**SHOPIFY 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.

*Answer:*

*You want the average order amount, but each order may have more than one sneaker, thus the large AOV. We have to weigh each order by the total # of items by dividing the order\_amount by total\_items first to get the price of the sneaker on a per sneaker basis, then average the orders.*

*Code:*

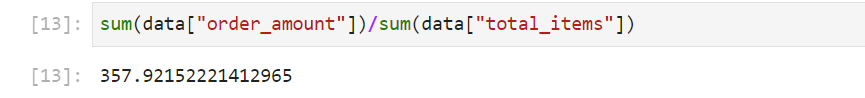
*Graphical user interface, text, application

Description automatically generated*

*I got on the average order value to be $387.74.*

*However, through this method, this does not account for the fact that some sneakers seem to be more popular than others, which is why some orders have many sneakers and some orders only have one sneaker. A way to account for this is to sum up the order\_amount and dividing it by the total number of sneakers purchased. This way, you also account for the amount of a sneaker being purchased. Through this method, I arrived at an AOV of $357.92.*

*Code:*



1. What metric would you report for this dataset?

*Answer: Average of price per sneaker per order.*

1. What is its value?

*Answer:*

*I got 387.74, assuming that you don’t account for the size (quantity of sneakers relative to other orders) of an order.*

*If you cared about the size of an order, you would get $357.92.*

**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?

***Answer:*** *68*

***Code:***

*SELECT COUNT(ShipperID)*

*FROM [Orders]*

*where ShipperID = 3*

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

***Answer:*** *Peacock*

***Code:***

*SELECT LastName FROM (SELECT LastName,COUNT(\*)*

*FROM (SELECT Orders.EmployeeID, Employees.LastName*

*FROM Orders*

*INNER JOIN Employees ON Orders.EmployeeID=Employees.EmployeeID)*

*group by LastName*

*order by COUNT(\*) desc)*

*LIMIT 1*

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

***Answer:*** *Boston Crab Meat*

***Code:***

*SELECT ProductName FROM(SELECT SUM(Quantity), ProductName FROM (SELECT Orders.CustomerID, Customers.CustomerID, Customers.Country, Orders.OrderID, OrderDetails.OrderID, Products.ProductID, OrderDetails.Quantity, OrderDetails.ProductID, 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 Country = 'Germany')*

*group by ProductName*

*Order by SUM(Quantity) desc)*

*LIMIT 1*