Exploratory Data Analysis (EDA) Report

1. Introduction:

This report provides an Exploratory Data Analysis (EDA) of an E-Commerce transaction Dataset. The dataset consisted of three files: customers.csv, transactions.csv and products.csv. The goal of this analysis was to uncover trends, get insights, fields of improvement, etc. To improve the growth of business.

For the analysis, first, I looked into the dataset, then cleaned it, then merged the datasets based on customerID, productID etc. to plot the graphs and gain valuable insights.

1. Dataset Overview

- Customers Dataset: Contains information about customers, including their ID, name, region, and signup date.

- Products Dataset: Contains information about products, including their ID, name, category, and price.

- Transactions Dataset: Contains information about transactions, including transaction ID, customer ID, product ID, transaction date, quantity, total value, and price.

2. Data Cleaning

- Missing Values: No missing values were found in any of the datasets.

- Duplicates: No duplicate rows were found in any of the datasets.

- Data Types: Date columns (`SignupDate` in Customers.csv and `TransactionDate` in Transactions.csv) were converted to datetime format for easier analysis.

3. Key Findings

3.1 Customer Demographics

- **Region Distribution:** The majority of the customers are located in South America, followed by Europe and North America. Last is Asia. This information suggests that South America is a key market for the business.

**Customer distribution by region**
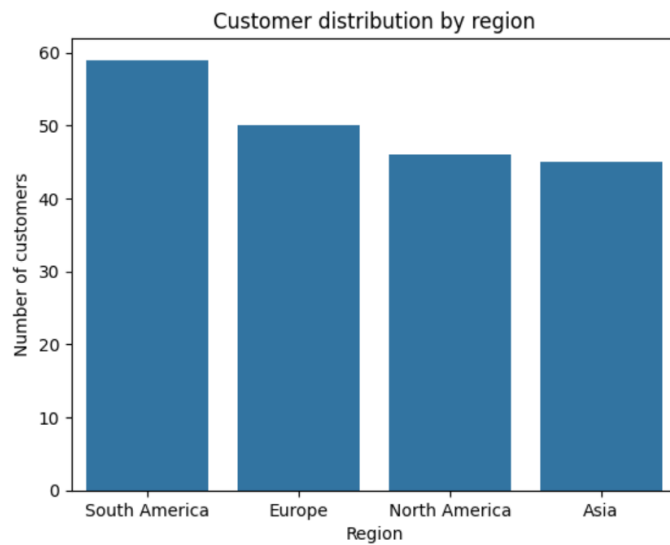
-Signup Trends:

Customer signups fluctuated over time, with a noticeable drop followed by steady growth from June to July. The decline may be due to reduced demand, while the growth aligns with holiday promotions or marketing efforts.
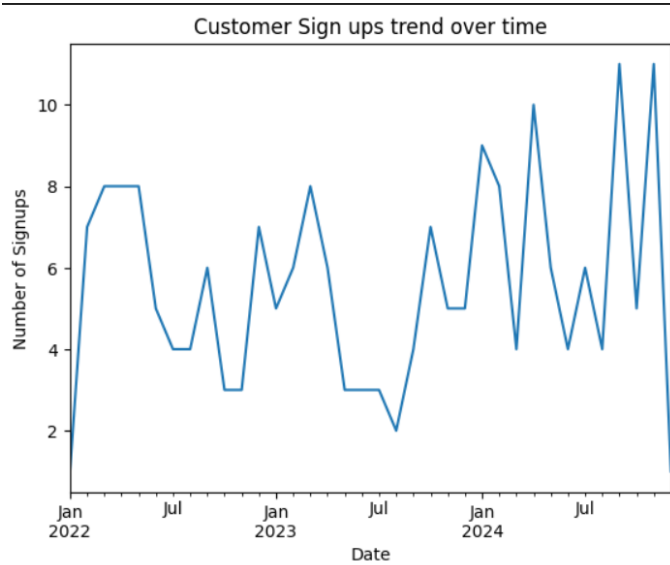
3.2 Product Analysis:

- Price Distribution:

Many products that are priced between $100 and $200, with a few high-priced items (something around $300) are purchased more frequently. This indicates that products in this price range are more in demand.

Price Distribution
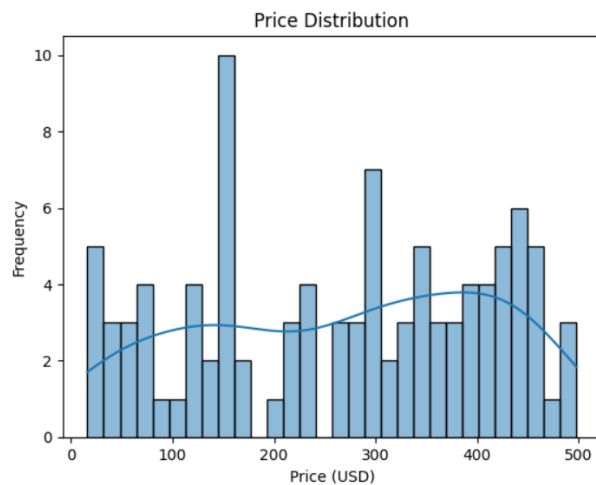
- Top-Selling Categories:

