# 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](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.
2. What metric would you report for this dataset?
3. What is its value?

**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?
2. What is the last name of the employee with the most orders?
3. What product was ordered the most by customers in Germany?

**Ans 1: Solution is in Google drive link with named as “Ans 1.ipnyb” with excel file named “Solved.xlsx”**

**Ans 2:**

(A)

Query:

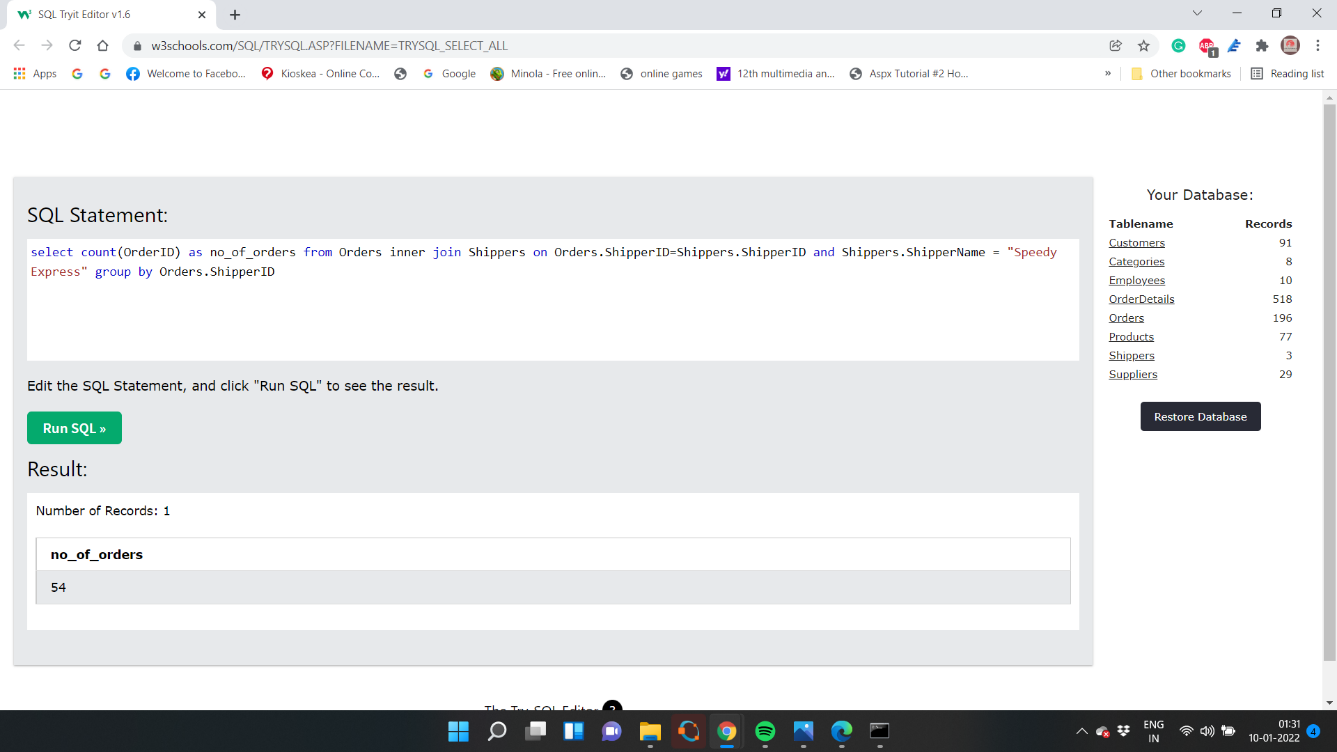
select count(OrderID) as no\_of\_orders

from Orders

inner join Shippers on Orders.ShipperID=Shippers.ShipperID and Shippers.ShipperName = "Speedy Express"

group by Orders.ShipperID

Output:



(b)

