Fall 2022 Data Science Intern Challenge

Please complete the following questions, and provide your thought process/work. You can attach your work in a text file, link, etc. on the application page. Please ensure answers are easily visible for reviewers!

Question 1: Given some sample data, write a program to answer the following: click here to access the required data set

On Shopify, we have exactly 100 sneaker shops, and each of these shops sells only one model of shoe. We want to do some analysis of the average order value (AOV). When we look at orders data over a 30 day window, we naively calculate an AOV of $3145.13. Given that we know these shops are selling sneakers, a relatively affordable item, something seems wrong with our analysis.

    a. Think about what could be going wrong with our calculation. Think about a better way to evaluate this data.

Looking at the data, apparently, the value 3145.13 is calculated by summing up all the order_amounts to get the total amount value. Then, dividing the amount by 5000, which is the number of orders. Assuming we want to know the average value per item, we shouldn't be dividing the number by 5000, but the sum of column total_items. The sum would give up the total amount of items ordered. Dividing the total order value by the total amount of items ordered would give us the average value of items in orders. So the mistake here is that we should use sum up items instead of counting the orders.

    b. What metric would you report for this dataset?

Therefore, we would be dealing with the variables/ columns "order_amount" and "total_items". The sum of each of those two columns would be our parameters of interest. Then we divide the sum of "order_amount" by the sum of "total_items".

    c. What is its value?

The average value is about 357.92

Question 2: For this question you'll need to use SQL. [Follow this link](#) to access the data set required for the challenge. Please use queries to answer the following questions. Paste your queries along with your final numerical answers below.

a. How many orders were shipped by Speedy Express in total?
   54

   > SELECT * FROM [Orders] WHERE ShipperID=1

b. What is the last name of the employee with the most orders?
   40

   >SELECT Orders.EmployeeID, Employees.LastName, count(Orders.OrderID)
         FROM Orders JOIN Employees ON
         Employees.EmployeeID=Orders.EmployeeID
         GROUP BY Orders.EmployeeID ORDER BY COUNT(*) desc;

c. What product was ordered the most by customers in Germany?
   Boston Crab Meat 160

   >CREATE VIEW Product_Orders_1 AS
         SELECT Orders.OrderID, Customers.Country,
         OrderDetails.Quantity,Products.ProductName FROM Orders
         JOIN OrderDetails On Orders.OrderID = OrderDetails.OrderID
         JOIN Customers On Customers.CustomerID = Orders.CustomerID
         JOIN Products On OrderDetails.ProductID = Products.ProductID
         WHERE Customers.country = 'Germany'

   >SELECT ProductName, sum(Quantity) AS TotalOrder
         FROM Product_Orders_1
         Group By ProductName
         ORDER BY TotalOrder DESC