

Summer 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.**
 - i. On inspecting the column stats for order amount within the data set, it is observed that there are sneakers sold in bulk that result in skewing the average order value. Separating the shops into bulk and not bulk could assist in evaluating the AOV effectively. Additionally, reevaluating the measures of central tendency using mean, median and mode may provide better insights.
- b. **What metric would you report for this dataset?**
 - i. Using the mode, we get a more accurate AOV
- c. **What is its value?**
 - i. \$153

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?**
 - a. 54 orders
- b. **What is the last name of the employee with the most orders?**
 - a. Peacock
- c. **What product was ordered the most by customers in Germany?**
 - a. Product Number 31 "Gorgonzola Telino" was ordered the most (5 times) by customers in Germany.

SQL Code Screenshots

SQL Statement:

```
SELECT ShipperID, COUNT (*) FROM [Orders]
WHERE ShipperID = 1
```

Edit the SQL Statement, and click "Run SQL" to see the result.

Run SQL »

Result:

Number of Records: 1

ShipperID	COUNT (*)
1	54

SQL Statement:

```
SELECT EmployeeID, COUNT(EmployeeID)
FROM Orders GROUP BY EmployeeID
HAVING COUNT (EmployeeID)=(
SELECT MAX(mycount)
FROM (
SELECT EmployeeID, COUNT(EmployeeID) mycount
```

Edit the SQL Statement, and click "Run SQL" to see the result.

[Run SQL »](#)

Result:

Number of Records: 1

EmployeeID	COUNT(EmployeeID)
4	40

SQL Statement:

```
SELECT Orders.OrderID, Customers.Country, OrderDetails.OrderID, Products.ProductID, Products.ProductName
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 Country = "Germany" ORDER BY Products.ProductID ASC;
```