# 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.

The incorrect AOV calculation of $3145.13 was most likely arrived at by mistakenly calculating the total\_items with the count() function instead of sum().count() will only provide the total count of the number of rows . Sum() will add together all of the values in the total\_items column. So will get Sum(order\_amount)/5000=3145.13, which is heavy number. If we use function Sum(), we will get Sum(order\_amount)/43936=357.92.

1. What metric would you report for this dataset?

for i in range(1,maxlum+1):

    if sheet.cell(1,i).value=='order\_amount':#get column of order\_amount

        oa=i

    elif sheet.cell(1,i).value=='total\_items':#get column of total\_items

        ti=i

for i in range(2,maxrow+1):

    oasum=oasum+sheet.cell(i,oa).value #get the sum of order\_amount

print(oasum)

for i in range(2,maxrow+1):

    tisum=tisum+sheet.cell(i,ti).value #get the sum of total\_items

1. What is its value?

357.92

**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(a.OrderID) FROM Orders a join Shippers b

where a.shipperID=b.shipperID and b.ShipperName="Speedy Express"In total, **54** Orders were shipped by Speedy Express.

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

SELECT a.LastName FROM Employees a join Orders b

where a.EmployeeID=b.EmployeeID group by a.EmployeeID

order by Count(b.OrderID) Desc limit 1

Last name is **Peacock**.

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

SELECT d.ProductName,Sum(c.Quantity) FROM Orders a join Customers b join OrderDetails c join Products d

where a.CustomerID=b.CustomerID and a.OrderID=c.OrderID and c.ProductID=d.ProductID and b.Country="Germany"

group by c.ProductID order by Sum(c.Quantity) Desc limit 1

**Boston Crab Meat** was most ordered by Germany-based customers with a total of 160 orders.