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!

> 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.

Answer: The data has an outlier and the calculation is mean, mean is more vulnerable to outliers thus considering median as a metric is more suitable for skewed data like this one.

b. What metric would you report for this dataset?
Answer: Median

c. What is its value?
Answer: It's value is 284, which gives more reasonable measure about given dataset.

a. How many orders were shipped by Speedy Express in total?

Query:

```
SELECT count(*)
AS 'Speedy Express Orders'
FROM Orders o
INNER JOIN Shippers s on o.ShipperID = s.ShipperID
WHERE s.ShipperName = 'Speedy Express'
```

Output: 54

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

Query:

```
SELECT LastName FROM     (
   SELECT count(*) as cnt,LastName
         FROM Orders o
         INNER JOIN Employees e ON o.EmployeeID = e.EmployeeID
         GROUP BY LastName
)
   WHERE cnt=(SELECT MAX(cnt) FROM     (
   SELECT count(*) as cnt,LastName
         FROM Orders o
         INNER JOIN Employees e ON o.EmployeeID = e.EmployeeID
         GROUP BY LastName
))
```

Output: Peacock

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

Query:

```
SELECT ProductName
FROM
(
 SELECT count(*) as cnt,ProductName
 FROM
 (
         SELECT *
         FROM
         (
         SELECT *
```

```
        FROM
        (
        SELECT *
        FROM Products P
        INNER JOIN OrderDetails O ON P.ProductID = O.ProductID
        ) A
        INNER JOIN Orders ord ON A.OrderID = ord.OrderID
        ) B
        INNER JOIN Customers C ON B.CustomerID = C.CustomerID
        WHERE Country = 'Germany'
 )
GROUP BY ProductName
)
WHERE cnt = (SELECT MAX(cnt)
FROM
(
 SELECT count(*) as cnt,ProductName
 FROM
 (
        SELECT *
        FROM
        (
        SELECT *
        FROM
        (
        SELECT *
        FROM Products P
        INNER JOIN OrderDetails O ON P.ProductID = O.ProductID
        ) A
        INNER JOIN Orders ord ON A.OrderID = ord.OrderID
        ) B
        INNER JOIN Customers C ON B.CustomerID = C.CustomerID
        WHERE Country = 'Germany'
 )
 GROUP BY ProductName
)
)
```

Output: Gorgonzola Telino