# 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.

There could be several large orders that massively skew the average order value, meaning they would make the AOV much larger than what we would anticipate. This method is thus very ineffective at telling us what we want to know: the “typical” order order size. A better way to evaluate the data for this problem would be to look at the median order size instead.

1. What metric would you report for this dataset?

We want to determine the AOV for these stores, but in this case, as stated above, a better way of finding the “typical” order size (the value that we actually want) would be to find the median order value. This would give us a sense of the midpoint of order value, with half of the orders being of higher value and the other half being of a lower value.

1. What is its value?

The median order value (MOV) is $284.

**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(OrderID), ShipperID FROM [Orders]

WHERE ShipperID in (SELECT ShipperID FROM [Shippers]

WHERE ShipperName = “Speedy Express”)

There were 54 orders shipped by Speedy Express

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

SELECT Count(Orders.OrderID) as numOrders, Orders.EmployeeID, Employees.LastName

FROM Orders

INNER JOIN Employees

ON Orders.EmployeeID=Employees.EmployeeID

GROUP BY Orders.EmployeeID

ORDER BY numOrders DESC

EmployeeID 4, Last name Peacock

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

SELECT Products.ProductID, Products.ProductName, MAX(OrderDetails.Quantity) as NumOrdered, OrderDetails.OrderID, Orders.CustomerID, Customers.Country

FROM Products

INNER JOIN OrderDetails

ON Products.ProductID=OrderDetails.ProductID

INNER JOIN Orders

ON OrderDetails.OrderID=Orders.OrderID

INNER JOIN Customers

ON Orders.CustomerID=Customers.CustomerID

WHERE Customers.Country = 'Germany'  
  
The most ordered product in Germany was the Steeleye Stout, with 100 of them being ordered.