

Exploratory Data Analysis

Jaiditya Nair

Introduction

The objective of this task was to perform an exploratory data analysis (EDA) on the eCommerce transactions dataset, consisting of three files:

Customers.csv, **Products.csv**, and **Transactions.csv**. The goal was to gain insights into the structure of the data, identify any patterns or trends, and explore relationships between customers, products, and transactions.

Data Overview

- **Customers.csv**: Contains information about customers, including their IDs, names, region, and signup dates.
- **Products.csv**: Contains details about the products, including product IDs, names, categories, and prices.
- **Transactions.csv**: Contains transaction data, including transaction IDs, customer IDs, product IDs, quantities, and total values.

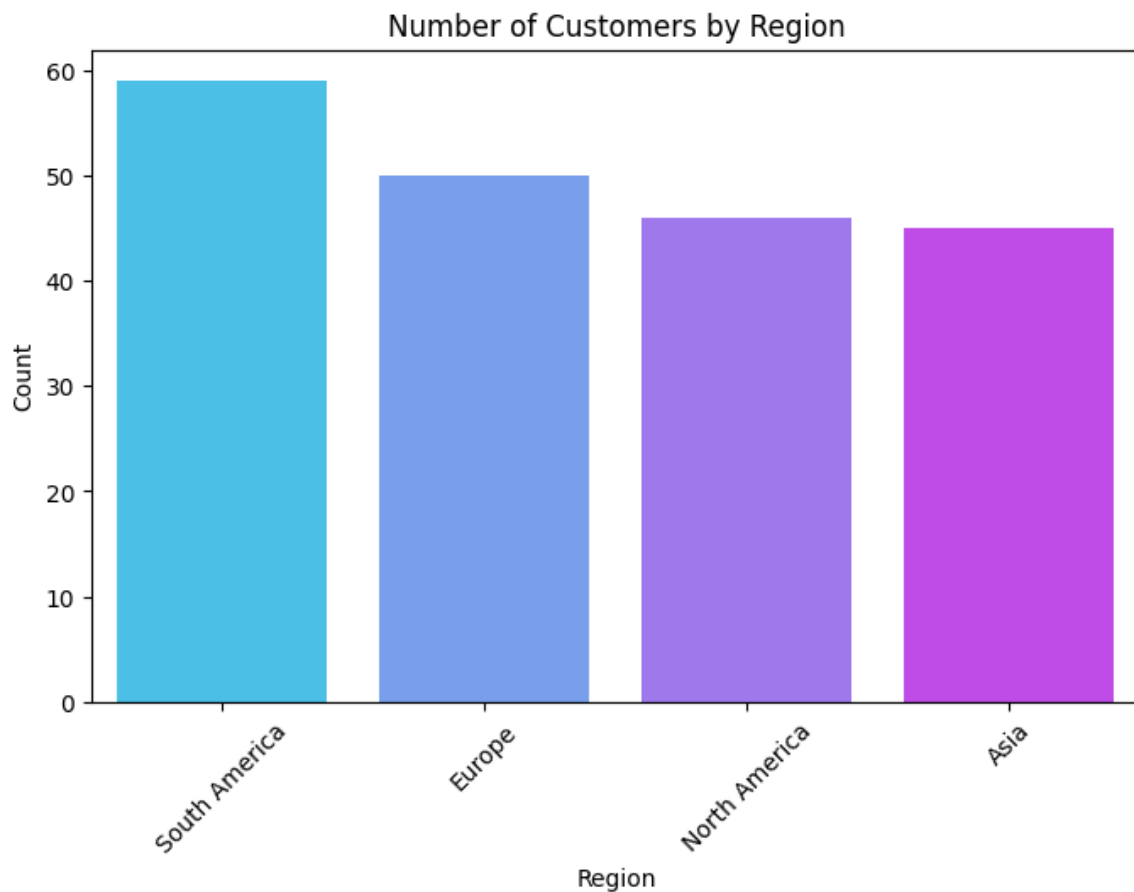
Data Preprocessing

- Missing Values:
 1. Checked for missing data in each dataset and found minimal missing values in the `Region` column of `Customers.csv`.
 2. Imputed missing values for regions with the mode (most frequent) region.
- Data Type Verification:
 1. Ensured that the data types for each column were appropriate (e.g., `CustomerID`, `ProductID`, `TransactionID` as strings, and `Price`, `Quantity`, `TotalValue` as numeric).
- Data Merging:
 1. Merged the three datasets based on common keys (`CustomerID` and `ProductID`) to create a consolidated dataset for analysis. This combined data allowed for richer insights by linking customer details with product purchases.

