# Winter 2022 Data Science Intern Challenge

Question 1:

The main issue with the calculation is that there is no filtering of the data. Looking at the dataset, there are two stores that are outliers: Store 42, where every order is for 2000 pairs of shoes, and Store 78, where the price per item is $25725. If possible, I would first investigate these two stores further, to see if the data has been corrupted somehow, or if these are just non-traditional shoe stores. In any case, the data from these two stores should not be included in this assessment, and so must be filtered out.

I think it is useful to know the AOV of all the sneaker shops, which I calculated to be $302.58. However, I think it would be more useful to see how the AOV changes per store. Rather than list 98 numbers, I created a histogram showing the AOV distribution. As seen, people typically spend between $250-$350, with the median value consistent with the combined AOV for all stores.
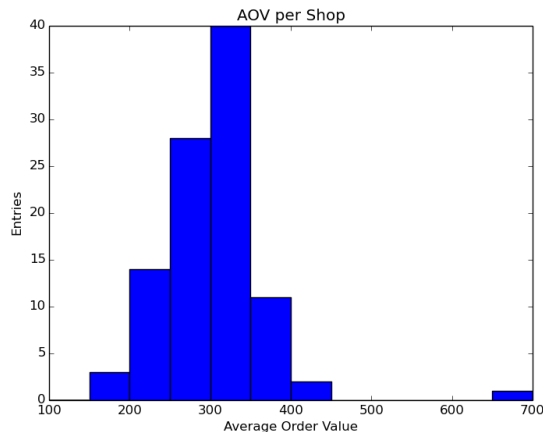
To see whether the AOV was dependent on who was buying the shoes, rather than the store selling them, I also plotted the AOV per customer (assuming that the user_id is consistent across stores). A similar trend emerges, where customers are most likely to spend between $250 - $350, and unlikely to spend more than $400 in a single transaction.
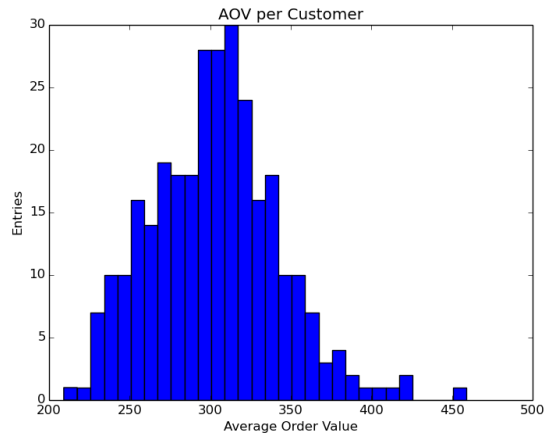
Finally, I calculated the AOVs separately for new vs returning customers, and found them to be $297.68 and $302.90 respectively. When compared to the cost of a pair of shoes, the difference in these two numbers is negligible, and so it seems that the AOV is independent of whether a customer has previously bought shoes on Shopify.

Question 2:

a.   How many orders were shipped by Speedy Express in total?

```
SELECT Shippers.ShipperName,
          COUNT(ShipperName) as Number_of_orders
FROM Shippers
JOIN Orders
ON Shippers.ShipperID = Orders.ShipperID
WHERE Shippers.ShipperName = 'Speedy Express'
GROUP BY ShipperName
```

Speedy Express shipped 54 orders.

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

```
SELECT Employees.LastName,
       Employees.FirstName,
       COUNT(Orders.OrderID) as Number_of_Orders
FROM Orders
JOIN Employees
ON Orders.EmployeeID = Employees.EmployeeID
GROUP BY Employees.LastName, Employees.FirstName
ORDER BY Number_of_Orders DESC
```

The last name of the employee with the most orders is Peacock

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

```
SELECT details.ProductID,
       details.Quantity,
       details.ProductName,
       COUNT(details.Quantity) as Number_of_orders,
       SUM(details.Quantity) as total_quantity
FROM (SELECT Orders.OrderID
       FROM Orders
       JOIN Customers
       ON Orders.CustomerID = Customers.CustomerID
       WHERE Customers.Country = 'Germany'
    ) german_orders
LEFT JOIN (Select Orderdetails.OrderID,
          Orderdetails.ProductID,
          Orderdetails.Quantity,
          Products.ProductID,
          Products.ProductName
          FROM Orderdetails
          JOIN Products
          ON Orderdetails.ProductID = Products.ProductID) details
ON german_orders.OrderID = details.OrderID
GROUP BY ProductID
ORDER BY total_quantity DESC
```

 The product that was ordered the most was Boston Crab Meat, with 160 units shipped over 4
orders. However, the product with the most orders, regardless of total quantity sold, was
Gorgonzola Telino with 5 orders placed.