# 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.
2. What metric would you report for this dataset?
3. What is its value?

**Question 1**

(a) Shop id 42 has large orders of 2000 units and Shop id 78 per unit price is coming as 25725 thus inflating the average of the whole dataset. Better way to evaluate data is to find per unit price of the sneaker sold by 100 shops, remove any noise and outliers before finding the AOV from the rest of the dataset.

(b) For the data set I would report the average selling price of sneaker to help analyse the average profit and the number of units sold can help to decide production to meet the demand.

(c) Average sneaker selling price (excluding shop id 42 and 78) is $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?
2. What is the last name of the employee with the most orders?
3. What product was ordered the most by customers in Germany?

**Question 2**

(a) SELECT COUNT(ORDERID) FROM [Orders] WHERE SHIPPERID = 1 GROUP BY SHIPPERID;

**Answer** 54

(b) select [employees].lastname, [orders].employeeid, count ([orders].orderid) from [orders] join [employees] on [orders].employeeid = [employees].employeeid group by [orders].employeeid order by count ([orders].orderid) DESC;

**Answer** Lastname Peacock employee id 4 number of orders 40

(c)SELECT [products].productname, [customers].country, [orders].customerid, [orderdetails].productid, count([orderdetails].productid) FROM [orders], [orderdetails], [customers], [products] where [orders].orderid = [orderdetails].orderid AND [orders].customerid = [customers].customerid AND [orderdetails].productid = [products].productid AND country = 'Germany' group by [products].productid order by count([orderdetails].productid) DESC;

**Answer** Product name Gorgonzola Telino