Online Retail Sales Data

Assignment Questions

PW SKILLS

Online Retail Sales Data

- **Dataset Link :***Online Retail Sales Dataset Link*
- **Rows:** 30
- **Columns:** 8
- **Description:**
- The dataset contains **sales transactions** from an online store. It includes product details, pricing, customer regions, and order dates. Some rows have **missing values**, allowing students to practice **data cleaning** techniques such as filling null values, validating data, and performing revenue and profit margin analysis.

Key Columns

1. **Order_ID**

- **Description:** Unique identifier for each sales transaction.
- **Purpose:**
  - Used to **track individual orders.**
  - Helps in identifying and removing duplicates during data cleaning.

2. **Product**

- **Description:** Name of the product sold (e.g., Yoga Mat, Wireless Mouse, etc.).
- **Purpose:**
  - Essential for **product-level sales analysis.**
  - Helps in identifying **top-selling products.**

3. **Category**

- **Description:** Product category (e.g., Electronics, Fitness, Kitchenware).
- **Purpose:**
  - Helps in **categorical analysis** of sales performance.
  - Useful for **grouping and filtering** data by category.

4. **Quantity**

- **Description:** Number of units purchased per transaction.
- **Purpose:**
  - Used in **revenue calculations** `(Total_Amount = Quantity × Price_Per_Unit).`
- Important for **profit margin analysis.**
- Contains some **null values**, making it useful for **data cleaning tasks.**

5. **Price_Per_Unit**

- **Description:** Price per unit of the product.
- **Purpose:**

- Used to calculate the **total revenue.**
- Helps in identifying **price variations** across categories.
- Includes **null values**, which need to be handled during data cleaning.

6. **Total_Amount**

- **Description:** Total revenue generated from the transaction.
  - **Formula:**
    `Total_Amount = Quantity × Price_Per_Unit`
- **Purpose:**
  - Used to measure **revenue by region and product.**
  - Essential for **trend analysis** and **profit margin calculations.**
  - Some rows contain **null values**, which can be derived by filling missing Quantity or Price values.

7. **Customer_Region**

- **Description:** The geographical region of the customer (e.g., North, South, East, West).
- **Purpose:**
  - Important for **regional sales analysis.**
  - Helps identify **high-revenue regions.**
  - Some rows contain null values, making it relevant for **data cleaning.**

8. **Order_Date**

- **Description:** The date when the order was placed.
- **Purpose:**
  - Used for **trend analysis** over time.
  - Helps in identifying **seasonal sales patterns.**
  -

Problem Statement

1. **Data Cleaning and Transformation:**
   - Identify and highlight the **null values** in the `Quantity, Price_Per_Unit`, and `Total_Amount` columns using **Conditional Formatting.**
   - Fill the missing `NULL` values with the **rounded average** of their respective columns.
   - Remove any **leading or trailing spaces** in the `Customer_Region` column using **TRIM().**
   - Convert all `Customer_Region` values to **uppercase.**
2. **PivotTable Analysis:**

- Create a **PivotTable** that displays the **total revenue** (`Total_Amount`) by `Customer_Region`.
- Identify the region with the **highest revenue.**
- Apply **Data Bars** in **Conditional Formatting** to visually highlight the region with the highest sales.

3. **VLOOKUP and INDEX/MATCH Operations:**
   - Use the `VLOOKUP` function to find the Total_Amount for `Order_ID = 1015`.
   - Use the `INDEX/MATCH` functions together to retrieve the Category for `Order_ID = 1027`.

- Compare the results of both methods and explain the difference between VLOOKUP and INDEX/MATCH.

4. **Trend Analysis:**

- **Create a Line Chart:**
  - Plot the **revenue trend over time** using `Order_Date` and `Total_Amount`.
- **Analyze the Trend:**
  - Identify if there are any seasonal patterns or spikes in sales.

5. **Profit Margin Calculation:**

- **Add a New Column:**
  - Create a calculated column:
    `Profit_Margin = Total_Amount - (Quantity × Price_Per_Unit × 0.6)`
- **Highlight High-Profit Products:**
  - Use **Conditional Formatting** to highlight products with a profit margin above **₹5000.**
  - Sort the dataset by `Profit_Margin` in **Descending Order.**
  - Identify the **top 3 products** with the highest `Profit_Margin` and display them in a **separate table.**