Key Findings from the EDA

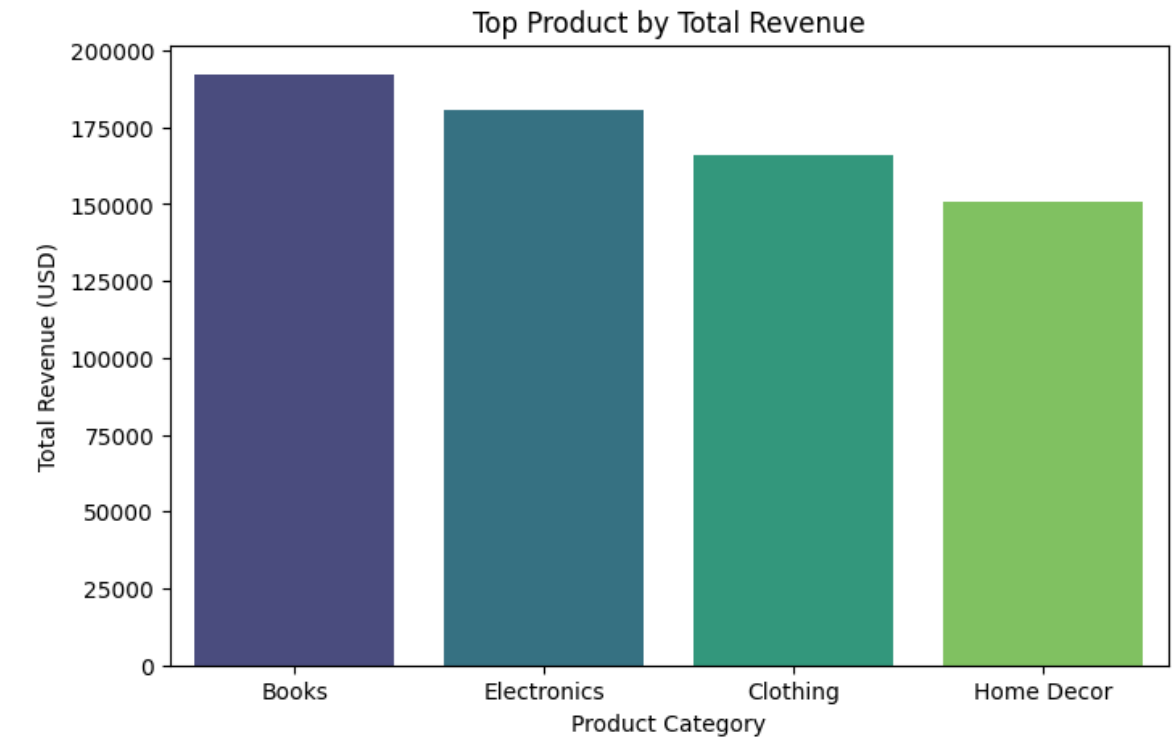
1. Customer Distribution by Region:

- Customers are distributed across several regions, with the largest customer base coming from **South America**. A noticeable portion of customers also come from **Europe** and **North America**.
- **Insight:** Targeted marketing strategies could be tailored for different regions to increase engagement and sales.



