# 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.

The data does not pass the eye test as the max order is an order of 2000 shoes.

This is something that an average customer will not do. This might be an order from a retailer

The total order amount is hence skewed, the max order amount is 704,000 again something that an average customer will not do. There are 17 orders containing an order of 2000 shoes per order. This clearly is causing the issue of having an high average order value

1. What metric would you report for this dataset?

I think the best course of action would be to use the median as average order value as this metric is more reliable when there are outliers and influential observations

1. What is its value?

The value of the median 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 ShipperName, COUNT(OrderID) As Orders FROM [Orders]

LEFT JOIN Shippers

ON Orders.ShipperID = Shippers.ShipperID

GROUP BY ShipperName

Speedy Express: 54 orders

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

SELECT LastName, Count(OrderID) AS Orders FROM [Orders]

LEFT JOIN Employees

ON Orders.EmployeeID = Employees.EmployeeID

Group by LastName Order by Orders Desc

Limit 1

Peacock 40 orders

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

Boston Crab Meat

SELECT ProductName

FROM Products, Orders, OrderDetails, Customers

WHERE Orders.CustomerID = Customers.CustomerID AND Country = 'Germany' AND Orders.OrderID = OrderDetails.OrderID AND OrderDetails.ProductID = Products.ProductID

GROUP BY ProductName

ORDER BY SUM(Quantity) DESC

Limit 1