

Stephen Zhu

Shopify Data Science Intern Challenge

Question 1

- a. Looking at the data, user 607 made 17 consistent purchases of 704,000 dollars, and 200 sneakers, in the 30-day timeframe. These purchases were all made at 4:00am on the dot, on different days. I would suspect that this was a way for business customers to purchase stock, or for a large sneaker collection to snatch up the hottest selections. If I were to obtain the average of the same 30-day window, I would exclude those abnormal transactions.
- b. It looks like there are a couple of very expensive orders of items. With this in mind, I would report the total number of items ordered and their average prices, instead of the average order size, since the 17 most expensive orders have 2000 items.
- c. The average price per item of small orders is 375.92\$, and 43936 items were sold.

Question 2

a.

```
SELECT * FROM [Orders]
WHERE ShipperID ==1
```

There were 54 orders shipped by Speedy Express.

b.

```
SELECT *, COUNT(*) as numorder
FROM Orders
INNER JOIN Employees ON Orders.EmployeeID=Employees.EmployeeID
GROUP BY Orders.EmployeeID
ORDER BY numorder DESC
```

Margaret Peacock is the employee with the most orders (40).

c.

```
SELECT p.ProductID, p.ProductName, sum(d.Quantity) as Quantity
FROM Products as p
INNER JOIN OrderDetails as d on d.ProductID=p.ProductID
INNER JOIN Orders as o on o.OrderID = d.OrderID
INNER JOIN Customers as c on c.CustomerID = o.CustomerID
WHERE c.Country = 'Germany'
GROUP BY d.ProductID
ORDER BY Quantity DESC
```

The most ordered product from customers in Germany is 'Boston Crab Meat', with an order quantity of 160.