**Question 1)**

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

A). In order amounts, we can see that the user with user id has bulk purchases, so a total of 17 orders have a bulk quantity of 2000, and that is why the total order amount is too high for these orders. Another thing is that in some stores, price of the shoe is too high. So, the total order amount is also high (In store with id 78, shoe price is 25,725 dollars, this seems unrealistic because we have data of same model of shoe). So, in short, we can say we have **many outliers** in this data. Here, we have used average(mean) to represent the central tendency of the dataset. A **mean** is **sensitive** to **outliers** and that is why our AOV is looking unrealistic. According to me, one can approach this problem in two ways, one is to find and remove the outlier and another way is to use **the median to represent central tendency** instead of the mean.

**Q-b) What metric would you report for this dataset?**

A). I will use **the median** to **represent the central tendency**of the dataset. The reason behind using median over mean is that **median is not sensitive to outliers**.

       Another method we can use is to remove the outliers. Here, as we can see, Q1 is 163 and Q3 is 390, so IQR is 227 for the order amount. So, all the orders with an amount greater than 730.5 are outliers (as per five-number summary). So, after removing those outliers, we can find the mean.

**Q-c) What is its value?**

A). The median of order amount is **284 dollars**.

If we remove the outliers and then find the AOV (mean), it is **293.71 dollars**.

**Question 2.)**

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

A). 54

Query:-

**SELECT**

**Shippers.ShipperName as shipper\_name,**

**COUNT(\*) as total\_orders**

**FROM Orders**

**INNER JOIN Shippers**

**ON Orders.ShipperId=Shippers.ShipperId**

**WHERE Shippers.ShipperName = "Speedy Express";**

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

A). Peacock (40 orders)

Query:-

**SELECT Employees.LastName as LastName,**

**COUNT(\*) as Total\_Orders FROM Orders**

**INNER JOIN Employees**

**ON Orders.EmployeeID = Employees.EmployeeID**

**GROUP BY Orders.EmployeeId**

**ORDER BY 2 DESC LIMIT 1;**

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

A). Boston Crab Meat (160 Quantity) (I have calculated total quantity of each product ordered by Customers in Germany and then find the maximum of it)

Query:-

**SELECT Products.ProductName, SUM(OrderDetails.Quantity) as Total\_Quantity**

**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 Customers.Country = "Germany"**

**GROUP BY OrderDetails.ProductID**

**ORDER BY 2 DESC LIMIT 1;**