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.

Calculation was included outliers. We must do it separately for individuals. (Main and Outliers) Details is attached in GitHub

b. What metric would you report for this dataset?

1. I would report these suspicious transactions at *Shop 42*. Which included exact same metric values in same time with 17 times occurrence.
It can be fraud or failed customer's automatic payment system (Need confirmation with our customer)

shop_id	user_id	order_amount	total_items	payment_method	created_at
42	607	704000	2000	credit_card	2017-03-07 4:00:00

2. We should focus on *Shop 78*. This shop has highest revenue from other shops except for *Shop 42*.
We have to do more deep analysis in this store. Probably we can find our most valued customers from this Shop.

c. What is its value?

293.72

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 sh.ShipperName, Count(1) No_Order
FROM [Orders] o, [Shippers] sh
WHERE ShipperName = 'Speedy Express' and o.ShipperID = sh.ShipperID
GROUP BY ShipperName;
```

ShipperName	No_Order
Speedy Express	54

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

```
SELECT e.LastName, Count(1) No_Order
FROM [Orders] o, [Employees] e
WHERE o.EmployeeID = e.EmployeeID
GROUP BY e.EmployeeID
ORDER BY No_Order desc
Limit 1;
```

LastName	No_Order
Peacock	40

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

ProductName	No_Order
Gorgonzola Telino	5

```
SELECT TEMP2.ProductName, COUNT(TEMP2.OrderID) No_Order

FROM

        (SELECT TEMP1.OrderID, P.ProductName

         FROM

        (SELECT O.OrderID

        FROM [Orders] O, [Customers] C

        WHERE O.CustomerID = C.CustomerID AND C.Country = 'Germany') TEMP1,
```

```
        [Orders] O INNER JOIN [OrderDetails] OD INNER JOIN [Products] P
ON TEMP1.OrderID = OD.OrderID AND OD.ProductID = P.ProductID
GROUP BY TEMP1.OrderID, P.ProductName) TEMP2
GROUP BY TEMP2.ProductName
ORDER BY No_Order DESC
LIMIT 1;
```