Store Inventory Management

You are given a store inventory dataset with the following columns:

product_id: Unique identifier for the product

product_name: Name of the product

stock_quantity: Quantity of the product currently in stock

restock_date: Date the product is due to be restocked

price: Price of the product

Your tasks are:

1. Find the products that need restocking, i.e., stock quantity is less than 10.

2. Find the total value of inventory (stock quantity * price).

3. Identify the top 3 products with the highest stock value.

Customer Lifetime Value (CLV) Calculation

You are working with a retail business and have customer purchase data with the following columns:

customer_id: Unique identifier for each customer

purchase_date: Date of the purchase

purchase_amount: The amount spent by the customer during a purchase

Your task is to:

1. Calculate the total purchase amount for each customer.
2. Calculate the average purchase amount per customer.
3. Determine the customers who have spent the most over time.

In the given DataFrame, calculate the average sales for each product, but only for the 'Home' category.

data = {

'Product': ['A', 'B', 'A', 'C', 'B', 'C'],

'Category': ['Electronics', 'Electronics', 'Home', 'Electronics', 'Home', 'Home'],

'Sales': [200, 300, 150, 400, 250, 100]

}

Case Study: Analyzing Sales Data for an E-commerce Company

Problem:

You are tasked with analyzing sales data for an e-commerce company to understand customer behavior, sales trends, and performance. The dataset consists of various attributes related to sales transactions. You need to analyze the data to extract insights like total sales and average order value.

Case Study : Patient Health Records

Scenario:

You're working with health data from a small clinic:

data = { 'PatientID': [1, 2, 3, 4, 5], 'Age': [34, 45, 23, 56, 67], 'Gender': ['F', 'M', 'F', 'M', 'F'], 'BloodPressure': [120, 140, 110, 160, 130], 'Cholesterol': ['Normal', 'High', 'Normal', 'Very High', 'High'] }

df = pd.DataFrame(data)

Questions:
- What is the average age of patients?
- Count the number of patients by gender.
- How many patients have "High" or "Very High" cholesterol?
- List patients with blood pressure over 130.
- Replace "Very High" with "Critical" in the cholesterol column.

Case Study : Employee Performance

Scenario:

A company wants to analyze performance data:

data = { 'Employee': ['Anna', 'Ben', 'Cara', 'David', 'Eva'], 'Department': ['HR', 'IT', 'IT', 'HR', 'Finance'], 'Score_2022': [88, 92, 85, 90, 75], 'Score_2023': [91, 89, 87, 93, 80] } df = pd.DataFrame(data)

Questions:

- Compute the improvement in score for each employee.
- Find the department with the highest average improvement.
- Who had the highest score in 2023?
- What is the average score in 2022 and 2023?
- Add a column with performance grade (A: >90, B: 80-90, C: <80) based on 2023 scores.

1) Given ,df = pd.DataFrame({
   'Department': ['HR', 'HR', 'IT', 'IT', 'Finance', 'Finance'],
   'Employee': ['Alice', 'Bob', 'Cara', 'Dan', 'Eve', 'Frank'],
   'Score': [88, 92, 95, 89, 85, 87]
})

Rank employees within each department based on their scores (highest first).

2) Given, df = pd.DataFrame({
   'Customer': ['Alice', 'Bob', 'Alice', 'Dave', 'Bob'],
   'OrderID': [1, 2, 1, 3, 2]
})

Question:

- How many duplicate orders are there?
- Remove them and keep the first occurrence.

3) Given,df = pd.DataFrame({

   'Region': ['North', 'South', 'North', 'South'],

   'Product': ['A', 'A', 'B', 'B'],

   'Sales': [200, 150, 100, 250]

})

Question:

- Pivot the table to show sales per region per product.
- Show only rows where the sales are above the average sales of their region.

4) How do you get the count of all unique values of a categorical column in a DataFrame?

5) Report the top 10 ranked songs by position. Output the track and the ranks in the descending order of ranks. If there are multiple songs with the same rank, sort them in alphabetical order of the track name.