

# 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](#)

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.

What we notice is that there are a few large orders which tends to skew the dataset. This is what causes the large AOV obtained. We get a more reasonable AOV of 301.07\$ once we remove all rows with an order amount above 1000\$.

**For more details and steps, see RMarkdown File.**

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

**We can examine the median order value rather than the mean which is a better indication in our situation when observing the entire dataset.**

- c. What is its value?

**The median order value obtained is 284\$ which is not far off from the AOV on the cleaned dataset.**

**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(*) as 'Total Orders Shipped', ShipperID
FROM Orders
WHERE ShipperID = 1
```

**Alternatively with Shipper Name:**

```
SELECT count(*) as 'Total Orders Shipped', Orders.ShipperID, ShipperName
FROM Orders
JOIN Shippers
ON Orders.ShipperID = Shippers.ShipperID
WHERE Shippers.ShipperName == 'Speedy Express'
```

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

The last name of the employee with the most orders: **Peacock**

```
SELECT Orders.EmployeeID, FirstName, LastName, count(*) as TotalOrders
FROM Orders
JOIN Employees
ON Orders.EmployeeID = Employees.EmployeeID
GROUP BY Orders.EmployeeID
ORDER BY TotalOrders DESC
```

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

Product most order by customers in Germany: **Boston Crab Meat**

```
SELECT Products.ProductID, ProductName, SUM(Quantity) as Sum
FROM Orders
JOIN OrderDetails ON OrderDetails.OrderID = Orders.OrderID
JOIN Customers ON Orders.CustomerID = Customers.CustomerID
JOIN Products ON OrderDetails.ProductID = Products.ProductID
WHERE Country = 'Germany'
GROUP BY Products.ProductID
ORDER BY Sum DESC
```