# Winter 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.

**The problem with the calculation is that the average is calculated using the mean when there are outliers that skew the data to the left. A better way to evaluate the data is to use the median, which would be 284$. This number more accurately represents the AOV.**

1. What metric would you report for this dataset?

**I would report the median value as the AOV because the data is skewed due to outliers.**

1. What is its value?

**The median value 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**

**s.ShipperName as Shipper, COUNT(distinct o.OrderID) as "Total Orders"**

**FROM**

**Shippers s, orders o**

**WHERE**

**s.ShipperID = O.ShipperID**

**AND s.ShipperID = 1**

**;**

**There was a total of 54 orders shipped by Speedy Express**

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

**SELECT**

**FirstName as "Employee First Name", LastName as "Employee Last Name", MAX(NumOrders) as Orders**

**FROM**

**(SELECT**

**e.FirstName, e.LastName, COUNT(distinct o.OrderID) as NumOrders**

**FROM**

**Employees e, orders o**

**WHERE**

**e.EmployeeID = o.EmployeeID**

**GROUP BY e.EmployeeID**

**)**

**;**

**The employee with the last name Peacock had the most orders with 40.**

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

**SELECT**

**Product, NumOrders as "Times Ordered", MAX(TotalQuantity) as "Total Quantity",**

**AvgQuantity as "AVG Quantity per Order"**

**FROM**

**(SELECT**

**p.ProductName as Product, count(od.productID) as NumOrders,**

**SUM(od.Quantity) as TotalQuantity, SUM(od.Quantity) / COUNT(od.productID) as AvgQuantity**

**FROM**

**Customers c, Orders o, OrderDetails od, Products p**

**WHERE**

**c.CustomerID = o.CustomerID**

**AND o.OrderID = od.OrderID**

**AND od.ProductID = p.ProductID**

**AND c.Country = "Germany"**

**GROUP BY 1**

**)**

**;**

**The product ordered the most by Germans was the Boston Crab Meat, which was ordered 4 times with an average order amount of 40, meaning the item was ordered 160 times.**