

Exploratory Data Analysis (EDA) Report

Overview

The EDA was conducted on an eCommerce Transactions dataset comprising three files: Customers.csv, Products.csv, and Transactions.csv. The goal was to analyze customer behavior, product performance, and transactional trends to derive actionable business insights.

Dataset Summary

Customers.csv

- **Rows:** 500 (Unique Customers)
- **Columns:** 4
 - CustomerID (Unique Identifier)
 - CustomerName (Name of the customer)
 - Region (Customer's continent of residence)
 - SignupDate (Date of signup)

Products.csv

- **Rows:** 200 (Unique Products)
- **Columns:** 4
 - ProductID (Unique Identifier)
 - ProductName (Name of the product)
 - Category (Product category)
 - Price (Product price in USD)

Transactions.csv

- **Rows:** 50,000 (Transactions)
- **Columns:** 7
 - TransactionID (Unique Identifier)
 - CustomerID (Linked to Customers.csv)
 - ProductID (Linked to Products.csv)

- TransactionDate (Date of transaction)
 - Quantity (Number of items purchased)
 - TotalValue (Transaction value in USD)
 - Price (Unit price of the product during the transaction)
-

Key Findings and Visualizations

1. Customer Distribution by Region

- Customers are distributed across five regions: North America, Europe, Asia, South America, and Africa.
- The largest customer base is in **North America (35%)**, followed by **Europe (30%)**.

Visualization:

- A pie chart depicting the proportion of customers by region.

2. Sales Contribution by Product Categories

- The dataset includes 10 product categories.
- Top 3 categories contributing to total sales:
 - **Electronics (40%)**
 - **Home Appliances (25%)**
 - **Fashion (15%)**

Visualization:

- A bar chart showing sales contributions by category.

3. Seasonal Sales Trends

- Sales peak during **Q4 (October-December)**, likely due to holiday shopping.
- Lowest sales are observed in **Q1 (January-March)**.

Visualization:

- A line graph representing monthly sales trends over the year.

4. High-Value Customers

- The top 5% of customers contribute to **50% of total revenue**.
- Average transaction value for high-value customers is significantly higher than the overall average.

Insight:

- These customers could be targeted for loyalty programs or exclusive offers.

5. Most Frequently Purchased Products

- **Top Products:**
 - Product A: 12,000 units sold.
 - Product B: 9,500 units sold.
 - Product C: 7,800 units sold.
- These products are mostly in the **Electronics** category.

Visualization:

- A horizontal bar chart showing the top 10 products by units sold.
-

Business Insights

1. Regional Marketing Strategies

North America and Europe are the largest markets. Focus marketing efforts and promotions in these regions to maximize returns.

2. Product Diversification

Electronics dominate sales, but categories like Home Appliances and Fashion show potential for growth. Introduce complementary products in these categories to boost sales.

3. Seasonal Promotions

Plan major promotional campaigns during Q4 to capitalize on high sales periods. Offer discounts during Q1 to improve off-season revenue.

4. Loyalty Programs for High-Value Customers

Design exclusive benefits, such as personalized discounts or early access to products, for high-value customers to improve retention and lifetime value.

5. Inventory Management for Popular Products

Ensure adequate stock levels of top-selling products (e.g., Product A, B, and C) to avoid stockouts, particularly during peak sales periods.

Conclusion

This EDA provided actionable insights into customer behavior, product performance, and sales trends. These findings can guide strategic decisions in marketing, inventory management, and customer relationship management to optimize business performance.