# 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.

Answer: The issue with the calculation is that the average is determine by only using values in the order\_amount column, where the average should be taken from the sum after dividing the order\_amount value with the total\_items value.

1. What metric would you report for this dataset?

Answer: While using AOV would be a useful metric, it would be beneficial to using multiple metrics such as AOV, Mean, Median, Mode and standard deviation.

1. What is its value?  
   The new AOV after applying the solution above is 387.74

**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?

Answer: 54 orders were shipped by Speedy Express in total.

Solution:

SELECT Shippers.ShipperName, Shippers.ShipperID, Orders.ShipperID, COUNT (Shippers.ShipperID)

FROM Shippers

JOIN Orders

ON Shippers.ShipperID=Orders.ShipperID

WHERE ShipperName ="Speedy Express"

GROUP BY Shippers.ShipperID

ORDER BY Shippers.ShipperID;



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

Answer: The last name of the employee with the most orders is Peacock.

Solution:

SELECT Employees.EmployeeID,Orders.EmployeeID, Employees.LastName,Count (Orders.EmployeeID) as NumberofOrders

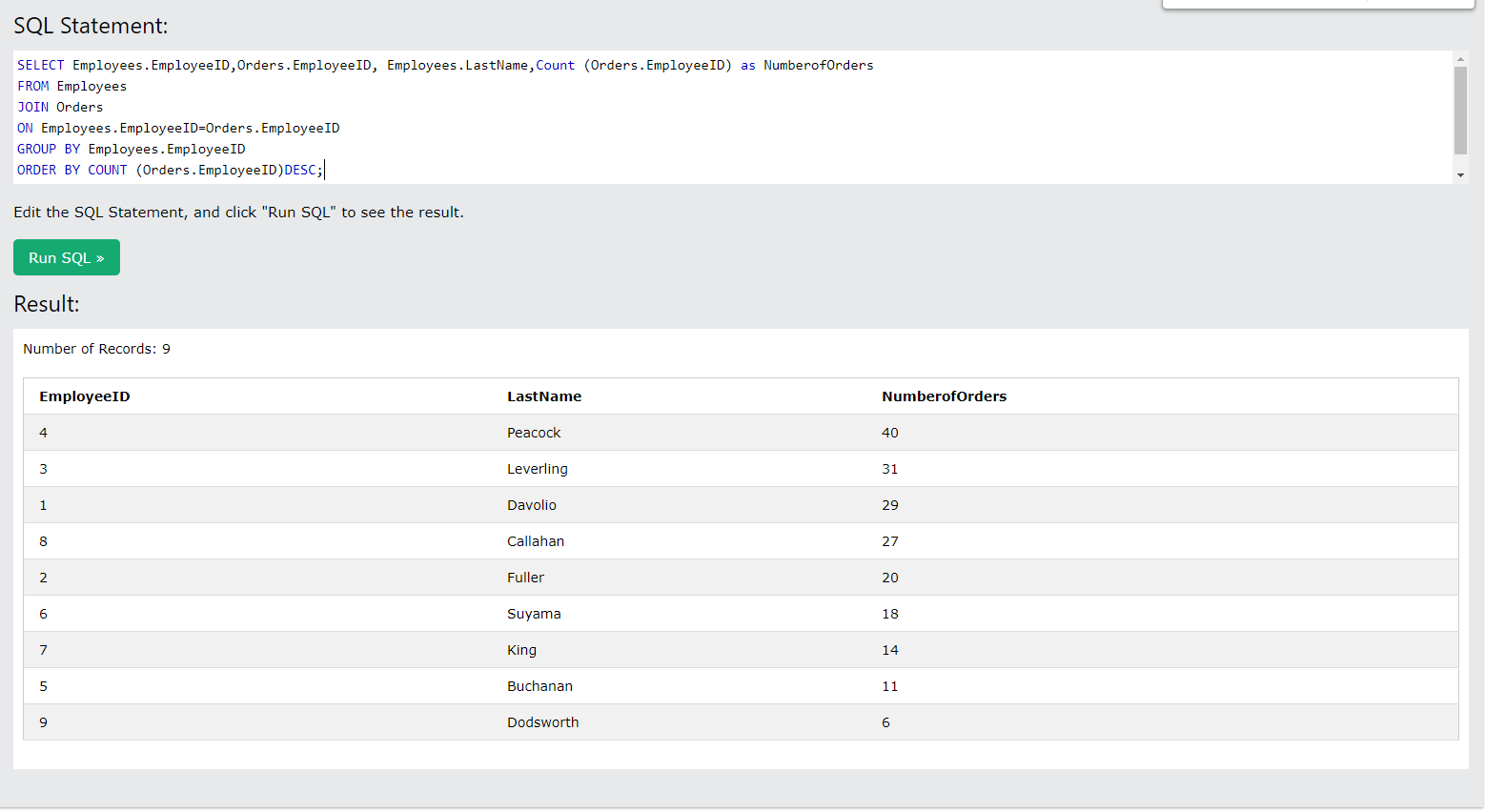
FROM Employees

JOIN Orders

ON Employees.EmployeeID=Orders.EmployeeID

GROUP BY Employees.EmployeeID

ORDER BY COUNT (Orders.EmployeeID)DESC;



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

Answer: The product that was ordered the most by customers in Germany is Gorgonzola Telino with the ProductID of “31” with the total of 5 orders.

Solution:

SELECT Customers.CustomerID,Customers.Country, OrderDetails.OrderID,OrderDetails.ProductID, Products.ProductName, COUNT (OrderDetails.ProductID) AS NumerofOrder

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 OrderDetails.ProductID

ORDER BY COUNT (OrderDetails.ProductID)DESC;

