# 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](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.

We are getting this result because we are using the mean and by observation of data we see we have some values that are outliers and thus push our mean up. Proof of this is a large standard deviation of 41282.539349.Such a large standard deviation makes the mean meaningless/useless. A better way to evaluate this data would be to check the skewness of the data and then exclude outliers when doing calculations.

1. What metric would you report for this dataset?

I would report the median of the order amount. This is because this is a better indicator of what a customer’s order amount usually is which is the purpose of an AOV.

1. What is its value?

Its value is 284

**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(\*)

FROM Shippers AS S JOIN Orders AS O

ON O.ShipperID=S.ShipperID

WHERE S.ShipperName="Speedy Express";

Answer is: 54

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

SELECT E.LastName

FROM (SELECT O.EmployeeID, COUNT(\*) AS numorder

FROM Orders AS O

GROUP BY O.EmployeeID) AS I JOIN Employees AS E

ON E.EmployeeID=I.EmployeeID

WHERE I.numorder=(SELECT MAX(I.numorder) AS max

FROM (SELECT O.EmployeeID, COUNT(\*) AS numorder

FROM Orders AS O

GROUP BY O.EmployeeID) AS I);

Answer: Peacock

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

SELECT \*

FROM(SELECT ProductID, SUM(Quantity) AS quant

FROM (SELECT \*

FROM (SELECT \*

FROM Customers AS C

WHERE C.Country="Germany") AS C JOIN Orders AS O

ON O.CustomerID=C.CustomerID) AS I JOIN OrderDetails AS OD

ON OD.OrderID=I.OrderID

GROUP BY ProductID) AS X JOIN Products AS P

ON P.ProductID=X.ProductID

WHERE quant= (SELECT MAX(quant) FROM (SELECT ProductID, SUM(Quantity) AS quant

FROM (SELECT \*

FROM (SELECT \*

FROM Customers AS C

WHERE C.Country="Germany") AS C JOIN Orders AS O

ON O.CustomerID=C.CustomerID) AS I JOIN OrderDetails AS OD

ON OD.OrderID=I.OrderID

GROUP BY ProductID))

;

Answer: Boston Crab Meat

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **ProductID** | **quant** | **ProductName** | **SupplierID** | **CategoryID** | **Unit** | **Price** |
| 40 | 160 | Boston Crab Meat | 19 | 8 | 24 - 4 oz tins | 18.4 |