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.
b. What metric would you report for this dataset?
c. What is its value?

Answer 1:

a. SELECT DISTINCT shop_id
   FROM dataset
   WHERE order_amount > 1200;
   SELECT Count(*)
   FROM dataset
   WHERE order_amount >1200;
   There are a total of 66 outliers with order amount larger than 1200 in the dataset provided. I chose 1200 to be the threshold as the sneakers should be affordable as mentioned in the question. There is something wrong with the data from Shop 42 and Shop 78. The order_amount is much larger than usual. When we calculate the average order value, the outlier will have a serious impact on the result.
b. I will choose the median to report for the dataset.
c. The answer 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?
b. What is the last name of the employee with the most orders?
c. What product was ordered the most by customers in Germany?

Answer:

a.
```
SELECT COUNT(*)
FROM Orders
WHERE ShipperID = (
        SELECT ShipperID
    FROM Shippers
    WHERE ShipperName = 'Speedy Express'
);
```

The numeric answer is 54.

b.
```
SELECT COUNT(*) as total_Number, e.LastName
FROM Employees e
JOIN Orders o
        ON e.EmployeeID = o.EmployeeID
GROUP BY e.EmployeeID
ORDER BY total_Number DESC
```

The answer is Peacock.

c.
```
SELECT SUM(Quantity) as prod_Quantity, ProductID
FROM Orders o
JOIN Customers c
        ON c.CustomerID = o.CustomerID
        AND c.Country = 'Germany'
JOIN Orderdetails od
        ON od.orderID = o.orderID
GROUP BY od.ProductID
ORDER BY prod_Quantity DESC
```

The numeric answer is **40.**