

# Data Science Assignment eCommerce Transactions Dataset

## Objective

To perform EDA on the provided datasets.

## Dataset Description

Files Description:

### 1. Customers.csv

- CustomerID: Unique identifier for each customer.
- CustomerName: Name of the customer.
- Region: Continent where the customer resides.
- SignupDate: Date when the customer signed up.

### 2. Products.csv

- ProductID: Unique identifier for each product.
- ProductName: Name of the product.
- Category: Product category.
- Price: Product price in USD.

### 3. Transactions.csv

- TransactionID: Unique identifier for each transaction.
- CustomerID: ID of the customer who made the transaction.
- ProductID: ID of the product sold.
- TransactionDate: Date of the transaction.
- Quantity: Quantity of the product purchased.
- TotalValue: Total value of the transaction.
- Price: Price of the product sold.

## Procedure

### 1.Importing Libraries

Libraries that were imported are:

## 2.Data Collection

Datasets used for this assignment are:

- Transactions
- Products
- Customers

## 3.EDA(Exploratory Data Analysis)

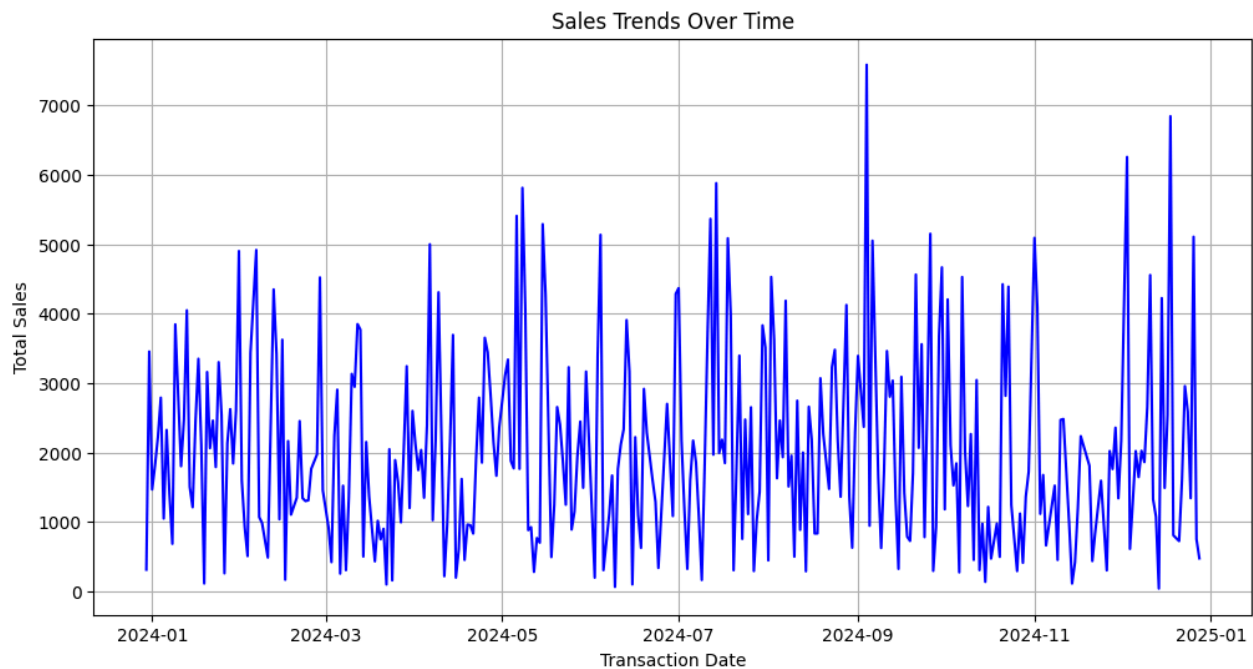


Figure: Sales Trends over time

- The sales trend shows fluctuations over time, indicating varying customer purchase behavior.
- Certain peaks suggest specific periods of higher sales activity, potentially due to promotions, seasonal trends, or other factors.

