Winter 2021 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.
      - This averaged value is including customers who are ordering large numbers of shoes, likely for resale, and do not represent the average consumer. This average is also including extremely expensive shoes, which would not be purchased by the average consumer.
      - A better way to evaluate this data would be to consider what the behavior of the majority of customers (*i.e.,* what is the behavior of 95% of the customers).
   b. What metric would you report for this dataset?
      - I believe a more representative value for the average order value is obtained when excluding the top 5% of two metrics: the number of shoes in an order, and the cost of each pair of shoes
      - First, I excluded orders where more than 4 pairs of shoes were purchased. Second, I excluded orders where the cost of a single pair of shoes exceeded $195.
   c. What is its value?
      - $270.30

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 COUNT(*) FROM Orders WHERE ShipperID = 1;
- 54

   b. What is the last name of the employee with the most orders?
- SELECT Orders.EmployeeID, COUNT(*) AS 'Total Order', Employees.LastName
  FROM Orders
  JOIN Employees
  ON Orders.EmployeeID = Employees.EmployeeID
  GROUP BY Orders.EmployeeID
  ORDER BY COUNT(*) DESC;
- Peacock

   c. What product was ordered the most by customers in Germany?
- SELECT Customers.Country, Products.ProductName,
  SUM(OrderDetails.Quantity) AS 'Total Quantity'
  FROM Customers
  JOIN Orders
  ON Customers.CustomerID = Orders.CustomerID
  JOIN OrderDetails
  ON Orders.OrderID = OrderDetails.OrderID
  JOIN Products
  ON OrderDetails.ProductID = Products.ProductID
  WHERE Customers.Country = 'Germany'
  GROUP BY Products.ProductName
  ORDER BY SUM(OrderDetails.Quantity) DESC;
- The product ordered most from customers in Germany was Boston Crab Meat with 160 quantities