Winter 2021 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.

*Working spreadsheet attached ->*****

1. **Think about what could be going wrong with our calculation. Think about a better way to evaluate this data.**

I found that the AOV of $3145.13 included some Outliers from bulk purchases that belong to multiple orders. A better way to evaluate this data is to filter out all of the Outliers. (Please have a **look at file above** ‘**Shopify-SneakerShop.xlsx’**) From the **FindingOutliers tabs** I calculated AOV without the Outlier data. I also performed calculations for AOV(cash), AOV(debit) and AOV(credit).

1. **What metric would you report for this dataset?**

The metric I would report on this data set is to setting the parameter to filter out all of the Outliers and I would apply multiple parameters depending on which AOV Payment type is being examined.

1. **What is its value?**

AOV Without Outliers (Cash, Debit, Credit): $754.09

AOV (Cash): $730.35

AOV (Debit): $966.84

AOV (Credit): $569.19

**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?**

The total Orders shipped by Speedy Express is **54.**

SQL Statement:

SELECT Shippers.ShipperName, COUNT(Orders.OrderID) AS TotalOrders

FROM Orders

LEFT JOIN Shippers ON Orders.ShipperID = Shippers.ShipperID

WHERE Shippers.ShipperName = 'Speedy Express'

GROUP BY ShipperName;

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

The last name of the employee with the most orders is: **Peacock**

SQL Statement:

SELECT Employees.LastName, COUNT(Orders.OrderID) AS TotalOrders

FROM Orders

LEFT JOIN Employees ON Orders.EmployeeID = Employees.EmployeeID

GROUP BY LastName

ORDER BY Count(\*) DESC;

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

The Product that was ordered the most by customers in Germany is: **Boston Crab Meat @ 160 units.**

SQL Statement:

SELECT SUM (OrderDetails.Quantity) as Total, Products.ProductName

FROM (((Customers

INNER JOIN Orders on Customers.CustomerID = Orders.CustomerID)

INNER JOIN OrderDetails on Orders.OrderID = OrderDetails.OrderID)

INNER JOIN Products on OrderDetails.ProductID = Products.ProductID)

WHERE Customers.Country = "Germany"

GROUP BY ProductName

ORDER BY SUM(OrderDetails.Quantity) DESC;