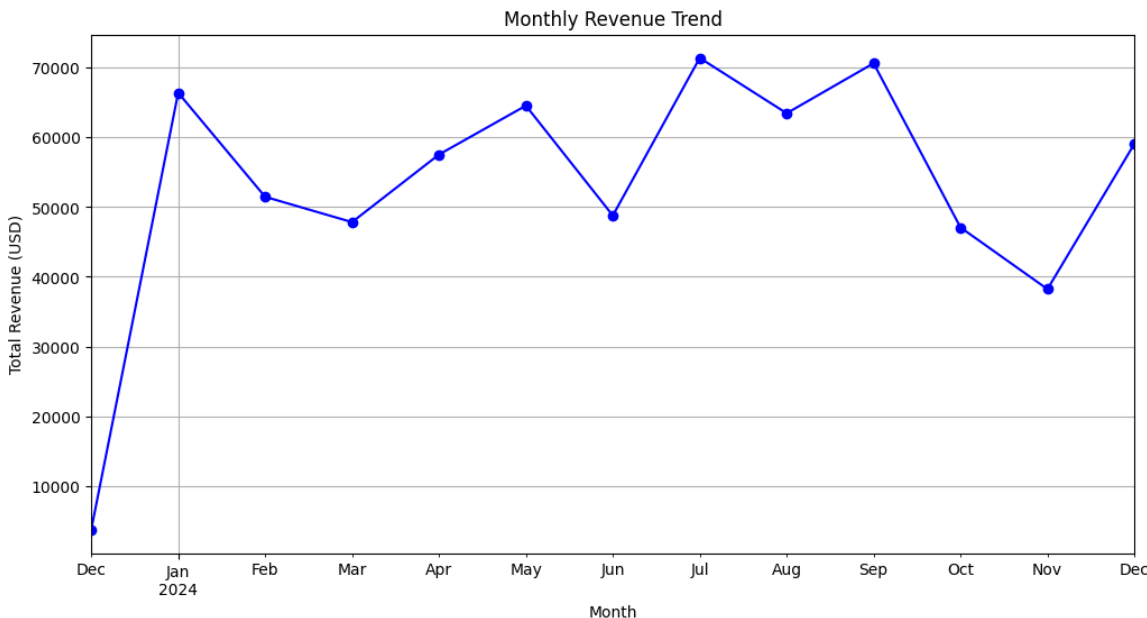
2. Product Revenue Distribution:

- The product dataset reveals that the most revenue generating product is **Books**.
- **Insight:** Business focus could be directed towards these top-selling product, particularly for promotions and cross-selling opportunities.



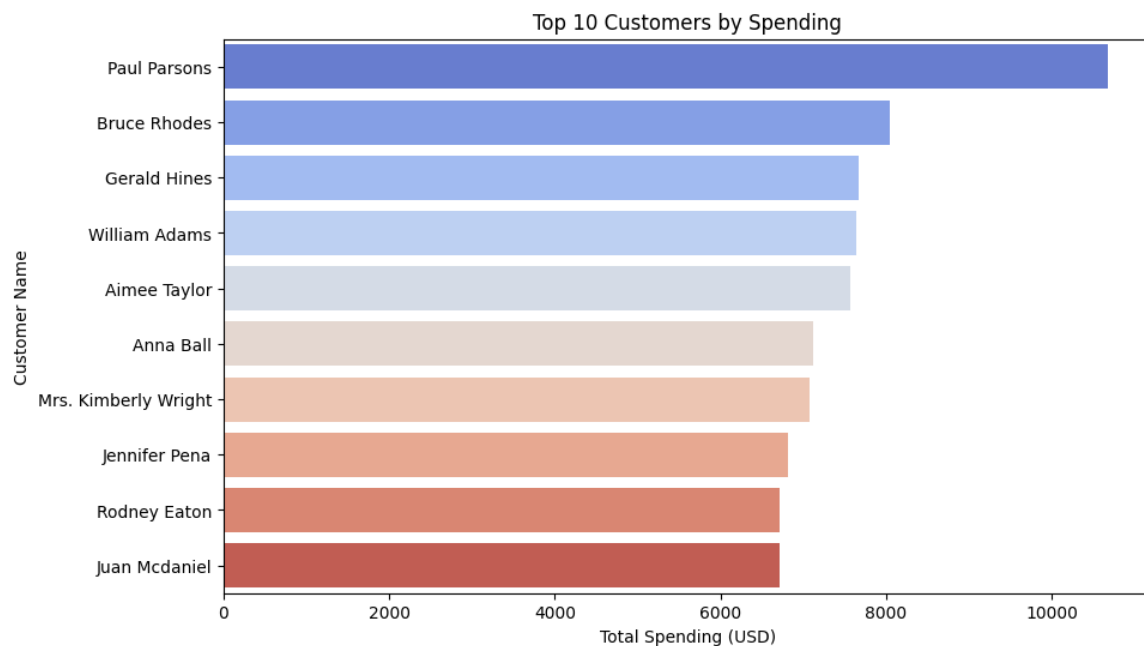
3. Monthly Revenue Trend:

- Display the monthly revenue being generated over the year.
- **Insight:** Shows how the company is performing over the year.



4. Top 10 Customers by Spending:

- Displays the customers spending the most amount of money.
- **Insight:** Can be used to analyse customer and their need for different products.



4. Total Quantity vs Price Range:

- Displays the quantity of products sold across price ranges.
- **Insight:** Can be used for inventory management.

