Question 1

a

The AOV is being influenced by outliers in the data caused by bulk orders and especially rare sneakers. Instead of looking at the total values of the orders as individuals, it would be better to analyze the value of the items sold.

b

This metric would be average item value (AIV) and would be calculated by taking the sum of the order values (SOV) and dividing by the total number of items sold (N).

c

```
from google.colab import drive
drive.mount('/content/gdrive')


import pandas as pd


path_to_data = "/content/gdrive/My Drive/2021_Data_Intern_Challenge/data.xlsx"
data         = pd.read_excel(path_to_data)


SOV = data.order_amount.sum()
N   = data.total_items.sum()
AIV = SOV/N

print(AIV)

    357.92152221412965
```

The AIV is $357.92

Question 2

```
SELECT count(*) FROM Orders where ShipperID=1;
```

b

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

```
SELECT LastName FROM Employees where EmployeeID IN
    (SELECT TOP 1
        EmployeeID
    FROM
      (SELECT EmployeeID, COUNT(*)
      FROM Orders
      GROUP BY EmployeeID
      ORDER BY COUNT(*) DESC));
```

c

The product ordered the most by customers in Germany was Gorgonzola Telino

```
SELECT ProductName FROM Products where ProductID IN
  (SELECT TOP 1
      ProductID
  FROM
    (SELECT ProductID, COUNT(Quantity) FROM OrderDetails
    where OrderID IN
      (SELECT OrderID
      FROM Orders
      where CustomerID IN
          (SELECT CustomerID FROM Customers where Country='Germany'))
    GROUP BY ProductID
    ORDER BY COUNT(Quantity) DESC));
```