Summer 2022 Data Science Intern Challenge

**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 wrong calculation is because we considered count instead of sum for our dataset.

Also, as each of the 100 shops sell different models, the average value might not give the correct results as it doesn’t take the frequency of the values. A weighted average or median maybe a better metric for this data which is also robust to any outliers.

We can calculate a column as AOV = order\_amount/total\_items and analyse it further to find the trend for 30 days for different models. We can also see that there are some outliers (extreme values of 25725 as AOV), we can’t remove it from our data but we need to treat it separately after further analysis.

1. What metric would you report for this dataset?

I would consider Median a better metric here giving weightage to all the models AOV and being robust to outliers.

1. What is its value?

The Median of calculated AOV is 153.

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(OrderID) FROM Orders, Shippers Where Orders.ShipperID = Shippers.ShipperID and ShipperName = "Speedy Express"

Ans: 54

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

SELECT [Employees].LastName, COUNT (\*) AS NOrders

FROM [Orders]

JOIN [Employees]

ON [Orders].EmployeeID = [Employees].EmployeeID

GROUP BY [Employees].LastName

ORDER BY NOrders DESC

Ans: Peacock – 40

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

SELECT Products.ProductName FROM Orders INNER JOIN Customers ON Orders.CustomerID = Customers.CustomerID INNER JOIN OrderDetails ON Orders.OrderID = OrderDetails.OrderID INNER JOIN Products ON OrderDetails.ProductID = Products.ProductID WHERE Customers.Country = 'Germany' GROUP BY Products.ProductName ORDER BY COUNT(Products.ProductName) DESC

Ans: Gorgonzola Telino