# Fall 2021 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.

**While sneakers are a relatively affordable item. After Doing some data analysis I found that there are shoes priced as high as $25725 and while that’s possible it can also be a mistake . as far as minimum orders. It is within range of sneakers. Of $90. And average/mean of at around $287. And a median of $284. If we calculate the AOV leaving the most expensive item as $352. The correct AOV would be. $2717.37. However, we do see order amounts of 2000 pair of shoes. Which also seems wrong. Therefore. Excluding those the AOV would be a $302.58 which is more likely for affordable shoes/sneakers.**

1. What metric would you report for this dataset?

**A better metric to calculate to make more profit would be to calculate the mode, or the amount that is order the most. By doing this, we can increase the aov better by upselling products that get sell the most and by the amount they get sell the most.**

1. What is its value?

**Mode for Original Dataset: $153**

**Mode for Fixed Dataset: $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?

**Total of 54 Orders where Shipped by Speedy Express**

**Queries:**

**SELECT \* FROM Orders WHERE ShipperID = 1;**

**Or**

**Select \* From Orders, Shippers WHERE Orders.ShipperID = Shippers.ShipperID AND Orders.ShipperID = 1;**

**Or**

**Select \* From Orders, Shippers WHERE Orders.ShipperID = Shippers.ShipperID AND Shippers.ShipperName = "Speedy Express";**

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

**Last Name is Peacock**

**Query:**

**SELECT Employees.LastName, COUNT(DISTINCT Orders.OrderID) as Total\_Orders,Employees.EmployeeID FROM Orders,Employees WHERE Orders.EmployeeID = Employees.EmployeeID Group by Orders.EmployeeID ORDER BY Total\_Orders DESC LIMIT 1;**

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

**Boston Crab Meat.**

**Query:**

**SELECT OrderDetails.OrderDetailID, OrderDetails.ProductID as Prodid,Products.ProductName, SUM(OrderDetails.Quantity) as total\_prod From OrderDetails,Orders,Customers,Products WHERE OrderDetails.ProductID = Products.ProductID AND Customers.Country = "Germany" AND Customers.CustomerID = Orders.CustomerID AND Orders.OrderID = OrderDetails.OrderID Group by OrderDetails.ProductID Order by total\_prod desc Limit 1;**