

Fall 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?

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?

--How many orders were shipped by Speedy Express in total?

```
SELECT Count(orderID) as Total_Speedy_Express_Orders
FROM Orders o
JOIN shippers s
ON s.shipperid=o.shipperID
WHERE shippername = "Speedy Express"
```

Total_Speedy_Express_Orders

54

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

--What is the last name of the employee with the most orders?

```
SELECT e.lastname, Count(o.orderID) as Most_Orders
FROM Orders o
JOIN Employees e
ON e.EmployeeID = o.employeeID
GROUP BY o.employeeID
ORDER BY Most_Orders Desc
LIMIT 1
```

LastName	Most_Orders
Peacock	40

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

Option 1: by product orders "most ordered"

--What product was ordered the most by customers in Germany?

```
SELECT count(p.productname) as Total_Orders, productname, Country
FROM OrderDetails od
JOIN Products p ON p.productID = od.productID
JOIN Orders o ON od.orderID = o.orderID
JOIN Customers c ON c.customerId = o.customerID
WHERE c.country = "Germany"
GROUP BY od.productID
Order by Total_Orders DESC
```

Limit 1

Total_Orders	ProductName	Country
5	Gorgonzola Telino	Germany

Option 2: by product quantity "most ordered"

```
SELECT count(p.productname) as Total_Orders, od.quantity, ProductName, Country
FROM OrderDetails od
JOIN Products p ON p.productID = od.productID
JOIN Orders o ON od.orderID = o.orderID
JOIN Customers c ON c.customerId = o.customerID
WHERE c.country = "Germany"
GROUP BY od.productID
order by od.quantity*Total_orders desc
limit 1
```

Total_Orders	Quantity	ProductName	Country
4	50	Boston Crab Meat	Germany