Operations:

1. Data Cleaning

• Remove duplicates from the dataset to ensure data integrity.
• Filter records where Country is 'United States' to focus on domestic orders.
• Extract the numeric part from Customer_ID to standardize customer identification. (EX: CH-1234, extract 1234)
• Concatenate Customer_ID and Customer_Name with '-' to create a unique identifier for each customer. (Ex: 1234-Charlies, Extracted_ID-Customer_name) and store it in Customer_Id_Name Column
• Drop the customer_id, Customer_name Columns

2. Sales Summary

• Summarize the sales data by calculating the sum of sales and store it in Total_Sales and average sales store it in an Avg_Sales column for each customer using their 'Customer_ID_Name'
• Order the summarized data in descending order based on the total sales ('Total_Sales').
• Filter customers with total sales greater than 3000 and average sales greater than 300 to focus on significant contributors.
• Drop the unnecessary columns, kindly check the sample output.

3. Customer Order Analysis

• Filter records for customers in the category 'Office Supplies' and City in 'San Francisco' to analyze local customer behavior.
• Create a new column orders_count, Calculate the count of orders for each customer to determine their order frequency.
• Sort the results by order count in descending order to identify the most frequent buyers and get only the top 10 records.
• Drop the unnecessary columns, kindly check the sample output.

4. Customer Geography Analysis

• Filter records for customers in state 'California' to focus on specific geographical areas.
• Group customers by region (North, South, East, West) based on their location data.
• Calculate the count of customers in each region to understand the geographical distribution.
• Drop the unnecessary columns, kindly check the sample output.

5. Order Processing Time Analysis

• Calculate the processing days for each order by subtracting the order date from the ship date and store it in the new column Processing_days.
• Categorize processing days (e.g., Less than 1 day then Immediate delivery, 1 to 3 days then Moderate Delivery, 3 or more days then Long-term delivery).
• Count the number of orders falling within each categorise processing days for each to analyze processing days distributions.
• Drop the unnecessary columns, kindly check the sample output.