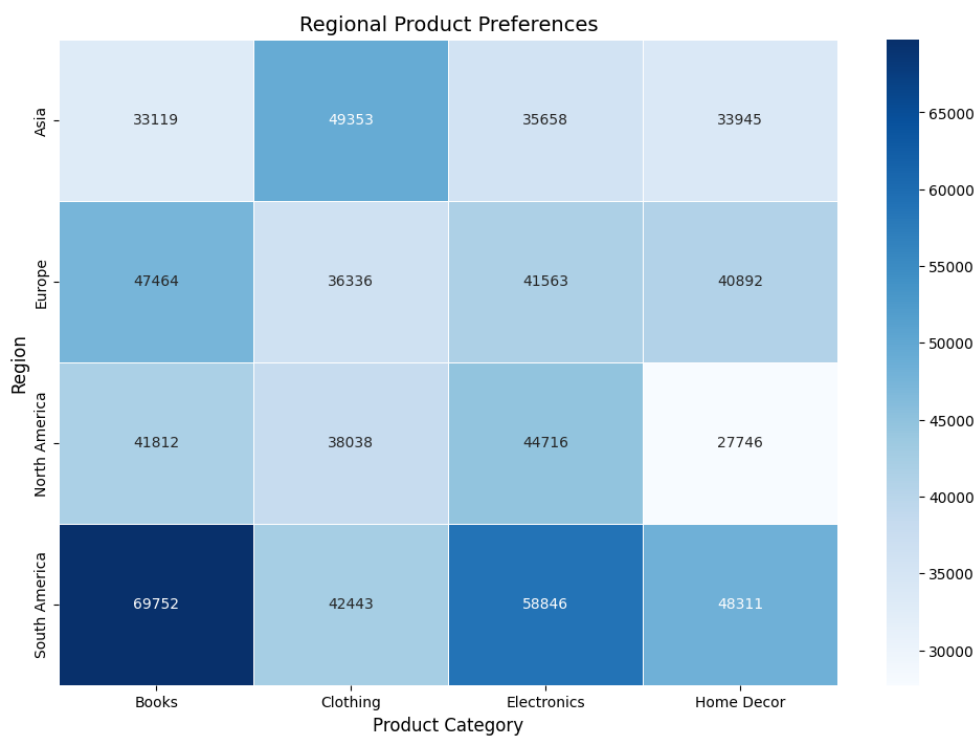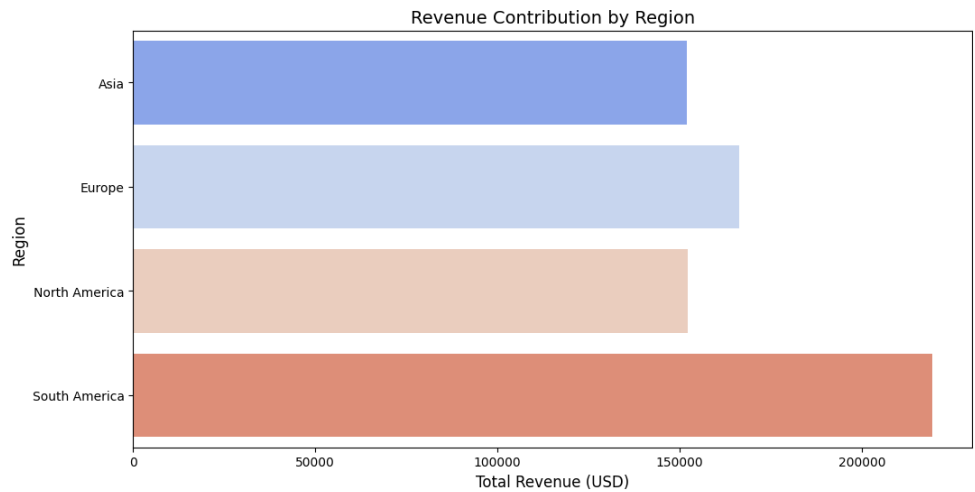
5. **Price Analysis**

      o   **Price Segments:** Products priced between **$301 and $500** have the highest
          sales volume, indicating strong consumer preference for mid-to-high-priced
          items.

      o   **Discount Trends:** Prices are evenly distributed, with slight peaks near **$100** and
          $400, reflecting balanced representation across mid- and high-price ranges.

