Q1

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) There are a few anomalous purchases, where the order amount is disproportionately high, and the ratio between the order amount : total items look too weird.

We see 17 out of the 5000 purchase instances where User 607 bought 2000 total items at an almost daily base, artificially inflating the AOV. We also see 46 out of the 5000 purchase instances where the amount is in the ten thousands while the total items bought is low, thus giving an incredibly high order_amount/total_items ratio. The typical price of a pair of shoes range from $352 to $90. Only "ultra-luxury" pairs would cost in the ten thousands!

I also noticed that the same model of shoes can cost wildly different (error rather than sale). This is true for shoes sold at shop_id 78. Differing prices for purchases of the same shoe can be done via a natural join and checking the number of matched tuples.

To fix the data, we associate each model with a shop and a price. All purchases should be flagged if necesssary denoting if it was sale during a sale. Then we should include only the sale entries which do not look suspicious based on the order_amount/total_items ratio and whether the ratio matches the regular shoe model price associated with the shop.

b) What metric would you report for this dataset?

```
I would report the weighted average order_amount/total_items ratio. That
is, the total order amount recorded divided by the total items from valid
purchases.
```

c) What is its value?

```
The average cost of a shoe purchhased is 151.01$, if the 2000-purchases
purchased daily have been discounted and the mismatched
order_amount/total_items purchases have been deleted.
```

Q2:

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

```
SELECT * FROM Orders WHERE ShipperID=1
```

There are 54 such orders by Speedy Express (ID=1)

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

```
SELECT * FROM
    (SELECT Orders.EmployeeID AS EmployeeID, Count(OrderID) AS
EmployeeOrderCount FROM (Orders INNER JOIN Employees ON
Orders.EmployeeID=Employees.EmployeeID) GROUP BY (Orders.EmployeeID)) AS A
    INNER JOIN
    Employees
    ON A.EmployeeID = Employees.EmployeeID
```

Peacock (ID=4) has the most orders

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

```
SELECT SUM(Quantity) AS Sold, ProductID FROM
    (Orders INNER JOIN Customers ON Customers.CustomerID=Orders.CustomerID)
INNER JOIN OrderDetails ON Orders.OrderID=OrderDetails.OrderID
WHERE Country="Germany" GROUP BY ProductID ORDER BY SUM(Quantity)
```

Item #40 was purchased 160 times by German customers