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.

The average order value is the average price of an order, in this case: sneakers. As is mentioned, sneakers are a “relatively affordable item” and should not be anywhere close to the calculated AOV of $3145.13.

Looking at the data there are 2 stores that must be ignored when doing any calculations. The 2 stores are anomalies regarding the unit price and the number of units sold per order.

b. What metric would you report for this dataset?

The metric I would report for this dataset is the median. Because anomalies would go off to the edges and essentially be disregarded.

c. What is its value?

The value of the median order amount is $284.

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?
   Query option 1:

```
SELECT COUNT(*) FROM Orders
WHERE shipperID= (SELECT ShipperID FROM Shippers WHERE ShipperName = "Speedy Express")
```

   Query option 2:

```
SELECT COUNT(*) FROM Orders
JOIN Shippers ON Shippers.ShipperID = Orders.ShipperID
WHERE Shippers.ShipperName = "Speedy Express"
```

Answer: **54 orders** were shipped by Speedy Express in total.

b. What is the last name of the employee with the most orders?
   Query option 1:

```
SELECT LastName FROM Employees
WHERE EmployeeID = (SELECT EmployeeID FROM Orders GROUP BY EmployeeID
ORDER BY COUNT (OrderID) DESC
LIMIT 1)
```

   QUERY option 2:

```
SELECT Employees.LastName FROM Employees
JOIN Orders ON Orders.EmployeeID = Employees.EmployeeID
GROUP BY Orders.EmployeeID
ORDER BY COUNT (OrderID) DESC
LIMIT 1
```

Answer: **Peacock** is the last name of the employee with the most orders.

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

```
SELECT ProductName FROM products WHERE ProductID = (
SELECT ProductID FROM OrderDetails
WHERE OrderID IN (
SELECT OrderID FROM Orders
WHERE CustomerID IN (SELECT CustomerID FROM Customers WHERE country = "Germany"))
```

```
GROUP BY ProductID
ORDER BY Sum(Quantity) DESC
LIMIT 1)
```

Answer: **Boston Crab Meat** was the product ordered most by customers in Germany.