Exploratory Data Analysis (EDA) Report for Task 1

Overview: In this task, we analyzed an eCommerce Transactions dataset to gain meaningful insights about customers, products, and transactions. Below is a summary of the techniques and methods used to perform EDA, along with a brief explanation of their purpose and significance.

---

1. Data Overview and Inspection:

- **Purpose:** To understand the structure, size, and basic characteristics of the dataset.
- **Techniques Used:**
    - info() function: Provided information about the dataset, such as column names, data types, non-null counts, and memory usage.
    - describe() function: Offered statistical summaries of numerical columns (e.g., mean, standard deviation, minimum, and maximum values).
    - Displayed unique values in key columns to understand data distribution.
- **Key Insights:**
    - There were 1,000 transactions, 199 unique customers, and 100 unique products.
    - Data was clean with no missing values.

---

2. Data Cleaning and Merging:

- **Purpose:** To prepare data for analysis by ensuring it is consistent and combined logically.
- **Techniques Used:**
    - Merged Customers, Products, and Transactions datasets using the merge() function based on common keys (e.g., CustomerID, ProductID).
    - Standardized date formats using pd.to_datetime() to ensure proper handling of transaction and signup dates.
- **Outcome:** A single consolidated dataset was created, making analysis more straightforward.

---

3. Summary Statistics:

- **Purpose:** To explore the distribution and variability of numerical columns.
- **Techniques Used:**
    - Analyzed the Quantity, TotalValue, and Price columns to identify trends and anomalies.

- Used groupby() to aggregate revenue by regions, product categories, and customers.

- **Key Insights:**
  - Average transaction value was approximately $689.99, with the highest value at $1,991.04.
  - South America generated the highest revenue among all regions.

---

4. Top-Selling Products and Revenue Analysis:

- **Purpose:** To identify the most popular products and revenue drivers.
- **Techniques Used:**
  - Calculated total quantities sold for each product using groupby() and sum().
  - Sorted products by quantity to determine the top 10 best sellers.
  - Visualized results using a bar chart (matplotlib and seaborn libraries).
- **Key Insights:**
  - The "ActiveWear Smartwatch" was the most sold product with 100 units sold.

---

5. Customer Signup Trends:

- **Purpose:** To examine customer growth over time.
- **Techniques Used:**
  - Extracted signup years from the SignupDate column.
  - Counted customer signups per year to identify trends.
  - Visualized trends using a line chart.
- **Key Insights:**
  - 2024 saw the highest customer signups, indicating recent growth in the customer base.

---

6. Visualizations:

- **Purpose:** To make patterns and trends more interpretable.

- **Techniques Used:**
    - Used bar plots to show top-selling products.
    - Visualized revenue distribution across regions using pie charts.
    - Displayed signup trends over time using line graphs.
- **Outcome:** Key trends and patterns were effectively communicated through easy-to-understand visuals.

Conclusion:

The EDA provided valuable insights into customer behavior, product performance, and regional revenue distribution. These insights can guide future business strategies, such as targeting high-revenue regions, promoting top-selling products, and focusing on customer acquisition during peak signup periods.