Winter 2021 Data Science Intern Challenge

Please complete the following questions, and provide your thought process/work. You can attach your work in a text file, link, etc. on the application page. Please ensure answers are easily visible for reviewers!

**Question 1:** Given some sample data, write a program to answer the following: [click here to access the required data set](https://docs.google.com/spreadsheets/d/16i38oonuX1y1g7C_UAmiK9GkY7cS-64DfiDMNiR41LM/edit#gid=0)

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.

1. Think about what could be going wrong with our calculation. Think about a better way to evaluate this data.
2. What metric would you report for this dataset?
3. What is its value?

The median value should be used because it is less impacted by large outlier values.

If you calculate the median of the order values you get $284 which is more accurate for sneaker purchases.

Program:

#include <iostream>

#include <algorithm>

int calcAverage() {

// dataset over 30 day period

int my\_order\_amounts[array\_size] = {dataset};

// sort the array in ascending order

std::sort(my\_order\_amounts[0], my\_order\_amounts[array\_size - 1]);

if (array\_size % 2) {

return my\_order\_amounts[array\_size / 2 + 1];

} else {

return (my\_order\_amounts[array\_size / 2] +

my\_order\_amounts[array\_size / 2 - 1] \ 2);

}

}

**Question 2:** For this question you’ll need to use SQL. [Follow this link](https://www.w3schools.com/SQL/TRYSQL.ASP?FILENAME=TRYSQL_SELECT_ALL) 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.

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

Final answer: 54

SELECT COUNT () FROM Orders WHERE ShipperID = (SELECT ShipperID FROM Shippers WHERE ShipperName='Speedy Express');

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

Final answer: Peacock

SELECT LastName FROM Employees WHERE EmployeeID = (SELECT EmployeeID FROM (SELECT `EmployeeID`, COUNT(`EmployeeID`) AS `occurrence` FROM `Orders` GROUP BY `EmployeeID` ORDER BY `occurrence` DESC));

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

Assumption: Counting of products ordered is by order and does not consider quantity ordered per order

Final answer; Gorgonzola Telino

SELECT ProductName FROM Products WHERE ProductID = (SELECT ProductID FROM (SELECT ProductID, COUNT(\*) AS `value\_occurrence` FROM (SELECT ProductID FROM Orders INNER JOIN OrderDetails ON Orders.OrderID = OrderDetails.OrderID WHERE CustomerID IN (SELECT CustomerID FROM Customers WHERE Country = 'Germany')) GROUP BY `ProductID` ORDER BY `value\_occurrence` DESC ));