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.
>   **What was going wrong was the weight of large order amounts skewing the AOV. To reduce the influence of these data points and therefore better evaluate the data, median should be used.**
> b. What metric would you report for this dataset?
>   **Median**
> c. What is its value?
>   **$284.00**

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?

SELECT S.ShipperName, COUNT(*) as c
FROM Products P, Orders O, OrderDetails OD, Shippers S
WHERE O.OrderID = OD.OrderID
AND P.ProductID = OD.ProductID
AND S.ShipperID = O.ShipperID
AND S.ShipperName = "Speedy Express"
ORDER BY c DESC

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

SELECT E.EmployeeID,
*   E.LastName,*
*   E.FirstName,*
*   COUNT(OrderID) AS TOTAL*
FROM Orders O, Employees E
WHERE O.EmployeeID = E.EmployeeID
GROUP BY E.EmployeeID
ORDER BY TOTAL DESC

Peacock

c. What product was ordered the most by customers in Germany?

SELECT P.ProductID, P.ProductName, COUNT(*) AS c
FROM Products P, Orders O, OrderDetails OD, Customers C
WHERE P.ProductID = OD.ProductID
AND O.CustomerID = C.CustomerID
AND O.OrderID = OD.OrderID
AND C.Country = "Germany"
GROUP BY P.ProductID, P.ProductName
ORDER BY c DESC

Gorgonzola Telino