

QUESTION 1:

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

Average order value (AOV) is an inaccurate representation of the average item value extracted from a customer, because in an order, a customer can buy more than one item/ sneaker. As a result of that, the AOV, which is calculated as the sum of amount of all orders/ number of orders, should be interpreted and evaluated as the average order/ transaction amount, regardless of the number of items in an order. The AOV was not calculated wrong, it is just the interpretation that was wrong. Therefore, the \$3145.13 amount just shows the total value of an order, not a sneaker/item.

However, if one want to track specifically the performance of sneakers, there are three ways:

1. Take the median of the order amount. This helps to reduce the magnitude/ effects of large orders, however, will only work in the case that most orders are of small amount(1 or 2 items most). This is not the best way to evaluate the company's performances.
2. For each order, divide the order amount by number of items to get the amount in dollar values per item. We can take the mean/ median of all orders dollar value per item to get the average/median item value in the month. However, this does not take into account the number of items sold in the month.
3. ***We calculate the sum of order amount and divide by sum of items sold in the month. This is a good way because it takes into account the number of items sold. This can be called AIV- Average Item Value.***

b. What metric would you report for this dataset?

- AIV, showing in a month period, what is average item value purchased by customers.
 - Number of customers in a month
 - Total order amount & total items sold in the month
- ➔ All of the above metrics can be used in the payment_method level as well as the shop levels, and can also go deeper vertically into daily statistics for easier tracking of performance. It might help also incorporate other statistics like standard deviation that measures uncertainty and variability, or unique count(as in cases of customers). For a larger dataset (in months or years), metrics like retention rate, FRM metrics (frequency, recency, monetary) can also play a greater role.

c. What is its value?

- AIV= \$357.92\$. Notice that in the dataset, there are some days with large orders of expensive sneakers/ items(or maybe bundles) of \$25,725. If we filter that out the AIV is \$307.01.
- There are 301 unique customers in the last month, although there are 5000 transactions, with all of them buying more than 1 order.
- Total order amount at \$15,725,640.
- Total items sold of 43,936.

For more insights/ graphs/ metrics, please follow this link:

[https://github.com/longhlerotman/ShopifyInternChallenge/blob/main/Shopify%20Intern%20Challenge%20\(1\).ipynb](https://github.com/longhlerotman/ShopifyInternChallenge/blob/main/Shopify%20Intern%20Challenge%20(1).ipynb)

QUESTION 2

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

Query:

```
WITH T2 AS (SELECT a.*,b.ShipperName FROM Orders a LEFT JOIN Shippers b on  
a.ShipperID=b.ShipperID) SELECT COUNT(*) FROM T2 WHERE ShipperName='Speedy  
Express';
```

Number of Records: 1

54 orders were shipped in total by
Speedy Express.

| COUNT(*) |
|----------|
| 54 |

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

Query:

```
SELECT LastName FROM employees WHERE EmployeeID = (SELECT  
EmployeeID from (SELECT EmployeeID, count(OrderID) from [Orders] Group By  
EmployeeID LIMIT 1));
```

Number of Records: 1

Last name of the employee with most
orders is **Davolio**.

| LastName |
|----------|
| Davolio |

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

Query:

```
WITH temptable as (SELECT a.*,b.ProductName, c.CustomerID,d.Country FROM  
OrderDetails a LEFT JOIN Products b on a.ProductID=b.ProductID LEFT JOIN  
Orders c on a.OrderID=c.OrderID LEFT JOIN Customers d on  
c.CustomerID=d.CustomerID WHERE d.Country='Germany')
```

```
SELECT ProductName,COUNT(*) AS NUMORDERS, SUM(Quantity) AS  
TOTALQUANTORDER FROM temptable  
GROUP BY ProductName  
ORDER BY COUNT(*) DESC,SUM(Quantity) DESC LIMIT 10;
```

Number of Records: 10

| ProductName | NUMORDERS | TOTALQUANTORDER |
|----------------------------|-----------|-----------------|
| Gorgonzola Telino | 5 | 125 |
| Boston Crab Meat | 4 | 160 |
| Lakkalikööri | 4 | 75 |
| Tunnbröd | 3 | 105 |
| Mozzarella di Giovanni | 3 | 86 |
| Inlagd Sill | 3 | 72 |
| Teatime Chocolate Biscuits | 2 | 95 |
| Chang | 2 | 84 |
| Camembert Pierrot | 2 | 75 |
| Flotemysost | 2 | 75 |

In terms of the number of orders, **Gorgonzola Telino** is the most ordered product of customers in Germany with 5 times appearance on the orders. In terms of number of quantities ordered regardless of number of orders, **Boston Crab Meat** is the 1st in the list (as illustrated in the table above) with total order quantity of 160.