

Exploratory Data Analysis (EDA) and Business Insights

This report summarizes insights derived from the exploratory data analysis (EDA) of three datasets: Customers, Products, and Transactions. The analysis focused on customer behaviors, product pricing, sales trends, and regional performance.

First I checked if there are any missing values in the dataset or not. Then I converted object datatype to int datatype. I also checked for any duplicate values present in the dataset.

Key Findings and Business Insights

1) Customer Data Analysis

Signup Trends:

- The maximum number of customer signups occurred in April during all three years.
- if we see year wise, then the most number of signups happened in 2024.

Region-wise Distribution:

- Customers are concentrated in regions like South America and Europe, while Asia has lower penetration.

2) Product Data Analysis

Category Popularity:

- Categories such as Books and Electronics have the highest number of products. These categories can be marketed to maximize variety and customer choice.

Pricing Trends:

- Count: There are 100 products in the dataset
- Mean: The average price of the products is 267.55
- Standard Deviation: The price variability is 143.22. A high standard deviation indicates that product prices are spread out over a wide range
- Minimum: The cheapest product costs 16.08
- 25% (1st Quartile): 25% of the products are priced below 147.77
- 50% (Median): Half of the products are priced below 292.88
- 75% (3rd Quartile): 75% of the products are priced below 397.09
- Maximum: The most expensive product costs 497.76

Wide Price Range: The prices range from 16.08 to 497.76, indicating products for both budget-conscious and premium customers.

Affordability: Products below the 25% quartile (147.77) are affordable options.

Premium Products: Products priced above the 75% quartile (397.09) are premium options, possibly targeting high-end customers.

3) Transaction Data Analysis

Monthly Revenue Trends: Monthly revenues peak in July, reflecting increased customer spending during these times. This finding suggests the importance of seasonal promotions and inventory planning.

Top-Selling Products: The top 10 products contributed significantly to total sales. These products should be prioritized in promotions to sustain their sales momentum.

Product Id of top 3 selling products -

- P059
- P054
- P029

Customer Purchase Frequency: Total of 42 customers made 5 transactions, indicating their loyalty. Offering targeted discounts or loyalty programs to this segment could further increase retention.

Regional Revenue: Regions like South America generate the highest revenue, emphasizing the need to replicate the strategies used there in other regions.

4) Combined Dataset Insights

Revenue by Category: Categories such as Books contributed the most to overall revenue. Expanding offerings in these categories could drive growth.

Customer Lifetime Value (CLV): Customer Id C0141, C0054, C0065 has the highest CLV. Businesses should develop retention strategies like exclusive offers for these high-value customers.

Customer-Product Interactions: Heatmap analysis reveals strong customer-product interactions, enabling personalized recommendations to boost sales.

5) Conclusion

The EDA reveals insights into customer behavior, product trends, and revenue opportunities, guiding businesses to refine strategies for growth, satisfaction, and profitability.