# Fall 2021 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.   
   I am noticing some outliers that could be the cause of such a high AOV. Shop 78 has shoes priced at $25,725 and shop 42 has a relatively high priced item at $352 and has many orders for 2000 items, which total > $700,000 each. These outliers would cause the rest of the data to become skewed and should be removed even if they are correct to report a more reasonable expectation of AOV.
2. What metric would you report for this dataset? The average item price per order
3. What is its value? $387.74

**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(\*)

FROM

Orders AS o

INNER JOIN Shippers AS s ON s.ShipperID = o.ShipperID

WHERE

ShipperName = 'Speedy Express'

54

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

SELECT Employees.LastName,

COUNT(Orders.EmployeeID) AS `order\_count`

FROM 'Orders'

INNER JOIN Employees ON Orders.EmployeeID = Employees.EmployeeID

GROUP BY Orders.EmployeeID

ORDER BY `order\_count` DESC

Peacock

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

SELECT p.ProductName, SUM(d.Quantity) AS Total\_Qua

FROM Customers AS c

INNER JOIN Orders AS o ON o.CustomerID = c.CustomerID

INNER JOIN OrderDetails AS d ON d.OrderID = o.OrderID

INNER JOIN Products AS p ON p.ProductID = d.ProductID

WHERE Country = 'Germany'

GROUP BY `ProductName`

ORDER BY `Total\_Qua` DESC

LIMIT 1;

|  |  |
| --- | --- |
| Boston Crab Meat | 160 |