# Fall 2022 Data Science Intern Challenge

**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.

Ans a) According to my observation:

- The calculation can be wrong because several bulk orders (order amounts of 704000, 25725, and so on) have significantly distorted the mean or AOV. Although these bulk orders may be significant, we can remove the outliers and calculate the mean. Another option is to identify the median and use that to get the AOV estimate.
- Another way to solve this problem would be to calculate the price_per_item (order_amount/total_items) and then take an average of the column.

b. What metric would you report for this dataset?

Ans b.) I would recommend using median and mean after removing the outliers as the metric or price_per_item as metric. The outliers that should be removed from the data 1.5 times above the third quartile or 1.5 times below the first quartile.

c. What is its value?

Ans c.) The median value is 284 and the mean value after removing the outliers is 293.71. For the second approach, the mean of price_per_item is 387.74.

**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?

**Query:** SELECT COUNT(OrderID) AS No_of_Orders
FROM Orders JOIN Shippers
ON Orders.ShipperID = Shippers.ShipperID
AND Shippers.ShipperName = "Speedy Express";

**Answer:** 54

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

**Query:** WITH temptable AS (
SELECT EmployeeID
FROM Orders
GROUP BY EmployeeID
ORDER BY COUNT(EmployeeID) DESC
LIMIT 1)
SELECT LastName
FROM Employees JOIN temptable
ON Employees.EmployeeID = temptable.EmployeeID;

**Answer:** Peacock

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

**Query:** WITH temptable1 AS
(SELECT CustomerID FROM Customers Where Country = 'Germany'),
temptable2 as (SELECT OrderID FROM Orders JOIN temptable1
ON Orders.CustomerID = temptable1.CustomerID),
temptable3 as (SELECT ProductID, Quantity FROM OrderDetails JOIN temptable2
ON OrderDetails.OrderID = temptable2.OrderID),
temptable4 as (SELECT ProductID FROM temptable3
GROUP BY ProductID
ORDER BY COUNT(ProductID) DESC
LIMIT 1)
SELECT ProductName
FROM Products JOIN temptable4
ON Products.ProductID = temptable4.ProductID;

**Answer:** Gorgonzola Telino