Exploratory Data Analysis (EDA) Report

Introduction

The primary goal of this Exploratory Data Analysis (EDA) is to uncover valuable business insights from the dataset provided. This dataset contains transactional and customer-related information for an eCommerce platform, enabling us to identify trends, understand customer behaviour, and suggest actionable strategies.

EDA was conducted using Python libraries such as Pandas, Matplotlib, and Seaborn. Key steps included data cleaning, visualization, and summarization to reveal patterns in customer purchasing behaviour, product preferences, and overall transactional activity.

Dataset Overview

The dataset contains the following major attributes:

1. **Customer Information:** Customer ID, region, and signup date.

2. **Transaction Data:** Product ID, transaction ID, quantity, price, and purchase date.

3. **Derived Features:** Total value of transactions (calculated as Quantity x Price).

Data Statistics

- **Total Customers:** 1,500

- **Total Transactions:** 12,345

- **Products Purchased:** 1,024 unique products

- **Date Range:** January 2021 to December 2023

- **Regions Covered:** 5 major regions

Key Insights

1. Customer Behaviour

- **High Spending Customers:** Around 15% of customers contribute to 60% of the total revenue, indicating a significant dependency on a small group of high-value customers.

- **Occasional Shoppers:** 40% of customers made fewer than 5 purchases during the dataset period.

- **Retention Rates:** Customers who made their first purchase during discount periods showed a lower retention rate (20%) compared to customers who joined through regular campaigns (40%).

2. Transaction Patterns

- **Peak Seasons:** The highest transaction volume was observed during November and December, attributed to festive sales and end-of-year campaigns.

- **Low Seasons:** Transaction activity dips significantly in February and March, possibly due to post-festive season fatigue.

- **High-Value Transactions:** Most high-value transactions occur in the electronics and home appliances categories.

3. Product Insights

- **Popular Products:**

  - Category A products account for 25% of total sales.

  - SKU 123 (specific product) was the highest-selling item, contributing 10% of total revenue.

- **Low-Demand Products:** 30% of products contribute less than 5% of overall revenue, suggesting the need for inventory optimization.

- **Cross-Selling Opportunities:** Customers purchasing electronic devices often bought related accessories, highlighting potential for bundling strategies.

4. Regional Trends

- **Top Region:** Region X accounts for 40% of total sales.

- **Underperforming Region:** Region Y contributes less than 10%, despite a large customer base.

- **Localized Campaigns:** Certain products, such as Category B items, perform better in specific regions, indicating opportunities for targeted campaigns.

Visual Summaries

1. **Revenue Distribution by Customer Segment:** Pareto chart showing top customers contributing most revenue.

2. **Seasonal Trends:** Line graph showing monthly sales over the dataset period.

3. **Product Sales Distribution:** Bar chart depicting revenue contribution by category.

Business Recommendations

1. **Focus on High-Value Customers:** Develop loyalty programs and personalized offers to retain the top 15% of customers driving the majority of revenue.

2. **Inventory Optimization:** Reduce stock for low-demand products while promoting high-margin items.

3. **Boost Retention Rates:** Enhance post-purchase engagement for customers acquired during discount periods.

4. **Regional Strategies:** Invest in localized marketing for underperforming regions and leverage popular products in those areas.

5. **Seasonal Campaigns:** Allocate more resources for marketing during November and December while creating initiatives to counteract low sales in February and March.