Business Insights from EDA of eCommerce Transactions Dataset

1. **Product Demand:**

   o **Insight**: Products with higher prices generally have fewer transactions, suggesting that customers prefer more affordable options. Retailers can consider offering discounts or promotional pricing to boost sales of premium products.

   o **Enhancement**: The data imputation process has ensured that missing price data doesn't distort the analysis, making the price-to-sales correlation more reliable.

2. **Region-Based Sales:**

   o **Insight**: The customer distribution by region reveals that the majority of customers are from North America, with significant contributions from Europe and Asia. This insight can guide region-specific marketing strategies and product offerings.

   o **Enhancement**: The automated handling of missing data in the customer regions allows for a more accurate regional distribution analysis, ensuring all regions are properly represented in visualizations.

3. **Pricing and Sales Correlation:**

   o **Insight**: There is a weak negative correlation between product price and the number of units sold, indicating that cheaper products tend to sell more frequently. This suggests that low-priced products are more popular among customers.

   o **Enhancement**: By using the SimpleImputer, missing values in price data were handled seamlessly, providing a cleaner and more precise correlation analysis between price and sales volume. This allows for more informed pricing strategies.

4. **Customer Retention:**

   o **Insight**: Newer customers (those who signed up recently) show a higher total transaction value, which implies that the customer base is growing and customers are making frequent purchases. Businesses can focus on retaining these newer customers by offering loyalty rewards or personalized offers.

   o **Enhancement**: Through automated imputation, even customers with incomplete transaction data were included in the analysis, ensuring that retention insights were based on comprehensive and accurate data.

5. **Seasonal Trends:**

   o **Insight**: A noticeable peak in sales occurred during the holiday season, particularly from November to December. This suggests that sales are driven by seasonal factors, and targeted campaigns during these months could further boost sales.

   o **Enhancement**: The clean, imputed data allowed for more detailed seasonal analysis without the skewing of missing values, ensuring that holiday trends are correctly represented in visualizations.

Key Features and Enhancements:

- **Automatic Missing Data Imputation**:

  - Utilized SimpleImputer for handling missing values in numerical columns across multiple datasets (customers, products, transactions), ensuring completeness in analysis without manual intervention.

- **Comprehensive Data Visualizations**:

  - In addition to standard visualizations like histograms and scatter plots, the report leverages imputed data to present accurate, region-specific, and seasonal insights.

  - Visualizations such as **Product Price Distribution**, **Quantity Sold vs Price**, and **Region-Based Customer Distribution** provide in-depth views that help businesses make informed decisions on pricing, customer segmentation, and regional marketing efforts.

- **Data Quality Checks**:

  - Imputation ensures that missing data in crucial columns like price, quantity, and customer region doesn't affect the overall analysis, providing accurate insights for decision-making.