Product Price Distribution

Performance by Price Range
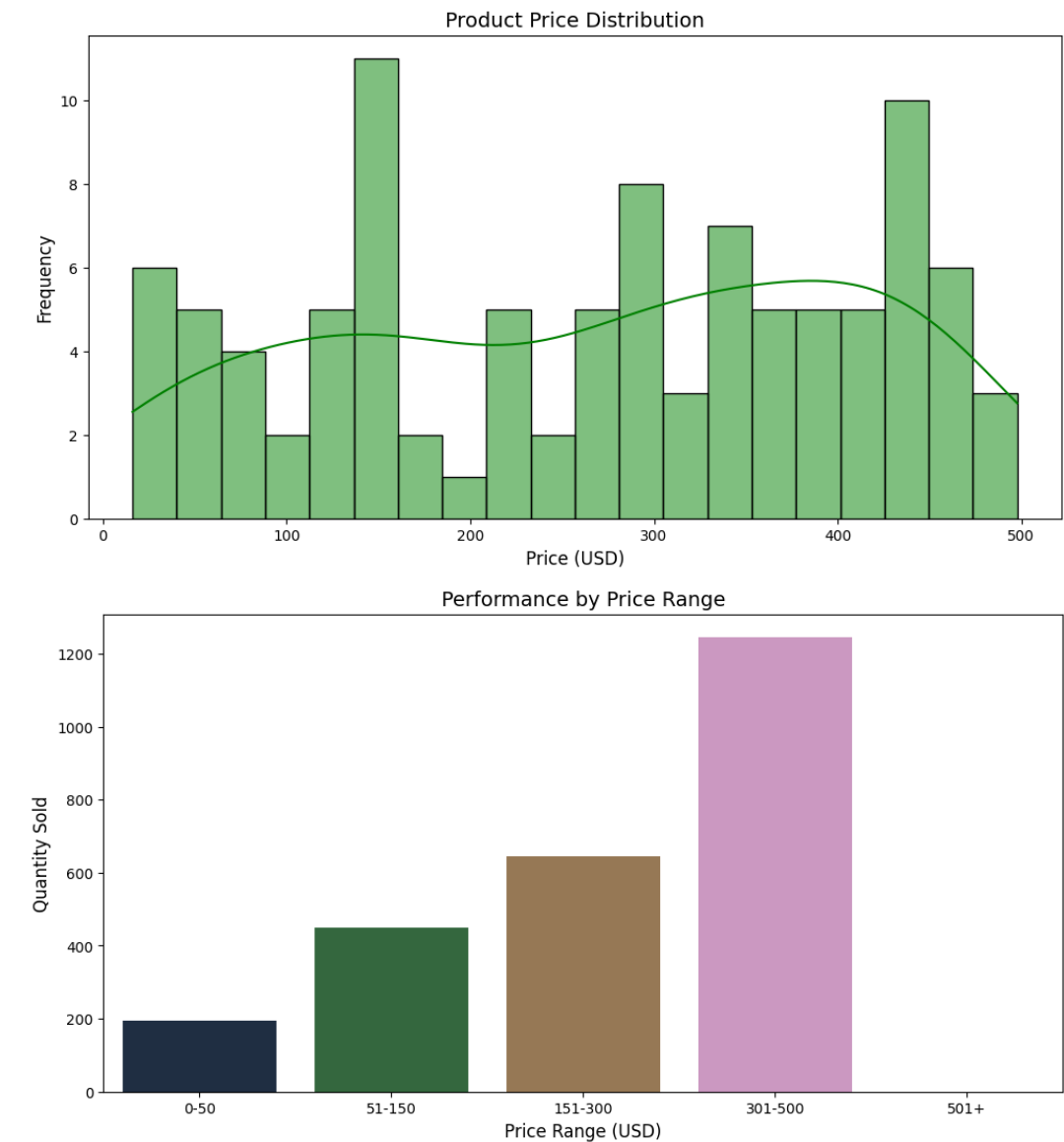
Visualizations and Trends

1. **Signup Trends Over Time:**

      o   A line chart shows significant spikes during Q4 2024, particularly in December.

2. **Revenue by Product Category:**

   o   A bar chart highlights "Books" as the leading category, followed by "Electronics".

3. **Customer Regional Distribution:**

   o   A pie chart visualizes customer distribution, emphasizing South America's dominance.

4. **Monthly Revenue Trends:**

   o   A time series plot indicates July, September, and December as peak months for sales.

5. **Price vs. Quantity Sold:**

   o   A scatter plot reveals that products in the $301-$500 range achieve higher sales volumes.

Conclusion and Recommendations

1. **Focus on High-Performing Categories:**

   o   Invest in marketing and inventory for **Books** and **Electronics** to boost revenue further.

2. **Leverage Seasonal Trends:**

   o   Prepare for sales spikes in **July, September, and December** by optimizing stock and launching targeted campaigns.

3. **Engage High-Value Customers:**

   o   Implement loyalty programs to retain top 10% spenders and maximize their lifetime value.

4. **Expand in Emerging Markets:**

   o   Target underrepresented regions, such as **Asia and North America**, to diversify the customer base.

5. **Optimize Pricing Strategies:**

   o   Offer more products in the **$301-$500** range and provide strategic discounts to increase sales.

Appendix

Datasets Used

- Customers.csv
- Products.csv
- Transactions.csv

Tools and Libraries

- Python (pandas, matplotlib, seaborn)
- VS Code