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

**Answer:**

1. The incorrect AOV calculation of $3145.13 was most likely arrived at by calculating the total items with the count() instead of sum(). Count() will provide the total count of the number of rows, while sum() will be required for the total items column in order to get the total number of orders.

Another reason for miscalculation of AOV is probable the average of order amount may be considered as AOV, instead of calculating AOV by total of order amount/total of total items

2. To determine the correct Average Order Value (AOV), the reporting metrics are the respective sums of both 'order\_amount' and 'total\_items':

OAmt = df.order\_amount.sum()

Titem = df.total\_items.sum()

Next, divide the total order amount (OAmt) by the total items amount (Titem):  
AOV = OAmt/Titem

print(AOV)

3. The Average Order Value (AOV) is: $357.92

VIEW CODE HERE: <https://github.com/DESAIKS/Shopify/blob/main/Shopify.ipynb>

**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.

**a. How many orders were shipped by Speedy Express in total?**

SELECT count(Orders.OrderID),Shippers.ShipperName

from Orders inner join Shippers

on Orders.ShipperID = Shippers.ShipperID

Where Shippers.ShipperName = "Speedy Express"

Answer: 54

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

SELECT Employees.LastName, count(Employees.LastName)

from Employees inner join Orders

on Employees.EmployeeID = Orders.EmployeeID

group by Employees.LastName

Answer: Peacock

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

Scenario1: The product that was ordered most frequently by customers in Germany.

SELECT Customers.Country, Products.ProductName, count(Products.ProductName)

from Customers

join Orders on Customers.CustomerID = Orders.CustomerID

join OrderDetails on Orders.OrderID = OrderDetails.OrderID

join Products on OrderDetails.ProductID = Products.ProductID

where Customers.Country = "Germany"

group by Products.ProductName order by count(Products.ProductName) desc limit 1

Answer: Gorgonzola Telino was ordered 5 times.

Scenario 2: If we are interested in the max Quantity of the product

SELECT Customers.Country, Products.ProductName, OrderDetails.Quantity

from Customers

join Orders on Customers.CustomerID = Orders.CustomerID

join OrderDetails on Orders.OrderID = OrderDetails.OrderID

join Products on OrderDetails.ProductID = Products.ProductID

where Customers.Country = "Germany"

group by Products.ProductID order by OrderDetails.Quantity desc limit 1

Answer: Steeleye Stout has been purchased 100 times