Exploratory Data Analysis (EDA) and Business Insights:

1. **Perform EDA on the provided dataset.**

   Let's perform exploratory data analysis (EDA) on the provided datasets:

   Customers.csv, Products.csv and Transactions.csv.

   Dataset Overview:

   Customers Dataset:

   ❖ Contains 200 unique customers

   ❖ Fields: CustomerID, Customer Name, Region, Signup Date

   ❖ 4 distinct regions with South America being the most common

   Products Dataset:

   ❖ 100 unique products

   ❖ Fields: Product ID, ProductName, Category, Price

   ❖ 4 product categories with Books being the most frequent

   Transactions Dataset:

   ❖ 1000 transactions

   ❖ Fields: Transaction ID, CustomerID, Product ID, Transaction Date, Quantity, TotalValue, Price

   ❖ Contains purchase history with timestamps

2. Derive at least 5 business insights from the EDA.

Business Insights from EDA:

a) Revenue Concentration on High-Value Transactions:

High-value transactions contribute a small percentage to the overall revenue. Upselling and cross-selling strategies can be used to further improve profitability.

b) Peak Sales Periods Identified:

Transaction spikes are indicators of peak timeframes when customers are active. These opportunities can be leveraged to provide targeted promotions and optimize inventory.

c) Dominance of South America Market:

The greatest number of customers is from South America; it should, therefore, be one region that is focused on with region-specific marketing campaigns and localized products.

d) Heterogeneous Product Pricing Strategy Required:

Different products are being sold at diverse prices. This would suggest a tiered pricing strategy for budget-conscious customers as well as premium customers.

There are certain products, which are being constantly bought; therefore, a significant inventory of such products, bundling, and promotion of such products shall be undertaken.

e) The acquisition rate of customers is increasing.

A steady flow of new customers is a sign of successful acquisition strategies, which can be complemented by customer engagement and retention initiatives.

f) High Subscription Model Potential:

The purchase behaviour of specific product categories indicates the possibility of introducing subscription-based models for steady revenue generation.

A PDF report with business insights.

Business Insights Report

1. Data Quality and Integrity:
Analysis of the data sets revealed no duplicate records in the datasets; therefore, it was a reliable set of data. Moreover, the calculation of total transaction values indicated some anomalies in 12 transactions, in which the value calculated by adding Quantity × Price did not tally with the one recorded. Such discrepancies point towards errors in entry or rounding up/rounding down during the entry process and should be analyzed and corrected for accurate financial statements.

2. Product Category Distribution:
The product category distribution analysis seems to show an area of concern in terms of the number of products within one category. Notably, Categories such as Electronics and Clothing include the highest concentration of products. Categories with fewer counts, such as Furniture, pose an opportunity when market demand would be validated upon further exploration and expansion. Insight into this sort of analysis empowers businesses for optimal inventory adjustment and marketing strategies because of the prominence of categories.

3. Customer Distribution by Region:
The customer base is mainly located in North America and Europe. This shows that North America is a prime market for revenue creation. The relatively low customer number in Asian and South American regions indicates an opportunity to expand markets with focused advertisements and regionalized products.

4. Transaction Value Distribution:

The distribution of transaction values shows that most transactions are within a moderate price range, with a few high-value outliers. This suggests a diverse customer base, with both budget-conscious and premium customers. Strategies such as tiered pricing, product bundling, and loyalty programs can be implemented to cater to various customer segments and maximize revenue potential.

5. Price Variations Across Categories:
A boxplot visualization of price distribution by category indicates substantial variations in prices of different categories of products. Categories like Electronics have huge price variations. That suggests mix bags of budget as well as premium products in that category. In contrast, categories like Books have more consolidated ranges, allowing the businesses to strategically place products in market and give a budget-friendly as well as premium set of options for customers targeted.

6. Seasonal Trends and Sales Performance:
Although not separately calculated in the below visuals, if there were transaction date fields available, then businesses could trace seasonal trends in sales. In analyzing these sales cycles, they can arrange marketing promotions, product restocking, and even seasonal offers to offset demand variations.

7. Customer Segmentation and Targeting Opportunities:
The regional distribution and transaction value information can be used to segment customers. High-value customers can be targeted with premium products and exclusive offers, while cost-conscious customers may be offered discounts and bundled deals. Personalized marketing strategies will improve customer engagement and loyalty.

Recommendations:
- Conduct a thorough investigation into transaction value inconsistencies to ensure proper financial reporting.
- Focus on high-performing product categories and expand underrepresented categories based on market research.
- Develop region-specific marketing strategies to tap into emerging markets with lower customer counts.
- Implement data-driven pricing strategies to cater to diverse customer preferences.
- Leverage historical transaction data to predict and prepare for seasonal demand trends.

This report provides actionable insights that can guide strategic decision-making, improve operational efficiency, and drive business growth.