### **Conclusion:**

- Focus on identifying the drivers behind these peaks (e.g., festive seasons, marketing campaigns).
- Strengthen marketing efforts during low-sales periods to maintain consistency.

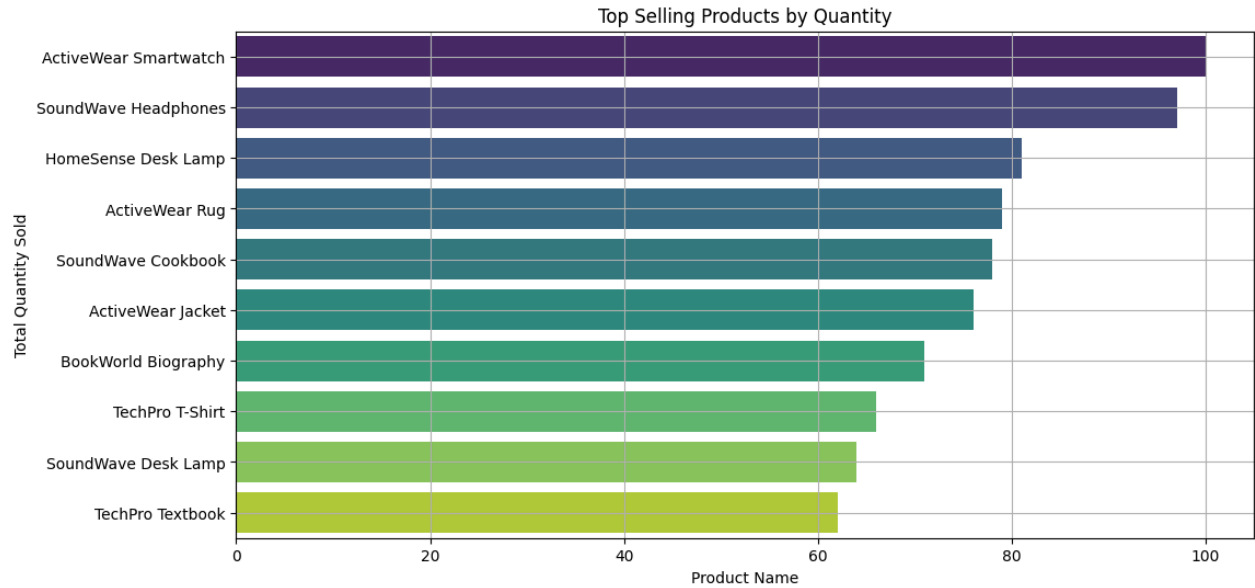


Figure: Top Selling Products by Quantity

- The top-selling products include a mix of categories (e.g., Electronics and Books), with the most popular product significantly outselling others.
- The distribution of quantities shows a steep drop after the top few products, indicating a highly concentrated demand.

**Conclusion:**

- Prioritize stocking and promoting these top products to maximize revenue.
- Analyze why some products have lower sales and assess if repositioning, discounts, or better marketing strategies could help.

