# 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.

Answer: Mathematically, nothing went wrong because we have some very large orders. For example, order 16 includes 2000 items which results in an extraordinary order value $70400. To provide a more comprehensive AOV analysis we can:

1． use the median value of the order\_amount instead of the average value;

2． only includes the orders where order items<10

1. What metric would you report for this dataset?

Answer: The media or mode value of the order amount should be a better evaluation

1. What is its value?

Answer: Media = 284 Mode = 153

**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: 54

Select count(OrderID) as num\_order

from Orders

join Shippers

using(ShipperID)

where ShipperName = 'Speedy Express';

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

Answer: Peacock 40 orders

Select count(OrderID) as num\_order,

EmployeeID,

LastName

from Orders

join Employees

using(EmployeeID)

group by EmployeeID

order by num\_order desc

limit 1;

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

Answer: assuming the product was ordered with the most quantity

Boston Crab Meat 160 total quantity ordered by customers in Germany

Select ProductName,

sum(Quantity) as TotalQuantity

from OrderDetails

join Products

using (ProductID)

join Orders

using (OrderID)

join Customers

using (CustomerID)

where Country='Germany'

group by ProductName

order by TotalQuantity desc

limit 1;

assuming the product was ordered with the most times

Gorgonzola Telino 5 orders

Select ProductName,

count(distinct OrderID) as TotalQuantity

from OrderDetails

join Products

using (ProductID)

join Orders

using (OrderID)

join Customers

using (CustomerID)

where Country='Germany'

group by ProductName

order by TotalQuantity desc

limit 1;