# Fall 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.

**See jupyter notebook for details:**

https://github.com/TonyHFu/shopify/blob/master/shopify.ipynb

1. Think about what could be going wrong with our calculation. Think about a better way to evaluate this data.
   * 1. **The mean is distorted because of extremely large values**
2. What metric would you report for this dataset?
   * 1. **The median**
3. What is its value?
   * 1. **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 Shippers.ShipperName, COUNT(\*)

FROM Orders

INNER JOIN Shippers

ON Orders.ShipperID=Shippers.ShipperID

GROUP BY Shippers.ShipperName;

* 1. **54**

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

SELECT Employees.LastName, COUNT(\*) as NumOrders

FROM Employees

INNER JOIN Orders

ON Orders.EmployeeID=Employees.EmployeeID

GROUP BY Employees.LastName

ORDER BY NumOrders DESC;

* 1. **Peacock**

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

SELECT ProductName, SUM(OrderDetails.Quantity) as QuantityOrdered, Customers.Country

FROM Customers

INNER JOIN Orders

ON Customers.CustomerID=Orders.CustomerID

JOIN OrderDetails

ON OrderDetails.OrderID=Orders.OrderID

JOIN Products

ON Products.ProductID=OrderDetails.ProductID

WHERE Customers.Country="Germany"

GROUP BY ProductName

ORDER BY QuantityOrdered DESC;

* 1. **Boston Crab Meat**