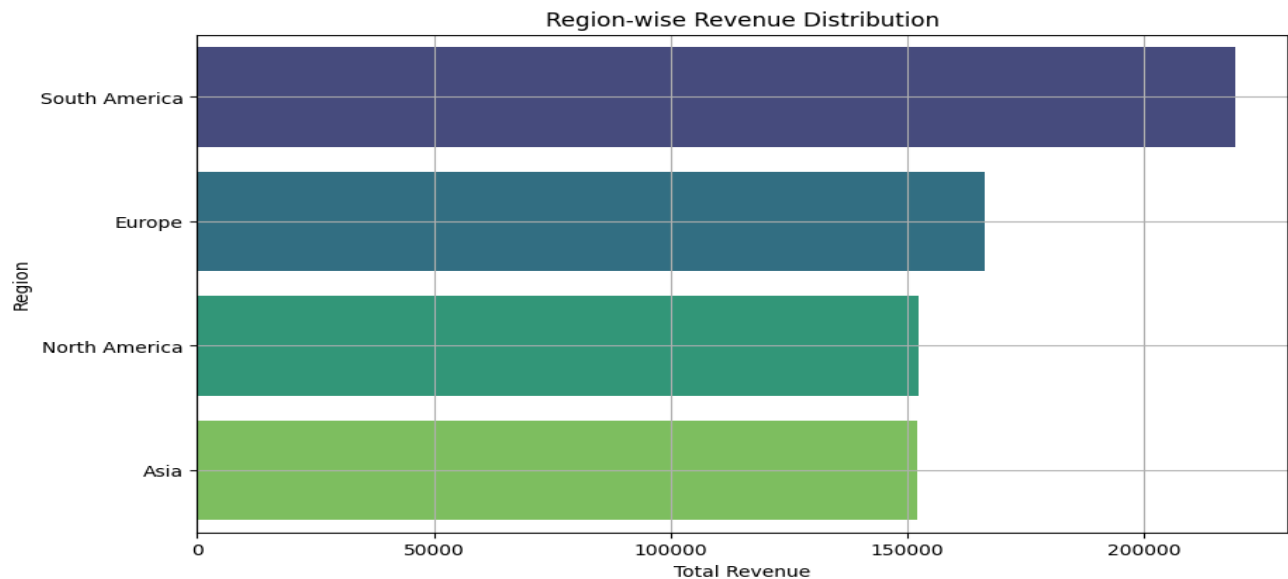


Figure: Region-Wise revenue distribution

- Revenue is unevenly distributed across regions, with one or two regions contributing the most to total sales.
- Regions with lower revenue may represent untapped or underperforming markets.

**Conclusion:**

- Focus on high-performing regions to maintain momentum.
- Investigate low-performing regions for potential improvement through tailored marketing, distribution strategies, or customer engagement.

