Winter 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.
   b. What metric would you report for this dataset?
   c. What is its value?
   SOLUTIONS
   a.The Average Order Value is misleading and skewed by a couple of extremely high($90k) or extremely low value orders.
   b. I would prefer the Revenue per visitor metric
   c. RPV is $157,256.40 per unique visito

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?
   b. What is the last name of the employee with the most orders?
   c. What product was ordered the most by customers in Germany?
      SOLUTIONS
   d. SELECT COUNT(*) FROM orders
      WHERE ShipperID = 1
      Result = **54**

e. SELECT DISTINCT O.EmployeeID AS Dist_emp, COUNT(O.EmployeeID) AS counter, E.LastName
   FROM Orders O, Employees E
   WHERE O.EmployeeID = E.EmployeeID
   GROUP BY Dist_emp
   ORDER BY counter DESC
   Result=**<u>Peacock</u>**

f. SELECT DISTINCT A.OrderID, COUNT(A.OrderID) AS most_orders, C.ProductName, D.CustomerName, D.Country
   FROM Orders A
           inner join Customers D
           ON A.CustomerID = D.CustomerID
     inner join OrderDetails B
           ON A.OrderID = B.OrderID
     inner join Products C
           ON B.ProductID = C.ProductID
   WHERE D.Country = "Germany"
   GROUP BY A.OrderID
   ORDER BY most_orders DESC
   results=**<u>Tunnbrof</u>**, **<u>Grandma's Boysenberry Spread</u>** and **<u>Ikura</u>**