

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?

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 count(OrderByID) from Orders where ShipperID = (select ShipperID from Shippers where ShipperName = 'Speedy Express')`
=54
- b. What is the last name of the employee with the most orders?
`select count(e.LastName)
from Employees e
inner join Orders o
on o.CustomerID = e.EmployeeID
group by e.LastName
order by e.LastName desc
limit 1;`
=4

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

```
select p.ProductName
from Orders o
inner join OrderDetails od
on od.OrderID = o.OrderID
inner join Products p
on p.ProductID = od.ProductID
inner join Suppliers s
on p.SupplierID = s.SupplierID
where s.Country = 'Germany'
group by o.OrderID
order by o.OrderID desc
limit 1;
=Rössle Sauerkraut
```