

**Name:** Mirdul Agrawal  
**Role:** Data Science Intern (Zeotap)  
**Email Id :** [mridulagarwal8998@gmail.com](mailto:mridulagarwal8998@gmail.com)  
**Phone:** 9696175898

## **Task 1: Exploratory Data Analysis (EDA) and Business Insights**

### **I. Introduction**

This report provides an analysis of three datasets: Customers, Products, and Transactions, with the goal of deriving actionable business insights. The analysis includes cleaning, merging, and visualizing the data to identify trends, customer behaviors, and revenue-generating patterns. The findings aim to guide data-driven decision-making for enhanced business performance.

### **II. EDA Process**

#### **Data Cleaning and Preprocessing:**

- The datasets were inspected for missing values, duplicates, and incorrect data types.
- Missing and duplicate records were negligible or non-existent, ensuring data quality.
- Date columns (SignupDate, TransactionDate) were converted to datetime formats for better handling.

#### **Merging Datasets:**

- The Transactions dataset was merged with the Customers and Products datasets using CustomerID and ProductID as keys.
- The resulting merged dataset enabled analysis across customers, products, and transaction details.

#### **Key Data Visualizations:**

- Customer Distribution by Region: A count plot displayed the number of customers per region, revealing the largest customer base in specific regions (e.g., Asia or Europe).
- Product Price Distribution: A histogram showed the spread of product prices, highlighting common price ranges.
- Top-Selling Products: A bar chart identified the most purchased products by quantity.
- Total Revenue by Region: A bar plot showed the revenue contribution of each region.
- Monthly Revenue Trend: A line plot visualized revenue trends over time, identifying seasonal peaks and troughs.

### III. Key Insights

- a) **Regional Customer Distribution:** The majority of customers are concentrated in specific regions, such as Asia and Europe, suggesting the need to tailor marketing efforts and strategies for these high-density regions.
- b) **Revenue Distribution Across Regions:** Revenue contributions varied by region, with certain regions generating significantly higher sales. This insight indicates potential for expansion in underperforming regions or focusing on high-performing ones for further growth.
- c) **Top-Selling Products:** The top 10 products contributed significantly to total sales volume. This insight underscores the importance of these products in revenue generation, suggesting a focus on inventory optimization for these items.
- d) **Product Price Trends:** The price distribution showed most products are priced within a specific range (e.g., \$20–\$50). This suggests a sweet spot for pricing strategies to maximize sales while remaining competitive.
- e) **Seasonal Revenue Trends:** The monthly revenue trend highlighted seasonal peaks during specific periods, such as holiday seasons. These trends provide opportunities for targeted promotions and stock management.

### IV. Conclusion

The analysis of the Customers, Products, and Transactions datasets reveals clear opportunities for enhancing regional strategies, optimizing inventory, and leveraging seasonal trends. By acting on these insights, the business can improve profitability and customer satisfaction while maintaining a competitive edge in the market.