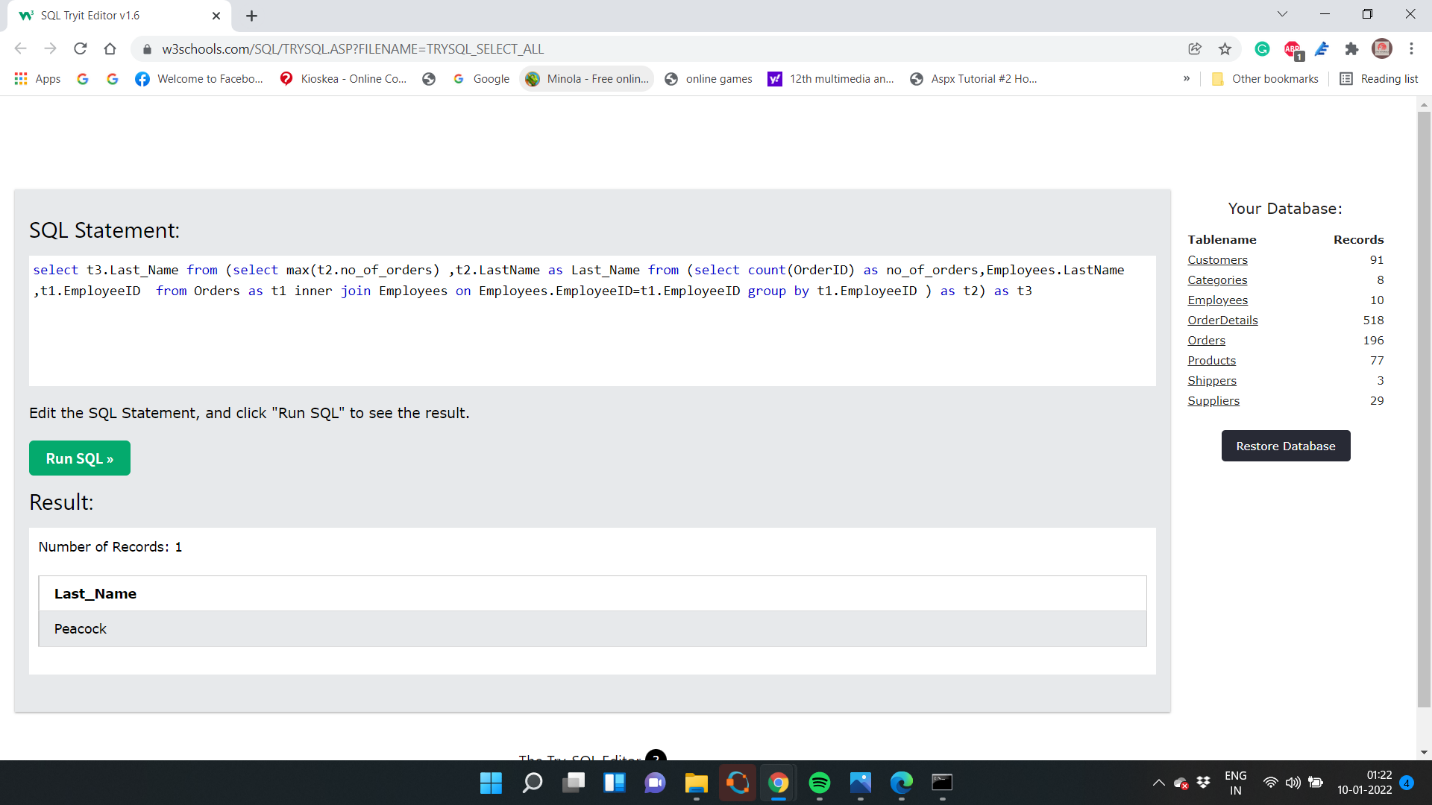
Query:

select t3.Last\_Name

from (select max(t2.no\_of\_orders) ,t2.LastName as Last\_Name

from (select count(OrderID) as no\_of\_orders,Employees.LastName ,t1.EmployeeID from Orders as t1 inner join Employees on Employees.EmployeeID=t1.EmployeeID group by t1.EmployeeID ) as t2) as t3

Output:



(c)

Query:

select t3.ProductName From (select max(t2.max),ProductName from Products inner join (SELECT Customers.CustomerID,t1.ProductID,t1.max FROM Customers inner join (SELECT max(Quantity) as max,Orders.CustomerID,Orders.OrderID,OrderDetails.ProductID FROM OrderDetails inner join Orders on OrderDetails.OrderID=Orders.OrderID group by OrderDetails.OrderID) as t1 on Customers.CustomerID=t1.CustomerID where Customers.Country="Germany" group by t1.OrderID)as t2 on Products.ProductID=t2.ProductID) as t3

Output: 