A Report on

Exploratory Data Analysis and Business Insights for eCommerce Transactions Dataset

Submitted By:

Alaina Umme Hafiza

B.Tech in Computer Science Engineering

CMR University

Submitted To:

Zeotap India Pvt. Ltd.

Introduction

The objective of this report is to analyze an eCommerce transactions dataset comprising customer, product, and transaction details. Through exploratory data analysis (EDA), we aim to uncover patterns, trends, and actionable business insights that can help optimize marketing, sales, and inventory strategies.

The dataset consists of three files:

1. Customers.csv: Contains customer demographic details such as region, signup date, and customer ID.
2. Products.csv: Includes product categories, names, and pricing information.
3. Transactions.csv: Details sales transactions, including transaction dates, quantities, and total values.

By merging and analyzing these datasets, we aim to identify high-value customers, best-performing product categories, regional sales contributions, and seasonal trends. The findings are expected to aid decision-making and improve overall business performance.

EDA – Exploratory Data Analysis Steps

1. Data Loading and Preparation

- Imported the datasets (Customers.csv, Products.csv, and Transactions.csv) using Python libraries such as pandas and numpy.
- Checked for missing values, duplicates, and inconsistencies in each dataset.
- Merged the three datasets into a single, cohesive dataset using common keys (CustomerID and ProductID).

2. Univariate Analysis

- Customers.csv:
    - Analyzed the regional distribution of customers to identify the top-performing regions.
    - Examined customer signup trends over time.

- Products.csv:
    - Studied the distribution of product prices and identified high-value products.
    - Evaluated the popularity of different product categories.
- Transactions.csv:
    - Assessed transaction frequency, total revenue, and purchase quantities.

3. Time-Series Analysis

- Analyzed TransactionDate to identify sales trends over time (e.g., monthly or quarterly trends).
- Highlighted seasonal peaks in sales.

4. Visualization

Created visualizations using libraries such as matplotlib and seaborn:

- Bar charts for regional sales contributions and product category performance.
- Line graphs for time-series analysis of sales.
- Box Plot representing the Price_x and Price_y from transactions.

Business Insights from EDA

Insight 1: Regional Sales Contribution

- **Analysis**: The sns.countplot visualization of transactions by customer region shows the distribution of customers across different regions.
- **Insight**: Regions with the highest number of customers or transactions should be prioritized for marketing campaigns and inventory optimization. Regions with fewer customers can be targeted with localized offers to improve sales.

Insight 2: Product Categories Performance

- **Analysis**: The sns.countplot visualization of product categories highlights the most and least popular product categories in terms of transactions.
- **Insight**: Focus inventory and promotions on the top-performing categories to maximize revenue. For underperforming categories, evaluate their pricing, marketing, or relevance to customer needs.

Insight 3: Monthly Sales Trends

- **Analysis**: The sns.countplot of transactions grouped by Month identifies monthly trends in sales activity.
- **Insight**: Sales peaks during certain months (e.g., due to holidays or promotions) indicate opportunities to plan seasonal campaigns. Low-sales months could benefit from targeted offers or bundled promotions to drive demand.

Insight 4: Product Pricing Variability

- **Analysis**: The box plots for Price_x (original product price) and Price_y (transaction price) reveal the distribution of product prices and transaction prices, including outliers.
- **Insight**: A significant difference between original prices and transaction prices suggests discounts or promotions. Businesses can evaluate whether these discounts effectively boost sales while maintaining profitability. Outliers in pricing may indicate premium products or data anomalies.

Insight 5: Customer Signup Trends

- **Analysis**: Converting SignupDate into a datetime format and analyzing customer acquisition trends over time can reveal when most customers joined.
- **Insight**: Peak signup periods suggest effective marketing campaigns during that time, which could be replicated. For slower periods, new strategies should be tested to increase customer acquisition.

Insight 6: Customer Value

- **Analysis**: By merging datasets, total spending by each customer can be calculated, helping identify high-value customers (those with high total transaction values).
- **Insight**: High-value customers should be rewarded with loyalty programs and personalized offers to retain them. These customers likely contribute a significant portion of the revenue.

Conclusion

The exploratory data analysis of the eCommerce transactions dataset provided valuable insights into customer behavior, product performance, and sales trends. The key findings highlight the importance of focusing on high-performing regions and product categories, leveraging seasonal sales trends, and retaining high-value customers through targeted loyalty programs. Additionally, analyzing pricing adjustments revealed opportunities to optimize promotional strategies while maintaining profitability.

By implementing these insights, the business can enhance its marketing, inventory, and pricing strategies to drive growth and improve overall efficiency. Continuous monitoring and analysis will be essential to adapt to changing customer preferences and market dynamics.