# 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) See below for responses in red.

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. Taking the average of the Order Amounts will not account for the variance in the size of the orders.
2. What metric would you report for this dataset? Average value per item
3. What is its value? $357.92

Python Code:

import pandas as pd  
df = pd.read\_csv(r'C:\Users\rache\OneDrive\Desktop\Shopify.csv')  
print(df)  
  
sum1 = df['order\_amount'].sum()  
sum2 = df['total\_items'].sum()  
quotient1 = sum1/sum2  
  
print ('Sum of Order Amount: ' + str(sum1))  
print ('Sum of Total Items: ' + str(sum2))  
print ('Average Value Per Item: ' + str(quotient1))

**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. See below for work in red and answers in blue.

1. How many orders were shipped by Speedy Express in total?
   * First SQL Statement:

**SELECT ShipperID**

**FROM [Shippers]**

**WHERE ShipperName = 'Speedy Express'**

* + - Speedy Express has Shipper ID 1
  + Second SQL Statement:

**SELECT COUNT (OrderID)**

**FROM [Orders]**

**WHERE ShipperID = 1**

* + - 54 orders

1. What is the last name of the employee with the most orders?
   * First SQL Statement:

**SELECT COUNT(OrderID), EmployeeID**

**FROM [Orders]**

**GROUP BY EmployeeID**

**ORDER BY COUNT(OrderID) DESC**

* + - Employee ID 4
  + Second SQL Statement

**SELECT LastName**

**FROM Employees**

**WHERE EmployeeID=4**

* + - Employee’s last name is Peacock

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

**SELECT OrderDetails.ProductID, OrderDetails.OrderID, Orders.CustomerID, Customers.Country, Products.ProductName**

**FROM OrderDetails**

**INNER JOIN Orders**

**ON OrderDetails.OrderID=Orders.OrderID**

**INNER JOIN Customers**

**ON Orders.CustomerID=Customers.CustomerID**

**INNER JOIN Products**

**ON OrderDetails.ProductID=Products.ProductID**

**GROUP BY Country**

**ORDER BY COUNT(ProductName) DESC**

* Boston Crab Meat