# 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](https://docs.google.com/spreadsheets/d/16i38oonuX1y1g7C_UAmiK9GkY7cS-64DfiDMNiR41LM/edit#gid=0)

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.

1. Think about what could be going wrong with our calculation. Think about a better way to evaluate this data.
2. What metric would you report for this dataset?
3. What is its value?

**Question 2:** For this question you’ll need to use SQL. [Follow this link](https://www.w3schools.com/SQL/TRYSQL.ASP?FILENAME=TRYSQL_SELECT_ALL) 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.

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

SELECT count(O.OrderID) as Total\_amount

FROM [Orders] as O

inner join [Shippers] as S

on O.ShipperID = S.ShipperID

where S.ShipperName = 'Speedy Express'

Total\_amount

54

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

SELECT E.LastName

FROM [Employees] as E

inner join (select count(O.OrderID) as amount\_of\_order, O.EmployeeID from Orders as O group by O.EmployeeID order by amount\_of\_order desc limit 0, 1) as OO

on E.EmployeeID = OO.EmployeeID

LastName

Peacock

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

select ProductName from Products as P, OrderDetails as OD, Orders as O, Customers as C

where P.ProductID = OD.ProductID and OD.OrderID = O.OrderID and O.CustomerID = C.CustomerID and C.Country = 'Germany'

group by P.ProductID

order by count(O.OrderID) desc

limit 0, 1

ProductName

Gorgonzola Telino