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.

There are some outliers in the data like the oder_id 16. These outliers will influence the accuracy of the AOV. My recommendation is to remove these outliers and recalculate the data with a normal distribution.

b. What metric would you report for this dataset?

I used the IQR method to find these outliers and remove them. If any value is smaller than the lower bound or higher than the upper bound, it would be considered an outlier.

Q1	163
Q3	390
IQR	227
Upper bound	730.5
Lower bound	-177.5

c. What is its value?

Now the AOV is 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 *
FROM [Orders]
WHERE ShipperID = (SELECT ShipperID FROM [Shippers] WHERE ShipperName
LIKE 'Speedy Express');
```

A: 54

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

```
SELECT COUNT(*) AS Number_Of_Order, [Orders].EmployeeID,
[Employees].LastName
FROM [Orders]
INNER JOIN Employees
ON Orders.EmployeeID = Employees.EmployeeID
GROUP BY [Orders].EmployeeID
ORDER BY Number_Of_Order DESC;
```

A: Peacock

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

```
SELECT COUNT(*) AS Number_Of_Product, [OrderDetails].ProductID,
[Products].ProductName, [Customers].Country
FROM [OrderDetails]
INNER JOIN [Orders] ON [OrderDetails].OrderID = [Orders].OrderID
INNER JOIN [Customers] ON [Orders].CustomerID = [Customers].CustomerID
INNER JOIN [Products] ON [OrderDetails].ProductID = [Products].ProductID
WHERE Country LIKE 'Germany'
GROUP BY [OrderDetails].ProductID
ORDER BY Number_Of_Product DESC;
```

A: Gorgonzola Telino