# 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.

Answer: There is one store that is bulk selling 2000 shoes per order, which will give a very large value for each of those orders and skew the data. There is also another outlier where a store has shoes that cost around 25k a pair. We can still use AOV to evaluate the data, but we should either remove the outliers or group stores based on their target market (general customer, high end shoes, and bulk order).

1. What metric would you report for this dataset?

Answer: AOV on stores based on target market

1. What is its value?

General target AOV: 302.58

High end AOV: 49213.04

Bulk Order AOV: 704000

**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?

Answer: 3575

SELECT ShipperID, SUM(Quantity) FROM Orders

JOIN OrderDetails

ON Orders.OrderID = OrderDetails.OrderID

GROUP BY ShipperID

HAVING ShipperID = (SELECT ShipperID FROM Shippers WHERE ShipperName = "Speedy Express");

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

Answer: Fuller

SELECT EmployeeID,

(SELECT LastName FROM Employees WHERE Orders.EmployeeID = EmployeeID),

MAX(Quantity)

FROM Orders

INNER JOIN OrderDetails

ON Orders.OrderID = OrderDetails.OrderID

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

Answer: Boston Crab Meat

SELECT ProductName,

SUM(Quantity) AS Total\_orders,

Country

FROM OrderDetails

INNER JOIN Products

ON OrderDetails.ProductID = Products.ProductID

INNER JOIN Orders

ON OrderDetails.OrderID = Orders.OrderID

INNER JOIN Customers

ON Orders.CustomerID = Customers.CustomerID

GROUP BY OrderDetails.ProductID

HAVING Country = "Germany"

ORDER BY SUM(Quantity) Desc LIMIT 1