  The category of Books generates the highest revenue, followed by Electronics and Clothing. However, Home decor has the lowest sales volume, suggesting that customers are more interested in Books, electronics and Clothing.
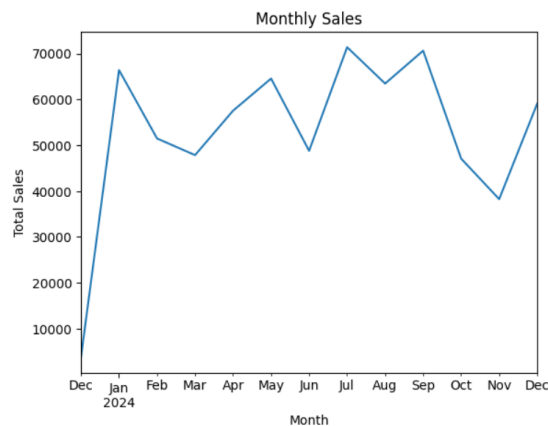
Total Sales by Product category
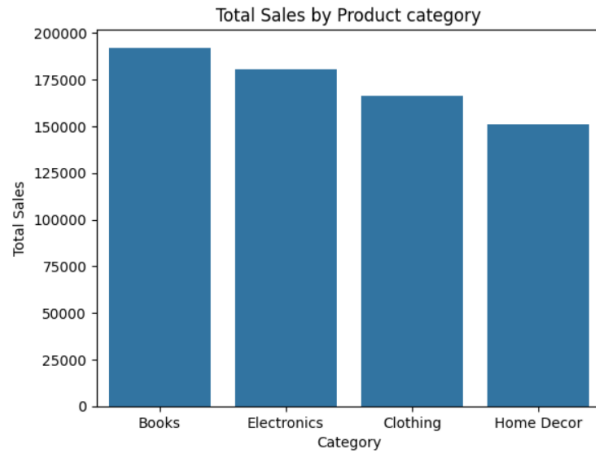
3.3 Transaction Analysis

- Sales Trends:

Sales fluctuate throughout the year with noticeable peaks in January, mid-year (July-September), and December.

Low points include February-March and October-November.

This pattern suggests a seasonal impact, possibly influenced by promotions, holidays, or market trends.

Monthly Sales
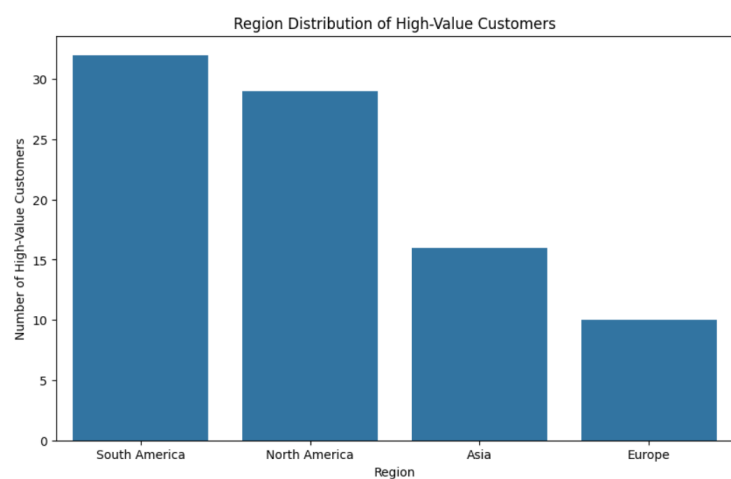
-Top customers:

Customer with ID C0141 spends the most with total spending of $10673.87.

```
Top 10 Customers by Total Spending:
    CustomerID  TotalSpending
140     C0141      10673.87
53      C0054       8040.39
64      C0065       7663.70
155     C0156       7634.45
81      C0082       7572.91
186     C0188       7111.32
58      C0059       7073.28
27      C0028       6819.57
98      C0099       6715.72
164     C0165       6708.10
```

- Regions of High value Customers:

According to the data, most of the High Valued customers are from South America region.

High Valued customers from south america are 32 and then 29 from North America.

4. Business Insights

4.1. South America is a Key Market:

Most customers are from South America, contributing significantly to revenue. The business should focus on expanding its presence in this region because it's a good market and high value customers are from the South America region.

4.2. Books Driven Revenue:

The Books category generates the highest revenue with highest sales volume. The business should consider optimizing pricing and marketing strategies for the category of Books.

4.3. Seasonal Sales Peaks:

Sales and customer signups peak in July and July, likely due to holiday shopping. The business should plan targeted marketing campaigns and promotions during this period.

4.4. High-Value Customers:

A small group of high-spending customers contributes something around 11.02% to the total revenue generated. The business should implement loyalty programs and give some extra rewards and coupons to retain these customers.

4.5. Affordable Products Dominate:

Most products priced between $100 and $200 indicate a focus on affordability because they have the highest frequency of purchase. The business should continue to offer a range of affordable products to attract a broad customer base.

5. Conclusion

This EDA has uncovered several key trends and insights that can inform business strategy. By focusing on high-value customers, optimizing pricing for key product categories, and targeting key regions, the business can drive growth and improve customer satisfaction.

6. Visualizations

Here are the visualizations included in the report:

6.1. Region Distribution: Bar plot showing the number of customers by region.

6.2. Signup Trends: Line plot showing customer signups over time.

6.3. Price Distribution: Histogram showing the distribution of product prices.

6.4. Top-Selling Categories: Bar plot showing total sales by product category.

6.5. Region Distribution of High-Value Customers: Bar plot showing regions of high value customers.

6.6. Product Category for High value customers: Bar plot showing categories which high value customers chose to make purchase from.

7. Suggestions

1. Targeted Marketing: Develop marketing campaigns for high-value customers and key regions.

2. Pricing Strategy: Optimize pricing for high-revenue categories like Books.

3. Seasonal Promotions: Plan promotions and discounts for the months of June and July to capitalize on holiday shopping trends.