

Ikwuonwu Jonathan Ukaegbu Shopify Fall 2022 Data Science Challenge

Question 1:

A. Think about what could be going wrong with our calculation. Think about a better way to evaluate this data.

- What went wrong here was that there were some outliers (values that were not within the range of the majority of values). From the dataset we observe, each order does not only have 1 item purchased. The issue is that we are judging sneaker affordability based off of total order amount, which is incorrect.
- The incorrect calculation for the average order mean is found by dividing by the total number of rows. The AOV is typically calculated by dividing the sum of sales (5725640) by the sum of items sold (43936). This would give us the actual mean value 357.92 . In addition, To get a holistic view, evaluating other metrics such as the mode, mean and median.

B. What metric would you report for this dataset?

- it is often incorrect to judge and analyze the dataset based off of the average. For instance, any majorly large number or outliers will skew the average. Hence, we can consider other metrics such as the median and mode. For this case I would report the **mode**. The mode lets us know which value(s) in a dataset is the most common. From the table below we see that most order amounts fall below the 1000 mark but there is an outlier with a value of 704000.



C. What is its value?

- Using the jupyter notebook to evaluate the metrics:
`mode = df['order_amount'].mode()[0]`
Mode
--Result = \$153.00

Question 2:

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

```
SELECT DISTINCT COUNT(OrderID)
FROM Orders
WHERE ShipperID = (SELECT ShipperID FROM Shippers WHERE ShipperName = 'Speedy Express');
-- Answer: COUNT(OrderID)= 54
```

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

```
SELECT emp.LastName, COUNT(emp.LastName) AS Occurrences
FROM Orders AS ord
JOIN Employees emp ON emp.EmployeeID = ord.EmployeeID
GROUP BY emp.LastName
ORDER BY Occurrences DESC
Limit 1
-- Answer: Peacock | 40 orders
```

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

```
SELECT prod.ProductName
FROM Products AS prod
JOIN OrderDetails AS details ON details.ProductID = prod.ProductID
JOIN Orders AS ord ON ord.OrderID = details.OrderID
JOIN Customers AS cust ON cust.CustomerID = ord.CustomerID
WHERE cust.Country = "Germany"
GROUP BY prod.ProductName
ORDER BY SUM(details.Quantity) DESC
LIMIT 1;
-- Answer: Boston Crab Meat
```