# 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.

In the problem, the average is calculated as which is little incorrect. The reason being, it treats every sneaker shop as a single shop i.e., all 100 different sneaker shops are considered as 1 single sneaker shop. The correct approach would be to calculate AOV for each shop i.e., for each shopid and then take an average of those 100 different AOV values.

For example, let’s suppose there are only 2 shopid with total 5000 orders where each shop has 2500 orders. Then, overall AOV will be equal to (AOV\_shop1 +AOV\_shop2)/2.

Also, AOV does not mean if an item’s price is way high or low. It indicates average order value of a store/shop. In case, a store have hi

1. What metric would you report for this dataset?

For this data set, the metric I will use to calculate an AOV is . Here, ‘i’ denotes individual shopid. Since there are 100 shops therefore, we will take the average of individual AOV’s.

For example, from the dataset, for shopid = 53, total order amount = $14560 and number of orders = 68, therefore, AOV = 14560/68 = $214.11. Similarly, we will calculate AOV for rest 99 shops and take average of all 100 AOV’s.

1. What is its value?

The value is $3089.19. The calculations are performed in the attached excel below.



**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? 54

SELECT COUNT(Orders.OrderID) AS 'Number of Orders', Shippers.ShipperName FROM Orders JOIN Shippers ON Orders.ShipperID = Shippers.ShipperID WHERE ShipperName = 'Speedy Express';

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

SELECT Orders.EmployeeID, Employees.LastName, COUNT(DISTINCT Orders.OrderID) AS 'No. of Orders' FROM Orders JOIN Employees ON Orders.EmployeeID = Employees.EmployeeID GROUP BY Orders.EmployeeID ORDER BY 3 DESC LIMIT 1;

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

SELECT Products.ProductID,Products.ProductName,SUM(OrderDetails.Quantity) AS 'TotalQuantity', Orders.CustomerID FROM OrderDetails JOIN Products ON Products.ProductID=OrderDetails.ProductID JOIN Orders ON Orders.OrderID=OrderDetails.OrderID JOIN Customers ON Orders.CustomerID = Customers.CustomerID WHERE Customers.Country = 'Germany' GROUP BY Products.ProductID ORDER BY TotalQuantity DESC LIMIT 1;