

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.

There are outliers from user 607 placing multiple orders of 2000 items per order. After removing the outliers the AOV decreased to 754.09.

- b. What metric would you report for this dataset?
I would consider using median after removing outliers as a metric for this dataset as the distributions of order amount and total items are very uneven.

- c. What is its value?
The median value of order amount after removing the outliers is 284.0.

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?

```
SELECT COUNT(OrderID) AS OrderCount
FROM Orders
WHERE ShipperID IN
(
  SELECT ShipperID
```

```

FROM Shippers
WHERE ShipperName = 'Speedy Express'
)

```

Number of Records: 1

OrderCount
54

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

SELECT e.LastName, COUNT(o.OrderID) AS NumberOfOrders
FROM Orders o
INNER JOIN Employees e
ON o.EmployeeID = e.EmployeeID
GROUP BY e.LastName
ORDER BY COUNT(o.OrderID) DESC
LIMIT 1

```

Number of Records: 1

| LastName | NumberOfOrders |
|----------|----------------|
| Peacock  | 40             |

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

SELECT p.ProductName, SUM(od.Quantity) AS OrderQuantity
FROM Customers c
INNER JOIN Orders o
ON c.CustomerID = o.CustomerID
INNER JOIN OrderDetails od
ON o.OrderID = od.OrderID
INNER JOIN Products p
ON od.ProductID = p.ProductID
WHERE c.Country = 'Germany'
GROUP BY p.ProductName
ORDER BY SUM(od.Quantity) DESC
LIMIT 1

```

Number of Records: 1

ProductName	OrderQuantity
Boston Crab Meat	160