Data Science Assignment: eCommerce Transactions Dataset

Task 1

Step – 1:
- Loaded and Cleaned Datasets: Imported the three datasets (Customers.csv, Products.csv, and Transactions.csv) and removed missing or inaccurate values by replacing them with mean or median values.
- Checked for Null Values: Verified the datasets for any null values and handled them to ensure completeness.
- Exploratory Data Analysis (EDA): Conducted initial analysis on the datasets to check for data consistency, structure, and readiness for further analysis.

Business Insights:

1. **Price Distribution Analysis: (BOXPLOT)**
   - The median product price is just below 300 USD, indicating a central tendency for affordable pricing.
   - The presence of outliers suggests a mix of both premium and budget products in the catalogue.
   - A significant portion of products is priced below the median, suggesting a focus on affordability.
   - The spread of prices indicates varying customer preferences, ranging from budget-conscious to premium buyers.
   - Understanding the price distribution can guide promotional strategies and identify opportunities for price adjustments.

2. **Category-wise Product Distribution:**
   - Certain product categories dominate the dataset, showing areas of high market concentration.
   - Categories with the most products may signify a business focus on those segments.
   - Analysing category distribution helps target inventory and marketing efforts toward high-demand areas.
   - Low-product categories may need more attention in product development or marketing to increase variety.
   - This insight aids in refining product portfolios by evaluating the performance and potential growth of each category.

3. **Transaction Trends over Time:**
   - Total transaction values fluctuate daily, with spikes on certain dates, indicating periodic demand.
   - Sales peaks correspond with specific dates, potentially linked to promotions or seasonality.
   - Identifying transaction spikes helps forecast future sales patterns for better planning.
   - This trend can guide inventory decisions to ensure products are available during high-demand periods.
   - The information is key for optimizing marketing efforts, timing sales, and planning promotional activities.

4. **Most Sold Product Analysis:**
   - The most sold product is identified by total quantity sold, revealing customer preferences and high demand.
   - Analysing customer purchases of this product can reveal patterns of repeat buyers and loyal customers.

- Transaction details for this product help identify high-value customers, valuable for personalized marketing.
- Tracking the total value over time for this product shows fluctuations in demand, guiding promotional planning.
- Understanding the most sold product can drive inventory management and targeted sales strategies for similar products.

5. **Daily Transaction Total Value:**
   - Peak Performance in September: The highest daily transaction value occurred in **September**, reflecting a significant surge in sales during this month, likely due to seasonal demand or targeted marketing campaigns.
   - Slower Sales in December: Conversely, **December** recorded the lowest daily transaction value, indicating a downturn in sales, which may be attributed to factors like market saturation or reduced customer engagement during the festive season.
   - Seasonal Insights: These variations suggest that the business experienced strong sales performance in the **9th month**, followed by a dip in the **12th month**, which could be useful for planning inventory and marketing strategies in future years

6. **Demand and Purchase Distribution:**
   - The demand analysis identifies both the most and least purchased products, showing areas of high and low interest.
   - The pie chart visualizes the top 5 most purchased products, with contrast to the least purchased, aiding decision-making.
   - This distribution helps in inventory planning, ensuring the best-selling products are always in stock.
   - For least purchased products, businesses can consider promotional activities or discontinuation.
   - Understanding demand distribution enhances marketing strategies, ensuring focus on high-performing products.

7. **Customer Conversion Time Analysis:**
   - The scatter plot shows the relationship between signup date and conversion time, highlighting how long customers take to convert.
   - The analysis reveals that customers signing up earlier tend to convert faster, aiding the identification of patterns.
   - Conversion time trends help improve customer onboarding processes, shortening conversion times.
   - Identifying customers with longer conversion times can help tailor strategies to accelerate their decision-making.
   - This insight allows businesses to optimize their sales funnel, ensuring quicker conversions and improving customer retention.

8. **Total Value Distribution Analysis:**
   - The distribution of total values highlights patterns in how much customers spend during transactions.
   - Peaks in the distribution suggest popular price points where customers are more likely to make purchases.
   - The histogram reveals the range of spending, helping identify which price ranges lead to the highest volume of sales.
   - The analysis can guide pricing strategies, ensuring that most products are priced within the most frequent spending ranges.
   - This insight helps in creating bundles or discounts that align with the spending habits of the majority of customers.