**Title:** Data Fusion and Reshaping for Complex Analytical

**Description:** Merge and reshape data from multiple DataFrames to create complex analytical datasets.

**Task 3 Explanation**

**Objective**

The goal of this task is to demonstrate advanced data fusion and reshaping techniques by merging and analyzing multiple datasets, ultimately creating an insightful and actionable Excel report. This analysis involves integrating information from multiple DataFrames, applying conditional transformations, creating new features, and reshaping the data into meaningful formats for decision-making.

**Data Overview**

**Input DataFrames**

1. **Product Data**
   * Contains details about products such as their IDs, names, and categories.
   * Key Columns: ProductID, ProductName, Category.
2. **Sales Data**
   * Records transactional details, including the product sold, customer involved, date of sale, units sold, and revenue generated.
   * Key Columns: SalesID, ProductID, CustomerID, Date, UnitsSold, Revenue.
3. **Customer Data**
   * Provides customer-specific details, such as their ID, region, and type (Regular or Premium).
   * Key Columns: CustomerID, Region, CustomerType.
4. **Discount Data**
   * Lists discounts applicable to specific products for defined periods.
   * Key Columns: ProductID, Discount, StartDate, EndDate.

**Data Processing Steps**

**1. Merging Sales and Product Data**

We merged the sales\_data and product\_data DataFrames using ProductID as the common key. This integration links each sale to the product name and category, enabling category-based analyses.

* **Output:** merged\_sales\_product DataFrame.

**2. Adding Customer Details**

The merged\_sales\_product DataFrame was further merged with the customer\_data DataFrame using CustomerID. This step associates each sale with the corresponding customer's region and type, supporting customer segmentation and regional analyses.

* **Output:** merged\_sales\_customer DataFrame.

**3. Incorporating Discounts**

To analyze the effect of discounts on sales, we merged the merged\_sales\_customer DataFrame with discount\_data using ProductID. The StartDate and EndDate columns were checked against the sale Date to determine whether the discount applied.

* **Key Transformation:**
  + Discounts were applied only if the sale date fell within the discount period.
  + If a discount was applicable, it was used to calculate the **Discounted Revenue**.
* **Output:** merged\_sales\_discount DataFrame.

**4. Feature Engineering**

We created new columns to enhance the dataset:

* **DiscountApplied:** Indicates the discount percentage applied to each sale.
* **DiscountedRevenue:** Calculates revenue after applying the discount (Revenue \* (1 - DiscountApplied)).
* **AvgRevenuePerUnit:** Represents average revenue per unit sold (Revenue / UnitsSold).

These features provide deeper insights into sales performance and customer purchasing behavior.

**5. Reshaping Data**

To extract actionable insights, the data was reshaped into summary formats:

* **Pivot Table:** Shows total discounted revenue (DiscountedRevenue) per product, segmented by customer type (Regular vs. Premium).
* **Sales Summary by Region:** Aggregates total discounted revenue by region, highlighting regional performance.

**Output: Excel Report**

The final processed data was saved to an Excel file named **Sales\_Analysis\_Report.xlsx** with the following sheets:

**1. Merged Sales Data**

Contains all merged and transformed data, including product, customer, sales, and discount details. Key insights include:

* Discounts applied to each sale.
* Revenue before and after discounts.
* Average revenue per unit sold.

**2. Pivoted Revenue by Customer Type**

A pivot table summarizing total discounted revenue for each product based on customer type:

* Useful for identifying which products are more popular among Regular vs. Premium customers.

**3. Sales by Region**

A summary table showing total discounted revenue grouped by region:

* Supports regional performance analysis to identify high and low-performing areas.

**Code Explanation**

**Key Functions Used**

1. **Merging DataFrames (pd.merge)**
   * Enables joining datasets on common keys (ProductID, CustomerID).
   * Supports both inner and left joins, depending on the need for complete or filtered datasets.
2. **Conditional Transformation (np.where)**
   * Applies logical conditions to calculate DiscountApplied based on date ranges.
3. **Feature Engineering**
   * Calculations such as DiscountedRevenue and AvgRevenuePerUnit provide additional metrics for deeper analysis.
4. **Pivot Table (pivot\_table)**
   * Reshapes data for specific analyses, enabling category-wise and customer-wise segmentation.
5. **ExcelWriter (pd.ExcelWriter)**
   * Saves multiple DataFrames to a single Excel file with organized sheets.

**Example Scenarios for Business Use**

1. **Product Promotions:** The DiscountedRevenue data helps assess the effectiveness of discounts on driving sales and revenue.
2. **Customer Insights:** The pivot table reveals spending patterns of Regular and Premium customers, aiding targeted marketing.
3. **Regional Performance:** The regional summary highlights high-performing regions, supporting resource allocation for sales strategies.

**Summary**

This analysis showcases the power of data fusion and reshaping in creating actionable insights from complex datasets. The generated Excel report can be directly used by business stakeholders for strategic decision-making, marketing, and performance evaluation.