# 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 problem with this calculation is that it is calculating the AOV without considering the number of items. Consider the table below:

A screenshot of a computer

Description automatically generated with medium confidence

Here the large values of order amount also have very large total\_items values. This could be for a large/bulk order for other resellers/retailers. Since the AOV values does not consider the total\_items column, it takes the values 704000 as the amount for one pair of shoes.

1. What metric would you report for this dataset?

The metric I would use is average order value per item (AOVI). This would be the average Order Value per Item in the dataset.

1. What is its value?

The value for AOVI for the dataset would be $ 357.921

Graphical user interface, website

Description automatically generated

**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

Table

Description automatically generated

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

Margaret Peacock

Table

Description automatically generated

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

Gorgonzola Telino

Select Table1.Country, Table2.ProductName, Count(Table1.CustomerID) As Quantity

From

(Select C.CustomerID, O.OrderID, Country

From Customers as C

Inner Join Orders as O

On C.CustomerID = O.CustomerID) as Table1

Inner Join

(SELECT P.ProductID, OD.OrderID, ProductName

From Products As P

Inner Join OrderDetails as OD

On P.ProductID = OD.ProductID) as Table2

On Table1.OrderID = Table2.OrderID

Where Country = "Germany"

Group By Table2.ProductName

Order By Quantity Desc;

Table

Description automatically generated with medium confidence