

Fall 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](#)

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.

Initial thoughts as to the cause is the potential for the data to contain outlier data that is skewing the average order amount by a large degree. Since this is a small dataset, I ran a query to select all the rows and to sort them by order_amount with descending values to see the largest values first. Doing this revealed 17 orders that contained an order_amount value of 704000 with a total_items value of 2000. To get a more accurate AOV we can make sure not to count these values to prevent skewing.

- b. What metric would you report for this dataset?

If we ignore the large order_amount orders, utilizing the average order value should now be more accurate.

c. What is its value?

Query Used to Calculate Average without outlier orders:

```
SELECT CAST(AVG(order_amount) AS DECIMAL(20,2)) FROM dbo.ChallengeData  
WHERE ChallengeData.order_amount < 704000;
```

The AOV value is now: 754.09

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 "Speedy Express Shipments" FROM Orders
JOIN Shippers ON Shippers.ShipperID = Orders.ShipperID
WHERE Shippers.ShipperName = "Speedy Express"
```

Numerical Answer: 54

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

Query:

```
SELECT Employees.LastName AS "Last Name", COUNT(OrderID) AS "Number of
Orders" FROM Employees
JOIN Orders ON Orders.EmployeeID = Employees.EmployeeID
GROUP BY Employees.EmployeeID
ORDER BY COUNT(Orders.OrderID) DESC
LIMIT 1
```

Numerical Answer: Peacock : 40

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

Query:

```
SELECT Products.ProductName AS "Product Name", SUM(OrderDetails.Quantity) AS
"Total Quantity Ordered" FROM Products
JOIN OrderDetails ON OrderDetails.ProductID = Products.ProductID
JOIN Orders ON Orders.OrderID = OrderDetails.OrderID
JOIN Customers ON Customers.CustomerID = Orders.CustomerID
WHERE Customers.Country = "Germany"
GROUP BY Products.ProductID
ORDER BY SUM(OrderDetails.Quantity) DESC
LIMIT 1
```

Numerical Answer: Boston Crab Meat : 160