Summer 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.

After checking the data, there are two outliers that can be noticed: average order_amount and price per pair of shoes have outliers which affect the results. So, we should exclude the outliers which can be found in shop_id 78 and 42. Then the AOV result would be more reasonable after they are excluded.

b. What metric would you report for this dataset?

Reporting price of shoes and the average amount of orders would be considered to be put into my report. I'd also report with the numbers and figures and check the median values. AOV would be : sum of order amount / order count

c. What is its value?

AOV : $300.16 (without outliers)

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?

A: 54

```
SELECT COUNT(OrderID), ShipperName
FROM [Orders] o
JOIN [Shippers] s
ON o.ShipperID = s.ShipperID
GROUP BY s.ShipperID
HAVING s.ShipperName = 'Speedy Express'
```

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

A: Peacock

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

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

A: Boston Crab Meat

```
SELECT SUM(OD.Quantity) AS SUMQ,  P.ProductID, P.ProductName, c.Country
FROM OrderDetails OD
JOIN Orders O, Products P
ON O.OrderID = OD.OrderID and P.ProductID = OD.ProductID
JOIN Customers C
ON C.CustomerID == O.CustomerID
WHERE C.Country == 'Germany'
GROUP BY P.ProductName
ORDER BY SUMQ DESC
LIMIT 1
```