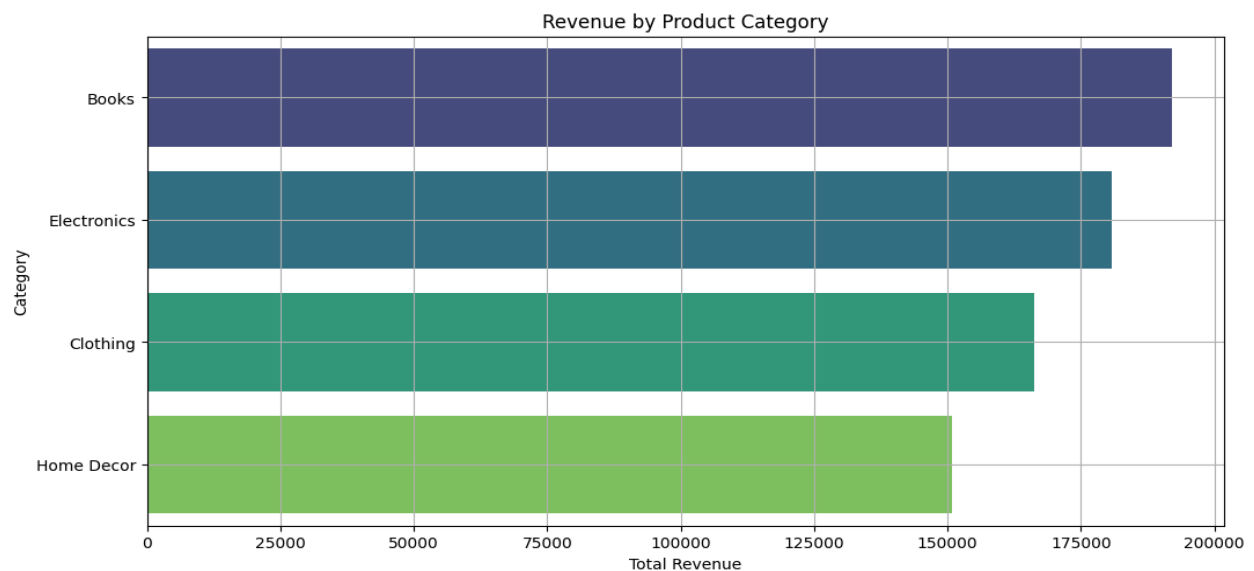


Figure:Revenue by Product Category

- Categories such as "Electronics" or "Books" (depending on the dataset) contribute the most to total revenue.
- Other categories show lower revenue, indicating either niche markets or less popular products.
- Strengthen product offerings and marketing in high-revenue categories to capitalize on demand.
- Reevaluate low-revenue categories to understand if the issue lies with pricing, demand, or competition.



- The business is highly dependent on a small group of high-value customers.
- Focus on **retaining these top customers** through loyalty programs, personalized offers, and excellent customer service.

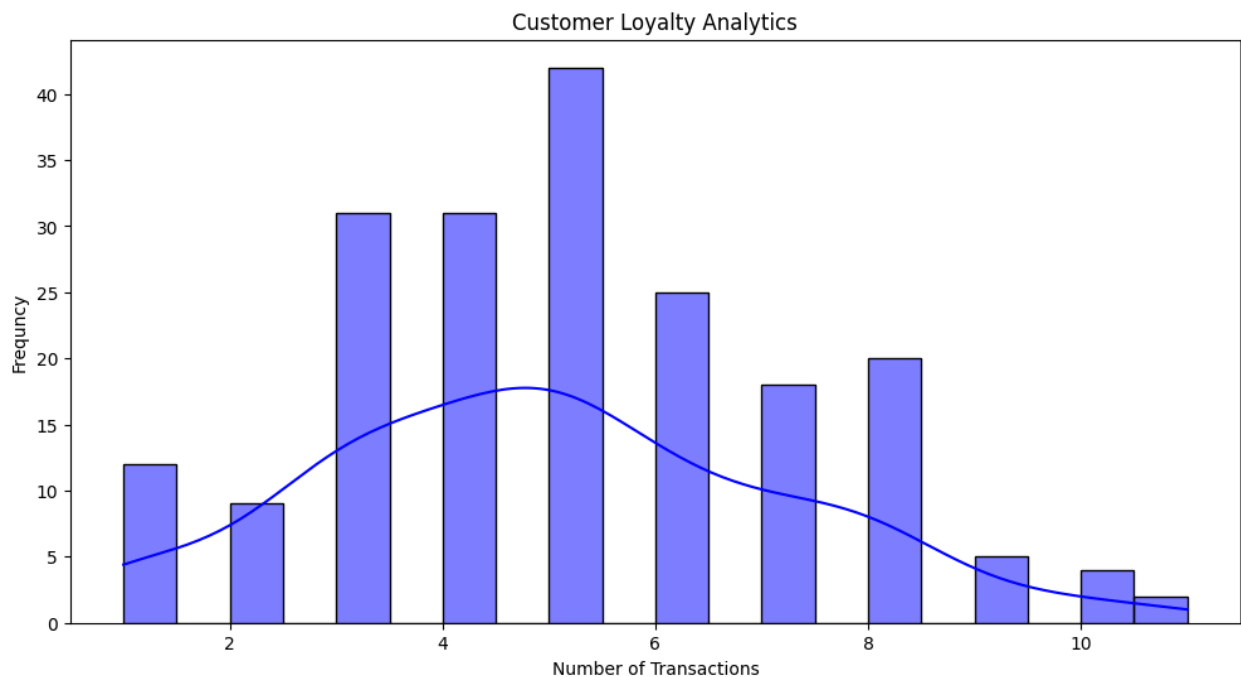


Figure:Customer Loyalty Analytics

- The business needs to focus on converting **one-time buyers into repeat customers** through targeted retention strategies (e.g., discounts for returning customers, subscription models, or loyalty programs).

- The loyal customers with higher transaction frequencies are key to sustainable revenue. Engage them further with exclusive benefits and personalized offers to maintain their loyalty.