# 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 issue with the current calculation is that every order is not of just one item but has multiple items in it. For example, order number 16 has a value of 704000 but that is for 2000 items. Hence the AOV of $3145.13 does not make any sense. Naively, it gives us the mean of order amounts disregarding the no of items ordered. This will serve us no purpose analysis. A better way of looking at this data is to calculate per item cost per order and then calculate the average sneaker price. Since each shop sells exactly one model of shoe, calculating the average sneaker price would help us investigate the range of shoe price and identify stores that are overselling or underselling.

1. What metric would you report for this dataset?

Average sneaker price

1. What is its value?

$387.74

This price is highly inflated due to few stores selling it for extremely nice price for example store no 42 sells it’s sneakers for $352. Store no 78 sells it’s sneakers for whooping 25725. These stores are selling the same model of shoe at a very high price, causing the average sneaker price to be way above what most of the stores sell it for. Removing the data for these two stores gives us a more acceptable average sneaker shoe price of $150.40

**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 count(\*) FROM Orders o Join Shippers s

on o.ShipperID = s.ShipperID

where s.ShipperID = 1;

Answer: 54

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

Select c.CustomerID,

SUBSTRING(c.CustomerName,LEN(c.CustomerName)-CHARINDEX(' ',REVERSE(c.CustomerName))+2, LEN(c.CustomerName)),

count(\*) as num\_orders from Customers c join Orders o

on c.CustomerID = o.CustomerID

group by c.CustomerID

order by num\_orders desc

limit 1

;

|  |  |  |
| --- | --- | --- |
| **CustomerID** | **CustomerName** | **num\_orders** |
| 20 | Handel | 10 |

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

SELECT P.ProductID,P.ProductName, count(\*) as maxOrder FROM OrderDetails OD, Products P, Orders O, Customers C

where OD.OrderID = O.OrderID and

OD.ProductID = P.ProductID and

O.CustomerID = C.CustomerID and

C.Country = "Germany"

group by P.ProductID

order by maxOrder desc

limit 1

;

|  |  |  |
| --- | --- | --- |
| **ProductID** | **ProductName** | **maxOrder** |
| 31 | Gorgonzola Telino | 5 |