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.

The issue is that a small minority of orders were very expensive, and these orders are significantly affecting the mean, which is the AOV calculated above. Examples of such orders include order_id 16 with value of $764,000 and order_id 161 with value of $25,725. A better way to evaluate the data would be either to remove the extreme outliers before calculation or to use the median instead of the mean.

b. What metric would you report for this dataset?

I would report the median for this dataset.

c. What is its value?

The value 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?

196 orders were shipped by Speedy Express in total.

```
SELECT COUNT(DISTINCT OrderID) AS CountOrders FROM Orders;
```

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

The last name is Peacock.

```
SELECT LastName FROM Employees as e JOIN
(SELECT COUNT(DISTINCT OrderID) AS CountOrders, EmployeeID FROM Orders
GROUP BY EmployeeID ORDER BY CountOrders DESC LIMIT 1) AS t
ON e.EmployeeID = t.EmployeeID;
```

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

The product ordered most by customers in Germany is Boston Crab Meat.

```
SELECT ProductName
FROM (SELECT OrderID FROM Customers AS c JOIN Orders AS o ON c.CustomerID =
o.CustomerID WHERE Country = "Germany") AS co
JOIN (SELECT p.ProductID, od.OrderID, ProductName, Quantity FROM OrderDetails AS od
JOIN Products AS p ON od.ProductID = p.ProductID) AS odp
ON co.OrderID = odp.OrderID
GROUP BY ProductID
ORDER BY SUM(Quantity) DESC
LIMIT 1;
```