# Summer 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.**

Going through the data set, it is immediately obvious that there are anomalous shops with abnormally high order value per order. It would make sense if the order item count is high, but after further calculation, shop 78 shows an abnormally high price per item of $25,725. To better evaluate this data, it would be best to set aside the data from shop 78, and evaluate the data based on average item value instead of average order value.

1. **What metric would you report for this dataset?**

The average item value across the 100 shops.

1. **What is its value?**

$153.58

**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

WHERE ShipperID = 1;

Order amount: 54

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

SELECT Employees.LastName, COUNT(\*)

FROM Orders, Employees

WHERE Orders.EmployeeID = Employees.EmployeeID

GROUP BY Employees.LastName

HAVING COUNT(\*) >= ALL(

SELECT COUNT(\*)

FROM Orders

GROUP BY EmployeeID);

Last Name: Peacock

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

Select ProductID, SUM(Expr1002)

FROM (

SELECT Orders.CustomerID, OrderDetails.ProductID, SUM(Quantity)

FROM Orders, OrderDetails

WHERE Orders.OrderID = OrderDetails.OrderID

GROUP BY Orders.CustomerID, OrderDetails.ProductID

Having CustomerID in(

SELECT CustomerID

FROM Customers

WHERE Country = 'Germany'))

Group BY ProductID

ProductID: 40