# 

# Kevin Teresi

January, 2022

# Summer 2022 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.

Taking a look thoroughly at the data, there are two major things going on with the data that need to be addressed. First of all, there are outlier sales that are completely messing up the average order value. In these, it appears that the customer is buying in bulk, for whatever reason. It could be that they are reselling the shoes, or maybe they are just huge fans of the shoes. The second thing that is going on is that there are some sales that make absolutely no sense. For example, there are some sales of 1 item for over $5000. This could have been an entry-error, but it does need to be addressed before the store continues to move on making sales.

When looking at data that contains outliers as such, or when looking at data that is significantly skewed, it doesn’t make much sense to take the mean value in any scenario. Clearly, the outliers cause the mean value to be much greater or smaller than a number that makes sense. In these kinds of situations, we want to use a different metric, such as Median.

1. What metric would you report for this dataset?

Median is typically the measure of central tendency when there are outliers in data. We want to take a look at the Median Order Value. The median denotes the value at the midpoint of all the sales values if they were lined up in order. This value has an equal number of values before and after it in a number line. In Excel, we will just use the Median() function over the entire column of order values.

1. What is its value?

MOV: $284.00. This makes a ton of sense. Looking at individual sales of 1 shoe, the sales typically range between $90 and $160. We can assume with the MOV of $284 that shoppers are buying two pairs of sneakers or less, half of the time.

**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?

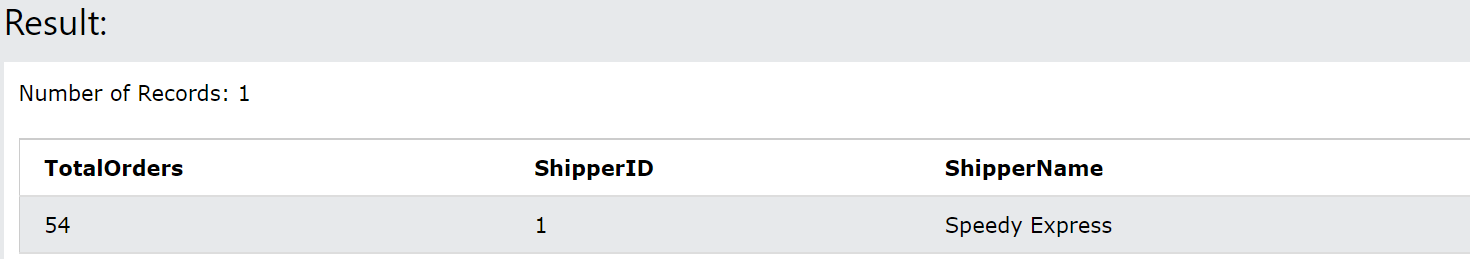
SELECT count(\*) AS TotalOrders, Shippers.ShipperID, Shippers.ShipperName AS ShipperName

FROM orders

INNER JOIN Shippers ON Orders.ShipperID = Shippers.ShipperID

WHERE Shippers.ShipperID == 1

GROUP BY Orders.ShipperID, Shippers.Shippername



Answer: 54 total orders by Speedy Express

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

SELECT count(\*) AS TotalOrders, Orders.employeeID, Employees.LastName AS LastName

FROM Orders

INNER JOIN Employees ON Orders.EmployeeID = Employees.EmployeeID

GROUP BY Orders.EmployeeID, Employees.LastName

ORDER BY TotalOrders DESC

LIMIT 1



Answer: Peacock

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

SELECT Country, SUM(OrderDetails.Quantity) as TotalQuantitySold, Products.ProductName

From Customers

INNER JOIN Orders ON Customers.CustomerID = Orders.CustomerID

INNER JOIN OrderDetails ON Orders.OrderID = OrderDetails.OrderID

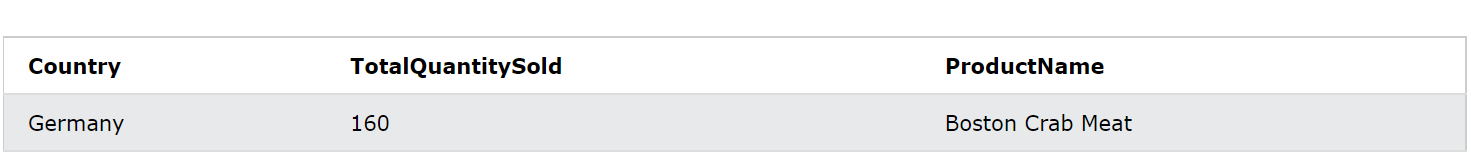
INNER JOIN Products on OrderDetails.ProductID = Products.ProductID

WHERE Customers.Country = "Germany"

Group By Country, Products.ProductName

Order By TotalQuantitySold DESC

LIMIT 1



Answer: Boston Crab Meat