

**Question 1:** Given some sample data, write a program to answer the following: [click here to access the required data set](#)

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.

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

There are a couple of different reasons explaining what is going wrong with the calculation. The following reasons all have to do with the user with User ID 607.

- One explanation could be that they have their order\_amount and total\_items values multiplied by 1000 in which case the actual values should have been 704 and 2, respectively.
- Another explanation could be that there is fraudulent activity going on with User 607. They use a credit card as their payment method, purchase 2000 shoes totaling \$704,000 and consistently make these purchases at 4AM from the exact same store each time (shop\_id = 42). Another indication that this may be the case is that it is not a glitch in the system that is repeating the same sale as each of the orders have different Order IDs.

Regardless of the reasoning, User ID 607 is clearly skewing the calculation and has a very low chance of being entirely legitimate (*obviously* there exists people who can spend nearly 12 million dollars on shoes during one month, but I am keeping it realistic), therefore the user should be investigated and also removed from the database in order to maintain accurate data.

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

Other metrics that can be highlighted amongst the different shops are:

- Total orders made overall
- Total sales made overall
- Average access time amongst all users

**c. What is its value?**

- Total orders made overall = 5000
- Total sales made overall = 15725640
- Average access time amongst all users = 2017-03-15 22:20:37

**Question 2:** For this question you'll need to use SQL. [Follow this link](#) 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(OrderID)
from Orders, Shippers
where Orders.ShipperID = Shippers.ShipperID and Shippers.ShipperName = 'Speedy Express'
```

**Answer: 54**

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

```
select LastName
from Employees, Orders
where Employees.EmployeeID = Orders.EmployeeID
group by LastName
having count(OrderID) = (
select max(OrderCount)
from (
select LastName, count(OrderID) as OrderCount
from Employees, Orders
where Employees.EmployeeID = Orders.EmployeeID
group by LastName
))
```

**Answer: Peacock**

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

```
select ProductName
from Customers, Products, Orders, OrderDetails
where Customers.CustomerID = Orders.CustomerID and Orders.OrderID =
OrderDetails.OrderID and OrderDetails.ProductID = Products.ProductID and
Customers.Country = 'Germany'
group by ProductName
having count(ProductName) = (
select max(ProductCount)
from (
select ProductName, count(ProductName) as ProductCount
from Customers, Products, Orders, OrderDetails
where Customers.CustomerID = Orders.CustomerID and Orders.OrderID =
OrderDetails.OrderID and OrderDetails.ProductID = Products.ProductID and
Customers.Country = 'Germany'
group by ProductName
))
```

**Answer: Gorgonzola Telino**